Genetic basis of immune-mediated susceptibility to infectious disease



Vivek Naranbhai

Keble College

Trinity 2015

Submitted in fulfillment of Doctor of Philosophy in Clinical Medicine

Supervised by:

Professor Adrian VS Hill

Professor Helen McShane

Professor Andrew P Morris

Vivek Naranbhai
Keble College
DPhil in Clinical Medicine, TT 2015

Abstract

Infectious diseases remain a major driver of morbidity and mortality globally. For example, *Mycobacterium tuberculosis* causes about 10 million cases of tuberculosis (TB) and 1.5 million deaths. In a series of studies I demonstrate that discovering the genetic underpinnings of how immune factors affect susceptibility to diseases such as TB reveal new insights into disease susceptibility. Previous transcriptome-wide studies suggested that the ratio of myeloid:lymphoid transcripts was elevated in infants who subsequently developed TB. These data were reminiscent of observations made 80 years ago where the ratio of monocytes:lymphocytes (ML) at time of challenge with mycobacteria, was predictive of survival in animal models. To establish the peripheral blood ML ratio as a novel intermediate phenotype for TB susceptibility, I demonstrated that the ML ratio predicts TB disease in three independent cohorts of adults (n=2207), infants (n=1336) and pregnant women (n=1202). Moreover, in collaboration with others, these findings were extended to other diseases including malaria. Turning attention to the pathophysiologic basis of this association, I found qualitative differences in monocytes according to ML ratio even amongst healthy individuals (n=136), both in functional response to mycobacteria and in transcriptomic profile at rest. Next, I found that the ML and neutrophil:lymphocyte (NL) ratio are highly phenotypically and genetically correlated, suggesting the presence of pleiotropic genetic regulators. I present formal estimates of heritability based on concordance and structural equation modeling of data from twin studies (1664 twin pairs), and amongst unrelated adults (n=5034). A meta-analysis of four genome-wide association studies (total n=5768) was performed as a resource to understand how the ML and NL ratios may be regulated, and to facilitate Mendelian randomization studies. Finally, in a more generalised approach I identified genetic regulators of gene expression in monocytes and neutrophils from healthy adults (n=101) and demonstrated how these give fundamental new insights into a wide array of infectious and non-infectious diseases. Taken together, these studies offer original contributions to the understanding of how genetic factors give rise to immune-mediated susceptibility to infectious disease.

“Science without humanity”

-MK Ghandhi’s fifth social sin

Acknowledgements

I am immensely grateful to many below who have each supported me in a myriad of ways; many that remain unmentioned but sincerely appreciated:

The patients and volunteers in these studies whose selfless contribution to science is made in the hope that it may eventually alleviate human suffering.

My mentors and supervisors Professors Adrian VS Hill, Helen McShane and Andrew Morris for their breadth of knowledge, wise counsel and for affording me the latitude to explore new ideas and approaches (even across geographies!).

My peers in the Hill, Knight, McShane and Morris groups for their companionship and intellectual discussion: Anna Rautanen, Benjamin Fairfax, Daniel Wong, Katherine Elliott, Rachel Tanner, Elena Stylianou, Matthew O'Shea, Peter Humburg, Seiko Makino, Evelyn Lau, Wanseon Lee, Esther Ng, Katherine Plant, Alexander Mentzer, James Gilchrist, Thomas Parks, Tara Mills and many others.

My friends and 'science-buddies' Doctors David Bauer, Mohammad Ali Rai, Mary Carrington and Leonore Herzenberg for our science chats.

An exemplary array of collaborators who have shared reagents, advice and data freely: in particular Professors Salim Abdool Karim, Quarraisha Abdool

Karim, Shabhir Madhi, Daya Moodley, Julian Knight, Benjamin Fairfax and members of the International Tuberculosis Host Genetics Consortium.

The Rhodes Trust, whose financial support is gratefully acknowledged and whose intellectual community fostered robust debate and helped focused my efforts.

My Viking 750 lathe, rusty and old, yet always able to spin away the worries of the world and permit deep meditation.

My parents, Raj and Bharti, and my siblings for inculcating in me a deep love of science and the conscience to attempt to apply it to human health; and Deepti, Manish, Aruna, Nagin and Kulsem for welcoming me to the UK.

Finally, my wife and companion, Christina Chang for her love, for challenging me to think harder and deeper about a problem, cheering me with good humour and wit and supporting me through a (second) PhD!

Attributions

This thesis is a report on some of the research work pursued at the Wellcome Trust Centre for Human Genetics, University of Oxford between September 2011 and July 2015.

All research presented in this thesis is original and was conducted by the author.

Contents

Abstract i

Acknowledgements iii

Attributions v

Contents vi

Dissertation outline x

Associated Publications xiv

List of Figures xix

List of Tables xxiv

List of Supplementary Tables xxvi

Abbreviations xxx

Chapter One. Literature Review 1

1.1 The epidemiology of human TB infection 1

1.2 Pathogenesis of TB 4

1.3 Diagnosis, prevention and treatment of infection 8

1.4 Correlates of susceptibility to TB 11

1.4.1 Heritability of host susceptibility to TB 13

1.4.2 Mendelian Susceptibility to Mycobacterial Disease (MSMD) 15

1.4.3 Candidate gene studies 16

1.4.4 Genome-wide association studies 27

1.4.5 Transcriptomic studies 29

1.4.7 Genome-wide cell-based or animal screens 30

1.4.8 Host-pathogen coevolution as a basis for immunity as a double-edged sword 31

1.5 Conclusions **33**

Chapter Two. Discovery and validation of the ratio of monocytes:lymphocytes as a correlate of TB susceptibility **35**

2.1 Introduction **35**

2.2 The Ratio of Monocytes to Lymphocytes in Peripheral Blood predicts the Risk of Incident Tuberculosis in HIV-Infected Adults Initiating Antiretroviral Therapy **36**

2.2.1 Background 36

2.2.2 Methods 38

2.2.3 Results 42

2.2.4 Discussion 58

2.3 The Ratio of Monocytes to Lymphocytes in Peripheral Blood at age 3-months predicts the Risk of Incident Tuberculosis in in the first two years of life **61**

2.3.1 Background 61

2.3.2 Methods 62

2.3.3 Results 64

2.3.4 Discussion 70

2.4 The Ratio of Monocytes to Lymphocytes in Peripheral Blood during pregnancy predicts the Risk of Incident in HIV-Infected Postpartum Women 73

2.4.1 Background 73

2.4.2 Methods 73

2.4.3 Results 75

2.4.4 Discussion 78

2.5 Conclusions **79**

Chapter Three. Functional basis for the association of myeloid:lymphoid ratio and human disease **80**

3.1 Introduction **80**

3.2 Distinct Transcriptional and Anti-Mycobacterial Profiles of Peripheral Blood Monocytes Dependent on the Ratio of Monocytes:Lymphocytes **81**

3.2.1 Background 81

3.2.2 Methods 83

3.2.3 Results 86

3.2.4 Discussion 95

3.3 Conclusions **97**

Chapter Four. The genetic architecture of myeloid:lymphoid cell ratios ... **99**

4.1 Introduction **99**

4.2 Heritability and co-heritability of absolute and relative leucocyte counts suggest coregulation of myeloid and lymphoid cell numbers **100**

4.2.1 Background 100

4.2.2 Methods 101

4.2.3 Results 103

4.2.4 Discussion 110

4.3 A genome wide association study of host genetic correlates of myeloid:lymphoid ratio **111**

4.3.1 Background 111

4.3.2 Methods 112

4.3.3 Results 114

4.3.4 Discussion 123

4.4 Conclusions **123**

Chapter Five. Genetics of gene expression in myeloid cells **125**

5.1 Introduction **125**

5.2 Genomic modulators of gene expression in human neutrophils **126**

5.2.1 Background 126

5.2.2 Methods 128

5.2.3 Results and Discussion 135

5.3 Conclusions 168

Chapter Six. Discussion and conclusions 169

6.1 Overall conclusions 169

6.2 Limitations 170

6.3 Perspectives and future research 172

Chapter Seven. References 173

Dissertation outline

Human tuberculosis (TB) remains a major scourge for humanity. It is a leading cause of adult and adolescent death in developing countries, and has become a recurring concern in developed countries due to recent increases in drug-resistance, population migration and the human immunodeficiency virus (HIV) pandemic. For new prophylactic, diagnostic and therapeutic interventions to be developed, elucidation of the underpinning pathophysiology is required. The overall aim of this work is to identify aspects of the genetic basis of immune-mediated susceptibility to infectious disease.

In chapter one, a brief general review of the literature relevant to TB, as the main disease-focus of this work is presented to contextualise the contributions made by this thesis.

Chapters two through five, the results chapters, each describe the literature that is directly relevant to the methods and results that are subsequently presented and discussed in that chapter.

Chapter two presents evidence that the ratio of monocytes to lymphocytes (ML ratio) in peripheral blood is associated with TB risk amongst HIV-infected adults, infants and peripartum women. It thus identifies a novel intermediate immunological phenotype for TB disease. In work not described in detail here, I contributed to a larger effort to identify immune correlates of TB risk in BCG-vaccinated infants (associated publication #8).

Chapter three reports on studies performed to identify possible pathophysiological mechanisms that underlie the association between ML ratio and disease. Through a combination of mycobacterial growth inhibition assays, transcriptomic profiling of monocytes from healthy adults and bioinformatic analyses we suggest that ML ratio reflects monocyte function, possibly as a consequence of ontogeny from myeloid-biased haematopoietic stem cells.

Although not discussed in detail, in collaboration with others, we also showed that the ML ratio is predictive of Malaria disease (associated publication #9), and of all-cause mortality amongst adult emergency room admittee's (associated publication #10). These latter studies reveal that the ML ratio may be of pathophysiological importance in other immune-mediated diseases. Moreover, during the course of this DPhil, an evolving literature on the association between ML ratio and other diseases became extant, and was supplemented by studies suggesting that the neutrophil to lymphocyte (NL) ratio may be of importance in some disorders. In light of the evolving literature on ML and NL ratios, the work in the following chapters was explicitly designed to examine whether the ML and NL ratio or their constituents reflect similar physiological roots.

In chapter four, attention is turned to the genetic architecture of the absolute and relative frequencies of leucocytes in peripheral blood. This is importance for two reasons. Firstly, partitioning of the proximal determinants of the ML and NL ratios into genetic and environmental determinants could help delineate the possibly causal pathways that lead to disease susceptibility.

Secondly, if specific genetic determinants of the ratio could be identified, they would facilitate Mendelian randomization studies as a powerful method of identifying whether the association between ML and/or NL ratio and disease is causal. Accordingly, chapter four presents a series of analyses of the heritability of the absolute or relative frequency of myeloid and lymphoid cells in peripheral blood in twins and unrelated adults, provides estimates of coheritability as a measure of shared genetic architecture, and decomposes the genetic heritability by genomic location. In the second part of chapter four, I present results of a genome-wide association study (GWAS) of both ML and NL ratio's.

Chapter five, the final results chapter details a study in which we attempted to unify genomic and transcriptomic studies in monocytes and neutrophils using approaches that I had previously developed. In work not detailed in this thesis (associated publication #11), I assisted in studies that dissected the genetic determinants of gene expression in monocytes subjected to various forms of stimulation. In other work not detailed in this thesis (associated publication #12), I helped to demonstrate that a novel non-coding RNA in which a genetic variant shows association with bactaraemia is expressed in neutrophils. In order to demonstrate this, I developed methods to sort neutrophils from peripheral blood suitable for transcriptomic analyses-a heretofore challenging endeavor. Hence in chapter five, I demonstrate how these approaches identify a large library of genetic variants that alter expression of genes in monocytes and neutrophils that are known to be involved in immune-mediated susceptibility to an array of infectious and non-infectious diseases. We demonstrate, through bioinformatics and functional assays, how the

epigenetic environment surrounding genomic variants results in differences in neutrophil and monocyte gene expression that may account for their involvement in, for example, leprosy-another mycobacterial disease of global significance.

Finally, although each chapter contextualises the relevance of that chapter's findings to the broader field, in chapter six I provide a more general discussion of the conclusions, limitations and implications for future research that the work described in this thesis collectively supports. In particular I discuss how the results from chapter four and five are being used in analyses of the International Tuberculosis Host Genetics Consortium, a collaborative study group including >35000 cases and controls with TB that we established during the course of my DPhil studies.

Several aspects of the work performed during this period of DPhil study are not detailed in this thesis. As an extension of the genomic approaches that help expose the physiological underpinnings of infectious disease susceptibility, I collaborated with investigators applying proteomic approaches to identify possible correlates of treatment response in TB (associated publication #13). Applying the approaches detailed here for TB to studies of HIV disease also led to associated publications #14 and #15. Similarly, I was fortunate to be able to contribute to application of some of the approaches developed here to other diseases (associated publications #16 and #17).

Associated Publications

Further detailed in this thesis

1. Naranbhai V, Hill AVS, Abdool Karim SS, Naidoo K, Abdool Karim Q, Warimwe G, McShane H, Fletcher H. Ratio of monocytes to lymphocytes in peripheral blood identifies adults at risk of incident tuberculosis among adults initiating antiretroviral therapy. **J Infect Dis.** 2014 Feb;209 (4):500-9.

   Contribution: conceived of idea, coordinated study, performed analysis, wrote first draft.

   Material from this manuscript is reproduced under copyright permission from Oxford University Press, License number 3678580888116

2. Naranbhai V, Kim S, Fletcher H, Cotton MF, Violari A, Mitchell C, Nachman S, McShane H, Hill AVS, Madhi SA. Association between the ratio of monocytes:lymphocytes at age 3 months to risk for tuberculosis (TB) in the first two years of life. **BMC Med.** 2014 Jul 17;12 (1):120

   Contribution: conceived of idea, coordinated study, performed analysis, wrote first draft.

   Material from this manuscript is reproduced under a Creative Commons CC-BY licence.

3. Naranbhai V, Moodley D, Chipato T, Stranix-Chibanda L, Nakabaiito C, Kamateeka M, Musoke P, Manji K, George K, Emel LM, Richardson P, Andrew P, Fowler M, Fletcher H, Mcshane H, Coovadia HM, Hill AV; for the HPTN 046 protocol team. The association between the ratio of monocytes:lymphocytes and risk of tuberculosis (TB) amongst HIV infected pregnant women. **J Acquir Immune Defic Syndr.** 2014 Sep 22.

Contribution: conceived of idea, coordinated study, performed analysis, wrote first draft.

4. Naranbhai V, Fletcher HA, Tanner R, O'Shea M, McShane H, Fairfax BP, Knight JC, Hill AVS. Distinct transcriptional and anti-mycobacterial profiles of peripheral blood monocytes dependent on the ratio of monocytes:lymphocytes. **Submitted.**

   Contribution: conceived of idea, coordinated study, performed analysis, wrote first draft

5. Naranbhai V, Cambridge Bioresource, United Kingdom Blood Service 1958 Birth cohort, TWINSUK, Morris AP, Hill AVS. Heritability and co-heritability of absolute and relative leucocyte counts suggest coregulation of myeloid and lymphoid cell numbers. **Manuscript in preparation.**

   Contribution: conceived of idea, performed analysis, wrote first draft

6. Naranbhai V, Cambridge Bioresource, United Kingdom Blood Service 1958 Birth cohort, TWINSUK,University of Tartu, Magi R, Morris AP, Hill AVS. A genome wide association study of host genetic correlates of monocyte:lymphocyte and neutrophil:lymphocyte ratios. **Manuscript in preparation.**

   Contribution: conceived of idea, performed analysis, wrote first draft

7. Naranbhai V, Fairfax BP, Makino S, Humburg P, Wong D, Ng E, Hill AVS, Knight JC. Genomic modulators of gene expression in human neutrophils. **Nature Communications.** 2015 Jul 7;6:7545.

   Contribution: Co-conceived of idea, performed laboratory experiments, performed analysis, wrote first draft.

Material from this manuscript is reproduced under a Creative Commons CC-BY licence.

Not further detailed in this thesis

8. Fletcher HA, Snowden MA, Landry B, Rida W, Satti I, Harris SA, Matsumiya M, Tanner R, O'Shea MK, Dheenadhayalan V, Bogardus L, Stockdale L, Marsay L, Chomska A, Harrington-Kandt R, Thomas ZM, Naranbhai V, Stylianou E, Hatherill E, Hussey G, Mahomed H, Tameris M, McClain JB, Evans TG, Hanekom WA, Scriba TJ, McShane H. Immune Correlates of Risk in BCG-vaccinated infants from the MVA85A efficacy trial. **Submitted**.

   Contribution: Assisted in laboratory assays, reviewed and edited final draft

9. Warimwe G, Murungi LM, Kumuyu G, Nyangweso GM, Wambua J, Naranbhai V, Fletcher HA, Hill AVS, Bejon P, Osier FHA, Marsh K. The ratio of monocytes to lymphocytes in peripheral blood correlates with increased susceptibility to malaria in Kenyan children. **PLoS One.** 2013;8 (2):e57320.

   Contribution: Contributed to study conception and analytical design, reviewed and edited final draft

10. Naranbai V, Infection Oxford Research Database (IORD), Hill AVS, Wyllie D.Peripheral blood myeloid/platelet:lymphoid ratios predict mortality risk amongst 70 917 emergency medical admissions: a cohort study. **Manuscript in preparation**.

    Contribution: co-conceived of idea, performed analysis, wrote first draft

11. Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, Lau E, Jostins L, Plant K, Andrews R, McGee C, Knight JC. Innate Immune Activity

Conditions Effects of Common Regulatory Variants upon Monocyte Gene Expression. **Science**. 2014 Mar 7;343 (6175):1246949.

Contribution: Assisted in performing laboratory assays and analyses, reviewed and edited final draft

12. The Kenyan Bacteraemia study group and Wellcome Trust Case Control Consortium. Genetic variation in a non-coding RNA is associated with a doubled risk of pneumococcal bacteraemia in Kenyan children. **Submitted.**

    Contribution: Optimised methods to isolate neutrophils, performed qPCR as functional follow-up experiments for manuscript, wrote corresponding section of first draft, reviewed and edited final draft

13. Kaewseekhao B, Naranbhai V, Roytrakul S, Namwat W, Paemanee A, Lulitanond V, Chaiprasert A, Faksri K. Comparative proteomics of macrophages infected with Mycobacterium tuberculosis identifies clearance biomarkers for tuberculosis treatment. **PLoS One**. 2015;10 (7):e0134168

    Contribution: Performed analyses, co-wrote first draft

14. Naranbhai V, de Assess Rosa D, Werner L, Moodley R, Hong H, Kharsany A, Mlisana K, Sibeko S, Garrett N, Chopera D, Carr WHC, Abdool Karim Q, Hill AVS, Abdool Karim SS, Altfeld M, Gray C, Ndungu T. Killer-cell-Immunoglobulin-like Receptor (KIR) gene profiles modify HIV disease course, not HIV acquisition in South African women. **Submitted**.

    Contribution: conceived of idea, coordinated study, performed analysis, wrote first draft

15. Gurdasani D, Iles L, Dillon DG, Young EH, Olsen AD, Naranbhai V, Fidler S, Gkania-Klotsas E, Post FA, Kellam P, Porter K, Sandhu MS on behalf

of the UK HIV Genomics Consortium. A Systematic Review of Definitions of Extreme Phenotypes of HIV Control and Progression. **AIDS.** 2014 Jan 14;28 (2):149-62.

Contribution: Contributed to data interpretation and reviewed and edited final draft

16. Mills TC, Elliot KS, Parks T, Naranbhai V, Leven M, Butler CC, Little P, Verheij T, Hinds C, Garrard CS, Goossens H, Rautanen A, Chapman S, Hill AVS. IFITM3 and susceptibility to respiratory viral infections. **J Infect Dis.** 2014 Apr 1;209 (7):1028-31.

Contribution: Contributed to study design and manuscript preparation

17. Wong D, Lee W, Humburg P, Makino S, Lau E, Naranbhai V, Fairfax BP, Chan K, Plant K, Knight JC. Genomic mapping of the MHC transactivator CIITA using an integrated ChIP-seq and genetical genomics approach. **Genome Biology.** 2014;15 (10):494

Contribution: Performed flow-cytometry follow-up experiments for manuscript, wrote corresponding section of first draft, reviewed and edited final draft

List of Figures

Figure 1: Global estimates of TB incidence rates in 2013. From Global Tuberculosis Report 2014, WHO.

Figure 2: TB pathogenesis. Reproduced with permission from Nature Publishing Group (License no. 3679040369264)(Nunes-Alves, Booty et al. 2014).

Figure 3: Phylogeographic and dating analysis of M.TB complex. Reproduced with permission from Nature Publishing Group (License no. 3679050891364)(Comas, Coscolla et al. 2013).

Figure 4: Kaplan-Meier estimates of probability of TB-disease free survival for individuals commencing cART, by category of baseline ML ratio

Figure 5: Kaplan-Meier estimates of probability of TB-disease free survival for individuals commencing cART, by centile category of baseline monocyte (A) or lymphocyte (B) count.

Figure 6: ML ratio at baseline does not predict immunological or virological response to cART.

Figure 7: ML ratios, monocyte counts and lymphocyte counts amongst healthy HIV-uninfected woman.

Figure 8: The impact of HIV acquisition and disease progression on ML ratios.

Figure 9: The impact of cART on ML ratio.

Figure 10: Distribution of monocyte:lymphocyte (ML) ratio at enrolment shown for HEU (n=796, dashed line) and HIV positive infants (n=540, solid line).

Figure 11: The ML ratio is associated with mycobacterial growth *ex vivo*, but the *in vivo* ML ratio of the donor dominates in explaining mycobacterial growth in a growth inhibition assay.

Figure 12: Monocyte transcriptomes are associated with the ML ratio of the human donor.

Figure 13: Monocyte genes associated with the ML ratio are enriched for genes in the interferon signalling pathway.

Figure 14: Monocyte genes associated with the ML ratio are enriched for transcription factor binding sites for IRF1, IRF2, RUNX1, RELA and Esrrb.

Figure 15: The distribution of peripheral blood white blood cell (WBC), neutrophil, lymphocyte or monocyte counts and ML and NL ratios according to participating cohort, and age group.

Figure 16: Correlation coefficients between peripheral blood measures.

Figure 17: (A) SNP-heritability (V_G/V_P) of each of the leucocyte measures according to cohort, or combined and (B) partitioned according to chromosome.

Figure 18: Genetic correlation or coheritability of leucocyte measures.

Figure 19: Manhattan plot showing GWS ($p<5x10^{-8}$) variants (blue) and non-GWS variants (red) associated with (A) monocyte, (B) lymphocyte or (C) neutrophil counts in 5768 adults.

Figure 20: Effect size of 39,125 variants nominally associated with both lymphocyte and monocyte counts.

Figure 21: Manhattan plot of meta-analysis results of genome-wide variants associated with (A) MLR or (B) NLR.

Figure 22: (A) Regional association plot showing variants on Chr9 associated with the MLR, and (B) a forest plot showing primary association statistics for rs10980797, the peak SNP across all five participating cohorts and in the meta-analysis.

Figure 23: Enrichment of transcription factor binding sites amongst genes most proximal to (A) MLR or (B) NLR associated genetic variants.

Figure 24: Study schema and overview of *cis* and *trans* eQTL in primary human neutrophils.

Figure 25: Effect size and location of *cis* eQTL in neutrophils relative to transcription start site (TSS) (a) or gene structure (b), partitioned by gene size.

Figure 26: Genes involved in the development, migration and function of neutrophils are frequent targets of genetic regulation.

Figure 27: Pathway analysis of genes with neutrophil specific *cis* eQTL.

Figure 28: Euler diagrams showing overlap of genes with an eQTL previously observed in lymphoid (a) or myeloid (b) cell types and their overlap with genes with eQTL in neutrophils (c).

Figure 29: Patterns of shared and cell-type specific *cis* eQTL in neutrophils and monocytes.

Figure 30: Effect sizes of eQTL.

Figure 31: Epigenetic and functional basis for shared and cell-type specific eQTL in neutrophils and monocytes.

Figure 32: eQTL in neutrophils and their association with complex disease or trait.

Figure 33: Shared and cell type specific *cis* eQTL are enriched for variants associated with diseases or traits in GWAS studies.

Figure 34: An example in *PADI4* demonstrates how integrated analysis of an eQTL informs understanding of rheumatoid arthritis (RA) risk.

Figure 35: eQTL in NOD2 and their involvement in Leprosy and Crohn's disease.

List of Tables

Table 1: Twin studies of TB susceptibility

Table 2: Genes associated with Tuberculosis disease in humans from candidate gene studies

Table 3: Baseline characteristics for the cohort of 1862 adults commencing cART in the study cohort.

Table 4: Cox-proportional hazards modeling of tuberculosis risk by pre-cART ML ratio category, n=1862.

Table 5: Cox-proportional hazards modeling of tuberculosis risk by pre-cART monocyte or lymphocyte category, n=1862.

Table 6: Stratified Cox-proportional hazards modeling of Tuberculosis risk amongst individuals with ML ratios above the 95^{th} centile or below the 5^{th} centile.

Table 7: Sensitivity analysis: the association between baseline ML ratio and TB outcomes holds across TB outcome groupings.

Table 8: Cox proportional hazards models of TB disease between 6 weeks and 18 months postpartum in HPTN 046 (n=1202)

Table 9: heritability estimates of six leucocyte parameters in TWINSUK (n=880 MZ and 784 DZ pairs)

Table 10: Baseline characteristics and quality metrics of genotypes for cohorts included in leucocyte GWAS

List of Supplementary Tables

Supplementary Tables are provided in electronic format on the electronic media card in this thesis. This is due to the size of the tables and hence the greater ease of reference in electronic format.

Supplementary Table 1: Results of association analysis of gene expression in monocytes with the host ML ratio or monocyte count.

Supplementary Table 2: Canonical pathways enriched amongst monocyte genes whose expression is associated with the ML-ratio from ingenuity pathway analysis.

Supplementary Table 3: Upstream regulators of ML-ratio associated transcriptomic changes from ingenuity pathway analysis.

Supplementary Table 4: ML-ratio associated signature enrichment amongst genes associated with other diseases.

Supplementary Table 5: Genes with variants that are reported to be associated with TB disease susceptibility and the association between their transcript levels in monocytes and ML ratio.

Supplementary Table 6: GWS loci associated with monocyte, lymphocyte or neutrophil counts, MLR or NLR

Supplementary Table 7: Summary of epigenetic features in neutrophils and monocytes from the BLUEPRINT project.

Supplementary Table 8: Primer and probe sequences for oligonucleotides used in this study.

Supplementary Table 9: Listing of data files and source for external files to which reference is made in this study.

Supplementary Table 10: *cis* eQTL in neutrophils ranked by the variance explained by the peak eQTL for each probe.

Supplementary Table 11: Association between copy-number variants and gene expression.

Supplementary Table 12: *trans* eQTL in neutrophils ranked by the variance explained by the peak eQTL for each probe.

Supplementary Table 13: eQTL in genes involved in neutrophil function categorised by role in neutrophil function and relationship, if any, with mendelian disorder.

Supplementary Table 14: Enrichment of eQTL in regions marked by histone modifications, DNA methylation, transcription factor bindigs, microRNA binding or transcribed enhancers.

Supplementary Table 15: Differential expression of transcription factors in neutrophils and monocytes ranked by log fold change in expression.

Supplementary Table 16: eQTL in neutrophils that are in regions bound by conserved microRNA.

Supplementary Table 17: eQTL in neutrophils that are in LD (r^2>0.8) with or are variants associated with a complex trait listed in the NHGRI GWAS catalog.

Supplementary Table 18: Enrichment of eQTL in neutrophils that are in LD (r^2>0.8) with or are variants associated with a complex trait listed in the NHGRI GWAS catalog according to disease category.

Supplementary Table 19: Results of droplet digital PCR for the ratio of rs1981760-C:rs1981760-T after chromatin immunoprecipitation of STAT3 from crosslinked neutrophil DNA from rs1981760 heterozygote individuals.

Supplementary Table 20: Neutrophil eQTL in regions of the genome identified as being subject to selection in Grossman et al.

Abbreviations

CAPRISA	Centre for the AIDS Programe of Research in South Africa
cART	combination Anti-Retroviral Therapy
cDNA	complementary Deoxyribonucleic Acid
DNA	Deoxyribonucleic Acid
EDTA	Ethylene-Diameno Tetra-acetic Acid
ENU	N-ethyl-N-nitrosourea
eQTL	expression Quantitative Trait Loci
FBC	Full Blood Count
GWAS	Genome-Wide Association Study
HEU	Human Immunodeficiency Virus (HIV) Exposed-Uninfected
HIV	Human Immunodeficiency Virus
HPTN	Human Immunodeficiency Virus (HIV) Prevention Trials Network
IFN	Interferon
IL	Interleukin
IMPAACT	International Maternal Pediatric Adolescent AIDS Clinical Trials Group
IGRA	Interferon Gamma Release Assay

INH	Isoniazid
IPT	Isoniazid Preventive Therapy
LTBI	Latent Tuberculosis (TB) Infection
M.TB	*Mycobacterium tuberculosis*
MDR	Multi-Drug Resistant
ML	Monocyte:lymphocyte ratio
MSMD	Mendelian susceptibility to Mycobacterial Disease
NL	Neutrophil:lymphocyte ratio
NNRTI	Non-nucleoside Reverse Transcriptase Inhibitor
NRTI	Nucleoside Reverse Transcriptase Inhibitor
PBMC	Peripheral Blood Mononuclear Cells
pTB	Pulmonary Tuberculosis
RNA	Ribonucleic Acid
siRNA	small interfering Ribonucleic Acid
TB	Tuberculosis
TST	Tuberculin Skin Test
XDR	Extensively-Drug Resistant

Chapter One. Literature Review

Sequencing studies suggest that *Mycobacterium tuberculosis (M.TB)* - the causative agent- has been a human pathogen since prior to our migration out of Africa(Comas, Coscolla et al. 2013). Human skeletal remains suggest that since at least 7,250 B.C.E.(Holloway, Henneberg et al. 2011) human TB has been a major scourge for humanity earning it a title as the “Captain of these men of death” in the 17th century (Bunyan 1680) when, as the cause of the “White Plague”, it was the principal cause of death amongst several infections. *M.TB* remains one of the top two causes of adult and adolescent death in most developing countries, and is becoming a recurring concern in developed countries - including the United Kingdom - due to increases in drug-resistance, population migration and the human immunodeficiency virus pandemic(WHO 2012). Yet current tools to prevent, diagnose and treat TB are blunt.

For new preventive, diagnostic and therapeutic interventions to be developed, elucidation of the underpinning pathophysiology is required. The overarching warrant of my work is that new genetic tools allow us to elucidate the biology of TB susceptibility and this can be translated into new clinical tools for the diagnosis, prevention or treatment of TB.

1.1 The epidemiology of human TB infection

Estimates in the 1990s suggested that about 2 billion people, or one-third of the extant population were latently infected with the pathogen(Dye, Scheele et al. 1999). Roughly 9 million children and adults develop active disease annually and in 2013, 1.5 million died(WHO 2014). Although the overall

incidence rate of TB probably peaked in 2004, the burden of disease is the highest in human history, concomitant with population expansion(Lawn and Zumla 2011). The majority of cases continue to occur in South-east Asia, the Western Pacific region and Africa (Figure 1) but incidence rates of disease in Eastern Europe, and South America are increasing. The six highest burden countries in 2009 were India, China, Nigeria, Pakistan, Indonesia and South Africa.

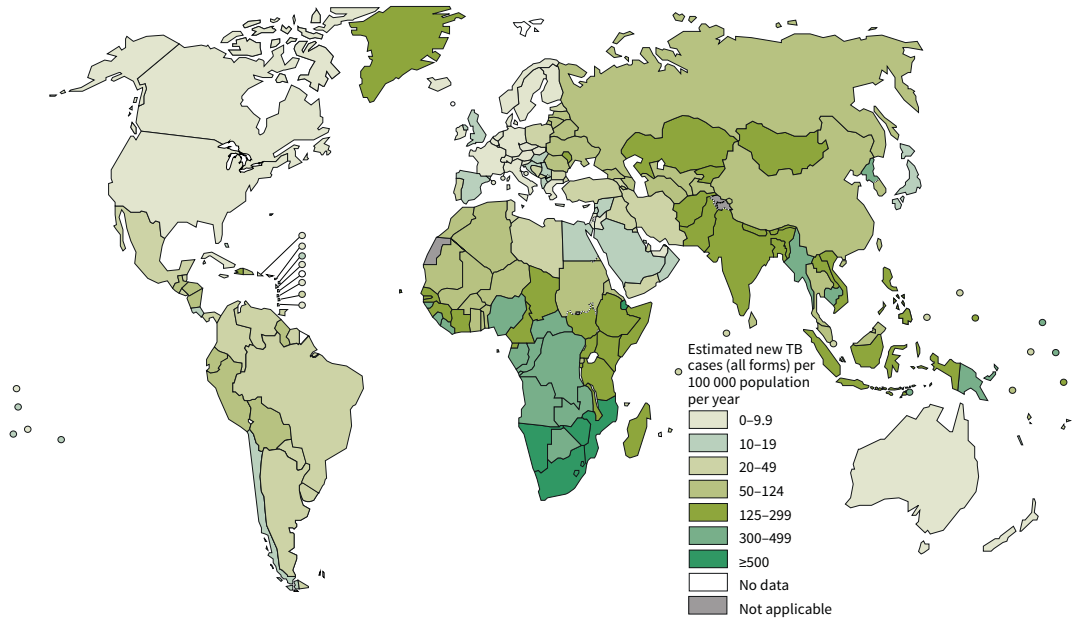


Figure 1: Global estimates of TB incidence rates in 2013. From Global Tuberculosis Report 2014, WHO.

The reasons for continued growth in burden are complex. Due to population expansion, widespread immunodeficiency secondary to infection with Human Immunodeficiency Virus (HIV)- a major risk factor for disease that accounts for 13% of TB cases - limited progress in improving hygiene globally, and suboptimal prevention strategies, incidence rates have been sustained in many developing countries. Next, methods to diagnose TB are suboptimal.

Consequently, the reproductive number (R_0) of infection for TB is high because of the period of infectiousness during undiagnosed and untreated infection. Finally, even where diagnosis and treatment are instituted, treatment success rates are disappointing and drug resistance a growing problem-in 2013 there were 480,000 incident MDR TB cases. Together, these factors make halting and beginning to reverse incidence of TB - Millennium Development Goal 6C(WHO)- a seemingly Sisyphean labour. Nevertheless, a reduction in the numbers of individuals with TB disease of about 1.5% since the early 2000's has been achieved and the 2015 Global TB Strategy's adopted resolution is ending TB by 2035 through a 90% decline in cases and 95% decline in mortality (WorldHealthAssembly 2014).

The reduction in TB on Europe though the 20th century, serves as an exemplar for tackling TB. Through the 17th-19th century Europe suffered incidence rates of up to 1,000 per 100,000 people, per year (Daniel 2009). But with substantial improvement in hygiene (better housing, nutrition and income), rates fell precipitously in the early 20th century. This trend was further consolidated with the discovery of therapeutic agents active against TB such as Streptomycin (1943), Isoniazid (1952) and Rifampicin (1970's) (Daniel 2009). The European experience suggests that the greatest impact on TB will probably be made by improved social and economic conditions but may also require biomedical innovation. In the United Kingdom, the burden fell precipitously from around 113,000 in 1913 to fewer than 5,000 in the 1980's but rose again thereafter and has settled at around 9,000 cases per year, concentrated in immigrant and HIV infected communities(HPA 2007). Thus

the second implication of the European experience is that even with improvement in conditions, it should be expected that fluxes in the numbers of patients at risk and numbers of migrants entering a population which may have active disease and be infectious, may sustain TB burden.

1.2 Pathogenesis of TB

Robert Koch identified *M.TB*, as the cause of TB in 1882 (Koch 1882). The bacillus is an aerobic, non spore-forming, non-motile and non-encapsulated bacterium with a lipid-rich mycolic acid cell-wall that retains dyes even after treatment with acidic solvents, making it “acid-fast” (Ryan and Ray 2010). *M.TB* is a facultative intracellular pathogen belonging to the *Mycobacteria* genus, which consists of more than 100 species that occur in the environment and amongst animals (Ryan and Ray 2010, Tortoli 2012). Recent studies of the 4.4 Mbp *M.TB* genome suggest that there are six/seven major phylogenetic strains circulating worldwide (Filliol, Motiwala et al. 2006, Comas, Coscolla et al. 2013). It appears that these strains may differ in infection efficiency, immunogenicity, virulence and mutability(Gagneux and Small 2007, Krishnan, Malaga et al. 2011, Reiling, Homolka et al. 2013{Ford, 2013 #138)}. Relative to other bacteria, *M.TB* has a slow replicative rate of about 18-54 hours(Gill, Harik et al. 2009). Further, *M.TB* is thought to be able enter a latent phase of growth during which replication ceases, and metabolic demands fall (Ryan and Ray 2010). The relevance of mycobacterial latency to clinical latency *in vivo* remains controversial, particularly in light of evidence of similar rates of *M.TB* evolution in latent and active disease in primates(Ford, Lin et al. 2011).

M.TB infects humans by aerosols, most efficiently by smaller droplets that are deposited into alveolar spaces in the lower lung (Figure 2). Bacilli are readily endocytosed by alveolar macrophages following macrophage recognition of mycobacteria via NOD2, Toll-like receptors (TLR) and likely other receptors(Pandey, Yang et al. 2009). This initial phase of infection is accompanied by a robust innate immune response involving neutrophils, macrophages, DCs and innate lymphoid cells. A growing body of evidence supports the idea that in the majority of cases this is sufficient for bacterial eradication. However, *M.TB* replicate within macrophages avoiding lysosome-endosome fusion and instigating necrosis as a mechanism of escape. *M.TB* have also been shown to survive acidification of the endosome should this ensue, and escape and survive in the cytosol. The ESX-1/ESAT-6 secretion system, present in only a few mycobacteria, is critical to many of these host-adaptation strategies (Cambier, Falkow et al. 2014). Therefore *M.TB* has evolved mechanisms for surviving macrophage defense mechanisms, and are also able to productively infect neutrophils and dendritic cells (DC) (Lowe, Redford et al. 2012). Moreover, mycobacteria enhance CCL2 secretion by macrophages thereby instigating recruitment of further macrophages as target cells(Cambier, Falkow et al. 2014). Macrophages may go on to survive with intracellular infection, die by apoptosis leading to mycobacterial death and antigen uptake by DCs leading to cross-presentation to B- and T-cells in regional lymph nodes, or die by necrosis leading to externalization of mycobacteria and potential dissemination to other cells or organs. M.TB induce host LXA4 secretion and inhibit PGE2 production as a means to push

macrophages towards necrosis, which would tend to favour bacterial survival over apoptosis(Cambier, Falkow et al. 2014).

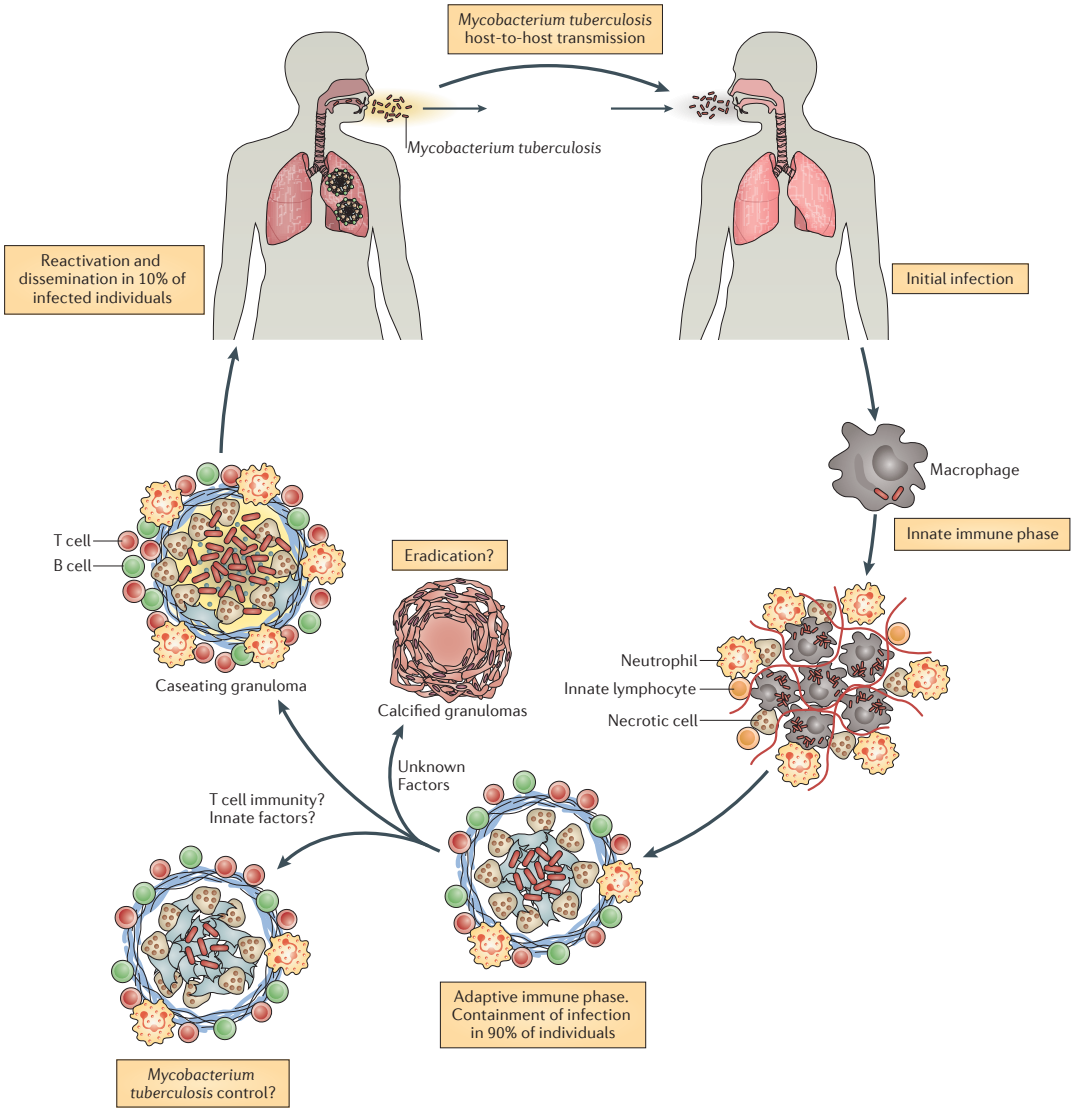


Figure 2: TB pathogenesis. Reproduced with permission from Nature Publishing Group (License no. 3679040369264)(Nunes-Alves, Booty et al. 2014).

Clinically, primary infection leads to active disease in about 1-5% of cases, characterized by immunopathology due to tissue destruction(Kaufmann, Hussey et al. 2010). In the majority of cases, immune mechanisms lead to control of infection prior to profound tissue destruction and disease. Innate and adaptive responses are presumed to be responsible for the establishment

of a granuloma-a complex pathological tissue consisting of mature macrophages that appear epitheloid in appearance and frequently become calcified to form a tubercle. This granuloma both protects the host by anatomically containing infection, but is exploited by the pathogen in that mycobacteria within granuloma's are probably better able to evade immune and pharmacological attack (Ramakrishnan 2012). A similar paradoxical relationship exists at the cellular level between macrophages and the bacillus. Although macrophages become activated by *M.TB* and are crucial effector cells in the defense against intracellular pathogens by stewarding additional innate and adaptive immune responses, bacilli subvert this response by impairing macrophage effector responses and can induce necrosis of macrophages (Ryan and Ray 2010, Ramakrishnan 2012). This leads to the characteristic caseous necrosis at the centre of the macrophage.

The majority of TB cases are pulmonary, but TB can infect most organs. Whether independent pathophysiological mechanisms are involved is unclear. Dogma holds that the establishment of a Ghon focus –a localized granuloma- visible on chest X-ray when calcified, indicates a primary response to TB infection (Kaufmann, Hussey et al. 2010). Enlargement of regional lymph nodes about 8-12 days after infection coincides with development of an adaptive TB-specific T-cell immune response that is measurable by day 14-17, which may lead to eradication or control of infection and establishment of clinical latency (Lawn and Zumla 2011). An effective immune response is then thought to keep the host free of symptoms and the bacillus in abeyance, but not necessarily eradication of infection. Instead bacilli are thought to themselves enter a latent state of infection within macrophages. Changes in

the immune response over time, immunosuppression or senescence leads to the breakdown of latency and consequent ‘active disease’. Recent animal and human studies demonstrate that the binary definition of latent and active disease is largely artificial and even within a single host, lesions show a spectrum between latency and active disease (Barry, Boshoff et al. 2009, Ford, Lin et al. 2011, Lin, Ford et al. 2014).

Nevertheless, the fact that host responses are sufficient in some humans to control infection is demonstrated by two observations. First, the lifetime risk of active disease amongst individuals with latent infection, defined by the presence of a T-cell response to *M.TB*, is estimated at about 10%. Although risk is increased by an order of magnitude or more if the host becomes immunocompromised (Lawn and Zumla 2011), this points to the fact that the majority of individuals are able to avoid clinical disease. The second observation suggesting the presence of protective immunity is from an unfortunate natural experiment that occurred in Lubeck, Germany in 1930, prior to antibiotic therapy availability, when 253 infants were accidentally infected with virulent *M.TB* through a contaminated culture of the Bacille Calmette-Guerin (BCG) vaccine (1930, 1932). Despite a large and direct inoculation, only 68 of the infants died, and the survivors remained alive up to 12 years later. Identifying the host factors that protect humans may help in developing better interventions for TB.

1.3 Diagnosis, prevention and treatment of infection

Current tools to diagnose infection rely on three approaches (Longo 2012). Firstly, detection of the pathogen in tissue/mucosal samples such as sputum,

by microscopy or nucleic acid amplification remains the cornerstone for diagnosis of active disease. Identification of tissue damage by imaging of a pattern consistent with TB is the second approach. Thoracic imaging may demonstrate characteristic reticulonodular shadows and hilar adenopathy in pulmonary TB, or a millet-seed-like pattern in miliary TB. Likewise, computerized tomographic scans of the head may demonstrate basal enhancement consistent with an inflammatory exudate that characteristically gathers at the base of the brain in TB meningitis. Finally, detection of a *M.TB* specific T-cell response is widely used to diagnose latent TB infection. This has classically been evaluated by a tuberculin-skin test (TST) but interferon-gamma release assays have been demonstrated to be more sensitive and specific for latent TB than TST (Pai, Zwerling et al. 2008).

There are two chief modalities for the prevention of TB: vaccination and prophylactic treatment with antituberculous drugs (Longo 2012). Bacillus Calmette-Guerrin (BCG) was developed by sequential passage of an *M. bovis* isolate on glycerinated ox-bile over 13 years by Calmette and Guerin in Paris in 1906 and became more widely used in Europe after the end of World War I. Due to variation in efficacy by latitude(Wilson, Fineberg et al. 1995), strain of vaccine (due to strain evolution)(Ritz, Hanekom et al. 2008, Ritz and Curtis 2009), co-infections and possibly host genetic factors, effectiveness varies between nil and 80% (Andersen and Doherty 2005, Roy, Eisenhut et al. 2014). Recent data suggest that the mechanism of BCG protection may be non-specific alteration of innate immune memory(Kleinnijenhuis, Quintin et al. 2012, Kleinnijenhuis, Quintin et al. 2014, Kleinnijenhuis, Quintin et al. 2014),

accounting for BCG effects on all-cause mortality, asthma(Arnoldussen, Linehan et al. 2011) and autoimmune disease(Ristori, Romano et al. 2014). Nevertheless, BCG is recommended for use by the WHO and is used in high burden countries based on evidence that it remains cost-effective to prevent severe childhood disease (Trunz, Fine et al. 2006). Several new vaccinations are in development and some have reached phase II/III clinical trials, giving some hope for an effective vaccine(Kaufmann, Hussey et al. 2010). The second modality for prevention relies on the administration of Isoniazid preventive therapy (IPT), to individuals at risk for infection. It appears that the effectiveness of this modality may be limited to those with latent TB infection (Bucher, Griffith et al. 1999, Samandari, Agizew et al. 2011) suggesting that the predominant mechanism of action is by treating latent infection. But data are conflicting as IPT success requires continued administration(Samandari, Agizew et al. 2011), which would suggest that it works to prevent or treat early infection. And in many high-incidence settings IPT appears to confer no benefit(Madhi, Nachman et al. 2011, Churchyard, Fielding et al. 2014). Hence guidelines for IPT use vary globally.

Treatment of individuals with active TB disease was revolutionized with the development of drugs effective against the pathogen. Prior to drug treatment for TB, medical therapy included quarantine (or in practice ostracisation) or admission to sanatoria for those who could afford it, an unproven medicament or, as was practiced for many years, iatrogenic pneumothoraces (Daniel 2009). In 1943, the antibiotic Streptomycin was hailed as a possible cure for TB, but emergence of resistance was already reported in the first clinical trial

of Streptomycin for TB in 1946 (MRC 1948). This observation foreshadowed the rapid emergence of resistance soon after each new antibiotic release, such as Isoniazid or Rifampicin in 1952 (MRC 1952) and the 1970's (Vall-Spinosa, Lester et al. 1970) respectively. Combination of both Isoniazid and Rifampicin was demonstrated to be more robust to resistance evolution, and therefore remains the backbone of current regimens to treat both latent and active TB. There are currently about 25 drugs used against TB in humans (Longo 2012). These drugs are combined into first-line regimens and regimens for retreatment or treatment of individuals with drug resistance. But, the tested regimens thus far have three drawbacks: a) resistance of mycobacteria against each of these drugs, including new drugs such as TMC207 has already been reported(Petrella, Cambau et al. 2006), b) almost all drugs have significant side effects and many are so toxic that despite efficacy they can be used only for salvage therapy and c) almost all tested combinations require prolonged therapy. Newer drugs are rapidly progressing through the development pipeline and entering trials as components of treatment-shortening regimens(Zumla, Nahid et al. 2013) or against multi drug resistant (MDR) TB, mycobacteria resistant to isoniazid and/or rifampicin or extensively drug resistant (XDR) TB that are, in addition, resistant to quinolones and an injectable anti-mycobacterial. Newer modalities, potentially ones that exploit host responses are desirable and are similarly being actively pursued (Zumla, Nahid et al. 2013).

1.4 Correlates of susceptibility to TB

The principal risk factors for TB infection are an increased exposure to infectious individuals and host immunity, affecting the probability of exposure

and the probability of clearance or disease respectively (Lawn and Zumla 2011, Longo 2012). Evidence for the former is largely from epidemiological studies and historical declines of TB disease tracking population density and hygiene(Daniel 2009). A role for host-factors in TB susceptibility has been appreciated throughout written history. Indeed, TB was widely presumed to be a familial disease. As a concrete example, early Sanskrit texts are reported to have advised against the marriage of high-caste Brahmins to women with a family history of TB, for it was presumed that the disease was familial. Only a small number of families with Mendelian susceptibility patterns have been described and the majority shows susceptibility to non-tuberculous mycobacteria. However, even amongst sporadic cases, when Robert Koch identified a germ-cause (Koch 1882), it became apparent that an interaction between environmental and genetic variables may underlie risk, and that this could plausibly be mediated through altered immunity.

Several diseases that alter immunity enhance the risk of TB. These include HIV infection, Diabetes Mellitus, malnutrition, smoking, silicosis, alcohol, chronic renal failure, malignancy, advanced age and immunosuppressive therapy (corticosteroids and tumour necrosis factor (TNF) antagonists) (Lawn and Zumla 2011). Noting these clinical observations, studies have examined specific aspects of immunity in TB disease pathogenesis amongst individuals with one of these causes of immunosuppression or without a secondary cause of immunodeficiency. Collectively, these studies led to a view that tumour necrosis factor (TNF) and interferon gamma (IFN-γ), likely produced by CD4+ T-cells, are necessary, but not sufficient, for protection from disease.

As a cautionary note however, a recent large prospective study failed to identify any cellular correlate of TB following BCG vaccination(Kagina, Abel et al. 2010). A range of other cells including neutrophils, regulatory T-cells, Th17 cells, myeloid-derived suppressor cells and CD1d-restricted T-cells and pathways such as autophagy have been suggested to be involved but their roles too remain inconclusive in the absence of prospective cohort study evidence for their involvement(O'Garra, Redford et al. 2013, Nunes-Alves, Booty et al. 2014).

Therefore in this section I describe how identification of specific cellular and genetic determinants of susceptibility to TB has formed the basis for identifying immune mechanisms of disease.

1.4.1 Heritability of host susceptibility to TB

Justification for a genetic component of risk for TB is demonstrated by higher-than expected concordance of disease amongst monozygotic than dizygotic twins. In addition, some recent studies have examined the heritability of traits intermediate between the exposure variable and TB risk.

Three major twin studies of TB susceptibility have been reported(Diehl and von Verschner 1933, Kallman and Reisner 1943, Comstock 1978). As shown in Table 1, the concordance rates amongst monozygotic twins are consistently higher than amongst dizygotic twins. Although varying by the differing underlying concordance rate estimates, the narrow-sense heritability is up to 80%. Van der Eijik and colleagues have however argued that greater concordance amongst monozygotic twins can be explained by greater

environmental sharing between monozygotic versus dizygotic twins and consequent greater exposure of an uninfected twin to an infected monozygotic twin (van der Eijk, van de Vosse et al. 2007). Their argument is weak, as there appears to be little empirical evidence to support the notion that the exposure of a monozygotic twin to its infected partner is sufficiently higher than between dizygotic twins to fully explain higher concordance. Nevertheless, their argument highlights the necessary contribution of TB exposure- an environmental factor- to TB susceptibility.

Table 1: Twin studies of TB susceptibility

No. of MZ twin pairs	Concordance in MZ twin pairs (R_{MZ})*	No. of DZ twin pairs	Concordance in DZ twin pairs (R_{DZ})*	Ref
80	0.65	125	0.25	(Diehl and von Verschner 1933)
78	0.88	230	0.28	(Kallman and Reisner 1943)
54	0.33	148	0.14	(Diehl and von Verschner 1933, Comstock 1978)

*R_{MZ} and R_{DZ} refer to concordance estimates of the phenotype amongst monozygotic and dizygotic twins respectively

Based on insights gleaned from studies of individuals with Mendelian Susceptibility to Mycobacterial Disease (MSMD, further discussed below) and immune-correlate studies, heritability of intermediate traits has been of recent interest. Twin and family studies of the heritability of IFN-γ, TNF-α and TGF-β production after *in vitro* antigen stimulation of whole blood demonstrate that TNF-α response is the most heritable of the three at 69%(Stein, Guwatudde

et al. 2003, Cobat, Gallant et al. 2010). Twin studies of the heritability of tuberculin reactivity after BCG vaccination estimate heritability at 28% (Sepulveda, Heiba et al. 1994). Understanding the genetic correlates of intermediate traits may therefore be an additional approach at delineating susceptibility pathways.

1.4.2 Mendelian Susceptibility to Mycobacterial Disease (MSMD)

Mendelian susceptibility to mycobacterial disease (MSMD) is an established clinical disorder described in more than a hundred individuals (Cottle 2011). The pattern of heritability differs between families and a variety of different genetic defects have been described in each. There are six genes in which one or more defects have been described in individuals with MSMD: *IFNGR1* (6q23-q24), *IFNGR2* (21q22.11) *STAT1* (2q32.2), *IL12RB1* (19q13.11), *IL12B* (5q33.3), all in the IFN-γ/IL12 pathway and *CYBB* (Xp11.4), on the X-chromosome (Casanova and Abel 2004, Kumararatne 2006, Cottle 2011). In addition there has been a case report of a patient with a defect in *TYK2* (19p13.2) with hyper-IgE syndrome and atypical mycobacteriosis (Minegishi, Saito et al. 2006). More recently two patients with susceptibility to mycobacterial but not viral disease were shown to have defects in *ISG15* that resulted in impaired IFN-γ production by NK cells and T-cells(Bogunovic, Byun et al. 2012). Taken together these studies strongly implicate the IFN-γ pathway as central to protection from mycobacterial disease. However, in almost all cases, susceptibility is to non-tuberculous mycobacterial disease as opposed to disease with *M.TB.* Therefore, the consequent immune defect may either be uninvolved in M.TB protection or, as is more likely, protection

from M.TB requires multiple redundant pathways and disabling a single node is therefore insufficient to cause disease.

1.4.3 Candidate gene studies

Studies to define the specific genetic modulators of host susceptibility to TB have been undertaken for several decades using genetic tools as they developed: linkage studies, microsatellite based genome scans, candidate gene polymorphism typing etc. Animal studies have played a major role either in the initial identification of gene or gene-regions involved in TB susceptibility and in the functional validation of some regions. In total, more than 30 genes have been implicated in TB susceptibility but results are largely inconclusive due to a combination of study design limitations, probable population specific polymorphisms and the differences between mouse and human genetic control of immunity. The genes that have been implicated, chromosomal location and their functional plausibility are tabulated in Table 2.

Table 2: Genes associated with TB disease in humans from candidate gene studies

Gene	Chromosomal location	Function	References
BTNL2	6p21.3	HLA-II related gene within which variants are known to be associated with sarcoidosis, also a pulmonary disease with granulomas.	(Moller, Kwiatkowski et al. 2007, Lian, Yue et al. 2010)
CCL2 (MCP-1)	17q11.2-12	Involved in recruitment of monocytes and probable granuloma formation	(Flores-Villanueva, Ruiz-Morales et al. 2005, Tamouza and Labie 2006, Buijtels, van de Sande et al. 2008, Alagarasu, Selvaraj et al. 2009, Moller, Nebel et al. 2009, Thye, Nejentsev et al. 2009, Xu, Xie et al. 2009, Yang, Zhuang et al. 2009, Ganachari, Ruiz-Morales et al. 2010, Ben-Selma, Harizi et al. 2011, Hussain, Ansari et al. 2011, Naderi, Hashemi et al. 2011, Arji, Busson et al. 2012, Feng, Flores-Villanueva et al. 2012, Mishra, Poojary et al. 2012, Velez Edwards, Tacconelli et al. 2012)
CCL3L1	17q21.1	Chemokine ligand 3-like 1	(Mamtani, Mummidi et al. 2011)
CCL5	17q11.2.-12	Chemokine for T-cells, monocytes, eosinophils	(Jamieson, Miller et al. 2004, Chu, Tam et al.

(RANTES)		and basophils. In animal models CCL5 appears to be important for granuloma formation.	2007, Sanchez-Castanon, Baquero et al. 2009, Ben-Selma, Harizi et al. 2011, Selvaraj, Alagarasu et al. 2011, Mishra, Poojary et al. 2012)
CCR2	3p21.31	Receptor for CCL2	(Alagarasu, Selvaraj et al. 2009, Mishra, Poojary et al. 2012)
CCR5	3p21.31	Receptor for CCR5	(Mamtani, Mummidi et al. 2011)
CD120b (TNFRSF1B)	1p36.2	Receptor for TNF cytokines and important protective anti M.TB responses.	(Ehlers, Kutsch et al. 2000, Moller, Flachsbart et al. 2010, Mokrousov, Wu et al. 2011)
CD206 (MRC1)	10p12.33	Uptake of M.TB by macrophages is partially regulated by MRC1 and engagement modulates phagosome-lysosome fusion.	(Zhang, Jiang et al. 2012)
CD209 (DC-SIGN)	19p13	Mediates phagocytosis of M.TB by DC's and is involved in receptor signaling and trafficking of ligand particles. Meta-analysis suggests no association except for one variant in Asians(Chang, Deng et al. 2012).	(Barreiro, Neyrolles et al. 2006, Ben-Ali, Barreiro et al. 2007, Olesen, Wejse et al. 2007, Vannberg, Chapman et al. 2008, Selvaraj, Alagarasu et al. 2009, Kobayashi, Yuliwulandari et al. 2011, Ogarkov, Mokrousov et al. 2012)
CD40	20q12-13.2	CD40 is a member of the TNF family of receptors	(Campbell, Sabeti et al. 2003, Lazarevic, Myers et al. 2003)

CISH	3p21.3	Cytokine responsive gene and one of many suppressors of cytokine signaling (SOCS)	(Khor, Vannberg et al. 2010, Wang and Wang 2010)
CXCL10 (IP-10)	4q21	Chemokine for multiple cell types and involved in NK cell and T-cell recruitment	(Tang, Fan et al. 2009)
HLA (mainly class II)	6p21.3	Known pleiotropic locus for infectious disease and likely function through peptide presentation to CD4+ or CD8+ T-cells or NK cells.	(Singh, Mehra et al. 1983, Singh, Mehra et al. 1984, Chandanayingyong, Maranetra et al. 1988, Hafez, el-Fiky et al. 1992, Sanjeevi, Narayanan et al. 1992, John, Murugesan et al. 1995, Pospelov, Matrakshin et al. 1996, Ravikumar, Dheenadhayalan et al. 1999, Dubaniewicz, Lewko et al. 2000, Vejbaesya, Chierakul et al. 2002, Amirzargar, Yalda et al. 2004, Balamurugan, Sharma et al. 2004, Kim, Park et al. 2005, Dubaniewicz, Dubaniewicz-Wybieralska et al. 2006, Lombard, Dalton et al. 2006, Selvaraj, Raghavan et al. 2008, Raghavan, Selvaraj et al. 2009, Lucena-Silva, Baliza et al. 2010, Magira, Papasteriades et al. 2012, Shahsavar, Mousavi et al. 2012)
IFN-γ	12q14	Cytokine crucial for protection against	(Rossouw, Nel et al. 2003, Tso, Ip et al. 2005,

		intracellular organisms. Meta-analysis suggest a single variant reproducibly associated with reduced susceptibility(Pacheco, Cardoso et al. 2008, de Albuquerque, Rocha et al. 2012).	Etokebe, Bulat-Kardum et al. 2006, Vidyarani, Selvaraj et al. 2006, Hwang, Kim et al. 2007, Moran, Ma et al. 2007, Sahiratmadja, Baak-Pablo et al. 2007, Sallakci, Coskun et al. 2007, Amim, Pacheco et al. 2008, Anuradha, Rakh et al. 2008, Ding, Li et al. 2008, Leandro, Rocha et al. 2009, Trajkov, Trajchevska et al. 2009, Vallinoto, Graca et al. 2010, Wang, Tang et al. 2010, Ansari, Hasan et al. 2011, Ben Selma, Harizi et al. 2011, Hashemi, Sharifi-Mood et al. 2011)
IFNGR1	6q23.3	Receptor for IFN-γ	(Bernard 1976, Donald 1998, Johnson, Braden et al. 2000, Bayele, Peyssonnaux et al. 2007, Homolka, Post et al. 2008, Magana-Arachchi, Perera et al. 2010, Oswald-Richter, Sato et al. 2010, Sharma, Rathored et al. 2010, Ben-Selma, Harizi et al. 2011, Ben-Selma, Ben-Abderrahmen et al. 2012)
IL-10	1q31-32	Suppresses proinflammatory responses but may also block phagosome maturation in	(Oh, Yang et al. 2007, Ates, Musellim et al. 2008, Pacheco, Cardoso et al. 2008, Thye,

		macrophages permitting latency. Meta-analysis of the three reported polymorphisms show no association overall, but protection amongst Caucasions(Zhang, Chen et al. 2011).	Browne et al. 2009, Ben-Selma, Harizi et al. 2011, Ben-Selma, Ben-Abderrahmen et al. 2012)
IL-12B	5q31.1-33.1	Induces IFN-γ production by T-cells via IL12 receptor signaling. Involved in MSMD.	(Ma, Reich et al. 2003, Tso, Lau et al. 2004, Morahan, Kaur et al. 2007, Prabhu Anand, Selvaraj et al. 2007, Sahiratmadja, Baak-Pablo et al. 2007, Wang, Tang et al. 2010, Morris, Edwards et al. 2011)
IL-18	11q22.2-22.3	Proinflammatory cytokine involved in IFN-γ production in synergy with IL12.	(Harishankar, Selvaraj et al. 2007, Lee, Choi et al. 2011, Taheri, Hashemi-Shahri et al. 2012)
IL-1B	2q14	Proinflammatory cytokine that triggers multiple additional inflammatory responses	(Awomoyi, Charurat et al. 2005, Gomez, Camargo et al. 2006, Taype, Shamsuzzaman et al. 2010, Bose and Jha 2012)
IL-6	7p21	Pro and inflammatory cytokine that mediates macrophage and T-cell differentiation.	(Zembrzuski, Basta et al. 2010, Lim, Chen et al. 2011)
IL-8	4q13-21	Chemotactic cytokine involved in recruitment of inflammatory cells to sites of inflammation	(Ma, Reich et al. 2003, de Wit, van der Merwe et al. 2011)
iNOS	17q11.2-q12	Involved in synthesis of Nitric Oxide (NO), an effector molecule against TB. Probably more	(MacMicking, North et al. 1997, Gomez, Anaya et al. 2007, Qu, Tang et al. 2007, Leandro,

		important in mice than humans due to differences in macrophage NO production.	Rocha et al. 2009, Velez, Hulme et al. 2009, Velez, Hulme et al. 2009)
Lta4h	12q22	Involved in leukotriene biosynthesis and modulates inflammatory TNF production	(Tobin, Vary et al. 2010, Curtis, Kopanitsa et al. 2011, Tobin, Roca et al. 2012)
MBL	10q11.2	Lectin receptor produced as an acute phase protein by the liver. Recognizes surface carbohydrates and results in opsonisation that enhances complement dependent and independent phagocytosis. Meta-analysis results suggest no association(Denholm, McBryde et al. 2010).	(Mombo, Lu et al. 2003, Soborg, Madsen et al. 2003, El Sahly, Reich et al. 2004, Alagarasu, Selvaraj et al. 2007, Cosar, Ozkinay et al. 2008, Solgun, Tastemir et al. 2011, Singla, Gupta et al. 2012)
NOD2 (CARD15)	16q21	Cytoplasmic sensor protein that binds strongly to N-glycolyl muramyl dipeptide (MDP) from M.TB	(Stockton, Howson et al. 2004, Austin, Ma et al. 2008, Pan, Dai et al. 2012, Zhao, Jiang et al. 2012)
P2RX7	12q24	Mediates ATP induced killing of mycobacteria within macrophages	(Li, Campbell et al. 2002, Mokrousov, Sapozhnikova et al. 2008)
SFTPA1 and SPTPA2	10q22.3	Calcium dependent lectin receptor that regulates expression and function of TLR2 and TLR4	(Madan, Saxena et al. 2002)
SLC26A1	4p16.3	Divalent cation transporter across phagosomal	(Blackwell, Barton et al. 1994, Liu, Fujiwara et

		membranes. Meta-analysis (Li, Zhang et al. 2006) suggests that polymorphisms may be susceptibility variants in Africans and Asians but not Caucasians.	al. 1995, Vidal, Tremblay et al. 1995, Blackwell, Black et al. 1997, Bellamy, Ruwende et al. 1998, Borgdorff 1998, Miles and McBride 1998, Cervino, Lakiss et al. 2000, Gao, Fujishima et al. 2000, Bellamy 2002, Abe, Iinuma et al. 2003, Kim, Lee et al. 2003, Hoal, Lewis et al. 2004, Zhang, Shao et al. 2005, Hsu, Chen et al. 2006, Li, Zhang et al. 2006)
SP110	2q37.1	Nuclear protein possibly involved in myeloid cell differentiation. Able to activate gene transcription. Meta-analysis suggests no association but may be complex gene-gene/ tissue-specific association(Lei, Zhu et al. 2012).	(Tosh, Campbell et al. 2006, Babb, Keet et al. 2007, Szeszko, Healy et al. 2007, Cong, Li et al. 2010, Abhimanyu, Jha et al. 2011, Liang, Zhao et al. 2011, Png, Alisjahbana et al. 2012)
TIRAP	11q24.2	Adaptor protein involved in TLR2 signaling pathway. Meta-analysis suggests no association exists (Miao, Li et al. 2011).	(Hawn, Dunstan et al. 2006, Khor, Chapman et al. 2007, Dissanayeke, Levin et al. 2009, Hamann, Kumpf et al. 2009, Selvaraj, Harishankar et al. 2010, Taype, Shamsuzzaman et al. 2010, Zhang, Xue et al. 2011)

TLR1	4p14	Forms a heterodimer with TLR2	(Ma, Liu et al. 2007, Uciechowski, Imhoff et al. 2011, Sun, Song et al. 2012)
TLR2	4q32	Mediates strong proinflammatory responses following binding by M.TB cell wall components	(Ben-Ali, Barbouche et al. 2004, Thuong, Hawn et al. 2007, Chen, Hsiao et al. 2010, Ma, Xie et al. 2010, Motsinger-Reif, Antas et al. 2010, Velez, Wejse et al. 2010, Xue, Zhao et al. 2010, Dalgic, Tekin et al. 2011)
TLR4	9q33.1	Recognizes components of M.TB and results in NFkB proinflammatory cascades initiation	(Abel, Thieblemont et al. 2002, Newport, Allen et al. 2004, Ferwerda, Kibiki et al. 2007, Speletas, Kalala et al. 2009, Velez, Hulme et al. 2009, Najmi, Kaur et al. 2010, Pulido, Leal et al. 2010, Xue, Zhao et al. 2010, Zaki, Leung et al. 2012)
TLR8	Xp22	Recognizes intracellular components of M.TB	(Davila, Hibberd et al. 2008, Dalgic, Tekin et al. 2011)
TLR9	3p21.3	Recognizes intracellular components of M.TB	(Velez, Wejse et al. 2010, Kobayashi, Yuliwulandari et al. 2012)
TNF	6p21.3	Involved in protective responses against TB through many different mechanisms and as a proinflammatory cytokine. Meta-analysis does	(Bean, Roach et al. 1999, Selvaraj, Sriram et al. 2001, Correa, Gomez et al. 2005, Amirzargar, Rezaei et al. 2006, Stein,

		not support presence of association(Pacheco, Cardoso et al. 2008).	Zalwango et al. 2007, Vejbaesya, Chierakul et al. 2007, Ates, Musellim et al. 2008, Merza, Farnia et al. 2009, Trajkov, Trajchevska et al. 2009, Fan, Wang et al. 2010, Ma, Xie et al. 2010, Sharma, Rathored et al. 2010, Ben-Selma, Harizi et al. 2011)
VDR	12q13.11	Vitamin D has been shown to suppress M.TB growth in vitro through production of the antimicrobial peptide, cathelicidin and possibly via modulation of TLR signaling. Polymorphisms associated with TB in meta-analysis(Zhao, Zhang et al. 2009).	(Selvaraj, Kurian et al. 2000, Selvaraj, Narayanan et al. 2000, Wilkinson, Llewelyn et al. 2000, Bornman, Campbell et al. 2004, Liu, Cao et al. 2004, Roth, Soto et al. 2004, Lewis, Baker et al. 2005, Babb, van der Merwe et al. 2007, Motsinger-Reif, Antas et al. 2010, Singh, Gaughan et al. 2011)

Several genes have been more well studied than others such as the HLA-DR locus because of its role in antigen presentation and the pleiotropic nature of HLA associations with infectious diseases; the *SLC11A1* gene (*NRAMP1*) because of it being the human homologue to the lsh/lty/bcg locus that affects mycobacterial susceptibility in mice; the vitamin D receptor because of functional and epidemiologic evidence for the importance of vitamin D in macrophage function and TB susceptibility respectively(Wilson, Fineberg et al. 1995, Berrington, Green et al. 2000, Luong and Nguyen 2011); *NOS2A* for its role in controlling inducible nitric oxide synthesis-an important macrophage response to infection in mice; and a variety of cytokines known to be important in anti-bacterial immunity. How substantial a role these genes play in TB susceptibility and which of these genes are true correlates of susceptibility remain to be determined. Many of the studies in which these genes were first implicated in susceptibility have serious limitations such as unmeasured population stratification, multiple-hypothesis testing and imprecise phenotype ascertainment.

With the advent of array-based genotyping technologies, unbiased searches for genetic correlates of susceptibility have been made possible: first by the use of case-control genome wide association scans and secondly by the use of microarrays to measure transcriptomic signatures across the genome. These approaches allow detection and correction for population stratification and strict correction for multiple hypothesis testing. Genome-wide gene silencing have also been useful alongside classical induced mutagenesis screens.

1.4.4 Genome-wide association studies

Five genome-wide association studies for pulmonary TB (pTB) have been reported(Thye, Vannberg et al. 2010, Mahasirimongkol, Yanai et al. 2012, Png, Alisjahbana et al. 2012, Thye, Owusu-Dabo et al. 2012). The Hill and Meyer groups previously studied HIV-negative adults with pTB in West Africa and compared them to healthy controls. In collaboration they reported a variant, rs4331426 (Chr18q 11.2, gene desert) to be associated with pTB (OR 1.19, 95%CI 1.13-1.27, $p=6.8x10^{-9}$) after meta-analysis(Thye, Vannberg et al. 2010). Imputation of variants across the genome and genotyping in 13,859 cases and 8,821 controls implicated a variant rs2057178 (Chr11p13 near *WT1*) as a protective variant (OR 0.77, 0.80, 0.84 and 0.91 in Ghanaians, Gambians, Indonesians and Russians respectively)(Thye, Owusu-Dabo et al. 2012). The function of these variants is not known.

More recently genome-wide association scans in 393 cases and 1,255 controls from Thailand and Japan, suggested at least one new region tagged by rs6071980 in *HSPEP/MAFB* to be associated with pTB amongst individuals <40 years of age(Mahasirimongkol, Yanai et al. 2012). This study's failure to identify many replicable associations, even amongst a population with less sparse linkage disequilibrium than Africans confirms that many common variants with large effect sizes are unlikely to explain pTB risk and larger sample sizes are necessary. Their findings are also interesting for the fact that the association is observed only in in individuals <40 but not amongst a further 906 cases and 1,786 controls older than 40 years; likely due to a gene-environment interaction such as exposure to TB.

Thirdly, a small two stage genome wide scan was reported in which 108 cases and 115 Indonesian controls were genotyped at ±100k SNP's(Png, Alisjahbana et al. 2012), then 2,453 SNPs with the lowest p-value for association were replicated in 1,189 further Indonesians and 3,760 Russians. Their results suggest that variants at 8 novel loci may be involved in TB susceptibility *JAG1, DYNLRB2, EBF1, TMEFF2, CCL17, HAUS6, PENK, TXNDC4* but none of the eight reported associations achieved levels for genome-wide statistical significance.

A fourth study, performed amongst the South African coloured population, was reported in 2014, in which 642 cases with pulmonary TB and 91 controls were genotyped and an association study performed. Although this study did not identify any new genome-wide-significant loci associated with disease, it replicated the Chr11p13 locus first reported in the West African study as being associated with disease.

Finally, a GWAS of 5,530 cases and 5,607 controls perfomed amongst Russian adults was reported earlier this year. In this study, the Chr11p13 locus was replicated, and a new variant in an intron of *ASAP1*, was shown to be associated with TB disease and demonstrated to alter *ASAP1* expression in DC.

Taken together these genome wide studies demonstrated that there are likely few common variants with large effect sizes that explain TB susceptibility

across different populations. It is plausible that different genetic effects prevail in different ethnic groups, due to host variation but also, as is discussed later, different strains of circulating *M.TB*. Larger studies, or meta-analyses may, however yield new variants. And newer methods, such as exome sequencing, may allow assessment of the contribution of rare variants and genetic burden testing.

1.4.5 Transcriptomic studies

An alternative approach to identify genes and pathways involved in TB is to directly assess whether expression of specific genes are associated with different TB outcomes. Several groups of investigators have searched for transcriptomic signals of response to *in vitro M.TB* infection such as *M.TB* infection of macrophages (Thuong, Dunstan et al. 2008) and *M.TB* stimulated peripheral blood mononuclear cells (PBMC) (Lesho, Forestiero et al. 2011). Several *in vivo* studies have also been performed. Berry et al., (Berry, Graham et al. 2010) identified an 86 gene-transcript signature in peripheral blood that differentiated adults infected with TB from other inflammatory diseases. The signature was shown to plausibly arise from neutrophils and was markedly enriched for genes involved in type-I and type-II interferon pathways. The transcriptomic results were confirmed by a second independent study (Maertzdorf, Ota et al. 2011), which further delineated novel trans-ethnic transcriptomic signatures of active TB disease, such as activation of the FcGR1 signaling pathway. Recent evidence in mouse models and in humans demonstrates that IL-1 induces eicosanoids that inhibit type-1 interferons leading to enhanced bacterial containment and confirm a role for type-1 interferons in risk of TB(Mayer-Barber, Andrade et al. 2014).

Some studies have evaluated whether similar signatures explain how know risk factors like Diabetes Mellitus affect TB risk (Qu, Rentfro et al. 2012). Finally, at least one study has attempted to unite genotypic data with transcriptomic data in the form of an expression quantitative trait locus (eQTLs) analysis. Barreiro et al., (Barreiro, Tailleux et al. 2012) identified 198 SNP's that were associated with variation in gene expression levels in either untreated or MTB treated DC but not both. These eQTL were overrepresented amongst nominally significant GWAS results of pTB relative to expectations.

1.4.7 Genome-wide cell-based or animal screens

A third way at unbiased identification of specific genes involved in TB pathogenesis is by experimental silencing of gene expression in host cells and evaluation of the consequences to *in vitro M.TB* growth. Kumar et al., (Kumar, Nath et al. 2010) performed a genome wide small interfering RNA (siRNA) screen for genes and networks that regulate intracellular survival of *M.TB* in macrophages derived from the THP1 monocytic leukemia cell line. They confirmed their findings using primary human monocyte derived macrophages and assessed the robustness of findings by using different *M.TB* isolates. Their results show that genetic regulation differs by organism strain and by specific hosts, but their results also delineate some invariant host specific regulatory pathways. In particular, genes that affect the autophagy pathway are crucial for macrophage inhibition of mycobacterial growth. Targeting of genes in model organisms and assessment of the consequence on infection phenotypes is complementary to siRNA screens. Classical non-specific methods such as N-ethyl-N-nitrosourea (ENU)

mutagenesis are being supplemented with specific *in vivo* silencing technologies to identify and validate genes involved in TB. Tobin et al., (Tobin, Vary et al. 2010) used this approach in *Mycobacterium marinarum*-infected Zebrafish larvae that serve as a model for human TB. They identified leukotriene A4 hydrolase (lta4h) as a modifier of *in vivo* infection in a forward ENU genetic screen, validated this by direct inhibition of translation using morpholino oligonucleotides and demonstrated that lta4h polymorphisms in humans are associated with differential TNF-a and eicosanoid responses and survival in tuberculous meningitis.

1.4.8 Host-pathogen coevolution as a basis for immunity as a double-edged sword

Recent studies in which sequencing of M.TB isolates from diverse geographic locations was performed, suggest that M.TB likely left with humans when we migrated out of Africa around 70,000 years ago, and has remained a primary human pathogen (Figure 3)(Comas, Coscolla et al. 2013). This observation suggests that humans and M.TB co-evolved. Therefore, it is plausible that M.TB may have been a powerful selective force over the course of human history. In this light, several recent observations in which immune responses to TB are both protective and harmful may be indicative of the evolved mechanisms in M.TB to subvert the protective immune responses.

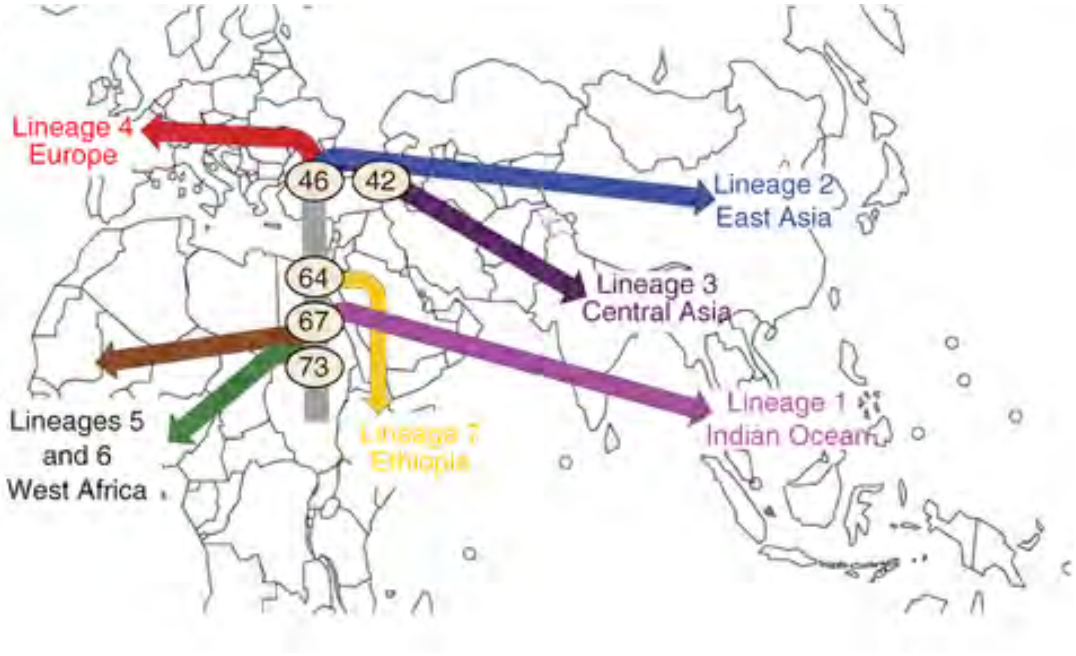


Figure 3: Phylogeographic and dating analysis of M.TB complex. Dating estimates (in KYA) for the origin of each of the seven lineages are shown and plotted according to origin and modern day occurrence. Reproduced with permission from Nature Publishing Group (License no. 3679050891364)(Comas, Coscolla et al. 2013).

Firstly, pathogen evolution studies have observed hyperconservation of antigenic epitopes, suggesting that *M.TB* may benefit from T-cells responding to these epitopes (Comas, Chakravartti et al. 2010) and yet a T-cell response to infection is probably necessary for protection and formation of an organized granuloma. Secondly, during the course of infection, M.TB appears to have evolved mechanisms to increase host CCL2 secretion and MMP9 expression resulting in enhanced macrophage recruitment enabling primary infection and granuloma organization. Thirdly, although IFN-y responses are thought to be necessary for microbicidal activity partly through induction of indoleamine 2,3-dioxygenase and consequent depletion of tryptophan, M.TB is relatively unique amongst pathogens in its ability to synthesize tryptophan. Finally, TNF is required for microbicidal activity of macrophages. Evolutionarily modern mycobacteria induce lower levels of TNF-α than ancient isolates suggesting

adaptation to ensure survival. However, excessive TNF appears to enhance macrophage necrosis by necroptosis and consequent externalization of replication competent mycobacteria.

A second implication of the host-pathogen coevolution is that geographic differentiation between mycobacteria into distinct lineages occurred together with evolution of host defense mechanisms. Therefore heterogeneity in association between susceptibility correlates and disease amongst different populations may plausibly be due to differences in lineage evolution. This may explain the relative absence of host genetic correlates in GWAS studies that replicate across different population groups. Moreover, host-pathogen coevolution may explain the relative success of a lineage that is introduced into a population that co-evolved with a heterologous lineage as is seen with the Beijing lineage in sub-Saharan Africa.

Finally, a third consequence of host-pathogen co-evolution is that the genetic factors which were selected for may come at the cost of altered risk for other diseases. In particular, enhanced immune responses may come at the cost of increased autoimmune disease. Although not yet shown for M.TB the architecture of inflammatory bowel disease has been shown to have been shaped by *M.leprae* (Jostins, Ripke et al. 2012).

1.5 Conclusions

In conclusion, TB is a disease of immense global importance. A more precise understanding of pathogenesis may allow new interventions for prophylaxis,

diagnosis or treatment of TB, all of which are desperately needed. An important contribution of genetics to TB susceptibility has been demonstrated through twin and family studies. Candidate gene studies and less biased screens have been used to identify genes that regulate immunity to TB. Using new genetic tools, the work summarized in this report aims to contribute to meeting the need for new insights into the contribution that genetics makes to the pathophysiology of TB.

Chapter Two. Discovery and validation of the ratio of monocytes:lymphocytes as a correlate of TB susceptibility

2.1 Introduction

In this chapter I describe three cohort studies in which I established the ML ratio as a predictor of TB disease risk. These studies have their origin in an unpublished observation where Fletcher et al., in a study of infants at risk for TB in Cape Town South Africa, noted that infants who developed TB by age two-years had an elevated ratio of myeloid to lymphoid transcripts at age 10-weeks. In clinical practice, monocytes and lymphocytes are enumerated routinely allowing examination of whether the transcript ratio was merely a reflection of cell frequency. Because blood count measures were not collected in Fletcher et al.'s study, I first examined whether the ML ratio predicted TB risk in a cohort of adult HIV-infected patients in whom I had previously reported high rates of early mortality due to TB and in whom extensive clinical profiling to identify cases of TB was performed(Naranbhai, Abdool Karim et al. 2012). The results of this study, including exploration of how HIV affects the ML ratio, are detailed in section 2.2. In section 2.3 I show how the ML ratio is of similar predictive value in both HIV-infected and HIV-uninfected infants who participated in a large randomized control trial (RCT) of IPT, in whom TB disease was a primary endpoint and therefore carefully ascertained. Finally, in section 2.4, I extend these results to pregnant women participating in a multi-centre RCT of methods to prevent mother-to-child transmission of HIV. These three cohort studies establish the ML ratio as a predictor of TB disease risk.

2.2 The Ratio of Monocytes to Lymphocytes in Peripheral Blood predicts the Risk of Incident Tuberculosis in HIV-Infected Adults Initiating Antiretroviral Therapy

2.2.1 Background

Tuberculosis (TB) is the leading cause of death in patients commencing combination antiretroviral therapy (cART) in sub Saharan Africa (Rana, Hawken et al. 2000, Cohen, Murray et al. 2010, Cox, Lukande et al. 2010), but practical methods to stratify risk in this population are lacking(Walzl, Ronacher et al. 2011). If individuals at risk for TB could be accurately identified, they could be targeted for preventative strategies such as chemoprophylaxis with IPT or vaccination. Risk factors for TB disease include immunosuppression, diabetes mellitus, smoking and the presence of latent *Mycobacterium tuberculosis* (*M.TB*) infection (LTBI) (Lawn and Zumla 2011). Tuberculin Skin Tests (TSTs) and the more recently developed Interferon-Gamma Release Assays (IGRAs) are poor at predicting which latently-infected individuals will subsequently go on to develop TB disease, particularly in HIV positive populations(Rangaka, Wilkinson et al. 2012). The need for targeting of groups at risk of TB is highlighted by the fact that Isoniazid Preventive Therapy (IPT) is most effective amongst TST positive patients(Lawn, Wood et al. 2010, Samandari, Agizew et al. 2011) as opposed TST negative individuals.

Gene expression studies have identified biomarkers of active TB disease that provide insight into immune mechanisms of disease and may be useful for improving the diagnosis of TB(Berry, Graham et al. 2010, Maertzdorf,

Repsilber et al. 2011). In an unpublished study, Fletcher et al. recently used whole-transcriptome microarrays to identify correlates of risk of TB disease up to 2 years before disease developed in infants who had been vaccinated with BCG at birth. Relative abundance of myeloid-specific gene transcripts and lymphoid-specific transcripts at 10 weeks of age were associated with subsequent TB after stratifying for BCG responses. The observation that the relative proportion of transcripts may be related to TB susceptibility is reminiscent of older studies relating to monocytes and lymphocyte cell numbers. Between 1921-1931, Florence Sabin and colleagues found that monocytes were increased following experimental infection of rabbits with *Mycobacterium bovis (M. bovis) (Cunningham, Sabin et al. 1925)*. They and others reported that the ratio of lymphocytes to monocytes in peripheral blood correlated with extent of disease in both rabbits(Sabin, Doan et al. 1926) and humans(Rogers 1928), although the numbers studied were small and the strength of the conclusions that could be reached in humans conceded to be modest. Finally, by experimentally reducing monocyte numbers in rabbits with a monocyte antiserum or conversely increasing the number of monocytes with A-3 phosphatide, and then challenging animals with *M.bovis*, they demonstrated that elevated monocyte:lymphocyte ratios were associated with enhanced lethality of mycobacterial infection(Doan and Sabin 1930). Lymphoid cells are thought to be the major effector cells in TB immunity and myeloid cells the major host cell for infection. Therefore the relative abundance may reflect a balance between effector and target cells. Amongst infants with altered myeloid:lymphoid transcript ratios in Fletcher et al.'s work however, HSC and inflammation associated transcripts were also altered.

Therefore an alternative explanation is that the relative abundance of these cell types could be a marker of hematopoietic parameters associated with TB. Collectively, these data support the hypothesis that a high or a low ratio of monocytes to lymphocytes (ML ratio hereafter) may be a correlate of risk for TB disease. I initially evaluated this hypothesis in a cohort of HIV-infected South African adults commencing cART followed prospectively.

2.2.2 Methods

Study design and setting

Data in this study were derived from several cohort studies or clinical trials(van Loggerenberg, Mlisana et al. 2008, Abdool Karim, Abdool Karim et al. 2010, Abdool Karim, Naidoo et al. 2010, Naranbhai, Abdool Karim et al. 2012) conducted at the at the Centre for the AIDS Program of Research in South Africa (CAPRISA) AIDS Treatment (CAT) clinic in Vulindlela, rural KwaZulu Natal or the CAPRISA eThekwini clinic in Durban, KwaZulu Natal. In each study, prospective follow-up with routine clinical assessments including specific assessment for Tuberculosis and routine measurement of monocyte and lymphocyte counts was performed. The UKZN Biomedical Research Ethic Committee approved the use of the data from each study.

Participants clinical care, TB diagnosis and data sources

For the primary analysis, participants were ambulatory, HIV-seropositive, cART-naive adults older than 18 years who enrolled in care and treatment between April 2004 and April 2011 at the CAPRISA CAT clinic in Vulindlela,

rural KwaZulu Natal(Naranbhai, Abdool Karim et al. 2012). Care for HIV and/or TB was provided per the South African Treatment guidelines of that time(Republic of South Africa Department of Health 2004, Republic of South Africa Department of Health 2009, Republic of South Africa Department of Health 2010). Briefly, patients who were clinically evaluated to have WHO stage IV HIV disease or had a CD4+ T-cell count <200 cells/μl were eligible for commencement on a cART regimen comprising of two nucleoside reverse transcriptase inhibitors (NRTI's) and one non-nucleoside reverse transcriptase inhibitor (NNRTI). During this period the South African guidelines did not support the routine use of IPT for patients on cART. For this analysis 1862 individuals who initiated cART and had at least one full blood count measurement available prior to cART were included. At enrolment all individuals underwent clinical examination including clinical screening for TB and other opportunistic infections. Individuals with active TB at enrolment were excluded. Participants were clinically evaluated at least twice prior to enrolment, and monthly after cART commencement. During the course of follow-up on cART, individuals with symptoms and/or signs of TB were assessed monthly and diagnosed per South African guidelines, which include clinical examination, sputum microscopy and/or culture and chest x-ray as indicated(Republic of South Africa Department of Health 2009).

The replication cohort was a cohort of cART-naïve adults who self-referred for care at the Prince Cyril Zulu Communicable Diseases Centre, a chest clinic which provides care for the eThekwini (Durban) municipality. Participants for this study were those in whom TB was clinically, microbiologically or

radiologically excluded at enrolment.

For analyses of the stability of the ML ratio in HIV uninfected participants in the same population, we obtained pseudonymised data from HIV uninfected women participants of the CAPRISA004 clinical trial(Abdool Karim, Abdool Karim et al. 2010), a phase IIB, placebo-controlled, double-blinded randomized clinical trial of topical 1% Tenofovir gel. This gel does not have any known effects on haematological indices when applied topically. Nevertheless, routine safety monitoring, including full blood count, was performed at screening (0 months) and at 3,12 and 24 months. For these analyses individuals who had signs or symptoms of TB were excluded.

For analyses of the impact of HIV acquisition and disease progression on ML ratio, data from the CAPRISA002 study(van Loggerenberg, Mlisana et al. 2008), a prospective cohort of HIV seroconverters, was used. In this study, full blood counts were performed at weekly intervals for 3 weeks, fortnightly until 3 months, monthly until 1 year post-enrollment and every 3 months thereafter. Participants who started cART were censored at the time of commencement.

For analyses of the impact of anti-tuberculous therapy on the ML ratio, data from the CAPRISA003/SAPiT trial were used(Abdool Karim, Naidoo et al. 2011). In this randomized controlled trial, HIV and TB co-infected participants received anti-tuberculous therapy followed by cART at one of three timepoints defined by randomization. Full blood count measures were performed prior to randomization and monthly throughout follow-up.

In all cohorts, data was collected in real-time during prospective follow-up on standardized case report forms (CRFs) and captured electronically using the DataFax system (Clinical DataFax Systems, USA) by at least two data encoders.

Full differential blood counts

Full blood counts (FBC) of peripheral blood collected in ethylene-diameno tetra-acetic acid containing (EDTA) tubes (Becton Dickinson) were performed by one of two clinical diagnostic laboratories each using a five-part differential hematology analyser (Sysmex Model XS, Hamburg, Germany). Full blood count measurement was subject to strict quality assurance procedures including twice-daily high and low internal quality control, fortnightly quality controls through the Thistle external quality assurance service (Thistle QA, Johannesburg, South Africa) and annual quality assurance as part of the College of American Pathologists QC scheme. Both laboratories are accredited by the South African National Accreditation System (SANAS) in accordance with international standards ISO 17025:2005 and ISO 15189/2007.

Statistical methods

To calculate the ML ratio, the absolute monocyte count was divided by the absolute lymphocyte count. For the primary analysis of TB incidence, a survival analysis was performed. Study duration was calculated as the time from cART initiation to the first diagnosed episode of TB or death, withdrawal

from the study, transfer to another facility or last visit date (for patients remaining in follow-up), whichever occurred first. The log rank test was used to compare survival curves stratified by ML ratio. To test the association between the ML ratio and TB incidence, a cox proportional hazards regression model was performed. Covariates included in the model are given in the text or table. For the analysis of continuous variables the Kruskal-Wallis rank sum test was used. All statistical tests were two-sided and $p<0.05$ was considered statistically significant. Poisson approximations were used to calculate confidence intervals (CIs) for estimations of the incidence rate. Boostrapped estimates of the adjusted HR across the ML ratio continuum were generated with the '*boot*' package. Statistical analyses were performed in R (The R Foundation for Statistical Computing) using the following packages epiR, survival, date and mfp.

2.2.3 Results

We assessed whether the ratio of peripheral blood monocytes to lymphocytes measured prior to commencement of cART was predictive of the development of TB disease whilst on cART in a prospective cohort of 1862 HIV-infected adults in South Africa. Baseline characteristics of this cohort are detailed in Table 3. Notably, the cohort was predominantly female (68.1%), was relatively young (mean age 24.5 years) and 11.1% of individuals disclosed a previous episode of treated TB. The median CD4+ T-cell count at baseline was 120 cells/ul. The median follow-up for this cohort was 17 months but extended up to seven years. The incidence of TB, diagnosed per South African national

guidelines, during follow up was 18.79 cases (95% CI 14.71-23.66) per 1000 patient-years on treatment, consistent with comparable cohorts (Lawn, Badri et al. 2005, Dembele, Saleri et al. 2010).

The ratio of monocytes: lymphocytes is associated with TB amongst adults starting cART

Based on Doan and Sabin's observation that extremes of the ML ratio were important for rabbit immunity to *M. bovis*(Doan and Sabin 1930), we stratified participants into categories derived from the distribution of baseline pre-cART ML ratio in the entire cohort. Participants with an ML ratio <5th percentile, between the 5th and 95th percentile or >95th percentile were similar for age and the proportion with a previous history of TB, but males, individuals with lower CD4 counts and higher WHO staging were overrepresented in the higher ML (>95th percentile) group relative to the other two groups. This observation is concordant with expectations of individuals with more severe lymphopenia having higher ML ratios (Table 3).

Table 3: Baseline characteristics for the cohort of 1862 adults commencing cART in the study cohort. Data are shown for the overall cohort and by category of ML ratio within the cohort.

	Overall	<5th percentile (n=93)	5-95th percentile (n=1671)	>95th percentile (n=98)	p-value§
Mean age (years)	24.5	24.5	24.6	22.6	0.228
Gender (n, % female)	1268/1862 (68.1%)	67 (72.04%)	1150 (68.8%)	52 (53.1%)	**0.002**
Median baseline CD4+ T-cell count (cells/ul, IQR)	120 (61-180)	138 (102-182)	124 (65-182)	36 (17-98)	**<0.0001**
WHO stage (n,%)*					
1	355 (19.34%)	26 (28.0%)	320 (19.2%)	9 (9.2%)	**0.004**
2	490 (26.69%)	22 (23.7%)	448 (26.8%)	20 (20.4%)	0.28
3	831 (45.26%)	41 (44.1%)	739 (44.2%)	51 (52.1%)	0.38
4	160 (8.72%)	3 (3.2%)	139 (8.3%)	18 (18.4%)	**0.005**
Previous history of TB (n, %)**	206 (11.11%)	10/93 (10.75%)	190 (11.37%)	6 (6.12%)	0.27

*26 patients missing baseline WHO staging

**7 patients missing baseline TB history

§ Kruskal-Wallis test

The pre-cART ML ratio was associated with significantly different TB-free survival probabilities (Figure 4, p=0.007 with the log rank test). After excluding cases diagnosed within 90 days of cART inititiation, which may plausibly represent unmasked TB, the association remained significant (p=0.02 with the log rank test). The overall incidence of TB amongst individuals with ML ratio<5th percentile was 32.61 cases (95% CI 15.38-61.54) per 1000 patient years, and for individuals with ML ratios>95th percentile was 51.80 cases (95% CI 23.10-101.71) per 1000 patient years, in comparison with an incidence rate of 16.36 (95% CI 12.39-21.23) per 1000 patient-years for individuals with ML ratio between the 5th and 95th percentiles (Table 4).

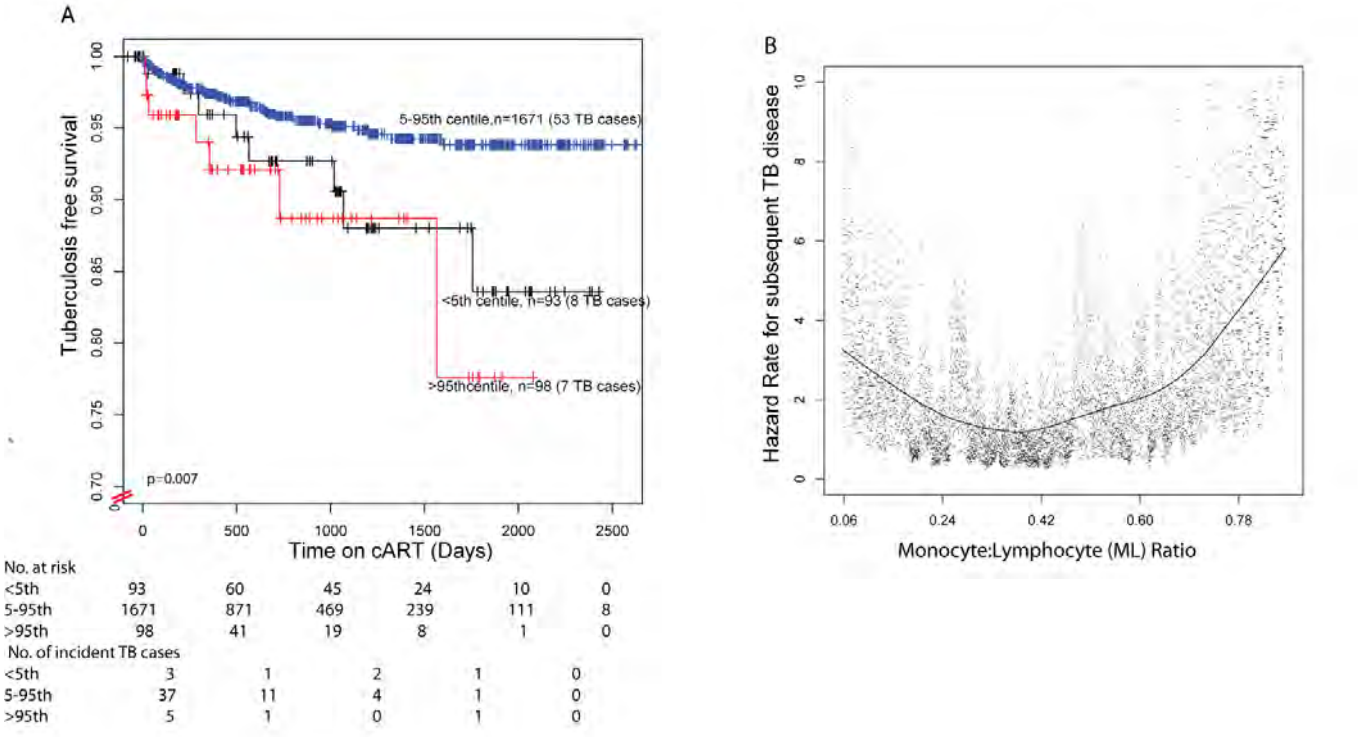


Figure 4: Kaplan-Meier estimates of probability of TB-disease free survival for individuals commencing cART, by category of baseline ML ratio (A) and bootstrapped HR estimates of TB disease across ML ratio continuum (B). HR's are adjusted for age, WHO stage and past history of TB, each dot denotes one estimate with LOESS curve overplotted.

Table 4: Cox-proportional hazards modeling of tuberculosis risk by pre-cART ML ratio category, n=1862.

Explanatory Variable	N	No. of patients developed TB/patient-years	Incidence Rate per 1000 patient years (95% CI)	Unadjusted HR (95% CI)	p-value	*Adjusted HR (95% CI)	p-value
ML ratio between 5th-95th percentile for patients commencing cART	1671	53/3238.8	16.36 (12.39-21.23)	1.00 (REFERENCE)	-	1.00 (REFERENCE)	-
ML ratio <5th percentile for patients commencing cART (0.1489619)	93	8/245.3	32.61 (15.38-61.54)	1.98 (0.945-4.14)	0.071	2.09 (0.99-4.39)	0.053
ML ratio >95th percentile for patients commencing cART (0.8746429)	98	7/135.1	51.80 (23.10-101.71)	**2.60 (1.19-5.68)**	**0.017**	**2.63 (1.18-5.83)**	**0.018**
ML ratio <5th or >95th percentile for patients commencing cART	191	15/380.5	39.44 (23.03-63.39)	**2.40 (1.35-4.25)**	**0.003**	**2.47 (1.39-4.40)**	**0.002**

*adjusted for sex, WHO stage, CD4 count, past history of TB

Overall, the incidence of TB amongst individuals with an ML ratio less than the 5th or greater than the 95th percentile was 39.44 (95% CI 23.03-63.39), HR 2.40 (95% CI 1.35-4.25, p=0.003). Neither the monocyte count nor lymphocyte count alone were predictive of TB on formal testing (Figure 5 and Table 5).

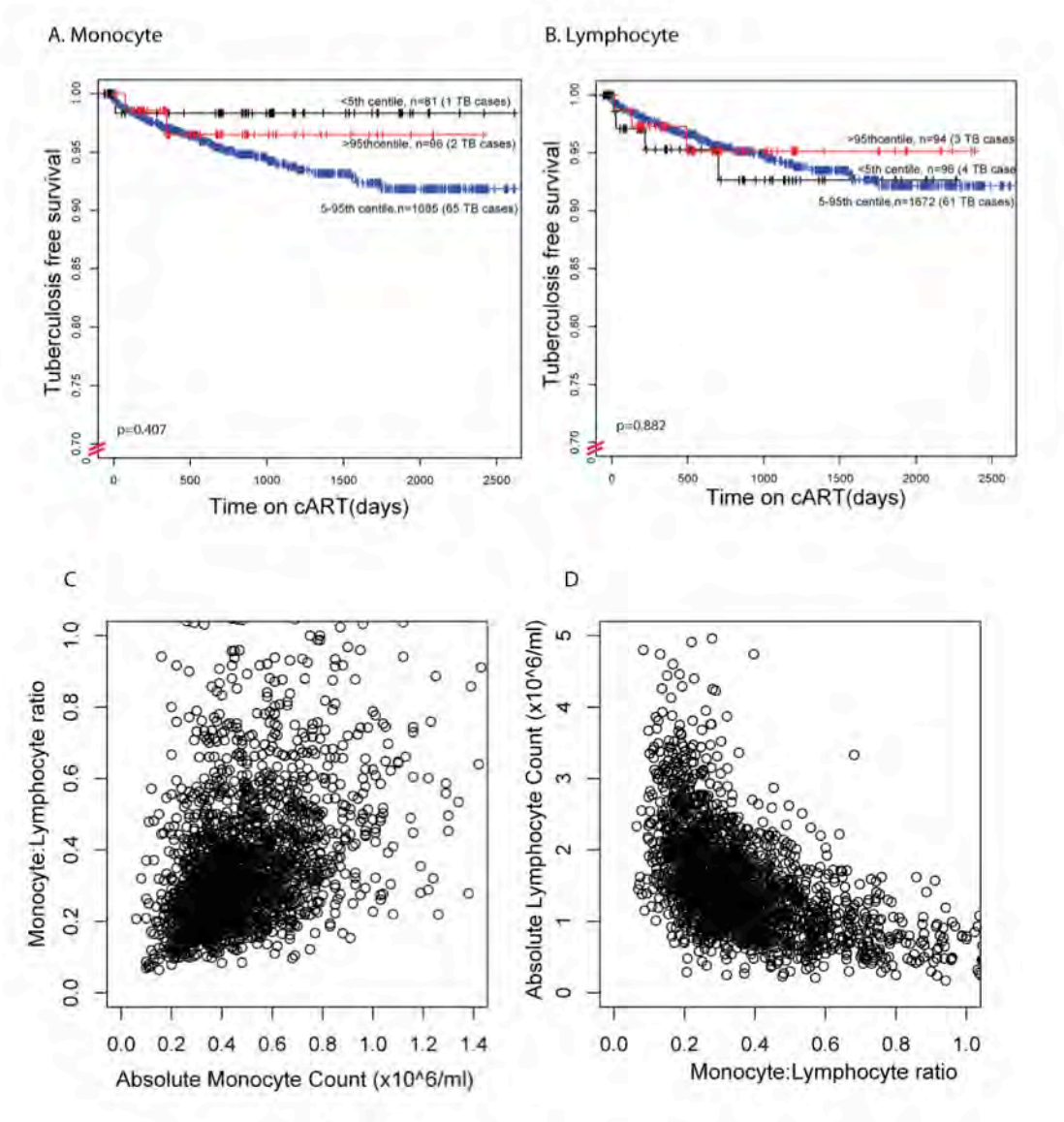


Figure 5: Kaplan-Meier estimates of probability of TB-disease free survival for individuals commencing cART, by centile category of baseline monocyte (A) or lymphocyte (B) count showing that neither the monocyte count nor the lymphocyte count are associated with TB-disease free survival for individuals commencing cART. Both the monocyte count (**C**) and the lymphocyte count (**D**) contribute to the Monocyte:Lymphocyte (ML) ratio.

Table 5: Cox-proportional hazards modeling of tuberculosis risk by pre-cART monocyte or lymphocyte category, n=1862.

Explanatory Variable	N	No. of patients developed TB/patient-years	Incidence Rate per 1000 patient years (95% CI)	Unadjusted HR (95% CI)	p-value	*Adjusted HR (95% CI)	p-value
Monocyte count 5th-95th percentile for patients commencing cART	1685	65/3301·04	19·69 (15·33-24·93)	1·00 (REFERENCE)	-	1·00 (REFERENCE)	-
Monocyte count <5th percentile for patients commencing cART	81	1/186·4	5·37 (0·49-25·02)	0·31 (0·04-2·21)	0·24	0·30 (0·04-2·18)	0·24
Monocyte count >95th percentile for patients commencing cART	96	2/179·23	15·16 (3·02-48·61)	0·68 (0·17-2·79)	0·60	0·65 (0·15-2·65)	0·54
Lymphocyte count 5th-95th percentile for patients commencing cART	1672	61/3291·4	18·53 (14·31-23·64)	1·00 (REFERENCE)	-	1·00 (REFERENCE)	-
Lymphocyte count <5th percentile for patients commencing cART	96	4/164·73	24·28 (8·12-57·73)	1·28 (0·47-3·52)	0·633	1·32 (0·47-3·69)	0·60
Lymphocyte count >95th percentile for patients commencing cART	94	3/163·2	18·38 (5·08-49·04)	0·94 (0·30-3·0)	0·92	0·90 (0·28-2·89)	0·85

*adjusted for sex, WHO stage, CD4 count, past history of TB

To validate these results the association between the baseline pre-cART ML ratio and risk of TB whilst on cART was assessed in a replication cohort of 345 adults commencing cART at an urban treatment site implementing the same care and treatment protocol as in the primary cohort. The overall incidence of TB after cART commencement in this cohort was 21.62 (95% CI 12.11-35.92) per 1000 patient-years. Consistent with the finding in the primary cohort, pre-treatment ML ratios less than the 5th or greater than the 95th percentile were independently associated with significantly greater incidence and hazards of TB (87.68, 95% CI 24.27-233.91, vs. 17.64, 95% CI 9.05-31.29 per 1000 patient-years; adjusted HR 3.94, 95% CI 1.05-14.73, p=0.04). To refine estimates of the association between the ML ratio and incident TB, we generated bootstrapped estimates of effect across the ML ratio continuum (Figure 4B). These data demonstrate a J-shaped association between ML ratio and incident TB with both low and high ML ratios being associated with subsequent TB

The association between the ratio of monocytes: lymphocytes and TB risk is not explained by confounding by conventional risk factors for TB.

Greater hazards of TB in those with highest ML ratios could potentially have been explained by known risk factors for TB in this population(Lawn, Badri et al. 2005, Seyler, Toure et al. 2005) which include a CD4+ T-cell count below 100 cells/μl, WHO stage three or four clinical disease, male sex and/or a past history of TB. To address these potential confounders, these covariates were included in a Cox proportional hazards model (Table 4 and 5). After adjusting for covariates, a pre-treatment ML ratio less than the 5th or greater than the

95th percentile remained independently associated with significantly greater hazards of TB (adjusted HR 2.47 95% CI 1.39-4.40, p=0.002). Moreover, the association was consistently observed in analyses stratified for each risk factor individually in the primary cohort (Table 6).

Table 6: Stratified Cox-proportional hazards modeling of Tuberculosis risk amongst individuals with ML ratios above the 95th centile or below the 5th centile. Selection of strata based on previously reported risk factors(Lawn, Badri et al. 2005).

	TB cases (n) /number at risk (N)	Unadjusted HR (95% CI)	p-value	*Adjusted HR (95% CI)	p-value
Gender					
Male	20/571	2.74 (0.996-7.55)	0.051	**2.71 (0.966-7.60)**	0.058
Female	48/1269	2.28 (1.13-4.57)	0.021	**2.47 (1.22-4.99)**	**0.012**
WHO clinical stage 3 or 4	59/1656	2.30 (1.25-4.27)	0.008	**3.18 (1.63-6.21)**	**0.0007**
CD4 count <100	28/656	3.52 (1.59-7.79)	0.002	**3.36 (1.55-7.69)**	**0.002**
Past history of TB					
Yes	9/206	3.13 (0.65-15.13)	0.16	3.45 (0.70-16.9)	0.13
No	56/1656	2.31 (1.25-4.27)	0.008	**2.40 (1.29-4.45)**	**0.005**

*adjusted for sex, WHO stage, CD4 count, past history of TB where that covariate is not the stratification variable

To exclude the possibility that the effect of ML ratios on TB risk was mediated by diiferent HIV treatment outcomes according to ML ratio, we confirmed that baseline ML ratios were not associated with virological or immunological outcomes (per WHO-2010 criteria) at six or twelve months on cART in the primary (Figure 6) or replication cohorts (data not shown).

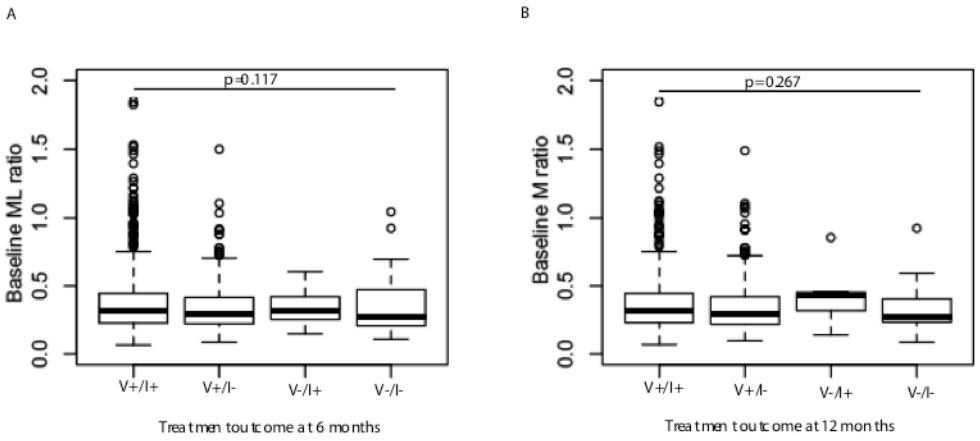


Figure 6: ML ratio at baseline does not predict immunological or virological response to cART. ML ratios at baseline (pre-cART commencement) are similar regardless of whether patients responded by (A) 6 months to cART with virological suppression (<5000 copies/ml) and immunological reconstitution (a 50 cells/µl increase from baseline) (V+/I+), virological suppression without immunological reconstitution (V+/I-), failure to suppress viraemia but successful immunological reconstitution (V-/I+) or failure to suppress viraemia and failure to reconstitute CD4+ T-cell counts (V-/I-). Box plots show baseline ML ratios for each category of treatment responders. ML ratios at baseline (pre-cART commencement) are similar regardless of whether patients respond to cART by 12 months. (B) Data shown as above but with outcomes at 12 months defined as virological suppression (<5000 copies/ml) and immunological reconstitution (a 100 cells/µl increase from baseline).

Monocyte:Lymphocyte ratios are stable in HIV uninfected women but are increased by HIV infection and disease progression and normalized by HIV or TB treatment

We assessed how the ML ratio is affected by changes in clinical condition. Intuitively, the ML ratio would be expected to vary following HIV acquisition, TB disease, cART or TB therapy because of changes in lymphocyte counts (Figure 5C and D). Since the index takes into account lymphocyte and monocyte values, and the latter may be independently regulated, variation of the ML ratio is of interest.

The stability of the ML ratio was assessed amongst 790 healthy, HIV uninfected, South African women who underwent routine safety monitoring over two years whilst enrolled in a topical microbicide gel trial(Abdool Karim, Abdool Karim et al. 2010). The ML ratio at visits 0 (enrolment), 3, 12 and 24 months were positively correlated (Spearman rho=0.35-0.47 and $p < 5 \times 10^{-7}$ for all comparisons, Figure 7A) and there was no difference in the distribution of ML ratio between any two visits (Figure 7B). Even when monocyte and/or lymphocyte counts were outside of their respective 5-95th centiles, the majority (65.7%) of individuals had an ML ratio within the 5^{-}95th centiles (both cell types increased or decreased). Conversely at 4.3% of visits the monocyte and lymphocyte counts were within the 5-95th percentiles but the ML ratio was not (Figure 7C). These data demonstrate that although the ML ratio is stable amongst healthy individuals, about two thirds of individuals with altered numbers (<5th or >95th centile) of either cell type retain apparently normal ML ratio's, and conversely about 1 in 20 individuals with normal cell numbers may still have altered ML ratio. Therefore, the ML ratio captures information not adequately captured in the monocyte or lymphocyte counts alone.

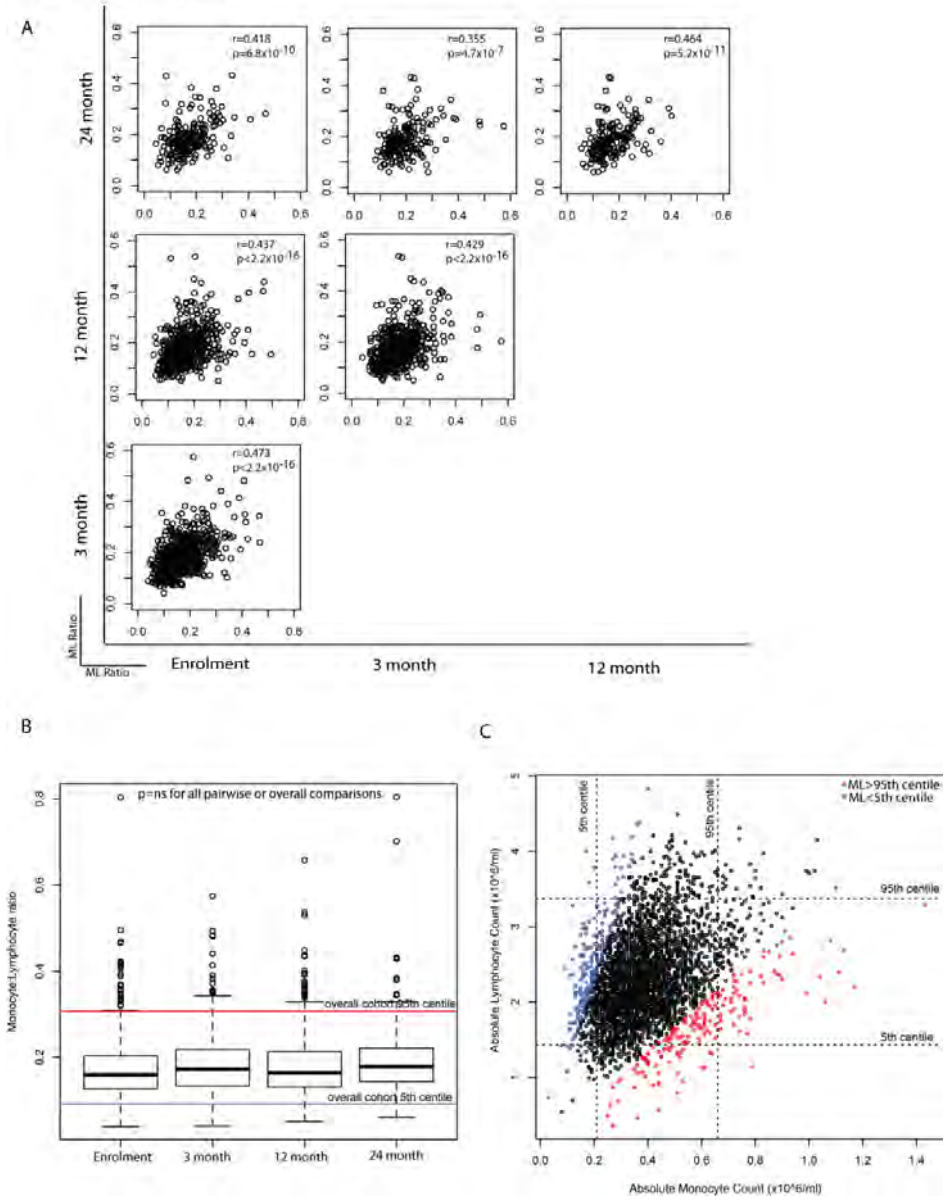


Figure 7: ML ratios, monocyte counts and lymphocyte counts amongst healthy HIV-uninfected woman. (A) The correlation between the ML ratio at enrolment and at 3, 12 and 24 months after enrolment amongst 790 participants. Results of a Spearman rank-correlation are shown as rho and p-values in the top right of each correlation plot. (B) The overall stability of ML ratios in healthy individuals shown as boxplots at each timepoint. The dotted horizontal lines denote the overall population 5th and 95th percentiles. Boxes show the 25th, 50th (median) and 75th percentiles, whiskers denote the 25th-1.5IQR and 75th+1.5IQR, and outliers are plotted as dots. (C) ML ratios may be abnormal even if the monocyte or lymphocyte count are normal. Monocyte and Lymphocyte counts are plotted on the x and y-axes respectively. Colors of datapoints denote the corresponding category of the ML ratio (black 5th-95th centiles, red>95th centile and blue<5th centile). Dotted horizontal lines denote the 5th and 95th percentile for lymphocyte counts. Dotted vertical lines denote the 5th and 95th percentile for monocyte counts.

Next, ML ratios were compared prior to and after HIV acquisition in a cohort of

160 high-risk women prospectively followed for HIV in several studies(van Loggerenberg, Mlisana et al. 2008, Abdool Karim, Abdool Karim et al. 2010, Abdool Karim, Kharsany et al. 2012). ML ratios were similar between the enrolment (median 258, IQR 126-432 days pre-infection) and last pre-infection timepoints (median 97, IQR 43-180 days pre-infection, Figure 8A). In contrast ML ratios were significantly increased following HIV acquisition (median time of sampling 29, IQR 15-54 days post-infection; median pre-infection 0.193 vs. median post infection 0.256, $p=8.2x10^{-6}$, Figure 8A), consistent with post-infection lymphopaenia. Moreover, the distribution of ML ratios broadened with HIV infection (Figure 8B). The ML ratio increased significantly over the course of HIV disease progression (Figure 8C and D; r for linear correlation with time post infection = 0.17, $p < 5 \times 10^{-12}$). The ML ratio was not associated with the rate of HIV disease progression defined as the time to CD4+ T-cell count < 350, death, or need for cART.

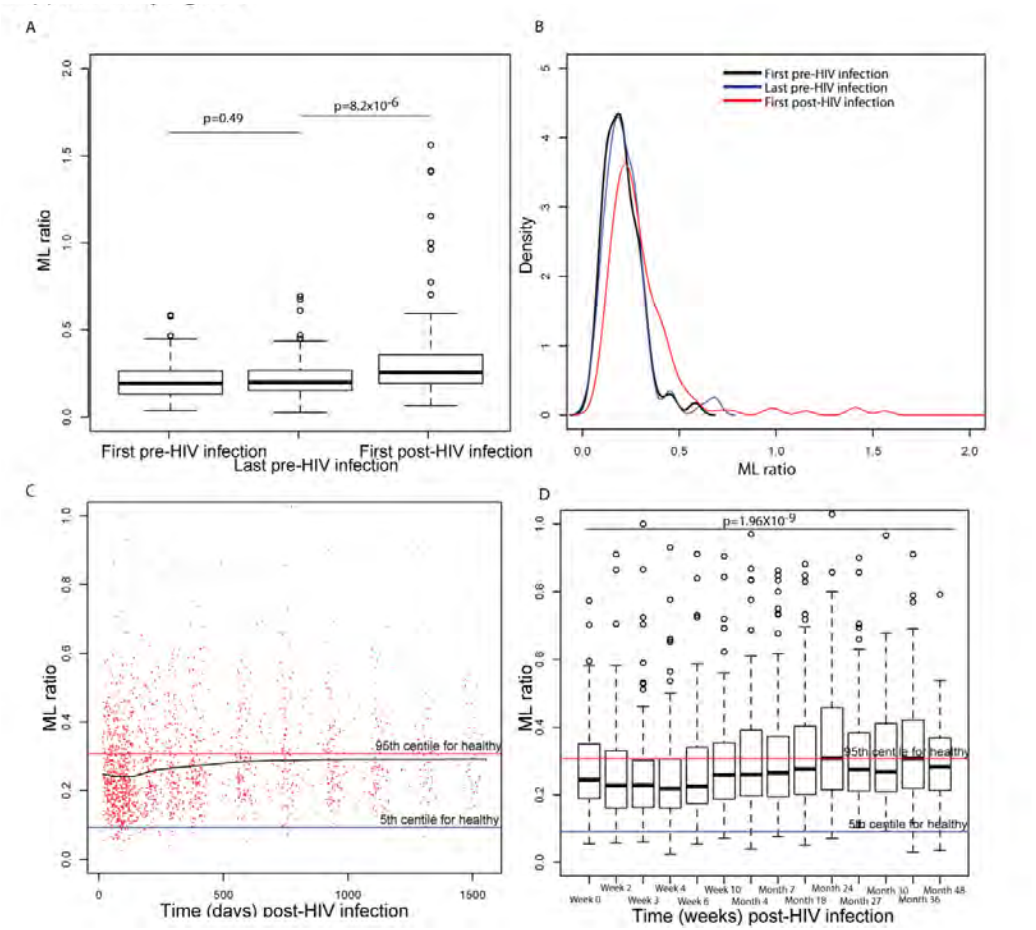


Figure 8: The impact of HIV acquisition and disease progression on ML ratios. (**A)** The ML ratio is stable pre=infection amongst 160 women who acquired HIV but is significantly increased by HIV acquisition. Boxes show the 25th, 50th (median) and 75th percentiles, whiskers denote the 25th-1.5IQR and 75th+1.5IQR, and outliers are plotted as dots. (**B)** The distribution of ML ratios is broadened by HIV acquisition. (**C)** ML ratios over the course of HIV acquisition are variable but increase overall. Data are shown as a scatterplot showing all ML measures over time amongst HIV acquirers. The black line is a smoothed curve computed by locally weighted scatterplot smoothing (LOESS). Horizontal lines denote the 5th and 95th percentile for ML ratios in HIV uninfected healthy individuals in red and blue respectively. (**D)** ML ratios over the course of HIV increase significantly with disease progression. Boxes show the 25th, 50th (median) and 75th percentiles, whiskers denote the 25th-1.5IQR and 75th+1.5IQR, and outliers are plotted as dots. Horizontal lines denote the 5th and 95th percentile for ML ratio in HIV uninfected healthy individuals in red and blue respectively.

Next, the change in ML ratio with duration of cART was assessed in a cohort of 2023 individuals screened and commenced on cART. The ML ratio was significantly altered by cART; over the first 24 months of therapy the median ML ratio decreased to within the range observed for HIV uninfected healthy volunteers and was maintained at this level for up to seven years (Figure 9). Finally, the impact of TB therapy on ML ratio was evaluated. Amongst 642 adults with TB and HIV infection who commenced anti-TB therapy prior to cART(Liu, Wilson et al. 2012), anti-TB therapy significantly reduced ML ratios (median pre-treatment ML ratio 0.373, IQR 0.273-0.530, median post-treatment 0.2753, IQR 0.203-0.376, $p=3.32X10^{-14}$). Prior to TB therapy 15% (96/642) of individuals had ML ratios in the normal range for an HIV uninfected population compared with 41.7% (189/642, median of 9 days on therapy) following TB therapy ($p<0.0001$).

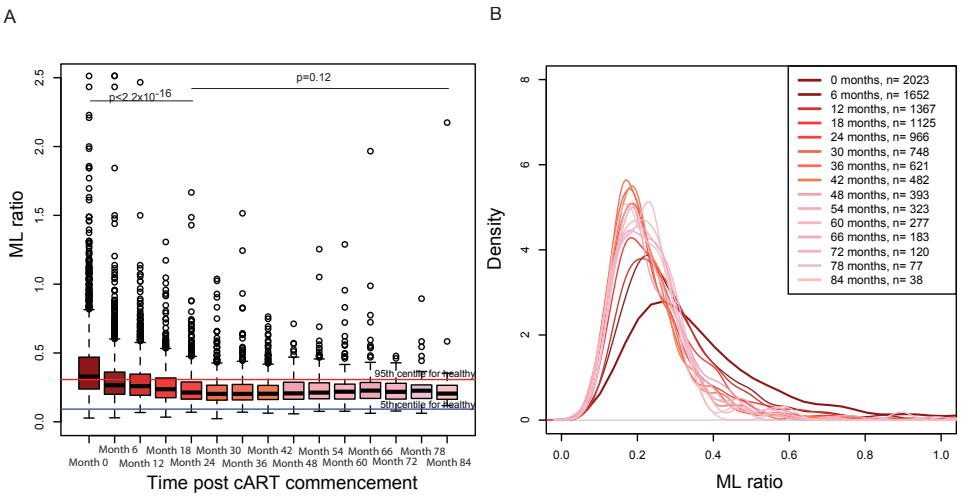


Figure 9: The impact of cART on ML ratio. (**A)** The ML ratio is reduced significantly over the course of the first two years of cART and stabilises within the normal range for HIV uninfected individuals until up to seven years post cART commencement. Box plots show the ML ratio pre-cART and over seven years of cART therapy. Boxes show the 25th, 50th (median) and 75th percentiles, whiskers denote the 25th-1.5IQR and 75th+1.5IQR, and outliers are plotted as dots. Horizontal lines denote the 5th and 95th percentile for ML ratios in HIV uninfected healthy individuals in red and blue respectively. (**B)** The distribution of ML ratios becomes narrower

2.2.4 Discussion

The ability to stratify patients starting cART for risk of TB during follow-up is necessary for targeting prevention interventions. Here we demonstrate that extremes in the ratio of peripheral blood monocytes to lymphocytes, but neither monocyte or lymphocyte counts alone, were associated with an increased risk of incident TB disease over several years. Our observation that the ML ratio is associated with TB disease risk supports unpublished findings of recent microarray studies in HIV negative infants (Fletcher et al.) and extends findings made over 80 years ago in rabbit models of mycobacterial infection(Doan and Sabin 1930). In cattle, the ratio of monocyte-derived macrophages to lymphocytes has been shown to correlate with inhibition of mycobacterial growth *in vitro*(Carpenter, Fray et al. 1997, Denis, Wedlock et al. 2004). The observation that risk is higher in individuals with either low or high ML ratios adds evidence to support recent findings that extremes of immunity are associated with TB(Casadevall and Pirofski 2003, Tobin, Roca et al. 2012). Collectively, the historic small animal, recent microarray data and these human studies demonstrate that the ML ratio may be an easily measured predictive biomarker of TB disease. This ratio could herald a previously unknown pathophysiologic driver of TB disease.

Whether the ML ratio is causatively involved in TB disease, or simply reflects underlying traits that affect TB risk requires further study. It is likely that the ML ratio captures underlying differences in immune microenvironment as has been implied by its prognostic value in influenza(Merekoulias, Alexopoulos et al. 2010) and malaria. I explore this hypothesis in chapter three. Recent

studies in mice(Muller-Sieburg, Cho et al. 2002) and humans(Pang, Price et al. 2011) have shown that subsets of hematopoietic stem cells (HSC) have distinct biases in the ratio of myeloid and lymphoid cells they give rise to. Therefore, differences in the proportions of myeloid-biased, balanced or lymphoid-biased HSC may underlie the peripheral differences in ML ratio. Recent studies have demonstrated that mycobacterial infection may alter haematopoiesis(Baldridge, King et al. 2010) or directly infect bone marrow mesenchymal stem cells(Das, Kashino et al. 2013). Therefore it is plausible that latent or prior TB infection may alter HSC such that the peripheral ML ratio is altered. In support of this is the observation that in the microarray studies, alongside the myeloid:lymphoid transcript ratio being a signature of risk, transcripts suggesting activation/quiescence of hematopoietic stem cells were more prevalent amongst infants that developed TB disease. Alternatively the ML ratio, and underlying HSC activity may be epiphenomena and the ML ratio may be a marker of where in the spectrum(Gideon and Flynn 2011) of TB disease between latency and active disease an individual is, consistent with hypothesis first proposed in 1935 by Kato(Kato 1935).

The clinical utility of the observation that the ML ratio is related to prospective risk of TB is promising but benefits from extension and validation in non TB-endemic regions, amongst children in section 2.3.

Although I observed that higher ML ratios were associated with TB, it is possible that TB disease was present but not diagnosed (reverse causation). But the association, even excluding early TB diagnoses, and that the

association appears to hold out for several years suggests that undiagnosed TB does not account for the association.

An important potential limitation of this study is that this was a secondary data analysis study and so there may plausibly have been misclassification of TB diagnosis if TB was not specifically and prospectively assessed. But the data used here were obtained during the conduct of prospective studies in which assessment for *any* clinical disorder, in particular TB, was performed. The diagnostic algorithm for TB used in these studies followed the national South African guidelines and so the diagnostic procedures in this study approximate those in clinical settings which are known to suffer impaired sensitivity for incident TB. A further limitation of this study is the absence of TST or IGRA testing of participants. Finally, at the time of clinical assessment, although clinicians were not blinded to the FBC result, they were not aware of this hypothesis thus diminishing diagnostic bias.

Amongst patients starting cART, the ability to predict the risk of TB may facilitate targeted interventions such as vaccination, treatment for LTBI to prevent TB disease or closer clinical monitoring to diagnose TB earlier. The ML ratio may be a useful, readily available tool to stratify risk of TB disease and provide a new insight into TB pathogenesis.

2.3 The Ratio of Monocytes to Lymphocytes in Peripheral Blood at age 3-months predicts the Risk of Incident Tuberculosis in in the first two years of life

2.3.1 Background

Globally, *Mycobacterium tuberculosis (MTB)* infects about 2 billion people, causing 10 million active cases, of whom about 500 000 cases are children(Lawn and Zumla 2011). Children with MTB infection are at risk for tuberculosis (TB) disease, therefore identification of these children is important. The current tools to identify children with MTB infection are TST or IGRA testing. Although isoniazid (INH) preventive therapy (IPT) is more effective in HIV-infected individuals with a positive TST result (Akolo, Adetifa et al. 2010) neither TST nor IGRA are sufficiently good at predicting TB disease(Rangaka, Wilkinson et al. 2012). Current World Health Organisation (WHO) guidelines do not support routine TST or IGRA testing before IPT provision in children (WHO 2014).

As introduced in Section 2.2, recent and historic studies suggest that the peripheral blood monocytes:lymphocytes ratio may be associated with subsequent mycobacterial disease outcomes. In the previous section, I explored how the ratio of monocytes:lymphocytes (ML) in HIV infected South African adults prior to combination antiretroviral therapy initiation (cART), was predictive of TB disease during the subsequent 5 years on cART (and called attention to the need for replication in children. In what appears to be the first and only study of the role of this ratio in childhood mycobacterial disease in 1928, Rogers studied the ML ratio in 41 children with a negative tuberculin

skin test, latent TB infection or active TB disease. He observed that the ML ratio correlated with the extent and severity of disease, but called for “the work [to] be checked up on a larger series”(Rogers 1928).

Here, we evaluated whether the peripheral blood ML ratio at between 3- and 4- months of age was associated with subsequent MTB infection or TB disease during the first two years of life in infants randomized to either IPT or placebo in the IMPAACT P1041 study(Madhi, Nachman et al. 2011).

2.3.2 Methods

Study population

In IMPAACT P1041 (clinicaltrials.gov reference NCT00080119), 544 HIV-infected and 810 HIV-exposed, uninfected infants (HEU) 91-120 days of age, born to HIV-infected mothers, were enrolled at one of three sites in South Africa (Chris Hani Baragwanath Hospital, Johannesburg; Tygerberg Hospital, Stellenbosch University, Cape Town and King Edward VIII Hospital, Durban) or at one center in Botswana (Princess Marina Hospital, Gaborone). Infants were randomized to receive Isoniazid 10 - 20mg per kilogram body weight daily or placebo. Results of the trial have been reported(Madhi, Nachman et al. 2011). Informed consent for participation in this study, specimen collection and future research was obtained from parents/guardians under protocols approved at each study site (University of the Witwatersrand: #020109; Stellenbosch University: #M04/05/026; and University of KwaZulu Natal: #T118/05) and by the Oxford Tropical Research Ethics Committee (Reference 508-13).

Definition of ML ratio

Investigations in P1041 included an enrolment full blood count, including white blood cell differential count, which was repeated three-monthly thereafter. Full blood counts were obtained using routine flow-cytometric assays in accredited laboratories.

Study endpoints

The primary outcome measure of this study was TB disease or death in HIV-infected children and latent MTB infection, TB disease or death in HEU children within 96 weeks (with 12 week window) after randomization. Infants were followed-up prospectively and routinely assessed for MTB exposure and outcomes as reported(Madhi, Nachman et al. 2011). An endpoint review committee of study-team clinicians unaware of the study-group assignments reviewed all deaths and potentially TB-related primary and secondary endpoints. Latent MTB infection amongst HIV-uninfected infants was diagnosed by TST at approximately 2 years of age (and again at 3 and 4 years for a subset), or when clinically indicated. Latent MTB infection was defined as a positive TST (induration ≥5 mm in horizontal diameter in HIV-infected children and ≥10 mm in HEU children) in the absence of evidence of active disease as per contemporary guidelines. The criteria used for categorizing TB as either possible, probable or microbiologic confirmed (definite) have been previously detailed(Madhi, Nachman et al. 2011).

Statistical methods

To calculate the ML ratio, the absolute monocyte count was divided by the absolute lymphocyte count. Analysis was performed according to a statistical analysis plan developed *a priori.* To model the association between ML ratio and the primary endpoint, Cox proportional hazards models were used. Scaled Schoenfeld residuals were inspected to evaluate the proportionality assumptions of the Cox model. Study duration was calculated as the time from randomization to the first diagnosed episode of TB or death, withdrawal from the study or end of week 108 (last visit date for patients remaining in follow-up), whichever occurred first. Where possible we preferred modeling continuous variables as fractional polynomials(Royston, Ambler et al. 1999, Royston and Sauerbrei 2004), an approach that allows continuous variables to be analyzed on their native scale without categorization, and allowing for non-linear fits. We explored non-linear fits with fractional polynomials, in all cases the best statistical fit was linear.

All statistical tests were two-sided at the 5% significance level. Poisson approximations were used to calculate confidence intervals (CIs) for estimations of the incidence rate. Bootstrapped estimates of the adjusted HR across the ML ratio continuum were generated with the '*boot*' package. Statistical analyses were performed in R (The R Foundation for Statistical Computing) using the following packages 'epiR', 'survival', 'date 'and 'mfp'.

2.3.3 Results

Between December 2004 and June 2008, 1354 infants born to HIV-infected mothers were enrolled at a median of 96 days of age (IQR 92-105 days) into IPAACT P1041. Their baseline characteristics have been reported(Madhi,

Nachman et al. 2011). 1336 infants were eligible for this analysis. The reasons for ineligibility were a missing full blood count result (9 participants), implausible results (2 participants), participant not receiving INH/placebo and not followed (3 participants) or missing covariates or outcomes (4 participants). The included infants were similar to those in the primary analysis (: 47.8% (638/1336) were male, 25.2% (136/540) of all HIV-infected infants were receiving cART at baseline, 7.2% (96/1336) had a maternal history of TB, and 9% (121/1215) were ever breastfed. The median WHO weight for age z-score was -0.55 (IQR -1.45-0.28).

The median ML ratio at baseline was 0.13 (IQR 0.09-0.18). As shown in Figure 10, the distribution of ML ratio in HIV-infected infants was significantly higher (median 0.15 vs. 0.12 $p<0.001$) than HEU infants. This was consistent with the higher absolute monocyte counts amongst HIV-infected infants (median 0.93 vs 0.71, $p<0.001$) but similar absolute lymphocyte counts (6.02 vs 5.89, $p=0.15$) compared to HEU infants.

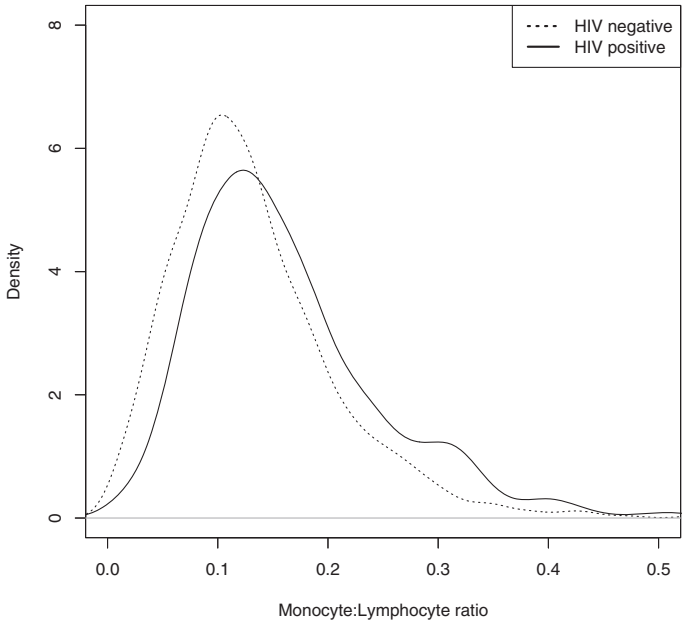


Figure 10: Distribution of monocyte:lymphocyte (ML) ratio at enrolment shown for HEU (n=796, dashed line) and HIV positive infants (n=540, solid line).

Baseline ML ratio is associated with TB disease free survival

Over 1997 infant-years, 187 infants reached the primary endpoint of this study: TB disease (probable, possible or definite) or death in HIV-infected infants and latent MTB infection, TB disease or death in HEU infants. Stratified by HIV status, the unadjusted HR per 1 unit increase in the ML ratio was 1.20 (95% CI 1.05-1.38, p=0.008). Adjusting for the weight-for-age z-score, receipt of cART at the time of sampling, a maternal history of TB, breastfeeding and receipt of INH prophylaxis, did not markedly alter the effect estimate (HR 1.17 (1.01-1.34), p=0.03). In contrast, neither the monocyte nor lymphocyte count alone was significantly associated with the primary study endpoint. Adjusting for CD4% amongst HIV-infected infants did not weaken the association in this group (HR 1.2; 95%CI: 1.01-1.43, p=0.04), nor did adjusting for absolute leucocyte counts in the entire cohort (HR 1.20; 95%CI: 1.01-1.43).

Baseline ML ratio is associated with probable or definite TB rather than all-cause mortality or latent MTB infection

To explore whether the ML ratio was specifically associated with TB disease or with other outcomes, we tested whether the association between the ML ratio and TB disease was robust to alternative endpoint definitions (Table 7). The ML ratio was significantly associated with elevated hazards of TB disease alone (HR 1.23, 95%CI 1.04-1.45, p=0.02) and with elevated hazards of TB disease plus all-cause death (HR 1.18, 95%CI 1.02-1.37, p=0.02) or TB disease plus latent MTB infection (HR 1.21, 95%CI 1.04-1.41, p=0.02). Increasing the stringency of the case definition by excluding individuals with possible TB and including only definite and probable TB diagnoses strengthened the association between ML ratio and TB disease (HR 1.50, 1.19-1.89, p=0.006). The ML ratio was not significantly associated with all-cause death (HR1.25, 95% CI 0.99-1.57, p=0.06). Amongst HEU children, with TST at 96 weeks, 33 of 725 (4.5%) had LTBI. There was no association between baseline ML ratio and LTBI (HR 1.00, 95% CI 0.66-1.5, p=0.99) neither was there evidence of interaction between ML ratio and INH prophylaxis. No INH effect was observable in a model of INH effect across the ML ratio spectrum.

The association between baseline ML ratio and TB disease free survival is similar in HIV infected and HEU children

The association between ML ratio and the primary endpoint was not statistically dissimilar between HIV infected and HEU infants (adjusted HR 1.2,

95%CI 1.01-1.42 vs. adjusted HR 1.07, 95%CI 0.83-1.4 respectively; Chi^2 test for heterogeneity with 1df: p=0.48, overall test for effect p=0.04).

Table 7: Sensitivity analysis: the association between baseline ML ratio and TB outcomes holds across TB outcome groupings.

	n (cases/ at risk)	Multivariable Regression Adjusted Hazard Ratio (95% CI)*	P value
Primary endpoint: TB disease[#] or death in HIV+ infants; and TB disease[#], death or latent MTB infection in HEU infants	187/1336	**1.17 (1.01-1.34)**	**0.03**
TB disease[#] or death	166[§]1336	**1.18 (1.02-1.37)**	**0.02**
TB disease[#] or latent MTB infection	152/1336	**1.21 (1.04-1.41)**	**0.02**
TB disease[#]	126/1336	**1.23 (1.04-1.45)**	**0.02**
Probable and definite TB only	49/1336	**1.50 (1.19-1.89)**	**0.006**
Death	50/1336	1.25 (0.99-1.57)	0.06
Latent MTB infection~	33/725	1.00 (0.66-1.5)	0.99

*Estimated from Cox Proportional Hazard regression models stratified by infant HIV status, modeling ML as linear variable following fractional polynomial exploration, adjusted for weight-for-age z-score, cART receipt at baseline, maternal history of TB, ever breastfed and randomized INH prophylaxis/placebo group. Hazard ratios are per unit increase in the ML ratio.

[#]Possible, probable or definite TB

[§]Participants who had TB disease (possible, probable or definite TB) and subsequently died were only counted once hence this number is not the sum of the TB disease and death numerators below.

~Latent MTB infection was only assessed in HEU infants

2.3.4 Discussion

Transcriptomic studies in infants suggested that the relative proportions of myeloid and lymphoid cells in infants have predictive value for TB risk (Fletcher et al., unpublished). Concordant animal studies support this notion(Doan and Sabin 1930), as does the study presented in section 2.2 performed in adults with AIDS(Naranbhai, Hill et al. 2013). Here, we found that the ML ratio in peripheral blood at around 3 months of age was associated with TB disease or MTB infection-free survival in South African children, notwithstanding conventional risk factors. Per 0.1 unit increase in ML ratio at 3 months of age, the hazard of probable or definite TB disease before 2 years increased by roughly 4% (~$1.5^{0.1}$). While significant, the modest effect size suggests that the ML ratio plays a modest role in predicting TB disease-free survival; its utility may therefore be limited to combination with other tools to stratify TB risk, or to study pathophysiologic determinants of TB disease.

Several factors support ML ratio being on the causal pathway for TB development. Firstly, monocytes are target cells for mycobacterial growth and lymphocytes are the major effectors for mycobacterial clearance. Second, there appears to be a dose or gradient effect with higher ratios being more predictive than lower ratios across the ML ratio gradient. Third, as we demonstrated, altered ML ratios precede active disease hence reverse causality is unlikely. Fourth, our data in children is consistent with *in vitro* findings. Fifth, there is overall coherence of this finding with experimental animal and observational adult studies. Finally, the association between ML ratio and subsequent TB disease has partial specificity. The association is stronger for ML ratio rather than monocyte counts and definite/probable TB

rather than possible TB. We also found that the ML ratio may be associated with childhood malaria incidence(Warimwe, Murungi et al. 2013), hence the ML ratio may actually have pleiotropic associations with some childhood infectious diseases. However, the effect size seems modest.

The mechanistic basis for the association is addressed in the chapter three.

The analysis presented has several limitations. Firstly, the definition of TB outcomes is challenging in infants. This analysis benefits from using data from a trial where outcomes were stringently defined and prospectively recorded, without knowledge of this hypothesis. Clearly specified definitions, including those delineating certainty of endpoints, allowed evaluation of the sensitivity of our finding to alternative outcome definitions, including more stringent definitions of TB disease. Secondly, this observational study cannot conclusively demonstrate a causal relationship between the ML ratio and TB disease. Mendelian randomization approaches, leveraging genetic correlates of the ratio as defined in chapter four, may facilitate more stringent evaluation of the association. Thirdly, although we observed association between ML ratio and incident TB disease in all HIV exposed infants regardless HIV infection status, further studies are required to establish whether the association is present in infants not exposed to HIV. Use of an *a priori* statistical analysis plan ameliorated the likelihood of the significant association between the monocyte:lymphocyte ratios and TB risk being a false discovery due to multiple hypothesis testing.

This study extends observations on ML ratio and risk of TB disease previously seen in HIV-infected adults to HIV-exposed infants. The modest effect size

suggests that the ML ratio plays a modest role in predicting TB disease risk in infants. Therefore, for clinical utility, the ML ratio may benefit from combination with other tools to identify at-risk infants. The replication of this finding across animal, infant and adult studies nevertheless suggests that the ML ratio is a pathophysiological predictor of tuberculosis disease and further study of this trait may yield insight into why some infants succumb to TB disease.

2.4 The Ratio of Monocytes to Lymphocytes in Peripheral Blood during pregnancy predicts the Risk of Incident in HIV-Infected Postpartum Women

2.4.1 Background

Pregnant women with HIV are at particularly high risk of TB disease due to suppression of Th1 responses during pregnancy and HIV-induced immunosuppression(Mathad and Gupta 2012). Although gratifying advances in prevention of mother-to-child transmission (pMTCT) of HIV have been achieved in sub-Saharan Africa(Govender and Coovadia 2014), TB in mothers has a profoundly negative direct effect on infants' risk of TB and indirect effect through maternal illness(Pillay, Sturm et al. 2004).

In section 2.2, I found that amongst combination antiretroviral therapy (cART)-naïve adults with AIDS, the ML ratio was predictive of incident TB disease during the subsequent five years on cART and in section 2.3, I showed that the ML ratio predicts TB in infants born to mothers with HIV. If these findings could be extended to pregnant women, they would add to the generalizability of the association and be of potential use in stratifying risk of TB in another high-risk group.

We hypothesized that increased ML ratio of in peripheral blood may be associated with incident TB disease in pregnant women. To test this I performed a secondary analysis of a large, multinational clinical trial.

2.4.2 Methods

Study design and participants

The HPTN 046 study (clinicaltrials.gov:NCT00074412), was a phase III randomized controlled trial conducted in South Africa, Uganda, Zimbabwe and Tanzania to determine the efficacy and safety of an extended regimen of nevirapine (NVP) to prevent breastfeeding transmission of HIV in infants born to HIV-infected women(Coovadia, Brown et al. 2012, Fowler, Coovadia et al. 2013). HIV-infected pregnant women (>18 years of age) who provided informed consent and were free of serious medical conditions that could interfere with study compliance were screened and enrolled in the third trimester of their pregnancy or before 7 days postpartum. The local ethics committee at each site approved the study.

Clinical care and outcome ascertainment

At screening, women provided a medical and obstetric history and underwent a physical examination. Contemporary HIV treatment guidelines informed maternal HIV treatment and women who were receiving cART or had previous NVP exposure were not excluded. At the time this study was conducted routine isoniazid preventive therapy was not implemented at the study sites but TB symptom screening was. At a minimum, to prevent MTCT, participants received the HIVNET 012 two-dose intrapartum/neonatal regimen of NVP(Guay, Musoke et al. 1999). A physical examination was performed and CD4+ T-cell enumeration and complete blood count including white blood cell differential was conducted in accredited laboratories using standardized automated counters. These clinical procedures were repeated in the women and their infants within 7 days of delivery, at 2 and 6 weeks, and at 3, 6, 12 and 18 months postpartum. For this analysis, participants with active TB or TB diagnosed in the puerperium were excluded to avoid reverse-causality.

Statistical analyses

A priori we specified that a fractional polynomial model would be used to model maternal monocyte: lymphocyte ratios against Tuberculosis outcomes, adjusting for baseline CD4+ T-cell count, WHO stage and cART receipt. Analyses were conducted in R using the *"mfp", "epiR"* and *"rms"* packages. Because of insufficient power, we elected not to evaluate infant ML ratio and TB disease.

2.4.3 Results

Between June 2008 and March 2010, 1678 pregnant women were screened in HPTN 046. For this analysis, 476 women were excluded because of loss to follow-up (n=146), withdrawal because the infant was not randomized or died (n=114 and n=64 respectively), no blood result at baseline (n=123) or active TB disease during pregnancy or puerperium (n=29). Therefore,1202 women were included in this analysis: the baseline blood specimen was obtained at a median of 40 days (IQR 20-59 days) prior to delivery. Their median age was 27 years (IQR 23-31 years) and their median CD4+ T-cell count was 435 cells/mm^3 (IQR 299-595). Clinical staging of women performed at the end of the puerperium was as follows: 1006 (83.7%) stage I, 172 (14.3%) stage II, 16 (1.3%) stage III and 3 (0.3%) stage IV (4 were missing staging information). 328 (27.3%) participants were receiving cART at enrolment, and through the course of follow-up. Further clinical characteristics have been reported(Watts, Brown et al. 2013).

The median ML ratio at enrolment was 0.253 (IQR 0.192-0.326), and the distribution was similar to that previously reported amongst non-pregnant HIV-infected women in South Africa of a similar age (p=0.76).

During 1788 years of follow-up, 12 women were diagnosed with TB (11 pulmonary TB and one TB adenitis): four developed TB by six months postpartum, seven between 6-12 months postpartum and one between 12 and 18 months postpartum. In fractional polynomial proportional hazards models, the untransformed ML ratio (i.e. exponent=1) was significantly associated with TB disease (Table 8). After adjusting for cART receipt, CD4+ T-cell count, and WHO stage, the association remained significant ($HR_{adjusted}$=1.22;95%CI 1.07-1.4;p=0.003). The association between increasing ML ratio and TB disease remained significant when absolute monocyte and absolute lymphocyte counts were included as covariates and when the model was stratified for study site.

Table 8: Cox proportional hazards models of TB disease between 6 weeks and 18 months postpartum in HPTN 046 (n=1202)*

Variable	Univariate Model		Multivariate Model	
	HR (95% CI)	**P**	**HR (95% CI)**	**P**
CD4+ T-cell count (per 100 cell increase)	**0.7 (0.5-0.98)**	**0.04**	0.61 (0.37-1.0)	0.05
WHO clinical stage (per stage increase)	0.94 (0.24-3.63)	0.93	0.73 (0.17-3.18)	0.68
cART receipt	1.44 (0.41-5.1)	0.57	0.75 (0.16-3.41)	0.71
ML ratio (per 0.1 unit increase)	**1.26 (1.12-1.41)**	**<0.0001**	**1.22 (1.07-1.4)**	**0.003**

* A multivariable fractional polynomial model, stratified by site, was used to model the ML ratio as a fractional polynomial so as to allow for non-linear fits. A linear fit, however was the most parsimonious and the variables are therefore presented untransformed.

2.4.4 Discussion

I performed a secondary analysis of women who were enrolled in HPTN 046, a large multinational study in postpartum women. The ML ratio in peripheral blood during the last trimester was positively associated with incident TB disease between 6-weeks and 18 months postpartum. A limitation of this study is the relatively low incidence rate suggesting that lack of active case finding may have led to outcome misclassification. The resultant bias towards the null is consistent with a small magnitude of effect. Nevertheless, the result is robust to adjustment for confounding. The ML ratio may be of modest use in stratifying risk of postpartum TB in pregnant women.

2.5 Conclusions

These three studies establish the ML ratio as a valid intermediate phenotype for further study**.** As discussed within each section they offer evidence first for the ML ratio as a potential risk stratification tool, and for the ML ratio as a pathophysiologic role-player in disease. However a mechanistic basis for the association was not addressed in these studies. Therefore in chapter three, I present studies aimed at understanding why the ML ratio is associated with TB disease?

Chapter Three. Functional basis for the association of myeloid:lymphoid ratio and human disease

3.1 Introduction

Whilst conducting the studies described in chapter two, I became aware of advances in the understanding of how myeloid cells develop from haematopoietic stem cells(Muller-Sieburg, Sieburg et al. 2012). These data led to the hypothesis that the ML ratio may reflect cellular function as a consequence of the ratio reflecting ontogeny. Alternatively, the ratio may reflect the frequency of target cells (monocytes) for growth relative to effector cells (lymphocytes). The next set of studies aimed to distinguish between these hypothesis by establishing the relative importance of monocyte quantity vs. quality in explaining mycobacterial disease outcomes.

3.2 Distinct Transcriptional and Anti-Mycobacterial Profiles of Peripheral Blood Monocytes Dependent on the Ratio of Monocytes:Lymphocytes

3.2.1 Background

The diagnostic and prognostic value of leucocyte differential counts was appreciated contemporaneously with the development of microscopy methods to distinguish leucocyte subsets in 1880 (Ehrlich 1880). A growing body of evidence suggests that the ratio of leucocyte subsets may be of similar, or greater, prognostic import than absolute counts. Although a full blood count is amongst the most frequently performed assays in clinical practice, the monocyte:lymphocyte (ML ratio) is not a widely-used parameter or biomarker in clinical care.

Rediscovering the animal experimental work of Sabin and colleages(Cunningham, Sabin et al. 1925, Sabin, Doan et al. 1926, Doan and Sabin 1930), I performed a series of prospective cohort studies of adults, infants and pregnant women in sub-Saharan Africa described in chapter two. In each study an elevated ML ratio was associated with the risk of subsequent tuberculosis disease. We, and others, made similar observations for malaria (Warimwe, Murungi et al. 2013) and extended these to demonstrate that the ML ratio stratifies efficacy of the candidate anti-malaria RTS,S vaccine(Warimwe, Fletcher et al. 2013). Other investigators have shown that an elevated ratio is associated with poor outcomes of nasopharyngeal carcinoma (Li, Jiang et al. 2013, Lin, Peng et al. 2014), diffuse large B-cell (Li,

Huang et al. 2012, Rambaldi, Boschini et al. 2013, Koh, Park et al. 2014, Li, Gu et al. 2014, Li, Pan et al. 2014, Markovic, Popovic et al. 2014, Porrata, Inwards et al. 2014, Porrata, Ristow et al. 2014, Watanabe, Tomita et al. 2014, Wei, Huang et al. 2014, Yan-Li, Kang-Sheng et al. 2014) and Hodgkin's (Koh, Kang et al. 2012, Porrata, Ristow et al. 2012, Romano, Vetro et al. 2012, Porrata, Inwards et al. 2013, Porrata, Ristow et al. 2013) lymphomas, multiple myeloma (Shin, Roh et al. 2013), clear-cell renal carcinoma (Hutterer, Stoeckigt et al. 2014), non-small cell lung cancer (Lin, Peng et al. 2014), soft tissue sarcoma (Szkandera, Gerger et al. 2014) and colon cancer (Stotz, Pichler et al. 2014).

The neutrophil:lymphocyte (NL) ratio, another measure of the myeloid:lymphoid cell proportion, has also been reported to be associated with cardiovascular and cancer outcomes (Guthrie, Charles et al. 2013, Templeton, McNamara et al. 2014, Wang, Zhang et al. 2014).. Since the ML and NL ratios are strongly correlated it is unclear whether one is a better predictor than the other. The pathophysiologic basis for association of myeloid:lymphoid cell ratios across diseases remains unclear.

Here we studied pathophysiologic mechanisms for the association between the ML ratio and mycobacterial disease susceptibility. We studied anti-mycobacterial and transcriptomic profiles of monocytes from healthy adult donors and found that qualitative differences in monocyte function partially explain the ML ratio association with mycobacterial growth *ex vivo*. Whole-transcriptome analysis of monocytes suggests that the ML ratio (but not the monocyte count) is a marker of monocyte function and that an elevated ratio is associated with an enrichment of interferon-associated transcripts in

monocytes. We show that the transcript signature of elevated ML ratio overlaps with mycobacterial and several other disease transcriptomic signatures such as atopy and inflammatory bowel disease. Taken together the data suggest that the ML ratio may be associated with disease by acting as a marker of monocyte function.

3.2.2 Methods

Study participants

We recruited 144 healthy adult Caucasian volunteers in Oxford, United Kingdom (Fairfax, Humburg et al. 2014). A healthcare professional (nurse and/or physician) conducted a verbal review of clinical history to determine eligibility based on the absence of major chronic illness, current medication administration or symptoms of infection.

Ethics statement

This study was approved by the Oxfordshire Research Ethics Committee (COREC reference 06/Q1605/55 & 10/H0505/3) and each individual gave written informed consent to participation.

Full blood counts

Leucocyte differential counts were performed at the Oxford Radcliffe Hospitals Pathology Laboratory, an accredited clinical laboratory, using standard procedures on a Sysmex automated haematology analyser. The ML ratio was calculated as the quotient of the absolute monocyte and lymphocyte counts.

Isolation of monocytes

Whole blood was collected into sodium-heparin containing blood collection tubes (Becton Dickinson) and processed with 4-6 hours after collection. The methods used in monocyte isolation have been previously described(Fairfax, Humburg et al. 2014) but briefly, we isolated peripheral blood mononuclear cells by density gradient centrifugation of blood-diluted with Hanks Buffered Saline solution (HBSS, Life Technologies, UK) layered on Lymphoprep (Axis-Shield, Norway), and sorted CD14+ monocytes using magnetic-activated cell sorting (MACS, Miltenyi Biotech). The CD14- fraction was used in recreating ML ratios after counting cells by microscopy.

Mycobacterial growth inhibition assays

Whole blood or mixed leucocyte growth inhibition assays were performed using the BACTEC mycobacteria growth indicator tube (MGIT) system (Becton Dickinson) as previously described(Wallis, Palaci et al. 2001), with the exception that whole blood was incubated with BCG (Pasteur) for 96 hours instead of 72 hours and mixed leucocytes (monocytes+non-monocytes = total $1x10^6$ live cells) incubated with BCG Pasteur for 72 hours. Growth inhibition was determined by calculating time to positivity (TTP) in the sample and TTP in the control (cell free culture) and converting to colony-forming units using a standard curve.

Gene expression

RNA was extracted using the Qiagen RNAeasy kit (Qiagen, UK). RNA was quantified and integrity assessed using a Bioanalyser RNA 6000 Nano kit (Agilent, UK). Gene expression was quantified using the Illumina HumanHT-12 v4 BeadChip gene expression array platform with 47,231 probes according

to the manufacturers instructions. Samples were randomly placed across expression chips and run in a single batch. Gene expression data were normalized using random-spline normalization, transformed by variance-stabilising transformation and sample outliers were iteratively removed and normalization repeated. We excluded 8 individuals as sample outliers.

Probe sequences mapping to more than one genomic location or regions with underlying polymorphisms frequent in >1% of the population were excluded from analysis (n=18220 probes). Only probes that were expressed and detected (GenomeStudio probe detection p<0.01) in monocytes were included in the analysis.

Transcription factor binding site analysis

We used oPOSSUM(Kwon, Arenillas et al. 2012) v3.0 to perform single-site analysis with the JASPAR core reference and the following search parameters: conservation cutoff=0.4, matrix score threshold=85% and upstream and downstream sequence lengths 10000 bp. Of 906 genes submitted for analysis 821 matched genes in the database.

Statistical methods

Analyses were conducted in *R,* using the packages: limma, lumi and 'vsn',. For data presentation the 'gridExtra', 'ggbio' and 'ggplot2' packages were used. Linear regression adjusting for age and sex was performed using the lm function in R. Enrichment analyses were performed using the fisher.test function in R comparing the observed vs. expected overlap in gene lists.

Ingenuity Pathway Analysis where used, was performed using a background set of all human genes on the Ilumina array. FDR where shown denotes the false discovery rate calculated using the Benjamini Hochberg procedure using the p.adjust function

3.2.3 Results

Qualitative differences in monocytes dependent on ML ratio dominate over quantitative differences in explaining the ML ratio association with mycobacterial growth

To evaluate whether the findings in chapter two are replicated *ex vivo*, we cultured whole blood from healthy adult donors with *BCG (Pasteur)* in a mycobacterial growth inhibition assay (Wallis, Palaci et al. 2001). The *in vivo* ML ratio in donors was significantly associated with enhanced mycobacterial growth *ex vivo* (β=2.23, SE0.91, p=0.02, Figure 11A).

To distinguish between a quantitative effect of increasing numbers of monocytes providing more target cells to support mycobacterial growth, from a qualitative effect of monocytes from individuals with higher ML ratio being relatively inferior at inhibiting mycobacterial growth, monocytes and lymphocytes were sorted from blood of 13 healthy adults, mixed at ratios *in vitro* that approximate the 25^{th}, 50^{th} and 75^{th} centile of ML ratios for adult populations and cultured with *BCG (*Pasteur). Higher *in vitro* ML ratio tended to be associated with elevated mycobacterial growth, however individuals with higher ML ratio *in vivo* supported greater mycobacterial growth regardless of the *in vitro* ML ratio that was created by mixing sorted cells (Figure 11B) suggesting that qualitative differences play a role beyond quantitative

differences. In a multivariate linear regression model, the *in vivo* ML ratio (β=3.22, SE 1.30, p=0.02) but not the *in vitro* ML ratio (β=1.12, SE 0.67, p=0.10) was significantly associated with mycobacterial growth (Figure 11C) suggesting that qualitative effects dominate over quantitative effects in explaining monocyte capacity to control mycobacterial growth.

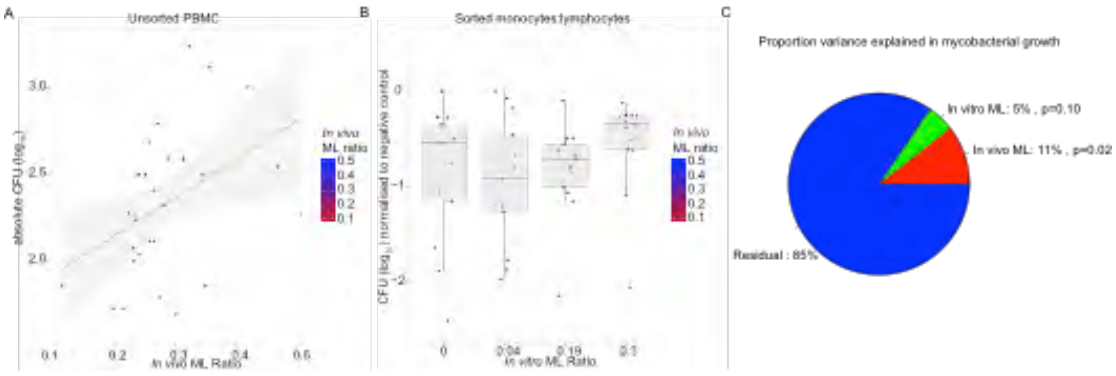


Figure 11: The ML ratio is associated with mycobacterial growth *ex vivo*, but the *in vivo* ML ratio of the donor dominates in explaining mycobacterial growth in a growth inhibition assay. **(**A) In a mycobacterial growth assay in which whole blood from healthy human adult donors (n=29) was co-cultured with BCG Pasteur for 72 hours and mycobacterial growth assessed by liquid culture, the ML ratio is associated with mycobacterial growth. Each donor is indicated by a dot that is coloured according to the *in vivo* ML ratio. A regression line and 95% CI is shown in red and grey respectively. (B) CD14+ monocytes and lymphocytes sorted from healthy human adult donors (n=13) were combined *in vitro* at ratios that approximate the 25th (0.04), 50th (0.19) and 75th (0.30) centiles of the background population and cultured with BCG Pasteur for 72 hours and mycobacterial growth assessed by liquid culture. Lymphocytes alone were added to the control tube (denoted as ML ratio=0). Each donor is indicated by a dot that is coloured according to the *in vivo* ML ratio. Boxes denote the 25th, 50th and 75th centiles and whiskers extend to median±1.5*IQR. (C) In a multivariate linear regression model of data from part B, the *in vivo* ML ratio significantly explains 11% of the variance in mycobacterial growth whilst the in vitro ratio explains 5% of variance, albeit not statistically significant.

Transcriptional profiles of monocytes dependent on the ML ratio

To understand the qualitative differences in monocytes from individuals with varying ML ratio, we analysed data for CD14+ monocytes from 136 healthy European volunteers from Oxfordshire UK (68 male, 68 female, median age=31.5 years, IQR 24-41) a subset of a cohort for whom we had performed transcriptome-wide profiling using the Illumina HT12v4 gene expression array. Donors were representative of a young healthy adult population in a region with low endemicity of mycobacterial disease. At baseline, the median monocyte count was 0.4 (IQR 0.3-0.48) $x10^9$/L and median lymphocyte count 1.81 (IQR 1.54-2.12) $x10^9$/L. After quality control procedures (Methods), we evaluated 10,104 genes (13,393 probes) that were robustly detected by high fidelity probes and were expressed in > 5% donors.

We tested for association between the level of each transcript and either the monocyte count alone, lymphocyte count alone or the ML ratio (in each case adjusting for age and sex). The expression level of 53 genes (55 probes) was significantly associated (false discovery rate (FDR) <0.05) with monocyte counts; 0 genes with lymphocyte count and 906 genes (991 probes) with ML ratio (Supplementary Table 1). For 51 of 53 genes (96%) whose expression was associated with monocyte counts, expression was also associated with the ML ratio; in these cases monocyte counts and ML ratio explained a similar proportion of overall gene expression variance (median 12.3% vs 13.4%, p=0.6). Taken together these data suggest that the ML ratio reflects a larger degree of the monocyte transcriptome than monocyte counts alone, and confirms, as per the null expectation, that the monocyte transcriptome is independent of lymphocyte count.

Of 906 genes (991 probes) associated with the ML ratio (Figure 12A), an elevated ML ratio was associated with reduced expression of 282 genes (296 probes) such as *ETS1* and *AFF3* (Figure 12B-C), and elevated expression of 624 genes (695 probes) such as *HP* (haptoglobin) and *RUNX1* (Figure 12D-E).

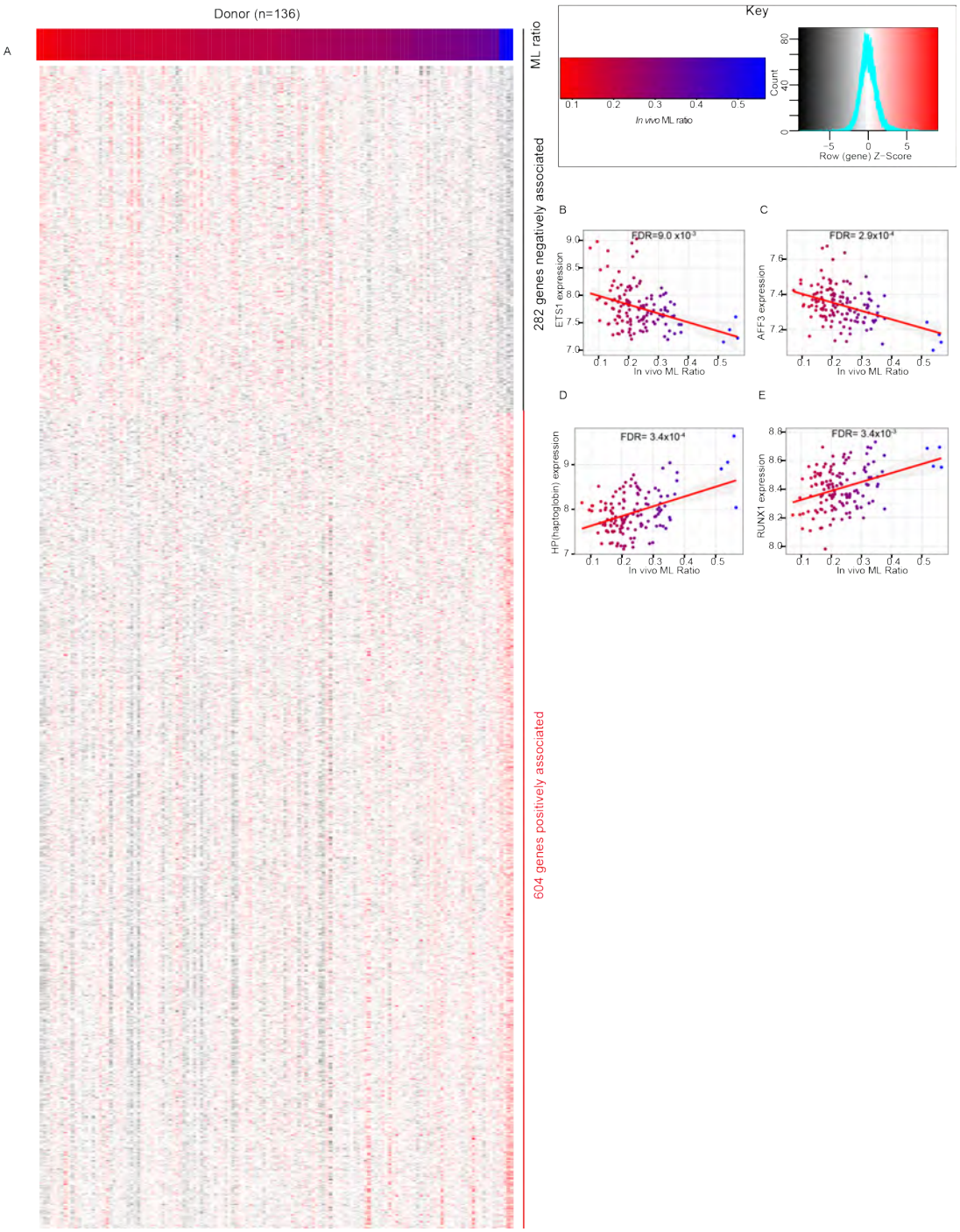


Figure 12: Monocyte transcriptomes are associated with the ML ratio of the human donor. **(**A) The ML ratio 139 individuals was associated with expression of 906 genes in CD14+ monocytes as shown in this heat map where donors are arranged according to ML ratio (along top), and direction of ML ratio-gene expression association (along right side). The key (top right) denotes colouring of individuals according to the ML ratio (top) and genes according to expression level (B-C) ETS1, encoding for a transcription factor involved in cytokine transcription, and AFF3, encoding for a transcription factor involved in lymphoid development, are examples of genes whose expression is negatively associated with ML ratio. In contrast (D-E) HP, which encodes for haptoglobin and RUNX1, encoding a transcription

factor, are positively associated with ML ratio. In B-E, the regression line and 95% CI are shown in red and grey respectively.

ML ratio associated gene transcripts are enriched for interferon signaling

Interferon signalling (Figure 13) was the most significantly enriched ($p=1.2x10^{-8}$) canonical pathway ($p<0.05$) (Supplementary Table 2). The five most significant upstream regulators were *IFNL1* (p=5.1x10-23), *IFNA2* (p-2.4x10-17), *TGM2* (p=9.2x10-16), *MAPK1* (p=5.1x10-15) and *IFNG* (p=1.1x10-13). Genes downstream of *IFNL1, IFNA2, TGM2* and *IFNG* were enriched and markedly activated whilst genes downstream of *MAPK1* were enriched and markedly inhibited. Several other regulators involved in interferon signalling are similarly enriched (Supplementary Table 3) including TLR3, IFNA, TNF, IFN1B, IFNAR2 and STAT1.

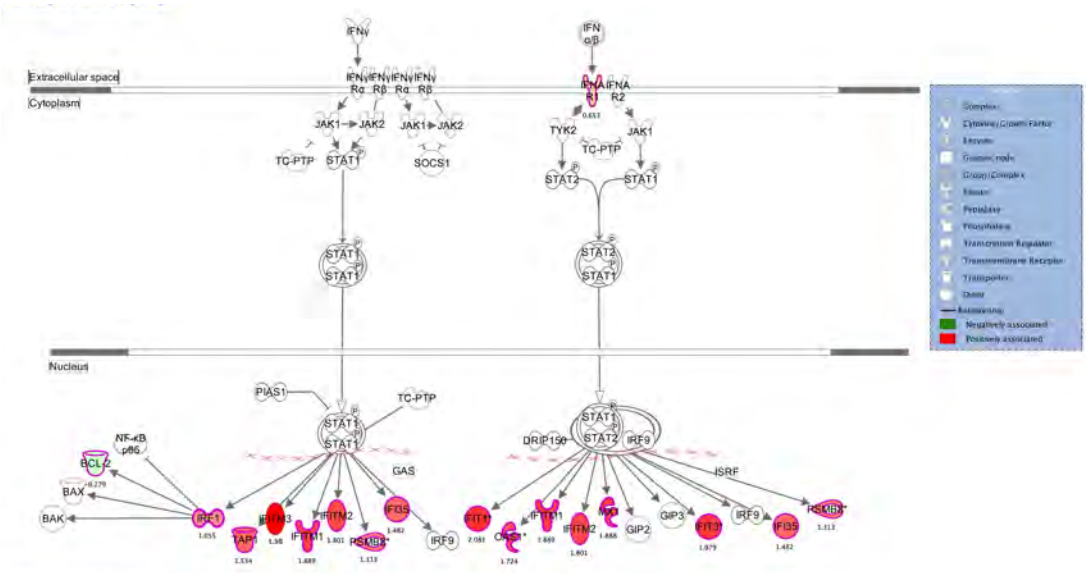


Figure 13: Monocyte genes associated with the ML ratio are enriched for genes in the interferon signalling pathway. The intensity of association with ML ratio is denoted in shades of red (increased) or green (decreased) proportionate to the level of expression and symbols denote the type of molecule. The text beneath each symbol denotes the effect size (beta) estimate between ML ratio and gene expression

Transcription factor binding analysis

To further investigate the possible pathways of association between ML ratio and monocyte transcriptomes, we considered whether the set of differentially expressed genes could be more parsimoniously explained by changes in expression of transcription factors. Of the 906 ML-ratio associated genes, 79 are human transcription factors, similar to the expected proportion (OR 0.85 95%CI 0.67-1.07) (Ravasi, Suzuki et al. 2010). Amongst the 38 transcription factors that were positively associated with ML ratio are *E2F4, STAT6, SP100, IRF1* and *IRF7*. Amongst the 41 transcription factors that were negatively associated with ML ratio are *ETS1, NOTCH2, KLF12* and *JUND* *(*Supplementary Table 1).

To characterise this further, we proceeded to test whether ML ratio-associated genes were enriched for specific transcription factor binding consensus sequences/binding sites (TFBS) using oPOSSUM(Kwon, Arenillas et al. 2012). TFBS for *IRF1/IRF2, RUNX1, RELA* and *ESRRB* were enriched for in the genes associated with the ML ratio (Figure 14). These data suggest that consequent upon upstream events such as interferon stimulation, transcription factors such as IRF1 may be involved in observed variation in monocyte transcriptomes.

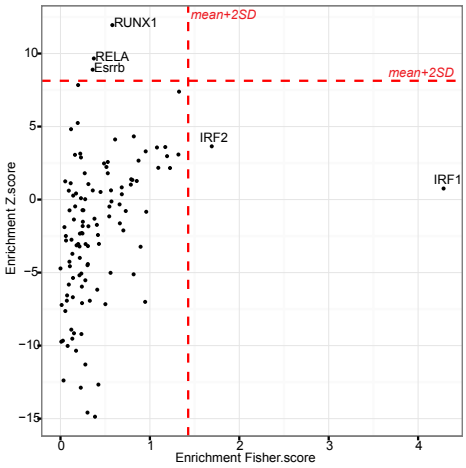


Figure 14: Monocyte genes associated with the ML ratio are enriched for transcription factor binding sites for IRF1, IRF2, RUNX1, RELA and Esrrb. The Fisher's exact test (x-axis) from oPossum disregards the number of times a TFBS occurs within a promoter in comparing observed vs expected TFBS frequency, whilst the z-score (y-axis) is a measure of the frequency of occurrence allowing for more than one TFBS motif within a gene promoted compared with the background expectation. The mean+2SD is plotted in red. Genes with enrichment scores greater than the mean+2SD are labelled.

Representation of ML associated transcripts in human diseases

In view of the apparent broad association of the ML ratio with malignant and non-malignant disease outcomes, we explored whether the set of ML associated genes in monocytes was enriched for amongst genes differentially expressed in blood in any of 151 human disease conditions curated in the Expression Atlas (Petryszak, Burdett et al. 2014). For 19 disorders, there is significant enrichment for the ML transcript signature (Supplementary Table 4). These include atopy (FDR=3.33×10^{-18}), HIV (FDR1.44×10^{-10}), the inflammatory bowel diseases (FDR=6.99×10^{-8}) ulcerative colitis and Crohn's disease, meningeal (FDR=4.97×10^{-4}), pulmonary (FDR=1.66×10^{-3}) and latent

(FDR=1.08x10^{-3}) tuberculosis, systemic lupus erythematosus (FDR=4.36x10^{-3}) and type 1 diabetes mellitus (FDR=5.1x10^{-3}).

ML associated transcripts are enriched for genes that are differentially expressed during TB disease and for TB-disease susceptibility associated genes.

On the basis of the epidemiologic associations, the *in vitro* data presented here and apparent enrichment for ML associated transcripts in a TB dataset curated in the Expression Atlas, we tested whether the ML associated gene set was similarly enriched in eight studies of tuberculosis not contained within the Expression Atlas that were recently collated and meta-analysed (Joosten, Fletcher et al. 2013). We observed a 3.50-fold (95% CI 2.64-4.61, p=1.95x10^{-16}) enrichment for TB disease-associated genes amongst ML associated genes, confirming that the transcriptomic changes during TB disease partially consist of ML-associated transcripts. These data are consistent with the observed elevation of the ML ratio during active TB disease that we showed in section 2.1.

In the absence of transcriptomic signatures that associate with TB susceptibility we examined genes that have been implicated in TB susceptibility by candidate gene studies (Supplementary Table 5). Of 95 genes for which variants have been reportedly associated with TB, 56 are expressed and detected in monocytes and were tested for association with ML ratio. 11/906 ML associated genes have been implicated in TB through candidate gene studies, a 2.5-fold (95% CI 1.16-4.93, p=0.02) enrichment

over background (45/9198). These were *CCL5, DEFB1, FCGR2A, HP, IRF1, MCL1, TAPBP, TIRAP, TLR4, TLR8* and *SP110*. This suggests that the ML ratio association with TB may, at least partly, overlap with genes in which genetic variation affects susceptibility.

3.2.4 Discussion

The relative frequency of myeloid and lymphoid cells is increasingly recognised as a variable distinct from its constituents, but the pathophysiologic basis of association remains unclear. Our observation of monocyte functional and transcriptional differences dependent on the ML ratio (but on neither constituent alone) suggests that qualitative differences in monocytes are better reflected by the ML ratio than by monocyte counts alone, potentially explaining epidemiologic associations of the ratio. The significant enrichment of interferon signalling we found supports a common role for type I and II interferons in altering the ML ratio and monocyte function sufficiently to explain altered disease course, consistent with the central role of interferons in mycobacterial and inflammatory diseases (Mayer-Barber, Andrade et al. 2014). Additionally, alteration in monocyte function may alter crosstalk with lymphocytes and adaptive immune responses (Zhong and Yazdanbakhsh 2013), hence changes in monocyte function alone may have repercussions on other aspects of the immune response that are detrimental. Finally, we demonstrate that the transcripts that are associated with the ML ratio are enriched in tuberculosis using external datasets and report evidence for a pathophysiologic role of ML-ratio associated genes in several other disease conditions.

Frequencies of leucocyte subsets in peripheral blood are a result of lineage commitment events during haematopoiesis. Clonal analysis in mice demonstrates that haematopoietic stem cells have distinct lineage biases and can be classified as myeloid-biased, balanced or lymphoid-biased depending on the relative proportions of myeloid and lymphoid progeny they give rise to(Muller-Sieburg, Sieburg et al. 2012). In mice, myeloid-biased HSC differentiation has been shown to be under the control of IFN-γ (Matatall, Shen et al. 2014). In humans, myeloid-biased HSC accumulate with age and explain the relative increase in myeloid cells in blood with age (Pang, Price et al. 2011). Therefore changes in ML ratio in blood are likely a marker of changes in the frequency of lineage-biased HSC. Intriguingly, mycobacterial infection has been shown to alter haematopoiesis through elevated IFN-γ levels (Baldridge, King et al. 2010). In murine models, ontogeny affects the functional capacity of the progeny of myeloid-biased HSC (Muller-Sieburg, Cho et al. 2004, Gekas and Graf 2013). This finding is recapitulated in humans in whom granulocyte-monocyte progenitors give rise to distinct myeloid subsets, are increased in the circulation, and portend impaired outcomes in a range of solid organ malignancies (Wu, Sun et al. 2014). The causal pathway is difficult to unambiguously ascertain but positioning our data in this context suggest a model in which inflammatory signals - particularly interferons - affect the ML ratio through lineage-bias during haematopoiesis and this has a lasting effect on the transcriptional and functional profile of the monocyte progeny.

Our study is limited in ability to exhaustively examine all potential mechanisms in operation. We studied only CD14+ monocytes and it is possible that

changes in other monocyte subsets may occur. Our evaluation of ML ratio transcripts in other diseases is limited to bioinformatic overlap with data curated in the Expression Atlas. Further studies will be required to confirm a role, if any, of the ML ratio in these disorders as it is plausible that the overlap is driven by common involvement of interferon pathways in these disorders. Finally, we cannot definitely determine the direction of association between the transcript profiles in monocytes and putative upstream events and it is likely that bidirectional relationships (for example between interferons and monocyte function) exist.

Collectively our data suggest that elevation in the ML ratio is associated with transcriptional changes in monocytes that may account for impaired anti-mycobacterial profiles of monocytes, potentially explaining some pathophysiologic disease associations of the ML ratio.

3.3 Conclusions

This study demonstrated that the ML ratio is a marker of monocyte function, and that the monocyte or lymphocyte counts alone do not explain monocyte transcriptomes as well as the ratio. An elevated ML ratio appears to be tightly linked to elevated type-I and type-II interferons and the monocytes concordantly show elevation in transcripts under the control of interferon responsive transcription factors. The study provides a possible mechanism for the prognostic value of the ML ratio and secondly, offers additional validation for the involvement of the ratio in tuberculosis disease and specifies novel disease associations. Further examination of the ML-ratio associated transcripts in other diseases is warranted.

Although there exists strength in the ML ratio association with TB and consistency, temporality, a biological gradient, coherence with *in vitro* evidence, and plausibility there are two elements that make causal inference incomplete according to the Bradford-Hill criteria for causation(Hill 1965). Firstly the ML ratio appears to be associated with a wide range of diseases and is not disease specific. Moreover, for some diseases the neutrophil:lymphocyte (NL) ratio has been reported as the predictive variable(Paramanathan, Saxena et al. 2014, Templeton, McNamara et al. 2014, Wang, Zhang et al. 2014). Secondly, although some experimental evidence exists this is either *in vitro* as in our study, or in animal models(Doan and Sabin 1930), robust experimental evidence is still lacking. Causal inference would be strengthened by a plausible hypothesis for the apparent lack of specificity and further experimental evidence. One reason for the lack of apparent specificity may be that the ML ratio reflects a monocyte functional profile that is central to many diseases. Indeed type-1 and type-II interferons, which we showed are linked to the ML ratio, are involved in many, if not most, inflammatory processes. Hence in the next set of studies I sought to define the extent to which the genetic architecture of the ML and NL ratios is shared, and whether this may be exploited in a Mendelian randomization study to test whether the ML ratio is causally associated with TB disease susceptibility.

Chapter Four. The genetic architecture of myeloid:lymphoid cell ratios

4.1 Introduction

Defining the genetic architecture of leucocytes, i.e. the pattern of genetic control of leucocyte frequency, may help to disentangle the relative contribution of different leucocyte subsets to disease susceptibility. In this chapter I describe what role genetic factors play in regulation leucocyte frequency, and go on to define specific genetic correlates of the absolute and relative frequency of leucocytes.

4.2 Heritability and co-heritability of absolute and relative leucocyte counts suggest coregulation of myeloid and lymphoid cell numbers

4.2.1 Background

A particular area of advancement in haematology since the first descriptions of leucocyte subsets(Ehrlich 1880) and their clinical relevance has been the delineation of their ontogeny from haematopoietic stem cells(Ema, Morita et al. 2014). Aided by advancements in genetic tools, understanding of the sequential branch-like commitment to each subset, and the attendant genetic and transcriptional events is burgeoning. Genome-wide association studies (GWAS) in humans have helped to outline the underlying genetic architecture of many individual leucocyte traits, including the total leucocyte count, and counts of lymphocytes, neutrophils and monocytes(Nalls, Couper et al. 2011, Okada, Hirota et al. 2011, Reiner, Lettre et al. 2011, Crosslin, McDavid et al. 2012, Kong and Lee 2013). The breadth of effect of a genetic variant (i.e. how many lineages it has an impact on) may intuitively be wider for variants that affect earlier events in leucopoiesis. However few studies have systematically examined evidence for pleiotropy of effect, whereby the same genetic variant affects more than one haematologic trait. Whether such pleiotropy exists, and to what extent, is an important question to answer if we are to understand how a genetic variant gives rise to a leucocyte trait and eventually to a clinical phenotype or disease.

The epidemiologic studies presented in chapters two and three find that the ML ratio, rather than the absolute counts, is associated with clinical outcomes.

By extension, these data suggest that factors involved in regulation of ratios may be of at least as great import as those of absolute counts.

Here, I estimate the heritability and genetic correlation of ML and NL ratios and their respective constituents. I present classical concordance-and structural equation modelling-based estimates of heritability in 1664 twin-pairs, as well as estimates of heritability and co-heritability using genome-wide genotyping and restricted maximum likelihood(Yang, Benyamin et al. 2010) in more than 5000 unrelated adults. The data provide a resource of heritability estimates, and suggest widespread pleiotropy in genetic regulation of leucocyte frequency.

4.2.2 Methods

Cohorts

We obtained genome-wide genotypes and full blood count measures for 1664 female twin pairs of ages 15-85 years (TWINSUK), 1074 female and male adults of age 15-65 (Cambridge Bioresource-CBR), and 2296 female and male adults born in 1958 (United Kingdom Blood Services-UKBS). The eligibility criteria and enrolment procedures for these cohorts have been previously described(Power and Elliott 2006, Soranzo, Spector et al. 2009, Moayyeri, Hammond et al. 2013).

Twin concordance

To estimate the twin heritability of leucocyte traits, we used Falconers formula(Falconer and MacKay 1996), given by $h^2=2 * (R_{MZ} - R_{DZ})$ where h^2

represents the heritability estimate, $R_{MZ\ and}$ $R_{DZ,}$ the concordance estimates of the phenotype amongst monozygotic and dizygotic twins respectively. We performed ACE modelling, a form of structural equation modelling, to calculate the additive genetic (A), common/shared environmental (C), and unshared environmental (E) effects, as implemented in the R package OpenMx(Boker, Neale et al. 2011).

Polygenic heritability

To estimate the proportion of phenotypic variance explained by genotypic variance ($h^2=V_G/V_P$, where V_G and V_P refer to genetic and phenotypic variances respectively) we used the restricted maximum likelihood (REML) method (implemented in the GCTA suite of tools(Yang, Lee et al. 2011). Estimates were calculated using genome-wide genotyping data for 1074 CBR participants, 2296 UKBS participants, a random selection of 1664 participants, one from each twin pair, in TWINSUK, and the combined dataset. Related individuals were excluded and 20 principal components were used as covariates to correct for population stratification. Phenotypes were transformed to z-scores. To partition SNP-heritability by autosome, we estimated genetic relationship matrices (GRM) for each autosome(Yang, Manolio et al. 2011). With 5964 individuals the probability of detecting $h^2>0$, assuming a true h^2 of at least 20%, at a type 1 error rate of 5%, was 0.96. Genetic correlation or co-heritability(Lee, Yang et al. 2012) was calculated using the --reml-bivar option in GCTA (v1.2.4).

Analyses and data plotting were performed in R unless otherwise indicated, using the following packages: ggplot2, gridExtra, corrplot, OpenMx.

4.2.3 Results

Monocyte:lymphocyte and neutrophil:lymphocyte ratios are correlated

The distribution of absolute and relative blood counts by cohort are shown in Figure 15. We calculated the correlation between the total leucocyte count, subset counts, and the ML and NL ratios (Figure 16). Constituents of WBC, ML and NL were, as expected, correlated with that composite measure. The monocyte and neutrophil counts were positively correlated ($r_{monocyte-neutrophil}$=0.29, 95%CI 0.26-0.32, $p<2.2x10^{-16}$), but normalising to the lymphocyte count reveals a stronger correlation between the ML and NL ratio (r_{ML-NL}=0.52, 95%CI 0.50-0.54, $p<2.2x10^{-16}$).

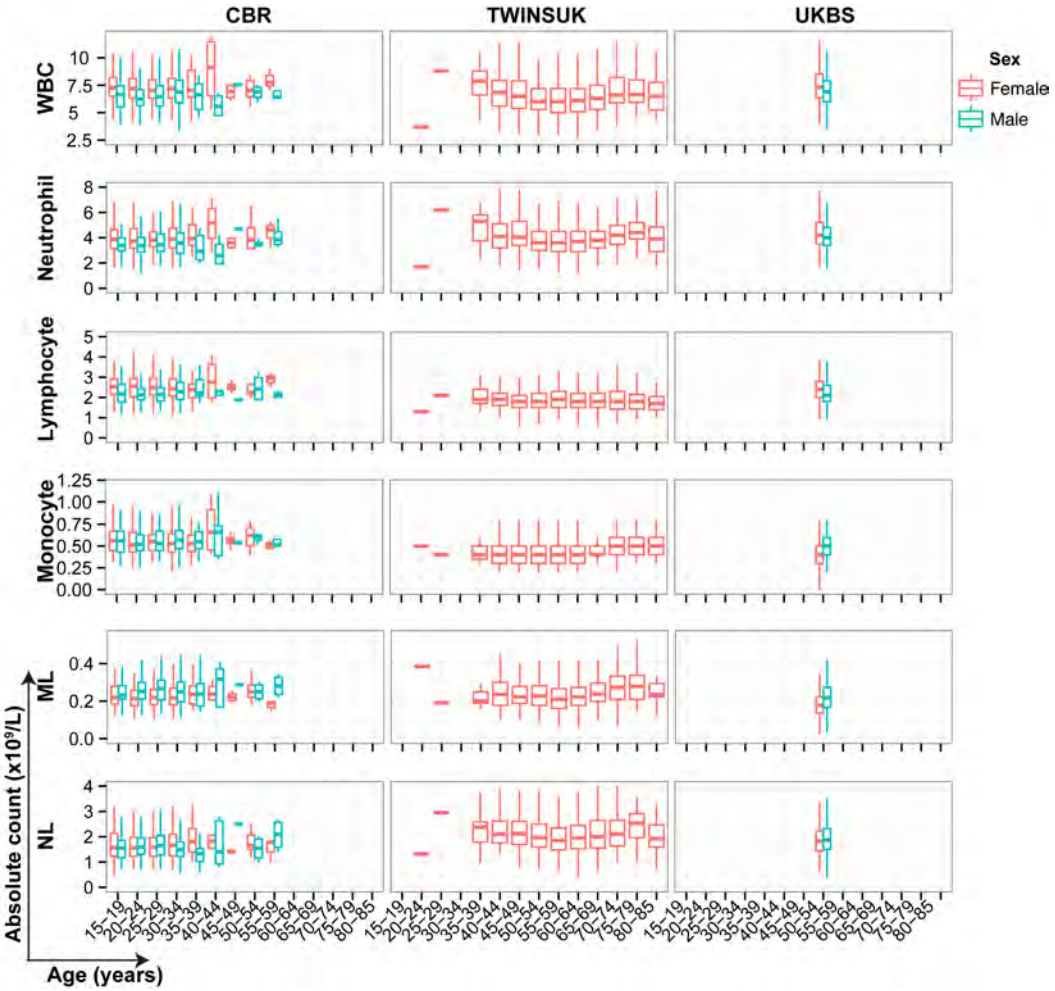


Figure 15: The distribution of peripheral blood white blood cell (WBC), neutrophil, lymphocyte or monocyte counts and ML and NL ratios according to participating cohort, and age group. Solid line at centre of box denotes median, box edges denote 25th and 75th centiles and whiskers extend to 1.67*IQR.



Figure 16: Correlation coefficients between peripheral blood measures. Squares in top right corner are drawn proportional in size and colour to the correlation coefficient (pearson r) given in bottom left of matrix.

Concordance of monocyte:lymphocyte or neutrophil:lymphocyte ratios in twins demonstrates substantial additive genetic control of these ratios

Comparing the concordance in leucocyte measures between 880 monozygotic and 784 dizygotic twin-pairs with non-missing yield twin-heritability estimates of between 44-67% (Table 9). Structural equation modelling to separate additive genetic (A), shared/common environmental (C) and unshared environmental (E) contributions using an ACE model, confirm a substantial role for additive genetic effects of about 50% for each trait and indicate that unshared environmental effects are responsible for the majority of the remaining variance in leucocyte subset counts. The NL and ML ratio have similar additive genetic effects (50.26%, 95%CI 41.31-54.79 and 53.24%, 95%CI 40.03-57.39 respectively).

Table 9: heritability estimates of six leucocyte parameters in TWINSUK (n=880 MZ and 784 DZ pairs)

	Falconer's method			Structural Equation model		
Trait	MZ concordance (R_{MZ})	DZ concordance (R_{DZ})	Heritability*	Additive genetic, A (95% CI)	Shared environment, C (95% CI)	Unshared environment, E (95% CI)
WBC	61.90	28.18	67.43	62.73 (55.46-66.30)	1.5x10-13 (8.7x10-19-6.0)	37.27 (33.70-41.19)
Neutrophil	55.87	22.39	66.97	52.59 (41.72-56.89)	8.3x10-13 (6.3x10-14-8.9)	47.07 (42.54-51.98)
Lymphocyte	63.46	41.21	44.48	54.90 (42.26-67.07)	9.6 (8.7x10-22-20.89)	35.45 (32.03-39.25)
Monocyte	59.20	28.29	62.02	52.93 (46.28-57.46)	4.2x10-14 (1.9x10-20-4.81)	47.07 (42.54-51.98)
NL	53.25	27.52	51.47	50.26% (41.31-54.79)	3x10-13 (2.1x10-20-6.99)	49.74 (45.21-54.61)
ML	53.22	25.84	54.77	53.24 (40.03-57.39)	4.6x10-13 (8.6x10-25-11.09)	46.76 (42.61-51.29)

Polygenic heritability of ML and NL ratios confirms a substantial role for genetics in control of the ratio

Next we estimated the SNP-heritability (V_G / V_P) of these leucocyte traits using the GREML method(Yang, Benyamin et al. 2010).

There are two noteworthy observations (Figure 17A). Firstly, the WBC, monocyte count and NL ratio are the most highly heritable. Secondly, the polygenic estimates are, in general, similar to the twin-heritability estimates. However, as expected by the imperfect tagging of causal variants by genotyped SNP's, polygenic estimates tend to underestimate the additive genetic effects calculated by ACE modelling amongst twins although this is not universal. For example the polygenic heritability of the NL ratio is 59.0% (95%CI 51.0-68.0) and the additive genetic effects from twin ACE modelling is 50.3%, (95%CI 41.31-54.79); this may indicate presence of non-additive genetic effects.

Partitioned by chromosome, the distribution of polygenic genetic effects is not proportional to the coding length of each chromosome (Figure 17B), suggesting that specific loci contribute to genetic control of these measures disproportionate to their length, in contrast to traits such as height(Yang, Manolio et al. 2011).

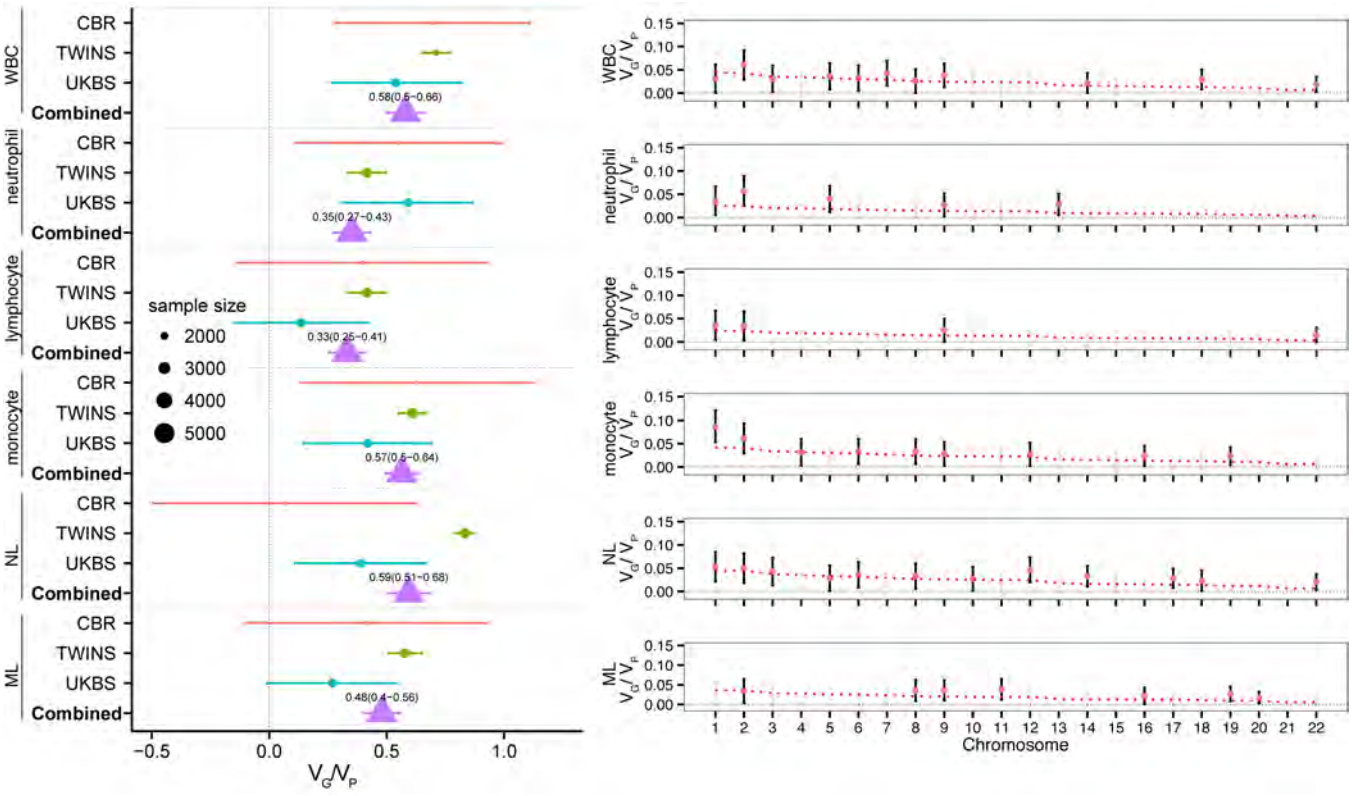


Figure 17: (A) SNP-heritability (V_G/V_P) of each of the leucocyte measures according to cohort, or combined and (B) partitioned according to chromosome. In (A), the size of the symbol at the centre is proportional to the sample size, and whiskers denote95% confidence interval. In (B), the red dotted line denotes the expected estimate if V_G/V_P was distributed according to the length of each chromosome. The observed estimates are plotted with a red dot denoting the point estimate, whiskers extending to 95% of the estimate and shown in low intensity if the CI crosses zero..

Genetic correlation between monocytes and lymphocytes suggests shared underlying architecture

Finally, to estimate genetic correlation (rG) of the ML and NL ratio as a baseline estimate of co-heritability and pleiotropy, we used the bivariate GREML method proposed by Visscher and colleagues(Lee, Yang et al. 2012). Consistent with expectations, the genetic correlation between constituents of a measure and that measure itself were high: neutrophils with WBC and NL ratio, monocytes with ML, and lymphocytes negatively correlated with both NL and ML although to differing extents (Figure 18). Less expectedly, the genetic correlation between lymphocyte and monocytes counts was 0.79 (95%CI:

0.67-0.91), indicating a large proportion of shared genetic architecture. The overall genetic correlation between ML and NL ratios was 63.3% (95%CI 55.4-21.2%) whereas the genetic correlation between absolute monocyte and neutrophil counts is just 13.6% (95%CI 5.1%-22.16%). Unsupervised hierarchical clustering of the co-heritability estimates supports a shared genetic basis for lymphocyte-monocyte counts, ML-NL ratio and WBC-neutrophil counts, although we note that the latter pair is fully nested (i.e neutrophils constitute the majority of WBC).

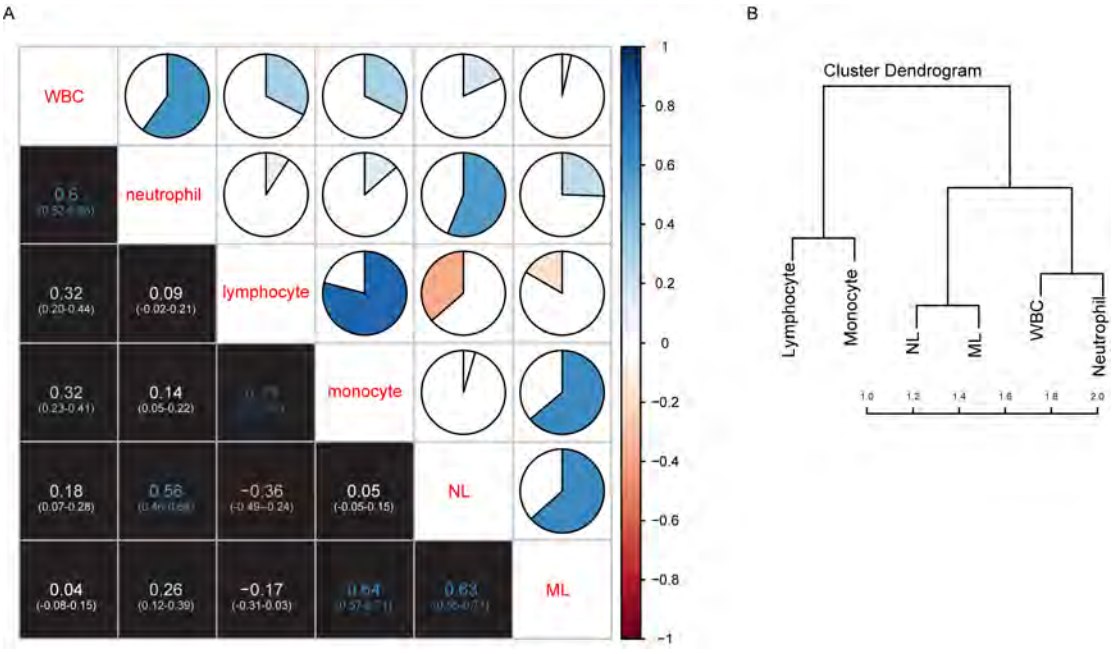


Figure 18: Genetic correlation or coheritability of leucocyte measures. In (A), the genetic correlation rG is shown at bottom left together with the 95% confidence interval of the estimate, is plotted as the according fraction of 1 in a pie plot at top right and is coloured according to the estimate. (B) Unsupervised hierarchical clustering of genetic correlation estimates performed using the Lance-Williams approach implemented in hclust (*R*)

4.2.4 Discussion

Here I examined the phenotypic and genetic correlation amongst six leucocyte traits: the WBC, neutrophil, lymphocyte and monocyte counts and the NL and ML ratios. The findings indicate that NL and ML ratios are phenotypically and genetically correlated, and therefore likely share common causal pathways. The observation that monocyte and lymphocyte counts are themselves highly genetically correlated suggests that a shared architecture for coregulation of these measures exists. We also estimated the heritability of the six leucocyte measures using twin studies and more recently developed SNP-heritability methods. Our findings provide evidence for substantial genetic effects on these six measures. Moreover, they provide insight into the architecture of genetic effects by demonstrating that the heritability is disproportionality partitioned to certain chromosomes in contrast to traits such as height or BMI(Yang, Manolio et al. 2011) where the small effects appear to be uniformly distributed across the genome.

More than 30 studies in diverse populations report that elevation in the monocyte:lymphocyte or the neutrophil:lymphocyte ratios is associated with poorer outcomes for infectious and malignant disease ((Paramanathan, Saxena et al. 2014, Templeton, McNamara et al. 2014, Wang, Zhang et al. 2014). Most also demonstrate that similar information is not conveyed by the constituent measures, indicating the fact that the ratio captures distinct prognostic value. Our results reconcile the ML and NL ratio in showing their phenotypic correlation. Validation and assessment of causality of the reported disease associations is desirable, but the feasibility of Mendelian randomisation studies to do so is reliant on identification of non-pleiotropic genetic determinants of the ratio that can be used as instrumental

variables(Sheehan, Didelez et al. 2008). The results here suggest that the monocyte:lymphocyte ratio may be extensively co-regulated by shared genetic effects. By implication, extensive pleiotropy exists, necessitating specific methods to isolate non-pleiotropic genetic instruments for Mendelian randomisation studies to validate the ML ratio association with disease. Revised models of long-term haematopoietic stem cells (HSC) indicate that the peripheral blood myeloid:lymphoid cell ratio may be regulated by the activity of lineage-biased HSC(Sieburg, Cho et al. 2006). We speculate that the genetic architecture that co-regulates monocyte and lymphocyte counts may operate through lineage-biased HSC.

Taken together, our data suggest an underlying genetic architecture that co-regulates monocytes and lymphocytes, and the ML and NL ratios. Genome-wide association studies to resolve the precise genetic determinants of the leucocyte measures will need to take into account the high degree of pleiotropy in genetic regulation of leucocytes.

4.3 A genome wide association study of host genetic correlates of myeloid:lymphoid ratio

4.3.1 Background

Leucocyte subset counts vary substantially even amongst healthy individuals. Highly aberrant counts were noted to be a feature of severe disease, in particular infectious(Ehrlich 1880), leading to development of normal ranges. Nevertheless, differences even within the apparent normal range are of recognised prognostic value for cardiovascular disease, malignant and all-cause mortality(Friedman, Klatsky et al. 1974, Grimm, Neaton et al. 1985, Danesh, Collins et al. 1998, Wheeler, Mussolino et al. 2004, Ganna and

Ingelsson 2015). Almost 80 years ago, the MLR was observed as a better predictive feature than either constituent count alone for the course of mycobacterial disease.

Identifying genetic correlates of peripheral blood leucocyte counts and ratios could facilitate MR studies to validate their association with disease (Solovieff, Cotsapas et al. 2013). There have been several studies of the genetic basis for leucocyte or subset counts(Soranzo, Spector et al. 2009, Nalls, Couper et al. 2011, Okada, Hirota et al. 2011, Reiner, Lettre et al. 2011, Crosslin, McDavid et al. 2012, Kong and Lee 2013) but no published assessments of the genetic determinants of the MLR or NLR. Here we performed a genome-wide association study of the MLR and NLR and their constituents. In view of the observations noted in section 4.1, of phenotypic covariance and substantial coheritability in the MLR and NLR, and the need for non-pleiotropic instrumental variables in MR studies, we specifically assessed evidence for pleiotropy.

4.3.2 Methods

Cohorts

We obtained genome-wide genotypes and full blood count measures for 1,304 females, each a randomly selected member of twin pair (TWINSUK), 1,072 female and male adults of age 15-65 (Cambridge Bioresource-CBR) and 2,296 female and male adults born in 1958 (United Kingdom Blood Services-UKBS). The eligibility criteria and enrolment procedures for these cohorts has been previously described(Power and Elliott 2006, Soranzo, Spector et al. 2009, Moayyeri, Hammond et al. 2013). We included 1,096 individuals from the Estonian Genome Centre University of Tartu cohorts who

were genotyped on one of two independent platforms (EGCUT370K and EGCUTOMNI), adults of Estonian descent(Leitsalu, Haller et al. 2014).

Data processing and association testing

For each cohort we performed quality control using standard procedures that are described elsewhere, phased genotypes using SHAPEIT(Delaneau, Marchini et al. 2012), and imputed missing genotypes using IMPUTE2(Howie, Marchini et al. 2011) with the 1000 Genomes phase 3 release as a reference set of haplotypes. Association with monocyte, lymphocyte and neutrophil counts or the MLR or NLR was performed in SNPTEST v2 with a frequentist additive model, correcting for age, sex and population structure using 4 principal components.

Meta-analysis

A fixed-effects meta-analysis with inverse variance weighting was performed in GWAMA(Magi and Morris 2010). Double genomic control was implemented to stringently control for population stratification. Single-nucleotide variants that were directly genotyped or well imputed (info>0.4), in at least 3 cohorts and 1,200 individuals were included in meta-analysis.

Transcription factor binding site analysis

A single-site analysis with the JASPAR core reference was performed in oPOSSUM(Kwon, Arenillas et al. 2012) v3.0 with the following search

parameters: conservation cutoff=0.4, matrix score threshold=85% and upstream and downstream sequence lengths 5,000 bp. Of 19 genes submitted for MLR analysis 10 matched genes in the database and of 57 submitted for NLR 32 matched genes in the database, those not matching encoded microRNA, long intergenic non-coding (LINC) RNA's or arose from unannotated regions (LOC).

Analyses and data plotting were performed in R unless otherwise indicated, using the following packages: ggplot2, gridExtra, ggbio and GenomicRanges.

4.3.3 Results

Through fixed-effects meta-analysis of primary association data from 5,768 adults of European descent enrolled into 5 cohorts, we aimed to identify genome-wide significant (GWS, $p<5x10^{-8}$) variants that passed QC (info>0.4, allele frequency>1%) in at least three of five studies and 1,200 individuals, and were associated with monocyte, lymphocyte or neutrophil counts, the MLR or the NLR (Table 10).

Table 10: Baseline characteristics and quality metrics of genotypes for cohorts included in leucocyte GWAS

Cohort	Cambridge Bioresource (CBR)	Estonian Genome Centre-University of Tartu 370K (EGCUT370K)	Estonian Genome Centre-University of Tartu OMNI (EGCUTOMNI)	TwinsUK	1958 Birth cohort-United Kingdom Blood Service (UKBS)
n (%female)	1,072 (58.7)	149 (49)	947 (52)	1,304 (100)	2,296 (50.2)

Median untransformed Monocyte count (IQR)	0.54 (0.45-0.65)	0.48 (0.39-0.63)	0.49 (0.4-0.61)	0.4 (0.3-0.5)	0.4 (0.4-0.5)
Median untransformed Lymphocyte count (IQR)	2.35 (2.0-2.8)	1.8 (1.49-2.28)	1.89 (1.56-2.29)	1.8 (1.50-3.68)	2.3 (1.9-2.7)
Median untransformed Neutrophil count (IQR)	3.67 (3.05-4.53)	3.4 (2.67-4.18)	3.46 (2.8-4.37)	3.8 (3.08-5.84)	4.2 (3.5-5.0)
Median untransformed MLR (IQR)	0.23 (0.19-0.28)	0.26 (0.22-0.32)	0.26 (0.21-0.33)	0.23 (0.18-0.29)	0.19 (0.15-0.25)
Median untransformed NLR (IQR)	1.58 (1.23-2.03)	1.77 (1.33-2.44)	1.84 (1.41-2.40)	2.06 (1.6-2.67)	1.85 (1.48-2.31)
Variants included after imputation	7995533	8156089	6109575	7572987	7857131
λ_M	1.00	1.03	1.02	1.00	1.01
λ_L	1.00	0.94	1.02	1.00	0.98
λ_N	1.02	1.02	1.02	1.01	1.01
λ_{NL}	1.00	1.02	1.03	1.00	1.00
λ_{ML}	1.01	1.01	1.02	1.01	1.00

Genome-wide identification of correlates of monocyte, lymphocyte and neutrophil counts

We identified 4 (3 independent loci), 28 (20 independent loci) and 54 (21 independent loci) GWS variants for the monocyte, lymphocyte and neutrophil

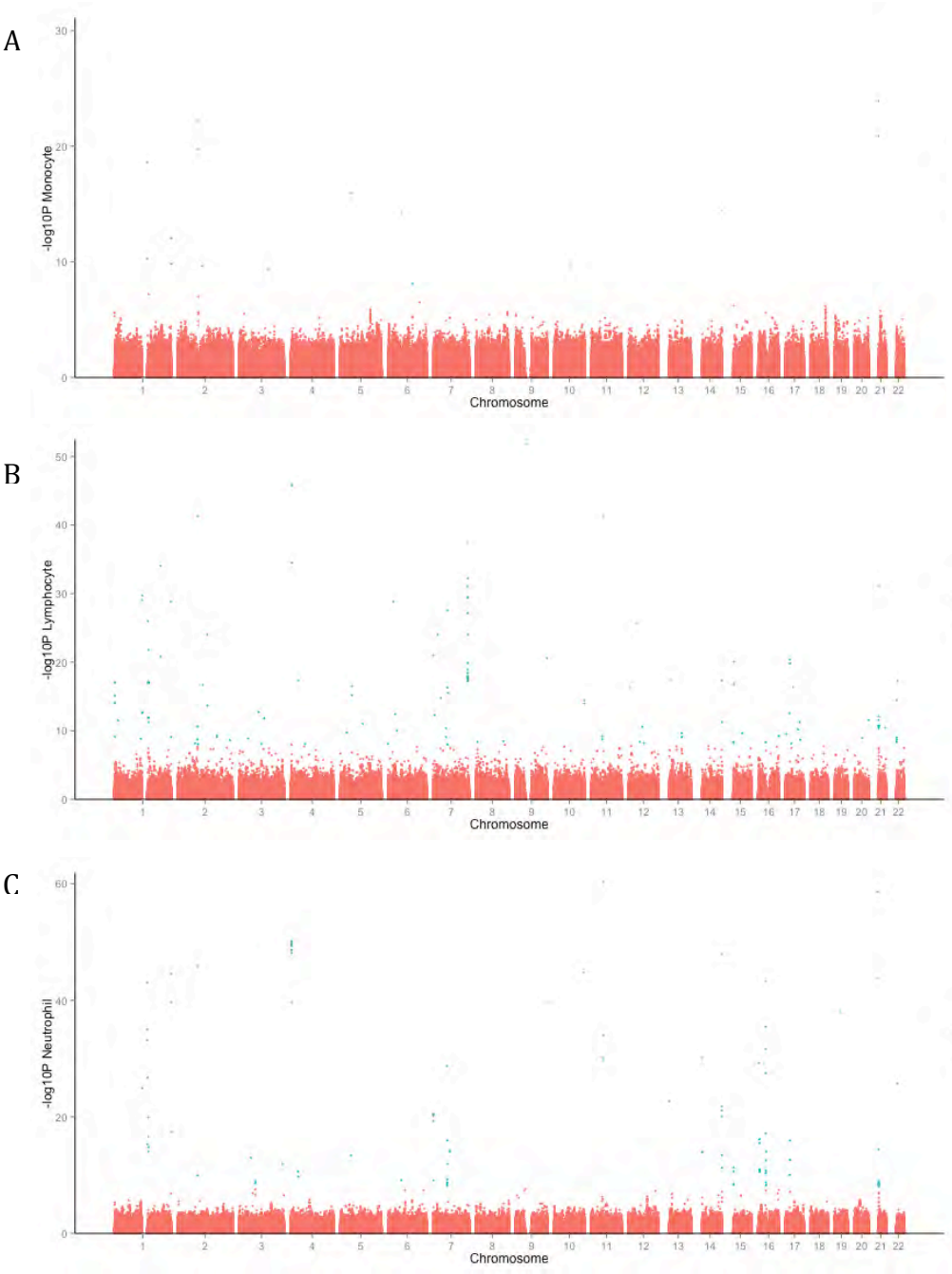


Figure 19: Manhattan plot showing GWS ($p<5x10^{-8}$) variants (blue) and non-GWS variants (red) associated with (A) monocyte, (B) lymphocyte or (C) neutrophil counts in 5768 adults.

counts respectively (Figure 19 and Supplementary Table 6).

Concordant with previous observation of genetic correlation between monocyte and lymphocytes(Naranbhai, Cambridge Bioresource et al.), considering 39,125 variants with nominally significant association with both monocyte and lymphocyte counts, there was a large excess, compared with the null expectation, of variants with the same direction of effect (38,275 directionally concordant, $p=4.9x10^{-324}$). This observation is consistent with the presence of genetic variants that would tend to act to alter monocyte and lymphocyte counts in a concordant direction, and hence possibly be distinct from variants that have discordant effects on monocytes and lymphocytes and hence influence the MLR. Interestingly, the effect size of the variant on monocyte and lymphocyte counts was correlated regardless of whether the direction of effect was concordant ($r=0.94$, $p<2.2x10^{-16}$) or discordant ($r=-0.80$,$p<2.2x10^{-16}$) (Figure 20).

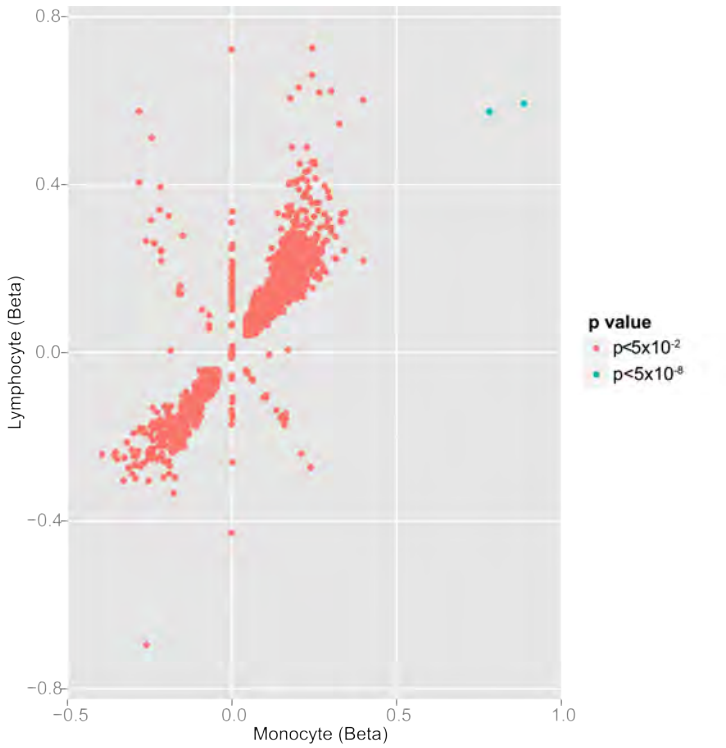


Figure 20: Effect size of 39,125 variants nominally associated with both lymphocyte and monocyte counts. Colour denotes GWS variants (blue) or associated but non-GWS variants (red).

Genome-wide identification of correlates of MLR and NLR

A total of 64 variants were significantly ($p<5x10^{-8}$) associated with the MLR and 183 with the NLR, representing 19 and 57 loci respectively (Figure 21 and Supplementary Table 6). Interestingly 4 variants (1 locus) were significantly associated with both the NLR and MLR.

We assessed whether the variants associated with the MLR or NLR may mediate their effects through association with a constituent of the MLR or NLR as evidenced by, at least, nominally significant ($p<0.05$) association in

the appropriate direction. Of 64 variants associated with the MLR, 28 variants (3 loci) had the same direction of effect on monocyte count, 0 had an opposing direction of effect on lymphocyte count and 36 (16 loci) were associated with the MLR but neither monocyte nor lymphocyte counts (Figure 3A). For the NLR, 32 variants (5 loci) had the same direction of effect on neutrophil counts, 9 variants (3 loci) had an opposing effect on lymphocyte counts and, 142 variants (49 loci) were associated with the NLR but neither neutrophil nor lymphocyte counts (Figure 21B). Taken together these data suggest that genetic modifiers of the MLR and NLR do not entirely operate through modulation of the ratios constituents, but that majority act to alter ratios with no significant effect on either constituent alone.

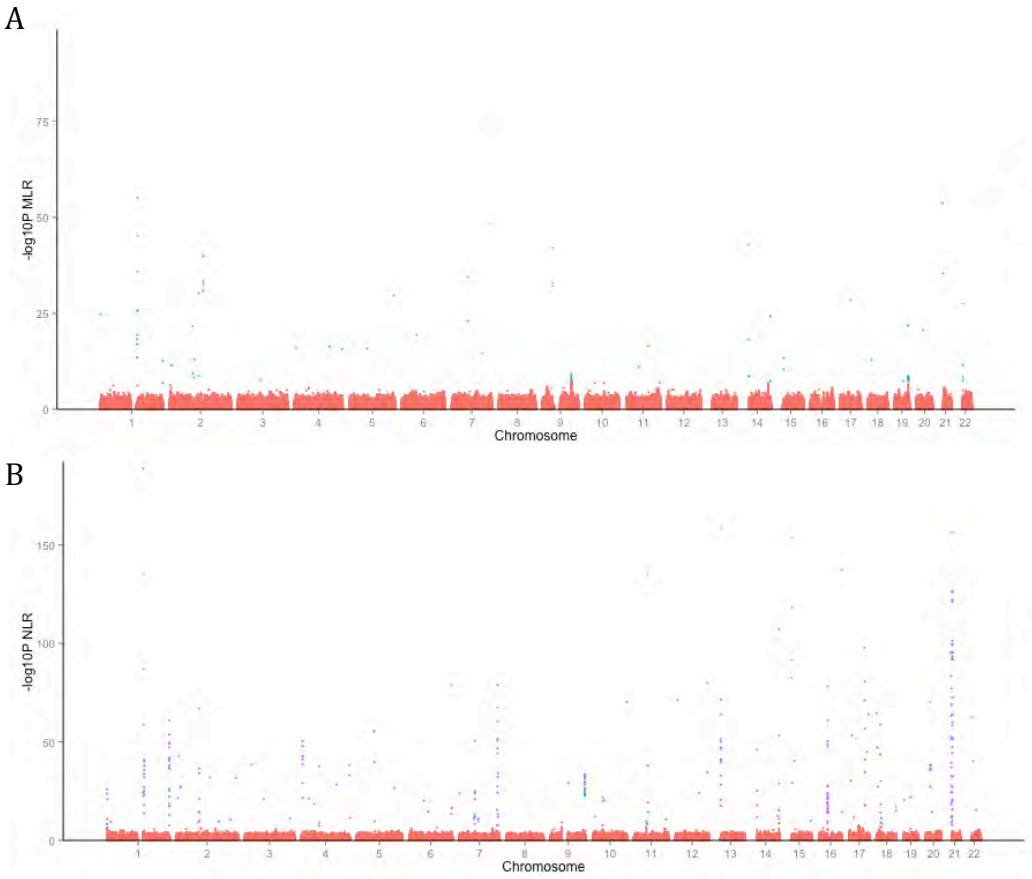


Figure 21: Manhattan plot of meta-analysis results of genome-wide variants associated with (A) MLR or (B) NLR. In A, blue variants are those GWS associated with the ML but neither nominally associated with monocyte nor lymphocytes, green variants are those associated with ML and nominally associated with monocyte counts. In B, purple variants are those GWS associated with the NL but neither nominally associated with neutrophil nor lymphocyte count, blue variants are those associated with the NL and neutrophil counts and green variants are those associated with the NL and lymphocyte counts.

For example a locus near LPAR1 on chr9 tagged by the rs10980797 variant is GWS associated with elevated MLR, and nominally significantly associated with monocyte counts in this, and a previous(Nalls, Couper et al. 2011) study (Figure 22) and is an eQTL in monocytes as shown in chapter five.

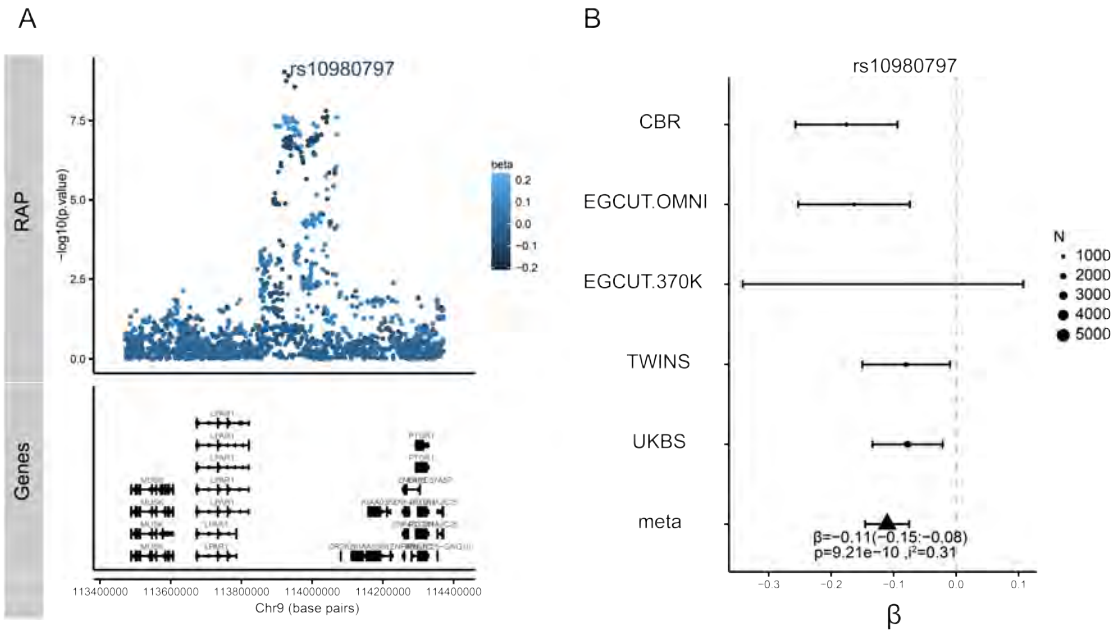


Figure 22: (A) Regional association plot showing variants on Chr9 associated with the MLR, and (B) a forest plot showing primary association statistics for rs10980797, the peak SNP across all five participating cohorts and in the meta-analysis. Whiskers in B extend to 95% of the beta estimate

Enrichment for transcription factor binding motifs amongst MLR and NLR associated loci

We speculate that variants that alter the MLR and NLR may exert the relevant effect in lineage-biased haematopoietic stem cells in order to give rise to altered ratios. Although the transcriptional landscape of these cells is yet to be described, expression of specific transcription factors is a distinguishing feature of different haematopoietic pathways. First about 13% (43/333) of all GWS variants associated with any leucocyte trait are bound by a transcription factor in the ENCODE database(Boyle, Hong et al. 2012). Secondly, we identified 8 and 10 transcription factors enriched for motifs proximal to MLR and NLR GWS variants respectively using oPOSSUM

(Figure 23A and B). Notably, the transcription factors with apparent motif enrichment proximal to MLR associated genes includes IRF2, a transcription factor that we previously implicated the MLR, and for NLR includes CEPBA, a transcription factor known to be necessary for emergency granulopoeisis(Zhang, Zhang et al. 1997). Similarly, involvement of retinoid X receptor alpha (RXRA) in MLR and NLR is consistent with the evidence that when RXRA/VDR is knocked out in mice, the effects on myelopoiesis and lymphopoeisis are opposing(Collins 2002). These data may help further refine understanding of the mechanism through which genetic variants regulate MLR and NLR, and help identify the transcription factors that could be implicated in lineage biased HSC.

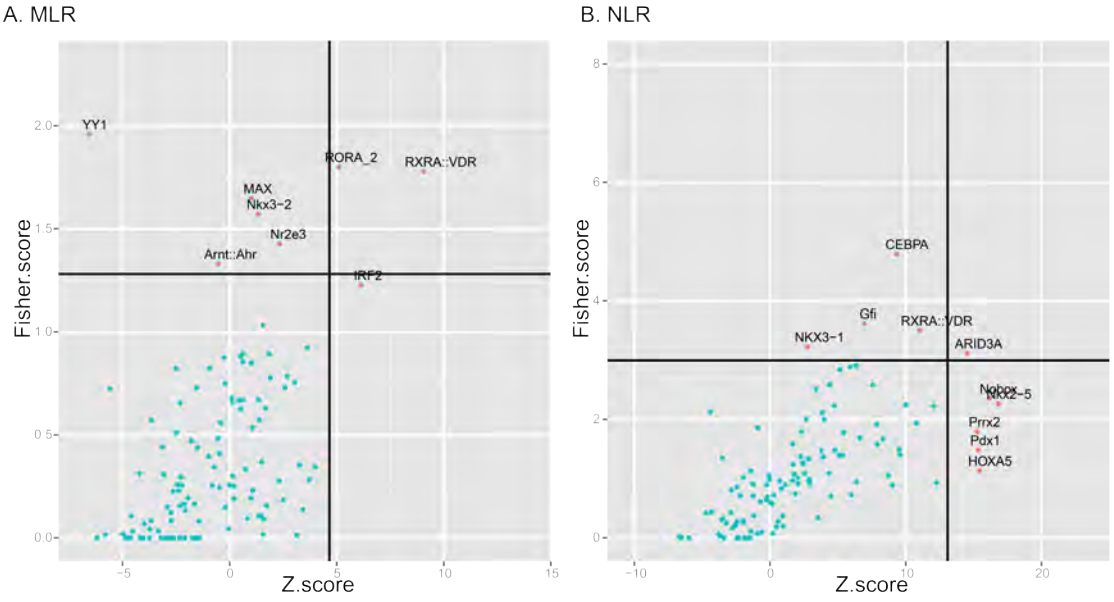


Figure 23: Enrichment of transcription factor binding sites amongst genes most proximal to (A) MLR or (B) NLR associated genetic variants. The Fisher's exact test (x-axis) from oPossum disregards the number of times a TFBS occurs within a promoter in comparing observed vs expected TFBS frequency, whilst the z-score (y-axis) is a measure of the frequency of occurrence allowing for more than one TFBS motif within a gene promoted compared with the background expectation. The mean+2SD is plotted in black. Genes with enrichment scores greater than the mean+2SD are labelled and denoted by red dots.

4.3.4 Discussion

Identification of the genetic correlates of leucocyte frequency has been undertaken before in view of these traits being involved in a wide range of disease states. In this section, I define host genetic correlates of the MLR and NLR, both of which are evolving biomarkers of infectious and malignant disease risk and outcome. We identify more than a dozen loci associated with each trait, define the extent to which the association is explained by effects on either constituent and indicate one promising aspect of plausible mechanism by which specific genetic variants give rise to altered MLR and NLR, namely lineage-constraint of transcription factors.

Our study has two notable limitations. Firstly, it is confined to Caucasian populations; whether the architecture is shared across other populations remains to be seen. Secondly, identification of pleiotropic loci is challenging and likely requires a more sophisticated approach than that used here, in particular to define a non-pleiotropic locus

In summary, these data suggest new genetic loci that regulate leucocyte counts. By identifying loci involved in the regulation of the MLR and NLR, these data delineate instrumental variables that may be useful in Mendelian Randomisation studies of the MLR and NLR in human disease.

4.4 Conclusions

In these two studies I show that the ML and NL ratio are highly genetically and phenotypically correlated, suggesting that their associated with disease may be collinear. Using a variety of approaches I demonstrate that there likely

exist specific genetic loci associated with leucocyte counts and that the architecture is focused on specific genetic loci as opposed diffusely distributed across the genome. In the second part of this chapter we find evidence for a large number of genetic loci potentially explaining the ML and/or NL ratio. These loci may be good candidates for Mendelian randomization studies.

In the next chapter I explore how genetic variants regulate gene expression in monocytes and neutrophils as a complementary approach to defining how genetic variants affect absolute numbers of these cells.

Chapter Five. Genetics of gene expression in myeloid cells

5.1 Introduction

Genetic effects on disease are mediated through alteration either in the quantity or quality of protein product that the gene encodes for. In a series of previous studies, genetic regulation of gene expression was shown to be cell-type specific and context specific. In the studies of cell-type specifity comparison was between monocytes and B-cells, leucocytes of myeloid and lymphoid lineages respectively. To explore how cell-type conditions the effect of genetic regulators of gene expression in two more similar cell-types, I performed an eQTL mapping study in monocytes and neutrophils, both of the myeloid lineage.

5.2 Genomic modulators of gene expression in human neutrophils

5.2.1 Background

Variation in the human genome is a major regulatory mechanism of gene transcription(Wright, Sullivan et al. 2014). Regulatory variants may modulate gene expression of local genes (*cis* eQTL, likely acting on the same chromosome) or genes at a distance on non-contiguous chromosomes (*trans* eQTL). Quantitative variation in transcription frequently leads to protein variation(Wilhelm, Schlegl et al. 2014) leading to phenotypic traits. Although cell-type specific effects were noted in early studies(Dimas, Deutsch et al. 2009), ease of sample availability may explain why the largest studies of eQTL are in cell lines(Stranger, Nica et al. 2007, Montgomery, Sammeth et al. 2010, Pickrell, Marioni et al. 2010, Stranger, Montgomery et al. 2012, Lappalainen, Sammeth et al. 2013) or mixed populations of leucocytes from peripheral blood(Fehrmann, Jansen et al. 2011, Mehta, Heim et al. 2013, Westra, Peters et al. 2013, Battle, Mostafavi et al. 2014). Studies of eQTL in monocytes(Zeller, Wild et al. 2010, Fairfax, Makino et al. 2012, Raj, Rothamel et al. 2014), monocyte-derived dendritic cells(Lee, Ye et al. 2014), B-lymphocytes(Fairfax, Makino et al. 2012), regulatory T-cells(Ferraro, D'Alise et al. 2014), CD4+ T-cells or bulk T-cells(Dimas, Deutsch et al. 2009) demonstrate that the effects of a genetic variant on gene expression differ according to cell type and these are further conditioned by cellular activation state and stimulatory environment(Barreiro, Tailleux et al. 2012, Fairfax, Humburg et al. 2014, Kim, Becker et al. 2014). Similar results comparing tissues support this conclusion(Schadt, Molony et al. 2008, Gibbs, van der

Brug et al. 2010, Nica, Parts et al. 2011, Ramasamy, Trabzuni et al. 2014). Collectively, these studies demonstrate the need for context-specific eQTL mapping in diverse primary populations of cells informative for disease.

Neutrophils make up 40-70% of the total circulating leucocyte pool and due to their abundance in blood and tissue, they are frequently observed in tissue specimens with minimal blood contamination. Nevertheless, challenges in isolating neutrophils(Grisham, Engerson et al. 1985) (as opposed bulk granulocytes that paradoxically bias RNA measures towards eosinophils(Stejskal, Koutna et al. 2010)) have limited the ability to study these cells leading to attempts at *in silico* prediction of neutrophil eQTL(Westra, Arends et al. 2014). About 10^{11} neutrophils are produced daily in the adult bone marrow and constitute first-responders in the innate immune response to a variety of infectious and non-infectious insults(Amulic, Cazalet et al. 2012). Neutrophils are characterised by multilobed nuclei and an abundance of primary, secondary and tertiary soluble-defence-mediator-filled granules. Although neutrophils are classically thought to be short-lived cells, they play roles in acute and chronic inflammation and can migrate into tissues or return to the blood compartment and survive(Kolaczkowska and Kubes 2013). Neutrophils have extensive crosstalk with each of the major blood cell subsets (megakaryocyte, myeloid and lymphoid)(Bennouna, Bliss et al. 2003, Pelletier, Maggi et al. 2010, von Bruhl, Stark et al. 2012, Diana, Simoni et al. 2013) further magnifying their function in health and disease. Neutrophils are thus central to orchestrating immune responses and understanding the regulation of gene expression in neutrophils may, we hypothesised, offer insights into disease biology.

5.2.2 Methods

Volunteer enrolment and cell purification

This study was approved by the Oxfordshire Research Ethics Committee (COREC reference 06/Q1605/55) and each individual gave informed consent to participation. The baseline characteristics of individuals has been previously described(Fairfax, Humburg et al. 2014). For this substudy we recruited 144 healthy Caucasian volunteers in Oxford, United Kingdom. A trained professional nurse and/or a medical doctor conducted a verbal review of clinical history to determine eligibility based on the absence of major chronic illness, current medication administration or symptoms of infection. We excluded 43 individuals because we did not have genotyping information at the time of this report (22 individuals) or the sample was an outlier after principal components analysis (20 individuals) or for whom expression profiling failed (1 individual). We used an automated outlier detection algorithm 'detectOutlier' (*lumi* package) based on Euclidean distance to the centre of the cluster with iterative outlier removal and normalisation. The median age of the 101 included participants was 31 years (IQR 24-41), 50 were male and 51 female. Whole blood was collected into sodium-heparin containing blood collection tubes (Becton Dickinson) and processed with 4-6 hours after collection.

The methods used in monocyte isolation have been previously described(Fairfax, Humburg et al. 2014) but briefly, we isolated peripheral blood mononuclear cells by density gradient centrifugation of blood-diluted with Hanks Buffered Saline solution (HBSS, Life Technologies, UK) layered

on Lymphoprep (Axis-Shield, Norway), and sorted CD14+ monocytes using magnetic-activated cell sorting (MACS, Miltenyi Biotech).

We isolated granulocytes using density gradient centrifugation in which blood was layered on Polymorphrpep (Axis-Shield, Norway) according to the manufacturers instructions. To remove red blood cell contamination we exposed granulocytes to endotoxin free cell-culture water (Life Technologies, UK) for 30 seconds and restored isosmosis using 2X HBSS. As neutrophils and eosinophils share density properties (1.085g/ml), and previous studies demonstrated that *ex vivo* granulocyte transcriptomes are dominated by eosinophil transcripts despite their relative paucity(Stejskal, Koutna et al. 2010) (<5% of granulocytes), we further isolated CD16+ granulocytes as a pure population of neutrophils using CD16+ microbeads. Purity assessed on a representative sample was >90%. We performed full blood counts on a Sysmex haematology analyser in a certified clinical laboratory; median neutrophil counts were 3.07 (IQR 2.52-3.78) $x10^9$ cells/L. We note that the use of beads directed against CD16 may, in principle, alter or stimulate neutrophils although all purification steps were carried out on ice or at 4°C. Cell viability after isolation was >80%. Lysed cell pellets were immediately cryopreserved until extraction.

Genotyping

The methods used for genome-wide genotyping have been previously described. Briefly, 733,202 variants were genotyped using the Illumina HumanOmniExpress-12v1.0 Beadchip. QC steps included removal of individuals that were outliers by PCA, had poor genotyping call rates or heterozygosity measures. Variant QC included removal of SNP's with

MAF<5%, HWE<$1x10^{-6}$ or poor call rates. Genome-wide imputation was performed by prephasing a scaffold of 588,170 SNP's with SHAPEIT(Delaneau, Marchini et al. 2012) and imputation in IMPUTE2(Howie, Donnelly et al. 2009) using the 1000G phase 1 release as the reference. Imputed SNP or indels with an info score <0.9, MAF<5% or departure from HWE ($1x10^{-3}$) were removed. In total 5,680,354 variants were included in the eQTL analysis. Locations, where reported, are according to the human genome build GRCh37. Measures of LD, where mentioned, are based on 1000G CEU(Genomes Project, Abecasis et al. 2012) or, if specifically mentioned, from this cohort of volunteers.

Gene expression analysis

During optimization of methods to isolate RNA from neutrophils we found that commonly used column-based isolation techniques were vastly outperformed by phenol-chloroform extraction using Trizol LS (Life Technologies) according to the manufacturers instructions. This may be because a higher concentration of chaotropic and reducing agent is required to fully inhibit the abundance of nucleases in neutrophils. Therefore RNA was extracted using Trizol LS and further cleaned using the RNA MinElute cleanup kit (Qiagen, UK). RNA was quantified and integrity assessed using a Bioanalyser RNA 6000 Nano kit (Agilent, UK). Gene expression was quantified using the Illumina HumanHT-12 v4 BeadChip gene expression array platform with 47,231 probes according to the manufacturers instructions. Samples were randomised across expression chips and run in a single batch.

Gene expression data were normalized using random-spline normalization, transformed by variance-stabilising transformation and sample outliers were

iteratively removed and normalization repeated. We note that the sample outliers were almost universally (19/20 samples) due to cRNA quality with these individuals having RNA size distribution<800 by BioAnalyser RNA 6000 Nano kit evaluation. We therefore did not obtain replacement arrays and hence all the data in this study derive from a single batch. Probe sequences mapping to more than one genomic location or regions with underlying polymorphisms frequent in >1% of the population were excluded from eQTL analysis (n=18220 probes). Only probes that were expressed and detected (GenomeStudio probe detection p<0.01) in neutrophils (n=11,023 probes for 9147 genes) were included in the primary analysis. For comparisons between neutrophils and monocytes, 10,012 probes for 8362 genes were included from 99 individuals.

eQTL mapping

We tested for association between genetic variation (using dosage estimates for imputed variants) within 1Mb of the expression probe (*cis*) or more distantly (*trans*), and gene expression in a linear regression framework implemented in 'MatrixEQTL'(Shabalin 2012) in *R (*Matrix_eQTL_main function*),* incorporating principal components as covariates to account for hidden confounders (empirically determined number of PCs =15). We filtered results by a false discovery rate of 5% for *cis* associations or 1% for *trans* associations. We caution that this study, due to its modest size, is not adequately powered to exhaustively identify eQTL in neutrophils, or to map *trans* eQTL.

Annotation of eQTL with genomic features

Identified eQTL were annotated by location relative to the transcription start site or gene structures using data from the UCSC genome browser (GRCh37/b19)(Karolchik, Barber et al. 2014).

We accessed whole genome bisulfite sequencing and histone chromatin immunoprecipitation data generated by the BLUEPRINT consortium in primary human monocytes and neutrophils from 4-8 individuals (depending on feature). We defined features marked by an epigenetic mark as those present in at least half of all individuals (supplementary table 7).

For analyses of eQTL overlap with the location of epigenetic marks(Adams, Altucci et al. 2012), enhancers(Andersson, Gebhard et al. 2014, Consortium, the et al. 2014), transcription factors(Boyle, Hong et al. 2012), conserved microRNA binding sites(Betel, Koppal et al. 2010) and in regions that have been under selection(Grossman, Andersen et al. 2013), the location of the given feature relative to an eQTL was calculated using the 'GenomicRanges' package in *R* with the *'distanceToNearest'* function. Fishers exact test was used to evaluate enrichment relative to all imputed variants tested for eQTL (ie those within 1Mb of a gene expressed in neutrophils and monocytes). The code used to generate the results is available from the authors.

Overlap between eQTL and variants associated with complex traits/disease

As previously described(Fairfax, Humburg et al. 2014) 1138 traits listed in the NHGRI Catalog Genome-Wide Association Studies (accessed 14 October 2014) were abstracted and expertly curated into one or more categories

according to the disease system affected (28 potential categories). A Fishers exact test comparing the proportion of eQTL that are trait associated variants or in linkage disequilibrium with these variants (r^2>0.8 in 1000G Caucasian populations) to the proportion of all tested variants that are associated variants was performed stratified by disease category.

Neutrophil stimulation

To identify the role of rs1981760 in NOD2 downstream effects, we stimulated 5-10 x10^6 freshly isolated neutrophils resuspended at 2x10^6 cells/ml in RPMI1640 supplemented with 10% foetal calf serum and L-glutamine with 1μg/ml muramyl dipeptide (n-acetyl) and 1 μg/ml Pam3CSK4, (both from Invivogen, UK) for two hours. Primer sequences for quantitative PCR are reported in Supplementary Table 8.

Chromatin immunoprecipitation and digital droplet PCR (ddPCR) assessment of allele-specific binding of STAT3 in Neutrophils

Crosslinked DNA from neutrophils was stored and subjected to chromatin immunoprecipitation (ChIP) following a protocol previously described(Wong, Lee et al. 2014) with one modification: fragmentation of crosslinked DNA was performed by sonication consisting two sets of 9 cycles of 30s each in a BioRuptor (Diagenode, Belgium) at high power. A mouse anti-human STAT3 antibody (Cell Signaling-NEB, UK) was used in conjunction with magnetic beads to immunoprecipitate STAT3 bound DNA.

For allele-specific ddPCR quantification of rs1981760-C compared with rs1981760-T we designed allele-specific probes (Supplementary Table 13), and performed amplification using the ddPCR Supermix for Probes (BioRad) according to the supplied protocol. The probe for the T and C allele were designed to fluoresce in different channels allowing simultaneous detection of both in a single reaction. Droplets were generated on a qx100 droplet generator (BioRad, UK) and droplets read on a qx100 droplet reader. Probes targeting SOCS3 and JAK3 were used as positive controls for ChIP of STAT3, and RAB4 as a negative control. Input DNA as well as DNA obtained after ChIP was amplified on the same plate. For allele-specific wells, we performed technical duplicates.

The proportion of positive droplets relative to negative droplets is associated with the absolute concentration of the product according to a poisson distribution. We therefore exploited this to empirically calculate the expected proportion of C vs. T droplets (so as to take into account possible probe efficiency differences) and compared this to the actual ratio after ChIP. P values using a chi-squared test with 2 degrees of freedom were used to estimate the probability of the observed vs. the expected ratio being due to chance.

Software used for analysis

Analyses were conducted in *R,* using the following packages: limma,lumi, 'annotate', 'vsn', 'GenomicRanges', 'qvalue', 'data.table'. For data presentation the 'gridExtra', 'ggbio', 'ggplot2' and 'VennDiagram' packages were used.

Local association plots were created using LocusZoom. Ingenuity Pathway Analysis where used, was performed using a background set of all human genes on the Ilumina array. A listing of external files and their sources is given in Supplementary Table 9. FDR where shown denotes the false discovery rate calculated using the Benjamini Hochberg procedure using the p.adjust function.

5.2.3 Results and Discussion

Defining eQTL in primary human neutrophils

To identify regulatory variants for gene expression in primary human neutrophils, we enrolled 101 healthy adult European volunteers in Oxford, United Kingdom (Figure 24a). Individuals were genotyped on the Human OmniExpress 12v1.0 chip, and we used genome-wide imputation with stringent quality checks to infer additional single-nucleotide polymorphism (SNP) or simple insertion/deletion genotypes with high confidence. CD16+ neutrophils, isolated by a two-step sequential gradient-density and immunomagnetic sorting procedure yielding a highly purified population, were subjected to whole-transcriptome characterization by hybridization of cRNA to Human HT-12 v4 Expression BeadChips (Illumina). We assessed only genes for which specific probes existed, and which were confidently detected in more than 5% of the cohort (GenomeStudio detection $p<0.01$). We used a widely-adopted linear-additive modeling approach (MatrixeQTL)(Shabalin 2012) adjusting for principal components which has been shown to enhance eQTL discovery and control for technical and demographic heterogeneity(Leek and Storey 2007, Fehrmann, Jansen et al. 2011, Fairfax, Humburg et al. 2014, Westra, Arends et al. 2014).

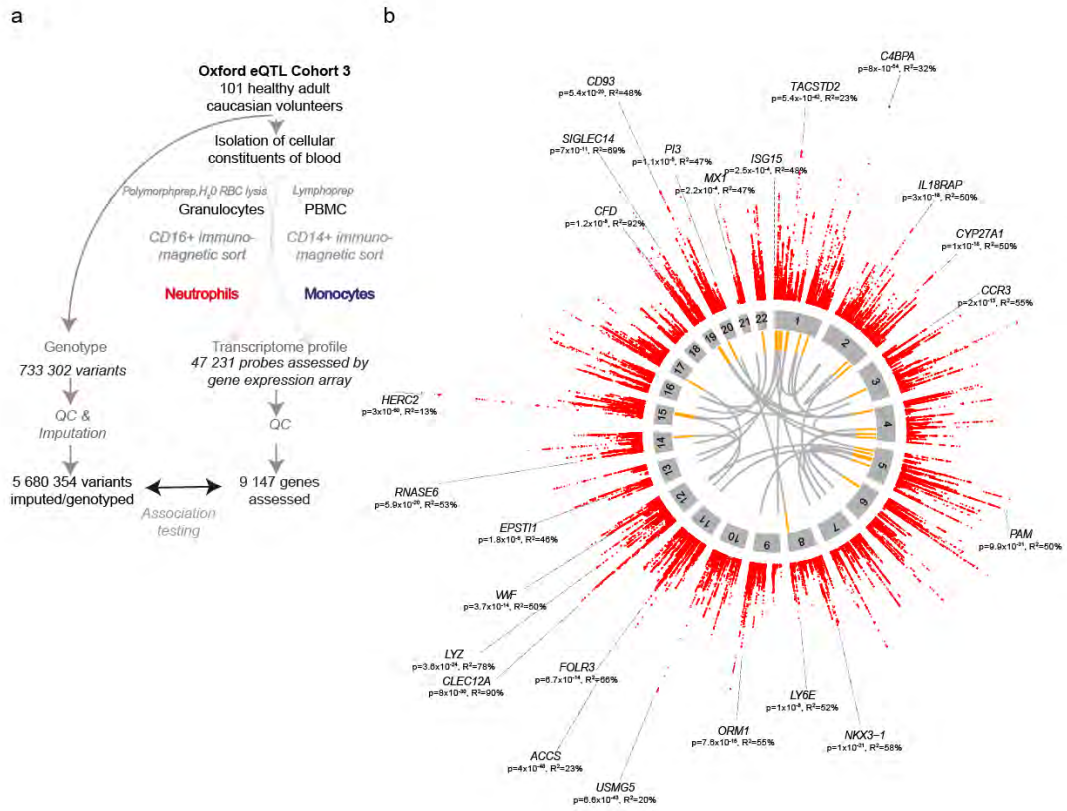


Figure 24: Study schema and overview of *cis* and *trans* eQTL in primary human neutrophils. (a) Flowchart showing the schema of this study to identify eQTL in human neutrophils from healthy adult volunteers of European ancestry with genotypes determined by array-genotyping and by imputation tested for association with global gene expression in *cis* (variant <1Mb of gene) or *trans* (variant on non-contiguous chromosome to gene). This analysis was contemporaneous to our previously reported study(Fairfax, Humburg et al. 2014) of monocytes enabling direct comparison of genetic correlates of gene expression between neutrophils and monocytes. (b) Circos plots for the neutrophil dataset. Outermost rim (red dots) shows a Manhattan plot for significant *cis* eQTL (FDR<0.05) with names of genes with large effect sizes (variance explained, R^2>45%) or lowest five p-values; the second rim (grey boxes) shows the chromosome ideogram with chromosome number within each box; the third rim (orange boxes) indicates a gene to which a significant (FDR<0.01) trans-eQTL was observed; and the innermost spokes (grey) connect trans-acting variants to the gene the variant is associated with.

After genome-wide imputation 3281 genes (~30% of 9147 genes tested and equivalent to 3675 probes) had one or more identifiable loci within 1Mb (defined as *cis*-eQTL) of a probe that was significantly associated with gene

expression at a false discovery rate (FDR) threshold of 5% (Figure 24b and Supplementary Table 10). The median proportion of variance explained by the most significant variant for each probe, was 11% but notably was greater than 50% for several genes, such as the anti-inflammatory monosodium urate receptor *CLEC12A(Neumann, Castineiras-Vilarino et al. 2014)*, complement factor d (*CFD*) and the antibacterial enzyme lysozyme (*LYZ*) (Additional genes with large effect sizes are highlighted in Figure 1b). Genes with an eQTL were more highly expressed (median normalised expression 8.14 vs. 7.7, $p=1.7x10^{-72}$) and had greater variance (median variance 0.049 vs 0.029, $p=1.07x10^{-122}$) than genes without an eQTL. The most significant eQTL per gene (denoted as the peak eQTL) clustered around the transcriptional start site (TSS), and the effect sizes increased with proximity to the TSS (Figure 25a) but, annotation and inspection at higher-resolution demonstrates distribution of eQTL across gene structures with clustering around TSS and transcription end sites, consistent with previous studies(Veyrieras, Kudaravalli et al. 2008) (Figure 25b).

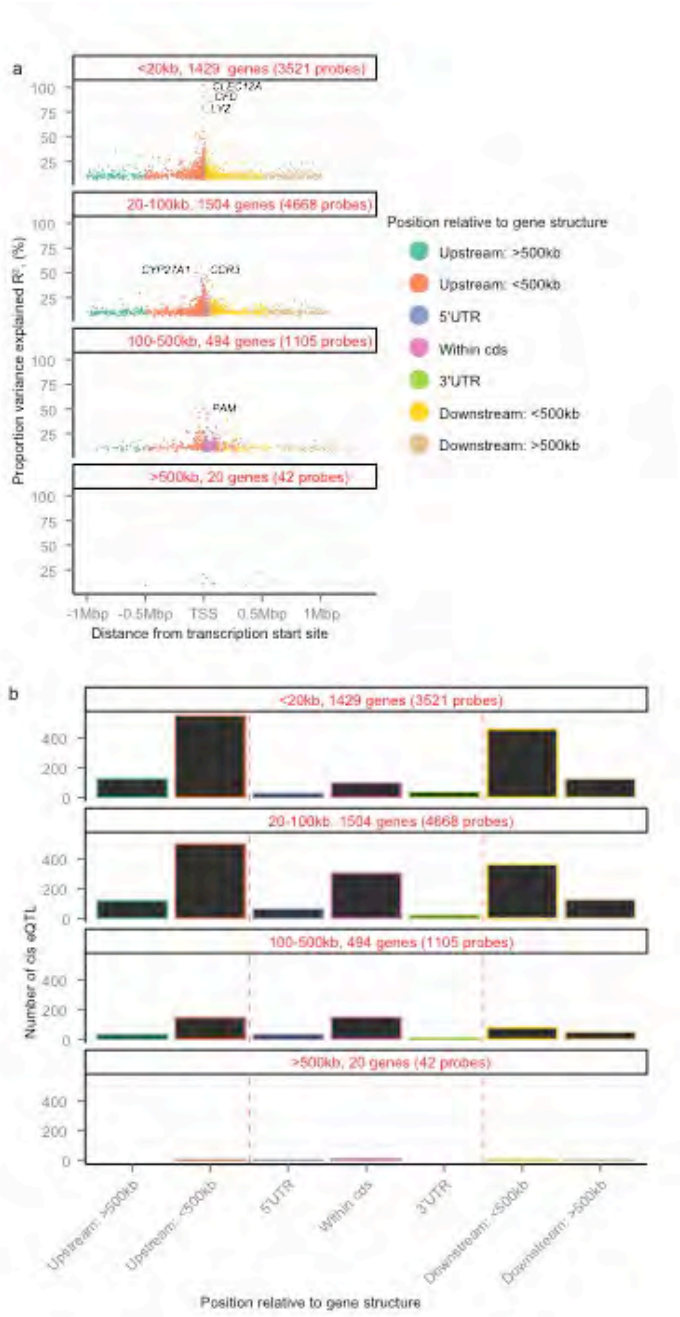


Figure 25: Effect size and location of *cis* eQTL in neutrophils relative to transcription start site (TSS) (a) or gene structure (b), partitioned by gene size. Each point in (a) denotes the peak eQTL for a gene and is plotted according to the proportion of variance explained by the variant with strongest statistical association with gene expression and location of the variant relative to the TSS. Several genes with large effect sizes are annotated. (b) shows the number of *cis* eQTL identified in neutrophils according to the location of the peak eQTL relative to gene structures.

We did not identify association between any of 168 copy-number variants, inferred from raw genotype calls, and gene expression after correcting for multiple comparisons (Supplementary Table 11).

Comprehensive identification of loci associated with gene expression on non-contiguous chromosomes (*trans* eQTL) typically requires large sample(Westra, Peters et al. 2013) sizes due to typically reduced effect sizes of variants acting *trans* allied to an increased burden of multiple hypothesis testing. Accordingly, we identified a smaller set of 33 genes regulated in *trans*, at a false discovery rate threshold of <1% (Figure 24b and Supplementary Table 12). These *trans* eQTL include rs10012416, which we find is a *cis* eQTL for *CRIPAK* (encoding an inhibitor of p21 activated kinase Pak1 important in neutrophil cytoskeletal dynamics involved in phagocytosis(Dharmawardhane, Brownson et al. 1999)) and has *trans* effects on *AVP*, *SPTBN3* and *IRF6*; and rs10784774, a *cis* eQTL for *LYZ* that is associated with *ZNF131* in neutrophils as we previously had reported in monocytes(Fairfax, Makino et al. 2012). We note that this study is not powered to identify trans eQTL with weak to moderate sized effects.

Genes involved in the development, migration and function of neutrophils are frequent targets of genetic regulation.

We proceeded to examine genes relevant to neutrophil function for eQTL as these may serve as a narrative resource to understanding the role of regulatory variants in neutrophil biology. In the absence of an unbiased list of genes relevant to neutrophil function, we assembled a list of 164 genes

described, with a verifiable reference, to be involved in a particular aspect of neutrophil function as being involved in aspects of neutrophil function in recent reviews(Borregaard, Sorensen et al. 2007, Amulic, Cazalet et al. 2012, Kolaczkowska and Kubes 2013, Bardoel, Kenny et al. 2014, McCracken and Allen 2014). Because this approach is at risk of narrative bias, we caution against interpretation of these results as evidence of enrichment for genes important in neutrophil function. Of the 113 genes for which an expression probe existed, passed QC and was included in analysis, 104 had an identifiable variant associated with expression ($p<0.05$) and 47 after false-discovery adjustment (FDR<0.05). We observe eQTL for genes involved in most aspects of neutrophil development and function (Figure 26, and Supplementary Table 13). Interestingly, several genes with an eQTL are implicated in Mendelian disorders involving neutrophils. For example rs933222 is associated with expression of *RAC2*, a gene encoding a Rho GTPase that is part of the NADPH oxidase complex involved in initiation of phagocytosis (Figure 26l) and is involved in neutrophil immunodeficiency syndrome(Ambruso, Knall et al. 2000) (OMIM #608203).

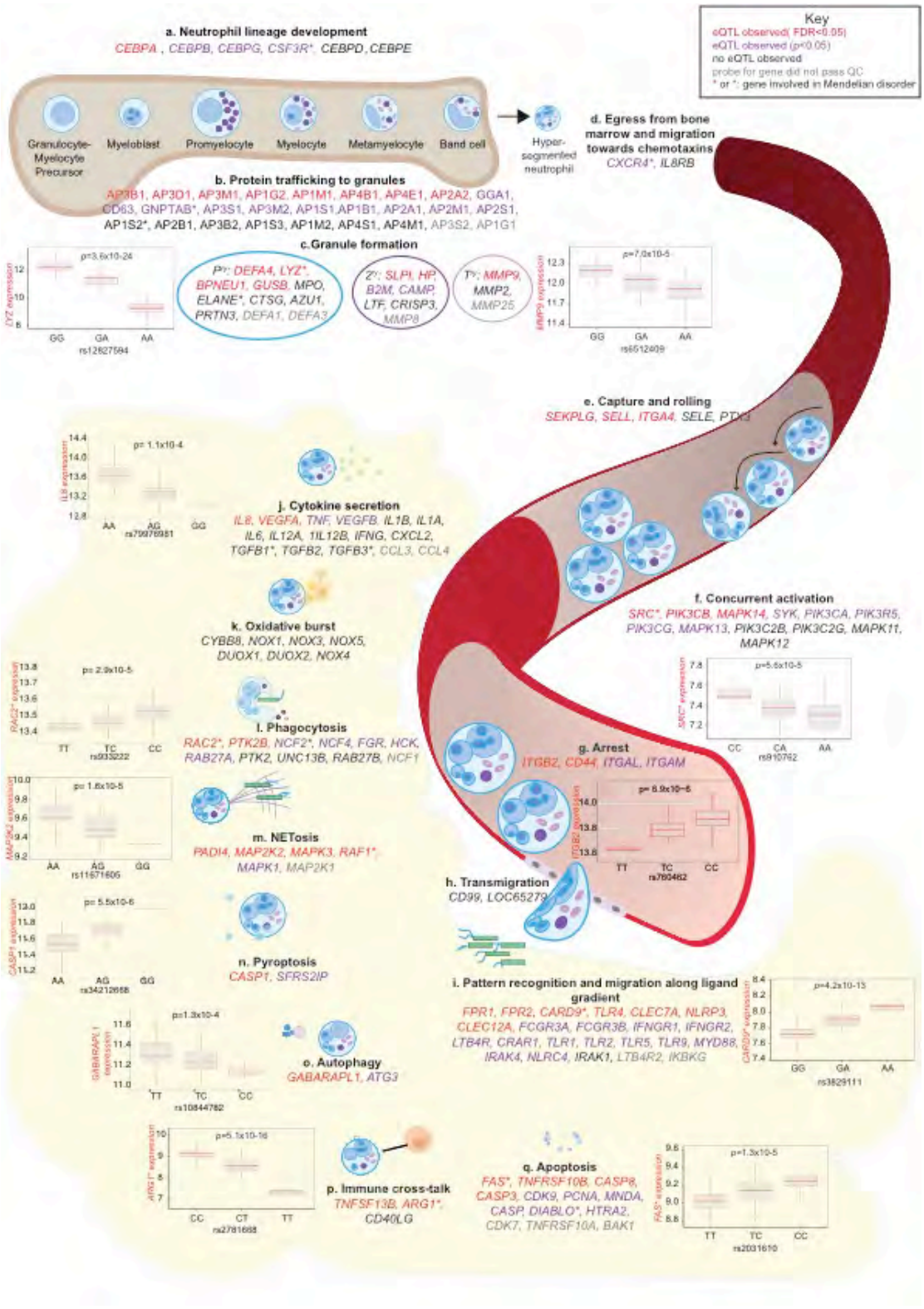


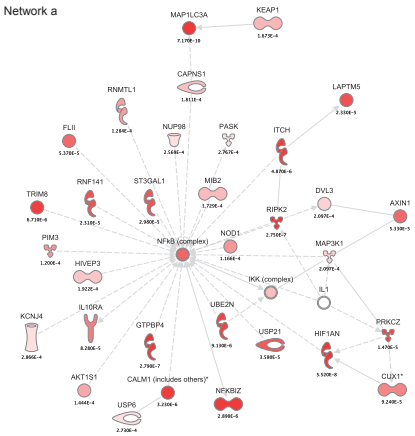
Figure 26: Genes involved in the development, migration and function of neutrophils are frequent targets of genetic regulation. We compiled a list of 164 genes from recent high-profile reviews(Borregaard, Sorensen et al. 2007, Amulic, Cazalet et al. 2012, Kolaczkowska and Kubes 2013, Bardoel, Kenny et al. 2014, McCracken and Allen 2014) on various aspects of neutrophil biology and prospectively annotated these with eQTL information and

associated mendelian disorders (asterisked genes). Probes for 113 of these genes passed QC (18 had non-specific probes, 31 genes were not expressed in >5% of the cohort or had poor array detection, grey gene names) of which 9 had no eQTL (black gene names) and 104 had an eQTL (47 with FDR<0.05 red gene names, 57 with p<0.05 but FDR>0.05 purple gene names). (a) Granulocyte-macrophage progenitor cells (GMP) divide in response to GM-CSF recognised by a receptor encoded by *CSFR3* (a gene with an eQTL and involved in hereditary neutrophilia) to form neutrophils under the instruction of CEBPα and other members of this family of transcription factors, several of which have eQTL. Neutrophil development from myeloblast into promyelocyte, myelocyte and metamyelocyte is coincidental with expression of proteins that traffic to primary secondary and tertiary granules. (b) Several genes encoding components of AP heterotetramers, involved in protein trafficking, have a significant eQTL, and expression of many of the (c) genes encoding granule contents such as lysozyme (LYZ, involved in renal amyloidosis) and matrix metalloproteinase 9 (MMP9, involved in metaphyseal osteolysis, nodulosis and arthropathy) are modified by one or more genetic variants. (d) Hyper-segmented neutrophils egress from the bone marrow in response to IL-8 and SDF-1, and are (e) captured within blood vessels through interaction of selectins and their ligands including several with an eQTL such as SEKPLG and SELL. (f) Capture, rolling and eventual arrest on endothelial surfaces is associated with intracellular events that lead to activation. Key modulators of this process include PI-3 and MAP kinases, for which many components have an eQTL. (g) Arrest of neutrophils is mediated by α1β2/αMβ2 integrin heterodimers encoded for by *ITGAL*, *ITGBA2*, *ITGAM*, for which we identify eQTL, and (h) rapid para- and trans-cellular migration ensues. (i) Neutrophils express pattern recognition receptors and receptors for host derived molecules allowing detection of bacteria and effective migration with eQTL identified including signaling adaptor proteins such as *CARD9* (involved in familial candidiasis and several complex diseases). In tissues, neutrophil response includes (j) cytokine secretion, (k) oxidative burst, (l) phagocytosis, (m) production of neutrophil extracellular traps (NETosis), (n) pyroptosis, (o) autophagy, (p) cross-talk with other immune, and eventual cell death via apoptosis should cell death not have eventuated from another response. As shown, genes involved in nearly every aspect of neutrophil response have identifiable eQTL and an example for each is shown including for *RAC2*, involved in neutrophil immunodeficiency syndrome and *FAS*, which is involved in autoimmune lymphoproliferative syndrome type 1A.

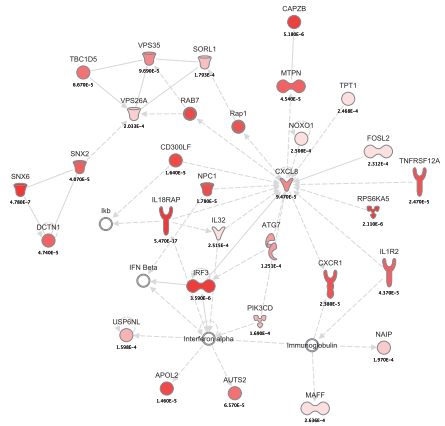
Ingenuity pathway analysis of 975 genes with an eQTL in neutrophils but not in monocytes (Figure 27, and detailed further below), revealed enrichment for functions relating to cell death ($p=1.4x10^{-7}$), apoptosis ($8.2x10^{-7}$), necrosis ($6.4x10^{-6}$); and infection, notably viral infection ($4.2x10^{-6}$) and

infection of cells ($2.2x10^{-5}$). The most significant upstream transcriptional regulator was TP53, which plays a critical role in cell proliferation and apoptosis as well as antimicrobial function highlighting how the consequences of TP53 for the individual may be modulated by eQTL in downstream mediator and effector genes. A number of cytokines were also identified as upstream regulators including IFNB1 ($p=5.9x10^{-4}$), IL15 ($3.0x10^{-3}$), IFNG ($4.1x10^{-3}$), CD40LG ($6.8x10^{-3}$) and TNF ($7.2x10^{-3}$).

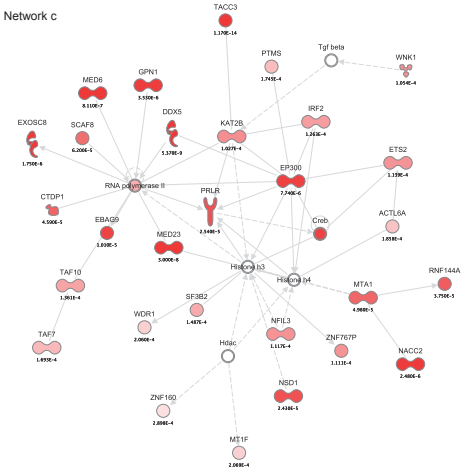
Network a



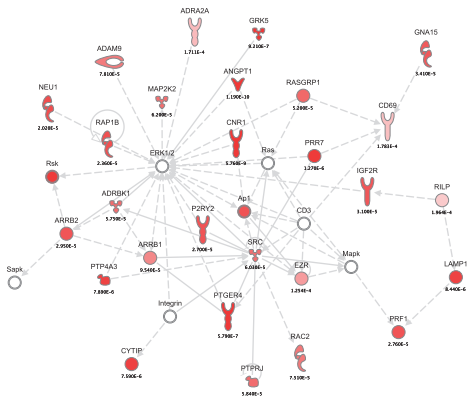
Network b



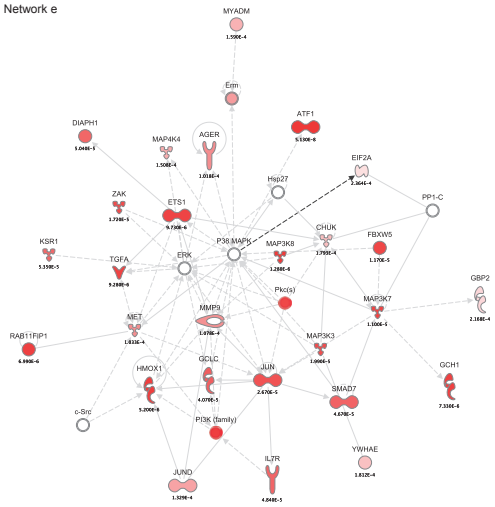
Network c



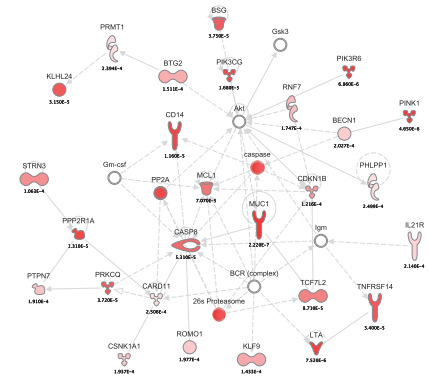
Network d



Network e



Network f



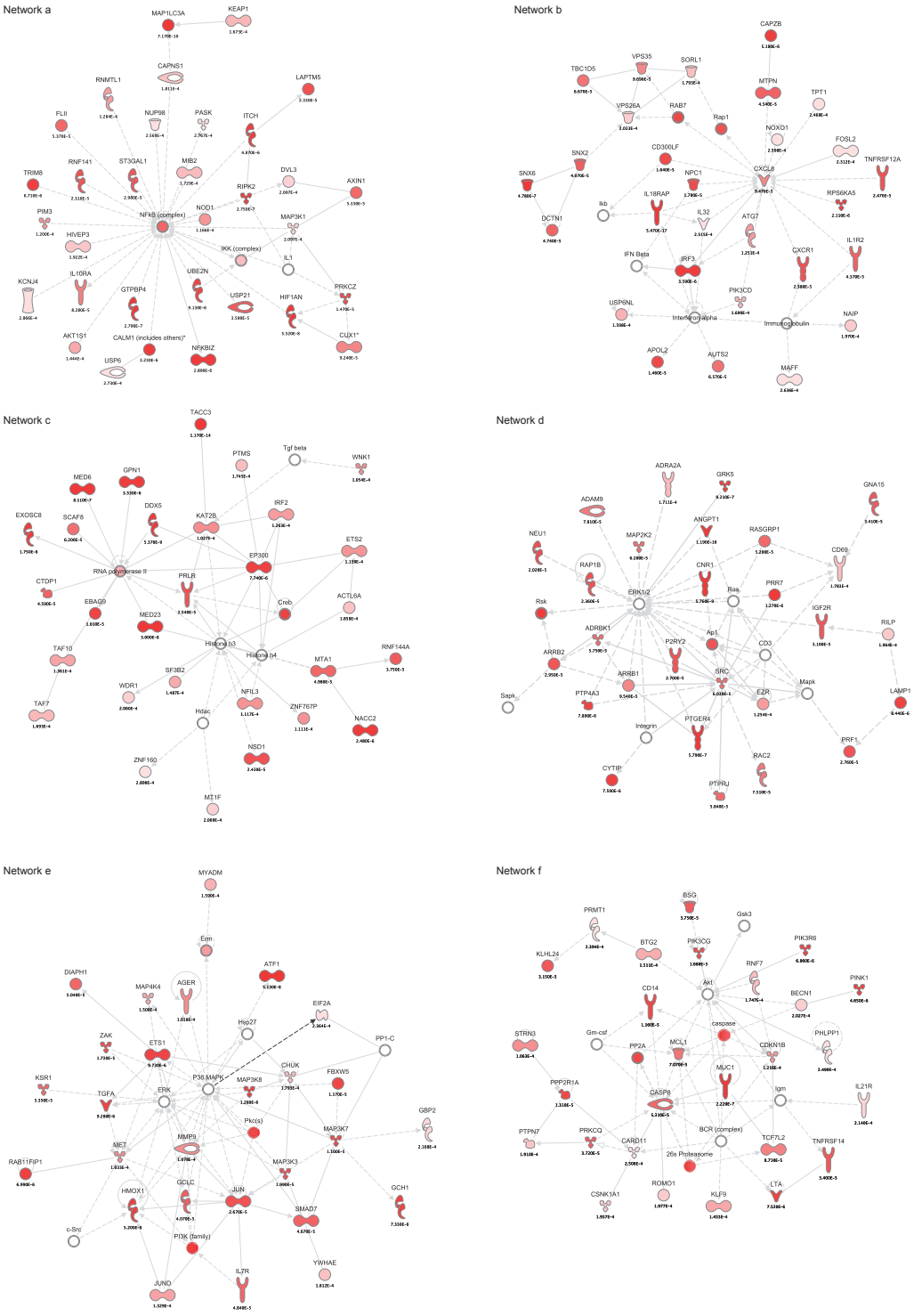


Figure 27: Pathway analysis of genes with neutrophil specific *cis* eQTL. Amongst 975 genes with an eQTL in neutrophils only, a number of gene networks are identified as enriched for the genes showing *cis* eQTL specific to neutrophils (top 10 shown here with intensity of red shading related to p value of eQTL, gene names and p values shown above and below molecular symbol). Network a ($p=1x10^{-36}$, Fishers exact test) involves genes associated with

NFkB such as *MAP1LC3A* encoding LC3 involed in autophagy ($p=7.1x10^{-10}$), *TRIM8* ($p=6.7x10^{-6}$) which modulates p53 activity and *CALM1* (calmodulin 1) ($p=3.2x10^{-6}$); network b ($p=1x10^{-30}$) involving *CXCL8* (IL8) ($p=9.5x10^{-5}$) a critical modulator of neutrophil recruitment and behaviour with associated genes showing eQTL including *IRF3* ($p=3.6x10^{-6}$), *IL18RAP* ($p=5.5x10^{-17}$) and *CXCR1* (IL8 receptor) ($p=2.4x10^{-5}$); network c ($p=1x10^{-30}$) centred on histones H3 and H4; network d ($p=1x10^{-26}$) involving genes linked with ERK1/2 MAP kinases; network e ($p=1x10^{-26}$) also focused on ERK and P38 MAPK and involving a network of genes including many transcription factors such as *ETS1* ($p=9.7x10^{-6}$), *ATF1* ($p=5.1x10^{-8}$) and *JUN* ($p=2.7x10^{-5}$); network f ($p=1x10^{-26}$) involving Akt and caspases; network g ($p=1x10^{-26}$) including eQTL involving *ATM* ($p=1.5x10^{-18}$), heat shock proteins *HSPA1A* and *HSPA1B* ($p=1x10^{-6}$) and *MCM5* ($p=3.4x10^{-11}$); network h ($p=1x10^{-23}$), network i ($p=1x10^{-19}$) involving genes linked to the nuclear transcriptional regulator NURP1 (also known as p8) showing eQTL such as the phopsholipid gene *PCTP* ($p=9.5x10^{-13}$); and network j ($p=1x10^{-18}$) involving IL12, PI3K and TCR which includes genes showing *cis*-eQTL such as *STAT5A* ($p=1.9x10^{-5}$) important in GM-CSF signaling and control of granulocyte homeostasis(Kimura, Rieger et al. 2009) and IDO1 ($p=4.8x10^{-6}$) which has a role in neutrophil self-regulation through tryptophan catabolism(El-Zaatari, Chang et al. 2014).

Patterns of eQTL in neutrophils relative to monocytes and other immune cell types

Regulation of gene expression may be constrained to specific cell-types and contexts. To delineate aspects of shared and unique regulatory genomic architecture in neutrophils, we pursued three complementary approaches.

First, we note that 63% (2069/3281) of genes with a *cis* eQTL in neutrophils are reported to have a *cis* eQTL in the largest blood eQTL meta-analysis to date(Westra, Peters et al. 2013) (obtained through the bloodeqtl browser) and recent *in silico* predictions(Westra, Arends et al. 2014) attribute 20% (443/2188) to neutrophils and 90% (1969/2188) as ‘generic’ (note that a

gene may be denoted as neutrophil *and* generic). Therefore many eQTL in neutrophils may not have been identified in whole blood studies, and even when they are, *in silico* deconvolution of cell types may not establish if a gene has an eQTL in a given cell type. Conversely, 84% of genes (415/495) with an eQTL in blood that are bioinformatically ascribed to neutrophils (and tested in our study) have an eQTL in purified neutrophils, providing evidence of cross-study validation. The estimated effect sizes in neutrophils and whole blood for genes with an eQTL in both are only moderately correlated (correlation estimate=0.52, 95%CI 0.43-0.60, $p<2.2x10^{-16}$); as we shall present later, directionally opposing eQTL amongst different cells are not uncommon and this may plausibly affect *in silico* effect size estimates. Secondly, for 40% (261/646) of genes with a *cis* eQTL in mouse CD4+ T-cells and/or neutrophils(Mostafavi, Ortiz-Lopez et al. 2014) whose human homologues we tested in humans, we observe a significant *cis* eQTL in neutrophils (a 5.72 fold enrichment over random expectation (646/9147), $p=1.02x10^{-80}$). Thirdly, we compared genes with an eQTL in neutrophils to those that have been reported previously in other primary immune cells: CD4 T-cells, regulatory T-cells(Ferraro, D'Alise et al. 2014) and B-cells(Fairfax, Makino et al. 2012) (grouped as lymphoid cells) or monocytes and monocyte derived dendritic cells(Zeller, Wild et al. 2010, Fairfax, Humburg et al. 2014, Lee, Ye et al. 2014, Raj, Rothamel et al. 2014) (grouped as myeloid cells). Of the 3281 genes with an eQTL in neutrophils, 1671 (51%) had an eQTL in both myeloid and lymphoid cells, 823 (25%) in non-neutrophil myeloid cells, 337 (10%) in lymphoid cells, and 450 have no reported eQTL in either myeloid or lymphoid cells (Figure 28). Although our analysis neither assesses whether the same

genetic variant regulates gene expression in all cell types nor whether the effect is the same, it demonstrates that, for at least 14% of the genes in which we observe an eQTL, we have identified novel regulatory variants and shown the utility of eQTL mapping in primary neutrophils.

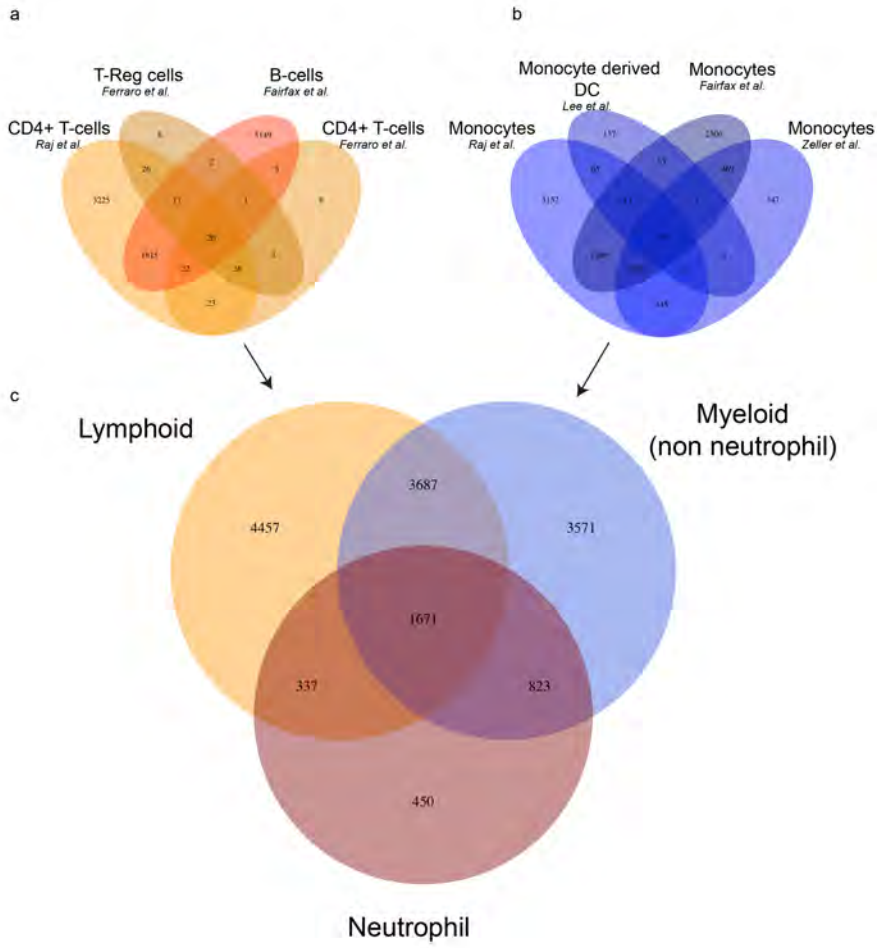


Figure 28: Euler diagrams showing overlap of genes with an eQTL previously observed in lymphoid (a) or myeloid (b) cell types and their overlap with genes with eQTL in neutrophils (c). We extracted lists of genes reported to have a significant eQTL from several previous studies and show the number of overlapping or non-overlapping genes for the four studies reporting eQTL in lymphoid cells(Fairfax, Makino et al. 2012, Ferraro, D'Alise et al. 2014, Raj, Rothamel et al. 2014) (a) or myeloid cells(Zeller, Wild et al. 2010, Fairfax, Makino et al. 2012, Lee, Ye et al. 2014, Raj, Rothamel et al. 2014) (b). Overlapping the lists of genes from A and B, with the list of 3281 genes with an eQTL in neutrophils yields (c) with number of overlapping genes noted.

To further examine the effect of cell-type on eQTL we performed a detailed analysis of monocytes and neutrophils isolated contemporaneously from 99 caucasian donors in which we confined analysis to 8362 genes expressed in both cell types. We identified 400,783 unique variant-gene associations across both cell types: 87,276 variants for 1031 genes in both neutrophils and monocytes, 118,817 variants for 2847 genes in neutrophils and 194,690 variants for 3674 genes in monocytes (Figure 29a).

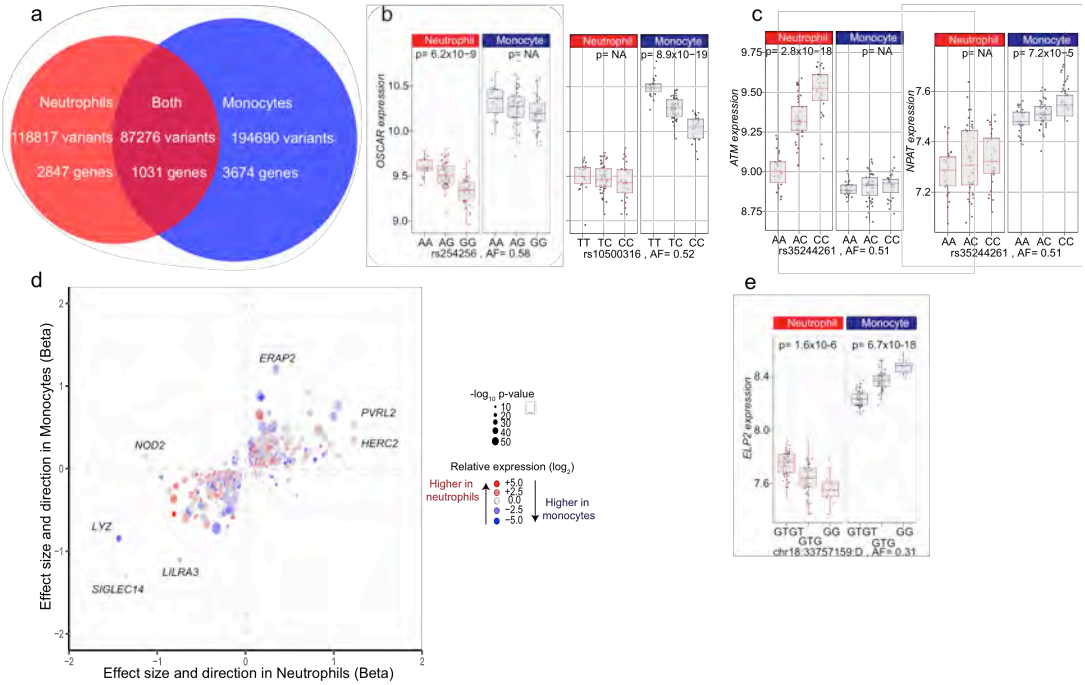


Figure 29: Patterns of shared and cell-type specific *cis* eQTL in neutrophils and monocytes. (a) Venn diagrams showing the number of variants and genes with cis-sQTL by cell type. Examples of cell-type constraint include (b) independent variants associated with expression of the same gene, *OSCAR*, specific to either neutrophils or monocytes (lead eSNPs rs254256 and rs1050031 respectively, r^2 <0.02). Conversely, (c) a single variant may be associated with different genes in each cell type as shown for rs35244261 (associated with elevated expression of *ATM* in neutrophils, and *NPAT* in monocytes). (d) Amongst 1939 genes that have an eQTL in both cell types, 1031 involve the same variant and are plotted with the effect

size in neutrophils and monocytes (shown on x- and y-axes respectively). Color denotes the relative expression in neutrophils and monocytes, and size denotes the minimum p value of an eQTL for that gene. Several eQTL show divergent direction of effects on the gene they regulate including *NOD2*, *THBD*, *TSTD1*, *TOM1L2* and *ELP2* (latter shown in e). Genes with large effect sizes are highlighted, including *HERC2*, the gene responsible for iris colour, which has amongst the most significant eQTL of all genes in neutrophils. P values greater than 0.05 are denoted as p=NA.

Higher expression of a gene in neutrophils compared with monocytes was a positive predictor of whether an eQTL was present in neutrophils or not, but normalised probe intensity explains just ~2% of variance and there are many genes in which an eQTL is observed in only one of neutrophils or monocytes despite similar levels of gene expression. The effect size of regulatory variants in both monocytes and neutrophils was larger for eQTL seen in both cell types compared with those seen only in one cell type ($p=7.4x10^{-196}$, Figure 30) confirming previous observations(Fairfax, Makino et al. 2012).

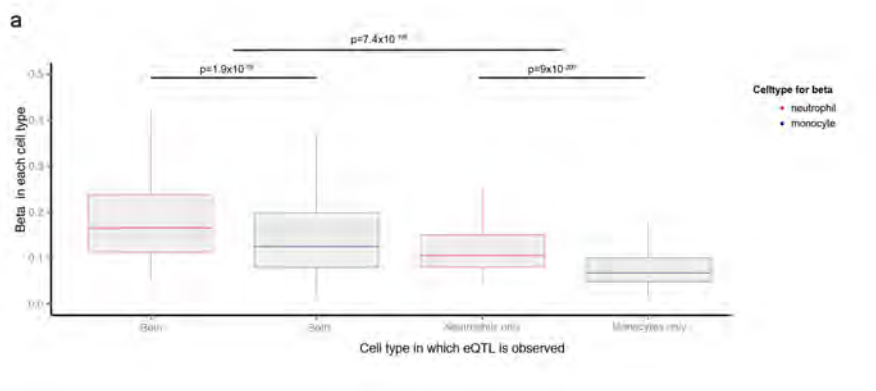


Figure 30: Effect sizes of eQTL. These tend to be larger for eQTL that are observed in both cell types (labelled as both) or those seen in one cell type only and are larger in neutrophils than monocytes regardless of whether the eQTL is observed in both cell types or one only.

Interestingly, whereas the number of unique eQTL is greater in monocytes, eQTL effect sizes were larger in neutrophils than in monocytes

regardless of whether the eQTL was seen in both cell types ($p=1.9x10^{-19}$) or just one ($p=9x10^{-200}$). This likely reflects a proportionally greater impact of genetic variation in governing gene expression in neutrophils that recapitulates their shorter lifespan and reduced exposure to environmental modifiers.

We highlight three patterns of cell environment modifying genetic effects on gene expression in addition to apparent cell-type constraint of regulatory activity. First, although 1939 genes have at least one eQTL in both cell types (and 1031 share an eQTL), for 908 genes the peak regulatory variants are different and for 840 (93%) of these, independent ($r^2<0.2$) demonstrating that the same gene may have independent regulatory variants of varying strengths and directions in different cell types. An example for the *OSCAR* gene is shown in Figure 29b. Conversely a particular variant may show pleiotropy in the gene it regulates in neutrophils and monocytes. About 4.5% (12661/278200) of variants that regulate a gene in one cell-type are associated to a different gene in the other cell type. For example, rs35244261 is associated with levels of *ATM* in neutrophils and *NPAT* in monocytes (Figure 29c). In gene-dense regions this could be due to cell type conditioning the effect of one or more variants that have regulatory effects on nearby genes. The third pattern we highlight is pleiotropy in direction of effect of a variant in different cell types on the same gene. We identified 2823 variants with significant but directionally opposing effects on 66 genes in both cell types (Figure 29d). Several are notable for the variant being associated with a disease directly or through a linked variant such as rs8066560 (*TOM1L2*, Parkinsons disease), chr18:3375159:D (Figure 29e, *ELP2*, oesophageal

squamous cell carcinoma), rs74058715 (*PADI4*, rheumatoid arthritis), and rs1981760 (*NOD2*, leprosy and CD), the latter two described in detail below. Collectively, the comparison of monocyte and neutrophil eQTL supports a model of widespread interaction between cellular milieu and genetic factors in regulation of gene expression even amongst cells of similar (myeloid) lineage. Moreover, delineation of cell-type restriction of eQTL in neutrophils may provide a tool to understand their involvement in variant-phenotype association.

Epigenetic and functional basis for eQTL in neutrophils and monocytes

The local genomic environment of a regulatory variant is cell type dependent and this may modify variant effect. Moreover, as has been shown for complex traits, leveraging epigenetic mark information may help fine-map genotype-phenotype associations(Trynka, Sandor et al. 2013). Therefore, to elucidate the role DNA methylation and histone modification play in genetic effects on gene expression, we compared eQTL maps in neutrophils and monocytes to maps of DNA methylation and several histone modifications generated by whole-genome bisulfite sequencing or chromatin-immunoprecipitation and sequencing (ChIP-Seq) respectively in neutrophils and monocytes from 4-8 individuals by the BLUEPRINT consortium(Adams, Altucci et al. 2012). As may be predicted by their shared ontology, monocytes and neutrophils have substantial overlap in regions of the genome that are methylated or subject to histone modifications, although this differs by the specific combination of epigenetic mark (Supplementary Table 14). Relative to

all imputed variants within 1MB of a gene (4812340 variants tested for being a cis eQTL), we observed marked and significant enrichment of peak *cis* eQTL in hypomethylated regions (p_{neut}=2.14x10^{-107}, p_{mono}= 2.52x10^{-124}) and depletion in hypermethylated regions (p_{neut}=6.36x10^{-46}, $p_{mono=}$7.55x10^{-23}) in the concordant cell type i.e. neutrophil eQTL in neutrophil hypomethylated regions (Figure 31a,31c and Supplementary Table 14). Similarly, as shown in Figure 31b (and Supplementary Table 14), peak *cis* eQTL were enriched in regions with histone marks associated with promoter activity (H3K4me3; p_{neut}= 8.23x10^{-231}, p_{mono}= 1.31x10^{-165}), active or poised enhancers (H3K27Ac; p_{neut}= 4.85x10^{-230}, p_{mono}= 1.34x10^{-80}, H3K4me1; $p_{neut=}$2.23x10^{-308}, p_{mono}= 4.84x10^{-229}) or activation (H3K36me3; p_{neut}= 2.01x10^{-238}, p_{mono}= 3.22x10^{-196}) and depleted in regions associated with repressive histone modifications (H3K27me3; p_{neut}= 3.78x10$_{-3}$, p_{mono}= 2.59x10^{-2}and H3K9me3; p_{neut}= 3.35x10^{-4}, p_{mono}= 2.57x10^{-2}). These data are replicated by an orthogonal approach, in which, despite vastly fewer regions, we observe 14-fold enrichment of neutrophil eQTL in neutrophil enhancer regions (p= 3.03x10^{-15}) and 6-fold enrichment of monocyte eQTL in monocyte enhancer regions (p=3.58x10^{-17}) using enhancer maps generated by CAGE-Seq in FANTOM5(Andersson, Gebhard et al. 2014, Consortium, the et al. 2014). These data support a model of epigenetic environment modifying regulatory effects in different cell types. An example for the *PADI2* gene is shown in Figure 31d-e.

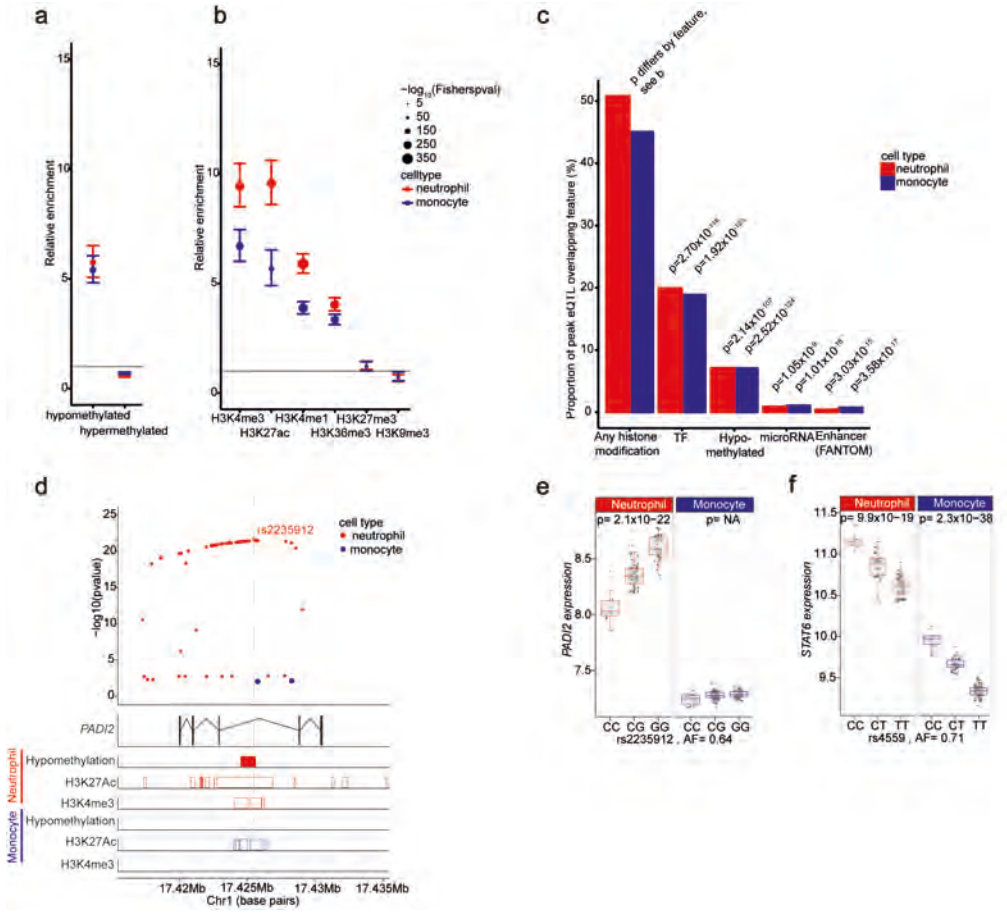


Figure 31: Epigenetic and functional basis for shared and cell-type specific eQTL in neutrophils and monocytes. We integrated eQTL data with whole-genome bisulfite sequencing and chromatin-immunoprecipitation sequencing data from primary neutrophils and monocytes from 4-8 individuals from the BLUEPRINT Consortium. (a) Enrichment of *cis* eQTL in neutrophils and monocytes, in hypo- and hyper-methylated regions of the genome in concordant cell types (neutrophil and monocyte respectively) relative to all SNP's tested for eQTL. (b) Enrichment of neutrophil and monocyte *cis* eQTL relative to all variants tested for colocalisation in regions of the genome associated with modified histones based on ChIP-Seq in concordant cell types: H3K4me3-histone 3 lysine 4 tri-methylation, H3K27Ac-histone 3 lysine 27 acetylation, H3K4me1-histone 3 lysine 4 mono-methylation, H3K36me3-histone 3 lysine 36 tri-methylation, H3K27me3-histone 3 lysine 27 tri-methylation or H3K9me3-histone 3 lysine 9 tri-methylation. (c) Proportion of *cis* eQTL in neutrophils (n=3774) and monocytes (n=4668) that overlie modified histones in the concordant cell type (BLUEPRINT), transcription factor binding sites in cell lines (ENCODE), sites that are hypomethylated in the concordant cell type (BLUEPRINT), target sites of conserved miRNAs with high mirSVR scores (microRNA.org) or CAGE-seq defined enhancer sites in neutrophils and monocytes (FANTOM5). Note that an eQTL may overlap more than one feature. (d-e) An example of a cell type specific eQTL that may be a result of methylation differences in cell types is tagged by rs2235912 that lies within an intron of *PADI2* in a region that is hypomethylated in

neutrophils but not monocytes. In addition, this locus is marked by the activating histone marks H3K27Ac and H3K4me3 in neutrophils but in monocytes has a smaller region marked by H3K27Ac only. As shown in (e), the gene is more highly expressed in neutrophils than monocytes, perhaps due to the greater activating histone marks in neutrophils. The G allele creates an additional putative CpG site which may explain the allele being associated with higher *PADI2* expression relative to the C allele. (f) STAT6 is an example of a gene with an eQTL in both monocytes and neutrophils where the eQTL lies in a predicted binding region for mir-18b, a microRNA. Tails in a and b show 95% CI of the enrichment estimate. P values greater than 0.05 are denoted as p=NA.

Mechanisms by which regulatory variants operate include alteration in transcription factor binding. Neutrophil and monocyte peak eQTL are 3.1- and 2.87-fold enriched (2.70×10^{-118} and $1.92 \times 10^{-123,}$ Supplementary Table 15) respectively in regions bound by a transcription factor in cell lines studied in the ENCODE project(Boyle, Hong et al. 2012). Expression levels of 44% (415/943) of reported human transcription factors(Ravasi, Suzuki et al. 2010) that are expressed in both cell types differ by >0.5 $\log_2$ between neutrophils and monocytes (Supplementary Table 15), making it plausible that alteration in transcription factor binding sites in addition to differences in transcription factor abundance could lead to eQTL being observed in one cell type and not the other. About 1% (52 in neutrophils, 35 in monocytes) of peak eQTL lie within microRNA binding sites, a 3.44 ($p=1.05 \times 10^{-9}$)- and 4.14-fold (1.01×10^{-16}) enrichment for neutrophils and monocyte respectively (Supplementary Table 16). For example, rs4559 is a common variant associated with *STAT6* expression (Figure 31f) in both cell types and is within the binding region of miR-18b. We note that although this analysis does not take into account whether the microRNA is expressed in the cell-type, the results are consistent

with some eQTL having their effect through microRNA-directed transcriptional gene silencing(Kim, Saetrom et al. 2008).

Neutrophil eQTL are enriched for GWAS disease-associated variants

eQTL mapping can provide functional insight into the basis for genetic association of complex traits. We systematically examined eQTL in neutrophils for association with complex traits. Of all eQTL in neutrophils 7.5% (14,390/191,767) are associated with one or more of 327 traits (of 1138 unique disease/traits curated in the NHGRI GWAS catalog(Welter, MacArthur et al. 2014)) directly or through linkage (r^2>0.8) with the trait-associated variant (Supplementary Table 17). This represents a 2.7-fold (95% CI 2.67-2.77, $p<2.2x10^{-16}$) enrichment relative to all SNPs originally tested for association with gene expression (eQTL). Notable examples are highlighted in Figure 32. Reciprocally, 7.8% (14390/183124) of trait-associated variants are, or tag, eQTL in neutrophils.

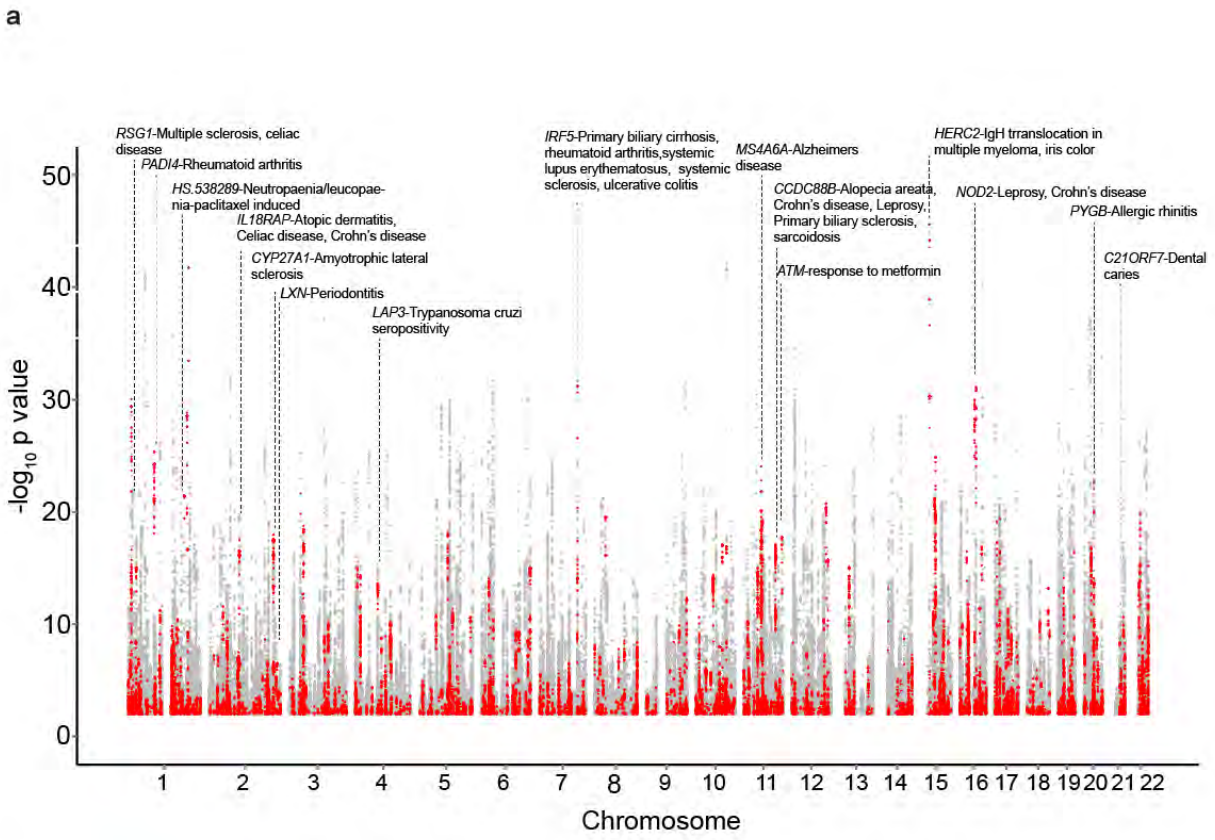


Figure 32: eQTL in neutrophils and their association with complex disease or trait. (a) Manhattan plot showing eQTL in neutrophils highlighting those associated with or in linkage disequilibrium (r^2>0.8) with a disease/trait associated variant listed in the NHGRI GWAS catalog. Each point denotes a single eQTL and is coloured red if the locus is associated with a complex trait or grey if not.

Collation of GWAS traits into disease categories demonstrates that enrichment is observed for gastrointestinal disorders (7.2-fold enrichment; $p<2x10^{-308}$), allergy (6.3 fold enrichment; $p=1.3x10^{-117}$), autoimmune (6.3-fold enrichment, $p<2x10^{-308}$) and infectious diseases (3.1-fold enrichment, $p=7.9x10^{-137}$) grouped together or disaggregated for viral (3.7-fold enrichment, $p=1.8x10^{-56}$), parasitic (5.1-fold enrichment, $p=6.4x10^{-96}$) and bacterial (3.1-fold enrichment, $p=2x10^{-34}$) diseases (Figure 33a, Supplementary table 18).

We note that many of the eQTL observed in neutrophils may also be observed in monocytes (Figure 33b) reinforcing the need for additional follow-up to resolve complexity of how a variant may predispose to disease through effects in one or more cell types. We therefore explored two eQTL in greater detail to demonstrate how integrated analysis can provide novel insights into disease.

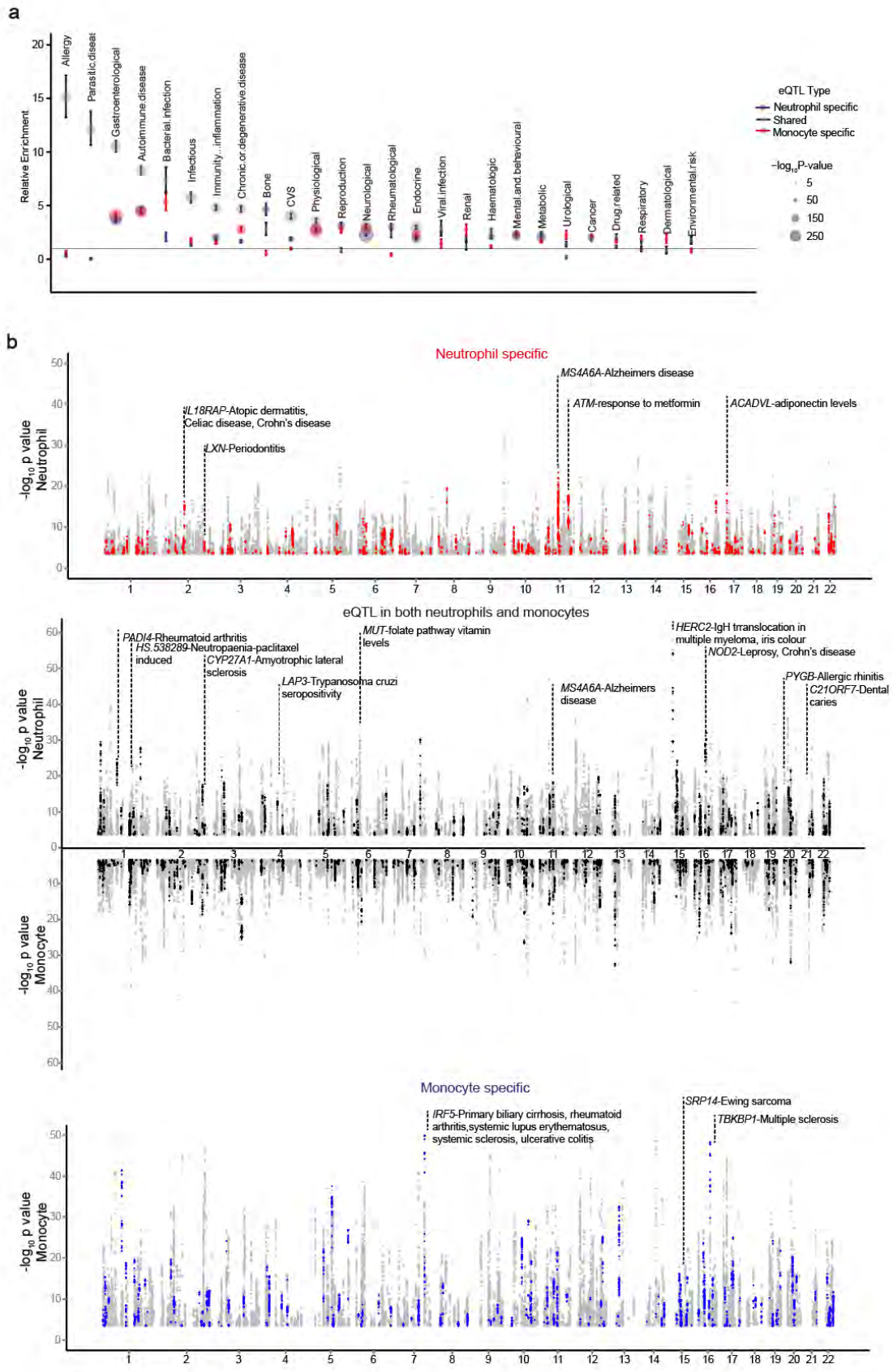


Figure 33: Shared and cell type specific cis eQTL are enriched for variants associated with diseases or traits in GWAS studies. (a) Enrichment of shared and cell type specific eQTL

for variants reported in the GWAS catalog as associated with a trait or disease by GWAS ontology category. For each category we compared the proportion of eQTL variants associated with a trait in the relevant category to the proportion of all variants tested that had this property by a Fisher's exact test. (b) Manhattan plots demonstrating significant (all FDR<0.05) shared (black), neutrophil-specific (red) or monocyte-specific (blue) eQTL that are either associated with a trait in the GWAS catalog (color) or not (grey). An arbitrary selection of diseases for which the risk variant (or a variant in linkage disequilibrium r^2>0.8) is also an eQTL are highlighted.

Integration of neutrophil and monocyte eQTL and epigenetic data with GWAS data implicates PADI4 expression in neutrophils in rheumatoid arthritis susceptibility

Rheumatoid arthritis (RA) is a systemic autoimmune disease afflicting ~0.5-1% of adults in which autoimmune destruction of synovial joints occurs(Scott, Wolfe et al. 2010). A major feature of disease is presence of autoantibodies directed against citrullinated proteins. We found that 13 of 101 loci (a 3-fold enrichment compared with background, 95%CI 1.5-5.4,$p=8x10^{-4}$) recently reported to be associated with RA risk in the largest meta-analysis to date(Okada, Wu et al. 2014) are eQTL in neutrophils. These include rs2240335, an eQTL for *PADI4* that is in near complete linkage disequilibrium with rs230188 (r^2=0.93 in our dataset). The A allele at rs2240335 is associated with elevated risk of RA and together with rs230188 is a genome-wide significant correlate of RA risk(Okada, Wu et al. 2014). Intriguingly,

rs2240335-A is associated with increased expression of *PADI4* in neutrophils but reduced expression in monocytes (Figure 34). rs74058715-T, a nearby SNP independent to rs2240335 (r^2=0.02), is associated with reduced *PADI4* expression in both cell types providing an additional instrument to probe the role of *PADI4* in neutrophils in RA risk (Figure 34). Examination of histone modifications in the region show that rs2240335 lies in a region marked by histone 3 lysine 27 acetylation (H3K27Ac) and histone 3 lysine 4 mono-methylation (H3K4me1) in neutrophils but not in monocytes whereas rs74058715 is in a region marked in both cell types (Figure 34), consistent with the cell type in which the eQTL is seen and suggesting rs2240335 as the functional variant as opposed to rs230188. Conditioning on rs2240335 does not reveal a secondary eQTL peak. Following from this prediction, rs74058715-T is nominally associated with reduced RA risk (p=0.05 in the largest meta-analysis of RA) showing directional consistency. We note that rs74058715 tags at least six other SNPs (LD>0.8) that are nominally associated with RA, and therefore may not itself be the causal SNP. Subsequent to oxidative responses by stimuli including rheumatoid factor, PADI4 post-translationally citrullinates histones and other proteins initiating NETosis, a process shown to be central to RA pathogenesis(Khandpur, Carmona-Rivera et al. 2013, Sohn, Onuma et al. 2014, Wright, Moots et al. 2014). Because PADI4 expression is confined to neutrophils and monocytes(Suzuki, Yamada et al. 2003), these data support a model in which rs230188 tags rs2240335-A, which alters PADI4 expression and RA risk likely through its actions in neutrophils.

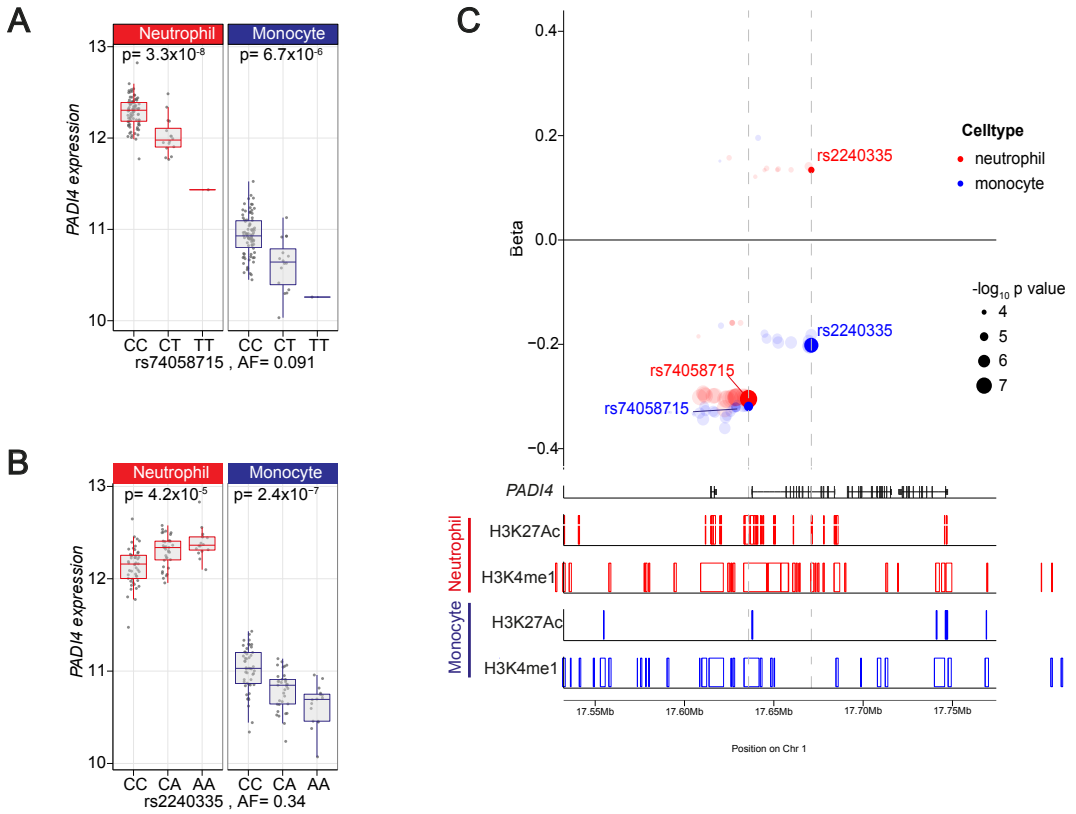


Figure 34: An example in *PADI4* demonstrates how integrated analysis of an eQTL informs understanding of rheumatoid arthritis (RA) risk. Expression level of *PADI4* is associated with two independent variants, (A) rs2240335 and (B) rs74058715 (r^2=0.02). rs2240335-A, the derived allele, is associated with elevated expression of *PADI4* in neutrophils and diminished levels in monocytes. rs7405871-T is associated with reduced *PADI4* expression in neutrophils and monocytes. rs2240335 is in near complete linkage with rs2301888, and this locus is associated with RA risk(Okada, Wu et al. 2014) (C) Compiled plot showing effect size estimate (beta) for variants associated with *PADI4* expression in neutrophils and monocytes (upper panel), relative to genic structures (track two), and BLUEPRINT ChIP-Seq reads for two histone marks in neutrophils (tracks three and four) and monocytes (tracks 5 and 6) for H3K27Ac and H3K4me1 demonstrating that rs2240335 lies in a region marked by both H3K27Ac and H3K4me1, a marker of active enhancers, in neutrophils but not in monocytes.

rs1981760, a SNP with pleiotropic association with leprosy and CD susceptibility, alters neutrophil inflammatory responses to NOD2 ligands through altered STAT3 binding and consequent NOD2 expression.

Leprosy, a disease caused by the infectious agent *Mycobacterium leprae*, and CD, an autoimmune inflammatory bowel disease, show profound overlap in genetic architecture(Jostins, Ripke et al. 2012). The ancestral T allele of rs1981760 is associated with increased susceptibility of leprosy, particularly multibacillary disease(Zhang, Huang et al. 2009) yet has weak protective effects in CD independent to the major CD-associated missense mutations(Jostins, Ripke et al. 2012). We observed a strong association between rs1981760-T and reduced expression of both *NOD2* ($p=8x10^{-30}$, variance explained=39%), and the adjacent gene *SNX20* ($p=7.8x10^{-10}$) in neutrophils and conversely elevated expression in monocytes ($p=3.3x10^{-10}$) (Figure 35a-b). The frequency of rs1981760-C is strikingly differentiated in Asians compared with other populations (78% in Asians vs. 26% in Europeans from the 1000G project) (Figure 35c). To further test the mechanism of this finding and functional consequences, we enrolled an independent cohort of 23 individuals. We were able to replicate the observed eQTL by quantitative PCR with two independent probe sets (Figure 35d). The rs1981760 polymorphism is reported to alter *STAT3* binding in some ENCODE cell-lines. We note that rs1981760 is in complete LD with rs9302752, the SNP first reported to be associated with leprosy and CD, but since rs1981760 is the primary eQTL signal, and in view of it potentially modifying transcription factor binding, we pursued rs1981760. Conditioning on rs1981760 does not reveal a secondary eQTL peak. To test whether the polymorphism alters STAT3 binding in neutrophils we performed chromatin immunoprecipitation with an antibody directed against STAT3 in 4 heterozygotes. Using allele-specific probes in a digital droplet PCR reaction

designed to accurately quantify the number of each allele, we observed significantly fewer copies of immunoprecipitated DNA containing T-alleles than C alleles after ChIP in each individual (Figure 35e and Supplementary Table 19) consistent with the hypothesis that the T allele reduces STAT3 binding either alone or potentially in a complex with other transcription factors. Moreover, *STAT3* is markedly differentially expressed in neutrophils and monocytes (Figure 35f). Finally, we stimulated neutrophils from individuals with the NOD2 ligand muramyl-dipeptide (supplemented with Pam3-CSK4 as a synergistic agonist). Individuals with the CC genotype expressed significantly greater levels of mRNA for *IFNB (*p=0.02) (Figure 35g) after stimulation. These data demonstrate the rs1981760 affects *NOD2* expression and subsequent IFNB responses to its ligand. Notably, eQTL in neutrophils are enriched for genes in the IFNB network ($p=5.9x10^{-4}$). Therefore these data suggest that type-1 interferons and neutrophils may be involved in leprosy as has been shown in tuberculosis(Mayer-Barber, Andrade et al. 2014).

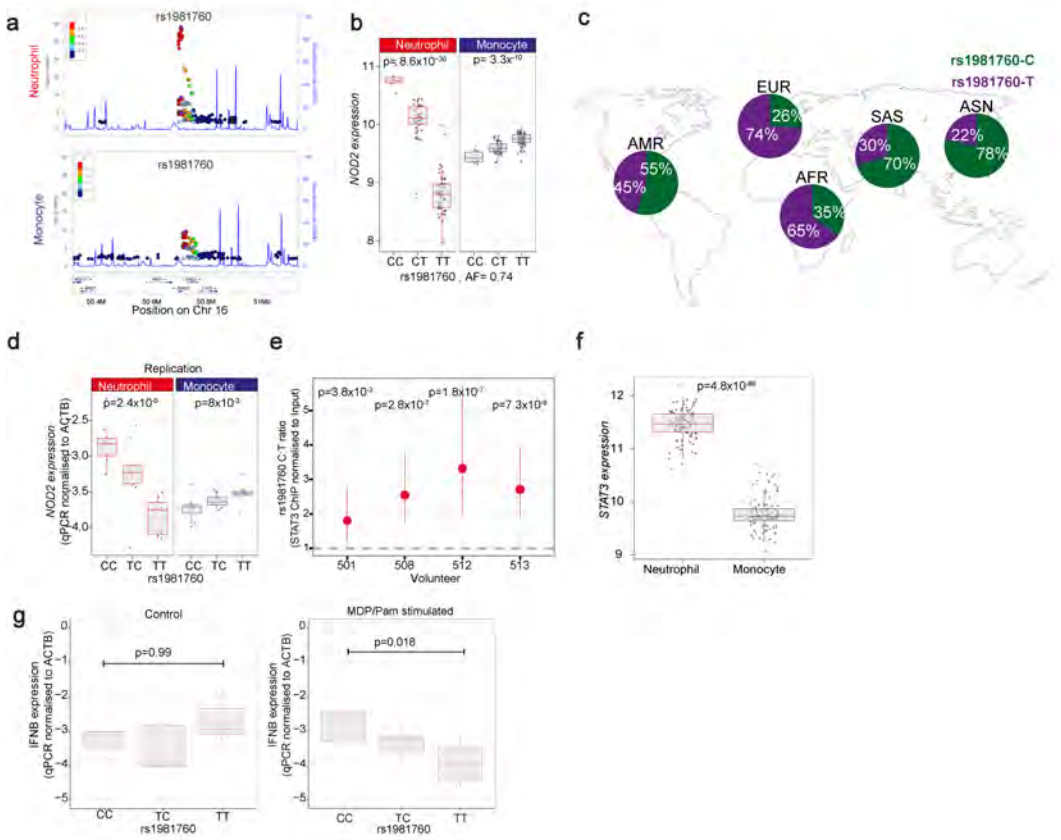


Figure 35: eQTL in NOD2 and their involvement in Leprosy and CD. (a) Regional association plots for rs1981760 and *NOD2* in Neutrophils (top) and monocytes (bottom). (b) The C (derived) allele at rs1981760, a variant located in an intergenic region between *NOD2* and *SNX20*, is associated with elevated *NOD2* expression in neutrophils and reduced *NOD2* expression in monocytes. (c) rs1981760 is remarkable for the profound differentiation this allele appears to have undergone in Asian populations with the derived C allele in higher frequency than the ancestral T allele in Asians according to data from the 1000 genomes project(Genomes Project, Abecasis et al. 2012). rs1981760 and rs9302752 are in high linkage disequilibrium (r^2=0.98 in our cohort), and the latter is associated with leprosy risk in GWAS reports from Chinese populations. The corresponding risk allele for leprosy at rs1981760 is C(Zhang, Huang et al. 2009). The C variant is also reported to have a mild protective effect in CD independent of the deleterious 3020insC frameshift mutation in *NOD2(Jostins, Ripke et al. 2012)*. (d) We replicated the observation of an eQTL in an independent set of 23 donors by quantitative PCR of *NOD2* with two different primer sets performed blinded to genotype status. Data shown represent one primer pair. (e) The rs1981760 SNP is located within a STAT3 transcription factor ChIP-seq binding site in MCF10A-Er-Src, an epithelial cell line (ENCODE). We investigated whether the variant may modulate allele-specific STAT3 binding in neutrophils by ChIP for four rs1981760

heterozygote healthy donors (screened to not also have the 3020insC frameshift mutation that ablates NOD2 expression). The frequency of C vs. T alleles after STAT3 ChIP normalised to the input is shown for allele-specific probes in a digital droplet PCR assay with the C allele significantly more frequently immunoprecipitated with STAT3 than the T-allele. p values for a chi-squared test comparing expected to actual ratio of positive and negative droplets for each donor are marked on the plot. (f) *STAT3* expression levels differ between neutrophils and monocytes as measured by gene expression array. The difference in expression by cell type is replicated by two additional probes targeting *STAT3*. (g) After stimulation with muramyldipeptide, the ligand for NOD2 and Pam3CSK4 (a co-stimulant) neutrophils from rs1981760-CC homozygous individuals had significantly higher levels of *IFNB* mRNA than TT homozygotes, with heterozygotes showing intermediate levels (left panel). No such association is apparent in mock-treated control neutrophils (right panel).

eQTL in neutrophils are enriched for regions subject to selection

Neutrophils are an evolutionarily ancient cell. In light of the rs1981760 observation, we systematically examined whether regulatory variants in neutrophils overlap regions that are identified as being subject to selection in systematic human surveys(Grossman, Andersen et al. 2013). Indeed, 1527 variants that are eQTL for a total of 39 genes lie in regions that have been subject to selection in humans (Supplementary Table 20), a 2.4-fold (95%CI 2.33-2.6) enrichment ($p=1.04x10^{-195}$) relative to the 20464 variants tested in our study that lie in annotated regions of selection. Many of the genes showing eQTL driven by variants in such regions are known to be involved in leukocyte biology, for example the chemokine receptors *CCR1(Lionakis, Fischer et al. 2012)* and *CCR3*, or *GUSB*, a gene in which mutations cause mucopolysaccharidosis VII(OMIM) in which recurrent respiratory infections are common. Others, such as *HERC2* involved in modifying eye colour(Sturm, Duffy et al. 2008) and for which the second strongest eQTL in neutrophils is observed, may suggest new hypotheses for operative selective forces and the

cells to yield the patterns observed. Therefore, regulatory variants in neutrophils are likely to have been subject to selection, reinforcing their biological relevance.

In conclusion, these data illustrate the utility of eQTL mapping in different leucocyte subsets and integration with epigenetic and disease-association data to reveal mechanisms of disease.

5.3 Conclusions

Genetic regulation of gene expression is central to explaining the effect of genetic variation on diseases. Here, in a careful study of monocytes and neutrophils I found that quantitative regulation of gene expression is a frequent across many genes in both cell types but is highly nuanced dependent on cell type, DNA methylation and local histone modification. Moreover, this exercise yielded many surprising insights into disease biology particularly rheumatoid arthritis, leprosy and Crohn's disease. Whilst we did not find any eQTL that are associated with TB-disease based on previous studies, this large resource allows functional annotation of TB associated loci in further studies. Therefore this chapter, in addition to chapter four will form an integral part of future studies of the genetic correlates of TB disease.

Chapter Six. Discussion and conclusions

6.1 Overall conclusions

The overall aim of this work was to identify aspects of the genetic basis of immune-mediated susceptibility to infectious disease. Collectively the studies presented demonstrate that understanding the genetic basis of immune mechanisms of disease yields fundamental new insights into disease pathogenesis. Specifically, the following outcomes and conclusions are apparent:

1. The monocyte:lymphocyte ratio is a novel correlate of TB susceptibility in diverse populations of at-risk individuals and is a correlate of malaria disease and outcomes for malignancies.
2. Functional and transcriptomic profile of monocytes differ according to monocyte:lymphocyte ratio.

    The data from chapter two and three collectively support a model in which differences in myeloid:lymphoid ratios act as a marker of monocyte ontogeny from myeloid or lymphoid biased haematopoietic stem cells.
3. The ML and NL ratios are co-heritable and both pleiotropic and non-pleiotropic genetic loci regulate both.
4. Finally, the genetic regulators of gene expression in monocytes and neutrophils operate in nuances ways conditioned by cell type and by epigenetic environment to give rise to their involvement in a wide range of infectious and non-infectious immune-mediated diseases and as targets of evolutionary selection.

6.2 Limitations

The series of studies presented here have several noteworthy limitations, some of which give guidance for future research.

Firstly, the studies presented in support of the ML ratio involvement in disease susceptibility in chapter two are limited to African populations. Although one other study in an Asian population has replicated these results(Wang, Yin et al. 2015), further studies in other populations would lend greater credence to the generalizability of these findings.

Secondly, a causal role for the ML ratio in TB disease susceptibility is plausible but is made less likely by the relatively weak effect size of the association and the apparent lack of specificity in that the ML ratio is associated with a range of other outcomes. Studies of whether the ML ratio is merely a marker of TB latency would assist, as would direct comparison of the relative prognostic value of the ML ratio alongside a positive Tuberculin skin test or IGRA.

An alternative method of establishing whether the ML ratio is causally associated with TB susceptibility as a consequence of it marking monocyte ontogeny, is the use animal models in which the immune system can be reconstituted from myeloid- or lymphoid-biased HSC's and the animal challenged with *M.TB*. Although early efforts in this regard were attempted during the conduct of these studies, the absence of definitive results highlights the third limitation: a lack of experimental evidence for myeloid-biased HSC's

giving rise to functionally distinct progeny. Even if this was to be achieved, it may yet be an imperfect indicator of human physiology.

Mendelian randomization studies that test whether genetic correlates of the ML ratio, used as instrumental variables, are associated with TB susceptibility are likely the most powerful method to demonstrate a causal association. However, delineation of non-pleoiotropic genetic correlates of the ML ratio is required. Consequently, the fourth limitation is that the genetic correlates of the ML and NL ratios that are defined here do not take advantage of more advanced methods for identifying pleiotropy. Future studies that explicitly model pleiotropy, for example using reverse regression and covariance-matrix meta-analysis as has been recently been applied to studies of body-mass index (Locke, Kahali et al. 2015) are indicated. This is currently underway.

Finally, the studies of genomic regulators of gene expression in monocytes and neutrophils is limited by sample size and by absence of comprehensive profiling of other cell types in order to confidently define what aspects of the regulatory architecture we define, is unique to neutrophils. Efforts to combine these results with those from another study of neutrophils in Han Chinese individuals, and with studies of B-cells, monocytes, macrophages and Natural Killer cells are underway.

The major strengths of this work are the use of comprehensive profiling methods of transcriptomic and genetic assessment, the integration of functional and bioinformatics analyses and the generation of resources that

may prove useful to understanding of other infectious and non-infectious diseases.

6.3 Perspectives and future research

The studies described in this dissertation explore the genomic basis of immune-mediated susceptibility to infectious disease. Using classical epidemiologic, genetic and transcriptomic studies largely in the context of mycobacterial disease and myeloid cells, I defined a new intermediate phenotype for TB, identified its possible pathophysiologic basis and delineated the genetic regulators for the phenotype. I also established a library of variants associated with gene expression in neutrophils and monocytes and show how this helps refine understanding of immune-mediated diseases.

The ML ratio offers utility as a biomarker of risk to identify individuals at highest risk for TB, and thus potentially the most likely to benefit from prophylactic intervention. It also suggests new mechanisms for TB susceptibility, which if manipulated may lead to new preventive or treatment strategies. Further studies to identify whether the ML ratio and TB disease risk are causally associated, and to identify nodes for intervention are required. The ability to perform a Mendelian randomization study critically depends on the presence of a well-powered genome-wide association study of TB. In pursuit of this, during the course of these studies we established the International Tuberculosis Host Genetics Consortium, including data from 14 cohorts (13 countries) and >30 000 cases and healthy controls. These studies will contribute to defining the shared and unshared genetic correlates of TB.

By interpolation with data in this thesis we will perform a Mendelian randomization study to establish whether variants that are associated with the ML ratio are also associated with TB risk to the extent predicted by the association between the variant and ML ratio, and ML ratio and TB.

Similarly, the library of genetic variants that alter expression of genes in monocytes and neutrophils form a resource for understanding how genetic variation affects disease susceptibility through its effect on immunity. Although we explored the causal pathways of relevance to leprosy, as an exemplar, this approach may be useful for functional annotation of TB or other disease associated variants.

Chapter Seven. References

(1930). "THE BACILLUS-CALMETTE-GUERIN TRAGEDY IN LÜBECK " Annals of Internal Medicine **4**(2): 190-192.

(1932). "BCG AND LUBECK." The Lancet **353**: 353-354.

Abdool Karim, Q., S. S. Abdool Karim, J. A. Frohlich, A. C. Grobler, C. Baxter, L. E. Mansoor, A. B. Kharsany, S. Sibeko, K. P. Mlisana, Z. Omar, T. N. Gengiah, S. Maarschalk, N. Arulappan, M. Mlotshwa, L. Morris and D. Taylor (2010). "Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women." Science **329**(5996): 1168-1174.

Abdool Karim, Q., A. B. Kharsany, J. A. Frohlich, L. Werner, M. Mlotshwa, B. T. Madlala and S. S. Abdool Karim (2012). "HIV Incidence in Young Girls in

KwaZulu-Natal, South Africa-Public Health Imperative for Their Inclusion in HIV Biomedical Intervention Trials." AIDS and behavior.

Abdool Karim, S. S., K. Naidoo, A. Grobler, N. Padayatchi, C. Baxter, A. Gray, T. Gengiah, G. Nair, S. Bamber, A. Singh, M. Khan, J. Pienaar, W. El-Sadr, G. Friedland and Q. Abdool Karim (2010). "Timing of initiation of antiretroviral drugs during tuberculosis therapy." The New England journal of medicine **362**(8): 697-706.

Abdool Karim, S. S., K. Naidoo, A. Grobler, N. Padayatchi, C. Baxter, A. L. Gray, T. Gengiah, S. Gengiah, A. Naidoo, N. Jithoo, G. Nair, W. M. El-Sadr, G. Friedland and Q. Abdool Karim (2011). "Integration of antiretroviral therapy with tuberculosis treatment." The New England journal of medicine **365**(16): 1492-1501.

Adams, D., L. Altucci, S. E. Antonarakis, J. Ballesteros, S. Beck, A. Bird, C. Bock, B. Boehm, E. Campo, A. Caricasole, F. Dahl, E. T. Dermitzakis, T. Enver, M. Esteller, X. Estivill, A. Ferguson-Smith, J. Fitzgibbon, P. Flicek, C. Giehl, T. Graf, F. Grosveld, R. Guigo, I. Gut, K. Helin, J. Jarvius, R. Kuppers, H. Lehrach, T. Lengauer, A. Lernmark, D. Leslie, M. Loeffler, E. Macintyre, A. Mai, J. H. Martens, S. Minucci, W. H. Ouwehand, P. G. Pelicci, H. Pendeville, B. Porse, V. Rakyan, W. Reik, M. Schrappe, D. Schubeler, M. Seifert, R. Siebert, D. Simmons, N. Soranzo, S. Spicuglia, M. Stratton, H. G. Stunnenberg, A. Tanay, D. Torrents, A. Valencia, E. Vellenga, M. Vingron, J. Walter and S. Willcocks (2012). "BLUEPRINT to decode the epigenetic signature written in blood." Nat Biotechnol **30**(3): 224-226.

Akolo, C., I. Adetifa, S. Shepperd and J. Volmink (2010). "Treatment of latent tuberculosis infection in HIV infected persons." Cochrane database of systematic reviews(1): CD000171.

Ambruso, D. R., C. Knall, A. N. Abell, J. Panepinto, A. Kurkchubasche, G. Thurman, C. Gonzalez-Aller, A. Hiester, M. deBoer, R. J. Harbeck, R. Oyer, G. L. Johnson and D.

Roos (2000). "Human neutrophil immunodeficiency syndrome is associated with an inhibitory Rac2 mutation." Proc Natl Acad Sci U S A **97**(9): 4654-4659.

Amulic, B., C. Cazalet, G. L. Hayes, K. D. Metzler and A. Zychlinsky (2012). "Neutrophil function: from mechanisms to disease." Annu Rev Immunol **30**: 459-489.

Andersen, P. and T. M. Doherty (2005). "The success and failure of BCG - implications for a novel tuberculosis vaccine." Nature reviews. Microbiology **3**(8): 656-662.

Andersson, R., C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidl, T. Suzuki, E. Ntini, E. Arner, E. Valen, K. Li, L. Schwarzfischer, D. Glatz, J. Raithel, B. Lilje, N. Rapin, F. O. Bagger, M. Jorgensen, P. R. Andersen, N. Bertin, O. Rackham, A. M. Burroughs, J. K. Baillie, Y. Ishizu, Y. Shimizu, E. Furuhata, S. Maeda, Y. Negishi, C. J. Mungall, T. F. Meehan, T. Lassmann, M. Itoh, H. Kawaji, N. Kondo, J. Kawai, A. Lennartsson, C. O. Daub, P. Heutink, D. A. Hume, T. H. Jensen, H. Suzuki, Y. Hayashizaki, F. Muller, F. Consortium, A. R. Forrest, P. Carninci, M. Rehli and A. Sandelin (2014). "An atlas of active enhancers across human cell types and tissues." Nature **507**(7493): 455-461.

Arnoldussen, D. L., M. Linehan and A. Sheikh (2011). "BCG vaccination and allergy: a systematic review and meta-analysis." J Allergy Clin Immunol **127**(1): 246-253, 253 e241-221.

Baldridge, M. T., K. Y. King, N. C. Boles, D. C. Weksberg and M. A. Goodell (2010). "Quiescent haematopoietic stem cells are activated by IFN-gamma in response to chronic infection." Nature **465**(7299): 793-797.

Bardoel, B. W., E. F. Kenny, G. Sollberger and A. Zychlinsky (2014). "The balancing act of neutrophils." Cell Host Microbe **15**(5): 526-536.

Barreiro, L. B., L. Tailleux, A. A. Pai, B. Gicquel, J. C. Marioni and Y. Gilad (2012). "Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection." Proc Natl Acad Sci U S A **109**(4): 1204-1209.

Barreiro, L. B., L. Tailleux, A. A. Pai, B. Gicquel, J. C. Marioni and Y. Gilad (2012). "Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection." Proceedings of the National Academy of Sciences of the United States of America **109**(4): 1204-1209.

Barry, C. E., 3rd, H. I. Boshoff, V. Dartois, T. Dick, S. Ehrt, J. Flynn, D. Schnappinger, R. J. Wilkinson and D. Young (2009). "The spectrum of latent tuberculosis: rethinking the biology and intervention strategies." Nature reviews. Microbiology **7**(12): 845-855.

Battle, A., S. Mostafavi, X. Zhu, J. B. Potash, M. M. Weissman, C. McCormick, C. D. Haudenschild, K. B. Beckman, J. Shi, R. Mei, A. E. Urban, S. B. Montgomery, D. F. Levinson and D. Koller (2014). "Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals." Genome Res **24**(1): 14-24.

Bennouna, S., S. K. Bliss, T. J. Curiel and E. Y. Denkers (2003). "Cross-talk in the innate immune system: neutrophils instruct recruitment and activation of dendritic cells during microbial infection." J Immunol **171**(11): 6052-6058.

Berrington, A., J. Green and R. Newton (2000). "Vitamin D deficiency and tuberculosis." Lancet **356**(9223): 74; author reply 74-75.

Berry, M. P., C. M. Graham, F. W. McNab, Z. Xu, S. A. Bloch, T. Oni, K. A. Wilkinson, R. Banchereau, J. Skinner, R. J. Wilkinson, C. Quinn, D. Blankenship, R. Dhawan, J. J.

Cush, A. Mejias, O. Ramilo, O. M. Kon, V. Pascual, J. Banchereau, D. Chaussabel and A. O'Garra (2010). "An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis." Nature **466**(7309): 973-977.

Betel, D., A. Koppal, P. Agius, C. Sander and C. Leslie (2010). "Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites." Genome Biol **11**(8): R90.

Bogunovic, D., M. Byun, L. A. Durfee, A. Abhyankar, O. Sanal, D. Mansouri, S. Salem, I. Radovanovic, A. V. Grant, P. Adimi, N. Mansouri, S. Okada, V. L. Bryant, X. F. Kong, A. Kreins, M. M. Velez, B. Boisson, S. Khalilzadeh, U. Ozcelik, I. A. Darazam, J. W. Schoggins, C. M. Rice, S. Al-Muhsen, M. Behr, G. Vogt, A. Puel, J. Bustamante, P. Gros, J. M. Huibregtse, L. Abel, S. Boisson-Dupuis and J. L. Casanova (2012). "Mycobacterial disease and impaired IFN-gamma immunity in humans with inherited ISG15 deficiency." Science **337**(6102): 1684-1688.

Boker, S., M. Neale, H. Maes, M. Wilde, M. Spiegel, T. Brick, J. Spies, R. Estabrook, S. Kenny, T. Bates, P. Mehta and J. Fox (2011). "OpenMx: An Open Source Extended Structural Equation Modeling Framework." Psychometrika **76**(2): 306-317.

Borregaard, N., O. E. Sorensen and K. Theilgaard-Monch (2007). "Neutrophil granules: a library of innate immunity proteins." Trends Immunol **28**(8): 340-345.

Boyle, A. P., E. L. Hong, M. Hariharan, Y. Cheng, M. A. Schaub, M. Kasowski, K. J. Karczewski, J. Park, B. C. Hitz, S. Weng, J. M. Cherry and M. Snyder (2012). "Annotation of functional variation in personal genomes using RegulomeDB." Genome Res **22**(9): 1790-1797.

Bucher, H. C., L. E. Griffith, G. H. Guyatt, P. Sudre, M. Naef, P. Sendi and M. Battegay (1999). "Isoniazid prophylaxis for tuberculosis in HIV infection: a meta-analysis of randomized controlled trials." AIDS **13**(4): 501-507.

Bunyan, T. (1680). The Life and Death of Mr. Badman; Presented to the World in a Familiar Dialogue Between Mr. Wiseman and Mr. Attentive

Cambier, C. J., S. Falkow and L. Ramakrishnan (2014). "Host evasion and exploitation schemes of Mycobacterium tuberculosis." Cell **159**(7): 1497-1509.

Carpenter, E., L. Fray and E. Gormley (1997). "Cellular responses and Mycobacterium bovis BCG growth inhibition by bovine lymphocytes." Immunology and cell biology **75**(6): 554-560.

Casadevall, A. and L. A. Pirofski (2003). "The damage-response framework of microbial pathogenesis." Nature reviews. Microbiology **1**(1): 17-24.

Casanova, J. L. and L. Abel (2004). "The human model: a genetic dissection of immunity to infection in natural conditions." Nature reviews. Immunology **4**(1): 55-66.

Churchyard, G. J., K. L. Fielding and A. D. Grant (2014). "A trial of mass isoniazid preventive therapy for tuberculosis control." N Engl J Med **370**(17): 1662-1663.

Cobat, A., C. J. Gallant, L. Simkin, G. F. Black, K. Stanley, J. Hughes, T. M. Doherty, W. A. Hanekom, B. Eley, N. Beyers, J. P. Jais, P. van Helden, L. Abel, E. G. Hoal, A. Alcais and E. Schurr (2010). "High heritability of antimycobacterial immunity in an area of hyperendemicity for tuberculosis disease." The Journal of infectious diseases **201**(1): 15-19.

Cohen, T., M. Murray, K. Wallengren, G. G. Alvarez, E. Y. Samuel and D. Wilson (2010). "The prevalence and drug sensitivity of tuberculosis among patients

dying in hospital in KwaZulu-Natal, South Africa: a postmortem study." PLoS medicine **7**(6): e1000296.

Collins, S. J. (2002). "The role of retinoids and retinoic acid receptors in normal hematopoiesis." Leukemia **16**(10): 1896-1905.

Comas, I., J. Chakravartti, P. M. Small, J. Galagan, S. Niemann, K. Kremer, J. D. Ernst and S. Gagneux (2010). "Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved." Nature genetics **42**(6): 498-503.

Comas, I., M. Coscolla, T. Luo, S. Borrell, K. E. Holt, M. Kato-Maeda, J. Parkhill, B. Malla, S. Berg, G. Thwaites, D. Yeboah-Manu, G. Bothamley, J. Mei, L. Wei, S. Bentley, S. R. Harris, S. Niemann, R. Diel, A. Aseffa, Q. Gao, D. Young and S. Gagneux (2013). "Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans." Nat Genet **45**(10): 1176-1182.

Comstock, G. W. (1978). "Tuberculosis in twins: a re-analysis of the Prophit survey." The American review of respiratory disease **117**(4): 621-624.

Consortium, F., R. P. the, Clst, A. R. Forrest, H. Kawaji, M. Rehli, J. K. Baillie, M. J. de Hoon, T. Lassmann, M. Itoh, K. M. Summers, H. Suzuki, C. O. Daub, J. Kawai, P. Heutink, W. Hide, T. C. Freeman, B. Lenhard, V. B. Bajic, M. S. Taylor, V. J. Makeev, A. Sandelin, D. A. Hume, P. Carninci and Y. Hayashizaki (2014). "A promoter-level mammalian expression atlas." Nature **507**(7493): 462-470.

Coovadia, H. M., E. R. Brown, M. G. Fowler, T. Chipato, D. Moodley, K. Manji, P. Musoke, L. Stranix-Chibanda, V. Chetty, W. Fawzi, C. Nakabiito, L. Msweli, R. Kisenge, L. Guay, A. Mwatha, D. J. Lynn, S. H. Eshleman, P. Richardson, K. George, P. Andrew, L. M. Mofenson, S. Zwerski, Y. Maldonado and H. p. team (2012). "Efficacy and safety of an extended nevirapine regimen in infant children of

breastfeeding mothers with HIV-1 infection for prevention of postnatal HIV-1 transmission (HPTN 046): a randomised, double-blind, placebo-controlled trial." Lancet **379**(9812): 221-228.

Cottle, L. E. (2011). "Mendelian susceptibility to mycobacterial disease." Clinical genetics **79**(1): 17-22.

Cox, J. A., R. L. Lukande, S. Lucas, A. M. Nelson, E. Van Marck and R. Colebunders (2010). "Autopsy causes of death in HIV-positive individuals in sub-Saharan Africa and correlation with clinical diagnoses." AIDS reviews **12**(4): 183-194.

Crosslin, D. R., A. McDavid, N. Weston, S. C. Nelson, X. Zheng, E. Hart, M. de Andrade, I. J. Kullo, C. A. McCarty, K. F. Doheny, E. Pugh, A. Kho, M. G. Hayes, S. Pretel, A. Saip, M. D. Ritchie, D. C. Crawford, P. K. Crane, K. Newton, R. Li, D. B. Mirel, A. Crenshaw, E. B. Larson, C. S. Carlson, G. P. Jarvik, R. Electronic Medical and N. Genomics (2012). "Genetic variants associated with the white blood cell count in 13,923 subjects in the eMERGE Network." Hum Genet **131**(4): 639-652.

Cunningham, R. S., F. R. Sabin, S. Sugiyama and J. A. Kindwall (1925). "The Role of the Monocyte in Tuberculosis." Bulletin of the Johns Hopkins Hospital **XXXVII**(4): 231-280.

Danesh, J., R. Collins, P. Appleby and R. Peto (1998). "Association of fibrinogen, C-reactive protein, albumin, or leukocyte count with coronary heart disease: meta-analyses of prospective studies." JAMA **279**(18): 1477-1482.

Daniel, T. M. (2009). The history of tuberculosis: past, present and challenges for the future. Tuberculosis: a comprehensive clinical reference. H. S. Schaaf and A. Zumla. Philadelphia, PA, Saunders.

Das, B., S. S. Kashino, I. Pulu, D. Kalita, V. Swami, H. Yeger, D. W. Felsher and A. Campos-Neto (2013). "CD271+ Bone Marrow Mesenchymal Stem Cells May

Provide a Niche for Dormant Mycobacterium tuberculosis." Science Translational Medicine **5**(170): 170ra113-170ra113.

Delaneau, O., J. Marchini and J. F. Zagury (2012). "A linear complexity phasing method for thousands of genomes." Nat Methods **9**(2): 179-181.

Dembele, M., N. Saleri, A. C. Carvalho, T. Saouadogo, A. D. Hien, I. Zabsonre, S. T. Koala, J. Simpore and A. Matteelli (2010). "Incidence of tuberculosis after HAART initiation in a cohort of HIV-positive patients in Burkina Faso." The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease **14**(3): 318-323.

Denis, M., D. N. Wedlock and B. M. Buddle (2004). "Ability of T cell subsets and their soluble mediators to modulate the replication of Mycobacterium bovis in bovine macrophages." Cellular immunology **232**(1-2): 1-8.

Dharmawardhane, S., D. Brownson, M. Lennartz and G. M. Bokoch (1999). "Localization of p21-activated kinase 1 (PAK1) to pseudopodia, membrane ruffles, and phagocytic cups in activated human neutrophils." J Leukoc Biol **66**(3): 521-527.

Diana, J., Y. Simoni, L. Furio, L. Beaudoin, B. Agerberth, F. Barrat and A. Lehuen (2013). "Crosstalk between neutrophils, B-1a cells and plasmacytoid dendritic cells initiates autoimmune diabetes." Nat Med **19**(1): 65-73.

Diehl, K. and O. F. von Verschner (1933). " Zwillingsforschung und erbliche Tuberkulosedisposition, Zwillingstuberkulose: Zwillingsforschung und erbliche Tuberkulosedisposition." **1**.

Dimas, A. S., S. Deutsch, B. E. Stranger, S. B. Montgomery, C. Borel, H. Attar-Cohen, C. Ingle, C. Beazley, M. Gutierrez Arcelus, M. Sekowska, M. Gagnebin, J. Nisbett, P. Deloukas, E. T. Dermitzakis and S. E. Antonarakis (2009). "Common regulatory

variation impacts gene expression in a cell type-dependent manner." Science **325**(5945): 1246-1250.

Doan, C. A. and F. R. Sabin (1930). "The Relation of the Tubercle and the Monocyte:Lymphocte ratio to resistance and susceptibility in Tuberculosis." J Exp Med **52**: 113-152.

Dye, C., S. Scheele, P. Dolin, V. Pathania and M. C. Raviglione (1999). "Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project." JAMA : the journal of the American Medical Association **282**(7): 677-686.

Ehrlich, P. (1880). "Methodologische Beiträge zur Physiologie und Pathologie der verschiedenen Formen der Leukocyten." Zeitschrift fuer klinische Medizin **1**: 553-560.

El-Zaatari, M., Y. M. Chang, M. Zhang, M. Franz, A. Shreiner, A. J. McDermott, K. F. van der Sluijs, R. Lutter, H. Grasberger, N. Kamada, V. B. Young, G. B. Huffnagle and J. Y. Kao (2014). "Tryptophan catabolism restricts IFN-gamma-expressing neutrophils and Clostridium difficile immunopathology." J Immunol **193**(2): 807-816.

Ema, H., Y. Morita and T. Suda (2014). "Heterogeneity and hierarchy of hematopoietic stem cells." Exp Hematol **42**(2): 74-82 e72.

Fairfax, B. P., P. Humburg, S. Makino, V. Naranbhai, D. Wong, E. Lau, L. Jostins, K. Plant, R. Andrews, C. McGee and J. C. Knight (2014). "Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression." Science **343**(6175): 1246949.

Fairfax, B. P., S. Makino, J. Radhakrishnan, K. Plant, S. Leslie, A. Dilthey, P. Ellis, C. Langford, F. O. Vannberg and J. C. Knight (2012). "Genetics of gene expression in

primary immune cells identifies cell type-specific master regulators and roles of HLA alleles." Nat Genet **44**(5): 502-510.

Falconer, D. S. and T. F. C. MacKay (1996). Introduction to Quantitative Genetics, . Longmans Green, Harlow, Essex, UK., Longman.

Fehrmann, R. S., R. C. Jansen, J. H. Veldink, H. J. Westra, D. Arends, M. J. Bonder, J. Fu, P. Deelen, H. J. Groen, A. Smolonska, R. K. Weersma, R. M. Hofstra, W. A. Buurman, S. Rensen, M. G. Wolfs, M. Platteel, A. Zhernakova, C. C. Elbers, E. M. Festen, G. Trynka, M. H. Hofker, C. G. Saris, R. A. Ophoff, L. H. van den Berg, D. A. van Heel, C. Wijmenga, G. J. Te Meerman and L. Franke (2011). "Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA." PLoS Genet **7**(8): e1002197.

Ferraro, A., A. M. D'Alise, T. Raj, N. Asinovski, R. Phillips, A. Ergun, J. M. Replogle, A. Bernier, L. Laffel, B. E. Stranger, P. L. De Jager, D. Mathis and C. Benoist (2014). "Interindividual variation in human T regulatory cells." Proc Natl Acad Sci U S A **111**(12): E1111-1120.

Filliol, I., A. S. Motiwala, M. Cavatore, W. Qi, M. H. Hazbon, M. Bobadilla del Valle, J. Fyfe, L. Garcia-Garcia, N. Rastogi, C. Sola, T. Zozio, M. I. Guerrero, C. I. Leon, J. Crabtree, S. Angiuoli, K. D. Eisenach, R. Durmaz, M. L. Joloba, A. Rendon, J. Sifuentes-Osornio, A. Ponce de Leon, M. D. Cave, R. Fleischmann, T. S. Whittam and D. Alland (2006). "Global phylogeny of Mycobacterium tuberculosis based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set." Journal of bacteriology **188**(2): 759-772.

Ford, C. B., P. L. Lin, M. R. Chase, R. R. Shah, O. Iartchouk, J. Galagan, N. Mohaideen, T. R. Ioerger, J. C. Sacchettini, M. Lipsitch, J. L. Flynn and S. M. Fortune (2011). "Use of whole genome sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection." Nat Genet **43**(5): 482-486.

Fowler, M. G., H. Coovadia, C. M. Herron, Y. Maldonado, T. Chipato, D. Moodley, P. Musoke, J. Aizire, K. Manji, L. Stranix-Chibanda, W. Fawzi, V. Chetty, L. Msweli, R. Kisenge, E. Brown, A. Mwatha, S. H. Eshleman, P. Richardson, M. Allen, K. George, P. Andrew, S. Zwerski, L. M. Mofenson, J. B. Jackson and H. P. T. for the (2013). "Efficacy and safety of an extended nevirapine regimen in infants of breastfeeding mothers with HIV-1 infection for prevention of HIV-1 transmission (HPTN 046): 18-month results of a randomized, double-blind, placebo-controlled trial." J Acquir Immune Defic Syndr.

Friedman, G. D., A. L. Klatsky and A. B. Siegelaub (1974). "The leukocyte count as a predictor of myocardial infarction." N Engl J Med **290**(23): 1275-1278.

Gagneux, S. and P. M. Small (2007). "Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development." Lancet Infect Dis **7**(5): 328-337.

Ganna, A. and E. Ingelsson (2015). "5 year mortality predictors in 498 103 UK Biobank participants: a prospective population-based study." Lancet.

Gekas, C. and T. Graf (2013). "CD41 expression marks myeloid-biased adult hematopoietic stem cells and increases with age." Blood **121**(22): 4463-4472.

Genomes Project, C., G. R. Abecasis, A. Auton, L. D. Brooks, M. A. DePristo, R. M. Durbin, R. E. Handsaker, H. M. Kang, G. T. Marth and G. A. McVean (2012). "An integrated map of genetic variation from 1,092 human genomes." Nature **491**(7422): 56-65.

Gibbs, J. R., M. P. van der Brug, D. G. Hernandez, B. J. Traynor, M. A. Nalls, S. L. Lai, S. Arepalli, A. Dillman, I. P. Rafferty, J. Troncoso, R. Johnson, H. R. Zielke, L. Ferrucci, D. L. Longo, M. R. Cookson and A. B. Singleton (2010). "Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain." PLoS Genet **6**(5): e1000952.

Gideon, H. P. and J. L. Flynn (2011). "Latent tuberculosis: what the host "sees"?" Immunologic research **50**(2-3): 202-212.

Gill, W. P., N. S. Harik, M. R. Whiddon, R. P. Liao, J. E. Mittler and D. R. Sherman (2009). "A replication clock for Mycobacterium tuberculosis." Nat Med **15**(2): 211-214.

Govender, T. and H. Coovadia (2014). "Eliminating mother to child transmission of HIV-1 and keeping mothers alive: recent progress." J Infect **68 Suppl 1**: S57-62.

Grimm, R. H., Jr., J. D. Neaton and W. Ludwig (1985). "Prognostic importance of the white blood cell count for coronary, cancer, and all-cause mortality." JAMA **254**(14): 1932-1937.

Grisham, M. B., T. D. Engerson, J. M. McCord and H. P. Jones (1985). "A comparative study of neutrophil purification and function." J Immunol Methods **82**(2): 315-320.

Grossman, S. R., K. G. Andersen, I. Shlyakhter, S. Tabrizi, S. Winnicki, A. Yen, D. J. Park, D. Griesemer, E. K. Karlsson, S. H. Wong, M. Cabili, R. A. Adegbola, R. N. Bamezai, A. V. Hill, F. O. Vannberg, J. L. Rinn, P. Genomes, E. S. Lander, S. F. Schaffner and P. C. Sabeti (2013). "Identifying recent adaptations in large-scale genomic data." Cell **152**(4): 703-713.

Guay, L. A., P. Musoke, T. Fleming, D. Bagenda, M. Allen, C. Nakabiito, J. Sherman, P. Bakaki, C. Ducar, M. Deseyve, L. Emel, M. Mirochnick, M. G. Fowler, L. Mofenson, P. Miotti, K. Dransfield, D. Bray, F. Mmiro and J. B. Jackson (1999). "Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: HIVNET 012 randomised trial." Lancet **354**(9181): 795-802.

Guthrie, G. J., K. A. Charles, C. S. Roxburgh, P. G. Horgan, D. C. McMillan and S. J. Clarke (2013). "The systemic inflammation-based neutrophil-lymphocyte ratio: experience in patients with cancer." Crit Rev Oncol Hematol **88**(1): 218-230.

Hill, A. B. (1965). "The Environment and Disease: Association or Causation?" Proc R Soc Med **58**: 295-300.

Holloway, K. L., R. J. Henneberg, M. de Barros Lopes and M. Henneberg (2011). "Evolution of human tuberculosis: a systematic review and meta-analysis of paleopathological evidence." Homo : internationale Zeitschrift fur die vergleichende Forschung am Menschen **62**(6): 402-458.

Howie, B., J. Marchini and M. Stephens (2011). "Genotype imputation with thousands of genomes." G3 (Bethesda) **1**(6): 457-470.

Howie, B. N., P. Donnelly and J. Marchini (2009). "A flexible and accurate genotype imputation method for the next generation of genome-wide association studies." PLoS Genet **5**(6): e1000529.

HPA (2007). Tuberculosis — Respiratory and Non-respiratory Notifications, England and Wales, 1913-2005.

Hutterer, G. C., C. Stoeckigt, T. Stojakovic, J. Jesche, K. Eberhard, K. Pummer, R. Zigeuner and M. Pichler (2014). "Low preoperative lymphocyte-monocyte ratio

(LMR) represents a potentially poor prognostic factor in nonmetastatic clear cell renal cell carcinoma." Urol Oncol.

Joosten, S. A., H. A. Fletcher and T. H. Ottenhoff (2013). "A helicopter perspective on TB biomarkers: pathway and process based analysis of gene expression data provides new insight into TB pathogenesis." PLoS One **8**(9): e73230.

Jostins, L., S. Ripke, R. K. Weersma, R. H. Duerr, D. P. McGovern, K. Y. Hui, J. C. Lee, L. P. Schumm, Y. Sharma, C. A. Anderson, J. Essers, M. Mitrovic, K. Ning, I. Cleynen, E. Theatre, S. L. Spain, S. Raychaudhuri, P. Goyette, Z. Wei, C. Abraham, J. P. Achkar, T. Ahmad, L. Amininejad, A. N. Ananthakrishnan, V. Andersen, J. M. Andrews, L. Baidoo, T. Balschun, P. A. Bampton, A. Bitton, G. Boucher, S. Brand, C. Buning, A. Cohain, S. Cichon, M. D'Amato, D. De Jong, K. L. Devaney, M. Dubinsky, C. Edwards, D. Ellinghaus, L. R. Ferguson, D. Franchimont, K. Fransen, R. Gearry, M. Georges, C. Gieger, J. Glas, T. Haritunians, A. Hart, C. Hawkey, M. Hedl, X. Hu, T. H. Karlsen, L. Kupcinskas, S. Kugathasan, A. Latiano, D. Laukens, I. C. Lawrance, C. W. Lees, E. Louis, G. Mahy, J. Mansfield, A. R. Morgan, C. Mowat, W. Newman, O. Palmieri, C. Y. Ponsioen, U. Potocnik, N. J. Prescott, M. Regueiro, J. I. Rotter, R. K. Russell, J. D. Sanderson, M. Sans, J. Satsangi, S. Schreiber, L. A. Simms, J. Sventoraityte, S. R. Targan, K. D. Taylor, M. Tremelling, H. W. Verspaget, M. De Vos, C. Wijmenga, D. C. Wilson, J. Winkelmann, R. J. Xavier, S. Zeissig, B. Zhang, C. K. Zhang, H. Zhao, I. B. D. G. C. International, M. S. Silverberg, V. Annese, H. Hakonarson, S. R. Brant, G. Radford-Smith, C. G. Mathew, J. D. Rioux, E. E. Schadt, M. J. Daly, A. Franke, M. Parkes, S. Vermeire, J. C. Barrett and J. H. Cho (2012). "Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease." Nature **491**(7422): 119-124.

Jostins, L., S. Ripke, R. K. Weersma, R. H. Duerr, D. P. McGovern, K. Y. Hui, J. C. Lee, L. P. Schumm, Y. Sharma, C. A. Anderson, J. Essers, M. Mitrovic, K. Ning, I. Cleynen, E. Theatre, S. L. Spain, S. Raychaudhuri, P. Goyette, Z. Wei, C. Abraham, J. P. Achkar, T. Ahmad, L. Amininejad, A. N. Ananthakrishnan, V. Andersen, J. M. Andrews, L. Baidoo, T. Balschun, P. A. Bampton, A. Bitton, G. Boucher, S. Brand, C. Buning, A. Cohain, S. Cichon, M. D'Amato, D. De Jong, K. L. Devaney, M. Dubinsky, C. Edwards, D. Ellinghaus, L. R. Ferguson, D. Franchimont, K. Fransen, R. Gearry, M. Georges, C. Gieger, J. Glas, T. Haritunians, A. Hart, C. Hawkey, M. Hedl, X. Hu, T. H. Karlsen, L. Kupcinskas, S. Kugathasan, A. Latiano, D. Laukens, I. C. Lawrance, C. W. Lees, E. Louis, G. Mahy, J. Mansfield, A. R. Morgan, C. Mowat, W. Newman, O. Palmieri, C. Y. Ponsioen, U. Potocnik, N. J. Prescott, M. Regueiro, J. I. Rotter, R. K. Russell, J. D. Sanderson, M. Sans, J. Satsangi, S. Schreiber, L. A. Simms, J. Sventoraityte, S. R. Targan, K. D. Taylor, M. Tremelling, H. W. Verspaget, M. De Vos, C. Wijmenga, D. C. Wilson, J. Winkelmann, R. J. Xavier, S. Zeissig, B. Zhang, C. K. Zhang, H. Zhao, M. S. Silverberg, V. Annese, H. Hakonarson, S. R. Brant, G. Radford-Smith, C. G. Mathew, J. D. Rioux, E. E. Schadt, M. J. Daly, A. Franke, M. Parkes, S. Vermeire, J. C. Barrett and J. H. Cho (2012). "Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease." Nature **491**(7422): 119-124.

Kagina, B. M., B. Abel, T. J. Scriba, E. J. Hughes, A. Keyser, A. Soares, H. Gamieldien, M. Sidibana, M. Hatherill, S. Gelderbloem, H. Mahomed, A. Hawkridge, G. Hussey, G. Kaplan, W. A. Hanekom and I. other members of the South African Tuberculosis Vaccine (2010). "Specific T cell frequency and cytokine expression profile do not correlate with protection against tuberculosis after bacillus

Calmette-Guerin vaccination of newborns." Am J Respir Crit Care Med **182**(8): 1073-1079.

Kallman, F. J. and D. Reisner (1943). "TWIN STUDIES ON GENETIC VARIATIONS IN RESISTANCE TO TUBERCULOSIS." Journal of Heridity **34**(9): 269-276.

Karolchik, D., G. P. Barber, J. Casper, H. Clawson, M. S. Cline, M. Diekhans, T. R. Dreszer, P. A. Fujita, L. Guruvadoo, M. Haeussler, R. A. Harte, S. Heitner, A. S. Hinrichs, K. Learned, B. T. Lee, C. H. Li, B. J. Raney, B. Rhead, K. R. Rosenbloom, C. A. Sloan, M. L. Speir, A. S. Zweig, D. Haussler, R. M. Kuhn and W. J. Kent (2014). "The UCSC Genome Browser database: 2014 update." Nucleic Acids Res **42**(Database issue): D764-770.

Kato, K. (1935). "Effect of tuberculin test on monocyte-lymphocyte ratio in children." The Journal of Pediatrics **7**(2): 238-244.

Kaufmann, S. H., G. Hussey and P. H. Lambert (2010). "New vaccines for tuberculosis." Lancet **375**(9731): 2110-2119.

Khandpur, R., C. Carmona-Rivera, A. Vivekanandan-Giri, A. Gizinski, S. Yalavarthi, J. S. Knight, S. Friday, S. Li, R. M. Patel, V. Subramanian, P. Thompson, P. Chen, D. A. Fox, S. Pennathur and M. J. Kaplan (2013). "NETs are a source of citrullinated autoantigens and stimulate inflammatory responses in rheumatoid arthritis." Sci Transl Med **5**(178): 178ra140.

Kim, D. H., P. Saetrom, O. Snove, Jr. and J. J. Rossi (2008). "MicroRNA-directed transcriptional gene silencing in mammalian cells." Proc Natl Acad Sci U S A **105**(42): 16230-16235.

Kim, S., J. Becker, M. Bechheim, V. Kaiser, M. Noursadeghi, N. Fricker, E. Beier, S. Klaschik, P. Boor, T. Hess, A. Hofmann, S. Holdenrieder, J. R. Wendland, H. Frohlich, G. Hartmann, M. M. Nothen, B. Muller-Myhsok, B. Putz, V. Hornung and J.

Schumacher (2014). "Characterizing the genetic basis of innate immune response in TLR4-activated human monocytes." Nat Commun **5**: 5236.

Kimura, A., M. A. Rieger, J. M. Simone, W. Chen, M. C. Wickre, B. M. Zhu, P. S. Hoppe, J. J. O'Shea, T. Schroeder and L. Hennighausen (2009). "The transcription factors STAT5A/B regulate GM-CSF-mediated granulopoiesis." Blood **114**(21): 4721-4728.

Kleinnijenhuis, J., J. Quintin, F. Preijers, C. S. Benn, L. A. Joosten, C. Jacobs, J. van Loenhout, R. J. Xavier, P. Aaby, J. W. van der Meer, R. van Crevel and M. G. Netea (2014). "Long-lasting effects of BCG vaccination on both heterologous Th1/Th17 responses and innate trained immunity." J Innate Immun **6**(2): 152-158.

Kleinnijenhuis, J., J. Quintin, F. Preijers, L. A. Joosten, D. C. Ifrim, S. Saeed, C. Jacobs, J. van Loenhout, D. de Jong, H. G. Stunnenberg, R. J. Xavier, J. W. van der Meer, R. van Crevel and M. G. Netea (2012). "Bacille Calmette-Guerin induces NOD2-dependent nonspecific protection from reinfection via epigenetic reprogramming of monocytes." Proc Natl Acad Sci U S A **109**(43): 17537-17542.

Kleinnijenhuis, J., J. Quintin, F. Preijers, L. A. Joosten, C. Jacobs, R. J. Xavier, J. W. van der Meer, R. van Crevel and M. G. Netea (2014). "BCG-induced trained immunity in NK cells: Role for non-specific protection to infection." Clin Immunol **155**(2): 213-219.

Koch, R. (1882). Die Ätiologie der Tuberculose. Berlin.

Koh, Y. W., H. J. Kang, C. Park, D. H. Yoon, S. Kim, C. Suh, H. Go, J. E. Kim, C. W. Kim and J. Huh (2012). "The ratio of the absolute lymphocyte count to the absolute monocyte count is associated with prognosis in Hodgkin's lymphoma: correlation with tumor-associated macrophages." Oncologist **17**(6): 871-880.

Koh, Y. W., C. S. Park, D. H. Yoon, C. Suh and J. Huh (2014). "Should the cut-off values of the lymphocyte to monocyte ratio for prediction of prognosis in diffuse large B-cell lymphoma be changed in elderly patients?" Eur J Haematol.

Kolaczkowska, E. and P. Kubes (2013). "Neutrophil recruitment and function in health and inflammation." Nat Rev Immunol **13**(3): 159-175.

Kong, M. and C. Lee (2013). "Genetic associations with C-reactive protein level and white blood cell count in the KARE study." Int J Immunogenet **40**(2): 120-125.

Krishnan, N., W. Malaga, P. Constant, M. Caws, T. H. Tran, J. Salmons, T. N. Nguyen, D. B. Nguyen, M. Daffe, D. B. Young, B. D. Robertson, C. Guilhot and G. E. Thwaites (2011). "Mycobacterium tuberculosis lineage influences innate immune response and virulence and is associated with distinct cell envelope lipid profiles." PLoS One **6**(9): e23870.

Kumar, D., L. Nath, M. A. Kamal, A. Varshney, A. Jain, S. Singh and K. V. Rao (2010). "Genome-wide analysis of the host intracellular network that regulates survival of Mycobacterium tuberculosis." Cell **140**(5): 731-743.

Kumararatne, D. S. (2006). "Mendelian susceptibility to mycobacterial disease." Respiration; international review of thoracic diseases **73**(3): 280-282.

Kwon, A. T., D. J. Arenillas, R. Worsley Hunt and W. W. Wasserman (2012). "oPOSSUM-3: advanced analysis of regulatory motif over-representation across genes or ChIP-Seq datasets." G3 (Bethesda) **2**(9): 987-1002.

Lappalainen, T., M. Sammeth, M. R. Friedlander, P. A. t Hoen, J. Monlong, M. A. Rivas, M. Gonzalez-Porta, N. Kurbatova, T. Griebel, P. G. Ferreira, M. Barann, T. Wieland, L. Greger, M. van Iterson, J. Almlof, P. Ribeca, I. Pulyakhina, D. Esser, T. Giger, A. Tikhonov, M. Sultan, G. Bertier, D. G. MacArthur, M. Lek, E. Lizano, H. P.

Buermans, I. Padioleau, T. Schwarzmayr, O. Karlberg, H. Ongen, H. Kilpinen, S. Beltran, M. Gut, K. Kahlem, V. Amstislavskiy, O. Stegle, M. Pirinen, S. B. Montgomery, P. Donnelly, M. I. McCarthy, P. Flicek, T. M. Strom, C. Geuvadis, H. Lehrach, S. Schreiber, R. Sudbrak, A. Carracedo, S. E. Antonarakis, R. Hasler, A. C. Syvanen, G. J. van Ommen, A. Brazma, T. Meitinger, P. Rosenstiel, R. Guigo, I. G. Gut, X. Estivill and E. T. Dermitzakis (2013). "Transcriptome and genome sequencing uncovers functional variation in humans." Nature **501**(7468): 506-511.

Lawn, S. D., M. Badri and R. Wood (2005). "Tuberculosis among HIV-infected patients receiving HAART: long term incidence and risk factors in a South African cohort." AIDS **19**(18): 2109-2116.

Lawn, S. D., R. Wood, K. M. De Cock, K. Kranzer, J. J. Lewis and G. J. Churchyard (2010). "Antiretrovirals and isoniazid preventive therapy in the prevention of HIV-associated tuberculosis in settings with limited health-care resources." Lancet Infect Dis **10**(7): 489-498.

Lawn, S. D. and A. I. Zumla (2011). "Tuberculosis." Lancet **378**(9785): 57-72.

Lee, M. N., C. Ye, A. C. Villani, T. Raj, W. Li, T. M. Eisenhaure, S. H. Imboywa, P. I. Chipendo, F. A. Ran, K. Slowikowski, L. D. Ward, K. Raddassi, C. McCabe, M. H. Lee, I. Y. Frohlich, D. A. Hafler, M. Kellis, S. Raychaudhuri, F. Zhang, B. E. Stranger, C. O. Benoist, P. L. De Jager, A. Regev and N. Hacohen (2014). "Common genetic variants modulate pathogen-sensing responses in human dendritic cells." Science **343**(6175): 1246980.

Lee, S. H., J. Yang, M. E. Goddard, P. M. Visscher and N. R. Wray (2012). "Estimation of pleiotropy between complex diseases using single-nucleotide

polymorphism-derived genomic relationships and restricted maximum likelihood." Bioinformatics **28**(19): 2540-2542.

Leek, J. T. and J. D. Storey (2007). "Capturing heterogeneity in gene expression studies by surrogate variable analysis." PLoS Genet **3**(9): 1724-1735.

Leitsalu, L., T. Haller, T. Esko, M. L. Tammesoo, H. Alavere, H. Snieder, M. Perola, P. C. Ng, R. Magi, L. Milani, K. Fischer and A. Metspalu (2014). "Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu." Int J Epidemiol.

Lesho, E., F. J. Forestiero, M. H. Hirata, R. D. Hirata, L. Cecon, F. F. Melo, S. H. Paik, Y. Murata, E. W. Ferguson, Z. Wang and G. T. Ooi (2011). "Transcriptional responses of host peripheral blood cells to tuberculosis infection." Tuberculosis **91**(5): 390-399.

Li, J., R. Jiang, W. S. Liu, Q. Liu, M. Xu, Q. S. Feng, L. Z. Chen, J. X. Bei, M. Y. Chen and Y. X. Zeng (2013). "A large cohort study reveals the association of elevated peripheral blood lymphocyte-to-monocyte ratio with favorable prognosis in nasopharyngeal carcinoma." PLoS One **8**(12): e83069.

Li, Y. L., K. S. Gu, Y. Y. Pan, Y. Jiao and Z. M. Zhai (2014). "Peripheral blood lymphocyte/monocyte ratio at the time of first relapse predicts outcome for patients with relapsed or primary refractory diffuse large B-cell lymphoma." BMC Cancer **14**: 341.

Li, Y. L., Y. Y. Pan, Y. Jiao, J. Ning, Y. G. Fan and Z. M. Zhai (2014). "Peripheral blood lymphocyte/monocyte ratio predicts outcome for patients with diffuse large B cell lymphoma after standard first-line regimens." Ann Hematol **93**(4): 617-626.

Li, Z. M., J. J. Huang, Y. Xia, J. Sun, Y. Huang, Y. Wang, Y. J. Zhu, Y. J. Li, W. Zhao, W. X. Wei, T. Y. Lin, H. Q. Huang and W. Q. Jiang (2012). "Blood lymphocyte-to-

monocyte ratio identifies high-risk patients in diffuse large B-cell lymphoma treated with R-CHOP." PLoS One **7**(7): e41658.

Lin, G. N., J. W. Peng, D. Y. Liu, J. J. Xiao, Y. Q. Chen and X. Q. Chen (2014). "Increased lymphocyte to monocyte ratio is associated with better prognosis in patients with newly diagnosed metastatic nasopharyngeal carcinoma receiving chemotherapy." Tumour Biol.

Lin, G. N., J. W. Peng, J. J. Xiao, D. Y. Liu and Z. J. Xia (2014). "Prognostic impact of circulating monocytes and lymphocyte-to-monocyte ratio on previously untreated metastatic non-small cell lung cancer patients receiving platinum-based doublet." Med Oncol **31**(7): 70.

Lin, P. L., C. B. Ford, M. T. Coleman, A. J. Myers, R. Gawande, T. Ioerger, J. Sacchettini, S. M. Fortune and J. L. Flynn (2014). "Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing." Nat Med **20**(1): 75-79.

Lionakis, M. S., B. G. Fischer, J. K. Lim, M. Swamydas, W. Wan, C. C. Richard Lee, J. I. Cohen, P. Scheinberg, J. L. Gao and P. M. Murphy (2012). "Chemokine receptor Ccr1 drives neutrophil-mediated kidney immunopathology and mortality in invasive candidiasis." PLoS Pathog **8**(8): e1002865.

Liu, T. T., D. Wilson, H. Dawood, D. W. Cameron and G. G. Alvarez (2012). "Inaccuracy of death certificate diagnosis of tuberculosis and potential underdiagnosis of TB in a region of high HIV prevalence." Clinical & developmental immunology **2012**: 937013.

Locke, A. E., B. Kahali, S. I. Berndt, A. E. Justice, T. H. Pers, F. R. Day, C. Powell, S. Vedantam, M. L. Buchkovich, J. Yang, D. C. Croteau-Chonka, T. Esko, T. Fall, T. Ferreira, S. Gustafsson, Z. Kutalik, J. Luan, R. Magi, J. C. Randall, T. W. Winkler, A.

R. Wood, T. Workalemahu, J. D. Faul, J. A. Smith, J. Hua Zhao, W. Zhao, J. Chen, R. Fehrmann, A. K. Hedman, J. Karjalainen, E. M. Schmidt, D. Absher, N. Amin, D. Anderson, M. Beekman, J. L. Bolton, J. L. Bragg-Gresham, S. Buyske, A. Demirkan, G. Deng, G. B. Ehret, B. Feenstra, M. F. Feitosa, K. Fischer, A. Goel, J. Gong, A. U. Jackson, S. Kanoni, M. E. Kleber, K. Kristiansson, U. Lim, V. Lotay, M. Mangino, I. Mateo Leach, C. Medina-Gomez, S. E. Medland, M. A. Nalls, C. D. Palmer, D. Pasko, S. Pechlivanis, M. J. Peters, I. Prokopenko, D. Shungin, A. Stancakova, R. J. Strawbridge, Y. Ju Sung, T. Tanaka, A. Teumer, S. Trompet, S. W. van der Laan, J. van Setten, J. V. Van Vliet-Ostaptchouk, Z. Wang, L. Yengo, W. Zhang, A. Isaacs, E. Albrecht, J. Arnlov, G. M. Arscott, A. P. Attwood, S. Bandinelli, A. Barrett, I. N. Bas, C. Bellis, A. J. Bennett, C. Berne, R. Blagieva, M. Bluher, S. Bohringer, L. L. Bonnycastle, Y. Bottcher, H. A. Boyd, M. Bruinenberg, I. H. Caspersen, Y. D. Ida Chen, R. Clarke, E. W. Daw, A. J. de Craen, G. Delgado, M. Dimitriou, A. S. Doney, N. Eklund, K. Estrada, E. Eury, L. Folkersen, R. M. Fraser, M. E. Garcia, F. Geller, V. Giedraitis, B. Gigante, A. S. Go, A. Golay, A. H. Goodall, S. D. Gordon, M. Gorski, H. J. Grabe, H. Grallert, T. B. Grammer, J. Grassler, H. Gronberg, C. J. Groves, G. Gusto, J. Haessler, P. Hall, T. Haller, G. Hallmans, C. A. Hartman, M. Hassinen, C. Hayward, N. L. Heard-Costa, Q. Helmer, C. Hengstenberg, O. Holmen, J. J. Hottenga, A. L. James, J. M. Jeff, A. Johansson, J. Jolley, T. Juliusdottir, L. Kinnunen, W. Koenig, M. Koskenvuo, W. Kratzer, J. Laitinen, C. Lamina, K. Leander, N. R. Lee, P. Lichtner, L. Lind, J. Lindstrom, K. Sin Lo, S. Lobbens, R. Lorbeer, Y. Lu, F. Mach, P. K. Magnusson, A. Mahajan, W. L. McArdle, S. McLachlan, C. Menni, S. Merger, E. Mihailov, L. Milani, A. Moayyeri, K. L. Monda, M. A. Morken, A. Mulas, G. Muller, M. Muller-Nurasyid, A. W. Musk, R. Nagaraja, M. M. Nothen, I. M. Nolte, S. Pilz, N. W. Rayner, F. Renstrom, R. Rettig, J. S. Ried, S. Ripke, N. R. Robertson, L. M. Rose, S.

Sanna, H. Scharnagl, S. Scholtens, F. R. Schumacher, W. R. Scott, T. Seufferlein, J. Shi, A. Vernon Smith, J. Smolonska, A. V. Stanton, V. Steinthorsdottir, K. Stirrups, H. M. Stringham, J. Sundstrom, M. A. Swertz, A. J. Swift, A. C. Syvanen, S. T. Tan, B. O. Tayo, B. Thorand, G. Thorleifsson, J. P. Tyrer, H. W. Uh, L. Vandenput, F. C. Verhulst, S. H. Vermeulen, N. Verweij, J. M. Vonk, L. L. Waite, H. R. Warren, D. Waterworth, M. N. Weedon, L. R. Wilkens, C. Willenborg, T. Wilsgaard, M. K. Wojczynski, A. Wong, A. F. Wright, Q. Zhang, S. LifeLines Cohort, E. P. Brennan, M. Choi, Z. Dastani, A. W. Drong, P. Eriksson, A. Franco-Cereceda, J. R. Gadin, A. G. Gharavi, M. E. Goddard, R. E. Handsaker, J. Huang, F. Karpe, S. Kathiresan, S. Keildson, K. Kiryluk, M. Kubo, J. Y. Lee, L. Liang, R. P. Lifton, B. Ma, S. A. McCarroll, A. J. McKnight, J. L. Min, M. F. Moffatt, G. W. Montgomery, J. M. Murabito, G. Nicholson, D. R. Nyholt, Y. Okada, J. R. Perry, R. Dorajoo, E. Reinmaa, R. M. Salem, N. Sandholm, R. A. Scott, L. Stolk, A. Takahashi, T. Tanaka, F. M. Van't Hooft, A. A. Vinkhuyzen, H. J. Westra, W. Zheng, K. T. Zondervan, A. D. Consortium, A.-B. W. Group, C. A. D. Consortium, C. K. Consortium, Glgc, Icbp, M. Investigators, T. C. Mu, M. I. Consortium, P. Consortium, C. ReproGen, G. Consortium, C. International Endogene, A. C. Heath, D. Arveiler, S. J. Bakker, J. Beilby, R. N. Bergman, J. Blangero, P. Bovet, H. Campbell, M. J. Caulfield, G. Cesana, A. Chakravarti, D. I. Chasman, P. S. Chines, F. S. Collins, D. C. Crawford, L. A. Cupples, D. Cusi, J. Danesh, U. de Faire, H. M. den Ruijter, A. F. Dominiczak, R. Erbel, J. Erdmann, J. G. Eriksson, M. Farrall, S. B. Felix, E. Ferrannini, J. Ferrieres, I. Ford, N. G. Forouhi, T. Forrester, O. H. Franco, R. T. Gansevoort, P. V. Gejman, C. Gieger, O. Gottesman, V. Gudnason, U. Gyllensten, A. S. Hall, T. B. Harris, A. T. Hattersley, A. A. Hicks, L. A. Hindorff, A. D. Hingorani, A. Hofman, G. Homuth, G. K. Hovingh, S. E. Humphries, S. C. Hunt, E. Hypponen, T. Illig, K. B. Jacobs, M. R. Jarvelin, K. H. Jockel, B. Johansen, P.

Jousilahti, J. W. Jukema, A. M. Jula, J. Kaprio, J. J. Kastelein, S. M. Keinanen-Kiukaanniemi, L. A. Kiemeney, P. Knekt, J. S. Kooner, C. Kooperberg, P. Kovacs, A. T. Kraja, M. Kumari, J. Kuusisto, T. A. Lakka, C. Langenberg, L. Le Marchand, T. Lehtimaki, V. Lyssenko, S. Mannisto, A. Marette, T. C. Matise, C. A. McKenzie, B. McKnight, F. L. Moll, A. D. Morris, A. P. Morris, J. C. Murray, M. Nelis, C. Ohlsson, A. J. Oldehinkel, K. K. Ong, P. A. Madden, G. Pasterkamp, J. F. Peden, A. Peters, D. S. Postma, P. P. Pramstaller, J. F. Price, L. Qi, O. T. Raitakari, T. Rankinen, D. C. Rao, T. K. Rice, P. M. Ridker, J. D. Rioux, M. D. Ritchie, I. Rudan, V. Salomaa, N. J. Samani, J. Saramies, M. A. Sarzynski, H. Schunkert, P. E. Schwarz, P. Sever, A. R. Shuldiner, J. Sinisalo, R. P. Stolk, K. Strauch, A. Tonjes, D. A. Tregouet, A. Tremblay, E. Tremoli, J. Virtamo, M. C. Vohl, U. Volker, G. Waeber, G. Willemsen, J. C. Witteman, M. C. Zillikens, L. S. Adair, P. Amouyel, F. W. Asselbergs, T. L. Assimes, M. Bochud, B. O. Boehm, E. Boerwinkle, S. R. Bornstein, E. P. Bottinger, C. Bouchard, S. Cauchi, J. C. Chambers, S. J. Chanock, R. S. Cooper, P. I. de Bakker, G. Dedoussis, L. Ferrucci, P. W. Franks, P. Froguel, L. C. Groop, C. A. Haiman, A. Hamsten, J. Hui, D. J. Hunter, K. Hveem, R. C. Kaplan, M. Kivimaki, D. Kuh, M. Laakso, Y. Liu, N. G. Martin, W. Marz, M. Melbye, A. Metspalu, S. Moebus, P. B. Munroe, I. Njolstad, B. A. Oostra, C. N. Palmer, N. L. Pedersen, M. Perola, L. Perusse, U. Peters, C. Power, T. Quertermous, R. Rauramaa, F. Rivadeneira, T. E. Saaristo, D. Saleheen, N. Sattar, E. E. Schadt, D. Schlessinger, P. E. Slagboom, H. Snieder, T. D. Spector, U. Thorsteinsdottir, M. Stumvoll, J. Tuomilehto, A. G. Uitterlinden, M. Uusitupa, P. van der Harst, M. Walker, H. Wallaschofski, N. J. Wareham, H. Watkins, D. R. Weir, H. E. Wichmann, J. F. Wilson, P. Zanen, I. B. Borecki, P. Deloukas, C. S. Fox, I. M. Heid, J. R. O'Connell, D. P. Strachan, K. Stefansson, C. M. van Duijn, G. R. Abecasis, L. Franke, T. M. Frayling, M. I. McCarthy, P. M. Visscher, A. Scherag, C. J. Willer, M. Boehnke, K. L.

Mohlke, C. M. Lindgren, J. S. Beckmann, I. Barroso, K. E. North, E. Ingelsson, J. N. Hirschhorn, R. J. Loos and E. K. Speliotes (2015). "Genetic studies of body mass index yield new insights for obesity biology." Nature **518**(7538): 197-206.

Longo, D. L. (2012). Tuberculosis, in Harrison's Principles of Internal Medicine. D. S. Longo, A. S. Fauci, D. Kasper et al., McGraw Hill.

Lowe, D. M., P. S. Redford, R. J. Wilkinson, A. O'Garra and A. R. Martineau (2012). "Neutrophils in tuberculosis: friend or foe?" Trends in immunology **33**(1): 14-25.

Luong, K. and L. T. Nguyen (2011). "Impact of vitamin D in the treatment of tuberculosis." The American journal of the medical sciences **341**(6): 493-498.

Madhi, S. A., S. Nachman, A. Violari, S. Kim, M. F. Cotton, R. Bobat, P. Jean-Philippe, G. McSherry and C. Mitchell (2011). "Primary isoniazid prophylaxis against tuberculosis in HIV-exposed children." The New England journal of medicine **365**(1): 21-31.

Madhi, S. A., S. Nachman, A. Violari, S. Kim, M. F. Cotton, R. Bobat, P. Jean-Philippe, G. McSherry, C. Mitchell and P. S. Team (2011). "Primary isoniazid prophylaxis against tuberculosis in HIV-exposed children." N Engl J Med **365**(1): 21-31.

Maertzdorf, J., M. Ota, D. Repsilber, H. J. Mollenkopf, J. Weiner, P. C. Hill and S. H. Kaufmann (2011). "Functional correlations of pathogenesis-driven gene expression signatures in tuberculosis." PloS one **6**(10): e26938.

Maertzdorf, J., D. Repsilber, S. K. Parida, K. Stanley, T. Roberts, G. Black, G. Walzl and S. H. Kaufmann (2011). "Human gene expression profiles of susceptibility and resistance in tuberculosis." Genes and immunity **12**(1): 15-22.

Magi, R. and A. P. Morris (2010). "GWAMA: software for genome-wide association meta-analysis." BMC Bioinformatics **11**: 288.

Mahasirimongkol, S., H. Yanai, T. Mushiroda, W. Promphittayarat, S. Wattanapokayakit, J. Phromjai, R. Yuliwulandari, N. Wichukchinda, A. Yowang, N. Yamada, P. Kantipong, A. Takahashi, M. Kubo, P. Sawanpanyalert, N. Kamatani, Y. Nakamura and K. Tokunaga (2012). "Genome-wide association studies of tuberculosis in Asians identify distinct at-risk locus for young tuberculosis." Journal of human genetics **57**(6): 363-367.

Markovic, O., L. Popovic, D. Marisavljevic, D. Jovanovic, B. Filipovic, D. Stanisavljevic, G. Matovina-Brko, J. Hajder, T. Matkovic, R. Zivkovic, N. Stanisavljevic, M. Todorovic, D. Petrovic and B. Mihaljevic (2014). "Comparison of prognostic impact of absolute lymphocyte count, absolute monocyte count, absolute lymphocyte count/absolute monocyte count prognostic score and ratio in patients with diffuse large B cell lymphoma." Eur J Intern Med **25**(3): 296-302.

Matatall, K. A., C. C. Shen, G. A. Challen and K. Y. King (2014). "Type II interferon promotes differentiation of myeloid-biased hematopoietic stem cells." Stem Cells **32**(11): 3023-3030.

Mathad, J. S. and A. Gupta (2012). "Tuberculosis in pregnant and postpartum women: epidemiology, management, and research gaps." Clin Infect Dis **55**(11): 1532-1549.

Mayer-Barber, K. D., B. B. Andrade, S. D. Oland, E. P. Amaral, D. L. Barber, J. Gonzales, S. C. Derrick, R. Shi, N. P. Kumar, W. Wei, X. Yuan, G. Zhang, Y. Cai, S. Babu, M. Catalfamo, A. M. Salazar, L. E. Via, C. E. Barry, 3rd and A. Sher (2014). "Host-directed therapy of tuberculosis based on interleukin-1 and type I interferon crosstalk." Nature **511**(7507): 99-103.

McCracken, J. M. and L. A. Allen (2014). "Regulation of human neutrophil apoptosis and lifespan in health and disease." J Cell Death **7**: 15-23.

Mehta, D., K. Heim, C. Herder, M. Carstensen, G. Eckstein, C. Schurmann, G. Homuth, M. Nauck, U. Volker, M. Roden, T. Illig, C. Gieger, T. Meitinger and H. Prokisch (2013). "Impact of common regulatory single-nucleotide variants on gene expression profiles in whole blood." Eur J Hum Genet **21**(1): 48-54.

Merekoulias, G., E. C. Alexopoulos, T. Belezos, E. Panagiotopoulou and D. M. Jelastopulu (2010). "Lymphocyte to monocyte ratio as a screening tool for influenza." PLoS currents **2**: RRN1154.

Minegishi, Y., M. Saito, T. Morio, K. Watanabe, K. Agematsu, S. Tsuchiya, H. Takada, T. Hara, N. Kawamura, T. Ariga, H. Kaneko, N. Kondo, I. Tsuge, A. Yachie, Y. Sakiyama, T. Iwata, F. Bessho, T. Ohishi, K. Joh, K. Imai, K. Kogawa, M. Shinohara, M. Fujieda, H. Wakiguchi, S. Pasic, M. Abinun, H. D. Ochs, E. D. Renner, A. Jansson, B. H. Belohradsky, A. Metin, N. Shimizu, S. Mizutani, T. Miyawaki, S. Nonoyama and H. Karasuyama (2006). "Human tyrosine kinase 2 deficiency reveals its requisite roles in multiple cytokine signals involved in innate and acquired immunity." Immunity **25**(5): 745-755.

Moayyeri, A., C. J. Hammond, A. M. Valdes and T. D. Spector (2013). "Cohort Profile: TwinsUK and healthy ageing twin study." Int J Epidemiol **42**(1): 76-85.

Montgomery, S. B., M. Sammeth, M. Gutierrez-Arcelus, R. P. Lach, C. Ingle, J. Nisbett, R. Guigo and E. T. Dermitzakis (2010). "Transcriptome genetics using second generation sequencing in a Caucasian population." Nature **464**(7289): 773-777.

Mostafavi, S., A. Ortiz-Lopez, M. A. Bogue, K. Hattori, C. Pop, D. Koller, D. Mathis, C. Benoist and C. Immunological Genome (2014). "Variation and genetic control of gene expression in primary immunocytes across inbred mouse strains." J Immunol **193**(9): 4485-4496.

MRC (1948). "STREPTOMYCIN treatment of pulmonary tuberculosis." British medical journal **2**(4582): 769-782.

MRC (1952). "TREATMENT of pulmonary tuberculosis with isoniazid; an interim report to the Medical Research Council by their Tuberculosis Chemotherapy Trials Committee." British medical journal **2**(4787): 735-746.

Muller-Sieburg, C. E., R. H. Cho, L. Karlsson, J. F. Huang and H. B. Sieburg (2004). "Myeloid-biased hematopoietic stem cells have extensive self-renewal capacity but generate diminished lymphoid progeny with impaired IL-7 responsiveness." Blood **103**(11): 4111-4118.

Muller-Sieburg, C. E., R. H. Cho, M. Thoman, B. Adkins and H. B. Sieburg (2002). "Deterministic regulation of hematopoietic stem cell self-renewal and differentiation." Blood **100**(4): 1302-1309.

Muller-Sieburg, C. E., H. B. Sieburg, J. M. Bernitz and G. Cattarossi (2012). "Stem cell heterogeneity: implications for aging and regenerative medicine." Blood **119**(17): 3900-3907.

Nalls, M. A., D. J. Couper, T. Tanaka, F. J. van Rooij, M. H. Chen, A. V. Smith, D. Toniolo, N. A. Zakai, Q. Yang, A. Greinacher, A. R. Wood, M. Garcia, P. Gasparini, Y. Liu, T. Lumley, A. R. Folsom, A. P. Reiner, C. Gieger, V. Lagou, J. F. Felix, H. Volzke, N. A. Gouskova, A. Biffi, A. Doring, U. Volker, S. Chong, K. L. Wiggins, A. Rendon, A. Dehghan, M. Moore, K. Taylor, J. G. Wilson, G. Lettre, A. Hofman, J. C. Bis, N. Pirastu, C. S. Fox, C. Meisinger, J. Sambrook, S. Arepalli, M. Nauck, H. Prokisch, J. Stephens, N. L. Glazer, L. A. Cupples, Y. Okada, A. Takahashi, Y. Kamatani, K. Matsuda, T. Tsunoda, T. Tanaka, M. Kubo, Y. Nakamura, K. Yamamoto, N. Kamatani, M. Stumvoll, A. Tonjes, I. Prokopenko, T. Illig, K. V. Patel, S. F. Garner, B. Kuhnel, M. Mangino, B. A. Oostra, S. L. Thein, J. Coresh, H. E. Wichmann, S. Menzel,

J. Lin, G. Pistis, A. G. Uitterlinden, T. D. Spector, A. Teumer, G. Eiriksdottir, V. Gudnason, S. Bandinelli, T. M. Frayling, A. Chakravarti, C. M. van Duijn, D. Melzer, W. H. Ouwehand, D. Levy, E. Boerwinkle, A. B. Singleton, D. G. Hernandez, D. L. Longo, N. Soranzo, J. C. Witteman, B. M. Psaty, L. Ferrucci, T. B. Harris, C. J. O'Donnell and S. K. Ganesh (2011). "Multiple loci are associated with white blood cell phenotypes." PLoS Genet **7**(6): e1002113.

Naranbhai, V., Q. Abdool Karim, K. Naidoo, N. Yende Zuma and S. S. Abdool Karim (2012). "Task-shifting for antiretroviral treatment is not only effective but also sustainable and improves over time." The Lancet **380**(9857): 1907-1908.

Naranbhai, V., Cambridge Bioresource, United Kingdom Blood Study, TwinsUK, A. P. Morris and A. V. S. Hill Heritability and co-heritability of absolute and relative leucocyte counts suggest coregulation of myeloid and lymphoid cell numbers

Naranbhai, V., A. V. S. Hill, S. S. Abdool Karim, K. Naidoo, Q. Abdool Karim, G. M. Warimwe, H. McShane and H. Fletcher (2013). "Blood monocyte – lymphocyte ratios identify adults at risk of incident tuberculosis amongst patients initiating antiretroviral therapy." Journal of Infectious Diseases **In Press**.

Neumann, K., M. Castineiras-Vilarino, U. Hockendorf, N. Hannesschlager, S. Lemeer, D. Kupka, S. Meyermann, M. Lech, H. J. Anders, B. Kuster, D. H. Busch, A. Gewies, R. Naumann, O. Gross and J. Ruland (2014). "Clec12a is an inhibitory receptor for uric acid crystals that regulates inflammation in response to cell death." Immunity **40**(3): 389-399.

Nica, A. C., L. Parts, D. Glass, J. Nisbet, A. Barrett, M. Sekowska, M. Travers, S. Potter, E. Grundberg, K. Small, A. K. Hedman, V. Bataille, J. Tzenova Bell, G. Surdulescu, A. S. Dimas, C. Ingle, F. O. Nestle, P. di Meglio, J. L. Min, A. Wilk, C. J. Hammond, N. Hassanali, T. P. Yang, S. B. Montgomery, S. O'Rahilly, C. M. Lindgren,

K. T. Zondervan, N. Soranzo, I. Barroso, R. Durbin, K. Ahmadi, P. Deloukas, M. I. McCarthy, E. T. Dermitzakis, T. D. Spector and T. C. Mu (2011). "The architecture of gene regulatory variation across multiple human tissues: the MuTHER study." PLoS Genet **7**(2): e1002003.

Nunes-Alves, C., M. G. Booty, S. M. Carpenter, P. Jayaraman, A. C. Rothchild and S. M. Behar (2014). "In search of a new paradigm for protective immunity to TB." Nat Rev Microbiol **12**(4): 289-299.

O'Garra, A., P. S. Redford, F. W. McNab, C. I. Bloom, R. J. Wilkinson and M. P. Berry (2013). "The immune response in tuberculosis." Annu Rev Immunol **31**: 475-527.

Okada, Y., T. Hirota, Y. Kamatani, A. Takahashi, H. Ohmiya, N. Kumasaka, K. Higasa, Y. Yamaguchi-Kabata, N. Hosono, M. A. Nalls, M. H. Chen, F. J. van Rooij, A. V. Smith, T. Tanaka, D. J. Couper, N. A. Zakai, L. Ferrucci, D. L. Longo, D. G. Hernandez, J. C. Witteman, T. B. Harris, C. J. O'Donnell, S. K. Ganesh, K. Matsuda, T. Tsunoda, T. Tanaka, M. Kubo, Y. Nakamura, M. Tamari, K. Yamamoto and N. Kamatani (2011). "Identification of nine novel loci associated with white blood cell subtypes in a Japanese population." PLoS Genet **7**(6): e1002067.

Okada, Y., D. Wu, G. Trynka, T. Raj, C. Terao, K. Ikari, Y. Kochi, K. Ohmura, A. Suzuki, S. Yoshida, R. R. Graham, A. Manoharan, W. Ortmann, T. Bhangale, J. C. Denny, R. J. Carroll, A. E. Eyler, J. D. Greenberg, J. M. Kremer, D. A. Pappas, L. Jiang, J. Yin, L. Ye, D. F. Su, J. Yang, G. Xie, E. Keystone, H. J. Westra, T. Esko, A. Metspalu, X. Zhou, N. Gupta, D. Mirel, E. A. Stahl, D. Diogo, J. Cui, K. Liao, M. H. Guo, K. Myouzen, T. Kawaguchi, M. J. Coenen, P. L. van Riel, M. A. van de Laar, H. J. Guchelaar, T. W. Huizinga, P. Dieude, X. Mariette, S. L. Bridges, Jr., A. Zhernakova, R. E. Toes, P. P. Tak, C. Miceli-Richard, S. Y. Bang, H. S. Lee, J. Martin, M. A. Gonzalez-Gay, L. Rodriguez-Rodriguez, S. Rantapaa-Dahlqvist, L. Arlestig, H. K.

Choi, Y. Kamatani, P. Galan, M. Lathrop, R. consortium, G. consortium, S. Eyre, J. Bowes, A. Barton, N. de Vries, L. W. Moreland, L. A. Criswell, E. W. Karlson, A. Taniguchi, R. Yamada, M. Kubo, J. S. Liu, S. C. Bae, J. Worthington, L. Padyukov, L. Klareskog, P. K. Gregersen, S. Raychaudhuri, B. E. Stranger, P. L. De Jager, L. Franke, P. M. Visscher, M. A. Brown, H. Yamanaka, T. Mimori, A. Takahashi, H. Xu, T. W. Behrens, K. A. Siminovitch, S. Momohara, F. Matsuda, K. Yamamoto and R. M. Plenge (2014). "Genetics of rheumatoid arthritis contributes to biology and drug discovery." Nature **506**(7488): 376-381.

OMIM. "Online Mendelian Inheritance in Man, OMIM®. MIM Number: 253220: 02/12/2014." 2014, from http://omim.org/.

Pai, M., A. Zwerling and D. Menzies (2008). "Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: an update." Annals of Internal Medicine **149**(3): 177-184.

Pandey, A. K., Y. Yang, Z. Jiang, S. M. Fortune, F. Coulombe, M. A. Behr, K. A. Fitzgerald, C. M. Sassetti and M. A. Kelliher (2009). "NOD2, RIP2 and IRF5 play a critical role in the type I interferon response to Mycobacterium tuberculosis." PLoS Pathog **5**(7): e1000500.

Pang, W. W., E. A. Price, D. Sahoo, I. Beerman, W. J. Maloney, D. J. Rossi, S. L. Schrier and I. L. Weissman (2011). "Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age." Proceedings of the National Academy of Sciences of the United States of America **108**(50): 20012-20017.

Pang, W. W., E. A. Price, D. Sahoo, I. Beerman, W. J. Maloney, D. J. Rossi, S. L. Schrier and I. L. Weissman (2011). "Human bone marrow hematopoietic stem

cells are increased in frequency and myeloid-biased with age." Proc Natl Acad Sci U S A **108**(50): 20012-20017.

Paramanathan, A., A. Saxena and D. L. Morris (2014). "A systematic review and meta-analysis on the impact of pre-operative neutrophil lymphocyte ratio on long term outcomes after curative intent resection of solid tumours." Surg Oncol **23**(1): 31-39.

Pelletier, M., L. Maggi, A. Micheletti, E. Lazzeri, N. Tamassia, C. Costantini, L. Cosmi, C. Lunardi, F. Annunziato, S. Romagnani and M. A. Cassatella (2010). "Evidence for a cross-talk between human neutrophils and Th17 cells." Blood **115**(2): 335-343.

Petrella, S., E. Cambau, A. Chauffour, K. Andries, V. Jarlier and W. Sougakoff (2006). "Genetic basis for natural and acquired resistance to the diarylquinoline R207910 in mycobacteria." Antimicrobial agents and chemotherapy **50**(8): 2853-2856.

Petryszak, R., T. Burdett, B. Fiorelli, N. A. Fonseca, M. Gonzalez-Porta, E. Hastings, W. Huber, S. Jupp, M. Keays, N. Kryvych, J. McMurry, J. C. Marioni, J. Malone, K. Megy, G. Rustici, A. Y. Tang, J. Taubert, E. Williams, O. Mannion, H. E. Parkinson and A. Brazma (2014). "Expression Atlas update--a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments." Nucleic Acids Res **42**(Database issue): D926-932.

Pickrell, J. K., J. C. Marioni, A. A. Pai, J. F. Degner, B. E. Engelhardt, E. Nkadori, J. B. Veyrieras, M. Stephens, Y. Gilad and J. K. Pritchard (2010). "Understanding mechanisms underlying human gene expression variation with RNA sequencing." Nature **464**(7289): 768-772.

Pillay, T., A. W. Sturm, M. Khan, M. Adhikari, J. Moodley, C. Connolly, D. Moodley, N. Padayatchi, A. Ramjee, H. M. Coovadia and J. L. Sullivan (2004). "Vertical transmission of Mycobacterium tuberculosis in KwaZulu Natal: impact of HIV-1 co-infection." Int J Tuberc Lung Dis **8**(1): 59-69.

Png, E., B. Alisjahbana, E. Sahiratmadja, S. Marzuki, R. Nelwan, Y. Balabanova, V. Nikolayevskyy, F. Drobniewski, S. Nejentsev, I. Adnan, E. van de Vosse, M. L. Hibberd, R. van Crevel, T. H. Ottenhoff and M. Seielstad (2012). "A genome wide association study of pulmonary tuberculosis susceptibility in Indonesians." BMC medical genetics **13**: 5.

Porrata, L. F., D. J. Inwards, S. M. Ansell, I. N. Micallef, P. B. Johnston, W. J. Hogan and S. N. Markovic (2013). "Day 100 Peripheral Blood Absolute Lymphocyte/Monocyte Ratio and Survival in Classical Hodgkin's Lymphoma Postautologous Peripheral Blood Hematopoietic Stem Cell Transplantation." Bone Marrow Res **2013**: 658371.

Porrata, L. F., D. J. Inwards, S. M. Ansell, I. N. Micallef, P. B. Johnston, W. J. Hogan and S. N. Markovic (2014). "Infused Autograft Lymphocyte to Monocyte Ratio and Survival in Diffuse Large B-Cell Lymphoma." Biol Blood Marrow Transplant.

Porrata, L. F., K. Ristow, T. M. Habermann, T. E. Witzig, J. P. Colgan, D. J. Inwards, S. M. Ansell, I. N. Micallef, P. B. Johnston, G. S. Nowakowski, C. Thompson and S. N. Markovic (2012). "Peripheral blood lymphocyte/monocyte ratio at diagnosis and survival in nodular lymphocyte-predominant Hodgkin lymphoma." Br J Haematol **157**(3): 321-330.

Porrata, L. F., K. M. Ristow, T. M. Habermann, W. R. Macon, T. E. Witzig, J. P. Colgan, D. J. Inwards, S. M. Ansell, I. N. Micallef, P. B. Johnston, G. Nowakowski, C. A. Thompson and S. N. Markovic (2013). "Peripheral blood absolute

lymphocyte/monocyte ratio recovery during ABVD treatment cycles predicts clinical outcomes in classical Hodgkin lymphoma." Blood Cancer J **3**: e110.

Porrata, L. F., K. M. Ristow, T. M. Habermann, T. E. Witzig, J. P. Colgan, D. J. Inwards, S. M. Ansell, I. N. Micallef, P. B. Johnston, G. Nowakowski, C. A. Thompson and S. N. Markovic (2014). "Peripheral blood absolute lymphocyte/monocyte ratio during rituximab, cyclophosphamide, doxorubicin, vincristine and prednisone treatment cycles predicts clinical outcomes in diffuse large B-cell lymphoma." Leuk Lymphoma.

Power, C. and J. Elliott (2006). "Cohort profile: 1958 British birth cohort (National Child Development Study)." Int J Epidemiol **35**(1): 34-41.

Qu, H. Q., A. R. Rentfro, Y. Lu, S. Nair, C. L. Hanis, J. B. McCormick and S. P. Fisher-Hoch (2012). "Host susceptibility to tuberculosis: insights from a longitudinal study of gene expression in diabetes." The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease **16**(3): 370-372.

Raj, T., K. Rothamel, S. Mostafavi, C. Ye, M. N. Lee, J. M. Replogle, T. Feng, M. Lee, N. Asinovski, I. Frohlich, S. Imboywa, A. Von Korff, Y. Okada, N. A. Patsopoulos, S. Davis, C. McCabe, H. I. Paik, G. P. Srivastava, S. Raychaudhuri, D. A. Hafler, D. Koller, A. Regev, N. Hacohen, D. Mathis, C. Benoist, B. E. Stranger and P. L. De Jager (2014). "Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes." Science **344**(6183): 519-523.

Ramakrishnan, L. (2012). "Revisiting the role of the granuloma in tuberculosis." Nature reviews. Immunology **12**(5): 352-366.

Ramasamy, A., D. Trabzuni, S. Guelfi, V. Varghese, C. Smith, R. Walker, T. De, U. K. B. E. Consortium, C. North American Brain Expression, L. Coin, R. de Silva, M. R.

Cookson, A. B. Singleton, J. Hardy, M. Ryten and M. E. Weale (2014). "Genetic variability in the regulation of gene expression in ten regions of the human brain." Nat Neurosci **17**(10): 1418-1428.

Rambaldi, A., C. Boschini, G. Gritti, F. Delaini, E. Oldani, A. Rossi, A. M. Barbui, D. Caracciolo, M. Ladetto, A. Gueli, A. De Crescenzo, R. Passera, L. Devizzi, C. Patti, A. M. Gianni and C. Tarella (2013). "The lymphocyte to monocyte ratio improves the IPI-risk definition of diffuse large B-cell lymphoma when rituximab is added to chemotherapy." Am J Hematol **88**(12): 1062-1067.

Rana, F. S., M. P. Hawken, C. Mwachari, S. M. Bhatt, F. Abdullah, L. W. Ng'ang'a, C. Power, W. A. Githui, J. D. Porter and S. B. Lucas (2000). "Autopsy study of HIV-1-positive and HIV-1-negative adult medical patients in Nairobi, Kenya." Journal of acquired immune deficiency syndromes **24**(1): 23-29.

Rangaka, M. X., K. A. Wilkinson, J. R. Glynn, D. Ling, D. Menzies, J. Mwansa-Kambafwile, K. Fielding, R. J. Wilkinson and M. Pai (2012). "Predictive value of interferon-gamma release assays for incident active tuberculosis: a systematic review and meta-analysis." The Lancet infectious diseases **12**(1): 45-55.

Ravasi, T., H. Suzuki, C. V. Cannistraci, S. Katayama, V. B. Bajic, K. Tan, A. Akalin, S. Schmeier, M. Kanamori-Katayama, N. Bertin, P. Carninci, C. O. Daub, A. R. Forrest, J. Gough, S. Grimmond, J. H. Han, T. Hashimoto, W. Hide, O. Hofmann, A. Kamburov, M. Kaur, H. Kawaji, A. Kubosaki, T. Lassmann, E. van Nimwegen, C. R. MacPherson, C. Ogawa, A. Radovanovic, A. Schwartz, R. D. Teasdale, J. Tegner, B. Lenhard, S. A. Teichmann, T. Arakawa, N. Ninomiya, K. Murakami, M. Tagami, S. Fukuda, K. Imamura, C. Kai, R. Ishihara, Y. Kitazume, J. Kawai, D. A. Hume, T. Ideker and Y. Hayashizaki (2010). "An atlas of combinatorial transcriptional regulation in mouse and man." Cell **140**(5): 744-752.

Reiling, N., S. Homolka, K. Walter, J. Brandenburg, L. Niwinski, M. Ernst, C. Herzmann, C. Lange, R. Diel, S. Ehlers and S. Niemann (2013). "Clade-specific virulence patterns of Mycobacterium tuberculosis complex strains in human primary macrophages and aerogenically infected mice." MBio **4**(4).

Reiner, A. P., G. Lettre, M. A. Nalls, S. K. Ganesh, R. Mathias, M. A. Austin, E. Dean, S. Arepalli, A. Britton, Z. Chen, D. Couper, J. D. Curb, C. B. Eaton, M. Fornage, S. F. Grant, T. B. Harris, D. Hernandez, N. Kamatini, B. J. Keating, M. Kubo, A. LaCroix, L. A. Lange, S. Liu, K. Lohman, Y. Meng, E. R. Mohler, 3rd, S. Musani, Y. Nakamura, C. J. O'Donnell, Y. Okada, C. D. Palmer, G. J. Papanicolaou, K. V. Patel, A. B. Singleton, A. Takahashi, H. Tang, H. A. Taylor, Jr., K. Taylor, C. Thomson, L. R. Yanek, L. Yang, E. Ziv, A. B. Zonderman, A. R. Folsom, M. K. Evans, Y. Liu, D. M. Becker, B. M. Snively and J. G. Wilson (2011). "Genome-wide association study of white blood cell count in 16,388 African Americans: the continental origins and genetic epidemiology network (COGENT)." PLoS Genet **7**(6): e1002108.

Republic of South Africa Department of Health (2004). South African National Antiretroviral Treatment Guidelines.

Republic of South Africa Department of Health (2009). "South African National Tuberculosis Management Guidelines."

Republic of South Africa Department of Health (2010). "South African National Antiretroviral Treatment Guidelines."

Ristori, G., S. Romano, S. Cannoni, A. Visconti, E. Tinelli, L. Mendozzi, P. Cecconi, R. Lanzillo, M. Quarantelli, C. Buttinelli, C. Gasperini, M. Frontoni, G. Coarelli, D. Caputo, V. Bresciamorra, N. Vanacore, C. Pozzilli and M. Salvetti (2014). "Effects of Bacille Calmette-Guerin after the first demyelinating event in the CNS." Neurology **82**(1): 41-48.

Ritz, N. and N. Curtis (2009). "Mapping the global use of different BCG vaccine strains." Tuberculosis (Edinb) **89**(4): 248-251.

Ritz, N., W. A. Hanekom, R. Robins-Browne, W. J. Britton and N. Curtis (2008). "Influence of BCG vaccine strain on the immune response and protection against tuberculosis." FEMS Microbiol Rev **32**(5): 821-841.

Rogers, P. M. (1928). "A Study of the Blood Monocytes in Children with Tuberculosis." The New England journal of medicine **198**(14): 740-749.

Romano, A., C. Vetro, D. Donnarumma, S. Forte, M. Ippolito and F. Di Raimondo (2012). "Early interim 2-(1)fluoro-2-deoxy-D-glucose positron emission tomography is prognostically superior to peripheral blood lymphocyte/monocyte ratio at diagnosis in classical Hodgkin's lymphoma." Haematologica **97**(6): e21-23; author reply e24.

Roy, A., M. Eisenhut, R. J. Harris, L. C. Rodrigues, S. Sridhar, S. Habermann, L. Snell, P. Mangtani, I. Adetifa, A. Lalvani and I. Abubakar (2014). "Effect of BCG vaccination against Mycobacterium tuberculosis infection in children: systematic review and meta-analysis." BMJ **349**: g4643.

Royston, P., G. Ambler and W. Sauerbrei (1999). "The use of fractional polynomials to model continuous risk variables in epidemiology." Int J Epidemiol **28**(5): 964-974.

Royston, P. and W. Sauerbrei (2004). "A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials." Stat Med **23**(16): 2509-2525.

Ryan, K. J. and G. Ray (2010). Sherris Medical Microbiology.

Sabin, F. R., C. A. Doan and R. S. Cunningham (1926). "Studies of the Blood in Experimental Tuberculosis: The Monocyte-Lymphocyte Ratio; The Anemia-

Leucopenia Phase." Transactions of the 22nd Annual Meeting of the National Tuberculosis Association (22): 252-256.

Samandari, T., T. B. Agizew, S. Nyirenda, Z. Tedla, T. Sibanda, N. Shang, B. Mosimaneotsile, O. I. Motsamai, L. Bozeman, M. K. Davis, E. A. Talbot, T. L. Moeti, H. J. Moffat, P. H. Kilmarx, K. G. Castro and C. D. Wells (2011). "6-month versus 36-month isoniazid preventive treatment for tuberculosis in adults with HIV infection in Botswana: a randomised, double-blind, placebo-controlled trial." Lancet **377**(9777): 1588-1598.

Schadt, E. E., C. Molony, E. Chudin, K. Hao, X. Yang, P. Y. Lum, A. Kasarskis, B. Zhang, S. Wang, C. Suver, J. Zhu, J. Millstein, S. Sieberts, J. Lamb, D. GuhaThakurta, J. Derry, J. D. Storey, I. Avila-Campillo, M. J. Kruger, J. M. Johnson, C. A. Rohl, A. van Nas, M. Mehrabian, T. A. Drake, A. J. Lusis, R. C. Smith, F. P. Guengerich, S. C. Strom, E. Schuetz, T. H. Rushmore and R. Ulrich (2008). "Mapping the genetic architecture of gene expression in human liver." PLoS Biol **6**(5): e107.

Scott, D. L., F. Wolfe and T. W. Huizinga (2010). "Rheumatoid arthritis." Lancet **376**(9746): 1094-1108.

Sepulveda, R. L., I. M. Heiba, C. Navarrete, R. C. Elston, B. Gonzalez and R. U. Sorensen (1994). "Tuberculin reactivity after newborn BCG immunization in mono- and dizygotic twins." Tubercle and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease **75**(2): 138-143.

Seyler, C., S. Toure, E. Messou, D. Bonard, D. Gabillard and X. Anglaret (2005). "Risk factors for active tuberculosis after antiretroviral treatment initiation in Abidjan." American journal of respiratory and critical care medicine **172**(1): 123-127.

Shabalin, A. A. (2012). "Matrix eQTL: ultra fast eQTL analysis via large matrix operations." Bioinformatics **28**(10): 1353-1358.

Sheehan, N. A., V. Didelez, P. R. Burton and M. D. Tobin (2008). "Mendelian randomisation and causal inference in observational epidemiology." PLoS Med **5**(8): e177.

Shin, S. J., J. Roh, M. Kim, M. J. Jung, Y. W. Koh, C. S. Park, D. H. Yoon, C. Suh, C. J. Park, H. S. Chi and J. Huh (2013). "Prognostic significance of absolute lymphocyte count/absolute monocyte count ratio at diagnosis in patients with multiple myeloma." Korean J Pathol **47**(6): 526-533.

Sieburg, H. B., R. H. Cho, B. Dykstra, N. Uchida, C. J. Eaves and C. E. Muller-Sieburg (2006). "The hematopoietic stem compartment consists of a limited number of discrete stem cell subsets." Blood **107**(6): 2311-2316.

Sohn, D. H., K. Onuma, C. Rhodes, X. Zhao, T. Gazitt, R. Shiao, J. F. Bober, D. Cheng, L. J. Lahey, H. Wong, J. van Eyk, W. H. Robinson and J. Sokolove (2014). Netosis Induced Histone Citrullination Facilitates Onset and Propagation of Rheumatoid Arthritis. American College of Rheumatology. Boston, USA, Wiley.

Solovieff, N., C. Cotsapas, P. H. Lee, S. M. Purcell and J. W. Smoller (2013). "Pleiotropy in complex traits: challenges and strategies." Nat Rev Genet **14**(7): 483-495.

Soranzo, N., T. D. Spector, M. Mangino, B. Kuhnel, A. Rendon, A. Teumer, C. Willenborg, B. Wright, L. Chen, M. Li, P. Salo, B. F. Voight, P. Burns, R. A. Laskowski, Y. Xue, S. Menzel, D. Altshuler, J. R. Bradley, S. Bumpstead, M. S. Burnett, J. Devaney, A. Doring, R. Elosua, S. E. Epstein, W. Erber, M. Falchi, S. F. Garner, M. J. Ghori, A. H. Goodall, R. Gwilliam, H. H. Hakonarson, A. S. Hall, N. Hammond, C. Hengstenberg, T. Illig, I. R. Konig, C. W. Knouff, R. McPherson, O.

Melander, V. Mooser, M. Nauck, M. S. Nieminen, C. J. O'Donnell, L. Peltonen, S. C. Potter, H. Prokisch, D. J. Rader, C. M. Rice, R. Roberts, V. Salomaa, J. Sambrook, S. Schreiber, H. Schunkert, S. M. Schwartz, J. Serbanovic-Canic, J. Sinisalo, D. S. Siscovick, K. Stark, I. Surakka, J. Stephens, J. R. Thompson, U. Volker, H. Volzke, N. A. Watkins, G. A. Wells, H. E. Wichmann, D. A. Van Heel, C. Tyler-Smith, S. L. Thein, S. Kathiresan, M. Perola, M. P. Reilly, A. F. Stewart, J. Erdmann, N. J. Samani, C. Meisinger, A. Greinacher, P. Deloukas, W. H. Ouwehand and C. Gieger (2009). "A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium." Nat Genet **41**(11): 1182-1190.

Stein, C. M., D. Guwatudde, M. Nakakeeto, P. Peters, R. C. Elston, H. K. Tiwari, R. Mugerwa and C. C. Whalen (2003). "Heritability analysis of cytokines as intermediate phenotypes of tuberculosis." The Journal of infectious diseases **187**(11): 1679-1685.

Stejskal, S., I. Koutna and Z. Rucka (2010). "Isolation of granulocytes: which transcriptome do we analyse - neutrophils or eosinophils?" Folia Biol (Praha) **56**(6): 252-255.

Stotz, M., M. Pichler, G. Absenger, J. Szkandera, F. Arminger, R. Schaberl-Moser, H. Samonigg, T. Stojakovic and A. Gerger (2014). "The preoperative lymphocyte to monocyte ratio predicts clinical outcome in patients with stage III colon cancer." Br J Cancer **110**(2): 435-440.

Stranger, B. E., S. B. Montgomery, A. S. Dimas, L. Parts, O. Stegle, C. E. Ingle, M. Sekowska, G. D. Smith, D. Evans, M. Gutierrez-Arcelus, A. Price, T. Raj, J. Nisbett, A. C. Nica, C. Beazley, R. Durbin, P. Deloukas and E. T. Dermitzakis (2012). "Patterns

of cis regulatory variation in diverse human populations." PLoS Genet **8**(4): e1002639.

Stranger, B. E., A. C. Nica, M. S. Forrest, A. Dimas, C. P. Bird, C. Beazley, C. E. Ingle, M. Dunning, P. Flicek, D. Koller, S. Montgomery, S. Tavare, P. Deloukas and E. T. Dermitzakis (2007). "Population genomics of human gene expression." Nat Genet **39**(10): 1217-1224.

Sturm, R. A., D. L. Duffy, Z. Z. Zhao, F. P. Leite, M. S. Stark, N. K. Hayward, N. G. Martin and G. W. Montgomery (2008). "A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color." Am J Hum Genet **82**(2): 424-431.

Suzuki, A., R. Yamada, X. Chang, S. Tokuhiro, T. Sawada, M. Suzuki, M. Nagasaki, M. Nakayama-Hamada, R. Kawaida, M. Ono, M. Ohtsuki, H. Furukawa, S. Yoshino, M. Yukioka, S. Tohma, T. Matsubara, S. Wakitani, R. Teshima, Y. Nishioka, A. Sekine, A. Iida, A. Takahashi, T. Tsunoda, Y. Nakamura and K. Yamamoto (2003). "Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis." Nat Genet **34**(4): 395-402.

Szkandera, J., A. Gerger, B. Liegl-Atzwanger, G. Absenger, M. Stotz, J. Friesenbichler, S. Trajanoski, T. Stojakovic, K. Eberhard, A. Leithner and M. Pichler (2014). "The lymphocyte/monocyte ratio predicts poor clinical outcome and improves the predictive accuracy in patients with soft tissue sarcomas." Int J Cancer **135**(2): 362-370.

Templeton, A. J., M. G. McNamara, B. Seruga, F. E. Vera-Badillo, P. Aneja, A. Ocana, R. Leibowitz-Amit, G. Sonpavde, J. J. Knox, B. Tran, I. F. Tannock and E. Amir

(2014). "Prognostic role of neutrophil-to-lymphocyte ratio in solid tumors: a systematic review and meta-analysis." J Natl Cancer Inst **106**(6): dju124.

Thuong, N. T., S. J. Dunstan, T. T. Chau, V. Thorsson, C. P. Simmons, N. T. Quyen, G. E. Thwaites, N. Thi Ngoc Lan, M. Hibberd, Y. Y. Teo, M. Seielstad, A. Aderem, J. J. Farrar and T. R. Hawn (2008). "Identification of tuberculosis susceptibility genes with human macrophage gene expression profiles." PLoS pathogens **4**(12): e1000229.

Thye, T., E. Owusu-Dabo, F. O. Vannberg, R. van Crevel, J. Curtis, E. Sahiratmadja, Y. Balabanova, C. Ehmen, B. Muntau, G. Ruge, J. Sievertsen, J. Gyapong, V. Nikolayevskyy, P. C. Hill, G. Sirugo, F. Drobniewski, E. van de Vosse, M. Newport, B. Alisjahbana, S. Nejentsev, T. H. Ottenhoff, A. V. Hill, R. D. Horstmann and C. G. Meyer (2012). "Common variants at 11p13 are associated with susceptibility to tuberculosis." Nature genetics **44**(3): 257-259.

Thye, T., F. O. Vannberg, S. H. Wong, E. Owusu-Dabo, I. Osei, J. Gyapong, G. Sirugo, F. Sisay-Joof, A. Enimil, M. A. Chinbuah, S. Floyd, D. K. Warndorff, L. Sichali, S. Malema, A. C. Crampin, B. Ngwira, Y. Y. Teo, K. Small, K. Rockett, D. Kwiatkowski, P. E. Fine, P. C. Hill, M. Newport, C. Lienhardt, R. A. Adegbola, T. Corrah, A. Ziegler, A. P. Morris, C. G. Meyer, R. D. Horstmann and A. V. Hill (2010). "Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2." Nature genetics **42**(9): 739-741.

Tobin, D. M., F. J. Roca, S. F. Oh, R. McFarland, T. W. Vickery, J. P. Ray, D. C. Ko, Y. Zou, N. D. Bang, T. T. Chau, J. C. Vary, T. R. Hawn, S. J. Dunstan, J. J. Farrar, G. E. Thwaites, M. C. King, C. N. Serhan and L. Ramakrishnan (2012). "Host genotype-specific therapies can optimize the inflammatory response to mycobacterial infections." Cell **148**(3): 434-446.

Tobin, D. M., J. C. Vary, Jr., J. P. Ray, G. S. Walsh, S. J. Dunstan, N. D. Bang, D. A. Hagge, S. Khadge, M. C. King, T. R. Hawn, C. B. Moens and L. Ramakrishnan (2010). "The lta4h locus modulates susceptibility to mycobacterial infection in zebrafish and humans." Cell **140**(5): 717-730.

Tortoli, E. (2012). "Phylogeny of the genus Mycobacterium: many doubts, few certainties." Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases **12**(4): 827-831.

Trunz, B. B., P. Fine and C. Dye (2006). "Effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness." Lancet **367**(9517): 1173-1180.

Trynka, G., C. Sandor, B. Han, H. Xu, B. E. Stranger, X. S. Liu and S. Raychaudhuri (2013). "Chromatin marks identify critical cell types for fine mapping complex trait variants." Nat Genet **45**(2): 124-130.

Vall-Spinosa, A., W. Lester, T. Moulding, P. T. Davidson and J. K. McClatchy (1970). "Rifampin in the treatment of drug-resistant mycobacterium tuberculosis infections." The New England journal of medicine **283**(12): 616-621.

van der Eijk, E. A., E. van de Vosse, J. P. Vandenbroucke and J. T. van Dissel (2007). "Heredity versus environment in tuberculosis in twins: the 1950s United Kingdom Prophit Survey Simonds and Comstock revisited." American journal of respiratory and critical care medicine **176**(12): 1281-1288.

van Loggerenberg, F., K. Mlisana, C. Williamson, S. C. Auld, L. Morris, C. M. Gray, Q. Abdool Karim, A. Grobler, N. Barnabas, I. Iriogbe and S. S. Abdool Karim (2008). "Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study." PloS one **3**(4): e1954.

Veyrieras, J. B., S. Kudaravalli, S. Y. Kim, E. T. Dermitzakis, Y. Gilad, M. Stephens and J. K. Pritchard (2008). "High-resolution mapping of expression-QTLs yields insight into human gene regulation." PLoS Genet **4**(10): e1000214.

von Bruhl, M. L., K. Stark, A. Steinhart, S. Chandraratne, I. Konrad, M. Lorenz, A. Khandoga, A. Tirniceriu, R. Coletti, M. Kollnberger, R. A. Byrne, I. Laitinen, A. Walch, A. Brill, S. Pfeiler, D. Manukyan, S. Braun, P. Lange, J. Riegger, J. Ware, A. Eckart, S. Haidari, M. Rudelius, C. Schulz, K. Echtler, V. Brinkmann, M. Schwaiger, K. T. Preissner, D. D. Wagner, N. Mackman, B. Engelmann and S. Massberg (2012). "Monocytes, neutrophils, and platelets cooperate to initiate and propagate venous thrombosis in mice in vivo." J Exp Med **209**(4): 819-835.

Wallis, R. S., M. Palaci, S. Vinhas, A. G. Hise, F. C. Ribeiro, K. Landen, S. H. Cheon, H. Y. Song, M. Phillips, R. Dietze and J. J. Ellner (2001). "A whole blood bactericidal assay for tuberculosis." J Infect Dis **183**(8): 1300-1303.

Walzl, G., K. Ronacher, W. Hanekom, T. J. Scriba and A. Zumla (2011). "Immunological biomarkers of tuberculosis." Nature reviews. Immunology **11**(5): 343-354.

Wang, J., Y. Yin, X. Wang, H. Pei, S. Kuai, L. Gu, H. Xing, Y. Zhang, Q. Huang and B. Guan (2015). "Ratio of monocytes to lymphocytes in peripheral blood in patients diagnosed with active tuberculosis." Braz J Infect Dis **19**(2): 125-131.

Wang, X., G. Zhang, X. Jiang, H. Zhu, Z. Lu and L. Xu (2014). "Neutrophil to lymphocyte ratio in relation to risk of all-cause mortality and cardiovascular events among patients undergoing angiography or cardiac revascularization: a meta-analysis of observational studies." Atherosclerosis **234**(1): 206-213.

Warimwe, G. M., H. A. Fletcher, A. Olotu, S. T. Agnandji, A. V. Hill, K. Marsh and P. Bejon (2013). "Peripheral blood monocyte-to-lymphocyte ratio at study

enrollment predicts efficacy of the RTS,S malaria vaccine: analysis of pooled phase II clinical trial data." BMC Med **11**: 184.

Warimwe, G. M., L. M. Murungi, G. Kamuyu, G. M. Nyangweso, J. Wambua, V. Naranbhai, H. A. Fletcher, A. V. Hill, P. Bejon, F. H. Osier and K. Marsh (2013). "The ratio of monocytes to lymphocytes in peripheral blood correlates with increased susceptibility to clinical malaria in Kenyan children." PLoS One **8**(2): e57320.

Watanabe, R., N. Tomita, M. Itabashi, D. Ishibashi, E. Yamamoto, S. Koyama, K. Miyashita, H. Takahashi, Y. Nakajima, Y. Hattori, K. Motohashi, H. Takasaki, R. Ohshima, C. Hashimoto, E. Yamazaki, K. Fujimaki, R. Sakai, S. Fujisawa, S. Motomura and Y. Ishigatsubo (2014). "Peripheral blood absolute lymphocyte/monocyte ratio as a useful prognostic factor in diffuse large B-cell lymphoma in the rituximab era." Eur J Haematol **92**(3): 204-210.

Watts, D. H., E. R. Brown, Y. Maldonado, C. Herron, T. Chipato, L. Reddy, D. Moodley, C. Nakabiito, K. Manji, W. Fawzi, K. George, P. Richardson, S. Zwerski, H. Coovadia, M. Fowler and H. P. Team (2013). "HIV disease progression in the first year after delivery among African women followed in the HPTN 046 clinical trial." J Acquir Immune Defic Syndr **64**(3): 299-306.

Wei, X., F. Huang, Y. Wei, H. Jing, M. Xie, X. Hao and R. Feng (2014). "Low lymphocyte-to-monocyte ratio predicts unfavorable prognosis in non-germinal center type diffuse large B-cell lymphoma." Leuk Res **38**(6): 694-698.

Welter, D., J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio, L. Hindorff and H. Parkinson (2014). "The NHGRI GWAS Catalog, a curated resource of SNP-trait associations." Nucleic Acids Res **42**(Database issue): D1001-1006.

Westra, H., D. Arends, T. Esko, M. J. Peters, C. Schurmann, K. Schramm, J. Kettunen, H. Yaghootkar, B. P. Fairfax, A. K. Andiappan, Y. Li, J. Fu, J. Karjalainen, M. Platteel, M. Visschedijk, R. Weersma, S. Kasela, L. Milani, L. Tserel, P. Peterson, E. Reinmaa, A. Hofman, A. G. Uitterlinden, F. Rivadeneira, G. Homuth, A. Petersmann, R. Lorbeer, H. Prokisch, T. Meitinger, C. Herder, M. Roden, H. Grallert, S. Ripatti, M. Perola, A. R. Wood, D. Melzer, L. Ferrucci, A. B. Singleton, D. G. Hernandez, J. C. Knight, R. Melchiotti, B. Lee, M. Poidinger, F. Zolezzi, A. Larbi, D. Y. Wang, L. H. van den Berg, J. H. Veldink, O. Rotzschke, S. Makino, T. Frayling, V. Salomaa, K. Strauch, U. Völker, J. B. J. van Meurs, A. Metspalu, C. Wijmenga, R. C. Jansen and L. Franke (2014). "Cell specific eQTL analysis without sorting cells." bioRxiv.

Westra, H. J., M. J. Peters, T. Esko, H. Yaghootkar, C. Schurmann, J. Kettunen, M. W. Christiansen, B. P. Fairfax, K. Schramm, J. E. Powell, A. Zhernakova, D. V. Zhernakova, J. H. Veldink, L. H. Van den Berg, J. Karjalainen, S. Withoff, A. G. Uitterlinden, A. Hofman, F. Rivadeneira, P. A. t Hoen, E. Reinmaa, K. Fischer, M. Nelis, L. Milani, D. Melzer, L. Ferrucci, A. B. Singleton, D. G. Hernandez, M. A. Nalls, G. Homuth, M. Nauck, D. Radke, U. Volker, M. Perola, V. Salomaa, J. Brody, A. Suchy-Dicey, S. A. Gharib, D. A. Enquobahrie, T. Lumley, G. W. Montgomery, S. Makino, H. Prokisch, C. Herder, M. Roden, H. Grallert, T. Meitinger, K. Strauch, Y. Li, R. C. Jansen, P. M. Visscher, J. C. Knight, B. M. Psaty, S. Ripatti, A. Teumer, T. M. Frayling, A. Metspalu, J. B. van Meurs and L. Franke (2013). "Systematic identification of trans eQTLs as putative drivers of known disease associations." Nat Genet **45**(10): 1238-1243.

Wheeler, J. G., M. E. Mussolino, R. F. Gillum and J. Danesh (2004). "Associations between differential leucocyte count and incident coronary heart disease: 1764

incident cases from seven prospective studies of 30,374 individuals." Eur Heart J **25**(15): 1287-1292.

WHO. "WHO| MDG 6: combat HIV/AIDS, malaria and other diseases." from http://www.who.int/topics/millennium_development_goals/diseases/en/.

WHO (2012). "Global tuberculosis report 2012."

WHO (2014). Global Tuberculosis Report.

WHO (2014). Guidance for national tuberculosis programmes on the management of tuberculosis in children –2nd ed.

Wilhelm, M., J. Schlegl, H. Hahne, A. Moghaddas Gholami, M. Lieberenz, M. M. Savitski, E. Ziegler, L. Butzmann, S. Gessulat, H. Marx, T. Mathieson, S. Lemeer, K. Schnatbaum, U. Reimer, H. Wenschuh, M. Mollenhauer, J. Slotta-Huspenina, J. H. Boese, M. Bantscheff, A. Gerstmair, F. Faerber and B. Kuster (2014). "Mass-spectrometry-based draft of the human proteome." Nature **509**(7502): 582-587.

Wilson, M. E., H. V. Fineberg and G. A. Colditz (1995). "Geographic latitude and the efficacy of bacillus Calmette-Guerin vaccine." Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **20**(4): 982-991.

Wong, D., W. Lee, P. Humburg, S. Makino, E. Lau, V. Naranbhai, B. P. Fairfax, K. Chan, K. Plant and J. C. Knight (2014). "Genomic mapping of the MHC transactivator CIITA using an integrated ChIP-seq and genetical genomics approach." Genome Biol **15**(10): 494.

WorldHealthAssembly (2014). Global strategy and targets for tuberculosis prevention, care and control after 2015. Geneva.

Wright, F. A., P. F. Sullivan, A. I. Brooks, F. Zou, W. Sun, K. Xia, V. Madar, R. Jansen, W. Chung, Y. H. Zhou, A. Abdellaoui, S. Batista, C. Butler, G. Chen, T. H. Chen, D. D'Ambrosio, P. Gallins, M. J. Ha, J. J. Hottenga, S. Huang, M. Kattenberg, J. Kochar, C.

M. Middeldorp, A. Qu, A. Shabalin, J. Tischfield, L. Todd, J. Y. Tzeng, G. van Grootheest, J. M. Vink, Q. Wang, W. Wang, W. Wang, G. Willemsen, J. H. Smit, E. J. de Geus, Z. Yin, B. W. Penninx and D. I. Boomsma (2014). "Heritability and genomics of gene expression in peripheral blood." Nat Genet **46**(5): 430-437.

Wright, H. L., R. J. Moots and S. W. Edwards (2014). "The multifactorial role of neutrophils in rheumatoid arthritis." Nat Rev Rheumatol **10**(10): 593-601.

Wu, W. C., H. W. Sun, H. T. Chen, J. Liang, X. J. Yu, C. Wu, Z. Wang and L. Zheng (2014). "Circulating hematopoietic stem and progenitor cells are myeloid-biased in cancer patients." Proc Natl Acad Sci U S A **111**(11): 4221-4226.

Yan-Li, L., G. Kang-Sheng, P. Yue-Yin, J. Yang and Z. Zhi-Min (2014). "The lower peripheral blood lymphocyte/monocyte ratio assessed during routine follow-up after standard first-line chemotherapy is a risk factor for predicting relapse in patients with diffuse large B-cell lymphoma." Leuk Res **38**(3): 323-328.

Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. E. Goddard and P. M. Visscher (2010). "Common SNPs explain a large proportion of the heritability for human height." Nat Genet **42**(7): 565-569.

Yang, J., S. H. Lee, M. E. Goddard and P. M. Visscher (2011). "GCTA: a tool for genome-wide complex trait analysis." Am J Hum Genet **88**(1): 76-82.

Yang, J., T. A. Manolio, L. R. Pasquale, E. Boerwinkle, N. Caporaso, J. M. Cunningham, M. de Andrade, B. Feenstra, E. Feingold, M. G. Hayes, W. G. Hill, M. T. Landi, A. Alonso, G. Lettre, P. Lin, H. Ling, W. Lowe, R. A. Mathias, M. Melbye, E. Pugh, M. C. Cornelis, B. S. Weir, M. E. Goddard and P. M. Visscher (2011). "Genome partitioning of genetic variation for complex traits using common SNPs." Nat Genet **43**(6): 519-525.

Zeller, T., P. Wild, S. Szymczak, M. Rotival, A. Schillert, R. Castagne, S. Maouche, M. Germain, K. Lackner, H. Rossmann, M. Eleftheriadis, C. R. Sinning, R. B. Schnabel, E. Lubos, D. Mennerich, W. Rust, C. Perret, C. Proust, V. Nicaud, J. Loscalzo, N. Hubner, D. Tregouet, T. Munzel, A. Ziegler, L. Tiret, S. Blankenberg and F. Cambien (2010). "Genetics and beyond--the transcriptome of human monocytes and disease susceptibility." PLoS One **5**(5): e10693.

Zhang, D. E., P. Zhang, N. D. Wang, C. J. Hetherington, G. J. Darlington and D. G. Tenen (1997). "Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice." Proc Natl Acad Sci U S A **94**(2): 569-574.

Zhang, F. R., W. Huang, S. M. Chen, L. D. Sun, H. Liu, Y. Li, Y. Cui, X. X. Yan, H. T. Yang, R. D. Yang, T. S. Chu, C. Zhang, L. Zhang, J. W. Han, G. Q. Yu, C. Quan, Y. X. Yu, Z. Zhang, B. Q. Shi, L. H. Zhang, H. Cheng, C. Y. Wang, Y. Lin, H. F. Zheng, X. A. Fu, X. B. Zuo, Q. Wang, H. Long, Y. P. Sun, Y. L. Cheng, H. Q. Tian, F. S. Zhou, H. X. Liu, W. S. Lu, S. M. He, W. L. Du, M. Shen, Q. Y. Jin, Y. Wang, H. Q. Low, T. Erwin, N. H. Yang, J. Y. Li, X. Zhao, Y. L. Jiao, L. G. Mao, G. Yin, Z. X. Jiang, X. D. Wang, J. P. Yu, Z. H. Hu, C. H. Gong, Y. Q. Liu, R. Y. Liu, D. M. Wang, D. Wei, J. X. Liu, W. K. Cao, H. Z. Cao, Y. P. Li, W. G. Yan, S. Y. Wei, K. J. Wang, M. L. Hibberd, S. Yang, X. J. Zhang and J. J. Liu (2009). "Genomewide association study of leprosy." N Engl J Med **361**(27): 2609-2618.

Zhong, H. and K. Yazdanbakhsh (2013). "Differential control of Helios(+/-) Treg development by monocyte subsets through disparate inflammatory cytokines." Blood **121**(13): 2494-2502.

Zumla, A., P. Nahid and S. T. Cole (2013). "Advances in the development of new tuberculosis drugs and treatment regimens." Nat Rev Drug Discov **12**(5): 388-404.