Systems biology of degenerative diseases

Lam, Simon

Awarding institution:
King's College London

Faculty of Dentistry, Oral &
Craniofacial Sciences
Centre for Host-Microbiome Interactions

Systems biology of degenerative diseases

Simon Lam

PhD in Dental and Health Sciences Research Thesis
Submitted for the Degree of Doctor of Philosophy in Systems Biology

SYSTEMS BIOLOGY OF DEGENERATIVE DISEASES

PhD IN DENTAL AND HEALTH SCIENCES RESEARCH
THESIS

SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN
SYSTEMS BIOLOGY

Simon Lam
(19049018)

Abstract

Technological advances in the digital age mean that researchers can now produce vast quantities of data. These data have the potential to shape the way that human health and disease are viewed. Mathematical and statistical analysis of high-dimensional biological molecular data, or omics data, reveals a paradigm shift that human disease is a product of faulty metabolism, be it caused by genetics or the environment. It therefore follows that metabolic modulation should be able to act as a treatment for such diseases, and metabolic considerations should be factored into prophylactic and curative medicines.

In this doctoral thesis incorporating papers, I describe the role of systems biology in the analysis of omics data and how it is used to drive discovery of treatments for human degenerative diseases. The current investigation describes a full journey around the cycle of data-driven biomarker identification, use of animal models of metabolism, and human clinical trials. This work introduces a focus shift away from single toxic species and towards reporter metabolites whose modulation is more tangible and have the potential to reverse the degenerative phenotype. I illustrate this with work focusing on Alzheimer's disease, Parkinson's disease, congenital generalised lipodystrophy, and muscle stem cells.

Keywords: systems biology, bioinformatics, degeneration, Alzheimer's disease, Parkinson's disease, lipodystrophy, muscle regeneration

Simon Lam

19049018

simon.1.lam@kcl.ac.uk

Centre for Host-Microbiome Interactions

Faculty of Dentistry, Oral & Craniofacial Sciences

King's College London

Abbreviations

AD	Alzheimer's disease
ADAS-COG	Alzheimer's Disease Assessment Scale – Cognitive
ALT	alanine aminotransferase
AST	aspartate aminotransferase
ARA	arachidonic acid
ATAC	assay for transposase-accessible chromatin
cDNA	complementary DNA
CGL	congenital generalised lipodystrophy
CMA	combined metabolic activators
COBRA	Constraint-Based Reconstruction and Analysis
EZH2	enhancer of zeste homologue 2
FANTOM5	Functional Annotation of the Mammalian Genome 5
FKPM	fragments per kilobase of sequence per million reads
GEM	genome-scale metabolic model
GEO	Gene Expression Omnibus
GGT	γ-glutamyl transferase
GO	Gene Ontology
GPR	gene-protein-reaction
GTEx	Genotype-Tissue Expression
HETE	hydroxyeicosatetraenoic acid
HPLC	high-performance liquid chromatography

KEGG	Kyoto Encyclopedia of Genes and Genomes
MG	myasthenia gravis
MND	motor neuron disease
MoCA	Montreal Cognitive Assessment
MS	mass spectrometry
NAD	nicotinamide adenine dinucleotide
NADP	nicotinamide adenine dinucleotide phosphate
NAFLD	non-alcoholic fatty liver disease
NASH	non-alcoholic steatohepatitis
NDD	neurodegenerative disease
NLD	neurological disease
OMIM	Online Mendelian Inheritance in Man
PD	Parkinson's disease
RAVEN	Reconstruction, Analysis and Visualisation of Metabolic Networks
ROS	reactive oxygen species
ROSMAP	Religious Orders Study and Rush Memory and Aging Project
SNP	single nucleotide polymorphism
TPM	transcripts per million reads

List of papers

Paper I: Systems analysis reveals ageing-related perturbations in retinoids and sex hormones in Alzheimer's and Parkinson's diseases

Lam S, Hartmann N, Benfeitas R, Zhang C, Arif M, Turkez H, Uhlén M, Englert C, Knight R, Mardinoglu A. 2021, *Biomedicines*, **9**(10): 1310.

This paper investigated molecular similarities between Alzheimer's and Parkinson's disease patients and reclassified patients into subclasses of molecularly similar disease using machine learning. A zebrafish ageing model was used to validate predicted the predicted contribution of ageing. In this paper, I was responsible for conceptualising the human component of the work. I performed all data analysis with the exception of the network analysis, interpreted all the results, and wrote the manuscript.

Paper II: Combined metabolic activators improve cognitive functions without altering motor scores in Parkinson's disease

Yulug B, Altay O, Li X, Hanoglu L, Cankaya S, **Lam S**, Yang H, Coskun E, İdil E, Nogaylar R, Hacımuftuoglu A, Arif M, Shoaie S, Zhang C, Nielsen J, Turkez H, Borén J, Uhlén M, Mardinoglu A. Manuscript.

This paper investigated the effect of combined metabolic activators in Parkinson's disease patients. In this study, I performed the statistical analysis into predicting which patients would respond the best to the treatment and contributed to the writing of the manuscript.

Paper III: Combined metabolic activators improve cognitive functions in Alzheimer's disease

Yulug B, Altay O, Li X, Hanoglu L, Cankaya S, **Lam S**, Yang H, Coskun E, İdil E, Nogaylar R, Bayram C, Bolat I, Öner S, Tozlu ÖÖ, Arslan ME, Hacımuftuoglu A, Yıldırım S, Arif M, Shoaie S, Zhang C, Nielsen J, Turkez H, Borén J, Uhlén M, Mardinoglu A. Amended in "Combined metabolic activators improve metabolic functions in the animal models of neurodegenerative diseases", 2023, *Life Sciences*, **314**: 121325; and in "Combined metabolic activators improve cognitive functions in Alzheimer's disease patients: A randomised, double-blinded, placebo-controlled phase-II trial", 2023, *Translational Neurodegeneration*, **12**: 4.

This paper investigated the effect of combined metabolic activators in Alzheimer's disease patients. In this work, I performed the metabolic modelling analysis and the statistical analysis into predicting which patients would respond best to the treatment and contributed to the writing of the manuscript.

Paper IV: Machine learning analysis reveals biomarkers for the detection of neurological diseases

Lam S, Arif M, Song X, Uhlen M, Mardinoglu A. 2022, *Frontiers in Molecular Neuroscience*, 15: 889728.

This paper analysed UK Biobank data from Alzheimer's, Parkinson's, motor neuron disease, and mysasthenia gravis patients for the prediction of biomarkers in biochemical measurements and the genome. I led this research. I conceptualised the project for this short report, performed all the analyses, interpreted all the results, and wrote the manuscript.

Paper V: Omics signatures show liver X receptor–driven NASH in congenital generalised lipodystrophy

Lam S, Benfeitas R, Mardinoglu A, Zeybel M. Manuscript.

This paper investigated the molecular differences between patients with NASH with and without congenital generalised lipodystrophy. For my part in this paper, I performed all the bioinformatic analyses and interpreted those results, and wrote them up.

Paper VI: Regulation of lipid signalling by the polycomb repressor complex protein EZH2 during regeneration drives muscle regeneration after injury

Lam S, Mardinoglu A, Knight R. Manuscript.

This paper studied the muscle injury regenerative pathway in zebrafish. I contributed to this paper by performing all the bioinformatic analyses, interpreting those results, and writing those sections in the manuscript.

List of figures and tables

collect proteomic and metabolomic data. (C) Other high-throughput assays can be used to capture other types of omics data, such as epigenomic data.

Figure 6. Static maps.

(A) The genetic code connects genetic information with protein sequences. (B) Proteins catalyse biochemical reactions. Sample chemical equations were taken from KEGG Reaction.

Figure 7. High-throughput data generation and processing.

(A) Nucleic acid sequencing produces sequencing reads which are aligned to a reference sequence. Illumina sequencing and pseudoalignment by Kallisto are shown. (B) Mass spectrometry can sequence peptides and identify metabolites in a sample.

Figure 8. Scope of data and methods by paper in the current investigation.

Zebrafish data were used in **Papers I** and **VI**. Genotype data were used in **Paper IV**. All **Papers I-VI** used transcriptomic differential analysis. Mathematical metabolic modelling was used in **Papers I**, **III**, and **VI**. Other omics were used in **Paper V** and **Paper VI**, namely, methylomics and lipidomics, respectively. Gene coexpression networks were explored in **Papers I** and **VI**. Machine learning was harnessed in **Papers I** and **IV**. Papers **II** and **III** report the results of a clinical trial.

Table of contents

1. Introduction

The human experience is defined by a state of constant flux. Voluntary and involuntary shifting sands define each individual and the species as a whole. Individuals differ from each other in space and themselves in time. Change is a force that cannot be resisted, and it is seemingly random, unpredictable, and irrevocable.

Such has been the natural order of humanity's interaction with the environment for millennia to the present day. Nature provides the selection pressures for evolutionary change. Nature bestows its resources for human consumption, shaping everything from human behaviour and interdependence to diet and personal decision-making.

But flux on the large scale, although operating with some stochasticity, has clear-cut causes and effects. Ancient cities sprung up in places with the most arable land. Widening highways does not ease long-term traffic congestion. People will flock to areas with the best transport links. For all its incessant variability and intricacy, flux between macro states is and has been measurable and tractable.

The natural order is to establish equilibrium. The "before" states – before discovery of fertile land, widening highways, and building new railway stations – and the "after" states – after migration, roadworks, and infrastructure improvement – describe macro systems at equilibrium and free from perturbation (**Figure 1A**). Uncertainty comes with the dynamic flux between these states – pioneering settlers looking for new pastures, city planners responding to a traffic crisis, and land economists supporting urban developers. Often the perturbation is chanced without a guarantee of the exact nature of the "after" state equilibrium. Other times, one is guided by a map or atlas consolidating all previous knowledge before embarking on an informed choice.

In the human body, nature follows these same rules. Equilibria are resting states or steady states and response to any perturbation to find its way back to the norm is called homeostasis[1]. Biological perturbation can take many forms: from pathogenic infection to dietary choices. In most cases, the human body can deal with external factors and insults and the ensuing changes to the internal environment. For example, the immune system may activate to deal with a bacterial incursion, or might modulate specific digestive

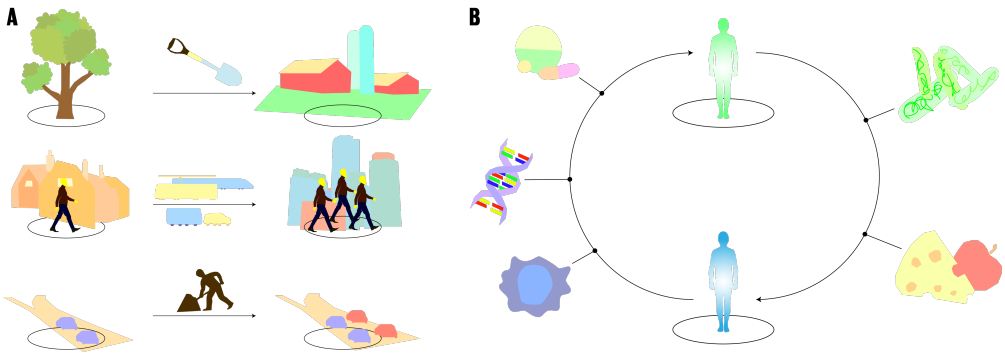


Figure 1. Perturbations between stable states. (A) Transitions between stable states in the macro world. (B) Transitions between and responses to stable states in the micro world.

functions in response to the nutritional content of food (**Figure 1B**). However, not all responses to perturbation are homeostatic. If the body fails to restore itself to the safe "before" state, then it may follow a path towards a dangerous "after" state[2–5]: a different steady state in which disease takes hold and where more external perturbations, such as antibiotic medication or EpiPen administration, might be necessary to reverse the process.

The body's ability to self-regulate homeostatic tasks is a function of the severity of external pressures and individual characteristics. A person's genetics determines the range and function of molecular machineries they can make, and therefore, how well equipped they are to deal with external threats on their own. In turn, the successful, appropriate, and faithful conversion of the DNA instructions into RNA blueprints and then protein machinery is required to respond to perturbations. Protein machineries have a variety of functions, ranging from dealing with the perturbation directly by neutralising foreign species or catalysing a metabolic reaction; to modulating the cycle anew by switching 'on' or 'off' other DNA instructions. Protein activities are finely tuned so that the response to a perturbation is appropriate and measured. Perhaps above all, the cellular metabolic profile is pertinent to perturbation due to metabolite concentrations responding to alterations to protein profiles, metabolites themselves acting as signalling agents[6–8], and the ability for metabolites and molecules to accumulate or be consumed as a result of the perturbation itself. Metabolite and molecular levels define the cellular environment, the range of biochemical reactions available to the cell, and the way that cells are presented

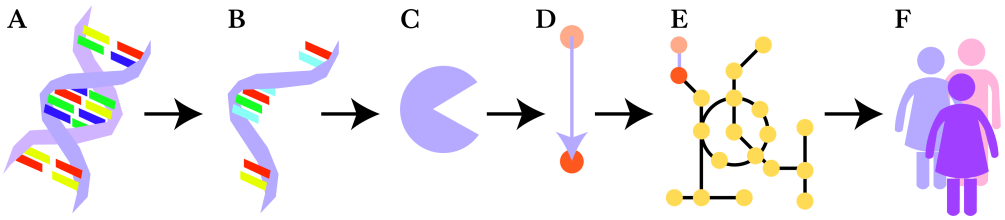


Figure 2. The central dogma of molecular biology. (A) DNA stores the genetic blueprints of all the proteins that the cell can make. (B) RNA is transcribed from DNA and are the instructions on how to make a certain protein that will actually be used by the cell at a given time. (C) Proteins are translated from RNA. (D) Proteins that are also enzymes catalyse biochemical reactions and change the levels of metabolites. (E) Metabolic changes lead to kinetic perturbations elsewhere in the metabolic network, which can alter the cellular state and lead to cell signalling. (F) The sum of expressed genetics contributes to an individual's phenotype.

to other cells[9,10] (**Figure 2**). It therefore stands to reason that those finely, delicately, sensitively balanced cellular states also have a profound effect on the outward phenotype and the individual's overall state of health and disease.

In fact, the consideration of metabolic activity is already considered in modern medicine[11–14]. Metabolic and molecular states influence the way that we define medical conditions[15–18]. Lack of molecular oxygen or oxidising equivalents puts cells into hypoxia, which favours anaerobic respiration over aerobic respiration for energy metabolism. Long-chain fatty acid synthesis and arachidonic acid production lead to a multifaceted inflammatory response involving recruitment of interleukins and leukotrienes. Invasion by viruses or bacteria is dealt with by the immune system. Diseases such as Alzheimer's and Parkinson's are defined based upon inappropriate accumulation of proteins that our cells can naturally make. None of these individual processes are inherently "good" or "bad" by their very nature, for all processes are a product of millennia of evolution, but it is the timing of, extent of, and interaction between these processes which determines whether they are desired and whether they lead to disease. Evolution forged the map of all processes, and it is the role of the regulatory machinery to ensure that the correct ones are engaged appropriately. Furthermore, these molecular definitions are useful for profiling cellular activity in health and disease. For example, a number of perfectly normal cellular processes are exploited in cancer, allowing tumour cells to establish a hypoxic microenvironment, hijack energy metabolism, evade the immune system, create its own blood supply, and enable limitless replication potential[19]. It is the inappropriate activation of "normal" processes that makes this after state dangerous. It is the identification of the

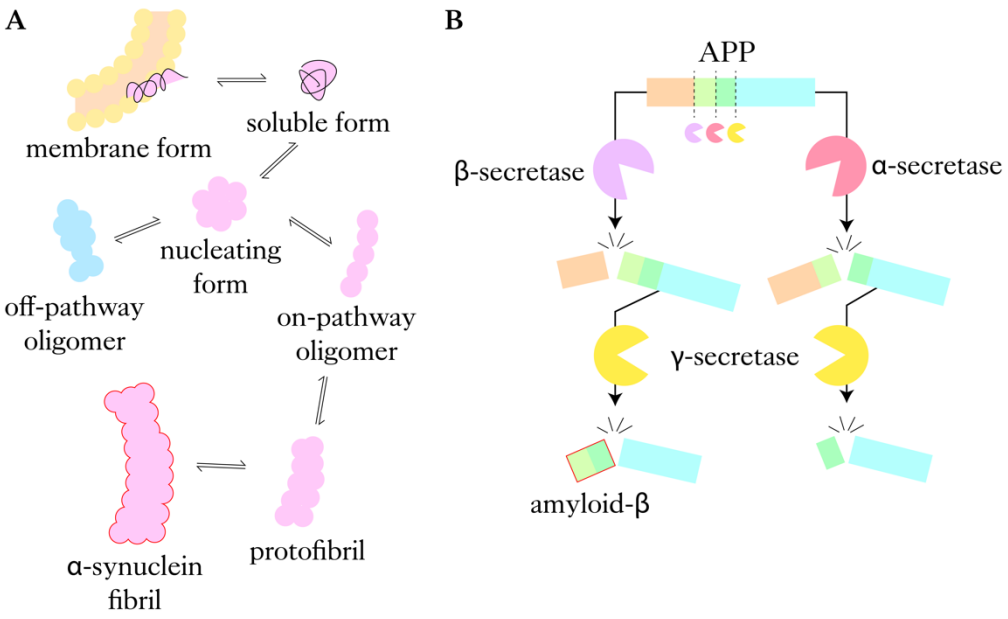


Figure 3. Neurotoxin formation in Parkinson's and Alzheimer's diseases. (A) In Parkinson's disease, neurotoxic α-synuclein form fibrils from aggregation of insoluble protein. (B) In Alzheimer's disease, neurotoxic amyloid-β is generated from amyloidogenic processing of the amyloid precursor protein (APP) via β-secretase and γ-secretase. The pathway involving α-secretase is non-amyloidogenic, but the function of the mature protein from this pathway is unclear.

set of "normal" processes leading to disease that is currently needed in order to develop, for example, precision medicine that targets those processes and regulates them when our own cells cannot.

Faults at any of the upper layers – DNA, RNA, protein – can result in an inappropriate metabolic environment and result in disease. Mistakes in DNA replication can occur due to ageing, which is intimately linked with metabolic rate[20]. Normal biochemical processes such as respiration jettison highly reactive free radicals, which can affect intracellular components and alter DNA instructions in cells[21]. Changes to the RNA landscape can come about due to changes to alternative splicing. This results in a different set of RNA blueprints from the same DNA instructions, and this has been shown to have prognostic implications in clear cell renal cell carcinoma[22] and hepatocellular carcinoma[23]. Altered protein profile is the basis of the aforementioned Alzheimer's (AD) and Parkinson's diseases (PD) as we understand them today. Improper processing of the amyloid precursor protein causes release of neurotoxic amyloid-β peptides in AD. Aggregation of

tau protein and α-synuclein results in tangles and inclusions of the protein in AD and PD, respectively, and these forms are also neurotoxic (**Figure 3**).

It is a medical triumph that so many medications exist for the treatment of disease and supplementation of the human body with the metabolic tools and molecular building blocks it needs to return to a healthy equilibrium[24–32]. However, it remains a medical challenge to design new medications for diseases for which patients present differing molecular responses to the disease perturbation and differing clinical responses to the medical perturbation. These heterogeneous, or complex diseases, are assessed by challenging the boundaries of diagnosed disease, stratifying patients into groups presenting similar responses, and identifying the biological markers of disease for each group. With the advent of tools aiding the measurement of DNA, RNA, protein, and metabolites, the once-hidden responses become visible, and the stratification step becomes attainable. By combining observations on these multiple "omics", one gains a highly comprehensive picture of the disease state of a group of patients with complex disease. The study of entire biological systems to achieve discoveries of medical importance is called systems medicine[33].

But it is neither desirable nor feasible to molecularly profile everyone. Although the cost of sequencing a human genome has fallen massively in recent years[34], practical reasons demand non-invasive collection of patient samples for diagnostic purposes.

The ultimate goal of systems medicine is to assemble huge databases of omics data, to combine those data and construct mathematically rigorous models of human disease, to execute or apply those models to uncover the basic biology underlying those diseases, to hypothesise, test, and validate basic biology insights at the bench or in the clinic, and then to offer personalised treatment solutions to patients presenting faults to their clinical or omics data[35] (**Figure 4**). These personalised solutions would come as dietary recommendations designed to perturb the biological state early on so that it does not follow a path towards disease, or as a curative medicine that aims to reverse the perturbation that has already taken hold in disease. Such investigations have only recently begun to become feasible due to pioneering efforts from those who have curated the first metabolic maps and those who have advanced such map-making tools. The combination

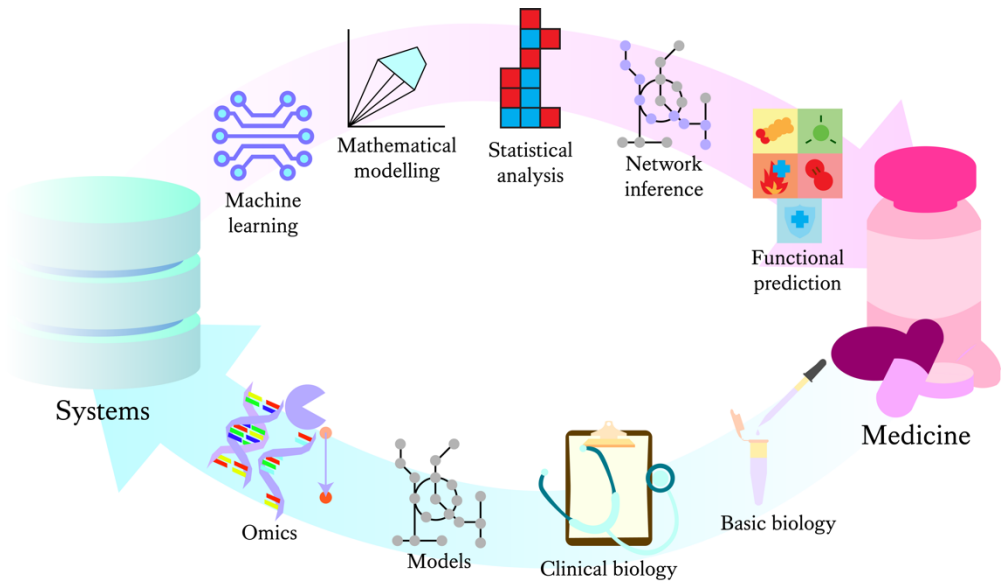


Figure 4. Overview of systems medicine. Big omics biological data are collected with clinical factors and knowledge from basic biology and biological models to form a systems-wide knowledge bank. These combined data can be analysed by mathematical, statistical, and network analysis and machine learning and AI to produce in silico *predictions on the cellular, tissue, and organismal state. Key analytes provide leads for drug discovery, and interventions are developed. The clinical and basic biological value of the intervention is assessed in the next round of data collection, leading to model refinement and discovery of the mechanism of action.*

of static maps with dynamic descriptions of biological states is what allows investigators today to navigate to health from disease.

This thesis presents and examines advancements contributing to each part of the systems medicine cycle. Here, I demonstrate how systems biology may harness big biological data to further medical discoveries. I use statistical and machine learning methods to identify molecular perturbations among AD and PD patients (**Paper I**) and construct a predictive model of neurological disease based on accessible blood and urine biochemistry (**Paper IV**). These endeavours highlight the power of systems biology methods for hypothesis generation and lead identification for further study, and suggested metabolic shifts resulting in a depletion of retinoid and androgen metabolites in a possible ageing-related mechanism.

Previous omics analysis and metabolic modelling have supported other leads, such as combined metabolic activator (CMA) therapy[36,37], comprising of metabolic supplements that are known to boost metabolic function. The present investigation examines the effect of CMA therapy on PD (**Paper II**) and AD (**Paper III**) patients, and showed that both

AD and PD patients exhibited cognitive benefits after CMA, compared to placebo. These studies show that the interface between systems biology and clinical science allows the rapid identification and validation of combination therapies, thus bypassing the expensive and time-consuming novel drug design stage.

The current investigation also shows the power of systems biology and omics analysis for its application to diverse cellular states and diverse data types. I demonstrate how multiple omics can be used to molecularly profile patients with non-alcoholic steatohepatitis (NASH) and congenital generalised lipodystrophy (CGL) (**Paper V**). Patients of CGL are susceptible to develop NASH[38,39], and this investigation showed that CGL patients with NASH have self-similar molecular profiles which is distinct from NASH patients from other causes, who present with heterogeneous molecular profiles. This study purported that CGL patients develop NASH via liver X receptor signalling, leading to inappropriate *de novo* lipogenesis and eventually fatty liver. As this is not the cause of all NASH, then, if validated, this study suggests separate precision medicine for NASH patients who also have a diagnosis of CGL.

Finally, the assessment of degeneration ends with an assessment of regeneration. The current investigation also used zebrafish model organism modelling to study the role of the muscle stem cell master regulator EZH2 in the regeneration of muscle after injury (**Paper VI**). Using transcriptomic, epigenomic, and lipidomic data analysis, this paper verifies previous studies suggesting the accumulation of lipids after EZH2 inactivation. This paper extends those findings by revealing long chain lipid elongation to favour fatty acid species such as arachidonic acid, which has a role in inflammation signalling and tissue healing.

Taken together, systems biology can greatly accelerate drug discovery lead identification via statistical, mathematical, and machine learning analysis that can assess tens or hundreds of thousands of biological analytes in a single big omics pipeline. This in turn accelerates and optimises the leads investigated at the bench or bedside and potentially allows for precision or personalised therapies. The metabolic nature of the proposed treatments leads to an overarching hypothesis of human disease being based on metabolic

imbalances which can also be restored on the metabolic level. This thesis explores the metabolic and molecular landscape for the analysis of degenerative diseases.

2. Systems biology

The study of entire biological systems is known as systems biology. It encompasses the treatment of biological entities as data and data objects and makes use of machine learning, mathematical modelling, and statistical testing. It can be either hypothesis-driven or hypothesis-free, flexibly allowing the use and re-use of biological data.

Systems biology today is built on a sound foundation of diverse biological data describing disease states and healthy states. Public and private efforts to map biochemistry and catalogue states of health and disease have culminated into the establishment of atlases and biobanks, huge repositories of data volunteered from healthy and diseased volunteers. Collections of software allow the rapid and accurate transformation of raw data into biologically relevant insights.

Collection of omics data is possible due to advancements in biotechnologies allowing the measurement of DNA, RNA, protein, and metabolites. Omics refers to big biological data which are collected via high-throughput methods (**Figure** 5). To illustrate individual omics, an individual's full complement of DNA, RNA, protein, and metabolites are known as their genome, transcriptome, proteome, and metabolome, respectively. In addition to the fine molecular definition of the individual's state, a description of their outward clinical presentation, or phenome (from phenotype), is also necessary, and this information comprises basic information such as age, sex, and background, but may also include blood and urine biochemical test results. The attachment of the phenome as metadata is necessary in order to contextualise the omics data, which is subject to the vicissitudes of health and disease and is often tied to the biological question.

Analysis of systems biology data involves treatment of omics data as computer data objects, which is a concept in bioinformatics. Mathematical, statistical, and probabilistic descriptions of biological systems enable inference on cellular functions, prediction of future flux, and outlooks for patient outcome. However, it is up to the investigator to ensure that models and methods are well selected for their purpose and that their resulting inferences are not overinterpreted[40–42]. *All models are wrong, but some are useful*[43]. For this reason, systems biology alone almost on purpose does not give a completely true picture

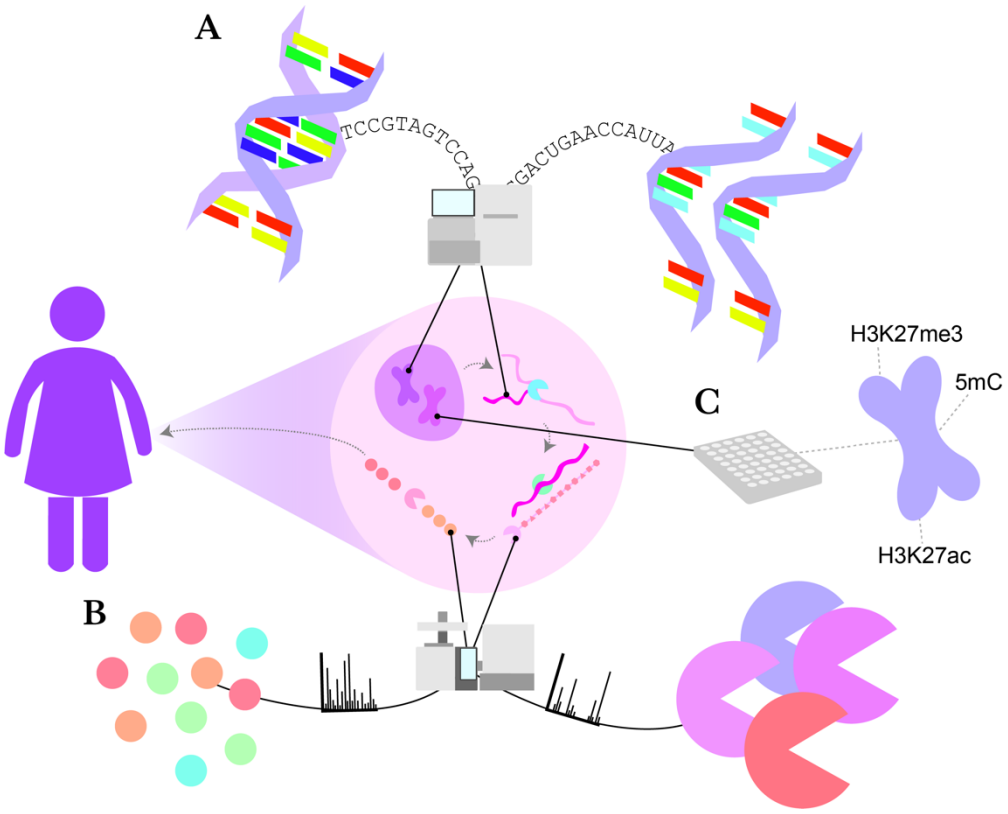


Figure 5. Omics data collection. (A) Nucleic acids can be sequenced to collect genomic and transcriptomic data. (B) Molecules such as proteins and metabolites can be analysed by mass spectrometry to collect proteomic and metabolomic data. (C) Other high-throughput assays can be used to capture other types of omics data, such as epigenomic data.

of a biological system due to model inaccuracy and the assumptions being made. Rather, systems biology leads the way for hypothesis generation and optimisation of leads to be validated at the bench. Assessment of entire omics allows investigation of full complements of species without bias, and these are excellent starting points for truly novel discovery.

Modelling and simulations can generate many more observations *in silico* than would be feasible *in vitro* or *in vivo*, but experimental validation remains extremely important to ensure that the observation is true biology. Systems biology therefore holds the promise for targeted allocation of bench time and resources.

2.1. Omics

Information in biology is encoded in sequences of chemically bonded molecular units. DNA and RNA are composed of nucleotides held together by a sugar-phosphate backbone. Proteins comprise of a string of amino acids joined by peptide bonds. There are four standard 'flavours' of nucleotide and twenty of amino acid (**Figure 6**).

The transfer of information from DNA to RNA is called transcription[44]. In the nucleus, pre-defined sections of double-stranded DNA, or genes, are transcribed into single-stranded RNA and released into the cytosol. Nucleotide sequences of DNA and the transcribed RNA are identical save for replacement of thymine by uracil, which can, for the purposes of this discussion, be considered the same 'flavour' of nucleotide. RNA molecules may also undergo splicing, whereby non-coding parts of transcripts (introns) are cut out of the message and discarded, leaving only the coding sequences (exons) in the mature transcript. There may be multiple ways to splice an RNA molecule. This alternative splicing means that multiple RNA molecules can result from a single DNA gene.

Information from mature RNA directly encodes a protein sequence. Each group of three nucleotides (a codon) in the RNA sequence corresponds to an amino acid in the protein sequence or a translation stop signal. This is known as the genetic code (**Figure 6A**). Translation from RNA to protein reliably produces a single protein product from a single RNA message, with few exceptions that are outside the scope of this discussion.

DNA, RNA, and protein can be considered as information storage molecules, and the rules governing their conversion are well established. Knowledge by the prior mapping of transcription and translation start and stop sides, intron/exon boundaries, and the genetic code allow prediction of nucleotide and amino acid sequences from one another. Consequences of DNA mutations on protein sequence can be deduced *in silico* since the mapping rules from DNA to protein are well established.

The full complement of genes in an individual is called the genome. The corresponding terms for transcripts and proteins are transcriptome and proteome, respectively. Increasingly, interactions in the interactome, metabolites in the metabolome, lipids in the

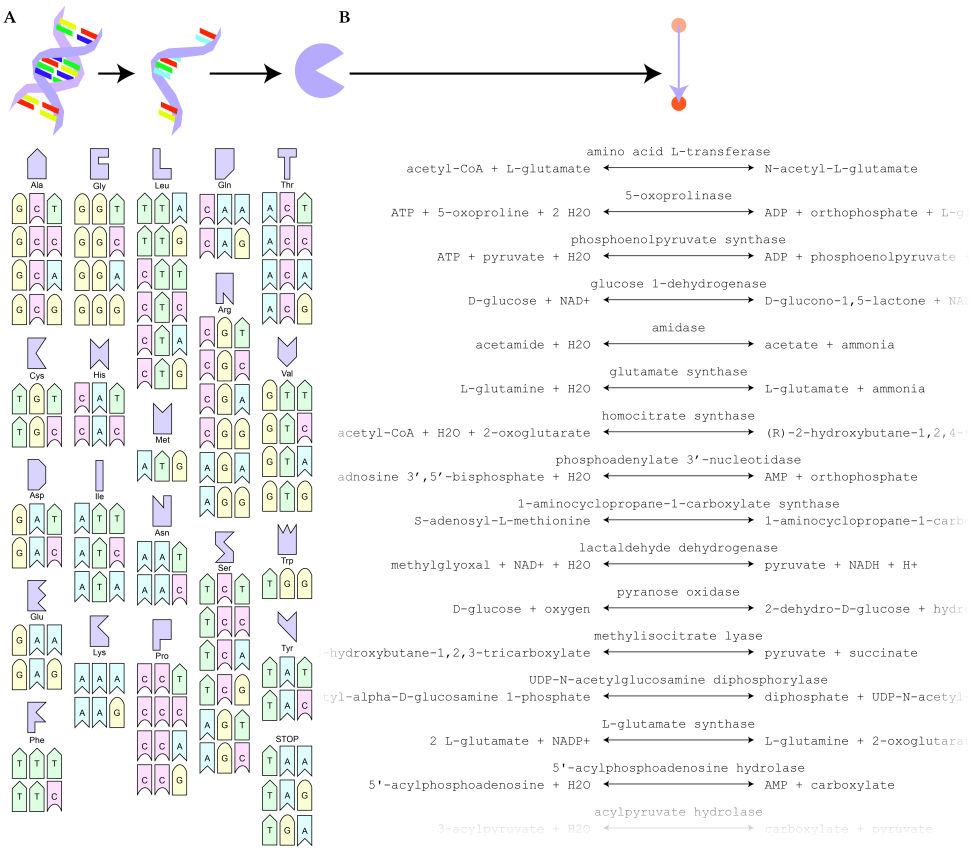


Figure 6. Static maps. (A) The genetic code connects genetic information with protein sequences. (B) Proteins catalyse biochemical reactions. Sample chemical equations were taken from KEGG Reaction.

lipidome, reaction rates in the fluxome, and phenotypes in the phenome have also become the subject of interest, as modulating interactions, metabolism, and lipid signalling has shown to have therapeutic value. Collectively known as omics, the collection of these data allows for a fuller picture of health and disease. Although certain interomic mappings are well established – for example, the central dogma of gene expression and the metabolic map of biochemical reactions – multiomic measurements are still highly important because it is a description of what *is* happening in cells in all their dynamicity; as opposed to a static map or theoretical predictions of what *can* happen.

However, this is not to say that static maps are without value. Static maps of metabolism embody our knowledge of the way metabolites are converted from one to another[45–48] (**Figure 6B**). They show all known possible paths and are supported by chemical stoichiometry and enzyme kinetics evidence. In fact, static maps are critical for

connecting omics in systems biology, from implementing the central dogma to connect the genome with the transcriptome and the proteome, to predicting the effect of changing enzyme production on metabolites. Entire sets of metabolic reactions are encapsulated as standardised tools known as genome-scale metabolic models[49] (GEMs). These describe the full set of reactions that are possible in a given organism, tissue, or cell. GEMs are carefully curated maps which can be used as a tool to estimate biochemical reactions because they embody gene-protein-reaction (GPR) rules. GPR rules connect transcriptomics and proteomics to enable predictions on fluxomics and metabolomics.

In fact, even the human reference genome can be considered a static map, as it contains all human genes and is shared across all individuals. The human reference genome is therefore a collection of all known genetic instructions, rather than a measured set of genes which are actually transcribed in a given environment – this is the transcriptome. Therefore, genomics studies into human health and disease rarely deal with full genomic sequences; rather, they focus on select genomic positions where variation can occur. These[50] include single nucleotide polymorphisms (SNPs), in which a single base in a DNA sequence differs from the reference genome; DNA methylation, in which nucleobases may be differentially methylated in different individuals; and histone modification, in which the scaffolding proteins around which DNA wraps and give it structure are themselves chemically altered. Genomic analysis of SNPs allows insights to population-level variations of SNPs with a direct consequence on the gene expression product, or co-variation between SNPs or between a SNP and a neighbouring gene or a gene in a topologically associated domain (i.e., close in 3-D space). This 'guilt by association' means that simple SNP measurement through the use of DNA probes can be enough to estimate a person's risk for a defective gene, and therefore disease[51].

The latter two genomic studies, measuring DNA methylation and chromatin histone modification, are referred to as the epigenome; and the former the methylome. These are parts of DNA which are not inherited, and can be dynamically modified. Proteins alter the epigenome by adding or removing chemical epigenetic marks on DNA and chromatin, and this can be in response to changes to the external environment. Classically, enrichment of DNA methylation in and around promoter regions, which is where transcription to RNA is initiated, indicates repression of gene expression at that

locus[52]. Histone modifications affect chromatin openness and therefore chromatin accessibility by RNA polymerase transcription machinery. Common modifications occur on histone N- and C-terminal tails, and can be methylations up to trimethylation, acetylation, or phosphorylation. Ubiquitylation, citrullination, and sumoylation are also possible but are less well studied. The consequences of histone modifications are summarised by the histone code hypothesis[53,54], where acetylation is generally associated with open chromatin and activation of genes at that locus; and trimethylation is associated with closed chromatin and repression of genes at that locus.

Omics analytes are variable, measurable properties of an individual which respond to fluxes between internal steady states and the external environment. Maps are established relationships which aid in the connection of omics layers and allow the prediction on hidden or hitherto unknown omics analyte levels. Models attempt to overlay measured omics data onto static maps in order to comprehensively define states in health and disease. Below, I summarise the ways in which omics were used and connected with static maps in the present investigation.

Paper I was an investigation into the ageing-related metabolic perturbations in AD and PD with the view of finding novel biomarkers as part of a paradigm shift away from the classical amyloid-β and α-synuclein molecules. In **Paper I**, I utilised publicly available transcriptomic and proteomic data and made inferences on reporter metabolites. This was made possible by using GEMs, which I reconstructed myself from a reference model for adipocytes. By overlaying expression data onto this GEM, I made predictions that retinoid and androgen metabolism pathways were dysregulated in subclasses of AD and PD. Transcriptomic data from a zebrafish model of ageing was then used to validate these findings, showing retinoid dysregulation on zebrafish GEMs. These investigations suggested that AD, PD, and ageing have effects on the metabolome and therefore may be metabolically treatable or preventable. These investigations further highlighted the similarities between AD and PD aetiology with reference to ageing as a contributor to disease.

Papers II and **III** assessed the effectiveness of CMA therapy in PD and AD, respectively. **Papers II** and **III** connected different parts of the phenome together using clinical trials as

a tool, and, using transcriptomics, GEMs, and experimental models for validation, showed that modulation of metabolome can lead to predicted changes in the phenome, namely, cognitive benefits in AD and PD patients. In these papers, I assessed clinical phenotypes of patients and generated hypotheses as to the clinical profile that would respond best to the treatment. These responses were justified by constructing a biological network integrating metabolomic and proteomic data, which revealed that metabolites associated with neurological and psychoactive drugs and proteins associated with the nervous system were central nodes in the network. These papers used data from multiple omics to suggest potential precision therapy.

Paper IV investigated clinical data for patients of neurological diseases to identify easily measurable biomarkers and their directional trends in disease. In **Paper IV**, UK Biobank SNP data and blood and urine biochemical data of AD, PD, motor neuron disease (MND), and myasthenia gravis (MG) participants were compared with those of healthy participants. Here, I examined genotype changes and trends which were associated with one or more neurological disease and identified heritable risk markers in genes which were also associated with cancer, which is a conclusion that has led to the well-established paradigm of degeneration—cancer antagonism. The biochemistry data were also used to construct a predictive machine learning model of neurological disease with an accuracy of 88.3%. Deconstruction of this model allowed inspection of the trends learned by the model, which were largely supported by the existing literature. This paper posits a machine learning method of rapid biomarker identification.

Paper V assessed gene expression of NASH patients with and without CGL through the analysis of methylomics and transcriptomics. Both omics suggested distinct molecular landscapes between NASH and CGL, suggesting that the NASH phenotype that CGL patients develop is distinct from heterogeneous NASH but is self-consistent across CGL patients. This combined omics analysis therefore supports a potential precision medicine solution for NASH patients who also have a diagnosis of CGL.

Paper VI studied zebrafish muscle regeneration following injury. **Paper VI** utilised zebrafish transcriptomic, epigenomic, and lipidomic data to investigate molecular trends before and after injury and in the presence and absence of the muscle master regulator

EZH2. Metabolic modelling with a zebrafish GEM revealed lipid accumulation, elongation, and flux through arachidonic acid. Epigenomic analysis confirmed chromatin accessibility at very long chain fatty acid elongation gene loci, while lipidomic analysis confirmed a trend towards longer chain lipid species. This paper guides the basic biology effort for the potential harnessing of natural metabolic circuits for muscle healing and regeneration.

2.2. Technologies

Systems biology owes itself to data. The power of systems biology analyses depends on the amount of data there is to be analysed. As a generality, systems biologists and bioinformaticians opine that the more data there are, the higher quality science can be done[55]. Therefore, collection, storage, and formatting of biological data on the scale of gigabytes or terabytes on disk is vital before the job of the systems biologist can even begin (**Figure 7**).

Nucleic acid data are obtained by next-generation sequencing[56]. Protocols vary, but the general principle is the same. Briefly, DNA is fragmented, ligated with barcode adapter sequences, and then amplified and sequenced in parallel. These sequencing reads are then assembled into contiguous sequences, or contigs, and then mapped onto a reference genome. For genotyping studies, variations from the reference template may be recorded as SNPs if the variation is found in a number of samples, although this is not the typical way to conduct genotyping studies. For transcriptome studies, complementary DNA (cDNA) is first generated from RNA via reverse transcription, and then reads are mapped to a transcriptome. Hits per gene are then recorded and then computed into raw expression counts or normalised expression values, such as fragments per kilobase of sequence per million reads (FPKM) or transcripts per million reads (TPM).

All **Papers I-VI** dealt with transcriptome data in some way. As the most direct way of measuring actual transcriptome-wide gene expression activity, RNA sequencing has become ubiquitous in systems biology. **Papers I** and **VI** required me to quantify zebrafish RNA abundances from raw sequencing reads. These raw reads were derived from whole mRNA and mapped to over 10,000 transcripts in the standard zebrafish transcriptome, easily allowing for hypothesis-free work.

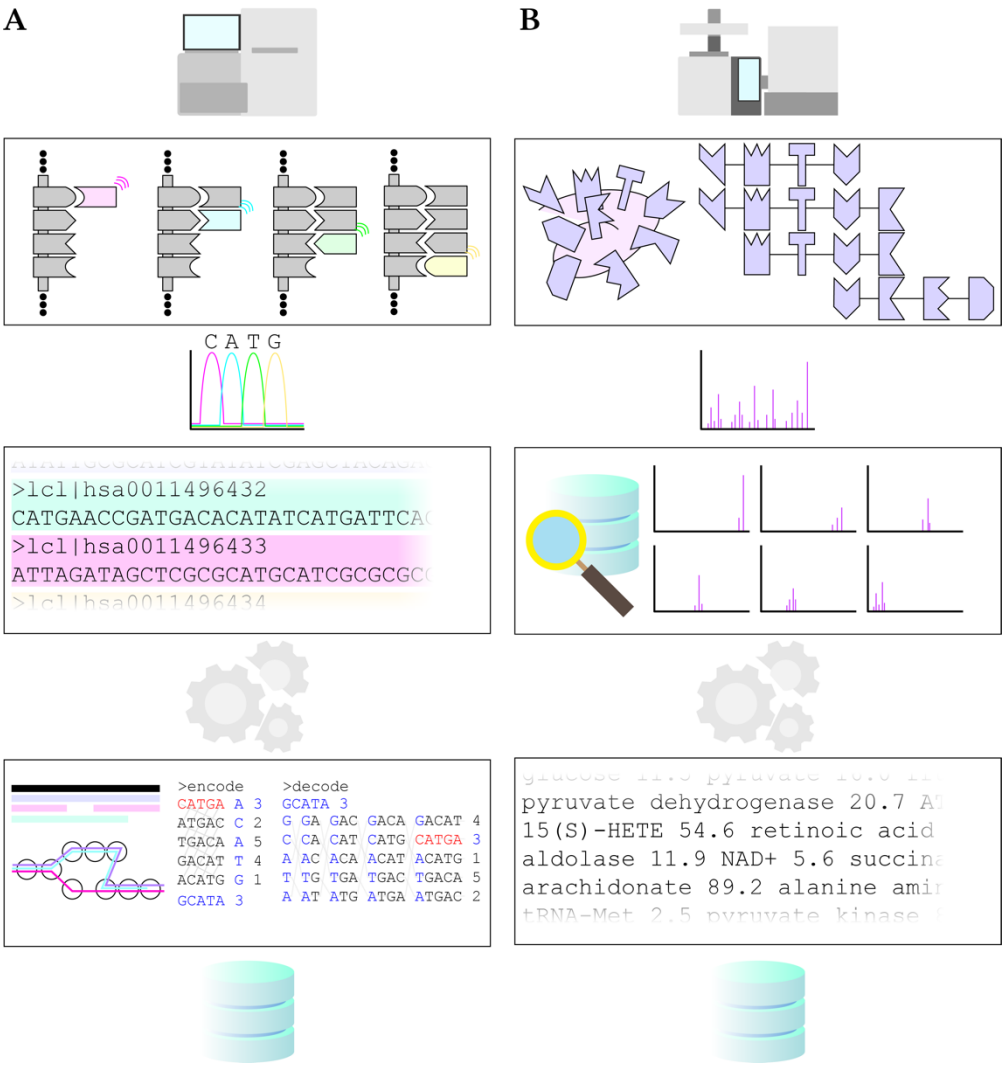


Figure 7. High-throughput data generation and processing. (A) Nucleic acid sequencing produces sequencing reads which are aligned to a reference sequence. Illumina sequencing and pseudoalignment by Kallisto are shown. (B) Mass spectrometry can sequence peptides and identify metabolites in a sample.

Genotyping studies using known SNPs does not rely on sequencing, but rather huge libraries of DNA probes which are sensitive to SNPs[57]. These are called SNP arrays, and can be used to detect hundreds of thousands of human SNPs across the human genome. As SNPs can co-vary with exonic mutations, the presence or absence of a SNP within close linkage with a gene can be a predictor for that gene's allele or mutations. Measurement of multiple SNPs therefore offers a probabilistic picture of an individual's genome without the need for whole genome sequencing. The full complement of SNPs for an individual is called a haplotype. Haplotypes of people from a variety of heritages

have been thoroughly studied and is the basis for associating disease risk with ethnic background[58]. Data from the Affymetrix UK Biobank Axiom array was used in **Paper IV**, in which I determined associations between over 800,000 SNPs with NLDs. Genotyping data allowed me to identify shared genetic risk loci among NLDs and between NLD and cancer. The use of a standard set of SNPs meant that associations with diseases in the Online Mendelian Inheritance in Man (OMIM) database were already annotated with the SNPs. The very large number of SNPs in the investigation allowed for a good amount of novel exploration and for the investigation to remain hypothesis-free.

Epigenome measures such as DNA methylation and chromatin accessibility can be used in conjunction with measurements on other omics to infer gene expression activity. To investigate DNA methylation, whole genome bisulphite sequencing works by chemically converting un-methylated cytosines, but not 5-methylcytosines, in DNA to uracil[59]. Upon sequencing, these methylation sites can be located at single-base precision. Differentially methylated sites and regions between disease states can inform differential transcription factor activity. For chromatin accessibility, ATAC-seq can be employed[60]. This technique uses a mutant Tn5 transposase to cleave DNA and insert sequencing adapters into the genome where the transposase is accessible – that is, in regions of open chromatin. The sequencing adapters allow next-generation sequencing of the intervening DNA in open chromatin, which can be mapped to a reference genome. **Paper V** dealt with scan data from the Infinium MethylationEPIC array, which detected over 850,000 methylation sites across the genome. When combined with gene expression counts by RNA sequencing, this approach allowed the unbiased discovery of altered lipid transport functions in the diseased patients compared to healthy control. In **Paper VI**, I re-analysed ATAC-seq peaks in zebrafish treated with an inhibitor for our gene of interest *ezh2* to determine changes to chromatin openness upon blockade, and compared these results with RNA sequencing data from *ezh2* mutants. The two datasets supported a hypothesis that lipid functions, including long chain lipid elongation, were activated upon *ezh2* loss by either mutation or inhibition.

Proteomic and metabolomic data can be obtained via mass spectrometry (MS)[61]. Analytes are ionised, accelerated to a constant kinetic energy, and detected in a mass spectrometer. In traditional MS, the magnetic field strength of the deflector at the time that ions hit the detector plate indicates the mass/charge ratio (m/z) of the ions, with lighter ions requiring a weaker magnetic field strength to be detected. In time-of-flight MS, the time it takes for ions to reach the detector indicates m/z, with lighter ions taking a shorter amount of time to reach the detector. On hitting the detector, the ion causes a current which is then amplified and recorded as a signal, with larger current indicating greater abundance. Mass spectra consist of the peaks representing the molecular ion, fragmentation products, and sometimes multimers if a light mode of ionisation was used. Identities of the analytes in the sample can be determined through the use of libraries of known mass spectra for small molecules. Amino acid sequences of proteins and peptides can be deduced by subtraction between peaks corresponding to fragments of the molecular ion arising from broken peptide bonds. Absolute abundances can be determined by spiking in known amounts of MS standards. Especially in metabolomics studies, high-performance liquid chromatography (HPLC) is often coupled with MS in a technique known as LC-MS. This has the advantage of physically separating analytes in a complex mixture by HPLC before identifying them by MS, allowing for high coverage of metabolites and high precision in identification. Each detected and identified analyte is recorded against its relative abundance. For proteins, higher abundance indicates increased gene expression, and for metabolites, higher abundance indicates increased biochemical activity. **Paper VI** also dealt with lipidomic data, a subset of metabolomics. From precision measurement of over a hundred lipid species, I was able to determine a trend favouring longer lipids in zebrafish with in an *ezh2* loss environment.

What unites these high-throughput technologies are two pillars: parallelisation and freedom from bias. Not having to run experiments for each analyte separately drastically speeds up data collection, reduces the risk of introducing operator-based variation, and allows the scaling-up of the number of analytes without linearly impacting time costs. This last point, enabling scaling-up, further means that analytes on the orders of hundreds of lipids, thousands of metabolites, tens of thousands of genes, and hundreds of thousands of genetic loci can be simultaneously measured. This untargeted, hypothesis-free

approach removes biases in the assumptions made at the experimental design stage, maximising the opportunity for discovery.

It is hypothesis-free data collection that produces the data that drives systems biology, which in turn has huge scope for making discoveries to be corroborated at the bench. Systems biologists are not confined to having a favourite gene.

2.3. Models

Huge efforts have been undertaken to collect and catalogue big biological data. However, data on its own has little value other than taking up space on a hard drive or server. In order to extract biological information from data, mathematical and statistical relationships must be harnessed, and the biases during data collection must be acknowledged.

Modern nucleic acid sequencing proceeds with the preparation of sequencing reads of typical length up to around 150 nucleotides as a raw output. In an RNA sequencing experiment, these sequencing reads must be mapped to a reference genome, which is 6.4 billion nucleotides for humans. The challenge comes in aligning the relatively short reads to a long reference, dealing with reads which span exon boundaries and therefore are not contiguous in the genome, and tolerating sequencing mismatches, gaps, and alternative splices. Bowtie[62] uses a technique from data compression called the Burrows-Wheeler transform on the reference genome to create a disk- and memory-efficient index onto which reads are mapped with high performance. However, only reads fully within one exon can be mapped by Bowtie – reads spanning multiple exons will be missed. TopHat[63] extends the functionality of Bowtie by predicting putative exons from contiguous mapped sequence reads and allowing the reads missed by Bowtie to overlap these predicted exons. In this way, novel splice sites can be discovered and differential splice site usage can be explored. Bowtie 2 further extends[64] mapping functionality by permitting gaps in the alignment of reads to the reference. Cufflinks[65] can take a TopHat alignment and use maximum likelihood methods to deduce alternative splice usage for each transcribed gene. In **Papers I** and **VI**, I opted to instead use Kallisto. Kallisto[66] creates indexes by constructing a transcriptome de Bruijn graph of k-mers corresponding to exonic sequences which could be mapped by a sequencing read. This pseudoalignment method

aligns *k*-mers in the input reads to probabilistically determine the transcripts where they could have come from rather than aligning lengths of sequence directly to the reference. The nodes in the de Bruijn graph are sensitive to transcripts resulting from alternative splicing and the pseudoalignment, as opposed to actual alignment, so Kallisto offers major computational time benefits.

Simultaneous identification and quantification of expressed genes in the transcriptome directly lends itself to differential expression analysis. In **Papers I**, **V**, and **VI**, I used the popular tool DESeq2[67,68], which models gene expression counts as following a negative binomial distribution and uses between-gene information sharing to shrink dispersions. The assumption of expression counts being distributed is fairly reasonable when dealing with animal data or data from a small sample size, but should be used with care when dealing with human population-scale data, in which the high variability in the data means that a negative binomial distribution is not appropriate[69]. This necessitates experimental or cross-validation. Nonetheless, the output of DESeq2 is a list of genes, often hundreds or thousands in length, which are predicted to be upregulated or downregulated between the disease states. Since genes can have redundant or co-operative functions, for example, multiple isoforms or genes along an operon, it is often useful to simplify the gene list according to their annotation. Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Online Mendelian Inheritance in Man (OMIM) are useful gene-level annotations, and functional analysis for the enrichment of annotations can be achieved through software packages such as Piano[70].

Gene expression quantification is the prelude to metabolic modelling. As mentioned previously, GEMs are computational tools that model global metabolism by observing GPR rules, which connect gene expression to protein amount and therefore biochemical reaction rate. It is with modelling with GEMs that metabolic and molecular inferences can be made *in silico* from gene expression data. Software such as Reconstruction, Analysis, and Visualisation of Metabolic Networks (RAVEN)[71,72] and Constraint-Based Reconstruction and Analysis (COBRA)[73–75] make predictions on reaction fluxes, metabolite concentrations, and steady states by treating the reactome as a linear programming problem. All biochemical reactions, uptake processes, and excretion processes encapsulated by the GEM are expressed in a stoichiometric matrix. Reactants,

products, limits, and directionalities are respected in this scheme, defining all the possible ways that metabolites can be converted or enter or leave the system, and are modelled as constraints on the system. Compartments are also respected, meaning that the metabolic and reaction profiles are distinct between subcellular compartments, and that metabolites need to be shuttled between compartments in order to participate in reactions that are dependent on subcellular location. In **Papers I** and **VI**, I used tools[76–78] within the RAVEN Toolbox 2 to predict changes to biochemical activity and steady state levels of metabolites given evidence for transcriptome-wide gene expression. In **Paper I**, this revealed three modes of perturbation within androgen biosynthesis and one within retinoid biosynthesis. Also by metabolic modelling, I further showed that the retinoid biosynthesis perturbation could be explained by ageing. **Paper VI** exploited a GEM for zebrafish, showing enhanced fatty acid elongation and increased metabolic flux towards arachidonic acid. These observations were consistent with observations from lipidomic analysis and gene set enrichment analysis.

Machine learning techniques can be used to partition data into clusters of similar profile or to make classifying predictions on unseen data. Partitioning data is a useful technique to quickly identify subgroupings within high-dimensional data for hypothesis generation on those individual subgroups, or to deconvolute heterogeneity in data purported to be from individuals of a single disease. Classification can be done by training a generalised linear model on data with the correct classifications, and the model can be tested using unseen data. Linear models can too be deconvoluted by inspecting the component weights assigned to each data variable in the model. **Paper I** made use of unsupervised clustering to generate putative AD and PD disease subclasses, which were then found to have fundamental different perturbations in androgen metabolism. **Paper IV** featured a multinomial generalised linear model which could distinguish between AD, PD, MND, and MG with an accuracy exceeding 80%. This model was trained on blood and urine biochemistry data from patients in the UK Biobank. The model was also inspected to find that the component weightings for each variable closely matched conclusions already reported in the literature. In both cases, machine learning accelerated the identification or verification of biomarkers and can efficient direct resources to the most promising leads.

In conclusion, modern bioinformatics allows detailed analysis on multiple omics levels. The connection from transcriptomic data to metabolic modelling is a powerful tool as it allows wide-ranging generation of biologically tractable hypotheses. Validation is conceptually simple and can be done by cross-validation of a second cohort or model organism, as was done in **Paper I**, or experimentally, as was done in **Paper VI**. Given the high-throughput nature of the input data and the general availability in repositories, metabolic modelling remains an extremely useful tool for rapid hypothesis generation.

2.4. Data

The demand for ever-faster and more powerful bioinformatics comes in no small part from a desire to match the high-throughput generation of data, culminating in an existing wealth of omics resources and repositories. Open data and standardised data management practices allow the researcher to quickly access and interoperate with omics data.

Public datasets include the Genotype-Tissue Expression (GTEx) project[79], Functional Annotation of the Mammalian Genome 5 (FANTOM5) project[80], The Cancer Genome Atlas, Gene Expression Omnibus (GEO)[81], ArrayExpress[82], UK Biobank[83], Human Protein Atlas[84], ProteomeXchange[85], and Expression Atlas[86], to name a few. They can comprise of curated or submitted data, targeted experiment or genome-wide data, single or multiple omics, and raw or processed formats. In addition to experimental datasets, repositories of static maps such as reference genomes and transcriptomes on Ensembl[87] and reference GEMs on Metabolic Atlas[88] are also crucial resources to enable alignment of experimental data, as discussed above.

In **Paper I**, I combined AD, PD, and control experimental data in GTEx, FANTOM5, GEO, and the bespoke brain diseases resource Religious Orders Study and Rush Memory and Aging Project[89] (ROSMAP). I also used protein-protein interaction data from the Human Reference Interactome[90] project to filter the data for protein-coding genes only. These data were selected due to their precision annotation to specific substructures in the brain, potentially allowing for region-specific investigation. To perform metabolic analysis, I obtained a GEM from Metabolic Atlas and modified it according to the brain-specific data. **Paper IV** aimed to benefit from adjacent genotyping and blood and urine biochemical data from NLD patients in the UK Biobank. I used

machine learning to construct a predictive model of NLD based on the biochemical data. Although the genotyping data were not able to improve the predictive model, observations arising from it were consistent with a view suggesting an ageing-related antagonism between neurodegeneration and cancer, which is a view that is gaining traction in the literature[91]. **Paper VI** used a wide range of single omics datasets deposited in GEO and the muscle-specific Muscle Gene Sets[92] database of muscle cell expression profiles. These multiple datasets were used to independently validate the lipid elongation trends that I observed in our in-house generated RNA sequencing data and in lipidomic data that was sent off for. **Papers II**, **III**, and **V** also made extensive use of in-house generated data: clinical and metabolomic data in the case of **Papers II** and **III** and RNA sequencing and DNA methylation data in the case of **Paper V**. Although public databases such as ROSMAP and UK Biobank are beginning to introduce multi-modal data, that is, multiple omics data originating from the same donor, it currently remains simpler to produce such data in-house.

3. Degenerative disease and metabolism

The metabolic basis for human disease is a widely studied concept and has been demonstrated to play a role in a range of non-communicable diseases[93–95], tissues[96–98], and cellular functions[99–103]. As metabolites can be easily and non-intrusively measured, for example, in blood and urine, there has been huge interest in this field. Furthermore, diverse metabolites are easily perturbed through the diet, paving the way to long-term preventive medicine through an individual's diet[104–106].

The scope for the current investigation was to determine whether there exists a metabolic basis for degenerative diseases, and if so, to characterise the metabolic and molecular biomarkers of disease (**Figure 8**). I generated metabolic models of degenerative diseases on the basis of high-throughput data and released them to the public.

3.1. Degeneration, cancer, and ageing

With ageing being a major risk factor for many human diseases[107,108], it is clear that ageing remains an important area of interest, particularly for countries with an ageing population. There is already growing evidence supporting antagonistic interplay between degeneration and cancer with regards ageing, that is, the incidence of many cancers peak at around age 75 years and then falls off with increasing age, whilst incidence of neurodegenerative disease (NDD) and cardiovascular disease expands at these ages[91,109]. The mechanistic details for the apparent inverse relationship between cancer and degenerative diseases are unclear.

In the current investigation, I exploit the ageing links between AD and PD in **Paper I**. The basis of this paper was that AD and PD are both age-onset diseases, and traditional investigations of the cytotoxic amyloid-β and α-synuclein have not resulted in significant progress in clinical trials and the amyloid hypothesis is attracting controversy[110]. I argued that this is due to the somewhat narrow-sense definitions of these diseases, which rely more on outward phenotype than molecular perturbations. Molecularly, AD and PD

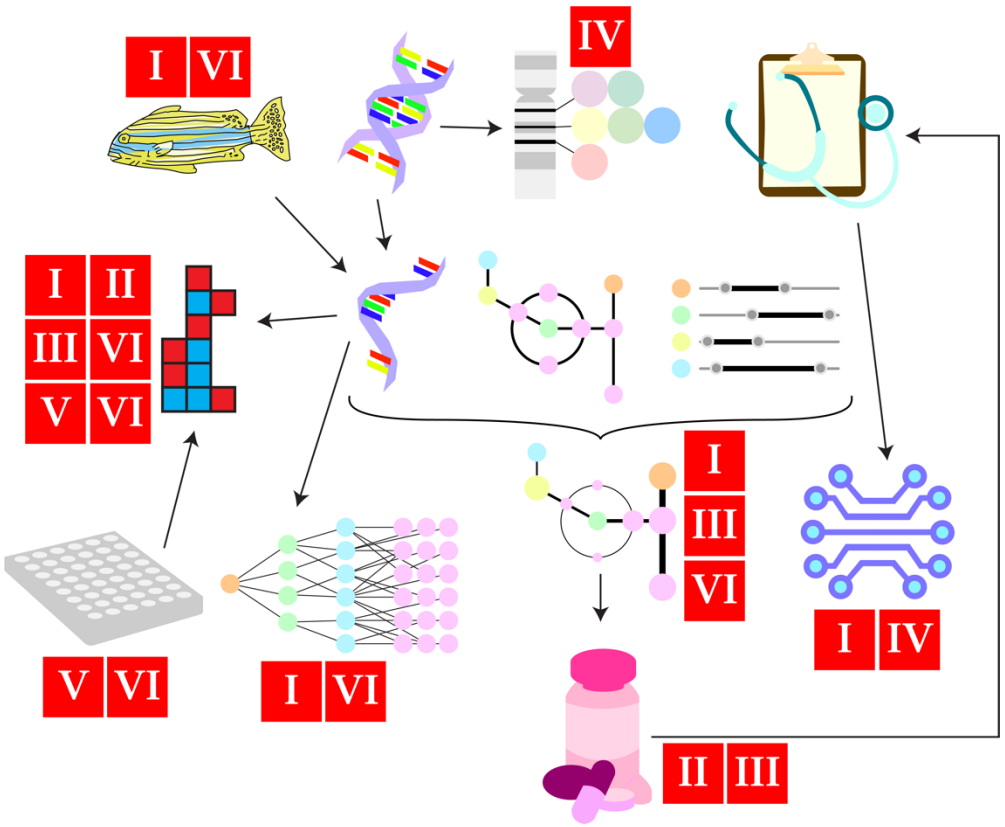


Figure 8. Scope of data and methods by paper in the current investigation. Zebrafish data were used in ***Papers I*** *and* ***VI****. Genotype data were used in* ***Paper IV****. All* ***Papers I-VI*** *used transcriptomic differential analysis. Mathematical metabolic modelling was used in* ***Papers I, III,*** *and* ***VI****. Other omics were used in* ***Paper V*** *and* ***Paper VI****, namely, methylomics and lipidomics, respectively. Gene coexpression networks were explored in* ***Papers I*** *and* ***VI****. Machine learning was harnessed in* ***Papers I*** *and* ***IV****. Papers* ***II*** *and* ***III*** *report the results of a clinical trial.*

patients present extremely heterogeneously, which is also likely the basis for their differing response to treatment. I therefore reasoned that AD and PD might correspond closely with ageing and therefore could be studied in tandem. Indeed, by combining AD and PD disease brain gene expression datasets, sourced from GTEx, FANTOM5, ROSMAP, I found that there exist three principal metabolic perturbation modes in androgen metabolism and one in retinoid metabolism. This was consistent with targeting the androgen[111–113] and retinoid[114–117] metabolism pathways in AD and PD treatment. However, the appearance of three distinct subgroups of disease in the current investigation would suggest that androgen replacement therapy would not be responded to identically by all patients, and therefore cannot be used as a blanket therapy. Indeed, androgen replacement therapy remains controversial in the field of AD and PD

treatment[118]. In the current investigation, I therefore suggest that the existence of multiple perturbation modes means that precision, targeted therapy is required to treat patients with different molecular presentations of disease.

In the zebrafish investigation in **Paper I**, I analysed in-house generated RNA sequencing data from homozygous and heterozygous *tert* mutants, as well as the wildtype. These mutants are used to model telomeric ageing[119] and were selected in the current investigation to explore the molecular perturbations in ageing and thereby validate the findings in the human investigation. Following transcriptomic analysis and metabolic modelling on a zebrafish GEM, we cross-validated our results from human data and concluded that the retinoid perturbation can be explained by ageing.

Inspired by the findings of the tandem AD/PD study, I expanded the scope of the current investigation to also include other NLDs. In **Paper IV**, I investigated UK Biobank data to identify trends in easily measurable blood and urine biochemistry in patients with AD, PD, MND, and MG. In the genome-wide association section of the investigation, I used genotyping data from an Affymetrix Axiom array to identify the genetic risk loci for the four NLDs. The results from this study concluded that many of the risk loci were shared between multiple diseases, thus suggesting a joint trajectory towards NLD and reinforcing the idea of tandem studies across NLDs. A number of the risk SNPs I identified were already annotated in the OMIM database as having associations with both brain cancers and non-brain cancers.

The results from these papers suggest a strong link between ageing, metabolic perturbation, and the degeneration/cancer antagonistic shift. The metabolic finding of **Paper I** suggests that dietary and/or lifestyle mitigations might be possibilities for the treatment and/or prevention of AD and PD. **Paper IV** confirmed the genetic similarities between NDDs and their association with cancers.

3.2. Metabolic neurodegeneration

As the impact of neurodegenerative and neurological diseases expands[120] with ageing populations worldwide, research focusing on NDDs have too been on the increase. Continuing with the theme of shifting the focus away from amyloid-β and α-synuclein,

an increasing amount of AD and PD research has shown links between neurodegeneration and metabolic disorders[121–129]. AD in particular has such strong links with comorbidities like type 2 diabetes that some commentators have dubbed it "diabetes of the brain" or "type 3 diabetes"[130]. Current frontline therapy only eases the symptoms of disease rather than reverse them[131], and emerging therapies such as aducanumab, which targets neurotoxic amyloid-β aggregates, has not avoided controversy surrounding interpretation of its efficacy[132]. The neurotoxic molecules in AD and PD have been known for decades, and the diseases themselves described a hundred years ago. Therefore, a fundamental paradigm shift is needed for the discovery of new targets in neurodegeneration.

The aim of the current investigation was to identify the metabolic changes in people with NDDs, and to see whether modulation of metabolism could be a viable means of treatment. As described above, **Papers I** and **IV** studied antagonism between degeneration and cancer with ageing. The three metabolic perturbation modes identified in **Paper I** demonstrated metabolic heterogeneity in AD and PD. The perturbations identified related to changes to increased cholesterol metabolism, decreased testosterone synthesis, and increased metabolism from oestrone to less potent forms of oestrone. This identification adds to the evidence suggesting a link between sex hormones and AD and PD and the role of hormone replacement therapy[112,113,118,133] , and the identification of multiple subgroups of disease could explain the heterogeneity in responses to these therapies[134].

A limitation of **Paper I** was that some of the specific metabolites that were predicted to be affected are not routinely detected in a clinical setting, requiring instead a technique such as MS to detect. Such techniques are not currently feasible for upscaling to population-level clinical screening, so I next shifted my attention to those metabolites which can be easily and non-intrusively detected. **Papers II-IV** aimed to analyse metabolism through blood and urine biochemistry. **Papers II** and **III** reported the effects of CMA therapy constituents nicotinamide adenine dinucleotide (NAD), L-carnitine tartrate, N-acetyl cysteine, and L-serine on cognition in AD and PD patients. **Paper IV** analysed the blood and urine biochemistry of populations with AD, PD, MND, and MG. I used a Monte Carlo method to select the clinical variables to be included as confounders

to construct a multinomial model which can predict NLD diagnosis based on clinical factors.

Consistent with the metabolic modelling results in **Paper I**, testosterone also featured as a confounder in the **Paper IV** multinomial model. When deconstructing the model weights, I found that AD and PD were among the less likely NLDs to be predicted by the model when presented with a sample with high testosterone. Furthermore, in the clinical trials in **Papers II** and **III**, it was found that patients with low alanine aminotransferase (ALT) exhibited greater improvements to cognition than patients with high ALT. Again, ALT featured in the **Paper IV** multinomial model, in which AD and PD were the only NDDs with a negative weighting, indicating that the model predicted deficits in ALT in these diseases. **Papers II** and **III** also demonstrated the strongest responses to CMA in patients with low aspartate aminotransferase (AST) and low γ-glutamyl transferase (GGT). ALT, AST, and GGT are enzymes found in the blood that can inform liver health. High levels of these enzymes can be indicative of liver diseases such as cirrhosis.

In the clinical trials described in **Papers II** and **III**, patients presented with a range of clinical conditions, as measured by the blood and urine biochemical tests. I was able to deconvolute the heterogeneity in clinical presentation by identifying the conditions of the patients who responded more strongly to CMA in terms of cognitive improvement, and these conditions included enzymes involved in liver metabolism. Furthermore, CMA is composed of metabolic activators designed to replenish cells with specific metabolites. That this treatment was able to lead to improved cognition in the patients who received it is further evidence that a metabolic solution to neurodegeneration is feasible.

It is, however, unlikely that CMA is the end-all-and-be-all for metabolic diseases. **Paper IV** demonstrated the ability to model NDDs by machine learning and discover accessible biomarkers by deconvolution of the model. I therefore posit that similar approaches are attractive approaches for the fast-tracking of biomarker leads and acceleration of drug discovery.

3.3. Reversal of degeneration

Secondary findings in **Paper I** was that the coenzymes NAD and nicotinamide adenine dinucleotide phosphate (NADP) and the eicosanoid arachidonic acid (ARA) were perturbed in AD, PD, and the zebrafish model of ageing. NAD and NADP are of great interest due to their involvement in carbon and energy metabolism in mitochondria, whereas ARA is involved with coordinating inflammation signalling. A natural next question in the current investigation was whether or not replenishing systems with the lost metabolites as a result of metabolic disease would represent a feasible therapeutic strategy against degenerative diseases.

The main finding of the clinical trials described in **Papers II** and **III** was that CMA therapy, which includes NAD, confers cognitive and metabolic benefits in PD and AD patients, respectively, over 84 days of oral administration. The CMA formulation was inspired by observed metabolic deficits in degenerative liver diseases such as non-alcoholic fatty liver disease (NAFLD) and NASH[135,136] and can be repositioned to AD and PD by virtue of similar metabolic perturbations in these diseases, such as in NAD and carnitine[122,137–141]. In contrast to current frontline treatments[131], the metabolic modulation approach to treating AD and PD aims to reverse the deleterious effects of neurodegeneration rather than simply relieve the symptoms. Indeed, in the clinical trials, we found enhanced improvements to the cognitive measures Alzheimer's Disease Assessment Scale – Cognitive (ADAS-COG) and Montreal Cognitive Assessment (MoCA) in patients receiving CMA compared to placebo. In addition to the primary endpoint, metabolomics also revealed increases in neuroprotective N-acetyl aspartate and sarcosine and decreases in neurotoxic quinolinic acid in AD patients after CMA; and increases in neuroprotective creatine and glycine, and decreases in neurotoxic S-adenosylhomocysteine, 2,3-dihydroxy-5-methylthio-4-pentenoate, and N-acetyltaurine in PD patients.

These effects were brought about by a relatively simple formulation of metabolic building blocks. The formulation of the therapy was based on observed metabolic deficits and administration of them was able to reverse those metabolic deficits and bring about wider neuroprotective changes to the metabolic landscape. Even more striking was the

independence from current AD and PD pathogenicity theory – that is, the amyloid cascade hypothesis and α-synuclein aggregation – and the non-reliance on current frontline therapy such as acetylcholinesterase inhibition and levodopa therapy. However, even after the clinical trial, it remains unclear whether cessation of CMA therapy would lead to relapse of AD and PD cognitive decline, and whether dietary supplementation could substitute the constituents of CMA.

Papers V and **VI** extended the investigation of a metabolic basis for degeneration to non-brain diseases. **Paper V** assessed the metabolic profiles of livers of patients with NASH and patients with CGL. CGL is characterised by lack of adipose tissue, lack of body fat, accumulation of fat in the liver, and hypertriglyceridaemia. It is an autosomal recessive disorder and can be caused by a number of genes, including *AGPAT2*, *BSCL2*, *CAV1*, and *PTRF*. Patients with CGL present with hepatosteatosis. In this work, I sought to characterise the molecular profiles of two lipodystrophies, *AGPAT2* and *BSCL2*, and compared these to a molecular profile of NASH. NASH is another liver disease with shared phenotypes with CGL, namely hepatosteatosis, but is not a congenital disease.

Through transcriptomic analysis of *AGPAT2* lipodystrophy and NASH patients, I found downregulations of the transcription factor FOXO1 and triglyceride synthesis enzyme GPAT3 in both NASH and CGL, but CGL-specific upregulation of the *de novo* lipogenesis enzymes SCD and ME1 and the lipid trafficking protein APOA4. From DNA methylation data, I saw very few differences in DNA methylation trends between *AGPAT2* and *BSCL2* lipodystrophy patients, but distinct differences between CGL and NASH. This suggests that disease progression in *AGPAT2* lipodystrophy patients might mirror that of *BSCL2* lipodystrophy patients, and if so, then the same therapies are likely to be equally effective on both. However, I also observed molecular differences between CGL and NASH, particularly in the aforementioned SCD and ME1 genes and liver X receptor activity. Patients with CGL presented with self-similar molecular profiles whereas NASH profiles were heterogeneous. This suggests that the fatty liver disease arising from CGL may represent a narrower spectrum of disease than those with NASH but without CGL. Therefore, this work suggests the potential for precision treatment of patients with NASH who already have a diagnosis of CGL. It is yet unknown whether

the fatty liver profile of CGL patients is represented by a subset of NASH patients without CGL.

Finally, **Paper VI** used a transgenic zebrafish model to comprehensively investigate regeneration of muscle following an injury event. This paper made use of a zebrafish stop-gain mutant in the *ezh2* gene, which encodes the polycomb repressor complex protein enhancer of zeste homologue 2, part of a histone methyltransferase complex which represses genes by closing chromatin. Following a muscle injury event, *ezh2* gene expression is reduced, leading to engagement of muscle developmental programs, proliferation of muscle stem cells, restoration of muscle fibres, and regeneration of tissue at the injury site. My hypothesis was that *ezh2* regulates the muscle regeneration by lipid signalling, as it has been shown by other groups that small molecule inhibition of EZH2 results in lipid accumulation. My approach to test this hypothesis was to analyse RNA sequencing data from wildtype and *ezh2* mutant zebrafish with and without muscle injury, use genome-scale metabolic modelling to predict metabolic alterations during injury, and identify probable gene targets that are controlled by *ezh2* that can result in those metabolic alterations.

From RNA sequencing analysis, I found upregulations in very long chain fatty acid elongation enzymes in the injured animals regardless of genotype. I also found evidence for cessation of core cellular processes, such as DNA replication and gene expression, and activation of protease activity, in the injured wildtype animals only. This indicated a worsening phenotype in the wildtype, possibly suggesting that these animals struggled to deal with the injury, whereas the *ezh2* mutant animals appeared to have a super-regenerative phenotype. These results were mirrored in metabolic analysis, where I identified a trend towards longer chain fatty acids and increased flux into ARA in the injured animals, this trend being more pronounced in the *ezh2* mutant.

Since ARA is a signalling molecule for inflammation, through its engagement of hydroxyeicosatetraenoic acid (HETE) and leukotrienes, and for reactive oxygen species (ROS) removal, through its engagement of glutathione, its accumulation in the *ezh2* mutant was consistent with prior observations of a super-regenerative phenotype. I next hypothesised that the *ezh2* mutation made it easier to engage ARA metabolism. In

exploration of a zebrafish interactome and gene co-expression analysis, I found a putative chain of gene regulation from *ezh2* to ARA metabolism genes within three regulatory steps, suggesting that *ezh2* does indeed regulate ARA. Therefore, presence of *ezh2*, as is the case in the wildtype, would result in a delay in the injury response to ARA regulation as *ezh2* needs to be removed first.

I therefore hypothesised that the *ezh2* mutant would have no such delay, but would rather be 'primed' and ready for the injury event, due to the lack of *ezh2* repressing ARA metabolism genes. If this were the case, I expected to find longer chain fatty acids in *ezh2* zebrafish muscle tissue even in the absence of any injury. To test this hypothesis, *ezh2* zebrafish muscle samples were sent for lipidomic analysis by MS, and indeed, I found evidence for a trend towards longer chain fatty acid species.

Paper VI demonstrated the role of ezh2 and fatty acid elongation for the regulation of the muscle regeneration response in zebrafish. This paper paves the way for the exploitation of this pathway for the engagement of endogenous processes to revert injured tissue and induce healing.

3.4. Present investigation

The present investigation covered well-known and less-well-studied disorders and degenerative diseases and conditions including those without an obvious link with metabolism. The breadth of the investigation is meant to demonstrate that the applicability of metabolic thinking to human disease is likewise broad. **Papers I**, **II**, and **III** sought to apply metabolic thinking to established fields of AD and PD, using classical omics techniques and metabolic modelling leading to the discovery of a role for retinoic acid as a possible therapeutic target, testosterone, oestrogen, and cholesterol as possible diagnostic biomarkers, and combined metabolic activators as possible dietary metabolic modulators. **Papers IV**, **V**, and **VI** took intellectual risks by applying metabolic thinking to identify predictive biomarkers in less-well-studied neurological disorders (NLDs) and in fields currently lacking a strong metabolic axis. The risk appeared to have paid off, leading to a potential for non-invasive screening of NLDs without the need of a specialist neurology unit, identification of metabolic similarities between the two outwardly different diseases lipodystrophy and NASH, and a role for the signalling metabolite

arachidonic acid in muscle repair. All of these papers demonstrate the value of metabolic thinking for diverse diseases and conditions.

4. Papers

This section contains **Papers I-VI** that constitute this thesis incorporating papers and that have been referred to as the current investigation. I am the copyright holder of all of the papers in this section and have permission from all co-authors to reproduce them in this section.

Due to image reproduction limitations of some of the visual data in these **Papers**, original display items, data items, and supplements have, wherever possible, also been made available for download at https://github.com/SimonLammmm/doctoral-thesis.

Paper I: Systems analysis reveals ageing-related perturbations in retinoids and sex hormones in Alzheimer's and Parkinson's diseases

The provided article below is adapted from the author's original version of the published paper available at:

Lam S, Hartmann N, Benfeitas R, Zhang C, Arif M, Turkez H, Uhlén M, Englert C, Knight R, Mardinoglu A. 2021, *Biomedicines*, **9**(10): 1310.

Systems analysis reveals ageing-related perturbations in retinoids and sex hormones in Alzheimer's and Parkinson's diseases

Simon Lam <simon.1.lam@kcl.ac.uk> ORCID: 0000-0002-4476-0971[1],

Nils Hartmann <nils.hartmann@unimedizin-mainz.de>[2,a],

Rui Benfeitas <rui.benfeitas@scilifelab.se> ORCID: 0000-0001-7972-0083[3],

Cheng Zhang <cheng.zhang@scilifelab.se> ORCID: 0000-0002-3721-8586[4],

Muhammad Arif <muhammad.arif@scilifelab.se> ORCID: 0000-0003-2261-0881[4],

Hasan Turkez <hasanturkez@gmail.com>[5],

Mathias Uhlén <mathias.uhlen@scilifelab.se> ORCID: 0000-0002-4858-8056[4],

Christoph Englert <Christoph.Englert@leibniz-fli.de>[2,6],

Robert Knight <robert.knight@kcl.ac.uk> ORCID: 0000-0001-9920-836X[1*],

and

Adil Mardinoglu <adilm@scilifelab.se> ORCID: 0000-0002-4254-6090[1,4*]

[1]Faculty of Dentistry, Oral and Craniofacial Sciences, King's College London, London, SE1 9RT, United Kingdom

[2]Leibniz Institute on Aging, Fritz Lipmann Institute, Jena, 07745, Germany

[3]National Bioinformatics Infrastructure Sweden (NBIS), Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, SE-17121, Sweden

[4]Science for Life Laboratory, KTH – Royal Institute of Technology, Stockholm, SE-17121, Sweden

[5]Department of Medical Biology, Faculty of Medicine, Atatürk University, Erzurum, 25240, Turkey

[6]Institute of Biochemistry and Biophysics, Freidrich-Schiller-University Jena, Jena, 07745, Germany

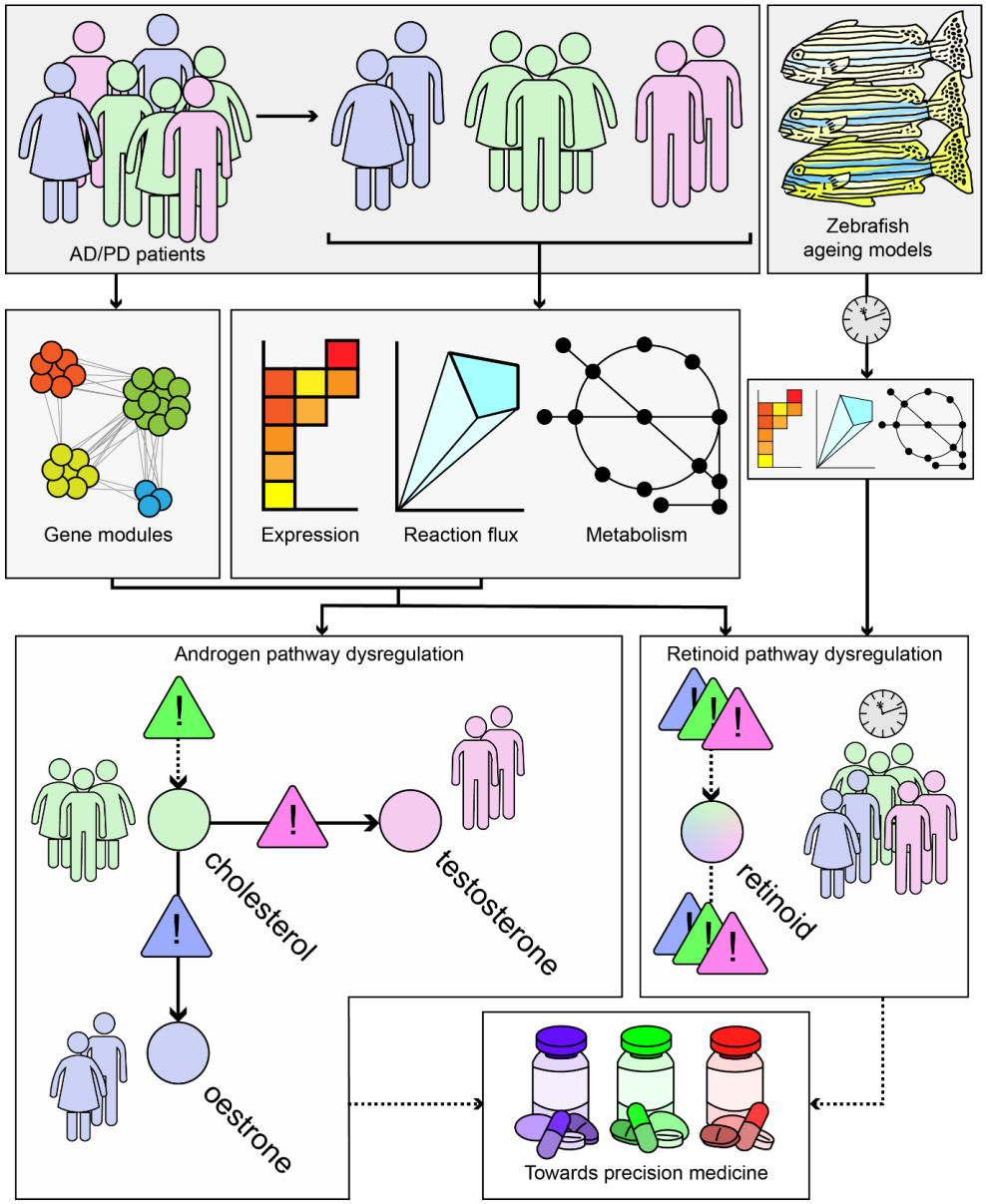
[a]Present address: Institute for Pathology, University Medical Center, Mainz, 55131, Germany

[*]Corresponding authors: adilm@scilifelab.se, robert.knight@kcl.ac.uk

Abstract

Neurodegenerative diseases (NDDs), including Alzheimer's (AD) and Parkinson's diseases (PD), are complex heterogeneous diseases with highly variable patient responses to treatment. Due to the growing evidence for ageing-related clinical and pathological commonalities between AD and PD, these diseases have recently been studied in tandem. In this study, we analyse transcriptomic data from AD and PD patients, and stratify these patients into three subclasses with distinct gene expression and metabolic profiles. Through integrating transcriptomic data with a genome-scale metabolic model and validating our findings by network exploration and co-analysis using a zebrafish ageing model, we identify retinoids as a key ageing-related feature in all subclasses of AD and PD. We also demonstrate that the dysregulation of androgen metabolism by three different independent mechanisms is a source of heterogeneity in AD and PD. Taken together, our work highlights the need for stratification of AD/PD patients and development of personalised and precision medicine approaches based on the detailed characterisation of these subclasses.

Graphical abstract



Keywords

Neurodegeneration, Alzheimer's, Parkinson's, ageing, systems biology

Abbreviations

AD, Alzheimer's disease; DEG, differentially expressed gene; FANTOM5, Functional Annotation of the Mammalian Genome 5; FBA, flux balance analysis; GEM, genome-scale metabolic model; GO, Gene Ontology; GSE, gene set enrichment; GTEx, Genotype-Tissue Expression; HuRI, Human Reference Protein Interactome; HPA, Human Protein Atlas; NDD, neurodegenerative disease; PD, Parkinson's disease; ROSMAP, Religious Orders Study and Rush Memory Aging Project

Introduction

Neurodegenerative diseases (NDDs), including Alzheimer's (AD) and Parkinson's diseases (PD), cause years of a healthy life to be lost. Much previous AD and PD research has focused on the causative neurotoxicity agents, namely amyloid β and α-synuclein, respectively. The current front-line therapies for AD and PD are cholinesterase inhibition and dopamine repletion, respectively, which are considered gold standards. Unfortunately, these therapies are not capable of reversing neurodegeneration (Liberini et al., 1996; Wijemanne and Jankovic, 2015), thus necessitating potentially lifelong dependence on the drug and risking drug-associated complications. Moreover, AD and PD are complex diseases with heterogeneous underlying molecular mechanisms involved in their progression (Greenland et al., 2019; Long and Holtzman, 2019). This variability can explain the differences in patient response to other treatments such as oestrogen replacement therapy (Baum, 2005; Meoni et al., 2020) and statin treatment (Shepardson et al., 2011; Jeong et al., 2019). Hence, we observed that there are distinct disease classes affecting specific cellular processes. Therefore, there is a need for the development of personalised treatment regimens.

In this study, we propose a holistic view of the mechanisms underlying the development of NDDs rather than focusing on amyloid β and α-synuclein (Lam et al., 2020). To date, complex diseases including liver disorders and certain cancers have been well studied through the use of metabolic modelling. This enabled the integration of multiple omics data for stratification of patients, discovery of diagnostic markers, identification of drug targets, and proposing of personalised or class-specific treatment strategies (Mardinoglu et al., 2018; Altay et al., 2019; Joshi et al., 2020; Lam et al., 2021). A similar approach may be applied for AD and PD since there is already a wealth of data from AD and PD patients from postmortem brain tissues and blood transcriptomics.

AD and PD share multiple clinical and pathological similarities, including comorbidities (Stampfer, 2006; De La Monte and Wands, 2008), inverse associations with cancer (Bajaj et al., 2010; Driver et al., 2012), and ageing as a risk factor (Hindle, 2010; Sengoku, 2020). One type of ageing is telomeric ageing, which is associated with the loss of telomeres, protein/nucleic acid structures that protect chromosome ends from degradation (Chakravarti et al., 2021). The enzyme telomerase is necessary for the maintenance of telomeres. In adults, telomerase activity is mostly limited to progenitor tissues such as in the ovaries, testes, and bone marrow. Loss of telomerase activity leads to telomere shortening, loss of sequences due to end-replication, and eventual degradation of sequences within coding regions, leading to telomeric ageing. Considering NDDs as a product of ageing, we can use an ageing model organism to study its effects on the brain. In our study, we used zebrafish (*Danio rerio*) as model organism since it has been used extensively used to study vertebrate ageing (Carneiro et al., 2016). For example, a zebrafish ageing model can harbour a nonsense mutation in the *tert* gene, which encodes the catalytic subunit of telomerase, and exhibit faster-than-normal ageing (Anchelin et al., 2013; Henriques et al., 2013).

In our study, we first analysed postmortem brain gene expression data and protein-protein interaction data from the Genotype-Tissue Expression (GTEx) database (GTEx Consortium, 2013), Functional Annotation of the Mammalian Genome 5 (FANTOM5) database (Forrest et al., 2014; Lizio et al., 2015, 2019; Marbach et al., 2016), Human Reference Protein Interactome (HuRI) database (Luck et al., 2019) and Human Protein Atlas (HPA) [http://www.proteinatlas.org, accessed 2021-03-09] (Uhlén et al., 2015) for characterization of normal brain tissue (**Figure 1A**). Secondly, we analysed transcriptomic data from the Religious Orders Study and Rush Memory Aging Project (ROSMAP) (Myers et al., 2007; Webster et al., 2009; Mostafavi et al., 2018) with published expression data from anterior cingulate cortices and dorsolateral prefrontal cortices of PD and Lewy body dementia patients, hereafter referred to as the Rajkumar dataset (Rajkumar et al., 2020), and from putamina, substantiae nigrae, and prefrontal cortices from patients with PD, hereafter referred to as the Zhang/Zheng dataset (Zhang et al., 2005; Zheng et al., 2010). On these data, we conducted differential gene expression and functional analysis, and then constructed biological networks to further explore coordinated patterns of gene expression. Next, we performed global metabolic analyses using genome-scale metabolic modelling. Alongside these analyses, we also leveraged zebrafish *tert* mutants to test the hypothesis that the identified changes may be associated with an ageing mechanism. Finally, based on our integrative systems analysis, we define three distinct disease subclasses within AD and PD and identified retinoids as a common feature of all three

subclasses and likely to be perturbed through ageing. We reveal subclass-specific perturbations at three separate processes in the androgen biosynthesis and metabolism pathway, namely oestradiol metabolism, cholesterol biosynthesis, and testosterone metabolism.

Results

Stratification of patients reveals three distinct disease classes

We retrieved gene expression and protein-protein interaction data from GTEx, FANTOM5, HuRI, HPA, and ROSMAP databases and integrated these data with the published datasets by Rajkumar and Zhang/Zheng. After performing quality control and normalisation (Materials and Methods), a total of 629 AD samples, 54 PD samples, and 889 control samples were included in the analysis (**Table 1**). To reveal transcriptomic differences between AD/PD samples compared to healthy controls, we identified differentially expressed genes (DEGs) and performed gene set enrichment (GSE) analyses. However, since AD and PD are complex diseases with no single cure, it is likely that multiple gene expression profiling exist, manifesting in numerous disease classes requiring distinct treatment strategies. We therefore used unsupervised clustering to elucidate these expression profiles and stratify the AD and PD patients based on the underlying molecular mechanisms involved in the disease occurrence.

Following unsupervised clustering with ConsensusClusterPlus (Wilkerson and Hayes, 2010), AD and PD samples separated into three clusters (**Figure 1B**, **Supplementary figure 1**). Clusters 1 and 2 contained samples from Zhang/Zheng and Rajkumar datasets, respectively, in addition to samples in the ROSMAP dataset. Cluster 3 contained only ROSMAP samples. Clusters did not form firmly along lines of sex, age, or brain tissues or brain subregion (**Supplementary figure 2**). Samples from non-diseased individuals were artificially added as a fourth, control cluster.

By differential expression analysis using DESeq2 (Love et al., 2014), we then characterised the distinct transcriptomic profiles within our disease clusters (**Figure 2A**). Cluster 1 showed mixed up- and downregulation of genes compared to control, whereas cluster 2 showed more downregulation and cluster 3 showed vast downregulation of genes compared to control.

To infer the functional differences between the subclasses, we performed GSE analysis using piano (Väremo et al., 2013) (**Figure 2B**, **Supplementary data 1**). Globally, DEGs in any cluster 1-3 were enriched in upregulated Gene Ontology (GO) terms for immune

response, olfaction, retinoid function, and apoptosis, but downregulated for copper ion transport and telomere organisation, compared to the control cluster. Considering individual clusters, cluster 1 DEGs were enriched in upregulated GO terms associated with immune signalling, cell signalling, and visual perception. We also found downregulation of GO terms associated with olfactory signalling and cytoskeleton. DEGs in cluster 2 were found to be enriched in downregulated GO terms associated with the cytoskeleton, organ development, cell differentiation, retinoid metabolism and response, DNA damage repair, inflammatory response, telomere maintenance, unfolded protein response, and acetylcholine biosynthesis and binding. On the other hand, we did not find any significantly enriched upregulated GO terms. In cluster 3, we found that DEGs were enriched in upregulated GO terms associated with neuron function, olfaction, cell motility, and immune system. DEGs in cluster 3 were found to be enriched in downregulated GO terms associated with DNA damage response, ageing, and retinoid metabolism and response.

The difference in expression profiles illustrate highly heterogeneous transcriptomics in AD and PD and that there are notable commonalities and differences between the subclasses of AD or PD samples. Interestingly, we found retinoid metabolism or function to be a common altered GO term in all subclasses. This was upregulated in cluster 1 but downregulated in clusters 2 and 3. We therefore observed that retinoid dysregulation appears to be a common ageing-related hallmark of NDD.

Metabolic analysis reveals retinoids and sex hormones as significantly dysregulated in AD and PD

Based on clustering and GSE analysis, we identified distinct expression profiles but these alone could not offer insights into metabolic activities of brain in AD and PD. To determine metabolic changes in the clusters compared to controls, we performed constraint-based genome-scale metabolic modelling. We reconstructed a brain-specific genome-scale metabolic model (GEM) based on the well-studied HMR2.0 (Mardinoglu et al., 2013) reference GEM by overlaying transcriptomic data from each cluster and applying brain-specific constraints as described previously (Baloni et al., 2020) using the tINIT algorithm (Agren et al., 2012, 2014) within the RAVEN Toolbox 2.0 (Wang et al., 2018). We generated a brain-specific GEM (*iBrain2845*) (**Supplementary file 1**) and used it as the reference GEM for reconstruction of cluster-specific GEMs in turn. We constructed the resulting context-specific *iADPD* series

GEMs *iADPD1*, *iADPD2*, *iADPD3*, and *iADPDControl*, corresponding to cluster 1, cluster 2, cluster 3, and the control cluster, respectively (**Supplementary file 2**).

We conducted flux balance analysis (FBA) by defining maximisation of ATP synthesis as the objective function. *iADPD1* and *iADPD2* both showed upregulation of fluxes in reactions involved in cholesterol biosynthesis and downregulation in O-glycan metabolism, with reaction flux changes being more pronounced in *iADPD2* than in *iADPD1* (**Table 2**, **Supplementary data 2**). We found that the fluxes in *iADPD1* were uniquely upregulated in oestrogen metabolism and the Kandustch-Russell pathway. *iADPD2* was uniquely upregulated in cholesterol metabolism, whereas *iADPD3* uniquely displayed roughly equal parts upregulation and downregulation in several pathways, including aminoacyl-tRNA biosynthesis, androgen metabolism, arginine and proline metabolism, cholesterol biosynthesis, galactose metabolism, glycine, serine, and threonine metabolism, and N-glycan metabolism.

In particular, we observed increased positive fluxes through reactions HMR_2055 and HMR_2059 in *iADPD1*, which convert oestrone to 2-hydroxyoestrone and then to 2-methoxyoestrone (**Figure 3**). In *iADPDControl*, these reactions carried zero flux. In *iADPD2*, we observed increased positive fluxes through HMR_1457 and HMR_1533, which produce geranyl pyrophosphate and lathosterol, respectively. Both of these molecules are precursors to cholesterol, and while we did not see a proportionate increase in the production of other molecules along the pathway (namely, farnesyl pyrophosphate and squalene), we did observe a general increase in fluxes through the androgen biosynthesis and metabolism pathway. Finally, we observed that *iADPD3* displayed a decreased production of testosterone from 4-androstene-3,17-dione via HMR_1974 despite an increase in production of 4-androstene-3,17-dione via HMR_1971.

Taken together, the obtained results indicate the existence of three distinct metabolic dysregulation profiles in AD and PD, with dysregulation being most pronounced in cluster 2 patients and least pronounced in cluster 3 patients. Furthermore, we found that all three feature dysregulations in or associated with sex hormone biosynthesis and metabolism, which might explain the heterogeneity in responses to sex hormone replacement therapy in AD and PD patients as extensively reported previously (Baum, 2005; Wahjoepramono et al., 2016; Resnick et al., 2017; Rajsombath et al., 2019). We also confirmed that dysregulations through sex hormone pathways in the *iADPD* series GEMs were not due to differences in relative frequencies between sexes in the main clusters 1-3 (Fisher's exact test, $p = 0.4700$).

In addition to metabolic inference and FBA, we performed reporter metabolite analysis (Patil and Nielsen, 2005) by overlaying DEG analysis results onto the reference GEM to identify hotspots of metabolism (**Table 3**, **Supplementary data 3**). In short, we uniquely identified oestrone as a reporter metabolite in cluster 1, and lipids such as acylglycerol and dolichol in cluster 2. No notable reporter metabolites were identified as significantly changed in cluster 3 only. In common to all clusters 1-3, retinoids and sex hormones such as androsterone and pregnanediol were identified as significantly changed reporter metabolites, which are generally in line with GSE and FBA results.

Network analysis supports retinoid and androgen dysregulation and suggests transcriptomic similarity between AD and PD

To further explore the gene expression patterns shown across AD and PD patients, we took expression data and constructed a weighted gene co-expression network for both groups (Spearman $\rho > 0.9$, FDR $< 10^{-9}$, Materials and Methods). Each network was compared against equivalent randomly-generated networks as null models. After quality control, the AD network contained 4861 nodes (genes) and ~397,000 edges (significant correlations), and the PD network contained 5857 nodes and ~394,000 edges (**Figure 4A, Figure 4B**, **Table 4**). A community analysis to identify modules of highly co-expressed genes (Traag et al., 2019) highlighted nine and fifteen communities with significant functional enrichment in AD and PD respectively.

In the AD network, gene module C3 was enriched for genes involved with neuron and synapse development, similar to patient cluster 3, C4 for genes involved with mRNA splicing, similar to patient cluster 2, and C5 for genes involved with the mitochondrial electron transport chain (**Figure 4C**, **Supplementary data 4**). C1 and C2 were the gene modules with the largest number of genes. C1 was enriched for gene expression quality control genes and development and morphogenesis genes, mirroring patient cluster 2, whereas C2 contained cytoskeleton-related genes, similar to patient cluster 1.

In the PD network, C1 was enriched for genes involved with retinoid metabolism, glucuronidation, and cytokine signalling. Since androgens are major targets of glucuronidation (Grosse et al., 2013), these results are in line with our main findings. Further, C2 contained DNA damage response and gene regulation genes, similar to patient cluster 2, C3 contained nuclear protein regulation genes, and C4 contained mRNA splicing genes, again similar to patient cluster 2.

Further, the two networks share a large number of enriched terms in common, and there is high similarity between the major gene modules, highlighting the similarity between AD and PD. In addition to this, enrichment analysis for KEGG terms alone was unable to assign "Alzheimer disease" and "Parkinson disease" to the correct gene modules from the respective networks, indicating functional similarity between the two diseases, and additional neurological disease terms such as "Huntington disease" and "Amyotrophic lateral sclerosis" were also identified by the analysis, further suggesting the transcriptomic similarity between neurological diseases. We found that AD C1 and PD C2 were frequently annotated with these disease terms, and these gene modules are also highly similar. Therefore, this gene module could constitute a core set of dysregulated genes in neurodegeneration.

Taken together, the network analysis supports our GSE findings. The functional consequences of differential expression in the patient clusters could be explained by differential modulation of gene modules identified in our network analysis together with dysregulation of a core set of genes implicated in both AD and PD.

Zebrafish transcriptomic and metabolic investigations suggest an association between brain ageing and retinoid dysregulation

To further validate our findings regarding the differences between clusters of human AD and PD samples, we analysed transcriptomic data from *tert* mutant zebrafish and reconstructed tissue-specific GEMs (**Figure 5A**). To ascertain that these effects of ageing were limited to the brain, we analysed the brain, liver, muscle, and skin of zebrafish as well as the whole animal.

We first repeated DEG and GSE analyses in the *tert* mutants using brain transcriptomic data. We found significant enrichment of GO terms associated with retinoid metabolism as well as eye development and light sensing, in which retinoids act as signalling molecules (Blomhoff and Blomhoff, 2006) (**Figure 5B**, **Supplementary figure 3**, **Supplementary data 5**). To further support our findings, we then reconstructed mutant- and genotype-specific GEMs by overlaying zebrafish *tert* mutant transcriptomic data onto a modified generic *ZebraGEM2* GEM (Van Steijn et al., 2019). We designated the modified GEM *ZebraGEM2.1* (**Supplementary file 3**) and used it as the reference GEM. We also generated zebrafish organ-specific GEMs and provide them to the interested reader (**Supplementary file 4**).

We then repeated reporter metabolite analysis using the transcriptomic data from zebrafish tissue-specific GEMs and found that retinoids were identified as significant reporter metabolites in *tert*$^{+/-}$ zebrafish (p = 0.045) but not in *tert*$^{-/-}$, where evidence was marginal (p = 0.084) (**Figure 5C**, **Table 5**, **Supplementary data 6**). We also observed this result in the skin of *tert*$^{-/-}$ mutants, where evidence was significant (p = 0.017). This result can be explained due to the susceptibility of skin as an organ to photoageing, for which topical application of retinol is a widely-used treatment (Riahi et al., 2016). However, we did not find evidence for significant changes in pregnanediol, and androsterone was significant only in the skin of *tert*$^{-/-}$ zebrafish (p = 0.017). This would suggest that either change in sex hormones are not ageing-related with regards AD and PD, or the changes were outside the scope of the zebrafish model that we used.

Taken together, these results indicated that ageing can largely explain alterations in retinoid metabolism in the brain but not alterations in sex hormone metabolism. These results also suggested that ageing has a differential effect on different organs, implying that metabolic changes due to ageing in the brain are associated with neurological disorders.

Discussion

In this work, we integrated gene expression data across diverse sources into context-specific GEMs and sought to identify and characterise disease subclasses of AD and PD. We used unsupervised clustering to identify AD/PD subclasses and employed DEG and GSE analysis to functionally characterise them. We used network exploration, constraint-based metabolic modelling, and reporter metabolite analysis to characterise flux and metabolic perturbations within basal metabolic functions and pathways. We then leveraged expression data from zebrafish ageing mutants to validate our findings that these perturbations might be explained by ageing. Our analysis concluded with the identification and characterisation of three AD/PD subclasses, each with distinct functional characteristics and metabolic profiles. All three subclasses showed depletion of retinoids by an ageing-related mechanism as a common characteristic.

We believe that a combined analysis that integrates AD and PD data is necessary to elucidate common attributes between the two diseases. However, we realised that such an analysis will likely obscure AD- and PD-specific factors, such as amyloid β and α-synuclein, but should aid the discovery of any factors in common. Since AD and PD share numerous risk factors and comorbidities such as old age, diabetes, and cancer risk, we believe that an AD/PD

combined analysis can identify factors in common to both diseases and prove valuable for the identification of treatment strategies which might be effective in the treatment of both diseases.

GSE analysis highlighted significant changes related to retinoid function or visual system function, in which retinol and retinal act as signalling molecules (Blomhoff and Blomhoff, 2006), in all clusters (**Figure 2**, **Supplementary data 1**). Together with the identification of multiple retinol derivatives as significant reporter metabolites in *iBrain2845* (**Table 3**, **Supplementary data 3**), we hypothesised that retinoids are a commonly dysregulated class of molecules in both AD and PD, and that this may be due to an ageing mechanism. Indeed, in our investigation with zebrafish telomerase mutants, we again found alterations in retinoid and visual system function in GSE analysis (**Figure 5B**, **Supplementary figure 3**, **Supplementary data 5**) and reporter metabolite analysis (**Figure 5C**, **Table 5**, **Supplementary data 6**).

Retinoids were identified as a reporter metabolite in all three clusters of patients in this study, and we believe that retinoid therapy is a potentially viable treatment for both AD and PD patients. Further, our zebrafish analysis highlighted the importance of retinoids in ageing of the brain and the skin (**Figure 5C**, **Table 5**, **Supplementary data 6**). Retinol, its derivatives, and its analogues are already used as topical anti-ageing therapies for aged skin (Riahi et al., 2016), and there is a growing body of evidence suggesting its efficacy for the treatment of AD (Shudo et al., 2009; Fukasawa et al., 2012; Das et al., 2019; Fitz et al., 2019). We add to the body of evidence with this *in silico* investigation involving zebrafish telomerase mutants, suggesting that the source of retinoid depletion in AD and PD is ageing-related. Interestingly, regarding our finding for skin ageing in zebrafish, lipid biomarkers have been proposed in a recent skin sebum metabolomics study in PD patients (Sinclair et al., 2021). This could be interpreted as co-ageing in brain and skin tissues, possibly allowing for cheap, non-invasive prognostic testing for PD.

In addition to retinoids, we found evidence for subclass-specific dysregulation within the androgen metabolism pathway in each of the three clusters in FBA (**Table 2**, **Supplementary data 2**) and reporter metabolite analysis (**Table 3**, **Supplementary data 3**). We found that *iADPD1* displayed increased oestrone conversion to the less potent (Martucci, 1983) 2-methoxyoestrone, *iADPD2* displayed increased production of the cholesterol precursor molecules geranyl pyrophosphate and lathosterol and increased androgen biosynthesis, and *iADPD3* displayed decreased conversion of 4-androstene-3,17-dione to testosterone. However, there was no definitive evidence to suggest an ageing-related basis for these observations based

on our zebrafish study, but this may be due to the diverse functional roles that sex hormones have, limitations within the *ZebraGEM2.1* model, or absence of an actual biological link between sex hormones and ageing of the brain. Despite this, given the widely reported variability in responses to sex hormone replacement therapy in AD and PD (Baum, 2005; Shepardson et al., 2011; Wahjoepramono et al., 2016; Resnick et al., 2017), we believe that this observation represents a possible explanation for the heterogeneity. Our observation regarding the dysregulation of the androgen pathway at three separate points suggests that dysregulation at other points might also be linked to AD and PD, thus implying that androgen metabolism dysregulation in general might be important for the development of AD and PD. Our finding via network community analysis of a gene module associated with glucuronidation activity points to a possible therapeutic strategy to combat androgen dysregulation. More work is needed to elucidate the importance of sex hormones and glucuronidation regarding AD and PD.

Identification of subclasses is desirable to address the heterogeneity in disease with regards transcriptomic profile and treatment response, but patients must be stratified in order to be diagnosed with the correct disease subclass and therefore administer the appropriate treatment. To this end, we used GSE analysis to functionally characterise the AD/PD subclasses (**Figure 2**, **Supplementary data 1**). Cluster 2, which was associated with a decreased immune and stress response, appeared to be most severe disease subclass, whereas cluster 3, which was associated with an increased sensory perception of smell, reduced haemostasis, and reduced immune and DNA damage response, seemed to be the least severe. Meanwhile, cluster 1 was associated with an increased immune and inflammatory responses and reduced sensory perception of smell. The functional terms are supported by community analysis of our AD and PD gene co-expression networks, which identified gene modules that roughly align with the GSE results (**Figure 4**, **Supplementary data 4**). The proposed severity ratings are supported by FBA findings, which show *iADPD2* as having the highest total flux dysregulation compared to control, and *iADPD3* as having the least (**Table 2**, **Supplementary data 2**). Although we did not attempt to characterise for stratifying and diagnosing patients in our study, our findings clearly showed that such stratification is possible. Given the differing nature of the proposed therapeutic strategies that we outline above, stratification of patients into distinct disease subclasses is desirable.

From a bioinformatic standpoint, we chose kallisto as our RNA-seq count quantifier and DESeq2 as our differential expression analysis method in favour over methods such as Bowtie, TopHat, and Cufflinks for quantifying reads and edgeR or limma-voom for differential expression. We chose kallisto due to the speed afforded in pseudoalignment as compared to

classical genome alignment of a large number of samples while maintaining accuracy. Kallisto was also used to generate normalised counts in TPMs, enabling integration with data sources from which we did not start with raw reads but with normalised counts, and enabling metabolic modelling analysis which requires TPMs. We chose DESeq2 as one of the most prevalent differential expression methods, but edgeR, which works on the same principle of assuming that most genes are not differentially expressed (Dillies et al, 2013), would be equally valid. Due to our custom normalisation strategy and manual setting of DESeq2 size factors to uniform, it is not anticipated that the choice of differential expression analysis method would adversely affect the overall conclusions. Finally, transcript-level analysis was not performed due to loss of analytical power and potential exaggeration when interpreting lowly expressed transcripts.

In conclusion, we report three distinct subclasses of AD and PD. The first subclass was identified as being associated with increased immune response, inflammatory response, and reduced sensory perception of smell, according to GSE results. We observed that this subclass exhibited increased oestradiol turnover, according to FBA results. We therefore propose that subjects consistent with the first subclass may be treatable with combined retinoid and oestradiol therapy. The second subclass was linked with increased cholesterol biosynthesis and general increased flux through the androgen biosynthesis and metabolism pathway. This subclass was characterised by reduced immune response. We therefore suggest that subjects consistent with the second subclass be studied further with combined retinoid and statin therapy. The third subclass was characterised by enrichment of GO terms indicating increased sensory perception of smell, reduced haemostasis, and reduced immune and DNA damage response. This subclass also exhibited reduced testosterone biosynthesis from androstenedione, as determined by FBA. We therefore hypothesise that subjects consistent with the third subclass may benefit from combined retinoid and testosterone therapy. For all subclasses of AD and PD, more investigation is required to verify the effectiveness of these stratification methods and precision therapies. To our knowledge this is the first meta-analysis at this scale highlighting the potential significance of NDD therapy using retinoids, oestradiol, and testosterone by studying AD and PD in combination. We observed that the existence of disease subclasses demands precision or personalised medicine and explains the heterogeneity in NDD response to single-factor treatments.

Materials and methods

Data acquisition and processing

Gene expression values of protein-coding genes from the ROSMAP dataset were determined using kallisto (Bray et al., 2016) by aligning raw RNA sequencing reads to the *Homo sapiens* genome in Ensembl release 96 (Yates et al., 2020). Raw single-cell RNA sequencing reads from ROSMAP were converted to counts in Cell Ranger 4.0 (10X Genomics, https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/installation) and aligned to the Cell Ranger *Homo sapiens* reference transcriptome version 2020-A. Single-cell expression values were compiled into pseudo-bulk expression profiles for each sample.

AD, PD, and control brain expression values of protein-coding genes from the ROSMAP dataset (Myers et al., 2007; Webster et al., 2009; Mostafavi et al., 2018), GTEx database version 8 (GTEx Consortium, 2013), FANTOM5 database (Forrest et al., 2014; Lizio et al., 2015, 2019) via Regulatory Circuits Network Compendium 1.0 (Marbach et al., 2016), HPA database (Uhlén et al., 2015), Rajkumar dataset (Rajkumar et al., 2020), and Zhang/Zheng dataset (Zhang et al., 2005; Zheng et al., 2010) were then combined. Genes from GTEx and FANTOM5 brain samples were filtered such that only genes whose products are known to participate in a protein-protein interaction described in the HuRI database (Luck et al., 2019) were included. Expression values were scaled and TMM normalised per sample, Pareto scaled per gene, and batch effects removed with the *removeBatchEffects* function from the limma (Ritchie et al., 2015) R package. After quality control and normalisation, a total of 64794 genes and 2055 samples resulted, of which 1572 samples corresponding to AD, PD, or control were accepted for analysis.

Projections onto 2-D space by PCA, t-SNE (Van Der Maaten and Hinton, 2008), and UMAP (McInnes et al., 2018) methods were generated on data after missing value imputation with data diffusion (van Dijk et al., 2018). t-SNE projections were generated with perplexity 20 and 1000 iterations. All other parameters were kept default. PCA and UMAP projections were generated using all default parameters.

Transcriptome analysis

Using normalised, imputed expression values, AD and PD samples were then arranged into clusters without supervision using ConsensusClusterPlus (Wilkerson and Hayes, 2010)

with maxK = 20 and rep = 1000. All other parameters were kept default. Clustering by $k = 3$ clusters was selected for downstream analysis. A fourth cluster containing only control samples was artificially added to the analysis.

For differential gene expression analysis, normalised, non-imputed counts were used. Genes were removed if expression values were missing in 40% or more of samples or were zero in all samples. Differential expression was then performed using DESeq2 (Love et al., 2014) with uniform size factors and all other parameters set to default. Genes with a Benjamini-Hochberg adjusted *p*-value at or below a cut-off of 1×10^{-10} were determined significantly differentially expressed genes.

Gene set enrichment analysis was performed using piano (Väremo et al., 2013) using all default parameters. GO term lists were obtained from Ensembl Biomart [https://www.ensembl.org/biomart/martview, accessed 2021-03-09] and were used as gene set collections. Enrichment of GO terms was determined by analysing GO terms of genes differentially expressed genes detected by DESeq2 as well as the parents of those GO terms. GO terms with an adjusted *p*-value at or below 0.05 for distinct-directional and/or mixed-directional methods were determined statistically significant.

Metabolic analysis

For each cluster, consensus gene expression values were determined by taking the geometric mean of normalised expression counts across all samples within each cluster.

A reference GEM was created by modifying the gene associations of all reactions within the adipocyte-specific GEM *iAdipocytes1850* (Mardinoglu et al., 2013) to match those within the generic human GEM HMR3 (Mardinoglu et al., 2014). The resulting GEM was designated *iBrain2845*. Cluster-specific GEMs were reconstructed using the RAVEN Toolbox 2.0 (Wang et al., 2018) tINIT algorithm (Agren et al., 2012, 2014) with *iBrain2845* as the reference GEM.

FBA was conducted on each cluster-specific GEM using the *solveLP* function from the RAVEN Toolbox 2.0 with previously reported constraints (Baloni et al., 2020) and defining ATP synthesis (*iBrain2845:* HMR_6916) as the objective function. All constraints were applied with the exception of the following reaction IDs, which were excluded: EX_ac[e] (*iBrain2845*: HMR_9086) and EX_etoh[e] (*iBrain2845*: HMR_9099).

Reporter metabolite analysis was conducted using the *reporterMetabolites* function (Patil and Nielsen, 2005) from the RAVEN Toolbox 2.0, using *iBrain2845* as the reference model.

Network analysis

To generate gene networks, normalised, non-imputed expression values from AD and PD samples were taken. Control samples and samples from blood were excluded. One network was generated each for AD and PD. For the AD model, all male samples were included and 171 female samples were chosen at random and included in order to ensure equal representation of male and female samples. Samples without sex recorded were disregarded. For the PD model, all samples were included. Genes with any missing values were dropped. Genes with the 15% lowest expression or 15% lowest variance were disregarded from further analysis. Spearman correlations were calculated for each pair of genes and the top 1% of significant correlations were used to generate gene co-expression networks. Random Erdős-Rényi models were created for the AD and PD models with the same numbers of nodes and edges to act as null models, and compared against their respective networks in terms of centrality distributions. Community analyses were performed through the Leiden algorithm (Traag et al., 2019) by optimizing CPMVertexPartition, after a resolution scan of 10,000 points between 10^{-3} and 10. The scan showed global maxima at resolutions = 0.077526 and 0.089074 for AD and PD networks, which were used for optimization. Enrichment analysis was performed on modules with >30 nodes using enrichr (Chen et al., 2013; Kuleshov et al., 2016) using GO Biological Process, KEGG, and Online Mendelian Inheritance in Man libraries and was explored using Revigo (Supek et al., 2011).

Zebrafish data acquisition and analysis

The *tert* mutant zebrafish line ($tert^{hu3430}$) was obtained from Miguel Godhino Ferreira (Henriques et al., 2013). Fish maintenance, RNA isolation, processing, and sequencing were conducted as described previously (Aramillo Irizar et al., 2018).

From n = 5 wildtype ($tert^{+/+}$), n = 5 heterozygous mutant ($tert^{+/-}$), and n = 3 homozygous mutant ($tert^{-/-}$), expression values were determined from RNA sequencing reads using kallisto by aligning to the *Danio rerio* genome in Ensembl release 96 (Yates et al., 2020). Expression values were generated for each extracted tissue as well as 'psuedo–whole animal', containing combined values across all tissues.

A reference zebrafish GEM was manually curated by modifying the existing *ZebraGEM2* model and was designated *ZebraGEM2.1*.

Differential expression analysis, gene set enrichment analysis, GEM reconstruction, FBA, and reporter metabolite analysis were conducted on *tert*$^{-/-}$ and *tert*$^{+/-}$ animals against a *tert*$^{+/+}$ reference using DESeq2, piano, and RAVEN Toolbox 2.0 with default parameters. Reporter metabolite analysis was conducted with *ZebraGEM2.1* as the reference GEM.

FBA was attempted as described for the human GEMs with the exception that the following metabolic constraints were excluded: r1391, HMR_0482 (*ZebraGEM2.1*: G3PDm), EX_ile_L[e] (*ZebraGEM2.1*: EX_ile_e), EX_val_L[e] (*ZebraGEM2.1*: EX_val_e), EX_lys_L[e] (*ZebraGEM2.1*: EX_lys_e), EX_phe_L[e] (*ZebraGEM2.1*: EX_phe_e), GLCt1r, EX_thr_L[e] (*ZebraGEM2.1*: EX_thr_e), EX_met_L[e] (*ZebraGEM2.1*: EX_met__L_e), EX_arg_L[e] (*ZebraGEM2.1*: EX_arg_e), EX_his_L[e] (*ZebraGEM2.1*: EX_his__L_e), EX_leu_L[e] (*ZebraGEM2.1*: EX_leu_e), and EX_o2[e] (*ZebraGEM2.1*: EX_o2_e). The objective function was defined as ATP synthesis (*ZebraGEM2.*1: ATPS4m). FBA results for zebrafish are not presented.

Data and code accessibility

All original computer code, models, and author-curated data files have been released under a Creative Commons Attribution ShareAlike 4.0 International Licence (https://creativecommons.org/licenses/by-sa/4.0/) and are freely available for download from <https://github.com/SimonLammmm/ad-pd-retinoid>.

Zebrafish *tert* mutant sequencing data have been deposited in the NCBI Gene Expression Omnibus (GEO) and are accessible through GEO Series accession numbers GSE102426, GSE102429, GSE102431, and GSE102434.

Ethics statement

Zebrafish were housed in the fish facility of the Leibniz Institute on Aging – Fritz Lipmann Institute (FLI) under standard conditions and a 14-h–light and 10-h–dark cycle. All animal procedures were performed in accordance with the German animal welfare guidelines and approved by the Landesamt für Verbraucherschutz Thüringen (TLV), Germany.

Conflicts of interest

The authors declare no competing financial interests.

Author contributions

A.M. supervised the study and designed the study on human data. R.K. and C.E. designed the study on zebrafish. N.H. performed the zebrafish RNA sequencing and generated the raw counts. With the exception of the network analysis, S.L. performed all *in silico* analysis

and M.A. and R.B. provided technical consultation. R.B. performed the network analysis. S.L. analysed all the results. S.L. wrote the manuscript with input from all the authors.

Funding

This work was supported by the German Ministry for Education and Research within the framework of the GerontoSys initiative (research core JenAge, funding code BMBF 0315581) to C.E; and the Knut and Alice Wallenberg Foundation (grant number 2017.0303) to A.M.

Acknowledgements

We are grateful to Catarina Henriques and Miguel Godinho Ferreira for sharing the *tert* mutant zebrafish line. We also thank Ivonne Heinze, Ivonne Görlich, and Marco Groth from the FLI sequencing facility for sequencing the zebrafish samples.

The authors acknowledge use of the research computing facility at King's College London, *Rosalind* (https://rosalind.kcl.ac.uk).

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 2019-12-06.

The results published here are in part based on data obtained from the AD Knowledge Portal (https://adknowledgeportal.synapse.org). Study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by NIA grants P30AG10161 (ROS), R01AG15819 (ROSMAP; genomics and RNAseq), R01AG17917 (MAP), R01AG30146, R01AG36042 (5hC methylation, ATACseq), RC2AG036547 (H3K9Ac), R01AG36836 (RNAseq), R01AG48015 (monocyte RNAseq) RF1AG57473 (single nucleus RNAseq), U01AG32984 (genomic and whole exome sequencing), U01AG46152 (ROSMAP AMP-AD, targeted proteomics), U01AG46161(TMT proteomics), U01AG61356 (whole genome sequencing, targeted proteomics, ROSMAP AMP-AD), the Illinois Department of Public Health (ROSMAP), and the Translational Genomics Research Institute (genomic). Additional phenotypic data can be requested at www.radc.rush.edu. We thank the patients and their families for their selfless donation to further understanding Alzheimer's disease. This project was supported by funding

from the National Institute on Aging (AG034504 and AG041232). Many data and biomaterials were collected from several National Institute on Aging (NIA) and National Alzheimer's Coordinating Center (NACC, grant #U01 AG016976) funded sites. Amanda J. Myers, PhD (University of Miami, Department of Psychiatry) prepared the series. The directors, pathologist and technicians involved include: Rush University Medical Center, Rush Alzheimer's Disease Center (NIH #AG10161): David A. Bennett, M.D. Julie A. Schneider, MD, MS, Karen Skish, MS, PA (ASCP)MT, Wayne T Longman. The Rush portion of this study was supported by National Institutes of Health grants P30AG10161, R01AG15819, R01AG17917, R01AG36042, R01AG36836, U01AG46152, R01AG34374, R01NS78009, U18NS82140, R01AG42210, R01AG39478, and the Illinois Department of Public Health. - Quality control checks and preparation of the gene expression data was provided by the National Institute on Aging Alzheimer's Disease Data Storage Site (NIAGADS, U24AG041689) at the University of Pennsylvania.

References

Agren R, Bordel S, Mardinoglu A, Pornputtapong N, Nookaew I, Nielsen J (2012) Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT Maranas CD, ed. PLoS Comput Biol 8:e1002518 Available at: http://dx.plos.org/10.1371/journal.pcbi.1002518 [Accessed January 15, 2020].

Agren R, Mardinoglu A, Asplund A, Kampf C, Uhlen M, Nielsen J (2014) Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. Mol Syst Biol 10:721.

Altay O, Nielsen J, Uhlen M, Boren J, Mardinoglu A (2019) Systems biology perspective for studying the gut microbiota in human physiology and liver diseases. EBioMedicine 49:364–373 Available at: https://linkinghub.elsevier.com/retrieve/pii/S2352396419306486 [Accessed November 21, 2019].

Anchelin M, Alcaraz-Pérez F, Martínez CM, Bernabé-García M, Mulero V, Cayuela ML (2013) Premature aging in telomerase-deficient zebrafish. DMM Dis Model Mech 6:1101–1112 Available at: /pmc/articles/PMC3759330/?report=abstract [Accessed January 14, 2021].

Aramillo Irizar P et al. (2018) Transcriptomic alterations during ageing reflect the shift from cancer to degenerative diseases in the elderly. Nat Commun 9:327 Available at: https://pubmed.ncbi.nlm.nih.gov/29382830/ [Accessed February 15, 2021].

Bajaj A, Driver JA, Schernhammer ES (2010) Parkinson's disease and cancer risk: A systematic review and meta-analysis. Cancer Causes Control 21:697–707.

Baloni P et al. (2020) Metabolic Network Analysis Reveals Altered Bile Acid Synthesis and Metabolism in Alzheimer's Disease. Cell Reports Med 1:100138 Available at: https://doi.org/10.1016/j.xcrm.2020.100138 [Accessed January 19, 2021].

Baum LW (2005) Sex, hormones, and Alzheimer's disease. Journals Gerontol - Ser A Biol Sci Med Sci 60:736–743 Available at: https://academic.oup.com/biomedgerontology/article-lookup/doi/10.1093/gerona/60.6.736 [Accessed January 21, 2021].

Blomhoff R, Blomhoff HK (2006) Overview of retinoid metabolism and function. J Neurobiol 66:606–630 Available at: https://pubmed.ncbi.nlm.nih.gov/16688755/

[Accessed February 17, 2021].

Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34:525–527.

Carneiro MC, De Castro IP, Ferreira MG (2016) Telomeres in aging and disease: Lessons from zebrafish. DMM Dis Model Mech 9:737–748 Available at: https://dmm.biologists.org/content/9/7/737 [Accessed November 17, 2020].

Chakravarti D, LaBella KA, DePinho RA (2021) Telomeres: history, health, and hallmarks of aging. Cell 184:306–322 Available at: http://www.ncbi.nlm.nih.gov/pubmed/33450206 [Accessed January 25, 2021].

Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles G V., Clark NR, Ma'ayan A (2013) Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14:128 Available at: https://pubmed.ncbi.nlm.nih.gov/23586463/ [Accessed March 24, 2021].

Das B, Dasgupta S, Ray S (2019) Potential therapeutic roles of retinoids for prevention of neuroinflammation and neurodegeneration in Alzheimer's disease. Neural Regen Res 14:1880–1892 Available at: https://pubmed.ncbi.nlm.nih.gov/31290437/ [Accessed January 21, 2021].

De La Monte SM, Wands JR (2008) Alzheimer's disease is type 3 diabetes-evidence reviewed. J Diabetes Sci Technol 2:1101–1113.

Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D, Estelle J, Guernec G, Jagla B, Jouneau L, Laloë D, Le Gall C, Schaëffer B, Le Crom S, Geudj M, Jaffrézic F, French StatOmique Consortium (2013) A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief Bioinform 14:671–683.

Driver JA, Beiser A, Au R, Kreger BE, Splansky GL, Kurth T, Kiel DP, Lu KP, Seshadri S, Wolf PA (2012) Inverse association between cancer and Alzheimer's disease: Results from the Framingham Heart Study. BMJ 344:e1442.

Fitz NF, Nam KN, Koldamova R, Lefterov I (2019) Therapeutic targeting of nuclear receptors, liver X and retinoid X receptors, for Alzheimer's disease. Br J Pharmacol 176:3599–3610 Available at: https://pubmed.ncbi.nlm.nih.gov/30924124/ [Accessed January 21, 2021].

Forrest ARR et al. (2014) A promoter-level mammalian expression atlas. Nature 507:462–470.

Fukasawa H, Nakagomi M, Yamagata N, Katsuki H, Kawahara K, Kitaoka K, Miki T, Shudo K (2012) Tamibarotene: A candidate retinoid drug for Alzheimer's disease. Biol Pharm Bull 35:1206–1212 Available at: https://pubmed.ncbi.nlm.nih.gov/22863914/ [Accessed January 21, 2021].

Greenland JC, Williams-Gray CH, Barker RA (2019) The clinical heterogeneity of Parkinson's disease and its therapeutic implications. Eur J Neurosci 49:328–338 Available at: http://doi.wiley.com/10.1111/ejn.14094 [Accessed January 25, 2021].

Grosse L, Pâquet S, Caron P, Fazli L, Rennie PS, Bélanger A, Barbier O (2013) Androgen glucuronidation: An unexpected target for androgen deprivation therapy, with prognosis and diagnostic implications. Cancer Res 73:6963–6971 Available at: http://cancerres.aacrjournals.org/ [Accessed March 7, 2021].

GTEx Consortium (2013) The Genotype-Tissue Expression (GTEx) project. Nat Genet 45:580–585.

Henriques CM, Carneiro MC, Tenente IM, Jacinto A, Ferreira MG (2013) Telomerase Is Required for Zebrafish Lifespan. PLoS Genet 9:1003214 Available at: /pmc/articles/PMC3547866/ [Accessed February 15, 2021].

Hindle J V. (2010) Ageing, neurodegeneration and Parkinson's disease. Age Ageing 39:156–161 Available at: https://academic.oup.com/ageing/article-lookup/doi/10.1093/ageing/afp223 [Accessed January 25, 2021].

Jeong SM, Jang W, Shin DW (2019) Association of statin use with Parkinson's disease: Dose–response relationship. Mov Disord 34:1014–1021 Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/mds.27681 [Accessed January 25, 2021].

Joshi A, Rienks M, Theofilatos K, Mayr M (2020) Systems biology in cardiovascular disease: a multiomics approach. Nat Rev Cardiol 18:313–330 Available at: https://www.nature.com/articles/s41569-020-00477-1 [Accessed January 25, 2021].

Kuleshov M V., Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A (2016) Enrichr: a comprehensive gene set enrichment analysis web

server 2016 update. Nucleic Acids Res 44:W90–W97 Available at: https://pubmed.ncbi.nlm.nih.gov/27141961/ [Accessed March 24, 2021].

Lam S, Bayraktar A, Zhang C, Turkez H, Nielsen J, Boren J, Shoaie S, Uhlen M, Mardinoglu A (2020) A systems biology approach for studying neurodegenerative diseases. Drug Discov Today 25:1146–1159 Available at: https://doi.org/10.1016/j.drudis.2020.05.010.

Lam S, Doran S, Yuksel HH, Altay O, Turkez H, Nielsen J, Boren J, Uhlen M, Mardinoglu A (2021) Addressing the heterogeneity in liver diseases using biological networks. Brief Bioinform 22:1751–1766 Available at: https://doi.org/10.1093/bib/bbaa002.

Liberini P, Valerio A, Memo M, Spano PF (1996) Lewy-body dementia and responsiveness to cholinesterase inhibitors: A paradigm for heterogeneity of Alzheimer's disease? Trends Pharmacol Sci 17:155–160.

Lizio M et al. (2015) Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol 16:22 Available at: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0560-6 [Accessed January 14, 2021].

Lizio M, Abugessaisa I, Noguchi S, Kondo A, Hasegawa A, Hon CC, De Hoon M, Severin J, Oki S, Hayashizaki Y, Carninci P, Kasukawa T, Kawaji H (2019) Update of the FANTOM web resource: Expansion to provide additional transcriptome atlases. Nucleic Acids Res 47:D752–D758 Available at: http://fantom.gsc.riken. [Accessed January 14, 2021].

Long JM, Holtzman DM (2019) Alzheimer Disease: An Update on Pathobiology and Treatment Strategies. Cell 179:312–339 Available at: /pmc/articles/PMC6778042/?report=abstract [Accessed January 25, 2021].

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550 Available at: http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8 [Accessed February 14, 2020].

Luck K et al. (2019) A reference map of the human protein interactome. bioRxiv.

Marbach D, Lamparter D, Quon G, Kellis M, Kutalik Z, Bergmann S (2016) Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. Nat Methods 13:366–370.

Mardinoglu A, Agren R, Kampf C, Asplund A, Nookaew I, Jacobson P, Walley AJ, Froguel P, Carlsson LM, Uhlen M, Nielsen J (2013) Integration of clinical data with a genome-scale metabolic model of the human adipocyte. Mol Syst Biol 9:649 Available at: https://onlinelibrary.wiley.com/doi/abs/10.1038/msb.2013.5 [Accessed January 17, 2020].

Mardinoglu A, Agren R, Kampf C, Asplund A, Uhlen M, Nielsen J (2014) Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. Nat Commun 5:3083.

Mardinoglu A, Boren J, Smith U, Uhlen M, Nielsen J (2018) Systems biology in hepatology: approaches and applications. Nat Rev Gastroenterol Hepatol 15:365–377 Available at: http://www.nature.com/articles/s41575-018-0007-8 [Accessed November 21, 2019].

Martucci CP (1983) The role of 2-methoxyestrone in estrogen action. J Steroid Biochem 19:635–638.

McInnes L, Healy J, Saul N, Großberger L (2018) UMAP: Uniform Manifold Approximation and Projection. J Open Source Softw 3:861 Available at: http://joss.theoj.org/papers/10.21105/joss.00861 [Accessed January 20, 2021].

Meoni S, Macerollo A, Moro E (2020) Sex differences in movement disorders. Nat Rev Neurol 16:84–96 Available at: https://doi.org/10.1038/ [Accessed January 25, 2021].

Mostafavi S et al. (2018) A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. Nat Neurosci 21:811–819.

Myers AJ et al. (2007) A survey of genetic human cortical gene expression. Nat Genet 39:1494–1499 Available at: https://www.nature.com/articles/ng.2007.16 [Accessed January 14, 2021].

Patil KR, Nielsen J (2005) Uncovering transcriptional regulation of metabolism by using metabolic network topology. Proc Natl Acad Sci U S A 102:2685–2689 Available at: www.pnas.orgcgidoi10.1073pnas.0406811102 [Accessed January 19, 2021].

Rajkumar AP, Bidkhori G, Shoaie S, Clarke E, Morrin H, Hye A, Williams G, Ballard C, Francis P, Aarsland D (2020) Postmortem Cortical Transcriptomics of Lewy Body Dementia Reveal Mitochondrial Dysfunction and Lack of Neuroinflammation. Am J Geriatr Psychiatry 28:75–86.

Rajsombath MM, Nam AY, Ericsson M, Nuber S (2019) Female Sex and Brain-Selective Estrogen Benefit α-Synuclein Tetramerization and the PD-like Motor Syndrome in 3K Transgenic Mice. J Neurosci 39:7628–7640 Available at: https://doi.org/10.1523/JNEUROSCI.0313-19.2019 [Accessed January 13, 2021].

Resnick SM et al. (2017) Testosterone treatment and cognitive function in older men with low testosterone and age-associated memory impairment. JAMA - J Am Med Assoc 317:717–727 Available at: /pmc/articles/PMC5433758/?report=abstract [Accessed January 21, 2021].

Riahi RR, Bush AE, Cohen PR (2016) Topical Retinoids: Therapeutic Mechanisms in the Treatment of Photodamaged Skin. Am J Clin Dermatol 17:265–276 Available at: https://pubmed.ncbi.nlm.nih.gov/26969582/ [Accessed January 21, 2021].

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47.

Sengoku R (2020) Aging and Alzheimer's disease pathology. Neuropathology 40:22–29 Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/neup.12626 [Accessed January 25, 2021].

Shepardson NE, Shankar GM, Selkoe DJ (2011) Cholesterol level and statin use in Alzheimer disease: I. Review of epidemiological and preclinical studies. Arch Neurol 68:1239–1244 Available at: /pmc/articles/PMC3211071/?report=abstract [Accessed January 21, 2021].

Shudo K, Fukasawa H, Nakagomi M, Yamagata N (2009) Towards Retinoid Therapy for Alzheimers Disease. Curr Alzheimer Res 6:302–311 Available at: https://pubmed.ncbi.nlm.nih.gov/19519313/ [Accessed January 21, 2021].

Sinclair E, Trivedi DK, Sarkar D, Walton-Doyle C, Milne J, Kunath T, Rijs AM, de Bie RMA, Goodacre R, Silverdale M, Barran P (2021) Metabolomics of sebum reveals lipid dysregulation in Parkinson's disease. Nat Commun 12:1–9 Available at: https://doi.org/10.1038/s41467-021-21669-4 [Accessed March 25, 2021].

Stampfer MJ (2006) Cardiovascular disease and Alzheimer's disease: Common links. J Intern Med 260:211–223 Available at: http://doi.wiley.com/10.1111/j.1365-2796.2006.01687.x [Accessed March 24, 2020].

Supek F, Bošnjak M, Škunca N, Šmuc T (2011) Revigo summarizes and visualizes long lists of gene ontology terms Gibas C, ed. PLoS One 6:e21800 Available at: https://dx.plos.org/10.1371/journal.pone.0021800 [Accessed March 24, 2021].

Traag VA, Waltman L, van Eck NJ (2019) From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep 9:1–12 Available at: https://doi.org/10.1038/s41598-019-41695-z [Accessed March 9, 2021].

Uhlén M et al. (2015) Tissue-based map of the human proteome. Science 347:1260419–1260419 Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.1260419 [Accessed January 7, 2020].

Van Der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579–2625 Available at: http://jmlr.org/papers/v9/vandermaaten08a.html [Accessed January 20, 2021].

van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr AJ, Burdziak C, Moon KR, Chaffer CL, Pattabiraman D, Bierie B, Mazutis L, Wolf G, Krishnaswamy S, Pe'er D (2018) Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. Cell 174:716-729.e27 Available at: https://doi.org/10.1016/j.cell.2018.05.061 [Accessed January 20, 2021].

Van Steijn L, Verbeek FJ, Spaink HP, Merks RMH (2019) Predicting Metabolism from Gene Expression in an Improved Whole-Genome Metabolic Network Model of Danio rerio. Zebrafish 16:348–362 Available at: https://pubmed.ncbi.nlm.nih.gov/31216234/ [Accessed January 19, 2021].

Väremo L, Nielsen J, Nookaew I (2013) Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. Nucleic Acids Res 41:4378–4391 Available at: https://academic.oup.com/nar/article/41/8/4378/2408999 [Accessed January 18, 2021].

Wahjoepramono EJ, Asih PR, Aniwiyanti V, Taddei K, Dhaliwal SS, Fuller SJ, Foster J, Carruthers M, Verdile G, Sohrabi HR, Martins RN (2016) The Effects of Testosterone Supplementation on Cognitive Functioning in Older Men. CNS Neurol Disord - Drug Targets 15:337–343 Available at: /pmc/articles/PMC5078598/?report=abstract [Accessed January 21, 2021].

Wang H, Marcišauskas S, Sánchez BJ, Domenzain I, Hermansson D, Agren R, Nielsen J, Kerkhoven EJ (2018) RAVEN 2.0: A versatile toolbox for metabolic network

reconstruction and a case study on Streptomyces coelicolor Ouzounis CA, ed. PLoS Comput Biol 14:e1006541 Available at: http://dx.plos.org/10.1371/journal.pcbi.1006541 [Accessed January 15, 2020].

Webster JA et al. (2009) Genetic Control of Human Brain Transcript Expression in Alzheimer Disease. Am J Hum Genet 84:445–458 Available at: http://www.cell.com/article/S0002929709001086/fulltext [Accessed January 14, 2021].

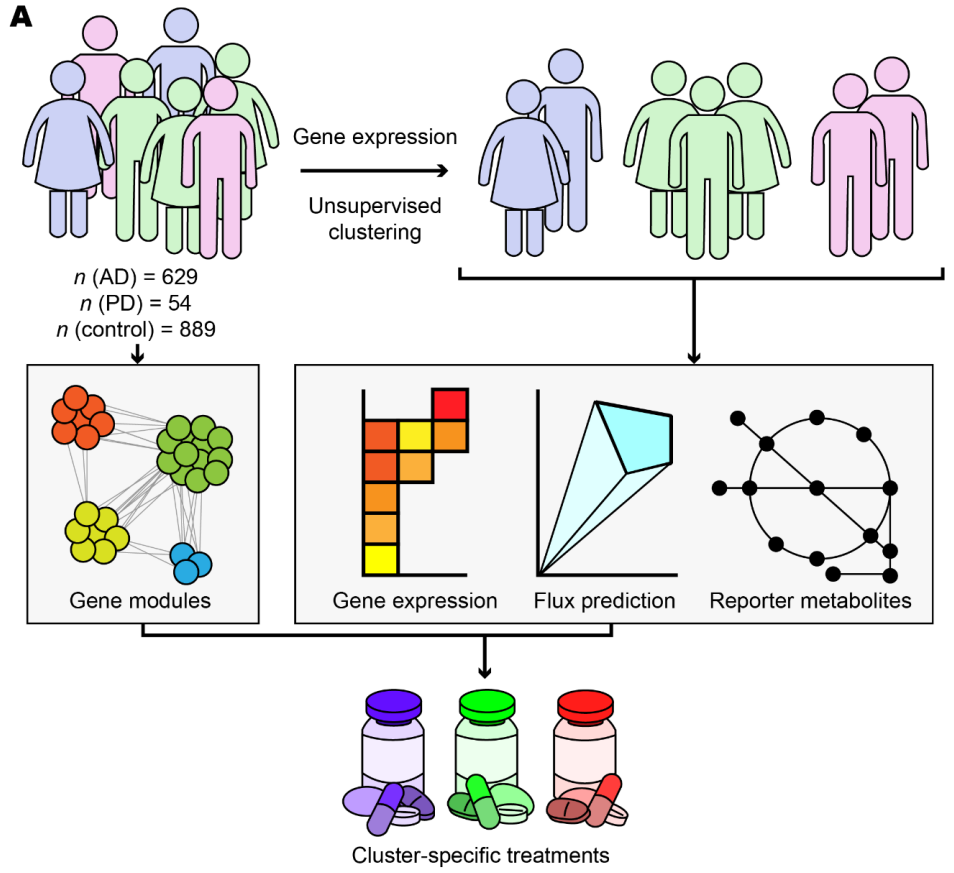
Wijemanne S, Jankovic J (2015) Dopa-responsive dystonia - Clinical and genetic heterogeneity. Nat Rev Neurol 11:414–424.

Wilkerson MD, Hayes DN (2010) ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. Bioinformatics 26:1572–1573 Available at: https://pubmed.ncbi.nlm.nih.gov/20427518/ [Accessed January 18, 2021].

Yates AD et al. (2020) Ensembl 2020. Nucleic Acids Res 48:D682–D688 Available at: https://academic.oup.com/nar/article/48/D1/D682/5613682 [Accessed January 20, 2021].

Zhang Y, James M, Middleton FA, Davis RL (2005) Transcriptional analysis of multiple brain regions in Parkinson's disease supports the involvement of specific protein processing, energy metabolism, and signaling pathways, and suggests novel disease mechanisms. Am J Med Genet - Neuropsychiatr Genet 137 B:5–16 Available at: https://pubmed.ncbi.nlm.nih.gov/15965975/ [Accessed July 9, 2020].

Zheng B et al. (2010) PGC-1α, a potential therapeutic target for early intervention in Parkinson's disease. Sci Transl Med 2:52ra73 Available at: https://pubmed.ncbi.nlm.nih.gov/20926834/ [Accessed July 9, 2020].


A
Gene expression
Unsupervised clustering
n (AD) = 629
n (PD) = 54
n (control) = 889
Gene modules
Gene expression
Flux prediction
Reporter metabolites
Cluster-specific treatments

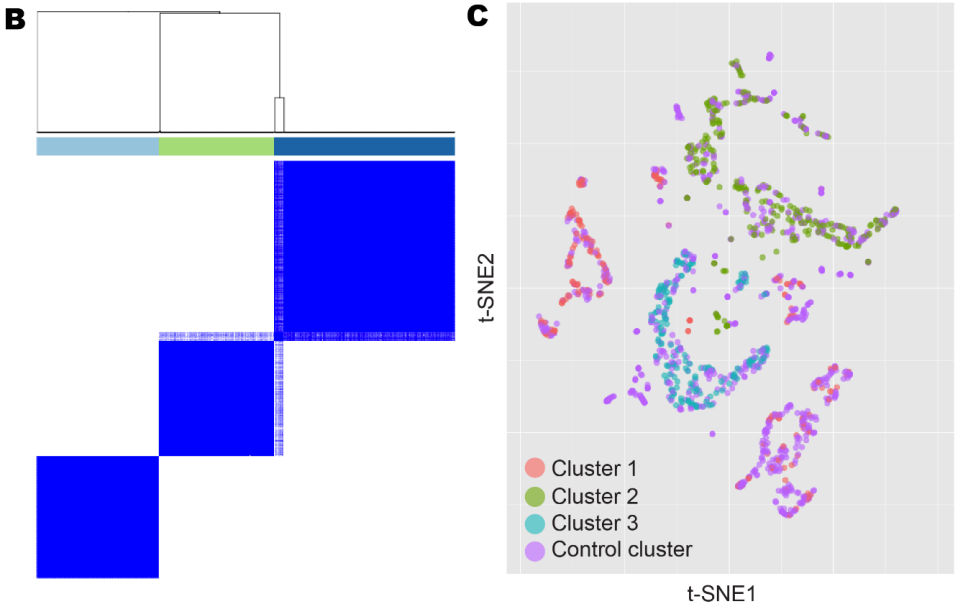

B
C
t-SNE2
t-SNE1
Cluster 1
Cluster 2
Cluster 3
Control cluster

Figure 1. Overview and exploratory data analysis. A) Workflow for the analysis of human AD and PD samples. **B)** AD and PD samples were clustered into *k* clusters without supervision on the basis of normalised expression counts. Results are shown $k = 3$ and 1000 bootstrap replicates. Colour bars indicate cluster identity for each sample. For $2 \leq k \leq 7$, refer to Supplementary figure 1. **C**) Normalised expression data from AD, PD, and control samples were projected onto 2-D space using t-distributed stochastic neighbour embedding (t-SNE). Points are coloured according to cluster assignment by unsupervised clustering. For further data visualisation, refer to Supplementary figure 2.

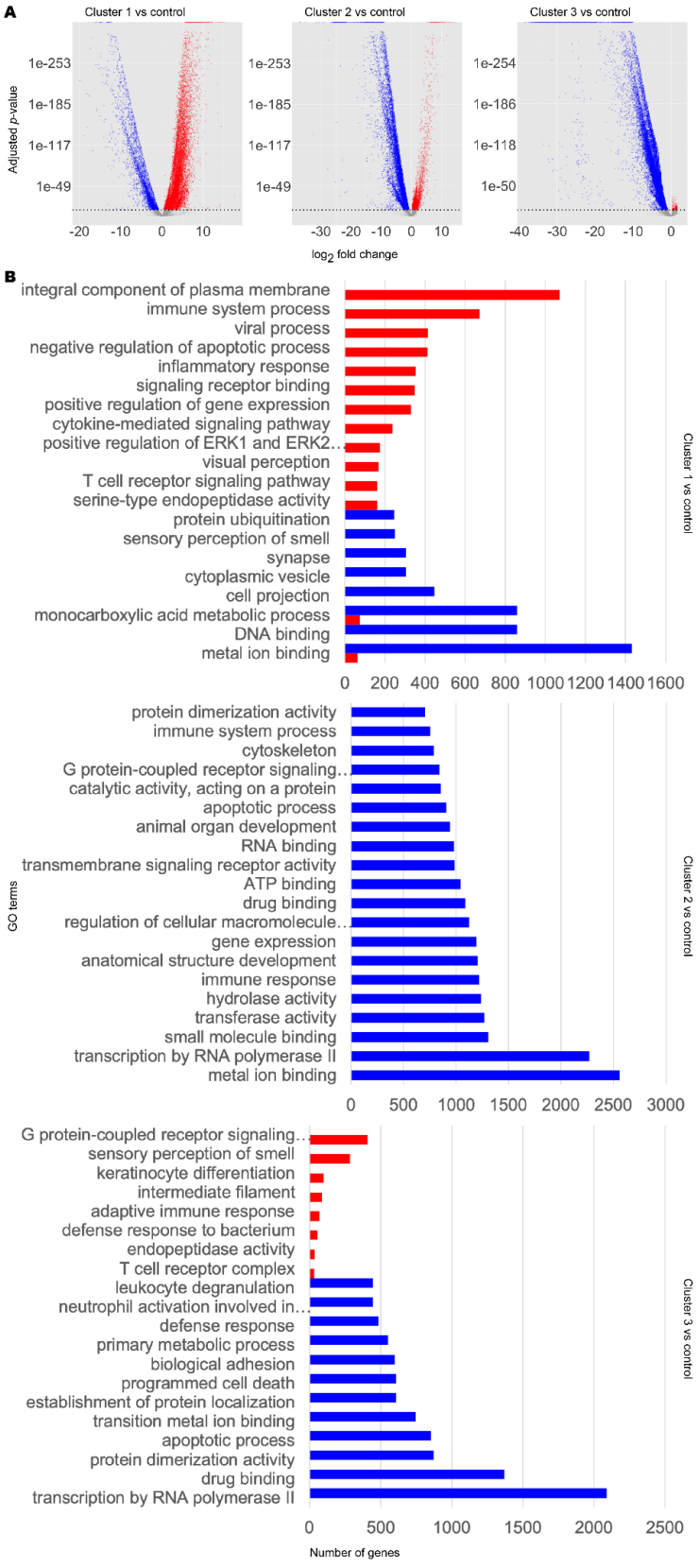

A
Cluster 1 vs control
Cluster 2 vs control
Cluster 3 vs control
Adjusted p-value
1e-253
1e-185
1e-117
1e-49
1e-254
1e-186
1e-118
1e-50
-20 -10 0 10
-30 -20 -10 0 10
-40 -30 -20 -10 0
log2 fold change
B
integral component of plasma membrane
immune system process
viral process
negative regulation of apoptotic process
inflammatory response
signaling receptor binding
positive regulation of gene expression
cytokine-mediated signaling pathway
positive regulation of ERK1 and ERK2…
visual perception
T cell receptor signaling pathway
serine-type endopeptidase activity
protein ubiquitination
sensory perception of smell
synapse
cytoplasmic vesicle
cell projection
monocarboxylic acid metabolic process
DNA binding
metal ion binding
0 200 400 600 800 1000 1200 1400 1600
Cluster 1 vs control
protein dimerization activity
immune system process
cytoskeleton
G protein-coupled receptor signaling…
catalytic activity, acting on a protein
apoptotic process
animal organ development
RNA binding
transmembrane signaling receptor activity
ATP binding
drug binding
regulation of cellular macromolecule…
gene expression
anatomical structure development
immune response
hydrolase activity
transferase activity
small molecule binding
transcription by RNA polymerase II
metal ion binding
0 500 1000 1500 2000 2500 3000
GO terms
Cluster 2 vs control
G protein-coupled receptor signaling…
sensory perception of smell
keratinocyte differentiation
intermediate filament
adaptive immune response
defense response to bacterium
endopeptidase activity
T cell receptor complex
leukocyte degranulation
neutrophil activation involved in…
defense response
primary metabolic process
biological adhesion
programmed cell death
establishment of protein localization
transition metal ion binding
apoptotic process
protein dimerization activity
drug binding
transcription by RNA polymerase II
0 500 1000 1500 2000 2500
Number of genes
Cluster 3 vs control

Figure 2. Transcriptomic and functional characterisation of AD and PD subclasses. Differentially expressed gene (DEG) analysis and gene set enrichment (GSE) analysis were performed for AD and PD and control samples for each disease cluster, using the control cluster as reference. **A)** DEG results. Significant DEGs were determined as those with a Benjamini-Hochberg adjusted *p*-value at or below a cut-off of 1×10^{-10}. Upregulated significant DEGs are coloured red. Downregulated significant DEGs are coloured blue. Non-significant DEGs are coloured grey. **B)** Selected significantly enriched GO terms by number of genes as determined by GSE analysis. Red bars indicate upregulated GO terms. Blue bars indicate downregulated GO terms. For full data, refer to Supplementary data 1.

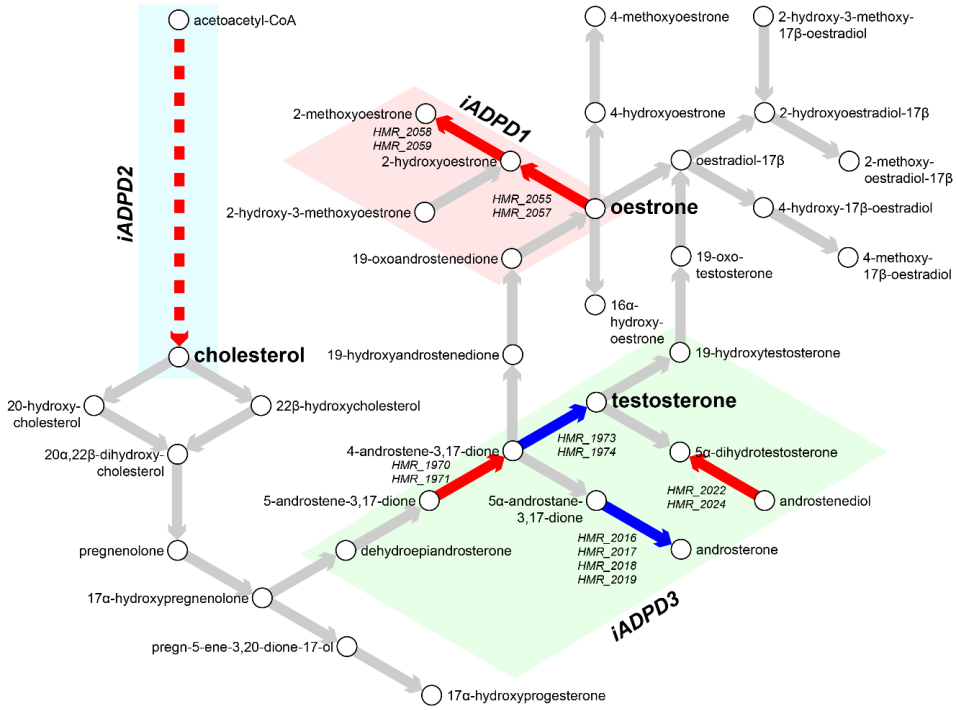


Figure 3. Metabolic characterisation of AD and PD subclasses. Flux balance analysis (FBA) was performed on *iADPD1-3* genome-scale metabolic models (GEMs) and flux values were compared with those of *iADPDControl*. Key metabolites and reactions within the androgen metabolism pathway are shown and key dysregulations are displayed as coloured arrows: red indicates increased flux compared to *iADPDControl*; blue indicates decreased flux compared to *iADPDControl*. Dysregulations associated to each GEM are shown in coloured boxes. The dashed line indicates multiple reactions are involved. Human Metabolic Reactions (HMR) identifiers are shown for androgen metabolism reactions with dysregulated fluxes. For full data, refer to Supplementary data 2.

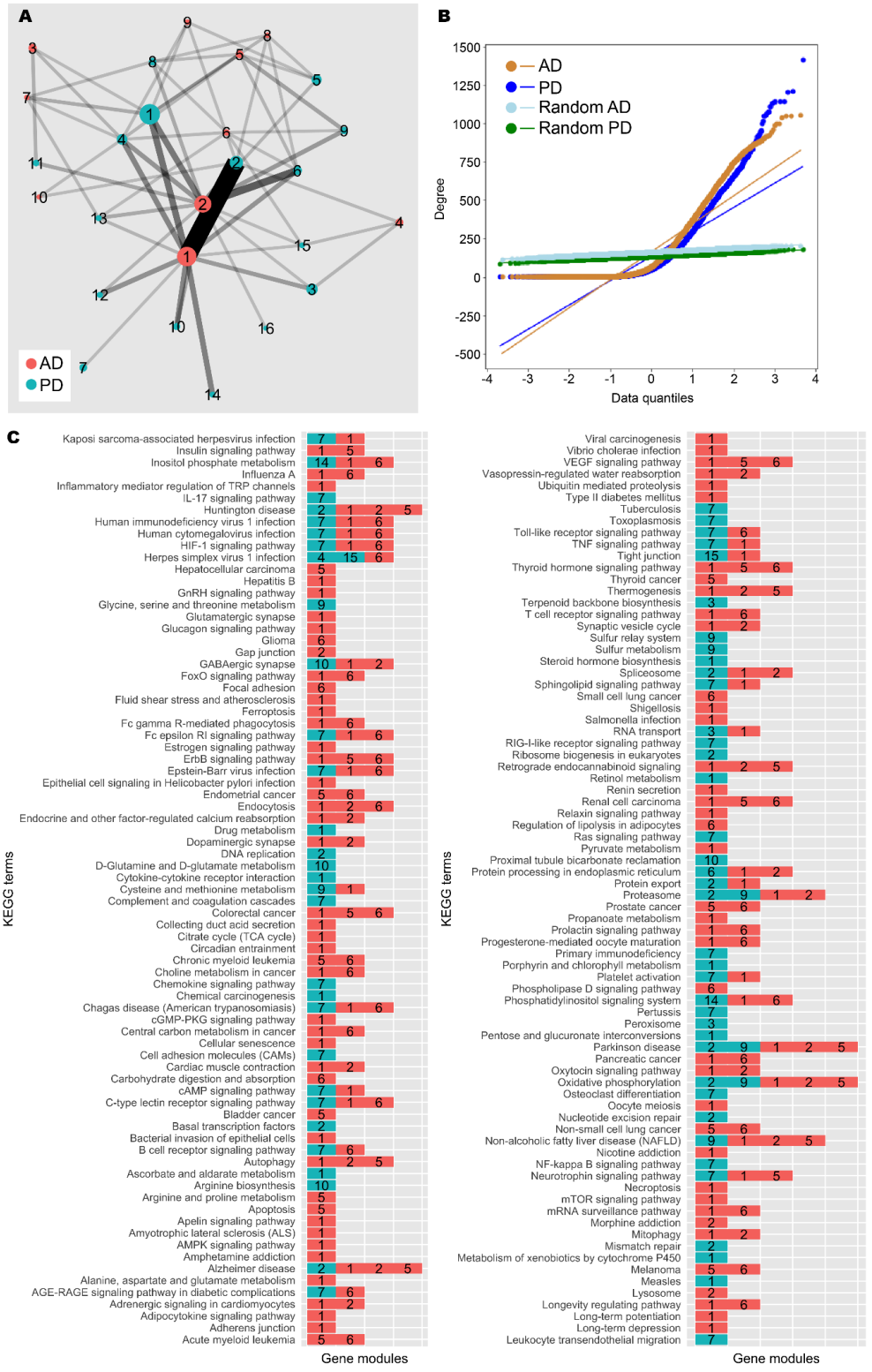


Figure 4. Network analysis of AD and PD gene co-expression modules. A) Gene co-

expression networks were constructed from transcriptomic data from AD and PD samples. Community analysis was used to identify gene modules (Methods). Modules with at least 30 genes are shown as nodes. Node size indicates number of genes. Nodes are coloured by network of origin and numbered in descending order of module size. Shared genes between modules are shown as edges. Edge weight indicates number of shared genes. **B**) Degree distribution of AD, PD, and random networks. **C)** Enrichment analysis was performed on gene modules containing at least 30 genes using the KEGG database (Methods). Significantly enriched gene modules are shown as coloured, numbered blocks. Colour and number keys are as in (A).

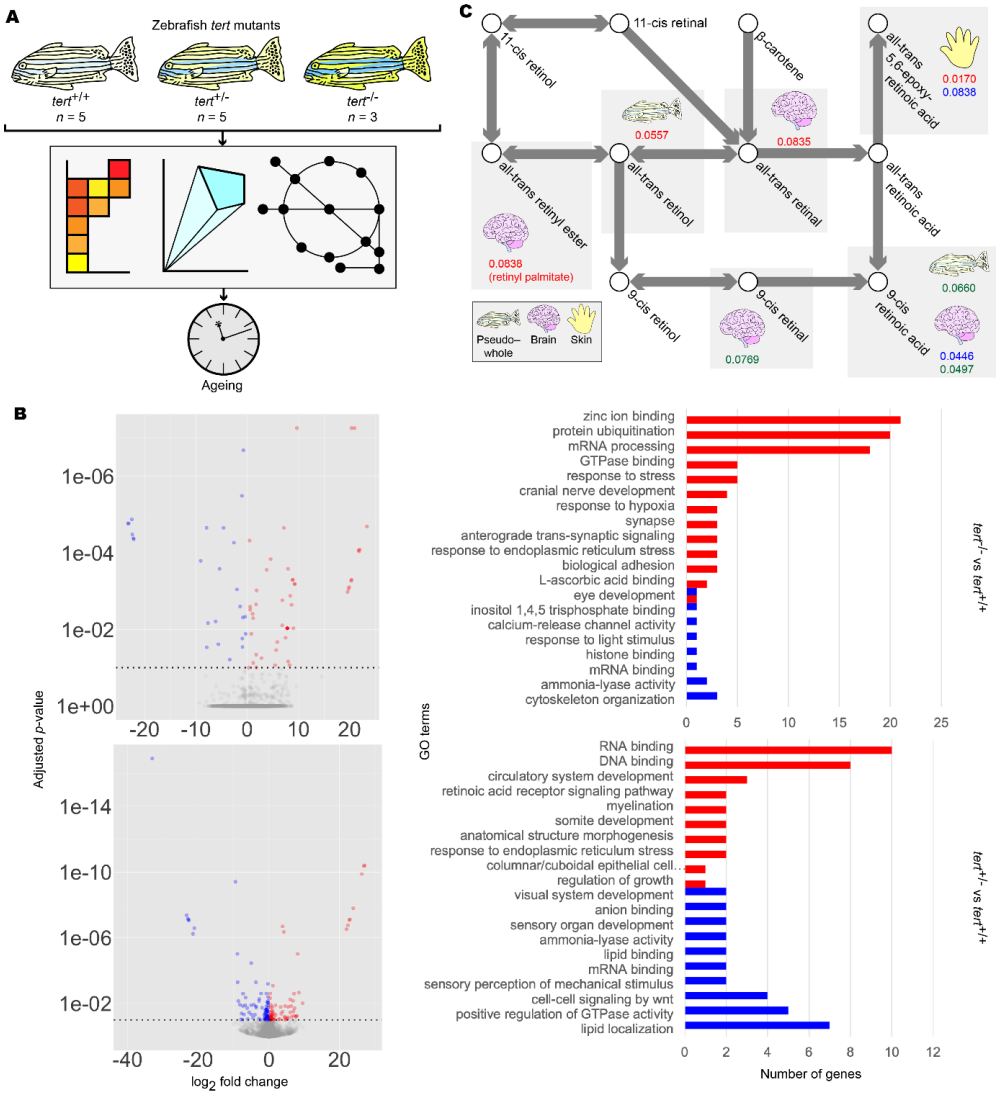


Figure 5. Summary of zebrafish *tert* mutant analysis. A) Workflow for the analysis of zebrafish *tert* mutants. **B)** Differentially expressed gene (DEG) (left panels) and gene set enrichment (GSE) analysis (right panels) of zebrafish brain samples. DEG and GSE analyses were performed on zebrafish *tert* mutant brain expression data for $tert^{-/-}$ (upper panels) and $tert^{+/-}$ (lower panels), using $tert^{+/+}$ as a reference. Methods and colour keys are as in Figure 2. For muscle, liver, skin, and pseudo–whole animal analyses, refer to Supplementary figure 3. For full data, refer to Supplementary data 5. **C)** Reporter metabolite analysis of zebrafish samples. DEG data were overlaid on *ZebraGEM2.1* to determine reporter metabolites. Shown are reporter metabolites with $p < 0.1$ within the retinoic acid metabolic pathway. Red numbers indicate *p*-values in $tert^{-/-}$ compared to $tert^{+/+}$. Blue numbers indicate *p*-values in $tert^{+/-}$ compared to $tert^{+/+}$. Green numbers indicate *p*-values in $tert^{-/-}$ compared to $tert^{+/-}$. Tissues are

indicated with icons. For full data, refer to Supplementary data 6.

Table 1. Summary of expression data sources. Expression data from AD and PD samples were obtained from the Genotype-Tissue Expression (GTEx) database, Functional Annotation of the Mammalian Genome 5 (FANTOM5) database, Human Protein Atlas (HPA), Religious Orders Study and Rush Memory Aging Project (ROSMAP), Rajkumar dataset, and Zhang/Zheng dataset.

Source	AD samples	PD samples	Control samples
GTEx/FANTOM5	0	0	67
HPA	0	0	52
Rajkumar	0	14	13
ROSMAP	629	0	704
Zhang/Zheng	0	40	53
Total	629	54	889

Table 2. Flux balance analysis of *iADPD1*, *iADPD2*, and *iADPD3* versus *iADPDControl*. Flux balance analysis was performed for each *iADPD*-series GEM and the predicted fluxes for the three disease cluster GEMs were compared against the predicted fluxes for the control cluster GEM. Reactions are grouped by subsystem and flux difference values are expressed as mean flux difference between disease clusters and the control cluster across all changed reactions within a subsystem. For full results, refer to Supplementary data 2.

Subsystem	***iADPD1***	***iADPD2***	***iADPD3***
Acyl-CoA hydrolysis	-0.001	0.001	0.000
Alanine, aspartate and glutamate metabolism	-0.148	0.014	0.000
Aminoacyl-tRNA biosynthesis	4.698	4.698	0.000
Androgen metabolism	-1.426	-0.399	-0.001
Arachidonic acid metabolism	-0.098	0.010	0.000
Arginine and proline metabolism	-0.182	-0.327	0.000
Beta oxidation of branched-chain fatty acids (mitochondrial)	-0.049	-0.049	-0.049
Beta oxidation of di-unsaturated fatty acids (n-6) (mitochondrial)	-0.636	0.002	-0.001
Beta oxidation of odd-chain fatty acids (mitochondrial)	0.001	-0.002	-0.002
Beta oxidation of poly-unsaturated fatty acids (mitochondrial)	0.709	0.024	0.000
Beta oxidation of unsaturated fatty acids (n-7) (mitochondrial)	-0.016	0.001	-0.003
Beta oxidation of unsaturated fatty acids (n-9) (mitochondrial)	0.011	0.000	0.007
Carnitine shuttle (cytosolic)	0.012	0.000	-0.001
Carnitine shuttle (mitochondrial)	0.003	0.000	0.002
Cholesterol biosynthesis 1 (Bloch pathway)	0.076	-0.983	0.001
Cholesterol biosynthesis 2	2.501	4.472	0.000
Cholesterol biosynthesis 3 (Kandustch-Russell pathway)	1.699	0.000	0.000

Cholesterol metabolism	0.067	4.482	0.000
Estrogen metabolism	2.085	0.000	0.000
Fatty acid activation (endoplasmic reticular)	0.000	0.000	0.000
Fatty acid biosynthesis (even-chain)	0.000	0.000	0.000
Fatty acid desaturation (even-chain)	0.785	0.000	0.000
Fatty acid elongation (odd-chain)	-0.042	-0.024	0.000
Formation and hydrolysis of cholesterol esters	-0.382	0.004	0.000
Fructose and mannose metabolism	-0.211	-0.007	0.000
Galactose metabolism	-0.008	0.035	0.000
Glycine, serine and threonine metabolism	0.276	0.557	0.000
Glycolysis / Gluconeogenesis	-0.213	0.022	0.033
Histidine metabolism	0.000	0.000	0.000
Leukotriene metabolism	-0.032	0.000	0.000
Lysine metabolism	0.000	0.000	0.000
N-glycan metabolism	-0.784	0.016	0.000
Nitrogen metabolism	0.000	0.000	0.000
Nucleotide metabolism	0.027	-0.028	0.000
O-glycan metabolism	-2.346	-4.738	0.000
Pentose phosphate pathway	0.127	0.000	0.000
Propanoate metabolism	-0.116	0.020	0.091
Protein degradation	0.000	0.000	0.000
Purine metabolism	0.112	-0.013	0.000
Pyrimidine metabolism	-0.071	-0.010	-0.001
Pyruvate metabolism	-0.183	-0.004	-0.077
Starch and sucrose metabolism	0.000	0.000	0.000
Steroid metabolism	-0.097	-0.295	0.003
Terpenoid backbone biosynthesis	0.398	0.187	0.020
Valine, leucine and isoleucine degradation	0.127	0.000	0.000

Table 3. Reporter metabolite analysis of AD and PD subclasses. Reporter metabolite analysis was performed for each AD/PD subclass by overlaying differential expression results onto *iBrain2845*. Top 10 unique reporter metabolites by *p*-value for each cluster compared to the control cluster are shown. For full results, refer to Supplementary data 3.

Reporter metabolite	Z-score	*p*-value
Cluster 1		
O2	6.111	4.95E-10
estrone	5.4557	2.44E-08
retinoate	5.3943	3.44E-08
NADP+	5.3667	4.01E-08
arachidonate	5.2822	6.38E-08
2-hydroxyestradiol-17beta	5.0999	1.70E-07
linoleate	5.0622	2.07E-07
10-HETE	5.0454	2.26E-07
11,12,15-THETA	5.0454	2.26E-07
11,14,15-theta	5.0454	2.26E-07
Cluster 2		
1-acylglycerol-3P-LD-PC pool	4.3322	7.38E-06
acyl-CoA-LD-PI pool	4.143	1.71E-05
phosphatidate-CL pool	4.0973	2.09E-05
thymidine	3.5852	0.00016843
uridine	3.5852	0.00016843
prostaglandin D2	3.2144	0.00065348
G10596	3.1354	0.0008581
G10597	3.1354	0.0008581
D-myo-inositol-1,4,5-trisphosphate	2.9988	0.0013552
dolichyl-phosphate	2.9655	0.001511
Cluster 3		

D-myo-inositol-1,4,5-trisphosphate	2.6543	0.0039734
13-cis-retinal	2.6537	0.0039806
heparan sulfate, precursor 9	2.5915	0.0047772
sn-glycerol-3-phosphate	2.578	0.0049682
DHAP	2.5353	0.0056173
porphobilinogen	2.4987	0.0062333
ATP	2.4838	0.0064998
L-glutamate 5-semialdehyde	2.4576	0.006994
prostaglandin D2	2.451	0.0071221
ribose	2.4133	0.0079045

Table 4. AD and PD network properties. Gene co-expression networks were generated for AD and PD samples. AD, PD, and random networks are shown.

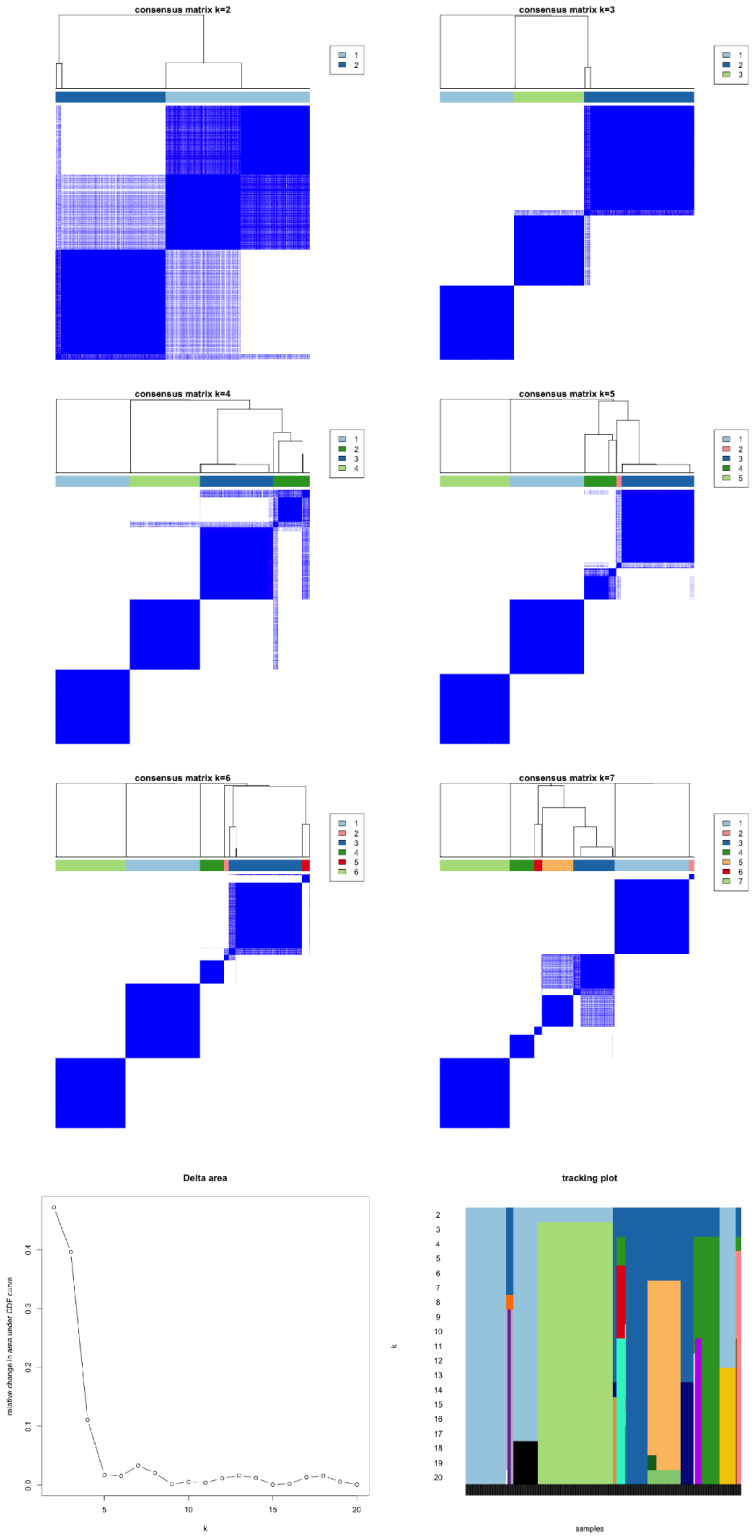
	Nodes	Edges	Diameter	Average path length	Density	Clustering coefficient	Connected network?	Minimum cut
AD	4861	396985	11	3.004	0.034	0.443	No	-
PD	5857	394405	18	3.598	0.023	0.397	No	-
Random AD	4861	396985	3	1.970	0.034	0.034	Yes	114
Random PD	5857	394405	3	2.021	0.023	0.023	Yes	89

Table 5. Reporter metabolite analysis of zebrafish *tert* mutants. Reporter metabolite analysis was performed for the brains of zebrafish *tert* mutant by overlaying differential expression results onto *ZebraGEM2.1*. Top 20 unique reporter metabolites by *p*-value for each cluster compared to wildtype $tert^{+/+}$ zebrafish are shown. For full results, refer to Supplementary data 6.

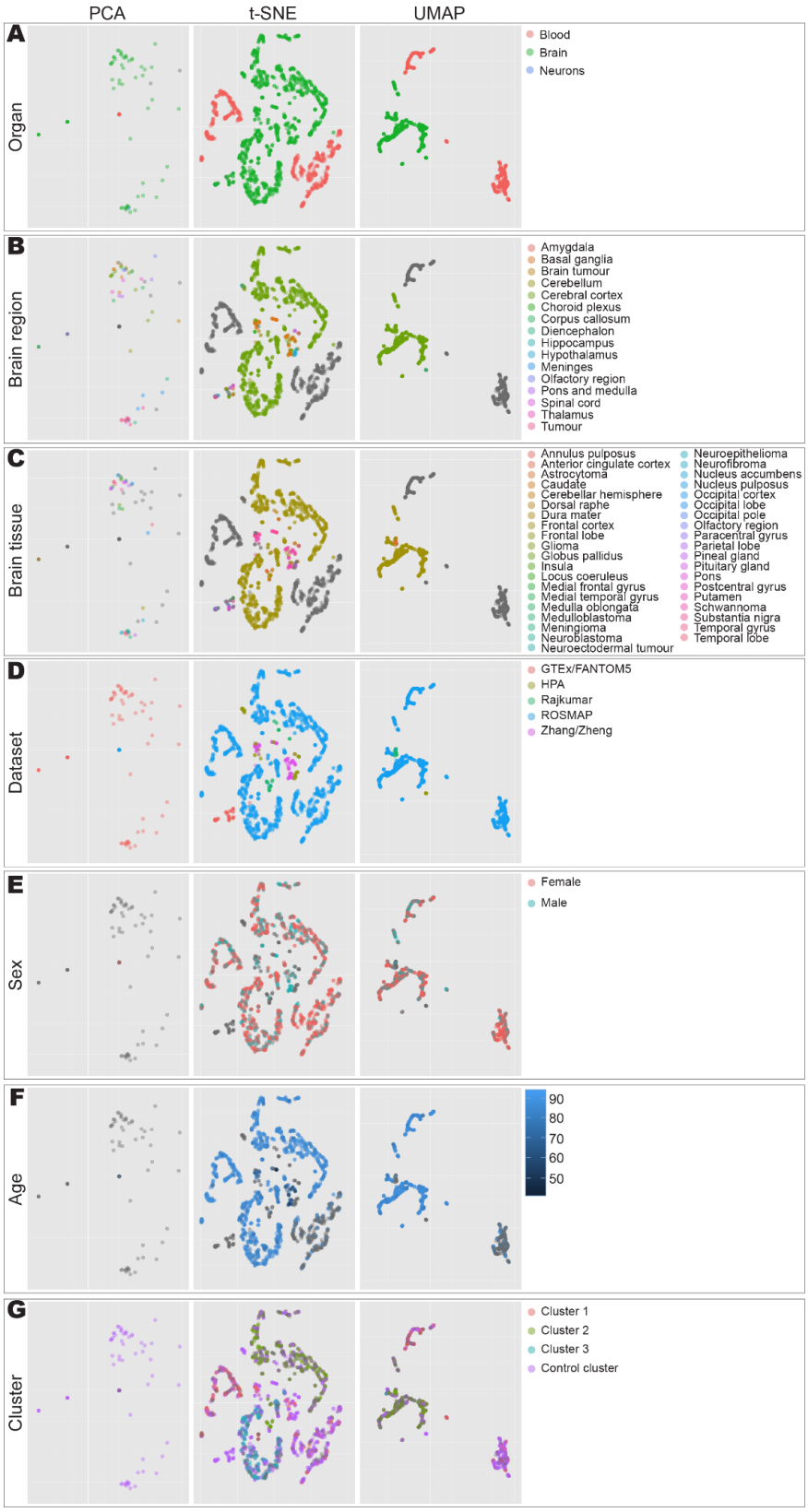
Reporter metabolite	**Z-score**	***p*-value**
***tert*$^{+/-}$**		
H+	3.911	4.60E-05
H2O	3.0672	0.0010804
L-Lysine	2.8564	0.0021424
Biocyt c	2.8564	0.0021424
Ubiquinone	2.5742	0.0050241
Nicotinamide adenine dinucleotide - reduced	2.3946	0.0083183
Phosphate	2.0562	0.019883
Superoxide anion	2.0365	0.020851
Sodium	1.9228	0.027254
TRNA (Glu)	1.8752	0.030381
Thiosulfate	1.7684	0.038493
Selenate	1.7684	0.038493
Reduced glutathione	1.7184	0.042862
ADP	1.6716	0.047305
L-Lysine	1.6625	0.04821
Benzo[a]pyrene-4,5-oxide	1.6042	0.054333
Formaldehyde	1.5955	0.055302
L-Glutamate	1.4622	0.071837
(1R,2S)-Naphthalene epoxide	1.4518	0.073276
Aflatoxin B1 exo-8,9-epozide	1.4518	0.073276
***tert*$^{+/-}$**		
H+	4.9585	3.55E-07

Ubiquinol	3.9938	3.25E-05
H2O	3.2078	0.00066883
Nicotinamide adenine dinucleotide - reduced	3.029	0.0012268
Superoxide anion	2.0908	0.018274
L-Lactate	2.0752	0.018983
O2	1.9958	0.022976
Lnlncgcoa c	1.9628	0.024834
Succinate	1.9449	0.025895
Ferricytochrome c	1.8352	0.033237
Phosphatidylinositol-3,4,5-trisphosphate	1.7494	0.040109
9-cis-Retinoic acid	1.7	0.044567
[(Gal)2 (GlcNAc)4 (LFuc)1 (Man)3 (Asn)1']	1.6672	0.047739
O-Phospho-L-serine	1.6601	0.048451
[(Glc)3 (GlcNAc)2 (Man)9 (Asn)1']	1.6276	0.051802
Protein serine	1.6078	0.053937
[(GlcNAc)1 (Ser/Thr)1']	1.6078	0.053937
Geranyl diphosphate	1.5912	0.055785
CTP	1.5625	0.059088
[(Gal)2 (GlcNAc)4 (LFuc)1 (Man)3 (Neu5Ac)2 (Asn)1']	1.5367	0.062179

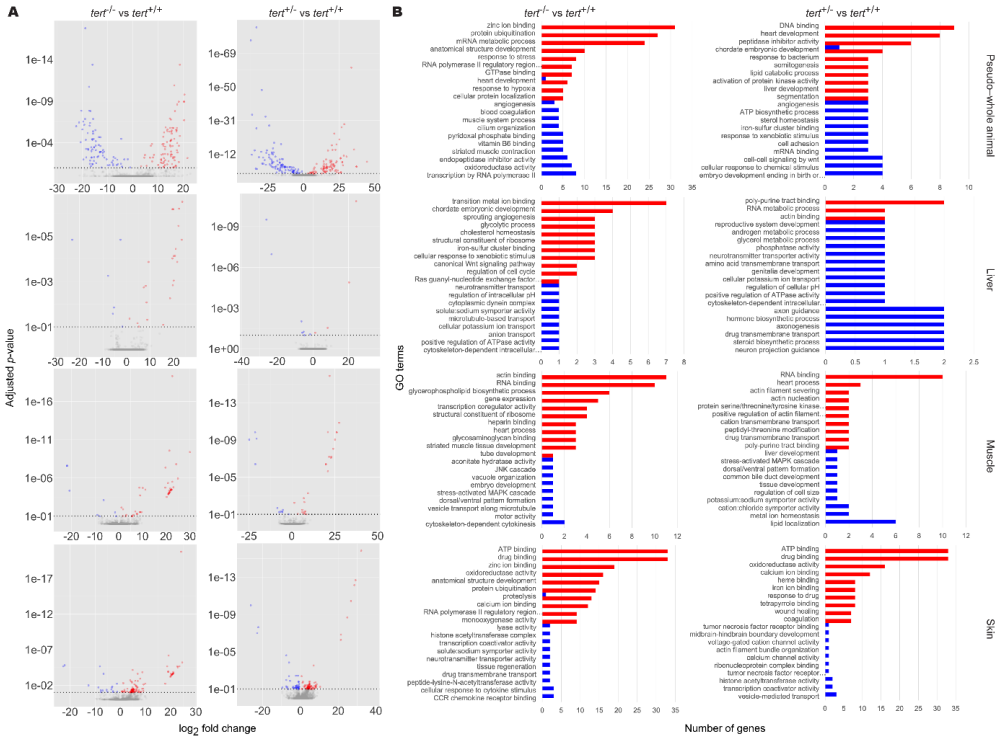
Supplementary figures and legends



Supplementary figure 1. Unsupervised clustering of AD and PD samples. AD and PD samples were clustered into k clusters without supervision on the basis of normalised expression counts. Clustering was performed with $2 \leq k \leq 20$. Consensus matrices for $2 \leq k \leq 7$ are shown. Parameters and colour keys are as in Figure 1b.


PCA
t-SNE
UMAP
A
Organ
Blood
Brain
Neurons
B
Brain region
Amygdala
Basal ganglia
Brain tumour
Cerebellum
Cerebral cortex
Choroid plexus
Corpus callosum
Diencephalon
Hippocampus
Hypothalamus
Meninges
Olfactory region
Pons and medulla
Spinal cord
Thalamus
Tumour
C
Brain tissue
Annulus pulposus
Anterior cingulate cortex
Astrocytoma
Caudate
Cerebellar hemisphere
Dorsal raphe
Dura mater
Frontal cortex
Frontal lobe
Glioma
Globus pallidus
Insula
Locus coeruleus
Medial frontal gyrus
Medial temporal gyrus
Medulla oblongata
Medulloblastoma
Meningioma
Neuroblastoma
Neuroectodermal tumour
Neuroepithelioma
Neurofibroma
Nucleus accumbens
Nucleus pulposus
Occipital cortex
Occipital lobe
Occipital pole
Olfactory region
Paracentral gyrus
Parietal lobe
Pineal gland
Pituitary gland
Pons
Postcentral gyrus
Putamen
Schwannoma
Substantia nigra
Temporal gyrus
Temporal lobe
D
Dataset
GTEx/FANTOM5
HPA
Rajkumar
ROSMAP
Zhang/Zheng
E
Sex
Female
Male
F
Age
90
80
70
60
50
G
Cluster
Cluster 1
Cluster 2
Cluster 3
Control cluster

Supplementary figure 2. Visualisation of AD and PD samples. Expression data from AD and PD and control samples were integrated, normalised, and projected onto 2-D space using principal component analysis (PCA), t-distributed stochastic neighbour embedding (t-SNE), and uniform manifold approximation and projection (UMAP). Points are coloured according to **A)** organ of sample origin, **B)** brain subregion of sample origin, **C)** brain tissue of sample origin, **D)** dataset, **E)** sex, **F)** age, or **G)** cluster assignment by unsupervised clustering. Points with no data available are shown in grey.



Supplementary figure 3. Transcriptomic and functional characterisation of zebrafish *tert* mutants. Differentially expressed gene (DEG) and gene set enrichment (GSE) analyses were performed on zebrafish *tert* mutant expression data for *tert*$^{-/-}$ and *tert*$^{+/-}$, using *tert*$^{+/+}$ as a reference. Methods and colour keys are as in Figure 2. **A)** DEG analyses. **B)** GSE analyses. Left panels, *tert*$^{-/-}$ vs *tert*+/+. Right panels, *tert*$^{+/-}$ vs *tert*$^{+/+}$. Panels top to bottom: pseudo–whole animal, liver, muscle, skin. For the brain, refer to Figure 5. For full results, refer to Supplementary data 5.

Paper II: Combined metabolic activators improve cognitive functions without altering motor scores in Parkinson's disease

The provided article below is adapted from the author's original version of the manuscript.

Combined Metabolic Activators Improve Cognitive Functions without Altering Motor Scores in Parkinson's Disease

Burak Yulug[1,#], Ozlem Altay[2,#], Xiangyu Li[2,#], Lutfu Hanoglu[3], Seyda Cankaya[1], Simon Lam[4], Hong Yang[2], Ebru Coskun[3], Ezgi İdil[1], Rahim Nogaylar[1], Ahmet Hacımuftuoglu[5], Muhammad Arif[2], Saeed Shoaie[2,4], Cheng Zhang[2,6], Jens Nielsen[7], Hasan Turkez[8], Jan Borén[9], Mathias Uhlén[2,*], Adil Mardinoglu[2,4,*]

[1] Department of Neurology and Neuroscience, Faculty of Medicine, Alanya Alaaddin Keykubat University, Antalya, Turkey

[2] Science for Life Laboratory, KTH - Royal Institute of Technology, Stockholm, Sweden

[3] Department of Neurology, Faculty of Medicine, Istanbul Medipol University, Istanbul, Turkey

[4] Centre for Host-Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, London, United Kingdom

[5] Department of Medical Pharmacology, Faculty of Medicine, Atatürk University, Erzurum, Turkey

[6] School of Pharmaceutical Sciences, Zhengzhou University, Zhengzhou, PR China

[7] Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

[8] Department of Medical Biology, Faculty of Medicine, Atatürk University, Erzurum, Turkey

[9] Department of Molecular and Clinical Medicine, University of Gothenburg and Sahlgrenska University Hospital, Gothenburg, Sweden

*Correspondence: adilm@scilifelab.se; mathias.uhlen@scilifelab.se

Emails: burakyulug@gmail.com; havva.altay@scilifelab.se; xiangyu.li@scilifelab.se; lhanoglu@kure.com.tr; cankayaseyda@hotmail.com; simon.1.lam@kcl.ac.uk; hong.yang@scilifelab.se; ebrucosskun@gmail.com; ezgi.idil92@gmail.com; ngylrrhm@gmail.com; ahmeth@atauni.edu.tr; muhammad.arif@scilifelab.se; saeed.shoaie@scilifelab.se; cheng.zhang@scilifelab.se; nielsenj@chalmers.se; hasanturkez@yahoo.com; Jan.Boren@wlab.gu.se; mathias.uhlen@scilifelab.se; adilm@scilifelab.se

ABSTRACT

The neuropathologic hallmarks of Parkinson's disease (PD) are associated with mitochondrial dysfunction and metabolic abnormalities. We have reported that combined metabolic activators (CMA) L-serine, nicotinamide riboside, N-acetyl-L-cysteine, and L-carnitine tartrate can be used in treating metabolic abnormalities. These metabolic activators are the precursors of nicotinamide adenine dinucleotide (NAD^+) and glutathione (GSH) and are used in activation of mitochondrial and global metabolism. We previously performed a placebo-controlled, phase-2 study in Alzheimer's disease (AD) patients and reported that cognitive functions in AD patients was significantly improved 29% in the CMA group whereas it was improved only 14% in the placebo group after 84 days. Here, we designed a randomized, double-blinded, placebo-controlled, phase-2 study in PD patients with CMA administration. We found that the cognitive functions in PD patients was significantly improved 21% in the CMA group, whereas it was improved only 11% in the placebo group after 84 days of CMA administration. We also found that the administration of CMA did not affect motor functions in PD patients. We performed a comprehensive multi-omics analysis of plasma proteins and metabolites, and revealed the molecular mechanism associated with the treatment of the patients. In conclusion, our results show that treating PD patients with CMAs leads to enhanced cognitive function, as recently reported in AD patients.

INTRODUCTION

Parkinson's disease (PD) is characterized by selective degeneration of dopaminergic neurons in the substantia nigra and the presence of fibrillar aggregates, which manifest in motor and non-motor features[1]. The prevalence of PD has surpassed that of Alzheimer's disease (AD) and many other neurodegenerative diseases[2,3]. Although most PD studies continue to focus on motor endpoints, PD is also being recognized for its complex range of non-motor symptoms[4], including cognitive impairment, which exist even in the prodromal stages of the disease. Indeed, growing data indicate that metabolic disorders associated with bioenergy failure of nerve cells might increase the risk of developing PD and lead to a higher degree of cognitive impairment and dementia[5]. Moreover, there is considerable evidence for the association between impaired glucose metabolism and PD[6–8], consistent with a predilection to cortical anaerobic glycolysis in patients with PD[7,9]. Growing evidence shows clinical benefits of metabolic treatments in patients with PD (e.g. reduced risk of PD in patients with diabetes using antidiabetics)[10,11], including some improvement in cognitive decline associated with PD[4].

To date, PD treatments are supportive and only provide symptom control; hence, there are still no curative treatments for PD that either inhibit or reverse the neurodegeneration. There is a need for new therapeutic agents exploiting newly defined mechanisms, such as brain energy metabolism in PD, to overcome these translational failures. As reported in many neurodegenerative diseases, several lines of evidence have implicated bioenergy deficiency as a critical element in the pathogenesis of PD[12]. This is closely associated with mitochondrial failure and increased oxidative stress that eventually leads to neurodegeneration[13,14]. Relatedly, reduction of mitochondrial activity and downregulation of target genes involved in mitochondrial biogenesis has already been reported in patients with PD[15]. Furthermore, numerous reports have highlighted specific defects in mitochondrial, electron transport system genes, and protein components[16]. Previous multi-omics studies in humans have shown reduced chaperone proteins[17] and dysregulated genes associated with mitochondrial energy metabolism such as complex I–IV, ATP synthase, and cytochrome C oxidase in PD[18,19]. Additionally, a very recent study evaluating the transcriptomic and proteomic levels of patients with PD suggested that PD should be considered as a disturbance of complex biologic systems[20].

We hypothesized that PD patients could be treated with combined metabolic cofactor activators (CMA) consisting of L-carnitine tartrate (to facilitate mitochondrial fatty acid uptake from the cytosol), nicotinamide riboside (the NAD^+ precursor to promote neuronal mitochondrial β-oxidation and facilitate fatty acid transfer through the mitochondrial membrane), and the potent glutathione precursors L-serine and N-acetyl-l-cysteine (to relieve oxidative stress)[21–23]. We further hypothesized that supplementation with these metabolic cofactors would activate mitochondria and improve global metabolism in the brain. In animal toxicology studies and a human calibration study for CMA, we found that metabolic cofactors were well tolerated and increased the plasma levels of cofactors and their associated metabolites[24]. Additionally, supplementation with CMA effectively increased fatty acid oxidation and de novo glutathione generation, as judged using metabolomic and proteomic profiling[21]. In this placebo-controlled phase 2 study, we tested our hypotheses and the efficacy and safety of CMA in patients with PD.

RESULTS

CMA Improves Cognition and Blood Parameters in PD Patients

In the double-blind, randomized, placebo-controlled phase-2 study, we screened 65 PD patients and recruited 48 patients. We included patients older than 40 years with mild-to-moderate PD according to Hoehn Yahr scale 2 to 4. Of the 48 AD patients, 32 were randomly assigned to the CMA group and 16 to the placebo group (**Figure 1A**, Dataset S1). Five patients dropped out of the study before Day 84 visit during the COVID-19 lockdown. On days 0, 28 and 84, we assessed the clinical variables and analyzed the differences between day 0 and day 28 and between day 0 and day 84 in the CMA and placebo groups (Dataset S2).

The mean age of the participants was 69.7 years (41–84 years), and 83.3% were men (Dataset S1). With regards to safety, no severe adverse events occurred, and five patients in the CMA group reported mild adverse events. All patients agreed to complete the study. The safety profile of CMA in these patients was consistent with the results of our previous one-day calibration study[24] and clinical trials[23,25], including only a single component of CMA. Our present study showed that CMA was safe and well-tolerated in patients with PD.

We measured clinical variables in all patients and analyzed the differences before and after CMA administration in the active and placebo groups (**Figure 1B** and **1C**, Dataset S1 and S2). We observed that the mean Montreal Cognitive Assessment (MoCA) scores were significantly higher in the CMA group both on Day 28 vs Day 0 (log2FoldChange(FC)=0.17, (13% improvement), p=0.001, higher score indicates better cognitive function) and on Day 84 vs Day 0 (log2FC=0.27, (21% improvement), p= 0.0001). We also observed significant increased on MoCA scores in the placebo group on Day 28 vs Day 0 (log2FC=0.16, (12% improvement), p=0.001) and on Day 84 vs day 0 (log2FC=0.15, (11% improvement), p=0.04) due to the recommendations of exercise and Mediterranean diet to all PD patients in the trial. Notably, the degree of increase of MoCA was much higher on Day 84 vs Day 0 in the CMA group than in the placebo group, suggesting the PD patients benefitted from CMA treatment after 84 days of treatment.

However, between the CMA and placebo arms of the study, there was no significant difference in the mean change in MoCA as a result of CMA. The mean change in MoCA among participants receiving CMA was an increase of 3.1 points, whereas the mean change in the placebo arm was 1.7 points. The mean benefit afforded by CMA was 1.3 points ($p = 0.20$, two-tailed t-test). The trial therefore did not reach its primary endpoint and the CMA and placebo arms were considered separately henceforth.

We also analyzed the differences of clinical parameters by stratifying the patients into high- and low-scoring MoCA groups (≥ 15 MoCA score is high, <15 is low). Interestingly, we observed a significant improvement only in the low-scoring patients in the CMA group both on Day 28 (log2FC=0.25, (19% improvement), p=0.003) and Day 84 (log2FC=0.32, (25% improvement), p=0.003), but no significance (p>0.05) was found in the low-scoring placebo group (**Figure 1B** and **1C**, Dataset S2). On the other hand, MoCA scores were significantly different in high-scoring patients in the CMA (log2FC=0.22, (17% improvement), p=0.015) and placebo (log2FC=0.13, (9% improvement), p=0.04) groups Day 84 vs Day 0. These results suggest that both moderate patients (low-scoring patients) and mild patients (high-scoring patients) have a better response to CMA administration at Day 84 vs Day 0 compared to placebo.

Analysis of secondary outcome variables on Day 84 vs Day 0 showed that serum aspartate aminotransferase (AST) (log2FC= −0.23, p=0.03), total bilirubin (log2FC= −0.27, p=0.008) and triglycerides (log2FC= −0.17, p=0.04) levels were significantly

lower in the CMA group (**Figure 1B**, Dataset S2). We also found significantly increased high-density lipoprotein (HDL) (log2FC= −0.07, p=0.04) only in the CMA group (**Figure 1B**, Dataset S2). In summary, we observed that the level of AST, total bilirubin, HDL and triglycerides were significantly improved due to the administration of CMA in PD patients, consistent with the previously performed NAFLD phase 2 clinical trial with CMA[26].

We also measured the level of complete blood count parameters and found that their levels were significantly changed in the CMA group (**Figure 1B**, Dataset S2). We found that the levels of white blood cells (log2FC= −0.14, p=0.01) as well as the absolute number of basophils (log2FC= −0.25, p=0.04) were significantly lower in the CMA group on Day 84 vs Day 0. Of note, a decrease in the inflammatory parameters was also significant in the MoCA low-scoring patients on Day 84 (**Figure 1B**, Dataset S2).

Response to CMA Is Affected By Patients' Clinical Profile

Variability in treatment response and clinical profiles among patients with PD is well documented[27]. As expected, we also observed this variability in our cohort. To harness this heterogeneity, we hypothesized that there exists some subset of patients, defined by clinical parameters, who would respond to CMA better than other patients.

We first determined whether alanine transferase (ALT), a marker for liver health status, could predict response to CMA. In order to do this, we stratified patients receiving CMA or placebo into high- and low ALT groups and recorded MoCA with visit time (**Figure 2A**). We found that only the low ALT group exhibited increased MoCA score, but only when given CMA. No other group showed improvement to the same or better significance.

We then conducted the same stratification for each of the other clinical measures to determine other conditions in which CMA treatment leads to the best response (**Figure 2B**). In addition to the aforementioned low ALT, we found that low ALP, high AST, low GGT, low HCT, low glucose, low HbA1c, low uric acid, low eosinophil count, and low MoCA could indicate better response to CMA. This would indicate that CMA might work better in patients with underlying blood and glucose conditions.

CMA Increases the Plasma Levels of Metabolites Associated with Metabolic Activators

We first analyzed the plasma levels of serine, carnitine, NR, and cysteine and found that administration of the CMA increased the plasma levels of serine, carnitine and nicotinamide proportionally on Day 84 vs Day 0 in the CMA group (**Figure 3A**, Dataset S3 & S4). In detail, the plasma levels of nicotinamide, nicotinurate, 1-methylnicotinamide, and N1-methyl-2-pyridone-5-carboxamide (associated with NR and NAD^+ metabolism); of serine, glycine and betaine (associated with serine and glycine metabolism); and of deoxycarnitine and carnitine (associated with carnitine metabolism) were significantly higher in the CMA group on Day 84.

Effect of CMA on Global Metabolism

We identified those plasma metabolites that were significantly (adj.p<0.05) different on Day 84 vs Day 0. We found that the plasma levels of 75 metabolites were significantly different in the CMA group (**Figure 3**, Dataset S4). Evaluation of plasma metabolites that differed significantly on Day 84 vs Day 0 in each group showed that the majority of metabolites related to amino acid (n=37) or lipid metabolism (n=16) and other metabolic pathways (n=22) were altered in the CMA group compared to the placebo group (**Figure 3**, Dataset S4).

Creatine is one of the most prevalent central nervous system (CNS) metabolites, and reduced levels have been associated with brain tissue injury[28]. Prior research has also shown that creatine is a central metabolite to maintain energy metabolism in brain[29,30]. In our study, we observed that plasma levels of creatine significantly increased on Day 84 vs Day 0 in the CMA group (**Figure 3B**, Dataset S4). Also upregulated on Day 84 vs Day 0 in the CMA group (**Figure 3B**, Dataset S4) was the metabolite glycine, which has been extensively investigated for its positive impact on cognitive performance[31,32].

Elevated plasma homocysteine levels are known found to be associated with PD[33] and several in-vivo studies previously suggested the beneficial results of a low methionine diet on neurodegenerative diseases[33]. In our clinical trial, plasma levels of S-adenosylhomocysteine as well as 2,3-dihydroxy-5-methylthio-4-pentenoate and N-acetyl taurine were significantly decreased on Day 84 vs Day 0 in the CMA group (**Figure 3B**, Dataset S4).

Higher plasma concentrations of kynurenine pathway metabolites were related to CNS disorders[34]. In our study, we found that kynurenate, indolepropionate, 8-methoxy kynurenate and tryptophan betaine were significantly decreased on Day 84 vs Day 0 in

the CMA group (**Figure 3B**, Dataset S4). Kynurenate is the product of tryptophan metabolism and is well known for its oxidative stress inducing effects by generating superoxide radicals and leading to cytochrome C depletion. According to the previous studies, high levels of kynurenine cause cell death in natural killer cells and decrease blood pressure in the systemic inflammatory response through reactive oxygen species[35,36].

Accumulating evidence suggest an association between kidney and brain disorders, but the causal relationship between renal function and cognitive impairment remains to be established[37]. Recent studies showed that plasma levels of N,N,N-trimethyl-5-aminovalerate involved in lysine metabolism is an indicator of elevated urinary albumin excretion[38]. Here, we found that the plasma level of N,N,N-trimethyl-5-aminovalerate was significantly decreased on Day 84 vs Day 0 in the CMA group (**Figure 3B**, Dataset S4).

Moreover, the plasma level of creatinine was also significantly decreased on Day 84 vs Day 0 in the CMA group (**Figure 3B**, Dataset S4). Additionally, our analysis revealed reduced levels of several metabolites related to histidine metabolism in the CMA group on Day 84 vs Day 0. Among those N-acetyl-1-methylhistidine is related to decreased renal function.

Also, we found that plasma levels of metabolites related to the urea cycle (3-amino-2-piperidone, pro-hydroxy-pro, trimethyl-alanylproline betaine and homoarginine) were significantly lower in the CMA group on Day 84 vs Day 0 (**Figure 3B**, Dataset S4).

Lipids are central players in the pathogenesis of neurodegenerative diseases, including PD. Sphingolipids and cholesterol are not only structural components of plasma membranes but are also recognised as important components of brain function[39]. In our study, plasma levels of a considerable number of metabolites associated with carnitine and fatty acid metabolism were significantly elevated on Day 84 vs Day 0 in the CMA group (**Figure 3C**, Dataset S4). Of note, plasma levels of pregnenolone steroids and dihydroxy fatty acids were significantly decreased on Day 84 vs Day 0 (**Figure 3C**, Dataset S4).

Our additional analysis showed significantly upregulated carnitine metabolites, and significantly decreased plasma bilirubin metabolites, such as biliverdin (**Figure 3D**, Dataset S4). It has been shown to exert neuroprotective and procognitive effects in

animal studies; this is especially appropriate given the altered levels found in our study. Carnitine, and NAD^+ metabolites were shown to be involved in restoring the mitochondrial function, oxidative status and synaptogenesis[40]; while, in contrast, increased bilirubin and related metabolites led to neurotoxic effects in many PD models.

Despite its concentration-dependent dual antioxidant role[41], studies on bilirubin in patients with PD have found increased levels of bilirubin[42-44], possibly linked with the increased oxidative stress enzyme activity in patients with PD[44]. In support of this, there is evidence of increased hemoxygenase activity of dopaminergic cells after oxidative stress[45,46], an enzyme responsible for the production of biliverdin[47].

Effect of CMA on Plasma Proteins

Plasma levels of 1466 protein markers were measured with the plasma proteome profiling platform Proximity Extension Assay quantifying the plasma level of target proteins. After quality control and exclusion of proteins with missing values in more than 50% of samples, 1463 proteins were analyzed (Dataset S5&S6). Proteins whose levels differed significantly between the visits in the CMA and placebo groups are listed in Dataset S6.

We analyzed the effect of CMA on plasma protein profile and found that the levels of 20 proteins were significantly ($p < 0.01$) changed in the CMA group on Day 84 vs Day 0. Thirteen of these proteins were significantly decreased, whereas seven of these proteins were significantly increased on Day 84 vs Day 0. Among these proteins, we found that the plasma levels of OSM, MMP9, RASSF2, GSTP1, GZMH, FEN1, NCF2, MNDA, AK1, AZU1, AARSD1 and RAPGAP1L were significantly downregulated. The plasma levels of KLB, GPA33, SLC39A14, IL17RB, LRIG1, ALPP, and SERPINB5 were significantly upregulated only in the CMA group (**Figure 4A** & Dataset S6). We observed that IL1B, CXCL6, TPT1, PIK3AP, ARHGAP1, CXCL11, PPME1, AKT1S1 and RILP were significantly ($p<0.01$) downregulated and KLB upregulated in the placebo group (**Figure 4A**, Dataset S6). Several experimental studies suggest that the altered proteins found in our study are involved in inflammation, membrane transport, DNA repair, membrane trafficking, synaptogenesis, oxidative injury, and protein aggregation. For instance, ALPP and LRIG1 are well-known molecules for their antioxidant and neurotrophic properties among the increased proteins. Similarly, SLC39A14 functions as a pivotal manganese transporter in vertebrates[48] and its

deficiency is associated with rapidly progressive childhood-onset parkinsonism–dystonia due to excessive accumulation of manganese in the brain. Also, GPA33, a protein strictly limited to the intestine and responsible for intestinal integrity with unknown central functions[49], was found to be increased, which might suggest the role of a gut-brain axis component in neurodegenerative disorders.

Similar relevant alterations were also observed for decreased protein levels. For instance, AK1, and ATP regulator protein, has been defined in postmortem PD brains as upregulated, indicating energy dysregulation in PD[50], which is especially relevant considering the increasing evidence of a strong link between protein aggregation, inflammation and energy deficiency in PD. We consistently found significantly reduced levels of MMP9, RASSF2, GSTP1, GNZMH, NCF-2, AARSD1, MNDA, OSM, and FEN1 which are proteins involved in neuroinflammation, apoptosis, oxidative stress, central and even peripheral immunologic responses.

Other notable observations include the down-regulated levels of OSM, MNDA, AZU1 and MMP9, which are well known neuroinflammatory markers, proven in several PD models. Similar beneficial alterations have also been observed for some other proteins such as RABGAP1L and single tRNA synthetase editing domain, involved directly or indirectly in misfolded protein aggregation. It should be noted that we observed changes in some other molecules that are also consistent with human data, showing either a significant improving effect on the neurodegenerative process of being involved in the pathogenetic process. For instance, AZU1 and MMP-9, both important cascades of a multifunctional neuroinflammatory process and blood-brain barrier breakdown, as mentioned above, were elevated in individuals with PD[51]. Also, dysregulated levels of RABGAP1L, involved in cellular membrane trafficking, and GSTP1, a well-known molecule with attenuating functions on oxidative and endoplasmic reticulum stress[52,53], have been found in human dopaminergic neurons and the synaptosomal fraction of patients with PD[54]. In addition, LRIG1, was increased in the present study, and was recently shown to be located in the soma and extends out into the apical dendrites of hippocampal pyramidal neurons, controlling brain-derived neurotrophic factor signalling[55], which is a neuroprotective and pro-cognitive molecule.

Integrative Multi-Omics Analysis

The integrations of multi-omics data have been previously shown to be beneficial in understanding diseases[56]. We generated a PD-specific network based on multi-omics (metabolomics and proteomics) data, complemented by clinical chemistry and anthropometrics data, generated in this study. The main goal of the network analysis was to elucidate the functional relationships between analytes within and between different omics and data types. The network was generated using the same pipeline as iNetModels[57], an interactive multi-omics network database and visualization tool, where we deposited the full network from this study. The generated network has ~2 million edges from 2295 nodes (40% network density, Dataset S7).

To understand the CMA and MoCA interactions, we extracted a subnetwork of those analytes and their top neighbours (**Figure 4B**). From the subnetwork, we observed that MoCA was associated with the plasma levels of serine, trimethylamine N-oxide (phospholipid), deoxycarnitine, creatine, and several nicotinate and nicotinamide metabolites (1-methylnicotinamide, nicotinurate, nicotinate ribonucleoside, and N1-methyl-2-pyridone-5-carboxamide). MoCA was also associated with the plasma levels of several proteins, including MMP9 and OSM, highlighted in the previous section. The same metabolites and proteins were shown to be positively correlated to the CMA.

We also performed a centrality analysis to identify the critical nodes in the network. The top 10 most central metabolites were dominated by xenobiotics metabolites, including those associated with neurological and psychoactive drugs (lamotrigine, O-desmethylvenlafaxine, venlafaxine, and diazepam). We also observe a nicotinamide-related metabolite (Adenosine diphosphate (ADP)-ribose) in that list. Meanwhile, the top 10 proteins included nervous system-related protein (NPY), a regulator of the TGF-beta pathway (ITGB6), and a T-cell activation regulator (PRKAR1A). These results showed that integrative multi-omics network analysis may elucidate the functional relationships between analytes, support the results from single omics data and add new insights by enabling the discovery of key analytes from the network.

Common Effect of CMA on Plasma Metabolites and Proteins in AD And PD Patients

We validated our findings in an independent AD cohort (n=60) who received the same protocol of CMA therapy and showed a significant improvement in cognitive functions[25]. Comparison of metabolomics data in the treated groups of both AD and PD cohorts than their placebo on Day 84 vs Day 0 revealed that the plasma levels of 60

metabolites were significantly altered in the same direction (**Figure 5**, Dataset S8). These changes were in amino acid (n=30), lipid metabolism (n=14) and other metabolic pathways (n=16) (**Figure 5**, Dataset S8).

Primarly, the plasma levels of CMA constituents and byproducts were increased in the CMA groups of both cohorts (**Figure 5A**, Dataset S8). Additionally, hypotaurine and trimethylamine N-oxide were also significantly increased in both CMA groups (**Figure 5A**, Dataset S8). On the other hand, plasma levels of numerous amino acids in key pathways (i.e glutamate, glutathione, histidine and tryptophan metabolism and urea cycle) were significantly decreased in the both CMA groups (**Figure 5A**, Dataset S8). Also, creatinine, betaine, N,N,N-trimethyl-5-aminovalerate and phenol sulfate plasma levels were significantly downregulated in CMA groups than placebo (**Figure 5A** Dataset S8). Furthermore, plasma levels of metabolites associated with fatty acids, choline and pregnenolone steroids were significantly decreased in the CMA groups of both cohorts (**Figure 5B**, Dataset S8). Interestingly, plasma levels of 2-O-methylascorbic acid and metabolites in purine metabolism were also significantly decreased in the both CMA groups (**Figure 5C**, Dataset S8).

We also compared the plasma proteomic profile between two cohorts and found that a significant increse in the plasma level of KLB and a significant decrease in the plasma level of OSM in both on Day 84 vs Day 0 in the CMA groups (Dataset S8). Our results collectively indicated that CMA improved cognitive functions in different patients groups by similar alterations on plasma profiles.

DISCUSSION

In our study, we observed that CMA treatment significantly improved cognitive function in patients with PD based on MoCA scores. We found that the cognitive functions in PD patients was improved 21% in the CMA group, whereas it was improved only 11% in the placebo group after 84 days of CMA administration. Our findings were in agreement with the results of our recently performed AD clinical phase 2 study, where we administed CMA to 60 AD patients using the same protocol **25**. In the AD study, we assessed the cognitive functions with AD Assessment Scale-cognitive subscale (ADAS-Cog) score and found that the ADAS-Cog scores were improved 29% in the CMA group whereas they were improved only 14% in the placebo group after 84 days of CMA administration. The 11-14% improvement in the placebo group in both PD and AD clinical trials can be explained with the recommendations of exercise and Mediterranean diet to all participated patients.

In evaluating the metabolomic parameters, we found that levels of plasma nicotinamide, carnitine, cysteine, and serine were significantly elevated. Considering the role of these metabolites in the pathogenesis of PD, it was not surprising to see such beneficial alterations after CMA treatment. For instance, NAD^+ has been shown to restore bioenergy imbalance in animal PD models and patients with PD, offering significant alleviation of motor symptoms correlated with the increase of L-dopa availability**58**,**59**. A similar situation also applies to other metabolites such as serine, cysteine, and carnitine.

Beyond their critical role in mitochondrial energy deficiency confirmed by many PD models, there are a growing number of studies suggesting that these metabolites are also involved in patients with PD. For instance, among many metabolites, plasma serine and carnitine levels have been shown to have the strongest inverse correlation with PD severity**60**. These clinical data accorded well with their beneficial role in oxidative injury**61** (increasing the synthesis of glutathione) and impaired mitochondrial bioenergetics in PD**62**. Also, a downregulation of carnitine and its metabolites has been shown in patients with PD**63**. As for NAD^+, recent human data suggest that it is also linked to improved behavioural and motor symptoms in patients with PD**64**, possibly mediated by the glutamatergic role of serine**65** when applied either separately or in combination.

Several animal and human studies provided strong evidence for countering some of the effects of disturbed carnitine metabolism with acetyl-L-carnitine**66**,**67**. Here, it is worth mentioning that acetyl-L-carnitine treatment seems to induce both cerebral energy metabolism and cholinergic neurotransmission**68**, which may have also led to improved cognition in our patients with PD. These findings suggest that in addition to increased carnitine levels, decreased choline metabolites may also have mediated the pro-cognitive effect of our treatment. Thus, our finding of improved lipid metabolites is of critical significance, as altered lipid metabolism and cognitive dysfunction are both known to occur during the neurodegenerative process.

Similar beneficial alterations were also observed in dihydroxybutyrate (DHBA) levels. Our finding of decreased DHBA levels might indicate the restoration of alternative energy production pathways, such as the GABA shunt**69**, activated during the cellular energy composition, as in AD**69** and PD**70**. Confirming this, in our recent AD study, we found similarly decreased post-therapeutic levels of DHBA**25**, suggesting an energy deficiency seen in both neurodegenerative diseases, but also the pro-energy role of CMA. Beyond their role in energy mechanisms, lipids are involved in many critical intracellular signaling and transporting processes, as the main component of cellular membranes, which make them a strong candidate for cognition, even in healthy individuals**71**. However, under neurodegenerative conditions characterized by disturbed lipid metabolism, their behavior shifts to become more pro-inflammatory and oxidative, rather than regulatory, and contributes significantly to the acceleration of the neurodegenerative process, as in PD**72**. An example is the link between lipid dysregulation and alpha-synuclein aggregation, leading to dopaminergic neuronal death in PD**72**.

Our analysis also revealed decreased purine metabolism in the CMA group. Plasma levels of hypoxanthine and adenosine were significantly reduced in the CMA group, consistent with a recent clinical study showing increased hypoxanthine levels in PD**73**. Furthermore, our findings of decreased peripheral levels of hypoxanthine could also be interpreted as increased CNS bioavailability of hypoxanthine considering its high blood-brain barrier penetrance and adenosine triphosphate (ATP)-enhancing role under energy crisis conditions.

Considering that creatine is an essential part of mitochondrial therapy for many neurodegenerative diseases**74** and exerts significant neuroprotection in several in-vitro

and in-vivo PD studies**75**, our findings of increased creatine levels were not surprising to us. Consistently, creatine supplementation improves cognition not only in healthy individuals**75**, but also in PD patients with cognitive impairment when combined with another mitochondrial energetic molecule, coenzyme q10[76]. Also, its combination with minocycline significantly is known to reduce PD progression in early stages**77**.

Human data have shown that creatine exerted its pro-cognitive effect through increasing the brain oxygen utilization**78** fitting well with its pro-energetic role. We also have found decreased creatinine levels, compatible with recent PD clinical data showing an inverse correlation between uric acid/creatinine ratios and PD progression**79**. Similarly, increased serum creatinine was related to incident dementia and cognitive impairment**80**.

Moreover, we found that the plasma level of 3-amino-2-piperidone associated with urea, especially in ornithine metabolism, was significantly decreased after CMA. Although it remains unclear how decreased urea metabolism is associated with our treatment effect, several studies have indicated that urea pathways might be perturbed in PD**81**–**83**, indicating increased ornithine levels in PD. This is consistent with the detrimental role of high urea on learning**84**.

We found decreased levels of homostachydrine, categorized under the xenobiotics class, which are generally not endogenously synthesized but can be produced by the gut microbiome**85**. Although no biological role has been defined for homostachydrine, a recent study reported that it was involved in the increased inflammatory process in an experimental autoimmune encephalomyelitis model in mice**86**. This observation agrees well with the identified decreased levels of homostachydrine after CMA treatment.

It is worth mentioning here that hypotaurine metabolism plays a vital role in fighting against neurodegeneration through the production of secondary bile acids, which exert a significant neuroprotective effect**87**. A recent study identified increased hypotaurine metabolism as a compensatory neuroprotective pathway in a mouse model with alpha-synuclein**88**. Also, N-stearyl serine, a lipoamino acid, with many pleiotropic signalling functions**89**, was recently shown to be positively associated with cognition in patients with AD, especially those with good adherence to a Mediterranean diet**90**. Thus, stearoyl-CoA desaturase, a rate-limiting enzyme in the biosynthesis of monounsaturated fatty acids, has been suggested as a new player mediating lipid metabolism with synuclein aggregation**90**.

It is difficult to explain why we observed only an improvement in cognitive scores but no improvement in Unified Parkinson Disease Rating Scale (UPDRS) motor scores; despite a robust cognitive response, we observed no significant alteration in motor scores (Dataset S2). This is, however, in line with previous studies showing that energy metabolic supplementations failed to show a clear benefit in PD**91**, suggesting that disturbed energy metabolism cannot be easily reversed in already dead circuits. Early stages of PD are likely to be responsive to such metabolic approaches. For instance, besides some critical differences in affected regions between patients with PD-with and without cognitive impairment**91**,**92**-the metabolic expression of both clinical patterns also increases at different rates. PD studies have suggested that cognitive expression has a slower metabolic deterioration rate than the pure motor pattern**93**, which would indicate probable later neurodegeneration, and explain the selective therapeutic response observed in our study.

Also, our improved MoCA scores, along with increased post-therapeutic alterations in lipid metabolites, fits well with recent metabolomics PD studies showing an inverse correlation between lipid metabolites and MoCA scores**94**, suggesting the role of lipid metabolite alterations is a key parameter in discriminating PD with cognitive impairment from regular PD patients.

A few limitations of the study need to be considered. First, we evaluated the treatment effect using only omics-based methods and clinical evaluation without neuroimaging. Thus, a clinical trial combined with neuroimaging methods to delineate the effects of CMA on functional and structural brain alterations would be informative. Second, the link between systemic and CNS alterations and their relations to the central protein aggregations defined as critical in PD has not been evaluated.

Further, when comparing outcomes between the two arms of the study as required by CONSORT guidelines, the mean benefit of 1.3 MoCA points afforded by CMA did not reach statistical significance. The trial did not meet its primary endpoint. The trial conformed to all other CONSORT guidelines and the overall trend of MoCA improvement, although not statistically significant, could be reconciled in our omics studies.

To date, there is no curative medical treatment for cognitive impairment in PD; therapeutic options are limited to cholinesterase inhibitors**95**, which may be transient to

replace the impaired cholinergic transmission**96**. Also, it is still under debate whether dopamine replacement medications have cognitive side effects**97** while improving motor symptoms in cognitively impaired PD patients. To that end, there is still no available drug able to revert PD cognitive deficits, and, unfortunately, current treatment approaches available for motor impairment are not devoid of cognitive side effects. Considering all of these findings, it is crucial to better delineate the neuropathology underlying cognitive symptoms through clinical and preclinical studies. In conclusion, CMA significantly improved cognition and serum markers of PD after 84 days. These findings suggest that targeting multiple pathways using CMA is a potentially effective therapeutic strategy for PD as previously shown in AD.

MATERIAL AND METHODS

Trial Design and Oversight

Patients for this randomized, double-blinded, placebo-controlled, phase-2 study were recruited at the Faculty of Medicine, Alanya Alaaddin Keykubat University, Antalya, Turkey and Faculty of Medicine, Istanbul Medipol University, Istanbul, Turkey. Written informed consent was obtained from all participants before the initiation of any trial-related procedures. An independent external data-monitoring committee oversaw the safety of the participants and the risk-benefit analysis. The trial was conducted in accordance with Good Clinical Practice guidelines and the principles of the Declaration of Helsinki. The study was approved by the ethics committee of Istanbul Medipol University, Istanbul, Turkey, and retrospectively registered at https://clinicaltrials.gov/ with Clinical Trial ID: NCT04044131.

Participants

Patients were enrolled in the trial if they were over 40 years of age with mild to moderate PD according to Hoehn Yahr scale 2 to 4. Patients who had a history of stroke, severe brain trauma, toxic drug exposure were excluded. Also, patients who indicate to Parkinson-Plus syndrome (i.e., pyramidal, cerebellar and autonomic dysfunction findings and gaze paralysis) in the neurological examination were omitted. The inclusion, exclusion, and randomization criteria are described in detail in the Supplementary Appendix.

Randomization, Interventions, and Follow-up

Patients were randomly assigned to receive CMA or placebo (2:1). Patient information (patient number, date of birth, initials) was entered into the web-based randomization system, and the randomization codes were entered into the electronic case report form. All clinical staff were blinded to treatment, as were the participants.

Treatment started on the day of diagnosis. Both placebo and CMA were provided in powdered form in identical plastic bottles containing a single dose to be dissolved in water and taken orally one dose in the morning after breakfast and one dose in the evening after dinner. Each dosage of CMA dose contained a 3.73 g L-carnitine tartrate, 2.55 g N-acetylcysteine, 1 g nicotinamide riboside chloride, and 12.35 g serine. All

patients returned for a follow-up visit on Day 84. Further information is provided in the Supplementary Appendix.

Outcomes

The primary endpoint in the original protocol was to assess the clinical efficacy of CMA in PD patients. For the primary purpose, the clinical differences in cognition of subjects receiving twelve-week treatment either with metabolic cofactors supplementation or placebo were determined. The primary analysis was on the difference in cognitive and motor function scores between the placebo and the treatment arms. The motor, cognitive and behavioural functions of PD patients were evaluated via Unified Parkinson Disease Rating Scale (UPDRS), The Montreal Cognitive Assessment (MoCA) and Neuropsychiatric Inventory (NPI), respectively. The secondary aim of this study was to evaluate the safety and tolerability of CMA. All protocol amendments were authorized and approved by the sponsor, the institutional review board or independent ethics committee, and the pertinent regulatory authorities.

The number and characteristics of adverse events, serious adverse events, and treatment discontinuation due to CMA were reported from the beginning of the study to the end of the follow-up period as key safety endpoints. The changes in vital signs, baseline values, and the status of treatment were recorded on day 0 and 84. A complete list of the endpoints is provided in the Supplementary Appendix.

Proteomics Analysis

Plasma levels of proteins were determined with the Olink panel (Olink Bioscience, Uppsala, Sweden). Briefly, each sample was incubated with DNA-labeled antibody pairs (proximity probes). When an antibody pair binds to its corresponding antigens, the corresponding DNA tails form an amplicon by proximity extension, which can be quantified by high-throughput, real-time PCR. Probe solution (3 µl) was mixed with 1 µl of sample and incubated overnight at 4°C. Then 96 µl of extension solution containing extension enzyme and PCR reagents for the pre-amplification step was added. The extension products were mixed with detection reagents and primers and loaded on the chip for qPCR analysis with the BioMark HD System (Fluidigm Corporation, South San Francisco, CA). To minimize inter- and intrarun variation, the data were normalized to both an internal control and an interplate control. Normalized data were expressed in arbitrary units (Normalized Protein eXpression, NPX) on a log2 scale and linearized

with the formula 2NPX. A high NPX indicates a high protein concentration. The limit of detection, determined for each of the assays, was defined as three standard deviations above the negative control (background).

Untargeted Metabolomics Analysis

Plasma samples were collected on Days 0 and 84 for nontargeted metabolite profiling by Metabolon (Durham, NC). The samples were prepared with an automated system (MicroLab STAR, Hamilton Company, Reno, NV). For quality control purposes, a recovery standard was added before the first step of the extraction. To remove protein and dissociated small molecules bound to protein or trapped in the precipitated protein matrix, and to recover chemically diverse metabolites, proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) and centrifuged. The resulting extract was divided into four fractions: one each for analysis by ultraperformance liquid chromatography– tandem mass spectroscopy (UPLC-MS/MS) with positive ion-mode electrospray ionization, UPLC-MS/MS with negative ion-mode electrospray ionization, and gas chromatography– mass spectrometry; one fraction was reserved as a backup.

Statistical Analysis

Paired t-tests were used to identify the differences in clinical parameters between time points, and one-way ANOVA was used to find the shifts between CMA and placebo groups at each time point. We removed the metabolite profiles with more than 50% missing values across all samples to analyse plasma metabolomics. Metabolite changes between time points were analyzed by paired t-test. One-way ANOVA analyzed metabolite changes between CMA and placebo groups. Missing values were removed in pairwise comparison. The p-values were adjusted by Benjamini & Hochberg method. Metabolites with a false-discovery rate of 5% were considered statistically significant.

We removed the protein profiles with more than 50% missing values across all samples to analyse plasma proteomics. A paired t-test was used to identify the changes between time points, and one-way ANOVA was used to determine the changes between different groups. $p<0.01$ was considered statistically significant.

For the identification of clinical variables that inform response to CMA, low-scored and high-scored patient groups were established for each clinical variable on the basis of the median score at day 0. Patients scoring equal to or less than the median score were

assigned “low”; patients scoring greater than the median score were assigned “high”. Statistical significance in the difference between MoCA score distributions over visit number was tested between visits using paired t-tests. Clinical variables were deemed informative to predict the response to CMA if exactly one group, low or high, showed more statistically significant changes in MoCA in the CMA group than in the placebo group.

Generation of Multi-Omics Network

A Multi-omics network was generated based on the Spearman correlations, and the significant associations (adj.p < 0.05) are presented. The analyses were performed with the SciPy package in Python 3.7. Centrality analysis on the network was performed using iGraph Python.

ACKNOWLEDGMENTS

This work was financially supported by ScandiBio Therapeutics and Knut and Alice Wallenberg Foundation. The authors would like to thank Metabolon Inc. (Durham, USA) for generating metabolomics data and ChromaDex Inc. (Irvine, CA, USA) for providing NR. AM and HY acknowledge support from the PoLiMeR Innovative Training Network (Marie Skłodowska-Curie Grant Agreement No. 812616), which received funding from the European Union's Horizon 2020 research and innovation programme.

The computations and data handling were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX, partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

CONFLICT OF INTEREST

AM, JB and MU are the founder and shareholders of ScandiBio Therapeutics. The other authors declare no competing interests.

SUPPORTING INFORMATION

Supporting Information includes 8 Supplementary Datasets.

FIGURES

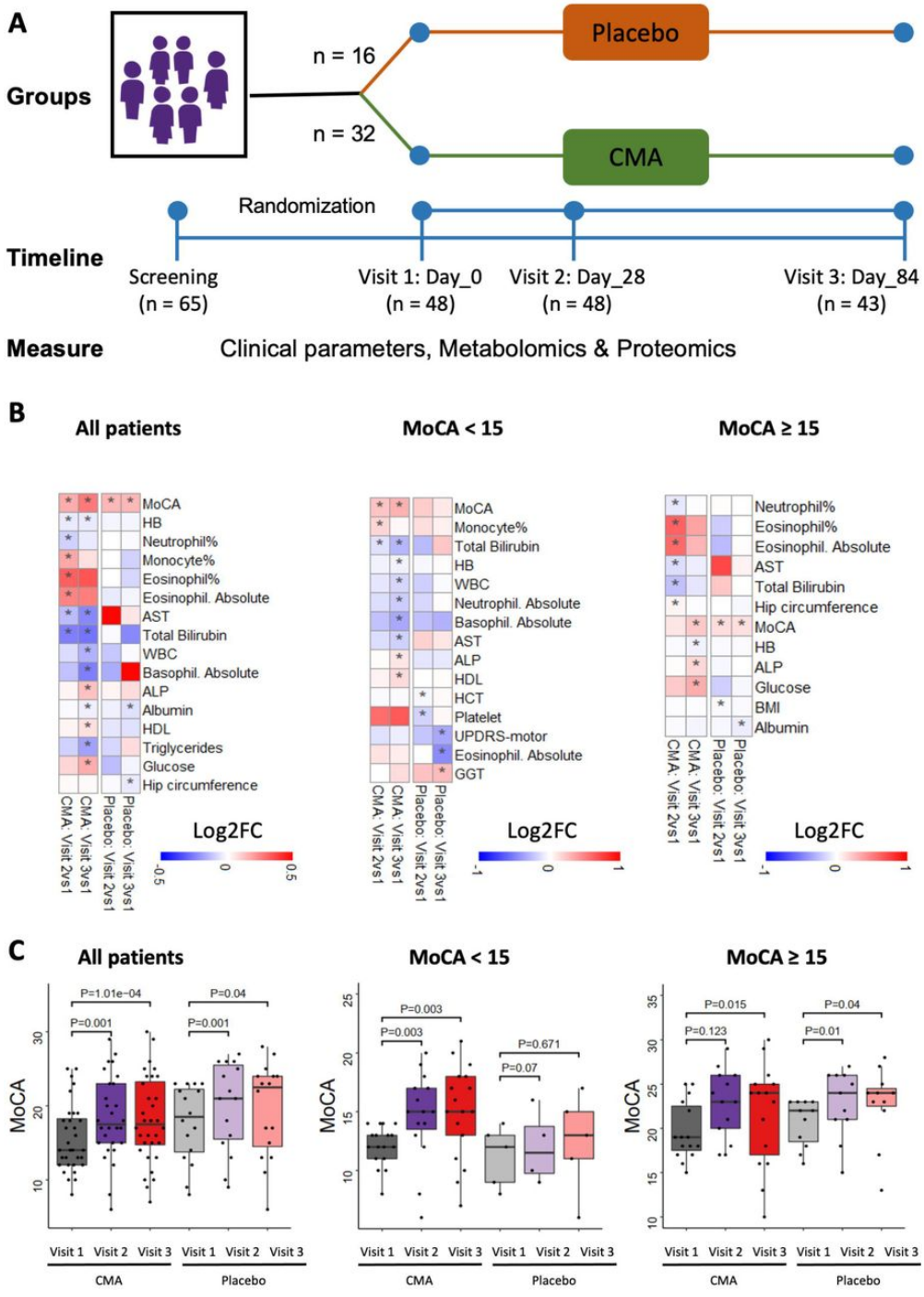


Figure 1. CMA Improves MoCA scores and clinical parameters. A) Study design for testing the effects of CMA in PD patients. B) Differences in MoCA scores in the CMA and placebo groups on Days 0, 28 and 84 are presented. Additionally, MoCA scores

were analyzed by stratifying the patients into high- and low-scored MoCA groups (≥ 15 MoCA score is high, <15 is low). C) Heatmaps shows log2FC based alterations of the clinical variables compared to the CMA administration in both drug and placebo groups. Asterisks indicate statistical significance based on Student's t-test. p-value <0.05. Log2FC: log2(fold change)

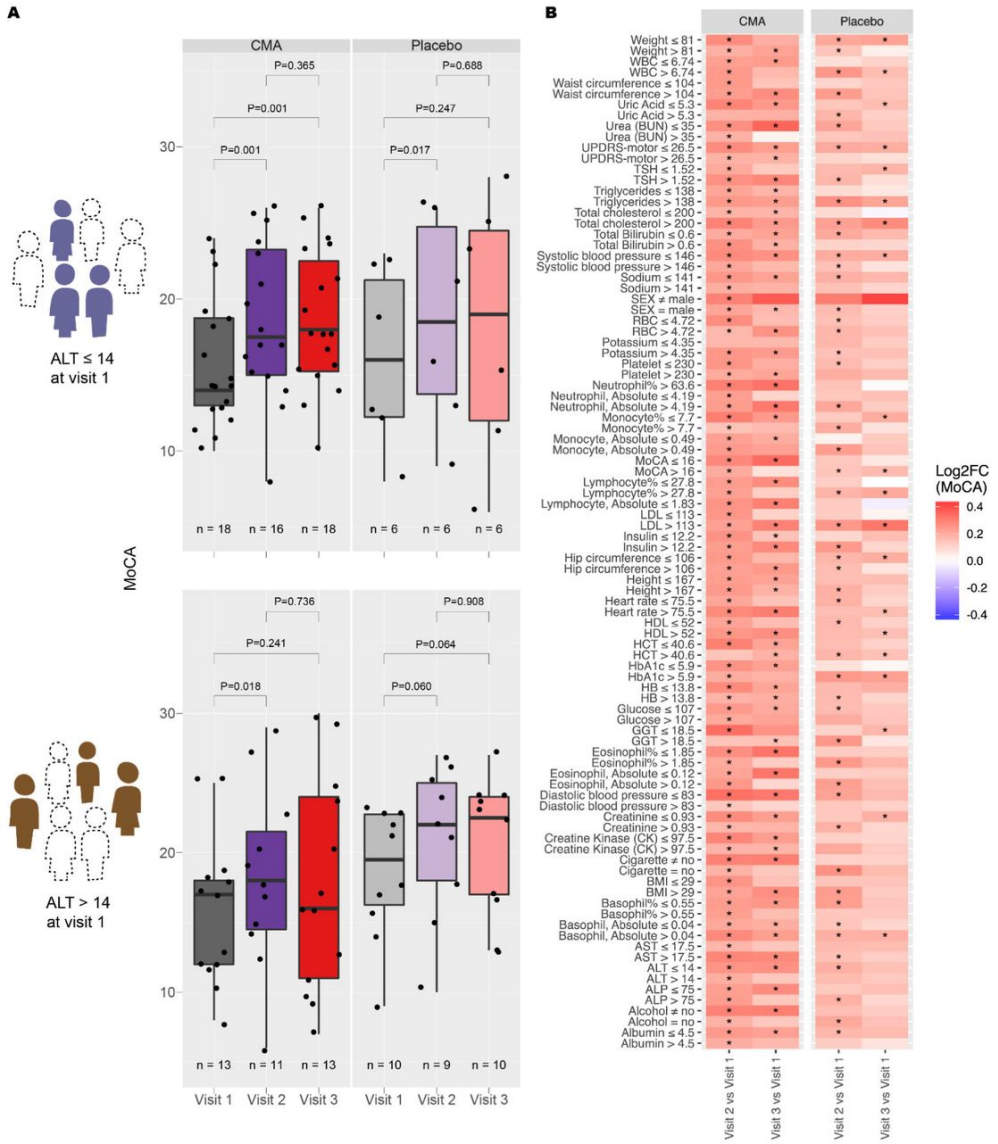


Figure 2. Identification of clinical measures informative for response to CMA. A) Distribution of MoCA scores over visit number for patients with ALT ≤ 14 at visit 1 (upper panel), and patients with ALT > 14 at visit 1 (lower panel). B) Between-visit changes to MoCA by clinical variable grouping. Only those groupings resulting in a more significant change to MoCA in CMA compared to placebo and with a p-value of 0.05 or better are shown. Colour scale indicates log2FC of MoCA between visits. Statistical significance between visits was determined by a paired t-test. *, $p<0.05$.

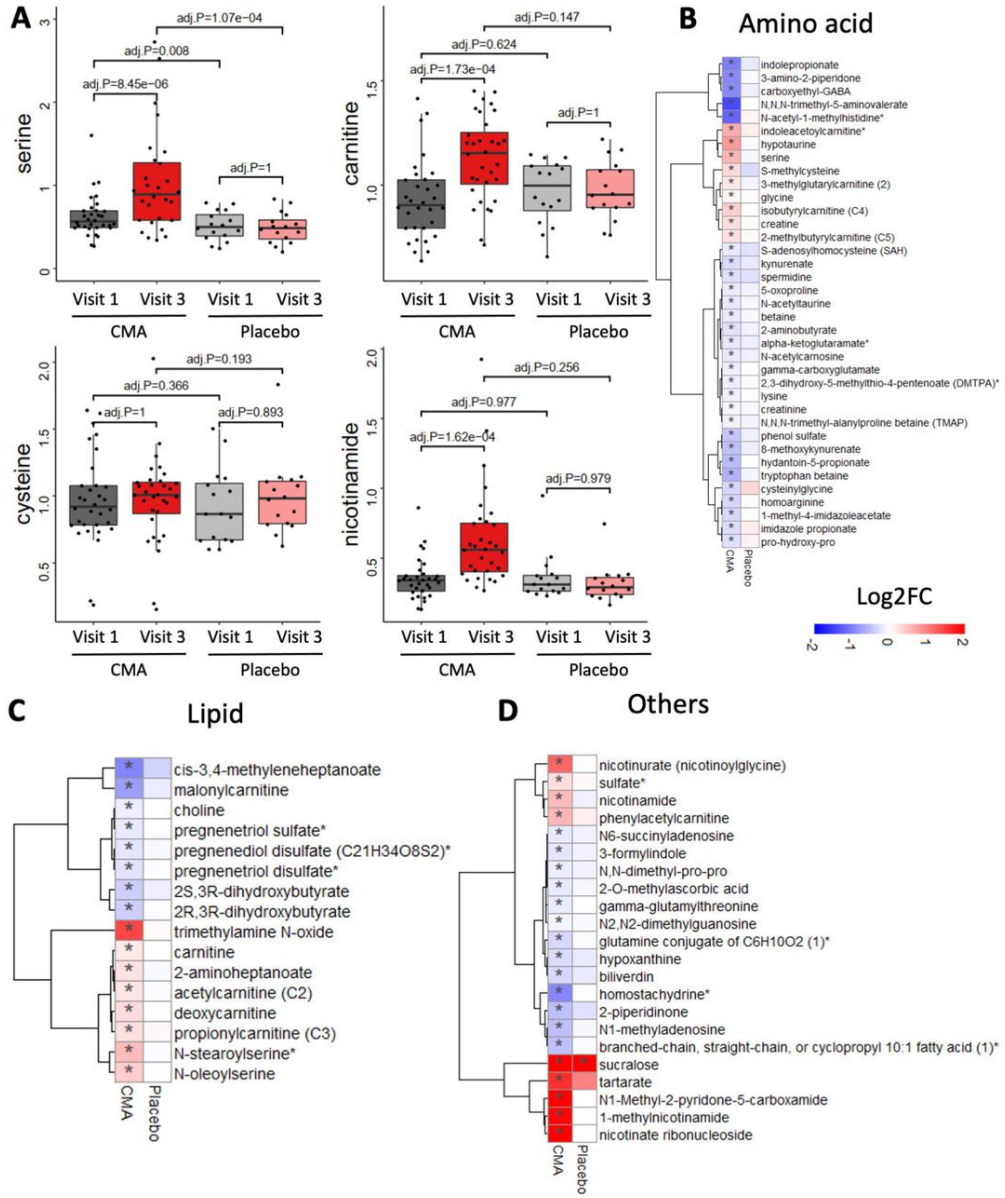


Figure 3. CMA alters plasma metabolite levels. A) Differences in the plasma levels of individual CMA, including serine, carnitine, cysteine and nicotinamide, are shown in the CMA and placebo groups on Days 0 and 84. Plasma levels of B) amino acids, C) lipids and D) other metabolites that are significantly different between Day 84 vs Day 0 in the CMA and placebo groups are presented. Adj. p< 0.05. Heatmap shows log2FC values of metabolites between Day 84 vs Day 0. Asterisks indicate statistical significance based on paired Student's t-test. adj.p< 0.05. Log2FC: log2(fold change).

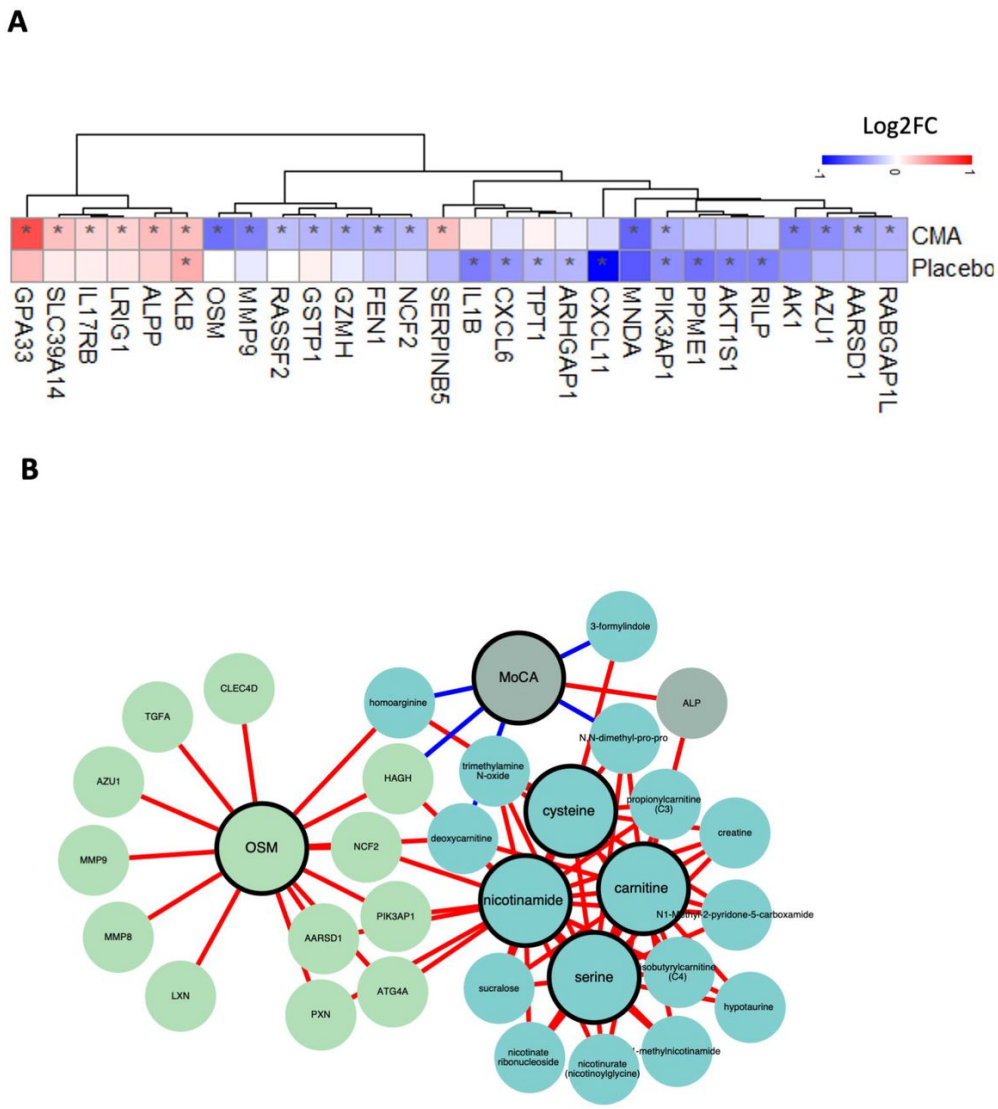


Figure 4. Altered plasma protein levels and integrated multi-omics network

A) Heatmap shows log2FC based alterations between the significantly different proteins on Day 84 vs Day 0 in the CMA and placebo groups. Asterisks indicate statistical significance based on paired Student's t-test. $p < 0.01$. B) Integrated multi-omics data based on network analysis represents the neighbours of the CMA, including serine, carnitine, nicotinamide and cysteine, and MoCA scores. Only analytes that are significantly altered in CMA Day 84 vs Day 0 are highlighted.

Figure 5. Common significantly altered metabolites in PD and AD clinical trials. Plasma levels of A) amino acids, B) lipids and C) other metabolites that are significantly different

between Day 84 vs Day 0 in the CMA group of this study and an independent AD trial25 are presented. Adj. $p < 0.05$. Heatmap shows log2FC values of metabolites between Day 84 vs Day 0. Asterisks indicate statistical significance based on paired Student's t-test. adj.$p < 0.05$. Log2FC: log2(fold change).

SUPPLEMENTARY DATASET LEGENDS

Dataset S1. Collection of samples of CMA and placebo groups and the measured values of clinical indicators before and after treatment.

Dataset S2. Statistical analysis of clinical indicators between different visits or groups.

Dataset S3. Plasma metabolomics data for each patient before and after treatment.

Dataset S4. Statistical analysis of plasma metabolites between different visits or groups.

Dataset S5. Plasma proteomics data was generated with the Olink cardiometabolic, inflammation, neurology and oncology panels for each patient before and after treatment.

Dataset S6. Statistical analysis of plasma proteins between different visits or groups.

Dataset S7. Multi-Omics Network Data, including edges and nodes information. The network is presented in the iNetModels (http://inetmodels.com).

Dataset S8. Common significant metabolites and proteins in this study and in an independent Alzheimer's study[25]

REFERENCES

1. Spillantini, M.G., Crowther, R.A., Jakes, R., Hasegawa, M. & Goedert, M. alpha-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with lewy bodies. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 6469–6473 (1998).
2. Feigin, V.L., et al. Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet Neurology* **16**, 877–897 (2017).
3. Dorsey, E.R. & Bloem, B.R. The Parkinson Pandemic-A Call to Action. *JAMA Neurol* **75**, 9–10 (2018).
4. Chaudhuri, K.R. & Sauerbier, A. Parkinson disease. Unravelling the nonmotor mysteries of Parkinson disease. *Nat Rev Neurol* **12**, 10–11 (2016).
5. Bosco, D., et al. Dementia is associated with insulin resistance in patients with Parkinson's disease. *J Neurol Sci* **315**, 39–43 (2012).
6. Pappatà, S., et al. Mild cognitive impairment in drug-naive patients with PD is associated with cerebral hypometabolism. *Neurology* **77**, 1357–1362 (2011).
7. Camargo Maluf, F., Feder, D. & Alves de Siqueira Carvalho, A. Analysis of the Relationship between Type II Diabetes Mellitus and Parkinson's Disease: A Systematic Review. *Parkinsons Dis* **2019**, 4951379 (2019).
8. Krikorian, R., et al. Nutritional ketosis for mild cognitive impairment in Parkinson's disease: A controlled pilot trial. *Clinical Parkinsonism & Related Disorders* **1**, 41–47 (2019).
9. Bohnen, N.I., et al. Cerebral glucose metabolic features of Parkinson disease and incident dementia: longitudinal study. *J Nucl Med* **52**, 848–855 (2011).
10. Brakedal, B., et al. Glitazone use associated with reduced risk of Parkinson's disease. *Mov Disord* **32**, 1594–1599 (2017).
11. Athauda, D., et al. Exenatide once weekly versus placebo in Parkinson's disease: a randomised, double-blind, placebo-controlled trial. *Lancet* **390**, 1664–1675 (2017).
12. González-Casacuberta, I., Juárez-Flores, D.L., Morén, C. & Garrabou, G. Bioenergetics and Autophagic Imbalance in Patients-Derived Cell Models of Parkinson Disease Supports Systemic Dysfunction in Neurodegeneration. *Front Neurosci* **13**, 894 (2019).
13. Schapira, A.H.V. & Gegg, M. Mitochondrial contribution to Parkinson's disease pathogenesis. *Parkinsons Dis* **2011**, 159160–159160 (2011).
14. Lin, M.T. & Beal, M.F. Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature* **443**, 787–795 (2006).
15. Siddiqui, A., et al. Mitochondrial Quality Control via the PGC1α-TFEB Signaling Pathway Is Compromised by Parkin Q311X Mutation But Independently Restored by Rapamycin. *The Journal of Neuroscience* **35**, 12833 (2015).
16. Saxena, U. Bioenergetics failure in neurodegenerative diseases: back to the future. *Expert Opin Ther Targets* **16**, 351–354 (2012).
17. Caudle, W.M., Bammler, T.K., Lin, Y., Pan, S. & Zhang, J. Using 'omics' to define pathogenesis and biomarkers of Parkinson's disease. *Expert Rev Neurother* **10**, 925–942 (2010).
18. Miller, R.M., et al. Robust dysregulation of gene expression in substantia nigra and striatum in Parkinson's disease. *Neurobiol Dis* **21**, 305–313 (2006).
19. Duke, D.C., et al. Transcriptome analysis reveals link between proteasomal and mitochondrial pathways in Parkinson's disease. *Neurogenetics* **7**, 139–148 (2006).
20. Ruffini, N., Klingenberg, S., Schweiger, S. & Gerber, S. Common Factors in Neurodegeneration: A Meta-Study Revealing Shared Patterns on a Multi-Omics Scale. *Cells* **9**(2020).

21. Mardinoglu, A., et al. An Integrated Understanding of the Rapid Metabolic Benefits of a Carbohydrate-Restricted Diet on Hepatic Steatosis in Humans. *Cell Metab* **27**, 559–571.*e555* (2018).
22. Mardinoglu, A., et al. The Potential Use of Metabolic Cofactors in Treatment of NAFLD. *Nutrients* **11**, 1578 (2019).
23. Altay, O., et al. Combined Metabolic Activators Accelerates Recovery in Mild-to-Moderate COVID-19. *Advanced Science n/a,* 2101222 (2021).
24. Zhang, C., et al. The acute effect of metabolic cofactor supplementation: a potential therapeutic strategy against non-alcoholic fatty liver disease. *Mol Syst Biol* **16**, e9495 (2020).
25. Yulug, B., et al. Combined Metabolic Activators Improves Cognitive Functions in Alzheimer's Disease. *medRxiv,* 2021.2007.2014.21260511 (2021).
26. Zeybel, M., et al. Combined Metabolic Activators Reduces Liver Fat in Nonalcoholic Fatty Liver Disease Patients. *medRxiv,* 2021.2005.2020.21257480 (2021).
27. Greenland, J.C., Williams-Gray, C.H. & Barker, R.A. The clinical heterogeneity of Parkinson's disease and its therapeutic = implications. *Eur J Neurosci* **49**, 328–338 (2019).
28. Béard, E. & Braissant, O. Synthesis and transport of creatine in the CNS: importance for cerebral functions. *Journal of Neurochemistry* **115**, 297–313 (2010).
29. Bonilla, D.A., et al. Metabolic Basis of Creatine in Health and Disease: A Bioinformatics-Assisted Review. *Nutrients* **13**, 1238 (2021).
30. Tachikawa, M., Hosoya, K., Ohtsuki, S. & Terasaki, T. A novel relationship between creatine transport at the blood-brain and blood-retinal barriers, creatine biosynthesis, and its use for brain and retinal energy homeostasis. *Subcell Biochem* **46**, 83–98 (2007).
31. de Bartolomeis, A., et al. Glycine Signaling in the Framework of Dopamine-Glutamate Interaction and Postsynaptic Density. Implications for Treatment-Resistant Schizophrenia. *Front Psychiatry* **11**, 369–369 (2020).
32. Avila, A., Nguyen, L. & Rigo, J.-M. Glycine receptors and brain development. *Frontiers in cellular neuroscience* **7**, 184–184 (2013).
33. Fan, X., et al. Role of homocysteine in the development and progression of Parkinson's disease. *Annals of Clinical and Translational Neurology* **7**, 2332–2338 (2020).
34. Huang, Y.-S., Ogbechi, J., Clanchy, F.I., Williams, R.O. & Stone, T.W. IDO and Kynurenine Metabolites in Peripheral and CNS Disorders. *Frontiers in Immunology* **11**(2020).
35. Darcy, C.J., et al. An observational cohort study of the kynurenine to tryptophan ratio in sepsis: association with impaired immune and microvascular function. *PLoS One* **6**, e21185 (2011).
36. Wang, Q., Liu, D., Song, P. & Zou, M.H. Tryptophan-kynurenine pathway is dysregulated in inflammation, and immune activation. *Front Biosci (Landmark Ed)* **20**, 1116–1143 (2015).
37. Davey, A., Elias, M.F., Robbins, M.A., Seliger, S.L. & Dore, G.A. Decline in renal functioning is associated with longitudinal decline in global cognitive functioning, abstract reasoning and verbal memory. *Nephrology Dialysis Transplantation* **28**, 1810–1819 (2013).
38. Haukka, J.K., et al. Metabolomic Profile Predicts Development of Microalbuminuria in Individuals with Type 1 Diabetes. *Scientific Reports* **8**, 13853 (2018).
39. Kraft, M.L. Sphingolipid Organization in the Plasma Membrane and the Mechanisms That Influence It. *Front Cell Dev Biol* **4**, 154 (2016).
40. Nezhadi, A., Sheibani, V., Esmaeilpour, K., Shabani, M. & Esmaeili-Mahani, S. Neurosteroid allopregnanolone attenuates cognitive dysfunctions in 6-OHDA-induced rat model of Parkinson's disease. *Behav Brain Res* **305**, 258–264 (2016).
41. Peng, F., et al. Low antioxidant status of serum uric acid, bilirubin and albumin in patients with neuromyelitis optica. *Eur J Neurol* **19**, 277–283 (2012).
42. Moccia, M., et al. Increased bilirubin levels in de novo Parkinson's disease. *Eur J Neurol* **22**, 954–959 (2015).
43. Scigliano, G., et al. Increased plasma bilirubin in Parkinson patients on L-dopa: evidence against the free radical hypothesis? Ital *J Neurol Sci* **18**, 69–72 (1997).

44. Macías-García, D., et al. Increased bilirubin levels in Parkinson's disease. *Parkinsonism Relat Disord* **63**, 213–216 (2019).
45. Yoo, M.S., et al. Oxidative stress regulated genes in nigral dopaminergic neuronal cells: correlation with the known pathology in Parkinson's disease. *Brain Res Mol Brain Res* **110**, 76–84 (2003).
46. Schipper, H.M., Song, W., Zukor, H., Hascalovici, J.R. & Zeligman, D. Heme oxygenase-1 and neurodegeneration: expanding frontiers of engagement. *J Neurochem* **110**, 469–485 (2009).
47. Yamamoto, N., et al. Elevation of heme oxygenase-1 by proteasome inhibition affords dopaminergic neuroprotection. *J Neurosci Res* **88**, 1934–1942 (2010).
48. Tuschl, K., et al. Mutations in SLC39A14 disrupt manganese homeostasis and cause childhood-onset parkinsonism-dystonia. *Nat Commun* **7**, 11601 (2016).
49. Williams, B.B., et al. Glycoprotein A33 deficiency: a new mouse model of impaired intestinal epithelial barrier function and inflammatory disease. *Dis Model Mech* **8**, 805–815 (2015).
50. Garcia-Esparcia, P., Hernández-Ortega, K., Ansoleaga, B., Carmona, M. & Ferrer, I. Purine metabolism gene deregulation in Parkinson's disease. *Neuropathol Appl Neurobiol* **41**, 926–940 (2015).
51. Lorenzl, S., et al. Tissue inhibitors of matrix metalloproteinases are elevated in cerebrospinal fluid of neurodegenerative diseases. *J Neurol Sci* **207**, 71–76 (2003).
52. Hughes, A.J., Daniel, S.E., Ben-Shlomo, Y. & Lees, A.J. The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. *Brain* **125**, 861–870 (2002).
53. Allocati, N., Masulli, M., Di Ilio, C. & Federici, L. Glutathione transferases: substrates, inihibitors and pro-drugs in cancer and neurodegenerative diseases. *Oncogenesis* **7**, 8 (2018).
54. Shi, M., et al. Identification of glutathione S-transferase pi as a protein involved in Parkinson disease progression. *Am J Pathol* **175**, 54–65 (2009).
55. Alsina, F.C., et al. Lrig1 is a cell-intrinsic modulator of hippocampal dendrite complexity and BDNF signaling. *EMBO Rep* **17**, 601–616 (2016).
56. Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. *Genome Biol* **18**, 83 (2017).
57. Arif, M., et al. iNetModels 2.0: an interactive visualization and database of multi-omics data. *Nucleic Acids Res* (2021).
58. Birkmayer, J.G., Vrecko, C., Volc, D. & Birkmayer, W. Nicotinamide adenine dinucleotide (NADH)--a new therapeutic approach to Parkinson's disease. *Comparison of oral and parenteral application. Acta Neurol Scand Suppl* **146**, 32–35 (1993).
59. Kuhn, W., et al. Parenteral application of NADH in Parkinson's disease: clinical improvement partially due to stimulation of endogenous levodopa biosynthesis. *J Neural Transm (Vienna)* **103**, 1187–1193 (1996).
60. LeWitt, P.A., Li, J., Lu, M., Guo, L. & Auinger, P. Metabolomic biomarkers as strong correlates of Parkinson disease progression. *Neurology* **88**, 862–869 (2017).
61. Meiser, J., et al. Loss of DJ-1 impairs antioxidant response by altered glutamine and serine metabolism. *Neurobiol Dis* **89**, 112–125 (2016).
62. Lewitt, P.A., et al. 3-hydroxykynurenine and other Parkinson's disease biomarkers discovered by metabolomic analysis. *Mov Disord* **28**, 1653–1660 (2013).
63. Zhao, H., et al. Potential biomarkers of Parkinson's disease revealed by plasma metabolic profiling. *J Chromatogr B Analyt Technol Biomed Life Sci* **1081-**1082, 101–108 (2018).
64. Covarrubias, A.J., Perrone, R., Grozio, A. & Verdin, E. NAD(+) metabolism and its roles in cellular processes during ageing. *Nat Rev Mol Cell Biol* **22**, 119–141 (2021).
65. Gelfin, E., et al. D-serine adjuvant treatment alleviates behavioural and motor symptoms in Parkinson's disease. *Int J Neuropsychopharmacol* **15**, 543–549 (2012).
66. Shigenaga, M.K., Hagen, T.M. & Ames, B.N. Oxidative damage and mitochondrial decay in aging. *Proc Natl Acad Sci U S A* **91**, 10771–10778 (1994).

67. Ferreira, G.C. & McKenna, M.C. L-Carnitine and Acetyl-L-carnitine Roles and Neuroprotection in Developing Brain. *Neurochem Res* **42**, 1661–1675 (2017).
68. Cherix, A., et al. Metabolic signature in nucleus accumbens for anti-depressant-like effects of acetyl-L-carnitine. *Elife* **9**(2020).
69. Salminen, A., Jouhten, P., Sarajärvi, T., Haapasalo, A. & Hiltunen, M. Hypoxia and GABA shunt activation in the pathogenesis of Alzheimer's disease. *Neurochem Int* **92**, 13–24 (2016).
70. Supandi, F. & van Beek, J. Computational prediction of changes in brain metabolic fluxes during Parkinson's disease from mRNA expression. *PLoS One* **13**, e0203687 (2018).
71. Chew, H., Solomon, V.A. & Fonteh, A.N. Involvement of Lipids in Alzheimer's Disease Pathology and Potential Therapies. *Frontiers in physiology* **11**, 598–598 (2020).
72. Castellanos, D.B., Martín-Jiménez, C.A., Rojas-Rodríguez, F., Barreto, G.E. & González, J. Brain lipidomics as a rising field in neurodegenerative contexts: Perspectives with Machine Learning approaches. *Front Neuroendocrinol* **61**, 100899 (2021).
73. Yakhine-Diop, S.M.S., et al. Metabolic alterations in plasma from patients with familial and idiopathic Parkinson's disease. *Aging (Albany NY)* **12**, 16690–16708 (2020).
74. Wyss, M. & Kaddurah-Daouk, R. Creatine and creatinine metabolism. *Physiol Rev* **80**, 1107–1213 (2000).
75. Matthews, R.T., et al. Creatine and cyclocreatine attenuate MPTP neurotoxicity. *Exp Neurol* **157**, 142–149 (1999).
76. Li, Z., et al. The effect of creatine and coenzyme q10 combination therapy on mild cognitive impairment in Parkinson's disease. *Eur Neurol* **73**, 205–211 (2015).
77. Couzin, J. Testing a Novel Strategy Against Parkinson's Disease. *Science* **315**, 1778 (2007).
78. Watanabe, A., Kato, N. & Kato, T. Effects of creatine on mental fatigue and cerebral hemoglobin oxygenation. *Neurosci Res* **42**, 279–285 (2002).
79. Zhong, L.-L., Song, Y.-Q., Tian, X.-Y., Cao, H. & Ju, K.-J. Level of uric acid and uric acid/creatinine ratios in correlation with stage of Parkinson disease. *Medicine (Baltimore)* **97**, e10967–e10967 (2018).
80. Elias, M.F., Dore, G.A. & Davey, A. Kidney disease and cognitive function. *Contrib Nephrol* **179**, 42–57 (2013).
81. Hare, T.A., Vanna, S., Beasley, B., Chambers, R. & Vogel, W.H. Amino acid and dopa levels in plasma and urine from L-dopa-treated patients with Parkinson's disease. *J Lab Clin Med* **77**, 319–325 (1971).
82. Manyam, B.V., Ferraro, T.N. & Hare, T.A. Cerebrospinal fluid amino compounds in Parkinson's disease. Alterations due to carbidopa/levodopa. *Arch Neurol* **45**, 48–50 (1988).
83. Hatano, T., Saiki, S., Okuzumi, A., Mohney, R.P. & Hattori, N. Identification of novel biomarkers for Parkinson's disease by metabolomic technologies. *J Neurol Neurosurg Psychiatry* **87**, 295–301 (2016).
84. Wang, H., et al. High urea induces depression and LTP impairment through mTOR signalling suppression caused by carbamylation. *EBioMedicine* **48**, 478–490 (2019).
85. Wikoff, W.R., et al. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proceedings of the National Academy of Sciences* **106**, 3698 (2009).
86. Mangalam, A., et al. Profile of Circulatory Metabolites in a Relapsing-remitting Animal Model of Multiple Sclerosis using Global Metabolomics. *J Clin Cell Immunol* **4**(2013).
87. Grant, S.M. & DeMorrow, S. Bile Acid Signaling in Neurodegenerative and Neurological Disorders. *Int J Mol Sci* **21**(2020).
88. Graham, S.F., et al. Biochemical Profiling of the Brain and Blood Metabolome in a Mouse Model of Prodromal Parkinson's Disease Reveals Distinct Metabolic Profiles. *J Proteome Res* **17**, 2460–2469 (2018).
89. Burstein, S.H. N-Acyl Amino Acids (Elmiric Acids): Endogenous Signaling Molecules with Therapeutic Potential. *Mol Pharmacol* **93**, 228–238 (2018).

90. Palacios, N., et al. Circulating Plasma Metabolites and Cognitive Function in a Puerto Rican Cohort. *Journal of Alzheimer's disease : JAD* **76**, 1267–1280 (2020).
91. Błaszczyk, J.W. The Emerging Role of Energy Metabolism and Neuroprotective Strategies in Parkinson's Disease. *Front Aging Neurosci* **10**, 301–301 (2018).
92. Brandão, P.R.P., et al. Cognitive impairment in Parkinson's disease: A clinical and pathophysiological overview. *J Neurol Sci* **419**, 117177 (2020).
93. Poston, K.L. & Eidelberg, D. FDG PET in the Evaluation of Parkinson's Disease. *PET Clin* **5**, 55–64 (2010).
94. Zhang, N., et al. Comprehensive serum metabolic and proteomic characterization on cognitive dysfunction in Parkinson's disease. *Ann Transl Med* **9**, 559 (2021).
95. Emre, M., et al. Rivastigmine for dementia associated with Parkinson's disease. *N Engl J Med* **351**, 2509–2518 (2004).
96. Solari, N., Bonito-Oliva, A., Fisone, G. & Brambilla, R. Understanding cognitive deficits in Parkinson's disease: lessons from preclinical animal models. *Learn Mem* **20**, 592–600 (2013).
97. Antonini, A. & Cilia, R. Behavioural adverse effects of dopaminergic treatments in Parkinson's disease: incidence, neurobiological basis, management and prevention. *Drug Saf* **32**, 475–488 (2009).

Paper III: Combined metabolic activators improve cognitive functions in Alzheimer's disease

The provided article below is adapted from an author's original version of the published papers which are available at:

Turkez H, Altay O, Yildirim S, Li X, Yang H, Bayram C, Bolat I, Oner S, Tozlu OO, Arslan ME, Arif M, Yulug B, Hanoglu L, Cankaya S, **Lam S**, Velioglu HA, Coskun E, Idil E, Nogaylar R, Ozsimsek A, Hacimuftuoglu A, Shoaie S, Zhang C, Nielsen J, Borén J, Uhlén M, Mardinoglu A. 2023. "Combined metabolic activators improve metabolic functions in the animal models of neurodegenerative diseases". *Life Sciences*, **314**: 121325

Yulug B, Altay O, Li X, Hanoglu L, Cankaya S, **Lam S**, Veligolu HA, Yang H, Coskun E, Idil E, Nogaylar R, Ozsimsek A, Bayram C, Bolat I, Oner S, Tozlu OO, Arslan ME, Hacimuftuoglu A, Yildirim S, Arif M, Shoaie S, Zhang C, Nielsen J, Turkez H, Borén J, Uhlén M, Mardinoglu A. 2023. "Combined metabolic activators improve cognitive functions in Alzheimer's disease patients: A randomised, double-blinded, placebo-controlled phase-II trial". *Translational Neurodegeneration*, **12**: 4.

Combined Metabolic Activators Improve Cognitive Functions in Alzheimer's Disease

Burak Yulug[1,#], Ozlem Altay[2,#], Xiangyu Li[2,#], Lutfu Hanoglu[3], Seyda Cankaya[1], Simon Lam[4], Hong Yang[2], Ebru Coskun[3], Ezgi İdil[1], Rahim Nogaylar[1], Cemil Bayram[5], Ismail Bolat[6], Sena Öner[7], Özlem Özdemir Tozlu[7], Mehmet Enes Arslan[7], Ahmet Hacımuftuoglu[5], Serkan Yıldırım[6], Muhammad Arif[2], Saeed Shoaie[2,4], Cheng Zhang[2,8], Jens Nielsen[9], Hasan Turkez[10], Jan Borén[11], Mathias Uhlén[2,*], Adil Mardinoglu[2,4,*,+]

[1] Department of Neurology and Neuroscience, Faculty of Medicine, Alanya Alaaddin Keykubat University, Antalya, Turkey

[2] Science for Life Laboratory, KTH - Royal Institute of Technology, Stockholm, Sweden

[3] Department of Neurology, Faculty of Medicine, Istanbul Medipol University, Istanbul, Turkey

[4] Centre for Host-Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, London, United Kingdom

[5] Department of Medical Pharmacology, Faculty of Medicine, Atatürk University, Erzurum, Turkey

[6] Department of Pathology, Veterinary Faculty, Ataturk University, Erzurum, Turkey

[7] Department of Molecular Biology and Genetics, Faculty of Science, Erzurum Technical University, Erzurum, Turkey

[8] School of Pharmaceutical Sciences, Zhengzhou University, Zhengzhou, PR China

[9] Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

[10] Department of Medical Biology, Faculty of Medicine, Atatürk University, Erzurum, Turkey

[11] Department of Molecular and Clinical Medicine, University of Gothenburg and Sahlgrenska University Hospital, Gothenburg, Sweden

* Correspondence: adilm@scilifelab.se; mathias.uhlen@scilifelab.se

[+] Lead author: Adil Mardinoglu (adilm@scilifelab.se)

Emails: burakyulug@gmail.com; havva.altay@scilifelab.se; xiangyu.li@scilifelab.se; lhanoglu@kure.com.tr; cankayaseyda@hotmail.com; simon.1.lam@kcl.ac.uk; hong.yang@scilifelab.se; ebrucosskun@gmail.com; ezgi.idil92@gmail.com; ngylrrhm@gmail.com; cemil489@gmail.com; ismail.bolat@atauni.edu.tr; senaoner02@gmail.com; ozlem.ozdemir@erzurum.edu.tr; enesiyte@gmail.com; ahmeth@atauni.edu.tr; syildirim@atauni.edu.tr; muhammad.arif@scilifelab.se; saeed.shoaie@scilifelab.se; cheng.zhang@scilifelab.se; nielsenj@chalmers.se; hasanturkez@yahoo.com; Jan.Boren@wlab.gu.se; mathias.uhlen@scilifelab.se; adilm@scilifelab.se

ABSTRACT

Alzheimer's disease (AD) is associated with metabolic abnormalities linked to critical elements of neurodegeneration. Here, we analysed the brain transcriptomics data of more than 600 AD patients using genome-scale metabolic models and provided supporting evidence of mitochondrial dysfunction related to the pathophysiologic mechanisms of AD progression. Subsequently, we investigated, in a rat model of AD, the oral administration of combined metabolic activators (CMAs), consisting of NAD+ and glutathione precursors, to explore the effect for improvement of biological functions in AD. CMAs consist of L-serine, nicotinamide riboside, N-acetyl-L-cysteine, and L-carnitine tartrate, the salt form of L-carnitine. The study revealed that supplementation of the CMAs improved the AD-associated histological parameters in the animals. Finally, we designed a randomized, double-blinded, placebo-controlled human phase 2 clinical trial and showed that the administration of CMAs improves cognitive functions in AD patients. As decreased AD Assessment Scale-cognitive subscale (ADAS-Cog) score is the indicator of the improved cognitive function in AD patients, we observed a significant decrease of ADAS-Cog scores on Day 84 vs Day 0 (Log2FC= -0.37, (29% improvement), p-value=0.00001) in the CMA group. We also observed a significant decrease in the placebo group on Day 84 vs Day 0 (Log2FC= -0.19, (14% improvement), p-value=0.001) due to the recommendations of exercise and Mediterranean diet to all AD patients participated in the trial. A comprehensive analysis of the human plasma metabolome and proteome revealed that plasma levels of proteins and metabolites associated with redox metabolism are significantly improved after treatment. In conclusion, our results show that treating AD patients with CMAs leads to enhanced cognitive functions, suggesting a role for such a therapeutic regime in treating AD and other neurodegenerative diseases.

HIGHLIGHTS

- Brain transcriptomics data of more than 600 AD patients were analysed.
- Performed an *in vivo* study using combined metabolic activators (CMAs) in AD rat models.
- We performed a randomized, double-blinded, placebo-controlled human phase 2 clinical trial.

- We showed that cognitive functions in AD patients is improved 29% in the CMA group whereas 14% in the placebo group.

KEYWORDS

Alzheimer's disease: combined metabolic activators (CMAs), Multi-omics analysis; Systems biology; Systems medicine

INTRODUCTION

Alzheimer's disease (AD) is characterized by progressive synaptic and axonal dysfunction, neuronal loss and cognitive decline (***1***). There is growing evidence that AD is closely associated with metabolic and oxidative stress linked to critical elements of neurodegeneration, such as mitochondrial dysfunctions and bioenergetic impairments (***2, 3***). Indeed, increasing data indicate that systemic metabolic disorders, such as insulin resistance, are strongly associated with bioenergetic failure of nerve cells(***4, 5***). This can manifest as cognitive impairment and brain-specific neuropathology while sharing common pathogenic mechanisms with AD, such as impaired glucose metabolism, increased oxidative stress, insulin resistance, and amyloidogenesis (***4, 6, 7***). Recent evidence accordingly suggests that patients with type 2 diabetes mellitus are at increased risk of developing AD (***6***).

Although the disease is still defined by the accumulation of abnormal amyloid and tau proteins (***8***), the mechanistic assumption of linear causality between the amyloid cascade and cognitive dysfunction in AD is flawed, since amyloid-lowering approaches have failed to provide cognitive benefits in human clinical trials (***9***). A growing body of evidence suggests that impaired brain energy metabolism in AD may contribute to cognitive decline. At the same time, therapeutic options, such as drugs typically prescribed for metabolic disorders that improve metabolic status, may slow cognitive decline or prevent dementia progression (***10***). This is suggested by positron emission tomography imaging studies revealing baseline cerebral glucose metabolism abnormalities before the onset of cognitive symptoms in patients with AD (***11***). In addition, recent preclinical data indicate that ageing and AD are associated with the reorganization of brain energy metabolism, including an overall increase in lactate secretion and the downregulation of bioenergetic enzymes (***12, 13***).

Although current research is paving the way for developing neuroprotective therapeutics, the results of early clinical trials of drugs targeting single pathways have been mostly unsuccessful. A divergent approach combining multiple compounds that simultaneously reduce oxidative injury and improve bioenergetics, in other words, targeting various pathways has been proposed as a therapeutic strategy associated more likely with successful translational outcomes (***14***). Previous research identified limited serine availability, reduced de novo glutathione synthesis, and altered NAD+ metabolism based on the combining multi-omics profiling of transgenic AD mouse

model of AD (***15***). Consistent with this, it also has been reported that age- and AD-associated metabolic shifts responded well to NAD(P)+/NAD(P)H redox-dependent reactions (***16, 17***). These findings were confirmed by human metabolomic data showing significantly altered cerebrospinal fluid (CSF) acylcarnitine levels in patients with AD, which correlated with the decline of cognitive functions and structural brain abnormalities (***18, 19***).

Based on integrative network analysis of non-alcoholic fatty liver disease multi-omics data, we have developed a mixture combined metabolic activators (CMAs) consisting of L-serine, N-acetyl cysteine (NAC), nicotinamide riboside (NR), and L-carnitine tartrate (LCAT, the salt form of L-carnitine) and showed that administration of CMAs activates mitochondria, improves inflammation markers in animals and humans (***20***–***24***). We have also found that the administration of CMAs promotes mitochondrial fatty acid uptake from the cytosol, facilitates the fatty acid oxidation in the mitochondria, and alleviates oxidative stress (***25***). Recently, we reported that CMAs administration effectively increased fatty acid oxidation and *de novo* glutathione generation, as evidenced by metabolomic and proteomic profiling (***20***). Moreover, plasma levels of metabolites associated with antioxidant metabolism and inflammatory proteins were improved in COVID-19 patients treated with CMAs compared to the placebo (***24***).

Here, we first analysed brain transcriptomics data obtained from 629 AD patients and 704 control subjects using genome-scale metabolic models and revealed that mitochondrial dysfunction is involved in the underlying molecular mechanisms associated with AD. Second, we tested the effect of CMAs, which has been shown to activate mitochondria in the AD rat models and showed that supplementation of the CMAs improved the AD and associated functions in animals. Next, we hypothesized that AD patients could be treated with the administration of the CMAs by activating the mitochondria in the brain tissue of the patients. Finally, we designed a randomized, double-blinded, placebo-controlled human phase 2 clinical study, studied the effect of administration on the global metabolism of AD patients and showed that administration of CMAs improves the cognitive functions in AD patients.

RESULTS

Analysis of transcriptomics data reveals mitochondrial dysfunction in the brain of AD patients

To identify the metabolic dysregulations based on brain transcriptomics data, we obtained the global mRNA expression profiling of 629 AD patients and 704 controls from the Religious Orders Study/Memory and Aging Project (ROSMAP) (***26***–***28***). We performed differential expression analysis and identified 914 significantly (p-adjusted < 1.0 E-10) upregulated and 1725 significantly (p-adjusted < 1.0 E-10) downregulated differentially expressed genes (DEGs) (**Figure 1A**, Dataset S1). We identified several upregulated metabolism-related genes, including pyruvate dehydrogenase kinase 4 (PDK4), carnitine palmitoyl transferase (CPT1A), hexokinase 2 (HK2), and spermine oxidase (SMOX), as well as downregulated genes including acyl-CoA dehydrogenases (ACADs), ATP synthase (ATP5MGL), and acetyl- and acyltransferases (GCNT7, MBOAT4, GALNT17). These results suggest that there may be some alterations associated with glycolysis, fatty acid biogenesis and the urea cycle in AD compared to control.

Gene set enrichment (GSE) analysis revealed that these DEGs were significantly enriched in the protein synthesis, ATP synthesis, lipid metabolism, cell cycle, cell migration, cell differentiation and cell adhesion pathways (**Figure 1B**, Dataset S1). Then, we performed reporter metabolite analysis to predict the significantly changed metabolites in each subcellular compartment using a genome-scale metabolic model of brain tissue (***29, 30***). We identified numerous reporter metabolites related to glycolysis, amino acid metabolism (e.g. glycine, serine, threonine, alanine and branched-chain amino acid valine, leucine and isoleucine), mitochondrial metabolism (acyl-CoA and ferredoxin) and TCA cycle (e.g. succinate and glutamine) (**Figure 1C** and **1D**, Dataset S1). These results suggested a widespread perturbation in energy metabolism related to brain cell survival and mitochondrial dysfunction in the brain.

Administration of CMAs to animal models of AD

To test the effect of CMAs in animals, we provided individual metabolic activators and CMAs to the rat model of AD which has been developed after intracerebroventricular-streptozotocin (STZ) injection. We observed that administration of all constituents of CMAs (in combination or separately) significantly ($p<0.05$) decreased the plasma levels

of triglycerides (TG) compared to the control group (Group 4; **Figure 2A**, Dataset S2). In parallel, administration of only serine or NR significantly decreased total cholesterol ($p=0.01$) and low-density lipoprotein (LDL; $p=0.03$) in rats (**Figure 2B**, Dataset S2). Additionally, a significant reduction in total cholesterol ($p=0.04$) and LDL ($p=0.03$) was observed with NAC-treated and LCAT-treated rats, respectively (**Figure 2B**, Dataset S2). Of note, we found significant reductions in plasma AST ($p=0.03$) and an increase in plasma ALP ($p=0.04$) concentrations only in LCAT-treated rats (Group 8) (**Figure 2B**, Dataset S2).

Additionally, the histological analyses and immunofluorescence imaging techniques showed a significant neuronal tissue damage in the high fat diet (HFD) and HFD+STZ groups' brains compared to those of the chow diet group (**Figure 2C**, Dataset S2). Specifically, the HFD groups developed more hyperaemia as well as more degeneration and necrosis in neurons (**Figure 2C**, Dataset S2). In parallel, DNA damage markers (namely 8-OHdG and H2A.X) and caspase 3 were elevated in the HFD groups (**Figure 2C**, Dataset S2). These animal models allowed us to examine each rat group's histopathological differences and assess the brain tissue response to CMAs administration compared to the HFD+STZ group (**Figure 2C**, Dataset S2). We observed that hyperemia, degeneration and necrosis in neurons were improved by serine, LCAT, or NR supplementation individually as well as in combination. However, we observed a better improvement after the administration of CMAs compared to individual metabolic activators. These findings were also supported by immunohistochemical evidence of decreased immunoreactivity seen in neurons (**Figure 2C**, Dataset S2). Of note, rats receiving combination therapy (consisting of serine, NAC, LCAT and NR) developed less hyperaemia in brain tissue and no necrosis in neurons (**Figure 2C**, Dataset S2). In parallel to the improvement in the brain, we also observed dramatic improvement in the liver after the supplementation of CMAs. Scoring of histopathological, immunohistochemical and immunofluorescence findings for brain and liver tissues are presented in **Table 1**.

CMAs Improve Cognition and Blood Parameters in Alzheimer's Disease Patients

To test the effect of the CMAs in AD patients, we performed a double-blind, randomized, placebo-controlled phase 2 study and screened 89 adults diagnosed with AD. We recruited 69 patients older than 50 years with mild to moderate AD according to ADAS-Cog (AD Assessment Scale-cognitive subscale; ADAS≥12) and the Clinical Dementia Rating Scale Sum of Boxes (CDR-SOB; CDR≤2). Of the 69 patients, 47 were randomly assigned to the CMA group and 22 to the placebo group and completed visit 2

after 28 days. Of these patients, 60 (40 in the CMA group and 20 in the placebo group) completed visit 3 after 84 days (**Figure 3A**, Dataset S3). We assessed the clinical variables on Days 0, 28 and 84, and analysed the differences between the time points in the CMA and placebo groups (Dataset S4).

The mean patient age was 70.8 years (56-86 years) and 52.1 % were men (Dataset S4). The baseline demographic and clinical characteristics were similar in the CMA and placebo groups (Dataset S4). Regarding safety, no severe adverse events occurred, and four patients (5.8 %) reported adverse events. All had a mild rash on the body and agreed to complete the study (Dataset S3).

We measured clinical variables in all patients and analysed the differences before and after administration in the active and placebo groups (**Figure 3B**, Dataset S4). ADAS-Cog scores were significantly decreased on Day 28 vs Day 0 (Log2FoldChange (FC)= -0.33, (26% improvement), p-value=0.0000003, lower indicates better cognitive function) and further decreased on Day 84 vs Day 0 (Log2FC= - 0.37, (29% improvement), p-value=0.00001) in the CMA group. A slight but statistically significant improvement was also found in the placebo group on Day 28 vs Day 0 (Log2FC= -0.16, (12% improvement), p-value=0.009) and Day 84 vs Day 0 (Log2FC= -0.19, (14% improvement), p-value=0.001) due to the recommendations of exercise and Mediterranean diet to all AD patients.

We also analysed the differences between clinical parameters by stratifying the patients into low-scoring (mild patients) and high-scoring (severe patients) ADAS-Cog groups (>20ADAS-Cog score is high, ≤20 is low). More interestingly, we found a significant improvement of ADAS-Cog scores between Day 28 vs Day 0 (Log2FC= -0.31, (24% improvement), p-value=0.002) and Day 84 vs Day 0 (Log2FC=-0.38, (30% improvement), p-value=0.003) in the severe CMA group and no significance difference in the severe placebo (p>0.05 in both time points) group (**Figure 3B**, Dataset S4). As shown on **Figure 3B**, we observed a significant difference in the baseline value distribution and mean of ADAS-Cog scores in the severe (ADAS-COG > 20) CMA and placebo group due to the randomization of the subjects. To verify our results, we selected 10 patients from the CMA group with matched ADAS-COG values to the placebo group (p-value=0.693) and present the ADAS-Cog scores in **Figure 3C**. We recalculated the differences in ADAS-COG scores and again found significant improvement in the CMA group, whereas no significant difference in the placebo group.

Our results indicated that the AD patients with high ADAS-Cog scores are more responsive to CMA.

However, when considering mean improvement to ADAS-Cog between the CMA and placebo arms of the study, the trial did not meet its primary endpoint. Participants receiving CMA had a mean change to ADAS-Cog of -5.6 points by day 84, whereas participants receiving placebo had a mean change of -4.2. The benefit of CMA to ADAS-Cog was -1.3 points compared to placebo ($p = 0.39$, two-tailed t-test). Although the trend suggests a promising improvement to ADAS-Cog, the change was not statistically significant, and the two arms of the study were examined separately henceforth.

Analysis of secondary outcome variables showed that serum alanine aminotransferase (ALT) levels (Log2FC=-0.38, (30% improvement), p-value=0.01) and the uric acid levels (Log2FC=-0.19, (14% improvement), p-value=0.001) were significantly lower on Day 84 vs Day 0 only in the CMA group (**Figure 3D**, Dataset S4). This reduction was seen both in high- and low-ALT level groups. In contrast, we found that no significantly altered parameters on Day 84 vs Day 0 in the placebo group (**Figure 3D**, Dataset S4). Hence, we observed that the ALT level and uric acid were significantly improved due to the administration of CMAs as previously reported in the phase 2 NAFLD and phase 3 Covid-19 clinical trials (***24, 31***).

We also measured the level of complete blood count parameters and found that their levels were significantly changed in the CMA group (**Figure 3D**, Dataset S4). We found that the levels of platelets, basophil% and absolute numbers of basophil and neutrophil were significantly lower on Day 84 vs Day 0 only in the CMA group. In contrast, we found that the levels of monocytes were significantly increased on Day 84 vs Day 0 in the CMA group (**Figure 3D**, Dataset S4). Hence, our analysis indicated that the administration of CMAs improved the clinical parameters in parallel to the improvement in cognitive functions in AD patients.

Blood Profile Informs the Response to CMAs

Treatment response variability and clinical heterogeneity in AD are well documented in the literature. We observed interindividual variability in responses to CMAs as well as in clinical measures. Therefore, we hypothesized that some of the patients would respond better to CMAs than others and that clinical measurements could define these subsets.

To determine whether alanine transferase (ALT), a marker for liver damage, could indicate a better response to CMAs, we stratified the patients into high and low ALT groups by the median value of ALT of all patients on Day 0. As shown in **Figure 4A**, the patients of the CMA group with low ALT achieved a significant improvement in ADAS-Cog score over different time points, while the patients in the placebo group had no improvement. In contrast, the patients of the CMA group with high ALT levels also exhibited an improved (i.e., decreased) ADAS-Cog score, but the degree of change was not as much as the patients in the CMA group with low ALT levels. Moreover, patients in the placebo group with high ALT levels also achieved an improved ADAS-Cog score. Thus, these results suggest that the patients with low ALT levels are more responsive to CMAs.

We repeated this stratification for each blood parameter to determine the patient conditions in which CMAs produces the greatest response (**Figure 4B**). In addition to low ALT, we identified high ALP, low GGT, high HCT, high HbA1c, high insulin, high uric acid, high ADAS-Cog, high basophil count, and high red blood cell count as indicators for better responsiveness to CMAs. We also found that individuals who do not drink alcohol or smoke also respond better to CMAs.

CMAs Increases the Plasma Levels of Metabolites Associated with Metabolic Activators

We first analysed the plasma levels of serine, carnitine, NR and cysteine, and their by-products. CMAs administration increased the plasma levels of metabolic activators on Day 84 vs Day 0 in the CMA group (**Figure 5A**, Dataset S5). Moreover, the plasma levels of NR, 1-methylnicotinamide, nicotinurate, N1-methyl-2-pyridone-5-carboxamide and nicotinamide (associated with NR and NAD+ metabolism); of serine, glycine and sarcosine (associated with serine and glycine metabolism); and of deoxycarnitine and carnitine (associated with carnitine metabolism) were significantly higher in the CMA group on Day 84 compared to Day 0.

Next, we investigated the relationship with the plasma level of administrated metabolic activators and other metabolites. We analysed 195 of the most significantly correlated plasma metabolites with serine, L-carnitine, NR, and cysteine (Dataset S6). We found two clusters of metabolites that are significantly correlated with cysteine only or together with serine, carnitine and NR. We observed that cysteine had a different plasma

changes compared to the other three metabolic activators as has been reported in previous clinical trials (***24, 31***).

Effect of CMAs on Global Metabolism

We identified the significantly (FDR<0.05) different plasma metabolites on Day 84 vs Day 0 and found that the plasma levels of 132 metabolites were significantly different in the CMA group (**Figure 5**, Dataset S5). Evaluation of plasma metabolites that differed significantly on Day 84 vs Day 0 in each group showed that a larger number of metabolites related to amino acids (n=53), lipid metabolism (n=42) and other metabolic pathways (n=37) were altered in the CMA group compared to the placebo group (**Figure 5**, Dataset S5).

N-acetyl aspartate (NAA) is one of the most abundant brain metabolites and its reduced plasma levels are associated with brain tissue damage. Previous research revealed the importance of NAA to maintain energy metabolism in the central nervous system (***32***). In our study, we observed that plasma levels of NAA significantly increased on Day 84 vs Day 0 in the CMA group (**Figure 5B**, Dataset S5). Another upregulated metabolite on Day 84 vs Day 0 in the CMA group is sarcosine (a derivative of glycine) which has been widely studied for its improving effects of cognitive symptoms by different pharmacological activities in neurons (***33***). Of note, quinolinic acid (an endogenous excitotoxin acting on NmMethyl-D-aspartate receptors leading to neurotoxic damage) levels significantly decreased on Day 84 vs Day 0 only in the CMA group (**Figure 5B**, Dataset S5).

Increased plasma homocysteine levels is a known risk factor for AD, and several animal studies implicated the promising results of methionine restriction (***34, 35***). In our clinical trial, plasma levels of S-adenosylhomocysteine as well as 2,3-dihydroxy-5-methylthio-4-pentenoate (DMTPA) and N-acetyl taurine were significantly downregulated on Day 84 vs Day 0 in the CMA group (**Figure 5B**, Dataset S5). Of note, reductions of these metabolites are significantly correlated with serine and NR supplementation (**Figure 6A**, Dataset S6).

Increased plasma levels of metabolites in the kynurenine pathway are associated with AD severity (***34***). In our study, we found that plasma levels of kynurenate and 8-methoxykynurenate were significantly lower on Day 84 vs Day 0 in the CMA group (**Figure 5B**, Dataset S5). Reduction in the plasma level of kynurenate was positively

correlated with plasma serine levels (Dataset S6). Kynurenate, which has a prooxidant effect, is the product of the tryptophan degradation pathway. Its aerobic irradiation produces superoxide radicals and leads to cytochrome C reduction (***36***). It has been reported that increased levels of kynurenine leads to cell death through the reactive oxygen species (ROS) pathway in nature killer (NK) cells (***37***) and lower blood pressure in systemic inflammation (***38***).

Emerging evidence indicates a link between the abnormal kidney function and AD, but the potential impact of kidney on cognitive impairment is still undetermined (***39***). Recent studies showed that plasma levels of N,N,N-trimethyl-5-aminovalerate are involved in lysine metabolism, and serve as an indicator of elevated urinary albumin excretion (***40***). Here, we found that the plasma level of N,N,N-trimethyl-5-aminovalerate was significantly decreased on Day 84 vs Day 0 in the CMA group (**Figure 5B**, Dataset S5) and significantly inversely correlated with the plasma level of serine and NR. Moreover, the plasma level of creatinine was also significantly decreased on Day 84 vs Day 0 in the CMA group (**Figure 5B**, Dataset S5). The plasma reduction on creatinine is inversely correlated with the plasma level of serine (Dataset S6). Additionally, our analysis revealed decreased levels of several metabolites belonging to histidine metabolism in the CMA group on Day 84 vs Day 0. Among those N-acetyl-1-methylhistidine is associated with chronic kidney disease and showed a significant negative correlation with serine supplementation (**Figure 5B**, Dataset S6). Also, we found that plasma levels of metabolites related to the urea cycle (3-amino-2-piperidone, arginine, homoarginine, N-alpha-acetylornithine, ornithine and pro-hydroxy-pro) were significantly decreased in the CMA group on Day 84 vs Day 0 (**Figure 5B**, Dataset S5) and inversely correlated with the plasma level of serine and NR (Dataset S6).

Lipids play a fundamental role in the pathophysiology of neurodegenerative diseases, including AD. Specific lipid species of cellular membranes (e.g., cholesterol and sphingolipids) are not only structural components of cell membranes but also regulate a plethora of critical aspects of brain functions (***41***). In our study, plasma levels of a considerable number of metabolites associated with sphingomyelins and fatty acid metabolism (acyl carnitines) were significantly increased on Day 84 vs Day 0 in the CMA group (**Figure 5C**, Dataset S5). Interestingly, plasma levels of pregnenolone steroids and 2R,3R-dihydroxybutyrate were significantly decreased on Day 84 vs Day 0

(**Figure 5C**, Dataset S5). These alterations were significantly positively correlated with carnitine and serine levels (Dataset S6).

Effect of CMAs on Plasma Proteins

Plasma levels of 1466 protein markers were measured with the plasma proteome profiling platform Proximity Extension Assay quantifying the plasma level of target proteins. After quality control and exclusion of proteins with missing values in more than 50% of samples, 1463 proteins were analysed (Dataset S7). Proteins whose levels differed significantly between the visits in the CMA and placebo groups are listed in Dataset S7.

We analysed the effect of CMAs on plasma protein profile and found that 22 proteins were significantly (p-value<0.01) different in the CMA group on Day 84 vs Day 0. Nineteen of these proteins were significantly decreased, whereas 3 of these proteins were significantly increased on Day 84 vs Day 0. After filtering out the proteins based on log2FC, we found that the plasma levels of PSPN, OSM, PADI4, PDGFC, SCGN, LTBP3, CLEC4G, MERTK, WNT9A, ISM1, ASAH2, CES3, HPGDS, NPY, THPO, SIGLEC6, GDNF, PADI2 and EGFL7 were significantly downregulated in the CMA group. The plasma level of KLB, BGN, and ST3GAL1 was significantly upregulated in the CMA group (**Figure 6B**, Dataset S7). We observed only one significantly (p-value<0.01) altered protein, EGFL7, which was upregulated in the placebo group (**Figure 6B**, Dataset S7). Hence, we observed that plasma levels of proteins are significantly altered in response to CMA treatment.

The proteomic analysis In this study revealed significant alteration in levels of several critical proteins that play an essential role in the pathogenesis of AD. For instance, levels of MertK (***42, 43***), EGFR (***44, 45***), oncostatin (***46–50***), PAD4 (***51, 52***), LTGF (***53–57***), and TPO (***58***), known as a strong inducer of neuro-inflammation, amyloid production and apoptosis, decreased. In contrast, proteins with neuroprotective and pro-cognitive properties, such as Klotho (***59, 60***) and ST3GAL1 (***61***), increased after CMA treatment. More interestingly, the majority of the analysed proteins were also found to be significantly altered in recent human AD studies (***53-57, 62-68***). KlothoB levels were also significantly altered after CMA treatment, consistent with their neuroprotective role as a cofactor and neurotrophic factor. In this context, recent studies have shown that KlothoB indirectly regulates glucose and energy metabolism through F2F1, expressed in

certain areas of the brain involved in learning and memory (***67***). Moreover, GABA signalling has also been shown to play a critical role in mediating the detrimental effects of increased dihydroxybutyrate levels in the progression of MCI (***69***). Interestingly, our metabolomic study indicated decreased post-therapeutic dihydroxybutyrate levels. Although the exact pathways involved in the metabolic generation of DHBA are still far from clear, it has been hypothesized that dihydroxybutyrate levels may be a compensatory response to increased cellular stress secondary to compromise of the Krebs cycle function, creating an alternative energy production pathway in AD (***69***). This represents indirect evidence to suggest that our treatment exhibits an energetic regulatory function.

Integrative Multi-Omics Analysis

Multi-omics data integrations have been proven to give novel insights and a more holistic view of the human body, in both healthy and disease states (***70***). In this study, we generated an integrative multi-omics network using metabolomics and proteomics data, coupled with detailed clinical variables, to understand the functional relationships between analytes from the same and different omics data types. We generated the network using the method used in iNetModels (***71***), to which we also deposited our network. The network consisted of 937,282 edges from 2,273 nodes (36.3% network density, Dataset S8).

We extracted a sub-network to highlight the interactions between the individual metabolic activators, cognitive function (ADAS-Cog scores), two highlighted proteins (OSM and PSPN), and their top neighbours (**Figure 6C**). From the sub-network, ADAS-Cog was shown to be negatively associated with carnitine (and its derivatives) and nicotinamide associated metabolites, where the metabolic cofactors were negatively associated with fatty acid and histidine metabolism. Finally, we observed that, among others, OSM and PSPN were positively associated with immune and cell cycle-related proteins.

Subsequently, we performed centrality analysis to identify the most central analytes in the networks. The top 20 most central metabolites were dominated by amino acid metabolites (tryptophan, glutamate, and branched-chain amino acid metabolism) and lipid metabolites (androgenic steroid pathway), where top proteins were related to,

among others, short- and long-term memory (CALB1), lipid metabolism (PLA2G10), and immune response (SELPLG, CLEC4D, and LGALS7).

Furthermore, we performed community analysis within the network using the Leiden algorithm. We discovered 3 modules that showed significant interaction among the members. In cluster-0, the biggest cluster, the top nodes were related to tryptophan metabolism (indoleacetate), fatty acid metabolism (3-hydroxyoctanoate), and steroid metabolism (11-ketoetiocholanolone glucuronide and 11-beta-hydroxyetiocholanolone glucuronide). Moreover, we found two top proteins in the same cluster, ACTA2 and IGFBP1, that have been associated with AD (***72, 73***). In cluster-1, the top nodes were associated with leucine metabolism (3-hydroxy-2-ethylpropionate), ceramide phosphatidylethanolamine, and a carnitine metabolite (erucoylcarnitine), meanwhile, cluster-2's central nodes were related to methionine metabolism and aminosugar metabolism (N-acetylglucosamine/N-acetylgalactosamine). These results showed that the integrative multi-omics network analysis can be used to strengthen the results from single omics analyses and identify key analytes associated with AD. Moreover, it provided new insights by elucidating the functional relationships within and between different omics data.

In evaluating the correlations between each cofactor (used in the present study for therapeutic purposes) and clinical, metabolic, and proteomic parameters, we identified significant correlations between serine, carnitine, cysteine, and nicotinamide levels and improved peripheral blood parameters, such as liver function, CBC, and glycated hemoglobin (HbA1c), which are relevant to the pathogenesis of AD. Accordingly, improved ADAS-Cog scores were also associated with changes in serum serine and carnitine, which fit well with their well-known pro-cognitive and energy-boosting effects. Similar results were also observed for metabolomic and proteomic data. The majority of the cofactors exhibited significant correlations with improved metabolites and proteins (either increased or decreased) relative to a slower degeneration process in AD. It is worth mentioning here that that two of the proteins, OSM and PSPN, most strongly associated with other beneficial protein metabolites, were also related to several critical amino acid alterations, such as spermidine and hypotaurine, which may suggest a metabolic shift from the protein to amino acid metabolism to compensate the energy deficit reported in AD.

DISCUSSION

Here, we show that oral administration of CMAs has profound effect on cognitive function after only 84 days of treatment in AD patients based on ADAS-Cog scores. We showed that cognitive functions in AD patients is improved 29% in the CMA group whereas 14% in the placebo group after 84 days of CMAs administration. We also recently tested the effect of CMAs administration on an independent cohort of Parkinson's disease patients in a phase 2 clinical trial, and found that oral CMAs administration has a profound effect on cognitive function without altering motor scores in Parkinson's disease (***74***). As the increased Montreal Cognitive Assessment (MoCA) scores is known as the indicator of the increased cognitive functions in PD patients, we observed that the mean MoCA scores were significantly higher in the CMA group both on Day 28 vs Day 0 (log2FoldChange(FC)=0.17, (13% improvement), p=0.001) and on Day 84 vs Day 0 (log2FC=0.27, (21% improvement), p= 0.0001). We also observed a significant increase on MoCA scores in placebo group on Day 28 vs Day 0 (log2FC=0.16, (12% improvement), p=0.001) and on Day 84 vs day 0 (log2FC=0.15, (11% improvement), p=0.04) due to the recommendations of exercise and Mediterranean diet to all PD patients participated in the trial. Notably, the degree of increase of MoCA was much higher on Day 84 vs Day 0 in the CMA group than in the placebo group, suggesting the PD patients benefitted from CMA treatment after 84 days of treatment. Even though Montreal Cognitive Assessment (MoCA) scores have been used to evaluate the cognitive function in the PD patients, the effect of CMAs administration on cognitive function has been verified in an independent patient group with different neurological diseases.

Clinically, when AD patients were stratified by high and low ADAS-Cog scorer, we observed that patients with lower ADAS-Cog scores in the placebo group also showed improved cognitive function similar to the CMA group. However, for the patients with higher ADAS-Cog scores, cognitive function was not improved in the placebo group, while a positive effect was observed for the patients treated with metabolic activators. This is particularly interesting, since this patient group lacks current therapeutic regimes, except for palliative support. Apart from clinical severity, we also observed that various clinical variables were also related to the treatment response. For example, patients with low ALT, who did not drink alcohol or smoke, and who had signs of an increased metabolic load (i.e., increased HbA1c and insulin levels) or impaired CBC values

responded better to treatment. Additionally, the treatment significantly improved the altered metabolic and CBC parameters described above.

The effect of oral administration of CMAs was substantiated with a comprehensive analysis of protein and metabolites in the plasma of the patients using a multi-omics analytical platform. The clinical results are consistent with the genome-scale metabolic modelling of more than 600 AD patients showing evidence of mitochondrial dysfunction. It is also consistent with the results from an animal model demonstrating improved AD-associated histological parameters in animals treated with an oral administration of CMAs. Thus, the present study suggests an attractive therapeutic regime for improving mitochondrial dysfunction in AD patients.

The metabolomics data confirmed the expected biological outcomes of CMA treatment. Levels of plasma nicotinamide and related metabolites increased, suggesting that NR provided sufficient substrate for mitochondrial fatty acid oxidation. In addition to its role as a cellular metabolite, NAD+ functions as an essential cofactor for the DNA repair protein PARP1 (***17***). Hyperactivation of PARP1 and decreased NAD+ have been already identified in the brains of patients with AD (***75, 76***). Serine plasma levels also increased, suggesting that CMA treatment improves the serine deficiency associated with AD. For instance, a recent study showed that the adenosine triphosphate (ATP)-reducing the effect of glucose hypometabolism was restored with oral serine supplementation, suggesting the potential use of oral serine as a ready-to-use therapy for AD (*77*). The exact mechanism of action also applies to cysteine. As a glutathione precursor, cysteine is even more specific in acting as an antioxidant and anti-inflammatory agent, maintaining the mitochondrial energetic homeostasis and key neurotransmitter systems, such as glutamate, involved in learning and memory (***78, 79***). Accordingly, NAC has been tested as a medication in AD and found to exhibit effects suggestive of future potential use as an alternative medication (***80***). More importantly, fatty acid oxidation and carnitine metabolism were significantly facilitated, as shown by the robust increase in plasma levels of carnitine. These findings fit well with recent human data showing that severe disturbances in carnitine metabolism frequently occur in individuals with AD, in association with severe mitochondrial dysfunction (***81, 82***). Cristofano et al. showed a progressive decrease in carnitine serum levels in individuals shifting from normal status to AD, suggesting that decreased serum concentrations of carnitine may predispose to AD (***83***). In support of this hypothesis, human clinical studies have

demonstrated the pro-cognitive effects of carnitine in mild cognitive impairment and AD (***84***–***86***). This, in turn, led to the suggestion that stabilizing the bioenergetic balance may slow or even reverse mild cognitive impairment and the progression of dementia in patients with AD.

In addition, the levels of tryptophan metabolites, including kynurenate, kynurenine, and tryptophan betaine, decreased significantly after CMA treatment. Increased levels of these metabolites were previously shown to be associated with increased neurodegeneration and clinical cognitive impairment through an increased oxidative load and the formation of neurofibrillary tangles (NFTs) (***87, 88***). For instance, recent data showed a synergistic relationship between β-amyloid 1-42 and enzymatic activations of the tryptophan kynurenine pathway, resulting in increased oxidative stress, which may be associated with the formation of NFTs and senile plaque development (***89***). Also, one recent study revealed that tryptophan-2,3-dioxygenase (TDO) was highly expressed in the brains of patients with AD and co-localized with quinolinic acid, NFTs, and amyloid deposits in the hippocampus of post-mortem brains of patients with AD (***90***).

We also observed significantly increased levels of NAA, sarcosine, methionine, cysteine, and S-adenosylmethionine (SAM) and decreased levels of histidine, tryptophan quinolate, and urea cycle metabolites, which play a critical role in cognitive and mitochondrial functions. For instance, increased NAA may provide an additional energy source for intercellular metabolite trafficking during the neurodegenerative process, especially when glucose metabolism is downregulated (***32***). Similarly, increased sarcosine levels may boost cognition, as previously shown in patients with schizophrenia, in which oxidative damage and impaired glucose metabolism play key roles (***91***). In addition, decreased histidine metabolism and other decreased markers, such as homocysteine and S-adenosylhomocysteine (SAH) found in our treatment group, have been already shown to slow the cognitive ageing process appropriately downregulated (***92***). For instance, increased plasma homocysteine levels are a known risk factor for AD, whereas a low leucine and arginine diet yield beneficial cognitive effects (***93***).

Unexpectedly, CMAs rapidly lowered uric acid and associated metabolite'' levels. Uric acid stimulates inflammation either directly or by activating NLRP3 inflammasomes (***94***). Although the extent to which uric acid reduction contributed to the regression in

cognitive impairment is unclear, it is likely that it is linked to the improvement in the metabolic homeostasis. A good example is a recent clinical study showing increased urea metabolism in patients with AD (***95***). Accordingly, decreased taurine levels and urea metabolites are associated with a diminished risk of dementia (***96***). The majority of clinical study findings collectively agree with our results, showing significantly dysregulated baseline metabolites, which normalized with treatment.

Of note, to date, a few studies aimed to identify global changes in metabolites and metabolic pathways in AD (***15, 97, 98***). Among these, some studies highlighting that lipid dysfunction also plays an important role in the pathophysiology of AD (***99***). In terms of lipid metabolism, significant differences in the levels of some compounds have been observed in patients with AD. Despite some discrepant trends in cross-sectional studies examining the levels of lipids in AD patients (***100, 101***), the plasma levels of sphingolipids, sphingomyelins (***102, 103***), acylcarnitines (***104***) and phosphatidylcholines (PC) (***105***–***107***) exhibited statistically lower concentrations in patients with AD, even in the preclinical stages of the disease (***18***). In addition, a significant correlation among different lipid metabolites, tau and amyloid pathology, brain atrophy and cognitive decline was observed in a AD human study (***18***). An autopsy study of frontal cortex metabolites from patients with AD showed that impaired glycerophospholipid metabolism was involved in six central metabolic pathways reported to be altered in the disease (***108***). In brief, we observed significantly increased post-therapeutic levels of lipid metabolites, previously reported to decrease in patients with AD, including sphingomyelin, carnitine and carnitine-related by-products.

Despite insufficient clinical AD data concerning cholesterol metabolites and dicarboxylic acids (DCAs), we observed significantly lower levels of these metabolites after CMA treatment (***109***). Levels of pregnanediol, a metabolite of pregnenolone, and DCAs, end-products of β- or omega oxidation, which were observed as decreased in the present study, were previously reported to be lower in the urine of patients with AD (***110, 111***). Considering the neurotoxic role of bile acids, along with the oxidative properties of DCAs, the detection of decreased levels of bile acid metabolites and DCA products in the present study is therefore not surprising. Similarly, allopregnanolone has already been reported to result in deleterious effects on cognitive functions through gamma-aminobutyric acid (GABA) signalling (***112***). Also, increased bile acid levels have been reported in mild cognitive impairment (MCI) and AD. In contrast, bile acids

strongly inhibited the cysteine catabolic pathway in the preclinical period, resulting in depletion of the free cysteine pool and reduction of antioxidant glutathione concentrations (***113***).

Although there has been no direct evidence relating plasma ascorbic acid (AA) levels to the pathogenesis of AD, our finding of decreased plasma levels after CMAs may be related to increased brain concentrations and central consumption of ascorbic acid in cognitively improved patients with AD. In this respect, and contrast to other metabolites, a direct transmission from the periphery to the brain suggested that AA might play a 'nourishing' role in the brain. A direct correlation between brain levels of AA and dementia was eventually confirmed by a recent study showing that a high CSF: plasma AA ratio was a marker of a "healthier" brain, better able to cope with the neurodegenerative process in AD (***114***).

However, despite these promising human data, plasma levels of several proteins in AD observed after CMA treatment in the present study were inconsistent with their established beneficial role under experimental conditions. For example, PSP, which decreased in the present study, is a novel neurotrophic factor exhibiting significant similarities to GDNF by exerting intense neurotrophic activity, specifically on central neurons (***115, 116***). Such discrepancies are not easily explained, even if the different results obtained in peripheral blood can be correlated with improved clinical status. A finding of decreased or increased post-therapeutic levels of these proteins in patients with AD may reflect a treatment-related alteration in their cellular processing in the central nervous system, resulting in variability in protein production and/or degradation in peripheral blood. In addition, controversial data have been reported regarding the concentrations of most of these proteins in patients with AD (***102, 117–120***). Also, due to the high heterogeneity of clinical cohorts and the restricted number of patients, the reproducibility of previous proteomics studies' results is noticeably low.

Our therapeutic data fits well with recent ROSMAP transcriptomic data in AD patients, showing impaired glycolysis, urea cycle, and fatty acid metabolism that was considerably normalized after the CMA treatment. For instance, our transcriptomic analysis showed increased expressions of pyruvate dehydrogenase kinase 4 (PDK4), carnitine palmitoyltransferase (CPT1A) and hexokinase 2 (HK2). It decreased expressions of acyl-CoA dehydrogenases (ACADs), ATP synthase (ATP5MGL), and acetyl- and acyltransferases (GCNT7, MBOAT4, GALNT17) which have been already

reported to link/associate with critical energetic deficiencies due to decreased glycolysis and lipid oxidation metabolism. Consistent with the above-mentioned expression levels, we observed increased metabolites in the cytosol, including carnitine, spermine, serine, alanine, glucose 6-phosphate, and acyl-CoA responsible for glycolytic and lipid metabolism-related energetic cascades. Also, our animal data confirmed the beneficial effects of our CMAs, showing the most substantial impacts on hyperemia, neuronal degeneration and necrosis in HFD animals.

A few limitations of the study need to be considered. First, the treatment effect was assessed by clinical evaluation and omics-analysis. Thus, our findings warrant a clinical trial with neuroimaging analysis to delineate the impact of CMAs on functional and structural brain alterations. Second, the link between systemic and CNS alterations and their relations to the AD pathology, i.e., amyloid and tau aggregation, has not been evaluated. Thus, further amyloid or tau-based neuroimaging combined with CSF evaluation should be pursued.

Further, CONSORT guidelines require comparison of outcomes between arms of the study, and this analysis showed no significant difference in the mean ADAS-Cog change in the CMA group compared to placebo. Therefore, the trial did not meet its primary endpoint. The trial conformed to all other CONSORT guidelines and the trend in ADAS-Cog improvement was reconciled by consideration of secondary endpoints.

In summary, the human phase 2 clinical study supports the data from animal models and genome-scale metabolic modelling suggesting that oral administration of metabolic activators can improve the mitochondrial dysfunction in AD patients. The safety profile of metabolic activators in these patients was consistent with the results of our previous one-day calibration study and clinical trials, including only a single component of the CMAs (***20***). Our present study showed that CMAs was safe and well-tolerated in patients with AD, and no major safety concerns were identified. Importantly, CMAs improved cognition and serum markers in these patients after only 12 weeks of treatment. These findings suggest that targeting multiple pathways by metabolic activators is a potentially effective therapeutic strategy for AD.

MATERIAL AND METHODS

Transcriptomics data analysis of the brain in AD patients

The mRNA expression profiles of 629 AD and 704 control samples were obtained from the Religious Orders Study and Rush Memory Aging Project (ROSMAP) datasets (***26***–***28***). The data has been normalized by quantile scaling, TMM normalization, Pareto scaling, and then limma *removeBatchEffect* (***121***). DESeq2 (***122***) was used to identify the differentially expressed genes (DEGs) with uniform size factors. Ensembl Biomart (***123***) was used to mapping different gene accession IDs or symbols. Gene set enrichment (GSE) was performed by using piano (***124***). The top 5% DEGs by DESeq2 statistic were accepted for GSE analysis. To identify the significantly changed metabolites in AD, we performed the Reporter metabolite analysis based on the RAVEN Toolbox 2.0 *reporterMetabolites* function (***125***), which uses DEGs information through the network topology of the reference metabolic model. The *iBrain2845* (***29***) genome-scale metabolic model was used as the reference metabolic model. KEGG Pathway Mapper (***126***) was used to identify predicted changed pathways based on maps M01200, M01212, and M01230

Animal study design

12-week-old female Sprague-Dawley rats (weighing 320-380g) were kept at a controlled temperature of 22 ± 2°C and a controlled humidity of 50 ± 5% on a 12-hour light/dark cycle. Food and water were available ad libitum. The animal experiments are performed at Medical Experimental Research Center, Atatürk University, Erzurum, Turkey. All experiments for the treatment of the animals were approved by the Ethics Committee of Atatürk University, and were conducted following the National Institutes of Health Guide for Care and Use of Laboratory Animals.

After acclimation to laboratory conditions for 1 week, rats were randomly divided into two dietary regimens receiving either a Chow diet (CHOW) or a high-fat diet (HFD. Group 1 (n=4) were fed with only regular CD for 5 weeks; Group 2 (n=4) were fed with CD and treated with STZ. The remaining animals in Group 3 (n=4) were fed with HFD for 3 weeks and sacrificed to verify the model development. Groups 4-10 were fed with HFD for 5 weeks. After 3 weeks, the animals in Groups 5-10 were treated with STZ and administered with individual or combined metabolic activators, including L-serine, NAC, LCAT and NR for 2 weeks.

Intracerebroventricular-streptozotocin (STZ) injection

Before surgical procedures, the rats were anesthetized by intraperitoneal administration of ketamine–xylazine (50 mg/kg ketamine and 5 mg/kg xylazine) and placed individually in the stereotaxic instrument (Stoelting, Illinois, USA). Stereotaxic coordinates for injection were 0.8 mm posterior to the bregma, 1.5 mm lateral to the sagittal suture and 3.6 mm below the brain surface (***127***). Then, 10 µL STZ (3 mg/kg, Sigma-Aldrich, Darmstadt, Germany) were injected over 3 min with a Hamilton micro-syringe into the bilateral ventricle. The injection needles were left in place for an additional 2 min to allow diffusion.

CMAs administration to animals

For carrying out the treatment experiments, the HFD rats were randomly separated into 9 (Groups 5-10) subgroups (n=4) and all treated with 3 mg/kg STZ (10 µL, icv). Group 5, treated only with STZ; Group 6, treated with STZ and serine (1000 mg/kg once daily by oral gavage); Group 7, treated with STZ and NAC (300 mg/kg once daily by oral gavage); Group 8, treated with STZ and LCAT (100 mg/kg once daily by oral gavage); Group 9, treated with STZ and NR (120 mg/kg once daily by oral gavage); Group 10, treated with STZ and serine (1000 mg/kg once daily by oral gavage), NAC (300 mg/kg once daily by oral gavage), LCAT (100 mg/kg/day) and NR (120 mg/kg once daily by oral gavage).

The body weight of the rats was recorded each week (Dataset S2). After 5 weeks, all animals were anesthetized with isoflurane and sacrificed. Blood samples were collected from the abdominal aorta and centrifuged at 8000 rpm for 15 min at 4°C for blood biochemistry analysis using automatic chemical analyser. The internal organs, including the heart, adipose tissues, liver, kidney, brain, muscle, intestine (duodenum, ileum, jejunum), pancreas, colon and stomach were immediately removed and then snap-frozen in liquid nitrogen and stored at -80°C.

Liver and brain tissue samples obtained due to the experimental procedure were fixed in 10% buffered formalin solution for 48 hours. Following the routine tissue procedure, the tissues were embedded in paraffin blocks and 4 µm thick sections were taken from each block. Preparations prepared for histopathological examination were stained with hematoxylin-eosin (HE) and examined with a light microscope (Olympus BX51, Germany). According to histopathological findings sections were evaluated by

independent pathologist and scored as absent (-), very mild (+), mild (++), moderate (+++) and severe (++++).

Immunohistochemical examination

Tissue sections taken on the adhesive (poly-L-lysine) slides for immunoperoxidase examination were deparaffinized and dehydrated. After washing, the tissues with suppressed endogenous peroxidase activity in 3% H_2O_2 were boiled in antigen retrieval solution. To prevent nonspecific background staining in the sections, protein block compatible with all primary and secondary antibodies was dropped and incubated for 5 minutes. Caspase 3 (Cat No: sc-56053 dilution ratio:1/100 US) for liver tissues and 8-OH-dG (Cat No: sc-66036 dilution ratio: 1/100 US) for brain tissues was used as the primary antibody. 3-3' Diaminobenzidine (DAB) chromogen was used as chromogen, and according to their immunopositivity, sections were evaluated by independent pathologist and scored as absent (-), very mild (+), mild (++), moderate (+++) and severe (++++).

Immunofluorescence examination

Tissue sections taken on the adhesive (poly-L-lysine) slides for immunoperoxidase examination were deparaffinized and dehydrated. After washing, the tissues with suppressed endogenous peroxidase activity in 3% H_2O_2 were boiled in antigen retrieval solution. To prevent nonspecific background staining in the sections, protein block compatible with all primary and secondary antibodies was dropped and incubated for 5 min. For liver tissues, primary antibody 8-OHdG (Cat No: sc-66036 dilution ratio: 1/100 US) was dropped and incubated at 37°C for 1 h. After washing, secondary FITC (Cat No: ab6717 dilution ratio: 1/500 UK) was dropped and incubated at 37°C for 30 min. The other primary antibody H2A.X (Cat No: I 0856-1 dilution ratio: 1/100 US) was dropped and incubated at 37°C for 1 h. After washing, secondary Texas Red (Cat No: sc-3917 dilution ratio: 1/100 US) was dropped and incubated at 37°C for 30 min. DAPI (Cat No: D-1306 dilution ratio:1/200 US) was dripped onto the washed tissues and incubated in the dark for 5 min, then glycerine was sealed. For brain tissues, primary antibody Caspase 3 (Cat No: sc-56053 dilution ratio: 1/100 US) was dropped and incubated for 1 h at 37°C. After washing, secondary FITC (Cat No: ab6717 dilution ratio: 1/500 UK) was dropped and incubated at 37°C for 30 min. The other primary antibody H2A.X (Cat No: I 0856-1 dilution ratio: 1/100 US) was dropped and incubated

at 37°C for 1 h. After washing, secondary Texas Red (Cat No: sc-3917 dilution ratio: 1/100 US) was dropped and incubated at 37°C for 30 min. DAPI (Cat No: D-1306 dilution ratio: 1/200 US) was dripped onto the washed tissues and incubated in the dark for 5 min, then glycerine was sealed. Sections were examined under a fluorescence microscope (Zeiss Germany) by an independent pathologist and, according to their immunopositivity evaluated as absent (-), very mild (+), mild (++), moderate (+++) and severe (++++).

Clinical Trial Design and Oversight

Patients for this randomized, double-blinded, placebo-controlled, phase 2 study were recruited at the Faculty of Medicine, Alanya Alaaddin Keykubat University, Antalya, Turkey and Faculty of Medicine, Istanbul Medipol University, Istanbul, Turkey. Written informed consent was obtained from all participants before the initiation of any trial-related procedures. The safety of the participants and the risk–benefit analysis was overseen by an independent external data-monitoring committee. The trial was conducted following Good Clinical Practice guidelines and the principles of the Declaration of Helsinki. The ethics committee approved the study of Istanbul Medipol University, Istanbul, Turkey, and retrospectively registered at https://clinicaltrials.gov/ with Clinical Trial ID: NCT04044131.

Participants

Patients were enrolled in the trial if they were over 50 years of age with mild to moderate AD according to ADAS-cog (AD Assessment Scale-cognitive subscale; ADAS≥12) and the Clinical Dementia Rating Scale Sum of Boxes (CDR-SOB; CDR≤2). Patients who had a history of stroke, severe brain trauma, toxic drug exposure were excluded. The main characteristics of the patients are summarized in Dataset S3. The inclusion, exclusion, and randomization criteria are described in detail in the Supplementary Appendix.

Randomization, Interventions, and Follow-up

Patients were randomly assigned to receive CMAs or placebo (2:1). Patient information (patient number, date of birth, initials) was entered into the web-based randomization system, and the randomization codes were entered into the electronic case report form. All clinical staff were blinded to treatment, as were the participants.

Treatment started on the day of diagnosis. Both placebo and CMAs were provided in powdered form in identical plastic bottles containing a single dose to be dissolved in water and taken orally one dose in the morning after breakfast and one dose in the evening after dinner. Each dose of CMAs contained 3.73 g L-carnitine tartrate, 2.55 g N-acetylcysteine, 1 g nicotinamide riboside chloride, and 12.35 g serine. All patients received one dose during the first 28 days and received two doses until Day 84. All patients came for a follow-up visit on Day 84. Further information is provided in the Supplementary Appendix.

Outcomes

The primary endpoint in the original protocol was to assess the clinical efficacy of CMAs in AD patients. For the primary purpose, the clinical differences in cognition of subjects receiving twelve-week treatment either with metabolic activators supplementation or placebo were determined. The primary analysis was on the difference in cognitive and daily living activity scores between the placebo and the treatment arms, which were assessed by Mini-Mental State Examination (MMSE), AD Assessment Scale-cognitive subscale (ADAS-Cog) and AD Cooperative Study - Activities of Daily Living (ADCS-ADL) in AD patients. The secondary aim in this study was to evaluate the safety and tolerability of CMAs. All protocol amendments were authorized and approved by the sponsor, the institutional review board or independent ethics committee, and the pertinent regulatory authorities.

Number and characteristics of adverse events, serious adverse events, and treatment discontinuation due to CMAs were reported from the beginning of the study to the end of the follow-up period as key safety endpoints. The changes in vital signs baseline values, and the status of treatment were recorded at Day 0 and 84. A complete list of the end points is provided in the Supplementary Appendix.

Proteomics Analysis

Plasma levels of proteins were determined with the Olink panel (Olink Bioscience, Uppsala, Sweden). Briefly each sample was incubated with DNA-labelled antibody pairs (proximity probes). When an antibody pair binds to its corresponding antigens, the corresponding DNA tails form an amplicon by proximity extension, which can be quantified by high-throughput, real-time PCR. Probe solution (3 µl) was mixed with 1 µl of sample and incubated overnight at 4°C. Then 96 µl of extension solution containing

extension enzyme and PCR reagents for the pre-amplification step was added, and the extension products were mixed with detection reagents and primers and loaded on the chip for qPCR analysis with the BioMark HD System (Fluidigm Corporation, South San Francisco, CA). To minimize inter- and intrarun variation, the data were normalized to both an internal control and an interplate control. Normalized data were expressed in arbitrary units (Normalized Protein eXpression, NPX) on a log2 scale and linearized with the formula 2NPX. A high NPX indicates a high protein concentration. The limit of detection, determined for each of the assays, was defined as three standard deviations above the negative control (background).

Untargeted Metabolomics Analysis

Plasma samples were collected on Days 0 and 84 for nontargeted metabolite profiling by Metabolon (Durham, NC). The samples were prepared with an automated system (MicroLab STAR, Hamilton Company, Reno, NV). For quality control purposes, a recovery standard was added before the first step of the extraction. To remove protein and dissociated small molecules bound to protein or trapped in the precipitated protein matrix, and to recover chemically diverse metabolites, proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) and centrifuged. The resulting extract was divided into four fractions: one each for analysis by ultraperformance liquid chromatography– tandem mass spectrometry (UPLC-MS/MS) with positive ion-mode electrospray ionization, UPLC-MS/MS with negative ion-mode electrospray ionization, and gas chromatography– mass spectrometry; one fraction was reserved as a backup.

Determination of Clinical Variables Informing Response to CMA Administration

The patient groups with low and high levels of each clinical parameter were established based on the median score for that clinical parameter across all patients on Day 0. Patients scoring at or below the median were placed in the low group; patients scoring above the median were placed in the high group. ADAS-Cog scores were measured over different time points and statistical significance was tested between time points by using a paired t-test. Clinical parameters were deemed informative for the response to CMAs if exactly one group (low or high) exhibited more statistically significant changes in ADAS-Cog in the CMA group than in the placebo group.

Statistical Analysis

Paired t-test was used to identify the differences in clinical parameters between time points and one-way ANOVA was used to find the shifts between CMA and placebo groups at each time point. For analysis of plasma metabolomics, we removed the metabolite profiles with more than 50% missing values across all samples. Metabolite changes between time points were analysed by paired t-test. Metabolite changes between CMA and placebo groups were analysed by one-way ANOVA. Missing values were removed in pairwise comparison. The p-values were adjusted by Benjamini & Hochberg method. Metabolites with a false-discovery rate of 5% were considered statistically significant. Two-sided Student's t-tests were used for statistical analyses of plasma parameters in animal model, statistical significance was considered $p < 0.05$.

For analysis of plasma proteomics, we removed the protein profiles with more than 50% missing values across all samples. Paired t-tests were used to identify the changes between time points and one-way ANOVA was used to identify the changes between different groups. $p<0.01$ was considered statistically significant. Spearman correlation analysis was used to analyse the association between CMAs and clinical parameters or metabolomics or proteomics.

Generation of Multi-Omics Network

Multi-omics network was generated based on the Spearman correlations and the significant associations (FDR < 0.05) are presented. The analyses were performed with SciPy package in Python 3.7. Centrality analysis on the network was performed using igraph Python.

ACKNOWLEDGMENTS

This work was financially supported by ScandiBio Therapeutics and Knut and Alice Wallenberg Foundation. The authors would like to thank the Metabolon Inc. (Durham, USA) for generation of metabolomics data, and ChromaDex Inc. (Irvine, CA, USA) for providing NR. AM and HY acknowledge support from the PoLiMeR Innovative Training Network (Marie Skłodowska-Curie Grant Agreement No. 812616) which has received funding from the European Union's Horizon 2020 research and innovation programme.

CONFLICT OF INTEREST

AM, JB and MU are the founder and shareholders of ScandiBio Therapeutics. The other authors declare no competing interests.

SUPPORTING INFORMATION

Supporting Information includes 8 Supplementary Datasets.

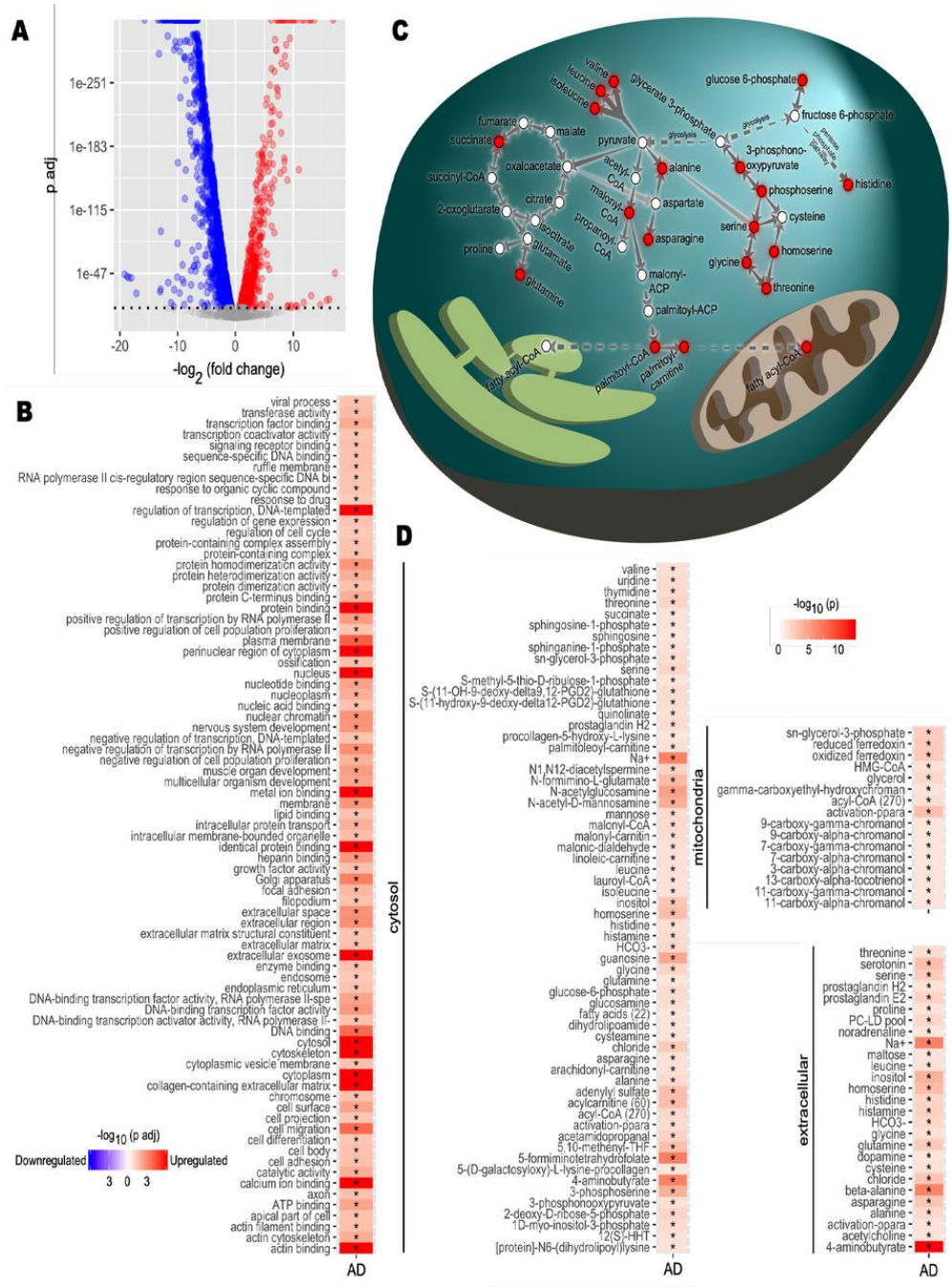


Figure 1. Transcriptomic analysis of brain tissue samples in ROSMAP datasets. A) Differential expression analysis for Alzheimer's disease (AD) brain samples compared to healthy controls was performed. Differentially expressed genes (DEGs) were determined from gene expression values. DEGs with a p-value of 1×10^{-10} or smaller after Benjamini-

Hochberg adjustment were determined statistically significant. Each point represents one gene. Red, significantly upregulated genes; blue, significantly downregulated genes; grey, not significant. B) Functional enrichment analysis was performed. Gene set enrichment analysis was applied on the DEGs to determine upregulated and downregulated GO terms compared to controls. Colour scale indicates direction of enrichment and p-value after Benjamini-Hochberg adjustment. Significance code: *, Adj.p< 0.05. C) Carbon metabolism pathway analysis. Alterations to pathways were inferred from reporter metabolite analysis. Key reactions and pathways linking reporter metabolites (red nodes) are shown. Reactions are simplified and arrows may represent multiple reactions. D) Reporter metabolite analysis. Statistics from DEG analysis were used to infer altered metabolites based on iBrain2845, a functional genome-scale metabolic model of brain. Asterisks indicate statistical significance based on Student's t-test. P value <0.05.

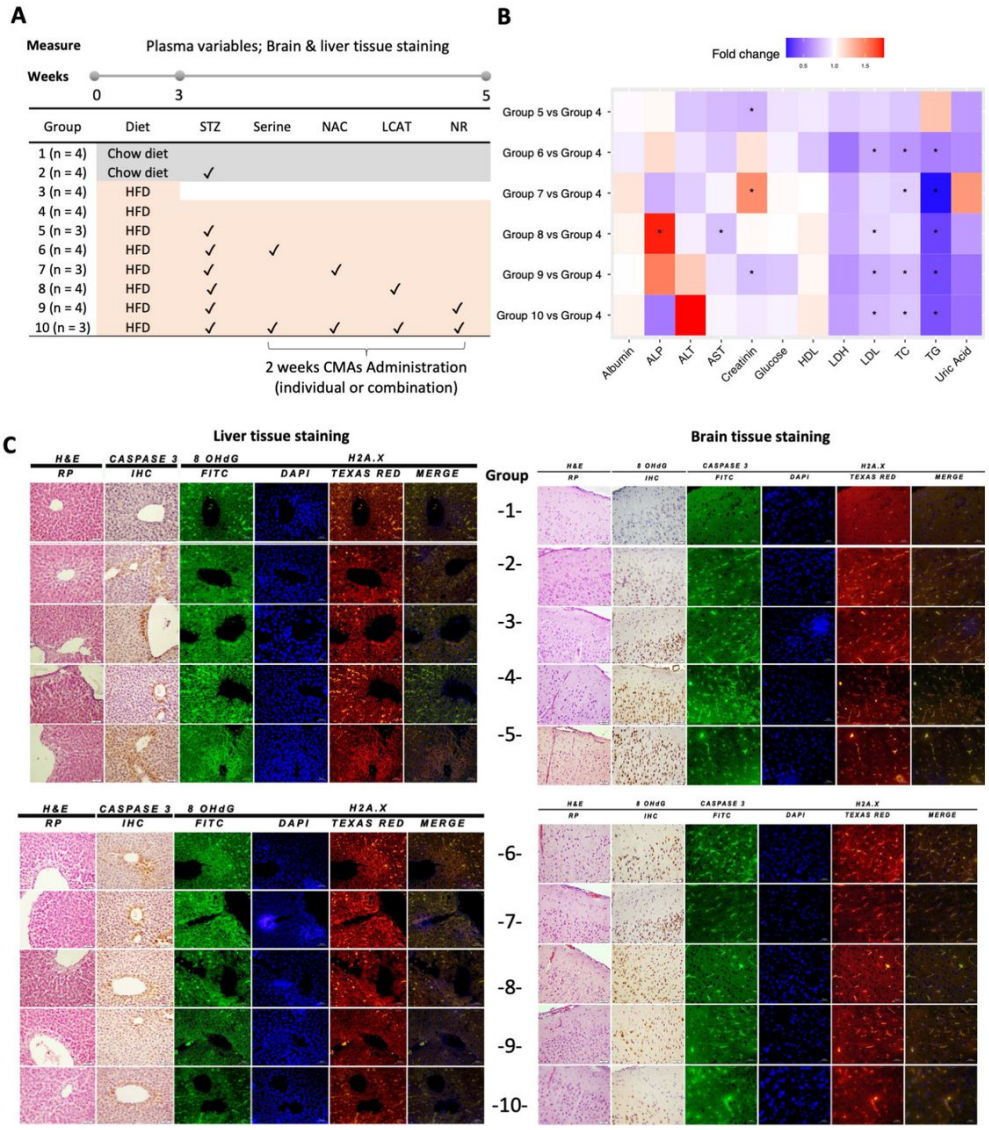


Figure 2. Animal study. A) Rat animal groups in in vivo experiments. Group 1 (n=4) were fed with only regular Chow diet (CHOW) for 5 weeks; Group 2 (n=4) were fed with CHOW and treated with streptozotocin (STZ). The remaining animals in Group 3 (n=4) were fed with high fat diet (HFD) for 3 weeks and sacrificed to verify model development. Groups 4-10 were fed with HFD for 5 weeks. After 3 weeks, the animals in Groups 5-10 were treated with STZ and administered with individual or combined metabolic activators, including L-serine, NAC, LCAT and NR for 2 weeks. B) Heatmap shows FC based alterations of the clinical variables in the rat study groups. Asterisks indicate statistical significance based on Student's t-test. P value <0.05. TG, triglyceride; TC, total cholesterol; ALP, alkaline phosphatase; AST, aspartate aminotransferase; ALT, alanine aminotransferase; HDL, high-density lipoprotein;

LDL, low-density lipoprotein; LDH, lactate dehydrogenase. C) Histopathological, immunohistochemical (caspase 3) and immunofluorescence (8-OHdG and H2A.X) images of rat brain (right side) and liver tissue (left side). Slides evaluated by independent pathologist and immunopositivity scores were: absent (-), very mild (+), mild (++), moderate (+++) and severe (++++).

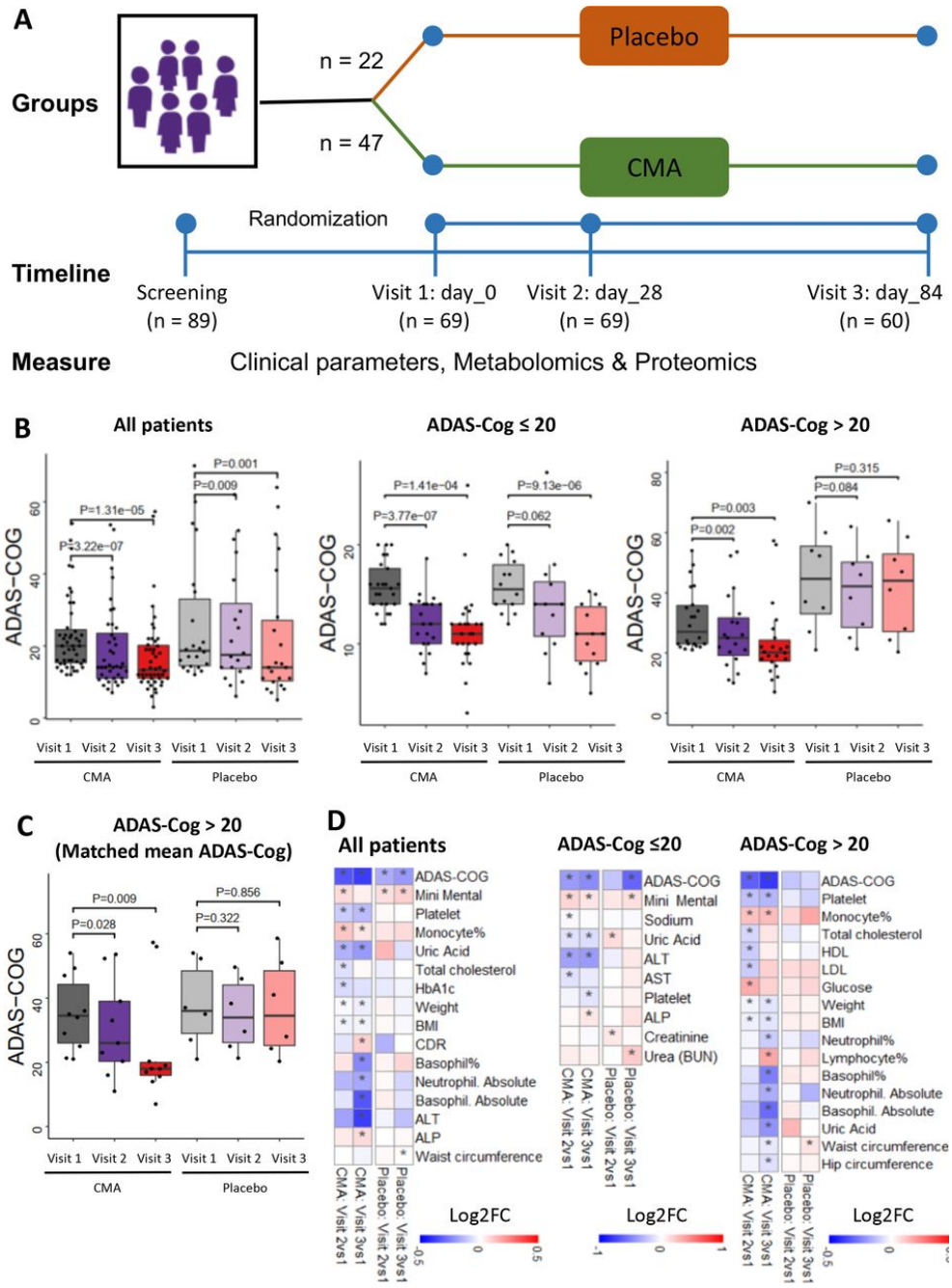


Figure 3. CMAs Improve ADAS-Cog scores and clinical parameters. A) Study design for testing the effects of CMAs in AD patients. B) Differences in ADAS-Cog scores in the CMA and placebo groups on Days 0, 28 and 84 are presented. Additionally, differences in ADAS-Cog scores were analysed by stratifying the patients into high and low levels of ADAS-Cog groups (> 20 ADAS-Cog is high, ≤ 20 is low). As decreased ADAS-Cog score is the indicator of the improved cognitive function in AD patients, the ADAS-Cog scores is significantly

decreased on Day 28 vs Day 0 (Log2FoldChange (FC)= -0.33, (26% improvement), p-value=0.0000003) and Day 84 vs Day 0 (Log2FC= -0.37, (29% improvement), p-value=0.00001) in the CMA group. A slight but significant decrease was found in the placebo group on Day 28 vs Day 0 (Log2FC= -0.16, (12% improvement), p-value=0.009) and Day 84 vs Day 0 (Log2FC= -0.19, (14% improvement), p-value=0.001) due to the recommendations of exercise and Mediterranean diet to all AD patients participated in the trial. The differences between clinical parameters have also been analysed by stratifying the patients into low-scoring (mild patients) and high-scoring (severe patients) ADAS-Cog groups (>20ADAS-Cog score is high, ≤ 0 is low). The ADAS-Cog scores were significantly decreased on Day 28 vs Day 0 (Log2FC= -0.31, (24% improvement), p-value=0.002) and Day 84 vs Day 0 (Log2FC=-0.38, (30% improvement), p-value=0.003) in the high scoring CMA group and no significance difference in the high-scoring placebo (p>0.05 in both time points) group. C) We selected ten patients from the severe (ADAS-COG > 20) CMA group with matched ADAS-COG values to the placebo group (p-value=0.693) and present the ADAS-Cog scores. We recalculated the differences in ADAS-COG scores, and found significant improvement in the CMA group whereas no significant difference in the placebo group. D) Heatmaps shows log2FC based alterations of the clinical variables, which are compared before and after the administration of CMA in both drug and placebo groups. Asterisks indicate statistical significance based on Student's t-test. p-value <0.05. Log2FC: log2(fold change).

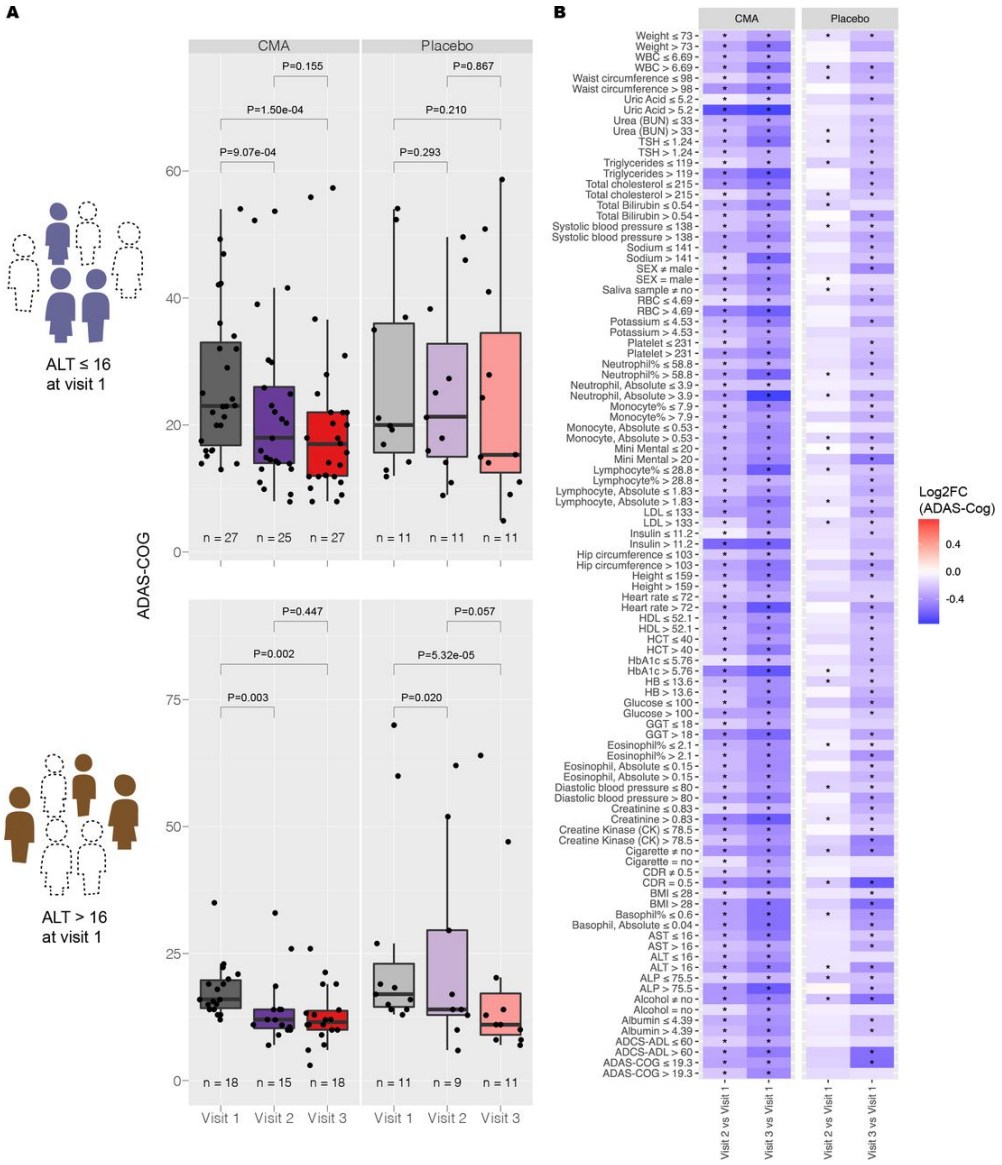


Figure 4. Identification of clinical variables informative for response to CMA administration. A) Distribution of ADAS-Cog scores over visit number for patients with ALT ≤ 16 IU/L at visit 1 (upper panel) and patients with ALT > 16 IU/L at visit 1 (lower panel). B) Between-visit changes to ADAS-Cog with various clinical variable groupings are shown. Only those clinical variable groupings resulting in a more significant change to ADAS-Cog in CMA group compared to placebo group (a p-value of 0.05 or better are shown). Colour scale indicates log2 fold change to ADAS-Cog between visits. Statistical significance between visits was determined by a paired t-test across individuals who attended both visits. Asterisks indicate statistical significance $p < 0.05$.

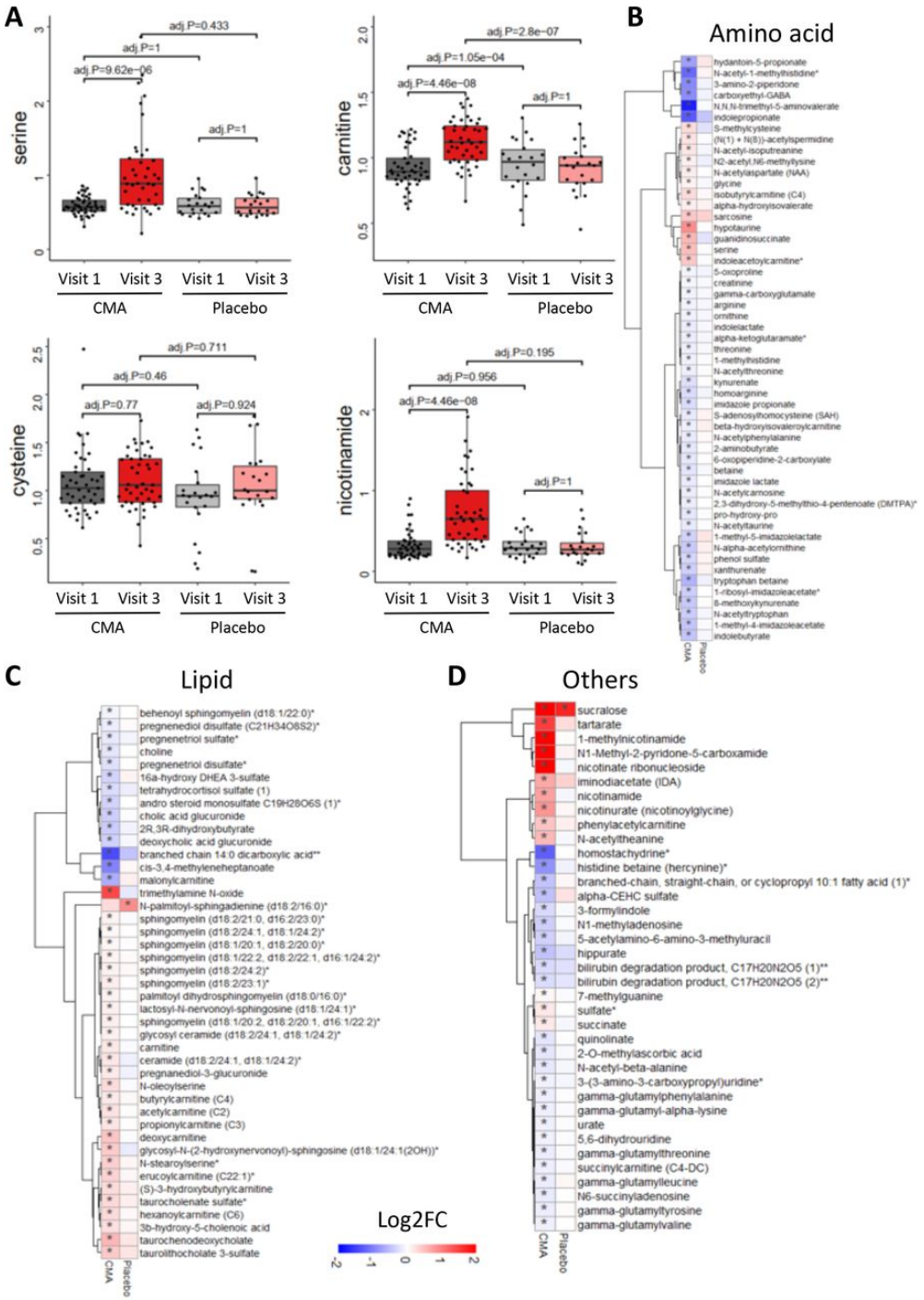


Figure 5. CMAs alter plasma metabolite levels. A) Differences in the plasma levels of individual CMAs including serine, carnitine, cysteine and nicotinamide are shown in the CMA and placebo groups on Days 0 and 84. Plasma level of B) amino acids, C) lipids and D) other metabolites that are significantly different between Day 84 vs Day 0 in the CMA and

placebo groups are presented. Adj. p< 0.05. Heatmap shows log2FC values of metabolites between Day 84 vs Day 0. Asterisks indicate statistical significance based on paired Student's t test. Adj.p< 0.05. Log2FC: log2(fold change).

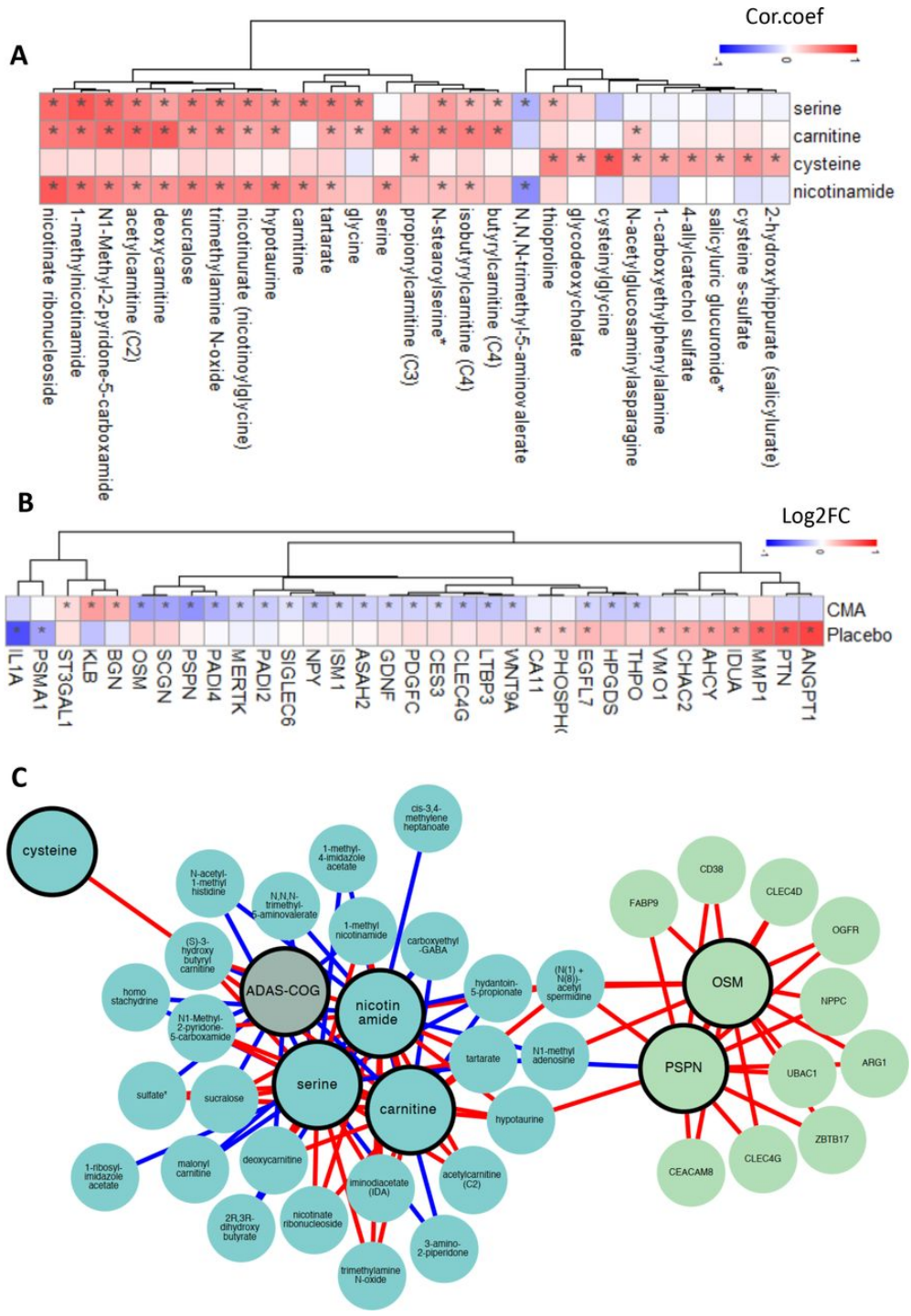


Figure 6. Correlation of CMAs with plasma metabolites and altered plasma protein levels. A) Associations between the plasma level of individual CMAs and the ten most significantly correlated plasma metabolites are presented. Asterisks indicate statistical significance (Adj.p < 0.05) based on Spearman correlation analysis. Cor.Coeff: Correlation coefficient B) Heatmap

shows log2FC based alterations between the significantly different proteins on Day 84 vs Day 0 in the CMA and placebo groups. Asterisks indicate statistical significance based on paired Student's t test. $p < 0.01$. C) Integrated multi-omics data based on network analysis represents the neighbours of the CMAs, including serine, carnitine, nicotinamide and cysteine, and ADAS-Cog scores. Only analytes that are significantly altered in CMA Day 84 vs Day 0 are highlighted.

Table 1. Scoring histopathological, immunohistochemical and immunofluorescence findings in brain tissues and liver tissues.

Organ	Groups	Hyperemia	Degeneration in Neurons	Necrosis in Neurons	8-OHdG (IHC	Caspase 3 (IFA FITC)	H2A.X (IFA Texas Red)
BRAIN	CHOW	-	-	-	-	-	-
	CHOW + STZ	+++	+++	+++	+++	+++	+++
	HFD 3w	+++	+++	+++	+++	+++	+++
	HFD 5w	++++	++++	++++	++++	++++	++++
	HFD+STZ	++++	++++	++++	++++	++++	++++
	HFD+STZ+Serine	+++	+++	++	+++	+++	+++
	HFD+STZ+NAC	++++	++++	+++	+++	+++	+++
	HFD+STZ+LCAT	+++	+++	++	+++	+++	+++
	HFD+STZ+NR	+++	+++	++	+++	+++	+++
	HFD+STZ+Serine+NAC+LCAT+NR	++	++	+	++	++	++
LIVER	CHOW	-	-	-	-	-	-
	CHOW + STZ	+++	+++	+++	+++	+++	+++
	HFD 3w	+++	+++	+++	+++	+++	+++
	HFD 5w	++++	++++	++++	++++	++++	++++
	HFD+STZ	++++	++++	++++	++++	++++	++++
	HFD+STZ+Serine	+++	+++	+++	+++	+++	+++
	HFD+STZ+NAC	+++	++++	+++	+++	+++	+++
	HFD+STZ+LCAT	+++	+++	+++	+++	+++	+++
	HFD+STZ+NR	+++	+++	++	+++	+++	+++
	HFD+STZ+Serine+NAC+LCAT+NR	+++	+++	+	+	++	++

REFERENCES

1. L. Trujillo-Estrada et al., In vivo modification of Abeta plaque toxicity as a novel neuroprotective lithium-mediated therapy for Alzheimer's disease pathology. *Acta Neuropathol Commun* **1**, 73 (2013).
2. A. Nunomura, G. Perry, RNA and Oxidative Stress in Alzheimer's Disease: Focus on microRNAs. *Oxidative Medicine and Cellular Longevity 2020*, **2638130** (2020).
3. S. Lam et al., A systems biology approach for studying neurodegenerative diseases. *Drug Discov Today* **25**, 1146–1159 (2020).
4. J. G. Mielke et al., A biochemical and functional characterization of diet-induced brain insulin resistance. *J Neurochem* **93**, 1568–1578 (2005).
5. S. E. Arnold et al., Brain insulin resistance in type 2 diabetes and Alzheimer disease: concepts and conundrums. *Nat Rev Neurol* **14**, 168–181 (2018).
6. Y. Wei et al., Ribosylation triggering Alzheimer's disease-like Tau hyperphosphorylation via activation of CaMKII. *Aging Cell* **14**, 754–763 (2015).
7. F. Masciopinto et al., Effects of long-term treatment with pioglitazone on cognition and glucose metabolism of PS1-KI, 3xTg-AD, and wild-type mice. *Cell Death Dis* **3**, e448–e448 (2012).
8. L. R. Wong, P. Wong, P. C. Ho, Metabolic Profiling of Female Tg2576 Mouse Brains Provides Novel Evidence Supporting Intranasal Low-Dose Pioglitazone for Long-Term Treatment at an Early Stage of Alzheimer's Disease. *Biomedicines* **8**, (2020).
9. S. Makin, The amyloid hypothesis on trial. *Nature* **559**, S4–s7 (2018).
10. B. M. Kuehn, In Alzheimer Research, Glucose Metabolism Moves to Center Stage. *JAMA* **323**, 297–299 (2020).
11. Z. Chen, C. Zhong, Decoding Alzheimer's disease from perturbed cerebral glucose metabolism: implications for diagnostic and therapeutic strategies. *Prog Neurobiol* **108**, 21–43 (2013).
12. C. L. Powell, A. R. Davidson, A. M. Brown, Universal Glia to Neurone Lactate Transfer in the Nervous System: Physiological Functions and Pathological Consequences. *Biosensors (Basel)* **10**, (2020).
13. D. Drulis-Fajdasz, A. Gizak, T. Wójtowicz, J. R. Wiśniewski, D. Rakus, Aging-associated changes in hippocampal glycogen metabolism in mice. Evidence for and against astrocyte-to-neuron lactate shuttle. *Glia* **66**, 1481–1495 (2018).
14. J. Cummings, G. Lee, A. Ritter, M. Sabbagh, K. Zhong, Alzheimer's disease drug development pipeline: 2020. *Alzheimers Dement (N Y)* **6**, e12050 (2020).
15. X. Pan et al., Alzheimer's disease-like pathology has transient effects on the brain and blood metabolome. *Neurobiol Aging* **38**, 151–163 (2016).
16. Y. Dong, G. J. Brewer, Global Metabolic Shifts in Age and Alzheimer's Disease Mouse Brains Pivot at NAD+/NADH Redox Sites. *J Alzheimers Dis* **71**, 119–140 (2019).
17. Y. Hou et al., NAD+ supplementation normalizes key Alzheimer's features and DNA damage responses in a new AD mouse model with introduced DNA repair deficiency. *Proceedings of the National Academy of Sciences* **115**, E1876–E1885 (2018).
18. J. B. Toledo et al., Metabolic network failures in Alzheimer's disease: A biochemical road map. *Alzheimers Dement* **13**, 965–984 (2017).
19. G. M. Sancesario, S. Bernardini, Alzheimer's disease in the omics era. *Clin Biochem* **59**, 9–16 (2018).
20. C. Zhang et al., The acute effect of metabolic cofactor supplementation: a potential therapeutic strategy against non-alcoholic fatty liver disease. *Mol Syst Biol* **16**, e9495 (2020).
21. A. Mardinoglu et al., The Potential Use of Metabolic Cofactors in Treatment of NAFLD. *Nutrients* **11**, 1578 (2019).
22. A. Mardinoglu et al., Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat Commun* **5**, 3083 (2014).

23. A. Mardinoglu, J. Boren, U. Smith, M. Uhlen, J. Nielsen, Systems biology in hepatology: approaches and applications. *Nature Reviews Gastroenterology & Hepatology* **15**, 365–377 (2018).
24. O. Altay et al., Combined Metabolic Activators Accelerates Recovery in Mild-to-Moderate COVID-19. *Advanced Science n/a*, **2101222** (2021).
25. A. Mardinoglu et al., An Integrated Understanding of the Rapid Metabolic Benefits of a Carbohydrate-Restricted Diet on Hepatic Steatosis in Humans. *Cell Metab* **27**, 559–571 *e555* (2018).
26. A. J. Myers et al., A survey of genetic human cortical gene expression. *Nature Genetics* **39**, 1494–1499 (2007).
27. J. A. Webster et al., Genetic control of human brain transcript expression in Alzheimer disease. *Am J Hum Genet* **84**, 445–458 (2009).
28. S. Mostafavi et al., A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat Neurosci* **21**, 811–819 (2018).
29. S. Lam, *et al.*, Systems analysis reveals ageing-related perturbations in retinoids and sex hormones in Alzheimer's and Parkinson's diseases. *bioRxiv, 2021.2006.2010.447367* (2021).
30. A. Mardinoglu et al., Integration of clinical data with a genome-scale metabolic model of the human adipocyte. *Molecular systems biology* **9**, 649–649 (2013).
31. M. Zeybel, et al., Combined Metabolic Activators Reduces Liver Fat in Nonalcoholic Fatty Liver Disease Patients. *medRxiv, 2021.2005.2020.21257480* (2021).
32. J. R. Moffett, P. Arun, P. S. Ariyannur, A. M. A. Namboodiri, N-Acetylaspartate reductions in brain injury: impact on post-injury neuroenergetics, lipid synthesis, and protein acetylation. *Front Neuroenergetics* **5**, 11–11 (2013).
33. S. P. Singh, V. Singh, Meta-analysis of the efficacy of adjunctive NMDA receptor modulators in chronic schizophrenia. *CNS Drugs* **25**, 859–885 (2011).
34. Y. Chen, G. J. Guillemin, Kynurenine pathway metabolites in humans: disease and healthy States. *Int J Tryptophan Res* **2**, 1–19 (2009).
35. C. Tapia-Rojas et al., Is L-methionine a trigger factor for Alzheimer's-like neurodegeneration?: Changes in Aβ oligomers, tau phosphorylation, synaptic proteins, Wnt signaling and behavioral impairment in wild-type mice. *Mol Neurodegener* **10**, 62–62 (2015).
36. Q. Wang, D. Liu, P. Song, M.-H. Zou, Tryptophan-kynurenine pathway is dysregulated in inflammation, and immune activation. *Frontiers in bioscience (Landmark edition)* **20**, 1116–1143 (2015).
37. K. Sas, E. Szabó, L. Vécsei, Mitochondria, Oxidative Stress and the Kynurenine System, with a Focus on Ageing and Neuroprotection. *Molecules* **23**, 191 (2018).
38. F. Fazio et al., Vasorelaxing Action of the Kynurenine Metabolite, Xanthurenic Acid: The Missing Link in Endotoxin-Induced Hypotension? *Front Pharmacol* **8**, 214 (2017).
39. Y. Shi, Z. Liu, Y. Shen, H. Zhu, A Novel Perspective Linkage Between Kidney Function and Alzheimer's Disease. *Front Cell Neurosci* **12**, 384 (2018).
40. J. K. Haukka et al., Metabolomic Profile Predicts Development of Microalbuminuria in Individuals with Type 1 Diabetes. *Scientific Reports* **8**, 13853 (2018).
41. M. M. Mielke et al., Plasma sphingomyelins are associated with cognitive progression in Alzheimer's disease. *Journal of Alzheimer's disease : JAD* **27**, 259–269 (2011).
42. K. A. Jhang, J. S. Park, H. S. Kim, Y. H. Chong, Sulforaphane rescues amyloid-β peptide-mediated decrease in MerTK expression through its anti-inflammatory effect in human THP-1 macrophages. *J Neuroinflammation* **15**, 75 (2018).
43. G. Tondo, D. Perani, C. Comi, TAM Receptor Pathways at the Crossroads of Neuroinflammation and Neurodegeneration. *Dis Markers 2019*, **2387614** (2019).
44. H. C. Chiang, L. Wang, Z. Xie, A. Yau, Y. Zhong, PI3 kinase signaling is involved in Abeta-induced memory loss in Drosophila. *Proc Natl Acad Sci U S A* **107**, 7060–7065 (2010).
45. L. Wang et al., Epidermal growth factor receptor is a preferred target for treating Amyloid-β–induced memory loss. *Proceedings of the National Academy of Sciences* **109**, 16743 (2012).

46. T. Owens, T. Renno, V. Taupin, M. Krakowski, Inflammatory cytokines in the brain: does the CNS shape immune responses? *Immunol Today* **15**, 566–571 (1994).
47. T. Kordula et al., Oncostatin M and the interleukin-6 and soluble interleukin-6 receptor complex regulate alpha1-antichymotrypsin expression in human cortical astrocytes. *J Biol Chem* **273**, 4112–4118 (1998).
48. J. V. Castell et al., Interleukin-6 is the major regulator of acute phase protein synthesis in adult human hepatocytes. *FEBS Lett* **242**, 237–239 (1989).
49. C. R. Abraham, D. J. Selkoe, H. Potter, Immunochemical identification of the serine protease inhibitor alpha 1-antichymotrypsin in the brain amyloid deposits of Alzheimer's disease. *Cell* **52**, 487–501 (1988).
50. D. J. Selkoe, The molecular pathology of Alzheimer's disease. *Neuron* **6**, 487–498 (1991).
51. R. Tu, H. M. Grover, L. P. Kotra, Peptidyl Arginine Deiminases *and* Neurodegenerative Diseases. Curr Med Chem 23, 104–114 (2016).
52. N. K. Acharya et al., Neuronal PAD4 expression and protein citrullination: possible role in production of autoantibodies associated with neurodegenerative disease. *J Autoimmun* **38**, 369–380 (2012).
53. T. Hamaguchi et al., Association of a polymorphism of the transforming growth factor-beta1 gene with cerebral amyloid angiopathy. *J Neurol Neurosurg Psychiatry* **76**, 696–699 (2005).
54. D. K. Lahiri, Y. W. Ge, Role of the APP promoter in Alzheimer's disease: cell type-specific expression of the beta-amyloid precursor protein. *Ann N Y Acad Sci* **1030**, 310–316 (2004).
55. E. K. Luedecking, S. T. DeKosky, H. Mehdi, M. Ganguli, M. I. Kamboh, Analysis of genetic polymorphisms in the transforming growth factor-beta1 gene and the risk of Alzheimer's disease. *Hum Genet* **106**, 565–569 (2000).
56. T. Burton, B. Liang, A. Dibrov, F. Amara, Transforming growth factor-beta-induced transcription of the Alzheimer beta-amyloid precursor protein gene involves interaction between the CTCF-complex and Smads. *Biochem Biophys Res Commun* **295**, 713–723 (2002).
57. S. Lesné et al., Transforming growth factor-beta 1 potentiates amyloid-beta generation in astrocytes and in transgenic mice. *J Biol Chem* **278**, 18408–18418 (2003).
58. H. Ehrenreich et al., A hematopoietic growth factor, thrombopoietin, has a proapoptotic role in the brain. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 862–867 (2005).
59. H. G. Woo, Y. Chang, D. R. Ryu, T. J. Song, Plasma Klotho concentration is associated with the presence, burden and progression of cerebral small vessel disease in patients with acute ischaemic stroke. *PLoS One* **14**, e0220796 (2019).
60. G. Paroni et al., Klotho at the Edge of Alzheimer's Disease and Senile Depression. *Mol Neurobiol* **56**, 1908–1920 (2019).
61. K. Yang, Z. Yang, X. Chen, W. Li, The significance of sialylation on the pathogenesis of Alzheimer's disease. *Brain Research Bulletin* **173**, 116–123 (2021).
62. I. Mohri et al., Hematopoietic prostaglandin D synthase and DP1 receptor are selectively upregulated in microglia and astrocytes within senile plaques from human patients and in a mouse model of Alzheimer disease. *J Neuropathol Exp Neurol* **66**, 469–480 (2007).
63. E. Tarkowski et al., Increased intrathecal levels of the angiogenic factors VEGF and TGF-beta in Alzheimer's disease and vascular dementia. *Neurobiol Aging* **23**, 237–243 (2002).
64. C. C. Chao et al., Serum cytokine levels in patients with Alzheimer's disease. *Clin Diagn Lab Immunol* **1**, 433–436 (1994).
65. D. R. Royall, R. F. Palmer, δ scores predict mild cognitive impairment and Alzheimer's disease conversions from nondemented states. *Alzheimers Dement (Amst)* **6**, 214–221 (2017).
66. V. B. Gupta et al., Altered levels of blood proteins in Alzheimer's disease longitudinal study: Results from Australian Imaging Biomarkers Lifestyle Study of Ageing cohort. *Alzheimers Dement (Amst)* **8**, 60–72 (2017).

67. R. Taliyan, S. K. Chandran, V. Kakoty, Therapeutic Approaches to Alzheimer's Type of Dementia: A Focus on FGF21 Mediated Neuroprotection. *Curr Pharm Des* **25**, 2555–2568 (2019).
68. G. Brombo et al., Lower Plasma Klotho Concentrations Are Associated with Vascular Dementia but Not Late-Onset Alzheimer's Disease. *Gerontology* **64**, 414–421 (2018).
69. A. Salminen, P. Jouhten, T. Sarajärvi, A. Haapasalo, M. Hiltunen, Hypoxia and GABA shunt activation in the pathogenesis of Alzheimer's disease. *Neurochem Int* **92**, 13–24 (2016).
70. Y. Hasin, M. Seldin, A. Lusis, Multi-omics approaches to disease. *Genome Biol* **18**, 83 (2017).
71. M. Arif et al., iNetModels 2.0: an interactive visualization and database of multi-omics data. *Nucleic Acids Res*, (2021).
72. B. Hutter-Schmid, C. Humpel, Alpha-Smooth Muscle Actin mRNA and Protein Are Increased in Isolated Brain Vessel Extracts of Alzheimer Mice. *Pharmacology* **98**, 251–260 (2016).
73. D. Åberg et al., Increased Cerebrospinal Fluid Level of Insulin-like Growth Factor-II in Male Patients with Alzheimer's Disease. *J Alzheimers Dis* **48**, 637–646 (2015).
74. B. Yulug, et al., Combined Metabolic Activators Improve Cognitive Functions without Altering Motor Scores in Parkinson's Disease. *medRxiv, 2021.2007.2028.21261293* (2021).
75. S. Martire et al., Bioenergetic Impairment in Animal and Cellular Models of Alzheimer's Disease: PARP-1 Inhibition Rescues Metabolic Dysfunctions. *J Alzheimers Dis* **54**, 307–324 (2016).
76. S. Martire, L. Mosca, M. d'Erme, PARP-1 involvement in neurodegeneration: A focus on Alzheimer's and Parkinson's diseases. *Mech Ageing Dev* 146–148, *53-64* (2015).
77. J. Le Douce et al., Impairment of Glycolysis-Derived l-Serine Production in Astrocytes Contributes to Cognitive Deficits in Alzheimer's Disease. *Cell Metab* **31**, 503–517.*e508* (2020).
78. R. Bavarsad Shahripour, M. R. Harrigan, A. V. Alexandrov, N-acetylcysteine (NAC) in neurological disorders: mechanisms of action and therapeutic opportunities. *Brain Behav* **4**, 108–122 (2014).
79. G. Tardiolo, P. Bramanti, E. Mazzon, Overview on the Effects of N-Acetylcysteine in Neurodegenerative Diseases. *Molecules* **23**, (2018).
80. Y. Hara, N. McKeehan, P. A. Dacks, H. M. Fillit, Evaluation of the Neuroprotective Potential of N-Acetylcysteine for Prevention and Treatment of Cognitive Aging and Dementia. *J Prev Alzheimers Dis* **4**, 201–206 (2017).
81. A. Kepka et al., Preventive Role of L-Carnitine and Balanced Diet in Alzheimer's Disease. *Nutrients* **12**, *1987* (2020).
82. J. L. Flanagan, P. A. Simmons, J. Vehige, M. D. Willcox, Q. Garrett, Role of carnitine in disease. *Nutr Metab (Lond)* **7**, 30–30 (2010).
83. A. Cristofano et al., Serum Levels of Acyl-Carnitines along the Continuum from Normal to Alzheimer's Dementia. *PLoS One* **11**, e0155694 (2016).
84. N. Chen et al., L-carnitine for cognitive enhancement in people without cognitive impairment. *Cochrane Database Syst Rev* **3**, *CD009374-CD009374* (2017).
85. K. A. Wollen, Alzheimer's disease: the pros and cons of pharmaceutical, nutritional, botanical, and stimulatory therapies, with a discussion of treatment strategies from the perspective of patients and practitioners. *Altern Med Rev* **15**, 223–244 (2010).
86. S. I. Gavrilova, B. Kalyn Ia, I. V. Kolykhalov, I. F. Roshchina, N. D. Selezneva, [Acetyl-L-carnitine (carnicetine) in the treatment of early stages of Alzheimer's disease and vascular dementia]. *Zh Nevrol Psikhiatr Im S S Korsakova* **111**, 16–22 (2011).
87. K. K. Ting, B. Brew, G. Guillemin, The involvement of astrocytes and kynurenine pathway in Alzheimer's disease. *Neurotox Res* **12**, 247–262 (2007).
88. K. O'Farrell, A. Harkin, Stress-related regulation of the kynurenine pathway: Relevance to neuropsychiatric and degenerative disorders. *Neuropharmacology* **112**, 307–323 (2017).
89. G. J. Guillemin, G. A. Smythe, L. A. Veas, O. Takikawa, B. J. Brew, A beta 1-42 induces production of quinolinic acid by human macrophages and microglia. *Neuroreport* **14**, 2311–2315 (2003).

90. W. Wu et al., Expression of tryptophan 2,3-dioxygenase and production of kynurenine pathway metabolites in triple transgenic mice and human Alzheimer's disease brain. *PLoS One* **8**, e59749 (2013).
91. A. Bryll et al., Oxidative-Antioxidant Imbalance and Impaired Glucose Metabolism in Schizophrenia. *Biomolecules* **10**, (2020).
92. S. Seshadri et al., Plasma homocysteine as a risk factor for dementia and Alzheimer's disease. *N Engl J Med* **346**, 476–483 (2002).
93. J. W. D. Griffin, P. C. Bradshaw, Amino Acid Catabolism in Alzheimer's Disease Brain: Friend or Foe? *Oxid Med Cell Longev 2017*, 5472792–5472792 (2017).
94. T. T. Braga et al., Soluble Uric Acid Activates the NLRP3 Inflammasome. *Scientific reports* **7**, 39884–39884 (2017).
95. F. Hansmannel et al., Is the urea cycle involved in Alzheimer's disease? *J Alzheimers Dis* **21**, 1013–1021 (2010).
96. V. Chouraki et al., Association of amine biomarkers with incident dementia and Alzheimer's disease in the Framingham Study. *Alzheimers Dement* **13**, 1327–1336 (2017).
97. S. F. Graham et al., Untargeted metabolomic analysis of human plasma indicates differentially affected polyamine and L-arginine metabolism in mild cognitive impairment subjects converting to Alzheimer's disease. *PLoS One* **10**, e0119452 (2015).
98. G. Wang et al., Plasma metabolite profiles of Alzheimer's disease and mild cognitive impairment. *J Proteome Res* **13**, 2649–2658 (2014).
99. H. Chew, V. A. Solomon, A. N. Fonteh, Involvement of Lipids in Alzheimer's Disease Pathology and Potential Therapies. *Frontiers in physiology* **11**, 598–598 (2020).
100. V. van der Velpen et al., Systemic and central nervous system metabolic alterations in Alzheimer's disease. *Alzheimer's Research & Therapy* **11**, 93 (2019).
101. D. Li et al., Prospective associations of plasma phospholipids and mild cognitive impairment/dementia among African Americans in the ARIC Neurocognitive Study. *Alzheimers Dement (Amst)* **6**, 1–10 (2016).
102. X. Han et al., Metabolomics in early Alzheimer's disease: identification of altered plasma sphingolipidome using shotgun lipidomics. *PLoS One* **6**, e21643 (2011).
103. D. Li et al., Plasma phospholipids and prevalence of mild cognitive impairment and/or dementia in the ARIC Neurocognitive Study (ARIC-NCS). *Alzheimers Dement (Amst)* **3**, 73–82 (2016).
104. D. Ciavardelli et al., Medium-chain plasma acylcarnitines, ketone levels, cognition, and gray matter volumes in healthy elderly, mildly cognitively impaired, or Alzheimer's disease subjects. *Neurobiol Aging* **43**, 1–12 (2016).
105. H. Oberacher et al., Targeted Metabolomic Analysis of Soluble Lysates from Platelets of Patients with Mild Cognitive Impairment and Alzheimer's Disease Compared to Healthy Controls: Is PC aeC40:4 a Promising Diagnostic Tool? *J Alzheimers Dis* **57**, 493–504 (2017).
106. B. N. Simpson et al., Blood metabolite markers of cognitive performance and brain function in aging. *J Cereb Blood Flow Metab* **36**, 1212–1223 (2016).
107. L. Whiley et al., Evidence of altered phosphatidylcholine metabolism in Alzheimer's disease. *Neurobiol Aging* **35**, 271–278 (2014).
108. G. Paglia et al., Unbiased Metabolomic Investigation of Alzheimer's Disease Brain Points to Dysregulation of Mitochondrial Aspartate Metabolism. *J Proteome Res* **15**, 608–618 (2016).
109. K. J. Castor et al., Urine dicarboxylic acids change in pre-symptomatic Alzheimer's disease and reflect loss of energy capacity and hippocampal volume. *PLoS One* **15**, e0231765 (2020).
110. N. Kurbatova et al., Urinary metabolic phenotyping for Alzheimer's disease. *Scientific reports* **10**, 21745–21745 (2020).
111. S. Passi et al., Saturated dicarboxylic acids as products of unsaturated fatty acid oxidation. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* **1168**, 190–198 (1993).

112. V. Birzniece et al., Neuroactive steroid effects on cognitive functions with a focus on the serotonin and GABA systems. *Brain Res Rev* **51**, 212–239 (2006).
113. Y. Wang, et al., Bile acids regulate cysteine catabolism and glutathione regeneration to modulate hepatic sensitivity to oxidative injury. *JCI Insight* **3**, e99676 (2018).
114. G. L. Bowman et al., Ascorbic acid and rates of cognitive decline in Alzheimer's disease. *J Alzheimers Dis* **16**, 93–98 (2009).
115. J. Milbrandt et al., Persephin, a Novel Neurotrophic Factor Related to GDNF and Neurturin. *Neuron* **20**, 245–253 (1998).
116. K. B. Zihlmann et al., The GDNF family members neurturin, artemin and persephin promote the morphological differentiation of cultured ventral mesencephalic dopaminergic neurons. *Brain Res Bull* **68**, 42–53 (2005).
117. C. Peña-Bautista et al., New screening approach for Alzheimer's disease risk assessment from urine lipid peroxidation compounds. *Scientific Reports* **9**, 14244 (2019).
118. C. Proto et al., Plasma levels of neuropeptides in Alzheimer's disease. *Gynecol Endocrinol* **22**, 213–218 (2006).
119. J. Olazarán, et al., A blood-based, 7-metabolite signature for the early diagnosis of Alzheimer's disease. *J Alzheimers Dis* **45**, 1157–1173 (2015).
120. R. Savica et al., Plasma sphingolipid changes with autopsy-confirmed Lewy Body or Alzheimer's pathology. *Alzheimers Dement (Amst)* **3**, 43–50 (2016).
121. M. E. Ritchie et al., limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
122. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
123. K. L. Howe, et al., Ensembl 2021. Nucleic Acids Research **49**, D884–D891 (2020).
124. L. Väremo, J. Nielsen, I. Nookaew, Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Res* **41**, 4378–4391 (2013).
125. R. Agren et al., The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum. *PLoS Comput Biol* **9**, e1002980 (2013).
126. M. Kanehisa, Y. Sato, KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci* **29**, 28–35 (2020).
127. L. J. Pellegrino, A. J. Cushman, A. S. Pellegrino, A stereotaxic atlas of the rat brain. (New York *(N.Y.) :* Plenum press, *ed. 2nd ed.* 1967, *1979)*.

Paper IV: Machine learning analysis reveals biomarkers for the detection of neurological diseases

The provided article below is adapted from the author's submitted version of the manuscript which has been accepted at Frontiers in Molecular Neuroscience.

Machine learning analysis reveals biomarkers for the detection of neurological diseases

Simon Lam[1], Muhammad Arif[2], Xiya Song[2], Mathias Uhlen[2], Adil Mardinoglu[1,2,*]

[1]Centre for Host-Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, London, SE1 9RT, United Kingdom

[2]Science for Life Laboratory, KTH – Royal Institute of Technology, Stockholm, 17165, Sweden

*Corresponding author: adilm@scilifelab.se

Abstract

It is critical to identify biomarkers for neurological diseases (NLDs) to accelerate drug discovery for effective treatment of patients of diseases that currently lack such treatments. In this work, we retrieved genotyping and clinical data from 1223 UK Biobank participants to identify genetic and clinical biomarkers for NLDs, including Alzheimer's disease (AD), Parkinson's disease (PD), motor neuron disease (MND), and myasthenia gravis (MG). Using a machine learning modelling approach with Monte Carlo randomisation, we identified a panel of informative diagnostic biomarkers for predicting AD, PD, MND, and MG, including classical liver disease markers such as alanine aminotransferase, alkaline phosphatase, and bilirubin. A multinomial model trained on accessible clinical markers could correctly predict an NLD diagnosis with an accuracy of 88.3%. We also explored genetic biomarkers. In a genome-wide association study of AD, PD, MND, and MG patients, we identified single nucleotide polymorphisms (SNPs) implicated in several craniofacial disorders such as apnoea and branchiootic syndrome. We found evidence for shared genetic risk loci among NLDs, including SNPs in cancer-related genes and SNPs known to be associated with non-brain cancers such as Wilms tumour, leukaemia, and colon cancer. This indicates overlapping genetic characterisations among NLDs which challenges current clinical definitions of the neurological disorders. Taken together, this work demonstrates the value of data-driven approaches to identify novel biomarkers in the absence of any known or promising biomarkers.

Significance statement

This study highlights the potential for data-driven, hypothesis-free mathematical modelling of easily measured clinical variables to identify diagnostic biomarkers for neurological diseases (NLDs). Prior to this study, the focus in NLD research has focused on toxic species such as amyloid-β and α-synuclein, but this approach has not enjoyed success at clinical trial. Here, we studied Alzheimer's disease, Parkinson's disease, motor neuron disease, and myasthenia gravis by constructing and inspecting a multinomial based on demographic data, cognitive scores, and blood and urine biochemistry. We found that cognitive measures and enzymes normally implicated in liver health were important for the predictive power of the model. By inspecting the model, we found that model weights correctly indicated multiple trends reported in the literature. We therefore posit that such model deconstruction could be useful to guide future NLD investigation. Separately, genome-wide association indicated a shared risk profile among NLDs, and between NLDs, cancer, and craniofacial disorders. This challenges the current clinical definitions of NLDs and suggests that the diseases are in fact overlapping syndromes.

Introduction

Neurological diseases (NLDs) pose a significant public health problem due to ageing populations and a widening global burden (Feigin et al., 2021; Nichols et al., 2019; Ray Dorsey et al., 2018). Currently, Alzheimer's disease (AD) is characterised by the accumulation of amyloid-β plaques and tau protein neurofibrillary tangles. These toxic species result in the destruction of cholinergic neurons and cause cognitive decline, leading to dementia. Parkinson's disease (PD) is characterised by α-synuclein inclusions in dopaminergic neurons in the substantia nigra, resulting in dopamine depletion and movement disorders. Motor neuron disease (MND) and myasthenia gravis (MG) are movement disorders that affect motor neurons and muscle, respectively, with MG also being an autoimmune disease. Individuals with PD or MND may also develop dementia.

Despite decades of research and clinical trials, there are currently no treatments to reverse the damage done by NLDs. Current research has focused on toxic species such as amyloid-β and tau in AD and α-synuclein in PD. Still, treatments derived from the research have been largely unsuccessful at clinical trials (Lam et al., 2020). Additional pharmacological and nonpharmacological targets need to be identified to enable early diagnosis and accelerate development of drugs and interventions to treat patients with NLDs effectively, and there is already great interest in this effort, revealing targets such as retinoid and androgen metabolism, YKL-40, AMIGO1, and GPRASP2; and suggesting that diet and lifestyle may also form part of an intervention (Baldacci et al., 2020; Bayraktar et al., 2021; Lam et al., 2021; Toschi et al., 2019; Zool et al., 2020).

Biomarkers such as those mentioned above are objectively measurable biological characteristics that give information on disease. Biomarkers can be used to measure susceptibility to disease, predict or evaluate the response to an intervention or environmental exposure, assess severity or other characteristics of a medical condition, or diagnose disease or stratify individuals into disease classes or subtypes. They can indicate correlation between diagnostic tests and disease, and they can provide mechanistic insight into disease progression. In addition to the targets mentioned above, there is also a need for diagnostic biomarkers which might be

present in easily collected samples such as blood and urine. Although such analytes might be less likely to provide rigorous mechanistic insight into disease progression, the need for easy and earlier detection prevails, enabling initial screening for suspected disease before onset and without requiring extensive laboratory equipment.

In the present study, we retrieved clinical and genotyping data from 1223 participants in the UK Biobank with NLDs (Sudlow et al., 2015), identified diagnostic markers for the predictions of NLDs, and identified shared genetic predispositions to NLDs (Figure 1). To this end, we used the clinical data to construct a predictive multinomial general linear model for AD, PD, MND, and MG and perform genome-wide association studies (GWAS). These diseases were chosen in order to represent a range of brain-related pathologies and a range of better-studied and less well-studied diseases.

The multinomial model was constructed using Monte Carlo randomisation to select the optimal clinical variables to be included in the model and was able to predict NLD diagnosis with 88.3% accuracy. In addition to the promising true positive rate, inspection of the model weights correctly identified directional trends in blood and urine biochemistry and cognitive function that confirm previous reports in the literature. These included biomarkers which are normally associated with liver disease, such as alanine aminotransferase (ALT), alkaline phosphatase (ALP), and bilirubin, confirming the possibility of a role for the liver-brain axis in NLDs such as AD, as previously reported (Bassendine et al., 2020; Jakhmola-Mani et al., 2021).

We also used genotype data to identify genetic predispositions to NLDs. In GWAS, we found hundreds of risk alleles that were shared across at least two NLDs. There was enrichment for single nucleotide polymorphisms (SNPs) associated with non-brain cancers, such as Wilms tumour, leukaemia, and colon cancer; and craniofacial syndromes, such as apnoea and branchiootic syndrome. We identified a panel of nine SNPs which were present in all four NLDs studied. These included two homeobox genes and three genes encoding biosynthetic enzymes. The identification of these shared SNPs suggests that the same SNP may be indicative of susceptibility

to more than NLD and could potentially suggest that there exists a basal genetic perturbation that is linked with developing NLDs generally.

Taken together, our work demonstrates the usefulness of data-driven approaches to quickly identify easily measurable diagnostic biomarkers and shared genetic risk loci. Our results suggest that cognitive measures and biochemical markers usually associated with liver diseases are informative for the prediction of NLDs, and that SNPs linked with non-brain cancers and craniofacial disorders may be indicative of multiple NLDs. We therefore recommend this data-driven approach to guide future investigation for the study of diseases of unknown aetiology or where promising biomarkers are yet unknown.

Results

Blood and urine biochemistry and cognitive trends inform NLD status

We retrieved clinical data from 1223 patients with AD, PD, MND, or MG, of which 1072 also had Affymetrix Axiom genotyping data on 820967 SNPs, from the UK Biobank (Supplementary Table 1). The mean age of participants with AD, PD, MND, and MG was 62.8, 62.1, 59.4, and 60.2, respectively, compared with 53.7 for control participants. The male proportions were 56.6%, 63.3%, 69.2%, and 46.6%, respectively, compared to 45.3% for controls. The cohort was made up of 89.5%, 91.3%, 90.8%, and 91.4% British participants, respectively, compared with 87.1% British for controls.

Patients with NLDs exhibited changes to blood and urine biochemical markers (Figure 2, Supplementary Table 2). Participants with AD had significantly higher ALT, ALP, and cystatin C levels and had the lowest performance in the cognitive tests. PD participants had lower ALT, albumin, apolipoprotein A (ApoA), calcium, cholesterol, low-density lipoprotein (LDL), and phosphate, but elevated ALP, bilirubin, glucose, microalbumin, sodium, and testosterone, compared to control. Participants with MND had significantly higher ALT, ALP, cystatin C, microalbumin, and testosterone compared to control, but lower ApoA. MG patients had decreased albumin, cholesterol, and LDL, but increased cystatin C compared to control. We noticed that a number of these clinical marker changes were involved in

markers usually associated with liver disease. Indeed, this result is consistent with reports associating liver-based correlates with cognitive function (Chen et al., 2021; Kellett et al., 2011) and movement disorders (Jeong et al., 2021; Lee & Yang, 2018; Zhang et al., 2021).

All NLD patients took longer to correctly identify matches in cognitive testing than control. In the prospective memory test, AD patients were the most likely not to recall the instruction after two attempts and showed the least improvement between subsequent visits. This result was to be expected since AD is the disease most strongly associated with dementia and cognitive decline out of the four NLDs under investigation.

A multinomial model classified neurological diseases and identified relevant biomarkers

From the empirical results suggesting clinical marker signatures in NLDs, we next sought to generate a predictive model of NLD diagnosis based on those markers. We therefore generated a multinomial generalised linear model to predict AD, PD, MND, MG, and control in a single test. We selected clinical measures with good coverage among participants to create this model and used Monte Carlo randomisation to optimise the clinical measures to be included in the model (see Methods). The confounders in our final model can be summarised as follows: demographics (age, ethnicity), blood test measures (ALT, albumin, apolipoprotein A, calcium, cholesterol, cystatin C, LDL, phosphate, testosterone), urine test measures (microalbumin, sodium), and cognitive test measures (prospective memory test, reaction test). The multinomial model had a true positive rate of 88.3% on unseen data (Supplementary Table 3A).

We performed leave-one-out cross validation to demonstrate the essentiality of each of the variables included in the model in terms of its contribution to predictive power (Supplementary Table 4A). We constructed and tested new multinomial models after singly omitting each clinical marker, separately. We found that the model was robust to single variable omission, with most models performing at around 87–88% accuracy. The biggest drops arose from omission of prospective memory result (first visit) (76.0% accuracy), age (85.8% accuracy), and testosterone

(87.0%). Omission of LDL, cystatin C, mean time to correctly identify matches, sodium, and apolipoprotein A all produced models with 87.7% accuracy. Age and cognition appeared to be the most important factors contributing to model accuracy.

We inspected the coefficients assigned to each of the clinical variables included in the model (Supplementary Table 5A). AD was more likely to be predicted for patients with low ALT, apolipoprotein A, cystatin C, LDL, and urine microalbumin. High calcium, cholesterol, and urine sodium were also characteristics of AD identified by the model. AD was the disease most tightly linked to advanced age, and patients of British, Chinese, or Irish background, were more likely to be predicted to have AD by the model, as did those who took the longest to identify matches. PD was assigned to those patients with advanced age, low ALT, low apolipoprotein A, and high testosterone. People of Chinese background or any other Asian background were less likely to be predicted to have PD. The model assigned MND to those patients with high ALT calcium, cholesterol, and testosterone, but low apolipoprotein A and phosphate. Patients who answered the ethnicity question with "any other white background" were more likely to be assigned MND by the model. The model linked MG with high ALT and cystatin C, but low albumin and LDL. Patients of British background were more likely to be assigned MG, whereas those of Irish and those who answered "any other white background" were less likely.

Across all diseases, the model applied a negative coefficient for the prospective memory test at the first visit, regardless of the result. Patients who took more attempts to recall correctly, or did not recall at all, were more likely to be assigned AD by the model than any other disease. Interestingly, the coefficients for PD were very similar regardless of the result. Together with the small coefficient given to the match identification test, the model suggests that cognitive measures were not very important for the classification of PD. In contrast, patients not recalling the instruction were given very negative coefficients for MND and MG, indicating that these patients are more likely to have a diagnosis of AD or PD instead.

Interestingly, at the second visit, the coefficients were always lower if the patient took two attempts rather than one, which appeared counter-intuitive for AD.

However, AD also had the least negative coefficient, indicating that patients who took two attempts to recall the instruction at the second visit were more likely to be assigned AD.

The model also identified numerous interactions with blood and urine biochemical levels, indicating elevated ALT in MND and MG, but decreased in AD and PD; sharply decreased apolipoprotein A in AD and MND, but more modest decreases in PD and MG; increases in calcium and cholesterol in AD, MND, and MG, but changes in these molecules not being significant in PD; decreases in LDL and phosphate in any disease, but most strikingly in MG and MND, respectively; and increases in sodium and testosterone in all diseases. That the model was able to suggest these directional changes, which could not be so clearly concluded by t-tests alone (Figure 2), indicates the potential for machine learning techniques to identify possible biomarkers for classification in the absence of an understanding by the model of the underlying biology.

To determine whether biomarkers alone could predict NLD diagnoses, we repeated the Monte Carlo multinomial model but excluded all demographics data (that is, sex, age, and ethnic background) from the analysis. We found that this biomarker-only model was able to predict NLDs with an accuracy of 85.3%, which is comparable to the full model (Supplementary Table 3B), and indicates that much of the predictive power of the model could be explained by biomarkers alone. This is also consistent with the leave-one-out cross validation results on the full model (Supplementary Table 4A). The nine variables included in the model were alkaline phosphatase, glucose, mean time to correctly identify matches, urine microalbumin, phosphate, prospective memory result (first and third visits), testosterone, and total bilirubin. Similar to the full model, the biomarker-only model predicted NLDs on the basis of biomarkers normally associated with liver health (alkaline phosphatase, glucose, and total bilirubin).

In addition, inspecting the model weights in the biomarker-only model revealed retained predicted effects of clinical measures on NLD prediction (Supplementary Table 5B). The biomarker-only model introduced four clinical measures that were not selected in the full model. Alkaline phosphatase was predictive for AD when

elevated but for MG when depleted, and glucose was predictive for MG when elevated but AD when depleted. Prospective memory test (third visit) was predictive for NLDs in a similar manner to second visit results in the full model. Total bilirubin was also predictive for AD when depleted.

In leave-one-out cross validation of the biomarker-only model, model accuracy was not impacted by single removal of alkaline phosphatase, glucose, urine microalbumin, or phosphate as predictors (Supplementary Table 4B). All other predictors only modestly reduced the model accuracy from 85.3% to 85.1%, with the exception of prospective memory result (first visit), whose removal resulted in a model accuracy of only 66.9%. This is in line with the full model and suggests that simple cognitive tests are highly informative when diagnosing NLDs.

Genome-wide association uncovered shared heritable factors between neurological diseases

In the multinomial model, we identified a heritable variable, ethnicity, to predict neurodegeneration. Therefore, we hypothesised that there might be more heritable factors that could alter a person's susceptibility to NLDs. To identify heritable factors, we analysed genotyping data originating from an Affymetrix Axiom array, which covers 836,727 SNPs (Figures 3 & 4, Supplementary Figure 1, Supplementary Table 6). We found significant SNPs which were also linked with non-brain cancers and craniofacial disorders. Many of these SNPs were found to be associated with more than one NLD. For example, AD, PD, and MND were all associated with SNPs in LAMA2, PTPN12, and SPATA7, which are implicated in muscular dystrophy, colon cancer, and retinitis pigmentosa, respectively. AD, PD, and MG all had SNPs in PTCH1 (associated with holoprosencephaly) and XKR9 (associated with otofaciocervical syndrome). AD, MND, and MG were all associated with SNPs in VPS41 (associated with Wilms tumour) and CHRM3 (associated with Eagle-Barrett syndrome). In total, we found 70 SNPs linked with at least three of the four NLDs under investigation.

Considering the functions of genes affected by the SNPs we identified, we found a vast number of significant SNPs within genes associated with cancer, its hallmarks, or neurotransmission (Supplementary Data 1). An intronic SNP in the scaffold

protein interactor GAB2 was associated with all four diseases. A missense mutation in the TNFα response gene TNFAIP3 gene was highly associated with AD and PD, but not MND or MG. AD was also associated with SNPs in the cytochrome P450 gene CYP2B6, adenylate cyclase ADCY8, and signalling gene PIK3C3 (PI3K). PD was highly associated with SNPs in dysferlin (DYSF), TP53, and numerous cytoskeleton genes such as AFAP1L1 and MYH1. MND was associated with SNPs in retinoblastoma protein interactor RBBP5, tumour necrosis factor family member TNFRSF25, and several cytoskeleton function genes, including tropomyosin TPM1 and troponin TNNI3. MND was also associated with SNPs in spermatogenesis associated genes such as SPATA8 and SPAG16. MG was associated with SNPs in hallmarks of cancer–related genes such as the microtubule-associated tumour suppressor MTUS1, leukocyte-associated gene LAIR2, and Cbl proto-oncogene CBLB. MG was also associated with SNPs in the brain- and neuron-specific genes, including cerebellin 4 precursor (CBLN4), potassium channel KCNH5, adenylate cyclase 8 (ADCY8), and autism susceptibility candidate gene AUTS2.

In addition to GAB2, we also identified a further eight SNPs which were associated with all four diseases (Figure 4, Supplementary Table 6). LINC00290, ACO1, HLA-G, SIX1, HS6ST1, GALNT10, ONECUT1, and HLA-DRB6 were all present in the top 1000 SNP associations by asymptotic p-value in all four diseases. Of these, HLA-G, SIX1, HS6ST1, ONECUT1, and HLA-DRB6 have existing annotations in the Online Mendelian Inheritance in Man (OMIM) database related to brain and craniofacial syndromes. ACO1, HS6ST1, and GALNT10 encode biosynthetic enzymes, strengthening our result in the multinomial models suggesting a role for biochemistry-based biomarkers. SIX1 and ONECUT1 are homeobox genes. SIX1 is a master regulator in multiple tissues and ONECUT1 is a transcription factor of the liver; and are associated with the craniofacial disorders branchiootic syndrome and amelogenesis imperfecta, respectively. This further strengthens the proposed link between NLDs and other tissues, and suggests that interactions with other tissues may also influence craniofacial disorders, but further investigation is required to confirm this.

These results indicate that the same SNPs may be associated with susceptibility to more than one NLD. Further, the identification of NLD risk SNPs linked with

cancer risk is significant as it is consistent with the reported degeneration/cancer antagonistic shift (Aramillo Irizar et al., 2018). The additional identification of NLD risk SNPs linked with craniofacial disorders suggests a potentially interesting axis of investigation for NLDs which is not well explored at present (Kamer et al., 2020). Taken together with the results from multinomial modelling, we propose a paradigm shift in NLD biomarker identification to focus on liver biochemistry, interactions with non-brain cancer, and association with craniofacial disorders.

Discussion

This study shows the value of big biological data in driving hypothesis-free studies towards the early diagnosis and prediction of genetic risk loci in NLDs. Inspired by empirical trends in clinical marker data (Figure 2, Supplementary Table 2), we first constructed a multinomial model that could predict AD, PD, MND, or MG with an accuracy of 88.3% (Supplementary Table 3). The clinical variables that we used to generate the model were selected by a Monte Carlo randomisation method, and the model weightings could be used to direct future NLD research.

Regarding the multinomial model, deconstruction of the model weightings revealed directional machine-learned associations between clinical markers and diseases (Supplementary Table 5). Some of these associations have already been reported in the literature: for instance, our model predicted low ALT in AD, which is consistent with the literature (Lu et al., 2021). Our model also correctly predicted low apolipoprotein A in AD, PD, and MND (Mariosa et al., 2017; Qiang et al., 2013; Zuin et al., 2021) and elevated calcium in AD and PD (Angelova et al., 2016; Ryan et al., 2020; Zakharov et al., 2007). Such model deconstruction after machine learning therefore appears to be a useful tool for identifying biomarkers and guiding further investigation in a manner that is agnostic to the assumptions of the underlying biology.

Leave-one-out cross validation (Supplementary Table 4A) on the model and assessment of non-demographic clinical markers only (Supplementary Tables 3B, 4B, & 5B) revealed that cognitive scores were the single most important variable in the model in terms of predictive power. Single removal of any other variable from the model did not adversely impact the predictive power of the model, indicating

that no single variable (apart from cognitive scores) could reasonably predict NLDs. This, along with the Monte Carlo randomised selection of biomarkers usually attributed to liver health, represents a paradigm shift away from single species in NLD research such as amyloid-β in AD and α-synuclein in PD.

Having also included ethnicity in the full multinomial model, we next carried out a GWAS to identify other heritable factors which might predispose an individual to NLDs (Figures 3 & 4). While we could not account for the variability in the data due to ethnicity, we found commonalities across all four diseases – particularly the appearance of SNPs associated with non-brain cancers and craniofacial disorders. The identification of hundreds of SNPs shared across at least two NLDs, dozens of SNPs shared across at least three NLDs, and nine SNPs shared across all four NLDs strikingly suggests some common but yet unknown genetic mechanism associated with NLD in general. We found a marked enrichment of SNPs annotated in the OMIM database to be linked with non-brain cancers and craniofacial disorders, as well as three SNPs in biosynthetic genes (ACO1, HS6ST1, and GALNT10) and two SNPs in homeobox genes (SIX1 and ONECUT1) among the nine SNPs shared by all four NLDs. Our GWAS results therefore suggest that the same SNPs may confer susceptibility to more than one NLD, and may also represent a risk factor for non-brain cancer and craniofacial disorders. More research must be conducted to uncover the genetic mechanism that appears to be common to all of these conditions.

Biomarker discovery proceeds through phases of preclinical exploration, clinical validation, longitudinal and retrospective study, and prospective screening. As this study was based on preclinical data from UK Biobank, the biomarkers proposed in this study are considered at the early exploratory phase and thus no direct biological mechanism linking them to disease is proposed. This study can be expanded by controlled clinical assessment of patients with or without a diagnosis of NLD for alterations in blood and urine biochemistry prior to a diagnosis or during the course of disease progression. This will also enable feasibility assessments for those analytes as early detection biomarkers.

It is important to acknowledge the limitations of the study. Firstly, the multinomial model weightings contradicted the literature in some places. An example of this was

the significant negative coefficient for LDL in MND, suggesting that MND patients might have lower serum LDL, whereas this is reported not to be the case (Chen et al., 2018; Zeng & Zhou, 2019). Secondly, although the model had an accuracy of 88.3%, it predicted some diseases better than others. Both of these limitations likely were due to the relatively low numbers of patients with AD, MND, and MG compared to PD and control. Because of this, we indicate that the model may be improved with the inclusion of more samples and more consistent data collection across all patients. This limitation, however, did not detract from the model's usefulness to be deconstructed to predict biomarkers, and does not change the fact that any predictions from the model must be treated as a starting point for future validation. Thirdly, the model did not include any biomarkers from cerebrospinal fluid (CSF). CSF measures are potentially informative due to circulating factors which are not detectable in blood. However, the UK Biobank does not collect such data, so we could not analyse these factors in the current study. Regarding the GWAS, again, our investigation would have benefited with more even patient numbers across NLDs. The studies conducted on AD and PD samples were more powerful than those on MND and MG. We overcame this limitation by inspecting the top 1000 SNPs per NLD rather than imposing a p-value cutoff, which also controlled false positive discoveries in the larger AD and PD patient pools. Fourthly, we identified a number of liver enzymes as having potentially predictive roles in the model. We did not consider alcohol consumption as a possible confounder of liver enzyme levels although it is known that alcohol can modulate liver enzyme levels. The model did not preferentially select alcohol consumption over liver enzyme levels as a predictive model feature, but we cannot exclude the possibility that the variance explained by alcohol consumption was captured by the variance in liver enzyme levels. Further studies would be required to exclude this possibility.

In conclusion, we propose data-driven machine learning and data exploration by GWAS as ideal first steps towards biomarker discovery for diseases of unknown aetiology or currently lacking promising biomarkers. Such data-driven approaches may be extended to bench experimental work and are expected to guide dynamic detection and quantification of target druggability, *in vivo* demonstration of mechanisms of action, and prediction of drug resistance mechanisms. In this work,

we demonstrated the value of a data-driven approach by identifying accessible blood, urine, and cognitive biomarkers in AD, PD, MND, and MG using a machine learning multinomial model with no knowledge of the underlying biology. This approach highlighted liver enzymes as potentially diagnostic biomarkers for NLDs and should now be targets for basic biology research in NLDs as outlined above. We also used GWAS to confirm shared genetic risk loci between NLDs, cancer, and craniofacial disorders. Although the ageing-related antagonistic switch between degeneration and cancer has been reported, the association between NLD and craniofacial disorders is a yet-untapped arena that demands further investigation.

Methods

Acquisition of data and inclusion criteria

Data were obtained from UK Biobank (Sudlow et al., 2015). Samples were accepted as Alzheimer's disease (AD), myasthenia gravis (MG), motor neuron disease (MND), or Parkinson's disease (PD) samples if any of the respective diseases were recorded by UK Biobank, even if other conditions were also recorded. UK Biobank records conditions and diseases on a self-reporting basis. Participants are asked to self-report by answering the question in a UK Biobank questionnaire: "You selected that you have been told by a doctor that you have other (non-cancer) serious illnesses or disabilities, could you now tell me what they are?" Control samples were accepted as those without any recorded diseases (Supplementary Table 1).

Generation of the multinomial model and leave-one-out cross validation

To generate the multinomial model, standard and easily-measurable clinical data including demographics, sight and hearing problems, diabetes diagnosis, stroke diagnosis, medication and treatment, illness, operations, cognitive and mental measures, brain measurements, blood and urine tests, and adverse events and death were obtained from UK Biobank. Clinical variables were removed from the analysis if less than 75% of samples had a value recorded. Categorical measures with only one category were also removed. After quality control, 40 clinical variables

remained in the analysis. A separate model was also generated excluding demographic data.

Monte Carlo randomisation was used to randomly sample clinical variables as independent variables in a multinomial model. Samples with any missing values across the selected clinical variables were dropped, yielding a final dataset consisting 37 AD, 328 PD, 17 MND, and 16 MG patients, and 823 healthy controls. The multinomial model was constructed using the R nnet package (version 7.3-14, https://cran.r-project.org/web/packages/nnet/index.html, accessed 2021-06-01). The dependent variable was disease. The dataset was randomly split into training (70%) and test sets (30%) to assess the accuracy (true positive rate) of the model on test set data. The model with the highest accuracy after 1000 random samplings was accepted for further consideration.

To find the model with the local maximum accuracy, clinical variables not in the model were added, and clinical variables already in the model were removed, individually. If, after each model change, the model's accuracy improved, then that change was kept; if otherwise, then that change was reverted. The final dataset contained 16 features (Supplementary Table 4A).

To perform leave-one-out cross validation on the multinomial model, variables were singly removed from the model, separately, and multinomial models were constructed and tested, as above.

Genotyping analysis

Raw genome-wide genotyping data were obtained from the UK Biobank Axiom Array. Axiom Analysis Suite (version 5.1.1, ThermoFisher, accessed 2021-06-01) was used to perform quality control and analysis on the raw data to determine patient genotypes at each single nucleotide polymorphism (SNP), using the Best Practices Workflow. Allele frequencies were computed from genotype data for each disease class. To detect significant differences in allele frequencies between disease and control, Cochran-Armitage trend tests were performed assuming a codominant allele model. The test statistics, exact p-values, and asymptotic p-values were recorded. CATTexact (version 0.1.1, https://cran.r-project.org/web/packages/CATTexact/index.html, accessed 2021-06-01) R package

was used for the computation of Cochran-Armitage p-values (Mehta et al., 1992). qqman (version 0.1.8, https://cran.r-project.org/web/packages/qqman/index.html, accessed 2021-06-01) R package was used to plot the Manhattan and quantile-quantile plots. The top 1000 SNPs by asymptotic p-value for each NLD were accepted for further visualisation in a Venn diagram. ggvenn (version 0.1.9, https://cran.r-project.org/web/packages/ggvenn/index.html, accessed 2022-03-31) R package was used to plot the Venn diagram. All bioinformatic and statistical analyses were performed in R (version 4.0.2) unless otherwise indicated.

Author contributions

M.U. and A.M. supervised the study. S.L. designed the study, performed all the analyses, and analysed all the results. S.L. wrote the manuscript with input from M.A., X.S., M.U., and A.M.

Acknowledgements

This research has been conducted using the UK Biobank Resource under application number 64488.

The authors acknowledge use of the research computing facility at King's College London, *Rosalind* (https://rosalind.kcl.ac.uk, accessed 2021-06-01).

This work was financially supported by Knut and Alice Wallenberg Foundation (grant number 2017.0303) to A.M.

Conflicts of interest

The authors declare no competing interests.

Figures

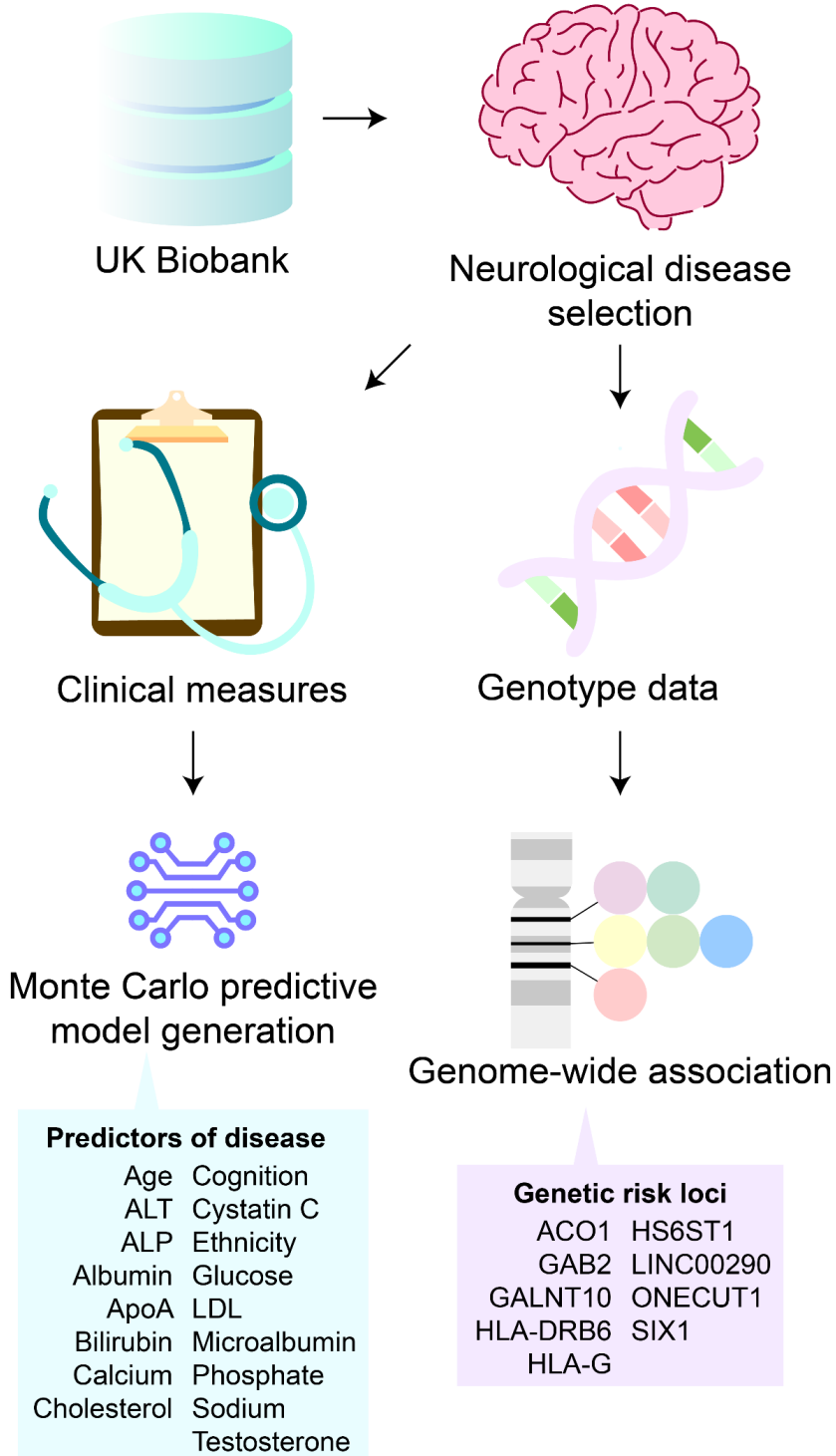

UK Biobank
Neurological disease selection
Clinical measures
Genotype data
Monte Carlo predictive model generation
Genome-wide association
Predictors of disease
Age
ALT
ALP
Albumin
ApoA
Bilirubin
Calcium
Cholesterol
Cognition
Cystatin C
Ethnicity
Glucose
LDL
Microalbumin
Phosphate
Sodium
Testosterone
Genetic risk loci
ACO1
GAB2
GALNT10
HLA-DRB6
HLA-G
HS6ST1
LINC00290
ONECUT1
SIX1

Figure 1. Study overview. From UK Biobank participants, Alzheimer's disease, Parkinson's disease, motor neuron disease, and myasthenia gravis clinical and genotype data were obtained. The clinical data were used to train a predictive model of neurological diseases. The model was deconstructed to reveal predictors of disease relating to demographics, cognitive scores, and liver health biomarkers. The genotype data were used to perform genome-wide association studies. This analysis revealed nine genetic risk loci that were shared among all four neurological diseases.

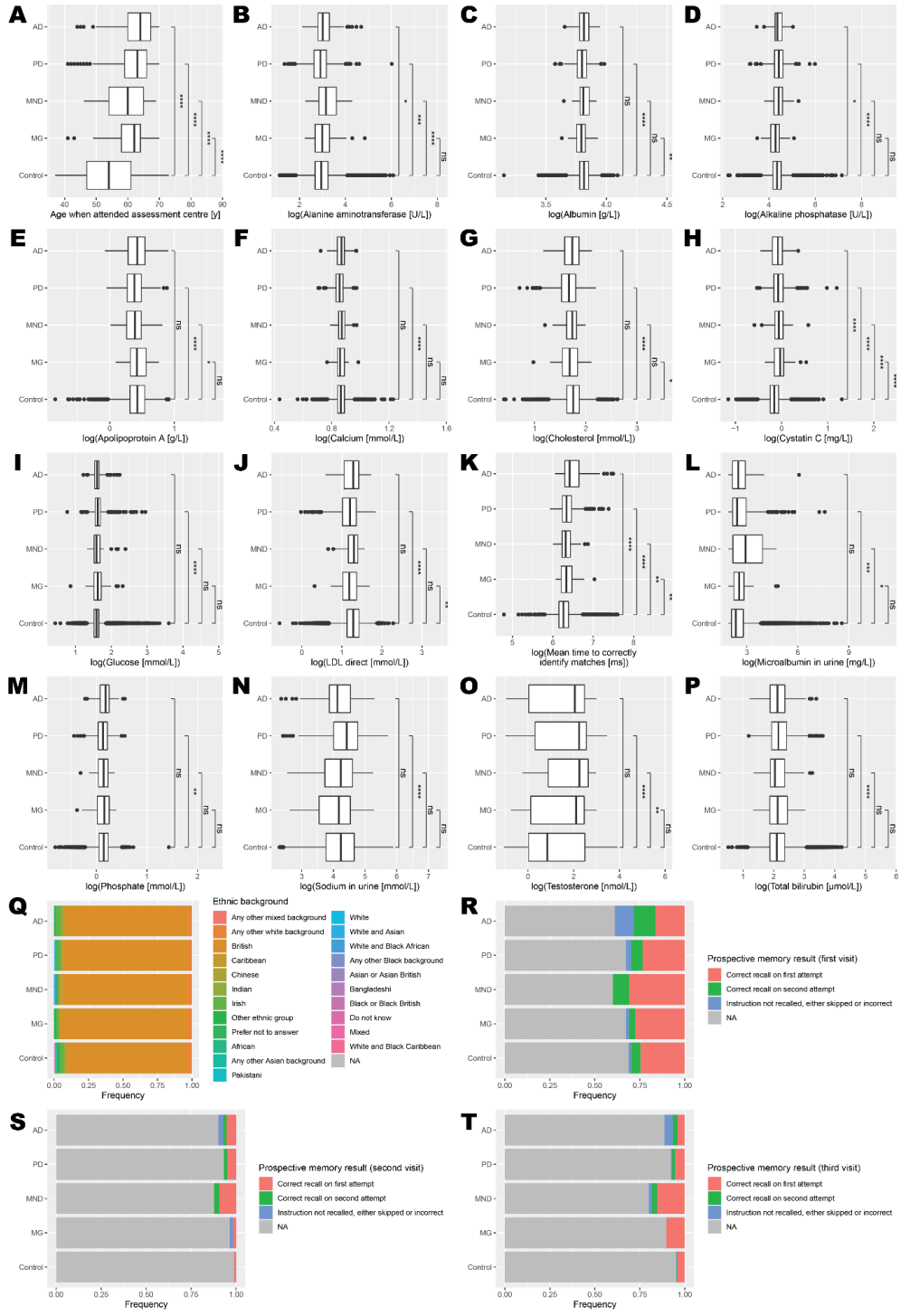


Figure 2. Distribution of clinical measures by diagnosis. A) Age when attended assessment centre. B) Alanine aminotransferase. C) Albumin. D) Alkaline phosphatase. E) Apolipoprotein A. F) Calcium. G) Cholesterol. H) Cystatin C. I)

Glucose. J) LDL direct. K) Mean time to correctly identify matches. L) Microalbumin in urine. M) Phosphate. N) Sodium in urine. O) Testosterone. P) Total bilirubin. Q) Ethnic background. R) Prospective memory result (first visit). S) Prospective memory result (second visit). T) Prospective memory result (third visit). Data represent all participants, including those whose samples which were not used in the training or testing of the multinomial model. AD, Alzheimer's disease (n = 152); PD, Parkinson's disease (n = 948); MND, motor neuron disease (n = 65); MG, myasthenia gravis (n = 58); Control (n = 116559). Mean comparisons were performed using t-tests after log transformation (except for age). *, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$; ****, $p \leq 0.0001$; ns, $p > 0.05$.

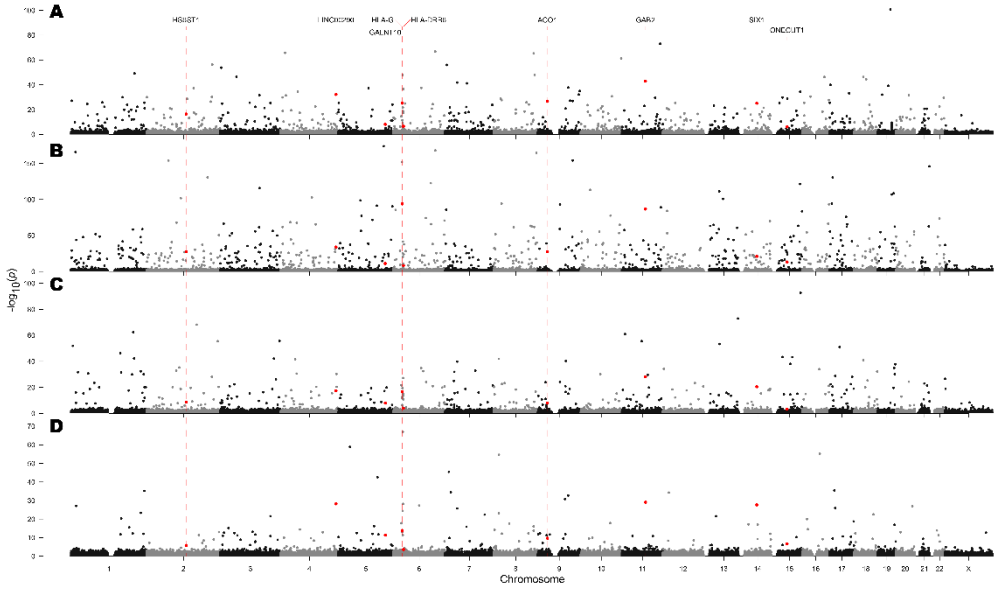


Figure 3. Genome-wide association with A) Alzheimer's disease, B) Parkinson's disease, C) motor neuron disease, and D) myasthenia gravis in the Affymetrix Axiom Biobank Array. Dashed red lines and red points indicate SNPs which appeared in the top 1000 associations by asymptotic Cochran-Armitage trend test p-value in all four neurological diseases.

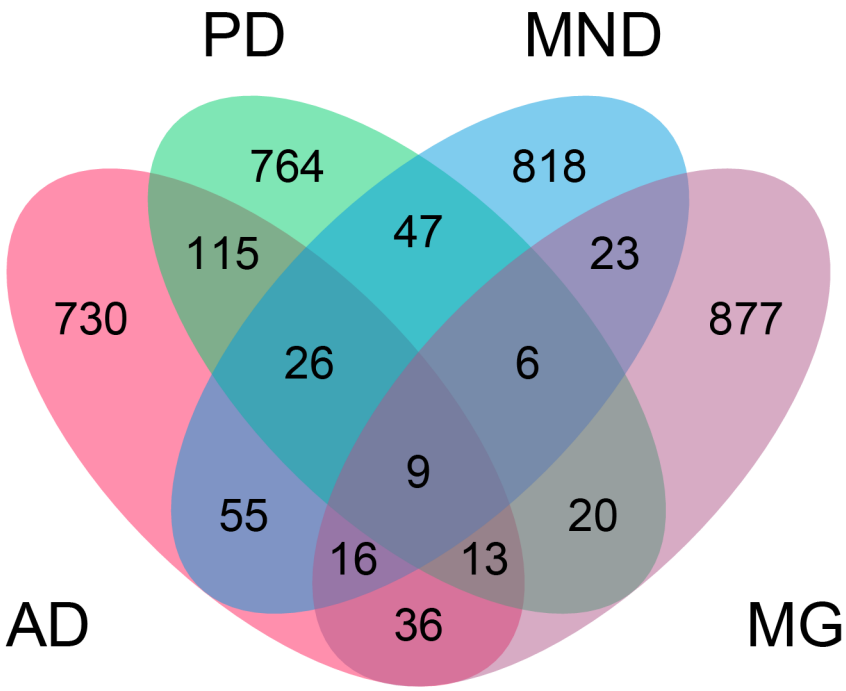
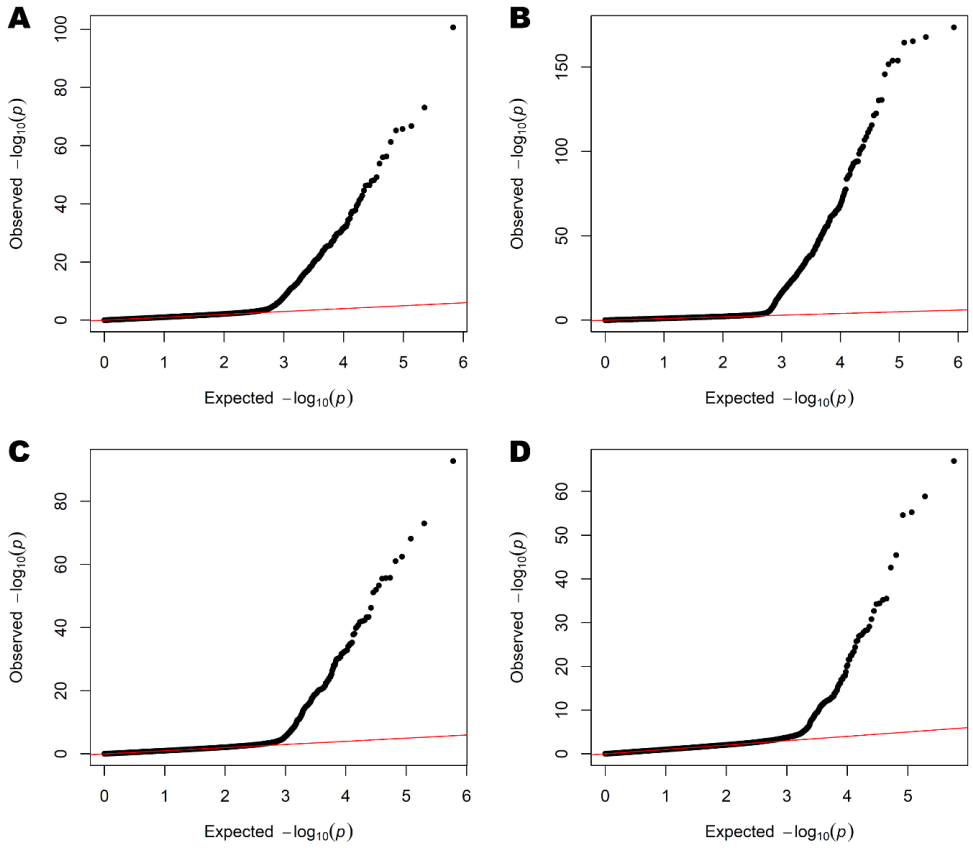


Figure 4. Distribution of shared SNPs across neurological diseases. The top 1000 SNPs identified by genome-wide association study with Alzheimer's disease (AD), Parkinson's disease (PD), motor neuron disease (MND), and myasthenia gravis (MG) were inspected for overlap across multiple diseases. Numbers indicate number of SNPs shared between diseases.

SI figures and tables



Supplementary Figure 1. Normal quantile-quantile plots of A) Alzheimer's disease, B) Parkinson's disease, C) motor neuron disease, and D) myasthenia gravis SNPs in the Affymetrix Axiom Biobank Array. All p-values are asymptotic Cochran-Armitage trend test p-values.

Supplementary Table 1. Characteristics of UK Biobank NDD dataset

Disease	Total	With genotyping data
AD	152	129
PD	948	832
MND	65	60
MG	58	51
Control	116559	446
Total (excluding controls)	1223	1072

Supplementary Table 2. Summary statistics of UK Biobank NDD dataset

Clinical marker	Disease	Mean	95% CI		p-value (t-test)*
Age when attended assessment centre [y]	AD	62.77	61.86	63.68	2.17×10^{-43}
	PD	62.13	61.79	62.48	1.80×10^{-255}
	MND	59.40	57.78	61.02	1.92×10^{-9}
	MG	60.17	58.50	61.84	2.12×10^{-10}
	Control	53.74	53.69	53.78	
Alanine aminotransferase [U/L]	AD	24.06	21.81	26.31	0.030
	PD	20.81	19.75	21.87	1.20×10^{-5}
	MND	28.21	24.80	31.62	1.82×10^{-4}
	MG	24.59	19.68	29.50	0.402
	Control	22.07	21.99	22.14	
Albumin [g/L]	AD	45.30	44.87	45.74	0.571
	PD	44.64	44.47	44.81	8.21×10^{-17}
	MND	45.40	44.81	45.99	0.951
	MG	44.41	43.74	45.07	5.61×10^{-3}
	Control	45.43	45.42	45.45	
Alkaline phosphatase [U/L]	AD	4.40	4.36	4.44	0.014
	PD	4.42	4.40	4.44	1.07×10^{-13}
	MND	4.42	4.35	4.49	0.044
	MG	4.28	4.20	4.37	0.146
	Control	4.35	4.34	4.35	
Apolipoprotein A [g/L]	AD	1.55	1.50	1.59	0.531
	PD	1.48	1.47	1.50	4.53×10^{-16}
	MND	1.47	1.41	1.53	0.0144
	MG	1.58	1.51	1.65	0.590
	Control	1.56	1.56	1.56	
Calcium [mmol/L]	AD	2.38	2.36	2.39	0.977
	PD	2.35	2.35	2.36	1.21×10^{-13}
	MND	2.39	2.37	2.42	0.225
	MG	2.36	2.34	2.39	0.255
	Control	2.38	2.38	2.38	
Cholesterol [mmol/L]	AD	5.70	5.52	5.88	0.0814
	PD	5.37	5.30	5.44	2.92×10^{-33}
	MND	5.64	5.42	5.85	0.115
	MG	5.48	5.17	5.80	0.0203
	Control	5.85	5.84	5.86	
Cystatin C [mg/L]	AD	0.94	0.91	0.96	1.03×10^{-7}
	PD	0.96	0.94	0.97	8.85×10^{-52}
	MND	0.96	0.91	1.00	3.22×10^{-4}
	MG	0.99	0.94	1.04	2.92×10^{-6}

	Control	0.87	0.87	0.87	
Glucose [mmol/L]	AD	1.61	1.59	1.63	0.056
	PD	1.64	1.63	1.65	3.76×10^{-17}
	MND	1.62	1.58	1.67	0.165
	MG	1.64	1.58	1.70	0.112
	Control	1.59	1.59	1.59	
LDL direct [mmol/L]	AD	3.57	3.43	3.71	0.125
	PD	3.36	3.30	3.41	1.50×10^{-25}
	MND	3.57	3.41	3.72	0.399
	MG	3.34	3.12	3.57	6.54×10^{-3}
	Control	3.67	3.66	3.67	
Mean time to correctly identify matches [ms]	AD	693.30	654.25	732.35	5.15×10^{-14}
	PD	591.66	583.36	599.96	2.99×10^{-34}
	MND	579.71	551.68	607.74	6.22×10^{-3}
	MG	588.43	555.41	621.45	3.54×10^{-3}
	Control	544.18	543.55	544.81	
Microalbumin in urine [mg/L]	AD	24.03	15.01	33.05	0.212
	PD	28.70	21.01	36.39	4.04×10^{-3}
	MND	31.02	23.76	38.28	0.0314
	MG	23.48	15.09	31.87	0.261
	Control	20.92	20.47	21.38	
Phosphate [mmol/L]	AD	1.18	1.16	1.21	0.106
	PD	1.14	1.13	1.15	7.17×10^{-3}
	MND	1.14	1.11	1.18	0.566
	MG	1.15	1.10	1.20	0.619
	Control	1.16	1.16	1.16	
Sodium in urine [mmol/L]	AD	74.39	67.77	81.02	0.447
	PD	89.43	86.57	92.29	6.70×10^{-19}
	MND	74.53	64.05	85.01	0.562
	MG	75.38	63.24	87.51	0.466
	Control	78.48	78.22	78.74	
Testosterone [nmol/L]	AD	7.19	6.29	8.10	0.114
	PD	8.47	8.08	8.85	1.86×10^{-20}
	MND	8.83	7.38	10.28	1.53×10^{-4}
	MG	7.63	6.05	9.21	0.174
	Control	6.66	6.62	6.70	
Total bilirubin [µmol/L]	AD	2.14	2.08	2.21	0.990
	PD	2.20	2.18	2.23	2.14×10^{-5}
	MND	2.13	2.03	2.24	0.852
	MG	2.15	2.04	2.25	0.955
	Control	2.14	2.14	2.15	

*Except for age, t-tests were performed after log transformation

	AD		PD		MND		MG		Control	
	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%
Ethnic background										
African	0	0.00	2	0.21	0	0.00	0	0.00	984	0.85
Any other Asian background	0	0.00	3	0.32	0	0.00	0	0.00	479	0.41
Any other Black background	0	0.00	0	0.00	0	0.00	0	0.00	22	0.02
Any other mixed background	1	0.66	0	0.00	0	0.00	0	0.00	234	0.20
Any other white background	3	1.97	23	2.43	2	3.08	2	3.45	4390	3.78
Asian or Asian British	0	0.00	0	0.00	0	0.00	0	0.00	9	0.01
Bangladeshi	0	0.00	0	0.00	0	0.00	0	0.00	53	0.05
Black or Black British	0	0.00	0	0.00	0	0.00	0	0.00	5	0.00
British	136	89.47	864	91.33	59	90.77	53	91.38	101087	87.10
Caribbean	1	0.66	5	0.53	1	1.54	1	1.72	1018	0.88
Chinese	1	0.66	2	0.21	1	1.54	0	0.00	566	0.49
Do not know	0	0.00	0	0.00	0	0.00	0	0.00	55	0.05
Indian	3	1.97	8	0.85	0	0.00	0	0.00	1408	1.21
Irish	5	3.29	23	2.43	0	0.00	1	1.72	3053	2.63
Mixed	0	0.00	0	0.00	0	0.00	0	0.00	13	0.01
Other ethnic group	1	0.66	5	0.53	0	0.00	1	1.72	1231	1.06
Pakistani	0	0.00	1	0.11	0	0.00	0	0.00	430	0.37
Prefer not to answer	1	0.66	4	0.42	1	1.54	0	0.00	400	0.34
White	0	0.00	3	0.32	0	0.00	0	0.00	115	0.10
White and Asian	0	0.00	3	0.32	0	0.00	0	0.00	227	0.20
White and Black African	0	0.00	0	0.00	1	1.54	0	0.00	118	0.10
White and Black Caribbean	0	0.00	0	0.00	0	0.00	0	0.00	160	0.14
Prospective memory result (first visit)										
Correct recall on first attempt	25	42.37	222	71.61	20	76.92	16	84.21	28604	78.35
Correct recall on second attempt	18	30.51	60	19.35	6	23.08	2	10.53	6319	17.31
Instruction not recalled, either skipped or incorrect	16	27.12	28	9.03	0	0.00	1	5.26	1584	4.34
Prospective memory result (second visit)										
Correct recall on first attempt	8	53.33	48	75.00	6	75.00	1	50.00	1171	86.48
Correct recall on second attempt	3	20.00	16	25.00	2	25.00	0	0.00	153	11.30

Instruction not recalled, either skipped or incorrect	4	26.67	0	0.00	0	0.00	1	50.00	30	2.22
Prospective memory result (third visit)										
Correct recall on first attempt	6	35.29	51	69.86	10	76.9	6	100.00	4662	84.17
Correct recall on second attempt	4	23.53	16	21.92	2	15.38	0	0.00	662	11.95
Instruction not recalled, either skipped or incorrect	7	41.18	6	8.22	1	7.69	0	0.00	215	3.88

Supplementary Table 3A. Prediction of training and test sets with the multinomial model

Training set						
	Actual diagnosis					
Predicted diagnosis	AD	PD	MND	MG	Control	**Totals**
AD	5	1	0	0	3	**9**
PD	13	170	7	5	8	**203**
MND	0	0	0	0	0	**0**
MG	0	0	0	1	0	**1**
Control	8	59	5	5	565	**642**
Totals	**26**	**230**	**12**	**11**	**576**	**855**
Test set						
	Actual diagnosis					
Predicted diagnosis	AD	PD	MND	MG	Control	**Totals**
AD	1	0	1	0	0	**2**
PD	8	75	2	3	1	**89**
MND	0	1	0	0	0	**1**
MG	0	0	0	1	0	**1**
Control	2	22	2	1	246	**273**
Totals	**11**	**98**	**5**	**5**	**247**	**366**

Supplementary Table 3B. Prediction of training and test sets with the multinomial model excluding demographic measures

Training set						
	Actual diagnosis					
Predicted diagnosis	AD	PD	MND	MG	Control	**Totals**
AD	1	2	0	0	0	**3**
PD	8	150	7	6	3	**174**
MND	0	0	0	0	0	**0**
MG	0	0	0	0	0	**0**
Control	17	78	6	5	579	**685**
Totals	**26**	**230**	**13**	**11**	**582**	**862**
Test set						
	Actual diagnosis					
Predicted diagnosis	AD	PD	MND	MG	Control	**Totals**
AD	1	0	0	0	0	**1**
PD	8	65	2	2	1	**78**
MND	0	1	0	0	0	**1**
MG	0	0	0	0	0	**0**
Control	2	32	3	3	248	**288**
Totals	**11**	**98**	**5**	**5**	**249**	**368**

Supplementary Table 4A. Leave-one-out cross validation of the multinomial model

Variable left out	Accuracy
None	0.8825
Age when attended assessment centre	0.8579
Alanine aminotransferase	0.8825
Albumin	0.8798
Apolipoprotein A	0.8770
Calcium	0.8825
Cholesterol	0.8798
Cystatin C	0.8770
Ethnic background	0.8798
LDL direct	0.8770
Mean time to correctly identify matches	0.8770
Microalbumin in urine	0.8825
Phosphate	0.8825
Prospective memory result (first visit)	0.7596
Prospective memory result (second visit)	0.8825
Sodium in urine	0.8770
Testosterone	0.8699

Supplementary Table 4B. Leave-one-out cross validation of the multinomial model when excluding demographic measures

Variable left out	Accuracy
None	0.8533
Alkaline phosphatase	0.8533
Glucose	0.8533
Mean time to correctly identify matches	0.8505
Microalbumin in urine	0.8533
Phosphate	0.8533
Prospective memory result (first visit)	0.6685
Prospective memory result (third visit)	0.8505
Testosterone	0.8505
Total bilirubin	0.8505

Supplementary Table 5A. Coefficients of the multinomial generalised linear model.

	Intercept	Age when attended assessment centre [y]	Alanine aminotransferase [U/L]	Albumin [g/L]	Apolipoprotein A [g/L]	Calcium [mmol/L]	Cholesterol [mmol/L]	Cystatin C [mg/L]
AD	-42.91	0.24	-8.45×10^{-3}	0.16	-3.70	3.18	1.91	-0.81
PD	7.82	0.17	-8.71×10^{-3}	0.05	-1.82	0.15	0.00	0.78
MND	-1.34	0.13	4.76×10^{-3}	-0.05	-2.53	2.37	2.19	-0.62
MG	-20.01	0.10	5.81×10^{-3}	-0.11	-0.60	1.48	1.82	4.64

	Ethnic background							
	African	British	Caribbean	Chinese	Indian	Irish	White and Asian	Any other Asian background
AD	-16.51	57.11	-31.42	61.05	-12.45	57.24	-13.62	-27.57
PD	29.52	28.88	27.92	-16.01	29.73	29.21	30.89	-23.03
MND	-8.03	37.11	-14.60	-4.80	-5.56	-23.95	-11.28	-12.03
MG	-2.45	54.63	-5.05	0.42	0.17	-17.48	-9.75	-0.60

	Ethnic background							Prospective memory result (first visit)
	Any other white background	Other ethnic group	Prefer not to answer	LDL direct [mmol/L]	Mean time to correctly identify matches [ms]	Microalbumin in urine [mg/L]	Phosphate [mmol/L]	Correct recall on first attempt
AD	-28.14	-38.01	-20.28	-2.84	4.76×10^{-3}	-3.84×10^{-3}	-0.11	-45.39
PD	28.18	10.05	-27.68	-1.01	8.34×10^{-4}	9.75×10^{-4}	-0.66	-46.37
MND	38.48	-32.02	40.74	-3.70	2.78×10^{-3}	1.98×10^{-3}	-3.00	-45.38
MG	-16.76	-16.96	-1.95	-3.92	2.47×10^{-3}	-3.36×10^{-5}	-0.82	-45.15

	Prospective memory result (first visit)		Prospective memory result (second visit)			
	Correct recall on second attempt	Instruction not recalled, either skipped or incorrect	Correct recall on first attempt	Correct recall on second attempt	Sodium in urine [mmol/L]	Testosterone [nmol/L]
AD	-45.21	-43.30	2.89	-29.00	0.0113	0.0049
PD	-46.22	-46.54	0.96	-30.93	0.0108	0.0187
MND	-45.21	-87.36	3.67	-72.33	0.0062	0.0247
MG	-45.79	-75.11	-40.22	-72.16	0.0125	0.0171

Supplementary Table 5B. Coefficients of the multinomial generalised linear model when trained without demographics data.

	Intercept	Alkaline phosphatase [U/L]	Glucose [mmol/L]	Mean time to correctly identify matches [ms]	Microalbumin in urine [mg/L]	Phosphate [mmol/L]
AD	16.45	1.42×10^{-2}	-0.36	6.78×10^{-3}	-1.78×10^{-3}	-0.26
PD	17.95	8.27×10^{-3}	0.26	3.29×10^{-3}	1.62×10^{-3}	-0.25
MND	15.12	1.84×10^{-2}	0.28	4.77×10^{-3}	6.73×10^{-3}	-1.97
MG	14.52	-9.16×10^{-3}	0.43	4.65×10^{-3}	-3.41×10^{-2}	0.37

	Prospective memory result (first visit)			Prospective memory result (third visit)		
	Correct recall on first attempt	Correct recall on second attempt	Instruction not recalled, either skipped or incorrect	Correct recall on first attempt	Correct recall on second attempt	Instruction not recalled, either skipped or incorrect
AD	-22.83	-22.33	-21.06	1.18	-6.22	-22.26
PD	-23.91	-23.65	-24.46	0.09	-9.29	-12.81
MND	-23.52	-22.92	-61.28	0.97	-26.32	-26.97
MG	-22.90	-23.70	-22.92	1.33	-26.67	-36.48

	Testosterone [mmol/L]	Total bilirubin [µmol/L]
AD	9.76×10^{-3}	-0.098
PD	5.31×10^{-2}	-0.018
MND	8.91×10^{-3}	-0.023
MG	1.09×10^{-1}	-0.088

Supplementary Table 6. Shared SNPs identified by genome-wide association studies in at least three neurological diseases.

Affymetrix SNP ID	dbSNP RS ID	Associated gene	OMIM annotation
AD + PD + MND + MG (9 SNPs)			
Affx-6002267	rs7109780	GAB2	N/A
Affx-24035197	rs74357320	LINC00290	N/A
Affx-37460503	rs78987863	ACO1	N/A
Affx-28411577	N/A	HLA-G	Asthma
Affx-35916822	rs75129440	SIX1	Branchiootic syndrome; Deafness
Affx-36372749	rs74725337	HS6ST1	Hypogonadotropic hypogonadism
Affx-25896904	rs11954189	GALNT10	N/A
Affx-11689538	rs11070930	ONECUT1	Amelogenesis imperfecta
Affx-28483230	rs36214709	HLA-DRB6	N/A
AD + PD + MND (26 SNPs)			
Affx-13427189	rs72809485	ZC3H18	N/A
Affx-37147036	rs117265661	LAMA2	Muscular dystrophy
Affx-31386906	rs72675959	TRPS1	Trichorhinophalangeal syndrome
Affx-2949815	rs61847486	PIP4K2A	N/A
Affx-34334471	rs10868716	SPATA31C2	N/A
Affx-30906527	rs75731297	PTPN12	Colon cancer
Affx-28174149	rs35018945	TCP10	N/A
Affx-3592936	rs12415301	PCBD1	Hyperphenylalaninemia
Affx-16808519	rs12624500	SULF2	N/A
Affx-11960689	rs7167705	LOXL1	Exfoliation syndrome
Affx-30291699	rs73687409	CHN2	N/A
Affx-28436162	N/A	DDR1	N/A
Affx-2829210	rs118164677	ADARB2	N/A
Affx-12378781	rs17675037	SNX29	N/A
Affx-4325871	rs7125270	TRIM29	Orofacial cleft; Cleft lip/palate-ectodermal dysplasia syndrome
Affx-2527556	rs76759472	GPAM	N/A
Affx-11085914	rs75517467	SPATA7	Retinitis pigmentosa; Leber congenital amaurosis
Affx-21725390	rs1351752	TBL1XR1	N/A
Affx-9625325	rs11840901	PCDH9	N/A
Affx-37379166	rs77875439	STAU2	N/A
Affx-22420864	rs67895930	DOCK3	N/A
Affx-10898469	rs11624916	ACOT2	N/A
Affx-13293139	rs76924295	CMIP	N/A
Affx-35631741	rs112947130	OR4A47	N/A
Affx-7393142	rs60140646	SLCO1C1	N/A
Affx-29199142	rs599903	HTR1E	N/A

AD + PD + MG (13 SNPs)			
Affx-36971881	rs116542297	LOC100133050	N/A
Affx-21140237	rs76149190	EFCC1	N/A
Affx-36670753	rs114373863	ZBTB20	N/A
Affx-18615618	rs12987681	DNAJC10	N/A
Affx-26144353	rs17076822	HMP19	N/A
Affx-34429450	rs2571346	PTCH1	Holoprosencephaly; Basal cell carcinoma; Basal cell nevus syndrome
Affx-36710670	rs114861488	BCHE	Apnoea
Affx-24011624	rs57948928	LINC00290	N/A
Affx-16041342	rs73543247	PTPRS	N/A
Affx-36393504	rs77796272	NR4A2	N/A
Affx-4132155	rs11226859	GRIA4	N/A
Affx-32549442	rs3103852	XKR9	Otofaciocervical syndrome; Branchiootic syndrome; Anterior segment anomalies; Branchiootorenal syndrome
Affx-14272718	rs76831832	FLJ37644	Campomelic dysplasia
AD + MND + MG (16 SNPs)			
Affx-30422455	rs117218839	VPS41	Wilms tumour susceptibility
Affx-29251151	rs78474144	MAP3K7	N/A
Affx-11505077	rs80243908	TMCO5A	N/A
Affx-7921171	rs4658397	MAP1LC3C	N/A
Affx-12105511	rs117254097	KLHL25	N/A
Affx-10143080	rs12589461	MIR203	N/A
Affx-28360450	rs75984220	BTN3A2	N/A
Affx-36729375	rs115672303	LPP	Lipoma; Leukaemia, acute myeloid
Affx-29549468	rs73211904	ST7	N/A
Affx-36781522	rs116617775	GNPDA2	N/A
Affx-33435971	rs72706624	PSIP1	N/A
Affx-7824085	rs77078859	CHRM3	Eagle-Barrett syndrome
Affx-34329363	rs72749767	CTSL3P	N/A
Affx-25331745	rs115455317	FBXL17	N/A
Affx-33426548	rs62534834	SMARCA2	Nicolaides-Baraitser syndrome
Affx-6287573	rs495344	NOX4	N/A
PD + MND + MG (6 SNPs)			
Affx-28184785	rs79097477	FRMD1	N/A
Affx-35812286	rs76152173	FGF9	Multiple synostoses syndrome
Affx-26957921	rs77838597	ANKRD31	N/A
Affx-36511074	rs79777109	CDH4	N/A
Affx-7008747	rs10849798	SPPL3	N/A
Affx-36416342	rs79916736	CALCRL	N/A

References

Angelova, P. R., Ludtmann, M. H. R., Horrocks, M. H., Negoda, A., Cremades, N., Klenerman, D., Dobson, C. M., Wood, N. W., Pavlov, E. v., Gandhi, S., & Abramov, A. Y. (2016). Ca2+ is a key factor in α-synuclein-induced neurotoxicity. *Journal of Cell Science*, *129*(9), 1792–1801. https://doi.org/10.1242/JCS.180737

Aramillo Irizar, P., Schäuble, S., Esser, D., Groth, M., Frahm, C., Priebe, S., Baumgart, M., Hartmann, N., Marthandan, S., Menzel, U., Müller, J., Schmidt, S., Ast, V., Caliebe, A., König, R., Krawczak, M., Ristow, M., Schuster, S., Cellerino, A., … Kaleta, C. (2018). Transcriptomic alterations during ageing reflect the shift from cancer to degenerative diseases in the elderly. *Nature Communications*, *9*(1). https://doi.org/10.1038/S41467-017-02395-2

Baldacci, F., Mazzucchi, S., della Vecchia, A., Giampietri, L., Giannini, N., Koronyo-Hamaoui, M., Ceravolo, R., Siciliano, G., Bonuccelli, U., Elahi, F. M., Vergallo, A., Lista, S., Giorgi, F. S., & Hampel, H. (2020). The path to biomarker-based diagnostic criteria for the spectrum of neurodegenerative diseases. *Expert Review of Molecular Diagnostics*, *20*(4), 421–441. https://doi.org/10.1080/14737159.2020.1731306

Bassendine, M. F., Taylor-Robinson, S. D., Fertleman, M., Khan, M., & Neely, D. (2020). Is Alzheimer's Disease a Liver Disease of the Brain? *Journal of Alzheimer's Disease*, *75*(1), 1. https://doi.org/10.3233/JAD-190848

Bayraktar, A., Lam, S., Altay, O., Li, X., Yuan, M., Zhang, C., Arif, M., Turkez, H., Uhlén, M., Shoaie, S., & Mardinoglu, A. (2021). Revealing the Molecular Mechanisms of Alzheimer's Disease Based on Network Analysis. *International Journal of Molecular Sciences 2021, Vol. 22, Page 11556*, *22*(21), 11556. https://doi.org/10.3390/IJMS222111556

Chen, X., Huang, Y., Bao, T., Jia, F., Ou, R., Wei, Q., Chen, Y., Liu, J., Yang, J., & Shang, H. (2021). Changes in Serum Cystatin C Levels and the Associations With Cognitive Function in Alzheimer's Disease Patients. *Frontiers in Aging Neuroscience*, *13*, 790939. https://doi.org/10.3389/FNAGI.2021.790939

Chen, X., Yazdani, S., Piehl, F., Magnusson, P. K. E., & Fang, F. (2018). Polygenic link between blood lipids and amyotrophic lateral sclerosis. *Neurobiology of Aging*, *67*, 202.e1-202.e6. https://doi.org/10.1016/J.NEUROBIOLAGING.2018.03.022

Feigin, V. L., Vos, T., Alahdab, F., Amit, A. M. L., Bärnighausen, T. W., Beghi, E., Beheshti, M., Chavan, P. P., Criqui, M. H., Desai, R., Dhamminda Dharmaratne, S., Dorsey, E. R., Wilder Eagan, A., Elgendy, I. Y., Filip, I., Giampaoli, S., Giussani, G., Hafezi-Nejad, N., Hole, M. K., … Murray, C. J. L. (2021). Burden of Neurological Disorders Across the US From 1990-2017: A Global Burden of Disease Study. *JAMA Neurology*, *78*(2), 1. https://doi.org/10.1001/JAMANEUROL.2020.4152

Jakhmola-Mani, R., Islam, A., & Katare, D. P. (2021). Liver-Brain Axis in Sporadic Alzheimer's Disease: Role of Ten Signature Genes in a Mouse Model. *CNS & Neurological Disorders Drug Targets*, *20*(9), 871–885. https://doi.org/10.2174/1871527319666201209111006

Jeong, S. M., Lee, H. R., Jang, W., Kim, D., Yoo, J. E., Jeon, K. H., Jin, S. M., Han, K., & Shin, D. W. (2021). Sex differences in the association between nonalcoholic fatty liver disease and Parkinson's disease. *Parkinsonism & Related Disorders*, *93*, 19–26. https://doi.org/10.1016/J.PARKRELDIS.2021.10.030

Kamer, A. R., Craig, R. G., Niederman, R., Fortea, J., & de Leon, M. J. (2020). Periodontal disease as a possible cause for Alzheimer's disease. *Periodontology 2000*, *83*(1), 242–271. https://doi.org/10.1111/PRD.12327

Kellett, K. A. B., Williams, J., Vardy, E. R. L. C., Smith, A. D., & Hooper, N. M. (2011). Plasma alkaline phosphatase is elevated in Alzheimer's disease and inversely correlates with cognitive function. *International Journal of Molecular Epidemiology and Genetics*, *2*(2), 114. /pmc/articles/PMC3110385/

Lam, S., Bayraktar, A., Zhang, C., Turkez, H., Nielsen, J., Boren, J., Shoaie, S., Uhlen, M., & Mardinoglu, A. (2020). A systems biology approach for studying neurodegenerative diseases. *Drug Discovery Today*, *25*(7), 1146–1159. https://doi.org/10.1016/J.DRUDIS.2020.05.010

Lam, S., Hartmann, N., Benfeitas, R., Zhang, C., Arif, M., Turkez, H., Uhlén, M., Englert, C., Knight, R., & Mardinoglu, A. (2021). Systems Analysis Reveals Ageing-Related Perturbations in Retinoids and Sex Hormones in Alzheimer's and Parkinson's Diseases. *Biomedicines 2021, Vol. 9, Page 1310*, *9*(10), 1310. https://doi.org/10.3390/BIOMEDICINES9101310

Lee, S. H., & Yang, E. J. (2018). Relationship between Liver Pathology and Disease Progression in a Murine Model of Amyotrophic Lateral Sclerosis. *Neuro-Degenerative Diseases*, *18*(4), 200–207. https://doi.org/10.1159/000491392

Lu, Y., Pike, J. R., Selvin, E., Mosley, T., Palta, P., Sharrett, A. R., Thomas, A., Loehr, L., Sidney Barritt, A., Hoogeveen, R. C., & Heiss, G. (2021). Low Liver Enzymes and Risk of Dementia: The Atherosclerosis Risk in Communities (ARIC) Study. *Journal of Alzheimer's Disease : JAD*, *79*(4), 1775–1784. https://doi.org/10.3233/JAD-201241

Mariosa, D., Hammar, N., Malmström, H., Ingre, C., Jungner, I., Ye, W., Fang, F., & Walldius, G. (2017). Blood biomarkers of carbohydrate, lipid, and apolipoprotein metabolisms and risk of amyotrophic lateral sclerosis: A more than 20-year follow-up of the Swedish AMORIS cohort. *Annals of Neurology*, *81*(5), 718–728. https://doi.org/10.1002/ANA.24936

Mehta, C. R., Patel, N., & Senchaudhuri, P. (1992). Exact Stratified Linear Rank Tests for Ordered Categorical and Binary Data. *Journal of Computational and Graphical Statistics*, *1*(1), 21. https://doi.org/10.2307/1390598

Nichols, E., Szoeke, C. E. I., Vollset, S. E., Abbasi, N., Abd-Allah, F., Abdela, J., Aichour, M. T. E., Akinyemi, R. O., Alahdab, F., Asgedom, S. W., Awasthi, A., Barker-Collo, S. L., Baune, B. T., Béjot, Y., Belachew, A. B., Bennett, D. A., Biadgo, B., Bijani, A., bin Sayeed, M. S., … Murray, C. J. L. (2019). Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet. Neurology*, *18*(1), 88. https://doi.org/10.1016/S1474-4422(18)30403-4

Qiang, J. K., Wong, Y. C., Siderowf, A., Hurtig, H. I., Xie, S. X., Lee, V. M. Y., Trojanowski, J. Q., Yearout, D., B. Leverenz, J., Montine, T. J., Stern, M., Mendick, S., Jennings, D., Zabetian, C., Marek, K., & Chen-Plotkin, A. S. (2013). Plasma Apolipoprotein A1 as a Biomarker for Parkinson's Disease. *Annals of Neurology*, *74*(1), 119. https://doi.org/10.1002/ANA.23872

Ray Dorsey, E., Elbaz, A., Nichols, E., Abd-Allah, F., Abdelalim, A., Adsuar, J. C., Ansha, M. G., Brayne, C., Choi, J. Y. J., Collado-Mateo, D., Dahodwala, N., Do, H. P., Edessa, D., Endres, M., Fereshtehnejad, S. M., Foreman, K. J., Gankpe, F. G., Gupta, R., Hankey, G. J., … Murray, C. J. L. (2018). Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet. Neurology*, *17*(11), 939. https://doi.org/10.1016/S1474-4422(18)30295-3

Ryan, K. C., Ashkavand, Z., & Norman, K. R. (2020). The Role of Mitochondrial Calcium Homeostasis in Alzheimer's and Related Diseases. *International Journal of Molecular Sciences*, *21*(23), 1–17. https://doi.org/10.3390/IJMS21239153

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*, *12*(3), e1001779. https://doi.org/10.1371/JOURNAL.PMED.1001779

Toschi, N., Lista, S., Baldacci, F., Cavedo, E., Zetterberg, H., Blennow, K., Kilimann, I., Teipel, S. J., Melo dos Santos, A., Epelbaum, S., Lamari, F., Genthon, R., Habert, M. O., Dubois, B., Floris, R., Garaci, F., Vergallo, A., Hampel, H., Bakardjian, H., … Younsi, N. (2019). Biomarker-guided clustering of Alzheimer's disease clinical syndromes. *Neurobiology of Aging*, *83*, 42–53. https://doi.org/10.1016/J.NEUROBIOLAGING.2019.08.032

Zakharov, S. D., Hulleman, J. D., Dutseva, E. A., Antonenko, Y. N., Rochet, J. C., & Cramer, W. A. (2007). Helical α-Synuclein Forms Highly Conductive Ion Channels†. *Biochemistry*, *46*(50), 14369–14379. https://doi.org/10.1021/BI701275P

Zeng, P., & Zhou, X. (2019). Causal effects of blood lipids on amyotrophic lateral sclerosis: a Mendelian randomization study. *Human Molecular Genetics*, *28*(4), 688–697. https://doi.org/10.1093/HMG/DDY384

Zhang, L., Ding, D., Yu, L., Qi, H., Han, C., Jiang, J., & Jiang, J. (2021). Primary biliary cirrhosis associated with myasthenia gravis after postpartum: a case report. *Journal of Medical Case Reports*, *15*(1). https://doi.org/10.1186/S13256-021-03092-X

Zool, T. J., Türkez, H., Enes ARSLAN, M., di STEFANO, A., Cacciatore, I., & Mardinoğlu, A. (2020). Nonpharmacological treatment options for Alzheimer's disease: from animal testing to clinical studies. *TURKISH JOURNAL OF ZOOLOGY*, *44*(2), 81–89. https://doi.org/10.3906/zoo-1911-32

Zuin, M., Cervellati, C., Trentini, A., Passaro, A., Rosta, V., Zimetti, F., & Zuliani, G. (2021). Association between serum concentrations of apolipoprotein a-i (Apoa-i)

and alzheimer's disease: Systematic review and meta-analysis. *Diagnostics*, *11*(6). https://doi.org/10.3390/DIAGNOSTICS11060984/S1

Paper V: Omics signatures show liver X receptor–driven NASH in congenital generalized lipodystrophy

The provided article below is the author's original version of the manuscript.

Omics signatures show liver X receptor–driven NASH in congenital generalised lipodystrophy

Simon Lam[1], Rui Benfeitas[2], Adil Mardinoglu[1,3], Mujdat Zeybel[4]

[1]Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, London, SE1 9RT, UK

[2]National Bioinformatics Infrastructure Sweden (NBIS), Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, SE-17121, Sweden

[3]Science for Life Laboratory, KTH – Royal Institute of Technology, Stockholm, SE-17121, Sweden

[4]Digestive Diseases Biomedical Research Centre, University of Nottingham, Nottingham, NG7 2UH, UK

Abstract

Lipodystrophy is a rare degenerative disorder characterised by the total or near-total loss of adipose cells. Individuals with lipodystrophy often have metabolic abnormalities, such as leptin deficiency and heightened *de novo* lipogenesis. Accumulation of fat in the liver can lead to nonalcoholic fatty liver disease and nonalcoholic steatohepatitis (NASH). This is a major health problem and liver complications occur in over 70% of lipodystrophy patients.

Congenital generalised lipodystrophy (CGL) is birth-onset and has been linked to mutations in the *AGPAT2* and *BSCL2* genes. Clinical presentations of *AGPAT2* and *BSCL2* lipodystrophy patients are heterogeneous, but important differences, such as onset of bone cysts, diabetes, and liver complications have been identified. Here, we

used an omics approach to molecularly profile *AGPAT2* and *BSCL2* patients and to identify molecular differences with NASH patients without lipodystrophy. Our findings show differential methylation of the liver X receptor gene *OSBLP9* and upregulation of *de novo* lipogenesis genes *SCD* and *ME1* in CGL compared to NASH. These results indicate a role for sterol signalling in the onset of NASH in CGL patients, and suggests that the CGL NASH phenotype may represent one subclass of heterogeneous NASH.

Introduction

Congenital generalised lipodystrophies (CGLs) are a group of rare heterogeneous degenerative disorders characterised by total or near-total loss of adipose tissue. Hallmarks of CGL include loss of adipocytokines such as leptin, a muscular appearance due to the loss of adipose tissue, and increased *de novo* lipogenesis. It has been shown that genetic mutations in the *AGPAT2* and *BSCL2* genes are causative for the vast majority of CGL cases (Agarwal et al., 2004; Magré et al., 2003). *AGPAT2* encodes an enzyme involved in *de novo* phospholipid biosynthesis, whereas *BSCL2* encodes seipin, a protein involved in lipid droplet formation. *CAV1* and *CAVIN1 (PTRF*) mutations are also linked to a smaller proportion of CGL cases (Kim et al., 2008; Shastry et al., 2010). *CAV1* and *CAVIN1* encode protein products involved in caveolae formation. CGL differs from acquired generalised lipodystrophy, which does not have a direct genetic cause, but rather is caused by autoimmune disorders or can be idiopathic.

Individuals with CGL commonly have organ complications, the most common of which being the liver (Akinci et al., 2019) with over 70% of patients developing complications. These complications are due to ectopic fat accumulation leading from increased *de novo* lipogenesis, which can progress to nonalcoholic fatty liver disease (NAFLD) and eventually nonalcoholic steatohepatitis (NASH) and hepatic cirrhosis (Polyzos et al., 2019).

Diabetes, fatty liver diseases such as NAFLD and NASH, and other obesity-associated conditions are sometimes called the silent epidemic due to their high global burden but uncorrespondingly low levels of public and scientific attention according to its severity (Rippe, 2021). The global prevalence of NAFLD is 24% (Younossi et al., 2017), and this share stands to increase as the silent epidemic of obesity deepens. However, NAFLD encompasses a range of conditions, including fibrosis, steatosis, cirrhosis, and hepatocellular carcinoma. It is a broadly defined disease with a wide set of presentations and manifestations, and is clinically and molecularly heterogeneous (Eslam et al., 2018).

Systems biology approaches using multi-omics have been effective in the past to study heterogeneous liver diseases. Here, we examined the methylome and

transcriptome, the full complement of DNA methylation states and RNA expression, respectively, of individuals with NASH, *AGPAT2* lipodystrophy, and *BSCL2* lipodystrophy. Briefly, our findings point to a role for sterol signalling via the lipid X receptor (LXR) encoded by *OSBLP9*, which we identified in differential methylation study. LXR signalling is accomplished via sterol group binding, which is driven by *de novo* lipogenesis genes such as *SCD* and *ME1*, which we identified as significantly upregulated in CGL compared to NASH in differential gene expression analysis. Taken together, these results may be interpreted as being a single subset of heterogeneous NASH which is converged upon in CGL. If proven true, then the implications of this result include the possibility for precision therapy for NASH patients who also have CGL.

Methods

DNA methylation analysis

From liver samples of four individuals with *AGPAT2* lipodystrophy, four with *BSCL2* lipodystrophy, twelve with NASH, and four healthy individuals, DNA methylation intensities at 833,136 methylation sites were determined by MethylationEPIC array (Infinium). Raw intensities were analysed and quality controlled with minfi (version 1.36.0) (Aryee et al., 2014; Fortin et al., 2017). Differential methylation was quantified pairwise between all disease states using limma (version 3.46.0) (Ritchie et al., 2015). Differential methylation within promoter regions and enriched GO terms was determined with mCSEA (version 1.14.0) (Martorell-Marugán et al., 2019) and missMethyl (version 1.28.0) (Phipson et al., 2016), respectively. Sites and promoters were considered significantly differentially methylated if they displayed a p-value of 0.05 or smaller after Benjamini-Hochberg adjustment. Revigo (http://revigo.irb.hr/, accessed 15-10-2021) (Supek et al., 2011) was used to remove redundant GO terms. Differential methylation of promoters was used to determine differentially active transcription factors using TRANSFAC and JASPAR gene regulatory networks obtained from Enrichr (https://maayanlab.cloud/Enrichr/, accessed 15-04-2021) (Chen et al., 2013; Kuleshov et al., 2016).

Gene expression analysis

From gene expression counts of four *AGPAT2* lipodystrophy, fourteen NASH, and two healthy samples, DESeq2 (version 1.30.1) (Love et al., 2014) was used to determine differentially expressed genes (DEGs) between all pairs of disease states. DEGs were analysed using piano (version 2.6.0) (Väremo et al., 2013) to identify up- and down-regulated enriched GO terms. Differentially expressed genes and enriched GO terms were considered statistically significant if they displayed a p-value of 0.05 or smaller after Benjamini-Hochberg adjustment. Revigo was used to remove redundant GO terms.

Additional analysis and visualisation

Molecular similarity between each pair of disease states was estimated by inspecting p-values obtained from analysis of differentially methylated probes, differentially methylated promoters, and differentially expressed genes.

Circos visualisations were generated using circlize (version 0.4.13).

All data analysis steps were performed in R (version 4.0.2) unless otherwise specified.

Data availability

Full results and all original computer code is available on request from the authors.

Results

Differential methylation analysis indicates molecular similarities between *AGPAT2* and *BSCL2* lipodystrophies

We obtained DNA methylation data from patients with *AGPAT2* or *BSCL2* lipodystrophy from an Infinium MethylationEPIC array. To identify epigenetic changes in lipodystrophy compared to NASH patients and healthy individuals, we performed statistical analysis on these data (Methods) and identified differentially methylated genes, promoters, and probes (**Figure 1**).

Overall, we found high similarity between *AGPAT2* and *BSCL2* lipodystrophies in terms of GO term enrichment of differentially methylated genes (**Figure 1a**) and differentially methylated probes (**Figure 1c**). Methylation profiles in *AGPAT2* and *BSCL2* lipodystrophies differed from that of NASH, with *AGPAT2* lipodystrophy featuring differential methylation in multiple tissue differentiation and development genes, such as in haematopoietic progenitor cells, kidney cells, neurons, and brown fat cells. *BSCL2* lipodystrophy patients had differential methylation compared to NASH patients in genes involved in DNA binding and transcription activity. However, the methylation profiles of *AGPAT2* and *BSCL2* lipodystrophy patients were very similar, with differential methylation only occurring in neurotransmitter genes.

Considering individual DNA methylation sites (**Figure 1c**), we found only three sites with significant differential methylation between *AGPAT2* and *BSCL2* lipodystrophy. These were cg23461147, cg21600611, and cg18211066, which correspond to the LXR gene *OSBPL9*, cytoskeletal gene *PALLD*, and miRNA-related gene *TNRC6C*, respectively. In comparisons with NASH profiles, *AGPAT2* had differential methylation in cg05176131, which is in the spermatogenesis gene *SPATA20*, and in cg09731288, which is in the tumour suppressor gene *PCDC4*. *BSCL2* lipodystrophy samples had differential methylation at the aforementioned cg23461147 and cg21600611 sites, and these were above the levels seen between *AGPAT2* and *BSCL2* lipodystrophies. Other highly differentially methylated sites in *BSCL2* lipodystrophy compared to NASH were cg18889307, which is within TGFB induced factor homeobox 1 (*TGIF1*), and cg01064804, which is within the antisense

RNA gene *CFAP418-AS1*. No probes within *AGPAT2* nor *BSCL2* gene loci were significantly differentially methylated in the respective diseases, indicating no role for methylation of the causative genes in CGL.

Methylation at promoters was highly variable between all comparisons (**Figure 1b**). This confirms a high degree of heterogeneity within the patient pools. We did not consider promoter methylation further.

Taken together, *AGPAT2* and *BSCL2* lipodystrophies appeared to be molecularly similar according to DNA methylation data, despite some departures when comparing with NASH patients. We therefore continued our investigation by considering *AGPAT2* lipodystrophy patients only.

Differential expression reveals distinct molecular profiles between lipodystrophy and NASH

In differential methylation analysis, we observed DNA methylation changes at genes involves in lipid metabolism and gene expression. In order to deduce whether any of these changes were associated with actual changes to gene expression, we obtained RNA sequencing data from the livers of *AGPAT2* lipodystrophy, NASH, and healthy individuals and performed differential expression analysis (**Figure 2**).

In differential expression, we found transcriptome-wide changes in expression in *AGPAT2* lipodystrophy, NASH, and healthy livers. There were more significantly differentially expressed genes between *AGPAT2* lipodystrophy versus healthy (**Figure 2a**) and between NASH versus healthy (**Figure 2c**) compared with between *AGPAT2* lipodystrophy and NASH (**Figure 2b**). In general, the identity of the differentially expressed genes were different with the exception of malic enzyme 1 (*ME1*) and apolipoprotein A4 (*APOA4*) genes, which were significantly upregulated in *AGPAT2* lipodystrophy versus both NASH and healthy individuals. Consistent with differential methylation data, *AGPAT2* was not significantly differentially expressed in AGPAT2 lipodystrophy samples compared to healthy. This suggests a post-transcriptional role for *AGPAT2* in the progression of AGPAT2 lipodystrophy, consistent with the disease being characterised by loss-of-function mutation rather than total gene loss.

Considering the coregulation of gene sets and GO term enrichment, we found that *AGPAT2* lipodystrophy and NASH shared few GO term changes compared to healthy, with the exception of upregulated cholesterol biosynthesis and lipid metabolism, and downregulated RNA processing (**Figure 2d**). In fact, we found large changes in cytoskeletal functions, energy metabolism, and lipid transport, and protein processing between *AGPAT2* lipodystrophy and NASH, indicating large molecular differences in the two diseases despite the relatively few differentially expressed genes.

Next, we inspected changes in expression of lipid metabolism and control genes (**Figure 3**). We found *AGPAT2* lipodystrophy–specific upregulation of stearoyl-CoA desaturase (*SCD*) gene expression, which, along with *ME1*, is responsible for *de novo* lipogenesis. Transcription factor *FOXO1* and glycerol-3-phosphate acyltransferase *GPAT3* genes were both downregulated in *AGPAT2* lipodystrophy and NASH alike. Most other lipid metabolism genes varied in the same direction in the two diseases compared to healthy individuals, with the exception of lipid transport genes *APOA1*, *APOB*, and *APOD*, and triglyceride synthesis genes *AGPAT1*, *AGPAT5*, *LCLAT1*, *MOGAT1*, *MOGAT3*, and *PLPP1*. This indicates differential utilisation of apolipoproteins and acylglycerols in *AGPAT2* lipodystrophy and NASH, however, these changes did not meet the threshold for statistical significance.

Taken together, our results indicate that fatty liver arising from CGL is molecularly distinct from other causes of NASH.

Discussion

In our work, we have shown that the molecular signatures of *AGPAT2* and *BSCL2* lipodystrophies are distinct from that of NASH (**Figure 4**). Causes of CGL by mutation in *AGPAT2* and *BSCL2* resulted in DNA methylation profiles which did not differ much from each other, but differed greatly from NASH. This work indicates that the fatty liver phenotype afforded by CGL may be distinct from other causes of NASH.

The DNA methylation differences in *AGPAT2* and *BSCL2* lipodystrophy revealed the potentially interesting finding of differential methylation at the *OSBPL9* gene, and the only differentially methylated gene in this comparison involved in lipid function (**Figure 1**). Its protein product is an intracellular oxysterol receptor whose function is bile synthesis. Methylation at this site may indicate modulation of bile synthesis. Indeed, it has been shown that enhanced bile acid levels by mitochondrial sterol pathways are associated with NASH-driven HCC (Conde de la Rosa et al., 2021) and that modulation of bile acid (Fiorucci et al., 2020; Xu et al., 2021) and LXRs (Aguayo-Orozco et al., 2018; Hong & Tontonoz, 2014) may have therapeutic value to NASH patients. The appearance of an LXR gene and bile acid synthesis gene *OSBPL9* in our CGL work is therefore a significant finding, potentially suggesting a genetic cause of NASH.

However, our work also showed significant differential expression profiles between *AGPAT2* lipodystrophy and NASH (**Figures 2 & 3**), indicating that despite the link to NASH via *OSBPL9* DNA methylation, there still exists a degree of transcriptomic changes between the two diseases. For instance, we identified a lipid metabolism and transcriptional control signature of upregulated *SCD*, *ME1*, and *APOA4* which was specific to *AGPAT2* lipodystrophy but not NASH. The upregulation of both *ME1* and *SCD* genes is consistent with the prior observation of differential methylation via *OSBPL9* due to the role of ME1 in regenerating NADPH for cholesterol biosynthesis (Simmen et al., 2020) and SCD in being transcriptionally regulated by sterols via binding of LXR, leading to generation of monounsaturated fatty acids and potential metabolic reprogramming towards a cancer state (Kikuchi & Tsukamoto, 2020). Although there was a signature for a tendency for *SCD* and

ME1 to be upregulated in NASH, this signature did not reach the threshold for statistical significance.

More generally, our differential expression results in *AGPAT2* lipodystrophy appeared to mirror the trends seen in NASH with striking similarity, even if statistical significance was not reached for the same genes in both diseases. Given the known molecular heterogeneity among NASH patients (Bosley et al., 2017), it is therefore likely that the NASH gene expression signature indicates common changes that are central to NASH, whereas the *AGPAT2* lipodystrophy gene expression signature highlights one possible, more unified, expression profile that can lead to NASH. Whether this expression profile is adopted by individuals with NASH but without *AGPAT2* lipodystrophy remains to be seen.

If confirmed, then this, to our knowledge, is the first evidence suggesting convergent metabolic perturbations in NASH and CGL, and that therapeutics could potentially be repositioned between the two, at least in metabolic resemblant subsets of disease. The identification of *OSBL9*, *SCD*, and *ME1* varying in common between the two diseases suggests that there may exist further liver genes that could be therapeutically exploited. Further experimentation in a NASH or CGL model and modulation of these and other liver function fenes could validate the drug discovery potential of these genes.

It is important to acknowledge the weaknesses of this study. This study did not use patient-matched DNA methylation and RNA sequencing samples, so we could not determine whether the patients who had differential *OSBPL9* methylation also had upregulated *SCD* and *ME1*. Knowing this would strengthen the link between LXR-mediated sterol signalling and CGL-driven NASH. Additionally, we did not attempt to address the molecular heterogeneity in our NASH sample pool owing to the fact that we only had 12 DNA methylation samples and 14 RNA sequencing samples from individuals with NASH. A larger sample size would have allowed us to perform unsupervised clustering for the identification of disease subclasses on the basis of methylation and gene expression profiles alone. It remains to be seen whether upregulated LXR and sterol signalling exists as a distinct molecular profile leading to NASH could be identified as a NASH subclass in this way.

Taken together, these results indicate that the NASH phenotype caused by *AGPAT2* lipodystrophy might be a subclass of heterogeneous NASH. More work needs to be done to ascertain whether *AGPAT2* mutations do indeed lead to LXR-mediated NASH, and whether alternative mechanisms of action may be involved. Future investigation should also focus on whether NASH patients with upregulated LXR and sterol signalling have a distinct molecular profile. If proven true, then this potentially enables precision therapies of NASH patients who also have a diagnosis of CGL.

Figures

Figure 1. Differential methylation of **a)** genes by GO term enrichment, **b)** promoters, and **c)** probes. Stars indicate statistical significance following Benjamini-Hochberg p-value adjustment: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.005$.

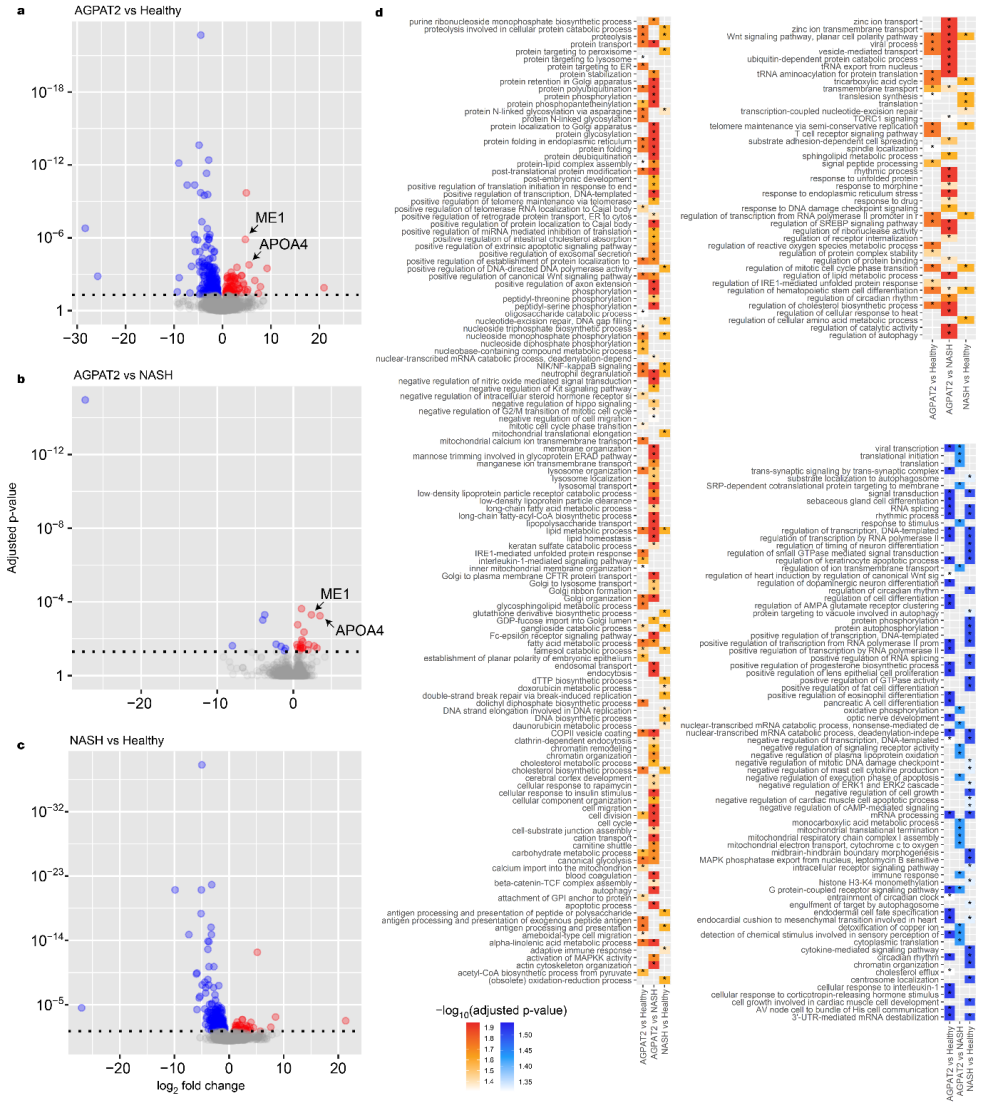


Figure 2. Differential expression. Differentially expressed genes in **a)** *AGPAT2* lipodystrophy vs healthy controls, **b)** *AGPAT2* lipodystrophy vs NASH, and **c)** NASH vs healthy controls. Dashed horizontal line indicates p-value cut-off for statistical significance ($p < 0.05$). **d)** GO term enrichment of differentially expressed genes. Red scale indicates upregulation; blue scale indicates downregulation. Stars indicate statistical significance following Benjamini-Hochberg p-value adjustment: *, $p < 0.05$.

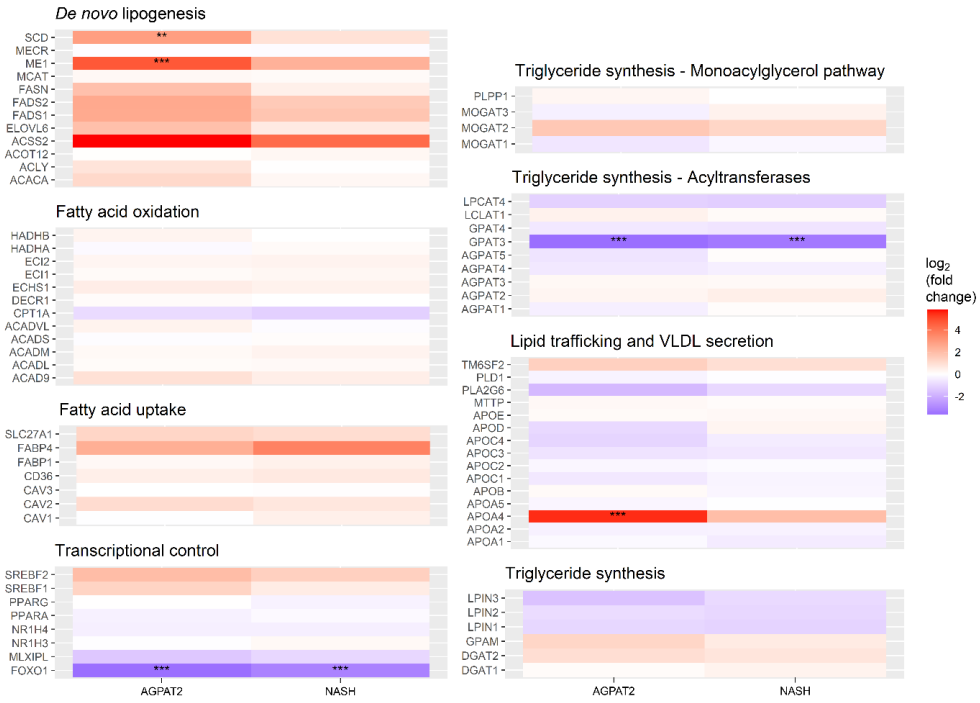


Figure 3. Expression changes compared to healthy controls in lipid metabolism genes. Red cells indicate upregulation; blue cells indicate downregulation. Stars indicate statistical significance following Benjamini-Hochberg p-value adjustment: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

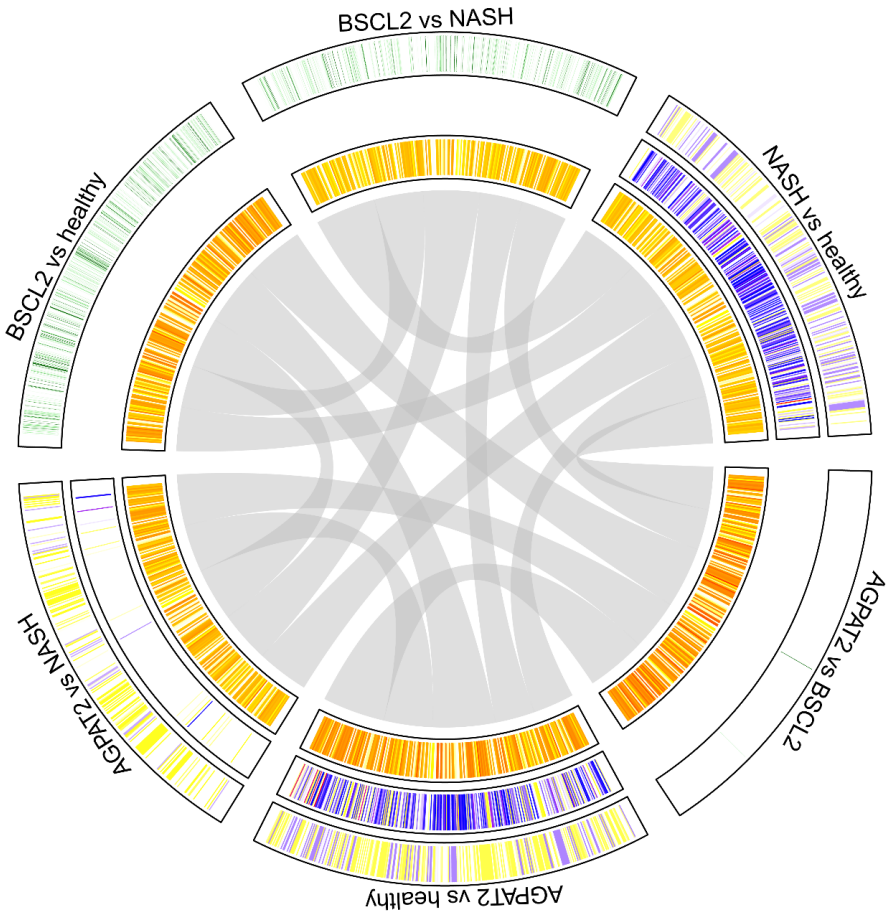


Figure 4. Overview of differential methylation and expression. Radial tracks (outer to inner): differential GO terms, expressed genes, transcription factors. Inset: shared changes in transcription factors; chord positions do not indicate actual transcription factors shared. For the GO term track, changed GO terms are based on differential methylation evidence for comparisons with *BSCL2* lipodystrophy and on differential gene expression evidence otherwise. GO term and expressed genes tracks: red, upregulated terms; purple, downregulated terms; green, nondirectional-specific changed terms; higher colour intensity indicates smaller adjusted p-value. Transcription factor track: higher red colour intensity indicates more differentially methylated promoters at which the transcription factor binds. Coordinates along tracks are the same in each sector.

Acknowledgements

This work was supported by the Knut and Alice Wallenberg Foundation (grant number 2017.0303) to A.M.

Author contributions

A.M. and M.Z. supervised the study. M.Z. designed the study. S.L. performed all analyses with technical consultation from R.B. S.L. wrote the manuscript with input from all the authors.

Conflicts of interest

The authors declare no competing financial interests.

References

Agarwal, A. K., Barnes, R. I., & Garg, A. (2004). Genetic basis of congenital generalized lipodystrophy. *International Journal of Obesity and Related Metabolic Disorders : Journal of the International Association for the Study of Obesity*, *28*(2), 336–339. https://doi.org/10.1038/SJ.IJO.0802487

Aguayo-Orozco, A., Bois, F. Y., Brunak, S., & Taboureau, O. (2018). Analysis of Time-Series Gene Expression Data to Explore Mechanisms of Chemical-Induced Hepatic Steatosis Toxicity. *Frontiers in Genetics*, *9*(SEP), 396. https://doi.org/10.3389/FGENE.2018.00396/FULL

Akinci, B., Oral, E. A., Neidert, A., Rus, D., Cheng, W. Y., Thompson-Leduc, P., Cheung, H. C., Bradt, P., Foss De Freitas, M. C., Montenegro, R. M., Fernandes, V. O., Cochran, E., & Brown, R. J. (2019). Comorbidities and Survival in Patients With Lipodystrophy: An International Chart Review Study. *The Journal of Clinical Endocrinology & Metabolism*, *104*(11), 5120–5135. https://doi.org/10.1210/JC.2018-02730

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., & Irizarry, R. A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*, *30*(10), 1363–1369. https://doi.org/10.1093/BIOINFORMATICS/BTU049

Bosley, J., Boren, C., Lee, S., Grøtli, M., Nielsen, J., Uhlen, M., Boren, J., & Mardinoglu, A. (2017). Improving the economics of NASH/NAFLD treatment through the use of systems biology. *Drug Discovery Today*, *22*(10), 1532–1538. https://doi.org/10.1016/J.DRUDIS.2017.07.005

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. v., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, *14*, 128. https://doi.org/10.1186/1471-2105-14-128

Conde de la Rosa, L., Garcia-Ruiz, C., Vallejo, C., Baulies, A., Nuñez, S., Monte, M. J., Marin, J. J. G., Baila-Rueda, L., Cenarro, A., Civeira, F., Fuster, J., Garcia-Valdecasas, J. C., Ferrer, J., Karin, M., Ribas, V., & Fernandez-Checa, J. C. (2021). STARD1 promotes NASH-driven HCC by sustaining the generation of bile acids through the alternative mitochondrial pathway. *Journal of Hepatology*, *74*(6), 1429–1441. https://doi.org/10.1016/J.JHEP.2021.01.028

Eslam, M., Valenti, L., & Romeo, S. (2018). Genetics and epigenetics of NAFLD and NASH: Clinical impact. *Journal of Hepatology*, *68*(2), 268–279. https://doi.org/10.1016/J.JHEP.2017.09.003

Fiorucci, S., Biagioli, M., Sepe, V., Zampella, A., & Distrutti, E. (2020). Bile acid modulators for the treatment of nonalcoholic steatohepatitis (NASH). *Expert Opinion on Investigational Drugs*, *29*(6), 623–632. https://doi.org/10.1080/13543784.2020.1763302

Fortin, J. P., Triche, T. J., & Hansen, K. D. (2017). Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi.

Bioinformatics, *33*(4), 558–560. https://doi.org/10.1093/BIOINFORMATICS/BTW691

Hong, C., & Tontonoz, P. (2014). Liver X receptors in lipid metabolism: opportunities for drug discovery. *Nature Reviews Drug Discovery 2014 13:6*, *13*(6), 433–444. https://doi.org/10.1038/nrd4280

Kikuchi, K., & Tsukamoto, H. (2020). Stearoyl-CoA Desaturase and Tumorigenesis. *Chemico-Biological Interactions*, *316*, 108917. https://doi.org/10.1016/J.CBI.2019.108917

Kim, C. A., Delépine, M., Boutet, E., el Mourabit, H., le Lay, S., Meier, M., Nemani, M., Bridel, E., Leite, C. C., Bertola, D. R., Semple, R. K., O'Rahilly, S., Dugail, I., Capeau, J., Lathrop, M., & Magré, J. (2008). Association of a homozygous nonsense caveolin-1 mutation with Berardinelli-Seip congenital lipodystrophy. *The Journal of Clinical Endocrinology and Metabolism*, *93*(4), 1129–1134. https://doi.org/10.1210/JC.2007-1328

Kuleshov, M. v, Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., & Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, *44*(W1), W90–W97. https://doi.org/10.1093/nar/gkw377

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9

Magré, J., Delépine, M., van Maldergem, L., Robert, J. J., Maassen, J. A., Meier, M., Panz, V. R., Kim, C. A., Tubiana-Rufi, N., Czernichow, P., Seemanova, E., Buchanan, C. R., Lacombe, D., Vigouroux, C., Lascols, O., Kahn, C. R., Capeau, J., & Lathrop, M. (2003). Prevalence of mutations in AGPAT2 among human lipodystrophies. *Diabetes*, *52*(6), 1573–1578. https://doi.org/10.2337/DIABETES.52.6.1573

Martorell-Marugán, J., González-Rumayor, V., & Carmona-Sáez, P. (2019). mCSEA: detecting subtle differentially methylated regions. *Bioinformatics*, *35*(18), 3257–3262. https://doi.org/10.1093/BIOINFORMATICS/BTZ096

Phipson, B., Maksimovic, J., & Oshlack, A. (2016). missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics (Oxford, England)*, *32*(2), 286–288. https://doi.org/10.1093/BIOINFORMATICS/BTV560

Polyzos, S. A., Perakakis, N., & Mantzoros, C. S. (2019). Fatty liver in lipodystrophy: A review with a focus on therapeutic perspectives of adiponectin and/or leptin replacement. *Metabolism*, *96*, 66–82. https://doi.org/10.1016/J.METABOL.2019.05.001

Rippe, J. M. (2021). The Silent Epidemic. *The American Journal of Medicine*, *134*(2), 164. https://doi.org/10.1016/J.AMJMED.2020.09.028

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47. https://doi.org/10.1093/nar/gkv007

Shastry, S., Delgado, M. R., Dirik, E., Turkmen, M., Agarwal, A. K., & Garg, A. (2010). Congenital Generalized Lipodystrophy, Type 4 (CGL4) Associated with Myopathy due to Novel PTRF Mutations. *American Journal of Medical Genetics. Part A*, *152A*(9), 2245. https://doi.org/10.1002/AJMG.A.33578

Simmen, F. A., Alhallak, I., & Simmen, R. C. M. (2020). Malic Enzyme 1 (ME1) in the Biology of Cancer: It's not Just Intermediary Metabolism. *Journal of Molecular Endocrinology*, *65*(4), R77. https://doi.org/10.1530/JME-20-0176

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*, *6*(7), e21800. https://doi.org/10.1371/journal.pone.0021800

Väremo, L., Nielsen, J., & Nookaew, I. (2013). Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Research*, *41*(8), 4378–4391. https://doi.org/10.1093/NAR/GKT111

Xu, Y., Zhu, Y., Hu, S., Xu, Y., Stroup, D., Pan, X., Bawa, F. C., Chen, S., Gopoju, R., Yin, L., & Zhang, Y. (2021). Hepatocyte Nuclear Factor 4α Prevents the Steatosis-to-NASH Progression by Regulating p53 and Bile Acid Signaling (in mice). *Hepatology (Baltimore, Md.)*, *73*(6), 2251–2265. https://doi.org/10.1002/HEP.31604

Younossi, Z., Anstee, Q. M., Marietti, M., Hardy, T., Henry, L., Eslam, M., George, J., & Bugianesi, E. (2017). Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nature Reviews Gastroenterology & Hepatology 2017 15:1*, *15*(1), 11–20. https://doi.org/10.1038/nrgastro.2017.109

Paper VI: Regulation of lipid signalling by the polycomb repressor complex protein EZH2 during regeneration drives muscle regeneration after injury

The provided article below is the author's original version of the manuscript.

Regulation of lipid signalling by the polycomb repressor complex protein EZH2 drives muscle regeneration after injury

Simon Lam[1], Adil Mardinoglu[1,2], Robert Knight[1]

[1]Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, London, United Kingdom

[2]Science for Life Laboratory, KTH – Royal Institute of Technology, Stockholm, Sweden

Abstract

Muscle stem cell (muSC) proliferation following injury is critical for generating a replacement cell population. In ageing and many diseases, this is perturbed, resulting in fibrosis and muscle weakness. To identify molecules regulating muSC proliferation, we focused on a master regulator of muSCs, the polycomb repressor complex protein EZH2. Using zebrafish zygotic mutants, we observed elevated proliferation of muSCs in *ezh2* mutants, resulting in delayed differentiation and elevated fusion. Similar phenotypes were observed when EZH2 function was temporally inhibited using 3-deazaneplanocin A (DZNep), a small molecule inhibitor of the catalytic S-adenosylmethionine-dependent methyltransferase (SAM) domain of EZH2.

To identify molecular signatures associated with loss of EZH2 function in regenerating muscle, we dissected tissue and performed RNA-seq. We identified an upregulation of genes associated with lipid transport and metabolism. Mapping of gene expression changes relative to metabolic association networks highlighted an enrichment of genes involved in very long chain fatty acid elongation and

arachidonic acid signalling. We confirmed trends favouring synthesis of longer lipid chains using lipidomic analysis.

Keywords: inflammation, arachidonic acid, degeneration, regeneration, injury, muscle, zebrafish, systems biology

Introduction

In ageing and disease, replenishment of muscle stem cells (muSCs) after injury is often defective, resulting in fibrosis and muscle weakness. Enhancer of zeste homologue 2 (EZH2) is a master regulator of muSCs, and has previously been studied in the context of cancers[1] and degenerative diseases[2,3]. It is required in adult cells for maintenance of cell identity and is lost after muscle injury to allow muSC proliferation. It has been shown that loss of EZH2 by maternal zygotic mutation or small molecule inhibition results in increased proliferation and elevated fusion[4,5], but its precise mechanism of action is unclear.

EZH2 is a member of the polycomb repressive complex 2 (PRC2). It works with suppressor of zeste 12 (Suz12) and embryonic ectoderm development (Eed) to place trimethyl groups on lysine 27 of histone 3 ($H3K27Me_3$)[6]. $H3K27Me_3$ is an epigenetic hallmark indicative of heterochromatin regions within which genes are silenced. In adult tissue, EZH2 is active and its target genes are transcriptionally silenced. Upon muscle injury, EZH2 is lost, enabling activation of genes at those target loci and engagement of a muscle regeneration programme. Loss of EZH2 by inhibition has been shown to enhance the extent of this programme, resulting in an enhanced muSC proliferative phenotype. Here, we used the zebrafish *ezh2*$^{hu5670/5670}$ stop-gain zygotic mutant to determine whether loss of functional Ezh2 would also result in a super-regenerative phenotype, and if so, to use RNA sequencing analysis and metabolic modelling to molecularly characterise the mutant. Upon muscle injury, we found activation of very long chain fatty acid elongation, arachidonic acid metabolism, inflammation signalling, and immune cell recruitment, and all of these characteristics were exacerbated in the *ezh2* mutant. These results suggest that very long chain fatty acid modulation could be potentially exploited to expedite muscle regeneration and reduce tissue healing times following injury.

Methods

Zebrafish sample acquisition and quality control

Zebrafish (*Danio rerio*) of either *ezh2*$^{hu5670/5670}$ (n = 6) or wildtype (n = 6) genotype were maintained in accordance to national and institutional animal care guidelines. Animals were randomly assigned injured (n = 6) or noninjured (n = 6) groups with even genotype distribution. Animals in the injured group were injured seven days post fertilisation. Animals were dissected 24 hours post injury (or 24 hours following the time the injury would have taken place, for the noninjured group).

Total RNA was obtained from dissected samples and analysed in a 2100 Bioanalyzer (Agilent) instrument for sample quality control. Following sample quality control, one sample corresponding to a noninjured *ezh2* mutant was discarded. All other samples passed quality control and were sequenced on a BGISeq-500 (BGI, Shenzhen, China) in paired-end mode.

Transcriptomic analysis

Raw sequencing reads were aligned to the Ensembl release 104 *Danio rerio* GRCz11 cDNA assembly[7] (http://ftp.ensembl.org/pub/release-104/fasta/danio_rerio/cdna/Danio_rerio.GRCz11.cdna.all.fa.gz, accessed 2021-05-10) and mRNA expression was quantified in transcripts per million (TPM) in paired-end mode using kallisto[8] (version 0.46.1, Pachter Lab).

TPMs were analysed to determine differential expression using DESeq2[9] (version 1.30.1). Differential expression results were used to determine functional enrichment using piano[10] (version 2.6.0), using GO terms obtained from Ensembl Biomart (https://www.ensembl.org/biomart/martview, accessed 2021-06-23) to generate the gene sets for enrichment testing. Differentially expressed genes and enriched GO terms were considered statistically significant if they displayed a p-value of 0.05 or smaller after Benjamini-Hochberg adjustment.

To filter differentially expressed genes of interest, genes related to inflammation were selected by searching for the keywords “TNF”, “caspase”, “interleukin”, and “chemokine” in the ZFIN gene description. Arachidonic acid pathway genes were selected from the *Zebrafish1* genome-scale metabolic model (GEM)[11] by searching

for the keyword “arachidon” in reaction equations. All genes associated with those reactions were selected as arachidonic acid pathway genes. Immune cell profiling was performed by inspecting enrichment of differentially expressed genes using the immune cell profiles described by Lyons et al[12].

All transcriptomic analysis steps were performed in the R programming language (version 4.0.2).

Metabolic modelling

Reporter metabolite analysis was performed using the *reporterMetabolites* function[13] in the RAVEN Toolbox 2.0[14], using the differential expression results and the *ZebraGEM2.1* [15] and *Zebrafish1* reference GEMs.

All metabolic modelling analysis steps were performed in MATLAB R2021a.

Network analysis

A previously described zebrafish interactome by Alexeyenko et al[16] was accessed and explored to find all first- and second-degree interactions with *ezh2*. All first- and second-orders interaction partners with *ezh2* were tested to see if they also interacted with an arachidonic acid pathway gene.

All network analysis steps were performed in the R programming language (version 4.0.2).

Bioinformatic cross-validation

Microarray gene expression data for maternal zygotic *ezh2* mutant and wildtype zebrafish embryos were obtained from the Gene Expression Omnibus (GEO) under accession number GSE64618[17]. Differential gene expression analysis, gene set enrichment analysis, and reporter metabolite analysis were performed as above.

Differential gene expression profiles from the present study were mapped onto expression profiles contained within the SysMyo Muscle Gene Sets[18]. Hypergeometric tests were performed to estimate the goodness of fit.

ATAC-seq chromatin accessibility data for zebrafish embryos treated and untreated with a methyltransferase inhibitor were obtained from the GEO under accession number GSE140233[5]. ATAC-seq peaks were mapped to the *Danio rerio* UCSC

TxDb genome annotation (version 10, available at: https://bioconductor.org/packages/release/data/annotation/html/TxDb.Drerio.UCSC.danRer10.refGene.html, accessed 2021-12-06) to deduce genes in open chromatin. The list of genes in open chromatin was inspected and compared with our differentially expressed genes.

Lipidomic analysis

Lipidomic data were collected for *ezh2*$^{hu5670/5670}$ (n = 4), *ezh2*$^{hu5670/+}$ (n = 4), and wildtype (n = 4) zebrafish embryos. Data collection and processing were conducted by Lipotype GmbH (Dresden, Germany). Mass spectra for 247 lipid species from 10 lipid classes were obtained on a hybrid quadrupole/Orbitrap Q-Exactive mass spectrometer (Thermo Scientific). Raw mass spectra were processed and normalised by Lipotype GmbH using LipotypeXplorer (Lipotype) to yield final lipid quantification values and concentrations.

Results

Transcriptomic characterisation of the zebrafish *ezh2*$^{hu5670/5670}$ mutant

The *ezh2*^{hu5670} mutant contains a nonsense point mutation which introduces a premature translation stop codon in the zebrafish *ezh2* gene[17]. The translation product is therefore truncated and omits the SET domain. The SET domain is a catalytic domain which is required for the H3K27 methyltransferase activity of *ezh2*.

A homologue of *ezh2*, known as *ezh1*, is thought to have been the result of a gene duplication event. Although it plays a smaller role than *ezh2*, it has been shown that *ezh1* also has methyltransferase activity[19] and can compensate for *ezh2* somewhat[20].

To test whether *ezh1* and *ezh2* expression changed with the *ezh2*$^{hu5670/5670}$ homozygous mutation (henceforth, the ezh2 mutation) and upon injury, we performed differential expression analysis on RNA-sequencing expression counts derived from injured and noninjured *ezh2* and wildtype zebrafish (**Figure 2A**). We did not identify any significant changes to either *ezh2* or *ezh1* expression with ezh2 mutation or injury status. Taken together, this suggests that neither the *ezh2*$^{hu5670/5670}$ mutation nor injury modulate *ezh2* or *ezh1* activity on the transcriptomic level.

However, we did find widespread changes to gene expression across the transcriptome upon injury, and found that there were some changes to the injury response between ezh2 mutants and wildtypes (**Figure 2B**, **Supplementary Data 1**). In particular, in the injured wildtype, we noticed marked decreases to genes involved in redox homeostasis (*mt-nd2*, *mt-nd4*, *ndufa4*), glyoxylate utilisation (*grhprb*, *agxt2*), glutathione utilisation (*gstt1a*, *gpx4b*, *gpx3*), and lipid utilisation (*apoba*, *pla2g12b*, *apoa1b*, *elovl8b*, *alox12*) compared to noninjured wildtypes. These animals also exhibited increased gene expression in degradation factors (*psme4a*, *cth*, *cela1.4*). In the injured ezh2 mutant, these trends were generally reversed, with increased redox homeostasis (*cyp2ad2*, *cyp1a*, *cyba*), increased glyoxylate utilisation (*agxta*, *agxtb*), increased lipid utilisation (*elovl6l*, *elovl7a*, *elovl7b*, *apobb.2*). Peptidase inhibitors were also upregulated (*serpina10a*, *spint2*).

Further, in the absence of injury, the expression profile of noninjured ezh2 mutants was not substantially different to that of noninjured wildtypes.

These observations taken together suggest that there are differences in the injury responses between ezh2 mutants and wildtypes. The ezh2 mutants appeared to respond better to injury by maintaining redox homeostasis and removing ROS. On the other hand, wildtype zebrafish appeared to favour protein degradation and were less able to respond to the effects of muscle injury.

Loss of cellular functions in injury was protected against in the *ezh2*$^{hu5670/5670}$ mutant

To assess the functional changes associated with differential expression, we performed gene set enrichment analysis and predicted the affected GO terms with ezh2 mutation and injury status (**Figure 3**, **Supplementary Figures 1-3**, **Supplementary Data 2**). In the wildtype, we found that following injury, growth functions were downregulated such as DNA replication (dUDP, dTTP, and dTDP biosynthetic process) and protein translation (cytoplasmic translation). Meanwhile, several protein and molecular degradation processes were upregulated (amyloid−beta clearance by cellular catabolic process, ubiquitin−protein transferase activity, histone mRNA catabolic process) and increased oxidative stress was evident (downregulated hydrogen peroxide catabolic process, upregulated pentose−phosphate shunt, oxidative branch). Energy utilisation was altered, favouring the urea cycle (upregulated citrulline biosynthetic process) over the TCA cycle and oxidative phosphorylation (downregulated glyoxylate catabolic process and downregulated FMN biosynthetic process). These results suggest that in the wildtype, the cellular functional state shifts towards cellular arrest and possible overwhelming by oxidative stress as a result of injury.

The ezh2 mutants responded to injury without loss of DNA replication and protein translation functions, but instead with enhanced fatty acid metabolism (upregulated cholesterol esterification and upregulated cellular response to fatty acid). We also predicted changes to developmental and cell differentiation pathways (upregulated third ventricle development, upregulated regulation of eye photoreceptor cell development, upregulated positive regulation of dendrite development, upregulated negative regulation of type B pancreatic cell development), which could be co-opted to aid tissue regeneration following injury, although the exact pathways were

unclear. These results suggest that the ezh2 mutants were primed to respond to injury, and that the *ezh2*^{hu5670} mutation was in some way protective against the effects of muscle injury.

In the uninjured state, the functional state of ezh2 and wildtypes varied very minimally, which is consistent with our differential expression results (**Figure 2B**). Taken together, our results show that the *ezh2*^{hu5670} mutation appeared to be tolerated by zebrafish and did not confer inappropriate cellular functional changes.

Mouse muscle gene signatures and chromatin accessibility studies confirm changes to lipid environment and redox homeostasis

We next sought to identify the conditions which might replicate the gene expression changes that we identified in our animals upon injury. We obtained gene expression signatures from muscle studies within SysMyo and tested for overlaps with the gene expression signatures from the present study (**Supplementary Data 9**). We found one study, comparing apoE-deficient mice with wildtype mice, in which the gene expression signature had significant overlaps with upregulated and downregulated genes in injured ezh2 mutants compared to injured wildtypes. In both cases, the upregulated genes were *ttr*, *gc*, *apoc1*, *apoa2*, *aldob*, *hp*, *c9*, *cfd*, *agxt2*, *serpinc1*, *ambp*, *gck*, *f2*, and *angptl3*; the downregulated genes were *plpp3*, *slc25a4*, *vdac*, *srl*, *nnt*, *acta2*, and *aco2*.

Mouse studies involving PGC1α knockout had significantly overlapping gene signatures with our injured wildtype zebrafish compared to noninjured. Although commonly upregulated genes were unclear, commonly downregulated gene families included *nduf*, *cox*, and *serpin*. Only *nduf* genes were commonly downregulated in injured ezh2 mutants compared to noninjured. This suggests that the response to injury in zebrafish might be associated with or use similar pathways to those involved in the response to PGC1α knockout in mice. This further suggests that loss of redox function and/or protease inhibition may be common to disruption of pathways involving lipid production.

We also considered changes to chromatin accessibility by mapping ATAC-seq peaks to the zebrafish genome. The ATAC-seq peak data were obtained from a study by

den Broeder et al[5], investigating the effects of an EZH2 inhibitor (EZH2i). We found many lipid-related genes and arachidonic acid-related genes were accessible in EZH2i but not control (**Supplementary Data 10**). These included *alox5b.3*, *elovl1a*, *elovl5*, and *elovl7b*. This suggests a causal link between loss of ezh2 and activation of lipid-related pathways. Den Broeder et al also confirmed changes to chromatin accessibility using differential expression using qRT-PCR. Their observations of increased *igfbp1a* and *lpl* were consistent with our differential expression results.

Metabolic shift towards pro-regeneration pathways was exacerbated in the *ezh2*$^{hu5670/5670}$ mutant

Having explored molecular changes in ezh2 mutant zebrafish on the transcriptomic and functional scale, and having gained supportive evidence for metabolic shifts in GABA, cholesterol, and fatty acids, we next hypothesised that the metabolic landscape might have changed in the injured ezh2 mutant compared to wildtype. To predict metabolic changes, we used genome-scale models (GEMs) of zebrafish metabolism. We selected the GEMs *ZebraGEM2.1*[21], an adaptation of *ZebraGEM2.0*[22] which has been applied to study neurodegeneration. In addition to this, we corroborated our results with *Zebrafish1*[11], a comprehensive model incorporating orthology-based and species-specific metabolic reactions. Both models allow for the *in-silico* prediction of reactions and metabolites based on expression or differential expression.

From analysis of GEMs, we found changes in multiple tRNAs, redox metabolites (ubiquinone, ubiquinol, NAD, NADH, cytochrome c), and reactive oxygen and nitrogen species (O_2, NH_3) in response to injury (**Figure 4**, **Supplementary Figure 3**, **Supplementary Data 3 and 4**). These findings were supported by both GEMs. While we observed all the above-described metabolic changes in the wildtype animals, the ezh2 mutants did not have the changes to tRNA. Injured animals also had increased very long chain fatty acid metabolism, increased production of hydroxyeicosatetraenoic acids (HETEs), cholesterols, and androgens. Upregulation of these metabolites was evidenced by *Zebrafish1* and did not differ with ezh2 mutation, indicating that they may be involved in the core muscle regeneration response. However, we also found ezh2-specific downregulation of malate,

spermine, spermidine, and fumarate, in injured ezh2 mutants compared to noninjured ezh2 mutants. On the other hand, the wildtype was specifically downregulated for glutathione and succinate.

Taken together, our metabolic analysis supports our earlier results from differential expression and gene set enrichment. The zebrafish response to muscle injury appears to be largely the same regardless of ezh2 mutation, involving production of very long chain fatty acids and HETEs. For this reason, we hypothesised that the likely pathways for injury response is via arachidonic acid. Arachidonic acid recruits HETEs to stimulate inflammation, immune cell recruitment, and removal of ROS. Our observation that the injured wildtypes arrested some core cellular functions, namely tRNA production, whereas the injured ezh2 mutants did not, indicates an enhanced regeneration phenotype in the ezh2 mutants. The ezh2 mutants' retention of glutathione and succinate indicates that this enhanced regeneration could be via improved ROS removal capabilities in the ezh2 mutant. Our differential expression results indicate that very long chain fatty acid synthesis is activated via *elovl6l*, *elovl7a*, and *elovl7b*, which were all significantly upregulated in the ezh2 mutant upon injury (**Supplementary data 1**). Interestingly, the wildtype responded by downregulating *elovl8b*. This is, however, consistent with the observation that the injured wildtype appeared to shut down central cellular processes.

Recruitment of immune cells accompanies upregulated arachidonic acid signalling

Having established a role for arachidonic acid and inflammation during muscle tissue regeneration, we next sought to test the hypothesis that injury altered the immune landscape differentially in the ezh2 mutant and wildtype. From our differential expression analysis, we inspected the change in expression for genes encoding arachidonic acid pathway enzymes, caspases, chemokines, interleukins, TNF-related proteins, and immune cell identity proteins (**Figure 5**). We found general activation of arachidonic acid pathway genes. Immune activation in response to muscle injury differed between the genotypes, with ezh2 mutants more stably activating immune cells than wildtypes. For the ezh2 mutants in particular, there was increased caspase, chemokine, interleukin, and TNF-related gene

activation, indicating a robust inflammatory response to injury which was absent in the wildtype. The ezh2 mutants also more robustly activated B cells, classical monocytes, intermediate monocytes, and mature dendritic cells, indicating that the immune cell population may be more enriched for these cell types in the ezh2 mutant than in the wildtype. Wildtype immune response to injury was unclear, with expression signatures for immune cell types being both up- and downregulated.

As expected, there were no changes or only incremental changes to the inflammation and immune response landscape between the noninjured ezh2 mutant and wildtype, indicating further that the mutation is tolerated.

Taken together, our results indicate a mechanism of muscle tissue regeneration after injury that relies on arachidonic acid signalling and interleukin activation. The immune landscape changed by recruitment of B cells, classical monocytes, M2 macrophages, and T cells. The ezh2 mutant enhanced the response and also recruited chemokines and TNF-related enzymes, as well as intermediate monocytes and mature dendritic cells. The ezh2 mutant therefore appeared to signal for robust, acute inflammation with the aim of enhanced healing, whereas the wildtype appeared to signal for chronic inflammation and eventual cell loss in the absence of any robust immune signalling.

Network exploration indicates regulatory link between *ezh2* and arachidonic acid

Our results suggest that arachidonic acid signalling is central to the injury response. Pathway activity appears to be modulated by ezh2 mutation, leading to differential changes to molecular, inflammation, and immune landscapes. We therefore hypothesised that the *ezh2* gene is a master regulator of arachidonic acid metabolism genes either directly or via a downstream interactor.

We explored a previously described zebrafish interactome to find possible regulatory links between *ezh2* and arachidonic acid pathway genes (**Figure 6**). We found that 23 arachidonic acid pathway genes were accessible within three regulatory steps from *ezh2*, and these included very long chain fatty acid elongases (*elovl2* and *elovl6*), solute channels (*slc25a20*), fatty acid binding proteins (*fabp3*,

fabp6), enzymes acting on acetyl-CoA (*acaa1*, *acadl*, *acadm*, *acads*, *acadvl*, *acot7*, *acox1*), and enzymes acting on arachidonic acid (*alox12*) (**Figure 6A**).

Observing that up to three regulatory steps were needed to connect *ezh2* with the 23 arachidonic acid pathway genes, we next hypothesised that a first- or second-order interactor with *ezh2* (that is, a gene that is regulated by *ezh2* directly or via one intermediary interactor) could regulate the arachidonic acid pathway genes specifically. We explored the connectivity of all first- and second-order interactors of *ezh2* with the arachidonic acid genes and found that up to 16 and 5 arachidonic acid pathway genes could be regulated via modulation of a single first- or second-order interactor of *ezh2*, respectively (**Figure 6B-C**).

Of the first-order interactors, most were housekeeping genes (*cse1l*, *tubb5*, *ran*) or were DNA-related interactors of the PRC2 complex within which the product of *ezh2* is a constituent (*fen1*, *mcm3*, *cdc45l*, *ccna2*, *mcm2*, *mcm5*). However, we also identified *zgc:76940* and *zgc:85729*, which encode serine/threonine protein phosphatases and each regulate 13 arachidonic acid pathway genes. In differential expression, *zgc:76940* was not significantly differentially expressed, whereas *zgc:85729* was downregulated in injured ezh2 mutants compared to noninjured ezh2 mutants with marginal evidence ($p = 0.059$). This suggests that interventional knockdown of *zgc:85729* might be able to phenocopy the specific effects of ezh2 mutation for arachidonic acid signalling.

Of the second-order interactors, we found the dehydrogenases *hsd17b10* and *zgc:77820*, each of which regulate four arachidonic acid pathway genes. *hsd17b10* was not significantly differentially expressed in our study, but *zgc:77820* was significantly upregulated in injured ezh2 mutant animals compared to injured wildtype animals, and significantly downregulated in injured wildtype animals compared to noninjured wildtype animals. This is consistent with our observation that core cellular processes, such as oxidative phosphorylation, were arrested in the injured wildtype.

Lipidomic analysis reveals fatty acid elongation

To validate our findings regarding very long chain fatty acid synthesis to arachidonic acid, we conducted lipidomic analysis on *ezh2* mutant, heterozygote,

and wildtype zebrafish embryos. We found that overall lipid compositions were very similar between genotypes, indicating that the mutation is tolerated (**Supplementary Figure 10**). However, we also observed dose-dependent trends towards longer chain length within cholesterol esters, diacylglycerols, phosphatidylcholines, phosphatidylglycerols, sphingomyelins, and triacylglycerols (**Supplementary Figure 11**, **Supplementary Data 11**). In these lipid classes, shorter chains tended to present in higher concentrations in wildtype embryos, whereas longer chains tended to be favoured in *ezh2* mutants, with the heterozygote being the intermediate state. This suggests that the *ezh2* phenotype might arise from priming the lipid landscape before injury occurs.

Discussion

In this work, we analysed transcriptomic changes in wildtype zebrafish and *ezh2*$^{hu5670/5670}$ mutant zebrafish to compare molecular alterations with and without injury. We observed transcriptional changes to *ezh2* as a result of the mutation as well as compensatory transcriptional changes to *ezh1* (**Figure 2A**). We found that *ezh1* compensated for the *ezh2* stop-gain mutation only in injury, and that in noninjury, the mutation was tolerated. Looking at whole-transcriptome changes (**Figure 2B**), we found that the *ezh2* mutation did not significantly alter the expression of many genes in the noninjury state, but it did differentially affect the molecular landscape upon injury. In our functional enrichment analysis (**Figure 3**), we found evidence in the injured zebrafish for loss of core cellular functions involving synthesis of DNA, RNA, and protein, whereas protein degradation, small molecule degradation, and oxidative stress were all increased. *ezh2* mutation appeared to rescue the deleterious degradation and oxidative stress effects and also exhibited enhanced fatty acid metabolism, activation of cell differentiation pathways, and increased chemokine and inflammatory response. Using GEMs, our reporter metabolite analysis (**Figure 4**) also predicted loss of tRNAs, decreased glutathione, as well as enhanced inflammatory signalling, increased arachidonic acid signalling, and increased fatty acid synthesis with injury, with the ezh2 mutant being able to rescue the decreased glutathione effect for better ROS removal. Our characterisation of the inflammatory and immune landscape (**Figure 5**) further supports a better inflammatory response in the ezh2 mutant in injury, leaving the noninjured mutant wildtype-like. Repeating all of these analyses with an independent dataset described by San et al yielded largely consistent results (**Supplementary Figures 4-8**). Finally, our network analysis (**Figure 6**) established a probable causal link between the *ezh2* gene and the arachidonic acid pathway.

In this work, we demonstrated changes to the lipid metabolism landscape upon muscle injury. Functional enrichment showed that ezh2 mutants had increased cholesterol metabolism (**Figure 3**), and metabolic analysis showed that cholesterols, very long chain fatty acids, and HETEs were upregulated upon injury (**Figure 4**). Differential expression showed that the very long chain fatty acid synthases *elovl6l*, *elovl7a*, and *elovl7b* were all upregulated in the injured ezh2 mutant, and this is

consistent with a previous chromatin accessibility study involving EZH2i[5]. Lipidomic analysis also confirmed a trend towards longer fatty acid chains with *ezh2* mutation.

Our analyses are consistent with previous studies reporting accumulation of lipids in the case of EZH2 downregulation or inhibition[5,23–25]. This suggests a role for activation of lipid metabolism in tissue injury. Further, although our wildtype zebrafish tended to downregulate *ezh2* upon injury, they did not exhibit the enhanced healing phenotype that we observed with the ezh2 mutants. This would suggest a protective role of *ezh2* interventional downregulation, and this is consistent with numerous reports highlighting the significance of *ezh2* in diverse diseases[26–30].

Given the enhancement of very long chain fatty acid synthesis pathways in injured ezh2 mutants, our work also showed that this activation likely stimulated arachidonic acid synthesis. Functional enrichment showed activation of molecular functions acting on arachidonic acid (**Figure 3**, **Supplementary Figure 2**). Arachidonate lipoxygenase function was activated in the injured wildtype, injured ezh2 mutant, and noninjured ezh2 mutant, that is, all conditions exhibiting downregulated *ezh2*. Differential expression showed that the arachidonate 12-lipoxygenase *alox12* was significantly upregulated in the injured wildtype compared to noninjured and tended to upregulation in the noninjured animals compared to wildtypes across both genotypes (**Supplementary Data 1**). Reporter metabolite analysis also indicated upregulation of arachidonic acid in the cytosol, lysosome, and endoplasmic reticulum of injured animals compared to wildtypes, again across both genotypes (**Supplementary Figure 4**). The lack of enhanced arachidonic acid synthesis in the ezh2 mutant upon injury suggests that the mutation may afford a higher baseline level of arachidonic acid synthesis compared to noninjured wildtypes. This is consistent with our network analysis, which demonstrated a regulatory link between *ezh2* and 22 arachidonic acid metabolism genes, one of which being *alox12* (**Figure 6**). To our knowledge, we are the first to report this link between *ezh2* and arachidonic acid.

We have shown changes to arachidonic acid synthesis as a likely endpoint to enhanced very long chain fatty acid synthesis in zebrafish with muscle injury. Our results show evidence for increased signalling leading to engagement of the inflammation and immune systems. Arachidonic acid is known for its inflammation signalling activity via leukotrienes and HETEs, and we have shown that all are engaged following injury regardless of genotype (**Supplementary Figure 4**). With respect to metabolic inflammation markers, the injury response was the same in both ezh2 mutants and wildtypes. However, ezh2 mutants did not downregulate central cellular functions such as protein synthesis and peptidase inhibitors (**Figure 3**). This is consistent with our observation that ezh2 mutants appeared to switch on different inflammatory and immune cell programs to the wildtype upon injury (**Figure 5**).

Our analyses also showed a priming behaviour of the *ezh2* mutant, that is, *ezh2* mutation was not enough to trigger the muscle regeneration response. An injury event was still required in order to change the transcriptomic (**Figure 2B**), cell functional (**Figure 3**), metabolic (**Figure 4**), and inflammatory and immune cell landscapes (**Figure 5**). Molecularly, the profiles of noninjured *ezh2* mutant zebrafish was not substantially different to noninjured wildtype zebrafish, with the exception of the lipid landscape trending towards longer chain species (**Supplementary Figure 11**), indicating that the mutation was well tolerated by the animals. Rather, the loss of functional EZH2 appeared to enable a heightened muscle regeneration response after injury. Our network exploration analysis showed feasible regulatory pathways between *ezh2* and arachidonic acid pathway genes (**Figure 6**), suggesting that *ezh2* mutants did not need to wait for *ezh2* gene regulation to be abrogated after injury. However, more work needs to be done in order to ascertain this link.

This work suggests a role for arachidonic acid in the zebrafish muscle injury response, mediated by EZH2 and giving rise to exacerbated response in the *ezh2* mutant, consistent with previous reports of lipid accumulation upon EZH2 inhibition. Although EZH2 is not a realistic therapeutic target due to its involvement in transcriptional regulation of multiple genes, it may be possible to phenocopy the EZH2-loss–mediated muscle injury response via the arachidonic acid synthesis and/or signalling pathway. If validated, then arachidonic acid, its derivatives, or its signalling targets, may constitute a therapeutic target for the accelerated healing of

muscle tissue. This could have translational benefit, for example, for people suffering from muscle degenerative disease or by shortening recovery times following a sports injury.

In conclusion, our work demonstrates that EZH2 regulates the muscle injury response in zebrafish by epigenetically silencing response genes in heterochromatin in the absence of injury. EZH2 is lost upon injury, leading to expression of response genes and activation of very long chain fatty acid synthesis and arachidonic acid metabolism. Arachidonic acid signalling induces inflammatory pathways and immune cell recruitment, leading to ROS clearance and tissue regeneration. These responses are heightened in a stop-gain mutant *ezh2* zebrafish mutant, and this is likely due to the pre-priming of the lipid landscape prior to injury.

Acknowledgements

The authors acknowledge use of the research computing facility at King's College London, *Rosalind*. (https://rosalind.kcl.ac.uk).

Ethics statement

All animal procedures were performed in accordance with local and national guidelines.

Conflicts of interest

The authors declare no competing financial interests.

Author contributions

R.K. conceived the study, performed the qRT-PCR experiments, and supervised the study. S.L. performed all bioinformatic analyses and interpreted the results. S.L. wrote the manuscript with input from A.M. and the rest of the authors.

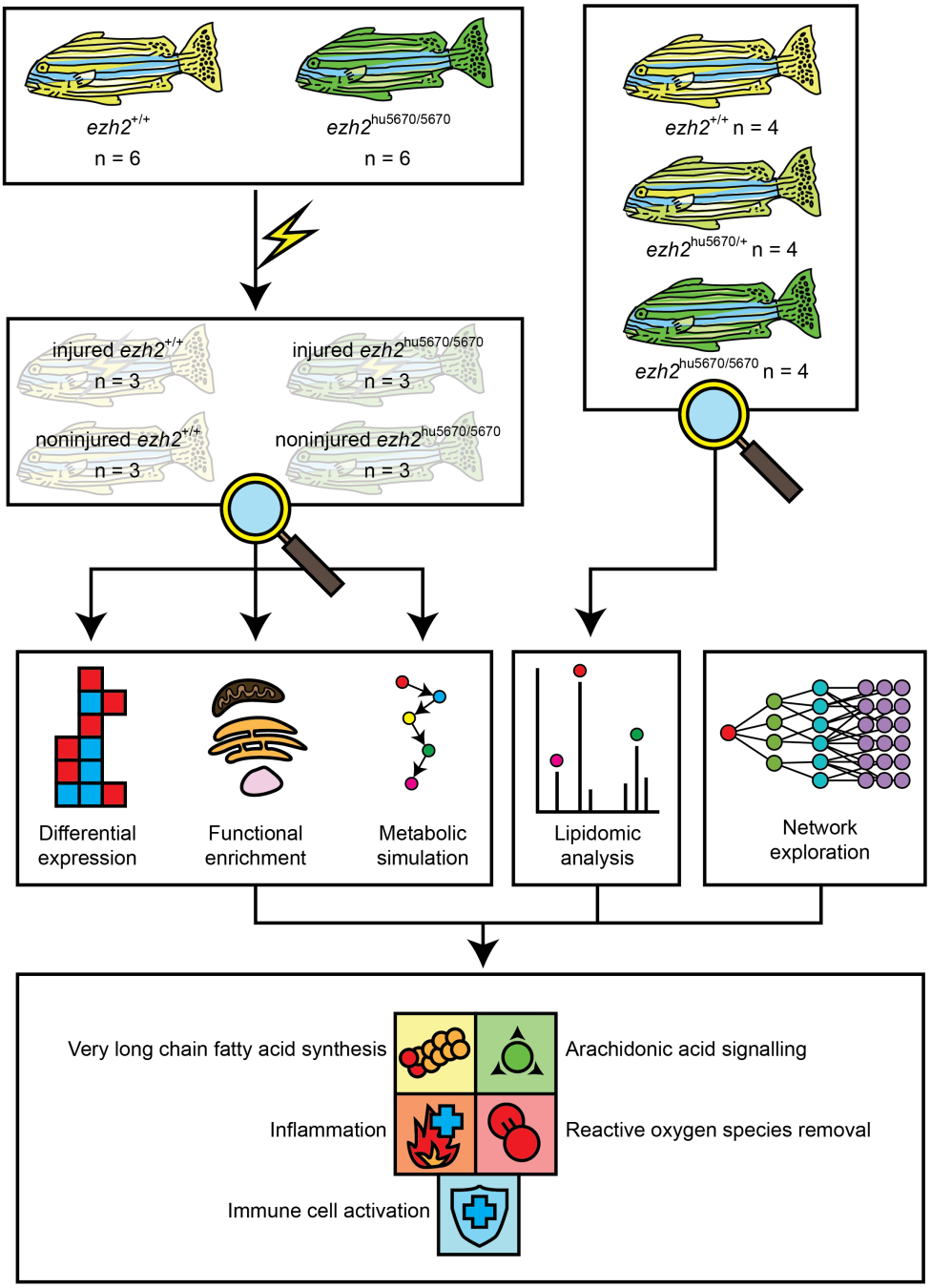


Figure 1. Study summary. ezh2 and wildtype zebrafish were analysed by differential expression, functional enrichment, and metabolic simulation. An interactome for *Danio rerio* was obtained and explored. From these results, five hallmarks of muscle regeneration were identified: very long chain fatty acid

synthesis, arachidonic acid signalling, acute inflammation, reactive oxygen species removal, immune system activation.

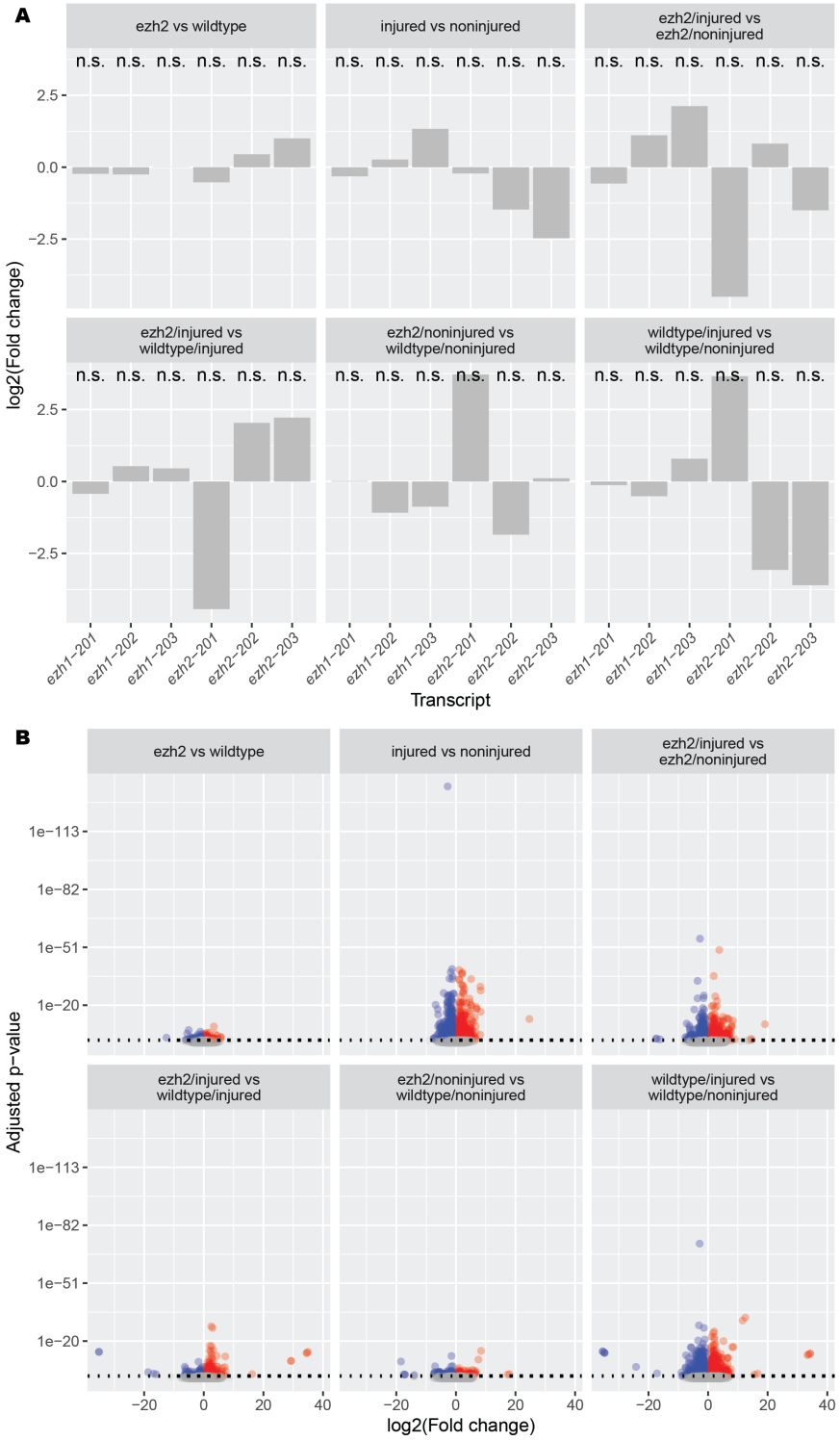


Figure 2. Differential expression. Changes to gene expression were deduced from RNA-seq. **A)** Changes to *ezh1* and *ezh2* transcript expression with ezh2 mutation

and injury. Grey bars indicate no statistical significance. n.s., not significant. **B)** Distribution of differentially expressed genes with ezh2 mutation and injury. Red points indicate significantly upregulated genes. Blue points indicate significantly downregulated genes. Grey points indicate no statistical significance. Horizontal dotted line indicates statistical significance threshold of $p < 0.05$. All p-values are Benjamini-Hochberg adjusted p-values. See also Supplementary Data 1.

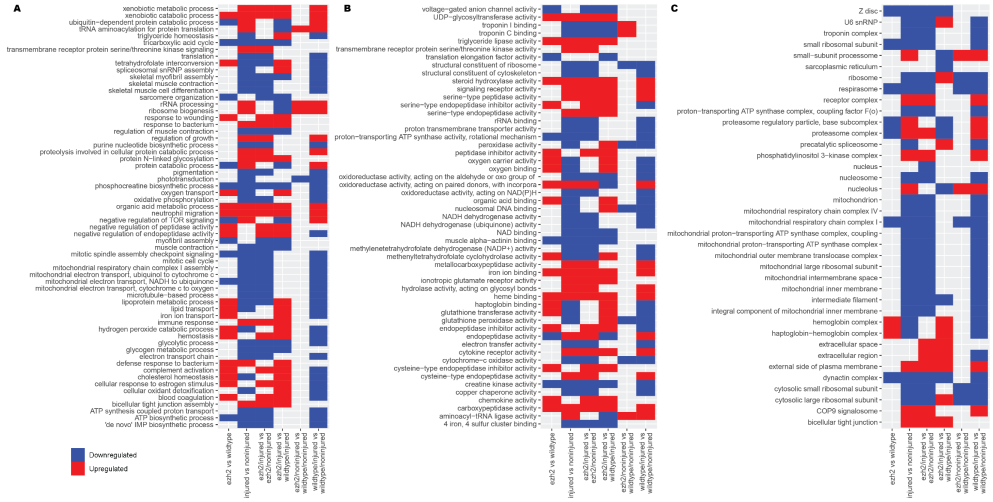


Figure 3. Gene set enrichment. Enrichment of **A)** Gene Ontology (GO) biological processes, **B)** GO molecular functions, and **C)** cellular compartments was deduced from differential expression. Scoring and removal of redundant GO terms was performed with Revigo. GO terms with a Revigo score of less than −2 are shown. Colour indicates direction of enrichment. See also Supplementary Figures 1-3 and Supplementary Data 2.

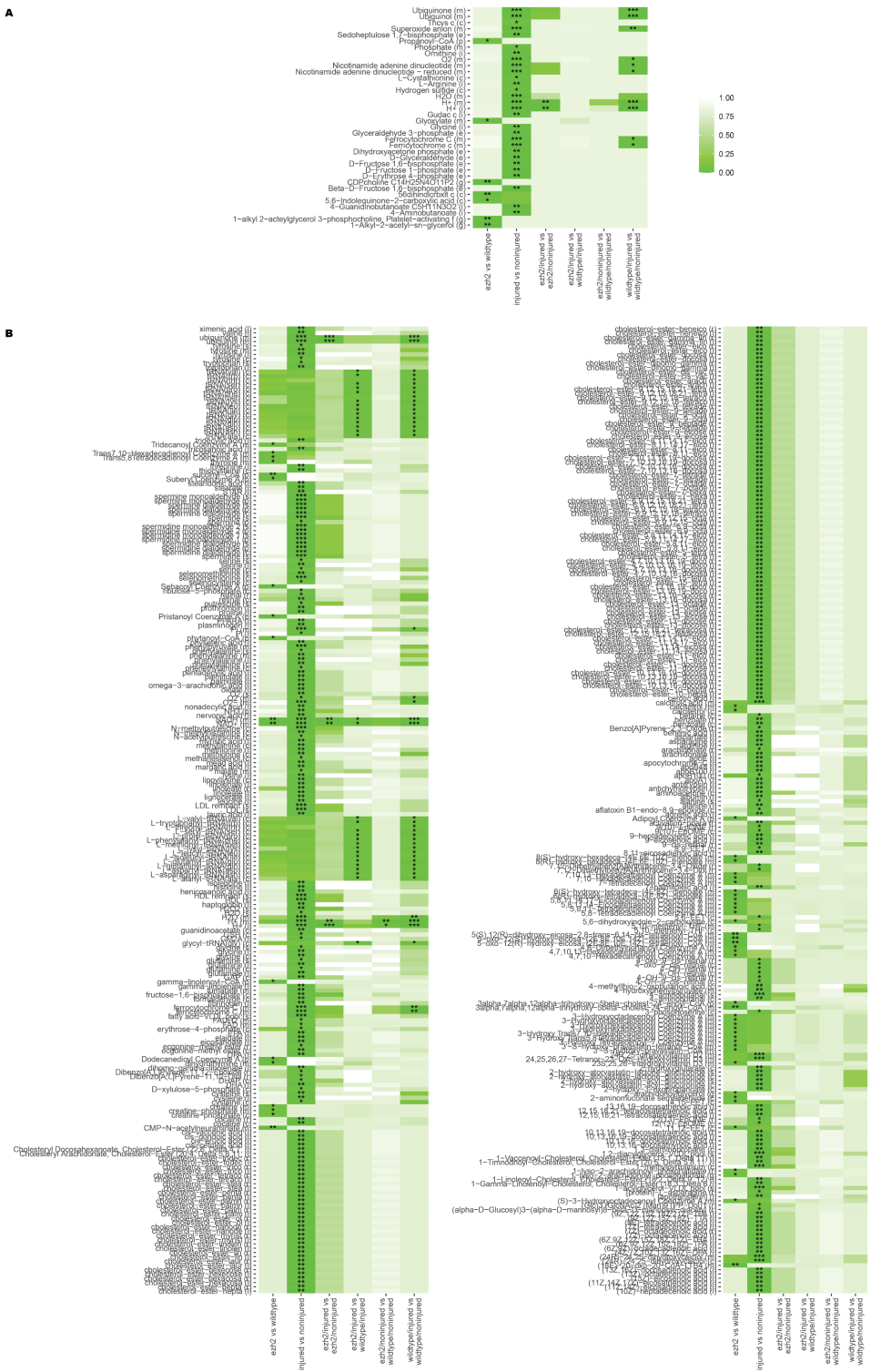


Figure 4. Reporter metabolites. Metabolites predicted to be changed were deduced from differential expression and simulation of **A)** *ZebraGEM2.1* or **B)** *Zebrafish1*

genome-scale metabolic models. Colour gradient indicates adjusted p-values. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. All p-values are Benjamini-Hochberg adjusted p-values. Letters in parentheses indicate subcellular compartment: c, cytosol; e, extracellular space; i, mitochondrial intermembrane space; m, mitochondrial matrix; mim, mitochondrial inner membrane; p, peroxisome; s, extracellular space. See also Supplementary Figure 4 and Supplementary Data 3 and 4.

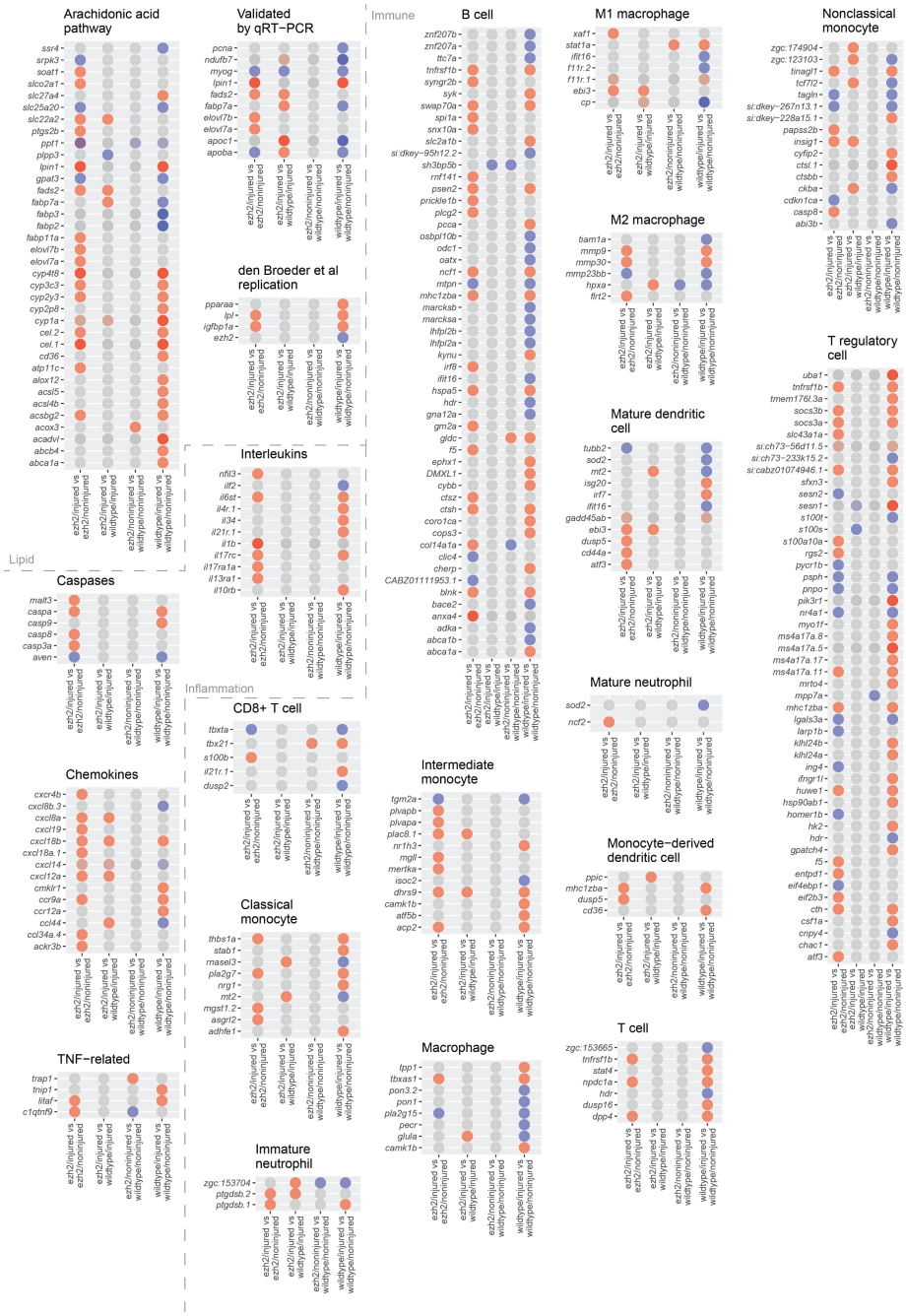


Figure 5. Differential expression of genes of interest. **A)** Arachidonic acid pathway and inflammation-related genes. **B)** Immune cell identity genes. Significantly differentially expressed genes are shown (Benjamini-Hochberg adjusted p-value < 0.05). Colour hue indicates direction of differential expression: red, upregulated; blue, downregulated; grey, not statistically significant. Intermediate hues indicate that some transcripts were upregulated and some transcripts were downregulated.

Colour intensity indicates number of significantly differentially expressed transcripts. See also Supplementary Data 1.

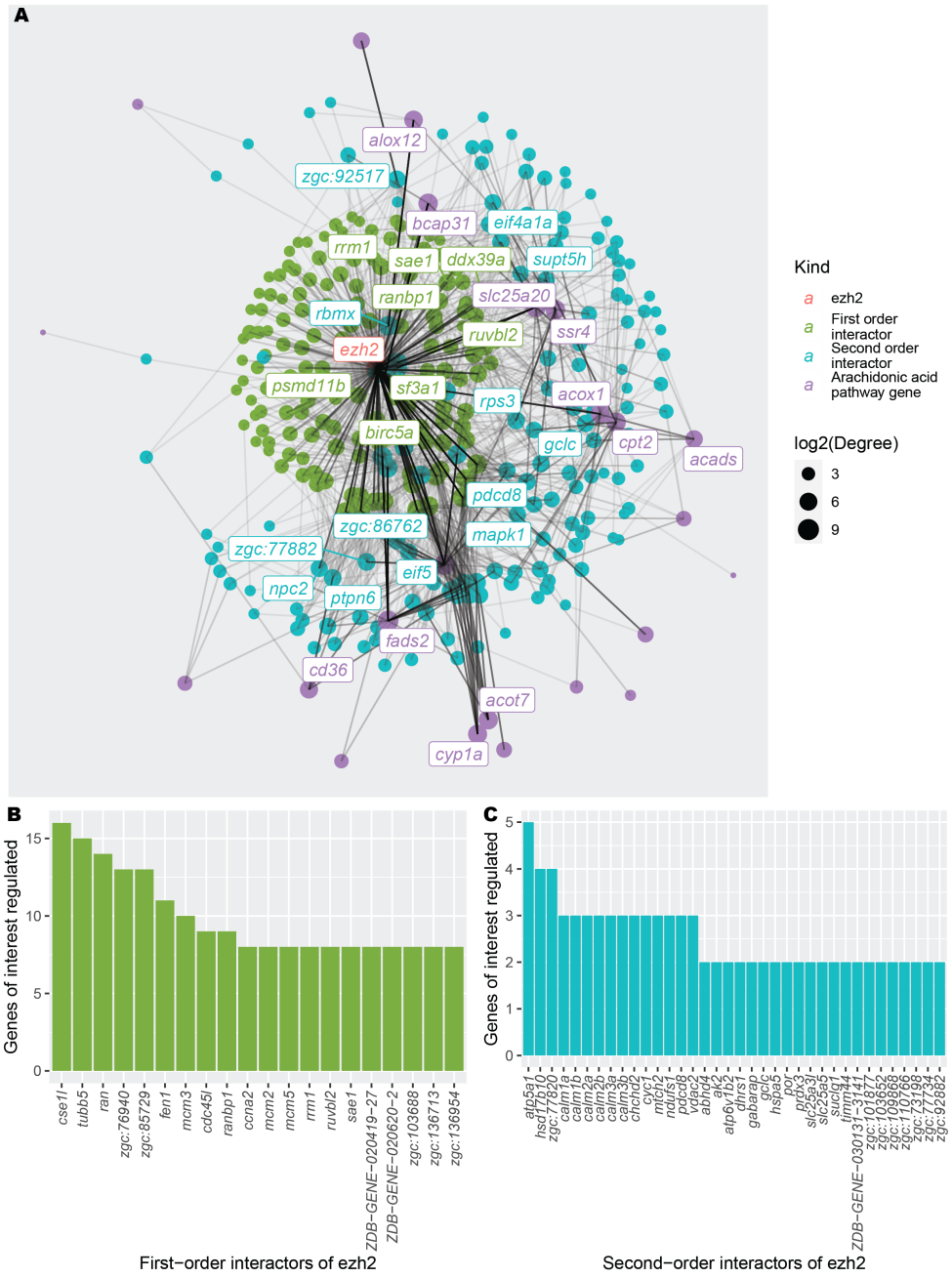


Figure 6. *ezh2* local interactome. **A)** Network of gene regulatory interactions between *ezh2* and arachidonic acid pathway genes. Connectivity of **B)** first-order and **C)** second-order interactors of *ezh2* to arachidonic acid pathway genes are shown.

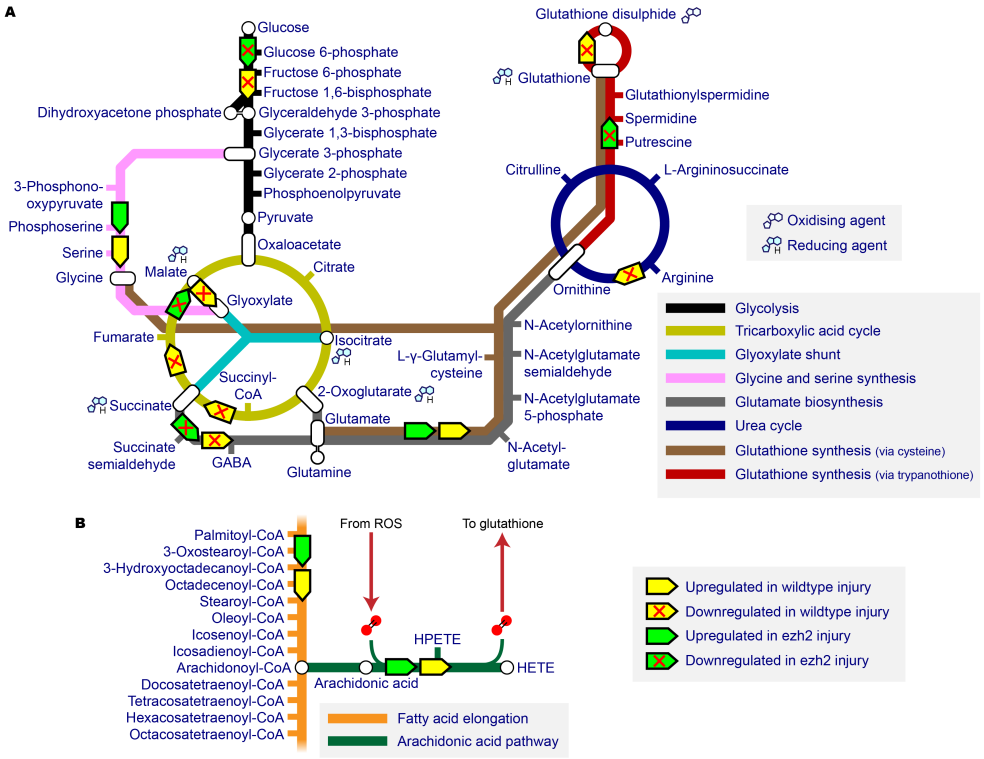
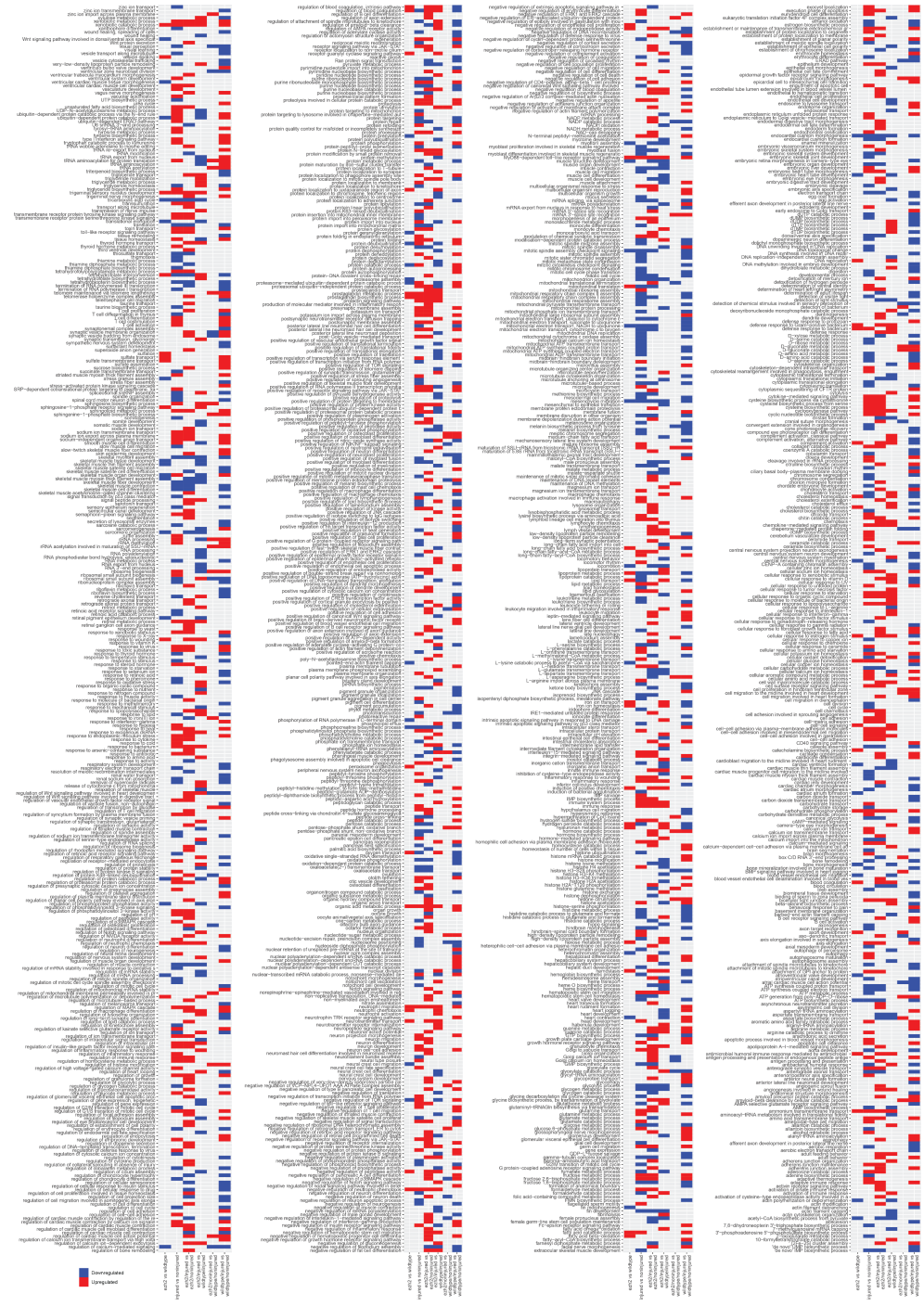
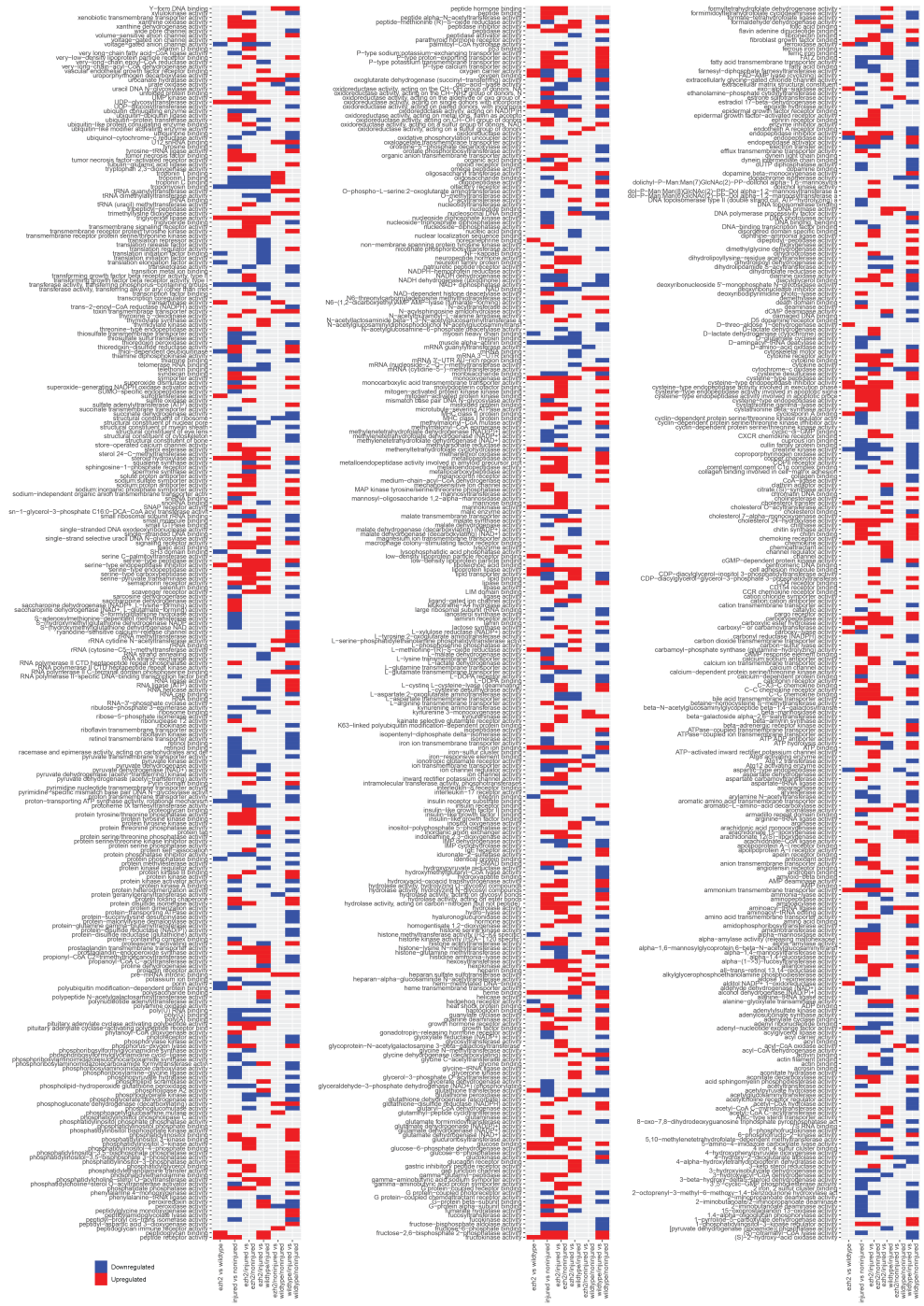


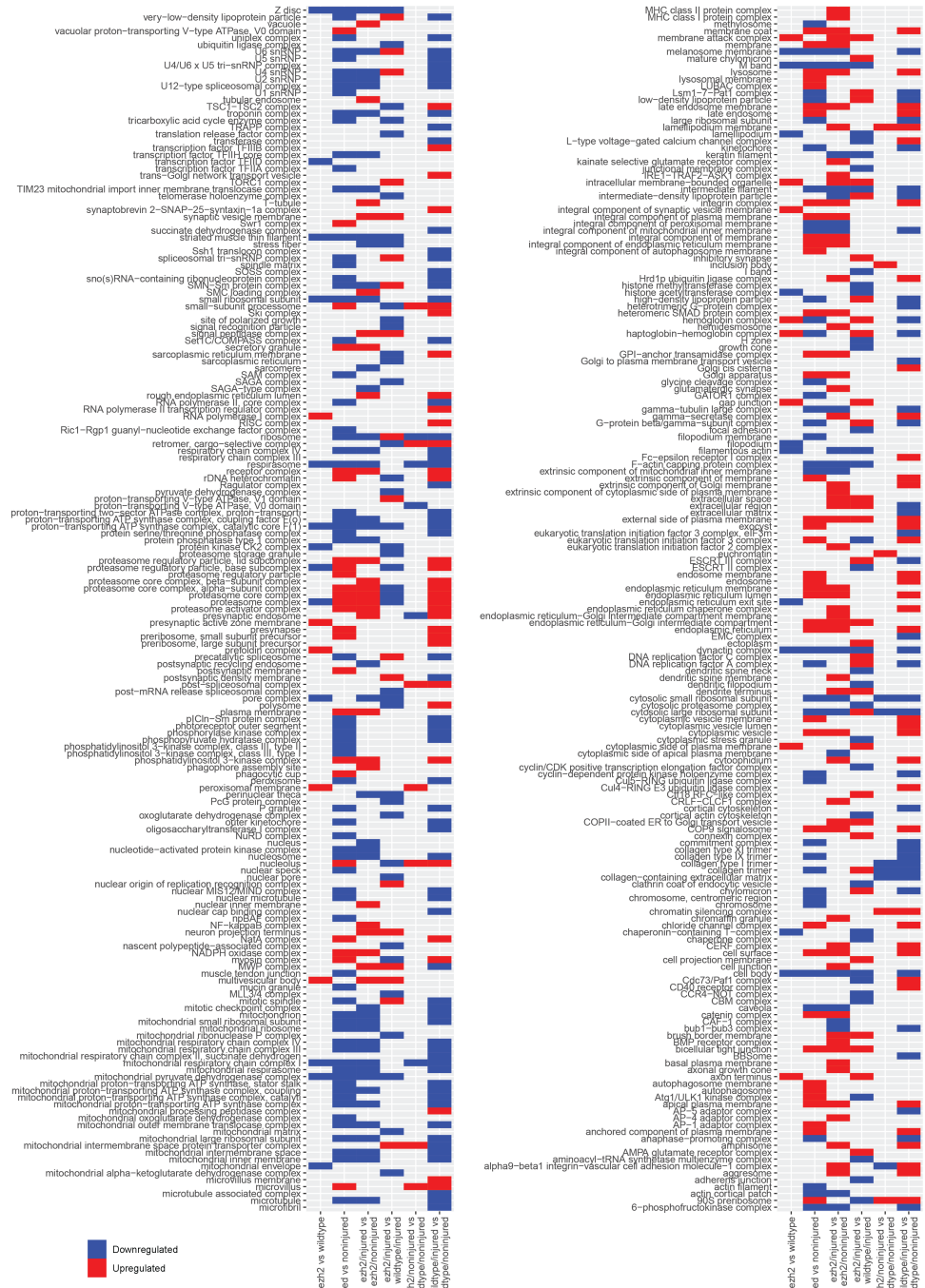
Figure 7. Summary of carbon metabolism changes with ezh2 mutation and injury. **A)** Energy metabolism. **B)** Fatty acid metabolism. Not all metabolites shown. Icons indicate changes with injury and ezh2 mutation.



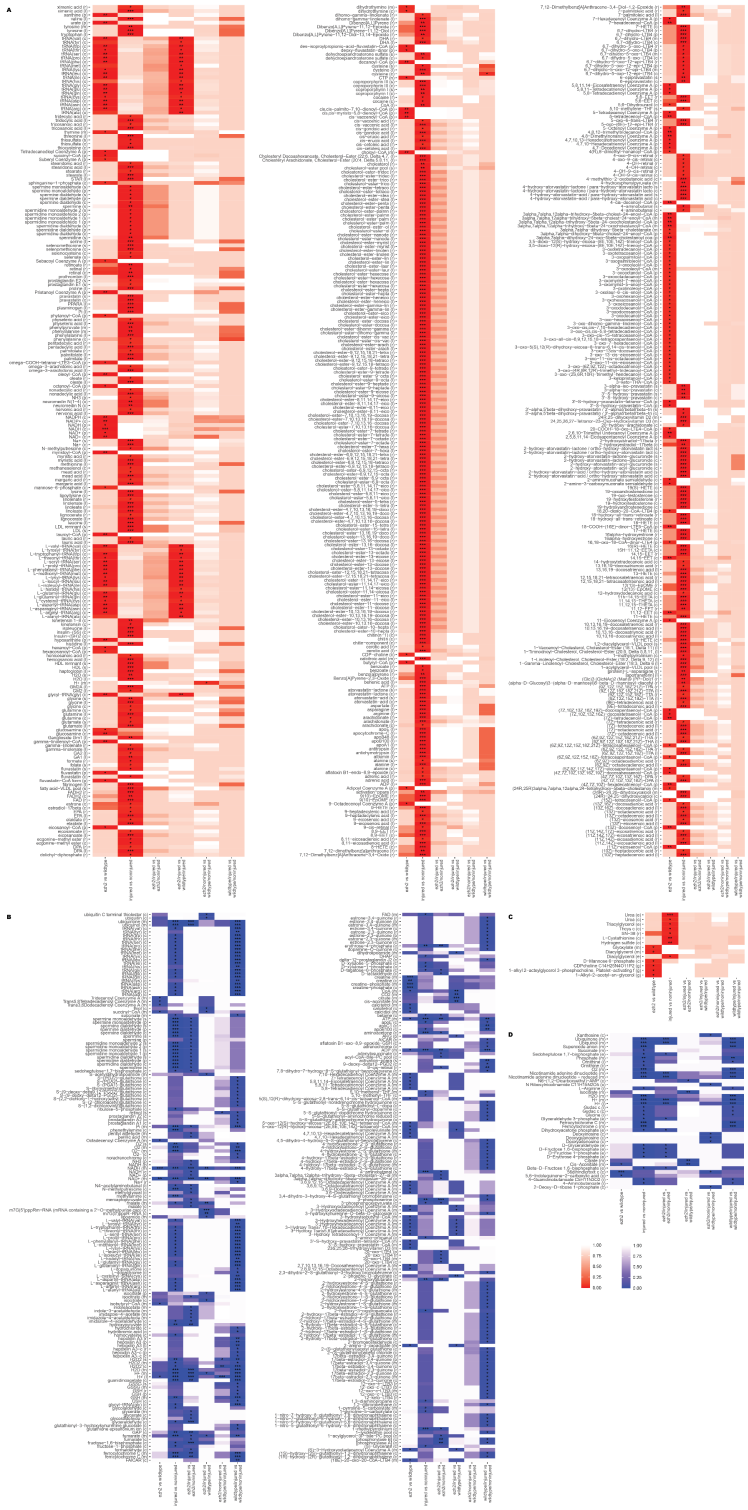
Supplementary Figure 1. Gene set enrichment. GO terms for biological processes are shown. Colour keys are as in Figure 3. See also Supplementary Data 2.



Supplementary Figure 2. Gene set enrichment. GO terms for molecular functions are shown. Colour keys are as in Figure 3. See also Supplementary Data 2.

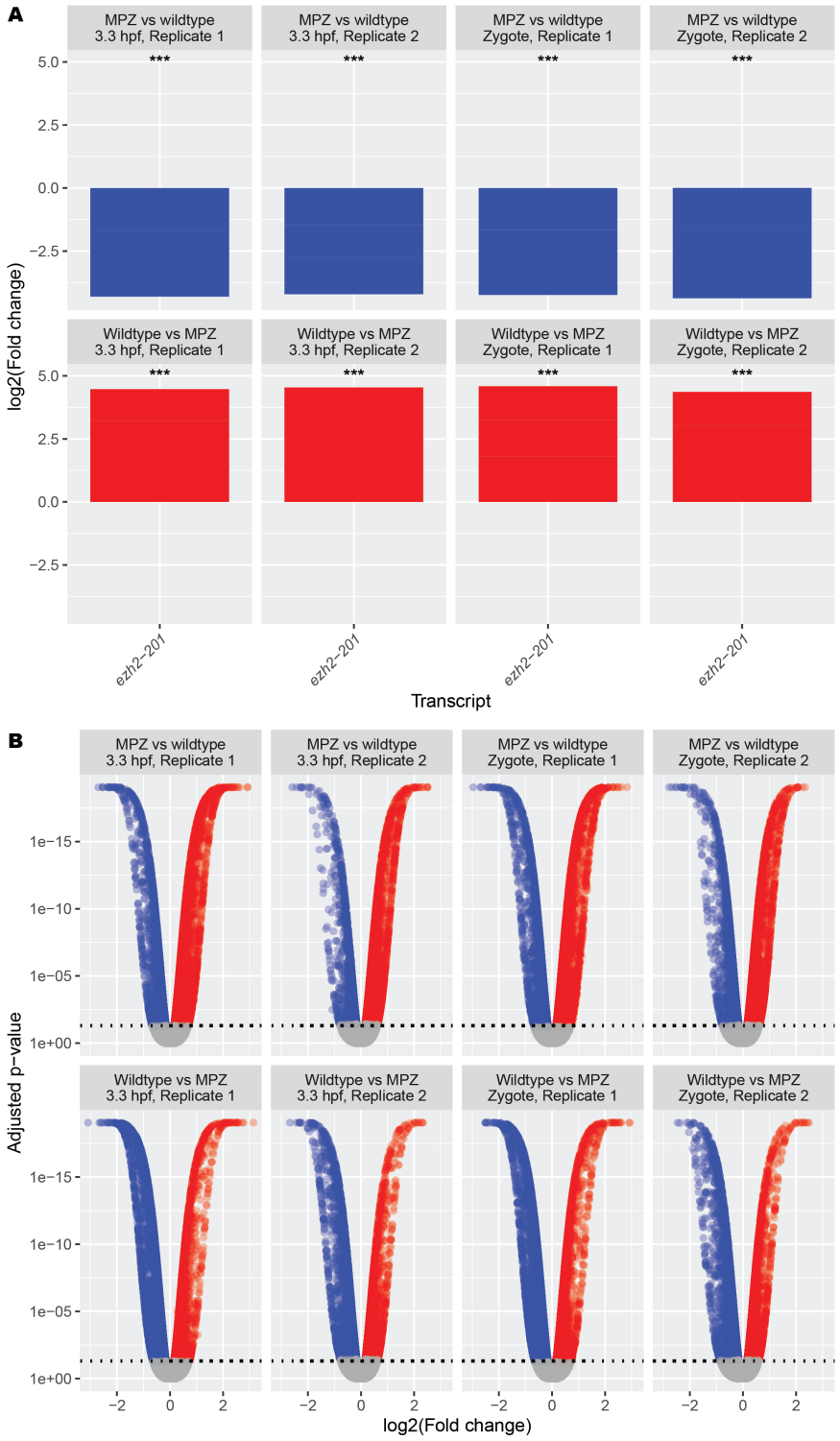


Supplementary Figure 3. Gene set enrichment. GO terms for cellular compartments are shown. Colour keys are as in Figure 3. See also Supplementary Data 2.



Supplementary Figure 4. Reporter metabolites. **A)** Upregulated reporter metabolites in *Zebrafish1*. **B)** Downregulated reporter metabolites in *Zebrafish1*. **C)**

Upregulated reporter metabolites in *ZebraGEM2.1*. **D)** Downregulated reporter metabolites in *ZebraGEM2.1*. Keys are as in Figure 4. See also Supplementary Data 3 and 4.

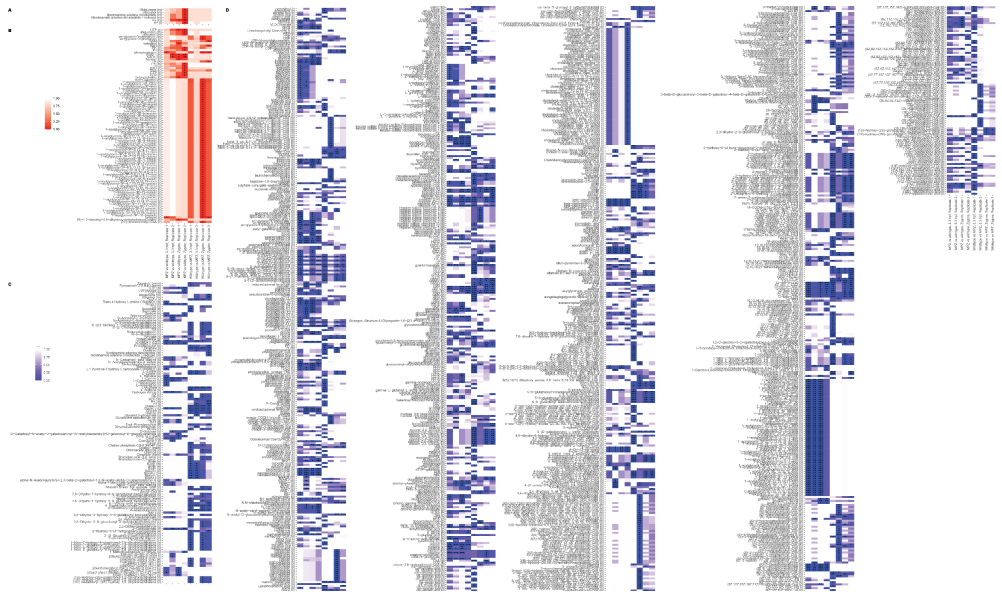


Supplementary Figure 5. Differential expression in the San et al dataset. **A)** Changes to *ezh1* and *ezh2* transcript expression with ezh2 mutation (MPZ). Red bars

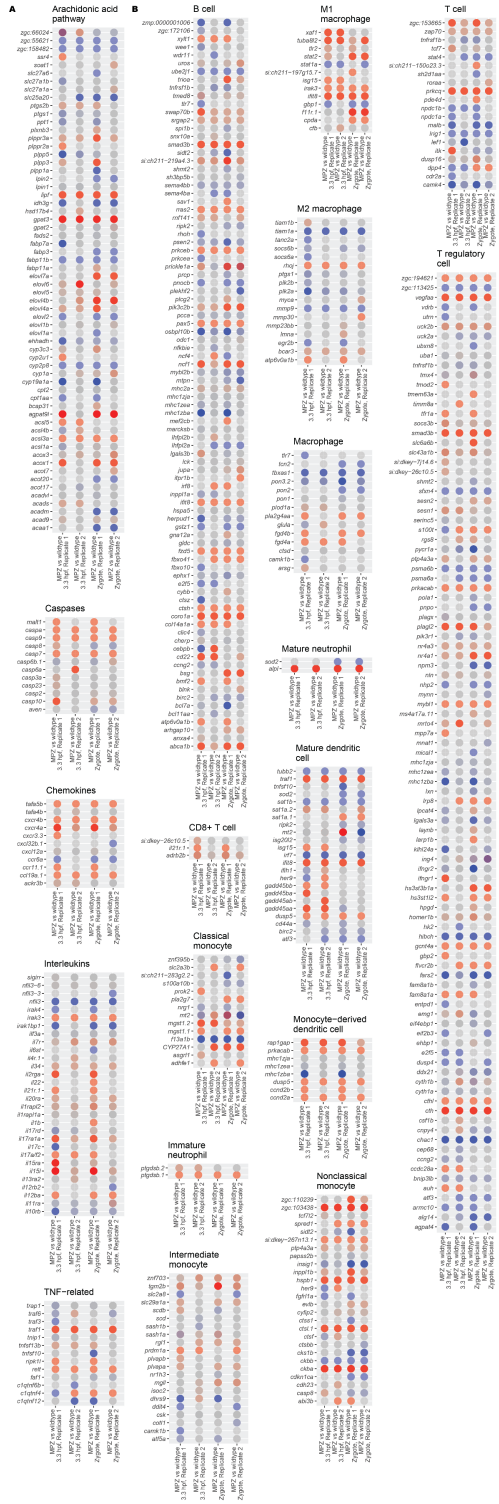
indicate significant upregulation. Blue bars indicate significant downregulation. ***, $p < 0.001$. All p-values are Benjamini-Hochberg adjusted p-values. **B)** Distribution of differentially expressed genes with MPZ. Red points indicate significantly upregulated genes. Blue points indicate significantly downregulated genes. Grey points indicate no statistical significance. Horizontal dotted line indicates statistical significance threshold of $p < 0.05$. All p-values are Benjamini-Hochberg adjusted p-values. See also Supplementary Data 5.

Supplementary Figure 6. Gene set enrichment in the San et al dataset. GO terms for biological processes are shown. Colour keys are as in Figure 3. See also Supplementary Data 6.

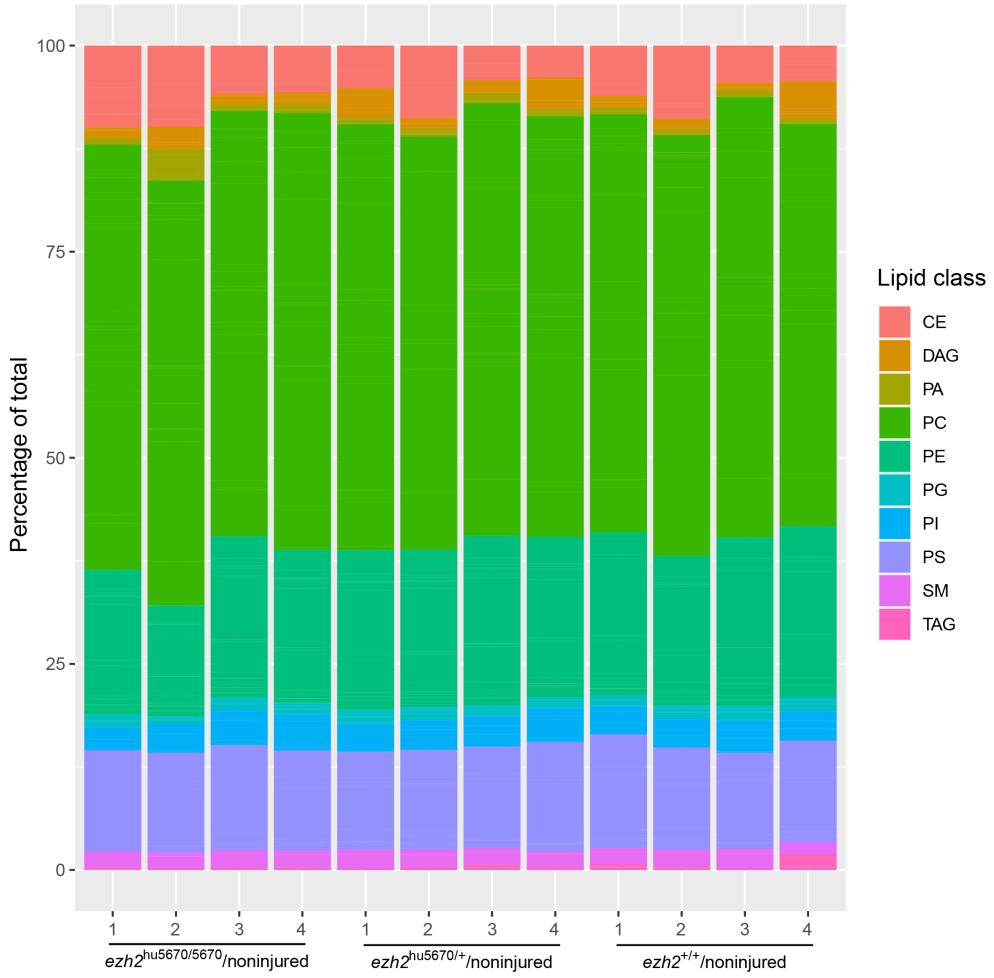
Supplementary Figure 7. Gene set enrichment in the San et al dataset. GO terms for molecular functions are shown. Colour keys are as in Figure 3. See also Supplementary Data 6.



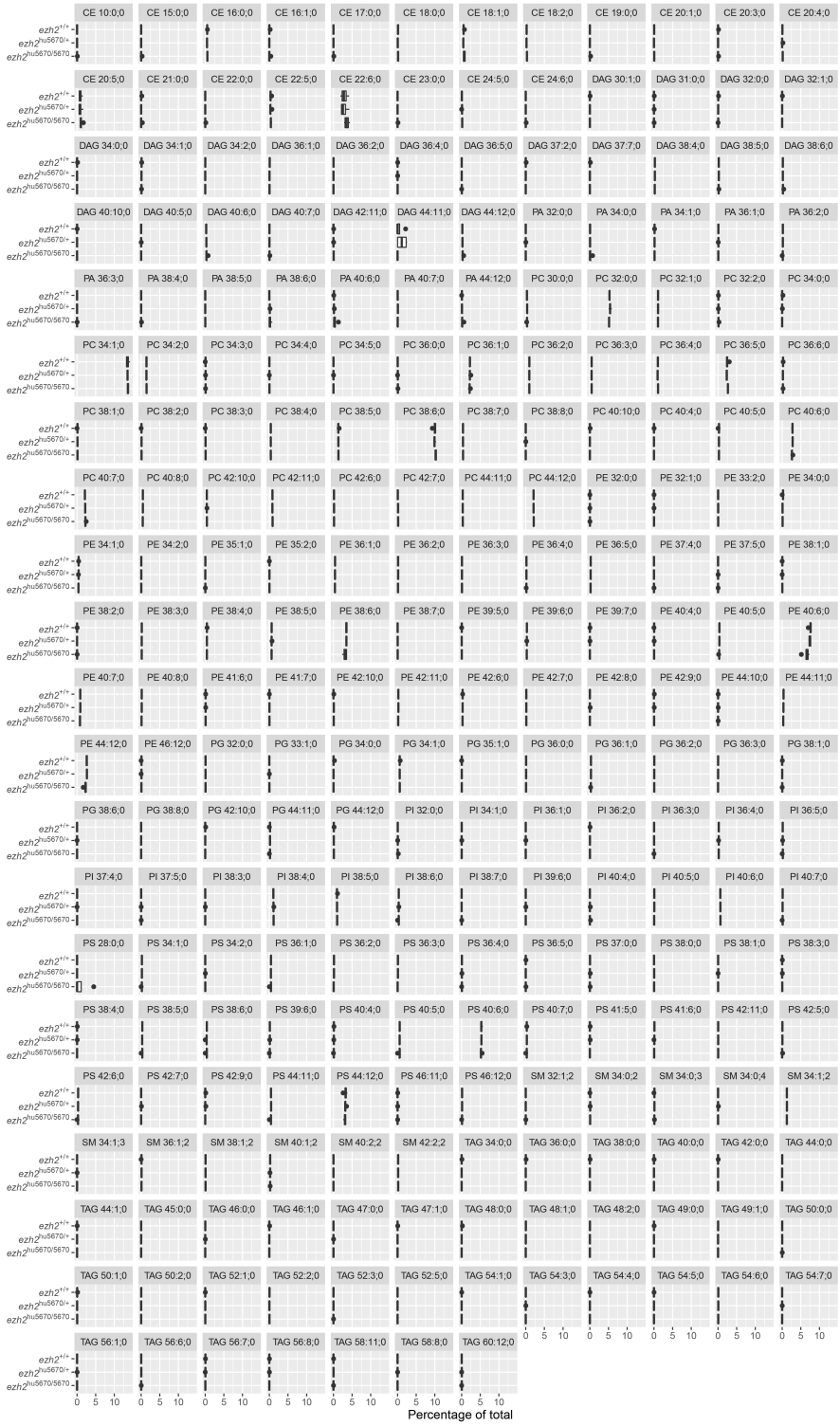
Supplementary Figure 8. Reporter metabolites in the San et al dataset. Upregulated reporter metabolites from **A)** *ZebraGEM2.1* and **B)** *Zebrafish1*. Downregulated reporter metabolites from **C)** *ZebraGEM2.1* **D)** *Zebrafish1*. Keys are as in Figure 4. See also Supplementary Data 7 and 8.


A
Arachidonic acid pathway
B
B cell
M1 macrophage
T cell
M2 macrophage
T regulatory cell
Caspases
Macrophage
Chemokines
Mature neutrophil
Interleukins
Mature dendritic cell
CD8+ T cell
Classical monocyte
Monocyte−derived dendritic cell
TNF−related
Immature neutrophil
Nonclassical monocyte
Intermediate monocyte
MZ7 vs wildtype 3.3 hpf, Replicate 1
MZ7 vs wildtype 3.3 hpf, Replicate 2
MZ7 vs wildtype Zygote, Replicate 1
MZ7 vs wildtype Zygote, Replicate 2

Supplementary Figure 9. Differential expression of genes of interest in the San et al dataset. **A)** Arachidonic acid pathway and inflammation-related genes. **B)** Immune cell identity genes. Keys are as in Figure 5. See also Supplementary Data 5.



Supplementary Figure 10. Lipid composition of *ezh2* mutant, heterozygote, and wildtype zebrafish embryos. CE, cholesterol ester; DAG, diacylglycerol; PA, phosphatidate; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; PI, phosphatidylinositol; PS, phosphatidylserine; SM, sphingomyelin; TAG, triacylglycerol.



Supplementary Figure 11. Detailed lipid composition of zebrafish embryos. Lipid class abbreviations are the same as in Supplementary Figure 10. Lipid species are

indicated using numerical representation: (carbon chain length):(carbon-carbon double bonds);(hydroxyl groups). See also Supplementary Data 11.

References

1. Laugesen, A., Højfeldt, J. W. & Helin, K. Role of the Polycomb Repressive Complex 2 (PRC2) in Transcriptional Regulation and Cancer. *Cold Spring Harbor Perspectives in Medicine* **6**, a026575 (2016).

2. Yan, N. *et al.* Postnatal onset of retinal degeneration by loss of embryonic Ezh2 repression of Six1. *Sci Rep* **6**, (2016).

3. Jiang, C. *et al.* Inhibition of EZH2 ameliorates cartilage endplate degeneration and attenuates the progression of intervertebral disc degeneration via demethylation of Sox-9. *EBioMedicine* **48**, 619–629 (2019).

4. San, B. *et al.* Normal formation of a vertebrate body plan and loss of tissue maintenance in the absence of ezh2. *Scientific Reports 2016 6:1* **6**, 1–16 (2016).

5. den Broeder, M. J. *et al.* Inhibition of methyltransferase activity of enhancer of zeste 2 leads to enhanced lipid accumulation and altered chromatin status in zebrafish. *Epigenetics Chromatin* **13**, (2020).

6. Meng, T. G. *et al.* PRC2 and EHMT1 regulate H3K27me2 and H3K27me3 establishment across the zygote genome. *Nature Communications* **11**, (2020).

7. Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Research* **48**, D682–D688 (2020).

8. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**, 525–527 (2016).

9. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).

10. Väremo, L., Nielsen, J. & Nookaew, I. Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Research* **41**, 4378–4391 (2013).

11. Wang, H. *et al.* Genome-scale metabolic network reconstruction of model animals as a platform for translational research. *Proc Natl Acad Sci U S A* **118**, 2102344118 (2021).

12. Lyons, Y. A., Wu, S. Y., Overwijk, W. W., Baggerly, K. A. & Sood, A. K. Immune cell profiling in cancer: molecular approaches to cell-specific identification. *npj Precision Oncology 2017 1:1* **1**, 1–8 (2017).

13. Patil, K. R. & Nielsen, J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci U S A* **102**, 2685–2689 (2005).

14. Wang, H. *et al.* RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor. *PLoS Computational Biology* **14**, e1006541 (2018).

15. Lam, S. *et al.* Systems Analysis Reveals Ageing-Related Perturbations in Retinoids and Sex Hormones in Alzheimer's and Parkinson's Diseases. *Biomedicines 2021, Vol. 9, Page 1310* **9**, 1310 (2021).

16. Alexeyenko, A. *et al.* Dynamic Zebrafish Interactome Reveals Transcriptional Mechanisms of Dioxin Toxicity. *PLoS ONE* **5**, (2010).

17. San, B. *et al.* Normal formation of a vertebrate body plan and loss of tissue maintenance in the absence of ezh2. *Scientific Reports* **6**, (2016).

18. Malatras, A., Duguez, S. & Duddy, W. Muscle Gene Sets: A versatile methodological aid to functional genomics in the neuromuscular field. *Skeletal Muscle* **9**, 1–12 (2019).

19. Shen, X. *et al.* EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol Cell* **32**, 491 (2008).

20. Völkel, P. *et al.* Ezh1 arises from Ezh2 gene duplication but its function is not required for zebrafish development. *Scientific Reports* **9**, (2019).

21. Lam, S. *et al.* Systems analysis reveals ageing-related perturbations in retinoids and sex hormones in Alzheimer's and Parkinson's diseases. *bioRxiv* 2021.06.10.447367 (2021) doi:10.1101/2021.06.10.447367.

22. van Steijn, L., Verbeek, F. J., Spaink, H. P. & Merks, R. M. H. Predicting Metabolism from Gene Expression in an Improved Whole-Genome Metabolic Network Model of Danio rerio. *Zebrafish* **16**, 348–362 (2019).

23. Vella, S. *et al.* EZH2 down-regulation exacerbates lipid accumulation and inflammation in in vitro and in vivo NAFLD. *Int J Mol Sci* **14**, 24154–24168 (2013).

24. Dong, Z. *et al.* LncRNA PU.1 AS regulates arsenic-induced lipid metabolism through EZH2/Sirt6/SREBP-1c pathway. *J Environ Sci (China)* **85**, 138–146 (2019).

25. Yiew, N. K. H. *et al.* Enhancer of zeste homolog 2 (EZH2) regulates adipocyte lipid metabolism independent of adipogenic differentiation: Role of apolipoprotein E. *J Biol Chem* **294**, 8577–8591 (2019).

26. Yomtoubian, S. *et al.* Inhibition of EZH2 Catalytic Activity Selectively Targets a Metastatic Subpopulation in Triple-Negative Breast Cancer. *Cell Rep* **30**, 755-770.e6 (2020).

27. Li, K. *et al.* EZH2 inhibition promotes ANGPTL4/CREB1 to suppress the progression of ulcerative colitis. *Life Sci* **250**, (2020).

28. Allas, L. *et al.* EZH2 inhibition reduces cartilage loss and functional impairment related to osteoarthritis. *Sci Rep* **10**, (2020).

29. McCabe, M. T. *et al.* EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* **492**, 108–112 (2012).

30. Kim, H. J., Cantor, H. & Cosmopoulos, K. Overcoming Immune Checkpoint Blockade Resistance via EZH2 Inhibition. *Trends Immunol* **41**, 948–963 (2020).

5. Conclusion

In the present investigation, I examined the metabolic and biomolecular characteristics of neurodegeneration (**Papers I – IV**), lipodystrophy (**Paper V**), and muscle injury response (**Paper VI**). In these following manuscripts, I demonstrate the potential of metabolic studies for the identification of diagnostic and predictive biomarkers in fields where classical research has so far not focused on metabolism, namely AD and PD research (**Papers I – III**), or where it does not so obviously play a direct role, namely muscle injury (**Paper VI**). To arrive at my conclusions, I employed statistical techniques alongside machine learning classification techniques (**Papers I** and **IV**) to identify trends and classes within clinical and omics data. I demonstrate that these techniques can be used to accelerate discovery in diseases less well studied.

Through these analyses, I discovered a role for retinoic acid metabolism deficiency in AD and PD which was also seen in an ageing zebrafish model (**Paper I**). Given additional time and resources, I would have sought or generated data on the effect of retinoic acid on amyloid, tau, or α-synuclein clearance; or on cognitive scores in humans, to be determined using clinical cognitive assessments such as the Mini Mental State Examination, Alzheimer's Disease Assessment Scale, or Montreal Cognitive Assessment. It would be interesting to reconcile the effect of retinoic acid on PD given that dementia is a lesser effect of PD than the movement disorders that characterise the disease. To take this hypothesis further, the effect of retinoic acid on motor and spatial cognitive functions could even be assessed in mice via methods such as the rotarod and Morris water maze.

Papers II and **III** demonstrated the potential therapeutic value of the combined metabolic activators (CMAs) L-carnitine, N-acetylcysteine, nicotinamide riboside, and L-serine. This formulation was initially designed for the treatment of non-alcoholic fatty liver disease (NAFLD) and each constituent was selected by metabolic modelling of NAFLD. However, owing to energy metabolism utilisation similarities between NAFLD, AD, and PD[130], and as informed via animal models, we decided to reposition CMAs to AD and PD after re-rationalising the formulation by metabolic modelling on ROSMAP participants. Although our results did not reach the threshold to support a clinical advantage of CMA, we found that CMAs led to cognitive improvements with higher

statistical confidence than in the placebo group. Given additional time and resources, a more closely controlled study based on targeted therapy of other metabolites, such as retinoic acid as discussed in **Paper I** of this thesis, would be tested for any beneficial effects on cognitive and motor function.

Paper IV proved the principle of data-driven identification of diagnostic biomarkers in neurological diseases (NLDs). This was achieved through Monte Carlo–based feature selection to maximise the predictive power of a machine learning model and validating the coefficients of each selected feature with trends in the literature. Selection of blood and urine biochemistry features – and especially liver enzymes – led to an attractive model of potential early detection during routine clinical tests to identify suspected NLDs without the need of a specialist neurology unit. This paper does not suggest that liver markers cause or are naively correlated with NLD; rather their alteration is a part of a larger metabolic perturbation involving multiple factors associated with NLD and that machine learning is required to detect such patterns. Given additional resources, an independent, ideally longitudinal, cohort would be sought to validate and refine the model, and the metabolic perturbations involved in the progression of different NLDs would be deduced from the model coefficients.

Papers V and **VI** were ambitious projects that inspected metabolic alterations in conditions that do not immediately apply metabolism in their response. In **Paper V**, I obtained evidence for convergent metabolic perturbations in NASH and lipodystrophy, two outwardly distinct diseases, in which a common liver X receptor utilisation characteristic was observed. The implication of this result is that treatments for lipodystrophies could potentially be beneficial to certain subsets of NASH patients displaying lipodystrophy-like metabolic perturbations. If proven true, then this paper challenges the wider definition of diseases based on symptoms and outward presentation and provides a strong basis for drug repositioning. In **Paper VI**, I found a role for arachidonic acid in the muscle injury response in Ezh2 mutant zebrafish in which I observed upregulated lipid elongation enzymes and enrichment of longer chain fatty acids. In my model, I suggested that this led to the accumulation of the signalling metabolite arachidonic acid, which promotes reactive oxygen species (ROS) removal via hydroxyeicosatetraenoic acid (HETE) and glutathione, consistent with metabolic

modelling conducted in parallel. Given additional resources, this study could be extended to inspect the effect of arachidonic acid and *elovl* family gene knockout on muscle repair. If proven true and translatable, then the clinical implications of this discovery could potentially improve muscle healing and shorten injury times after exercise and sports injuries.

As a result of the present investigation, I can conclude that a metabolic perturbation basis for human degenerative diseases is a feasible and attractive philosophy that is able to unlock both basic biological insights and realistic therapeutic strategies. This approach is realized by biotechnological and computational advancements in systems biology which enable the simulation of cellular metabolism *in silico*. These efforts, while not perfect and require some assumptions in the data, are a good first step to generating tractable leads and targeted hypotheses. Importantly, systems biology has shown to be able to prioritise and direct resources and attention to the strongest metabolic candidates on the basis of the wealth of data that already exist after more than a decade of high-throughput data collection. The interface of systems biology and disease biology, known as systems medicine, therefore holds promise for the acceleration of drug development and repositioning.

In the current investigation, I have detailed the hypothesis-free identification of candidate metabolic biomarkers in **Papers I**, **IV**, **V**, and **VI**. I have shown the usefulness of zebrafish experimental models of metabolism in **Papers I** and **VI**. I have demonstrated the translational potential for therapies based on metabolic activation in **Papers II** and **III**. Taken together, the current investigation represents a full trip around the systems medicine cycle, from hypothesis-free data exploration for the identification of biomarkers to the experimental validation in *in vivo* models of metabolism and clinical trials in humans.

My closing remark is that systems medicine will hold promise not only for therapeutic medicine, but also for preventive medicine. This vision requires longitudinal health, nutritional, and lifestyle data collection, ideally from wearables technology, and will be able to model health trajectories, provide personalised risk scores for multiple diseases, and give dietary and lifestyle suggestions based on an individual's own data. The common

denominator between this vision for the “holy grail” of data-driven preventive medicine and the current investigation is the underlying understanding that metabolic perturbations, which are caused by genetics, age, lifestyle, and diet, hold the key to the difference between health and disease.

Ethnography

Reflecting on my personal experience as a research student at the College, I will say that my foremost impression is a sense of belonging in my department. Distinct from my experience as an undergraduate student, where I had relatively fleeting, ephemeral links with disparate departments, I really benefitted from the support structures afforded by being grounded in a single department. The Faculty is sufficiently large enough so that the types of support, skillsets, and research are wide and varied, but also sufficiently small enough to feel like a close-knit community. Activities such as Faculty mindfulness sessions set up by Katya Griffin and my collaboration with Dr Robert Knight, which led to **Papers I** and **VI** in this current investigation, are examples of the Faculty's size working to benefit and unite.

Being a doctoral candidate during the COVID-19 pandemic and the introduction of working from home guidance across the UK proved to me just how much of my practical work as a bioinformatician and systems biologist can be done on the computer. From being pleasantly surprised by the success of the Cold Spring Harbor Neurodegenerative Diseases: Biology & Therapeutics symposium being held completely online, to my own personal productivity by cutting out travel and commute and improving my work-life balance, the experience has firmly proven to me that remote and hybrid working arrangements are feasible and desirable. With the appreciation that not all science can be done from home, and so some collaborators will be found at the bench, I also welcome a phased return to the workplace and in-person symposia for the interactions which cannot be replicated on a computer screen. Still, I see myself discussing with future managers working from home at least some of the time for the foreseeable future, as a change of scenery is a break from monotony, and I think that this in no small part was another contributor to my boosted productivity during lockdown.

Finally, I give credit for the additional and extracurricular opportunities available to postgraduate research students at the College, including but not limited to module teaching, skills training, and the Associateship of King's College. These were highly enriching experiences, and yet another extremely welcome break from potential

monotony. Module teaching as a Graduate Teaching Assistant firmly placed me on the "other side" of university education as I had seen it during my undergraduate years, and it also went a long way to break down my own imposter syndrome, built up over numerous manuscript rejections and temperate-to-scathing remarks from peer reviewers, which I have also learned to be the norm, and ultimately for my own benefit as a researcher. Skills training, particularly the Adobe Illustrator and Photoshop courses led by Alex Greenhough, will stay with me, and these experiences too were excellent enrichments to College life. And I only have praise for the Associateship of King's College, the team behind which seem to have a knack for inviting thought leaders from across the College and beyond.

Taken together, I see my experience as a doctoral candidate at the College as a representative prelude to academic life. The College has provided the support and enrichment I need to work at my best.

Acknowledgements

A great number of people have contributed towards the success of my investigations, and I have had the privilege to enjoy the support from colleagues within the College and further afield. Our collaborations are immortalised in the by-line of manuscripts which made it to publication, yet these are outnumbered by valuable connections with individuals un-named.

Still, I feel duty-bound to personally acknowledge a few individuals. My first thanks go to Prof Adil Mardinoglu, my supervisor throughout my doctoral studies, for his vision of the future of systems medicine and his seemingly unfailing ability to very rapidly interpret conclusions from results. Adil has also been extremely supportive of me for changes to the research plan as the science allows it, and has given me a great deal of research freedom and autonomy while keeping the ultimate vision firmly in mind.

Under Adil's supervision, I had the benefit of collaborating with lab members from different affiliations. I recognise colleagues from the Science for Life Laboratory at KTH – Royal Institute of Technology in Stockholm, Sweden. I thank my colleagues and friends Cheng Zhang, Muhammad Arif, Ozlem Altay, Rui Benfeitas, Woonghee Kim, and Xiangyu Li, all of whom have been extremely knowledgeable and supportive when it comes to anything from technical to practical matters. Given the distance, most of my contact with the KTH team has been via virtual channels, and thus I have always welcomed any opportunity to meet up for conference or training.

My journey as a doctoral student was not taken alone, and it is important to acknowledge those on their own journeys. Abdulahad Bayraktar and Ali Kaynar are doctoral students at the College, whose work has been an inspiration for my own and whose company and empathy have given me the will to carry on in tough times.

Final thanks are reserved for my wife Xuemeng, who has never let me lose sight of the bigger picture.

References

1. Kellum JA. The modern concept of homeostasis. *Minerva Anestesiol*. 2002;68(1-2):3-11.

2. Clancy J, McVicar A. Hypoxia as a failure of respiratory homeostasis. *Br J Theatre Nurs*. 1997;6(10):16-20.

3. Crimeen-Irwin B, Scalzo K, Gloster S, Mottram PL, Plebanski M. Failure of immune homeostasis -- the consequences of under and over reactivity. *Curr Drug Targets Immune Endocr Metabol Disord*. 2005;5(4):413-422. doi:10.2174/156800805774912980

4. Dursun I, Sahin M. Difficulties in maintaining potassium homeostasis in patients with heart failure. *Clin Cardiol*. 2006;29(9):388-392. doi:10.1002/clc.4960290904

5. Koudinov AR, Koudinova NV. Cholesterol homeostasis failure as a unifying cause of synaptic degeneration. *J Neurol Sci*. 2005;229-230:233-240. doi:10.1016/j.jns.2004.11.036

6. Wang YP, Li JT, Qu J, Yin M, Lei QY. Metabolite sensing and signaling in cancer. *J Biol Chem*. 2020;295(33):11938-11946. doi:10.1074/jbc.REV119.007624

7. Ghyselinck NB, Duester G. Retinoic acid signaling pathways. *Development*. 2019;146(13):dev167502. doi:10.1242/dev.167502

8. Wang T, Fu X, Chen Q, et al. Arachidonic Acid Metabolism and Kidney Inflammation. *Int J Mol Sci*. 2019;20(15):3683. doi:10.3390/ijms20153683

9. Nieto Gutierrez A, McDonald PH. GPCRs: Emerging anti-cancer drug targets. *Cell Signal*. 2018;41:65-74. doi:10.1016/j.cellsig.2017.09.005

10. Ray PD, Huang BW, Tsuji Y. Reactive oxygen species (ROS) homeostasis and redox regulation in cellular signaling. *Cell Signal*. 2012;24(5):981-990. doi:10.1016/j.cellsig.2012.01.008

11. Valko M, Leibfritz D, Moncol J, Cronin MTD, Mazur M, Telser J. Free radicals and antioxidants in normal physiological functions and human disease. *Int J Biochem Cell Biol*. 2007;39(1):44-84. doi:10.1016/j.biocel.2006.07.001

12. Frayn KN, Arner P, Yki-Järvinen H. Fatty acid metabolism in adipose tissue, muscle and liver in health and disease. *Essays Biochem*. 2006;42:89-103. doi:10.1042/bse0420089

13. Tirado SMC, Yoon KJ. Antibody-dependent enhancement of virus infection and disease. *Viral Immunol*. 2003;16(1):69-86. doi:10.1089/088282403763635465

14. Cox D, Raeburn C, Sui X, Hatters DM. Protein aggregation in cell biology: An aggregomics perspective of health and disease. *Semin Cell Dev Biol*. 2020;99:40-54. doi:10.1016/j.semcdb.2018.05.003

15. Jouanne M, Rault S, Voisin-Chiret AS. Tau protein aggregation in Alzheimer's disease: An attractive target for the development of novel therapeutic agents. *Eur J Med Chem*. 2017;139:153-167. doi:10.1016/j.ejmech.2017.07.070

16. Ghosh D, Mehra S, Sahay S, Singh PK, Maji SK. α-synuclein aggregation and its modulation. *Int J Biol Macromol*. 2017;100:37-54. doi:10.1016/j.ijbiomac.2016.10.021

17. Blokhuis AM, Groen EJN, Koppers M, van den Berg LH, Pasterkamp RJ. Protein aggregation in amyotrophic lateral sclerosis. *Acta Neuropathol*. 2013;125(6):777-794. doi:10.1007/s00401-013-1125-6

18. Huang PL. A comprehensive definition for metabolic syndrome. *Dis Model Mech*. 2009;2(5-6):231-237. doi:10.1242/dmm.001180

19. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000;100(1):57-70. doi:10.1016/S0092-8674(00)81683-9

20. Azzu V, Valencak TG. Energy Metabolism and Ageing in the Mouse: A Mini-Review. *Gerontology*. 2017;63(4):327-336. doi:10.1159/000454924

21. Pomatto LCD, Davies KJA. Adaptive homeostasis and the free radical theory of ageing. *Free radical biology & medicine*. 2018;124:420-430. doi:10.1016/J.FREERADBIOMED.2018.06.016

22. Li X, Turanli B, Juszczak K, et al. Classification of clear cell renal cell carcinoma based on PKM alternative splicing. *Heliyon*. 2020;6(2). doi:10.1016/J.HELIYON.2020.E03440/ATTACHMENT/62304833-6822-4AE4-9C42-0145A440244D/MMC1

23. Ma WK, Voss DM, Scharner J, et al. ASO-based PKM splice-switching therapy inhibits hepatocellular carcinoma growth. *Cancer Research*. 2021;82(5):canres.0948.2020. doi:10.1158/0008-5472.CAN-20-0948/674382/AM/ASO-BASED-PKM-SPLICE-SWITCHING-THERAPY-INHIBITS

24. Pekala J, Patkowska-Sokoła B, Bodkowski R, et al. L-carnitine--metabolic functions and meaning in humans life. *Curr Drug Metab*. 2011;12(7):667-678. doi:10.2174/138920011796504536

25. Gérard P. Gut microbiota and obesity. *Cell Mol Life Sci*. 2016;73(1):147-162. doi:10.1007/s00018-015-2061-5

26. Vrzáčková N, Ruml T, Zelenka J. Postbiotics, Metabolic Signaling, and Cancer. *Molecules*. 2021;26(6):1528. doi:10.3390/molecules26061528

27. Sumien N, Shetty RA, Gonzales EB. Creatine, Creatine Kinase, and Aging. *Subcell Biochem*. 2018;90:145-168. doi:10.1007/978-981-13-2835-0_6

28. Gnoni A, Longo S, Gnoni GV, Giudetti AM. Carnitine in Human Muscle Bioenergetics: Can Carnitine Supplementation Improve Physical Exercise? *Molecules*. 2020;25(1):182. doi:10.3390/molecules25010182

29. Tanabe Y, Fujii N, Suzuki K. Dietary Supplementation for Attenuating Exercise-Induced Muscle Damage and Delayed-Onset Muscle Soreness in Humans. *Nutrients*. 2021;14(1):70. doi:10.3390/nu14010070

30. Mehmel M, Jovanović N, Spitz U. Nicotinamide Riboside-The Current State of Research and Therapeutic Uses. *Nutrients*. 2020;12(6):1616. doi:10.3390/nu12061616

31. Gyanwali B, Lim ZX, Soh J, et al. Alpha-Ketoglutarate dietary supplementation to improve health in humans. *Trends Endocrinol Metab*. 2022;33(2):136-146. doi:10.1016/j.tem.2021.11.003

32. de Sire A, de Sire R, Curci C, Castiglione F, Wahli W. Role of Dietary Supplements and Probiotics in Modulating Microbiota and Bone Health: The Gut-Bone Axis. *Cells*. 2022;11(4):743. doi:10.3390/cells11040743

33. Mardinoglu A, Nielsen J. Systems medicine and metabolic modelling. *Journal of Internal Medicine*. 2012;271(2):142-154. doi:10.1111/j.1365-2796.2011.02493.x

34. Rehm HL. Evolving health care through personal genomics. *Nat Rev Genet*. 2017;18(4):259-267. doi:10.1038/nrg.2016.162

35. Kirschner M. Systems Medicine: Sketching the Landscape. *Methods Mol Biol*. 2016;1386:3-15. doi:10.1007/978-1-4939-3283-2_1

36. Yang H, Mayneris-Perxachs J, Boqué N, et al. Combined Metabolic Activators Decrease Liver Steatosis by Activating Mitochondrial Metabolism in Hamsters Fed with a High-Fat Diet. *Biomedicines*. 2021;9(10):1440. doi:10.3390/biomedicines9101440

37. Zeybel M, Altay O, Arif M, et al. Combined metabolic activators therapy ameliorates liver fat in nonalcoholic fatty liver disease patients. *Mol Syst Biol*. 2021;17(10):e10459. doi:10.15252/msb.202110459

38. Lightbourne M, Brown RJ. Genetics of Lipodystrophy. *Endocrinol Metab Clin North Am*. 2017;46(2):539-554. doi:10.1016/j.ecl.2017.01.012

39. Li X, Ding K, Li X, et al. Deficiency of WTAP in hepatocytes induces lipoatrophy and non-alcoholic steatohepatitis (NASH). *Nat Commun*. 2022;13(1):4549. doi:10.1038/s41467-022-32163-w

40. Curtin F, Schulz P. Multiple correlations and Bonferroni's correction. *Biol Psychiatry*. 1998;44(8):775-777. doi:10.1016/s0006-3223(98)00043-2

41. Karran JC, Moodie EEM, Wallace MP. Statistical method use in public health research. *Scand J Public Health*. 2015;43(7):776-782. doi:10.1177/1403494815592735

42. Mishra P, Pandey CM, Singh U, Keshri A, Sabaretnam M. Selection of appropriate statistical methods for data analysis. *Ann Card Anaesth*. 2019;22(3):297-301. doi:10.4103/aca.ACA_248_18

43. Wasserstein R. George Box: a model statistician. *Significance*. 2010;7(3):134-135. doi:10.1111/j.1740-9713.2010.00442.x

44. Crick F. Central dogma of molecular biology. *Nature*. 1970;227(5258):561-563. doi:10.1038/227561a0

45. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017;45(D1):D353-D361. doi:10.1093/nar/gkw1092

46. Fabregat A, Jupe S, Matthews L, et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Research*. 2018;46(D1):D649-D655. doi:10.1093/nar/gkx1132

47. Caspi R, Billington R, Keseler IM, et al. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Res*. 2020;48(D1):D445-D453. doi:10.1093/nar/gkz862

48. Stobbe MD, Jansen GA, Moerland PD, van Kampen AHC. Knowledge representation in metabolic pathway databases. *Brief Bioinform*. 2014;15(3):455-470. doi:10.1093/bib/bbs060

49. Uhlén M, Hallström BM, Lindskog C, Mardinoglu A, Pontén F, Nielsen J. Transcriptomics resources of human tissues and organs. *Molecular Systems Biology*. 2016;12(4):862. doi:10.15252/MSB.20155865

50. Wang KC, Chang HY. Epigenomics: Technologies and Applications. *Circ Res*. 2018;122(9):1191-1199. doi:10.1161/CIRCRESAHA.118.310998

51. Wray NR, Visscher PM. Quantitative genetics of disease traits. *J Anim Breed Genet*. 2015;132(2):198-203. doi:10.1111/jbg.12153

52. Wang M, Ngo V, Wang W. Deciphering the genetic code of DNA methylation. *Brief Bioinform*. 2021;22(5):bbaa424. doi:10.1093/bib/bbaa424

53. Flaus A, Downs JA, Owen-Hughes T. Histone isoforms and the oncohistone code. *Curr Opin Genet Dev*. 2021;67:61-66. doi:10.1016/j.gde.2020.11.003

54. Zhang Y, Sun Z, Jia J, et al. Overview of Histone Modification. *Adv Exp Med Biol*. 2021;1283:1-16. doi:10.1007/978-981-15-8104-5_1

55. Ledolter J, Kardon RH. Focus on Data: Statistical Design of Experiments and Sample Size Selection Using Power Analysis. *Investigative Ophthalmology & Visual Science*. 2020;61(8):11. doi:10.1167/iovs.61.8.11

56. Mardis ER. Next-Generation DNA Sequencing Methods. *Annual Review of Genomics and Human Genetics*. 2008;9(1):387-402. doi:10.1146/annurev.genom.9.081307.164359

57. Ding C, Jin S. High-throughput methods for SNP genotyping. *Methods Mol Biol*. 2009;578:245-254. doi:10.1007/978-1-60327-411-1_16

58. Stumpf MP. Haplotype diversity and SNP frequency dependence in the description of genetic variation. *Eur J Hum Genet*. 2004;12(6):469-477. doi:10.1038/sj.ejhg.5201179

59. Krueger F, Kreck B, Franke A, Andrews SR. DNA methylome analysis using short bisulfite sequencing data. *Nat Methods*. 2012;9(2):145-151. doi:10.1038/nmeth.1828

60. Grandi FC, Modi H, Kampman L, Corces MR. Chromatin accessibility profiling by ATAC-seq. *Nat Protoc*. 2022;17(6):1518-1552. doi:10.1038/s41596-022-00692-9

61. Alseekh S, Aharoni A, Brotman Y, et al. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nat Methods*. 2021;18(7):747-756. doi:10.1038/s41592-021-01197-1

62. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*. 2009;10(3):1-10. doi:10.1186/GB-2009-10-3-R25/TABLES/5

63. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25(9):1105. doi:10.1093/BIOINFORMATICS/BTP120

64. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods 2012 9:4*. 2012;9(4):357-359. doi:10.1038/nmeth.1923

65. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology 2010 28:5*. 2010;28(5):511-515. doi:10.1038/nbt.1621

66. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*. 2016;34(5):525-527. doi:10.1038/nbt.3519

67. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biology*. 2010;11(10):1-12. doi:10.1186/GB-2010-11-10-R106/COMMENTS

68. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014;15(12):550. doi:10.1186/s13059-014-0550-8

69. Li Y, Ge X, Peng F, Li W, Li JJ. Exaggerated false positives by popular differential expression methods when analyzing human population samples. *Genome Biology 2022 23:1*. 2022;23(1):1-13. doi:10.1186/S13059-022-02648-4

70. Väremo L, Nielsen J, Nookaew I. Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Research*. 2013;41(8):4378-4391. doi:10.1093/nar/gkt111

71. Wang H, Marcišauskas S, Sánchez BJ, et al. RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor. Ouzounis CA, ed. *PLoS Computational Biology*. 2018;14(10):e1006541. doi:10.1371/journal.pcbi.1006541

72. Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I, Nielsen J. The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for Penicillium chrysogenum. *PLOS Computational Biology*. 2013;9(3):e1002980. doi:10.1371/JOURNAL.PCBI.1002980

73. Becker SA, Feist AM, Mo ML, Hannum G, Palsson B, Herrgard MJ. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nature Protocols 2007 2:3*. 2007;2(3):727-738. doi:10.1038/nprot.2007.99

74. Schellenberger J, Que R, Fleming RMT, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols 2011 6:9*. 2011;6(9):1290-1307. doi:10.1038/nprot.2011.308

75. Heirendt L, Arreckx S, Pfau T, et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols*. 2019;14(3):639-702. doi:10.1038/s41596-018-0098-2

76. Patil KR, Nielsen J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102(8):2685-2689. doi:10.1073/pnas.0406811102

77. Agren R, Mardinoglu A, Asplund A, Kampf C, Uhlen M, Nielsen J. Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Molecular Systems Biology*. 2014;10(3):721. doi:10.1002/msb.145122

78. Agren R, Bordel S, Mardinoglu A, Pornputtapong N, Nookaew I, Nielsen J. Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. Maranas CD, ed. *PLoS Computational Biology*. 2012;8(5):e1002518. doi:10.1371/journal.pcbi.1002518

79. Lonsdale J, Thomas J, Salvatore M, et al. The Genotype-Tissue Expression (GTEx) project. *Nature genetics*. 2013;45(6):580. doi:10.1038/NG.2653

80. Forrest ARR, Kawaji H, Rehli M, et al. A promoter-level mammalian expression atlas. *Nature*. 2014;507(7493):462-470. doi:10.1038/nature13182

81. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: Archive for functional genomics data sets - Update. *Nucleic Acids Research*. 2013;41(D1):D991-D995. doi:10.1093/nar/gks1193

82. Athar A, Füllgrabe A, George N, et al. ArrayExpress update - From bulk to single-cell expression data. *Nucleic Acids Research*. 2019;47(D1):D711-D715. doi:10.1093/nar/gky964

83. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*. 2015;12(3):e1001779. doi:10.1371/JOURNAL.PMED.1001779

84. Uhlén M, Fagerberg L, Hallström BM, et al. Tissue-based map of the human proteome. *Science*. 2015;347(6220):1260419-1260419. doi:10.1126/science.1260419

85. Deutsch EW, Bandeira N, Sharma V, et al. The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. *Nucleic Acids Research*. 2020;48(D1):D1145-D1152. doi:10.1093/NAR/GKZ984

86. Papatheodorou I, Fonseca NA, Keays M, et al. Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Research*. 2018;46(Database issue):D246. doi:10.1093/NAR/GKX1158

87. Cunningham F, Allen JE, Allen J, et al. Ensembl 2022. *Nucleic acids research*. 2022;50(D1):D988-D995. doi:10.1093/NAR/GKAB1049

88. Wang H, Robinson JL, Kocabas P, et al. Genome-scale metabolic network reconstruction of model animals as a platform for translational research. *Proceedings of the National Academy of Sciences of the United States of America*. 2021;118(30):2102344118. doi:10.1073/PNAS.2102344118/-/DCSUPPLEMENTAL

89. De Jager PL, Ma Y, McCabe C, et al. A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Scientific Data 2018 5:1*. 2018;5(1):1-13. doi:10.1038/sdata.2018.142

90. Luck K, Kim DK, Lambourne L, et al. A reference map of the human binary protein interactome. *Nature*. 2020;580(7803):402-408. doi:10.1038/s41586-020-2188-x

91. Aramillo Irizar P, Schäuble S, Esser D, et al. Transcriptomic alterations during ageing reflect the shift from cancer to degenerative diseases in the elderly. *Nature Communications*. 2018;9(1):327. doi:10.1038/s41467-017-02395-2

92. Malatras A, Duguez S, Duddy W. Muscle Gene Sets: A versatile methodological aid to functional genomics in the neuromuscular field. *Skeletal Muscle*. 2019;9(1):1-12. doi:10.1186/S13395-019-0196-Z/TABLES/1

93. Nishida N. Metabolic disease as a risk of hepatocellular carcinoma. *Clinical and molecular hepatology*. 2021;27(1):87-90. doi:10.3350/CMH.2020.0302

94. Zeigerer A. NAFLD - A rising metabolic disease. *Molecular metabolism*. 2021;50. doi:10.1016/J.MOLMET.2021.101274

95. Juszczak F, Caron N, Mathew AV, Declèves AE. Critical Role for AMPK in Metabolic Disease-Induced Chronic Kidney Disease. *International journal of molecular sciences*. 2020;21(21):1-23. doi:10.3390/IJMS21217994

96. Lilleker JB, Keh YS, Roncaroli F, Sharma R, Roberts M. Metabolic myopathies: a practical approach. *Practical neurology*. 2018;18(1):14-26. doi:10.1136/PRACTNEUROL-2017-001708

97. Boulangé CL, Neves AL, Chilloux J, Nicholson JK, Dumas ME. Impact of the gut microbiota on inflammation, obesity, and metabolic disease. *Genome medicine*. 2016;8(1). doi:10.1186/S13073-016-0303-2

98. O'Rourke RW. Adipose tissue and the physiologic underpinnings of metabolic disease. *Surgery for obesity and related diseases : official journal of the American Society for Bariatric Surgery*. 2018;14(11):1755-1763. doi:10.1016/J.SOARD.2018.07.032

99. Gabriel BM, Zierath JR. Circadian rhythms and exercise - re-setting the clock in metabolic disease. *Nature reviews Endocrinology*. 2019;15(4):197-206. doi:10.1038/S41574-018-0150-X

100. Gonzalez FJ, Jiang C, Patterson AD. An Intestinal Microbiota-Farnesoid X Receptor Axis Modulates Metabolic Disease. *Gastroenterology*. 2016;151(5):845-859. doi:10.1053/J.GASTRO.2016.08.057

101. Scherer T, Sakamoto K, Buettner C. Brain insulin signalling in metabolic homeostasis and disease. *Nature reviews Endocrinology*. 2021;17(8):468-483. doi:10.1038/S41574-021-00498-X

102. Lee YS, Olefsky J. Chronic tissue inflammation and metabolic disease. *Genes & development*. 2021;35(5-6):307-328. doi:10.1101/GAD.346312.120

103. Shimizu I, Yoshida Y, Suda M, Minamino T. DNA damage response and metabolic disease. *Cell metabolism*. 2014;20(6):967-977. doi:10.1016/J.CMET.2014.10.008

104. Zhu S, Surampudi P, Rosharavan B, Chondronikola M. Intermittent fasting as a nutrition approach against obesity and metabolic disease. *Current opinion in clinical nutrition and metabolic care*. 2020;23(6):387-394. doi:10.1097/MCO.0000000000000694

105. Kahleova H, Levin S, Barnard N. Cardio-Metabolic Benefits of Plant-Based Diets. *Nutrients*. 2017;9(8). doi:10.3390/NU9080848

106. Zinöcker MK, Lindseth IA. The Western Diet-Microbiome-Host Interaction and Its Role in Metabolic Disease. *Nutrients*. 2018;10(3). doi:10.3390/NU10030365

107. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. *Cell*. 2013;153(6):1194. doi:10.1016/J.CELL.2013.05.039

108. James SL, Abate D, Abate KH, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet (London, England)*. 2018;392(10159):1789. doi:10.1016/S0140-6736(18)32279-7

109. Harding C, Pompei F, Wilson R. Peak and decline in cancer incidence, mortality, and prevalence at old ages. *Cancer*. 2012;118(5):1371-1386. doi:10.1002/CNCR.26376

110. Makin S. The amyloid hypothesis on trial. *Nature*. 2018;559(7715):S4-S7. doi:10.1038/D41586-018-05719-4

111. Rajsombath MM, Nam AY, Ericsson M, Nuber S. Female Sex and Brain-Selective Estrogen Benefit α-Synuclein Tetramerization and the PD-like Motor Syndrome in 3K Transgenic Mice. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2019;39(38):7628-7640. doi:10.1523/JNEUROSCI.0313-19.2019

112. Resnick SM, Matsumoto AM, Stephens-Shields AJ, et al. Testosterone treatment and cognitive function in older men with low testosterone and age-associated memory impairment. *JAMA - Journal of the American Medical Association*. 2017;317(7):717-727. doi:10.1001/jama.2016.21044

113. Wahjoepramono EJ, Asih PR, Aniwiyanti V, et al. The Effects of Testosterone Supplementation on Cognitive Functioning in Older Men. *CNS & Neurological Disorders - Drug Targets*. 2016;15(3):337-343. doi:10.2174/1871527315666151110125704

114. Das B, Dasgupta S, Ray S. Potential therapeutic roles of retinoids for prevention of neuroinflammation and neurodegeneration in Alzheimer's disease. *Neural Regeneration Research*. 2019;14(11):1880-1892. doi:10.4103/1673-5374.259604

115. Shudo K, Fukasawa H, Nakagomi M, Yamagata N. Towards Retinoid Therapy for Alzheimers Disease. *Current Alzheimer Research*. 2009;6(3):302-311. doi:10.2174/156720509788486581

116. Fukasawa H, Nakagomi M, Yamagata N, et al. Tamibarotene: A candidate retinoid drug for Alzheimer's disease. *Biological and Pharmaceutical Bulletin*. 2012;35(8):1206-1212. doi:10.1248/bpb.b12-00314

117. Fitz NF, Nam KN, Koldamova R, Lefterov I. Therapeutic targeting of nuclear receptors, liver X and retinoid X receptors, for Alzheimer's disease. *British Journal of Pharmacology*. 2019;176(18):3599-3610. doi:10.1111/bph.14668

118. Baum LW. Sex, hormones, and Alzheimer's disease. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*. 2005;60(6):736-743. doi:10.1093/gerona/60.6.736

119. Carneiro MC, De Castro IP, Ferreira MG. Telomeres in aging and disease: Lessons from zebrafish. *DMM Disease Models and Mechanisms*. 2016;9(7):737-748. doi:10.1242/dmm.025130

120. Feigin VL, Neurosciences A, Vos T. Global , regional , and national burden of neurological disorders 1990 – 2016 : a systematic analysis for the Global Burden of Disease Study Research in context. *Lancet Neurol*. 2016;16:877-897.

121. Pappatá S, Santangelo G, Aarsland D, et al. Mild cognitive impairment in drug-naive patients with PD is associated with cerebral hypometabolism. *Neurology*. 2011;77(14):1357-1362. doi:10.1212/WNL.0B013E3182315259

122. Camargo Maluf F, Feder D, Alves De Siqueira Carvalho A. Analysis of the Relationship between Type II Diabetes Mellitus and Parkinson's Disease: A Systematic Review. *Parkinson's disease*. 2019;2019. doi:10.1155/2019/4951379

123. Krikorian R, Shidler MD, Summer SS, et al. Nutritional ketosis for mild cognitive impairment in Parkinson's disease: A controlled pilot trial. *Clinical parkinsonism & related disorders*. 2019;1:41-47. doi:10.1016/J.PRDOA.2019.07.006

124. Bohnen NI, Koeppe RA, Minoshima S, et al. Cerebral glucose metabolic features of Parkinson disease and incident dementia: longitudinal study. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 2011;52(6):848-855. doi:10.2967/JNUMED.111.089946

125. Brakedal B, Flønes I, Reiter SF, et al. Glitazone use associated with reduced risk of Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society*. 2017;32(11):1594-1599. doi:10.1002/MDS.27128

126. Mielke JG, Taghibiglou C, Liu L, et al. A biochemical and functional characterization of diet-induced brain insulin resistance. *Journal of neurochemistry*. 2005;93(6):1568-1578. doi:10.1111/J.1471-4159.2005.03155.X

127. Arnold SE, Arvanitakis Z, Macauley-Rambach SL, et al. Brain insulin resistance in type 2 diabetes and Alzheimer disease: concepts and conundrums. *Nature reviews Neurology*. 2018;14(3):168-181. doi:10.1038/NRNEUROL.2017.185

128. Wei Y, Han C, Wang Y, et al. Ribosylation triggering Alzheimer's disease-like Tau hyperphosphorylation via activation of CaMKII. *Aging cell*. 2015;14(5):754-763. doi:10.1111/ACEL.12355

129. Masciopinto F, Di Pietro N, Corona C, et al. Effects of long-term treatment with pioglitazone on cognition and glucose metabolism of PS1-KI, 3xTg-AD, and wild-type mice. *Cell Death & Disease*. 2012;3(12):e448. doi:10.1038/CDDIS.2012.189

130. Nguyen TT, Ta QTH, Nguyen TKO, Nguyen TTD, Giau VV. Type 3 Diabetes and Its Role Implications in Alzheimer's Disease. *International journal of molecular sciences*. 2020;21(9). doi:10.3390/IJMS21093165

131. Lam S, Bayraktar A, Zhang C, et al. A systems biology approach for studying neurodegenerative diseases. *Drug Discovery Today*. 2020;25(7):1146-1159. doi:10.1016/j.drudis.2020.05.010

132. Knopman DS, Jones DT, Greicius MD. Failure to demonstrate efficacy of aducanumab: An analysis of the EMERGE and ENGAGE trials as reported by Biogen, December 2019. *Alzheimer's & Dementia*. 2021;17(4):696-701. doi:10.1002/ALZ.12213

133. Shepardson NE, Shankar GM, Selkoe DJ. Cholesterol level and statin use in Alzheimer disease: I. Review of epidemiological and preclinical studies. *Archives of Neurology*. 2011;68(10):1239-1244. doi:10.1001/archneurol.2011.203

134. Greenland JC, Williams-Gray CH, Barker RA. The clinical heterogeneity of Parkinson's disease and its therapeutic implications. *European Journal of Neuroscience*. 2019;49(3):328-338. doi:10.1111/ejn.14094

135. Mardinoglu A, Ural D, Zeybel M, Yuksel HH, Uhlén M, Borén J. The Potential Use of Metabolic Cofactors in Treatment of NAFLD. *Nutrients*. 2019;11(7):1578. doi:10.3390/nu11071578

136. Zhang C, Bjornson E, Arif M, et al. The acute effect of metabolic cofactor supplementation: a potential therapeutic strategy against non-alcoholic fatty liver disease. *Molecular Systems Biology*. 2020;16(4):e9495. doi:10.15252/MSB.209495

137. Pan X, Nasaruddin MB, Elliott CT, et al. Alzheimer's disease-like pathology has transient effects on the brain and blood metabolome. *Neurobiology of aging*. 2016;38:151-163. doi:10.1016/J.NEUROBIOLAGING.2015.11.014

138. Hou Y, Lautrup S, Cordonnier S, et al. NAD+ supplementation normalizes key Alzheimer's features and DNA damage responses in a new AD mouse model with introduced DNA repair deficiency. *Proceedings of the National Academy of Sciences of the United States of America*. 2018;115(8):E1876-E1885. doi:10.1073/PNAS.1718819115/-/DCSUPPLEMENTAL

139. Dong Y, Brewer GJ. Global Metabolic Shifts in Age and Alzheimer's Disease Mouse Brains Pivot at NAD+/NADH Redox Sites. *Journal of Alzheimer's disease : JAD*. 2019;71(1):119-140. doi:10.3233/JAD-190408

140. Toledo JB, Arnold M, Kastenmüller G, et al. Metabolic network failures in Alzheimer's disease: A biochemical road map. *Alzheimer's & dementia : the journal of the Alzheimer's Association*. 2017;13(9):965-984. doi:10.1016/J.JALZ.2017.01.020

141. Bosco D, Plastino M, Cristiano D, et al. Dementia is associated with insulin resistance in patients with Parkinson's disease. *Journal of the neurological sciences*. 2012;315(1-2):39-43. doi:10.1016/J.JNS.2011.12.008

Index