



## King's Research Portal

DOI:

[10.1371/journal.pgen.1006587](https://doi.org/10.1371/journal.pgen.1006587)

*Document Version*

Publisher's PDF, also known as Version of record

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

Soderquest, K., Hertweck, A., Giambartolomei, C., Henderson, S., Mohamed, R., Goldberg, R., Perucha, E., Franke, L., Herrero, J., Plagnol, V., Jenner, R. G., & Lord, G. M. (2017). Genetic variants alter T-bet binding and gene expression in mucosal inflammatory disease. *PLoS Genetics*, *13*(2), Article e1006587. <https://doi.org/10.1371/journal.pgen.1006587>

### **Citing this paper**

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### **Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

RESEARCH ARTICLE

# Genetic variants alter T-bet binding and gene expression in mucosal inflammatory disease

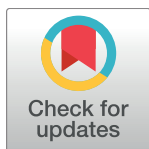
Katrina Soderquest<sup>1,2</sup>, Arnulf Hertweck<sup>3</sup>, Claudia Giambartolomei<sup>4</sup>, Stephen Henderson<sup>3</sup>, Rami Mohamed<sup>1,2,5</sup>, Rimma Goldberg<sup>1,2</sup>, Esperanza Perucha<sup>1a</sup>, Lude Franke<sup>6</sup>, Javier Herrero<sup>3</sup>, Vincent Plagnol<sup>4</sup>, Richard G. Jenner<sup>3‡\*</sup>, Graham M. Lord<sup>1,2‡\*</sup>

**1** Department of Experimental Immunobiology, King's College London, London, United Kingdom, **2** NIHR Biomedical Research Centre at Guy's and St Thomas' Hospital and King's College London, London, United Kingdom, **3** UCL Cancer Institute, University College London, London, United Kingdom, **4** UCL Genetics Institute, University College London, London, United Kingdom, **5** The Francis Crick Institute, London, United Kingdom, **6** Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

<sup>a</sup> Current address: Department of Rheumatology, King's College London, London, United Kingdom

<sup>‡</sup> These authors are joint senior authors on this work.

\* [graham.lord@kcl.ac.uk](mailto:graham.lord@kcl.ac.uk) (GML); [r.jenner@ucl.ac.uk](mailto:r.jenner@ucl.ac.uk) (RGJ)



**OPEN ACCESS**

**Citation:** Soderquest K, Hertweck A, Giambartolomei C, Henderson S, Mohamed R, Goldberg R, et al. (2017) Genetic variants alter T-bet binding and gene expression in mucosal inflammatory disease. *PLoS Genet* 13(2): e1006587. doi:10.1371/journal.pgen.1006587

**Editor:** Christine A. Wells, The University of Queensland, AUSTRALIA

**Received:** July 24, 2016

**Accepted:** January 17, 2017

**Published:** February 10, 2017

**Copyright:** © 2017 Soderquest et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** T-bet ChIP-seq data from donors 1 and 2 heterozygous for rs1465321 are publicly available at GEO under accession number GSE81881.

**Funding:** This study was supported by grants awarded by the Wellcome Trust ([wellcome.ac.uk](http://wellcome.ac.uk), RGJ, GML grant number 091009), the MRC ([mrc.ac.uk](http://mrc.ac.uk), GML, RGJ, grant number MR/M003493/1, GML, grant number MR/K002996/1) and the BBSRC ([bbsrc.ac.uk](http://bbsrc.ac.uk), RGJ, GML, grant number (BB/L009277/1). Research was also supported by

## Abstract

The polarization of CD4+ T cells into distinct T helper cell lineages is essential for protective immunity against infection, but aberrant T cell polarization can cause autoimmunity. The transcription factor T-bet (TBX21) specifies the Th1 lineage and represses alternative T cell fates. Genome-wide association studies have identified single nucleotide polymorphisms (SNPs) that may be causative for autoimmune diseases. The majority of these polymorphisms are located within non-coding distal regulatory elements. It is considered that these genetic variants contribute to disease by altering the binding of regulatory proteins and thus gene expression, but whether these variants alter the binding of lineage-specifying transcription factors has not been determined. Here, we show that SNPs associated with the mucosal inflammatory diseases Crohn's disease, ulcerative colitis (UC) and celiac disease, but not rheumatoid arthritis or psoriasis, are enriched at T-bet binding sites. Furthermore, we identify disease-associated variants that alter T-bet binding *in vitro* and *in vivo*. ChIP-seq for T-bet in individuals heterozygous for the celiac disease-associated SNPs rs1465321 and rs2058622 and the IBD-associated SNPs rs1551398 and rs1551399, reveals decreased binding to the minor disease-associated alleles. Furthermore, we show that rs1465321 is an expression quantitative trait locus (eQTL) for the neighboring gene *IL18RAP*, with decreased T-bet binding associated with decreased expression of this gene. These results suggest that genetic polymorphisms may predispose individuals to mucosal autoimmune disease through alterations in T-bet binding. Other disease-associated variants may similarly act by modulating the binding of lineage-specifying transcription factors in a tissue-selective and disease-specific manner.

the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's and St Thomas and King's College London ([www.guysandstthomasbrc.nihr.ac.uk](http://www.guysandstthomasbrc.nihr.ac.uk), KS, RM, RG, GML), and the NIHR University College London Hospitals Biomedical Research Centre ([www.uclhospitals.brc.nihr.ac.uk](http://www.uclhospitals.brc.nihr.ac.uk), AH, RGJ). The UCL Cancer Institute genomics core facility is supported by the Cancer Research UK – UCL Centre. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Author summary

Research to date has identified many genetic variants that are more common in people with a particular disease. However, in conditions that reflect multiple genetic and environmental factors, it is difficult to know with certainty if and why any particular genetic variant is causative and the mechanism that may underlie this. Such variants are often outside of protein-coding exons, instead falling in regions that regulate gene expression. In these cases, the genetic variation may alter transcription factor binding and subsequent gene expression. In this study, we have examined how genetic variation affects T-bet binding to DNA, as a key transcriptional regulatory mechanism in the immune response. An inability to mount this response effectively can result in increased susceptibility to infections or cancer, while a response that is too strong, or wrongly targeted, can result in uncontrolled/chronic inflammatory and autoimmune conditions. We have found that T-bet binding sites are specifically enriched in genetic variants associated with the mucosal autoimmune-inflammatory diseases UC, Crohn's disease and celiac disease. We also identify genetic variants that alter T-bet binding and gene expression. This discovery thus identifies a molecular mechanism through which genetic variants can be associated with increased risk of mucosal autoimmune disease.

## Introduction

The differentiation of naïve CD4<sup>+</sup> T cells into distinct T helper cell (Th) lineages is essential for adaptive immunity. The original paradigm of interferon-gamma (IFN- $\gamma$ ) producing T-helper 1 (Th1), and type-2 (Interleukin 4, 5, and 13) cytokine producing Th2 cells has expanded to include both Interleukin-17 (IL-17) producing Th17 and anti-inflammatory T-regulatory (Treg) cells. Th cell differentiation is controlled by a set of master regulatory or lineage-specifying transcription factors, with the T-box family member T-bet necessary and sufficient for Th1 cell differentiation. GATA3, ROR $\gamma$ T and FOXP3 perform parallel roles in Th2, Th17 and Treg cells, respectively [1]. Importantly, T-bet inhibits alternative lineage fate specification, repressing both the Th17 and Th2 lineages [2–4].

Inappropriate Th cell activation and polarization can lead to autoimmunity. Worldwide, autoimmune and auto-inflammatory diseases are now estimated to affect nearly 10% of the population [5]. The incidence of inflammatory bowel diseases (IBD), including Crohn's disease and UC, and celiac disease, is rising rapidly, with more than 1.4 million people affected in the USA alone [6]. A role for T-bet is particularly apparent in the mucosal immune system and has been linked to IBD and celiac disease [7]. The expression of T-bet is upregulated in lamina propria T cells of patients with Crohn's and celiac disease and *ex vivo* culture of biopsies from untreated celiac patients with gliadin increases T-bet expression through STAT1 activation [8,9]. In addition to this, it is now apparent that mucosal inflammation is also driven by IL-17, which is enhanced by IL-23 receptor signals in effector T cells [10]. Loss of T-bet in the innate immune system leads to a transmissible form of ulcerative colitis in the TRUC (T-bet and Rag deficient Ulcerative Colitis) model, driven by transcriptional derepression of *TNF* in colonic mononuclear phagocytes [11–13]. This susceptibility has also been shown to be dependent on IL-17 and mediated via repression of IL-7 receptor expression by T-bet in innate lymphoid cells (ILCs) [11]. T-bet has subsequently been shown to play a role in the development of the NKp46<sup>+</sup> CCR6<sup>-</sup> subset of IL-22 expressing ILCs that, in turn, are important for protecting the epithelial barrier during *Salmonella enterica* infection [14,15].

Autoimmune diseases cluster in families, suggesting a large genetic component [16]. Genome-wide association studies (GWAS) have identified hundreds of risk loci for autoimmune diseases, including for IBD and celiac disease [16–23]. The majority of autoimmune disease-associated SNPs lie outside of gene coding regions in intergenic or intronic regions [24]. This can make it challenging to understand the molecular basis of how a genetic variant predisposes to disease. Furthermore, the causal variant can be difficult to identify from the large clusters of SNPs in linkage disequilibrium that tend to be identified by GWAS. Thus, efforts have been made to identify SNPs located within regulatory elements marked by open chromatin, histone modifications associated with active enhancers or known/predicted transcription factor binding sites [21,24–32]. Some of these variants have been shown to modulate transcription factor binding or epigenetic regulation. Genetic variants that alter DNase I hypersensitivity [27,33,34], DNA methylation [35–38], histone modification [27,39–43], and the binding of transcriptional regulators to DNA [27,33,34,44–51], have been identified, suggesting potential causal mechanisms.

Although previous studies have demonstrated enrichment of transcription factor binding sites at disease-associated polymorphisms, whether specific disease causing variants act to alter the binding of T cell lineage-specifying factors has not been investigated. Having previously mapped T-bet binding across the genome in human Th1 cells [52–54] we used a systematic functional GWAS (fGWAS) approach to determine the degree to which disease-associated SNPs were enriched within T-bet binding sites. SNPs were then tested for effects on T-bet binding *in vitro* using a novel flow cytometric assay and *in vivo* by allele-specific ChIP-seq. We report here that SNPs associated with mucosal inflammatory diseases are selectively enriched at T-bet binding sites. Furthermore, we show that the celiac disease associated variants of rs1465321 and rs2058622, and the IBD-associated variants of rs1551398 and rs1551399, exhibit decreased T-bet binding *in vivo*. We further demonstrate that the genes associated with these SNPs, *IL18RAP* and *TRIB1*, respectively, are transcriptionally regulated by T-bet and that rs1465321 is an expression quantitative trait locus (eQTL) for *IL18RAP*. Taken together, these data mechanistically link alterations in T-bet binding to disease predisposition.

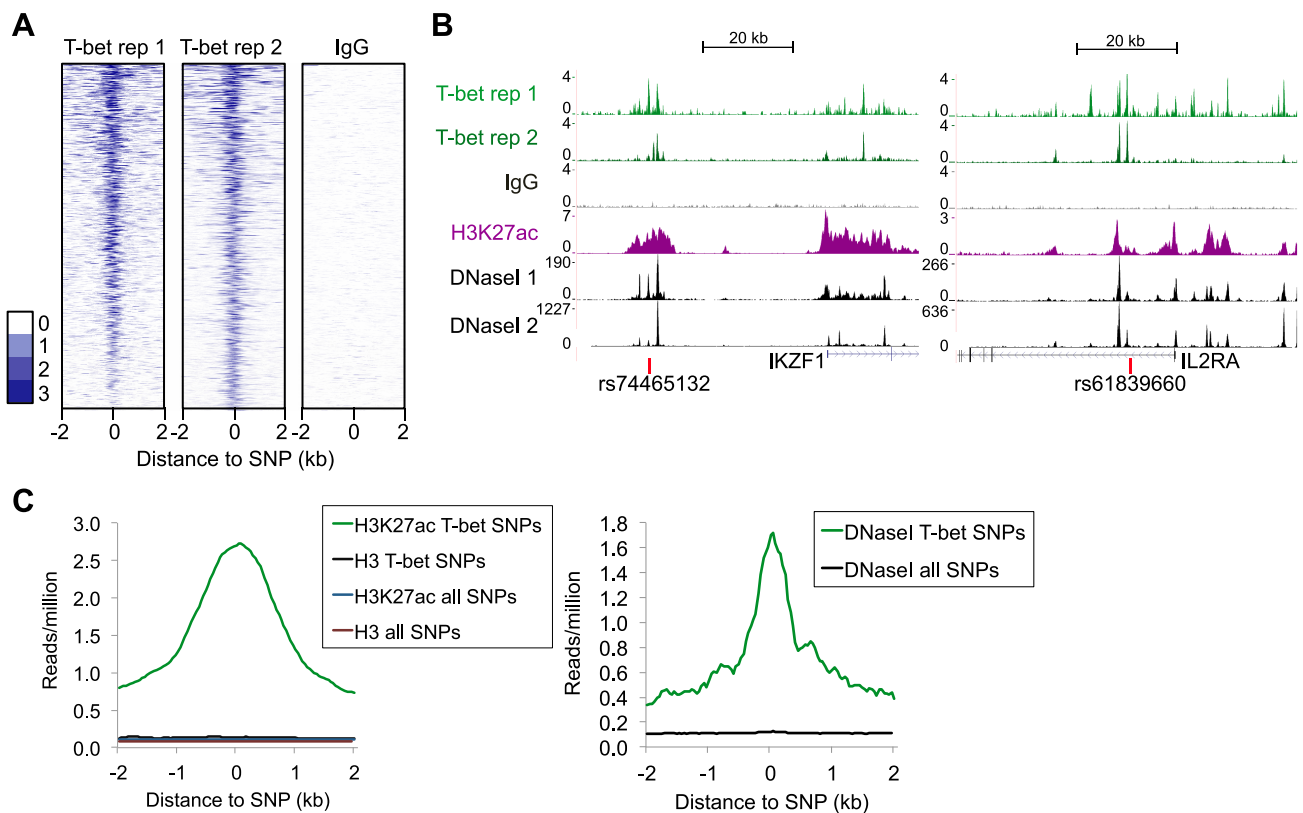
## Results

### Disease-associated SNPs at T-bet binding sites

To identify disease-associated polymorphisms at T-bet binding sites, we compared the locations of GWAS hits listed in the National Human Genome Research Institute (NHGRI) catalogue [55] with binding sites for T-bet in primary human Th1 cells [52–54]. As the published trait-associated SNP may not be the most functionally relevant [28], SNPs in high linkage disequilibrium LD ( $r^2 > 0.8$ ) were also examined. This returned a list of 926 unique SNPs located at T-bet binding sites (hereafter referred to as T-bet hit-SNPs; Fig 1A and 1B, S1A Fig and S1 Table). In line with previous reports, a minority (143) of the T-bet hit-SNPs were the putative causal SNP from GWAS data, with the others being in high LD (total of 621 independent LD blocks). Examination of the location of T-bet hit-SNPs in relation to protein-coding genes revealed that the majority (63%) were distal (>1 kb) to gene promoters. As expected, H3K27ac and DNaseI hypersensitivity were highly enriched in Th1 cells at T-bet hit-SNPs compared with all disease-associated SNPs, consistent with these being located within active regulatory elements (Fig 1C).

### SNPs associated with mucosal immune diseases are enriched at T-bet binding sites

As T-bet is only expressed in cells of the immune system, we hypothesised that T-bet hit-SNPs would be primarily associated with autoimmune diseases. To test this, we used fGWAS [56], a

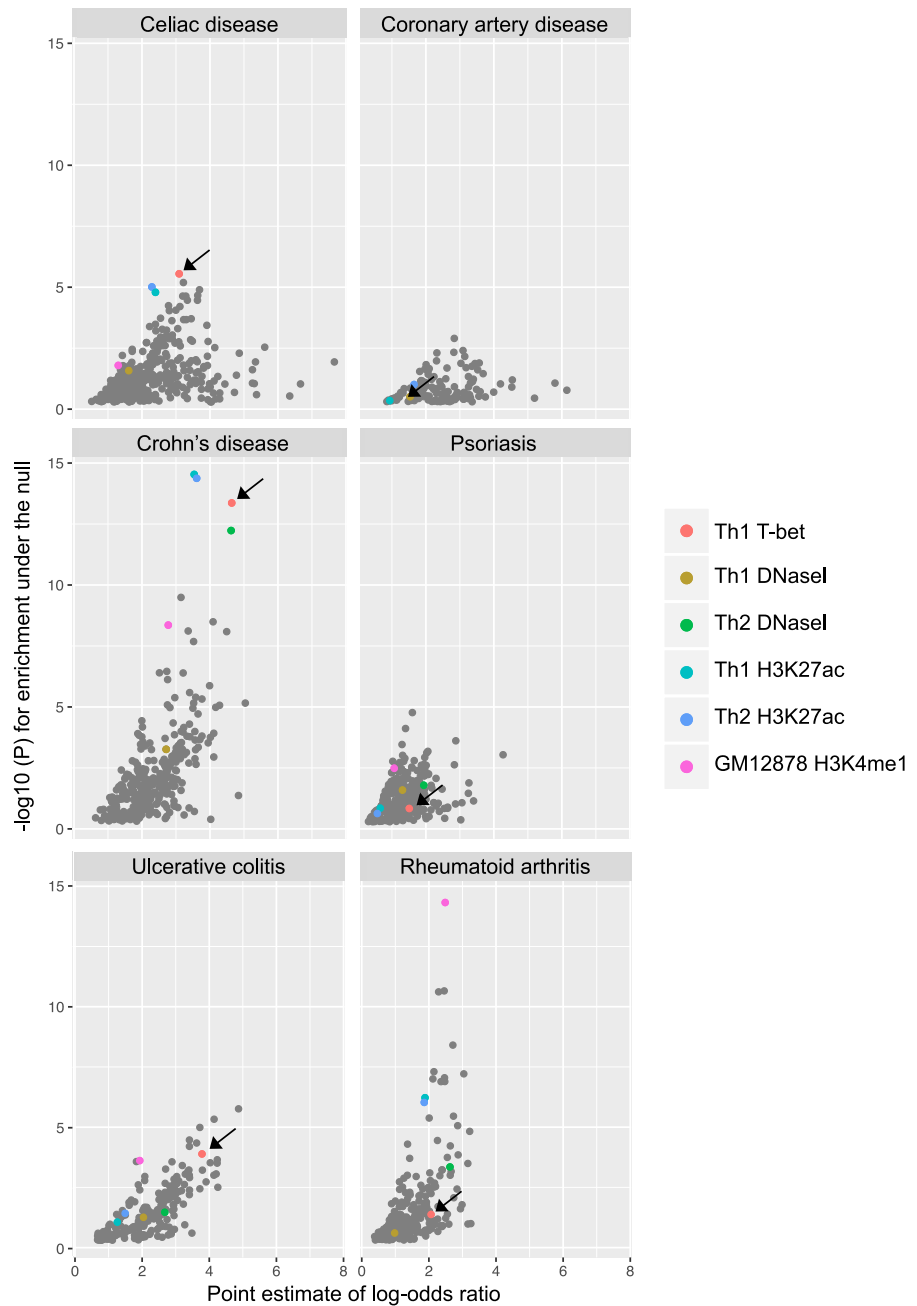


**Fig 1. T-bet binding at polymorphic sites.** **A.** Heat map showing T-bet occupancy around SNPs located within T-bet binding sites (T-bet hit-SNPs). Each row is centred on a single SNP, with T-bet binding shown across the genomic region stretching 2 kb up and downstream. Sequence reads (per million total reads) at each position are represented by colour, according to the scale on the left. Negative IgG ChIP-seq data are shown on the right at the same loci. **B.** T-bet binding at two example T-bet hit-SNPs. The number of sequencing reads from T-bet, IgG control and H3K27ac ChIP-enriched DNA are plotted per million input-subtracted total reads and aligned with the human genome. DNaseI hypersensitivity data (2 replicates) are from ENCODE. **C.** Left: Average number of ChIP-seq reads for H3K27ac and control total H3 in human Th1 cells plotted against the genomic distance from T-bet hit-SNPs or the complete set of GWAS SNPs plus those in high LD. Right: Average number of sequencing reads measuring DNaseI hypersensitivity plotted against genomic distance.

doi:10.1371/journal.pgen.1006587.g001

hierarchical model that assesses relative enrichment of GWAS associations within various functional elements. This model splits the genome into large blocks (larger than regions of linkage disequilibrium), assesses whether each block contains a SNP associated with the trait of interest or not and then searches among supplied functional annotations for those that improve the likelihood of predicting the presence of a trait-associated SNP, finally predicting which SNP in the block is most likely causal.

To test whether disease-associated SNPs were enriched at T-bet binding sites, we gathered GWAS data for the Th1-associated auto-inflammatory conditions celiac disease, Crohn's disease, UC, rheumatoid arthritis (RA), psoriasis and, as a negative non-immune control, coronary artery disease (Fig 2). We compared T-bet binding sites with a number of other relevant functional annotations, including Th1 and Th2 cell DHS [57], H3K27ac [58], and sites of histone modification and transcription factor binding in immune cell lines from the ENCODE project [26] and other sources (S2 Table). Notably, we found that SNPs associated with all of the mucosal immune-mediated diseases tested (Crohn's disease, UC and celiac disease) were enriched at T-bet binding sites, with the effect in Crohn's disease being particularly striking. Enrichment at T-bet binding sites was similar to, or stronger than, DHS and H3K27ac and, in



**Fig 2. SNPs associated with mucosal autoimmune diseases are enriched at T-bet binding sites.** Scatter plots showing log-odds ratio against  $-\log_{10}$  p-value for the enrichment of disease-associated SNPs at different functional annotation datasets (DHS, histone modification, FAIRE-seq and transcription factor binding). Selected enriched functional annotation datasets are highlighted. GM12878 H3K4me1 indicates sites of H3K4me1 in the GM12878 lymphoblastoid cell line. Celiac disease, Crohn's disease and UC-associated SNPs, but not RA, psoriasis or coronary artery disease-associated SNPs, are strongly enriched at T-bet binding sites (red dots with arrows).

doi:10.1371/journal.pgen.1006587.g002

the case of Crohn's and celiac disease, stronger than any other sets of transcription factor binding sites. As expected, SNPs associated with coronary artery disease were not enriched at T-bet binding sites. Of interest, no enrichment for T-bet binding sites was observed for RA- or psoriasis-associated SNPs, suggesting a specific role for altered T-bet binding in mucosal inflammatory disease.

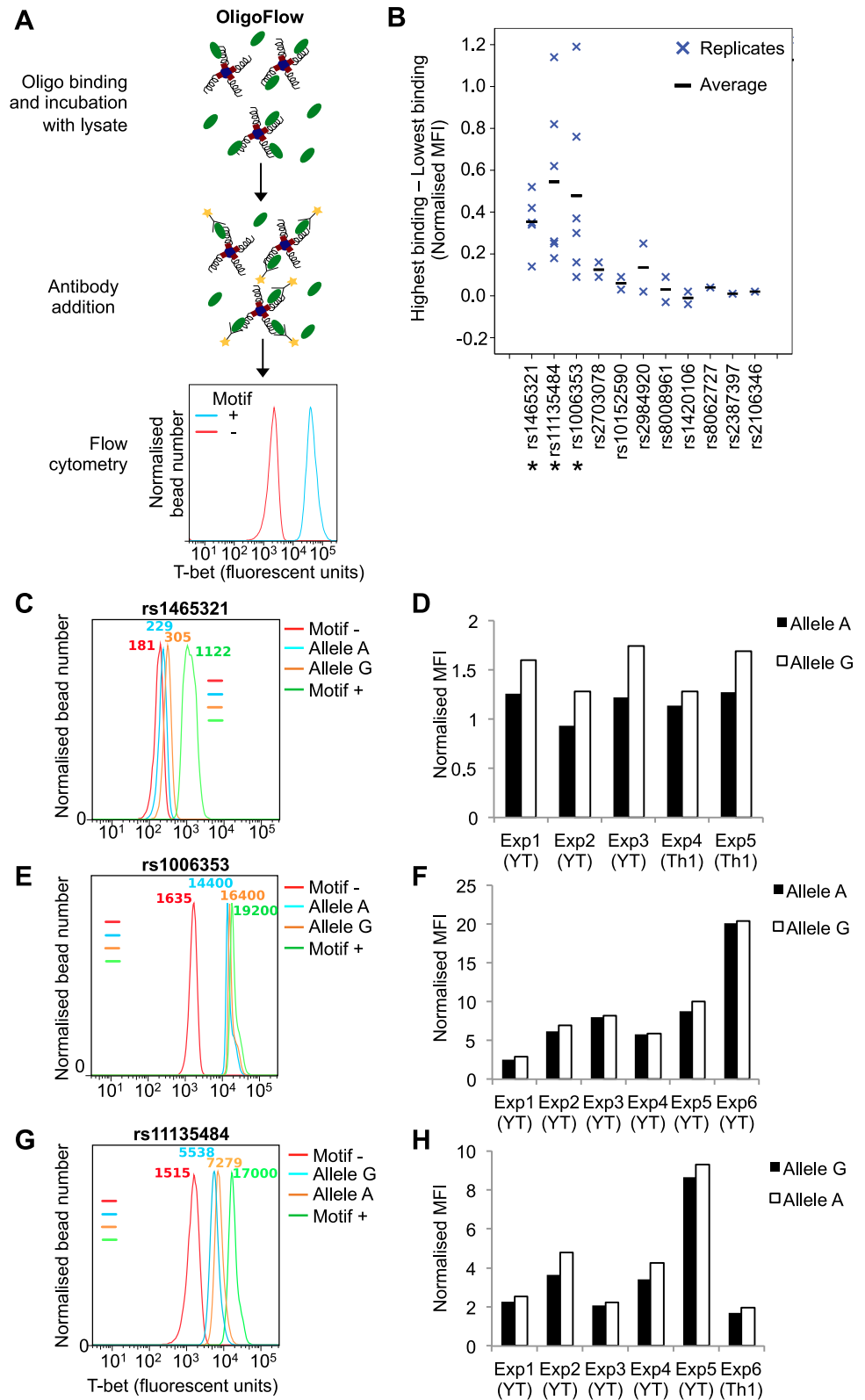
To confirm that T-bet binding is enriched at IBD-associated SNPs, we compared T-bet binding sites to a set of credible SNPs identified at 94 IBD-associated loci [21]. We found that T-bet binding sites were more highly associated with credible SNPs than other SNPs at the same loci (93 bound by T-bet,  $p = 1.4 \times 10^{-5}$ , Fisher exact test). Furthermore, within the set of credible SNPs, the higher the posterior probability for causality, the more likely that the SNP overlapped a T-bet binding site ( $p = 6.3 \times 10^{-6}$ , continuous binomial regression, S1B Fig). The association of T-bet binding with causal SNPs is highlighted by the finding that, of the 93 credible SNPs bound by T-bet, 11 are the lead variants for their loci. Three of these (rs74465132, rs1887428 and rs61839660) have a posterior probability for causality of greater than 95%. These data suggest that the strong association of these SNPs with IBD is related to T-bet binding at these sites.

## Detection of altered T-bet binding at disease-associated variants by OligoFlow

Having identified a set of SNPs overlapping T-bet binding sites, we next asked whether these sequence variants altered T-bet binding. The traditional pull-down technique is time intensive and semi-quantitative. Therefore, we explored whether transcription factor binding could be assayed using a flow cytometric readout. In this technique, which we call OligoFlow, a fluorochrome-labelled antibody for the transcription factor of interest is added to the oligonucleotide-bead / lysate mix, and the Median Fluorescence Intensity (MFI) of the beads is assessed by flow cytometry as a quantitative measure of binding efficiency (Fig 3A).

To validate this new technique, a positive control oligonucleotide (Motif+) was designed to incorporate the previously identified consensus sequence [54] surrounded by non-specific sequence (S3 Table). A negative control oligo (Motif-) incorporated mutations of two key residues within the motif. OligoFlow was conducted with lysate from either the YT human cell line, which constitutively expresses T-bet [59], or lysate from primary human CD4<sup>+</sup> cells polarised under Th1 conditions in culture. The positive and negative control oligonucleotides showed a clear difference in MFI (Fig 3A) and thus OligoFlow can successfully discriminate positive and negative transcription factor binding events.

We then proceeded to test a subset of our T-bet hit-SNPs that were also associated with H3K27ac or near genes of immunological interest. SNPs that showed differential binding were tested at least five times. Within each experiment, the MFI of each allele was normalised to the MFI of the negative control and significantly altered binding between alleles across all experiments was assessed using a paired t-test. Three T-bet hit-SNPs exhibited significantly different binding to the two alleles; rs1465321, located within the second intron of *IL18R1*, rs1006353, 22.5 kb upstream of *MTIF3*, and rs11135484, within an intron of *ERAP2* (Fig 3B). Differential T-bet binding to the two alleles of rs1465321 were confirmed by traditional oligonucleotide pull-down (S2 Fig). All 3 SNPs are [A/G] with A as the minor allele. In each case, allele A is also in LD with alleles associated with for the trait under investigation. rs1465321 is in high LD with multiple SNPs associated with celiac disease, including rs13015714 and rs917997, identified as the strongest risk alleles for celiac disease in 2q12.1 [18,60], with the disease-associated alleles linked to reduced *IL18RAP* expression [60]. rs1465321 and rs11135484 have also been associated with Crohn's disease [18,22,60,61], but not in a more recent study [21]. For



**Fig 3. Genetic variants alter T-bet binding *in vitro*.** **A.** Outline of the OligoFlow method. Double-stranded oligonucleotides are annealed to beads and incubated with cell lysate containing the transcription factor of interest. Fluorescently labelled antibody is added and MFI of the beads measured by flow cytometry. The



histograms show the MFI of beads coated with oligonucleotides containing a T-bet binding motif (Motif +) or a mutated sequence (Motif -) after incubation with YT lysate, normalised for the number of beads acquired. **B.** Summary of OligoFlow results for the 11 SNPs tested. In each case, MFI for both alleles is normalised such that the negative control equals 1. Normalised MFI for the lowest binding allele was then subtracted from the value for the highest binding allele. Each cross represents one experiment, with the average difference between alleles represented by a horizontal line. \* Significantly different binding between the two alleles ( $p < 0.05$ , paired t-test.) **C.** Representative experiment measuring the binding of T-bet to the A and G alleles of rs1465321. Data for the different oligonucleotide probes are separated according to the key on the right and the MFI is also shown. **D.** Bar chart showing all replicate experiments for rs1465321. The y-axis shows MFI for each allele normalised to the MFI of the negative control oligonucleotide (set to 1). Each pair of bars represents one experiment, performed with either YT cells (YT) or Th1-polarised primary CD4<sup>+</sup> cells (Th1). **E.** As C but for rs1006353. **F.** As D but for rs1006353. **G.** As C but for rs1135484. **H.** As D but for rs1135484.

doi:10.1371/journal.pgen.1006587.g003

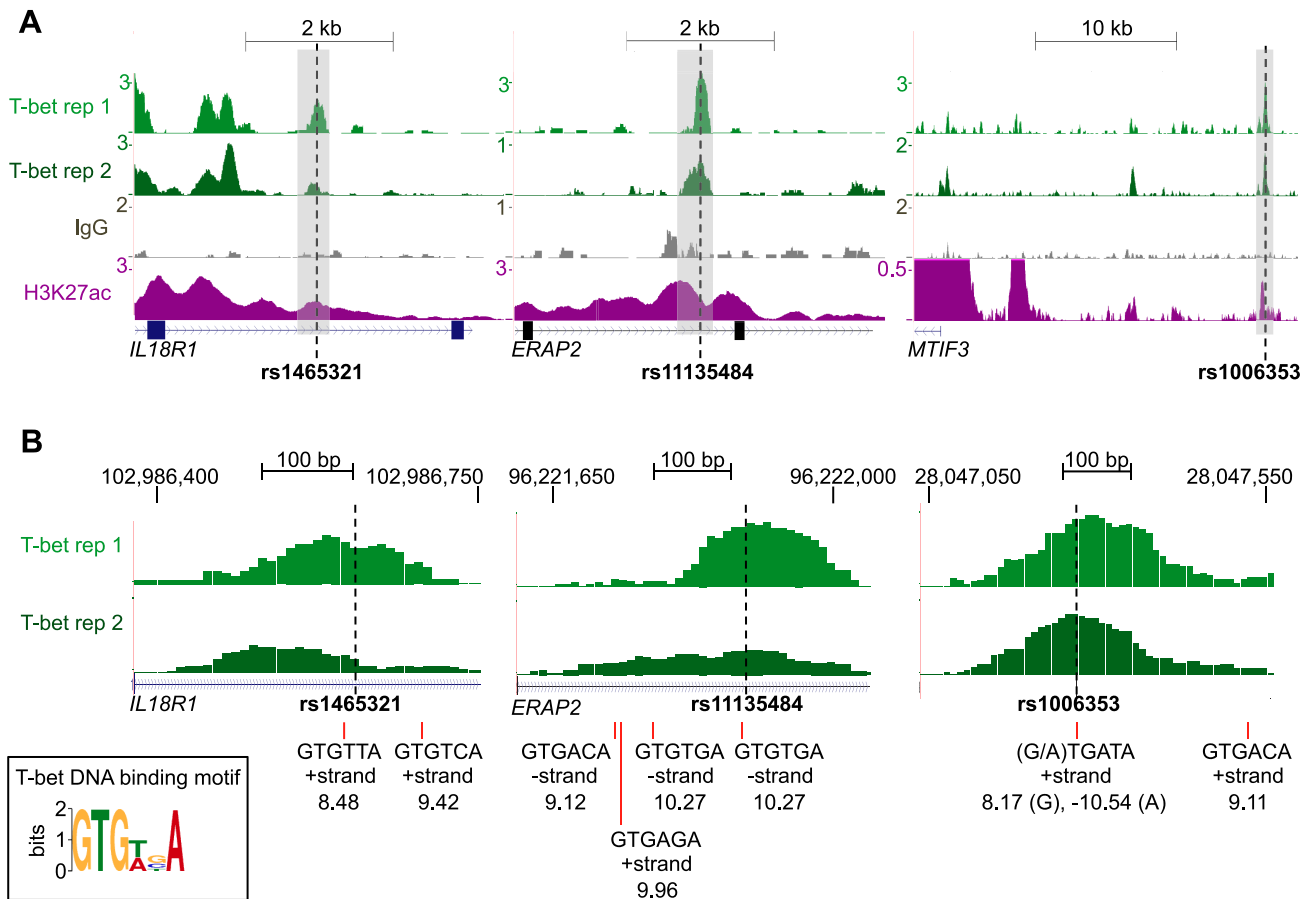
rs1465321 and rs1006353, the minor disease-associated A allele binds T-bet less strongly than the G allele (Fig 3C–3F). In contrast, for rs1135484, the A allele binds T-bet more strongly than the G allele (Fig 3G and 3H). We conclude that disease-associated genetic variants can alter T-bet binding to DNA *in vitro*.

### SNPs affecting T-bet binding cannot reliably be identified by motif analysis

Motif analysis has often been used to predict transcription factor binding sites affected by genetic variants. We previously derived a consensus T-bet motif from T-bet binding sites in human Th1 cells [54] and repeated this analysis with duplicate T-bet ChIP-seq data (Fig 4). The three T-bet hit-SNPs that showed altered binding in OligoFlow were then examined for whether they disrupted such a T-bet binding motif. In the case of rs1006353, the G allele formed part of a T-bet binding motif, whereas the A allele abolished this binding site (Fig 4B). However, neither of the other two SNPs, rs1465321 and rs1135484, overlapped a predicted T-bet binding motif (Fig 4B). Thus, over-reliance on motif analysis can result in SNPs with the potential to alter transcription factor binding sites being missed and highlights the importance of using experimental validation to confirm binding of the relevant transcription factor.

### Differential T-bet binding at disease variants *in vivo*

We next sought to confirm that T-bet exhibited differential binding to disease-associated SNPs *in vivo*. We focused on rs1465321, because it lies within the *IL18R1/IL18RAP* gene locus that we have previously identified as a T-bet target [54] and because disease-associated alleles in high LD are associated with reduced *IL18RAP* expression and disease [60]. Primary naive CD4<sup>+</sup> T cells were purified from the peripheral blood of two individuals heterozygous for this SNP and were polarised into the Th1 lineage. We then performed ChIP-seq for T-bet in these cells, as previously described [54]. We aligned the reads for the T-bet ChIP-enriched DNA and input controls to the reference human genome and then counted the number of reads matching the major or minor alleles in the inputs and ChIP samples. In the input DNA samples, there were approximately equal numbers of reads for the two alleles in both individuals. In comparison, the T-bet ChIP reads showed significantly lower enrichment for the minor A allele in both donors (Fig 5A and 5B). There was also a significant allelic imbalance for T-bet binding at the neighbouring SNP rs2058622, which is in high LD ( $r^2 = 1.0$ ) with rs1465321 (Fig 5A and 5B). To determine whether T-bet exhibited allelic imbalanced binding at any other loci, we identified all SNPs that exhibited heterozygosity in both individuals. Of the heterozygous SNPs that overlapped a T-bet binding site, 19 exhibited significant allelic imbalanced binding in both donors after adjustment for multiple hypothesis testing



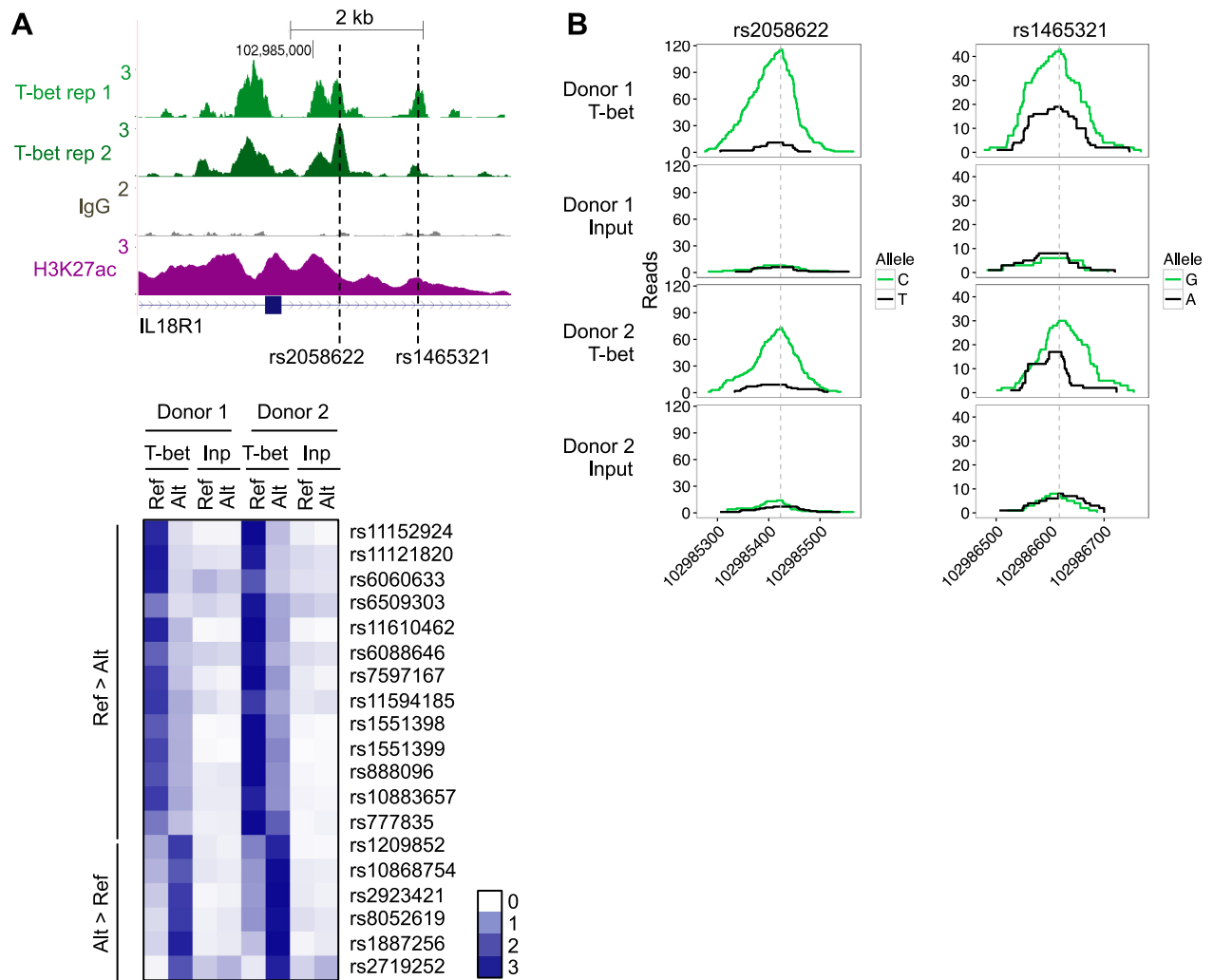
**Fig 4. Motif analysis does not reliably predict impact on T-bet binding.** **A.** T-bet binding, IgG control and H3K27ac modification (ChIP-seq reads/million) at the genomic regions surrounding the SNPs rs1465321 (left), rs11135484 (center) and rs1006353 (right). The location of the SNPs are indicated by dashed vertical lines. The regions highlighted in grey are expanded in B. **B.** Expanded view of T-bet binding at the regions highlighted in grey in A. The locations of sequences matching the identified T-bet DNA binding motif (inset) are marked by red lines, together with their score (a negative value indicates a poor match). Only rs1006353 overlaps a T-bet DNA binding motif and the A allele is predicted to disrupt the motif and T-bet binding.

doi:10.1371/journal.pgen.1006587.g004

(Fig 5C, S4 Table). These included the IBD-associated SNPs rs1551398 and rs1551399 [21], situated 86bp apart and downstream of *TRIB1* (Fig 5C, S1 and S3 Figs). We conclude that the two alleles of rs1465321 exhibit different levels of T-bet binding *in vivo*, with the disease associated A allele bound significantly less, and that the credible IBD variants rs1551398 and rs1551399 also influence T-bet binding.

### Regulation of *Il18rap* and *Trib1* expression in T-bet<sup>-/-</sup> Th1 cells

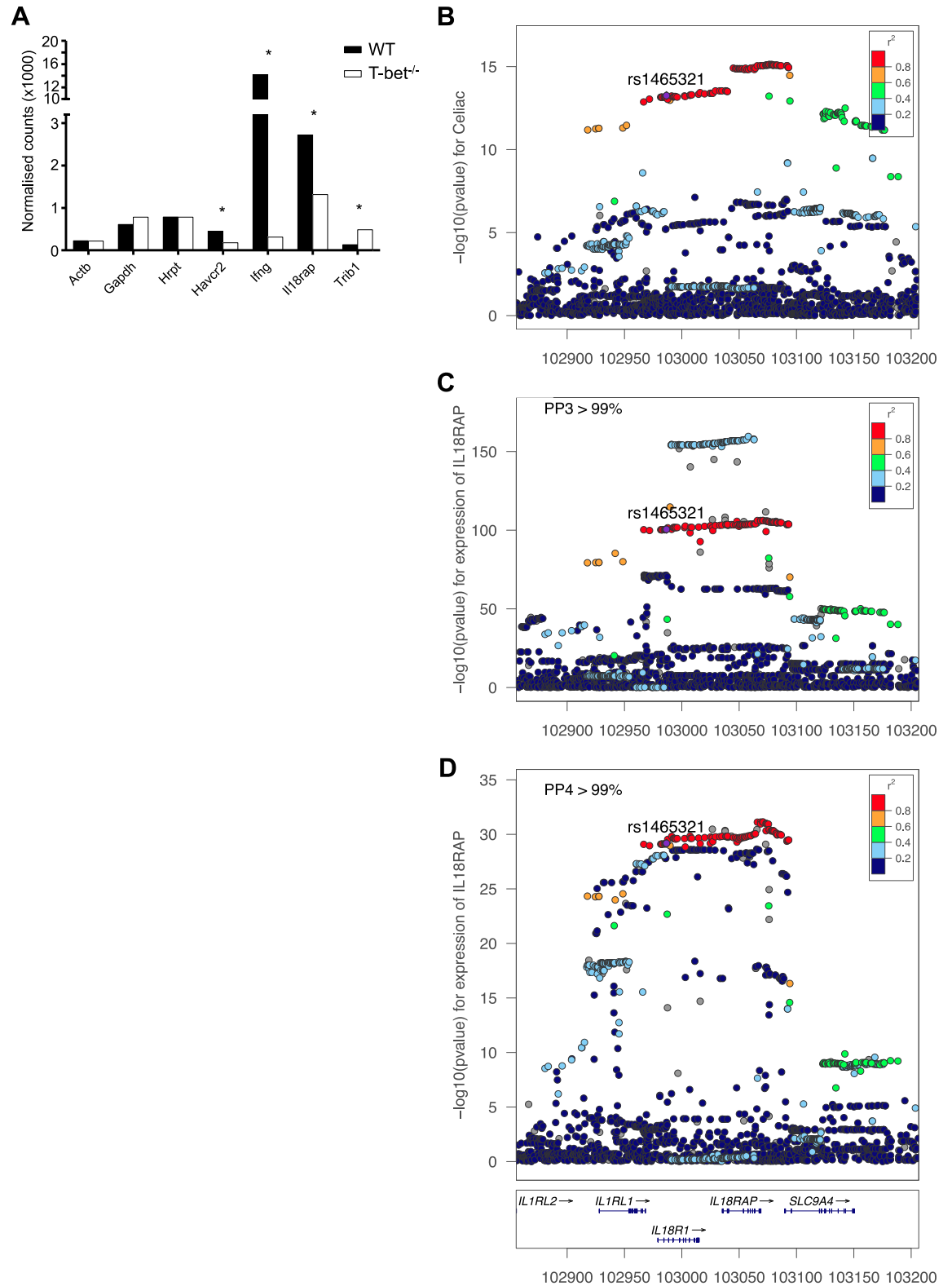
Having identified rs1465321, rs2058622, rs1551398 and rs1551399 as disease associated SNPs that modulate T-bet binding *in vivo*, we next determined whether there was a functional relationship between T-bet binding and the genes associated with these SNPs. rs1465321 and rs2058622 are in high LD with SNPs associated with low expression of *IL18RAP* in celiac disease [60]. The IBD-associated SNPs rs1551398 and rs1551399 are associated with *TRIB1* [21]. To determine whether there was a functional relationship between T-bet binding and *IL18RAP* and *TRIB1* expression, we compared gene expression profiles of wild type and T-bet<sup>-/-</sup> naïve CD4<sup>+</sup> T cells polarised in Th1 conditions. As expected, genes known to be



**Fig 5. Genetic variants alter T-bet binding *in vivo*.** **A.** Genomic context of rs1465321 and rs2058622, which is in high LD ( $r^2 = 1.0$ ) with rs1465321. **B.** T-bet ChIP and input sequencing reads that cross rs2058622 (chr2: 102985274–102985565; left) or rs1465321 (chr2: 102986477–102986768; right) in two donors heterozygous for rs1465321. In each case, the number of reads that match the reference allele are shown in black and the alternative allele in green. **C.** T-bet ChIP and input (Inp) sequencing reads at the set of 19 additional heterozygous SNPs that exhibited allelic imbalanced T-bet binding. For each SNP, the color shows fold-enrichment in the number of sequencing reads matching the Ref or Alt allele, relative to the average number of reads across all samples, as indicated by the scale on the right hand side. SNPs are divided into those exhibiting greater T-bet binding to the reference (Ref) allele (Ref > Alt, top) or the alternative (Alt) allele (Alt > Ref, bottom).

doi:10.1371/journal.pgen.1006587.g005

positively regulated by T-bet were significantly downregulated in T-bet<sup>-/-</sup> cells, including Interferon- $\gamma$  (*Ifng*) and *Tim-3* (*Havcr2*), while the housekeeping genes *Gapdh*, *Actb* and *Hprt* remained unchanged (Fig 6A). *Il18rap* was also significantly downregulated in the absence of T-bet, implying a positive regulatory role for T-bet in modulating its expression (Fig 6A). In contrast, *Trib1* was significantly upregulated in T-bet<sup>-/-</sup> cells, implying that T-bet functions to repress this gene. Consistent with a direct role for T-bet in regulating *Il18rap* and *Trib1* expression, multiple T-bet binding sites were located within intronic regions of murine *Il18rap* and downstream of *Trib1* (S4 Fig). Thus, these data support a direct role for T-bet binding in the regulation of *IL18RAP* and *TRIB1* expression.



**Fig 6. rs1465321 is an eQTL for *IL18RAP* and celiac disease.** **A.** RNA abundance (size factor-normalised counts) for selected genes in wild type (WT) and T-bet deficient (T-bet KO) naive lymphocytes cultured under Th1 polarising conditions. \* significant change in expression ( $p < 0.05$ , Wald test after Benjamini-Hochberg correction). **B.** "Locus-zoom" plot showing the distribution of association p-values for celiac disease in the *IL18RAP/IL18R1* chromosome region (genes shown below

panel D). The x-axis shows the chromosome position. Colors show the level of linkage disequilibrium with rs1465321, which is indicated with a purple spot. **C.** “Locus-zoom” plot showing the distribution of association p-values for *IL18RAP* eQTL in 1,214 whole blood RNA samples. The Bayesian statistic for colocalisation with the celiac disease signal shows a posterior probability *against* colocalisation (PP3) greater than 99%, indicating that this primary whole blood *IL18RAP* eQTL signal is not compatible with a shared causal variant with celiac disease. **D.** “Locus-zoom” plot showing the secondary *IL18RAP* eQTL (conditional on rs1985329) signal in the same 1,214 whole blood RNA samples. For this secondary signal, the Bayesian statistic for colocalisation with the celiac disease signal shows a posterior probability *in favour of* colocalisation (PP4) greater than 99%, indicating that this secondary whole blood *IL18RAP* eQTL signal is compatible with a shared causal variant with celiac disease.

doi:10.1371/journal.pgen.1006587.g006

## rs1465321 is associated with *IL18RAP* expression and co-localises with celiac disease risk

We next explored whether the genotype of rs1465321 could control the expression of nearby genes and how this potential eQTL related to celiac disease susceptibility (Fig 6B). Celiac disease association was based on a case control association study of 12,041 celiac disease cases and 12,228 controls [23]. Using a gene expression dataset of 1,214 samples [62] we found a strong correlation between rs1465321 genotypes and *IL18RAP* expression level ( $p < 10^{-100}$ , Fig 6C). No other gene showed a significant association with rs1465321. However, this SNP did not display the greatest eQTL association compared with other variants in the region, which could suggest a lack of a causal role. Moreover, using a previously developed methodology [63], we established that the eQTL and disease association signals in the *IL18RAP* regions were unlikely to be driven by the same genetic variant (posterior probability supporting a shared variant  $< 1\%$ , Fig 6C). However, a stepwise regression analysis of the eQTL data shows that after accounting for the primary eQTL signal (conditional on rs1985329), a second eQTL association was clearly detectable ( $p < 10^{-30}$ ). This suggested that at least two independent variants, with distinct biological mechanisms, are controlling *IL18RAP* mRNA expression. Interestingly, this secondary eQTL signal co-localized with the celiac disease risk signal (Fig 6D, posterior probability supporting a shared variant  $> 99\%$ ). Moreover, rs1465321 is one of the most strongly associated genetic variants for this secondary eQTL signal, with the disease-associated A allele, which exhibited reduced T-bet binding, associated with reduced *IL18RAP* expression. Therefore, our combined fine-mapping disease eQTL data are consistent with rs1465321 affecting *IL18RAP* expression through altered binding of T-bet.

## Discussion

We have found that IBD and celiac disease-associated SNPs are significantly enriched at T-bet binding sites. Surprisingly, this association is not observed for RA or psoriasis, suggesting it may be specific for mucosal inflammatory disease. Furthermore, we have identified genetic variants that alter T-bet binding to DNA, both *in vitro* and *in vivo*, including rs1465321, which we also identify as an eQTL for *IL18RAP* and celiac disease. Thus, these data provide a mechanistic explanation for why a single base change at this locus is associated with changes in gene expression and disease risk.

Although some studies have identified sequence variants that modulate transcription factor binding, alterations in the binding of Th lineage-specifying factors at disease-associated variants has not previously been identified. Our discovery that SNPs associated with IBD and celiac disease alter T-bet occupancy reveals that genetic variants can have a significant impact on the function of key master regulator transcription factors that govern cell fate. The strong association of T-bet binding sites with mucosal autoimmune/inflammatory diseases suggests that other disease-associated variants also act to alter the binding of this critical immune

regulator, with important consequences for T cell polarisation and lineage-specific gene expression.

That T-bet binding sites are associated with mucosal autoimmune disease, but not with RA or psoriasis is somewhat surprising, because all of these diseases have been linked to aberrant Th1 responses [2]. However, mucosal disease is more strongly associated with aberrant Th17 responses, which are repressed by T-bet [3,4,64,65], providing a mechanistic rationale for our findings. We and others have recently shown that T-bet plays a critical and non-redundant role in the function of ILCs [2,7,11–15]. It is therefore feasible that the association of mucosal autoimmune disease-associated SNPs with T-bet binding sites reflects alterations to T-bet binding in ILCs, which have a key regulatory role at mucosal surfaces. Expanding our fGWAS analysis to other autoimmune conditions will be necessary to fully establish the specificity of the association of T-bet with SNPs associated with mucosal disease.

Significantly, we have demonstrated that T-bet binding is enriched at disease-associated SNPs that have high posterior probabilities [21]. This suggests that more T-bet bound variants will be discovered when other IBD loci are subjected to fine-mapping analysis. We further found that the disease-associated alleles of rs1551398 and rs1551399 both reduce T-bet binding *in vivo*. These SNPs are located upstream of *TRIB1*, a gene that is upregulated in the mucosa of both UC and CD patients [66]. Consistent with this, we find that T-bet functions to repress *Trib1* expression, suggesting that the disease-associated alleles may increase disease risk by abrogating T-bet-mediated repression of this gene. T-bet also binds at 2 other sites near *TRIB1* (rs28510097 and rs1551400) and, together, these 4 SNPs account for 55% posterior probability of association for this locus [21].

We also identified rs1465321, located within an intron of *IL18R1*, to exhibit allele-imbalanced T-bet binding. This SNP is an eQTL for *IL18RAP* and celiac disease risk, with the minor disease-associated allele linked with reduced T-bet binding and *IL18RAP* gene expression. *IL18RAP* and *IL18R1* together form the IL-18 receptor. Signaling through this receptor, IL-18 synergizes with IL-12 to induce IFN $\gamma$ . rs1465321 is in high LD with the lead SNP in this locus for celiac disease [60]. Although our data are consistent with rs1465321 altering *IL18RAP* expression through altered binding of T-bet, we cannot rule out that variants in strong LD with rs1465321 could also be causal, such as rs2058622 that also exhibits allele-imbalanced T-bet binding. Given that T-bet acts through multiple sites to regulate its target genes [52, 54, 67, 68], it is likely to be the combined effect of the haplotype that is relevant. ChIP-seq for T-bet in individuals heterozygous for other disease-associated SNPs will likely reveal further examples of genetic variants that modulate T-bet binding.

Our finding that there are two independent eQTLs for *IL18RAP*, and that only one of these is associated with celiac disease (Fig 6), suggests that the level of *IL18RAP* expression may not be functionally relevant for disease susceptibility. Alternatively, it is possible that the two independent eQTLs for *IL18RAP* represent different enhancers that mediate transcriptional activation in different cells or in response to different stimuli, and that *IL18RAP* expression level is only relevant for celiac disease in one cell type or in response to a particular signal.

Attempts to determine the likely effect of non-coding sequence variants have mostly focused on identifying overlapping transcription factor binding motifs or overlapping sites of transcription factor binding, DNase I hypersensitivity or DNA and histone modification. Our analysis of allele-specific T-bet ChIP-seq data shows that genetic variants within transcription factor binding sites do not necessarily alter transcription factor binding. Similarly, genetic variants that do impact transcription factor binding do not necessarily lie within the predicted motif. Thus, confirmation of allele-specific binding events is necessary to confirm that a SNP does indeed impact transcription factor function and provides a mechanistic link between genetic variation and disease risk.

We have established the feasibility of using flow cytometry to assay allelic effects on transcription factor binding, and validated this technique through both the traditional pull-down assay and allele-specific ChIP-seq. As flow cytometric methods can be easily automated, this method provides a more rapid means to assay large numbers of allelic variants compared to traditional pull-down methods. Using this OligoFlow method, we identified alterations in T-bet binding at rs11135484, in high LD with a SNP associated with Crohn's disease and with rs1006353, the closest neighbor of which is *MTIF3*, associated with body mass index [69]. Interestingly, T-bet has been linked with regulation of insulin sensitivity and visceral adiposity [70].

In summary, we have identified a specific association between T-bet binding sites and mucosal autoimmune disease variants and determined that such genetic variants modulate T-bet binding in cells. This suggests that altered binding of T cell master regulators can predispose individuals to specific autoimmune and inflammatory conditions. This study establishes a scalable method that can be used to explore the impact of genetic variation on the function of other lineage-specifying transcriptional factors. These insights will identify molecular mechanisms that underlie the genetic basis of autoimmune diseases and suggest new therapies for their treatment.

## Materials and methods

### T-bet hit-SNP identification

ChIP-seq for T-bet in human Th1 cells was performed previously [53–54] (GEO accessions: GSE31320 and GSE62486) and binding sites were identified from the merged dataset with MACS 1.4 ( $p < 10^{-7}$ ) [71]. The positions of T-bet peaks were identified relative to gene transcription start sites annotated in RefSeq. The GWAS catalogue was downloaded from the NHGRI [55] on December 12th 2014. SNPs were checked against dbSNP and 4 SNPs called 'suspect' removed. SNPs that had been merged with other IDs were checked against HapMap3 and the ID given in HapMap3 used in downstream analysis. SNPs not in HapMap3 were removed, giving 13,936 autosomal SNPs in the final analysis. Data were analysed using the bioconductor *snpMatrix* programme (recently updated to *snpStats*) [72, 73]. SNPs in high LD ( $r^2 > 0.8$  with a SNP from the GWAS catalogue) were obtained from HapMap3 [74], giving a total of 127,594 SNPs. These were then overlapped with the T-bet binding sites. To identify the number of independent LD blocks were represented by the 926 T-bet bound SNPs, we used the *SNPclip* module of LDlink to reduce any SNPs in high LD to a single tag SNP, using a  $R^2$  threshold of 0.8 and a MAF (Minimum Allele Frequency) threshold of 0.01.

### Comparison to H3K27ac and DHS

ChIP-seq data for IgG, H3K27ac and total H3 were taken from GSE62486. Sequence reads were trimmed to remove low quality bases and to remove adapters and aligned using Bowtie (default settings) to hg19. Peaks of H3K27ac were identified with MACS ( $p < 10^{-7}$ ) [71]. DHS data were obtained from ENCODE (GEO accession GSM736592) [75,76]. Average binding profiles were calculated across 4 kb windows centred on hit-SNPs using *ngsplot* [77]. Data were visualized on the UCSC genome browser by calculating tag density in 10bp windows, normalizing to reads per million total reads and subtracting background (input for T-bet and H3 for H3K27ac), as described [54].

### Human SNP genotyping

Individuals heterozygous for rs1465321 were identified from the Twins UK cohort at the Guy's and St Thomas' NHS Foundation Trust (GSTT) Bioresource, where HumanHap610Q

Illumina array data is available for all registered participants. The Illumina calling algorithm [78] was used to assign genotypes from array data. Before imputation, quality controls were applied, with exclusion of all samples with: (1) call rate <98%, (2) heterozygosity across all SNPs  $\geq 2$  standard deviations from the sample mean; (3) evidence of non-European ancestry as assessed by PCA comparison with HapMap3 populations; (4) observed pairwise IBD probabilities suggestive of sample identity errors. We also corrected zygosity based on IBD probabilities. Quality controls were also applied to each individual SNP using the following exclusion criteria: (1) Hardy-Weinberg p-value <  $10^{-6}$  (assessed in a set of unrelated samples); (2) MAF < 1% (assessed in a set of unrelated samples); (3) SNP call rate <97% (SNPs with MAF  $\geq 5\%$ ) or < 99% (for  $1\% \leq \text{MAF} < 5\%$ ). Finally all the alleles were aligned to the forward strand of HapMap2. After completion of both sample and SNP quality controls checks, imputation was performed using the IMPUTE software package (v2) [79] using HapMap2 as a reference panel (HapMap2, rel. 22, combined CEU+YRI+ASN panels). Heterozygous SNPs were selected using PLINK (version 1.0.7) [80] “—recode-rlist” option on the imputed dataset. A final QC stage was applied on all the heterozygous SNPs, excluding all those polymorphisms with an imputation quality score  $\leq 0.8$ .

### Allele-specific T-bet ChIP-seq

In accordance with the Department of Health’s Research Governance Framework for Health and Social Care, ethical approval for this study was gained from the South London Research Ethics Committee (Ref:15/LO/0151), and from the Department of Research and Development at GSTT NHS Trust (Ref:RJ115/N122). Approval was also gained from the GSTT National Institute of Health Research (NIHR) Bioresource for recruitment of individuals registered on the Bioresource and heterozygous for rs1465321. All of the subjects in this study gave written consent. Blood was taken from two individuals heterozygous for the desired SNP. CD4+ T cells were purified from whole blood leukocytes using CD4 microbeads (Miltenyi Biotec) and naïve CD4+ T-cells sorted by FACS selection for CD4+ CD45RA+ CD4RO- CD25- CCR7+ cells. Sorted naïve T-cells were activated with anti-CD3/CD28 and polarized under Th1 conditions (IL2, IL12 and anti-IL4) for 7 days [54]. Cells were then crosslinked and ChIP-seq for T-bet performed with a custom-made polyclonal antibody [54]. Libraries were quantified using the KAPA library quantification kit and sequenced (150 bp single-end) with an Illumina NextSeq. Sequence reads were trimmed and aligned to hg19 as before. Peak regions for both donor 1 and 2 were identified separately using MACS 1.4. Broad shallow peaks were filtered, intersecting peaks identified with Bedtools (n = 8185), and then narrowed to the central 400 bp. Potential SNP variants within these intersecting peak regions were extracted from dbSNP version 138 (assembly hg19, n = 490,310). SNP sites for further analysis were determined from the Bowtie aligned bam files as containing >1 reads with both Ref and Alt bases in both ChIP and Input samples from both donors (n = 9058). This list was then compared to the set of heterozygous SNPs identified by the SNP array analysis (n = 2621 high confidence heterozygous SNPs). Reads surrounding these sites were extracted into R using the Bioconductor Rsamtools and GenomicRanges packages. The reads were split by Ref and Alt alignment for visualization using the GenomicAlignments package.

To test whether T-bet exhibited allelic imbalanced binding at rs1465321 and at SNPs in high LD, we used a binomial test. Donor 1 and 2 p-values were combined using the Fisher method. rs1465321 and rs2058622 showed significant allelic imbalance ( $p < 0.01$ ) in the T-bet ChIP samples and allelic balance ( $p > 0.1$ ) in the Input samples. To identify other heterozygous SNPs that exhibited allelic-imbalanced T-bet binding, we used a binomial test to identify heterozygous SNPs at which significantly more reads were reported for one allele compared to



the other in both T-bet ChIP samples (Benjamini-Hochberg adjusted  $p < 0.05$ ) but not imbalanced in the Input samples from either donor (unadjusted  $p > 0.4$ ). This produced a list of 19 additional SNPs (S4 Table).

T-bet ChIP-seq data from donors 1 and 2 heterozygous for rs1465321 are available at GEO under accession GSE81881.

## RNA-seq analysis

Data-sets for wild-type and T-bet deficient CD4<sup>+</sup> T cells polarised in Th1 and Th2 conditions were obtained from GEO (GSE38808). Raw reads were aligned to the mm10 build of the murine genome using Subread [81], and subsequently mapped to RefSeq genes using featureCounts [82]. DESeq2 was used to normalise read counts by size factors, and call differentially regulated genes using an empirical Bayes model and the Wald test followed by Benjamini-Hochberg correction for multiple testing [83].

## Motif analysis

The presence of T-bet motifs was assessed using FIMO [84] using previously compiled matrices for T-bet binding obtained by ChIP-seq [54]. Sequences for T-bet binding sites were obtained from the hg19 reference genome and SNPs were manually altered to the alternative allele.

## fGWAS

*fGWAS* analysis was performed as described in [56] using *fGWAS* version 0.3.3 with case control setting. Data were prepared for *fGWAS* using R and the *GenomicRanges* package to compute overlap between binding sites and SNPs. Publicly available GWAS data were downloaded from the websites of the relevant consortiums for UC, Crohn's disease (<http://www.ibdgenetics.org/downloads.html>), coronary artery disease (<http://www.cardiogramplusc4d.org/downloads/>), and rheumatoid arthritis (<http://plaza.umin.ac.jp/~yokada/datasource/software.htm>). Psoriasis data are from [85]. T-bet binding sites were identified as described above. ENCODE ChIP-seq, FAIRE-seq and DNaseI hypersensitivity datasets were obtained from the ENCODE website in bed format ([http://ftp.ebi.ac.uk/pub/databases/ensembl/encode/integration\\_data\\_jan2011](http://ftp.ebi.ac.uk/pub/databases/ensembl/encode/integration_data_jan2011)). The complete ENCODE datasets combines DNaseI (125 annotations), FAIRE-seq (24 annotations), histone marks (117 annotations) and transcription factor binding site datasets (S2 Table). In addition, we included GATA3 binding sites in Th1 and Th2 cells (GSE31320) [54], FOXP3 binding sites in Tregs [86], NF- $\kappa$ B binding sites in lymphoblastoid cells [47] (GSE19486), and H3K27ac [58] and DHS [57] in Th1 and Th2 cells.

## eQTL analysis

Celiac disease association was based on a case control association study of 12,041 celiac disease cases and 12,228 controls [23]. Gene expression data were taken from [62]. eQTL analysis was performed as described [63]. eQTL p-values were obtained by fitting a linear trend test regression between the expression of each gene and all variants 200 kb upstream and downstream from each probe. Posterior computation was performed as described [63]. The regional association plots for the eQTL and Biomarker datasets were created using LocusZoom [87] (<http://csg.sph.umich.edu/locuszoom/>).

Colocalisation analysis was performed using the R package COLOC [63] based on single variant summary statistics (log odds ratio, standard error for the log odds ratio for case control

and effect size and standard error for effect size for eQTL study, in addition to MAF and physical position for each variant) and with the default settings provided with the R package.

### Cell culture for oligonucleotide pull-down assays

Human CD4<sup>+</sup> cells were isolated from buffy coats (UK National Blood Service, used under REC reference number 10/H0804/65 from SE London Research Ethics Committee 2) using RosetteSep human CD4<sup>+</sup> T cell enrichment cocktail (STEMCELL Technologies) according to manufacturer's instructions and polarised towards a Th1 phenotype in supplemented RPMI as described in above. Cells were harvested after a total of seven days of culture. YT cells were cultured in RPMI medium (PAA) supplemented with 50 units/ml penicillin, 50 µg/ml streptomycin (Gibco), 10 mM HEPES buffer solution (Fisher Scientific), 1 mM sodium pyruvate (Gibco), 1 × minimum essential medium-non essential amino acids (Gibco), 2 mM L-glutamine (Gibco) and 10% foetal bovine serum (PAA). All cells were maintained at 37°C in 5% CO<sub>2</sub>.

### OligoFlow and pull-downs

Forward and reverse single-stranded oligos (Integrated DNA Technologies, [S3 Table](#)) for each allele of each SNP were annealed by incubating at 94°C for 5 mins, 65°C for 10 mins, 25°C for 10 mins and 4°C thereafter in annealing buffer (50 mM Tris pH 8, 7 mM MgCl<sub>2</sub> and 1 mM DTT). For pull-down and western blot, 20 µl of streptavidin agarose beads (Sigma) were used per sample. For OligoFlow, 50 µl of Sphero streptavidin polystyrene particles (Spherotech #SVP-100-4) were used per sample. Beads were washed twice in PBS and then once in annealing buffer. Beads were then incubated with double-stranded oligonucleotides for 1 hr at 4°C, washed twice in oligo buffer (10 mM Tris pH 8, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 5% glycerol, 1 mg/ml BSA Fraction V, 20 µg/ml dI/dC (Sigma, P4929) and Complete protease inhibitor (Roche) and finally resuspended in 450 µl oligo buffer. Cells (30 million per sample) were washed twice in PBS and lysed in 1 ml hypotonic buffer (20 mM HEPES pH 8, 10 mM KCl, 1 mM MgCl<sub>2</sub>, 0.1% Triton X-100, 5% glycerol, 1 mM DTT and Complete protease inhibitor) on ice for 5 mins. Lysed cells were pelleted and resuspended in 150 µl hypertonic buffer (20 mM HEPES pH 8, 400 mM NaCl, 1 mM EDTA, 0.1% Triton X-100, 5% glycerol, 1 mM DTT and Complete protease inhibitor). Debris was pelleted, 180 µl of supernatant containing nuclear extract added to the beads and incubated on a rotor for 1 hour at 4°C. For western blotting, samples were then washed three times in oligo buffer and resuspended in SDS loading buffer. For OligoFlow, 0.25 µg of anti-T-bet Alexa647 antibody (clone 4B10, BioLegend) was added and samples incubated for a further 1 hr at 4°C. Data (at least 30,000 events) were acquired on a FACSCanto flow cytometer (BD Biosciences).

### Western blotting

Oligonucleotide pull-down samples were heated in SDS loading buffer before transfer to nitrocellulose membrane. Samples were blocked in 5% milk in TBS-T (1 hr, RT) and incubated with 1:1000 anti-T-bet (clone eBio4B10 (eBioscience); 4°C overnight). Blots were washed before addition of anti-mouse-HRP (GE Healthcare) and visualised with Enhanced Chemiluminescent Substrate (PerkinElmer) and exposed to film.

### Supporting information

**S1 Fig. T-bet binding at disease-associated SNPs. A.** T-bet binding at further example T-bet hit-SNPs. The number of sequencing reads from T-bet, IgG control and H3K27ac ChIP-enriched DNA are plotted per million input-subtracted total reads and aligned with the human

genome. DNaseI hypersensitivity data (2 replicates) are from ENCODE. **B.** The frequency distribution of posterior probabilities for association with IBD (from [21]) for SNPs [21] that do not overlap (left) or that do overlap (right) a T-bet binding site. SNPs that overlap a T-bet binding site tend to have a higher posterior probability (binomial regression,  $p = 6.3 \times 10^{-6}$ ).  
(PDF)

**S2 Fig. Confirmation of altered T-bet binding at rs1465321 by oligonucleotide pull-down.** **A** The effect of the different rs1465321 alleles on T-bet binding was assessed by oligonucleotide pull-down followed by immunoblotting—a representative blot for rs1465321 is shown. **B** Quantification of immunoblot band density, normalised to allele A. Error bars show standard deviation ( $n = 4$ ). The difference in T-bet binding to the G compared to A allele of rs1465321 was significant ( $p = 0.048$ , paired t-test).  
(PDF)

**S3 Fig. Heterozygous SNPs showing imbalanced T-bet binding.** T-bet ChIP and input sequencing reads that cross rs888096, rs1551399 and rs1551398 in two donors. In each case, the number of reads that match the reference allele are shown in black and the alternative allele in green.  
(PDF)

**S4 Fig. T-bet binding at the mouse *Il18r1/Il18rap* and *Trib1* loci.** The number of sequencing reads from T-bet ChIP-enriched DNA from WT (GSM998272 and GSM836124) and T-bet KO mouse Th1 cells (GSM998273) plotted per million input-subtracted total reads and aligned with the mouse genome (mm9) at the *Il18r1/Il18rap* (A) and the *Trib1* (B) loci.  
(PDF)

**S1 Table. Summary data for T-bet hit-SNPs.**  
(XLSX)

**S2 Table. Functional annotation datasets used for fGWAS.**  
(XLS)

**S3 Table. Oligonucleotide sequences used for OligoFlow.**  
(DOCX)

**S4 Table. Heterozygous SNPs showing imbalanced T-bet binding *in vivo*.**  
(XLSX)

## Acknowledgments

We thank the UCL Cancer Institute Genomics Core Facility for ChIP-seq. Thanks to John Ambrose for data submission to GEO. We gratefully acknowledge the assistance of the National Institute for Health Research (NIHR) BioResource and TwinsUK at the Biomedical Research Centre at Guy's and St Thomas' Hospital and King's College London. We gratefully acknowledge the participation of all NIHR BioResource volunteers, and thank the NIHR BioResource Centre and staff for their contribution. We thank both the NIHR and NHS Blood and Transplant. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health.

## Author Contributions

**Conceptualization:** VP RGJ GML.

**Data curation:** KS SH RM VP RGJ.

**Formal analysis:** KS CG SH RM VP RGJ.

**Funding acquisition:** RGJ GML.

**Investigation:** KS AH RGJ.

**Methodology:** KS EP RGJ.

**Project administration:** RGJ GML.

**Resources:** LF RG.

**Supervision:** JH VP RGJ GML.

**Validation:** KS AH.

**Visualization:** KS CG SH RM VP RGJ.

**Writing – original draft:** KS RGJ GML.

**Writing – review & editing:** KS RM VP RGJ GML.

## References

1. Zhu J, Paul WE. CD4 T cells: fates, functions, and faults. *Blood* 2008; 112(5):1557–1569. doi: [10.1182/blood-2008-05-078154](https://doi.org/10.1182/blood-2008-05-078154) PMID: [18725574](https://pubmed.ncbi.nlm.nih.gov/18725574/)
2. Lazarevic V, Glimcher LH, Lord GM. T-bet: a bridge between innate and adaptive immunity. *Nat Rev Immunol* 2013; 13(11):777–789. doi: [10.1038/nri3536](https://doi.org/10.1038/nri3536) PMID: [24113868](https://pubmed.ncbi.nlm.nih.gov/24113868/)
3. Lazarevic V, Chen X, Shim JH, Hwang ES, Jang E, Bolm AN et al. T-bet represses T(H)17 differentiation by preventing Runx1-mediated activation of the gene encoding ROR $\gamma$ t. *Nat Immunol*. 2011 Jan; 12(1):96–104. doi: [10.1038/ni.1969](https://doi.org/10.1038/ni.1969) PMID: [21151104](https://pubmed.ncbi.nlm.nih.gov/21151104/)
4. Gökmen MR, Dong R, Kanhere A, Powell N, Perucha E, Jackson I, et al. Genome-Wide Regulatory Analysis Reveals That T-bet Controls Th17 Lineage Differentiation through Direct Suppression of IRF4. *The Journal of Immunology* 2013; 191(12):5925–5932. doi: [10.4049/jimmunol.1202254](https://doi.org/10.4049/jimmunol.1202254) PMID: [24249732](https://pubmed.ncbi.nlm.nih.gov/24249732/)
5. Cooper GS, Bynum ML, Somers EC. Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. *J Autoimmun* 2009; 33(3–4):197–207. doi: [10.1016/j.jaut.2009.09.008](https://doi.org/10.1016/j.jaut.2009.09.008) PMID: [19819109](https://pubmed.ncbi.nlm.nih.gov/19819109/)
6. Kappelman MD, Rifas-Shiman SL, Porter CQ, Ollendorf DA, Sandler RS, Galanko JA, et al. Direct health care costs of Crohn's disease and ulcerative colitis in US children and adults. *Gastroenterology* 2008 Dec; 135(6):1907–1913. doi: [10.1053/j.gastro.2008.09.012](https://doi.org/10.1053/j.gastro.2008.09.012) PMID: [18854185](https://pubmed.ncbi.nlm.nih.gov/18854185/)
7. Powell N, Canavan JB, MacDonald TT, Lord GM. Transcriptional regulation of the mucosal immune system mediated by T-bet. *Mucosal Immunol* 2010 Nov; 3(6):567–577. doi: [10.1038/mi.2010.53](https://doi.org/10.1038/mi.2010.53) PMID: [20844482](https://pubmed.ncbi.nlm.nih.gov/20844482/)
8. Monteleone I, Monteleone G, Del Vecchio Blanco G, Vavassori P, Cucchiara S, MacDonald TT, et al. Regulation of the T helper cell type 1 transcription factor T-bet in coeliac disease mucosa. *Gut* 2004 Aug; 53(8):1090–1095. doi: [10.1136/gut.2003.030551](https://doi.org/10.1136/gut.2003.030551) PMID: [15247173](https://pubmed.ncbi.nlm.nih.gov/15247173/)
9. Neurath MF, Weigmann B, Finotto S, Glickman J, Nieuwenhuis E, Iijima H, et al. The Transcription Factor T-bet Regulates Mucosal T Cell Activation in Experimental Colitis and Crohn's Disease. *J Exp Med* 2002; 195(9):1129–1143. doi: [10.1084/jem.20011956](https://doi.org/10.1084/jem.20011956) PMID: [11994418](https://pubmed.ncbi.nlm.nih.gov/11994418/)
10. Gaffen SL, Jain R, Garg AV, Cua DJ. The IL-23-IL-17 immune axis: from mechanisms to therapeutic testing. *Nat Rev Immunol* 2014; 14(9):585–600. doi: [10.1038/nri3707](https://doi.org/10.1038/nri3707) PMID: [25145755](https://pubmed.ncbi.nlm.nih.gov/25145755/)
11. Powell N, Walker AW, Stolarczyk E, Canavan JB, Gokmen MR, Marks E, et al. The transcription factor T-bet regulates intestinal inflammation mediated by interleukin-7 receptor+ innate lymphoid cells. *Immunity* 2012 Oct 19; 37(4):674–684. doi: [10.1016/j.immuni.2012.09.008](https://doi.org/10.1016/j.immuni.2012.09.008) PMID: [23063332](https://pubmed.ncbi.nlm.nih.gov/23063332/)
12. Ermann J, Staton T, Glickman JN, de Waal Malefyt R, Glimcher LH. Nod/Ripk2 signaling in dendritic cells activates IL-17A-secreting innate lymphoid cells and drives colitis in T-bet $^{-/-}$ .Rag2 $^{-/-}$  (TRUC) mice. *Proc Natl Acad Sci U S A* 2014 Jun 24; 111(25):E2559–66. doi: [10.1073/pnas.1408540111](https://doi.org/10.1073/pnas.1408540111) PMID: [24927559](https://pubmed.ncbi.nlm.nih.gov/24927559/)
13. Garrett WS, Lord GM, Punit S, Lugo-Villarino G, Mazmanian SK, Ito S, et al. Communicable ulcerative colitis induced by T-bet deficiency in the innate immune system. *Cell* 2007 Oct 5; 131(1):33–45. doi: [10.1016/j.cell.2007.08.017](https://doi.org/10.1016/j.cell.2007.08.017) PMID: [17923086](https://pubmed.ncbi.nlm.nih.gov/17923086/)

14. Klose CSN, Kiss EA, Schwierzeck V, Ebert K, Hoyler T, d'Hargues Y, et al. A T-bet gradient controls the fate and function of CCR6-ROR $\gamma$ t<sup>+</sup> innate lymphoid cells. *Nature* 2013 02/14; 494(7436):261–265. doi: [10.1038/nature11813](https://doi.org/10.1038/nature11813) PMID: [23334414](https://pubmed.ncbi.nlm.nih.gov/23334414/)
15. Sciume G, Hirahara K, Takahashi H, Laurence A, Villarino AV, Singleton KL, et al. Distinct requirements for T-bet in gut innate lymphoid cells. *J Exp Med* 2012 Dec 17; 209(13):2331–2338. doi: [10.1084/jem.20122097](https://doi.org/10.1084/jem.20122097) PMID: [23209316](https://pubmed.ncbi.nlm.nih.gov/23209316/)
16. Gutierrez-Arcelus M, Rich SS, Raychaudhuri S. Autoimmune diseases—connecting risk alleles with molecular traits of the immune system. *Nat Rev Genet* 2016; 17(3):160–174. doi: [10.1038/nrg.2015.33](https://doi.org/10.1038/nrg.2015.33) PMID: [26907721](https://pubmed.ncbi.nlm.nih.gov/26907721/)
17. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007 06/07; 447(7145):661–678. doi: [10.1038/nature05911](https://doi.org/10.1038/nature05911) PMID: [17554300](https://pubmed.ncbi.nlm.nih.gov/17554300/)
18. Dubois PCA, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* 2010; 42(4):295–302. doi: [10.1038/ng.543](https://doi.org/10.1038/ng.543) PMID: [20190752](https://pubmed.ncbi.nlm.nih.gov/20190752/)
19. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, et al. A Genome-Wide Association Study Identifies IL23R as an Inflammatory Bowel Disease Gene. *Science* 2006; 314(5804):1461–1463. doi: [10.1126/science.1135245](https://doi.org/10.1126/science.1135245) PMID: [17068223](https://pubmed.ncbi.nlm.nih.gov/17068223/)
20. Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet* 2007 print; 39(2):207–211. doi: [10.1038/ng1954](https://doi.org/10.1038/ng1954) PMID: [17200669](https://pubmed.ncbi.nlm.nih.gov/17200669/)
21. Huang H, Fang M, Jostins L, Mirkov MU, Boucher G, Anderson CA, et al. Association mapping of inflammatory bowel disease loci to single variant resolution. *bioRxiv* 2015.
22. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012 11/01; 491(7422):119–124. doi: [10.1038/nature11582](https://doi.org/10.1038/nature11582) PMID: [23128233](https://pubmed.ncbi.nlm.nih.gov/23128233/)
23. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet* 2011; 43(12):1193–1201. doi: [10.1038/ng.998](https://doi.org/10.1038/ng.998) PMID: [22057235](https://pubmed.ncbi.nlm.nih.gov/22057235/)
24. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 2015 02/19; 518(7539):337–343. doi: [10.1038/nature13835](https://doi.org/10.1038/nature13835) PMID: [25363779](https://pubmed.ncbi.nlm.nih.gov/25363779/)
25. Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 2011 May 5; 473(7345):43–49. doi: [10.1038/nature09906](https://doi.org/10.1038/nature09906) PMID: [21441907](https://pubmed.ncbi.nlm.nih.gov/21441907/)
26. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012 Sep 6; 489(7414):57–74. doi: [10.1038/nature11247](https://doi.org/10.1038/nature11247) PMID: [22955616](https://pubmed.ncbi.nlm.nih.gov/22955616/)
27. Maurano MT, Haugen E, Sandstrom R, Vierstra J, Shafer A, Kaul R, et al. Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. *Nat Genet* 2015; 2015; 47(12):1393–1401. doi: [10.1038/ng.3432](https://doi.org/10.1038/ng.3432) PMID: [26502339](https://pubmed.ncbi.nlm.nih.gov/26502339/)
28. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. *Genome Res* 2012 Sep; 22(9):1748–1759. doi: [10.1101/gr.136127.111](https://doi.org/10.1101/gr.136127.111) PMID: [22955986](https://pubmed.ncbi.nlm.nih.gov/22955986/)
29. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. *Cell* 2013 Nov 7; 155(4):934–947. doi: [10.1016/j.cell.2013.09.053](https://doi.org/10.1016/j.cell.2013.09.053) PMID: [24119843](https://pubmed.ncbi.nlm.nih.gov/24119843/)
30. Parker SC, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci U S A* 2013 Oct 29; 110(44):17921–17926. doi: [10.1073/pnas.1317023110](https://doi.org/10.1073/pnas.1317023110) PMID: [24127591](https://pubmed.ncbi.nlm.nih.gov/24127591/)
31. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet* 2013 Feb; 45(2):124–130. doi: [10.1038/ng.2504](https://doi.org/10.1038/ng.2504) PMID: [23263488](https://pubmed.ncbi.nlm.nih.gov/23263488/)
32. Onengut-Gumuscu S, Chen WM, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC, et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet* 2015 Apr; 47(4):381–386. doi: [10.1038/ng.3245](https://doi.org/10.1038/ng.3245) PMID: [25751624](https://pubmed.ncbi.nlm.nih.gov/25751624/)
33. McDaniel R, Lee B, Song L, Liu Z, Boyle AP, Erdos MR, et al. Heritable Individual-Specific and Allele-Specific Chromatin Signatures in Humans. *Science* 2010; 328(5975):235–239. doi: [10.1126/science.1184655](https://doi.org/10.1126/science.1184655) PMID: [20299549](https://pubmed.ncbi.nlm.nih.gov/20299549/)

34. Degner JF, Pai AA, Pique-Regi R, Veyrieras J, Gaffney DJ, Pickrell JK, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* 2012 02/16; 482(7385):390–394. doi: [10.1038/nature10808](https://doi.org/10.1038/nature10808) PMID: [22307276](https://pubmed.ncbi.nlm.nih.gov/22307276/)
35. Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A, et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife* 2013 Jun 4; 2: e00523. doi: [10.7554/eLife.00523](https://doi.org/10.7554/eLife.00523) PMID: [23755361](https://pubmed.ncbi.nlm.nih.gov/23755361/)
36. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol* 2013; 31(2):142–147. doi: [10.1038/nbt.2487](https://doi.org/10.1038/nbt.2487) PMID: [23334450](https://pubmed.ncbi.nlm.nih.gov/23334450/)
37. Banovich NE, Lan X, McVicker G, van dG, Degner JF, Blischak JD, et al. Methylation QTLs Are Associated with Coordinated Changes in Transcription Factor Binding, Histone Modifications, and Gene Expression Levels. *PLoS Genet* 2014 09/18; 10(9):e1004663. doi: [10.1371/journal.pgen.1004663](https://doi.org/10.1371/journal.pgen.1004663) PMID: [25233095](https://pubmed.ncbi.nlm.nih.gov/25233095/)
38. Javierre BM, Fernandez AF, Richter J, Al-Shahrour F, Martin-Subero JI, Rodriguez-Ubreva J, et al. Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res* 2010 Feb; 20(2):170–179. doi: [10.1101/gr.100289.109](https://doi.org/10.1101/gr.100289.109) PMID: [20028698](https://pubmed.ncbi.nlm.nih.gov/20028698/)
39. del Rosario RC, Poschmann J, Rouam SL, Png E, Khor CC, Hibberd ML, et al. Sensitive detection of chromatin-altering polymorphisms reveals autoimmune disease mechanisms. *Nat Methods* 2015 May; 12(5):458–464. doi: [10.1038/nmeth.3326](https://doi.org/10.1038/nmeth.3326) PMID: [25799442](https://pubmed.ncbi.nlm.nih.gov/25799442/)
40. Karczewski KJ, Dudley JT, Kukurba KR, Chen R, Butte AJ, Montgomery SB, et al. Systematic functional regulatory assessment of disease-associated variants. *Proc Natl Acad Sci U S A* 2013 Jun 4; 110(23):9607–9612. doi: [10.1073/pnas.1219099110](https://doi.org/10.1073/pnas.1219099110) PMID: [23690573](https://pubmed.ncbi.nlm.nih.gov/23690573/)
41. McVicker G, van de Geijn B, Degner JF, Cain CE, Banovich NE, Raj A, et al. Identification of genetic variants that affect histone modifications in human cells. *Science* 2013 Nov 8; 342(6159):747–749. doi: [10.1126/science.1242429](https://doi.org/10.1126/science.1242429) PMID: [24136359](https://pubmed.ncbi.nlm.nih.gov/24136359/)
42. Kilpinen H, Waszak SM, Gschwind AR, Raghav SK, Witwicki RM, Orioli A, et al. Coordinated Effects of Sequence Variation on DNA Binding, Chromatin Structure, and Transcription. *Science* 2013; 342(6159):744–747. doi: [10.1126/science.1242463](https://doi.org/10.1126/science.1242463) PMID: [24136355](https://pubmed.ncbi.nlm.nih.gov/24136355/)
43. Kasowski M, Kyriazopoulou-Panagiotopoulou S, Grubert F, Zaugg JB, Kundaje A, Liu Y, et al. Extensive variation in chromatin states across humans. *Science* 2013 Nov 8; 342(6159):750–752. doi: [10.1126/science.1242510](https://doi.org/10.1126/science.1242510) PMID: [24136358](https://pubmed.ncbi.nlm.nih.gov/24136358/)
44. Linnell J, Mott R, Field S, Kwiatkowski DP, Ragoussis J, Udalova IA. Quantitative high-throughput analysis of transcription factor binding specificities. *Nucleic Acids Res* 2004 Feb 27; 32(4):e44. doi: [10.1093/nar/gnh042](https://doi.org/10.1093/nar/gnh042) PMID: [14990752](https://pubmed.ncbi.nlm.nih.gov/14990752/)
45. Larsson L, Rymo L, Berglundh T. Sp1 binds to the G allele of the -1087 polymorphism in the IL-10 promoter and promotes IL-10 mRNA transcription and protein production. *Genes Immun* 2010; 11(2):181–187. doi: [10.1038/gene.2009.103](https://doi.org/10.1038/gene.2009.103) PMID: [20072143](https://pubmed.ncbi.nlm.nih.gov/20072143/)
46. Hummelshoj T, Ryder LP, Madsen HO, Odum N, Svejgaard A. A functional polymorphism in the Eta-1 promoter is associated with allele specific binding to the transcription factor Sp1 and elevated gene expression. *Mol Immunol* 2006 Mar; 43(7):980–986. doi: [10.1016/j.molimm.2005.05.012](https://doi.org/10.1016/j.molimm.2005.05.012) PMID: [16009426](https://pubmed.ncbi.nlm.nih.gov/16009426/)
47. Kasowski M, Grubert F, Heffelfinger C, Hariharan M, Asabere A, Waszak SM et al. Variation in transcription factor binding among humans. *Science* 2010 Apr 9; 328(5975):232–5. doi: [10.1126/science.1183621](https://doi.org/10.1126/science.1183621) PMID: [20299548](https://pubmed.ncbi.nlm.nih.gov/20299548/)
48. Schodel J, Bardella C, Sciesielski LK, Brown JM, Pugh CW, Buckle V, et al. Common genetic variants at the 11q13.3 renal cancer susceptibility locus influence binding of HIF to an enhancer of cyclin D1 expression. *Nat Genet* 2012 Mar 11; 44(4):420–5, S1–2. doi: [10.1038/ng.2204](https://doi.org/10.1038/ng.2204) PMID: [22406644](https://pubmed.ncbi.nlm.nih.gov/22406644/)
49. Zhang X, Cowper-Salari R, Bailey SD, Moore JH, Lupien M. Integrative functional genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24.3 prostate cancer risk locus. *Genome Res* 2012 Aug; 22(8):1437–1446. doi: [10.1101/gr.135665.111](https://doi.org/10.1101/gr.135665.111) PMID: [22665440](https://pubmed.ncbi.nlm.nih.gov/22665440/)
50. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* 2010 Aug 5; 466(7307):714–719. doi: [10.1038/nature09266](https://doi.org/10.1038/nature09266) PMID: [20686566](https://pubmed.ncbi.nlm.nih.gov/20686566/)
51. Adrianto I, Wen F, Templeton A, Wiley G, King JB, Lessard CJ, et al. Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus. *Nat Genet* 2011 Mar; 43(3):253–258. doi: [10.1038/ng.766](https://doi.org/10.1038/ng.766) PMID: [21336280](https://pubmed.ncbi.nlm.nih.gov/21336280/)
52. Hertweck A, Evans CM, Eskandarpour M, Lau JCH, Oleinika K, Jackson I, et al. T-bet Activates Th1 Genes through Mediator and the Super Elongation Complex. *Cell Rep* 2016; Jun 21; 15:2756–2770 doi: [10.1016/j.celrep.2016.05.054](https://doi.org/10.1016/j.celrep.2016.05.054) PMID: [27292648](https://pubmed.ncbi.nlm.nih.gov/27292648/)

53. Jenner RG, Townsend MJ, Jackson I, Sun K, Bouwman RD, Young RA, et al. The transcription factors T-bet and GATA-3 control alternative pathways of T-cell differentiation through a shared set of target genes. *Proc Natl Acad Sci U S A* 2009 Oct 20; 106(42):17876–17881. doi: [10.1073/pnas.0909357106](https://doi.org/10.1073/pnas.0909357106) PMID: [19805038](https://pubmed.ncbi.nlm.nih.gov/19805038/)
54. Kanhere A, Hertweck A, Bhatia U, Gokmen MR, Perucha E, Jackson I, et al. T-bet and GATA3 orchestrate Th1 and Th2 differentiation through lineage-specific targeting of distal regulatory elements. *Nat Commun* 2012; 3:1268. doi: [10.1038/ncomms2260](https://doi.org/10.1038/ncomms2260) PMID: [23232398](https://pubmed.ncbi.nlm.nih.gov/23232398/)
55. Hindorf L, MacArthur J, Morales J, Junkins H, Hall P, Klemm A, et al. A Catalog of Published Genome-Wide Association Studies. Accessed: <http://www.genome.gov/gwastudies>.
56. Pickrell JK. Joint Analysis of Functional Genomic Data and Genome-wide Association Studies of 18 Human Traits. *The American Journal of Human Genetics* 2014; 95(1):126.
57. Sekimata M, Pérez-Melgosa M, Miller SA, Weinmann AS, Sabo PJ, Sandstrom R, et al. CCCTC-Binding Factor and the Transcription Factor T-bet Orchestrate T Helper 1 Cell-Specific Structure and Function at the Interferon- $\gamma$  Locus. *Immunity* 2009 10/16; 31(4):551–564. doi: [10.1016/j.immuni.2009.08.021](https://doi.org/10.1016/j.immuni.2009.08.021) PMID: [19818655](https://pubmed.ncbi.nlm.nih.gov/19818655/)
58. Hawkins RD, Larjo A, Tripathi SK, Wagner U, Luu Y, Lonnberg T, et al. Global chromatin state analysis reveals lineage-specific enhancers during the initiation of human T helper 1 and T helper 2 cell polarization. *Immunity* 2013 Jun 27; 38(6):1271–1284. doi: [10.1016/j.immuni.2013.05.011](https://doi.org/10.1016/j.immuni.2013.05.011) PMID: [23791644](https://pubmed.ncbi.nlm.nih.gov/23791644/)
59. Szabo SJ, Kim ST, Costa GL, Zhang X, Fathman CG, Glimcher LH. A Novel Transcription Factor, T-bet, Directs Th1 Lineage Commitment. *Cell* 2000; 100(6):655–669. PMID: [10761931](https://pubmed.ncbi.nlm.nih.gov/10761931/)
60. Hunt KA, Zhernakova A, Turner G, Heap GAR, Franke L, Bruinenberg M, et al. Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet* 2008; 40(4):395–402. doi: [10.1038/ng.102](https://doi.org/10.1038/ng.102) PMID: [18311140](https://pubmed.ncbi.nlm.nih.gov/18311140/)
61. Franke A, McGovern DPB, Barrett JC, Wang K, Radford-Smith GL, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 2010 Dec 42: 1118–1125. doi: [10.1038/ng.717](https://doi.org/10.1038/ng.717) PMID: [21102463](https://pubmed.ncbi.nlm.nih.gov/21102463/)
62. Cheung VG, Fehrmann RSN, Jansen RC, Veldink JH, Westra H, Arends D, et al. Trans-eQTLs Reveal That Independent Genetic Variants Associated with a Complex Phenotype Converge on Intermediate Genes, with a Major Role for the HLA. *PLoS Genetics* 2011; 7(8):e1002197. doi: [10.1371/journal.pgen.1002197](https://doi.org/10.1371/journal.pgen.1002197) PMID: [21829388](https://pubmed.ncbi.nlm.nih.gov/21829388/)
63. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet* 2014 05/15; 10(5):e1004383. doi: [10.1371/journal.pgen.1004383](https://doi.org/10.1371/journal.pgen.1004383) PMID: [24830394](https://pubmed.ncbi.nlm.nih.gov/24830394/)
64. Zimmermann J, Kuhl AA, Weber M, Grun JR, Loffler J, Haftmann C, et al. T-bet expression by Th cells promotes type 1 inflammation but is dispensable for colitis. *Mucosal Immunol* 2016 Feb 17.
65. Krausgruber T, Schiering C, Adelmann K, Harrison OJ, Chomka A, Pearson C, et al. T-bet is a key modulator of IL-23-driven pathogenic CD4+ T cell responses in the intestine. *Nat Commun* 2016; 7:11627.
66. Peloquin JM, Goel G, Kong L, Huang H, Haritunians T, et al. Characterization of candidate genes in inflammatory bowel disease-associated risk loci. *JCI Insight*. 2016 Aug 18; 1(13):e87899. doi: [10.1172/jci.insight.87899](https://doi.org/10.1172/jci.insight.87899) PMID: [27668286](https://pubmed.ncbi.nlm.nih.gov/27668286/)
67. Shnyreva M, Weaver WM, Blanchette M, Taylor SL, Tompa M, et al. Evolutionarily conserved sequence elements that positively regulate IFN-gamma expression in T cells. *Proc Natl Acad Sci* 2004 Aug 24; 101(34):12622–12627 doi: [10.1073/pnas.0400849101](https://doi.org/10.1073/pnas.0400849101) PMID: [15304658](https://pubmed.ncbi.nlm.nih.gov/15304658/)
68. Schoenborn JR, Dorschner MO, Sekimata M, Santer DM, Shnyreva M, et al. Comprehensive epigenetic profiling identifies multiple distal regulatory elements directing transcription of the gene encoding interferon-gamma. *Nat Immunol* 2007 8, 732–742. doi: [10.1038/ni1474](https://doi.org/10.1038/ni1474) PMID: [17546033](https://pubmed.ncbi.nlm.nih.gov/17546033/)
69. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010 print; 42(11):937–948. doi: [10.1038/ng.686](https://doi.org/10.1038/ng.686) PMID: [20935630](https://pubmed.ncbi.nlm.nih.gov/20935630/)
70. Stolarczyk E, Vong C, Perucha E, Jackson I, Cawthorne M, Wargent E, et al. Improved Insulin Sensitivity despite Increased Visceral Adiposity in Mice Deficient for the Immune Cell Transcription Factor T-bet. *Cell Metabolism* 2013 4/2; 17(4):520–533. doi: [10.1016/j.cmet.2013.02.019](https://doi.org/10.1016/j.cmet.2013.02.019) PMID: [23562076](https://pubmed.ncbi.nlm.nih.gov/23562076/)
71. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* 2008; 9(9):1–9.
72. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004; 5(10):1–16.
73. Clayton D, Leung H-. An R Package for Analysis of Whole-Genome Association Studies. *Hum Hered* 2007; 64(1):45–51. doi: [10.1159/000101422](https://doi.org/10.1159/000101422) PMID: [17483596](https://pubmed.ncbi.nlm.nih.gov/17483596/)

74. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010 09/02; 467(7311):52–58. doi: [10.1038/nature09298](https://doi.org/10.1038/nature09298) PMID: [20811451](https://pubmed.ncbi.nlm.nih.gov/20811451/)
75. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. *Nature* 2012 09/06; 489(7414):75–82. doi: [10.1038/nature11232](https://doi.org/10.1038/nature11232) PMID: [22955617](https://pubmed.ncbi.nlm.nih.gov/22955617/)
76. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* 2012 09/06; 337(6099):1190–1195. doi: [10.1126/science.1222794](https://doi.org/10.1126/science.1222794) PMID: [22955828](https://pubmed.ncbi.nlm.nih.gov/22955828/)
77. Shen L, Shao N, Liu X, Nestler E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* 2014; 15(1):1–14.
78. Teo YY, Inouye M, Small KS, Gwilliam R, Deloukas P, Kwiatkowski DP, et al. A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics* 2007; 23(20):2741–2746. doi: [10.1093/bioinformatics/btm443](https://doi.org/10.1093/bioinformatics/btm443) PMID: [17846035](https://pubmed.ncbi.nlm.nih.gov/17846035/)
79. Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genet* 2009 06/19; 5(6):e1000529. doi: [10.1371/journal.pgen.1000529](https://doi.org/10.1371/journal.pgen.1000529) PMID: [19543373](https://pubmed.ncbi.nlm.nih.gov/19543373/)
80. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* 2007; 81(3):559–575. doi: [10.1086/519795](https://doi.org/10.1086/519795) PMID: [17701901](https://pubmed.ncbi.nlm.nih.gov/17701901/)
81. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research* 2013 May 01; 41(10):e108–e108. doi: [10.1093/nar/gkt214](https://doi.org/10.1093/nar/gkt214) PMID: [23558742](https://pubmed.ncbi.nlm.nih.gov/23558742/)
82. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014 April 01; 30(7):923–930. doi: [10.1093/bioinformatics/btt656](https://doi.org/10.1093/bioinformatics/btt656) PMID: [24227677](https://pubmed.ncbi.nlm.nih.gov/24227677/)
83. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; 15(12):1–21.
84. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics* 2011 04/01; 27(7):1017–1018. doi: [10.1093/bioinformatics/btr064](https://doi.org/10.1093/bioinformatics/btr064) PMID: [21330290](https://pubmed.ncbi.nlm.nih.gov/21330290/)
85. Tsoi LC, Spain SL, Knight J, Ellinghaus E, Stuart PE, et al. Identification of 15 new 23 psoriasis susceptibility loci highlights the role of innate immunity. *Nature Genetics* 2012; 24 44, 1341–1348.
86. Birzele F, Fauti T, Stahl H, Lenter MC, Simon E, Knebel D, et al. Next-generation insights into regulatory T cells: expression profiling and FoxP3 occupancy in Human. *Nucleic Acids Res* 2011 10/01; 39(18):7946–7960. doi: [10.1093/nar/gkr444](https://doi.org/10.1093/nar/gkr444) PMID: [21729870](https://pubmed.ncbi.nlm.nih.gov/21729870/)
87. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, et al. Locuszoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010; 26: 2336–2337. doi: [10.1093/bioinformatics/btq419](https://doi.org/10.1093/bioinformatics/btq419) PMID: [20634204](https://pubmed.ncbi.nlm.nih.gov/20634204/)