UNIVERSITY OF CALIFORNIA

Los Angeles

Examining the impact of Neanderthal DNA on modern human biology

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Human Genetics

by

Christopher Raymond Robles

2022

ABSTRACT OF THE DISSERTATION

Examining the impact
of Neanderthal DNA
on modern human biology

by

Christopher Raymond Robles
Doctor of Philosophy in Human Genetics
University of California, Los Angeles, 2022
Professor Sriram Sankararaman, Co-Chair
Professor Paivi Elisabeth Pajukanta, Co-Chair

Recent genomic studies have revealed that all present-day non-African populations inherit $1-4\%$ of their genetic ancestry from a population related to the Neanderthals. Due to the high divergence of Neanderthals and modern humans, this introgression event introduced many novel mutations into non-African populations. Although introgression between archaic and modern humans have previously been documented, the biological consequences of these admixture events on modern humans are not fully understood. In this thesis, we examine these introgressed mutations to help elucidate the biological differences between Neanderthals and modern humans, as well as determine the selective forces that have acted on our genomes in the approximately 50,000 years since Neanderthal introgression occurred.

We first examine the impact Neanderthal introgression has on a single phenotype, Major Depressive Disorder, in more than 10,000 Han Chinese individuals who were diagnosed using structured clinical interviews. A recent examination of Neanderthal informative mutations

(NIMs) among a large cohort of Europeans showed that these markers explained some proportions of the phenotypic risk of depression. Our results found no association between the NIMs and MDD, as well as a difference in the variance explained by the NIMs upon the phenotype vs Europeans. These results demonstrate the impact due to Neanderthal ancestry on one type of trait, MDD, appears to differ between Han Chinese and Europeans.

We expanded upon a single phenotype by looking at the impact of Neanderthal introgression on wide range of phenotypes in white British individuals in the UK Biobank. We present a fine mapping strategy for prioritizing causal NIMs that enables the identification of sets of NIMs that can credibly exert influence on specific phenotypes. We identify NIMs in a number of functionally important genes, including a premature stop codon in the FCGR2A gene, and a start codon loss in COQ10, pinpointing introgressed alleles in immune-related genes that could have and continue to modulate human phenotype.

We then propose a pipeline that is able leverage information from Neanderthal introgressed alleles in order to identify Human-specific mutations that rose to near fixation in modern human populations, and examine how they impact modern human phenotypes. We discover two regions of the genome that are confidently fixed-derived and have associations with a number of phenotypes, and are in close proximity to genes that are related to immunity suggesting that these regions may have importance in development of immune response in modern humans.

Finally, we look to highlight possible mechanisms for how complex traits evolved in human history by examining the genetic contribution of Neanderthal ancestry. We performed forward-in-time population genetic simulations to model the evolution of Neanderthal and non-Neanderthal alleles according to a demographic model relating modern humans and Neanderthals, to try an understand the evolutionary models that could explain observations of enriched or depleted contributions of Neanderthal ancestry to phenotypic variance seen in real data. We develop simulations of models of directional, stabilizing and disruptive selection, as well as propose two modified directional and stabilizing selection models, and find

that most models lead to NIM heritability that is comparable or lower than non-introgressed SNPs, however our two modified models are able to present an increase in NIM heritability, suggesting possible mechanisms for how complex traits evolved in human history by examining the genetic contribution of Neanderthal ancestry.

The dissertation of Christopher Raymond Robles is approved.

Bogdan Pasaniuc

Eleazar Eskin

Paivi Elisabeth Pajukanta, Committee Co-Chair

Sriram Sankararaman, Committee Co-Chair

University of California, Los Angeles

2022

For my family

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Graduate school at UCLA has been both a rewarding and trying experience, and I would like to thank all of the people who have helped me along my journey and seen all the ups and downs. I would first like to thank my advisor Sriram, who has been the most supportive advisor I could ever ask for. As Sriram's first student, I have seen the lab grow from just him and I in the basement of Boyer hall, to a large group of students in the new Engineering 6 building. I vividly remember my first year working with Sriram, where he stayed late at night with me in the old lab, going through his CRF code line by line with me, patiently explaining everything. I started in his lab with a limited knowledge of statistics and machine learning, and his constant support, patience, and teachings have helped me grow tremendously as a mathematician, bioinformatician, scientist, and researcher. I have been so lucky to work with such a great scientist, mentor, and person overall.

I would also like to the thank my committee: Eleazar Eskin, for always encouraging me and showing me that research can be fun and rewarding. His CGSI community has helped me build a large network, and allowed me to informally interact with and discuss reserach problems with many people who I consider to not only be my colleagues, but friends. Paivi Pajukanta, for always being there for me when I had any sort of problem, she was just a phone call away. Bogdan, for all the great discussion we had in journal clubs and at different events, as well as some cool soccer games.

I would also like to thank my lab mates, Alec, Brandon, Leah, Mike, Arun, Ariel, Ali, Ruthie, Elior, Liat, Brian, Rob Brown, Nadav and Johnson. All those great late nights shooting hoops around while we were going crazy with trying to figure out our codes.

I would like to thank my partner Sheila Rahimi, whose support during my toughest times means the world to me. I would also like to thank my friends at UCLA, Cat, Eric, Nate, Wenting, and Adriana. You guys made me stay sane through all my ups and downs, just being able to grab a beer a talk at the pub.

I would like to thank my sister, Danielle Robles, she always had my back, its me and you against the world.

Last and not least, I would like to thank my parents Raymond and Lorraine Robles. They sacrificed so much for me, just to get me where I am today. Without their love and support I could never be where I was. I hope to one day pay you guys back for all the unconditional love, support, and cheering you have provided me with for all time. I love you both so much.

VITA

2007–2011	B.S. in Bioengineering *cum laude honors*, University of California Santa Cruz, Santa Cruz, CA
2010	Undergraduate Research Assistant, Broad institute of Harvard and MIT, Cambridge, MA.
2011–2013	NIH Academy Research Fellow, National Human Genome Research Institute (NHGRI),Bethesda, MD
2013–2014	MHIRT Research Associate, University of Otago, Dunedin, New Zealand
2014–2015	Bioinformatics Analyst, Mount Sinai School of Medicine, New York, NY
2015–2022	Graduate Student Fellow, University of California Los Angeles, Los Angeles, California

PUBLICATIONS

Xinzhu Wei*, **Christopher R. Robles*,** , Ali Pazokitoroudi, Andrea Ganna, Alexander Gusev, Arun Durvasula, Steven Gazal, Po-Ru Loh, David Reich and Sriram Sankararaman."The lingering effects of Neanderthal introgression on human complex traits". *In submission. *Equally contributed*,
https://www.biorxiv.org/content/10.1101/2022.06.07.495223v1.full.pdf

Charleston W. K. Chiang, Serghei Mangul, **Christopher R. Robles**, Sriram Sankararam"A comprehensive map of genetic variation in the world's largest ethnic group - Han Chinese".

Molecular Biology and Evolution, Volume 35, Issue 11, 1 November 2018, Pages 2736–2750, https://doi.org/10.1093/molbev/msy170

Kristina M. Garske, David Z. Pan, Zong Miao, Yash V. Bhagat, Caroline Comenho, **Christopher R. Robles**, Jihane N. Benhammou, Marcus Alvarez, Arthur Ko, Chun Jimmie Ye, Joseph R. Pisegna, Karen L. Mohlke, Janet S. Sinsheimer, Markku Laakso, Päivi Pajukanta."Reverse gene-environment interaction approach to identify variants influencing body mass index in humans".*Nat Metab 1*, 630–642 (2019), https://doi.org/10.1038/s42255-019-0071-6

CHAPTER 1

Introduction

Where did we come from, and where will we go? Throughout history, one of humankind's greatest questions is how we came to be, and what might our future hold? The relationship of modern humans and our archaic hominid ancestors, such as Neanderthals, has been debated for quite some time. Historically, there were two theories of evolution and migration of modern humans, the multi-regional and out-of-Africa models, with evolutionary research seeming to favor the latter [3]. Around the time that modern humans left Africa for Eurasia, archaeological evidence shows that other archaic hominids also inhabited these regions, and may have come into contact with modern humans [4]. In recent years, advances in genomic technologies allowed for extraction and analysis of DNA from several of these ancient ancestors including Neanderthals [5, 6, 7], illuminating that fact that admixture occurred between these two species [8]. Further analysis revealed that all present-day non-African populations inherit 1-4% of their genetic ancestry from a population related to the Neanderthals [5], and that Neanderthals had lower genetic diversity than any modern human population [7]. Due to this high divergence between the two species, this introgression event introduced many novel mutations into the non-African population. Around the time of this introgression event, archaeological records suggest that modern humans were experiencing behavioral modernity, or cognitive traits such as abstract thinking, which distinguish humans from closely related species.

Systematically studying these mutations has the potential to provide clues about the biological differences between Neanderthals and modern humans, as well as the selective

forces that have acted on our genomes in the approximately 50,000 years since Neanderthal introgression occurred. The fact that the period of time since Neanderthal introgression coincides with the period of behavioral modernity evident in the archaeological record [9] suggests that studying the evolution of Neanderthal-derived mutations in modern humans over this period, will give us insight into the nature of natural selection during this critical period of our species' evolution.

Analysis of how these Neanderthal segments are distributed in the non-African genome indicates that Neanderthal variants underwent various types of selective pressures [10, 11]. Genomic regions of reduced Neanderthal ancestry are enriched in genes and imply a negative selection of Neanderthal genetic material. One such region is the X chromosome which shows a five-fold reduction of Neanderthal ancestry. This observation is notable as the X chromosome is a region known to harbor many male hybrid sterility genes suggesting that Neanderthal alleles caused decreased fertility in males. This is consistent with the hypothesis that the bulk of Neanderthal variants were deleterious in the modern human genetic background [10, 11]. On the other hand, the frequency of Neanderthal haplotypes is substantially elevated in a small number of genomic locations suggesting evidence for archaic adaptive introgression [10, 11, 12, 13, 14]. Analyses of these genomic locations have suggested that Neanderthal variants could have had an important impact on immune-related as well as skin and hair-related traits, However, the effects of these Neanderthal variants on phenotypes, and selections is still not understood. In principle a powerful approach to assessing the biological impact of Neanderthal interbreeding is to study Neanderthal-derived mutations in very large cohorts of individuals measured for diverse phenotypes. A recent study employed such an approach to analyze electronic medical records and genotypes in about 28,000 individuals to show that Neanderthal variants modulate risk for disease traits such as major depression, blood-clotting disorders and tobacco use [15]. A difficulty with this approach is that variants introgressed from Neanderthals are rare on average (due to the low proportion of Neanderthal ancestry in present-day genomes) and the genotypes for most

rare variants cannot be reliably inferred with the arrays typically used in genetic association studies. Another study analyzed about 112,000 individuals from the interim release of the UK Biobank and identified Neanderthal variants that are individually associated with skin tone, hair color, height, sleeping patterns, mood, and smoking [16]. However, beyond identifying the associations of individual Neanderthal variants, the systematic impact of these variants on a broad spectrum of phenotypes remains to be rigorously assessed. Knowing this, leads us to other relevant questions in how these homonid species impacted our modern human biology. (1) How has neanderthal introgression impacting genetic and phenotypic variation in modern humans? (2) Do these suggest any functional relevance? (3) Was this genomic material harmful or beneficial?

In this dissertation, I will discuss how the movement of ancient hominid DNA impacts the genomic landscape of modern humans and in turn how this impacts our modern human biology through variation in phenotype. In chapter 2, I present work that looks at how neanderthal introgression impacts a single phenotype that has been extremely well cataloged. We examined 10,000 Han Chinese individuals that were diagnosed for Major Depressive Disorder and Melancholia. This chapter contains excerpts of my contributions from a previously published paper, Chiang et al. [17]. In chapter 3, I present the bulk of my PhD work in which we examine how Neanderthals ancestry in the modern human genome impacts a wide range of phenotypes in white British individuals in the UK Biobank. In chapter 4, I build upon the analysis of this dataset by looking at how modern human specific regions in the human genome impact our biology. Lastly, in chapter 5, I develop simulated models of selective forces acting upon a evolutionary history of Neanderthal introgression, aiming to explain changes in heritbaility we see in previous sections. Taken together, these chapters elucidate how humankind's interweaving history with our closest hominid relatives continue to impact us today.

CHAPTER 2

Impact of Neanderthal DNA on depression in Han Chinese individuals

2.1 Introduction

This chapter contains excerpts of a published paper Chiang et al. [17] presented here with permission from the authors. Previous analyses of the locations of Neanderthal segments within the genomes of non-African individuals indicated that some of the Neanderthal variants were adaptively beneficial while the bulk of Neanderthal variants were deleterious in the modern human genetic background [18, 19]. Specifically, a recent examination of Neanderthal-informative markers (NIMs) among a large cohort of Europeans showed that these markers explained some proportions of the phenotypic risk of a number of diseases in the electronic health record [15], including depression. A challenge with these studies is that depression is a challenging disease to accurately phenotype. We sought to replicate this finding in East Asians as our data set was originally ascertained as a case-control study of Major Depressive Disorder (MDD) in Han Chinese women [20]. In this dataset, the depression status was determined using structured clinical interviews to increase diagnostic accuracy [21].

To date, a range of strategies has been employed to characterize populations across large numbers of individuals. In Genomes of the Netherlands (GoNL) [22], the trio design allowed estimation of high-quality genotypes for both single nucleotide and structural variations with intermediate ($\sim 13\times$) sequencing coverage and enabled the investigation of de novo mutations. In Sardinian [23] and Icelandic [24] population cohorts, extensive haplotype

sharing within populations was used to inform accurate genotype calling among low- ($\sim 4-6\times$) and intermediate- ($\sim 20\times$) coverage sequencing of $\sim 2,000-3,000$ individuals. In the UK10K project [25], low ($\sim 7x$) whole-genome sequencing in 3,781 healthy samples from two British cohorts was combined with deep ($80\times$) exome sequencing in three disease cohorts to accurately detect low frequency and rare variants associated with quantitative traits. However, much like the genome-wide association studies preceding the current era of sequencing studies, most sequencing efforts are biased toward European populations.

The whole-genome sequencing data set of the Han Chinese analyzed here adopted a different approach. We analyzed genetic variants from very low-coverage whole-genome sequencing data of 11,670 Han Chinese women, previously generated to study MDD [20]. With a median coverage of $1.7\times$, this data set is predicted to identify rare ($< 0.5\%$) single nucleotide polymorphisms (SNPs) with high confidence [23] and obtain accurate estimates of allele frequencies in a large sample.

2.2 Results

A total of 11,670 Han Chinese women were previously sequenced at a median coverage of $1.7\times$ per individual, of which 10,640 individuals and 25,057,223 SNPs remained after quality control (QC) [20, 21]. Despite the low median coverage genome-wide, individual level genotype calls showed high concordance ($> 96-97\%$) in validation experiments [21]. Furthermore, the allele frequencies in this data set are highly correlated (mean $r = 0.995$ across chromosomes) to those from the East Asian sample in Exome Aggregation Consortium (ExAC; Lek et al. 2016). This observation suggests that any batch effect due to genotype calling in very low-coverage sequencing should not impact allele frequency estimates and their use in downstream analyses. Restricting analysis to variants with minor allele counts (MAC) ≥ 10 (9,888,655 variants), we found that the alternate alleles of 477,792 (4.8%), 567,731 (5.7%), and 868,251 (8.8%) variants are not seen in 1KG (phase 3), 1KG East Asians, and

1KG CHB+CHS panels, respectively (Table 2.3). We defined three minor allele frequency (MAF) categories: Common ($MAF \geq 0.05$), low frequency ($0.005 \leq MAF < 0.05$), and rare ($MAF < 0.005$). As expected, a large proportion (66-79%) of novel alleles are rare in the population, and an additional 11-17% of them are of low frequency (supplementary table S1, Supplementary Material online). We also identified $\sim 82,000$ variants with $MAF \geq 0.05$ in our data set that are not seen in the 1KG CHB+CHS populations. Even though this class of variants is likely enriched for sequencing errors, a subset of these variants were identified in limited number of East Asians included in other recent large-scale sequencing efforts [26] and the frequency estimates are highly concordant (r=0.78 and 0.90 when compared with ExAC and GNOMAD databases, Fig 2.2). Taken together, these observations suggest that our data set currently consists of the largest variant map in Han Chinese both in terms of span of the genome and in sample size.

2.2.1 Genetic Relationship with Archaic Hominin Individuals

Past studies of Neanderthal genomes have shown that the East Asians have inherited $\sim 20\%$ more Neanderthal ancestry than Europeans and that this excess ancestry may reflect a second pulse of admixture in East Asians or a dilution of Neanderthal admixture in Europeans [7, 10, 11, 27]. We largely recapitulated the relationship of a number of Neanderthal samples and Denisovan to the Han Chinese as previously reported (supplementary fig. S9, Supplementary Material online). We observed subtle differences in allele-sharing pattern and estimated Neanderthal ancestry $\sim (1.8 - -2\%)$ across China, though the difference is not significant after correcting for multiple testing (Fig 2.1).

2.2.2 Analysis of introgressed Neanderthal variants

We extracted 75,539 SNPs that were previously identified to tag Neanderthal haplotypes in East Asian individuals in the 1KG project [10], and assessed the contribution of these NIMs

to depression in our cohort consisting 5,224 cases of MDD and 5,218 controls. The allele frequencies of these NIMs are highly correlated ($r = 0.951$) between our cohort and 1KG, suggesting that the NIMs are not overt outliers from the rest of the variants in our data set in terms of data quality. We tested the association between the NIMs and depression by performing a logistic regression of depression, controlling for age and the first ten PCs, for MDD and Melancholia. Using the current sample size and sequence data, we found no association surviving the Bonferroni correction (Fig 2.1) and the QQ plots did not reveal any systematic inflation nor significant enrichment among top associated SNPs (data were not shown).

We also calculated the proportion of phenotypic variance explained by these NIMs using GCTA [28] for MDD. We used a prevalence of 7.5% to transform the heritability to the liability scale. We found that the variance explained by the NIMs is $\sim 1\%$, which is different from that reported in Simonti et al. ($\sim 2\%$) and is not significantly different from 0 ($P = 0.12$). Repeating the analysis with NIMs with MAF >0.01 as well as with no covariates did not qualitatively alter the results (Tables 2.1, 2.2). Finally, we found that the heritability explained by NIMs is not significantly different from that of a background set of SNPs chosen at random to match the NIMs by derived allele frequency decile and by Linkage Disequilibrium (LD) scores ($P > 0.4$). Our analysis may be under-powered given the smaller sample size and low coverage, but the results could suggest that the impact of Neanderthal ancestry on MDD differs between European and Han Chinese. Future investigation in larger cohorts will be informative.

2.3 Discussion

We demonstrated how the impact due to Neandertal ancestry on one type of trait, MDD and Melancholia, appears to differ between Han Chinese and Europeans. In general, these unique histories undoubtedly contributed to the variation of phenotype within Han Chinese as well

as between Han Chinese and other global populations. Therefore, a better understanding of Han Chinese history will help in conducting and interpreting future medical genetic studies within the largest ethnic group of mankind. On the other hand, our analysis also provides a starting point to investigate systematically the contribution of archaic introgression to clinically relevant disease phenotypes.

2.4 Figures

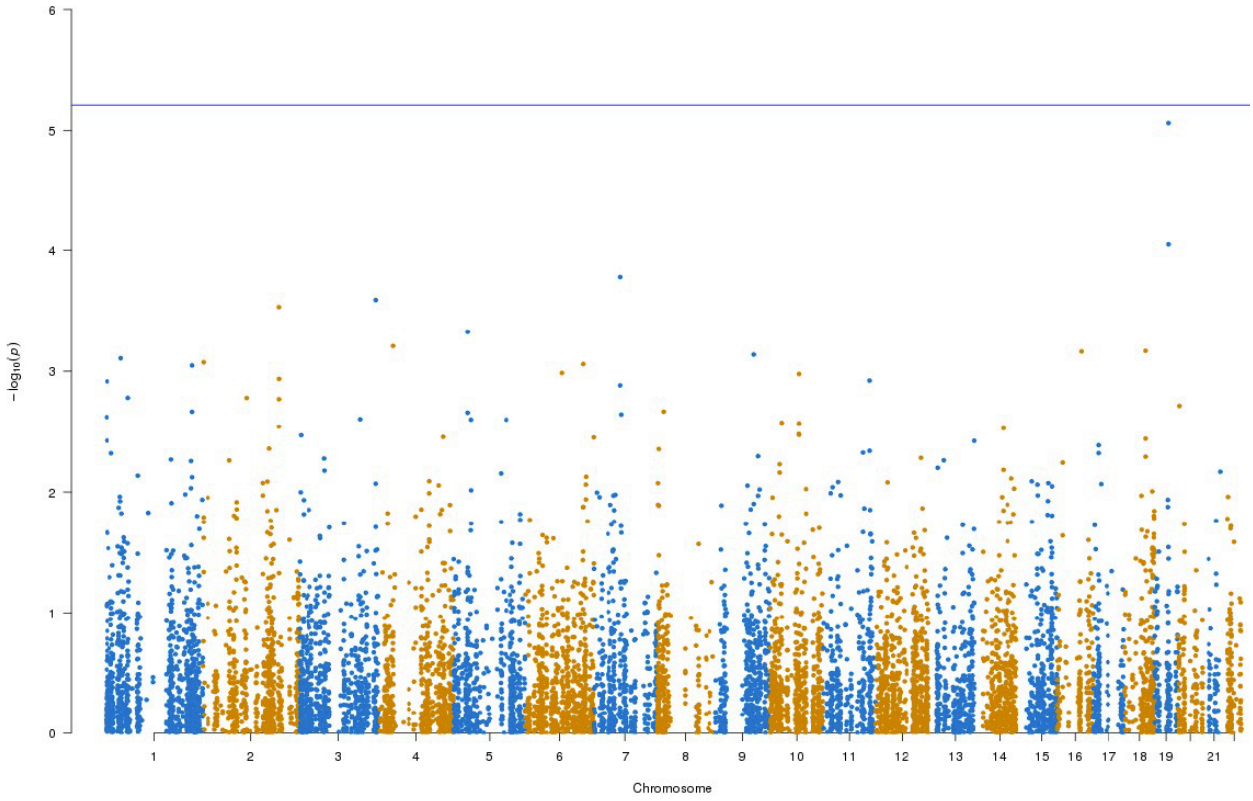

Figure 2.1: Manhattan plot for association of Neandertal-informative markers (NIMs) and MDD. We find no NIMs significantly associate with MDD. The blue line represents the Bonferroni correction threshold.

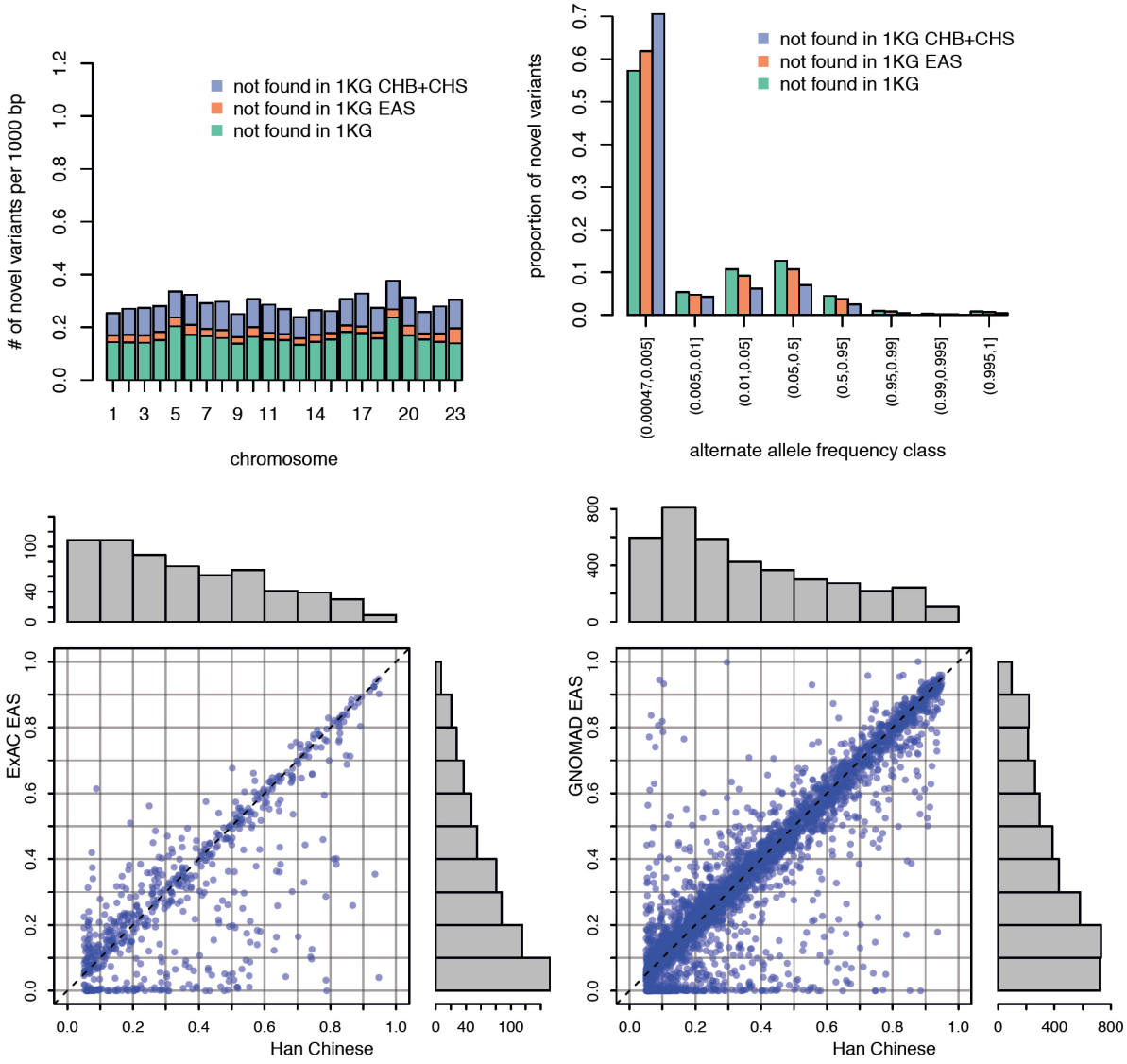

Figure 2.2 (Caption on next page.)

Figure 2.2: Variant density, frequency spectra, and allele frequency comparisons for novel variants. We defined novel variants with respect to three increasing level of inclusiveness: those with called alternative alleles that are not found among 1KG (phase3) CHB+CHS populations, 1KG EAS super-population, and all 1KG populations. Note that if the alternative allele identified in our dataset represents a previously unobserved allele in 1KG, it is considered a novel variant, even if the site is otherwise variable in 1KG. (Top Left) Density of novel alleles discovered per 1000 bp, per chromosome. Chromosome 23 signifies the X chromosome. (Top Right) Distribution of alternate allele frequency of the novel alleles. For each novel allele that was common ($MAF \geq 0.05$) in our dataset, we compared its frequency to that estimated from ExAC (Bottom Left) and GNOMAD (Bottom Right) if the same allele passed quality control filters and was called in at least 2,000 (out of 8,654) or 800 (out of 1,622) East Asian chromosomes in ExAC and GNOMAD, respectively. In total, out of 82,626 common novel alleles not found in 1KG CHB+CHS, frequency comparisons were made for 644 alleles in ExAC and 40,544 alleles in GNOMAD (random 10% of these alleles are shown here). The correlations were 0.78 and 0.90, respectively.

2.5 Tables

	H^2_{NIM}	Standard error	p-value
All NIM	0.009455	0.008175	0.1177
$MAF \geq 0.01$	0.008081	0.007664	0.1386
$MAF \geq 0.01$, no covariates	0.013443	0.007650	0.0295

Table 2.1: Heritability associated with Neanderthal-informative mutations (NIMs) for MDD

	H^2_{NIM}	Standard error	p-value
All NIM	0.010055	0.008867	0.1223
$MAF \geq 0.01$	0.006631	0.008247	0.2051
$MAF \geq 0.01$, no covariates	0.010721	0.008249	0.086

Table 2.2: Heritability associated with Neanderthal-informative mutations (NIMs) for Melancholia.

Chr	FREQ (0, 0.005]	FREQ (0.005, 0.05]	FREQ (0.05, 0.5]	MAC ≥ 10	Novel 1KG	Novel EAS	Novel CHB+ CHS
1	1,343,061	163,396	400,089	746,698	35,854	42,022	63,290
2	1,489,539	170,075	420,949	793,269	34,855	41,666	65,670
3	1,212,220	146,233	372,469	685,664	28,081	33,632	54,029
4	1,155,002	145,299	383,272	690,219	28,936	34,813	53,651
5	1,091,572	129,719	331,457	610,644	36,821	42,802	60,665
6	1,044,524	137,817	360,191	652,006	29,219	35,802	55,266
7	962,541	113,402	310,643	556,322	26,689	30,905	46,405
8	961,513	104,469	286,434	519,804	23,278	27,624	43,402
9	732,741	89,897	227,055	418,914	19,519	23,029	35,290
10	837,611	99,357	267,928	484,979	22,215	27,179	41,506
11	851,586	91,851	260,208	466,401	20,793	24,230	38,504
12	792,885	100,571	248,101	457,006	20,327	23,264	36,063
13	593,970	73,948	190,863	345,579	15,415	18,237	27,483
14	552,395	67,394	173,684	320,535	15,551	18,346	28,421
15	506,764	62,374	153,593	286,766	15,794	18,335	26,796
16	571,041	60,257	161,897	297,775	16,521	18,633	27,761
17	484,841	53,251	136,736	258,266	14,421	16,463	26,634
18	475,960	57,359	149,977	273,187	12,392	14,085	21,350
19	373,806	48,412	119,786	224,128	13,998	15,854	22,261
20	395,775	43,630	112,815	210,807	10,641	12,990	19,724
21	222,299	25,482	76,266	133,070	7,369	8,492	12,446
22	232,448	30,142	71,407	136,040	7,461	9,005	14,299
X	720,083	69,725	153,166	320,576	21,642	30,323	47,335
Tot	17,604,177	2,084,060	5,368,986	9,888,655	477,792	567,731	868,251

Table 2.3: summary of variants discovered per chromosome. Novel variants are the subset of variants with minor allele count (MAC) ≥ 10 in the current study that are also monomorphic or not reported in 1KG (phase 3), 1KG EAS, or 1KG CHB+CHS populations.

CHAPTER 3

Impact of Neanderthal DNA on UK Biobank Phenotypes

This chapter contains excerpts of a published paper Wei and Robles et al. [29] presented here with permission from the authors.

3.1 Introduction

Genomic analyses have revealed that present-day non-African human populations inherit 1-4% of their genetic ancestry from introgression with Neanderthals [5, 7]. This introgression event introduced uniquely Neanderthal variants into the ancestral out-of-Africa human gene pool, which may have helped this bottleneck population survive the new environments they encountered [30, 31, 10, 11, 14, 32]. On the other hand, the bulk of Neanderthal variants appear to have been deleterious in the modern human genetic background leading to a reduction in Neanderthal ancestry in conserved genomic regions [10, 11, 18, 19, 33]. Systematically studying these variants can provide insights into the biological differences between Neanderthals and modern humans and the evolution of human phenotypes in the 50,000 years since introgression.

In principle, studying Neanderthal-derived mutations in large cohorts of individuals measured for diverse phenotypes can help understand the biological impact of Neanderthal introgression. Previously, Dannemann and Kelso [16] showed that some Neanderthal introgressed variants are significantly associated with traits such as skin tone, hair color, and height based on Genome-Wide Association Studies (GWAS) in British samples. However, using

data from Iceland, Skov et al. [34] found that most of the significantly associated Neanderthal introgressed SNPs are in the proximity of strongly associated non-archaic variants. They suggested that these associations at Neanderthal introgressed SNPs were driven by the associations at linked non-archaic variants, indicating a limited contribution to modern human phenotypes from Neanderthal introgression. In contrast to these attempts to associate individual introgressed variants with a trait, studies have attempted to measure the aggregate contribution of introgressed Neanderthal SNPs to trait variation [15, 35]. A recent study by McArthur and colleagues [35] estimated the proportion of heritable variation that can be attributed to introgressed variants though their approach is restricted to common variants (minor allele frequency $> 5\%$) that represent a minority of introgressed variants. Despite these attempts, assessing the contribution of introgressed Neanderthal variants towards specific phenotypes remains challenging. The first challenge is that variants introgressed from Neanderthals are rare on average (due to the low proportion of Neanderthal ancestry in present-day genomes). The second challenge arises from the unique evolutionary history of introgressed Neanderthal variants resulting in distinct population genetic properties at these variants which can, in turn, confound attempts to characterize their effects. As a result, attempts to characterize the systematic impact of introgressed variants on complex phenotypes need to be rigorously assessed.

To enable analyses of genome-wide introgressed Neanderthal variants in large sample sizes, we selected and added Single Nucleotide Polymorphism (SNPs) that tag introgressed Neanderthal variants to the UKBiobank Axiom Array that was used to genotype the great majority of the approximately 500,000 individuals in the UK Biobank (UKBB) [36]. We used a previously compiled map of Neanderthal haplotypes in the 1000 Genomes European populations [10] to identify introgressed SNPs that tag these haplotypes. After removing SNPs that are well-tagged by those previously present on the UK Biobank (UKBB) array, we used a greedy algorithm to select 6,027 SNPs that tag the remaining set of introgressed SNPs at $r^2 > 0.8$ which were then added to the UKBB genotyping array to better tag Neanderthal

ancestry. These SNPs allow variants of Neanderthal ancestry to be confidently imputed and allow us to identify a list of 235,592 mutations that are likely to be Neanderthal-derived (termed Neanderthal Informative Mutations or NIMs) out of a total of 7,774,235 QC-ed SNPs in UKBB (see Sections 3.3.1, 3.7.1).

The goals of our study are threefold: 1) to estimate the contribution of NIMs to phenotypic variation in modern humans, 2) to test the null hypothesis that a NIM has the same contribution to phenotypic variation as a non-introgressed modern human SNP, and 3) to pinpoint regions of the genome at which NIMs are highly likely to modulate phenotypic variation. We develop rigorous methodology for each of these goals which we validate in simulations. We then applied these methods to 96 distinct phenotypes measured in about 300,000 unrelated white British individuals in UKBB.

3.2 Results

3.2.1 The contribution of Neanderthal introgressed variants to trait heritability

To understand the contribution of Neanderthal introgressed variants to trait variation, we aim to estimate the proportion of phenotypic variance attributed to NIMs (NIM heritability) and to test the null hypothesis that per-NIM heritability is the same as the heritability of a non-introgressed modern human (MH) SNP. We first annotated each of the 7,774,235 QC-ed SNPs in UKBB as either a NIM or a MH SNP (see Sections 3.3.2, 3.3.3). NIMs include SNPs created by mutations which likely originated in the Neanderthal lineage after the human-Neanderthal split. SNPs that are not defined as NIMs are annotated as MH SNPs which likely originated in the modern human lineage or the human-Neanderthal common ancestor.

To estimate NIM heritability, we used a recently proposed method (RHE-mc) that can partition the heritability of a phenotype measured in large samples across various genomic annotations [37]. We applied RHE-mc with genomic annotations that correspond to the ancestry of each SNP (NIM vs MH) to estimate NIM heritability (h^2_{NIM}). We also attempted

to estimate whether per-NIM heritability is the same as the per-SNP heritability of MH SNPs (Δ_{h^2}). A positive (negative) value of Δ_{h^2} indicates that, on average, a NIM makes a larger (smaller) contribution to phenotypic variation relative to a MH SNP.

To assess the accuracy of this approach, we performed simulations where NIMs are neither enriched nor depleted in heritability (true $\Delta_{h^2} = 0$). Following previous studies of the genetic architecture of complex traits [38, 39], we simulated phenotypes (across 291,273 unrelated white British individuals and 7,774,235 SNPs) with different architectures where we varied heritability, polygenicity, and how the effect size at a SNP is coupled to its population genetic properties (the minor allele frequency or MAF at the SNP and the linkage disequilibrium or LD around a SNP). We explored different forms of MAF-LD coupling where BASELINE assumes that SNPs with phenotypic effects are chosen randomly, RARE (COMMON) assumes that rare (common) variants are enriched for phenotypic effects, and HIGH (LOW) assumes that SNPs with high (low) levels of LD (as measured by the LD score [40]) are enriched for phenotypic effects (see Section 3.3.5). Estimates of h^2_{NIM} and Δ_{h^2} tend to be miscalibrated (Fig 3.1). The miscalibration is particularly severe when testing Δ_{h^2} so that a test of the null hypothesis has a false positive rate of 0.55 across all simulations (at a p-value threshold of 0.05).

To understand these observations, we compared the Minor Allele Frequencies (MAF) and LD scores at NIMs to MH SNPs. We observe that NIMs tend to have lower MAF (Fig 3.2 a) and higher LD scores compared to MH SNPs (Fig 3.2 b) (the average MAF of NIMs and MH SNPs are 3.9% and 9.9%, respectively while their average LD scores are 170.6 and 64.9). Among the QC-ed SNPs, 76.9% of NIMs have $MAF > 1\%$, and 27.7% have $MAF > 5\%$, in contrast to 61.6% and 41.6% of MH SNPs. Distinct from MH SNPs, the MAF and LD score of NIMs tend not to increase with each other (Fig 3.2 c,d).

To account for the differences in the MAF and LD scores across NIMs and MH SNPs, we applied RHE-mc with annotations corresponding to the MAF and the LD score at each SNP (in addition to the ancestry annotation that classifies SNPs as NIM vs. MH) to estimate

NIM heritability (h^2_{NIM}) and to test whether per-NIM heritability is the same as the per-SNP heritability of MH SNPs i.e., $\Delta_{h^2} = 0$ (see Section 3.3.6, 3.7.4). Our simulations show that RHE-mc with SNPs assigned to annotations that account for both MAF and LD (in addition to the ancestry annotation that classifies SNPs as NIM vs. MH) is accurate both in the estimates of h^2_{NIM} (Fig 3.1 a,c) and in testing the null hypothesis that $\Delta_{h^2} = 0$ (the false positive rate of a test of $\Delta_{h^2} = 0$ is 0.017 at a p-value threshold of 0.05; Fig 3.1 b,d). On the other hand, not accounting for either MAF or LD leads to poor calibration (Fig 3.1; we observe qualitatively similar results when estimating genome-wide SNP heritability; Fig 3.7).

We then applied RHE-mc with ancestry+MAF+LD annotations to analyze a total of 96 UKBB phenotypes that span 14 broad categories (Data S2). In all our analyses, we include the top five PCs estimated from NIMs (NIM PCs) as covariates in addition to the top twenty genetic PCs estimated from common SNPs, sex, and age (see Section 3.3.7). The inclusion of NIM PCs is intended to account for stratification at NIMs that may not be adequately corrected by including genotypic PCs estimated from common SNPs (we also report concordant results from our analyses when excluding NIM PCs; Section 3.7.3 and Fig 3.9 and Fig 3.10).

We first examined NIM heritability to find six phenotypes with significant NIM heritability (Z-score $(\hat{h^2_{NIM}} = 0) > 3$): body fat percentage, trunk fat percentage, whole body fat mass, overall health rating, gamma glutamyltransferase (a measure of liver function), and forced vital capacity (FVC) (Fig 3.5). Meta-analyzing within nine categories that contain at least four phenotypes, we find that $meta - \hat{h^2_{NIM}}$ is significantly larger than zero for anthropometry, blood biochemistry, bone densitometry, kidney, liver, and lung but not for blood pressure, eye, lipid metabolism ($p < 0.05$ accounting for the number of hypotheses tested). Meta-analyzing across all phenotypes with low correlation, we obtain overall NIM heritability estimates ($meta - \hat{h^2_{NIM}}$)= 0.1% (one-sided $p = 9.59 \times 10^{-9}$). The estimates of NIM heritability are modest as would be expected from traits that are highly polygenic and given that NIMs account for a small percentage of all SNPs in the genome (see Section 3.3.1).

We next tested whether the average heritability at a NIM is larger or smaller compared to a MH SNP ($\hat{\Delta_{h^2}} = 0$). We find seventeen phenotypes with significant evidence of depleted NIM heritability that include standing height, body mass index, and HDL cholesterol ($Z - score < -3$; Fig 3.5 b,d). Five phenotypic categories show significant NIM heritability depletion (anthropometry, blood biochemistry, blood pressure, lipid metabolism, lung) in meta-analysis. Meta-analyzing across phenotypes, we find a significant depletion in NIM heritability ($meta - \hat{\Delta_{h^2}} = -1.4 \times 10^{-3}, p = 2.55 \times 10^{-11}$). On average, we find that heritability at NIMs is reduced by about 57% relative to a modern human variant with matched MAF and LD characteristics. In contrast to the evidence for depletion in NIM heritability, we find no evidence for traits with elevated NIM heritability across the phenotypes analyzed. Despite the observation that NIMs have been primarily under purifying selection for thousands of generations [18, 33], they still make a substantial contribution to phenotypic variation in present-day humans.

Finally, we investigated the impact of controlling for MAF and LD on our findings in UKBB. Analyses that do not control for MAF and LD tend to broadly correlate with our results that control for both (Pearson's $r = 0.96, 0.68$, and 0.65 and $p < 10 - 12$ among $\hat{h^2}$, $\hat{h^2_{NIM}}$, and $\hat{\Delta_{h^2}}$). However, these analyses underestimate both heritability (Fig 3.4 a) and NIM heritability (Fig 3.4 b), resulting in apparent NIM heritability depletion ($Z - score < -3$) in 83 of the 96 phenotypes (Fig 3.4 c). While yielding qualitatively similar conclusions about the depletion in heritability at NIMs relative to MH SNPs, prior knowledge that per SNP heritability of complex traits can be MAF and LD dependent [38] coupled with our extensive simulations lead us to conclude that controlling for MAF and LD lead to more accurate results.

3.2.2 Identifying genomic regions at which introgressed variants influence phenotypes

Having documented an overall contribution of NIMs to phenotypic variation, we focus on identifying individual introgressed variants that modulate variation in complex traits. We first tested individual NIMs for association with each of 96 phenotypes (controlling for age, sex, twenty genetic PCs (estimated from common SNPs), and five NIM PCs (that account for potential stratification that is unique to NIMs). We obtained a total of 13,075 significant NIM-phenotype associations in 64 phenotypes with 8,018 unique NIMs ($p < 10^{-10}$ that accounts for the number of SNPs and phenotypes tested) from which we obtain 348 significant NIM-phenotype associations with 294 unique NIMs after clumping associated NIMs by LD (see Section 3.3.8).

A limitation of the association testing approach is the possibility that a NIM might appear to be associated with a phenotype simply due to being in LD with a non-introgressed variant [34]. We formally assessed this approach in simulations of phenotypes with diverse genetic architectures described previously where the identities of causal SNPs are known. A NIM that was found to be associated with a phenotype ($p < 10^{-10}$) was declared a true positive if the 200 kb region surrounding the associated NIM contains any NIM with a non-zero effect on the phenotype and a false positive otherwise. Averaging across all genetic architectures, the False Discovery Proportion (FDP; the fraction of false positives among the significant NIMs) of the association testing approach is around 30% (Fig 3.5). Hence, finding NIMs that are significantly associated with a phenotype does not confidently localize regions at which introgressed variants affect phenotypes.

To improve our ability to identify NIMs that truly modulate phenotype, we designed a customized pipeline that combines association testing with a fine-mapping approach that integrates over the uncertainty in the identities of causal SNPs to identify sets of NIMs that plausibly explain the association signals at a region (Fig 3.5 a)). Our pipeline starts with a

subset of significantly associated NIMs that are relatively independent ($p < 10^{-10}$) followed by the application of a statistical fine-mapping method (SuSiE [41]) within the 200kb window around each NIM signal and additional post-processing to obtain a set of NIMs that have an increased probability of being causal for a trait. We term the NIMs within this set credible NIMs while the shortest region that contains all credible NIMs in a credible set is termed the credible NIM region (see Section 3.3.9; Fig 3.5 a).

We employed the same simulations as previously described to evaluate our fine-mapping approach. The fine mapping approach yields a reduction in the FDP relative to association mapping (FDP of 15.6% on average; Fig 3.5 b) while attributing the causal effect to a few dozen NIMs within the credible NIM set (mean: 79, median: 54 NIMs across all simulations). Applying our pipeline to the set of 96 UKBB phenotypes, we identified a total of 112 credible NIM regions containing 4,303 unique credible NIMs across 47 phenotypes (Fig 3.6). The median length of credible NIM regions, 65.7kb (95% CI: [4.41kb, 469.3 kb]) is close to the expected length of Neanderthal introgressed segments [34] suggesting that the resolution of our approach is that of an introgressed LD block (Fig 3.5 c). While fine mapping generally attributes the causal signal to a subset of the tested NIMs (mean: 55.8, median: 37 NIMs across phenotypes), the degree of this reduction varies across regions likely reflecting differences in the LD among NIMs (Fig 3.5 d). We do not detect any credible NIM in 49 out of 96 phenotypes potentially due to the limited power of our procedure that aims to control the FDR (Fig 3.5 e). The sensitivity of our method is affected by both total heritability (Fig 3.5 f, Pearson's $r = 0.49, p = 3.3 \times 10^{-7}$) and NIM heritability (Fig 3.5 g, Pearson's $r = 0.36, p = 3.3 \times 10^{-4}$). A linear model that uses both total heritability and NIM heritability to predict the number of credible sets yields $r^2 = 0.29, p = 1.3 \times 10^{-5}$ and 0.015, respectively), while linear models with only total heritability or only NIM heritability result in statistically lower r^2 (0.24 and 0.13, respectively).

3.2.3 Examination of the functional impact of credible NIMs

We annotated all 4,303 unique credible NIMs using SnpEff [1] to identify a total of 26 NIMs with high (e.g., start codon loss, stop codon gain) or moderate impact (nonsynonymous variants) on genes (Fig 3.6, SI Data S7). We identified two credible NIMs, rs9427397 ($1:161,476,204C>T$) and rs60542959 (12:56,660,905 $G>T$), that have a high impact on protein sequences. The 1:161,476,204 C>T mutation, a NIM that is associated with increased gamma glutamyltransferase and aspartate aminotransferase (enzymes associated with liver function) and decreased total protein levels in blood, introduces a premature stop codon in the FCGR2A gene (Fig. S7). FCGR2A codes for a receptor in many immune cells, such as macrophages and neutrophils, and is involved in the process of phagocytosis and clearing of immune complexes. This NIM is in a region that contains SNPs shown in several GWAS linked to rheumatoid arthritis [42, 43]. The other high impact mutation, $12:56,660,905$
$G>T(rs60542959)$, results in the loss of the start codon in COQ10A, and this SNP is a credible NIM for both mean platelet volume and standing height (Fig 3.6 c) . COQ10 genes (A and B) are important in respiratory chain reactions. Deficiencies of CoQ10 (MIM 607426) have been associated with encephalomyopathy, infantile multisystemic disease; cerebellar ataxia, and pure myopathy [44]. The start codon in COQ10A is conserved among mammals with its loss having a potentially significant effect on COQ10A expression in immune cells [45].

In addition, we detect 24 credible NIMs that function as missense mutations in 19 genes. Seven out of the 19 genes are known to have immune related functions (FCGR2A, PCDHG (A8, A9, B7, C4), STAT2, and IKZF3). The NIM in STAT2 (rs2066807, 12:56,740,682 $C>G$) was the first adaptive introgression locus to be identified [30]. The STAT2 introgressed variant segregates at 0.066 frequency in the UKBB white British and leads to an I594M amino acid change in the corresponding protein. STAT2 gene and COQ10A are neighboring genes thereby providing an example of an introgressed region that potentially impacts function at

multiple genes (Fig 3.6 c).

At least seven of the 12 genes not known to be immune related have other important functions documented in the literature, such as DNA replication/damage (FANCA, CCDC8), transition in meiosis (FBXO34), detoxification/metabolism (AKR1C4), and neurological/developmental (ZNF778, ANKRD11, TBC1D32) functions. rs17134592 (10:5260682 $C > G$) is a non-synonymous mutation in AKR1C4, a gene that is involved in the metabolism of ketone-containing steroids in the liver. The NIM is associated with increased serum bilirubin levels ($p = 3e - 11$) (Fig. S7a) while also being associated with increased levels of alkaline phosphatase, insulin-like growth factor 1 (IGF1) and decreased apolipoprotein A, sex hormone binding globulin (SHBG) and triglyceride levels. rs17134592 has been identified to be a splicing QTL that is active in the liver and testis in the GTeX data (Fig. S7b). This NIM alters Leucine to Valine (L311V) which, in combination with the tightly-linked non-synonymous variant rs3829125 (S145C) in the same gene, have been shown to confer a three-to-five-fold reduction in catalytic activity of the corresponding enzyme (3-alpha hydroxysteroid dehydrogenase) in human liver [46]). Interestingly, the single amino acid change S145C did not significantly alter enzyme activity suggesting the importance of the amino acid residue at position 311 for the substrate binding of the enzyme.

3.3 Methods

3.3.1 Identification and design of SNPs that tag Neanderthal ancestry on the UK Biobank Axiom array

We chose a subset of SNPs to add to the UKBiobank Axiom array that would tag introgressed Neanderthal alleles segregating in present-day European populations.

We began with a list of 95,462 SNPs that are likely to be Neanderthal-derived from Sankararaman et al. 2014 [10]. These SNPs were identified to tag confidently inferred Neanderthal haplotypes in the European individuals identified in the 1000 Genomes Phase

1 data (Note S1).

We winnowed down this list to 43,026 SNPs after removing ones already tagged at $r^2 > 0.8$ by SNPs on the UKBiLEVE array. We then designed a greedy algorithm to capture the remaining untagged SNPs that could still be accommodated on the array (we determined the number of oligonucleotide features that would be needed to genotype each SNP as well as the total number of features available on the array through discussions with UKBiobank Axiom array design team).

Specifically, we computed LD between all pairs of Neanderthal-derived SNPs and then iteratively picked SNPs with the highest score to add to the array where the score was computed as:

$$Score_{SNPj} = \frac{\sum_{i=1}^{n}[\delta_{r^2>0.80}(i,j) Derived\ frequency_{SNPi}}{[Features\ required\ to\ genotype\ SNPj]}$$

Here, $\delta_{r^2>0.80}(i,j)$ is an indicator variable that is 1 if the squared correlation coefficient between SNPs i and j is > 0.80 and zero otherwise. Thus, SNP j is scored higher if it tags other untagged SNPs on the array. The other two terms upweight SNPs that tag other Neanderthal-derived SNPs with high derived allele frequency in Europeans and downweight SNPs by the number of oligonucleotide features required to genotype the SNP.

We iteratively chose SNPs until we obtained 6,027 SNPs (requiring 16,674 features) that fully tagged the remaining set of Neanderthal-derived SNPs. These 6,027 SNPs were then added to the UKBiobank Axiom array.

3.3.2 UK Biobank (UKBB) genotype QC

We restricted all our analyses to a set of high-quality imputed SNPs (with a hard call threshold of 0.2 and an info score greater than or equal to 0.8), which, among the 291,273 imputed genotypes of UKBB unrelated white British individuals, 1) have MAF higher than 0.001, 2) are under Hardy-Weinberg equilibrium ($p > 10^{-7}$), and 3) are confidently imputed in more than 99% of the genomes. Additionally, we excluded SNPs in the MHC region,

resulting in a total of 7,774,235 SNP which we refer to as QC-ed SNPs.

3.3.3 Identification of Neanderthal Informative Mutations

We intersected the 95,462 Neanderthal-derived SNPs identified in the 1000 Genomes European individuals with UKBB QC-ed SNPs, resulting in 70,374 mutations that we term confident Neanderthal Informative Mutations (NIM). SNPs in high Linkage Disequilibrium (LD) with this set are likely introduced through Neanderthal introgression. We expanded this set by including all QC-ed SNPs, which 1) have an r^2 of 0.99 or higher with any confident NIM, and 2) are located in the proximal neighborhood of any confident NIM (within 200kb). We term this set of SNPs as expanded NIMs. On average, 80.58% of expanded NIMs match the corresponding Altai Neanderthal allele, in contrast to 2.18% of the remaining SNPs, suggesting that these SNPs are also highly informative about Neanderthal ancestry. This treatment expands the number of NIMs in the UKBB QC-ed SNPs from 70,374 (confident NIMs) to 235,592 (expanded NIMs). We primarily use this more inclusive set of SNPs in our analyses, and refer to them as NIMs in the main results. SNPs that were not part of the expanded NIMs are termed modern human (MH) SNPs.

3.3.4 Annotating QC-ed SNPs by MAF and LD

In addition to ancestry (Neanderthal vs MH), we annotate each QC-ed SNP by its minor allele frequency (MAF) and LD. We define five MAF-based annotations by dividing all QC-ed SNPs into five equal-sized bins by their MAFs. We similarly define five LD-based annotations by dividing all QC-ed SNPs into five equal-sized bins based on their LD-score computed from 291,273 imputed unrelated white British genotypes. In-sample LD-score is computed on QC-ed genotypes using GCTA (https://cnsgenomics.com/software/gcta/) with flags "–ld-score –ld-wind 10000".

After each QC-ed SNP is annotated with three properties – ancestry (NIMs vs MH),

MAF, and LD, we use them to construct three additional sets of annotations: ancestry + MAF, ancestry + LD, and ancestry + MAF + LD annotations, by intersecting MAF annotation with ancestry annotation, LD annotation with ancestry annotation, and all three annotations, respectively. For example, for ancestry + MAF annotation, we intersect the previously defined MAF annotation with the ancestry annotation and divide SNPs into ten non-overlapping bins – from low to high MAF with Neanderthal ancestry (five bins) and from low to high MAF with modern human ancestry (five bins). Similarly, when SNPs are annotated with LD + ancestry, we have five LD bins with Neanderthal ancestry corresponding to five LD groups with modern human ancestry.

Because NIMs tend to have low MAF and high LD-score (Fig 3.2), the sizes of the annotation bins are highly uneven. To enable reliable downstream heritability analyses, we remove the annotation bins in their entirety if they include fewer than 30 SNPs. Such exceptions only occur when SNPs are annotated based on all three annotations, i.e., ancestry + MAF + LD.

3.3.5 Whole-genome simulations

We simulated phenotypes based on QC-ed UKBB genotypes with the same sample size (291,273) and number of SNPs (7,774,235). In each simulation, either 10,000 variants (mimicking moderate polygenicity) or 100,000 (mimicking high polygenicity) are sampled from the QC-ed SNPs to have causal phenotypic effects while the rest of the variants have zero effect. Causal effects and phenotypes are simulated with GCTA assuming either a high SNP heritability of 0.5 or a moderate SNP heritability of 0.2.

With the simulated causal NIM variants, true NIM heritability h^2_{NIM} can be computed as

$$h^2_{NIM} = \sum_i \beta^2_{NIM,i} / Var(y)$$

where phenotypes y are simulated based on a set of standardized genotype data with a simple

additive genetic model

$$y_j = \sum_i w_{ij}\beta_i + \epsilon_j$$

and

$$w_{ij} = (x_{ij} - 2p_i)/\sqrt{2p_i(1-p_i)}$$

with x_{ij} being the number of reference alleles for the i^{th} causal variant of the j^{th} individual and p_i being the frequency of the i^{th} causal variant, β_i is the allelic effect of the i^{th} causal variant and ϵ_j is the residual effect generated from a normal distribution with mean 0 and variance $Var(\sum_i w_{ij}\beta_i)/(1/h^2 - 1)$.

Following previous work (Evans 2018), we chose causal variants according to five different MAF and LD-dependent genetic architecture : 1) BASELINE: baseline architecture, where SNPs are randomly selected to be causal variants, 2) COMMON: common SNPs are enriched for phenotypic effects so that SNPs with MAF > 0.05 contribute 90% of causal variants while rare SNPs contribute 10%, 3) RARE: rare variants are enriched for phenotypic effects such that SNPs with MAF $<= 0.05$ contribute to 90% of causal variants while the rest contribute 10%, 4) LOW: low LD SNPs are enriched for phenotypic effects, realized as SNPs whose LD-score $<= 10$ contribute 90% of causal variants, and the rest contribute 10%, and 5) HIGH: high LD SNPs are enriched for phenotypic effects, such that SNPs with LD-score > 10 contribute 90% causal variants while the rest contribute 10%. We simulated three replicates, for each genetic architecture with two different values of SNP heritability (0.2 and 0.5) and two different levels of polygenicity (10,000 and 100,000 causal variants). Thus, we simulated a total of 60 genetic architectures.

3.3.6 Estimating NIM heritability with RHE-mc

We are interested in estimating the proportion of phenotypic variance attributed to NIMs (true NIM heritability h^2_{NIM}) and evaluating if the heritability at a NIM (per-NIM heritability) is larger or smaller than that of a background MH SNP. To this end, we used a

variance components model that partitions phenotypic variance across genomic annotations that include ancestry (NIM vs MH) as one of the input annotations.

We use RHE-mc, a method that can partition genetic variance across large sample sizes, to estimate NIM heritability [37]. For each phenotype, we run RHE-mc, in turn, with four types of input annotations: ancestry alone, ancestry + MAF, ancestry + LD, and ancestry + MAF + LD as described above. The ancestry+ MAF, ancestry + LD, and ancestry + MAF + LD annotations are intended to account for the differences in the MAF and LD properties of NIMs compared to MH SNPs.

To estimate NIM heritability, $\hat{h^2_{NIM}}$, we combine the heritability of each bin corresponding to Neanderthal ancestry:

$$\hat{h^2_{NIM}} = \sum_i \hat{h^2_{NIM,i}}$$

and the heritability estimates for any bins with modern human ancestry are used to compute the total heritability from MH. Thus, when we estimate NIM heritability from RHE-mc run with ancestry + MAF annotations, we add the heritability estimates from five bins of low to high MAF NIMs.

To compare the average heritability at a NIM to the heritability of a background MH SNP that is chosen to match the NIM in terms of MAF and LD profiles, we compute the following statistic:

$$\hat{\Delta_{h^2}} = \hat{h^2_{NIM}} - \hat{h^2_{MH}}$$

where $\hat{h^2_{MH}} = \sum_i \frac{M_{NIM,i}}{M_{MH,i}} \hat{h^2_{MH,i}}$, is the heritability of the background set matched for the MAF and LD profile of the set of NIMs. Here $M_{MH,i}$ denotes the number of MH SNPs in bin i (defined according to MAF and/or LD of the MH SNPs) while $M_{NIM,i}$ denotes the number of NIMs in the corresponding bin. A more detailed justification of this statistic is provided in Section 3.7.4.

The standard errors (s.e.) of these statistics are computed using 100 jackknife blocks using an extension of RHE-mc that takes into account the covariance among different annotations. This new version of the RHE-mc is now available at https://github.com/alipazokit/RHEmc-coeff.

3.3.7 NIM heritability and META-analysis using UKBB phenotypes

We applied RHE-mc to a total of 96 UKBB phenotypes. These phenotypes fall into 14 broader phenotypic categories (Data S1): anthropometry, autoimmune disorders, blood biochemistry, blood pressure, bone densitometry, environmental factors, eye, general medical information, glucose metabolism, kidney, lipid metabolism, liver, lung, and skin and hair. For each phenotype, we use RHE-mc to estimate the NIM heritability $\hat{h^2_{NIM}}$ and the difference between per-NIM heritability and the per-SNP heritability of MH SNPs $\hat{\Delta_{h^2}}$ while controlling for age, sex, the first 20 genetic Principal Components (PCs) estimated from common SNPs, and the first five PCs estimated from NIMs (NIM PCs). The five NIM PCs are computed using all NIMs in unrelated white British samples with ProPCA (Agrawal 2020).

To improve power to detect patterns that are shared across groups of phenotypes, we combined analyses across groups of phenotypes and across all phenotypes analyzed. We performed random effect meta-analysis on each phenotypic category containing at least four phenotypes. We assume that the phenotypes within each category i have their $\hat{h^2_{NIM}}$ drawn from the same distribution so that we can estimate the mean $(meta - h^2_{NIM})$ and variance of distribution i, based on the sampled $\hat{h^2_{NIM}}$and the s.e.$(\hat{h^2_{NIM}})$. From there, we computed the meta analysis Z-score to test if the $meta - h^2_{NIM}$is equal to zero. Similarly, we assume the phenotypes within each category i have their $\hat{\Delta_{h^2}}$ drawn from the same distribution, and compute the Z-score to test if the $meta - \Delta_{h^2}$is equal to zero. In addition to the meta-analysis within the phenotypic category, we also performed meta-analysis across all phenotypes where we used a subset of 32 phenotypes that were chosen to have low correlation

(Pearson's $r^2 \leq 0.25$).

3.3.8 Identifying individual NIMs associated with phenotype

To identify individual NIMs associated with a phenotype, we fit a linear regression model using plink 2.0 –glm and included covariates controlling for age, sex and the first 20 genotypic PCs, and first five NIM PCs. We used a stringent p-value threshold of 10-10 to correct for the number of NIMs and phenotypes tested. For each phenotype, we clumped all significant NIMs that lie within 250 kb and with an LD threshold (r^2) of 0.5 using a significance threshold for the index SNP of 10^{-10}.

3.3.9 Identifying NIMs that modulate phenotype

To assess our ability to identify introgressed variants that truly modulate a phenotype, we first tested each NIM for association with the simulated phenotype. A challenge with such an approach is the possibility that a NIM can be found to be associated with a phenotype due to being in LD with a non-introgressed variant. To exclude settings where the association signal at a NIM might be driven by LD with a non-introgressed variant, we applied a Bayesian statistical fine-mapping method (SuSiE, https://stephenslab.github.io/susie-paper/index.html) that analyzes both NIM and MH SNPs in the region surrounding an associated NIM to output a set of SNPs that can explain the association signal at the region. Furthermore, we processed these credible sets to obtain a set of credible NIMs.

We performed simulations to test the accuracy of such an approach in identifying truly causal NIMs. In particular, we first ran an association test with plink (https://www.cog-genomics.org/plink/) to identify significant NIMs (p-value $<$ 10-10). We then LD-pruned significant NIMs to get a subset of NIMs which are approximately uncorrelated with each other (using the plink flag "–indep-pairwise 100kb 1 0.99"). For each LD-pruned significant NIM, we considered all the QC-ed SNPs in its 200kb neighborhood as input to fine mapping.

We ran SuSiE with $\rho = 0.95$ and $L = 10$, such that it returns credible sets that have at least 0.95 probability to contain one causal variant and outputs at most ten credible sets for each tested region. If there are more than one credible set for a tested region, we merge them into one set. We then removed the credible sets which have 50% or more MH SNPs in their credible set. The remaining credible sets all have majority NIMs (i.e. positive results), and they are further merged together with other such regions it overlaps with, resulting in distinct regions with evidence of NIM causal effects. We termed the set of all resulting NIMs as the credible NIM set and all NIMs that lie in the credible set as credible NIMs. The region containing the credible NIM set is termed credible NIM region. If there is at least one true causal NIM within the set of credible NIMs, this credible NIM region is counted as a True Positive (TP). If there is no causal NIM in the credible NIMs, this credible NIM region is counted as a False Positive (FP).

We adopted the same approach when analyzing UKBB phenotypes while incorporating covariates. Because the SuSiE package does not directly incorporate covariates, we used regression residuals from linear regression between each UKBB phenotype and UKBB covariates (age, sex, 20 regular PCs, 5 NIM PCs), as the input phenotype to SuSiE.

3.3.10 Annotating NIMs

We annotated all unique credible NIMs using SnpEff [1] which uses Sequence Ontology (http://www.sequenceontology.org/) to assign standardized terminology for assessing sequence change and impact. We primarily focused on examining the high (e.g., start codon loss, stop codon gain) and moderate impact SNPs (nonsynonymous variants) which are coding variants that alter protein sequences.

3.4 Discussion

Our analysis demonstrates the complex influence of Neanderthal introgression on complex human phenotypes. The assessment of the overall contribution of introgressed Neanderthal alleles to phenotypic variation indicates a pattern where, taken as a group, these alleles tend to be depleted in their impact on phenotypic variation (with about a third of the studied phenotypes showing evidence of depletion). This pattern is consistent with these alleles having entered the modern human population roughly 50,000 years ago and being subject to purifying selection. Selection to purify deleterious introgressed variants, coupled with stabilizing selection on human complex traits, could result in introgressed heritability depletion such that the remaining introgressed variants in present-day humans tend to have smaller phenotypic effects compared to other modern human variants.

Nevertheless, we document a modest but significant contribution of introgressed alleles to variation in a number of phenotypes. In contrast to the previous heritability analyses by McArthur et al. [35], we did not find any NIM heritability enrichment in the 96 phenotypes. This discrepancy could be due to the different methods and NIMs used in the two studies. McArthur et al. estimate the heritability associated with common NIMs (NIMs with $MAF > 5\%$) using stratified LD Score Regression (S-LDSR) with LD scores computed from 1KG (see Section 3.7.2). Because more than 70% of NIMs have $MAF < 5\%$, this approach may not extrapolate to understand the heritability from all NIMs. An additional potential concern with analyses of NIMs is the possibility of confounding due to population structure among these introgressed variants. Typical approaches to account for population stratification based on the inclusion of principal components (PCs) may not be adequate as these PCs are computed from common SNPs on the UK Biobank genotyping array and may not account for stratification at the NIMs that tend to be rare on average (Mathieson 2012). Since our analyses work directly on individual genotype data, we are able to control for stratification specific to NIMs by including PCs estimated from NIMs in addition to PCs estimated from

common SNPs. Our analyses are broadly consistent when including NIM PCs than without (see Section 3.7.3).

Beyond characterizing aggregate effects of NIMs, we also attempted to identify individual NIMs that modulate phenotypic variation. A challenge in identifying such variants comes from the fact that NIMs tend to have lower MAF and higher LD compared to MH SNPs. Lower MAF tends to limit the power to detect a genetic effect while higher LD makes it harder to identify the causal variant. These challenges led us to design a fine mapping strategy for prioritizing causal NIMs that enables the identification of sets of NIMs that can credibly exert influence on specific phenotypes. Using this approach, we identified credible NIMs in a number of functionally important genes, including a premature stop codon in the FCGR2A gene, and a start codon loss in COQ10A. In addition, mutations in STAT2 are found to be highly pleiotropic. As many of the genes are relevant to immune, metabolic, and developmental disorders, with functions relevant to the transition to new environments, the credible NIMs reported in our study offer a starting point for detailed investigation of the biological effects of introgressed variants. Greenbaum et al. hypothesized that introgression-based transmission of alleles related to the immune system could have helped human out-of-Africa expansion in the presence of new pathogens [47]. While our results do not directly support this hypothesis, they pinpoint introgressed alleles in immune-related genes that could have and continue to modulate human phenotypes. Although we identified a number of likely causal NIMs in fine mapping, our strategy likely only picks up a small fraction of the functional NIMs suggesting that additional NIMs that are causal for specific traits remain to be discovered.

Our study has several limitations due to the current availability of data and statistical methods. First, all of our analyses focus on the white British individuals in the UKBB due to the large sample size that permits the interrogation of low-frequency NIMs and our choice of NIMs based on introgressed mutations segregating in European populations. Whole-genome sequencing data in diverse populations can potentially elucidate the impact of Neanderthal

introgression in other out-of-African populations that harbor substantial Neanderthal ancestry. Alternatively, designing arrays that have SNPs informative of archaic ancestry followed by genotype imputation could be a fruitful strategy to leverage large Biobanks to systematically explore the contribution of archaic introgression. Second, while our approach to localize credible NIMs yields a list of NIMs that are highly likely to modulate variation in a trait, our method only identifies a subset of causal variants. The design of fine mapping methods to study introgressed mutations while taking into account the ancestry (as well as better incorporating other measures such as posterior inclusion probabilities) is an important direction for future work. More broadly, the unique evolutionary history of introgressed variants motivate the development of methods tailored to their population genetic properties. While our results suggest potential evolutionary models that explain our observations of depleted heritability at introgressed alleles, evolutionary models that can comprehensively explain our observations are lacking. A major challenge is the large space of potential models that need to be explored. Nevertheless, proposing and validating such models will be an important direction for future work.

3.5 Figures

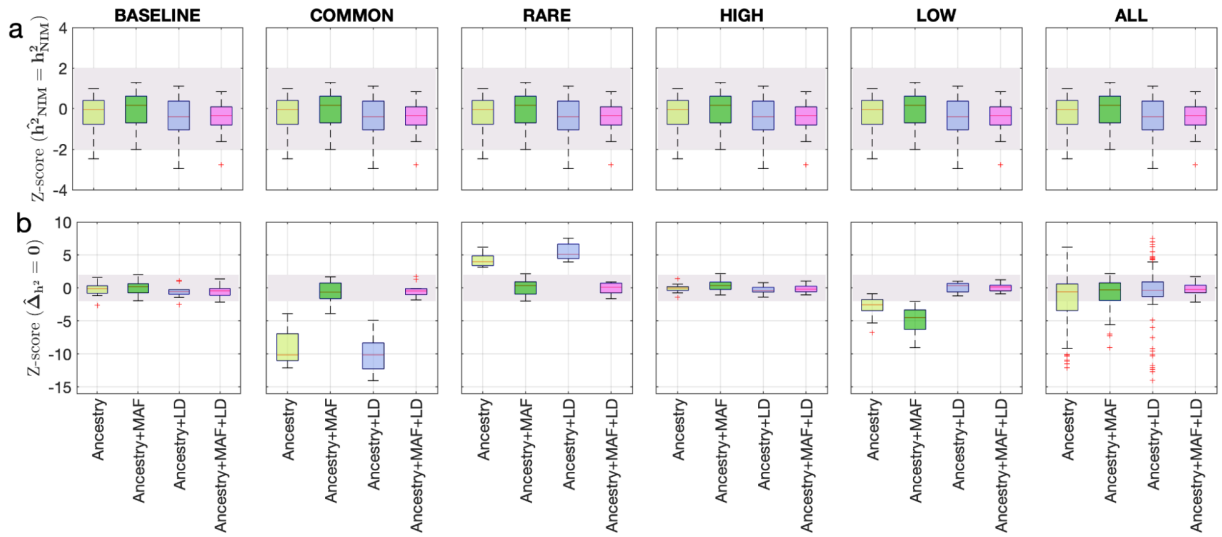

Figure 3.1: Benchmarking approaches for estimating the heritability components of Neanderthal introgression. We group simulations by relationships between minor allele frequency (MAF) and local linkage disequilibrium at a SNP on effect size (MAF-LD coupling): BASELINE, COMMON, RARE, HIGH, LOW. In each group, we perform 12 simulations with varying polygenicity and heritability (see Methods). Additionally, we combine results from all simulations together as ALL. We plot the distributions of two Z-scores (y-axis), one on each row: (a) Z-score (=) tests whether the estimated and true NIM heritability are equal, and (b) Z-score () tests whether the estimated per-NIM heritability is the same as the per-SNP heritability of MH SNPs (see Methods). In each panel, we present results from a variance components analysis method (RHE-mc) using four different input annotations: ancestry only where ancestry is either NIM or MH, ancestry + MAF, ancestry + LD, ancestry + MAF + LD. A calibrated method is expected to have Z-scores distributed around zero and within 2 (shaded region). Among all tested approaches, only RHE-mc with ancestry + MAF + LD annotations is calibrated across simulations.

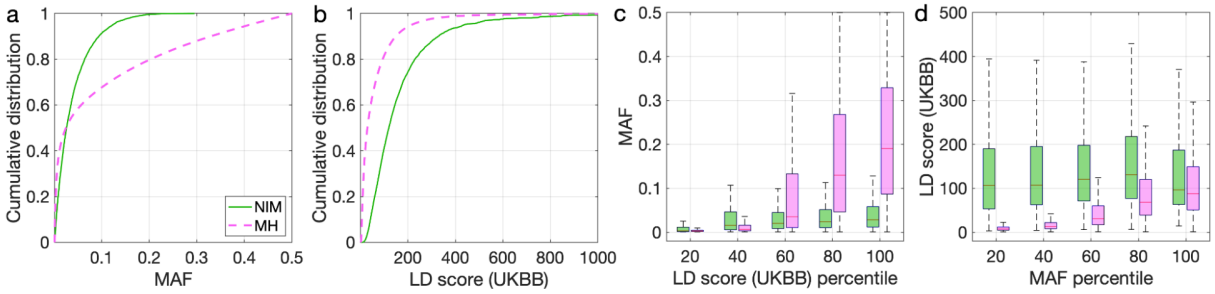

Figure 3.2: Distributions of minor allele frequency (MAF) and LD-score in NIMs and MH SNPs. Empirical cumulative distribution functions of (a) MAF and (b) LD scores of NIMs (in solid green line) and MH SNPs (in pink dashed line) estimated in the UK Biobank (UKBB). (c) Boxplots of MAFs of NIMs (on the left filled in green) and MH SNPs (on the right side filled in pink) while controlling for LD score (UKBB). (d) Boxplots of LD score (UKBB) of NIMs and MH SNPs while controlling for MAF. NIMs and MH SNPs are divided by the 20, 40, 60, 80, 100 (c) LD score (UKBB) percentile or MAF percentile (d) based on all QC-ed SNPs (7,774,235 imputed SNPs with MAF > 0.001). The lower and upper edges of a box represent the first and third quartile (qu1 and qu3), respectively; the horizontal red line inside the box indicates median (md); the whiskers extend to the most extreme values inside inner fences, $md \pm 1.5(qu3 - qu1)$.

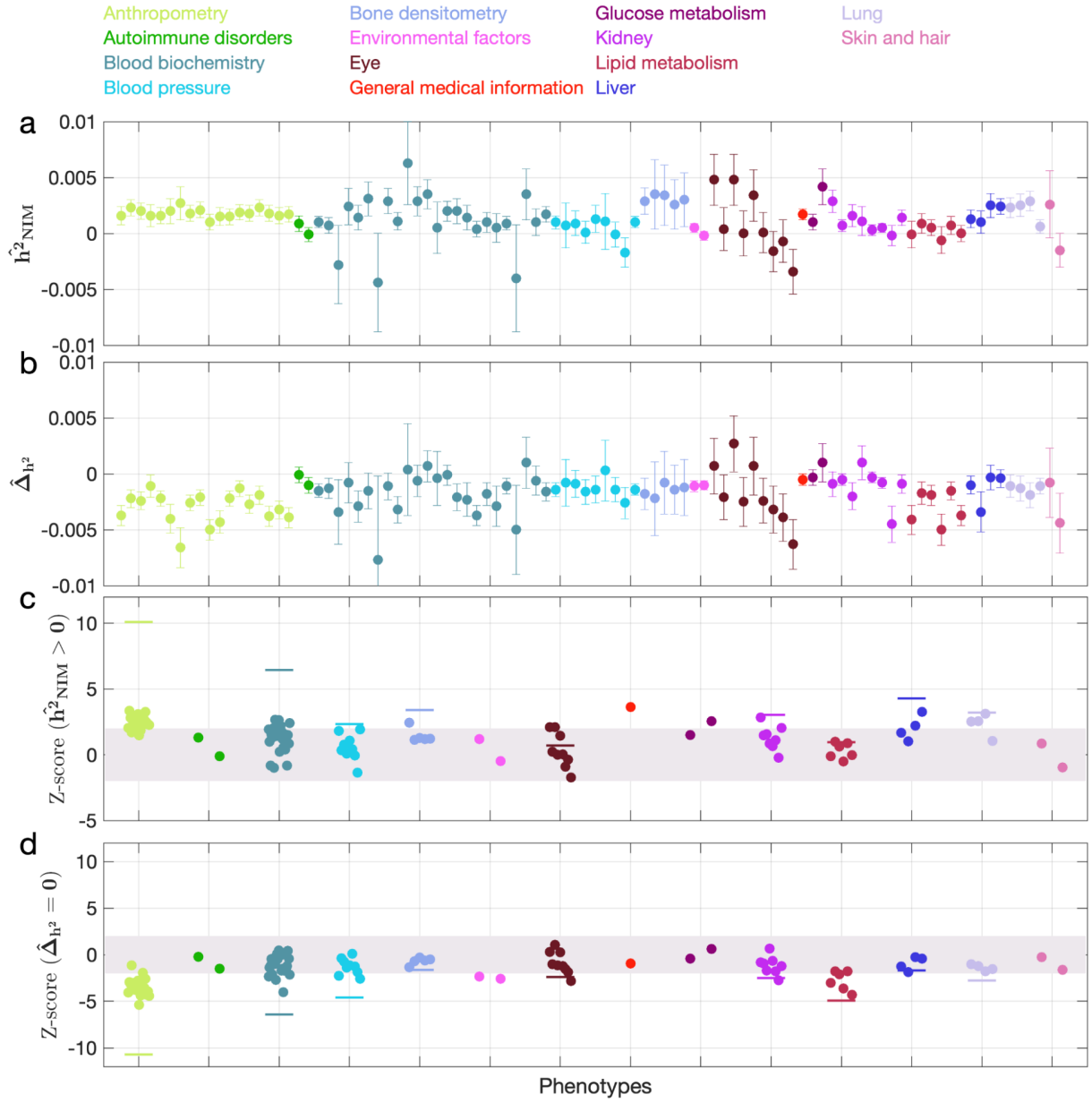

Figure 3.3(Caption on next page.)

Figure 3.3: NIM heritability in UKBB phenotypes. (a) Estimates of NIM heritability ($\hat{h^2_{NIM}}$) and (c) the Z-score of $\hat{h^2_{NIM}}$ (testing the hypothesis that NIM heritability is positive) for each UKBB phenotype. Analogously, (b) estimates of Δ_{h^2} and Z-score (d) of Δ_{h^2} (testing the hypothesis that per-NIM heritability is equal to per-SNP heritability at MH SNPs after controlling for MAF and LD). Phenotypic categories are shown in alphabetical order and listed on the top of panel (a) in the same color and alphabetical order (from top to bottom, and left to right) as they are in the figure. The estimate for each phenotype is shown as one colored dot, on the x-axis based on its phenotypic category, and on the y-axes based on its Z-score ($(\hat{h^2_{NIM}} = 0)$ and Z-score ($\Delta_{h^2} = 0$), for panels (c) and (d) respectively. For each phenotypic category with at least four phenotypes, their Z-scores from random effect meta-analysis are plotted with the flat colored lines (see Methods). The color shades cover Z-scores around zero and within ± 2.

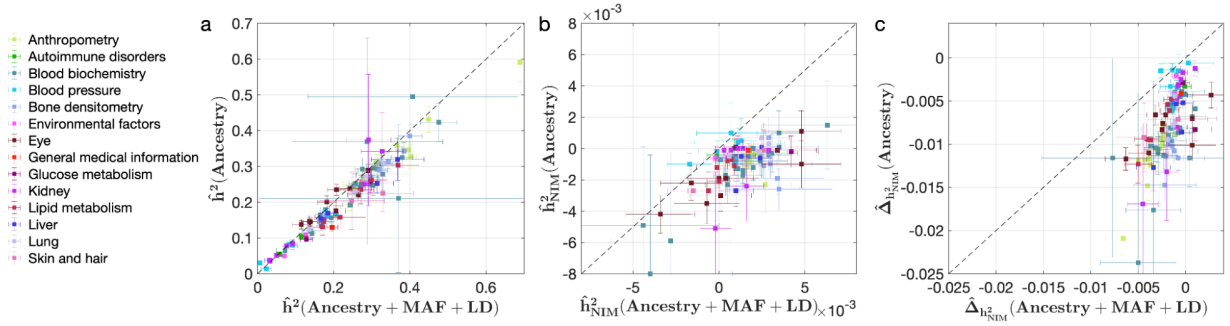

Figure 3.4: Comparing heritability analyses with and without controlling for MAF and LD in UKBB phenotypes. Each phenotype is shown with one dot colored by the phenotypic category it belongs to, on the y-axis based on its point estimate and standard error (estimated by RHE-mc with Ancestry annotation) and on the x-axis based on its point estimate and standard error (estimated by RHE-mc with ancestry + MAF + LD annotation). Estimates shown are (a) total heritability $\hat{h^2}$, (b) NIM heritability $\hat{h^2_{NIM}}$, and (c) the difference between per-NIM heritability and matched MH SNPs heritability Δ_{h^2}. Not controlling for MAF and LD leads to underestimation of NIM heritability, which leads to false positives when testing whether heritability at a NIM is elevated or depleted relative to a MH SNP.

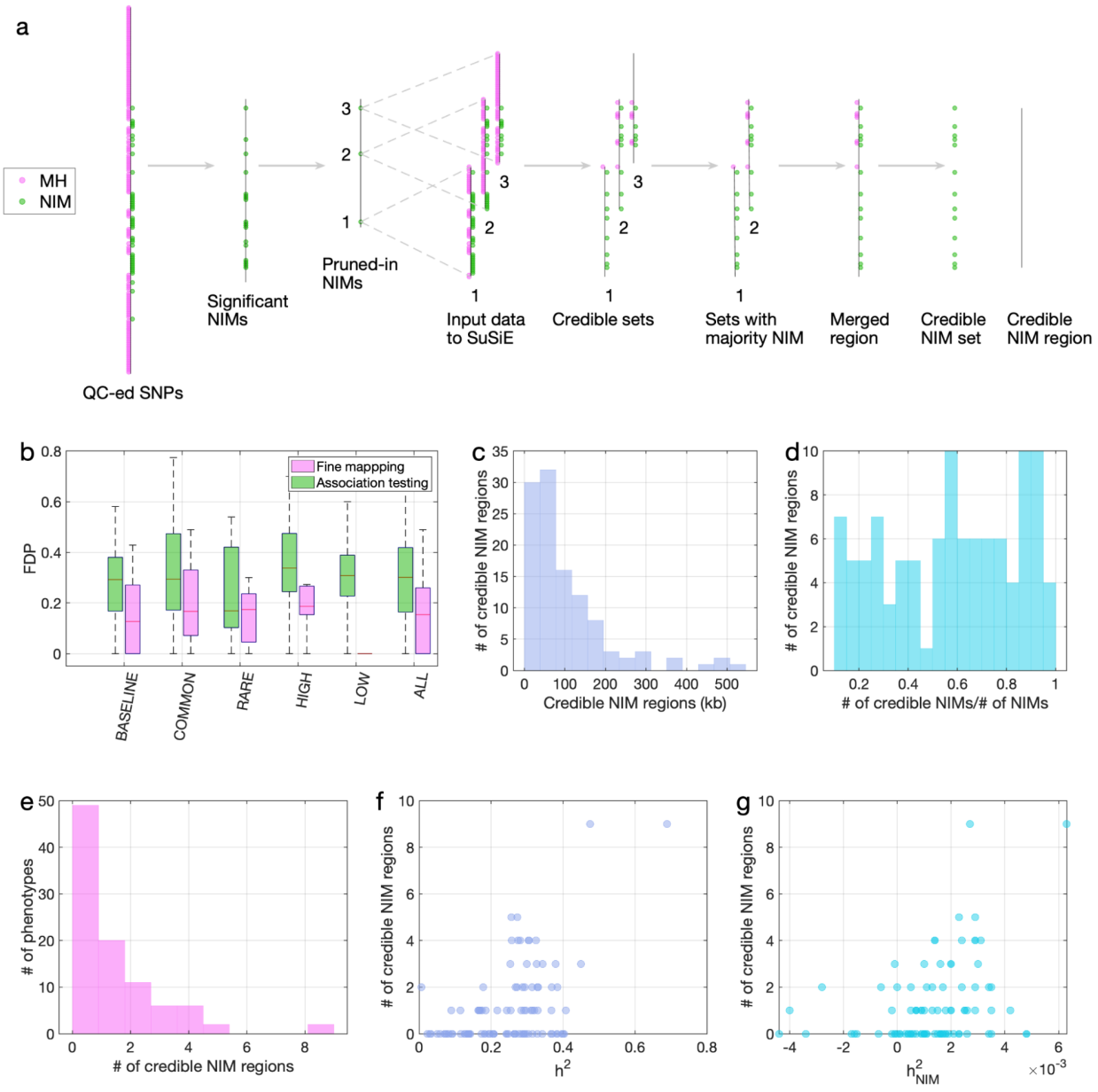

Figure 3.5 (Caption on next page.)

Figure 3.5: Fine mapping of NIMs in simulations and the UKBB. (a) Fine mapping pipeline to identify NIMs that aims to identify genomic regions at which NIMs are likely to modulate phenotypic variation (credible NIM regions). (b) Comparison of approaches for identifying credible NIM regions. For each simulation, False Discovery Proportion (FDP) is computed for association testing compared to our pipeline (combining association testing and fine-mapping). The distributions of the FDP are shown across genetic architectures (summarized across groupings of coupling of effect size, MAF and LD) and summarized across architectures (ALL). Our approach to identifying credible NIMs decreases FDP in all studied architectures (the LOW LD setting has a median and quartiles of zero across replicates). (c) The distribution of the length of credible NIM regions across 96 UKBB phenotypes. (d) Distribution of the ratio between the number of credible NIMs and number of tested NIMs (in the example of panel (a), the number of tested NIMs is the union of NIMs in input to the fine-mapping software (SuSiE) 1 and 2). This figure shows that our fine mapping approach is effective in prioritizing NIMs that affect phenotype. (e) The distribution of the number of credible NIM regions among phenotypes. The number of credible NIM regions is positively correlated with (f) heritability (g) NIM heritability.

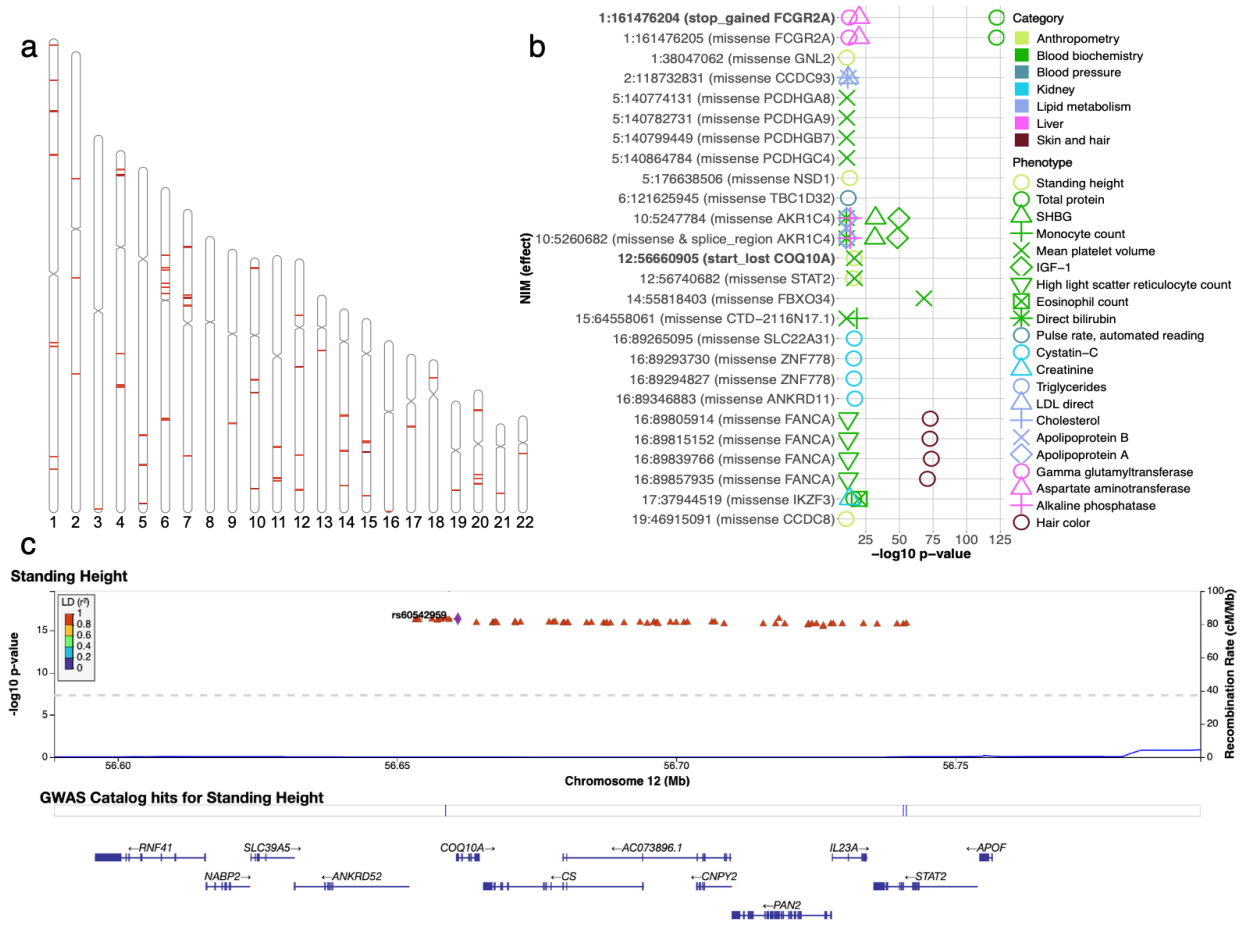

Figure 3.6 (Caption on next page.)

Figure 3.6: Analysis of credible NIMs. (a) Distribution of credible NIMs across the genome (b) High and moderate impact credible NIMs annotated by SnpEff software ([1]). A total of 26 credible NIMs have high (marked in bold) or moderate impact effects on nearby genes (chromosome number and hg19 coordinates). The effect of the SNP and the gene name are displayed. This plot shows significant associations of these NIMs with specific phenotypes (color denotes the phenotype category). (c) Plot of 300kb region surrounding rs60542959 (marked in black diamond; hg19 coordinates), a credible NIM for standing height that results in loss of the start codon in COQ10A. The plot displays other significantly associated NIMs in the region along with their LD (r^2) to rs60542969 in 1000 Genomes Europeans ([2]).

3.6 Supplement

3.6.1 Supplementary Figures

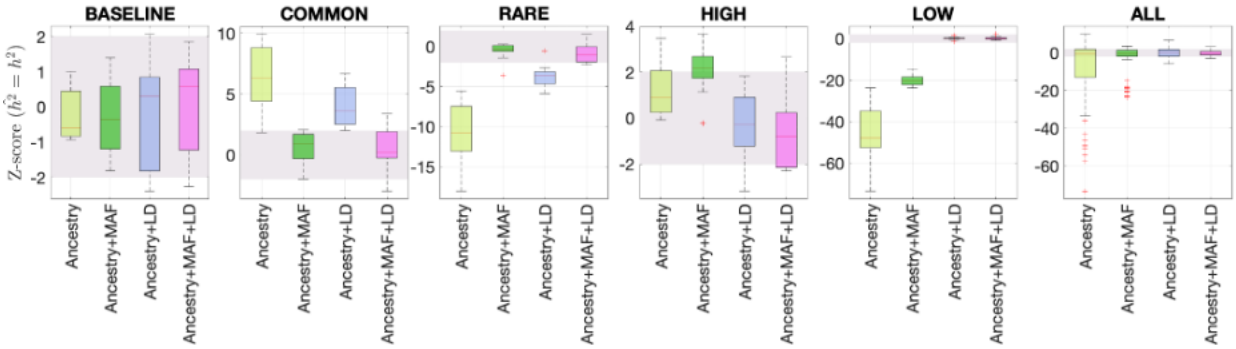

Figure 3.7: Benchmarking different methods for estimating the total SNP heritability. We grouped the simulations by the five different MAF-LD coupling: BASELINE, COMMON, RARE, HIGH, LOW, as labeled on top of each column. In each group, there are 12 simulations with different levels of polygenicity and heritability (see Methods). Additionally, we combined simulations from all five architectures together as ALL for the sixth column. On the y-axis, Z-score ($\hat{h^2} = h^2$) tests whether the estimated and simulated total heritability are equal. In each panel, the results from RHE-mc with four different annotations, ancestry only, ancestry + MAF, ancestry + LD, ancestry + MAF + LD are shown on the x-axis. A calibrated method is expected to have all Z-scores distributed around zero and within 2 (shaded region). Among all tested methods, only RHE-mc with annotation ancestry + MAF + LD satisfies this criterion.

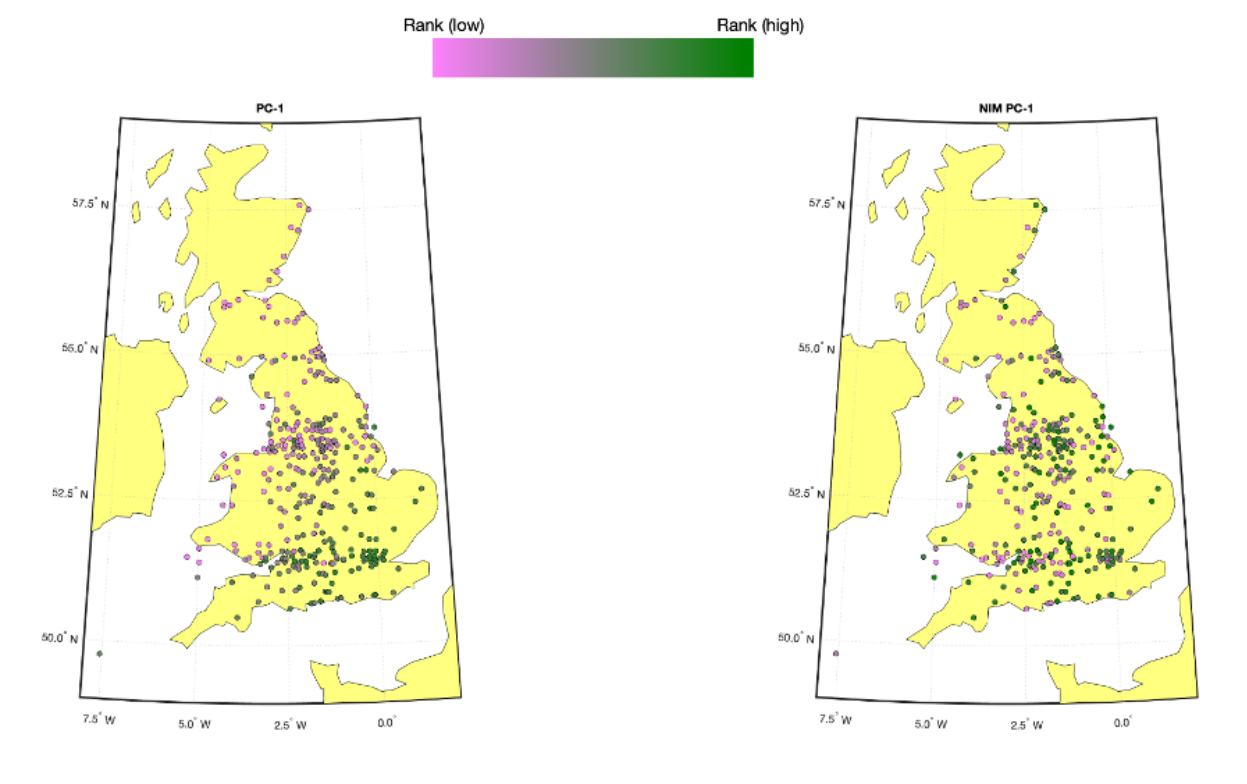

Figure 3.8: Population structure within white British samples. PC-1 from the whole genome genotypes (released by UKBB) is shown on the left, and NIM PC-1 is shown on the right. We used a 20-by-20 grid along the latitude and longitude, dividing the map into 400 colonies. We then computed the average PC projection as well as the median longitude and latitude among the individuals belonging to each colony, if there are at least 10 individuals in a colony. Each color-filled circle with a 5 kilometer radius represents one colony on the map. To maximize the visible differences, we sorted the colonies by their PC values and used the rank to determine the color of the colony. Compared to NIM PC-1, PC-1 shows a much stronger correlation with geographical location.

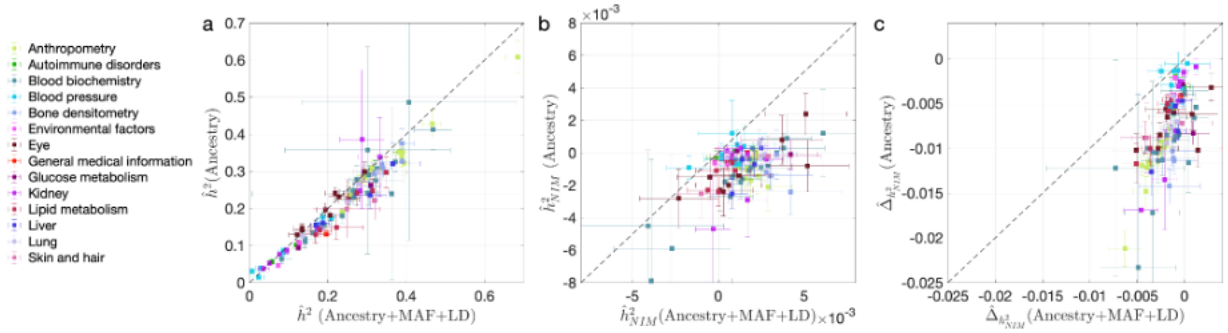

Figure 3.9: Comparing heritability estimates from RHE-mc without controlling for NIM PCs with Ancestry+MAF+LD annotation and RHE-mc with Ancestry annotation in UKBB phenotypes. This figure is plotted in the same way as Fig. 4. The trend that not controlling for MAF and LD lead to underestimation of (a) total heritability $\hat{h^2}$, (b) NIM heritability $\hat{h^2_{NIM}}$, and stronger NIM heritability depletion (c) $\hat{\delta_{h^2}}$ is also apparent when NIM PCs are not controlled for.

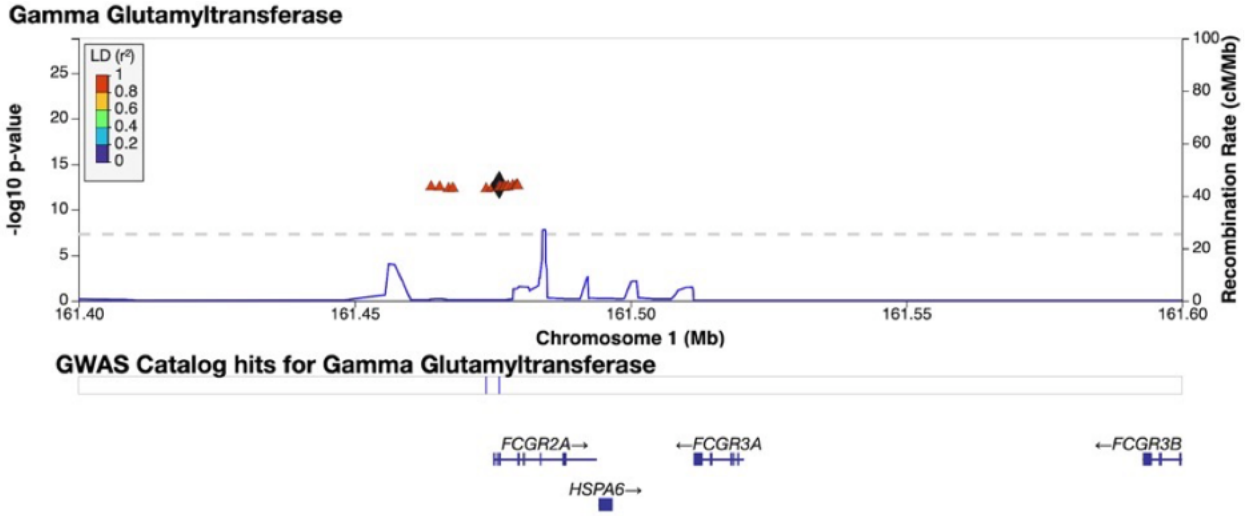

Figure 3.10: Credible NIM in the FCGR2A gene associated with gamma-glutamyl transferase levels. Plot of 200kb region surrounding rs9427397 (marked in black diamond; hg19 coordinates), a credible NIM in FCGR2A that introduces a premature stop codon and is associated with increased levels of gamma glutamyltransferase (while also associated with increased levels of aspartate aminotransferase and decreased total protein). The plot displays other NIMs in the region along with their LD (r^2) to rs9427397 computed in 1000 Genomes Europeans.

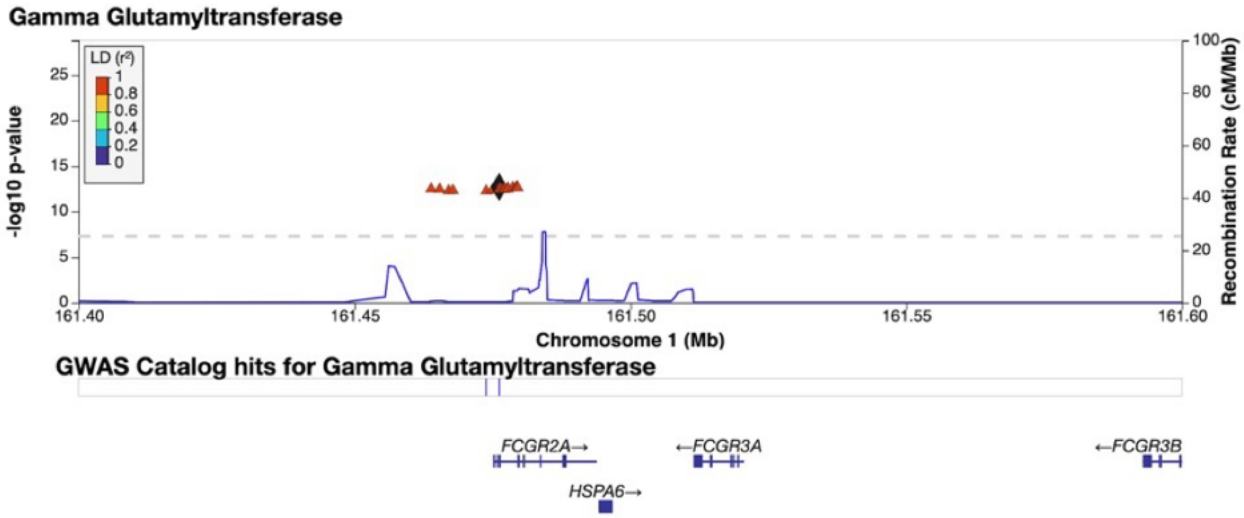

Figure 3.11: Credible NIM in the AKR1C4 gene is associated with bilirubin levels. (a) Plot of 300kb region surrounding rs17134592 (marked in black diamond; hg19 coordinates), a non-synonymous NIM in AKR1C4, that is associated with increased serum bilirubin levels. The plot displays other NIMs in the region along with their LD (r^2) to rs17134592. (b) rs17134592 is a splicing QTL in liver (AKR1C8P) and testis (AKR1C4) identified in GTEx v8.

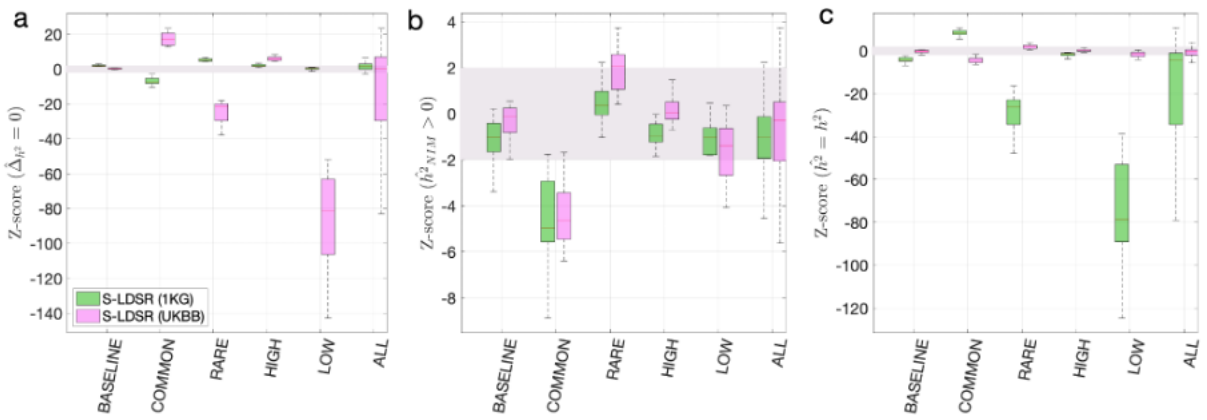

Figure 3.12: Appendix 2 - Figure 1. Benchmarking stratified LDSC regression (S-LDSR) with in-sample and out-of-sample LD scores. We group the simulations by the MAF-LD coupling: BASELINE, COMMON, RARE, HIGH, LOW, and ALL, as labeled on the x-axis. We plot the distributions of three Z-scores (y-axis), one on each panel: (a) Z-score ($\hat{\delta_{h^2}} = 0$) tests whether the estimated NIM heritability is different from the matched MH heritability, (b) Z-score ($\hat{h^2_{NIM}} = h^2_{NIM}$) tests whether the estimated and expected NIM heritability are equal, and (c) Z-score ($\hat{h^2} = h^2$) tests whether the estimated and simulated total heritability are equal. In each panel, S-LDSR with the out-of-sample LD score from 1000 Genomes (1KG) is shown in green and S-LDSR with in-sample LD score from UKBB in pink. In S-LDSR, only ancestry annotation is used. The Z-scores within 2 are color shaded. S-LDSR (1KG) is not calibrated even for BASELINE architecture.

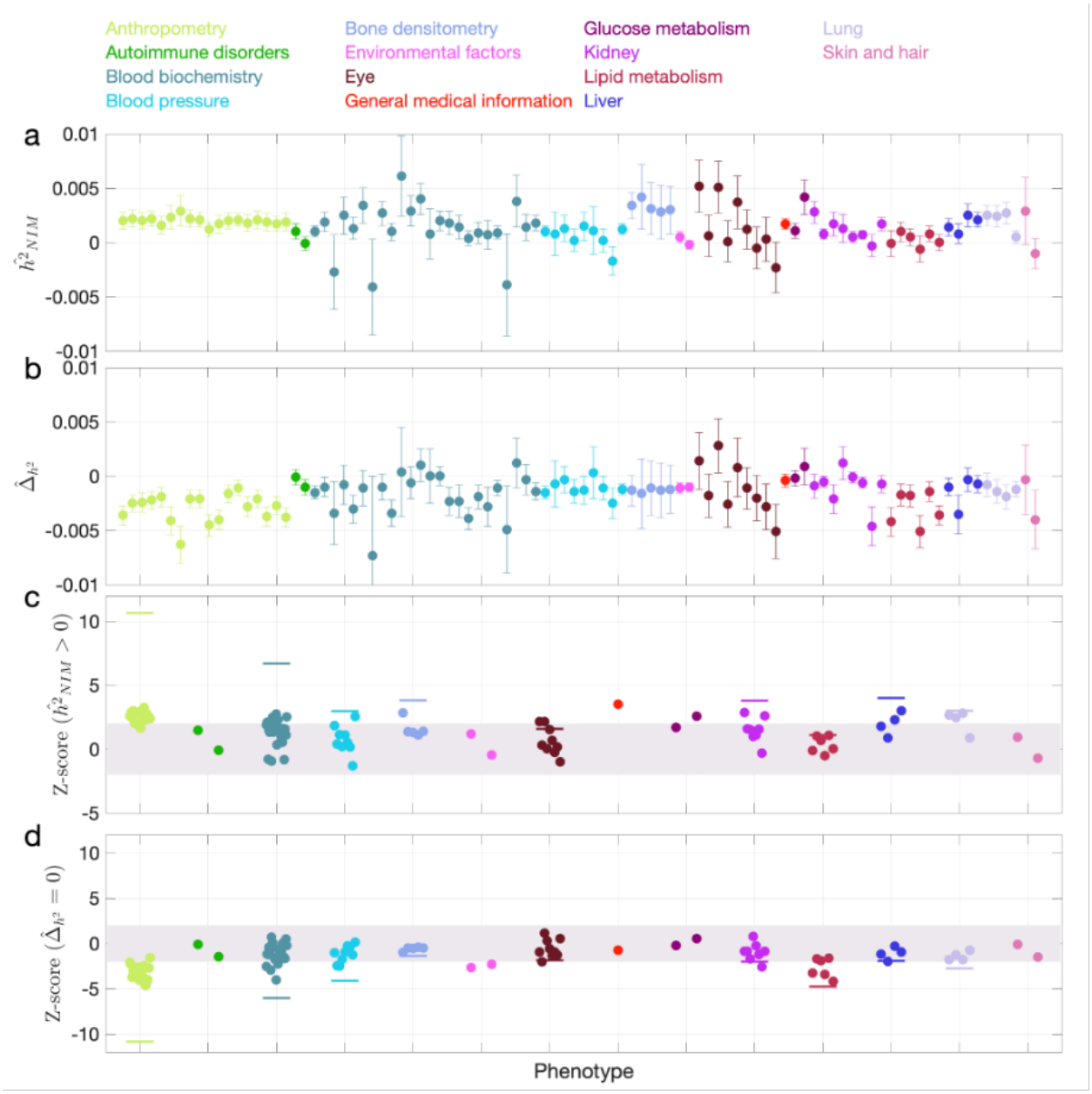

Figure 3.13: Appendix 3 - Figure 1. NIM heritability in the 96 UKBB phenotypes without controlling for NIM PCs. This figure is plotted in the same way as Fig. 3. Heritability estimates are largely similar, but fewer phenotypes are significant. Three phenotypes have significant positive NIM heritability (Z-score $(\hat{\delta_{h^2_{NIM}}} = 0) > 3$): overall health rating, waist-hip-ratio, and gamma glutamyltransferase. Fourteen phenotypes (standing height, sitting height, weight, body fat percentage, whole body fat-free mass, whole body water mass, trunk fat-free mass, trunk predicted mass, basal metabolic rate, RBC count, apolipoprotein A, HDL cholesterol, triglycerides) are significantly depleted for NIM heritability (Z-score $<$ -3).

3.6.2 Supplementary Data

Supplementary data include 10 items, Data S1-S10 are available at:
https://github.com/AprilWei001/NIM

1. Data S1. UKBB phenotype annotation.
2. Data S2: RHE-mc results in simulated data.
3. Data S3: RHE-mc results with Ancestry+MAF+LD annotations and NIM PCs included in covariates applied to 96 UKBB phenotypes
4. Data S4: RHE-mc results with Ancestry only annotation and NIM PCs included in covariates applied to 96 UKBB phenotypes
5. Data S5: RHE-mc results with Ancestry+MAF+LD annotation without NIM PC in covariates applied to 96 UKBB phenotypes
6. Data S6: Fine mapping FDP in simulated data.
7. Data S7: 112 credible NIM sets and credible NIMs
8. Data S8: SnpEff annotation of all unique credible NIMs
9. Data S9. Stratified LD score regression results in simulated data using LD score from 1KG
10. Data S10. Stratified LD score regression results in simulated data using LD score from UKBB

3.7 Appendix

3.7.1 Appendix 1: Identification of SNPs that tag Neanderthal ancestry on the UK Biobank Axiom array

Starting with the confidently inferred Neanderthal haplotypes identified in Sankararaman et al. 2014 [10], we identified whether a SNP segregating in a target modern human population owes its origin to the Neanderthal gene flow event as follows: 1. We identified sets of haplotypes that are confidently labeled as Neanderthal, N by the Conditional Random Field (CRF) method proposed in Sankararaman et al. 2014 scanning for runs of consecutive SNPs with marginal probability of Neanderthal ancestry ≥ 0.90. We also identified sets of haplotypes that are confidently labeled as non-Neanderthal, MH by scanning for SNPs with marginal probability ≥ 0.1). We also required the Neanderthal haplotype to be at least 0.02 cM long. 2. For each SNP called in the 1000 Genomes dataset, we required that none of the derived alleles at this SNP falls on one of the modern human haplotypes in the set MH and all of the haplotypes in N carry the derived allele. This procedure allows for some false negatives in the predictions of the CRF. 3. We ran this procedure on the combined calls from the European ancestry populations (CEU, GBR, FIN, IBS and TSI) in the 1000 Genomes Project.

This procedure yielded a total of 95,462 SNPs that are likely to be Neanderthal-derived. We winnowed down this list to 43,026 SNPs after removing ones already tagged at $r^2 > 0.8$ by SNPs on the UKBiLEVE array. We then designed a greedy algorithm to capture the remaining untagged SNPs that could still be accommodated on the array (we determined the number of oligonucleotide features that would be needed to genotype each SNP as well as the total number of features available on the array through discussions with UKBiobank Axiom array design team).

Specifically, we computed LD between all pairs of Neanderthal-derived SNPs and then iteratively picked SNPs with the highest score to add to the array where the score was

computed as:

$$Score_{SNPj} = \frac{\sum_{i=1}^{n}[\delta_{r^2>0.80}(i,j)][\text{Derived frequency}_{SNPi}]}{\text{Features required to genotype SNP j}}$$

Here $\delta_{r^2>0.80}(i,j)$ is an indicator variable that is 1 if the squared correlation coefficient between SNPs i and j is > 0.80 and zero otherwise. Thus, SNP j is scored higher if it tags other untagged SNPs on the array. The other two terms upweight SNPs that tag other SNPs with high derived allele frequency in Europeans and downweight SNPs by the number of oligonucleotide features required to genotype it.

We iteratively chose SNPs until we obtained 6,027 SNPs (requiring 16,674 features) that fully tagged the remaining set of Neanderthal-derived SNPs. These 6,027 SNPs were then added to the UKBiobank Axiom array.

3.7.2 Appendix 2: Estimating NIM heritability with partitioned LD-score regression

We considered two candidate methods for estimating the NIM heritability in large datasets and testing the related hypotheses to NIM heritability, S-LDSR [40] and RHE-mc (see Main text) [37]. S-LDSR can speedily estimate partitioned heritability given GWAS statistics and LD scores without any individual-level data. S-LDSR can be used with either in-sample LD scores (i.e., computed from the same data as for GWAS) or out-of sample LD scores (i.e., computed from an external and often much smaller data set). Out-of-sample LD scores from 1000 Genomes (1KG) is often used in S-LDSR (McArthur 2021, Koller 2021) because 1) it is computationally much cheaper to compute than using the GWAS cohorts, and 2) individual-level data from GWAS cohorts are not always accessible; despite that, S-LDSR with in-sample LD scores is more accurate in theory.

Previous studies [35] used S-LDSR to estimate the heritability from archaic ancestries. They computed the stratified LD scores using the 1000 Genomes (1KG) EUR and EAS

samples and performed LD score regression against the GWAS statistics from a different cohort. If the ancestry from 1KG samples does not match well with the GWAS cohort, it could lead to biased heritability estimates. Additionally, the LD score distribution and MAF distributions of NIMs are very different from the distributions of MH SNPs (Fig 3.2), which might also affect the heritability estimates if not taken into account. Finally, LD score regression is restricted to a subset of SNPs (typically with MAF $> 5\%$) which substantially reduces the number of NIMs analyzed. Here, we benchmarked S-LDSR on the simulated data with both out-of-sample LD scores from 1KG and the in-sample LD scores from all UKBB QC-ed data, stratified by ancestry (NIM vs MH).

First, we used the aforementioned simulations to evaluate the partitioned LD score regression in estimating NIM heritability. We downloaded the 1KG EUR data (from **this site**) that is typically used for LD score regression. There are 9,997,231 SNPs in the data, and 5,789,471 of them are shared with the UKBB QC-ed SNPs. Out of the 235,592 expanded NIMs defined in UKBB QC-ed data, 210,962 are present in 1KG EUR, and we refer to these SNPs as the 1KG NIMs. We defined the 9,786,269 SNPs in 1KG EUR that are not expanded NIMs as 1KG MH SNPs. We then computed the stratified LD score using GCTA software with flags *–ld-score –ld-wind 10000* for all the 1KG SNPs, all the 1KG NIMs, and all the 1KG MH SNPs. From here, we computed the stratified LD score with $ldsc_{SNP} = ldsc_{NIM} + ldsc_{MH}$, such that each SNP has two stratified LD scores, one due to its LD with NIMs and another due to its LD with MH SNPs. We then intersected the 1KG SNPs with UKBB QC-ed SNPs, and used the shared 5,789,471 SNPs to perform stratified LD score regression. The S-LDSR is then performed with all SNPs that overlap between 1KG and UKBB. For each simulation, we ran S-LDSR to estimate $h^2_{NIM}, h^2_{MH}, \Delta_{h^2}$, and their standard errors from 200 jackknife blocks. We found that the results from using out-of-sample LD are biased even when heritability does not depend on MAF and LD (i.e., BASELINE) (Fig 3.12).

As a comparison, we computed the in-sample stratified LD score using the UKBB QC-ed

data and applied S-LDSR with these in-sample LD scores. In contrast to the previous results, the results are well calibrated for BASELINE, suggesting that the previous biases observed with BASELINE are due to the disagreement between the out-of-sample LD score and the in-sample LD score (Fig 3.12). Not surprisingly, the results for MAF and LD-dependent architectures are still biased, as these factors are not taken into account. We caution that our simulations are based on UKBB QC-ed SNPs, where non QC-ed SNPs do not have an impact on the simulated phenotypes. This setting will favor S-LDSR based on UKBB QC-ed SNPs more than in actual settings, and disfavor S-LDSR 1KG more than in actual settings. It is possible that in reality, the biases with in-sample LD score will become larger, and the biases with out-of-sample LD score will become smaller. Nonetheless, because it is often expensive to compute in-sample LD scores, the accuracy will largely depend on how well the external panel resembles the GWAS cohort.

The out-of-sample LD score could be particularly biased for low MAF SNPs, hence S-LDSR recommends not using annotations with fewer than 5% of SNPs as best practice. This practice will necessarily exclude more than 70% of NIMs and about half of the MH SNPs, and the heritability estimates from high MAF SNPs may not extrapolate to low MAF SNPs. Therefore, S-LDSR, under the best practice, is not suitable for studying Neanderthal introgressed variants.

3.7.3 Appendix 3: The impact of inclusion of NIM PCs on NIM heritability estimates

We computed the first five NIM PCs using all NIMs in unrelated white British samples with ProPCA (Agrawal 2020). Compared to the regular genetic PCs (estimated from common SNPs), NIM PCs are only weakly correlated with birth GPS locations (Fig 3.8), consistent with the fact that Neanderthal introgression occurred soon after the out-of-African migration before population expansion.

When NIM PCs were not being controlled for (with remaining regular covariates still

used), we found three phenotypes with significant NIM heritability (Z-score ($\hat{h^2_{NIM}} = 0$) > 3): overall health rating, waist-hip-ratio (WHR), and gamma glutamyltransferase (a measure of liver function). We also combined phenotypes into broader phenotypic categories and performed random effect meta-analysis on the nine categories that contain at least four phenotypes (see Methods). We found that $meta - \hat{h^2_{NIM}}$ is significantly larger than zero (Z-score > 2.53 for one-tail $p = 0.05$ level) for all but two categories (eye, lipid metabolism), meaning that NIMs heritability is generally nonzero (Fig 3.13 a,c). We then tested whether NIM heritability is larger or smaller compared to MH SNPs ($\hat{\Delta_{h^2}} = 0$). Fourteen phenotypes (standing height, sitting height, weight, body fat percentage, whole body fat-free mass, whole body water mass, trunk fat-free mass, trunk predicted mass, basal metabolic rate, RBC count, apolipoprotein A, HDL cholesterol, triglycerides) remain significantly depleted (Z-score < -3), among which eight are anthropometric phenotypes, and three are related to lipid metabolism. This is in contrast to seventeen phenotypes when NIM PCs are controlled for (body mass index, hip circumference, waist circumference, standing height, sitting height, weight, whole body fat-free mass, whole body water mass, whole body impedance, trunk fat mass, trunk fat-free mass, trunk predicted mass, basal metabolic rate, RBC count, apolipoprotein A, HDL cholesterol, triglycerides).

Four phenotypic categories show significant NIM heritability depletion (anthropometry, blood biochemistry, blood pressure, lipid metabolism), and five are not significantly different with meta analysis (Appendix 3 - Figure 1bd). In contrast to the evidence for depletion in NIM heritability, we found no evidence for traits with elevated NIM heritability even when excluding NIM PCs (Fig 3.13 d).

3.7.4 Appendix 4: Statistic to compare per-NIM heritability to per-SNP heritability at a set of background MH SNPs

In this note, we provide additional intuition behind our statistic to compare difference between the heritability at a NIM (per-NIM heritability) and the per-SNP heritability of a

background set of MH SNPs:

$$\hat{\Delta_{h^2}} = \hat{h^2_{NIM}} - \hat{h^2_{MH}}$$

Let $\sigma^2_{a,i} = \frac{h^2_{a,i}}{M_{a,i}}$ where $a \in \{NIM, MH\}, i$ denotes one of the annotations (MAF,LD), $h^2_{a,i}$ denotes the heritability attributed to annotation (a,i) , and $M_{a,i}$ denotes the number of SNPs in annotation (a,i). Thus $\sigma^2_{a,i}$ denotes the per-SNP heritability associated with annotation (a,i).

The per-SNP heritability associated with NIMs is given by:

$$\sigma^2_{NIM} = \sum_i \frac{\sigma^2_{NIM,i} M_{NIM,i}}{M_{NIM}} = \sum_i \frac{h^2_{NIM,i}}{M_{NIM}} = \frac{1}{M_{NIM}} \sum_i h^2_{NIM,i}$$

where M_{NIM} denotes the total number of NIMs.

To choose a background set of MH SNPs that match the NIMs in terms of their MAF and LD distribution, we would pick a given bin i with probability $\frac{M_{NIM,i}}{M_{NIM}}$. The per-SNP heritability associated with this background set of MH SNPs is then given by:

$$\sigma_{MH} = \sum_i \frac{\sigma^2_{MH,i} M_{NIM,i}}{M_{NIM}} = \frac{1}{M_{NIM}} \sum_i h^2_{MH,i} \frac{M_{NIM,i}}{M_{MH,i}}$$

Thus, we are interested in testing the null hypothesis that the per-NIM heritability is equal to the per-SNP heritability of the background set of MH SNPs.

$$\sigma^2_{NIM} - \sigma^2_{MH} = 0$$

$$\Rightarrow \frac{1}{M_{NIM}} \sum_i h^2_{NIM,i} - \frac{1}{M_{NIM}} \sum_i h^2_{MH,i} \frac{M_{NIM,i}}{M_{MH,i}} = 0$$

$$\Rightarrow \sum_i h^2_{NIM,i} - \sum_i \frac{M_{NIM,i}}{M_{MH,i}} h^2_{MH,i} = 0$$

Defining our parameter of interest: $\Delta_{h^2} = \sum_i h^2_{NIM,i} - \sum_i \frac{M_{NIM,i}}{M_{MH,i}} h^2_{MH,i}$ our null hypothesis

is that $\Delta_{h^2} = 0$.

We estimate the relative reduction in NIM heritability as:

$$\delta_{h^2} = \frac{\sigma^2_{NIM} - \sigma^2_{MH}}{\sigma^2_{MH}}$$

$$= \frac{\sum_i h^2_{NIM,i} - \sum_i \frac{M_{NIM,i}}{M_{MH,i}} h^2_{MH,i}}{\sum_i \frac{M_{NIM,i}}{M_{MH,i}} h^2_{MH,i}}$$

CHAPTER 4

Impact of human specific variants on modern human biology

4.1 Introduction

Uncovering genetic changes that make anatomically modern humans *unique* is critical for a comprehensive understanding of human evolution. Recent genomic advances have began to elucidate the nuances of human evolution and how humankind relates to some of our closest hominid relatives. These advances are starting to reveal how intermixing between species has shaped modern human biology. Research has been done examining how regions of DNA that have been passed into homosapiens from our closest relatives, suggesting that some of this introgressed DNA may be adaptively beneficial or undergoing negative selection. On the other hand, there have been studies showing variant regions of DNA that are human-specific, or lacking in our closest relatives, suggesting that these mutations may have been important in developing modern human biology. One study uses ancestral recombination graphs to determine regions of the genome that are uniquely human [48]. Another study used massively parallel reporter assays to look at regulatory effects of modern human specific variants [49] in embryonic stem cells, neural progenitor cells, and bone osteoblasts finding (13%) of sequences containing these variants showed active regulatory activity, and (23%) of these drove differential expression between human groups. However, there hasn't been a comprehensive analysis determining and analysing the effect these mutations have on a broad range of human traits. By analyzing genome sequences from our closest evolution-

ary relatives, Neanderthals and Denisovans, we can functionally characterize these genetic changes. Towards this end, we identified 50,505 mutations that are nearly fixed for the derived allele in African individuals from the 1000 Genomes project (> 99% derived allele frequency) but are absent in all of the deeply sequenced Altai and Vindija Neanderthal and Denisovan genomes. Here we look at some of these human specific regions and how they impact modern human phenotypes.

To understand the phenotypic impact of these fixed derived mutations (FDMs), we leverage the observation that interbreeding with Neanderthals likely re-introduced the ancestral allele at a number of these sites. We estimate that 39% of FDMs are polymorphic in European populations so that their phenotypic impact can be analyzed by genotyping these mutations in large cohorts with phenotypic information. These Fixed Derived alleles (FDs) are mutations that rise to high frequency in modern humans since the split from archaic humans, and may give us clues to the biology that cause modern humans to differ from our closest relatives.

4.2 Results

4.2.1 Identifying genomic regions at which fixed derived mutations influence phenotypes

To understand how human specific mutations influence trait variation we first annotated the fixed derived mutations in the UK biobank. One approach we can use is to look at genetic components of traits that modern humans share, and our archaic relatives, such as the Neanderthal, don't. In other words, we wanted to identify derived mutations in the modern human genome that differed from our ancient relatives, and rose to high frequency. These mutations occur after modern humans split from Archaic individuals such as Neanderthals, and then rise to a high frequency in modern humans. In order to overcome the problem of these mutations becoming fixed, or at nearly 100 % allele frequency, we leverage the fact that

introgression may have reintroduced variation into these mutations in European individuals (Figure 4.1). We first identified 25,448 (50,505) mutations that were $> 99\%$($> 95\%$) for the derived allele in 1000 genomes phase 3 African populations and ancestral in the Altai Neanderthal or Denisovan and deemed Fixed Derived mutations (FD99 and FD95). These mutations are heterozygous in a number of White British individuals in the UK Biobank (4.2 4.3). After identifying the $FDMs > 99\%$, we used plink2 to run GLM on 96 phenotypes to find 464 FDMs associated with 39 phenotypes at a p-value threshold of $p < 10^{-10}$. For each phenotype, we clumped all significant FDMs that lie within 250 kb and with an LD threshold (r^2) of 0.5 using a significance threshold for the index SNP of 10^{-10}. After clumping analysis we find 70 independent FD99-phenotype associations of the 39 phenotypes. We repeated clumping analysis on $FDMs > 95\%$ and found 1457 FD95-phenotype associations over 67 phenotypes (Figure 4.4). We find that FDMs are associated with a wide range of phenotypic categories including Anthropomery, Blood-related, Bone Density, Metabolism, Kidney, Liver, Lung, Skin and Hair phenotypes.

4.2.2 Finemapping and functional annotation of FDMs

We then applied our finemapping protocol as referenced in the Section 3.3.1. Our pipeline starts with a subset of significantly associated FDMs that are relatively independent ($p < 10^{-10}$) followed by the application of a statistical fine-mapping method (SuSiE) within the 200kb window around each NIM signal (Wang 2020) and additional post-processing to obtain a set of NIMs that have an increased probability of being causal for a trait. In our post processing step, we removed the credible regions that have >50% non-FDMs in their credible set. The remaining credible sets all have majority FDMs (i.e. positive results), and they are further merged together with other such regions it overlaps with, resulting in distinct regions with evidence of FDMs causal effects. We termed the set of all resulting FDMs as the confident-credible FDM set and all FDMs that lie in the credible set as credible FDMs. There were two confident-credible FDM sets containing 11 confident-credible FDMs

that associated with 6 phenotypes (Table 4.1). The first confident-credible FDM set region (chr1:161378366-161579657) was associated with gamma-glutamyl transferase (GGT) and contained 5 confident-credible FDMs associated with GGT and Aspartate aminotransferase. The second confident-credible FDM set region (chr12:56764867-56967108) was associated with albumin, urea, and urate and contained 3 confident credible FDMs that were associated with all three phenotypes. We then sought to determine functional effect of the confident credible FDMs. We first looked at genes nearby to our loci of interest and found that our first locus (chr1:161378366-161579657) was in proximity to FCGR2A and RP11-25K21.6 genes (Figure 4.5). We looked at functional effects of our specific FDMs, however found that they were downstream and upstream gene variants of the two genes respectively 4.2. The FCGR2A codes for a receptor in many immune cells, such as macrophages and neutrophils, and is involved in the process of phagocytosis and clearing of immune complexes. Interestingly, this gene was also found to be of importance in our previous chapter. We next looked at the locus in chromosome 12 (chr12:56764867-56967108) and found that it was in close proximity to SPRYD4 and GLS2 genes (Figure 4.6). One study shows that SPRY-domain containing protein 4 (SPRYD4) inhibits tumor progression in hepatocellular carcinoma by inducing apoptotic cell death [50]. The SPRY-domain has been proposed to act as a protein-interaction module that is present in multiple proteins with diverse functions in different biological processes. Mutations in SPRY domain-containing genes have been reported in diseases like Opitz syndrome and familial Mediterranean fever, but as yet limited information is available on their association with the onset and progression of cancer. The GLS2 encodes a protein that regulates the metabolism of glutamine, a regulatory process that activates white blood cells. One study shows that variation in GLS2 is associated with development of complicated Staphylococcus aureus bacteremia [51]. Another study shows that the GLS2 also has tumor suppression activity in human hepatocellular carcinoma [52] similar to SPRYD4 gene. Theses genes are in close proximity providing an example of a region that potentially impacts function at multiple genes.

4.3 Methods

4.3.1 UK Biobank (UKBB) genotype QC

We restricted all our analyses to a set of high-quality imputed SNPs (with a hard call threshold of 0.2 and an info score greater than or equal to 0.8), which, among the 291,273 imputed genotypes of UKBB unrelated white British individuals, 1) have MAF higher than 0.001, 2) are under Hardy-Weinberg equilibrium ($p > 10^{-7}$), and 3) are confidently imputed in more than 99% of the genomes. Additionally, we excluded SNPs in the MHC region, resulting in a total of 7,774,235 SNP which we refer to as QC-ed SNPs.

4.3.2 Identifying Fixed Derived Mutations

We first determined derived allele frequencies from 504 African individuals in the 1000 genomes project phase3. We combined allele frequencies from ESN, GWD, LWK, MSL, YRI individuals in 1000 genomes in order to determine an African derived allele frequency for. We then determine the allele frequency in archaic individuals combining Vindija, and Altai neanderthals as well as the Denisovan individual. Combining the datasets, we examined 41,864,101 SNPs. We then filtered out SNPs with an African derived allele frequency >0.99 and >0.95 to find 25,448 and 50,505 mutations that are likely human specific. We then intersected these variants with those QC-ed SNPs in the UK Biobank to find 19,632 and 9,407 confident fixed derived mutations able to be tested. We expanded this set by including all QCed SNPs that were withing 200kb of a FDM95 and had an $r^2 > 0.8$, 0.9 or 0.99 to get 73,129, 71,280 and 62,933 expanded FDMs.

4.3.3 Association Testing

To identify individual FDMs associated with a phenotype, we fit a linear regression model using plink 2.0 –glm and included covariates controlling for age, sex and the first 20 genotypic

PCs. We used a stringent p-value threshold of 10^{-10} to correct for the number of FDMs and phenotypes tested. We found 6573, 6411, and 5804 for expanded sets with $r^2 > 0.8$, 0.9 or 0.99. For each phenotype, we clumped all significant FDMs that lie within 250 kb and with an LD threshold (r^2) of 0.5 using a significance threshold for the index SNP of 10^{-10}.

4.3.4 Annotating FDMs

We annotated all unique credible FDMs using SnpEff [1] which uses Sequence Ontology (http://www.sequenceontology.org/) to assign standardized terminology for assessing sequence change and impact.

4.4 Discussion

Our research shows that we are able to use the effect of Neanderthal introgression on the modern human genome to determine how fixed derived mutations impact modern human phenotypes. We present a new pipeline able to identify these fixed derived mutations and exploit their heterozygosity in Europeans to determine a signal of association in a wide range of phenotypes. We also use our fine mapping pipeline to discover two regions of the genome that are confidently fixed-derived and have associations with a number of phenotypes. We find that the two regions are in close proximity to genes that are related to immunity suggesting that these regions may have importance in development of immune response in modern humans, thus leading to fixation. In addition, literature review of the two genes in one of the regions are related to tumor suppression activity in human hepatocellular carcinoma, which is interesting considering the relationship between cancers and the immune response. One limitation of our study is that many regions containing significantly associated FDMs were removed potentially due to the limited power of our procedure that aims to control the FDR. Our current ongoing work involves fine-tuning our procedure to determine if can more accurately represent what is defined as "fixed derived". One approach is to

run evolutionary simulations to determine how the genetic architecture of these variants may evolve under a demography representing neanderthal introgressed DNA and fixation of variants. We can then fine tune the parameters of our pipeline of what levels of frequency might be needed to be considered "fixed" in African populations. Further work is also needed to determine the effect these fixed derived mutations have on heritability of modern human phenotypes.

4.5 Figures

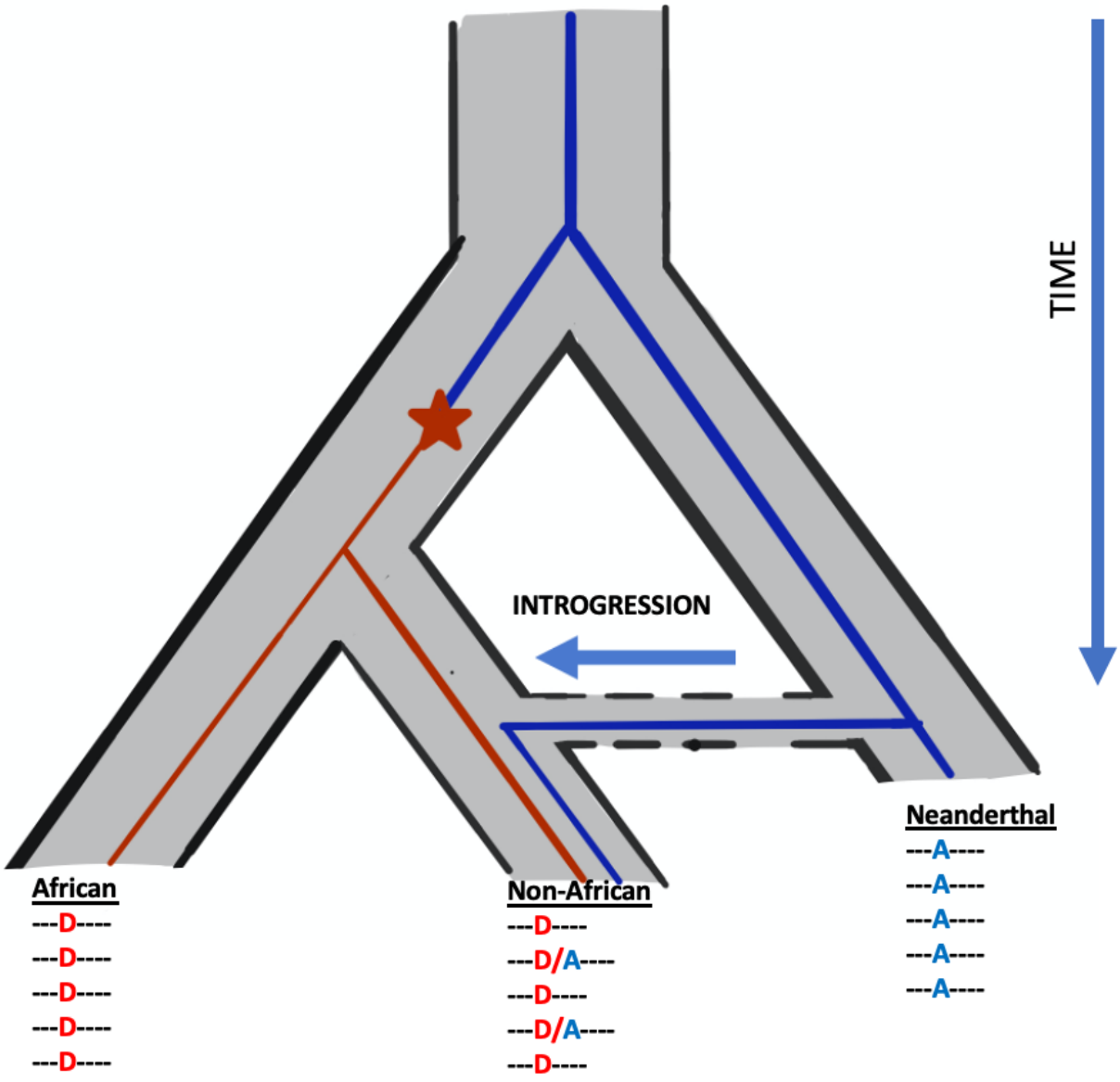

Figure 4.1: Cartoon depicting how Fixed Derived Mutations are determined

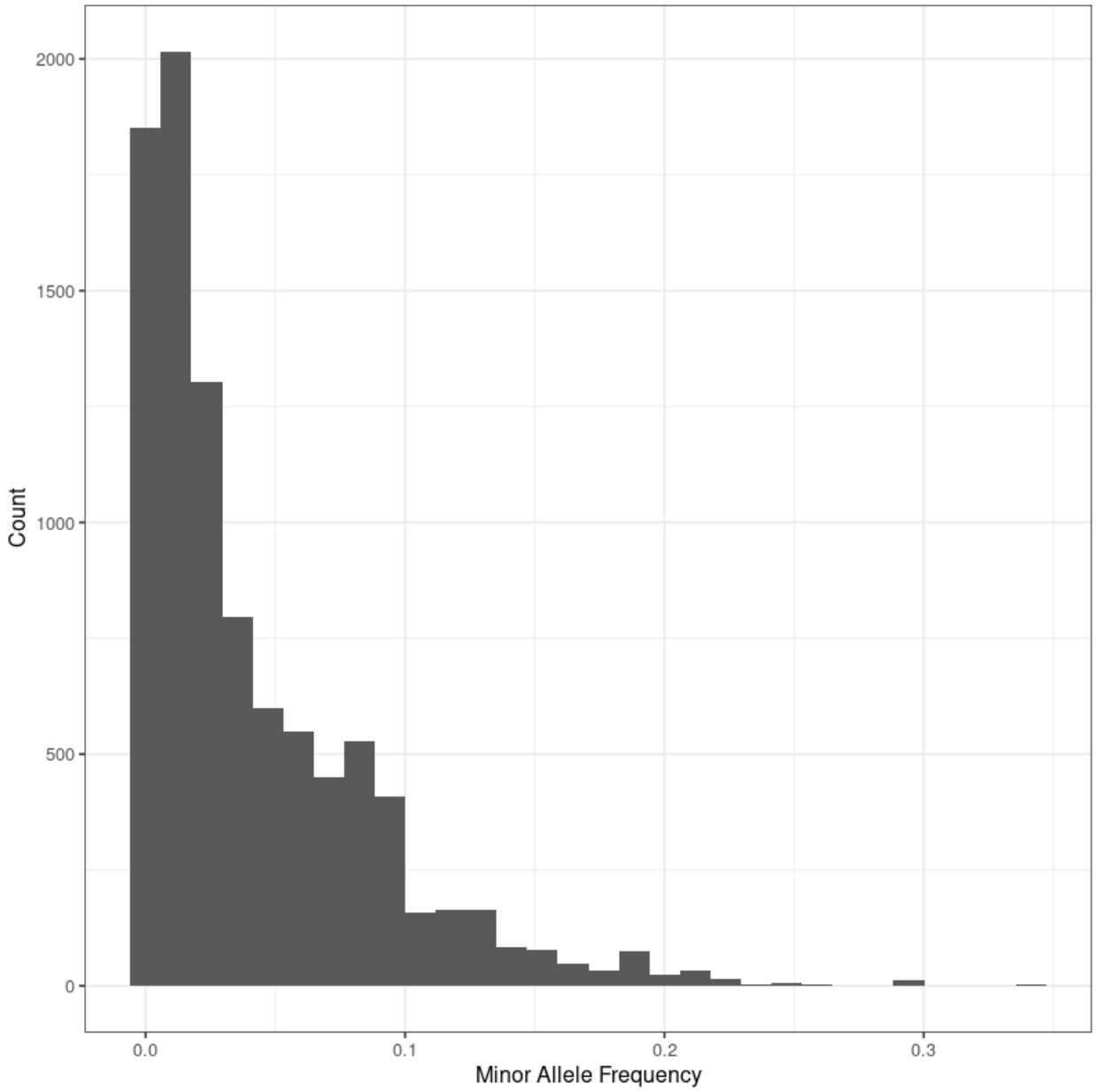

Figure 4.2: European MAF for mutations that were $> 99\%$ for the derived allele in 1000 genomes phase 3 African population.

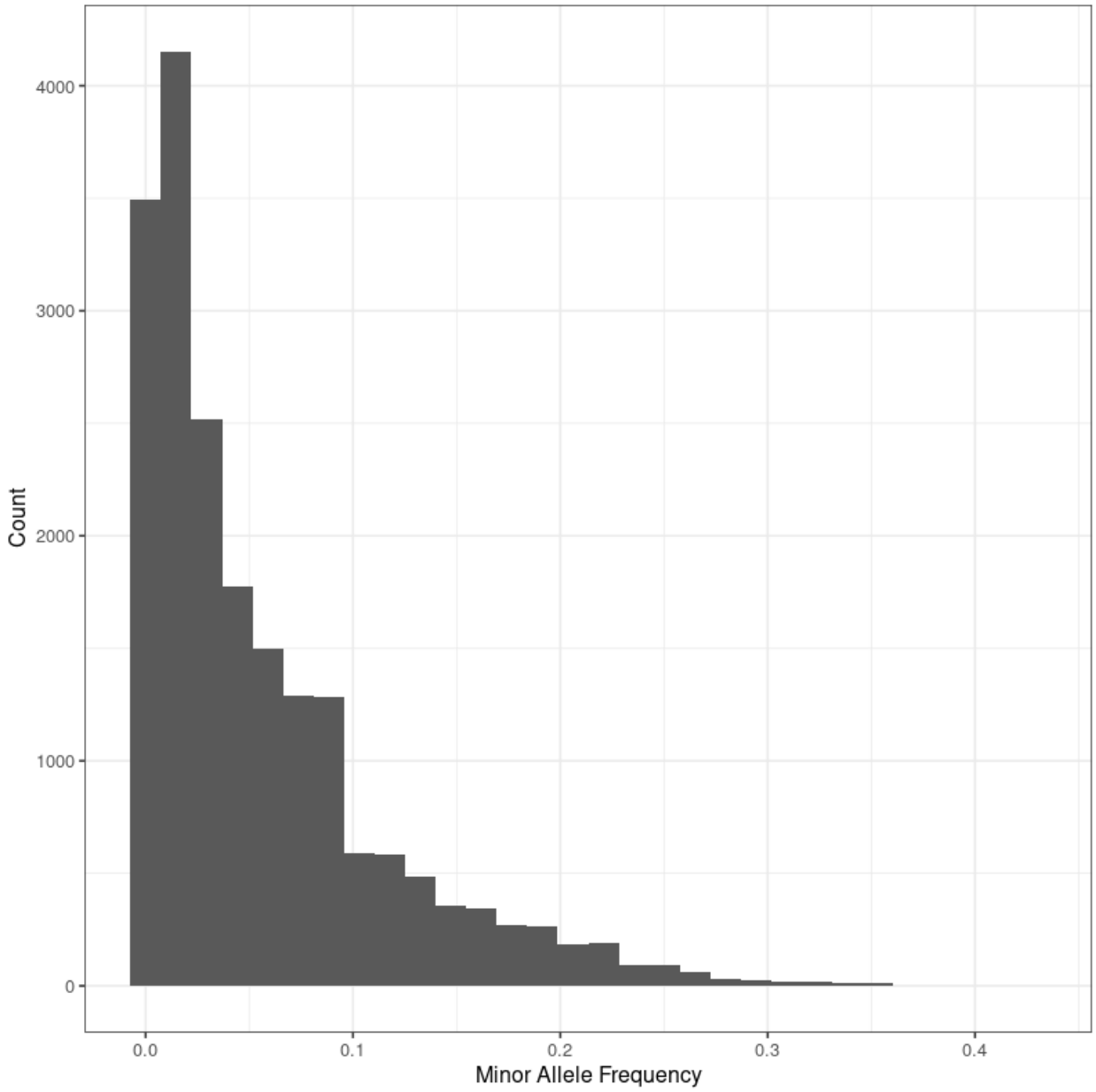

Figure 4.3: European MAF for mutations that were $> 95\%$ for the derived allele in 1000 genomes phase 3 African population.

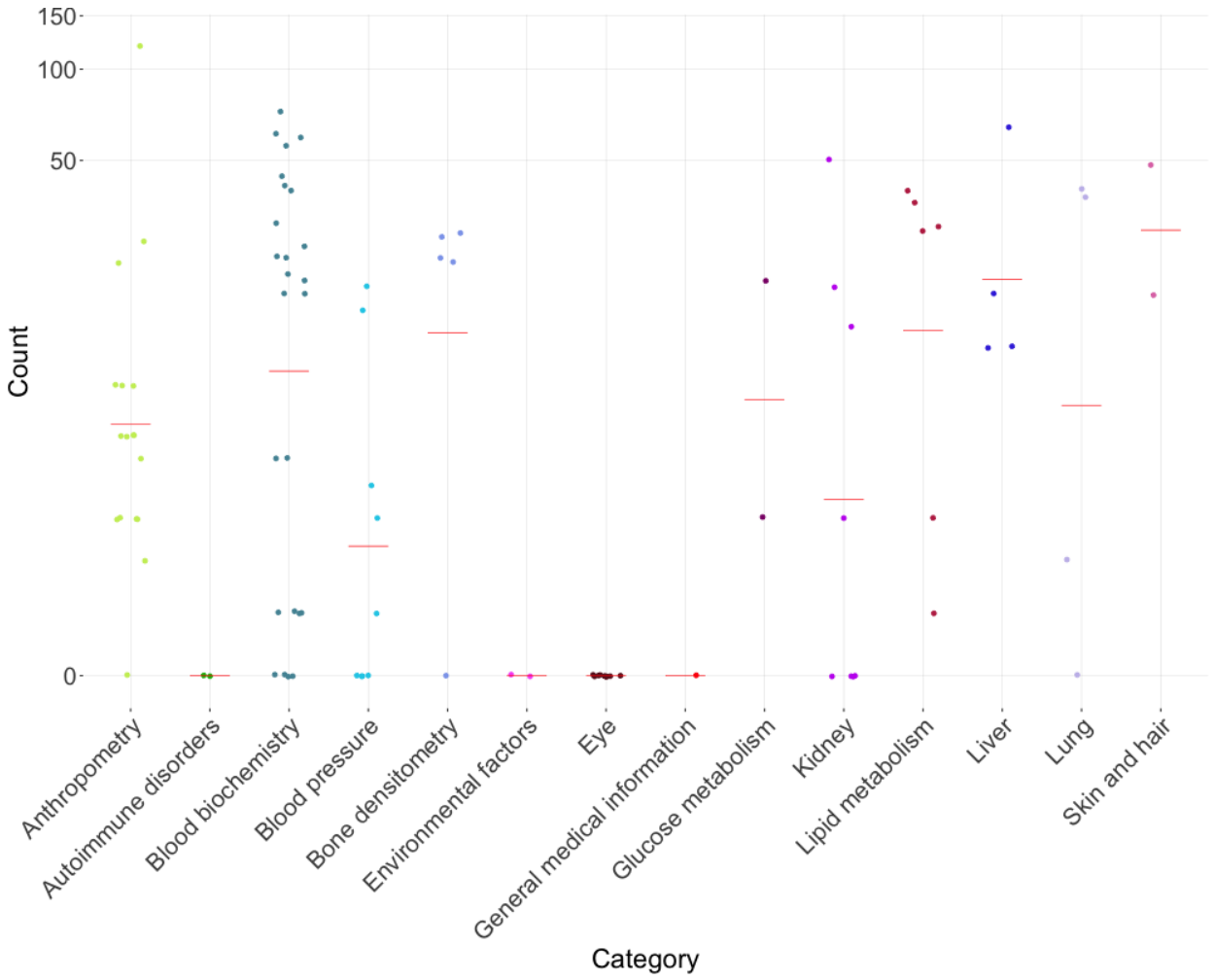

Figure 4.4: Number of FDMs (FD95) significantly associated with phenotypes grouped by Phenotypic category. Each dot in a category represents a unique phenotype.

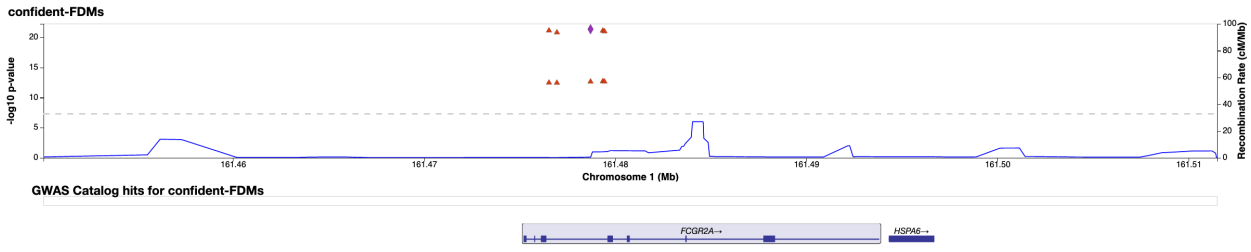

Figure 4.5: Zoomed in view of the confident credible FDM (FD95) region on chromosome 1

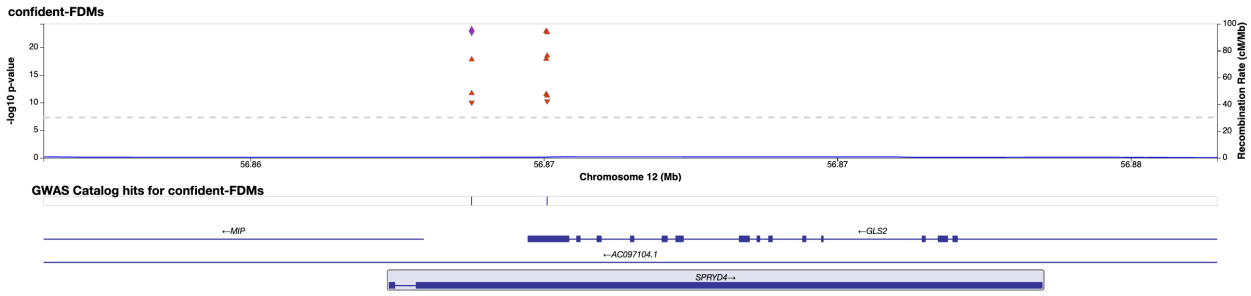

Figure 4.6: Zoomed in view of the confident credible FDM (FD95) region on chromosome 12

4.6 Tables

CHR	POS	REF	A1	BETA	P	Phenotype
12	56863770	G	C	0.0312767	1.69E-18	albumin
12	56865040	G	T	0.0313671	1.34E-18	albumin
12	56865056	G	C	0.0318923	3.58E-19	albumin
1	161476533	T	C	0.0385679	7.71E-22	asp_at
1	161476949	C	G	0.0382341	1.58E-21	asp_at
1	161478708	G	A	0.0386901	4.62E-22	asp_at
1	161479352	G	A	0.0383936	8.03E-22	asp_at
1	161479438	C	T	0.0382705	1.06E-21	asp_at
12	56863770	G	C	-0.0222723	9.99E-11	c_reactive_prot
12	56865056	G	C	-0.0225582	5.72E-11	c_reactive_prot
1	161476533	T	C	0.0282701	3.37E-13	ggt
1	161476949	C	G	0.0282119	3.58E-13	ggt
1	161478708	G	A	0.0284026	2.30E-13	ggt
1	161479352	G	A	0.0284242	2.00E-13	ggt
1	161479438	C	T	0.0283498	2.27E-13	ggt
12	56863770	G	C	0.0206617	2.17E-12	urate
12	56865040	G	T	0.0205196	3.03E-12	urate
12	56865056	G	C	0.0202435	5.92E-12	urate
12	56863770	G	C	0.0332519	1.24E-23	urea
12	56865040	G	T	0.0332406	1.25E-23	urea
12	56865056	G	C	0.0330631	2.17E-23	urea

Table 4.1: Confident-credible FDMs and their associations with phenotypes

CHR	POS	REF	ALT	Annotaion	Gene
1	161476533	T	C	downstream_gene_variant	FCGR2A
1	161476949	C	G	downstream_gene_variant	FCGR2A
1	161478708	G	A	upstream_gene_variant	RP11-25K21.6
1	161479352	G	A	upstream_gene_variant	RP11-25K21.6
1	161479438	C	T	upstream_gene_variant	RP11-25K21.6
12	56863770	G	C	3_prime_UTR_variant	SPRYD4
12	56865040	G	T	3_prime_UTR_variant	GLS2
12	56865056	G	C	3_prime_UTR_variant	GLS2

Table 4.2: Confident FDMs and their functional effect on genes

CHAPTER 5

Evolutionary modeling of the differential contribution of Neanderthal ancestry to complex traits provides insights into selective forces that shape trait variation

5.1 Introduction

We recently developed a methodology to assess whether Neanderthal ancestry is over- or under-represented in the genetic component of complex phenotypes compared to random genetic variation. Based on 500,000 individuals from the UK Biobank, we found the estimated contribution of Neanderthal alleles (NIMs) to phenotypic variation (NIM heritability) is significantly depleted in the great majority of the phenotypes. This is consistent with the observation that in general, natural selection has acted to remove Neanderthal alleles since introgression.

To understand the evolutionary models that could explain these observations, we performed forward-in-time population genetic simulations to model the evolution of Neanderthal and non-Neanderthal alleles according to a demographic model relating modern humans and Neanderthals. We chose parameters used in a previous study [33] analyzing the fitness cost of Neanderthal introgression. Specifically, an ancestral population of size 10,000 diploid individuals splits into a human population and a Neanderthal population, each one evolves separately before a single pulse of Neanderthal admixture followed by subsequent random mating. Under this demography, we modeled evolution of phenotypes subject to differ-

ent forces including directional, stabilizing, and disruptive selection. We estimated a NIM heritability Z-score, a measure of whether NIM heritability deviates significantly from the background non-introgressed alleles. We found under most models of selection, the NIM heritability Z-score is near zero or negative, indicating NIM heritability is neutral or depleted. Interestingly, we were able to recreate a positive NIM heritability Z-score, indicating an elevated Neanderthal contribution to heritability in two separate models of stabilizing and directional selection. In the stabilizing selection model, the optimal value of the trait is decreased in the human branch during the split between humans and Neanderthals leading to a positive NIM heritability Z-score. We also observe a positive NIM heritability Z-score in a directional selection model in which the parameter that couples SNP effect size and fitness is reduced after introgression. This observation highlights possible mechanisms for how complex traits evolved in human history by examining the genetic contribution of Neanderthal ancestry.

5.2 Results

Our goal was to examine how different evolutionary models affect NIMs and their contribution to phenotypic heritability. Specifically we sought to determine if variants of Neanderthal ancestry contribute to an enriched amount of heritability (NIM heritability) compared to a background set of MAF and LD matching SNPS under different conditions. To determine this we ran population genetic simulations to model evolution of phenotypes under several evolutionary forces including directional, stabilizing, and disruptive selection.

5.2.1 Examining how models of stabilizing, directional and disruptive selection impact NIM heritability

We first created a simple demographic model to quickly test different types of selection (Figure 5.1). We used forward-in-time simulation software SLiM 3.0 [53] in order to create a

demographic model of a common ancestral population followed by a split between Neanderthals and modern humans and finally a single pulse of introgression. Using this model, we were able to simulate how selective forces impacted evolution of polygenic phenotypes by looking at the quantitative trait loci (QTLs). We then examined the QTLs in order to quantify how they were affecting heritability of a phenotype.

We first examined how the evolutionary force of stabilizing selection impacted heritability in our simple demographic model. Under the force of stabilizing selection extreme values of a phenotype are not favored, while there is an optimum value of said quantitative phenotype (Fig 5.1 panel 2). We used a model of stabilizing selection developed by Lande et al. [54] in which each phenotype (y) has an optimum value (ω) which would impact fitness ($f(y)$) (Fig 5.2,5.3). We then ran our forward-in-time simulations under the simple demographic model and examined the QTLs to see how they would impact heritability in modern humans. We partitioned the QTLs into those of Neanderthal ancestry (NIMs) or those not of neanderthal ancestry (*see chapter 3 for full definition*). We calculated heritability of NIMs and compared them to a matching background set matched by MAF and LD to determine enrichment of NIM heritability or a positive NIM heritability Z-score. Under this model of stabilizing selection, we were not able to see any significant enrichment in NIM heritability. (*as defined in chapter 3*).

We then examined how directional selection impacts NIM heritability in our simple demography. Directional selection occurs when selection favours the phenotype at one extreme of the range of phenotypes increasing the fitness of those individuals with phenotypes nearer to that extreme (Fig 5.1 panel 1). We used a previously published model of directional selection by Eyre-Walker et al. [55] in which SNP effect size β and fitness (S) are coupled by a parameter Tau (τ). We again performed forward in time simulations under our simple demography to calculate a NIM heritability Z-score. We saw no significant enrichment in NIM heritability. Finally we looked at how disruptive selection impacts NIM heritability in a disruptive selection setting. Disruptive selection occurs when extreme values for a trait

are favored over intermediate values (Fig 5.1 panel 3). We based our simulation off a model from Zeng et al. [56] that examines disruptive selection for a quantitative trait by relating the normally distributed phenotype (y) to fitness (S) through a hypothetical function. After performing simulations under our simple demography and modeling disruptive selection, we found no significant enrichment in NIM heritability Z-score.

5.2.2 Modified models of stabilizing and directional selection present enriched NIM heritability

After we observed no enrichment of NIM heritability, we wanted to see if we modified the models by adjusting different parameters at different times in the Neanderthal introgression demography would allow us to see enrichment. We developed a model of stabilizing selection that relaxed the parameter of a constant QTL optima (ω) at different stages in the evolutionary history. We created several different cases where the optima would shift at different branches of our demographic model following the split of modern humans and neanderthals. We defined ω_A as the optima in the shared common ancestors, and shifted the optima immediately after the split between humans (ω_H) and Neanderthals (ω_N) and immediately after introgression into modern humans ω_{MH}. We explored many different combinations of changing that parameters to be constant, have a positive or negative change in value, and/or a reversal in sign ($-\omega$) of the optima (Fig 5.3). We found that a decrease in QTL optima value in the human population (Neanderthal optima unchanged) after the split between Neanderthals and modern humans produced an enrichment in NIM heritability (*i.e.*, $\omega_A = \omega_N > \omega_H$; Fig 5.4). Under this modified model of directional selection, we observed a positive NIM heritability Z-score indicating an enrichment of NIM heritability in our simulated phenotype (Fig 5.5).

We next modified the previously stated model of directional selection that relaxed the assumption of a constant relationship between SNP effect size β and fitness S throughout an evolutionary history. In the previous model, β and S are coupled through a parameter τ,

quantifying the relationship between the two. In our modified model, we allowed for the value of τ to change before (τ_a) and immediately after (τ_i) introgression (Fig 5.6 and Fig 5.7). We found that a reduction in the value of the initial coupling parameter (τ_a) to a smaller value after introgression (τ_i) produces a positive NIM heritability Z-score (i.e. $\tau_a > \tau_i$) (Fig 5.8). This positive NIM heritability Z-score again indicates that under this specific evolutionary scenario, we observe an enrichment of NIM heritability in our simulated phenotype.

We repeated these experiments under a more realistic demography than our simple demographic model based on parameters from a previously published model by Harris and Nielsen [18]. In this model, there is a development of genetic background in a common ancestor for a number of generations, a split between neanderthals and modern humans, a second split within modern humans simulating the out of Africa movement, followed immediately by introgression into a non-African population (Fig 5.3). Under this more realistic model, both results of NIM heritability enrichment in our modified directional and selection models persist. Lastly, we explored variations in several demographic parameters to observe how they may impact NIM heritability. We simulated models in which there is (1) changes in values of effective population sizes of Neanderthals or modern humans (2) changes in genomic element size (3) changes in admixture proportion and (4) changes in causal variant proportion. In all of these demographic models, we observe no significant enrichment of NIM heritability.

5.3 Methods

5.3.1 Simple demographic model

We created a simple demographic model to simulate Neanderthal introgression into modern humans. We used SLiM 3.0 [53] to run a forward in time simulations in a normally distributed phenotype with quantitative trait loci. The mutation rate was set to 10^{-7} and the chromosome size was set to 100 kb with neutral and QTL mutations occurring at equal

proportions with a rate of 10−7. In our model we begin with an ancestral population size of $N_A = 10,000$ and allowed this population to develop over 3,980 generations. From here, we had the population split into two separate sized populations of humans ($N_H = 10,000$) and Neanderthals ($N_N = 1,000$). These two populations randomly mated separately for 10,000 generations followed by an introgression event of 20%. Finally, modern humans evolved for another 2,000 generations and effect sizes of QTLs were observed.

5.3.2 Stabilizing selection model

We used a model of stabilizing selection previous published by Lande et al.[54] in which a normally distributed continuous phenotype (y) had an optimum value of (ω) and a phenotypic variance of (σ^2). The fitness of an individual $f(y)$ can be determined by the equation:

$$f(y) = 10 * \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\omega - y)^2}{2\sigma^2}\right\}$$

We used this calculation of fitness in the simulation mentioned in section 5.3.1

5.3.3 Modified model of stabilizing selection

We relaxed the assumption of a constant phenotype optima ω in the section above and allowed it to vary at different points in the evolutionary history. We let ω_i be the phenotypic optima occurring in ancestral population, humans and neanderthals after split, and modern humans i.e. $i = \{A, H, N, MH\}$ representing different values in those populations. Our new model of fitness for individuals in a population i can be determined:

$$f(y) = 10 * \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\omega_i - y)^2}{2\sigma^2}\right\}$$

We used this calculation of fitness in the simulation mentioned in section 5.3.1

5.3.4 Directional selection model

We used a model of directional selection previously published by Eyre-Walker [55] in which a the effect size of a QTL (β) is coupled with fitness (S) by a parameter τ where $S = 4N_E s$ and ϵ is normally distributed with a mean 0 and a SD σ. In this model δ transforms the distribution of effects such that mutations have equal probabilities of increasing or decreasing the trait.

$$\beta(S, \epsilon, \delta, t) = \delta S^{\tau}(1 + \epsilon)$$

In this model the strength of association between the effects of mutations on the trait and fitness is dependent upon the parameter τ. We used this calculation of fitness in the simulation mentioned in section 5.3.1.

5.3.5 Modified directional selection model

We relaxed the assumption of a constant coupling parameter τ in the model of directional selection above and allowed the value to change immediately following introgression (Fig 5.7). In our model, the strength of association (τ_j) between effect size β and selection S changes in a population i before introgression and after introgression i.e. $j = \{a, i\}$. Our new model can be determined:

$$\beta(S, \epsilon, \delta, t) = \delta S^{\tau_j}(1 + \epsilon)$$

We used this model of fitness and effect size in the simulation mentioned in section 5.3.1.

5.3.6 Realistic demographic model

We created a more realistic demographic model to simulate Neanderthal introgression into modern humans based on a model previously published by Harris and Nielsen [18]. We used SLiM 3.0 [53] to run a forward in time simulations in a normally distributed phenotype with quantitative trait loci. The mutation rate was set to 10^{-8} and the chromosome size was set to 100kb with neutral and QTL mutations occurring at equal proportions with a rate of 10−8. In our model we begin with an ancestral population size of $N_A = 10,000$ and allowed this population to burn in for 70,000 generations. From here, we had the population split into two separate sized populations of humans ($N_H = 10,000$) and Neanderthals ($N_N = 1,000$). These two populations randomly mated separately for 17,800 generations followed followed by a splitting of modern humans into two equal sized populations of Africans ($N_{AFR} = 10,000$) and Europeans ($N_{EUR} = 10,000$). One generation after the split, there was an introgression event of 10%. Finally, modern humans evolved for another 2,200 generations and effect sizes of QTLs were observed. We used this model to simulate the same forces of selection in previous sections.

5.4 Discussion

Our research provides new insights into how selective forces acting upon the genetic architecture of modern human and Neanderthal populations can affect phenotypes under an introgressed demography. Looking at phenotypic variance data in modern humans, partitioned by modern human and neanderthal introgressed DNA, may allow us to determine what selective forces might be acting upon a population. We can compare and contrast the partitioned heritability estimates in real data and compare them to simulated evolutionary models to determine how these genetic architectures may have evolved over time and their influence on phenotypic heritability. In general, we find that most models of selection (including stabilizing and directional selection) lead to NIM heritability that is compara-

ble or lower than non-introgressed SNPs consistent with our empirical results in Chapter 3 that document that across a large number of phenotypes, their NIM heritability is depleted relative to matched non-introgressed SNPs.

5.5 Figures

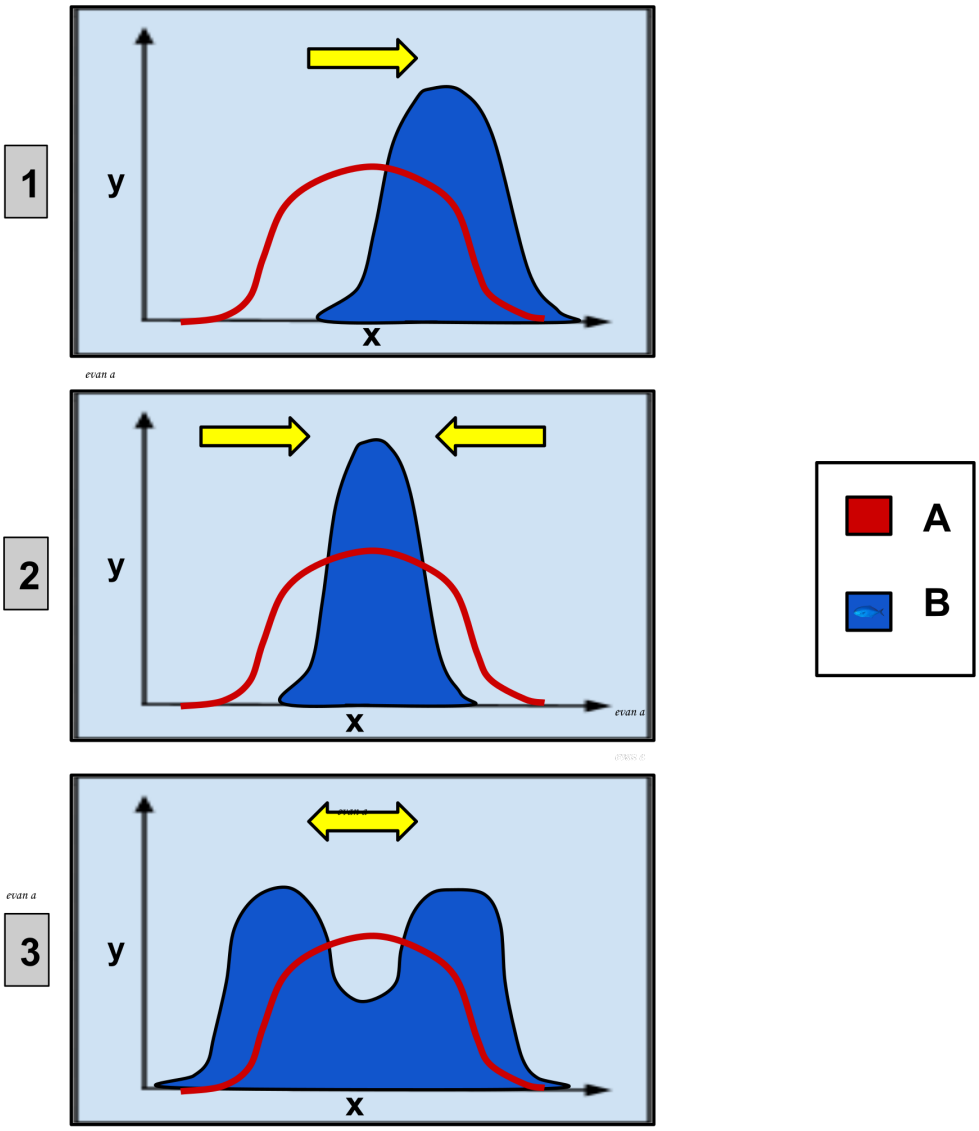

Figure 5.1: Different models of selection: In each panel, x-axis is the phenotype value and y-axis is the number of organisms. Group A is the population distribution before selection and group B is after selection. 1) Directional section, single extreme phenotype value is favored to be the QTL optima. 2) Stabilizing selection, where an intermediate phenotype is favored and 3) Disruptive selection, where an extreme phenotype on either end is favored. Source: Ealbert17, Genetic Distribution.svg. CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons.

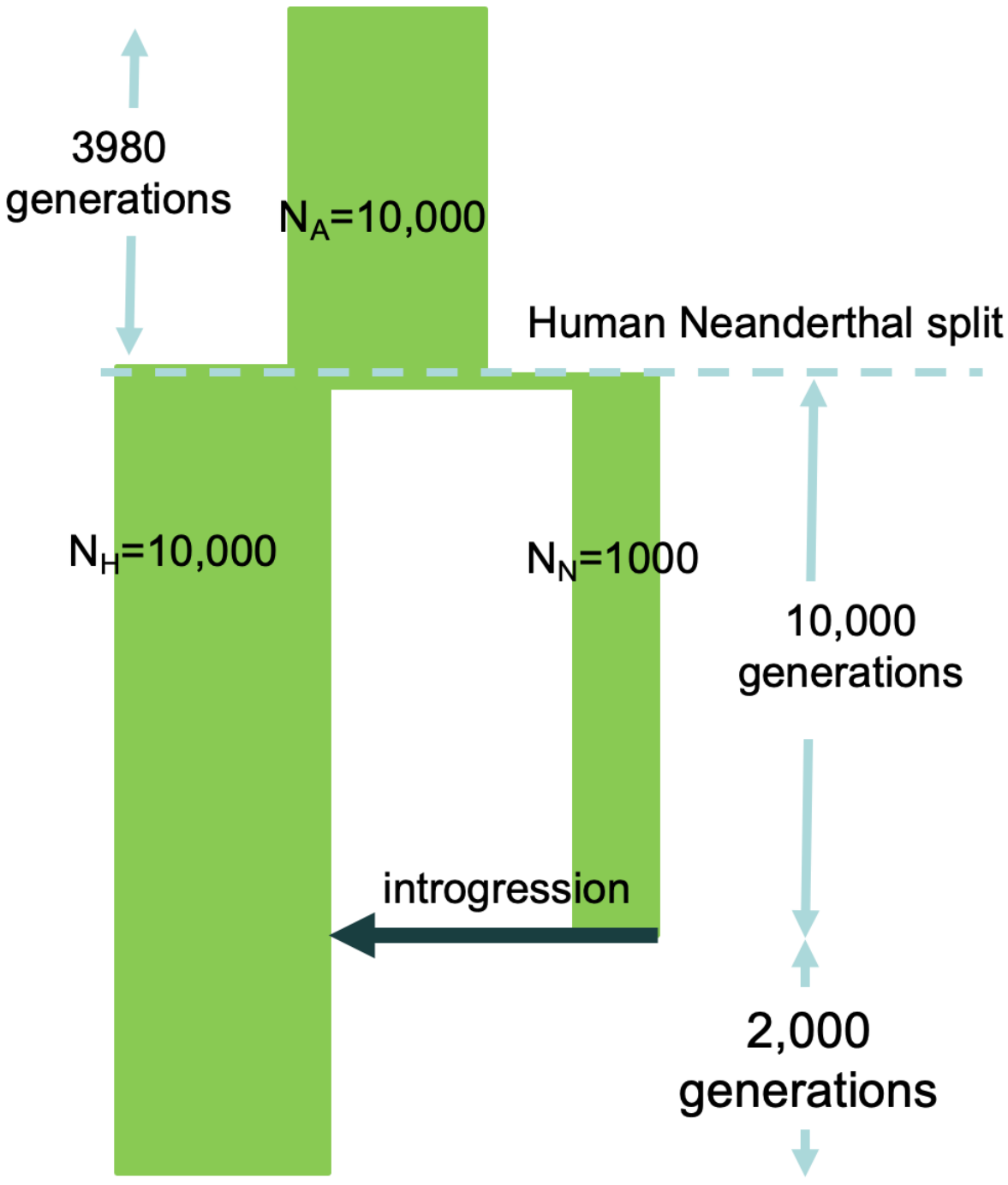

Figure 5.2: Simple demographic model

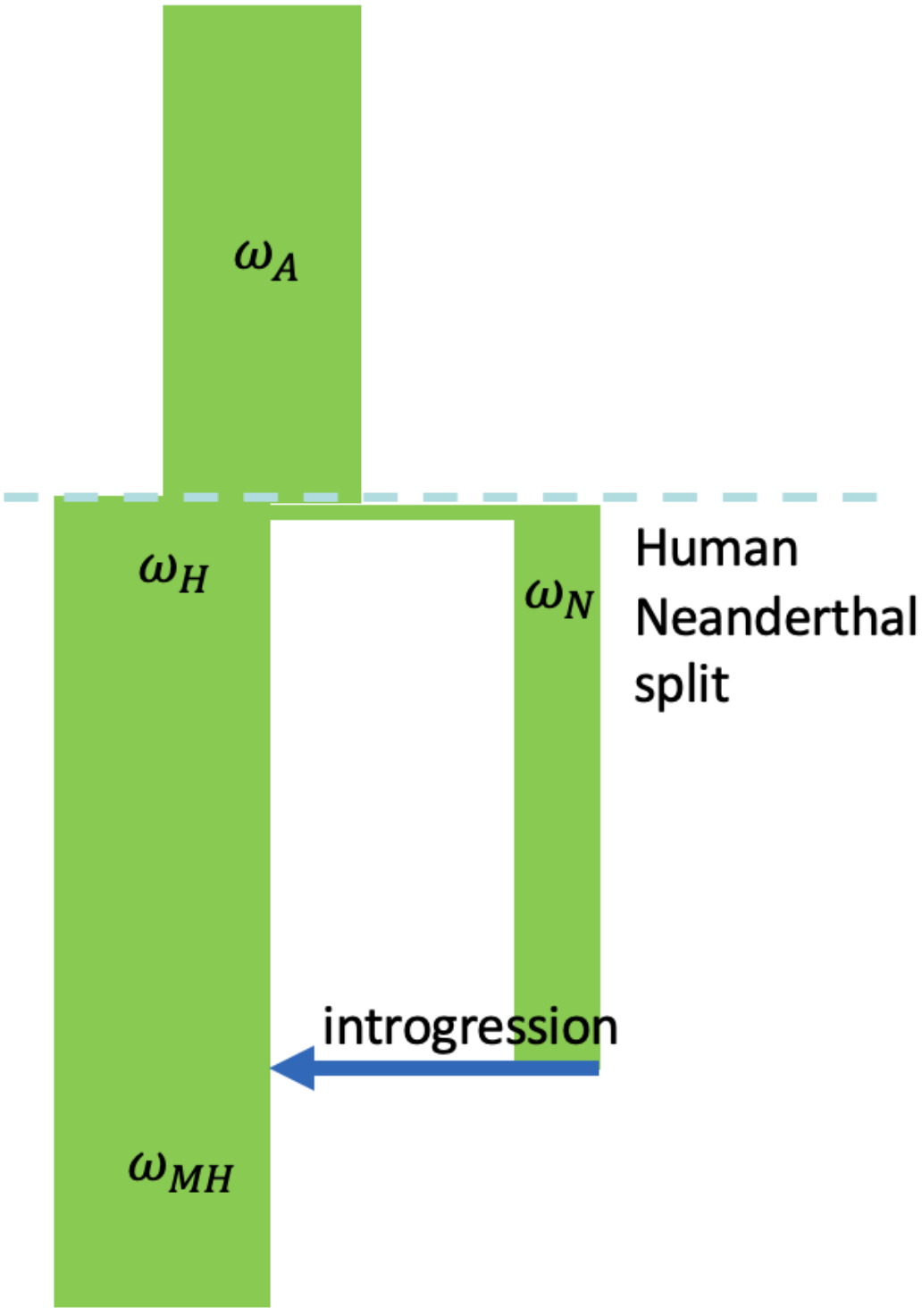

Figure 5.3: Simple demographic model with stabilizing selection

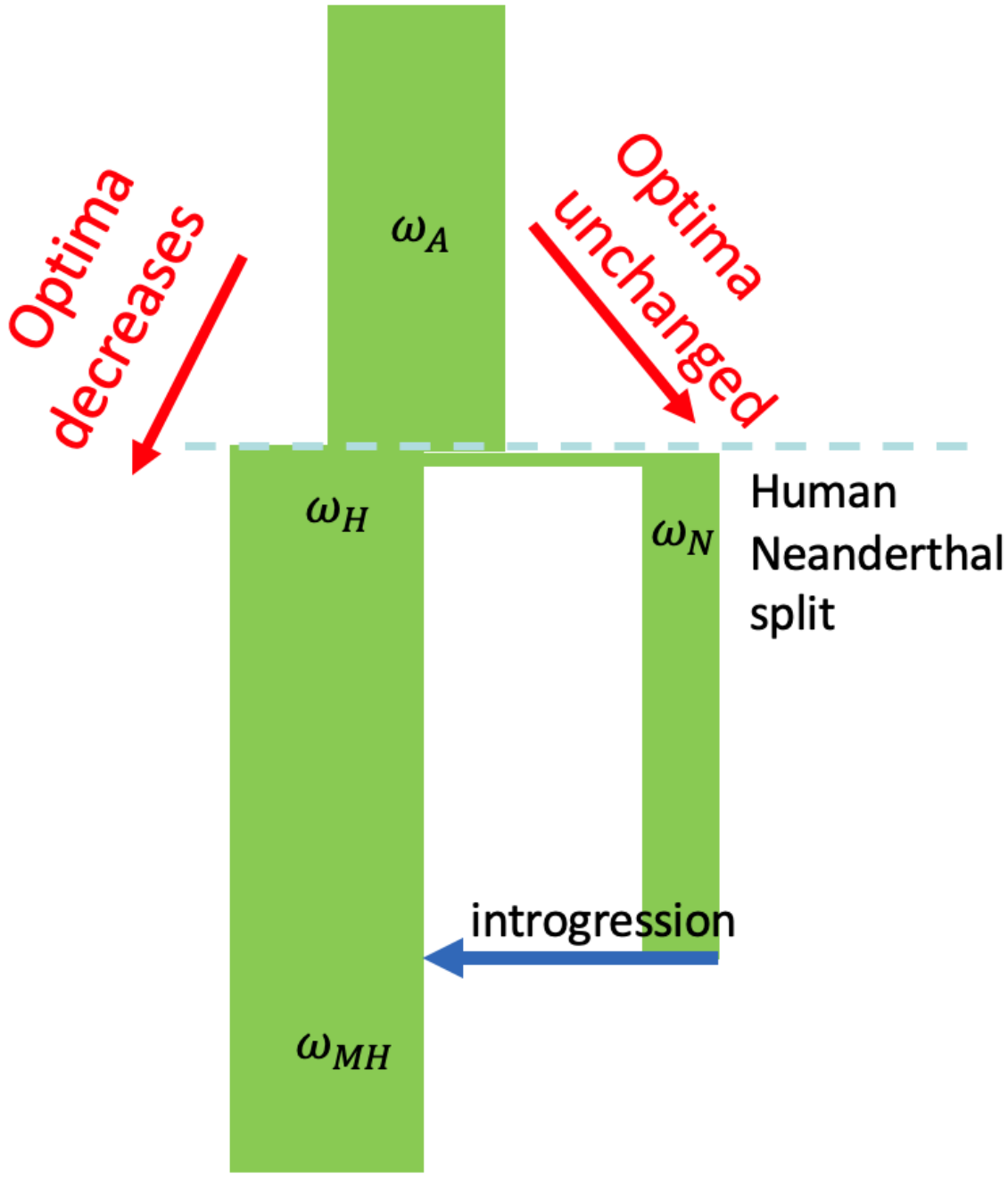

Figure 5.4: Modified stabilizing selection model

$\omega = \{\omega_A, \omega_H, \omega_N, \omega_{MH}\}$	Z-score
ω = {200,200,1,200}	-10.8399
ω = {200,200,1,1}	-6.16116
ω = {200,200,1,50}	-8.02057
ω = {200,1,200,200}	5.90213
ω = {200,1,200,1}	5.93435
ω = {200,1,200,50}	5.44799
ω = {200,200,200,200}	-0.472652
ω = {200,200,200,50}	2.03464
ω = {200,200,200,1}	1.5733

Figure 5.5: Modified stabilizing selection model produces positive NIM heritability Z-scores

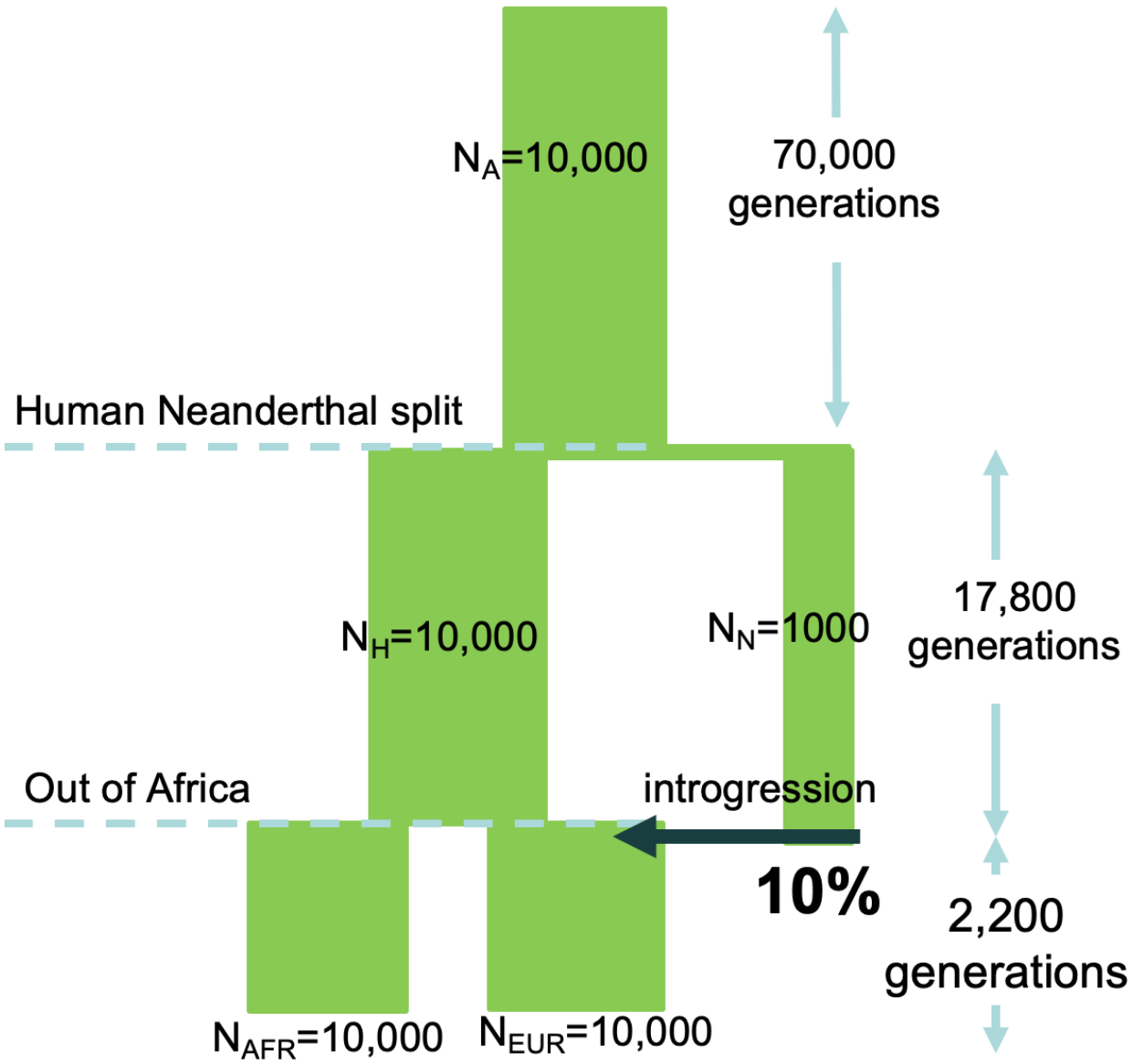

Figure 5.6: Realistic demographic model

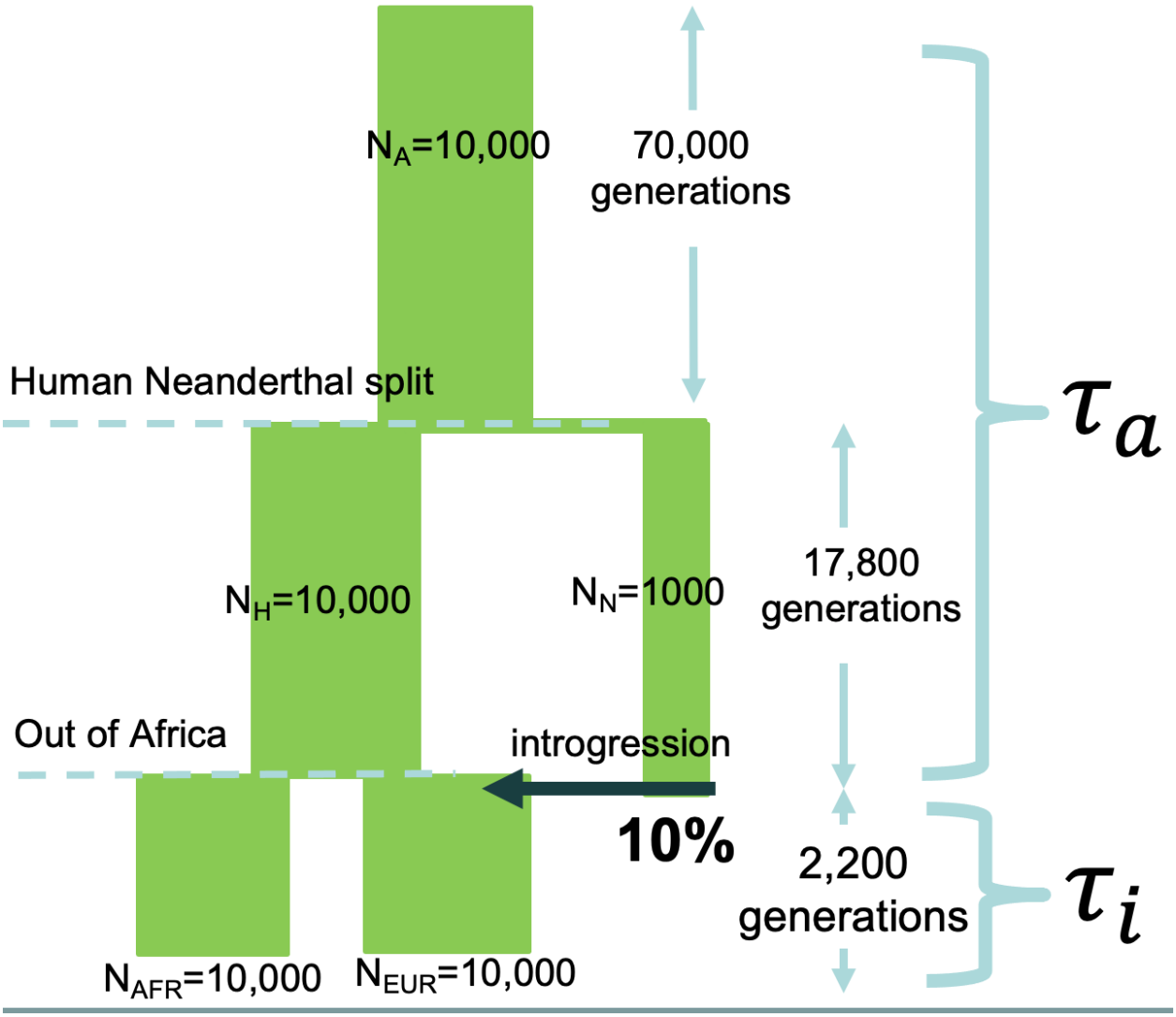

Figure 5.7: Directional selection - modified model

τ_a	τ_i	Z-score
0	0	-0.3506
0.5	0.25	26.2896
1	1	-1.542402
2	4	-0.493292

Figure 5.8: Modified directional selection model produces NIM heritability Z-scores

REFERENCES

[1] P. Cingolani, A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden. A program for annotating and predicting the effects of single nucleotide polymorphisms, snpeff: Snps in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. *Fly.*, 6:80–92, 2012.

[2] A. P. Boughton, R. P. Welch, M. Flickinger, P. VandeHaar, D. Taliun, G. R. Abecasis, and M. Boehnke. Locuszoom.js: interactive and embeddable visualization of genetic association study results. *Bioinformatics.*, 37:3017–3018, 2021.

[3] Sohini Ramachandran, Omkar Deshpande, Charles C Roseman, Noah A Rosenberg, Marcus W Feldman, and L Luca Cavalli-Sforza. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in africa. *Proceedings of the National Academy of Sciences*, 102(44):15942–15947, 2005.

[4] Stefano Benazzi, Katerina Douka, Cinzia Fornai, Catherine C Bauer, Ottmar Kullmer, Jiří Svoboda, Ildikó Pap, Francesco Mallegni, Priscilla Bayle, Michael Coquerelle, et al. Early dispersal of modern humans in europe and implications for neanderthal behaviour. *Nature*, 479(7374):525–528, 2011.

[5] R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H.-Y. Fritz, N. F. Hansen, E. Y. Durand, A.-S. Malaspinas, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Pr"ufer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. H"ober, B. H"offner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, and D. Brajkovic. A draft sequence of the neandertal genome. *Science*, 328:710–722, 2010.

[6] David Reich, Richard E Green, Martin Kircher, Johannes Krause, Nick Patterson, Eric Y Durand, Bence Viola, Adrian W Briggs, Udo Stenzel, Philip LF Johnson, et al. Genetic history of an archaic hominin group from denisova cave in siberia. *nature*, 468(7327):1053–1060, 2010.

[7] K. Pr"ufer, F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. de Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu, M. Kircher, M. Kuhlwilm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P. Moorjani, J. Pickrell, J. C. Mullikin, S. H. Vohr, R. E. Green, I. Hellmann, P. L. F. Johnson, H. Blanche, H. Cann, J. O. Kitzman, J. Shendure, E. E. Eichler, E. S. Lein, T. E. Bakken, L. V. Golovanova, V. B. Doronichev, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Slatkin, D. Reich, J. Kelso, and S. P"a"abo. The complete genome sequence of a neanderthal from the altai mountains. *Nature*, 505:43–49, 2014.

[8] Sriram Sankararaman, Nick Patterson, Heng Li, Svante Pääbo, and David Reich. The date of interbreeding between neandertals and modern humans. *PLoS Genetics*, 2012.

[9] Richard G Klein and Blake Edgar. *The dawn of human culture*. Wiley New York, 2002.

[10] S. Sankararaman, S. Mallick, M. Dannemann, K. Pr"ufer, J. Kelso, S. P"a"abo, N. Patterson, and D. Reich. The genomic landscape of neanderthal ancestry in present-day humans. *Nature*, 507:354–357, 2014.

[11] B. Vernot and J. M. Akey. Resurrecting surviving neandertal lineages from modern human genomes. *Science.*, 343:1017–1021, 2014.

[12] Sriram Sankararaman, Swapan Mallick, Nick Patterson, and David Reich. The combined landscape of denisovan and neanderthal ancestry in present-day humans. *Current Biology*, 26(9):1241–1247, 2016.

[13] Benjamin Vernot, Serena Tucci, Janet Kelso, Joshua G Schraiber, Aaron B Wolf, Rachel M Gittelman, Michael Dannemann, Steffi Grote, Rajiv C McCoy, Heather Norton, et al. Excavating neandertal and denisovan dna from the genomes of melanesian individuals. *Science*, 352(6282):235–239, 2016.

[14] F. Racimo, S. Sankararaman, R. Nielsen, and E. Huerta-Sánchez. Evidence for archaic adaptive introgression in humans. *Nat Rev Genet.*, 16:359–371, 2015.

[15] C. N. Simonti, B. Vernot, L. Bastarache, E. Bottinger, D. S. Carrell, R. L. Chisholm, D. R. Crosslin, S. J. Hebbring, G. P. Jarvik, I. J. Kullo, R. Li, J. Pathak, M. D. Ritchie, D. M. Roden, S. S. Verma, G. Tromp, J. D. Prato, W. S. Bush, J. M. Akey, J. C. Denny, and J. A. Capra. The phenotypic legacy of admixture between modern humans and neandertals. *Science*, 351:737–741, 2016.

[16] M. Dannemann and J. Kelso. The contribution of neanderthals to phenotypic variation in modern humans. *The American Journal of Human Genetics*, 101:578–589, 2017.

[17] Charleston WK Chiang, Serghei Mangul, Christopher Robles, and Sriram Sankararaman. A comprehensive map of genetic variation in the world's largest ethnic group—han chinese. *Molecular biology and evolution*, 35(11):2736–2750, 2018.

[18] Kelley Harris and Rasmus Nielsen. The genetic cost of neanderthal introgression. *Genetics*, 203(2):881–891, 2016.

[19] I. Juric, S. Aeschbacher, and G. Coop. The strength of selection against neanderthal introgression. *PLoS Genet*, 12, 2016.

[20] Na Cai, Tim B Bigdeli, Warren Kretzschmar, Yihan Li, Jieqin Liang, Li Song, Jingchu Hu, Qibin Li, Wei Jin, Zhenfei Hu, et al. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature*, 523(7562):588–591, 2015.

[21] N Cai, TB Bigdeli, WW Kretzschmar, Y Li, J Liang, J Hu, RE Peterson, S Bacanu, BT Webb, B Riley, et al. 11,670 whole-genome sequences representative of the han chinese population from the converge project. sci. data 4: 170011, 2017.

[22] Dorret I Boomsma, Cisca Wijmenga, Eline P Slagboom, Morris A Swertz, Lennart C Karssen, Abdel Abdellaoui, Kai Ye, Victor Guryev, Martijn Vermaat, Freerk Van Dijk, et al. The genome of the netherlands: design, and project goals. *European Journal of Human Genetics*, 22(2):221–227, 2014.

[23] Carlo Sidore, Fabio Busonero, Andrea Maschio, Eleonora Porcu, Silvia Naitza, Magdalena Zoledziewska, Antonella Mulas, Giorgio Pistis, Maristella Steri, Fabrice Danjou, et al. Genome sequencing elucidates sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. *Nature genetics*, 47(11):1272–1281, 2015.

[24] Daniel F Gudbjartsson, Hannes Helgason, Sigurjon A Gudjonsson, Florian Zink, Asmundur Oddson, Arnaldur Gylfason, Soren Besenbacher, Gisli Magnusson, Bjarni V Halldorsson, Eirikur Hjartarson, et al. Large-scale whole-genome sequencing of the icelandic population. *Nature genetics*, 47(5):435–444, 2015.

[25] K Walter, JL Min, J Huang, L Crooks, Y Memari, S McCarthy, JRB Perry, C Xu, M Futema, D Lawson, et al. Management committee (2015). the uk10k project identifies rare variants in health and disease. *Nature*, 526(7571):82–90, 2015.

[26] Monkol Lek, Konrad J Karczewski, Eric V Minikel, Kaitlin E Samocha, Eric Banks, Timothy Fennell, Anne H O'Donnell-Luria, James S Ware, Andrew J Hill, Beryl B Cummings, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536(7616):285–291, 2016.

[27] Bernard Y Kim and Kirk E Lohmueller. Selection and reduced population size cannot explain higher amounts of neandertal ancestry in east asian than in european human populations. *The American Journal of Human Genetics*, 96(3):454–461, 2015.

[28] Jian Yang, S Hong Lee, Michael E Goddard, and Peter M Visscher. Gcta: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, 88(1):76–82, 2011.

[29] Xinzhu Wei, Christopher R Robles, Ali Pazokitoroudi, Andrea Ganna, Alexander Gusev, Arun Durvasula, Steven Gazal, Po-Ru Loh, David Reich, and Sriram Sankararaman. The lingering effects of neanderthal introgression on human complex traits. *bioRxiv*, 2022.

[30] F. L. Mendez, J. C. Watkins, M. F. Hammer, and A. Haplotype. at stat2 introgressed from neanderthals and serves as a candidate of positive selection in papua new guinea. *The American Journal of Human Genetics*, 91:265–274, 2012.

[31] L. Abi-Rached, M. J. Jobin, S. Kulkarni, A. McWhinnie, K. Dalva, L. Gragert, F. Babrzadeh, B. Gharizadeh, M. Luo, F. A. Plummer, J. Kimani, M. Carrington, D. Middleton, R. Rajalingam, M. Beksac, S. G. E. Marsh, M. Maiers, L. A. Guethlein, S. Tavoularis, A.-M. Little, R. E. Green, P. J. Norman, and P. Parham. The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science.*, 334:89–94, 2011.

[32] R. M. Gittelman, J. G. Schraiber, B. Vernot, C. Mikacenic, M. M. Wurfel, and J. M. Akey. Archaic hominin admixture facilitated adaptation to out-of-africa environments. *Current Biology*, 26:3375–3382, 2016.

[33] M. Petr, S. P"a"abo, J. Kelso, and B. Vernot. Limits of long-term selection against neandertal introgression. *Proc Natl Acad Sci USA*, 116:1639–1644, 2019.

[34] L. Skov, M. Coll Macià, G. Sveinbj"ornsson, F. Mafessoni, E. A. Lucotte, M. S. Einarsdóttir, H. Jonsson, B. Halldorsson, D. F. Gudbjartsson, A. Helgason, M. H. Schierup, and K. Stefansson. The nature of neanderthal introgression revealed by 27,566 icelandic genomes. *Nature*, 582:78–83, 2020.

[35] E. McArthur, D. C. Rinker, and J. A. Capra. Quantifying the contribution of neanderthal introgression to the heritability of complex traits. *Nat Commun.*, 12:4481, 2021.

[36] C. Bycroft, C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O. Delaneau, J. O'Connell, A. Cortes, S. Welsh, A. Young, M. Effingham, G. McVean, S. Leslie, N. Allen, P. Donnelly, and J. Marchini. The uk biobank resource with deep phenotyping and genomic data. *Nature.*, 562:203–209, 2018.

[37] A. Pazokitoroudi, Y. Wu, K. S. Burch, K. Hou, A. Zhou, B. Pasaniuc, and S. Sankararaman. Efficient variance components analysis across millions of genomes. *Nat Commun.*, 11:4020, 2020.

[38] L. M. Evans, R. Tahmasbi, S. I. Vrieze, G. R. Abecasis, S. Das, S. Gazal, D. W. Bjelland, T. R. de Candia, Haplotype Reference Consortium, M. E. Goddard, B. M. Neale, J. Yang, P. M. Visscher, and M. C. Keller. Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nat Genet*, 50:737–745, 2018.

[39] Steven Gazal, Hilary K Finucane, Nicholas A Furlotte, Po-Ru Loh, Pier Francesco Palamara, Xuanyao Liu, Armin Schoech, Brendan Bulik-Sullivan, Benjamin M Neale, Alexander Gusev, et al. Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. *Nature genetics*, 49(10):1421, 2017.

[40] Hilary K Finucane, Brendan Bulik-Sullivan, Alexander Gusev, Gosia Trynka, Yakir Reshef, Po-Ru Loh, Verneri Anttila, Han Xu, Chongzhi Zang, Kyle Farh, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics*, 47(11):1228–1235, 2015.

[41] G. Wang, A. Sarkar, P. Carbonetto, and M. Stephens. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. B*, 82:1273–1300, 2020.

[42] Y. Okada, 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. H. Coenen, P. L. C. M. van Riel, M. A. F. J. van de Laar, H.-J. Guchelaar, T. W. J. Huizinga, P. Dieudé, X. Mariette, S. L. Bridges, A. Zhernakova, R. E. M. Toes, P. P. Tak, C. Miceli-Richard, S.-Y. Bang, H.-S. Lee, J. Martin, M. A. Gonzalez-Gay, L. Rodriguez-Rodriguez, S. Rantap"a"a-Dahlqvist, L. Arlestig, H. K. Choi, Y. Kamatani, P. Galan, and M. Lathrop. netics of rheumatoid arthritis contributes to biology and drug discovery. *Nature.*, 506:376–381, 2014.

[43] V. A. Laufer, H. K. Tiwari, R. J. Reynolds, M. I. Danila, J. Wang, J. C. Edberg, R. P. Kimberly, L. C. Kottyan, J. B. Harley, T. R. Mikuls, P. K. Gregersen, D. M. Absher, C. D. Langefeld, D. K. Arnett, and S. L. Bridges Jr. Genetic influences on susceptibility to rheumatoid arthritis in african-americans. *Human Molecular Genetics*, 28:858–874, 2019.

[44] C. M. Quinzii, L. C. López, A. Naini, S. DiMauro, and M. Hirano. Human coq 10 deficiencies. *BioFactors*, 32:113–118, 2008.

[45] N. Kubota and M. Suyama. An integrated analysis of public genomic data unveils a possible functional mechanism of psoriasis risk via a long-range errfi1 enhancer. *BMC Med Genomics.*, 13:8, 2020.

[46] T. Kume, H. Iwasa, H. Shiraishi, T. Yokoi, K. Nagashima, M. Otsuka, T. Terada, T. Takagi, A. Hara, and T. Kamataki. Characterization of a novel variant (s145c/l311v) of 3alpha-hydroxysteroid/dihydrodiol dehydrogenase in human liver. *Pharmacogenetics.*, 9:763–771, 1999.

[47] G. Greenbaum, W. M. Getz, N. A. Rosenberg, M. W. Feldman, E. Hovers, and O. Kolodny. Disease transmission and introgression can explain the long-lasting contact zone of modern humans and neanderthals. *Nat Commun.*, 10:5003, 2019.

[48] Nathan K Schaefer, Beth Shapiro, and Richard E Green. An ancestral recombination graph of human, neanderthal, and denisovan genomes. *Science Advances*, 7(29):eabc0776, 2021.

[49] Carly V Weiss, Lana Harshman, Fumitaka Inoue, Hunter B Fraser, Dmitri A Petrov, Nadav Ahituv, and David Gokhman. The cis-regulatory effects of modern human-specific variants. *Elife*, 10:e63713, 2021.

[50] Kashif Rafiq Zahid, Shiming Han, Fuling Zhou, and Umar Raza. Novel tumor suppressor spryd4 inhibits tumor progression in hepatocellular carcinoma by inducing apoptotic cell death. *Cellular Oncology*, 42(1):55–66, 2019.

[51] William K Scott, Felix Mba Medie, Felicia Ruffin, Batu K Sharma-Kuinkel, Derek D Cyr, Shengru Guo, Derek M Dykxhoorn, Robert L Skov, Niels E Bruun, Anders Dahl, et al. Human genetic variation in gls2 is associated with development of complicated staphylococcus aureus bacteremia. *PLoS genetics*, 14(10):e1007667, 2018.

[52] Juan Liu, Cen Zhang, Meihua Lin, Wei Zhu, Yingjian Liang, Xuehui Hong, Yuhan Zhao, Ken H Young, Wenwei Hu, and Zhaohui Feng. Glutaminase 2 negatively regulates the pi3k/akt signaling and shows tumor suppression activity in human hepatocellular carcinoma. *Oncotarget*, 5(9):2635, 2014.

[53] Benjamin C Haller and Philipp W Messer. Slim 3: forward genetic simulations beyond the wright–fisher model. *Molecular biology and evolution*, 36(3):632–637, 2019.

[54] Russell Lande. Natural selection and random genetic drift in phenotypic evolution. *Evolution*, pages 314–334, 1976.

[55] Adam Eyre-Walker. Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proceedings of the National Academy of Sciences*, 107(suppl 1):1752–1756, 2010.

[56] Jian Zeng, Ronald De Vlaming, Yang Wu, Matthew R Robinson, Luke R Lloyd-Jones, Loic Yengo, Chloe X Yap, Angli Xue, Julia Sidorenko, Allan F McRae, et al. Signatures of negative selection in the genetic architecture of human complex traits. *Nature genetics*, 50(5):746–753, 2018.