The Integration of Large Biological and Clinical Datasets towards the Understanding of Human Disease

Thesis submitted for the degree of Doctor of Philosophy (Ph.D.)
In the College of Medical, Veterinary and Life Sciences
27th July 2020

Scott William Robinson, B.Sc. (Hons), M.Res.

BHF Glasgow Cardiovascular Research Centre
Institute of Cardiovascular and Medical Sciences
College of Medical, Veterinary and Life Sciences
University of Glasgow

Abstract

As the cost of high-throughput techniques reduces, and new more powerful equipment is designed, more highly-dimensional biological data will be available – and a lot of data is already in the public domain. The aim of this thesis is to investigate three case studies with interesting opportunities for the integration of large molecular datasets, corresponding clinical data, and publicly available data.

Genome-wide DNA methylation was studied with respect to hypertension. Genomic location data was used both to group individual methylation sites into meaningful functional groups such as promoter regions, and to report the results in a genomic context. Genome-wide SNP data was used to help rule out potential false positives where SNPs interfere with detection of DNA methylation.

Left ventricular hypertrophy is an intermediate cardiovascular phenotype associated with the development of heart failure. This phenotype was studied as a continuous variable – left ventricular mass index (LVMI) – using multiple sample types, in the context of a large cohort, using datasets with different classes of biomolecules and varying genomic coverage. Two alternative analysis approaches were compared, and a linear model was generated showing that a signature of molecular and clinical markers in combination best describes LVMI.

A multi-omics respiratory dataset was investigated, which includes high-throughput data for mRNA, miRNA, proteins, and metabolites and has measurements in two relevant sample types. Test statistics were performed on all datasets, identifying molecules dysregulated with asthma, COPD, and smoking. An asthma molecular interaction network was created with the significant molecules, and the links between them were formed using a variety of public data. Comparisons were made between asthma and COPD, and between asthma in smokers and non-smokers. Correlations with cell type counts may indicate cell type of origin in samples with multiple cell types like induced sputum.

Table of Contents

List of Tables

List of Figures

Acknowledgements

Firstly I would like to thank Dr Holger Husi and Prof Christian Delles whose supervision was invaluable throughout my studentship. Dr Husi contributed his expertise in all things molecular and high-throughput, and Prof Delles helped with clinical interpretation and by leading various projects I was involved in. Both were very motivational and incredibly patient throughout.

Special thanks to Dr Tony McBryan and Prof Sandosh Padmanabhan for their guidance throughout the analysis described in chapter three and to Charles McSharry for the opportunity to work on the respiratory dataset described in chapter five. Also special thanks to Dr John McClure for help with use of 'R' and with various statistics questions. Thanks to the fellow students I have had fun working alongside, especially Marco Fernandes, Yoann Gloaguen, Sophia Tsiropoulou, Mohammed Dashti, Conor Diffin, and Safaa Al-Sanosi.

A big thank you to my friends and family for being so kind and supportive while I worked on this and for helping me through a long and difficult period of illness which coincided with the studentship – in particular my 'Acoustic Night family', Emma Fitzgibbon, Ali Dunwoodie, Natalie Boyle, Rieke Dieckhoff, Becca Watson, Nina Doherty, Sergei Miller-Pomphrey, John Ferguson, Roger Wotherspoon, and my parents Brenda and Ray.

Funding for this studentship was provided by EU-MASCARA.

Author's Declaration

I declare that this thesis has been written entirely by me. As described in the text, all data were gathered by other groups and individuals and all analysis was performed by me unless otherwise stated.

Abbreviations

Alpha-2-HS-glycoprotein precursor (AHSG)

Arachidonate 15-lipoxygenase (ALOX15)

Bayesian information criterion (BIC)

Benjamini-Hochberg multi-test corrected p-values (pBH)

Body mass index (BMI)

Brain associated protein 1 (BASP1)

Ca2+ induced inactivation (CDI)

Calcyphosin-Like Protein (CAPSL)

Calmodulin (CALM)

Calmodulin-dependent protein kinase II (CAMKII)

Carboxypeptidase A3 (CPA3)

Cardiovascular disease (CVD)

C-C motif chemokine ligand 17 (CCL17)

Congestive heart failure (CHF)

Chronic obstructive pulmonary disease (COPD)

Coagulation Factor XIII (F13A1)

Cofilin-1 (CFL1)

Complement C3 (C3)

CpG islands (CGIs)

Cystatin 1 (CST1)

Cystatin A (CSTA)

Deoxyribonucleic acid (DNA)

Differential methylation (DM)

DNA Data Bank of Japan (DDBJ)

Enolase 2 (ENO2)

European Bioinformatics Institute (EBI)

Ezrin (EZR)

Fc fragment of IgE receptor II (FCER2)

Forced expiratory volume in one second (FEV1)

Fibroblast-to-myofibroblast transition (FMT)

Fibronectin (FN1)

G protein subunit beta 2 (GNB2)

Gene Expression Omnibus (GEO)

Gene Ontology (GO)

Genome-wide association studies (GWAS)

Glutathione S-transferase alpha 3 (GSTA3)

Glycoside hydrolase 18 (GH18)

Graphical user interface (GUI)

Human Proteome Organisation's Proteomics Standards Initiative (HUPO-PSI)

Illumina Methylation Analyser (IMA)

KIT proto-oncogene, receptor tyrosine kinase (KIT)

Lactate dehydrogenase A (LDHA)

Lactoferrin (LTF)

Left ventricular hypertrophy (LVH)

Left ventricular mass (LVM)

Linkage disequilibrium (LD)

Lipocalin 1 (LCN1)

Lipocalin 2 (LCN2)

LVM index (LVMI)

Lysozyme (LYZ)

Mann-Whitney-Wilcoxon (MWW)

Mitogen-activated protein kinase (MAPK)

Mass spectrometry (MS)

Messenger RNA (mRNA)

Metabolomics Standards Initiative (MSI)

Metalloproteinase inhibitor 1 precursor (TIMP1)

microRNA (miRNA)

Minimum Information About a Microarray Experiment (MIAME)

Minimum Information about Sequencing Experiments (MINSEQE)

Matrix metalloproteinase 10 (MMP10)

Moesin (MSN)

Mucin 5AC (MUC5AC)

MYC proto-oncogene (MYC)

NADPH oxidases (NOX)

National Centre for Bioinformatics (NCBI)

Next-generation sequencing (NGS)

Nuclear factor kappa B (NF-κB)

Nicotinamide adenine dinucleotide phosphate (NADP),

Normalised unscaled standard error (NUSE)

Nucleobindin 1 (NUCB1)

Nucleotide Sequence Database Collaboration (INSDC)

Oviductal glycoprotein 1 (OVGP1)

p value, Benjamini-Hochberg multi-test corrected (pBH)

p value, Bonferroni multi-test corrected (pBON)

Phosphoinositide 3-kinase (PI3K)

Pyruvate kinase M1/2 (PKM)

Principal components analysis (PCA)

Proline rich 4 (PRR4)

Prominin 1 (PROM1)

Proteomics Identification Database (PRIDE)

Reduced nicotinamide adenine dinucleotide phosphate (NADPH)

Relative log expression (RLE)

Ribonucleic acid (RNA)

Robust multi-array average (RMA)

Secretoglobin family 2A member 1 (SCGB2A1)

Sequence Read Archive (SRA)

Single nucleotide polymorphism (SNP)

Stomatin (STOM)

Submaxillary gland androgen regulated protein 3B (SMR3B)

Systolic blood pressure (SBP)

Thioredoxin (TXN)

Transaldoase 1 (TALDO1)

Transcription start site (TSS)

Trimethylamine (TMA)

Trimethylamine-n-oxide (TMAO)

Universal protein resource (UniProt)

Uromodulin (UMOD)

Wnt family member 5A (WNT5A)

1. Introduction

Until around the turn of the century most molecular biology relied upon a relatively distinct hypothesis to be tested (hypothesis-based research). The design of the experiment was based heavily on information in the literature and the understanding of it, e.g. if a particular protein was previously shown to be strongly correlated with a disease state then an experiment might be designed to yield evidence on whether the relationship is causal.

A less specific hypothesis might concern a small group of molecules, perhaps related in structure or function, but still the designer of the experiment is looking at the literature, constructing a hypothesis and testing it. If a certain family of genes is involved in a molecular function that is thought to be crucial to the understanding of the disease then every member of the family might be tested. The tests may be numerous but are specific and the selection of biomolecules is influenced by the experiments which came before.

Now, with the use of a number of high-throughput techniques, the boundaries of an experiment can move beyond these hypotheses by testing 'globally' – i.e. testing as many biomolecules of a certain class as a given high-throughput technique will allow. In doing so the previous work does not bias the future work in terms of what is tested – though it is still helpful for interpretation – and there is greater opportunity for finding novel variation. This is especially important for complex diseases where a large number of variables come into effect, and different populations may have different profiles of genetic, epigenetic and environmental causes.

This is not to say that hypothesis-free research is simply better than hypothesis-driven research but that it serves a different and complementary purpose (Fig. 1-1). Rather than starting with a hypothesis and testing it, testing 'all' of the data generates novel hypotheses to be tested with follow-up experiments. In addition to generating hypotheses, hypothesis-free experiments may also identify biomarkers or indicate how the system functions as a whole. It has been suggested that there are 'hidden hypotheses' that exist in elements of the experimental design, e.g. that the chosen tissue is relevant to the disease (1).

Alternatively it might be suggested that every iteration of the statistical test in question is essentially questioning a hypothesis. Regardless, the meaning of the phrase and the benefits of the approach which it describes are clear.

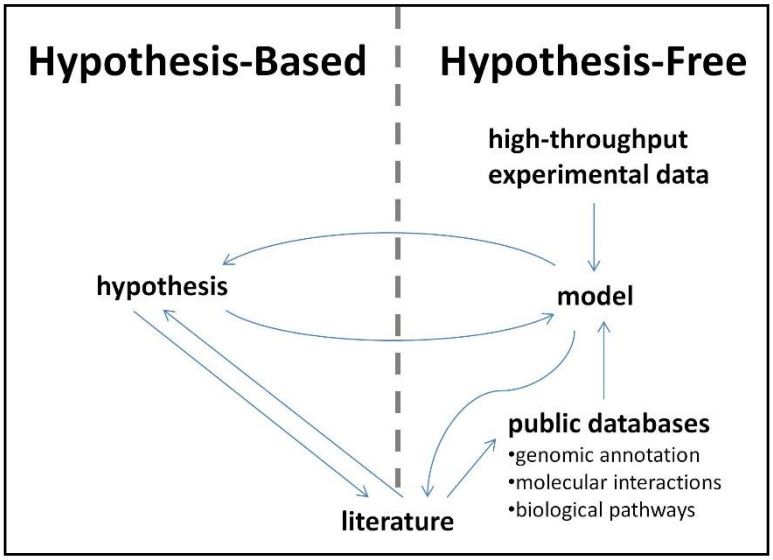

Fig. 1-1: Hypothesis-Based and Hypothesis-Free
Hypothesis-based and hypothesis-free experiments fulfil complementary roles. Hypothesis-free experiments yield large amounts of data, the interpretation of which may lead to a better understanding of the system or to new hypotheses to test. These hypotheses will then confirm or alter the model of the system.

1.1 Bioinformatics

Bioinformatics is a field at the intersection of biology, information technology and statistics - programmatic approaches are used to efficiently process, store, and analyse large biological data sets which would otherwise be unmanageable. Bioinformatics approaches are usually required with hypothesis-free research.

The data sets gathered by high-throughput techniques are known as ‘omics’ data sets – genomics, epigenomics, transcriptomics, proteomics and so on, e.g. genomics data describes genetic variation throughout the genome, proteomics describes variation in levels of proteins derived from throughout the genome and owing to various processes including alternative splicing and post-translational modification. The main sources of such information are high-throughput DNA sequencing, microarrays and mass spectrometry.

1.1.1 Modern Sequencing Technologies

Massively parallel modern sequencing technologies are capable of generating a vast amount of information compared to older technologies at a fraction of the cost and have revolutionised genetic research. This explosive growth exceeds the growth rate of storage capacity and has resulted in new challenges for data storage, retrieval, and analysis. The primary resource for raw sequence data, both DNA and RNA, is the Sequence Read Archive (SRA) which is a public repository for data from Next Generation Sequencing (NGS) machines. The SRA is part of the International Nucleotide Sequence Database Collaboration (INSDC) which is a collaboration of three partners, namely the European Bioinformatics Institute (EBI), the National Centre for Bioinformatics (NCBI) and the DNA Data Bank of Japan (DDBJ) (2).

1.1.2 Gene Expression and Microarrays

Microarrays are flat surfaces or ‘chips’ on which a large number of probes are located. Manufacturing differs from company to company but the general concept remains the same: a probe or a set of genomically proximate probes is designed to hybridize to a transcript and fluoresce upon hybridization, and the level of fluorescence will inform on the abundance of the target molecule. Each ‘spot’ on the array tests for the presence of a specific biomolecule of a particular type and the libraries of probes used have approximate genome-wide coverage.

The devices are perhaps best known for assaying gene expression though they have many other uses, including SNP genotyping, protein-DNA binding, protein-protein binding and DNA methylation. In practical terms a gene expression microarray allows the researcher to see which genes have an altered expression e.g. depending on different phenotypes or conditions.

Often it is the case that different probe sets for the same genes are highly correlated and where they are not, it suggests that either an un-annotated

cross-hybridization or alternative splicing is occurring. RNA-seq is an emerging alternative method for estimating gene expression by sequencing complementary DNA and aligning it to a reference genome. RNA-seq is a more powerful technology in that it can detect much more variation on the level of genetics and alternative splicing.

Microarrays are still a valuable technology, being cheaper, less complicated, less data-heavy, and simpler to analyse. RNA-seq, however, can detect SNPs and other genetic variation, fusion genes, and levels of different splice variants. It has also been suggested that the two technologies can be used together to detect the maximum coverage of the transcriptome (3).

NCBI's Gene Expression Omnibus (GEO) (4) and EBI's ArrayExpress (5) are the main databases for array-based data, also including some RNA-seq data and MS data. One problem with hosting data online like this is the lack of consensus on whether additional consent is required to submit the datasets to such a database. In a study on the issue participants stated that it was very (69%) or somewhat (21%) important that they were asked for their permission, indicating an ethical consent issue (6).

All data in GEO and ArrayExpress must adhere to Minimum Information About a Microarray Experiment (MIAME) (7) or Minimum Information about Sequencing Experiments (MINSEQE), which are standards dictating that enough information must be included for a second party to sufficiently analyse and/or recreate the experiment. In 2012 ArrayExpress reported having over 30,000 datasets, 27,000 more than two years previous. Both GEO and ArrayExpress continue to grow quickly and both tools, and ways to link to popular tools, are in development from their parent groups and others (8).

In some cases rather than searching for experimental data relating to a specific disease one may simply wish to know about the normal expression across different tissues of an organism. This can help identify possible sources of cross-talk, or help to study normal function or to compare systems in multiple species. BioGPS and EBI's Gene Expression Atlas offer similar services in this area, including graphical output for each gene, and useful related links (9;10). In order

to identify genes which are normally similarly expressed and so potentially share similar role or regulatory pressures, the expression profile may also be searched against other expression profiles for one with a correlation surpassing a pre-selected cut-off (11).

1.1.3 Proteomics and Metabolomics

While any large dataset may be useful, the final functional products – mostly proteins and metabolites – are perhaps the most useful. Mass spectrometry (MS) is most often the tool to collect this data on a large scale. It is used in combination with a method to separate the components of a sample prior to MS such as liquid chromatography, gas chromatography or capillary electrophoresis.

MS itself involves ionizing the molecules, fragmenting them and measuring the mass/charge (m/z) ratios of the resulting fragment molecules. Each profile of these m/z ratios relates to a specific molecule and can be used for identification. Different types of separation and different MS equipment have different strengths and limitations. For example different separation techniques will be more/less suitable for differently sized or charged molecules and offer a different degree of chromatographic resolution (12;13).

The Universal Protein Resource (UniProt) is a comprehensive database of protein structures and annotations. The Human Metabolome Database describes small molecule metabolites in humans. It contains over 40,000 metabolites and links three types of data: chemical, clinical, and molecular biology/biochemistry (14). Several similar databases exist for other species.

There is currently much less support for sharing of proteomics and metabolomics datasets than for transcriptomics, despite the growing need, and relatively few public datasets available. The Proteomics Identification Database (PRIDE) contains over 25,000 proteomics experiments (15). Metabolights is a repository hosted by EBI and launched in 2012, which currently houses 39 experiments (16). Standards for reporting proteomics and metabolomics experiments are

coordinated by Human Proteome Organisation's Proteomics Standards Initiative (HUPO-PSI), and Metabolomics Standards Initiative (MSI) respectively.

1.1.4 Pathway Analysis and Molecular Annotation

Various interaction and pathway databases can be important for downstream analysis of a wide range of experiments. Some databases are simply for storage and searching whereas others provide certain extra functionality. KEGG Pathway is a large well-curated pathway database with impressive zoomable 'global' maps which show large sections of metabolism in one view, however, part of its service is commercial (17). A similar database with fewer compounds but more pathways and reactions, and slightly more extensive pathway report pages is MetaCyc (18). Pathway Commons, collects data from a range of free access databases and provides a link for interactive viewing of the pathways by Cytoscape (19). Reactome allows mapping of a list of IDs to its database of pathways, allowing the user to see if a set of biomolecules-of-interest are involved heavily in a particular pathway (20). PathVisio is a tool which can be used to view and edit the pathways in WikiPathways (21;22). WikiPathways is non-commercial and unrelated to Wikipedia, but is also an open and collaborative platform. As PathVisio is written in Java it is cross-platform and its website provides tutorials to introduce the user to its installation process and features. These are just a few of the interactions and pathway databases available, listing just some of their 'front-end' functionality. To give scope as to the scale of the available data it is worth mentioning Pathguide, which lists over 700 biological pathway and molecular interaction resources (23).

One of the most common types of annotation used in downstream analysis is Gene Ontology (GO) (24). GO terms are a standardised way of classifying genes and their products and are split into three broad types – biological process, cellular component and molecular function. Each GO term is part of a hierarchy, e.g. the term "mRNA splicing, via spliceosome" is a child term of "mRNA processing". Gene ontologies are useful for term enrichment analysis where an algorithm detects if any terms are over-represented in a list of genes of interest,

such as a list of differentially expressed genes. The more similar a model is the more useful it is which is why gene ontologies are not static, changing with each version subject to the Gene Ontology Consortium and with advice from the scientific community. When a term is associated with a gene, a type of evidence is also recorded so that the reader or algorithm may filter out those associations with weaker evidence.

1.2 Systems Biology

In 1958 Francis Crick first discussed the central dogma of molecular biology: that information is transferred sequentially in one direction from nucleic acid to protein and cannot move in the opposite direction, which is often summarized by the phrase “DNA makes RNA makes protein” (25). While the central dogma is still a core part of our understanding of the molecular machinery that facilitates life, the picture is of course now much more complex (Fig. 1-2), as has been previously discussed (26;27).

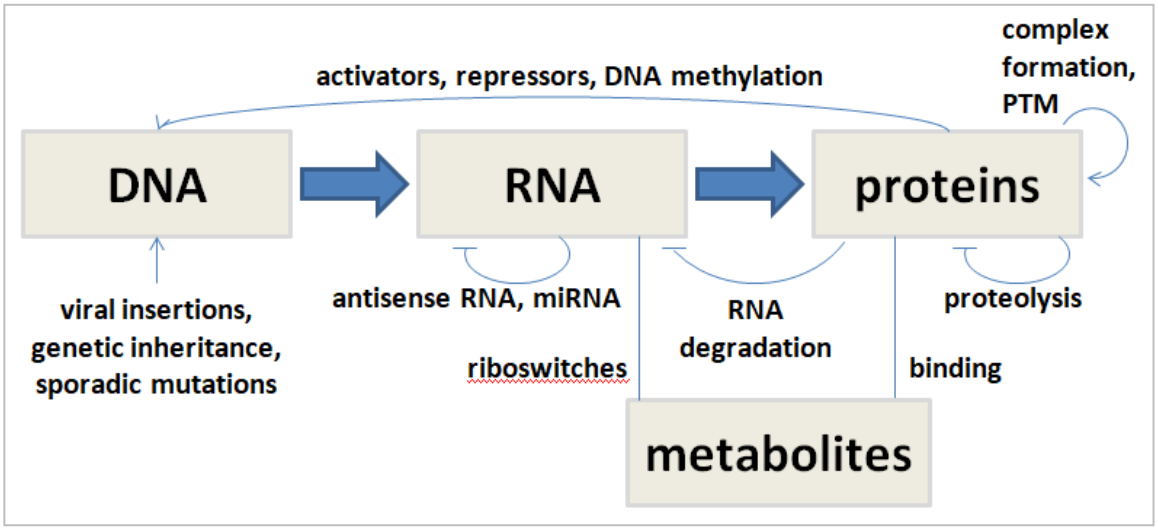

Fig. 1-2: A Summary of the Regulation of Major Classes of Biomolecule and their Associated Interactions.

We now know that in addition to the genetic information stored as the code in the form of the four bases guanine, thymine, adenine and cytosine, there is also information carried in by the modification of these bases, e.g. methylation or

hydroxymethylation of cytosine. Another type of epigenetic (above genetic) modification is to the histones, the proteins which bind to DNA to form chromatin. These epigenetic changes can act in concert (28) and they contribute to changes in levels of gene expression (29) and can direct which parts of the genetic code form a resulting mature transcript via alternative promoter selection and alternative splicing (30;31). While epigenetic changes are mostly erased during gametogenesis, they have been shown in some cases to persist through generations (32).

Similarly RNA transcripts may undergo base modifications although these are much less extensively studied (33). While most transcripts are protein-coding as suggested by the summarization of the central dogma quoted above, many are non-coding. One class of transcripts called microRNAs serve to downregulate gene expression by cleaving their specific target mRNA sequences. Some of these miRNAs seem to target thousands of specific RNAs and are extremely highly conserved across eukaryotes (34). As transcripts can undergo alternative splicing one gene may encode a large number of proteins by the removal of exons from pre-mRNA (35). The protein products of these transcripts also undergo post-translational modifications before forming a mature protein product (36), so that along with splice variants and alternative start sites, one multi-exon gene has the potential to form of a vast array of proteins. Proteins of course interact with each other and with metabolites, but also assist in various nucleic acid related processes such as transcription (37) and miRNA-directed downregulation (38).

One of the initial challenges of working with large datasets is the association between useful identifiers and the data itself. Many different sets of IDs exist for both, and some have become defunct over time as support for a database is dropped. Of those which have remained current, some databases are more inconsistent over time than others, e.g. the same ID being used for different molecules. While some conversion tools are available these issues often cannot properly be addressed without a concerted effort of someone with a biological background.

Systems biology is the study of biological systems using large biological datasets by the use of computational and mathematical approaches. While the large

datasets used for systems biology are often generated from samples gathered specifically for the study, public datasets may also be used either in conjunction with a new dataset or as the main dataset(s) for the analyses themselves.

Re-use of existing data by systematic data mining and re-stratification, one of the cornerstones of integrative systems biology, is also gaining attention. While tremendous efforts using a systems methodology have already yielded excellent results, it is apparent that a lack of suitable analytic tools and purpose-built databases poses a major bottleneck in applying a systematic workflow. Statistical meta-analysis of multiple public datasets can be very powerful. Raw data from different experiments is merged into a larger data set and reanalysed.

As more of a system is measured it should become easier to discern the pathways and structures involved. Conversely the more of the system measured the more information must be integrated and the more complex the analysis. This describes both the promise and challenges of multi-omics studies.

The most obvious way to integrate various omics experimental datasets is simply to analyse each set separately and retain only the positives from each set for further downstream analysis. The alternative to this set-by-set approach is to integrate the data prior to analysis. Generally, this can be done in one of two ways — either simply adding all datasets into one large matrix, or identifying biological relationships between the molecules and analyzing the resulting network. Specialized tools for this type of analysis are currently limited and are only beginning to emerge.

1.3 Chronic Disease

Chronic diseases comprise many of the most common and most costly challenges in healthcare today, affecting people of all ages, nationalities, and socioeconomic status. The four main types of chronic disease as outlined by WHO are: cardiovascular diseases, cancers, respiratory diseases, and diabetes.

These four groups represent 56% of all deaths globally, and another 15% of deaths are caused by other chronic diseases (39).

Nine of the top ten causes of death in high-income countries are chronic diseases whereas only two of the top ten are chronic diseases in low-income countries (40). Though they affect people of varied backgrounds their effect can be disproportionate against the less economically-privileged – 85% of the premature deaths caused by chronic diseases occur in low and middle income countries. This differential vulnerability is also found between people of different socioeconomic backgrounds within the same country (41;42).

Numerous definitions of chronic disease and lists of examples exist throughout medical and scientific communities, leading to some confusion over exactly what constitutes a chronic disease (43-45). Definitions may include a duration of over three months, increased occurrence with aging, or resistance to treatment. WHO uses the term 'non-communicable' interchangeably with 'chronic' in its literature and describes them as diseases which "tend to be of long duration and are the result of a combination of genetic, physiological, environmental and behaviours factors".

The inclusion of certain diseases under the four main types of chronic disease could be misleading, e.g. some types of cancer in particular can be fatal in the near-term. However under WHO's definition chronic/non-communicable diseases only "tend" to be of long duration. While the definition is somewhat vague with regards duration the diseases included also have shared risk factors and challenges to treatment, and many of them are complex diseases.

The four main types of chronic disease are related to four common behavioural risk factors (alcohol use, tobacco use, physical inactivity, and unhealthy diet) which lead to key physiological changes (e.g. increases in blood pressure, weight, blood glucose, and blood lipids). Governments and health agencies therefore tackle chronic disease in part through implementing policies which aim to reduce these behavioural risks e.g. through campaigns to educate the public about nutrition and disease, or the taxation of alcohol and tobacco.

Air pollution is another modifiable risk factor for many chronic diseases and one which may be more difficult to tackle for technological, economic, and political reasons. The urban centres where outdoor air pollution is highest may rely economically on emission-producing vehicles and factories. While it is described as modifiable (or potentially modifiable), there may be little an individual can realistically do to modify their own risk. Indoor air pollution is also a problem, specifically in rural areas of low-income countries where fires are needed for many daily processes and may be kept lit for long periods of time in structures with poor ventilation.

Some risk factors, unlike behavioural risk factors or air pollution, are described as non-modifiable, e.g. age, sex, socioeconomic status, genetic risk. Discovering which genetic variants are involved in a complex disease is extremely difficult due to the scope of the human genome, genetic variation throughout the species, and the variation in the molecular basis/development of a disease. Risk factors may have interactive effects, further complicating discovery.

Like environmental risk factors, individual genetic risk factors can also affect the risk of multiple chronic diseases, like HLA-DQB1 which Zenin et al found to be associated with COPD, diabetes, cancer, and dementia (46). This study used genome-wide association study (GWAS) data on over 300,000 patients and discovered 12 SNPs associated with 'healthspan' – the duration until a chronic disease is first developed. While the SNP in HLA-DQB1 was associated with multiple chronic diseases, another SNP was not associated with any chronic diseases in particular and yet still affected healthspan.

Chronic diseases sharing both environmental and genetic risk factors, other risk factors affecting healthspan in general, and one disease being a risk factor for another, all drive multimorbidity (the co-occurrence of two or more chronic conditions). Multimorbidity is rising in prevalence and poses a considerable challenge to healthcare, both in terms of challenging treatment and worsening mental health and quality-of-life with additional conditions (47).

As more risk factors are identified, and the relationships between them and with subtypes of disease are elucidated, we come closer to the era of precision

medicine, when a combination of clinical and molecular measurements will predict the optimal treatment based on an individual's genetics, epigenetics, and environment.

1.4 Cardiovascular Disease

Cardiovascular diseases (CVDs) are disorders of the heart and blood vessels. They are the leading cause of death worldwide, representing 31% of all annual deaths, most of which are caused by stroke or coronary heart disease (48). Low and middle income countries are most affected by CVD deaths, accounting for 80% of CVD deaths (49). In high-income countries like the UK - where CVD mortalities have been substantially reduced - CVDs remain a leading cause of mortality and CVD prevalence remains high so it continues to be a significant health burden and economic burden (50). Worldwide CVDs are estimated to have cost $863 billion US dollars in 2010 and are predicted to cost $1044 billion in 2030 (51).

In addition to the four common behavioural risk factors of chronic disease mentioned above, some other common risk factors for CVD are: obesity, diabetes, abnormal blood lipids (which are all linked to unhealthy diet and/or physical inactivity); age, sex, and socioeconomic status (also mentioned above); psychological stress, ethnicity, and family history. Family history is a representation of both genetic factors and behaviours/environment, so it is a useful predictor for identifying at-risk groups (52). Treatment for those with established disease may include medications (aspirin, beta-blockers, angiotensin-converting enzyme inhibitors, and statins), medical devices such as pacemakers, or costly surgical operations.

Primary hypertension is the occurrence of persistently elevated blood pressure without any identifiable cause and it is the leading mortality risk factor world-wide. It accounts for most hypertensive patients and while it does not normally cause symptoms itself, it is a major risk factor for cardiovascular disease, and is also a risk factor for chronic kidney disease.

Several tissues and pathways are involved in the maintenance of blood pressure. The renin–angiotensin–aldosterone system is one of the central pathways by which blood pressure is maintained, and involves the kidneys, lungs, brain and vasculature. Angiotensin II increases blood pressure by its vasoconstrictor action, and also stimulates the secretion of aldosterone. Aldosterone causes reabsorption of sodium and water into the blood causing a raise in blood pressure. In the vasomotor system endothelin and endothelial nitric oxide induce vasoconstriction and vasodilation respectively, in turn causing an increase/decrease in blood pressure. The sympathetic nervous system is another major controller of blood pressure, through its effect on cardiac output and peripheral resistance.

Instances of genetic variation in multiple genes in all of these systems, e.g. genes relating to sodium regulation such as the three genes for the subunits which constitute the epithelial sodium channel, may contribute to differences in blood pressure (53). Natriuretic peptides are released by the heart in response to atrial and ventricular distension, and neurohumoral stimulation from heart failure. These peptides act by inducing natriuresis (excretion of sodium in the urine by action of the kidneys), thus lowering blood pressure. SNPs associated with blood pressure have been identified both in the genes that encode the precursors of these peptides and in their receptors (54-56). Uromodulin is a gene expressed exclusively in the thick ascending limb of the loop of Henle. Although its precise function is not known, its expression results in increased co-transporter activity, and sodium retention – thereby increasing blood pressure. The minor allele of UMOD SNP rs13333226 has been shown to elicit a protective (blood pressure lowering) effect (57).

Recent GWAS studies have elucidated some of the genetic basis of the targets of current anti-hypertensive drugs, and have allowed for the discovery of new predictors and new disease mechanisms. GWAS experiments have revealed new associations between blood pressure and genes involved in calcium transport/homeostasis (e.g. *ATP2B1, SLC8A1,* and *SLC14A2* whose protein is targeted by calcium channel blocker drug nifedipine) (56;58). A SNP associated with genes for subunits of a mediator of vasodilation, guanylate cyclase, have been shown by GWAS to be associated with blood pressure (56). Guanylate

cyclase propagates nitric oxide signalling, and is involved with blocking calcium influx and dephosphorylating myosin light chains, resulting in smooth muscle relaxation. A nonsense SNP in the gene *ENPEP* is associated with hypertension and may represent a novel RAAS target (56). It encodes glutamyl aminopeptidase, an enzyme which converts angiotensin II to angiotensin III, leading to downstream vasodilation.

The genetic component of blood pressure has been estimated at between approximately 20% and 70% (59;60). However, until recently less than 3% of variation of blood pressure could be described by known genetic variants (61;62). One possible reason for this discrepancy is interaction effects, which are difficult to detect due to the large number of possible gene-gene and gene-environment combinations. An example of work in this field is a GWAS which investigated salt-sensitive effects on blood pressure in a cohort of 1,876 patients, and identified 8 novel salt-sensitive blood pressure loci (63).

Rare variants, which are difficult to detect by GWAS, may constitute a large proportion of the unknown genetic predictors of blood pressure. One solution to this issue is to use small isolated populations to search for rare variants, although increasing sample size through meta-analysis can also improve rare variant detection. A recent meta-analysis of GWAS hypertension data using the largest cohort to date (>1 million patients) confirmed all 274 previously identified genetic risk loci, confirmed 92 previously reported but not replicated loci, and identified another 535 novel loci (60). In total this set of 901 loci explained 5.7% of the variation in blood pressure, a considerable increase on the 2.8% explained by the previous 274 loci. Based on the estimate of heritability specifically for this cohort that means ~27% of the genetic heritability was explained.

The discrepancy between the explained variation and estimated heritability may also be partly attributable to epigenetic modifications – histone modifications and DNA methylation. Where GWAS is used to compare genome-wide genetic variation, epigenome-wide association studies compare epigenetic variation. In addition to epigenetic variation causing disease, disease can also cause epigenetic variation. Unlike with genetic variation in GWAS, the epigenetic

variations assessed in EWAS vary between tissues. So far studies have focussed on blood samples, however other relevant tissues may contain further important data.

The largest EWAS study on hypertension assessed DNA methylation in 17,010 individuals and identified 13 sites where differential methylation was associated with blood pressure (64). These methylations are heritable and explained 1.4% and 2% of variation in SBP and DBP respectively in a model alongside established predictors age, sex and BMI. This data in combination with gene expression analysis identified six genes (*TSPAN2*, *SLC7A11*, *UNC93B1*, *CPT1A*, *PTMS*, and *LPCAT3*) with associations between DNA methylation, gene expression and blood pressure. These results point to the modulation of vascular contractility and inflammatory processes as functionally and causally connecting DNA methylation and BP.

Left ventricular hypertrophy (LVH) is the enlargement of the wall of the left ventricle of the heart, which can cause weakening and stiffening, and acts both as a metric of cardiovascular strain and as a prognostic marker of cardiovascular events. While it is most often described in the context of left ventricular dysfunction and disease it may also result as a physiological adaption to strenuous exercise, however LVH is usually benign in this context and regresses upon reduction of exercise.

LVH is an intermediate cardiovascular phenotype which is associated with the development of heart failure, and is a risk factor for several other cardiovascular outcomes such as stroke and coronary heart disease (65). It can both affect the severity of myocardial infarction, and can be induced by myocardial infarctions (66).

At the cellular level LVH is characterised by enlarged cardiomyocytes (causing the increase in LVM), fibrosis, and changes in the extra-cellular matrix. Collagen is deposited by myofibroblasts which are transformed from fibroblasts in a process mediated by aldosterone, angiotensin II (Ang II), endothelin-1 (ET-1) and tissue growth factor-B1 (TGF-B1) (67). In particular Ang II drives this process by its secretion from activated macrophages in response to hemodynamic stress and

apoptosis. Ang II has pro-fibrotic and hypertrophic effects in part by stimulating TGFβ signalling (68). Myofibroblasts are responsible for an increase in the ratio of matrix metalloproteinases (MMPs) to tissue inhibitors of metalloproteinases (TIMPs) which has been implicated in fibrosis and heart failure (69). Increased circulating Ang II is associated with LVH, along with other hormones epinephrine and aldosterone, independent of BP (70).

LVH is more common in those with hypertension as it can be caused by hypertension-induced left ventricular remodelling, and because they share numerous risk factors. While hypertension can drive left ventricular remodelling, LVH is an independent predictor of cardiovascular morbidity and mortality. Other predictors include demographic variables (e.g. body mass index, height, age, and sex), hemodynamic variables (e.g. blood pressure, stroke volume), contractility, and history of CVD (aneurysm of the abdominal aorta, myocardial infarction) (71;72). Since LVM increases with body size it is more reflective of disease prediction when indexed by a relevant measurement or score such as height, body mass index, or body surface area. The variable chosen to index LVM affects modelling outcomes and can complicate the relationship between LVH and obesity - which is a risk factor for LVH, along with other comorbidities such as hypertension, diabetes mellitus and chronic kidney disease.

A meta-analysis of 30 studies and 37,700 patients found that LVH has a prevalence of 10-19% in the general population, 19-48% in hypertensive cohorts, and 58-77% in those with severe hypertension and CVD (73). Left ventricular mass (LVM) can be estimated using electrocardiography, and used to categorise LVH. A substantial genetic component underlies LVM, demonstrated by monozygotic twins showing substantially more similar LVM than dizygotic twins. Estimates of heritability of LVM calculated as low as 5% and as high as 84%, probably mainly due to differences in cohort recruitment, adjustment for covariates, and issues with the repeatability echocardiography in general and between modes (74-76).

Some of the genetic basis and molecular mechanisms of LVH have been discerned. Kruppel-like factor 15 (KLF15) is a transcription factor which acts to reduce LVM through inhibition of pro-hypertrophic transcription regulators which

reduce the activity of the promoters of the natriuretic peptide genes ANP and BNP (77). It is also involved in negative regulation of TGFß which is a key mediator of fibrosis. Interestingly KLF15 is reduced in pathological LVH, but not physiological LVH.

ACE acts to promote LVH through Ang II and aldosterone and its inhibition is one of the main pharmacological means of protection against LVH. A meta-analysis has identified genetic variation in the ACE gene which appears to predispose patients to LVH (78). Single nucleotide polymorphisms elsewhere in the RAAS system have also been associated with LVM (79).

GWAS studies have identified SNPs associated with LVM which pertain to genes associated with a range of established and potentially relevant processes including obesity/energy metabolism (*ATRN, NMB*), inflammation (*ATRN*), ion channels (*KCNB1, SCN5A*), and IGF signalling (*IGF1R1*) (80-82).

As LVH is a pre-clinical condition it is often only diagnosed after patients present with serious related conditions such as heart failure. As a major cardiovascular risk factor the regression of LVH is a priority for researchers, however the cardioprotective efficacy of current drugs is modest. Further elucidation of the complex processes involved in LVH is important both for the discovery of new treatments and the identification of prognostic biomarker candidates.

1.5 Respiratory Disease

Chronic respiratory diseases are diseases which affect the airways and other structures of the lung, and include asthma, COPD, pulmonary hypertension, respiratory allergies, and occupational lung diseases. Collectively these chronic respiratory conditions account for 7% of deaths globally and result in significant economic burden via both cost-of-illness and lost output (51). Although these diseases are currently incurable, various treatments can improve airflow and reduce symptom severity, improving quality of life. There are numerous environmental risk factors, including tobacco smoke, air pollution, occupational

chemicals and dust. A history of frequently occurring lower respiratory infections during childhood is also an important risk factor.

Asthma is a chronic respiratory disease which is characterised by variable airflow obstruction, mucous hypersecretion, and airway inflammation and hyperresponsiveness, resulting in coughing, shortness of breath, difficulty breathing, chest tightness, and wheezing. These symptoms can be triggered by allergens, irritants, infections, exercise and some medications (e.g. beta blockers), and are reversible either spontaneously or with treatment. In people with asthma one or several of these triggers in combination can lead to inflammation and constriction of the bronchial tubes.

Asthma affects ~300 million globally (with rising prevalence) and causes 250,000 deaths annually (83). Unlike many chronic diseases asthma is common in childhood. Another way it differs from many chronic diseases is its relatively low mortality rate (383,000 deaths globally in 2015), although it has a large impact on quality-of-life and is responsible for work and school absenteeism. It is commonly treated using inhaled ß-agonists which act as bronchodilators and corticosteroids which inhibit inflammation, however other medications for treatment of asthma are emerging. Asthma is described as severe if the symptoms are not well controlled by bronchodilators and corticosteroids. While often figures of 5-10% are quoted there is little evidence for these estimates as discussed by Chung et al (84).

Asthma is a complex multifactorial disease and mechanisms underlying the development of it, including age-of-onset and severity, are not well understood. Aeroallergen sensitisation and viral infections have been identified as key causal factors in its development, working independently or synergistically (85). There is a large genetic component to the disease - hundreds of genes are significantly associated with it (86) and heritability estimates ranging from 35% to 95% (87). Asthma is often categorised as allergic/non-allergic or atopic/non-atopic, however there are many established and emerging categorisations based on clinical phenotypes (e.g. exercise-induced, obesity-associated, infection-related), endotypes categorised by particular immune cell recruitment (e.g. eosinophilic, neutrophilic, Th2, Th2-low), and molecular biomarkers (88-90).

Atopy describes an exaggerated immune response mediated by immunoglobulin E (IgE), as determined by skin prick tests. IgE binding to receptor FcεRI, mediates acute degranulation of mast cells and basophils, and may also drive inflammation by facilitating antigen presentation. Around a third of asthmatics are non-allergic and around a quarter are non-atopic. Atopic and non-atopic asthma have been shown by multiple studies to be very similar in terms of immunopathology and symptomatic presentation, however there are numerous differences including age-of-onset and male:female ratio (91). There are also differences in the cellular and molecular level, including various signs of epithelial damage and different abundances of various cell types infiltrating the bronchial mucosa - mast cells, basophils, eosinophils, lymphocytes, and macrophages. In non-atopic patients macrophage infiltration is driven by differential expression of the α-subunit of the granulocyte macrophage colony stimulating factor receptor (GM-CSFr). These differences may be moderated by differences in levels of cytokines and their receptors. Interleukins 4, 5 and 8 have been detected with greater abundance in atopic asthma and IL-2 and γ-IFN in non-atopic asthma. Not all of these associations have proven consistent, possibly pointing to the heterogeneity of asthma even within currently-recognised subtypes.

The drug omalizumab interferes with the binding between IgE and its receptors and has proven effective at treating atopic asthma (92). Interestingly IgE is also overexpressed in non-atopic asthmatics and omalizumab has had promising results in this context also (93). In addition to classical induction of the IgE pathway through allergen response, viral infections and air pollution can also induce IgE. Omalizumab's impact on respiratory symptoms are mainly believed to be mediated through its inhibitory effect on the allergic inflammation cascade, however there is evidence to suggest that there are other distinct IgE-asthma mechanisms also affected, such as airway remodelling (94).

Th2 lymphocytes are involved in both atopic and non-atopic asthma via Th2 cytokine release. Th2 cytokines IL-4 and IL-13 induce B-cells to synthesise IgE. Anti-IL-4 medications have failed to have efficacy in regressing symptoms of asthma possibly due to the redundancy found in IL-13 due to having several shared actions. This is further supported by the successful action of dupilumab,

a monoclonal antibody directed against IL-4Rα, affecting both IL-4 and IL-13 signalling. In one study it has led to an 87% reduction in asthma exacerbations amongst moderate and severe atopic asthmatics. Chemokines are cytokines which induce chemotaxis. They are mainly released by macrophages and the epithelium, and are important in asthma particularly for the recruitment of the pro-inflammatory eosinophils and neutrophils. Antibodies against specific chemokines and their receptors can be effective targets for treatment of asthma.

Eosinophilic inflammation occurs mainly via the release of granule-associated substances in both atopic and non-atopic asthma. Eosinophils are also responsible for the release of other relevant classes of molecules like leukotrienes, cytokines, and growth factors. IL-5 is involved in eosinophil differentiation, survival and activation. Anti-IL5 therapies are therefore effective against eosinophilic asthma.

Th1 cytokine IFN-γ has been associated with asthma severity and was found to have higher expression in non-allergic asthmatics (95). Th2-low subtypes of asthma are also common and occur particularly in severe asthma (96). Th9 cytokines such as IL-9 can also lead to asthma symptoms. Genetic and epigenetic variation in Th1 and Th2-associated cytokine genes in asthmatic patients is well established (97). Th1 and Th17 cells are responsible for the neutrophilic inflammation in the context of severe asthmatic airways. Th17 cells release IL-17 (which helps regulate IgE synthesis) and viral infections are also associated with neutrophilic inflammation.

Bronchoconstriction in the context of atopic asthma triggers the release of pro-fibrotic cytokines and the deposition of collagen (98). This is similar to the processes of fibrosis and collagen deposition seen in LVH - also involving TGF-β signalling and cytokines. Hypertrophy also occurs in asthma, with respect to smooth muscle of the airways, and is also mediated by TGF-β amongst others, including: IL-1β, IL-6, histamine, serotonin, leukotrienes and vascular endothelial growth factor (VEGF). This structural remodelling of the airways occurs because of persistent inflammation, and is mediated by TGF-α as well as TGF-β.

Defective cell adhesion components (affected by genetic variation, viruses, or pollutants) such as desmosomes, and tight junctions reduce the barrier function of the airways. This increases the number of allergens with the access to exacerbate symptoms via dendritic (antigen-presenting) cells.

A cluster analysis based on spirometry variables and age-of-onset, and another cluster analysis based on relevant cytokines both indicate the existence of somewhat distinct subgroups of asthma (99;100). Different existing categorisations of asthma have varied responses to a particular class of drugs, however variation exists within categorisations, overlaps exist between categorisations, and some asthmatics still suffer from severe asthma. Further understanding of the various molecular mechanisms of asthma is therefore appealing in terms of unveiling new drug targets and better predictions of drug response.

GWAS experiments have revealed hundreds of SNPs associated with asthma, including many SNPs in genes already known to be involved in asthma (101). Transcriptomics is useful for providing a mechanistic explanation of the effect of relevant SNPs and in some cases showing downstream effects. Due to harsh statistical correction in GWAS and the rarity of variety of SNPs which are risk SNPs for asthma, transcriptomics can be more powerful on the gene-level. Independently or in combination, transcriptomics along with other omics analyses – e.g. proteomics, metabolomics, and epigenomics – enhance the molecular description of the pathology of asthma.

Chronic Obstructive Pulmonary Disease (COPD) and lower respiratory infections are the third and fourth leading causes of mortality world-wide, and both are among the top causes of mortality regardless of national income level (40). In the context of COPD lower respiratory tract infections are more frequent and there is a 3 to 10-fold increase in lung cancer risk in comparison with smokers without COPD. There is also a greater risk of cardiovascular mortality in COPD patients, particularly through an increased rate of myocardial infarction and stroke (102). Over 3 million people die each year from COPD, over 90% of which occur in low and middle-income countries. COPD is usually the result of smoking, or exposure to air pollution, and the amount of exposure correlates

with symptom severity. Although smoking is the biggest risk factor for COPD, only a fraction of smokers develop the disease, indicating the importance of genetic susceptibility.

COPD is a complex, heterogeneous obstructive airway disease like asthma. It is categorised by many of the same symptoms seen in asthma and the two diseases share several biomarkers and biological pathways in common, and a moderate genetic correlation (103). But Asthma and COPD also contrast with each other in several ways. Unlike asthma, the airway obstruction in COPD is largely irreversible, meaning that it responds poorly to anti-inflammatory treatment. The airway obstruction in COPD is slowly progressive and is associated with enhanced infiltration of immune cells, a thickening alveolar wall, and damage to the epithelium – it may be described as a combination of emphysema and bronchitis in an environment of inflammation in the small airways.

Oxidative stress is increased in COPD, which can lead to mucous hypersecretion and the inactivation of antiproteases. The imbalance of proteases and anti-proteases is one of the main mechanisms by which emphysema develops in the COPD-affected lung. There are many possible proteases involved, and they are of different classes – serine proteases (in particular A1AT), matrix metalloproteinases, cysteine proteases, and aspartic proteases (104). Emphysema is also induced via the disruption of homeostatic maintenance and repair. The system can be skewed towards apoptosis by the reduction in vascular endothelial growth factor (VEGF), increasing the sensitivity of the alveolar wall to oxidative stress and proteases (105).

Like asthma, TGF-β signalling and WNT signalling are involved in pathogenesis of COPD. These pathways may be controlled via the miR-15/107 family, as discovered through a multi-omics study using miRNA and mRNA microarrays (106). Like in asthma and in LVH, TGF-β is involved in fibrotic remodelling and it is involved in VEGF expression in fibroblasts. Also like asthma there are various types of T cells which can help describe the molecular basis of an individual's condition – including Th2 and Th17 responses in both asthma and COPD (107).

Reductions in levels of miR-146a and let-7c have been observed in COPD patients (108). The reduction in miR-146a leads to increases in cyclooxygenase-2 and subsequently prostaglandin E2 which is correlated with COPD severity - it acts to maintain inflammation of the airway epithelium and cell growth and senescence in lung fibroblasts, causing a reduction in the repair capacity of the lung (109). The reduced level of let-7c leads to increased tumor necrosis factor receptor 2 which can have a pro-inflammatory outcome and is involved in the pathogenesis of COPD.

Inflammation in the airways in COPD is mostly neutrophil-associated. Neutrophilic airway inflammation is usually steroid-resistant and is not mediated by Th2 mechanisms. However up to 40% of COPD patients have eosinophilic COPD and may benefit more from asthma medications with Th2-related targets (110;111).

The most commonly studied molecular biomarkers of COPD exacerbation are C-reactive protein (CRP), IL-6 and TNF-α (112). CRP is produced by the liver in response to IL-6 secretion by macrophages and T-cells. It binds to the surface of dead or dying bacterial cells in order to activate the complement system and promote phagocytosis by macrophages. CRP is associated with increased mortality in patients with COPD, along with molecular marker fibrinogen, and clinical markers: shorter six minute walk distance, elevated heart rate, and white cell count (113).

1.6 Aims

Many serious diseases are poorly understood and difficult to treat, seemingly due to a large number of independent risk factors operating in a complex system. High throughput methods allow more of the system to be measured which could lead to better biological understanding of a disease and ultimately more patients being targeted with effective treatment or preventative measures (relevant to their disease subtype, genetic background and/or environmental factors).

Ideally this kind of research would use large cohorts and would not be restricted to a single class of biomolecules. Multi-omics research requires the establishment of multi-omics workflows applying existing methods and possibly the development of new methods. Equally when it comes to downstream analysis we can use various public databases to help with interpretation.

The aim of this work is to study cardiovascular and respiratory disease by approaches which rely on the integration of large datasets, whether integrating different omics layers (genomics, transcriptomics, proteomics) or contextualising results with information from various public databases or with clinical measurements. These approaches can be compared in some cases, highlighting potential strengths and weaknesses.

2. Materials and Methods

2.1 General Methodology

A wide variety of methods may be used for both preprocessing and downstream analysis, found as stand-alone software or on shared platforms. ‘R’ is a scripting language and environment primarily developed for statistical computing. It is particularly useful in bioinformatics and systems biology both because of the base statistics functions and the more specific bioinformatics packages, many of which are available through the open source Bioconductor project (114).

R has a diverse range of uses, from preprocessing, to statistical testing and on to downstream analysis, and in particular it has great utility in microarray processing and many other omics data pre-processing. Many graphical user interface (GUI) alternatives exist, however often what is gained in speed and simplicity is lost in flexibility and power, and many of these GUI applications such as Partek, SPSS and IPA are commercial.

2.1.1 Data Pre-Processing

One popular pre-processing procedure is called Robust Multi-array Average (RMA), which background adjusts, quantile normalizes, log-transforms and summarizes from individual probe values down to probe set values (115). A log2 transformation is performed in order to acquire a more normal distribution to allow the use of parametric tests. The log2 scale is also beneficial to the interpretation of fold changes as upregulations and downregulations are scaled equally around zero, as opposed to raw downregulations being found between zero and one and upregulations being found between one and infinity.

There are now a diverse range of mass spectrometers used to generate MS data in proteomics, with various advantages and limitations. Similarly there are also a considerable number of algorithms developed to query and cross compare the

tandem MS data (116). IDEOM is an Excel interface used for the analysis of LC/MS and GC/MS metabolomics data (117). It alleviates the requirement for either scripting skills or in-depth understanding of preprocessing procedures in obtaining a filtered, interpretable list of metabolites from a raw input file. IDEOM uses XCMS (118) to extract raw peaks and mzMatch (119) for peak matching, noise filtering, gap-filling and annotation of related peaks. After this preprocessing and identification, worksheets are populated with metabolite data and graphs of statistical output.

2.1.2 Dimensionality Reduction

After pre-processing the data should be in the format of a large matrix with rows by biomolecule and columns by patient (or the transpose of this). Distance matrices may be useful at this stage, showing either distances between samples or between molecules and often displayed as a heatmap. The 'distance' is most commonly Euclidean distance, Manhattan distance or some type of correlation (Pearson, Spearman etc.).

Dimensionality reduction methods reduce the number of rows describing each patient so that the data may be plotted on a 1D, 2D or 3D graph. The points on the graph are often coloured according to various features — plotting these points and colouring by certain variables may prove informative as to quality control by identifying samples which are extremely different from others, or displaying the association between clusters and various variables.

Sammon mapping compresses a highly dimensional dataset down to a plottable number of dimensions (120). It does this while minimizing what is described as stress, which is a representation of the error in the distances between points in the new data space as compared to the original. Principal Components Analysis (PCA) on the other hand seeks to essentially tilt the axis through the data space, such that the 'first' principle component (denoted "PC1") captures the maximum amount of variance possible and all components remain orthogonal to each other (121). The order of the naming of the components is determined by

ranking the list based on the amount of variance described by each, i.e. PC2 captures the second most variance. While all the data is maintained (unlike in Sammon mapping) scree plots may be used to show the amount of variance for each component and to decide how many principle components are worthy of examination.

Each PC can be described as correlating to a certain extent with each original input variable, such that if a particular PC separates the samples into two clusters, those variables mostly responsible may be identified. A technique which can be used in combination with PCA is varimax rotation, in which the top PCs are selected and further rotated such that for each varimax-rotated component the variance across the correlations with the input variables is maximized — therefore each PC may be said to largely correlate to a small number of input variables (122;123). In some cases, where data is highly dimensional and highly correlated, it can be useful to input these data into statistical tests — this way each varimax-rotated PC would be representative of a group of biomolecules and each group can be tested alongside the other in the same model, rather than doing iterative tests in which different molecules may be accounting for the same difference in the dependant variable. This is, in a sense, a more powerful approach — in that less statistical power is lost through multi-test adjustment (Section 2.1.3).

2.1.3 Significance Testing

Several possibilities exist for statistical analysis on a list of biomolecules or groupings of biomolecules, each with its own set of assumptions. Standard statistical tests to compare two groups such as the nonparametric Mann-Whitney-Wilcoxon (MWW) and parametric Student's t-test may be used depending on the assumptions appropriate to the data. Alternatively the independent variable of interest may be continuous in which case a different technique such as simple linear regression is used. Linearity of relationships should be considered and non-linear regression and mutual information employed where appropriate (124). Non-linear relationships are poorly

accounted for by linear approaches and can be responsible for apparent noise in a system (125).

These may be iterated over every available probe/biomolecule and the results should be multi-test adjusted. These tests seek to identify those results with a low (e.g. < 0.05) probability of occurring by chance, and so if 100 tests were done then 5 results would be expected to show positive without multitest correction. Bonferroni correction is direct and intuitive and simply involves multiplying each p-value by the number of tests done. It is very stringent however, especially with a very large number of tests, so other less stringent methods have been developed (such as the Benjamini–Hochberg correction), which often take into account the rank of each test as ordered by p value (126).

New methods have been developed with the advent of genome-wide technologies. Limma is an R package which was developed to facilitate gene expression microarray analysis (127). With Limma one may use linear modelling to adjust for batch effects, include technical replicates, and analyse complex multifactor experiments. It also provides the option to compute a moderated t-statistic, essentially borrowing information from other genes.

2.1.4 Downstream Analysis

Two of the most common procedures in downstream analysis of omics data sets are term enrichment and pathway analysis. Term enrichment analysis relies on utilising a database of information which links to the biomolecules being studied. Perhaps most commonly this would be the Gene Ontology database. In this case a group of genes with statistical significance could be described as being ‘enriched for a gene ontology’ — i.e. a larger number of the genes are linked to that gene ontology than one would expect by chance, indicating that a particular biological process, molecular function, or cellular component is of particular importance. This enrichment itself is also given a statistical value, which can be used as a cut-off. The same analysis can also identify terms that are under-represented rather than over-represented. Another way to do term

enrichment analysis is with ranked lists – i.e. rather than describing a group of genes of interest the entirety of the data is taken into consideration and certain terms are identified as enriched towards the top or bottom of the list.

Cytoscape is a tool primarily designed for network visualization and analysis; it makes use of a wide variety of plug-ins to extend its functionality which are designed by the scientific community. ClueGO (128) is a popular Cytoscape plug-in used for term enrichment analysis. It calculates enchrichment/depletion tests based on the hypergeometric distribution. As the name suggests this is done with GOs as the terms, and allows the user to subcategorize based on the three main categories or by evidence codes. It also provides the capability to analyze with KEGG, WikiPathways and Reactome terms. For additional functionality, CluePedia (129) can be added to ClueGO to produce networks with custom correlation scores and other data plotted as edges (lines) between nodes representing genes and gene ontologies.

Cytoscape has a “pathway database” ‘app category’ containing plug-ins which derive data from a variety of information sources and provide some appropriate tools for pathway editing and enrichment analysis: CyKEGGParser manipulates KEGG files (130); ReactomeFIPlugIn facilitates pathway enrichment analysis based on the Reactome database (131); an alternative interface to WikiPathways is provided; and Metscape allows users to build and analyze networks of genes and compounds, use gene expression and metabolomics data to identify enriched pathways and their metabolic consequences and rely on data from several different sources (132). These are just a few examples of the tools available in this category. There are many similar tools under several other related categories and many of them are found repeatedly across categories.

2.2 DNA Methylation in Essential Hypertension

2.2.1 Design

A hypertension genotype risk score comprising 35 SNPs identified from meta-analysis of various genome-wide association studies (GWAS) (133-135) was developed by Sandosh Padmanabhan, who designed the project for this chapter. Each unit in this risk score represents presence of a disease-associated allele. Peripheral whole blood samples from a previous GWAS – covering patients with a wide range of blood pressure phenotypes ranging from severe hypertension to low blood pressure (57) - was restricted to those which were homozygous for the uromodulin *(UMOD)* SNP rs13333226, as it has a greater effect on blood pressure. The ages and genotype risk scores were similar between hypertensives and normotensives (mean ages: 55, 58; mean genotype risk scores: 36.3, 35; Table 2-1). Body Mass Index (BMI) was categorically different between hypertensives ('overweight' on average, with a mean BMI of 29.2 kg/m^2) and normotensives ('normal weight' on average, with a mean BMI of 24.0 kg/m^2).

Variable	Normotensive Patients	Hypertensive Patients
Age (years)	58	55
Body Mass Index (kg/m^2)	24.0	29.2
Genotype Risk Score	35.0	36.3

Table 2-1: Patient Demographics of the DNA Methylation Cohort.
The hypertensive group has a substantially higher BMI at similar genetic risk score and age. Each group comprises 12 patients.

12 hypertensive and 12 control samples were selected on the basis of having an equal number of high and low genotype risk score and *UMOD* genotype, such that every combination, or 'subgroup', of blood pressure, genotype risk score and *UMOD* genotype has n=3 (Fig 2-1). *UMOD* genotype was analysed distinctly from the genotype risk score because the homozygous selection gives two clear discordant groups. While the sample size is small compared to the number of tests the cohort selection is quite homogenous – all patients being male Swedes of a similar age.

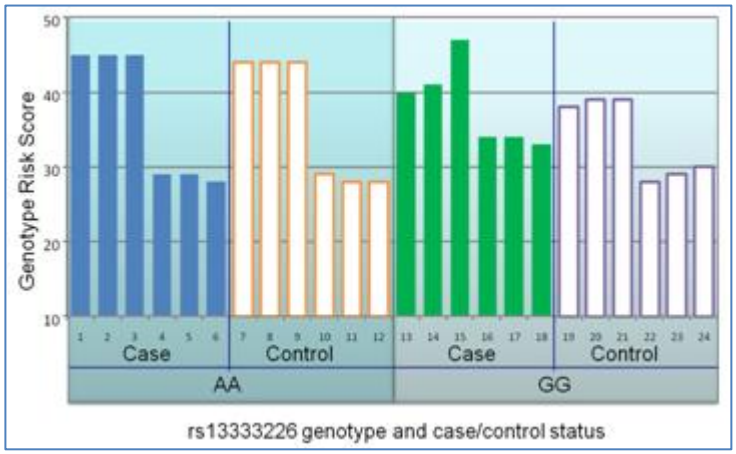

Fig. 2-1: Histogram Describing Experimental Design.
24 patients are selected for having the most extreme hypertension, genetic risk scores and being homozygous for the minor or major allele of the *UMOD* SNP 'rs13333226'.

The most extremely discordant hypertensive then would have high BP, low genotype risk and the minor, protective *UMOD* allele (G), whereas the most discordant normotensive group would have low BP, high genotype risk and the major *UMOD* allele (A). The high genetic risk group had a mean score of 41.6 (SD = 3.7) and the low genetic risk group had a mean score of 29.7 (SD = 2.0).

2.2.2 Statistical Analysis

Peripheral leukocyte DNA from these samples was interrogated using Illumina's Infinium HumanMethylation 450 BeadChip. Microarray experiments were conducted by BGI (previously Beijing Genomics Institute) and summarised into a standard tab-delimited GenomeStudio output file. Illumina Methylation Analyzer (IMA) (136) was used to import the data into 'R' and filter it using the defaults laid out in the pipeline provided:

- Sites with missing data were removed
- Sites in which more than 50% of samples had detection p-values greater than 0.05 were removed
- Samples in which more than 75% of sites had detection p-values less than 1×10^{-5} were kept

After filtering 461,856 sites remained from the original 485,577 and no samples were removed. Peak correction normalization was also performed using IMA to correct for the different Illumina probe types used in the arrays (137). Beta values for each site were logit-transformed as in previous studies (138;139).

The 'ComBat' function from the R package 'sva' (140) was used to remove the batch effect from the data prior to clustering using Principal Component Analysis. Euclidean distances between each sample were calculated and used to empirically determine distances between groups. Shapiro-Wilk test showed that the distances were not normally distributed ($W=0.95$, $p=3.05 \times 10^{-5}$) and normal Q-Q plot revealed a negative skew, so MWW was used to identify clustering – by testing distances within groups against distances between groups.

Limma (127) was used to assess differential methylation (DM) in every remaining site. The R function 'stepAIC' was used to select various combinations of variables suspected of influencing DNA methylation and linear modelling was then used to assess how well the models explained the data. Linear regression was iterated over all sites and the model with the highest mean adjusted-R^2 was selected. This model involved taking high/low blood pressure, *UMOD* genotype and high/low genetic risk as one categorical variable – the subgroup. Additionally the batch effect of the experiment being conducted over two chips and BMI were included in the model. BMI was included as it was less well controlled between groups, compared to other confounders such as age. Several comparisons were made between different subgroups and combinations of subgroups. Contrasts made were 'discordant cases – others', 'discordant controls – others', 'hypertensives – normotensives', 'UMOD_AA – UMOD_GG' and 'high genetic risk – low genetic risk'. Resulting p values were multi-test corrected with the Benjamini and Hochberg method, over all sites within each contrast.

Gene regions (i.e. TSS, gene body etc.) were analysed rather than entire genes as DNA methylation has been associated with different changes to expression depending on the gene region, as described in Section 3.1.2. If a gene were to have hypermethylation in one region and hypomethylation in another the 'effect' could be in the same direction, so averaging over a whole gene seems inappropriate. Thus DM is detected by site or region, however some downstream

analyses were done on a gene level. Not only is this a biologically relevant way to group CpGs, potentially showing a more complete picture, it also allows for more statistical power than testing by individual CpG site allows (due to multi-test correction). This is particularly important since due to the relatively small sample size (24) for the number of data points (450,000), only those CpGs with extreme raw significance will persist through multi-test correction.

Values were summarised, by geometric mean (141), for all gene regions and CGI regions as annotated by Illumina. CGI regions which later showed to have significant DM were inspected manually on UCSC Genome Browser (142) to assign potentially relevant genes to them and give the results biological meaning. Geometric mean was used to summarise the extent and potential effect of methylation of gene regions as it treats outliers in a conservative manner. The same model and contrasts were used and resulting p-values were multi-test corrected (Benjamini-Hochberg, p_{BH}) over all region types (1^{st} exon, north CGI shore, etc.), for each contrast. Lists of gene names ordered by p value were generated based on these analyses and input into Gorilla (143), a gene ontology term enrichment analysis web application. The number of genes related to hits were categorised by chromosome and chromosome 1 was tested for enrichment in gene-level hits by chi-squared test.

The mean of the genome-wide methylation was calculated for each patient and multi-factor ANOVA was used to analyse differences by hypertension status, genetic risk and *UMOD*, taking batch effect, age and BMI into account. Welch's t-test was used to test whether a global increase or decrease could be contributing to discordant cases or discordant controls. Linkage disequilibrium (LD) blocks have been shown to be biologically relevant for studying DNA methylation as in genetic variation (144). MWW was used to check if differential methylation was higher within LD blocks containing the risk SNPs comprising the genetic risk score, and p-values were Bonferroni multi-test corrected.

2.3 Integrative Analysis of Variation in Left Ventricular Mass Index

2.3.1 Design

The InGenious HyperCare cohort was established by first identifying index patients, and through them selecting families to be included. Index patients were diagnosed with hypertension before the age of 50 years and either had a blood pressure of at least 160/95mmHg on two occasions or were on treatment with at least two antihypertensive drugs. Families were included if a family member of an index patient had identical age criteria, at least 140/90mmHg on two occasions or treatment with at least one antihypertensive drug. Minimum family size was four, and normotension was defined as blood pressure below 140/90 mmHg in absence of treatment. The cohort consists of 1589 participants belonging to 460 families recruited between 2008 and 2010 in 19 study centers in Europe. Cardiovascular phenotypes from 535 of these individuals have been previously published (145). For the present analysis 270 participants from four sites in Gdansk, Krakov, Glasgow and Prague were selected for detailed molecular phenotyping. In order to represent a continuum of blood pressure and LVM, patients with hypertension and normotensive controls were included. This study has been approved by local ethics committees and all participants gave written informed consent.

There were approximately equal numbers of males and females at a mean of 48 years of age, with overall normal mean diastolic blood pressure and BMI and heart rate and borderline high systolic blood pressure and borderline low estimated glomerular filtration rate eGFR (Table 2-2). As LVMI in the cohort takes a normal distribution containing 34 cases of LVH we in this study focused on the continuous analysis of LVMI data. Significant results with linear regression were also tested with logistic regression to show that these results extend to the clinical categorization into presence and absence of LVH.

Clinical Variable	Summary (ratio or mean)	Standard Deviation	Range
Sex (M:F)	134:136	-	-
Age (years)	48	14.6	17-80
BMI (kg/m^2)	28.7	5.0	18.9-51.7
HR (bpm)	70.0	10.7	46-103
DBP (mmHg)	84.5	11.6	59-117
SBP (mmHg)	141.0	20.8	95-223
eGFR (mL/min)	88.4	19.9	18.9-181.4
Diabetes Count (%)	6.7	-	-
LVMI (g/m^2)	88.2	18.4	42.5-139.7
LVH Count (%)	12.6	-	-
Blood Pressure Medication (%)	47.8	-	-
RAAS Count (%)	41.5	-	-

Table 2-2: Clinical demographics of the subset of the InGenious HyperCare cohort used for this study.
Mean values are given in the summary column unless stated otherwise. BMI: body mass index, HR: heart rate, DBP: diastolic blood pressure, SBP: systolic blood pressure, eGFR: estimated glomerular filtration rate, LVMI: left ventricular mass index, LVH: left ventricular hypertrophy, antihypertensive medications, RAAS: RAAS-blocking medications

2.3.2 Echocardiography

Echocardiograms were performed by experts in the recruiting sites according to strict acquisition procedures detailed in the InGenious HyperCare protocol. In brief, echocardiographic instruments equipped with 2.5 to 3.5 MHZ transducer with M-mode, 2D and Doppler capability had to be used. Images had to be acquired in sequence along parasternal long-axis view, parasternal short-axis view, apical four-chamber view, apical two-chamber view and apical three-chamber view.

In order to minimize observer dependency and to improve overall quality, echocardiographic tracings from the 270 individuals included in the present analyses were sent in electronic format to the central laboratory at Istituto Auxologico Italiano, Ospedale San Luca, Milan, for centralized reading according to American Society of Echocardiography standards (146). Left ventricular mass was calculated using the formula recommended by the American Society of Echocardiography (147).

2.3.3 Molecular Testing

The miRNAs together with one technical normalization miRNA (cel-miR-39) were measured in plasma samples by standard Q-PCR using a LightCycler® 480 (Roche) quantitative real time PCR machine, and employing general methodology as directed by the manufacturer. miRNAs were amplified using a miRNA specific forward primer and a universal reverse primer (miScript kit, Qiagen). Raw qRT-PCR data were analyzed using free qRT-PCR analysis software: LinReg (148) which calculates an arbitrary value which is representative to the amount of a specific miRNA that was present at the start of the amplification.

Capillary electrophoresis mass spectrometry analysis was performed on urinary samples by Mosaiques Diagnostics using a P/ACE MDQ capillary electrophoresis system (Beckman Coulter, Brea, CA) coupled on line to micro-TOF-MS instrument (Bruker Daltonics, Bremen, Germany) (149). Serum metabolomic profile was assessed by INCLIVA using a Bruker Advance DRX 600 spectrometer (150). Serum carboxy-terminal propeptide of procollagen type I (PICP) was measured by using the METRA EIA kit (Quidel Corporation) and serum carboxy-terminal telopeptide of collagen type I (CITP) was measured by an ELISA method (Orion Diagnostica) (151).

2.3.4 Bioinformatics analysis

Simple and multiple linear regression was performed in the statistical software environment R, and diagnostic plots produced for each regression, to show the regression plot itself, residuals vs fit, normal Q-Q, scale location, residuals vs leverage and influence. Plots for each regression were visually inspected in order to ascertain whether the assumptions of the method were violated. The R package leaps (152) was used to interrogate models via an all-subsets approach, using Bayesian information criterion (BIC; a measure of how well the data fits the model) and adjusted-R^2 as assessment criteria. Backwards stepwise selection on the variables identified with leaps was used to select the model with the greatest number of independently significant terms.

Two alternative approaches were used to analyze the data. The **first approach** (Fig. 2-2) involved an initial step of establishing a linear model using the clinical dataset. Each molecular dataset was then interrogated with univariate linear regression, multiple linear regression with confounders (age, sex and BMI) and multiple linear regression with the above-mentioned clinical model. Including BMI in models may seem to be redundant or a case of overfitting, however there is only a moderate correlation (Spearman's Rho: 0.56) between BMI and Body Surface Area – the variable used in this case to calculate LVMI.

Points of high influence (where Cook's distance > 4/n) were removed and a record was made of these removals so that those samples which had consistently high influence across datasets could be identified for removal in the modeling stage. All non-detections were treated as missing except NMR metabolomics data and CE-MS peptidomics data where zero values were deemed appropriate. Negative values in the metabolomics data were interpreted as noise and also set to zero.

Any variable passing this screening procedure with a Benjamini-Hochberg adjusted p value < 0.05 in any iteration was used as a potential predictor in the final modeling stage, thereby reducing the number of terms to analyze in the final modeling step. In the final modeling step two further models were generated – a molecular model and a mixed molecular-clinical model. In order to retain all variables and reduce the bias towards those of high detection data was imputed and linear models generated again for comparison. Missing points were imputed using the defaults of the Bioconductor package impute (153), using K-nearest neighbour or the mean value where appropriate (where missing data exceeds 50% in a variable).

The **second approach** (Fig. 2-2) uses PCA with varimax rotation, which results in principal components (PCs) which may each be described as mostly composed of a small subset of the input variables. The resulting varimax-rotated PCs were then tested with linear regression in the same manner as the molecular variables in the first approach. Samples were filtered prior to PCA for missing data (where >80% missing) and variables were filtered for low variance (< 0.1*mean). Remaining missing points were imputed as above.

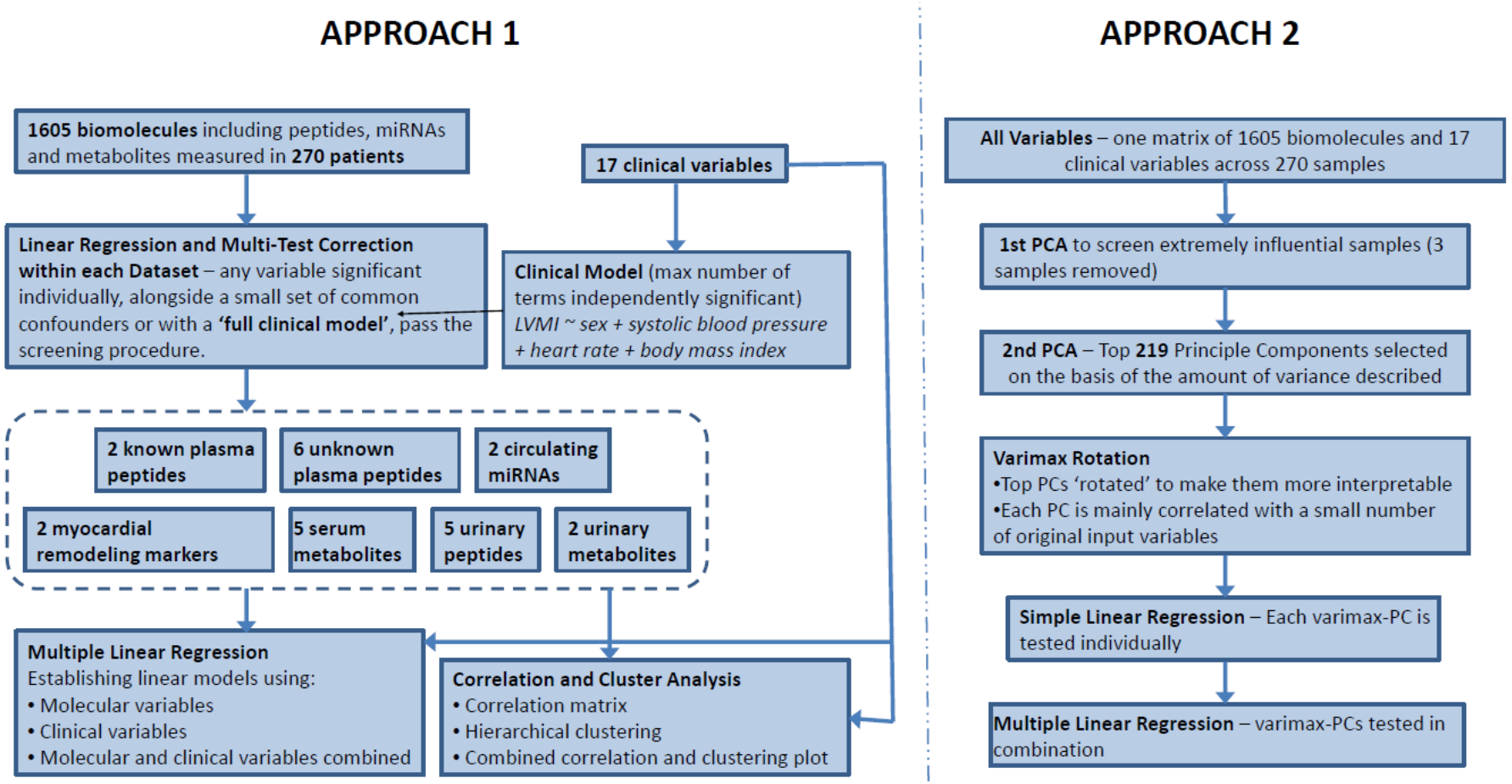

Fig. 2-2: The Two Approaches Used to Study LVMI
The first approach utilizes a linear regression based screening step to reduce dimensionality of the dataset prior to multiple linear regression, whereas the second relies on PCA to reduce dimensionality based on the amount of variance described. The other major difference is that where additional tests are required in the first approach to determine how predictor variables relate to each other, in the second approach the relationships between original input variables are inherently described in the resulting varimax-rotated principal components. PCA: Principal Components Analysis, LVMI: Left Ventricular Mass Index.

2.4 Multi-Omics Analysis of Respiratory Data

2.4.1 Design

The main aims of the COPD and Asthma Biomarker (CAB) study were to compare severe (treatment-resistant) asthma and COPD, and investigate the effect of smoking on those respiratory conditions, hence the six groups studied were all combinations of disease status (healthy, asthma, COPD) and smoking status (smoker, never-smoker). While this work focused on the omics subsets, the larger CAB cohort amounts to 220 patients. A dataset of mostly clinical and lab measurements (and associated metadata) on these patients contains 1,818 variables, however for many variables data is only available for a subset of the patients.

Induced sputum and nasal epithelium samples collected from subsets of the full cohort were used for 'omics' analyses. Sputum samples contain a varied mixture of cell types so our apparent 'differential expression' may come from the same mixture of cells undergoing 'true' differential expression or different cell types being recruited under different conditions. Correlation between each probe set and cell type were calculated in order to aid interpretation results in an empirical manner. The cell counts used were highly reproducible as shown by the close correlation between counts done on the same samples (Fig 2-3).

Enrichment of specific cell types in sputum have been associated with asthma (154), COPD(155) and smoking (156), and are defined by a selected cut-off percentile for the percentage of a particular cell type in the sample. For example often eosinophilia is defined as >3% eosinophils and neutrophilia is defined as >60% which can be seen in asthma and COPD respectively in Figure 2-4. A weakness of this classification is that a difference in the raw numbers of one cell type necessarily alters the percentage of another cell type even if their raw numbers remain steady. By the same token neutrophilia has been shown to mask eosinophilia (157), so it is unclear whether the COPD group truly exhibits less eosinophilia on average than the asthma group.

Both asthma and smoking have been implicated in epithelial shedding which explains increases in epithelial cells. With the large differences in cell type across groups we might expect to find some effects which are dependent on differential recruitment of cell types. Correlations between molecular abundance and the number of cells of a particular type may suggest which cells are involved in production of the molecule, either directly or by stimulating another cell.

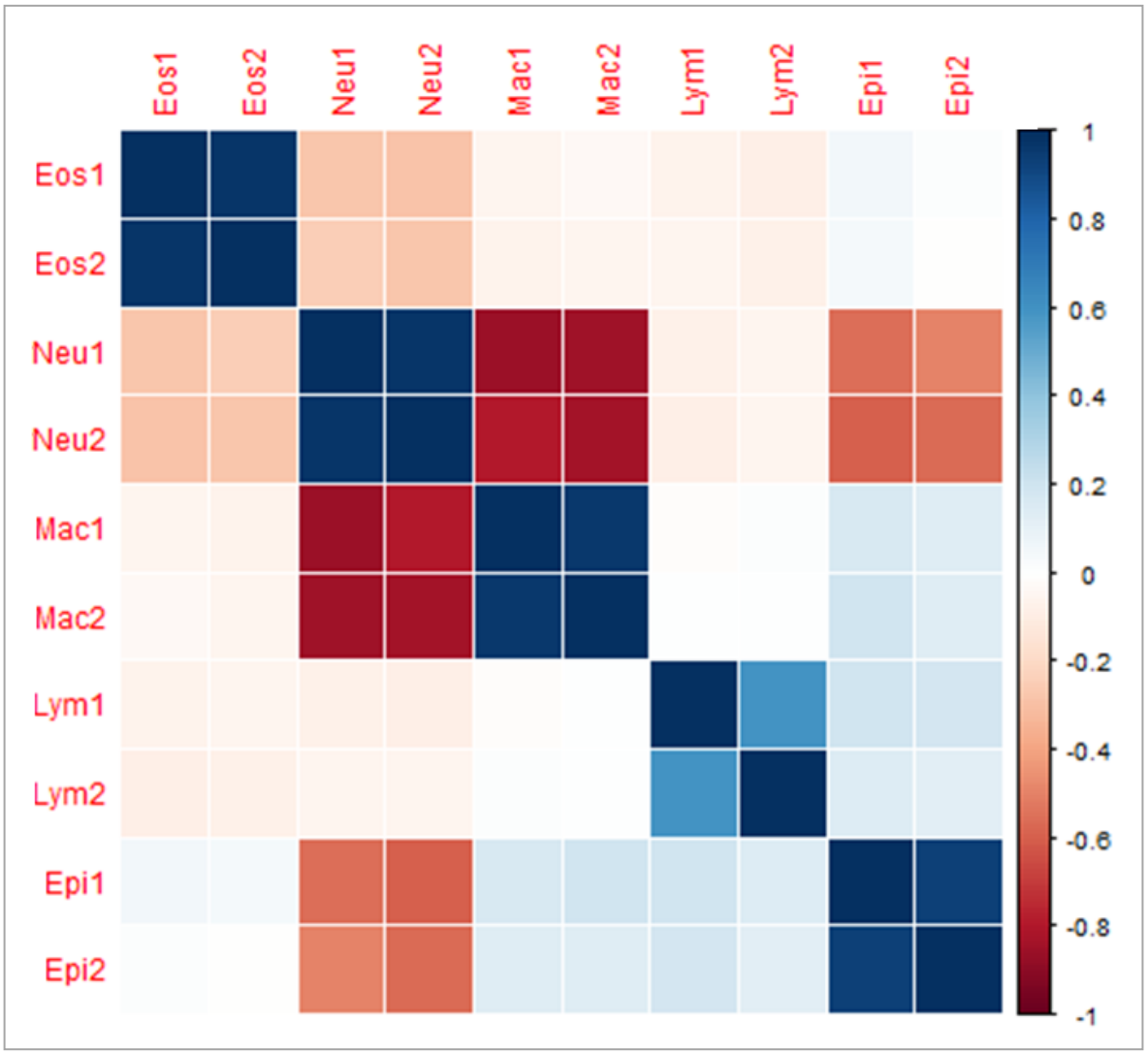

Fig. 2-3: Reproducibility of Cell Counts in CAB Dataset.
High correlation is found between iterations of counting each cell type in the same samples. Slightly less correlation is found between lymphocyte counts presumably due to v small counts of that cell type (as seen in Fig 2-4). Neutrophil cell count is in strong anti-correlation with macrophage cell count and moderate anti-correlation with epithelial cell count.

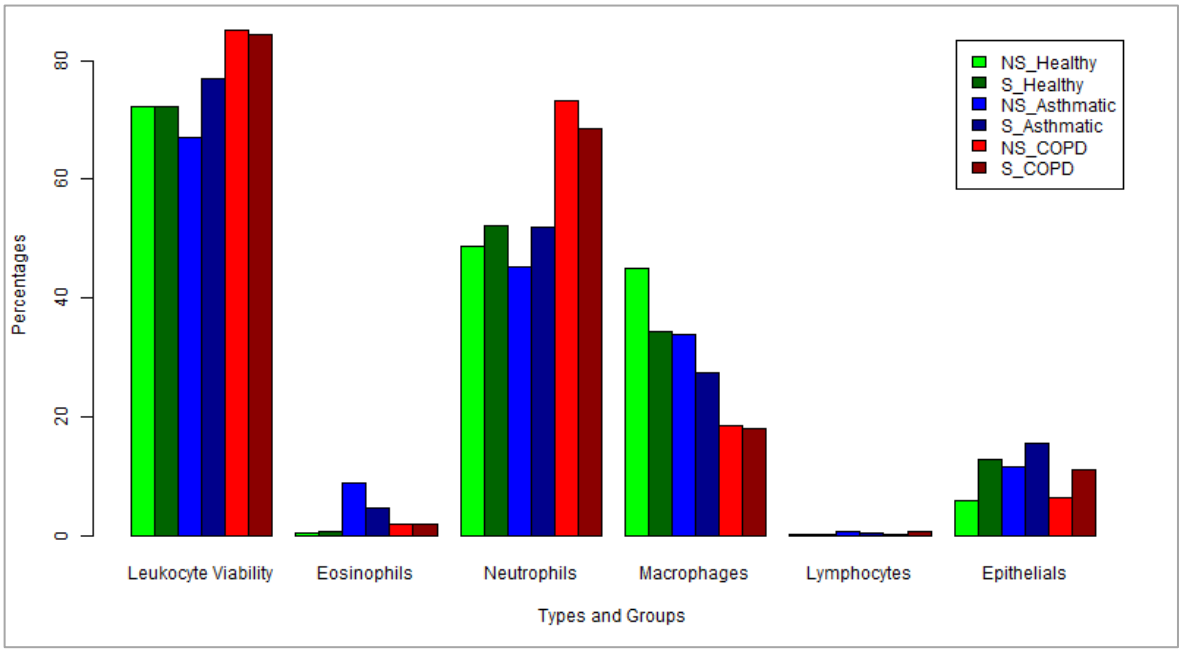

Fig. 2-4: Leukocyte Viability and Cell Type Counts of Induced Sputum of Smokers and Never-Smokers in Asthma, COPD, and Healthy Controls
Cell recruitment appears to occur with particular conditions, such as recruitment of eosinophils in asthmatics and recruitment of neutrophils in COPD patients.

Sex, age and BMI were compared between groups (Fig 2-5) to identify potential confounders. There is heterogeneity in that both males and females were used in the study however groups were quite well balanced by sex. There are various sex-specific differences with regards asthma, including more severe and frequent symptoms in boys vs girls, women vs men, and in particular women in a state of pregnancy or menopause. While our groups are well balanced for sex, due to different ages dictating different sex-responses there may be interaction effects between age and sex which will go un-modelled.

16 of the 100 patients used for omics-level analysis across the study had missing height or weight so BMI could not be calculated. Across the cohort there is a broad range of BMI from clinically underweight to obese in each. BMI is well balanced between groups, except for in smokers with COPD – a large proportion of whom were obese. Without statistical correction a comparison between this group and any other would be particularly prone to obesity-related false positives. Ages are extremely varied within and between groups.

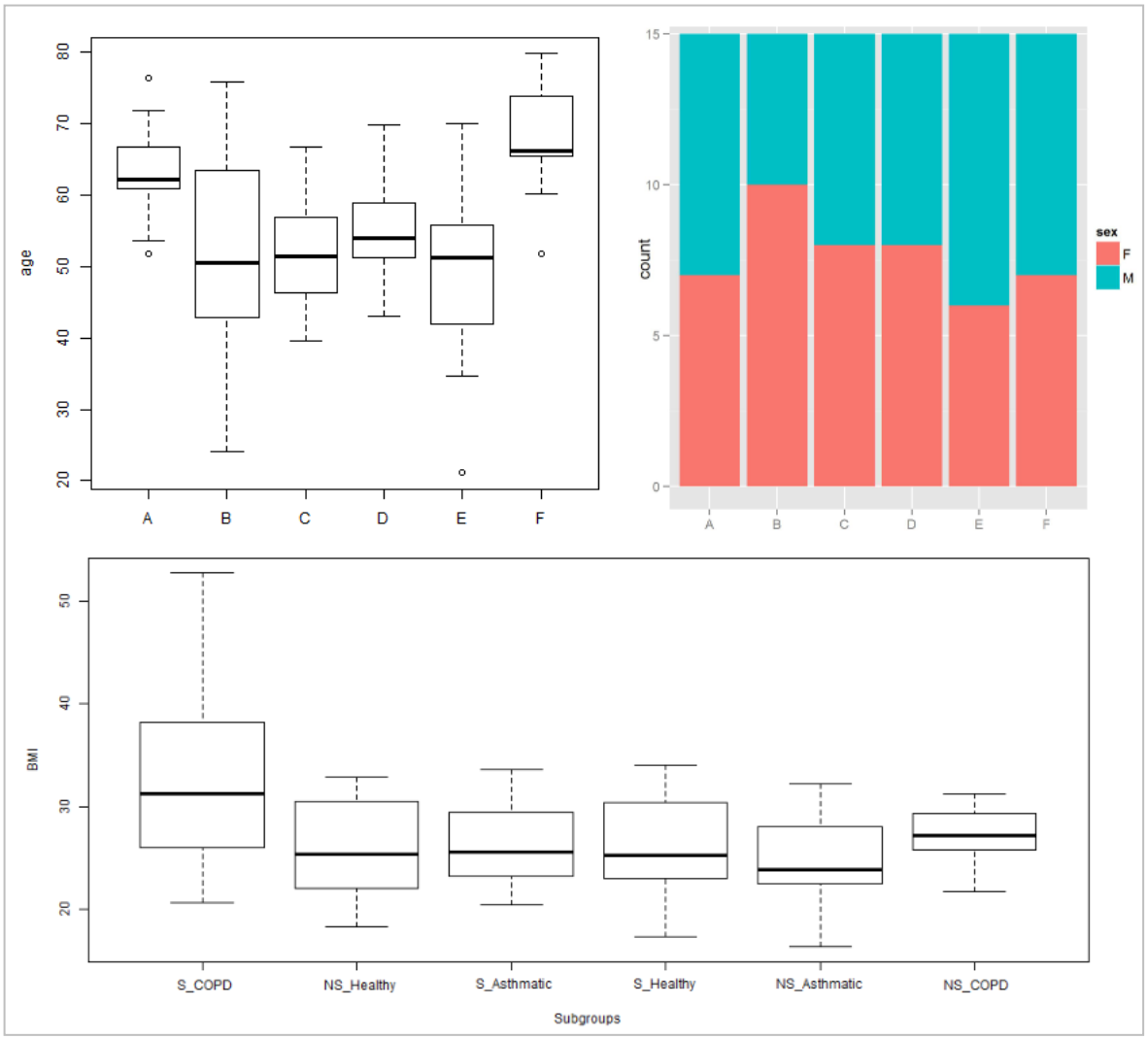

Fig. 2-5: Age, Sex and BMI of Groups.
BMI is poorly controlled in asthmatic non-smokers. Subset used to generate graphs was that used in the sputum microarray samples, however graphs of sets used in other comparisons are similar. NS: non-smoker, S: smoker. Groups A-F correspond to the order of subgroups in the BMI boxplot.

We might also expect obscuring effects from medications (Fig. 2-6), particularly those used for treatment of asthma or COPD. Some terminology was noted to be inconsistent - so any further text analysis should merge e.g. “tiotropium” and the brand name “Spiriva” - however there are too many relevant prescriptions to work into a useful statistical model given the number of samples. To compensate for this while still maintaining statistical power, analyses were repeated with a steroid dose normalised to beclomethasone. 13 of the 57 patients (of 100 ‘omics’ data patients) on steroids were missing a steroid dose value and for analysis the median value was substituted.

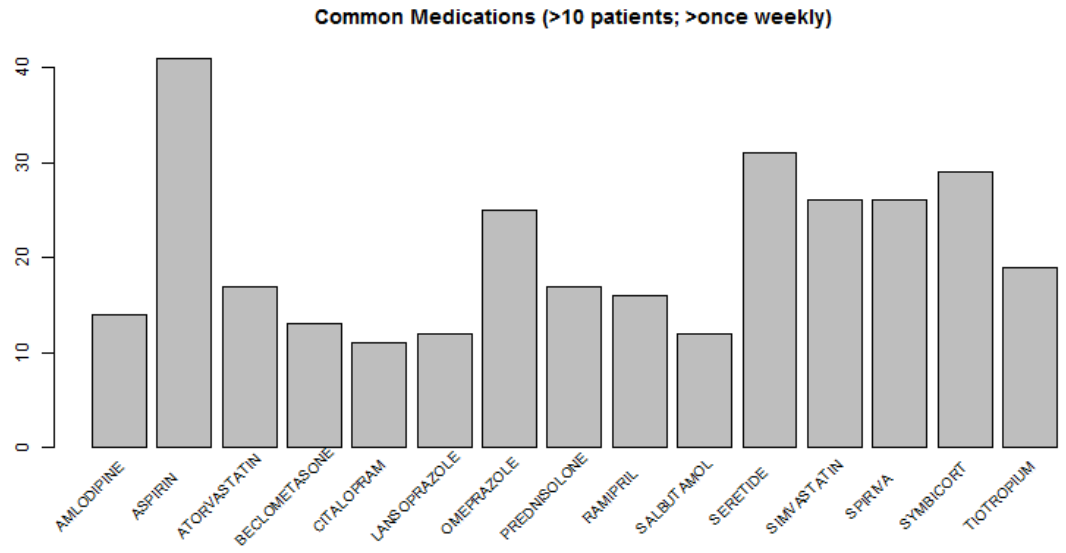

Fig. 2-6: All Prescription Medications Taken Once Weekly by at least Ten Patients. Some medications appear repeatedly in the database due to usage of brand names (e.g. “tiotropium”/“Spiriva”). Given the number of different medications, even after cleaning the data, all steroids were normalised to beclomethasone and other medications were excluded from analysis.

Microarrays for testing the genome-wide transcription of mRNA and miRNA were used for both the nasal epithelium samples and the induced sputum samples. The same 89 patients were tested in the mRNA and miRNA nasal epithelium microarray data, with an extra patient with only miRNA tested. A similar set of 90 patients was tested in induced sputum, with 80 patients at the intersection. Comparing information between files highlighted several typos with sample IDs used in the mRNA/miRNA work which were corrected. Proteomic and metabolomic data were only available for healthy and asthmatic induced sputum samples (not for COPD patients). 34 patients are found at the intersection of all four omics datasets (Fig 2-7).

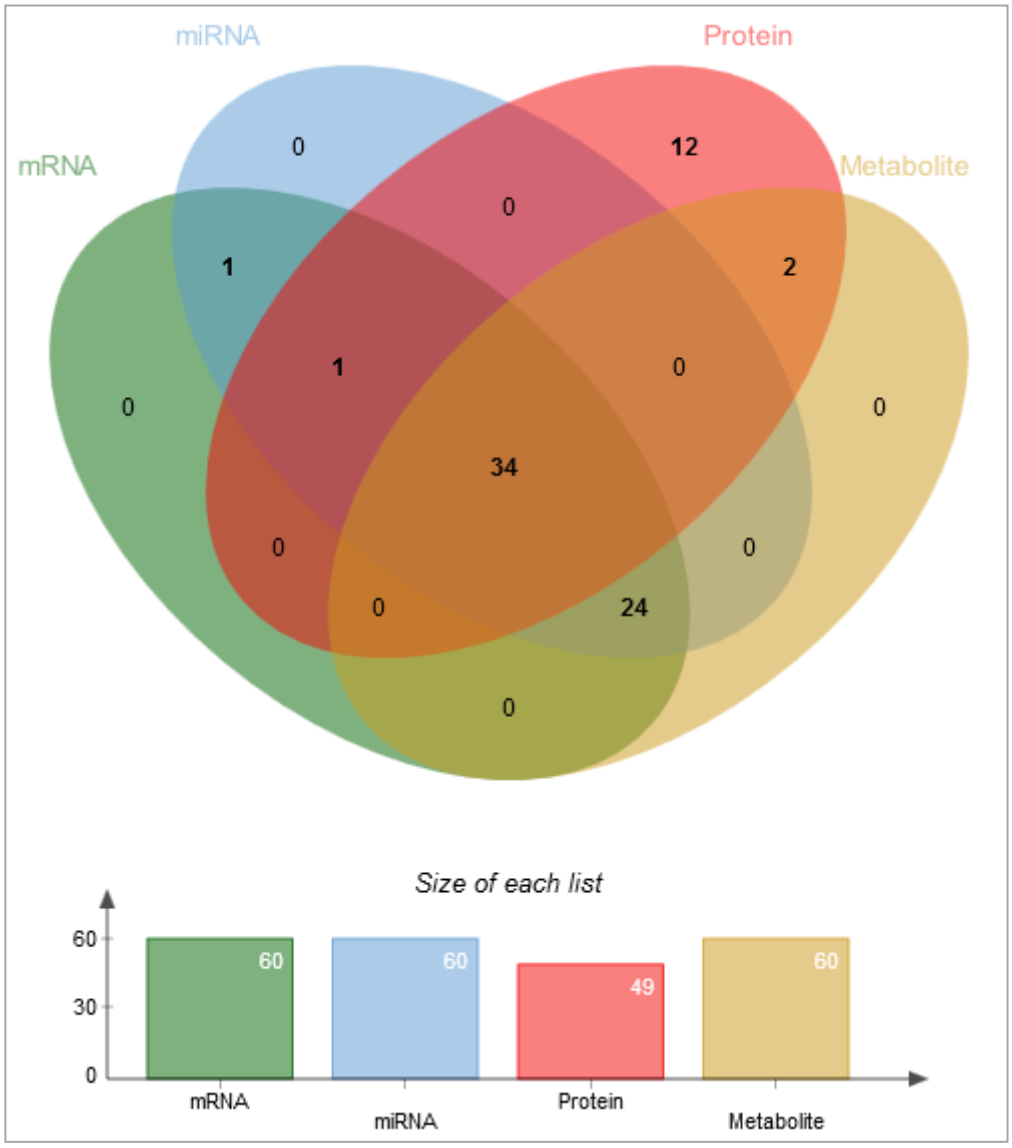

Fig. 2-7: Counts of Induced Sputum Samples from Asthmatic and Healthy Individuals in Microarray and Mass Spectrometry Datasets
All microarray samples in sputum work were shared, however 'mismatched' samples in the proteomics set and metabolomics set resulted in only 34 patients at the intersection of all four techniques.

2.4.2 mRNA Microarray

Genome-wide transcription in nasal epithelium and induced sputum samples was analysed using the Human Genome U133 Plus 2.0 Microarray from Affymetrix. Analysis was performed in R using various Bioconductor packages (158). The package simpleaffy (159) was used for CEL file parsing and affyPLM (160) used in conjunction with generic methods towards quality control (161;162). These methods resulted in various diagnostic plots including RNA degradation plots,

PCA clustering, normalised unscaled standard error (NUSE) plots, relative log expression (RLE) plots, box plots and correlation plots.

The mRNA microarrays were run in two batches over a year apart. 16 technical replicates of samples from the first batch were tested along with the second batch to assess the associated technical effect from e.g. sample degradation, subtle differences in application of protocol or equipment maintenance. Technical replicate pairs were shown to cluster closely (Fig. 2-8), having a median Manhattan distance 22.2% the magnitude of the median Manhattan distance of all other pair-wise combinations.

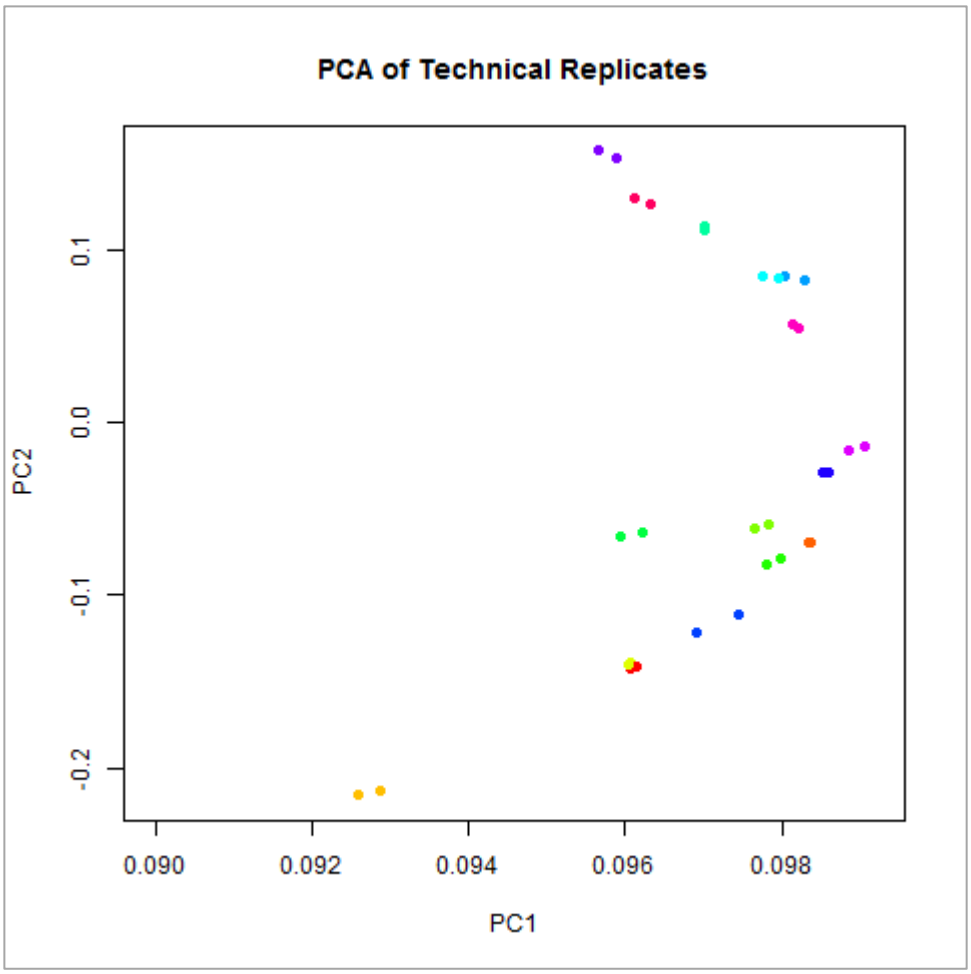

Fig. 2-8: Validation of Technical Replicates by PCA.
Close clustering of technical replicate pairs (indicated by colour) shows that a batch effect of testing samples at different times has not adversely affected analysis.

Spatial technical effects constitute areas of a microarray where intensity is particularly high – since the probes are randomly distributed across the array there should be no prevailing patterns. These were identified and replaced with "NA" using Harshlight, which relies on a combination of statistical methods and image-processing methods to identify both small localised 'blemishes' where every outlier probe is adjacent to one another, and larger more diffuse defects (163). Chips were then median-scaled and suspect values replaced by the median of the other values using a custom Java application. There is an option in the Harshlight package to replace technical effects with the median directly (Fig. 2-9) however this worsened or introduced new technical effects in some cases (Fig. 2-10). All images were examined and this improved protocol was used to remove technical effects.

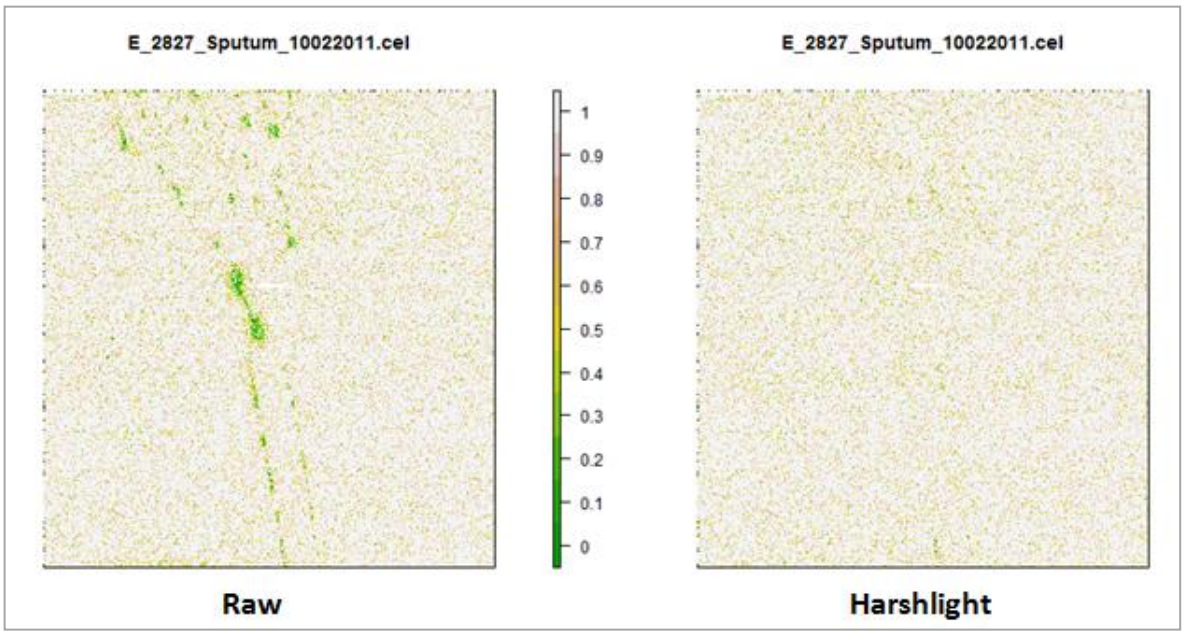

Fig. 2-9: Usage of R Package Harshlight Identifies and Removes Technical Effects.
In this case the option was selected to directly replace those affected areas with the median value on other arrays.

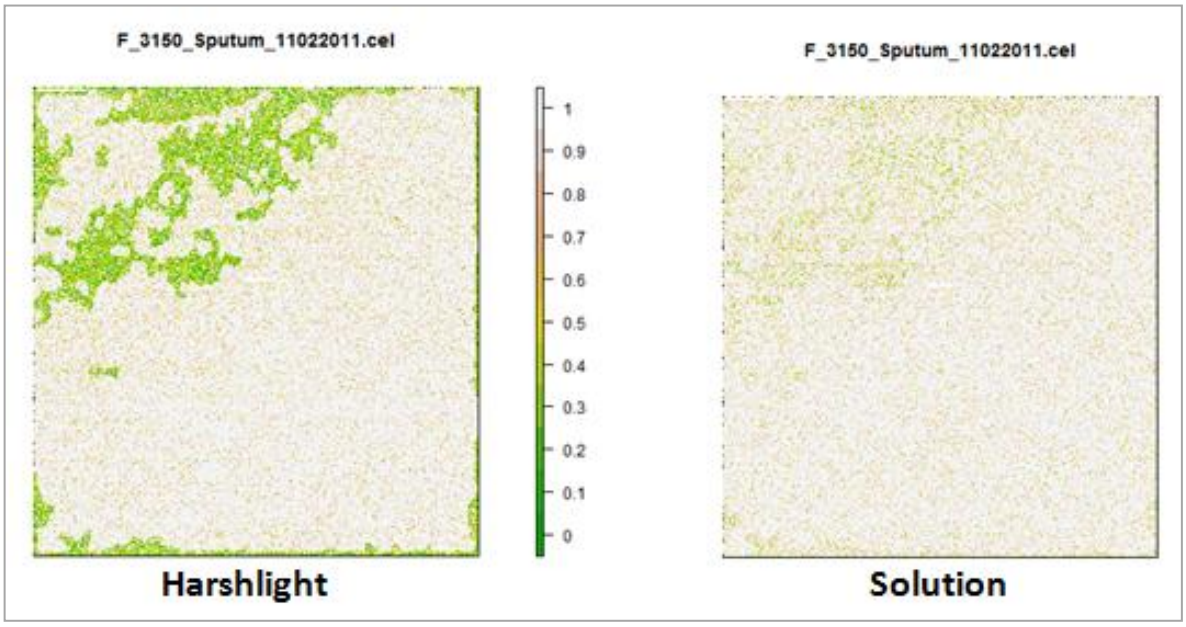

Fig. 2-10: Harshlight Was Used for Detection but Not Replacement.
Using Harshlight for the replacement makes some technical artefacts worse, so this was done separately with a custom solution written in Java, leaving Harshlight only responsible for detection.

CDF files describe the physical location of each probe on the microarray and how they are grouped into 'probe sets'. The HGU133Plus2_Hs_ENSG v18 chip definitions file (CDF) from BrainArray (164) was used for this analysis. Affymetrix's probe set selection relied on early genome and transcriptome annotation, significantly different from today's annotation. Brainarray is a project which periodically generates new CDFs based on current annotation. Affymetrix annotation has 54,613 probe sets, many of which cross-hybridise, hybridise the same gene multiple times (sometimes catching different transcripts) or target no known gene/transcript. Brainarray excludes cross-hybridising probes, and designs one probe set per gene (19,947 Ensembl Gene IDs; 58.6% of probes). To pick up alternative splicing events there is also a transcript-level grouping (86,103 Ensembl Transcript IDs; 58.2% of probes).

GCRMA(165) was used to quantile normalize, background correct and summarise from probes to probe sets. After pre-processing the nsFilter function from the R package genefilter (166) was used to screen the Affymetrix controls and target genes with low variance. After pre-processing, the nsFilter function from the R package genefilter (23) was used to screen the Affymetrix controls and 50% of target genes based on their having low variance. Y chromosome genes were also removed since the analysis is not restricted to men.

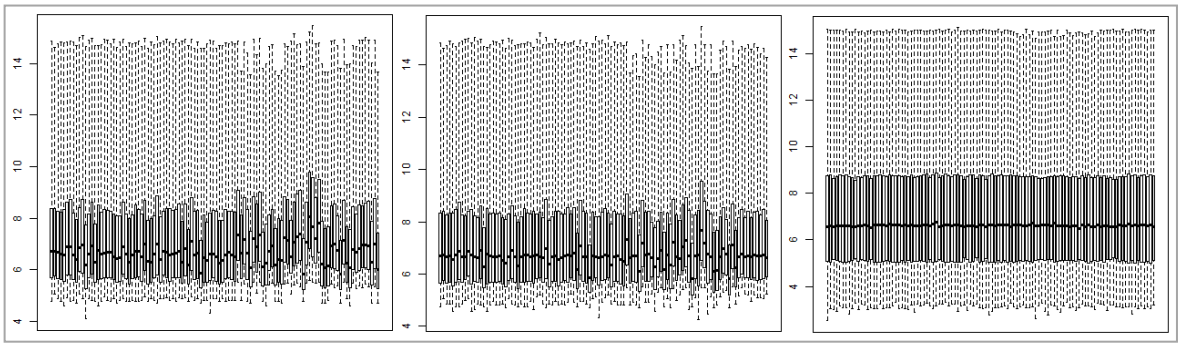

Fig. 2-11: Distributions of Sputum mRNA Before Normalisation, After Spatial Normalisation, and After Quantile Normalisation
Quantile normalisation adjusts expression distributions of each array, making them more uniform and accounting for technical effects.

Steroid dose was normalised to beclomethasone, with missing values (5/30) substituted with the mean dose. Statistics were generated from iterated linear regression, testing a model with asthma disease status and smoking status always included, and variations of this model with the addition of: age, sex, BMI or steroid dose.

Data was $\log_2$-transformed for input to limma as standard. Log-fold-change values given here are those reported by limma and are calculated as the difference between the means of the $\log_2$ values for each group rather than the $\log_2$-transformation of the fold change. The CEL files for the above-mentioned technical replicates were also included the limma analysis (167).

2.4.3 miRNA Microarray

The miRNA 2.0 Array from Affymetrix was used to assess the levels of miRNAs and other non-coding RNAs. The samples used pertained to the exact same group of patients as the mRNA microarrays, however no technical replicates were available. Spatial normalisation was not possible by the same procedure for this array type, though none required removal upon visual inspection of plots. The CEL files were processed and checked for quality control in a similar manner, but using the VSNRMA function of the VSN package instead of GCRMA for normalisation and summarisation. A custom CDF was created by merging those probe sets which are assigned to different species in the Affymetrix annotation

but contain exactly identical probes. Other non-human probes were removed except for viral miRNAs which are distinctly different from human miRNAs and may be interesting as an indicator of infection. Data was filtered for low variance as with the mRNA microarrays. Screening, modelling and significance-testing were all done as with the mRNA microarray analysis.

With the miRNA chip each ‘probe set’ consists of repeats of the same probe as mature miRs are too short to require multiple probes. Some mature miR sequences from different human miRNA genes are identical. The sequence of some miRs is conserved entirely between several species. In total 3245/6703 miRs on the chip are redundant. All non-unique probe sets were merged in a custom CDF, also removing all non-human species entries except for a few viral, human-infecting species (potentially useful as measures of infection).

2.4.4 Proteomics

Proteomics data were provided in two files which used different scales. A reference protein present in both datasets was used to appropriately scale the data so it could all be included. Protein IDs were mapped from the defunct IPI using the internal CluSO database. For the purposes of generating a summary value, protein quantities were summarised to the gene-level however the data for individual peptides were used for statistical testing so that alternative splicing and protein modifications could also be identified. MWW was used for significance testing, taking zero values as an equal bottom-rank so as not to exclude results where a protein is detected only in one group (but treating them as truly missing for calculation of summary statistics such as the mean of a group).

2.4.5 Metabolomics

Metabolites were extracted from 60 airway fluid samples using a standard organic phase extraction protocol. LC/MS analysis was performed using Exactive/Qexactive mass spectrometer and pHilic column. 200 standard metabolites were used for identification as described by the MSI (metabolomics standard initiative). IDEOM provides Students t-test results however it does not do multiple test adjustment on those figures and as the documentation itself points out the t-test is often not the optimal statistical method – often the t-test is not appropriate for large portions of these types of datasets due to non-normality or heteroschedasticity (168). As such, MWW was used on output from IDEOM, as in proteomics. Statistics were calculated for all three asthma contrasts and multi-test adjusted as per the other omics data sets. Unfortunately different mediums were used for different samples which may have affected data quality.

2.4.6 Downstream Analysis

The p values from all experiments were Benjamini-Hochberg multi-test corrected (by dataset) and an alpha of 0.05 was selected for all. Variables were (Spearman-) correlated with cell count percentages of eosinophils, neutrophils, macrophages, lymphocytes and epithelial cells. For the transcriptomic and proteomic data ClueGO (169) and CluePedia (170) were used (with default settings) for term enrichment analysis utilizing terms from gene ontology, KEGG, WikiPathways and Reactome.

KEGG Mapper (171) and PathVisio (172) were used for pathway analysis. Relationships between miRNA and transcript/protein levels were scored using CoMir, where all those with probability >90% were included. Additional relationships between significant variables were also extracted from GeneMania (173) and MiMI (174).

Integration of different omics using various databases:

- ITFP (transcription factors targets)
- CoMir and MirTarBase (predicted and validated miR targets)
- MiMI & GeneMania (protein-protein interactions)
- KEGG & WikiPathways (protein-protein, protein-metabolite)
- Allergy and Asthma Portal

These analyses were done for several contrasts ('all asthmatics vs all non-asthmatics', 'asthmatic non-smokers vs healthy non-smokers', and 'asthmatic smokers vs healthy smokers'), thus focusing on asthmatic status with different smoking backgrounds. The results were combined into a large asthma network, where all nodes have been found significant in at least one contrast and are joined by the edges gathered from the databases listed above.

Different colour schemes were attributed onto the network to denote up/down-regulation and cell type correlation. Comparing with different fold changes set against a large combined map is more informative than comparing gene sets, as it can be seen how different results are (e.g. by fold change), not simply whether they pass the selected statistical criterion.

Another version was produced extending out to also include non-significant but biologically consistent networks (assessed by inspecting relationship and fold change for each node and edge). This allowed the inclusion of some metabolomic data, giving it context in terms of the other results and bridge gaps where data is simply unavailable due to technological limitations.

3. Using SNP Data to Facilitate Enhanced Interpretation of Genome-Wide DNA Methylation Data in Essential Hypertension

3.1 Introduction

3.1.1 Heritability of Essential Hypertension

Hypertension is the leading world-wide mortality risk factor and is the fifth greatest risk factor for disability-adjusted life years (175). Although there are rare monogenic forms of hypertension it is usually a multifactorial disease – so-called essential hypertension. Hypertension is associated with modification of the vasculature and is itself an influential risk factor for several cardiovascular diseases such as stroke, heart disease and kidney failure (176;177). Many environmental and life-style risk factors for hypertension are well established such as high salt diets, high alcohol consumption, lack of exercise and obesity. Using familial studies the genetic component of essential hypertension has been estimated at between approximately 30% and 70% (59). So far, however, less than 2% of variation of blood pressure in the general population is described by known genetic variants (61). Incomplete concordance of essential hypertension between monozygotic twins reported in the range of 38%-52% (178;179) further indicates that epigenetic factors may be involved.

3.1.2 DNA Methylation and Gene Expression

DNA methylation is an epigenetic modification which is prone to occur at sites in the genome where a cytosine is followed by a guanine, and affects about 70-80% of these 'CpG' sites in the genome, however CpG sites are depleted throughout most of the genome, being found mostly in ~1kb regions called CpG islands (CGIs) often within or close to genes and regulatory elements (180). CpG sequences are palindromes and may be methylated *de novo* (by DNMT3a and

DNMT3b), or methylation may be maintained by the action of an enzyme (DNMT1) which recognises hemimethylated DNA after somatic cell division. DNA methyltransferase enzymes are required both for this methylation pattern maintenance and for *de novo* methylation and in the latter case are influenced by histone modifications (181;182), whereas demethylation occurs by base excision and by lack of maintenance methylation (183).

It has long been held that DNA methylation can silence gene expression, and that the presence of methylated CpGs may be interpreted by proteins with methyl-binding domains (184;185). However, with the recent advances of high throughput technologies the picture of how these epigenetic modifications operate is growing more complex. Methylation of a CGI at the transcription start site (TSS) appears to either 'lock in' a silenced state (186) or direct that state to occur by other mechanisms (187), and is usually found in the context of stable, long-term silencing (e.g. X-inactivation, imprinting). Methylation at shores (2kb island-flanking the islands) and shelves (2kb beyond shores) have been shown to have different levels of methylation, and have also been associated with differential expression (188;189). Non-CGI TSS methylation is more dynamic and, although it is less well studied, it is also associated with silencing (190).

Methylation of gene bodies, on the other hand, has been shown to be positively correlated with transcription (191), possibly via repression of intragenic transcription initiation (192) and may influence alternative splicing (30;31). One must also remember that often genes may have additional, downstream, TSSs so that a region may be considered to be both a TSS and a gene body and that DNA methylation could be considered a mechanism for alternative promoter selection. Methylation also been shown to inhibit the activity of enhancers (193) and insulator methylation appears to either be cause or effect of insulator activity (194;195).

3.1.3 DNA Methylation Microarrays in Hypertension and the Integration of SNP Data

There are a multitude of available technologies for assessing DNA methylation (196). The technology selected for this study was Illumina's Infinium HumanMethylation450 BeadChip which is a microarray which interrogates over 485,000 CpGs throughout the genome. It is one of many technologies which rely on treating the sample DNA with sodium bisulphite which converts methylated cytosine residues into uracil residues. While this is a commonly-used technique it does have a drawback in that it also converts hydroxymethylated cytosine residues into uracil, i.e. some apparent DNA methylation results will instead be DNA hydroxymethylation which is a less common and less well-studied epigenetic mark (197). Another issue is that SNPs within the probe sequences could interfere with binding and also generate false positives, providing further challenges to interpretation of results (198).

A decreased level of global methylation has been associated with hypertension in whole blood DNA in one study (199), whereas another study has found results to the contrary in repeat regions of DNA extracted from leukocytes (200). A similar lack of consensus can be seen in placental DNA of women with preeclampsia (201;202). These apparent disparities may in some cases merely be due to the fact that different technologies were used in different studies, and so different sets of CpGs were being targeted e.g. some target repetitive regions or CGIs at promoters and others operate in a more truly global sense. Alternatively, or in addition to this, it might be concluded that DNA methylation in hypertension is a gene, gene region, or CGI-specific phenomenon, rather than a global one, and so any differences in global DNA methylation related to hypertension may be either spurious or determined by the combination of gene-specific differences inherent to that group (203).

There have been many studies on the epigenetics of hypertension (204), however there have only been two genome-wide DNA methylation studies on essential hypertension, and two studies which instead focussed on blood pressure as a continuous variable. Wang et al used Illumina's HumanMethylation27 BeadChip (which targets 27,000 CpGs) to compare whole blood samples from 8

hypertensive patients and 8 age-matched controls, all of whom were male (205). The study did not find any statistically significant results after multi-test correction, however the most significant CpG (in gene *SULF1*) showed significance after correction in a meta-analysis of two replication cohorts, along with an additional CpG found within the same gene. These results remained significant after adjusting for age but not when adjusting for both age and BMI. Wang et al suggest that SULF1 may be influential in hypertension by regulating inflammation via IL6 production.

The other hypertension study had a much larger sample size - 126 hypertensive patients out of a total of 712 in the cohort (206). Patients of European and South Asian ethnicity were recruited and analyses were done both within these groups and across all patients. Results were presented both unadjusted for confounders, and adjusted for a large number of variables: age, BMI, smoking status, social class, estimated cell counts and batch effects. There are two significant CpGs in the trans-ancestry analysis, found in genes *PPP1R2* and *LOC100132354*, however their significance markedly drops (beyond the significance threshold) upon addition of the confounders to the model. *PPP1R2* encodes a protein which binds serine/threonine phosphotase PP1, strongly inhibiting its activity. PP1 has hundreds of potential targets but the target with most relevance is $nNOS^{ser852}$ which in the vascular endothelium is dephosphorylated by PP1 leading to subsequent NO generation and vasodilation (207). In addition the activity of PP1 can be inhibited by oxidative inactivation by *NOX4* which is implicated in hypertension pathogenesis (208). *LOC100132354* is a long non-coding RNA which has been shown to promote angiogenesis which can affect blood pressure (209).

In this study blood pressure was also analysed as a continuous variable. Unadjusted for confounders there were four CpGs associated with SBP (in genes *FHL2, MYO5C, ELOVL2, KLF14*) and two associated with DBP (in *AHRR, MYO1G)*. With adjustment none are significant however one distinct CpG is significant in the presence of confounders: cg07598370 near *OR5AP2*. This gene encodes an olfactory receptor which may influence blood pressure via renal expression, as other olfactory receptors have been shown to do (210). There were three CpGs associated with DBP in the European subset, unadjusted for confounders. Two of which remained after adjustment and a third additional CpG was also associated

in this context. A CpG was also associated with SBP in the South Asian subset. In addition to analysing CpGs individually, they were also analysed as agglomerated into differentially methylated regions by the R package DMRcate. This analysis identified 395 differentially methylated regions (mapped to 326 genes) which were associated with SBP, and 237 differentially methylated regions (mapped to 157 genes) which were associated with DBP. Pathway analysis of differentially methylated regions indicated Notch signalling in relation to SBP in the European subset and insulin-like growth factor-2 mRNA binding proteins in relation to DBP in the South Asian subset.

Boström et al studied DNA methylation in relation to SBP in 11 obese patients undergoing gastric surgery, and found 24 differentially methylated CpGs correlated to changes in SBP between before and after the surgery (211). A replication cohort showed that two of these CpGs (in genes EHMT2 and SKOR2) were significant independent of age, BMI, ethnicity and sex, although the statistics in the replication cohort instead addressed the categorical variable hypertension. Boström et al suggest that EHMT2 may be involved in hypertension via regulation of the pro-inflammatory cytokine IL-17.

Richard et al also focussed on blood pressure rather than hypertension, devising a two-stage meta-analysis using a total of 17,010 patients (212). Of 31 discovery stage CpGs (found significant in a model along with age, sex and technical covariates) 13 were replicated. These 13 sites are heritable ($h^2 > 30\%$) and independent of known BP genetic variants. 4 of these were found to have one or several cis-located genes (*TSPAN2*, *SLC7A11*, *UNC93B1*, *CPT1A*, *PTMS*, and *LPCAT3*) whose expression was associated with both CpG methylation and SBP, DBP or hypertension. In particular Richard et al describe *TSPAN2* as a candidate gene for blood pressure which is regulated by heritable DNA methylation. *TSPAN2* was found to have the strongest associations with methylation level and blood pressure out of all transcripts tested, and it is significantly associated with SBP, DBP and hypertension. *TSPAN2* is highly expressed in the vascular tissues and two different SNPs in the gene have been associated with large artery atherosclerosis-related stroke (213), and migraine (214) – which can be driven by changes in the vasculature and is itself a risk factor for cerebrovascular and cardiovascular disease.

Genome-wide differential methylation studies in related conditions such as preeclampsia may also prove to be insightful. Using the NimbleGen 'Human CpG Island Plus Promoter' microarray (385K sites in promoter regions) Jia et al found 296 genes significantly differentially methylated in the placentas of patients with preeclampsia. The greatest overrepresentation of these genes was in chromosome 1 (10.5%, P=0.005) (215).

Epigenetics of hypertension has also been studied with a candidate gene approach, specifically analysing genes which are already implicated. Many associations between hypertension and differential methylation at these genes have been detected, particularly related to pathways and processes central to our current understanding of blood pressure control (e.g. renin–angiotensin-system genes *ACE* and *AGTR1*) (216). There is even strong evidence indicating that fine particulate matter air pollution can induce *ACE* DNA methylation, ACE expression, and blood pressure elevation (217).

While the candidate gene approach has been successful, epigenome-wide studies have identified genes which had not previously been linked with hypertension at any level, and perhaps reflect the heterogeneity of hypertension. In addition to furthering biological understanding of the molecular mechanisms of blood pressure control, epigenomic data may be important to fill a gap in knowledge moving towards personalised medicine – predicting predisposition to disease or to a particular treatment, in combination with other data such as SNPs.

The integration of SNP data regarding the same cohort facilitates enhanced interpretation of the DNA methylation data, by identifying those results which are likely to be false positives due to SNPs. Where data is unavailable for a particular SNP, allele frequencies integrated from SNPDB can at least give an estimation of how common the SNP is in a general population. This aids in the interpretation of the DNA methylation data, and the selection of DM sites/regions for verification. Another possibility is to run an analysis with sites with all proximate common SNPs (e.g. within 10bp of CpGs) excluded altogether. In addition to integrating SNP data, one could also integrate transcriptome data to show the effect on genes proximate to the significant differential methylations, however this data was unavailable.

3.2 Results

3.2.1 Data Pre-Processing

The data was initially filtered by missing data and detection p values using the defaults of the R package IMA. The quality of the samples was assessed by the number of sites whose detection p value exceeds 1 X 10^{-05}. All samples have less than 0.08% sites exceeding this. Had any samples exceeded 75% of sites exceeding the cut-off they would have warranted exclusion, i.e. quality is good and no samples were rejected. A similar procedure was carried out checking for any sites with high rates (again 75%) of large detection p values (>0.05). Finally, those sites with missing values were removed.

Distributions are similar across samples both in shape and median values (Fig. 3-1), with a slight left skew. Similar medians between samples are to be expected as the same amount of DNA from each sample is used in the lab protocol. The distributions should also be similar, assuming a relatively small proportion of the methylome is altered in certain clinical groups (while comparing samples of the same type). No distribution stands out in particular in terms of quality control.

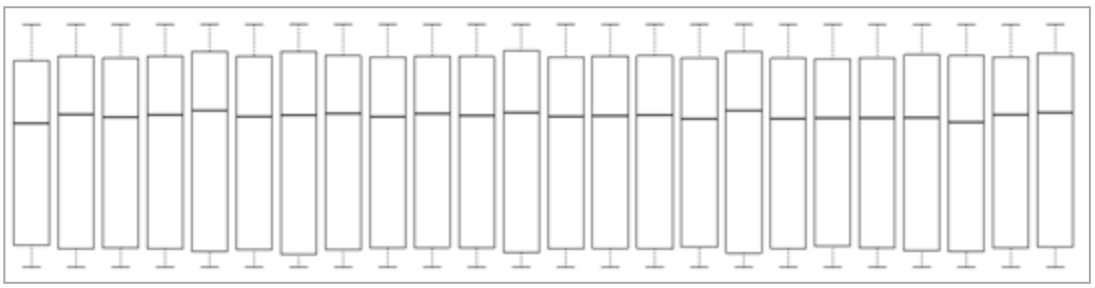

Fig. 3-1: Box Plot Showing Distributions for each Sample.
Samples have similar medians and distributions as would be expected – no need to remove any samples.

The cleaning of the data can be summarised as:

- 0 samples removed with at least 75% of sites having detection p value greater than 1 X 10^{-05}
- 12 sites had at least 75% samples with detection p value greater than 0.05 and were removed
- 23,709 sites contain missing values and were removed
- 461,856 sites were retained from the original 485,577 sites

The remaining values were peak-normalised, logit-transformed and finally grouped into gene regions and CGI regions by definitions from UCSC provided in the official Illumina annotation, resulting in 206,326 regions. Peak normalisation was effective at normalising signals between the Illumina Infinium I probe type and Infinium II probe type (Figures 3-2 and 3-3), which are both used on this microarray. Both probe types follow binomial distributions in each sample – one peak for those mostly methylated sites and another for the mostly unmethylated sites. The differences in beta between the peaks of type I vs type II probes are artefacts of differences in probe design; the differences represent a technical effect to be corrected.

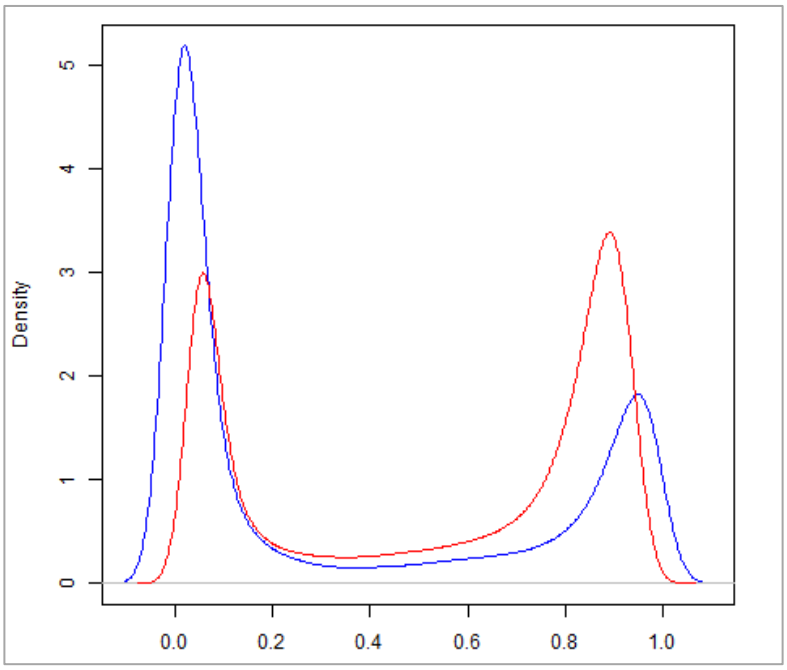

Fig. 3-2: Density Plot of Beta in one Sample ('948.151') Before Peak Normalization.
Each line represents one of the two probe types used on the array. The offsets of the peaks represent a technical effect to be removed prior to statistical analysis.

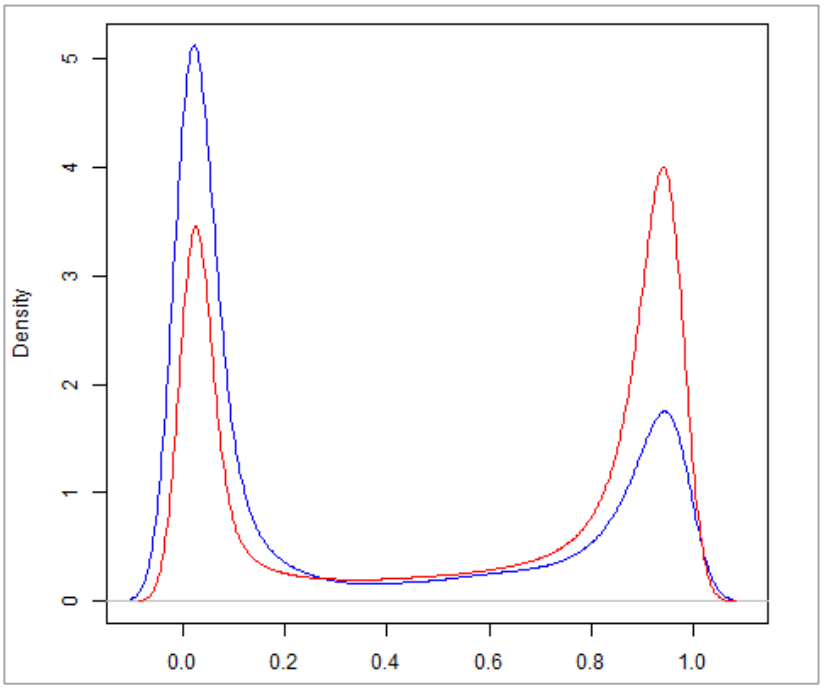

Fig. 3-3: Density Plot of Beta in one Sample ('948.151') After Peak Normalization.
Each line represents one of the two probe types used on the array. The data has been corrected for the effect of the different probe types used on the array, which can be seen by the alignment of the peaks.

3.2.2 Sites

DM of CpG sites within the genes *PAX7*, *NADK*, *IFT140* and *HLA-G* were found to be significantly associated with genetic risk (p_{BH}=4.11 X 10^{-3}, 0.01, 0.02, 0.05; Table 3-1, Figure 3-4). The same *NADK* and *PAX7* sites were found to be borderline-associated (p_{BH} = 0.05) with discordant controls. The *NADK* site ('cg27433479') has the highest change in beta value of 0.57 between discordant controls and others. It is one of 28 CpGs in the gene body of *NADK* which is found to be significantly differentially methylated with respect to genetic risk ($\Delta\beta$=0.02, p_{BH}=0.02).

By merging these results with data from dbSNP we find that some of the significant CpGs overlap with locations of SNPs, which may interfere with probe binding and create misleading results (Table 3-2). In particular SNPs associated with cg27433479 and cg26544072 have large minor allele frequencies – indicating a commonly-occurring variant in the general population.

Site	Gene	CGI	Δβ	p	p_{BH}	Contrast
cg27433479	NADK	chr1:1685373-1685971	-0.57	2.01×10^{-7}	5.03×10^{-2}	Discordant Control
cg27433479	NADK	chr1:1685373-1685971	0.02	1.30×10^{-7}	1.99×10^{-2}	Genetic Risk
cg26544072	HLA-G	chr6:29795553-29796594	-0.01	4.29×10^{-7}	4.95×10^{-2}	Genetic Risk
cg00065215	PAX7	chr1:18956895-18959829	0.01	8.89×10^{-9}	4.11×10^{-3}	Genetic Risk
cg00427553	IFT140	chr16:1604964-1605345	<0.01	5.05×10^{-8}	1.17×10^{-2}	Genetic Risk
cg00065215	PAX7	chr1:18956895-18959829	<0.01	2.18×10^{-7}	5.03×10^{-2}	Genetic Risk

Table 3-1: Details of Significant Results Across all CpG sites in all the Contrasts studied.
Gene and CGI occupied by CpG is shown as well as statistics and the contrast to which they relate. All sites were found in the "body" regions of their respective genes. cg27433479 was in the south shore region of CGI 'chr1:1685373-1685971' and all other CpGs were in the island regions of their respective CGI. Δβ: change in beta, p_{BH}: Benjamini-Hochberg adjusted p values.

site	Gene	SNPs	SNP Distance	MAF
cg27433479	NADK	rs2076328	1	0.36
cg26544072	HLA-G	rs1130355;rs41555713	1;37	0.50;<0.01
cg00427553	IFT140	rs143047330	5	<0.01

Table 3-2: SNPs Documented in dbSNP which may Interfere with Probe Binding and Result in a False Methylation Signal.
'SNP distance' gives the distance to the nearest SNP in number of bases. MAF: Minor Allele Frequency, SNP: Single Nucleotide Polymorphism.

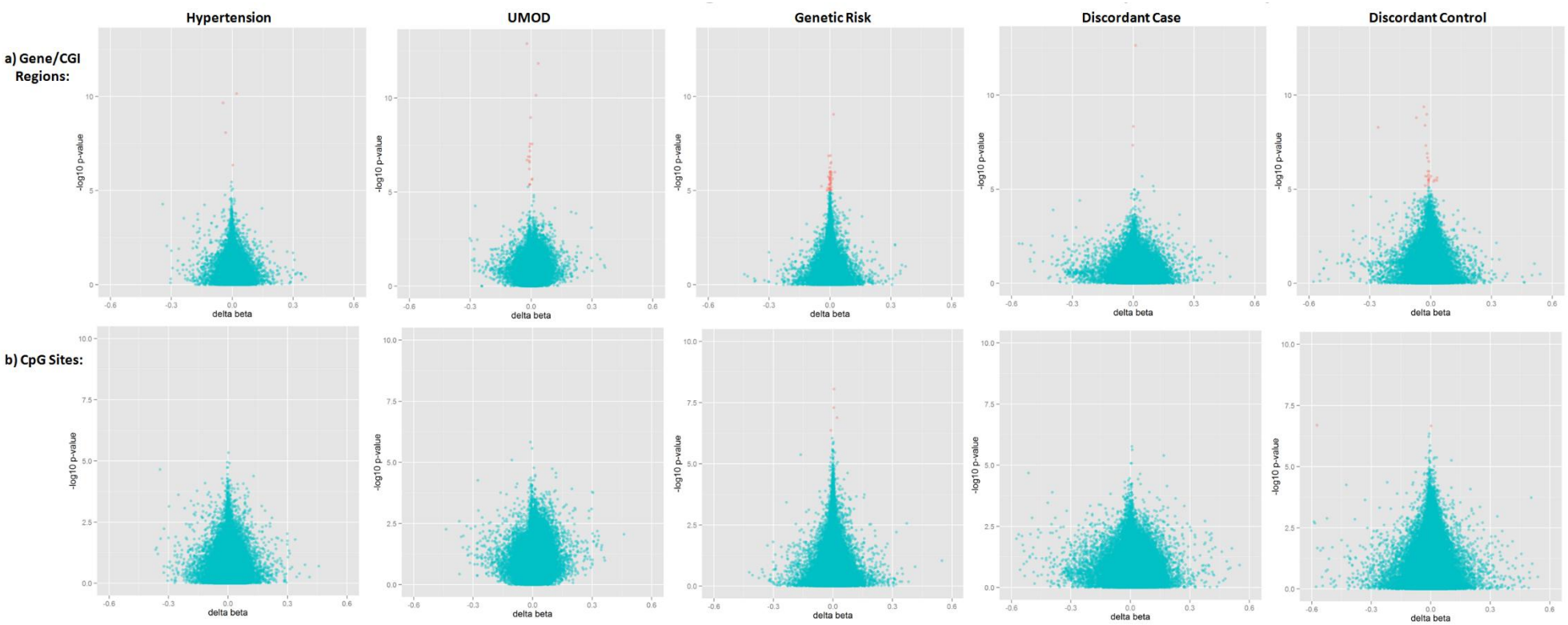

Fig. 3-4: Volcano Plots of Sites and Regions of Differential Methylation by Contrast.
p_{BH} values shown in log scale. Those with $p_{BH} < 0.05$ are highlighted red. Deviation from the centre along the horizontal axis shows the associated increase or decrease in beta.

3.2.3 Regions

Significant DM was found with all contrasts on a region level and the intersections of the genes and CGIs involved are shown in the form of a Venn diagram in Figure 3-5. The total significant hits for each contrast were: genetic risk = 48; hypertension = 4; *UMOD* = 19; discordant cases = 4; discordant controls = 27, although some genes and CGIs have DM in several regions so are listed multiple times within these lists. Almost half of the significant gene regions are located proximate to the transcription start site (Fig. 3-6). Approximately 2/3 of CGI-related regions were within the island themselves, and more regions were found to be demethylated compared with other samples. 38 unique genes were associated with the gene regions results and 31 with the CGI regions results, with an intersection of 15 between the sets and 54 genes represented in total. Reanalysing after filtering out CpGs with SNPs <10bp away reduces this list to 32. Chromosome 1 was overrepresented in the list of 54 genes (10.11%, p=0.04).

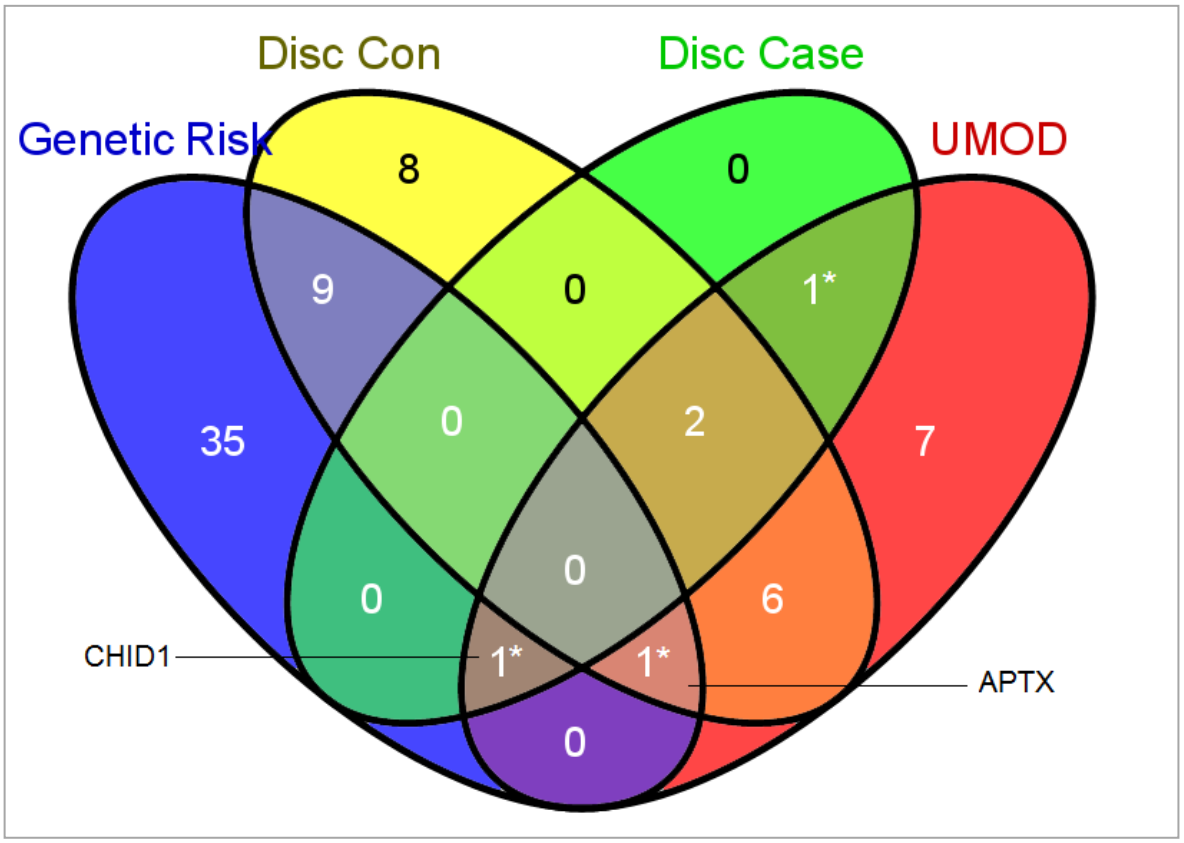

Fig. 3-5: Venn Diagrams Showing Intersections of Significant Hits from Various Contrasts in Terms of Whole Genes or CGIs, as Detected in Regions.
Genes identified in the hypertension contrast are designated by an asterisk to reduce the complexity of a fifth set in the diagram.

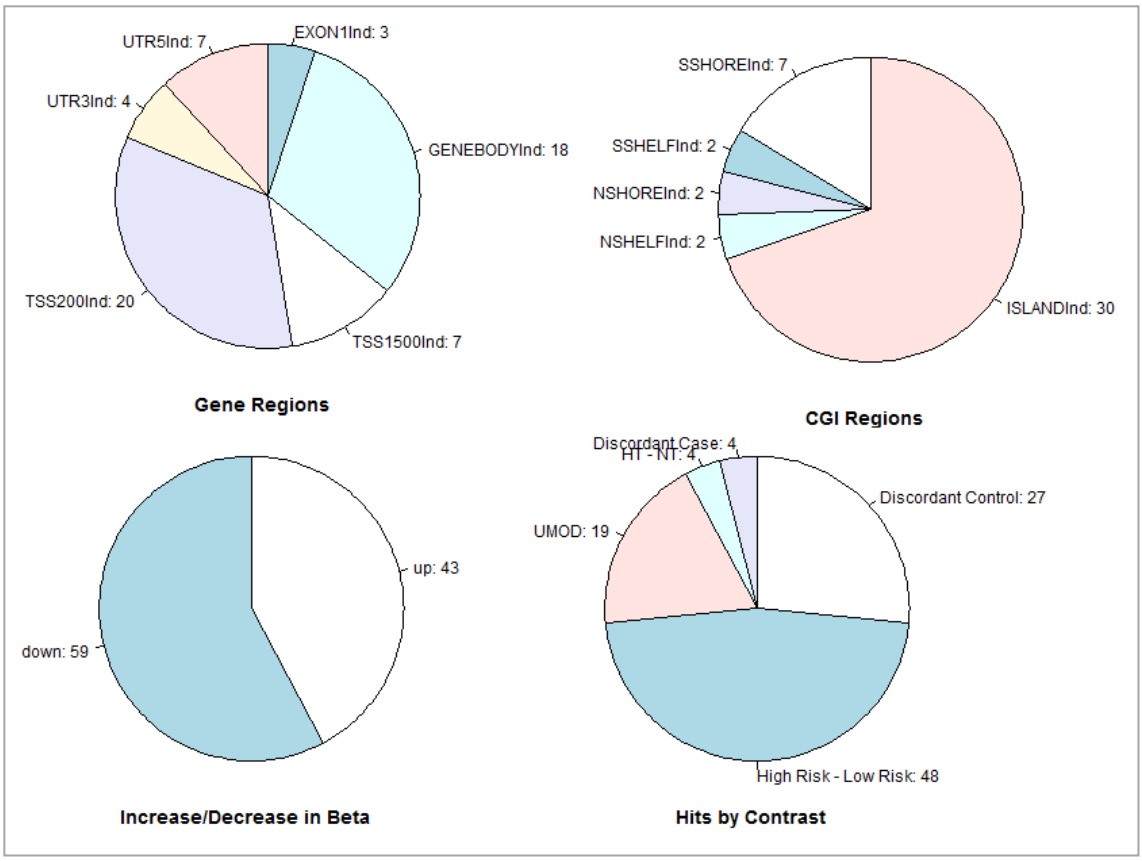

Fig. 3-6: Summaries of Significant Gene Regions and CGI Regions. CGI: CpG islands. Most of the significant regions were within 1500 bases of the transcription start site ("TSS200Ind", "TSS1500Ind") or within the gene body.

The largest change in beta was in the south shore of the CGI at 1685373-1685971 of chromosome 1 and measured from three CpGs (Table 3-3). The same region was also significantly associated with genetic risk ($\Delta\beta$=0.02, p_{BH}=0.01). The region with the greatest statistical significance is the 5' UTR of *CHID1* ($\Delta\beta$=0.08, p_{BH}=4.36 X 10^{-12}), and comprises 18 CpG sites (Table 3-4). *CHID1* and *APTX* are found at the intersection of four contrasts, and *CNIH4* and a CGI of chromosome 1 at 224544523-224545224 are found at the intersection of 3 contrasts. Interestingly, all nine results at the intersection of the discordant controls contrast and the *UMOD* genotype contrast are in the same direction, displaying a similarity in pattern of DM. In these two contrasts DM were mostly seen as hypomethylation (16/18 and 11/12 respective to the order above).

Gene/Region ID	Region Type	Δβ	p	p_{BH}	Site Count	Contrast	Survive SNP Filter
chr1:1685373-1685971 (NADK)	SSHORElnd	-0.26	5.24×10^{-9}	2.16×10^{-4}	3	Discordant Control	
CHID1	UTR5lnd	0.08	2.11×10^{-17}	4.36×10^{-12}	18	Discordant Case	Y
MBOAT2	TSS200lnd	-0.07	1.58×10^{-9}	1.09×10^{-4}	3	Discordant Control	Y
APTX	GENEBODYlnd	-0.05	2.19×10^{-10}	2.26×10^{-5}	6	HT – NT	Y
APTX	GENEBODYlnd	-0.04	5.93×10^{-6}	3.60×10^{-2}	6	High Risk - Low Risk	
APTX	GENEBODYlnd	0.04	1.46×10^{-12}	1.50×10^{-7}	6	UMOD	Y
chr6:83775374-83775766 (UBE3D)	ISLANDlnd	-0.04	4.23×10^{-10}	8.73×10^{-5}	9	Discordant Control	Y
APTX	UTR5lnd	-0.03	8.23×10^{-9}	5.66×10^{-4}	5	HT – NT	Y
APTX	GENEBODYlnd	0.03	2.46×10^{-6}	3.08×10^{-2}	6	Discordant Control	
chr1:26560449-26561028 (CEP85)	NSHELFlnd	0.03	3.39×10^{-6}	3.18×10^{-2}	1	Discordant Control	Y

Table 3-3: Top Ten Significant Changes in Geometric Mean Beta Value of Genomic Regions, Ranked by Absolute Change in Beta.
Δβ: the change in methylation level 'β' where 0=not methylated and 1=fully methylated. $p_{BH:}$ the p value multi-test corrected by the Benjamini-Hochberg method.

Gene/Region ID	Region Type	Δβ	p	p_{BH}	Site Count	Contrast	Survive SNP Filter
CHID1	UTR5lnd	0.08	2.11×10^{-17}	4.36×10^{-12}	18	Discordant Case	Y
chr11:910242-910500 (CHID1)	ISLANDlnd	0.01	2.31×10^{-13}	2.38×10^{-8}	3	Discordant Case	Y
CHID1	UTR5lnd	-0.02	1.30×10^{-13}	2.67×10^{-8}	18	UMOD	Y
APTX	GENEBODYlnd	0.04	1.46×10^{-12}	1.50×10^{-10}	6	UMOD	Y
APTX	UTR5lnd	0.02	7.38×10^{-11}	5.07×10^{-6}	5	UMOD	Y
CHID1	UTR5lnd	0.02	6.98×10^{-11}	1.44×10^{-5}	18	HT - NT	Y
APTX	GENEBODYlnd	-0.05	2.19×10^{-10}	2.26×10^{-5}	6	HT - NT	Y
chr11:910242-910500 (CHID1)	ISLANDlnd	-0.01	1.09×10^{-9}	5.62×10^{-5}	3	UMOD	Y
chr6:83775374-83775766 (UBE3D)	ISLANDlnd	-0.03	4.23×10^{-10}	8.73×10^{-5}	9	Discordant Control	Y
chr14:58666567-58667198 (ACTR10;C14orf37)	ISLANDlnd	-0.02	1.05×10^{-9}	1.08×10^{-4}	6	Discordant Control	Y

Table 3-4: Top Ten Significant Changes in Geometric Mean Beta Value of Genomic Regions, Ranked by p.
$p_{BH:}$ the p value multi-test corrected by the Benjamini-Hochberg method.

3.2.4 Genomic Distribution

A summary of results by chromosome in Figure 3-7 shows observed and expected counts of statistically differentially methylated genes by chromosome. Chromosome 1 was found to have an overrepresentation of differentially methylated genes (10.11%, p=0.04).

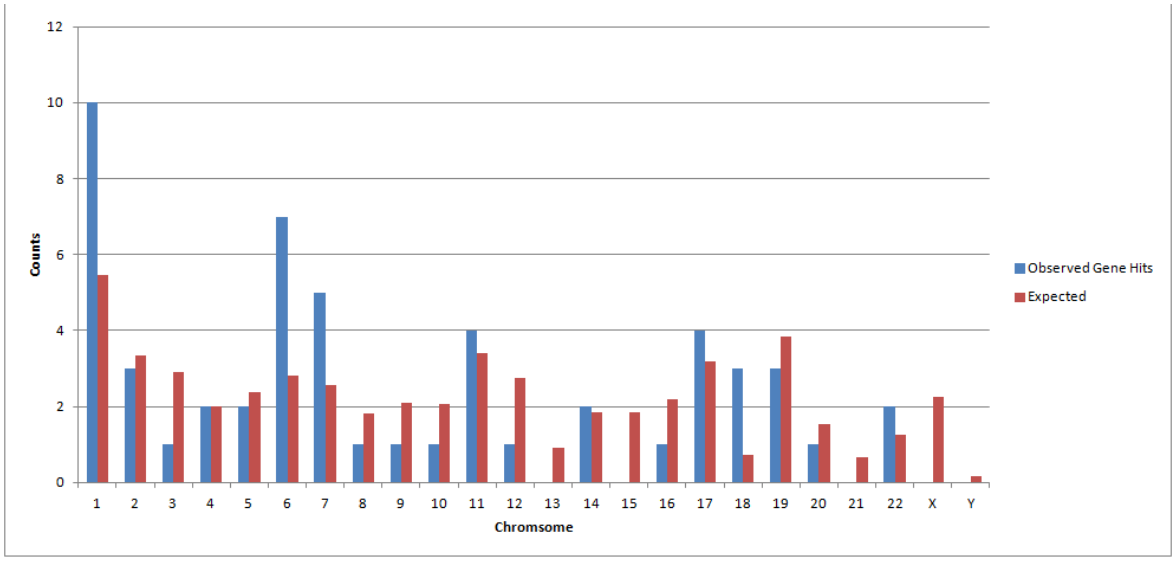

Fig. 3-7: The Genomic Distribution of Differentially Methylated Genes: showing disparity between observed results and those expected under a null hypothesis.
A significant increase in differential methylation noted in chromosome 1 in particular.

ANOVA revealed a significant association ($\Delta\beta = 3.54 \times 10^{-3}$, $p = 0.04$) between the *UMOD* phenotype and global methylation and a borderline-significant association ($\Delta\beta = 3.82 \times 10^{-3}$, $p = 0.05$) with the hypertension status and global methylation. Discordant controls and discordant cases were not found to have significant differences in global methylation as compared to the remaining samples (p = 0.63, 0.08).

The extent of differential methylation found in discordant controls and individuals of high genetic risk is greater within the LD blocks surrounding those SNPs which comprise the genetic risk score (Table 3-5). The group with high genetic risk has the greatest change in DM (23% increase in mean DM). There is only one risk-SNP-containing LD block which contains regions shown to exhibit DM (in genes *HLA-G*, *HLA-DQB1*, *ZNRD1-AS1*; Fig. 3-8).

Contrast	P Value	P_{BON}	Difference (%)
Discordant Control	2.63×10^{-11}	1.31×10^{-10}	0.87
Discordant Case	8.57×10^{-1}	1.00	1.83
Genetic Risk	9.48×10^{-3}	4.74×10^{-2}	23.3
Hypertension	3.19×10^{-1}	1.00	11.3
UMOD SNP	4.97×10^{-2}	2.48×10^{-1}	12.7

Table 3-5: Comparison of Differential Methylation in LD Blocks Containing Risk SNPs and Those Without Risk SNPs
DM was higher in risk SNP LD blocks for all contrasts, though only discordant controls and individuals of high genetic risk were significantly different. p_{BON} = bonferroni-corrected p values

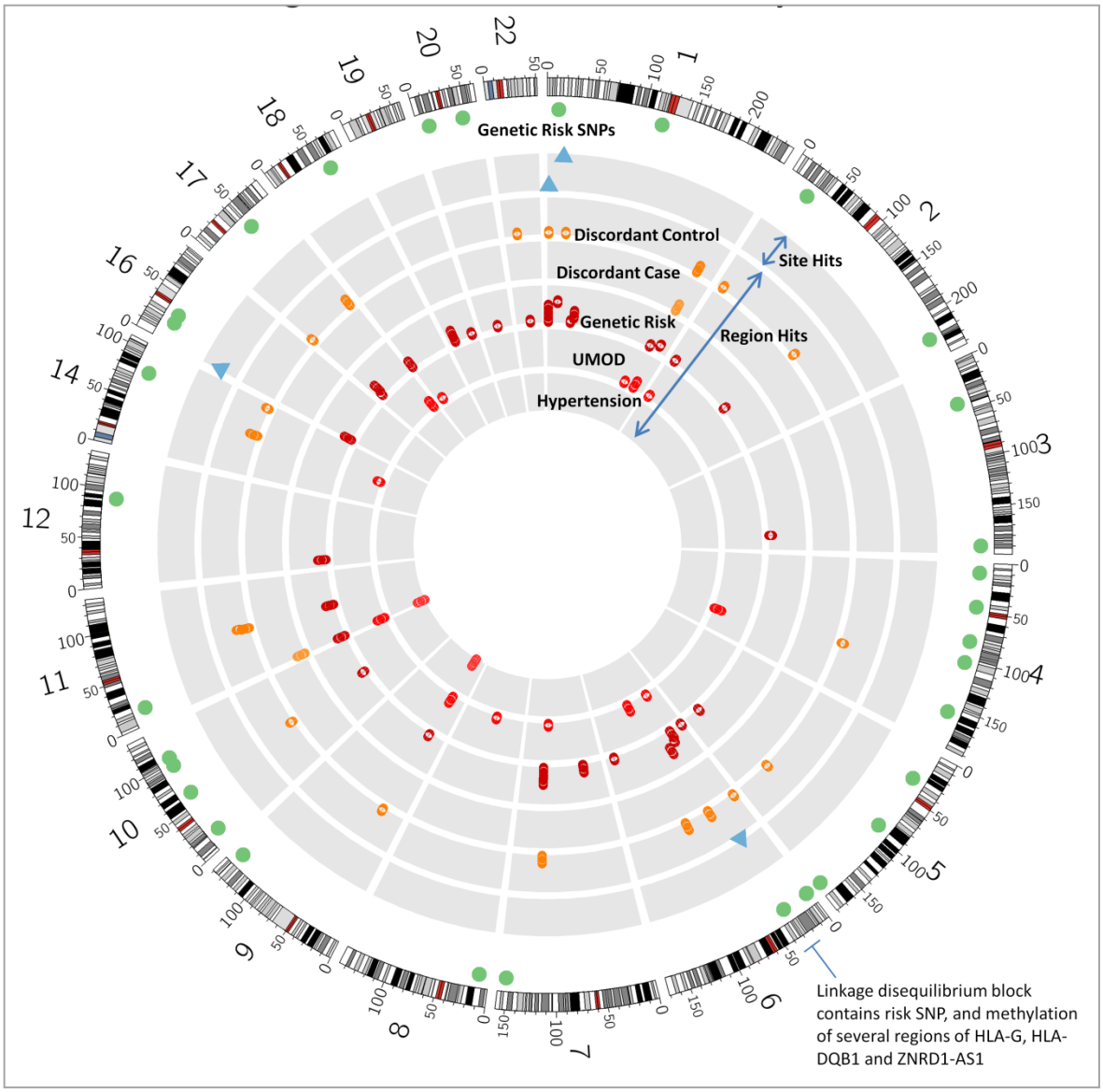

Fig. 3-8: More Detailed View of the Genomic Distribution of the Significant Events
Locations of the DM sites and region, risk SNPs, and LD block of interest is shown. Sizes of regions shown are not representative, they are enlarged for clarity. Chromosome numbers are shown around the outside of the image.

3.2.5 Clustering Analysis

Hierarchical clustering applied to a heatmap of hits from all contrasts was used to graphically represent the most important data (Fig. 3-9). It can be seen that the statistically significant differences between groups are mostly subtle ones - having little difference in methylation level. Related results cluster closely, e.g. the *NADK* gene body as a whole and the significant CpG site within it (Fig. 3-10).

PC1 explains 98.02% of the variance in the data (Fig. 3-11), and graphing the PC1 data ordered by sample group showed no obvious trends (Fig. 3-12 to 3-15). Euclidean distances within subgroups are not significantly different from distances between subgroups ($p= 0.07$), perhaps due to the small sample size. No significant clustering was detected by hypertension, genetic risk or *UMOD* genotype.

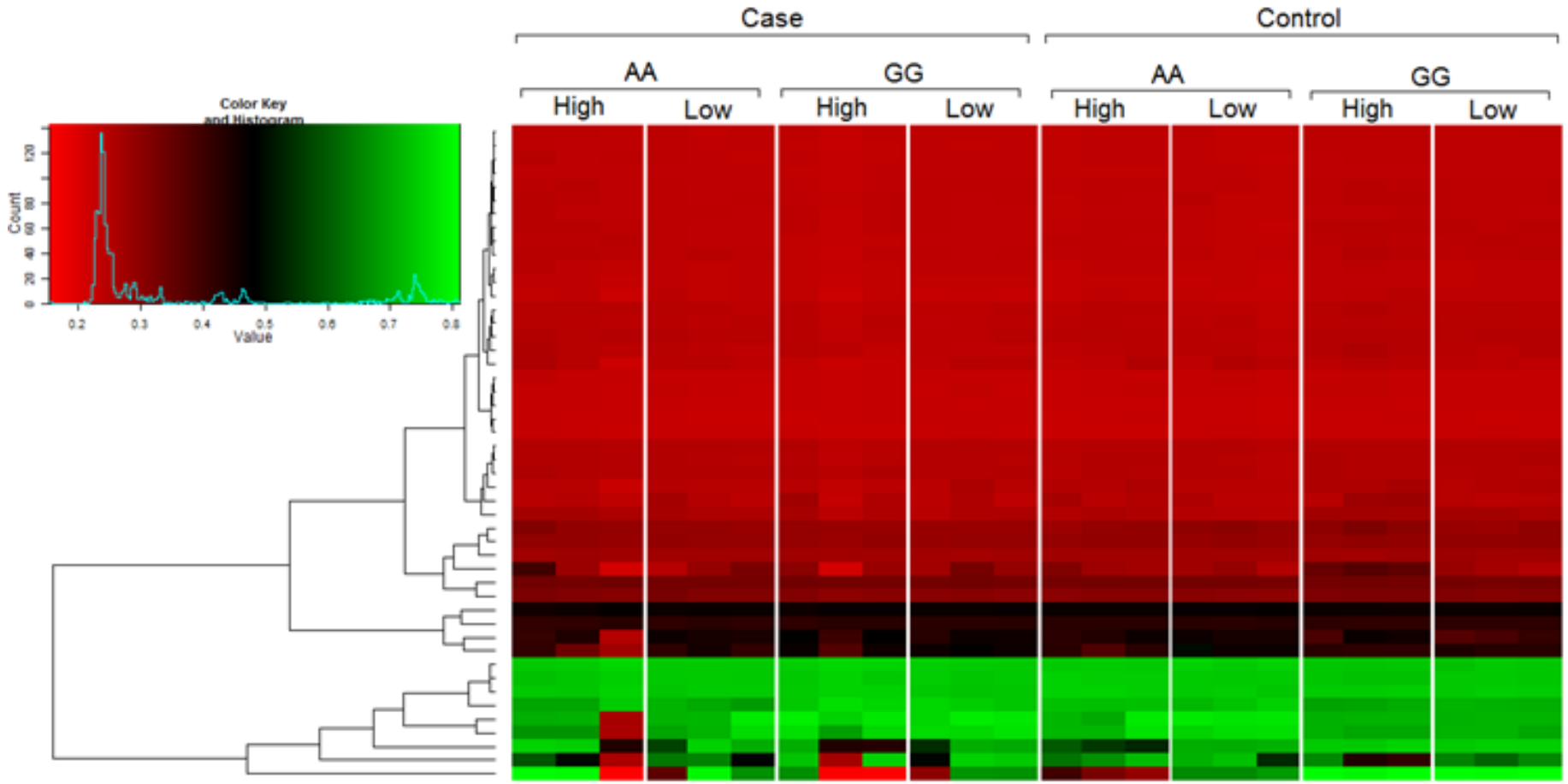

Fig. 3-9: A Hierarchical Clustering of the Average beta of all Significant Gene Regions.
Includes results from all contrasts tested. Representation of data in this format quickly conveys that most of the significant differential methylations detected are subtle or somewhat inconsistent.

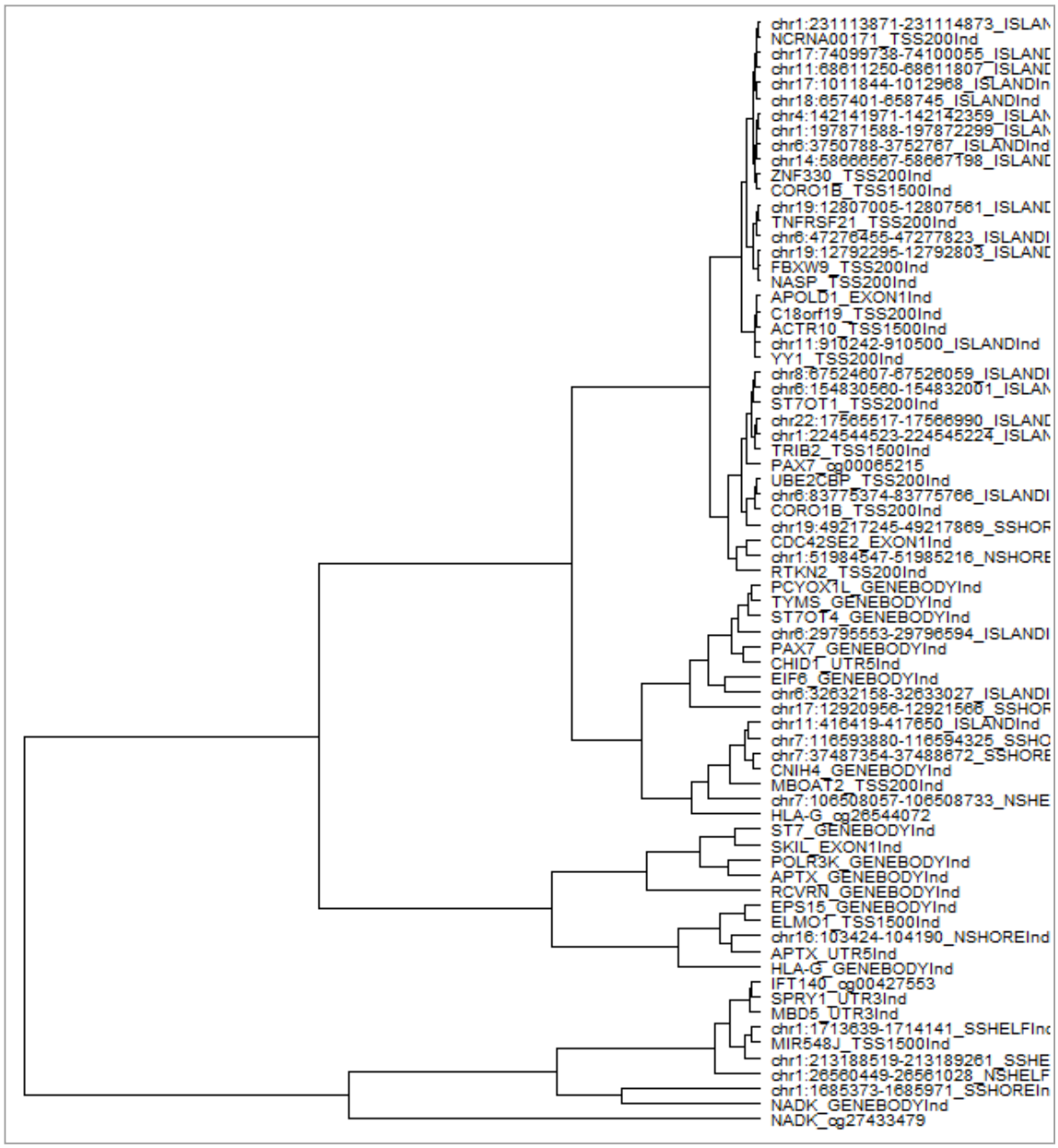

Fig. 3-10 A Hierarchical Clustering of the Average beta of all Significant Gene Regions.
Includes results from all contrasts tested.

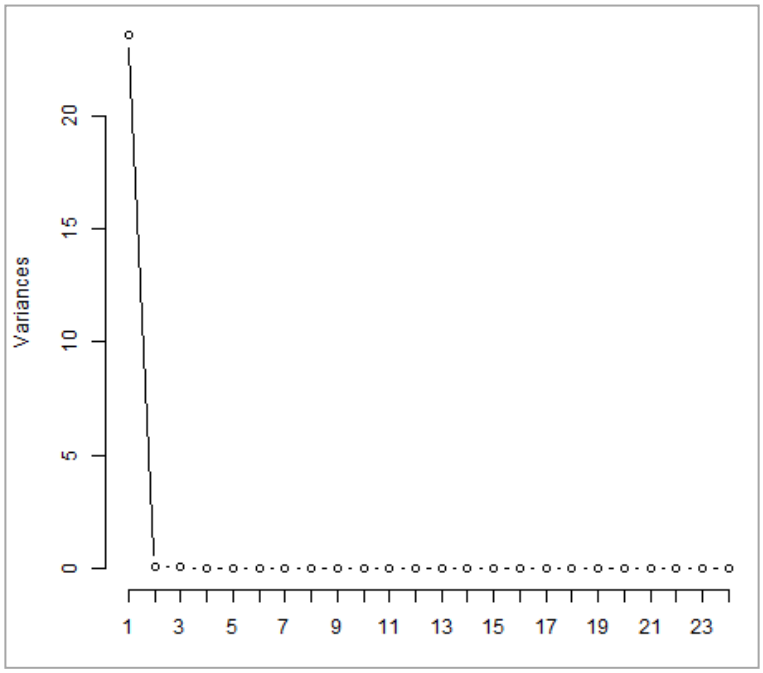

Fig. 3-11: Scree Plot showing that the Second and Third Components Describe Relatively Little Variance

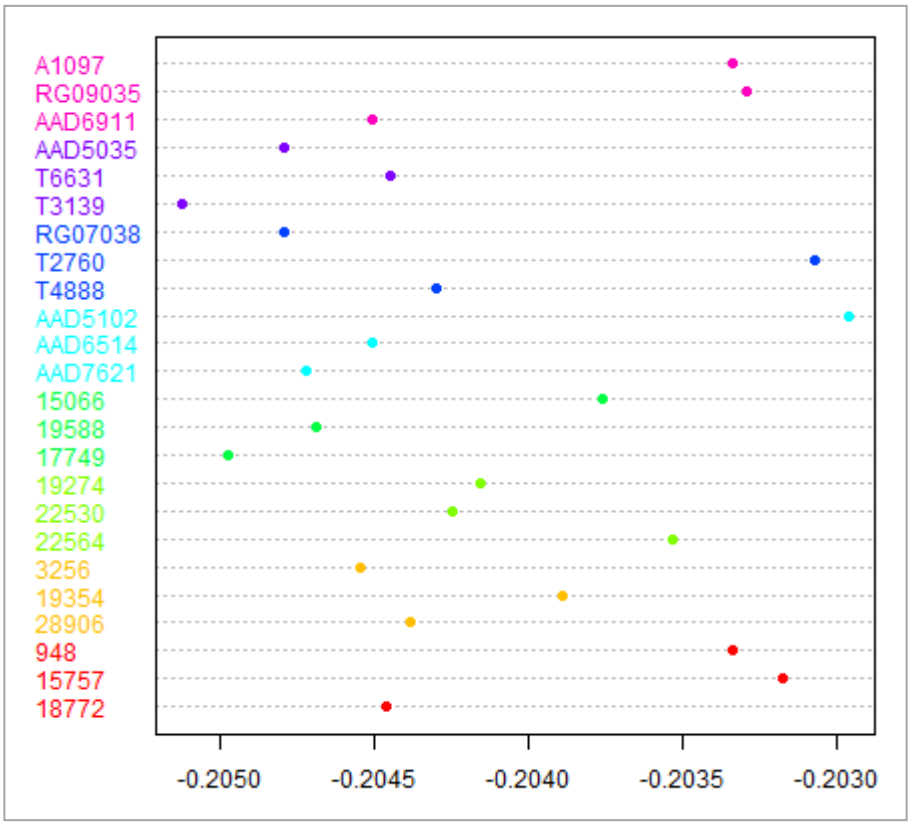

Fig. 3-12: Principle Component 1, Showing Possible Clustering of Subgroups.
Coloured by subgroup. It is difficult to assess clustering in such small subgroups, but it does not appear that any subgroup is noticeably distinct from the others. Sample IDs are on the vertical margin.

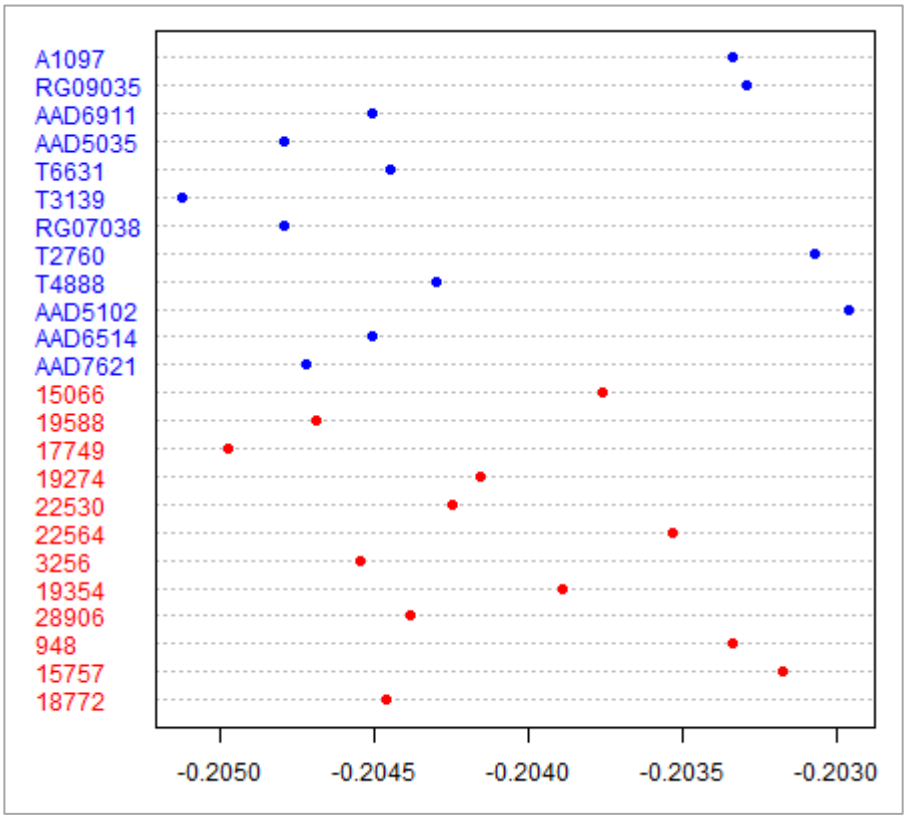

Fig. 3-13: Principal Component 1, Coloured by Hypertensive Status.
Sample IDs are on the vertical margin. There may be some separation by hypertensive status.

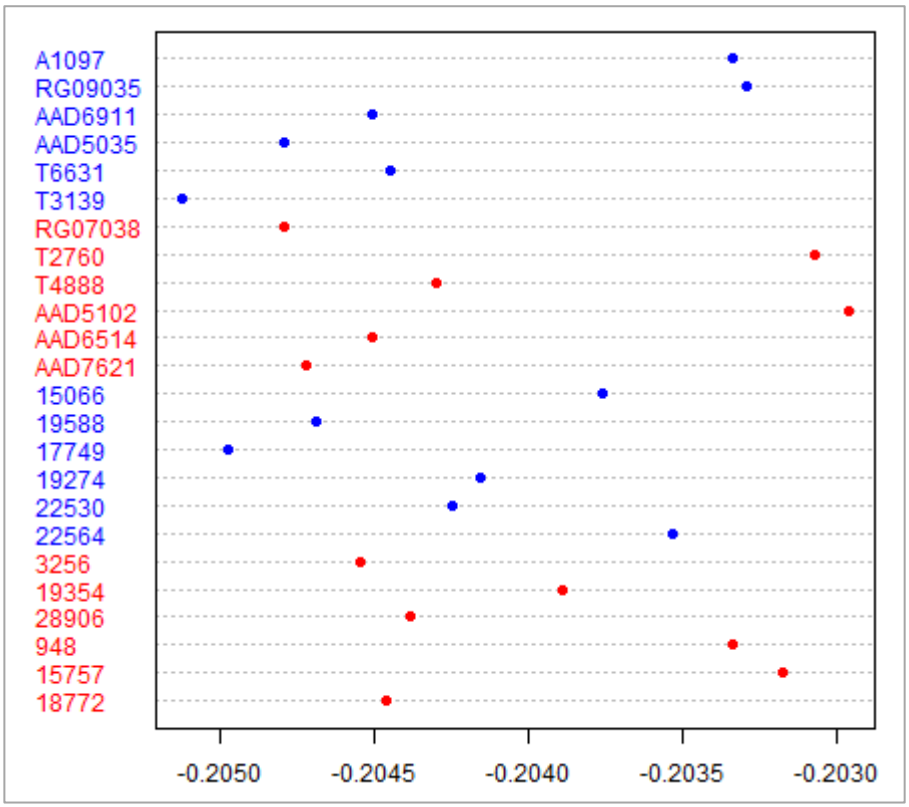

Fig. 3-14: Principal Component 1, Coloured by UMOD Allele.
Sample IDs are on the vertical margin.

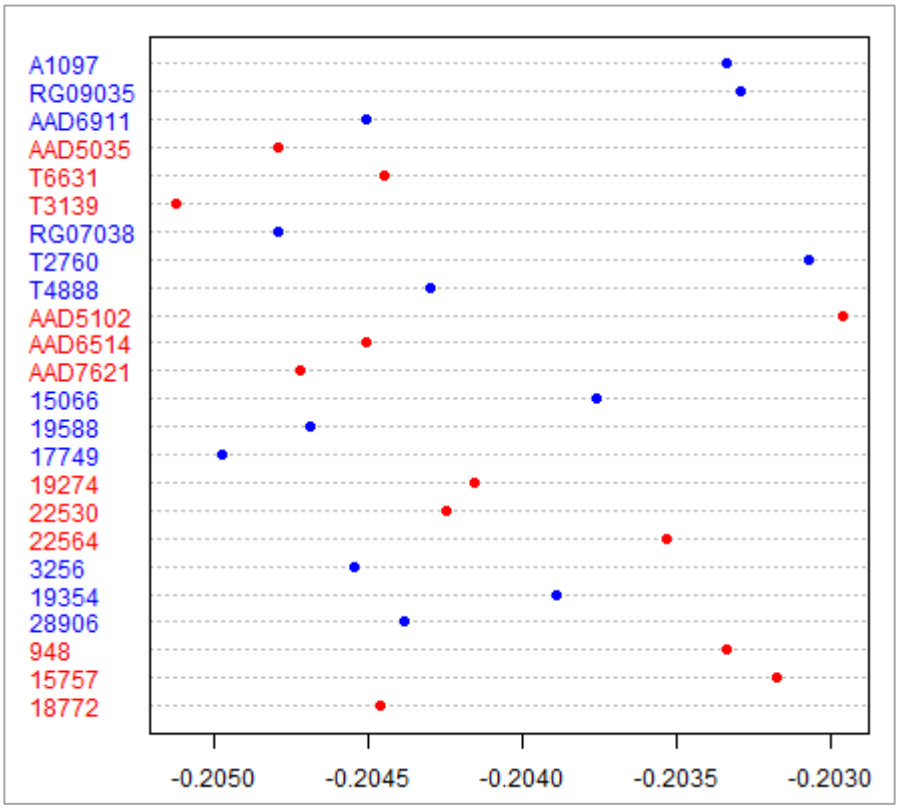

Fig. 3-15: Principal Component 1, Coloured by Genotype Risk Score.
Sample IDs are on the vertical margin.

3.2.6 Term Enrichment Analysis

GO Term enrichment analysis showed many general disturbances across broad gene ontologies. Single ranked lists of the discordant control and discordant case lists were shown to be significantly enriched in 290 and 326 GO terms respectively, after applying multiple test correction. In both analyses (and their intersection) the biological processes terms formed the majority of significant results (Fig. 3-16).

The discordant controls contrast was particularly enriched in terms related to regulation, development and morphogenesis, e.g. “anatomical structure morphogenesis” (p_{BH}: 4.77 X 10^{-15}). A text analysis of the discordant control terms shows 118 occurrences of “regulation”; 37 of “morphogenesis”; 32 of “development”; 21 of “binding”; 21 of “metabolic”.

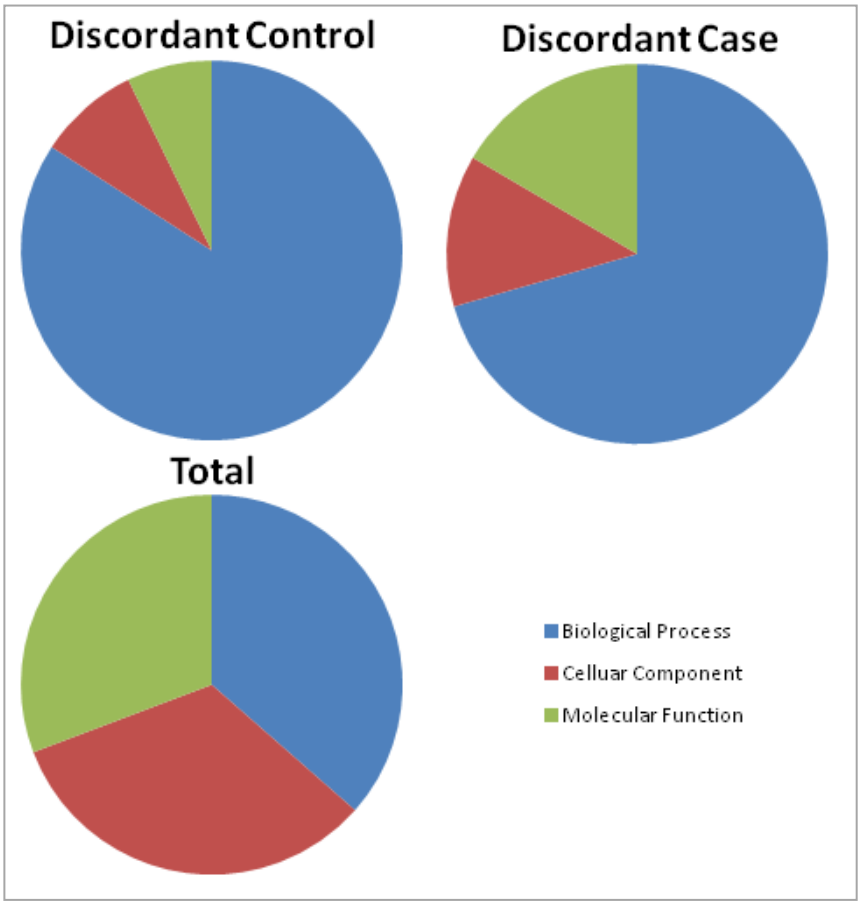

Fig. 3-16: Discordant Control and Discordant Case Appear to be Enriched for Biological Process Gene Ontologies
Ratios of Different Types of Significantly Enriched Gene Ontologies found in Two Contrasts in this Experiment contrasted with the Ratio for the Entire Gene Ontology Database.

All 21 molecular function gene ontologies were related to transcription factor activity and transcription, e.g. “sequence-specific DNA binding transcription factor activity” (p_{BH}: 1.17×10^{-09}). The cellular component terms were mostly related to “cell junctions” (p_{BH}: 9.75×10^{-06}), “organelles” (p_{BH}: 1.71×10^{-04}), and “dendrites” (p_{BH}: 2.80×10^{-3}). Inclusion of dendrite/neuron-related terms may be indicative of enriched differential methylation of genes related to dendritic cells or a related immune cell type.

The discordant cases contrast was particularly enriched with metabolic terms, e.g. “cellular macromolecule metabolic process” (p_{BH}: 5.42×10^{-26}). Text analysis reveals 46 occurrences of “metabolic” – over twice as many as in the discordant controls. Some other words with high occurrences were: “catabolic” 41 times; “binding” 35 times; “regulation” 33 times; “ligase” 18 times. There are 33 terms shared between the two contrasts, however they are too general to be very informative.

3.3 Discussion

This study benefits from genotype data which aids interpretation of the results in cases where probes may truly be indicating SNPs rather than DM. Rather than only assessing DM in individual probes, summarised values have been generated for gene and CGI regions, which in addition to more powerfully picking up subtle but consistent DM across a region, should be less prone to SNPs confounding results through differences in hybridisation affinity.

An extreme case-control design was employed as in the original GWAS which, through that design, successfully identified a novel SNP associated with lowering blood pressure (*UMOD*). As all samples are from male patients of the same ethnicity and region it is unclear whether the trends identified extend across the population or are unique to this group, although this is also beneficial since the selected group is more homogeneous and statistical adjustments are not required for confounding factors such as sex and genetic background. In addition to assessing DM in hypertension, this study makes use of nested subgroups of genetic risk to test the hypothesis that DM may be involved in the discordant phenotypes present in the cohort. While the probes in the 450K have been selected in such a way as to try to avoid cross-hybridisation, and may have been more focussed towards transcribed regions, the global DNA methylation levels will be less biased in its description of global methylation than a technology specifically targeting promoter regions or repeats.

3.3.1 Differential Methylation of NADK and its Potential Role in Hypertension

Nicotinamide adenine dinucleotide phosphate is a coenzyme which, in its reduced form (NADPH), acts as a reducing agent in several different pathways (218-220). NADPH oxidases (NOX) are enzymes which transfer electrons from NADPH to molecular oxygen, creating superoxide anions. Superoxide is a reactive oxygen species which can form hydrogen peroxide and other reactive oxygen species. This process was first discovered to occur in phagocytes as a non-

specific host defence and in that instance is called an oxidative burst or respiratory burst (221;222). It has since been discovered to be constitutively active at low levels in other cell types (223).

NADPH is responsible for the restoration of all known innate defence systems against oxidative stress (224-226). In doing so, through the transfer of an electron, it is oxidised to $NADP^+$. While human cells seem ultimately to rely primarily on NADP-dependant dehydrogenases for their protection against oxidative stress (by reducing $NADP^+$), it has been shown that the activity of NAD kinase (NADK) can also contribute a moderate enhancement (227). NADK is an enzyme which converts NAD to NADP by phosphorylation. An increased amount of NADK activity allows more conversion of NAD to NADP and thus there is a greater pool specifically of NADPH, since NADP exists largely in its reduced form in human cells (228). The hypomethylation of *NADK* detected in discordant controls may be responsible for an increase in expression of *NADK* and in turn a greater protective effect against the potentially hypertension-inducing effects of oxidative stress - potentially explaining why individuals with a higher genetic risk for hypertension might not express the phenotype.

Oxidative stress is associated with many diseases and appears to play a central role in the pathophysiology of hypertension. Oxidative stress leads to a reduction in nitric oxide bioavailability which is a major factor in controlling vascular tone and therefore blood pressure. The action of several vasoconstrictor peptides such as angiotensin II, endothelin-1 and urotensin II lead to oxidative stress by activating enzymes like NOX and xanthine oxidase which generate reactive oxygen species. NOX enzymes in particular appear to be involved in blood pressure control via several mechanisms and across multiple relevant tissues (229). The effective reduction of oxidative stress in the relevant tissue(s) – either by introduction of antioxidants, or by interfering with the pathways involved in reactive oxygen species - is an active area of research (230). Several existing hypertension medications such as ACE inhibitors and calcium channel blockers may work in part by reducing oxidative stress.

A SNP (rs1130355) for which there is no data within the significant *NADK* CpG site with a high minor allele frequency (MAF) of 0.36 could be the true source of

the association with hypertension. This hypothesis is backed up by the fact that the *NADK*-related regions are filtered out by the SNP-filtered analysis. If it were this SNP itself it would be a novel association. The apparent association with hypertension may also be a nearby SNP in close linkage with it - it can be seen from Figure 3-17 that there is another potential occurrence of DM in a close genomic proximity to *NADK*, within the *GNB1* gene.

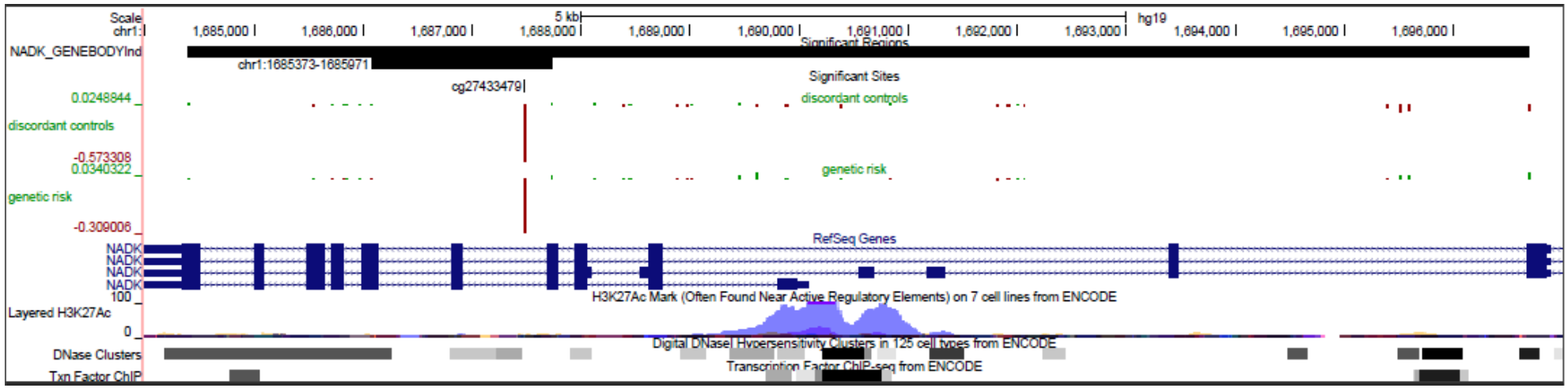

Fig. 3-17: Genomic context of *NADK*, and the methylation regions and CpG cg27433479

Fig. 3-18: Genomic context of *CHID1*.
Location of methylation regions consistent with downregulation by promoter activity.

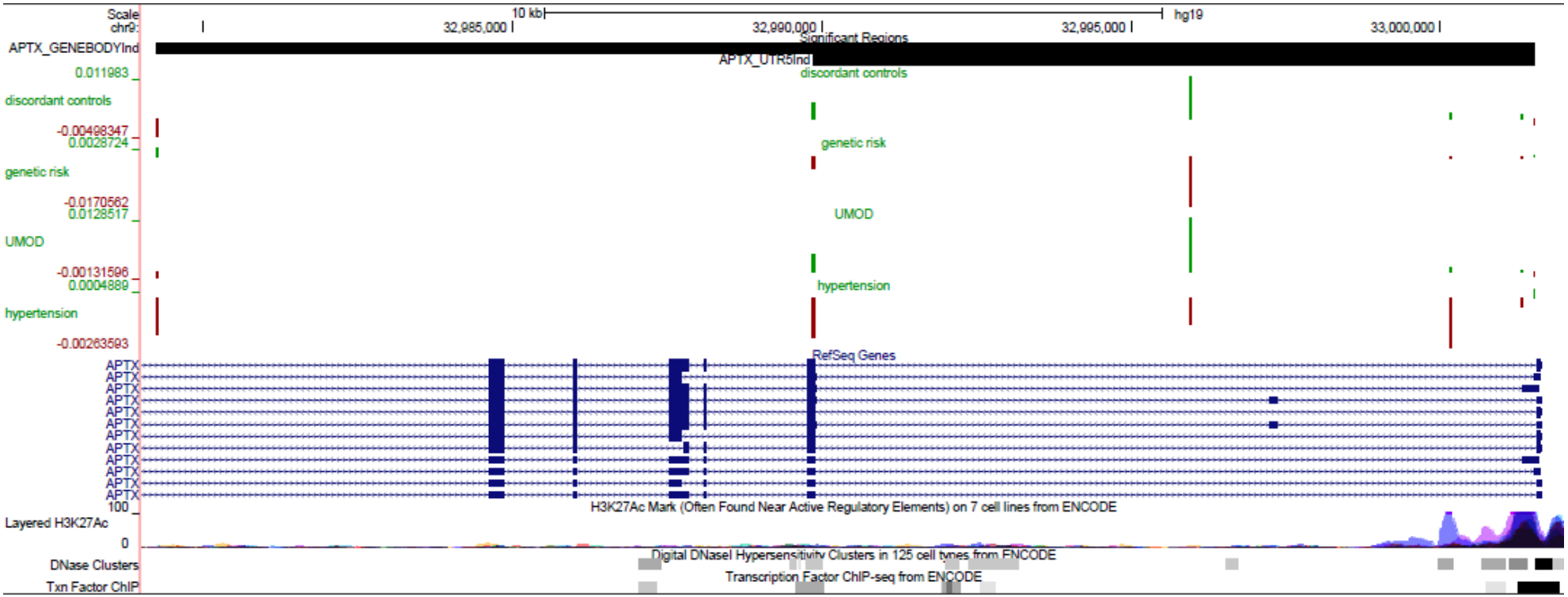

Fig. 3-19: Genomic context of *APTX*.
Location of methylation regions consistent with downregulation at promoter, upregulation in gene body or splicing – many variants are referenced in RefSeq.

3.3.2 Genes CHID1 and APTX are found at the Intersection of Results

While not all of the directions of the DM (i.e. positive or negative Δβ) are as expected, the relationships between the discordant cases and controls and the hypertension contrast appear to be in concordance in the two most prominent hits, as identified by relatively large, significant DM, across several contrasts and regions – *CHID1* and *APTX*. *CHID1* (Fig. 3-18) and *APTX* (Fig. 3-19) show DM in four of the five contrasts investigated (Fig. 3-5) and the latter shows DM in two gene regions. Regions of these genes are also shown repeatedly in the top ten significant regions ranked by beta (Table 3-3) and by significance (Table 3-4). DM of the gene body of *APTX* could potentially have a protective effect for discordant controls (hypertension: Δβ=-0.05, $p=2.26 \times 10^{-5}$; discordant control: Δβ=0.03, p=0.03) and DM of the 5' UTR of *CHID1* may have a predisposing effect on discordant cases (hypertension: Δβ=0.02, $p=1.44 \times 10^{-05}$; discordant case: Δβ=0.08, $p=4.35 \times 10^{-12}$).

CHID1 is part of a family of the glycoside hydrolase 18 (GH18) family of chitinases and followed a different evolutionary path from all others of the family, remaining conserved and unduplicated in humans (231). It is upregulated in macrophages, its protein has been detected in lysosomes, and it is secreted into the extracellular region. While little is known about its function it has the ability, *in vitro*, to bind a variety of saccharides and saccharide-containing molecules such as chitin and lipopolysaccharides (LPS). Chitin is found in fungal pathogens and LPS is a potent driver of inflammation of bacterial origin. Inflammation is thought to contribute to hypertension by via oxidative stress and endothelial dysfunction.

LPS in particular can induce an immune response from multiple cell types, including macrophages, and it has been shown to affect blood pressure in animal models (232). To resolve LPS-driven inflammation the bacterial cells responsible need to be cleared and the LPS neutralised. Although CHID1 has not been shown to bind LPS *in vivo* and its binding affinity is relatively low, recombinant CHID1 protein can reduce the inflammatory response to LPS challenge in macrophages – specifically regarding cytokines IL-1β, IL-8, TNFα, and IL-6 (233) – indicating an involvement in LPS neutralisation.

CHID1 has been shown to interact with the trans-membrane receptor stabilin-1 (234), which is known to be expressed in macrophages and endothelial cells (including angiogenic endothelial cells). In macrophages it appears to be responsible for intracellular sorting of endogenously synthesised CHID1 into lysosomes, leading to CHID1 secretion (235), potentially facilitating an effect of CHID1 on LPS neutralisation. Stabilin-1 mediates apoptotic cell clearance in alternatively activated macrophages (236). Apoptotic cells can induce inflammation, and in addition to the removal of this source of inflammation apoptotic cell clearance triggers anti-inflammatory mechanisms (237). Stabilin-1 is capable of binding gram-negative and gram-positive bacteria, which may indicate involvement in bacterial cell clearance also (238).

Stabilin-1 acts as a scavenger receptor for a variety of molecules including acetylated low density lipoproteins (acLDL) and oxidated low density lipoproteins (oxLDL). oxLDL is normally cleared by macrophages however an excess of oxLDL can cause their transformation into foam cells which play a role in the initiation and progression of atherosclerosis - physically contributing to atherosclerotic lesions and driving inflammation. oxLDL is a major risk factor for cardiovascular disease with multiple putative mechanisms (239). It affects other relevant cell types beyond macrophages, including endothelial cells which also express stabilin-1.

SPARC is an acLDL which modulates angiogenesis and is mainly endocytosed by stabilin-1 in macrophages which appear to be responsible for clearance of SPARC from the extracellular space (240). Stabilin-1 has also shown angiogenesis-modulating activities (238). Anti-angiogenesis medications (targeting VEGF) are used in the treatment of cancer and have been demonstrated to cause hypertension as a side-effect. This may occur by multiple possible mechanisms including vasoconstriction induced by activation of endothelial cells, increased endothelin signalling and decreased NO signalling leading to vasoconstriction, and by reduced lymphangiogenesis modulating the salt sensitivity of blood pressure (241).

APTX is a gene involved in DNA repair and expressed highest in lymphoblasts. DNA damage is associated with hypertension, seen as a product of the increased oxidative stress the system is under (242). Increased expression of DNA repair genes such as *PARP1* – a gene known to interact with *APTX* – have been associated with hypertension, possibly as a response to the increased ROS-related DNA damage(243). A CGI within the gene body of *PARP1* and of *CNIH4* (chr1:224544523-224545224) is shown to be differentially methylated across three contrasts. An overlapping region annotated by IMA as being the gene body of *CNIH4* however being in the region of both genes is also found to have DM across these contrasts. In addition to passing the SNP-filtered version of the analysis, no significant associations with hypertension were found in these regions in the previous GWAS study, indicating that the result is novel and not a false positive induced by a SNP. It seems plausible that *APTX* and *PARP1* are differentially methylated/expressed as a consequence of the same oxidative stress which drives hypertension, rather than being a cause of hypertension themselves.

3.3.3 Other Genes of Interest and Global Methylation

UBE3D encodes a ubiquitin protein ligase, an enzyme which transfers ubiquitin to substrates promoting their degradation by the proteosome. DM of a CGI found at the transcription start site of the UBE3D gene has the 8th greatest difference with the 9th greatest significance ($\Delta\beta$=-0.03, p=8.73 X 10^{-05}). Other ubiquitin protein ligases, *UBE3A* and *UBE4A* were amongst the 296 genes detected by Jia et al. *MBOAT* and *CEP85* are involved in lipid metabolism and the centrosome respectively and as such are related to the two largest functional categories described by Jia et al.

PAX7 is one of the family of paired box (PAX) family of transcription factors involved in fetal development and a number of diseases (244). *IFT140* encodes a subunit of intraflagellar transport (IFT) complex A and mutations of the gene have been associated with kidney disease(245). *HLA-G* is a leukocyte antigen gene which is relatively highly expressed in both blood and placenta, is downregulated in the

placentas of preeclamptic women(246) and hypermethylation of its promoter is associated with that downregulation(247). It is one of several genes in the risk-SNP LD block shown in Figure 3-8. As with *NADK*, the significant site of *HLA-G* has a common SNP (rs2076328, MAF= 0.50) within the CpG sequence itself, and the association may truly be with a SNP rather than DM. The probes corresponding to significant sites of *PAX7* and *IFT140* however do not contain SNPs annotated by Illumina.

Jia et al found that in the placenta of patients with preeclampsia the greatest overrepresentation of differentially methylated genes was in chromosome 1 (10.5%, p= $5X10^{-3}$). Despite the differences in condition and sample type it may be interesting that this observation was somewhat mirrored in this data (10.1%, p=0.04) (215). The over or under representation of other chromosomes is less interpretable as numbers for expected counts are lower, however chromosomes 12 and 19 do not seem to follow the less pronounced over-representation also described by Jia et al (Fig. 3-7).

A significant association was found between global methylation and hypertension, showing a ~0.3% increase in hypertensives. This result may be somewhat biased by the selection criteria used by Illumina to assemble the 450,000 CpGs tested on their microarray. It is in agreement with a study by Kim et al which focussed on repeat regions, however it is also in disagreement with a study by Smolarek which used a smaller sample size (60 compared with 286) but a less targeted ((i.e. more global) methodology (199;200).

The results of the term enrichment analysis show similarity with the functional analysis of Jia et al (215), in particular the two largest groups, one relating to metabolic processes and the other to replication, repair etc. Numerous terms related to the immune system were also shared between both sets of results. Some structural terms which were significantly enriched may bear relevance to vascular remodelling, e.g. “anatomical structural development”.

The increased differential methylation within the LD blocks surrounding those SNPs which comprise the genetic risk score similar to previous findings of an association between hypertension risk SNPs and nearby CpGs (248).

3.3.4 Conclusion

The findings regarding *SULF1, PPP1R2,* and *LOC100132354* from the previous studies of DM in essential hypertension could not be replicated, possibly because of samples being sourced from groups of different ancestry, different environment, small sample sizes, or simply due to the heterogeneity of hypertension. As essential hypertension is a complex, multifactorial disease, with many potential causes and effects it is perhaps not surprising that the results only comprise parts of some components of the disease – in particular changes to metabolism, efficiency of combating over-oxidation and the resulting DNA repair response, immune responses, and structural development (in particular angiogenesis). *CHID1* and *APTX* are repeated across several contrasts, several regions, and/or show close interactions with other identified genes, lending additional confidence to the relevance of these results in particular. *APTX* (and another DNA repair gene, *PARP1*) may be differentially methylated in response to the increased oxidative stress found amongst people with hypertension.

CHID1 appears to be responsible for the neutralisation of LPS, a molecule of bacterial origin which otherwise would cause inflammation. Inflammation can in turn cause oxidative stress and endothelial dysfunction, ultimately resulting in increased blood pressure. The hypermethylation at *CHID1* in hypertensive patients may cause a decrease in expression which could be partly responsible for their increased blood pressure. It may also affect blood pressure through its interaction with stabilin-1, which is involved in several relevant processes including angiogenesis, atherosclerosis, and the clearance of pro-inflammatory cells. CHID1 protein could modulate its activity by binding to it or could block its binding sites for other molecules.

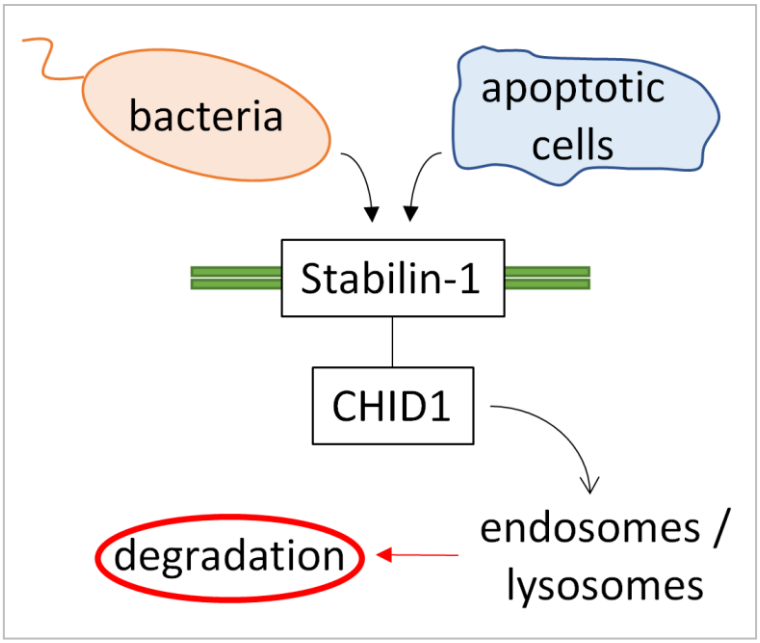

Fig 3-20 CHID1 Interacts with the Stabilin-1 and is Involved in LPS Neutralisation
CHID1 is packaged in the lysosomes by stabilin-1 for secretion into the extracellular region where it binds LPS and other antigens to inhibit inflammation

Several statistical tests indicate a protective effect of DM in discordant normotensive patients – i.e. those who are normotensive despite high genetic risk - and that term enrichment analysis would suggest that its role is primarily related to morphogenesis. By far the largest change in beta is in a CpG of *NADK*, whose expression has been shown to have an impact on reducing oxidative stress by the phosphorylation of NAD to NADP, which exists primarily as the protective NADPH. However this may truly be association with a SNP which was not covered in the preceding GWAS study. Although further studies are needed to confirm these associations, DNA methylation appears to play a role in hypertension – on a CpG site level, gene region level, and global level – whether it is causative or simply a downstream effect. Knowledge derived from this experiment and similar experiments may help both in the understanding of the pathogenesis of hypertension and in the future of personalised medicine where epigenetic marks may be insightful, adding novel points to signatures of disease predisposition or drug response.

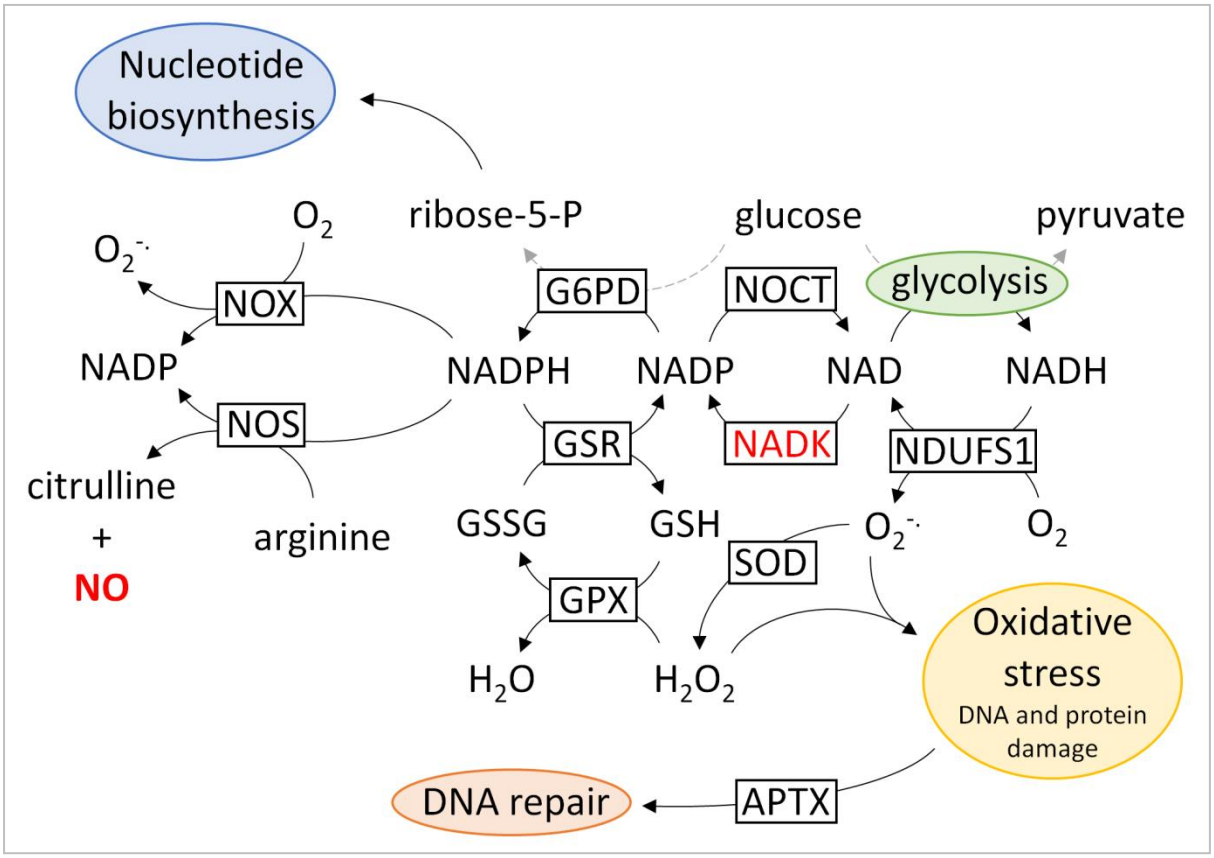

Fig 3-21 NADK Converts NAD to NADP, and Subsequently NADPH Combats Oxidative Stress
The addition of a phosphate by NADK means an increase specifically of NADPH, since NADP exists largely in its reduced form in human cells. NADPH combats oxidative stress via reduction of its targets. Meanwhile APTX responds to oxidative stress by its involvement in DNA repair.

The interpretation of the main results is enriched by integrating various additional datasets. Gene ontologies, a common addition to omics experiments, summarised the results in terms of biological processes, cellular components and molecular function. Extensive SNP location data based from dbSNP and experimental SNP data from the previous GWAS study on this cohort were useful in determining which of them were the most reliable results in terms of ruling out interference from SNPs. Integration of gene region data and linkage disequilibrium blocks allow the data to be analysed with different boundaries and categorisations which is useful since still relatively little is known about the causes and effects of differential methylation.

4. Integration of Diverse Datasets Towards Studying Mechanisms Underlying Variation in Left Ventricular Mass Index

4.1 Introduction

With the advances in omics technologies and analysis methodology, and with rapidly reducing costs, experiments involving multiple large datasets are increasingly attractive. As biomolecules of different classes interact it follows that we should aim to combine data of different types in order to test whether predictors are independent of each other, to describe relationships between different parts of the system and to put results in a greater context.

Heart failure is a major global health issue; it is both common and rising worldwide (249). Left ventricular hypertrophy (LVH) is an intermediate cardiovascular phenotype associated with the development of heart failure (65). A previous study in 536 patients with hypertension integrated clinical variables in relation to left ventricular mass (LVM), and found systolic blood pressure (SBP), body mass index (BMI), height, sex and history of aneurysm of the abdominal aorta to be important predictors (71). This model accounted for 45% (adjusted R^2) of the variation in the data set, and the inclusion of electrocardiography data added a small increase of 2%.

Other studies have shown similar findings in various contexts – each case including relatively small numbers of variables and only including clinical measurements or assessments (250;251). Although only clinical models of LVM have been developed, some molecular predictors have been identified and could be modelled with each other or along with clinical predictors, potentially describing additional variation in the data, or providing novel insights. Single nucleotide polymorphisms in the angiotensinogen and apolipoprotein B genes have been shown to significantly predict changes in LVM index (LVMI) (79) and omics experiments of various types have detected many significant associations with LVH (252;253).

Within the Network of Excellence “InGenious HyperCare” patients with hypertension and normotensive controls belonging to the same families were characterised for LVM and other clinical variables (254). The full cohort consists of 1589 participants which were recruited from 19 study centers in Europe. While no studies have been published before on this combined cohort, cardiovascular phenotypes from 535 of these individuals have been previously published (145). For the present analysis 270 participants from four sites in Gdansk, Krakov, Glasgow and Prague were selected for detailed molecular phenotyping. These 270 patients exhibit an approximately normal distribution of left ventricular mass index, and include 34 cases of LVH. There were approximately equal numbers of males and females at a mean of 48 years of age. Further details on the patients which participated can be found in section 2.3.1. LVMI was analysed with respect to one clinical dataset and several molecular datasets. The molecular datasets include data on 1,605 biomolecules including peptides, miRNAs and metabolites. They were acquired with both targeted and untargeted methodologies and derived from blood and urine samples. Lab work and some initial data processing was done in various labs specialising in different methodologies, before being sent to Glasgow for this combined analysis.

The large number of predictor variables, the large amount of missing data for some predictors and the varying scales and sizes of the individual datasets present challenges to analysis (255;256). These challenges are not unique to our dataset but are a common problem for integration of molecular and clinical features. We present two alternative linear regression-based approaches to deal with such large diverse datasets in order to identify mechanistic biomarkers for the underlying pathways of increased LVM. Analysing variables from molecular and clinical data in combination could enhance discovery of significant associations and allow for the description of relationships between the relevant molecules.

The main aims of this analysis were to identify novel molecular predictors, to compare these with clinical predictors, and to see whether molecular predictors explained additional variation in the data which clinical predictors alone did not. The first approach involves a screening step to reduce dimensionality before using a range of established methods to further test those significant predictors of LVM

and how they correlate with each other. First, as had been done in other studies, a linear model was established using only clinical variables. Each molecular dataset was then interrogated with univariate linear regression, multiple linear regression with confounders included (age, sex and BMI) and multiple linear regression with the above-mentioned clinical model. Any variable with a Benjamini-Hochberg adjusted p value < 0.05 in any iteration of testing was used as a potential predictor of LVMI in the final modeling stage, where two further models were generated – a molecular model and a mixed molecular-clinical model. Correlations between variables of interest were mapped out, and complimentary clustering analysis done.

The strength of this approach is in the detail and relative robustness of the results, however with the numerous rounds of various statistical tests and data from different sources which all need to be treated appropriately, the analysis requires time and varied expertise. The length of time to thoroughly analyse data in this manner increases with the size of the datasets and number of significant results. It would therefore be attractive if there were a simpler way to analyse large mixed datasets such as this, which was not so affected by multi-test correction. The second approach uses principal components analysis (PCA) with varimax rotation to reduce dimensionality and capture relationships between predictors inherently in the resulting variables - each varimax-rotated principal component corresponds to several highly correlated molecules.

4.2 Results

16 of the 17 clinical variables in this dataset were shown to be significantly associated with LVMI (Table 4-1, Fig. 4-1 and 4-2). A multiple linear regression model was also developed from these 17 variables. The leaps output (Fig. 4-3) demonstrates not only what the best model is in terms of adjusted R^2 and BIC, but also how well other models perform – whether the best model is significantly better or only marginally better than others by the selected criteria. By backwards

step selection the model with the greatest number of significant p values is also identified (Table 4-2).

All three models describe a similar amount of variance (when corrected for number of variables) and the models based on the number of significant variables and minimal BIC had a better and similar fit to the data and lower p values. Five terms were shown to be independently statistically significant, namely sex, presence of congestive heart failure (CHF), SBP, heart rate (HR) and BMI. This model was selected to be taken forward and used for modelling with molecular variables in 'Approach 1' below, however the factor CHF had to be dropped as there were few cases and in many instances there was an intersection between cases and missing molecular data.

Variable	Coefficient	p	R^2	Model p
Sex	8.18	2.68×10^{-4}	0.05	2.68×10^{-4}
Age	0.35	4.70×10^{-6}	0.08	4.70×10^{-6}
Age > 50 years	7.38	1.03×10^{-3}	0.04	1.03×10^{-3}
Hypertension	9.45	2.53×10^{-5}	0.07	2.53×10^{-5}
Diabetes	9.57	3.74×10^{-2}	0.02	3.74×10^{-2}
CAD	13.53	9.24×10^{-3}	0.03	9.24×10^{-3}
MI	18.71	2.36×10^{-2}	0.02	2.36×10^{-2}
CHF	24.51	7.76×10^{-3}	0.03	7.76×10^{-3}
BMI	0.88	6.88×10^{-5}	0.06	6.88×10^{-5}
BMI > 30 kg/m^2	7.41	1.47×10^{-3}	0.04	1.47×10^{-3}
SBP	0.35	1.12×10^{-11}	0.16	1.12×10^{-11}
DBP	0.39	7.78×10^{-5}	0.06	7.78×10^{-5}
HR	-0.30	4.85×10^{-3}	0.03	4.85×10^{-3}
Creatinine	0.14	1.20×10^{-2}	0.02	1.20×10^{-2}
eGFR	-0.07	2.52×10^{-1}	0.01	2.52×10^{-1}
BPMED	2.24	7.83×10^{-4}	0.04	7.83×10^{-4}
RAAS	8.70	1.40×10^{-4}	0.05	1.40×10^{-4}

Table 4-1: Simple Linear Regression of Clinical Variables.
16/17 variables were shown to be significantly predictive of LVMI. eGFR: Estimated glomerular filtration rate, RAAS: Renin-angiotensin-aldosterone system, BPMED: number of antihypertensive medications, RAAS: RAAS-blocking medications, BMI: body mass index, HR: heart rate, DBP: diastolic blood pressure, SBP: systolic blood pressure, CAD: Coronary Artery Disease, MI: Myocardial Infarction, CHF: Congestive Heart Failure.

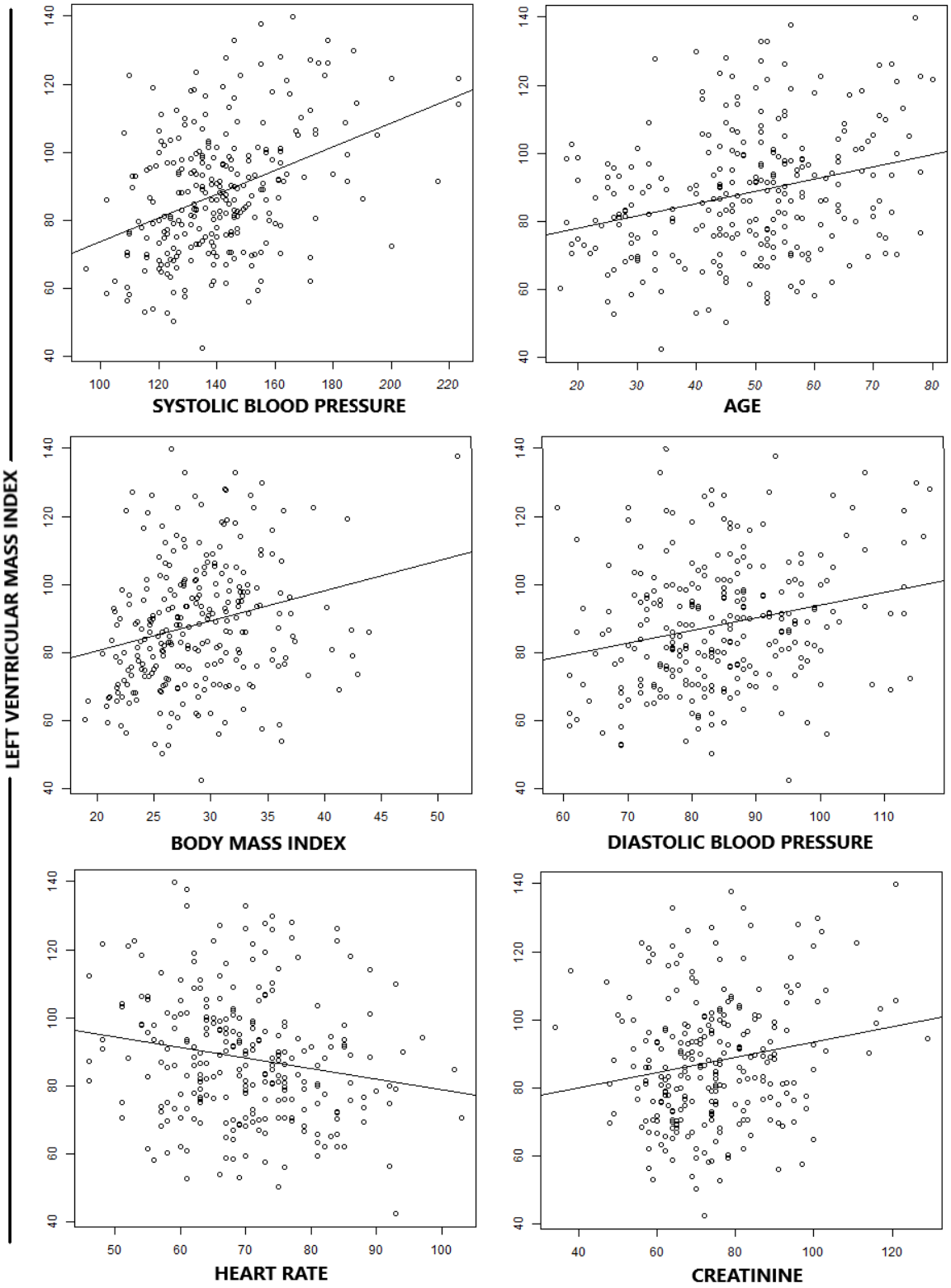

Fig. 4-1: Simple Linear Regression of Clinical Variables.
16/17 variables were shown to be significantly predictive of LVMI including six continuous variables: systolic blood pressure ($p = 1.12 \times 10^{-11}$), age ($p = 4.70 \times 10^{-6}$), body mass index ($p = 6.88 \times 10^{-5}$), diastolic blood pressure ($p = 7.78 \times 10^{-5}$), heart rate ($p = 4.85 \times 10^{-3}$), creatinine ($p = 1.20 \times 10^{-2}$). Variables are listed in order of statistical significance and match graphs reading left-to-right, top-to-bottom. LVMI is shown on the vertical axes. One data point with an implausible creatinine value of 300 was removed from the plot as its validity could not be verified.

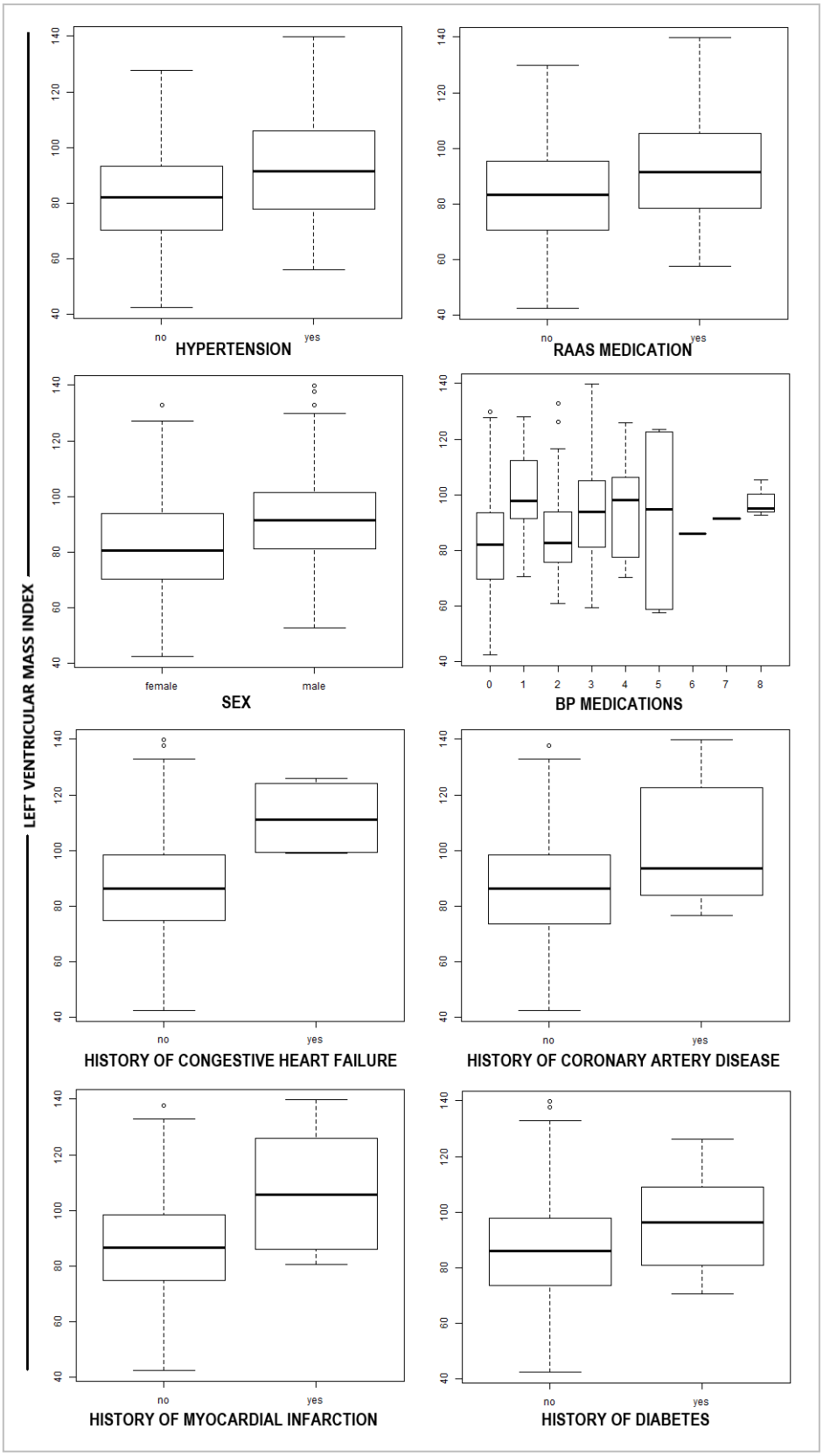

Fig. 4-2: Significant Categorical Clinical Variables.
16/17 variables were shown to be significantly predictive of LVMI including ten categorical variables: hypertension ($p = 2.53 \times 10^{-5}$), renin-angiotensin-aldosterone system medications ($p = 1.4 \times 10^{-4}$), sex ($p = 2.68 \times 10^{-4}$), blood pressure medication ($p = 7.83 \times 10^{-4}$), history of congestive heart failure ($p = 7.76 \times 10^{-3}$), history of coronary artery disease ($p = 9.24 \times 10^{-3}$), history of myocardial infarction ($p = 2.36 \times 10^{-2}$), history of diabetes ($p = 3.74 \times 10^{-2}$)

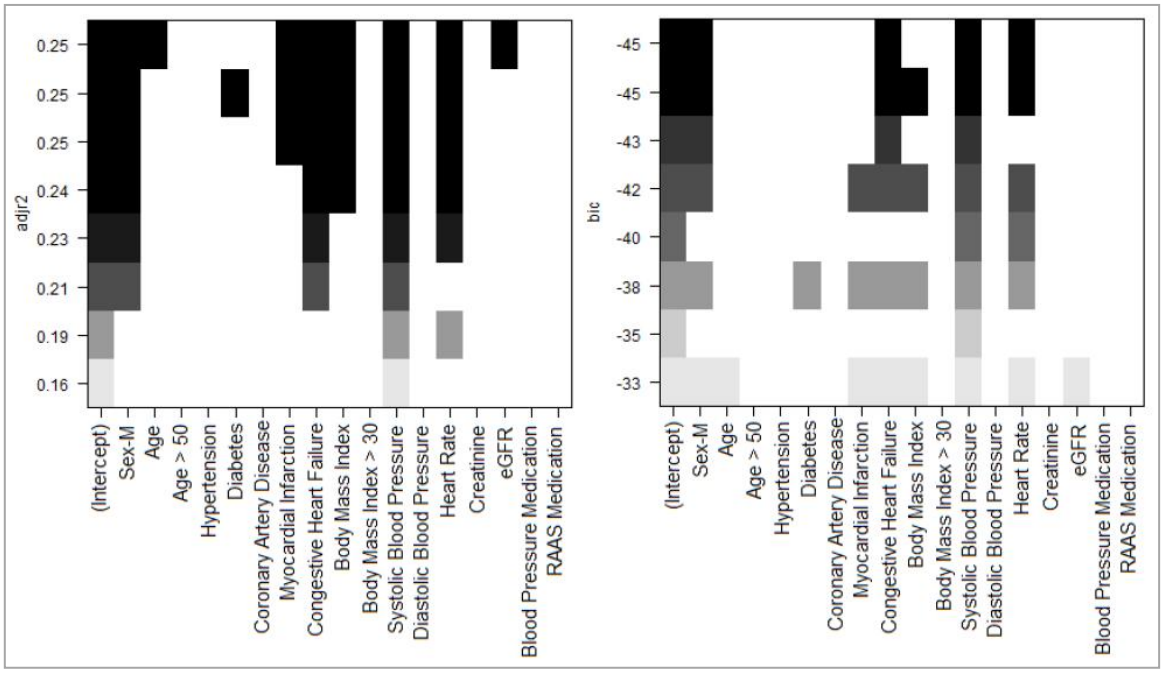

Fig. 4-3: Development of a Clinical Model using an All-Subsets Model Selection with the R Package Leaps.

The selection criteria used were the maximum adjusted R^2 (adjr2) and the minimum Bayesian Inference Criterion (BIC). Each row describes a model, and colouring indicates the inclusion of variables in each model. Where directional model selection methods are heuristic the all-subsets approach runs all possible combinations of variables, and it can be seen from the output how the top models compare.

Criterion	Terms	Adj-R^2	BIC	Model p
Max Terms p < 0.05	sex, CHF, SBP, HR, BMI	0.24	-45	4.02×10^{-15}
Max Adj-R^2	sex, age, MI, CHF, SBP, HR, BMI, eGFR	0.24	-33	2.44×10^{-14}
Min BIC	sex, CHF, SBP, HR	0.23	-45	8.74×10^{-15}

Table 4-2: Modelling with a Combination of an All-subsets Approach and Backwards step selection.

The R package leaps determined the models of lowest Bayesian Information Criterion (BIC) and highest adjusted R^2. Backwards step selection was run on the maximal adjusted R^2 model to find the model with the maximum number of significant terms. All three of these models describe a similar amount of variance (adjusted for number of terms in the model). The maximal terms model and best fit model have a lower fit and more significant model p value than that of the maximal adjusted R^2 model. eGFR: Estimated glomerular filtration rate, CHF: Congestive Heart Failure, SBP: systolic blood pressure, HR: heart rate, BMI: body mass index, MI: Myocardial Infarction.

4.2.1 Approach 1: Testing Molecular and Clinical Variables Separately then Combining the Data

All constituent datasets contained at least one significant molecule at the screening step except the Randox Cytokine Array. Positive associations with LVM were found (Table 4-3) with two identified plasma peptides (Angiotensin A, Angiotensin II), 6 unidentified plasma peptides (m/z: 266.16, 288.13, 567.14, 834.18, 831.18, 553.10), two circulating miRNAs (has-miR-18a-3p, has-miR-92b-3p), two myocardial remodeling markers (PICP:CITP ratio, PICP), three serum metabolite signals (D-glucose, D-glucose + L/D-proline, 'D-glucose + alanine + glutamine') and one urinary metabolite (phenylacetylglycine). Negative associations were found with two serum metabolite signals (trimethylamine, 'unsaturated fatty acids + isoleucine + L-proline'), one unidentified urinary metabolite and five urinary peptides (a Haemoglobin Subunit Beta peptide, a Collagen Alpha-1(III) Chain peptide and three Collagen Alpha-1(I) Chain peptides). For these molecules the coefficients were consistently signed regardless of the other variables included in the model.

Molecule	Dataset (size)	Simple Linear Regression			With Confounders (Age, Sex, BMI)			With Clinical Model (Sex, SBP, HR, BMI)		
		Std. Beta	p_{BH}	Adj. R^2	Std. Beta	p_{BH}	Adj. R^2	Std. Beta	p_{BH}	Adj. R^2
Angiotensin II	IPP (3)	0.18	1.00E-01	0.01	**0.28**	**5.91E-04**	**0.14**	**0.22**	**4.89E-03**	**0.25**
Angiotensin A	IPP(3)	0.17	1.00E-01	0.01	**0.24**	**5.91E-04**	**0.15**	**0.19**	**4.89E-03**	**0.25**
Unknown Peptide 266.16m/z	UPP (18)	**0.43**	**1.05E-03**	**0.17**	0.29	9.36E-02	0.23	**0.33**	**1.03E-02**	**0.30**
Unknown Peptide 288.13m/z	UPP (18)	**0.30**	**4.18E-02**	**0.05**	0.10	9.38E-01	0.17	0.15	5.77E-01	0.28
Unknown Peptide 567.14m/z	UPP (18)	**0.37**	**4.18E-02**	**0.16**	0.16	6.79E-01	0.26	0.13	7.29E-01	0.39
Unknown Peptide 834.18m/z	UPP (18)	**0.32**	**4.18E-02**	**0.11**	0.16	8.17E-01	0.29	0.27	1.68E-01	0.38
Unknown Peptide 831.18m/z	UPP (18)	**0.31**	**4.18E-02**	**0.20**	0.21	6.79E-01	0.31	0.09	8.51E-01	0.19
Unknown Peptide 553.10m/z	UPP (18)	**0.32**	**3.06E-02**	**0.08**	0.16	6.79E-01	0.08	0.21	3.00E-01	0.21
hsa-miR-18a-3p	miRNA (5)	**0.23**	**4.30E-03**	**0.05**	0.06	9.76E-01	0.12	0.01	9.50E-01	0.21
hsa-miR-92b-3p	miRNA (5)	**0.18**	**2.60E-02**	**0.03**	0.03	9.76E-01	0.11	0.00	9.50E-01	0.20
PICP/CITP (collagen turnover)	MRM (4)	0.14	7.72E-02	0.01	**0.17**	**1.62E-02**	**0.21**	0.13	5.47E-02	0.31
PICP	MRM (4)	0.13	7.72E-02	0.01	**0.14**	**3.41E-02**	**0.20**	0.11	8.34E-02	0.32
isoleucine + L-proline + unsat f.a.	Serum Met (50)	**-0.27**	**1.31E-02**	**0.06**	-0.14	2.74E-01	0.14	-0.15	5.40E-01	0.26
D-Glucose	Serum Met (50)	**0.24**	**1.71E-02**	**0.04**	0.11	4.22E-01	0.13	0.10	5.40E-01	0.22
D-Glucose + L/D-Proline	Serum Met (50)	**0.27**	**1.71E-02**	**0.04**	0.15	3.82E-01	0.15	0.12	5.40E-01	0.25
D-Glucose + alanine + glutamine	Serum Met (50)	**0.27**	**1.42E-02**	**0.05**	0.10	4.82E-01	0.14	0.07	5.70E-01	0.23
Trimethylamine	Serum Met (50)	**-0.24**	**1.71E-02**	**0.05**	-0.06	5.50E-01	0.14	-0.08	5.40E-01	0.23
'Unknown8'	Urinary Met (168)	**-0.35**	**9.49E-03**	**0.07**	-0.19	4.94E-01	0.17	-0.14	3.40E-01	0.27
Phenylacetylglycine	Urinary Met (168)	**0.47**	**8.20E-02**	**0.03**	**0.50**	**2.15E-02**	**0.19**	**0.46**	**1.99E-02**	**0.29**
Collagen alpha-1(I) chain (50172)	Urinary Pep (1340)	**-0.28**	**2.26E-02**	**0.06**	-0.18	5.24E-01	0.17	-0.17	1.33E-01	0.27
Collagen alpha-1(III) chain (84440)	Urinary Pep (1340)	**-0.28**	**2.26E-02**	**0.06**	**-0.24**	**2.17E-02**	**0.21**	-0.17	1.33E-01	0.30
Haemoglobin subunit beta (110333)	Urinary Pep (1340)	**-0.30**	**2.75E-02**	**0.06**	-0.24	1.41E-01	0.22	-0.22	1.33E-01	0.31
Collagen alpha-1(I) chain (18393)	Urinary Pep (1340)	**-0.32**	**3.07E-02**	**0.05**	-0.17	5.24E-01	0.20	-0.12	4.93E-01	0.31
Collagen alpha-1(I) chain (124886)	Urinary Pep (1340)	**-0.23**	**3.75E-02**	**0.05**	-0.14	5.24E-01	0.18	-0.13	3.55E-01	0.31

Table 4-3: Results from simple linear regression and multiple linear regression of molecular variables with respect to left ventricular mass index (LVMI).

Coefficients have been standardised to aid comparison between variables of different scales and units. The value resulting from this standardization ("Std Beta") represents the number of standard deviations of LVMI0.5 increase/decrease with every 1 SD increase/decrease of the predictor. Associations with plasma peptides Angiotensin II and Angiotensin A and urinary metabolite phenylacetylglycine would not have been detected with simple linear regression but strong associations were found with multiple linear regression. A confounder model (age, sex, body mass index) and a clinical model (sex, body mass index, systolic blood pressure, heart rate) were used for multiple linear regression. IPP: identified plasma peptides, UPP: unidentified plasma peptides, MRM: myocardial remodeling markers, Std. Beta: standardized beta, p_{BH}: Benjamini-Hochberg corrected p value, Adj-R^2: adjusted R^2, BMI: body mass index, HR: heart rate, SBP: systolic blood pressure, PICP: carboxy-terminal propeptide of procollagen type I, CITP: carboxy-terminal telopeptide of collagen type I.

These results were taken forward and used as a basis for cluster/correlation analysis (Fig. 4-4) and multiple linear regression. Correlations performed between each variable were organized into a correlation matrix. The columns/rows of this matrix were sorted by the resulting dendrogram from the hierarchical clustering analysis, to aid visual inspection of the data. These results helped group the variables into three clusters, the three clades at the root of the graph.

The largest cluster is cluster C and it includes LVMI. This cluster contains 133 significant positive correlations and only one significant negative correlation – between diastolic blood pressure and coronary artery disease. Cluster C clusters closely with cluster B (phenylacetylglycine, angiotensin II and angiotensin A) whose pattern of correlation is more complex. While all significant correlations between phenylacetylglycine and cluster C are positive, the Angiotensin molecules are positively related to LVMI, but negatively related to miR-18a-3p, miR-92b-3p, D-glucose, 'Sex-M' and creatinine. Strong correlations are detected between the two Angiotensin molecules, between two miRs (miR-18a-3p, miR-92b-3p), and between three serum metabolites (D-Glucose, Met.S38, Met.S42).

Cluster Acontains nine significant positive correlations and two significant negative correlations, most of which are weak. Three serum metabolites (trimethylamine, Met.S17, Lipids) exhibit strong positive correlations with each other. Significant correlations between the variables of Cluster A and Cluster C are mostly negative (count: 93). Strong anti-correlations include those between two sets of serum metabolites (Cluster A: trimethylamine, Met.S17, Lipids; Cluster C: D-Glucose, Met.S38, Met.S42), and between PICP/CITP and both Angiotensin molecules. 13 correlations were positive including the correlations between diastolic blood pressure and heart rate, and between Sex-M and eGFR. 'PICP/CITP' and 'Hx of CHF' were consistently positively correlated with cluster D and are responsible for the remaining 11 positive correlations.

A dataset with recurring points of high influence removed was used for multiple linear regression analysis (Table 4-4). Using molecular variables alone it was determined that four molecules are independently predictive of LVMI: collagen type I alpha 1 peptide (internal ID: 50172), phenylacetylglycine, D-glucose and

Angiotensin A (from urine, urine, serum and plasma respectively). The clinical modelling based on this slightly smaller dataset results in a similar model to the maximal significant variables model identified above, however BMI as a predictive variable was slightly over the selected significance threshold.

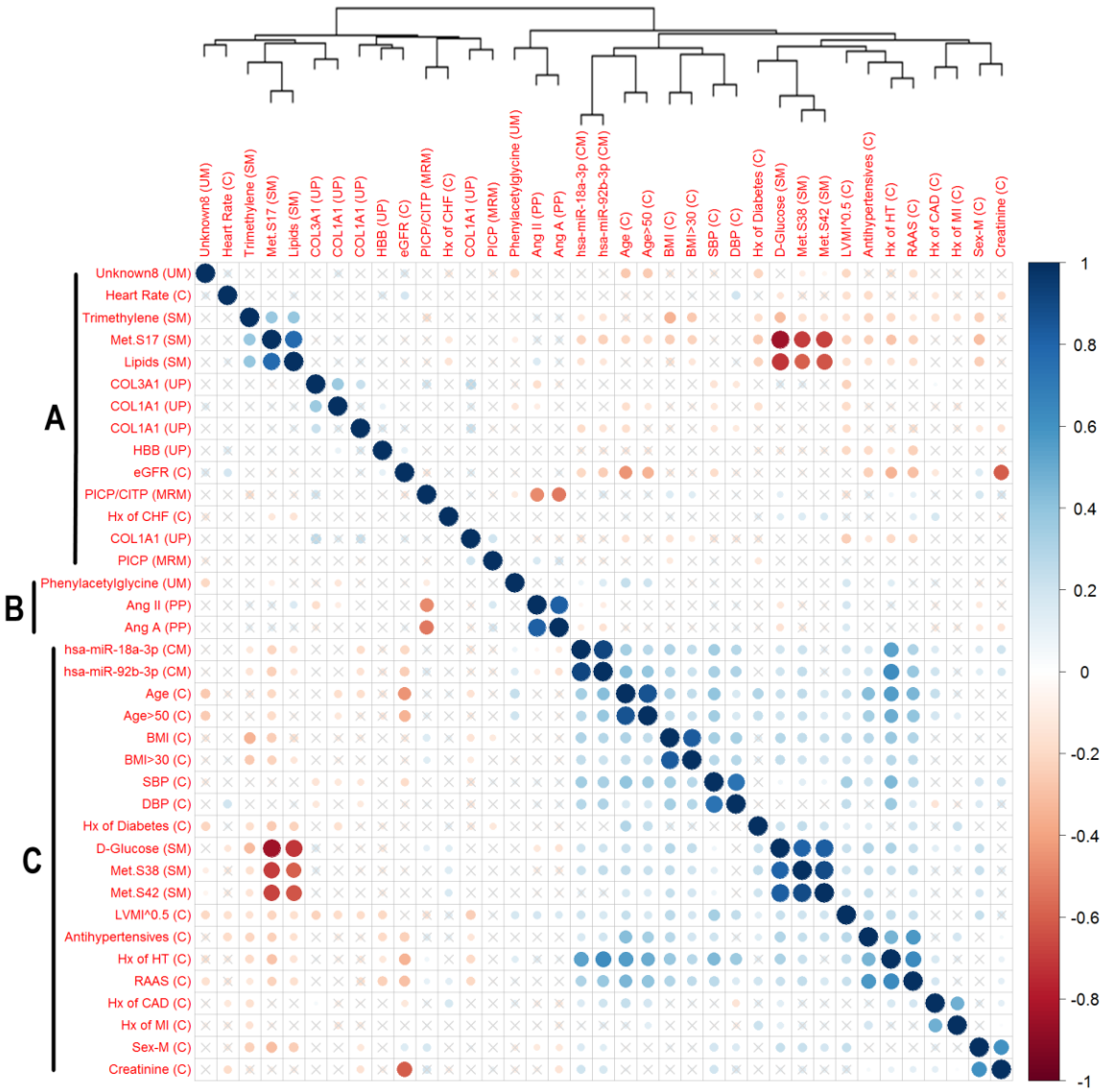

Fig. 4-4: A Clustered Correlation Map of the Variables Identified in the Screening Step of Approach 1.
This figure shows significant correlations between molecular and clinical variables ordered by hierarchical clustering. Many expected correlations are shown between variables, e.g. Angiotensin A and Angiotensin II are strongly positively correlated. Associations with non-specific metabolite signals can be derived: MetS17 (unsaturated fatty acids); MetS38 and MetS42 (glucose). The unsaturated fatty acids, lipids, and trimethylamine cluster is in anticorrelation with the glucose cluster. This relationship may be due to control by insulin which suppresses FMO3 expression. The 'lipids' metabolite was not significant, but added to assess its correlation with significant variables and aid in understanding other metabolite signals. A blue-red colour scale for correlation rho values is shown in the legend to the right. Grey crosses indicate insignificant correlations.

Criterion	Terms	Adj-R^2	BIC	Model p
Clinical	sex, CHF, SBP, HR	0.20	-33	3.61×10^{-12}
Molecular	collagen type I alpha 1 peptide (50172), phenylacetylglycine, D-glucose, angiotensin A	0.20	-9.4	1.30×10^{-7}
Clinical and Molecular	collagen type I alpha 1 peptide (50172), haemoglobin beta peptide (110333), phenylacetylglycine, angiotensin A, sex, SBP	0.29	-25	1.99×10^{-12}

Table 4-4: Modelling with Clinical Variables and Significant Molecular Variables.
The model using both molecular and clinical variables describes approximately 9% more variance than those using either dataset alone. Models were selected here on the criterion of having the maximum number of independently predictive variables. CHF: congestive heart failure, SBP: systolic blood pressure, HR: heart rate, Adj-R^2: adjusted R^2, BIC: Bayesian information criterion.

When both molecular and clinical terms were used as predictors the emerging model consisted of: collagen type I alpha 1 peptide (internal ID: 50172), haemoglobin beta peptide (internal ID: 110333), phenylacetylglycine, angiotensin A, sex and SBP. Using both clinical and molecular variables in combination it could be seen that there is a greater proportion of the data described – an adjusted R^2 of 0.20 for either set alone and 0.29 for the combined model.

Data imputation results in a similar combined model describing a similar amount of variance but with a better fit, lower model p value and more independent predictors (Table 4-5). The two models share five variables and both include an angiotensin molecule. While SBP and HR are found to be significant predictors in the second model this slightly different approach does not yield more variance explained after adjustment for the number of terms included. Logistic regression with LVH categorisation, while underpowered, showed six of the molecules identified in the linear regression screening step (i.e. shown in Table 4-2) were also significant for LVH and that SBP, Angiotensin A and phenylacetylglycine were independently significant.

Criterion	Terms	Adj-R^2	BIC	Model p
Clinical	sex, CHF, SBP, HR	0.20	-33	3.61×10^{-12}
Molecular	collagen type I alpha 1 peptide (50172), collagen type III alpha 1 peptide (84440), phenylacetylglycine, D-glucose, angiotensin II, 533.10m/z peptide	0.18	-24	5.21×10^{-10}
Clinical and Molecular	collagen type I alpha 1 peptide (50172), haemoglobin beta peptide (110333), phenylacetylglycine, angiotensin II, sex, SBP, BMI, HR	0.29	-47	2.20×10^{-16}

Table 4-5: Modelling with Clinical Variables and Significant Molecular Variables Using Imputed Data.
Using K-nearest neighbour and mean imputation missing data points were imputed. This allows the resulting molecular model to include one of the terms with poor representation (533.10m/z peptide). The resulting combined model describes a similar amount of data to that from the non-imputed dataset, but associating one more independent predictor and has a better fit and lower model p value. CHF: congestive heart failure, SBP: systolic blood pressure, HR: heart rate, Adj-R^2: adjusted R^2, BIC: Bayesian information criterion.

4.2.2 Approach 2: Combine Variables Prior to Testing

Four patients and six variables were screened based on having a high proportion of missing data, and missing data in the remaining variables were imputed. The imputation of missing data can alter datasets, changing distributions and artificially reducing variance. The more data points are imputed for a particular variable the worse these problems get hence the removal of some variables altogether. While not perfect this screening and imputation step is necessary for the following statistical analysis which is incompatible with missing data. Initial PCA clearly showed that a large proportion of the data was mainly describing three extreme outliers, and therefore these were removed (Fig. 4-5). It seems likely that these outliers are so different due to some unknown technical effect. If not then it could be argued that the analysis is being biased by their removal, however the results will still reflect the vast majority of the sample – and they will reflect the majority better un-skewed by the outliers.

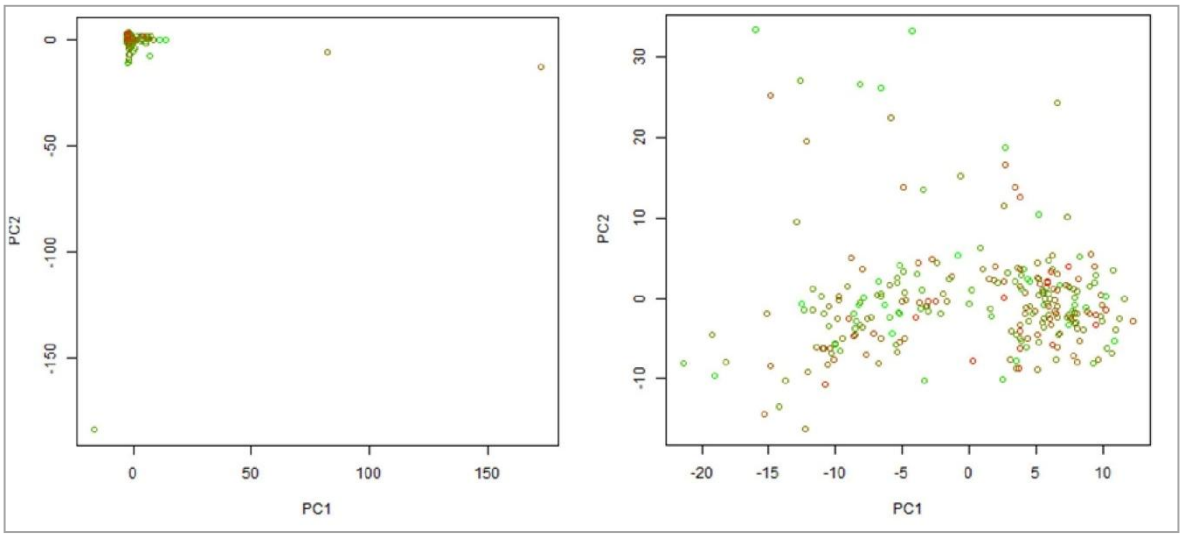

Fig. 4-5: Screening Extreme Outliers with Principal Components Analysis.
The first two principal components (PC1 and PC2) are shown. This initial principal components analysis (PCA) revealed that a large proportion of the variation discriminated only three outliers so these three points were removed for the subsequent PCA analysis. Red = high LVMI, green = low LVMI.

219 components were selected from a subsequent PCA analysis by the Kaiser criterion (257) and those components were varimax-rotated. Four were found to be significantly associated with LVMI. These four components correlate with many clinical and molecular variables associated in the first approach (Fig. 4-6). All four are independently predictive and this model outperforms using only the

clinical or only the molecular data from approach 1, but is itself outperformed by the combined model from approach 1 (Table 4-6). PC44, PC204 and PC209 were all found to be significantly associated with left ventricular hypertrophy by simple logistic regression.

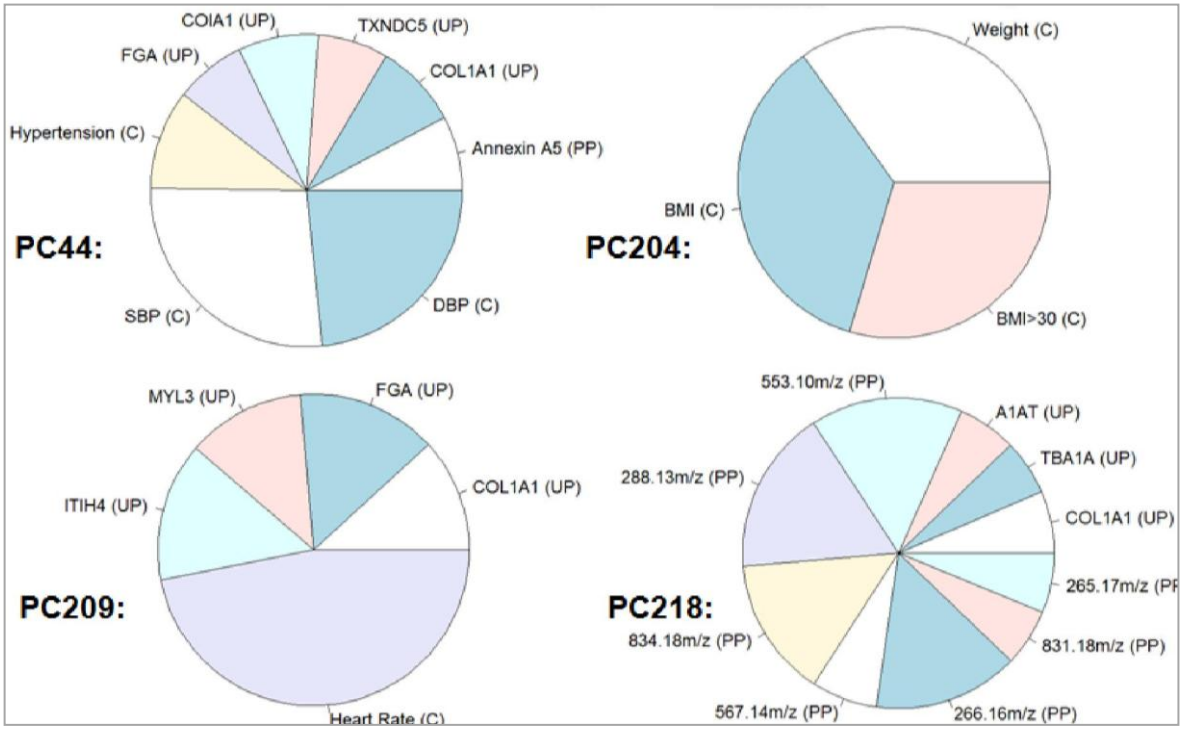

Fig 4-6: Contributions of Original Variables to Significant Varimax-Rotated Components. Each pie segment represents a correlation of 0.1 or greater and those with a lower correlation are removed for clarity. They are independently predictive and seem to be related to blood pressure, BMI, heart rate, and the unidentified plasma peptides (PC 44, PC204, PC209, PC218 respectively). Contributions to these come from only three of the constituent datasets and there are several weak contributions (i.e. slightly above the inclusion threshold) from the largest of these (urinary peptides). C: clinical, UP: urinary peptides, PP: plasma peptides.

PC	Estimate	Standard Error	p	p_{BH}	Adj. R^2	model p
PC44	0.16	0.03	5.48×10^{-8}	1.20×10^{-5}	0.11	5.48×10^{-8}
PC204	-0.15	0.03	2.59×10^{-7}	2.84×10^{-5}	0.10	2.59×10^{-7}
PC209	0.16	0.03	1.39×10^{-6}	1.01×10^{-4}	0.09	1.39×10^{-6}
PC218	0.13	0.04	3.47×10^{-4}	1.90×10^{-2}	0.05	3.47×10^{-4}
All four PCs	-	-	-	-	0.21	8.13×10^{-13}

Table 4-6: Linear Regression was used to test for Associations Between LVMI and Varimax-Rotated Principal Components.
The results of simple linear regression analysis of each individual PC is shown along with those of a multiple linear regression with all four PCs wherein all four PCs are significant. p_{BH}: Benjamini-Hochberg corrected p value, Adj-R^2: adjusted R^2.

4.3 Discussion

4.3.1 Combined Molecular and Clinical Model

Using both molecular and clinical variables in multiple linear regression describes approximately 29% of the variation in LVMI. This model contains collagen type I alpha 1 peptide (50172), haemoglobin beta peptide (110333), phenylacetylglycine, angiotensin A, sex, SBP. Sex and SBP were determined to be important independent clinical predictors in the initial clinical model along with BMI, HR and CHF. There were few cases of the fifth clinical variable (CHF) in our data which made analysis more complicated – when some molecular variables were tested along with the clinical model, missing molecular data coincided with CHF cases meaning that CHF had to be dropped from the model in those particular tests. This also means that its presence in the clinical model vs its absence in mixed clinical-molecular model should not be over-interpreted – perhaps if there were a few more cases the molecular variables would not have 'displaced it' from the model. Interestingly when missing data is imputed the same combined model is derived but with the addition of remaining two clinical variables, BMI and HR, suggesting that all of the molecular variables in our final model are independently predictive of the clinical model (albeit it with the alternative but highly correlated angiotensin II in place of angiotensin A). Thus the molecular variables can be shown to add variation to LVMI which cannot be explained by clinical data alone and vice versa, assuming the data imputation is reliable. This indicates that the molecular processes of LVH underlying these molecular variables are at least partially distinct from the molecular processes related to the clinical variables. Clinical variables such as blood pressure clearly drive the development of LVH, however not all hypertensive patients develop LVH. Molecules which describe variation which is distinct from easily-measured well-understood clinical measures are perhaps the most promising for discovery of novel mechanisms to describe the differences between those patients which are susceptible to development of LVH.

Urinary phenylacetylglycine appears to be the most important molecular predictor which has been identified. It was found to be positively associated, has

the greatest coefficient of the significant results, is predictive of LVMI independently of the clinical model, and its addition to the clinical model used for the screening step of Approach 1 explains approximately the same amount of variation in LVMI (when adjusted for the number of variables in the model). It is associated with heart failure (258) and is a biomarker for phospholipidosis (Fig. 4-7). Phospholipidosis is the accumulation of abnormally high levels of phospholipids in lysosomes or cytoplasm and it can be induced by many commonly prescribed cationic amphiphilic drugs - which includes drugs of many classes used to treat a wide range of conditions (259). In drug-induced phospholipidosis the inhibition of lysosomal phospholipase leads to the formation of large indigestible complexes which are visible with electron microscopy. Organs affected by phospholipidosis show signs of inflammation and histopathological changes (260), and there are associations with various modulations to immune response (261). In cardiac tissue this process may lead to cell hypertrophy and fibrosis (262). Furthermore anzoline-induced renal phospholipidosis has been suggested as a cause of heart failure in a case study (263).

Phospholipidosis can occur in many organs so another possible connection with LVH may be found elsewhere in the body. Gentamicin-induced renal phospholipidosis has been demonstrated to have a causal link to nephrotoxicity (264-266). This may explain the association with heart failure since eGFR is associated with LVM (267-269), notably in a longitudinal study of living kidney donors where reduced renal function appeared to be causative for increased LVM (270). However if urinary phenylacetylglycine were indicating reduced kidney function in this study we would not have expected so little correlation between it and eGFR. Alternatively the association may not underlie a causative role for development of LVM, rather it could be a biomarker of LVH or simply of medication use – e.g. those at high risk of developing LVH may already be using certain medications which are capable of inducing phospholipidosis, such as propanolol. There is also some debate whether it truly is phenylacetylglycine or phenylacetylglutamine being detected in the studies described - while phenylacetylglycine has been detected in human urine using NMR but could not be detected using MS which has a higher sensitivity and is generally considered

to be more reliable (271) – in this case however phenylacetylglycine was detected with MS, suggesting the NMR results were accurate.

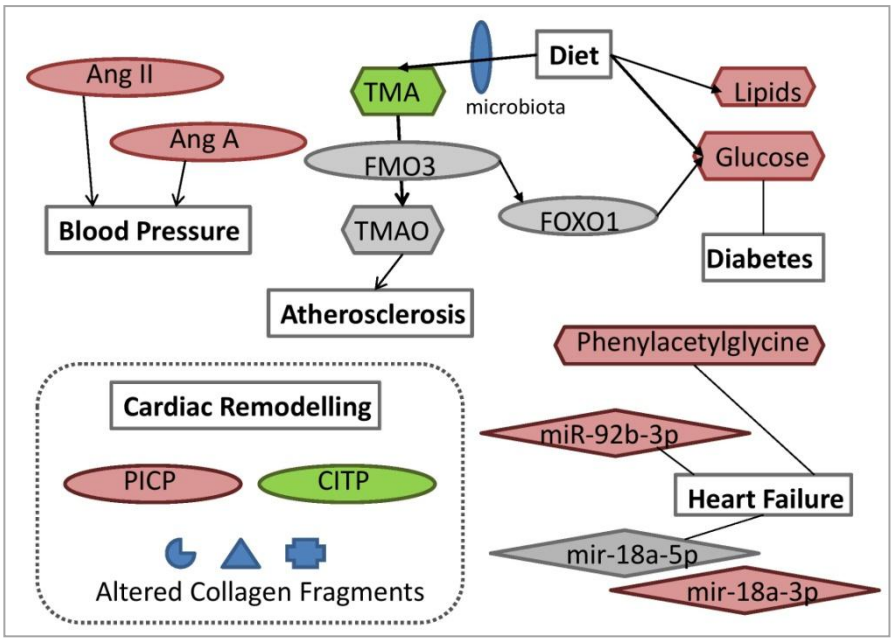

Fig. 4-7: A Summary of the Molecular Variables Identified by the Screening Step of Approach 1 and how they Relate to Various Relevant Clinical Terms.
Unsat: unsaturated, Ang: angiotensin, TMA: trimethylamine, TMAO: trimethylamine oxide, FMO3: flavin containing monooxygenase 3, FOXO1: forkhead box protein O1, PICP: collagen I propeptide, CITP: C-telopeptide for type I collagen. Circle: protein, diamond: miRNA, hexagon: metabolite, red: upregulated, green: downregulated, blue: up/downregulated in different forms.

Angiotensin II is well known to increase blood pressure and has been shown in rats to have a blood pressure independent effect on LVM (272) – an association also demonstrated in humans in the data presented here. In hypertensive patients blockers of the renin-angiotensin system have been shown to be more effective in producing reduction in LVM than beta-blockers (273;274). Angiotensin II also drives hypertrophy via inflammation and oxidative stress, and by interactions with the sympathetic nervous system (275). Angiotensin A is identical to angiotensin II but for a single amino acid difference and it is formed via the enzymatic decarboxylation of an aspartic acid residue of Angiotensin II. While the effects of angiotensin A appear to be very similar to angiotensin II, it has a higher affinity for the angiotensin II receptor type 2 receptor (276). The same study by Jankowski et al found the relative abundance of angiotensin A to angiotensin II to be 9.7% in plasma samples from a group of healthy human controls. Angiotensin A was only first described in 2007 and research in humans in very limited - most information about its function so far is drawn from animal models. In rats it has very similar cardiovascular effects to angiotensin II,

including an increase in blood pressure and vasoconstriction, and a decrease in cardiac contractility and heart rate (277). Also similarly to angiotensin II its effects are partially reduced by an angiotensin II receptor type 1 antagonist (278). Angiotensin A is less effective at inducing vasoconstriction compared to angiotensin II. This is probably due to the vasodilatory effects it has through (i) its increased affinity for angiotensin II receptor type 2, and (ii) production of alamandine which binds with 'mas-related G-protein coupled receptor D', increasing levels of atrial natriuretic peptides (279;280). Separately angiotensin II and angiotensin A were both predictive of LVMI, but since they were so highly correlated in our data no reliable inference can be made based on which one of them is included in the combined model.

The two remaining molecules in the combined model were a urinary collagen (type I) peptide and a urinary haemoglobin peptide. The inclusion of a haemoglobin beta peptide in the model may represent the causal link between anaemia and left ventricular mass in patients with chronic kidney disease and patients with end-stage renal failure – treatment with erythropoietin regresses LVH independent of blood pressure (281;282).

Four of the significant urinary peptides were collagen peptides, one of which was significant independent of common confounders, and another is described above as part of the combined model. Cardiac fibrosis is a process which occurs during the development of LVH and it is characterised by an increase of collagen and other extracellular matrix components. The composition of the extracellular matrix has an influence on heart function, and the accumulation of collagen can cause stiffness and impairs function. Collagen build-up can occur by an increase in procollagen peptide gene expression, a reduction in collagenases, or by an increase in collagenase inhibition. Collagen I and collagen III are the main myocardial collagens, both of which can be synthesised at an increased rate in patients with LVH (283;284). Collagen type I synthesis is driven partly by angiotensin II, and has been associated with hypertensive heart failure (285;286). The significant correlations between angiotensin II and two collagen peptides in our data support this. While angiotensin A has significant correlations with most of the same variables (PICP/CITP, miR-18a-3p, miR-92b-3p, glucose, sex), it is not significantly correlated with the collagen peptides. While the

collagen peptides identified here are inversely correlated this may be representative of reduced activity of a specific protease(s) rather than a decrease in the parent molecule itself - different peptidases which can cleave collagen alter the levels of these peptides. A reduced level of collagenase would lead to collagen build-up as part of the process of cardiac fibrosis. The activity of certain peptidases, especially matrix metalloproteinase proteins – many of which are already known to contribute directly towards LVH – may reflect predisposition to LVH (287). The expression of particular known myocardial collagenases such as MMP-1, MMP-8, MMP-9, and MMP-13 could be tested and compared against the collagen fragments data to determine which collagenase or combination of collagenases is responsible for a particular collagen fragment. The collagen fragments are positively but weakly correlated. While some of the correlations are not significant this would suggest that more than one collagenase is differentially expressed.

4.3.2 Additional Insights

Other molecules identified in the screening step are also worthy of discussion because: (i) they may interact with molecules described above which is why they add nothing to the statistical model but may be interesting from a biological function/pathways perspective, (ii) they may not describe enough variation in the sample to be significant with this sample size but be more relevant in other samples/populations or have a strong influence in certain individuals, (iii) high rates of missing data can be misleading, resulting in reduced power especially when modelling alongside other more complete variables, (iv) relationships between variables within the study may generally add to knowledge from the literature (v) they may be highly correlated with a well-described clinical variable, so while they may not describe differences in proneness to LVH in isolation they could add to the understanding of a clinical variables relationship with LVMI.

miR-18a-5p is negatively associated with age-related heart failure in mice and has been shown to repress anti-angiogenic factors connective tissue growth factor (CTGF) and thrombospondin-1 (TSP-1) (288). We found the other (3p) arm to be positively associated, indicating that ‘arm switching’ is responsible for the

decrease in miR-18a-5p rather than differential expression of the precursor (289;290). The other miRNA hsa-miR-92b-3p is positively associated with heart failure in human (291).

Of the myocardial remodelling markers procollagen type I (PICP) – a precursor to mature collagen and therefore indicative of collagen synthesis – was the only significant molecule. PICP:CITP ratio has been previously suggested as an index of the coupling between collagen synthesis and collagen degradation (i.e. with maximum coupling at approximately 1 and net synthesis and degradation above and below this respectively) (292) and been shown to correlate to changes in left ventricular stiffness upon treatment with angiotensin II receptor blockers (293). LVH regression induced by blockers of the renin-angiotensin system has been found to be associated with a significant decrease in echocardiographic indexes of myocardiac fibrosis (294;295). PICP:CITP was calculated and found to be more significant and with a higher coefficient (both positive) and adjusted R^2 than PICP alone, showing collagen turnover to be a more important indicator of LVMI than collagen synthesis. Interestingly while PICP is significantly positively correlated with angiotensin II – as previously shown (296) – the PICP:CITP ratio is more strongly correlated and unexpectedly in the opposite direction.

Several studies have shown blood glucose to be associated with LVMI. However, in a well-powered study it was shown that significance does not persist when accounting for BMI (297), which is correlated with blood glucose levels. This indicates that blood glucose may not be important itself, but rather is acting as an indicator of BMI which is a known predictor. Our results support both the association between glucose and LVMI, and the result that shows it is not predictive independently of BMI. One significant metabolite was resolved to D-glucose alone, however another two were compound measures of several metabolites – the signal representing several molecules combined with energetically similar hydrogens – but included D-glucose and cluster and correlate highly with the glucose signal, so the most parsimonious explanation is that the association found with all three measures can be attributed to D-glucose. Similarly another serum metabolite association is unresolved, with a value representing levels of isoleucine, L-proline and unsaturated fatty acids. Using clustering we can see that the variations in this molecule closely resemble

that of a (non-significant) measure of lipids which suggests that the association detected is with the unsaturated fatty acids (rather than isoleucine or L-proline), as we might expect from the literature (298).

Trimethylamine (TMA) is produced by gut bacteria from dietary choline, phosphatidylcholine, and carnitine. In the liver it is oxidised by FMO3 into trimethylamine-n-oxide (TMAO) which promotes development of atherosclerosis (299) and is associated with heart failure (300) and poorer prognosis in chronic heart failure patients (300;301). We detected a negative LVMI association with molecules with TMA, perhaps due to increased FMO3 activity converting TMA to TMAO. FMO3 varies greatly between individuals (302) and FMO3 knock-out in insulin-resistant mice prevents atherosclerosis (303). TMA clusters with lipids and unsaturated fatty acids. This may be due to the effect of FMO3 on lipid metabolism or microbiota composition altering in response to dietary changes (304) and affecting TMA production and metabolism (305;306). The unsaturated fatty acids, lipids and TMA cluster (Fig 4-4) is in anticorrelation with the glucose cluster. This relationship may be due to control by insulin which suppresses FMO3.

Some peptides which were found to be associated with LVMI could not be identified and so they can only be reported rather than interpreted. An unidentified urinary metabolite, ‘Unknown8’, is negatively associated with LVMI and clusters most closely with heart rate and has significant negative correlations with several variables. Six of the 18 unidentified peptides tested were shown to be significantly related to LVMI, in particular the one with an m/z value of 266.16, which is robust to the addition of the clinical variables; in addition to it being independently predictive of the selected clinical variables it also has one of the greatest coefficients. The identity of this molecule is currently unknown. This molecule is not found in our final models however this may be due to the high frequency of missing data associated with these unidentified peptides relative to the rest of the data. Another unidentified peptide which stands out is ‘831.18m/z’ which has an adjusted R^2 approximately equal to that of the clinical model and the molecular model.

4.3.3 Comparison of Approaches and Limitations of Study

The results from the second approach are similar to the first however less comprehensive and with some additional urinary peptides. In general it seems they can largely be described as related to blood pressure, BMI, heart rate, and the unidentified plasma peptides (PC44, PC204, PC209, PC218 respectively). As we can see there is a tendency towards weak contributions (slightly above the set rho>0.1 inclusion threshold) from the largest constituent dataset.

Missing data is problematic for performing certain statistical tests (including linear regression and PCA), particularly if it is found dispersed rather than in certain samples or variables that should obviously be removed. In linear regression individual missing data points may be excluded or imputed – the former results in differently-powered tests for different variables and the latter introduces a small violation of the linear regression assumption of independent observations and potentially introduces erroneous data. For the screening step of Approach 1 the primary aim was not to compare results so removal was preferable. For the modelling step both methods of dealing with missing data were employed, and while there were some differences in results the final models were very similar. For Approach 2 data had to be imputed for PCA analysis, however the multivariate nature of this method (the resulting principal components are composed of data from several variables) means that erroneous imputation should have less impact on results.

Some molecules found to be associated with LVM were not identified however all of those which were in the final linear models were identified except the '533.10m/z' peptide. While it was not the aim of this paper to make these identifications the identities of those molecules corresponding to two clusters of metabolites could be speculated and supported with statistics from the clustering and correlation analysis of Approach 1. Perhaps with the introduction of another data set with a wider genomic scope such as gene expression microarray or RNA-seq similar speculations about unknown peptides could be made and tested. While no replication or validation is available for these results,

the potential role in the pathophysiology is discussed and it is supported by the published literature.

4.3.4 Conclusions

Our modelling shows that the combination of clinical and molecular data can explain several times the variance of most individual biomarkers, and approximately an additional 10% of total variance than that of either set (clinical or molecular). While several molecular markers are included in the combined models, a similar amount of variance can be significantly explained by clinical measures combined with just urinary phenylacetylglycine, which stands out as the most useful identified biomarker of the dataset and was also shown to be significantly associated with the clinical categorisation of normal and increased LVM. Phenylacetylglycine is associated with phospholipidosis which is a disorder that can be induced in the kidneys by various drugs. Multiple terms relevant to left ventricular mass were highlighted by these results, including diet, blood pressure, diabetes, atherosclerosis, anaemia, cardiac remodelling and heart failure.

Interestingly some well-established associations with markers of LVH and related cardiovascular disease, such as angiotensin II, only become apparent in the presence of the confounders or the full clinical model. This demonstrates the utility of analysing putative predictors of different types in combination when analysing complex diseases/processes. While we cannot draw firm conclusions our study provides interesting insights into the molecular genesis of LVM. Without intending to over-interpret the data we were still pleased to find that many of the signals that derived from this approach are in keeping with and in part extending current biological knowledge.

While the PCA-varimax solution is substantially less time-intensive and is appealing in its inherent description of relationships between input variables and original components, it is far less comprehensive than the screening-modelling approach, and in a sense it is biased towards the largest datasets. The screening-modelling approach was found to be superior, identifying markers belonging to a range of molecule classes and relating to a wide range of

biological processes known to be involved in LVM or heart failure, and providing more statistical output to analyse. It also performed better in terms the amount of variance described in the combined (clinical and molecular) multiple regression models produced. PCA-varimax may however be more effective in an animal model where there is less noise due to heterogeneity.

5. Analysis of Multi-Omics Data in Conjunction with a Large Clinical Respiratory Dataset

5.1 Introduction

Asthma is a chronic respiratory disease which causes bouts of coughing, shortness of breath, difficulty breathing, chest tightness and wheezing. Symptoms are triggered by various environmental stimuli and can resolve spontaneously or with treatment. Severe asthmatics have frequent exacerbations and are more likely to have irreversible airflow obstruction associated with airway remodelling despite high doses of corticosteroids. They account for only 5-10% of asthmatics however require substantially greater healthcare resource use (307;308). Severe asthmatic patients' symptoms are exacerbated by different triggers and they express different profiles of biomarkers, both suggesting multiple subtypes of disease or complex heterogeneity.

Many current categorisations of severe asthma relate to a phenotype such as a trigger or an enriched cell type (e.g. eosinophilic). Another important phenotype to consider regarding the treatment of asthma is smoking, which increases asthma severity and reduces steroid efficacy – and the same is true for COPD. Developing better understanding of the molecular processes underpinning these difficult-to-treat cases could lead to improved testing and treatment.

Several recent attempts to develop asthma drugs beyond beta agonists and corticosteroids have failed to deliver consistent effective results through clinical trials (309;310). However more recently a move towards personalised medicine has shown several promising results. Monoclonal antibodies (mAb) have been developed against specific targets and more are in development (311). These can be applied in a phenotype-specific manner (e.g. based on cell counts) or based on biomarker measurements. Type 2 inflammation is characterised in part by high levels of IL-5, which is heavily involved in the proliferation of eosinophils. mAbs for IL-5, Mepolizumab and Reslizumab, are given as supplementary treatments to patients with severe eosinophilic asthma, and they

have been demonstrated to improve lung function and quality-of-life (312;313). In allergic asthma IgE is central to allergic response – when allergens bind IgE it triggers the release of histamine and other inflammatory mediators. The mAb Omalizumab can be used for severe asthmatics with high IgE and a positive skin prick test and reduces symptoms and inhaled steroid use (314).

Asthma and COPD are both common obstructive pulmonary diseases and share many of the same features, such as inflammation and airflow limitation (315). COPD is usually the result of sustained exposure to cigarette smoke or environmental pollution and it causes over 3 million deaths per year. It typically presents with many of the same symptoms as asthma – such as wheezing, breathlessness and coughing – however they are less reversible. In addition to reversibility COPD patients have other differences to asthma patients, including age-of-onset, counts of different cell types, and abundances of various mediators of inflammation. Generally sputum neutrophilia is detected in COPD and sputum eosinophilia is detected in asthma, however severe asthmatics exhibit more COPD-like inflammation - they are more neutrophilic and share some mediators of inflammation with COPD such as IL-8 and TNF-α (316).

The diagnosis of both asthma and COPD is often described as a distinct syndrome: Asthma-COPD Overlap Syndrome (ACOS). Due to similarities in the two diseases this syndrome may be difficult to diagnose if it is brought into practice (317). ACOS patients suffer more frequent and severe symptoms (318;319), experience a lower health-related quality of life (320), are hospitalised more frequently, and present a greater healthcare burden (321). The description of these diseases co-occurring as a separate syndrome unto itself has drawn some criticism but the similarities and differences between the diseases are of great interest regardless of categorisation (322).

Gene expression microarrays have been used to study asthma and COPD in different states and in a range of sample types including induced sputum and various types of airway samples (323;324). This kind of work has allowed the identification of biomarkers and the beginnings of a better molecular understanding of disease, e.g. expression profiles which can predict steroid response and discriminate between subtypes of asthma (325-327). In a multi-

tissue gene expression microarray study by Singhania et al (328), 19 healthy controls were compared to 46 asthmatic patients of varied severity, across epithelial brushings and flow-sorted $CD3^+$ T cells from sputum and BAL. Amongst severe asthmatics there was a neutrophilic phenotype and upregulation of genes relating to neutrophilia, mucin, and oxidative stress response. The majority of the disease signature was present in the 267 genes significantly dysregulated in the T cells of the severe asthmatics. The IL-13-inducible chemokines (POSTN, SERPINB2, and CLCA1) were found to be upregulated in the epithelium of mild and moderate asthmatics, but not severe asthmatics. Alternatively IL-17-inducible chemokines (CXCL1, CXCL2, CXCL3, IL8, and CSF3) were only found significantly upregulated in the sputum of severe asthmatics.

miRNA microarray studies and candidate gene studies have identified miRNAs which are dysregulated in asthma and COPD in various sample types, including several which are dysregulated in both conditions (329). In induced sputum miR-145-5p and miR-338-3p are both expressed with a higher abundance in both asthmatic patients and those with COPD (330). In an animal model of asthma, antagonism of miR-145-5p results in inhibited eosinophilic inflammation, mucus hypersecretion, Th2 cytokine production, and airway hyperresponsiveness (AHR) – with effects comparable to steroid treatment (331). Interestingly miR-145-5p, and several other asthma associated miRs also have anti-inflammatory potential, depending on the biological context (332).

There are relatively few proteomics and metabolomics studies compared to genomics and transcriptomics ones (333). Proteomics studies in asthma, COPD and smoking have identified several dysregulated molecules in common across phenotypes and studies (334). Gharib et al performed LC-MS/MS on five healthy subjects and ten asthmatic subjects and analysis identified ten proteins which were upregulated and seven proteins which were downregulated (335). Upregulation of serpin peptidase inhibitor (SERPINA1) and downregulation of SMR3B and SCGB1A1 were confirmed by western blot. Gharib et al suggest that SERPINA1 causes inflammation and airway remodelling by degradation of elastin in the extra-cellular matrix. SCGB1A1 inhibits phospholipase A2 which is involved in inflammation in asthma and a polymorphism in this gene has been shown to confer a risk for asthma susceptibility (336). In this same study exercise was

shown to induce increased levels of hemopoxin (HPX) and activated complement component 3 (C3a), where only C3a (which is involved in allergen response) was validated. Another proteomics study was performed by Lee et al to determine dysregulated proteins involved in severe neutrophilic asthma and identified S100A9 as a putative marker (337). Interestingly C3 (which hydrolyses to form C3a and C3b) was found downregulated in this context, but it is not clear whether this is due to increased C3 hydrolysis or decreased C3 production.

Several metabolomics studies in asthma and COPD have shown dysregulation within lipid metabolism, fatty acid metabolism, and mitochondrial beta-oxidation (338). COPD metabolomics studies have also shown dysregulation in the glycolysis and TCA cycle. Studies have taken place using several relevant or readily available tissues and a large number of putative biomarkers identified (339), however there are no studies published to date on the metabolome of induced sputum in asthma or COPD.

In this study microarrays for gene expression and miRNA expression were used to analyse the similarities and differences between severe asthma and COPD, and the effect of smoking on airway disease. While not featuring ACOS patients this data will highlight the molecular similarities that exist between the constituent diseases. The samples used were from 100 people (Table 5-1) representing all six combinations of respiratory health status (severe asthma/COPD/healthy) and smoking status (smoker/non-smoker). These patients are a subset of the CAB cohort which features a large volume of corresponding clinical data. Metabolomics and proteomics techniques were also used for healthy patients and asthmatic patients to look more closely at severe asthma and smoking in a wider molecular context. Not all of the patients with microarray data available were also used for proteomics and/or metabolomics, however a core set of 34 healthy and asthmatic patients have all four of these omics datasets available (Table 5.2). Some groups in this table have particularly small numbers, especially for omics-level analysis, however many published omics studies are similarly underpowered but still valuable (proteomics study above by Gharib et al used only five healthy controls), and combining groups for analysis provides a more reasonable sample size – e.g. asthmatics of both smoking status (n=25) compared with non-asthmatics of both smoking status (n=9).

Group (n)	M:F	GOLD (II:III)	Age	BMI	ICS	Steroid Dose
S_COPD (15)	8:7	10:5	63	24.8	9	1800
NS_Healthy (18)	5:13	-	50	26.2	0	-
S_Asthmatic (17)	8:9	-	51	26	17	1557
S_Healthy (16)	7:9	-	54	26.2	0	-
NS_Asthmatic (17)	11;6	-	49	32.5	17	1892
NS_COPD (17)	8:9	8:9	68	25.7	17	1410

Table 5-1: The 100 Patients of the CAB Dataset with Omics Data Available
Microarray data (mRNA, miRNA) corresponding to these patients exists for both induced sputum and nasal epithelium samples whereas mass spectrometry data (proteomics, metabolomics) is only available for asthmatics and healthy participants. M: male, F: female, S: smoker, NS: non-smoker, GOLD: Global Initiative for Chronic Obstructive Lung Disease, GOLD II: moderate COPD, GOLD III: severe COPD, BMI: body mass index, ICS: inhaled corticosteroid (number of patients prescribed this class of drug).

Group (n)	M:F	Age	BMI	ICS	Steroid Dose
NS_Healthy (4)	2:2	46	25.3	-	-
S_Asthmatic (11)	5:6	51	26.3	11	1420
S_Healthy (5)	2:3	59	25.4	0	-
NS_Asthmatic (14)	8:6	48	32.7	14	1982

Table 5-2: The 34 Patients of the CAB Dataset with Four Types of Omics Data Available
Metabolomics, Proteomics, and Transcriptomics – mRNA and miRNA – data sets are all available for induced sputum samples in the same set of patients. M: male, F: female, S: smoker, NS: non-smoker, BMI: body mass index, ICS: inhaled corticosteroid (number of patients prescribed this class of drug).

In both asthma and COPD the cells of the bronchial epithelium and the proximate leukocytes react to those external stimuli which may cause or exacerbate symptoms or features of disease. Nasal epithelium was used as a surrogate for bronchial epithelium, as has been done previously for studies of asthma and bronchial inflammation (340;341), because the samples are easier to obtain than bronchial samples - both procedurally and in terms of gaining consent. It forms a continuous tract with the bronchial epithelium, shares several mediators of inflammation with a strong correlation (342), and exhibits a similar response to viral infection (343). They also share similar gene expression changes in response to smoking (344). Genome-wide DNA methylation experiments show that nasal epithelium has a more similar epigenetic profile than other putative surrogates for bronchial epithelium such as blood or buccal cells (345). While many features of the nasal epithelium are similar there are some important differences such as lower proliferation rates, lower percentages of goblet cells and ciliated cells, and altered responses to IFN and IL-13 (346;347). In addition there is evidence suggesting that nasal epithelium is

unsuitable for use as a surrogate for bronchial epithelium in patients with COPD (348).

Induced sputum was used to study leukocytes, and counts were made for neutrophils, eosinophils, macrophages and lymphocytes to calculate differentials. These cell counts can be used as features/classifiers of disease (e.g. neutrophilic asthma) and taken into account when making clinical decisions (349), although cut-offs are quite varied - sputum eosinophilia has been defined as 1-3% of sample content and neutrophilia as 40-80% depending on the study (350-352). Using 3% and 60% for cut-offs of eosinophilia and neutrophilia respectively, non-smoking asthmatics had the highest proportion of eosinophilic patients (10/17), and non-smoking COPD patients had the most neutrophilic patients (10/17) (Table 5-3).

Target (Total)	Eos >3%	Neut >60%
S_COPD (15)	5	6
NS_Healthy (18)	1	6
S_Asthmatic (17)	7	7
S_Healthy (16)	1	7
NS_Asthmatic (17)	10	4
NS_COPD (17)	6	10

Table 5-3: Cell Type Percentages in Induced Sputum of the 100 Omics CAB Patients – Eosinophilia and Neutrophilia
Metabolomics, Proteomics, and Transcriptomics – mRNA and miRNA – data sets are all available for induced sputum samples in the same set of patients. S: smoker, NS: non-smoker, Eos: eosinophil count expressed as a percentage of total cells counted, Neut: neutrophil count expressed as a percentage of total cells counted.

Not only is macrophage count important but also ‘polarisation’, in which macrophages adopt different patterns of expression and different dominant functions, e.g. the pro-inflammatory M1 macrophages are involved in microbicidal activity, whereas the anti-inflammatory M2 macrophages are involved in wound healing (353). M2 macrophages are further divided into four subtypes with different gene expression profiles and functions (354). In asthma T-lymphocytes release Th2 cytokines such as IL-13, inducing airway hyperresponsiveness (355). T-lymphocytes express different cell surface markers in the context of smoking and in the context of COPD (356).

When viewed over long time-scales, age relates to many biological and clinical variables in a non-linear way, often tracking with disease events or physiological changes such as childhood development and functional decline. Age and bodyweight affect lung function even in healthy lungs. Asthma is more severe in older adults and age itself seems to contribute more towards risk of severe asthma than does the duration since diagnosis (357). Age has also been implicated in a steroid-resistant inflammatory effect, which could easily be mistaken for an asthma effect or a COPD effect (358).

Obesity reduces apoptotic cell clearance in airways (359), causes systemic inflammation and results in reduced lung function by several metrics including a reduced FEV1 and FVC. Even patients being overweight rather than obese was enough to detect a significant difference in lung function in several studies (360). Obesity in the context of an asthmatic patient makes their asthma more severe and resistant to treatment (361). On the other hand obesity in the context of COPD appears to have both deleterious and protective effects (362;363).

As age and obesity were not well controlled between groups they were modelled as confounders where possible to reduce their effect on the main statistical contrasts of interest. Despite these limitations it is a valuable multi-omic, multi-tissue, and multi-disease dataset which allows a lot of different biological questions to be asked.

5.2 Results

5.2.1 Induced Sputum

After pre-processing and screening the mRNA microarray data, 9957 genes remained for significance testing. The model including BMI had the lowest median p value, highest adjusted R^2, lowest median AIC, and almost as many genes were best-modelled with BMI as without. Since BMI is distinctly different between groups, and modelling statistics demonstrate the impact on the data across a large number of genes (Table 5-4) it was included in the model. This model was used with the R package LIMMA to determine which genes were differentially expressed in relation to asthma (364;365).

Model	Median p	Median Adj-R^2	Median AIC	# Genes Lowest AIC
~group	6.78×10^{-2}	6.05×10^{-2}	-162.73	4127
~group + bmi	5.62×10^{-2}	7.19×10^{-2}	-162.76	3505
~group + age	6.87×10^{-2}	6.58×10^{-2}	-162.08	2325

Table 5-4: Comparison of Statistical Models in Relation to mRNA Data
This shows that the resulting statistics from modelling with BMI included are on average more significant and explain more of the data as highlighted above. They also provide on average the best fit across all genes tested and a large proportion of genes are better-modelled with BMI included. Adj-R^2: adjusted R^2, AIC: Akaike Information Criterion

Filtering by variance (nsFilter) reduces the non-coding RNA (ncRNA; mostly mature miRNAs, miRs) list to 935. The same modelling procedure was followed as for the mRNA microarrays, showing that the model with age had the lowest median p value and highest median adjusted R^2, whereas the model with BMI has a lower median AIC (Table 5-5). Since there is the same number of samples but less than ten percent of the tests (to multi-test adjust for) including both is not statistically prohibitive.

Model	Median p	Median Adj-R^2	Median AIC	# Genes Lowest AIC
~group	1.00×10^{-1}	4.90×10^{-2}	-191.154	487
~group + bmi	1.07×10^{-1}	5.17×10^{-2}	-190.235	170
~group + age	8.25×10^{-2}	6.01×10^{-2}	-189.74	278

Table 5-5: Comparison of statistical models in relation to miRNA data.
While the best fit is seen with the simplest model a large number of the genes are better described by the inclusion of age or BMI. However, the inclusion of these terms while looking at miRNA data is less statistically costly given the number of variables being analysed. Adj-R^2: adjusted R^2, AIC: Akaike Information Criterion. Adj-R^2: adjusted R^2, AIC: Akaike Information Criterion.

Despite 'spending statistical power' by adjusting for age and BMI, there is still a large number of significant molecules (especially large number of miRNAs) the interaction network for the second becomes too complex:

	mRNA	ncRNA	Peptides	Metabolites
Asthma	49	42	158	0
Asthma in Non-Smokers	78	163	69	0
Asthma in Smokers	0	5	48	0
Smoking	1578	17	1	5

Table 5-6: Counts of Statistically Significant Variables Across Four Different Sets of 'Omics' Data.
For the purposes of interaction networks the number of hits highlighted likely covers too many molecules to provide useful results.

So for these purposes we restricted this list to those with two-fold change (as well as being multi-test adjusted $p < 0.05$):

	mRNA	ncRNA	Peptides	Metabolites
Asthma	12	4	105	0
Asthma in Non-Smokers	36	36	69	0
Asthma in Smokers	0	0	48	0
Smoking	80	2	1	5

Table 5-7: Counts of Statistically Significant Variables Across Four Different Sets of 'Omics' Data After Restriction by Two-Fold Change in Abundance.
Restriction by both fold change and multi-test adjusted p value which provides numbers of molecules which should be useful for generating meaningful interaction networks

5.2.1.1 Asthma

49 mRNA probe sets were found to be significantly associated with asthma.12 of these had at least a two-fold change, and they were all upregulated in asthma (F13A1, CCL17, ALOX15, MMP10, CPA3, CCL26, CST1, CD1A, CD1B, BPIFB1, IL18R1, TSPAN8; Table 5-8). 46 of the 49 significant genes were correlated with eosinophils (mean rho = 0.49), 4 with neutrophils (mean rho = 0.38), 1 with macrophages (rho = 0.53), 5 with lymphocytes (mean rho = 0.29) and 10 with epithelial cells (mean rho = 0.38).

ENSG	Log_2(FC)	Ave Exp	p	p_{BH}	Gene Symbol
ENSG00000124491	1.29	8.78	1.32×10^{-8}	1.32×10^{-4}	F13A1
ENSG00000102970	1.17	6.98	9.55×10^{-8}	4.75×10^{-4}	CCL17
ENSG00000166670	1.43	7.91	4.39×10^{-6}	5.20×10^{-3}	MMP10
ENSG00000161905	1.60	6.82	5.35×10^{-6}	5.20×10^{-3}	ALOX15
ENSG00000163751	1.41	6.75	1.12×10^{-5}	6.81×10^{-3}	CPA3
ENSG00000006606	1.23	6.20	1.4×10^{-5}	7.72×10^{-3}	CCL26
ENSG00000170373	1.38	5.72	2.27×10^{-5}	1.13×10^{-2}	CST1
ENSG00000158477	1.21	8.20	2.75×10^{-5}	1.19×10^{-2}	CD1A
ENSG00000158485	1.09	7.66	4.18×10^{-5}	1.49×10^{-2}	CD1B
ENSG00000125999	1.46	9.48	1.05×10^{-4}	2.83×10^{-2}	BPIFB1
ENSG00000115604	1.15	8.66	1.75×10^{-4}	4.16×10^{-2}	IL18R1
ENSG00000127324	1.31	6.26	1.82×10^{-4}	4.18×10^{-2}	TSPAN8

Table 5-8: Significant mRNAs in Asthmatics with a >2-fold Change.
These results are generated by asthmatics and non-asthmatics, regardless of smoking status, but with the effects of smoking statistically accounted for.

Term enrichment analysis identified two clusters (lipid antigen binding, and leukotriene metabolic process) and an individual term (positive regulation of production of molecular mediator of immune response) which were enriched. These all corresponded to genes upregulated in asthma and correlated with eosinophils (Fig 5-1). 80% of all genes in the genome associated with 'lipid antigen binding' were enriched.

158 peptides were found to be statistically significantly dysregulated in asthma, and restricted to those with a two-fold change 68 were downregulated and 37 upregulated. Correlations between two-fold altered peptides and cell types showed: 17 with eosinophils (mean rho = 0.37), 3 with macrophages (mean rho = 0.36), and 5 with epithelials (mean rho = 0.40). Proteins corresponding to this list of two-fold dysregulated peptides were split into 'upregulation' and 'downregulation' groups for term enrichment. 21 terms were significantly associated with downregulated genes including regulation of actin cytoskeleton, defence response to fungus, glucose metabolism, and negative regulation of apoptosis. Terms for staphylococcus aureus infection and complement and coagulation cascades were upregulated.

Just two of the significant proteins were also found to be significant at the transcript level (CST1, BPIFB1) and both were consistent in direction of fold change between the protein level and the transcript level.

38 mature miRNA sequences (miRs) and small nucleolar RNAs (snoRNAs) were found negatively associated with asthma - of which four had two-fold changes (hsa-miR-146a-5p, hsa-let-7f-5p, hsa-let-7g-5p, hsa-miR-146b-5p). Two correlated with eosinophils (mean rho = 0.28), two with neutrophils (mean rho = 0.43) and 21 with macrophages (mean rho = 0.42), and 1 with epithelials (mean rho = 0.26). 4 miRs were found positively associated with asthma with a two-fold change, two of which were correlated with macrophages. The miRNA with the largest fold change and the lowest p value by over a degree of magnitude is hsa-miR-146a-5p ($Log_2(FC) = -1.35$, $p_{BH} = 1.03 \times 10^{-4}$), which has previously been discussed as a putative M1 macrophage marker. It is similarly downregulated in all six subgroups (combinations of disease state and smoking status) except the healthy non-smoker (Fig 5-2).

Two other putative M1 macrophage markers followed the same trend, one of which was also statistically significant (miR-155-5p: $Log_2(FC) = -0.76$, $p_{BH} = 0.02$; miR-187-3p: $Log_2(FC) = -0.72$, p_{BH}: 0.09), whereas statistics for two M2 macrophage markers showed little difference (miR-511-5p: $Log_2(FC) = -0.07$, $p_{BH} = 0.68$; miR-193b-3p: $Log_2(FC) = -0.13$, p_{BH}: 0.75).

None of the 1019 identified metabolites were significant after multi-test correction, however 23 were downregulated two-fold and 67 were upregulated two-fold. The most significant changes in identified metabolite levels were a >13-fold increase in toluene-4-sulfonate (p: 2.77×10^{-3}) and a >7-fold increase in quinate (p: 7.12×10^{-3}). Three unidentified metabolites were significant after multi-test correction (76.02Da, 85.02Da, 85.02Da).

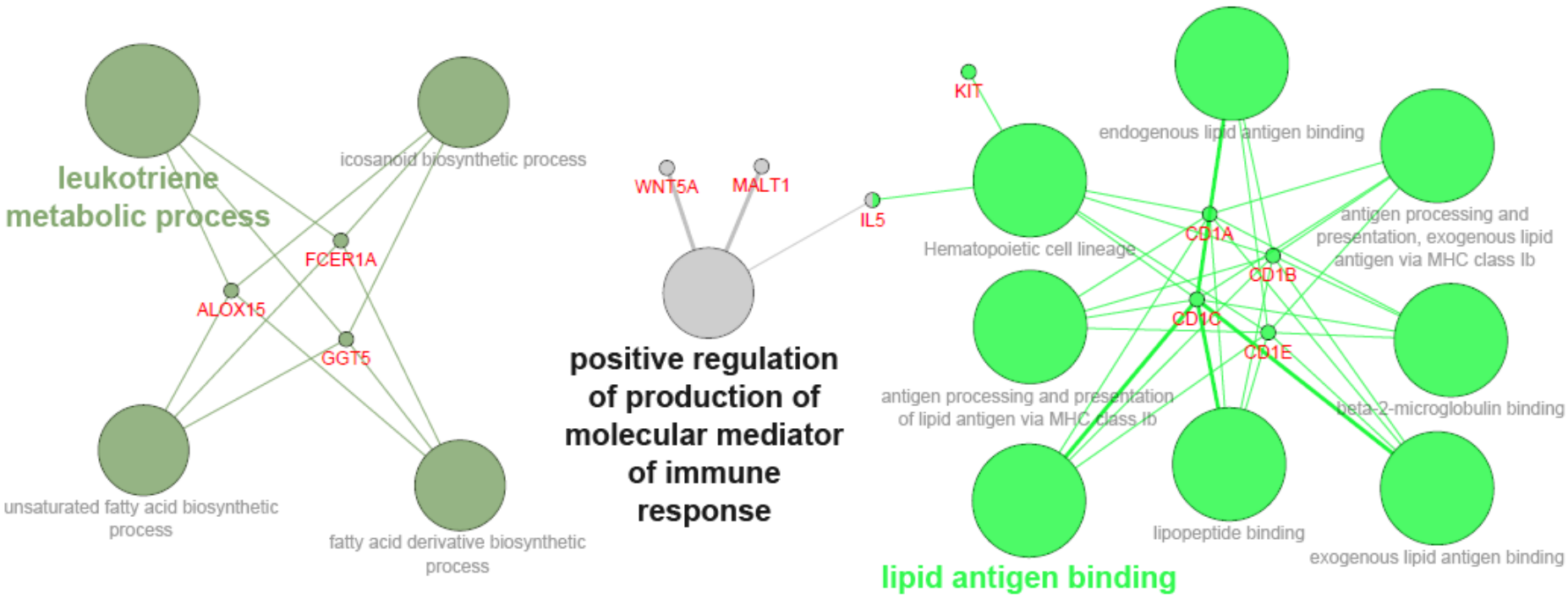

Fig. 5-1: Term Enrichment Analysis of Asthma using mRNA Microarray Data.
Two clusters of terms (dark green, light green) and one single individual term (grey) were identified as being enriched in asthma. Colours indicate clustering of terms by ClueGO and genes contributing to enrichment are shown. All genes shown are upregulated in asthma and correlated with eosinophil

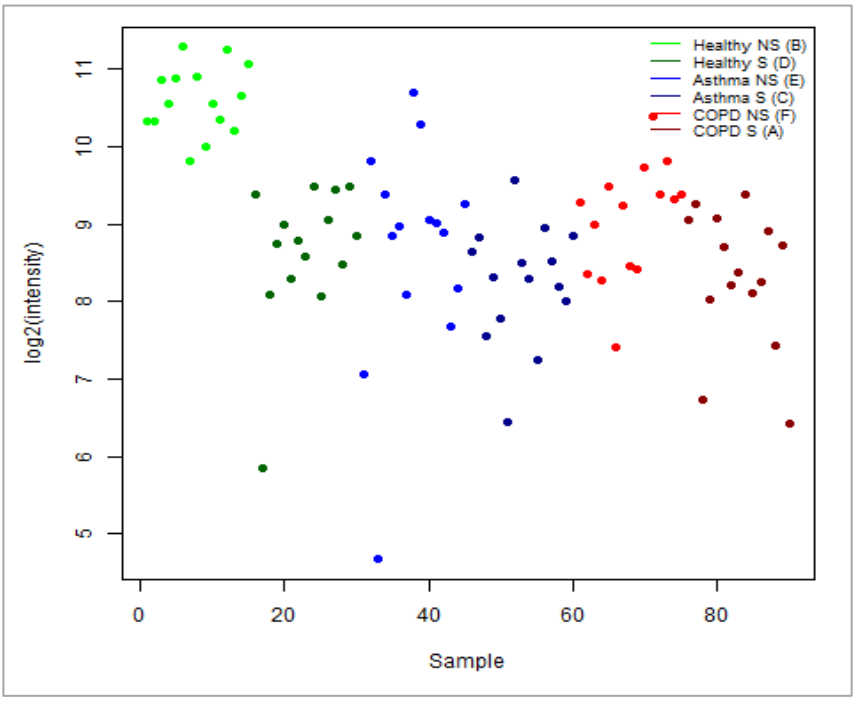

Fig. 5-2: miR-146a is Downregulated in Sputum of Smokers and Patients with Asthma or COPD
miR-146a is downregulated with asthma, COPD, and with smoking status in induced sputum. Legend shows colours for each group, with group code in brackets (S: smoker, NS: non-smoker). Light green: healthy smoker, dark green: healthy non-smoker, blue: asthma non-smoker, dark blue: asthma smoker, red: COPD non-smoker, dark red: COPD smoker.

Taking all statistically significant data points with at least a two-fold change and integrating them in a contextualised manner using interaction databases led to the construction of the interaction network shown in Figure 5-3. Relationships demonstrating upregulation or downregulation were removed where the fold changes were inconsistent with that action. In this network graph there are 3 small clusters and one large cluster containing fibronectin (FN1), 24 molecules it interacts with, and other 'connected' molecules.

Fibronectin is the node (molecule) with the highest degree, i.e. the molecule with the most interactions with other molecules. 16 of these 24 interacting molecules were downregulated. One of the eight upregulated interacting molecules is known to degrade it (matrix metalloproteinase 10; MMP10), possibly explaining its change in abundance. While sputum fibronectin has previously been found upregulated and eosinophil-correlated in asthmatic patients, here it is downregulated and macrophage-correlated.

The node with the next-highest degree (eight interactions) is beta-actin (ACTB) which is usually considered to be a ubiquitous well-regulated component of the

cytoskeleton but can also be secreted into the extracellular space and can be dysregulated in lung conditions - and like fibronectin it is also downregulated here.

The several of the next-most interconnected molecules of the network are connected with fibronectin, beta-actin, and with each other. Cofilin 1 (CFL1), ezrin (EZR), and plasminogen (PLG) all interact with fibronectin and beta-actin, and with two additional proteins each. Fibrinogen alpha chain (FGA) and pyruvate kinase M1/2 (PKM) interact with four molecules including fibronectin. Tubulin alpha 4A (TUBA4A) interacts with four molecules including PKM.

Lactoferrin (LTF) also interacts with four other molecules, including the upregulated lysozyme (LYZ) which acts in concert with it for its antibacterial activity. In addition to a reduction in LTF – which has anti-bacterial and anti-inflammatory properties – there is also a reduced amount of calmodulin to deactivate it. It also interacts with mucin 7 (MUC7) which itself interacts with four molecules in the network.

Both miRs shown (miR-146a-5p and miR-146b-5p) can downregulate the same protein – a type II Golgi transmembrane protein also found in the extracellular space. The reduction in cathepsin G (CTSG) may explain the increase in complement 3 (C3). CTSG is also known to deactivate AGT however there is increased AGT and an increased amount of another molecule known to activate it – carboxypeptidase (CPA3).

The same network was subsequently used to view correlations between molecules of interest and the cell types which may be responsible for producing them (Fig 5-4), indicating which cells of the mixed sample type are relevant in which processes and highlighting potential signalling between different cell types. The macrophage-associated reductions in fibronectin, miR-146a-5p, and miR-146b-5p, and lack of macrophage-associations in other molecules may suggest that these molecules are secreted by macrophages for interaction with molecules originating from other cell types - although 19 of the molecules which interact with fibronectin do not have predicted cell types by this method. The other five molecules connected to fibronectin are all eosinophil-correlated, two

of which are also lymphocyte-correlated (MMP10) and epithelial-correlated (KRT10). Secretion of fibronectin (which forms part of the extracellular matrix) and miRs (secreted in vesicles) are consistent with the literature.

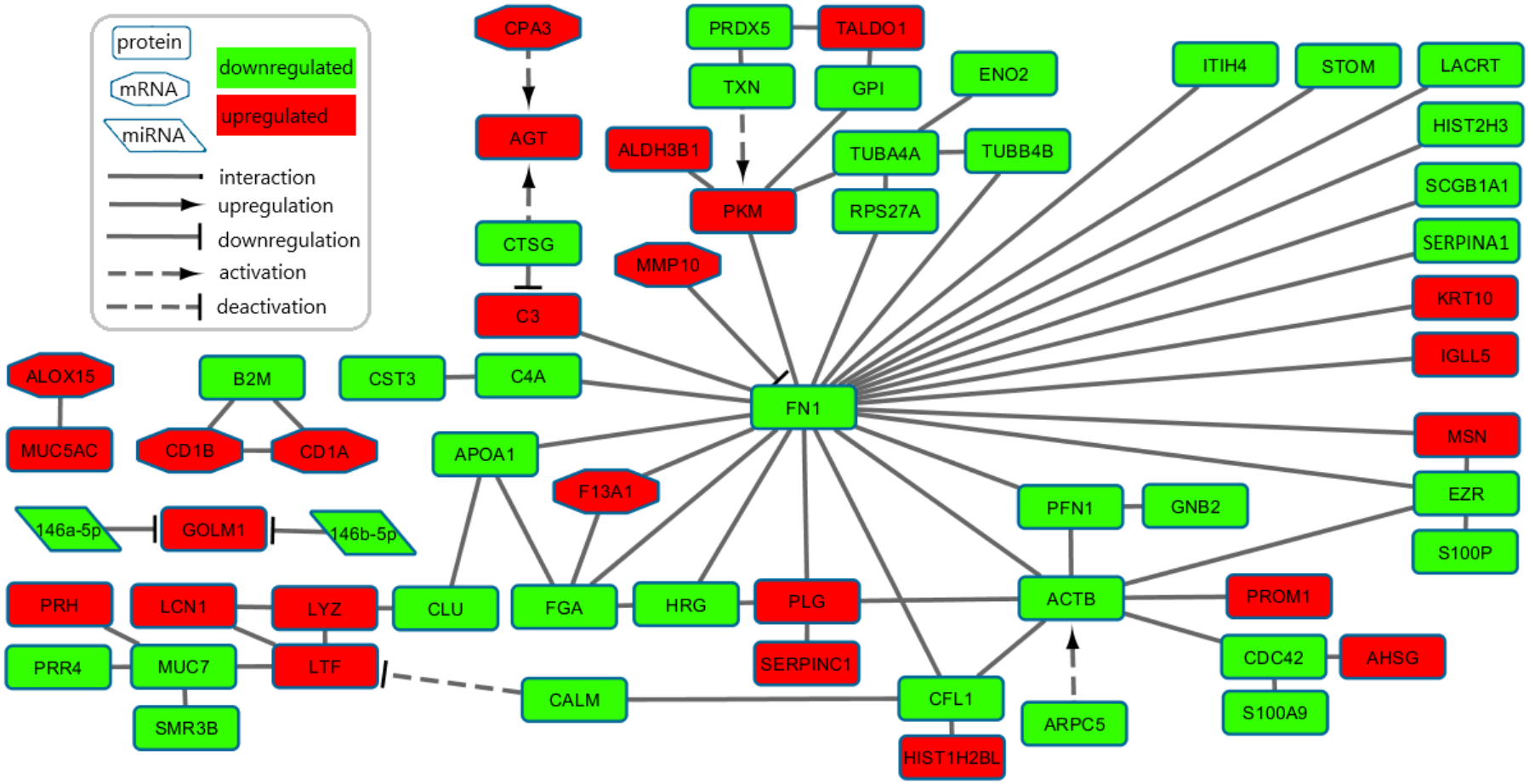

Fig. 5-3: Interaction Network Consisting of Molecules Significantly Differentially Regulated in Asthma.
This interaction network graph shows three small networks and one large network with fibronectin, and to a lesser extent beta-actin, as highly-connected nodes, indicating a significant dysregulation of the extracellular matrix in asthmatic sputum. Red: upregulation in asthma, green: downregulation, cyan: different transcripts/proteins of the same gene are regulated in different direction. Solid lines with arrow/flat heads indicate up/downregulation. Dashed lines denote activation/inactivation. Rectangle: protein, octagon: mRNA, parallelogram: miR.

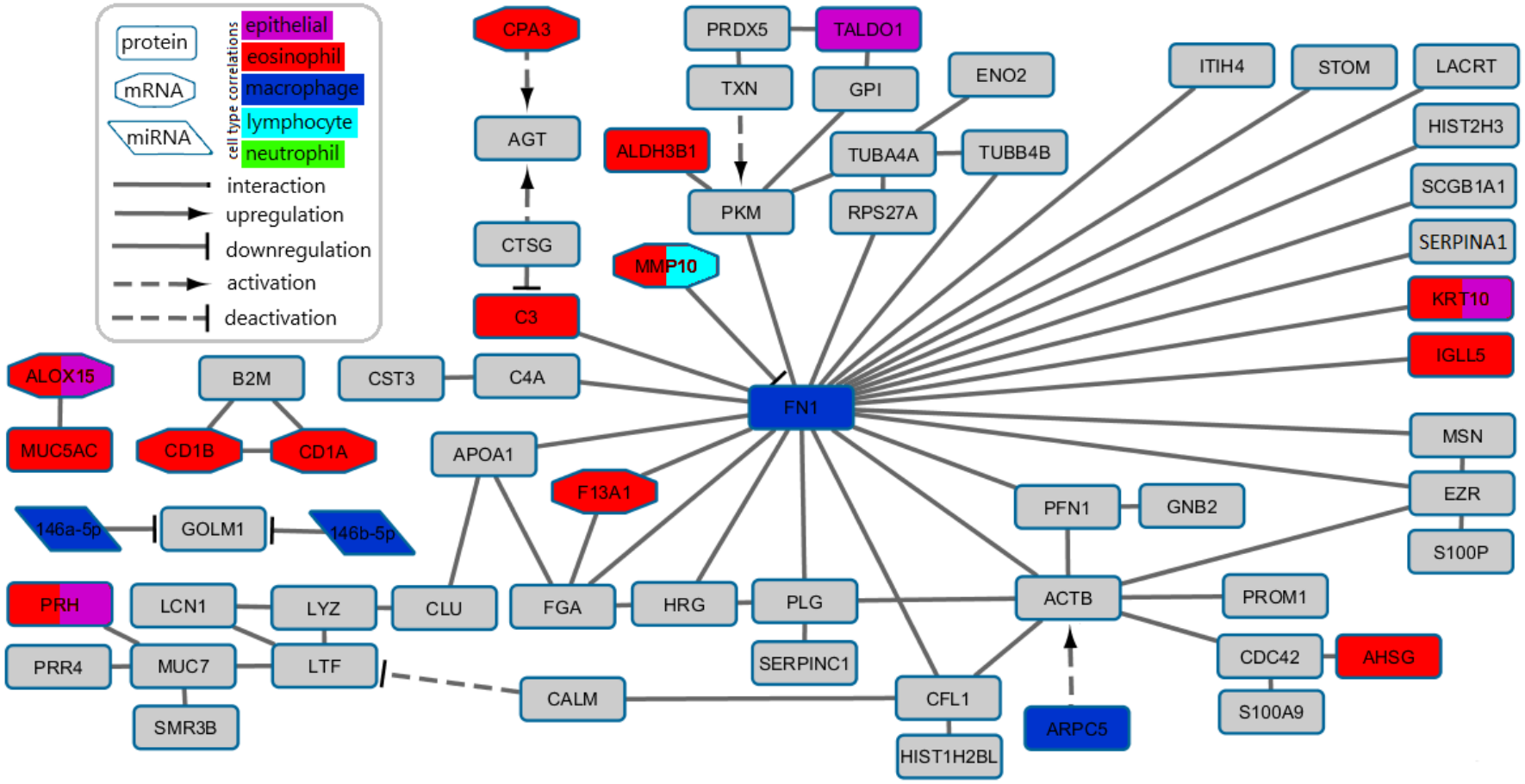

Fig. 5-4: Asthma Interaction Network Coloured by Cell Type Correlations.
Most significant molecules were not correlated with any particular cell type. Fibronectin which is at the centre of the network is macrophage-correlated, and interacts with molecules without correlation or eosinophil-correlated. Coloured by correlation with cell type, red = eosinophil, purple = epithelial, blue = macrophage, cyan = lymphocyte. Solid lines with arrow/flat heads indicate up/downregulation. Dashed lines denote activation/inactivation. Rectangle: protein, octagon: mRNA, parallelogram: miR.

5.2.1.2 Asthma in Non-Smokers

78 mRNA probe sets were found to be significantly associated with asthma – 8 downregulated and 70 upregulated. 36 of these were differentially regulated at least two-fold. 70 genes were correlated with eosinophils (mean rho = 0.63), 3 with neutrophils (mean rho = 0.46), 5 with macrophages (rho = 0.58), 6 with lymphocytes (mean rho = 0.41) and 17 with epithelials (mean rho = 0.47). ClueGO identifies four clusters of terms (unsaturated fatty acid biosynthetic process, coated vesicle membrane, chemokine receptor binding, and lipid antigen binding) and two individual terms (regulation of fibroblast growth factor receptor, and activation of matrix metalloproteinases) upregulated in asthmatic non-smokers in contrast to healthy non-smokers (Fig 5-5). All but three of the genes are eosinophil-correlated: *CD207, HLA-DQB2 and SLC16A2* – which are all related to the 'coated vesicle membrane' term.

69 peptides were found to be significantly associated with asthma in non-smokers, all but one with two-fold difference in expression – 47 downregulated and 22 upregulated. 19 were correlated with eosinophils (mean rho = 0.51), 1 with macrophages (rho = 0.68), and 1 with epithelials (rho = 0.56). Terms relating to salivary gland secretion, G13 signalling and glucose metabolism were mostly downregulated and complement and coagulation cascades were enriched (Fig. 5-6). Proteins were split by correlations with cell type for further analysis which revealed a cluster of macrophage-correlated terms (prostaglandin synthesis and regulation, phospholipase inhibitor activity, lipase inhibitor activity) amongst individual epithelial-correlated terms (myeloid cell development, amyloids, salivary secretion, cysteine-type endopeptidase inhibitor activity) and a cluster of 28 epithelial-correlated terms whose main term is RHO GTPases activate PKNs. *MUC5AC* was significantly associated both at the protein and mRNA level, both enriched in asthmatics.

163 ncRNAs were dysregulated with asthma in non-smokers, all but 21 of which were downregulated. 36 of these downregulated ncRNAs were downregulated more than two-fold, all of which were miRs. These miRs were mostly macrophage-associated (18 probes, mean rho = 0.46) and some were neutrophil-associated (2 probes, mean rho = 0.43).

None of the identified metabolites were significant after multi-test correction, however with at least a two-fold change: 25 were downregulated and 61 were upregulated.

Figure 5-7 shows an interaction network comprising significant results for asthmatic non-smokers, with colours indicating the direction of regulation in asthmatics. The largest network consists of 59 peptides, mRNAs and miRs, and the remaining two networks are 3 and 2 molecules in size and consist only of peptide results. In the large network there are many instances of putative regulation by downregulated miRs, perhaps explaining the increased abundance of target genes at the mRNA or protein level. Most miRs appear to be expressed by macrophages and neutrophils and target exclusively eosinophil-correlated genes (Fig. 5-8). Four miRs are shown to target four different genes (151a-5p, 34a-3p, 181c-5p, 148b-3p), and four genes are targeted by four different miRs (*PRH, KIT, MAOA, CTTNBP2*).

In comparison with the previous network most notably fibronectin is missing, though several elements are found in common. Again PKM is a highly connected molecule (nine interactions), and seven of the eight peptides in the upper-left of the graph, centred around LTF are maintained from the previous network. With this contrast however three of these molecules have a different direction of fold change – LCN1 and LYZ are downregulated instead of upregulated, and there are peptides of LTF regulated in both directions.

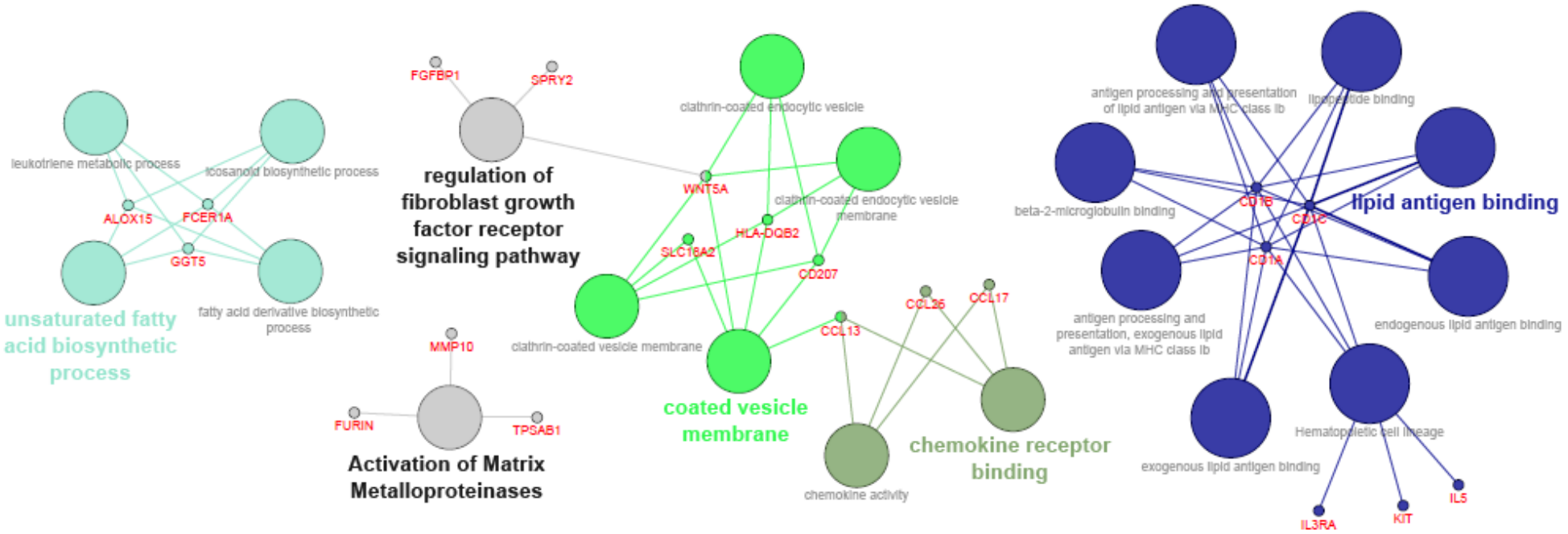

Fig. 5-5 Term Enrichment Analysis of Asthma in Non-Smokers Using mRNA Microarray Data.
Four clusters of terms (shown by light blue, green, dark green and dark blue left-to-right above) and two individual terms (grey) were identified as being enriched in asthma. Colours indicate clustering of terms by ClueGO and genes contributing to enrichment are shown. All genes shown are upregulated in asthma and correlated with eosinophils 2/3 of the genes described by the term “regulation of fibroblast growth factor receptor signalling pathway” are also correlated with epithelials.

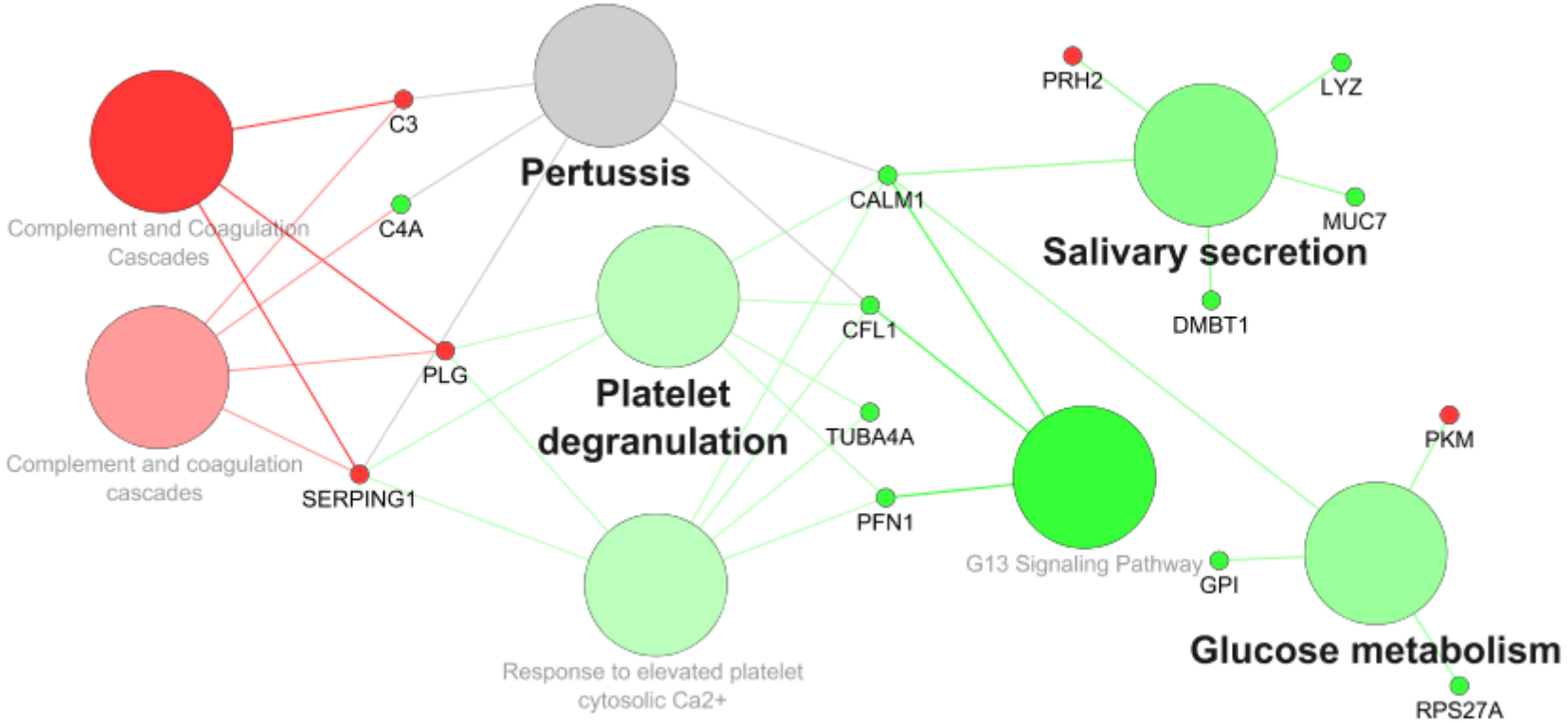

Fig. 5-6: A Term Enrichment Analysis of Asthma in Non-Smokers Using Proteomics Data and Performed with ClueGO.
Terms relating to salivary gland secretion, G13 signalling and glucose metabolism were mostly downregulated and complement and coagulation cascades were enriched. colours indicate direction of fold change of respective genes, where green signifies downregulation, red signifies upregulation with grey in the centre of the scale.

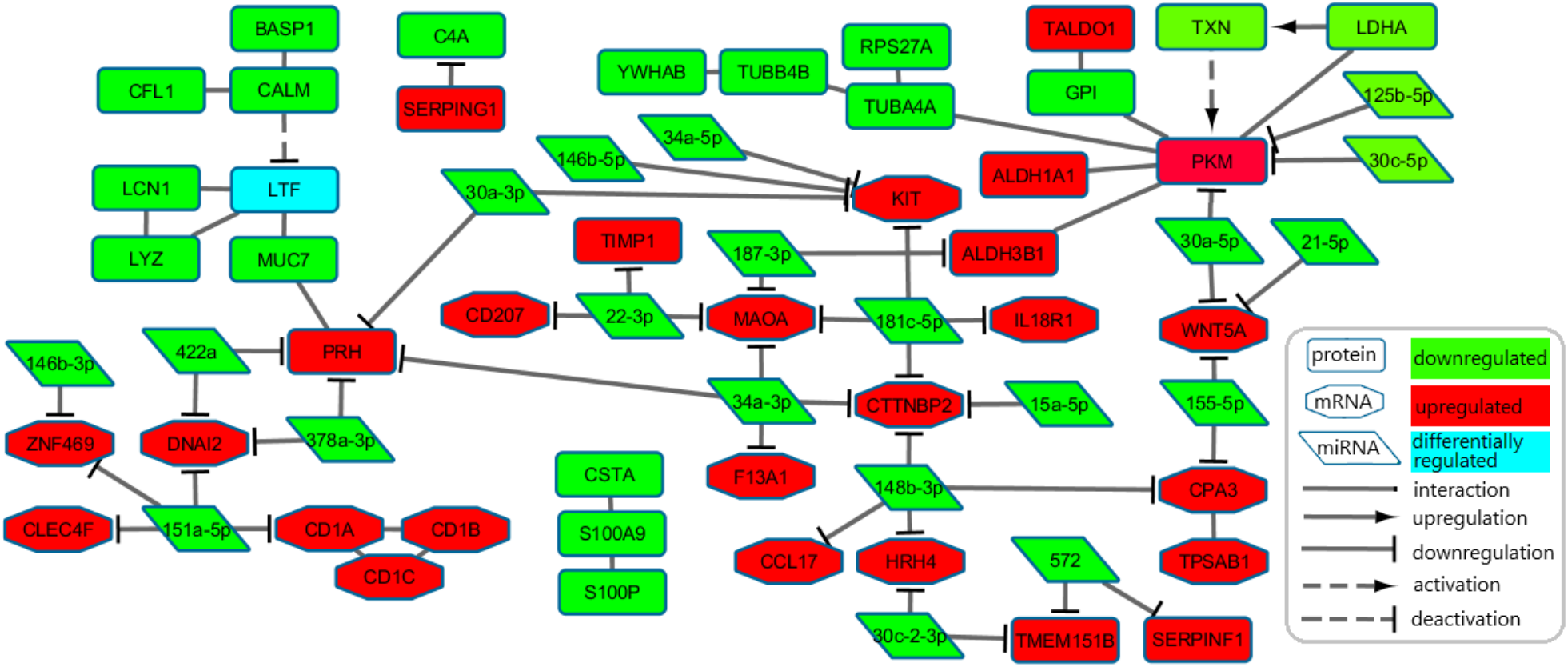

Fig. 5-7: Interaction Network Consisting of Molecules Significantly Differentially Regulated in Asthma Amongst Non-Smokers.
Red: upregulation in asthma, green: downregulation, cyan: different proteins of the same gene are regulated in different direction. Solid lines with arrow/flat heads indicate up/downregulation. Dashed lines denote activation/inactivation. Rectangle: protein, octagon: mRNA, parallelogram: miR.

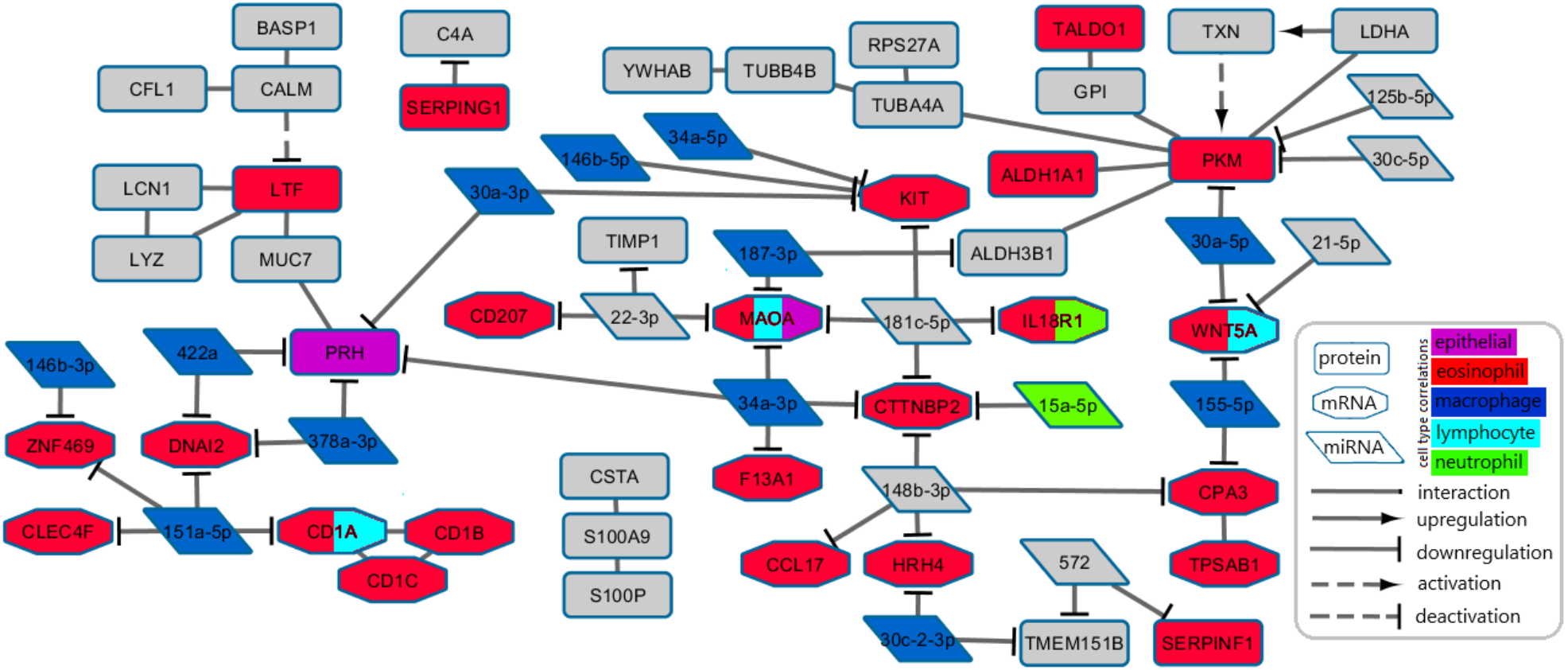

Fig. 5-8: Interaction Network Consisting of Molecules Significantly Differentially Regulated in Asthma Amongst Non-Smokers Coloured by Cell Type Correlation.

Significant mRNA and protein levels were mostly eosinophil and epithelial-correlated, while miRNAs were mostly neutrophil and macrophage-correlated. Coloured by correlation with cell type, red = eosinophil, purple = epithelial, blue = macrophage, cyan = lymphocyte. Solid lines with arrow/flat heads indicate up/downregulation. Dashed lines denote activation/inactivation. Rectangle: protein, octagon: mRNA, parallelogram: miR

5.2.1.3 Asthma in Smokers

No mRNA probe sets were found significantly associated with asthma in smokers, even when BMI was removed from the model. 93 mRNAs had a raw p value < 0.05, four of which had a two-fold change. This list was enriched for several gene ontologies (Fig. 5-9) related to genes which were all upregulated in asthmatics amongst smokers.

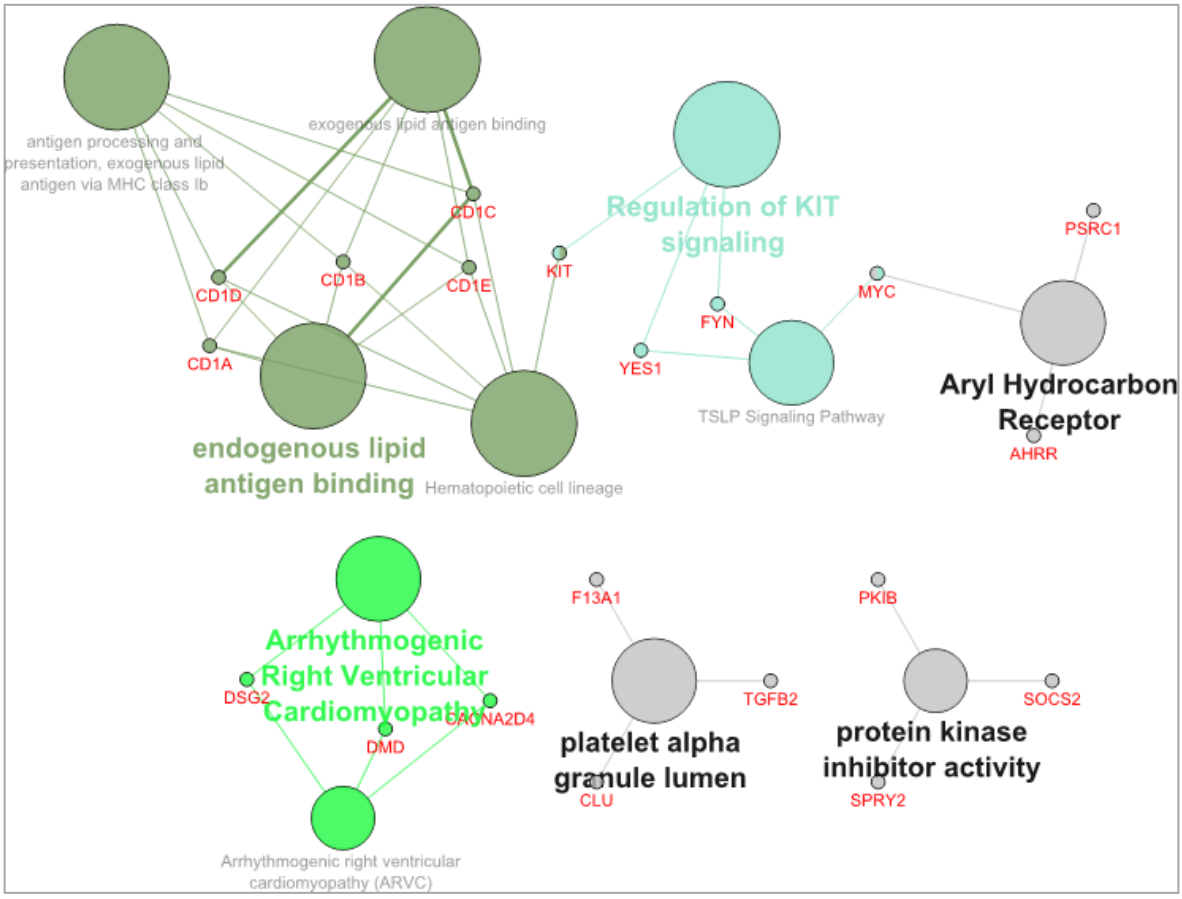

Fig. 5-9: Term Enrichment Analysis of Asthma in Smokers Using mRNA Microarray Data and Performed with ClueGO.
Three clusters of terms (shown by dark green, light blue, green) and three individual terms (grey) were identified as being enriched. This analysis was done on those genes with raw p values < 0.05 so it is not as robust as the other analyses which use multi-test adjusted gene lists.

50 peptides were found significant with asthma in smokers, all but two of which were differentially expressed two-fold – 35 downregulated and 15 upregulated. Term enrichment analysis showed similar patterns of term enrichment as in non-smoking asthmatics (Fig. 5-10). The genes associated with 'blood microparticle' and 'acute-phase response' were all eosinophil-correlated except for *SERPINA1* which was lymphocyte-correlated and *STOM* and *KV402_HUMAN* which were not correlated to a cell type.

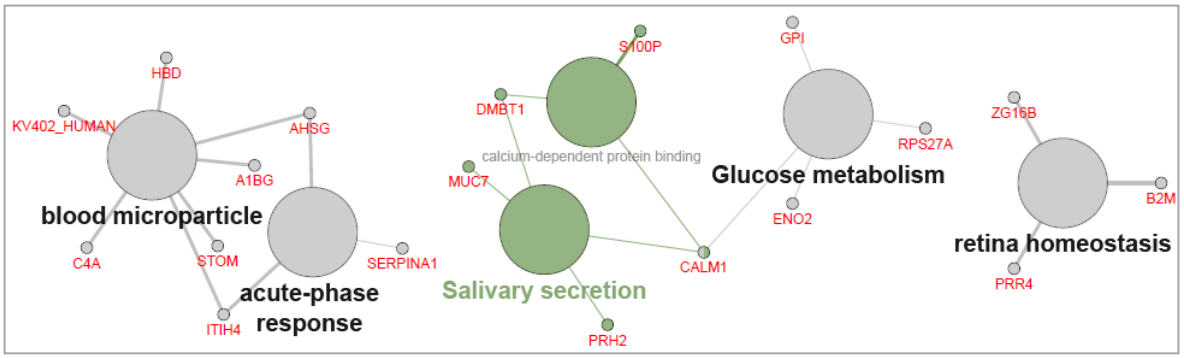

Fig. 5-10: Term Enrichment Analysis of Asthma in Smokers Using Proteomics Data.
Retina homeostasis, calcium-dependent protein binding and glucose metabolism are all consistently associated with downregulated genes.

3 miRs and 2 snoRNAs were found significantly upregulated with asthma in smokers. Three of these ncRNAs were eosinophil-associated (mean rho = 0.49), one was lymphocyte-associated (rho = 0.43). No metabolites were significant after multi-test correction although quinate had a large fold change and significant raw p value (>6-fold, p=0.017). 30 identified metabolites were decreased more than two-fold and 63 were increased more than two-fold. Two clusters are found in the interaction network which is smaller than the previous two and consists only of proteomics results (Fig. 5-11). Of the larger six-molecule network four of these molecules were also found downregulated in the asthma contrast and three of those were also downregulated in the asthma in non-smokers contrast. Similarly the MUC7-centered network was previously found as part of the asthma contrast and two of those molecules were found in the asthma in non-smokers contrast.

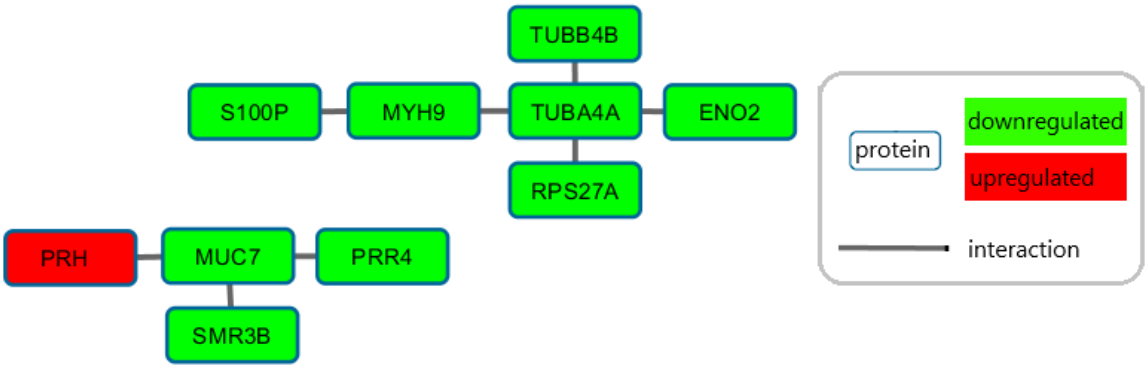

Fig. 5-11: Interaction network consisting of molecules significantly differentially regulated in asthma amongst smokers.
Both networks shown in this graph were found to an extent in previous networks based on similar asthma contrasts. Red = upregulation in asthma, green = downregulation.

5.2.1.4 Smoking

Of the 1578 genes that were found to be significantly associated with smoking only 80 had a greater than two-fold change. A term enrichment analysis of the full set of 1578 shows 217 significant terms clustered into 23 groups with 11 remaining individual terms. Two of these 217 terms were represented by over 50% of their associated genes ('fat-soluble vitamin biosynthetic process' and 'Transcriptional activation by NRF2', downregulated and upregulated respectively). Term analysis split by cell type indicated increased macrophage cell division.

Of the smaller more stringent set of 80: 45 were upregulated and 35 were downregulated. 6 were correlated with eosinophils (mean rho = 0.37), 5 with neutrophils (mean rho = 0.32), 15 with macrophages (mean rho = 0.38) and 15 with epithelials (mean rho = 0.35). Term enrichment analysis of this fold-change-restricted data shows upregulation of mast cell activation, oestrogen and glutathione metabolism, and the downregulation of chemokine receptor binding.

One protein, NAD(P)H dehydrogenase [quinone] 1 (NQO1) was found highly upregulated (over 9-fold) in smokers. The same upregulation and correlation is detected at the transcript level also. In smokers 11 miRNAs were downregulated, three of which correlated with macrophages (mean rho = 0.38). 5 miRs were upregulated, three of which were correlated with neutrophils (mean rho = 0.29) and one RNA-binding E3 ubiquitin ligase was upregulated and correlated with epithelials (mean rho = 0.51).

5.2.1.5 COPD

3078 mRNA signals were statistically significant, 67 of which had a two-fold change. Term enrichment using the full set of significant mRNAs shows 219 significant terms clustering into 20 clusters with 10 individual terms remaining. There is upregulation in terms relating to cell activation, cytokine production, and signalling by NF-κB, interleukins, and CD209/DC-SIGN. Downregulation occurs with terms relating to lipid/lipoprotein metabolism, mitochondria and microtubules. The upregulated terms are predominantly neutrophil-correlated and the downregulated terms are predominantly macrophage-correlated. Two terms were represented by over 50% of their genes: 'I-kappaB/NF-kappaB complex' and 'CD209 (DC-SIGN) signalling', which are both upregulated. A cluster of mitosis terms and an NRF2 term are upregulated and a cell activation cluster was downregulated. A similar cluster of mitosis terms is macrophage-correlated.

The 67 with at least two-fold change were also used for term enrichment analysis (Fig. 5-12). The matrix metalloproteinase cluster is consistently upregulated in COPD across all four associated genes, whereas the other two terms have genes both up- and down-regulated.

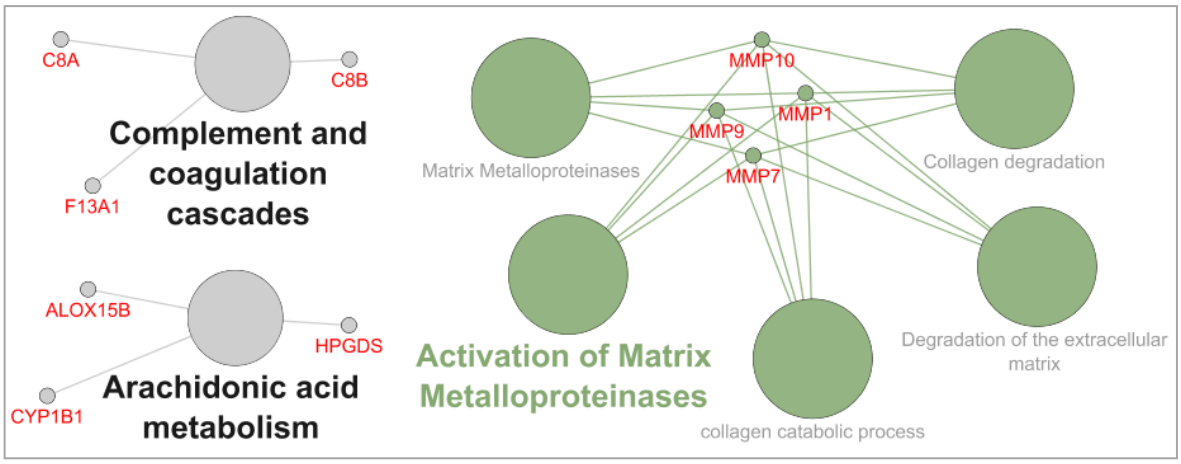

Fig. 5-12: Term Enrichment Analysis of COPD Using mRNA Microarray Data.
Complement and coagulation cascades, arachidonic acid metabolism, and activation of matrix metalloproteinases were the gene ontologies enriched with COPD mRNA results. Green = cluster of terms, grey = individual terms.

The vast number of molecules was used to generate an interaction graph which showed no discernible features due to extremely high interconnectivity. A version truncated by a two-fold cut-off shows the relationships between some of

the most dysregulated genes (Fig 5-13). Four of these molecules are matrix metalloproteinases, which can activate each other. These are well-known to be dysregulated in COPD, probably participating in proteolytic attack on the alveolar wall matrix. MMP10 in particular was also significantly dysregulated in asthma. The other three molecules are chemokines (CCL7, CCL24) and a chemokine receptor (CCR3). One miR (hsa-miR-663a) was statistically significant after multi-test correction (pBH: 0.02, log2FC: -0.60).

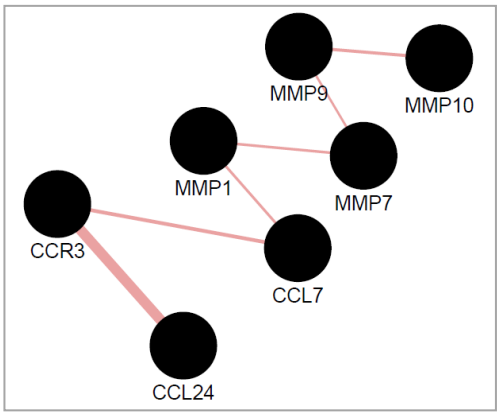

Fig. 5-13 GeneMania Analysis of COPD-Associated Genes
Relationships between those significant genes with a two-fold difference between COPD patients and healthy participants were identified with GeneMania. Molecules include matrix metalloproteinases, chemokines and a chemokine receptor.

5.2.1.6 Comparing Contrasts

More significant mRNA results were found with the smaller non-smoking comparison than the comparison also including smokers (78 in non-smokers, 49 in all asthmatics, Fig. 7-1). 19 of the 1568 smoking-associated genes overlap with asthma contrasts. Of the 80 smoking-associated genes with >2-fold change, four genes overlap with one of the asthma contrasts (*FCER1A, CD1A, CD1B, HLA-DQB2*). 55 of the 3078 COPD-associated genes overlap with asthma contrasts. Of the 67 COPD-associated genes with >2-fold change, three genes overlap with one of the asthma contrasts (*F13A1, MMP10, IL18R1*). 12 peptides were significant in asthmatic smokers but not in asthmatic non-smokers, versus 33 in non-smokers but not in smokers (Fig. 7-2).

One miR, hsa-miR-335-5p, was found to be significant for asthma amongst both non-smokers (Log_2FC: -0.54, p_{BH}: 0.01) and smokers (Log_2FC: 0.74, p_{BH}: 7.86 X 10^{-3}), and interestingly the change is in a different direction depending on smoking status (Fig 7-3). 3/17 smoking miRNAs were also significant in asthma contrasts (hsa-miR-146a-5p, hsa-miR-708-5p, hsa-miR-187-3p).

The induced sputum asthma and COPD hits were used for clustering analysis and heatmap generation (Fig 5-14), and displayed numerous small clades of consistent phenotypes and combinations of phenotypes (i.e. disease state and smoking status), which may indicate distinct subtypes of disease. Interestingly there is a group of four healthy non-smokers distinct from the other healthy subjects, who perhaps have a shared phenotype with a similar expression pattern to those with lung disease. The few asthmatic patients with similar expression patterns to healthy subjects may achieve a severe phenotype through a smaller set of influential molecular predictors, or by signals which have gone unmeasured. Generally the small clusters of patients with the same phenotype being distant from each other indicates a complex molecular signature or relatively distinct subtypes of disease – however it must be assumed that important information is missing, including metabolomics data which was not suitably statistically powered, peptides missed by the particular technique used, and splice variant level data.

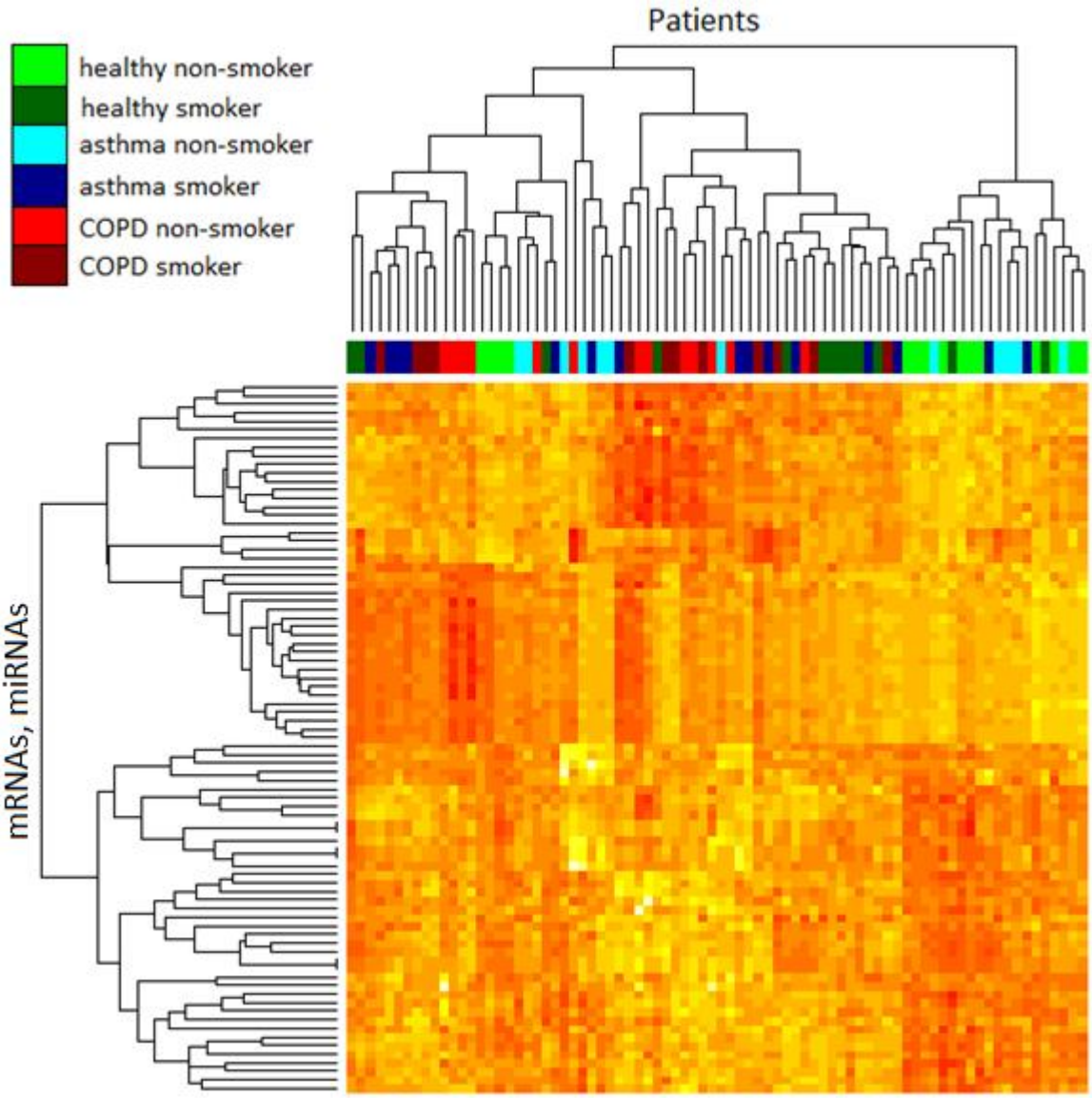

Fig. 5-14 Clustering and Heatmap of Induced Sputum Asthma and COPD Hits
Hits were used for clustering analysis and heatmap generation, which displayed numerous small clades of consistent phenotypes and combinations of phenotypes (i.e. disease state and smoking status). Clustering of patient groups indicates a complex molecular signature or relatively distinct subtypes of disease.

5.2.1.7 Merged Interaction Network

The interaction network for asthma in smokers is much smaller than the other two asthma contrasts, but this may be misleading. Molecules in this contrast may have only fallen slightly short of the specified thresholds, so to check the extent of the differences identified between the three asthma contrasts (i.e. 'Asthma', 'Asthma in Non-Smokers', 'Asthma in Smokers'), and to summarise the pertinent relationships the respective interaction networks were merged into one large network. All significant molecules across any of these contrasts were included and the data describing relationships between these molecules were gathered from the same sources as in the previous interaction networks.

A large number of molecules were shown to interact with fibronectin (FN1), so to reduce the complexity of the diagram those interactions were summarised by making the FN1 node a 'container' for those nodes with which it interacts. This step reduces complexity of the image by removing 26 edges from the network without any loss of information content. Fold change values for asthma, asthma in non-smokers, and asthma in smokers, were mapped onto this network (Fig 7-4 to 7-6).

While 'directionless' interactions between molecules of increased and decreased abundance can be informative (showing what interactions may not occur when one of the molecules is missing, and also which molecules are therefore more available for interactions with other molecules) these were also removed to form a more concise network, with fewer edges crossing over each other (Fig 5-15 to 5-17).

19 of the 30 instances of miRNA targeting mRNA transcripts in the network are confirmed by significant correlations (mean rho: -0.36). Even though these verified interactions exist in a complex network, where there are several occurrences of particular miRs targeting multiple genes, and several occurrences of particular genes targeted by those miRs. In particular KIT is targeted by four miRs in the network, three of which have significant correlations, and two of which are both significant alongside each other in a linear model (miR-30a-3p, miR-146b-5p). Although fewer samples were available with both proteomic and

miRNA microarray data were available 1 of 15 instances of miRNA downregulation was significant at the protein-level. One of the four interactions shown at the mRNA/protein level had a significant correlation – TXN upregulating LDHA (rho: 0.75, p: 1.85 X 10^{-07}).

The differences between the asthma contrast and the asthma in non-smokers contrast when viewed in this way appear minimal, and suggest that many of the differences are merely due to a reduced statistical power, putting the adjusted p-value below the threshold. As noted above (section 5.2.1.2) when looking at all asthmatics LCN1, LTF and LYZ are upregulated, however specifically amongst non-smokers LCN1 and LYZ are downregulated and LTF is differentially regulated (peptides of the same gene are regulated in different directions). Amongst smokers they are all upregulated, which is responsible for the difference when looking at all asthmatics. While few of the molecules in the interaction network were significant amongst non-smokers most of the fibronectin-associated and adjacent genes maintained strong fold changes in the same directions. The fold changes of most of the miRs and many of their targets were significantly reduced however, which may be because smoking itself induces the same fold changes in many of the molecules as seen with miR-146a-5p (Fig 5-2).

Cell type correlations from all three contrasts were combined and added to the network graph (Fig 5-18). Many genes at both the peptide and mRNA level are eosinophil-correlated, and all of these are upregulated in asthma. Some of these eosinophil-correlated molecules were also found correlated with neutrophils, lymphocytes, or epithelial cells. miRs were mostly macrophage-correlated and were all downregulated in asthma. One of the three peptides (LDHA) which were macrophage-correlated had a greatly-reduced fold change in asthma amongst smokers - the same pattern as seen with miRs.

To gauge the effect of steroid treatment and its influence on the model linear modelling was used with the same background confounders as used for each dataset previously, then the process was repeated with steroid dose included – normalised to beclomethasone. The percentage increase or decrease in the asthma coefficient when including steroid dose in the model was mapped to the network, and those molecules for which steroid dose was significant were

highlighted (Fig 5-19). This methodology may not be as appropriate for proteomics data as for mRNA and miRNA data and each molecule type used different confounders so results are not directly comparable. Nonetheless it indicates that in most cases steroid effect on asthma dysregulation in the network was relatively low. It also indicates quite consistently that steroids do have an effect in the "correct" direction with regards those molecules which interact with fibronectin and their interacting molecules, with the exception of MMP10, TIMP1 and SERPING1. Most of the molecules where the steroid dose itself is significant were also close to fibronectin in the network and all but one of them are peptides (HRH4). With the exception of LYZ (which was found regulated differently depending on smoking status) HRH4 was also the only molecule which had a significant steroid dose association and which appears to be further dysregulated rather than reducing the effect of asthma. The miR which targets it also has follows this pattern of regulation.

Clinical variables were also mapped to this network. Strong correlations exist between FEV1 and many molecules of the network (Fig 5-20). In particular again the fibronectin portion of the network is consistently correlated. Low FEV1 is indicative of asthma and as this portion of the network was negatively associated with asthma a positive correlation results between FEV1 and these molecules.

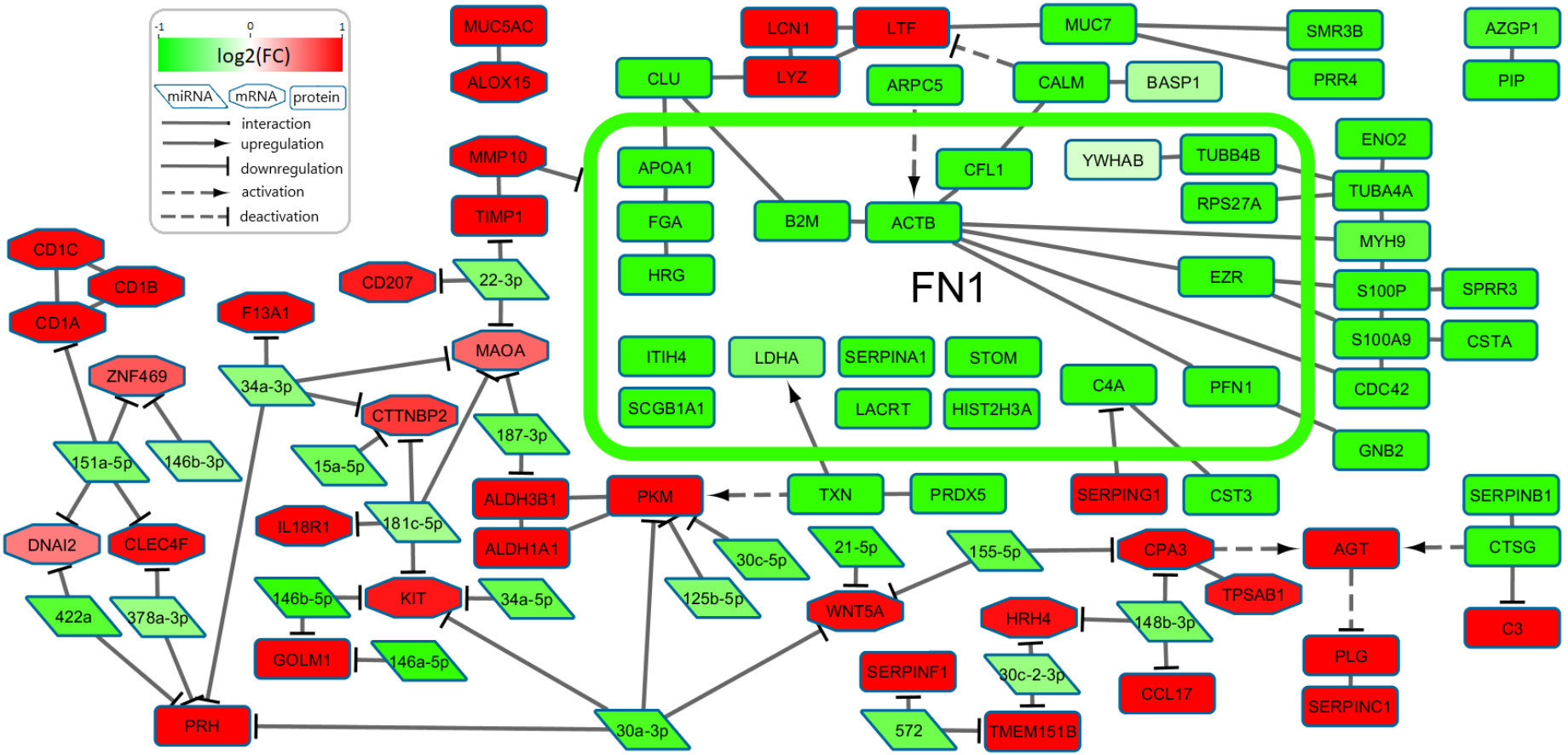

Fig. 5-15 Interaction Network Constructed from Hits from all Three Asthma Contrasts – Coloured by Asthma Fold Change
This network shows interactions between molecules with a significant two-fold difference in at least one of the three asthma contrasts. Interactions were gathered from interaction databases and directly from the literature and only inconsistent directions of fold change were excluded. The FN1 node is represented as a container for many of the nodes it interacts with to simplify the representation of the network. This graph shows that many of the molecules only found significant amongst non-smokers – including the majority of the miRs – maintain the same direction of fold change although they do not quite meet the significance threshold.

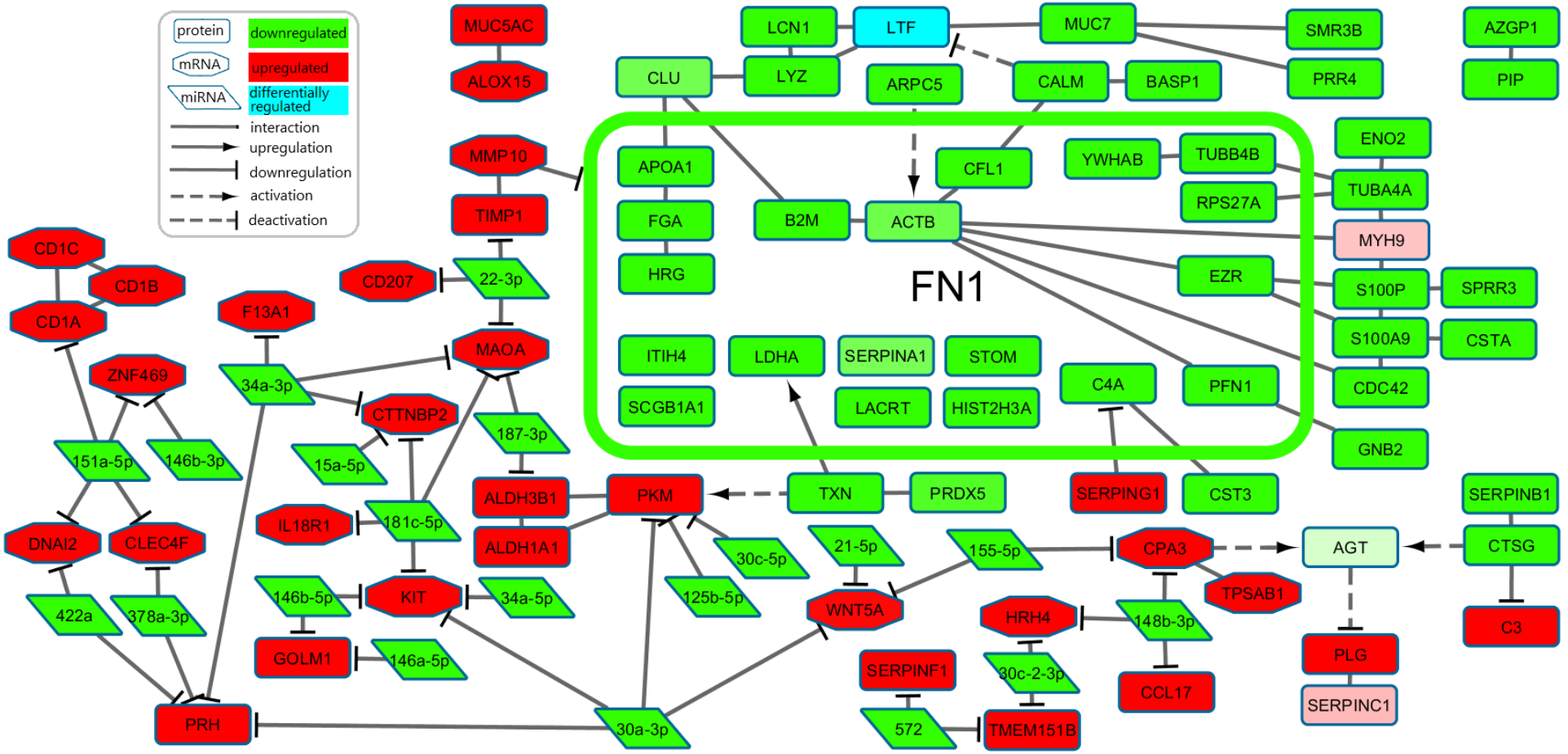

Fig. 5-16 Interaction Network Constructed from Hits from all Three Asthma Contrasts – Coloured by Fold Change of Asthma Amongst Non-Smokers
This network shows interactions between molecules with a significant two-fold difference in at least one of the three asthma contrasts. Interactions were gathered from interaction databases and directly from the literature and only inconsistent directions of fold change were excluded. The FN1 node is represented as a container for many of the nodes it interacts with to simplify the representation of the network. This graph shows that fibronectin and its associated molecules maintain the same direction of fold change although they do not quite meet the significance threshold. However LCN1, LYZ and LTF follow different patterns of expression.

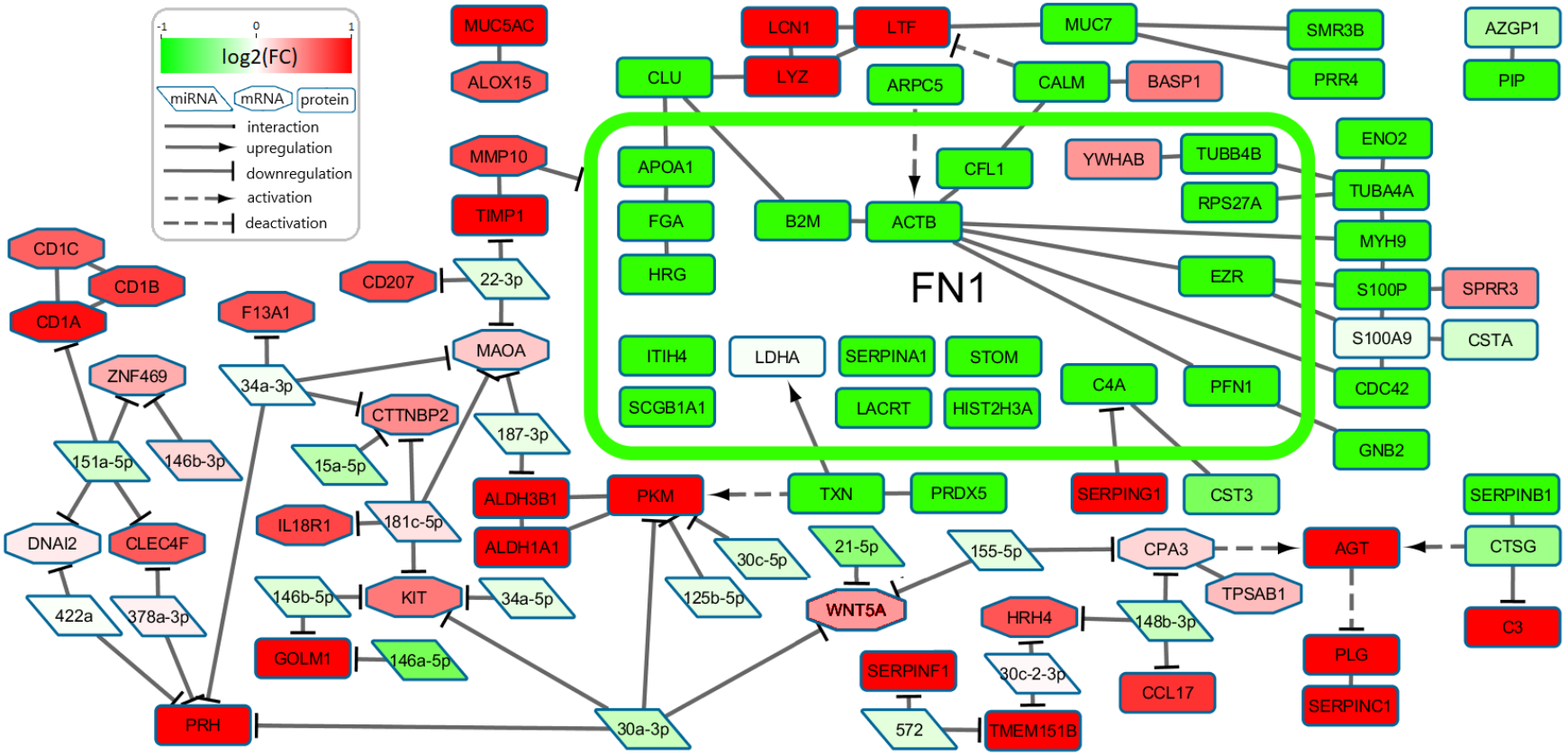

Fig. 5-17 Interaction Network Constructed from Hits from all Three Asthma Contrasts – Coloured by Fold Change of Asthma Amongst Smokers
This network shows interactions between molecules with a significant two-fold difference in at least one of the three asthma contrasts. Interactions were gathered from interaction databases and directly from the literature and only inconsistent directions of fold change were excluded. The FN1 node is represented as a container for many of the nodes it interacts with to simplify the representation of the network. This graph shows many molecules which show strong fold changes despite not passing the significant threshold (particularly the fibronectin-associated molecules). Alternatively many molecules, in particular the miRs, are much less dysregulated here than in the other contrasts. This may be due smoking status being responsible for similar changes.

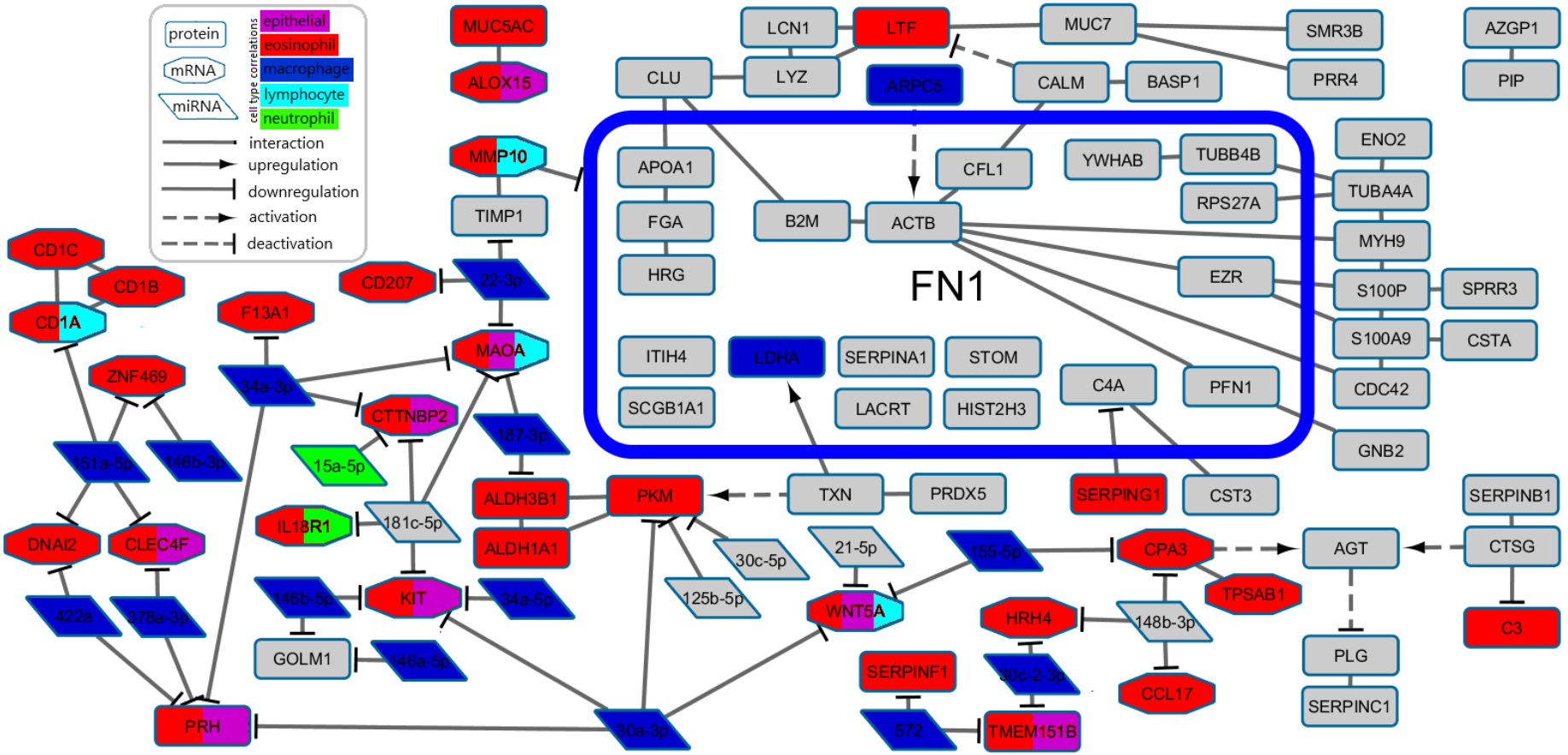

Fig. 5-18: Interaction Network Constructed from Hits from all Three Asthma Contrasts – Coloured by Cell Type Correlation
This network shows interactions between molecules with a significant two-fold difference in at least one of the three asthma contrasts. Interactions were gathered from interaction databases and directly from the literature and only inconsistent directions of fold change were excluded. The FN1 node is represented as a container for many of the nodes it interacts with to simplify the representation of the network. Most molecules with significant positive correlations with a cell type are eosinophil-correlated (peptides and mRNAs) or macrophage-correlated (mostly miRs but also some peptides including the central node fibronectin). Correlations with lymphocytes, neutrophils and epithelial cells were also detected.

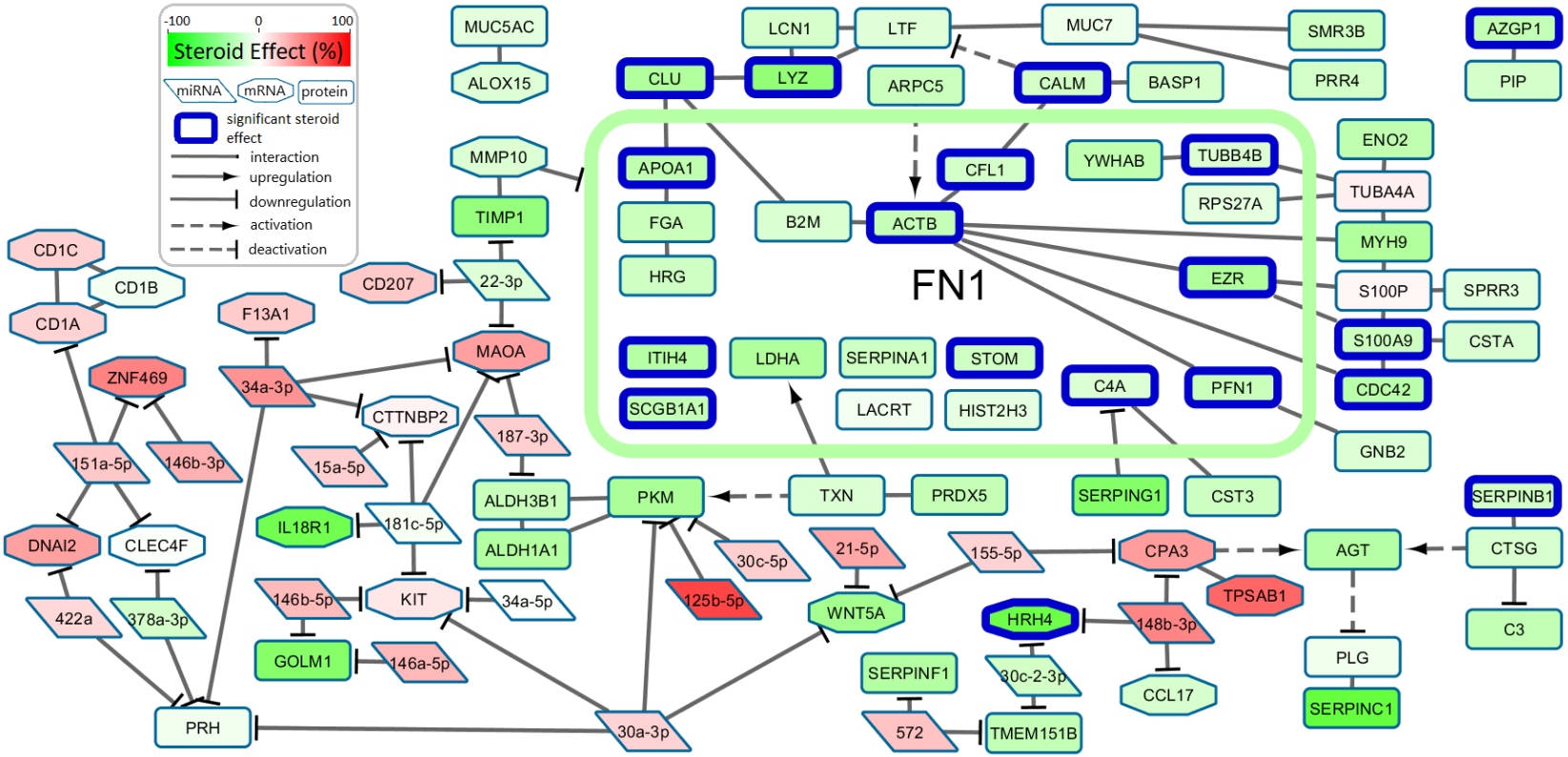

Fig. 5-19: Interaction Network Constructed from Hits from all Three Asthma Contrasts – Coloured by Difference in Asthma Estimate with the Inclusion of Steroid Dose

This network shows interactions between molecules with a significant two-fold difference in at least one of the three asthma contrasts. Interactions were gathered from interaction databases and directly from the literature and only inconsistent directions of fold change were excluded. The FN1 node is represented as a container for many of the nodes it interacts with to simplify the representation of the network. When steroid dose is modelled alongside asthma fibronectin and associated molecules are associated with a greater dysregulation due to the steroid dose working in opposition to asthmatic dysregulation. In many cases the steroid dose effect is significant.

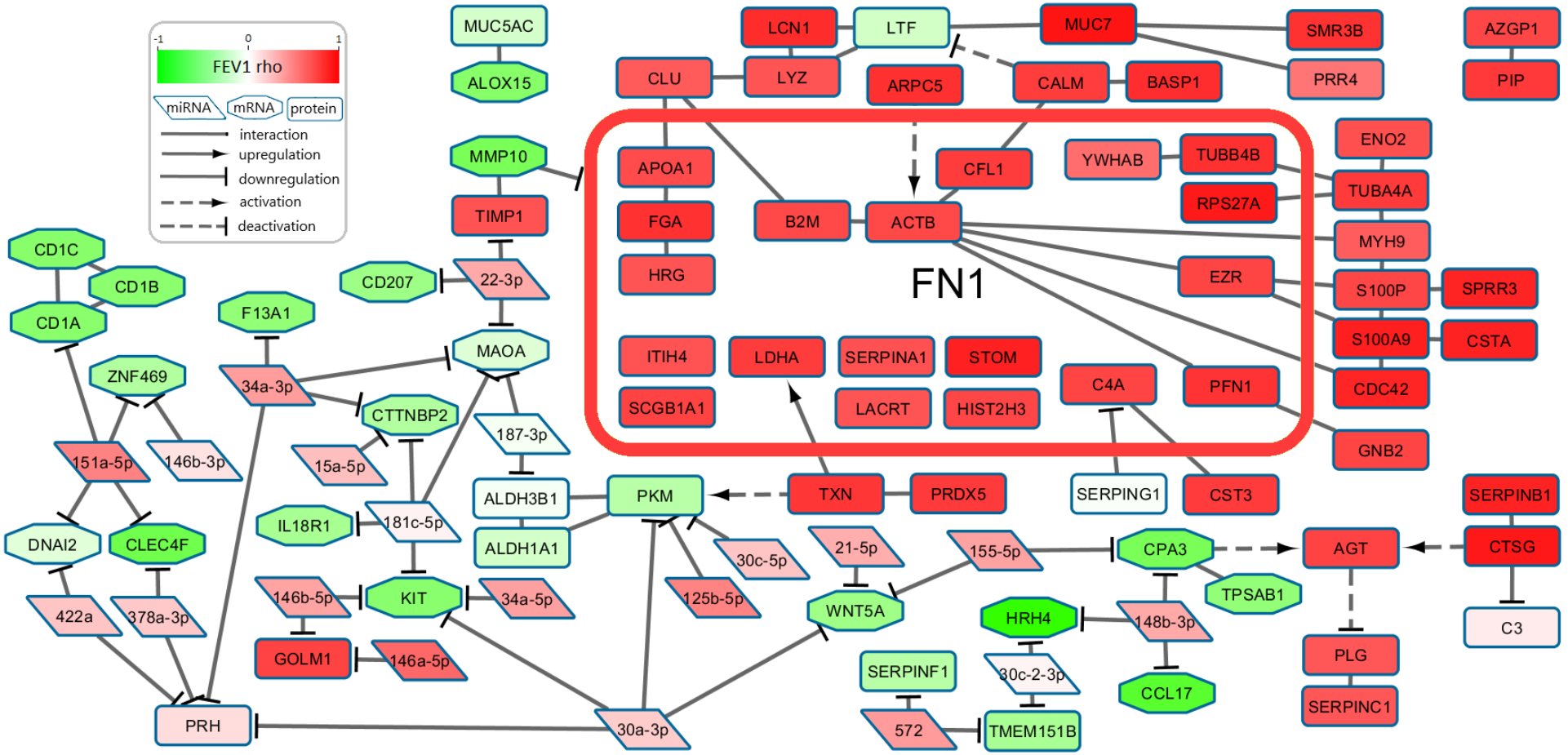

Fig. 5-20: Interaction Network Constructed from Hits from all Three Asthma Contrasts – Coloured by correlation with FEV1
This network shows interactions between molecules with a significant two-fold difference in at least one of the three asthma contrasts. Interactions were gathered from interaction databases and directly from the literature and only inconsistent directions of fold change were excluded. The FN1 node is represented as a container for many of the nodes it interacts with to simplify the representation of the network. Many molecules are seen to have a strong correlation with FEV1, particularly the fibronectin section of the network.

5.2.2 Nasal Epithelium

A large proportion of the genes in this gene set were best-modelled with the BMI parameter, resulting in a model with the lowest median p-value and the highest median adjusted R^2. Since modelling statistics (Table 5-9) demonstrate the impact on the data across a large number of genes BMI was included in the model (as in the induced sputum mRNA model).

Model	Median p	Median adj-R^2	Median AIC	# Genes Lowest AIC
~group	1.80E-01	3.09E-02	-199.28	6091
~group + bmi	1.67E-01	3.73E-02	-198.95	2712
~group + age	2.10E-01	2.90E-02	-198.14	1150

Table 5-9: Comparison of Statistical Models in Relation to Nasal mRNA data.
While the best fit is seen with the simplest model a large number of the genes are better described by the inclusion of BMI, and the inclusion of BMI explains more of the variation in the data. Adj-R^2: adjusted R^2, Adj-R^2: adjusted R^2, AIC: Akaike Information Criterion.

The model for miRNA with BMI had the highest median adjusted R^2, and while the simplest model has the best fit on average, both the BMI model and the model with age account for the best fit of data for a large proportion of the genes (Table 5-10). As with the sputum samples, both were used in the statistical model since it is not statistically prohibitive.

model	median p	median adj-R2	median aic	# Genes Lowest AIC
~group	0.265427	0.017407	-108.529	644
~group + bmi	0.2794	0.017977	-107.233	143
~group + age	0.293264	0.016085	-107.002	148

Table 5-10: Comparison of Statistical Models in Relation to Nasal miRNA data.
While the best fit is seen with the simplest model a large number of the genes are better described by the inclusion of BMI, and the inclusion of BMI explains more of the variation in the data. Adj-R^2: adjusted R^2, AIC: Akaike Information Criterion. Adj-R^2: adjusted R^2, AIC: Akaike Information Criterion.

5.2.2.1 Asthma

Four genes showed differential expression in asthma - TDRKH-AS1, FREM2, PBX3 and SIN3A (Table 5-11). No miRNAs showed differential expression even with both confounders removed from the model.

ENSG	Log_2(FC)	Ave Exp	p	p_{BH}	Gene Symbol
ENSG00000203288	-0.48	8.50	5.10×10^{-6}	0.03	TDRKH-AS1
ENSG00000150893	-0.56	9.41	6.83×10^{-6}	0.03	FREM2
ENSG00000167081	-0.31	7.84	1.25×10^{-5}	0.04	PBX3
ENSG00000169375	-0.22	10.38	1.61×10^{-5}	0.04	SIN3A

Table 5-11: Significant mRNAs in Asthmatics
These results are generated by asthmatics and non-asthmatics, regardless of smoking status, but with the effects of smoking statistically accounted for.

5.2.2.2 Asthma in Non-Smokers

OVGP1 was significantly downregulated in asthmatics amongst never-smokers (logFC: -0.70, pBH: 0.017). Eight non-coding RNAs were also found to have significant differential expression – six human miRs, one viral miR and one small nucleolar RNA (Table 5-12). miR-210-3p and miR-140-3p are both increased two-fold in never-smoker asthmatics.

ID	Log_2(FC)	p	p_{BH}
hsa-miR-374a-3p	0.81	6.01×10^{-6}	5.62×10^{-3}
hsa-miR-210	1.02	7.24×10^{-5}	2.32×10^{-2}
hsa-miR-548p	0.48	9.40×10^{-5}	2.32×10^{-2}
U18C	-0.41	9.94×10^{-5}	2.32×10^{-2}
kshv-miR-K12-6-5p	0.61	1.84×10^{-4}	3.45×10^{-2}
hsa-miR-140-3p	1.05	3.44×10^{-4}	4.40×10^{-2}
hsa-miR-193a-5p	0.40	3.75×10^{-4}	4.40×10^{-2}
hsa-miR-487b	0.46	3.76×10^{-4}	4.40×10^{-2}

Table 5-12: Significant mRNAs in Non-Smoking Asthmatics
These results are generated by non-smoking asthmatics and non-asthmatics.

5.2.2.3 Asthma in Smokers

No mRNAs or miRNAs were found significantly dysregulated in asthmatics amongst smokers after multi-test correction regardless of adjusting for BMI and/or age.

5.2.2.4 Smoking

585 genes were significantly downregulated and 458 significantly upregulated (7 and 38 respectively when restricted to two-fold changes). Of the larger set 304 terms were significantly enriched, including 'lipid antigen binding' and 'regulation of dendritic cell differentiation' at over 50%. An analysis of the smaller set of 44 genes identified enriched upregulation of six clusters of terms, two individual terms and the enriched downregulation of an "Interferon-gamma-mediated signalling pathway" term cluster.

No miRs were found significant when adjusted for confounders. Without adjustment however miR-193b-3p was found significantly upregulated in smokers (log2FC: 0.52, pBH: 0.03) and miR-424-3p significantly downregulated (log2FC: -0.33, pBH: 0.03).

5.2.2.5 COPD

459 genes were significantly downregulated and 657 significantly upregulated (6 and 29 respectively when restricted to two-fold changes). Nine clusters and five individual terms were enriched in this set (Fig 5-35), most of which are upregulated. "Fatty acid elongation, saturated fatty acid" has 80% of its genes found to be significant.

Eight miRs were upregulated in patients with COPD, four of which had more than a two-fold increase (Table 5-13).

miR	Probe ID	logFC	AveExpr	P.Value	adj.P.Val
hsa-miR-199a-3p	hsa-miR-199a-3p_st	1.423595	4.947135	1.30E-05	0.012193
hsa-miR-126-3p	hsa-miR-126_st	1.772569	6.273463	7.23E-05	0.023803
hsa-miR-363-3p	hsa-miR-363_st	0.55563	3.821824	8.58E-05	0.023803
hsa-miR-143-3p	hsa-miR-143_st	1.53381	6.001103	0.000162	0.023803
hsa-miR-145-5p	hsa-miR-145_st	1.603344	6.524047	0.000165	0.023803
hsa-miR-199a-5p	hsa-miR-199a-5p_st	0.985083	4.268569	0.000194	0.023803
hsa-miR-192-5p	hsa-miR-192_st	0.575629	4.680521	0.000202	0.023803
hsa-miR-195-5p	hsa-miR-195_st	0.962701	5.250416	0.000204	0.023803

Table 5-13: miRNAs Upregulated in Association with COPD

5.2.2.6 Comparing Contrasts

SIN3A was downregulated in both asthma and COPD, and PBX3 was downregulated in both asthma and smoking. 49 terms were shared between asthma and COPD. None of the miRs found significant with smoking or COPD overlapped with each other or with those from the asthma amongst non-smokers contrast (regardless of using confounders or not). Asthma and COPD hits in the nasal epithelium were clustered alongside a heatmap of the data (Fig. 5-21). This heatmap is based on fewer molecules of interest than the induced sputum heatmap and may show some clustering, but less than that which was seen in the induced sputum. This may be due to the sample type obtained being from the upper rather than lower respiratory tract. The clade of nine molecules at the bottom of the graph includes all eight of the miRs included.

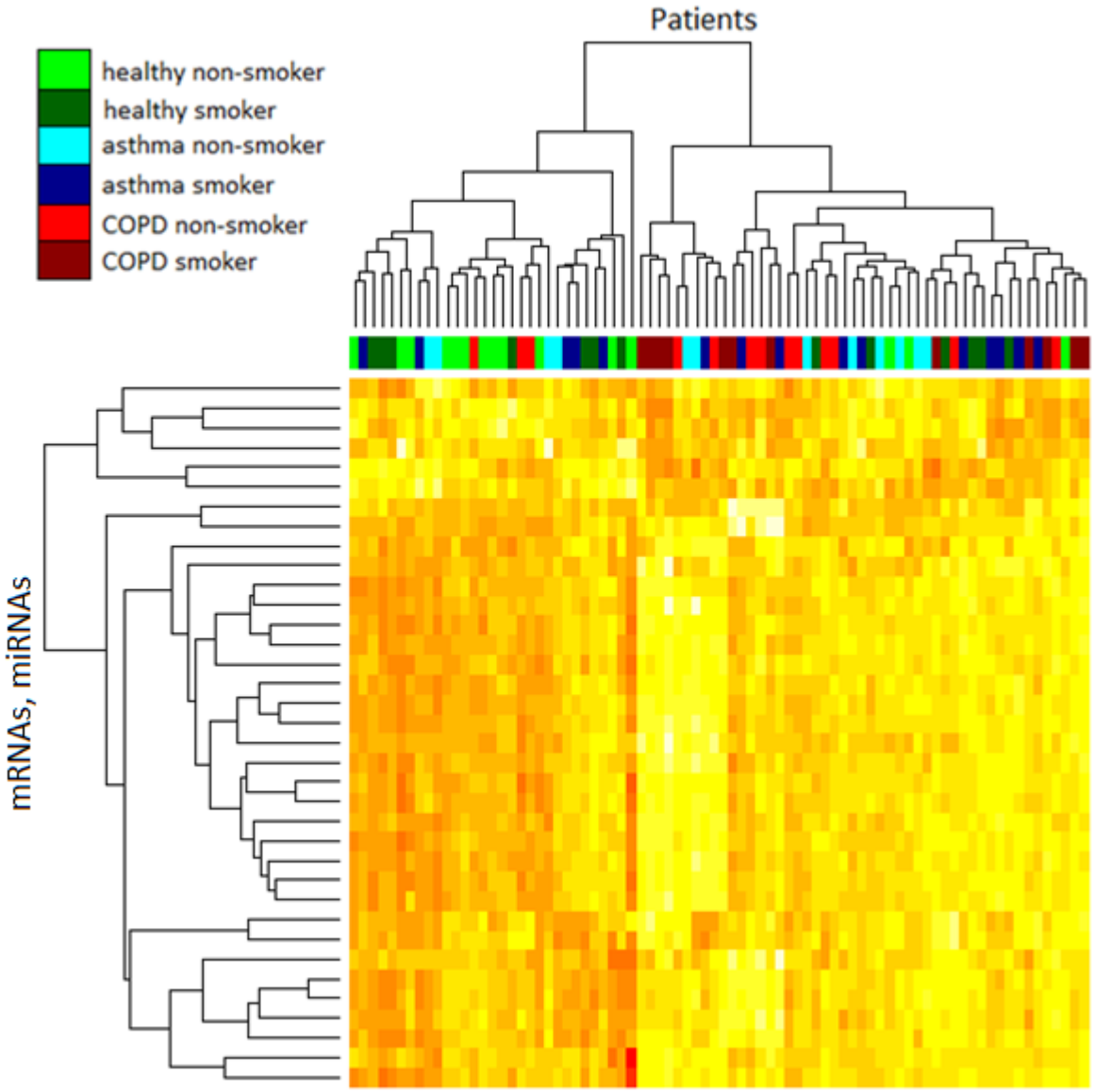

Fig. 5-21: Heatmap Showing Dysregulation of Genes in the Nasal Epithelium
Heatmap rows (genes) and columns (patients) were clustered. As with induced sputum small groups of phenotypes appear to cluster, however overall there appears to be less clustering of phenotypes than with induced sputum results. mRNA: messenger RNA, ncRNA: non-coding RNA.

5.2.3 Sample Types

Clustered heatmaps were produced to represent the significant microarray results across both sample types. One was produced for asthma, COPD, and both conditions together (see Figures 5-22, 5-23, 5-24 respectively). All show clustering of diseases and to a lesser extent combinations of disease and smoking status. Grouping is more apparent in when viewing the clustering results of healthy controls, asthmatic patients, and COPD patients together.

Each phenotype clusters to into small groups in Fig 5-22, particularly a large group of mostly healthy controls on the left of the graph. In some cases small clusters of asthmatics of a certain smoking status can be seen adjacent to healthy patients of the same smoking status. The same four miRs which clustered together based on induced sputum data alone (Fig 5-14; miR-146a-5p, miR-146b-5p, let-7f, and let-7g) again cluster together. A greater degree of clustering is seen in the cross-sample-type COPD data in Fig 5-23. The largest patient clusters in Fig 5-24 are those of healthy controls, reflecting the similarities in the molecular dysregulation between asthma, COPD and smoking status.

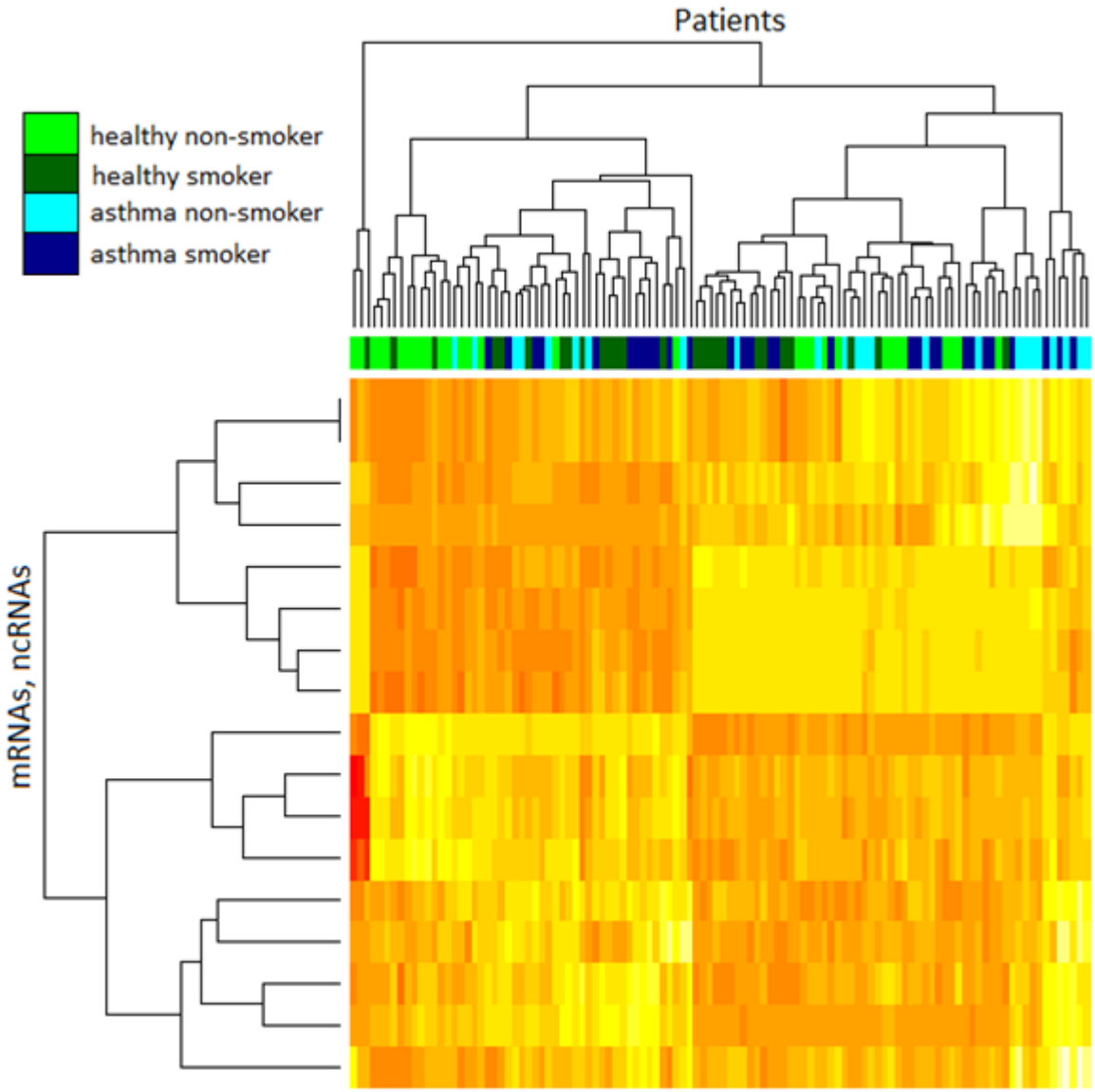

Fig. 5-22: Heatmap Showing Dysregulation of Genes in the Induced Sputum and Nasal Epithelium with Asthma
Heatmap rows (genes) and columns (patients) were clustered. Colours at the top denote the subgroup of the patient – green: healthy non-smoker, dark green: healthy smoker, blue: asthma non-smoker, dark blue: asthma smoker.

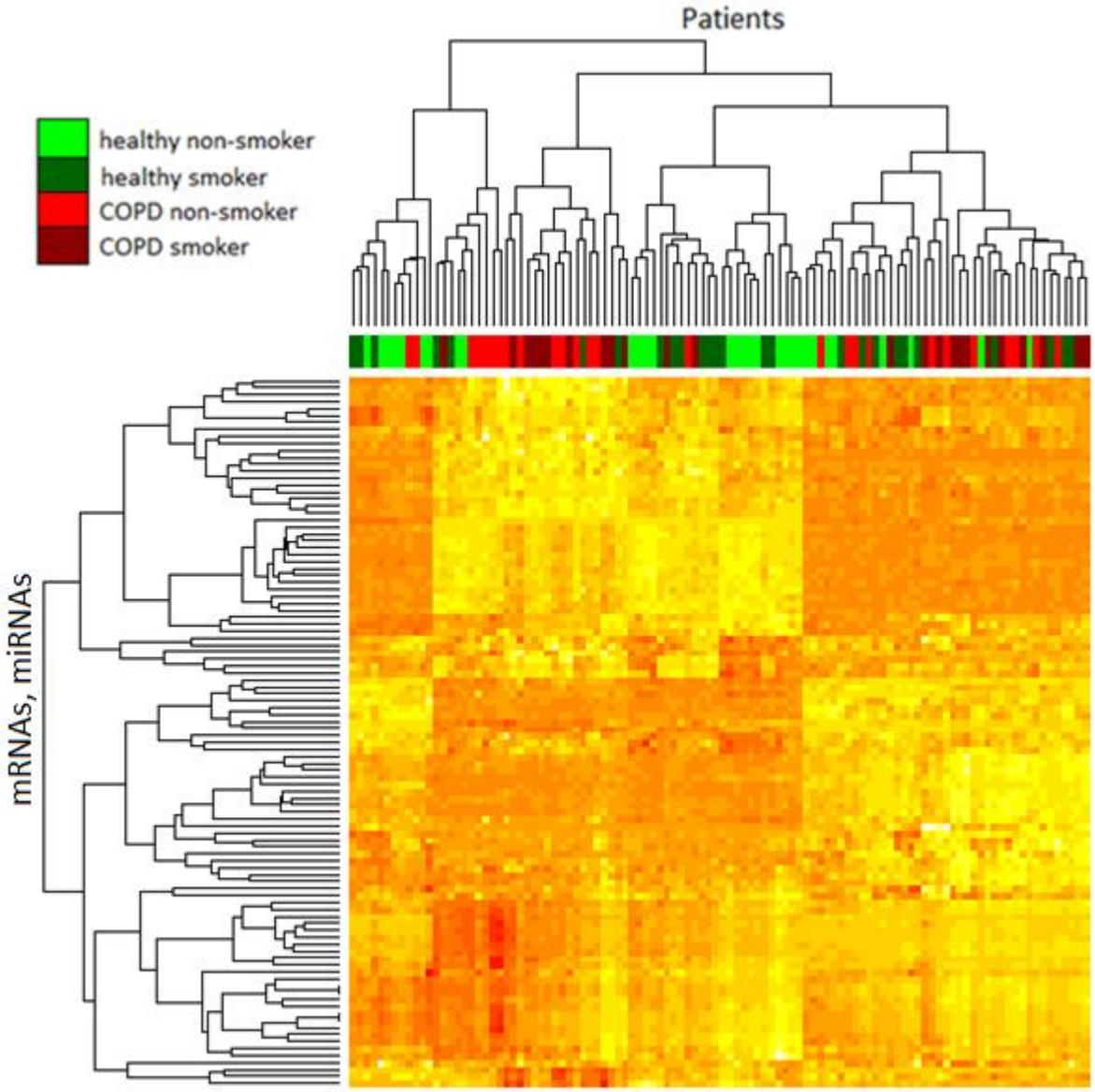

Fig. 5-23: Heatmap Showing Dysregulation of Genes in the Induced Sputum and Nasal Epithelium with COPD

Heatmap rows (genes) and columns (patients) were clustered. Colours at the top denote the subgroup of the patient – green: healthy non-smoker, dark green: healthy smoker, red: COPD, dark red: COPD smoker.

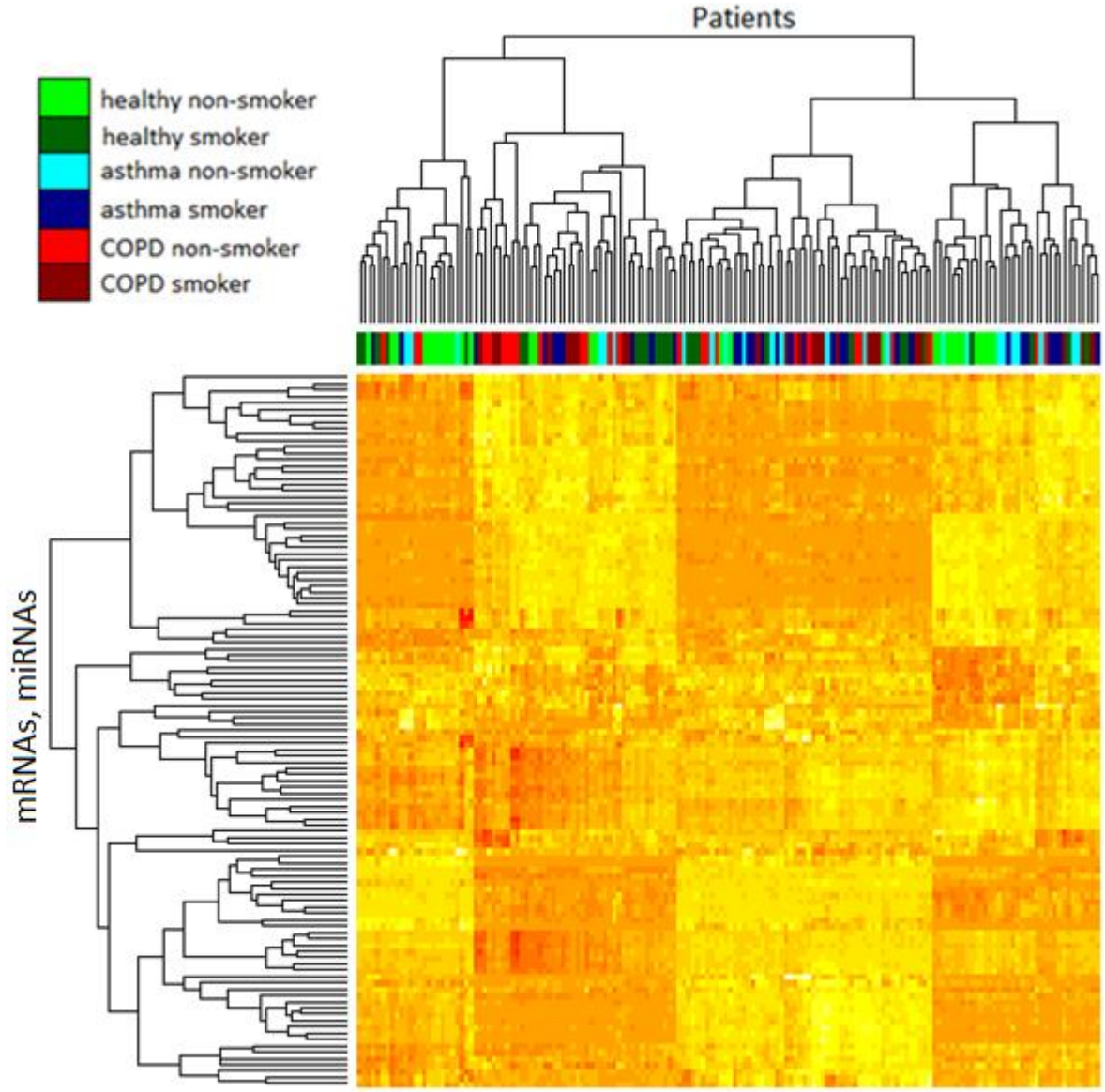

Fig. 5-24: Heatmap Showing Dysregulation of Genes in the Induced Sputum and Nasal Epithelium with COPD and Asthma

Heatmap rows (genes) and columns (patients) were clustered. Colours at the top denote the subgroup of the patient – green: healthy non-smoker, dark green: healthy smoker, blue: asthma non-smoker, dark blue: asthma smoker, red: COPD, dark red: COPD smoker.

5.3 Discussion

Numerous significant results were found via mRNA and miRNA microarray investigations pertaining to asthma, COPD, and smoking status. Metabolomic and proteomic data were compromised by incorrect sample selection and by metabolite samples being stored in different media. Despite this proteomics analysis still yielded numerous statistically significant results, but metabolomics analysis did not show any identified metabolites of significance after multiple-test correction.

Results of term enrichment analysis indicated consistently enriched gene sets in asthmatic patients, upregulated for classic immune system terms related to general mediators of immune response, chemokines, antigen binding, and leukotrienes, and with other terms including the complement system, fibroblast growth factor, matrix metalloproteinase (MMP) activation, and vesicle endocytosis. Term enrichment also indicated consistently downregulated gene groups related to glucose metabolism, platelet degranulation, and the secretion of proteins normally found in the saliva.

Smoking and COPD contrasts for induced sputum show significant widespread dysregulation across thousands of genes, however less than a hundred in each have a two-fold difference. Term enrichment analysis of significant COPD genes truncated by fold change also showed dysregulation of MMP activation and the complement system as in asthma, but with different genes involved. In general there were many genes which overlapped between asthma and COPD contrasts but only a few when restricted to two-fold changes.

Results of the interaction network generated for the sputum results show a large cluster of downregulated extracellular matrix fibronectin-associated genes, and a large number of genes potentially knocked down by miRNAs. The while much of the fibronectin cluster does not retain significance in the smoking asthma contrast it does maintain large fold changes for most of its constituents and nearest neighbours. This section of the network is shown to correlate strongly with FEV1 and while steroid dose does seem to have a consistent positive effect

in returning it to healthy levels of expression the effect is mild and only significant for some molecules.

The scope of the results is vast, however this discussion will focus more on the asthmatic sputum work as there were the most types of biomolecules available for that sample type and disease (proteomic and metabolomic data were unavailable for the COPD samples and for nasal epithelium samples), and because that sample type is more reliable in terms of being from the lower respiratory tract rather than the surrogate nasal tissue. The discussion focuses on two-fold dysregulated molecules significant after multi-test correction and comparisons are drawn against these data to smoking and COPD where appropriate.

5.3.1 Calcium Signalling and Homeostasis

Calmodulin (CALM) is a ubiquitously-expressed protein encoded by three separate genes and it is downregulated in asthmatics regardless of smoking status. It regulates the activity of a large number of proteins by Ca^{2+} signalling, affecting various relevant processes such as inflammation, metabolism, apoptosis, immune response, secretion, and smooth muscle contraction. CALM regulates levels of cytosolic Ca^{2+} both by CALM-mediated Ca^{2+} induced inactivation (CDI) and by CALM-mediated Ca^{2+} induced facilitation, both in terms of cellular uptake and intracellular transport across internal membranes (366). While Ca^{2+} signalling is required for many central processes and the functioning of many proteins besides CALM, CDI is also essential as excessive intracellular levels of Ca^{2+} are toxic (367).

Although CALM (complexed with Ca^{2+}) is essential to priming smooth muscle for contraction, and increased smooth muscle contraction is a major feature of asthma, the levels of CALM in induced sputum samples (primarily comprising leukocytes and epithelial cells) may not be reflective of those in the smooth muscle cells, so a reduction in CALM here is likely not to be contradictory.

A recently discovered role of CALM in the airway epithelium is its regulation of ciliary beat frequency, as studied in the nasal epithelium (368). Inhibition of CALM was shown to reduce ciliary beat frequency which in turn could reduce mucociliary clearance of mucous (excessively-produced in asthma) and harmful foreign substances. This excess mucous obstructs the airways and provides an advantageous environment for bacteria.

The interaction of CALM with calmodulin-dependent protein kinase II (CAMKII) induces the activation of numerous transcription factors, including NF-κB, which has hundreds of known targets. NF-κB is typically described as pro-inflammatory due to its induction of pro-inflammatory cytokines and chemokines, and it has been associated with several inflammatory diseases including asthma and COPD. Similarly its role in relation to oxidative stress is complex and context-dependent, being capable of inducing both pro- and anti-oxidative responses (369). A protective effect of NF-κB has been demonstrated in the epithelium in which it promotes epithelial cell survival and mucosal barrier integrity (370). The dysregulation of a number of other genes (e.g. genes related to tight junctions) discussed later also contribute to reduced epithelial barrier function, possibly via NF-κB whose expression has been shown to repair epithelial tight junctions in lung epithelia (371).

In addition many of the studies which implicate NF-κB as a central driver of pathological oxidative stress and inflammation in asthma rely on data from murine models (otherwise-healthy animals exposed to a variety of possible allergens) and *in vitro* studies, which may be misleading. Studies in murine models show inhibition of CAMKII activity inhibiting NF-κB activity and suggest that it is essential for proasthmatic effects such as airway hyper-reactivity, eosinophilia, and mucin 5AC (MUC5AC) expression (372;373). Despite a large reduction in CALM, which CAMKII is dependent on for many of its interactions including NF-κB activation, the airway hyperreactivity is unresolved in our cohort, eosinophilia is present, and MUC5AC is greatly upregulated. In contrast to the idea of increased CALM/NF-κB activity as a main driver of a pathology-inducing level of oxidative stress and inflammation, a reduced level of CALM expression and NF-κB activity (implied by reduced CALM and other NF-κB activators) may contribute to pathology by the reduced mucociliary clearance

action of CALM, the disruption of the epithelial barrier function by reduced NF-κB, and a reduced capability of macrophages to kill bacteria – CALM and NF-κB are both implicated in macrophage survival and bacterial killing (374-377). This would be consistent with an allergen-sensitised respiratory environment and with the tendency for bronchitis to occur in the context of severe asthma.

Over 30 small clinical trials of calcium channel blockers such as verapamil or nifedipine have shown little if any effect on asthma symptoms in response to exercise or histamine challenge (378;379), and no effect as maintenance therapy (380). A small effect on FEV1 in exercise-induced asthma has been confirmed by a recent meta-analysis (381). Interestingly, one study showed a paradoxical bronchoconstrictor effect with administration of doses greater than 10mg of verapamil, which could be representative of the state of dysregulation in this severe asthma cohort (382).

20 known targets of NF-κB were found significantly dysregulated amongst asthmatic patients, and the expression of 10 of these (*SERPINB1, C4A, B2M, FN1, S100A6, LCN2, LTF, SERPINA1, ENO2,* and *CCL17*) were significantly correlated with CALM expression. Overall CALM was significantly correlated with 81 of the 181 other two-fold dysregulated molecules (mRNA/miRNA/peptides), including all of the three proteins shown to interact directly with CALM in the interaction network (BASP1, CFL1, and LTF).

Lactoferrin (LTF) has a wide range of effects in host defence, including anti-bacterial, anti-fungal, and anti-viral effects (383). These functions are at least partly explained by its interactions with cell surface receptors of cells of innate and adaptive immunity, and it also interacts with the surface of epithelial cells. It is has a bacteriocidal effect working in concert with lysozyme (LYZ; an NF-κB target which was downregulated in non-smoking asthma, but upregulated in smoking asthma), and acting alone each protein is bacteriostatic (384). Its bacteriocidal effect may be due to its ability to induce bacterial agglutination which can promote phagocytic killing. Its bacteriostatic effect may be due to its binding free iron ions which otherwise would allow bacterial growth. This iron-binding also gives it the capability to reduce oxidative stress. LTF also binds and neutralises LPS and other pathogen-derived molecules giving it a role in anti-

inflammation and antigen presentation (383), it attracts eosinophils (which is it correlated with in our data) (385), and appears to induce wound healing by promoting fibroblast proliferation and migration and stimulating re-epithelialisation (386).

While it is upregulated with asthma in this dataset amongst smokers, there are lactoferrin peptides regulated in both directions amongst non-smokers - and several peptides of LTF have been reported to be active with respect to host defence (387;388). These two peptides are not simply in anti-correlation with each other, as some asthmatic patients had very high levels of both, relative to the rest of the samples.

CALM is known to bind lactoferrin in a Ca^{2+}-dependent manner, and inhibit its ability to agglutinate bacterial cells (389). CALM acting both to increase its transcription through NF-κB and acting as an antagonist forms a kind of negative feedback loop, at least for the purposes of its bacteriocidal capacity. Regardless of which LTF peptide is responsible for this agglutination effect (or both), CALM is downregulated so it is less likely to work as an antagonist. Beyond being an antagonistic interaction the CALM-LTF complex has a secondary functional interface of unknown specificity, and could be involved in one of the other known LTF effects (390). Human LTF has been shown to induce asthma symptoms in an atopic mouse line, also resulting in increased MUC5AC and IL5 as also seen in this dataset (391). The presence of allergens *in vitro* causes asthmatic neutrophils to release LTF, but not neutrophils from healthy patients, and the amount released increases with asthma severity (392).

Cofilin-1 (CFL1) encodes an actin-binding protein involved in both polymerisation and de-polymerisation, and although it is downregulated in this data is has been previously found upregulated in severe asthma (393). Brain associated protein 1 (BASP1) is proposed to act as a CALM antagonist, interfering with the action of other CALM-bound transcription factors, particularly proteins of the pro-proliferation *MYC* gene (394). It is also involved in the inhibition of WT1 which is another gene which is involved in cell proliferation amongst other activities, but CALM is not as yet implicated in this interaction (395). BASP1 also binds

phospholipids and actin (including beta actin, downregulated in our data), and may be involved in the maturation of actin filaments (396).

Calcium signalling regulates tight junction permeability and eicosanoid production by eosinophils (397;398). Extracellular Ca^{2+} is also required at the epithelium for cell polarisation and the formation of tight junctions. Interestingly Cd^{2+}, a component of cigarette smoke and air pollution, blocks formation of tight junctions and the influx of Ca^{2+} (399). Calcium requirements are different between females and males and across different ages, which may account for some of the variation in asthma frequency and severity observed. Differences in levels of various hormones may also account for some of the demographic differences recorded (400). Estradiol stimulates growth factor signalling and causes an increase in intracellular Ca^{2+} and drives some of the known molecular processes involved in asthma including IgE-induced mast cell degranulation, and leukotriene production (401). The estrogen-dependant OVGP1 (oviductal glycoprotein 1) was downregulated in nasal epithelium of non-smoking asthmatics.

Patients with various lung diseases often have low vitamin D, which is required for calcium absorption (402). In asthma vitamin D levels have been found reduced, positively correlated with FVC and FEV1, and negatively correlated with exercise-induced bronchoconstriction (403), however levels of serum Ca2+ do not correlate with this difference in vitamin D level and symptom severity in asthmatic patients so calcium absorption/deficiency is likely not responsible (404). Vitamin D may instead be beneficial through various other routes such as the suppression of IL-5 (upregulated in this dataset) secretion, in an effect which may be distinct from corticosteroid action (405). Vitamin D supplementation is somewhat effective against asthma and COPD (406;407).

28 of the dysregulated asthma genes' products are calcium-binding. Several of these (other than CALM) are involved in calcium homeostasis, including AHSG, PRH1/PRH2, CAPSL, GNB2, NUCB1, S100 proteins, STOM, and WNT5A.

Alpha-2-HS-glycoprotein precursor (AHSG), which is upregulated and eosinophil-correlated can inhibit calcium salt precipitation (408). It also inhibits the

phosphorylation of the insulin receptor and insulin receptor substrate 1 (409), which in turn inhibits the PI3K and MAPK pathways, and processes of energy metabolism. It may also lead to the inhibition of T-cell immunity during inflammation and infection (410). An upregulated peptide of salivary acidic proline-rich phosphoprotein 1/2 precursor (encoded by genes PRH1/PRH2; PRH in network graph) encodes a proline-rich protein normally secreted into the saliva in a number of forms due to genetic variation and extensive modification such as glycosylation and phosphorylation. In the saliva it is involved in calcium homeostasis by inhibiting calcium phosphate precipitation (411).

Calcyphosin-Like Protein (CAPSL) does not have a known function but as a paralog of calcyphosin it is predicted to be involved in ion transport. G Protein Subunit Beta 2 (GNB2) is involved in the PI3K pathway, and it is normally involved in inhibition of Ca^{2+} channels and it is downregulated in asthmatics in our data. Nucleobindin 1 (NUCB1) is a major calcium binding protein of the Golgi and it may have a role in calcium homeostasis. S100 proteins (discussed more extensively later) are regulatory molecules involved in many processes including calcium homeostasis. Ca^{2+} channels may also be affected by the downregulation of stomatin (STOM), which interacts with calcium pump Ca-ATPase-4. Wnt Family Member 5A (WNT5A), acts through the non-canonical WNT pathway to increase levels of Ca2+ via downstream release through channels of the endoplasmic reticulum.

5.3.2 Fibronectin, the Extracellular Matrix, and Airway Remodelling

Decreased fibronectin (FN1) production and secretion by airway epithelial cells contributes to the dysregulated epithelium repair, possibly due to its importance to cell adhesion and migration (412). This downregulation may be related to the reduced CALM/NF-κB signalling mentioned above. Fibronectin has also been reduced in macrophages, fibroblasts and bronchial epithelial cells in response to bacterial and viral infections (413-415). Other groups however have reported increased fibronectin expression in the airway epithelium cells of asthmatic

patients (416), and suggested this may be responsible for pathological airway remodelling. Fibronectin is also involved in binding and clearance of fibrin by macrophages (417). Fibronectin is cleaved by plasmin, the activated form of plasminogen (PLG; a peptide of which is highly upregulated in our data), along with fibrin, thrombospondin and von Willebrand factor. Fibrin and fibronectin were both found downregulated. Angiotensin II, a downstream peptide of *AGT* expression, inhibits activation of plasminogen, but it is unclear which peptide related to *AGT* was detected as upregulated.

Fibronectin can form cross-links with other molecules by action of a transglutaminase. In the study by Hallstrand et al transglutaminase 2 was the only gene expressed differently at baseline and it was suggested that its role in asthma was to cause a sustained increase in the activity of secreted phospholipase A_2 (demonstrated *in vitro*) which regulates the release of arachidonic acid leading to eicosanoid production. Factor XIII (F13A1) is the only transglutaminase included in our list of putative markers. Like many of the eosinophil-correlated enzymes it is Ca^{2+} dependent and interestingly it was one of three genes also found significant and two-fold increased in COPD.

Coagulation factor XIII cross-links fibrin chains in a calcium-dependent manner as the final step of the classic coagulation cascade, generating a clot to which additional proteins are attached by cross-linking, including complement C3 which was upregulated and eosinophil-correlated in our data. It can also cross-link fibronectin or alpha-2-plasmin inhibitor (a serine protease inhibitor responsible for inactivating plasmin) to fibrin, fibronectin to itself, fibronectin to collagen and lipoproteins to fibrinogen (418;419). Factor XIII has been associated with asthma previously and its expression is correlated with type 2 immune response and airway obstruction (420;421). Factor XIII enhances the proliferation and migration of fibroblasts, and inhibits their apoptosis (422). It binds to the fibroblast ECM and matrix assembly sites, where it remains active until it is eventually internalized and degraded (423). It also appears to reduce the proteolytic breakdown of collagen precursors (424). Factor XIII has been shown to be highly correlated with other significant, eosinophil-correlated genes *CCL17* and *CD207* in induced sputum (425).

MMP10 is an extra-cellular calcium-dependant enzyme which degrades fibronectin, gelatins and collagen. It is upregulated in eosinophilic asthma where it is suspected of being involved in airway remodelling (426). It is expressed in response to damage or infection, and can modulate macrophage activation (427). MMP10 drives the conversion of M1 macrophages to M2 macrophages and may be responsible for fibrotic clearance by M2 macrophages (428). This shift to M2 macrophages is also supported by a reduction of M1 macrophage miRs such as the most significant miR hsa-miR-146a-5p. MMP10 also has increased abundance in bronchial biopsies with high mucosal eosinophils (429). In our data MMP10 is upregulated and correlated with eosinophils and lymphocytes, but not macrophages. It is also found significantly upregulated two-fold in COPD. Despite the upregulation of MMP10, metalloproteinase inhibitor 1 precursor (TIMP1), an inhibitor of MMP10 is upregulated (430).

5.3.3 Actin

The actin cytoskeleton has several relevant roles in asthma including smooth muscle contraction and airway hyperresponsiveness and remodelling (431). It also plays a role in cell adhesion via actin filaments at adherens junctions. ACTB (actin beta) is downregulated in asthmatics in our data. ALDH1A1 is strongly upregulated and eosinophil-correlated and is involved the retinoid cascade which regulates genes involved in actin assembly/disassembly such as gelsolin and actinins. A number of dysregulated molecules are involved in actin binding or actin (de)polymerisation including BASP1, CFL1, CDC42, MSN, PFN1, ARPC5 and EZR – all of which except MSN (moesin) were downregulated. Cellular calcium concentrations affect also actin interactions/conformation and antagonism of CALM has been shown to inhibit polymerisation (432).

Ezrin (EZR) is a membrane-actin cytoskeleton linker, involved in cell-to-cell adhesion and microvilli formation, and it is also involved in numerous signalling pathways (433). It is also biomarker of asthma - reduced levels of this protein are present in exhaled breath condensate and serum of asthmatic patients and the (IL-13-driven) reduction in EZR is correlated with reduction in lung function

(434). Knockdown of EZR also results in increased epithelial permeability, and similarly inhibition of AKT (which activates EZR) and NF-κB (which is activated by EZR) can block epithelial-to-mesenchymal transition (EMT) (435). EMT creates a mesenchymal cell which detaches from the basement membrane, generates fibroblasts, and is involved in inflammation-driven wound healing (436). In cases of chronic inflammation wound healing goes unchecked and it leads to fibrosis (437). This mechanism may be involved in asthma (438). In airway epithelial cells EZR is downregulated by IL13 (439).

Ezrin binds F-actin and is involved its assembly, and in interactions between it and the cell membrane. Activated Ezrin provides membrane tension (440), whereas fluid shear stress induces Ezrin activation via Ca^{2+} dependant Akt phosphorylation (441). It has been suggested that EZR is secreted in epithelial-derived exosomes, however in our data EZR is negatively correlated with epithelial cells. Intracellularly EZR activation by binding of an S100P dimer precedes its translocation to the cell membrane and induction of microvilli (442;443).

5.3.4 Cell Adhesion

Tight junctions and adherens junctions bind the cells of the airway epithelium, forming a barrier against the external environment including allergens and pathogens. Disruptions to this barrier open the airway tissues to allergens causing inflammatory damage, caused by the recruitment and activation of immune cells. Fibronectin and actin have a profound effect on cell adhesion, amongst other structural molecules and regulators, many of which were found dysregulated in this dataset.

Microtubules are important for maintaining tight junctions and products of two microtubule genes (*TUBA4A* and *TUBB4B*) and a microtubule-associated gene (*KLC3*) are all downregulated.

Trefoil factor 3 (TFF3) has been variously reported to cause downregulation and internalisation of E-cadherin, and the phosphorylation of ß-catenin (CTNNB1) which induces dissociation of the E-cadherin/catenin complex from the actin cytoskeleton (444-446). These disruptions of the E-cadherin/catenin complexes of adherens junctions cause reduced intercellular adhesion, promoting cell migration in a manner which aids epithelial repair (447;448). It has been shown that inhibition of E-cadherin function can prevent tight junction and desmosome formation in vitro however TFF3 also affects tight junctions by causing the redistribution of ZO-1 from the cytoplasm to tight junctions, which may add structural support (449). Further, TFF3 increases expression of the 'tightening' claudin 1 and decreases expression of the cation channel forming claudin 2 which results in reduced permeability in vitro. Presumably these changes allow maintenance of barrier function despite the reduced cell adhesion from adherens junctions.

Tetraspanin 8 (TSPAN8) is a transmembrane protein which is known to complex with integrins. Tryptase Alpha/Beta 1 (TPSAB1) is the major neutral protease released by mast cells to degrade the ECM (450). While the study by Hallstrand et al found no difference at baseline, of the aforementioned molecules, some of them were found differently expressed at baseline in another study by Dougherty et al (451), however only within a high Th2 subgroup of asthmatics – CPA3, CST1 and TPSAB1. They are part of a mast cell associated cluster of differentially expressed genes along with some other significant genes from our data – KIT, PRR4 and SPRR3.

Dorscheid et al found that in an airway epithelial cell line, corticosteroids induced epithelial apoptosis (452). This suggests that while corticosteroids clearly provide a protective effect in the short-term by reducing inflammation, their chronic use may promote the epithelial damage partly responsible for driving that same inflammation. Corticosteroid treatment is also responsible from thickening of the epithelium (453). Epithelial shedding triggers induction of myofibroblasts and fibrogenesis (454).

In our data cystatin 1 (CST1) is upregulated, and cystatin 3 is downregulated allowing cystatin 1 to act unrestricted and preventing inhibition of cathepsin B.

Cystatin A (CSTA) being downregulated is also consistent with preventing inhibition of cathepsin B. CSTA is a type I cystatin (or stefin), and it is found both intracellularly and extracellularly (455). It inhibits cathepsins B, H and L and can play an important role in desmosome-mediated cell-to-cell adhesion (456). Inhibition of cathepsin B results in activation of glycogen synthase and induction of glycogen accumulation (457). A desmosome subunit gene (DSG3) was also downregulated, leading to further disruption to desmosome formation and loss of cell adhesion. Calcium binding at these subunits normally acts to strengthen connections.

5.3.5 miRNAs

Many miRNAs have been implicated in lung diseases and several have been shown to be produced by one cell type, and transferred to its target cell type via microvesicles and exosomes (458). All significant miRs were downregulated and 22/36 miRs were correlated with macrophages. While most miRs shown in the interaction network were macrophage correlated, none of their targets were. This may indicate macrophages producing miRs and targeting products in/from other cell types after transport. Most target genes were correlated with eosinophils, but proteins from all other measured cell types except macrophages were ‘targeted’ by this rationale.

hsa-miR-146a-5p inhibits the NOTCH1 pathway and drives macrophages towards an M2 phenotype (459). It is a regulator of immune response, has anti-inflammatory function in human airway smooth muscle and polymorphisms in this miR are associated with asthma. In mice miR-146a-5p deficiency is associated with autoimmune deficiency and their macrophages are hypersensitive to LPS.

17 of the significant miRs were previously found downregulated in asthma (460-469). Previous evidence shows that NF-κB can regulate nine of the significant miRs and is regulated by ten of them (470-473).

In some studies in mice it was demonstrated that miR-155-5p can attenuate Th2 response and promote Th1 response (474-476), however there is also evidence in that model organism to suggest it is essential for Th2 response in some circumstances (477;478). In mouse macrophages it is induced by exposure to LPS (479). It is predicted to target CPA3 and WNT5A both of which show significant correlations.

Although it has previously been described as pro-fibrotic in several organs including the lungs (480;481) possibly by involvement in EMT and FMT (fibroblast-to-myofibroblast transition) (482;483), the 5p arm of the ubiquitously expressed miR-21 is downregulated in asthma in our data. However, corresponding with our data, this miR is downregulated in the exosomes from the BALF of asthmatic patients (484). In general it is known to inhibit apoptosis and to promote proliferation through activation of the MAPK pathway (485). In mice it has been shown to regulate the Th1/Th2 balance in immune cells in favour of Th2 response (486). Like miR-155-5p, this miR is also predicted to target WNT5A and shows a significant correlation with it.

Although not two-fold dysregulated one miR is interesting in that it is significantly downregulated in asthmatic non-smokers and significantly upregulated in asthmatic smokers. miR-335-5p has been implicated in the inhibition of TGFß1-induced epithelial–mesenchymal transition in evidence from lung cancer cell lines. While it is unclear what process causes this alternate direction of dysregulation, smoking inhibits TGF-ß1 via a different signalling route providing a compensatory effect in asthmatic smokers (487;488).

30 miR-gene interactions are shown in the interaction network (some validated interactions and others predicted at over 90% confidence), and 19 of these have significant correlations between miR and target across all asthmatic and healthy samples. This list of interactions with significant correlations provides a list of 13 miRs which target a total of 12 genes, whose relationships could be confirmed with antagomirs. miR-151-5p, miR-155-5p, miR-30a-3p, and miR-34a-3p had more than one target in this list, perhaps making these the highest priority for study using antagomirs to replicate their mRNA knockdown effects in isolation. Similarly WNT5A, KIT, MAOA, DNAI2, and ZNF469 were all targeted by more than

one miR. WNT5A (which has been found upregulated in other asthma datasets) and KIT (whose siRNA knockdown in a murine model of asthma led to a decrease in eosinophils and lymphocytes infiltration, and a reduction in Th2 cytokines IL4 and IL5) present as particularly interesting therapeutic targets for future treatments utilising synthetic miRNA mimics, having previous evidence of asthma association and being collectively targeted by three of these four miRs.

5.3.6 IL-4, IL-5, and IL-13

While not sufficient for expression on their own, the classical Th2-cytokines IL-4 and IL-13 - which are involved in asthma and in the recruitment of eosinophils (489;490) – highly upregulate expression of arachidonate 15-lipoxygenase (ALOX15) (491). IL-4 and IL-13 were not significantly upregulated in asthmatics in our data set, however they were both increased as expected in asthma, and modest IL-4 concentrations may be sufficient to induce ALOX15 expression as strong (>300-fold) upregulation has been observed (492). Exposure to IL-4/13 strongly upregulates ALOX15, fibronectin, MAOA, CD1c, FCER2 (Fc Fragment Of IgE Receptor II) and the coagulation factor XIII (F13A1, a transglutaminase) (493) – all of these except FCER2 were significant in our data and only fibronectin was significant with a fold change in the opposite direction expected.

While FCER2 was not significant in our data, another IgE receptor FCER1A was significantly increased. FCER1A is responsible for initiating the allergic response and polymorphisms of its gene are associated with variations in levels of IgE expression and may drive IgE in the context of asthma (494). Although IgE is not measured, increase in expression of one of FCER1A of a receptor for histamine (released downstream of IgE) imply that its inhibition may also be beneficial.

IL-4 and IL-13 have distinct signalling cascades with several shared components. Under energy-deficient conditions AMPK is activated which suppresses IL4-induced ALOX15 expression (495). IL-13 induces p38 MAPK activation that upregulates STAT1 and STAT3 phosphorylation, which in turn activates *ALOX15* expression (496). IL-13 induction of 15-HETE-PE enhances MUC5AC expression in

human airway epithelial cells. Significant high levels of MUC5AC have been observed in asthma previously (497;498) and in our data. MUC5AC is the highest expressed gel-type mucin in the airways and hypersecretion of mucins can generate mucus which is viscous and difficult to clear (499;500). MUC5AC expression by bronchial epithelial cells in vitro can be induced by mechanical stress representative of bronchoconstriction (501). Our data suggest that eosinophils may be involved in the regulation of MUC5AC in the epithelium.

While IL-5 (statistically significant in our data) is coexpressed with IL-4 and IL-13 (all of which part of the same gene cluster) it does not appear to upregulate ALOX15A. IL5 is involved in eosinophil-related airway hyperreactivity (502) via increased eosinophil recruitment, differentiation and survival (503;504). It is associated with asthma, and in particular with severe asthma in which it is already a successful drug target (505;506). IL-5 expression may also be necessary for recruitment of eosinophils by eotaxins (507).

5.3.7 Arachidonic Acid Pathway, ALOX15, and Leukotrienes

ALOX15 (Arachidonate 15-lipoxygenase) encodes an iron-containing fatty acid dioxygenase which has previously been shown to be increased in eosinophils of severe asthmatics (508). It is one of 6 different functional human ‘LOX’ genes which each produce distinct isoforms of an enzyme which converts polyunsaturated fatty acids (PUFA) into their hydroperoxy derivatives. LOX enzymes have a greater affinity for free PUFAs so the reaction often is preceded by the liberation of membrane-bound PUFAs. However, ALOX15 has also been shown to bind to membranes, in a calcium-dependant manner, which increases its oxygenase activity and allows oxygenation of membrane-bound PUFAs (509;510). Action on cholesterol esters has also been observed (511).

ALOX15 is involved in three catalytic activities – as a lipoxygenase (above), a lipohydroperoxidase and a leukotriene synthase. The lipoxygenase function of this enzyme converts arachidonic acid to 15-hydroxyeicosatetraenoic acid (15-

HETE), and to a lesser extent 12-HETE (inserting molecular oxygen to the 12th carbon instead). Another, more preferred substrate is linoleic acid which it converts to 13-HODE. These three major ALOX15 products 15-HETE, 12-HETE and 13-HODE are pro-inflammatory (512). The hydroperoxy lipids resulting from a lipoxygenase reaction may be further converted in the same manner if they still contain bisallylic methylene groups, leading to production of (the anti-inflammatory) lipoxins, protectins, resolvins and maresins (513).

Lipohydroperoxidase activity of ALOX15 further converts hydroperoxy lipids into hepoxilins (hydroxy epoxy eicosanoids) and leukotriene synthase activity further converts hydroperoxy lipids into epoxy leukotrienes. Lipohydroperoxidase activity requires free polyenoic acids as reducing agents and is strongly favoured over lipoxygenase activity under anaerobic conditions, whereas leukotriene synthase activity occurs under aerobic and anaerobic conditions. Lipohydroperoxidase can occur under aerobic conditions with limited concentrations of acid substrate or oxygen or both (514). Eoxins are pro-inflammatory metabolites formed in eosinophils by leukotriene synthase activity (515). Leukotrienes may be involved in recruiting eosinophils to the asthmatic airway following exercise challenge (516).

Research into the various functions of this enzyme in differing environments is complicated by the fact that orthologs in animal models show different affinities, e.g. in mouse ALOX15 has a greater affinity for the 12th residue and ALOX12 is more similar to human ALOX15 in function and regulation. While ALOX15 can accept a variety of polyenoic acids as substrates, monoenoic acids and saturated fatty acids are not oxygenated and act as weak competitive inhibitors. It is constitutively expressed at high levels in reticulocytes, eosinophils and airway epithelium (517), which is replicated in our cell type correlation results for both eosinophils and airway epithelium. It is also expressed at low levels in spermatozoa where it is implicated in infertility via the propagation of oxidative stress cascades (518). miRNA mir-125b-5p is induced by NF-κB and causes the downregulation of ALOX15 (519), and is significantly downregulated in our data which would be consistent with the observed ALOX15 upregulation. In a mouse model it has been shown to reduce inflammation, levels of IL4 and IL13, goblet cell differentiation, and production

of MUC5AC (520), possibly via ALOX15. This goblet cell differentiation may also be due to ALOX15 as 15-HETE-PE is involved in goblet cell differentiation (521). hsa-mir-125b also inhibits the expression of TNF-α, which has been implicated in asthma.

The ALOX15 promoter contains STAT6 binding sites, and the phosphorylation and acetylation of this transcription factor by histone acetyltransferase CREB-binding protein/p300 is required for ALOX15 production (522). Interestingly, despite adenovirus infections having been shown to exacerbate asthma symptoms, adenovirus protein E1A binds to the CREB-binding protein/p300 complex preventing it from phosphorylating STAT6 and p53 (which appears to induce ALOX15 expression via SAT1 (523)) and thereby downregulating ALOX15 (524). E1A binding CREB-binding protein/p300 also inhibits hypoxia response via HIF1A (525).

It has been shown that ALOX15 in eosinophils can drive fibrin formation in individuals with thrombotic disease by providing the required procoagulant phospholipid surface (526;527). Eosinophils from severe asthmatic patients have increased 15-HETE and eoxin C4 compared to mild asthmatic patients or healthy volunteers (528).

Epithelial ALOX15 expression is correlated with asthma severity (529). 15-HETE has been shown to induce endothelial barrier permeability by the phosphorylation of zonula occludens 1 and zonula occludens 2 proteins, causing their dissociation from tight junctions (530). ALOX15 inhibition has been shown to prevent this effect in tight junctions of arterial endothelium (531). In airway epithelium increased barrier permeability results in increased penetration of allergens into the intercellular space allowing direct disruption of apical junction complexes (complexes of tight junctions and adherens junctions), and the indirect disruption via IL4 and IL13, which forms a positive feedback loop (532). Tight junctions can also act as scaffolding platforms for cell signalling and docking stations for transport vesicles (533;534). Viral infection of respiratory epithelial cells has also been shown to disrupt AJCs and increase permeability. This epithelial barrier disruption involves actin cytoskeletal remodelling, possibly

dependent on cortactin activation (535). In our data Cortactin Binding Protein 2 (CTTNBP2) is significantly upregulated in asthma.

WNT5A initiates an age-related cascade in asthma which contributes to leukotriene formation and is steroid-resistant (536). WNT5A has previously been reported upregulated in asthmatic samples (537;538). WNT5A can drive or inhibit beta-catenin signalling, depending on the receptor context (539), and can drive fibroblast ECM deposition, proliferation, and resistance to apoptosis (540). It also facilitates ß-catenin/p300 interaction in rat (541).

In addition to fibroblasts, myofibroblasts are also involved in airway remodelling. These can differentiate from fibroblasts by FMT. This process is stimulated by IL-4 and IL-13 via multiple pathways, by CCL26 (eotaxin 3), and by cysteinyl leukotrienes (LTC_4, LTD_4, and LTE_4) produced via the arachidonic acid pathway (542). Cysteinyl leukotrienes also regulate collagen production by bronchial fibroblasts (543). GGT5 cleaves glutathione and converts leukotriene C4 into leukotriene D4. It was found correlated with eosinophils in non-smoking asthmatics and neutrophils in smoking asthmatics.

ALDH3B1 is an enzyme which oxidises unsaturated medium- and long-chain aldehydes such as hexadecanal, which is an intermediate in glycosphingolipid metabolism (544). It is protective against oxidative stress (545). Transcription is driven by transcription factor NRF2 which itself is expressed in response to oxidative stress (546) and induces NQO1 which was found dysregulated in smokers.

5.3.8 Metabolism

An increase in glycolysis has been reported in asthma previously, and it appears to drive T-cell activation, airway inflammation and hyperreactivity, and production of several molecules involved in asthma including IL-5 (547). Pyruvate Kinase M1/2 (PKM) is a glycolytic enzyme involved in production of ATP from ADP and is upregulated in asthmatics in our data. ALDH1A1 and ALDH3B1

are also involved in glycolysis – causing the oxidation of acetaldehyde to acetate – and are both upregulated. Transaldoase 1 (TALDO1) is an enzyme of the non-oxidative phase of the pentose phosphate pathway which is involved in NADPH generation, oxidative stress, inflammation, and it is closely related to glycolysis, providing for it the substrates fructose-6-phosphate and glyceraldehyde-3-phosphate (548;549). TALDO1 is upregulated and epithelial-correlated in our data.

On the other hand many genes involved in glycolysis were also downregulated. Lactate dehydrogenase A (LDHA) catalyzes the conversion of L-lactate and NAD to pyruvate and NADH (and the reverse reaction) in the final step of anaerobic glycolysis. It is upregulated in response to viral infection, and downregulated with steroid treatment in severe asthma (550;551). In this data it is downregulated with asthma amongst non-smokers but not smokers, possibly due to a downregulated gene (TXN) which is predicted to upregulate LDHA via HIF-1α activity (552) – expression of TXN and LDHA is >75% correlated. Glucose-6-Phosphate Isomerase (GPI) which converts between glucose-6-phosphate and fructose-6-phosphate is downregulated. Enolase 2 (ENO2), which promotes cell survival in a Ca^{2+}-dependant manner and is part of glycolysis is also downregulated, possibly via reduced NF-κB signalling.

In addition to these energy metabolism enzymes, alpha-2-glycoprotein 1, zinc-binding (AZGP1) which is involved in glucose transport was also downregulated. As mentioned above AHSG also inhibits several processes of energy metabolism through inhibition of insulin receptor and insulin receptor substrate 1.

While metabolomics data did not yield any significant results after multi-test adjustment there were large increases of toluene-4-sulfonate and quinate in asthmatic patients, which were significant before multi-test adjustment. Quinate (>13-fold increase in asthmatics) is the conjugate base of quinic acid, which is normally found in plant sources but is used in the production of pharmaceuticals. Toluene-4-sulfonate (>3-fold increase in asthmatics) is the conjugate base of toluene-4-sulfonic acid which is a strong organic acid, likely to act as a respiratory irritant.

Exposure to another highly reactive toluene compound, toluene diisocyanate, has many harmful effects on the body including in the lungs where it can induce asthma (553). In animal models it has shown to increase epithelial shedding (554) possibly via E-cadherin distribution (555). In addition to toluene-4-sulphonate, two other toluene compounds had a greater-than-twofold change in asthma contrasts - 2,4-diamino-6-nitrotoluene (asthma amongst smokers, log_2FC: 1.23); 4-toluenesulfonamide (asthma amongst non-smokers, log_2FC: 1.41).

Assuming these identifications are accurate they may be relevant to asthma as indicators of lung irritation. Although normally studied in the context of bicarbonate levels resulting from respiratory acidosis or metabolic acidosis, acid-base disturbances in the lung are detectable in asthma and are associated with reduction in FEV1 (556).

5.3.9 Chemokines

C-C Motif Chemokine Ligand 17 (CCL17) is a T-lymphocyte chemoattractant and has been shown to attract Th2 cells and induce airway inflammation in a humanised mouse model (557). In a mouse model it is required for NOD1-mediated asthma exacerbation (558). It is upregulated by lung macrophages in asthmatics (559) and in our data it is significantly enriched in asthmatics amongst non-smokers and consistently eosinophil-correlated. CCL17 expression is corticosteroid-resistant but is suppressed by PI3Kinase enzyme inhibitors.

CCL13, CCL17 and CCL26 are all increased in eosinophilic asthma and COPD (560). CCL13 is highly expressed in numerous cell types during inflammation, including eosinophils and lymphocytes with which it is correlated in our data (561). It has also been detected in (and was found correlated with) airway epithelial cells where it is associated with pulmonary fibrosis (562), and in airway smooth muscle where it is upregulated in asthmatics and where it is downregulated by vitamin D (563). CCR3 is a receptor for CCL13 which is expressed in airway smooth muscle and induces intracellular calcium mobilization and ASMC migration (564). CCL13 and CCL17, along with IgE, are

effectively reduced in severe asthma via an IL13 antibody called lebrikzumab (565).

CCL26 is involved in eosinophil chemotaxis and is overexpressed even more in severe asthma than mild/moderate asthma. Patients with increased CCL26 may also respond to this lebrikzumab treatment, being that CCL26 is upregulated by IL13, expressed in airway epithelial cells and is upregulated in asthma and so may also respond to this treatment (566). In our data CCL26 is correlated with eosinophils and lymphocytes, both of which are recruited by it (567;568), and the latter is involved in its expression (569). It also acts to repel macrophages via an antagonistic interaction with an alternative receptor CCR2 (570). CCL26 is also associated with persistent (>48 hours) eosinophil-driven inflammation (571) and has been reported to induce fibroblast migration (572). CCL26 (AKA eotaxin 3) is more effective than eotaxin 1 and 2, and unlike them it induces an additional phase of cell migration which is resistant to blockade (573).

Clusterin (CLU), which normally modulates airway inflammation by attenuation of CCL20 was downregulated in our data, suggesting a possible increase in CCL20 activity also.

5.3.10 Minor Discussion Points

5.3.10.1 The Complement System

Complement C3 (C3) was upregulated at the protein level and eosinophil-correlated. C3 is the central component of the complement system, as its activation products are involved as mediators in most of the systems immune regulatory effects (574). IL4 and IL13 stimulate its production in an airway epithelial cell line (575). C3 acts as a danger sensor – DAMPs cause a nucleophillic attack of its thioester. C3 cleavage is initiated by lectin binding mannose on a pathogen surface or by C1 complex activation through binding of C1 complex to antigen-antibody complexes to apoptotic cells or to C reactive protein. Eosinophil granule proteins have been shown to strongly inhibit C3

convertase function by interfering with C1 (576). The initial cleavage products of C3 are C3a and C3b. In addition to its numerous functions across various cell types it also is responsible for the proliferation and differentiation of leukocytes. Cathepsin G, which was downregulated, cleaves C3, and may be involved in connective tissue remodelling at sites of inflammation. It may also exhibit bacteriocidal properties.

C3a is an anaphylatoxin, a complement molecule that is proinflammatory via several pathways. It is upregulated in induced sputum and BAL of asthmatics, particularly during exacerbations (577;578). It is involved in neutrophil, eosinophil and mast cell chemotaxis and induces histamine release from mast cells and basophilic leukocytes. C3a can stimulate the increase of free intracellular Ca^{2+} by the influx of extracellular Ca^{2+} via a G-protein mediated process (579). It is also responsible for the activation of respiratory burst in neutrophils (580) and can stimulate MUC5AC production in airway epithelial cells in vitro (581). Interestingly it can trigger smooth muscle contraction. C3b can also activate respiratory burst (582) and assists in phagocytosis as an opsonin. C3b can form subunits for the alternate C3 convertase (causing positive feedback) or for C5 convertase. Interestingly it has also been shown that C5 can regulate airway hyperreactivity and pulmonary eosinophilia by downregulating expression of the C3a receptor (583). In addition to its numerous functions across various cell types C3 also is responsible for the proliferation and differentiation of leukocytes.

C4 protein, in contrast, was found heavily downregulated via one of its two highly similar genes, *C4A*, and not associated with any particular cell type. As this was a downregulation at the peptide level it may be due to reduction in NF-κB transcription regulation or enhanced cleavage. Its cleavage product C4b is a subunit of one of the C3 convertase enzymes which forms C3a and C3b; C3b is a subunit of the other C3 convertase. C4a is also an anaphylatoxin and C4b can operate as an opsonin. C4a upregulation has been detected in asthma, but only in blood samples (584;585).

Karp suggests that the balance of C3a to C5a is important in the pathogenesis of asthma – that a C3a increase and C5a decrease potentially caused by allergens

such as smoke and by viral infections initiates the Th2 environment described in asthmatic lungs (586). Further, that the myeloid dendritic cells present under these conditions which produce CCL17 contribute to the development of this Th2 environment.

5.3.10.2 The PI3K/AKT Pathway

Phosphoinositide 3-kinase (PI3K) is implicated in asthma pathogenesis in several ways and its inhibition has been shown to reduce mucous secretion, mast cell degranulation, and immune cell recruitment, and to cause bronchodilation, making this pathway an attractive target – however it also has protective effects on the airway (587).

In the airway smooth muscle PI3K affects contraction, accumulation of contractile proteins, airway tone, chemokine/cytokine secretion, proliferation, migration, ß-2 adrenergic receptor resensitization, and levels of CD38 – a calcium signalling protein involved in AHR (588-592). In the epithelium PI3K is involved in: responses to environmental stimuli such as viruses (593;594), inflammation via iNOS and VEGF (595;596), and IL-13 driven mucous hypersecretion (597). There is a high level of expression of PI3K in eosinophils, induced in part by Th2 cytokines IL4 and IL5 (598). PI3K is also responsible for neutrophilic and eosinophilic degranulation and cell migration (599;600).

On the other hand PI3K negatively regulates both IgE – normally secreted by B cells – and its receptor – localised on the surface of mast cells – reducing mast cell degranulation (601) which normally results in the release of mediators of bronchoconstriction such as histamine and prostaglandin D2. It can also play an anti-inflammatory role, inhibiting pro-inflammatory cytokines and increasing macrophage production of anti-inflammatory IL10 (602).

Somewhat surprisingly (as with NF-κB) in this cohort the direction of dysregulation of several molecules would indicate a reduced PI3K signalling - CALM, which is downregulated promotes PI3K activation; AHSG which is upregulated inhibits PI3K; GNB2 which is part of the pathway is downregulated; downregulated S100P activates the PI3K pathway; knockdown of EZR which is

downregulated can result in inhibition of AKT; dysregulation of KIT, TXN and PIP, below. A downregulation of the PI3K may itself contribute to downregulation of the NF-κB pathway.

KIT Proto-Oncogene, Receptor Tyrosine Kinase (KIT) acts as a cell surface receptor for cytokine KITLG/SCF, and causes phosphorylation of many targets including PIK3R1 (the regulatory subunit of phosphatidylinositol 3-kinase, PI3K). KIT is correlated with eosinophils and epithelials in our data, and it is upregulated with asthma, suggesting a deactivation of PI3K. Interestingly its target phosphatidylinositol is a 'self' target of CD1B and it is also bound by SCGB1A1 which is a protein involved in phospholipase A2 inhibition and whose downregulation in induced sputum and bronchial brushings is linked with asthma (as in our data) (603-605). KIT also activates transcription factors STAT1, STAT3, STAT5A and STAT5B and activates the MAPK pathway. In mice it has been shown to be essential for alveolar maintenance and resistance to an emphysema-like disease (606). Knockdown of KIT in a murine model by siRNA reduces infiltration of eosinophils and lymphocytes to the lung tissue and BALF and reduces levels of IL-4 and IL-5 (found upregulated in this dataset) (607).

Thioredoxin (TXN) is a redox-active protein which is dysregulated in childhood asthma (608), and in animal models it inhibits eosinophilic inflammation and the production of chemokines and Th2 cytokines, thus a decrease as in our data would lead to an increase in airway hyperresponsiveness (609). Inhibition of TXN in chickens results in the inhibition of the P13K pathway. Prolactin Induced Protein (PIP) is involved in AKT signalling (and actin binding) and has previously been found downregulated in chronic rhinitis, the same direction of change as with asthma in our data (610).

5.3.10.3 S100 Proteins

The S100 proteins are calcium binding proteins of a broad range of intracellular and extracellular functions and all three of the significant S100 proteins in this dataset were downregulated. S100P is a calcium-binding protein which forms

dimers with itself and with other S100 proteins. These dimers interact with the extracellular region of the AGER (aka RAGE) cell surface receptor. This S100P-RAGE interaction activates NF-κB, MAPK, JNK, and PI3K/AKT pathways (611). miRNA-155 and miRNA-21 are regulated by S100P-RAGE signalling through these pathways (612;613). Interestingly an asthma medication, cromolyn sodium, binds S100P which blocks the S100P-RAGE interaction (614;615), which in this cohort may contribute to dysfunction.

S100P dimers also bind and inactivate p53 protein (616) which triggers apoptosis. Although it inactivates p53 it also prevents its degradation by binding HDM2. S100P is increased in induced sputum of asthmatics after a high-fat meal (617).

S100A6 has been detected in many cell types including lymphocytes, but most abundantly in epithelial cells and fibroblasts (618-620). It drives proliferation in pulmonary fibroblasts and affects cell morphology and cytoskeletal organisation (621). Its modulation of the cytoskeleton may occur through its binding structural proteins tropomyosin, various annexins, caldesmon and calponin (622-625). It is involved in cell adhesion in various studies, with contrasting effects depending on cell type (626). In healthy rather than cancerous cells it seems to promote cell adhesion, particularly in conjunction with fibronectin and fibrinogen which are both also reduced in our data. This downregulation and loss of cell adhesion may occur as a result of decreased NF-κB signalling for which S100A6 is a target.

S100A6 activates the JNK pathway, driving apoptosis (627), and it also seems to be involved in proliferation via MAPK pathway and others (628;629). This dual functionality may represent alternative functions in different cellular types and conditions or it may regulate proliferation and apoptosis simultaneously – both regulating apoptosis in target cells and driving proliferation in nearby cell of a different type or state (630). It has also however been shown to inhibit both processes under certain circumstances (631;632). Like S100P it also binds intracellularly to p53, inactivating but preventing degradation (633).

S100A6 is correlated with eosinophils in our asthmatic non-smokers and healthy non-smokers, but no significant correlation is present regarding the asthmatic

smokers and healthy smokers. NF-κB acts as a transcription factor for S100A6 and p53 inhibits transcription (634;635). It is upregulated under certain stress conditions, including oxidative stress (636) and under mechanical stress in fibroblasts (637). Mechanical stress can occur in the airway during bronchoconstriction, which without additional inflammation is capable of inducing airway remodelling (638). The compressed bronchial epithelium has been shown to produce collagens I and III (639) and to release exosomes (640). Other mechanical forces present might include stress from hypertrophy, hyperplasia or the periodic stresses from cell jamming (641).

S100A6 is downregulated by asthma medication amlexanox (642). It binds with high affinity to GADPH (643). In Rat it has been shown to inhibit histamine and actin secretion in mast cells (644) and in Chicken it binds lysozyme which is also significant in our data (645). Lysozyme is a bacteriocidal enzyme which is expressed highly in mucosal secretions and neutrophil granules, and its bacteriocidal activity is inhibited by high levels of calcium (646). While lysozyme is effective at improving airway function in COPD it is not effective in asthma (647). It is involved in cell spreading and cell anchorage in fibroblasts (647) and promotes EMT under signalling by Sonic hedgehog-Gli1 (648).

S100A9 stimulates NF-κB signalling (649) and MAPK signalling, via RAGE, recruiting fibroblasts to the lung (650), and has been implicated in neutrophillic inflammation in asthmatic patients (651). However, in asthmatic patients without increased counts of inflammatory cells S100A9 has been found to be downregulated (652). It also activates natural killer cells via RAGE (653) and induces MUC5AC production in airway epithelials (654). It forms a heterodimer with S100A8 which may be involved in neutrophil chemotaxis and adhesion to fibrinogen (655). This dimer binds and transports arachidonic acid, linoleic acid and others. When this heterodimer is phosphorylated it appears to alter its ion-binding specificity, thus it accepts Cu^{2+} or Zn^{2+} which blocks binding with arachidonic acid and may induce alternative pathways. Equimolar tetramers can also form bound to Ca^{2+} ions – essential for formation of microtubules (656) – or bound to Zn^{2+} ions (657).

S100A9 activates NADPH oxidase which produces inflammatory superoxide ions, however paradoxically it may also be involved in a protective anti-inflammatory effect via oxidant scavenging, possibly as part of a negative feedback loop to control excessive inflammation (658;659). This interaction also forms a link with energy metabolism, particularly in activated neutrophils where NADPH oxidase activity is increased and accounts for some of the increased glycolysis rate in response to inflammation (660;661).

As a damage-associated molecular pattern (DAMP), S100A9 is one of a group of diverse molecules which respond to non-infectious damage and pro-inflammatory stressors, signalling IFNs through pattern recognition receptors like TLR4. The inflammasome reaction generated by DAMPs induces IL-18 expression, which in turn leads to the activation of NF-κB (662-664).

5.3.10.4 Dendritic Cells

CD1a has previously been shown to be upregulated in macrophages in BAL from asthmatic patients, and CD1a+ dendritic cells are increased in asthmatic bronchial epithelium. CD1C is a cell surface marker which is involved in antigen presentation and is usually expressed in dendritic cells. Airway dendritic cells express more of CD1C when presented with an allergen challenge (665) and it is upregulated in asthmatic patients (666). CD1A, CD1B and CD1C are all increased in asthmatics and correlated with eosinophils. It is possible that some of the eosinophil-correlated results relate to dendritic cell activity as the counts of activated eosinophils and dendritic cells in induced sputum are correlated (667). However eosinophils can also act as antigen-presenting cells (668). Each CD1 protein complexes with a large set of specific self and non-self targets (669). Some self targets appear to be involved in blocking the large hydrophobic pockets of CD1 structures during trafficking to endosomes and then to the cell membrane.

Another dendritic cell protein involved in antigen presentation and which is normally targeted by NF-κB, B2M, was downregulated in our data. It is known to

form heterodimers with CD1A, CD1B, and CD1C, and its absence has been shown to cause an immunodeficiency related to the major histocompatibility complex, and various respiratory conditions (670).

CD207 (langerin) is expressed in langerin+ dendritic cells (Langerhans cells) which are responsible for LPS-induced reactivation of allergen-specific Th2 responses in postasthmatic mice (671). These cells are responsible for another respiratory condition called pulmonary Langerhans cell histiocytosis which also causes inflammation and fibrosis.

5.3.10.5 Hormonal Regulation

IGFBP7, which binds IGF-I and IGF-II, was significantly downregulated in our data, though the relevance to asthma may be its promotion of cell adhesion. While growth factors such as IGF-I or EGF cannot directly induce androgen receptor activity, they can increase the effect of androgens, yielding large responses in the presence of low levels of androgens (672). SMR3B (Submaxillary Gland Androgen Regulated Protein 3B), an androgren-related protein was downregulated in asthma. SCGB2A1 (Secretoglobin Family 2A Member 1) may bind androgens and other steroids, and appears to undergo alternative splicing/modification, with one significant peptide upregulated and another downregulated. A SNP in another gene of this family, SCGB3A1, is implicated in asthma, which is consistent with the downregulation in our data. AKR1C2 (Aldo-Keto Reductase Family 1 Member C2) catalyses the inactivation of androgens, was upregulated in our data and correlated with epithelial cells in asthmatics amongst smokers.

MAO-A (Monoamine Oxidase A) degrades norepinephrine, serotonin and dopamine, which are all correlated/involved with asthma and/or bronchodilation (673-676), and it produces hydrogen peroxide. A significant gene involved in reducing hydrogen peroxide levels, PRDX5, is downregulated potentially leading to even greater levels. Hydrogen peroxide is increased in the exhaled air of asthmatic patients (677) is inflammatory and in epithelials (one of

the cell types it was correlated with) it drives EMT. MAO-A also appears to be involved in apoptosis (678;679).

Glutathione S-Transferase Alpha 3 (GSTA3), which is involved in testosterone and progesterone synthesis, was downregulated significantly in asthmatics amongst non-smokers. Lipocalin 1 (LCN1) was upregulated in asthma and another lipocalin (LCN2) was downregulated, possibly as a result of reduced NF-κB signalling. Both are involved in the extracellular transport of small hydrophobic metabolites. LCN2 has been shown in a murine model of allergic airway disease to prevent airway inflammation (680).

5.3.10.6 Other Interesting Results

While MUC5AC was upregulated two other mucins were downregulated, however a reduction in these particular mucins may not be advantageous to asthma pathology as initially expected. MUC7 is a small mucin which facilitates bacterial clearance, and which an allele of was associated with asthma (681). This adds to the effect of reduced mucociliary clearance from downregulated CALM. MUC1 is a mucin associated with anti-inflammatory effects in the lungs (682). Trefoil Factor 3 (TFF3) is a trefoil – a family of mucin-associated genes, which have mostly been studied in the context of the intestinal epithelium. Trefoils are involved in various signalling pathways (683) and affect physical properties of mucous which they partly comprise, possibly by linking mucins. TFF3 is expressed in goblet cells in various epithelia of the body including the airway epithelium (684) and its presence in mucous appears to increase viscosity (685). It is expressed more in neutrophilic asthma and paucigranulocytic asthma than eosinophilic asthma (686), and it has also been upregulated in severe asthma, as in our data (687). Prominin 1 (PROM1), which is involved in mucocilliary differentiation in the epithelium, was upregulated, though its precise effects are unknown (688).

Proline Rich 4 (PRR4) is a gene of unknown function which is secreted and appears to have a protective effect. Downregulation of this gene has been

implicated in other pathologies (689;690) as in our data, however its upregulation has previously been detected in asthma (691;692). SPRR3 is a cornified envelope precursor protein, mutations in which have previously been implicated in asthma (693).

SERPINF1 is known to normally be involved in anti-angiogenesis, so we might expect it to be downregulated in asthma, however it is upregulated in our data. It has previously been upregulated in three studies of emphysema (694). SERPINC1 and SERPING1 were also upregulated in asthma. SERPING1 inhibits the complement system via components of the C1 complex, and also inhibits factor XIIa, chymotripsin and kallikrein. SERPINC1 is also an inhibitor of thrombin, along with blood clotting factors IX, X and XI. Thrombin is often implicated in tissue remodelling, including in asthmatic airways (695). SERPINB1 is an NF-κB target and is downregulated in this dataset, and through its protease targets normally acts to prevent damage from neutrophil-associated inflammation. SERPINA1 – also a target of NF-κB - is an inhibitor of elastase, plasmin, thrombin, trypsin, chymotrypsin, thrombin, and plasminogen activator, and has proteolytic activity against insulin. It has previously been found upregulated in asthmatic sputum, contrary to our data, however this is perhaps not surprising given its multiple targets. Another serine protease which is downregulated, KLK11, targets bz-Phe-Arg-4-methylcoumaryl-7-amide and other kallikrein and trypsin substrates. Previously an exercise-induced upregulation has been reported (696).

Histamine is involved in both early and late phase allergic response and is released by basophils and mast cells. Levels of histamine in BALF correlate with bronchial hyperresponsiveness (697). Histamine receptor H4 (HRH4) reduces LPS-induced TNF production in vivo. Polymorphisms in this gene are associated with asthma - particularly infection-induced asthma (698). HDC converts histidine to histamine and was strongly downregulated and correlated with eosinophils in asthmatics amongst non-smokers.

IL-18 is increased in induced sputum of severe asthmatic patients (699) and IL-18R1 has also been found upregulated in severe asthmatics (700) and asthmatics with eosinophilia (701). IL-18R1 is increased in asthmatic patients in our data and correlated with eosinophils and neutrophils. IL-18 has had contradictory

reports on its function in neutrophils – e.g. one study reporting an anti-apoptotic effect and expression of almost exclusively the beta receptor (702), another reporting enhanced adhesion molecule expression, respiratory burst, and chemokine production – but no apoptotic effect – and constitutive expression of both the alpha and beta receptors (703). Two SNPs in IL-18R1 encode variants with a higher rate of transcription, which are associated with asthma (704). IL-18 can activate NK cells and co-stimulate production of IFNγ via immunoglobulin receptors (705).

CST1 (Cystatin SN/Cystatin 1) and CPA3 (Carboxypeptidase A3) were both upregulated and eosinophil-correlated, and have previously been found dysregulated with exercise-induced bronchoconstriction in asthmatics, in an airway environment of high levels of leukotrienes and increased epithelial shedding (706). Cystatins inhibit the proteolytic activity of endogenous and exogenous cysteine proteases, which normally act to degrade proteins and modify proteins/proproteins. They are part of a complicated system tightly regulating proteolytic burst under particular circumstances, and the disturbance of this system is implicated in several diseases. CST1 is extracellular and inhibits cathepsin C. Cathepsins can be secreted to degrade the extracellular matrix. CST1 also binds and inhibits cystatin 3 (CST3), which was downregulated in our dataset, interfering with its inhibition of cathepsin B (707). In fact the apparent downregulation of CST3 at the protein level may be due to this dimerisation, driven by the increase in *CST1* transcription. CST3 can form inactive homodimers and amyloid fibrils (708), and has been shown to inhibit the replication of herpes simplex virus (709). Like CST1 it is a type II cystatin and is extracellular, however it can be re-internalised to the intracellular compartment - in particular to lysosomes (710).

CPA3 is a protease of exogenous and endogenous targets and is a marker for a connective tissue subtype of mast cells which are not commonly found in the lung but are associated with severe asthma and COPD. Presumably this is the product of mast cells in our data also, despite our correlation with eosinophils – there is a correlation between mast cells expressing CPA3 and eosinophil recruitment (711). Indeed CPA3 and TPSAB1 (also known to be secreted by mast cells) have been suggested as biomarkers of eosinophilic asthma (712;713).They

are expressed more in peripheral airways in severe asthma and in normal patients (714). In a study using them as markers of mast cell subtypes indicative of asthma (715), 7/14 genes differentially expressed across mast cell subtypes were found also in our results (*CD1A, CD1B, CD1C, CLEC4F, DNASE1L3, FCER1A, HDC*). CLEC4F is a receptor which is correlated with epithelial cells and eosinophils and upregulated in our data. It was until recently thought to be solely expressed in the liver, where it responds to galactose, fucose and glucosphingolipids (716). In a respiratory and particularly an asthmatic context it may be a receptor which responds to environmental stimuli (perhaps mostly bacterial in origin), leading to secretions from mast cells involved in eosinophil recruitment. Genetic variation in CLEC4F has been linked with susceptibility to *Pseudomonas aeruginosa* lung infection in mice (717).

DNASE1L3 is a calcium-dependant enzyme which hydrolyses DNA. Our results correspond with the increase in asthmatic sputum reported for this gene previously and the correlation with eosinophils corresponds to it being a marker of eosinophilic asthma (718). DNASE1L3 and CPA3 also discriminate between steroid responders and non-responders (719).

BPIFA2 plays role in antibacterial response in the upper respiratory airways by binding LPS and inhibiting bacterial growth. The upregulated GOLM1 has an anti-viral response. The class II major histocompatibility complex gene HLA-DQB2 which is involved in antigen presentation was downregulated. A component of class I histocompatibility complex was also downregulated. BPIFB1 binds and modulates responses to LPS, and is an autoantigen. Similarly to TFF3, it is upregulated in asthma, it is eosinophil-correlated and epithelial-correlated, and it is expressed more in neutrophilic asthma and paucigranulocytic asthma (720). BPIFB2 is a member of the lipid transfer/lipopolysaccharide binding protein gene family, and was downregulated in asthma and correlated with eosinophils.

DMBT1 (Deleted In Malignant Brain Tumors 1) is a calcium-dependent enzyme which may be involved in anti-bacterial defence and epithelial differentiation. Lacritin (LACRT) is a Ca^{2+} binding protein normally secreted in tears and saliva and is involved in epithelial proliferation, survival, and wound healing. YWHAB

was strongly downregulated in asthma amongst non-smokers, and a SNP in this gene has been previously reported as associated with asthma (721).

Five out of eleven significantly dysregulated immunoglobulin peptides (IGG1, IGHG1, IGLV1-51, IGLV1-44, IGHV3-23) were upregulated and correlated with eosinophils. Of the remaining immunoglobulin peptides one was upregulated and the remaining were downregulated, and none were correlated with cell types.

A peptide derived from *AGT*, whose initial protein product (angiotensinogen) is converted into the active protein angiotensin II, is upregulated in asthma. While CPA3 (which can cause the first step of this conversion into angiotensin I) is upregulated, CTSG which can cause the second step is downregulated – however other enzymes can cause this conversion. Angiotensin II is mostly studied in respect to cardiovascular phenotypes (as in chapter 4) however it also is known to cause bronchoconstriction. Haemoglobin subunits HBB and HBD were found upregulated with asthma in our dataset, and HBD is correlated with eosinophil. HBB has previously been found upregulated in asthmatic lung biopsies (682).

ENSG00000258752, FGFBP1, KRT10, PRH1/PRH2 and TMEM151B are all correlated with epithelials and eosinophils, and they are all upregulated in asthma in our data. ENSG00000258752 is an anti-sense RNA gene which has been previously detected in BALF, where its downregulation coincided with severity of pneumonia infection (722). It inhibits translation of the transcription factor FOXN3, which otherwise acts to inhibit cell proliferation (723). It has also been shown that FOXN3 binds beta-catenin, blocking its interaction with at least one other protein (724). Little is known about the transmembrane protein TMEM151B, but in this context, and given its correlation with epithelials and eosinophils it might be logical to assume that it is a receptor on the surface of the epithelium, affected by environmental stimuli and whose downstream effects include eosinophil recruitment or function. FGFBP1 codes for a secreted fibroblast growth factor carrier protein which binds FGFs, particularly FGF2, facilitating its release from ECM storage enhancing their mitogenic activity and aiding in epithelial repair. It has previously been detected upregulated in severe asthma (725) and its target FGF2 has been implicated in atopic asthma (726;727). KRT10 encodes a type I cytokeratin, which form part of the

cytoskeleton (the intermediate filament) in epithelial cells. It has previously been shown upregulated in asthmatic sputum (728). Interestingly it interacts with AKT1 – a protein whose phosphorylation by activated PI3K can cause many downstream effects including mTOR activation – apparently preventing its translocation and therefore subsequent activation. A type II cytokeratin, KRT4, is upregulated in our data and correlated with epithelial cells. The acidic KRT10 and basic KRT4 may interact with each other in epithelial cells to form the acidic-basic keratin heterodimers which chain to form keratin filaments. KRT4, like KRT10, is epithelial-correlated in our data.

Interestingly a number of genes which have previously been identified as biomarkers of response to exercise challenge in asthmatics with exercise-induced bronchoconstriction were recorded as dysregulated without exercise challenge in our cohort of severe asthmatics – *CST1, CPA3, CD207, MUC5AC, TFF3, TPSAB1,* and *TSPAN8* (729). All were upregulated in both cohorts and eosinophil-correlated in our data (*MUC5AC, TFF3,* and *TSPAN8* were all also epithelial-correlated).

5.4 Conclusions

While measuring cell types as percentages of a number of cells counted rather than as a number of cells per unit volume could be somewhat problematic – due to large variation in one cell type obscuring variations in other cell types – cell type percentages still seem beneficial to downstream analysis. In some cases cell type correlations indicate which cell type might be involved in regulation or production of the molecule, and in the case of miRs they seem to indicate a large percentage produced by macrophages and endocytosed to interact with gene products of other cell types. Something similar might also be helpful with other relevant, less abundant, cell types such as mast cells, dendritic cells and goblet cells.

Many of the molecules identified verify previous work in other cohorts, and many others seem to be totally novel, and in need of further research. Smoking and

COPD contrasts for induced sputum show significant widespread dysregulation across thousands of genes, however less than a hundred in each have a two-fold difference. Of these two-fold differentially expressed molecules there were three in common between asthma and COPD (F13A1, MMP10, IL18R1), and four in common between asthma and smoking (FCER1A, CD1A, CD1B, HLA-DQB2). Results demonstrated how important smoking status is when investigating the asthmatic airway. While the contrast of all asthmatic patients vs all healthy patients gave almost twice as big a sample size the results for non-smokers alone had more significant results. And not only were some results significant in one set and not another, some molecules were differently regulated depending on smoking status , e.g. lysozyme which has bacteriostatic and bacteriocidal properties. One gene which works in concert with lysozyme for its bacteriocidal effect, lactoferrin has many relevant functions for the asthmatic airway, however it is difficult to discern the effect of its differential regulation as one peptide is significantly increased and the other significantly decreased with regards asthma. A fourth asthma contrast may even have been helpful, comparing smoking asthmatic patients to healthy non-smokers.

The interaction networks show the high connectivity of the pathway(s) underlying asthma, which highlights the complexity of the system. While these interaction networks were informative the available databases only contain a fraction of the information available in the wider literature and only indicated direct relations without ‘gaps’ where data may be missing.

Large disruptions exist in relation to calcium signalling (particularly calmodulin), the extracellular matrix (with a focus on fibronectin), and actin – each spanning a large number of dysregulated mRNAs, peptides and miRs. These dysregulations and others appear to underlie a system wherein bacterial killing, the epithelium has reduced barrier function, mucous is upregulated and it (along with allergens and pathogens therein) is poorly cleared from the respiratory tract. Surprisingly there are several indicators of NF-κB being downregulated despite its known involvement in inflammation and previous implication in asthma through studies with murine models – and similarly with PI3K. This potential reduction in NF-κB activity may be due to the reduced levels of calmodulin or other factors.

Several drugs are being developed and tested towards targeting well-studied molecules with respect to asthma, including IgE and interleukins (730-732), and calcium signalling (733). The poor effect of calcium channel blockers in human trials may be reflected by the reduced CALM expression seen in the induced sputum of this cohort. Mepolizumab and Reslizumab should have efficacy for this cohort, as it exhibits high IL-5 expression. There is also some indirect and weak evidence that mAbs for IgE and IL-13 could have beneficial effects. The knockdown of upregulated genes by siRNAs or synthetic miR mimics may be interesting future therapeutic targets, particularly those which have multiple significant targets backed up by significant correlations.

6. General Discussion

While in hypothesis-based research we ask ourselves “why do this”, in hypothesis-free we should instead ask “why not do this?”. Diseases are not limited to an individual class of biomolecule so ideally we should look at multiple omics technologies simultaneously. While a single-gene disorder might not require such an extensive approach, many of the diseases which have the highest mortality and biggest effect on quality-of-life are complex diseases which are still poorly understood at the molecular level (including those respiratory and cardiovascular diseases discussed).

DNA methylation microarrays are relatively inexpensive ways to assess differential methylation at a large number of sites across the genome, however they are prone to false positives where a CpG site coincides with a SNP. By combining these data with SNP data, this drawback is eliminated. The shortcoming with this approach is when there is incomplete coverage of SNP data, as with the dataset in chapter two - where SNP data was also assessed by microarray rather than by sequencing. In this way data from one omics experiment supplemented the analysis of another. Another method to overcome this issue was to summarise methylation values of CpGs by regions of the genes in which they are annotated - rather than grouping them across the entire gene since DNA methylation in different gene regions appears to correlate to different effects, be it causative or not. This way they are again less prone to the interference of a single SNP and there is less of a statistical cost in multi-test adjustment. This was helpful with the dataset regarding DNA methylation in essential hypertension as sample size was extremely small for an omics experiment. While the sample size was small the samples originated from a much larger cohort and their selection made use of extremes of phenotype and genotype.

The results of this experiment implicated *NADK, CHID1,* and *APTX* as being differentially methylated in essential hypertension. Nicotinamide adenine dinucleotide phosphate is a coenzyme which, in its reduced form (NADPH), acts as a reducing agent to restore all known innate defence systems against

oxidative stress. In doing so it is itself oxidised and can be returned to its reduced form by NADP-dependant dehydrogenases – by this action these enzymes form the main protection against oxidative stress. However, since NADP mainly exists in its reduced state in human cells, by increasing the overall pool of NADP the amount of reduced NADP is increased - and so NADK is responsible for a moderate protective effect also. Hypomethylation of NADK detected in discordant controls may be responsible for an increase in expression of NADK and in turn a greater protective effect against the potentially hypertension-inducing effects of oxidative stress - potentially explaining why individuals with a higher genetic risk for hypertension might not express the phenotype. A SNP (rs1130355) for which there is no data within the significant *NADK* CpG site with a high minor allele frequency (MAF) of 0.36 could be the true source of the association with hypertension. If it were this SNP itself it would be a novel association.

CHID1 and *APTX* were also differentially methylated across several group comparisons of hypertension phenotype and genotype. CHID1 may function to combat inflammation and its effect on hypertension - also ultimately by oxidative stress - by neutralising LPS and other pathogen-derived antigens. APTX is involved in DNA repair in response to oxidative stress, along with another differentially methylated gene *PARP1*, which it is known to interact with. *APTX* and *PARP1* may simply be genes which are activated in response to DNA damage occurring in the context of hypertension rather than being causative themselves.

Further work is required to verify these data, ideally with the gold-standard bisulfite sequencing method. The next steps would be to assess whether these instances of confirmed DNA methylation truly lead to differential expression of the identified genes or genes which are genomically proximate. Finally this work would be validated in another cohort if possible, potentially elucidating mechanisms of hypertension and identifying epigenetic marks which could be used towards patient stratification for personalised treatment.

Using both molecular and clinical variables in multiple linear regression, a linear model for the development of left ventricular mass index (LVMI) was derived - collagen peptide, a haemoglobin beta peptide, phenylacetylglycine, angiotensin

A/angiontensin II, sex, and systolic blood pressure (SBP) - which describes approximately 29% of the variation in LVMI. Angiotensin A and Angiotensin II were too highly correlated to make a meaningful distinction between them. Sex and SBP were determined to be important independent clinical predictors in a model derived from only clinical predictors, along with body mass index, heart rate and history of congestive heart failure (CHF). With missing data imputed the same combined model is derived but with the addition of BMI and HR, demonstrating that all of the molecular variables in the combined model are independently predictive of the clinical model - except CHF which may have been missing due to having very few occurrences across the dataset.

This indicates that the molecular processes of left ventricular hypertrophy (LVH) underlying these molecular variables are at least partially distinct from the molecular processes related to the clinical variables. Clinical variables such as blood pressure clearly drive the development of LVH, however not all hypertensive patients develop LVH, and these molecules which describe variation which is distinct from easily-measured and well-understood clinical measures are promising for the discovery of novel LVH mechanisms and drugs which could target these mechanisms.

The combination of clinical and molecular data can explained several times the variance of most individual biomarkers, and approximately an additional 10% of total variance than clinical markers or molecular markers separately. Some well-established markers of LVH, such as angiotensin II, only become apparent in the presence of the confounders or the full clinical model, demonstrating the utility of analysing putative predictors of different types in combination when analysing complex diseases/processes.

Urinary phenylacetylglycine appears to be the most important molecular predictor. It is associated with heart failure (an end-point of LVH) and phospholipidosis which may be responsible for causing cardiac cell hypertrophy and fibrosis. Angiotensin II is well known to increase blood pressure and has been shown in rats to have a blood pressure independent effect on LVM – an association also demonstrated in humans in the data presented here. This independent association may be through driving hypertrophy via inflammation

and oxidative stress, and by interactions with the sympathetic nervous system. The inclusion of a haemoglobin beta peptide in the model may represent the causal link between anaemia and left ventricular mass in patients with chronic kidney disease and patients with end-stage renal failure. The collagen peptide may be indicative of cardiac fibrosis which is a process which occurs during the development of LVH and it is characterised by an increase of collagen and other extracellular matrix components, whose accumulation causes stiffness and impairs function.

Some other significant molecules not included in the model were also of interest. miR-18a-3p whose 5p arm is negatively associated with age-related heart failure in mice was instead positively associated LVMI. The other miRNA miR-92b-3p is positively associated with heart failure in human. PICP:CITP ratio has been previously suggested as an index of the coupling between collagen synthesis and collagen degradation. This composite value representing collagen turnover was found to be more strongly predictive of LVMI than the marker of collagen synthesis alone. Our results confirm a previous finding that while blood glucose is predictive of LVMI it is not predictive independently of BMI. Trimethylamine (TMA) is produced by gut bacteria from dietary choline, phosphatidylcholine, and carnitine. In the liver it is oxidised by FMO3 into trimethylamine-n-oxide (TMAO) which is associated with heart failure and poorer prognosis in chronic heart failure patients. We detected a negative LVMI association with molecules with TMA, perhaps due to increased FMO3 activity converting TMA to TMAO.

While the PCA-varimax solution is substantially less time-intensive and is appealing in its inherent description of relationships between input variables and original components, it is far less comprehensive than the screening-modelling approach, it is less well-known and harder to describe the underpinnings of it to other researchers, it provides less useful statistical output, and in a sense it is biased towards the largest datasets when using multiple sources of data (i.e. testing different types of biomolecules with different methodologies).

Where significantly differentially regulated asthma genes comprised a relatively small list which were mostly two-fold dysregulated, there was a broad pattern of

dysregulation associated with COPD and smoking, consisting of thousands genes with <2-fold changes. While only three genes were found statistically significantly dysregulated two-fold in both asthma and COPD, term enrichment analysis showed dysregulation of similar terms indicating the involvement of the complement system and matrix metalloproteinase activity. Metabolomic analysis did not show any significant results after multi-test correction partly due to small sample sizes and also due to incorrect sample selection and by metabolite samples being stored in different media.

Results of the interaction network generated for the induced sputum results indicate a large cluster of downregulated fibronectin and actin associated genes which may correspond to a reduction in epithelial barrier function (via downregulated peptides relating to tight junctions, adherens junctions, desmosomes and microtubules), allowing ongoing sensitisation from allergens and an increased risk of infection from pathogens. While much of the fibronectin associated section of the graph does not maintain significance in asthma amongst smokers, it does mostly maintain large fold changes consistent with the direction of the other asthma contrasts. It also correlates strongly with FEV1 and while steroid dose is not predicted to have a strong effect on expression it does seem to 'act against' the dysregulation of these molecules in asthma.

The interaction network also indicates a number of miRs potentially produced by macrophages, secreted, and targeting gene products in or from other cell types. This section of the graph is much more altered by smoking status, and less consistent with regards steroid dose effect. Many of these miR targets are known or suspected to be involved in asthma and many of the interactions show significant correlations between miR and target.

Somewhat surprising, given data from murine models of asthma and its effect on airway smooth muscle cells, calmodulin is downregulated. Calmodulin is a ubiquitously-expressed protein which regulates a wide variety of relevant processes including inflammation, metabolism, apoptosis, immune response, secretion, mucociliary clearance, and smooth muscle contraction. Calmodulin, along with many other dysregulated genes is also involved in calcium homeostasis. Cell staining could provide valuable insight into the effect of these

dysregulations, in terms of calcium content intracellularly, and within particular compartments like the endoplasmic reticulum.

There are also indications that NF-κB and PI3K pathways may be downregulated - possibly by calmodulin, and other dysregulated molecules with known effects on these pathways - which would also be surprising given their involvement in inflammation and evidence from murine models. NF-κB may instead be involved in pathogenesis (along with many other dysregulated genes) through a contribution to reduced barrier function and a reduced ability for macrophages to kill bacteria. Calmodulin may be involved in these processes too along with a reduction in mucociliary clearance which results in excess mucous and a pro-bacterial environment. The activity of NF-κB and PI3K pathways should be confirmed along with the level of glycolysis, which is normally high in asthmatic airways but in term enrichment indicated the opposite and further analysis of the results had mixed indications – reduction in glycolysis would be consistent with reduced PI3 signalling however.

While its overall expression may not be altered much between groups on an mRNA level, there are two lactoferrin peptides which are dysregulated in different directions. Lactoferrin has many relevant effects including bacteriocidal and bacteriostatic activity, anti-inflammatory activity, wound healing, and attracting of eosinophils (which are increased in number). Lysozyme which it requires for its bacteriocidal activity is regulated differently depending on smoking status, along with lipocalin-1 which both lactoferrin and lysozyme interact with.

In addition to macrophages being reduced in number and there being molecular indications of their reduced capability for bacteriocidal activity, there is also evidence from both miRNAs and from MMP10 mRNA that they are converted towards an M2 (anti-inflammatory) type.

Many of the molecules identified verify previous work in other cohorts, and many others seem to be totally novel and some which require further research to validate and verify - particularly those which appear to run counter to previous findings. Novel asthma-associated miRs and their known/likely targets should be

validated and verified in another cohort, perhaps starting with those miRs with multiple targets with significant correlations (miR-151-5p, miR-155-5p, miR-30a-3p, and miR-34a-3p). Their identity as exosome-contained miRs could be confirmed also (rather than acting intracellularly), as has been done in previous work on macrophage-derived miRNA exosomes. Antagomirs could allow study of their effects in isolation in airway cells *in vitro* and synthetic miRs may prove to be useful for treatment in the future. The results of this study should also be taken in combination with other analyses of asthma (particularly those using human samples) to generate a list of biomarkers which could be used on a much larger sample set, towards identifying subtypes of disease.

Regardless of the scale of the omics study there may be potential for wider usage of data beyond the usual fold changes, term enrichments, and clustering usually reported in the literature. Counts of cell type appear useful for performing correlations to help determine which cell of a mixed sample is responsible for the production or regulation of the molecule-of-interest. The same could be interesting for the effect of drugs, measurements of phenotypes, and anything which is relatively easily quantified in the clinic or lab. Even when relatively few associated data are available, typically-recorded demographic variables can be useful for downstream analysis.

In addition to the extra work that could be routinely done in omics datasets in order to generate hypotheses, a corresponding increase would also be required in the lab to test those hypotheses - the barriers to research in biology are quickly transitioning from technical limitations to logistical problems with their roots in politics and economy. Even many of the statistical methods which have only relatively recently come into use were developed decades ago before the data and computing power were available.

One specific recurring issue which would seem relatively trivial to improve is the completeness of the accompanying data. It is to be expected that some variables are more difficult to gather due to time constraints, patient compliance, and technical difficulties, however simple measurements such as height or age should not be missing, and the effect on the power and accuracy of downstream statistics may not be trivial considering the low samples sizes often involved.

Care should be taken to identify all the relevant clinical variables prior to a study and to place appropriate focus on the lab and clinical work involved in gathering those data.

Ideally a bioinformatician or systems biologist would work in cooperation with other molecular biologists, in order to do more extensive literature searches, though an extensive molecular interactions database and a suite of attached tools would help in this direction. Such a database should cover all types of biomolecular interactions, be free for academic use, be manually curated, and it have an indicated species, sample type, evidence type (e.g. yeast two-hybrid), and publication source for each interaction. An associated mapping tool should allow the stylising of edges to show different types of relationships, and facilitate the application of multiple variables to the nodes, e.g. by shape, fill colour, border colour, and asterisks. It should also allow various additional features such as the inclusion of neighbouring molecules which meet certain conditions (e.g. relation with ‘x’ number of dysregulated molecules), or the displaying of results in the context of a pre-defined pathway.

Doing multi-omics analysis in a collaborative effort is more advantageous to several omics experiments being funded to analyse different sets of biomolecules in different cohorts, but towards the same ends. A multi-omics approach allows for relationships between molecules to be elucidated, which allows for enhanced interpretation and more informed selection of molecules for validation and verification. Similarly the addition of clinical data as in chapters 4 and 5 can help describe more variation than the molecular variables alone, and help more fully describe the molecular results.

Non-linear relationships are still a challenge to analysis, with no single obvious solution. An area of analysis which could be improved is the access to biological knowledge in the literature, and how that can be used to provide a greater context for the results. Heuristic methods should be developed to aid researchers in this respect, but their efficacy will rely on the quality of the underlying database. Ideally these methods for pathway assembly should be flexible to any number of omics sets, and should be trained and tested against well understood systems where possible.

Each disease may have several distinct subtypes and each subtype potentially caused by a multitude of possible polymorphisms within the same set of genes or pathways. This may be the case even where patients are clinically identical. Or perhaps the subtypes are not so distinct, and a disease comprises and large number of possible combinations of disease predictors. Until such diseases are better understood at the molecular level it may be wise to act towards larger, more collaborative efforts (such as EU-MASCARA in chapter 4), using multiple omics methodologies on the same cohort (as in the CAB study in chapter 5), rather than splitting focus into smaller studies with less statistical power and less scope. Using multiple omics methodologies not only allows a greater view of the 'molecular landscape' of a disease but generates data which would not exist if experiments were done separately – e.g. evidence of SNP interference in DNA methylation experiments, correlations between miRNAs and the mRNAs they target, and the metabolic effects of combinations of differentially expressed proteins. Unfortunately in chapter 4 the omics methodologies chosen did not reflect a wide spread of the biology, and in chapter 5 there were small sample sizes and errors in sample storage and selection which compounded problems. In addition the comparison to COPD was hampered by the lack of proteomics samples for that disease type.

Due to the variety of the available datasets this thesis covered a wide spectrum of techniques which were used to analyse a variety of diseases and disease-related variables, some of which were respiratory and others cardiovascular in nature, and therefore it does not come with a clear-cut message. However, a range of approaches were explored in relation to omics-level analysis which not only contribute to the knowledgebase of each individual field – furthering future research therein – but also identify useful methods of integrative analysis to be used for the systems biology analysis of complex human disease.

7. Appendix

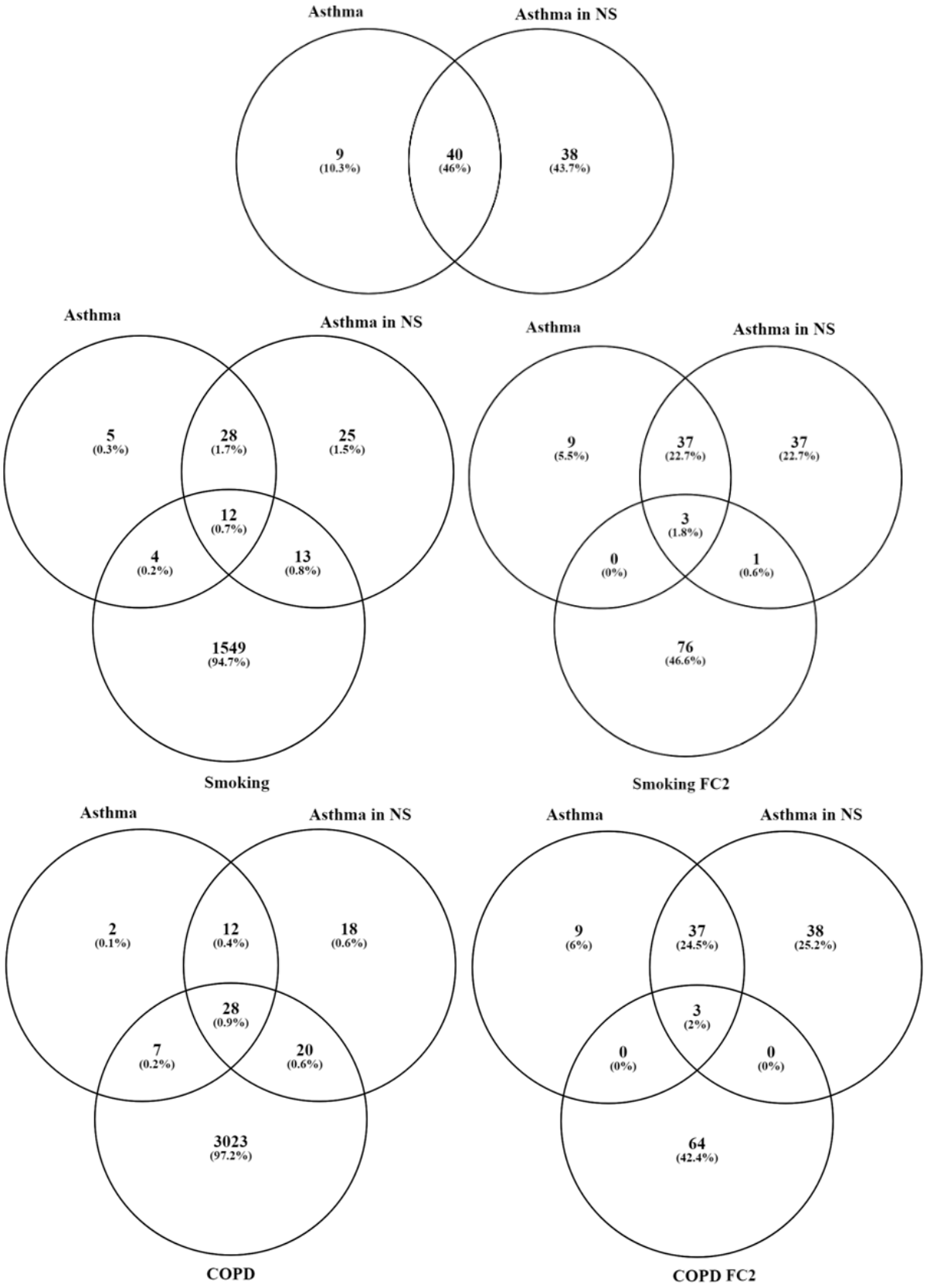

Fig. 7-1 Comparing Results of Different Statistical Contrasts When Applied to mRNA Data. The numbers of mRNA microarray hits are shown for each contrast. NS: non-smoker, FC2: two-fold change criteria also met.

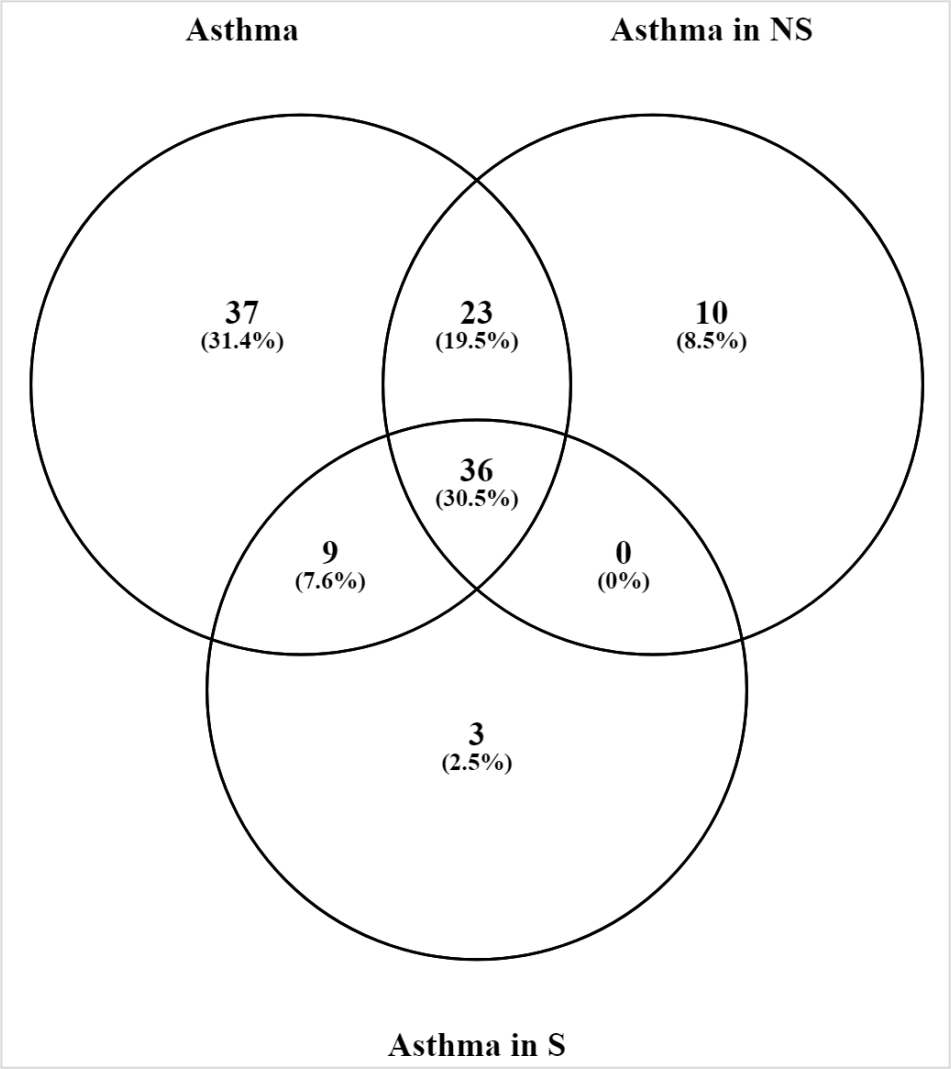

Fig. 7-2 Comparing Results of Different Statistical Contrasts When Applied to Protein Data. The numbers of proteomics hits are shown for each contrast. NS: non-smoker, S: smoker.

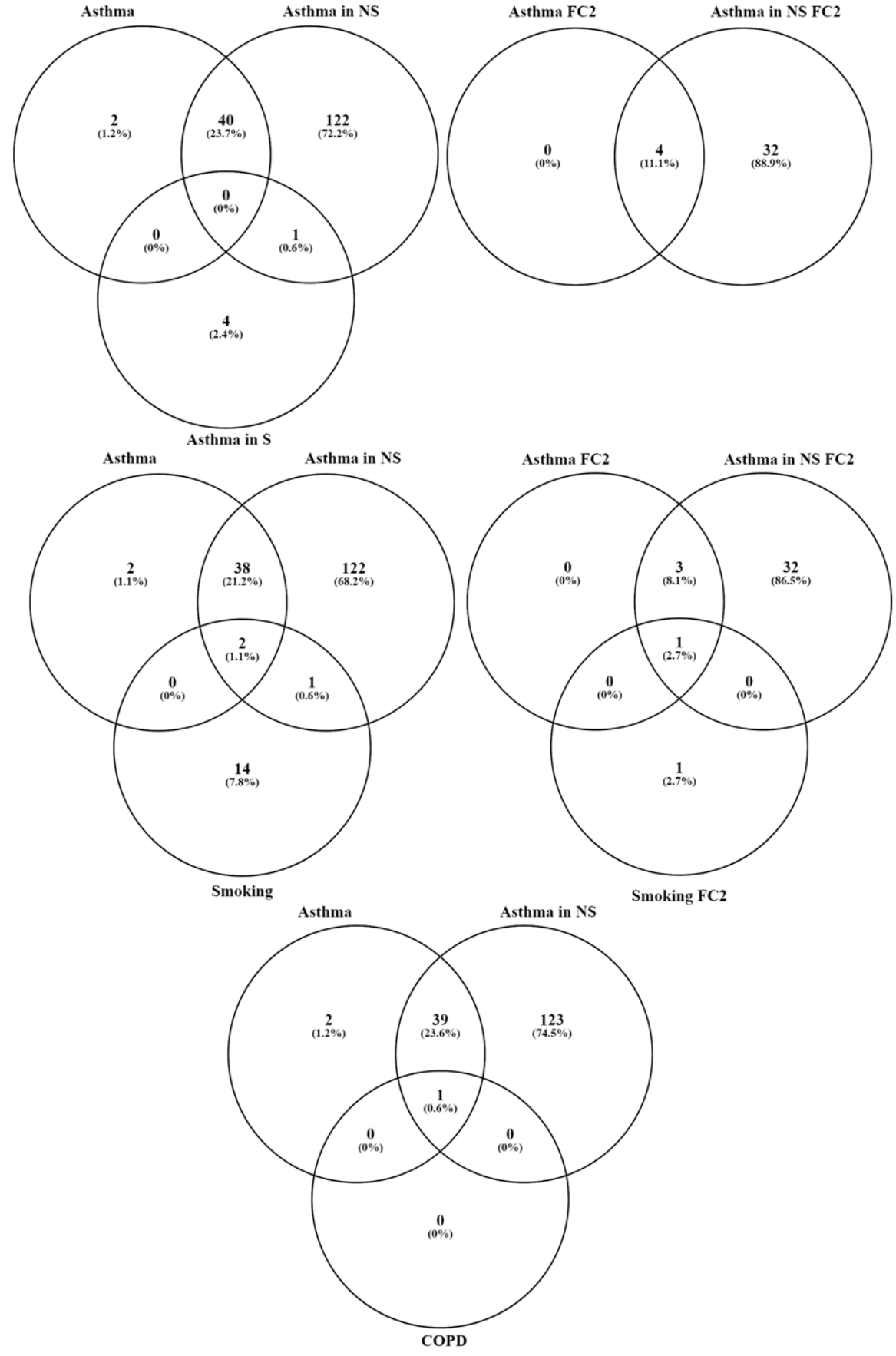

Fig. 7-3 Comparing Results of Different Statistical Contrasts When Applied to miRNA Data. The numbers of miRNA microarray hits are shown for each contrast. NS: non-smoker, S: smoker, FC2: two-fold change criteria also me

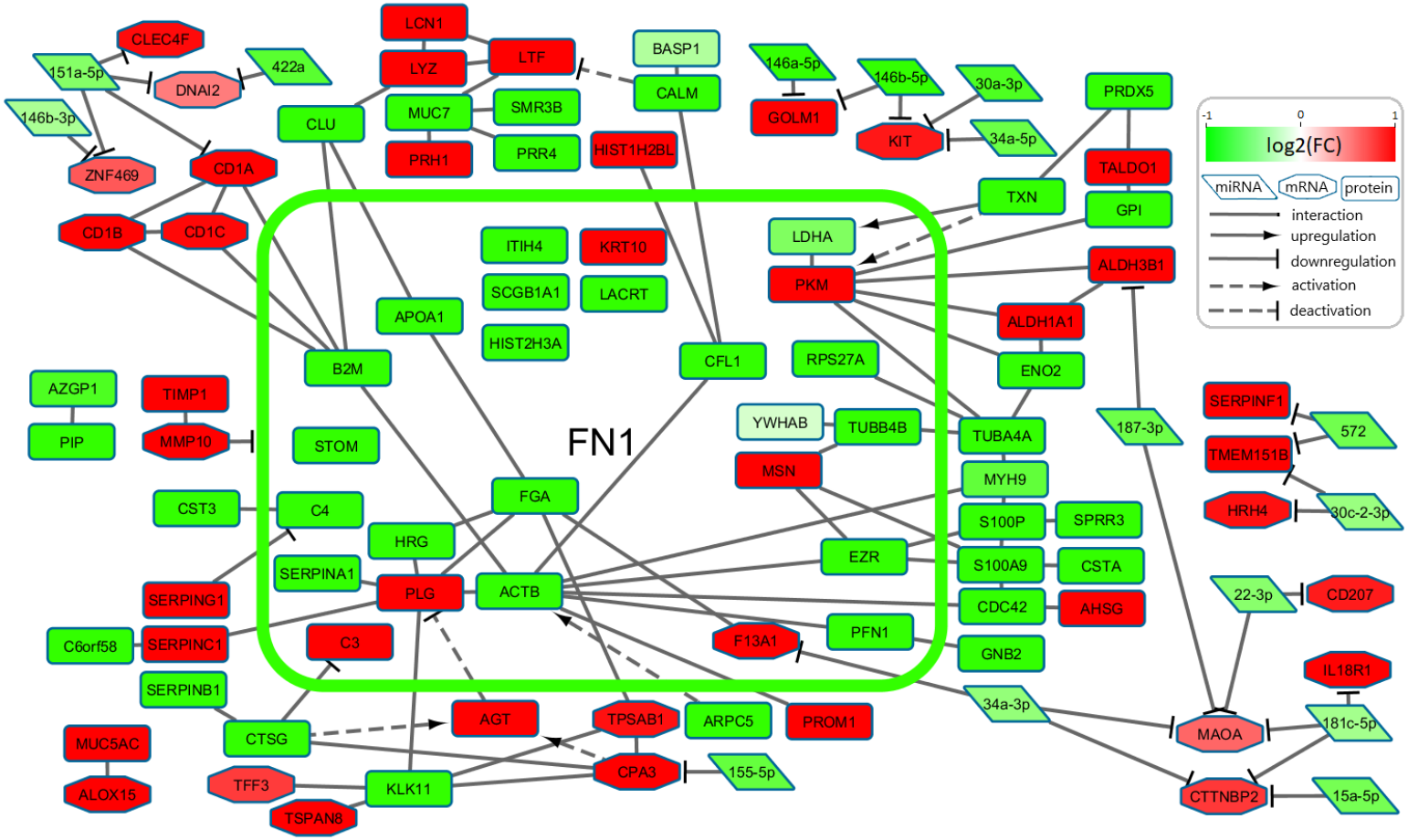

Fig. 7-4 Interaction Network Constructed from Hits from all Three Asthma Contrasts – Coloured By Asthma Fold Change
This network shows interactions between molecules with a significant two-fold difference in at least one of the three asthma contrasts. Interactions were gathered from interaction databases and directly from the literature and only inconsistent directions of fold change were excluded. The FN1 node is represented as a container for many of the nodes it interacts with – reducing the complexity of the network by 26 edges.

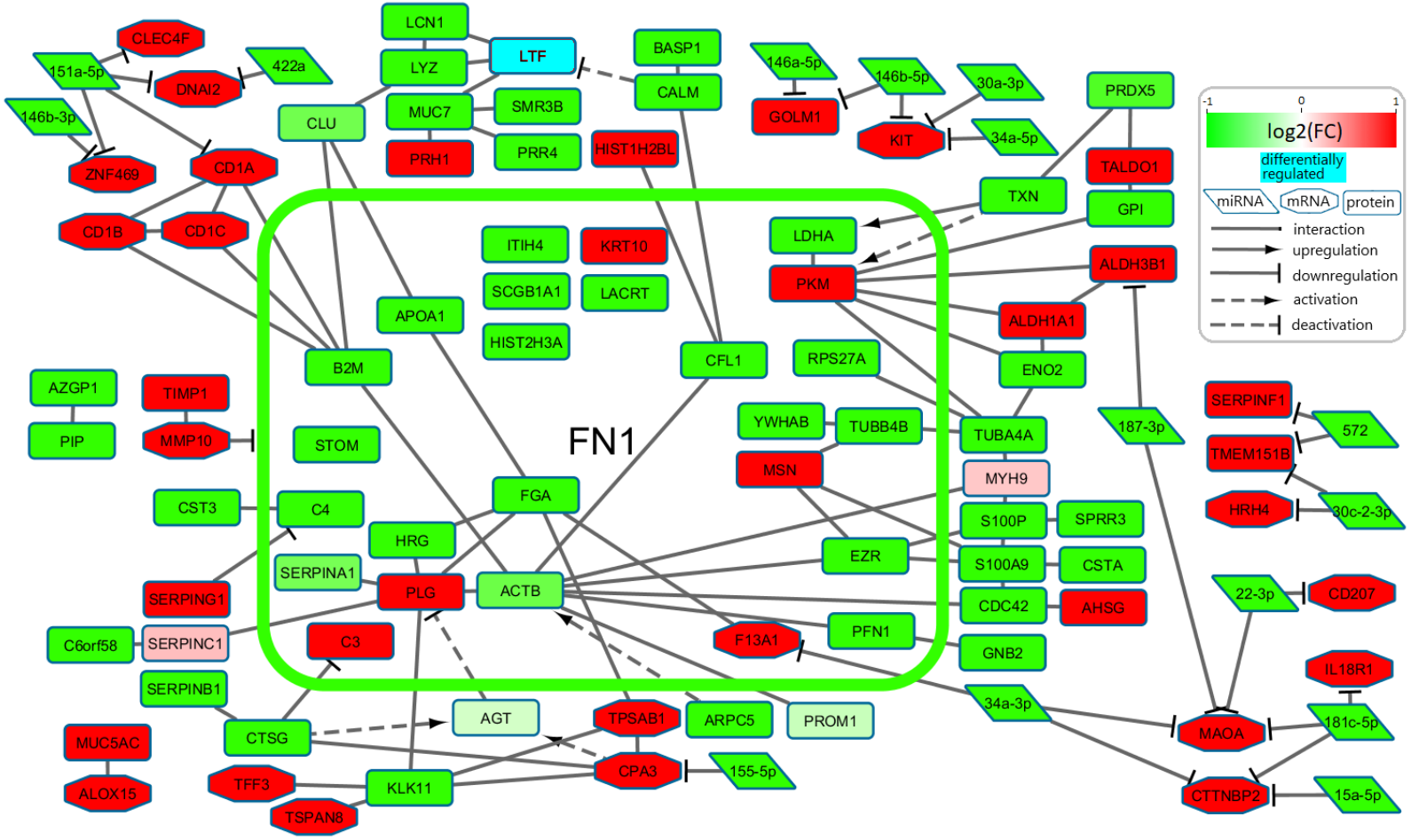

Fig. 7-5 Interaction Network Constructed from Hits from all Three Asthma Contrasts – Coloured By Fold Change of Asthma in Non-Smokers
This network shows interactions between molecules with a significant two-fold difference in at least one of the three asthma contrasts. Interactions were gathered from interaction databases and directly from the literature and only inconsistent directions of fold change were excluded. The FN1 node is represented as a container for many of the nodes it interacts with – reducing the complexity of the network by 26 edges.

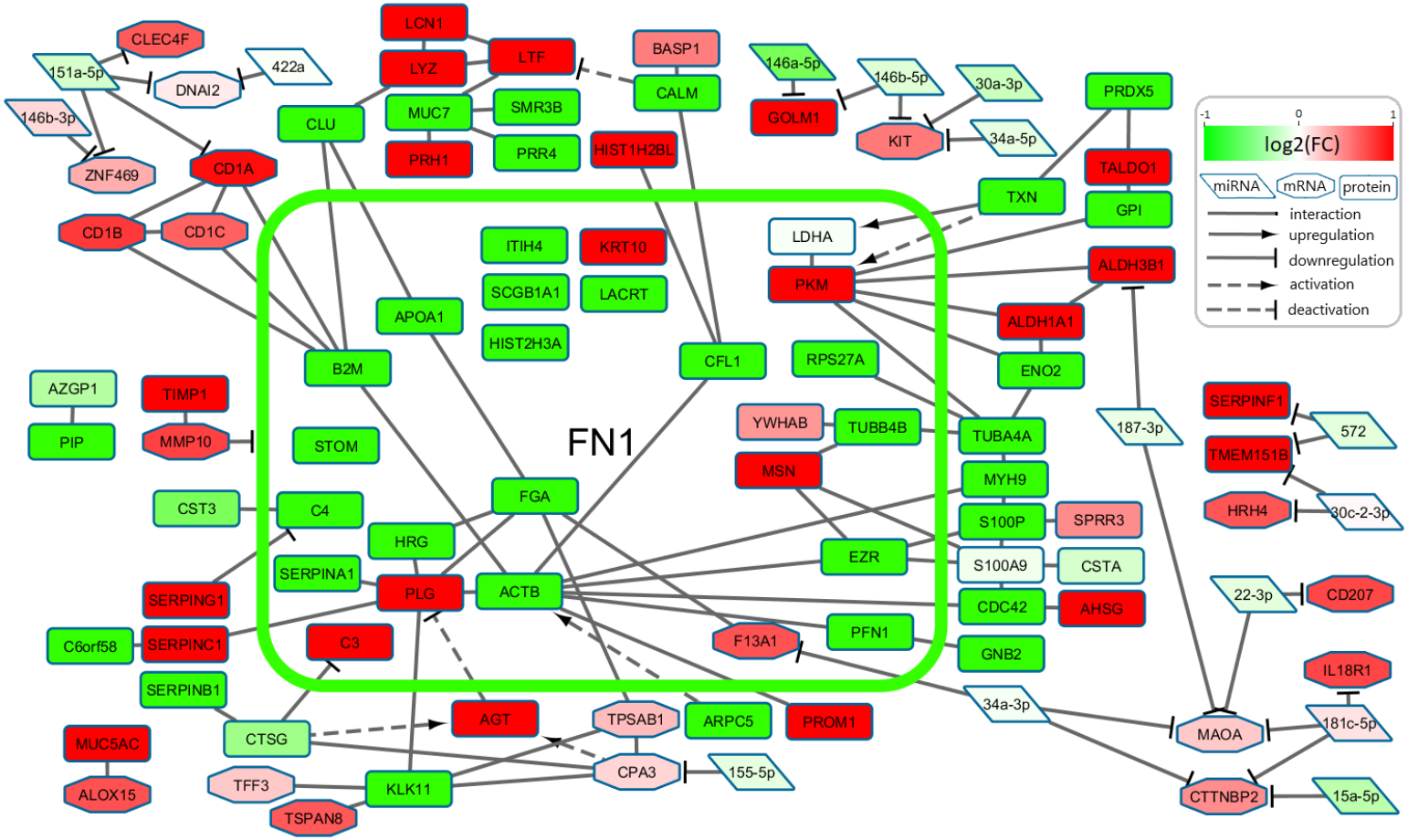

Fig. 7-6 Interaction Network Constructed from Hits from all Three Asthma Contrasts – Coloured By Fold Change of Asthma in Smokers
This network shows interactions between molecules with a significant two-fold difference in at least one of the three asthma contrasts. Interactions were gathered from interaction databases and directly from the literature and only inconsistent directions of fold change were excluded. The FN1 node is represented as a container for many of the nodes it interacts with – reducing the complexity of the network by 26 edges.

8. References

(1) Barker ED, Roberts S, Walton E. Hidden hypotheses in 'hypothesis-free' genome-wide epigenetic associations. Curr Opin Psychol 2018 Jul 25;27:13-7.

(2) Kodama Y, Shumway M, Leinonen R. The Sequence Read Archive: explosive growth of sequencing data. Nucleic Acids Res 2012 Jan;40(Database issue):D54-D56.

(3) Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009 Jan;10(1):57-63.

(4) Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res 2013 Jan;41(Database issue):D991-D995.

(5) Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, Farne A, et al. ArrayExpress--a public database of microarray experiments and gene expression profiles. Nucleic Acids Res 2007 Jan;35(Database issue):D747-D750.

(6) Ludman EJ, Fullerton SM, Spangler L, Trinidad SB, Fujii MM, Jarvik GP, et al. Glad you asked: participants' opinions of re-consent for dbGap data submission. J Empir Res Hum Res Ethics 2010 Sep;5(3):9-16.

(7) Brazma A. Minimum Information About a Microarray Experiment (MIAME)--successes, failures, challenges. ScientificWorldJournal 2009 May 29;9:420-3.

(8) Rustici G, Kolesnikov N, Brandizi M, Burdett T, Dylag M, Emam I, et al. ArrayExpress update--trends in database growth and links to data analysis tools. Nucleic Acids Res 2013 Jan;41(Database issue):D987-D990.

(9) Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, et al. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol 2009;10(11):R130.

(10) Papatheodorou I, Fonseca NA, Keays M, Tang YA, Barrera E, Bazant W, et al. Expression Atlas: gene and protein expression across multiple studies and organisms. Nucleic Acids Res 2018 Jan 4;46(D1):D246-D251.

(11) Robinson SW, Herzyk P, Dow JA, Leader DP. FlyAtlas: database of gene expression in the tissues of Drosophila melanogaster. Nucleic Acids Res 2013 Jan;41(Database issue):D744-D750.

(12) Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. Mass Spectrom Rev 2007 Jan;26(1):51-78.

(13) Griffin TJ, Goodlett DR, Aebersold R. Advances in proteome analysis by mass spectrometry. Curr Opin Biotechnol 2001 Dec;12(6):607-12.

(14) Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0--The Human Metabolome Database in 2013. Nucleic Acids Res 2013 Jan;41(Database issue):D801-D807.

(15) Vizcaino JA, Cote RG, Csordas A, Dianes JA, Fabregat A, Foster JM, et al. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res 2013 Jan;41(Database issue):D1063-D1069.

(16) Steinbeck C, Conesa P, Haug K, Mahendraker T, Williams M, Maguire E, et al. MetaboLights: towards a new COSMOS of metabolomics data management. Metabolomics 2012 Oct;8(5):757-60.

(17) Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000 Jan 1;28(1):27-30.

(18) Altman T, Travers M, Kothari A, Caspi R, Karp PD. A systematic comparison of the MetaCyc and KEGG pathway databases. BMC Bioinformatics 2013 Mar 27;14:112.

(19) Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, Anwar N, et al. Pathway Commons, a web resource for biological pathway data. Nucleic Acids Res 2011 Jan;39(Database issue):D685-D690.

(20) Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. The Reactome pathway knowledgebase. Nucleic Acids Res 2014 Jan;42(Database issue):D472-D477.

(21) van Iersel MP, Kelder T, Pico AR, Hanspers K, Coort S, Conklin BR, et al. Presenting and exploring biological pathways with PathVisio. BMC Bioinformatics 2008 Sep 25;9:399.

(22) Kelder T, van Iersel MP, Hanspers K, Kutmon M, Conklin BR, Evelo CT, et al. WikiPathways: building research communities on biological pathways. Nucleic Acids Res 2012 Jan;40(Database issue):D1301-D1307.

(23) Bader GD, Cary MP, Sander C. Pathguide: a pathway resource list. Nucleic Acids Res 2006 Jan 1;34(Database issue):D504-D506.

(24) Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000 May;25(1):25-9.

(25) CRICK FH. On protein synthesis. Symp Soc Exp Biol 1958;12:138-63.

(26) Baluska F, Witzany G. At the dawn of a new revolution in life sciences. World J Biol Chem 2013 May 26;4(2):13-5.

(27) Brenner S. History of science. The revolution in the life sciences. Science 2012 Dec 14;338(6113):1427-8.

(28) Fuks F. DNA methylation and histone modifications: teaming up to silence genes. Curr Opin Genet Dev 2005 Oct;15(5):490-5.

(29) Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet 2013 Mar;14(3):204-20.

(30) Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, et al. Dynamic changes in the human methylome during differentiation. Genome Research 2010 Mar;20(3):320-31.

(31) Schwartz S, Meshorer E, Ast G. Chromatin organization marks exon-intron structure. Nature Structural & Molecular Biology 2009 Sep;16(9):990-U117.

(32) Sharma A. Transgenerational epigenetic inheritance: focus on soma to germline information transfer. Prog Biophys Mol Biol 2013 Dec;113(3):439-46.

(33) Saletore Y, Meyer K, Korlach J, Vilfan ID, Jaffrey S, Mason CE. The birth of the Epitranscriptome: deciphering the function of RNA modifications. Genome Biol 2012 Oct 31;13(10):175.

(34) Tanzer A, Stadler PF. Molecular evolution of a microRNA cluster. J Mol Biol 2004 May 28;339(2):327-35.

(35) McManus CJ, Graveley BR. RNA structure and the mechanisms of alternative splicing. Curr Opin Genet Dev 2011 Aug;21(4):373-9.

(36) Prabakaran S, Lippens G, Steen H, Gunawardena J. Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. Wiley Interdiscip Rev Syst Biol Med 2012 Nov;4(6):565-83.

(37) Todeschini AL, Georges A, Veitia RA. Transcription factors: specific DNA binding and specific gene regulation. Trends Genet 2014 Jun;30(6):211-9.

(38) Jaskiewicz L, Filipowicz W. Role of Dicer in posttranscriptional RNA silencing. Curr Top Microbiol Immunol 2008;320:77-97.

(39) Noncommunicable Diseases Country Profiles 2018. World Health Organization . 2018.

(40) The Top 10 Causes of Death Fact Sheet. World Health Organization . 24-5-2018.

(41) Gulley SP, Rasch EK, Bethell CD, Carle AC, Druss BG, Houtrow AJ, et al. At the intersection of chronic disease, disability and health services research: A scoping literature review. Disabil Health J 2018 Apr;11(2):192-203.

(42) Allen L, Williams J, Townsend N, Mikkelsen B, Roberts N, Foster C, et al. Socioeconomic status and non-communicable disease behavioural risk factors in low-income and lower-middle-income countries: a systematic review. Lancet Glob Health 2017 Mar;5(3):e277-e289.

(43) Bernell S, Howard SW. Use Your Words Carefully: What Is a Chronic Disease? Front Public Health 2016;4:159.

(44) Gulley SP, Rasch EK, Bethell CD, Carle AC, Druss BG, Houtrow AJ, et al. At the intersection of chronic disease, disability and health services research: A scoping literature review. Disabil Health J 2018 Apr;11(2):192-203.

(45) Currie G, Delles C. Vascular biomedicine in an era of chronic disease and multimorbidity. Clin Sci (Lond) 2019 May 15;133(9):1137-43.

(46) Zenin A, Tsepilov Y, Sharapov S, Getmantsev E, Menshikov LI, Fedichev PO, et al. Identification of 12 genetic loci associated with human healthspan. Commun Biol 2019;2:41.

(47) Pefoyo AJ, Bronskill SE, Gruneir A, Calzavara A, Thavorn K, Petrosyan Y, et al. The increasing burden and complexity of multimorbidity. BMC Public Health 2015 Apr 23;15:415.

(48) Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, et al. Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association. Circulation 2017 Mar 7;135(10):e146-e603.

(49) Cardiovascular diseases (CVDs) Fact Sheet. World Health Organization . 17-5-2017.

(50) Bhatnagar P, Wickramasinghe K, Wilkins E, Townsend N. Trends in the epidemiology of cardiovascular disease in the UK. Heart 2016 Dec 15;102(24):1945-52.

(51) Bloom DE, Cafiero ET, Jané-Llopis E, et al. The Global Economic Burden of Noncommunicable Diseases. World Economic Forum, Geneva 2011.

(52) Stamler R, Stamler J, Riedlinger WF, Algera G, Roberts RH. Family (parental) history and prevalence of hypertension. Results of a nationwide screening program. JAMA 1979 Jan 5;241(1):43-6.

(53) Wei LK, Au A, Teh LK, Lye HS. Recent Advances in the Genetics of Hypertension. Adv Exp Med Biol 2017;956:561-81.

(54) Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, et al. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet 2009 Jun;41(6):666-76.

(55) Newton-Cheh C, Larson MG, Vasan RS, Levy D, Bloch KD, Surti A, et al. Association of common variants in NPPA and NPPB with circulating natriuretic peptides and blood pressure. Nat Genet 2009 Mar;41(3):348-53.

(56) Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 2011 Sep 11;478(7367):103-9.

(57) Padmanabhan S, Melander O, Johnson T, Di Blasio AM, Lee WK, Gentilini D, et al. Genome-Wide Association Study of Blood Pressure Extremes Identifies Variant near UMOD Associated with Hypertension. Plos Genetics 2010 Oct;6(10).

(58) Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nat Genet 2017 Mar;49(3):403-15.

(59) Ehret GB. Genome-Wide Association Studies: Contribution of Genomics to Understanding Blood Pressure and Essential Hypertension. Current Hypertension Reports 2010 Feb;12(1):17-25.

(60) Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nat Genet 2018 Oct;50(10):1412-25.

(61) Taal HR, Verwoert GC, Demirkan A, Janssens ACJW, Rice K, Ehret G, et al. Genome-Wide Profiling of Blood Pressure in Adults and Children. Hypertension 2012 Feb;59(2):241-+.

(62) Hoffmann TJ, Ehret GB, Nandakumar P, Ranatunga D, Schaefer C, Kwok PY, et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. Nat Genet 2017 Jan;49(1):54-64.

(63) Li C, He J, Chen J, Zhao J, Gu D, Hixson JE, et al. Genome-Wide Gene-Sodium Interaction Analyses on Blood Pressure: The Genetic Epidemiology Network of Salt-Sensitivity Study. Hypertension 2016 Aug;68(2):348-55.

(64) Richard MA, Huan T, Ligthart S, Gondalia R, Jhun MA, Brody JA, et al. DNA Methylation Analysis Identifies Loci for Blood Pressure Regulation. Am J Hum Genet 2017 Dec 7;101(6):888-902.

(65) Hawkins NM, Wang D, McMurray JJ, Pfeffer MA, Swedberg K, Granger CB, et al. Prevalence and prognostic implications of electrocardiographic left ventricular hypertrophy in heart failure: evidence from the CHARM programme. Heart 2007 Jan;93(1):59-64.

(66) Stiermaier T, Poss J, Eitel C, de WS, Fuernau G, Desch S, et al. Impact of left ventricular hypertrophy on myocardial injury in patients with ST-segment elevation myocardial infarction. Clin Res Cardiol 2018 Nov;107(11):1013-20.

(67) Berk BC, Fujiwara K, Lehoux S. ECM remodeling in hypertensive heart disease. J Clin Invest 2007 Mar;117(3):568-75.

(68) Murphy AM, Wong AL, Bezuhly M. Modulation of angiotensin II signaling in the prevention of fibrosis. Fibrogenesis Tissue Repair 2015;8:7.

(69) LeGrice IJ, Pope AJ, Sands GB, Whalley G, Doughty RN, Smaill BH. Progression of myocardial remodeling and mechanical dysfunction in the spontaneously hypertensive rat. Am J Physiol Heart Circ Physiol 2012 Dec 1;303(11):H1353-H1365.

(70) Olsen MH, Wachtell K, Hermann KL, Frandsen E, Dige-Petersen H, Rokkedal J, et al. Is cardiovascular remodeling in patients with essential hypertension related to more than high blood pressure? A LIFE substudy. Losartan Intervention For Endpoint-Reduction in Hypertension. Am Heart J 2002 Sep;144(3):530-7.

(71) Meijs MF, Vergouwe Y, Cramer MJ, Vonken EJ, Velthuis BK, Verton DJ, et al. A prediction model for left ventricular mass in patients at high cardiovascular risk. Eur J Cardiovasc Prev Rehabil 2010 Dec;17(6):621-7.

(72) Devereux RB, Roman MJ, de SG, O'Grady MJ, Paranicas M, Yeh JL, et al. Relations of left ventricular mass to demographic and hemodynamic variables in American Indians: the Strong Heart Study. Circulation 1997 Sep 2;96(5):1416-23.

(73) Cuspidi C, Sala C, Negri F, Mancia G, Morganti A. Prevalence of left-ventricular hypertrophy in hypertension: an updated review of echocardiographic studies. J Hum Hypertens 2012 Jun;26(6):343-9.

(74) Nethononda RM, McGurk KA, Whitworth P, Francis J, Mamasoula C, Cordell HJ, et al. Marked variation in heritability estimates of left ventricular mass depending on modality of measurement. Sci Rep 2019 Sep 19;9(1):13556.

(75) Nadour W, Biederman RW. Is left ventricular hypertrophy regression important? Does the tool used to detect it matter? J Clin Hypertens (Greenwich) 2009 Aug;11(8):441-7.

(76) Bella JN, Goring HH. Genetic epidemiology of left ventricular hypertrophy. Am J Cardiovasc Dis 2012;2(4):267-78.

(77) Patel SK, Ramchand J, Crocitti V, Burrell LM. Kruppel-Like Factor 15 Is Critical for the Development of Left Ventricular Hypertrophy. Int J Mol Sci 2018 Apr 27;19(5).

(78) Fajar JK, Pikir BS, Sidarta EP, Berlinda Saka PN, Akbar RR, Heriansyah T. The Gene Polymorphism of Angiotensin-Converting Enzyme Intron Deletion and Angiotensin-Converting Enzyme G2350A in Patients With Left Ventricular Hypertrophy: A Meta-analysis. Indian Heart J 2019 May;71(3):199-206.

(79) Liljedahl U, Kahan T, Malmqvist K, Melhus H, Syvanen AC, Lind L, et al. Single nucleotide polymorphisms predict the change in left ventricular mass in response to antihypertensive treatment. J Hypertens 2004 Dec;22(12):2321-8.

(80) Arnett DK, Li N, Tang W, Rao DC, Devereux RB, Claas SA, et al. Genome-wide association study identifies single-nucleotide polymorphism in KCNB1 associated with left ventricular mass in humans: the HyperGEN Study. BMC Med Genet 2009 May 19;10:43.

(81) Parry HM, Donnelly LA, Van ZN, Doney AS, Elder DH, Morris AD, et al. Genetic variants predicting left ventricular hypertrophy in a diabetic population: a Go-DARTS study including meta-analysis. Cardiovasc Diabetol 2013 Jul 23;12:109.

(82) Shah S, Nelson CP, Gaunt TR, van der Harst P, Barnes T, Braund PS, et al. Four genetic loci influencing electrocardiographic indices of left ventricular hypertrophy. Circ Cardiovasc Genet 2011 Dec;4(6):626-35.

(83) Masoli M, Fabian D, Holt S, Beasley R. The global burden of asthma: executive summary of the GINA Dissemination Committee report. Allergy 2004 May;59(5):469-78.

(84) Chung KF, Wenzel SE, Brozek JL, Bush A, Castro M, Sterk PJ, et al. International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. Eur Respir J 2014 Feb;43(2):343-73.

(85) Holt PG, Sly PD. Viral infections and atopy in asthma pathogenesis: new rationales for asthma prevention and treatment. Nat Med 2012 May 4;18(5):726-35.

(86) Bijanzadeh M, Mahesh PA, Ramachandra NB. An understanding of the genetic basis of asthma. Indian J Med Res 2011 Aug;134:149-61.

(87) Ober C, Yao TC. The genetics of asthma and allergic disease: a 21st century perspective. Immunol Rev 2011 Jul;242(1):10-30.

(88) Svenningsen S, Nair P. Asthma Endotypes and an Overview of Targeted Therapy for Asthma. Front Med (Lausanne) 2017;4:158.

(89) Medrek SK, Parulekar AD, Hanania NA. Predictive Biomarkers for Asthma Therapy. Curr Allergy Asthma Rep 2017 Sep 19;17(10):69.

(90) Menzella F, Galeone C, Bertolini F, Castagnetti C, Facciolongo N. Innovative treatments for severe refractory asthma: how to choose the right option for the right patient? J Asthma Allergy 2017;10:237-47.

(91) Pillai P, Corrigan CJ, Ying S. Airway epithelium in atopic and nonatopic asthma: similarities and differences. ISRN Allergy 2011;2011:195846.

(92) Menzella F, Piro R, Facciolongo N, Castagnetti C, Simonazzi A, Zucchi L. Long-term benefits of omalizumab in a patient with severe non-allergic asthma. Allergy Asthma Clin Immunol 2011 May 24;7(1):9.

(93) Pillai P, Chan YC, Wu SY, Ohm-Laursen L, Thomas C, Durham SR, et al. Omalizumab reduces bronchial mucosal IgE and improves lung function in non-atopic asthma. Eur Respir J 2016 Dec;48(6):1593-601.

(94) Samitas K, Delimpoura V, Zervas E, Gaga M. Anti-IgE treatment, airway inflammation and remodelling in severe allergic asthma: current knowledge and future perspectives. Eur Respir Rev 2015 Dec;24(138):594-601.

(95) Truyen E, Coteur L, Dilissen E, Overbergh L, Dupont LJ, Ceuppens JL, et al. Evaluation of airway inflammation by quantitative Th1/Th2 cytokine mRNA measurement in sputum of asthma patients. Thorax 2006 Mar;61(3):202-8.

(96) Gaga M, Zervas E, Humbert M. Targeting immunoglobulin E in non-atopic asthma: crossing the red line? Eur Respir J 2016 Dec;48(6):1538-40.

(97) Kumar A, Ghosh B. Genetics of asthma: a molecular biologist perspective. Clin Mol Allergy 2009 May 6;7:7.

(98) Grainge CL, Lau LC, Ward JA, Dulay V, Lahiff G, Wilson S, et al. Effect of bronchoconstriction on airway remodeling in asthma. N Engl J Med 2011 May 26;364(21):2006-15.

(99) Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, et al. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. Am J Respir Crit Care Med 2010 Feb 15;181(4):315-23.

(100) Ghebre MA, Bafadhel M, Desai D, Cohen SE, Newbold P, Rapley L, et al. Biological clustering supports both "Dutch" and "British" hypotheses of asthma and chronic obstructive pulmonary disease. J Allergy Clin Immunol 2015 Jan;135(1):63-72.

(101) Hernandez-Pacheco N, Pino-Yanes M, Flores C. Genomic Predictors of Asthma Phenotypes and Treatment Response. Front Pediatr 2019;7:6.

(102) Reilev M, Pottegard A, Lykkegaard J, Sondergaard J, Ingebrigtsen TS, Hallas J. Increased risk of major adverse cardiac events following the onset of acute exacerbations of COPD. Respirology 2019 Dec;24(12):1183-90.

(103) Hobbs BD, de JK, Lamontagne M, Bosse Y, Shrine N, Artigas MS, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. Nat Genet 2017 Mar;49(3):426-32.

(104) Dey T, Kalita J, Weldon S, Taggart CC. Proteases and Their Inhibitors in Chronic Obstructive Pulmonary Disease. J Clin Med 2018 Aug 28;7(9).

(105) Kanazawa H. Role of vascular endothelial growth factor in the pathogenesis of chronic obstructive pulmonary disease. Med Sci Monit 2007 Nov;13(11):RA189-RA195.

(106) Ezzie ME, Crawford M, Cho JH, Orellana R, Zhang S, Gelinas R, et al. Gene expression networks in COPD: microRNA and mRNA regulation. Thorax 2012 Feb;67(2):122-31.

(107) Garudadri S, Woodruff PG. Targeting Chronic Obstructive Pulmonary Disease Phenotypes, Endotypes, and Biomarkers. Ann Am Thorac Soc 2018 Dec;15(Suppl 4):S234-S238.

(108) Angulo M, Lecuona E, Sznajder JI. Role of MicroRNAs in lung disease. Arch Bronconeumol 2012 Sep;48(9):325-30.

(109) Dagouassat M, Gagliolo JM, Chrusciel S, Bourin MC, Duprez C, Caramelle P, et al. The cyclooxygenase-2-prostaglandin E2 pathway maintains senescence of chronic obstructive pulmonary disease fibroblasts. Am J Respir Crit Care Med 2013 Apr 1;187(7):703-14.

(110) Singh D, Kolsum U, Brightling CE, Locantore N, Agusti A, Tal-Singer R. Eosinophilic inflammation in COPD: prevalence and clinical characteristics. Eur Respir J 2014 Dec;44(6):1697-700.

(111) Narendra DK, Hanania NA. Targeting IL-5 in COPD. Int J Chron Obstruct Pulmon Dis 2019;14:1045-51.

(112) Chen YW, Leung JM, Sin DD. A Systematic Review of Diagnostic Biomarkers of COPD Exacerbation. PLoS One 2016;11(7):e0158843.

(113) Fermont JM, Masconi KL, Jensen MT, Ferrari R, Di Lorenzo VAP, Marott JM, et al. Biomarkers and clinical outcomes in COPD: a systematic review and meta-analysis. Thorax 2019 May;74(5):439-46.

(114) Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5(10):R80.

(115) Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003 Apr;4(2):249-64.

(116) Titulaer MK, Siccama I, Dekker LJ, van Rijswijk AL, Heeren RM, Sillevis Smitt PA, et al. A database application for pre-processing, storage and comparison of mass spectra derived from patients and controls. BMC Bioinformatics 2006 Sep 5;7:403.

(117) Creek DJ, Jankevics A, Burgess KE, Breitling R, Barrett MP. IDEOM: an Excel interface for analysis of LC-MS-based metabolomics data. Bioinformatics 2012 Apr 1;28(7):1048-9.

(118) Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 2006 Feb 1;78(3):779-87.

(119) Scheltema RA, Jankevics A, Jansen RC, Swertz MA, Breitling R. PeakML/mzMatch: a file format, Java library, R library, and tool-chain for mass spectrometry data analysis. Anal Chem 2011 Apr 1;83(7):2786-93.

(120) Sammon J. **A Non-Linear Mapping for Data Structure Analysis**. IEEE Trans Comput 1969;18:401.

(121) Pearson K. **On lines and planes of closest fit to systems of points in space**. Philos Mag 1901;2:559-72.

(122) Kasier H. The varimax criterion for analytic rotation in factor-analysis. Psychometrika 1958;23:187-200.

(123) Shah SH, Hauser ER, Bain JR, Muehlbauer MJ, Haynes C, Stevens RD, et al. High heritability of metabolomic profiles in families burdened with premature cardiovascular disease. Mol Syst Biol 2009;5:258.

(124) Slonim N, Atwal GS, Tkacik G, Bialek W. Information-based clustering. Proc Natl Acad Sci U S A 2005 Dec 20;102(51):18297-302.

(125) Wagner CD, Persson PB. Chaos in the cardiovascular system: an update. Cardiovasc Res 1998 Nov;40(2):257-64.

(126) Hochberg Y. **A sharper Bonferroni procedure for multiple tests of significance**. Biometrika 1988;75:800-2.

(127) Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004;3:Article3.

(128) Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 2009 Apr 15;25(8):1091-3.

(129) Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. Bioinformatics 2013 Mar 1;29(5):661-3.

(130) Arakelyan A, Nersisyan L. KEGGParser: parsing and editing KEGG pathway maps in Matlab. Bioinformatics 2013 Feb 15;29(4):518-9.

(131) Wu G, Feng X, Stein L. A human functional protein interaction network and its application to cancer data analysis. Genome Biol 2010;11(5):R53.

(132) Gao J, Tarcea VG, Karnovsky A, Mirel BR, Weymouth TE, Beecher CW, et al. Metscape: a Cytoscape plug-in for visualizing and interpreting metabolomic data in the context of human metabolic networks. Bioinformatics 2010 Apr 1;26(7):971-3.

(133) Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 2011 Oct 6;478(7367):103-9.

(134) Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, et al. Genome-wide association study of blood pressure and hypertension. Nature Genetics 2009 Jun;41(6):677-87.

(135) Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, et al. Genome-wide association study identifies eight loci associated with blood pressure. Nature Genetics 2009 Jun;41(6):666-76.

(136) Wang D, Yan L, Hu Q, Sucheston LE, Higgins MJ, Ambrosone CB, et al. IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data. Bioinformatics 2012 Mar 1;28(5):729-30.

(137) Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450K technology. Epigenomics 2011 Dec;3(6):771-84.

(138) Du P, Zhang XA, Huang CC, Jafari N, Kibbe WA, Hou LF, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. Bmc Bioinformatics 2010 Nov 30;11.

(139) Koestler DC, Avissar-Whiting M, Houseman EA, Karagas MR, Marsit CJ. Differential DNA methylation in umbilical cord blood of infants exposed to low levels of arsenic in utero. Environ Health Perspect 2013 Aug;121(8):971-7.

(140) sva: Surrogate Variable Analysis. R package version 3.6.0. [computer program]. 2013.

(141) Xiong ZG, Wu AH, Bender CM, Tsao JL, Blake C, Shibata D, et al. Mismatch repair deficiency and CpG island hypermethylation in sporadic colon adenocarcinomas. Cancer Epidemiology Biomarkers & Prevention 2001 Jul;10(7):799-803.

(142) Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Research 2002 Jun;12(6):996-1006.

(143) Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. Bmc Bioinformatics 2009 Feb 3;10.

(144) Shoemaker R, Deng J, Wang W, Zhang K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. Genome Research 2010 Jul 1;20(7):883-9.

(145) Neisius U, Bilo G, Taurino C, McClure JD, Schneider MP, Kawecka-Jaszcz K, et al. Association of central and peripheral pulse pressure with intermediate cardiovascular phenotypes. J Hypertens 2012 Jan;30(1):67-74.

(146) Douglas PS, DeCara JM, Devereux RB, Duckworth S, Gardin JM, Jaber WA, et al. Echocardiographic imaging in clinical trials: American Society of Echocardiography Standards for echocardiography core laboratories: endorsed by the American College of Cardiology Foundation. J Am Soc Echocardiogr 2009 Jul;22(7):755-65.

(147) Lang RM, Bierig M, Devereux RB, Flachskampf FA, Foster E, Pellikka PA, et al. Recommendations for chamber quantification. Eur J Echocardiogr 2006 Mar;7(2):79-108.

(148) Ruijter JM, Ramakers C, Hoogaars WM, Karlen Y, Bakker O, van den Hoff MJ, et al. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Res 2009 Apr;37(6):e45.

(149) Pontillo C, Filip S, Borras DM, Mullen W, Vlahou A, Mischak H. CE-MS-based proteomics in biomarker discovery and clinical application. Proteomics Clin Appl 2015 Apr;9(3-4):322-34.

(150) Marrachelli VG, Monleon D, Rentero P, Mansego ML, Morales JM, Galan I, et al. Genomic and metabolomic profile associated to microalbuminuria. PLoS One 2014;9(2):e98227.

(151) Cauwenberghs N, Ravassa S, Thijs L, Haddad F, Yang WY, Wei FF, et al. Circulating Biomarkers Predicting Longitudinal Changes in Left Ventricular Structure and Function in a General Population. J Am Heart Assoc 2019 Jan 22;8(2):e010430.

(152) Lumley T. leaps. http://cran.r-project.org/web/packages/leaps/leaps.pdf. 2009.

(153) Hastie T, Tibshirani R, Narasimhan B, Chu G. impute: Imputation for microarray data. R package version 1.42.0.

(154) Bandyopadhyay A, Roy PP, Saha K, Chakraborty S, Jash D, Saha D. Usefulness of induced sputum eosinophil count to assess severity and treatment outcome in asthma patients. Lung India 2013 Apr;30(2):117-23.

(155) Stanescu D, Sanna A, Veriter C, Kostianev S, Calcagni PG, Fabbri LM, et al. Airways obstruction, chronic expectoration, and rapid decline of FEV1 in smokers are associated with increased levels of sputum neutrophils. Thorax 1996 Mar;51(3):267-71.

(156) Rovina N, Dima E, Gerassimou C, Kollintza A, Gratziou C, Roussos C. IL-18 in induced sputum and airway hyperresponsiveness in mild asthmatics: effect of smoking. Respir Med 2009 Dec;103(12):1919-25.

(157) D'silva L, Allen CJ, Hargreave FE, Parameswaran K. Sputum neutrophilia can mask eosinophilic bronchitis during exacerbations. Can Respir J 2007 Jul;14(5):281-4.

(158) Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5(10):R80.

(159) Wilson CL, Miller CJ. Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. Bioinformatics 2005 Sep 15;21(18):3683-5.

(160) Brettschneider J, Collin F, Bolstad B, Speed T. Quality assessment for short oligonucleotide microarray data. Technometrics 2008;50(3):241-64.

(161) affyQCReport: QC Report Generation for affyBatch objects [computer program]. Version R package version 1.38.0 2014.

(162) Corrplot: Visualization of a Correlation Matrix [computer program]. Version R package version 0.73 2013.

(163) Suarez-Farinas M, Pellegrino M, Wittkowski KM, Magnasco MO. Harshlight: a "corrective make-up" program for microarray chips. BMC Bioinformatics 2005;6:294.

(164) Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, et al. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res 2005;33(20):e175.

(165) Wu Z, Irizarry R, Gentleman RC, Martinez-Murillo F, Spencer F. A model-based background adjustment for oligonucleotide expression arrays. J Am Stat Assoc 2014;99:909-18.

(166) Bourgon R, Gentleman R, Huber W. Independent filtering increases detection power for high-throughput experiments. Proc Natl Acad Sci U S A 2010 May 25;107(21):9546-51.

(167) Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics 2005 May 1;21(9):2067-75.

(168) Vinaixa M, Samino S, Saez I, Duran J, Guinovart JJ, Yanes O. A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data. Metabolites 2012 Oct 18;2(4):775-95.

(169) Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 2009 Apr 15;25(8):1091-3.

(170) Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. Bioinformatics 2013 Mar 1;29(5):661-3.

(171) Ogata H, Goto S, Fujibuchi W, Kanehisa M. Computation with the KEGG pathway database. Biosystems 1998 Jun;47(1-2):119-28.

(172) van Iersel MP, Kelder T, Pico AR, Hanspers K, Coort S, Conklin BR, et al. Presenting and exploring biological pathways with PathVisio. BMC Bioinformatics 2008;9:399.

(173) Montojo J, Zuberi K, Rodriguez H, Kazi F, Wright G, Donaldson SL, et al. GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. Bioinformatics 2010 Nov 15;26(22):2927-8.

(174) Jayapandian M, Chapman A, Tarcea VG, Yu C, Elkiss A, Ianni A, et al. Michigan Molecular Interactions (MiMI): putting the jigsaw puzzle together. Nucleic Acids Res 2007 Jan;35(Database issue):D566-D571.

(175) World Health Organisation. Global Health Risks. 2009.

(176) Sudano I, Roas S, Noll G. Vascular abnormalities in essential hypertension. Curr Pharm Des 2011;17(28):3039-44.

(177) Carretero OA, Oparil S. Essential hypertension. Part I: definition and etiology. Circulation 2000 Jan 25;101(3):329-35.

(178) Carmelli D, Cardon LR, Fabsitz R. Clustering of Hypertension, Diabetes, and Obesity in Adult Male Twins - Same Genes Or Same Environments. American Journal of Human Genetics 1994 Sep;55(3):566-73.

(179) Carmelli D, Robinette D, Fabsitz R. Concordance, Discordance and Prevalence of Hypertension in World-War-Ii Male Veteran Twins. Journal of Hypertension 1994 Mar;12(3):323-8.

(180) Bird A. DNA methylation patterns and epigenetic memory. Genes & Development 2002 Jan 1;16(1):6-21.

(181) Jones PA, Liang GN. OPINION Rethinking how DNA methylation patterns are maintained. Nature Reviews Genetics 2009 Nov;10(11):805-11.

(182) Ooi SKT, Qiu C, Bernstein E, Li KQ, Jia D, Yang Z, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 2007 Aug 9;448(7154):714-U13.

(183) Popp C, Dean W, Feng SH, Cokus SJ, Andrews S, Pellegrini M, et al. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature 2010 Feb 25;463(7284):1101-U126.

(184) Holliday R, Pugh JE. Dna Modification Mechanisms and Gene Activity During Development. Science 1975;187(4173):226-32.

(185) Riggs AD. X-Inactivation, Differentiation, and Dna Methylation. Cytogenetics and Cell Genetics 1975;14(1):9-25.

(186) Lock LF, Takagi N, Martin GR. Methylation of the Hprt Gene on the Inactive-X Occurs After Chromosome Inactivation. Cell 1987 Jan 16;48(1):39-46.

(187) Challen GA, Sun DQ, Jeong M, Luo M, Jelinek J, Berg JS, et al. Dnmt3a is essential for hematopoietic stem cell differentiation. Nature Genetics 2012 Jan;44(1):23-U43.

(188) Rechache NS, Wang YH, Stevenson HS, Killian JK, Edelman DC, Merino M, et al. DNA Methylation Profiling Identifies Global Methylation Differences and Markers of Adrenocortical Tumors. Journal of Clinical Endocrinology & Metabolism 2012 Jun;97(6):E1004-E1013.

(189) Irizarry RA, Ladd-Acosta C, Wen B, Wu ZJ, Montano C, Onyango P, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nature Genetics 2009 Feb;41(2):178-86.

(190) Han H, Cortez CC, Yang XJ, Nichols PW, Jones PA, Liang GN. DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter. Human Molecular Genetics 2011 Nov 15;20(22):4299-310.

(191) Jones PA. The DNA methylation paradox. Trends in Genetics 1999 Jan;15(1):34-7.

(192) Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 2010 Jul 8;466(7303):253-U131.

(193) Schmidl C, Klug M, Boeld TJ, Andreesen R, Hoffmann P, Edinger M, et al. Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. Genome Research 2009 Jul;19(7):1165-74.

(194) Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 2000 May 25;405(6785):482-5.

(195) Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature 2011 Dec 22;480(7378):490-5.

(196) Laird PW. Principles and challenges of genome-wide DNA methylation analysis. Nature Reviews Genetics 2010 Mar;11(3):191-203.

(197) Nestor C, Ruzov A, Meehan R, Dunican D. Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. Biotechniques 2010 Apr;48(4):317-9.

(198) Price EM, Cotton AM, Lam LL, Farre P, Emberly E, Brown CJ, et al. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenetics & Chromatin 2013 Mar 3;6.

(199) Smolarek I, Wyszko E, Barciszewska AM, Nowak S, Gawronska I, Jablecka A, et al. Global DNA methylation changes in blood of patients with essential hypertension. Medical Science Monitor 2010;16(3):CR149-CR155.

(200) Kim M, Long TI, Arakawa K, Wang RW, Yu MC, Laird PW. DNA Methylation as a Biomarker for Cardiovascular Disease Risk. Plos One 2010 Mar 15;5(3).

(201) Kulkarni A, Chavan-Gautam P, Mehendale S, Yadav H, Joshi S. Global DNA Methylation Patterns in Placenta and Its Association with Maternal Hypertension in Pre-Eclampsia. Dna and Cell Biology 2011 Feb;30(2):79-84.

(202) Nomura Y, Lambertini L, Rialdi A, Lee M, Mystal EY, Grabie M, et al. Global Methylation in the Placenta and Umbilical Cord Blood From Pregnancies With Maternal Gestational Diabetes, Preeclampsia, and Obesity. Reproductive Sciences 2014 Jan 1;21(1):131-7.

(203) Yuen RKC, Penaherrera MS, von Dadelszen P, McFadden DE, Robinson WP. DNA methylation profiling of human placentas reveals promoter hypomethylation of multiple genes in early-onset preeclampsia. European Journal of Human Genetics 2010 Sep 10;18(9):1006-12.

(204) Millis RM. Epigenetics and Hypertension. Current Hypertension Reports 2011 Feb;13(1):21-8.

(205) Wang XL, Falkner B, Zhu HD, Shi HD, Su SY, Xu XJ, et al. A Genome-Wide Methylation Study on Essential Hypertension in Young African American Males. Plos One 2013 Jan 10;8(1).

(206) Kazmi N, Elliott HR, Burrows K, Tillin T, Hughes AD, Chaturvedi N, et al. Associations between high blood pressure and DNA methylation. PLoS One 2020;15(1):e0227728.

(207) Navia-Pelaez JM, Campos GP, Araujo-Souza JC, Stergiopulos N, Capettini LSA. Modulation of nNOS(ser852) phosphorylation and translocation by PKA/PP1 pathway in endothelial cells. Nitric Oxide 2018 Jan 30;72:52-8.

(208) Santos CX, Hafstad AD, Beretta M, Zhang M, Molenaar C, Kopec J, et al. Targeted redox inhibition of protein phosphatase 1 by Nox4 regulates eIF2alpha-mediated stress signaling. EMBO J 2016 Feb 1;35(3):319-34.

(209) Wang Y, Zhang F, Wang J, Hu L, Jiang F, Chen J, et al. lncRNA LOC100132354 promotes angiogenesis through VEGFA/VEGFR2 signaling pathway in lung adenocarcinoma. Cancer Manag Res 2018;10:4257-66.

(210) Pluznick JL. Renal and cardiovascular sensory receptors and blood pressure regulation. Am J Physiol Renal Physiol 2013 Aug 15;305(4):F439-F444.

(211) Bostrom AE, Mwinyi J, Voisin S, Wu W, Schultes B, Zhang K, et al. Longitudinal genome-wide methylation study of Roux-en-Y gastric bypass patients reveals novel CpG sites associated with essential hypertension. BMC Med Genomics 2016 Apr 22;9:20.

(212) Richard MA, Huan T, Ligthart S, Gondalia R, Jhun MA, Brody JA, et al. DNA Methylation Analysis Identifies Loci for Blood Pressure Regulation. Am J Hum Genet 2017 Dec 7;101(6):888-902.

(213) Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. Lancet Neurol 2016 Feb;15(2):174-84.

(214) Anttila V, Winsvold BS, Gormley P, Kurth T, Bettella F, McMahon G, et al. Genome-wide meta-analysis identifies new susceptibility loci for migraine. Nat Genet 2013 Aug;45(8):912-7.

(215) Jia RZ, Zhang X, Hu P, Liu XM, Hua XD, Wang X, et al. Screening for differential methylation status in human placenta in preeclampsia using a CpG island plus promoter microarray. International Journal of Molecular Medicine 2012 Jul;30(1):133-41.

(216) Gonzalez-Jaramillo V, Portilla-Fernandez E, Glisic M, Voortman T, Bramer W, Chowdhury R, et al. The role of DNA methylation and histone modifications in blood pressure: a systematic review. J Hum Hypertens 2019 Oct;33(10):703-15.

(217) Wang C, Chen R, Cai J, Shi J, Yang C, Tse LA, et al. Personal exposure to fine particulate matter and blood pressure: A role of angiotensin converting enzyme and its DNA methylation. Environ Int 2016 Sep;94:661-6.

(218) Iyanagi T. Structure and function of NADPH-cytochrome P450 reductase and nitric oxide synthase reductase domain. Biochemical and Biophysical Research Communications 2005 Dec 9;338(1):520-8.

(219) Ziboh VA, Dreize MA, Hsia SL. Inhibition of Lipid Synthesis and Glucose-6-Phosphate Dehydrogenase in Rat Skin by Dehydroepiandrosterone. Journal of Lipid Research 1970;11(4):346-&.

(220) Gaskin F, Clayton RB. Interstrain Difference in Cholesterol-Synthesis In-Vitro in Mice, Dependent Upon A Difference in Endogenous Nadph-Generating Capacity. Journal of Lipid Research 1972;13(1):106-&.

(221) Baldridge CW, Gerard RW. The extra respiration of phagocytosis. American Journal of Physiology 1933 Jan;103(1):235-6.

(222) Rossi F, Zatti M. Biochemical Aspects of Phagocytosis in Polymorphonuclear Leucocytes . Nadh + Nadph Oxidation by Granules of Resting + Phagocytizing Cells. Experientia 1964;20(1):21-&.

(223) Li JM, Shah AM. ROS generation by nonphagocytic NADPH oxidase: Potential relevance in diabetic nephropathy. Journal of the American Society of Nephrology 2003 Aug;14:S221-S226.

(224) Mustacich D, Powis G. Thioredoxin reductase. Biochemical Journal 2000 Feb 15;346:1-8.

(225) Dringen R. Metabolism and functions of glutathione in brain. Progress in Neurobiology 2000 Dec;62(6):649-71.

(226) Kirkman HN, Rolfo M, Ferraris AM, Gaetani GF. Mechanisms of protection of catalase by NADPH - Kinetics and stoichiometry. Journal of Biological Chemistry 1999 May 14;274(20):13908-14.

(227) Pollak N, Niere M, Ziegler M. NAD kinase levels control the NADPH concentration in human cells. Journal of Biological Chemistry 2007 Nov 16;282(46):33562-71.

(228) Kirkman HN, Gaetani GD, Clemons EH, Mareni C. Red-Cell Nadp+ and Nadph in Glucose-6-Phosphate-Dehydrogenase Deficiency. Journal of Clinical Investigation 1975;55(4):875-8.

(229) Santillo M, Colantuoni A, Mondola P, Guida B, Damiano S. NOX signaling in molecular cardiovascular mechanisms involved in the blood pressure homeostasis. Front Physiol 2015;6:194.

(230) Togliatto G, Lombardo G, Brizzi MF. The Future Challenge of Reactive Oxygen Species (ROS) in Hypertension: From Bench to Bed Side. Int J Mol Sci 2017 Sep 15;18(9).

(231) Huang QS, Xie XL, Liang G, Gong F, Wang Y, Wei XQ, et al. The GH18 family of chitinases: Their domain architectures, functions and evolutions. Glycobiology 2012 Jan;22(1):23-34.

(232) Huang L, Wang A, Hao Y, Li W, Liu C, Yang Z, et al. Macrophage Depletion Lowered Blood Pressure and Attenuated Hypertensive Renal Injury and Fibrosis. Front Physiol 2018;9:473.

(233) Meng G, Zhao Y, Bai X, Liu Y, Green TJ, Luo M, et al. Structure of human stabilin-1 interacting chitinase-like protein (SI-CLP) reveals a saccharide-binding cleft with lower sugar-binding selectivity. J Biol Chem 2010 Dec 17;285(51):39898-904.

(234) Kzhyshkowska J, Mamidi S, Gratchev A, Kremmer E, Schmuttermaier C, Krusell L, et al. Novel stabilin-1 interacting chitinase-like protein (SI-CLP) is up-regulated in alternatively activated macrophages and secreted via lysosomal pathway. Blood 2006 Apr 15;107(8):3221-8.

(235) Kzhyshkowska J, Mamidi S, Gratchev A, Kremmer E, Schmuttermaier C, Krusell L, et al. Novel stabilin-1 interacting chitinase-like protein (SI-CLP) is up-regulated in alternatively activated macrophages and secreted via lysosomal pathway. Blood 2006 Apr 15;107(8):3221-8.

(236) Park SY, Jung MY, Lee SJ, Kang KB, Gratchev A, Riabov V, et al. Stabilin-1 mediates phosphatidylserine-dependent clearance of cell corpses in alternatively activated macrophages. J Cell Sci 2009 Sep 15;122(Pt 18):3365-73.

(237) Szondy Z, Sarang Z, Kiss B, Garabuczi Ã, KÃ¶rÃ¶skÃ©nyi K. Anti-inflammatory Mechanisms Triggered by Apoptotic Cells during Their Clearance. Front Immunol 2017;8:909.

(238) Adachi H, Tsujimoto M. FEEL-1, a novel scavenger receptor with in vitro bacteria-binding and angiogenesis-modulating activities. Journal of Biological Chemistry 2002 Sep 13;277(37):34264-70.

(239) Trpkovic A, Resanovic I, Stanimirovic J, Radak D, Mousa SA, Cenic-Milosevic D, et al. Oxidized low-density lipoprotein as a biomarker of cardiovascular diseases. Crit Rev Clin Lab Sci 2015;52(2):70-85.

(240) Kzhyshkowska J, Workman G, CardÃ³-Vila M, Arap W, Pasqualini R, Gratchev A, et al. Novel function of alternatively activated macrophages: stabilin-1-mediated clearance of SPARC. J Immunol 2006 May 15;176(10):5825-32.

(241) Lankhorst S, Saleh L, Danser AJ, van den Meiracker AH. Etiology of angiogenesis inhibition-related hypertension. Curr Opin Pharmacol 2015 Apr;21:7-13.

(242) Negishi H, Ikeda K, Kuga S, Noguchi T, Kanda T, Njelekela M, et al. The relation of oxidative DNA damage to hypertension and other cardiovascular risk factors in Tanzania. Journal of Hypertension 2001 Mar;19(3):529-33.

(243) Meloche J, Pflieger A, Vaillancourt M, Paulin R, Potus F, Zervopoulos S, et al. Role for DNA Damage Signaling in Pulmonary Arterial Hypertension. Circulation 2013 Nov 22.

(244) Vorobyov E, Mertsalov I, Dockhorn-Dworniczak B, Dworniczak B, Horst J. The genomic organization and the full coding region of the human PAX7 gene. Genomics 1997 Oct 1;45(1):168-74.

(245) Schmidts M, Frank V, Eisenberger T, Al TS, Bizet AA, Antony D, et al. Combined NGS approaches identify mutations in the intraflagellar transport gene IFT140 in skeletal ciliopathies with early progressive kidney Disease. Hum Mutat 2013 May;34(5):714-24.

(246) Colbern GT, Chiang MH, Main EK. Expression of the nonclassic histocompatibility antigen HLA-G by preeclamptic placenta. Am J Obstet Gynecol 1994 May;170(5 Pt 1):1244-50.

(247) Tang Y, Liu H, Li H, Peng T, Gu W, Li X. Hypermethylation of the HLA-G promoter is associated with preeclampsia. Mol Hum Reprod 2015 Sep;21(9):736-44.

(248) Kato N, Loh M, Takeuchi F, Verweij N, Wang X, Zhang W, et al. Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. Nat Genet 2015 Nov;47(11):1282-93.

(249) Sanderson JE, Tse TF. Heart failure: a global disease requiring a global response. Heart 2003 Jun;89(6):585-6.

(250) Gardin JM, Arnold A, Gottdiener JS, Wong ND, Fried LP, Klopfenstein HS, et al. Left ventricular mass in the elderly. The Cardiovascular Health Study. Hypertension 1997 May;29(5):1095-103.

(251) Shub C, Klein AL, Zachariah PK, Bailey KR, Tajik AJ. Determination of left ventricular mass by echocardiography in a normal population: effect of age and sex in addition to body size. Mayo Clin Proc 1994 Mar;69(3):205-11.

(252) Kurland L, Liljedahl U, Karlsson J, Kahan T, Malmqvist K, Melhus H, et al. Angiotensinogen gene polymorphisms: relationship to blood pressure response to antihypertensive treatment. Results from the Swedish Irbesartan Left Ventricular Hypertrophy Investigation vs Atenolol (SILVHIA) trial. Am J Hypertens 2004 Jan;17(1):8-13.

(253) Heymans S, Schroen B, Vermeersch P, Milting H, Gao F, Kassner A, et al. Increased cardiac expression of tissue inhibitor of metalloproteinase-1 and tissue inhibitor of metalloproteinase-2 is related to cardiac fibrosis and dysfunction in the chronic pressure-overloaded human heart. Circulation 2005 Aug 23;112(8):1136-44.

(254) Neisius U, Bilo G, Taurino C, McClure JD, Schneider MP, Kawecka-Jaszcz K, et al. Association of central and peripheral pulse pressure with intermediate cardiovascular phenotypes. J Hypertens 2012 Jan;30(1):67-74.

(255) Gligorijevic V, Przulj N. Methods for biological data integration: perspectives and challenges. J R Soc Interface 2015 Nov 6;12(112).

(256) Aittokallio T. Dealing with missing values in large-scale studies: microarray data imputation and beyond. Brief Bioinform 2010 Mar;11(2):253-64.

(257) Kaiser H. The application of electronic computers to factor analysis. Educational and Psychological Measurement 1960.

(258) Kang SM, Park JC, Shin MJ, Lee H, Oh J, Ryu dH, et al. (1)H nuclear magnetic resonance based metabolic urinary profiling of patients with ischemic heart failure. Clin Biochem 2011 Mar;44(4):293-9.

(259) Breiden B, Sandhoff K. Emerging mechanisms of drug-induced phospholipidosis. Biol Chem 2019 Dec 18;401(1):31-46.

(260) Anderson N, Borlak J. Drug-induced phospholipidosis. FEBS Lett 2006 Oct 9;580(23):5533-40.

(261) Reasor MJ, Kacew S. Drug-induced phospholipidosis: are there functional consequences? Exp Biol Med (Maywood) 2001 Oct;226(9):825-30.

(262) Di GF, Claver E, Olive M, Salazar-Mendiguchia J, Manito N, Cequier A. Dilated Cardiomyopathy and Hydroxychloroquine-induced Phospholipidosis: From Curvilinear Bodies to Clinical Suspicion. Rev Esp Cardiol (Engl Ed) 2018 Jun;71(6):491-3.

(263) Scheurle C, Dammrich M, Becker JU, Baumgartel MW. Renal phospholipidosis possibly induced by ranolazine. Clin Kidney J 2014 Feb;7(1):62-4.

(264) Kaloyanides GJ, Pastoriza-Munoz E. Aminoglycoside nephrotoxicity. Kidney Int 1980 Nov;18(5):571-82.

(265) Samadian T, Dehpour AR, Amini S, Nouhnejad P. Inhibition of gentamicin-induced nephrotoxicity by lithium in rat. Histol Histopathol 1993 Jan;8(1):139-47.

(266) Kishore BK, Ibrahim S, Lambricht P, Laurent G, Maldague P, Tulkens PM. Comparative assessment of poly-L-aspartic and poly-L-glutamic acids as protectants against gentamicin-induced renal lysosomal phospholipidosis, phospholipiduria and cell proliferation in rats. J Pharmacol Exp Ther 1992 Jul;262(1):424-32.

(267) Bansal N, Lin F, Vittinghoff E, Peralta C, Lima J, Kramer H, et al. Estimated GFR and Subsequent Higher Left Ventricular Mass in Young and Middle-Aged Adults With Normal Kidney Function: The Coronary Artery Risk Development in Young Adults (CARDIA) Study. Am J Kidney Dis 2016 Feb;67(2):227-34.

(268) Aumann N, Baumeister SE, Werner A, Wallaschofski H, Hannemann A, Nauck M, et al. Inverse association of estimated cystatin C- and creatinine-based glomerular filtration rate with left ventricular mass: Results from the Study of Health in Pomerania. Int J Cardiol 2013 Sep 10;167(6):2786-91.

(269) Nerpin E, Ingelsson E, Riserus U, Sundstrom J, Andren B, Jobs E, et al. The association between glomerular filtration rate and left ventricular function in two independent community-based cohorts of elderly. Nephrol Dial Transplant 2014 Nov;29(11):2069-74.

(270) Moody WE, Ferro CJ, Edwards NC, Chue CD, Lin EL, Taylor RJ, et al. Cardiovascular Effects of Unilateral Nephrectomy in Living Kidney Donors. Hypertension 2016 Feb;67(2):368-77.

(271) Matsumoto M, Zhang CH, Kosugi C, Matsumoto I. Gas chromatography-mass spectrometric studies of canine urinary metabolism. J Vet Med Sci 1995 Apr;57(2):205-11.

(272) De Mello WC, Monterrubio J. Intracellular and extracellular angiotensin II enhance the L-type calcium current in the failing heart. Hypertension 2004 Sep;44(3):360-4.

(273) Agabiti-Rosei E, Ambrosioni E, Dal PC, Muiesan ML, Zanchetti A. ACE inhibitor ramipril is more effective than the beta-blocker atenolol in reducing left ventricular mass in hypertension. Results of the RACE (ramipril cardioprotective evaluation) study on behalf of the RACE study group. J Hypertens 1995 Nov;13(11):1325-34.

(274) Dahlof B, Zanchetti A, Diez J, Nicholls MG, Yu CM, Barrios V, et al. Effects of losartan and atenolol on left ventricular mass and neurohormonal profile in patients with essential hypertension and left ventricular hypertrophy. J Hypertens 2002 Sep;20(9):1855-64.

(275) Kurdi M, Booz GW. New take on the role of angiotensin II in cardiac hypertrophy and fibrosis. Hypertension 2011 Jun;57(6):1034-8.

(276) Coutinho DC, Foureaux G, Rodrigues KD, Salles RL, Moraes PL, Murca TM, et al. Cardiovascular effects of angiotensin A: a novel peptide of the renin-angiotensin system. J Renin Angiotensin Aldosterone Syst 2014 Dec;15(4):480-6.

(277) Coutinho DC, Foureaux G, Rodrigues KD, Salles RL, Moraes PL, Murca TM, et al. Cardiovascular effects of angiotensin A: a novel peptide of the renin-angiotensin system. J Renin Angiotensin Aldosterone Syst 2014 Dec;15(4):480-6.

(278) Yang R, Smolders I, Vanderheyden P, Demaegdt H, Van EA, Vauquelin G, et al. Pressor and renal hemodynamic effects of the novel angiotensin A peptide are angiotensin II type 1A receptor dependent. Hypertension 2011 May;57(5):956-64.

(279) Park BM, Ai Phuong HT, Li W, Kim SH. Similarity and dissimilarity between angiotensin A and angiotensin II in cardiovascular functions in a rat model. Peptides 2020 Mar 10;127:170298.

(280) Hrenak J, Paulis L, Simko F. Angiotensin A/Alamandine/MrgD Axis: Another Clue to Understanding Cardiovascular Pathophysiology. Int J Mol Sci 2016 Jul 20;17(7).

(281) Portoles J, Torralbo A, Martin P, Rodrigo J, Herrero JA, Barrientos A. Cardiovascular effects of recombinant human erythropoietin in predialysis patients. Am J Kidney Dis 1997 Apr;29(4):541-8.

(282) Parfrey PS, Lauve M, Latremouille-Viau D, Lefebvre P. Erythropoietin therapy and left ventricular mass index in CKD and ESRD patients: a meta-analysis. Clin J Am Soc Nephrol 2009 Apr;4(4):755-62.

(283) Querejeta R, LÃ³pez B, GonzÃ¡lez A, SÃ¡nchez E, Larman M, MartÃ-nez Ubago JL, et al. Increased collagen type I synthesis in patients with heart failure of hypertensive origin: relation to myocardial fibrosis. Circulation 2004 Sep 7;110(10):1263-8.

(284) Diez J, Laviades C. Insulin-like growth factor I and collagen type III synthesis in patients with essential hypertension and left ventricular hypertrophy. J Hum Hypertens 1994 Sep;8 Suppl 1:S21-S25.

(285) Querejeta R, Lopez B, Gonzalez A, Sanchez E, Larman M, Martinez Ubago JL, et al. Increased collagen type I synthesis in patients with heart failure of hypertensive origin: relation to myocardial fibrosis. Circulation 2004 Sep 7;110(10):1263-8.

(286) Chen K, Chen J, Li D, Zhang X, Mehta JL. Angiotensin II regulation of collagen type I expression in cardiac fibroblasts: modulation by PPAR-gamma ligand pioglitazone. Hypertension 2004 Nov;44(5):655-61.

(287) Zile MR, Desantis SM, Baicu CF, Stroud RE, Thompson SB, McClure CD, et al. Plasma biomarkers that reflect determinants of matrix composition identify the presence of left ventricular hypertrophy and diastolic heart failure. Circ Heart Fail 2011 May;4(3):246-56.

(288) van Almen GC, Verhesen W, van Leeuwen RE, van d, V, Eurlings C, Schellings MW, et al. MicroRNA-18 and microRNA-19 regulate CTGF and TSP-1 expression in age-related heart failure. Aging Cell 2011 Oct;10(5):769-79.

(289) Li SC, Tsai KW, Pan HW, Jeng YM, Ho MR, Li WH. MicroRNA 3' end nucleotide modification patterns and arm selection preference in liver tissues. BMC Syst Biol 2012;6 Suppl 2:S14.

(290) Li SC, Liao YL, Ho MR, Tsai KW, Lai CH, Lin WC. miRNA arm selection and isomiR distribution in gastric cancer. BMC Genomics 2012;13 Suppl 1:S13.

(291) Goren Y, Kushnir M, Zafrir B, Tabak S, Lewis BS, Amir O. Serum levels of microRNAs in patients with heart failure. Eur J Heart Fail 2012 Feb;14(2):147-54.

(292) Lopez B, Gonzalez A, Varo N, Laviades C, Querejeta R, Diez J. Biochemical assessment of myocardial fibrosis in hypertensive heart disease. Hypertension 2001 Nov;38(5):1222-6.

(293) Kawasaki D, Kosugi K, Waki H, Yamamoto K, Tsujino T, Masuyama T. Role of activated renin-angiotensin system in myocardial fibrosis and left ventricular diastolic dysfunction in diabetic patients--reversal by chronic angiotensin II type 1A receptor blockade. Circ J 2007 Apr;71(4):524-9.

(294) Ciulla MM, Paliotti R, Esposito A, Diez J, Lopez B, Dahlof B, et al. Different effects of antihypertensive therapies based on losartan or atenolol on ultrasound and biochemical markers of myocardial fibrosis: results of a randomized trial. Circulation 2004 Aug 3;110(5):552-7.

(295) Ciulla MM, Paliotti R, Esposito A, Cuspidi C, Muiesan ML, Rosei EA, et al. Effects of antihypertensive treatment on ultrasound measures of myocardial fibrosis in hypertensive patients with left ventricular hypertrophy: results of a randomized trial comparing the angiotensin receptor antagonist, candesartan and the angiotensin-converting enzyme inhibitor, enalapril. J Hypertens 2009 Mar;27(3):626-32.

(296) Ishimitsu T, Kobayashi T, Honda T, Takahashi M, Minami J, Ohta S, et al. Protective effects of an angiotensin II receptor blocker and a long-acting calcium channel blocker against cardiovascular organ injuries in hypertensive patients. Hypertens Res 2005 Apr;28(4):351-9.

(297) Ferrara AL, Vaccaro O, Cardoni O, Panarelli W, Laurenzi M, Zanchetti A. Is there a relationship between left ventricular mass and plasma glucose and lipids independent of body mass index? Results of the Gubbio Study. Nutr Metab Cardiovasc Dis 2003 Jun;13(3):126-32.

(298) Den Ruijter HM, Verkerk AO, Schumacher CA, Houten SM, Belterman CN, Baartscheer A, et al. A diet rich in unsaturated fatty acids prevents progression toward heart failure in a rabbit model of pressure and volume overload. Circ Heart Fail 2012 May 1;5(3):376-84.

(299) Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature 2011 Apr 7;472(7341):57-63.

(300) Lever M, George PM, Slow S, Bellamy D, Young JM, Ho M, et al. Betaine and Trimethylamine-N-Oxide as Predictors of Cardiovascular Outcomes Show Different Patterns in Diabetes Mellitus: An Observational Study. PLoS One 2014;9(12):e114969.

(301) Tang WH, Wang Z, Shrestha K, Borowski AG, Wu Y, Troughton RW, et al. Intestinal microbiota-dependent phosphatidylcholine metabolites, diastolic dysfunction, and adverse clinical outcomes in chronic systolic heart failure. J Card Fail 2015 Feb;21(2):91-6.

(302) Nagashima S, Shimizu M, Yano H, Murayama N, Kumai T, Kobayashi S, et al. Inter-individual variation in flavin-containing monooxygenase 3 in livers from Japanese: correlation with hepatic transcription factors. Drug Metab Pharmacokinet 2009;24(3):218-25.

(303) Miao J, Ling AV, Manthena PV, Gearing ME, Graham MJ, Crooke RM, et al. Flavin-containing monooxygenase 3 as a potential player in diabetes-associated atherosclerosis. Nat Commun 2015;6:6498.

(304) Mendelsohn AR, Larrick JW. Dietary modification of the microbiome affects risk for cardiovascular disease. Rejuvenation Res 2013 Jun;16(3):241-4.

(305) Brugere JF, Borrel G, Gaci N, Tottey W, O'Toole PW, Malpuech-Brugere C. Archaebiotics: proposed therapeutic use of archaea to prevent trimethylaminuria and cardiovascular disease. Gut Microbes 2014 Jan;5(1):5-10.

(306) Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, Sheehy BT, et al. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. Nat Med 2013 May;19(5):576-85.

(307) Chung KF, Wenzel SE, Brozek JL, Bush A, Castro M, Sterk PJ, et al. International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. Eur Respir J 2014 Feb;43(2):343-73.

(308) Kerkhof M, Tran TN, Soriano JB, Golam S, Gibson D, Hillyer EV, et al. Healthcare resource use and costs of severe, uncontrolled eosinophilic asthma in the UK general population. Thorax 2018 Feb;73(2):116-24.

(309) Rennard S, Knobil K, Rabe KF, Morris A, Schachter N, Locantore N, et al. The efficacy and safety of cilomilast in COPD. Drugs 2008;68 Suppl 2:3-57.

(310) Lotvall J, Bateman ED, Busse WW, O'Byrne PM, Woodcock A, Toler WT, et al. Comparison of vilanterol, a novel long-acting beta2 agonist, with placebo and a salmeterol reference arm in asthma uncontrolled by inhaled corticosteroids. J Negat Results Biomed 2014 Jun 13;13(1):9.

(311) Santus P, Saad M, Damiani G, Patella V, Radovanovic D. Current and future targeted therapies for severe asthma: Managing treatment with biologics based on phenotypes and biomarkers. Pharmacol Res 2019 Jun 4;146:104296.

(312) Chupp GL, Bradford ES, Albers FC, Bratton DJ, Wang-Jairaj J, Nelsen LM, et al. Efficacy of mepolizumab add-on therapy on health-related quality of life and markers of asthma control in severe eosinophilic asthma (MUSCA): a randomised, double-blind, placebo-controlled, parallel-group, multicentre, phase 3b trial. Lancet Respir Med 2017 May;5(5):390-400.

(313) Santus P, Ferrando M, Baiardini I, Radovanovic D, Fattori A, Braido F. Patients beliefs on intravenous and subcutaneous routes of administration of biologics for severe asthma treatment: A cross-sectional observational survey study. World Allergy Organ J 2019;12(4):100030.

(314) Humbert M, Busse W, Hanania NA, Lowe PJ, Canvin J, Erpenbeck VJ, et al. Omalizumab in asthma: an update on recent developments. J Allergy Clin Immunol Pract 2014 Sep;2(5):525-36.

(315) Jeffery PK. Remodeling and inflammation of bronchi in asthma and chronic obstructive pulmonary disease. Proc Am Thorac Soc 2004;1(3):176-83.

(316) Cukic V, Lovre V, Dragisic D, Ustamujic A. Asthma and Chronic Obstructive Pulmonary Disease (COPD) - Differences and Similarities. Mater Sociomed 2012;24(2):100-5.

(317) Tho NV, Park HY, Nakano Y. Asthma-COPD overlap syndrome (ACOS): A diagnostic challenge. Respirology 2016 Apr;21(3):410-8.

(318) Kurashima K, Takaku Y, Ohta C, Takayanagi N, Yanagisawa T, Sugita Y. COPD assessment test and severity of airflow limitation in patients with asthma, COPD, and asthma-COPD overlap syndrome. Int J Chron Obstruct Pulmon Dis 2016;11:479-87.

(319) Alshabanat A, Zafari Z, Albanyan O, Dairi M, FitzGerald JM. Asthma and COPD Overlap Syndrome (ACOS): A Systematic Review and Meta Analysis. PLoS One 2015;10(9):e0136065.

(320) Kauppi P, Kupiainen H, Lindqvist A, Tammilehto L, Kilpelainen M, Kinnula VL, et al. Overlap syndrome of asthma and COPD predicts low quality of life. J Asthma 2011 Apr;48(3):279-85.

(321) Andersen H, Lampela P, Nevanlinna A, Saynajakangas O, Keistinen T. High hospital burden in overlap syndrome of asthma and COPD. Clin Respir J 2013 Oct;7(4):342-6.

(322) Kostikas K, Clemens A, Patalano F. The asthma-COPD overlap syndrome: do we really need another syndrome in the already complex matrix of airway disease? Int J Chron Obstruct Pulmon Dis 2016;11:1297-306.

(323) Sordillo J, Raby BA. Gene expression profiling in asthma. Adv Exp Med Biol 2014;795:157-81.

(324) Chen ZH, Kim HP, Ryter SW, Choi AM. Identifying targets for COPD treatment through gene expression analyses. Int J Chron Obstruct Pulmon Dis 2008;3(3):359-70.

(325) Berthon BS, Gibson PG, Wood LG, MacDonald-Wicks LK, Baines KJ. A sputum gene expression signature predicts oral corticosteroid response in asthma. Eur Respir J 2017 Jun;49(6).

(326) Baines KJ, Simpson JL, Wood LG, Scott RJ, Fibbens NL, Powell H, et al. Sputum gene expression signature of 6 biomarkers discriminates asthma inflammatory phenotypes. J Allergy Clin Immunol 2014 Apr;133(4):997-1007.

(327) Baines KJ, Simpson JL, Wood LG, Scott RJ, Gibson PG. Transcriptional phenotypes of asthma defined by gene expression profiling of induced sputum samples. J Allergy Clin Immunol 2011 Jan;127(1):153-60, 160.

(328) Singhania A, Wallington JC, Smith CG, Horowitz D, Staples KJ, Howarth PH, et al. Multitissue Transcriptomics Delineates the Diversity of Airway T Cell Functions in Asthma. Am J Respir Cell Mol Biol 2018 Feb;58(2):261-70.

(329) Szymczak I, Wieczfinska J, Pawliczak R. Molecular Background of miRNA Role in Asthma and COPD: An Updated Insight. Biomed Res Int 2016;2016:7802521.

(330) Lacedonia D, Palladino GP, Foschino-Barbaro MP, Scioscia G, Carpagnano GE. Expression profiling of miRNA-145 and miRNA-338 in serum and sputum of patients with COPD, asthma, and asthma-COPD overlap syndrome phenotype. Int J Chron Obstruct Pulmon Dis 2017;12:1811-7.

(331) Collison A, Mattes J, Plank M, Foster PS. Inhibition of house dust mite-induced allergic airways disease by antagonism of microRNA-145 is comparable to glucocorticoid treatment. J Allergy Clin Immunol 2011 Jul;128(1):160-7.

(332) Tahamtan A, Teymoori-Rad M, Nakstad B, Salimi V. Anti-Inflammatory MicroRNAs and Their Potential for Inflammatory Diseases Treatment. Front Immunol 2018;9:1377.

(333) Kan M, Shumyatcher M, Himes BE. Using omics approaches to understand pulmonary diseases. Respir Res 2017 Aug 3;18(1):149.

(334) Terracciano R, Pelaia G, Preiano M, Savino R. Asthma and COPD proteomics: current approaches and future directions. Proteomics Clin Appl 2015 Feb;9(1-2):203-20.

(335) Gharib SA, Nguyen EV, Lai Y, Plampin JD, Goodlett DR, Hallstrand TS. Induced sputum proteome in healthy subjects and asthmatic patients. J Allergy Clin Immunol 2011 Dec;128(6):1176-84.

(336) Nie W, Xue C, Chen J, Xiu Q. Secretoglobin 1A member 1 (SCGB1A1) +38A/G polymorphism is associated with asthma risk: a meta-analysis. Gene 2013 Oct 10;528(2):304-8.

(337) Lee TH, Jang AS, Park JS, Kim TH, Choi YS, Shin HR, et al. Elevation of S100 calcium binding protein A9 in sputum of neutrophilic inflammation in severe uncontrolled asthma. Ann Allergy Asthma Immunol 2013 Oct;111(4):268-75.

(338) Nambiar S, Bong HS, Gummer J, Trengove R, Moodley Y. Metabolomics in chronic lung diseases. Respirology 2019 Mar 25.

(339) Nobakht MGB, Aliannejad R, Rezaei-Tavirani M, Taheri S, Oskouie AA. The metabolomics of airway diseases, including COPD, asthma and cystic fibrosis. Biomarkers 2015 Feb;20(1):5-16.

(340) Poole A, Urbanek C, Eng C, Schageman J, Jacobson S, O'Connor BP, et al. Dissecting childhood asthma with nasal transcriptomics distinguishes subphenotypes of disease. J Allergy Clin Immunol 2014 Mar;133(3):670-8.

(341) Schaefer N, Li X, Seibold MA, Jarjour NN, Denlinger LC, Castro M, et al. The effect of BPIFA1/SPLUNC1 genetic variation on its expression and function in asthmatic airway epithelium. JCI Insight 2019 Apr 18;4(8).

(342) McDougall CM, Blaylock MG, Douglas JG, Brooker RJ, Helms PJ, Walsh GM. Nasal epithelial cells as surrogates for bronchial epithelial cells in airway inflammation studies. Am J Respir Cell Mol Biol 2008 Nov;39(5):560-8.

(343) Roberts N, Al MR, Francisco D, Kraft M, Chu HW. Comparison of paired human nasal and bronchial airway epithelial cell responses to rhinovirus infection and IL-13 treatment. Clin Transl Med 2018 May 2;7(1):13.

(344) Zhang X, Sebastiani P, Liu G, Schembri F, Zhang X, Dumas YM, et al. Similarities and differences between smoking-related gene expression in nasal and bronchial epithelium. Physiol Genomics 2010 Mar 3;41(1):1-8.

(345) Brugha R, Lowe R, Henderson AJ, Holloway JW, Rakyan V, Wozniak E, et al. DNA methylation profiles between airway epithelium and proxy tissues in children. Acta Paediatr 2017 Dec;106(12):2011-6.

(346) Giovannini-Chami L, Paquet A, Sanfiorenzo C, Pons N, Cazareth J, Magnone V, et al. The "one airway, one disease" concept in light of Th2 inflammation. Eur Respir J 2018 Oct;52(4).

(347) Thavagnanam S, Parker JC, McBrien ME, Skibinski G, Shields MD, Heaney LG. Nasal epithelial cells can act as a physiological surrogate for paediatric asthma studies. PLoS One 2014;9(1):e85802.

(348) Comer DM, Elborn JS, Ennis M. Comparison of nasal and bronchial epithelial cells obtained from patients with COPD. PLoS One 2012;7(3):e32924.

(349) Hargreave FE, Nair P. Point: Is measuring sputum eosinophils useful in the management of severe asthma? Yes. Chest 2011 Jun;139(6):1270-3.

(350) Wan XC, Woodruff PG. Biomarkers in Severe Asthma. Immunol Allergy Clin North Am 2016 Aug;36(3):547-57.

(351) Arron JR, Choy DF, Laviolette M, Kelsen SG, Hatab A, Leigh R, et al. Disconnect between sputum neutrophils and other measures of airway inflammation in asthma. Eur Respir J 2014 Feb;43(2):627-9.

(352) O'Neil SE, Sitkauskiene B, Babusyte A, Krisiukeniene A, Stravinskaite-Bieksiene K, Sakalauskas R, et al. Network analysis of quantitative proteomics on asthmatic bronchi: effects of inhaled glucocorticoid treatment. Respir Res 2011 Sep 22;12:124.

(353) Girodet PO, Nguyen D, Mancini JD, Hundal M, Zhou X, Israel E, et al. Alternative Macrophage Activation Is Increased in Asthma. Am J Respir Cell Mol Biol 2016 Oct;55(4):467-75.

(354) Arora S, Dev K, Agarwal B, Das P, Syed MA. Macrophages: Their role, activation and polarization in pulmonary diseases. Immunobiology 2018 Apr;223(4-5):383-96.

(355) Kay AB. The role of T lymphocytes in asthma. Chem Immunol Allergy 2006;91:59-75.

(356) Barcelo B, Pons J, Ferrer JM, Sauleda J, Fuster A, Agusti AG. Phenotypic characterisation of T-lymphocytes in COPD: abnormal CD4+CD25+ regulatory T-lymphocyte response to tobacco smoking. Eur Respir J 2008 Mar;31(3):555-62.

(357) Zein JG, Dweik RA, Comhair SA, Bleecker ER, Moore WC, Peters SP, et al. Asthma Is More Severe in Older Adults. PLoS One 2015;10(7):e0133490.

(358) Dietz K, de Los Reyes JM, Gollwitzer ES, Chaker AM, Zissler UM, Radmark OP, et al. Age dictates a steroid-resistant cascade of Wnt5a, transglutaminase 2, and leukotrienes in inflamed airways. J Allergy Clin Immunol 2017 Apr;139(4):1343-54.

(359) Fernandez-Boyanapalli R, Goleva E, Kolakowski C, Min E, Day B, Leung DY, et al. Obesity impairs apoptotic cell clearance in asthma. J Allergy Clin Immunol 2013 Apr;131(4):1041-7, 1047.

(360) Melo LC, Silva MA, Calles AC. Obesity and lung function: a systematic review. Einstein (Sao Paulo) 2014 Jan;12(1):120-5.

(361) Marko M, Pawliczak R. Obesity and asthma: risk, control and treatment. Postepy Dermatol Alergol 2018 Dec;35(6):563-71.

(362) Spelta F, Fratta Pasini AM, Cazzoletti L, Ferrari M. Body weight and mortality in COPD: focus on the obesity paradox. Eat Weight Disord 2018 Feb;23(1):15-22.

(363) Zewari S, Vos P, van den Elshout F, Dekhuijzen R, Heijdra Y. Obesity in COPD: Revealed and Unrevealed Issues. COPD 2017 Dec;14(6):663-73.

(364) Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015 Apr 20;43(7):e47.

(365) Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004;3:Article3.

(366) Halling DB, Aracena-Parks P, Hamilton SL. Regulation of voltage-gated Ca2+ channels by calmodulin. Sci STKE 2006 Jan 17;2006(318):er1.

(367) Farber JL. The role of calcium ions in toxic cell injury. Environ Health Perspect 1990 Mar;84:107-11.

(368) Do BH, Nguyen TN, Baba R, Ohbuchi T, Ohkubo JI, Kitamura T, et al. Calmodulin and protein kinases A/G mediate ciliary beat response in the human nasal epithelium. Int Forum Allergy Rhinol 2019 Nov;9(11):1352-9.

(369) Lingappan K. NF-Î°B in Oxidative Stress. Curr Opin Toxicol 2018 Feb;7:81-6.

(370) Lawrence T. The nuclear factor NF-kappaB pathway in inflammation. Cold Spring Harb Perspect Biol 2009 Dec;1(6):a001651.

(371) Ward C, Schlingmann B, Stecenko AA, Guidot DM, Koval M. NF-Î°B inhibitors impair lung epithelial tight junctions in the absence of inflammation. Tissue Barriers 2015;3(1-2):e982424.

(372) Sanders PN, Koval OM, Jaffer OA, Prasad AM, Businga TR, Scott JA, et al. CaMKII is essential for the proasthmatic effects of oxidation. Sci Transl Med 2013 Jul 24;5(195):195ra97.

(373) Sebag SC, Koval OM, Paschke JD, Winters CJ, Jaffer OA, Dworski R, et al. Mitochondrial CaMKII inhibition in airway epithelium protects against allergic asthma. JCI Insight 2017 Feb 9;2(3):e88297.

(374) Lasbury ME, Durant PJ, Liao CP, Lee CH. Effects of decreased calmodulin protein on the survival mechanisms of alveolar macrophages during Pneumocystis pneumonia. Infect Immun 2009 Aug;77(8):3344-54.

(375) Smallwood HS, Shi L, Squier TC. Increases in calmodulin abundance and stabilization of activated inducible nitric oxide synthase mediate bacterial killing in RAW 264.7 macrophages. Biochemistry 2006 Aug 15;45(32):9717-26.

(376) Bai X, Feldman NE, Chmura K, Ovrutsky AR, Su WL, Griffin L, et al. Inhibition of nuclear factor-kappa B activation decreases survival of Mycobacterium tuberculosis in human macrophages. PLoS One 2013;8(4):e61925.

(377) Hill AA, Anderson-Baucum EK, Kennedy AJ, Webb CD, Yull FE, Hasty AH. Activation of NF-Î°B drives the enhanced survival of adipose tissue macrophages in an obesogenic environment. Mol Metab 2015 Oct;4(10):665-77.

(378) Fanta CH. Calcium-channel blockers in prophylaxis and treatment of asthma. Am J Cardiol 1985 Jan 25;55(3):202B-9B.

(379) Boner AL, Antolini I, Andreoli A, De SG, Sette L. Comparison of the effects of inhaled calcium antagonist verapamil, sodium cromoglycate and ipratropium bromide on exercise-induced bronchoconstriction in children with asthma. Eur J Pediatr 1987 Jul;146(4):408-11.

(380) Ann TM, Harman E, Chesrown S, Hendeles L. Efficacy of calcium channel blockers as maintenance therapy for asthma. Br J Clin Pharmacol 2002 Mar;53(3):243-9.

(381) Chiu KY, Li JG, Lin Y. Calcium channel blockers for lung function improvement in asthma: A systematic review and meta-analysis. Ann Allergy Asthma Immunol 2017 Dec;119(6):518-23.

(382) Harman E, Hill M, Pieper JA, Hendeles L. Inhaled verapamil-induced bronchoconstriction in mild asthma. Chest 1991 Jul;100(1):17-22.

(383) Legrand D, Mazurier J. A critical review of the roles of host lactoferrin in immunity. Biometals 2010 Jun;23(3):365-76.

(384) Ellison RT, III, Giehl TJ. Killing of gram-negative bacteria by lactoferrin and lysozyme. J Clin Invest 1991 Oct;88(4):1080-91.

(385) Kruzel ML, Bacsi A, Choudhury B, Sur S, Boldogh I. Lactoferrin decreases pollen antigen-induced allergic airway inflammation in a murine model of asthma. Immunology 2006 Oct;119(2):159-66.

(386) Tang L, Wu JJ, Ma Q, Cui T, Andreopoulos FM, Gil J, et al. Human lactoferrin stimulates skin keratinocyte function and wound re-epithelialization. Br J Dermatol 2010 Jul;163(1):38-47.

(387) Gifford JL, Hunter HN, Vogel HJ. Lactoferricin: a lactoferrin-derived peptide with antimicrobial, antiviral, antitumor and immunological properties. Cell Mol Life Sci 2005 Nov;62(22):2588-98.

(388) Lutaty A, Soboh S, Schif-Zuck S, Zeituni-Timor O, Rostoker R, Podolska MJ, et al. A 17-kDa Fragment of Lactoferrin Associates With the Termination of Inflammation and Peptides Within Promote Resolution. Front Immunol 2018;9:644.

(389) de LA, Tejerina JM, Fierro JF. Interaction of calmodulin with lactoferrin. FEBS Lett 1992 Feb 24;298(2-3):195-8.

(390) Gifford JL, Ishida H, Vogel HJ. Structural characterization of the interaction of human lactoferrin with calmodulin. PLoS One 2012;7(12):e51026.

(391) Nagaoka K, Ito T, Ogino K, Eguchi E, Fujikura Y. Human lactoferrin induces asthmatic symptoms in NC/Nga mice. Physiol Rep 2017 Aug;5(15).

(392) FernÃ¡ndez-Delgado L, Vega-Rioja A, Ventura I, Chamorro C, Aroca R, Prados M, et al. Allergens Induce the Release of Lactoferrin by Neutrophils from Asthmatic Patients. PLoS One 2015;10(10):e0141278.

(393) Vasavda N, Eichholtz T, Takahashi A, Affleck K, Matthews JG, Barnes PJ, et al. Expression of nonmuscle cofilin-1 and steroid responsiveness in severe asthma. J Allergy Clin Immunol 2006 Nov;118(5):1090-6.

(394) Hartl M, Puglisi K, Nist A, Raffeiner P, Bister K. The brain acid-soluble protein 1 (BASP1) interferes with the oncogenic capacity of MYC and its binding to calmodulin. Mol Oncol 2020 Mar;14(3):625-44.

(395) Carpenter B, Hill KJ, Charalambous M, Wagner KJ, Lahiri D, James DI, et al. BASP1 is a transcriptional cosuppressor for the Wilms' tumor suppressor protein WT1. Mol Cell Biol 2004 Jan;24(2):537-49.

(396) Hartl M, Schneider R. A Unique Family of Neuronal Signaling Proteins Implicated in Oncogenesis and Tumor Suppression. Front Oncol 2019;9:289.

(397) Bleich M, Shan Q, Himmerkus N. Calcium regulation of tight junction permeability. Ann N Y Acad Sci 2012 Jul;1258:93-9.

(398) Hallstrand TS, Lai Y, Henderson WR, Jr., Altemeier WA, Gelb MH. Epithelial regulation of eicosanoid production in asthma. Pulm Pharmacol Ther 2012 Dec;25(6):432-7.

(399) Contreras RG, Miller JH, Zamora M, GonzÃ¡lez-Mariscal L, Cereijido M. Interaction of calcium with plasma membrane of epithelial (MDCK) cells during junction formation. Am J Physiol 1992 Aug;263(2 Pt 1):C313-C318.

(400) Postma DS. Gender differences in asthma development and progression. Gend Med 2007;4 Suppl B:S133-S146.

(401) Zaitsu M, Narita S, Lambert KC, Grady JJ, Estes DM, Curran EM, et al. Estradiol activates mast cells via a non-genomic estrogen receptor-alpha and calcium influx. Mol Immunol 2007 Mar;44(8):1977-85.

(402) Herr C, Greulich T, Koczulla RA, Meyer S, Zakharkina T, Branscheidt M, et al. The role of vitamin D in pulmonary disease: COPD, asthma, infection, and cancer. Respir Res 2011 Mar 18;12:31.

(403) Chinellato I, Piazza M, Sandri M, Peroni DG, Cardinale F, Piacentini GL, et al. Serum vitamin D levels and exercise-induced bronchoconstriction in children with asthma. Eur Respir J 2011 Jun;37(6):1366-70.

(404) Shaikh MN, Malapati BR, Gokani R, Patel B, Chatriwala M. Serum Magnesium and Vitamin D Levels as Indicators of Asthma Severity. Pulm Med 2016;2016:1643717.

(405) Poon AH, Mahboub B, Hamid Q. Vitamin D deficiency and severe asthma. Pharmacol Ther 2013 Nov;140(2):148-55.

(406) Mathyssen C, Gayan-Ramirez G, Bouillon R, Janssens W. Vitamin D supplementation in respiratory diseases: evidence from randomized controlled trials. Pol Arch Intern Med 2017 Nov 30;127(11):775-84.

(407) Jolliffe DA, Greenberg L, Hooper RL, Griffiths CJ, Camargo CA, Jr., Kerley CP, et al. Vitamin D supplementation to prevent asthma exacerbations: a systematic review and meta-analysis of individual participant data. Lancet Respir Med 2017 Nov;5(11):881-90.

(408) Heiss A, DuChesne A, Denecke B, GrÃ¶tzinger J, Yamamoto K, RennÃ© T, et al. Structural basis of calcification inhibition by alpha 2-HS glycoprotein/fetuin-A. Formation of colloidal calciprotein particles. J Biol Chem 2003 Apr 11;278(15):13333-41.

(409) Mathews ST, Chellam N, Srinivas PR, Cintron VJ, Leon MA, Goustin AS, et al. Alpha2-HSG, a specific inhibitor of insulin receptor autophosphorylation, interacts with the insulin receptor. Mol Cell Endocrinol 2000 Jun;164(1-2):87-98.

(410) Tsai S, Clemente-Casares X, Zhou AC, Lei H, Ahn JJ, Chan YT, et al. Insulin Receptor-Mediated Stimulation Boosts T Cell Immunity during Inflammation and Infection. Cell Metab 2018 Dec 4;28(6):922-34.

(411) Schwartz SS, Hay DI, Schluckebier SK. Inhibition of calcium phosphate precipitation by human salivary statherin: structure-activity relationships. Calcif Tissue Int 1992 Jun;50(6):511-7.

(412) Kicic A, Hallstrand TS, Sutanto EN, Stevens PT, Kobor MS, Taplin C, et al. Decreased fibronectin production significantly contributes to dysregulated repair of asthmatic epithelium. Am J Respir Crit Care Med 2010 May 1;181(9):889-98.

(413) Chen HY, Lin MH, Chen CC, Shu JC. The expression of fibronectin is significantly suppressed in macrophages to exert a protective effect against Staphylococcus aureus infection. BMC Microbiol 2017 Apr 13;17(1):92.

(414) Baumert J, Schmidt KH, Eitner A, Straube E, Rodel J. Host cell cytokines induced by Chlamydia pneumoniae decrease the expression of interstitial collagens and fibronectin in fibroblasts. Infect Immun 2009 Feb;77(2):867-76.

(415) Bochkov YA, Hanson KM, Keles S, Brockman-Schneider RA, Jarjour NN, Gern JE. Rhinovirus-induced modulation of gene expression in bronchial epithelial cells from subjects with asthma. Mucosal Immunol 2010 Jan;3(1):69-80.

(416) Ge Q, Zeng Q, Tjin G, Lau E, Black JL, Oliver BG, et al. Differential deposition of fibronectin by asthmatic bronchial epithelial cells. Am J Physiol Lung Cell Mol Physiol 2015 Nov 15;309(10):L1093-L1102.

(417) Sherman LA. Binding of soluble fibrin to macrophages. Ann N Y Acad Sci 1983 Jun 27;408:610-20.

(418) Mosher DF, Schad PE. Cross-linking of fibronectin to collagen by blood coagulation Factor XIIIa. J Clin Invest 1979 Sep;64(3):781-7.

(419) Romanic AM, Arleth AJ, Willette RN, Ohlstein EH. Factor XIIIa cross-links lipoprotein(a) with fibrinogen and is present in human atherosclerotic lesions. Circ Res 1998 Aug 10;83(3):264-9.

(420) Sharma S, Zhou X, Thibault DM, Himes BE, Liu A, Szefler SJ, et al. A genome-wide survey of CD4(+) lymphocyte regulatory genetic variants identifies novel asthma genes. J Allergy Clin Immunol 2014 Nov;134(5):1153-62.

(421) Esnault S, Kelly EA, Sorkness RL, Evans MD, Busse WW, Jarjour NN. Airway factor XIII associates with type 2 inflammation and airway obstruction in asthmatic patients. J Allergy Clin Immunol 2016 Mar;137(3):767-73.

(422) Dardik R, Krapp T, Rosenthal E, Loscalzo J, Inbal A. Effect of FXIII on monocyte and fibroblast function. Cell Physiol Biochem 2007;19(1-4):113-20.

(423) Barry EL, Mosher DF. Binding and degradation of blood coagulation factor XIII by cultured fibroblasts. J Biol Chem 1990 Jun 5;265(16):9302-7.

(424) Paye M, Nusgens B, Lapiere CM. Factor XIII of blood coagulation decreases the susceptibility of collagen precursors to proteolysis. Biochim Biophys Acta 1991 Apr 9;1073(3):437-41.

(425) Esnault S, Kelly EA, Sorkness RL, Evans MD, Busse WW, Jarjour NN. Airway factor XIII associates with type 2 inflammation and airway obstruction in asthmatic patients. J Allergy Clin Immunol 2016 Mar;137(3):767-73.

(426) Kuo CS, Pavlidis S, Zhu J, Loza M, Baribaud F, Rowe A, et al. Contribution of airway eosinophils in airway wall remodeling in asthma: Role of MMP-10 and MET. Allergy 2019 Jun;74(6):1102-12.

(427) McMahan RS, Birkland TP, Smigiel KS, Vandivort TC, Rohani MG, Manicone AM, et al. Stromelysin-2 (MMP10) Moderates Inflammation by Controlling Macrophage Activation. J Immunol 2016 Aug 1;197(3):899-909.

(428) Wight TN, Frevert CW, Debley JS, Reeves SR, Parks WC, Ziegler SF. Interplay of extracellular matrix and leukocytes in lung inflammation. Cell Immunol 2017 Feb;312:1-14.

(429) Kuo CS, Pavlidis S, Zhu J, Loza M, Baribaud F, Rowe A, et al. Contribution of airway eosinophils in airway wall remodeling in asthma: Role of MMP-10 and MET. Allergy 2019 Jun;74(6):1102-12.

(430) Bissonnette EY, Madore AM, Chakir J, Laviolette M, Boulet LP, Hamid Q, et al. Fibroblast growth factor-2 is a sputum remodeling biomarker of severe asthma. J Asthma 2014 Mar;51(2):119-26.

(431) Tang DD. Critical role of actin-associated proteins in smooth muscle contraction, cell proliferation, airway hyperresponsiveness and airway remodeling. Respir Res 2015 Oct 30;16:134.

(432) Kang BN, Ha SG, Bahaie NS, Hosseinkhani MR, Ge XN, Blumenthal MN, et al. Regulation of serotonin-induced trafficking and migration of eosinophils. PLoS One 2013;8(1):e54840.

(433) Yin LM, Duan TT, Ulloa L, Yang YQ. Ezrin Orchestrates Signal Transduction in Airway Cells. Rev Physiol Biochem Pharmacol 2018;174:1-23.

(434) Jia M, Yan X, Jiang X, Wu Y, Xu J, Meng Y, et al. Ezrin, a Membrane Cytoskeleton Cross-Linker Protein, as a Marker of Epithelial Damage in Asthma. Am J Respir Crit Care Med 2019 Feb 15;199(4):496-507.

(435) Wang Y, Lin Z, Sun L, Fan S, Huang Z, Zhang D, et al. Akt/Ezrin Tyr353/NF-kappaB pathway regulates EGF-induced EMT and metastasis in tongue squamous cell carcinoma. Br J Cancer 2014 Feb 4;110(3):695-705.

(436) Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. J Clin Invest 2009 Jun;119(6):1420-8.

(437) Pain M, Bermudez O, Lacoste P, Royer PJ, Botturi K, Tissot A, et al. Tissue remodelling in chronic bronchial diseases: from the epithelial to mesenchymal phenotype. Eur Respir Rev 2014 Mar 1;23(131):118-30.

(438) Hackett TL. Epithelial-mesenchymal transition in the pathophysiology of airway remodelling in asthma. Curr Opin Allergy Clin Immunol 2012 Feb;12(1):53-9.

(439) Laoukili J, Perret E, Willems T, Minty A, Parthoens E, Houcine O, et al. IL-13 alters mucociliary differentiation and ciliary beating of human respiratory epithelial cells. J Clin Invest 2001 Dec;108(12):1817-24.

(440) Rouven BB, Pietuch A, Nehls S, Rother J, Janshoff A. Ezrin is a Major Regulator of Membrane Tension in Epithelial Cells. Sci Rep 2015 Oct 5;5:14700.

(441) Miura S, Sato K, Kato-Negishi M, Teshima T, Takeuchi S. Fluid shear triggers microvilli formation via mechanosensitive activation of TRPV6. Nat Commun 2015 Nov 13;6:8871.

(442) Koltzscher M, Neumann C, Konig S, Gerke V. Ca2+-dependent binding and activation of dormant ezrin by dimeric S100P. Mol Biol Cell 2003 Jun;14(6):2372-84.

(443) Austermann J, Nazmi AR, Heil A, Fritz G, Kolinski M, Filipek S, et al. Generation and characterization of a novel, permanently active S100P mutant. Biochim Biophys Acta 2009 Jun;1793(6):1078-85.

(444) Efstathiou JA, Noda M, Rowan A, Dixon C, Chinery R, Jawhari A, et al. Intestinal trefoil factor controls the expression of the adenomatous polyposis coli-catenin and the E-cadherin-catenin complexes in human colon carcinoma cells. Proc Natl Acad Sci U S A 1998 Mar 17;95(6):3122-7.

(445) Meyer zum BD, Hoschutzky H, Tauber R, Huber O. Molecular mechanisms involved in TFF3 peptide-mediated modulation of the E-cadherin/catenin cell adhesion complex. Peptides 2004 May;25(5):873-83.

(446) Durer U, Hartig R, Bang S, Thim L, Hoffmann W. TFF3 and EGF induce different migration patterns of intestinal epithelial cells in vitro and trigger increased internalization of E-cadherin. Cell Physiol Biochem 2007;20(5):329-46.

(447) Buda A, Jepson MA, Pignatelli M. Regulatory function of trefoil peptides (TFF) on intestinal cell junctional complexes. Cell Commun Adhes 2012 Oct;19(5-6):63-8.

(448) Durer U, Hartig R, Bang S, Thim L, Hoffmann W. TFF3 and EGF induce different migration patterns of intestinal epithelial cells in vitro and trigger increased internalization of E-cadherin. Cell Physiol Biochem 2007;20(5):329-46.

(449) Meyer zum BD, Tauber R, Huber O. TFF3-peptide increases transepithelial resistance in epithelial cells by modulating claudin-1 and -2 expression. Peptides 2006 Dec;27(12):3383-90.

(450) Jackson NE, Wang HW, Bryant KJ, McNeil HP, Husain A, Liu K, et al. Alternate mRNA splicing in multiple human tryptase genes is predicted to regulate tetramer formation. J Biol Chem 2008 Dec 5;283(49):34178-87.

(451) Dougherty RH, Sidhu SS, Raman K, Solon M, Solberg OD, Caughey GH, et al. Accumulation of intraepithelial mast cells with a unique protease phenotype in T(H)2-high asthma. J Allergy Clin Immunol 2010 May;125(5):1046-53.

(452) Dorscheid DR, Wojcik KR, Sun S, Marroquin B, White SR. Apoptosis of airway epithelial cells induced by corticosteroids. Am J Respir Crit Care Med 2001 Nov 15;164(10 Pt 1):1939-47.

(453) Vignola AM, Chiappara G, Siena L, Bruno A, Gagliardo R, Merendino AM, et al. Proliferation and activation of bronchial epithelial cells in corticosteroid-dependent asthma. J Allergy Clin Immunol 2001 Nov;108(5):738-46.

(454) Morishima Y, Nomura A, Uchida Y, Noguchi Y, Sakamoto T, Ishii Y, et al. Triggering the induction of myofibroblast and fibrogenesis by airway epithelial shedding. Am J Respir Cell Mol Biol 2001 Jan;24(1):1-11.

(455) Abrahamson M, Barrett AJ, Salvesen G, Grubb A. Isolation of six cysteine proteinase inhibitors from human urine. Their physicochemical and enzyme kinetic properties and concentrations in biological fluids. J Biol Chem 1986 Aug 25;261(24):11282-9.

(456) Blaydon DC, Nitoiu D, Eckl KM, Cabral RM, Bland P, Hausser I, et al. Mutations in CSTA, encoding Cystatin A, underlie exfoliative ichthyosis and reveal a role for this protease inhibitor in cell-cell adhesion. Am J Hum Genet 2011 Oct 7;89(4):564-71.

(457) Oh SS, Park S, Lee KW, Madhi H, Park SG, Lee HG, et al. Extracellular cystatin SN and cathepsin B prevent cellular senescence by inhibiting abnormal glycogen accumulation. Cell Death Dis 2017 Apr 6;8(4):e2729.

(458) Chen J, Hu C, Pan P. Extracellular Vesicle MicroRNA Transfer in Lung Diseases. Front Physiol 2017;8:1028.

(459) Huang C, Liu XJ, QunZhou, Xie J, Ma TT, Meng XM, et al. MiR-146a modulates macrophage polarization by inhibiting Notch1 pathway in RAW264.7 macrophages. Int Immunopharmacol 2016 Mar;32:46-54.

(460) Levanen B, Bhakta NR, Torregrosa PP, Barbeau R, Hiltbrunner S, Pollack JL, et al. Altered microRNA profiles in bronchoalveolar lavage fluid exosomes in asthmatic patients. J Allergy Clin Immunol 2013 Mar;131(3):894-903.

(461) Oglesby IK, McElvaney NG, Greene CM. MicroRNAs in inflammatory lung disease--master regulators or target practice? Respir Res 2010 Oct 28;11:148.

(462) Solberg OD, Ostrin EJ, Love MI, Peng JC, Bhakta NR, Hou L, et al. Airway epithelial miRNA expression is altered in asthma. Am J Respir Crit Care Med 2012 Nov 15;186(10):965-74.

(463) Newcomb DC, Cephus JY, Boswell MG, Fahrenholz JM, Langley EW, Feldman AS, et al. Estrogen and progesterone decrease let-7f microRNA expression and increase IL-23/IL-23 receptor signaling and IL-17A production in patients with severe asthma. J Allergy Clin Immunol 2015 Oct;136(4):1025-34.

(464) Tsitsiou E, Williams AE, Moschos SA, Patel K, Rossios C, Jiang X, et al. Transcriptome analysis shows activation of circulating CD8+ T cells in patients with severe asthma. J Allergy Clin Immunol 2012 Jan;129(1):95-103.

(465) Comer BS, Camoretti-Mercado B, Kogut PC, Halayko AJ, Solway J, Gerthoffer WT. MicroRNA-146a and microRNA-146b expression and anti-inflammatory function in human airway smooth muscle. Am J Physiol Lung Cell Mol Physiol 2014 Nov 1;307(9):L727-L734.

(466) Jimenez-Morales S, Gamboa-Becerra R, Baca V, Del Rio-Navarro BE, Lopez-Ley DY, Velazquez-Cruz R, et al. MiR-146a polymorphism is associated with asthma but not with systemic lupus erythematosus and juvenile rheumatoid arthritis in Mexican patients. Tissue Antigens 2012 Oct;80(4):317-21.

(467) Nakano T, Inoue Y, Shimojo N, Yamaide F, Morita Y, Arima T, et al. Lower levels of hsa-mir-15a, which decreases VEGFA, in the CD4+ T cells of pediatric patients with asthma. J Allergy Clin Immunol 2013 Nov;132(5):1224-7.

(468) Chen Y, Qiao J. Protein-protein interaction network analysis and identifying regulation microRNAs in asthmatic children. Allergol Immunopathol (Madr) 2015 Nov;43(6):584-92.

(469) Kishore A, Borucka J, Petrkova J, Petrek M. Novel insights into miRNA in lung and heart inflammatory diseases. Mediators Inflamm 2014;2014:259131.

(470) Ma X, Becker Buscaglia LE, Barker JR, Li Y. MicroRNAs in NF-kappaB signaling. J Mol Cell Biol 2011 Jun;3(3):159-66.

(471) Wu J, Ding J, Yang J, Guo X, Zheng Y. MicroRNA Roles in the Nuclear Factor Kappa B Signaling Pathway in Cancer. Front Immunol 2018;9:546.

(472) Zhao Y, Bhattacharjee S, Jones BM, Hill J, Dua P, Lukiw WJ. Regulation of neurotropic signaling by the inducible, NF-κB-sensitive miRNA-125b in Alzheimer's disease (AD) and in primary human neuronal-glial (HNG) cells. Mol Neurobiol 2014 Aug;50(1):97-106.

(473) Yang Y, Wang JK. The functional analysis of MicroRNAs involved in NF-kappaB signaling. Eur Rev Med Pharmacol Sci 2016 May;20(9):1764-74.

(474) Banerjee A, Schambach F, DeJong CS, Hammond SM, Reiner SL. Micro-RNA-155 inhibits IFN-gamma signaling in CD4+ T cells. Eur J Immunol 2010 Jan;40(1):225-31.

(475) Yoshimura A, Naka T, Kubo M. SOCS proteins, cytokine signalling and immune regulation. Nat Rev Immunol 2007 Jun;7(6):454-65.

(476) Rodriguez A, Vigorito E, Clare S, Warren MV, Couttet P, Soond DR, et al. Requirement of bic/microRNA-155 for normal immune function. Science 2007 Apr 27;316(5824):608-11.

(477) Malmhall C, Alawieh S, Lu Y, Sjostrand M, Bossios A, Eldh M, et al. MicroRNA-155 is essential for T(H)2-mediated allergen-induced eosinophilic inflammation in the lung. J Allergy Clin Immunol 2014 May;133(5):1429-38, 1438.

(478) Okoye IS, Czieso S, Ktistaki E, Roderick K, Coomes SM, Pelly VS, et al. Transcriptomics identified a critical role for Th2 cell-intrinsic miR-155 in mediating allergy and antihelminth immunity. Proc Natl Acad Sci U S A 2014 Jul 29;111(30):E3081-E3090.

(479) Tili E, Michaille JJ, Cimino A, Costinean S, Dumitru CD, Adair B, et al. Modulation of miR-155 and miR-125b levels following lipopolysaccharide/TNF-alpha stimulation and their possible roles in regulating the response to endotoxin shock. J Immunol 2007 Oct 15;179(8):5082-9.

(480) Cho JH, Gelinas R, Wang K, Etheridge A, Piper MG, Batte K, et al. Systems biology of interstitial lung diseases: integration of mRNA and microRNA expression changes. BMC Med Genomics 2011 Jan 17;4:8.

(481) Zhong X, Chung AC, Chen HY, Meng XM, Lan HY. Smad3-mediated upregulation of miR-21 promotes renal fibrosis. J Am Soc Nephrol 2011 Sep;22(9):1668-81.

(482) Yamada M, Kubo H, Ota C, Takahashi T, Tando Y, Suzuki T, et al. The increase of microRNA-21 during lung fibrosis and its contribution to epithelial-mesenchymal transition in pulmonary epithelial cells. Respir Res 2013 Sep 24;14:95.

(483) Gong C, Nie Y, Qu S, Liao JY, Cui X, Yao H, et al. miR-21 induces myofibroblast differentiation and promotes the malignant progression of breast phyllodes tumors. Cancer Res 2014 Aug 15;74(16):4341-52.

(484) Levanen B, Bhakta NR, Torregrosa PP, Barbeau R, Hiltbrunner S, Pollack JL, et al. Altered microRNA profiles in bronchoalveolar lavage fluid exosomes in asthmatic patients. J Allergy Clin Immunol 2013 Mar;131(3):894-903.

(485) Liu F, Zheng S, Liu T, Liu Q, Liang M, Li X, et al. MicroRNA-21 promotes the proliferation and inhibits apoptosis in Eca109 via activating ERK1/2/MAPK pathway. Mol Cell Biochem 2013 Sep;381(1-2):115-25.

(486) Murugaiyan G, Garo LP, Weiner HL. MicroRNA-21, T helper lineage and autoimmunity. Oncotarget 2015;6(12):9644-5.

(487) Samanta D, Gonzalez AL, Nagathihalli N, Ye F, Carbone DP, Datta PK. Smoking attenuates transforming growth factor-Î²-mediated tumor suppression function through downregulation of Smad3 in lung cancer. Cancer Prev Res (Phila) 2012 Mar;5(3):453-63.

(488) Du W, Tang H, Lei Z, Zhu J, Zeng Y, Liu Z, et al. miR-335-5p inhibits TGF-Î²1-induced epithelial-mesenchymal transition in non-small cell lung cancer via ROCK1. Respir Res 2019 Oct 21;20(1):225.

(489) Shi HZ, Deng JM, Xu H, Nong ZX, Xiao CQ, Liu ZM, et al. Effect of inhaled interleukin-4 on airway hyperreactivity in asthmatics. Am J Respir Crit Care Med 1998 Jun;157(6 Pt 1):1818-21.

(490) Pope SM, Brandt EB, Mishra A, Hogan SP, Zimmermann N, Matthaei KI, et al. IL-13 induces eosinophil recruitment into the lung by an IL-5- and eotaxin-dependent mechanism. J Allergy Clin Immunol 2001 Oct;108(4):594-601.

(491) Brinckmann R, Topp MS, Zalan I, Heydeck D, Ludwig P, Kuhn H, et al. Regulation of 15-lipoxygenase expression in lung epithelial cells by interleukin-4. Biochem J 1996 Aug 15;318 (Pt 1):305-12.

(492) Chaitidis P, O'Donnell V, Kuban RJ, Bermudez-Fajardo A, Ungethuem U, Kuhn H. Gene expression alterations of human peripheral blood monocytes induced by medium-term treatment with the TH2-cytokines interleukin-4 and -13. Cytokine 2005 Jun 21;30(6):366-77.

(493) Mandarapu R, Ajumeera R, Venkatesan V, Prakhya BM. Proliferation and TH1/TH2 cytokine production in human peripheral blood mononuclear cells after treatment with cypermethrin and mancozeb in vitro. J Toxicol 2014;2014:308286.

(494) Potaczek DP, Michel S, Sharma V, Zeilinger S, Vogelberg C, von BA, et al. Different FCER1A polymorphisms influence IgE levels in asthmatics and non-asthmatics. Pediatr Allergy Immunol 2013 Aug;24(5):441-9.

(495) Namgaladze D, Snodgrass RG, Angioni C, Grossmann N, Dehne N, Geisslinger G, et al. AMP-activated protein kinase suppresses arachidonate 15-lipoxygenase expression in interleukin 4-polarized human macrophages. J Biol Chem 2015 Oct 2;290(40):24484-94.

(496) Xu B, Bhattacharjee A, Roy B, Xu HM, Anthony D, Frank DA, et al. Interleukin-13 induction of 15-lipoxygenase gene expression requires p38 mitogen-activated protein kinase-mediated serine 727 phosphorylation of Stat1 and Stat3. Mol Cell Biol 2003 Jun;23(11):3918-28.

(497) Zhao J, Maskrey B, Balzar S, Chibana K, Mustovich A, Hu H, et al. Interleukin-13-induced MUC5AC is regulated by 15-lipoxygenase 1 pathway in human bronchial epithelial cells. Am J Respir Crit Care Med 2009 May 1;179(9):782-90.

(498) Hallstrand TS, Debley JS, Farin FM, Henderson WR, Jr. Role of MUC5AC in the pathogenesis of exercise-induced bronchoconstriction. J Allergy Clin Immunol 2007 May;119(5):1092-8.

(499) Fahy JV, Dickey BF. Airway mucus function and dysfunction. N Engl J Med 2010 Dec 2;363(23):2233-47.

(500) Evans CM, Kim K, Tuvim MJ, Dickey BF. Mucus hypersecretion in asthma: causes and effects. Curr Opin Pulm Med 2009 Jan;15(1):4-11.

(501) Park JA, Tschumperlin DJ. Chronic intermittent mechanical stress increases MUC5AC protein expression. Am J Respir Cell Mol Biol 2009 Oct;41(4):459-66.

(502) Shi HZ, Xiao CQ, Zhong D, Qin SM, Liu Y, Liang GR, et al. Effect of inhaled interleukin-5 on airway hyperreactivity and eosinophilia in asthmatics. Am J Respir Crit Care Med 1998 Jan;157(1):204-9.

(503) Clutterbuck EJ, Hirst EM, Sanderson CJ. Human interleukin-5 (IL-5) regulates the production of eosinophils in human bone marrow cultures: comparison and interaction with IL-1, IL-3, IL-6, and GMCSF. Blood 1989 May 1;73(6):1504-12.

(504) Sanderson CJ. Interleukin-5, eosinophils, and disease. Blood 1992 Jun 15;79(12):3101-9.

(505) Mukherjee M, Bulir DC, Radford K, Kjarsgaard M, Huang CM, Jacobsen EA, et al. Sputum autoantibodies in patients with severe eosinophilic asthma. J Allergy Clin Immunol 2018 Apr;141(4):1269-79.

(506) Drick N, Seeliger B, Welte T, Fuge J, Suhling H. Anti-IL-5 therapy in patients with severe eosinophilic asthma - clinical efficacy and possible criteria for treatment response. BMC Pulm Med 2018 Jul 18;18(1):119.

(507) Kraneveld AD, Folkerts G, Van Oosterhout AJ, Nijkamp FP. Airway hyperresponsiveness: first eosinophils and then neuropeptides. Int J Immunopharmacol 1997 Sep;19(9-10):517-27.

(508) James A, Daham K, Backman L, Brunnstrom A, Tingvall T, Kumlin M, et al. The influence of aspirin on release of eoxin C4, leukotriene C4 and 15-HETE, in eosinophilic granulocytes isolated from patients with asthma. Int Arch Allergy Immunol 2013;162(2):135-42.

(509) Brinckmann R, Schnurr K, Heydeck D, Rosenbach T, Kolde G, Kuhn H. Membrane translocation of 15-lipoxygenase in hematopoietic cells is calcium-dependent and activates the oxygenase activity of the enzyme. Blood 1998 Jan 1;91(1):64-74.

(510) Walther M, Wiesner R, Kuhn H. Investigations into calcium-dependent membrane association of 15-lipoxygenase-1. Mechanistic roles of surface-exposed hydrophobic amino acids and calcium. J Biol Chem 2004 Jan 30;279(5):3717-25.

(511) Belkner J, Wiesner R, Kuhn H, Lankin VZ. The oxygenation of cholesterol esters by the reticulocyte lipoxygenase. FEBS Lett 1991 Feb 11;279(1):110-4.

(512) Kuhn H. Biosynthesis, metabolization and biological importance of the primary 15-lipoxygenase metabolites 15-hydro(pero)XY-5Z,8Z,11Z,13E-eicosatetraenoic acid and 13-hydro(pero)XY-9Z,11E-octadecadienoic acid. Prog Lipid Res 1996 Sep;35(3):203-26.

(513) Ivanov I, Kuhn H, Heydeck D. Structural and functional biology of arachidonic acid 15-lipoxygenase-1 (ALOX15). Gene 2015 Nov 15;573(1):1-32.

(514) Kuhn H, Salzmann-Reinhardt U, Ludwig P, Ponicke K, Schewe T, Rapoport S. The stoichiometry of oxygen uptake and conjugated diene formation during the dioxygenation of linoleic acid by the pure reticulocyte lipoxygenase. Evidence for aerobic hydroperoxidase activity. Biochim Biophys Acta 1986 Apr 15;876(2):187-93.

(515) Feltenmark S, Gautam N, Brunnstrom A, Griffiths W, Backman L, Edenius C, et al. Eoxins are proinflammatory arachidonic acid metabolites produced via the 15-lipoxygenase-1 pathway in human eosinophils and mast cells. Proc Natl Acad Sci U S A 2008 Jan 15;105(2):680-5.

(516) Kivity S, Argaman A, Onn A, Shwartz Y, Man A, Greif J, et al. Eosinophil influx into the airways in patients with exercise-induced asthma. Respir Med 2000 Dec;94(12):1200-5.

(517) Nadel JA, Conrad DJ, Ueki IF, Schuster A, Sigal E. Immunocytochemical localization of arachidonate 15-lipoxygenase in erythrocytes, leukocytes, and airway cells. J Clin Invest 1991 Apr;87(4):1139-45.

(518) Walters JLH, De Iuliis GN, Dun MD, Aitken RJ, McLaughlin EA, Nixon B, et al. Pharmacological inhibition of arachidonate 15-lipoxygenase protects human spermatozoa against oxidative stress. Biol Reprod 2018 Jun 1;98(6):784-94.

(519) Lukiw WJ. NF-small ka, CyrillicB-regulated micro RNAs (miRNAs) in primary human brain cells. Exp Neurol 2012 Jun;235(2):484-90.

(520) Liu Z, Chen X, Wu Q, Song J, Wang L, Li G. miR-125b inhibits goblet cell differentiation in allergic airway inflammation by targeting SPDEF. Eur J Pharmacol 2016 Jul 5;782:14-20.

(521) Zhao J, Minami Y, Etling E, Coleman JM, Lauder SN, Tyrrell V, et al. Preferential Generation of 15-HETE-PE Induced by IL-13 Regulates Goblet Cell Differentiation in Human Airway Epithelial Cells. Am J Respir Cell Mol Biol 2017 Dec;57(6):692-701.

(522) Shankaranarayanan P, Chaitidis P, Kuhn H, Nigam S. Acetylation by histone acetyltransferase CREB-binding protein/p300 of STAT6 is required for transcriptional activation of the 15-lipoxygenase-1 gene. J Biol Chem 2001 Nov 16;276(46):42753-60.

(523) Ou Y, Wang SJ, Li D, Chu B, Gu W. Activation of SAT1 engages polyamine metabolism with p53-mediated ferroptotic responses. Proc Natl Acad Sci U S A 2016 Nov 1;113(44):E6806-E6812.

(524) O'Connor MJ, Zimmermann H, Nielsen S, Bernard HU, Kouzarides T. Characterization of an E1A-CBP interaction defines a novel transcriptional adapter motif (TRAM) in CBP/p300. J Virol 1999 May;73(5):3574-81.

(525) Arany Z, Huang LE, Eckner R, Bhattacharya S, Jiang C, Goldberg MA, et al. An essential role for p300/CBP in the cellular response to hypoxia. Proc Natl Acad Sci U S A 1996 Nov 12;93(23):12969-73.

(526) Uderhardt S, Ackermann JA, Fillep T, Hammond VJ, Willeit J, Santer P, et al. Enzymatic lipid oxidation by eosinophils propagates coagulation, hemostasis, and thrombotic disease. J Exp Med 2017 Jul 3;214(7):2121-38.

(527) Nemerson Y. The phospholipid requirement of tissue factor in blood coagulation. J Clin Invest 1968 Jan;47(1):72-80.

(528) James A, Daham K, Backman L, Brunnstrom A, Tingvall T, Kumlin M, et al. The influence of aspirin on release of eoxin C4, leukotriene C4 and 15-HETE, in eosinophilic granulocytes isolated from patients with asthma. Int Arch Allergy Immunol 2013;162(2):135-42.

(529) Zhao J, Maskrey B, Balzar S, Chibana K, Mustovich A, Hu H, et al. Interleukin-13-induced MUC5AC is regulated by 15-lipoxygenase 1 pathway in human bronchial epithelial cells. Am J Respir Crit Care Med 2009 May 1;179(9):782-90.

(530) Chattopadhyay R, Dyukova E, Singh NK, Ohba M, Mobley JA, Rao GN. Vascular endothelial tight junctions and barrier function are disrupted by 15(S)-hydroxyeicosatetraenoic acid partly via protein kinase C epsilon-mediated zona occludens-1 phosphorylation at threonine 770/772. J Biol Chem 2014 Feb 7;289(6):3148-63.

(531) Kundumani-Sridharan V, Dyukova E, Hansen DE, III, Rao GN. 12/15-Lipoxygenase mediates high-fat diet-induced endothelial tight junction disruption and monocyte transmigration: a new role for 15(S)-hydroxyeicosatetraenoic acid in endothelial cell dysfunction. J Biol Chem 2013 May 31;288(22):15830-42.

(532) Georas SN, Rezaee F. Epithelial barrier function: at the front line of asthma immunology and allergic airway inflammation. J Allergy Clin Immunol 2014 Sep;134(3):509-20.

(533) Schneeberger EE, Lynch RD. The tight junction: a multifunctional complex. Am J Physiol Cell Physiol 2004 Jun;286(6):C1213-C1228.

(534) Shin K, Fogg VC, Margolis B. Tight junctions and cell polarity. Annu Rev Cell Dev Biol 2006;22:207-35.

(535) Rezaee F, DeSando SA, Ivanov AI, Chapman TJ, Knowlden SA, Beck LA, et al. Sustained protein kinase D activation mediates respiratory syncytial virus-induced airway barrier disruption. J Virol 2013 Oct;87(20):11088-95.

(536) Dietz K, de Los Reyes JM, Gollwitzer ES, Chaker AM, Zissler UM, Radmark OP, et al. Age dictates a steroid-resistant cascade of Wnt5a, transglutaminase 2, and leukotrienes in inflamed airways. J Allergy Clin Immunol 2017 Apr;139(4):1343-54.

(537) Choy DF, Modrek B, Abbas AR, Kummerfeld S, Clark HF, Wu LC, et al. Gene expression patterns of Th2 inflammation and intercellular communication in asthmatic airways. J Immunol 2011 Feb 1;186(3):1861-9.

(538) Singhania A, Rupani H, Jayasekera N, Lumb S, Hales P, Gozzard N, et al. Altered Epithelial Gene Expression in Peripheral Airways of Severe Asthma. PLoS One 2017;12(1):e0168680.

(539) Mikels AJ, Nusse R. Purified Wnt5a protein activates or inhibits beta-catenin-TCF signaling depending on receptor context. PLoS Biol 2006 Apr;4(4):e115.

(540) Vuga LJ, Ben-Yehudah A, Kovkarova-Naumovski E, Oriss T, Gibson KF, Feghali-Bostwick C, et al. WNT5A is a regulator of fibroblast proliferation and resistance to apoptosis. Am J Respir Cell Mol Biol 2009 Nov;41(5):583-9.

(541) Rieger ME, Zhou B, Solomon N, Sunohara M, Li C, Nguyen C, et al. p300/beta-Catenin Interactions Regulate Adult Progenitor Cell Differentiation Downstream of WNT5a/Protein Kinase C (PKC). J Biol Chem 2016 Mar 18;291(12):6569-82.

(542) Michalik M, Wojcik-Pszczola K, Paw M, Wnuk D, Koczurkiewicz P, Sanak M, et al. Fibroblast-to-myofibroblast transition in bronchial asthma. Cell Mol Life Sci 2018 Nov;75(21):3943-61.

(543) Eap R, Jacques E, Semlali A, Plante S, Chakir J. Cysteinyl leukotrienes regulate TGF-beta(1) and collagen production by bronchial fibroblasts obtained from asthmatic subjects. Prostaglandins Leukot Essent Fatty Acids 2012 Mar;86(3):127-33.

(544) Kitamura T, Naganuma T, Abe K, Nakahara K, Ohno Y, Kihara A. Substrate specificity, plasma membrane localization, and lipid modification of the aldehyde dehydrogenase ALDH3B1. Biochim Biophys Acta 2013 Aug;1831(8):1395-401.

(545) Marchitti SA, Brocker C, Orlicky DJ, Vasiliou V. Molecular characterization, expression analysis, and role of ALDH3B1 in the cellular protection against oxidative stress. Free Radic Biol Med 2010 Nov 15;49(9):1432-43.

(546) Nguyen T, Nioi P, Pickett CB. The Nrf2-antioxidant response element signaling pathway and its activation by oxidative stress. J Biol Chem 2009 May 15;284(20):13291-5.

(547) Ostroukhova M, Goplen N, Karim MZ, Michalec L, Guo L, Liang Q, et al. The role of low-level lactate production in airway inflammation in asthma. Am J Physiol Lung Cell Mol Physiol 2012 Feb 1;302(3):L300-L307.

(548) Perl A, Hanczko R, Telarico T, Oaks Z, Landas S. Oxidative stress, inflammation and carcinogenesis are controlled through the pentose phosphate pathway by transaldolase. Trends Mol Med 2011 Jul;17(7):395-403.

(549) Banki K, Hutter E, Colombo E, Gonchoroff NJ, Perl A. Glutathione levels and sensitivity to apoptosis are regulated by changes in transaldolase expression. J Biol Chem 1996 Dec 20;271(51):32994-3001.

(550) Wark PA, Johnston SL, Moric I, Simpson JL, Hensley MJ, Gibson PG. Neutrophil degranulation and cell lysis is associated with clinical severity in virus-induced asthma. Eur Respir J 2002 Jan;19(1):68-75.

(551) O'Neil SE, Sitkauskiene B, Babusyte A, Krisiukeniene A, Stravinskaite-Bieksiene K, Sakalauskas R, et al. Network analysis of quantitative proteomics on asthmatic bronchi: effects of inhaled glucocorticoid treatment. Respir Res 2011 Sep 22;12:124.

(552) Zhao L, Li W, Zhou Y, Zhang Y, Huang S, Xu X, et al. The overexpression and nuclear translocation of Trx-1 during hypoxia confers on HepG2 cells resistance to DDP, and GL-V9 reverses the resistance by suppressing the Trx-1/Ref-1 axis. Free Radic Biol Med 2015 May;82:29-41.

(553) Padoan M, Pozzato V, Simoni M, Zedda L, Milan G, Bononi I, et al. Long-term follow-up of toluene diisocyanate-induced asthma. Eur Respir J 2003 Apr;21(4):637-40.

(554) Gordon T, Sheppard D, McDonald DM, Distefano S, Scypinski L. Airway hyperresponsiveness and inflammation induced by toluene diisocyanate in guinea pigs. Am Rev Respir Dis 1985 Nov;132(5):1106-12.

(555) Waller JA. Injury as disease. Accid Anal Prev 1987 Feb;19(1):13-20.

(556) Vasileiadis I, Alevrakis E, Ampelioti S, Vagionas D, Rovina N, Koutsoukou A. Acid-Base Disturbances in Patients with Asthma: A Literature Review and Comments on Their Pathophysiology. J Clin Med 2019 Apr 25;8(4).

(557) Perros F, Hoogsteden HC, Coyle AJ, Lambrecht BN, Hammad H. Blockade of CCR4 in a humanized model of asthma reveals a critical role for DC-derived CCL17 and CCL22 in attracting Th2 cells and inducing airway inflammation. Allergy 2009 Jul;64(7):995-1002.

(558) Ait YS, Azzaoui I, Everaere L, Vorng H, Chenivesse C, Marquillies P, et al. CCL17 production by dendritic cells is required for NOD1-mediated exacerbation of allergic asthma. Am J Respir Crit Care Med 2014 Apr 15;189(8):899-908.

(559) Staples KJ, Hinks TS, Ward JA, Gunn V, Smith C, Djukanovic R. Phenotypic characterization of lung macrophages in asthmatic patients: overexpression of CCL17. J Allergy Clin Immunol 2012 Dec;130(6):1404-12.

(560) Ghebre MA, Pang PH, Diver S, Desai D, Bafadhel M, Haldar K, et al. Biological exacerbation clusters demonstrate asthma and chronic obstructive pulmonary disease overlap with distinct mediator and microbiome profiles. J Allergy Clin Immunol 2018 Jun;141(6):2027-36.

(561) Garcia G, Godot V, Humbert M. New chemokine targets for asthma therapy. Curr Allergy Asthma Rep 2005 Mar;5(2):155-60.

(562) Belperio JA, Dy M, Murray L, Burdick MD, Xue YY, Strieter RM, et al. The role of the Th2 CC chemokine ligand CCL17 in pulmonary fibrosis. J Immunol 2004 Oct 1;173(7):4692-8.

(563) Himes BE, Koziol-White C, Johnson M, Nikolos C, Jester W, Klanderman B, et al. Vitamin D Modulates Expression of the Airway Smooth Muscle Transcriptome in Fatal Asthma. PLoS One 2015;10(7):e0134057.

(564) Joubert P, Lajoie-Kadoch S, Labonte I, Gounni AS, Maghni K, Wellemans V, et al. CCR3 expression and function in asthmatic airway smooth muscle cells. J Immunol 2005 Aug 15;175(4):2702-8.

(565) Agrawal S, Townley RG. Role of periostin, FENO, IL-13, lebrikzumab, other IL-13 antagonist and dual IL-4/IL-13 antagonist in asthma. Expert Opin Biol Ther 2014 Feb;14(2):165-81.

(566) Larose MC, Chakir J, Archambault AS, Joubert P, Provost V, Laviolette M, et al. Correlation between CCL26 production by human bronchial epithelial cells and airway eosinophils: Involvement in patients with severe eosinophilic asthma. J Allergy Clin Immunol 2015 Oct;136(4):904-13.

(567) Gerber BO, Zanni MP, Uguccioni M, Loetscher M, Mackay CR, Pichler WJ, et al. Functional expression of the eotaxin receptor CCR3 in T lymphocytes co-localizing with eosinophils. Curr Biol 1997 Nov 1;7(11):836-43.

(568) Petkovic V, Moghini C, Paoletti S, Uguccioni M, Gerber B. Eotaxin-3/CCL26 is a natural antagonist for CC chemokine receptors 1 and 5. A human chemokine with a regulatory role. J Biol Chem 2004 May 28;279(22):23357-63.

(569) Conroy DM, Williams TJ. Eotaxin and the attraction of eosinophils to the asthmatic lung. Respir Res 2001;2(3):150-6.

(570) Ogilvie P, Paoletti S, Clark-Lewis I, Uguccioni M. Eotaxin-3 is a natural antagonist for CCR2 and exerts a repulsive effect on human monocytes. Blood 2003 Aug 1;102(3):789-94.

(571) Ravensberg AJ, Ricciardolo FL, van SA, Rabe KF, Sterk PJ, Hiemstra PS, et al. Eotaxin-2 and eotaxin-3 expression is associated with persistent eosinophilic bronchial inflammation in patients with asthma after allergen challenge. J Allergy Clin Immunol 2005 Apr;115(4):779-85.

(572) Kohan M, Puxeddu I, Reich R, Levi-Schaffer F, Berkman N. Eotaxin-2/CCL24 and eotaxin-3/CCL26 exert differential profibrogenic effects on human lung fibroblasts. Ann Allergy Asthma Immunol 2010 Jan;104(1):66-72.

(573) Provost V, Larose MC, Langlois A, Rola-Pleszczynski M, Flamand N, Laviolette M. CCL26/eotaxin-3 is more effective to induce the migration of eosinophils of asthmatics than CCL11/eotaxin-1 and CCL24/eotaxin-2. J Leukoc Biol 2013 Aug;94(2):213-22.

(574) Erdei A, Sandor N, Macsik-Valent B, Lukacsi S, Kremlitzka M, Bajtay Z. The versatile functions of complement C3-derived ligands. Immunol Rev 2016 Nov;274(1):127-40.

(575) Khirwadkar K, Zilow G, Oppermann M, Kabelitz D, Rother K. Interleukin-4 augments production of the third complement component by the alveolar epithelial cell line A549. Int Arch Allergy Immunol 1993;100(1):35-41.

(576) Weiler JM, Edens RE, Bell CS, Gleich GJ. Eosinophil granule cationic proteins regulate the classical pathway of complement. Immunology 1995 Feb;84(2):213-9.

(577) Gharib SA, Nguyen EV, Lai Y, Plampin JD, Goodlett DR, Hallstrand TS. Induced sputum proteome in healthy subjects and asthmatic patients. J Allergy Clin Immunol 2011 Dec;128(6):1176-84.

(578) Nakano Y, Morita S, Kawamoto A, Suda T, Chida K, Nakamura H. Elevated complement C3a in plasma from patients with severe acute asthma. J Allergy Clin Immunol 2003 Sep;112(3):525-30.

(579) Norgauer J, Dobos G, Kownatzki E, Dahinden C, Burger R, Kupper R, et al. Complement fragment C3a stimulates Ca2+ influx in neutrophils via a pertussis-toxin-sensitive G protein. Eur J Biochem 1993 Oct 1;217(1):289-94.

(580) Elsner J, Oppermann M, Czech W, Kapp A. C3a activates the respiratory burst in human polymorphonuclear neutrophilic leukocytes via pertussis toxin-sensitive G-proteins. Blood 1994 Jun 1;83(11):3324-31.

(581) Dillard P, Wetsel RA, Drouin SM. Complement C3a regulates Muc5ac expression by airway Clara cells independently of Th2 responses. Am J Respir Crit Care Med 2007 Jun 15;175(12):1250-8.

(582) Hoogerwerf M, Weening RS, Hack CE, Roos D. Complement fragments C3b and iC3b coupled to latex induce a respiratory burst in human neutrophils. Mol Immunol 1990 Feb;27(2):159-67.

(583) Melendi GA, Hoffman SJ, Karron RA, Irusta PM, Laham FR, Humbles A, et al. C5 modulates airway hyperreactivity and pulmonary eosinophilia during enhanced respiratory syncytial virus disease by decreasing C3a receptor expression. J Virol 2007 Jan;81(2):991-9.

(584) Lee SH, Rhim T, Choi YS, Min JW, Kim SH, Cho SY, et al. Complement C3a and C4a increased in plasma of patients with aspirin-induced asthma. Am J Respir Crit Care Med 2006 Feb 15;173(4):370-8.

(585) Abdel FM, El BM, Sherif A, Adel A. Complement components (C3, C4) as inflammatory markers in asthma. Indian J Pediatr 2010 Jul;77(7):771-3.

(586) Wills-Karp M. Complement activation pathways: a bridge between innate and adaptive immune responses in asthma. Proc Am Thorac Soc 2007 Jul;4(3):247-51.

(587) Koziol-White CJ, Yoo EJ, Cao G, Zhang J, Papanikolaou E, Pushkarsky I, et al. Inhibition of PI3K promotes dilation of human small airways in a rho kinase-dependent manner. Br J Pharmacol 2016 Sep;173(18):2726-38.

(588) Ge Q, Moir LM, Trian T, Niimi K, Poniris M, Shepherd PR, et al. The phosphoinositide 3'-kinase p110delta modulates contractile protein production and IL-6 release in human airway smooth muscle. J Cell Physiol 2012 Aug;227(8):3044-52.

(589) Halayko AJ, Kartha S, Stelmack GL, McConville J, Tam J, Camoretti-Mercado B, et al. Phophatidylinositol-3 kinase/mammalian target of rapamycin/p70S6K regulates contractile protein accumulation in airway myocyte differentiation. Am J Respir Cell Mol Biol 2004 Sep;31(3):266-75.

(590) Goncharova EA, Ammit AJ, Irani C, Carroll RG, Eszterhas AJ, Panettieri RA, et al. PI3K is required for proliferation and migration of human pulmonary vascular smooth muscle cells. Am J Physiol Lung Cell Mol Physiol 2002 Aug;283(2):L354-L363.

(591) Gupta MK, Asosingh K, Aronica M, Comhair S, Cao G, Erzurum S, et al. Defective Resensitization in Human Airway Smooth Muscle Cells Evokes beta-Adrenergic Receptor Dysfunction in Severe Asthma. PLoS One 2015;10(5):e0125803.

(592) Jude JA, Tirumurugaan KG, Kang BN, Panettieri RA, Walseth TF, Kannan MS. Regulation of CD38 expression in human airway smooth muscle cells: role of class I phosphatidylinositol 3 kinases. Am J Respir Cell Mol Biol 2012 Oct;47(4):427-35.

(593) Lau C, Wang X, Song L, North M, Wiehler S, Proud D, et al. Syk associates with clathrin and mediates phosphatidylinositol 3-kinase activation during human rhinovirus internalization. J Immunol 2008 Jan 15;180(2):870-80.

(594) Cakebread JA, Haitchi HM, Xu Y, Holgate ST, Roberts G, Davies DE. Rhinovirus-16 induced release of IP-10 and IL-8 is augmented by Th2 cytokines in a pediatric bronchial epithelial cell model. PLoS One 2014;9(4):e94010.

(595) Cakebread JA, Haitchi HM, Xu Y, Holgate ST, Roberts G, Davies DE. Rhinovirus-16 induced release of IP-10 and IL-8 is augmented by Th2 cytokines in a pediatric bronchial epithelial cell model. PLoS One 2014;9(4):e94010.

(596) Lee KS, Park SJ, Kim SR, Min KH, Lee KY, Choe YH, et al. Inhibition of VEGF blocks TGF-beta1 production through a PI3K/Akt signalling pathway. Eur Respir J 2008 Mar;31(3):523-31.

(597) Atherton HC, Jones G, Danahay H. IL-13-induced changes in the goblet cell density of human bronchial epithelial cell cultures: MAP kinase and phosphatidylinositol 3-kinase regulation. Am J Physiol Lung Cell Mol Physiol 2003 Sep;285(3):L730-L739.

(598) Bracke M, van de Graaf E, Lammers JW, Coffer PJ, Koenderman L. In vivo priming of FcalphaR functioning on eosinophils of allergic asthmatics. J Leukoc Biol 2000 Nov;68(5):655-61.

(599) Winkler DG, Faia KL, DiNitto JP, Ali JA, White KF, Brophy EE, et al. PI3K-delta and PI3K-gamma inhibition by IPI-145 abrogates immune responses and suppresses activity in autoimmune and inflammatory disease models. Chem Biol 2013 Nov 21;20(11):1364-74.

(600) Kampe M, Lampinen M, Stolt I, Janson C, Stalenheim G, Carlson M. PI3-kinase regulates eosinophil and neutrophil degranulation in patients with allergic rhinitis and allergic asthma irrespective of allergen challenge model. Inflammation 2012 Feb;35(1):230-9.

(601) Xu H, Gu LN, Yang QY, Zhao DY, Liu F. MiR-221 promotes IgE-mediated activation of mast cells degranulation by PI3K/Akt/PLCgamma/Ca(2+) pathway. J Bioenerg Biomembr 2016 Jun;48(3):293-9.

(602) Stark AK, Sriskantharajah S, Hessel EM, Okkenhaug K. PI3K inhibitors in inflammation, autoimmunity and cancer. Curr Opin Pharmacol 2015 Aug;23:82-91.

(603) Cheng D, Di H, Xue Z, Zhen G. CC16 gene A38G polymorphism and susceptibility to asthma: an updated meta-analysis. Intern Med 2015;54(2):155-62.

(604) Gharib SA, Nguyen EV, Lai Y, Plampin JD, Goodlett DR, Hallstrand TS. Induced sputum proteome in healthy subjects and asthmatic patients. J Allergy Clin Immunol 2011 Dec;128(6):1176-84.

(605) Singhania A, Rupani H, Jayasekera N, Lumb S, Hales P, Gozzard N, et al. Altered Epithelial Gene Expression in Peripheral Airways of Severe Asthma. PLoS One 2017;12(1):e0168680.

(606) Lindsey JY, Ganguly K, Brass DM, Li Z, Potts EN, Degan S, et al. c-Kit is essential for alveolar maintenance and protection from emphysema-like disease in mice. Am J Respir Crit Care Med 2011 Jun 15;183(12):1644-52.

(607) Wu W, Wang T, Dong JJ, Liao ZL, Wen FQ. Silencing of c-kit with small interference RNA attenuates inflammation in a murine model of allergic asthma. Int J Mol Med 2012 Jul;30(1):63-8.

(608) Chen Y, Qiao J. Protein-protein interaction network analysis and identifying regulation microRNAs in asthmatic children. Allergol Immunopathol (Madr) 2015 Nov;43(6):584-92.

(609) Imaoka H, Hoshino T, Okamoto M, Sakazaki Y, Sawada M, Takei S, et al. Endogenous and exogenous thioredoxin 1 prevents goblet cell hyperplasia in a chronic antigen exposure asthma model. Allergol Int 2009 Sep;58(3):403-10.

(610) Stankovic KM, Goldsztein H, Reh DD, Platt MP, Metson R. Gene expression profiling of nasal polyps associated with chronic sinusitis and aspirin-sensitive asthma. Laryngoscope 2008 May;118(5):881-9.

(611) Arumugam T, Simeone DM, Schmidt AM, Logsdon CD. S100P stimulates cell proliferation and survival via receptor for activated glycation end products (RAGE). J Biol Chem 2004 Feb 13;279(7):5059-65.

(612) Mercado-Pimentel ME, Onyeagucha BC, Li Q, Pimentel AC, Jandova J, Nelson MA. The S100P/RAGE signaling pathway regulates expression of microRNA-21 in colon cancer cells. FEBS Lett 2015 Aug 19;589(18):2388-93.

(613) Onyeagucha BC, Mercado-Pimentel ME, Hutchison J, Flemington EK, Nelson MA. S100P/RAGE signaling regulates microRNA-155 expression via AP-1 activation in colon cancer. Exp Cell Res 2013 Aug 1;319(13):2081-90.

(614) Penumutchu SR, Chou RH, Yu C. Interaction between S100P and the anti-allergy drug cromolyn. Biochem Biophys Res Commun 2014 Nov 21;454(3):404-9.

(615) Storms W, Kaliner MA. Cromolyn sodium: fitting an old friend into current asthma treatment. J Asthma 2005 Mar;42(2):79-89.

(616) Gibadulinova A, Pastorek M, Filipcik P, Radvak P, Csaderova L, Vojtesek B, et al. Cancer-associated S100P protein binds and inactivates p53, permits therapy-induced senescence and supports chemoresistance. Oncotarget 2016 Apr 19;7(16):22508-22.

(617) Li Q, Baines KJ, Gibson PG, Wood LG. Changes in Expression of Genes Regulating Airway Inflammation Following a High-Fat Mixed Meal in Asthmatics. Nutrients 2016 Jan 7;8(1).

(618) Ferrari S, Tagliafico E, Manfredini R, Grande A, Rossi E, Zucchini P, et al. Abundance of the primary transcript and its processed product of growth-related genes in normal and

leukemic cells during proliferation and differentiation. Cancer Res 1992 Jan 1;52(1):11-6.

(619) Lesniak W, Swart GW, Bloemers HP, Kuznicki J. Regulation of cell specific expression of calcyclin (S100A6) in nerve cells and other tissues. Acta Neurobiol Exp (Wars) 2000;60(4):569-75.

(620) Kuznicki J, Kordowska J, Puzianowska M, Wozniewicz BM. Calcyclin as a marker of human epithelial cells and fibroblasts. Exp Cell Res 1992 Jun;200(2):425-30.

(621) Breen EC, Tang K. Calcyclin (S100A6) regulates pulmonary fibroblast proliferation, morphology, and cytoskeletal organization in vitro. J Cell Biochem 2003 Mar 1;88(4):848-54.

(622) Nedjadi T, Kitteringham N, Campbell F, Jenkins RE, Park BK, Navarro P, et al. S100A6 binds to annexin 2 in pancreatic cancer cells and promotes pancreatic cancer cell motility. Br J Cancer 2009 Oct 6;101(7):1145-54.

(623) Zeng FY, Gerke V, Gabius HJ. Identification of annexin II, annexin VI and glyceraldehyde-3-phosphate dehydrogenase as calcyclin-binding proteins in bovine heart. Int J Biochem 1993 Jul;25(7):1019-27.

(624) Mani RS, Kay CM. Isolation and characterization of a novel molecular weight 11,000 Ca2(+)-binding protein from smooth muscle. Biochemistry 1990 Feb 13;29(6):1398-404.

(625) Wills FL, McCubbin WD, Kay CM. Smooth muscle calponin-caltropin interaction: effect on biological activity and stability of calponin. Biochemistry 1994 May 10;33(18):5562-9.

(626) Slomnicki LP, Lesniak W. S100A6 (calcyclin) deficiency induces senescence-like changes in cell cycle, morphology and functional characteristics of mouse NIH 3T3 fibroblasts. J Cell Biochem 2010 Feb 15;109(3):576-84.

(627) Leclerc E, Fritz G, Weibel M, Heizmann CW, Galichet A. S100B and S100A6 differentially modulate cell survival by interacting with distinct RAGE (receptor for advanced glycation end products) immunoglobulin domains. J Biol Chem 2007 Oct 26;282(43):31317-31.

(628) Duan L, Wu R, Zou Z, Wang H, Ye L, Li H, et al. S100A6 stimulates proliferation and migration of colorectal carcinoma cells through activation of the MAPK pathways. Int J Oncol 2014 Mar;44(3):781-90.

(629) Ning X, Sun S, Zhang K, Liang J, Chuai Y, Li Y, et al. S100A6 protein negatively regulates CacyBP/SIP-mediated inhibition of gastric cancer cell proliferation and tumorigenesis. PLoS One 2012;7(1):e30185.

(630) Ryoo HD, Bergmann A. The role of apoptosis-induced proliferation for regeneration and cancer. Cold Spring Harb Perspect Biol 2012 Aug 1;4(8):a008797.

(631) Jurewicz E, Goral A, Filipek A. S100A6 is secreted from Wharton's jelly mesenchymal stem cells and interacts with integrin beta1. Int J Biochem Cell Biol 2014 Oct;55:298-303.

(632) Lyu XJ, Li HZ, Ma X, Li XT, Gao Y, Ni D, et al. Elevated S100A6 (Calcyclin) enhances tumorigenesis and suppresses CXCL14-induced apoptosis in clear cell renal cell carcinoma. Oncotarget 2015 Mar 30;6(9):6656-69.

(633) Slomnicki LP, Nawrot B, Lesniak W. S100A6 binds p53 and affects its activity. Int J Biochem Cell Biol 2009 Apr;41(4):784-90.

(634) Joo JH, Kim JW, Lee Y, Yoon SY, Kim JH, Paik SG, et al. Involvement of NF-kappaB in the regulation of S100A6 gene expression in human hepatoblastoma cell line HepG2. Biochem Biophys Res Commun 2003 Jul 25;307(2):274-80.

(635) Kroliczak W, Pietrzak M, Puzianowska-Kuznicka M. P53-dependent suppression of the human calcyclin gene (S100A6): the role of Sp1 and of NFkappaB. Acta Biochim Pol 2008;55(3):559-70.

(636) Lesniak W, Szczepanska A, Kuznicki J. Calcyclin (S100A6) expression is stimulated by agents evoking oxidative stress via the antioxidant response element. Biochim Biophys Acta 2005 May 15;1744(1):29-37.

(637) Breen EC, Fu Z, Normand H. Calcyclin gene expression is increased by mechanical strain in fibroblasts and lung. Am J Respir Cell Mol Biol 1999 Dec;21(6):746-52.

(638) Grainge CL, Lau LC, Ward JA, Dulay V, Lahiff G, Wilson S, et al. Effect of bronchoconstriction on airway remodeling in asthma. N Engl J Med 2011 May 26;364(21):2006-15.

(639) Park JA, Fredberg JJ, Drazen JM. Putting the Squeeze on Airway Epithelia. Physiology (Bethesda) 2015 Jul;30(4):293-303.

(640) Park JA, Sharif AS, Tschumperlin DJ, Lau L, Limbrey R, Howarth P, et al. Tissue factor-bearing exosome secretion from human mechanically stimulated bronchial epithelial cells in vitro and in vivo. J Allergy Clin Immunol 2012 Dec;130(6):1375-83.

(641) Park JA, Atia L, Mitchel JA, Fredberg JJ, Butler JP. Collective migration and cell jamming in asthma, cancer and development. J Cell Sci 2016 Sep 15;129(18):3375-83.

(642) Tamai H, Yamaguchi H, Miyake K, Takatori M, Kitano T, Yamanaka S, et al. Amlexanox Downregulates S100A6 to Sensitize KMT2A/AFF1-Positive Acute Lymphoblastic Leukemia to TNFalpha Treatment. Cancer Res 2017 Aug 15;77(16):4426-33.

(643) Filipek A, Wojda U, Lesniak W. Interaction of calcyclin and its cyanogen bromide fragments with annexin II and glyceraldehyde 3-phosphate dehydrogenase. Int J Biochem Cell Biol 1995 Nov;27(11):1123-31.

(644) Fujii T, Kuzumaki N, Ogoma Y, Kondo Y. Effects of calcium-binding proteins on histamine release from permeabilized rat peritoneal mast cells. Biol Pharm Bull 1994 May;17(5):581-5.

(645) Filipek A, Wojda U. Chicken gizzard calcyclin--distribution and potential target proteins. Biochem Biophys Res Commun 1996 Aug 5;225(1):151-4.

(646) Ellison RT, III, Giehl TJ. Killing of gram-negative bacteria by lactoferrin and lysozyme. J Clin Invest 1991 Oct;88(4):1080-91.

(647) Ohbayashi H, Setoguchi Y, Fukuchi Y, Shibata K, Sakata Y, Arai T. Pharmacological effects of lysozyme on COPD and bronchial asthma with sputum: A randomized, placebo-controlled, small cohort, cross-over study. Pulm Pharmacol Ther 2016 Apr;37:73-80.

(648) Satta G, Azzarone B, Varaldo PE, Fontana R, Valisena S. Stimulation of spreading of trypsinized human fibroblasts by lysozymes from Staphylococcus aureus, hen egg white, and human urine. In Vitro 1980 Sep;16(9):738-50.

(649) Lee TH, Song HJ, Park CS. Role of inflammasome activation in development and exacerbation of asthma. Asia Pac Allergy 2014 Oct;4(4):187-96.

(650) Xu X, Chen H, Zhu X, Ma Y, Liu Q, Xue Y, et al. S100A9 promotes human lung fibroblast cells activation through receptor for advanced glycation end-product-mediated extracellular-regulated kinase 1/2, mitogen-activated protein-kinase and nuclear factor-kappaB-dependent pathways. Clin Exp Immunol 2013 Sep;173(3):523-35.

(651) Lee TH, Jang AS, Park JS, Kim TH, Choi YS, Shin HR, et al. Elevation of S100 calcium binding protein A9 in sputum of neutrophilic inflammation in severe uncontrolled asthma. Ann Allergy Asthma Immunol 2013 Oct;111(4):268-75.

(652) Gharib SA, Nguyen EV, Lai Y, Plampin JD, Goodlett DR, Hallstrand TS. Induced sputum proteome in healthy subjects and asthmatic patients. J Allergy Clin Immunol 2011 Dec;128(6):1176-84.

(653) Narumi K, Miyakawa R, Ueda R, Hashimoto H, Yamamoto Y, Yoshida T, et al. Proinflammatory Proteins S100A8/S100A9 Activate NK Cells via Interaction with RAGE. J Immunol 2015 Jun 1;194(11):5539-48.

(654) Kang JH, Hwang SM, Chung IY. S100A8, S100A9 and S100A12 activate airway epithelial cells to produce MUC5AC via extracellular signal-regulated kinase and nuclear factor-kappaB pathways. Immunology 2015 Jan;144(1):79-90.

(655) Ryckman C, Vandal K, Rouleau P, Talbot M, Tessier PA. Proinflammatory activities of S100: proteins S100A8, S100A9, and S100A8/A9 induce neutrophil chemotaxis and adhesion. J Immunol 2003 Mar 15;170(6):3233-42.

(656) Leukert N, Vogl T, Strupat K, Reichelt R, Sorg C, Roth J. Calcium-dependent tetramer formation of S100A8 and S100A9 is essential for biological activity. J Mol Biol 2006 Jun 16;359(4):961-72.

(657) Vogl T, Leukert N, Barczyk K, Strupat K, Roth J. Biophysical characterization of S100A8 and S100A9 in the absence and presence of bivalent cations. Biochim Biophys Acta 2006 Nov;1763(11):1298-306.

(658) Gomes LH, Raftery MJ, Yan WX, Goyette JD, Thomas PS, Geczy CL. S100A8 and S100A9-oxidant scavengers in inflammation. Free Radic Biol Med 2013 May;58:170-86.

(659) Lim SY, Raftery MJ, Goyette J, Hsu K, Geczy CL. Oxidative modifications of S100 proteins: functional regulation by redox. J Leukoc Biol 2009 Sep;86(3):577-87.

(660) Theron AJ, Steenkamp KJ, Anderson R. NADPH-oxidase activity of stimulated neutrophils is markedly increased by serum. Inflammation 1994 Oct;18(5):459-67.

(661) Baillet A, Hograindleur MA, El BJ, Grichine A, Berthier S, Morel F, et al. Unexpected function of the phagocyte NADPH oxidase in supporting hyperglycolysis in stimulated neutrophils: key role of 6-phosphofructo-2-kinase. FASEB J 2017 Feb;31(2):663-73.

(662) Newton K, Dixit VM. Signaling in innate immunity and inflammation. Cold Spring Harb Perspect Biol 2012 Mar 1;4(3).

(663) Feldman N, Rotter-Maskowitz A, Okun E. DAMPs as mediators of sterile inflammation in aging-related pathologies. Ageing Res Rev 2015 Nov;24(Pt A):29-39.

(664) Tan X, Zheng X, Huang Z, Lin J, Xie C, Lin Y. Involvement of S100A8/A9-TLR4-NLRP3 Inflammasome Pathway in Contrast-Induced Acute Kidney Injury. Cell Physiol Biochem 2017;43(1):209-22.

(665) Jahnsen FL, Moloney ED, Hogan T, Upham JW, Burke CM, Holt PG. Rapid dendritic cell recruitment to the bronchial mucosa of patients with atopic asthma in response to local allergen challenge. Thorax 2001 Nov;56(11):823-6.

(666) McCarthy NE, Jones HA, Marks NA, Shiner RJ, Ind PW, Al-Hassi HO, et al. Inhaled allergen-driven CD1c up-regulation and enhanced antigen uptake by activated human respiratory-tract dendritic cells in atopic asthma. Clin Exp Allergy 2007 Jan;37(1):72-82.

(667) Koh YI, Lee JB, Lee SR, Ji SG, Choi IS. Relationship between dendritic cells and activated eosinophils in induced sputum of asthmatics. J Korean Med Sci 2005 Jun;20(3):384-9.

(668) Padigel UM, Lee JJ, Nolan TJ, Schad GA, Abraham D. Eosinophils can function as antigen-presenting cells to induce primary and secondary immune responses to Strongyloides stercoralis. Infect Immun 2006 Jun;74(6):3232-8.

(669) Le NJ, Shahine A, Gras S. Molecular features of lipid-based antigen presentation by group 1 CD1 molecules. Semin Cell Dev Biol 2018 Dec;84:48-57.

(670) Ardeniz Ã, Unger S, Onay H, Ammann S, Keck C, Cianga C, et al. Î²2-Microglobulin deficiency causes a complex immunodeficiency of the innate and adaptive immune system. J Allergy Clin Immunol 2015 Aug;136(2):392-401.

(671) Ortiz-Stern A, Kanda A, Mionnet C, Cazareth J, Lazzari A, Fleury S, et al. Langerin+ dendritic cells are responsible for LPS-induced reactivation of allergen-specific Th2 responses in postasthmatic mice. Mucosal Immunol 2011 May;4(3):343-53.

(672) Orio F, Jr., Terouanne B, Georget V, Lumbroso S, Avances C, Siatka C, et al. Potential action of IGF-1 and EGF on androgen receptor nuclear transfer and transactivation in normal and cancer human prostate cell lines. Mol Cell Endocrinol 2002 Dec 30;198(1-2):105-14.

(673) Menard G, Turmel V, Bissonnette EY. Serotonin modulates the cytokine network in the lung: involvement of prostaglandin E2. Clin Exp Immunol 2007 Nov;150(2):340-8.

(674) Lechin F, van der Dijs B, Orozco B, Lechin M, Lechin AE. Increased levels of free serotonin in plasma of symptomatic asthmatic patients. Ann Allergy Asthma Immunol 1996 Sep;77(3):245-53.

(675) Scanlon RT, Chang S. Brain norepinephrine: a possible role in bronchial asthma. Ann Allergy 1988 Apr;60(4):333-8.

(676) Cabezas GA, Israili ZH, Velasco M. The actions of dopamine on the airways. Am J Ther 2003 Nov;10(6):477-86.

(677) Jobsis Q, Raatgeep HC, Hermans PW, de Jongste JC. Hydrogen peroxide in exhaled air is increased in stable asthmatic children. Eur Respir J 1997 Mar;10(3):519-21.

(678) Ou XM, Chen K, Shih JC. Monoamine oxidase A and repressor R1 are involved in apoptotic signaling pathway. Proc Natl Acad Sci U S A 2006 Jul 18;103(29):10923-8.

(679) Naoi M, Maruyama W, Akao Y, Yi H, Yamaoka Y. Involvement of type A monoamine oxidase in neurodegeneration: regulation of mitochondrial signaling leading to cell death or neuroprotection. J Neural Transm Suppl 2006;(71):67-77.

(680) Dittrich AM, Krokowski M, Meyer HA, Quarcoo D, Avagyan A, Ahrens B, et al. Lipocalin2 protects against airway inflammation and hyperresponsiveness in a murine model of allergic airway disease. Clin Exp Allergy 2010 Nov;40(11):1689-700.

(681) Kirkbride HJ, Bolscher JG, Nazmi K, Vinall LE, Nash MW, Moss FM, et al. Genetic polymorphism of MUC7: allele frequencies and association with asthma. Eur J Hum Genet 2001 May;9(5):347-54.

(682) O'Neil SE, Sitkauskiene B, Babusyte A, Krisiukeniene A, Stravinskaite-Bieksiene K, Sakalauskas R, et al. Network analysis of quantitative proteomics on asthmatic bronchi: effects of inhaled glucocorticoid treatment. Respir Res 2011 Sep 22;12:124.

(683) Taupin D, Podolsky DK. Trefoil factors: initiators of mucosal healing. Nat Rev Mol Cell Biol 2003 Sep;4(9):721-32.

(684) Wiede A, Jagla W, Welte T, Kohnlein T, Busk H, Hoffmann W. Localization of TFF3, a new mucus-associated peptide of the human respiratory tract. Am J Respir Crit Care Med 1999 Apr;159(4 Pt 1):1330-5.

(685) Babyatsky MW, deBeaumont M, Thim L, Podolsky DK. Oral trefoil peptides protect against ethanol- and indomethacin-induced gastric injury in rats. Gastroenterology 1996 Feb;110(2):489-97.

(686) Baines KJ, Simpson JL, Wood LG, Scott RJ, Fibbens NL, Powell H, et al. Sputum gene expression signature of 6 biomarkers discriminates asthma inflammatory phenotypes. J Allergy Clin Immunol 2014 Apr;133(4):997-1007.

(687) Singhania A, Rupani H, Jayasekera N, Lumb S, Hales P, Gozzard N, et al. Altered Epithelial Gene Expression in Peripheral Airways of Severe Asthma. PLoS One 2017;12(1):e0168680.

(688) Ross AJ, Dailey LA, Brighton LE, Devlin RB. Transcriptional profiling of mucociliary differentiation in human airway epithelial cells. Am J Respir Cell Mol Biol 2007 Aug;37(2):169-85.

(689) Grus FH, Podust VN, Bruns K, Lackner K, Fu S, Dalmasso EA, et al. SELDI-TOF-MS ProteinChip array profiling of tears from patients with dry eye. Invest Ophthalmol Vis Sci 2005 Mar;46(3):863-76.

(690) Matheis N, Okrojek R, Grus FH, Kahaly GJ. Proteomics of tear fluid in thyroid-associated orbitopathy. Thyroid 2012 Oct;22(10):1039-45.

(691) Singhania A, Rupani H, Jayasekera N, Lumb S, Hales P, Gozzard N, et al. Altered Epithelial Gene Expression in Peripheral Airways of Severe Asthma. PLoS One 2017;12(1):e0168680.

(692) Staples KJ, Hinks TS, Ward JA, Gunn V, Smith C, Djukanovic R. Phenotypic characterization of lung macrophages in asthmatic patients: overexpression of CCL17. J Allergy Clin Immunol 2012 Dec;130(6):1404-12.

(693) Epstein TG, LeMasters GK, Bernstein DI, Ericksen MB, Martin LJ, Ryan PH, et al. Genetic variation in small proline rich protein 2B as a predictor for asthma among children with eczema. Ann Allergy Asthma Immunol 2012 Mar;108(3):145-50.

(694) Francis SM, Larsen JE, Pavey SJ, Bowman RV, Hayward NK, Fong KM, et al. Expression profiling identifies genes involved in emphysema severity. Respir Res 2009 Sep 2;10:81.

(695) Gabazza EC, Taguchi O, Tamaki S, Takeya H, Kobayashi H, Yasui H, et al. Thrombin in the airways of asthmatic patients. Lung 1999;177(4):253-62.

(696) Hallstrand TS, Wurfel MM, Lai Y, Ni Z, Gelb MH, Altemeier WA, et al. Transglutaminase 2, a novel regulator of eicosanoid production in asthma revealed by genome-wide expression profiling of distinct asthma phenotypes. PLoS One 2010 Jan 5;5(1):e8583.

(697) Bousquet J, Godard P, Michel FB. Antihistamines in the treatment of asthma. Eur Respir J 1992 Oct;5(9):1137-42.

(698) Simon T, Semsei AF, Ungvari I, Hadadi E, Virag V, Nagy A, et al. Asthma endophenotypes and polymorphisms in the histamine receptor HRH4 gene. Int Arch Allergy Immunol 2012;159(2):109-20.

(699) Mukherjee M, Bulir DC, Radford K, Kjarsgaard M, Huang CM, Jacobsen EA, et al. Sputum autoantibodies in patients with severe eosinophilic asthma. J Allergy Clin Immunol 2018 Apr;141(4):1269-79.

(700) Rossios C, Pavlidis S, Hoda U, Kuo CH, Wiegman C, Russell K, et al. Sputum transcriptomics reveal upregulation of IL-1 receptor family members in patients with severe asthma. J Allergy Clin Immunol 2018 Feb;141(2):560-70.

(701) Oda H, Kawayama T, Imaoka H, Sakazaki Y, Kaku Y, Okamoto M, et al. Interleukin-18 expression, CD8(+) T cells, and eosinophils in lungs of nonsmokers with fatal asthma. Ann Allergy Asthma Immunol 2014 Jan;112(1):23-8.

(702) Hirata J, Kotani J, Aoyama M, Kashiwamura S, Ueda H, Kuroda Y, et al. A role for IL-18 in human neutrophil apoptosis. Shock 2008 Dec;30(6):628-33.

(703) Leung BP, Culshaw S, Gracie JA, Hunter D, Canetti CA, Campbell C, et al. A role for IL-18 in neutrophil activation. J Immunol 2001 Sep 1;167(5):2879-86.

(704) Mete F, Ozkaya E, Aras S, Koksal V, Etlik O, Baris I. Association between gene polymorphisms in TIM1, TSLP, IL18R1 and childhood asthma in Turkish population. Int J Clin Exp Med 2014;7(4):1071-7.

(705) Srivastava S, Pelloso D, Feng H, Voiles L, Lewis D, Haskova Z, et al. Effects of interleukin-18 on natural killer cells: costimulation of activation through Fc receptors for immunoglobulin. Cancer Immunol Immunother 2013 Jun;62(6):1073-82.

(706) Hallstrand TS, Lai Y, Henderson WR, Jr., Altemeier WA, Gelb MH. Epithelial regulation of eicosanoid production in asthma. Pulm Pharmacol Ther 2012 Dec;25(6):432-7.

(707) Kim JT, Lee SJ, Kang MA, Park JE, Kim BY, Yoon DY, et al. Cystatin SN neutralizes the inhibitory effect of cystatin C on cathepsin B activity. Cell Death Dis 2013 Dec 19;4:e974.

(708) Ekiel I, Abrahamson M. Folding-related dimerization of human cystatin C. J Biol Chem 1996 Jan 19;271(3):1314-21.

(709) Bjorck L, Grubb A, Kjellen L. Cystatin C, a human proteinase inhibitor, blocks replication of herpes simplex virus. J Virol 1990 Feb;64(2):941-3.

(710) Wallin H, Bjarnadottir M, Vogel LK, Wasselius J, Ekstrom U, Abrahamson M. Cystatins--Extra- and intracellular cysteine protease inhibitors: High-level secretion and uptake of cystatin C in human neuroblastoma cells. Biochimie 2010 Nov;92(11):1625-34.

(711) Wang G, Baines KJ, Fu JJ, Wood LG, Simpson JL, McDonald VM, et al. Sputum mast cell subtypes relate to eosinophilia and corticosteroid response in asthma. Eur Respir J 2016 Apr;47(4):1123-33.

(712) Baines KJ, Simpson JL, Wood LG, Scott RJ, Fibbens NL, Powell H, et al. Sputum gene expression signature of 6 biomarkers discriminates asthma inflammatory phenotypes. J Allergy Clin Immunol 2014 Apr;133(4):997-1007.

(713) Wang G, Baines KJ, Fu JJ, Wood LG, Simpson JL, McDonald VM, et al. Sputum mast cell subtypes relate to eosinophilia and corticosteroid response in asthma. Eur Respir J 2016 Apr;47(4):1123-33.

(714) Singhania A, Rupani H, Jayasekera N, Lumb S, Hales P, Gozzard N, et al. Altered Epithelial Gene Expression in Peripheral Airways of Severe Asthma. PLoS One 2017;12(1):e0168680.

(715) Wang G, Baines KJ, Fu JJ, Wood LG, Simpson JL, McDonald VM, et al. Sputum mast cell subtypes relate to eosinophilia and corticosteroid response in asthma. Eur Respir J 2016 Apr;47(4):1123-33.

(716) Yang CY, Chen JB, Tsai TF, Tsai YC, Tsai CY, Liang PH, et al. CLEC4F is an inducible C-type lectin in F4/80-positive cells and is involved in alpha-galactosylceramide presentation in liver. PLoS One 2013;8(6):e65070.

(717) De SM, Spagnuolo L, Lore NI, Cigana C, De F, I, Broman KW, et al. Mapping genetic determinants of host susceptibility to Pseudomonas aeruginosa lung infection in mice. BMC Genomics 2016 May 11;17:351.

(718) Baines KJ, Simpson JL, Wood LG, Scott RJ, Fibbens NL, Powell H, et al. Sputum gene expression signature of 6 biomarkers discriminates asthma inflammatory phenotypes. J Allergy Clin Immunol 2014 Apr;133(4):997-1007.

(719) Berthon BS, Gibson PG, Wood LG, MacDonald-Wicks LK, Baines KJ. A sputum gene expression signature predicts oral corticosteroid response in asthma. Eur Respir J 2017 Jun;49(6).

(720) Baines KJ, Simpson JL, Wood LG, Scott RJ, Fibbens NL, Powell H, et al. Sputum gene expression signature of 6 biomarkers discriminates asthma inflammatory phenotypes. J Allergy Clin Immunol 2014 Apr;133(4):997-1007.

(721) Freidin MB, Bragina EI, Fedorova OS, Deev IA, Kulikov ES, Ogorodova LM, et al. [Genome-wide association study of allergic diseases in Russians of Western Siberia]. Mol Biol (Mosk) 2011 May;45(3):464-72.

(722) Wang K, Gao M, Yang M, Meng F, Li D, Lu R, et al. Transcriptome analysis of bronchoalveolar lavage fluid from children with severe Mycoplasma pneumoniae pneumonia reveals novel gene expression and immunodeficiency. Hum Genomics 2017 Mar 16;11(1):4.

(723) Sun J, Li H, Huo Q, Cui M, Ge C, Zhao F, et al. The transcription factor FOXN3 inhibits cell proliferation by downregulating E2F5 expression in hepatocellular carcinoma cells. Oncotarget 2016 Jul 12;7(28):43534-45.

(724) Dai Y, Wang M, Wu H, Xiao M, Liu H, Zhang D. Loss of FOXN3 in colon cancer activates beta-catenin/TCF signaling and promotes the growth and migration of cancer cells. Oncotarget 2017 Feb 7;8(6):9783-93.

(725) Singhania A, Rupani H, Jayasekera N, Lumb S, Hales P, Gozzard N, et al. Altered Epithelial Gene Expression in Peripheral Airways of Severe Asthma. PLoS One 2017;12(1):e0168680.

(726) Redington AE, Roche WR, Madden J, Frew AJ, Djukanovic R, Holgate ST, et al. Basic fibroblast growth factor in asthma: measurement in bronchoalveolar lavage fluid basally and following allergen challenge. J Allergy Clin Immunol 2001 Feb;107(2):384-7.

(727) Xie B, Tassi E, Swift MR, McDonnell K, Bowden ET, Wang S, et al. Identification of the fibroblast growth factor (FGF)-interacting domain in a secreted FGF-binding protein by phage display. J Biol Chem 2006 Jan 13;281(2):1137-44.

(728) Staples KJ, Hinks TS, Ward JA, Gunn V, Smith C, Djukanovic R. Phenotypic characterization of lung macrophages in asthmatic patients: overexpression of CCL17. J Allergy Clin Immunol 2012 Dec;130(6):1404-12.

(729) Hallstrand TS, Wurfel MM, Lai Y, Ni Z, Gelb MH, Altemeier WA, et al. Transglutaminase 2, a novel regulator of eicosanoid production in asthma revealed by genome-wide expression profiling of distinct asthma phenotypes. PLoS One 2010 Jan 5;5(1):e8583.

(730) Agrawal S, Townley RG. Role of periostin, FENO, IL-13, lebrikzumab, other IL-13 antagonist and dual IL-4/IL-13 antagonist in asthma. Expert Opin Biol Ther 2014 Feb;14(2):165-81.

(731) Papathanassiou E, Loukides S, Bakakos P. Severe asthma: anti-IgE or anti-IL-5? Eur Clin Respir J 2016;3:31813.

(732) Shrimanker R, Pavord ID. Interleukin-5 Inhibitors for Severe Asthma: Rationale and Future Outlook. BioDrugs 2017 Apr;31(2):93-103.

(733) Yarova PL, Stewart AL, Sathish V, Britt RD, Jr., Thompson MA, AP PL, et al. Calcium-sensing receptor antagonists abrogate airway hyperresponsiveness and inflammation in allergic asthma. Sci Transl Med 2015 Apr 22;7(284):284ra60.