A bioinformatics approach to building an otic gene regulatory network

Anwar, Maryam

Awarding institution:
King's College London

A bioinformatics approach to building an otic gene regulatory network

by

Maryam Anwar

A thesis submitted to the King's College London's Higher Degree Office in partial fulfilment for the Degree of

Doctor of Philosophy

Department of Craniofacial Development and Stem Cell Biology
King's College London
London, UK, SE1 9RT

February 2016

Maryam Anwar, London. UK

Abstract

During development, the coordinated and sequential action of signals and regulatory factors controls how cells become different from each other and acquire specific fates. This information can be integrated in gene regulatory networks (GRNs) that model these processes over time and consider temporal and spatial changes of gene expression and how these are regulated.

During early development, vertebrate sensory organs arise from the pre-placodal region at the border of the neural plate. Subsequently, FGF signalling plays a crucial role in inducing otic-epibranchial progenitors that ultimately give rise to the otic and epibranchial placodes. Downstream of FGF signalling, many transcription factors are activated. However, their regulatory relationships are not very clear. This project uses a bioinformatics approach to establish a GRN to model how multipotent progenitors transit through sequential regulatory states until they are committed to the ear lineage.

To this end, using systematic perturbation experiments, new ear-specific genes have been identified some of which respond early to FGF. Focussing on these early genes, I have used phylogenetic footprinting combined with histone ChIP-seq to identify novel enhancers. Subsequently, I have investigated transcription factor binding sites within these enhancers to identify a small group of common regulators. In parallel, using mRNA-seq and perturbation data, I have reverse-engineered GRNs that recapitulate known interactions and predict new ones. Using a combination of these approaches, I have ultimately enriched a preliminary literature-based GRN by placing otic genes and their interactions into a hierarchy. Thus, this network is a resource for identifying key otic regulators and their targets and provides guidelines for future experiments.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Four years in the Streit lab...it's been a long eventful journey, full of learning, life experiences and experimentation, forging new friendships, waging wars against sleep (mostly victorious, but sometimes lost) and full blown crazy moments. The good, the bad and everything in between would not have been possible without the unwavering support and help of some of these great people that I am lucky to have in my life.

I would like to thank my supervisor, Prof. Andrea Streit, for giving me the opportunity to work with her. This would not have been possible without her constant guidance and her willingness to devote her time and efforts. I would also like to thank my second supervisor, Dr. Reiner Schulz, for his expert advice on bioinformatics. A special thanks to Prof. Claudio Stern and members of his lab for great scientific discussions and advice. A special mention to Ramya Ranganathan and Monica Tambalo for showing me that good friends are hard to come by. Four years we stuck together, discussed science, helped each other and put up with each other's meltdowns, I could not have asked for better work partners. A big thank you to Jingchen for countless useful discussions on analyzing ChIP-seq data!!! I would also like to thank Mark Hintze, Mohi Ahmed, Ravindra Prajapati, Alice Gervasoni and Anneliese Norris for their support, advice and friendship. It was great working with you! Grateful to Ewa Kolano-Merlin and Fiorella Guerra for technical assistance.

I would like to thank my parents, my sister, the Burneys in Chicago and the Mirzas in Pakistan for their constant support, encouragement, prayers and love. A very special thanks to my husband for always being there and for having faith in me even in those times when I would be willing to give up, helping me out when I got side-tracked and cheering me up with jokes after a monotonous work day. Without his unwavering motivation and presence, it would not have been possible.

1. Introduction

In this study, the domestic chicken (*Gallus gallus*) has been used as a model organism to study the early stages of development of the inner ear. The molecular and biological processes of organ formation in chick are comparable to mammals which makes it a good model to study development.

This project focuses on the early stages of development of the inner ear. The inner ear which is responsible for hearing and balance has a complex structure consisting of cochlea for sound perception and vestibular sensory organs (semi-circular canals and vestibule) for balance (reviewed in: Forge and Wright, 2002, Kelley, 2006). The inner ear is composed of different types of cells such as hair cells that convert external auditory input into electrical signals, sensory neurons that innervate the hair cells and transmit information to vestibular and auditory neurons in the brain stem, supporting cells which act as scaffold for the sensory epithelium and endolymph-secreting cells that maintain the ionic environment for proper transmission of information.

The entire inner ear forms from the otic placode, which is one of the sensory placodes in the vertebrate head. Sensory placodes are also called cranial placodes and these give rise to a diverse group of structures such as the olfactory epithelium, inner ear, lens of the eye and distal portions of the cranial sensory ganglia (reviewed in Kelley, 2006, Streit, 2008, Schlosser, 2010). Despite the differences in their structure, all sensory placodes share common features at early stages of development and originate from a common domain that surrounds the neural plate, a territory called the pre-placodal region (PPR) (reviewed in Schlosser, 2006, Streit, 2007, Streit, 2008, Schlosser, 2010, Patthey et al., 2014). At Hamburger and Hamilton stage 6 (HH6) (Hamburger and Hamilton, 1951), the PPR is a

mixture of precursors for neural crest, neural plate, placodes and epidermis, which gradually become distinct from each other as development proceeds (Kozlowski et al., 1997, Streit, 2002, Streit, 2008, Bhattacharyya and Bronner-Fraser, 2008, Xu et al., 2008, Schlosser, 2010). The placode precursors express a common set of genes namely *Six1*, *Six4*, *Eya1* and *Eya2* (Mishima and Tomarev, 1998, Sahly et al., 1999, Esteve and Bovolenta, 1999, Pandur and Moody, 2000, Kobayashi et al., 2000, Schlosser and Ahrens, 2004, Bessarab et al., 2004, Ahrens and Schlosser, 2005, Sato et al., 2010, Pieper et al., 2011). At HH7, when somitogenesis starts, neural crest cells begin to separate from the placode precursor cells (Kozlowski et al., 1997, Streit, 2002) and contribute to e.g. melanocytes, craniofacial cartilage and bone, peripheral nervous system and glia (for review see: Sauka-Spengler and Bronner-Fraser, 2008, Betancur et al., 2010a). The precursors for different sensory placodes begin to separate from each other around HH8 and become fully restricted by HH15 (Streit, 2002, Bhattacharyya et al., 2004, Xu et al., 2008, Schlosser, 2010, Pieper et al., 2011). For this thesis, the main interest is to study specification of the otic placode which becomes morphologically visible at HH10, anterior to the first somite at the level of rhombomeres 5 and 6 (reviewed in: Baker and Bronner-Fraser, 2001, Streit, 2007, Schlosser, 2010). In the next sections, I will discuss the developmental events leading from PPR to the otic placode.

1.1 The pre-placodal region (PPR)

As described earlier, all sensory placodes originate from the ectoderm that surrounds the neural plate in a horse-shoe shaped territory called the pre-placodal region (PPR) (reviewed in Schlosser, 2006, Streit, 2007, Streit, 2008, Schlosser, 2010, Patthey et al., 2014) (Figure 1.1. A). At this stage, sensory placode precursors share similar features but ultimately differentiate to form distinct placodes (Whitlock and Westerfield, 2000, Streit, 2002, Xu et al., 2008, Bhattacharyya and Bronner, 2013). The first ever model of the PPR

was proposed by Jacobson in the 1960s (Jacobson, 1963a, Jacobson, 1963b, Jacobson, 1963c) after he performed a series of embryological experiments in Newt and identified a horse-shoe shaped structure at the border of the neural plate as a domain that could generate sensory placodes. Jacobson showed that rotating the domain along the rostro-caudal axis at early neural stages causes the precursor cells to acquire placode fate according to their new location. However, when the same experiment is done at late neural stages, the precursor cells are already specified to give rise to a particular placode and develop according to their original fate irrespective of their new location (Jacobson, 1963c). These experiments revealed that placode progenitors are initially competent to give rise to any placode and thus may share common features, but also pointed to the time when they become restricted to particular fates. Subsequent studies in amphibians and fish provided morphological evidence for the PPR as a visible thickening of the epithelium at the border of the neural plate (Verwoerd and van Oostrom, 1979, Miyake, 1997), whereas in chick, axolotl and frog, this thickening is absent (Couly and Le Douarin, 1985, Northcutt and Brandle, 1995, Schlosser and Northcutt, 2000). However, fate map studies indicate a corresponding location for placode precursors surrounding the neural plate (Hatada and Stern, 1994, Kozlowski et al., 1997, Whitlock and Westerfield, 2000, Streit, 2002, Bhattacharyya et al., 2004, Xu et al., 2008). Thus, the PPR is defined by morphology in some species, gene expression (Six and Eya family members) and the location of placode progenitors.

While these observations together with Jacobson's transplantation experiments provide good support for the idea that PPR cells initially share common properties, later experiments in the chick confirmed this. Asking the question whether precursors for different placodes are already specified at PPR stages, Bailey and colleagues conducted specification assays (Bailey et al., 2006). When different regions of the PPR are cultured

in isolation, they do not adopt otic, trigeminal or olfactory fate, but instead initiate the 'lens programme' indicated by gene expression similar to normal lens development in the embryo (Bailey et al., 2006). PPR explants begin to express the lens markers *Pax6*, *L-Maf*, *Foxc1* as well as the differentiation genes $\alpha-$ and *δ-crystallin* and acquire a morphology similar to a lens vesicle with elongated cells (Bailey et al., 2006). Thus, all placode progenitors irrespective of their later fate are initially specified as lens. How is lens fate restricted to the actual lens? Differential signalling is required to suppress lens specification and to promote the development of placodes other than lens (Bailey et al., 2006). FGF signalling appears to play an important role (Bailey et al., 2006, Lleras-Forero et al., 2013). For example, FGF signalling from the anterior neural ridge promotes olfactory identity, while in the posterior PPR it represses lens fate and leads to the development of the otic-epibranchial progenitor domain, which ultimately gives rise to otic and epibranchial placodes (Martin and Groves, 2006, Lleras-Forero et al., 2013, Tambalo, 2015). In summary, these findings suggest that indeed placode progenitors not only come from a common domain, but also initially share common properties.

What is the functional significance of the PPR? The importance of the "PPR-state" as a vital step towards placode formation is evident from the literature. Misexpression studies of placode specific transcription factors suggest that indeed only the PPR is competent to give rise to placodes. Over expression of *Sox3*, *Six3* and *Pax6* in frog, zebrafish and mouse leads to ectopic lenses, while misexpression of *Sall4* generates ectopic otic vesicles (Oliver et al., 1996, Altmann et al., 1997, Chow et al., 1999, Koster et al., 2000, Lagutin et al., 2001, Barembaum and Bronner-Fraser, 2007). However, this only occurs closely associated with the neural tube, presumably within the PPR. More direct proof that cells must go through a 'PPR-state' before they become placodes comes from experiments in the chick: the anterior epiblast of primitive streak stage embryos does not

give rise to placodes and when treated with the otic inducing signal FGF2 does not turn on otic makers (Martin and Groves, 2006). However, if transplanted into the PPR it begins to express the PPR markers *Dlx5/6* and *Eya.* When this tissue is now cultured with FGF2 it adopts an otic fate. These experiments demonstrate that in order to respond to placode inducing signals, cells must first have acquired PPR identity. Together, these observations show that the formation of placodes is at least a two-step process where cells initially adopt PPR identity and only PPR cells later become placodes when exposed to appropriate signals.

As described earlier, PPR cells express *Six1*, *Six4* and *Eya2* (chick) and *Eya1* (frog and zebrafish) (Mishima and Tomarev, 1998, Esteve and Bovolenta, 1999, Sahly et al., 1999, Kobayashi et al., 2000, Pandur and Moody, 2000, Bessarab et al., 2004, Schlosser and Ahrens, 2004, Ahrens and Schlosser, 2005, Ishihara et al., 2008a). These factors continue to express in the placodes, except for the lens (Esteve and Bovolenta, 1999, Kobayashi et al., 2000, Ghanbari et al., 2001, Bessarab et al., 2004, Schlosser and Ahrens, 2004). The importance of *Six* and *Eya* factors in the PPR is indicated by the fact that if these are knocked down or mutated, the development of all sense organs is affected (Xu et al., 1999, Laclef et al., 2003, Zheng et al., 2003, Ozaki et al., 2004, Brugmann et al., 2004, Friedman et al., 2005, Kozlowski et al., 2005, Konishi et al., 2006, Ikeda et al., 2007, Christophorou et al., 2009). In chick and frog, repression of all *Six1* target genes using a constitutive repressor form (Six1-EnR) leads to the loss of PPR markers (Christophorou et al., 2009), while misexpression of *Six1* or *Six1* and *Eya2* induces ectopic PPR gene expression, while repressing neural and neural crest genes (Christophorou et al., 2009). In addition, activation of *Six1* target genes is required for the expression of the olfactory-lens progenitor marker *Pax6*, the otic epibranchial marker *Pax2* and the ophthalmic trigeminal

marker *Pax3* (Christophorou et al., 2009). Thus, the *Six* and *Eya* cassette seems to play an important role in conferring PPR identity to ectodermal cells.

Moreover, various mouse mutants indicate a crucial role of these factors in all sensory placodes and their derivatives. In *Six1* knockout mice, the development of the inner ear and nose is severely affected along with craniofacial defects, missing thymus and a missing kidney (Laclef et al., 2003, Zheng et al., 2003, Ozaki et al., 2004, Ikeda et al., 2007). Likewise, loss of *Eya1* function in mouse affects the epibranchial placode, hence the cranial sensory ganglia fail to develop and apoptosis is increased (Zou et al., 2004). Finally, in humans, mutations in *Six1* and *Eya1* have been associated with Branchio-Oto-Renal syndrome (BOR) characterized by hearing loss and branchial and renal defects (Abdelhak et al., 1997, Johnson et al., 1999, Ruf et al., 2004, Zhang et al., 2004, Krug et al., 2011, Song et al., 2013). Together, these observations highlight the importance of *Six* and *Eya* genes for normal development of all sensory placodes.

As development proceeds, the PPR starts to be regionalized. The first genes to be regionally restricted include the homeobox transcription factors *Otx2* and *Gbx2* (Bally-Cuif et al., 1995, von Bubnoff et al., 1996, Acampora et al., 2001, Tour et al., 2001). Already at primitive streak stages these genes are broadly expressed in partially overlapping domains and over time become restricted. Within the PPR *Otx2* is expressed in the anterior PPR, while *Gbx2* is present in the posterior PPR (Simeone et al., 1993, Bally-Cuif et al., 1995, von Bubnoff et al., 1996, Acampora et al., 2001, Tour et al., 2001, Li et al., 2009, Steventon et al., 2012). *Gbx2* and *Otx2* have been implicated in rostro-caudal patterning of the ectoderm by mutually repressing each other in the PPR and in the neural tube (Wassarman et al., 1997, Hidalgo-Sanchez et al., 1999, Millet et al., 1999, Broccoli et al., 1999, Katahira et al., 2000, Joyner et al., 2000, Castro et al., 2006,

Steventon et al., 2012). As *Otx2* is expressed in the anterior PPR, activation of its targets is required for the formation of anterior placodes including lens, trigeminal and olfactory placode. *Gbx2*, on the other hand, is necessary for posterior placode formation: the otic and epibranchial placodes (Steventon et al., 2012). Over time the PPR becomes further regionalized ultimately leading to the formation of distinct placodes expressing sets of genes characteristic for each (reviewed in Schlosser, 2006, Schlosser, 2010). Some of the earliest genes correspond to different parts of the PPR and include the members of the paired-box family of transcription factors where *Pax6* encompasses olfactory, lens and trigeminal precursors, *Pax2* corresponds to future otic-epibranchial cells and *Pax3* to the ophthalmic portion of the trigeminal placode (Li et al., 1994, Zygar et al., 1998, Bailey et al., 2006, Canning et al., 2008, McCarroll et al., 2012). In the next sections, I will discuss the events leading to placode segregation with a particular interest in the formation of the otic placode.

1.2 The cranial sensory placodes

Following the specification of the PPR and its regionalization, sensory placodes appear as regions of epithelial thickening beside the neural tube (Figure 1.1 B). The location of each placode has been determined from extensive fate map studies (Rawles, 1936, D'Amico-Martel and Noden, 1980, D'Amico-Martel, 1982, Couly and Le Douarin, 1985, Tam, 1989, Eagleson et al., 1995, Kozlowski et al., 1997, Whitlock and Westerfield, 2000, Streit, 2002, Xu et al., 2008, Modrell et al., 2014).

The olfactory, lens and otic placodes give rise to sensory organs: the olfactory epithelium, the lens of the eye and the inner ear (Figure 1.1 C). The other placodes are neurogenic patches that produce sensory ganglia. The epibranchial placodes (geniculate, petrosal and nodose) give rise to the distal portions of the VI, IX and X cranial ganglia, while the

anterior and posterior lateral line placodes give rise to the lateral line system in the head and along the body axis, and the trigeminal placode (maxillomandibular and ophthalmic regions) to the trigeminal ganglion (Figure 1.1 C). After epithelial thickening, olfactory, lens and otic placodes invaginate and form vesicles or cavities, and otic and olfactory epithelia undergo extensive morphogenesis to form complex structures (for review see: Baker and Bronner-Fraser, 2001, Schlosser, 2006, Streit, 2008, Schlosser, 2010). The olfactory placode is found next to the future olfactory bulb and gives rise to the epithelial lining of the nasal cavity. It produces olfactory sensory neurons, supporting cells and a population of stem cells that renew sensory neurons throughout life, as well as migratory neurons like GnRH or Calbindin^{+} that migrate from the placode into the central nervous system (Schlosser, 2010). The lens is a non-neurogenic placode, gives rise to the lens of the eye and is located next to the optic vesicles (Schlosser, 2010). It invaginates, separates from the surface ectoderm and forms only two cell types: lens fibre cells and lens epithelial cells, which generate lens fibres throughout life. The otic placode located next to rhombomeres 4 and 5 of the hindbrain, also invaginates, and forms the otic vesicle which subsequently generates the entire inner ear. This comprises the complex architecture of the vestibular portion with the semicircular canals and their associated sensory patches, the cochlear portion with the cochlea or cochlear duct, and finally the neurons that innervate the ear, the cochlear-vestibular ganglion (Torres and Giraldez, 1998). Formation of the otic placode will be discussed in detail below.

In contrast, epibranchial and trigeminal placodes are neurogenic. Neuronal precursors delaminate from the placodes, migrate away and coalesce to form ganglia (reviewed in: Baker and Bronner-Fraser, 2001, Schlosser, 2006, Streit, 2008, Schlosser, 2010). The trigeminal placode is situated next to the midbrain and consists of ophthalmic placode, which gives rise to the ophthalmic ganglion that innervates the eyeball, eye muscles,

conjunctiva, nose and skin of the head. The maxillomandibular region gives rise to the maxillar nerve that innervates the upper teeth, palate and pharynx and the mandibular nerve that innervates the lower teeth, gums, floor of the oral cavity and mucosa of the tongue (reviewed in Baker and Bronner-Fraser, 2001). The epibranchial placodes (geniculate, petrosal and nodose) lie in the lateral ectoderm at the level of the hindbrain (Figure 1.1 B). The geniculate placode gives rise to the distal parts of the VIIth cranial nerve and innervates taste buds. In birds, it contributes to the paratympanic organ. The petrosal placode gives rise to the glossopharyngeal ganglion and the distal part of IXth cranial nerve. Finally, the nodose placode gives rise to the nodose ganglion and the distal part of the IXth cranial nerve (Figure 1.1 B-C).

Thus, placodes form a diverse set of structures and cell types, yet because of their origin from the PPR, their placodal morphology and their contribution to sensory structures share certain features. Because this thesis focuses on the earliest steps of inner ear formation, in the next sections I will discuss the development of the otic-epibranchial progenitor domain (OEPD).

1.3 The Otic-Epibranchial Progenitor Domain (OEPD)

Following the beginning of PPR regionalization, local signals induce the posterior PPR cells to form the otic-epibranchial progenitor domain next to the hindbrain. This domain ultimately gives rise to otic and epibranchial placodes (Figure 1.1 B). At the OEPD stage, progenitor cells for otic and epibranchial placodes are intermingled with each other and with epidermal and possibly neural crest cells (for review see (Ladher et al., 2010)). Some of the first markers of the OEPD are the paired-box transcription factors *Pax2* and *Pax8* (Nornes et al., 1990, Krauss et al., 1991, Pfeffer et al., 1998, Terzic et al., 1998, Heller and Brandli, 1999, Hutson et al., 1999, Streit, 2002, Li et al., 2004, Burton et al., 2004,

Hans et al., 2004, Mackereth et al., 2005, Sanchez-Calderon et al., 2005, Christophorou et al., 2010, McCarroll et al., 2012). The chromosomal region containing *Pax8* locus has undergone considerable rearrangements during evolution (Fan et al., 2002). In amphibians, medaka and stickleback, the *Pax8*-containing region corresponds to that in mammals whereas zebrafish *Pax8* locus shows no synteny with this region and in birds and reptiles, the entire region is missing (Christophorou et al., 2010).

FGF signalling from the mesoderm underlying the future otic region plays an important role in the OEPD induction. It is triggered when FGF ligands bind to the extracellular domain of tyrosine kinase receptors leading to receptor dimerization and trans-phosphorylation of the tyrosine in the intracellular domain of the receptor (Ornitz et al., 1995). The signal transduction cascade works either through MAP kinase activation, mostly thought to be involved in proliferation and differentiation, Atk which seems to mediate cell survival or protein kinase C, which is involved in morphogenesis and cell migration (reviewed in Dorey and Amaya, 2010). The MAP kinase pathway plays a fundamental role in otic induction as its inhibition leads to loss of OEPD markers and consequently of otic placode formation (Yang et al., 2013). This indicates that FGF signalling is crucial for otic induction although in different species, different FGF members are involved (for review see: Schimmang, 2007).

In frog, *FGF3* and *FGF8* are expressed in the hindbrain and paraxial mesoderm respectively (Lombardo et al., 1998, Fletcher et al., 2006). *FGF8* is also later expressed in the future midbrain-hindbrain boundary, otic placode and pharyngeal arches (Fletcher et al., 2006).

In birds, *FGF3* and *FGF19* are expressed in the cranial paraxial mesoderm and later in the hindbrain (Mahmood et al., 1995, Ladher et al., 2000, Karabagli et al., 2002, Kil et al., 2005), while *FGF8* is expressed in the endoderm underlying the cranial mesoderm at early stages and in the pharyngeal endoderm at somite stage 5 (Karabagli et al., 2002, Stolte et al., 2002, Ladher et al., 2005). *FGF3* and *FGF19* appear in the pharyngeal endoderm from somite stages 5 and 6 respectively (Mahmood et al., 1995). OEPD cells themselves express different FGF receptors with the highest occurrence of *FGFR1* (Walshe and Mason, 2000, Nishita et al., 2011) and also show expression of ERK/MAP kinase responsive genes such as *Etv4/5* indicating that FGF is indeed active in the OEPD as it becomes induced (Lunn et al., 2007).

In mouse, *FGF3* is expressed in the PPR from somite stage 3 (McKay et al., 1996, Alvarez et al., 2003, Wright and Mansour, 2003), while *FGF8* is initially expressed in the PPR, the mesoderm and the endoderm at somite stage 0 (Crossley and Martin, 1995, Ladher et al., 2005) and later in the pharyngeal endoderm (Crossley and Martin, 1995). Finally, *FGF10* is expressed in the anterior and ventral mesoderm between 0 and 4 somite stages and later in the hindbrain (Wright and Mansour, 2003).

In zebrafish, *FGF3* is initially present in the pre-placodal ectoderm at 75% epiboly and later in the cranial paraxial mesoderm (Phillips et al., 2001, Leger and Brand, 2002, Maroon et al., 2002, Liu et al., 2003). Apart from being expressed in the mesoderm at 80% epiboly, *FGF8* is also expressed in the hindbrain (Phillips et al., 2001, Walshe et al., 2002), while *FGF10b* is found in the cranial paraxial mesoderm underneath the epibranchial placode at tail bud stage (Maulding et al., 2014).

While these expression studies support a potential role for FGF signalling in OEPD induction, functional experiments provide evidence that this pathway is crucial for formation of the otic placode. In chick, the first demonstration for the role of FGF signalling came from experiments where future otic ectoderm was exposed to *FGF19* (initially expressed in the underlying mesoderm) in vitro; as a result, cells turned on otic markers like *Pax2* and *FGF3* (Ladher et al., 2000). Subsequently, it was shown that inhibition of FGF signalling by SU5402, an inhibitor of FGF receptors (Mohammadi et al., 1997) results in the loss of OEPD markers and subsequent formation of otic placode (Martin and Groves, 2006, Yang et al., 2013). Moreover, *FGF3* (from the neural tube) knockdown using siRNA does not stop otic placode formation, but subsequent development into an otic vesicle (Zelarayan et al., 2007). On the other hand, *FGF3* misexpression induces ectopic otic vesicles (Vendrell et al., 2000, Zelarayan et al., 2007). Finally, knockdown of endodermal *FGF8* using siRNA results in loss of *Pax2* and OEPD induction (Ladher et al., 2005). However, *FGF8* only affects OEPD formation indirectly: it is sufficient and required for *FGF19* expression in the mesoderm, which in turn induces OEPD markers in the overlying ectoderm (Ladher et al., 2005, reviewed in Ladher et al., 2010). Thus, a cascade of FGF signalling is responsible for OEPD induction in the chick.

In mouse, *FGF3* and *FGF10* loss of function results in the formation of small otic vesicles (Ohuchi et al., 2000a, Pauley et al., 2003, Wright and Mansour, 2003), while inactivation of both FGFs leads to the complete absence of the placode. Placodal markers like *Pax2* and *Dlx5* are never expressed. *FGF8* homozygous mice die at early embryonic stages (Meyers et al., 1998), however in animals that are null for *FGF3* and hypomorphic for *FGF8,* otic placodes do not develop suggesting a similar signalling loop as described in chick (Ladher et al., 2005). In humans, *FGF3* plays a significant role in inner ear

development, where homozygous mutations in *FGF3* can cause syndromic deafness and lack of development of the inner ear (Tekin et al., 2007).

In zebrafish, *FGF10b* knockdown does not affect otic induction but impairs accumulation of otic cells (Maulding et al., 2014), whereas simultaneous *FGF3* and *FGF8* loss-of-function results in the loss of otic markers like *Pax2* and *Pax8* (Phillips et al., 2001, Leger and Brand, 2002, Liu et al., 2003). Additionally, inhibition of FGF receptors using SU5402 causes the loss of OEPD markers *Pax2, Dlx3, Pax8* and *Spry4* (Maroon et al., 2002, Leger and Brand, 2002, Solomon et al., 2004). On the other hand, gain-of-function experiments in zebrafish also point to a key role of FGF in otic induction. Misexpression of *FGF3* and *FGF8* causes an expansion of the OEPD leading to ectopic otic vesicles (Phillips et al., 2004, Padanad et al., 2012). Similarly, treatment with retinoic acid leads to OEPD expansion through the expansion of *FGF3* and *FGF8* expression, which ultimately generates ectopic otic vesicles (Phillips et al., 2001).

In summary, these findings demonstrate a fundamental role for FGF signalling in OEPD induction in mammals, birds and fish and a recent model summarises its activity (reviewed in Ladher et al., 2010). Initially, FGF from the endoderm induces FGF in the overlying mesoderm, which in turn induces the expression of other FGF and WNT ligands (see below) in the hindbrain and OEPD specific transcripts in the overlying ectoderm. The next section describes how otic and epibranchial precursors segregate after OEPD induction.

1.4 The otic placode and its segregation from the epibranchial placode

While FGF plays an important role in OEPD induction, FGF signalling must be reduced or inactive for cells to progress towards otic fate. Simultaneously, canonical WNT

signalling is required to complete otic placode formation (Litsiou et al., 2005, Ohyama et al., 2006, Park and Saint-Jeannet, 2008). As otic and epibranchial precursors segregate, FGFs promote the epibranchial fate by enhancing the expression of epibranchial genes such as *Phox2b* and *Sox3*, while simultaneously inhibiting late otic genes such as *Soho1* and *Nkx5.1* (Nikaido et al., 2007, Freter et al., 2008, Abello et al., 2010, McCarroll and Nechiporuk, 2013). Moreover, BMP signalling from the pharyngeal endoderm promotes epibranchial fate together with FGF (Begbie et al., 1999). Inhibition of *BMP7* in chick results in the absence of epibranchial neurons (Begbie et al., 1999). On the other hand addition of *BMP7* to ectodermal explants induces the formation of these neurons (Begbie et al., 1999). Likewise, in zebrafish, BMP signalling from the pharyngeal endoderm has also been shown to induce an epibranchial fate (Holzschuh et al., 2005). Thus, prolonged FGF signalling in OEPs favours epibranchial, but represses otic fate.

An early response to FGF signalling from the cranial mesoderm is the induction of WNT ligands in the hindbrain, and this activity is necessary for otic development (Litsiou et al., 2005). In mouse, WNT activity is observed in the medial placode close to the hindbrain as evidenced by the TCF/LacZ reporter line (Ohyama et al., 2006) consistent with a role for this pathway in promoting otic identity. In $FGF3^{-/-}$ and $FGF10^{-/-}$ mutants, *WNT8a* expression is reduced in the hindbrain and as a consequence, the embryos lack the OEPD and in turn epidermal genes are expanded (Urness et al., 2010). This suggests that indeed WNT signalling is under the control of FGF. Furthermore, loss of WNT through conditional deletion of β-catenin in Pax2+ cells causes expansion of epidermal markers (*Foxi2)* at the expense of otic fate (*Pax2, Pax8* and *Dlx5*), whereas β-catenin activation in Pax2+ cells results in expansion of the otic fate at the expense of the epidermal fate (Ohyama et al., 2006). Thus, WNT downstream of FGF promotes otic development.

Likewise chick experiments indicate an important role for WNT signalling. Ectodermal explants begin to express *Wnt8c* in the presence but not in the absence of FGF19 (Ladher et al., 2000), indicating that WNT is under the control of FGF. Moreover, activation of canonical WNT signalling leads to the inhibition of epibranchial development (*Foxi2* and *Phox2b),* while inhibition of WNT signalling prevents the expression of the otic markers *Soho1* and *Nkx5.1* (Freter et al., 2008). However, it does not compromise the expression of OEPD genes *Pax2* and *Foxi3* (Freter et al., 2008). This suggests that WNT activity is required after OEPD induction. Finally, WNT stimulation in zebrafish leads to high levels of *Pax2a* which is a characteristic feature of otic cells (McCarroll et al., 2012). Similarly, hindbrain-derived WNT promotes otic fate in frog embryos (Park and Saint-Jeannet, 2008).

Thus, in all vertebrate species examined so far the canonical WNT pathway appears to be under the control of FGF signalling, and together they promote otic fate at the expense of epibranchial identity. However, prolonged exposure to FGFs inhibits otic development, while simultaneously promoting epibranchial character. Thus, the balance between FGF and WNT signalling is critical to determine the identity of OEPD cells.

Apart from FGF and WNT, Notch also plays a role in otic induction (Jayasena et al., 2008). In chick and mouse, Notch effectors such as *Hes1*, *Hes5* and *Jag1* are expressed in the otic region and are induced by WNT (Jayasena et al., 2008, Paxton et al., 2010). In mouse, activation of Notch in Pax2-expressing cells leads to the expansion of otic placode at the expense of epibranchial/epidermal fates whereas its inhibition leads to reduction of the otic placode (Jayasena et al., 2008). Moreover, conditional overexpression of Notch in the otic placode promotes WNT targets *Tcf* and *Lef1* (Jayasena et al., 2008). This indicates that WNT and Notch signalling work together to promote an otic fate. The

interplay between these two pathways appears to be a conserved mechanism in other systems such as the intestine (reviewed in Watt et al., 2008).

While the otic and epibranchial placodes share a common origin and require FGFs, shortly after OEPD specification, both develop independently. Otic precursors become independent of FGF whereas epibranchial cells require sustained FGF signalling (Nikaido et al., 2007, Sun et al., 2007, Freter et al., 2008, McCarroll and Nechiporuk, 2013). Sustained FGF signalling emanates from pharyngeal endoderm and comes into contact with the lateral portion of OEPD (reviewed in Ladher et al., 2010), thus inhibiting otic fate by downregulating *Soho1* and *Nkx5.1* (otic genes) and upregulating *Phox2b* (epibranchial gene) (Freter et al., 2008). The pharyngeal endoderm is also a source of BMPs in addition to FGFs. In zebrafish, *BMP2b* and *BMP5* are expressed in the pharyngeal endoderm and induce *Phox2b* (Holzschuh et al., 2005). Manipulation of BMP signalling has no effect on the development of trigeminal and otic placodes (Holzschuh et al., 2005). Similarly, in chick, *BMP7* is expressed in the pharyngeal endoderm; its inhibition prevents the formation of epibranchial neurons and its addition to ectodermal explants induces their formation (Begbie et al., 1999). As only a few epibranchial markers have been characterised, a precise time-resolution of differentiation between the otic and epibranchial placodes remains unclear.

To summarise, interplay of various signals (FGF, WNT and Notch) is required to segregate otic and epibranchial fates following the OEPD formation. Specifically, FGF induces WNT signalling in the hindbrain, which inhibits epibranchial but promotes otic fate together with Notch signalling. In the lateral OEPD, FGF and BMP from the endoderm work together to promote epibranchial identity.

1.5 Transcriptional control from PPR to the otic placode: A Gene regulatory network perspective

In response to the signalling events described above the expression of different transcription factors is initiated as cells adopt OEPD and later otic fate. With *Pax2* and *Pax8* being the earliest OEPD markers, many studies have concentrated on their regulation and induction. As described earlier, *Gbx2* and *Otx2* are among the earliest genes that subdivide the PPR along the rostro-caudal axis and they repress each other to form a boundary between trigeminal and epibranchial progenitors, with *Gbx2* being restricted posteriorly (Simeone et al., 1992, Acampora et al., 1995, von Bubnoff et al., 1996, Acampora et al., 2001, Li et al., 2009, Steventon et al., 2012). Additionally, members of the *Irx, Dlx* (*Dlx5/6* in chick) and *Foxi* (*Foxi3* in chick; *Foxi1* in fish) family are expressed in the posterior PPR, where otic and epibranchial precursors reside (Goriely et al., 1999, Solomon, 2003, Solomon et al., 2003b, Ohyama and Groves, 2004, Brown et al., 2005, Khatri et al., 2014). *Gbx2* plays an important role upstream of *Pax2* and *Pax8*: knockdown of *Gbx2* leads to a reduction or loss of these genes, although it is not sufficient on its own to induce them (Steventon et al., 2012). Therefore, other factors must cooperate with *Gbx2* to promote OEPD identity. One good candidate is *Foxi3*, which is expressed in the posterior PPR, the OEPD and the epibranchial and trigeminal placodes, but downregulated in the maturing otic placode (Ohyama and Groves, 2004, Khatri and Groves, 2013, Khatri et al., 2014). In chick, its knockdown results in loss of the OEPD and otic markers *Pax2* and *Foxg1* (Khatri et al., 2014). In zebrafish, loss of *Foxi1* reduces the PPR marker *Six1* and the OEPD markers *Pax2a, Dlx3b* and *Pax8* causing development of very small or absent otic vesicles (Solomon et al., 2003a, Nissen et al., 2003, Bricaud and Collazo, 2006). The identification of an otic enhancer for Six1 (Six1-21) revealed binding sites for Foxi family members, which when mutated decreased enhancer activity (Sato et al., 2012), suggesting that Foxi factors directly regulate otic

Six1 expression. Thus, in both fish and chick, *Foxi1/3* seems to play a role in regulating PPR and OEPD genes and therefore is a crucial factor for otic placode specification.

In fish and chick, *Dlx* genes regulate their own expression (Solomon and Fritz, 2002, McLarren et al., 2003, Aghaallaei et al., 2007), and zebrafish mutants and morpholino experiments indicate *Dlx3b/4b* to be upstream of *Pax2* and *Pax8* and required for otic placode formation (Solomon and Fritz, 2002, Hans et al., 2004, Solomon et al., 2004). However, as the placode forms, *Foxi1/3* expression is lost and it becomes confined to the epibranchial and trigeminal territory. Indeed, *Foxi3* is required for epibranchial fate as its downregulation results in the loss of the epibranchial gene *Sox3* (Sun et al., 2007). The dynamic changes in *Foxi1/3* expression during this process are not well understood, however a regulatory interaction between *Foxi* and *Dlx* genes is possible. *Foxi1/3* and *Dlx3b/5* mutually promote each other (Solomon et al., 2003b, Solomon, 2003, Pieper et al., 2011, Khatri et al., 2014). While *Dlx* genes remain expressed in the otic region (Pera and Kessel, 1999, Brown et al., 2005, Khudyakov and Bronner-Fraser, 2009), *Foxi* is later expressed in the epibranchial placode (Khatri and Groves, 2013, Khatri et al., 2014). Thus it is likely that FGF signalling acts through *Foxi* and *Dlx* genes which work together to specify the OEPD, however, later both play different roles and lead to segregation of otic and epibranchial placodes.

The OEPD formation is characterized by the expression of *Pax2* and *Pax8* (Torres et al., 1996, Burton et al., 2004, Nechiporuk et al., 2007, Christophorou et al., 2010, Bouchard et al., 2010). *Pax8* is absent in chick, but expressed before *Pax2* in frog, fish and mouse (Pfeffer et al., 1998, Heller and Brandli, 1999, Christophorou et al., 2010, Freter et al., 2012). In fish, loss of *Pax8* causes reduction of *Pax2* (Hans et al., 2004, Mackereth et al., 2005) whereas in mouse, *Pax8* deficient mice do not show an otic phenotype (Bouchard

et al., 2010). In contrast, in the absence of *Pax2* function, the otic vesicle develops normally, but the cochlea, which arises from its anterior-ventral portion, is defective (Burton et al., 2004). However, *Pax2/Pax8* double mutants arrest development after the otic vesicle stage (Bouchard et al., 2010) suggesting that they compensate for each other at least in part. In humans, mutations in *Pax2* lead to sensorineuronal deafness (Sanyanusin et al., 1995a, Sanyanusin et al., 1995b, Favor et al., 1996, Schimmenti et al., 1997) suggesting that it plays an important role in ear formation. In chick, loss of *Pax2* causes reduction of the otic genes *Eya1* and *Gata3* although OEPD formation is not affected (Christophorou et al., 2010). In addition, *Pax2* has been implicated in controlling proliferation of otic precursors rather than cell identity (Freter et al., 2012). Together these findings implicate *Pax2/8* as some of the key regulators of otic development: at early stages *Pax* proteins appear to control specification of otic-epibranchial progenitors, and later continue to be important for the formation of particular structures such as the cochlea.

Experiments in zebrafish indicate that levels of *Pax2* differentiate between the otic and epibranchial fates (McCarroll et al., 2012). Gain and loss-of-function experiments reveal that cells expressing high levels of *Pax2a/Pax8* contribute to the otic placode, whereas cells with low levels will form epibranchial placodes (McCarroll et al., 2012). Moreover, WNT plays a role in inducing high levels of *Pax2a/Pax8* in the OEPD (McCarroll et al., 2012) in line with the observations that WNT signalling from the hindbrain promotes otic identity.

Pax2 expression appears to be under the control of *Six* and *Eya* genes. Misexpression of a constitutive repressive form of *Six1* in chick and *Six1* knockdown in zebrafish lead to loss of *Pax2* (Bricaud and Collazo, 2006, Christophorou et al., 2009). Conversely, *Pax2b* and

Pax8 morpholino injection in zebrafish reduces *Six1* expression (Bricaud and Collazo, 2006). The *Six1* otic enhancer (Six1-21) has a *Pax* binding site which upon mutation causes decreased enhancer activity (Sato et al., 2012). This indicates that *Six1* and *Pax2* may maintain otic specificity by mutually promoting each other.

Another important gene that is essential for otic placode formation is *Sall4*, which is initially expressed in the neural plate and PPR, and later expressed in the olfactory, lens and otic placodes (Barembaum and Bronner-Fraser, 2007). In chick, *Sall4* misexpression is sufficient to induce invagination of the ectoderm and formation of an otic vesicle (Barembaum and Bronner-Fraser, 2007). Moreover, *Sall4* levels seem to be crucial as both up-regulation and down-regulation of *Sall4* results in otic defects (Barembaum and Bronner-Fraser, 2007). Additionally, *Etv4* and *Pax2* were found to directly regulate *Sall4* through binding to its otic enhancer (CR-F fragment) (Barembaum and Bronner-Fraser, 2007) indicating that *Sall4* is regulated by early PPR and OEPD genes. In mouse, *Sall4* inactivation causes conductive hearing loss (Warren et al., 2007) and in humans, mutations in the *Sall4* locus are associated with an autosomal dominant condition called Okihiro syndrome characterized by deafness and eye and kidney disorders (Kohlhase et al., 2005).

Multiple genes are co-expressed in the otic region. *Foxg1,* a telencephalon marker is expressed in the otic placode at the 10 somite stage and increases in the otic vesicle (Khatri et al., 2014). In mouse, a similar expression pattern has been described, where *Foxg1* is required for otic morphogenesis and innervation of the vestibular system (Hwang et al., 2009). Furthermore, *Foxg1* is regulated by FGF (Urness et al., 2010, Yang et al., 2013).

Among the later otic genes are *Soho1* and *Nkx5.1,* which are induced in response to WNT signalling (Freter et al., 2008). *Soho1* is one of the first genes to be expressed in the newly formed otic placode at somite stage 9-10 in the chick embryo (Deitcher et al., 1994, Kiernan et al., 1997, Freter et al., 2008). Additionally, *Soho1* and *Nkx5.1* are co-expressed in the otic vesicle in fish (Adamska et al., 2000, Adamska et al., 2001).

Sox10 is expressed in the otic placode from $HH9^{+}$ stage onwards (Betancur et al., 2010b), and a *Sox10* enhancer (*Sox*10E2) has been identified which drives reporter gene expression in the otic territory. The enhancer presents with binding sites for *Etv4, Sox8, Myb* and others (Betancur et al., 2010b, Betancur et al., 2011) and knockdown of these three transcription factors blocks enhancer activity (Betancur et al., 2011) indicating that these are direct regulators.

As the epibranchial placode separates from the otic territory, it starts to express unique makers including *Sox3, Phox2a* and *Phox2b* along with *Foxi3*. *Phox2a* is expressed in epibranchial neuroblasts before and during delamination and *Phox2b* is expressed in delaminating neuroblasts (Begbie, 2002). In *Phox2a* homozygous mutant mice, *Phox2b* expression is lost in epibranchial neuroblasts affecting their specification (Pattyn et al., 1997, Pattyn et al., 1999, Pattyn et al., 2000), placing *Phox2b* downstream of *Phox2a*. Thus together they play a role in neuronal differentiation. In chick, *Sox3* is initially expressed in the lateral OEPD at somite stage 10 and is later confined to the epibranchial territory (Abu-Elmagd et al., 2001). When epibranchial cells initiate neurogenesis, *Sox3* expression is lost and instead *NeuroD* and *Phox2a* expression is gained (Abu-Elmagd et al., 2001). Indeed, overexpression of *Sox3* causes defects in neurogenesis suggesting that its downregulation is required for neuroblast formation (Abu-Elmagd et al., 2001).

Lastly, both *Six1* and *Eya1* continue to be expressed in the otic and epibranchial placodes with *Six1* enhancers (Six1-12 and Six1-21) and *Eya1* enhancers (CS1-3 and CS1-5) active in both otic and epibranchial regions (Ishihara et al., 2008b, Sato et al., 2012). The enhancer Six1-21 has binding sites for *Foxi, Pax* and *Sox* family members that decrease its activity upon mutation (Sato et al., 2012). It is however unclear which transcription factors drive activity in otic or epibranchial or both placodes.

In conclusion, the otic and epibranchial placodes originate from a common region, the OEPD, and along their developmental path they progressively differentiate in response to different transcriptional and signalling inputs. All interactions are shown in Table 1.1 where gene names from species other than chick are reported in brackets and experimentally verified direct regulatory interactions are reported with an asterisk (*). Key regulatory events from PPR to otic placode as described above are shown in a gene regulatory network (GRN; Figure 1.2).

Table 1.1 Known PPR to otic placode regulatory relationships

Source	Interaction	Target	Organism	Evidence
Regionalisation of the PPR				
Otx1/2/5	Represses	Gbx2	Xenopus	(Steventon et al., 2012)
Otx1/2/5	Promotes	Dmrt4	Xenopus	(Steventon et al., 2012)
SSTR5	Promotes	Noc	Chick	(Lleras-Forero et al., 2013)
Noc	Promotes	SSTR5	Chick	(Lleras-Forero et al., 2013)
Pax6*, Noc, Otx1/2/5, Six1/Eya2, Six3*, SST	Promotes	Pax6	Chick, mouse Xenopus	(Ashery-Padan et al., 2000*, Liu et al., 2006*, Christophorou et al., 2009, Steventon et al., 2012, Lleras-Forero et al., 2013)
Dlx5/6 (Dlx3b/4b), FGF, Pax3, TGFβ, WNT	Represses	Pax6	Chick, mouse	(Bhattacharyya et al., 2004, Smith et al., 2005, Bailey et al., 2006, Grocott et al., 2011, Wakamatsu, 2011)
Pax6	Promotes	Six3	Mouse	(Ashery-Padan et al., 2000)
FGF, Pax3, PDGF, Six1/Eya2, WNT	Promotes	Pax3	Chick	(Lassiter et al., 2007, Canning et al., 2008, McCabe and Bronner-Fraser, 2008, Christophorou et al., 2009, Dude et al., 2009)
Pax6	Represses	Pax3	Chick	(Wakamatsu, 2011)
Pax3, SST	Promotes	Eya2	Chick	(Dude et al., 2009, Lleras-Forero et al., 2013)
Pax3	Represses	Otx1/2/5	Xenopus	(Steventon et al., 2012)
Posterior PPR interactions				
Dlx5/6 (Dlx3b/4b), FGF, Foxi1/3	Promotes	Dlx5/6 (Dlx3b/4b)	Chick, Xenopus, Zebrafish	(Solomon and Fritz, 2002, Nissen et al., 2003, Solomon et al., 2003b, Hans et al., 2004, Litsiou et al., 2005, Bailey et al., 2006, Hans et al., 2007, Aghaallaei et al., 2007, Pieper

				et al., 2012, Khatri et al., 2014)
Dlx5/6 (Dlx3b/4b), FGF	Promotes	Foxi3 (Foxi1)	Chick, Xenopus, Zebrafish	(Nissen et al., 2003, Solomon et al., 2003a, Phillips et al., 2004, Hans et al., 2004, Hans et al., 2007, Pieper et al., 2012, Khatri et al., 2014)
FGF	Promotes	Etv5	Chick, Xenopus, Zebrafish	(Raible and Brand, 2001, Roehl and Nusslein-Volhard, 2001, Lunn et al., 2007, Kwon et al., 2010)
FGF	Promotes	Etv4	Chick, Zebrafish	(Raible and Brand, 2001, Roehl and Nusslein-Volhard, 2001, Lunn et al., 2007)
Gbx2	Represses	Otx1/2/5	Xenopus	(Steventon et al., 2012)
Dlx5/6 (Dlx3b/4b), FGF, Foxi3 (Foxi1), Irx1	Promotes	Six1	Chick, Xenopus, Zebrafish	(Woda, 2003, Glavic et al., 2004, Ahrens and Schlosser, 2005, Pieper et al., 2012, Khatri et al., 2014)
Dlx5/6 (Dlx3b/4b), FGF, Foxi3 (Foxi1), Six1/Eya2	Promotes	Eya1/2	Chick, Medaka, Zebrafish	(Leger and Brand, 2002, Litsiou et al., 2005, Esterberg and Fritz, 2009, Christophorou et al., 2009, Kwon et al., 2010, Pieper et al., 2012, Khatri et al., 2014)
Dlx5/6 (Dlx3b/4b), FGF, Foxi3 (Foxi1), Six1/Eya2	Promotes	Six4	Chick, Xenopus, Zebrafish	(Leger and Brand, 2002, Litsiou et al., 2005, Esterberg and Fritz, 2009, Christophorou et al., 2009, Kwon et al., 2010, Pieper et al., 2012, Khatri et al., 2014)
PPR to OEPD interactions				
FGF, Gbx2, Foxi3 (Foxi1)	Promotes	Pax8	Xenopus, zebrafish	(Phillips et al., 2001, Leger and Brand, 2002, Nissen, 2003, Solomon et al., 2003b, Hans et al., 2004, Phillips et al., 2004, Mackereth et al., 2005, Park and Saint-Jeannet, 2008, Padanad and Riley, 2011, Padanad et al., 2012, Steventon

				et al., 2012)
Dlx5/6 (Dlx3b/4b), Gbx2, Foxi3 (Foxi1), FGF, Six1/Eya2, Pax8	Promotes	Pax2	Chick, mouse, Xenopus, zebrafish	(Ladher et al., 2000, Leger and Brand, 2002, Maroon et al., 2002, Solomon and Fritz, 2002, Wright and Mansour, 2003, Hans et al., 2004, Phillips et al., 2004, Solomon et al., 2004, Mackereth et al., 2005, Ladher et al., 2005, Bricaud and Collazo, 2006, Martin and Groves, 2006, Hans et al., 2007, Sun et al., 2007, Freter et al., 2008, Christophorou et al., 2009, Padanad and Riley, 2011, Steventon et al., 2012, Padanad et al., 2012, Khatri et al., 2014)
Pax2	Represses	Pax3	Chick	(Dude et al., 2009)
Dlx5/6 (Dlx3b/4b)	Represses	Pax6	Chick	(Bhattacharyya et al., 2004)
Pax2	Promotes	Six1	Zebrafish	(Bricaud and Collazo, 2006)
FGF	Promotes	*Sox*8	Chick	(Yang et al., 2013)
FGF, Foxi3, Pax2 (Foxi1)	Promotes	Foxg1	Chick, mouse	(Urness et al., 2010, Freter et al., 2012, Yang et al., 2013, Khatri et al., 2014)
Pax2*, Etv4*	Promotes	Spalt4	Chick	(Barembaum and Bronner-Fraser, 2010)
Pax2	Promotes	Gata3	Chick	(Christophorou et al., 2010)
FGF, Six1/Eya2	Promotes	Foxi2/3 (Foxi1)	Chick, zebrafish	(Christophorou, 2008, Christophorou et al., 2009)
Otic and Epibranchial interactions				
WNT	Promotes	TCF/Lef1	Mouse	(Ohyama et al., 2006)
WNT	Promotes	Notch	Mouse	(Jayasena et al., 2008)
Notch	Promotes	TCF/Lef1	Mouse	(Jayasena et al., 2008)

WNT	Promotes	Pax2	Chick, mouse, zebrafish	(Ohyama et al., 2006, Freter et al., 2008, McCarroll et al., 2012)
Notch, WNT	Promotes	Pax8	Mouse, Xenopus, zebrafish	(Phillips et al., 2004, Ohyama et al., 2006, Jayasena et al., 2008, Park and Saint-Jeannet, 2008, McCarroll et al., 2012)
Etv4*, cMyb*, *Sox*8*	Promotes	*Sox*10	Chick	(Betancur et al., 2011*)
Pax2	Promotes	Eya1	Chick	(Christophorou et al., 2010)
Dlx5/6 (Dlx3b/4b)	Promotes	Six1	Medaka	(Aghaallaei et al., 2007)
WNT	Promotes	Dlx5/6 (Dlx3b/4b)	Mouse	(Ohyama et al., 2006)
WNT	Promotes	Gbx2	Mouse	(Ohyama et al., 2006)
WNT	Promotes	Nkx5.1	Chick	(Freter et al., 2008)
FGF	Represses	Nkx5.1	Chick	(Freter et al., 2008)
Pax2,WNT	Promotes	Soho1	Chick	(Freter et al., 2008, Freter et al., 2012)
FGF	Represses	Soho1	Chick	(Freter et al., 2008)
FGF, Six1	Promotes	Foxi2/3 (Foxi1)	Chick, mouse, zebrafish	(Bricaud and Collazo, 2006, Freter et al., 2008, Rogers et al., 2011, Khatri et al., 2014)
Notch, Pax2, Pax8, WNT	Represses	Foxi2/3 (Foxi1)	Mouse, zebrafish	(Ohyama et al., 2006, Jayasena et al., 2008, Freter et al., 2008, Padanad and Riley, 2011)
FGF, Foxi2/3 (Foxi1)	Promotes	*Sox*3	Chick, zebrafish	(Sun et al., 2007, Nikaido et al., 2007, Abello et al., 2010, McCarroll and Nechiporuk, 2013)
Pax2, Pax8	Represses	*Sox*3	Zebrafish	(Padanad and Riley, 2011)
Foxi2/3 (Foxi1), *Sox*3	Promotes	Phox2a	Chick, zebrafish	(Abu-Elmagd et al., 2001, Lee et al., 2003)

BMP, FGF, Foxi2/3 (Foxi1), Phox2a	Promotes	Phox2b	Chick, mouse, zebrafish	(Pattyn et al., 1997, Holzschuh et al., 2005, Freter et al., 2008)
WNT	Represses	Phox2b	Chick	(Freter et al., 2008)

1.6 Enhancers: properties and epigenetic signatures

The regulatory interactions shown in Table 1.1 have been identified mainly through loss or gain-of-function experiments with the exception of a few where enhancers have been identified (Barembaum and Bronner-Fraser, 2010, Betancur et al., 2011, Sato et al., 2012). It is therefore unclear which of these relationships are direct or indirect. To build a GRN, it is not only important to identify the targets of a specific protein, but also whether this involves direct interaction of a protein and its target by binding to the relevant enhancer. To accomplish this, it is essential to identify enhancers. In the next sections, I'll discuss the features of enhancers and some of the ways of identifying them.

1.6.1 Properties of enhancers

In eukaryotes, gene regulation is brought about by a multitude of factors including the general transcription machinery, binding of transcription factors (TFs) and co-factors to regulatory elements, modifications of the nucleosomes and splicing (Erwin et al., 2014). Enhancers play a fundamental role in gene regulation as these are bound by transcription factors and influence the timing and amount of gene expression in a particular tissue (reviewed in (Ong and Corces, 2011, Bulger and Groudine, 2011, Shlyueva et al., 2014)). Moreover, many enhancers act in a tissue specific manner, mostly regulating only one of the many expression domains of a specific gene. Therefore, each gene generally is controlled by different enhancers driving its expression in different tissues (Visel, 2009, Visel, 2013). Enhancers have been found in introns, intergenic regions and can be located upstream or downstream of the transcription start site (Noonan and McCallion, 2010).

Enhancers range in size from a few hundred bp to about 50 kb (clusters of enhancers; also called super-enhancers) (Whyte et al., 2013). The first identified enhancer was a 72 bp sequence of the SV40 virus genome that could enhance the transcription of a reporter gene by several hundred fold in HeLa cells (Banerji, 1981). Enhancers can be proximal, located close to their target genes (a few kb away) (Wei et al., 2005, Erwin et al., 2014) or distal in which case these are located many kilobases, even megabases away (Vavouri et al., 2006). For example, an enhancer for Sonic hedgehog (*Shh)* which is conserved between fish and mammals is located 1 Mb upstream in the intron of another gene (Lettice et al., 2003). Similarly, some enhancers for *Sox9* are located over 1 Mb upstream of its transcription start site (Bagheri-Fam et al., 2006, Gordon et al., 2009). In such cases, enhancers are able to act on their target genes through the process of looping (Bartolomei et al., 1991, DeChiara et al., 1991, Splinter et al., 2006, Hou et al., 2008, Amano et al., 2009, Kagey et al., 2010, Zhang et al., 2013b) which brings the enhancer close to its target gene's promoter.

Typically, before an enhancer is accessible to its interacting transcription factors the so-called pioneer factors bind to condensed chromatin and displace nucleosomes to make the target sequence accessible (for review see Zaret and Carroll, 2011) (Figure 1.3 A-B). For example, the pioneer factor *Foxa1* is required for normal liver development (Lee et al., 2005) and is reported to access its binding sites in nucleosomal DNA. Subsequently, it decompacts chromatin and displaces nucleosomes from its target enhancer region. In *Drosophila*, *Zelda* acts as a pioneer factor of early zygotic enhancers although the underlying mechanism of its action and how it facilitates other factors to bind is unclear (Liang et al., 2008, Harrison et al., 2011). The binding of pioneer factors is often accompanied by H3K4 mono-methylation, which leads to the opening of chromatin thereby facilitating the binding of other transcription factors (Xu et al., 2007, Lupien et

al., 2008, Xu et al., 2009, reviewed in Zaret and Carroll, 2011, Serandour et al., 2011, reviewed in (Spitz and Furlong, 2012, Calo and Wysocka, 2013)) (Figure 1.3 B-C). Thus, tissue-specific TFs bind to active enhancers followed by the binding of co-factors, which lack DNA-binding domains and hence interact with the bound TFs forming an enhancer complex (Erokhin et al., 2015) (Figure 1.3 D). Additionally, mediator, a conserved multiprotein co-activator complex, interacts with the already formed enhancer complex as well as RNA polymerase II at the promoter leading to the regulation of transcription (Szutorisz et al., 2005, Malik and Roeder, 2010, Ansari and Morse, 2013, Poss et al., 2013) (Figure 1.3 E). This process is mediated by DNA looping which is thought to be controlled by the synergistic action of the mediator complex and the chromosome-associated multi-subunit protein complex cohesin to bring enhancers close to promoters, thus allowing the regulation of transcription (Kagey et al., 2010, Zhang et al., 2013b) (Figure 1.3 E).

The importance of enhancers in development is revealed from several examples. A single bp change in an enhancer of *Shh* results in polydactyly and other limb abnormalities in human , mouse , cat (Lettice et al., 2008) and chicken (Maas et al., 2011). Hirschsprung (HSCR) disease is a complex genetic disorder resulting in the failure of the enteric neural crest cells to form ganglia in the hindgut (Erokhin et al., 2015). HSCR is associated with a single-nucleotide polymorphism (SNP) in the *RET* enhancer (Emison et al., 2005). Van Buchem (VB) disease is an autosomal skeletal dysplasia that causes bone overgrowth (Erokhin et al., 2015). This disease is associated with the deletion of a 52 kb region in the enhancer of *SOST* gene (Staehling-Hampton et al., 2002). Thus, enhancers seem to play a vital role in development.

Advances in enhancer identification methods have now allowed genome-wide enhancer discovery which will be discussed further in the next sections.

1.6.2 Histone modifications associated with enhancers

Eukaryotic DNA is packaged into chromatin whose basic subunit is the nucleosome. A nucleosome is formed from DNA wrapped around an octamer of four core histones H2A, H2B, H3 and H4 in ~1.7 turns equalling 146 bp of DNA. Adjacent nucleosomes are separated by linker DNA, which is ~20-50 bp in length (reviewed in (Annunziato, 2008, Herold et al., 2012)). It has been suggested that a number of covalent modifications of different amino acids on the N-terminal tails of histones changes chromatin configuration, which in turn allows binding of activating or repressing factors for functions such as transcription.

Genome-wide mapping of histone modifications has revealed that specific marks are associated with transcriptionally active and repressed regions (Roh et al., 2005, Heintzman et al., 2007, Heintzman et al., 2009, Rada-Iglesias et al., 2011, Bonn et al., 2012). Active promoters are marked by H3K4me3 and H3K27ac whereas active enhancers are marked by H3K4me1 and H3K27ac (Rada-Iglesias et al., 2011, Bonn et al., 2012, Arnold et al., 2013) (Figure 1.3 D-E). Alternatively, repressed or silent promoters and enhancers are marked by H3K27me3 (Simon and Kingston, 2009, Tie et al., 2009).

H3K4me1 was the first histone modification linked to enhancers (Heintzman et al., 2007). In mammals, methyltransferases *MLL3/4* are mainly responsible for mono-methylation of H3 Lysine 4 (for review see Calo and Wysocka, 2013), whereas the demethylase *LSD1* is responsible for the removal of methylation from H3 Lysine 4 rendering the enhancer inactive (Whyte et al., 2012). Some studies indicate that H3K4me1 pre-marks regions that

have the potential of being active and subsequently guides and facilitates pioneer factors like *Foxa1* to these regions (Lupien et al., 2008). Thus for H3K4me1, a potential role in priming the enhancer is suggested where it marks broad regions and facilitates enhancer activation possibly through the binding of pioneer factors and nucleosome displacement which then allows other TFs to come and bind to the primed enhancer (Figure 1.3 B-C).

In addition to H3K4me1, H3K27ac marks the active enhancers. During transcription, several co-activator proteins which have histone acetyltransferase (HAT) activity are recruited to enhancers. P300 and CBP are homologous, ubiquitously expressed HATs that work cooperatively and are recruited to enhancers where they acetylate H3K27 (Goodman and Smolik, 2000, Tie et al., 2009, Pasini et al., 2010, Jin et al., 2011) (Figure 1.3 D). The presence of H3K27ac distinguishes an active enhancer from a poised enhancer where the latter is not functionally active but can be activated in response to external stimuli (Creyghton et al., 2010). In human embryonic stem cells (ESCs), poised enhancers are occupied by p300, but lack H3K27ac marks (Heintzman et al., 2009, Creyghton et al., 2010, Zentner et al., 2011, Rada-Iglesias et al., 2011, Bonn et al., 2012) (Figure 1.3 F). This indicates that acetylation of enhancers is carried out in two steps: the recruitment of p300/CBP and then regulation of its enzymatic activity directly or indirectly. As a result of H3K27 acetylation, the lysine residues are recognized by bromodomains of nuclear proteins including TFIID complex, HATs (e.g. p300, CBP) and other factors that regulate transcription (Filippakopoulos and Knapp, 2012, Filippakopoulos et al., 2012). Thus through acetylation, HATs promote their further recruitment maintaining these regions in an active state. Since charges on both DNA and histone maintain electrostatic interactions necessary for the formation of a compact nucleosome, it has been suggested that acetylation of lysine residues on histones may

disrupt this compact structure leading to nucleosome displacement and facilitating binding of TFs to enhancers (reviewed in Calo and Wysocka, 2013).

What causes the enhancers to become inactive? Polycomb group (PcG) proteins act as a complex to silence genes (Schwartz and Pirrotta, 2007). The Polycomb repressor complex 2 (PRC2) contains H3K27-specific histone methyltransferase called E(Z) along with other proteins and is responsible for mono, di and tri-methylation of H3K27 (Tie et al., 2009) (Figure 1.3 F). However, only H3K27me3 (not H3K27me1/2) is implicated in silencing of *PRC2* target genes (Wang et al., 2008b). The balance between *CBP*-mediated acetylation of H3K27 and Polycomb mediated tri-methylation seems to determine whether or not an enhancer is active (Tie et al., 2009, Tie et al., 2014). Knockdown of the histone deacetylase *RPD3* (associated with PcG complex) by RNAi in *Drosophila* S2 cells elevated H3K27ac levels indicating that silencing of the Polycomb target genes may be mediated by *RPD3* through deacetylation. Conversely, partial knockdown of *CBP* using *Gal4*-driven-*CBP* RNAi transgene (Kumar et al., 2004) showed a decrease in H3K27ac and a 49% increase in H3K27me3 (Tie et al., 2009). In these animals, the abdominal segment was truncated in adults, a phenotype attributed to reduced expression of the homeotic gene *Abd-B*. Thus, a global reduction of H3K27ac and increase of H3K27me3 is associated with silencing of *Abd-B*. In addition, H3K27ac and H3K27me3 profiles appear to be complementary to each other: as Polycomb mediated silencing begins with an increase in H3K27me3 and decrease in H3K27ac suggesting an antagonistic relationship (Tie et al., 2009, Pasini et al., 2010, Tie et al., 2014). Indeed, when the methyltransferase *E(Z)* as part of the *PRC2* complex is knocked down using RNAi, H3K27me3 decreases which is accompanied by a 3-fold increase in H3K27ac globally.

Thus, H3K4me1 and H3K27ac appear to mark active enhancers whereas H3K27me3 marks inactive enhancers.

1.7 Methods for the identification of enhancers

There are various methods for enhancer identification: some are strictly computational, others experimental. In this section I will discuss some of the widely-used methods of enhancer discovery and highlight their advantages and limitations.

1.7.1 Identification of enhancers using histone modifications

Recently there has been an increase in the use of histone marks in genome-wide enhancer prediction studies (Roy et al., 2010, Rada-Iglesias et al., 2011, Kharchenko et al., 2011, Ernst et al., 2011, Bonn et al., 2012, Shen et al., 2012) which has led to the identification of several features associated to active and repressed enhancers (see review Shlyueva et al., 2014). For example, poised enhancers are thought to carry histone modifications for both active and repressed states (H3K4me1 and H3K27me3) (Bernstein et al., 2006), latent enhancers are not marked by H3K4me1 or H3K27ac, but acquire these marks after stimulation of signalling pathways (Ostuni et al., 2013, Zhang et al., 2013a), while active enhancers bind the transcriptional cofactor p300 in addition to being marked with H3K27ac and H3K4me1 (Ghisletti et al., 2010, Blow et al., 2010, Orom et al., 2010, May et al., 2012) .

There is no strong consensus about which histone modifications should be used for finding active enhancers. This is due to the reason that none of the known histone modifications correlates perfectly with enhancer activity (for review see Shlyueva et al., 2014). Even so, the most widely used identification marks for active enhancers are H3K4me1 and H3K27ac (Heintzman et al., 2007, Rada-Iglesias et al., 2011) and it has

been reported that a combination of H3K4me1 and H3K27ac is a good predictor of *Drosophila* developmental enhancers (Bonn et al., 2012). Typically for enhancer prediction, in addition to ChIP-seq, data collected from DNA accessibility studies including DNase-seq (Boyle et al., 2008a), MNase-seq (Yuan et al., 2005), ATAC-seq (Buenrostro et al., 2015) or FAIRE-seq (Giresi et al., 2007) are incorporated to increase the chances of enhancer discovery.

Chromatin immunoprecipitation (ChIP) is the most powerful and widely used experimental technique for the mapping of DNA-associated proteins (Ji et al., 2008, Barski and Zhao, 2009). In a typical ChIP-seq experiment, the tissue of interest is crosslinked using a fixative (such as formaldehyde). This is followed by lysis of cells and DNA fragmentation by sonication (Ji et al., 2008, Barski and Zhao, 2009). Following this, immunoprecipitation is carried out using an antibody against the protein of interest. Finally, the DNA is sequenced by high-throughput sequencing. In a typical ChIP-seq experiment, the sequenced reads are mapped to a reference genome either to unique or multiple sites. Keeping only the uniquely mapped reads may result in the loss of some true sites of occupancy located in duplicated regions or repeats therefore the choice of mapping entirely depends upon the user and the biological question (see review Pepke et al., 2009). After mapping, using any of the available programs such as Bowtie (Langmead et al., 2009) or Novoalign (Novocraft 2.08.01), the next challenge is to identify true binding locations (narrow or broad) in comparison to a background (control reaction) either provided by the user or modelled by the algorithm in the absence of a control reaction (Ladunga, 2010). 60-90% of the sequenced reads come from background due to the binding of antibody to untargeted proteins (see review Pepke et al., 2009). However, this can be managed in part by estimating background noise levels and control of the false discovery rate (FDR) (Benjamini, 1995). A smoothed signal profile across the genome is

produced by using a sliding window of user-specified length and associating each position with the number of tags within the window. Within this profile, a position with a locally maximal read density is termed a summit. Regions and summits in a signal profile that pass a minimum enrichment threshold relative to the background (e.g., a fold-change threshold) and/or an FDR (False Discovery Rate) or p-value threshold are reported as peaks.

There are many algorithms that can be used for peak-calling such as cisGenome (Ji et al., 2008), SiSSRs (Jothi et al., 2008), MACS (Zhang et al., 2008), Homer (Heinz et al., 2010) and SICER (Zang et al., 2009). Although most of the peak-calling programs have been developed to identify sharp peaks (characteristic of a TF), there are some specifically designed to identify broad peaks (characteristic of epigenetic modifications). Some of these include Homer (Heinz et al., 2010), ChromaSig (Hon et al., 2008) and BroadPeak (Wang et al., 2013). In a recent review (Wilbanks and Facciotti, 2010), the performance of 11 peak calling programs was assessed on three published TF ChIP-seq datasets including human neuron-restrictive silencer factor (NRSF) (Johnson et al., 2007), growth-associated binding protein (GABP) (Valouev et al., 2008) and hepatocyte nuclear factor 3α (Foxa1) (Zhang et al., 2008). These TFs were selected as each has a well-defined binding site that can be used to assess the peak-calling programs. The peak calling programs that were tested included CisGenome (Ji et al., 2008), PeakFinder (Johnson et al., 2007), E-RANGE (Mortazavi et al., 2008), MACS (Zhang et al., 2008), QuEST (Valouev et al., 2008), HPeak (Qin et al., 2010), Sole-Search (Blahnik et al., 2010), PeakSeq (Rozowsky et al., 2009), SISSRS (Jothi et al., 2008) and spp (Boyle et al., 2008b). All of these programs were run with the recommended parameter settings and it was found that for each dataset, although the number of peaks reported by each program was different, there was a small list of peaks that was commonly reported by all

11 programs. To detect the sensitivity of each program in identifying true binding sites, the results were compared to qPCR-validated binding sites for NRSF and GABP. The performance levels of all programs were comparable except for Sole-Search and cisGenome that missed several true binding sites which were picked up by other programs. Additionally, to check the specificity of each program, 30 qPCR-determined negative sites for NRSF (enrichment less than 3 fold) were used. Nine out of the 11 programs called 2 false positives whereas cisGenome and QuEST called none. The main difference in the programs was found at the level of estimation of the location of TF-DNA binding event where MACS and spp provided the best estimates. While analysing ChIP-seq peaks, a user may decide to examine either a stringent list of peaks or a comprehensive list of peaks (with more false-positives) depending on the biological question. One way is to set an FDR or p-value threshold to call peaks and then rank the peaks by fold-change (FC) to examine the top *n* peaks where the choice of *n* depends upon the user's desired stringency.

To conclude, several peak-calling programs are available; each having some advantages and limitations. In the present study, MACS and Homer were used due to their overall good performance.

1.7.2 Identification of enhancers using sequence conservation

A substantial number of enhancer prediction methods rely on nucleotide sequence conservation between orthologous species (for review see (Ureta-Vidal et al., 2003, Wasserman and Sandelin, 2004)). Such methods are categorised under phylogenetic footprinting and operate under the assumption that sequence comparison of orthologous genomic regions in closely related species can predict important biological functions. It also assumes that the regulation of orthologous genes in different species uses the same

mechanisms and is under similar evolutionary pressures. The availability of several eukaryotic genome sequences has made it possible to identify conserved regulatory regions by comparing orthologous sequences (Adams et al., 2000, Lander et al., 2001, Aparicio et al., 2002, Woolfe et al., 2005). In recent years, phylogenetic footprinting has become a gold standard in predicting regulatory regions which can then motivate validation experiments. Two major classes of phylogenetic footprinting have been suggested 1) alignment-based methods 2) motif discovery methods. One critical aspect of phylogenetic footprinting is the appropriate choice of species for enhancer prediction as very closely related species such as human and chimpanzee will show high sequence similarity while widely divergent species such as primates and fish will show low similarity (Lenhard et al., 2003). Although comparisons between highly divergent species such as human and puffer fish can indeed reveal regulatory regions in early embryonic development (Aparicio et al., 1995, Bagheri-Fam et al., 2001, Woolfe et al., 2005), it seems that a good choice would be to include both closely-related and divergent species while identifying regulatory regions (reviewed in Pennacchio and Rubin, 2001). This is due to the fact that different regions of a genome within a species evolve at different rates. For example, the beta-globin locus control region (LCR) has evolved rapidly, making it easier to identify conserved regulatory sequences in closely related mammals (Jimenez et al., 1992, Loots et al., 2000, Gottgens et al., 2000). On the other hand, T-cell receptor loci have evolved very slowly (Koop and Hood, 1994, Hood et al., 1995) in which case, distantly related species such as marsupials, birds, reptiles and fish may give a good indication of conserved regulatory sequences. This reveals that to some extent knowing the evolution rates of regions in the genome can help to discern the choice of species for comparison and identification of regulatory sequences.

Once the species have been selected for the comparative analysis, there are two options for the prediction of conserved regulatory regions: alignment or motif discovery. In addition to pair-wise alignments that compare two sequences and are ideally used for the comparison of orthologous regions between two species, it is possible to construct multiple alignments that compare several sequences and are used for comparison of orthologous regions from multiple species. Several multiple alignment tools are available including CLUSTALW (Thompson et al., 2002), MULTALIGN (Barton and Sternberg, 1987), MULTAL (Taylor, 1988), PRRP (Gotoh, 1996), DIALIGN (Morgenstern et al., 1998), MGA (Hohl et al., 2002) and LAGAN and multi-LAGAN (Brudno et al., 2003a). Additionally, pre-computed multiple alignments are available from UCSC via PhastCons and GERP (Meyer et al., 2013), Ensembl PECAN (Paten et al., 2008), ECR base (Loots and Ovcharenko, 2007), ECR Browser (Ovcharenko et al., 2004) and VISTA Browser (Frazer et al., 2004), which have made it easier to use this information for the retrieval of orthologous sequences from different species for comparison and conservation analysis. VISTA additionally provides multiple tools for comparative genomics including GenomeVISTA and mVISTA for pairwise and multiple alignments between different species and rVISTA for searching transcription factor binding sites from TRANSFAC in addition to comparative sequence analysis.

An important aspect to consider using the above methods is the type of alignment used to study conservation. There are two major classes of alignment: local and global (Brudno et al., 2003a). Local alignments maximise similarity by potentially discarding prefixes and/or suffixes of the provided sequences and thus, are able to identify rearrangements; e.g., the Smith-Waterman algorithm (Smith and Waterman, 1981), BLAST and BLASTZ (Altschul et al., 1990, Schwartz et al., 2003), BLAT (Kent, 2002) and Shuffle-LAGAN (Brudno et al., 2003b). On the other hand, global alignments reveal conserved features in

an order-dependent way since the alignment includes both sequences in their entirety. Example algorithms include Needleman-Wunsch (Needleman and Wunsch, 1970), PECAN (Paten et al., 2008), DIALIGN (Morgenstern et al., 1998) and LAGAN and multi-LAGAN (Brudno et al., 2003a). These programs assume that important functional sequences have maintained their order and orientation during evolution.

Many studies have reported the use of multiple sequence alignment in the identification of regulatory regions; some of which have been carried out on a genome-wide scale. For example conserved elements have been identified among four yeast species using BLAST (Kellis et al., 2003), in vertebrates using a multitude of alignment programs: CLUSTALW, MLAGAN, DIALIGN and TBA (Prakash and Tompa, 2005), MegaBLAST followed by identification of regional conservation using MLAGAN (Woolfe et al., 2005) and in mammals using GERP (Cooper et al., 2004). Some studies have reported the use of multiple alignments on a gene level such as the identification of conserved elements for *Eya2* and *Six1* (Ishihara et al., 2008b, Sato et al., 2012). An important consideration while making a choice of the method of enhancer prediction is the fact that TFBSs may be lost, gained or re-arranged during the course of evolution (Ludwig et al., 2000). Because of the dependency of global alignments on order and orientation, global aligners tend to struggle with the identification of conserved segments in case of rearrangements. While local aligners circumvent this problem, there are other methods in phylogenetic footprinting that address this more effectively. These are based on the hypothesis that there are clusters of TFBS motifs in regulatory regions that can be present in any orientation or number and are responsible for binding of TFs in order to regulate transcription. These clusters of TFBS motifs have been termed *cis*-regulatory modules (CRMs) (reviewed in (Wasserman and Sandelin, 2004, Ladunga, 2010, Hardison and Taylor, 2012, Shlyueva et al., 2014)). Hence, a class of enhancer predictors called the

"motif discovery tools" aim to find CRMs in orthologous sequences irrespective of the order or orientation of the motifs within these regions. Some of these tools include FootPrinter (Blanchette and Tompa, 2003), PhyME (Sinha et al., 2004) and the more recent DREiVe [(Sosinsky et al., 2007) (http://dreive.cryst.bbk.ac.uk/)]. While FootPrinter and DREiVe share a similar underlying principle of motif discovery by not relying on alignment for the identification of CRMs, PhyME computes regions of high local similarity through alignment of the orthologous sequences and then identifies conserved motifs within these. DREiVE, on the other hand uses a traditional pattern matching algorithm called SPLASH to identify conserved motifs (Califano, 2000). Both DREiVE and FootPrinter allow flexibility in the identification of the motifs by letting the user specify parameters related to the size of the motif and the number of conserved residues within the motif. As a final step in both, a window of user-specified size screens the input sequences to identify orthologous regions with the highest scoring motif clusters. These are then reported as putative CRMs. Additionally, DREiVE allows the specification of the minimum number of species in which a motif should be found to be considered conserved. In a recent study (Khan et al., 2013), it was shown that DREiVE was able to predict 18 out of 25 (72%) previously known enhancers of *Sox2* (Uchikawa et al., 2003) and an additional nine new putative enhancers. This indicates that phylogenetic footprinting methods are useful in predicting regions that may be involved in the regulation of transcription. More recently, enhancer identification using phylogenetic footprinting coupled to transcription factor binding site analysis has been suggested to improve predictions (Khan et al., 2013).

Thus, a number of different methods are available for enhancer prediction. In the present study, DREiVE was used for enhancer prediction due to its overall good performance as shown by Khan et al. (2013).

1.7.3 Transcription factor binding site analysis of enhancers

Gene regulation is brought about through the cooperative binding of transcription factors that recognise short sequences (6-10 bp) within promoters and/or enhancers (reviewed in Shlyueva et al., 2014). The binding site of a transcription factor can be represented as a consensus sequence or a position frequency matrix (Tjian, 1978, Giniger et al., 1985, Pavesi et al., 2004a). A consensus describes a set of aligned oligonucleotides with the most frequent nucleotide in each position (Figure 1.4 A-B), while a position frequency matrix (PFM) represents the frequency of each nucleotide at each position in the oligonucleotides in a 4 x *m* matrix; *m* being the length of the aligned oligonucleotides (Figure 1.4 C). The PFMs can additionally be converted to position weight matrices (PWMs) using a formula that converts normalised frequency values to a logarithmic scale (see review Wasserman and Sandelin, 2004). The largest collection of PFMs and PWMs of transcription factor binding sites are available from JASPAR (Sandelin et al., 2004), TRANSFAC (Matys et al., 2006) and UniPROBE (Newburger and Bulyk, 2009) where JASPAR and UniPROBE contain collections of experimentally validated TFBSs obtained using high-throughput techniques such as ChIP-sequencing, Protein Binding Microarray or SELEX .

While phylogenetic footprinting predicts enhancers, it does so without employing TFBS data (PFMs or PWMs). Therefore, it is common practice to predict enhancers using phylogenetic footprinting and then predicting TFBSs using TF libraries (reviewed in Wasserman and Sandelin, 2004). There are again two widely used methods for TFBS analysis which will be discussed below.

Typically for transcription factor binding site (TFBS) analysis, the first set of methods use available PFMs or PWMs to screen predicted enhancer sequences. These include RSAT

matrix-scan (Thomas-Chollier et al., 2008), ANN-Spec (Workman and Stormo, 2000), MSCAN (Alkema et al., 2004), MatInspector (Cartharius et al., 2005) and Clover (Frith et al., 2004a). Using the TFBS matrix, a quantitative score can be calculated for the TF at a given location in the DNA sequence by summing the values that correspond to the observed nucleotide at each position. Since each TFBS tool has its own scoring scheme to identify enriched binding sites in regulatory sequences as compared to a background, the choice of the tool often depends on the type of biological question being asked (Khan et al., 2013). For instance, in the RSAT suite, matrix-scan allows the user to find all TFBSs in individual sequences as well as over-represented TFs in the entire set of sequences (Thomas-Chollier et al., 2008). On the other hand, Clover (Frith et al., 2004a) can only detect over-represented TFBSs in the entire set of sequences. Thus, if one is interested in finding all binding sites in a single sequence, it would be ideal to use matrix-scan.

One of the drawbacks of using TFBS libraries to screen regulatory sequences is the increase in false positives in case of short and/or degenerate binding sites (reviewed in Pennacchio and Rubin, 2001). An example of this is the TATAA sequence that is short and therefore, its rate of occurrence is very high. This can be circumvented by giving more weight to occurrences of TATAA within 30 bp of a transcription initiation site. However, it is not always this simple for all TFBSs and some ways have been suggested to tackle this problem. For example, if a transcription factor cooperates with another transcription factor, then instead of searching for single binding events of the TF, composite binding events of the two TFs can be detected in the regulatory sequences (van Helden et al., 1998, Wagner, 1999). However, despite very stringent criteria, there is always a possibility of predicting TFBSs that may be non-functional. Additionally, TFBS libraries are not comprehensive and therefore do not include binding sites for all annotated TFs. There is thus a need for generation of data that can be used to produce

good-quality TF PFM models. Recently, different methods have been developed and applied to characterise DNA binding preferences of some of the annotated transcription factors of *Drosophila* and humans but the databases are still far from complete (Noyes et al., 2008, Wei et al., 2010, Jolma et al., 2013).

The second type of methods find over-represented motifs (*de novo* motif discovery) within regulatory sequences as compared to the background without prior knowledge of TFBSs but subsequently these motifs can be matched to TFBS libraries such as JASPAR and TRANSFAC (Tompa et al., 2005). Some tools of this type include MEME (Bailey and Elkan, 1995, Bailey, 2002), AlignACE (Hughes et al., 2000), ANN-spec (Workman and Stormo, 2000), Consensus (Hertz and Stormo, 1999), GLAM (Frith et al., 2004b), MITRA (Eskin and Pevzner, 2002), MotifSampler (Thijs et al., 2001), oligo-dyad-analysis (van Helden et al., 1998, van Helden et al., 2000), Weeder (Pavesi et al., 2004b) and YMF (Sinha and Tompa, 2003). Some of these methods such as Weeder and MITRA are based on Suffix trees which are efficient in finding short motifs and allow some degree of variation within the motif. However, these methods are less effective in finding long words in which case probabilistic methods such as those using Gibbs sampling (Lawrence et al., 1993) or Expectation maximization (EM) (Bailey and Elkan, 1994) perform better (for review see Ladunga, 2010). MEME uses EM to find over-represented motifs in a set of sequences as compared to the background. Although the general assumption in EM is that each of the input sequences should at least have one occurrence of the motif, this can be bypassed in MEME. It also takes into account the multiple occurrences of a motif within a sequence. On the other hand, GLAM , AlignACE and MotifSampler use Gibbs sampling and unlike MEME, the assumption is that each input sequence has at least one occurrence of the motif, which makes these methods quite stringent (reviewed in Ladunga, 2010). Where some *de novo* motif discovery tools limit

the length of the motifs to between 6 and 12 bp, others such as MEME allow the identification of motifs of any length between 5 and 50 bp which is useful in case of bacterial sequences where TFBSs are frequently longer (for review see Tompa et al., 2005).

Thus, a large number of tools are available for the identification of TFBSs but the question is which performs the best. In a review on the performance of 13 motif discovery tools, binding sites for 56 TFs from TRANSFAC were used to produce 56 different datasets (one for each TF) (see review Tompa et al., 2005). Each dataset included a mix of four different types of sequences: 1) actual promoter sequences with known location and orientation of the binding sites of the TF, 2) randomly chosen promoter sequences, 3) simulated sequences with binding sites for the TF planted in them, and 4) negative controls with no binding sites for that TF. The reviewed tools were used to report the best over-represented motif in each dataset. It was found that MEME and Weeder outperformed the other tools in reproducing the motif corresponding to the right TF. However, since this study was constrained to the reporting of the best motif rather than several top motifs, it may explain the compromised performance of some tools.

One of the limitations of finding TFBSs through motif-discovery tools is that these are constrained by the size of the motif, while the actual TF binding sites in JASPAR and TRANSFAC are variable with some binding sites being long. Another drawback is that these methods readily identify short duplicated regions as over-represented motifs but in reality, these are not necessarily biologically relevant. However, an advantage is that one may be able to find biologically relevant motifs that do not correspond to any TF due to the lack of comprehensive TFBS libraries, but may still be very important in imparting function to the regulatory sequences.

To conclude, both identification methods have advantages and limitations. It is difficult to truly measure the correctness of any tool's predictions due to a lack of thorough understanding of the underlying regulatory mechanisms. Moreover, even if a binding site has been predicted, it may not necessarily be bound by a TF in a given cell type as TF binding is highly context-dependent (Yanez-Cuna et al., 2012). Therefore, instead of using a single tool for TFBS identification, a much better approach is to use a combination of different TFBS tools and consider the overlap (Hu et al., 2005, Khan et al., 2013). In the present study, RSAT and Clover were used for TFBS analysis using JASPAR and TRANSFAC libraries. Clover was selected for its overall good performance (McLeay and Bailey, 2010) and RSAT was selected because it can deal with large datasets (sequences >3000 in number) which was useful in identifying TFBSs in ChIP-seq identified enhancers (Chapter 4).

1.7.4 Integrative approaches to enhancer discovery

Advances in techniques such as ChIP-seq (Johnson et al., 2007), RNA sequencing (RNA-seq), sequencing of DNaseI-digested chromatin (DNase-seq) (Boyle et al., 2008a) and formaldehyde-assisted isolation of regulatory elements (FAIRE-seq) (Giresi et al., 2007) have enabled the genome-wide measurement of histone modifications, transcription levels, binding sites of regulatory proteins and structural conformation of DNA (Erwin et al., 2014). Additionally, ENCODE (Consortium, 2012) and FANTOM (Andersson et al., 2014) have made such functional genomics data publicly available. Recently, there have been advancements in machine learning approaches that can use such datasets to identify DNA sequence features of experimentally determined enhancers (Heintzman et al., 2007, Thurman et al., 2012, Cotney et al., 2012) and use these to predict novel enhancers. Typically, in machine learning techniques, a classification algorithm is trained on two types of datasets, e.g., enhancers and non-enhancers from which it learns the sequence

features (e.g., evolutionary conservation, chromatin signature, DNA motifs) of each set (Erwin et al., 2014). A trained classifier is then used to assign uncharacterised genomic sequences to either of the two categories (e.g., enhancer or non-enhancer). Some examples of machine learning approaches include support vector machines (SVMs) that have been trained on known enhancers and have successfully identified novel enhancers in the heart (Narlikar et al., 2010), muscle (Busser et al., 2012) and the hindbrain (Burzynski et al., 2012). Some studies have used histone modification data (H3K4me1) or p300 binding data to train these methods and to predict enhancers in human embryonic stem cells and lung fibroblasts (Gorkin et al., 2012, Rajagopal et al., 2013). Although different types of data are available to train machine learning programs, most approaches use a single histone mark or a particular pattern such as p300 to distinguish between enhancers and non-enhancers. A recent program, EnhancerFinder, is trained on evolutionary conservation, chromatin signature as well as DNA motifs (Erwin et al., 2014) and has been shown to predict four times as many VISTA enhancers when a combination of datasets is used compared to when only one dataset is used.

Thus, there are many different methods of enhancer identification. Some of the most widely used methods have been discussed here. With the vast availability of sequenced genomes, it has become possible to predict evolutionarily conserved enhancers through comparative analyses. Additionally, histone modification data from ENCODE and other sources have allowed the identification of many features of enhancers that were previously unknown. Current efforts to develop a vocabulary for enhancers from different datasets will ultimately contribute to an increase in understanding of regulation and regulatory mechanisms. One of the pressing problems in identification of enhancers is: How far should the upstream and downstream regions of a gene be analyzed to find

enhancers? To answer this, it is important to understand the role of the CCCTC-binding factor (CTCF) in regulation which is discussed in the next section.

1.8 CCTC-binding factor (CTCF) and its role in insulation

Insulators are DNA elements that prevent inappropriate interactions between neighbouring regions of the genome (Defossez and Gilson, 2002, West et al., 2002, Cuddapah et al., 2009). Insulators can be divided into two classes: enhancer-blocking and barriers. Enhancer-blockers prevent genes from interacting with neighbouring unrelated enhancers (Figure 1.5 A), whereas barriers prevent genes and their regulatory elements from the repressive influence of the neighbouring heterochromatin by forming boundaries around them and facilitating interactions between them (Gerasimova and Corces, 1996, Bell et al., 1999, Felsenfeld et al., 2004) (Figure 1.5 B). Insulators have been identified in different organisms including vertebrates, *Drosophila* and yeast and are known to bind proteins that mediate the insulator activity (Chung et al., 1993, Bi and Broach, 2001, Gerasimova and Corces, 2001, Dhillon and Kamakaka, 2002, Donze and Kamakaka, 2002). Several insulator binding proteins have been identified in *Drosophila* but the major insulator binding protein in vertebrates is CTCF (CCCTC-binding factor) (Bell et al., 1999, Gerasimova and Corces, 2001, West et al., 2002). CTCF is a multi-functional protein and binds to insulator elements via its highly conserved zinc finger domain. It has been shown to play an important role in transcription activation and repression, imprinting, long-range chromatin interactions and organization of higher order chromatin by looping (Baniahmad et al., 1990, Filippova et al., 1996, Bell and Felsenfeld, 2000, Yusufzai et al., 2004, Splinter et al., 2006, Hou et al., 2008, Phillips and Corces, 2009, Barkess and West, 2012, Ghirlando et al., 2012).

One of the most prominent examples of enhancer-blocking and looping mediated by CTCF is at the imprinting control region (ICR) downstream of the maternal gene *Igf2* (Figure 1.6). The Insulin-like growth factor 2 *Igf2/H19* locus was one of the first imprinted gene loci identified (Bartolomei et al., 1991, DeChiara et al., 1991). In insects and vertebrates, genetic information is stored in two copies; one received from each parent and usually for each gene both copies are expressed (Herold et al., 2012). In mammals, however, there is a small number of genes (~100) with a mono-allelic expression due to a parental imprint (i.e. some genes are expressed only if inherited from the mother and others when inherited from the father). These genes are highly conserved in mammals and are arranged such that *H19* lies downstream of *Igf2*. In mice, gene knockout studies showed that *Igf2* enhances foetal growth, whereas *H19* retards it (DeChiara et al., 1991). In both mouse and human, the same enhancer activates *H19* on the maternal allele and *Igf2* on the paternal allele (Herold et al., 2012), and this enhancer is downstream of *H19* (Figure 1.6). In contrast, the ICR is located between *H19* and *Igf2* and regulates differential activation of each gene. Several groups have shown that on the maternal allele, the ICR binds CTCF, which insulates the enhancer from the *Igf2* promoter, but allows the expression of *H19* instead (Bell and Felsenfeld, 2000, Hark et al., 2000, Kanduri et al., 2000, Szabo et al., 2000). On the paternal allele, both the ICR and the *H19* promoter are methylated and therefore CTCF does not bind the ICR (Figure 1.6). Instead an insulator site upstream of *Igf2* and downstream of the enhancer forms a loop bringing the enhancer close to the *Igf2* promoter and allowing its expression. Deletion of CTCF binding sites in maternal DNA results in biallelic expression of *Igf2*, thus confirming that CTCF bound to ICR leads to enhancer-blocking on the maternal allele (Engel et al., 2006).

Recent Hi-C experiments indicate that genomes of higher eukaryotes are arranged into topologically associating domains (TADs) with a high degree of interactions within domains and a low degree of interactions between domains (for review see: Ong and Corces, 2014). These domains or TADs are insulated from each other to prevent erroneous gene expression (Filion et al., 2010, Ernst et al., 2011). The boundaries of TADs are enriched for CTCF binding sites implicating its possible role in establishing or maintaining these topological domains (Sofueva et al., 2013, Seitan et al., 2013, Zuin et al., 2014). Depletion of CTCF at TAD boundaries leads to a decrease in intra-domain interactions and increase in inter-domain interactions (Zuin et al., 2014), which suggests a role for CTCF in insulating one TAD domain from another. Within a particular TAD domain, CTCF may direct enhancers to the appropriate gene promoter.

Moreover, genome-wide mapping of histone modifications in different cell types has revealed repressed and active chromatin regions characterised by the presence of H3K27me3 and H3K27ac marks respectively (Barski et al., 2007, Guelen et al., 2008, Filion et al., 2010, Ernst et al., 2011). Combining CTCF and chromatin feature analysis reveals that CTCF demarcates boundaries of repressed chromatin marked with H3K27me3 and active chromatin marked with H3K27ac. This indicates a barrier function for CTCF where active and repressed regions are insulated from each other (Barski et al., 2007, Guelen et al., 2008, Cuddapah et al., 2009, Negre et al., 2010). In a recent study using CTCF and histone ChIP-seq data from HeLa, Jurkat and CD4+ T cells, repressed domains were identified as regions of up to 25 kb marked by H3K27me3. The overlap between CTCF binding sites and H3K27me3 boundaries reveals cell type specific patterns. Although there was an extensive overlap in all CTCF binding sites between different cell types (40%-60%) the overlap of barrier CTCF sites is very small indicating that these are highly cell-type specific (Cuddapah et al., 2009). These findings point

towards a potential role of CTCF in insulating active domains from the neighbouring repressed domains.

The different functions of CTCF may be attributed to its interaction with many other proteins (Donohoe et al., 2007, Wallace and Felsenfeld, 2007, Rubio et al., 2008, Stedman et al., 2008, Wendt et al., 2008). Genome-wide analysis has revealed that cohesin that mediates contact between sister chromatids during cell division co-localizes with CTCF and may function as a transcriptional insulator (Parelho et al., 2008, Rubio et al., 2008, Wendt et al., 2008). Cohesin interacts with the carboxy-terminal region of CTCF through its SA2 subunit and stabilizes the contact of CTCF with the DNA (Wendt et al., 2008, Parelho et al., 2008, Rubio et al., 2008, Stedman et al., 2008, Nativio et al., 2009, Hou et al., 2010, Xiao et al., 2011). Down regulation of cohesin through RNAi causes disruption of CTCF-mediated chromosomal interactions (Nativio et al., 2009, Hou et al., 2010) indicating the importance of CTCF-cohesin interaction.

A number of other features of CTCF binding have been observed in various studies (Kim et al., 2007, Martin et al., 2011, Li et al., 2013). In a study involving the analysis of ENCODE CTCF ChIP-seq data for 56 human cell lines, ~24000 CTCF sites were found in 90% of the cell lines and were termed as constitutive sites, while others seem to be cell type specific (Li et al., 2013). Additionally, ChIP in combination with tiling arrays has revealed distribution patterns of CTCF as 46% intergenic, 22% intronic, 12% exonic, and 20% within 2.5 kb of promoters (Kim et al., 2007) indicating that CTCF has a high occurrence in non-coding regions as compared to coding regions. Recently, it has also been shown that CTCF boundaries are conserved and encompass syntenic regions in human, mouse and chick (Martin et al., 2011). Several genome-wide mapping experiments (ChIP-seq) for CTCF binding and data from ENCODE have allowed the

analysis and comparison of CTCF binding sites between different species (Mukhopadhyay et al., 2004, Kim et al., 2007, Cuddapah et al., 2009, Martin et al., 2011). Finally, CTCF binding site prediction in non-coding DNA regions has recently been used to predict insulator boundaries (Xie et al., 2007, Khan et al., 2013). For computational analysis, the 19 bp position frequency matrix (PFM) of CTCF is available from the JASPAR database (Sandelin et al., 2004) which can be used to screen DNA sequences for identification of CTCF binding sites using the available transcription factor binding site tools.

In summary, CTCF plays multiple roles. It forms boundaries around genes and prevents them from interacting with enhancers outside of the boundaries (enhancer-blocking function) thus allowing genes to interact only with the enhancers within the boundaries. Moreover, it prevents genes and their regulatory elements from the repressive influence of the neighbouring heterochromatin (barrier function).

1.9 Inferring gene regulatory networks

In previous sections, I discussed different ways of identifying enhancers, which ultimately contain the 'code' for temporal and spatially-restricted gene expression. Their analysis allows the identification of regulator-target relationships which thus enrich gene networks. An alternate approach to identifying regulatory relationships between genes involves computational strategies like the inference of GRNs from expression data. Here, I will discuss the importance of inferring GRNs from expression data and some of the methods involved.

During development, genes and proteins interact with each other in complex biological processes that are ultimately involved in determining cell fate (Levine and Davidson,

2005). Gene regulation is one such process where expression and repression of genes is controlled in a systematic manner by TFs. With the availability of large scale genomic data, new methods are being developed for understanding the complexities of regulation. One of these methods is the inference of GRNs (Levine and Davidson, 2005, Huang et al., 2009c, Zhang et al., 2012, Gong et al., 2015). An inferred GRN represents coordinated regulation of gene expression in a cell where genes are represented as nodes in the network and regulatory relationships between genes are represented as edges (Huang et al., 2009c, Zhang et al., 2012). GRNs are useful in giving a systems-level insight into the flow of information in a biological system and help to identify circuits within the network that may be involved in different biological processes (Levine and Davidson, 2005, Stathopoulos and Levine, 2005, Walhout, 2006, Long et al., 2008, Macneil and Walhout, 2011). Moreover, GRNs can be useful in understanding relationships between genes that have similar phenotypes or which are involved in disease (Macneil and Walhout, 2011), as well as pinpoint evolutionary changes (Levine and Davidson, 2005). GRN inference which is also termed as 'reverse engineering' is one of the most challenging tasks in systems biology (Basso et al., 2005, Margolin et al., 2006). The main aim of reverse engineering is to identify interactions between sets of genes from expression datasets according to specific criteria and thus to uncover novel regulatory interactions. The most pressing problem in reverse engineering is the choice of reliable inference algorithms that can reproduce known interactions because only then can their predictions be accepted with high confidence (Levine and Davidson, 2005, Huang et al., 2009c). Additionally, the quality and accuracy of data used for reverse engineering is equally important (Huang et al., 2009c). The most-widely used data for reverse engineering is gene expression where the level of expression of a gene is an important indicator of its activity under a given condition (Schlitt and Brazma, 2007). Gene expression data mostly comes from cDNA microarrays, NanoString and recently, next-

generation sequencing (RNA-seq) (Noor et al., 2013). Among these, RNA-seq which maps and quantifies transcriptomes is expected to replace contemporary methods because of its superior performance in terms of time, complexity and accuracy (Noor et al., 2013). Additionally RNA-seq appears to be more reproducible and less noisy as compared to microarrays, which were traditionally used for reverse engineering (Noor et al., 2013). Once a network is constructed using expression data, the different predicted regulatory relationships can be analyzed in detail. In a typical GRN, the regulatory relationships in a network are directional where the role of regulators is solely taken up by TFs (Macneil and Walhout, 2011). To start analyzing a network, its architecture and topology must be studied in detail as it can give important insights into the underlying biology (Levine and Davidson, 2005, Schlitt and Brazma, 2007). This includes analyzing the 'node degree', a measure for the number of relationships of a given node. The 'out-degree' of a node indicates the number of genes regulated by it (outgoing edges) and the 'in-degree' of a node is the number of genes regulating it (incoming edges). Nodes with a high out-degree are referred to as 'TF hubs' as they regulate a large number of target genes (Luscombe et al., 2004, Deplancke et al., 2006, Yu and Gerstein, 2006). Alternatively, nodes with a high in-degree are referred to as 'gene hubs' as they are regulated by many TFs. A collection of hubs is termed a module which is useful in giving an insight into tightly regulated biological processes (Levine and Davidson, 2005, Huang et al., 2009c, Macneil and Walhout, 2011). Interestingly, the modular structure of GRNs has been particularly useful in identifying circuits that control the same developmental processes in different organisms and have been retained through evolution (Levine and Davidson, 2005). An example is the small regulatory circuit consisting of the genes *Krox, Otx* and *Gata* which are involved in sea-urchin endomesoderm development (Davidson et al., 2002). This small circuit is retained in starfish even after independently evolving for 500 million years (Hinman et al., 2003). This shows the importance of GRNs in understanding

developmental processes and their conservation between different species although it requires the generation of high-quality GRNs for diverse developmental processes in different species (Levine and Davidson, 2005).

An important feature of GRNs is loops which are of three major types (Walhout, 2006). 'Feed-forward' loops where a regulator controls another regulator and both regulate the same target; 'feed-back' loops where a target controls its regulator and 'self-loops' where a regulator controls itself and is maybe involved in auto-activation or auto-repression (Schlitt and Brazma, 2007, Macneil and Walhout, 2011). An example of the importance of loops is indicated by the circuit of interacting genes involved in the stabilization and promotion of endoderm specification in the sea-urchin embryo (Davidson et al., 2002). This circuit was identified using perturbation experiments (morpholinos) leading to loss of function. Within this circuit, the transcription factor *Krox* activates *Otx* which regulates genes that in turn target endoderm regulators including the *Gata* gene. *Gata* gene in turn positively regulates *Otx* in a 'feed-back' loop that promotes endoderm specification by locking the cells in a particular state.

Another important property of the network architecture is 'node betweenness' which is a measure of centrality and importance of a node in the network i.e. the number of shortest paths from one part of the network to another that pass through this particular node (Macneil and Walhout, 2011). It has been suggested that nodes with a high betweenness centrality can connect different modules.

While analyzing a GRN, it is important to keep in mind that inference methods may not necessarily indicate direct relationships between TFs and their target genes (Macneil and Walhout, 2011). To discover direct regulatory relationships, it has been suggested to

combine TFBS information with network inference (Levine and Davidson, 2005, Gong et al., 2015). Additionally, it has been argued that although GRNs are very useful in giving an insight into possible regulatory relationships, the only way to authenticate the GRN is through experimental manipulation of the predicted relationships (Levine and Davidson, 2005).

There are many different inference programs that use gene expression data to construct a GRN (Tegner et al., 2003, Margolin et al., 2006, Bansal M, 2007, Wang et al., 2008a, Mordelet and Vert, 2008, Huynh-Thu et al., 2010). The Dialogue for Reverse Engineering Assessments and Methods (DREAM) program has been established to assess the performance of different GRN inference methods (Stolovitzky et al., 2007). Some of the most widely used methods include information-theoretic approaches such as those implemented in ARACNE (Margolin et al., 2006), CLR (Faith et al., 2007) and R package minet (Meyer et al., 2008). These methods compute mutual information (MI) between pairs of genes which depends heavily on the correlation between them. One of the advantages of these methods is their ability to deal with thousands of genes even with a limited number of samples. However, the major disadvantage of these methods is that these produce an undirected network (Hache et al., 2009). Moreover, these assume that gene expression levels in different samples are independent of each other which ignores the dependencies in case of time-series data where expression levels of a gene at later time-points depend on its initial levels as well as the levels of other genes in the previous time-points. Bayesian networks are another category of GRN inference methods that use conditional probability distributions and represent nodes as random variables with edges as conditional dependencies between them (Yu et al., 2004). One of the drawbacks of Bayesian networks is their incapacity to predict 'feed-back' loops although the inferred network is directional unlike ARACNE (Hache et al., 2009). Dynamic Bayesian networks

are a class of Bayesian networks that are capable of predicting 'feed-back' loops but these can only work with time-series data (Huynh-Thu et al., 2010). Graphical Gaussian models (Schafer and Strimmer, 2005a, Schafer and Strimmer, 2005b) are probabilistic models like Bayesian networks but unlike them, they result in undirected GRNs. Regression trees have also been used by several groups for network inference (Phuong et al., 2004, Ruan and Zhang, 2006, Xiao and Segal, 2009, Huynh-Thu et al., 2010). GENIE3 (Huynh-Thu et al., 2010) is one such method that works well with large number of genes and also supports 'feed-back' loops. GENIE3 was the best performer in DREAM4 challenge where it outperformed other popular GRN inference programs [CLR (Faith et al., 2007), ARACNE (Margolin et al., 2006), MRNET (Meyer et al., 2007) and GGMs (Schafer and Strimmer, 2005b)] (Dialogue for Reverse Engineering Assessments and Methods) challenge http://dreamchallenges.org/2010-publications/). In a recent review (Hache et al., 2009), six different programs including Bayesian, Graphical Gaussian and relevance based methods were assessed on simulated time-series gene expression data. No method was capable of reconstructing the true network structure for all datasets used. In general all methods exhibited low precision i.e. a small number of true regulatory relationships were recovered with a high number of false positives.

Keeping this in mind, there are no 'best' inference programs and each method has its limitations and advantages as discussed here. Some methods work best with a small number of samples (Schafer and Strimmer, 2005a) whereas others can handle large datasets (Huynh-Thu et al., 2010). Recently, new approaches are emerging that incorporate multiple types of data for GRN inference such as TFBS and perturbation data along with gene expression but these are still in their infancy (Gong et al., 2015). Thus, it can be concluded that there is a need for improvement of current algorithms so that they

can integrate different types of data as well as a need for improvement of quality and quantity of expression data.

Due to its overall good performance in the DREAM4 challenge, GENIE3 has been used as the inference program in the present study.

1.10 Aims of the project

This work aims to understand the regulatory events as cells transit from sensory progenitor cells in the pre-placodal region to specified cells in the otic placode. While FGF signalling plays an important role in the initiation of this process, the downstream gene regulatory network is not very well understood. This project aims to provide a deeper understand of the early steps in otic induction and will address the following questions:

1. Which otic placode specific transcripts respond to FGF signalling?
2. What are otic specific enhancers for FGF-response genes?
3. What are the regulators and targets of FGF-response genes and how can they be placed in a hierarchy in an otic gene regulatory network?

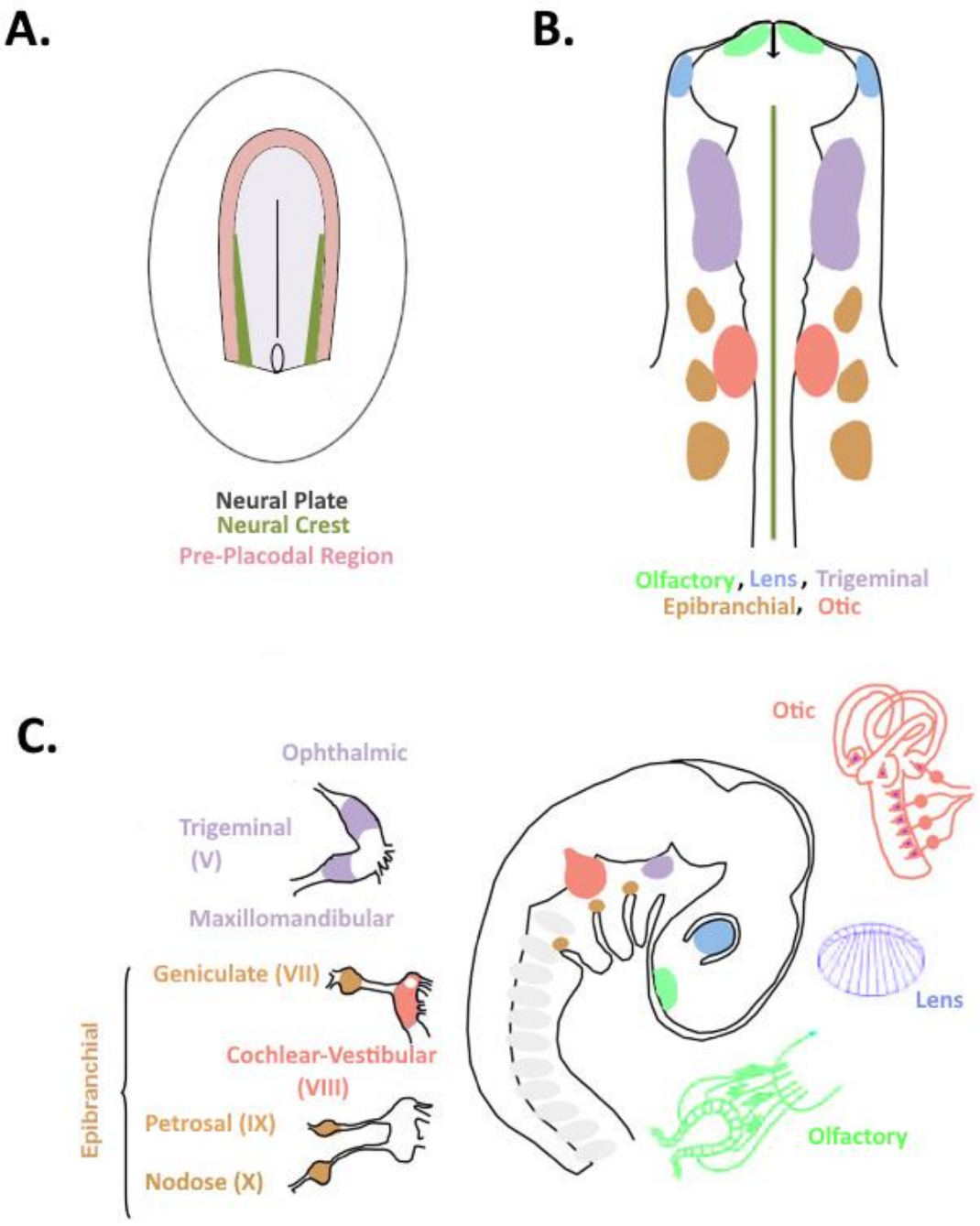


Figure 1.1 From PPR to sensory placodes
(A) At stage HH6, the pre-placodal region (light pink) is specified at the border of the neural plate (grey) and neural crest (dark green). At this stage the precursors for all sensory placodes are intermingled in the PPR. **(B)** Schematic of a 10-11 somite stage chick embryo. Individual placodes are located at different rostrocaudal locations and become distinct as ectodermal thickenings. **(C)** A 3-day old chick embryo with sensory placodes. Figure adapted from Grocott et al. 2012.

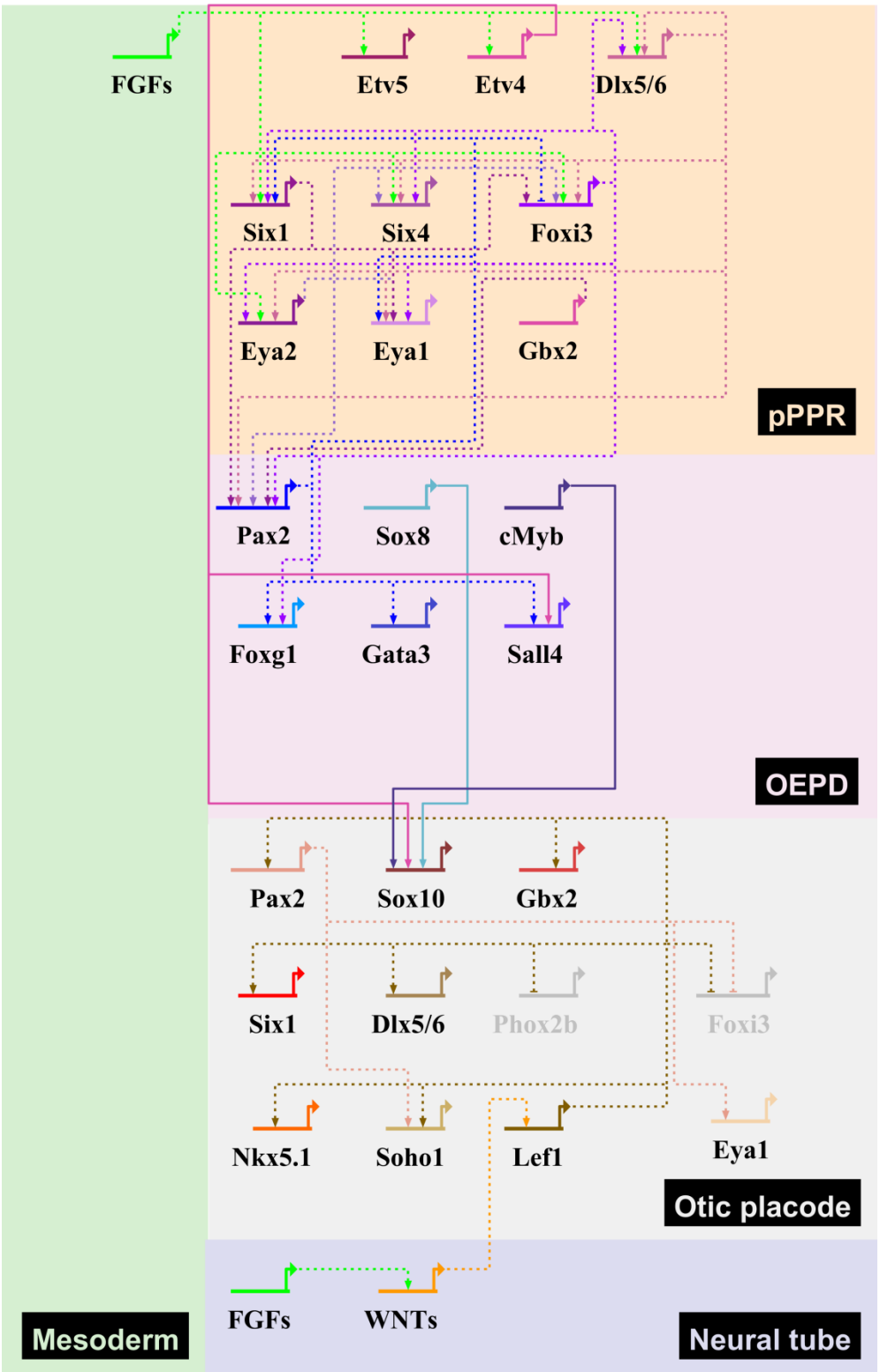


Figure 1.2 Gene regulatory network highlighting key events from PPR to otic placode

The posterior PPR transcription factors work together with mesodermal FGFs to promote the OEPD formation. At otic placode stage, WNT from the neural tube promotes otic fate whereas BMP promotes epibranchial fate (not shown here). Solid lines represent experimentally verified direct interactions. Where there is no information about direct binding, dotted lines are placed.

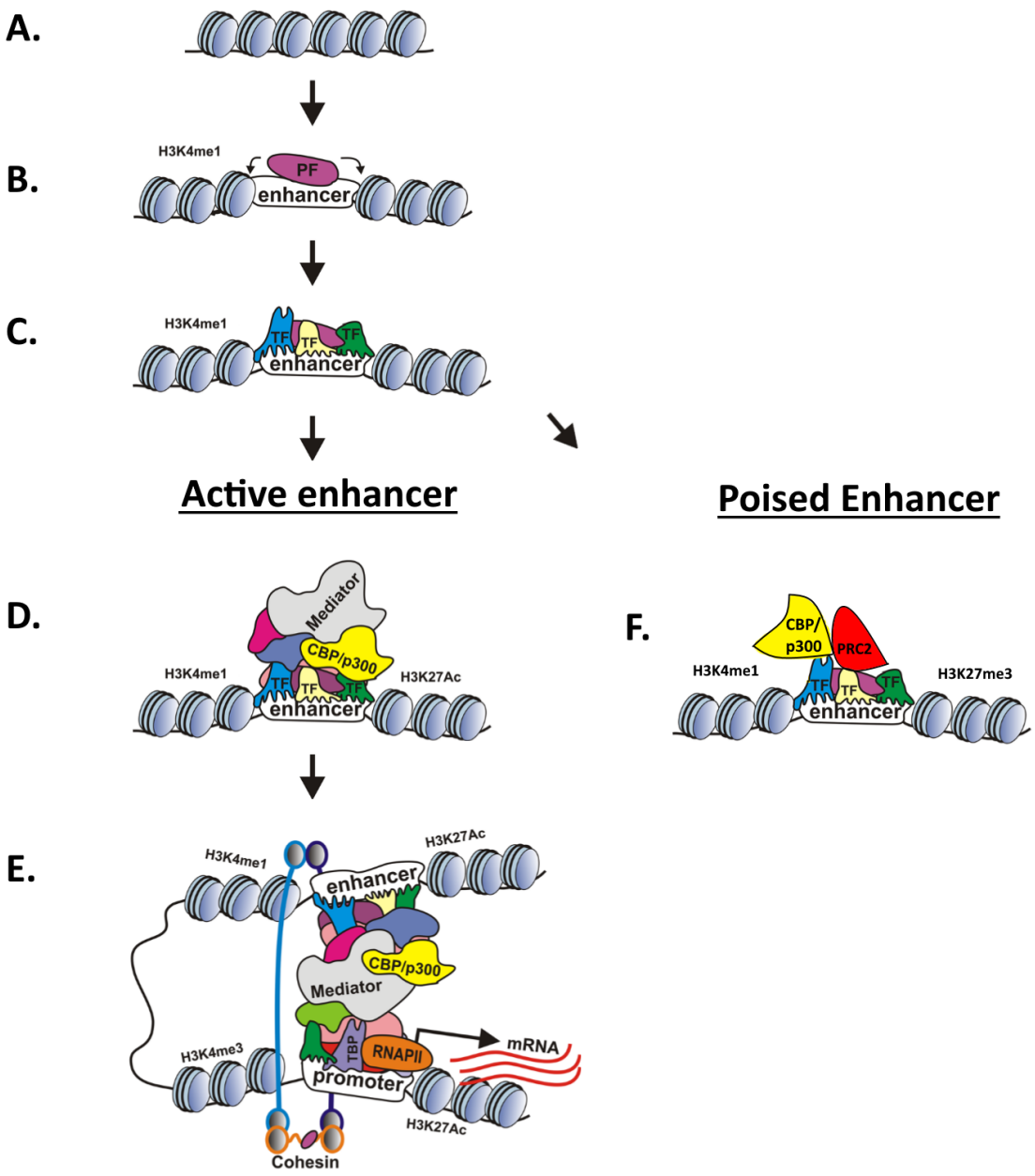


Figure 1.3 Steps of enhancer complex formation
(A) Initially chromatin is in a condensed state. **(B)** Binding of pioneer factors and mono-methylation of H3K4 displaces nucleosomes. **(C)** Pioneer factors facilitate the binding of other TFs to enhancer. **(D)** Co-activators of transcription are recruited such as the mediator complex and CBP/p300 (mediates acetylation of H3K27). **(E)** The mediator complex connects the enhancer with its corresponding promoter through looping which is facilitated by cohesin. **(F)** Poised enhancer where PRC2 complex mediates tri-methylation of H3K27 and repressive TFs keep the enhancer in a poised state. Additionally, CBP/p300 is also present at poised enhancers but not enzymatically active to deposit acetylation on H3K27. Figure adapted from Erokhin et al. 2015.

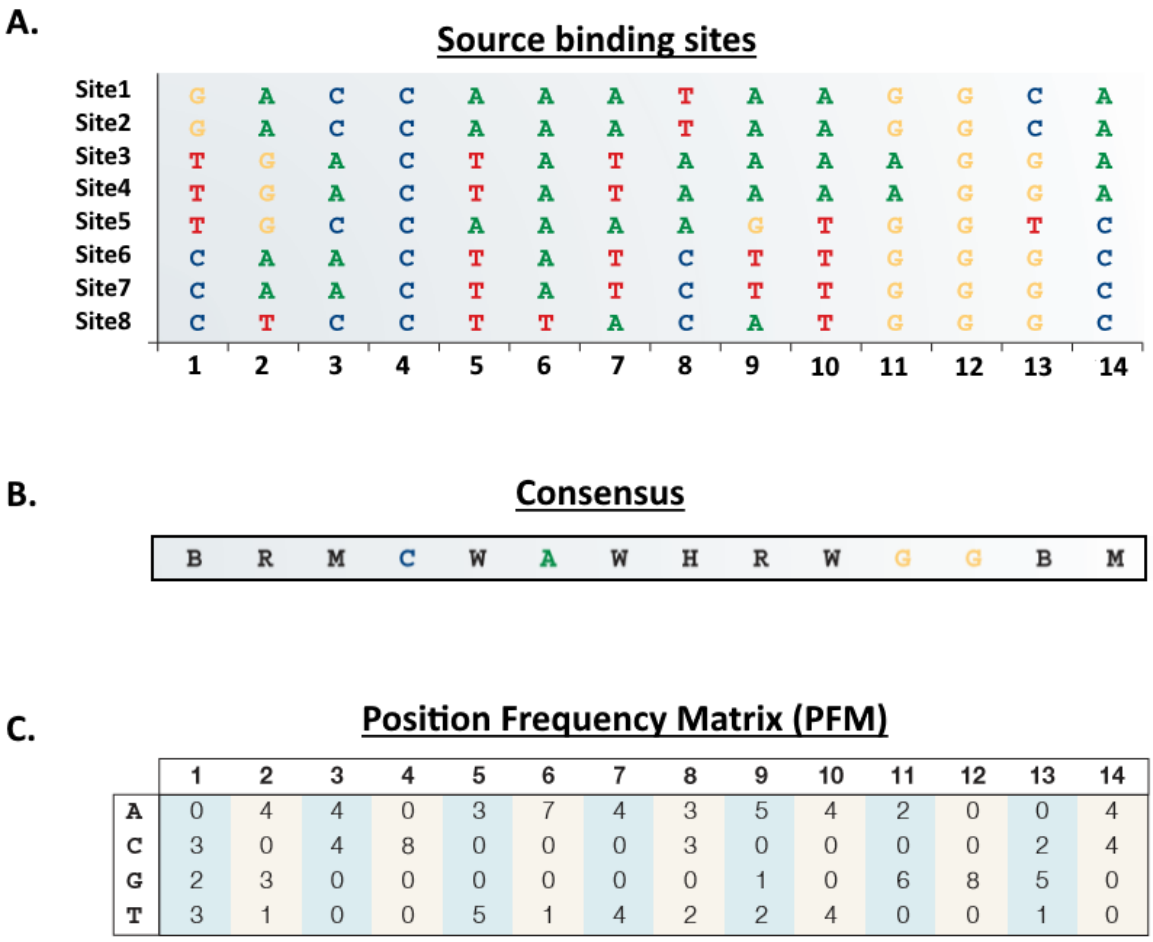


Figure 1.4 Ways of representing transcription factor binding sites
(A) A set of experimentally validated TFBSs are aligned. **(B)** Consensus sequence model: Based on alignment in **(A)**, each position is represented by a symbol that explains the nucleotides present at that position using the following code: B: (C, G, T), R: (A, G, T), M: (A, C), W: (A, T), H: (A, T, C). **(C)** To accurately reflect the nucleotides at each position, a matrix of length *m* where *m* = length of the aligned oligonucleotides is created. It records the number of occurrence of each nucleotide at each position in the binding site. Figure from Wasserman et al. 2004.

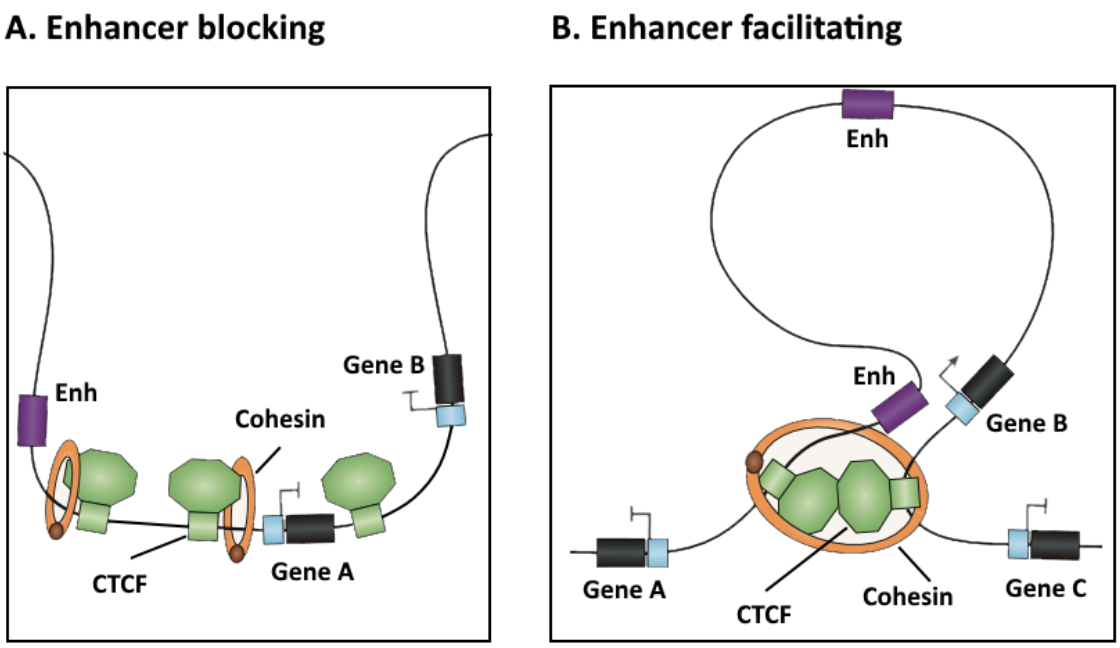


Figure 1.5 Multiple roles of CTCF
(A) CTCF and cohesin occupy regions between a gene (Gene A) and the enhancer (purple) blocking its interaction with the promoter of Gene A. **(B)** Two CTCF bound-regions loop to communicate with each other facilitating Gene B to come close to an enhancer within the loop thus regulating transcription. Figure adapted from Ong et al. 2014.

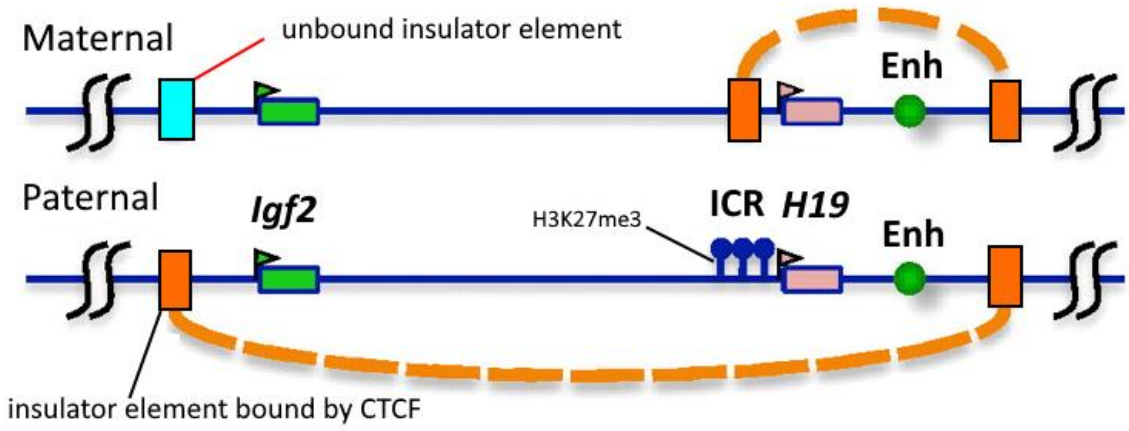


Figure 1.6 The imprinted *Igf2/H19* locus requires insulation by CTCF for normal development
In normal conditions in human and mouse, only the maternally derived allele shows *H19* expression induced by the downstream enhancer. On the paternal allele, DNA of the imprinted control region (ICR) and *H19* promoter is methylated and prevents the binding of CTCF abrogating insulator function at this site. Another upstream insulator bound by CTCF contacts the downstream insulator bound by CTCF to bring the enhancer close to the *Igf2* promoter hence activating it. Figure from Herold et al. 2012.

2. Materials and Methods

2.1 Embryo collection

Fertile chick eggs, from Winter Farm (UK), were incubated at 38°C in a humid incubator until they reached the required stage, based on Hamburger and Hamilton (Hamburger and Hamilton, 1951). To isolate the embryos, the shell was partially removed to create an opening followed by removal of excess albumen using blunt forceps. The yolk was rotated until the embryo was positioned in the centre and using scissors the vitelline membrane was cut in a square around the embryo. Using a spoon, the embryo was isolated and immersed in a dish with phosphate buffered saline (PBS) solution. The embryos were detached from the vitelline membrane and cleaned by gently blowing saline on the embryo using a Pasteur pipette. The embryos were then fixed for whole mount in situ hybridisation (see section 2.2.2).

2.2 Whole mount in situ hybridisation (WISH)

2.2.1 Preparation of Digoxigenin (DIG) – labelled riboprobes

For genes of interest, expressed sequence tags (ESTs) were obtained from Source Bioscience and used to make antisense probes (in pBlueScript II SK vector, see Table 2.1). A few plasmids were obtained from other sources (Table 2.1). Plasmid identity was verified by sequencing (DBS Genomics, Durham University; Source Bioscience Sequencing, Cambridge). The insert was amplified using M13 forward and reverse primers and Taq DNA polymerase (see Table 2.2 for PCR reaction). An aliquot (1/20th) of the PCR reaction was analyzed by agarose gel electrophoresis to check the amplified product. For ESTs, antisense DIG-labelled probes were generated using T3 RNA polymerase (Promega). Transcription reaction was set up according to Table 2.3 and

incubated for 2 hrs at 37°C. Next, 1µl of RQ1 DNase (RNase free, Promega) was added and incubated for 30 minutes at 37°C to remove the DNA template. An aliquot (1/20th) of the transcribed product was run on agarose gel to verify size and quantity. The volume of the transcription reaction was made up to 80µl with nuclease-free water and then precipitated using 1/10th volume of 4M LiCl and 2.5 volumes of 100% ethanol. Samples were incubated overnight at -20°C or for 1 hour at -80°C. The reaction was then centrifuged at maximum speed (10,000xg) to obtain a pellet and washed in 300µl of 70% ethanol. The pellet was then dried and dissolved in 80µl nuclease-free water. To purify the probe, the transcript was precipitated a second time. The pellet was then dissolved in 100µl of nuclease-free water at 65°C for 15 minutes, denatured at 95°C for 3 minutes and cooled on ice for 5 minutes. After centrifugation, 10 volumes of hybridisation buffer were added to the probe before storing it at -20°C.

Table 2.1 DIG-antisense riboprobes

Gene	Vector	Insert size (bp)	RNA polymerase	Reference
CXCL14	pBlueScript II SK	1200	T3	ChEST896P24
LMX1A	pBlueScript II SK	726	T3	ChEST609m14
SOX13	pBlueScript II SK	788	T3	ChEST437d11
ETV4	pBlueScript II SK	1500	T3	(Barembaum and Bronner-Fraser, 2010)
FOXI3	pBlueScript II SK	1000	T3	(Khatri and Groves, 2013)
GBX2	pBlueScript II SK	500	T3	(Martyn Goulding)
HESX1	pBlueScript II SK	750	T3	Obtained from lifetechnologies (http://www.lifetechnologies.com)
SPRY2	pBlueScript II SK	700	T3	(Blentic et al., 2008)
SPRY1	pBlueScript II SK	797	T3	Obtained from lifetechnologies (http://www.lifetechnologies.com)

Table 2.2 PCR reaction

PCR composition	
5x GoTaq Buffer (Promega)	2µl
dNTP Mix (10mM each) (Roche)	0.2µl
M13 Forward Primer (10µM)	1µl
M13 Reverse Primer (10µM)	1µl
GoTaq DNA Polymerase (5µg/µl) (Promega)	0.2µl
Plasmid DNA (1µg/µl)	0.5µl
Nuclease-free H20	5.1µl
Total	**10µl**
PCR condition	
1. 95°C for 3 minutes 2. 95°C for 1 minute 3. 55°C for 1 minute 4. 72°C for 1 minute 5. Repeat 2-4, 24 times 6. 72°C for 10 minutes 7. 4°C	

Table 2.3 Transcription reaction

Transcription Reaction	
Template DNA (PCR product)	0.5µl
Nuclease-free H_20	13µl
5xTranscription Buffer (Promega)	5µl
DTT (100mM) (Promega)	2.5µl
10x DIG-UTP Labelling Mix (Roche)	2.5µl
1-2mg/µl RNasin (Promega)	0.5µl
RNA polymerase (T3, T7 or SP6) (Promega)	1µl
Total	**25µl**

2.2.2 Whole mount in situ hybridisation

Chick embryos were collected as described in section 2.1 and fixed in paraformaldehyde with 2mM EGTA (Sigma) in PBS for 4 hours at room temperature or overnight at 4°C. After fixation, the embryos were stored in 100% methanol at -20°C for a maximum of one week. Embryos were re-hydrated in decreasing concentration (75%, 50% and 25%) of methanol in PTW (PBS with 0.1% Tween-20, BDH) and washed twice in PTW for 10

minutes. Embryos older than 2 days were bleached for 1 hour in 6% H_2O_2 in PTW, and further rinsed in PTW three times for 10 minutes. The embryos were then incubated in proteinase-K (10µg/ml, Sigma) according to their stage: HH4-7 for 16 minutes, HH8-10 for 20 minutes and older stages for 30 minutes. Further to this, they were washed in PTW and incubated in post-fixing solution (4% formaldehyde in PTW and 0.1% glutaraldehyde) for 30 minutes at room temperature. After rinsing with PTW, these were put in hybridisation solution [50% formamide (BDH), 5mM EDTA (pH 8.0), 50µg/ml yeast RNA (Promega), 2mg/l Tween-20 (100%, BDH), 5mg/ml CHAPS (Sigma), 1.3X SSC (Sodium Chloride Sodium Citrate, BHD) and 100µg/ml Heparin (Sigma)] for 2 hours at 70°C. The hybridisation solution was replaced with pre-warmed DIG-labelled antisense probe and incubated at 70°C overnight. Following day, embryos were washed at 70°C in hybridisation solution (3x30 minutes) and 20 minutes in 1:1 hybridisation solution: Tris-buffered saline containing 1% Tween-20 (TBST; 0.05M Tris, 0.15M NaCl, 1% Tween-20). The embryos were then washed in TBST (2x15 minutes) at room temperature and then incubated for 3 hrs in blocking buffer (5% heat inactivated sheep serum (Sigma), 1mg/ml BSA (Sigma) in TBST). Blocking buffer was then replaced with anti-DIG antibody solution (0.2/0.4µg/ml sheep IgG-AP, Roche, diluted in 1:5000 blocking buffer) and incubated overnight at 4°C. The next day, embryos were washed all day with TBST to remove un-bound antibody and later incubated twice for 10 minutes in developing buffer NTMT (5M NaCl, 2M Tris-HCl (pH 9.5), 2M $MgCl_2$, 1% Tween-20) and then with NTMT containing NBT (Nitro Blue Tetrazolium, Sigma) and BCIP (5-Bromo-4 Chloro-3 Indodyl Phosphate, Sigma) as substrates (4.5µl NBT, 50mg/ml in 70% N, N-Dimethylformamide (DMF); 3.5µl BCIP, 50mg/ml in 100% DMF, per 1.5ml of NTMT). Embryos were protected from light and left to develop at room temperature until a dark blue colour appeared. To stop developing, stained embryos were washed in PBS

and then stored in 4% PFA at 4°C. Pictures were taken with Olympus SZX12 dissecting microscope and an AxioCam HR digital camera.

2.2.3 Wax sectioning

Embryos were incubated in absolute methanol for 10 minutes at room temperature and then in propan-2-ol for 5 minutes. Next, they were incubated in tetrahydronapthalene: wax at 60°C for 30 minutes and then 3 times in wax. The embryos were then put in a mould, covered with wax left to solidify at 4°C overnight, before being processed for 12 µm sections using a Leica RM2245 microtome. Sections were de-waxed by 2 washes Histoclear for 10 minutes each and mounted using DPX (Solmedia Laboratory Suppliers) medium. Sections were viewed using Zeiss Axiovert 300M inverted microscope and photographed using AxioCam HR digital camera.

2.3 Immunohistochemistry

Cells electroporated with fluorescein labelled morpholino (MO) were visualised using immunohistochemistry. After insitu hybridisation, embryos were fixed in 4% formaldehyde in PBS, then they were rinsed well in PBS to remove residual fixative. Embryos were then blocked in 1% goatserum, 0.5% Triton X100 in PBS for 1 hr at room temperature and incubated in anti-fluorescein antibody coupled to POD for 2-3 days. They were then washed (5x30 minutes) in PBS and further in 100mM Tris (pH 7.4) for 15 minutes, before being incubated for 5 minutes in 1 ml DAB solution (in the dark). 10µl H_2O_2 solution was added to start the reaction and embryos were left to develop in the dark until a brown colour appeared. Embryos were then rinsed several times with water to stop the reaction and fixed in 4% formaldehyde in PBS.

2.4 Whatman filter culture

The filter culture method (Chapman et al., 2001) uses a square-cut filter paper with a hole in its centre to support embryos in culture. Using a paper puncher, a hole of about 7.5mm was made into filter papers of 1.5x1.5 cm. The paper was then autoclaved. The filter paper serves as an optimal material for providing support and attachment to the vitelline membrane allowing the embryo to grow normally. First, the egg was opened and thick albumen removed as described in section 2.1. Some thin albumen was collected as culture medium (see section 2.5). After rotating the yolk to bring the embryo in the centre, a filter paper was placed on top of the vitelline membrane with the embryo in the centre of the hole. Using scissors, the vitelline membrane was cut around the filter paper and using fine forceps the filter was gently removed from the yolk. The filter carrying the embryo was then immersed in a petri dish containing Tyrode's saline and further cleaned under the microscope to remove excess of yolk. Ideally, the filter culture should not be left in Tyrode's solution for longer than half an hour as the vitelline membrane starts to detach. The embryos cultured using this method were used for electroporation (section 2.5).

2.5 Electroporation

Electroporation is used to introduce morpholinos or DNA plasmids into the chick embryo. The embryos were prepared as described in section 2.4 and transferred to an electroporation chamber (2x2mm platinum electrode) ventral side upwards. DNA [general DNA mixture: 3µg/µl reporter DNA (pTK Citrine/Cherry) and 1.5 µg/µl pCAB RFP/GFP in H_2O, 0.1% fast green] or moprholino [MO (1 mM), 0.1% fast green plus 50ng/µl pCAB vector used as a carrier; for MOs see Table 2.4)] was injected between the vitelline membrane and dorsal side of the embryo using a glass needle and air pressure. A silver electrode (2x1mm) was placed on top of the target area without touching the embryo and 5 pulses of 4V and 50ms duration with intervals of 750ms were applied using

Intracel TSS20 OVODYNE pulse generator. The embryo was then placed into a petri dish (35mm) with 1 ml of albumen collected during culture preparation. The lid of the petri dish was sealed using albumen and the embryos were then placed in the incubator at 38°C until they reached the required stage. The embryos were further processed for fixation followed by in situ hybridisation (section 2.2) or imaging. Electroporated embryos were analyzed using Zeiss Axiovert 300M inverted microscope and photographed using a Hamamatsu C4742-9S camera (Digital Pixel software). Images were processed using ImageJ 1.480 and Adobe® Photoshop® CS6 (Adobe).

2.6 LMX1A Morpholinos

Morpholinos are artificial antisense oligonucleotides (length = 25 nucleotides) that are used to perform loss of function experiments. These are conjugated to fluorescein thus allowing easy visualization of MO-electroporated cells. LMX1A gene has a single transcript containing a total of 8 exons. Two LMX1A MOs were designed using the gene tools website (http://www.gene-tools.com/) that target exon-intron boundaries and lead to exon deletion; specifically, MO LMX1A-E3 deletes exon 3 and LMX1A-E4 deletes exon 4 of LMX1A gene. The details of LMX1A MOs and control MO targeting β-globin RNA are given in Table 2.4.

Table 2.4 Morpholinos used for LMX1A *in vivo* knockdown

MO	Sequence	[MO]
Control (Lleras-Forero et al., 2013)	5'CCTCTTACCTCAGTTACAATTTATA3'	1mM
LMX1A-E3	5'TGCATCAGGCAGCCCCTTACCGGAA3'	1mM
LMX1A-E4	5'ACCCCCAGTGTCCCCATACCTTCCT3'	1mM

2.6.1 Tissue dissection, mRNA extraction and 1-step RT-PCR

To test the efficiency of exon deletion by LMX1A MO, embryos were electroporated as described above with control or LMX1A MOs 1-2 otic tissues were dissected at HH9-HH12 using a fine syringe needle (3mm, 30 half gauge; BD Microlance^{TM}3) and collected in Tyrode's saline. Tissue dissection was performed by Dr. Monica Tambalo and Ramya Ranganathan. Placodes were lysed in 100µl of lysis buffer, mRNA was extracted using RNAqueous-Micro Kit (Ambion) and eluted in 10µl elution buffer. mRNA concentration was measured with NanoDrop 2000 (Thermo Scientific). Using the QIAGEN Rotor Gene Q, 1-step RT-PCR was set up as given in Table 2.5. Primers for LMX1A were designed spanning exons 1 and 5-6 (Table 2.6) using tools from IDT (http://eu.idtdna.com/scitools/Applications/RealTimePCR) and Sigma-Aldrich (http://www.oligoarchitect.com), and tested by PCR on cDNA obtained from whole chick embryo.

Table 2.5 1-step RT-PCR reaction

PCR composition	
Reaction mix RT buffer	2µl
dNTP mix (10mM)	0.4µl
Primer mix (2mM)	2.5µl
RNase-fee H_2O	2.2µl
QIAGEN enzyme mix	0.4µl
RNasin	0.5µl
RNA template	2µl
Total	**10µl**
PCR condition	
1. Reverse-transcription for 30 minutes 2. 95^{o}C for 15 minutes 3. 94^{o}C for 1 minute 4. 60^{o}C for 1 minute 5. 72^{o}C for 1 minute 6. Repeat 3-5, 40 times 7. 72^{o}C for 10 minutes	

Table 2.6 Primers for LMX1A RT-PCR

Gene	Primers	Amplification size (cDNA)
GAPDH (control)	F 5'-TCTCTGGCAAAGTCCAAGTG-3' R 5'-TCACAAGTTTCCCGTTCTCAG-3'	135bp
LMX1A (wild-type)	F 5'-GCTTGAAGATGGAGGAGACTTT-3' R 5'-CAGAACCAGAGAGCAAAGATGA-3'	750bp
LMX1A-E3	F 5'-GCTTGAAGATGGAGGAGACTTT-3' R 5'-CAGAACCAGAGAGCAAAGATGA-3'	500bp (exon 3 deleted)
LMX1A-E4	F 5'-GCTTGAAGATGGAGGAGACTTT-3' R 5'-CAGAACCAGAGAGCAAAGATGA-3'	580bp(exon 4 deleted)

2.7 Molecular cloning of putative enhancers

Sequences of putative enhancers identified from the bioinformatics pipeline described in sections 2.10 and 2.11 were retrieved from UCSC genome browser. Primers were designed using the IDT primer quest tool (http://eu.idtdna.com/PrimerQuest; primers shown in Table 2.7). Putative enhancers were amplified from chick genomic DNA using PCR conditions in Table 2.8. The PCR products were run on a 1% agarose gel to check the correct band size. Bands were cut out and purified using Agarose gel DNA extraction kit (Roche). The candidate enhancer regions were cloned upstream of a minimal promoter (tyrosine kinase promoter, pTK) and the coding sequence for yellow fluorescent protein (pTK-Citrine). To prepare the pTK vector for T-A cloning, the plasmid was linearized using blunt-end restriction enzyme SmaI at 25°C overnight, then purified and tailed with dTTP using GoTag DNA Polymerase (Promega) and dTTP for 2 hrs at 72°C. During PCR Taq polymerase adds dATP to fragment ends, thus creating an A overhang complementary to the dTTP-tailed vector. Ligation was set up using a 1:1 molar ratio of vector and PCR product, which was found optimal. The reaction was set up using linearized vector, purified PCR product, 1µl T4 DNA ligase enzyme (Promega) and 1µl 10x T4 DNA ligase Buffer (Promega) and incubated overnight at 16°C. For transformation, 3-5µl of the ligation product were mixed with 50-100µl of DH5α competent cells and kept on ice for 25 minutes, incubated at 42°C for 30 seconds and

returned to ice for a further 2 minutes. To this 600µl LB was added and incubated in a shaker at 225 rpm, 37°C for 1 hour. This was followed by centrifugation at maximum speed. 500µl of supernatant was discarded, the pellet was resuspended and plated on an LB agar plate with ampicillin (100µg/ml). After overnight incubation at 37°C, 5-10 colonies were picked for colony PCR and liquid culture (3ml LB medium plus ampicillin). PCR was performed as indicated in Table 2.8. The product was run on a 1% agarose gel to determine size and the colonies with the correct insert were grown overnight at 37°C. Plasmids were purified using peqGOLD miniprep kit, Peqlab or peqGOLD XChange Plasmid midiprep, Peqlab (for high plasmid DNA concentration) and verified by sequencing. DNA was then used for electroporation (see section 2.3). Cloning and electroporation of putative enhancers was performed in collaboration with Dr. Monica Tambalo.

Table 2.7 Primers for putative enhancer cloning

Putative Enhancer	Size (bp)	Primers	Coordinates
CXCL14 E1	1730	F 5'-AGCCTACCAGTTGTCCTAGA-3' R 5'-CACAGTGTATTGCTTGGCTTT-3	Chr13:14642163-14643893
SPRY1 E1	1553	F 5'-CTGCCAGCTGTTTCCATTTC-3' R 5'-CTGGGCTGCATGTTGTATTTC-3'	Chr4:52750797-52752350
SPRY1 E5	494	F 5'-ACGCCTCTCTACCCTCTTT-3' R 5'-GCTGGAAGCTAGAGCCATATC-3'	Chr4:52768022-52768515
SPRY2 E1	1727	F 5'-GCAAGAGTTACATTTAAGACCCTTAG-3' R 5'-TGCCAAGATGAACTGTCTCTC-3'	Chr1:151929286-151931013
HESX1 E1	496	F 5'-CAACTGCTTTCTATAATGTGTACCAG-3' R 5'-GCGTTTGATTATCGTGCTGTC-3'	Chr12:8576385-8576880
FOXI3 E1	2138	F 5'-AAGGAACTTGGGCAGGATG-3' R 5'-GAGTTCGTTCAGGAAAGACAGA-3'	Chr4: 85594801-85596939
FOXI3 E1.A	1000	F 5'-TCTGACATTTCATCATGGCTTCA-3' R 5'-CCCTTCTTGTTGTTGTTGTGT-3'	Chr4: 85595131-85596131
FOXI3 E1.B	647	F 5'- TCTGACATTTCATCATGGCTTCA-3' R 5'- GGTCATCTGAATGACAACTGTCTC-3'	Chr4: 85595131-85595778
FOXI3 E2	1095	F 5'-GCCTTGTATGGATGTTGCTGGA-3' R 5'-AGCTGGTGAACTCAATGGTGATG-3'	Chr4: 85611260-85612355
FOXI3 E2.A	510	F 5'-TTTGGCCCTGTTCAAATGG-3' R 5'-CAGTTTGTTGATACCTTCAGTGT-3'	chr4:85611260-85611770

Table 2.8 PCR reaction

PCR composition	
5x GoTaq Buffer (Promega) dNTP Mix (10mM each) (Roche) Forward Primer (10µM) Reverse Primer (10µM) GoTaq DNA Polymerase (5µg/µl) (Promega) genomic DNA (100ng/µl) Nuclease-free H_20	10µl 1µl 1.25µl 1.25µl 0.25µl 2µl 34.25µl
Total	**50 µl**
PCR condition	
1. 95°C for 3 minutes 2. 95°C for 10 seconds 3. 60°C for 45 seconds 4. 72°C for 4 minute 5. Repeat 2-4, 29 times 6. 72°C for 10 minutes 7. 4°C	

2.8 NanoString analysis

Posterior PPR tissue was dissected and cultured in the presence and absence of FGF2 and then processed for NanoString by Dr. Monica Tambalo. She processed three replicates for each experiment and further analyzed the data according to nCounter Data Analysis Guidelines (Tambalo, 2015). The NanoString code set contains positive and negative controls (sequences with no homology to any known organisms) that were used for normalization. Positive controls account for differences in hybridisation, purification and binding efficiency. The counts for each positive control in a lane were first summed to estimate the overall hybridisation efficiency and recovery for each individual lane. The individual positive control sums were then averaged; for each column the average was divided by the sum of that column creating a normalisation factor for each lane. Next, the count of each gene in a particular lane was multiplied by the normalisation factor. Negative controls were used to remove background reads. For each lane, counts for negative controls were summed and then the sums were averaged. Then column-wise

standard deviation of all negative controls was calculated and added to the average. This constituted background correction which was then subtracted from the gene counts giving the final detected mRNA count. To consider differences in amounts of starting material, the data was further normalized using the total mRNA content in each sample. For downstream analysis, the average normalized values of the three replicates were used. A fold difference was calculated between treated and control samples and a cut-off of 1.25 was used for up-regulated genes and 0.75 for down-regulated genes. P-values were calculated using an un-paired t-test and the cut-off of <=0.05 was used to identify significant results. To identify genes with similar expression patterns, average expression values were calculated for each gene from the triplicates and hierarchical clustering was carried out in R using the package gplots (Warnes, 2015). I also employed these data to infer a gene regulatory network (for explanations, see section 2.13).

2.9 mRNA-sequencing

pPPR, OEPD and otic placode tissues were dissected and collected for mRNA-seq by Dr. Jingchen Chen and Dr. Monica Tambalo. Lens, trigeminal and non-neural ectoderm tissues were collected by Ramya Ranganathan. For pPPR, OEPD and otic samples, library preparation and paired-end sequencing was carried out in the Division of Biology, California Institute of Technology, USA. For all other samples, library preparation and sequencing was carried out at UCL Genomics Centre, Institute of Child Health.

2.9.1 Bioinformatics analysis

For the purpose of this project, mRNA-seq data were used for gene regulatory network inference (section 2.13). First, the sequence quality of mRNA-seq samples was determined using FastQC (Andrews, 2010). The first nucleotide at the 5' end was trimmed as mispriming during reverse transcription may affect accuracy at the first position of

each read. The reads were aligned using TopHat 2 (v2.0.7) to chicken genome (Galgal4.71). This was performed by Dr. Jingchen Chen. Gene annotations from Ensembl (Galgal4.71) and refGene (Nov. 2011 ICGSC Gallus_gallus-4.0/galGal4) were used to assemble cDNA fragments to transcripts using Cufflinks v2.1.1 (Trapnell et al., 2010) and differentially expressed genes were identified using Cuffdiff v2.1.1 (Trapnell et al., 2010). The resulting FPKM (Fragments Per Kilobase of exon per Million fragments mapped) values for all samples were used to infer a gene regulatory network (see section 2.15).

2.10 Chromatin immuno-precipitation (ChIP) sequencing

Posterior PPR tissue was cultured with or without FGF2 for 6 hrs and used to perform a histone ChIP-seq. The antibodies used for immunoprecipitation included anti-IgG (control), anti-H3K27ac and anti-H3K27me3. This was performed by Dr. Monica Tambalo (Tambalo, 2015). Library preparation and paired-end sequencing was done at UCL Genomics Centre, Institute of Child Health.

2.10.1 Bioinformatics analysis

First, the sequence quality was assessed using FastQC (Andrews, 2010). During each ChIP-seq experiment, an amplification step was carried out that is reported to produce mismatches at the first 9bp due to random priming (Adli and Bernstein, 2011). Therefore, these 9 bp were trimmed as the sequence quality was poor. Additionally, if the sequence quality from the 3' end was poor then further trimming at this end was done to improve alignment. The reads were aligned using Novoalign (Novocraft 2.08.01, http://www.novocraft.com/products/novoalign/) to the chick genome Galgal4.71 and uniquely aligned sequences were used for peak calling using Homer (Heinz et al., 2010). A fold change of 1.5 relative to input and a False Discovery Rate (FDR) of 0.01 were

used, as using these parameters, a known otic enhancer for *Spalt4* was retrieved. This enhancer was a good candidate to assess the ChIP-seq data as it has been shown to be regulated by *Etv4* which is a downstream effector of FGF signalling (Barembaum and Bronner-Fraser, 2010). Using Homer output, putative enhancers were identified in the following way: Regions of up to 3 kb flanked by H3K27ac peaks and devoid of H3K27me3 peaks were identified and assigned to the nearest gene using gene annotation files as described in section 2.9.1 using the 'annotatePeaks' function in Homer. Read distributions around transcription start site (TSS) or the centre of a putative enhancer were also plotted using 'annotatePeaks'. Following this, the putative enhancers for +FGF2 and -FGF2 were compared to find common and unique putative enhancers using the R package ChIPpeakAnno (Zhu et al., 2010). Putative enhancers in +FGF2 and -FGF2 were considered to be overlapping and therefore common if they had a 0 bp gap between them, otherwise they were considered to be unique to the respective condition. At this point, we had identified differentially expressed genes from the OEPD mRNA-seq by comparing it to control (whole embryo). As the +FGF2 ChIP-seq sample corresponds to the OEPD stage mRNA-seq samples, it was interesting to find out if putative +FGF2-specific enhancers found by ChIP-seq were assigned to any of the genes that are enriched in the OEPD mRNA-seq. To do so, the FPKM values for genes with putative +FGF2-specific enhancer were retrieved from OEPD stage mRNA-seq and then compared to the FPKM values of all genes in the OEPD sample using the one-sided Wilcoxon test. It was hypothesized that the mean FPKM of genes with putative +FGF2-specific enhancers is greater than mean FPKM of all OEPD genes. A p-value of 0.01 was used as cut-off. Further to this, genes with unique +FGF2 and -FGF2 putative enhancers were subjected to Gene Ontology (GO) analysis using DAVID (DAVID Bioinformatics Resources 6.7) (Huang et al., 2009a, Huang et al., 2009b).

To gain confidence in the ChIP-seq results, a second peak-caller MACS2 (Zhang et al., 2008) was used. For MACS2, an FDR of 0.05 as suggested in the MACS manual to obtain broad peaks (characteristic of histone peaks) and a default p-value of 1e-5 were used. Using these parameters, MACS2 also retrieved the *Spalt4* enhancer. Following this, the MACS2 output was overlapped with the Homer output for a few genes of interest and these overlapping putative enhancers were subsequently prioritized for experimental verification. All ChIP-seq data were viewed in the IGB browser (Nicol et al., 2009).

2.11 Enhancer prediction

Parallel to histone ChIP-seq, enhancers were predicted from DNA sequence for some FGF2-response genes (*Etv4, Foxi3, Gbx2, Cxcl14, Sox13, Spry1, Spry2* and *Hesx1*). It has been reported that regulatory elements are evolutionarily conserved (Ishihara et al., 2008b, Jumlongras et al., 2012, Huang and Ovcharenko, 2014). Here, two different methods that exploit evolutionary conservation were used to predict enhancers.

2.11.1 Identification of syntenic regions

Before predicting enhancers for FGF-response genes, regions around the gene of interest were analyzed for synteny between human, mouse and chick. Synteny is the physical co-localisation of genes in blocks that are conserved across different species (Ghiurcuta and Moret, 2014). First, taking chick as reference, regions of 10 Mb upstream and 10 Mb downstream of the FGF-response gene were obtained from the Ensembl website. Then the corresponding regions in human and mouse were obtained from the Ensembl website. For each FGF-response gene, the obtained regions between human, chick and mouse were analyzed. Within these regions, gene blocks that were highly conserved (>=75% common genes) in human, mouse and chick were considered syntenic regions. These were then prepared in the appropriate format to view in the GSV synteny viewer (Revanna et al.,

2011) to produce images where common genes in the three species are highlighted. These regions were then used for insulator binding site analysis (section 2.11.2). For FGF-response genes without syntenic neighbourhoods, 300 kb upstream and 300 kb downstream of the gene were analyzed for enhancer prediction and no insulator binding site analysis was carried out.

2.11.2 Prediction of Insulator (CTCF) binding sites

To limit regions around a gene of interest for enhancer prediction, insulator binding sites were identified. CCCTC-binding factor (CTCF) belongs to a class of architectural proteins that form boundaries around genes and facilitate interactions of regulatory elements with genes within the boundaries (reviewed in Ong and Corces, 2014). To identify insulator binding sites, synteny between human, mouse and chick was analyzed as described in section 2.11.1 and the sequences of syntenic regions were acquired from UCSC browser. The CTCF position weight matrix (PWM) was obtained from the JASPAR Transcription Factor Database (Sandelin et al., 2004) and sequences were screened for CTCF binding sites using matrix-scan (Turatsinze et al., 2008). The default p-value of 1e-4 was used to obtain significant results. Next, constitutive CTCF binding sites (CTCF sites bound across multiple cell types) were identified in human using CTCF ChIP-seq samples (52 samples from different cell lines) from ENCODE (available at University of California, Santa Cruz (UCSC) website) (Meyer et al., 2013). For human, predicted CTCF binding sites from matrix-scan were overlapped with CTCF binding sites identified from ChIP-seq (ENCODE) and only those sites were analyzed further that were present in all ChIP samples. CTCF sites closest to the 5' and 3' end of the gene of interest were considered as putative boundaries. These boundaries were then used to acquire corresponding conserved sites in mouse by analysing Multiz alignments in UCSC (Meyer et al., 2013) and using predicted CTCF binding sites in mouse. In chick, CTCF ChIP-seq

datasets (2 samples) from Martin and colleagues (Martin et al., 2011) were downloaded and using UCSC liftOver (Meyer et al., 2013), CTCF peak coordinates were converted from galGal3 to galGal4 and then overlapped with predicted CTCF binding sites. Only binding sites that overlapped between CTCF ChIP and predicted sites were taken for further analysis. After this, the constitutive CTCF binding sites in human were again used as reference to obtain the corresponding chick CTCF sites. After boundaries for each gene of interest were identified, the genes were processed for enhancer prediction.

2.11.3 Prediction of enhancers

For each gene, the human sequence within the CTCF-defined boundaries were submitted to DREiVe [(Sosinsky et al., 2007); (http://dreive.cryst.bbk.ac.uk/)] as the reference to predict conserved regulatory regions between human, horse, cow, rabbit, mouse, opossum, platypus, chick and lizard. DREiVe uses a motif-discovery algorithm to identify putative regulatory regions as clusters of short conserved motifs (8 bp) in a 300 bp window. DREiVe does not depend on sequence alignment; it is able to identify re-arrangements of motifs within regulatory elements and does not require prior information of transcription factor binding sites. Regions that were found to be conserved in 7 out of 9 species were returned as output. From this output, orthologous sequences for human, chick and mouse were retrieved for further analysis. DREiVe-predicted chick enhancers were overlapped with ChIP-identified enhancers (see section 2.10.1) in R to prioritize verification. In parallel, I obtained multiple alignments between 21 amniotes from Ensembl PECAN (Paten et al., 2008) as a second source of conservation. While selecting putative enhancers for experimental verification, either or both DREiVe and PECAN output was used for conservation analysis.

2.12 Prediction of transcription factor binding sites (TFBSs) in putative enhancers

After obtaining putative enhancers using the bioinformatics pipeline (section 2.10 and 2.11), transcription factor binding site analysis was carried out using RSAT matrix-scan (Turatsinze et al., 2008) and Clover (Frith et al., 2004a), full JASPAR (Sandelin et al., 2004) and TRANSFAC (Matys et al., 2006) libraries and customized libraries containing enriched PPR, otic, lens and trigeminal transcription factors (as compared to the whole embryo from mRNA-seq). As control, sequence shuffling was carried out a 1000 times and subsequently p-values were calculated to determine significant binding sites. For RSAT matrix-scan, the default p-value of 1e-4 and for Clover, a p-value of 0.01 was used. Once the binding sites were identified, the number of occurrences of each transcription factor in each analyzed sequence was plotted as a heatmap with hierarchical clustering using the R package gplots (Warnes, 2015). This allowed the visualization of the transcription factors with the highest number of binding sites in the sequences and also the identification of clusters of sequences having similar transcription factor binding sites.

2.13 Gene regulatory network (GRN) inference

GENIE3 R implementation (Huynh-Thu et al., 2010) was used to predict an otic gene regulatory network from various expression datasets in the lab (NanoString, mRNA-seq). GENIE3 was a good choice for GRN inference as it outperformed other popular GRN inference programs [CLR (Faith et al., 2007), ARACNE (Margolin et al., 2006), MRNET (Meyer et al., 2007) and GGMs (Schafer and Strimmer, 2005b)] in the DREAM4 (Dialogue for Reverse Engineering Assessments and Methods) challenge (http://dreamchallenges.org/2010-publications/, (Stolovitzky et al., 2007)). To create a GRN, each gene in the input file is taken as a target gene that can potentially be regulated by other genes in the input file. GENIE3 attempts to explain the expression profile of a

target gene from the expression profiles of all other genes. It then calculates the importance of each input gene (regulator) in the prediction of target gene's expression profile. The importance measure is then taken as an indication of a putative regulatory link. In this way, all regulatory links are calculated and ranked according to an importance measure where larger values indicate greater significance. To use GENIE3, first NanoString and mRNA-seq datasets were analyzed as described in sections 2.8 and 2.9. Additionally, genes with a very low value of expression (<0.00004 for NanoString and FPKM <10 for mRNA-seq) were treated as absent and their values set to 0 as their expression cannot be detected by *in situ* hybridisation. As GENIE3 input, a gene expression file and a list of transcription factors (potential regulators in the dataset) were provided. For the NanoString network, the importance measure threshold was kept at 0.005 and for mRNA-seq at 0.001 as using these thresholds, known interactions were retrieved. Both networks were viewed and analyzed in Cytoscape v 3.0.2 (Shannon et al., 2003) and overlapped with known interactions using the union and intersection functions in Cytoscape. Size and colour of the nodes were assigned according to the centrality or out-degree of each node in the network. Community clustering (Newman and Girvan, 2004) was performed on both the networks using the GLay plugin (Su et al., 2010) in Cytoscape. The advantage to using Girvan and Newman's clustering algorithm is that it does not require the number of clusters to be fixed as in other clustering techniques such as K-means (MacQueen, 1967), thus finding the natural community structure within the network. Following clustering, the resulting communities or modules were annotated with Gene Ontology (GO) and KEGG pathways using Cytoscape plugin BiNGO (Maere et al., 2005).

3. FGF-response genes during the induction of otic-epibranchial progenitors and prediction of their regulatory elements

3.1 Introduction

At neurula stages, sensory placode progenitors are located in the pre-placodal region (PPR), a strip of ectoderm surrounding the neural plate. PPR cells have the potential to contribute to different sense organs and the cranial sensory ganglia (reviewed in Schlosser, 2006, Streit, 2007, Streit, 2008, Schlosser, 2010, Patthey et al., 2014) . Initially, they express a specific set of genes such as members of the Six and Eya families (Mishima and Tomarev, 1998, Sahly et al., 1999, Esteve and Bovolenta, 1999, Pandur and Moody, 2000, Kobayashi et al., 2000, Schlosser and Ahrens, 2004, Bessarab et al., 2004, Ahrens and Schlosser, 2005, Sato et al., 2010, Pieper et al., 2011). As development proceeds, other transcription factors begin to be expressed in a subset of PPR cells thus subdividing this region along the rostro-caudal axis: for example *Pax6* is initiated in lens and olfactory precursors, *Pax3* in trigeminal and *Pax2* in otic-epibranchial (Streit, 2002, Bailey et al., 2006, Lassiter et al., 2007, Canning et al., 2008). Ultimately, a "transcription factor code" may imbue cells with a specific identity that ultimately results in the formation of different sensory placodes. It is the OEPD that gives rise to the otic and epibranchial placodes, but also contains some neural crest and epidermal precursors. It has previously been shown that the paraxial mesoderm underlying the posterior part of PPR plays an important role in inducing the OEPD (Jacobson, 1963a, Orts et al., 1971, Mendonsa and Riley, 1999, Ladher et al., 2000, Phillips et al., 2001, Leger and Brand, 2002, Kil et al., 2005) and evidence from different species implicates FGF signalling as a crucial pathway (reviewed in: Ohyama et al., 2007, Schimmang, 2007, Ladher et al., 2010). In addition, FGFs from the future hindbrain contribute to OEPD induction. However, different FGF ligands have been identified in different species with *FGF3* and *FGF10* being important in mouse (Wright and Mansour, 2003, Ladher et al., 2005,

Urness et al., 2010), *FGF3* and *FGF19* in chick (Ladher et al., 2000, Karabagli et al., 2002, Kil et al., 2005, Ladher et al., 2005) and *FGF3* and *FGF8* in zebrafish (Phillips et al., 2001, Leger and Brand, 2002, Liu et al., 2003). Abolishing FGF signalling in different species has confirmed its significant role in OEPD induction. *FGF3* and *FGF10* knockout mice form smaller otic vesicles (Ohuchi et al., 2000b, Pauley et al., 2003, Wright and Mansour, 2003, Alvarez et al., 2003). Similarly, loss of FGF receptor 2 (receptor for *FGF3* and *FGF10*) results in smaller otic vesicles in mice (Pirvola et al., 2000). In chick, *FGF8* from the endoderm promotes *FGF19* in the mesoderm that in turn promotes the OEPD induction (Ladher et al., 2010), while knockdown of *FGF8* causes loss of the OEPD and reduces levels of *Pax2* (Ladher et al., 2005). Indeed, otic and epibranchial cells show activity of ERK1/2 and ERK/MAP kinase responsive genes (Lunn et al., 2007) and inhibition of the ERK/MAP kinase pathway results in loss of *Pax2* and in the absence of the otic placode (Yang et al., 2013) indicating the importance of FGF signalling via ERK/MAP kinase. In zebrafish, *FGF3* or *FGF8* loss-of-function causes a reduction of otic markers, while loss of both leads to the almost complete absence of the placode (Phillips et al., 2001, Maroon et al., 2002, Leger and Brand, 2002, Liu et al., 2003). Furthermore, inhibition of FGF receptors by SU5402 in zebrafish causes the loss of the otic markers *Pax2, Pax8, Spry4* and *Dlx3* thus preventing otic placode formation (Leger and Brand, 2002, Maroon et al., 2002, Solomon et al., 2004).

While the evidence for FGF involvement in OEPD induction is overwhelming, the downstream gene regulatory network is less well understood. Given the importance of FGF signalling, the aim of this study is to investigate and dissect the downstream network. To place FGF regulated genes into a hierarchy, it is not only essential to provide a list of all targets, but also to identify regulatory links between them. To this end

identification of the regulatory elements that drive their expression is crucial as is the identification of the transcription factor binding sites within these.

Vertebrate genomes have a small ratio of genes to noncoding DNA and it is hypothesized that this noncoding DNA houses transcriptional regulatory signals that control the expression of a gene (Loots, 2008). To identify these regulatory elements, the most commonly applied methods include evolutionary comparisons where sequences between different species are compared for sequence similarity. These methods are facilitated with the availability of large amounts of sequence data from numerous organisms and thus have become popular as a method for the identification of regulatory elements. Phylogenetic footprinting is one such method and includes alignment-based as well as motif-discovery methods (Blanchette et al., 2002, Chiang et al., 2003, Wasserman and Sandelin, 2004, Pennacchio et al., 2007). Several studies have reported the use of evolutionary comparisons for the detection of regulatory elements (Gottgens et al., 2000, Nobrega et al., 2003, Pennacchio et al., 2007, Ishihara et al., 2008b, Sato et al., 2012, Clarke et al., 2012). Enhancers are elements that increase the transcriptional rate and typically range from 100 base pairs (bp) (Banet et al., 2000, Catena et al., 2004) to several kilobases (kb) in length (Chi et al., 2005). In addition, an enhancer can be located close to the gene or several megabases (Mb) away (Nobrega et al., 2003, Sagai et al., 2005), it may reside in intergenic or intronic regions, upstream or downstream of the transcription start site it interacts with or within another gene, and even may control multiple genes (Zuniga et al., 2004). This creates a problem of how far should the regions upstream and downstream of a gene be analyzed for identifying putative enhancers using computational methods? To answer this, it is important to understand insulation mechanisms in the regulation of gene expression.

It has been reported that CCCTC-binding factor (CTCF), a transcription factor that belongs to a class of architectural proteins, harbours insulator activity when positioned between an enhancer and a gene promoter thus preventing their communication (Bell et al., 1999, Hark et al., 2000, Phillips and Corces, 2009, Giles et al., 2010, Li et al., 2013). It forms boundaries on each side of a gene, hence facilitating its interactions with enhancers that are present only within these boundaries (for review see: Ong and Corces, 2014). To allow distant enhancers within the boundaries to interact with the promoter of the corresponding gene, CTCF mediates DNA looping thus bringing the enhancers in contact with their respective promoter (Herold et al., 2012). CTCF is highly conserved in higher eukaryotes displaying 100% homology between mouse, chicken and human (Ohlsson et al., 2001). Systematic chromatin immunoprecipitation experiments combined with high-throughput sequencing (ChIP-seq) have allowed genome-wide mapping of CTCF binding events in many tissues of different species (Schmidt et al., 2012, Wang et al., 2012) and it has been reported that one-third of these binding events are conserved across different cell types (Wang et al., 2012).

This chapter particularly focuses on identifying the earliest FGF-responsive genes and the prediction of their regulatory elements. To accomplish this, Dr. Monica Tambalo performed a series of perturbation experiments manipulating FGF signalling (activation and inhibition) and measuring the response of OEPD and PPR-specific genes at different time points (Tambalo, 2015). Having identified FGF-responsive genes, I have then used computational tools to predict CTCF boundaries around FGF-response genes, their regulatory elements and their transcriptional inputs. This will be the first step towards defining the gene regulatory network downstream of FGF signalling.

3.2 NanoString probe sets and experiments

NanoString nCounter is a method that allows the quantification of gene expression changes comparable to RTqPCR (Geiss et al., 2008). Two different NanoString probe sets were designed to assess the response of posterior PPR cells to FGF signalling. The first set includes 126 genes put together from a previous microarray screen (Tambalo, 2015) and includes known and new transcription factors expressed in the PPR and OEPD, as well as markers for specific placodes, the neural plate and neural crest cells and read-outs for different signalling pathways and housekeeping genes. The second NanoString probe set (otic set) consists of 221 genes and was designed using transcriptome data corresponding to different stages of otic development (Tambalo, 2015). This set consists of known and new otic and epibranchial transcription factors, chromatin modifiers, olfactory, lens, trigeminal, neural and neural crest transcription factors, read-outs for signalling pathways and housekeeping genes. This mix of genes from different regions of the embryo allows the analysis of the effects of FGF signalling on cell fate decisions and other signalling pathways.

Two sets of experiments were performed by Dr. Monica Tambalo. The first set assessed the response of sensory progenitors to FGF: posterior PPR (HH6) was isolated and cultured with or without FGF2 for 6, 12 and 24 hours to identify the temporal response to FGF signalling. It has been reported that *Pax2* is induced in the OEPD in response to FGF signalling (Abello et al., 2010, Yang et al., 2013), so to confirm the efficiency of the assay, *Pax2* expression was analyzed using *in situ* hybridisation (Figure 3.1 A). Although endogenously, it is *FGF19* from the mesoderm that induces OEPD formation, *FGF2* mimics this activity in vitro (Martin and Groves, 2006).

The second set of experiments was designed to assess which transcripts require FGF signalling during OEPD formation: the posterior PPR was isolated together with the underlying mesoderm (source of FGF) and cultured in the presence of DMSO (control) or SU5402 (20μM) to block FGF receptor signalling for 6, 12 and 24 hrs (Figure 3.2 A). Again *Pax2* expression was visualized using in situ hybridisation to ensure the efficiency of the assay. Three replicates were performed for each time point and the data were analyzed by Dr. Monica Tambalo as described in section 2.8. The average of normalized values for each gene was used in all downstream analyses.

3.3 FGF signalling is sufficient to induce OEPD genes

A number of genes were found to be up and downregulated after treatment with FGF2. The expression levels of these genes are plotted in Figures 3.1 B, C and D. Among these genes, it is evident that PPR and OEPD genes (*Etv4/5, Foxi3, Gbx2 and Pax2*) are upregulated rapidly just after 6 hours, as is the chemokine *Cxcl14* (Figure 3.1 B). These transcripts are maintained after 12 and 24 hours FGF exposure (Figure 3.1 C-D). Other late otic and epibranchial factors known to be associated with FGF signalling such as *Hesx1*(Abe et al., 2006), *Hey2* (Doetzlhofer et al., 2009) and *Foxg1* (Yang et al., 2013) also respond positively to FGF. On the other hand, FGF represses *Pax6* at the earliest time point thus preventing the initiation of the lens programme, which is reported to be the ground state of PPR (Bailey et al., 2006, Lleras-Forero et al., 2013). Genes like Follistatin-like 4 (*Fstl4*) and *Ptpru* are initially downregulated at 6 hours, but become upregulated after 24 hours' treatment with FGF. *Fstl4* is initially present in the anterior PPR but, at later stages, it is expressed in the otic placode (Lleras-Forero, 2011, Lleras-Forero et al., 2013), again indicating that FGF promotes posterior PPR and OEPD at the expense of anterior fate. Other genes repressed by FGF at 6 hours include *Sstr5, Gpr160* and *Kremen1* all of which are expressed in the anterior PPR (Lleras-Forero, 2011, Lleras-

Forero et al., 2013, Tambalo, 2015). After 12 hours of FGF treatment, the transcriptional repressor *Sall1,* which is expressed in the otic placode at stage HH11 (Sweetman et al., 2005), is upregulated. Additionally, experiments performed with the otic NanoString set (Figure 3.3 A-B), showed *Spry1* and *Spry2* to be upregulated by FGF within 6 hrs, consistent with them being ERK/MAP kinase targets as are *Etv4* and *Etv5* (Raible and Brand, 2001, Ozaki et al., 2001). Further experiments involving the treatment of posterior PPR tissue with FGF2 and a protein synthesis blocker CycloHeXimide (10μM) confirmed both *Etv4* and *Etv5* as direct targets of FGF signalling (Tambalo, 2015). In addition, FGF also seems to modulate other signalling pathways: the WNT targets *Lef1* and *Axin2* are downregulated after FGF exposure. Together, these observations indicate that FGF potentially plays multiple roles: it promotes OEPD fate, represses anterior character and may also modulate other signalling pathways. A large number of transcripts remain unchanged after FGF treatment. These include 86 genes at 6 hrs, 101 genes at 12 hrs and 81 genes at 24 hrs in the PPR NanoString set and 178 transcripts from the otic NanoString. Thus only a small number of genes are modulated by FGF during OEPD induction.

3.4 FGF is required for expression of OEPD genes

Dr. Monica Tambalo isolated the posterior PPR together with the underlying mesoderm as the source of FGFs (see section 3.2 and Figure 3.2 A); explants were cultured in the presence of SU5402 to block FGF receptor signalling and changes in gene expression were compared to DMSO-treated controls.

As described earlier a small group of genes is induced and maintained by FGF signalling. Blocking FGF signalling confirms that FGF is not only sufficient to induce these PPR and OEPD transcripts (*Foxi3, Gbx2, Etv4/5, Cxcl14 and Pax2*), but is also necessary for their

expression. They are downregulated upon treatment with SU5402 after 6 hrs although the *Pax2* values are not significant (Figure 3.2 B). Likewise, the otic genes *Hesx1 and Eya2* are significantly downregulated after 6 hours (Figure 3.2 B) whereas some members of the *Sox* family are downregulated upon SU5402 treatment later including *Sox10* at 12 hrs and *Sox3* at 24 hrs (Figure 3.2 C-D). It has been shown already that *Sox3* requires FGF signalling (Abello et al., 2010, Yang et al., 2013) and this finding confirms that. On the other hand, *Pax6* expression levels increase when FGF receptor signalling is inhibited indicating that repression of *Pax6* requires FGF signalling during OEPD induction. Similarly, levels of *Axin2* and *Cited2* increase upon FGF inhibition (Figure 3.2 B) indicating that FGF signalling may play a role in modulating WNT signalling at this time point. Some non-neural genes such as *Gata2* and *Keratin19* also seem to be downregulated after 12 hrs (Figure 3.2 C and D). These observations indicate that among the earliest FGF-response genes, there is a small subset including *Foxi3, Gbx2, Etv4/5* and *Pax2* that require FGF to maintain or induce their expression during otic induction.

3.5 Identification of co-expressed genes during OEPD formation

To further understand the process of OEPD induction and to identify FGF-response genes with a similar expression profile, hierarchical clustering was performed using NanoString normalized gene expression data (average values from triplicates at each time point) as described in section 2.8. The clustering for FGF2-treated and control samples is shown in Figure 3.4 A. A total of 11 large clusters were obtained with cluster 3 containing FGF-dependent early OEPD genes (*Cxcl14, Foxi3, Gbx2, Etv4/5* and *Pax2*) as well as *Hesx1* suggesting that they are indeed co-regulated during otic induction. Their response to FGF is illustrated in detail when plotting individual expression profiles; expression levels are considerably increased at all time points when compared to control levels (Figure 3.4 B), However, while some genes like *Pax2* and *Gbx2* increase continuously, other transcripts

like *Cxcl14* and *Foxi3* are reduced from 12 hrs FGF exposure onwards. This reflects their normal expression in otic cells with *Cxcl14* and *Foxi3* being absent from the placode (Figure 3.6 F and O). In contrast, several members of cluster 10 are normally expressed anteriorly, in the PPR and the neural plate, and are repressed by FGF including *Pax6* (Bailey et al., 2006), *Dlx5/6* (Brown et al., 2005) and *Zfhx1b* (Dady et al., 2012). Comparison of the individual expression profiles across all time points reveals that the expression levels are considerably reduced in FGF-treated as compared to the control sample (Figure 3.4 C). This confirms previous findings that FGF represses anterior character in the PPR (Bailey et al., 2006, Lleras-Forero et al., 2012). In addition, this cluster contains the WNT target *Lef1*, which is only later expressed in the otic placode (Tambalo, 2015).

Cluster 11 and 6 mostly consist of genes that do not change in response to FGF. Overall cluster 6 transcripts are expressed at higher levels than genes in cluster 11. Cluster 2 contains transcripts, whose expression in the embryo varies, including anterior PPR markers like *SSTR5* (Lleras-Forero, 2011, Lleras-Forero et al., 2013), neural plate and neural plate border genes like *Dbx2* (Tambalo, 2015), *Zic1* (Khudyakov and Bronner-Fraser, 2009, McMahon and Merzdorf, 2010), *Sox9* (Watanabe et al., 2009), *Kremen1* and *Ptpru* (McKeown et al., 2005, Tambalo, 2015). Like cluster 10 members, transcripts in cluster 4 are repressed by FGF at early time points including *Fstl4, Irx2, Gata3* and *Foxm1*; their normal expression patterns in the embryo are dynamic. Finally, the early PPR markers *Six1* and its co-factor *Eya2* (Ishihara et al., 2008a) appear together in cluster 7, while *Six4* is in cluster 11. All three genes are already expressed at the start of the experiment. While *Eya2* responds rapidly to FGF signalling, *Six1* is only upregulated later, consistent with its increase at otic placode stages, and *Six4* does not respond to FGF at all. Thus, all three PPR genes appear to be regulated differently.

The larger otic NanoString probe set was used to analyse the response of pPPR cells to 6 hrs FGF exposure. Clustering of the data (Figure 3.3 C) reveals that *Spry1/2* (Figure 3.6 A-C; S-U) and *Sox13* (Figure 3.6 G-I) cluster with the early FGF-response genes *Gbx2, Hesx1, Pax2* and *Cxcl14* (cluster 2). In addition, members of cluster 11 are also upregulated in response to FGF, including *Rai1, Ezrin, Eya2* and *Epha4* albeit to much lower levels than cluster 2 genes (Figure 3.3 C). Thus there is a larger group of genes that responds positively to FGF at the earliest time point (Figure 3.3 A-B) and displays similar expression profiles. On the other hand, anterior (*Pax6*, *Dlx5/6*) and non-neural ectoderm genes (*Gata2, Tfap2c*) cluster together in cluster 7; likewise cluster 10 also contains aPPR transcripts like *pNoc* (Lleras-Forero et al., 2013) in addition to a number of other genes like *Gli2, Slit1* and *Gata3*. Genes in cluster 7 and 10 are negatively regulated by FGF. The majority of genes show insignificant or very small changes upon FGF treatment (clusters 3, 4, 7 and 8).

The clustering of DMSO and SU5402 samples is shown in Figure 3.5 A where cluster 10 consists of genes that respond very early to FGF including *Etv4/5*, *Gbx2* and *Foxi3*. Their expression profiles are plotted in Figure 3.5 C. All three of these are significantly downregulated upon treatment with SU5402 (Figure 3.2 B-D) although *Gbx2* is significantly affected only after 12 and later 24 hrs. Likewise cluster 11 consists of genes that are downregulated after 24 hrs including the neural genes *Sox2/3* (Rex et al., 1997), *Dlx5* (Bhattacharyya et al., 2004) and *Zhx2* and genes with dynamic expression such as *Irx1* and *Lmx1b* which are expressed in the neural as well as otic tissues (Khudyakov and Bronner-Fraser, 2009, Abello et al., 2010) (Figure 3.5 A). Some members of cluster 9 and cluster 7 are also downregulated upon SU5402 treatment including the anterior gene *pNoc* (Lleras-Forero et al., 2013), late otic gene *Hesx1* (Figure 3.6 X) and the OEPD marker *Pax2* (Streit, 2002). No significant changes are observed in clusters 1, 3, 4, 5 and 14.

Cluster 8 houses genes that are anteriorly expressed such as *Pax6* (Bailey et al., 2006) and neural genes such as *Kremen1* (Tambalo, 2015) and *Zic1* (Khudyakov and Bronner-Fraser, 2009). The individual profiles of cluster 8 are shown in Figure 3.5 B. At 6 hrs, *Pax6* is upregulated upon treatment with SU5402 (Figure 3.5 A, cluster 8 and 3.2 B) whereas *Kremen1* and *Cxcl14* are upregulated after 24 hrs (Figure 3.5 A-B, cluster 8).

In summary, these results show that there is an early group of genes including *Etv4/5, Gbx2, Foxi3, Cxcl14, Hesx1, Sox13,* and *Spry1/2* that consistently cluster together. Thus, these transcripts may be co-regulated in response to FGF signalling. The next sections will focus on identifying the regulatory elements of these genes to understand how their expression is controlled.

3.6 Synteny

Synteny is the physical co-localisation of genes into blocks that are conserved across different species (Ghiurcuta and Moret, 2014). To identify if the loci surrounding FGF-response genes are syntenic in human, mouse and chick, the following steps were taken: FGF-response genes and their flanking regions (10 Mb upstream and 10 Mb downstream of the gene) were obtained for chick from the Ensembl website. Then the corresponding regions were obtained for human and mouse from the Ensembl website and analyzed as described in section 2.11.1. Regions that had >=75% common genes between human, mouse and chick were termed as syntenic. Figures 3.7 A-H show loci containing the FGF-response genes in human, mouse and chick. Examining the *Foxi3, Gbx2, Cxcl14, Hesx1* and *Spry2* loci reveals that the regions surrounding these genes are conserved in human, mouse and chick and were therefore classified as syntenic regions (Figure 3.7 B-F). On the other hand, *Etv4, Spry1* and *Sox13* loci were not found to be conserved between

human, mouse and chick (Figure 3.7 A, G, H). Following this, the next step was to identify insulator boundaries.

3.7 Identification of insulating boundaries around FGF-response genes

As a first step towards enhancer prediction, I identified putative insulating boundaries using available CTCF ChIP-seq data as well as by predicting CTCF binding sites using RSAT matrix-scan (Turatsinze et al., 2008) and a position weight matrix for CTCF (PWM) (Figure 3.8). To predict CTCF binding sites, syntenic regions of human, chicken and mouse were acquired from UCSC browser (Meyer et al., 2013) and screened with the CTCF PWM using matrix-scan (Turatsinze et al., 2008). However, as these are only predicted CTCF sites and may or may not be occupied, CTCF ChIP-seq data for 52 human cell lines was obtained from ENCODE (available at UCSC website (Meyer et al., 2013)) was used to augment the predictions. Constitutive sites were defined as sites that were occupied in all ChIP-seq samples. Additionally, it was observed that all constitutive sites in human were also predicted by matrix-scan which increased the confidence in prediction of CTCF binding sites in chicken and mouse for which such extensive ChIP-seq data is not available. The constitutive CTCF sites closest to the 5' and 3' end of the FGF-response genes were considered as putative insulating boundaries in human (Figures 3.9-3.13 A-C). Next the corresponding insulating boundaries in mouse were obtained by analyzing predicted CTCF binding sites and conservation data (Multiz alignments, UCSC) in mouse (Meyer et al., 2013) (Figures 3.9-3.13 D).

In chick, CTCF predicted sites were first overlapped to the available CTCF ChIP-seq data from Martin and colleagues (2 samples (Martin et al., 2011)) as described in section 2.11.2. Next, the human insulating boundaries were used to obtain the corresponding chick boundaries by analyzing the overlapped CTCF sites in chick (Figure 3.9-3.13 E)

The coordinates for CTCF boundaries, identified as described above for human, mouse and chick, and the regions further analyzed for enhancer prediction, are given in Table 3.1.

Table 3.1 Predicted CTCF boundary coordinates in human, mouse and chick

Gene	Human	size	Mouse	size	Chick	size
Foxi3	Chr2:86623738-86623757 Chr2:89039706-89039725	2.4 Mb	Chr6:70768950-70768969 Chr6:71633365-71633384	864 Kb	Chr4:85137869-85137889 Chr4:85701354-85701373	563 Kb
Gbx2	Chr2:234620245-234620264 Chr2:238832240-238832260	4.2 Mb	Chr1:87944328-87944347 Chr1:91216640-91216659	3.2 Mb	Chr7:4508945-4508964 Chr7:5829008-5829027	1.3 Mb
Cxcl14	Chr5:134241475-134241494 Chr5:137412967-137412986	3.1 Mb	Chr13:55728121-55728140 Chr13:58130530-58130549	2.4 Mb	Chr13:13670903-13671303 Chr13:14988702-14988721	1.3 Mb
Hesx1	Chr3:53304332-53304351 Chr3:57850743-57850762	4.5 Mb	Chr14:26431281-26431300 Chr14:30432831-30432850	4 Mb	Chr12:6960299-6960318 Chr12:8778376-8778395	1.8 Mb
Spry2	Chr13:79965617-79965636 Chr13:86402404-86402423	6.4 Mb	Chr14:105176411-105176430 Chr14:110762779-110762798	5.5 Mb	Chr1:148646794-148646813 Chr1:152319638-152319657	3.6 Mb

3.8 Identification of enhancers using phylogenetic footprinting

Having defined putative insulator regions, I used a motif-discovery tool DREiVe [(Sosinsky et al., 2007; (http://dreive.cryst.bbk.ac.uk/)] to predict enhancers. As described in section 2.11.3, DREiVe searches for clusters of short conserved motifs of about 8 bp in a window of 300 bp. Apart from human, mouse and chick, additional species (horse, cow, rabbit, opossum, platypus and lizard) were selected to include variation and to obtain highly conserved regions as predicted enhancers.

The following regions were used as input for DREiVe: for genes where putative CTCF boundaries were predicted (*Foxi3, Gbx2, Cxcl14, Hesx1* and *Spry2*), coordinates within these boundaries were used. For *Etv4, Spry1* and *Sox13*, where no boundaries could be predicted due to the lack of synteny between human, mouse and chick, 300 kb upstream and 300 kb downstream of each gene were used for enhancer prediction. Regions that were conserved in 7 out of the 9 species used were returned as output. From the output, predicted enhancer sequences for human, mouse and chick were retrieved. DREiVe-predicted enhancers are shown in Figures 3.14 A-H, their number for each FGF response gene in Table 3.2 and their coordinates in Appendix 8.1. A large number of enhancers (Table 3.2) was predicted for some genes, particularly *Cxcl14* and *Hesx1*. This reflects the fact that in both cases CTCF boundaries are several Mbs away from the TSS and thus a large region was used for prediction.

Table 3.2 Number of predicted enhancers of FGF response genes conserved in 7 species

Gene	No. of predicted enhancers
Etv4	**8**
Foxi3	**8**
Gbx2	**6**
Cxcl14	**169**
Hesx1	**280**
Spry2	**24**
Spry1	**26**
Sox13	**6**

However, at this stage, these are only predictions and given the large number of putative enhancers, it is not useful to test their activity in vivo. To prioritize enhancers for verification, the results from predictions were complemented by experimental strategies for enhancer identification i.e. ChIP-seq for histone modifications (see Chapter 4).

As described earlier, *Foxi3, Gbx2, Etv4, Cxcl14, Hesx1, Spry1* and *Spry2* cluster together in +/- FGF experiments (Figures 3.3 C and 3.4 A). Upon analyzing the expression patterns of these genes by *in situ* hybridisation (Figure 3.6), it is evident that *Foxi3, Etv4* and *Gbx2* are expressed in the PPR, the OEPD and later in the otic placode. This suggests that they may be co-regulated and may share common upstream regulators, although *Etv4* is a direct target of FGF signalling. Because of this and because of the small number of predicted enhancers, I analyzed their transcription factor binding sites as a first step to identify potential transcriptional inputs.

3.9 Identification of transcription factor binding sites in predicted enhancers for *Foxi3, Etv4* and *Gbx2*

Putative transcription factor binding sites (TFBS) were identified in the predicted enhancers of *Foxi3, Etv4* and *Gbx2* using the JASPAR and TRANSFAC libraries and TFBS tools Clover and RSAT matrix-scan (section 2.12). A default p-value of 1e-4 for RSAT matrix-scan and 0.01 for Clover were used to identify significant binding sites in input sequences as compared to control sequences (input sequences shuffled 1000 times). The results are plotted in Figure 3.15 as a heatmap containing only TFBSs found with both Clover and RSAT matrix-scan and clustered according to TFBS. The heatmap reveals three main enhancer clusters. Putative enhancers in cluster 1 (Figure 3.15) contain few TFBSs. However, one predicted *Etv4* enhancer shows an *Ap1* binding site, which corresponds to binding of the *c-Fos* and *c-Jun* complex (Neuberg et al., 1989, Glover and Harrison, 1995) activated by the ERK/MAP kinase pathway, downstream of FGF signalling (Gruda et al., 1994, Hurd et al., 2002, Lopez-Bergami et al., 2007). This suggests that this particular enhancer may be a direct target of FGF signalling. The most striking cluster is cluster 2 (Figure 3.15): the majority of predicted enhancers have *AP1* binding sites as well as binding sites for *Sox* family members several of which are

expressed in the otic placode (McKeown et al., 2005, Abello et al., 2007, Abello et al., 2010, Tambalo, 2015). Additionally, some putative enhancers have binding sites for *Six1* and *Six4* both of which start to express at PPR stage and continue to express in the otic placode (Esteve and Bovolenta, 1999, Kobayashi et al., 2000, Pandur and Moody, 2000, Bessarab et al., 2004, Ahrens and Schlosser, 2005, Sato et al., 2010). Interestingly, there are a few putative enhancers that have binding sites for the repressor *Sall1* which starts to express at the otic placode stage (Sweetman et al., 2005). Cluster 3 (Figure 3.15) contains putative enhancers with binding sites for the otic gene *Pax2* as well as anterior genes such as *Pax6* and *Dlx*. In addition, one potential *Foxi3* enhancer contains *Etv5* binding sites suggesting that *Foxi3* is an indirect FGF target whose activation is mediated by *Etv5*.

TFBS analysis reveals that while cluster 1 has very few binding sites, the major difference between clusters 2 and 3 is that cluster 3 has binding sites for transcription factors with an anterior expression such as *Pax6* and *Otx2* whereas cluster 2 has binding sites for transcription factors with a posterior expression such as *Six* and *Sox* family members. A small hypothetical model based on information from this binding site analysis is given in Figure 3.16. Further discussion on enhancers of FGF response genes and their transcriptional inputs will be carried out in Chapter 4 after overlap of predicted enhancers of FGF-response genes with ChIP-identified enhancers.

3.10 Discussion

In this chapter, I have identified a small set of genes that are promoted by FGF early in OEPD induction (*Etv4, Foxi3, Gbx2, Cxcl14, Hesx1, Spry1/2* and *Sox13*). This subset of genes also seems to be co-expressed and to understand how these genes are regulated, I predicted their enhancers. Within this subset of genes, *Etv4, Foxi3* and *Gbx2* seem to share some transcriptional inputs.

3.10.1 FGF promotes a subset of genes during OEPD induction

As described earlier, FGF signalling plays a crucial role in the induction of OEPD (reviewed in: Ohyama et al., 2007, Schimmang, 2007, Ladher et al., 2010) and experiments on inhibition of FGF signalling in different species has confirmed its significant role in the OEPD induction, such as smaller otic vesicles in *FGF3* and *FGF10* knockout mice (Ohuchi et al., 2000b, Pauley et al., 2003, Wright and Mansour, 2003, Alvarez et al., 2003), loss of OEPD induction and reduced levels of *Pax2* in chick (Ladher et al., 2005) and reduction or loss of otic markers in *FGF3/FGF8* loss-of-function studies in zebrafish (Phillips et al., 2001, Maroon et al., 2002, Leger and Brand, 2002, Liu et al., 2003). Several otic markers have been linked to FGF signalling before (Urness et al., 2010, Yang et al., 2013). *Etv4/5*, *Gbx2, Pax2* and *Spry1/2* have been reported previously to be downstream of FGF signalling in mouse (Urness et al., 2010) and these findings are confirmed with the data presented here. Moreover, clustering the FGF-response data has made it easier to identify FGF-responsive and non-responsive genes thus providing a means to explore the data in detail. For example, a group of positively regulated genes including *Gbx2, Pax2, Spry1/2*, *Cxcl14, Hesx1, Sox13* and *Foxi3* have been found to cluster together in the PPR and otic NanoString datasets (Figures 3.3 and 3.4). Additionally, groups of neural genes such as *Zfhx1b*, *Zic1* and *Kremen1* and anteriorly expressed genes such as *Pax6* and *Dlx* which respond negatively to FGF have been found to cluster together (Figures 3.4 and 3.5). Moreover, clustering also provides a visual means to analyze the response time to FGF or SU5402 treatments indicating which genes are potential early positively-regulated targets of FGF signalling (*Etv4/5, Gbx2, Foxi3*, *Pax2, Hesx1, Cxcl14, Spry1/2, Sox13*) and potential early negatively-regulated targets (*Pax6, Sstr5, Ptpru*) and which genes are late targets of FGF signalling (*Zfhx1b, Sox10, Mynn, pNoc*). This indicates that clustering of expression data can not only help to identify genes that are possibly co-expressed but also indicate early

and late targets of FGF signalling. Using this data, a small group of co-expressed, positively regulated, early FGF targets were selected for enhancer identification.

3.10.2 Identification of regulatory elements of early FGF-response genes

Here, I've used the tool DREiVe that identifies clusters of conserved motifs between different species irrespective of their order and orientation. This circumvents the problem of identifying conserved segments in cases of rearrangements and duplications where global aligners struggle because of their dependency on order and orientation of sequences (Wasserman and Sandelin, 2004). Additionally, in a recent study (Khan et al., 2013), it was shown that DREiVe was able to predict 18 out of 25 (72%) previously known enhancers of *Sox2* (Uchikawa et al., 2003) and 9 new putative enhancers. This indicates that DREiVe performs quite well in predicting regions that may be involved in regulation of transcription. I also predicted CTCF boundaries in order to limit regions for enhancer prediction. As described in section 1.8, when placed between the promoter of a gene and an enhancer, CTCF acts as an insulating factor thus preventing communication between the promoter and enhancer (Bell et al., 1999, Hark et al., 2000, Phillips and Corces, 2009, Giles et al., 2010, Li et al., 2013). It was thus ideal to predict these boundaries to limit the number of putative enhancers for further analysis and for experimental verification. A similar pipeline was also used by Khan et al where they used computational tools and ENCODE CTCF ChIP-seq data to identify CTCF boundaries in the *Sox2* locus and found known enhancers to be present within these boundaries (Khan et al., 2013). Adopting a similar pipeline, CTCF binding sites were predicted in human, mouse and chick using RSAT matrix-scan (Turatsinze et al., 2008) and CTCF PWM (Sandelin et al., 2004). Constitutive CTCF sites were found in human by comparing ENCODE ChIP data for 52 cell lines. Then CTCF sites closest to the FGF-response gene on either end were selected as putative boundaries. The corresponding conserved CTCF

boundaries in mouse and chick were obtained using human CTCF boundaries as reference. Recently, Hi-C technology has been used to unravel the 3D architecture of human genome in nine different cell types and the data displayed in an application called "Juicebox" (Rao et al., 2014). This has allowed the identification of domain boundaries that bind CTCF. A brief comparison of the predicted CTCF boundaries with the TAD boundaries using "Juicebox" indicates an overlap for *Foxi3* in 5 cell types and for *Cxcl14* in 1 cell type. In future, a thorough investigation will be carried out for identifying insulating boundaries by incorporating data from Hi-C and CTCF ChIP-seq.

The data presented here indicates that for some genes particularly *Cxcl14* and *Hesx1*, the number of predicted enhancers exceeds 100. It is not ideal to verify this high number of predicted enhancers. Therefore, further investigation is required to prioritize verification of these. For this purpose, a ChIP-seq experiment was designed to allow identification of enhancers in otic tissue (Chapter 4). Subsequently, the predicted enhancers identified here will be overlapped with ChIP-identified enhancers to set up a priority for enhancer verification. This strategy is not only useful in reducing the number of putative otic enhancers but also more fruitful as it uses the combination of two different approaches. This will be discussed in detail in Chapter 4.

3.10.3 Early FGF response genes: *Foxi3, Etv4* and *Gbx2* share common transcriptional inputs

Since the number of predicted enhancers for *Foxi3, Etv4* and *Gbx2* were small, I predicted TFBSs in these to get an idea about their transcriptional inputs. Three large clusters were observed from the TFBS analysis (Figure 3.15) where cluster 2 seemed the most interesting with binding sites for many otic-specific transcription factors including *Sox8, Sox13, Six1* and *Six4*. This indicates that these putative enhancers are good

candidates for being active in otic cells. Additionally, it was found that most of the putative enhancers have binding sites for *Ap1* (complex formed of *c-Fos* and *c-Jun*) which is known to be activated by ERK/MAP kinase pathway, downstream of FGF signalling (Gruda et al., 1994, Hurd et al., 2002, Lopez-Bergami et al., 2007). Upon careful analysis, putative *Etv4* enhancers in this cluster are among those that have *Ap1* binding sites. It is known already that *Etv4* along with *Etv5* is one of the direct targets of FGF signalling (Tambalo, 2015) therefore it is possible that FGF is upregulating *Etv4* through the binding of *Ap1* to these enhancers. Within this cluster, *Sall1* is a repressor that begins to express at the otic placode stage (Sweetman et al., 2005) and has binding sites in *Etv4, Foxi3* and *Gbx2* enhancers. *Foxi3* is expressed in the PPR, OEPD and later in epibranchial and trigeminal regions but not in the otic placode so it can be speculated that the putative *Foxi3* enhancers with *Sall1* binding sites may be active in the PPR and OEPD but may be shut down when *Sall1* begins to express, hence removing *Foxi3* expression from the otic placode (Figure 3.6 O). However, *Etv4* and *Gbx2* are expressed in the otic placode (Figure 3.6 P-R; J-L) and therefore it is possible that their putative enhancers with *Sall1* binding sites are not active in the otic region but elsewhere as both *Etv4* and *Gbx2* are also expressed in other tissues (Figure 3.6 P-R; J-L). While cluster one did not have many binding sites, cluster 3 mostly had binding sites for anteriorly expressed genes such as *Pax6* (Bailey et al., 2006), *Dlx* (Bhattacharyya et al., 2004) and *Pax3* (McCabe and Bronner-Fraser, 2008, Khudyakov and Bronner-Fraser, 2009), neural crest genes such as *Msx1* (Phillips et al., 2006) and neural genes such as those from the *Zic* family (Hong and Saint-Jeannet, 2007). These indicate that the putative *Etv4, Gbx2* and *Foxi3* enhancers with such binding sites may be active in tissues other than otic, for example *Foxi3* is expressed in the trigeminal region possibly through the regulation of its putative enhancer with *Pax3* binding site; *Pax3* being expressed in the trigeminal region (McCabe and Bronner-Fraser, 2008, Khudyakov and Bronner-Fraser, 2009).

This indicates that TFBS analysis is useful in identifying the putative region of activity of a predicted enhancer based on its transcriptional inputs. In the next chapter, I will discuss enhancer identification through histone ChIP-seq and its overlap with predicted enhancers of FGF-response genes to prioritize verification. Further to this, I will investigate transcription factor binding sites in the overlapped enhancers which will bring us one step closer to an otic gene regulatory network.

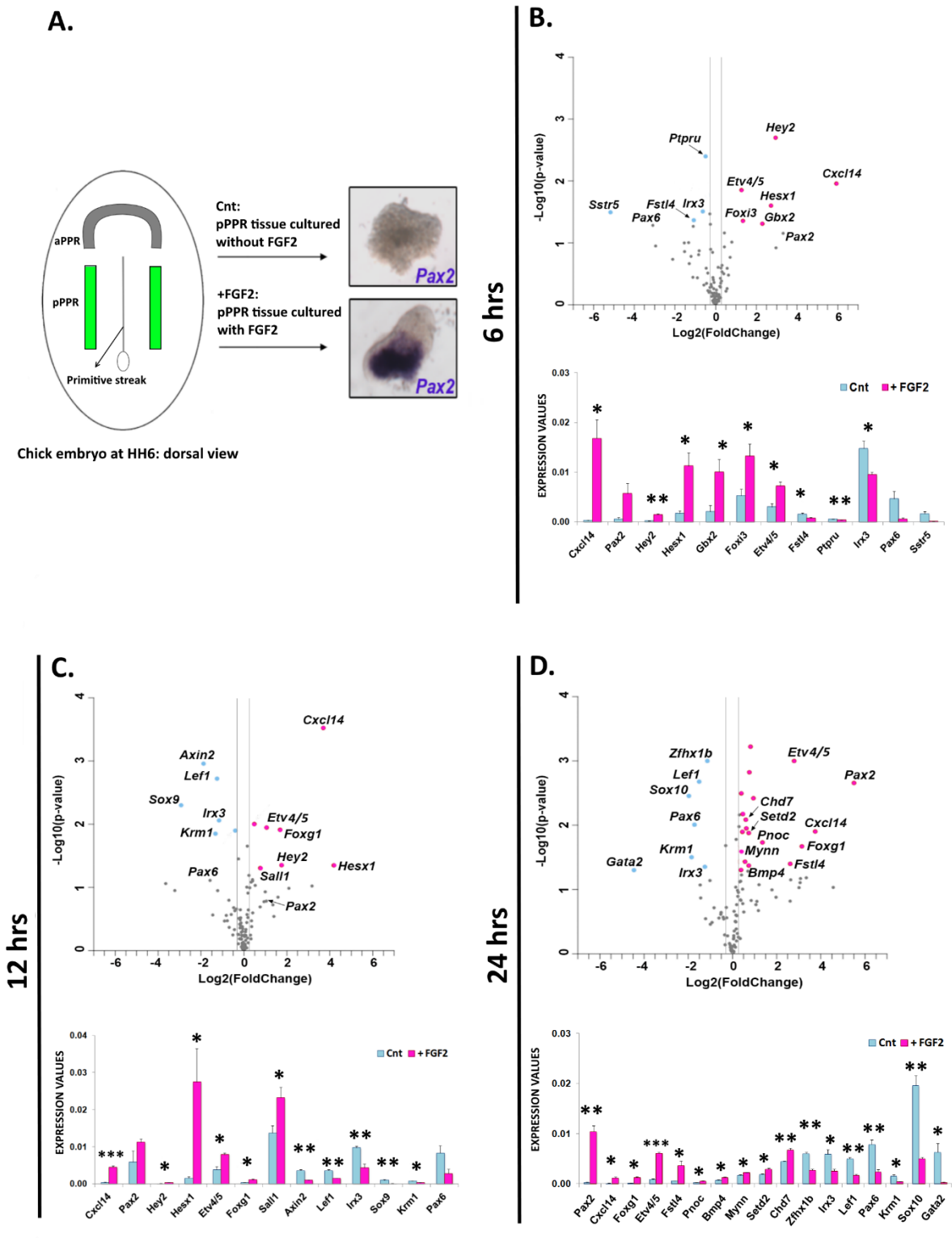


Figure 3.1 FGF2-regulated transcripts after 6, 12 and 24 hrs treatment in pPPR explants

Figure 3.1 FGF2-regulated transcripts after 6, 12 and 24 hrs treatment in pPPR explants
(A) NanoString experiment; schematic of a HH6 stage embryo; the pre-placodal region is highlighted; anterior PPR in grey and posterior PPR in green. Dissected pPPR cultured in isolation does not express *Pax2* however addition of FGF2 leads to *Pax2* expression just after 6 hrs. **(B-D)** Significantly up and downregulated genes (p-value<0.05) in +FGF2 are shown in a volcano plot (Log2FoldChange on x-axis; -Log10pvalue on y-axis). Genes that are significantly upregulated in +FGF2 are shown in pink and significantly downregulated genes are shown in blue. All other genes are shown in grey. The average expression levels of the same genes are plotted in a bar graph (Control in blue; +FGF2 in pink). Error bars represent the standard error. Asterisks (***, ** and *) indicate significance 0.001, 0.01 and 0.05, respectively.

Figure 3.2 Genes that require FGF signalling during OEPD induction
(A) NanoString experiment; schematic of a HH6 stage embryo where the pre-placodal region is highlighted: anterior PPR in grey and posterior PPR in green. The paraxial mesoderm underneath the posterior PPR is the source of FGF19 signalling in chick (yellow). Dissected pPPR and underlying mesoderm were cultured in the control condition (DMSO) where *Pax2* was induced. Alternatively, inhibition of FGF by SU5042 reduces *Pax2* expression. **(B-D)** Significantly up and downregulated genes (p-value<0.05) are shown in a volcano plot (Log2FoldChange on x-axis; -Log10pvalue on y-axis). Genes that are significantly upregulated after SU5402 treatment (or inhibited by FGF2) are shown in blue whereas genes that are significantly downregulated after SU5402 treatment (alternatively upregulated by FGF2) are shown in pink. All other genes are shown in grey. The average expression levels of the same genes are plotted in a bar graph (Control (DMSO) in pink; SU5402 in blue). Error bars represent the standard error. Asterisks (***, ** and *) indicate significance 0.001, 0.01 and 0.05, respectively.

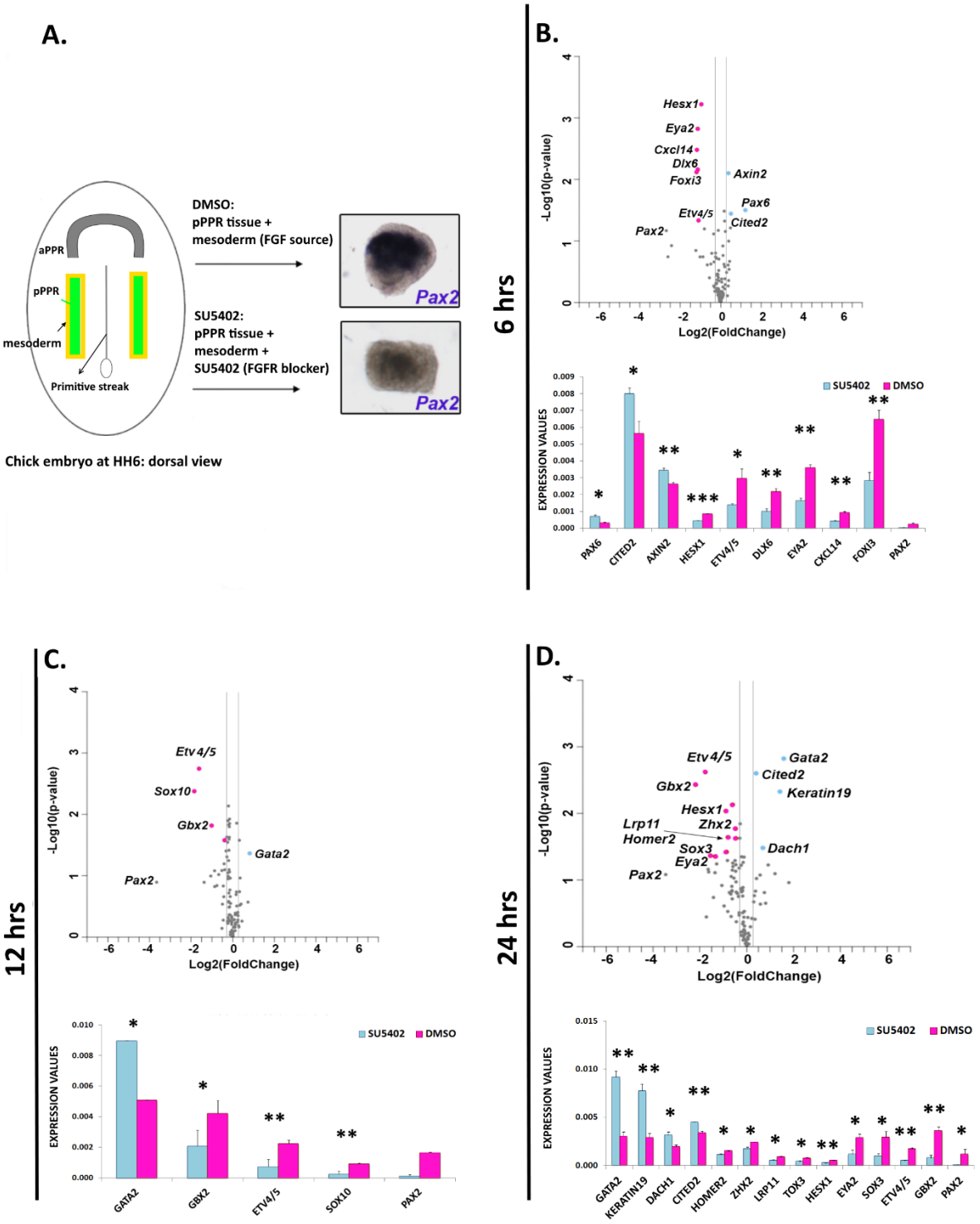


Figure 3.2 Genes that require FGF signalling during OEPD induction

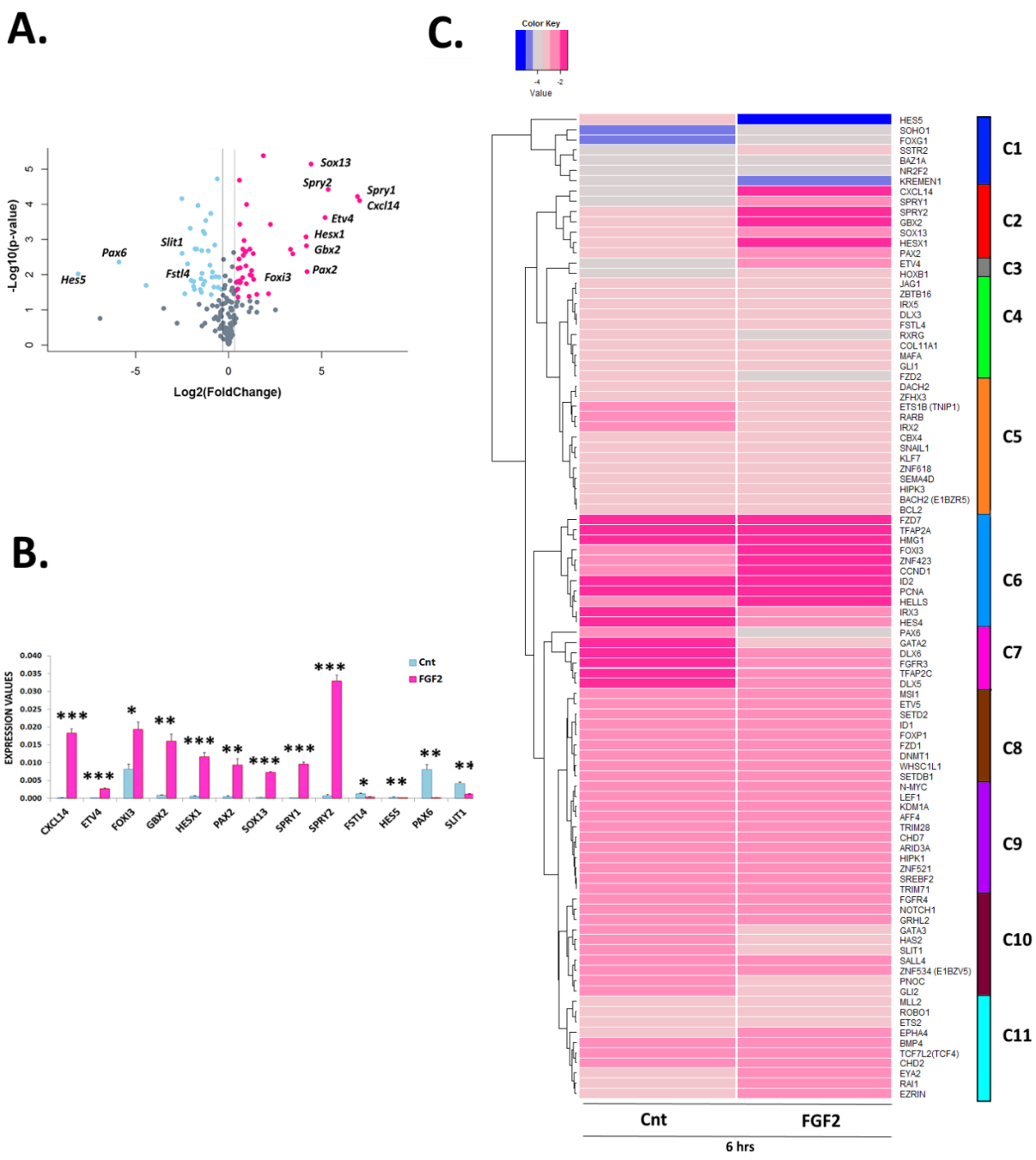


Figure 3.3 Otic NanoString set reveals new FGF response genes after 6 hrs treatment in pPPR explants

Figure 3.3 Otic NanoString set reveals new FGF response genes after 6 hrs treatment in pPPR explants
(A-B) Significantly up and downregulated genes (p-value<0.05) in +FGF2 are shown in a volcano plot (Log2FoldChange on x-axis; -Log10pvalue on y-axis). Genes that are significantly upregulated in +FGF2 are shown in pink and significantly downregulated genes are shown in blue. All other genes are shown in grey. The average expression levels of the same genes are plotted in a bar graph (Control in blue; +FGF2 in pink). Error bars represent the standard error. Asterisks (***, ** and *) indicate significance 0.001, 0.01 and 0.05, respectively. Comparison of results obtained using the PPR and otic NanoString probe set shows a robust correlation. **(C)** Hierarchical clustering of the Log2 (average expression level) of genes in +FGF2 and Control reveals 11 clusters. Low levels of expression are shown in blue whereas high levels of expression are shown in a gradient of pink.

Figure 3.4 Clustering of PPR NanoString at 6, 12 and 24 hrs treatment with and without FGF reveals co-expressed genes
(A) Hierarchical clustering of the Log2 (average expression level) of genes in +FGF2 and Control at 6,12 and 24 hrs reveals 11 clusters. Low levels of expression are shown in blue whereas high levels of expression are shown in a gradient of pink. **(C-D)** Expression profiles of the most interesting clusters C3 and C10 in +FGF2 and Control.

Figure 3.5 Clustering of PPR NanoString at 6, 12 and 24 hrs treatment with DMSO and SU5402 reveals co-expressed genes
(A) Hierarchical clustering of the Log2 (average expression level) of genes in DMSO and SU5402 at 6, 12 and 24 hrs reveal 14 clusters. Low levels of expression are shown in blue whereas high levels of expression are shown in a gradient of pink **(B-C)** Expression profiles of the most interesting clusters C8 and C10 in DMSO and SU5402.

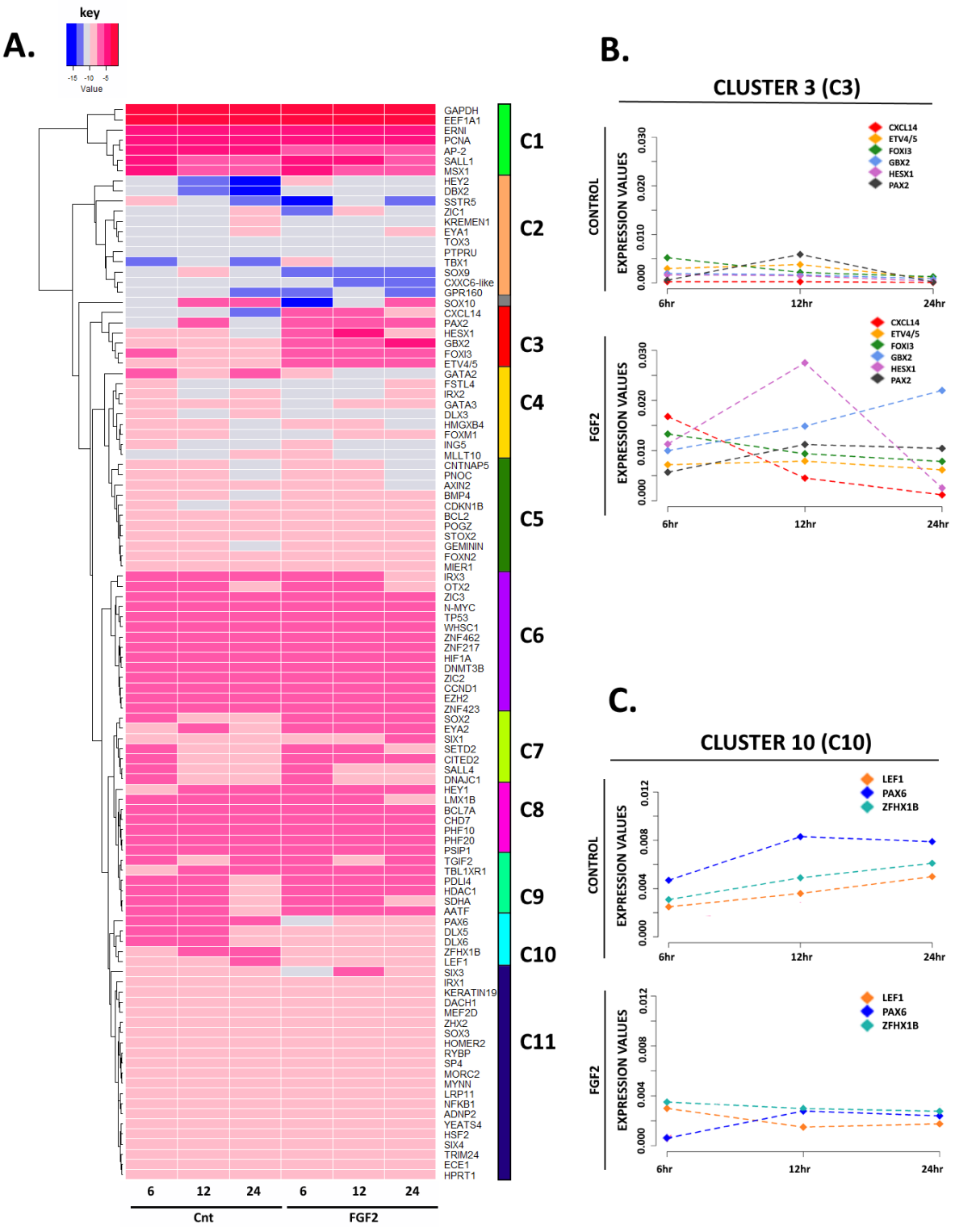


Figure 3.4 Clustering of PPR NanoString at 6, 12 and 24 hrs treatment with and without FGF reveals co-expressed genes.

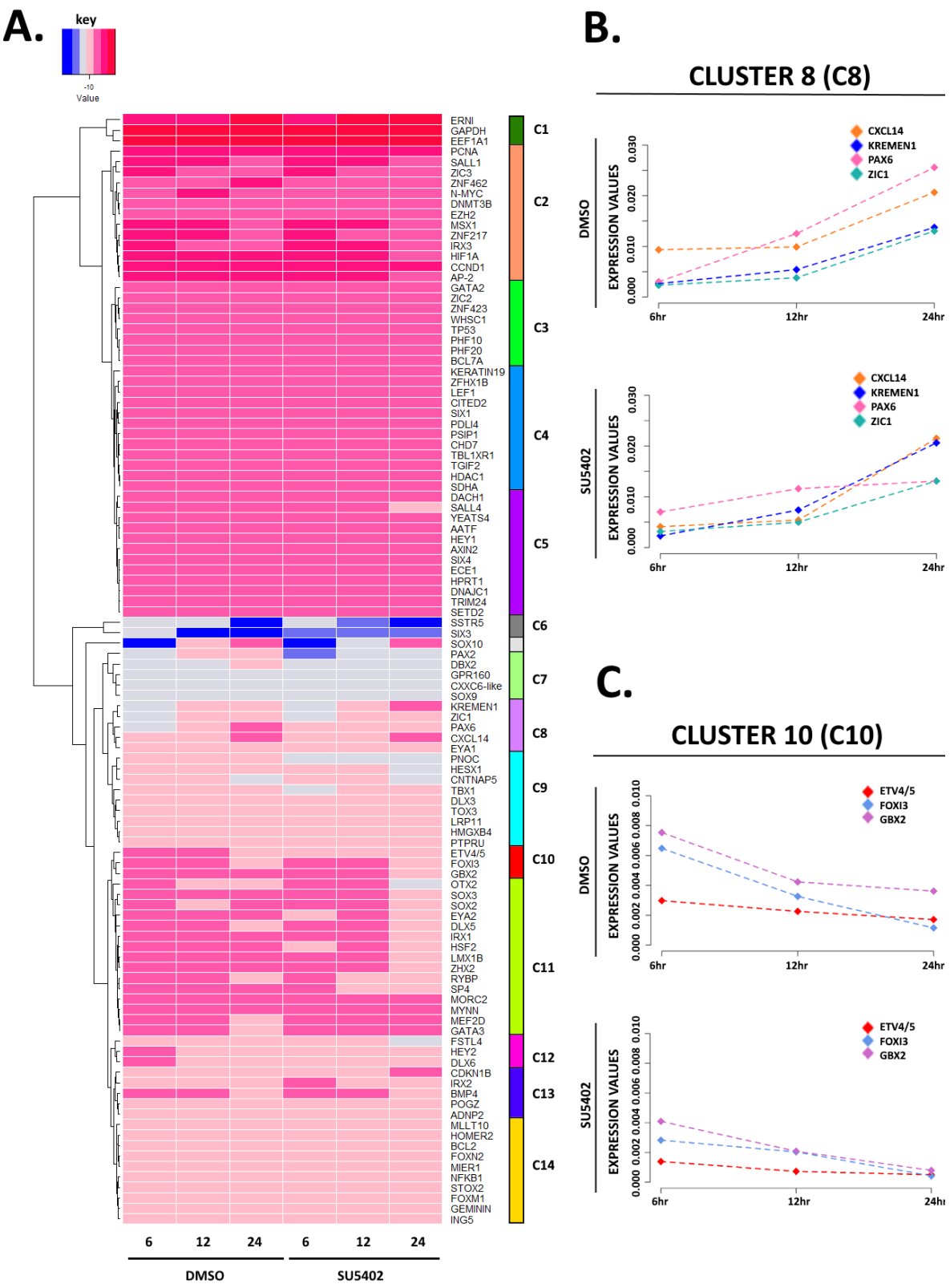


Figure 3.5 Clustering of PPR NanoString at 6, 12 and 24 hrs treatment with DMSO and SU5402 reveals co-expressed genes

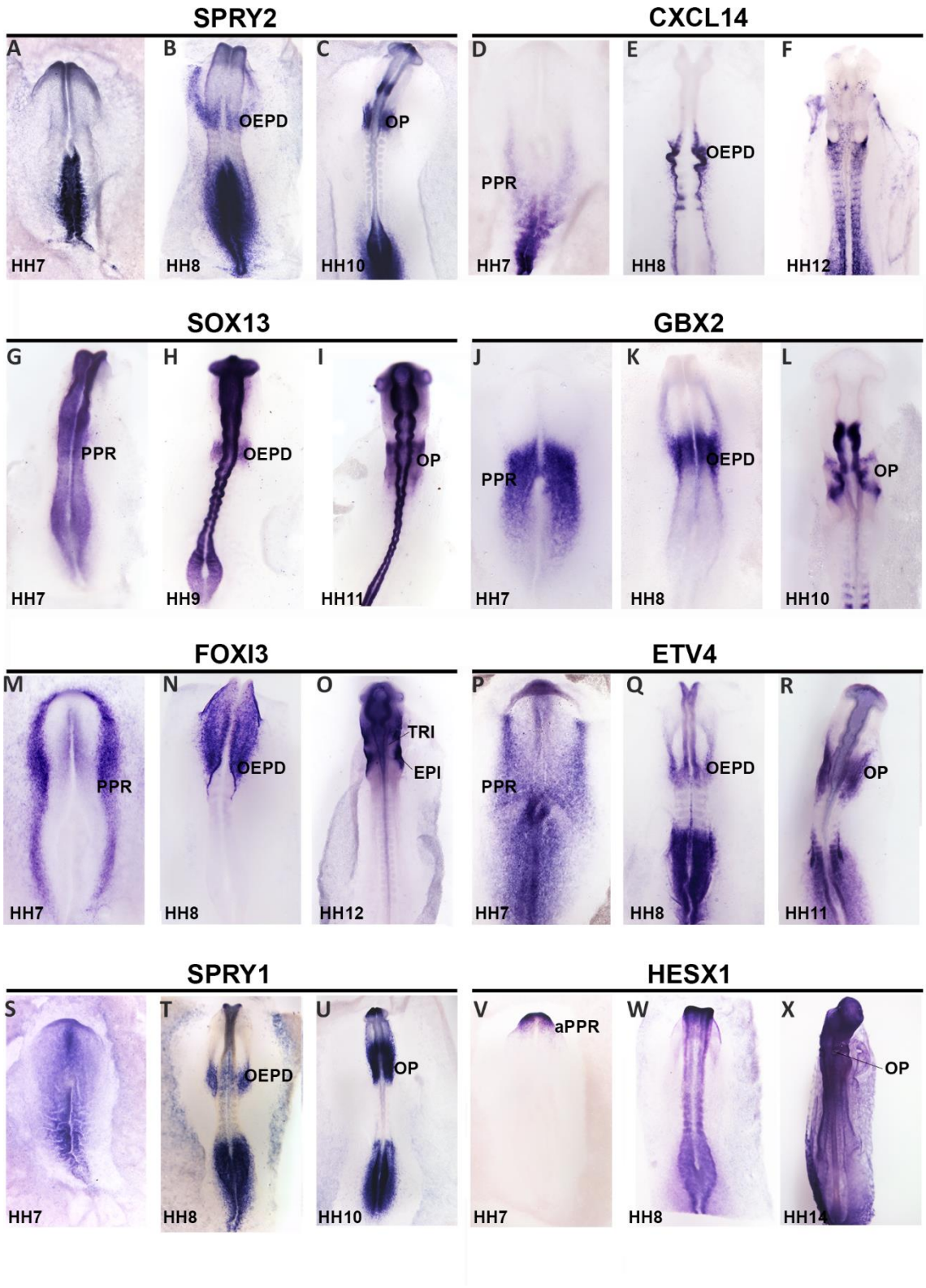


Figure 3.6 Expression patterns of early FGF response genes

Figure 3.6 Expression patterns of early FGF response genes
The known FGF target *Spry2* starts to express in the OEPD at HH8 and later expresses in the otic placode at HH10 **(A-C)**. The chemokine ligand *Cxcl14* is present in the posterior PPR at HH7 **(D)**, in the OEPD at HH8 **(E)** and is restricted to the ectoderm surrounding the otic placode at HH12 **(F)**. *Sox13* is expressed in the entire PPR and neural plate **(G)**, in the OEPD **(H)** and strongly in the otic and epibranchial placodes as well as neural tube **(I)**. *Gbx2* is expressed in the PPR and neural plate at HH7 **(J)**, then in OEPD **(K)** in the otic placode at HH10/11 **(L)**. *Foxi3* is expressed in the entire posterior PPR at HH7 **(M)** and in the OEPD **(N)** but is later lost from the otic placode; it is then expressed in the epibranchial and trigeminal placodes **(O)**. Another known FGF target *Etv4* is expressed in the PPR **(P)**, the entire OEPD at HH8 **(Q)** and is later enriched in the otic placode **(R)**. *Spry1* starts to express in the OEPD at HH8 and later expresses in the otic placode at HH10/11 **(S-U)**. *Hesx1* is initially expressed in the anterior PPR (aPPR) **(V)** and in the neural tissues at HH8 **(W)** and much later in the otic placode **(X)**.
Pre-placodal region: PPR; anterior Pre-placodal region: aPPR; otic-epibranchial progenitor domain: OEPD; epibranchial: EPI; trigeminal: TRI

Figure 3.7 Synteny in the loci containing FGF-response genes
(A-H) The regions surrounding the early FGF response genes (*Etv4, Foxi3, Gbx2, Cxcl14, Hesx1, Spry2, Spry1* and *Sox13*) in human, mouse and chick are shown in GSV synteny browser. Genes are represented as black boxes and coloured lines connect same genes from the three species. The FGF response genes are highlighted in red. There are certain cases **(A, G, H)** where there is no synteny (<75% common genes). For all other FGF response genes **(B-F),** syntenic regions were identified with >75% similarity.

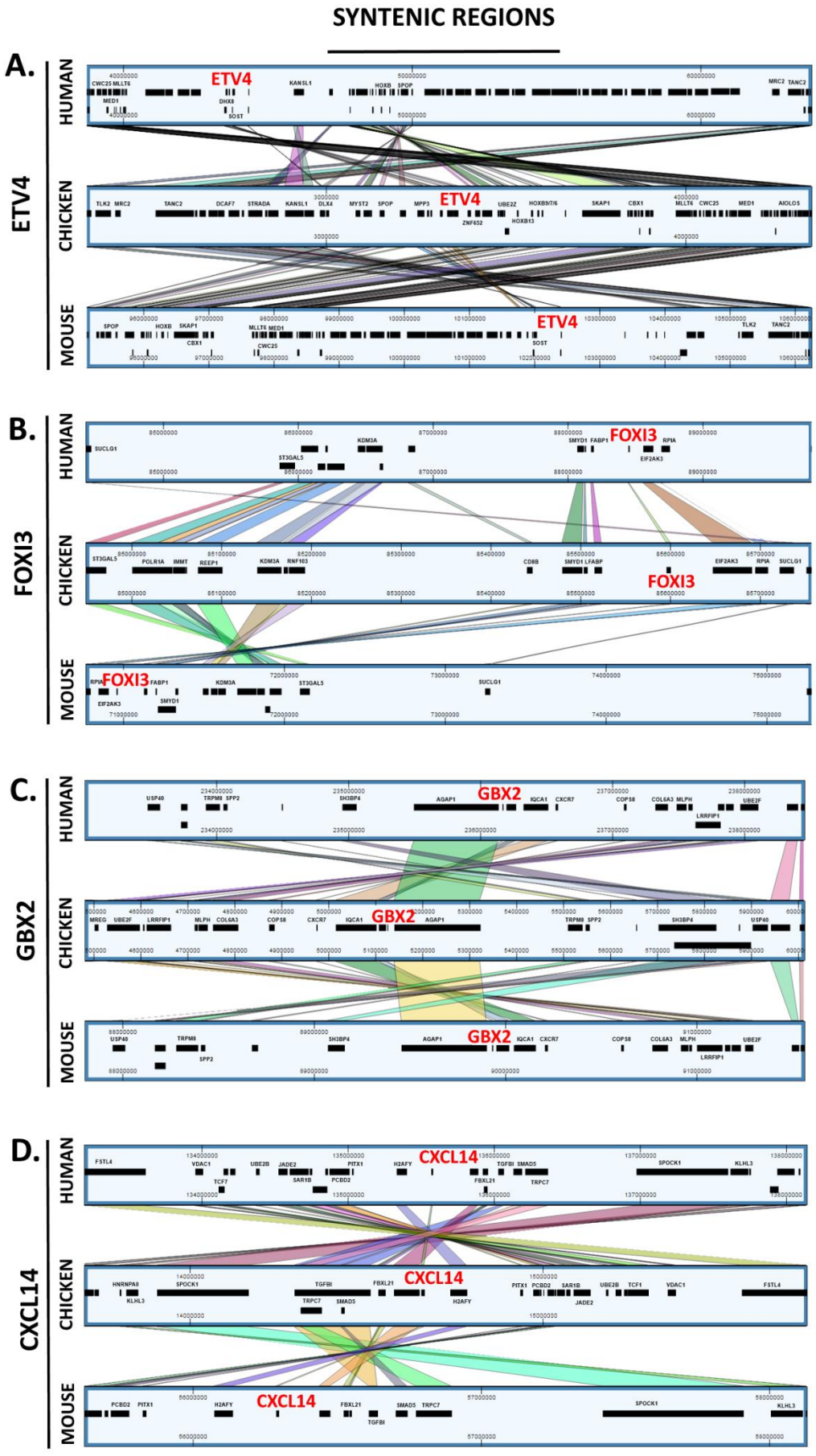


Figure 3.7 Synteny in the loci containing FGF-response genes

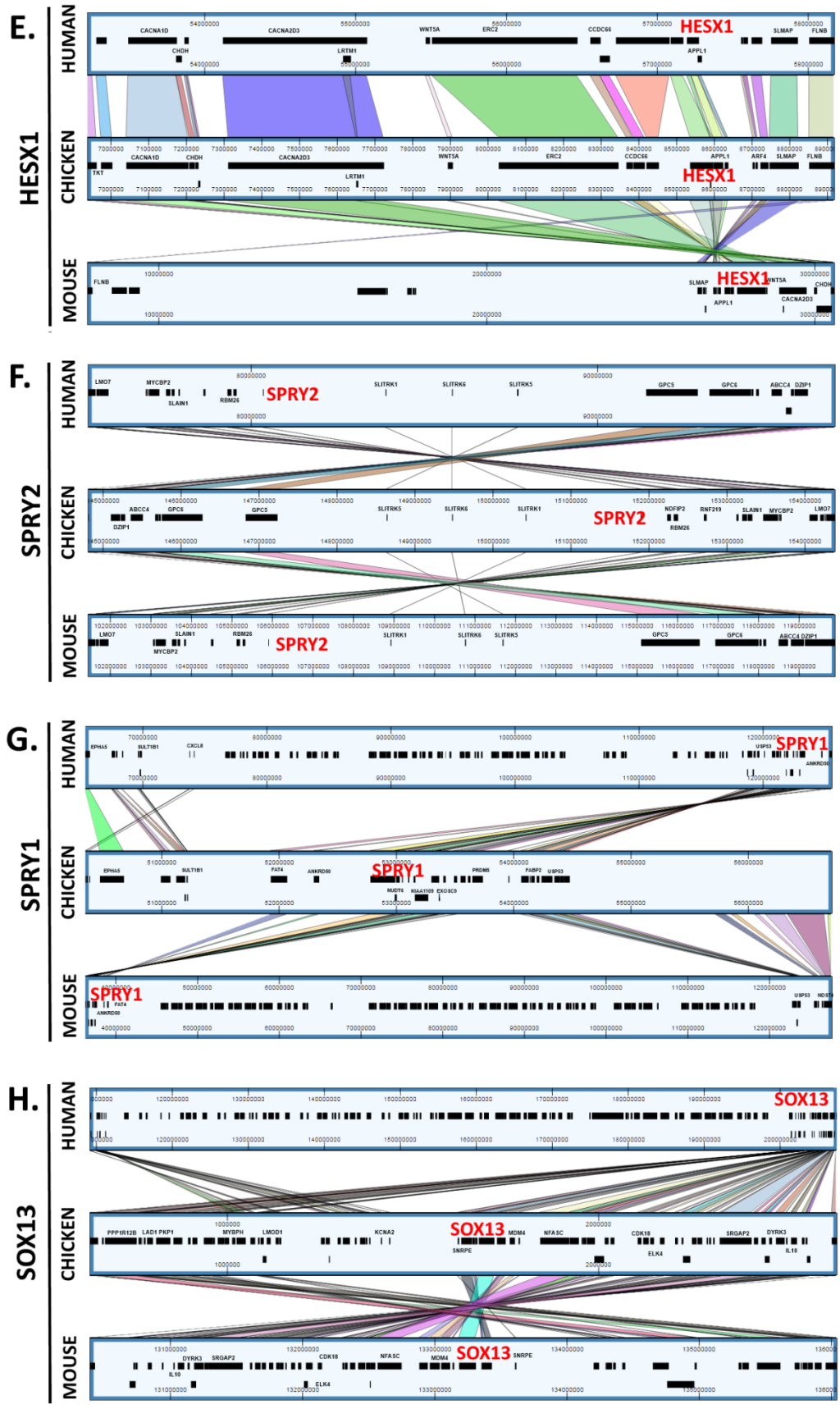


Figure 3.7 Synteny in the loci containing FGF-response genes

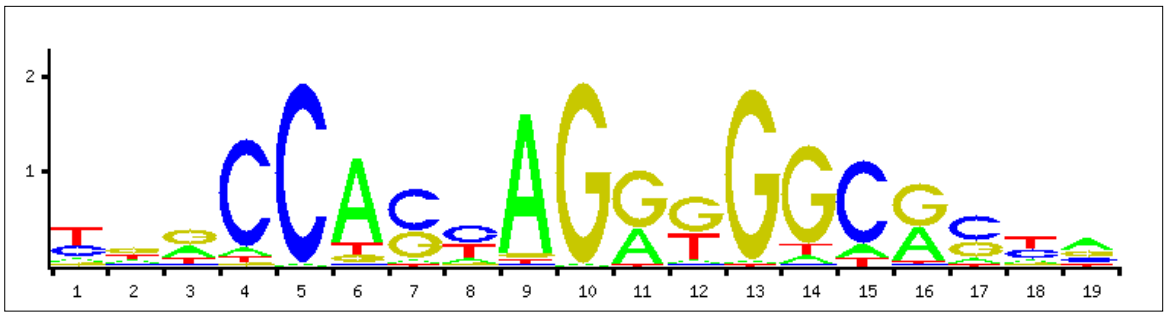


Figure 3.8 CTCF binding site
The CTCF binding site logo obtained from JASPAR transcription factor database to screen sequences surrounding and encompassing early FGF response genes for insulator boundaries using RSAT matrix-scan. The CTCF binding site is 19 bp long with variable 5' and 3' ends but a central CG-rich conserved motif.

Figure 3.9 Insulator boundaries for *Foxi3*
(A) In human, CTCF binding sites were predicted using CTCF PWM (Figure 3.8) and matrix-scan. The top track in the browser is predicted CTCF binding sites. ENCODE ChIP data from 52 cell lines (tracks in blue, green, pink, orange and black) were used to identify constitutive CTCF sites and then overlapped with predicted CTCF sites to get CTCF boundaries closest to the 5' and 3' end of the gene. The second track is RefSeq genes where *Foxi3* is highlighted in red. CTCF boundaries are also highlighted in red and a zoomed in image of each CTCF boundary is given in **B-C. (D)** The top track in the browser is predicted CTCF binding sites in mouse followed by RefSeq track where *Foxi3* is highlighted in red and then Multiz conservation track. Human CTCF boundaries were used as reference to obtain mouse CTCF boundaries (highlighted in red). **(E)** For chick, 2 CTCF ChIP datasets (top 2 tracks) from Martin and colleagues (Martin et al., 2011) were overlapped with predicted CTCF binding sites (third track) and then the overlapped sites were compared to human to obtain corresponding CTCF boundaries in chick (highlighted in red).

Figure 3.10 Insulator boundaries for *Gbx2*
(A) In human, CTCF binding sites were predicted using CTCF PWM (Figure 3.7) and matrix-scan. The top track in the browser is predicted CTCF binding sites. ENCODE ChIP data from 52 cell lines (tracks in blue, green, pink, orange and black) were used to identify constitutive CTCF sites and then overlapped with predicted CTCF sites to get CTCF boundaries closest to the 5' and 3' end of the gene. The second track is RefSeq genes where *Gbx2* is highlighted in red. CTCF boundaries are also highlighted in red and a zoomed in image of each CTCF boundary is given in **B-C. (D)** The top track in the browser is predicted CTCF binding sites in mouse followed by RefSeq track where *Gbx2* is highlighted in red and then Multiz conservation track. Human CTCF boundaries were used as reference to obtain mouse CTCF boundaries (highlighted in red). **(E)** For chick, 2 CTCF ChIP datasets (top 2 tracks) from Martin and colleagues (Martin et al., 2011) were overlapped with predicted CTCF binding sites (third track) and then the overlapped sites were compared to human to obtain corresponding CTCF boundaries in chick (highlighted in red).

Figure 3.11 Insulator boundaries for *Cxcl14*
(A) In human, CTCF binding sites were predicted using CTCF PWM (Figure 3.7) and matrix-scan. The top track in the browser is predicted CTCF binding sites. ENCODE ChIP data from 52 cell lines (tracks in blue, green, pink, orange and black) were used to identify constitutive CTCF sites and then overlapped with predicted CTCF sites to get CTCF boundaries closest to the 5' and 3' end of the gene. The second track is RefSeq genes where *Cxcl14* is highlighted in red. CTCF boundaries are also highlighted in red and a zoomed in image of each CTCF boundary is given in **B-C. (D)** The top track in the browser is predicted CTCF binding sites in mouse followed by RefSeq track where *Cxcl14* is highlighted in red and then Multiz conservation track. Human CTCF boundaries were used as reference to obtain mouse CTCF boundaries (highlighted in red). **(E)** For chick, 2 CTCF ChIP datasets (top 2 tracks) from Martin and colleagues (Martin et al., 2011) were overlapped with predicted CTCF binding sites (third track) and then the overlapped sites were compared to human to obtain corresponding CTCF boundaries in chick (highlighted in red).

Figure 3.12 Insulator boundaries for *Hesx1*
(A) In human, CTCF binding sites were predicted using CTCF PWM (Figure 3.7) and matrix-scan. The top track in the browser is predicted CTCF binding sites. ENCODE ChIP data from 52 cell lines (tracks in blue, green, pink, orange and black) were used to identify constitutive CTCF sites and then overlapped with predicted CTCF sites to get CTCF boundaries closest to the 5' and 3' end of the gene. The second track is RefSeq genes where *Hesx1* is highlighted in red. CTCF boundaries are also highlighted in red and a zoomed in image of each CTCF boundary is given in **B-C. (D)** The top track in the browser are predicted CTCF binding sites in mouse followed by RefSeq track where *Hesx1* is highlighted in red and then Multiz conservation track. Human CTCF boundaries were used as reference to obtain mouse CTCF boundaries (highlighted in red). **(E)** For chick, 2 CTCF ChIP datasets (top 2 tracks) from Martin and colleagues (Martin et al., 2011) were overlapped with predicted CTCF binding sites (third track) and then the overlapped sites were compared to human to obtain corresponding CTCF boundaries in chick (highlighted in red).

Figure 3.13 Insulator boundaries for *Spry2*
(A) In human, CTCF binding sites were predicted using CTCF PWM (Figure 3.7) and matrix-scan. The top track in the browser is predicted CTCF binding sites. ENCODE ChIP data from 52 cell lines (tracks in blue, green, pink, orange and black) were used to identify constitutive CTCF sites and then overlapped with predicted CTCF sites to get CTCF boundaries closest to the 5' and 3' end of the gene. The second track is RefSeq genes where *Spry2* is highlighted in red. CTCF boundaries are also highlighted in red and a zoomed in image of each CTCF boundary is given in **B-C. (D)** The top track in the browser are predicted CTCF binding sites in mouse followed by RefSeq track where *Spry2* is highlighted in red and then Multiz conservation track. Human CTCF boundaries were used as reference to obtain mouse CTCF boundaries (highlighted in red). **(E)** For chick, 2 CTCF ChIP datasets (top 2 tracks) from Martin and colleagues (Martin et al., 2011) were overlapped with predicted CTCF binding sites (third track) and then the overlapped sites were compared to human to obtain corresponding CTCF boundaries in chick (highlighted in red).

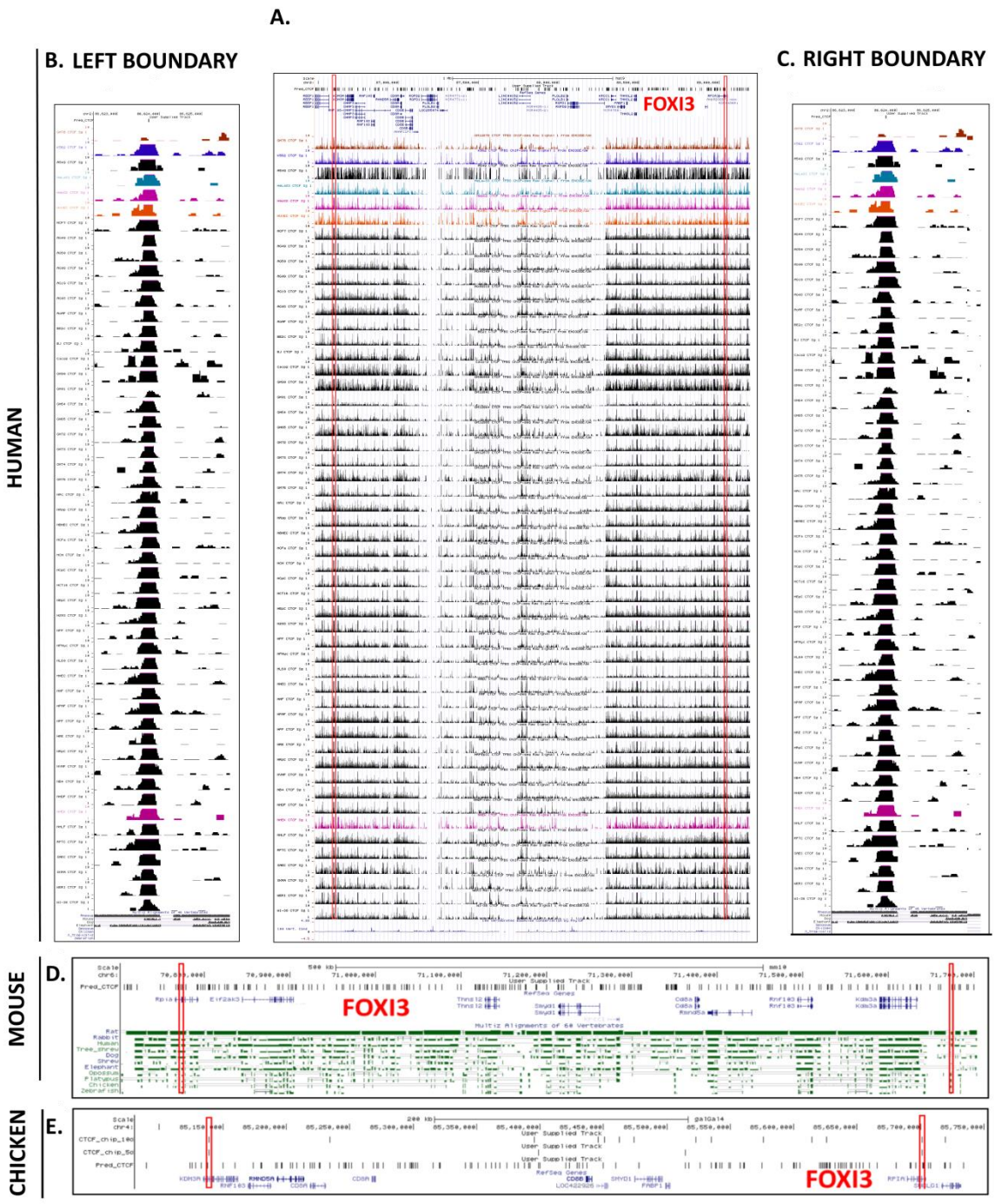


Figure 3.9 Insulator boundaries for *Foxi3*

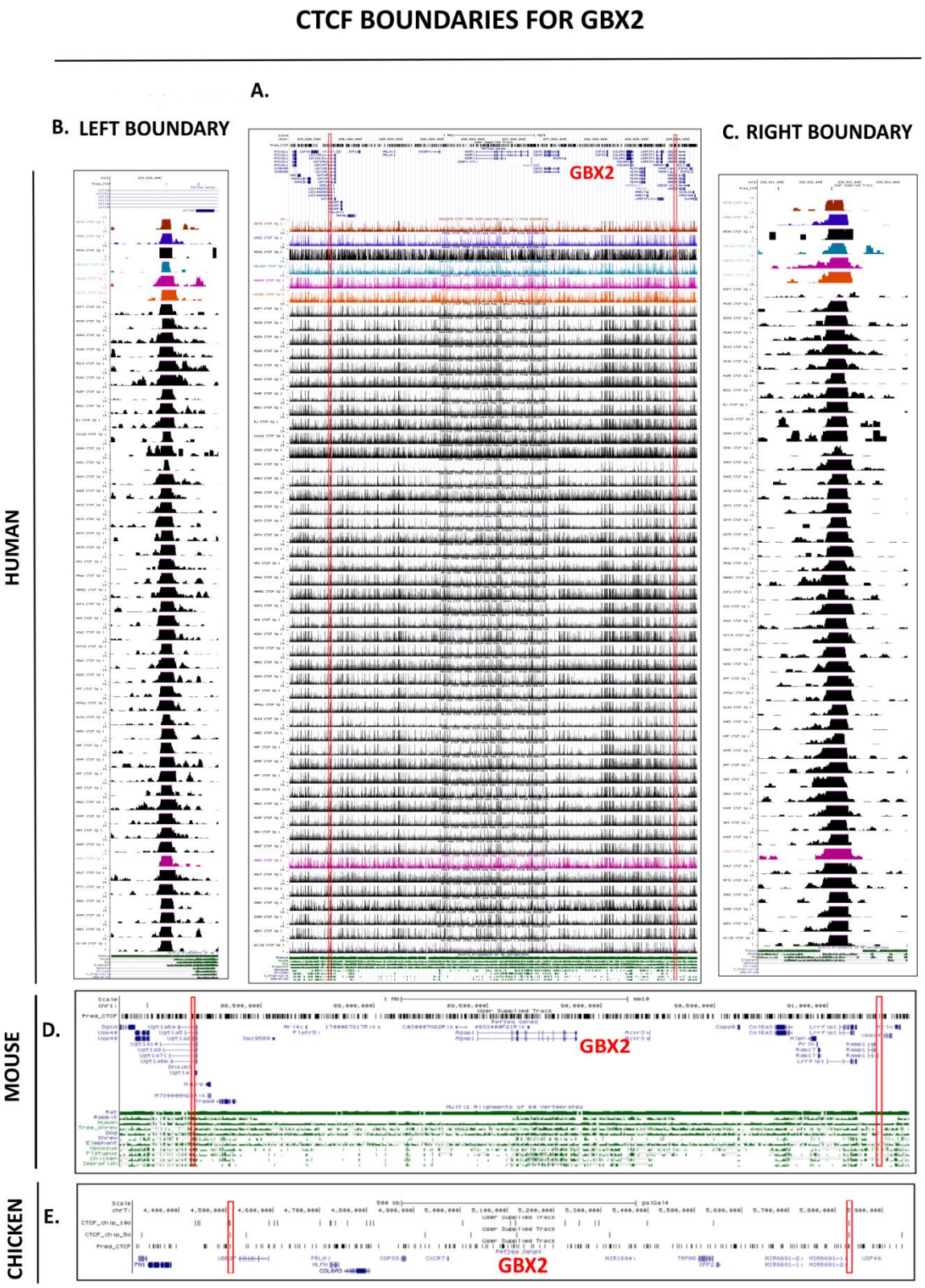


Figure 3.10 Insulator boundaries for *Gbx2*

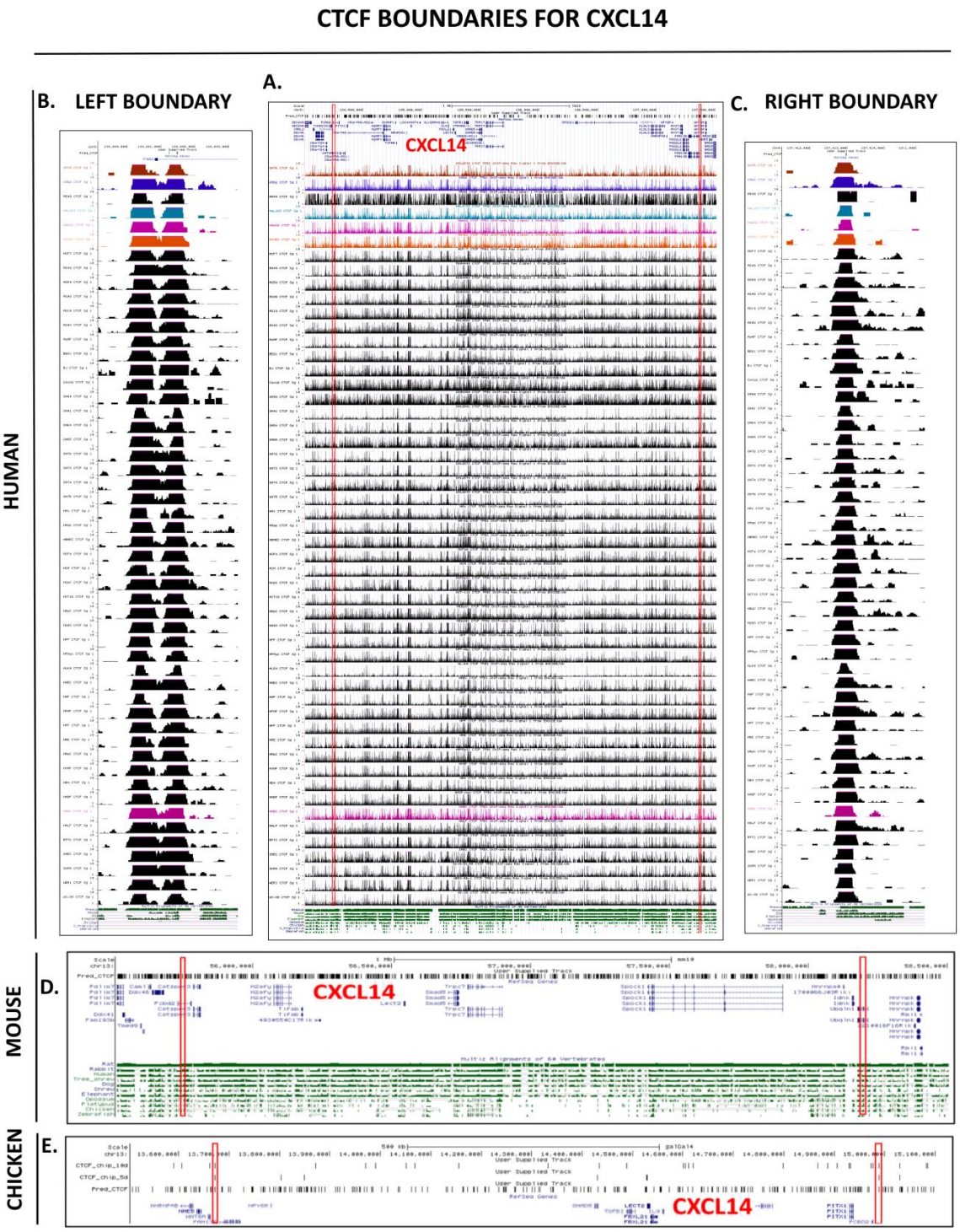


Figure 3.11 Insulator boundaries for *Cxcl14*

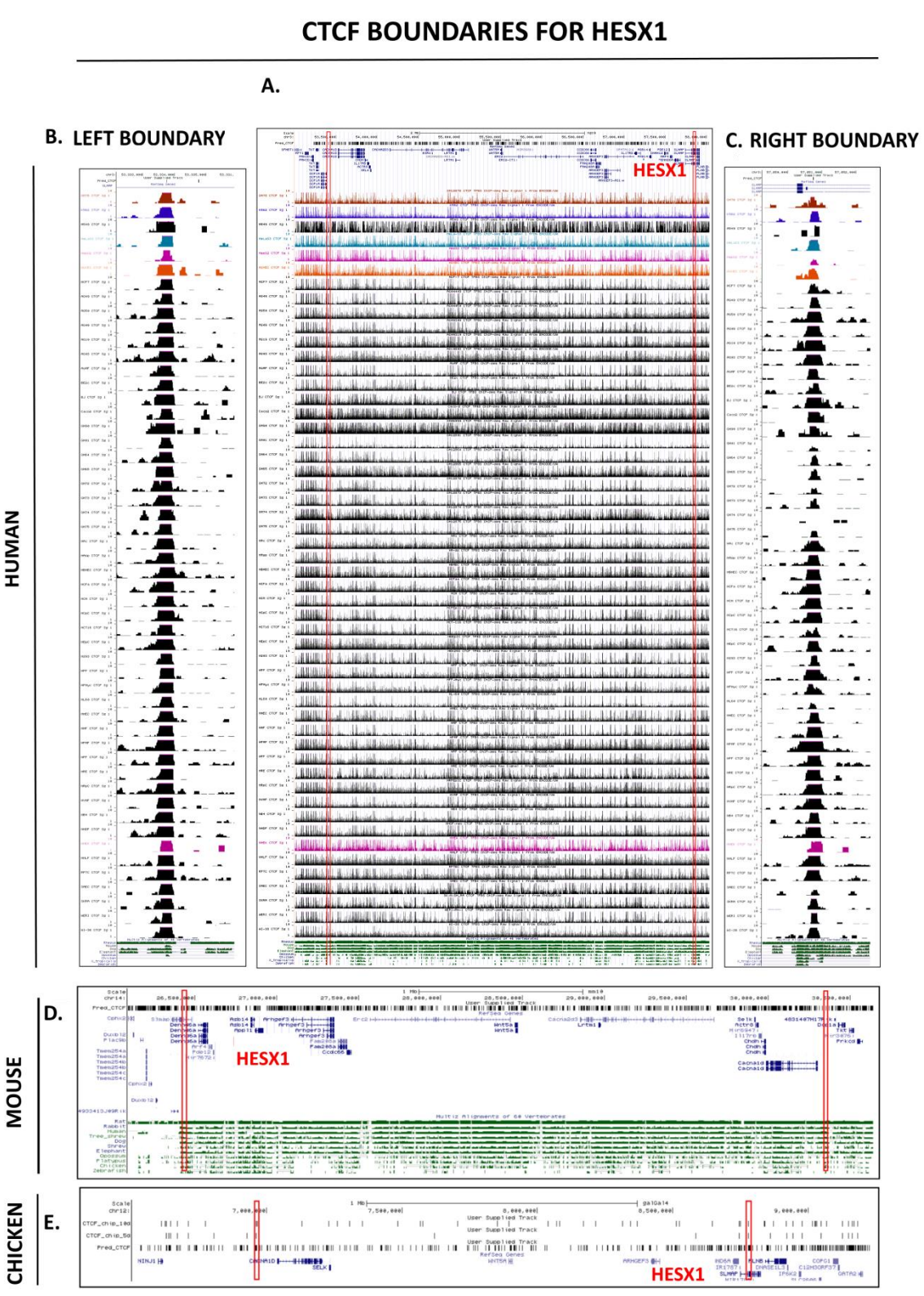


Figure 3.12 Insulator boundaries for *Hesx1*

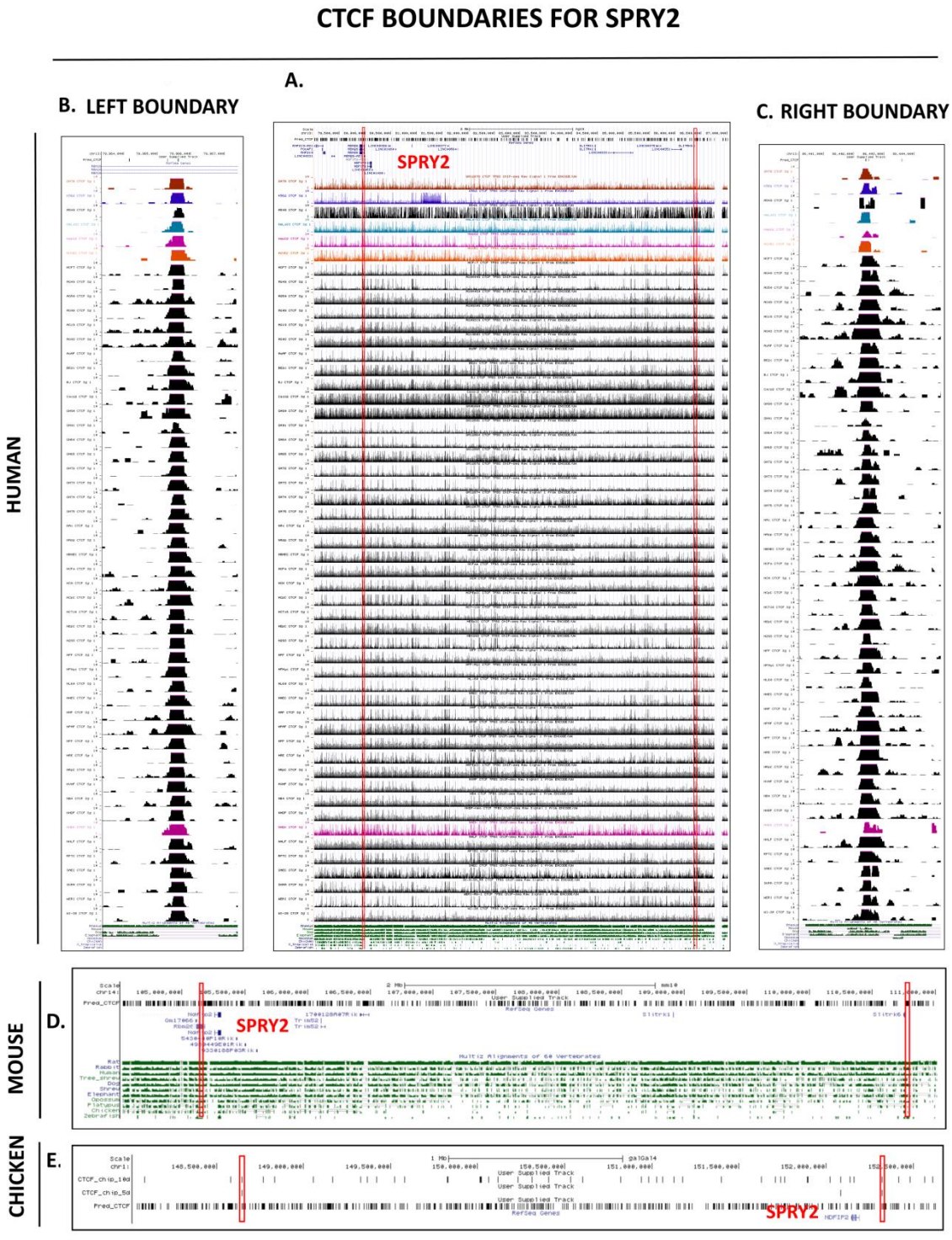


Figure 3.13 Insulator boundaries for *Spry2*

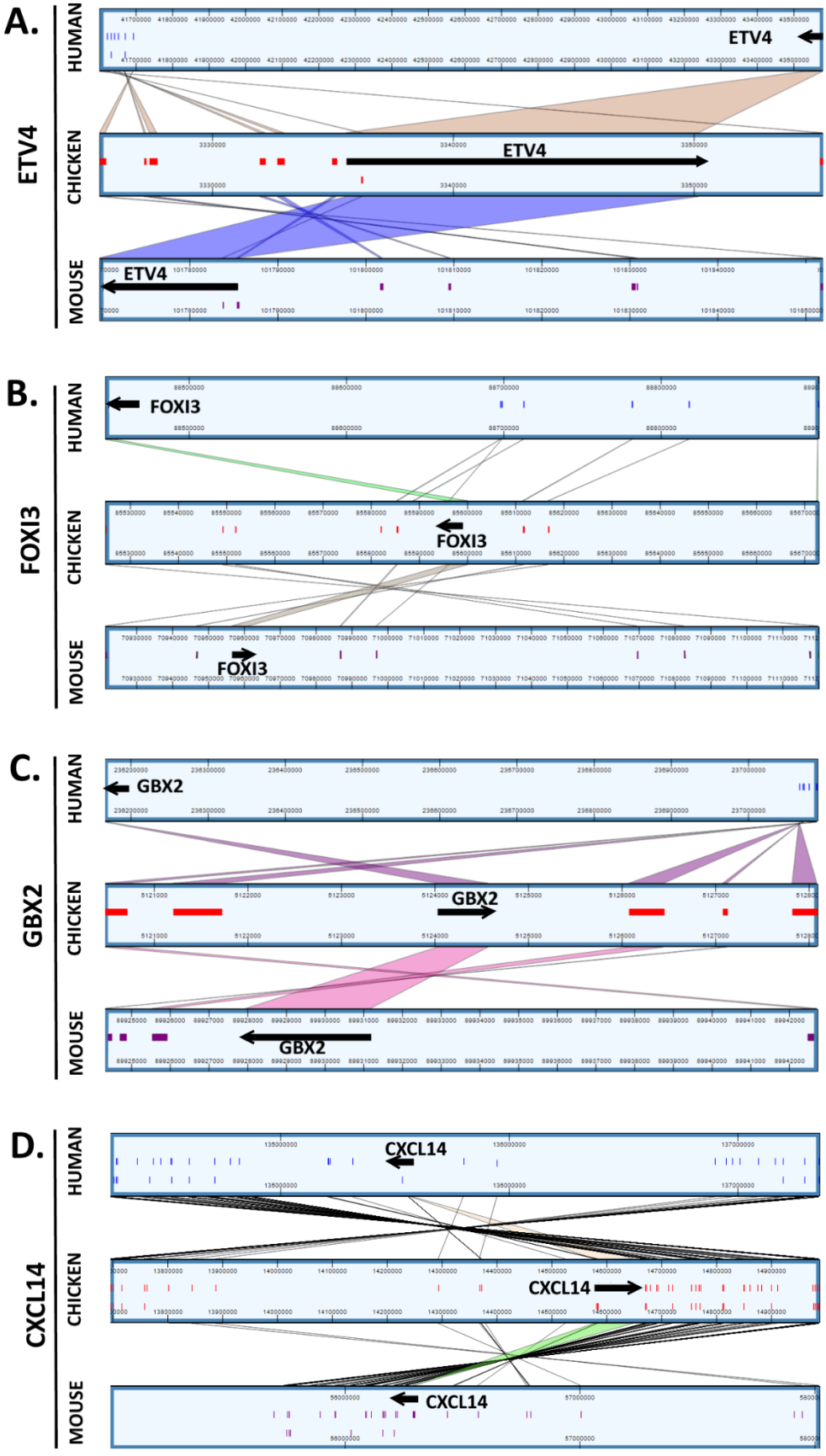


Figure 3.14 Predicted enhancers of early FGF response genes

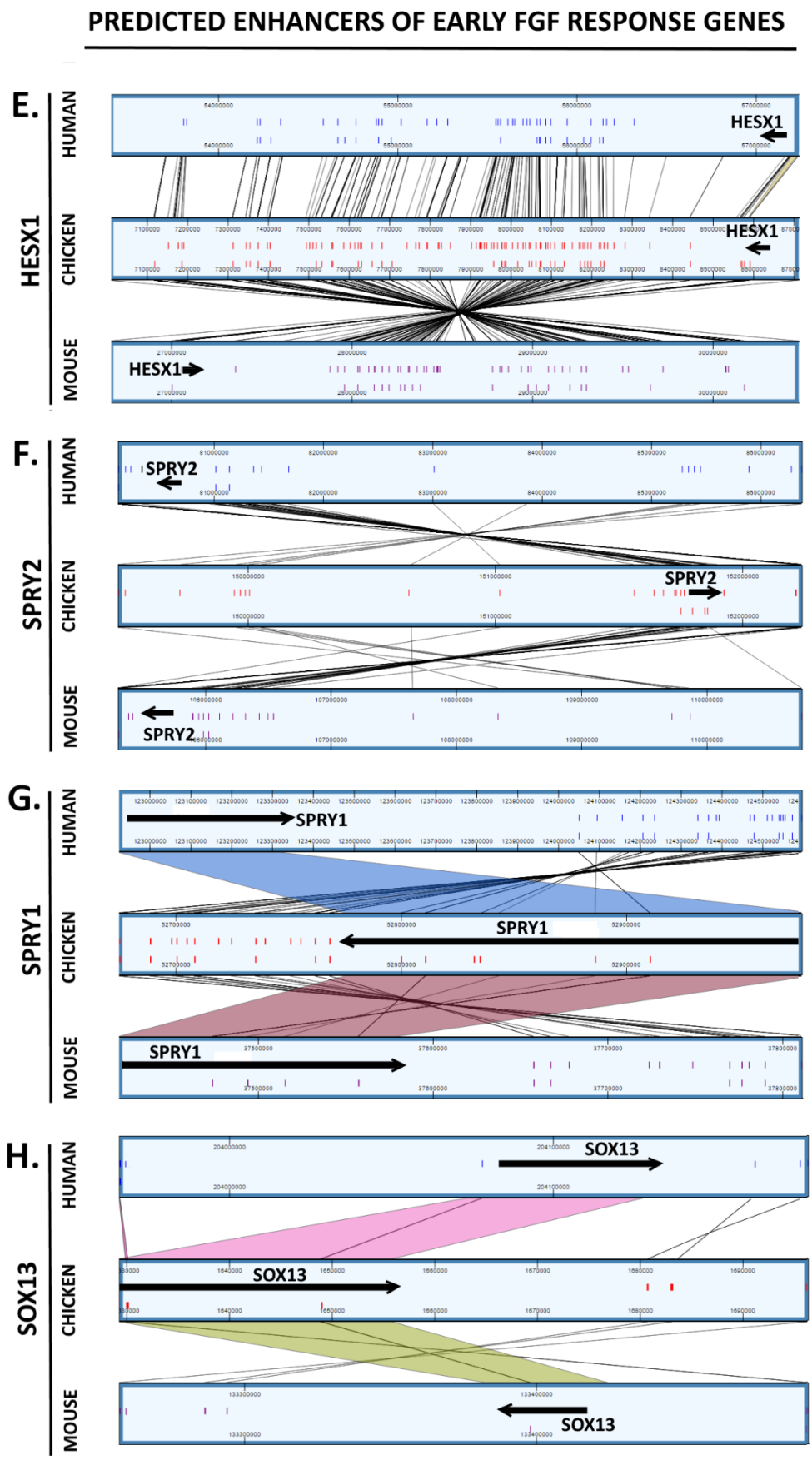


Figure 3.14 Predicted enhancers of early FGF response genes

Figure 3.14 Predicted enhancers of early FGF response genes
(A-H) The regions between the CTCF boundaries identified for human, mouse and chick were used to predict enhancers using DREiVe. The predicted enhancers are displayed in GSV browser. For genes without CTCF boundaries (*Etv4, Spry1* and *Sox13*), 300 kb upstream and 300 kb downstream regions were analyzed for predicting enhancers. Genes are shown as black arrows. The predicted enhancers are shown for human, chick and mouse as blue, red and purple boxes respectively. The corresponding conserved enhancers in the three species are linked through lines.

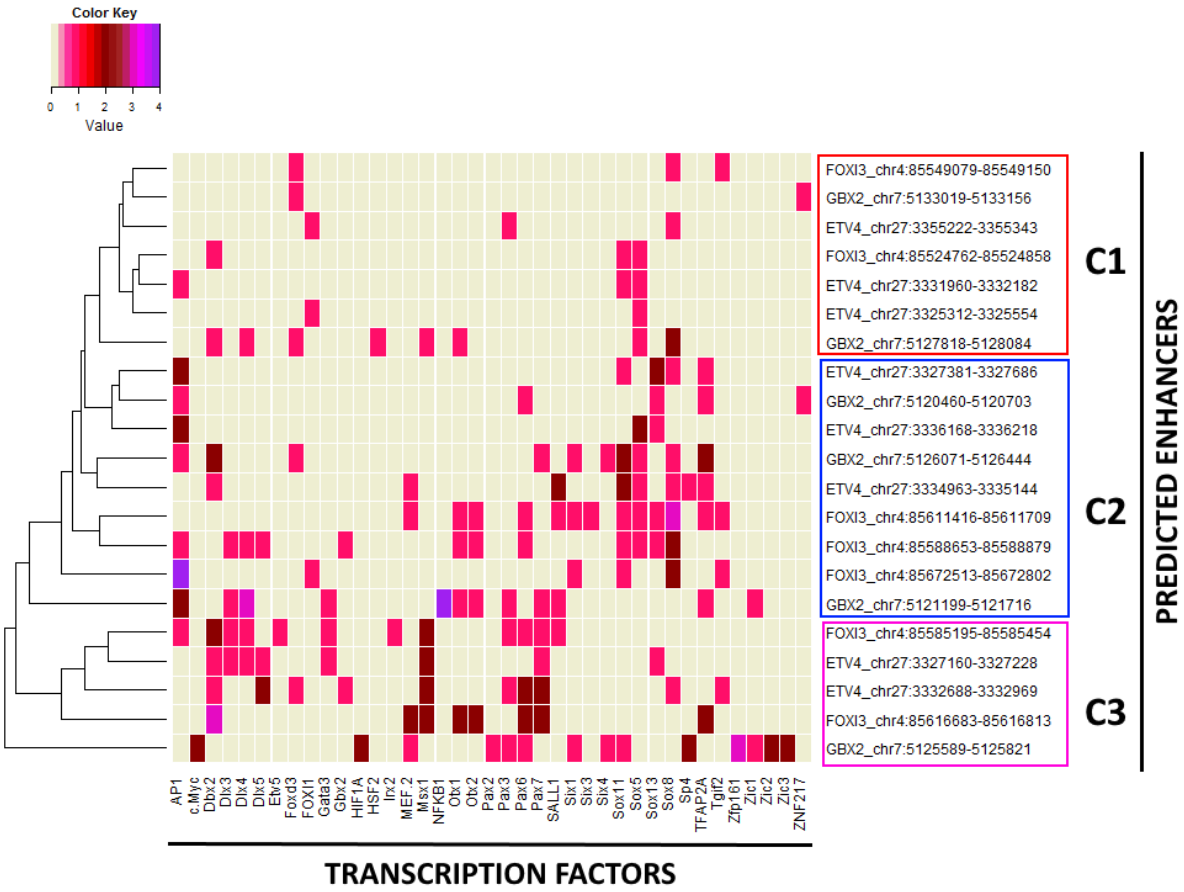


Figure 3.15 Transcription factor binding site analysis of the co-expressed early genes: *Foxi3*, *Etv4* and *Gbx2*
Transcription factor binding sites in *Foxi3, Etv4* and *Gbx2* were predicted using Clover and RSAT matrix-scan. Binding sites that were commonly reported in predicted enhancers by Clover and RSAT matrix-scan were subjected to hierarchical clustering to identify groups of predicted enhancers with common binding sites. X-axis represents transcription factors and y-axis represents the predicted enhancers. Number of binding sites of each transcription factor in the corresponding predicted enhancer range from 0-4 (colours: cream to purple). Three clusters were found: C1(red), C2(blue) and C3(pink).

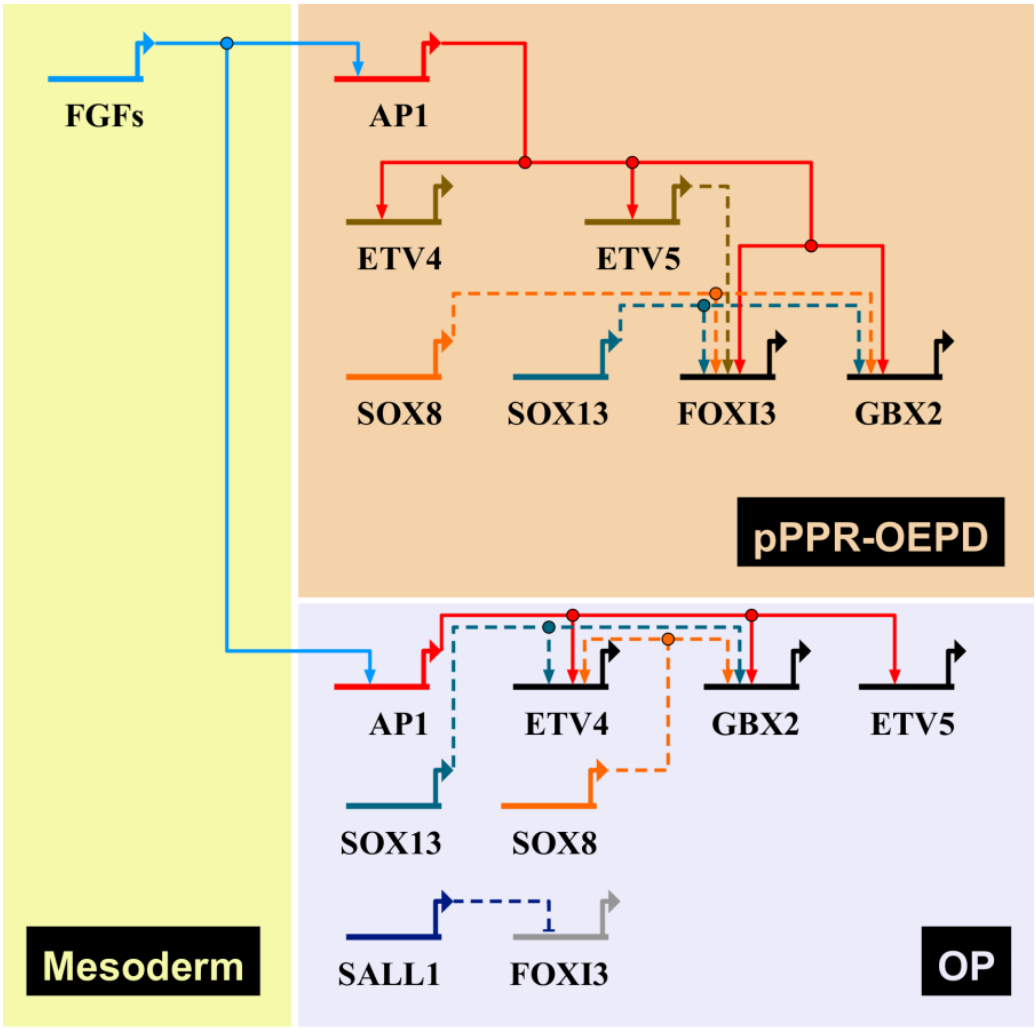


Figure 3.16 A hypothetical model of regulation for *Foxi3, Etv4* and *Gbx2* during otic placode formation.
A preliminary model in BioTapestry: The model represents events from pPPR-OEPD and OP separately as most of the represented genes are expressed from pPPR-OP stages except for *Foxi3* which is not expressed in the OP. All predicted interactions are represented as dashed lines. FGFs from the underlying mesoderm signal via *Ap1* to upregulate early FGF-response genes *Etv4, Foxi3* and *Gbx2* in the PPR. *Etv4* and *Etv5* are known direct targets of FGF signalling and hence are placed above *Foxi3* and *Gbx2* in the pPPR-OEPD stage. From the transcription factor binding site analysis, *Etv5* seems to be regulating *Foxi3* and hence a putative link is put from *Etv5* to *Foxi3*. Similarly, *Sox* family members (particularly *Sox8* and *Sox13*) which are expressed in the PPR and continue to express in the OEPD and otic placode (OP) were found to be regulating *Foxi3, Etv4* and *Gbx2* and hence putative links are put from *Sox* to these in PPR-OEPD as well as in OP. As *Foxi3* is not expressed in the OP, no links are put between *Sox8/13* and *Foxi3* in the OP. *Sall1* binding site was found in *Foxi3* and hence it can be speculated that *Sall1* may be involved in removing *Foxi3* from the OP through shutting its enhancer activity.

4. FGF signalling affects the chromatin landscape during OEPD induction and reveals novel enhancers

4.1 Introduction

In the previous chapter, I identified the earliest FGF-response genes and predicted their regulatory elements. However, as the number of predicted enhancers is high, it is essential to devise a strategy for prioritizing experimental verification of these enhancers. In order to do so and to understand the epigenetic state of the PPR cells upon FGF treatment, ChIP-seq for histone modifications was performed. This allowed the genome-wide identification of FGF-responsive regulatory elements during OEPD induction and further prioritized enhancer verification.

The epigenetic state of a cell plays an important role in regulating gene expression and ultimately defining the fate of a cell. The interpretation of genomic information is influenced by multiple factors such as the history of the cell and environmental cues. Ultimately, the control of gene expression occurs at the level of the chromatin and is brought about by regulatory elements such as promoters, enhancers, silencers and insulators. Among these, enhancers can regulate transcription of their target genes at great distances ranging from hundreds to thousands of kilobases away (Nobrega et al., 2003, Sagai et al., 2005). As described earlier, an enhancer is activated when different proteins called the pioneer factors bind to condensed chromatin and displace nucleosomes [see Chapter 1; section 1.6.2; (for review see Zaret and Carroll, 2011)]. This is accompanied by deposition of activation marks such as mono-methylation of H3K4 (H3K4me1) and acetylation of H3K27 (H3K27ac) which facilitate the binding of other transcription factors (Xu et al., 2007, Lupien et al., 2008, Xu et al., 2009, reviewed in Zaret and Carroll, 2011, Serandour et al., 2011, Asp et al., 2011, Ernst et al., 2011, Kharchenko et al., 2011, Spitz and Furlong, 2012, Bonn et al., 2012, Buecker and Wysocka, 2012,

reviewed in Calo and Wysocka, 2013) leading to interaction with the basic transcriptional machinery, to facilitate activation of their target genes (for review see: Ong and Corces, 2011, Shlyueva et al., 2014). In contrast, tri-methylation of H3K27 (H3K27me3) is associated with inactive or repressed enhancers (Rada-Iglesias et al., 2011, Zentner et al., 2011).

To understand how PPR cells become specified as otic cells, it is important to investigate whether FGF signalling changes the epigenetic landscape and, if so this will help to identify putative enhancers responsible for regulating the expression of FGF target genes. The aim of this chapter is three-fold: first, to understand if FGF signalling modulates the epigenetic state of PPR cells; second, to identify putative enhancers for FGF-response genes from FGF-treated ChIP-seq and overlap them with the predicted enhancers identified in Chapter 3; third, to investigate the transcription factor binding sites in validated enhancers. This will enhance the gene regulatory network downstream of FGF signalling and determine putative links between the earliest FGF-response genes through the identification of transcriptional inputs in their otic enhancers.

4.2 A histone ChIP-seq experiment after FGF2 treatment

To address whether FGF signalling affects the chromatin landscape and, to define regulatory elements of some of the FGF target genes Dr. Monica Tambalo designed the following experiment. PPR cells that are not yet specified as otic placode, but will form the placode later, were cultured in the presence or absence of FGF2 for 6 hours. Explants were harvested for ChIP-seq for histone modifications. She investigated the distribution of H3K27ac (associated with active enhancers) and H3K27me3 (associated with repressed regions) following the nano-ChIP protocol of Adli and Bernstein (Adli and Bernstein, 2011). The ChIP DNA was used for library preparation and sent for 100 bp

paired-end sequencing as described in section 2.10. For details of experiment see (Tambalo, 2015).

4.3 Quality assessment, alignment and peak-calling

To assess the quality of sequencing, FastQC (Andrews, 2010) was used, which produces a number of plots. Of those, the per base sequence quality plot is the most useful as it indicates the sequence quality at the base pair level. These plots are included in Appendix 8.2 and 8.3 for control and FGF2 ChIP-seq experiments, respectively.

Because of the high frequency of mismatches the first 9 bp were trimmed as suggested by Adli and Bernstein (Adli and Bernstein, 2011). In addition, if the sequence quality was poor at the 3' end as assessed by QC analysis (below phred score = 10; base call accuracy < 90%), further trimming was carried out (Appendix 8.2 and 8.3). Trimmed sequences were then aligned to the chick genome Galgal4.71 using Novoalign (Novocraft 2.08.01, (http://www.novocraft.com/products/novoalign/) and the number of uniquely aligned reads (>10 million for each sample) is given in Appendix 8.2 and 8.3. The sequence quality was found to be in line with the standard reported in literature (reviewed in Furey, 2012). Uniquely aligned reads were further used for peak calling using Homer (Heinz et al., 2010) as described in section 2.10.1. A fold change of 1.5 and a False Discovery Rate (FDR) of 0.01 were used to identify significant peaks in +FGF2 and control samples. Using these parameters, a known otic enhancer for *Spalt4* was retrieved. This enhancer is a good candidate to assess the ChIP-seq data as it is regulated by *Etv4*, a downstream effector of FGF signalling (Barembaum and Bronner-Fraser, 2010). The *Spalt4* enhancer was found to be flanked by H3K27ac peaks and depleted in H3K27me3 in +FGF2, whereas the same region is depleted of H3K27ac in the control sample (Figure 4.1). To gain additional confidence in ChIP-seq results, a second peak-caller MACS2 (Zhang et

al., 2008) was used. For MACS2, an FDR of 0.05 as suggested in MACS manual to obtain broad peaks (characteristic of histone peaks) and a default p-value of 1e-5 was used. Using these parameters, MACS2 also retrieved *Spalt4* enhancer. Both Homer and MACS2 estimate the fragment size from the aligned reads provided so the peak size was not specified. Following peak calling, enhancers were identified as described in the next section.

4.4 Genome-wide identification of putative enhancers in FGF2 and control samples

After peak calling, a few criteria were used to identify enhancers from ChIP-seq as described in section 2.10.1. One feature of the known *Spalt4* enhancer was that it was flanked by H3K27ac peaks and devoid of H3K27me3. This feature, where active promoters and enhancers are flanked by H3K27ac has been reported previously (Kimura, 2013, Joshi, 2014). Keeping this in mind, regions of up to 3 kb flanked by H3K27ac peaks and devoid of H3K27me3 peaks were identified and annotated to the nearest gene using gene annotation files (Ensembl and RefSeq) as described in section 2.9.1 and the 'annotatePeaks' function in Homer.

In order to get insight into the distribution of histone modifications, read densities for both H3K27ac and H3K27me3 in +FGF2 and control were plotted centred around the transcription start site (TSS) of the nearest gene using 'annotatePeaks' (Figure 4.2 A-B). This shows that the TSS is flanked by H3K27ac and H3K27me3 peaks on either side with the peak height for H3K27me3 being substantially lower than that for H3K27ac consistent with previous reports (see review: Kimura, 2013). Figure 4.2 shows that the TSS is histone free; a similar distribution is expected for active enhancers (Figure 4.3). Figure 4.3 shows the distribution of read densities for H3K27ac and H3K27me3 from the centre of all enhancers identified in +FGF2-treated and control samples. Since putative

enhancers are flanked by both marks, although the overall density of H3K27ac is more than 2 times greater than H3K27me3, it is possible that some enhancers are poised rather than active. Alternatively, this may also reflect cell heterogeneity within each explant; for example, cells in the OEPD appear to be heterogeneous in terms of their expression of *Pax2* (Streit et al., 2000).

Overall, 3691 putative enhancers were identified for +FGF2 and 3259 for control tissues. To test whether these +FGF2 enhancers are associated with genes that are indeed enriched in the endogenous OEPD (obtained from mRNA-seq analyzed as compared to the control (whole embryo) by Dr. Jingchen Chen), a one-sided Wilcoxon test was carried out. The mean FPKM of genes with putative +FGF2-specific enhancers was found to be greater than mean FPKM of all OEPD genes with a significant p-value of 0.01 indicating that the genes with +FGF2-specific enhancers are among those that are highly expressed in the OEPD (Figure 4.4). This indicates that FGF2-treated pPPR may recapitulate the transcription profile of an endogenous OEPD as also indicated by Yang et al. (2013).

4.5 Increase in H3K27ac upon FGF2 treatment

During enhancer identification from the ChIP-seq data, a difference was observed in the number of H3K27ac peaks around FGF-response genes in +FGF2 and control samples. To begin to address whether this difference may indeed result from FGF signalling, a genome-wide strategy was devised to quantify the number of H3K27ac peaks around the transcription start site (TSS). All H3K27ac peaks were annotated to the nearest gene in both +FGF2 and control samples; for each gene the number of assigned peaks was summed. A total of 13,637 genes were associated with at least one H3K27ac peak in at least one of the two samples. This analysis revealed that H3K27 acetylation is highly dynamic and responds to FGF treatment: 1278 genes presented more, while 652 showed

less H3K27ac peaks as compared to the control. Genes with a 2 or more peak difference between FGF2+ and control are displayed in a genome-plot as yellow vertical lines (Figure 4.5). In Chapter 3, a group of genes was identified that is upregulated by FGF just after 6 hrs including *Spry1, Spry2, Cxcl14, Hesx1* and *Foxg1*. Investigating their acetylation patterns in +FGF2 and control samples reveals a difference of 5 or more in the number of H3K27ac peaks (Figure 4.5 pink lines). For example, *Spry1* and *Spry2* are strongly upregulated by FGF and are known to be early FGF targets (Minowada et al., 1999, Chambers and Mason, 2000, Ozaki et al., 2001). 38 H3K27ac peaks are associated with each of *Spry1* and *Spry2* after FGF2 treatment, compared to only 3 and 10 H3K27ac peaks in controls. Thus, both transcripts are among the top candidates for high levels of H3K27 acetylation as compared to control.

Likewise, *Hesx1* is associated with 4 H3K27ac peaks after FGF2 treatment, but shows none in control and 9 peaks were identified for *Cxcl14* in +FGF2 versus 5 in control tissue. Some genes such as *Foxi3* and *Etv4/5* though densely acetylated did not present a difference of 2 or more peaks between FGF and control samples and are thus not highlighted in the plot.

In contrast, a number of genes show more H3K27 acetylation in the control as compared to +FGF2 samples. These include *Dlx6, Gata2/3* and *Axin2* (Figure 4.5 blue lines). This is consistent with the finding that these transcripts are downregulated upon FGF treatment after 6 hrs (Chapter 3) and hence the reduced acetylation marks in +FGF2 ChIP-seq.

In conclusion, it seems that genes upregulated upon FGF treatment show an increase in H3K27 acetylation, while downregulated transcripts display a decrease. This finding

suggests that a subset of FGF-response genes is subject to epigenetic changes in response to FGF signalling, which in turn may reflect in transcriptional changes.

4.6 Dense H3K27ac peaks flank Ap1 binding sites in FGF2-treated sample

As described in Chapter 3, the *Ap1* complex consisting of *c-Fos* and *c-Jun* (Neuberg et al., 1989, Glover and Harrison, 1995) is activated by the ERK/MAP kinase pathway, downstream of FGF signalling (Gruda et al., 1994, Hurd et al., 2002, Lopez-Bergami et al., 2007). This suggests that in direct FGF targets putative enhancers identified in +FGF2 samples should contain *Ap1* binding sites. To investigate this, all +FGF2 enhancers were screened for *Ap1* binding sites using available *Ap1, Fos* and *Jun* binding sites from JASPAR and RSAT matrix-scan. The total number of significant matches for *Ap1/Fos/Jun* found in +FGF2 putative enhancers as compared to shuffled sequences (control) was 12061 in 2249 putative enhancers of 2212 unique genes. To see if *Ap1* binding has any association with H3K27 acetylation, the H3K27ac peaks from +FGF2 and control were plotted in a window of -/+5 kb around the *Ap1* binding sites in a heatmap using seqMINER (Ye et al., 2011) (Figure 4.6 A). The peaks are clustered into groups based on their distance from the *Ap1/Fos/Jun* binding site. In the heatmap, each line corresponds to a genomic location being compared between +FGF2 and control conditions. The heatmap reveals five different clusters (C1-C5); C1 and C2 show a very tight association of H3K27 acetylation with *Ap1* binding sites with active marks found at a distance of 1 and 1.5 kb. In cluster C3 H3K27ac peaks are found at about 500 bp upstream of *Ap1* binding sites, while in C4 and C5 they are enriched downstream (~800 bp, ~500 bp, respectively). In contrast, the same genes show very low levels of H3K27 acetylation around *Ap1* binding sites in non-treated controls indicating a correlation between potential enhancers of FGF target genes and their histone acetylation.

The peak density plots (Figure 4.6 B) are in line with this suggestion and demonstrate that in each cluster the H3K27ac peak heights are greater in +FGF2 than in controls. Further analysis reveals that clusters C1 and C3 include putative enhancers for some of the earliest FGF-response genes *Foxi3, Cxcl14* and *Spry1* (for expression patterns see Chapter 3, sections 3.3 and 3.5), while cluster C4 includes putative *Hesx1* enhancers. Functional annotation of these clusters using DAVID (Huang et al., 2009b) shows enrichment of terms like ear morphogenesis, inner ear development, sense organ development and, importantly, MAPK signalling (Figure 4.6 C). These findings may point towards a potential role for FGF signalling in regulating its target genes by modifying histone acetylation during OEPD induction which must be explored experimentally.

4.7 Identification of unique putative enhancers in response to FGF2 signalling

In total, analysis of the ChIP-seq data identified 3691 putative enhancers for +FGF2 and 3259 for control tissues (see section 4.4) with many of the former containing *Ap1* binding sites (see section 4.6). I next sought to establish unique and common putative enhancers in each tissue using the R package ChIPpeakAnno (Zhu et al., 2010). Putative enhancers were considered to overlap if they had a gap of 0 bp between them, otherwise they were considered to be unique. A total of 2451 unique enhancers was identified in control and 2883 in +FGF2 conditions, as well as 808 common putative enhancers (Figure 4.7 A). Their distribution across the genome is shown in a genome plot (Figure 4.7 C; red: +FGF2 unique enhancers; green: control unique enhancers; brown: overlapping enhancers). To gain insight into the potential function of genes associated with these putative enhancers, enriched GO terms were identified using DAVID for both samples (Huang et al., 2009b). Genes associated with unique enhancers in FGF2-treated cells recover terms such as inner ear development and MAP kinase signalling (Figure 4.7 B,

red), whereas the same terms were absent in genes with unique enhancers in controls (Figure 4.7 B, green). Thus, OEPD and ear-related terms are exclusively found in +FGF2 samples.

To assess similarities and differences of the transcriptional inputs for these putative enhancers and to determine motifs that may be enriched in putative FGF activated enhancers, I performed a transcription factor binding site analysis. The sequences for unique +FGF2 and control putative enhancers were retrieved from UCSC and scanned using RSAT matrix-scan with transcription factor binding sites present in the full JASPAR library. This analysis did not reveal a significant difference between the two samples. One possible explanation for this observation is that so far the identified enhancers are only putative otic enhancers as none of them has been verified experimentally. As such, there is a high probability of including regions in the TFBS analysis that may not be active at all. This introduces a certain degree of noise in both sets that may not allow the proper identification of differential binding sites in both samples. As an example, the top 50 TFBSs in +FGF2 associated enhancers are plotted in Figure 4.8 showing the number of occurrences of the corresponding TFs in +FGF2 (red) and control (green). As expected, *Ap1* binding sites occur twice as often in +FGF2 unique enhancers when compared to controls. However, this difference is much smaller for other factors like *Sox, Zic, Lhx* and *Lmx* family members.

In summary, sets of putative enhancers can be identified after FGF2 treatment of the pPPR, which clearly differ from control conditions. When associated with the nearest gene, GO term analysis reveals strong association of potential FGF-regulated transcripts with ear formation indicating that this strategy indeed identified putative enhancers

relevant to otic development. However, TFBS analysis does not reveal a big difference in the potential upstream regulators of FGF2 associated and control enhancers.

4.8 DREiVe-predicted enhancers overlap with ChIP-identified enhancers

To prioritize putative enhancers for in vivo verification, in addition to H3K27ac patterns, I considered the presence of motifs conserved across species. Therefore, I focused further analysis on a small subset of FGF-response genes, whose enhancers were predicted using DREiVe (see Chapter 3): *Etv4, Foxi3, Gbx2, Cxcl14, Sox13, Hesx1, Spry1* and *Spry2*. DREiVe-predicted enhancers were overlapped with putative enhancers identified by ChIP-seq for histone modifications after FGF treatment of pPPR using the R package ChIPpeakAnno (Zhu et al., 2010). The putative enhancers identified from ChIP-seq spanned 100 bp to 3 kb while predicted enhancers were generally smaller ranging from 50 bp to 600 bp. An overlap of 1 bp or more between predicted and ChIPseq-identified enhancers was considered. As the length of the predicted enhancers was much smaller, in certain instances more than 1 predicted enhancer overlapped with a ChIPseq-identified enhancer, e.g. for *Gbx2* and *Sox13*. In a few cases part of a predicted enhancer overlapped with the flanking H3K27ac peak. These cases were also considered overlapping to accommodate any inaccuracies of peak calling in defining the exact nucleosome position. The total number of predicted enhancers that overlap with putative enhancers of FGF response genes is 13 (Table 4.1, Figure 4.9). Additionally, some predicted enhancers overlapped with ChIPseq-identified enhancers for other genes and will not be discussed here. The coordinates of overlapping DREiVe and +FGF2 enhancers are given in Appendix 8.4.

It is worthwhile to note that the number of predicted enhancers is fairly large as predictions were carried out within CTCF boundaries (identified in Chapter 3). However,

in case of ChIP-seq, only proximal enhancers were identified by associating putative enhancers with the closest gene. For *Cxcl14*, no overlap was found and hence it is not shown in the chromosome plot. Predicted enhancers lie upstream of *Cxcl14*, and appear to be marked by repressive marks which will be discussed further in the next sections. This analysis has helped to reduce the number of putative enhancers to be verified. Additionally, conserved regions in chick, identified from Ensembl PECAN multiple alignments, were compared with DREiVe and ChIP-seq identified enhancers. The two *Foxi3* enhancers were found to overlap with +FGF2 enhancers and DREiVe predictions (Figure 4.11 A). In the cases of *Hesx1, Spry1* and *Cxcl14*, conservation in non-coding regions was not observed using alignment (Figures 4.14 A, 4.15 A, 4.16 A). As a result of this, DREiVe predictions were solely considered for conservation analysis.

Thus, in summary, of the 23 ChIPseq-identified enhancers for the six FGF-response genes, 10 overlap with 13 predicted enhancers i.e. 10 ChIPseq-identified enhancers are evolutionarily conserved. They are therefore likely to represent enhancers that drive gene expression in OEPD cells in different species, and useful for further investigation. In addition to being conserved and marked by active histone modifications, these were found to be unique to the +FGF2 ChIP-seq and hence were good candidates to start the validation process.

Before proceeding to experimental verification, putative overlapping enhancers for *Foxi3, Spry1, Spry2, Sox13, Gbx2* and *Hesx1* were analyzed for transcription factor binding sites. Additionally, a proximal ChIPseq-identified enhancer for *Cxcl14* was selected as there was no overlap between predicted and ChIPseq-identified enhancers for *Cxcl14*. Moreover, no enhancers were identified for *Etv4* from ChIP-seq therefore one proximal predicted *Etv4* enhancer was also included in the TFBS analysis. PFMs of enriched PPR,

otic, lens and trigeminal transcription factors (as compared to the whole embryo) from mRNA-seq data of Dr. Jingchen Chen and Ramya Ranganathan were used to create a customised library which was then used to screen these selected putative enhancers using RSAT matrix-scan (Turatsinze et al., 2008). TFBS enrichment in the enhancer sequences was determined using randomised sequences as controls.

Table 4.1 Overlap between DREiVe-predicted and +FGF2 putative enhancers

Gene	DREiVe	+FGF2	Overlap
Etv4	8	0	0
Foxi3	8	2	2
Gbx2	6	2	3
Cxcl14	169	4	0
Hesx1	280	4	1
Sox13	6	3	3
Spry1	26	4	2
Spry2	24	4	2

A heatmap was generated from the results of RSAT to identify common transcriptional inputs of the putative enhancers (Figure 4.10). It is evident that the TFs with most number of binding sites in the putative enhancers are concentrated to the left of the heatmap. Putative enhancers for *Spry1* and *Spry2* share many binding sites for transcription factors, consistent with the fact that they are often co-expressed (Chambers and Mason, 2000, Ozaki et al., 2001). Some of the TFs with binding sites in most of the putative enhancers include *Ap1, Wt1, Sall1, Zic3, Sox* family *(Sox3/5/8/13), Otx2, Irx2/4/5* and *Lmx1a/b.* The presence of *Ap1* binding sites suggests that these could be potential FGF-responsive enhancers. It was discussed previously in Chapter 3 during the TFBS analysis of *Foxi3, Etv4* and *Gbx2* that *Sox* family members particularly *Sox8* and *Sox13* which are expressed early in the PPR and OEPD were among some of the prime regulators (Chapter 3; Figures

3.15 and 3.16). Here, again binding sites of *Sox8* and *Sox13* along with *Sox3/5* are found in putative enhancers of *Spry1/2, Foxi3* and *Gbx2*. *Sox3/5* are earlier expressed in the PPR and neural tissues but begin to express in the otic placode at HH12 (Matsumata et al., 2005) therefore these are good candidates for regulating *Spry1/2* and *Gbx2* both of which are expressed in the otic placode (Chambers and Mason, 2000, Sanchez-Calderon et al., 2004). Additionally, *Lmx1a* and *Wt1* have binding sites in *Sox13, Spry1, Spry2* and *Hesx1*. Activators *Lmx1a* and *Wt1* begin to express at the OEPD and otic placode stage respectively (Tambalo, 2015) and hence are good candidates for regulating these genes. The Irx family members Irx2/4/5 have very similar TFBSs hence all three have been found in putative enhancers of *Spry1/2, Gbx2, Foxi3* and *Hesx1.* Of these, *Irx2* and *Irx4* are strongly expressed in the posterior PPR and later in the otic region (Goriely et al., 1999, Khudyakov and Bronner-Fraser, 2009) and therefore may be involved in regulating *Spry1/2, Gbx2* and *Hesx1* in the otic placode by binding to their putative enhancers. However, members of the Irx family have also been implicated in repression (Bilioni et al., 2005) and as *Foxi3* is not expressed in the otic placode, it is possible that *Irx2/4* are responsible for removing its expression via binding to its putative enhancer.

Otx2, which is a transcriptional repressor, is expressed in the anterior PPR (Sanchez-Calderon et al., 2007) and may be important in repressing *Spry1/2, Foxi3, Gbx2* and *Sox13* expression in the anterior PPR via their putative enhancers. Similarly, *Sall1* is a repressor that begins to express in the otic placode at HH11 (Sweetman et al., 2005) and has binding sites in the majority of the putative enhancers including those of *Cxcl14* and *Foxi3*. Both *Foxi3* and *Cxcl14* are expressed in the PPR and OEPD but removed from the otic placode (Chapter 3; Figure 3.6) and it is possible that this is mediated by *Sall1* through binding to their putative enhancers. Lastly, *Zic3* binding sites were found in putative enhancers of *Gbx2, Hesx1* and *Sox13. Zic3* is expressed in the neural tissues

(McMahon and Merzdorf, 2010) and may be involved in repressing *Gbx2, Hesx1* and *Sox13* expression in the neural tissues via binding to their putative enhancers.

In conclusion, a small group of TFs seem to be involved in positively regulating the FGF-response genes, particularly *Lmx1a* and *Sox* family members. However, experimental verification of these selected putative enhancers is required to draw any further conclusions.

4.9 Experimental verification of selected putative enhancers

Ultimately, it is critical to assess whether putative enhancers are active in vivo and specific to the cell type under investigation. To do this the candidate enhancer regions are cloned upstream of a minimal promoter (tyrosine kinase promoter, pTK) and the coding sequence for yellow fluorescent protein (pTK-Citrine) (see section 2.7). The reporter is then electroporated together with a plasmid containing RFP driven by a ubiquitous β-actin enhancer (pCAB RFP). The latter serves as control for electroporation, but importantly to show wide-spread targeting, while the test enhancer should only be active in a specific cell type. HH4/4^{+} chick embryos were electroporated and analyzed by fluorescent microscopy from HH6^{+} until they had reached placode stage.

A few candidate enhancers were selected for *Foxi3, Hesx1* and *Spry1* (Figures 4.11 A, 4.14 A and 4.15 A; see section 2.7 for coordinates). For *Cxcl14*, no proximal enhancers were found to overlap between the predicted enhancers and ChIPseq-identified enhancers however, the region around *Cxcl14* was found to be densely acetylated and was therefore selected for in vivo testing (Figure 4.16 A; see section 2.7 for coordinates).

Of these, both *Foxi3* enhancers (Figure 4.12 A- F and 4.13 A-D), the *Hesx1* (Figure 4.14 B), *Spry1* (Figure 4.15 B-C) and *Cxcl14* enhancers (Figure 4.16 B-C) were found to be active and their specific profile will be described below. The second *Spry1* enhancer was inactive (Figure 4.15 D-E).

The two *Foxi3* (Foxi3 E1 and E2) enhancers show distinct spatial and temporal activity (Figures 4.12 and 4.13), which correlates with the different endogenous expression domains of *Foxi3*. Initially, *Foxi3* is expressed in the pPPR, then in the OEPD and later lost from the otic placode and restricted to trigeminal and epibranchial regions and by HH15 expressed in the pharyngeal arches (Khatri and Groves, 2013). Initially, the entire 2.2 kb region between the H3K27ac peaks was cloned and named Foxi3 E1 (Figure 4.11 A). However, TFBS analysis using a customised library containing enriched PPR, otic, lens and trigeminal transcription factors (as compared to the whole embryo from mRNA-seq) showed that many TFBSs were concentrated in the centre of the enhancer. Therefore, different sub-clones were generated named Foxi3 E1.A and Foxi3 E1.B, respectively (Figure 4.11 B). The shortest region Foxi3 E1.B was sufficient to drive citrine expression; activity is first observed very weakly at PPR stages (HH7; Figure 4.12), which then increases as the OEPD is induced. No enhancer activity is detected at HH10 suggesting that Foxi3 E1.B is responsible for *Foxi3* expression in the pPPR and OEPD, and mimics the loss of *Foxi3* transcription as the otic placode is established (Khatri and Groves, 2013). Interestingly, the DREiVe-predicted enhancer was not included in Foxi3 E1.B (Figure 4.11 B).

In contrast, Foxi3 E2 contains (Figure 4.11 A, C) the DREiVe-predicted enhancer, which was sub-cloned into the reporter vector (Foxi3 E2.A; Figure 4.11 C). Electroporation of this construct reveals that it was sufficient to drive citrine expression, which was first

detected at HH9 in the trigeminal region and continued to be active in the trigeminal and epibranchial territories at HH11, as well as in pharyngeal arches at HH15 (Figure 4.13).

Hesx1 is expressed in the otic placode from HH12 onwards and remains in the otic vesicle until at least HH19 (Abe et al., 2006). Four unique enhancers were discovered from ChIPseq in +FGF2 pPPR cells, but not in controls, out of which E1 overlaps with a DREiVe-predicted enhancer (Figure 4.14 A). The DREiVe-predicted enhancer (conserved part of E1) was cloned. When electroporated into HH4/4^{+} chick embryos, enhancer activity was first observed at the otic placode at HH12, despite widespread electroporation (Figure 4.14 B). In addition, there is activity in the adjacent ectoderm or neural tube, although sections will be needed to determine the precise location, and in the posterior neural tube. *Hesx1* is expressed in the neural plate at earlier stages and in the neural tube (Chapman et al., 2002) therefore it can be concluded that the enhancer contains elements that are responsible for its expression in neural as well as otic tissues. Further characterisation is required to elucidate temporal and spatial aspects of Hesx1 E1 activity.

Spry1 is initially expressed in the PPR, then in the OEPD and otic placode until HH21. ChIP-seq identified five unique potential Spry1 enhancers, present in +FGF2, but not in control cells. Two of these (E1 and E5) overlap with DREiVe-predicted enhancers (Figure 4.15 A). The DREiVe-predicted enhancer within Spry1 E5 was cloned and electroporated. It showed strong activity in the forming otic placode at HH10 and at HH12 (Figure 4.15 B-C). The entire Spry1 E1 was cloned and electroporated and was found to be inactive in the otic placode or otic vesicle (Figure 4.15 D-E).

Cxcl14 is initially expressed in the posterior PPR and the posterior OEPD and the adjacent neural folds; at HH10 it is present in the ectoderm encircling the otic placode (Chapter 3, Figure 3.6; D-F). ChIP-seq highlighted four possible Cxcl14 enhancers (E1-4), while Cxcl14 E1 and E4 are unique to FGF treated PPR cells, E2 and E3 are not (Figure 4.16 A); none of them overlapped with DREiVe predictions. Predicted putative enhancers lie upstream of *Cxcl14*, and appear to be marked by repressive marks (Figure 4.16 A). Cxcl14 E1 was chosen for cloning for two reasons; firstly, it was proximal to *Cxcl14* and unique to +FGF2; secondly, it had binding sites for OEPD and neural transcription factors (Figure 4.16 D) and as already mentioned *Cxcl14* is expressed in the OEPD and neural tissues, therefore this enhancer was a good candidate for validation. Electroporation into chick embryos shows weak activity with only a few citrine+ cells seen in the neural folds at HH9 (Figure 4.16 B). At HH12, activity is observed in the ectoderm, which may or may not include the otic placode (Figure 4.16 C). Further characterisation is required to determine whether Cxcl14 E1 truly reflects the *Cxcl14* mRNA localisation.

Overall, it seems that combining enhancer predictions with ChIP-seq analysis allows prioritization for *in vivo* validation experiments, and reduces the number of potential enhancers to be tested. The success of this strategy is demonstrated by the number of enhancers with conservation that were found to be active in the otic tissues.

4.10 New otic enhancers share common transcriptional inputs

Once the enhancers for *Foxi3, Hesx1, Spry1* and *Cxcl14* had been validated, they were analyzed for transcription factor binding sites using the customized library as described in sections 2.12 and 4.8. Foxi3 E1.B had binding sites for *Tead1, Sall1* and *Sox* family members, particularly *Sox8* and *Sox13*, for *Pax2, Jun* and *Otx2/3* (Figure 4.12 G). *Jun* is

an essential element of the Ap1 complex, suggesting that FGF may regulate *Foxi3* expression through binding of *Ap1* to E1.B. *Sox8* and *Sox13* are expressed in the posterior PPR and OEPD respectively (McKeown et al., 2005, Tambalo, 2015), while *Sall1* is a repressor that begins to express at the placode stage (Sweetman et al., 2005). It is therefore possible that *Sox8* and *Sox13* promote *Foxi3* expression at the PPR and OEPD stages by binding to Foxi3 E1.B whereas *Sall1* removes *Foxi3* from the otic placode by shutting down Foxi3 E1.B. On the other hand Foxi3 E2.A presents different binding sites to E1.B (Figure 4.13 E). It has *Six1, Nr2f2, Rxra* and *Tcf4* binding sites. *Nr2f2* is expressed in the epibranchial region whereas the repressor *Tcf4* is expressed in the otic placode (Tambalo, 2015); thus *Tcf4* may be responsible for removing *Foxi3* expression from the otic placode and *Nr2f2* may instead promote *Foxi3* expression in the epibranchial region.

Transcription factor binding site analysis of Hesx1 E1 reveals binding sites for *Six1, Ap1*, *Lmx1a/b* and *Gbx2* (Figure 4.14 C). As *Hesx1* is expressed quite late in the otic placode (Abe et al., 2006), some of these factors may regulate its expression through Hesx1 E1. For example, *Six1*, *Gbx2* and *Lmx1a* are all expressed prior to *Hesx1* (Litsiou et al., 2005, Bok et al., 2005, Barembaum and Bronner-Fraser, 2007, Paxton et al., 2010, Sato et al., 2012) and thus may be involved in promoting its expression in the otic placode.

Spry1 E5 presents a number of interesting binding sites (Figure 4.15 F). *Lmx1a, Sox* and *Gbx2* binding sites are all enriched with *Sox8, Sox13* and *Gbx2* already expressed at PPR and OEPD stages. These factors are therefore good candidates as regulators of *Spry1*.

Likewise, Cxcl14 E1 is also enriched in binding sites of *Six1* and *Gata3* which may be regulating its expression in the ectoderm surrounding the otic placode (Figure 4.16 D).

Upon careful analysis, it is possible that the Cxcl14 E1 sequence comprises two different elements; the 3' end seems to contain binding sites for neural transcription factors which may regulate expression in the neural folds such as *Sox2, Sox5* and *Hoxa2*. On the other hand, the 5' portion of Cxcl14 E1 harbours binding sites for otic or ectodermal transcription factors such as *Lmx1a, Gata3* and *Sox13*.

In conclusion, TFBS analysis reveals the presence of a few transcription factors which are shared by most enhancers, mainly *Sox* family members. Based on the TFBS analysis, a model of regulation for *Foxi3, Hesx1* and *Spry1* is presented in Figure 4.17 which confirms and adds upon the links in the previous hypothetical model (Chapter 3; Figure 3.16).

4.11 Discussion

This chapter addresses four main aspects of the transition of pPPR cells to otic-epibranchial progenitor cells under the influence of FGF signalling. First, I investigated a potential role of FGF signalling in modifying the epigenetic state of cells during this process. Second, histone modifications, particularly H3K27ac and H3K27me3 were analyzed at the genomic level to identify enhancers uniquely present in +FGF2 treated cells as they become specified as OEPD. Third, to demonstrate potential functional relevance of putative enhancers, a selection of candidate enhancers for FGF-response genes was assayed for their *in vivo* activity using reporter assays. Fourth, to identify candidate genes involved in regulating FGF-response genes through binding to their enhancers, transcription factor binding site analysis was carried out.

An important consideration while analyzing ChIP-seq data is the presence of biological replicates as this allows assessing the reproducibility of data. In a typical ChIP-seq

experiment, non-specific binding of the antibody or biases in library preparation and sequencing introduce random background noise which in part can be circumvented by the use of control experiments but cannot be eliminated altogether. In such a case, replication is useful to correctly identify actual biological events. In the present study, such assessment could not be made due to the lack of biological replicates. As a consequence, a single ChIP-ed sample was compared to the input sample to detect regions enriched in acetylation or methylation. As shown in Figure 4.1, for H3K27ac and to a lesser extent for H3K27me3, the read tracks between +FGF2 and control samples display a similar trend. As a consequence, the peaks identified for the two samples are not very different. The presence of replicates would not only have allowed the removal of noise in each sample by considering the concordant peaks among the replicates but would have also given confidence in the identified peaks and subsequently in identifying enhancers. For this particular project, to deal with the lack of replicates and to prioritize experimental verification, DREiVe-predicted enhancers (Chapter 3) were overlapped with ChIPseq-identified enhancers to find conserved elements which were then screened for binding sites for otic-specific transcription factors.

4.11.1 FGF signalling: A potential role in modulating the epigenetic state of the cells during OEPD induction

The role of signalling in modulating the epigenetic state of cells has been investigated before in endothelial development where highly dynamic H3K27ac was associated with VEGFA signalling and it was shown that increased H3K27ac levels at specific loci corresponded to increased gene expression (Zhang et al., 2013a). Similar to this, here, the role of FGF signalling during OEPD induction is used to study an influence, if any, on the epigenetic state of the posterior PPR cells during OEPD induction. As described in the previous chapter, FGF signalling leads to the up regulation of otic and epibranchial genes

and down regulation of anterior PPR and lens genes. To understand how this transcriptional response compares with the epigenetic changes in the cell, a ChIP-seq for histone modifications (H3K27ac and H3K27me3) was carried out by using FGF2-treated and untreated posterior PPR cells. Upon FGF2 treatment, indeed H3K27ac (active marks) is increased considerably in the vicinity of genes that rapidly respond to FGF (*Spry1, Spry2, Foxi3, Cxcl14, Hesx1* and others; see Chapter 3, section 3.3), P300/CREB (histone acetyltransferase) is known to acetylate histone-3 lysine-27 (for review see: Karamouzis et al., 2007, Holmqvist and Mannervik, 2013) and its activity is stimulated by phosphorylation (Ait-Si-Ali et al., 1999) suggesting a role for FGF signalling in the phosphorylation of P300/CREB.

It has also been shown that P300/CREB function is enhanced when it interacts with phosphorylated c-Jun and c-Fos (Chen et al., 2004). As c-Fos and c-Jun are essential components of the Ap1 complex, this raises the possibility that the increase of H3K27 acetylation is due to CREB interaction with *Ap1* and that *Ap1* recruits p300 to the enhancer regions. This is supported by the observation that *Ap1* is flanked by H3K27ac peaks in PPR cells exposed to FGF2 specifically in regions near FGF-response genes (*Spry1, Spry2, Foxi3, Hesx1* and *Cxcl14*), whereas the same regions are depleted in H3K27ac peaks in the control (Figure 4.6 A). It can thus be speculated that after deposition of the Ap1 complex and H3K27 acetylation, additional PPR and OEPD transcription factors are recruited to the enhancer regions.

4.11.2 Conserved regions flanked by H3K27ac peaks and depleted in H3K27me3 reveal novel enhancers

In this project, enhancers have been identified by combining phylogenetic footprinting with histone ChIP-seq. In the previous chapter, enhancers were predicted for a small set

of early FGF-response genes. In this chapter, H3K27ac and H3K27me3 distributions were analyzed on a genome-wide scale to identify putative enhancers, which were then overlapped with DREiVe-predicted enhancers focussing on the earliest FGF-response genes to select a few for experimental verification.

As described earlier, active enhancers can be marked by a number of features including H3K27ac, H3K4me1 and P300 (for review see: Maston et al., 2012, Calo and Wysocka, 2013, Kimura, 2013). Due to difficulties encountered during antibody validation, H3K4me1 is not shown here and a ChIP quality P300 antibody was not available for chick. Therefore, to make enhancer identification easier, phylogenetic footprinting was employed together with histone ChIP-seq. Although phylogenetic footprinting finds conserved regions, it generally predicts many enhancers. Since most genes are expressed in different tissues at different times, the predicted enhancers may correspond to different expression domains of a gene. Thus combining it with histone ChIP-seq for the required tissue can reveal novel enhancers. This strategy was not only useful in reducing the number of putative otic enhancers to verify but also more fruitful as four out of six conserved enhancers were found to be active. Many enhancers are evolutionarily conserved (Prabhakar et al., 2006, Clarke et al., 2012) and this is consistent with the results presented here. All except one conserved candidate enhancer for *Spry1* were found to be active. It is possible that this enhancer is active in otic tissues at stages different from the ones analyzed or it may be active in tissues other than otic. This will require further investigation. In future, prediction of enhancers for other FGF-response genes will be carried out and compared to FGF ChIP-seq before verification.

One of the challenges in identification of enhancers from ChIP-seq is the assignment of enhancers to genes. This can be overcome by predicting insulator boundaries for genes of

interest as done in this study. This is helpful in associating enhancers to genes as it allows the identification of both proximal and distal enhancers of a gene. A similar strategy has been reported to be used before (Khan et al., 2013).

Thus it is shown here that computational predictions together with histone ChIP-seq improve identification of otic enhancers.

4.11.3 Otic enhancers share transcription factor binding sites

Upon TFBS analysis of the novel enhancers, binding sites for *Sox* family members were identified as particularly enriched. Both *Sox8* and *Sox13* are expressed in the posterior tissues prior to otic placode formation and therefore make good candidates as regulators of *Foxi3 E1.B, Spry1* and *Cxcl14.* Although Hesx1 E1 has no *Sox* binding sites, it has binding sites for *Ap1, Six1, Six4, Gbx2* and *Lmx1a*; all of which are expressed earlier at the PPR and OEPD stages. *Lmx1a* is a particularly good candidate which has been identified as a new transcription factor expressed at the OEPD stage (mRNA-seq; Dr. Jingchen Chen). A preliminary TFBS analysis of all +FGF2 unique enhancers also identified *Lmx1a* as one of the enriched transcription factors. It is a transcriptional activator and it has been shown that *Lmx1a* null mice have malformed ears (Nichols D. H., 2008). The presence of *Lmx1a* binding sites in *Spry1* and *Hesx1* enhancers indicates that it may be involved in positively regulating otic genes ultimately contributing to normal ear development. The role of *Lmx1a* will be further discussed in the next chapter.

Foxi3 E1.B also presents with *Sall1* binding sites. *Sall1* is a repressor that is expressed in the anterior PPR and later in the otic placode (Sweetman et al., 2005). At this time, *Foxi3* expression is lost from the otic placode thus *Sall1* may be shutting off the activity of Foxi3 E1.B at the otic placode stage. In humans, mutations in *Sall1* are associated with

the Townes-Brocks syndrome (TBS) which is a rare, autosomal dominant malformation that presents with anal, renal, limb and ear anomalies (Kohlhase et al., 1998). Moreover, heterozygous mice for *Sall1* mutation also mimic TBS patients by showing sensorineural hearing loss, renal cystic hypoplasia and wrist bone abnormalities (Kiefer et al., 2003). As *Foxi3* promotes the epibranchial fate at later stages (Khatri and Groves, 2013) it can be speculated that *Sall1* may be playing a vital role in removing *Foxi3* expression from the otic placode through binding to Foxi3 E1.B, thus promoting the otic fate at the expense of epibranchial fate.

Foxi3 E2.A which was identified as a trigeminal and epibranchial enhancer presents with binding sites for *Nr2f2* which is expressed in the epibranchial region and *Tcf4* which is expressed in the otic placode (Tambalo, 2015). *Tcf4,* a repressor, may therefore be involved in shutting down Foxi3 E2.A activity in the otic placode. A model for the regulation of *Foxi3, Hesx1* and *Spry1* is given in Figure 4.17.

Further experiments are required to link these TFs with the enhancer activity. This will involve gene knockdown as well as mutation of binding sites in enhancer sequences to assess if enhancer activity is affected. Additionally, a computational approach called reverse-engineering can be utilized to give weight to the link between these TFs and the corresponding gene before designing experiments. This involves using expression data (NanoString, mRNA-seq) to predict interactions between TFs and genes and will be discussed in detail in the next chapter. Using a combination of such approaches will ultimately allow building a high-resolution otic gene regulatory network.

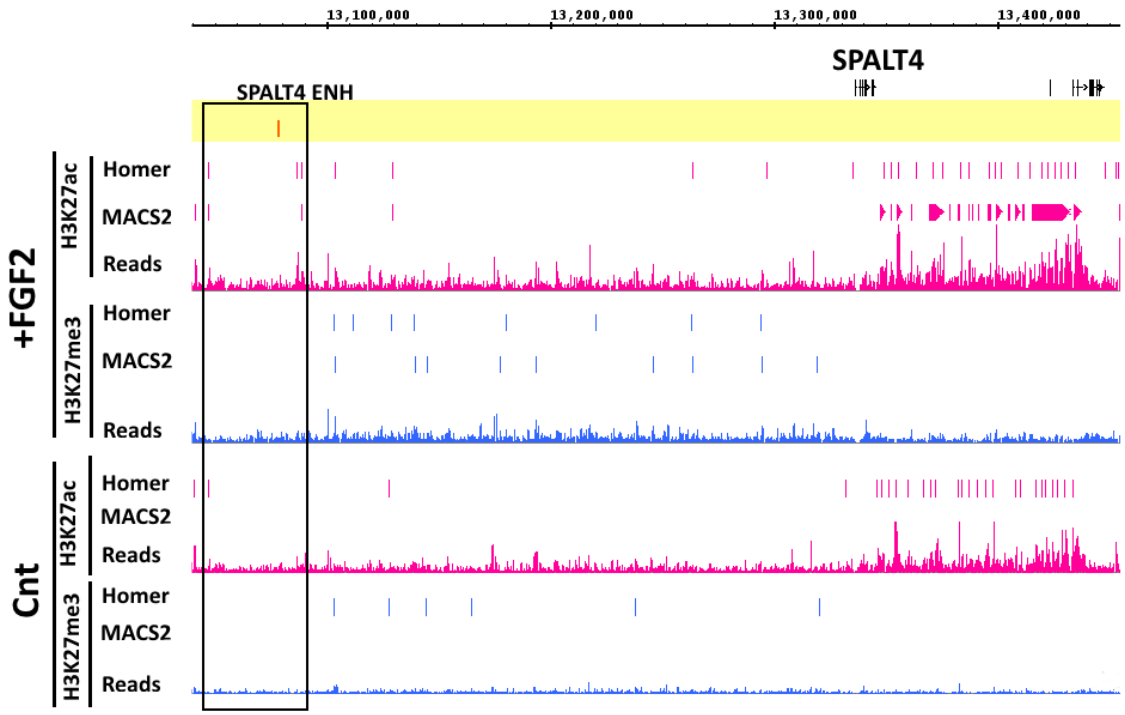


Figure 4.1 *Spalt4* otic enhancer is recovered upon +FGF2 treatment
A browser image for H3K27ac (pink) and H3K7me3 (blue) tracks of +FGF2 and control PPR explants. Peaks are called using both Homer and MACS. The known otic enhancer for *Spalt4* is recovered in +FGF2 ChIP-seq and is flanked by H3K27ac peaks and depleted in H3K27me3. The same region is depleted in H3K27ac in control ChIP-seq.

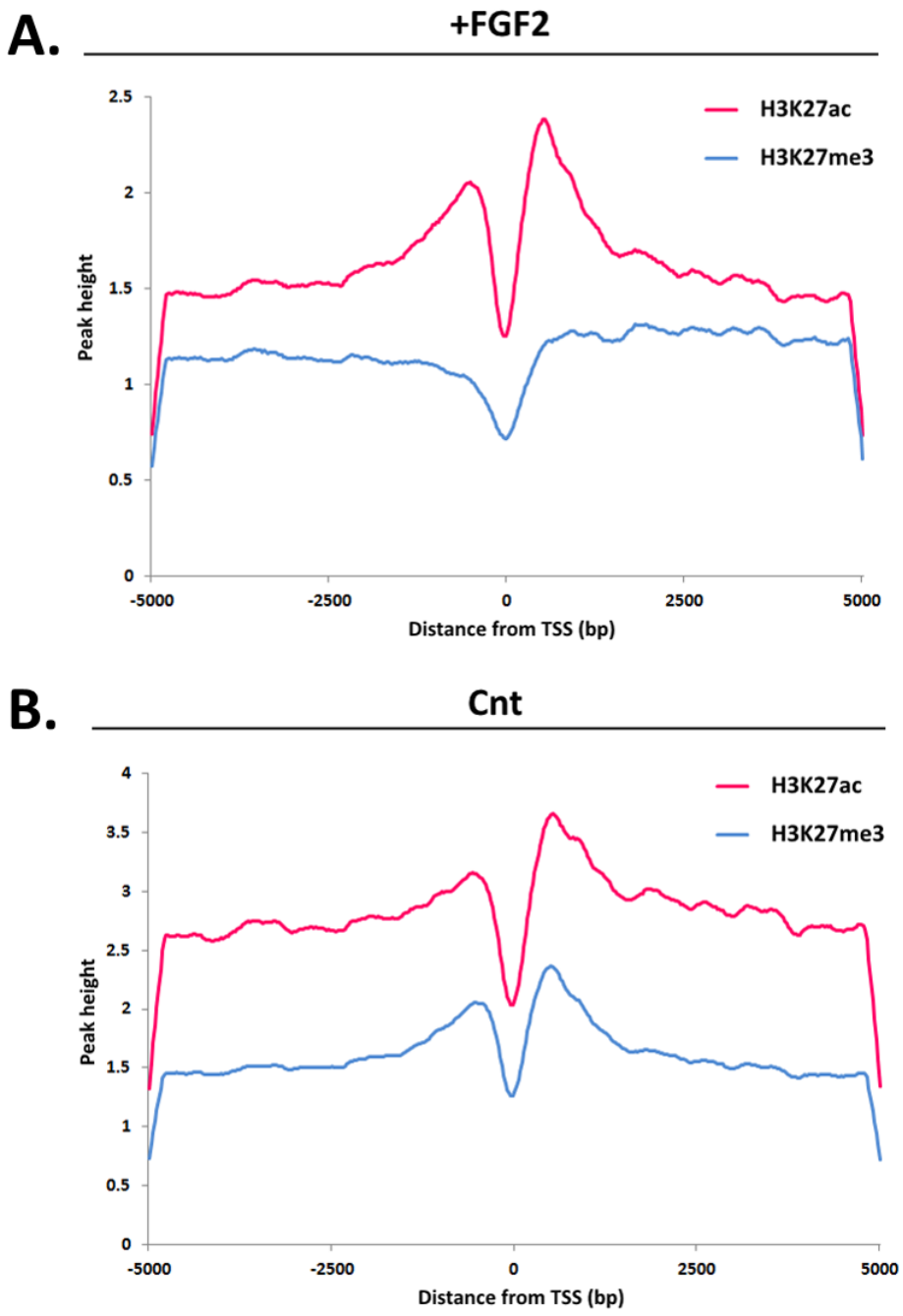


Figure 4.2 Read density distributions for H3K27ac and H3K27me3 around transcription start site (TSS)

Average read density for H3K27ac and H3K27me3 is plotted around TSS. **(A)** FGF2-treated sample. **(B)** Control sample. The x-axis represents distance from the TSS in base pairs and the y-axis represents the read density or height. Read densities are highest close to the TSS and display a peak-dip-peak pattern where the dip-region may bind transcriptional activators or repressors.

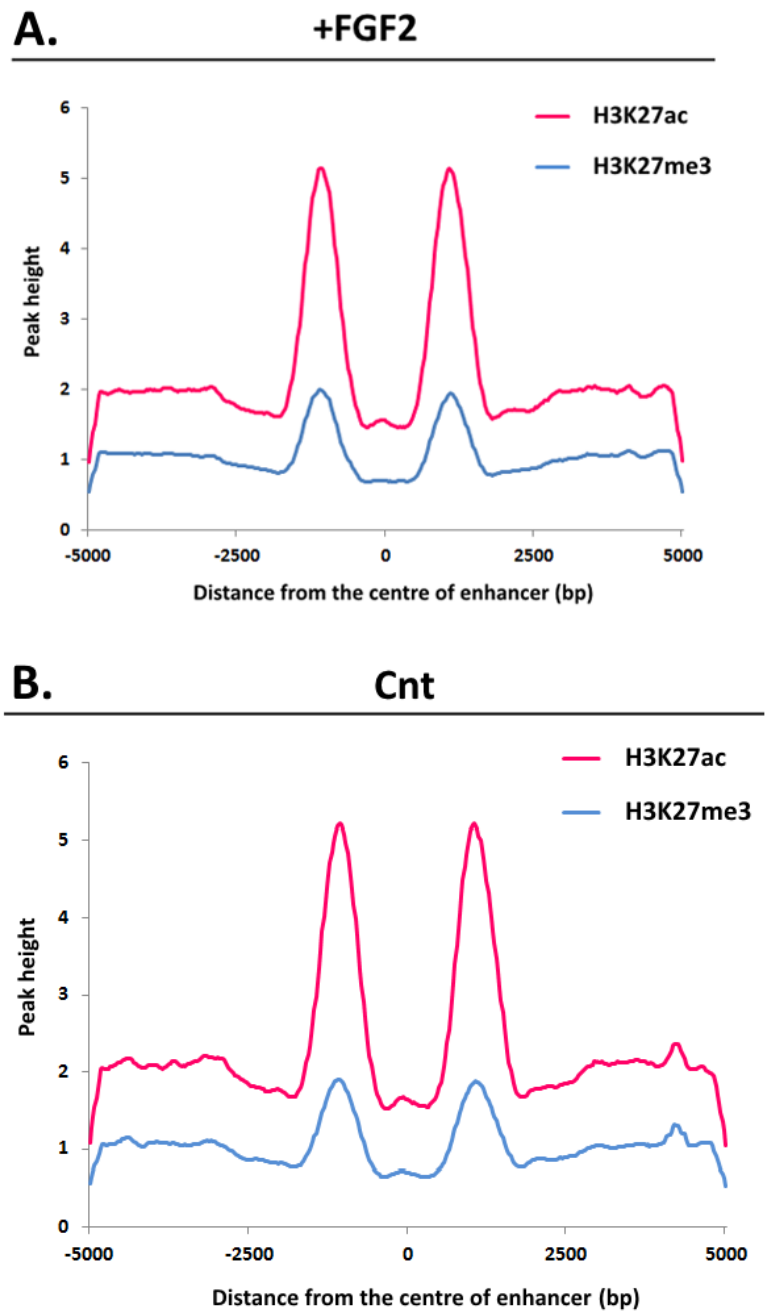


Figure 4.3 Read density distributions for H3K27ac and H3K27me3 from the centre of enhancer

Average read density for H3K27ac and H3K27me3 is plotted from the centre of enhancer. **(A)** FGF2-treated sample. **(B)** Control sample. The x-axis represents distance from the centre of enhancer in base pairs and the y-axis represents the read density or peak height. Similar to the TSS, putative enhancers exhibit a bimodal distribution of H3K27ac and H3K27me3 where the free regions may bind regulators of transcription.

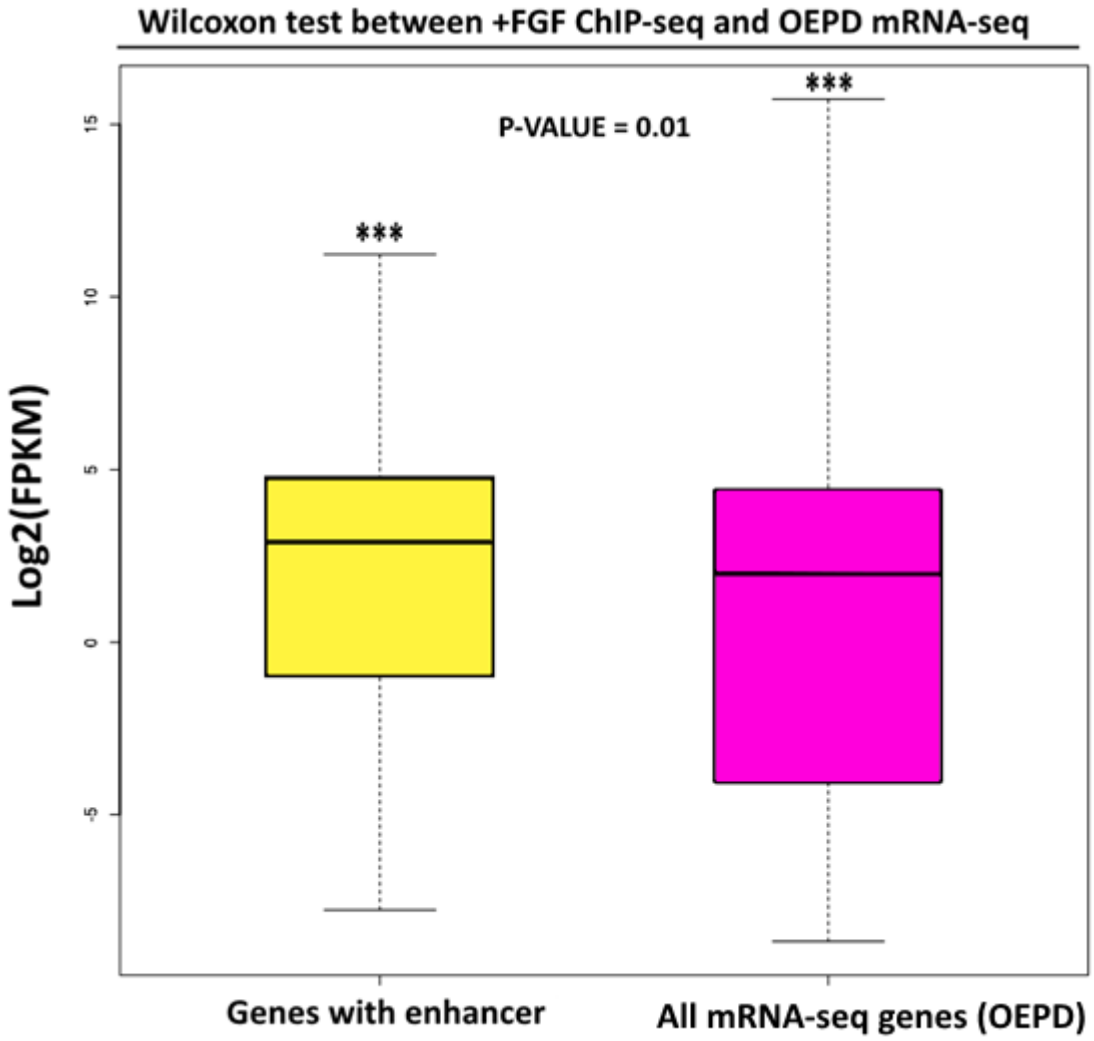


Figure 4.4 Wilcoxon test reveals a significant correlation between +FGF2 ChIP-seq and OEPD mRNA-seq

Putative enhancers for +FGF2 were identified as up to 3 kb regions flanked by H3K27ac peaks and devoid of H3K27me3 peaks. These were then annotated to the nearest gene. mRNA-seq (Chen & Tambalo, unpublished; Tambalo, 2015) identified transcripts in the OEPD. A Wilcoxon test was carried out to test if the mean FPKM of genes with putative enhancers in +FGF2 is greater than the mean FPKM of all OEPD genes. This analysis shows that indeed enhancer associated genes have high expression levels in the OEPD with a significant p-value of 0.01.

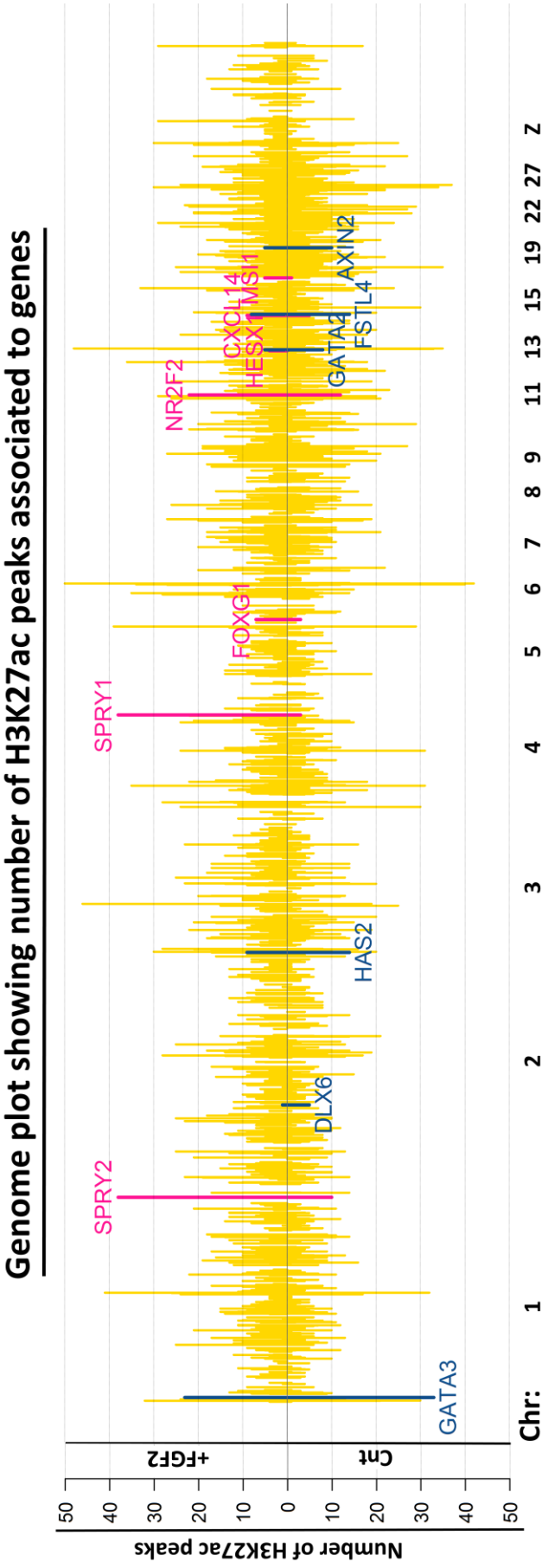


Figure 4.5 Genome-wide distribution of H3K27ac peaks in +FGF2 and control

Figure 4.5 Genome-wide distribution of H3K27ac peaks in +FGF2 and control
After peak-calling, all H3K27ac peaks were assigned to the nearest gene in both +FGF2 and control. In each sample, H3K27ac peaks for each gene were summed. On the x-axis are all chromosomes. The y-axis represents the number of H3K27ac peaks of a gene in +FGF2 and control samples. Some of the FGF2 induced targets display greater acetylation in the presence of FGF2 than in controls including *Spry1/2, Foxg1, Cxcl14* and others. In contrast, transcripts that do not respond to FGF like *Gata3, Gata2, Fstl4* and others display greater acetylation in the control sample.

Figure 4.6 Dense H3K27ac peaks flank AP1 binding sites in +FGF2 ChIP-seq
(A) SeqMINER view of the heatmap for comparison of the distribution of H3K27ac peaks around *Ap1/Fos/Jun* binding sites in +FGF2 and control. First *Ap1* binding sites were identified in +FGF2 putative enhancers. Then to compare these regions with control and to assess if FGF signalling plays a role in recruitment of active marks (H3K27ac) through *Ap1* complex, the H3K27ac peaks of +FGF2 and control were plotted around the *Ap1* binding sites in a window of -/+5kb. **(B)** H3K27ac peak profiles around *Ap1* in +FGF2 (blue) and control (pink) in a window of -/+5kb. The heatmap and profiles indicate that in +FGF2; due to the possible binding of *Ap1* in response to FGF signalling, the regions flanking the *Ap1* sites are H3K27ac-dense whereas the same regions are depleted of H3K27ac in control. **(C)** To assess each cluster Gene Ontology (GO) and KEGG terms were identified using DAVID. GO terms are coloured according to the colours of each cluster. Cluster 3 and 4 contain putative enhancers for some of the earliest FGF-target genes during OEPD induction hence the enriched terms: inner ear morphogenesis and MAPK signalling.

Figure 4.7 Common and unique enhancers in +FGF2 and control cells
(A) Venn diagram showing common and unique control (green) and +FGF2 (red) putative enhancers from ChIP-seq. **(B)** Functional annotation of the unique putative enhancers of +FGF2 and control using DAVID reveals relevant GO and KEGG terms. A p-value cut-off of 0.05 was used corresponding to -Log10(p-value) of 1.3. **(C)** A genome plot provides a genome-wide view of the location of common and unique enhancers in +FGF2 and control samples.

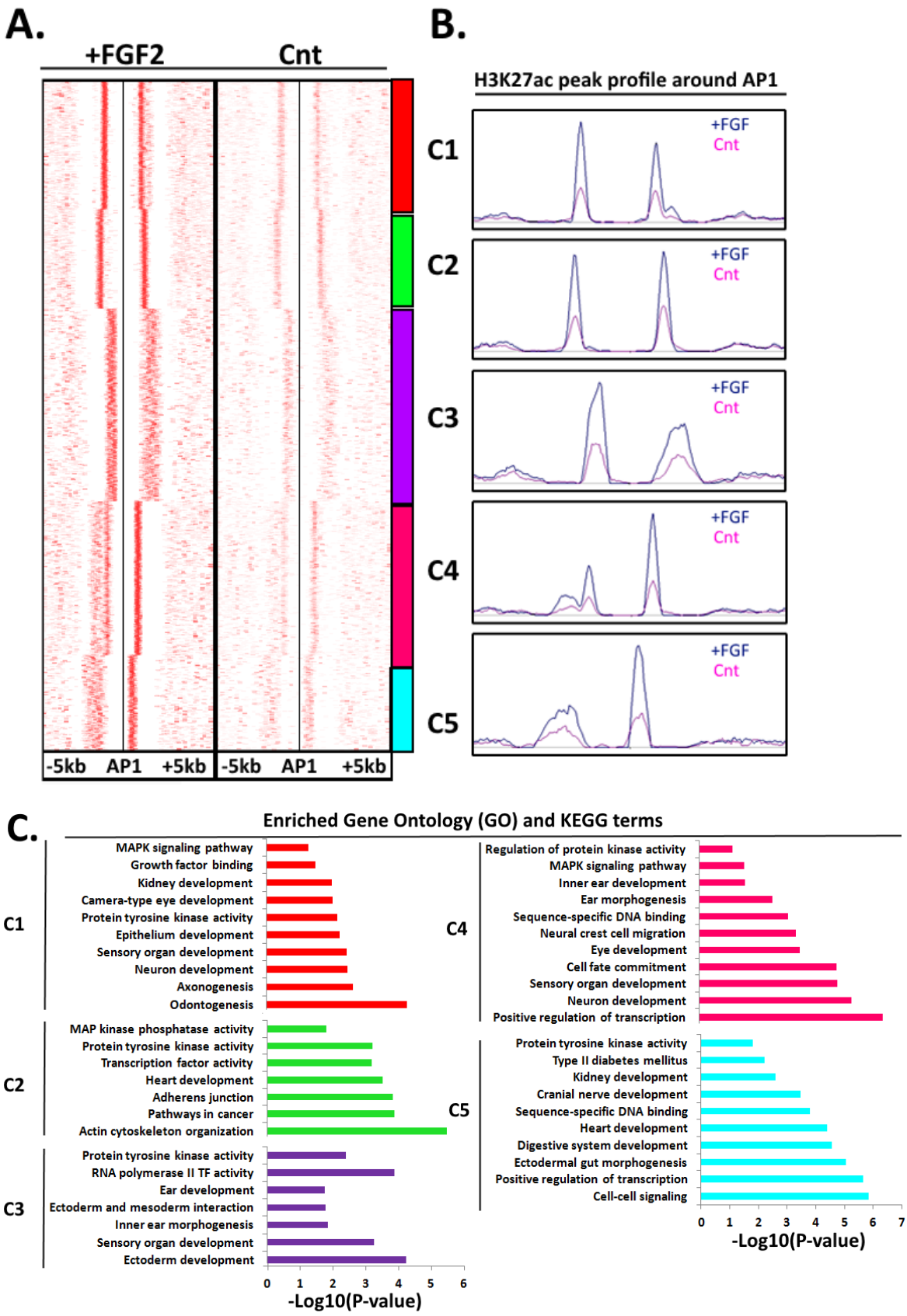


Figure 4.6 Dense H3K27ac peaks flank AP1 binding sites in +FGF2 ChIP-seq

Figure 4.7 Common and unique enhancers in +FGF2 and control cells

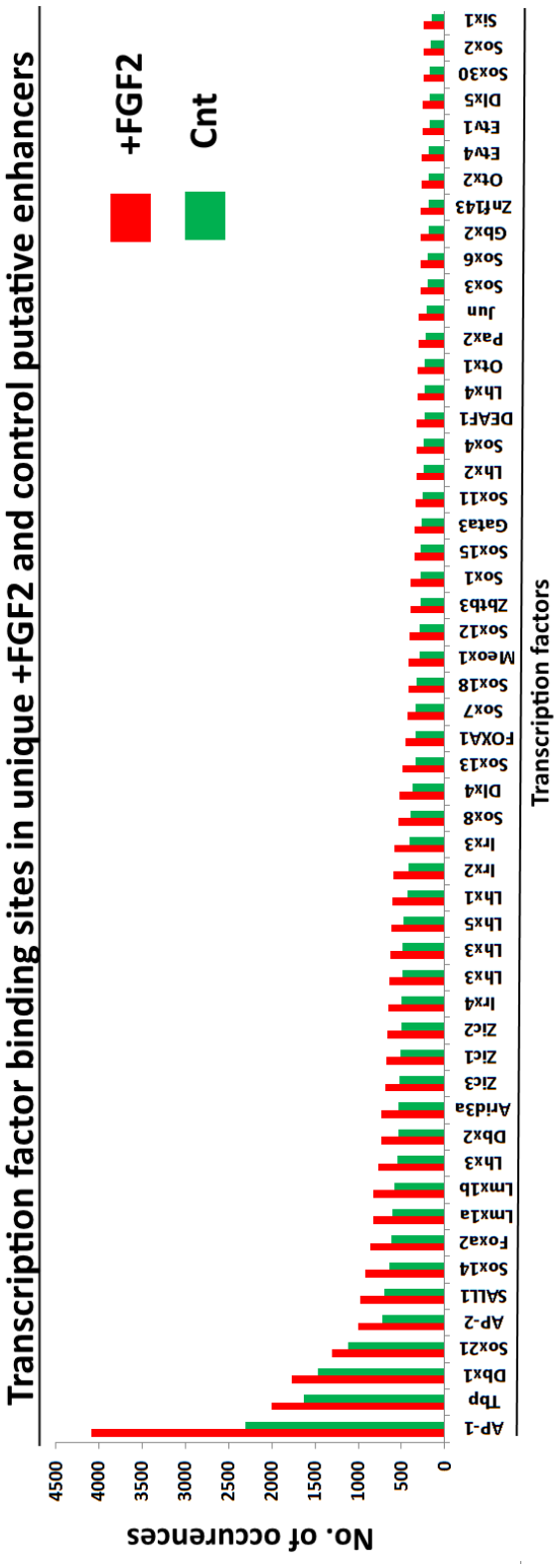


Figure 4.8 Transcription factor binding sites in unique +FGF2 and control putative enhancers

Transcription factor binding site analysis of all putative +FGF2 and control enhancers using RSAT matrix-scan and JASPAR full TF library. Matches of some of the top TFs enriched in +FGF2 putative enhancers and the corresponding matches in control putative enhancers are plotted as a bar graph. As indicated, the most enriched TFBSs in +FGF2 are those for *Ap1, Lmx1a* and *Sox* family members among others.

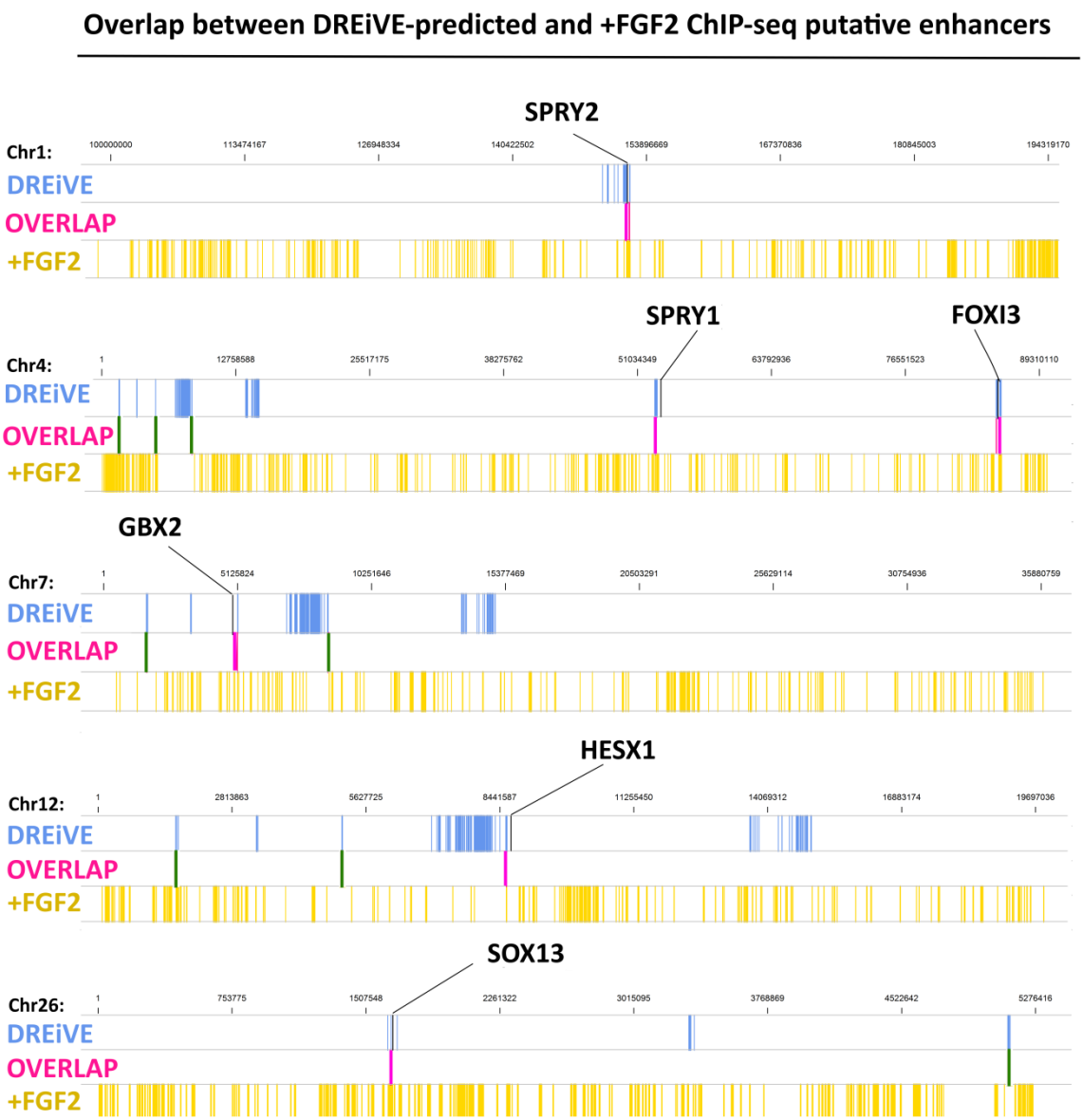


Figure 4.9 Overlap between DREiVe-predicted and +FGF2 putative enhancers
Overlap between DREiVe-predicted and +FGF2 putative enhancers are shown as chromosome plots. DREiVe-predicted enhancers are shown in blue and +FGF2 putative enhancers are shown in gold. In total, 13 predicted enhancers overlap with 10 ChIPseq-identified enhancers for FGF-response genes (shown in pink). Predicted enhancers that overlap with ChIPseq-identified enhancers of other genes are shown in green.

Figure 4.10 Selected putative unique +FGF2 enhancers are enriched in binding sites of otic transcription factors as compared with randomized controls

Figure 4.10 Selected putative unique +FGF2 enhancers are enriched in binding sites of otic transcription factors as compared with randomized controls
Transcription factor binding site analysis of selected putative +FGF2 enhancers using RSAT matrix-scan and customised library (from mRNA-seq). Results of TFBS are represented as a heatmap. The horizontal axis represents transcription factors whereas the vertical axis represents putative selected +FGF2 enhancers. The colours in the heatmap indicate the number of binding sites in each enhancer for the corresponding TF. The enhancers are clustered based on their binding sites. The heatmap indicates that a few TFs are over-represented in a group of putative enhancers (TFs on the left-side of heatmap). these include *Sox* family members, *Sall1* and *Lmx1a* among others.

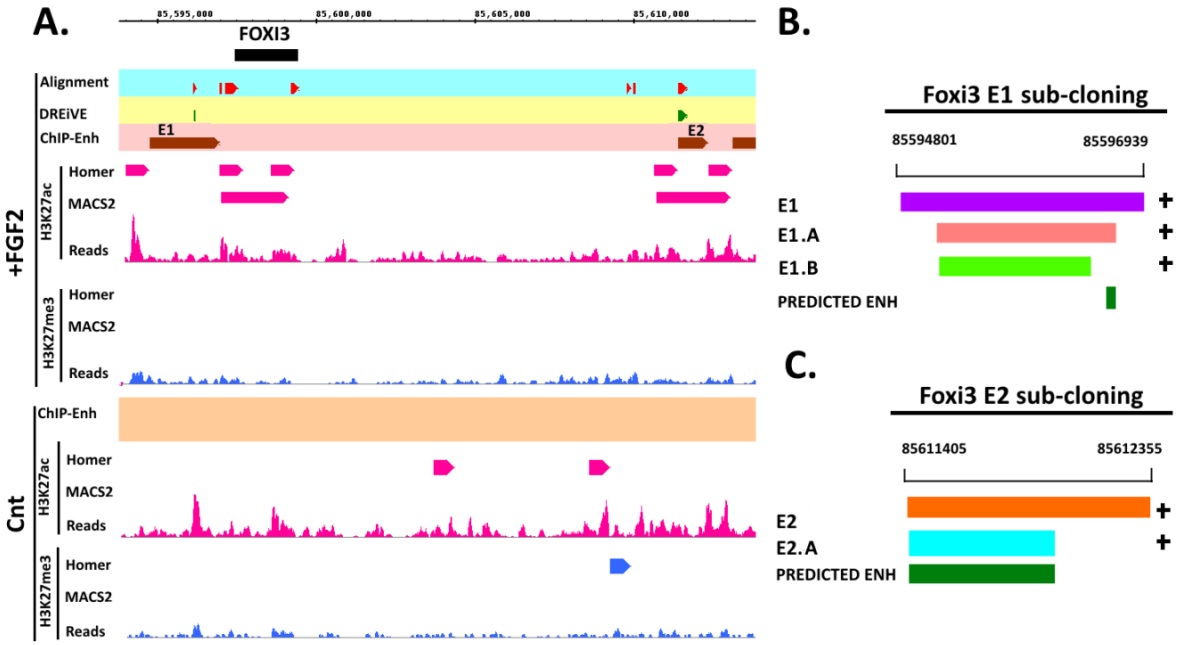


Figure 4.11 Foxi3 enhancers are conserved
(A) IGB browser view of the selected *Foxi3* enhancers. The first track represents conservation from PECAN alignments (ENSEMBL). The second track represents DREiVe-predicted enhancers. The third track represents ChIP-identified enhancers for +FGF2. Tracks in pink and blue are H3K27ac and H3K27me3 respectively. Two peak-callers were used: Homer and MACS2. The two enhancers identified and marked as E1 and E2 (both overlap with predictions and conservation). **(B)** Information on sub-cloning of E1: E1.A is within E1 and contains predicted enhancer; E1.B is within E1.A and lacks the predicted enhancer **(C)** Information on sub-cloning of E2: E2 is the full length enhancer; E2.A is just the predicted enhancer.

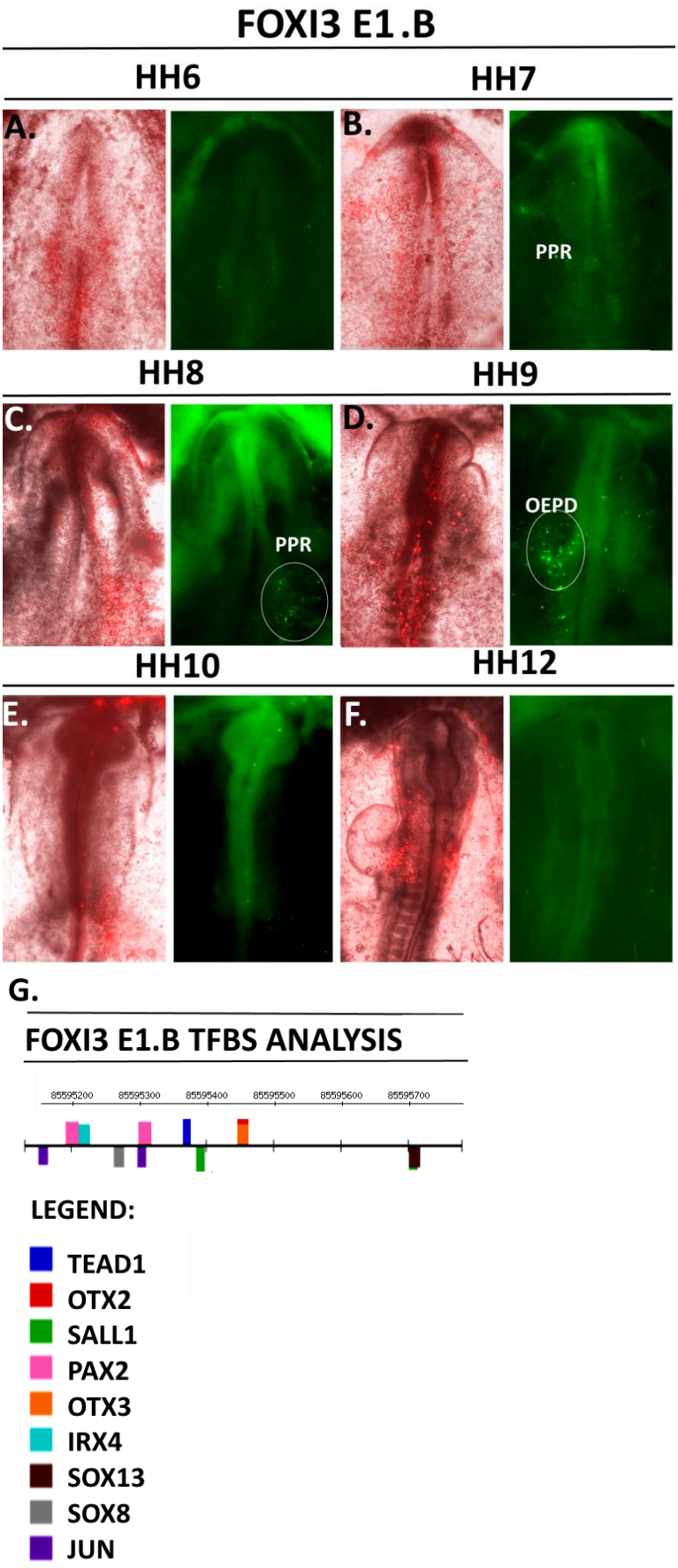


Figure 4.12 In vivo activity of Foxi3 E1.B and its predicted transcriptional inputs

Figure 4.12 In vivo activity of Foxi3 E1.B and its predicted transcriptional inputs
E1.B was electroporated bilaterally into HH4 embryos; while RFP expression is observed widespread, citrine expression is only observed in the future otic territory. E1.B activity was initially detected at HH7, where a few GFP+ cells were observed in the PPR **(B)**. The enhancer activity becomes strong at HH8 and later in HH9 in the OEPD **(C-D)** Enhancer becomes inactive at HH10 and stays off at HH12 **(E-F)**. **(G)** TFBS analysis of Foxi3 E1.B was carried out using customized library containing otic and anterior genes. Each TFBS is coloured on the enhancer sequence at the appropriate location.

Figure 4.13 In vivo activity of Foxi3 E2.A and its predicted transcriptional inputs
Characterization of the in vivo activity of E2 revealed that the sub-clone E2.A (predicted enhancer) is sufficient to drive GFP expression. The enhancer was initially found to be active in the trigeminal placode (TRI) at HH9 **(B)**. The enhancer is also expressed in the epibranchial region (EPI) at HH11 **(C)** and later in the pharyngeal arches (PH) at HH15 **(D)**. The enhancer activity is consistent with the expression pattern of *Foxi3* at these stages. **(E)** TFBS analysis of Foxi3 E2.A was carried out using customized library containing otic and anterior genes. Each TFBS is coloured on the enhancer sequence at the appropriate location.

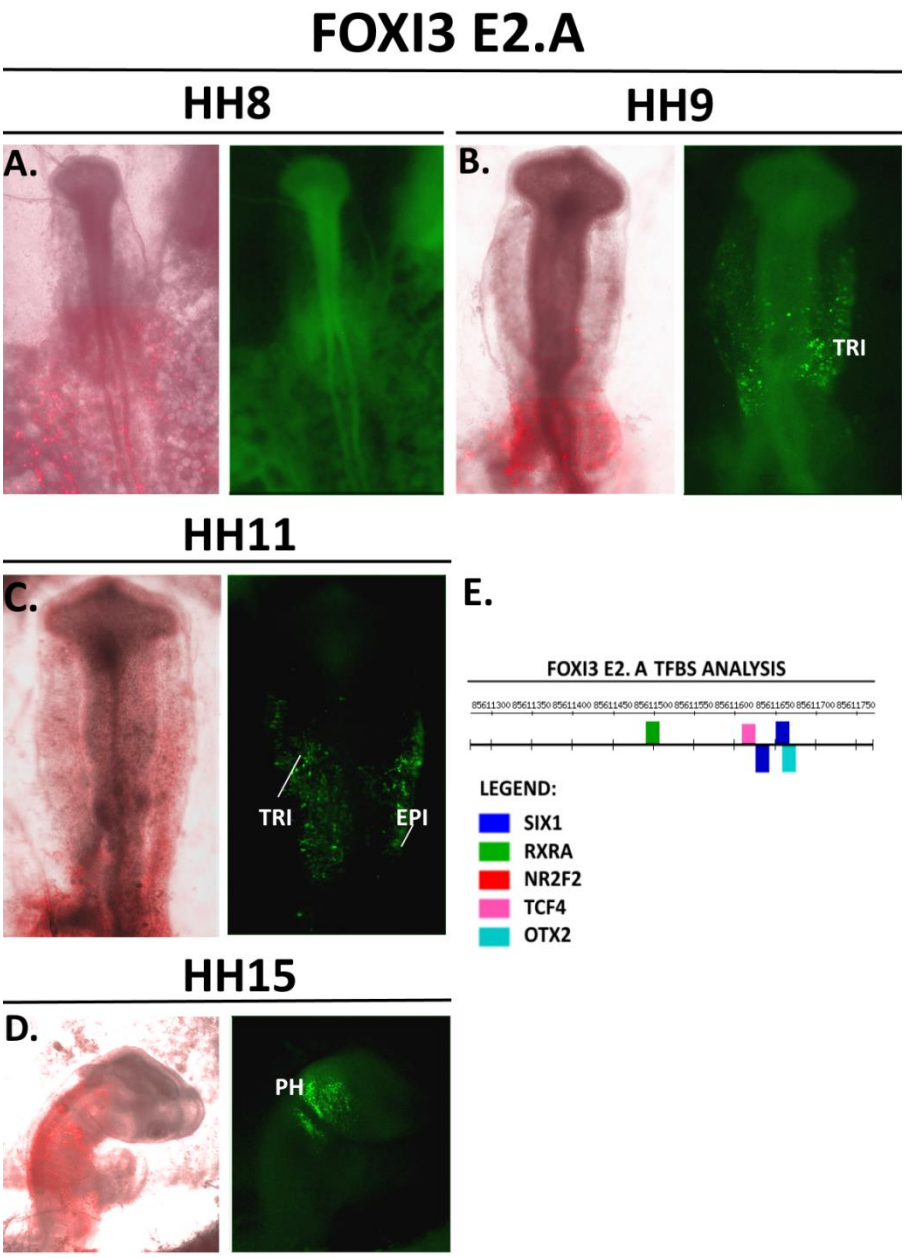


Figure 4.13 In vivo activity of Foxi3 E2.A and its predicted transcriptional inputs

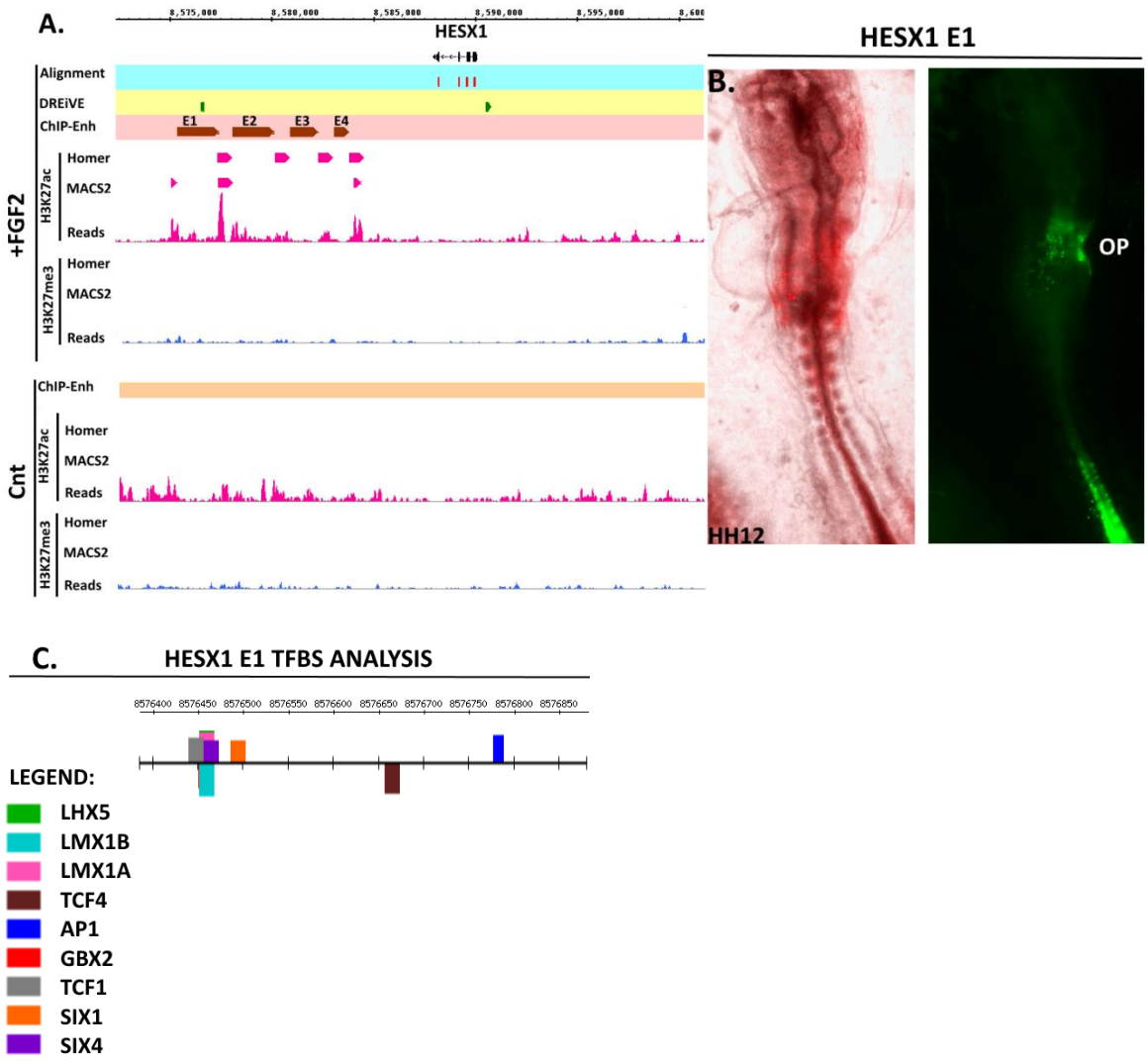


Figure 4.14 In vivo activity of the conserved Hesx1 E1 enhancer and its transcriptional inputs

(A) IGB browser view of *Hesx1* enhancers. The first track represents conservation from PECAN alignments (ENSEMBL). The second track represents DREiVe-predicted enhancers. The third track represents ChIP-identified enhancers for +FGF2. Tracks in pink and blue are H3K27ac and H3K27me3 respectively. Two peak-callers were used: Homer and MACS2. Four unique putative enhancers were found from +FGF2 out of which E1 overlaps with DREiVe-predicted enhancer. **(B)** *In vivo* activity of Hesx1 E1 reveals that it is expressed in otic placode (OP) at HH12. A detailed characterization is required to assess the time window of its activity **(D)** TFBS analysis of Hesx1 E1 was carried out using customized library containing otic and anterior genes. Each TFBS is coloured on the enhancer sequence at the appropriate location.

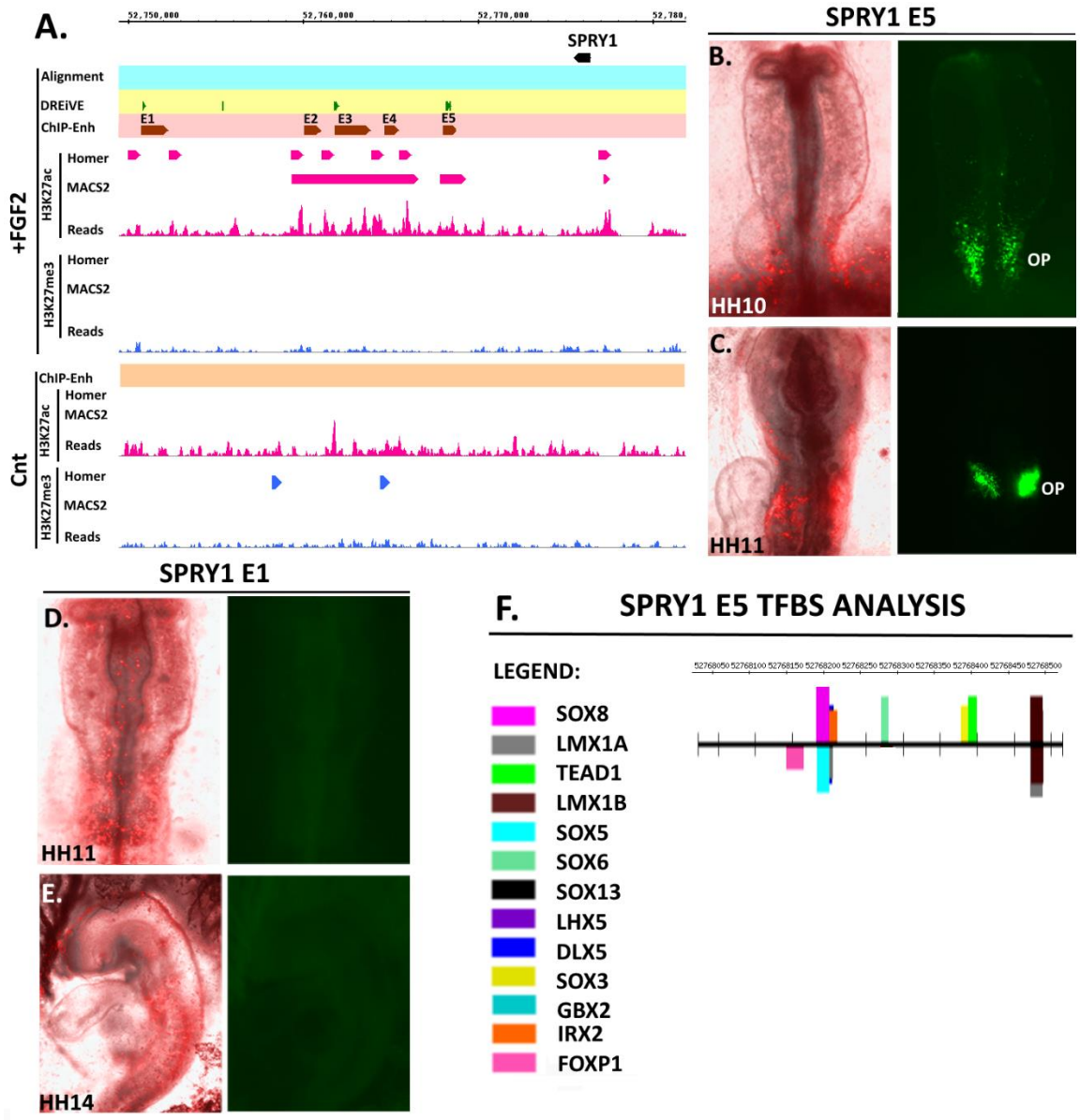


Figure 4.15 In vivo activity of conserved Spry1 E1 and E5 and transcriptional inputs of Spry1 E5

(A) IGB browser view of *Spry1* enhancers. The first track represents conservation from PECAN alignments (ENSEMBL). The second track represents DREiVe-predicted enhancers. The third track represents ChIP-identified enhancers for +FGF2. Tracks in pink and blue are H3K27ac and H3K27me3 respectively. Two peak-callers were used: Homer and MACS2. Five unique putative enhancers were found from +FGF2 out of which E1 and E5 overlap with DREiVe-predicted enhancers. **(B-C)** In vivo activity of Spry1 E5 reveals that it is strongly expressed in the otic placode (OP) at HH10 and HH11 **(D-E)** Spry1 E1 is inactive at the otic placode and otic vesicle stage. **(F)** TFBS analysis of Spry1 E5 was carried out using customized library containing otic and anterior genes. Each TFBS is coloured on the enhancer sequence at the appropriate location.

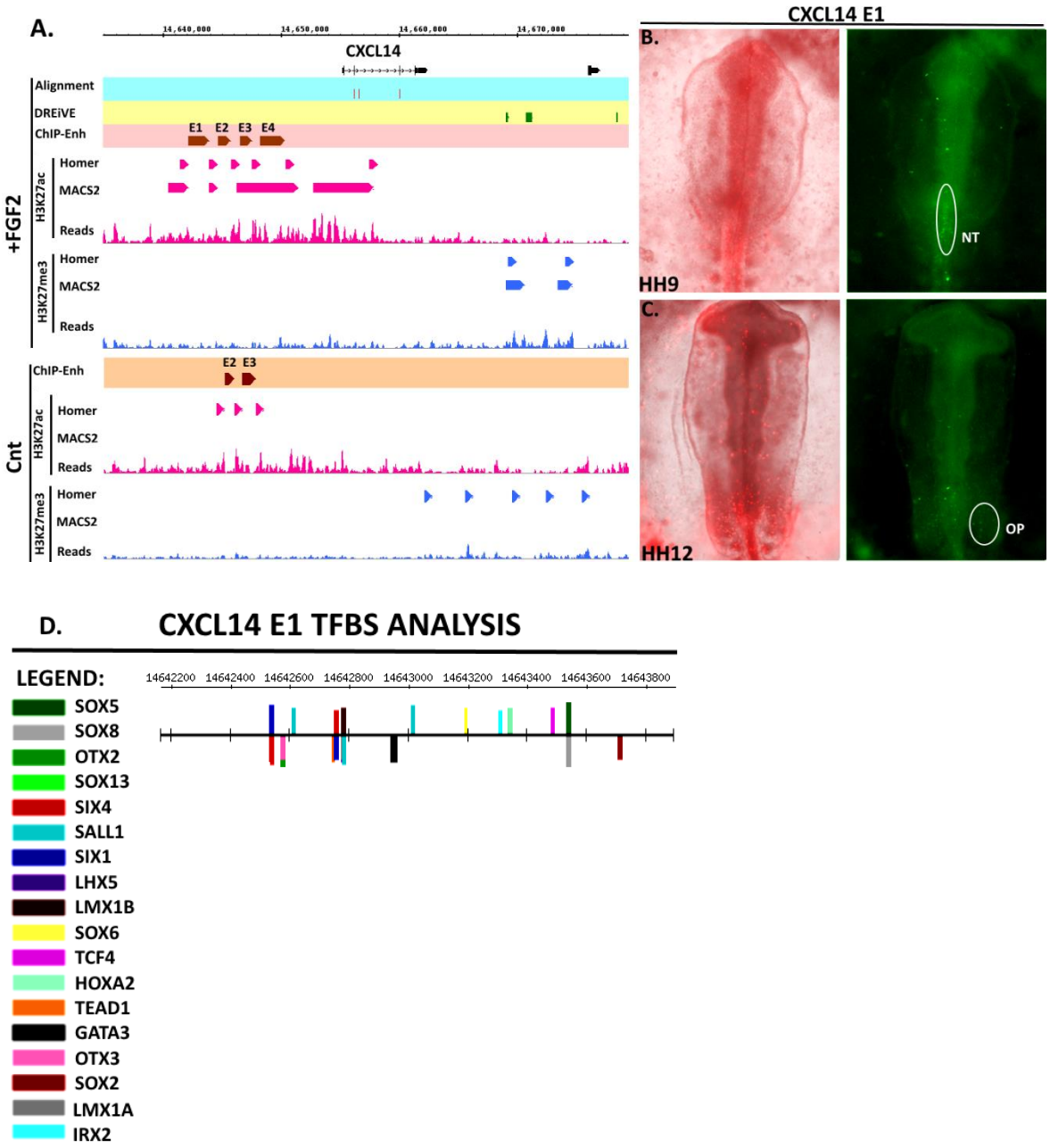


Figure 4.16 In vivo activity of Cxcl14 E1 and its transcriptional inputs
(A) IGB browser view of *Cxcl14* enhancers. The first track represents conservation from PECAN alignments (ENSEMBL). The second track represents DREiVe-predicted enhancers. The third track represents ChIP-identified enhancers for +FGF2. Tracks in pink and blue are H3K27ac and H3K27me3 respectively. Two peak-callers were used: Homer and MACS2. Four putative enhancers were found from +FGF2 out of which E2 and E3 overlap with the putative enhancers in control. No overlap between DREiVe-predicted enhancers and ChIP-identified enhancers was found for *Cxcl14*. A proximal enhancer E1 was thus selected that was unique in +FGF2. **(B-C)** Characterization of *in vivo* activity of Cxcl14 E1 reveals that it is initially expressed in the neural tube (NT) at HH9 and later at HH12; it shows weak activity in the ectoderm surrounding the otic placode (OP). **(D)** TFBS analysis of Cxcl14 E1 was carried out using customized library containing otic and anterior genes Each TFBS is coloured on the enhancer sequence at the appropriate location.

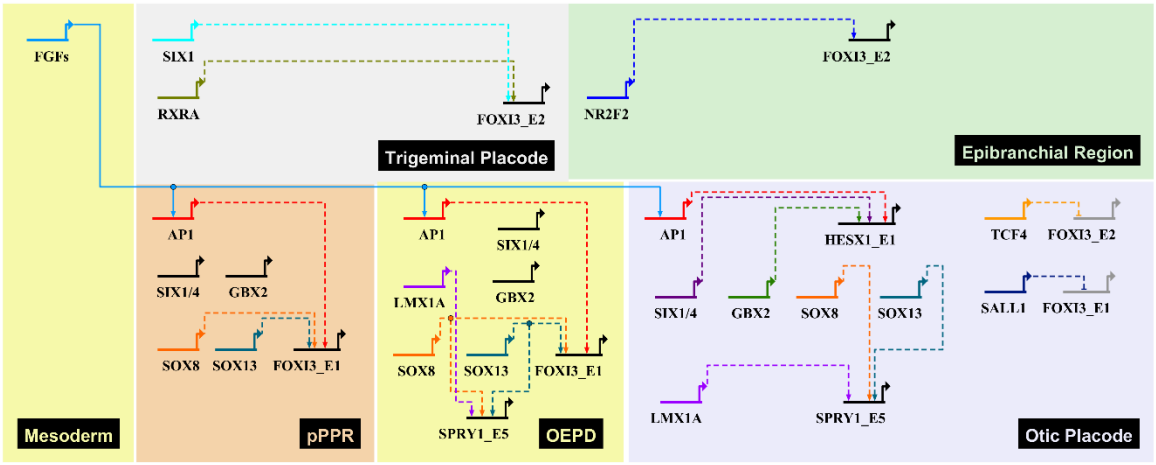


Figure 4.17 A model displaying common transcriptional inputs for the otic enhancers of FGF-response genes

This BioTapestry model represents the common transcriptional inputs for the otic enhancers of FGF-response genes *Foxi3, Hesx1* and *Spry1*. All predicted interactions are represented as dashed lines. FGFs from the underlying mesoderm signal via *Ap1* and up regulate *Foxi3*, *Hesx1* and *Spry1*. However, for this model an interaction between *Ap1* and the corresponding enhancers is shown only if an *Ap1* binding site was found. *Foxi3* expression is recapitulated by two enhancers; E1 which is active from pPPR to OEPD and E2 which is active in the trigeminal and epibranchial regions. Foxi3_E1 is regulated by *Sox8* and *Sox13* and its activity is shut off at the otic placode stage by *Sall1*. Foxi3_E2 is driven by *Six1* and *Rxra* in the trigeminal placode and by *Nr2f2* in the epibranchial region. Its activity is prevented in the otic placode via *Tcf4*. *Hesx1*_E1 is active in the otic placode and is regulated by *Ap1, Six1/4* and *Gbx2*. Spry1_E5 is active in the OEPD and otic placode and is regulated by *Lmx1a, Sox8 and Sox13*.

5. Towards an otic gene regulatory network

5.1 Introduction

Cells adapt to changing environments by altering their gene expression patterns in response to different cues. There are various levels at which gene expression is controlled including transcription, mRNA stability, translation and protein stability (Macneil and Walhout, 2011). Gene transcription is under complex control mediated by different factors including TFs, their co-factors, chromatin modifiers and RNA binding proteins. Transcription factors together with their co-factors control mRNA production by interacting with the regulatory elements of their target genes, which in turn determine the rate of transcription by interacting with the general transcriptional machinery. In all organisms, specific transcriptional programs determine proper development and ensure appropriate responses to environmental cues such as the expression of *Hox* genes that determine anterior-posterior positioning and segmentation in *Drosophila* (Lewis, 1978, Harding et al., 1985). Often these transcriptional programs are complex networks of interactions between multiple TFs and their targets, where many TFs act synergistically to regulate one or more genes. Building gene regulatory networks (GRNs) allows a better understanding of such interactions providing insight into the systems-level mechanisms of gene regulation that control growth and development (Levine and Davidson, 2005, Long et al., 2008, Macneil and Walhout, 2011). Recent developments in computational approaches have made it possible to study these regulatory interactions by modelling GRNs using large-scale expression data (Gardner and Faith, 2005, Margolin et al., 2006, Bansal M, 2007, Markowetz and Spang, 2007, Huynh-Thu et al., 2010). This is referred to as “Reverse engineering” or “Inference”. Typically, a GRN is composed of nodes (genes) joined by edges (regulatory interactions) going from the TF to the target gene. GRNs are useful in studying the flow of information in a biological system and help to

identify circuits within the network that may be used for specific purposes or under specific conditions (Levine and Davidson, 2005, Stathopoulos and Levine, 2005, Macneil and Walhout, 2011).

In the previous chapters, I discussed the identification of enhancers for otic genes using computational methods and histone ChIP-seq. This resulted in the identification of five novel otic enhancers for early FGF response genes. Subsequently, these otic enhancers were screened for transcription factor binding sites to determine their regulators. In this chapter, I will use a systems-level approach and discuss the inference of GRNs using expression data from NanoString and mRNA-seq and highlight the importance of this strategy in understanding the underlying biology. This will also allow confirmation of the TF to target interactions proposed in previous chapters and predict potential otic regulatory interactions that can then help to form plausible hypotheses and drive further experiments. At the end I will utilize this information to present a gene regulatory network specifying stages of otic development.

5.2 A predicted GRN for NanoString data and mRNA-seq using GENIE3

As described in Chapter 3, section 3.2, two sets of NanoString experiments including 126 genes were performed by Dr. Monica Tambalo. In the first set, posterior PPR (HH6) was cultured with or without FGF2 for 6, 12 and 24 hours to identify the temporal response to FGF signalling as cells acquire otic identity. In the second set of experiments, the posterior PPR with the underlying mesoderm (source of FGF) was dissected and treated with DMSO (control) or SU5402 to block FGF signalling and changes in gene expression were analyzed in a time course. The normalised data (Chapter 2, section 2.8) were used to infer a GRN using GENIE3 (Chapter 2, section 2.13). As the NanoString probeset consists of TFs, read-outs for different signalling pathways and housekeeping genes, a list

of TFs from the NanoString data was provided separately as input to GENIE3 to aid GRN prediction by specifying only TFs as regulators. Moreover, housekeeping genes and genes with low values (<0.00004) were removed before GRN prediction. This reduced the number of genes to 109 out of which 58 were TFs. To create a GRN, each gene in the expression file is considered as a target gene that can potentially be regulated by one or more TFs. GENIE3 attempts to explain the expression profile of a target gene from the expression profiles of all TFs, and then calculates the importance measure of each TF in the prediction of a target gene's expression profile. The importance measure represents the significance of a putative regulatory link. In this way, all regulatory links are calculated and ranked according to the importance measure, where larger values indicate higher significance. For NanoString data, a threshold of 0.005 was used as it allowed the recovery of 40% of the known otic interactions (20 out of 49 for 14 genes). The same threshold was used recently by others and is deemed stringent (Potier et al., 2014). The resulting network was further analyzed as described in the next sections.

The NanoString data was limited in providing a systems-level view of the development from PPR to placode stages because of a limited number of genes included. Therefore, data from an mRNA-seq experiment were used for GRN inference; mRNA seq was carried out from anterior and posterior PPR, the OEPD, the otic placode from different stages, the future lens-olfactory region ($Pax6^+$ region at HH8), the future trigeminal/epibranchial territory ($Pax3^+$ region at HH8) and non-neural ectoderm from HH6 (Dr. Jingchen Chen, Dr. Monica Tambalo and Ramya Ranganathan). To gain an overall insight into the regulatory mechanisms that govern how cells go from a simple PPR stage to specific fates such as otic, lens and trigeminal placodes, all samples were used for GRN inference. It was hypothesized that inherent differences in gene expression among these tissues will allow the segregation of modules that are specific to their

development and fate. To prepare mRNA-seq datasets for GENIE3, the data were first analyzed as follows (also see Chapter 2, section 2.9.1): the sequenced reads were aligned using TopHat 2 (v2.0.7) to the chicken genome (Galgal4.71). Gene annotations from Ensembl (Galgal4.71) were used to assemble transcripts using Cufflinks v2.1.1 and differentially expressed genes were identified using Cuffdiff v2.1.1. From the resulting FPKM values, those below 10 were set to 0. This value was chosen because the expression of these genes cannot be detected using *in situ* hybridisation and they were therefore considered to be absent. Since the number of genes in mRNA-seq was about 17,000, a predicted GRN incorporating all genes is very difficult to analyze. Therefore, in the first instance, only TFs (identified from the Animal Transcription Factor Database (Animal TFDB) http://www.bioguo.org/AnimalTFDB) were considered for GRN prediction with a total of 534. All TFs were used as regulators. The importance measure threshold for mRNA-seq was kept as 0.001, which recovered 35% (26 out of 76 interactions for 23 genes) of known otic interactions. In contrast, a more stringent threshold of 0.005 recovered only 10% of known interactions and was therefore considered too stringent.

First, the analysis of the NanoString network will be discussed followed by the mRNA-seq network and a final model specifying the regulatory interactions from PPR to otic placode.

5.3 GENIE3 recovers known otic interactions in the predicted NanoString network

The first step towards establishing the reliability of the predicted network is to test if it recovers known interactions. Known interactions from PPR to otic placode stages are discussed in detail in Chapter 1 and summarized in Table 1.1 and Figure 1.2.

After applying a threshold of 0.005, GENIE3 produces a directed network consisting of 109 nodes and 3439 edges. The resultant network was displayed and coloured in Cytoscape (Figure 5.1 A). Each node is coloured according to its out-degree to identify highly active TFs regulating most of the other genes in the network. Low out-degrees are depicted in green and high out-degrees in orange and red. The width of each edge is mapped to the importance measure of the regulatory link. Thicker edges indicate higher significance. The network shows a number of red and orange nodes indicating that these genes may be some of the most active TFs. However, the network will be dissected and analyzed in detail only after some degree of reliability has been established. To do so, GENIE3 network was overlapped to known interactions. A Cytoscape view of the known PPR to otic placode interactions is shown in Figure 5.1 B. The interactions are coloured according to the region in which they are expressed in the normal embryo and the interaction takes place (see Table 1.1), with PPR interactions in green, posterior PPR interactions in purple, OEPD interactions in orange and otic and epibranchial interactions in pink. The overlap of GENIE3 network and known interactions is shown in Figure 5.1 C using the same colour code. Out of the 49 known interactions for 14 genes, 20 interactions were recovered (40% recovery; see Table 5.1). As most of the known interactions are not confirmed to be direct or indirect, the predictions can be used to organize the overlapping interactions and their corresponding genes into a hierarchy. After the recovery of known interactions, the network was deemed reliable to carry out further analysis.

Table 5.1 Overlap between GENIE3 NanoString network and known PPR to otic placode interactions

Source	Target	Interaction	IM	Region
DLX5	EYA1	Activates	0.012964	pPPR
DLX5	PAX-6	Represses	0.010109	PPR to OEPD
DLX5	DLX6	Activates	0.04266	pPPR
DLX5	SIX4	Activates	0.008157	pPPR
DLX6	DLX5	Activates	0.098702	pPPR
DLX6	PAX-6	Represses	0.033914	PPR to OEPD
DLX6	EYA2	Activates	0.010585	pPPR
EYA2	FOXI3	Activates	0.031732	PPR to OEPD
EYA2	PAX2	Activates	0.08152	PPR to OEPD
FOXI3	DLX6	Activates	0.006777	pPPR
FOXI3	EYA2	Activates	0.031732	pPPR
FOXI3	FOXG1	Activates	0.011318	PPR to OEPD
FOXI3	SIX4	Activates	0.043597	pPPR
GBX2	PAX2	Activates	0.029614	PPR to OEPD
IRX1	SIX1	Activates	0.019387	Regionalization of PPR
OTX2	PAX-6	Activates	0.019503	Regionalization of PPR
PAX2	FOXI3	Represses	0.007842	PPR to OEPD
PAX2	FOXG1	Activates	0.065659	PPR to OEPD
PAX2	GATA3	Activates	0.014371	PPR to OEPD
SIX1	PAX-6	Activates	0.048144	Regionalization of PPR

5.4 Dissecting the predicted NanoString network: Top 500 interactions

In the first instance, to interpret the network, all interactions within the threshold of 0.005 were sorted in decreasing order with the interaction with highest importance measure at the top. Then the top 500 interactions were viewed and analyzed in Cytoscape (Figure 5.2). The nodes were coloured and sized according to their centrality in the network and the edges were weighted according to their importance measure. The network clearly

indicates three major circuits or modules C1 (pink), C2 (blue) and C3 (purple). In Chapter 3, section 3.3 and section 3.4, FGF responsive up- and downregulated genes were identified and I therefore investigated the position of these genes in the network. Careful analysis reveals that the network indeed organizes genes according to their temporal response to FGF. C1 seems to house some of the earliest FGF response genes (6 hrs) including the positively regulated transcripts *Etv5, Foxi3, Gbx2, Pax2, Hesx1* and *Eya2* (highlighted in pink). Other positively regulated genes include *Hey1* and *Znf462*. Some genes present in this module are not significantly upregulated by FGF after 6 hrs, but inhibition of FGF results in their loss indicating that FGF is not sufficient to induce their expression, but yet required. These include *Six3, Irx1* and *Hsf2*. Thus, transcripts regulated rapidly by FGF are clustered in the same module. It is also interesting to note that some of the earliest negatively regulated genes are also present in this module including *Gata2, Follistatin-like 4, Foxm1* and *Irx2*. Within this module, *Hesx1,* a repressor, is the most central or active gene. Although *Foxi3, Gbx2, Etv5* and *Pax2* are not as central, they are co-expressed in the OEPD. Thus, it is possible that together they repress non-OEPD transcripts like *Irx2*, and *Follistatin-like 4*. As described in Chapter 1, FGF signalling promotes otic at the expense of other fates (Martin and Groves, 2006). While GENIE3 cannot distinguish between negative and positive interactions, the architecture of this module clearly recapitulates this role of FGF signalling as cells rapidly transit to OEPD identity.

The second module in the network C2 harbours genes that respond to FGF at later time points (after 12 or 24 hrs), particularly some that are negatively regulated. These genes include *Geminin, Sox9, Gata3, Sox10, Irx3* and *Zic1*, which are all down regulated after 12 hrs (highlighted in blue), as well as *Zfhx1b* which is repressed 24 hrs after FGF treatment. It is important to note here that *Geminin, Zic1* and *Zfhx1b* are normally

expressed in the neural plate (Papanayotou et al., 2008, Khudyakov and Bronner-Fraser, 2009, Dady et al., 2012), where they do not overlap with otic transcripts. The only genes in this module that are activated by FGF are *Sall1* (highlighted in blue, upregulated after 12 hrs), *Sall4* and *Tox3* (up regulated after 24 hrs). Some genes in this cluster do not change upon treatment with FGF but are expressed in the otic placode. These include *Eya1, Six4* and *Sox3* (Litsiou et al., 2005, Matsumata et al., 2005, Christophorou et al., 2010). However, inhibition of FGF results in downregulation of *Eya1*and *Sox3* after 6 and 12 hrs respectively indicating that FGF is not sufficient to induce their expression, but yet required. The most central gene in this module is *Lef1,* a WNT target (Ohyama et al., 2006) which is initially upregulated by FGF after 6 hrs but downregulated after 24 hrs. As discussed previously (Chapter 1, section 1.4), mesodermal FGFs control WNT signalling in the hindbrain which is crucial for promotion of the otic fate (Ladher et al., 2000, Ohyama et al., 2006, Park and Saint-Jeannet, 2008, Urness et al., 2010). Since *Lef1* is the most central gene, it can be concluded that FGF may be modulating WNT signalling hence leading to promotion of otic development at later stages.

The third module in the network is C3 (highlighted in purple) and contains transcripts that are normally expressed in the anterior placode domain including *Pax6, Sstr5, Pnoc, Nfkb1, Otx2* and *Dlx6.* Some genes in this module are expressed in the entire PPR such as *Six1* and *Dach1* (Litsiou et al., 2005, Barembaum and Bronner-Fraser, 2007, Sato et al., 2012) whereas *Lmx1b* is initially expressed anteriorly but later in the otic placode (Abello et al., 2010). Although *Six1* does not change in response to FGF treatment, *Pax6, Pnoc, Sstr5, Ptpru, Lmx1b, Otx2, Lrp11* and *Dlx6* are reduced after 6 hr FGF treatment (see Chapter 3, section 3.3) and *Cited2* and *Dach1* are up regulated upon FGF inhibition after 24 hrs (Chapter 3, section 3.4). *Pax6* and *Dach1* appear to be the most central genes in

this module followed by *Lmx1b* and *Nfkb1*. Thus it can be concluded that this module mainly houses genes that are downregulated after FGF treatment.

This analysis indicates that modelling a GRN from time-series expression data allows a systems-level view of the flow of biological information during development and helps in the identification of modules that are specific to certain processes. For example, module C1 including most OEPD genes may mediate otic development, while module C3 including anterior and lens genes may antagonise OEPD specification and promote anterior fates. The co-clustering of genes in a module allows us to formulate a new hypothesis for their function which we can then test. Although the network clearly indicates three modules, it is important to note that there is also communication between them as indicated by interactions going from one module to another. However, there are clearly more interactions within a particular module than between modules. Some of the predicted regulatory interactions will be analyzed in the next section.

5.5 Top regulatory interactions of FGF response genes in predicted NanoString network

After analyzing the network architecture, the next step is to identify the regulatory relationships among individual genes in detail. In Chapters 3 and 4, putative enhancers were identified for a few FGF response genes including *Foxi3, Gbx2, Etv4, Hesx1, Cxcl14*, *Sox13, Spry1* and *Spry2* using computational methods and histone ChIP-seq and a few enhancers were indeed active in the otic placode. Here, I compare regulatory relationships predicted through TFBS analysis of the otic enhancers and those predicted in the GENIE3 network. Furthermore, as the predicted GRN is directional, it can be used to establish a hierarchy among genes in the known otic network.

The predicted GRN only contains a limited number of genes as the early FGF response genes *Sox13, Etv4, Spry1* and *Spry2* are not present in the NanoString probe set. Therefore, the predicted interactions for *Foxi3, Gbx2, Hesx1* and *Cxcl14* were analyzed along with some selected FGF-response genes from each of the modules discussed above. For each gene, the top 25 interactions are presented in Figure 5.3.

Although *Cxcl14* is a chemokine and not a TF, its regulators were analyzed because it responds strongly and rapidly to FGF signalling (Figure 5.3 A). *Cxcl14* is expressed in the pPPR and OEPD but later restricted to a ring of ectoderm surrounding the OP (Chapter 3; Figure 3.6). In total there are seven TFs predicted to regulate *Cxcl14*; among those *Irx2* and *Sall1* (red edges) also have binding sites in the enhancer of *Cxcl14* (Chapter 4; Figure 4.16). *Etv5, Gbx2, Foxi3* and *Mynn* are all expressed in the PPR (Lunn et al., 2007, Paxton et al., 2010, Khatri and Groves, 2013, Tambalo, 2015) and therefore good candidates for regulating *Cxcl14. Irx2* is mainly expressed in the pPPR and neural tissues (Goriely et al., 1999) and the presence of a binding site in *Cxcl14* enhancer makes it a plausible candidate for further investigation. In contrast, both *Sall1* and *Hesx1* are repressors (Carvalho et al., 2003, Sweetman et al., 2005) and are initially expressed anteriorly but later expressed in the otic placode [(Sweetman et al., 2005), Chapter 3; Figure 3.6] thus these may be involved in preventing *Cxcl14* expression in the otic placode. To summarise, the top choices for further investigation are the two interactions that are also identified from TFBS analysis of *Cxcl14* enhancer (Figure 5.3 A, red).

Three TFs *Gbx2, Pax2* and *Etv5* are predicted to regulate *Hesx1* expression (Figure 5.3 B). All are expressed in the otic placode prior to *Hesx1* (Streit, 2002, Lunn et al., 2007, Paxton et al., 2010) and thus good candidate upstream regulators. In support of this, the

Hesx1 enhancer (Chapter 3, Figure 4.14) harbours a *Gbx2* binding site. In contrast, *Hesx1* is a transcriptional repressor (Carvalho et al., 2003) and is predicted to control *Cxcl14*, *Foxi3* and *Etv5* as well as the epidermal keratin *Keratin19*. *Etv5* is initially co-expressed with *Hesx1* in the otic placode, but is later reduced, suggesting that *Etv5* may activate its own repressor. In contrast, *Foxi3* (pPPR) is initially present in a fashion complementary to *Hesx1* (aPPR and anterior neural plate, Chapter 3, Figure 3.6), and later lost from the placode (Chapter 3, Figure 3.6) as *Hesx1* begins to be expressed indicating a repressive interaction. Likewise, it is possible that *Hesx1* negatively regulates epidermal *Keratin19* (Figure 5.3 B; green), which is repressed by FGF signalling. *Hesx1*, as an FGF-response gene, may mediate this FGF function. In summary, the top choice for further investigation is *Gbx2* as it has a binding site in *Hesx1* enhancer.

Sox10 (Figure 5.3C) is a key gene in otic specification (Barembaum and Bronner-Fraser, 2007), although it is initially downregulated by FGF signalling (C2, Figure 5.2). *Sox10* is predicted to provide positive input for *Sall1* and *Sall4, Six4* and *Mynn*; all four genes are expressed in the otic placode (Melnick et al., 2000, Streit, 2002, Sweetman et al., 2005, Barembaum and Bronner-Fraser, 2007) and it is therefore feasible that *Sox10* controls their expression. Interestingly, while the GRN predicts a positive regulatory loop between *Sall4* and *Sox10, Sox10* appears to promote the expression of its own repressor *Sall1* in a negative feed-back loop. Finally, the transcriptional repressor *Zfhx1b* (Dady et al., 2012) is predicted to repress *Sox10* and vice versa; this is consistent with the fact that both factors do not co-localise. While *Zfhx1b* is present in the neural tube (Figure 5.3C, cyan; (Dady et al., 2012)), *Sox10* is confined to the otic placode and neural crest cells (Barembaum and Bronner-Fraser, 2007). Thus, *Sox10* seems to promote the expression of otic genes and repress neural fate.

Figure 5.3 D shows interactions for one of the fundamental otic genes, *Foxi3*. *Foxi3* is initially expressed in the PPR and OEPD but later in the trigeminal and epibranchial regions (Chapter 3; Figure 3.6). It is predicted to have positive regulatory loops with *Etv5* and *Gbx2* with which it is co-expressed in the PPR (Lunn et al., 2007, Paxton et al., 2010, Khatri and Groves, 2013). It is also predicted to regulate PPR and OEPD genes *Six4* (reported before (Kwon et al., 2010)), *Sall4* (Streit, 2002, Barembaum and Bronner-Fraser, 2007), *Cxcl14* (Chapter 3; Figure 3.6) and the epidermal gene *Keratin19*. While it could be promoting the expression of various PPR genes (above), it may be involved in repressing *Keratin19* consistent with the fact that these two do not co-localise. On the other hand, *Pax2*, an OEPD marker (Streit, 2002) is predicted to regulate *Foxi3* expression (Figure 5.3 D), has binding sites in Foxi3 E1.B (Chapter 4; Figure 4.12) and was confirmed to promote *Foxi3* by MO data (Dr. Jingchen Chen and Dr. Monica Tambalo). In contrast, both *Sall1* and *Hesx1* are repressors (Carvalho et al., 2003, Sweetman et al., 2005) and are initially expressed anteriorly but later expressed in the otic placode [(Sweetman et al., 2005), Chapter 3; Figure 3.6] thus these may be involved in preventing *Foxi3* expression in the otic placode. Additionally *Sall1* was also predicted to have binding sites in Foxi3 E1.B (Chapter 4; Figure 4.12) making it a plausible candidate to investigate further. Lastly, *Foxi3* and *Zfhx1b* may be mutually repressing each other consistent with the fact that these do not co-localise with *Zfhx1b* expressed in the neural tube (Dady et al., 2012). In summary, top choices for further investigation include interactions with *Sall1* and *Pax2* mainly because of their binding sites in Foxi3 E1.B and confirmation from Pax2 MO data (Dr. Jingchen Chen and Dr. Monica Tambalo).

Figure 5.3 E shows interactions for Pax2 which is an OEPD marker (Streit, 2002). *Pax2* is predicted to be regulated by PPR genes *Etv5, Gbx2* (reported before (Steventon et al., 2012)) and *Mynn*. The presence of positive regulatory loops with *Mynn* and *Gbx2* indicate

that the cells may be locked in a transcriptional state possibly promoting OEPD formation. As described earlier, *Pax2* was predicted to have binding sites in Foxi3 E1.B and this interaction was also confirmed from MO data (Dr. Jingchen Chen and Dr. Monica Tambalo). Moreover, *Pax2* was predicted to regulate *Hesx1*, a late otic gene (Chapter 3; Figure 3.6), and genes expressed in the PPR and otic placode including *Eya1* (Barembaum and Bronner-Fraser, 2007), *Homer2* (Tambalo, 2015), *Gata3* (Sheng and Stern, 1999), *Tfap2a* (Khudyakov and Bronner-Fraser, 2009) and *Foxg1* (Ahlgren et al., 2003). Thus it seems that *Pax2* is promoting otic fate. Some of these interactions have been confirmed by MO data (Dr. Jingchen Chen and Dr. Monica Tambalo) (Figure 5.3 E, blue) while others are already known such as *Pax2* to *Gata3* and *Foxg1* and *Gbx2* to *Pax2* (Christophorou et al., 2010, Freter et al., 2012, Steventon et al., 2012).

Foxg1 is expressed in the forebrain at HH7 and later starts to appear in the otic placode around HH11 stage (Khatri et al., 2014). It is a known repressor (Roth et al., 2010) and is predicted to regulate *Pax2*. *Pax2* is expressed in the OEPD and otic placode and later restricted to a sub-domain within the otic cup (Streit, 2002) hence it is possible that *Foxg1* is repressing *Pax2* at later otic stages. Lastly, *Pax2* may be repressing the non-neural genes *Gata2* (Sheng and Stern, 1999) and *Keratin19* consistent with the fact that they do not co-localize. In summary, *Pax2* may be promoting the otic fate by up-regulating otic genes; some of which have been confirmed by MO. An interesting interaction to explore further is with *Foxi3* where the enhancer of *Foxi3* also has binding sites for *Pax2*.

Gbx2 is expressed early in the PPR (Chapter 3, Figure 3.6). There seems to be mutual regulation between *Gbx2* and its fellow co-expressing genes *Etv5* (PPR), *Foxi3* (PPR) and *Pax2* (OEPD). Interestingly, *Gbx2* seems to promote its own repressor *Foxg1* in a negative feed-back loop. *Foxg1* becomes strong in the otic placode at HH11 (Khatri et al.,

2014) which coincides with the time when *Gbx2* expression starts to become faint in the otic placode (Chapter 3, Figure 3.6). Thus *Foxg1* may be responsible for removing its expression from the otic placode. *Hey1* (Leimeister et al., 2000) and *Fstl4* (Lleras-Forero et al., 2013) are expressed late in the otic placode and therefore it is feasible that these are regulated by *Gbx2* although at this point *Gbx2* expression begins to fade. Similarly, *Gbx2* was also predicted to regulate *Cxcl14* which is expressed in the PPR and OEPD (Chapter 3; Figure 3.6). Lastly, non-neural gene *Gata2* (Sheng and Stern, 1999) and the neural gene *Zfhx1b* (Dady et al., 2012) may be repressing *Gbx2* consistent with the fact that these do not co-localise. In summary, top choices for further investigation include interactions with the co-expressed genes *Etv5* and *Foxi3* as well as with *Pax2*.

Anterior genes that are down-regulated by FGF (Figure 5.3 G, orange) show a small circuit of interactions involving *Dlx5/6, Nfkb1, Pax6, Fstl4* and *Sstr5* (Bhattacharyya et al., 2004, Bailey et al., 2006, Lleras-Forero et al., 2013). PPR genes *Eya2, Six1* and *Six4* (Streit, 2002, Litsiou et al., 2005, Barembaum and Bronner-Fraser, 2007, Sato et al., 2010) are predicted to regulate *Fstl4* and *Pax6* possibly promoting these in the anterior PPR. *Pax6* is predicted to regulate *Dlx5* and *Dlx6* which is feasible as all three are co-expressed in the lens and olfactory precursor cells. Alternatively, *Dlx6* is predicted to regulate *Pax6* (reported as repressive (Bhattacharyya et al., 2004)). It has been shown that initially *Dlx5/6* and *Pax6* co-express but later play a role in segregation of lens and nasal precursors (Bhattacharyya et al., 2004). Once lens cells acquire columnar morphology, *Dlx* family members are downregulated whereas *Pax6* expression is lost from the olfactory cells. Moreover, ectopic expression of *Dlx5* in lens precursor cells results in repression of *Pax6* and hence of the lens fate. Therefore, it can be concluded that initially these promote each other towards an anterior fate but later play a role in segregating lens and olfactory fates by repressing each other. Lastly, *Dlx5/6* are predicted to regulate the

anterior genes *Nfkb1* and *Sstr5* as well as each other where feed-back loops possibly lock the transcriptional state thus promoting an anterior fate.

In conclusion, analysis of the interactions of FGF-response genes allows formulating new hypotheses which can then be tested.

5.6 Clustering of the predicted NanoString network confirms modularity and reveals sub-networks

The predicted network has a modular structure and the top 500 interactions indicate three large modules. However, there may be more than three modules in the network, which is difficult to identify when the network exists as a single connected component. In order to unveil the true modular structure of the network, community clustering (Newman and Girvan, 2004) was performed using GLay plugin (Su et al., 2010) in Cytoscape. As discussed before, the advantage to using Girvan and Newman's clustering algorithm is that it does not require the number of clusters to be fixed as in other clustering techniques such as K-means. Thus, it allows finding the natural community structure within the network. Clustering of the top 500 interactions of GENIE3 network reveals 5 clusters (Figure 5.4). For further analysis, each cluster was annotated to GO and KEGG terms using BiNGO (Maere et al., 2005). The components of each cluster were then coloured according to the mapped GO and KEGG terms, with genes that did not map to any terms coloured white. Moreover, to understand the relationship between members of a cluster, Pearson's correlation was calculated for all NanoString genes and plotted as a heatmap (Figure 5.5), which indicates positive (blue) and negative (red) correlation between genes.

Cluster 1 (Figure 5.4 A, purple) largely consists of the same members as the module C3 in Figure 5.2 containing anterior PPR and lens expressed genes including *Pax6, Sstr5,*

Nfkb1, Dlx5 and *Dlx6*. Among the significant GO and KEGG terms (P-value < 0.05) mapping to this cluster are some general terms such as transcription regulation, anatomical and, importantly, sensory organ development (Figure 5.4 A, grey, red, cyan and dark green). In particular, camera-type eye development and eye-photoreceptors are enriched in this cluster, largely because of *Pax6* and its interactions with some of the anterior genes *Otx2, Nfkb1, Dlx5/6* and *Lmx1b*. There are also terms related to histone modifications, because of the presence of DNA methyltransferase *Dnmt3b*. The only signalling pathway associated with this cluster is canonical WNT receptor signalling. This is interesting because activation of the WNT pathway leads to loss of the PPR markers *Six1, Eya2* and *Six4*, while WNT inhibition expands the placode territory (Litsiou et al., 2005). Keeping this in mind, analysing the correlation between members of this cluster reveals two main groups. One group, in which genes are positively correlated to each other (Figure 5.5, purple box) includes *Pax6, Dlx5/6, Sstr5, Nfkb1, Otx2, Lrp11, Ap2, Lmx1b, Setd2, Zic2, Sp4* and *Mier1*. The second group of genes correlates negatively with the first (Figure 5.5, purple box) and includes the WNT target *Cited2* (Schlange et al., 2000) along with *Cxxc6, Tbl1xr1, Ccnd1, Ptpru, Ece1, Pogz* and *Six1*. Based on this analysis, the regulatory interactions between both groups could be repressive and hence these are shown as negative interactions in pink (Figure 5.4 A). Additionally, interactions of the repressor *Otx2* (Bai et al., 2012) are also shown in pink (Figure 5.4 A).

Cluster 2 (Figure 5.4 B, pink) contains the same genes as the module C1 in Figure 5.2 and includes genes that respond positively to 6 hr FGF treatment (Chapter 3, section 3.3). These are *Foxi3, Etv5, Gbx2, Pax2, Hesx1* and *Eya2* with *Hesx1* being a central node (Figure 5.2). Clustering confirms that these FGF response genes form a small circuit where the genes regulate each other and possibly drive the posterior PPR towards OEPD and otic fate. General GO and KEGG terms mapping to many genes include regulation of

transcription and gene expression (Figure 5.4 B, pink, green). Importantly, inner ear morphogenesis, inner ear development and sensory perception of sound (Figure 5.4 B, light green, cyan, dark green) are enriched in this cluster and correspond to *Gbx2, Foxg1, Hesx1, Cdkn1b* and *Chd7*, while *Hesx1* is mapped to otic vesicle development (Figure 5.4 B, red). Correlation analysis reveals two transcript groups. Genes in the first group positively correlate with each other (Figure 5.5, pink box) and contain *Pax2, Six3, Hesx1, Foxg1*, *Gbx2, Hey1, Homer2, Sox2, Chd7, Etv5, Eya2, Hsf2, Foxi3 and Znf423*. The vast majority of these factors is expressed in the developing otic placode. The second group shows negative correlation to the first group and includes the WNT target *Lef1* along with *Zfhx1b, Gata2, Gata3* and *Tgif2*. As discussed above, WNT signalling must be repressed for PPR formation (Litsiou et al., 2005), and therefore the links between *Lef1* and the first group are shown as negative (pink) in Figure 5.4 B. Additionally, the links within each group are shown as activating and between the two different groups (shown in pink) as repressing as there is negative correlation between the two groups. *Foxg1, Hey1* and *Hesx1* are repressors (Carvalho et al., 2003, Yada et al., 2006, Roth et al., 2010) therefore their links are also shown as pink (Figure 5.4 B).

Cluster 3 (Figure 5.4 C) contains the same genes as C2 in Figure 5.2 and mostly consists of genes responding to FGF after 12 and 24 hrs (Chapter 3; section 3.3). General GO and KEGG terms mapping to many genes include regulation of gene expression (Figure 5.4 C, pink). Ear morphogenesis and inner ear morphogenesis (Figure 5.4 C, red, purple) correspond to *Six4, Zic1* and *Sall1* while histone modifications and histone acetylation (Figure 5.4 C, green, light green) correspond to *Ing5* and *Yeats4*. Neurogenesis and forebrain development (Figure 5.4 C, light purple, dark purple) correspond to *Bmp4, Sall1*, *Msx1* and *Sox3*. Correlation analysis reveals two transcript groups. Genes in the first group positively correlate with each other (Figure 5.5, blue box) and contain *Sox10,*

Zic1 and *Tox3*. All three are expressed late in the otic placode (Sun Rhodes and Merzdorf, 2006, Betancur et al., 2010b, Tambalo, 2015). The second group shows negative correlation to the first group and includes *Sall4, Trim24, Morc2, Zic3, Six4, Sall1*, *Znf217, Bmp4, Msx1, Irx3, Mef2d, Phf10, Yeats4, Hey2* and *Ing5*. Therefore, the links between the first and the second group are shown as negative (pink) in Figure 5.4 C. Within the second group, the links for repressors *Sall1* (Sweetman et al., 2005) and *Msx1* (Catron et al., 1995) are also shown in pink (Figure 5.4 C).

The fourth and the fifth cluster seem to correspond to genes that were mostly present around the periphery of the network with very few interactions with the central network (Figure 5.2). Mapping to GO and KEGG terms reveals enrichment of middle ear, inner ear and cochlear development in cluster 4 (Figure 5.4 D) mainly because of the presence of *Eya1* and *Tbx3*. Most other terms correspond to general processes such as cell adhesion, cellular processes, chromatin organization, aging and anatomical structure development. Careful analysis of the correlation shows that most cluster 4 genes correlate positively including *Dnajc1, Hdac1, Rybp* and *Geminin* (Figure 5.5, red box). Additionally, these are negatively correlated to *Tbx3, Eya1* and *Kremen1* (Figure 5.5, red box). Hence repressive links between these are shown in pink in Figure 5.4 D. Since *Tbx3* is a transcriptional repressor (Kunasegaran et al., 2014), its interactions are also represented in pink (Figure 5.4 D).

Cluster 5 (Figure 5.4 E) only consists of genes with positive correlation to each other (Figure 5.5, green) and an enrichment of GO and KEGG terms like regulation of cellular processes, post-replication repair and biosynthetic processes.

Overall, clustering confirms that there are three modules in the NanoString network, one corresponding to genes that are negatively regulated by FGF (Figure 5.2 C3; Figure 5.4 cluster 1), the second corresponding to genes that are positively regulated by FGF at the earlier time-point of 6 hrs (Figure 5.2 C1; Figure 5.4 cluster 2) and the third corresponding to genes that respond to FGF after 12 or 24 hrs (Figure 5.2 C2; Figure 5.4 cluster 3). This clearly indicates the usefulness of network modelling and correlation analysis in understanding the regulatory relationships among a set of genes through the identification of modules specific to different developmental processes.

5.7 A gene regulatory network for otic development

Now that the predicted network has been analyzed thoroughly, the information gathered so far will be organized into an otic developmental model incorporating i) known interactions which have been confirmed by GENIE3, ii) the top 25 interactions for selected FGF-response genes (Figure 5.3) and iii) predicted interactions for other FGF response genes from the top 500 predicted interactions in the whole network (Figure 5.2). The reason for not including interactions below the top 500 is to focus on the most important interactions defining the network structure and to keep the model simple. The model was built in BioTapestry (Longabaugh et al., 2009). Within the model (Figure 5.6) the genes are organized according to their spatial and temporal expression patterns. Additionally, where an enhancer was identified (Chapter 4; Figure 4.12-4.16) and its regulators predicted through both TFBS analysis and network inference, interactions are placed as solid lines (Figure 5.6).

The pPPR to otic placode region is subdivided into pPPR, OEPD and OP. In the pPPR sub-region, *Etv5* is a direct target of FGF signalling (Tambalo, 2015) and therefore it is placed at the top of the hierarchy being regulated by FGF signalling from the mesoderm.

Etv5 is predicted to activate *Mynn, Foxi3* and *Gbx2*; these transcripts are not directly regulated by FGF signalling and are therefore placed downstream of *Etv5*. *Foxi3* and *Gbx2* mutually regulate each other and together with *Mynn* activate *Cxcl14*. While *Etv5* is also predicted to regulate *Cxcl14*, it is likely that *Mynn, Gbx2* and *Foxi3* mediate this interaction downstream of *Etv5*. Hence *Cxcl14* is placed downstream of these three genes. As a chemokine *Cxcl14* does not have any targets.

Sall4 is another PPR gene that is regulated by *Foxi3* and *Sox10*; in turn it regulates *Sox10* in the otic placode. *Homer2* is only regulated by *Pax2* (also confirmed from our MO knockdown), however since it is expressed in the PPR prior to *Pax2*, it is placed in the appropriate position. It is possible that *Pax2* is involved in *Homer2* maintenance. The PPR gene *Six4* is regulated by *Foxi3* (hence placed below *Foxi3;* also reported previously (Khatri et al., 2014)) and *Sox10.* It is possible that *Sox10* is involved in its maintenance.

Pax2 is an OEPD gene which interacts with a number of PPR and otic placode genes. It is predicted to be activated by *Mynn* and *Gbx2*; the latter has previously been reported (Steventon et al., 2012). Although *Etv5* is predicted to regulate *Pax2*, this may be mediated by *Mynn* and *Gbx2*. *Pax2* in turn regulates the otic gene *Foxg1* (also confirmed from our MO knockdown and reported before (Freter et al., 2012)). In turn, *Foxg1,* a repressor regulates *Pax2* and *Gbx2* possibly removing their expression at late otic stages. Furthermore, *Pax2* is predicted to activate *Foxi3* (known from *Pax2* MO data and TFBS analysis), *Gbx2* (known from *Pax2* MO data) and *Gata3* (reported before (Christophorou et al., 2010).

As the otic placode forms, new transcripts become expressed; these include the transcriptional repressor *Sall1* (Sweetman et al., 2005), as well as *Sox10* and *Hesx1*. The

repressor *Sall1* is a target of FGF signalling, however no other regulators were found. It takes 12 hrs to be induced by FGF (Chapter 3; Figure 3.1 C) and hence it is unlikely to be a direct FGF target as by that time FGF signalling has been shut down in the otic placode (Freter et al., 2008). *Sall1* represses *Foxi3*, a finding also predicted by the TFBS analysis of the Foxi3 enhancer 1 (Chapter 4; Figure 4.12). Like *Foxi3* transcripts, enhancer activity is lost as *Sall1* appears in the otic placode. Together, these findings make *Sall1* a good candidate to ensure that *Foxi3* is depleted from the otic placode.

Lastly, *Hesx1,* a repressor regulates *Foxi3, Cxcl14* and *Etv5*. Both *Foxi3* and *Cxcl14* transcripts are lost from the otic placode and *Hesx1* is a good candidate together with *Sall1* to bring this about.

The model presented here is an attempt towards defining the hierarchy downstream of FGF signalling by incorporating some of the FGF response genes and their regulatory relationships identified from the top interactions in the NanoString network. It will be further improved by adding new genes and their regulatory relationships identified from the predicted mRNA-seq network discussed in the next sections.

5.8 GENIE3 recovers known otic interactions in the predicted mRNA-seq network

As discussed already, the NanoString probe set contains a limited set of genes. To investigate how cells, go from PPR to otic commitment and how cells that form other placodes are different from the otic lineage, mRNA-seq was performed from different tissues including aPPR, pPPR, OEPD, OP, future trigeminal, future lens and non-neural ectoderm (Dr. Jingchen Chen, Dr. Monica Tambalo and Ramya Ranganathan). At this point, it was therefore ideal to use the mRNA-seq data to get a picture of regulatory relationships between the newly identified otic genes (Dr. Jingchen Chen). For this

purpose, the mRNA-seq data were prepared and a GENIE3 network was produced as described in section 5.2. After applying the threshold of 0.001, the mRNA-seq network consisted of 534 genes and 98983 interactions. In the predicted mRNA-seq network (Figure 5.7 A) each node is coloured according to its out-degree to identify highly active TFs that are likely to regulate many other genes. In addition, the width of each edge is mapped to the importance measure of the regulatory link with thicker edges indicating higher significance. The network shows a number of highly active nodes (Figure 5.7 A, red and orange nodes). As the mRNA-seq network contains almost 4 times as many genes as the NanoString dataset, it offers a lot more information. At this stage, this is a preliminary attempt to study the general architecture of the predicted network and to place a few new otic genes and their predicted interactions into the model in Figure 5.6.

To assess the reliability of predictions the GENIE3 network was overlapped with known interactions from the literature (shown in Chapter 1; Table 1.1; Figure 1.2). The Cytoscape view of the known regulatory relationships is shown in Figure 5.7 B, in which interactions are coloured according to the region where they occur (see Table 1.1) with PPR interactions in green, posterior PPR interactions in purple, OEPD interactions in orange and otic and epibranchial interactions in pink. Out of 76 known interactions for 23 genes, 26 interactions were recovered (36% recovery; Figure 5.7 C); this is summarised in Table 5.2. Out of the 26 predicted and experimentally verified interactions, 10 are common with the predicted NanoString network, while the remaining 16 are not found in the NanoString network. Conversely, out of the 20 predicted and experimentally verified interactions in NanoString network 50% are also recovered in the mRNA-seq network.

In addition, MO knockdown data for *Pax2, Sox8* and *Etv4* were available from Dr. Jingchen Chen and Dr. Monica Tambalo. The predictions from the mRNA-seq network

were compared to the experimental outcome from these experiments to establish which regulatory relationships overlap (Figure 5.8). The GENIE3 mRNA-seq network is shown in Figure 5.8 A where node colours are mapped to the out-degree of a node thus identifying highly active TFs in the network as red and less active TFs in green. Figure 5.8 B illustrates the results of experimental reduction of *Pax2* (pink), *Sox8* (cyan) and *Etv4* (green). Positive relationships are shown as red edges whereas negative relationships are shown as black edges. The overlap of the predicted and experimental data (Figure 5.8 C) reveals that in total 27 out of 105 interactions (26% recovery) are common; these are summarised in Table 5.3. Using this overlap, a hierarchy will be established and assembled into a model later in the chapter.

5.9 Clustering segregates anterior and posterior genes

As discussed previously, the mRNA-seq samples used for GRN inference not only include tissues from various stages of otic development but also from anterior PPR, the future lens/olfactory region and the future trigeminal placode. It is thus expected that clustering will be able to segregate these different groups. To investigate this, Newman's community clustering was carried out on the mRNA-seq network which divided into two sub-networks (Figure 5.9); one containing genes that are normally expressed in a posterior domain (pPPR, OEPD and otic placode genes) and the other containing genes normally confined anteriorly (aPPR, trigeminal and lens genes). The nodes in the network are coloured according to their out-degrees with red being highly active TFs and green being less active TFs. Among the most active nodes are early PPR genes like *Six1, Etv4, Etv5, Znf385c* and *Sox13* (red-orange), while OEPD genes like *Pax2* and *Lmx1a* (orange-yellow) have intermediate out-degrees and otic placode genes like *Sox10, Soho-1* and *Rere* (yellow-green) have lower out-degrees. Together, the temporal sequence of gene expression and node out-degrees point to a possible epistatic hierarchy of these factors.

The interactions of a few selected genes from the otic cluster will be analyzed in detail in the next section in order to understand their regulatory relationships. At this point and for this project, the anterior cluster will not be discussed. However, in future, it can be analyzed in detail to understand regulatory relationships that lead to an anterior fate.

Table 5.2 Overlap between GENIE3 mRNA-seq network and known PPR to otic placode interactions

Source	Target	Interaction	IM	Region	Common with NanoString network
DLX5	EYA1	Activates	0.001303	pPPR	YES
DLX5	PAX-6	Represses	0.005156	PPR to OEPD	YES
DLX5	SIX1	Activates	0.002272	pPPR	-
DLX6	EYA2	Activates	0.001805	pPPR	YES
DLX6	PAX-6	Represses	0.002708	PPR to OEPD	YES
ETV4	SOX10	Activates	0.003276	Otic and epibranchial	-
DLX5	EYA2	Activates	0.001671	pPPR	-
EYA2	PAX-6	Activates	0.001255	Regionalization of PPR	-
FOXI2	DLX6	Activates	0.003044	pPPR	YES
FOXI2	DLX5	Activates	0.004219	pPPR	-
GBX2	PAX2	Activates	0.004132	PPR to OEPD	YES
IRX1	SIX1	Activates	0.001267	Regionalization of PPR	YES
JUN	ETV5	Activates	0.001184	pPPR	-
JUN	EYA1	Activates	0.008174	pPPR	-
JUN	FOXG1	Activates	0.002778	PPR to OEPD	-
JUN	PAX-6	Represses	0.001479	Regionalization of PPR	-
JUN	SIX1	Activates	0.001133	pPPR	-
JUN	SOHO-1	Represses	0.002811	Otic and epibranchial	-
OTX2	PAX-6	Activates	0.005238	Regionalization of PPR	YES
PAX2	FOXG1	Activates	0.003337	PPR to OEPD	YES
PAX2	SOHO-1	Activates	0.001676	Otic and epibranchial	-
PAX2	GATA3	Activates	0.002192	PPR to OEPD	YES
PAX2	SALL4	Activates	0.001154	PPR to OEPD	-
SIX1	FOXI2	Activates	0.002572	PPR to OEPD	-
SIX1	EYA2	Activates	0.001342	pPPR	-
SIX1	PAX2	Activates	0.001819	PPR to OEPD	-

Table 5.3 Overlap between GENIE3 mRNA-seq network and Pax2, Sox8 and Etv4 MO data

Source	Target	Interaction	IM
ETV4	PAX2	Activates	0.00131
ETV4	EYA2	Activates	0.002452
ETV4	GATA3	Activates	0.001728
ETV4	SPRY2	Activates	0.001877
ETV4	SIX1	Activates	0.006204
ETV4	GRHL2	Activates	0.001933
ETV4	ZNF385C	Activates	0.00385
PAX2	FOXG1	Activates	0.003337
PAX2	GATA3	Activates	0.002192
SOX8	HESX1	Activates	0.001355
SOX8	PAX2	Activates	0.001543
SOX8	GATA3	Activates	0.008605
ETV4	RXRG	Represses	0.01
ETV4	IRX5	Represses	0.00996
ETV4	IRX4	Represses	0.00445
ETV4	SOX10	Represses	0.003276
ETV4	HOXA2	Represses	0.00175
ETV4	LEF1	Represses	0.00144
PAX2	IRX5	Represses	0.011947
PAX2	LEF1	Represses	0.007029
PAX2	SOX10	Represses	0.005421
PAX2	HOXA2	Represses	0.004
PAX2	SNAI2	Represses	0.001316
PAX2	MEIS1	Represses	0.001731
SOX8	NR2F2	Represses	0.001182
SOX8	IRX5	Represses	0.005797
SOX8	SOHO-1	Represses	0.001535

5.10 Predicted regulatory interactions for otic genes in mRNA-seq network

The predicted interactions for some FGF- response genes such as *Foxi3, Gbx2, Hesx1*, and *Pax2* have been discussed in the predicted NanoString network. However, other FGF response genes such as *Spry1, Spry2* and *Sox13* were not present in the NanoString network. Most FGF-response genes analyzed in the previous chapters (Chapters 3 and 4), for which enhancers were predicted, are present in the mRNA-seq data. It is thus possible to investigate their regulatory relationships. After applying a threshold of 0.001, the predicted interactions of FGF-response genes and some otic genes were analyzed to identify relationships that may define otic development. As discussed earlier, this is a preliminary attempt to enrich the model in Figure 5.6.

Figure 5.10 shows the regulatory interactions for selected otic genes with blue edges highlighting interactions that overlap with MO data, edges marked with a black asterisk (*) showing known interactions (Chapter 1; Table 1.1; Figure 1.2), red edges showing those that are also predicted by TFBS analysis of the appropriate enhancer and a pink asterisk (*) indicating interactions in common with the predicted NanoString network. Three nodes are encircled in red throughout: *Lmx1a, Sox8* and *Sox13*. In the previous chapter, I showed that *Sox* family members and *Lmx1a* are among the top regulators of the FGF-response genes (Chapter 4; Figure 4.8 and 4.10) and this is consistent with the predicted network presented here. Thus, two independent approaches lead to the same prediction that *Sox* factors and *Lmx1a* may be crucial regulators of early otic genes and thus important for otic fate determination.

The GRN predicts *Foxi3* to be regulated by *Sox13, Six1* and *TCF7L2* (*Tcf4*) and binding sites for all three factors were found in *Foxi3* enhancers. Foxi3 E1.B contains a *Sox13* binding site, whereas *Six1* and *TCF7L2* (*TCF4*) binding sites were found in the second

enhancer Foxi3 E2.A (Chapter 4, Figure 4.12 and 4.13). *Tcf4* is a repressor and may contribute to the loss of *Foxi3* expression from the otic placode. *Six1* has previously been shown to positively regulate *Foxi3* (Christophorou et al., 2010).

A number of PPR genes interact with *Foxi3* including *Dlx5, Dlx6, Etv4, Etv5, Sall4*, *Mynn* and *Gbx2*. *Etv5* and *Foxi3* and *Dlx5* and *Foxi3* mutually regulate each other. *Spry2* is expressed slightly later in the OEPD, is predicted to activate *Foxi3* and may thus maintain its expression. It is also evident that *Foxi3* in turn regulates some OEPD genes like *Zbtb16* (Tambalo, 2015) and *Lmx1a* (Figure 5.11 B), as well as the expression of late otic genes including *Sox10* and *Soho-1*.

As evident from Figure 5.10 *Etv4* is a node with high out-degrees within the otic cluster and *Etv4* shows more putative interactions (Figure 5.10 B) than other otic genes (Figure 5.9). Since *Etv4* is one of the mediators of FGF signalling, many of its interactions should correspond to FGF regulated genes (see Chapter 3; section 3.3 for FGF targets). Among all 25 predicted interactions, 12 correspond to interactions seen in MO knockdown experiments (blue edges; Figure 5.10 B and Figure 5.8 C).

Although *Etv4* is a transcriptional activator (Lunn et al., 2007), a number of genes appear to be repressed by it suggesting that these interactions may be indirect. Among these are *Irx4/5, Sox10* and *Lef1* which are expressed later in the otic placode (Barembaum and Bronner-Fraser, 2007, Olivera-Martinez and Storey, 2007) and it is possible that these are initially repressed by *Etv4* to avoid premature expression. *Rxrg* (Cui et al., 2003) and *Hoxa2* (Paxton et al., 2010) are not co-expressed with *Etv4*, but instead present in neural tissues and therefore could be repressed by *Etv4*.

Like *Foxi3, Etv4* seems to interact with a number of PPR genes including *Gata3, Etv5, Foxi3, Gbx2, Eya2, Six1, Mynn, Znf385c* and *Sox13* (Sheng and Stern, 1999, Lunn et al., 2007, Barembaum and Bronner-Fraser, 2007, Sato et al., 2010, Sato et al., 2012, Khatri and Groves, 2013, Tambalo, 2015). It is interesting to note here that with many of these, *Etv4* seems to form positive feed-back loops (*Sox13, Znf385c, Mynn, Six1*) suggesting that these interactions may stabilise PPR identity even in the absence of continued FGF signalling. *Etv4* also forms a feedback loop with *Pax2,* an early OEPD marker (Streit, 2002), with the importance measure of the link from *Etv4* to *Pax2* being higher than vice versa. In addition, *Etv4* also regulates *Sox8* an OEPD gene (Tambalo, 2015) as well as the FGF targets *Spry1* and *Spry2* (Rogers et al., 2011), and the epibranchial gene *Nr2f2* (Tambalo, 2015) , although it cannot be predicted at this stage whether this is a repressive or activating interaction. In contrast, OEPD gene *Lmx1a* (Figure 5.11 B) is predicted to regulate *Etv4* as do the late otic gene *Soho-1* (Bok et al., 2011) and the repressors *Foxg1* (Roth et al., 2010), *Rere* (Tambalo, 2015) and *Zbtb16* (Wasim et al., 2010), which may be involved in removing *Etv4* expression at later stages.

Pax2 is among the earliest OEPD genes (Streit, 2002) and as such central to otic specification; its interactions are depicted in Figure 5.10 C. *Pax2* function has been studied previously (Barembaum and Bronner-Fraser, 2010, Christophorou et al., 2010, Freter et al., 2012) and several of its known interactions are recovered in the predicted network (Figure 5.10 C, black asterisk *), while others were verified by our own MO knockdown experiments (see Figure 5.8 B-C; Figure 5.10 C, blue edges) or predicted in the NanoString-based network (Figure 5.10 C, pink asterisk *).

Pax2 can function both as a repressor and an activator (Abraham et al., 2015). Network predictions are consistent with the evidence from MO knockdown experiments; both

indicate that *Pax2* represses later otic and neural genes like *Irx5, Sox10, Hoxa2, Meis1* (Barembaum and Bronner-Fraser, 2007, Paxton et al., 2010, Sanchez-Guardado et al., 2011) and *Lef1* (Olivera-Martinez and Storey, 2007) as well as the neural crest gene *Snai2* (Marin and Nieto, 2004). Thus, *Pax2* participates in preventing the specification of alternate fates and premature expression of late otic markers. This is consistent with *Pax2* expression, which at otic cup stages becomes restricted to a sub-domain (Baker and Bronner-Fraser, 2000, Streit, 2002). Among the transcripts positively regulated by *Pax2* are the PPR factors *Gata3* (Sheng and Stern, 1999), which has already been reported in the literature (Christophorou et al., 2010), the OEPD genes *Spry1, Spry2* and *Foxg1* with the latter being confirmed by MO knockdown, and the otic placode genes *Sall1, Soho-1* and *Rere*.

In contrast, *Pax2* is regulated by PPR genes including *Six1, Six4, Gbx2, Etv5, Etv4* and *Sox13* as well as the OEPD factor *Sox8* (consistent with our MO data (Figure 5.8 C)). *Foxg1,* a repressor (Roth et al., 2010) is expressed in the otic placode (Khatri et al., 2014) and predicted to regulate *Pax2.* While *Pax2* is expressed in the OEPD and otic placode, later it is restricted to a sub-domain within the otic cup (Baker and Bronner-Fraser, 2000, Streit, 2002) hence it is possible that *Foxg1* represses *Pax2* expression at the otic cup stages thus restricting it to a sub-domain. An interesting regulator of *Pax2* is the OEPD gene *Lmx1a*. As discussed earlier, *Lmx1a* motifs are among the top TFBSs enriched in putative enhancers regulated by FGF response, along with *Sox* family motifs. This fits well with potential regulation of *Pax2* by these factors.

Gbx2 is a PPR gene that acts both as a transcriptional activator and repressor (Steventon et al., 2012). *Gbx2* is predicted to interact with a number of PPR genes including *Etv4, Mynn, Foxi3, Six1, Sox13, Etv5* and *Dlx5* (Figure 5.8 D). Feed-back loops from PPR

genes *Mynn, Etv5, Dlx5* and *Sox13* indicate that these interactions may stabilise PPR identity even in the absence of continued FGF signalling. *Gbx2* is also predicted to regulate OEPD genes *Pax2* (also known from (Steventon et al., 2012)*), Spry2, Sox8* and *Zbtb16* (Streit, 2002, Rogers et al., 2011, Tambalo, 2015) and the otic gene *Hesx1* (Chapter 3; Figure 3.6) indicating its role in promoting the otic fate.

Conversely, *Gbx2* is predicted to be regulated by the PPR genes *Gata3* and *Znf385c* (Sheng and Stern, 1999, Tambalo, 2015) and the OEPD gene *Spry1* (Rogers et al., 2011) thus promoting the otic fate.

Hesx1 is a transcriptional repressor that is initially expressed in the anterior PPR but later expressed in the otic placode (Chapter 3; Figure 3.6). *Hesx1* is predicted to be regulated by a number of PPR genes (Figure 5.10 E) including *Mynn, Sall4, Foxi3, Dlx5, Etv5* and *Gbx2* (also harbours binding sites in *Hesx1* enhancer) and OEPD genes *Spry1* and *Sox8* (consistent with MO data; Figure 5.8 C). Conversely *Hesx1* represses *Sall4*, *Etv5, Foxi3* and *Dlx5* via feed-back loops (Figure 5.8 C) which is consistent with the expression pattern of these genes at later stages. At later stages, *Sall4* is removed from the otic cup whereas *Dlx5* is expressed in the olfactory region (Barembaum and Bronner-Fraser, 2007, Khudyakov and Bronner-Fraser, 2009). Similarly *Foxi3* is later expressed in the trigeminal and epibranchial placodes and *Etv5* becomes faint at the otic cup stages (Lunn et al., 2007, Blentic et al., 2008, Khatri and Groves, 2013). Thus *Hesx1* may be responsible for removing the expression of these genes from the otic region at later stages. It also represses PPR genes *Etv4* and *Gata3,* OEPD gene *Spry2* and otic genes *Sall1* and *Foxg1*. *Etv4* and *Spry2* become faint at later otic stages (Chambers and Mason, 2006, Lunn et al., 2007), whereas the repressors *Sall1* and *Foxg1* are removed at otic

vesicle stage (Ahlgren et al., 2003, Sweetman et al., 2005). Thus, the predicted regulation of these genes via *Hesx1* fits well with their expression pattern.

Sox8 is expressed in the OEPD (McKeown et al., 2005, Tambalo, 2015) and is a transcriptional activator (Schepers et al., 2000). Several of its interactions verified by our own MO knockdown experiments are recovered (see Figure 5.8 B-C; Figure 5.10 F, blue edges) or predicted in the NanoString-based network (Figure 5.10 F, pink asterisk *). Many PPR genes *Etv4, Gbx2, Dlx5, Sox13, Etv5* (Streit, 2002, Lunn et al., 2007, Paxton et al., 2010, Tambalo, 2015) and OEPD genes *Lmx1a* (Figure 5.11 B), *Spry1* and *Spry2* (Rogers et al., 2011) are predicted to regulate *Sox8* thus promoting the otic fate. Additionally, feed-back loops from *Sox8* to *Sox13, Lmx1a, Spry1* (also *Spry1* enhancer harbours a *Sox8* binding site; Figure 4.15 F; Figure 5.10 F, red) and *Spry2* indicates that such interactions may maintain OEPD stability even in the absence of FGF signalling. Conversely, *Sox8* is predicted to regulate its own repressors *Zbtb16* (Wasim et al., 2010) and *Foxg1* (Roth et al., 2010) via feed-back loops (Figure 5.10 F) which may be involved in removing *Sox8* at later stages. Several PPR genes including *Znf385c, Six1, Gata3* (Sheng and Stern, 1999, Barembaum and Bronner-Fraser, 2007), OEPD gene *Pax2* (Streit, 2002) and otic genes *Rere* (Kee et al., 2007) and *Hesx1* (Chapter 3; Figure 3.6) are predicted to be regulated by *Sox8* (Figure 5.10 F). Lastly, *Sox8* does not act as a repressor itself but is predicted to regulate the epibranchial gene *Nr2f2* (Tambalo, 2015) and late otic genes *Irx5* and *Soho-1* (Bok et al., 2011) [repressive interaction determined from MO knockdown]. Thus *Sox8* may activate an intermediate repressor such as *Foxg1* and *Zbtb16* (Figure 5.10 F) that could bring about the repression.

Sox13, a PPR gene (Tambalo, 2015) was among the top regulators of FGF-response genes along with *Sox8* and *Lmx1a* (Chapter 4; Figure 4.8; Figure 4.10; Figures 4.12; Figure

4.15-4.16). A number of PPR genes including *Etv4, Six1, Znf385c, Gbx2, Dlx5, Mynn, Etv5* and *Foxi3* (Streit, 2002, Lunn et al., 2007, Paxton et al., 2010, Sato et al., 2012, Khatri and Groves, 2013, Tambalo, 2015) are regulated by *Sox13* (Figure 5.10 G). The presence of feed-back loops from many of these genes to *Sox13* indicates stabilization of the PPR identity even in the absence of continued FGF signalling. Additionally, OEPD genes *Sox8 , Spry1, Pax2, Spry2* (Streit, 2002, McKeown et al., 2005, Rogers et al., 2011) and *Lmx1a* (Figure 5.11 B) also interact positively with *Sox13* via feed-back loops indicating stabilization of the OEPD identity. Thus it seems that *Sox13* promotes the otic fate through interacting with various PPR and OEPD genes. Conversely, repressors *Foxg1, Rere* and *Zbtb16* may be involved in removing *Sox13* expression at later stages (Figure 5.10 G).

Spry1 is an OEPD gene (Chapter 3; Figure 3.6); an otic enhancer was identified in the present study (Chapter 4; Figure 4.15). Several of its interactions were predicted in the TFBS analysis of its otic enhancer (Chapter 4; Figure 4.15 F; Figure 5.10 H, red). *Spry1* is predicted to interact with several PPR genes including *Etv5, Sox13, Etv4, Six1, Dlx5, Znf385c, Gbx2, Mynn, Gata3* (Sheng and Stern, 1999, Lunn et al., 2007, Paxton et al., 2010, Sato et al., 2012, Tambalo, 2015) and OEPD genes *Spry2, Sox8, Pax2* (Streit, 2002, McKeown et al., 2005, Rogers et al., 2011) and *Lmx1a* (Figure 5.11 B). The presence of feed-back loops indicates stabilization of the OEPD identity (Figure 5.10 H). *Spry1* can also act as a repressor (Chatterjee et al., 2012) and may be involved in repressing the epibranchial gene *Nr2f2* (Tambalo, 2015) in the otic placode (Figure 5.10 H).

Spry1 seems to promote its own repressors *Rere* and *Zbtb16* that are expressed in the otic placode (Tambalo, 2015) (Figure 5.10 H). *Spry1* expression is lost at the otic vesicle stage (Rogers et al., 2011) and therefore *Rere* and *Zbtb16* may be responsible for this.

Additionally, the repressors *Blimp-1* and *Otx2* are predicted to regulate *Spry1*. *Blimp-1* is expressed in the lateral part of OEPD and later in the epibranchial region (Tambalo, 2015) whereas *Otx2* is initially expressed in the anterior PPR and much later in the otic placode (Abe et al., 2006). It is thus possible that *Otx2* may be involved in removing *Spry1* expression at the otic vesicle stage whereas *Blimp-1* may repress *Spry1* in the epibranchial region.

Spry2 is an OEPD gene (Rogers et al., 2011). One of the predicted interactions is consistent with *Etv4* MO knockdown data (Figure 5.8 B-C; Figure 5.10 I, blue). Like *Spry1,* it is predicted to be regulated by many PPR genes including *Mynn, Six1, Etv4, Gbx2, Sox13* and *Dlx5* (Streit, 2002, Lunn et al., 2007, Paxton et al., 2010, Sato et al., 2012, Tambalo, 2015) and OEPD genes *Sox8, Lmx1a, Pax2* and *Spry1* (Streit, 2002, McKeown et al., 2005, Rogers et al., 2011, Tambalo, 2015) indicating that these interactions promote otic fate (Figure 5.10 I). The repressors *Rere, Zbtb16, Hesx1* and *Foxg1* are expressed in the otic placode and predicted to regulate *Spry2*. Like *Spry1, Spry2* expression is lost at the otic vesicle stage (Rogers et al., 2011) and hence the predicted interactions are consistent with *Spry2* expression pattern at later stages.

Conversely, *Spry2* is predicted to regulate several of the PPR and OEPD genes via feed-back loops including *Mynn, Sox8, Lmx1a, Six1, Etv4, Pax2, Gbx2, Sox13* and *Spry1* indicating that such interactions may be involved in stabilizing the OEPD identity (Figure 5.10 I). Lastly, *Spry2* is predicted to promote the PPR genes *Etv5, Foxi3, Six4, Znf385c* and *Gata3*.

In conclusion, this analysis helped in the identification of the key interactions of otic genes which will be used later in the chapter to enrich the model in Figure 5.6.

5.11 Lmx1a: A potential regulator of otic genes

As discussed earlier, *Lmx1a* emerges as a potential key regulator not only of FGF-response genes but also of other otic transcripts (Figure 5.10 A-I). TFBS analysis of FGF-responsive enhancers shows significant enrichment of an *Lmx1a* binding motif as compared to control, while network analysis predicts *Lmx1a* to regulate many OEPD and otic genes. To assess whether indeed *Lmx1a* function could lie at the top of the hierarchy during otic specification, the first step is to characterise its temporal expression pattern (Figure 5.11 A-C). *Lmx1a* is a transcriptional activator and at PPR stages (HH6/7), it is strongly expressed in the notochord and weakly at the edge of the neural plate, including the aPPR, but seems to be absent in the pPPR. As the OEPD is specified, *Lmx1a* transcripts become highly enriched in this region, and continue to be expressed at the edge of the folding neural plate and in the notochord (Figure 5.11 B). At HH10/11, there is strong expression in otic placode (Figure 5.11 C). Thus it seems a plausible candidate to regulate OEPD and otic genes. GENIE3 predicts *Lmx1a* to regulate *Etv4, Sox8, Sox13, Pax2, Spry1* and *Spry2* (Figure 5.10). To investigate its interactions in more detail, the top 100 *Lmx1a* interactions were retrieved from the predicted mRNA-seq network after applying the threshold of 0.001 (Figure 5.11 D). This analysis predicts that *Lmx1a* activates a number of PPR genes including *Six1, Sox13, Etv4, Etv5, Dlx5* and *Gata3* (Sheng and Stern, 1999, Streit, 2002, Lunn et al., 2007, Sato et al., 2012, Tambalo, 2015), as well as the OEPD genes *Sox8, Spry1, Spry2* and *Pax2* (Streit, 2002, McKeown et al., 2005, Rogers et al., 2011) and the placode genes *Sox10, Rere* and *Soho-1* (McKeown et al., 2005, Bok et al., 2011, Tambalo, 2015). TFBS analysis of the *Spry1* otic enhancer indeed reveals the presence of an *Lmx1a* binding site. Finally, *Lmx1a* also interacts with the epibranchial gene *Nr2f2* (Tambalo, 2015).

Interestingly, most of the *Lmx1a* targets are also its regulators including *Spry1, Sox8, Sox10* and *Six1*. This suggests, as found with many other interactions (see above, Figure 5.10) that positive feedback loops are characteristic features that may help to stabilise otic fate. *Foxg1* and *Zbtb16* are repressors (Roth et al., 2010, Wasim et al., 2010) and therefore down-regulate *Lmx1a* but do not show a feedback loop indicating that they receive different input and may be part of another regulatory circuit (Figure 5.11 D). Additionally, *Rere* is also a repressor and although it is promoted by *Lmx1a,* it represses *Lmx1a* (Figure 5.11 D).

Can predicted interactions be verified experimentally? To do this *Pax2,* as a key OEPD gene, was selected for verification as potential *Lmx1a* target. Since the *Spry1* otic enhancer has an *Lmx1a* binding site, this was also selected for verification and *Spry2* was selected because it ranked among the top 100 targets of *Lmx1a*.

5.11.1 Lmx1a morpholino leads to the reduction of *Pax2, Spry1* and *Spry2*

Two *Lmx1a* morpholinos (MOs) were designed as described in Chapter 2; section 2.6. The first MO was designed to target the exon3-intron3 boundary (LMX1A-E3) and the second MO was designed to target the exon4-intron4 boundary (LMX1A-E4). *Lmx1a* and control MOs (Chapter 2; section 2.6) were electroporated at HH4+ stage and embryos were analyzed at OEPD or placode stage. The targeted tissue was dissected and analyzed by RT-PCR to test efficient deletion of exon 3 and 4, respectively. As electroporation leads to a mosaic expression with wild-type cells mixed with cells carrying the transgene, the PCR will detect the presence of both wild-type and exon-deleted transcripts. The wild-type transcript of *Lmx1a* is 750 bp, while the size of LMX1A-E3 with exon 3 deleted is 500 bp and the size of LMX1A-E4 with exon 4 deleted is 580 bp. After electroporation with LMX1A-E4 MO, a 750 bp wild-type transcript as well as a 580 bp

fragment is observed indicating exon 4 deletion (Figure 5.11 E, Lane 3). In contrast, electroporation of control MOs show the presence of wild-type transcript only (Figure 5.11 E, Lane 7). Primers for the housekeeping gene *Gapdh* amplify a strong band of expected size (Figure 5.11 E, Lane 2, 125 bp). Unfortunately, there was insufficient cDNA to assess the efficiency of LMX1A-E3 MO since *Gapdh* amplification shows a very faint band only in both control and experimental MO electroporated tissue (Figure 5.11 E, Lanes 5, 6). In agreement with this, *Lmx1a* amplification reveals a weak wild type band of 750 bp; since there is generally more wild type transcript than exon-deleted transcript (see Lane 3), a 580 bp band would be below the detection limit. Therefore, only LMX1A-E4 MO was used for further experiments.

To validate interactions from *Lmx1a* to *Pax2, Spry1* and *Spry2*, embryos were electroporated at stage HH4+ and cultured overnight until they reached at least OEPD stages. Expression of *Pax2, Spry1* and *Spry2* was then assessed by *in situ* hybridisation, followed by immunohistochemistry using an anti-fluorescein antibody to visualize MO carrying cells (Figure 5.11 F-K). As predicted *Lmx1a* knockdown leads to loss of *Pax2* (n=5/6), *Spry2* (n=3/4) and *Spry1* (n=5/7) as compared to the control side of the same embryo or electroporation of control MOs (Figure 5.11: *Pax2* (F, I), *Spry2* (G, J), *Spry1* (H, K)). Sections confirm this phenotype (Figure 5.11 *Pax2* (i'), *Spry2* (j'), *Spry1* (k') showing reduction in both gene expression and placode thickening. While *Pax2* is almost completely lost, *Spry1* and *Spry2* expression is affected to a slightly lesser degree, but in both cases the placode has lost its characteristic thickened morphology (Figure 5.11 j', k'). This confirms that *Lmx1a* plays a fundamental role in regulating *Spry1, Spry2* and *Pax2*. Together with the network analysis, *Lmx1a* can be placed at the top of the otic hierarchy upstream of *Spry1, Spry2* and *Pax2*. In future, it will be interesting to see if loss

of *Lmx1a* has any effect on the expression of late otic genes *Soho-1* and *Rere* (Figure 5.11 D).

Although this is a very small number of tested interactions, it provides good support for the network approach and confidence that GENIE3 can predict reliable regulatory relationships. This approach therefore generates an excellent resource to explore more regulatory relationships and a wealth of information for future experiments.

5.12 An improved gene regulatory network for otic development

New otic genes were identified in mRNA-seq and their interactions have been predicted in the network. To improve the initial model in Figure 5.6, new genes and their regulatory relationships were added using interactions identified through network prediction (Figure 5.10). Additionally, where an enhancer was identified (Chapter 4; Figure 4.12-4.16) and its regulators predicted through both TFBS analysis and network inference, interactions are placed as solid lines (Figure 5.12). As discussed previously, *Etv4* and *Etv5* are at the top of the hierarchy being direct targets of FGF signalling via *Ap1*. The interactions in Figure 5.10 were carefully analyzed to define hierarchy downstream of *Etv4* and *Etv5*.

Etv4 and *Etv5* are predicted to activate a number of PPR genes including *Mynn, Foxi3, Six1* (also verified by *Etv4* MO KD), *Dlx5/6, Sall4* (also regulated by OEPD gene *Pax2;* reported before (Barembaum and Bronner-Fraser, 2007)), *Sox13, Gbx2, Gata3* and *Znf385c* (latter two also verified by *Etv4* MO KD). Downstream of *Etv4* and *Etv5, Foxi3* is regulated by *Sox13, Gbx2, Sall4* and *Six1* (reported before (Christophorou, 2008, Christophorou et al., 2009); also harbours binding sites in Foxi3 E2.A). On the other hand, *Foxi3* is repressed by the late otic genes *Hesx1* and *Sall1* (harbours binding site in Foxi3 E1.B; also predicted in NanoString network) which is consistent with the loss of

Foxi3 transcripts at the otic placode stage). *Foxi3,* in turn regulates other PPR genes *Mynn,* the chemokine *Cxcl14, Sall4, Six4* and *Dlx5/6* (reported before (Hans et al., 2004, Hans et al., 2007)). The chemokine *Cxcl14,* whose enhancer was identified in the present study (Chapter 4; Figure 4.16), is predicted to be regulated by *Sall1* (NanoString network; harbours binding site in *Cxcl14* enhancer). Like *Foxi3, Cxcl14* is removed from the otic placode at later stages and expresses in the ectoderm surrounding the otic placode. Thus, *Sall1* may be involved in removing its expression.

Etv4 activates the OEPD gene *Pax2* (verified by *Etv4* MO KD), however the OEPD genes *Sox8* and *Lmx1a* also activate *Pax2* (verified by *Sox8* (Figure 5.8) and *Lmx1a* MO KD (Figure 5.11)), therefore it is possible that *Etv4* may mediate this activation through *Sox8* and *Lmx1a* (where interaction from *Etv4* to *Sox8* is predicted from the network and *Sox8* and *Lmx1a* are predicted to mutually activate each other; Figure 5.10 F). Hence, *Pax2* is placed below *Sox8* and *Lmx1a* in the OEPD region. *Lmx1a* is faint at the start of the OEPD stage but becomes stronger at HH8 (Figure 5.11 B). The repressor *Zbtb16* which is activated by *Foxi3* is predicted to regulate *Lmx1a* and may repress it at early OEPD stages to avoid its premature expression thus *Zbtb16* may be above *Lmx1a* in the hierarchy and could be repressed by other factors (not identified) to allow *Lmx1a* expression in the OEPD. In the present study, *Lmx1a* was identified to be upstream of both *Spry1* and *Spry2* (predicted in network; harbours binding sites in *Spry1* enhancer; verified by *Lmx1a* MO KD). *Sox8* can also be placed upstream of *Spry1* and *Spry2* (predicted in network; harbours binding site in *Spry1* enhancer). Both *Spry1* and *Spry2* interact with *Pax2* and are repressed by *Zbtb16* and the late otic gene *Rere* which are possibly involved in removing *Spry1* and *Spry2* expression at the otic vesicle stage (Rogers et al., 2011).

At the otic placode stage, *Sox8* is predicted to activate *Hesx1* (verified by *Sox8* MO KD), *Rere* and *Sox10* (harbours binding site in *Sox10* enhancer; reported in (Betancur et al., 2011)) and repress *Soho-1* (verified by *Sox8* MO KD). The repressor *Foxg1* is activated by *Pax2* whereas the repressor *Hesx1* is activated by *Pax2, Spry1* and the PPR gene *Mynn* besides *Sox8.* It is interesting to find that *Hesx1* is predicted to repress the repressors *Sall1* and *Foxg1* both of which have been predicted to repress PPR genes. This indicates that *Hesx1* may be involved in preventing untimely repression of PPR genes by repressing *Sall1* and *Foxg1*. Lastly, *Lef1,* a WNT target is activated by WNT signalling from the neural tube which is required at later stages to promote otic fate (Ohyama et al., 2006). However, before the right time, *Pax2* may be involved in repressing *Lef1* (verified by *Pax2* MO KD).

Thus, many previously known interactions are recapitulated by the network; some of the indirect interactions are elucidated by using network predictions and MO KD data. A number of new interactions have been predicted that help in explaining the temporal and spatial expression of otic genes. Hence, this network can be used as an information resource for planning future experiments.

5.13 Discussion

In this chapter, I discussed the inference of GRNs using two different expression datasets: NanoString and mRNA-seq and presented a model at the end incorporating a few otic genes and their regulatory relationships. Over the last few years, computational identification of interactions between genes and TFs through inference techniques has become very popular (Gardner and Faith, 2005, Margolin et al., 2006, Bansal M, 2007, Markowetz and Spang, 2007, Huynh-Thu et al., 2010). In this study, GENIE3 was used to infer networks from NanoString and mRNA-seq datasets. GENIE3 was a good choice as

it outperformed other popular GRN inference programs [CLR (Faith et al., 2007), ARACNE (Margolin et al., 2006), MRNET (Meyer et al., 2007) and GGMs (Schafer and Strimmer, 2005b)] in the DREAM4 (Dialogue for Reverse Engineering Assessments and Methods) challenge (http://dreamchallenges.org/2010-publications/, (Stolovitzky et al., 2007) and it was able to recover 40% of the known interactions. It is easier to identify true positives (sensitivity) among a set of predicted interactions as compared to true negatives (specificity) as data specifying that a regulator affects the expression of a target gene is more readily available in biological literature as compared to data specifying that a regulator does not affect the expression of a target gene. For this reason, only sensitivity was estimated for the predicted networks in present study. It is, however, important to consider a balance between sensitivity and specificity. This is due to the fact that in an inferred network, sensitivity tends to increase with an increase in the number of overall predicted interactions. This is accompanied by an increase in the number of false positives which indirectly reduces specificity. ROC (Receiver Operating Characteristic) curves are an efficient means to estimate the predictive power of an algorithm and can be used to determine an optimum threshold where sensitivity is high without overly compromising the specificity. In future, specificity in the predicted networks can be estimated to some extent using two approaches. The first is by estimating the false positive rate. This can be achieved by assuming that all predicted interactions other than true positives are false. Then at varying IM (weight of interactions) thresholds, the percentage of such interactions among total predictions can be plotted against the percentage of true positives retrieved. Alternatively, at varying IM thresholds, the total number of predicted interactions can be directly plotted against percentage of true positives retrieved. Such plots can therefore give an indication of the trade-off between sensitivity and specificity.

Graphic GRN diagrams were produced for each predicted network using Cytoscape to gain insight into mechanisms of differential gene expression at a systems level. The networks were analyzed at two levels in this study. At the level of overall network architecture where genes with the highest out-degree and centrality were identified (Figure 5.1, 5.2, 5.7, 5.8 and 5.9). Secondly, at the level of a single gene, the regulatory relationships were analyzed in detail to assess a gene's place in the hierarchy of otic development (Figure 5.3 and 5.10).

In general, a network contains central regions called "hubs" or "modules" that are highly connected and give an indication of shared functionality between the genes within the modules (Luscombe et al., 2004, Deplancke et al., 2006, Yu and Gerstein, 2006). This was also observed in the present study (Figure 5.2 C1, C2, C3 and Figure 5.9 posterior and anterior). Thus it can be hypothesized that different modules may represent different functional units whereby compartmentalization into modules insulates expression changes within the module from the rest of the network. Indeed, the temporal and spatial expression of the components in each module suggests different biological functions for each module, while highlighting similarities within each module. Thus, each module may represent different developmental processes leading to different fates [NanoString network (Figure 5.2, C1, C2 and C3)]. Clustering of the network confirmed modularity (Figure 5.4). Additionally, GO term analysis identified different terms associated with different clusters that matched the gene expression in each cluster such as lens specification for cluster 1 and inner ear development for cluster 2 (Figure 5.4 A, B).

Upon detailed analysis of individual genes' interactors, an interesting observation was the presence of feed-back loops (Figure 5.3; Figure 5.10). These have been reported in gene regulatory networks as circuits that may be involved in stable co-expression of both genes

(Milo et al., 2002, Shen-Orr et al., 2002) and may lock the cells in a particular transcriptional state in the absence of continued signalling (Davidson et al., 2002, Levine and Davidson, 2005, Oliveri et al., 2008). Thus it can be concluded that the presence of feed-back loops between the various PPR genes may lock cells in a state where the PPR identity is stabilised until some trigger makes the cells go to the next stage (OEPD) where the same is repeated until another trigger makes the cells go to the next stage (OP). This may be associated to the varying levels of FGF and WNT at different stages of otic development.

GRN analysis also provides insight into the hierarchical organization of the regulatory interactions. It is clear from the mRNA-seq network (Figure 5.9) that there are certain nodes with very high out-degrees. Among these are genes (example: *Etv4* and *Etv5*) that are expressed early at the PPR stages and analysis of their predicted interactions shows that indeed these are responsible for regulating many OEPD and otic genes (Figure 5.10). The OEPD genes *Sox8* and *Lmx1a* emerge as intermediate factors (Figure 5.9; yellow-orange node) that are regulated by *Etv4* and *Sox13* and in turn regulate other OEPD and late otic genes. Thus, *Sox8* and *Lmx1a* may be the link between PPR and otic genes. This shows that analyzing out-degrees of nodes in a network gives an indication about hierarchy, whereby genes with the highest out-degrees are at the top, genes with intermediate out-degrees in the middle and genes with very low out-degrees at the bottom. This also explains why the nodes with the highest out-degrees have feed-back loops amongst them in order to maintain stable co-expression and thus co-regulate the downstream genes.

Finally, the combination of experimental and predictive approaches offers a major step forward to determine epistatic relationships and the hierarchy of interactions. Previously

established regulatory relationships are mostly indirect, as they have been identified mainly through knockout or knockdown strategies of otic genes followed by assessment of the effect on downstream targets without knowledge of enhancer regions or considering timing and sequence of gene expression (Chapter 1; Table 1.1; Figure 1.2). Since the predicted network is directional, it was possible to confirm some known interactions (Figure 5.1, 5.7 and 5.8) and determine a hierarchical relationship (Figure 5.12) together with enhancer analysis and identification of TFBSs. For example, *Lmx1a* binding sites were found in the otic enhancer of *Spry1* and knockdown of *Lmx1a* leads indeed to reduction of *Spry1* expression. *Lmx1a* was also found to be one of the key regulators of other FGF response genes (Chapter 4 Figures 4.8 and 4.10) and Figure 5.10. From these, *Pax2* and *Spry2* were also confirmed to be targets of *Lmx1a*. This is a very small number of interactions that have been tested but it indicates that GRN inference from expression data is useful in indicating the flow of information during the developmental process and may help to elucidate regulatory relationships which can then be confirmed with experiments.

At this point, only a preliminary analysis of the mRNA-seq network has been carried out with a main focus on understanding regulatory relationships between a few selected FGF response and otic genes. In future, to fully understand the relationships predicted in the mRNA-seq network between various genes, correlation heatmaps can be generated to determine negative and positive relationships between genes. Additionally, the posterior cluster identified in Figure 5.9 can be further clustered to separate pPPR, OEPD and otic genes and analyze their regulatory relationships in detail. To conclude, the predicted GRNs have allowed elucidating the known regulatory relationships as well as predict new ones through various analyses which are finally presented in the model in Figure 5.12.

Now, this model can be used to form plausible hypothesis and to test some of the interactions in the otic network.

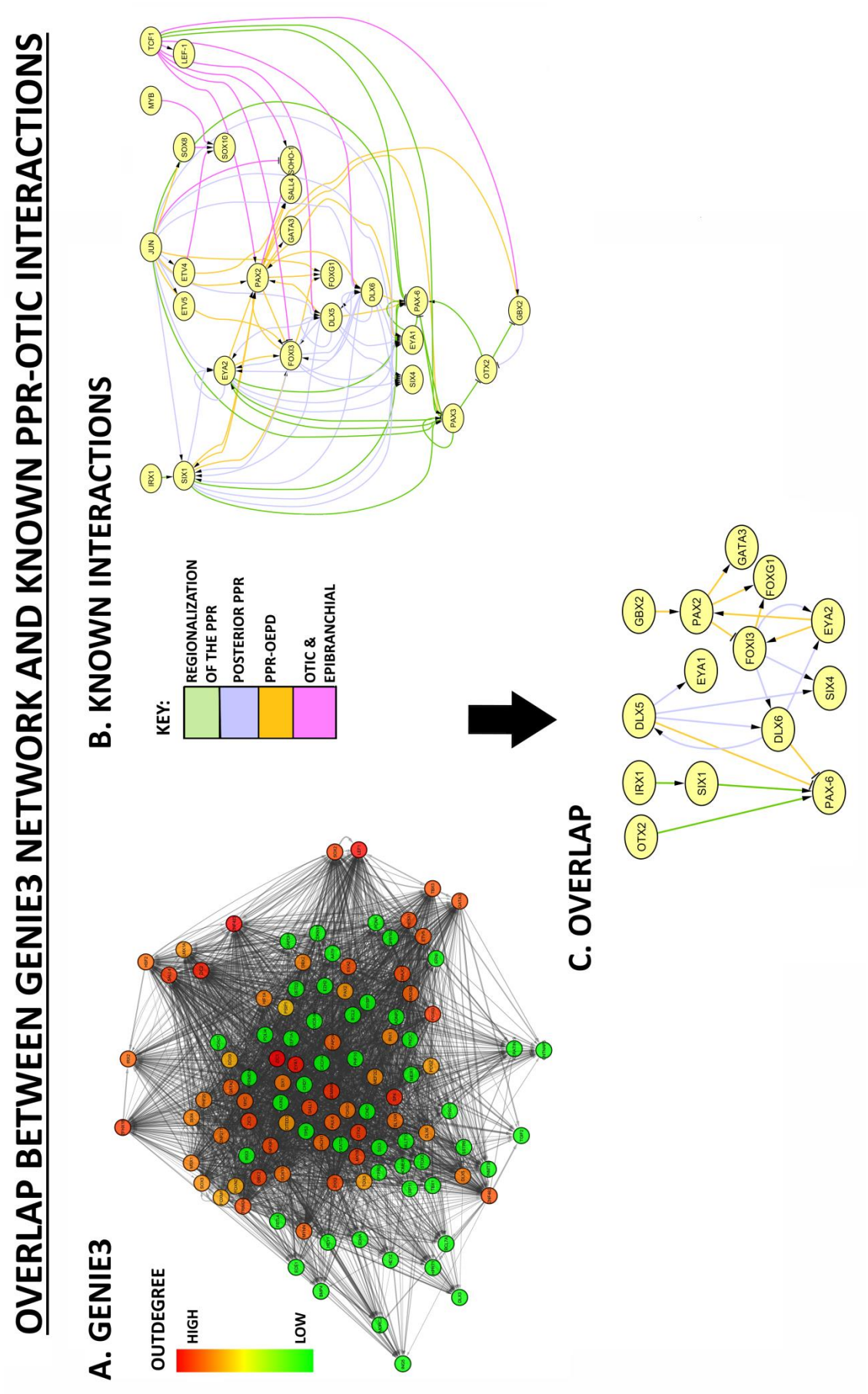


Figure 5.1 Overlap between GENIE3 NanoString network and known PPR to otic placode interactions

Figure 5.1 Overlap between GENIE3 NanoString network and known PPR to otic placode interactions
(A) Cytoscape view of the directed GENIE3 NanoString network (IM>=0.005). Nodes are coloured according to their out-degrees; Nodes with higher out-degrees in red and nodes with low out-degrees in green. **(B)** Cytoscape view of the known interactions from the literature; interactions during regionalization of PPR in green; posterior PPR in purple; PPR-OEPD in orange; otic and epibranchial in pink. **(C)** Overlap between the predicted GENIE3 network and known interactions recovers 40% of the interactions.

Figure 5.2 Top 500 GENIE3 NanoString interactions reveal a modular structure
Top 500 predicted interactions in the NanoString data reveal three modules highlighted in pink, blue and purple (C1, C2 and C3). The nodes are coloured and sized according to their centrality in the network and the edges are weighted according to the importance measure (IM). C1 consists of genes that respond positively to FGF and are involved in the OEPD induction (nodes encircled in pink). *Hesx1* seems to be the most central gene in this module. C2; highlighted in blue consists mainly of genes that respond to FGF after 12 or 24 hrs (nodes encircled in blue). The most central gene seems to be *Lef-1* which is down regulated by FGF after 12 hrs. C3; highlighted in purple consists of genes that respond negatively to FGF and are mainly anterior or lens genes (nodes encircled in purple). *Pax6* which is a lens marker is the most central gene in this module.

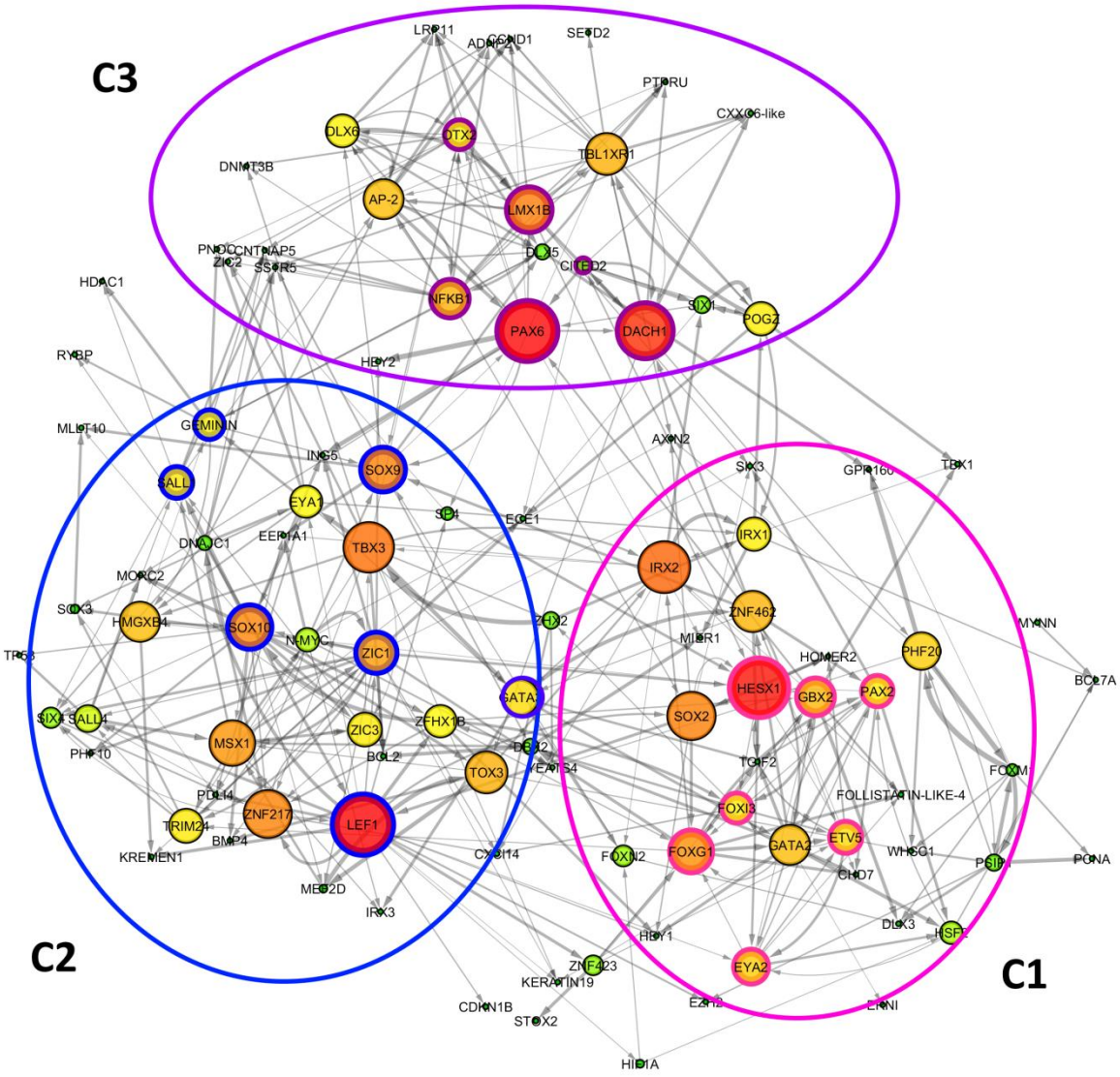


Figure 5.2 Top 500 GENIE3 NanoString interactions reveal a modular structure

TOP 25 INTERACTIONS FOR SELECTED GENES

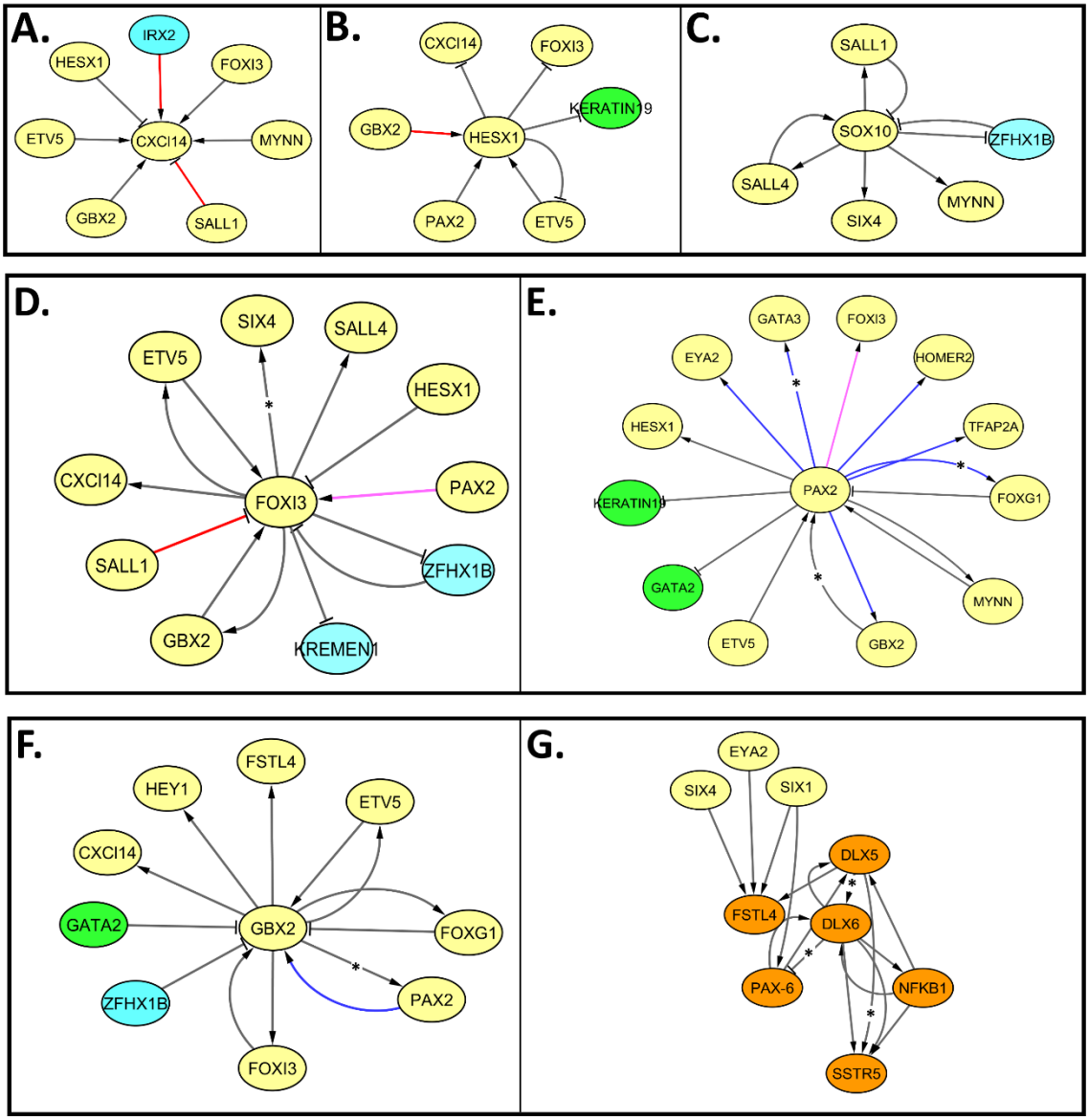


Figure 5.3 Top 25 GENIE3 NanoString interactions for selected FGF-response genes From the GENIE3 NanoString network, top 25 interactions for selected FGF response gene were plotted and coloured in Cytoscape. Neural genes are shown in cyan and non-neural genes in green. Interactions in red have been predicted both in GENIE3 and TFBS analysis of the enhancer of the respective gene. Interactions marked with an asterisk (*) are known regulatory interactions found by GENIE3. Interactions in blue are confirmed by MO data (Dr. Jingchen Chen and Dr. Monica Tambalo). Interactions in pink are shown by GENIE3, TFBS analysis and MO data. **(A)** Interactions for Cxcl14 **(B)** Interactions for Hesx1 **(C)** Interactions for Sox10 **(D)** Interactions for Foxi3 **(E)** Interactions for Pax2 **(F)** Interactions for Gbx2 **(G)** Interactions for anterior genes (brown).

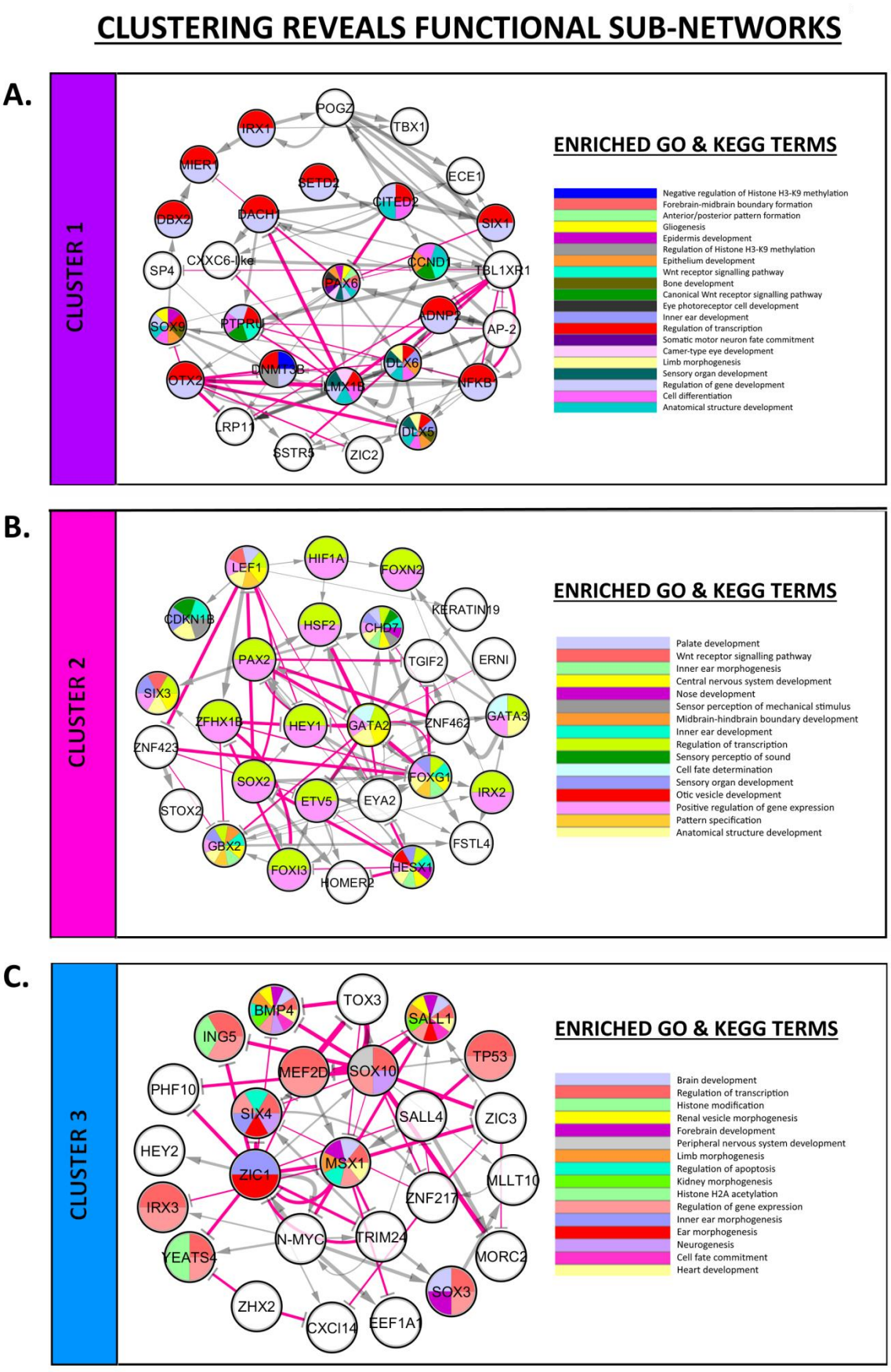


Figure 5.4 Clustering of the top 500 GENIE3 NanoString interactions reveals sub-networks

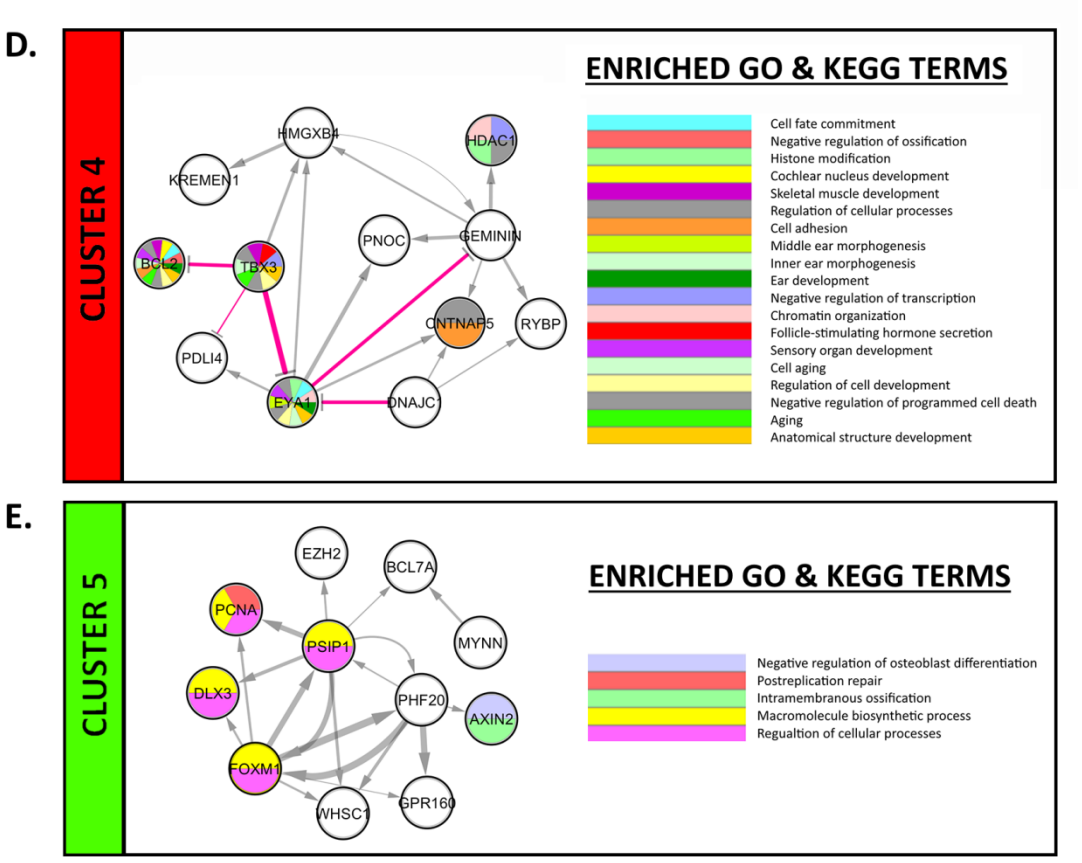


Figure 5.4 Clustering of the top 500 GENIE3 NanoString interactions reveals sub-networks

Clustering of the top 500 interactions in the predicted NanoString network using community clustering (GLay Plugin in Cytoscape) confirms modularity in the network and reveals 5 sub-networks. Each cluster was mapped to enriched GO and KEGG terms (P-value <0.05) and nodes coloured accordingly. Genes that do not map to any terms are coloured white. Repressive interactions are shown in pink and were determined from the correlation values between the NanoString genes (Figure 5.5). Edges are weighted according to the IM values **(A)** Cluster 1 includes anterior and lens-fate genes that respond negatively to FGF. Some of the corresponding GO terms include eye development and anterior/posterior pattern formation. **(B)** Cluster 2 corresponds to OEPD and otic genes that respond positively to FGF. Some of the corresponding terms include inner ear development and sensory perception of sound. **(C)** Cluster 3 corresponds to genes that respond to FGF after 12 or 24 hrs. The corresponding terms include forebrain and nervous system development. **(D)** Cluster 4 contains genes that were generally found around the periphery of the network (Figure 5.2) and correspond to general GO and KEGG terms such as cell adhesion and cellular processes**. (E)** Cluster 5 corresponds to general GO terms such as regulation of cellular processes and like Cluster 4 contains genes that are present at the periphery of the network (Figure 5.2).

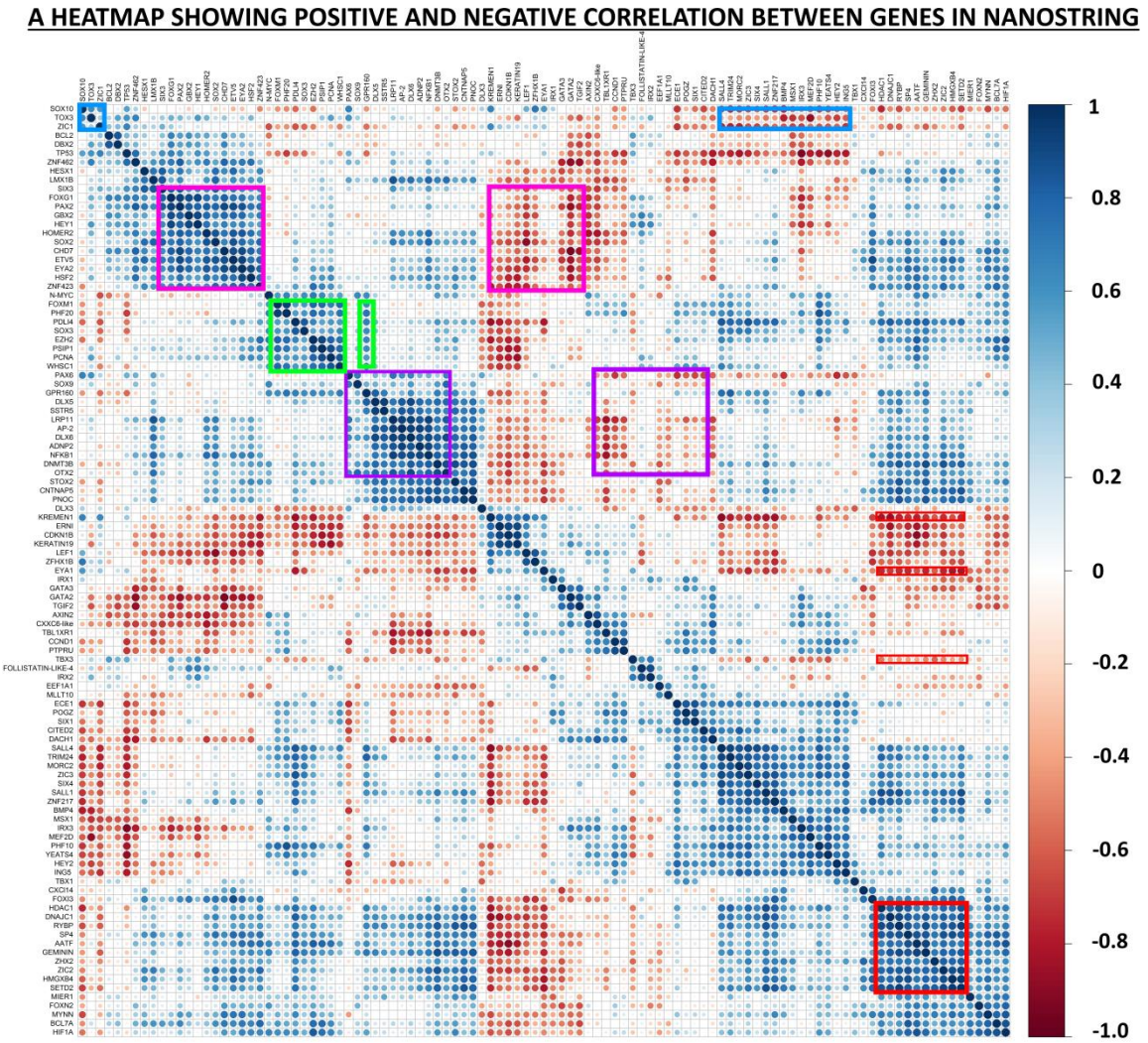


Figure 5.5 Correlation heatmap for NanoString data

Pearson's correlation was calculated among all the genes in NanoString data and plotted as a heatmap. All NanoString genes are shown along the X-axis and the Y-axis. Negative correlation is shown in red and positive correlation is shown in blue. The diagonal in blue shows correlation of a gene with itself. Highlighted regions of the heatmap show Clusters 1-5 (Figure 5.4); Cluster 1 in purple showing two groups of genes (positively and negatively correlated); Cluster 2 in pink showing two groups of genes (positively and negatively correlated); Cluster 3 in blue showing two groups of genes (positively and negatively correlated); Cluster 4 in red showing a group of positively correlated genes and three negatively correlating genes; Cluster 5 in green showing a single positively correlated group of genes.

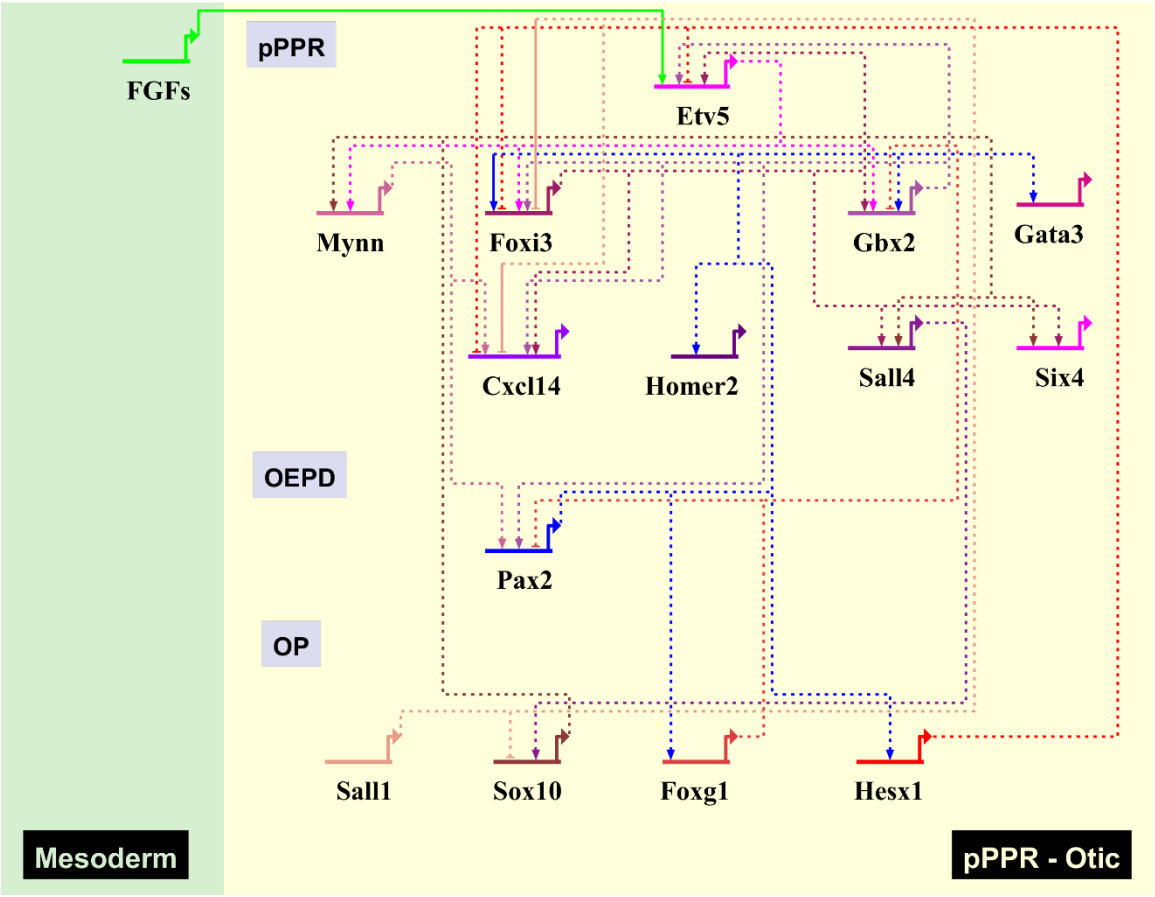


Figure 5.6 A GRN for otic development from predicted NanoString network
A BioTapestry model for otic development: The pPPR-Otic region is subdivided into pPPR, OEPD and OP depending upon the expression of genes. FGF from the mesoderm directly up regulates *Etv5* which is organized at the top of the hierarchy in the pPPR. *Etv5* regulates other PPR genes (shades of pink) which in turn regulate the OEPD gene *Pax2* (blue). Together with PPR genes, *Pax2* regulates otic genes (shades of brown). Solid lines indicate binding of the TF to the enhancer of its target gene. See text for details.

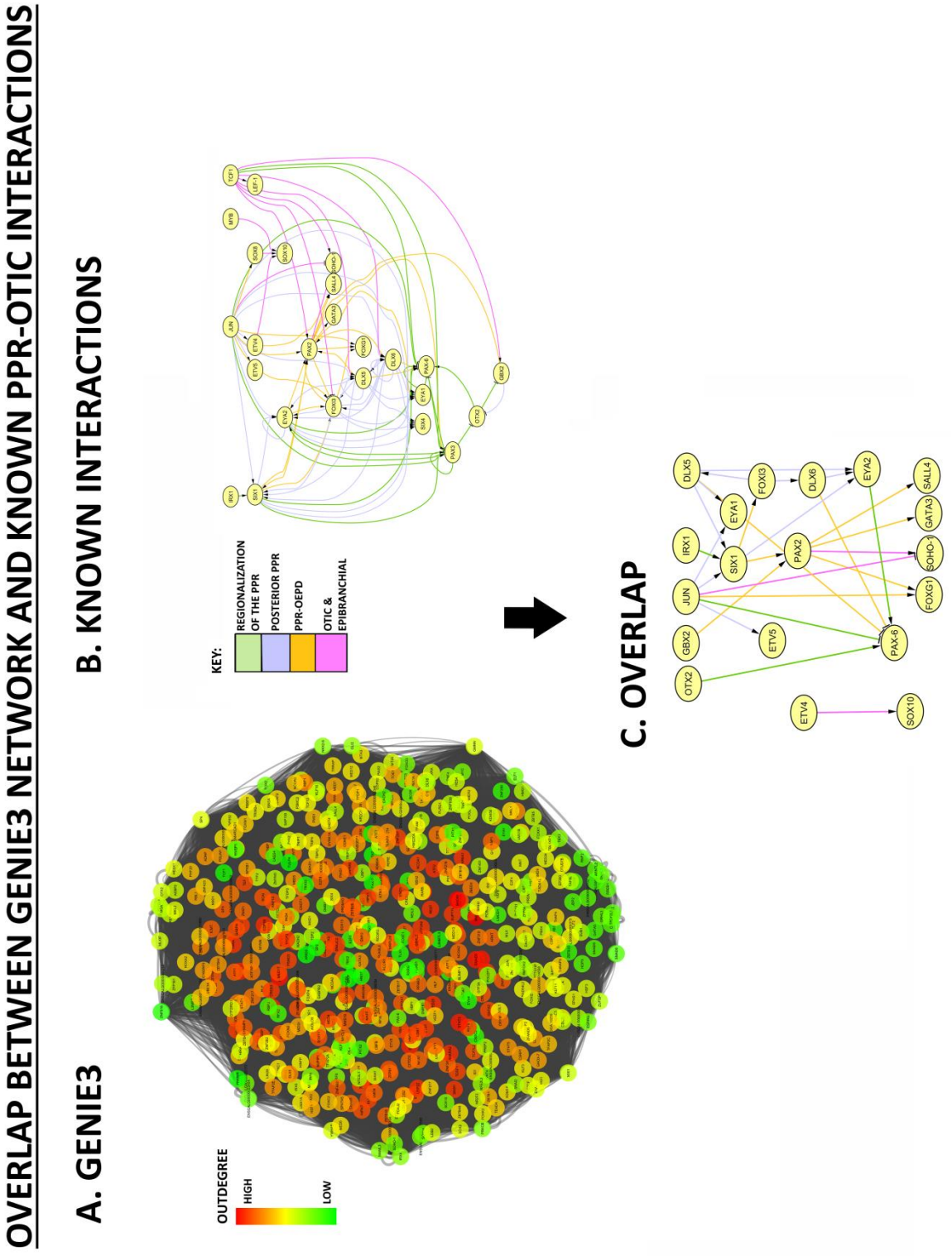


Figure 5.7 Overlap between GENIE3 mRNA-seq network and known PPR to otic interactions

(A) Cytoscape view of the directed GENIE3 mRNA-seq network (IM>=0.001). Nodes are coloured according to their out-degrees; Nodes with higher out-degrees in red and nodes with low out-degrees in green. **(B)** Cytoscape view of the known interactions from the literature; interactions during regionalization of PPR in green; posterior PPR in purple; PPR-OEPD in orange; otic and epibranchial in pink. **(C)** Overlap between the predicted GENIE3 mRNA-seq network and known interactions recovers 36% of the interactions.

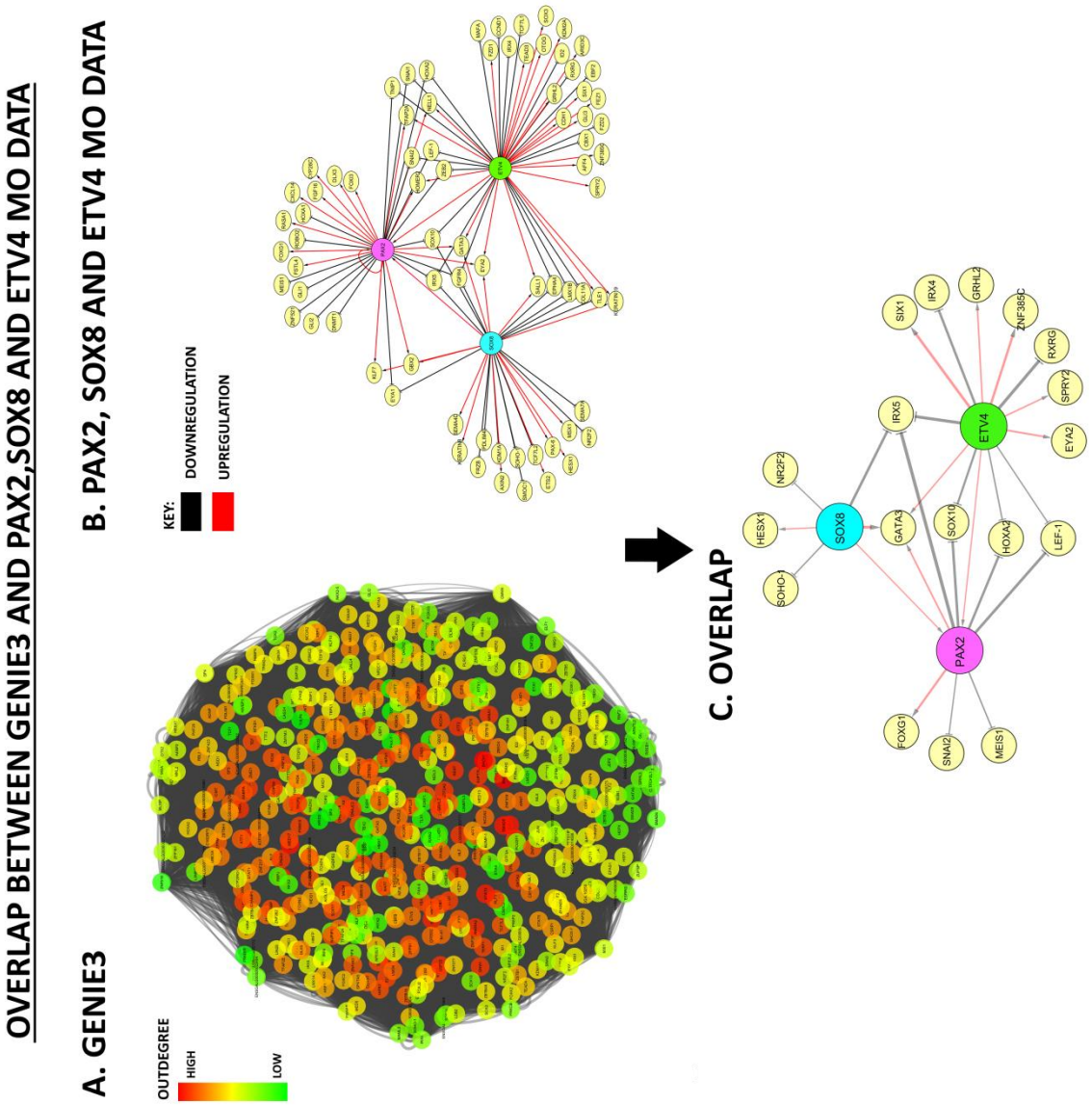


Figure 5.8 Overlap between GENIE3 mRNA-seq network and Pax2, *Sox*8 and Etv4 MO data

(A) Cytoscape view of the directed GENIE3 mRNA-seq network (IM>=0.001). Nodes are coloured according to their out-degrees; Nodes with higher out-degrees in red and nodes with low out-degrees in green. **(B)** Cytoscape view of *Pax2, Sox8* and *Etv4* MO data (Dr. Jingchen Chen and Dr. Monica Tambalo); down regulated interactions are shown in black and up regulated interactions are shown in red. **(C)** Overlap between the predicted GENIE3 mRNA-seq network and MO data recovers 26% of the interactions.

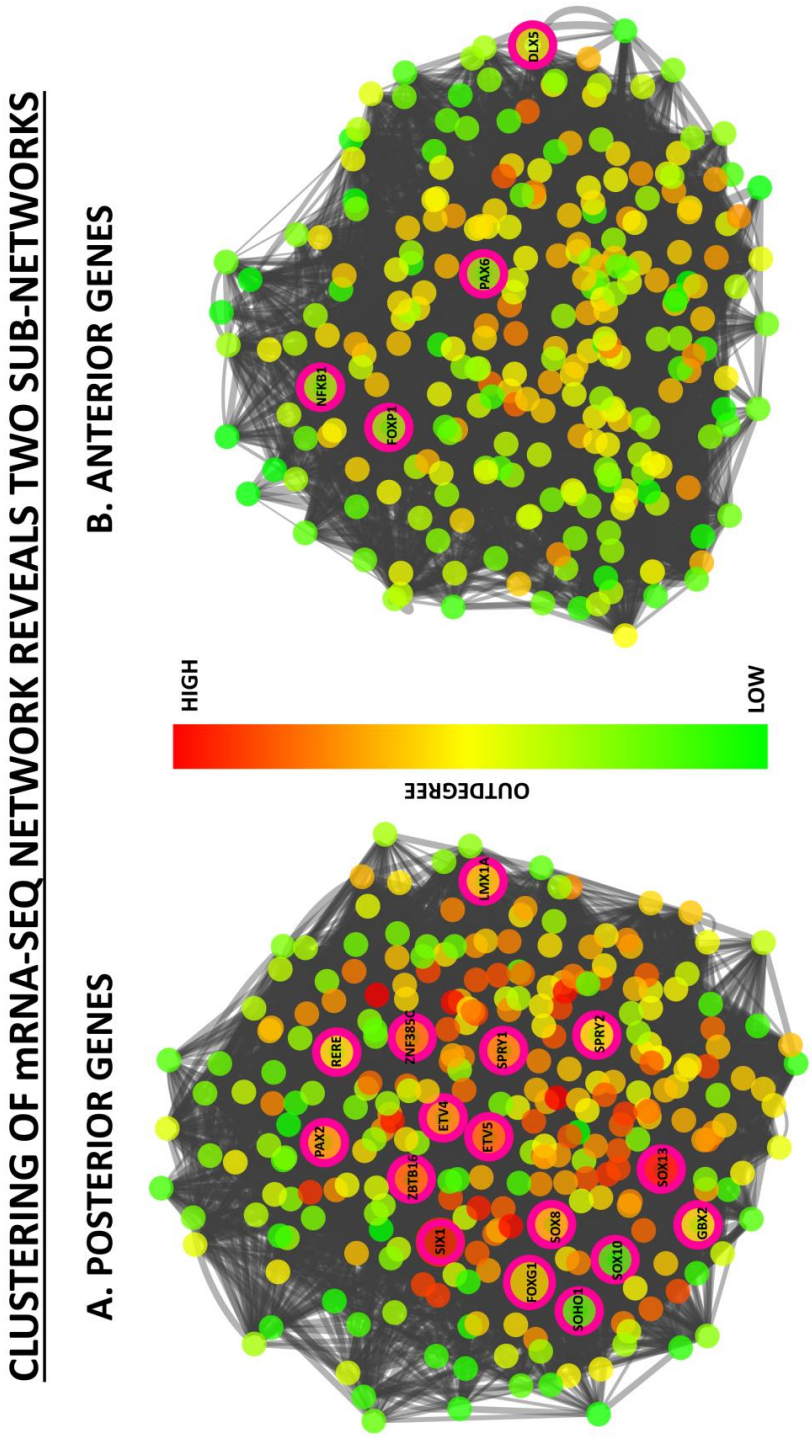


Figure 5.9 Clustering of mRNA-seq predicted network segregates posterior and anterior genes into two sub-networks

Clustering of the mRNA-seq network (IM>0.001) was carried out using Newman's clustering (GLay Plugin in Cytoscape). Nodes are coloured according to their out-degree; Red nodes indicating a higher out-degree; Green nodes indicating a lower out-degree. Edges are weighted according to the IM values. Clustering reveals two main sub-networks; **(A)** Posterior genes: Genes expressed in the posterior PPR, OEPD and the otic placode; some otic genes are highlighted in pink. It is evident that some of the PPR genes such as *Etv4, Etv5* and *Sox13* have the highest out-degrees with OEPD genes such as *Sox8, Foxg1, Pax2* and *Lmx1a* having intermediate out-degrees and late otic genes such as *Soho-1* and *Sox10* having the lowest out-degrees.

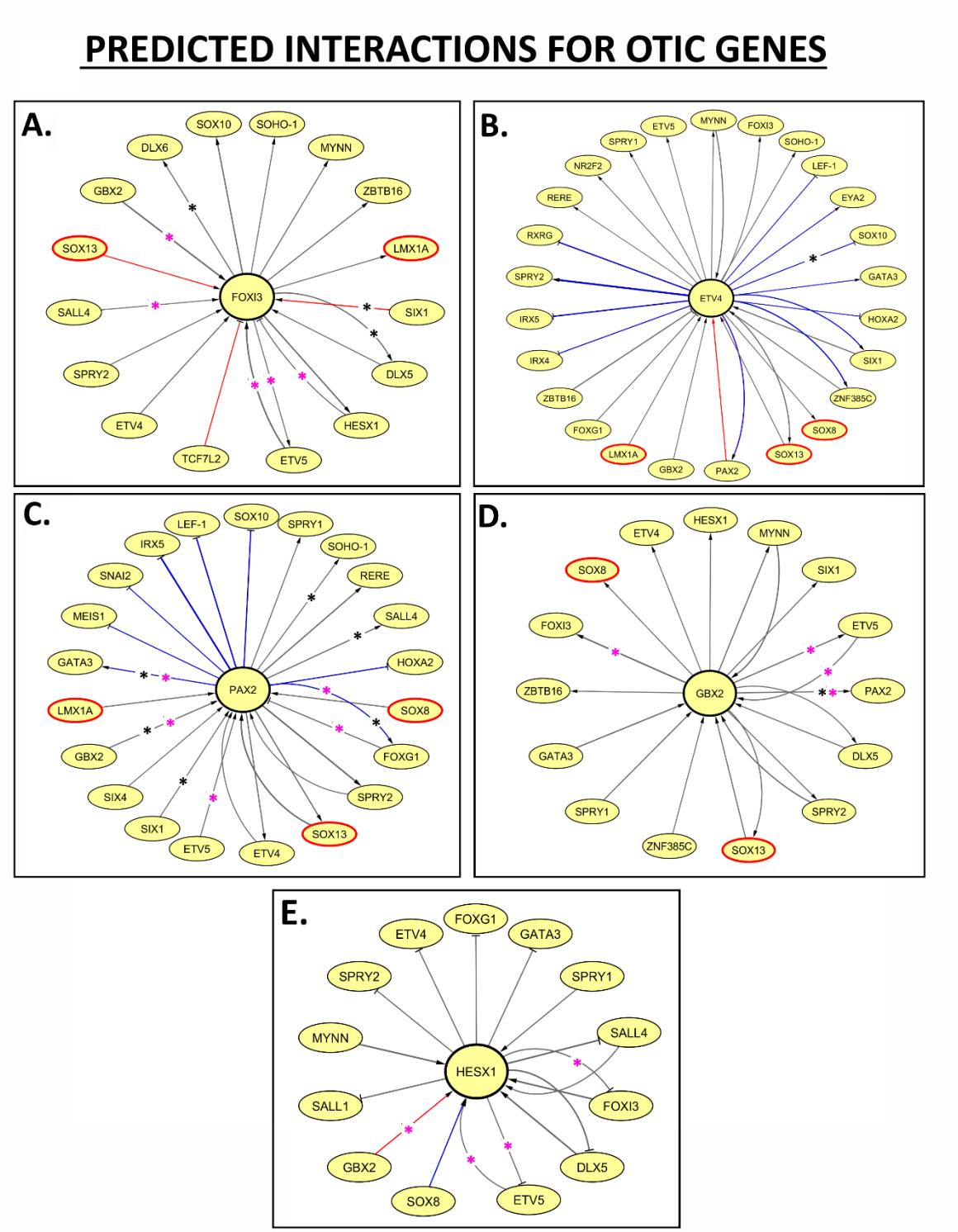


Figure 5.10 Otic-specific interactions in GENIE3-predicted mRNA-seq network

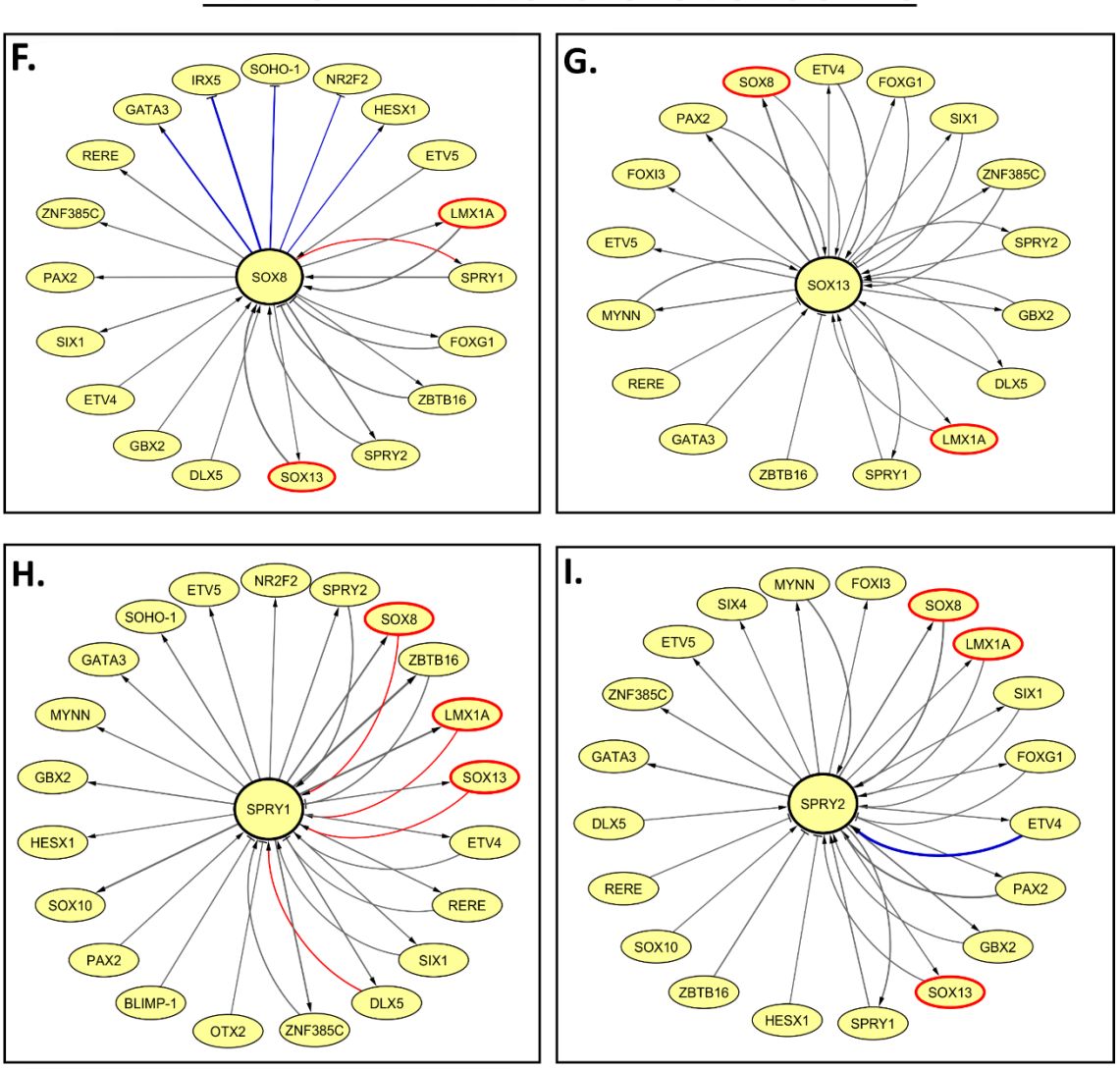


Figure 5.10 Otic-specific interactions in GENIE3-predicted mRNA-seq network
From the GENIE3 mRNA-seq network (IM>0.001), interactions for selected otic genes were plotted and coloured in Cytoscape. Interactions in red have been predicted both in GENIE3 and TFBS analysis of the enhancer of the respective gene. Interactions marked with a black asterisk (*) are known regulatory interactions found by GENIE3. Interactions marked with a pink asterisk (*) are common with predicted NanoString network. Interactions in blue are found from MO data (Dr. Jingchen Chen and Dr. Monica Tambalo). The top regulators of FGF response genes *Sox8, Sox13* and *Lmx1a* (Chapter 4, Figure 4.8; Figure 4.10) are encircled in red **(A)** Interactions for *Foxi3* **(B)** Interactions for *Etv4* **(C)** Interactions for *Pax2* **(D)** Interactions for *Gbx2* **(E)** Interactions for *Hesx1* **(F)** Interactions for *Sox8* **(G)** Interactions for *Sox13* **(H)** Interactions for *Spry1* and **(I)** Interactions for *Spry2*.

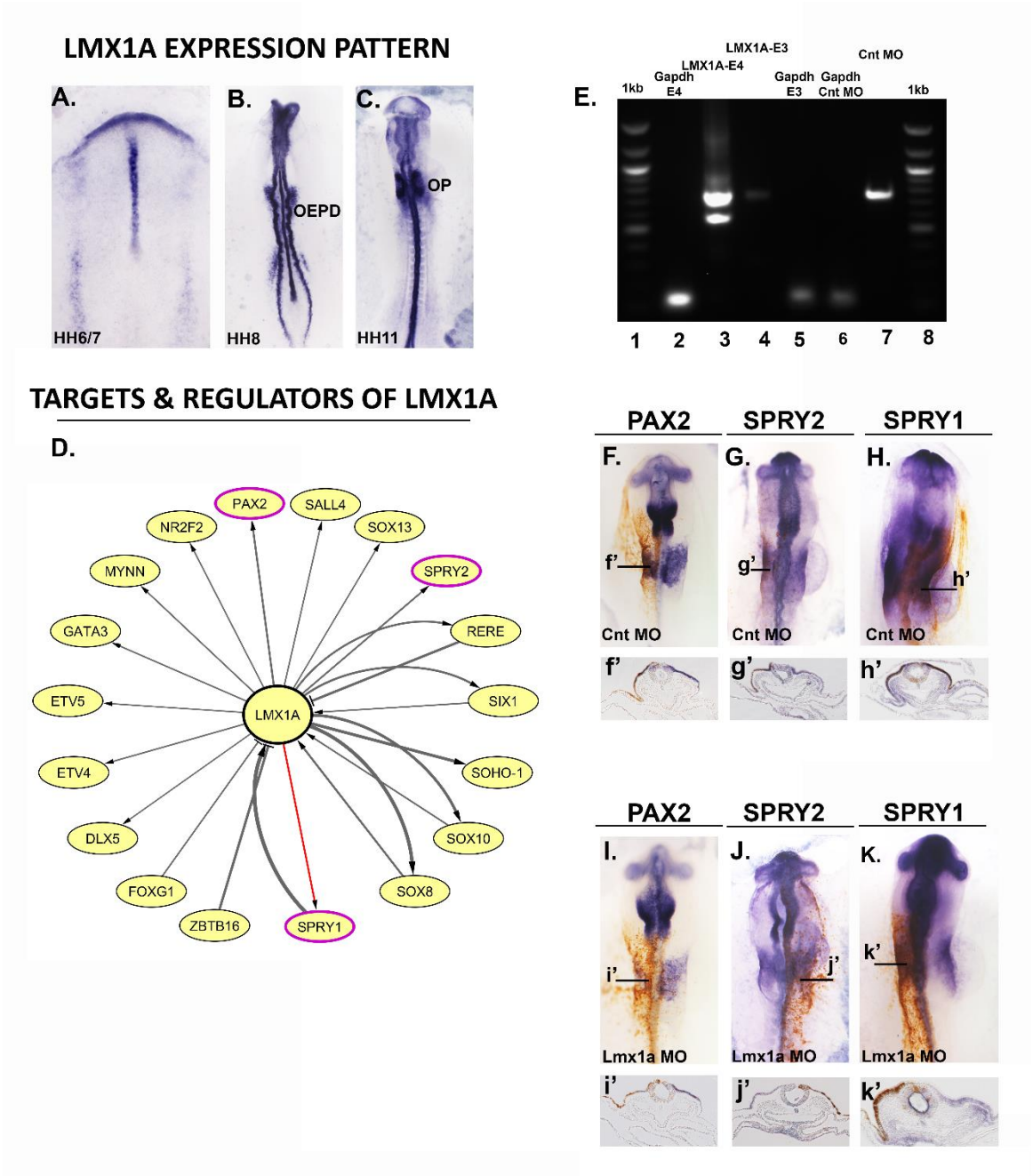


Figure 5.11 Lmx1a: A potential regulator of otic genes

Figure 5.11 Lmx1a: A potential regulator of otic genes
(A-C) *Lmx1a* expression pattern (in situ hybridisation performed by Ramya Ranganathan. At PPR stages (HH6/7), *Lmx1a* is strongly expressed in the notochord and weakly at the edge of the neural plate, including the aPPR, but seems to be absent in the pPPR **(A)**. *Lmx1a* becomes strongly expressed in the OEPD **(B)**. At HH10/11, there is strong expression in otic placode **(C). (D)** Cytoscape view of the *Lmx1a* regulators and targets from the predicted mRNA-seq network (IM>0.001). Interaction in red from *Lmx1a* to *Spry1* has been predicted both in GENIE3 and TFBS analysis of the *Spry1* enhancer. The edges are weighted according to their IM value. Interactions from *Lmx1a* to *Spry1, Spry2* and *Pax2* were selected for experimental verification and therefore these are encircled in purple. **(E)** Two different morpholinos were designed to reduce *Lmx1a* expression by targeting exon-intron boundaries. LMX1A-E4 results in the deletion of exon4 (580 bp) in Lane3. Additionally, the wild-type transcript (750 bp) is also shown. There was insufficient cDNA to assess the efficiency of LMX1A-E3 MO. *Lmx1a* amplification reveals a weak wild type band of 750 bp (Lane 4) but a 580 bp band would be below the detection limit. Morpholinos do not affect the expression of housekeeping gene *Gapdh* (Lanes 2, 5 and 6). MO electroporated embryos are shown in **(F-K).** Electroporated side is shown in brown. In the control morpholino, expression of *Pax2* **(F, f')**, *Spry2* **(G, g')** and *Spry1* **(H, h')** is comparable in the electroporated and non-electroporated side. *Lmx1a* knockdown causes loss of placode thickening and reduced expression of *Pax2* **(I, i'),** *Spry2* **(J, j')** and *Spry1* **(K, k').**

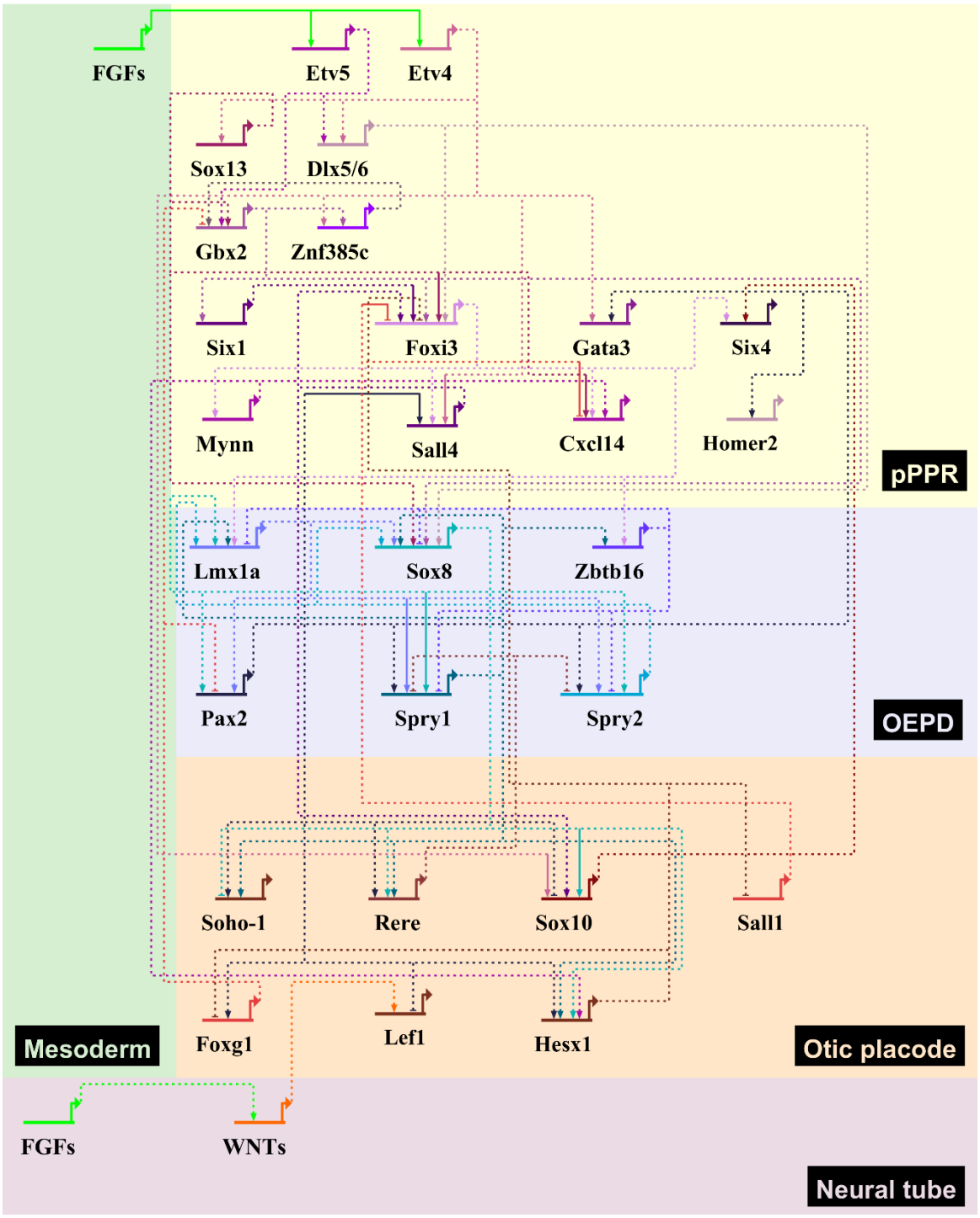


Figure 5.12 An improved GRN for otic development
This figure shows an improved otic GRN built in BioTapestry which incorporates otic-specific interactions from Figure 5.6 and Figure 5.10. The regulatory relationships are divided into three stages: posterior PPR (pPPR), Otic-Epibranchial Progenitor Domain (OEPD) and otic placode. FGF from the underlying mesoderm directly regulates *Etv4* and *Etv5* which in turn regulate other PPR (shades of pink) and OEPD genes (shades of blue). In the OEPD, *Sox8* and *Lmx1a* seem to be the link between PPR and otic genes (shades of brown). Solid lines indicate binding to the enhancer of the target gene. See text for details.

6. Discussion

The main goal of this work was to unravel the GRN that governs the process of development from PPR to otic placode stages. To this end, using a combination of computational and experimental approaches, new otic genes, their enhancers and their transcriptional inputs have been identified. This information was used to enrich a preliminary network (Figure 1.2) now presented as an improved otic GRN (Figure 5.12).

6.1 Computational and experimental methods reveal novel otic enhancers of FGF-response genes

Despite the diversity of sense organs, their precursors arise from a common progenitor domain in the ectoderm called the PPR (Streit, 2002, Bhattacharyya et al., 2004, reviewed in Schlosser, 2006, Xu et al., 2008, Streit, 2008, Schlosser, 2010, Pieper et al., 2012). The next step is the induction of OEPD. It has previously been shown that the paraxial mesoderm underlying the posterior part of PPR plays an important role in inducing the OEPD (Jacobson, 1963a, Orts et al., 1971, Mendonsa and Riley, 1999, Ladher et al., 2000, Phillips et al., 2001, Leger and Brand, 2002, Kil et al., 2005) and evidence from different species implicates FGF signalling as a crucial pathway (for review see: Ohyama et al., 2007, Schimmang, 2007, Ladher et al., 2010). However, the sequence of events downstream of FGF is not very clearly understood. In the present work, many OEPD markers have been identified by analysing different time points (6, 12 and 24 hrs) of *in vitro* culture of PPR with FGF (Chapter 3; section 3.3; Figure 3.1). Hierarchical clustering of these data helped in the identification of a small set of co-expressed transcripts (*Foxi3, Gbx2, Etv4, Sox13, Spry1/2, Cxcl14* and *Hesx1*) that is induced rapidly by FGF (Chapter 3; Figures 3.3 and 3.4). To understand how these genes are regulated and to identify their otic specific enhancers, an integrated approach was used involving identification of

insulator elements, enhancer predictions using phylogenetic footprinting and ChIP-seq for histone modifications (Chapter 3 and 4). The combined use of such methods for enhancer identification during otic development has not been reported before. Some studies have only employed phylogenetic footprinting methods (alignment-based) to identify enhancers for *Six1, Eya2* and *Sox10* (Ishihara et al., 2008b, Sato et al., 2010, Betancur et al., 2011, Sato et al., 2012) while others have identified enhancers using ChIP-seq for histone modifications together with p300 binding (Visel et al., 2009, Ghisletti et al., 2010, Blow et al., 2010, Blum et al., 2012). This project is therefore unique in the sense that data from multiple methods have been combined in order to identify enhancers. An important point is the use of DREiVe (Sosinsky et al., 2007) in this study which is superior to alignment-based methods in predicting enhancers because of its ability to determine clusters of conserved motifs in any orientation making it possible to find conservation even in evolutionarily re-arranged sequences (Sosinsky et al., 2007). Its credibility in predicting enhancers is revealed from the fact that it identified 72% of previously known *Sox2* enhancers (Khan et al., 2013). Although evolutionary conservation is widely used for the prediction of enhancers, one concern about predictions is that these identify many putative enhancers (Chapter 3; Table 3.2), which may correspond to different expression domains of a gene. Hence, there is a need to couple these predictions with other methods such as histone ChIP-seq for the required tissue (Chapter 4) in order to pick enhancers that are both conserved and specific to the tissue.

A histone ChIP-seq was thus carried out on FGF-treated and control PPR tissues (Chapter 4), which is one of the first genome-wide studies to identify FGF-responsive enhancers during OEPD induction. Signalling pathways have been linked before to epigenetic changes at gene loci (Mosimann et al., 2009, Wamstad et al., 2012, Patel et al., 2013) and

recently Zhang and colleagues have demonstrated that H3K27ac is dynamically deposited in response to VEGFA stimulation leading to increased gene expression (Zhang et al., 2013a). In the present study, a similar increase in H3K27ac level was observed upon FGF treatment at the locus of FGF-response genes particularly those that are rapidly induced by FGF (*Etv4, Foxi3, Gbx2, Spry1/2, Sox13, Hesx1* and *Cxcl14).* It has been reported that p300/CREB acetylates H3K27 (for review see: Karamouzis et al., 2007, Holmqvist and Mannervik, 2013) and MAP kinase, downstream of FGF, can stimulate p300/CREB by phosphorylating it (Ait-Si-Ali et al., 1999). This points towards a potential role of FGF in deposition of active marks to its target genes. However, this will require further experiments. One possibility is to treat PPR tissue with FGF and simultaneously block the activity of p300. If H3K27 acetylation is the cause for gene activation, then FGF will no longer be able to activate its target genes when p300 is blocked. Conversely, if acetylation is the consequence of gene activity, then the activation of OEPD genes will not be affected when p300 is blocked.

From the comparison between FGF-treated and control tissues, a number of unique proximal enhancers were identified for *Etv4, Foxi3, Gbx2, Spry1/2, Sox13, Hesx1* and *Cxcl14* (Chapter 4). Predicted enhancers were overlapped with the ChIP-seq enhancers to identify putative conserved, otic-specific enhancers (Chapter 4; Table 4.1). The advantage of identifying enhancers using conservation along with histone modification profiles is revealed from the fact that four out of the five overlapped enhancers were found to be active (Chapter 4; Figures 4.12-4.15). Thus, this supports the idea of using conservation along with ChIP-seq in order to find otic enhancers.

In the present work, only proximal enhancers were tested for *in vivo* activity, however, the prediction of insulating boundaries (Chapter 3; Table 3.1) has allowed us to define

regions within which other putative enhancers can be found for experimental validation in future.

Finally, enhancer identification can be further improved through machine learning approaches that can learn DNA sequence features of experimentally determined enhancers (Heintzman et al., 2007, Thurman et al., 2012, Cotney et al., 2012) and use these to predict novel enhancers (Erwin et al., 2014). This strategy can be very powerful if the training dataset is not too small. For example in a recent study (Narlikar et al., 2010), 77 vertebrate embryonic heart enhancers were used to train the classifier which was then able to predict 40,000 novel heart enhancers in the human genome. Some of these were validated in a zebrafish *in vivo* reporter assay and demonstrated a success rate of 62% (Narlikar et al., 2010). At present, a very small set of experimentally validated otic enhancers is available but in future with the availability of more data, such methods can be implemented.

6.2 *Sox* family members, *Lmx1a* and *Sall1* are key regulators of FGF-response genes

TFBS analysis of unique FGF-responsive enhancers revealed *Sox8, Sox13* and *Lmx1a* as the top TFs (Chapter 4; Figure 4.10). All three are transcriptional activators with *Sox13* expressed at the PPR stage (Tambalo, 2015) and *Lmx1a* (Figure 5.11 B) and *Sox8* (McKeown et al., 2005) expressed at the OEPD stage. *Sox8* MO knockdown causes reduction of the OEPD marker *Pax2* (Tambalo, 2015) demonstrating the importance of *Sox8* in otic development. Similarly, *Lmx1a* null mice present with ear malformations demonstrating an important role of *Lmx1a* in the normal development of the ears (Nichols D. H., 2008). Its importance in otic induction is further strengthened by the fact that *Lmx1a* MO knockdown causes reduction of the OEPD genes *Pax2, Spry1* and *Spry2* (Figure 5.11 F-K). It can thus be speculated that *Sox8* and *Lmx1a* are upstream of the

OEPD marker *Pax2* and therefore important players in activating the downstream otic fate. On the other hand, *Sox* family members have also been implicated in recruiting p300/CREB to specific genomic loci (Tsuda et al., 2003, Chen et al., 2004, Furumatsu et al., 2005a, Furumatsu et al., 2005b). The abundance of *Sox8* and *Sox13* binding sites in FGF-responsive enhancers (Chapter 4; Figures 4.12, 4.15, 4.16) could mean that these recruit p300/CREB to the corresponding enhancers causing their acetylation and subsequently activation to allow other factors to bind. Conversely, *Sall1,* a repressor which is initially expressed in the anterior PPR and later expressed in the otic placode (Sweetman et al., 2005), had binding sites in enhancers of *Foxi3* and *Cxcl14* (Chapter 4; Figure 4.12; Figure 4.16). Both *Foxi3* and *Cxcl14* are initially expressed in the PPR and OEPD, but *Foxi3* is later expressed in the epibranchial and trigeminal regions (Khatri and Groves, 2013) whereas *Cxcl14* is expressed in a ring of ectoderm around the otic placode (Lleras-Forero et al., 2013). Thus, *Sall1* may be responsible for removing their mRNA from the otic placode. The importance of *Sall1* in normal otic development is revealed by the fact that mutations in *Sall1* are associated with the Townes-Brocks syndrome (TBS) in humans which is a rare, autosomal dominant malformation that presents with anal, renal, limb and ear anomalies (Kohlhase et al., 1998). Moreover, heterozygous mice for *Sall1* mutation mimic TBS patients by showing sensorineural hearing loss, renal cystic hypoplasia and bone abnormalities (Kiefer et al., 2003). Because *Foxi3* promotes the epibranchial fate at later stages (Khatri and Groves, 2013), it can be speculated that *Sall1* may be playing a vital role in suppressing *Foxi3* expression in the otic placode through binding to its enhancer, thus promoting the otic fate at the expense of epibranchial fate.

In conclusion, *Sox8, Sox13, Lmx1a* and *Sall1* are some of the key players promoting OEPD and otic fate. However, further experiments are required to link these TFs to their

respective enhancers. One possibility is to mutate the binding sites within the respective enhancers and assess enhancer activity.

6.3 Gene regulatory network inference elucidates regulatory relationships downstream of FGF signalling during otic induction

To characterise the process of otic specification, gene expression was analyzed by means of both NanoString and mRNA-seq which generates large datasets. To get a systems-level view of the data, gene regulatory networks were inferred for each of these expression datasets using GENIE3 (Huynh-Thu et al., 2010). Such approaches have been implemented before (D'Haeseleer et al., 2000, de Hoon et al., 2003, Wang et al., 2006) using an array of different techniques that infer networks from expression data (Gardner and Faith, 2005, Margolin et al., 2006, Bansal M, 2007, Markowetz and Spang, 2007, Huynh-Thu et al., 2010). However, in this study, to gain confidence in the predicted network interactions, TFBS analysis of otic enhancers (Chapter 4; Figure 4.12-4.16), known interactions (Chapter 1; Table 1.1; Figure 1.2) and MO data (Dr. Jingchen Chen and Dr. Monica Tambalo) have been compared to the network. Moreover, clustering of the network helped in identification of various modules in the network such as those responding positively or negatively to FGF from NanoString data (Chapter 5; Figure 5.2) and those consisting of anterior or posterior genes from mRNA-seq data (Chapter 5; Figure 5.9), which allowed the identification of key otic players and their targets within each module (Chapter 5; section 5.4; section 5.9).

Upon detailed analysis of the modules and top interactions of FGF-response genes, it appeared that *Sox8, Sox13* and *Lmx1a* were among the key regulators of many OEPD and otic genes including *Pax2, Spry1, Spry2* and *Hesx1,* which is consistent with the TFBS analysis of otic enhancers (Chapter 4; Figures 4.12-4.16). Moreover, network analysis

helped to elucidate the preliminary network downstream of FGF signalling that is responsible for otic induction (Chapter 1, Figure 1.2). For example, *Etv4* and *Etv5* are direct targets of FGF signalling (Tambalo, 2015). It is thus possible that other FGF-response genes are targets of *Etv4* and *Etv5*. To this end, predictions from the network as well as *Etv4* MO data (Dr. Jingchen Chen and Dr. Monica Tambalo) were compared. Indeed, positively-regulated early FGF-response genes *Foxi3, Pax2, Spry1* and *Spry2* and negatively-regulated FGF-response genes *Sox10* and *Lef-1* are targets of *Etv4* (Chapter 5; Figure 5.10 B) indicating that FGF may be acting on these genes via *Etv4.*

The OEPD marker *Pax2* has a central role in otic induction whereby mutations in its locus have been associated with human sensorineural deafness (Sanyanusin et al., 1995b, Schimmenti et al., 1997). It is therefore important to understand how this gene is regulated. In the present study, the *Pax2* locus was not analyzed and therefore no enhancers have been identified for *Pax2* as yet, however, *Etv4, Sox8* and *Lmx1a* MO knockdown result in reduction of *Pax2* (Chapter 5; Figure 5.10 B; Figure 5.11 I), showing that these are required for defining the otic fate through the regulation of *Pax2*. Interestingly, network predictions help to place *Etv4, Sox8, Lmx1a* and *Pax2* in a hierarchy. Upon analysis of the network, interactions were found from *Etv4* to *Sox8, Sox8* to *Lmx1a* and from both *Sox8* and *Lmx1a* to *Pax2* indicating that *Etv4* indirectly acts on *Pax2* by possibly activating the intermediate genes *Sox8* and *Lmx1a.* This hypothesis could be further tested by analyzing the *Pax2* locus in detail and identifying its otic enhancers and transcriptional inputs within these enhancers to see if the analysis produces the same predictions as above.

While no targets are known for *Lmx1a* during otic induction*, Sox10* and *Sall4* are known targets of *Etv4* (Barembaum and Bronner-Fraser, 2010, Betancur et al., 2011) and *Sox10*

is a known target of *Sox8* (Betancur et al., 2011). This study has contributed to the elucidation of relationships between direct targets (*Etv4/5*) of FGF signalling and other downstream genes including *Sox8, Lmx1a* and *Pax2*. A few interactions have also been validated (Figure 5.11 D, nodes highlighted in purple) which provide good support for the network approach.

In conclusion the predicted networks offer a wealth of information and highlight new regulatory relationships that can be tested. One of the current concerns in gene regulatory network inference is the accurate prediction of direct targets (Yip et al., 2010). This can be improved by developing algorithms that can integrate expression data with protein-DNA binding data such as that from ChIP-chip or ChIP-seq experiments (Gong et al., 2015). However, these methods can be implemented only after the large-scale availability of such binding data in chick. Until then, it is suggested to adopt a strategy similar to the present work where various methods were employed independently and subsequently combined to build a GRN.

6.4 Association of otic genes with deafness

This project also gives some insight into the relationship of new otic genes with deafness-associated loci. At present, many human genomic loci have been characterised at the mutational level and associated with non-syndromic hearing loss (Van Camp and Smith, 2014). For the vast majority of these loci, the causal gene is still missing. Therefore, in this study, it was interesting to find out if any of the otic genes for which enhancers and regulatory relationships have been identified are associated with any of these loci. A preliminary analysis was carried out using the hearing loss database (http://hereditaryhearingloss.org/) which revealed that indeed *Foxi3* is associated with the autosomal recessive deafness locus DFNB47 (Hassan et al., 2006), while *Spry1* is

associated with the autosomal dominant deafness locus DFNA24 (Hafner et al., 2000). Similarly, *Lmx1a* which was identified as one of the key OEPD genes regulating many downstream otic genes is also associated with the autosomal dominant deafness locus DFNA49 (Moreno-Pelayo et al., 2003). Since the causal genes for these loci have not been identified, it is possible that *Foxi3, Spry1* and *Lmx1a* are putative causal genes. In future, a systematic analysis of the loci of newly identified otic genes can help to elucidate their association with deafness loci. Moreover, regulatory relationships from the predicted GRNs can further help to understand their circuit of interactions.

Overall, this thesis combines various computational and experimental approaches to present an improved otic gene regulatory network (Figure 5.12). This network is a resource to identify regulatory relationships between various otic genes and provides guidelines for future experiments.

7. Bibliography

ABDELHAK, S., KALATZIS, V., HEILIG, R., COMPAIN, S., SAMSON, D., VINCENT, C., WEIL, D., CRUAUD, C., SAHLY, I., LEIBOVICI, M., BITNER-GLINDZICZ, M., FRANCIS, M., LACOMBE, D., VIGNERON, J., CHARACHON, R., BOVEN, K., BEDBEDER, P., VAN REGEMORTER, N., WEISSENBACH, J. & PETIT, C. 1997. A human homologue of the Drosophila eyes absent gene underlies branchio-oto-renal (BOR) syndrome and identifies a novel gene family. *Nat Genet,* 15**,** 157-64.

ABE, Y., CHEN, W., HUANG, W., NISHINO, M. & LI, Y. P. 2006. CNBP regulates forebrain formation at organogenesis stage in chick embryos. *Dev Biol,* 295**,** 116-27.

ABELLO, G., KHATRI, S., GIRALDEZ, F. & ALSINA, B. 2007. Early regionalization of the otic placode and its regulation by the Notch signaling pathway. *Mech Dev,* 124**,** 631-45.

ABELLO, G., KHATRI, S., RADOSEVIC, M., SCOTTING, P. J., GIRALDEZ, F. & ALSINA, B. 2010. Independent regulation of Sox3 and Lmx1b by FGF and BMP signaling influences the neurogenic and non-neurogenic domains in the chick otic placode. *Dev Biol,* 339**,** 166-78.

ABRAHAM, S., PAKNIKAR, R., BHUMBRA, S., LUAN, D., GARG, R., DRESSLER, G. R. & PATEL, S. R. 2015. The Groucho Associated Phosphatase PPM1B Displaces Pax Transactivation Domain Interacting Protein (PTIP) to Switch the transcription factor Pax2 from a Transcriptional Activator to a Repressor. *J Biol Chem*.

ABU-ELMAGD, M., ISHII, Y., CHEUNG, M., REX, M., LE ROUEDEC, D. & SCOTTING, P. J. 2001. cSox3 expression and neurogenesis in the epibranchial placodes. *Dev Biol,* 237**,** 258-69.

ACAMPORA, D., GULISANO, M., BROCCOLI, V. & SIMEONE, A. 2001. Otx genes in brain morphogenesis. *Prog Neurobiol,* 64**,** 69-95.

ACAMPORA, D., MAZAN, S., LALLEMAND, Y., AVANTAGGIATO, V., MAURY, M., SIMEONE, A. & BRULET, P. 1995. Forebrain and midbrain regions are deleted in Otx2-/- mutants due to a defective anterior neuroectoderm specification during gastrulation. *Development,* 121**,** 3279-90.

ADAMS, M. D., CELNIKER, S. E., HOLT, R. A., EVANS, C. A., GOCAYNE, J. D., AMANATIDES, P. G., SCHERER, S. E., LI, P. W., HOSKINS, R. A., GALLE, R. F., GEORGE, R. A., LEWIS, S. E., RICHARDS, S., ASHBURNER, M., HENDERSON, S. N., SUTTON, G. G., WORTMAN, J. R., YANDELL, M. D., ZHANG, Q., CHEN, L. X., BRANDON, R. C., ROGERS, Y. H., BLAZEJ, R. G., CHAMPE, M., PFEIFFER, B. D., WAN, K. H., DOYLE, C., BAXTER, E. G., HELT, G., NELSON, C. R., GABOR, G. L., ABRIL, J. F., AGBAYANI, A., AN, H. J., ANDREWS-PFANNKOCH, C., BALDWIN, D., BALLEW, R. M., BASU, A., BAXENDALE, J., BAYRAKTAROGLU, L., BEASLEY, E. M., BEESON, K. Y., BENOS, P. V., BERMAN, B. P., BHANDARI, D., BOLSHAKOV, S., BORKOVA, D., BOTCHAN, M. R., BOUCK, J., BROKSTEIN, P., BROTTIER, P., BURTIS, K. C., BUSAM, D. A., BUTLER, H., CADIEU, E., CENTER, A., CHANDRA, I., CHERRY, J. M., CAWLEY, S., DAHLKE, C., DAVENPORT, L. B., DAVIES, P., DE PABLOS, B., DELCHER, A., DENG, Z., MAYS, A. D., DEW, I., DIETZ, S. M., DODSON, K., DOUP, L. E., DOWNES, M., DUGAN-ROCHA, S., DUNKOV, B. C., DUNN, P., DURBIN, K. J., EVANGELISTA, C. C., FERRAZ, C., FERRIERA, S., FLEISCHMANN, W., FOSLER, C.,

GABRIELIAN, A. E., GARG, N. S., GELBART, W. M., GLASSER, K., GLODEK, A., GONG, F., GORRELL, J. H., GU, Z., GUAN, P., HARRIS, M., HARRIS, N. L., HARVEY, D., HEIMAN, T. J., HERNANDEZ, J. R., HOUCK, J., HOSTIN, D., HOUSTON, K. A., HOWLAND, T. J., WEI, M. H., IBEGWAM, C., et al. 2000. The genome sequence of Drosophila melanogaster. *Science,* 287**,** 2185-95.

ADAMSKA, LEGER, S., BRAND, M., HADRYS, T., BRAUN, T. & BOBER, E. 2000. Inner ear and lateral line expression of a zebrafish Nkx5-1 gene and its downregulation in the ears of FGF8 mutant, ace. *Mech Dev,* 97**,** 161-5.

ADAMSKA, WOLFF, A., KREUSLER, M., WITTBRODT, J., BRAUN, T. & BOBER, E. 2001. Five Nkx5 genes show differential expression patterns in anlagen of sensory organs in medaka: insight into the evolution of the gene family. *Development Genes and Evolution,* 211**,** 338-349.

ADLI, M. & BERNSTEIN, B. E. 2011. Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq. *Nat Protoc,* 6**,** 1656-68.

AGHAALLAEI, N., BAJOGHLI, B. & CZERNY, T. 2007. Distinct roles of Fgf8, Foxi1, Dlx3b and Pax8/2 during otic vesicle induction and maintenance in medaka. *Dev Biol,* 307**,** 408-20.

AHLGREN, S., VOGT, P. & BRONNER-FRASER, M. 2003. Excess FoxG1 causes overgrowth of the neural tube. *J Neurobiol,* 57**,** 337-49.

AHRENS, K. & SCHLOSSER, G. 2005. Tissues and signals involved in the induction of placodal Six1 expression in Xenopus laevis. *Dev Biol,* 288**,** 40-59.

AIT-SI-ALI, S., CARLISI, D., RAMIREZ, S., UPEGUI-GONZALEZ, L. C., DUQUET, A., ROBIN, P., RUDKIN, B., HAREL-BELLAN, A. & TROUCHE, D. 1999. Phosphorylation by p44 MAP Kinase/ERK1 stimulates CBP histone acetyl transferase activity in vitro. *Biochem Biophys Res Commun,* 262**,** 157-62.

ALKEMA, W. B., JOHANSSON, O., LAGERGREN, J. & WASSERMAN, W. W. 2004. MSCAN: identification of functional clusters of transcription factor binding sites. *Nucleic Acids Res,* 32**,** W195-8.

ALTMANN, C. R., CHOW, R. L., LANG, R. A. & HEMMATI-BRIVANLOU, A. 1997. Lens induction by Pax-6 in Xenopus laevis. *Dev Biol,* 185**,** 119-23.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

ALVAREZ, Y., ALONSO, M. T., VENDRELL, V., ZELARAYAN, L. C., CHAMERO, P., THEIL, T., BOSL, M. R., KATO, S., MACONOCHIE, M., RIETHMACHER, D. & SCHIMMANG, T. 2003. Requirements for FGF3 and FGF10 during inner ear formation. *Development,* 130**,** 6329-38.

AMANO, T., SAGAI, T., TANABE, H., MIZUSHINA, Y., NAKAZAWA, H. & SHIROISHI, T. 2009. Chromosomal Dynamics at the Shh Locus: Limb Bud-Specific Differential Regulation of Competence and Active Transcription. *Developmental Cell,* 16**,** 47-57.

ANDERSSON, R., GEBHARD, C., MIGUEL-ESCALADA, I., HOOF, I., BORNHOLDT, J., BOYD, M., CHEN, Y., ZHAO, X., SCHMIDL, C., SUZUKI, T., NTINI, E., ARNER, E., VALEN, E., LI, K., SCHWARZFISCHER, L., GLATZ, D., RAITHEL, J., LILJE, B., RAPIN, N., BAGGER, F. O., JORGENSEN, M., ANDERSEN, P. R., BERTIN, N., RACKHAM, O., BURROUGHS, A. M., BAILLIE, J. K., ISHIZU, Y., SHIMIZU, Y., FURUHATA, E., MAEDA, S., NEGISHI, Y., MUNGALL, C. J., MEEHAN, T. F., LASSMANN, T., ITOH, M., KAWAJI, H., KONDO, N., KAWAI, J., LENNARTSSON, A., DAUB, C. O., HEUTINK, P., HUME, D. A., JENSEN, T. H., SUZUKI, H., HAYASHIZAKI, Y., MULLER, F., CONSORTIUM, F.,

FORREST, A. R., CARNINCI, P., REHLI, M. & SANDELIN, A. 2014. An atlas of active enhancers across human cell types and tissues. *Nature,* 507**,** 455-61.

ANDREWS, S. 2010. FastQC: A quality control tool for high throughput sequence data. .

ANNUNZIATO, A. 2008. DNA Packaging: Nucleosomes and Chromatin. *Nature Education,* 1.

ANSARI, S. A. & MORSE, R. H. 2013. Mechanisms of Mediator complex action in transcriptional activation. *Cell Mol Life Sci,* 70**,** 2743-56.

APARICIO, S., CHAPMAN, J., STUPKA, E., PUTNAM, N., CHIA, J. M., DEHAL, P., CHRISTOFFELS, A., RASH, S., HOON, S., SMIT, A., GELPKE, M. D., ROACH, J., OH, T., HO, I. Y., WONG, M., DETTER, C., VERHOEF, F., PREDKI, P., TAY, A., LUCAS, S., RICHARDSON, P., SMITH, S. F., CLARK, M. S., EDWARDS, Y. J., DOGGETT, N., ZHARKIKH, A., TAVTIGIAN, S. V., PRUSS, D., BARNSTEAD, M., EVANS, C., BADEN, H., POWELL, J., GLUSMAN, G., ROWEN, L., HOOD, L., TAN, Y. H., ELGAR, G., HAWKINS, T., VENKATESH, B., ROKHSAR, D. & BRENNER, S. 2002. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. *Science,* 297**,** 1301-10.

APARICIO, S., MORRISON, A., GOULD, A., GILTHORPE, J., CHAUDHURI, C., RIGBY, P., KRUMLAUF, R. & BRENNER, S. 1995. Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, Fugu rubripes. *Proc Natl Acad Sci U S A,* 92**,** 1684-8.

ARNOLD, C. D., GERLACH, D., STELZER, C., BORYN, L. M., RATH, M. & STARK, A. 2013. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science,* 339**,** 1074-7.

ASHERY-PADAN, R., MARQUARDT, T., ZHOU, X. & GRUSS, P. 2000. Pax6 activity in the lens primordium is required for lens formation and for correct placement of a single retina in the eye. *Genes Dev,* 14**,** 2701-11.

ASP, P., BLUM, R., VETHANTHAM, V., PARISI, F., MICSINAI, M., CHENG, J., BOWMAN, C., KLUGER, Y. & DYNLACHT, B. D. 2011. Genome-wide remodeling of the epigenetic landscape during myogenic differentiation. *Proc Natl Acad Sci U S A,* 108**,** E149-58.

BAGHERI-FAM, S., BARRIONUEVO, F., DOHRMANN, U., GÜNTHER, T., SCHÜLE, R., KEMLER, R., MALLO, M., KANZLER, B. & & SCHERER, G. 2006. Long-range upstream and downstream enhancers control distinct subsets of the complex spatiotemporal Sox9 expression pattern. *Dev Biol.,* 291**,** 382-397.

BAGHERI-FAM, S., FERRAZ, C., DEMAILLE, J., SCHERER, G. & PFEIFER, D. 2001. Comparative genomics of the SOX9 region in human and Fugu rubripes: conservation of short regulatory sequence elements within large intergenic regions. *Genomics,* 78**,** 73-82.

BAI, R. Y., STAEDTKE, V., LIDOV, H. G., EBERHART, C. G. & RIGGINS, G. J. 2012. OTX2 represses myogenic and neuronal differentiation in medulloblastoma cells. *Cancer Res,* 72**,** 5988-6001.

BAILEY, A. P., BHATTACHARYYA, S., BRONNER-FRASER, M. & STREIT, A. 2006. Lens specification is the ground state of all sensory placodes, from which FGF promotes olfactory identity. *Developmental Cell,* 11**,** 505-517.

BAILEY, T. L. 2002. Discovering novel sequence motifs with MEME. *Curr Protoc Bioinformatics,* Chapter 2**,** Unit 2 4.

BAILEY, T. L. & ELKAN, C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol,* 2**,** 28-36.

BAILEY, T. L. & ELKAN, C. 1995. The value of prior knowledge in discovering motifs with MEME. *Proc Int Conf Intell Syst Mol Biol,* 3**,** 21-9.
BAKER, C. V. & BRONNER-FRASER, M. 2000. Establishing neuronal identity in vertebrate neurogenic placodes. *Development,* 127**,** 3045-56.
BAKER, C. V. & BRONNER-FRASER, M. 2001. Vertebrate cranial placodes I. Embryonic induction. *Dev Biol,* 232**,** 1-61.
BALLY-CUIF, L., GULISANO, M., BROCCOLI, V. & BONCINELLI, E. 1995. c-otx2 is expressed in two different phases of gastrulation and is sensitive to retinoic acid treatment in chick embryo. *Mech Dev,* 49**,** 49-63.
BANERJI, J., RUSCONI, S. & SCHAFFNER, W. 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA. *Cell,* 27**,** 299-308.
BANET, G., BIBI, O., MATOUK, I., AYESH, S., LASTER, M., KIMBER, K. M., TYKOCINSKI, M., DE GROOT, N., HOCHBERG, A. & OHANA, P. 2000. Characterization of human and mouse H19 regulatory sequences. *Mol Biol Rep,* 27**,** 157-65.
BANIAHMAD, A., STEINER, C., KOHNE, A. C. & RENKAWITZ, R. 1990. Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. *Cell,* 61**,** 505-14.
BANSAL M, B. V., AMBESI-IMPIOMBATO A, DI BERNARDO D 2007. How to infer gene networks from expression profiles. *Mol Syst Biol* 3.
BAREMBAUM, M. & BRONNER-FRASER, M. 2007. Spalt4 mediates invagination and otic placode gene expression in cranial ectoderm. *Development,* 134**,** 3805-14.
BAREMBAUM, M. & BRONNER-FRASER, M. 2010. Pax2 and Pea3 synergize to activate a novel regulatory enhancer for spalt4 in the developing ear. *Dev Biol,* 340**,** 222-31.
BARKESS, G. & WEST, A. G. 2012. Chromatin insulator elements: establishing barriers to set heterochromatin boundaries. *Epigenomics,* 4**,** 67-80.
BARSKI, A., CUDDAPAH, S., CUI, K., ROH, T. Y., SCHONES, D. E., WANG, Z., WEI, G., CHEPELEV, I. & ZHAO, K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell,* 129**,** 823-37.
BARSKI, A. & ZHAO, K. 2009. Genomic location analysis by ChIP-Seq. *J Cell Biochem,* 107**,** 11-8.
BARTOLOMEI, M. S., ZEMEL, S. & TILGHMAN, S. M. 1991. Parental imprinting of the mouse H19 gene. *Nature,* 351**,** 153-5.
BARTON, G. J. & STERNBERG, M. J. 1987. A strategy for the rapid multiple alignment of protein sequences. Confidence levels from tertiary structure comparisons. *J Mol Biol,* 198**,** 327-37.
BASSO, K., MARGOLIN, A. A., STOLOVITZKY, G., KLEIN, U., DALLA-FAVERA, R. & CALIFANO, A. 2005. Reverse engineering of regulatory networks in human B cells. *Nat Genet,* 37**,** 382-90.
BEGBIE, J. 2002. Early Steps in the Production of Sensory Neurons by the Neurogenic Placodes. *Molecular and Cellular Neuroscience,* 21**,** 502-511.
BEGBIE, J., BRUNET, J. F., RUBENSTEIN, J. L. & GRAHAM, A. 1999. Induction of the epibranchial placodes. *Development,* 126**,** 895-902.
BELL, A. C. & FELSENFELD, G. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature,* 405**,** 482-5.
BELL, A. C., WEST, A. G. & FELSENFELD, G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell,* 98**,** 387-96.
BENJAMINI, Y., AND HOCHBERG, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple hypothesis testing. *J R Stat Soc B* 57**,** 289–300.

BERNSTEIN, B. E., MIKKELSEN, T. S., XIE, X., KAMAL, M., HUEBERT, D. J., CUFF, J., FRY, B., MEISSNER, A., WERNIG, M., PLATH, K., JAENISCH, R., WAGSCHAL, A., FEIL, R., SCHREIBER, S. L. & LANDER, E. S. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell,* 125**,** 315-26.

BESSARAB, D. A., CHONG, S. W. & KORZH, V. 2004. Expression of zebrafish six1 during sensory organ development and myogenesis. *Dev Dyn,* 230**,** 781-6.

BETANCUR, P., BRONNER-FRASER, M. & SAUKA-SPENGLER, T. 2010a. Assembling neural crest regulatory circuits into a gene regulatory network. *Annu Rev Cell Dev Biol,* 26**,** 581-603.

BETANCUR, P., BRONNER-FRASER, M. & SAUKA-SPENGLER, T. 2010b. Genomic code for Sox10 activation reveals a key regulatory enhancer for cranial neural crest. *Proc Natl Acad Sci U S A,* 107**,** 3570-5.

BETANCUR, P., SAUKA-SPENGLER, T. & BRONNER, M. 2011. A Sox10 enhancer element common to the otic placode and neural crest is activated by tissue-specific paralogs. *Development,* 138**,** 3689-98.

BHATTACHARYYA, S., BAILEY, A. P., BRONNER-FRASER, M. & STREIT, A. 2004. Segregation of lens and olfactory precursors from a common territory: cell sorting and reciprocity of Dlx5 and Pax6 expression. *Dev Biol,* 271**,** 403-14.

BHATTACHARYYA, S. & BRONNER-FRASER, M. 2008. Competence, specification and commitment to an olfactory placode fate. *Development,* 135**,** 4165-77.

BHATTACHARYYA, S. & BRONNER, M. E. 2013. Clonal analyses in the anterior pre-placodal region: implications for the early lineage bias of placodal progenitors. *Int J Dev Biol,* 57**,** 753-7.

BI, X. & BROACH, J. R. 2001. Chromosomal boundaries in S. cerevisiae. *Curr Opin Genet Dev,* 11**,** 199-204.

BILIONI, A., CRAIG, G., HILL, C. & MCNEILL, H. 2005. Iroquois transcription factors recognize a unique motif to mediate transcriptional repression in vivo. *Proc Natl Acad Sci U S A,* 102**,** 14671-6.

BINDEA, G., MLECNIK, B., HACKL, H., CHAROENTONG, P., TOSOLINI, M., KIRILOVSKY, A., FRIDMAN, W. H., PAGES, F., TRAJANOSKI, Z. & GALON, J. 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics,* 25**,** 1091-3.

BLAHNIK, K. R., DOU, L., O'GEEN, H., MCPHILLIPS, T., XU, X., CAO, A. R., IYENGAR, S., NICOLET, C. M., LUDASCHER, B., KORF, I. & FARNHAM, P. J. 2010. Sole-Search: an integrated analysis program for peak detection and functional annotation using ChIP-seq data. *Nucleic Acids Res,* 38**,** e13.

BLANCHETTE, M., SCHWIKOWSKI, B. & TOMPA, M. 2002. Algorithms for phylogenetic footprinting. *J Comput Biol,* 9**,** 211-23.

BLANCHETTE, M. & TOMPA, M. 2003. FootPrinter: A program designed for phylogenetic footprinting. *Nucleic Acids Res,* 31**,** 3840-2.

BLENTIC, A., TANDON, P., PAYTON, S., WALSHE, J., CARNEY, T., KELSH, R. N., MASON, I. & GRAHAM, A. 2008. The emergence of ectomesenchyme. *Developmental Dynamics,* 237**,** 592-601.

BLOW, M. J., MCCULLEY, D. J., LI, Z., ZHANG, T., AKIYAMA, J. A., HOLT, A., PLAJZER-FRICK, I., SHOUKRY, M., WRIGHT, C., CHEN, F., AFZAL, V., BRISTOW, J., REN, B., BLACK, B. L., RUBIN, E. M., VISEL, A. & PENNACCHIO, L. A. 2010. ChIP-Seq identification of weakly conserved heart enhancers. *Nat Genet,* 42**,** 806-10.

BLUM, R., VETHANTHAM, V., BOWMAN, C., RUDNICKI, M. & DYNLACHT, B. D. 2012. Genome-wide identification of enhancers in skeletal muscle: the role of MyoD1. *Genes Dev,* 26**,** 2763-79.

BOK, J., BRONNER-FRASER, M. & WU, D. K. 2005. Role of the hindbrain in dorsoventral but not anteroposterior axial specification of the inner ear. *Development,* 132**,** 2115-24.

BOK, J., RAFT, S., KONG, K. A., KOO, S. K., DRAGER, U. C. & WU, D. K. 2011. Transient retinoic acid signaling confers anterior-posterior polarity to the inner ear. *Proc Natl Acad Sci U S A,* 108**,** 161-6.

BONN, S., ZINZEN, R. P., GIRARDOT, C., GUSTAFSON, E. H., PEREZ-GONZALEZ, A., DELHOMME, N., GHAVI-HELM, Y., WILCZYNSKI, B., RIDDELL, A. & FURLONG, E. E. 2012. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet,* 44**,** 148-56.

BOUCHARD, M., DE CAPRONA, D., BUSSLINGER, M., XU, P. & FRITZSCH, B. 2010. Pax2 and Pax8 cooperate in mouse inner ear morphogenesis and innervation. *BMC Dev Biol,* 10**,** 89.

BOYLE, A. P., DAVIS, S., SHULHA, H. P., MELTZER, P., MARGULIES, E. H., WENG, Z., FUREY, T. S. & CRAWFORD, G. E. 2008a. High-resolution mapping and characterization of open chromatin across the genome. *Cell,* 132**,** 311-22.

BOYLE, A. P., GUINNEY, J., CRAWFORD, G. E. & FUREY, T. S. 2008b. F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics,* 24**,** 2537-8.

BRICAUD, O. & COLLAZO, A. 2006. The transcription factor six1 inhibits neuronal and promotes hair cell fate in the developing zebrafish (Danio rerio) inner ear. *J Neurosci,* 26**,** 10438-51.

BROCCOLI, V., BONCINELLI, E. & WURST, W. 1999. The caudal limit of Otx2 expression positions the isthmic organizer. *Nature,* 401**,** 164-8.

BROWN, S. T., WANG, J. & GROVES, A. K. 2005. Dlx gene expression during chick inner ear development. *J Comp Neurol,* 483**,** 48-65.

BRUDNO, M., DO, C. B., COOPER, G. M., KIM, M. F., DAVYDOV, E., PROGRAM, N. C. S., GREEN, E. D., SIDOW, A. & BATZOGLOU, S. 2003a. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res,* 13**,** 721-31.

BRUDNO, M., MALDE, S., POLIAKOV, A., DO, C. B., COURONNE, O., DUBCHAK, I. & BATZOGLOU, S. 2003b. Glocal alignment: finding rearrangements during alignment. *Bioinformatics,* 19 Suppl 1**,** i54-62.

BRUGMANN, S. A., PANDUR, P. D., KENYON, K. L., PIGNONI, F. & MOODY, S. A. 2004. Six1 promotes a placodal fate within the lateral neurogenic ectoderm by functioning as both a transcriptional activator and repressor. *Development,* 131**,** 5871-81.

BUECKER, C. & WYSOCKA, J. 2012. Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet,* 28**,** 276-84.

BUENROSTRO, J. D., WU, B., CHANG, H. Y. & GREENLEAF, W. J. 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol,* 109**,** 21 29 1-9.

BULGER, M. & GROUDINE, M. 2011. Functional and mechanistic diversity of distal transcription enhancers. *Cell,* 144**,** 327-39.

BURTON, Q., COLE, L. K., MULHEISEN, M., CHANG, W. & WU, D. K. 2004. The role of Pax2 in mouse inner ear development. *Developmental Biology,* 272**,** 161-175.

BURZYNSKI, G. M., REED, X., TAHER, L., STINE, Z. E., MATSUI, T., OVCHARENKO, I. & MCCALLION, A. S. 2012. Systematic elucidation and in

vivo validation of sequences enriched in hindbrain transcriptional control. *Genome Res,* 22**,** 2278-89.

BUSSER, B. W., TAHER, L., KIM, Y., TANSEY, T., BLOOM, M. J., OVCHARENKO, I. & MICHELSON, A. M. 2012. A machine learning approach for identifying novel cell type-specific transcriptional regulators of myogenesis. *PLoS Genet,* 8**,** e1002531.

CALIFANO, A. 2000. SPLASH: structural pattern localization analysis by sequential histograms. *Bioinformatics,* 16**,** 341-57.

CALO, E. & WYSOCKA, J. 2013. Modification of enhancer chromatin: what, how, and why? *Mol Cell,* 49**,** 825-37.

CANNING, C. A., LEE, L., LUO, S. X., GRAHAM, A. & JONES, C. M. 2008. Neural tube derived Wnt signals cooperate with FGF signaling in the formation and differentiation of the trigeminal placodes. *Neural Dev,* 3**,** 35.

CARTHARIUS, K., FRECH, K., GROTE, K., KLOCKE, B., HALTMEIER, M., KLINGENHOFF, A., FRISCH, M., BAYERLEIN, M. & WERNER, T. 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics,* 21**,** 2933-42.

CARVALHO, L. R., WOODS, K. S., MENDONCA, B. B., MARCAL, N., ZAMPARINI, A. L., STIFANI, S., BRICKMAN, J. M., ARNHOLD, I. J. & DATTANI, M. T. 2003. A homozygous mutation in HESX1 is associated with evolving hypopituitarism due to impaired repressor-corepressor interaction. *J Clin Invest,* 112**,** 1192-201.

CASTRO, L. F., RASMUSSEN, S. L., HOLLAND, P. W., HOLLAND, N. D. & HOLLAND, L. Z. 2006. A Gbx homeobox gene in amphioxus: insights into ancestry of the ANTP class and evolution of the midbrain/hindbrain boundary. *Dev Biol,* 295**,** 40-51.

CATENA, R., TIVERON, C., RONCHI, A., PORTA, S., FERRI, A., TATANGELO, L., CAVALLARO, M., FAVARO, R., OTTOLENGHI, S., REINBOLD, R., SCHOLER, H. & NICOLIS, S. K. 2004. Conserved POU binding DNA sites in the Sox2 upstream enhancer regulate gene expression in embryonic and neural stem cells. *J Biol Chem,* 279**,** 41846-57.

CATRON, K. M., ZHANG, H., MARSHALL, S. C., INOSTROZA, J. A., WILSON, J. M. & ABATE, C. 1995. Transcriptional repression by Msx-1 does not require homeodomain DNA-binding sites. *Mol Cell Biol,* 15**,** 861-71.

CHAMBERS, D. & MASON, I. 2000. Expression of sprouty2 during early development of the chick embryo is coincident with known sites of FGF signalling. *Mech Dev,* 91**,** 361-4.

CHAMBERS, D. & MASON, I. 2006. A high throughput messenger RNA differential display screen identifies discrete domains of gene expression and novel patterning processes along the developing neural tube. *BMC Dev Biol,* 6**,** 9.

CHAPMAN, S. C., COLLIGNON, J., SCHOENWOLF, G. C. & LUMSDEN, A. 2001. Improved method for chick whole-embryo culture using a filter paper carrier. *Developmental Dynamics,* 220**,** 284-289.

CHAPMAN, S. C., SCHUBERT, F. R., SCHOENWOLF, G. C. & LUMSDEN, A. 2002. Analysis of spatial and temporal gene expression patterns in blastula and gastrula stage chick embryos. *Dev Biol,* 245**,** 187-99.

CHATTERJEE, R., RAMOS, E., HOFFMAN, M., VANWINKLE, J., MARTIN, D. R., DAVIS, T. K., HOSHI, M., HMIEL, S. P., BECK, A., HRUSKA, K., COPLEN, D., LIAPIS, H., MITRA, R., DRULEY, T., AUSTIN, P. & JAIN, S. 2012. Traditional and targeted exome sequencing reveals common, rare and novel functional deleterious variants in RET-signaling complex in a cohort of living US patients with urinary tract malformations. *Hum Genet,* 131**,** 1725-38.

CHEN, L. C., CHEN, B. K., CHANG, J. M. & CHANG, W. C. 2004. Essential role of c-Jun induction and coactivator p300 in epidermal growth factor-induced gene expression of cyclooxygenase-2 in human epidermoid carcinoma A431 cells. *Biochim Biophys Acta,* 1683**,** 38-48.

CHI, X., CHATTERJEE, P. K., WILSON, W., 3RD, ZHANG, S. X., DEMAYO, F. J. & SCHWARTZ, R. J. 2005. Complex cardiac Nkx2-5 gene expression activated by noggin-sensitive enhancers followed by chamber-specific modules. *Proc Natl Acad Sci U S A,* 102**,** 13490-5.

CHIANG, D. Y., MOSES, A. M., KELLIS, M., LANDER, E. S. & EISEN, M. B. 2003. Phylogenetically and spatially conserved word pairs associated with gene-expression changes in yeasts. *Genome Biol,* 4**,** R43.

CHOW, R. L., ALTMANN, C. R., LANG, R. A. & HEMMATI-BRIVANLOU, A. 1999. Pax6 induces ectopic eyes in a vertebrate. *Development,* 126**,** 4213-22.

CHRISTOPHOROU, N. 2008. The role of Eya2 and Six1 in early placode development. *A thesis submitted to the King's College London's Higher Degree Office in partial fulfilment for the Degree of Doctor of Philosophy*.

CHRISTOPHOROU, N. A., BAILEY, A. P., HANSON, S. & STREIT, A. 2009. Activation of Six1 target genes is required for sensory placode formation. *Dev Biol,* 336**,** 327-36.

CHRISTOPHOROU, N. A., MENDE, M., LLERAS-FORERO, L., GROCOTT, T. & STREIT, A. 2010. Pax2 coordinates epithelial morphogenesis and cell fate in the inner ear. *Dev Biol,* 345**,** 180-90.

CHUNG, J. H., WHITELEY, M. & FELSENFELD, G. 1993. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell,* 74**,** 505-14.

CLARKE, S. L., VANDERMEER, J. E., WENGER, A. M., SCHAAR, B. T., AHITUV, N. & BEJERANO, G. 2012. Human developmental enhancers conserved between deuterostomes and protostomes. *PLoS Genet,* 8**,** e1002852.

CONSORTIUM, E. P. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

COOPER, G. M., BRUDNO, M., STONE, E. A., DUBCHAK, I., BATZOGLOU, S. & SIDOW, A. 2004. Characterization of evolutionary rates and constraints in three Mammalian genomes. *Genome Res,* 14**,** 539-48.

COTNEY, J., LENG, J., OH, S., DEMARE, L. E., REILLY, S. K., GERSTEIN, M. B. & NOONAN, J. P. 2012. Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. *Genome Res,* 22**,** 1069-80.

COULY, G. F. & LE DOUARIN, N. M. 1985. Mapping of the early neural primordium in quail-chick chimeras. I. Developmental relationships between placodes, facial ectoderm, and prosencephalon. *Dev Biol,* 110**,** 422-39.

CREYGHTON, M. P., CHENG, A. W., WELSTEAD, G. G., KOOISTRA, T., CAREY, B. W., STEINE, E. J., HANNA, J., LODATO, M. A., FRAMPTON, G. M., SHARP, P. A., BOYER, L. A., YOUNG, R. A. & JAENISCH, R. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A,* 107**,** 21931-6.

CROSSLEY, P. H. & MARTIN, G. R. 1995. The mouse Fgf8 gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo. *Development,* 121**,** 439-51.

CUDDAPAH, S., JOTHI, R., SCHONES, D. E., ROH, T. Y., CUI, K. & ZHAO, K. 2009. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res,* 19**,** 24-32.

CUI, J., MICHAILLE, J. J., JIANG, W. & ZILE, M. H. 2003. Retinoid receptors and vitamin A deficiency: differential patterns of transcription during early avian development and the rapid induction of RARs by retinoic acid. *Dev Biol,* 260**,** 496-511.

D'AMICO-MARTEL, A. 1982. Temporal patterns of neurogenesis in avian cranial sensory and autonomic ganglia. *Am J Anat,* 163**,** 351-72.

D'AMICO-MARTEL, A. & NODEN, D. M. 1980. An autoradiographic analysis of the development of the chick trigeminal ganglion. *J Embryol Exp Morphol,* 55**,** 167-82.

D'HAESELEER, P., LIANG, S. & SOMOGYI, R. 2000. Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics,* 16**,** 707-26.

DADY, A., BLAVET, C. & DUBAND, J. L. 2012. Timing and kinetics of E- to N-cadherin switch during neurulation in the avian embryo. *Dev Dyn,* 241**,** 1333-49.

DAVIDSON, E. H., RAST, J. P., OLIVERI, P., RANSICK, A., CALESTANI, C., YUH, C. H., MINOKAWA, T., AMORE, G., HINMAN, V., ARENAS-MENA, C., OTIM, O., BROWN, C. T., LIVI, C. B., LEE, P. Y., REVILLA, R., RUST, A. G., PAN, Z., SCHILSTRA, M. J., CLARKE, P. J., ARNONE, M. I., ROWEN, L., CAMERON, R. A., MCCLAY, D. R., HOOD, L. & BOLOURI, H. 2002. A genomic regulatory network for development. *Science,* 295**,** 1669-78.

DE HOON, M. J., IMOTO, S., KOBAYASHI, K., OGASAWARA, N. & MIYANO, S. 2003. Inferring gene regulatory networks from time-ordered gene expression data of Bacillus subtilis using differential equations. *Pac Symp Biocomput***,** 17-28.

DECHIARA, T. M., ROBERTSON, E. J. & EFSTRATIADIS, A. 1991. Parental imprinting of the mouse insulin-like growth factor II gene. *Cell,* 64**,** 849-59.

DEFOSSEZ, P. A. & GILSON, E. 2002. The vertebrate protein CTCF functions as an insulator in Saccharomyces cerevisiae. *Nucleic Acids Res,* 30**,** 5136-41.

DEITCHER, D. L., FEKETE, D. M. & CEPKO, C. L. 1994. Asymmetric expression of a novel homeobox gene in vertebrate sensory organs. *J Neurosci,* 14**,** 486-98.

DEPLANCKE, B., MUKHOPADHYAY, A., AO, W., ELEWA, A. M., GROVE, C. A., MARTINEZ, N. J., SEQUERRA, R., DOUCETTE-STAMM, L., REECE-HOYES, J. S., HOPE, I. A., TISSENBAUM, H. A., MANGO, S. E. & WALHOUT, A. J. 2006. A gene-centered C. elegans protein-DNA interaction network. *Cell,* 125**,** 1193-205.

DHILLON, N. & KAMAKAKA, R. T. 2002. Breaking through to the other side: silencers and barriers. *Curr Opin Genet Dev,* 12**,** 188-92.

DOETZLHOFER, A., BASCH, M. L., OHYAMA, T., GESSLER, M., GROVES, A. K. & SEGIL, N. 2009. Hey2 regulation by FGF provides a Notch-independent mechanism for maintaining pillar cell fate in the organ of Corti. *Dev Cell,* 16**,** 58-69.

DONOHOE, M. E., ZHANG, L. F., XU, N., SHI, Y. & LEE, J. T. 2007. Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. *Mol Cell,* 25**,** 43-56.

DONZE, D. & KAMAKAKA, R. T. 2002. Braking the silence: how heterochromatic gene repression is stopped in its tracks. *Bioessays,* 24**,** 344-9.

DOREY, K. & AMAYA, E. 2010. FGF signalling: diverse roles during early vertebrate embryogenesis. *Development,* 137**,** 3731-42.

DUDE, C. M., KUAN, C. Y., BRADSHAW, J. R., GREENE, N. D., RELAIX, F., STARK, M. R. & BAKER, C. V. 2009. Activation of Pax3 target genes is necessary but not sufficient for neurogenesis in the ophthalmic trigeminal placode. *Dev Biol,* 326**,** 314-26.

EAGLESON, G., FERREIRO, B. & HARRIS, W. A. 1995. Fate of the anterior neural ridge and the morphogenesis of the Xenopus forebrain. *J Neurobiol,* 28**,** 146-58.

EMISON, E. S., MCCALLION, A. S., KASHUK, C. S., BUSH, R. T., GRICE, E., LIN, S., PORTNOY, M. E., CUTLER, D. J., GREEN, E. D. & CHAKRAVARTI, A. 2005. A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature,* 434**,** 857-63.

ENGEL, N., THORVALDSEN, J. L. & BARTOLOMEI, M. S. 2006. CTCF binding sites promote transcription initiation and prevent DNA methylation on the maternal allele at the imprinted H19/Igf2 locus. *Hum Mol Genet,* 15**,** 2945-54.

ERNST, J., KHERADPOUR, P., MIKKELSEN, T. S., SHORESH, N., WARD, L. D., EPSTEIN, C. B., ZHANG, X., WANG, L., ISSNER, R., COYNE, M., KU, M., DURHAM, T., KELLIS, M. & BERNSTEIN, B. E. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature,* 473**,** 43-9.

EROKHIN, M., VASSETZKY, Y., GEORGIEV, P. & CHETVERINA, D. 2015. Eukaryotic enhancers: common features, regulation, and participation in diseases. *Cell Mol Life Sci,* 72**,** 2361-75.

ERWIN, G. D., OKSENBERG, N., TRUTY, R. M., KOSTKA, D., MURPHY, K. K., AHITUV, N., POLLARD, K. S. & CAPRA, J. A. 2014. Integrating diverse datasets improves developmental enhancer prediction. *PLoS Comput Biol,* 10**,** e1003677.

ESKIN, E. & PEVZNER, P. A. 2002. Finding composite regulatory patterns in DNA sequences. *Bioinformatics,* 18 Suppl 1**,** S354-63.

ESTERBERG, R. & FRITZ, A. 2009. dlx3b/4b are required for the formation of the preplacodal region and otic placode through local modulation of BMP activity. *Dev Biol,* 325**,** 189-99.

ESTEVE, P. & BOVOLENTA, P. 1999. cSix4, a member of the six gene family of transcription factors, is expressed during placode and somite development. *Mech Dev,* 85**,** 161-5.

FAITH, J. J., HAYETE, B., THADEN, J. T., MOGNO, I., WIERZBOWSKI, J., COTTAREL, G., KASIF, S., COLLINS, J. J. & GARDNER, T. S. 2007. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *Plos Biology,* 5**,** 54-66.

FAN, Y., LINARDOPOULOU, E., FRIEDMAN, C., WILLIAMS, E. & TRASK, B. J. 2002. Genomic structure and evolution of the ancestral chromosome fusion site in 2q13-2q14.1 and paralogous regions on other human chromosomes. *Genome Res,* 12**,** 1651-62.

FAVOR, J., SANDULACHE, R., NEUHAUSER-KLAUS, A., PRETSCH, W., CHATTERJEE, B., SENFT, E., WURST, W., BLANQUET, V., GRIMES, P., SPORLE, R. & SCHUGHART, K. 1996. The mouse Pax2(1Neu) mutation is identical to a human PAX2 mutation in a family with renal-coloboma syndrome and results in developmental defects of the brain, ear, eye, and kidney. *Proc Natl Acad Sci U S A,* 93**,** 13870-5.

FELSENFELD, G., BURGESS-BEUSSE, B., FARRELL, C., GASZNER, M., GHIRLANDO, R., HUANG, S., JIN, C., LITT, M., MAGDINIER, F., MUTSKOV, V., NAKATANI, Y., TAGAMI, H., WEST, A. & YUSUFZAI, T. 2004. Chromatin boundaries and chromatin domains. *Cold Spring Harb Symp Quant Biol,* 69**,** 245-50.

FILION, G. J., VAN BEMMEL, J. G., BRAUNSCHWEIG, U., TALHOUT, W., KIND, J., WARD, L. D., BRUGMAN, W., DE CASTRO, I. J., KERKHOVEN, R. M., BUSSEMAKER, H. J. & VAN STEENSEL, B. 2010. Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell,* 143**,** 212-24.

FILIPPAKOPOULOS, P. & KNAPP, S. 2012. The bromodomain interaction module. *FEBS Lett,* 586**,** 2692-704.

FILIPPAKOPOULOS, P., PICAUD, S., MANGOS, M., KEATES, T., LAMBERT, J. P., BARSYTE-LOVEJOY, D., FELLETAR, I., VOLKMER, R., MULLER, S., PAWSON, T., GINGRAS, A. C., ARROWSMITH, C. H. & KNAPP, S. 2012. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell,* 149**,** 214-31.

FILIPPOVA, G. N., FAGERLIE, S., KLENOVA, E. M., MYERS, C., DEHNER, Y., GOODWIN, G., NEIMAN, P. E., COLLINS, S. J. & LOBANENKOV, V. V. 1996. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol Cell Biol,* 16**,** 2802-13.

FLETCHER, R. B., BAKER, J. C. & HARLAND, R. M. 2006. FGF8 spliceforms mediate early mesoderm and posterior neural tissue formation in Xenopus. *Development,* 133**,** 1703-14.

FORGE, A. & WRIGHT, T. 2002. The molecular architecture of the inner ear. *Br Med Bull,* 63**,** 5-24.

FRAZER, K. A., PACHTER, L., POLIAKOV, A., RUBIN, E. M. & DUBCHAK, I. 2004. VISTA: computational tools for comparative genomics. *Nucleic Acids Res,* 32**,** W273-9.

FRETER, S., MUTA, Y., MAK, S. S., RINKWITZ, S. & LADHER, R. K. 2008. Progressive restriction of otic fate: the role of FGF and Wnt in resolving inner ear potential. *Development,* 135**,** 3415-24.

FRETER, S., MUTA, Y., O'NEILL, P., VASSILEV, V. S., KURAKU, S. & LADHER, R. K. 2012. Pax2 modulates proliferation during specification of the otic and epibranchial placodes. *Dev Dyn,* 241**,** 1716-28.

FRIEDMAN, R. A., MAKMURA, L., BIESIADA, E., WANG, X. & KEITHLEY, E. M. 2005. Eya1 acts upstream of Tbx1, Neurogenin 1, NeuroD and the neurotrophins BDNF and NT-3 during inner ear development. *Mech Dev,* 122**,** 625-34.

FRITH, M. C., FU, Y., YU, L., CHEN, J. F., HANSEN, U. & WENG, Z. 2004a. Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Res,* 32**,** 1372-81.

FRITH, M. C., HANSEN, U., SPOUGE, J. L. & WENG, Z. 2004b. Finding functional sequence elements by multiple local alignment. *Nucleic Acids Res,* 32**,** 189-200.

FUREY, T. S. 2012. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet,* 13**,** 840-52.

FURUMATSU, T., TSUDA, M., TANIGUCHI, N., TAJIMA, Y. & ASAHARA, H. 2005a. Smad3 induces chondrogenesis through the activation of SOX9 via CREB-binding protein/p300 recruitment. *J Biol Chem,* 280**,** 8343-50.

FURUMATSU, T., TSUDA, M., YOSHIDA, K., TANIGUCHI, N., ITO, T., HASHIMOTO, M., ITO, T. & ASAHARA, H. 2005b. Sox9 and p300 cooperatively regulate chromatin-mediated transcription. *J Biol Chem,* 280**,** 35203-8.

GARDNER, T. S. & FAITH, J. J. 2005. Reverse-engineering transcription control networks. *Phys Life Rev,* 2**,** 65-88.

GEISS, G. K., BUMGARNER, R. E., BIRDITT, B., DAHL, T., DOWIDAR, N., DUNAWAY, D. L., FELL, H. P., FERREE, S., GEORGE, R. D., GROGAN, T., JAMES, J. J., MAYSURIA, M., MITTON, J. D., OLIVERI, P., OSBORN, J. L., PENG, T., RATCLIFFE, A. L., WEBSTER, P. J., DAVIDSON, E. H., HOOD, L. & DIMITROV, K. 2008. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol,* 26**,** 317-25.

GERASIMOVA, T. I. & CORCES, V. G. 1996. Boundary and insulator elements in chromosomes. *Curr Opin Genet Dev,* 6**,** 185-92.

GERASIMOVA, T. I. & CORCES, V. G. 2001. Chromatin insulators and boundaries: effects on transcription and nuclear organization. *Annu Rev Genet,* 35**,** 193-208.

GHANBARI, H., SEO, H. C., FJOSE, A. & BRANDLI, A. W. 2001. Molecular cloning and embryonic expression of Xenopus Six homeobox genes. *Mech Dev,* 101**,** 271-7.

GHIRLANDO, R., GILES, K., GOWHER, H., XIAO, T., XU, Z., YAO, H. & FELSENFELD, G. 2012. Chromatin domains, insulators, and the regulation of gene expression. *Biochim Biophys Acta,* 1819**,** 644-51.

GHISLETTI, S., BAROZZI, I., MIETTON, F., POLLETTI, S., DE SANTA, F., VENTURINI, E., GREGORY, L., LONIE, L., CHEW, A., WEI, C. L., RAGOUSSIS, J. & NATOLI, G. 2010. Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity,* 32**,** 317-28.

GHIURCUTA, C. G. & MORET, B. M. 2014. Evaluating synteny for improved comparative studies. *Bioinformatics,* 30**,** i9-18.

GILES, K. E., GOWHER, H., GHIRLANDO, R., JIN, C. & FELSENFELD, G. 2010. Chromatin boundaries, insulators, and long-range interactions in the nucleus. *Cold Spring Harb Symp Quant Biol,* 75**,** 79-85.

GINIGER, E., VARNUM, S. M. & PTASHNE, M. 1985. Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell,* 40**,** 767-74.

GIRESI, P. G., KIM, J., MCDANIELL, R. M., IYER, V. R. & LIEB, J. D. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res,* 17**,** 877-85.

GLAVIC, A., MARIS HONORE, S., GLORIA FEIJOO, C., BASTIDAS, F., ALLENDE, M. L. & MAYOR, R. 2004. Role of BMP signaling and the homeoprotein Iroquois in the specification of the cranial placodal field. *Dev Biol,* 272**,** 89-103.

GLOVER, J. N. & HARRISON, S. C. 1995. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature,* 373**,** 257-61.

GONG, W., KOYANO-NAKAGAWA, N., LI, T. & GARRY, D. J. 2015. Inferring dynamic gene regulatory networks in cardiac differentiation through the integration of multi-dimensional data. *BMC Bioinformatics,* 16**,** 74.

GOODMAN, R. H. & SMOLIK, S. 2000. CBP/p300 in cell growth, transformation, and development. *Genes Dev,* 14**,** 1553-77.

GORDON, C. T., TAN, T. Y., BENKO, S., FITZPATRICK, D., LYONNET, S. & & FARLIE, P. G. 2009. Long-range regulation at the SOX9 locus in development and disease. *J Med Genet,* 46**,** 649-656.

GORIELY, A., DIEZ DEL CORRAL, R. & STOREY, K. G. 1999. c-Irx2 expression reveals an early subdivision of the neural plate in the chick embryo. *Mech Dev,* 87**,** 203-6.

GORKIN, D. U., LEE, D., REED, X., FLETEZ-BRANT, C., BESSLING, S. L., LOFTUS, S. K., BEER, M. A., PAVAN, W. J. & MCCALLION, A. S. 2012. Integration of ChIP-seq and machine learning reveals enhancers and a predictive regulatory sequence vocabulary in melanocytes. *Genome Res,* 22**,** 2290-301.

GOTOH, O. 1996. Significant improvement in accuracy of multiple protein sequence alignments by iterative refinement as assessed by reference to structural alignments. *J Mol Biol,* 264**,** 823-38.

GOTTGENS, B., BARTON, L. M., GILBERT, J. G., BENCH, A. J., SANCHEZ, M. J., BAHN, S., MISTRY, S., GRAFHAM, D., MCMURRAY, A., VAUDIN, M., AMAYA, E., BENTLEY, D. R., GREEN, A. R. & SINCLAIR, A. M. 2000. Analysis of vertebrate SCL loci identifies conserved enhancers. *Nat Biotechnol,* 18**,** 181-6.

GROCOTT, T., JOHNSON, S., BAILEY, A. P. & STREIT, A. 2011. Neural crest cells organize the eye via TGF-beta and canonical Wnt signalling. *Nat Commun,* 2**,** 265.

GRUDA, M. C., KOVARY, K., METZ, R. & BRAVO, R. 1994. Regulation of Fra-1 and Fra-2 phosphorylation differs during the cell cycle of fibroblasts and phosphorylation in vitro by MAP kinase affects DNA binding activity. *Oncogene,* 9**,** 2537-47.

GUELEN, L., PAGIE, L., BRASSET, E., MEULEMAN, W., FAZA, M. B., TALHOUT, W., EUSSEN, B. H., DE KLEIN, A., WESSELS, L., DE LAAT, W. & VAN STEENSEL, B. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature,* 453**,** 948-51.

HACHE, H., LEHRACH, H. & HERWIG, R. 2009. Reverse engineering of gene regulatory networks: a comparative study. *EURASIP J Bioinform Syst Biol***,** 617281.

HAFNER, F. M., SALAM, A. A., LINDER, T. E., BALMER, D., BAUMER, A., SCHINZEL, A. A., SPILLMANN, T. & LEAL, S. M. 2000. A novel locus (DFNA24) for prelingual nonprogressive autosomal dominant nonsyndromic hearing loss maps to 4q35-qter in a large Swiss German kindred. *Am J Hum Genet,* 66**,** 1437-42.

HAMBURGER, V. & HAMILTON, H. L. 1951. A series of normal stages in the development of the chick embryo. *J Morphol,* 88**,** 49-92.

HANS, S., CHRISTISON, J., LIU, D. & WESTERFIELD, M. 2007. Fgf-dependent otic induction requires competence provided by Foxi1 and Dlx3b. *BMC Dev Biol,* 7**,** 5.

HANS, S., LIU, D. & WESTERFIELD, M. 2004. Pax8 and Pax2a function synergistically in otic specification, downstream of the Foxi1 and Dlx3b transcription factors. *Development,* 131**,** 5091-102.

HARDING, K., WEDEEN, C., MCGINNIS, W. & LEVINE, M. 1985. Spatially regulated expression of homeotic genes in Drosophila. *Science,* 229**,** 1236-42.

HARDISON, R. C. & TAYLOR, J. 2012. Genomic approaches towards finding cis-regulatory modules in animals. *Nat Rev Genet,* 13**,** 469-83.

HARK, A. T., SCHOENHERR, C. J., KATZ, D. J., INGRAM, R. S., LEVORSE, J. M. & TILGHMAN, S. M. 2000. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature,* 405**,** 486-9.

HARRISON, M. M., LI, X. Y., KAPLAN, T., BOTCHAN, M. R. & EISEN, M. B. 2011. Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet,* 7**,** e1002266.

HASSAN, M. J., SANTOS, R. L., RAFIQ, M. A., CHAHROUR, M. H., PHAM, T. L., WAJID, M., HIJAB, N., WAMBANGCO, M., LEE, K., ANSAR, M., YAN, K., AHMAD, W. & LEAL, S. M. 2006. A novel autosomal recessive non-syndromic hearing impairment locus (DFNB47) maps to chromosome 2p25.1-p24.3. *Hum Genet,* 118**,** 605-10.

HATADA, Y. & STERN, C. D. 1994. A fate map of the epiblast of the early chick embryo. *Development,* 120**,** 2879-89.

HEINTZMAN, N. D., HON, G. C., HAWKINS, R. D., KHERADPOUR, P., STARK, A., HARP, L. F., YE, Z., LEE, L. K., STUART, R. K., CHING, C. W., CHING, K. A., ANTOSIEWICZ-BOURGET, J. E., LIU, H., ZHANG, X., GREEN, R. D., LOBANENKOV, V. V., STEWART, R., THOMSON, J. A., CRAWFORD, G. E., KELLIS, M. & REN, B. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature,* 459**,** 108-12.

HEINTZMAN, N. D., STUART, R. K., HON, G., FU, Y., CHING, C. W., HAWKINS, R. D., BARRERA, L. O., VAN CALCAR, S., QU, C., CHING, K. A., WANG,

W., WENG, Z., GREEN, R. D., CRAWFORD, G. E. & REN, B. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet,* 39**,** 311-8.

HEINZ, S., BENNER, C., SPANN, N., BERTOLINO, E., LIN, Y. C., LASLO, P., CHENG, J. X., MURRE, C., SINGH, H. & GLASS, C. K. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell,* 38**,** 576-89.

HELLER, N. & BRANDLI, A. W. 1999. Xenopus Pax-2/5/8 orthologues: novel insights into Pax gene evolution and identification of Pax-8 as the earliest marker for otic and pronephric cell lineages. *Dev Genet,* 24**,** 208-19.

HEROLD, M., BARTKUHN, M. & RENKAWITZ, R. 2012. CTCF: insights into insulator function during development. *Development,* 139**,** 1045-57.

HERTZ, G. Z. & STORMO, G. D. 1999. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics,* 15**,** 563-77.

HIDALGO-SANCHEZ, M., MILLET, S., SIMEONE, A. & ALVARADO-MALLART, R. M. 1999. Comparative analysis of Otx2, Gbx2, Pax2, Fgf8 and Wnt1 gene expressions during the formation of the chick midbrain/hindbrain domain. *Mech Dev,* 81**,** 175-8.

HINMAN, V. F., NGUYEN, A. T., CAMERON, R. A. & DAVIDSON, E. H. 2003. Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. *Proc Natl Acad Sci U S A,* 100**,** 13356-61.

HOHL, M., KURTZ, S. & OHLEBUSCH, E. 2002. Efficient multiple genome alignment. *Bioinformatics,* 18 Suppl 1**,** S312-20.

HOLMQVIST, P. H. & MANNERVIK, M. 2013. Genomic occupancy of the transcriptional co-activators p300 and CBP. *Transcription,* 4**,** 18-23.

HOLZSCHUH, J., WADA, N., WADA, C., SCHAFFER, A., JAVIDAN, Y., TALLAFUSS, A., BALLY-CUIF, L. & SCHILLING, T. F. 2005. Requirements for endoderm and BMP signaling in sensory neurogenesis in zebrafish. *Development,* 132**,** 3731-42.

HON, G., REN, B. & WANG, W. 2008. ChromaSig: a probabilistic approach to finding common chromatin signatures in the human genome. *PLoS Comput Biol,* 4**,** e1000201.

HONG, C. S. & SAINT-JEANNET, J. P. 2007. The activity of Pax3 and Zic1 regulates three distinct cell fates at the neural plate border. *Mol Biol Cell,* 18**,** 2192-202.

HOOD, L., ROWEN, L. & KOOP, B. F. 1995. Human and mouse T-cell receptor loci: genomics, evolution, diversity, and serendipity. *Ann N Y Acad Sci,* 758**,** 390-412.

HOU, C., DALE, R. & DEAN, A. 2010. Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc Natl Acad Sci U S A,* 107**,** 3651-6.

HOU, C., ZHAO, H., TANIMOTO, K. & DEAN, A. 2008. CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc Natl Acad Sci U S A,* 105**,** 20398-403.

HU, J., LI, B. & KIHARA, D. 2005. Limitations and potentials of current motif discovery algorithms. *Nucleic Acids Res,* 33**,** 4899-913.

HUANG, D. & OVCHARENKO, I. 2014. Genome-Wide Analysis of Functional and Evolutionary Features of Tele-Enhancers. *G3-Genes Genomes Genetics,* 4**,** 579-593.

HUANG, D. W., SHERMAN, B. T. & LEMPICKI, R. A. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research,* 37**,** 1-13.

HUANG, D. W., SHERMAN, B. T. & LEMPICKI, R. A. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols,* 4**,** 44-57.

HUANG, Y., TIENDA-LUNA, I. M. & WANG, Y. 2009c. A Survey of Statistical Models for Reverse Engineering Gene Regulatory Networks. *IEEE Signal Process Mag,* 26**,** 76-97.

HUGHES, J. D., ESTEP, P. W., TAVAZOIE, S. & CHURCH, G. M. 2000. Computational identification of cis-regulatory elements associated with groups of functionally related genes in Saccharomyces cerevisiae. *J Mol Biol,* 296**,** 1205-14.

HURD, T. W., CULBERT, A. A., WEBSTER, K. J. & TAVARE, J. M. 2002. Dual role for mitogen-activated protein kinase (Erk) in insulin-dependent regulation of Fra-1 (fos-related antigen-1) transcription and phosphorylation. *Biochem J,* 368**,** 573-80.

HUTSON, M. R., LEWIS, J. E., NGUYEN-LUU, D., LINDBERG, K. H. & BARALD, K. F. 1999. Expression of Pax2 and patterning of the chick inner ear. *J Neurocytol,* 28**,** 795-807.

HUYNH-THU, V. A., IRRTHUM, A., WEHENKEL, L. & GEURTS, P. 2010. Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. *Plos One,* 5.

HWANG, C. H., SIMEONE, A., LAI, E. & WU, D. K. 2009. Foxg1 is required for proper separation and formation of sensory cristae during inner ear development. *Dev Dyn,* 238**,** 2725-34.

IKEDA, K., OOKAWARA, S., SATO, S., ANDO, Z., KAGEYAMA, R. & KAWAKAMI, K. 2007. Six1 is essential for early neurogenesis in the development of olfactory epithelium. *Dev Biol,* 311**,** 53-68.

ISHIHARA, T., IKEDA, K., SATO, S., YAJIMA, H. & KAWAKAMI, K. 2008a. Differential expression of Eya1 and Eya2 during chick early embryonic development. *Gene Expr Patterns,* 8**,** 357-67.

ISHIHARA, T., SATO, S., IKEDA, K., YAJIMA, H. & KAWAKAMI, K. 2008b. Multiple evolutionarily conserved enhancers control expression of Eya1. *Dev Dyn,* 237**,** 3142-56.

JACOBSON, A. G. 1963a. The Determination and Positioning of the Nose, Lens and Ear. I. Interactions within the Ectoderm, and between the Ectoderm and Underlying Tissues. *J Exp Zool,* 154**,** 273-83.

JACOBSON, A. G. 1963b. The Determination and Positioning of the Nose, Lens and Ear. Ii. The Role of the Endoderm. *J Exp Zool,* 154**,** 285-91.

JACOBSON, A. G. 1963c. The Determination and Positioning of the Nose, Lens and Ear. Iii. Effects of Reversing the Antero-Posterior Axis of Epidermis, Neural Plate and Neural Fold. *J Exp Zool,* 154**,** 293-303.

JAYASENA, C. S., OHYAMA, T., SEGIL, N. & GROVES, A. K. 2008. Notch signaling augments the canonical Wnt pathway to specify the size of the otic placode. *Development,* 135**,** 2251-61.

JI, H., JIANG, H., MA, W., JOHNSON, D. S., MYERS, R. M. & WONG, W. H. 2008. An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol,* 26**,** 1293-300.

JIMENEZ, G., GALE, K. B. & ENVER, T. 1992. The mouse beta-globin locus control region: hypersensitive sites 3 and 4. *Nucleic Acids Res,* 20**,** 5797-803.

JIN, Q., YU, L. R., WANG, L., ZHANG, Z., KASPER, L. H., LEE, J. E., WANG, C., BRINDLE, P. K., DENT, S. Y. & GE, K. 2011. Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *EMBO J,* 30**,** 249-62.

JOHNSON, D. S., MORTAZAVI, A., MYERS, R. M. & WOLD, B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science,* 316**,** 1497-502.

JOHNSON, K. R., COOK, S. A., ERWAY, L. C., MATTHEWS, A. N., SANFORD, L. P., PARADIES, N. E. & FRIEDMAN, R. A. 1999. Inner ear and kidney anomalies caused by IAP insertion in an intron of the Eya1 gene in a mouse model of BOR syndrome. *Hum Mol Genet,* 8**,** 645-53.

JOLMA, A., YAN, J., WHITINGTON, T., TOIVONEN, J., NITTA, K. R., RASTAS, P., MORGUNOVA, E., ENGE, M., TAIPALE, M., WEI, G., PALIN, K., VAQUERIZAS, J. M., VINCENTELLI, R., LUSCOMBE, N. M., HUGHES, T. R., LEMAIRE, P., UKKONEN, E., KIVIOJA, T. & TAIPALE, J. 2013. DNA-binding specificities of human transcription factors. *Cell,* 152**,** 327-39.

JOSHI, A. 2014. Mammalian transcriptional hotspots are enriched for tissue specific enhancers near cell type specific highly expressed genes and are predicted to act as transcriptional activator hubs. *BMC Bioinformatics,* 15**,** 412.

JOTHI, R., CUDDAPAH, S., BARSKI, A., CUI, K. & ZHAO, K. 2008. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res,* 36**,** 5221-31.

JOYNER, A. L., LIU, A. & MILLET, S. 2000. Otx2, Gbx2 and Fgf8 interact to position and maintain a mid-hindbrain organizer. *Curr Opin Cell Biol,* 12**,** 736-41.

JUMLONGRAS, D., LACHKE, S. A., O'CONNELL, D. J., ABOUKHALIL, A., LI, X., CHOE, S. E., HO, J. W. K., TURBE-DOAN, A., ROBERTSON, E. A., OLSEN, B. R., BULYK, M. L., AMENDT, B. A. & MAAS, R. L. 2012. An Evolutionarily Conserved Enhancer Regulates Bmp4 Expression in Developing Incisor and Limb Bud. *Plos One,* 7.

KAGEY, M. H., NEWMAN, J. J., BILODEAU, S., ZHAN, Y., ORLANDO, D. A., VAN BERKUM, N. L., EBMEIER, C. C., GOOSSENS, J., RAHL, P. B., LEVINE, S. S., TAATJES, D. J., DEKKER, J. & YOUNG, R. A. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature,* 467**,** 430-5.

KANDURI, C., PANT, V., LOUKINOV, D., PUGACHEVA, E., QI, C. F., WOLFFE, A., OHLSSON, R. & LOBANENKOV, V. V. 2000. Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr Biol,* 10**,** 853-6.

KARABAGLI, H., KARABAGLI, P., LADHER, R. K. & SCHOENWOLF, G. C. 2002. Comparison of the expression patterns of several fibroblast growth factors during chick gastrulation and neurulation. *Anat Embryol (Berl),* 205**,** 365-70.

KARAMOUZIS, M. V., KONSTANTINOPOULOS, P. A. & PAPAVASSILIOU, A. G. 2007. Roles of CREB-binding protein (CBP)/p300 in respiratory epithelium tumorigenesis. *Cell Res,* 17**,** 324-32.

KATAHIRA, T., SATO, T., SUGIYAMA, S., OKAFUJI, T., ARAKI, I., FUNAHASHI, J. & NAKAMURA, H. 2000. Interaction between Otx2 and Gbx2 defines the organizing center for the optic tectum. *Mech Dev,* 91**,** 43-52.

KEE, Y., HWANG, B. J., STERNBERG, P. W. & BRONNER-FRASER, M. 2007. Evolutionary conservation of cell migration genes: from nematode neurons to vertebrate neural crest. *Genes Dev,* 21**,** 391-6.

KELLEY, M. W. 2006. Regulation of cell fate in the sensory epithelia of the inner ear. *Nat Rev Neurosci,* 7**,** 837-49.

KELLIS, M., PATTERSON, N., ENDRIZZI, M., BIRREN, B. & LANDER, E. S. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature,* 423**,** 241-54.

KENT, W. J. 2002. BLAT--the BLAST-like alignment tool. *Genome Res,* 12**,** 656-64.

KHAN, M. A., SOTO-JIMENEZ, L. M., HOWE, T., STREIT, A., SOSINSKY, A. & STERN, C. D. 2013. Computational tools and resources for prediction and analysis of gene regulatory regions in the chick genome. *Genesis,* 51**,** 311-24.

KHARCHENKO, P. V., ALEKSEYENKO, A. A., SCHWARTZ, Y. B., MINODA, A., RIDDLE, N. C., ERNST, J., SABO, P. J., LARSCHAN, E., GORCHAKOV, A. A., GU, T., LINDER-BASSO, D., PLACHETKA, A., SHANOWER, G., TOLSTORUKOV, M. Y., LUQUETTE, L. J., XI, R., JUNG, Y. L., PARK, R. W., BISHOP, E. P., CANFIELD, T. K., SANDSTROM, R., THURMAN, R. E., MACALPINE, D. M., STAMATOYANNOPOULOS, J. A., KELLIS, M., ELGIN, S. C., KURODA, M. I., PIRROTTA, V., KARPEN, G. H. & PARK, P. J. 2011. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature,* 471**,** 480-5.

KHATRI, S. B., EDLUND, R. K. & GROVES, A. K. 2014. Foxi3 is necessary for the induction of the chick otic placode in response to FGF signaling. *Dev Biol,* 391**,** 158-69.

KHATRI, S. B. & GROVES, A. K. 2013. Expression of the Foxi2 and Foxi3 transcription factors during development of chicken sensory placodes and pharyngeal arches. *Gene Expr Patterns,* 13**,** 38-42.

KHUDYAKOV, J. & BRONNER-FRASER, M. 2009. Comprehensive spatiotemporal analysis of early chick neural crest network genes. *Dev Dyn,* 238**,** 716-23.

KIEFER, S. M., OHLEMILLER, K. K., YANG, J., MCDILL, B. W., KOHLHASE, J. & RAUCHMAN, M. 2003. Expression of a truncated Sall1 transcriptional repressor is responsible for Townes-Brocks syndrome birth defects. *Hum Mol Genet,* 12**,** 2221-7.

KIERNAN, A. E., NUNES, F., WU, D. K. & FEKETE, D. M. 1997. The expression domain of two related homeobox genes defines a compartment in the chicken inner ear that may be involved in semicircular canal formation. *Dev Biol,* 191**,** 215-29.

KIL, S. H., STREIT, A., BROWN, S. T., AGRAWAL, N., COLLAZO, A., ZILE, M. H. & GROVES, A. K. 2005. Distinct roles for hindbrain and paraxial mesoderm in the induction and patterning of the inner ear revealed by a study of vitamin-A-deficient quail. *Dev Biol,* 285**,** 252-71.

KIM, T. H., ABDULLAEV, Z. K., SMITH, A. D., CHING, K. A., LOUKINOV, D. I., GREEN, R. D., ZHANG, M. Q., LOBANENKOV, V. V. & REN, B. 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell,* 128**,** 1231-45.

KIMURA, H. 2013. Histone modifications for human epigenome analysis. *J Hum Genet,* 58**,** 439-45.

KOBAYASHI, M., OSANAI, H., KAWAKAMI, K. & YAMAMOTO, M. 2000. Expression of three zebrafish Six4 genes in the cranial sensory placodes and the developing somites. *Mech Dev,* 98**,** 151-5.

KOHLHASE, J., CHITAYAT, D., KOTZOT, D., CEYLANER, S., FROSTER, U. G., FUCHS, S., MONTGOMERY, T. & ROSLER, B. 2005. SALL4 mutations in Okihiro syndrome (Duane-radial ray syndrome), acro-renal-ocular syndrome, and related disorders. *Hum Mutat,* 26**,** 176-83.

KOHLHASE, J., WISCHERMANN, A., REICHENBACH, H., FROSTER, U. & ENGEL, W. 1998. Mutations in the SALL1 putative transcription factor gene cause Townes-Brocks syndrome. *Nat Genet,* 18**,** 81-3.

KONISHI, Y., IKEDA, K., IWAKURA, Y. & KAWAKAMI, K. 2006. Six1 and Six4 promote survival of sensory neurons during early trigeminal gangliogenesis. *Brain Res,* 1116**,** 93-102.

KOOP, B. F. & HOOD, L. 1994. Striking sequence similarity over almost 100 kilobases of human and mouse T-cell receptor DNA. *Nat Genet,* 7**,** 48-53.

KOSTER, R. W., KUHNLEIN, R. P. & WITTBRODT, J. 2000. Ectopic Sox3 activity elicits sensory placode formation. *Mech Dev,* 95**,** 175-87.

KOZLOWSKI, D. J., MURAKAMI, T., HO, R. K. & WEINBERG, E. S. 1997. Regional cell movement and tissue patterning in the zebrafish embryo revealed by fate mapping with caged fluorescein. *Biochem Cell Biol,* 75**,** 551-62.

KOZLOWSKI, D. J., WHITFIELD, T. T., HUKRIEDE, N. A., LAM, W. K. & WEINBERG, E. S. 2005. The zebrafish dog-eared mutation disrupts eya1, a gene required for cell survival and differentiation in the inner ear and lateral line. *Dev Biol,* 277**,** 27-41.

KRAUSS, S., JOHANSEN, T., KORZH, V. & FJOSE, A. 1991. Expression of the zebrafish paired box gene pax[zf-b] during early neurogenesis. *Development,* 113**,** 1193-206.

KRUG, P., MORINIERE, V., MARLIN, S., KOUBI, V., GABRIEL, H. D., COLIN, E., BONNEAU, D., SALOMON, R., ANTIGNAC, C. & HEIDET, L. 2011. Mutation screening of the EYA1, SIX1, and SIX5 genes in a large cohort of patients harboring branchio-oto-renal syndrome calls into question the pathogenic role of SIX5 mutations. *Hum Mutat,* 32**,** 183-90.

KUMAR, J. P., JAMAL, T., DOETSCH, A., TURNER, F. R. & DUFFY, J. B. 2004. CREB binding protein functions during successive stages of eye development in Drosophila. *Genetics,* 168**,** 877-93.

KUNASEGARAN, K., HO, V., CHANG, T. H., DE SILVA, D., BAKKER, M. L., CHRISTOFFELS, V. M. & PIETERSEN, A. M. 2014. Transcriptional repressor Tbx3 is required for the hormone-sensing cell lineage in mammary epithelium. *PLoS One,* 9**,** e110191.

KWON, H. J., BHAT, N., SWEET, E. M., CORNELL, R. A. & RILEY, B. B. 2010. Identification of early requirements for preplacodal ectoderm and sensory organ development. *PLoS Genet,* 6**,** e1001133.

LACLEF, C., SOUIL, E., DEMIGNON, J. & MAIRE, P. 2003. Thymus, kidney and craniofacial abnormalities in Six1 deficient mice. *Mechanisms of Development,* 120**,** 669-679.

LADHER, R. K., ANAKWE, K. U., GURNEY, A. L., SCHOENWOLF, G. C. & FRANCIS-WEST, P. H. 2000. Identification of synergistic signals initiating inner ear development. *Science,* 290**,** 1965-7.

LADHER, R. K., O'NEILL, P. & BEGBIE, J. 2010. From shared lineage to distinct functions: the development of the inner ear and epibranchial placodes. *Development,* 137**,** 1777-85.

LADHER, R. K., WRIGHT, T. J., MOON, A. M., MANSOUR, S. L. & SCHOENWOLF, G. C. 2005. FGF8 initiates inner ear induction in chick and mouse. *Genes Dev,* 19**,** 603-13.

LADUNGA, I. 2010. *Computational biology of transcription factor binding,* New York, NY, Humana Press.

LAGUTIN, O., ZHU, C. C., FURUTA, Y., ROWITCH, D. H., MCMAHON, A. P. & OLIVER, G. 2001. Six3 promotes the formation of ectopic optic vesicle-like structures in mouse embryos. *Dev Dyn,* 221**,** 342-9.

LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., DEVON, K., DEWAR, K., DOYLE, M., FITZHUGH, W., FUNKE, R., GAGE, D., HARRIS, K., HEAFORD, A., HOWLAND, J., KANN, L., LEHOCZKY, J., LEVINE, R., MCEWAN, P., MCKERNAN, K., MELDRIM, J., MESIROV, J. P., MIRANDA, C., MORRIS, W., NAYLOR, J., RAYMOND, C., ROSETTI, M., SANTOS, R., SHERIDAN, A., SOUGNEZ, C., STANGE-

THOMANN, N., STOJANOVIC, N., SUBRAMANIAN, A., WYMAN, D., ROGERS, J., SULSTON, J., AINSCOUGH, R., BECK, S., BENTLEY, D., BURTON, J., CLEE, C., CARTER, N., COULSON, A., DEADMAN, R., DELOUKAS, P., DUNHAM, A., DUNHAM, I., DURBIN, R., FRENCH, L., GRAFHAM, D., GREGORY, S., HUBBARD, T., HUMPHRAY, S., HUNT, A., JONES, M., LLOYD, C., MCMURRAY, A., MATTHEWS, L., MERCER, S., MILNE, S., MULLIKIN, J. C., MUNGALL, A., PLUMB, R., ROSS, M., SHOWNKEEN, R., SIMS, S., WATERSTON, R. H., WILSON, R. K., HILLIER, L. W., MCPHERSON, J. D., MARRA, M. A., MARDIS, E. R., FULTON, L. A., CHINWALLA, A. T., PEPIN, K. H., GISH, W. R., CHISSOE, S. L., WENDL, M. C., DELEHAUNTY, K. D., MINER, T. L., DELEHAUNTY, A., KRAMER, J. B., COOK, L. L., FULTON, R. S., JOHNSON, D. L., MINX, P. J., CLIFTON, S. W., HAWKINS, T., BRANSCOMB, E., PREDKI, P., RICHARDSON, P., WENNING, S., SLEZAK, T., DOGGETT, N., CHENG, J. F., OLSEN, A., LUCAS, S., ELKIN, C., UBERBACHER, E., FRAZIER, M., et al. 2001. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol,* 10**,** R25.

LASSITER, R. N., DUDE, C. M., REYNOLDS, S. B., WINTERS, N. I., BAKER, C. V. & STARK, M. R. 2007. Canonical Wnt signaling is required for ophthalmic trigeminal placode cell fate determination and maintenance. *Dev Biol,* 308**,** 392-406.

LAWRENCE, C. E., ALTSCHUL, S. F., BOGUSKI, M. S., LIU, J. S., NEUWALD, A. F. & WOOTTON, J. C. 1993. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science,* 262**,** 208-14.

LEE, C. S., FRIEDMAN, J. R., FULMER, J. T. & KAESTNER, K. H. 2005. The initiation of liver development is dependent on Foxa transcription factors. *Nature,* 435**,** 944-7.

LEE, S. A., SHEN, E. L., FISER, A., SALI, A. & GUO, S. 2003. The zebrafish forkhead transcription factor Foxi1 specifies epibranchial placode-derived sensory neurons. *Development,* 130**,** 2669-79.

LEGER, S. & BRAND, M. 2002. Fgf8 and Fgf3 are required for zebrafish ear placode induction, maintenance and inner ear patterning. *Mech Dev,* 119**,** 91-108.

LEIMEISTER, C., DALE, K., FISCHER, A., KLAMT, B., HRABE DE ANGELIS, M., RADTKE, F., MCGREW, M. J., POURQUIE, O. & GESSLER, M. 2000. Oscillating expression of c-Hey2 in the presomitic mesoderm suggests that the segmentation clock may use combinatorial signaling through multiple interacting bHLH factors. *Dev Biol,* 227**,** 91-103.

LENHARD, B., SANDELIN, A., MENDOZA, L., ENGSTROM, P., JAREBORG, N. & WASSERMAN, W. W. 2003. Identification of conserved regulatory elements by comparative genome analysis. *J Biol,* 2**,** 13.

LETTICE, L. A., HEANEY, S. J., PURDIE, L. A., LI, L., DE BEER, P., OOSTRA, B. A., GOODE, D., ELGAR, G., HILL, R. E. & DE GRAAFF, E. 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet,* 12**,** 1725-35.

LETTICE, L. A., HILL, A. E., DEVENNEY, P. S. & HILL, R. E. 2008. Point mutations in a distant sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial polydactyly. *Hum Mol Genet,* 17**,** 978-85.

LEVINE, M. & DAVIDSON, E. H. 2005. Gene regulatory networks for development. *Proc Natl Acad Sci U S A,* 102**,** 4936-42.

LEWIS, E. B. 1978. A gene complex controlling segmentation in Drosophila. *Nature,* 276**,** 565-70.

LI, B., KURIYAMA, S., MORENO, M. & MAYOR, R. 2009. The posteriorizing gene Gbx2 is a direct target of Wnt signalling and the earliest factor in neural crest induction. *Development,* 136**,** 3267-78.

LI, H., LIU, H., CORRALES, C. E., MUTAI, H. & HELLER, S. 2004. Correlation of Pax-2 expression with cell proliferation in the developing chicken inner ear. *J Neurobiol,* 60**,** 61-70.

LI, H. S., YANG, J. M., JACOBSON, R. D., PASKO, D. & SUNDIN, O. 1994. Pax-6 is first expressed in a region of ectoderm anterior to the early neural plate: implications for stepwise determination of the lens. *Dev Biol,* 162**,** 181-94.

LI, Y., HUANG, W., NIU, L., UMBACH, D. M., COVO, S. & LI, L. 2013. Characterization of constitutive CTCF/cohesin loci: a possible role in establishing topological domains in mammalian genomes. *BMC Genomics,* 14**,** 553.

LIANG, H. L., NIEN, C. Y., LIU, H. Y., METZSTEIN, M. M., KIROV, N. & RUSHLOW, C. 2008. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature,* 456**,** 400-3.

LITSIOU, A., HANSON, S. & STREIT, A. 2005. A balance of FGF, BMP and WNT signalling positions the future placode territory in the head. *Development,* 132**,** 4051-62.

LIU, D., CHU, H., MAVES, L., YAN, Y. L., MORCOS, P. A., POSTLETHWAIT, J. H. & WESTERFIELD, M. 2003. Fgf3 and Fgf8 dependent and independent transcription factors are required for otic placode specification. *Development,* 130**,** 2213-24.

LIU, W., LAGUTIN, O. V., MENDE, M., STREIT, A. & OLIVER, G. 2006. Six3 activation of Pax6 expression is essential for mammalian lens induction and specification. *EMBO J,* 25**,** 5383-95.

LLERAS-FORERO, L. 2011. Novel roles for neuropeptides in early sensory organ development. *A thesis submitted to the King's College London's Higher Degree Office in partial fulfilment for the Degree of Doctor of Philosophy*.

LLERAS-FORERO, L., TAMBALO, M., CHRISTOPHOROU, N., CHAMBERS, D., HOUART, C. & STREIT, A. 2013. Neuropeptides: developmental signals in placode progenitor formation. *Dev Cell,* 26**,** 195-203.

LOMBARDO, A., ISAACS, H. V. & SLACK, J. M. 1998. Expression and functions of FGF-3 in Xenopus development. *Int J Dev Biol,* 42**,** 1101-7.

LONG, T. A., BRADY, S. M. & BENFEY, P. N. 2008. Systems approaches to identifying gene regulatory networks in plants. *Annu Rev Cell Dev Biol,* 24**,** 81-103.

LONGABAUGH, W. J., DAVIDSON, E. H. & BOLOURI, H. 2009. Visualization, documentation, analysis, and communication of large-scale gene regulatory networks. *Biochim Biophys Acta,* 1789**,** 363-74.

LOOTS, G. & OVCHARENKO, I. 2007. ECRbase: database of evolutionary conserved regions, promoters, and transcription factor binding sites in vertebrate genomes. *Bioinformatics,* 23**,** 122-4.

LOOTS, G. G. 2008. Genomic identification of regulatory elements by evolutionary sequence comparison and functional analysis. *Adv Genet,* 61**,** 269-93.

LOOTS, G. G., LOCKSLEY, R. M., BLANKESPOOR, C. M., WANG, Z. E., MILLER, W., RUBIN, E. M. & FRAZER, K. A. 2000. Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science,* 288**,** 136-40.

LOPEZ-BERGAMI, P., HUANG, C., GOYDOS, J. S., YIP, D., BAR-ELI, M., HERLYN, M., SMALLEY, K. S., MAHALE, A., EROSHKIN, A., AARONSON,

S. & RONAI, Z. 2007. Rewired ERK-JNK signaling pathways in melanoma. *Cancer Cell,* 11**,** 447-60.
LUDWIG, M. Z., BERGMAN, C., PATEL, N. H. & KREITMAN, M. 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature,* 403**,** 564-7.
LUNN, J. S., FISHWICK, K. J., HALLEY, P. A. & STOREY, K. G. 2007. A spatial and temporal map of FGF/Erk1/2 activity and response repertoires in the early chick embryo. *Developmental Biology,* 302**,** 536-552.
LUPIEN, M., EECKHOUTE, J., MEYER, C. A., WANG, Q., ZHANG, Y., LI, W., CARROLL, J. S., LIU, X. S. & BROWN, M. 2008. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell,* 132**,** 958-70.
LUSCOMBE, N. M., BABU, M. M., YU, H., SNYDER, M., TEICHMANN, S. A. & GERSTEIN, M. 2004. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature,* 431**,** 308-12.
MAAS, S. A., SUZUKI, T. & FALLON, J. F. 2011. Identification of spontaneous mutations within the long-range limb-specific Sonic hedgehog enhancer (ZRS) that alter Sonic hedgehog expression in the chicken limb mutants oligozeugodactyly and silkie breed. *Dev Dyn,* 240**,** 1212-22.
MACKERETH, M. D., KWAK, S. J., FRITZ, A. & RILEY, B. B. 2005. Zebrafish pax8 is required for otic placode induction and plays a redundant role with Pax2 genes in the maintenance of the otic placode. *Development,* 132**,** 371-82.
MACNEIL, L. T. & WALHOUT, A. J. 2011. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Res,* 21**,** 645-57.
MACQUEEN, J. B. 1967. Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability,* 1**,** 281-297.
MAERE, S., HEYMANS, K. & KUIPER, M. 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics,* 21**,** 3448-9.
MAHMOOD, R., KIEFER, P., GUTHRIE, S., DICKSON, C. & MASON, I. 1995. Multiple roles for FGF-3 during cranial neural development in the chicken. *Development,* 121**,** 1399-410.
MALIK, S. & ROEDER, R. G. 2010. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat Rev Genet,* 11**,** 761-72.
MARGOLIN, A. A., WANG, K., LIM, W. K., KUSTAGI, M., NEMENMAN, I. & CALIFANO, A. 2006. Reverse engineering cellular networks. *Nature Protocols,* 1**,** 663-672.
MARIN, F. & NIETO, M. A. 2004. Expression of chicken slug and snail in mesenchymal components of the developing central nervous system. *Dev Dyn,* 230**,** 144-8.
MARKOWETZ, F. & SPANG, R. 2007. Inferring cellular networks--a review. *BMC Bioinformatics,* 8 Suppl 6**,** S5.
MAROON, H., WALSHE, J., MAHMOOD, R., KIEFER, P., DICKSON, C. & MASON, I. 2002. Fgf3 and Fgf8 are required together for formation of the otic placode and vesicle. *Development,* 129**,** 2099-108.
MARTIN, D., PANTOJA, C., FERNANDEZ MINAN, A., VALDES-QUEZADA, C., MOLTO, E., MATESANZ, F., BOGDANOVIC, O., DE LA CALLE-MUSTIENES, E., DOMINGUEZ, O., TAHER, L., FURLAN-MAGARIL, M., ALCINA, A., CANON, S., FEDETZ, M., BLASCO, M. A., PEREIRA, P. S., OVCHARENKO, I., RECILLAS-TARGA, F., MONTOLIU, L., MANZANARES, M., GUIGO, R., SERRANO, M., CASARES, F. & GOMEZ-SKARMETA, J. L. 2011. Genome-wide CTCF distribution in vertebrates defines

equivalent sites that aid the identification of disease-associated genes. *Nat Struct Mol Biol,* 18**,** 708-14.

MARTIN, K. & GROVES, A. K. 2006. Competence of cranial ectoderm to respond to Fgf signaling suggests a two-step model of otic placode induction. *Development,* 133**,** 877-87.

MASTON, G. A., LANDT, S. G., SNYDER, M. & GREEN, M. R. 2012. Characterization of enhancer function from genome-wide analyses. *Annu Rev Genomics Hum Genet,* 13**,** 29-57.

MATSUMATA, M., UCHIKAWA, M., KAMACHI, Y. & KONDOH, H. 2005. Multiple N-cadherin enhancers identified by systematic functional screening indicate its Group B1 SOX-dependent regulation in neural and placodal development. *Dev Biol,* 286**,** 601-17.

MATYS, V., KEL-MARGOULIS, O. V., FRICKE, E., LIEBICH, I., LAND, S., BARRE-DIRRIE, A., REUTER, I., CHEKMENEV, D., KRULL, M., HORNISCHER, K., VOSS, N., STEGMAIER, P., LEWICKI-POTAPOV, B., SAXEL, H., KEL, A. E. & WINGENDER, E. 2006. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res,* 34**,** D108-10.

MAULDING, K., PADANAD, M. S., DONG, J. & RILEY, B. B. 2014. Mesodermal Fgf10b cooperates with other Fgfs during induction of otic and epibranchial placodes in zebrafish. *Dev Dyn*.

MAY, D., BLOW, M. J., KAPLAN, T., MCCULLEY, D. J., JENSEN, B. C., AKIYAMA, J. A., HOLT, A., PLAJZER-FRICK, I., SHOUKRY, M., WRIGHT, C., AFZAL, V., SIMPSON, P. C., RUBIN, E. M., BLACK, B. L., BRISTOW, J., PENNACCHIO, L. A. & VISEL, A. 2012. Large-scale discovery of enhancers from human heart tissue. *Nat Genet,* 44**,** 89-93.

MCCABE, K. L. & BRONNER-FRASER, M. 2008. Essential role for PDGF signaling in ophthalmic trigeminal placode induction. *Development,* 135**,** 1863-74.

MCCARROLL, M. N., LEWIS, Z. R., CULBERTSON, M. D., MARTIN, B. L., KIMELMAN, D. & NECHIPORUK, A. V. 2012. Graded levels of Pax2a and Pax8 regulate cell differentiation during sensory placode formation. *Development,* 139**,** 2740-50.

MCCARROLL, M. N. & NECHIPORUK, A. V. 2013. Fgf3 and Fgf10a work in concert to promote maturation of the epibranchial placodes in zebrafish. *PLoS One,* 8**,** e85087.

MCKAY, I. J., LEWIS, J. & LUMSDEN, A. 1996. The role of FGF-3 in early inner ear development: an analysis in normal and kreisler mutant mice. *Dev Biol,* 174**,** 370-8.

MCKEOWN, S. J., LEE, V. M., BRONNER-FRASER, M., NEWGREEN, D. F. & FARLIE, P. G. 2005. Sox10 overexpression induces neural crest-like cells from all dorsoventral levels of the neural tube but inhibits differentiation. *Dev Dyn,* 233**,** 430-44.

MCLARREN, K. W., LITSIOU, A. & STREIT, A. 2003. DLX5 positions the neural crest and preplacode region at the border of the neural plate. *Developmental Biology,* 259**,** 34-47.

MCLEAY, R. C. & BAILEY, T. L. 2010. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics,* 11**,** 165.

MCMAHON, A. R. & MERZDORF, C. S. 2010. Expression of the zic1, zic2, zic3, and zic4 genes in early chick embryos. *BMC Res Notes,* 3**,** 167.

MELNICK, A., AHMAD, K. F., ARAI, S., POLINGER, A., BALL, H., BORDEN, K. L., CARLILE, G. W., PRIVE, G. G. & LICHT, J. D. 2000. In-depth mutational analysis of the promyelocytic leukemia zinc finger BTB/POZ domain reveals

motifs and residues required for biological and transcriptional functions. *Mol Cell Biol,* 20**,** 6550-67.

MENDONSA, E. S. & RILEY, B. B. 1999. Genetic analysis of tissue interactions required for otic placode induction in the zebrafish. *Dev Biol,* 206**,** 100-12.

MEYER, L. R., ZWEIG, A. S., HINRICHS, A. S., KAROLCHIK, D., KUHN, R. M., WONG, M., SLOAN, C. A., ROSENBLOOM, K. R., ROE, G., RHEAD, B., RANEY, B. J., POHL, A., MALLADI, V. S., LI, C. H., LEE, B. T., LEARNED, K., KIRKUP, V., HSU, F., HEITNER, S., HARTE, R. A., HAEUSSLER, M., GURUVADOO, L., GOLDMAN, M., GIARDINE, B. M., FUJITA, P. A., DRESZER, T. R., DIEKHANS, M., CLINE, M. S., CLAWSON, H., BARBER, G. P., HAUSSLER, D. & KENT, W. J. 2013. The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res,* 41**,** D64-9.

MEYER, P. E., KONTOS, K., LAFITTE, F. & BONTEMPI, G. 2007. Information-theoretic inference of large transcriptional regulatory networks. *EURASIP J Bioinform Syst Biol***,** 79879.

MEYER, P. E., LAFITTE, F. & BONTEMPI, G. 2008. minet: A R/Bioconductor package for inferring large transcriptional networks using mutual information. *BMC Bioinformatics,* 9**,** 461.

MEYERS, E. N., LEWANDOSKI, M. & MARTIN, G. R. 1998. An Fgf8 mutant allelic series generated by Cre- and Flp-mediated recombination. *Nat Genet,* 18**,** 136-41.

MILLET, S., CAMPBELL, K., EPSTEIN, D. J., LOSOS, K., HARRIS, E. & JOYNER, A. L. 1999. A role for Gbx2 in repression of Otx2 and positioning the mid/hindbrain organizer. *Nature,* 401**,** 161-4.

MILO, R., SHEN-ORR, S., ITZKOVITZ, S., KASHTAN, N., CHKLOVSKII, D. & ALON, U. 2002. Network motifs: simple building blocks of complex networks. *Science,* 298**,** 824-7.

MINOWADA, G., JARVIS, L. A., CHI, C. L., NEUBUSER, A., SUN, X., HACOHEN, N., KRASNOW, M. A. & MARTIN, G. R. 1999. Vertebrate Sprouty genes are induced by FGF signaling and can cause chondrodysplasia when overexpressed. *Development,* 126**,** 4465-75.

MISHIMA, N. & TOMAREV, S. 1998. Chicken Eyes absent 2 gene: isolation and expression pattern during development. *Int J Dev Biol,* 42**,** 1109-15.

MIYAKE, T., VON HERBING, I. H. AND HALL, B. K. 1997. Neural ectoderm, neural crest, and placodes: contribution of the otic placode to the ectodermal lining of the embryonic opercular cavity in Atlantic cod (Teleostei). . *J.Morphol.* **,** 231-253.

MODRELL, M. S., HOCKMAN, D., UY, B., BUCKLEY, D., SAUKA-SPENGLER, T., BRONNER, M. E. & BAKER, C. V. 2014. A fate-map for cranial sensory ganglia in the sea lamprey. *Dev Biol,* 385**,** 405-16.

MOHAMMADI, M., MCMAHON, G., SUN, L., TANG, C., HIRTH, P., YEH, B. K., HUBBARD, S. R. & SCHLESSINGER, J. 1997. Structures of the tyrosine kinase domain of fibroblast growth factor receptor in complex with inhibitors. *Science,* 276**,** 955-60.

MORDELET, F. & VERT, J. P. 2008. SIRENE: supervised inference of regulatory networks. *Bioinformatics,* 24**,** i76-82.

MORENO-PELAYO, M. A., MODAMIO-HOYBJOR, S., MENCIA, A., DEL CASTILLO, I., CHARDENOUX, S., FERNANDEZ-BURRIEL, M., LATHROP, M., PETIT, C. & MORENO, F. 2003. DFNA49, a novel locus for autosomal dominant non-syndromic hearing loss, maps proximal to DFNA7/DFNM1 region on chromosome 1q21-q23. *J Med Genet,* 40**,** 832-6.

MORGENSTERN, B., FRECH, K., DRESS, A. & WERNER, T. 1998. DIALIGN: finding local similarities by multiple sequence alignment. *Bioinformatics,* 14**,** 290-4.

MORTAZAVI, A., WILLIAMS, B. A., MCCUE, K., SCHAEFFER, L. & WOLD, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods,* 5**,** 621-8.

MOSIMANN, C., HAUSMANN, G. & BASLER, K. 2009. Beta-catenin hits chromatin: regulation of Wnt target gene activation. *Nat Rev Mol Cell Biol,* 10**,** 276-86.

MUKHOPADHYAY, R., YU, W., WHITEHEAD, J., XU, J., LEZCANO, M., PACK, S., KANDURI, C., KANDURI, M., GINJALA, V., VOSTROV, A., QUITSCHKE, W., CHERNUKHIN, I., KLENOVA, E., LOBANENKOV, V. & OHLSSON, R. 2004. The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide. *Genome Res,* 14**,** 1594-602.

NARLIKAR, L., SAKABE, N. J., BLANSKI, A. A., ARIMURA, F. E., WESTLUND, J. M., NOBREGA, M. A. & OVCHARENKO, I. 2010. Genome-wide discovery of human heart enhancers. *Genome Res,* 20**,** 381-92.

NATIVIO, R., WENDT, K. S., ITO, Y., HUDDLESTON, J. E., URIBE-LEWIS, S., WOODFINE, K., KRUEGER, C., REIK, W., PETERS, J. M. & MURRELL, A. 2009. Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. *PLoS Genet,* 5**,** e1000739.

NECHIPORUK, A., LINBO, T., POSS, K. D. & RAIBLE, D. W. 2007. Specification of epibranchial placodes in zebrafish. *Development,* 134**,** 611-23.

NEEDLEMAN, S. B. & WUNSCH, C. D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol,* 48**,** 443-53.

NEGRE, N., BROWN, C. D., SHAH, P. K., KHERADPOUR, P., MORRISON, C. A., HENIKOFF, J. G., FENG, X., AHMAD, K., RUSSELL, S., WHITE, R. A., STEIN, L., HENIKOFF, S., KELLIS, M. & WHITE, K. P. 2010. A comprehensive map of insulator elements for the Drosophila genome. *PLoS Genet,* 6**,** e1000814.

NEUBERG, M., SCHUERMANN, M., HUNTER, J. B. & MULLER, R. 1989. Two functionally different regions in Fos are required for the sequence-specific DNA interaction of the Fos/Jun protein complex. *Nature,* 338**,** 589-90.

NEWBURGER, D. E. & BULYK, M. L. 2009. UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. *Nucleic Acids Res,* 37**,** D77-82.

NEWMAN, M. E. & GIRVAN, M. 2004. Finding and evaluating community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys,* 69**,** 026113.

NICHOLS D. H., P. S., JAHAN I., BEISEL K. W., MILLEN K. J. & FRITZSCH, B. 2008. Lmx1a is required for segregation of sensory epithelia and normal ear histogenesis and morphogenesis. *Cell Tissue Res,* 334**,** 339–358.

NICOL, J. W., HELT, G. A., BLANCHARD, S. G., RAJA, A. & LORAINE, A. E. 2009. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics,* 25**,** 2730-2731.

NIKAIDO, M., DOI, K., SHIMIZU, T., HIBI, M., KIKUCHI, Y. & YAMASU, K. 2007. Initial specification of the epibranchial placode in zebrafish embryos depends on the fibroblast growth factor signal. *Dev Dyn,* 236**,** 564-71.

NISHITA, J., OHTA, S., BLEYL, S. B. & SCHOENWOLF, G. C. 2011. Detection of isoform-specific fibroblast growth factor receptors by whole-mount in situ hybridization in early chick embryos. *Dev Dyn,* 240**,** 1537-47.

NISSEN, R. M. 2003. Zebrafish foxi one modulates cellular responses to Fgf signaling required for the integrity of ear and jaw patterning. *Development,* 130**,** 2543-2554.

NISSEN, R. M., YAN, J., AMSTERDAM, A., HOPKINS, N. & BURGESS, S. M. 2003. Zebrafish foxi one modulates cellular responses to Fgf signaling required for the integrity of ear and jaw patterning. *Development,* 130**,** 2543-54.

NOBREGA, M. A., OVCHARENKO, I., AFZAL, V. & RUBIN, E. M. 2003. Scanning human gene deserts for long-range enhancers. *Science,* 302**,** 413.
NOONAN, J. P. & MCCALLION, A. S. 2010. Genomics of long-range regulatory elements. *Annu Rev Genomics Hum Genet,* 11**,** 1-23.
NOOR, A., SERPEDIN, E., NOUNOU, M., NOUNOU, H., MOHAMED, N. & CHOUCHANE, L. 2013. An overview of the statistical methods used for inferring gene regulatory networks and protein-protein interaction networks. *Adv Bioinformatics,* 2013**,** 953814.
NORNES, H. O., DRESSLER, G. R., KNAPIK, E. W., DEUTSCH, U. & GRUSS, P. 1990. Spatially and temporally restricted expression of Pax2 during murine neurogenesis. *Development,* 109**,** 797-809.
NORTHCUTT, R. G. & BRANDLE, K. 1995. Development of branchiomeric and lateral line nerves in the axolotl. *J Comp Neurol,* 355**,** 427-54.
NOYES, M. B., CHRISTENSEN, R. G., WAKABAYASHI, A., STORMO, G. D., BRODSKY, M. H. & WOLFE, S. A. 2008. Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell,* 133**,** 1277-89.
OHLSSON, R., RENKAWITZ, R. & LOBANENKOV, V. 2001. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet,* 17**,** 520-7.
OHUCHI, H., HORI, Y., YAMASAKI, M., HARADA, H., SEKINE, K., KATO, S. & ITOH, N. 2000a. FGF10 acts as a major ligand for FGF receptor 2 IIIb in mouse multi-organ development. *Biochem Biophys Res Commun,* 277**,** 643-9.
OHUCHI, H., KIMURA, S., WATAMOTO, M. & ITOH, N. 2000b. Involvement of fibroblast growth factor (FGF)18-FGF8 signaling in specification of left-right asymmetry and brain and limb development of the chick embryo. *Mech Dev,* 95**,** 55-66.
OHYAMA, T. & GROVES, A. K. 2004. Expression of mouse Foxi class genes in early craniofacial development. *Dev Dyn,* 231**,** 640-6.
OHYAMA, T., GROVES, A. K. & MARTIN, K. 2007. The first steps towards hearing: mechanisms of otic placode induction. *Int J Dev Biol,* 51**,** 463-72.
OHYAMA, T., MOHAMED, O. A., TAKETO, M. M., DUFORT, D. & GROVES, A. K. 2006. Wnt signals mediate a fate decision between otic placode and epidermis. *Development,* 133**,** 865-75.
OLIVER, G., LOOSLI, F., KOSTER, R., WITTBRODT, J. & GRUSS, P. 1996. Ectopic lens induction in fish in response to the murine homeobox gene Six3. *Mech Dev,* 60**,** 233-9.
OLIVERA-MARTINEZ, I. & STOREY, K. G. 2007. Wnt signals provide a timing mechanism for the FGF-retinoid differentiation switch during vertebrate body axis extension. *Development,* 134**,** 2125-35.
OLIVERI, P., TU, Q. & DAVIDSON, E. H. 2008. Global regulatory logic for specification of an embryonic cell lineage. *Proc Natl Acad Sci U S A,* 105**,** 5955-62.
ONG, C. T. & CORCES, V. G. 2011. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet,* 12**,** 283-93.
ONG, C. T. & CORCES, V. G. 2014. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet,* 15**,** 234-46.
ORNITZ, D. M., HERR, A. B., NILSSON, M., WESTMAN, J., SVAHN, C. M. & WAKSMAN, G. 1995. FGF binding and FGF receptor activation by synthetic heparan-derived di- and trisaccharides. *Science,* 268**,** 432-6.
OROM, U. A., DERRIEN, T., BERINGER, M., GUMIREDDY, K., GARDINI, A., BUSSOTTI, G., LAI, F., ZYTNICKI, M., NOTREDAME, C., HUANG, Q.,

GUIGO, R. & SHIEKHATTAR, R. 2010. Long noncoding RNAs with enhancer-like function in human cells. *Cell,* 143**,** 46-58.

ORTS, F., JIMENEZ-COLLADO, L. & JIMENEZ-COLLADO, J. 1971. Regulation of the embryo after the extirpation of Hensen's node. Consequences on the differentiation of the otic placode. *Arch Anat Histol Embryol,* 54**,** 1-11.

OSTUNI, R., PICCOLO, V., BAROZZI, I., POLLETTI, S., TERMANINI, A., BONIFACIO, S., CURINA, A., PROSPERINI, E., GHISLETTI, S. & NATOLI, G. 2013. Latent enhancers activated by stimulation in differentiated cells. *Cell,* 152**,** 157-71.

OVCHARENKO, I., NOBREGA, M. A., LOOTS, G. G. & STUBBS, L. 2004. ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. *Nucleic Acids Res,* 32**,** W280-6.

OZAKI, H., NAKAMURA, K., FUNAHASHI, J., IKEDA, K., YAMADA, G., TOKANO, H., OKAMURA, H. O., KITAMURA, K., MUTO, S., KOTAKI, H., SUDO, K., HORAI, R., IWAKURA, Y. & KAWAKAMI, K. 2004. Six1 controls patterning of the mouse otic vesicle. *Development,* 131**,** 551-62.

OZAKI, K., KADOMOTO, R., ASATO, K., TANIMURA, S., ITOH, N. & KOHNO, M. 2001. ERK pathway positively regulates the expression of Sprouty genes. *Biochem Biophys Res Commun,* 285**,** 1084-8.

PADANAD, M. S., BHAT, N., GUO, B. & RILEY, B. B. 2012. Conditions that influence the response to Fgf during otic placode induction. *Dev Biol,* 364**,** 1-10.

PADANAD, M. S. & RILEY, B. B. 2011. Pax2/8 proteins coordinate sequential induction of otic and epibranchial placodes through differential regulation of foxi1, sox3 and fgf24. *Dev Biol,* 351**,** 90-8.

PANDUR, P. D. & MOODY, S. A. 2000. Xenopus Six1 gene is expressed in neurogenic cranial placodes and maintained in the differentiating lateral lines. *Mech Dev,* 96**,** 253-7.

PAPANAYOTOU, C., MEY, A., BIROT, A. M., SAKA, Y., BOAST, S., SMITH, J. C., SAMARUT, J. & STERN, C. D. 2008. A mechanism regulating the onset of Sox2 expression in the embryonic neural plate. *PLoS Biol,* 6**,** e2.

PARELHO, V., HADJUR, S., SPIVAKOV, M., LELEU, M., SAUER, S., GREGSON, H. C., JARMUZ, A., CANZONETTA, C., WEBSTER, Z., NESTEROVA, T., COBB, B. S., YOKOMORI, K., DILLON, N., ARAGON, L., FISHER, A. G. & MERKENSCHLAGER, M. 2008. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell,* 132**,** 422-33.

PARK, B. Y. & SAINT-JEANNET, J. P. 2008. Hindbrain-derived Wnt and Fgf signals cooperate to specify the otic placode in Xenopus. *Dev Biol,* 324**,** 108-21.

PASINI, D., MALATESTA, M., JUNG, H. R., WALFRIDSSON, J., WILLER, A., OLSSON, L., SKOTTE, J., WUTZ, A., PORSE, B., JENSEN, O. N. & HELIN, K. 2010. Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res,* 38**,** 4958-69.

PATEL, N. S., RHINN, M., SEMPRICH, C. I., HALLEY, P. A., DOLLE, P., BICKMORE, W. A. & STOREY, K. G. 2013. FGF signalling regulates chromatin organisation during neural differentiation via mechanisms that can be uncoupled from transcription. *PLoS Genet,* 9**,** e1003614.

PATEN, B., HERRERO, J., BEAL, K., FITZGERALD, S. & BIRNEY, E. 2008. Enredo and Pecan: Genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Research,* 18**,** 1814-1828.

PATTHEY, C., SCHLOSSER, G. & SHIMELD, S. M. 2014. The evolutionary history of vertebrate cranial placodes--I: cell type evolution. *Dev Biol,* 389**,** 82-97.

PATTYN, A., HIRSCH, M., GORIDIS, C. & BRUNET, J. F. 2000. Control of hindbrain motor neuron differentiation by the homeobox gene Phox2b. *Development,* 127**,** 1349-58.

PATTYN, A., MORIN, X., CREMER, H., GORIDIS, C. & BRUNET, J. F. 1997. Expression and interactions of the two closely related homeobox genes Phox2a and Phox2b during neurogenesis. *Development,* 124**,** 4065-75.

PATTYN, A., MORIN, X., CREMER, H., GORIDIS, C. & BRUNET, J. F. 1999. The homeobox gene Phox2b is essential for the development of autonomic neural crest derivatives. *Nature,* 399**,** 366-70.

PAULEY, S., WRIGHT, T. J., PIRVOLA, U., ORNITZ, D., BEISEL, K. & FRITZSCH, B. 2003. Expression and function of FGF10 in mammalian inner ear development. *Dev Dyn,* 227**,** 203-15.

PAVESI, G., MAURI, G. & PESOLE, G. 2004a. In silico representation and discovery of transcription factor binding sites. *Brief Bioinform,* 5**,** 217-36.

PAVESI, G., MEREGHETTI, P., MAURI, G. & PESOLE, G. 2004b. Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Res,* 32**,** W199-203.

PAXTON, C. N., BLEYL, S. B., CHAPMAN, S. C. & SCHOENWOLF, G. C. 2010. Identification of differentially expressed genes in early inner ear development. *Gene Expr Patterns,* 10**,** 31-43.

PENNACCHIO, L. A., LOOTS, G. G., NOBREGA, M. A. & OVCHARENKO, I. 2007. Predicting tissue-specific enhancers in the human genome. *Genome Res,* 17**,** 201-11.

PENNACCHIO, L. A. & RUBIN, E. M. 2001. Genomic strategies to identify mammalian regulatory sequences. *Nat Rev Genet,* 2**,** 100-9.

PEPKE, S., WOLD, B. & MORTAZAVI, A. 2009. Computation for ChIP-seq and RNA-seq studies. *Nat Methods,* 6**,** S22-32.

PERA, E. & KESSEL, M. 1999. Expression of DLX3 in chick embryos. *Mech Dev,* 89**,** 189-93.

PFEFFER, P. L., GERSTER, T., LUN, K., BRAND, M. & BUSSLINGER, M. 1998. Characterization of three novel members of the zebrafish Pax2/5/8 family: dependency of Pax5 and Pax8 expression on the Pax2.1 (noi) function. *Development,* 125**,** 3063-74.

PHILLIPS, B. T., BOLDING, K. & RILEY, B. B. 2001. Zebrafish fgf3 and fgf8 encode redundant functions required for otic placode induction. *Dev Biol,* 235**,** 351-65.

PHILLIPS, B. T., KWON, H. J., MELTON, C., HOUGHTALING, P., FRITZ, A. & RILEY, B. B. 2006. Zebrafish msxB, msxC and msxE function together to refine the neural-nonneural border and regulate cranial placodes and neural crest development. *Dev Biol,* 294**,** 376-90.

PHILLIPS, B. T., STORCH, E. M., LEKVEN, A. C. & RILEY, B. B. 2004. A direct role for Fgf but not Wnt in otic placode induction. *Development,* 131**,** 923-31.

PHILLIPS, J. E. & CORCES, V. G. 2009. CTCF: master weaver of the genome. *Cell,* 137**,** 1194-211.

PHUONG, T. M., LEE, D. & LEE, K. H. 2004. Regression trees for regulatory element identification. *Bioinformatics,* 20**,** 750-7.

PIEPER, M., AHRENS, K., RINK, E., PETER, A. & SCHLOSSER, G. 2012. Differential distribution of competence for panplacodal and neural crest induction to non-neural and neural ectoderm. *Development,* 139**,** 1175-87.

PIEPER, M., EAGLESON, G. W., WOSNIOK, W. & SCHLOSSER, G. 2011. Origin and segregation of cranial placodes in Xenopus laevis. *Dev Biol,* 360**,** 257-75.

PIRVOLA, U., SPENCER-DENE, B., XING-QUN, L., KETTUNEN, P., THESLEFF, I., FRITZSCH, B., DICKSON, C. & YLIKOSKI, J. 2000. FGF/FGFR-2(IIIb) signaling is essential for inner ear morphogenesis. *J Neurosci,* 20**,** 6125-34.

POSS, Z. C., EBMEIER, C. C. & TAATJES, D. J. 2013. The Mediator complex and transcription regulation. *Crit Rev Biochem Mol Biol,* 48**,** 575-608.

POTIER, D., DAVIE, K., HULSELMANS, G., NAVAL SANCHEZ, M., HAAGEN, L., HUYNH-THU, V. A., KOLDERE, D., CELIK, A., GEURTS, P., CHRISTIAENS, V. & AERTS, S. 2014. Mapping gene regulatory networks in Drosophila eye development by large-scale transcriptome perturbations and motif inference. *Cell Rep,* 9**,** 2290-303.

PRABHAKAR, S., POULIN, F., SHOUKRY, M., AFZAL, V., RUBIN, E. M., COURONNE, O. & PENNACCHIO, L. A. 2006. Close sequence comparisons are sufficient to identify human cis-regulatory elements. *Genome Res,* 16**,** 855-63.

PRAKASH, A. & TOMPA, M. 2005. Discovery of regulatory elements in vertebrates through comparative genomics. *Nat Biotechnol,* 23**,** 1249-56.

QIN, Z. S., YU, J., SHEN, J., MAHER, C. A., HU, M., KALYANA-SUNDARAM, S., YU, J. & CHINNAIYAN, A. M. 2010. HPeak: an HMM-based algorithm for defining read-enriched regions in ChIP-Seq data. *BMC Bioinformatics,* 11**,** 369.

RADA-IGLESIAS, A., BAJPAI, R., SWIGUT, T., BRUGMANN, S. A., FLYNN, R. A. & WYSOCKA, J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature,* 470**,** 279-83.

RAIBLE, F. & BRAND, M. 2001. Tight transcriptional control of the ETS domain factors Erm and Pea3 by Fgf signaling during early zebrafish development. *Mech Dev,* 107**,** 105-17.

RAJAGOPAL, N., XIE, W., LI, Y., WAGNER, U., WANG, W., STAMATOYANNOPOULOS, J., ERNST, J., KELLIS, M. & REN, B. 2013. RFECS: a random-forest based algorithm for enhancer identification from chromatin state. *PLoS Comput Biol,* 9**,** e1002968.

RAO, S. S., HUNTLEY, M. H., DURAND, N. C., STAMENOVA, E. K., BOCHKOV, I. D., ROBINSON, J. T., SANBORN, A. L., MACHOL, I., OMER, A. D., LANDER, E. S. & AIDEN, E. L. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell,* 159**,** 1665-80.

RAWLES, M. E. 1936. *A study in the localization of organ-forming areas in the chick blastoderm of the head-process state.* Ph D, University of Chicago.

REVANNA, K. V., CHIU, C. C., BIERSCHANK, E. & DONG, Q. F. 2011. GSV: a web-based genome synteny viewer for customized data. *Bmc Bioinformatics,* 12.

REX, M., ORME, A., UWANOGHO, D., TOINTON, K., WIGMORE, P. M., SHARPE, P. T. & SCOTTING, P. J. 1997. Dynamic expression of chicken Sox2 and Sox3 genes in ectoderm induced to form neural tissue. *Dev Dyn,* 209**,** 323-32.

ROEHL, H. & NUSSLEIN-VOLHARD, C. 2001. Zebrafish pea3 and erm are general targets of FGF8 signaling. *Curr Biol,* 11**,** 503-7.

ROGERS, A. A. M., ZHANG, J. & SHIM, K. 2011. Sprouty1 and Sprouty2 limit both the size of the otic placode and hindbrain Wnt8a by antagonizing FGF signaling. *Dev Biol,* 353**,** 94-104.

ROH, T. Y., CUDDAPAH, S. & ZHAO, K. 2005. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev,* 19**,** 542-52.

ROTH, M., BONEV, B., LINDSAY, J., LEA, R., PANAGIOTAKI, N., HOUART, C. & PAPALOPULU, N. 2010. FoxG1 and TLE2 act cooperatively to regulate ventral telencephalon formation. *Development,* 137**,** 1553-62.

ROY, S., ERNST, J., KHARCHENKO, P. V., KHERADPOUR, P., NEGRE, N., EATON, M. L., LANDOLIN, J. M., BRISTOW, C. A., MA, L., LIN, M. F., WASHIETL, S., ARSHINOFF, B. I., AY, F., MEYER, P. E., ROBINE, N.,

WASHINGTON, N. L., DI STEFANO, L., BEREZIKOV, E., BROWN, C. D., CANDEIAS, R., CARLSON, J. W., CARR, A., JUNGREIS, I., MARBACH, D., SEALFON, R., TOLSTORUKOV, M. Y., WILL, S., ALEKSEYENKO, A. A., ARTIERI, C., BOOTH, B. W., BROOKS, A. N., DAI, Q., DAVIS, C. A., DUFF, M. O., FENG, X., GORCHAKOV, A. A., GU, T., HENIKOFF, J. G., KAPRANOV, P., LI, R., MACALPINE, H. K., MALONE, J., MINODA, A., NORDMAN, J., OKAMURA, K., PERRY, M., POWELL, S. K., RIDDLE, N. C., SAKAI, A., SAMSONOVA, A., SANDLER, J. E., SCHWARTZ, Y. B., SHER, N., SPOKONY, R., STURGILL, D., VAN BAREN, M., WAN, K. H., YANG, L., YU, C., FEINGOLD, E., GOOD, P., GUYER, M., LOWDON, R., AHMAD, K., ANDREWS, J., BERGER, B., BRENNER, S. E., BRENT, M. R., CHERBAS, L., ELGIN, S. C., GINGERAS, T. R., GROSSMAN, R., HOSKINS, R. A., KAUFMAN, T. C., KENT, W., KURODA, M. I., ORR-WEAVER, T., PERRIMON, N., PIRROTTA, V., POSAKONY, J. W., REN, B., RUSSELL, S., CHERBAS, P., GRAVELEY, B. R., LEWIS, S., MICKLEM, G., OLIVER, B., PARK, P. J., CELNIKER, S. E., HENIKOFF, S., KARPEN, G. H., LAI, E. C., MACALPINE, D. M., STEIN, L. D., WHITE, K. P. & KELLIS, M. 2010. Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science,* 330**,** 1787-97.

ROZOWSKY, J., EUSKIRCHEN, G., AUERBACH, R. K., ZHANG, Z. D., GIBSON, T., BJORNSON, R., CARRIERO, N., SNYDER, M. & GERSTEIN, M. B. 2009. PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nat Biotechnol,* 27**,** 66-75.

RUAN, J. & ZHANG, W. 2006. A bi-dimensional regression tree approach to the modeling of gene expression regulation. *Bioinformatics,* 22**,** 332-40.

RUBIO, E. D., REISS, D. J., WELCSH, P. L., DISTECHE, C. M., FILIPPOVA, G. N., BALIGA, N. S., AEBERSOLD, R., RANISH, J. A. & KRUMM, A. 2008. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A,* 105**,** 8309-14.

RUF, R. G., XU, P. X., SILVIUS, D., OTTO, E. A., BEEKMANN, F., MUERB, U. T., KUMAR, S., NEUHAUS, T. J., KEMPER, M. J., RAYMOND, R. M., JR., BROPHY, P. D., BERKMAN, J., GATTAS, M., HYLAND, V., RUF, E. M., SCHWARTZ, C., CHANG, E. H., SMITH, R. J., STRATAKIS, C. A., WEIL, D., PETIT, C. & HILDEBRANDT, F. 2004. SIX1 mutations cause branchio-oto-renal syndrome by disruption of EYA1-SIX1-DNA complexes. *Proc Natl Acad Sci U S A,* 101**,** 8090-5.

SAGAI, T., HOSOYA, M., MIZUSHINA, Y., TAMURA, M. & SHIROISHI, T. 2005. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development,* 132**,** 797-803.

SAHLY, I., ANDERMANN, P. & PETIT, C. 1999. The zebrafish eya1 gene and its expression pattern during embryogenesis. *Dev Genes Evol,* 209**,** 399-410.

SANCHEZ-CALDERON, H., FRANCISCO-MORCILLO, J., MARTIN-PARTIDO, G. & HIDALGO-SANCHEZ, M. 2007. Fgf19 expression patterns in the developing chick inner ear. *Gene Expr Patterns,* 7**,** 30-8.

SANCHEZ-CALDERON, H., MARTIN-PARTIDO, G. & HIDALGO-SANCHEZ, M. 2004. Otx2, Gbx2, and Fgf8 expression patterns in the chick developing inner ear and their possible roles in otic specification and early innervation. *Gene Expr Patterns,* 4**,** 659-69.

SANCHEZ-CALDERON, H., MARTIN-PARTIDO, G. & HIDALGO-SANCHEZ, M. 2005. Pax2 expression patterns in the developing chick inner ear. *Gene Expr Patterns,* 5**,** 763-73.

SANCHEZ-GUARDADO, L. O., FERRAN, J. L., RODRIGUEZ-GALLARDO, L., PUELLES, L. & HIDALGO-SANCHEZ, M. 2011. Meis gene expression patterns in the developing chicken inner ear. *J Comp Neurol,* 519**,** 125-47.

SANDELIN, A., ALKEMA, W., ENGSTROM, P., WASSERMAN, W. W. & LENHARD, B. 2004. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res,* 32**,** D91-4.

SANYANUSIN, P., MCNOE, L. A., SULLIVAN, M. J., WEAVER, R. G. & ECCLES, M. R. 1995a. Mutation of PAX2 in two siblings with renal-coloboma syndrome. *Hum Mol Genet,* 4**,** 2183-4.

SANYANUSIN, P., SCHIMMENTI, L. A., MCNOE, L. A., WARD, T. A., PIERPONT, M. E., SULLIVAN, M. J., DOBYNS, W. B. & ECCLES, M. R. 1995b. Mutation of the PAX2 gene in a family with optic nerve colobomas, renal anomalies and vesicoureteral reflux. *Nat Genet,* 9**,** 358-64.

SATO, S., IKEDA, K., SHIOI, G., NAKAO, K., YAJIMA, H. & KAWAKAMI, K. 2012. Regulation of Six1 expression by evolutionarily conserved enhancers in tetrapods. *Dev Biol,* 368**,** 95-108.

SATO, S., IKEDA, K., SHIOI, G., OCHI, H., OGINO, H., YAJIMA, H. & KAWAKAMI, K. 2010. Conserved expression of mouse Six1 in the pre-placodal region (PPR) and identification of an enhancer for the rostral PPR. *Dev Biol,* 344**,** 158-71.

SAUKA-SPENGLER, T. & BRONNER-FRASER, M. 2008. A gene regulatory network orchestrates neural crest formation. *Nat Rev Mol Cell Biol,* 9**,** 557-68.

SCHAFER, J. & STRIMMER, K. 2005a. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics,* 21**,** 754-64.

SCHAFER, J. & STRIMMER, K. 2005b. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat Appl Genet Mol Biol,* 4**,** Article32.

SCHEPERS, G. E., BULLEJOS, M., HOSKING, B. M. & KOOPMAN, P. 2000. Cloning and characterisation of the Sry-related transcription factor gene Sox8. *Nucleic Acids Res,* 28**,** 1473-80.

SCHIMMANG, T. 2007. Expression and functions of FGF ligands during early otic development. *Int J Dev Biol,* 51**,** 473-81.

SCHIMMENTI, L. A., CUNLIFFE, H. E., MCNOE, L. A., WARD, T. A., FRENCH, M. C., SHIM, H. H., ZHANG, Y. H., PROESMANS, W., LEYS, A., BYERLY, K. A., BRADDOCK, S. R., MASUNO, M., IMAIZUMI, K., DEVRIENDT, K. & ECCLES, M. R. 1997. Further delineation of renal-coloboma syndrome in patients with extreme variability of phenotype and identical PAX2 mutations. *Am J Hum Genet,* 60**,** 869-78.

SCHLANGE, T., ANDREE, B., ARNOLD, H. & BRAND, T. 2000. Expression analysis of the chicken homologue of CITED2 during early stages of embryonic development. *Mech Dev,* 98**,** 157-60.

SCHLITT, T. & BRAZMA, A. 2007. Current approaches to gene regulatory network modelling. *BMC Bioinformatics,* 8 Suppl 6**,** S9.

SCHLOSSER, G. 2006. Induction and specification of cranial placodes. *Dev Biol,* 294**,** 303-51.

SCHLOSSER, G. 2010. Making Senses: Development of Vertebrate Cranial Placodes. *International Review of Cell and Molecular Biology,* 283**,** 129-234.

SCHLOSSER, G. & AHRENS, K. 2004. Molecular anatomy of placode development in Xenopus laevis. *Dev Biol,* 271**,** 439-66.

SCHLOSSER, G. & NORTHCUTT, R. G. 2000. Development of neurogenic placodes in Xenopus laevis. *J Comp Neurol,* 418**,** 121-46.

SCHMIDT, D., SCHWALIE, P. C., WILSON, M. D., BALLESTER, B., GONCALVES, A., KUTTER, C., BROWN, G. D., MARSHALL, A., FLICEK, P. & ODOM, D. T. 2012. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell,* 148**,** 335-48.

SCHWARTZ, S., KENT, W. J., SMIT, A., ZHANG, Z., BAERTSCH, R., HARDISON, R. C., HAUSSLER, D. & MILLER, W. 2003. Human-mouse alignments with BLASTZ. *Genome Res,* 13**,** 103-7.

SCHWARTZ, Y. B. & PIRROTTA, V. 2007. Polycomb silencing mechanisms and the management of genomic programmes. *Nat Rev Genet,* 8**,** 9-22.

SEITAN, V. C., FAURE, A. J., ZHAN, Y., MCCORD, R. P., LAJOIE, B. R., ING-SIMMONS, E., LENHARD, B., GIORGETTI, L., HEARD, E., FISHER, A. G., FLICEK, P., DEKKER, J. & MERKENSCHLAGER, M. 2013. Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Res,* 23**,** 2066-77.

SERANDOUR, A. A., AVNER, S., PERCEVAULT, F., DEMAY, F., BIZOT, M., LUCCHETTI-MIGANEH, C., BARLOY-HUBLER, F., BROWN, M., LUPIEN, M., METIVIER, R., SALBERT, G. & EECKHOUTE, J. 2011. Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res,* 21**,** 555-65.

SHANNON, P., MARKIEL, A., OZIER, O., BALIGA, N. S., WANG, J. T., RAMAGE, D., AMIN, N., SCHWIKOWSKI, B. & IDEKER, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res,* 13**,** 2498-504.

SHEN-ORR, S. S., MILO, R., MANGAN, S. & ALON, U. 2002. Network motifs in the transcriptional regulation network of Escherichia coli. *Nat Genet,* 31**,** 64-8.

SHEN, Y., YUE, F., MCCLEARY, D. F., YE, Z., EDSALL, L., KUAN, S., WAGNER, U., DIXON, J., LEE, L., LOBANENKOV, V. V. & REN, B. 2012. A map of the cis-regulatory sequences in the mouse genome. *Nature,* 488**,** 116-20.

SHENG, G. & STERN, C. D. 1999. Gata2 and Gata3: novel markers for early embryonic polarity and for non-neural ectoderm in the chick embryo. *Mech Dev,* 87**,** 213-6.

SHLYUEVA, D., STAMPFEL, G. & STARK, A. 2014. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet,* 15**,** 272-86.

SIMEONE, A., ACAMPORA, D., GULISANO, M., STORNAIUOLO, A. & BONCINELLI, E. 1992. Nested expression domains of four homeobox genes in developing rostral brain. *Nature,* 358**,** 687-90.

SIMEONE, A., ACAMPORA, D., MALLAMACI, A., STORNAIUOLO, A., D'APICE, M. R., NIGRO, V. & BONCINELLI, E. 1993. A vertebrate gene related to orthodenticle contains a homeodomain of the bicoid class and demarcates anterior neuroectoderm in the gastrulating mouse embryo. *EMBO J,* 12**,** 2735-47.

SIMON, J. A. & KINGSTON, R. E. 2009. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol,* 10**,** 697-708.

SINHA, S., BLANCHETTE, M. & TOMPA, M. 2004. PhyME: a probabilistic algorithm for finding motifs in sets of orthologous sequences. *BMC Bioinformatics,* 5**,** 170.

SINHA, S. & TOMPA, M. 2003. YMF: A program for discovery of novel transcription factor binding sites by statistical overrepresentation. *Nucleic Acids Res,* 31**,** 3586-8.

SMITH, A. N., MILLER, L. A., SONG, N., TAKETO, M. M. & LANG, R. A. 2005. The duality of beta-catenin function: a requirement in lens morphogenesis and signaling suppression of lens fate in periocular ectoderm. *Dev Biol,* 285**,** 477-89.

SMITH, T. F. & WATERMAN, M. S. 1981. Identification of common molecular subsequences. *J Mol Biol,* 147**,** 195-7.

SOFUEVA, S., YAFFE, E., CHAN, W. C., GEORGOPOULOU, D., VIETRI RUDAN, M., MIRA-BONTENBAL, H., POLLARD, S. M., SCHROTH, G. P., TANAY, A. & HADJUR, S. 2013. Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J,* 32**,** 3119-29.

SOLOMON, K. S. 2003. Zebrafish foxi1 mediates otic placode formation and jaw development. *Development,* 130**,** 929-940.

SOLOMON, K. S. & FRITZ, A. 2002. Concerted action of two dlx paralogs in sensory placode formation. *Development,* 129**,** 3127-36.

SOLOMON, K. S., KUDOH, T., DAWID, I. B. & FRITZ, A. 2003a. Zebrafish foxi1 mediates otic placode formation and jaw development. *Development,* 130**,** 929-40.

SOLOMON, K. S., KWAK, S. J. & FRITZ, A. 2004. Genetic interactions underlying otic placode induction and formation. *Dev Dyn,* 230**,** 419-33.

SOLOMON, K. S., LOGSDON, J. M., JR. & FRITZ, A. 2003b. Expression and phylogenetic analyses of three zebrafish FoxI class genes. *Dev Dyn,* 228**,** 301-7.

SONG, M. H., KWON, T. J., KIM, H. R., JEON, J. H., BAEK, J. I., LEE, W. S., KIM, U. K. & CHOI, J. Y. 2013. Mutational analysis of EYA1, SIX1 and SIX5 genes and strategies for management of hearing loss in patients with BOR/BO syndrome. *PLoS One,* 8**,** e67236.

SOSINSKY, A., HONIG, B., MANN, R. S. & CALIFANO, A. 2007. Discovering transcriptional regulatory regions in Drosophila by a nonalignment method for phylogenetic footprinting. *Proc Natl Acad Sci U S A,* 104**,** 6305-10.

SPITZ, F. & FURLONG, E. E. 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet,* 13**,** 613-26.

SPLINTER, E., HEATH, H., KOOREN, J., PALSTRA, R. J., KLOUS, P., GROSVELD, F., GALJART, N. & DE LAAT, W. 2006. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev,* 20**,** 2349-54.

STAEHLING-HAMPTON, K., PROLL, S., PAEPER, B. W., ZHAO, L., CHARMLEY, P., BROWN, A., GARDNER, J. C., GALAS, D., SCHATZMAN, R. C., BEIGHTON, P., PAPAPOULOS, S., HAMERSMA, H. & BRUNKOW, M. E. 2002. A 52-kb deletion in the SOST-MEOX1 intergenic region on 17q12-q21 is associated with van Buchem disease in the Dutch population. *Am J Med Genet,* 110**,** 144-52.

STATHOPOULOS, A. & LEVINE, M. 2005. Genomic regulatory networks and animal development. *Dev Cell,* 9**,** 449-62.

STEDMAN, W., KANG, H., LIN, S., KISSIL, J. L., BARTOLOMEI, M. S. & LIEBERMAN, P. M. 2008. Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *EMBO J,* 27**,** 654-66.

STEVENTON, B., MAYOR, R. & STREIT, A. 2012. Mutual repression between Gbx2 and Otx2 in sensory placodes reveals a general mechanism for ectodermal patterning. *Dev Biol,* 367**,** 55-65.

STOLOVITZKY, G., MONROE, D. & CALIFANO, A. 2007. Dialogue on reverse-engineering assessment and methods: the DREAM of high-throughput pathway inference. *Ann N Y Acad Sci,* 1115**,** 1-22.

STOLTE, D., HUANG, R. & CHRIST, B. 2002. Spatial and temporal pattern of Fgf-8 expression during chicken development. *Anat Embryol (Berl),* 205**,** 1-6.

STREIT, A. 2002. Extensive Cell Movements Accompany Formation of the Otic Placode. *Developmental Biology,* 249**,** 237-254.

STREIT, A. 2007. The preplacodal region: an ectodermal domain with multipotential progenitors that contribute to sense organs and cranial sensory ganglia. *Int J Dev Biol,* 51**,** 447-61.

STREIT, A. 2008. The cranial sensory nervous system: specification of sensory progenitors and placodes. *StemBook*.
STREIT, A., BERLINER, A. J., PAPANAYOTOU, C., SIRULNIK, A. & STERN, C. D. 2000. Initiation of neural induction by FGF signalling before gastrulation. *Nature,* 406**,** 74-8.
SU, G., KUCHINSKY, A., MORRIS, J. H., STATES, D. J. & MENG, F. 2010. GLay: community structure analysis of biological networks. *Bioinformatics,* 26**,** 3135-7.
SUN RHODES, L. S. & MERZDORF, C. S. 2006. The zic1 gene is expressed in chick somites but not in migratory neural crest. *Gene Expr Patterns,* 6**,** 539-45.
SUN, S. K., DEE, C. T., TRIPATHI, V. B., RENGIFO, A., HIRST, C. S. & SCOTTING, P. J. 2007. Epibranchial and otic placodes are induced by a common Fgf signal, but their subsequent development is independent. *Dev Biol,* 303**,** 675-86.
SWEETMAN, D., SMITH, T. G., FARRELL, E. R. & MUNSTERBERG, A. 2005. Expression of csal1 in pre limb-bud chick embryos. *Int J Dev Biol,* 49**,** 427-30.
SZABO, P., TANG, S. H., RENTSENDORJ, A., PFEIFER, G. P. & MANN, J. R. 2000. Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. *Curr Biol,* 10**,** 607-10.
SZUTORISZ, H., DILLON, N. & TORA, L. 2005. The role of enhancers as centres for general transcription factor recruitment. *Trends Biochem Sci,* 30**,** 593-9.
TAM, P. P. 1989. Regionalisation of the mouse embryonic ectoderm: allocation of prospective ectodermal tissues during gastrulation. *Development,* 107**,** 55-67.
TAMBALO, M. 2015. Towards a gene regulatory network for otic and epibranchial specification. *A thesis submitted to the King's College London's Higher Degree Office in partial fulfillment for the Degree of Doctor of Philosophy*.
TAYLOR, W. R. 1988. A flexible method to align large numbers of biological sequences. *J Mol Evol,* 28**,** 161-9.
TEGNER, J., YEUNG, M. K., HASTY, J. & COLLINS, J. J. 2003. Reverse engineering gene networks: integrating genetic perturbations with dynamical modeling. *Proc Natl Acad Sci U S A,* 100**,** 5944-9.
TEKIN, M., HISMI, B. O., FITOZ, S., OZDAG, H., CENGIZ, F. B., SIRMACI, A., ASLAN, I., INCEOGLU, B., YUKSEL-KONUK, E. B., YILMAZ, S. T., YASUN, O. & AKAR, N. 2007. Homozygous mutations in fibroblast growth factor 3 are associated with a new form of syndromic deafness characterized by inner ear agenesis, microtia, and microdontia. *Am J Hum Genet,* 80**,** 338-44.
TERZIC, J., MULLER, C., GAJOVIC, S. & SARAGA-BABIC, M. 1998. Expression of PAX2 gene during human development. *Int J Dev Biol,* 42**,** 701-7.
THIJS, G., LESCOT, M., MARCHAL, K., ROMBAUTS, S., DE MOOR, B., ROUZE, P. & MOREAU, Y. 2001. A higher-order background model improves the detection of promoter regulatory elements by Gibbs sampling. *Bioinformatics,* 17**,** 1113-22.
THOMAS-CHOLLIER, M., SAND, O., TURATSINZE, J. V., JANKY, R., DEFRANCE, M., VERVISCH, E., BROHEE, S. & VAN HELDEN, J. 2008. RSAT: regulatory sequence analysis tools. *Nucleic Acids Res,* 36**,** W119-W127.
THOMPSON, J. D., GIBSON, T. J. & HIGGINS, D. G. 2002. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics,* Chapter 2**,** Unit 2 3.
THURMAN, R. E., RYNES, E., HUMBERT, R., VIERSTRA, J., MAURANO, M. T., HAUGEN, E., SHEFFIELD, N. C., STERGACHIS, A. B., WANG, H., VERNOT, B., GARG, K., JOHN, S., SANDSTROM, R., BATES, D., BOATMAN, L., CANFIELD, T. K., DIEGEL, M., DUNN, D., EBERSOL, A. K., FRUM, T., GISTE, E., JOHNSON, A. K., JOHNSON, E. M., KUTYAVIN, T., LAJOIE, B., LEE, B. K., LEE, K., LONDON, D., LOTAKIS, D., NEPH, S., NERI, F., NGUYEN, E. D., QU, H., REYNOLDS, A. P., ROACH, V., SAFI, A.,

SANCHEZ, M. E., SANYAL, A., SHAFER, A., SIMON, J. M., SONG, L., VONG, S., WEAVER, M., YAN, Y., ZHANG, Z., ZHANG, Z., LENHARD, B., TEWARI, M., DORSCHNER, M. O., HANSEN, R. S., NAVAS, P. A., STAMATOYANNOPOULOS, G., IYER, V. R., LIEB, J. D., SUNYAEV, S. R., AKEY, J. M., SABO, P. J., KAUL, R., FUREY, T. S., DEKKER, J., CRAWFORD, G. E. & STAMATOYANNOPOULOS, J. A. 2012. The accessible chromatin landscape of the human genome. *Nature,* 489**,** 75-82.

TIE, F., BANERJEE, R., SAIAKHOVA, A. R., HOWARD, B., MONTEITH, K. E., SCACHERI, P. C., COSGROVE, M. S. & HARTE, P. J. 2014. Trithorax monomethylates histone H3K4 and interacts directly with CBP to promote H3K27 acetylation and antagonize Polycomb silencing. *Development,* 141**,** 1129-39.

TIE, F., BANERJEE, R., STRATTON, C. A., PRASAD-SINHA, J., STEPANIK, V., ZLOBIN, A., DIAZ, M. O., SCACHERI, P. C. & HARTE, P. J. 2009. CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development,* 136**,** 3131-41.

TJIAN, R. 1978. The binding site on SV40 DNA for a T antigen-related protein. *Cell,* 13**,** 165-79.

TOMPA, M., LI, N., BAILEY, T. L., CHURCH, G. M., DE MOOR, B., ESKIN, E., FAVOROV, A. V., FRITH, M. C., FU, Y., KENT, W. J., MAKEEV, V. J., MIRONOV, A. A., NOBLE, W. S., PAVESI, G., PESOLE, G., REGNIER, M., SIMONIS, N., SINHA, S., THIJS, G., VAN HELDEN, J., VANDENBOGAERT, M., WENG, Z., WORKMAN, C., YE, C. & ZHU, Z. 2005. Assessing computational tools for the discovery of transcription factor binding sites. *Nat Biotechnol,* 23**,** 137-44.

TORRES, M. & GIRALDEZ, F. 1998. The development of the vertebrate inner ear. *Mech Dev,* 71**,** 5-21.

TORRES, M., GOMEZ-PARDO, E. & GRUSS, P. 1996. Pax2 contributes to inner ear patterning and optic nerve trajectory. *Development,* 122**,** 3381-91.

TOUR, E., PILLEMER, G., GRUENBAUM, Y. & FAINSOD, A. 2001. The two Xenopus Gbx2 genes exhibit similar, but not identical expression patterns and can affect head formation. *FEBS Lett,* 507**,** 205-9.

TRAPNELL, C., WILLIAMS, B. A., PERTEA, G., MORTAZAVI, A., KWAN, G., VAN BAREN, M. J., SALZBERG, S. L., WOLD, B. J. & PACHTER, L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology,* 28**,** 511-U174.

TSUDA, M., TAKAHASHI, S., TAKAHASHI, Y. & ASAHARA, H. 2003. Transcriptional co-activators CREB-binding protein and p300 regulate chondrocyte-specific gene expression via association with Sox9. *J Biol Chem,* 278**,** 27224-9.

TURATSINZE, J. V., THOMAS-CHOLLIER, M., DEFRANCE, M. & VAN HELDEN, J. 2008. Using RSAT to scan genome sequences for transcription factor binding sites and cis-regulatory modules. *Nature Protocols,* 3**,** 1578-1588.

UCHIKAWA, M., ISHIDA, Y., TAKEMOTO, T., KAMACHI, Y. & KONDOH, H. 2003. Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals. *Dev Cell,* 4**,** 509-19.

URETA-VIDAL, A., ETTWILLER, L. & BIRNEY, E. 2003. Comparative genomics: genome-wide analysis in metazoan eukaryotes. *Nat Rev Genet,* 4**,** 251-62.

URNESS, L. D., PAXTON, C. N., WANG, X., SCHOENWOLF, G. C. & MANSOUR, S. L. 2010. FGF signaling regulates otic placode induction and refinement by controlling both ectodermal target genes and hindbrain Wnt8a. *Dev Biol,* 340**,** 595-604.

VALOUEV, A., JOHNSON, D. S., SUNDQUIST, A., MEDINA, C., ANTON, E., BATZOGLOU, S., MYERS, R. M. & SIDOW, A. 2008. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat Methods,* 5**,** 829-34.

VAN CAMP, G. & SMITH, R. J. H. 2014. Hereditary Hearing Loss Homepage.

VAN HELDEN, J., ANDRE, B. & COLLADO-VIDES, J. 1998. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J Mol Biol,* 281**,** 827-42.

VAN HELDEN, J., RIOS, A. F. & COLLADO-VIDES, J. 2000. Discovering regulatory elements in non-coding sequences by analysis of spaced dyads. *Nucleic Acids Res,* 28**,** 1808-18.

VAVOURI, T., MCEWEN, G. K., WOOLFE, A., GILKS, W. R. & ELGAR, G. 2006. Defining a genomic radius for long-range enhancer action: duplicated conserved non-coding elements hold the key. *Trends Genet,* 22**,** 5-10.

VENDRELL, V., CARNICERO, E., GIRALDEZ, F., ALONSO, M. T. & SCHIMMANG, T. 2000. Induction of inner ear fate by FGF3. *Development,* 127**,** 2011-9.

VERWOERD, C. D. & VAN OOSTROM, C. G. 1979. Cephalic neural crest and placodes. *Adv Anat Embryol Cell Biol,* 58**,** 1-75.

VISEL, A., AKIYAMA, J.A., SHOUKRY, M., AFZAL, V. & RUBIN, E. M. 2009. Functional autonomy of distant-acting human enhancers. *Genomics,* 93**,** 509–513.

VISEL, A., BLOW, M. J., LI, Z., ZHANG, T., AKIYAMA, J. A., HOLT, A., PLAJZER-FRICK, I., SHOUKRY, M., WRIGHT, C., CHEN, F., AFZAL, V., REN, B., RUBIN, E. M. & PENNACCHIO, L. A. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature,* 457**,** 854-8.

VISEL, A., TAHER, L., GIRGIS, H., MAY, D. & GOLONZHKA, O. 2013. A high resolution enhancer atlas of the developing telencephalon. *Cell,* 152**,** 895-908.

VON BUBNOFF, A., SCHMIDT, J. E. & KIMELMAN, D. 1996. The Xenopus laevis homeobox gene Xgbx-2 is an early marker of anteroposterior patterning in the ectoderm. *Mech Dev,* 54**,** 149-60.

WAGNER, A. 1999. Genes regulated cooperatively by one or more transcription factors and their identification in whole eukaryotic genomes. *Bioinformatics,* 15**,** 776-84.

WAKAMATSU, Y. 2011. Mutual repression between Pax3 and Pax6 is involved in the positioning of ophthalmic trigeminal placode in avian embryo. *Dev Growth Differ,* 53**,** 994-1003.

WALHOUT, A. J. 2006. Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. *Genome Res,* 16**,** 1445-54.

WALLACE, J. A. & FELSENFELD, G. 2007. We gather together: insulators and genome organization. *Curr Opin Genet Dev,* 17**,** 400-7.

WALSHE, J., MAROON, H., MCGONNELL, I. M., DICKSON, C. & MASON, I. 2002. Establishment of hindbrain segmental identity requires signaling by FGF3 and FGF8. *Curr Biol,* 12**,** 1117-23.

WALSHE, J. & MASON, I. 2000. Expression of FGFR1, FGFR2 and FGFR3 during early neural development in the chick embryo. *Mech Dev,* 90**,** 103-10.

WAMSTAD, J. A., ALEXANDER, J. M., TRUTY, R. M., SHRIKUMAR, A., LI, F., EILERTSON, K. E., DING, H., WYLIE, J. N., PICO, A. R., CAPRA, J. A., ERWIN, G., KATTMAN, S. J., KELLER, G. M., SRIVASTAVA, D., LEVINE, S. S., POLLARD, K. S., HOLLOWAY, A. K., BOYER, L. A. & BRUNEAU, B. G. 2012. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell,* 151**,** 206-20.

WANG, H., MAURANO, M. T., QU, H., VARLEY, K. E., GERTZ, J., PAULI, F., LEE, K., CANFIELD, T., WEAVER, M., SANDSTROM, R., THURMAN, R. E.,

KAUL, R., MYERS, R. M. & STAMATOYANNOPOULOS, J. A. 2012. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res,* 22**,** 1680-8.

WANG, J., LUNYAK, V. V. & JORDAN, I. K. 2013. BroadPeak: a novel algorithm for identifying broad peaks in diffuse ChIP-seq datasets. *Bioinformatics,* 29**,** 492-3.

WANG, Y., JOSHI, T., ZHANG, X. S., XU, D. & CHEN, L. 2006. Inferring gene regulatory networks from multiple microarray datasets. *Bioinformatics,* 22**,** 2413-20.

WANG, Z., YANG, F., HO, D. W., SWIFT, S., TUCKER, A. & LIU, X. 2008a. Stochastic dynamic modeling of short gene expression time-series data. *IEEE Trans Nanobioscience,* 7**,** 44-55.

WANG, Z., ZANG, C., ROSENFELD, J. A., SCHONES, D. E., BARSKI, A., CUDDAPAH, S., CUI, K., ROH, T. Y., PENG, W., ZHANG, M. Q. & ZHAO, K. 2008b. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet,* 40**,** 897-903.

WARNES, G. R., BOLKER, B., LUMLEY, T 2015. gplots: Various R programming tools for plotting data.

WARREN, M., WANG, W., SPIDEN, S., CHEN-MURCHIE, D., TANNAHILL, D., STEEL, K. P. & BRADLEY, A. 2007. A Sall4 mutant mouse model useful for studying the role of Sall4 in early embryonic development and organogenesis. *Genesis,* 45**,** 51-8.

WASIM, M., CARLET, M., MANSHA, M., GREIL, R., PLONER, C., TROCKENBACHER, A., RAINER, J. & KOFLER, R. 2010. PLZF/ZBTB16, a glucocorticoid response gene in acute lymphoblastic leukemia, interferes with glucocorticoid-induced apoptosis. *J Steroid Biochem Mol Biol,* 120**,** 218-27.

WASSARMAN, K. M., LEWANDOSKI, M., CAMPBELL, K., JOYNER, A. L., RUBENSTEIN, J. L., MARTINEZ, S. & MARTIN, G. R. 1997. Specification of the anterior hindbrain and establishment of a normal mid/hindbrain organizer is dependent on Gbx2 gene function. *Development,* 124**,** 2923-34.

WASSERMAN, W. W. & SANDELIN, A. 2004. Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet,* 5**,** 276-87.

WATANABE, T., SATO, Y., SAITO, D., TADOKORO, R. & TAKAHASHI, Y. 2009. EphrinB2 coordinates the formation of a morphological boundary and cell epithelialization during somite segmentation. *Proc Natl Acad Sci U S A,* 106**,** 7467-72.

WATT, F. M., ESTRACH, S. & AMBLER, C. A. 2008. Epidermal Notch signalling: differentiation, cancer and adhesion. *Curr Opin Cell Biol,* 20**,** 171-9.

WEI, G. H., BADIS, G., BERGER, M. F., KIVIOJA, T., PALIN, K., ENGE, M., BONKE, M., JOLMA, A., VARJOSALO, M., GEHRKE, A. R., YAN, J., TALUKDER, S., TURUNEN, M., TAIPALE, M., STUNNENBERG, H. G., UKKONEN, E., HUGHES, T. R., BULYK, M. L. & TAIPALE, J. 2010. Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *EMBO J,* 29**,** 2147-60.

WEI, X. C., DOHKAN, J., KISHI, H., WU, C. X., KONDO, S. & MURAGUCHI, A. 2005. Characterization of the proximal enhancer element and transcriptional regulatory factors for murine recombination activating gene-2. *Eur J Immunol,* 35**,** 612-621.

WENDT, K. S., YOSHIDA, K., ITOH, T., BANDO, M., KOCH, B., SCHIRGHUBER, E., TSUTSUMI, S., NAGAE, G., ISHIHARA, K., MISHIRO, T., YAHATA, K., IMAMOTO, F., ABURATANI, H., NAKAO, M., IMAMOTO, N., MAESHIMA, K., SHIRAHIGE, K. & PETERS, J. M. 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature,* 451**,** 796-801.

WEST, A. G., GASZNER, M. & FELSENFELD, G. 2002. Insulators: many functions, many mechanisms. *Genes Dev,* 16**,** 271-88.

WHITLOCK, K. E. & WESTERFIELD, M. 2000. The olfactory placodes of the zebrafish form by convergence of cellular fields at the edge of the neural plate. *Development,* 127**,** 3645-53.

WHYTE, W. A., BILODEAU, S., ORLANDO, D. A., HOKE, H. A., FRAMPTON, G. M., FOSTER, C. T., COWLEY, S. M. & YOUNG, R. A. 2012. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature,* 482**,** 221-5.

WHYTE, W. A., ORLANDO, D. A., HNISZ, D., ABRAHAM, B. J., LIN, C. Y., KAGEY, M. H., RAHL, P. B., LEE, T. I. & YOUNG, R. A. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell,* 153**,** 307-19.

WILBANKS, E. G. & FACCIOTTI, M. T. 2010. Evaluation of algorithm performance in ChIP-seq peak detection. *PLoS One,* 5**,** e11471.

WODA, J. M. 2003. Dlx proteins position the neural plate border and determine adjacent cell fates. *Development,* 130**,** 331-342.

WOOLFE, A., GOODSON, M., GOODE, D. K., SNELL, P., MCEWEN, G. K., VAVOURI, T., SMITH, S. F., NORTH, P., CALLAWAY, H., KELLY, K., WALTER, K., ABNIZOVA, I., GILKS, W., EDWARDS, Y. J., COOKE, J. E. & ELGAR, G. 2005. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol,* 3**,** e7.

WORKMAN, C. T. & STORMO, G. D. 2000. ANN-Spec: a method for discovering transcription factor binding sites with improved specificity. *Pac Symp Biocomput***,** 467-78.

WRIGHT, T. J. & MANSOUR, S. L. 2003. Fgf3 and Fgf10 are required for mouse otic placode induction. *Development,* 130**,** 3379-90.

XIAO, T., WALLACE, J. & FELSENFELD, G. 2011. Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Mol Cell Biol,* 31**,** 2174-83.

XIAO, Y. & SEGAL, M. R. 2009. Identification of yeast transcriptional regulation networks using multivariate random forests. *PLoS Comput Biol,* 5**,** e1000414.

XIE, X., MIKKELSEN, T. S., GNIRKE, A., LINDBLAD-TOH, K., KELLIS, M. & LANDER, E. S. 2007. Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. *Proc Natl Acad Sci U S A,* 104**,** 7145-50.

XU, J., POPE, S. D., JAZIREHI, A. R., ATTEMA, J. L., PAPATHANASIOU, P., WATTS, J. A., ZARET, K. S., WEISSMAN, I. L. & SMALE, S. T. 2007. Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. *Proc Natl Acad Sci U S A,* 104**,** 12377-82.

XU, J., WATTS, J. A., POPE, S. D., GADUE, P., KAMPS, M., PLATH, K., ZARET, K. S. & SMALE, S. T. 2009. Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells. *Genes Dev,* 23**,** 2824-38.

XU, P. X., ADAMS, J., PETERS, H., BROWN, M. C., HEANEY, S. & MAAS, R. 1999. Eya1-deficient mice lack ears and kidneys and show abnormal apoptosis of organ primordia. *Nat Genet,* 23**,** 113-7.

XU, X., DUDE, C. M. & BAKER, C. V. 2008. Fine-grained fate maps for the ophthalmic and maxillomandibular trigeminal placodes in the chick embryo. *Dev Biol,* 317**,** 174-86.

YADA, K., DO, E., SAKAKIBARA, S., OHSAKI, E., ITO, E., WATANABE, S. & UEDA, K. 2006. KSHV RTA induces a transcriptional repressor, HEY1 that represses rta promoter. *Biochem Biophys Res Commun,* 345**,** 410-8.

YANEZ-CUNA, J. O., DINH, H. Q., KVON, E. Z., SHLYUEVA, D. & STARK, A. 2012. Uncovering cis-regulatory sequence requirements for context-specific transcription factor binding. *Genome Res,* 22**,** 2018-30.

YANG, L., O'NEILL, P., MARTIN, K., MAASS, J. C., VASSILEV, V., LADHER, R. & GROVES, A. K. 2013. Analysis of FGF-dependent and FGF-independent pathways in otic placode induction. *PLoS One,* 8**,** e55011.

YE, T., KREBS, A. R., CHOUKRALLAH, M. A., KEIME, C., PLEWNIAK, F., DAVIDSON, I. & TORA, L. 2011. seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res,* 39**,** e35.

YIP, K. Y., ALEXANDER, R. P., YAN, K. K. & GERSTEIN, M. 2010. Improved reconstruction of in silico gene regulatory networks by integrating knockout and perturbation data. *PLoS One,* 5**,** e8121.

YU, H. & GERSTEIN, M. 2006. Genomic analysis of the hierarchical structure of regulatory networks. *Proc Natl Acad Sci U S A,* 103**,** 14724-31.

YU, J., SMITH, V. A., WANG, P. P., HARTEMINK, A. J. & JARVIS, E. D. 2004. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics,* 20**,** 3594-603.

YUAN, G. C., LIU, Y. J., DION, M. F., SLACK, M. D., WU, L. F., ALTSCHULER, S. J. & RANDO, O. J. 2005. Genome-scale identification of nucleosome positions in S. cerevisiae. *Science,* 309**,** 626-30.

YUSUFZAI, T. M., TAGAMI, H., NAKATANI, Y. & FELSENFELD, G. 2004. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol Cell,* 13**,** 291-8.

ZANG, C., SCHONES, D. E., ZENG, C., CUI, K., ZHAO, K. & PENG, W. 2009. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics,* 25**,** 1952-8.

ZARET, K. S. & CARROLL, J. S. 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev,* 25**,** 2227-41.

ZELARAYAN, L. C., VENDRELL, V., ALVAREZ, Y., DOMINGUEZ-FRUTOS, E., THEIL, T., ALONSO, M. T., MACONOCHIE, M. & SCHIMMANG, T. 2007. Differential requirements for FGF3, FGF8 and FGF10 during inner ear development. *Dev Biol,* 308**,** 379-91.

ZENTNER, G. E., TESAR, P. J. & SCACHERI, P. C. 2011. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res,* 21**,** 1273-83.

ZHANG, B., DAY, D. S., HO, J. W., SONG, L., CAO, J., CHRISTODOULOU, D., SEIDMAN, J. G., CRAWFORD, G. E., PARK, P. J. & PU, W. T. 2013a. A dynamic H3K27ac signature identifies VEGFA-stimulated endothelial enhancers and requires EP300 activity. *Genome Res,* 23**,** 917-27.

ZHANG, H., JIAO, W., SUN, L., FAN, J., CHEN, M., WANG, H., XU, X., SHEN, A., LI, T., NIU, B., GE, S., LI, W., CUI, J., WANG, G., SUN, J., FAN, X., HU, X., MRSNY, R. J., HOFFMAN, A. R. & HU, J. F. 2013b. Intrachromosomal looping is required for activation of endogenous pluripotency genes during reprogramming. *Cell Stem Cell,* 13**,** 30-5.

ZHANG, X., ZHAO, X. M., HE, K., LU, L., CAO, Y., LIU, J., HAO, J. K., LIU, Z. P. & CHEN, L. 2012. Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information. *Bioinformatics,* 28**,** 98-104.

ZHANG, Y., KNOSP, B. M., MACONOCHIE, M., FRIEDMAN, R. A. & SMITH, R. J. 2004. A comparative study of Eya1 and Eya4 protein function and its implication in branchio-oto-renal syndrome and DFNA10. *J Assoc Res Otolaryngol,* 5**,** 295-304.

ZHANG, Y., LIU, T., MEYER, C. A., EECKHOUTE, J., JOHNSON, D. S., BERNSTEIN, B. E., NUSBAUM, C., MYERS, R. M., BROWN, M., LI, W. & LIU, X. S. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol,* 9**,** R137.

ZHENG, W., HUANG, L., WEI, Z. B., SILVIUS, D., TANG, B. & XU, P. X. 2003. The role of Six1 in mammalian auditory system development. *Development,* 130**,** 3989-4000.

ZHU, L. J., GAZIN, C., LAWSON, N. D., PAGES, H., LIN, S. M., LAPOINTE, D. S. & GREEN, M. R. 2010. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *Bmc Bioinformatics,* 11.

ZOU, D., SILVIUS, D., FRITZSCH, B. & XU, P. X. 2004. Eya1 and Six1 are essential for early steps of sensory neurogenesis in mammalian cranial placodes. *Development,* 131**,** 5561-72.

ZUIN, J., DIXON, J. R., VAN DER REIJDEN, M. I., YE, Z., KOLOVOS, P., BROUWER, R. W., VAN DE CORPUT, M. P., VAN DE WERKEN, H. J., KNOCH, T. A., VAN, I. W. F., GROSVELD, F. G., REN, B. & WENDT, K. S. 2014. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci U S A,* 111**,** 996-1001.

ZUNIGA, A., MICHOS, O., SPITZ, F., HARAMIS, A. P., PANMAN, L., GALLI, A., VINTERSTEN, K., KLASEN, C., MANSFIELD, W., KUC, S., DUBOULE, D., DONO, R. & ZELLER, R. 2004. Mouse limb deformity mutations disrupt a global control region within the large regulatory landscape required for Gremlin expression. *Genes Dev,* 18**,** 1553-64.

ZYGAR, C. A., COOK, T. L. & GRAINGER, R. M., JR. 1998. Gene activation during early stages of lens induction in Xenopus. *Development,* 125**,** 3509-19.

8. Appendix

8.1 DREiVe-predicted enhancers for FGF response genes in human, chicken and mouse

Gene	Human enhancers	Chicken enhancers	Mouse enhancers
SPRY1	chr4:124620858-124621170	chr4:52674650-52674961	chr3:37816650-37816961
SPRY1	chr4:124620827-124621129	chr4:52674691-52674993	chr3:37816618-37816920
SPRY1	chr4:124620858-124620940	chr4:52674881-52674961	chr3:37816650-37816730
SPRY1	chr4:124571017-124571298	chr4:52688478-52688744	chr3:37789464-37789749
SPRY1	chr4:124571011-124571269	chr4:52688507-52688750	chr3:37789459-37789720
SPRY1	chr4:124553145-124553334	chr4:52697985-52698177	chr3:37780476-37780666
SPRY1	chr4:124547664-124547852	chr4:52700265-52700439	chr3:37776344-37776531
SPRY1	chr4:124547664-124547758	chr4:52700360-52700439	chr3:37776344-37776437
SPRY1	chr4:124538176-124538473	chr4:52708139-52708432	chr3:37769260-37769567
SPRY1	chr4:124538296-124538458	chr4:52708154-52708316	chr3:37769390-37769552
SPRY1	chr4:124538349-124538407	chr4:52708205-52708263	chr3:37769443-37769501
SPRY1	chr4:124509486-124509596	chr4:52724522-52724632	chr3:37748122-37748232
SPRY1	chr4:124476735-124476922	chr4:52735189-52735377	chr3:37729294-37729479
SPRY1	chr4:124476626-124476850	chr4:52735261-52735484	chr3:37729162-37729407
SPRY1	chr4:124467415-124467635	chr4:52739466-52739686	chr3:37723296-37723515
SPRY1	chr4:124391834-124392013	chr4:52750844-52751029	chr3:37677493-37677673
SPRY1	chr4:124365650-124365719	chr4:52761855-52761924	chr3:37666849-37667051
SPRY1	chr4:124339788-124339941	chr4:52768169-52768322	chr3:37657314-37657466
SPRY1	chr4:124339634-124339915	chr4:52768195-52768476	chr3:37657157-37657440
SPRY1	chr4:124205437-124205716	chr4:52810668-52810943	chr3:37556935-37557209
SPRY1	chr4:124205436-124205651	chr4:52810730-52810944	chr3:37556934-37557148
SPRY1	chr4:124205436-124205611	chr4:52810769-52810944	chr3:37556934-37557108
SPRY1	chr4:124154635-124154755	chr4:52843417-52843536	chr3:37514962-37515084

SPRY1	chr4:124093281-124093426	chr4:52886159-52886301	chr3:37493605-37493750
SPRY1	chr4:124049161-124049401	chr4:52910388-52910632	chr3:37473231-37473462
SPRY1	chr4:124049291-124049363	chr4:52910424-52910498	chr3:37473358-37473430
FOXI3	chr2:88697848-88698115	chr4:85585195-85585454	chr6:70986738-70987003
FOXI3	chr2:88698850-88698930	chr4:85596160-85596222	chr6:70996738-70996838
FOXI3	chr2:88781396-88781688	chr4:85611416-85611709	chr6:70946592-70946872
FOXI3	chr2:88817591-88817720	chr4:85616683-85616813	chr6:70921276-70921403
FOXI3	chr2:88712465-88712695	chr4:85588653-85588879	chr6:70987854-70989999
FOXI3	chr2:88899741-88900023	chr4:85672513-85672802	chr6:70934242-70934881
FOXI3	chr2:88696560-88698250	chr4:85524762-85524858	chr6:71119843-71119947
FOXI3	chr2:88725832-88725900	chr4:85549079-85549150	chr6:71082472-71082543
GBX2	chr2:237087627-237088182	chr7:5121199-5121716	chr1:89927432-89927900
GBX2	chr2:237065691-237065966	chr7:5127818-5128084	chr1:89928534-89929000
GBX2	chr2:237088747-237088990	chr7:5120460-5120703	chr1:89942476-89942720
GBX2	chr2:237071382-237071755	chr7:5126071-5126444	chr1:89925534-89925907
GBX2	chr2:237069996-237070042	chr7:5125589-5125821	chr1:89924292-89924338
GBX2	chr2:237087626-237087958	chr7:5133019-5133156	chr1:89942915-89943125
HESX1	chr3:53393344-53393455	chr12:7007271-7007382	chr14:30469943-30470055
HESX1	chr3:53707027-53707216	chr12:7117917-7118107	chr14:30171892-30172081
HESX1	chr3:53707037-53707142	chr12:7117927-7118032	chr14:30171988-30172071
HESX1	chr3:53707038-53707090	chr12:7117928-7117980	chr14:30172018-30172070
HESX1	chr3:53707047-53707121	chr12:7117937-7118011	chr14:30171987-30172061
HESX1	chr3:53707049-53707142	chr12:7117939-7118032	chr14:30171966-30172059
HESX1	chr3:53707053-53707147	chr12:7117943-7118037	chr14:30171961-30172055
HESX1	chr3:53707070-53707135	chr12:7117960-7118025	chr14:30171973-30172038
HESX1	chr3:53707098-53707151	chr12:7117988-7118041	chr14:30171957-30172010
HESX1	chr3:53707712-53707987	chr12:7118505-7118783	chr14:30171148-30171426
HESX1	chr3:53774295-53774371	chr12:7151755-7151831	chr14:30107780-30107856
HESX1	chr3:53789043-53789174	chr12:7165265-7165380	chr14:30093968-30094091
HESX1	chr3:53796509-53796673	chr12:7167959-7168126	chr14:30089283-30089449

HESX1	chr3:53796576-53796618	chr12:7168029-7168071	chr14:30089338-30089380
HESX1	chr3:53796582-53796644	chr12:7168035-7168097	chr14:30089331-30089374
HESX1	chr3:53796588-53796646	chr12:7168041-7168099	chr14:30089310-30089368
HESX1	chr3:53796602-53796650	chr12:7168055-7168103	chr14:30089306-30089354
HESX1	chr3:53804156-53804405	chr12:7176225-7176483	chr14:30082584-30082840
HESX1	chr3:53816864-53817137	chr12:7185122-7185397	chr14:30069738-30070003
HESX1	chr3:53817035-53817134	chr12:7185295-7185394	chr14:30069741-30069840
HESX1	chr3:53820806-53821093	chr12:7189468-7189739	chr14:30065919-30066206
HESX1	chr3:54158052-54158182	chr12:7311758-7311889	chr14:29720307-29720437
HESX1	chr3:54158056-54158368	chr12:7311762-7312058	chr14:29720134-29720433
HESX1	chr3:54158080-54158320	chr12:7311787-7312030	chr14:29720169-29720409
HESX1	chr3:54215186-54215308	chr12:7344198-7344322	chr14:29664191-29664312
HESX1	chr3:54230548-54230682	chr12:7353413-7353549	chr14:29650672-29650806
HESX1	chr3:54290923-54291191	chr12:7373556-7373802	chr14:29599840-29600092
HESX1	chr3:54291112-54291401	chr12:7373729-7374009	chr14:29599606-29599895
HESX1	chr3:54291117-54291433	chr12:7373734-7374041	chr14:29599574-29599890
HESX1	chr3:54291227-54291411	chr12:7373835-7374019	chr14:29599596-29599780
HESX1	chr3:54291328-54291439	chr12:7373936-7374047	chr14:29599568-29599679
HESX1	chr3:54291366-54291439	chr12:7373974-7374047	chr14:29599568-29599641
HESX1	chr3:54346595-54346753	chr12:7396053-7396279	chr14:29541519-29541677
HESX1	chr3:54358283-54358353	chr12:7402162-7402231	chr14:29530462-29530532
HESX1	chr3:54360005-54360287	chr12:7403228-7403503	chr14:29528735-29529017
HESX1	chr3:54360153-54360287	chr12:7403369-7403503	chr14:29528735-29528869
HESX1	chr3:54653529-54653727	chr12:7509416-7509616	chr14:29313306-29313503
HESX1	chr3:54665674-54665879	chr12:7517457-7517662	chr14:29296513-29296718
HESX1	chr3:54665846-54666016	chr12:7517629-7517929	chr14:29296376-29296546
HESX1	chr3:54665984-54666195	chr12:7517766-7517973	chr14:29296194-29296408
HESX1	chr3:54666121-54666195	chr12:7517904-7517973	chr14:29296194-29296269
HESX1	chr3:54704306-54704582	chr12:7531315-7531590	chr14:29268684-29268957
HESX1	chr3:54704383-54704653	chr12:7531392-7531661	chr14:29268614-29268880

HESX1	chr3:54704458-54704653	chr12:7531467-7531661	chr14:29268614-29268805
HESX1	chr3:54704561-54704653	chr12:7531570-7531661	chr14:29268614-29268704
HESX1	chr3:54767204-54767485	chr12:7556051-7556328	chr14:29204781-29205066
HESX1	chr3:54767206-54767537	chr12:7556053-7556380	chr14:29204729-29205064
HESX1	chr3:54767226-54767541	chr12:7556073-7556318	chr14:29204725-29205044
HESX1	chr3:54767232-54767292	chr12:7556075-7556135	chr14:29204974-29205038
HESX1	chr3:54767257-54767574	chr12:7556100-7556417	chr14:29204692-29205013
HESX1	chr3:54767275-54767337	chr12:7556118-7556180	chr14:29204929-29204991
HESX1	chr3:54767278-54767480	chr12:7556121-7556323	chr14:29204786-29204988
HESX1	chr3:54767376-54767477	chr12:7556219-7556320	chr14:29204789-29204890
HESX1	chr3:54767417-54767702	chr12:7556260-7556546	chr14:29204564-29204849
HESX1	chr3:54767490-54767789	chr12:7556333-7556634	chr14:29204469-29204776
HESX1	chr3:54767496-54767777	chr12:7556339-7556622	chr14:29204481-29204770
HESX1	chr3:54767506-54767562	chr12:7556349-7556405	chr14:29204704-29204760
HESX1	chr3:54767509-54767768	chr12:7556352-7556613	chr14:29204490-29204757
HESX1	chr3:54767512-54767768	chr12:7556355-7556531	chr14:29204579-29204754
HESX1	chr3:54767514-54767666	chr12:7556357-7556510	chr14:29204600-29204752
HESX1	chr3:54767761-54767895	chr12:7556606-7556740	chr14:29204367-29204497
HESX1	chr3:54768868-54768966	chr12:7558613-7558711	chr14:29203033-29203135
HESX1	chr3:54768873-54768966	chr12:7558618-7558711	chr14:29203033-29203130
HESX1	chr3:54789909-54790026	chr12:7572519-7572631	chr14:29180424-29180539
HESX1	chr3:54807477-54807695	chr12:7585490-7585708	chr14:29163987-29164205
HESX1	chr3:54807519-54807602	chr12:7585532-7585615	chr14:29164080-29164163
HESX1	chr3:54807523-54807689	chr12:7585536-7585702	chr14:29163993-29164159
HESX1	chr3:54807616-54807690	chr12:7585629-7585703	chr14:29163992-29164066
HESX1	chr3:54855359-54855455	chr12:7602429-7602523	chr14:29120094-29120190
HESX1	chr3:54879573-54879732	chr12:7613057-7613218	chr14:29098563-29098722
HESX1	chr3:54879675-54879732	chr12:7613160-7613218	chr14:29098563-29098620
HESX1	chr3:54880252-54880298	chr12:7613641-7613687	chr14:29097999-29098045
HESX1	chr3:54892815-54893093	chr12:7621140-7621448	chr14:29085796-29086076

HESX1	chr3:54893038-54893263	chr12:7621364-7621591	chr14:29085624-29085851
HESX1	chr3:54893107-54893263	chr12:7621433-7621591	chr14:29085624-29085782
HESX1	chr3:54911870-54911953	chr12:7629837-7629922	chr14:29065520-29065604
HESX1	chr3:54911871-54912061	chr12:7629838-7630028	chr14:29065412-29065603
HESX1	chr3:54911935-54912061	chr12:7629904-7630028	chr14:29065412-29065538
HESX1	chr3:54962279-54962346	chr12:7655660-7655726	chr14:29018168-29018235
HESX1	chr3:54962317-54962615	chr12:7655697-7655990	chr14:29017900-29018197
HESX1	chr3:54994776-54994861	chr12:7672513-7672598	chr14:28988842-28988928
HESX1	chr3:55017905-55018232	chr12:7680043-7680373	chr14:28972684-28973015
HESX1	chr3:55017907-55018232	chr12:7680045-7680373	chr14:28972684-28973013
HESX1	chr3:55017909-55018232	chr12:7680047-7680373	chr14:28972684-28973011
HESX1	chr3:55017914-55018232	chr12:7680052-7680373	chr14:28972684-28973006
HESX1	chr3:55017917-55018232	chr12:7680055-7680373	chr14:28972684-28973003
HESX1	chr3:55017919-55018232	chr12:7680057-7680373	chr14:28972684-28973001
HESX1	chr3:55018137-55018337	chr12:7680278-7680478	chr14:28972577-28972779
HESX1	chr3:55062320-55062467	chr12:7705389-7705536	chr14:28934890-28935037
HESX1	chr3:55062334-55062389	chr12:7705403-7705458	chr14:28934968-28935023
HESX1	chr3:55162491-55162645	chr12:7741957-7742118	chr14:28863667-28863824
HESX1	chr3:55203884-55203970	chr12:7763551-7763638	chr14:28827291-28827378
HESX1	chr3:55217200-55217478	chr12:7771842-7772117	chr14:28816291-28816564
HESX1	chr3:55277256-55277561	chr12:7791140-7791434	chr14:28776867-28777184
HESX1	chr3:55278343-55278677	chr12:7792414-7792749	chr14:28775728-28776063
HESX1	chr3:55278344-55278438	chr12:7792415-7792509	chr14:28775967-28776062
HESX1	chr3:55278367-55278630	chr12:7792438-7792701	chr14:28775775-28776039
HESX1	chr3:55278369-55278622	chr12:7792440-7792693	chr14:28775783-28776037
HESX1	chr3:55278408-55278599	chr12:7792479-7792670	chr14:28775783-28775997
HESX1	chr3:55278412-55278592	chr12:7792483-7792663	chr14:28775813-28775993
HESX1	chr3:55278414-55278673	chr12:7792485-7792745	chr14:28775775-28775991
HESX1	chr3:55278415-55278509	chr12:7792486-7792580	chr14:28775896-28775990
HESX1	chr3:55278422-55278659	chr12:7792493-7792730	chr14:28775746-28775983

HESX1	chr3:55278424-55278712	chr12:7792495-7792787	chr14:28775746-28775981
HESX1	chr3:55278434-55278481	chr12:7792505-7792552	chr14:28775924-28775971
HESX1	chr3:55278438-55278624	chr12:7792509-7792695	chr14:28775781-28775967
HESX1	chr3:55278439-55278667	chr12:7792510-7792738	chr14:28775738-28775966
HESX1	chr3:55278459-55278697	chr12:7792530-7792771	chr14:28775709-28775946
HESX1	chr3:55278463-55278559	chr12:7792534-7792630	chr14:28775846-28775942
HESX1	chr3:55278545-55278655	chr12:7792616-7792726	chr14:28775750-28775860
HESX1	chr3:55278551-55278621	chr12:7792622-7792692	chr14:28775784-28775854
HESX1	chr3:55278552-55278661	chr12:7792623-7792732	chr14:28775744-28775853
HESX1	chr3:55278553-55278620	chr12:7792624-7792691	chr14:28775785-28775852
HESX1	chr3:55278574-55278632	chr12:7792645-7792703	chr14:28775773-28775831
HESX1	chr3:55278605-55278680	chr12:7792676-7792752	chr14:28775725-28775800
HESX1	chr3:55348065-55348336	chr12:7818751-7818971	chr14:28687718-28688033
HESX1	chr3:55348241-55348363	chr12:7818875-7818998	chr14:28687692-28687824
HESX1	chr3:55356936-55357269	chr12:7827041-7827370	chr14:28675308-28675640
HESX1	chr3:55357101-55357255	chr12:7827203-7827356	chr14:28675322-28675478
HESX1	chr3:55357116-55357421	chr12:7827218-7827522	chr14:28675156-28675463
HESX1	chr3:55357123-55357269	chr12:7827225-7827370	chr14:28675308-28675457
HESX1	chr3:55357125-55357269	chr12:7827227-7827426	chr14:28675252-28675455
HESX1	chr3:55357132-55357374	chr12:7827234-7827475	chr14:28675203-28675448
HESX1	chr3:55357133-55357367	chr12:7827235-7827468	chr14:28675342-28675447
HESX1	chr3:55357143-55357269	chr12:7827245-7827370	chr14:28675308-28675437
HESX1	chr3:55357158-55357457	chr12:7827260-7827558	chr14:28675120-28675422
HESX1	chr3:55357163-55357253	chr12:7827265-7827354	chr14:28675324-28675417
HESX1	chr3:55357167-55357402	chr12:7827269-7827503	chr14:28675175-28675413
HESX1	chr3:55357194-55357323	chr12:7827295-7827424	chr14:28675254-28675383
HESX1	chr3:55357209-55357361	chr12:7827310-7827462	chr14:28675216-28675368
HESX1	chr3:55357213-55357256	chr12:7827314-7827357	chr14:28675321-28675364
HESX1	chr3:55357215-55357458	chr12:7827316-7827559	chr14:28675119-28675362
HESX1	chr3:55357236-55357369	chr12:7827337-7827470	chr14:28675208-28675341

HESX1	chr3:55357238-55357364	chr12:7827339-7827465	chr14:28675213-28675339
HESX1	chr3:55395259-55395502	chr12:7849236-7849480	chr14:28630554-28630801
HESX1	chr3:55395414-55395502	chr12:7849392-7849480	chr14:28630554-28630642
HESX1	chr3:55514971-55515025	chr12:7901745-7901799	chr14:28511744-28511797
HESX1	chr3:55537022-55537148	chr12:7913505-7913631	chr14:28484425-28484551
HESX1	chr3:55546368-55546546	chr12:7921806-7921984	chr14:28474216-28474394
HESX1	chr3:55549834-55550131	chr12:7922687-7922988	chr14:28470336-28470631
HESX1	chr3:55549953-55550138	chr12:7922806-7922995	chr14:28470329-28470512
HESX1	chr3:55552498-55552721	chr12:7924341-7924565	chr14:28468101-28468324
HESX1	chr3:55552532-55552678	chr12:7924376-7924522	chr14:28468144-28468290
HESX1	chr3:55552533-55552677	chr12:7924377-7924521	chr14:28468145-28468289
HESX1	chr3:55552536-55552701	chr12:7924380-7924545	chr14:28468121-28468286
HESX1	chr3:55552537-55552682	chr12:7924381-7924526	chr14:28468140-28468285
HESX1	chr3:55552552-55552630	chr12:7924396-7924474	chr14:28468192-28468270
HESX1	chr3:55552553-55552679	chr12:7924397-7924523	chr14:28468143-28468269
HESX1	chr3:55552556-55552636	chr12:7924400-7924480	chr14:28468186-28468266
HESX1	chr3:55552557-55552701	chr12:7924401-7924545	chr14:28468121-28468265
HESX1	chr3:55552562-55552628	chr12:7924406-7924472	chr14:28468194-28468260
HESX1	chr3:55552569-55552678	chr12:7924413-7924522	chr14:28468144-28468253
HESX1	chr3:55552574-55552633	chr12:7924418-7924477	chr14:28468189-28468248
HESX1	chr3:55552575-55552633	chr12:7924419-7924745	chr14:28468189-28468247
HESX1	chr3:55552576-55552703	chr12:7924420-7924547	chr14:28468119-28468246
HESX1	chr3:55552582-55552908	chr12:7924426-7924750	chr14:28467914-28468240
HESX1	chr3:55552584-55552908	chr12:7924428-7924750	chr14:28467914-28468238
HESX1	chr3:55552585-55552908	chr12:7924429-7924750	chr14:28467914-28468237
HESX1	chr3:55552599-55552857	chr12:7924443-7924699	chr14:28467965-28468223
HESX1	chr3:55552603-55552695	chr12:7924447-7924539	chr14:28468127-28468219
HESX1	chr3:55552608-55552896	chr12:7924452-7924738	chr14:28467926-28468214
HESX1	chr3:55552609-55552859	chr12:7924453-7924701	chr14:28467963-28468213
HESX1	chr3:55552611-55552657	chr12:7924455-7924501	chr14:28468165-28468211

HESX1	chr3:55552612-55552769	chr12:7924456-7924611	chr14:28468053-28468210
HESX1	chr3:55552614-55552845	chr12:7924458-7924687	chr14:28467977-28468208
HESX1	chr3:55552622-55552704	chr12:7924466-7924548	chr14:28468142-28468200
HESX1	chr3:55552625-55552699	chr12:7924469-7924543	chr14:28468123-28468197
HESX1	chr3:55552626-55552942	chr12:7924470-7924785	chr14:28467880-28468196
HESX1	chr3:55552635-55552908	chr12:7924479-7924750	chr14:28467914-28468187
HESX1	chr3:55552639-55552692	chr12:7924483-7924536	chr14:28468130-28468183
HESX1	chr3:55552642-55552898	chr12:7924486-7924740	chr14:28467924-28468180
HESX1	chr3:55552643-55552704	chr12:7924487-7924548	chr14:28468118-28468179
HESX1	chr3:55552683-55552871	chr12:7924527-7924713	chr14:28467951-28468139
HESX1	chr3:55555151-55555267	chr12:7926233-7926349	chr14:28466208-28466324
HESX1	chr3:55555171-55555331	chr12:7926253-7926411	chr14:28466143-28466304
HESX1	chr3:55572854-55573096	chr12:7938648-7938876	chr14:28450067-28450299
HESX1	chr3:55572927-55572988	chr12:7938709-7938769	chr14:28450173-28450232
HESX1	chr3:55572928-55573091	chr12:7938710-7938871	chr14:28450072-28450231
HESX1	chr3:55596872-55597065	chr12:7950535-7950728	chr14:28426419-28426523
HESX1	chr3:55613649-55613810	chr12:7956542-7956704	chr14:28410832-28410992
HESX1	chr3:55613702-55614005	chr12:7956595-7956896	chr14:28410638-28410940
HESX1	chr3:55629633-55629730	chr12:7966008-7966106	chr14:28396121-28396218
HESX1	chr3:55640669-55640790	chr12:7975818-7975914	chr14:28384028-28384128
HESX1	chr3:55640803-55640961	chr12:7975927-7976088	chr14:28383858-28384018
HESX1	chr3:55640867-55640932	chr12:7975993-7976059	chr14:28383887-28383954
HESX1	chr3:55640869-55640998	chr12:7975995-7976126	chr14:28383819-28383952
HESX1	chr3:55640872-55640929	chr12:7975998-7976056	chr14:28383890-28383949
HESX1	chr3:55640932-55640998	chr12:7976059-7976126	chr14:28383819-28383887
HESX1	chr3:55642320-55642485	chr12:7976830-7976992	chr14:28382464-28382629
HESX1	chr3:55642372-55642485	chr12:7976879-7976992	chr14:28382464-28382577
HESX1	chr3:55649286-55649392	chr12:7982896-7983002	chr14:28376582-28376688
HESX1	chr3:55649951-55650178	chr12:7983632-7983858	chr14:28375807-28376032
HESX1	chr3:55650101-55650210	chr12:7983781-7983889	chr14:28375775-28375884

HESX1	chr3:55651351-55651636	chr12:7986491-7986778	chr14:28374235-28374518
HESX1	chr3:55651352-55651636	chr12:7986492-7986778	chr14:28374235-28374517
HESX1	chr3:55651550-55651761	chr12:7986692-7986903	chr14:28374111-28374321
HESX1	chr3:55653241-55653328	chr12:7987768-7987852	chr14:28371970-28372057
HESX1	chr3:55668485-55668627	chr12:8002969-8003110	chr14:28357259-28357410
HESX1	chr3:55681300-55681374	chr12:8008357-8008431	chr14:28347697-28347771
HESX1	chr3:55681778-55681824	chr12:8008848-8008894	chr14:28347244-28347290
HESX1	chr3:55698606-55698730	chr12:8015813-8015942	chr14:28332543-28332667
HESX1	chr3:55698683-55698889	chr12:8015895-8016092	chr14:28332391-28332590
HESX1	chr3:55720081-55720299	chr12:8029109-8029323	chr14:28314853-28315065
HESX1	chr3:55725917-55726124	chr12:8034809-8035010	chr14:28309536-28309739
HESX1	chr3:55736051-55736285	chr12:8043793-8044025	chr14:28300226-28300457
HESX1	chr3:55736109-55736253	chr12:8043851-8043997	chr14:28300251-28300400
HESX1	chr3:55736113-55736285	chr12:8043855-8044024	chr14:28300226-28300396
HESX1	chr3:55738229-55738443	chr12:8047079-8047290	chr14:28298137-28298352
HESX1	chr3:55749433-55749584	chr12:8050397-8050544	chr14:28288870-28289016
HESX1	chr3:55753628-55753685	chr12:8051709-8051766	chr14:28285651-28285708
HESX1	chr3:55775803-55776097	chr12:8060391-8060682	chr14:28266749-28267041
HESX1	chr3:55776031-55776302	chr12:8060616-8060863	chr14:28266554-28266815
HESX1	chr3:55776652-55776783	chr12:8061206-8061339	chr14:28266065-28266196
HESX1	chr3:55782090-55782177	chr12:8066446-8066530	chr14:28260410-28260501
HESX1	chr3:55789497-55789749	chr12:8071092-8071344	chr14:28252661-28252911
HESX1	chr3:55789593-55789910	chr12:8071188-8071505	chr14:28252500-28252816
HESX1	chr3:55789756-55790061	chr12:8071351-8071656	chr14:28252349-28252654
HESX1	chr3:55789867-55789931	chr12:8071462-8071526	chr14:28252479-28252543
HESX1	chr3:55789887-55790166	chr12:8071482-8071760	chr14:28252244-28252523
HESX1	chr3:55789891-55789971	chr12:8071486-8071566	chr14:28252439-28252519
HESX1	chr3:55789896-55790166	chr12:8071491-8071760	chr14:28252244-28252514
HESX1	chr3:55825994-55826135	chr12:8085921-8086062	chr14:28223555-28223696
HESX1	chr3:55826003-55826186	chr12:8085930-8086111	chr14:28223506-28223687

HESX1	chr3:55837633-55837755	chr12:8090384-8090505	chr14:28216101-28216223
HESX1	chr3:55853133-55853377	chr12:8096427-8096671	chr14:28202716-28202962
HESX1	chr3:55853213-55853443	chr12:8096507-8096737	chr14:28202650-28202882
HESX1	chr3:55894012-55894286	chr12:8108088-8108364	chr14:28165205-28165479
HESX1	chr3:55894013-55894257	chr12:8108089-8108335	chr14:28165234-28165478
HESX1	chr3:55894042-55894257	chr12:8108119-8108335	chr14:28165234-28165449
HESX1	chr3:55894066-55894366	chr12:8108143-8108445	chr14:28165125-28165425
HESX1	chr3:55894067-55894266	chr12:8108144-8108344	chr14:28165225-28165424
HESX1	chr3:55894070-55894366	chr12:8108147-8108445	chr14:28165125-28165421
HESX1	chr3:55894149-55894366	chr12:8108227-8108445	chr14:28165125-28165342
HESX1	chr3:55894152-55894366	chr12:8108230-8108445	chr14:28165125-28165339
HESX1	chr3:55894158-55894256	chr12:8108236-8108440	chr14:28165130-28165333
HESX1	chr3:55894180-55894264	chr12:8108258-8108342	chr14:28165227-28165311
HESX1	chr3:55937737-55938015	chr12:8126469-8126753	chr14:28126942-28127223
HESX1	chr3:55945434-55945595	chr12:8130855-8131013	chr14:28120901-28121056
HESX1	chr3:55945442-55945684	chr12:8130863-8131102	chr14:28120811-28121048
HESX1	chr3:55945473-55945612	chr12:8130891-8131030	chr14:28120884-28121022
HESX1	chr3:55976994-55977180	chr12:8150695-8150873	chr14:28089816-28090014
HESX1	chr3:56013171-56013233	chr12:8168183-8168242	chr14:28053431-28053493
HESX1	chr3:56027128-56027261	chr12:8177355-8177488	chr14:28039359-28039490
HESX1	chr3:56037141-56037447	chr12:8182007-8182311	chr14:28029108-28029454
HESX1	chr3:56037165-56037475	chr12:8182031-8182339	chr14:28029080-28029430
HESX1	chr3:56037220-56037493	chr12:8182085-8182357	chr14:28029062-28029375
HESX1	chr3:56037252-56037531	chr12:8182117-8182395	chr14:28029024-28029343
HESX1	chr3:56037286-56037531	chr12:8182150-8182395	chr14:28029024-28029269
HESX1	chr3:56046777-56046914	chr12:8189035-8189175	chr14:28019399-28019535
HESX1	chr3:56046821-56046936	chr12:8189084-8189197	chr14:28019399-28019491
HESX1	chr3:56077971-56078134	chr12:8197564-8197727	chr14:28002726-28002884
HESX1	chr3:56077972-56078065	chr12:8197565-8197659	chr14:28002790-28002883
HESX1	chr3:56120596-56120657	chr12:8217647-8217708	chr14:27963085-27963143

HESX1	chr3:56126464-56126765	chr12:8221047-8221347	chr14:27956046-27956342
HESX1	chr3:56126467-56126594	chr12:8221050-8221177	chr14:27956213-27956339
HESX1	chr3:56145907-56146128	chr12:8229198-8229420	chr14:27938121-27938344
HESX1	chr3:56145932-56146194	chr12:8229223-8229486	chr14:27938055-27938317
HESX1	chr3:56145956-56146145	chr12:8229247-8229437	chr14:27938104-27938293
HESX1	chr3:56145972-56146140	chr12:8229263-8229432	chr14:27938109-27938277
HESX1	chr3:56145974-56146096	chr12:8229265-8229388	chr14:27938153-27938275
HESX1	chr3:56145996-56146139	chr12:8229287-8229431	chr14:27938110-27938253
HESX1	chr3:56145997-56146195	chr12:8229288-8229487	chr14:27938054-27938252
HESX1	chr3:56146024-56146146	chr12:8229315-8229438	chr14:27938103-27938225
HESX1	chr3:56146103-56146197	chr12:8229395-8229489	chr14:27938052-27938146
HESX1	chr3:56146125-56146182	chr12:8229417-8229474	chr14:27938067-27938124
HESX1	chr3:56164974-56165208	chr12:8238453-8238673	chr14:27919660-27919882
HESX1	chr3:56207554-56207807	chr12:8254151-8254401	chr14:27876328-27876580
HESX1	chr3:56318352-56318538	chr12:8281785-8281961	chr14:27789714-27789895
HESX1	chr3:56501176-56501258	chr12:8342624-8342706	chr14:27623590-27623672
HESX1	chr3:56817672-56817885	chr12:8443067-8443284	chr14:27351858-27352069
HESX1	chr3:56817796-56817985	chr12:8443193-8443387	chr14:27351758-27351947
HESX1	chr3:57185149-57185346	chr12:8567083-8567281	chr14:27051614-27051811
HESX1	chr3:57185190-57185419	chr12:8567125-8567356	chr14:27051541-27051770
HESX1	chr3:57185271-57185352	chr12:8567206-8567287	chr14:27051608-27051689
HESX1	chr3:57188860-57189029	chr12:8570707-8570869	chr14:27048371-27048530
HESX1	chr3:57188905-57189029	chr12:8570753-8570869	chr14:27048371-27048485
HESX1	chr3:57193349-57193520	chr12:8576526-8576698	chr14:27045164-27045334
HESX1	chr3:57193417-57193543	chr12:8576594-8576721	chr14:27045142-27045265
HESX1	chr3:57234385-57234519	chr12:8590512-8590645	chr14:27000162-27000295
HESX1	chr3:57234494-57234663	chr12:8590620-8590789	chr14:27000017-27000187
CXCL14	chr5:137356787-137356848	chr13:13694427-13694488	chr18:34499753-34499814
CXCL14	chr5:137354791-137354837	chr13:13695492-13695537	chr18:34498115-34498161
CXCL14	chr5:137354788-137354834	chr13:13695495-13695540	chr18:34498112-34498158

CXCL14	chr5:137354733-137354832	chr13:13695497-13695595	chr18:34498057-34498156
CXCL14	chr5:137354647-137354808	chr13:13695521-13695681	chr18:34497971-34498132
CXCL14	chr5:137354641-137354805	chr13:13695524-13695687	chr18:34498049-34498129
CXCL14	chr5:137347498-137347624	chr13:13699543-13699669	chr18:34494640-34494766
CXCL14	chr5:137347548-137347621	chr13:13699546-13699619	chr18:34494690-34494763
CXCL14	chr5:137347554-137347617	chr13:13699550-13699613	chr18:34494696-34494759
CXCL14	chr5:137347550-137347615	chr13:13699552-13699608	chr18:34494701-34494757
CXCL14	chr5:137347552-137347605	chr13:13699562-13699615	chr18:34494694-34494747
CXCL14	chr5:137292105-137292302	chr13:13715046-13715242	chr18:34459298-34459494
CXCL14	chr5:137292193-137292301	chr13:13715047-13715268	chr18:34459384-34459493
CXCL14	chr5:137292078-137292299	chr13:13715049-13715268	chr18:34459273-34459491
CXCL14	chr5:137292178-137292274	chr13:13715074-13715170	chr18:34459369-34459466
CXCL14	chr5:137292176-137292272	chr13:13715076-13715172	chr18:34459367-34459464
CXCL14	chr5:137292168-137292269	chr13:13715079-13715180	chr18:34459359-34459461
CXCL14	chr5:137292186-137292268	chr13:13715080-13715162	chr18:34459377-34459460
CXCL14	chr5:137183481-137183532	chr13:13761381-13761432	chr18:44323195-44323246
CXCL14	chr5:137089456-137089655	chr13:13800286-13800485	chr11:54746583-54746782
CXCL14	chr5:136975546-136975619	chr13:13842947-13843020	chr13:58019366-58019439
CXCL14	chr5:136899856-136899999	chr13:13886393-13886539	chr13:57945388-57945533
CXCL14	chr5:135800578-135800684	chr13:14292546-14292654	chr13:57004011-57004117
CXCL14	chr7:135947492-135947668	chr13:14367516-14367692	chr13:56789608-56789784
CXCL14	chr5:135561814-135561938	chr13:14371006-14371130	chr13:56783715-56783839
CXCL14	chr5:135561791-135561868	chr13:14371076-14371153	chr13:56783692-56783769
CXCL14	chr5:135561774-135561838	chr13:14371106-14371170	chr13:56783675-56783739
CXCL14	chr5:135561756-135561830	chr13:14371114-14371188	chr13:56783657-56783731
CXCL14	chr5:135315400-135315490	chr13:14524346-14524430	chr13:56566527-56566609
CXCL14	chr5:135215654-135215729	chr13:14580241-14580316	chr13:56470347-56470422
CXCL14	chr5:135209714-135209832	chr13:14582030-14582148	chr13:56464959-56465077
CXCL14	chr5:135207278-135207353	chr13:14583301-14583376	chr13:56463616-56463691
CXCL14	chr5:134880489-134880801	chr13:14669057-14669267	chr13:56259355-56259658

CXCL14	chr5:134878848-134879178	chr13:14670748-14671083	chr13:56258148-56258479
CXCL14	chr5:134878838-134879132	chr13:14670796-14671093	chr13:56258138-56258434
CXCL14	chr5:134878834-134879130	chr13:14670798-14671097	chr13:56258134-56258432
CXCL14	chr5:134878802-134879102	chr13:14670826-14671129	chr13:56258102-56258404
CXCL14	chr5:134879020-134879095	chr13:14670833-14670908	chr13:56258322-56258397
CXCL14	chr5:134878848-134879067	chr13:14670861-14671083	chr13:56258148-56258369
CXCL14	chr5:134878836-134879041	chr13:14670887-14671095	chr13:56258136-56258343
CXCL14	chr5:134878691-134879014	chr13:14670914-14671241	chr13:56257992-56258316
CXCL14	chr5:134878652-134878970	chr13:14670958-14671280	chr13:56257953-56258272
CXCL14	chr5:134878860-134878960	chr13:14670968-14671071	chr13:56258160-56258262
CXCL14	chr5:134878772-134878951	chr13:14670977-14671159	chr13:56258073-56258253
CXCL14	chr5:134878752-134878917	chr13:14671014-14671179	chr13:56258053-56258216
CXCL14	chr5:134878802-134878896	chr13:14671035-14671084	chr13:56258147-56258196
CXCL14	chr5:134878847-134878893	chr13:14671038-14671174	chr13:56258058-56258193
CXCL14	chr5:134878757-134878885	chr13:14671046-14671174	chr13:56258058-56258185
CXCL14	chr5:134878810-134878883	chr13:14671048-14671121	chr13:56258110-56258183
CXCL14	chr5:134878754-134878876	chr13:14671055-14671177	chr13:56258055-56258176
CXCL14	chr5:134878749-134878866	chr13:14671065-14671182	chr13:56258050-56258166
CXCL14	chr5:134878652-134878846	chr13:14671085-14671280	chr13:56257953-56258146
CXCL14	chr5:134878767-134878818	chr13:14671113-14671164	chr13:56258068-56258118
CXCL14	chr5:134878751-134878802	chr13:14671129-14671180	chr13:56258052-56258102
CXCL14	chr5:134844393-134844545	chr13:14678417-14678551	chr13:56236534-56236601
CXCL14	chr5:134818477-134818673	chr13:14689557-14689749	chr13:56214882-56215071
CXCL14	chr5:134813064-134813251	chr13:14692053-14692210	chr13:56208725-56208890
CXCL14	chr5:134812901-134813211	chr13:14692094-14692373	chr13:56208562-56208849
CXCL14	chr5:134813010-134813055	chr13:14692219-14692264	chr13:56208671-56208716
CXCL14	chr5:134812998-134813052	chr13:14692222-14692274	chr13:56208659-56208713
CXCL14	chr5:134812909-134813036	chr13:14692238-14692365	chr13:56208570-56208697
CXCL14	chr5:134779783-134779901	chr13:14711367-14711487	chr13:56170975-56171090
CXCL14	chr5:134771042-134771146	chr13:14718573-14718676	chr13:56160142-56160245

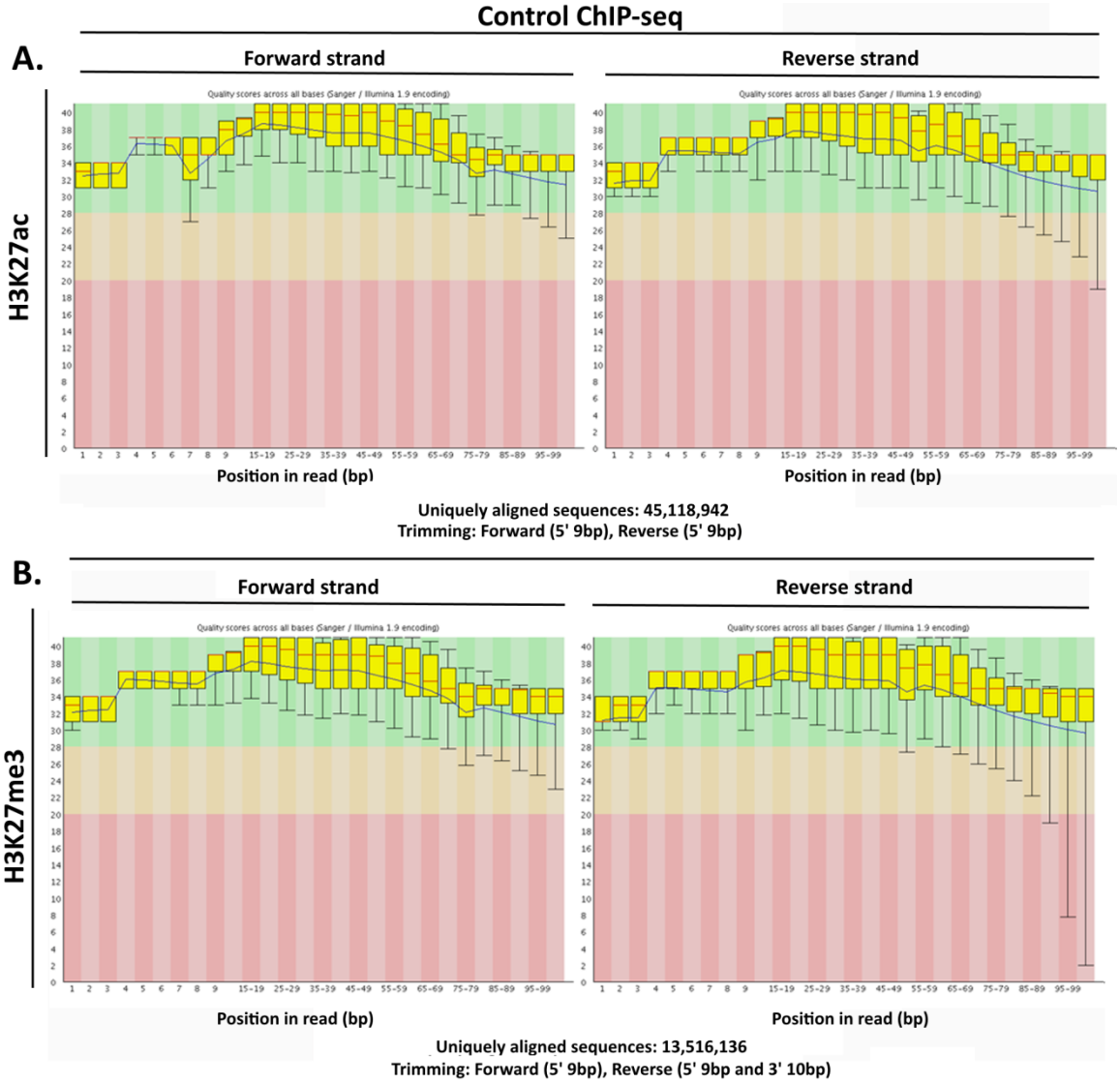
CXCL14	chr5:134711473-134711589	chr13:14753052-14753165	chr13:56111511-56111624
CXCL14	chr5:134711260-134711489	chr13:14753149-14753357	chr13:56111311-56111528
CXCL14	chr5:134702919-134703077	chr13:14761111-14761268	chr13:56102730-56102888
CXCL14	chr5:134688735-134688790	chr13:14767008-14767228	chr13:56089857-56089911
CXCL14	chr5:134686891-134687016	chr13:14768203-14768329	chr13:56088668-56088794
CXCL14	chr5:134686848-134687012	chr13:14768207-14768368	chr13:56088621-56088791
CXCL14	chr5:134686607-134686907	chr13:14768313-14768604	chr13:56088346-56088684
CXCL14	chr5:134686391-134686631	chr13:14768580-14768815	chr13:56088133-56088370
CXCL14	chr5:134686556-134686630	chr13:14768581-14768655	chr13:56088295-56088369
CXCL14	chr5:134686555-134686623	chr13:14768588-14768656	chr13:56088294-56088362
CXCL14	chr5:134686555-134686612	chr13:14768599-14768656	chr13:56088294-56088351
CXCL14	chr5:134686545-134686607	chr13:14768604-14768666	chr13:56088284-56088346
CXCL14	chr5:134686533-134686602	chr13:14768609-14768678	chr13:56088272-56088341
CXCL14	chr5:134686512-134686597	chr13:14768614-14768699	chr13:56088251-56088336
CXCL14	chr5:134683980-134684067	chr13:14770086-14770163	chr13:56086287-56086374
CXCL14	chr5:134659667-134659754	chr13:14791599-14791687	chr13:56063464-56063551
CXCL14	chr5:134601132-134601279	chr13:14810393-14810635	chr13:56025495-56025737
CXCL14	chr5:134600133-134600405	chr13:14811706-14811979	chr13:56024629-56024902
CXCL14	chr5:134600133-134600404	chr13:14811707-14811979	chr13:56024629-56024901
CXCL14	chr5:134600163-134600403	chr13:14811708-14811949	chr13:56024659-56024900
CXCL14	chr5:134600166-134600402	chr13:14811709-14811946	chr13:56024662-56024899
CXCL14	chr5:134600258-134600401	chr13:14811710-14811944	chr13:56024664-56024898
CXCL14	chr5:134600270-134600400	chr13:14811711-14811842	chr13:56024767-56024897
CXCL14	chr5:134600160-134600376	chr13:14811735-14811952	chr13:56024656-56024873
CXCL14	chr5:134600266-134600343	chr13:14811768-14811903	chr13:56024793-56024840
CXCL14	chr5:134600178-134600339	chr13:14811772-14811946	chr13:56024674-56024836
CXCL14	chr5:134600253-134600338	chr13:14811773-14811859	chr13:56024750-56024835
CXCL14	chr5:134600187-134600336	chr13:14811775-14811925	chr13:56024683-56024833
CXCL14	chr5:134600236-134600298	chr13:14811813-14811876	chr13:56024732-56024795
CXCL14	chr5:134600133-134600291	chr13:14811821-14811979	chr13:56024629-56024788

CXCL14	chr5:134600134-134600286	chr13:14811826-14811978	chr13:56024630-56024783
CXCL14	chr5:134600182-134600276	chr13:14811836-14811930	chr13:56024678-56024773
CXCL14	chr5:134600158-134600268	chr13:14811844-14811954	chr13:56024654-56024765
CXCL14	chr5:134600181-134600267	chr13:14811845-14811931	chr13:56024677-56024764
CXCL14	chr5:134600143-134600224	chr13:14811888-14811969	chr13:56024639-56024720
CXCL14	chr5:134523199-134523366	chr13:14848421-14848589	chr13:55959248-55959416
CXCL14	chr5:134523163-134523282	chr13:14848505-14848625	chr13:55959212-55959332
CXCL14	chr5:134520287-134520447	chr13:14849216-14849377	chr13:55956479-55956638
CXCL14	chr5:134451233-134451338	chr13:14874173-14874279	chr13:55901836-55901941
CXCL14	chr5:134442554-134442865	chr13:14881065-14881350	chr13:55893315-55893630
CXCL14	chr5:134401501-134401610	chr13:14899035-14899176	chr13:55859048-55859188
CXCL14	chr5:134401469-134401541	chr13:14899104-14899177	chr13:55859047-55859119
CXCL14	chr5:134387928-134388024	chr13:14910348-14910444	chr13:55847203-55847297
CXCL14	chr5:134286487-134286670	chr13:14974544-14974716	chr13:55767125-55767308
CXCL14	chr5:134285648-134285917	chr13:14975583-14975854	chr13:55766281-55766551
CXCL14	chr5:134285533-134285774	chr13:14975727-14975969	chr13:55766166-55766407
CXCL14	chr5:134285645-134285734	chr13:14975768-14975857	chr13:55766278-55766367
CXCL14	chr5:134285612-134285694	chr13:14975808-14975890	chr13:55766245-55766327
CXCL14	chr5:134280634-134280854	chr13:14978353-14978566	chr13:55761785-55762007
CXCL14	chr5:134280496-134280823	chr13:14978384-14978705	chr13:55761647-55761976
CXCL14	chr5:134280494-134280812	chr13:14978395-14978507	chr13:55761845-55761965
CXCL14	chr5:134280526-134280809	chr13:14978398-14978675	chr13:55761677-55761963
CXCL14	chr5:134280458-134280730	chr13:14978469-14978743	chr13:55761610-55761884
CXCL14	chr5:134280480-134280694	chr13:14978505-14978721	chr13:55761632-55761847
CXCL14	chr5:134280501-134280692	chr13:14978507-14978700	chr13:55761652-55761845
CXCL14	chr5:134280551-134280691	chr13:14978508-14978649	chr13:55761702-55761844
CXCL14	chr5:134280602-134280684	chr13:14978516-14978598	chr13:55761753-55761835
CXCL14	chr5:134280597-134280661	chr13:14978539-14978603	chr13:55761748-55761812
CXCL14	chr5:134280553-134280655	chr13:14978545-14978647	chr13:55761704-55761806
CXCL14	chr5:134280558-134280654	chr13:14978546-14978642	chr13:55761709-55761805

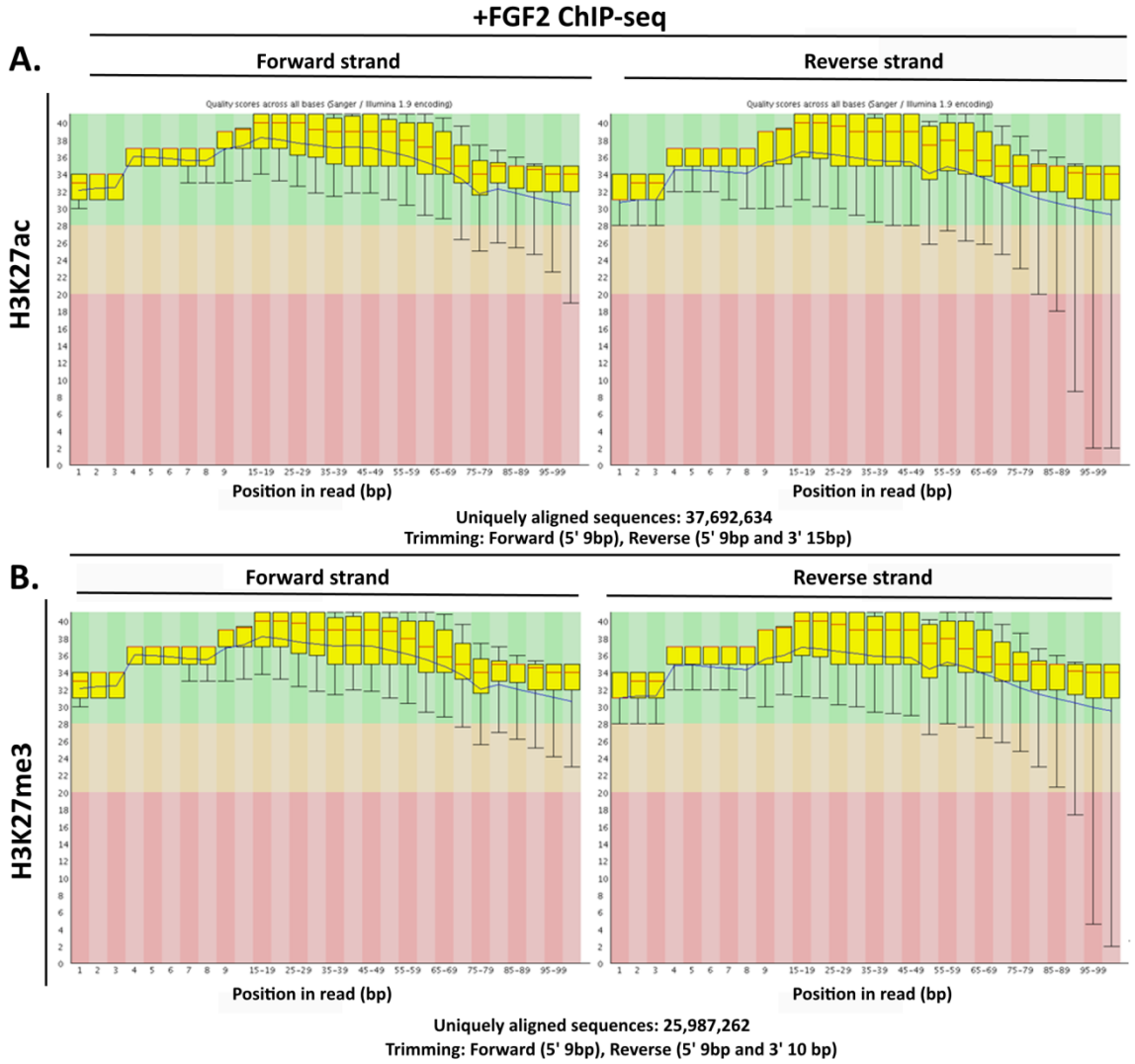
CXCL14	chr5:134280553-134280652	chr13:14978548-14978642	chr13:55761704-55761803
CXCL14	chr5:134280544-134280648	chr13:14978552-14978656	chr13:55761695-55761799
CXCL14	chr5:134280499-134280644	chr13:14978556-14978702	chr13:55761650-55761795
CXCL14	chr5:134280597-134280642	chr13:14978558-14978603	chr13:55761748-55761793
CXCL14	chr5:134280507-134280641	chr13:14978559-14978694	chr13:55761658-55761792
CXCL14	chr5:134280548-134280639	chr13:14978561-14978652	chr13:55761699-55761790
CXCL14	chr5:134280560-134280637	chr13:14978563-14978640	chr13:55761711-55761788
CXCL14	chr5:134280505-134280629	chr13:14978571-14978696	chr13:55761656-55761780
CXCL14	chr5:134280548-134280622	chr13:14978578-14978652	chr13:55761699-55761773
CXCL14	chr5:134280550-134280609	chr13:14978591-14978650	chr13:55761701-55761760
CXCL14	chr5:134280502-134280604	chr13:14978596-14978699	chr13:55761653-55761755
CXCL14	chr5:134280558-134280603	chr13:14978597-14978642	chr13:55761709-55761754
CXCL14	chr5:134280500-134280570	chr13:14978630-14978701	chr13:55761651-55761721
CXCL14	chr5:134280501-134280562	chr13:14978638-14978700	chr13:55761652-55761713
CXCL14	chr5:134280458-134280522	chr13:14978679-14978743	chr13:55761610-55761673
CXCL14	chr5:134273380-134273523	chr13:14982215-14982360	chr13:55754628-55754770
CXCL14	chr5:134273447-134273508	chr13:14982231-14982509	chr13:55754695-55754755
CXCL14	chr5:134271087-134271151	chr13:14984407-14984473	chr13:55750755-55750818
CXCL14	chr5:134270530-134270847	chr13:14984706-14985019	chr13:55750234-55750552
CXCL14	chr5:134270503-134270819	chr13:14984734-14985049	chr13:55750207-55750524
CXCL14	chr5:134270508-134270816	chr13:14984737-14985050	chr13:55750212-55750521
CXCL14	chr5:134270495-134270811	chr13:14984742-14985057	chr13:55750199-55750516
CXCL14	chr5:134270472-134270789	chr13:14984764-14985080	chr13:55750176-55750494
CXCL14	chr5:134270471-134270787	chr13:14984766-14985081	chr13:55750175-55750492
CXCL14	chr5:134270449-134270727	chr13:14984826-14985103	chr13:55750153-55750432
CXCL14	chr5:134270432-134270724	chr13:14984829-14985121	chr13:55750135-55750429
CXCL14	chr5:134270561-134270721	chr13:14984832-14984991	chr13:55750265-55750426
CXCL14	chr5:134270445-134270717	chr13:14984836-14985107	chr13:55750149-55750422
CXCL14	chr5:134270544-134270716	chr13:14984837-14985008	chr13:55750248-55750421
CXCL14	chr5:134270450-134270710	chr13:14984843-14985102	chr13:55750154-55750415

CXCL14	chr5:134270442-134270709	chr13:14984844-14985110	chr13:55750146-55750414
CXCL14	chr5:134270515-134270695	chr13:14984858-14985037	chr13:55750219-55750400
CXCL14	chr5:134270630-134270693	chr13:14984860-14984922	chr13:55750334-55750398
CXCL14	chr5:134270503-134270633	chr13:14984919-14985049	chr13:55750207-55750337
CXCL14	chr5:134270513-134270618	chr13:14984934-14985118	chr13:55750217-55750322
CXCL14	chr5:134270448-134270615	chr13:14984937-14985104	chr13:55750152-55750319
CXCL14	chr5:134270560-134270613	chr13:14984939-14984992	chr13:55750264-55750317
CXCL14	chr5:134270475-134270609	chr13:14984943-14985077	chr13:55750179-55750313
CXCL14	chr5:134270493-134270590	chr13:14984962-14985059	chr13:55750197-55750294
CXCL14	chr5:134270535-134270574	chr13:14984978-14985257	chr13:55750239-55750278
CXCL14	chr5:134270432-134270566	chr13:14984986-14985122	chr13:55750134-55750270
CXCL14	chr5:134270514-134270558	chr13:14984994-14985038	chr13:55750218-55750262
CXCL14	chr5:134270475-134270557	chr13:14984995-14985077	chr13:55750179-55750261
CXCL14	chr5:134270432-134270520	chr13:14985032-14985121	chr13:55750135-55750224
CXCL14	chr5:134254943-134255006	chr13:14986592-14986655	chr13:55740365-55740428
CXCL14	chr5:134254962-134255002	chr13:14986596-14986636	chr13:55740384-55740424
SOX13	chr1:203965961-203966197	chr26:1629902-1630138	chr1:133492470-133492706
SOX13	chr1:204077895-204078023	chr26:1648890-1649018	chr1:133397778-133397907
SOX13	chr1:204077984-204078046	chr26:1648979-1649041	chr1:133397755-133397817
SOX13	chr1:204078003-204078061	chr26:1648998-1649055	chr1:133397741-133397798
SOX13	chr1:204212268-204212470	chr26:1682983-1683179	chr1:133286170-133286369
SOX13	chr1:203966014-203966217	chr26:1629956-1630154	chr1:133492456-133492654
ETV4	chr17:41692008-41692253	chr27:3325312-3325554	chr11:101851685-101851931
ETV4	chr17:41669698-41669771	chr27:3327160-3327228	chr11:101830775-101830854
ETV4	chr17:41669152-41669512	chr27:3327381-3327686	chr11:101830199-101830587
ETV4	chr17:41651175-41651397	chr27:3331960-3332182	chr11:101809379-101809602
ETV4	chr17:41640431-41640731	chr27:3332688-3332969	chr11:101801615-101801913
ETV4	chr17:41622054-41622106	chr27:3336168-3336218	chr11:101783723-101783775
ETV4	chr17:41669245-41669560	chr27:3334963-3335144	chr11:101785330-101785549
ETV4	chr17:41600775-41600892	chr27:3355222-3355343	chr11:101784564-101784892

8.2 Plots showing per base sequence quality for H3K27ac and H3K27me3 in control ChIP-seq.



8.3 Plots showing per base sequence quality for H3K27ac and H3K27me3 in +FGF2 ChIP-seq.



8.4 Coordinates of overlapping DREiVe and +FGF2 enhancers

S.No	Gene	FGF2-treated ChIP enhancers	DREiVE enhancers
1	Sox13	chr26:1648949-1651648	chr26:1648890-1649018
2	Sox13	chr26:1648949-1651648	chr26:1648979-1649041
3	Sox13	chr26:1648949-1651648	chr26:1648998-1649055
4	Spry1	chr4:52750774-52752350	chr4:52750844-52751029
5	Spry1	chr4:52767820-52769341	chr4:52768195-52768476
6	Spry2	chr1:151881753-151884331	chr1:151884330-151884610
7	Spry2	chr1:152213343-152214723	chr1:152213668-152213729
8	Foxi3	chr4:85594771-85596968	chr4:85596160-85596222
9	Foxi3	chr4:85611408-85612363	chr4:85611416-85611709
10	Gbx2	chr7:5120747-5122576	chr7:5121199-5121716
11	Gbx2	chr7:5123327-5126074	chr7:5126071-5126444
12	Gbx2	chr7:5123327-5126074	chr7:5125589-5125821
13	Hesx1	chr12:8575392-8577407	chr12:8576526-8576698