Mitochondrial genome copy number in human blood-derived DNA is strongly associated with insulin levels and related metabolic traits and primarily reflects cell-type composition differences

Liron Ganel[1,2], Lei Chen[1,2], Ryan Christ[1], Jagadish Vangipurapu[3], Erica Young[1,4], Indraniel Das[1], Krishna Kanchi[1], David Larson[1,5], Allison Regier[1,2], Haley Abel[1,2,5], Chul Joo Kang[1], Aki Havulinna[6,7], Charleston W. K. Chiang[8,9], Susan Service[10], Nelson Freimer[10], Aarno Palotie[6,11,12], Samuli Ripatti[6,12,13], Johanna Kuusisto[3,14], Michael Boehnke[15], Markku Laakso[3,14], Adam Locke[1,2], Nathan O. Stitziel[1,4,5], Ira M. Hall[1,2,16]

[1] McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO, United States of America

[2] Department of Medicine, Washington University School of Medicine, St. Louis, MO, United States of America

[3] Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland

[4] Department of Medicine, Cardiovascular Division, Washington University School of Medicine, St. Louis, MO, United States of America

[5] Department of Genetics, Washington University School of Medicine, St. Louis, MO, United States of America

[6] Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland

[7] Finnish Institute for Health and Welfare (THL), Helsinki, Finland

[8] Center for Genetic Epidemiology, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States of America

[9] Quantitative and Computational Biology Section, Department of Biological Sciences, University of Southern California, Los Angeles, CA, United States of America

[10] Center for Neurobehavioral Genetics, Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA, United States of America

[11] Analytical and Translational Genetics Unit (ATGU), Psychiatric & Neurodevelopmental Genetics Unit, Departments of Psychiatry and Neurology, Massachusetts General Hospital, Boston, MA, United States of America

[12] Broad Institute of MIT and Harvard, Cambridge, MA, United States of America

[13] Faculty of Medicine, University of Helsinki, Helsinki, Finland

[14] Department of Medicine, Kuopio University Hospital, Kuopio, Finland

[15] Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI, United States of America

[16] Department of Genetics, Yale University School of Medicine, New Haven, CT, United States of America

* Corresponding authors

E-mail: ira.hall@yale.edu (IMH); nstitziel@wustl.edu (NOS)

Abstract

Mitochondrial copy number is known to vary among humans and across tissues, and a population-based study of mitochondrial genome copy number (MT-CN) in blood – as estimated from genome sequencing data – observed strong (~54%) heritability. However, the genetic causes and phenotypic consequences of MT-CN variation in humans are not well-studied. Here, we studied MT-CN variation in blood-derived DNA from 19,184 Finnish individuals with deep cardiometabolic trait measurements using a combination of whole genome (N = 4,163) and exome sequencing (N = 19,034) data as well as imputed array genotypes (N = 17,718). We confirmed that MT-CN in blood is highly heritable (31% by GREML). We identified two loci in the nuclear genome that are significantly associated with MT-CN variation: a common variant at the *MYB-HBS1L* locus ($P = 1.6\times10^{-8}$), which has previously been associated with numerous hematological parameters; and a burden of rare variants in the *TMB1M1* gene ($P = 3.0\times10^{-8}$), which has been reported to protect against non-alcoholic fatty liver disease in model organisms. We also found that MT-CN is strongly associated with insulin levels ($P = 2.0\text{x}10^{-21}$), fat mass ($P = 4.5\text{x}10^{-16}$), and other related traits. Using a Mendelian randomization framework, we constructed a genetic instrument for MT-CN using penalized regression with adjustment for potentially confounding covariates and found a significant association with insulin levels, which suggests that our MT-CN measurement in blood may be causally related to metabolic syndrome. Finally, we computed our genetic instrument in UK Biobank participants and tested it against a set of cell count and cardiometabolic traits. We found significant associations between MT-CN and both neutrophil and platelet counts ($P = 1.8\times10^{-8}$ and $P = 1.2\times10^{-3}$). While the association between MT-CN and metabolic syndrome traits was replicated in the UK Biobank, adjusting for cell counts largely eliminated these signals, suggesting that MT-CN is actually a proxy measurement for neutrophil and platelet counts in its effect on metabolic syndrome. Taken together,

these results suggest that measurements of MT-CN in blood-derived DNA may primarily reflect differences in cell-type composition, and that these differences may be causally linked to insulin and related traits.

Author summary

The number of mitochondria per cell is variable between tissues and between individuals, and prior studies have shown that mitochondrial genome copy number in blood (MT-CN) – as estimated indirectly from sequencing of blood-derived DNA – is a genetically determined trait that varies among humans. We studied genetic data from approximately 19,000 Finnish individuals and showed that MT-CN is significantly associated with insulin and related metabolic traits, providing evidence that MT-CN levels play a role in determining these traits. Consistent with a previous study, we showed that genetics play a significant role in determining blood-derived MT-CN. We also found new evidence linking several regions of the genome to MT-CN, including a gene known to be associated with non-alcoholic fatty liver disease. Finally, we found that in the link between blood-derived MT-CN and metabolic syndrome, MT-CN likely represents the relative quantities of circulating immune cells, providing further evidence for the role of inflammation in metabolic syndrome.

Introduction

There are many reported links between mitochondrial content and cardiometabolic phenotypes in various tissues, including adipose [1–3], liver [1,4,5], skeletal muscle [1,6–10], and blood [11–17]. Traits associated with mitochondrial (MT) content include coronary heart disease (CHD), type 2 diabetes, and metabolic syndrome traits such as insulin sensitivity/resistance, obesity, and blood triglycerides. However, these studies have generally been limited by small sample sizes and low

statistical power. This, in addition to the use of heterogeneous mitochondrial quantification methods, has led to inconsistencies in the literature about the strength and directions of effect between mitochondrial content and metabolic syndrome traits. In the only large WGS study of mitochondrial genome copy number (MT-CN) known to us, Ding *et al*. used data from 2,077 Sardinians to estimate the heritability of MT-CN at 54% and detect significant associations between MT-CN and both waist circumference and waist-hip ratio, but not body mass index (BMI) [11].

Although variations in MT-CN measured from whole blood can in principle be attributed to either variability of MT copy number within cells or the cell type composition of the blood (given that different cell types have varying MT content [18–20]), the literature on this subject is inconclusive. Using CpG methylation data, a large (N = 11,443), low-coverage (1.7x autosomal; 102x mitochondrial) sequencing study of the link between MT-CN and major depressive disorder using buccal DNA from Chinese women concluded that variability of MT-CN from buccal swabs was not due to differences in cell type composition [21]. However, this study did not do a similar experiment in blood. Two small (N = 756 and N = 400) studies identified an association between MT content and CHD that they attributed to variable MT-CN within leukocytes, but they did not directly investigate the possibility of cell type composition being the true driver of the association [12,17]. For brevity, we will use the term “MT-CN” to refer to the underlying phenotype reflected by measuring this quantity for the remainder of this work, with these caveats.

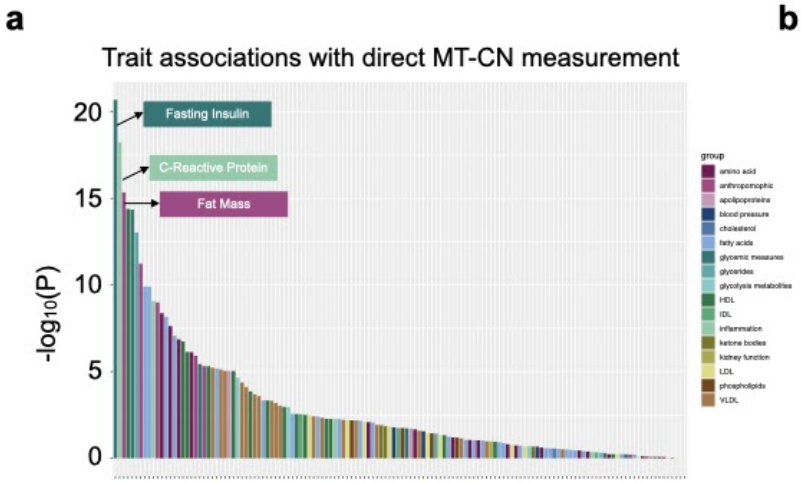
While several studies have found that peripheral blood MT content is heritable, only a small number of MT-CN associated loci have been identified [22,23]. In one of these studies, Curran *et al*. used linkage analysis in Mexican Americans to find an MT-CN associated locus near a marker previously associated with triglyceride levels [23–25], providing further indirect evidence for the link between MT-CN and metabolic syndrome.

110 Here, we take advantage of large-scale genome, exome, and array genotype data to
111 investigate the causes and effects of MT-CN in a large, deeply phenotyped Finnish cohort. Our
112 results reveal novel links with metabolic syndrome and provide evidence supporting a causal role for
113 MT-CN.
114

115 ## Results

116 ### Association of MT-CN with metabolic traits

117 We estimated MT-CN in 4,163 individuals from the METSIM and FINRISK studies based on
118 deep (>20x coverage) WGS data. We did so by measuring the mean coverage depth of reads
119 mapped to the mitochondrial genome in each sample, and normalizing it to the mean autosomal
120 coverage (see Methods). We performed batch normalization separately for METSIM and for two
121 FINRISK batches separated by survey years (see Methods). Each measurement was adjusted for
122 age, age^2, and sex, then inverse rank normalized separately before combining across batches. We
123 tested the resulting MT-CN estimates for association with 137 quantitative traits that were collected
124 and normalized according to the procedures described previously [26]. MT-CN was strongly
125 associated with fat mass ($P = 4.48 \times 10^{-16}$) and fasting serum insulin ($P = 2.02 \times 10^{-21}$), as well as
126 numerous additional quantitative traits, many related to metabolic syndrome (Fig 1a, S1 Table).
127 Notably, BMI was significantly associated with MT-CN, although Ding *et al*. did not find evidence of
128 this association [11]. Since population structure was a potential confounder in this analysis
129 considering the presence of mtDNA polymorphisms that might adversely affect short-read alignment,
130 we included SNP-inferred mitochondrial haplogroup as a covariate and reran the tests (Fig 1b). The
131 association signals retained significance even after this adjustment.



b

Trait	EMMAX model		Conditional on mitochondrial haplogroup	
	P	Beta	P	Beta
Fat mass	4.48×10^{-16}	-0.137	$< 2 \times 10^{-16}$	-0.171
ln(Fasting insulin)	2.02×10^{-21}	-0.182	$< 2 \times 10^{-16}$	-0.226

Fig 1. Cardiometabolic trait associations with MT-CN in WGS data. (a) Phenome-wide association study of normalized MT-CN against 137 cardiometabolic traits in the 4,163 sample data set. Traits are grouped into 17 categories, represented by the color of each bar. The top three most significant traits are, in order: fasting serum insulin, C-reactive protein, and fat mass. Exact P values and effect estimates, as calculated by EMMAX, are listed in S1 Table. (b) Association tests between normalized MT-CN and both fat mass and fasting serum insulin using WGS data (N = 4,163). Results are shown for the EMMAX test and a permutation test in which mitochondrial haplogroups were adjusted for.

To understand the connection between MT-CN and more clinically relevant phenotypes, we tested our MT-CN estimate against Matsuda ISI and disposition index (Table 1), which measure insulin sensitivity and secretion, respectively, and were not included in the initial screen. MT-CN was strongly associated with both insulin phenotypes. Notably, the Matsuda ISI signals survived adjustment for fat mass percentage after excluding diabetic individuals, which indicates that the association of peripheral blood MT-CN with insulin sensitivity was independent of fat mass.

To test for this association signal in a larger cohort, we developed a method to estimate mitochondrial genome copy number using 19,034 samples with whole exome sequencing (WES) data from the METSIM and FINRISK studies that included most of the WGS samples [26] (see Methods). Consistent with the WGS-based analysis, WES-estimated MT-CN was significantly associated with both fat mass and fasting serum insulin levels, even after removing the samples with WGS data, with

151 **Table 1. Associations of normalized MT-CN with disposition index and Matsuda ISI in METSIM.**

	Disposition index									
	Baseline					Follow-up[b]				
	N	β	SE	P	P*[a]	N	β	SE	P	P*[a]
All subjects	2975	0.094	0.004	3.0×10^{-7}	0.0004	2492	0.062	0.004	0.002	0.068
Excludes diabetic subjects at baseline	2842	0.091	0.003	1.3×10^{-6}	0.0007	2452	0.067	0.004	0.0009	0.041
Excludes diabetic subjects at baseline and during follow-up	2453	0.069	0.003	0.0007	0.023	2449	0.067	0.004	0.0009	0.042
	Matsuda ISI									
	Baseline					**Follow-up[b]**				
	N	**β**	**SE**	**P**	**P*[a]**	**N**	**β**	**SE**	**P**	**P*[a]**
All subjects	2975	0.192	0.005	4.3×10^{-26}	7.3×10^{-17}	2492	0.157	0.006	3.7×10^{-15}	8.7×10^{-10}
Excludes diabetic subjects at baseline	2842	0.191	0.005	1.0×10^{-24}	2.4×10^{-16}	2452	0.161	0.006	1.3×10^{-15}	3.0×10^{-10}
Excludes diabetic subjects at baseline and during follow-up	2453	0.173	0.005	7.2×10^{-18}	5.3×10^{-12}	2449	0.16	0.006	1.7×10^{-15}	3.8×10^{-10}

152 Testing was done by linear regression using disposition and Matsuda ISI, respectively, as the dependent variable.

153 [a]P* columns represent the P value from linear regression with additional adjustment for fat mass.

154 [b]Follow-up measurements were taken at a later time point.

155

156 identical directions of effect (S2 Table).

157 Anecdotally, it is interesting to note that these MT association signals can also be detected
158 using read-depth analysis of the nuclear genome (S1 Fig; manuscript in prep.), where reads derived
159 from mtDNA align erroneously to several nuclear loci based on homology between the MT genome
160 and ancient nuclear mitochondrial insertions. This result provides additional evidence for the reported
161 trait associations using an independent MT-CN estimation method, and indicates that these
162 homology-based signals need to be taken into account in future CNV association studies.

163

164 **Heritability analysis**

165 To assess the extent to which MT-CN is genetically determined, we estimated the heritability of
166 mitochondrial genome copy number using GREML (Table 2). We explored two different approaches
167 available: (1) analysis of the 4,149 samples with WGS data that passed quality control measures,
168 where both nuclear genotypes and MT-CN are measured directly from the WGS data, and (2)
169 analysis of the set of 17,718 samples with imputed genotype array data, where MT-CN is estimated
170 from WES data. Of these, (1) benefited from more accurate measurement of genotype and
171 phenotype, whereas (2) had noisier measurements but benefited from larger sample size. We
172 focused primarily on the METSIM cohort, both because of the homogeneity of this cohort (see
173 Methods) and because the number of FINRISK samples with WGS data was small.

174

175 **Table 2. GREML heritability estimates in each cohort separately and in joint analysis.**

		WGS-measured MT-CN			WES-measured MT-CN		
		N[a]	h^2	SE	N[a]	h^2	SE
Joint Analysis	WGS genotypes	4149	0.17	0.06	3916	0.11	0.06
	Imputed genotypes	3916	0.16	0.06	17718	0.09	0.02
METSIM	WGS genotypes	3065	0.31	0.07	2974	0.20	0.08
	Imputed genotypes	2974	0.27	0.08	9791	0.11	0.03
FINRISK	WGS genotypes	1084	0.20	0.22	942	0.24	0.27
	Imputed genotypes	942	0.35	0.27	7927	0.08	0.03

176 [a]All analyses in this table were limited to sample sets with available imputed genotype data, yielding slightly lower sample
177 sizes than in other tables.

178

In the WGS analysis, the GREML-estimated heritability of MT-CN in METSIM was 31% - somewhat less than the 54% value reported in the only prior large-scale study of peripheral blood MT-CN heritability, which was based on low-coverage WGS [11]. For comparison, we used this same approach to estimate heritability of LDL in METSIM WGS data, which yielded an estimate of 34% with a standard error of 7.9% (Table 3). This is broadly consistent with prior work [27,28], including analysis of the same Finnish sample set using distinct methods [26] (20.2% heritability). These results show that mitochondrial genome copy number is a genetically determined trait with significant heritability, comparable to that of LDL and other quantitative cardiometabolic traits [26].

Table 3. GREML and GREML-LDMS heritability estimates for normalized MT-CN and low-density lipoprotein (LDL).

Trait	Genotype source (phenotype source)	N[a]	GREML		GREML-LDMS[b]	
			h^2	SE	h^2	SE
Normalized MT-CN	WGS (WGS-measured MT-CN)	3065	0.31	0.07	0.31	0.09
	Imputed array[c] (WES-measured MT-CN)	6789	0.11	0.04	0.14	0.05
LDL	WGS	3062	0.34	0.08	0.38	0.10
	Imputed array[c]	6787	0.25	0.04	0.32	0.05

[a]All analyses are limited to METSIM data

[b]GREML-LDMS heritability estimates are calculated using PCs 1-10 as fixed-effect covariates.

[c]Analyses of imputed array data exclude samples with WGS data

The analysis of imputed METSIM genotypes using WES-estimated MT-CN yielded an estimated heritability of 11%, which is much lower than the WGS-based estimate (Table 2). To understand this discrepancy, we repeated the GREML analysis with the other two combinations of phenotype source (WGS vs. WES estimation) and genotype source (WGS vs. imputed array). When using the WGS-measured phenotype, the estimated heritability decreased only slightly (31% to 27%)

when switching from the WGS to imputed genotypes. This suggests that the difference in genotyping method was not the main driver of the observed heritability disparity between the WGS and imputed array datasets. Conversely, when analyzing the imputed METSIM genotypes, switching from WGS-measured to WES-measured MT-CN resulted in a large drop (27% to 11%) in estimated heritability. This suggests that the extra noise inherent in WES-based MT-CN estimates was responsible for the reduction in the GREML-estimated heritability despite the increased sample size of the imputed array dataset.

Identification of genetic factors associated with MT-CN

Previous studies have identified three autosomal quantitative trait loci (QTL) for MT-CN in other populations [22,23]. We conducted single variant GWAS for MT-CN (see Methods). Analysis of WGS (N = 4,149) and WES (N = 19,034) genotypes yielded no variants exceeding the respective significance thresholds of 5×10^{-8} and 5×10^{-7} (S2 Fig). However, despite the increased noise in the WES-measured phenotype, GWAS of imputed array genotypes from METSIM (N = 9,791) yielded two loci with genome-wide significant associations, identified by lead markers rs2288464 and rs9389268 (Fig 2, Table 4). Of the previously reported QTLs, we observed a suggestive signal at rs445 ($P = 7.95\times10^{-4}$), but no signal at rs11006126 (P = 0.206) or within the linkage peak reported on

Table 4. Single marker association results for rs2288464 and rs9389268

		FINRISK[a]				METSIM				Joint Analysis			
		N	MAF	P[b]	Beta	N	MAF	P[b]	Beta	N	MAF	P[b]	Beta
rs2288464	Imputed	7927	0.148	0.613	0.0113	9791	0.165	**2.55×10^{-9}**	0.119	17718	0.158	9.77×10^{-7}	0.075
	WES	9221	0.150	0.376	0.0186	9813	0.166	**6.75×10^{-9}**	0.118	19034	0.158	9.34×10^{-7}	0.0734
	WGS	1084	0.142	0.383	0.0532	3065	0.161	0.113	0.0561	4149	0.156	0.0655	0.0562
rs9389268	Imputed	7927	0.354	0.189	0.0216	9791	0.347	**1.26×10^{-10}**	0.0973	17718	0.35	**1.62×10^{-8}**	0.0634
	WGS	1084	0.351	0.788	0.0121	3065	0.347	**3.24×10^{-8}**	0.150	4149	0.348	7.87×10^{-7}	0.115

[a]Analyses of imputed FINRISK array data were performed with covariates for FINRISK genotyping batch.

[b] Bolded results are significant at the appropriate threshold for the given test (see Methods).

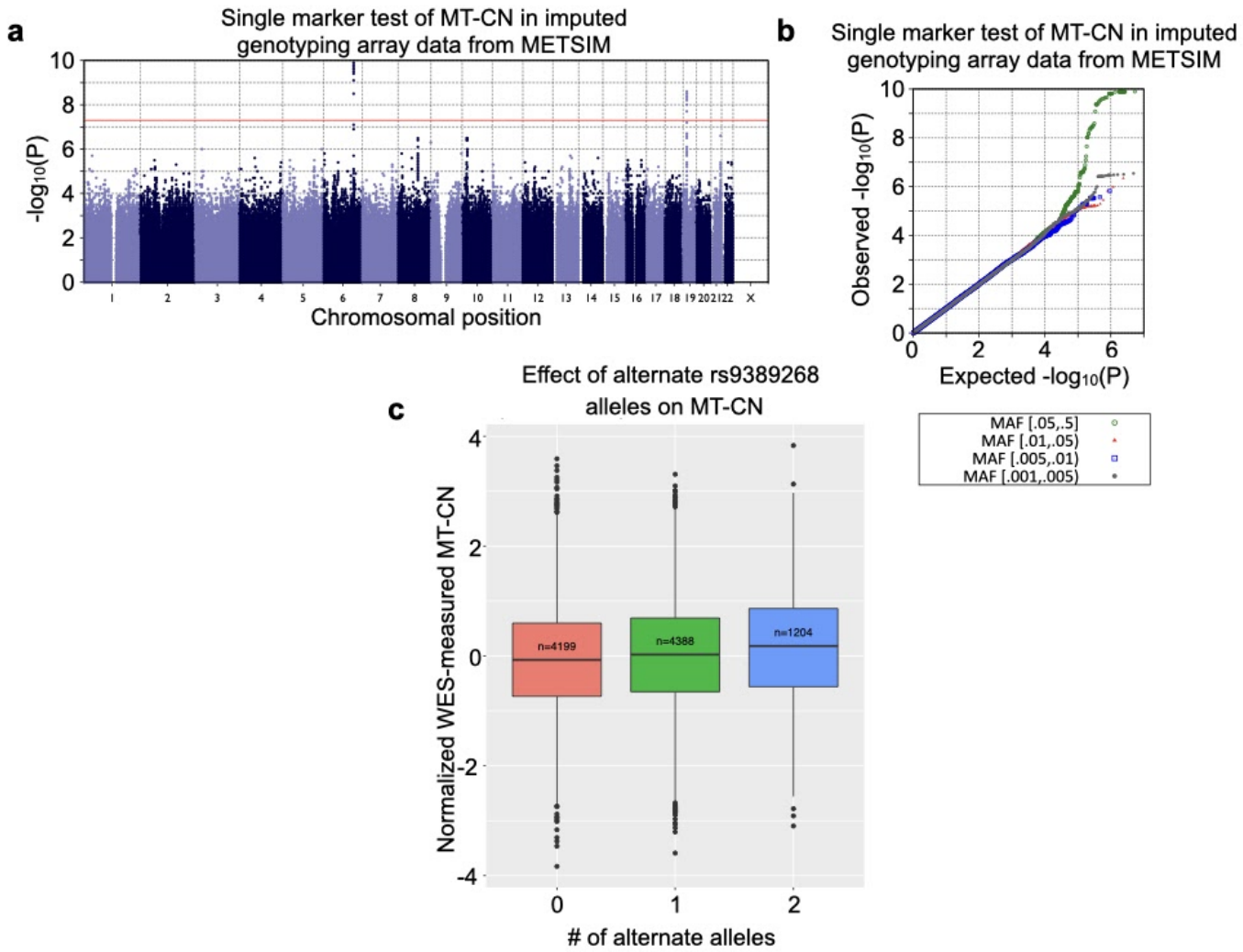


Fig 2. Single-marker genetic associations with MT-CN in imputed array data. (a) Manhattan plot for a genome-wide association test of normalized, WES-measured MT-CN using imputed array genotype data from METSIM (N = 9,791). Two loci markers reached the genome-wide significance of 5×10^{-8}, identified by lead markers rs2288464 and rs9389268. (b) Quantile-quantile (QQ) plot for the association test shown in (a). This plot is separated by minor allele frequency bin, as indicated by the colors and shapes of the points. (c) Boxplot showing the distributions of normalized WES-measured MT-CN in METSIM separated by the number of rs9389268 alternate alleles as detected by imputed array genotyping (N = 9,791). The EMMAX P value for this variant was 1.62×10^{-8} in imputed data.

chromosome 10 (S3 Fig) [22,23].

rs9389268 was the only marker that was strongly associated with MT-CN in the METSIM analyses of both WGS and imputed array data ($P = 7.87\times10^{-7}$ and $P = 1.62\times10^{-8}$, respectively). Although this variant was not significantly associated with MT-CN in FINRISK or in a random-effect

231 meta-analysis of both cohorts (P = 0.115), the lack of signal in FINRISK is likely the product of
232 lower-quality MT-CN measurements in FINRISK, which displayed heterogeneity across survey years
233 (S4 Fig). This variant is located in an intergenic region between the *MYB* and *HBS1L* genes, is
234 common across many populations, and is slightly more frequent in Finns compared to non-Finnish
235 Europeans (gnomAD v3 MAF 34.4% vs. 26.0%). *MYB* and *HBS1L* are hematopoietic regulators
236 [29,30], and the region between them is known to be associated with many hematological parameters
237 including fetal hemoglobin levels, hematocrit, and erythrocyte, platelet, and monocyte counts [31–34].
238 It has been suggested that these intergenic variants function by disrupting *MYB* transcription factor
239 binding and disrupting enhancer-promoter looping [35]. Conditioning the METSIM-only imputed array
240 GWAS on rs9399137 – a tag SNP shown to be associated with many of these hematological
241 parameters [33] – resulted in elimination of the rs9389268 signal entirely (P = 0.408), suggesting that
242 the haplotype responsible for the association of rs9389268 with MT-CN in our data is the same one
243 previously known to be associated with numerous hematological phenotypes.

244 This result is not surprising considering our approach for normalizing MT-CN. Because our
245 MT-CN estimate was based on the ratio of mtDNA coverage to nuclear DNA coverage, changes in
246 the cell type composition of blood could result in changes in our normalized measurement if the
247 underlying cell types have different average numbers of mitochondria. This is especially true of
248 platelets, which can contain mitochondria but not nuclei, and whose counts are known to be
249 associated with rs9399137.

250 rs2288464 seemed to be a good candidate due to its location in the 3' untranslated region of
251 *MRPL34*, which codes for a large subunit protein of the mitochondrial ribosome. While the association
252 signal at this marker was not observed in the WGS data (P = 0.0655), based on the observed effect
253 size of this variant in WES and imputed data as well as the number of WGS datasets available, there

254 was insufficient power (~0.5% at $\alpha = 5\times10^{-7}$) to robustly detect this association in the WGS data alone
255 [36].

256 We next performed rare variant association (RVAS) analyses using a mixed-model version of
257 SKAT-O [37] to test for genes in which the presence of high-impact rare variants might be associated
258 with MT-CN levels (see Methods; Fig 3, Table 5). Using WES data, the only gene passing the
259 Bonferroni-adjusted P value threshold of 2.16×10^{-6} was *TMBIM1* ($P = 2.96\times10^{-8}$), a member of a gene
260 family thought to regulate cell death pathways [38]. *TMBIM1* has been shown to be protective against
261 non-alcoholic fatty liver disease (NAFLD), progression to non-alcoholic steatohepatitis, and insulin
262 resistance in mice and macaques [39]. Interestingly, in our analysis, in which a burden test was
263 determined to be optimal by SKAT-O, putative loss-of-function variants in *TMBIM1* were associated
264 with a higher MT-CN (Fig 3c), which was in turn associated with less severe metabolic syndrome,
265 suggesting that *TMBIM1* is actually a risk gene, not a protective one. Thus, the published function of
266 *TMBIM1* makes it a strong candidate, although the direction of effect in our data disagreed with the
267 direction suggested by prior work in model organisms [39].

268 **Table 5. Gene-based rare variant association results for *TMBIM1***

	FINRISK			METSIM			Joint Analysis		
	N	Fraction with rare allele	P[a]	N	Fraction with rare allele	P[a]	N	Fraction with rare allele	P[a]
WES	9221	0.016	1.57×10^{-3}	9813	0.013	**1.44×10^{-6}**	19034	0.014	**2.96×10^{-8}**
WGS	1084	0.028	0.489	3065	0.014	0.01	4149	0.013	0.01

269 *TMBIM1* was the only genome-wide significant gene in the WES rare-variant association tests of METSIM and the whole
270 dataset.

271 [a] Bolded results are significant at the appropriate threshold for the given test (see Methods).

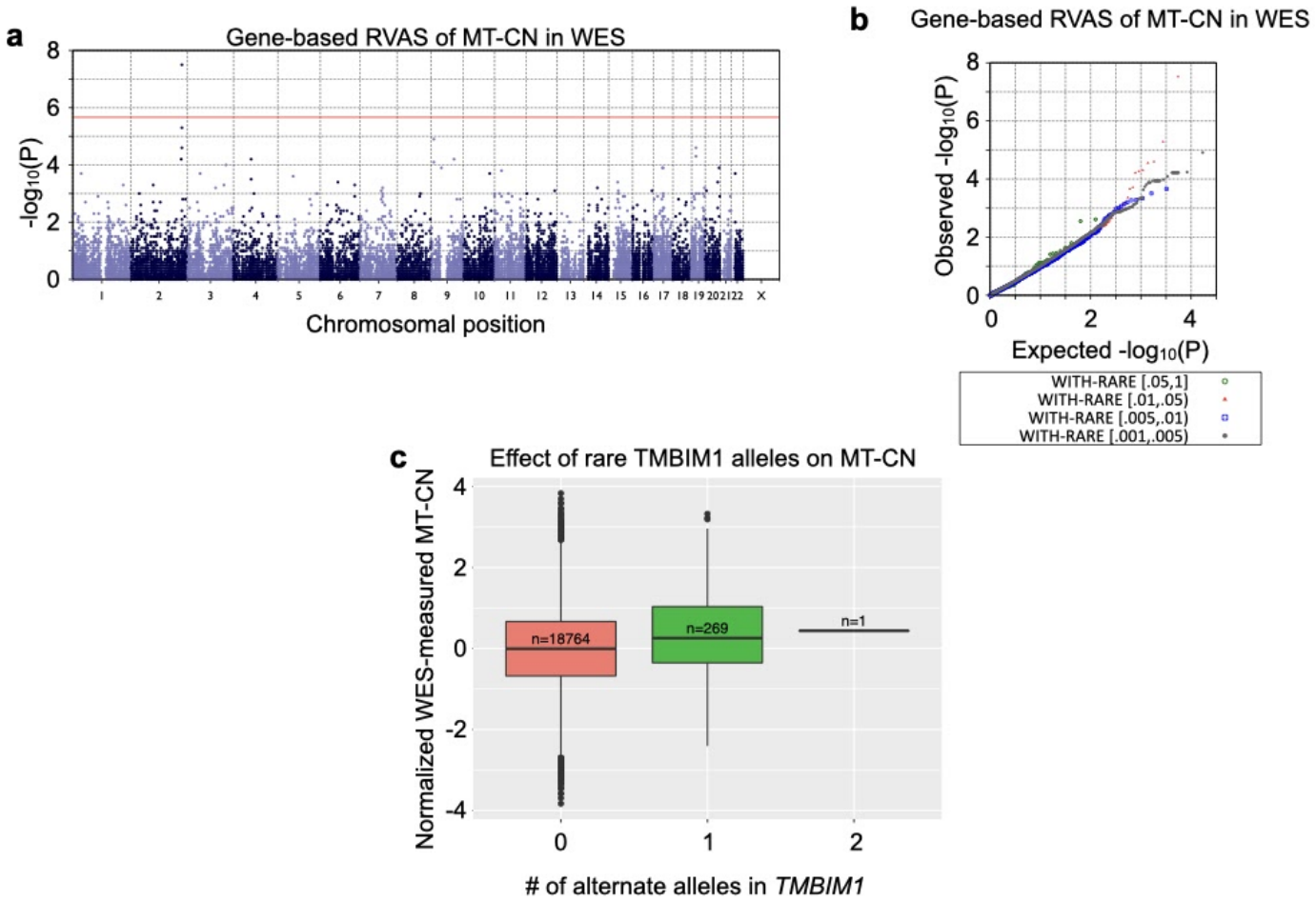


Fig 3. Gene-based associations with MT-CN in WES data. (a) Manhattan plot and quantile-quantile (QQ) plot for a gene-based rare variant association test of normalized, WES-measured MT-CN using WES data from both METSIM and FINRISK (N = 19,034). The red line represents a Bonferroni significance level of 2.164×10^{-6}, as 23,105 genes were included in this test. *TMBIM1* is the only gene to reach significance at this level. (b) QQ plot for the test shown in (a). This plot is separated by minor allele frequency bin, as indicated by the colors and shapes of the points. (c) Boxplot showing the distributions of normalized WES-measured MT-CN, separated by the number of WES-detected alternate alleles in *TMBIM1* with MAF < 0.01 and CADD score > 20 (N = 19,034).

Inference of causality in the association between MT-CN and insulin

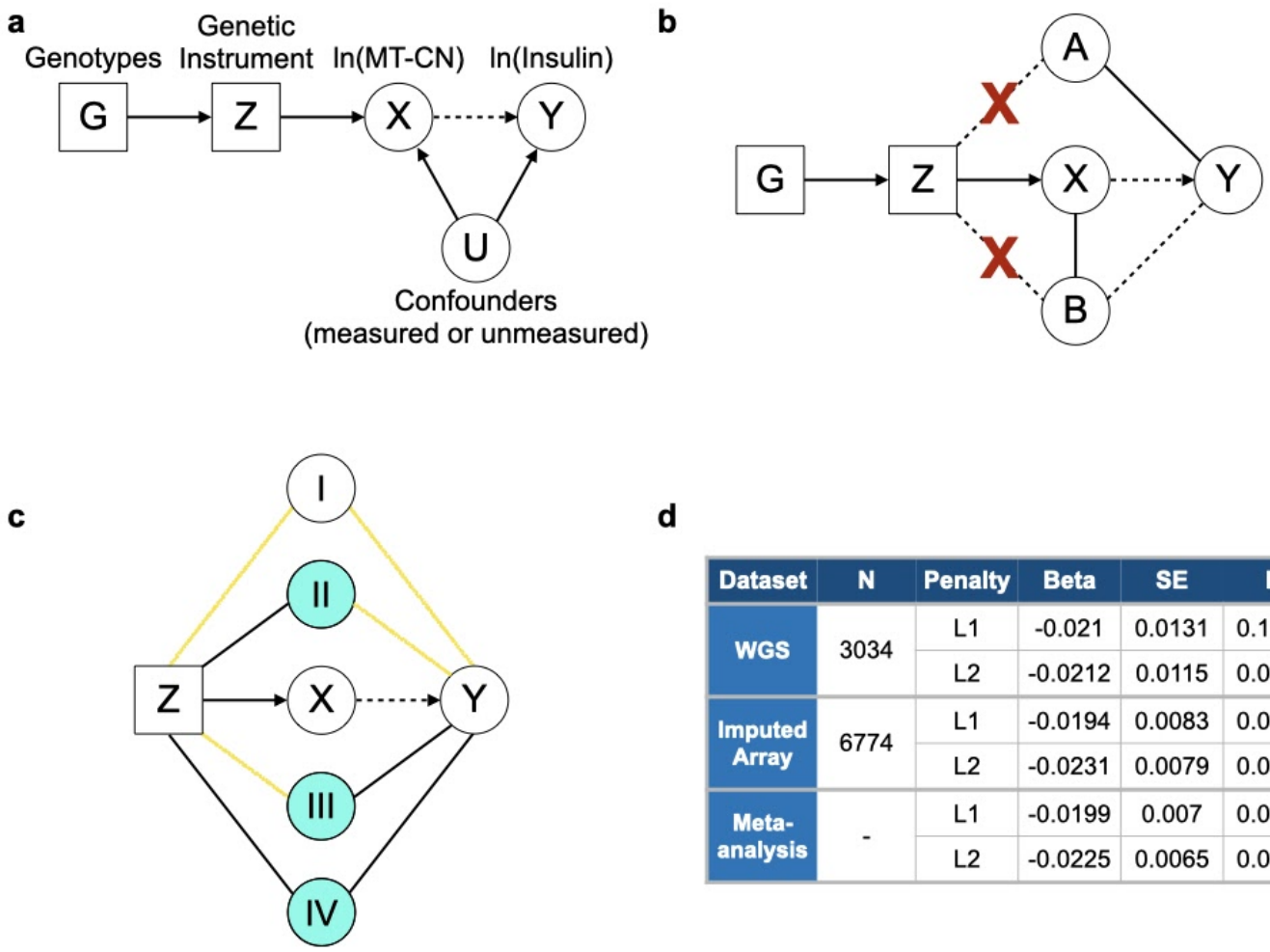
To further understand the association between MT-CN and fasting serum insulin, we employed a Mendelian randomization (MR) approach with MT-CN as the exposure and insulin as the outcome. Using penalized regression, we leveraged our extensive phenotype data to build a genetic instrument

284 from a large number of genetic variants and adjust for possible confounders via a novel approach
285 (see Methods; Fig 4). We believe this approach to be more robust to violations of key MR
286 assumptions than other methods in situations where limited data are available and few robust
287 genotype-exposure associations are known. We restricted our analysis to METSIM samples due to
288 batch effects and inconsistencies in available quantitative trait data observed across FINRISK survey
289 years (S4 Fig). The effect sizes of the instrument in the causality test for insulin levels are shown in
290 Fig 4d. We calculated our instrument using either L1 or L2 regularization. In both cases, the MT-CN
291 instrument was not a significant predictor ($\alpha = 0.05$) of insulin when we constructed our instrument
292 from WGS variants, but was significant when the instrument was constructed from imputed array
293 variants. This was likely due to the larger sample size of the imputed array data set. However, the
294 effect estimates were remarkably similar across all four cases. As a result, inverse-variance weighted
295 meta-analysis across datasets yielded highly significant P values for both penalties. In summary, our
296 analysis provided evidence for a significant causal role for MT-CN in determining fasting serum insulin
297 levels that was robust to the choice of regression penalty when building the genetic instrument. We
298 note that this evidence for causality comes with some caveats (see Methods).
299

300 **Replication and biological interpretation**

301 In principle, changes in MT-CN can be caused by changes in the number of mitochondrial
302 genome copies within cells or by changes in the blood cell type composition. Based on the
303 association with rs9389268 and the nuances of the normalization procedure described above, we
304 sought to test the hypothesis that our MT-CN measurement primarily reflects the cell type
305 composition of the blood rather than the number of mitochondria per cell. We used imputed array
306 genotype and phenotype data from the UK Biobank (N = 357,656) for this purpose [40].

307 We first tested cell counts from the UK Biobank (UKBB) against a polygenic risk score (PRS)



Dataset	N	Penalty	Beta	SE	P
WGS	3034	L1	-0.021	0.0131	0.1087
		L2	-0.0212	0.0115	0.0645
Imputed Array	6774	L1	-0.0194	0.0083	0.0192
		L2	-0.0231	0.0079	0.0037
Meta-analysis	-	L1	-0.0199	0.007	0.0045
		L2	-0.0225	0.0065	0.0006

Fig 4. Mendelian randomization approach and results. (a) Formulation of the Mendelian Randomization causality test. G represents genotypes, Z is a genetic instrument value constructed from G, X represents ln(MT-CN), Y represents ln(Insulin), and U represents any confounders of the association between X and Y. The arrow from X to Y is dashed to indicate that although an association is known, the relationship is not known to be causal. In this formulation, a significant association between Z and Y would provide evidence that X is casual for Y. (b) Strategy for choosing variables to adjust for when building Z in order to enforce MR assumptions. A represents those columns of covariate matrix W that are associated with Y (represented by the solid line between A and Y) and B represents those columns of matrix W that are associated with X conditional on A (represented by the solid line between B and X). Dashed lines represent possible, but unproven associations. The penalized regression of X on G used to build Z is adjusted for A and B (with no penalty) in an attempt to prevent any associations between Z and either A or B (represented by the blue X's). While an association between B and Y is unlikely (represented by the dashed line between B and Y) because B is not contained in A, B is still adjusted for in the penalized regression to be as conservative as possible. (c) Strategy for choosing covariates to adjust for in the causality test of Y against Z in another attempt to reduce the impact of any remaining associations between Z and assumption-violating variables. Covariate sets I, II, III, and IV are defined by the presence of known first-order

associations (represented by black lines) with Z and Y (see Methods). Yellow lines represent relationships where a first-order association is not known, but a higher-order association is possible. Covariate sets II, III, and IV (colored blue) are adjusted for in the causality test because there is at least one first order association linking them to Z and Y, so they risk violating MR assumptions 2 or 3. (d) Results of Mendelian randomization test for causality of MT-CN on fasting serum insulin.

for MT-CN built using the genetic instrument from the Finnish data. Leukocyte, neutrophil, and platelet counts were all significantly associated with MT-CN PRS conditional on age, age^2, and sex (see Methods, Table 6). However, adjusting for neutrophil counts in the leukocyte regression eliminated the signal (PRS regression coefficient P = 0.839), suggesting that the leukocyte count signal was driven by the effect of neutrophil count. We removed any high leverage, large residual samples and repeated the neutrophil and platelet count regressions to ensure that this result was robust to outliers and found no appreciable change in significance (Table 6). As a result, we concluded that our MT-CN measurement was significantly associated with neutrophil and platelet counts. Subsequent analyses were performed both with and without adjustment for these variables, as described below.

Table 6. Association results between blood cell count traits and MT-CN polygenic risk score in 357,656 UK Biobank samples.

Cell Type	All samples			No *post hoc* high leverage outliers		
	β[a]	SE	P[b]	β[a]	SE	P[b]
Leukocyte	-0.00856	0.00170	**4.42×10^{-7}**	-	-	-
Monocyte	-0.00119	0.00170	0.482	-	-	-
Lymphocyte	-0.00250	0.00170	0.142	-	-	-
Neutrophil	-0.00954	0.00169	**1.80×10^{-8}**	-0.00948	0.00169	**2.17×10^{-8}**
Platelet	0.00548	0.00170	**1.24×10^{-3}**	0.00548	0.00170	**1.25×10^{-3}**

[a]β refers to the regression coefficient of MT-CN in a linear regression of cell type onto MT-CN PRS and other covariates (see Methods)

[b]P values below a nominal α = 0.05 are shown in bold

We note that the effect directions of the associations of platelet counts with metS and MT-CN PRS seem inconsistent at first glance, as platelet counts were positively correlated with MT-CN PRS (Table 6) and metS (S3 Table) while MT-CN and insulin (a proxy for metS) were negatively correlated (Fig 1b). However, the FinMetSeq regression model in Fig 1b was not conditional on any other covariates (although age, age^2, and sex were regressed out of the MT-CN measurement prior to this analysis), while the UKBB models that gave rise to Table 6 and S3 Table adjusted for many additional covariates, including 20 PCs and age-sex interaction terms. As a result, the effect directions for the analyses in the two datasets are not directly comparable.

We next tested for associations between MT-CN PRS and several cardiometabolic phenotypes from the test in Fig 1a (see Methods). With the exception of C-reactive protein, which showed no significant association, all tested phenotypes showed nominal association with MT-CN PRS at $\alpha = 0.05$, with total triglycerides and HDL being the only traits surviving Bonferroni correction (Table 7). We interpret this as replication of the link between mitochondrial genome copy number and metabolic syndrome in a large, independent data set.

Table 7. Association results between metabolic syndrome traits and MT-CN polygenic risk score in 357,656 UK Biobank samples.

Trait	Without platelet and neutrophil adjustment			With platelet and neutrophil adjustment		
	β[b]	SE	P[c]	β[b]	SE	P[c]
Type 2 Diabetes	-0.1681	0.0826	**0.0419**	-0.1467	0.0844	0.0823
BMI	-0.0372	0.0175	**0.0331**	-0.0233	0.0175	0.1836
Fat Mass	-0.0452	0.0176	**0.0100**	-0.0318	0.0177	0.0716
C-Reactive Protein	-0.0015	0.0178	0.9305	0.0206	0.0171	0.2291
HDL	0.0573	0.0187	**0.0021**	0.0416	0.0187	**0.0262**
Total Triglycerides	-0.0500	0.0176	**0.0046**	-0.0330	0.0175	0.0603
Weight[a]	-0.0359	0.0176	**0.0414**	-0.0218	0.0178	0.2189

[a]The weight phenotype tested was that which was measured at the time of impedance measurement.

[b]β refers to the regression coefficient of MT-CN in a linear regression of cell type onto MT-CN PRS and other covariates (see Methods)

[c]P values below a nominal α = 0.05 are shown in bold

To determine whether there was any association between MT-CN and metabolic syndrome not mediated through cell counts, we repeated the tests of cardiometabolic trait association with MT-CN PRS with adjustment for platelet and neutrophil counts. HDL was the only trait with a nominal (α = 0.05) association with PRS under this adjustment, but this signal was not strong enough to survive Bonferroni correction (Table 7). This suggests that the associations we observed between MT-CN and metabolic traits arose simply because MT-CN is a proxy for platelet and neutrophil count. This was supported by the fact that direct testing of platelets and neutrophils against triglycerides, fat mass, and HDL yielded remarkably significant associations, which survived *post-hoc* removal of high-leverage, high-residual outlier samples (S3 Table). This evidence for MT-CN as a proxy for platelet and neutrophil counts strongly suggests that the causal relationship observed in the Mendelian randomization experiment (see above) in fact represents a causative role for neutrophils and platelet counts in setting serum insulin levels.

Given the strong observed associations between blood cell count phenotypes and MT-CN PRS, we used these blood phenotypes to seek replication of the genetic associations detected in Finnish data. Using imputed UKBB genotype data, we tested the expected alternate allele dosage of both rs2288464 and rs9389268 against the same blood cell traits mentioned above, using linear regression (S4 Table; expected alternate allele dosage was calculated from genotype call probabilities as $DS = P(0/1) + 2P(1/1)$). As expected given its known associations with multiple hematological parameters (see above), rs9389268 showed strong associations with all tested blood cell phenotypes. rs2288464 was not significantly associated with any of the five phenotypes after

correction for multiple testing, although a nominal association was detected with total leukocyte count. This further strengthens our belief that rs9389268 is truly associated with MT-CN through blood cell composition. We also tested *TMBIM1* against the same blood cell traits using SKAT-O [37], and found no significant associations (S4 Table). This may mean that *TMBIM1* affects MT-CN through a mechanism other than altering blood cell type composition.

We further sought to generate hypotheses for other phenotypic associations with MT-CN. To this end, we performed a phenome-wide screen of MT-CN PRS against all of the UKBB phenotypes available to us. To curate and transform these phenotypes, we used a modified version of PHESANT [41,42], which outputs all continuous variables in both raw and inverse rank-normalized form. We chose to interpret the results from the normalized continuous variables (S5 Table) to be conservative and robust to outliers, although the results of the raw continuous variable analyses were similar (S6 Table). No metabolic syndrome traits appeared among the tested traits with $q < 0.05$. However, the tests for HDL cholesterol, self-reported heart attack, and doctor-diagnosed heart attack did yield somewhat suggestive results (q = 0.123, 0.176, and 0.176, respectively). We also repeated this screen with adjustment for neutrophil and platelet counts (S7 Table and S8 Table), resulting again in no metabolic syndrome phenotypes achieving $q < 0.05$. The addition of neutrophil and platelet counts as covariates attenuated the suggestive signals for HDL cholesterol, self-reported heart attack, and doctor-diagnosed heart attack (q = 0.284, 0.391, and 0.402, respectively).

Discussion

We have described one of the most well-powered studies to date of the genetic relationship between MT-CN measurements in blood and cardiometabolic phenotypes. Our study is one of very few of which we are aware to utilize sequencing data for this purpose, and our samples are more deeply phenotyped than those used in prior work. Our data show highly significant associations

410 between blood-derived MT-CN measurements and several cardiometabolic traits, particularly insulin
411 and fat mass. We observed strong heritability of MT-CN (31%), on par with other widely studied
412 cardiometabolic traits such as LDL, and identified one single marker association on a haplotype
413 previously associated with several hematological parameters [31–34]. We also report one gene with a
414 rare-variant association with MT-CN, *TMBIM1*, that has a known link to non-alcoholic fatty liver
415 disease in model organisms [39]. Interestingly, rare variants in this gene were not associated (in
416 aggregate) with blood cell counts in UK Biobank, suggesting that this gene might modulate NAFLD
417 risk via MT-CN, but through a mechanism not involving blood composition. More work must be done
418 to confirm this tentative genetic association and mechanistic hypothesis.

419 Using a novel multiple-variant instrument-building method, we report evidence from Mendelian
420 randomization supporting a causal role for MT-CN in metabolic syndrome. Further, we used UK
421 Biobank data to show that not only does the link between MT-CN and metabolic syndrome replicate in
422 an independent data set using a polygenic risk score approach, but also that this association is
423 mediated by neutrophil and platelet counts, contrary to previous claims that variability in the number
424 of mitochondria per cell is responsible for CHD risk [12]. Taken together, these results suggest that
425 neutrophil and platelet counts in peripheral blood are causally related to fasting insulin and related
426 traits.

427 There is prior evidence to support the role of inflammation – specifically via innate immune
428 cells such as neutrophils – in the etiology of type 2 diabetes (T2D) and insulin resistance [43–45],
429 which suggests a plausible model by which peripheral blood neutrophil count could influence
430 metabolic syndrome. Nutrient excess and high fat diets are known to recruit neutrophils into tissues,
431 which then cause insulin resistance both by releasing TNF-α and IL-6 and by upregulating
432 cyclooxygenase [43]. This leads to increased LTB4 and subsequent upregulation of NF-κB, a central
433 regulator of inflammation. Moreover, free fatty acids also cause neutrophils to stay in tissues longer,

resulting in persistent inflammation and leading to insulin resistance [43]. While it is known that inflammation, and particularly neutrophils, play a role in metabolic syndrome, our results strongly suggest that peripheral blood neutrophil count causally contributes to this process and is associated with heritable genetic variation in the human population. Overall, our work provides further insight into the role that inflammation plays in metabolic syndrome and supports the idea that targeting inflammation may be a fruitful avenue of investigation in developing future therapeutics.

Methods

Genotype and phenotype data

Whole genome sequencing (WGS) was performed on a cohort of 4,163 samples comprising 3,074 male samples from the METSIM study [46] and 1,089 male and female samples from the FINRISK study [47]. Genomic DNA was fragmented on the Covaris LE220 instrument targeting 375 bp inserts. Automated Illumina libraries were constructed with the TruSeq PCR-free (Illumina) or KAPA Hyper PCR-free library prep kit (KAPA Biosystems/Roche) on the SciClone NGS platform (Perkin Elmer). The fragmented genomic DNA was size-selected on the SciClone instrument with AMPure XP beads to tighten the distribution of fragmented DNA to ensure the average insert of the libraries was 350-375 bp. We followed the manufacturer’s protocol as provided by Perkin Elmer, with the following exception: post ligation, the libraries were purified twice with a 0.7x AMPure bead/sample ratio to eliminate any residual adaptors present. An aliquot of the final libraries was diluted 1:20 and quantitated on the Caliper GX instrument (Perkin Elmer). The concentration of each library was accurately determined through qPCR utilizing the KAPA library Quantification Kit according to the manufacturer's protocol (KAPA Biosystems/Roche) to produce cluster counts appropriate for the Illumina HiSeqX instrument. Libraries were pooled and run over a few lanes of the HiSeq X to ensure the libraries within the pool were equally balanced. The final pool of balanced

libraries was loaded over the remaining number of HiSeq X lanes to achieve the desired coverage for this project. 2x150 paired end sequence data were demultiplexed using a single index, which was a restriction on the HiSeqX instrument at this time. A minimum of 19.5x coverage was achieved per sample.

The quality of the aligned sequence data was assessed using metrics generated by Picard [48] v2.4.1, Samtools v1.3.1 (http://www.htslib.org/) and VerifyBamID v1.1.3 (https://genome.sph.umich.edu/wiki/VerifyBamID). Based on the output files from Picard, the following alignment statistics were collected for review: PF_MISMATCH_RATE, PF_READS, PF_ALIGNED_BASES, PCT_ADAPTER, PCT_CHIMERAS, PCT_PF_READS_ALIGNED, PCT_READS_ALIGNED_IN_PAIRS, PF_HQ_ALIGNED_BASES, PF_HQ_ALIGNED_Q20_BASES, PF_HQ_ALIGNED_READS, MEAN_INSERT_SIZE, STANDARD_DEVIATION, MEDIAN_INSERT_SIZE, TOTAL_READS, PCT_10x, and PCT_20x. Alignment rate was calculated as PF_READS_ALIGNED/TOTAL_READS. The formula for haploid coverage was as follows: $Haploid\ coverage = MEAN_COVERAGE * \frac{1-PCT_EXC_DUPE}{1-PCT_EXC_TOTAL}$. From the Samtools output, inter-chromosomal rate was calculated as: $\frac{reads_mapped_in_interchromosomal_pairs}{reads_mapped_in_pair}$ and discordant rate was calculated as: $reads_mapped_percentage - reads_mapped_in_proper_pairs_percentage$. Properly paired percentage (reads_mapped_in_proper_pairs_percentage) and singleton percentage (reads_mapped_as_singleton_percentage) were also reviewed. From VerifyBamID, the Freemix value was reviewed.

The metrics for judgement of passing data quality were: FIRST_OF_PAIR_MISMATCH_RATE < .05, SECOND_OF_PAIR_MISMATCH_RATE < 0.05, haploid coverage ≥ 19.5, interchromosomal rate < .05, and discordant rate < 5. All of the above metrics must have been met in order for the sample to be assigned as QC pass. If a sample did not meet the passing criteria, the following failure

481 analysis was performed: a) If the Freemix score was at least 0.05, the sample or the library was
482 considered contaminated, and both the library and the sample were abandoned; b) if the discordant
483 rate was over 5 and/or the inter-chromosomal rate was over 0.05, the quality of DNA was considered
484 poor and the sample was removed from the sequencing pipeline; and c) in the case of a) and b), the
485 collaborator was contacted to determine if selection of a replacement sample from the same cohort
486 was desired or feasible.

487 Separately, whole exome sequencing (WES) data (N = 19,034), genotyping array data (N =
488 17,718) imputed using the Haplotype Reference Consortium panel [49] v1.1, and transformed,
489 normalized quantitative cardiometabolic trait data were obtained from an earlier study [26]. FINRISK
490 array data came in nine genotyping batches, two of which were excluded from the present study due
491 to small sample size. The traits, normalization and transformation procedures, and sample sizes are
492 described in a previous publication [26]. The WES and imputed sample sets contained 4,013 and
493 3,929 of the 4,163 WGS samples included in the present study.

494 All participants in both the METSIM and FINRISK studies provided informed consent, and
495 study protocols were approved by the Ethics Committees at participating institutions (National Public
496 Health Institute of Finland; Hospital District of Helsinki and Uusimaa; Hospital District of Northern
497 Savo). All relevant ethics committees approved this study.

498

499 **WGS callset generation and quality control**

500 Single nucleotide polymorphisms (SNPs) and small insertions and deletions were called from
501 the full set of 4,163 samples using GATK [50] v3.5. GVCFs containing SNVs and Indels from GATK
502 HaplotypeCaller (`-ERC GVCF -GQB 5 -GQB 20 -GQB 60 -variant_index_type LINEAR`
503 `-variant_index_parameter 128000`) were first processed to ensure no GVCF blocks crossed
504 boundaries every 1 Mb (CombineGVCFs; `--breakBandsAtMultiplesOf 1000000`). The

resulting GVCFs were then processed in 10 Mb shards across each chromosome. Each shard was combined (CombineGVCFs), genotyped (GenotypeGVCFs; `-stand_call_conf 30 -stand_emit_conf 0`), hard filtered to remove alternate alleles uncalled in any individual removed (SelectVariants; `--removeUnusedAlternates`), and hard filtered to remove lines solely reporting symbolic deletions in parallel. All shards were jointly recalibrated (VariantRecalibrator) and then individually filtered (ApplyRecalibration) based on the recalibration results. All of the above methods were performed using GATK v3.5. SNP variant recalibration was performed using the following options to VariantRecalibrator and all resources were drawn from the GATK hg38 resource bundle (v0):

```
-mode SNP
-resource:hapmap,known=false,training=true,truth=true,prior=15.0
-resource:omni,known=false,training=true,truth=true,prior=12.0
-resource:1000G,known=false,training=true,truth=false,prior=10.0
-resource:dbsnp,known=true,training=false,truth=false,prior=2.0
-an QD -an DP -an FS -an MQRankSum -an ReadPosRankSum
-tranche 100.0 -tranche 99.9 -tranche 99.0 -tranche 90.0
```

Indel variant recalibration was performed using the following options to VariantRecalibrator (with the same resource bundle as with SNPs):

```
-mode INDEL
-resource:mills,known=true,training=true,truth=true,prior=12.0
-an DP -an FS -an MQRankSum -an ReadPosRankSum
--maxGaussians 4
-tranche 100.0 -tranche 99.9 -tranche 99.0 -tranche 90.0
```

When applying the variant recalibration the following options were used:

For SNPs: `--ts_filter_level 99.0`

For Indels: `--ts_filter_level 99.0`

Following SNP and INDEL variant recalibration, multiallelic variants were decomposed and normalized with vt [51] v0.5. Duplicate variants and variants with symbolic alleles were subsequently removed. The bottom tranche of variants identified by GATK's Variant Quality Score Recalibration tool and variants with missingness greater than 2% were removed as well, although variants with allele balance between 0.3 and 0.7 were rescued. Variants with Hardy-Weinberg equilibrium (on a second-degree unrelated subset of 3,969 individuals, as determined by KING [52]) P value less than 10^{-6} and those with allele balance less than 0.3 or greater than 0.7 were also removed.

Sample-level quality control was also undertaken on this dataset; 13 samples were identified for exclusion because of singleton counts that were at least eight median absolute deviations above the median. Separately, 12 sex-discordant samples were flagged using `plink --check-sex`, and after examining chromosome Y missingness and F coefficient values for these samples, only the one that clearly differed from its reported sex was marked for exclusion. No samples were excluded based on missingness fraction or the first five principal components. In total, 14 samples were excluded from the heritability, GWAS, and Mendelian randomization analyses; the other analyses were performed without exclusion of these samples. As a result, the former analyses were performed with N = 4,149 while the latter had N = 4,163.

Mitochondrial genome copy number estimation

We estimated mitochondrial genome copy number (MT-CN) from both WGS and WES data. In WGS data, we used BEDTools [53] to calculate per-base coverage on the mitochondrial genome from the latest available 4,163 WGS CRAM files. MT-CN was then calculated by normalizing the mean

553 coverage of the mitochondrial genome to the "haploid coverage" of the autosomes as calculated by
554 Picard [48]. The result was then doubled to account for the diploidy of the autosomal genome. This
555 normalization is summarized by the following equation: $MT_CN_{WGS} = 2 \times \frac{mean\ mtDNA\ coverage}{haploid\ autosomal\ coverage}$.

556 The output from the above equation served as the raw measurement of per sample MT-CN. To
557 reduce batch effects, we separated the 4,163 samples into three groups: METSIM, FINRISK collected
558 in 1992 or 1997, and FINRISK collected in 2002 or 2007 (the FINRISK batching decisions were made
559 based on the means shown in S4 Fig). Within each cohort, the raw estimates were regressed on age,
560 age^2 and sex (FINRISK only) and the residuals were inverse-normal transformed. We combined the
561 three batches of normalized MT-CN values and inverse-normal transformed the combined values for
562 downstream analysis.

563 We used a similar procedure to estimate MT-CN from WES data, with mean autosomal
564 coverage estimates taken from XHMM [54]. However, as mitochondrial genomic coverage was
565 nonuniform due to the use of hybrid capture probes, mean mtDNA coverage was not an obvious
566 choice of metric for MT-CN estimation (S5 Fig). To summarize this nonuniform mitochondrial genomic
567 coverage into a single number, we tried taking the mean and the maximum depth of reads that
568 aligned to the mitochondrial chromosome; the resulting values were then processed in the same way
569 as the WGS-estimated values. We evaluated the approaches by measuring the R^2 between
570 WGS-estimated and WES-estimated MT-CN in the 4,013 samples for which both data types were
571 available (S6 Fig). While R^2 was fairly high using both approaches, the maximum coverage method
572 was ultimately selected for use as it yielded a higher R^2 (0.445 vs 0.380). As a result, the WES
573 MT-CN estimate was calculated as follows: $MT_CN_{WES} = 2 \times \frac{maximum\ mtDNA\ coverage}{mean\ haploid\ autosomal\ coverage}$.

574

575 **Mitochondrial haplogroup estimation**

576 We assigned mitochondrial haplogroups using HaploGrep [55] v1.0. Mitochondrial SNP/indel
577 variants were genotyped using GATK GenotypeGVCFs, and a customized filter based on allele
578 balance was applied to the combined callset. HaploGrep was then used to call mitochondrial
579 haplotypes for each individual. We adjusted for major haplogroups in the same linear regressions of
580 metabolic traits onto MT-CN (see Results) and calculated the summary statistics from a permutation
581 test as implemented in lmPerm (https://github.com/mtorchiano/lmPerm).
582

583 **Heritability analysis**

584 To estimate heritability of MT-CN, a genomic relatedness-based restricted maximum-likelihood
585 (GREML) method was used as implemented in GCTA [56]. The original GREML [57] method was
586 used first, followed by GREML-LDMS [58] to account for biases arising from differences in minor
587 allele frequency (MAF) spectrum or linkage disequilibrium (LD) properties between the genotyped
588 variants and the true causal variants [59]. For both analyses, MT-CN values were normalized and
589 residualized for sex, age, and age^2 as described above. Heritability estimation was performed jointly
590 and separately for METSIM and FINRISK samples using WGS and imputed array genotypes. In all
591 cases, a minimum MAF threshold of 1% was applied. Beyond those covariates already adjusted for in
592 the normalization process, sensitivity analyses were performed on imputed array data to determine
593 whether heritability estimates were sensitive to inclusion of covariates. In these experiments, either
594 cohort or FINRISK genotyping array batch were included as fixed-effect covariates in joint analyses of
595 imputed array data; in neither case was the final heritability estimate significantly affected ($h^2 = 0.09$,
596 SE = 0.02 in both cases). In GREML-LDMS, genotypes were split into four SNP-based LD score
597 quartiles and two MAF bins (1% > MAF > 5% and MAF > 5%), and genetic relatedness matrices
598 (GRMs) were estimated separately for each of the eight combinations. The GREML algorithm was
599 then run on all eight GRMs simultaneously using the first ten principal components (PCs) of the

genotype matrix (as calculated by smartPCA v13050; https://github.com/DReichLab/EIG) as fixed covariates [58]. The use of GREML-LDMS over GREML also did not affect estimated heritability values (Table 3), suggesting that the properties of the causal variants for this trait do not lead to significant biases when using the standard GREML approach.

We observed that WGS heritability estimates decrease when analyzing FINRISK and METSIM data together compared to analysis of METSIM alone (Table 2) (note that FINRISK-only heritability estimates are not reliable as they have large standard errors resulting from the small number of FINRISK samples sequenced). One potential explanation for this is that there exists substantial heterogeneity across FINRISK survey years (S4 Fig), and between the FINRISK and METSIM cohorts, with respect to the reliability with which mtDNA was captured (likely due to different DNA preparation protocols).

Genome-wide association analyses

Genome-wide association studies (GWAS) were performed using the same normalized phenotype used in heritability analyses. Single-variant GWAS were conducted using EMMAX as implemented in EPACTS (https://genome.sph.umich.edu/wiki/EPACTS). Kinship matrices required by EMMAX were generated by EPACTS; kinship matrices for WGS GWAS were generated from WGS data, while those for WES and imputed array based GWAS were generated from WES data. A P value threshold of 5×10^{-8} was used for the WGS and imputed array GWAS while 5×10^{-7} was used for significance in the WES GWAS. Single-variant association analyses of WGS and WES data did not include any covariates in the EMMAX model, although all association analyses were performed using MT-CN values that adjusted for age, age^2, sex, and cohort (see “Mitochondrial Genome Copy Number Estimation”).

Gene-based variant aggregation studies (RVAS) were done using a mixed-model version of SKAT-O [37] as implemented in EPACTS. Variants with CADD [60] score greater than 20 and minor allele frequency less than 1% were grouped into genes as annotated by VEP [61] (which annotates a variant with a gene name if the gene falls within 5k of that gene by default). For gene-based RVAS, genome-wide significance thresholds varied slightly due to differing the number of genes with at least two variants meeting the above criteria in each test, but were approximately 2×10^{-6} in all cases.

Mendelian randomization

To assess the evidence for a causal relationship between mitochondrial genome copy number and fasting serum insulin levels, the METSIM cohort alone was used due to its homogeneity of sex, collection procedures, and location. A penalized regression based, multiple variant Mendelian randomization (MR) approach was employed to enforce the necessary assumptions of MR methods. While some MR studies have tested one or more assumptions *post hoc*, to our knowledge, there is no published method that tries to enforce these assumptions during the process of building the genetic instrument in the absence of a large set of known genotype-exposure associations. In our formulation (Fig 4a), X, the natural log of MT-CN (adjusted for nuclear genomic coverage but not for age, age^2, or sex), and a genotype matrix G were used to build a genetic instrument Z, which was then tested against Y, the natural log of fasting serum insulin. The goal of the MR approach was to use a large number of common variants to build a genetic instrument Z that satisfied the three assumptions of MR [62]:

1. Association of Z with X
2. Independence of Z from any variables U confounding the relationship between X and Y
3. Independence of Z and Y given X and U

646 To attempt to build a genetic instrument satisfying assumptions 2 and 3, the deep METSIM
647 phenotype data were leveraged. A matrix W was constructed using the 75 measured traits and first
648 20 PCs of the genotype matrix (including a third-degree polynomial basis for PC 1). From these
649 variables, covariates that could violate one of these two assumptions were chosen by selecting
650 columns of W associated with X or Y (Fig 4b). These columns were selected using two successive
651 LASSO feature selection procedures. First, a set A of covariates associated with Y was chosen by
652 using LASSO to regress Y onto W. In this regression, age and the third-degree polynomial basis for
653 PC 1 were left unpenalized to ensure that A contains these covariates. The shrinkage parameter was
654 chosen by tenfold cross-validation as the largest value that gives a mean squared error (MSE) within
655 one standard error of the minimum observed MSE. Next, the columns of W associated with X
656 conditional on A were chosen using a similar LASSO procedure in the regression of X onto W. In this
657 step, however, the variables in set A were left unpenalized in order to only capture associations that
658 are conditionally independent of A. The selected variables from this regression were designated set
659 B.

660 The instrument was built using a penalized regression (using either an L1 or L2 penalty, as
661 implemented in glmnet [63]) of the form $X \sim G + W_A + W_B$, where W_A and W_B are the columns of W
662 representing sets A and B, respectively, and G is a genotype matrix containing the alternate allele
663 dosage (missing alleles are replaced with the MAF, similarly to PLINK [64]) of all variants with MAF
664 greater than 1% and marginal GWAS P value below 0.01. As X was the target vector for this
665 regression, assumption 1 of MR was trivial. In the penalized regression, W_A and W_B were unpenalized
666 in an effort to orthogonalize the regression coefficients of the genotypes to these covariates in an
667 effort to enforce assumptions 2 and 3. glmnet was run with a convergence threshold of 1×10^{-10} and
668 maximum number of iterations of 200,000. To avoid the overfitting that would result from calculating
669 instrument values on the same samples on which regression coefficients are learned [65], the

670 penalized regression model was fit on independent subsets of the data as follows. Five models were
671 fit, each by holding out a different 20% of samples, such that the instrument value computed for each
672 sample was calculated using the regression coefficient vector learned without that sample. The vector
673 of possible shrinkage parameters λ for all five models was supplied as (10^3,10^2,…,10^{-13},10^{-14}), and the
674 λ value which minimized the joint residual sum of squares of all five models was chosen for
675 instrument calculation.

676 Formally, we randomly partitioned the set of samples S with nonmissing insulin measurements
677 into five nonoverlapping sets S_j for $j = \{1, \ldots, 5\}$. We denote set complements as $S_j^C = S \setminus S_j$, such
678 that each S_j^C contained 80% of the training samples. The instrument vector Z_j for each S_j was
679 computed as follows: $Z_j = G_j \times \beta_G^{(-j)}$, where Z_j is the instrument vector for S_j, G_j is the genotype matrix
680 of S_j, and $\beta_G^{(-j)}$ is the vector of genotype regression coefficients from the model described above,
681 trained on S_j^C. The instrument values within each S_j were inverse rank-normalized using a Blom
682 transformation [66,67] before being concatenated across the values of j to give the final instrument
683 vector Z. Because samples with missing insulin values could not be included in the causality test
684 anyway, these samples were excluded from S but safely included in the training sets of all five
685 models. The instrument values of these samples were never calculated or used in downstream
686 analyses.

687 Often, the inclusion of unpenalized covariate sets A and B in the instrument-building regression
688 was not sufficient to completely orthogonalize Z to these covariates (see below). As a result, the test
689 for association between Z and Y was performed conditional on a set of potentially
690 assumption-violating covariates chosen using the newly constructed instrument Z in another attempt
691 to account for possible violations of MR assumptions in the causality test (Fig 4c). To choose this set
692 of covariates C, a final feature selection step was performed using LASSO regression of Z on W with
693 covariate set A excluded from the penalty. As in the previous feature selection steps, the shrinkage

694 parameter was chosen via tenfold cross-validation as the largest value with MSE within one standard
695 error of the minimum observed MSE. Once this set, D, of covariates associated with Z was chosen,
696 the covariates in W were partitioned into sets I, II, III, and IV based on their membership in A and D
697 (see Fig 4). Formally, this partitioning was done as follows: $I = W \setminus (A \cup D)$, $II = D \cap A^C$,
698 $III = A \cap D^C$, and $IV = A \cap D$, where $A^C = W \setminus A$ and $D^C = W \setminus D$. Then, the test for causality
699 came from the regression coefficient of Z in the multiple regression $Y \sim Z + C$, where C is the union of
700 sets II, III, and IV (colored blue in Fig 4c).

701 To account for missing data in W, missing values were multiply imputed using regression trees
702 as implemented in the R package mice [68] v3.4.0 (`maxit=25`). This imputation was repeated 1000
703 times in parallel, with each set of imputed values being carried through the entire procedure
704 described above. The resulting 1000 computed instrument effect sizes and standard errors were
705 combined using Rubin's method as implemented in the R package Amelia [69] v1.7.5. The combined
706 effect size and standard error were then tested for significance using a t-test with 998 degrees of
707 freedom.

708 The above procedure was performed separately for METSIM samples with WGS data (N =
709 3,034) and METSIM samples with only imputed array data (N = 6,774) using an L1 penalty in the
710 instrument-building regression, and again using an L2 penalty. Both sample sets were limited to those
711 for which relevant quantitative traits were available. An inverse-variance weighted meta-analysis was
712 performed across data sets for L1 and L2-penalized regression separately. The resulting effect size
713 and standard error were tested for significance using a Z test.

714 To ensure that our results were not driven by outlier samples, we removed outliers in two
715 stages. Before the MR analyses, we used principal components analysis (PCA), Mahalanobis
716 distance, and multi-trait extreme outlier identification to remove 5 WGS samples and 15 imputed
717 array samples based on quantitative trait data. We also removed high leverage, high residual outliers

718 from the causality test regression (see below) *post hoc* and recomputed the instrument effect sizes to
719 ensure that there was no significant change in the results. In each of the 1,000 multiple imputation
720 runs, among the samples with standardized residual greater than 1, the top 10 samples by leverage
721 were recorded. Any sample that was recorded in this way in at least one run was then excluded from
722 the re-analysis as a *post hoc* outlier. The results of this additional analysis showed only very small
723 differences in effect estimates, and their interpretation remained the same (S9 Table). Thus, we
724 concluded that our causal inference results were not driven by outlier samples.

725 One caveat of this method is that, as mentioned above, exclusion of sets A and B from the
726 regression penalty did not perfectly orthogonalize the resulting instrument from these variables in
727 practice (S7 Fig). Reasons for this may include relatively low levels of shrinkage in the
728 instrument-building regression or higher order associations between MT-CN and the confounding
729 variables. However, our method still represents an improvement over the current standard, which is
730 not to adjust for these covariates at all. Another caveat is that it is impossible to determine the perfect
731 set of covariates for which adjustment is appropriate. Lack of adjustment for truly confounding
732 variables can result in an instrument which does not satisfy MR assumptions 2 and/or 3, yielding a
733 biased effect estimate. Conversely, unnecessary adjustment for certain variables can also result in
734 biases. For example, adjusting for an intermediate phenotype that truly lies along the path from Z to X
735 to Y can cause a false negative signal, making the causality test overly conservative. Alternatively,
736 adjusting for some variables can result in collider biases [70]. That is, if both Z and Y are causal for a
737 confounder U, then adjusting for U can induce a dependency between Z and Y (S8 Fig) that did not
738 previously exist.

739 We note that a known source of bias in MR studies is the selection of samples based on
740 case-control status for a related disease [71]. While METSIM is a population-based study, samples
741 were selected for WGS based on cardiovascular disease case-control status so as to enrich the

742 sequenced samples for cases. This has the potential to bias a MR experiment if both the exposure
743 and the outcome are associated with the disease, which is certainly possible. However, in our design,
744 all of the METSIM samples not chosen for WGS were tested in the imputed array experiment. The
745 consistency of effect estimates between the WGS and imputed array samples both in the L1 and L2
746 penalty cases (Fig 4d) suggests that there is little to no bias arising from sample selection in this
747 experiment.
748

749 **Calculation and testing of polygenic risk score in the UK Biobank**

750 To search for associations between MT-CN and other phenotypes, the genetic instrument
751 calculated in Finnish imputed array data was computed and treated as a polygenic risk score (PRS)
752 in a relatively homogenous subset of 357,656 UK Biobank samples identified by a previous study
753 [42]. We calculated $\overline{\beta_G}$, the average of the five values of $\beta_G^{(-j)}$ across all 1000 multiple imputation runs
754 using an L2 penalty and imputed array data – the L2 penalty was chosen because it performed better
755 than the L1 on both METSIM data types, and the imputed array data set was chosen due to its larger
756 sample size than the WGS set (Fig 4d). Next, to keep the procedure as consistent as possible with
757 the imputation protocol used for METSIM – which used haploid dosage values to call imputed
758 genotypes [26] – we called imputed genotypes using the expected alternate allele dosage from the
759 UK Biobank by setting thresholds of 0.5 and 1.5. Using the resulting imputed variant calls, we
760 calculated our PRS as $\tilde{Z} = \overline{\beta_G} \times \tilde{G}$, where $\tilde{G}$ is the UK Biobank genotype dosage matrix constructed
761 in the same way as G in METSIM.

762 To test for associations with MT-CN PRS in the UK Biobank, we employed two approaches: a
763 hypothesis-driven analysis targeted to the phenotypes associated with MT-CN in the Finnish data as
764 well as a hypothesis-free screen of all the phenotypes available to us.

765 In the targeted analysis, we used our genetic instrument from the MR experiment as a PRS for
766 MT-CN in our chosen subset of the UK Biobank and tested for associations with several blood cell
767 count and metabolic syndrome traits. Given the association of MT-CN with rs9389268 (see Results),
768 we selected as cell count traits total leukocyte count as well as lymphocyte, neutrophil, monocyte,
769 and platelet counts for testing (because lymphocyte count was not readily available, it was calculated
770 as the product of leukocyte count and lymphocyte percentage). We did not include basophils and
771 eosinophils in this analysis considering that they comprise a small minority of white blood cells and
772 are unlikely to affect MT-CN measured from whole blood. All cell count traits were log-transformed
773 and standardized separately by sex.

774 We took several steps to eliminate outlier samples in the dataset. Through three iterations of
775 PCA on the cell count matrix and subsequent outlier removal, we removed 1,637 outlier samples. We
776 then fit null linear models of the form $cell\ count \sim age + age^2 + sex + age : sex + age^2 : sex + PCs$ (the first
777 20 PCs were included) for each cell count trait and subsequently removed samples with either large
778 residuals or high leverage and moderate residuals in at least one model (following the example of
779 [42]). Through two iterations of null model fitting and outlier removal, we removed 7 additional
780 samples based on null model fit. Tests of association between cell counts and MT-CN PRS were
781 based on the PRS regression coefficient in linear models of the form
782 $cell\ count \sim PRS + age + age^2 + sex + age : sex + age^2 : sex + PCs$.

783 We repeated this process for those cardiometabolic traits found to be suggestively associated
784 with MT-CN in the Finnish dataset ($P < 10^{-6}$) that were also readily available in the UK Biobank (Fig
785 1a, S1 Table); these phenotypes were body mass index (BMI), fat mass, C-reactive protein,
786 high-density lipoprotein, total triglycerides, and weight. We also chose to include T2D status because
787 of the lack of insulin measurement in the UK Biobank. Except for T2D, a binary trait, all traits were
788 log-transformed before further analysis (after removing 817 samples with negative values for T2D,

789 representing missing information). The above outlier removal steps were repeated for the
790 cardiometabolic traits after excluding the outliers already identified from the cell count data, with the
791 only major modification being the use of logistic regression for the T2D models. This process resulted
792 in the removal of 42 and 53 samples from PCA and null model fitting, respectively.

793 SKAT-O tests of association between *TMBIM1* and the cell count traits identified above were
794 also performed. Similarly to the RVAS in Finnish data, variants within 5kb of *TMBIM1* with MAF < 1%
795 and CADD v1.6 score > 20 were selected for inclusion in this analysis. Rather than the mixed-model
796 version of SKAT-O used in the Finnish data, standard SKAT-O was used due to the lower expected
797 level of cryptic relatedness in the UK Biobank population.

798 We also performed a hypothesis-free, phenome-wide screen of UK Biobank traits to which we
799 had access (S10 Table), to search for other associations with MT-CN PRS. The statistical models
800 used in this screen were of the same form as those described above, both with and without
801 adjustment for neutrophil and platelet counts. To curate and transform phenotypes, we used an
802 adapted version of PHESANT [41,42]. A few further modifications were made to the pipeline, the
803 most significant being the direct use of logistic regression for testing categorical unordered variables,
804 the inclusion of cancer phenotypes, and the exclusion of sex-specific (or nearly sex-specific)
805 categorical traits. The PHESANT pipeline we used [42] outputs continuous variables both in their raw
806 form and after applying an inverse rank normal transformation. For the sake of being conservative
807 and robust to outliers, we chose to interpret the results from the normalized continuous variables. To
808 control false discovery rate, we performed a Benjamini-Hochberg procedure with Storey correction as
809 implemented in the R package qvalue [72] v2.18.0 on the categorical and normalized continuous
810 variables together. As a secondary analysis, this same correction was applied to the categorical and
811 raw continuous variables together.

812

Author contributions

LG led trait association, heritability, and MR analyses. LC discovered the original links between NumtS, MT-CN, and metabolic traits, and developed the MT-CN estimation method. RC and LG conceived the MR covariate correction approach, and RC provided critical guidance on statistical methods. HJA, DL, ID, AR, and LG led variant callset creation and QC. JV and ML performed the analysis of insulin sensitivity and resistance. AP, SR, ML, JK, and AH contributed samples and phenotypic data. All authors edited the manuscript and/or provided intellectual contributions. IH and NOS conceived the study and supervised the overall project. LG and IH wrote the paper, with contributions from LC.

Acknowledgements

We thank H. Jung for her help in identifying outlier individuals as well as the WashU data production team, in particular R. Fulton, L. Fulton, C. Fronick, A. Wollam and S.K. Dutcher. This research has been conducted using the UK Biobank Resource under Application Number 56546. The FINRISK samples used for the research were obtained from the FINRISK Study and from THL Biobank. We thank all study participants for their generous participation at THL Biobank and the FINRISK Study.

References

1. Koliaki C, Roden M. Alterations of Mitochondrial Function and Insulin Sensitivity in Human Obesity and Diabetes Mellitus | Annual Review of Nutrition. Annu Rev Nutr. 2016;36: 337–367.
2. Weisberg SP, McCann D, Desai M, Rosenbaum M, Leibel RL, Ferrante AW. Obesity is associated with macrophage accumulation in adipose tissue. J Clin Invest. 2003;112: 1796–1808.
3. Kim J-A, Wei Y, Sowers JR. Role of Mitochondrial Dysfunction in Insulin Resistance. Circ Res. 2008;102: 401–414.

838 4. Burgueño AL, Cabrerizo R, Gonzales Mansilla N, Sookoian S, Pirola CJ. Maternal high-fat intake during
839 pregnancy programs metabolic-syndrome-related phenotypes through liver mitochondrial DNA copy
840 number and transcriptional activity of liver PPARGC1A. J Nutr Biochem. 2013;24: 6–13.

841 5. Sookoian S, Rosselli MS, Gemma C, Burgueño AL, Gianotti TF, Castaño GO, et al. Epigenetic regulation
842 of insulin resistance in nonalcoholic fatty liver disease: Impact of liver methylation of the peroxisome
843 proliferator–activated receptor γ coactivator 1α promoter. Hepatology. 2010;52: 1992–2000.

844 6. Begriche K, Igoudjil A, Pessayre D, Fromenty B. Mitochondrial dysfunction in NASH: Causes,
845 consequences and possible means to prevent it. Mitochondrion. 2006;6: 1–28.

846 7. Zhou X, Li R, Liu X, Wang L, Hui P, Chan L, et al. ROCK1 reduces mitochondrial content and irisin
847 production in muscle suppressing adipocyte browning and impairing insulin sensitivity. Sci Rep. 2016;6:
848 29669.

849 8. Ren J, Pulakat L, Whaley-Connell A, Sowers JR. Mitochondrial biogenesis in the metabolic syndrome and
850 cardiovascular disease. J Mol Med. 2010;88: 993–1001.

851 9. Stephenson EJ, Hawley JA. Mitochondrial function in metabolic health: A genetic and environmental tug of
852 war. Biochimica et Biophysica Acta (BBA) - General Subjects. 2014;1840: 1285–1294.

853 10. Szendroedi J, Phielix E, Roden M. The role of mitochondria in insulin resistance and type 2 diabetes
854 mellitus. Nat Rev Endocrinol. 2012;8: 92–103.

855 11. Ding J, Sidore C, Butler TJ, Wing MK, Qian Y, Meirelles O, et al. Assessing Mitochondrial DNA Variation
856 and Copy Number in Lymphocytes of ~2,000 Sardinians Using Tailored Sequencing Analysis Tools. PLoS
857 Genet. 2015;11: e1005306.

858 12. Chen S, Xie X, Wang Y, Gao Y, Xie X, Yang J, et al. Association between leukocyte mitochondrial DNA
859 content and risk of coronary heart disease: A case-control study. Atherosclerosis. 2014. pp. 220–226.
860 doi:10.1016/j.atherosclerosis.2014.08.051

861 13. Lee HK, Song JH, Shin CS, Park DJ, Park KS, Lee KU, et al. Decreased mitochondrial DNA content in
862 peripheral blood precedes the development of non-insulin-dependent diabetes mellitus. Diabetes Res Clin
863 Pract. 1998;42: 161–167.

864 14. Shoar Z, Goldenthal MJ, De Luca F, Suarez E. Mitochondrial DNA content and function, childhood obesity,
865 and insulin resistance. Endocr Res. 2016;41: 49–56.

866 15. Song J, Oh JY, Sung Y-A, Pak YK, Park KS, Lee HK. Peripheral Blood Mitochondrial DNA Content Is
867 Related to Insulin Sensitivity in Offspring of Type 2 Diabetic Patients. Diabetes Care. 2001;24: 865–869.

868 16. Weng S-W, Lin T-K, Liou C-W, Chen S-D, Wei Y-H, Lee H-C, et al. Peripheral blood mitochondrial DNA
869 content and dysregulation of glucose metabolism. Diabetes Res Clin Pract. 2009;83: 94–99.

870 17. Liu L-P, Cheng K, Ning M-A, Li H-H, Wang H-C, Li F, et al. Association between peripheral blood cells
871 mitochondrial DNA content and severity of coronary heart disease. Atherosclerosis. 2017;261: 105–110.

872 18. Robin ED, Wong R. Mitochondrial DNA molecules and virtual number of mitochondria per cell in
873 mammalian cells. J Cell Physiol. 1988;136: 507–513.

874 19. Maianski NA, Geissler J, Srinivasula SM, Alnemri ES, Roos D, Kuijpers TW. Functional characterization of
875 mitochondria in neutrophils: a role restricted to apoptosis. Cell Death Differ. 2004;11: 143–153.

20. Zharikov S, Shiva S. Platelet mitochondrial function: from regulation of thrombosis to biomarker of disease. Biochem Soc Trans. 2013;41: 118–123.

21. Cai N, Chang S, Li Y, Li Q, Hu J, Liang J, et al. Molecular signatures of major depression. Curr Biol. 2015;25: 1146–1156.

22. Cai N, Li Y, Chang S, Liang J, Lin C, Zhang X, et al. Genetic Control over mtDNA and Its Relationship to Major Depressive Disorder. Curr Biol. 2015;25: 3170–3177.

23. Curran JE, Johnson MP, Dyer TD, Göring HHH, Kent JW, Charlesworth JC, et al. Genetic determinants of mitochondrial content. Hum Mol Genet. 2007;16: 1504–1514.

24. Pajukanta P, Terwilliger JD, Perola M, Hiekkalinna T, Nuotio I, Ellonen P, et al. Genomewide scan for familial combined hyperlipidemia genes in finnish families, suggesting multiple susceptibility loci influencing triglyceride, cholesterol, and apolipoprotein B levels. Am J Hum Genet. 1999;64: 1453–1463.

25. Pajukanta P, Allayee H, Krass KL, Kuraishy A, Soro A, Lilja HE, et al. Combined analysis of genome scans of dutch and finnish families reveals a susceptibility locus for high-density lipoprotein cholesterol on chromosome 16q. Am J Hum Genet. 2003;72: 903–917.

26. Locke AE, Steinberg KM, Chiang CWK, Service SK, Havulinna AS, Stell L, et al. Exome sequencing of Finnish isolates enhances rare-variant association power. Nature. 2019;572: 323–328.

27. Kaess B, Fischer M, Baessler A, Stark K, Huber F, Kremer W, et al. The lipoprotein subfraction profile: heritability and identification of quantitative trait loci. J Lipid Res. 2008;49: 715–723.

28. Weiss LA, Pan L, Abney M, Ober C. The sex-specific genetic architecture of quantitative traits in humans. Nat Genet. 2006;38: 218–222.

29. Wang X, Angelis N, Thein SL. MYB - A regulatory factor in hematopoiesis. Gene. 2018;665: 6–17.

30. Pandit RA, Svasti S, Sripichai O, Munkongdee T, Triwitayakorn K, Winichagoon P, et al. Association of SNP in exon 1 of HBS1L with hemoglobin F level in beta0-thalassemia/hemoglobin E. Int J Hematol. 2008;88: 357–361.

31. Thein SL, Menzel S, Peng X, Best S, Jiang J, Close J, et al. Intergenic variants of HBS1L-MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults. Proceedings of the National Academy of Sciences. 2007;104: 11346–11351.

32. Ganesh SK, Zakai NA, van Rooij FJA, Soranzo N, Smith AV, Nalls MA, et al. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. Nat Genet. 2009;41: 1191–1198.

33. Menzel S, Jiang J, Silver N, Gallagher J, Cunningham J, Surdulescu G, et al. The HBS1L-MYB intergenic region on chromosome 6q23.3 influences erythrocyte, platelet, and monocyte counts in humans. Blood. 2007;110: 3624–3626.

34. Lin BD, Carnero-Montoro E, Bell JT, Boomsma DI, de Geus EJ, Jansen R, et al. 2SNP heritability and effects of genetic variants for neutrophil-to-lymphocyte and platelet-to-lymphocyte ratio. J Hum Genet. 2017;62: 979–988.

35. Stadhouders R, Aktuna S, Thongjuea S, Aghajanirefah A, Pourfarzad F, van Ijcken W, et al. HBS1L-MYB intergenic variants modulate fetal hemoglobin via long-range MYB enhancers. J Clin Invest. 2014;124: 1699–1710.

36. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics. 2003;19: 149–150.

37. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet. 2012;91: 224–237.

38. Zhou J, Zhu T, Hu C, Li H, Chen G, Xu G, et al. Comparative genomics and function analysis on BI1 family. Comput Biol Chem. 2008;32: 159–162.

39. Zhao G-N, Zhang P, Gong J, Zhang X-J, Wang P-X, Yin M, et al. Tmbim1 is a multivesicular body regulator that protects against non-alcoholic fatty liver disease in mice and monkeys by targeting the lysosomal degradation of Tlr4. Nat Med. 2017;23: 742–752.

40. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12: e1001779.

41. Millard LAC, Davies NM, Gaunt TR, Davey Smith G, Tilling K. Software Application Profile: PHESANT: a tool for performing automated phenome scans in UK Biobank. Int J Epidemiol. 2018;47: 29–35.

42. Neale B. Neale Lab UK Biobank Analysis. Available: http://www.nealelab.is/uk-biobank/

43. Welty FK, Alfaddagh A, Elajami TK. Targeting inflammation in metabolic syndrome. Transl Res. 2016;167: 257–280.

44. Creely SJ, McTernan PG, Kusminski CM, Fisher ff M, Da Silva NF, Khanolkar M, et al. Lipopolysaccharide activates an innate immune system response in human adipose tissue in obesity and type 2 diabetes. Am J Physiol Endocrinol Metab. 2007;292: E740–7.

45. Hotamisligil GS. Inflammation, metaflammation and immunometabolic disorders. Nature. 2017;542: 177–185.

46. Laakso M, Kuusisto J, Stančáková A, Kuulasmaa T, Pajukanta P, Lusis AJ, et al. The Metabolic Syndrome in Men study: a resource for studies of metabolic and cardiovascular diseases. J Lipid Res. 2017;58: 481–493.

47. Borodulin K, Tolonen H, Jousilahti P, Jula A, Juolevi A, Koskinen S, et al. Cohort Profile: The National FINRISK Study. Int J Epidemiol. 2018;47: 696–696i.

48. Picard Toolkit. Broad Institute, GitHub repository; 2019. Available: http://broadinstitute.github.io/picard/

49. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48: 1279–1283.

50. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43: 491–498.

51. Tan A, Abecasis GR, Kang HM. Unified representation of genetic variants. Bioinformatics. 2015;31: 2202–2204.

52. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010;26: 2867–2873.

53. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics.

952 2010;26: 841–842.

953 54. Fromer M, Moran JL, Chambert K, Banks E, Bergen SE, Ruderfer DM, et al. Discovery and statistical
954 genotyping of copy-number variation from whole-exome sequencing depth. Am J Hum Genet. 2012;91:
955 597–607.

956 55. Kloss-Brandstätter A, Pacher D, Schönherr S, Weissensteiner H, Binna R, Specht G, et al. HaploGrep: a
957 fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Hum Mutat.
958 2011;32: 25–32.

959 56. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J
960 Hum Genet. 2011;88: 76–82.

961 57. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from
962 genome-wide association studies. Am J Hum Genet. 2011;88: 294–305.

963 58. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, et al. Genetic variance estimation with
964 imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet.
965 2015;47: 1114–1120.

966 59. Evans LM, Tahmasbi R, Vrieze SI, Abecasis GR, Das S, Gazal S, et al. Comparison of methods that use
967 whole genome data to estimate the heritability and genetic architecture of complex traits. Nat Genet.
968 2018;50: 737–745.

969 60. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating
970 the relative pathogenicity of human genetic variants. Nat Genet. 2014;46: 310–315.

971 61. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect
972 Predictor. Genome Biol. 2016;17: 122.

973 62. Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, et al. Using multiple genetic
974 variants as instrumental variables for modifiable risk factors. Stat Methods Med Res. 2012;21: 223–242.

975 63. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate
976 Descent. J Stat Softw. 2010;33: 1–22.

977 64. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the
978 challenge of larger and richer datasets. Gigascience. 2015;4: 7.

979 65. Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. Int
980 J Epidemiol. 2013;42: 1134–1144.

981 66. Beasley TM, Erickson S, Allison DB. Rank-based inverse normal transformations are increasingly used,
982 but are they merited? Behav Genet. 2009;39: 580–595.

983 67. Blom G. Statistical estimates and transformed beta-variables. Almqvist & Wiksell. 1958. Available:
984 http://www.diva-portal.org/smash/record.jsf?pid=diva2:516729

985 68. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. Journal
986 of Statistical Software, Articles. 2011;45: 1–67.

987 69. Honaker J, King G, Blackwell M. Amelia II: A Program for Missing Data. Journal of Statistical Software,
988 Articles. 2011;45: 1–47.

989 70. Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, et al. Illustrating bias due to

990 conditioning on a collider. Int J Epidemiol. 2010;39: 417–420.

991 71. VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in mendelian
992 randomization. Epidemiology. 2014;25: 427–435.

993 72. Storey JD, Bass AJ, Dabney A, Robinson D. qvalue: Q-value estimation for false discovery rate control.
994 2019. Available: http://github.com/jdstorey/qvalue

995 73. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at
996 UCSC. Genome Res. 2002;12: 996–1006.

997 74. International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. A second
998 generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449: 851–861.

999

1000 Supporting information

1001 **S1 Fig. Insulin association signal at chromosome 1 NumtS.** Manhattan plot from an analysis of
1002 copy number variation at nuclear mitochondrial insertion sites (NumtS) using CNVnator read-depth
1003 measurements in 1 kb genomic windows, using an initial set of 2,049 samples from an earlier data
1004 freeze analyzed for an unrelated CNV association study (manuscript in prep.). Shown at the bottom
1005 are two tracks from the UCSC Genome Browser [73]. The black track represents an assembly gap
1006 upstream of the association signal, and the blue track shows the location of the NumtS region where
1007 the association peak is located.

1008

1009 **S2 Fig. Joint (METSIM and FINRISK) single-marker association tests with MT-CN in WGS and**
1010 **WES data.** (a) Manhattan plot and quantile-quantile (QQ) plot for an exome-wide association test of
1011 normalized, WES-measured MT-CN using WES genotype data (N = 19,034). The red line represents
1012 an exome-wide significance level of 5×10^{-7}; no tested markers achieved this level of significance. The
1013 QQ plot is separated by minor allele frequency bin, as indicated by the colors and shapes of the
1014 points. (b) Manhattan plot and quantile-quantile (QQ) plot for a genome-wide association test of
1015 normalized, WGS-measured MT-CN using WGS genotype data (N = 4,149). The red line represents

1016 a genome-wide significance level of 5×10^{-8}; no tested markers achieved this level of significance. The
1017 QQ plot is separated by minor allele frequency bin, as indicated by the colors and shapes of the
1018 points.
1019
1020 **S3 Fig. Single-marker association test results at chromosome 10 QTL identified by Curran *et*
1021 *al.*** Chromosome 10 Manhattan plot for association with WES-estimated MT-CN in Finnish imputed
1022 array data from METSIM and FINRISK. The blue highlighted region represents the approximate
1023 location of the linkage peak reported for MT-CN in Curran *et al*., 2007 [23]. This region is larger than
1024 the reported region in the Curran study (~30 Mb vs. ~24 Mb) because the exact coordinates of the
1025 1-LOD support interval were not reported explicitly, so the genetic coordinates were approximated
1026 from a figure and converted to physical coordinates using a genetic map from HapMap Phase 2 [74].
1027 Some error may have been introduced due to differing populations between the HapMap data and the
1028 Mexican-American samples used by Curran.
1029
1030 **S4 Fig. Raw WGS-based MT-CN estimate distributions for METSIM and all four FINRISK**
1031 **surveys.** Facet labels “1992”, “1997”, “2002”, and “2007” refer to individual FINRISK survey years.
1032
1033 **S5 Fig. Nonuniform WES coverage across the mitochondrial genome for four representative**
1034 **samples.**
1035
1036 **S6 Fig. Comparison of two methods of summarizing nonuniform mitochondrial coverage in**
1037 **WES into a single measurement.** Both panels show the correlation with WGS mean coverage for
1038 the 4,013 samples for which both WGS and WES data were available. (a) shows the results of
1039 summarizing WES mitochondrial coverage using the maximum coverage value across the

1040 mitochondrial chromosome while (b) shows the results of using the mean mitochondrial coverage
1041 instead. R^2 values are shown on each panel.
1042
1043 **S7 Fig. Associations between Z and traits in set A in imputed data**. P values computed from
1044 linear regression using the output from a randomly chosen run of multiple imputation of phenotype
1045 data. (a) and (c) show the marginal P values of the traits in set A from regressions of Z onto each trait
1046 separately, while (b) and (d) show the P values from multiple regression of Z onto all traits in set A.
1047 The instrument in (a) and (b) was computed using an L1 penalty, while that in (c) and (d) was
1048 computed using an L2 penalty.
1049
1050 **S8 Fig. Example of collider bias.** If Z and Y are independently causal for a variable U, then adjusting
1051 for U can induce an association between Z and Y that did not previously exist.
1052
1053 **S1 Table. Cardiometabolic trait associations with MT-CN in WGS data.** Phenome-wide EMMAX
1054 association results of normalized MT-CN against 137 cardiometabolic traits (N = 4,163). These are
1055 the data underlying Fig 1a. OGTT: Oral glucose tolerance test, NMR: nuclear magnetic resonance.
1056
1057 **S2 Table. MT-CN associations with insulin and fat mass in WES data.** Results of EMMAX test of
1058 association between normalized MT-CN and both fat mass and fasting serum insulin using WES data.
1059 Association tests were performed in all samples and also separately among samples with and without
1060 WGS data.
1061
1062 **S3 Table. Direct associations between platelet/neutrophil counts and metabolic syndrome**
1063 **phenotypes in the UK Biobank.** Results of direct testing of platelet and neutrophil counts against

1064 metabolic syndrome phenotypes chosen based on marginally significant associations for MT-CN
1065 without adjustment for cell counts. Testing was done in the UK Biobank (N = 357,656) using linear
1066 regression of traits onto cell counts conditional on the same covariates as the MT-CN analyses (see
1067 Methods). The bottom two rows show the results after removal of high-residual, high-leverage outliers
1068 as determined by Cook's distance.

1069

1070 **S4 Table. Associations in the UK Biobank between cell counts and loci mapped to MT-CN in**
1071 **Finnish data.** Association tests in the UK Biobank (N = 357,656) of cell counts against
1072 common-variant and gene-based rare-variant associations found in Finnish GWAS.

1073

1074 **S5 Table. MT-CN phenome-wide screen using normalized continuous variables with no cell**
1075 **count adjustment.** Results of phenome-wide screen for association with MT-CN PRS in the UK
1076 Biobank (N = 357,656), with FDR correction applied to all categorical variables and normalized
1077 continuous variables. These tests were performed without adjustment for platelet and neutrophil
1078 counts. A dark line separates the results at q = 0.05.

1079

1080 **S6 Table. MT-CN phenome-wide screen using raw continuous variables with no cell count**
1081 **adjustment.** Results of phenome-wide screen for association with MT-CN PRS in the UK Biobank (N
1082 = 357,656), with FDR correction applied to all categorical variables and raw continuous variables.
1083 These tests were performed without adjustment for platelet and neutrophil counts. A dark line
1084 separates the results at q = 0.05.

1085

1086 **S7 Table. MT-CN phenome-wide screen using normalized continuous variables with**
1087 **adjustment for platelet and neutrophil counts.** Results of phenome-wide screen for association

with MT-CN PRS in the UK Biobank (N = 357,656), with FDR correction applied to all categorical variables and normalized continuous variables. These tests were performed with adjustment for platelet and neutrophil counts. A dark line separates the results at $q = 0.05$.

S8 Table. MT-CN phenome-wide screen using raw continuous variables with adjustment for platelet and neutrophil counts. Results of phenome-wide screen for association with MT-CN PRS in the UK Biobank (N = 357,656), with FDR correction applied to all categorical variables and raw continuous variables. These tests were performed with adjustment for platelet and neutrophil counts. A dark line separates the results at $q = 0.05$.

S9 Table. Mendelian randomization results after removal of high leverage *post hoc* outliers. Results of Mendelian randomization test for causality of MT-CN on fasting serum insulin, after removal of high leverage *post hoc* outliers (see “Inference of causality in the association between MT-CN and insulin”). These results are nearly identical to those computed before outlier removal (Fig 4d), indicating that the detected signals were not driven by outlier samples.

S10 Table. Phenotypes in the UK Biobank included in the phenome-wide screen after filtering and transformation by PHESANT. Variable type is reported by PHESANT, where "Categorical unordered" corresponds to a binary variable representing a single possible answer. Samples are generally only excluded as a result of sex imbalances, although there were four phenotypes for which a model was unable to be fit when adjusting for cell counts.