Supplementary note 1: Testing the robustness of h^2_{WGS} estimates

The following section extends the analysis performed to ensure the robustness of our heritability estimates for height and BMI. We used a set of N=25,465 unrelated samples of European ancestry without excess of heterozygosity. Variants were removed using several quality filters (genotypes missingness rate, Hardy-Weinberg equilibrium test *P* value, MAF, quality classifier) and grouped into 8 bins according to their MAF (0.0001 – 0.001, 0.001 – 0.01, 0.01 – 0.1, 0.1 – 0.5) and individual LD score value (median-based). The phenotypes were pre-adjusted for age and standardized to a mean of 0 and a variance of 1 in each sex and cohort group. See Online Methods for more details on the dataset and QC performed.

Influence of LD definition on variance estimates

We observed minor differences in heritability estimates when replicating previous studies using only genotypes from SNP mimicking arrays followed by imputation (h^2_{G+IMP}). Estimates of h^2_{G+IMP} for height and BMI were in the range of 0.50-0.56 (SE 0.06-0.07) and 0.16-0.21 (SE 0.07) respectively (Figure 1). Previous estimates[1] were 0.56 (SE 0.02) for height and 0.27 (SE 0.02) for BMI, with a set of 44,126 unrelated samples and ~17.6M imputed variants on 1000 Genomes panel with no filtering on imputation quality score, a segment-based LD definition to define LD bins and no LD pruning on the variants used to compute PCs. Moreover, our height variance estimate using segment-based LD definition using Axiom array SNPs prior to imputation was at 0.51 (SE 0.04) compared to 0.56 (SE 0.07) using individual SNP LD score (Supplementary Figure 9). While a difference in heritability estimates between LD definition was expected from previous study[2], the difference between current and prior LD-based estimates for height could be explained by the PCs from LD pruned variants capturing better potential population stratification, a different imputation panel, a more stringent imputation quality threshold or differences in base population heritabilities. Differences in BMI could also be explained by the aforementioned reasons and the different LD definitions.

Comparing LD / MAF structure using UK10K data

We also tested if there was any bias due to a specific LD or MAF structure in the TOPMed dataset by using a different sequenced dataset for SNP stratification. We used the UK10K dataset[3] and analysed the TOPMed data using the MAF and LD stratification from either TOPMed or UK10K data. We converted the UK10K WGS data[3] to GRCh38 coordinates using LiftMap, a wrapper Python script for LiftOver[4]. There are 3,781 individuals in the UK10K dataset. As with TOPMed, we performed a quality control of the genotypes using PLINK with the following filtering thresholds: individuals with missingness rate < 0.05 and variants with missingness rate < 0.05, Hardy-Weinberg equilibrium test *P* value $<1 \times 10^{-6}$, or minor allele frequency < 0.0001, and retained 3,781 individuals and 42.68M variants. From these variants, we selected 20.6M variants in common with those in the TOPMed dataset. Using the UK10K genotypes of the 20.6M variants, we defined 4 MAF bins (0.0001 < MAF < 0. 001, 0.001 < MAF < 0.01, 0.01 < MAF < 0.1 and 0.1 < MAF<0.5) and further split the variants in each MAF bin into 2 LD bins according to their LD scores (calculated using a window size of 10Mb in either direction). We also defined 8 MAF- and LD-stratified variant bins using the TOPMed genotypes of the 20.6M variants. We estimated and partitioned additive genetic variance in the TOPMed based on these 20.6M variants, using either the MAF and LD annotation from the UK10K or TOPMed, fitting 160 PCs computed from LD-

pruned WGS variants. The estimates were highly consistent between the two analyses (Supplementary Figure 34). For height, the estimates were 0.54 (SE 0.08) when using the TOPMed annotation and 0.55 (SE 0.07) using SNP stratification from the UK10K, and the corresponding estimates for BMI were 0.27 (SE 0.07 – 0.08) in both cases. The similarity between the estimates from the two references for MAF and LD stratification suggests that our inference from the TOPMed annotation is not biased by using MAF and LD stratification from another dataset.

Effect of rank-inverse normal transformation

We analysed both height and BMI (both adjusted for age and standardized within each sex and cohort group) with a rank-based inverse normal transformation (RINT). We used the set of N=25,465 unrelated Europeans samples and 33.7M high quality variants. We used GREML-LDMS fitting 8 MAF/LD bins and 48 PCs capturing population stratification. The estimate for the RINT-transformed phenotypes were similar when compared to those from the untransformed trait for height (0.63 (SE 0.09)) but lower for BMI (0.20 (SE 0.10)) (Supplementary Figure 6). Since BMI naturally has a skewed distribution (Supplementary Figure 4), results may be sensitive to the scale of analysis. Past estimates of heritability from pedigree and GWAS designs are not consistent in the scale, with some using the actual scale of measurement and other performing a pre-analysis logarithm or RINT transformation of the data. We performed analyses on the actual and RINT scale and found that the estimates from the RINT-transformed data appeared more sensitive to the model, although the RINT estimates seems to be more consistent in analyses considering a reduced number of rare variants (such as down-sampling or analysing exome data). These differences could be due to sampling variation and would need to be investigated further with a larger sample size.

Influence of SNP quality and GRM estimators

As a further check, we compared GREML-LDMS estimates by selecting the complete set of genotyped variants in place of the high-quality variants identified with a classifier trained using a support vector machine algorithm (SVM) and additional hard filters (Online Methods). That increased the total number of variants used to compute GRMs from 33.7M to 44.2M variants. The estimate of h^2_{WGS} for height was 0.62 when fitting the 20 HM3 PCs and 0.63 when fitting the 160 PCs. When investigating the differences between these estimates and the ones from high quality variants, we noticed some large differences in diagonal and off-diagonal values of the low MAF GRMs, potentially due to sequencing errors or batch effects in these ultra-rare variants (Supplementary Figure 26, Supplementary Figure 27). We therefore concluded that the choice of stringent variant QC was justified.

To quantify the effect of differential SNP weighting within each GRM, we investigated the influence of calculating the GRM estimator A_{jk} using the average ratio over loci ("average of ratio")[5] instead of the ratio of total SNP covariance and total SNP heterozygosity over loci ("ratio of averages")[6] (Supplementary Figure 28, Supplementary Figure 29) on the GREML-LDMS estimates (Online Methods, Supplementary Figure 30). Note that the default GCTA method (average of ratios) assumes an inverse relationship between MAF and variant effect size whereas the ratio of averages assumes no relationship between MAF and variants effect sizes. For height and BMI, the average of ratio method showed slightly higher estimates (0.63, SE 0.10 for height and 0.29 (SE 0.11) for BMI, respectively) than the ratio of averages (0.61 (SE 0.09) and 0.25 (SE 0.10)), accounting for population stratification. These results show that our current estimates seem robust to the GRM estimator (see Discussion and the Supplementary Note 3 for more discussion), consistent with prior work[1].

Investigating the influence of direct GRM QC on heritability estimates

We investigated how outlier values in the rare variants GRMs influence the estimates of h^2_{WGS}. From the 28,755 unrelated individuals of European ancestry (before filtering individuals based on their heterozygosity values in each bin, Supplementary Figure 3, Supplementary Figure 31, Supplementary Figure 32), we filtered individuals based either on their off-diagonal values (removing pairs > 0.1), diagonal values (< 0.7 and > 1.3), or both (Online Methods, Supplementary Table 6, Supplementary Figure 33). For height, removing one of each pair of individuals with large off-diagonal values in any of the rare variants GRMs did not change the estimates regardless of the number of PCs fitted in the model (~0.58 – 0.71 (SE 0.08)). However, excluding individuals with extremely large diagonal values yielded a substantial increase in h^2_{WGS} estimates (0.71 – 0.81 (SE 0.09)). Removing individuals based on both diagonals and off-diagonals values still showed this large increase in heritability estimates (0.71 – 0.82 (SE 0.09 – 0.10)). For BMI, filtering samples did not increase the estimates but revealed a larger contribution from rare variants in the low-LD bins; and a large overall increase was observed compared to estimates from QC based on sample heterozygosity. The reasons why excluding individuals with extreme diagonal values affects heritability estimates for height and not for BMI are unclear and warrant further investigation. Heritability differences between height and BMI may contribute to inflate the effect of extreme GRM values, by making them more visible for height. Note that observed differences remain within standard errors. We also investigated the influence of a more stringent common-variant relatedness threshold of 0.025 (instead of 0.05 used previously), but the estimates did not change substantially (Supplementary Table 7). Therefore, extreme GRM diagonal values observed in rare variant GRMs seem to have a larger influence on estimates than extreme off-diagonal values.

Impact of LD stratification

Finally, we investigated the influence of LD on GREML-LDMS estimate by increasing the number of LD bins. Prior work suggest that 2 LD bins are sufficient to estimate and partition genetic variance[1]. However, when using WGS variants there is substantial variation in LD score within a MAF group (Supplementary Figure 35). We therefore investigated whether further partitioning according to LD could better capture genetic variance. We divided each MAF bin into 3 and 4 LD bins instead of 2, thereby increasing the total number of GRMs fitted in the model from 8 to 12 (when dividing each MAF bin into 3 LD tertile groups) or 16 (when dividing each MAF bin into 4 LD quartile groups). Compared to the h^2_{WGS} estimates based on 2 LD bins, the estimates based on 3 and 4 LD bins were consistently higher, ranging between 0.67 – 0.68 (SE 0.10) for height and 0.28 – 0.32 (SE 0.10) for BMI (Figure 2), with the 48 PCs fitted as fixed covariates. This increase is consistent with larger LD heterogeneity within MAF bins, which was not fully accounted for by 2 LD bins. While using 3 LD bins substantially increased estimates for both traits, we did not observe such increase when increasing the number of LD bins to 4 (Supplementary Figure 36). Comparing the Akaike Information Criterion (AIC) between different models, the 4 LD grouping strategy consistently performed the best for height and BMI, followed by the 3 LD grouping indicating that the 2 LD grouping strategy is not the most appropriate when considering rare variants (Supplementary Figure 37). These estimates are also closer to the estimates of h^2_{WGS} without outlying diagonal and off-diagonal elements (see above). However, additional partitioning comes at the cost of decreased precision of the estimate of heritability.

Supplementary note 2: Investigating potential bias of h^2_{WGS} estimates due to localized environmental effects

Simulation parameters

We investigated potential bias in the heritability estimates due to localized and sharply distributed environmental effects correlating with rare variants stratification. Such environmental effects have been previously shown to cause a bias in PGS estimated effect, which may not be fully corrected for by fitting PCs[7,8]. We used a simple simulation setting with the UKB WES data where we identified groups of samples based either on their birthplace administrative regions or visually, based on the correlation between PCs computed from rare variants on the odd and even chromosomes (Supplementary Figure 24). We selected 971 individuals born near Rotherham (administrative region with the largest sample size excluding the main cities in the UK) or 841 individuals born in the North of the UK (hereafter denoted the "North" sample; normalized birth coordinates (Online Methods) between 0.5 and 0.7 on the North axis and between 0.25 and 0.55 on the East axis) (Supplementary Figure 17). The mean Fst between the "North" samples and the rest of the dataset was of 1.45×10^{-4} for common variants (MAF > 0.01) and $3.7 \times 10^{-5}$for rare variants (MAF < 0.01 and MAC > 3). For the Rotherham samples Fst was of 2.1×10^{-5} for common variants and 1.8×10^{-5} for rare variants, indicating a lower level of genetic heterogeneity with the rest of the dataset for the Rotherham samples as compared to the "North" samples.

We generated standardized phenotypes from randomly sampling causal variants (either 10,000 with MAF > 0.01; or 10,000 with MAC 3 < MAF < 0.01; or 5,000 variants from both the common and rare variants groups). Each simulation setting was replicated 10 times. To simulate a localized environmental effect, we added an effect $\beta \in \{0.5, 1, 2\}$ to the standardized phenotypes of Rotherham or "North" samples (accounting for 0.7%, 2.6%, 10.5% and 0.6%, 2.3%, 9.2% of the total phenotypic variance for Rotherham and North, respectively, Supplementary Figure 18). We performed a GREML-LDMS analysis with a rare variants GRM (MAC3 < MAF < 0.01) and a common variants GRM (0.01 < MAF < 0.5). We fitted as fixed effects either 40, 200 or 1000 PCs from LD pruned SNPs (i.e. 20, 100 or 500 per bin) depending on the experiment. We also investigated whether using a RINT transformation could affect the estimates. Finally, we conducted an experiment assigning a constant term to N=971 samples selected randomly from the dataset, as a negative control.

Effect of localized environmental effect

On the phenotypes without any added effect and the negative control (effect added at random) we observed a heritability estimate of 0.70, consistent with the simulated trait value (Supplementary Figure 19). We observed an inflation coming from the rare-variant GRM when a localized effect was added on the North or Rotherham (Supplementary Figure 20, Supplementary Figure 21). The magnitude of this effect was greater when all the causal variants were common. In some instances, this effect could be partially corrected by applying a RINT transformation to correct for the phenotype skewness or by PCs when the populations were genetically differentiated from the rest of the samples ("North" samples). For the Rotherham samples, fitting a large number of PCs could not fully account for the environmental effect and the estimate of genetic variance was biased. Adding a localized effect of 2 standard deviations resulted in a large inflation of the estimates for rare variants (~0.10), confirming previous observations[7,8]. However, these simulations settings do not reflect the observed distribution of regional height effects in the UKB. When investigating

the mean phenotypic difference per administrative region (Supplementary Figure 18), the largest effect we could observe in the whole UKB was of the order of 0.5 standard deviations when considering effect size, and 0.2% when considering phenotypic variance explained. The simulated value of 2 standard deviations (equivalent to about 14 cm for height) accounting for ~10% of the phenotypic variance, while useful to study how much PCs correct for localized effects, appears unrealistic when compared to real phenotypes. We then replicated the simulations with an effect of 0.5 standard deviations accounting for ~2% of the variance (still larger than what was observed in the UKB) and could not find any substantial inflation in the estimation of variance explained by the rare variants bin. Moreover, any real regional phenotypic differences might also partially reflect genetic effects which could be accounted for by fitting a sufficient number of PCs.

Inflation of heritability estimates for whole-genome estimates

Our simulations confirm a potential inflation of heritability estimates when a localized environmental effect affects the phenotype of a small number of individuals and these individual share rare variants. When quantifying the magnitude of this effect, there was no substantial inflation when simulating an effect corresponding to the highest effect in the UKB or TOPMed (the mean differences between cohort within TOPMed have also been corrected for in the rest of our analysis). The Fst in between the cohorts sampled for these simulations are very low, and a higher Fst between populations could help PCs to account for localized effects. These simulations have investigated only one localized effect and might not fully reflect all the environmental effects found within a population, such as a combination of effects of small magnitude. However, from the distribution of mean phenotypes across administrative regions, it does not seem likely that any real regional environmental effects are large. From these simulations, our estimates seem broadly robust to localized environmental effects. Large datasets of WGS or WES with environmental information could help investigate this question in the future.

Supplementary note 3: the effect of rare variants population stratification on genomic relationship matrices

Here, we describe various analyses aiming at better understanding the impact of rare variants stratification on estimates of SNP-based heritability from WGS data. The following notes expand in greater details some of the analyses briefly described in the Online Methods section.

Extreme values in rare-variants GRMs

After performing an initial QC to select unrelated adults (of age > 18) samples of European ancestry (N=28,755) based on the first 4 PCs for common and rare variants and 1000 Genomes reference populations[9], we performed QC on variants using several filters (excluding variants with genotypes missingness rate > 0.05, Hardy-Weinberg equilibrium test *P* value $<1 \times 10^{-6}$, MAF < 0.0001 and high quality variants from a SVM classifier). On the 36.9M variants left after sample and genotype QC, defined 8 variants bins based on their MAF (0.0001 – 0.001, 0.001 – 0.01, 0.01 – 0.1, 0.1 – 0.5) and individual LD score value (median-based). We then computed GRM using these 8 variants groupings. We observed extreme diagonal and off-diagonals values in GRMs (Online methods) calculated from variants with 0.0001 < MAF < 0.01. Prior to performing any further sample/variant filtering, we first investigated the influence of the GRM estimator on these extreme values. We compared two GRM estimators implemented in GCTA: the default GRM estimator[5], which calculate average of per-SNP relatedness estimates; and the VanRaden estimator[10], which estimates relatedness as the ratio of SNP covariance and SNP heterozygosity. We found small differences between the diagonal elements of these two matrices (Supplementary Figure 28) and larger differences in off-diagonal elements of the GRM (Supplementary Figure 29). The average of ratios estimator shows off-diagonal values consistently higher for extreme values. As a whole, we observe, for variants with 0.0001 < MAF < 0.01, high relatedness for some pairs (with relatedness values consistent with a first-degree relationship) and large GRM diagonal values that are inconsistent with the expected sampling variance of rare variants GRMs with a larger number of effective markers[1]. The GRM estimator alone does not explain why a set of unrelated (estimated common variants relatedness < 0.05) samples using HM3 common SNPs exhibits very high relatedness from rare variants GRMs.

Influence of IBD segments on relatedness

The relatedness threshold of 0.05 using HM3 common SNPs, while being widely used in previous studies[1], might not be sufficient to remove residual relatedness. For example, cousins-once-removed have an expected relatedness value of 0.0625 (SD 0.017)[11]. A relatedness threshold of 0.05 could then include such pairs who would share segments IBD as shown through simulations in previous studies[12]. We first selected samples with a more stringent threshold on HM3 SNPs of 0.025. That removed an additional 2282 samples. When investigating the rare variants GRMs we could still observe extreme values in the diagonals (~4) and off-diagonals (~1) elements, despite the more stringent threshold on common variants relatedness.

In the absence of sample relatedness for common variants and genome-wide population stratification, large values in rare variants GRMs could be driven by shared IBD segments. To identify the shared IBD segments, we selected 987,393 LD-pruned variants from the full WGS data (window size of 50kb and LD r^2 threshold of 0.1) with MAF > 0.01 on N=28,755 samples and identified pairs shared IBD segments > 1MB on each of the 22 chromosomes

with KING 2.2.6 software[13]. Across all chromosomes, 56.6M segments were shared IBD with 3.26M unique pairs sharing one or more segments IBD (Supplementary Figure 38). For each pair of individuals (sharing at least 1 segment IBD), we compared the cumulative length of IBD segments with its corresponding GRM off-diagonal values. We found that some pairs sharing the longest IBD segments accumulatively did not contribute disproportionately to the GRM off-diagonal elements for high-LD variants with $0.0001 < MAF < 0.01$. We also noticed that some pairs had an extreme relatedness value (large off-diagonals), yet did not share a large proportion of their genome IBD. These pairs were enriched with individuals exhibiting a high level of MAF/LD bin-specific heterozygosity (Supplementary Figure 39). These results suggest that high off-diagonal values are caused by high mean heterozygosity due to variants that are disproportionally shared but not in long IBD segments.

We confirmed these findings by investigating, for a pair of individuals showing an extreme off-diagonal value in the GRM for high-LD variants with $0.0001 < MAF < 0.001$, the distribution of the heterozygous rare and oppositive homozygous common variants for the chromosome enriched for rare variants shared by the pair (Supplementary Figure 40). For this specific pair, we identified a region with many clumped rare variants and no oppositive homozygotes common variants, suggesting a region shared IBD.

We then defined 3 groups of samples (Supplementary Figure 41) selecting either the first 100 pairs with the highest off-diagonals (for high-LD variants with $0.0001 < MAF < 0.001$), the first 100 pairs sharing the longest proportion of their genome IBD, and 100 pairs selected at random (within 0.01 SD from the median of IBD length shared and off-diagonal values) as control. For each group, we used a global ancestry reference[14,15] to infer the genome-wide proportion ancestry of the group of samples (Online methods). There was no difference in mean ancestry across sample groups for each population with the exception that samples with high GRM off-diagonals had a significantly higher proportion of African ancestry, indicating that these pairs might be sharing few shorts IBD segments from a different ancestry, thereby strongly impacting their relatedness. We further investigated the source of heterozygosity by comparing the number of heterozygous variants from either the IBD segments or the entire genome, for the very rare high-LD variants (Supplementary Figure 41). The group showing high relatedness had a very high number of heterozygotes variants (as expected by their high mean heterozygosity rate) with most of these variants on their shared IBD segments. The group sharing a higher proportion of genome IBD had a substantially lower number of heterozygotes variants. Finally, the control group showed a small number of heterozygous variants overall and more spread across the entire genome.

In summary, we show that short IBD segments from a different ancestry have a strong impact on the estimated relatedness values for some pairs in the rare variants GRMs. Including a step in the QC pipeline to identify and remove such pairs allows the correction for extreme relatedness which could bias the variance components estimates. While directly investigating the origin of these segments specifically (through segment-based PCA, for example) is computationally demanding, we used the sample heterozygosity to identify the pairs sharing such segments and remove them. We used multiple rounds of heterozygosity QC to remove the extreme samples (Online methods). Shared IBD segments are expected in an outbred population[16-18], but we tried to control for the large effects it can have when considering very rare variants. We performed a GREML-LDMS analysis with the proportion of African ancestry fitted as an additional covariate and did not observe a significant change in any of the heritability estimates (Supplementary Table 7), because the proportion of African ancestry did not fully capture the outlier samples with high diagonal values.

Linear regression of phenotype on PCs

We performed a number of additional analyses to test the robustness of estimates of h^2_{WGS} to population stratification. The standard correction method using PCs may not be effective in controlling for local population stratification in spatially structured populations[7,8]. To estimate how much phenotypic variance is explained by PCs, we regressed the phenotype on a large number of PCs computed from common and rare variants (Online methods). This analysis showed that increasing the number of PCs from 20 (from common variants) to 160 (common and rare variants) modestly increased the regression R^2 from 0.012 to 0.018 for height and from 0.001 to 0.004 for BMI (Supplementary Figure 14). These results imply that population structure, as quantified by linear regression on PCs from common and rare variants, in total may contribute only about 1.5% of phenotypic variance for height and nearly none for BMI. However, fitting them as fixed effects in a linear mixed model might not affect heritability estimates commensurately (Supplementary Figure 12).

Using birthplace coordinates to capture potential population stratification in the UKB

We lacked any direct information on spatial substructure in the TOPMed dataset, and therefore turned to the UKB where such information is available. We selected a sample of 35,867 unrelated individuals of European ancestry with both whole-exome sequence (WES) data and birthplace coordinates available (Online methods). We investigated how well the potential population stratification could be captured by the birth coordinates in the UKB. The estimates from a GREML-LDMS analysis fitting 14 MAF/LD bins (i.e., 7 MAF x 2 LD) from the WES data and an additional bin of HM3 common SNPs from imputed data were 0.62 (SE 0.04) for height and 0.33 (SE 0.04) for BMI. When including the birthplace coordinates as fixed covariates in the GREML-LDMS analysis, the estimates were essentially the same, 0.61 (SE 0.04) for height and 0.33 (SE 0.04) for BMI (Supplementary Figure 15), showing either no evidence of a strong effect of local stratification on the GREML estimates for rare WES variants in the UKB or that fitting birth coordinates does not fully capture population stratification. Estimates between TOPMed and UKB WES datasets were also similar when selecting variants present in both datasets after equalising the sample sizes (Online methods, Supplementary Figure 16), although we observed UKB WES-based estimates were consistently higher, possibly due to a difference in trait heritability across TOPMed and UKB samples.

Residual population stratification

Finally, we further our investigation to the influence of ancestry proportion on the diagonal elements of the GRMs. Our analysis is limited to individuals already selected to be from European ancestry by a PCA filtering on both common and rare variants then further filtered to stay within 3 SD from main reference populations using a global ancestry inference[14,15] (Online methods). When investigating the relationship between the diagonal elements and the proportion of each ancestry for each MAF/LD grouping, we could not notice a meaningful relationship with the exception of the impact of African ancestry proportion positively correlated ($R^2 = 0.69$) with diagonal values for the high-LD variants with $0.0001 < \text{MAF} < 0.001$ (Supplementary Figure 43). The larger genetic diversity for African populations has been documented by previous studies[9] and is reflected in the diagonal values for rare variants GRMs.

The samples in our study have, after QC steps, a maximum of 1.5% of their genome that is estimated to be derived from African ancestry. Lowering this threshold would remove a large

number of samples. Moreover, the sample identified as having African ancestry IBD segments still show an average proportion of African ancestry genome-wide of around 0.75%, far below our current threshold. Thus, while it is important to consider removing sample showing an excess of any ancestry other than the population under study (here, people of European ancestry), this criterion alone cannot be used to control for subtle population stratification and needs to be used in conjunction with more specific metrics (such as the sample MAF/LD bin heterozygosity rate).

Quality control for WGS data

Much is yet to be learned when using WGS data with very rare variants for genetic analysis. While the set of 28,755 unrelated European ancestry samples did not present any specific issue when considering only common variants, including very rare variants in our analysis necessitated additional QC steps. Ensuring a homogeneous ancestry with PCA and global ancestry inference helps removing any cryptic population stratification that could influence the diagonals elements of the rare variants GRMs. Moreover, it is important to identify pairs sharing short IBD segments coming from a different ancestry and having a large influence on GRM elements for rare variants. Removing samples based on their MAF/LD bin-specific heterozygosity helps to correct for potentially biased estimates of variance components from rare variant GRMs. While this does not fully address the issue of samples showing large GRM element values, it is an efficient way to correct for the bias identified while trying to maximise the sample size. A WGS dataset from a more homogeneous population might not exhibit such patterns among rare variants.

References for Supplementary Notes

1 Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet* **47**, 1114-1120, doi:10.1038/ng.3390 (2015).

2 Evans, L. M. *et al.* Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nat Genet* **50**, 737-745, doi:10.1038/s41588-018-0108-x (2018).

3 UK10K Consortium. The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82-90, doi:10.1038/nature14962 (2015).

4 Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* **34**, D590-598, doi:10.1093/nar/gkj144 (2006).

5 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82, doi:10.1016/j.ajhg.2010.11.011 (2011).

6 Goudet, J., Kay, T. & Weir, B. S. How to estimate kinship. *Mol Ecol* **27**, 4121-4135, doi:10.1111/mec.14833 (2018).

7 Mathieson, I. & McVean, G. Differential confounding of rare and common variants in spatially structured populations. *Nat Genet* **44**, 243-246, doi:10.1038/ng.1074 (2012).

8 Zaidi, A. A. & Mathieson, I. Demographic history mediates the effect of stratification on polygenic scores. *Elife* **9**, doi:10.7554/eLife.61548 (2020).

9 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).

10 VanRaden, P. M. Efficient methods to compute genomic predictions. *J Dairy Sci* **91**, 4414-4423, doi:10.3168/jds.2007-0980 (2008).

11 Hill, W. G. & Weir, B. S. Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genet Res (Camb)* **93**, 47-64, doi:10.1017/S0016672310000480 (2011).

12 Hill, W. G. & White, I. M. Identification of pedigree relationship from genome sharing. *G3 (Bethesda)* **3**, 1553-1571, doi:10.1534/g3.113.007500 (2013).

13 Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-2873, doi:10.1093/bioinformatics/btq559 (2010).

14 Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290-299, doi:10.1038/s41586-021-03205-y (2021).

15 Maples, B. K., Gravel, S., Kenny, E. E. & Bustamante, C. D. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet* **93**, 278-288, doi:10.1016/j.ajhg.2013.06.020 (2013).

16 Browning, S. R. & Browning, B. L. High-resolution detection of identity by descent in unrelated individuals. *Am J Hum Genet* **86**, 526-539, doi:10.1016/j.ajhg.2010.02.021 (2010).

17 Palamara, P. F., Lencz, T., Darvasi, A. & Pe'er, I. Length distributions of identity by descent reveal fine-scale demographic history. *Am J Hum Genet* **91**, 809-822, doi:10.1016/j.ajhg.2012.08.030 (2012).

18 Nait Saada, J. *et al.* Identity-by-descent detection across 487,409 British samples reveals fine scale population structure and ultra-rare variant associations. *Nat Commun* **11**, 6130, doi:10.1038/s41467-020-19588-x (2020).

Supplementary Information: Biobanks and study-specific acknowledgements

UK10K (EGA accessions: EGAS00001000108 and EGAS00001000090): was funded by the Wellcome Trust award WT091310. Twins UK (TUK): TUK was funded by the Wellcome Trust and ENGAGE project grant agreement HEALTH-F4-2007–201413. The study also receives support from the Department of Health via the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St. Thomas' NHS Foundation Trust in partnership with King's College London. Dr Spector is an NIHR senior Investigator and ERC Senior Researcher. Funding for the project was also provided by the British Heart Foundation grant PG/12/38/29615 (Dr Jamshidi). A full list of the investigators who contributed to the UK10K sequencing is available from https://www.uk10k.org/

UK Biobank:
This research has been conducted using the UK Biobank Resource under project 12505.

TOPMed:
WGS for "NHLBI TOPMed: Genetics of Cardiometabolic Health in the Amish" (phs000956.v3.p1.c999) was performed at the Broad Institute of MIT and Harvard (HHSN268201500014C). WGS for TOPMed "NHLBI TOPMed: Trans-Omics for Precision Medicine Whole Genome Sequencing Project: ARIC" (phs001211.v1.p1.c999) was performed at the Broad Institute of MIT and at the Baylor Human Genome Sequencing Center (3R01HL092577-06S1, HHSN268201500015C, 3U54HG003273-12S). WGS for TOPMed "NHLBI TOPMed: Mount Sinai BioMe Biobank (BioMe)" (phs001644.v1.p1.c999) was performed at the McDonnell Genome Institute and at the Baylor Human Genome Sequencing Center (HHSN268201600037I, HHSN268201600033I).WGS for TOPMed "NHLBI TOPMed: Coronary Artery Risk Development in Young Adults (CARDIA)" (phs001612.v1.p1.c999) was performed at the Baylor Human Genome Sequencing Center and at the Keck Molecular Genomics Core Facility (HHSN268201600038I, HHSN268201600033I). WGS for "NHLBI TOPMed: The Cleveland Clinic Atrial Fibrillation Study of the CV/Arrhythmia Biobank" (phs001189.v1.p1.c999) was performed at the Broad Institute of MIT and Harvard (3R01HL092577- 06S1). WGS for "NHLBI TOPMed: The Cleveland Family Study (WGS)" (phs000954.v2.p1.c999) was performed at the University of Washington Northwest Genomics Center (3R01HL098433-05S1). WGS for "NHLBI TOPMed: Cardiovascular Health Study" (phs001368.v1.p1.c999) was performed at the Baylor Human Genome Sequencing Center (HHSN268201500015C, 75N92021D00006). WGS for "NHLBI TOPMed: Genetic Epidemiology of COPD (COPDGene) in the TOPMed Program" (phs000951.v2.p2.c999) was performed at the Broad Institute of MIT and Harvard and the University of Washington Northwest Genomics Center (HHSN268201500014C). WGS for "NHLBI TOPMed: Whole Genome Sequencing and Related Phenotypes in the Framingham Heart Study" (phs000974.v3.p2.c999) was performed at the Broad Institute of MIT and Harvard (3R01HL092577-06S1). WGS for "NHLBI TOPMed: GeneSTAR (Genetic Study of Atherosclerosis Risk)" (phs001218.v1.p1.c999) was performed at the Broad Institute of MIT and Harvard, at Macrogen Corp (HHSN268201500014C) and at Illumina (HL112064).WGS for "NHLBI TOPMed: Genetics of Lipid Lowering Drugs and Diet Network (GOLDN)" (phs001359.v1.p1.c999) was

performed at the University of Washington Northwest Genomics Center (3R01HL104135-04S1). WGS for "NHLBI TOPMed: Heart and Vascular Health Study (HVH)" (phs000993.v2.p2.c999) was performed at the Broad Institute of MIT and Harvard and the Baylor Human Genome Sequencing Center (3R01HL092577-06S1, 3U54HG003273-12S2). WGS for "NHLBI TOPMed: Whole Genome Sequencing of Venous Thromboembolism (WGS of VTE)" (phs001402.v1.p1.c999) was performed at the Baylor Human Genome Sequencing Center (HHSN268201500015C, 3U54HG003273-12S2). WGS for "NHLBI TOPMed: MESA and MESA Family AA-CAC" (phs001416.v1.p1.c999) was performed at the Broad Institute of MIT and Harvard (3U54HG003067-13S1, HHSN268201500014C). WGS for "NHLBI TOPMed: MGH Atrial Fibrillation Study" (phs001062.v3.p2.c999) was performed at the Broad Institute of MIT and Harvard (3R01HL092577-06S1). WGS for "NHLBI TOPMed: Partners Healthcare Biorepository (Partners)" (phs001024.v1.p1.c999) was performed at the Broad Institute of MIT and Harvard (3R01HL092577-06S1). WGS for "NHLBI TOPMed: The Vanderbilt AF Ablation Registry" (phs000997.v3.p2.c999) was performed at the Broad Institute of MIT and Harvard (3R01HL092577-06S1). WGS for "NHLBI TOPMed: The Vanderbilt Atrial Fibrillation Registry" (phs001032.v3.p2.c999) was performed at the Broad Institute of MIT and Harvard (3R01HL092577-06S1). WGS for "NHLBI TOPMed: The Women's Genome Health Study" (phs001040.v3.p1.c999) was performed at the Broad Institute of MIT and Harvard (3R01HL092577-06S1). WGS for "NHLBI TOPMed: Women's Health Initiative (WHI)" (phs001237.v1.p1.c999) was performed at the Broad Institute of MIT and Harvard (HHSN268201500014C). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1). Phenotype harmonization, data management, sample-identity QC, and general study coordination, were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed.

The full TOPMed Consortium authorship list can be found at:
https://topmed.nhlbi.nih.gov/topmed-banner-authorship

COHORT SPECIFIC ACKNOWLEDGEMENTS:
Amish Research Program: This research has been supported in part by NIH grants U01 HL072515, U01 HL072515, R01 AG18728, and P30 DK072488.

Atherosclerosis Risk in Communities Study: The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, under Contract nos. (HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700005I, HHSN268201700004I). The authors thank the staff and participants of the ARIC study for their important contributions.

Coronary Artery Risk Development in Young Adults: Molecular data for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). Whole Genome Sequencing for the NHLBI TOPMed: CARDIA Study (phs001612) was performed at the Baylor College of Medicine Human genome Sequencing Center (contract HHSN268201600033I). Core support including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1;

contract HHSN268201800002I). Core support including phenotype harmonization, data management, sample-identity QC, and general program coordination were provided by the TOPMed Data Coordinating Center (R01HL-120393; U01HL-120393; contract HHSN268201800001I). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed.
The Coronary Artery Risk Development in Young Adults Study (CARDIA) is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the University of Alabama at Birmingham (HHSN268201800005I & HHSN268201800007I), Northwestern University (HHSN268201800003I), University of Minnesota (HHSN268201800006I), and Kaiser Foundation Research Institute (HHSN268201800004I).

Cardiovascular Health Study: This research was supported by contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, and grants U01HL080295, U01HL130114, and R01HL059367 from the National Heart, Lung, and Blood Institute (NHLBI), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Cleveland Family Study (CFS): CFS was supported by National Institutes of Health grants National Institutes of Health grants R01-HL046380-15, HL113338, and KL2-RR024990-05, and R35HL135818 [5-R01-HL046380, and 5-KL2-RR024990-05.

Genetic Epidemiology of COPD Study (COPDGene): The COPDGene project described was supported by Award Number U01 HL089897 and Award Number U01 HL089856 from the National Heart, Lung, and Blood Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health. The COPDGene project is also supported by the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Novartis, Pfizer, Siemens and Sunovion. A full listing of COPDGene investigators can be found at: http://www.copdgene.org/directory

Framingham Heart Study: The Framingham Heart Study (FHS) acknowledges the support of contracts NO1-HC-25195 and HHSN268201500001I from the National Heart, Lung and Blood Institute and grant supplement R01 HL092577-06S1 for this research. We also acknowledge the dedication of the FHS study participants without whom this research would not be possible. Dr. Vasan is supported in part by the Evans Medical Foundation and the Jay and Louis Coffman Endowment from the Department of Medicine, Boston University School of Medicine.

Genetics of Lipid Lowering Drugs and Diet Network (GOLDN): GOLDN biospecimens, baseline phenotype data, and intervention phenotype data were collected with funding from National Heart, Lung and Blood Institute (NHLBI) grant U01 HL072524. Whole-genome sequencing in GOLDN was funded by NHLBI grant R01 HL104135 and supplement R01 HL104135-04S1.

Massachusetts General Atrial Fibrillation Study: This work was supported by grants from the National Institutes of Health to Dr. Ellinor (1RO1HL092577, R01HL128914, K24HL105780). Dr. Ellinor is also supported by the American Heart Association (18SFRN34110082) and by the Fondation Leducq (14CVD01).

Mayo Clinic Venous Thromboembolism Study (Mayo_VTE): Funded, in part, by grants from the National Institutes of Health, National Heart, Lung and Blood Institute (HL66216 and HL83141). the National Human Genome Research Institute (HG04735, HG06379), and research support provided by Mayo Foundation.

Multi-Ethnic Study of Atherosclerosis (MESA): The MESA project is supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420. Also supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

The Partners Healthcare Biorepository (Partners): Dr. Lubitz is supported by NIH grant 1R01HL139731 and American Heart Association 18SFRN34250007.

The Johns Hopkins Genetic Study of Atherosclerosis Risk (GeneSTAR): was supported by grants from the National Institutes of Health through the National Heart, Lung, and Blood Institute (HL49762, HL071025, U01HL72518, HL087698, HL092165, HL099747, K23HL105897, HL112064) and the National Institute of Nursing Research (NR0224103), by a grant from the National Center for Research Resources (M01-RR000052) to the Johns Hopkins General Clinical Research Center, and by a grant from the National Center for Research Resources and the National Center for Advancing Translational Sciences (UL1 RR 025005) to the Johns Hopkins Institute for Clinical and Translational Research.

The Mount Sinai BioMe Biobank (BioME): is supported by The Andrea and Charles Bronfman. We thank the participants in the BioMe Biobank for their invaluable contribution to biomedical research.
Ruth Loos is supported by the NIH (X01HL134588; R01DK110113; R01HL142302; R01DK107786; R01DK124097).

The Women's Health Initiative (WHI): The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C. Investigators: *Program Office*: (National Heart, Lung, and Blood Institute, Bethesda, Maryland) Jacques Rossouw, Shari Ludlam, Joan McGowan, Leslie Ford, and Nancy Geller. *Clinical Coordinating Center*: (Fred Hutchinson Cancer Research Center, Seattle, WA) Garnet Anderson, Ross Prentice, Andrea LaCroix, and Charles Kooperberg. *Investigators and Academic Centers*: (Brigham and Women's Hospital, Harvard Medical School, Boston, MA) JoAnn E. Manson; (MedStar Health Research Institute/Howard University, Washington, DC)

Barbara V. Howard; (Stanford Prevention Research Center, Stanford, CA) Marcia L. Stefanick; (The Ohio State University, Columbus, OH) Rebecca Jackson; (University of Arizona, Tucson/Phoenix, AZ) Cynthia A. Thomson; (University at Buffalo, Buffalo, NY) Jean Wactawski-Wende; (University of Florida, Gainesville/Jacksonville, FL) Marian Limacher; (University of Iowa, Iowa City/Davenport, IA) Jennifer Robinson; (University of Pittsburgh, Pittsburgh, PA) Lewis Kuller; (Wake Forest University School of Medicine, Winston-Salem, NC) Sally Shumaker; (University of Nevada, Reno, NV) Robert Brunner. *Women's Health Initiative Memory Study*: (Wake Forest University School of Medicine, Winston-Salem, NC) Mark Espeland

Supplementary material

Supplementary Table 1: List of studies included in the analysis with corresponding sample sizes comprised of participants with both genotypic and phenotypic information available before quality control, along with their TOPMed accession number.

Study name and code	**Sample size**	**TOPMed accession**
Genetics of Cardiometabolic Health in the Amish (Amish)	1111	phs000956.v3.p1.c999
Atherosclerosis Risk in Communities (ARIC)	8128	phs001211.v1.p1.c999
Mount Sinai BioMe Biobank (BioMe)	11193	phs001644.v1.p1.c999
Coronary Artery Risk Development in Young Adults (CARDIA)	3087	phs001612.v1.p1.c999
Cleveland Clinic Atrial Fibrillation (CCAF) Study	363	phs001189.v1.p1.c999
Cleveland Family Study (CFS)	1259	phs000954.v2.p1.c999
Cardiovascular Health Study (CHS)	3537	phs001368.v1.p1.c999
Genetic Epidemiology of COPD (COPDGene)	10283	phs000951.v2.p2.c999
The Framingham Heart Study (FHS)	4166	phs000974.v3.p2.c999
Genetic Studies of Atherosclerosis Risk (GeneSTAR)	1763	phs001218.v1.p1.c999
Genetics of Lipid Lowering Drugs and Diet Network (GOLDN)	945	phs001359.v1.p1.c999
Heart and Vascular Health Study (HVH)	693	phs000993.v2.p2.c999
Whole Genome Sequencing of Venous Thromboembolism (Mayo_VTE)	1345	phs001402.v1.p1.c999
Multi-Ethnic Study of Atherosclerosis (MESA)	5351	phs001416.v1.p1.c999
Massachusetts General Hospital Atrial Fibrillation (MGH_AF)	989	phs001062.v3.p2.c999
Partners Healthcare Biorepository (Partners)	128	phs001024.v1.p1.c999
The Vanderbilt AF Ablation Registry (VAFAR)	173	phs000997.v3.p2.c999
The Vanderbilt Atrial Fibrillation Registry (VU_AF)	1128	phs001032.v3.p2.c999
Women's Genome Health Study (WGHS)	117	phs001040.v3.p1.c999
Women's Health Initiative (WHI)	11031	phs001237.v1.p1.c999
Total	66790	

Supplementary Table 2: Additional cohort specific information on age, height, BMI, total sample size and proportion of male/females in the study after QC.

Study code	Age		Height		BMI		Sample size	Sex proportion	
	Mean	SD	Mean	SD	Mean	SD		M	F
ARIC	54.0	5.7	169.0	9.6	26.7	4.7	4219	0.45	0.55
BioMe	56.0	14.2	172.7	10.2	27.0	6.1	965	0.52	0.48
CARDIA	25.9	3.1	171.6	9.2	23.6	4.0	1093	0.45	0.55
CCAF	54.5	9.1	179.9	9.7	30.4	6.1	254	0.79	0.21
CFS	46.9	13.1	169.8	9.7	31.5	7.4	139	0.47	0.53
CHS	72.7	5.4	165.6	9.5	26.3	4.3	1668	0.42	0.58
COPDGene	62.2	8.8	170.0	9.5	28.7	6.2	4887	0.52	0.48
FHS	39.7	10.9	169.2	9.3	25.4	4.4	939	0.46	0.54
GeneSTAR	47.2	12.0	170.5	10.2	28.5	6.7	217	0.42	0.58
HVH	62.5	12.0	175.1	10.2	31.8	7.7	515	0.66	0.34
Mayo_VTE	56.3	16.4	171.7	10.7	31.0	7.6	1030	0.49	0.51
MESA	61.4	9.8	170.1	9.6	27.9	5.1	1128	0.48	0.52
MGH_AF	54.7	10.1	179.3	9.9	28.4	5.4	631	0.80	0.20
Partners	51.1	7.4	175.7	10.4	31.3	7.0	68	0.66	0.34
VAFAR	58.4	8.3	177.5	10.8	32.8	6.9	118	0.67	0.33
VU_AF	53.7	10.7	178.0	10.1	31.2	6.8	750	0.74	0.26
WGHS	49.5	3.4	167.8	6.1	28.4	6.3	86	0.00	1.00
WHI	67.3	6.5	161.9	6.0	28.3	5.8	6758	0.00	1.00

Supplementary Table 3: Number of SNPs on each of the Illumina HumanCore24, GSA 24 and Affymetrix Axiom arrays before merging with TOPMed dataset (probes), after merging and after preparing the files for imputation.

Array name	Number of probes on the array	Number of SNPs after merging with TOPMed	Final number of SNPs after preparing files for imputation
Affymetrix Axiom	784,849	742,749	735,920
Illumina HumanCore24	263,947	257,901	257,392
Illumina GSA	648,327	523,370	520,235

Supplementary Table 4: Imputation statistics after imputing SNPs on Illumina InfiniumCore24, GSA 24 and Affymetrix Axiom arrays using the HRC reference panel on Sanger imputation servers, with SNP number with imputation info score above 0.3 after QC.

Array	**Imputation info scores**				**Number of SNPs with $r^2 > 0.3$ after QC**
	Min.	**Median**	**Mean**	**Max.**	
Axiom	0.3	0.829	0.782	1	20,393,950
Infinium	0.3	0. 750	0. 727	1	18,993,608
GSA	0.3	0.791	0.755	1	19,822,722

Supplementary Table 5: Number of variants and LD properties (based on individual variant LD score) of each of the four groupings of the TOPMed dataset according to the allele frequency of the variant. The number of SNPs decrease as the MAF increases as rare variants makes up for most of the variants in the dataset.

MAF bin	Number of SNPs	LD score properties		
		Mean	Median	SD
0.0001 – 0.001	19,583,648	27.1	13.6	46.3
0.001 – 0.01	5,283,043	40.7	22.5	76.7
0.01 – 0.1	3,935,389	122.6	71.5	260.8
0.1 – 0.5	4,902,760	193.6	136.3	241.7

Supplementary Table 6: Values of mean GRM diagonals +/- 3 standard deviations before and after sample heterozygosity QC for each MAF/LD bin.

MAF bin	LD bin	Sample size			
		28,755 (no HET QC)		25,465 (HET QC)	
		Mean – 3*sd	Mean + 3*sd	Mean – 3*sd	Mean + 3*sd
0.0001 – 0.001	Low	0.85	1.17	0.87	1.14
	High	0.21	1.81	0.54	1.45
0.001 – 0.01	Low	0.95	1.05	0.96	1.05
	High	0.85	1.16	0.87	1.13
0.01 – 0.1	Low	0.97	1.03	0.97	1.03
	High	0.92	1.09	0.92	1.08
0.1 – 0.5	Low	0.98	1.02	0.98	1.02
	High	0.96	1.04	0.96	1.03

Supplementary Table 7: Summary of the estimates of heritability for height and BMI from all main analyses performed.

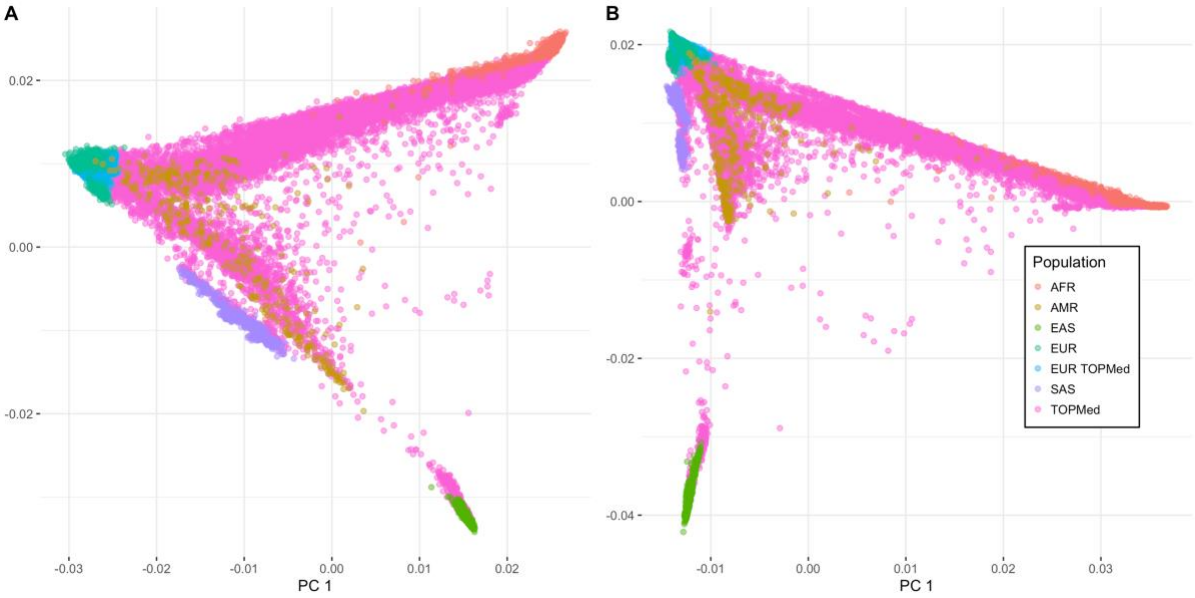
Dataset	Experiment info	SNPs	Sample size	GRM Algorithm	N MAF bins	N LD bins	Random effects (GRMs)	Fixed effects (PCs)	Estimates	
									Height	BMI
TOPMed	GREML-SC (HM3 SNPs)	SVM HM3	25465	Alg1	1	1		20	0.48 (0.02)	0.24 (0.02)
	GREML-MS (WGS SNPs)	High quality SVM WGS SNPs			4	1	4	20 HM3 SNPs	0.48 (0.05)	0.24 (0.05)
					4	1	4	160 WGS	0.45 (0.05)	0.23 (0.05)
	GREML-LDMS (WGS SNPs) (median based)				4	2	8	20 HM3	0.70 (0.09)	0.29 (0.09)
					4	2	8	48 WGS	0.61 (0.09)	0.25 (0.10)
					4	2	8	160 WGS	0.60 (0.09)	0.23 (0.10)
	GREML-LDMS – WGS SNPs – 3 LD bins (tertile based)				4	3	12	20 HM3	0.78 (0.09)	0.31 (0.10)
					4	3	12	48 WGS	0.68 (0.09)	0.32 (0.10)
					4	3	12	160 WGS	0.68 (0.10)	0.30 (0.10)
	GREML-LDMS – WGS SNPs – 4 LD bins (quantile based)				4	4	16	48 WGS	0.68 (0.10)	0.30 (0.10)
					4	4	16	160 WGS	0.67 (0.10)	0.29 (0.10)
					4	4	16	320 WGS (16*20)	0.68 (0.10)	0.28 (0.10)
	Enrichment analysis (splitting low MAF and low LD bins into protein-altering and non-protein-altering)				4	2(+1)	11	20 HM3	0.70 (0.09)	0.29 (0.09)
					4	2(+1)	11	48 WGS	0.61 (0.09)	0.24 (0.10)
	Enrichment and removing extreme diagonal samples		22100		4	2(+1)	11	20 HM3	0.79 (0.10)	0.26 (0.10)
					4	2(+1)	11	48 WGS	0.73 (0.10)	0.21 (0.10)
	GREML-LDMS – Different GRM estimator: Average of ratios		25465	Alg 0	4	2	8	20 HM3	0.74 (0.10)	0.29 (0.11)
					4	2	8	160 WGS	0.63 (0.10)	NA
	GREML-LDMS – No High quality filter on variants	All variants WGS		Alg 1	4	2	8	20 HM3	0.62 (0.06)	NA
					4	2	8	160 WGS	0.62 (0.06)	NA

	GREML-LDMS – TOPMed LD/MAF reference	Intersection UK10K / TOPMed			4	2	8	20 HM3	0.60 (0.08)	0.32 (0.08)
	GREML-LDMS – UK10K LD/MAF reference				4	2	8	20 HM3	0.60 (0.06)	0.30 (0.07)
	GREML-LDMS – Imputation from Axiom	Imputed SNPs Rsq > 0.3			4	2	8	20 HM3	0.56 (0.07)	0.21 (0.07)
	GREML-LDMS – Imputation from GSA				4	2	8	20 HM3	0.55 (0.07)	0.16 (0.07)
	GREML-LDMS – Imputation from Infinium				4	2	8	20 HM3	0.50 (0.06)	0.18 (0.07)
	GREML-LDMS – Imputation from Axiom -LD score segment for LD reference				4	2	8	20 HM3	0.51 (0.04)	NA
	GREML-LDMS – Comparison Axiom imputed – WGS – Axiom genotypes	Intersection Axiom imputed / TOPMed WGS			4	2	8	20 HM3	0.55 (0.07)	0.18 (0.07)
					4	2	8	160 WGS	0.50 (0.07)	0.13 (0.07)
	GREML-LDMS – Comparison Axiom imputed – WGS – TOPMed genotypes				4	2	8	20 HM3	0.62 (0.07)	0.25 (0.07)
					4	2	8	160 WGS	0.56 (0.07)	0.22 (0.07)
	GREML – MS – UKB Exome	Intersection UKB WES TOPMed WGS		Alg1	2	1	2	20 HM3	0.35 (0.02)	0.10 (0.02)
	GREML – MS – TOPMed				2	1	2	20 HM3	0.30 (0.02)	0.07 (0.02)
	GREML – LDMS – UKB Exome				2	2	4	20 HM3	0.37 (0.02)	0.13 (0.02)
	GREML – LDMS – TOPMed				2	2	4	20 HM3	0.31 (0.02)	0.09 (0.02)
	GREML -LDMS – No heterozygosity QC step	SVM WGS	28754		4	2	8	20 HM3	0.70 (0.08)	0.36 (0.08)
					4	2	8	160 WGS	0.58 (0.08)	0.31 (0.08)
	GREML -LDMS – No heterozygosity QC step – 3 LD bins analysis				4	3	12	20 HM3	0.76 (0.08)	0.40 (0.09)
					4	3	12	160 WGS	0.64 (0.08)	0.35 (0.09)
	GREML -LDMS – No heterozygosity QC step but removal of samples with diagonal values > 1.3 and < 0.7		25333		4	2	8	20 HM3	0.81 (0.09)	0.37 (0.09)
					4	2	8	160 WGS	0.71 (0.09)	0.31 (0.10)
	GREML -LDMS – No heterozygosity QC step but removed pairs with off-diagonal values across all WGS GRMs > 0.1		26698		4	2	8	20 HM3	0.71 (0.08)	0.35 (0.09)
					4	2	8	160 WGS	0.59 (0.09)	0.29 (0.09)
			24233		4	2	8	20 HM3	0.82 (0.09)	0.38 (0.10)

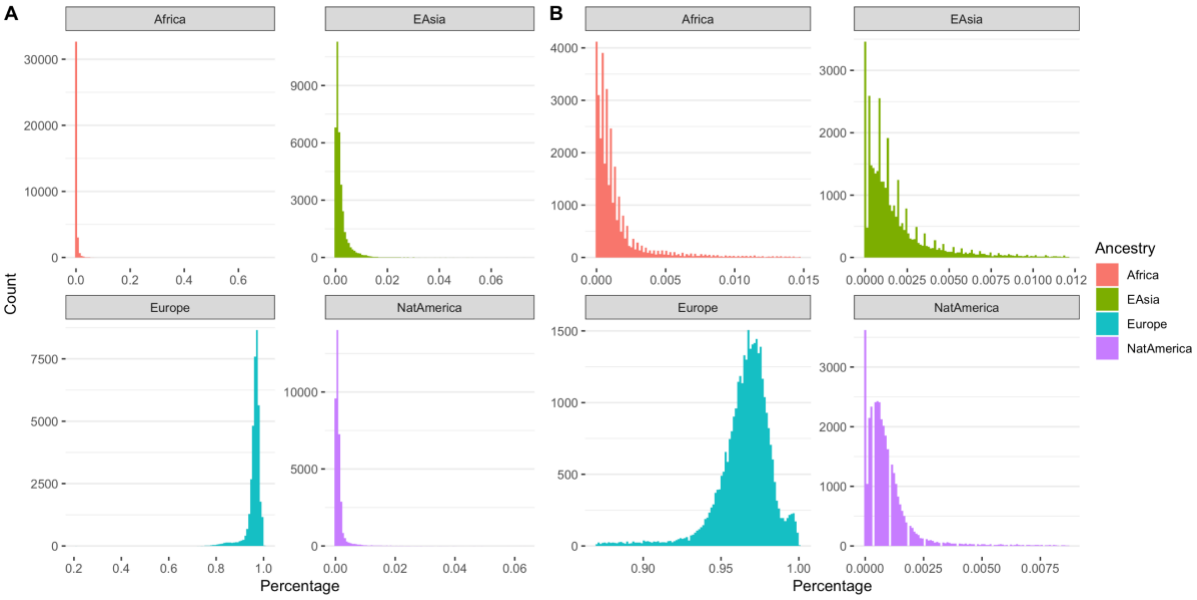
	GREML -LDMS – No heterozygosity QC step but removed diagonals and off-diagonals extreme values across all GRMs				4	2	8	160 WGS	0.71 (0.10)	0.31 (0.10)
	GREML-LDMS – Relatedness cutoff 0.025 – (median based)	SVM WGS	24210	Alg1	4	2	8	48 WGS	0.64 (0.10)	0.23 (0.10)
	GREML-LDMS – Relatedness cutoff 0.025 – (tertile based)				4	3	12	48 WGS	0.72 (0.10)	0.31 (0.10)
	GREML-LDMS – Relatedness cutoff 0.025 – (quartile based)				4	4	16	48 WGS	0.72 (0.10)	0.29 (0.11)
	GREML-LDMS – WGS SNPs – 2 LD bins (median based) + African ancestry covariate		25465		4	2	8	48 WGS + % African ancestry	0.60 (0.09)	0.24 (0.10)
	GREML-LDMS – WGS SNPs – 3 LD bins (tertile based) + African ancestry covariate				4	3	12	48 WGS + % African ancestry	0.68 (0.09)	0.31 (0.10)
	GREML-LDMS – WGS SNPs – 4 LD bins (quartile based) + African ancestry covariate				4	4	16	48 WGS + % African ancestry	0.68 (0.10)	0.30 (0.10)
UK Biobank Whole Exome	GREML-LDMS + HM3 GRM	WES + HM3	35867	Alg0	7	2	14 + 1 (HM3)	20 HM3	0.62 (0.04)	0.33 (0.04)
	GREML-LDMS + HM3 GRM – fitting birth coordinates				7	2	14 + 1 (HM3)	20 HM3	0.61 (0.04)	0.33 (0.04)
	GREML-LDMS + HM3 GRM				7	2	14 + 1 (HM3)	280 WES (20 per bin)	0.59 (0.04)	0.31 (0.04)
	GREML-LDMS + HM3 GRM – fitting birth coordinates				7	2	14 + 1 (HM3)	280 WES (20 per bin)	0.56 (0.04)	0.25 (0.05)
	GREML-LDMS + HM3 GRM – fitting sequencing center				7	2	14 + 1 (HM3)	280 WES (20 per bin)	0.57 (0.04)	0.28 (0.04)

Supplementary Table 8: Putative impacts and enriched GREML-LDMS analysis bin of variant effects as predicted by SnpEff v4.1 annotation software.

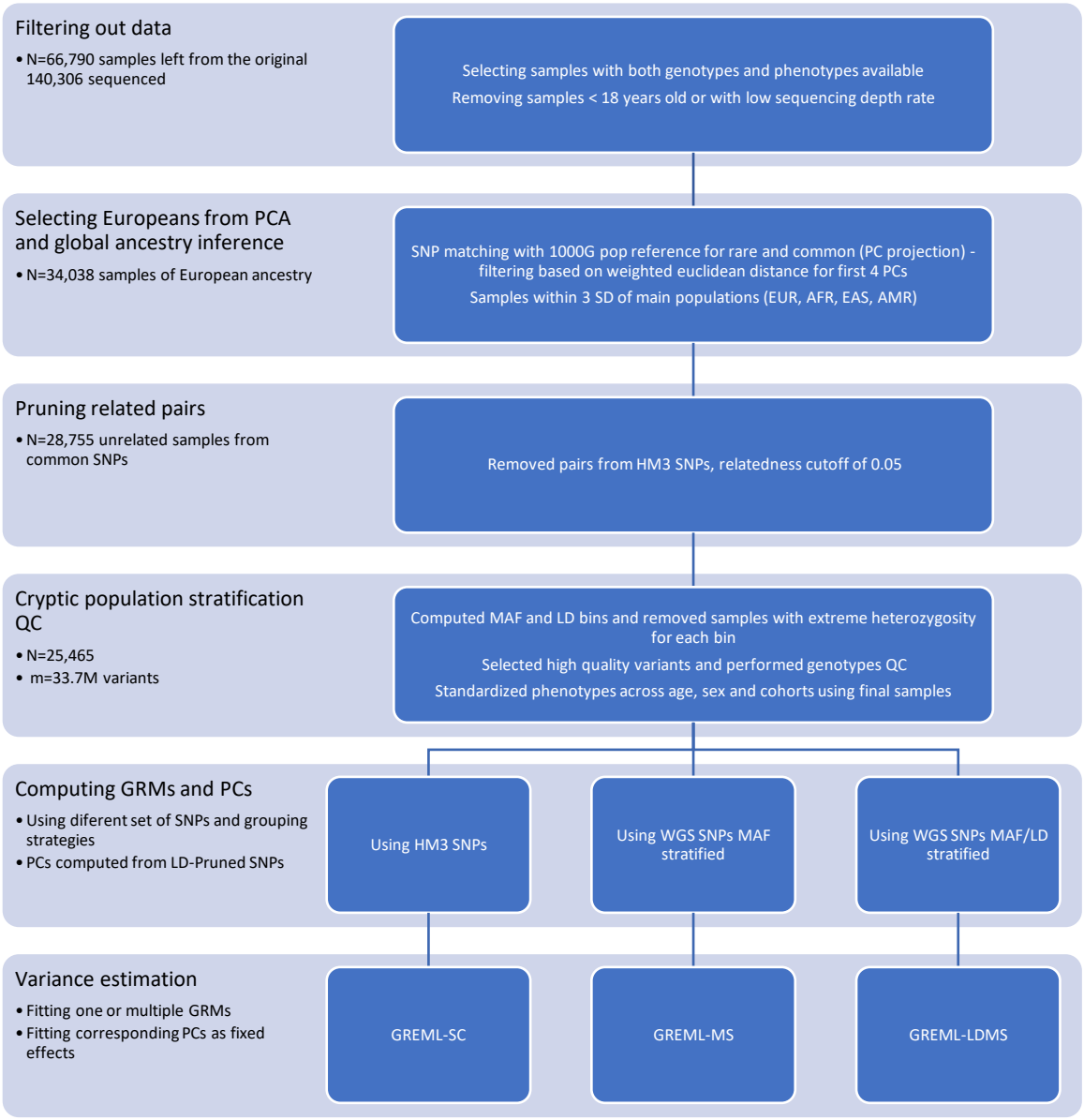
Enriched GREML-LDMS bin analysis	Putative impact	Sequence Ontology term
Protein-altering variants	HIGH	chromosome_number_variation
		exon_loss_variant
		frameshift_variant
		rare_amino_acid_variant
		splice_acceptor/donor_variant
		start_lost
		stop_gained/lost
		transcript_ablation
	MODERATE	3_or_5_prime_UTR_truncation & exon_loss
		coding_sequence_variant
		conservative_inframe_insertion/deletion
		disruptive_inframe_insertion/deletion
		missense_variant
		regulatory_region_ablation
		splice_region_variant
		TFBS_ablation
Non-protein-altering variants	LOW	5_prime_UTR_premature_start_codon_gain_variant
		initiator_codon_variant
		splice_region_variant
		start/stop_retained_variant
		synonymous_variant
	MODIFIER	3_or_5_prime_UTR_variant
		coding_sequence_variant
		conserved_intergenic_variant
		conserved_intron_variant
		downstream_gene_variant
		exon_variant
		feature_elongation/truncation
		gene_variant
		intragenic/intergenic_region
		intron_variant
		mature_miRNA_variant
		miRNA
		NMD_transcript_variant
		non_coding_transcript_exon_variant
		non_coding_transcript_variant
		regulatory_region_amplification
		regulatory_region_variant
		TF_binding_site_variant
		TFBS_amplification
		transcript_amplification
		transcript_variant
		upstream_gene_variant



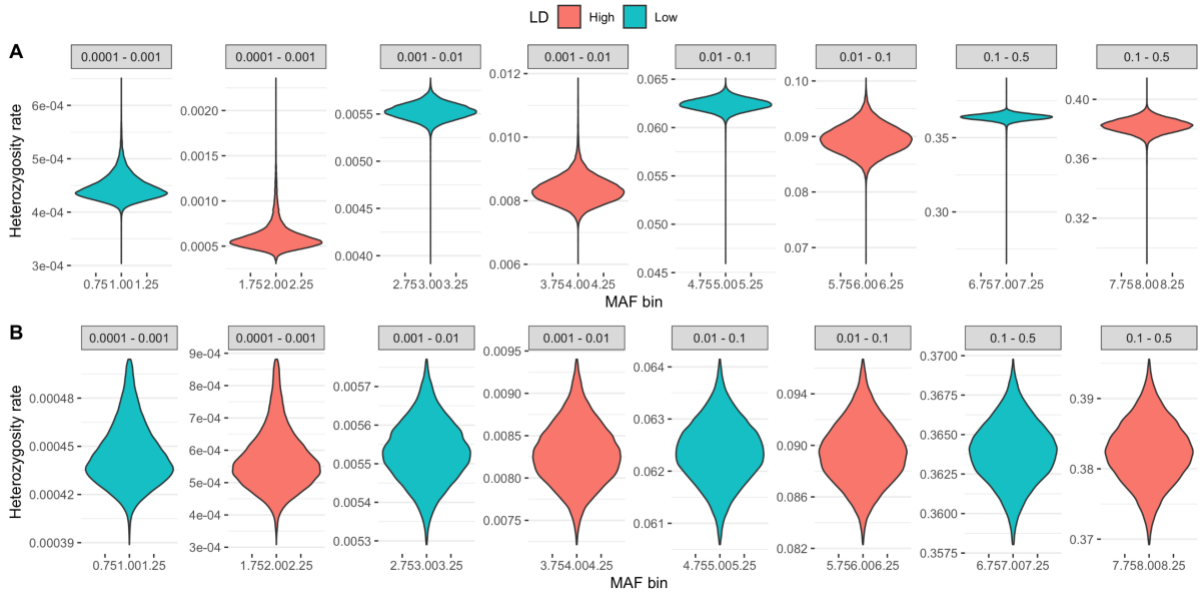
Supplementary Figure 1: Principal Component Analysis plot of individuals compared to 1000 Genomes populations. (A) PCA of TOPMed samples and 1000G populations using 580k common SNPs. (B) PCA of TOPMed samples and 1000G populations using ~1.3M rare SNPs.



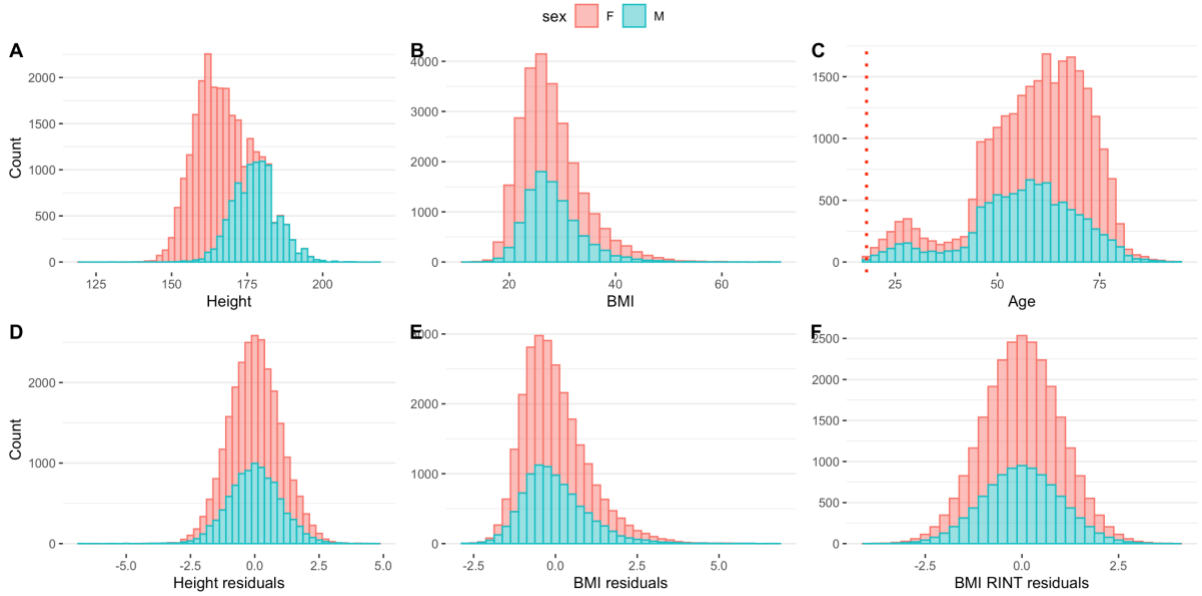
Supplementary Figure 2: Proportion of ancestry according to RFMix reference before and after filtering samples further away from 3 standard deviations of each reference population. (A) Before filtering samples (N=36,938). (B) After filtering samples (N=34,038).



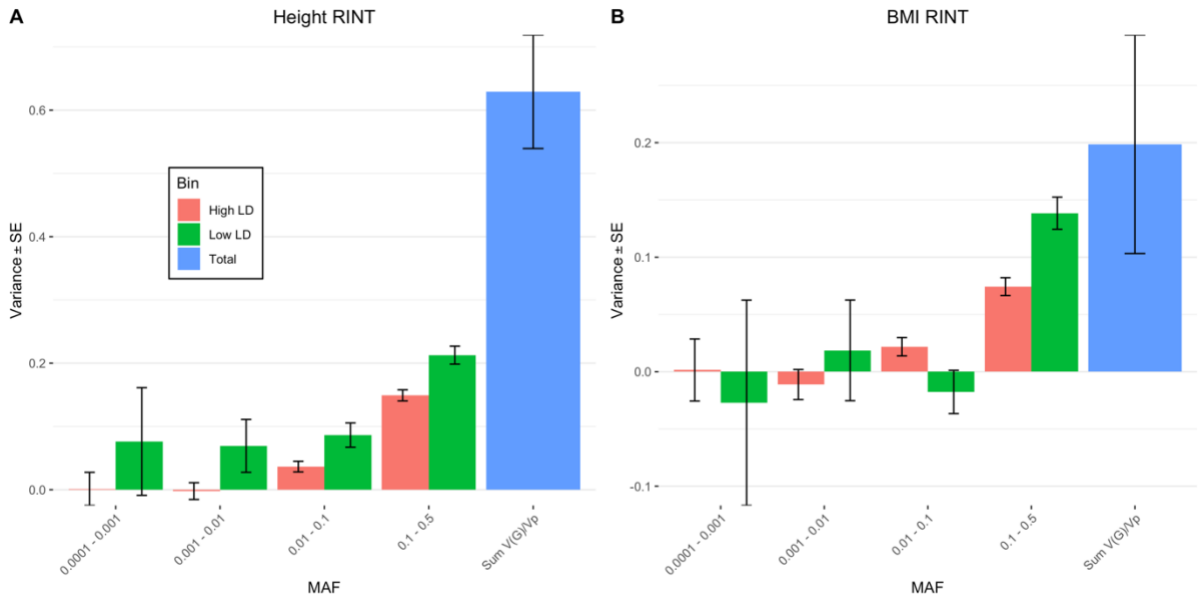
Supplementary Figure 3: Quality control and analysis pipeline for TOPMed WGS data to estimate trait variance.



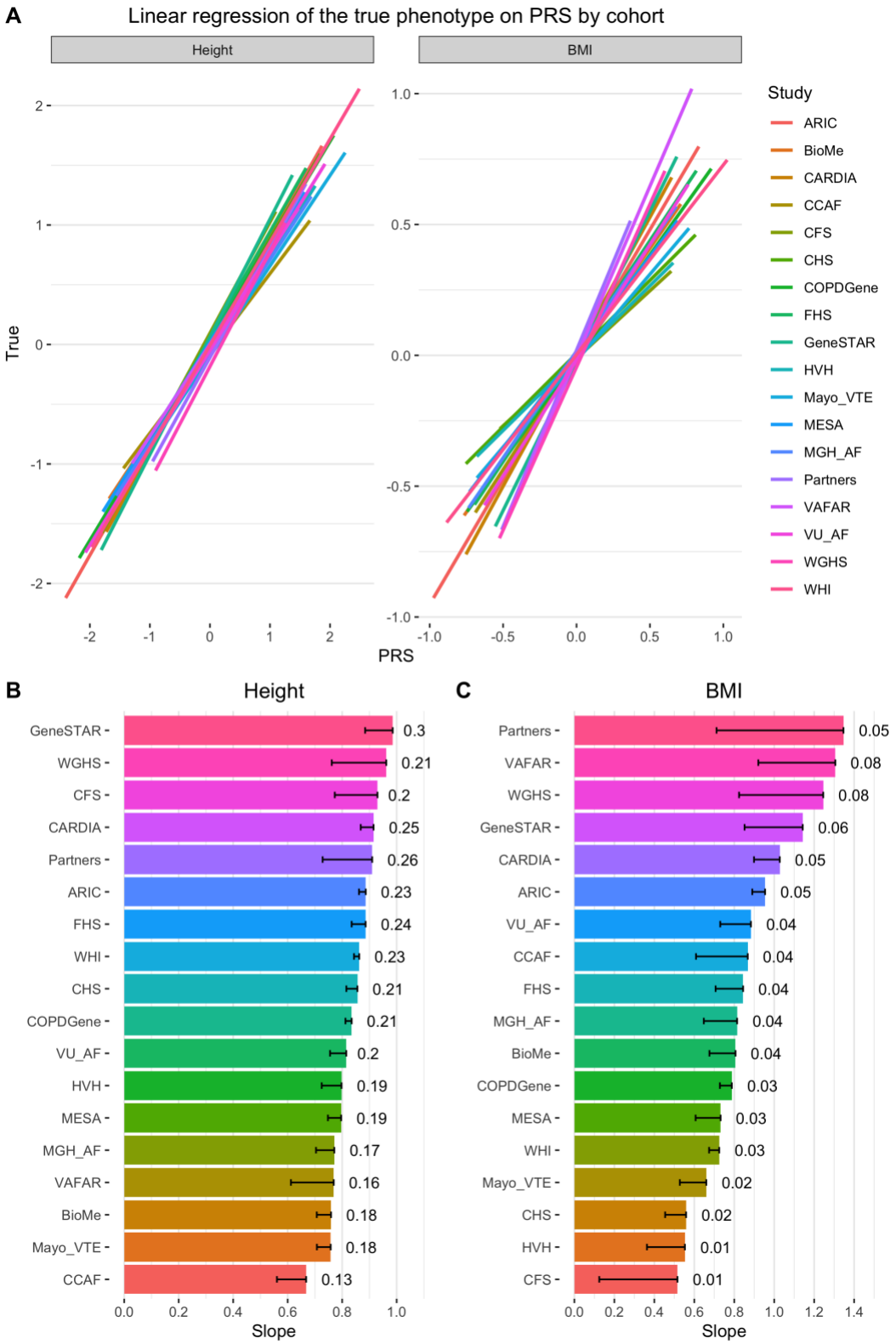
Supplementary Figure 4: Distribution of sample heterozygosity for each MAF and LD grouping. (A) Distribution before filtering (N=28,755). (B) Distribution after 4 rounds of filtering out samples further away than 3 standard deviations for each distribution mean (N=25,465).



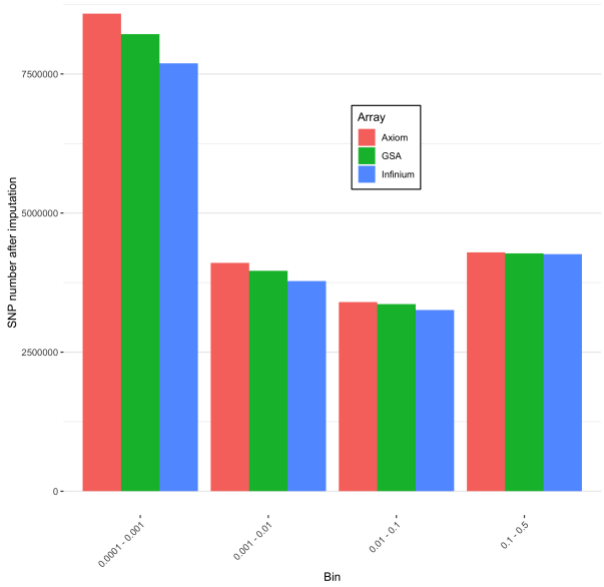
Supplementary Figure 5: Distribution of N=25,465 phenotypes (unrelated Europeans) in the dataset before and after the QC process. (A) Distribution of height before standardization. (B) Distribution of BMI before standardization. (C) Distribution of age in the dataset with indivuduals < 18 years old removed. (D) Distribution of the residuals for height after QC standardization. (E) Distribution of the residuals for BMI after QC standardization. (F) Distribution of the residuals for BMI with a rank inverse normal transformation after QC standardization. The skewness in distribution of the BMI is removed by the transformation.



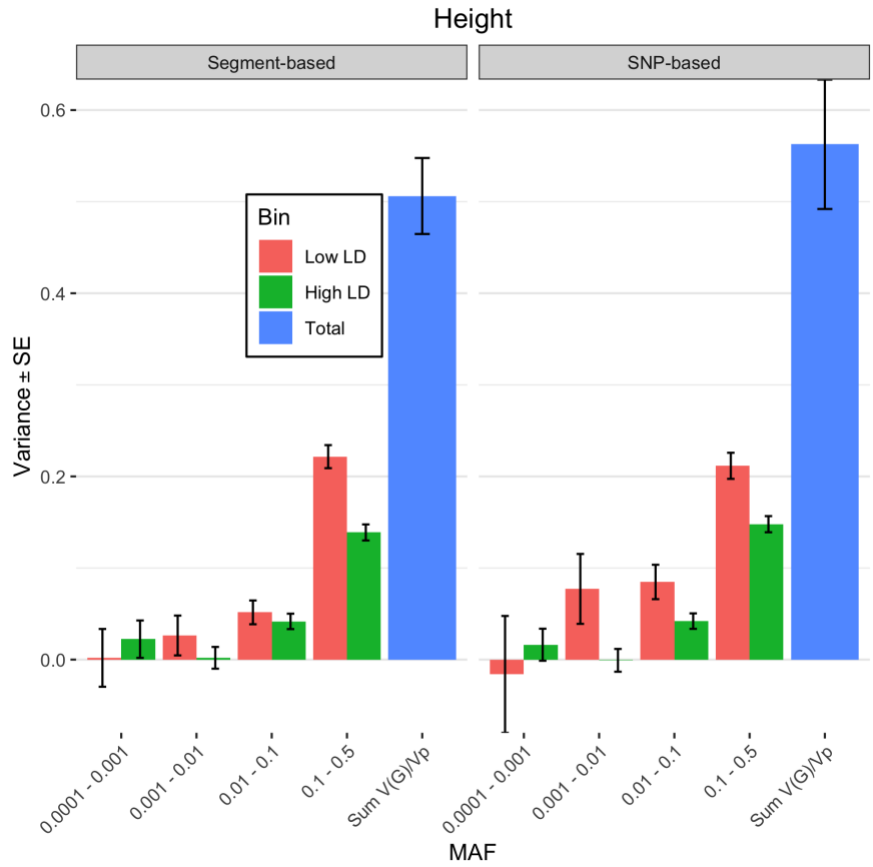
Supplementary Figure 6: GREML-LDMS estimates using WGS (~33.7M variants, N=25,465) with a rank inverse normal transformation (RINT) correcting by 48 PCs. (A) Estimates for height$_{RINT}$ ~0.63 (SE 0.09) are consistent with untransformed trait. (B) Total estimates for BMI$_{RINT}$ are lower than BMI ~0.20 (0.10).



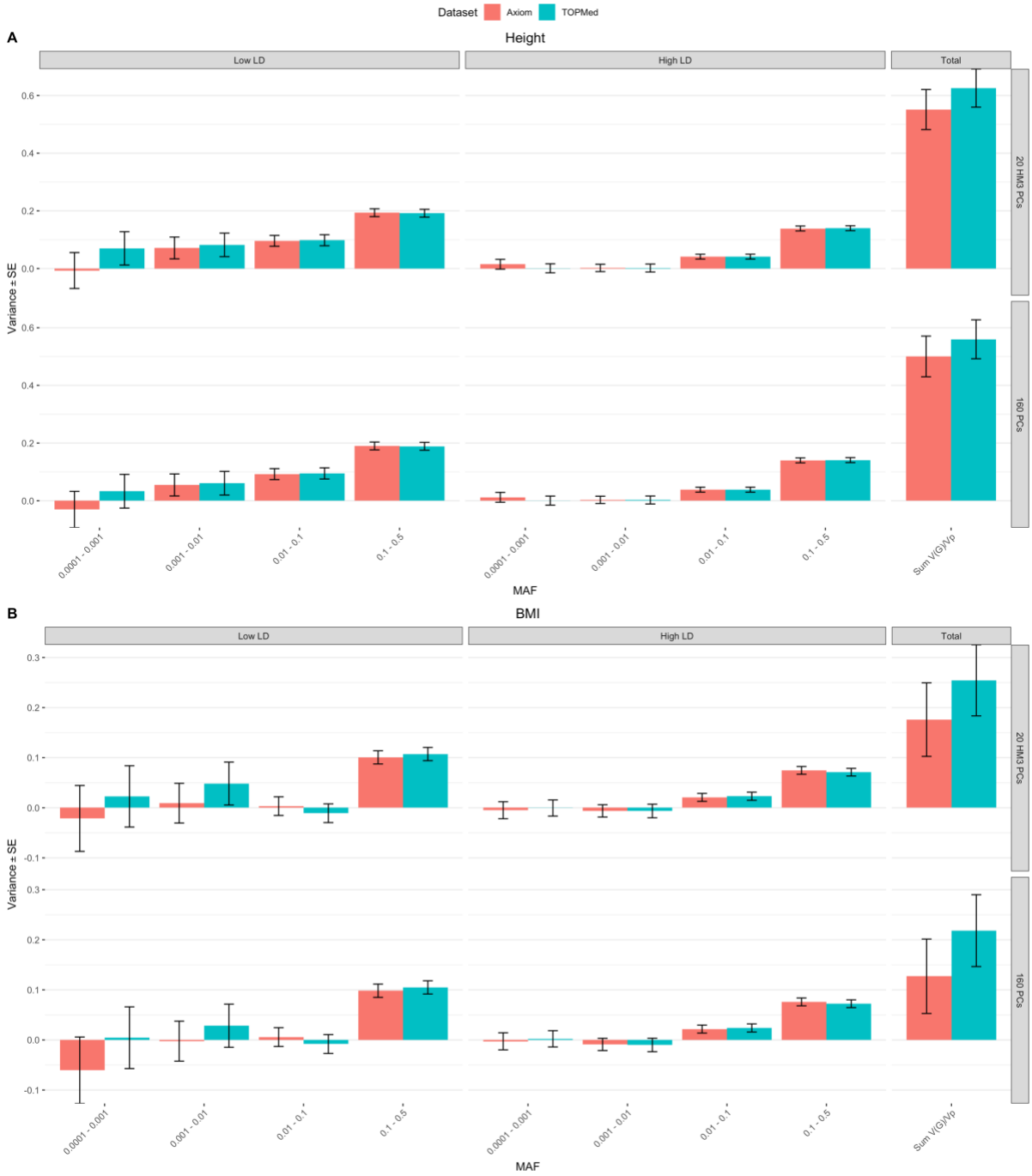
Supplementary Figure 7: PGS in n=25,465 samples from 1360 and 449 independent SNPS associated with height and BMI respectively in the UKB matching TOPMed dataset. (A). Fitted regression slopes for samples in each 18 cohorts. (B-C). Estimates of the individual slopes (x-axis) and the proportion of phenotypic variance explained (numbers displayed) for each TOPMed cohort, error bars represent the SE associated to the ß estimate (slope).



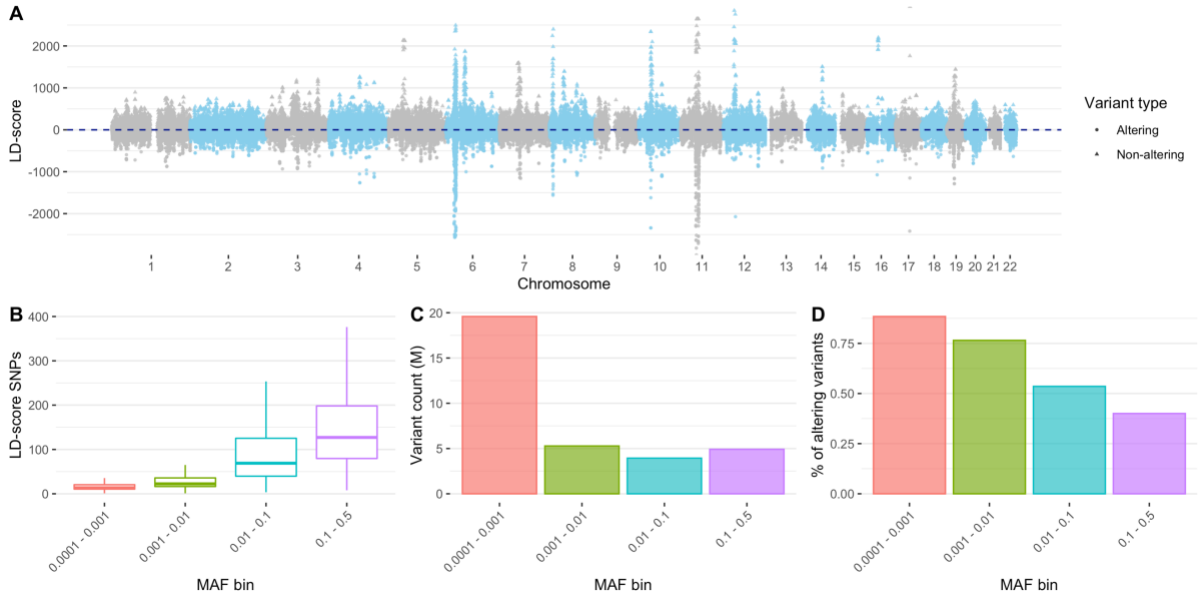
Supplementary Figure 8: Number of SNPs in each MAF bin after imputation and QC process, for each of the Illumina InfiniumCore24, GSA-24 and Affymetrix Axiom arrays. Total number of SNPs is between ~19.0 and ~20.0 M SNPs.



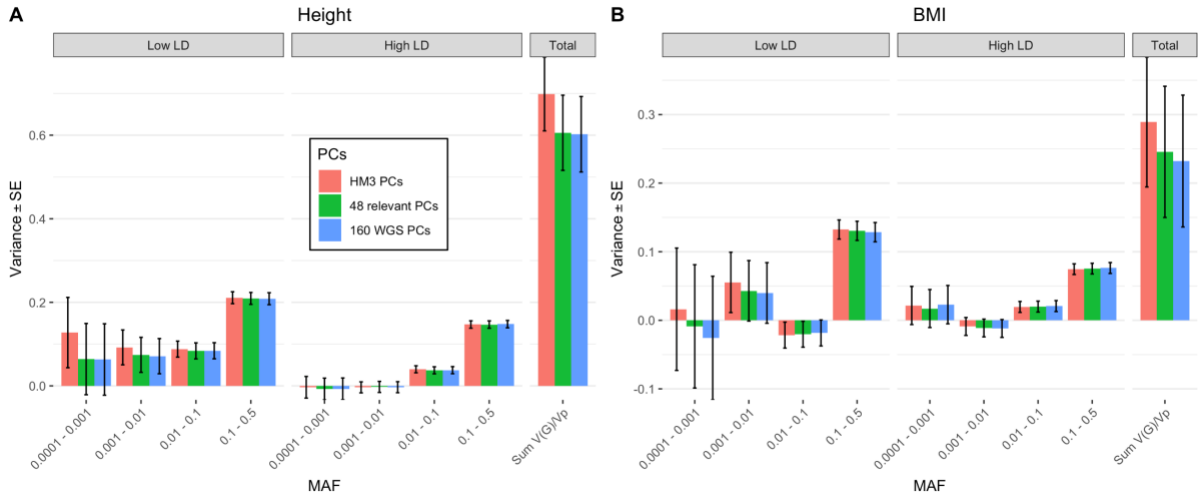
*Supplementary Figure 9: Estimates of SNP-based heritability (h_{SNP}^2) from a GREML-LDMS analysis based on 8 bins (4 MAF * 2 LD bins, N=25,465) for SNPs imputed on Axiom array (~20.4M SNPs), corrected for 20 HM3 SNPs PCs. LD stratification was done using either on segment-based LD value or individual-SNP LD value. Estimates went from h_{SNP}^2 ~0.51 (SE 0.04) using a segment-based LD value stratification to h_{SNP}^2 ~0.56 (SE 0.07) using an individual-SNP LD value stratification.*



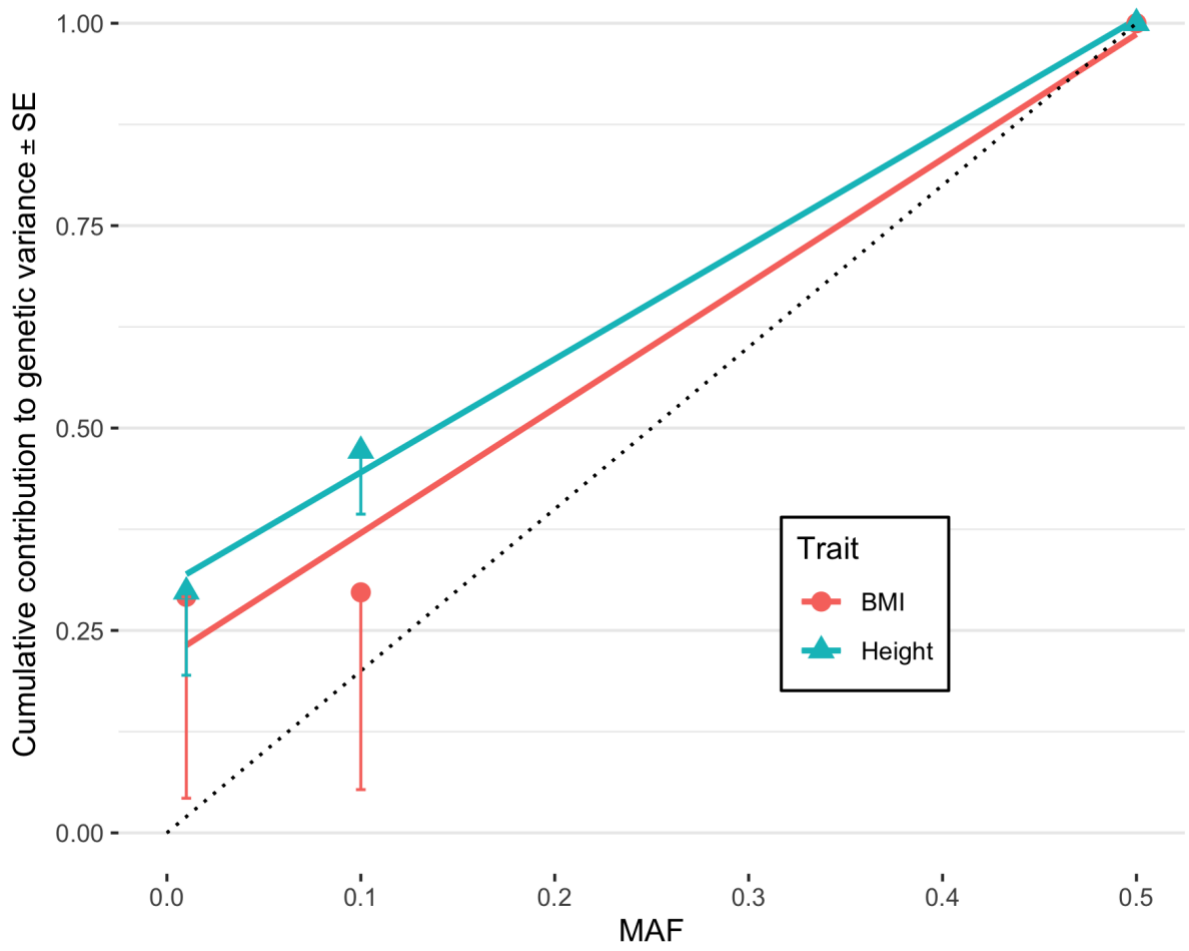
Supplementary Figure 10: GREML-LDMS estimates stratified in 14 MAF and LD bins, using subset of SNPs in common between TOPMed dataset and imputed dataset from Axiom array SNPs (~17.9M SNPs in common, N=25,465) and correcting by either 20 PCs computed from HM3 independent SNPs or 160 PCs from each MAF/LD bin. Estimates are consistently higher for the TOPMed dataset, with most of the difference coming from the variants in the 0.0001 to 0.001 MAF range in low LD. This difference could be due to imputation errors as common variant bins seems consistent between the two datasets.



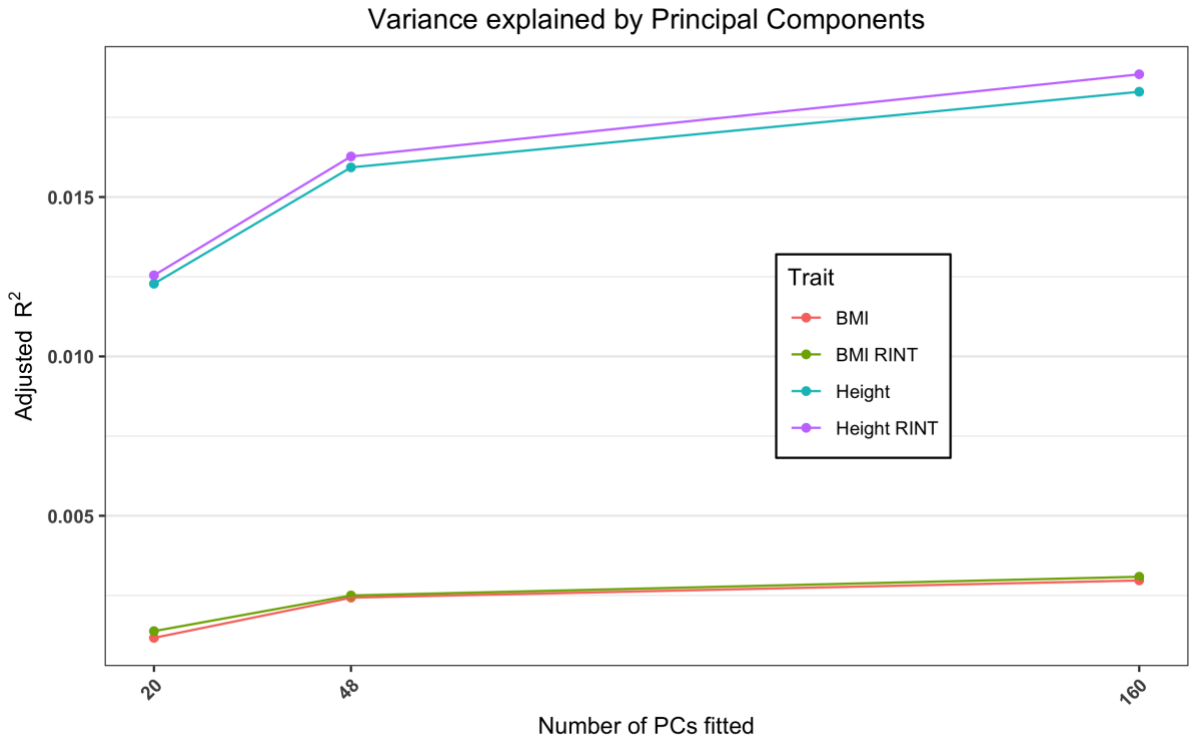
*Supplementary Figure 11: (A) LD score value across the genome for a subset of 350k protein-altering and non-protein-altering variants. LD value were calculated within each of the 4 MAF bin, LD values > 3000 are not shown (B) Boxplot of the distribution of individual SNPs LD values within each bin, N=33.7M. Lower and upper hinges correspond to the first and third quartiles and whiskers extend to +/- 1.5 * inter-quartile range. (C) Number of SNPs for each of the four MAF bin. (D) Fraction of the high impacting variants as a percentage of the total number of variants in the bin.*



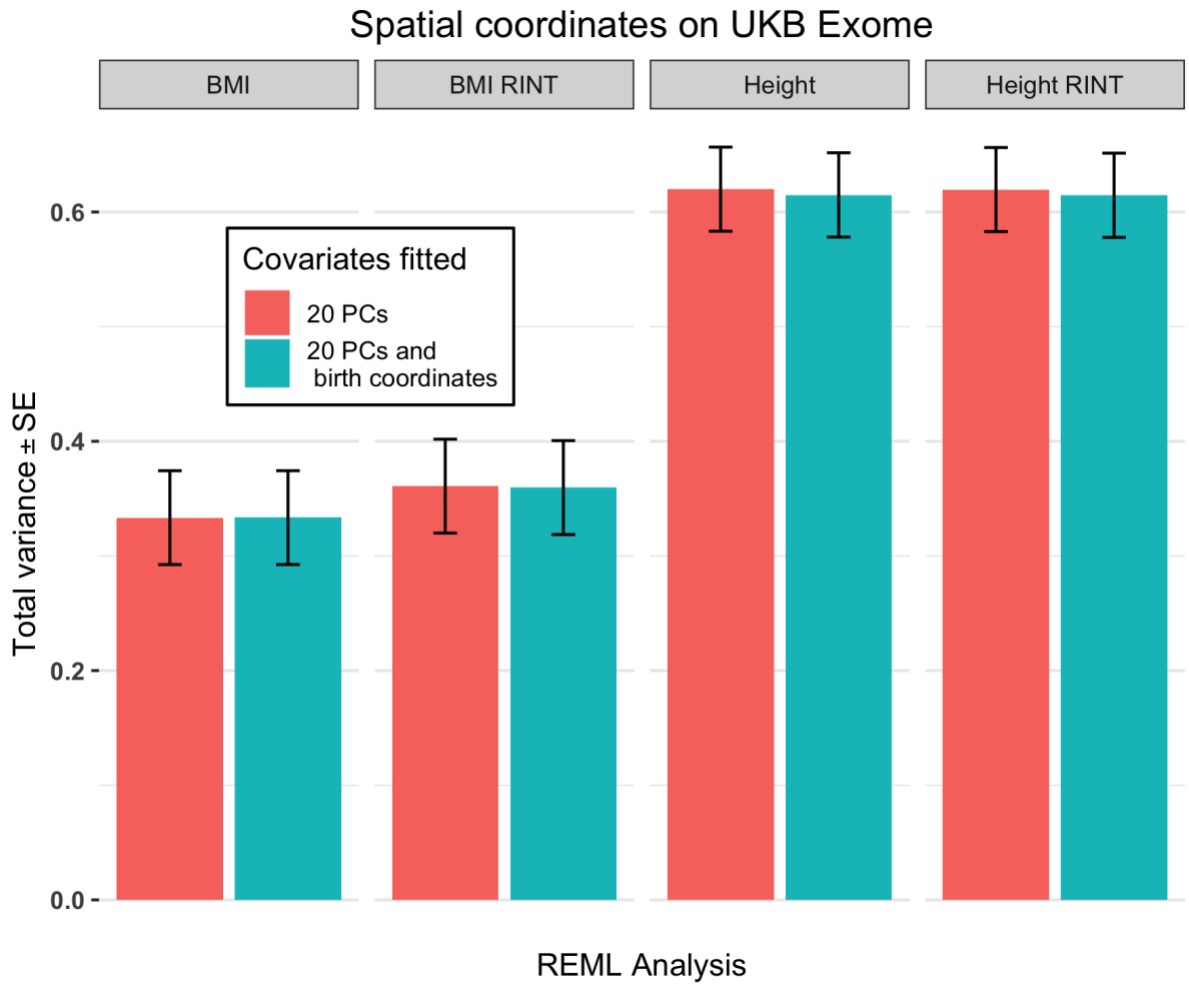
*Supplementary Figure 12: GREML-LDMS estimates from WGS data (~33.7M variants, N=25,465) stratified in 8 bins (4 MAF bins in 2 LD bins) with correction for 20 PCs (on HM3 SNPs), 48 PCs reflecting population stratification (Supplementary Figure 23) or 160 PCs (20 * 8 bins). (A) Estimates for height with h^2_{WGS} at ~0.60 - ~0.70 (SE ~0.09). (B) Estimates for BMI with hWGS2 at ~0.25 - ~0.29 (SE 0.09 – 0.10). The number of variants in each of the 4 MAF bins (twice the number in each LD bin) is, from the lowest to highest MAF bins, 19.3M, 5.3M, 3.9M and 4.9M, respectively (Supplementary Table 5 and Supplementary Figure 11).*



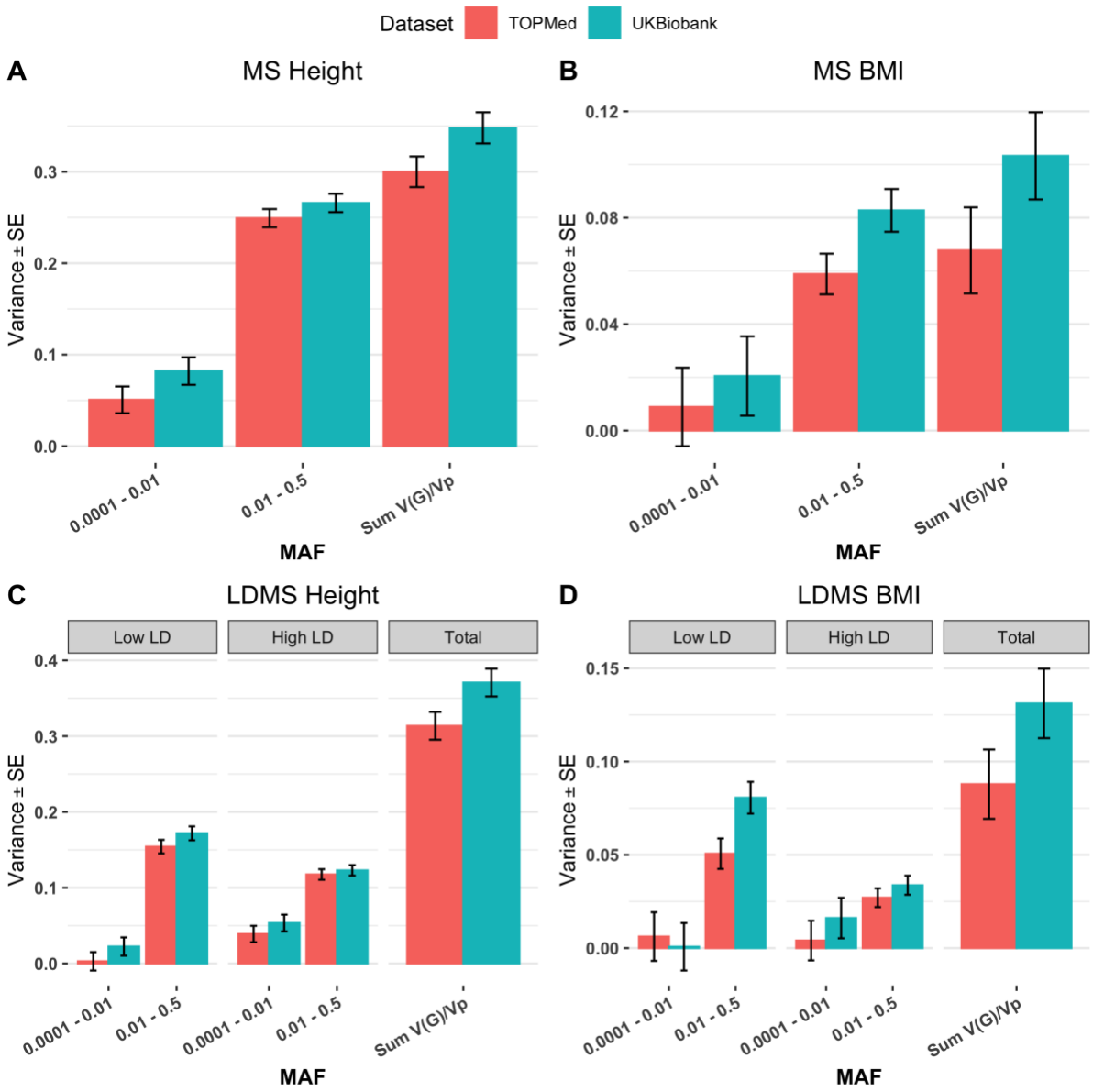
Supplementary Figure 13: Estimate of the cumulative contribution of variants, for height and BMI, from GREML-LDMS analysis (N=25,465). The dotted line represents the expected contribution under a neutral evolutionary model. The deviation from this dotted line suggest that height and BMI are under negative selection. For each trait, a linear model has been fitted to better visualise the trait selection. The first bin represents variants with MAF < 0.01 (cumulative contribution to genetic variance for height and BMI of 0.30 (SE 0.10) and 0.29 (SE 0.25), respectively) when second bin shows all variants with MAF < 0.1 (cumulative contribution to genetic of 0.47 (SE0.08) for height and 0.30 (SE 0.24) for BMI). The third MAF bin includes all variants in the dataset.



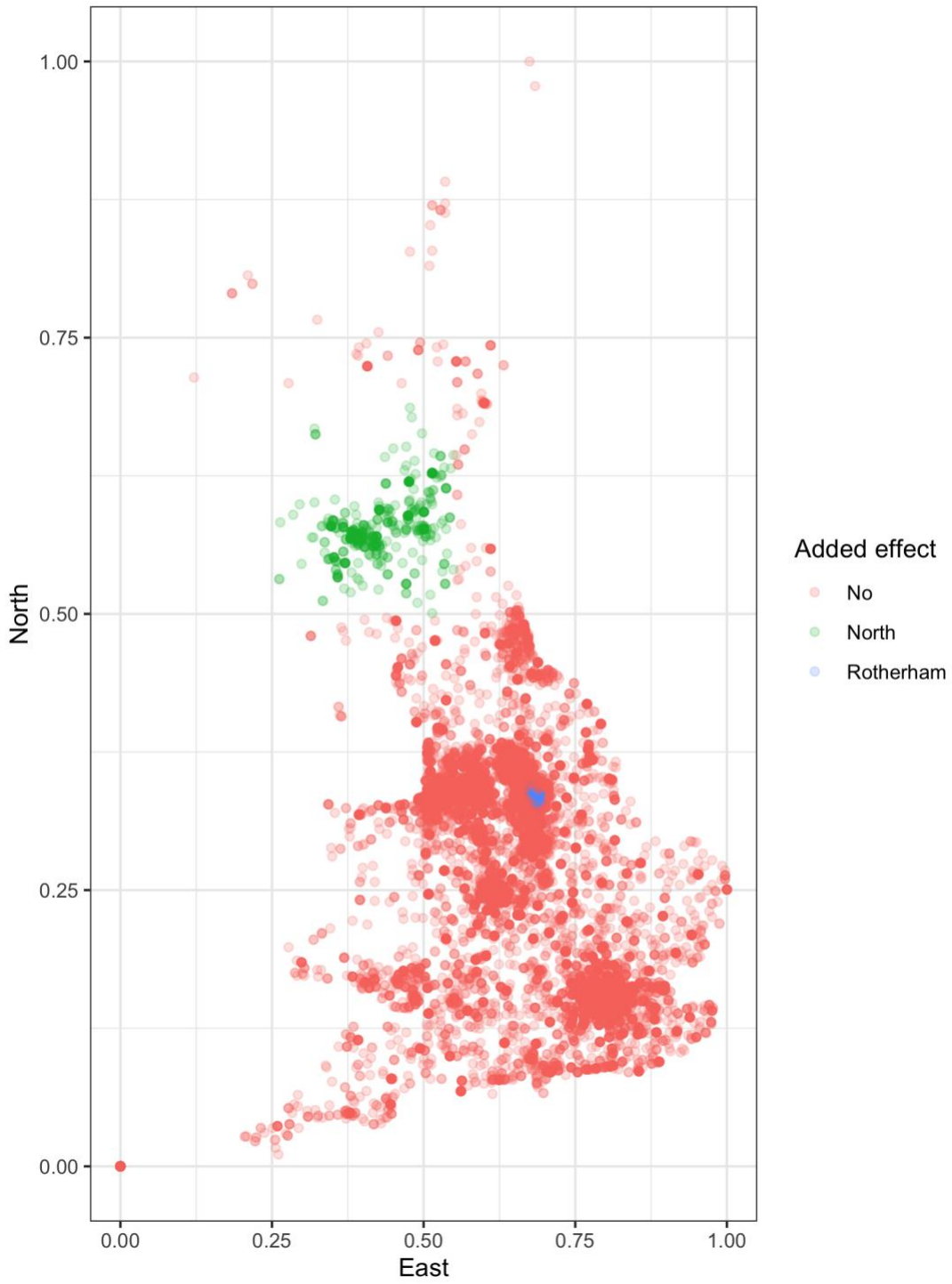
*Supplementary Figure 14: Adjusted R^2 of a linear model regressing the phenotype against 20 PCs from LD-pruned HM3 SNPs and 48 (calculated from 8 MAF/LD bins of LD-pruned SNPs) or 160 (calculated from 4 MAF bins * 2 LD bins, on independent SNPs) PCs calculated from TOPMed WGS dataset.*



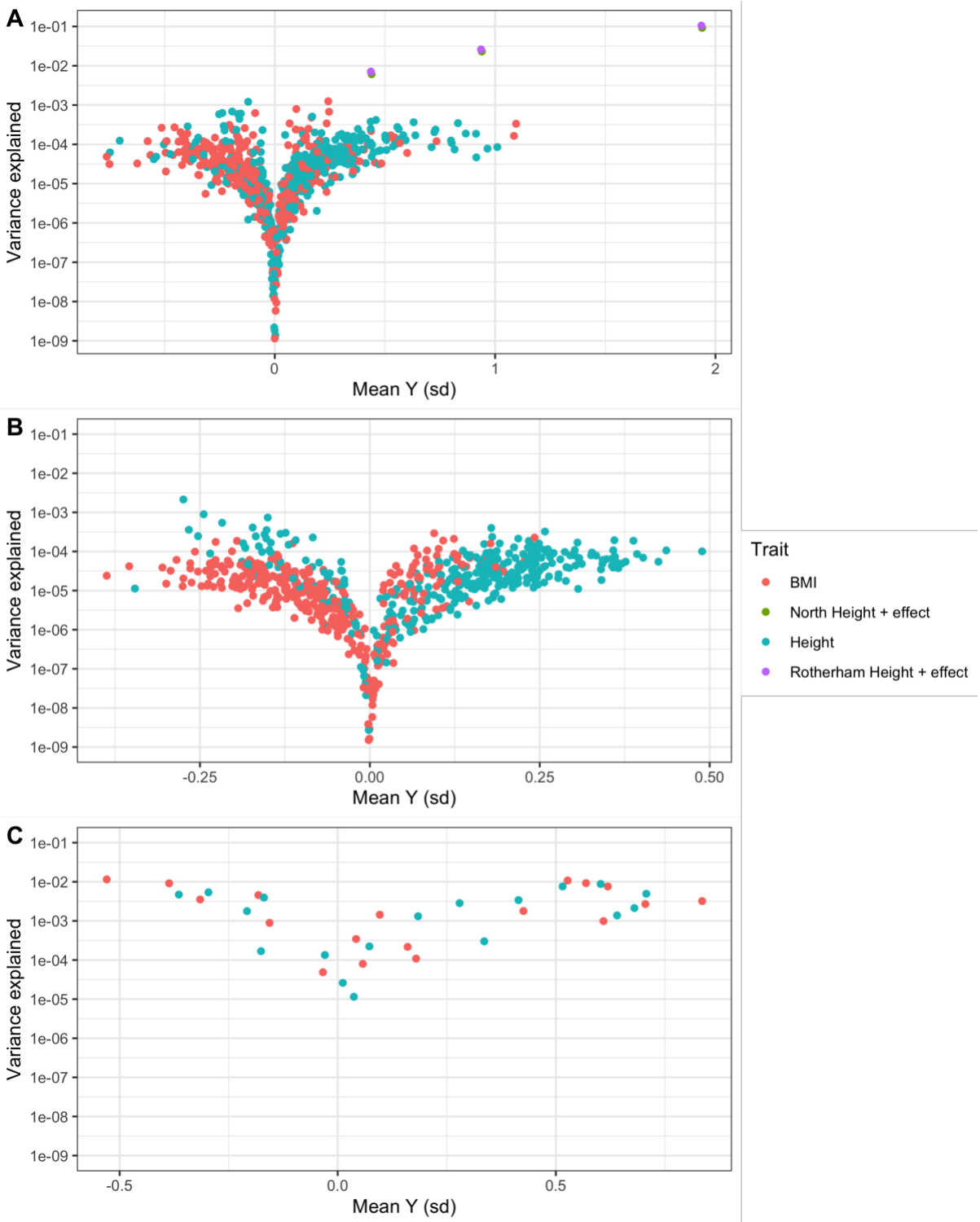
Supplementary Figure 15: GREML-LDMS estimates fitting 14 bins from UKB Exome data and one bin of HM3 imputed common SNPs. Estimates fitting either 20 PCs calculated from HM3 SNPs or 20 PCs and the north and east birth coordinates scaled on a 0,1 range. Estimates are of ~0.61-0.62 (SE 0.04) for height and height$_{RINT}$, 0.33 (SE 0.04) for BMI and 0.36 (SE 0.04) for BMI$_{RINT}$.



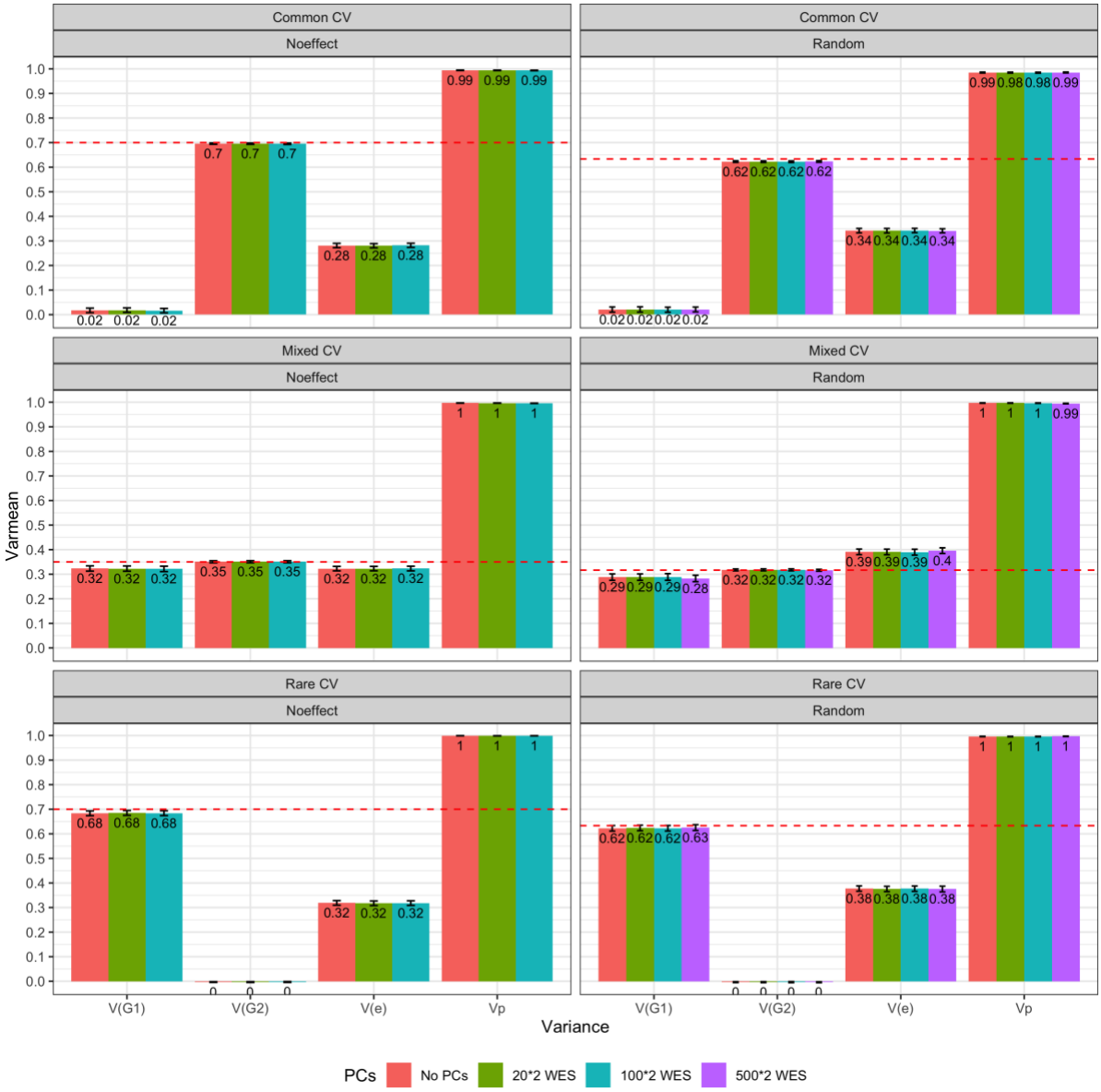
Supplementary Figure 16: GREML estimates using exome SNPs present in both TOPMed and UKB Exome dataset. Only 2 MAF groupings are fitted in this mode with rare (0.0001 < MAF < 0.01) or common (0.1 < MAF < 0.5) variants. The sample size of UKB Exome dataset was downsampled to match TOPMed's sampe size (N=25,465 unrelated european individuals). Rare variants not following a normal distirbution were removed (Online Methods). We further investigated the effect of LD stratification with the same set of variants in (C) and (D). For TOPMed and UKB Exome respectively. (A) Total estimates for height of 0.30 (SE 0.02) and 0.35 (SE 0.02). (B) Estimates for BMI of 0.07 (SE 0.02) and 0.10 (SE 0.02) for BMI. (C) GREML-LDMS estimates for height 0.31 (SE 0.02) and 0.37 (SE 0.02). (D) Estimates for BMI of 0.09 (SE 0.02) and 0.13 (SE 0.02).



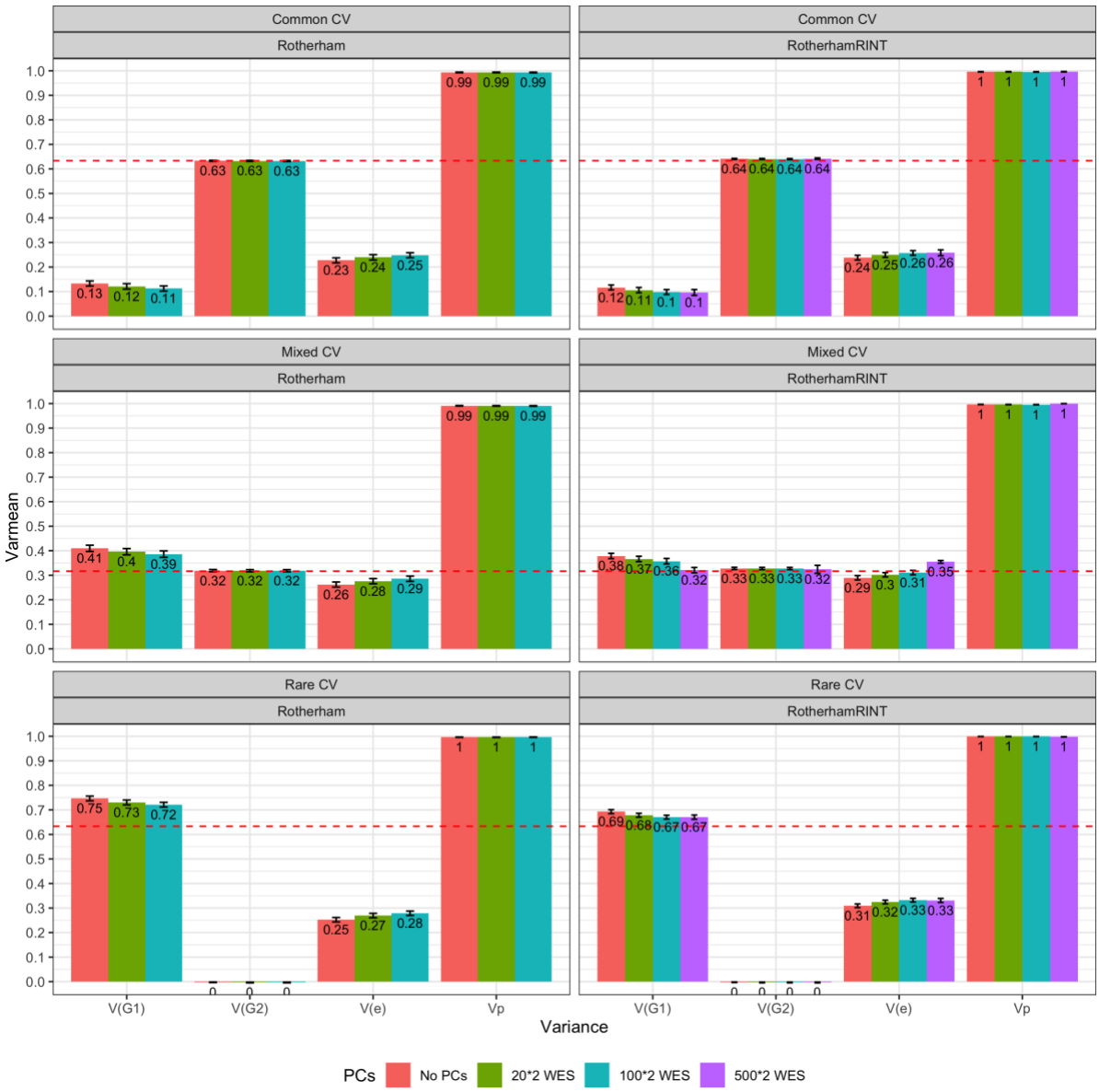
Supplementary Figure 17: Geographical distribution of UKB samples used to simulate a sharp environmental effect associated with their birth coordinates. To simulate a sharp environmental effect, we selected samples either from the Rotherham region or based on regional coordinates defined by rare variants PC stratification (Supplementary Figure 24 A).



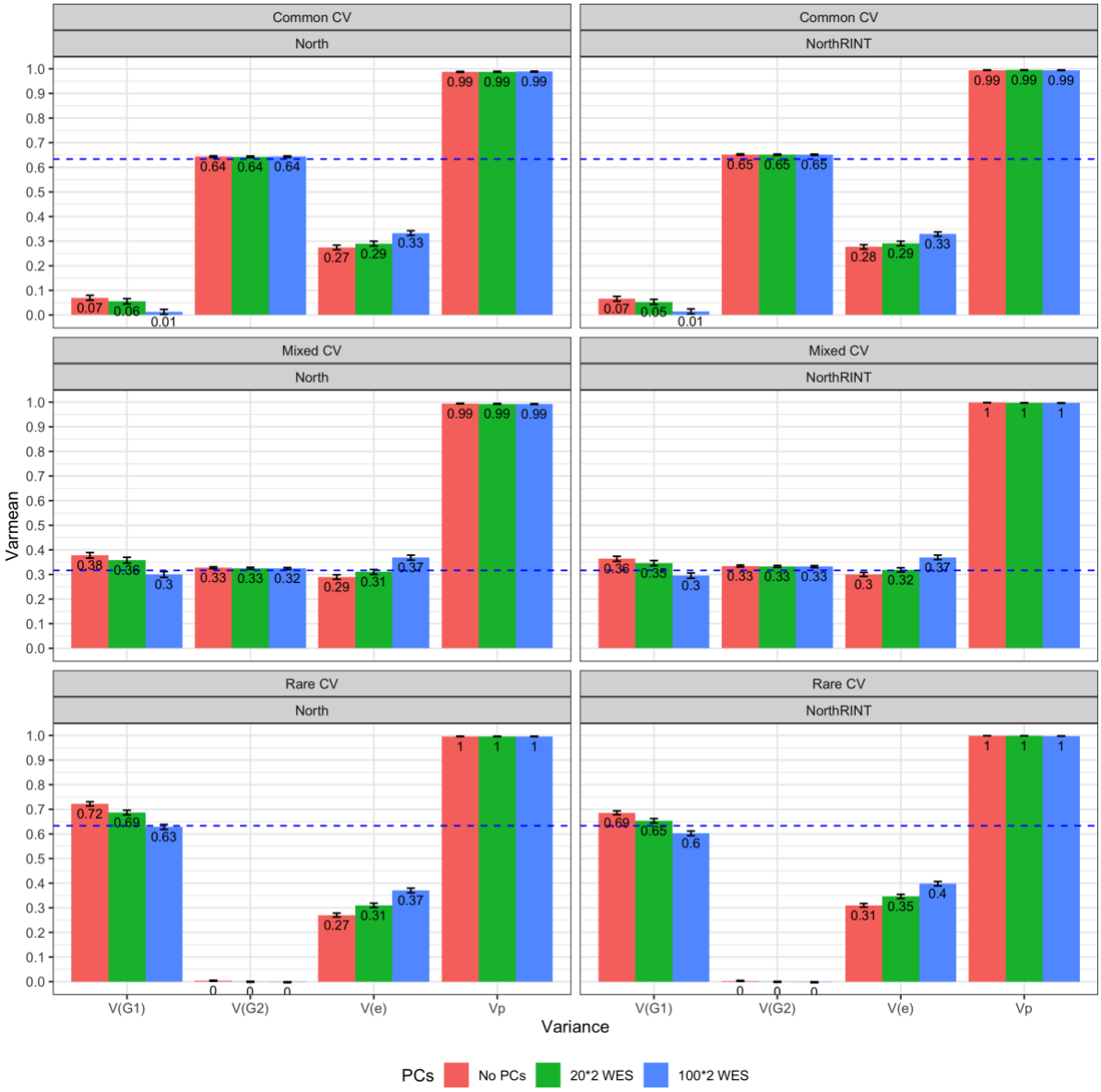
Supplementary Figure 18: Distribution of the mean standardized phenotype (sex and age adjusted) per administrative region/cohort in the UKB and TOPMed samples and their contribution to phenotypic variance, which was defined as $f(1-f)\beta^2$, where f is the proportion of the total sample for a region/cohort, and β the estimated mean phenotype. (A) Phenotypes distribution for the N=35,867 UKB WES samples for each of the 378 administrative regions. Also shown are the Rotherham and North samples after adding an environmental effect of 0.5, 1 and 2 sd to the mean height cohort value. (B) Mean phenotype per administrative region for the Europeans within the full UKB data (N=424,700) (C) Distribution of the mean height and BMI per TOPMed cohort, after correcting for sex and age. These phenotypes have also been standardized by cohort for the REML analysis. Note that the y-axis is on a log scale.



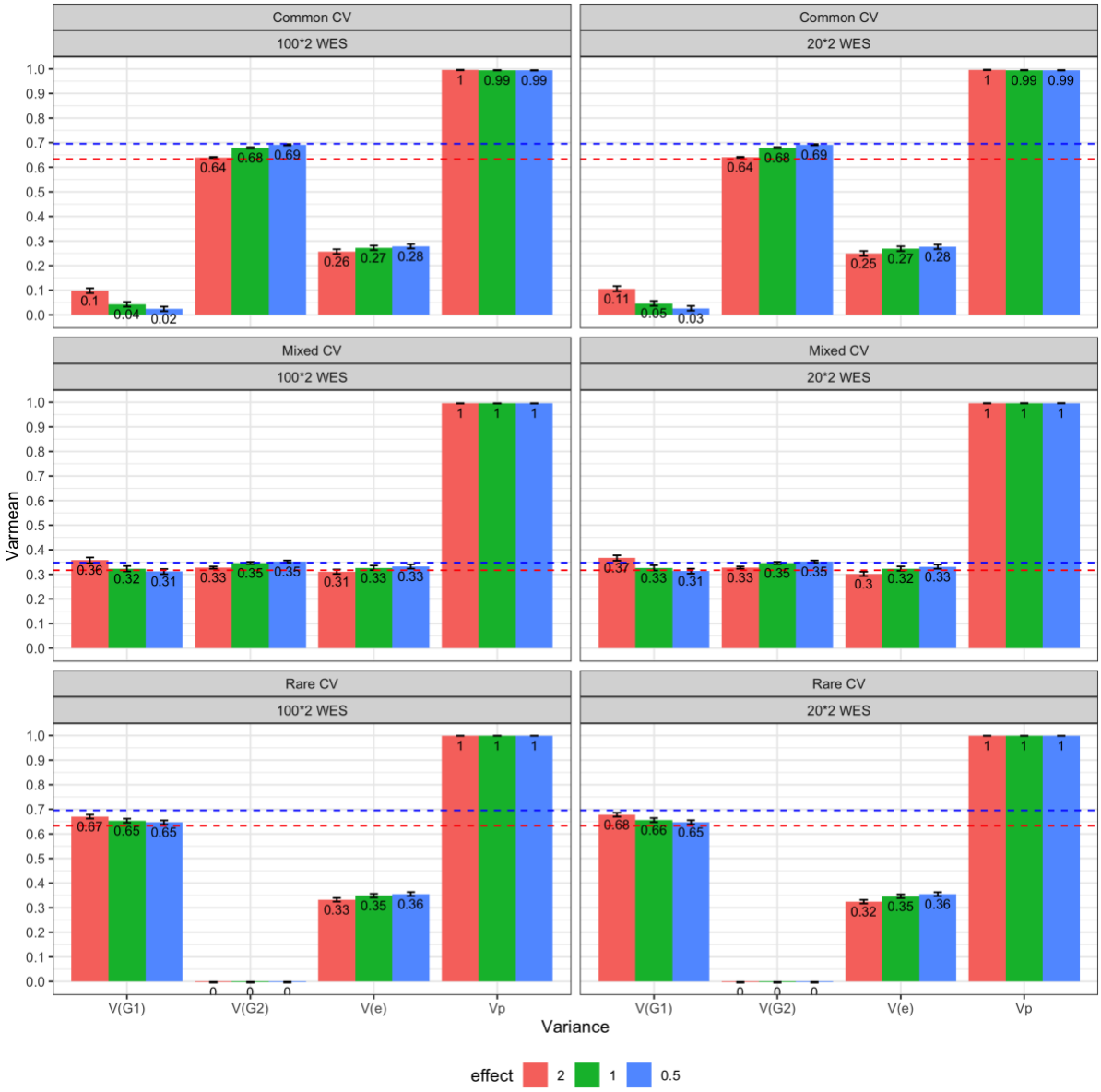
Supplementary Figure 19: Variance components when simulating a phenotype $h^2_{simu} = 0.70$ from 10,000 causal variants, either from common variants, rare variants, or a mixture of both (5,000 variants from each bin) over 10 iterations with N=35,867. Estimates without any environmental effect in the left panel, and with an effect of 2 standard deviations in the right panel, on 971 samples selected at random in the dataset. We fitted either 0 PCs, or 20/100/500 from each of the 2 sets of variants (after LD pruning). V(G1) corresponds to variants with MAC 3 < MAF < 0.01, when V(G2) is 0.01 < MAF < 0.5. The contribution from each GRM to genetic variance is as expected (dotted line) when adding a large environmental effect at random or no effect at all.



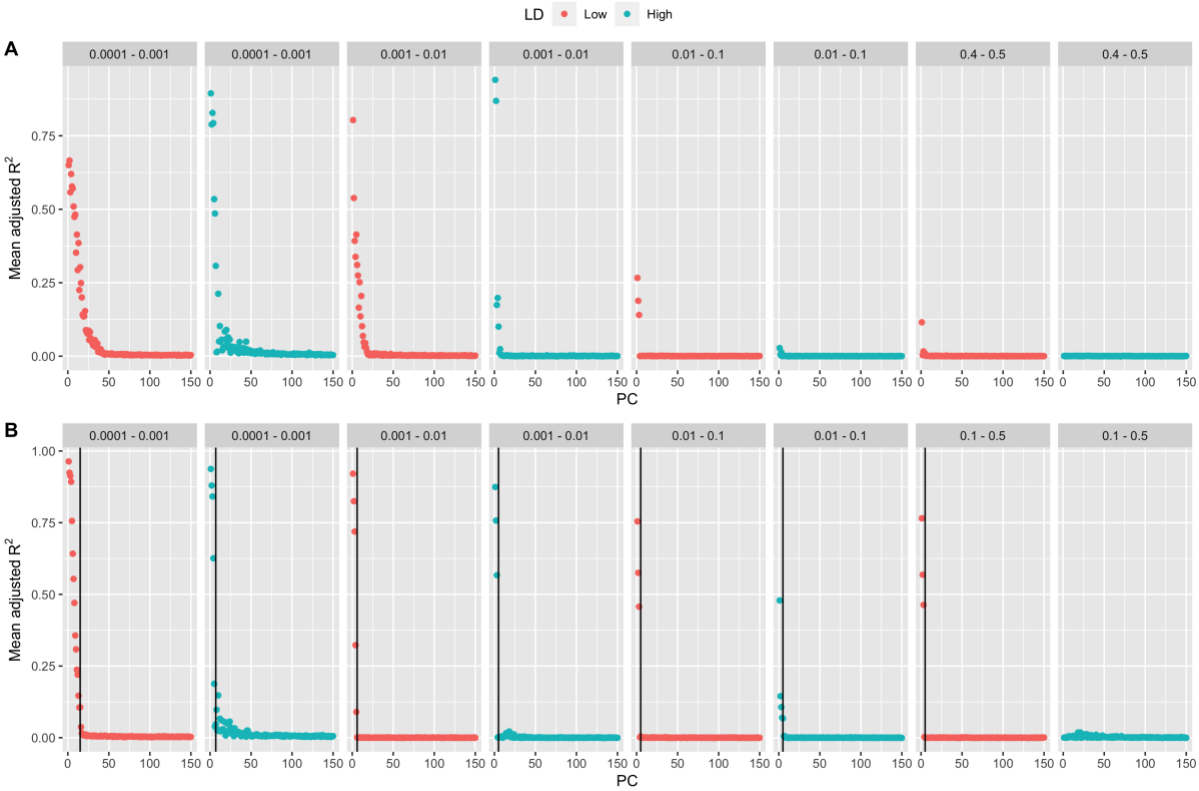
Supplementary Figure 20: Variance components when simulating a phenotype $h^2_{Simu} = 0.70$ from 10,000 causal variants, either from common variants, rare variants, or a mixture of both (5,000 variants from each bin) over 10 iterations with N=35,867. Estimates with a 2 standard deviations effect added on 971 samples from the Rotherham administrative region in the UKB in the left panel, and after a RINT transformation of the phenotype in the right panel. We fitted either 0 PCs, or 20/100/500 from each of the 2 sets of variants (after LD pruning). The variance contribution from the rare variants bin is inflated, especially with a phenotype simulated from common variants (by ~0.10). A RINT transformation or a large number of PCs fitted can only partially correct for this inflation.



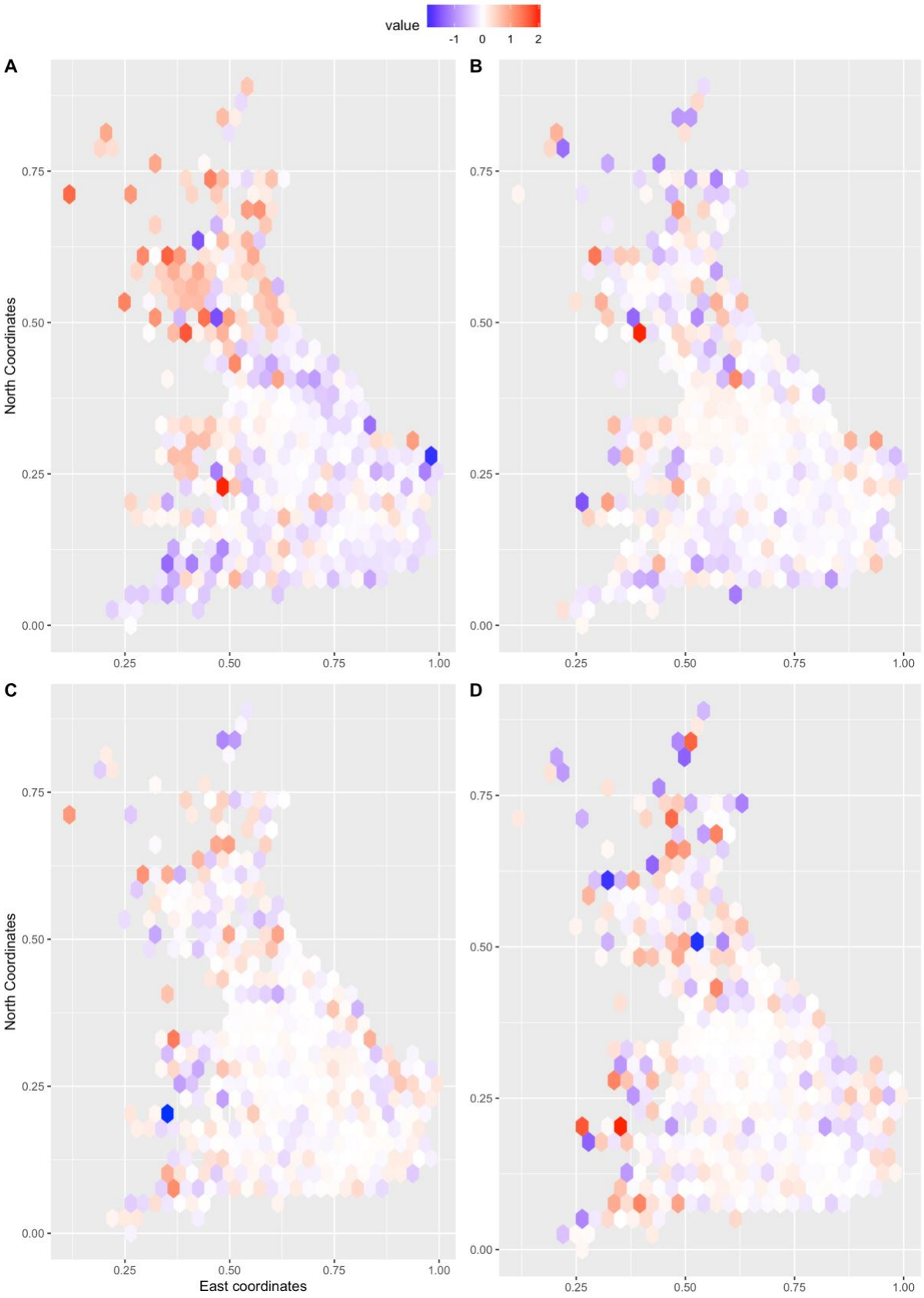
Supplementary Figure 21: Variance components for phenotypes with a 2sd effect added on the North samples (Supplementary Figure 24) in the left panel, and the phenotype after a RINT transformation in the right panel (N=35,867). As in Supplementary Figure 20, the variance contribution from rare variants is inflated. A RINT transformation partially corrects this inflation and fitting 100 PCs per bin fully captures the environmental effect.



Supplementary Figure 22: Variance components for a sharp environmental effect of 2, 1 or 0.5 standard deviations added the samples from the Rotherham administrative region. Phenotype was generated from different sets of causal variants, and either 20 or 100 PCs from LD pruned MAF bin were fitted (N=35,867). The magnitude of the bias induced by this effect is proportional to the effect added. Estimates corresponding to a sharp environmental effect of 0.5 standard deviations are similar to the expected estimates without inflation from rare variants.

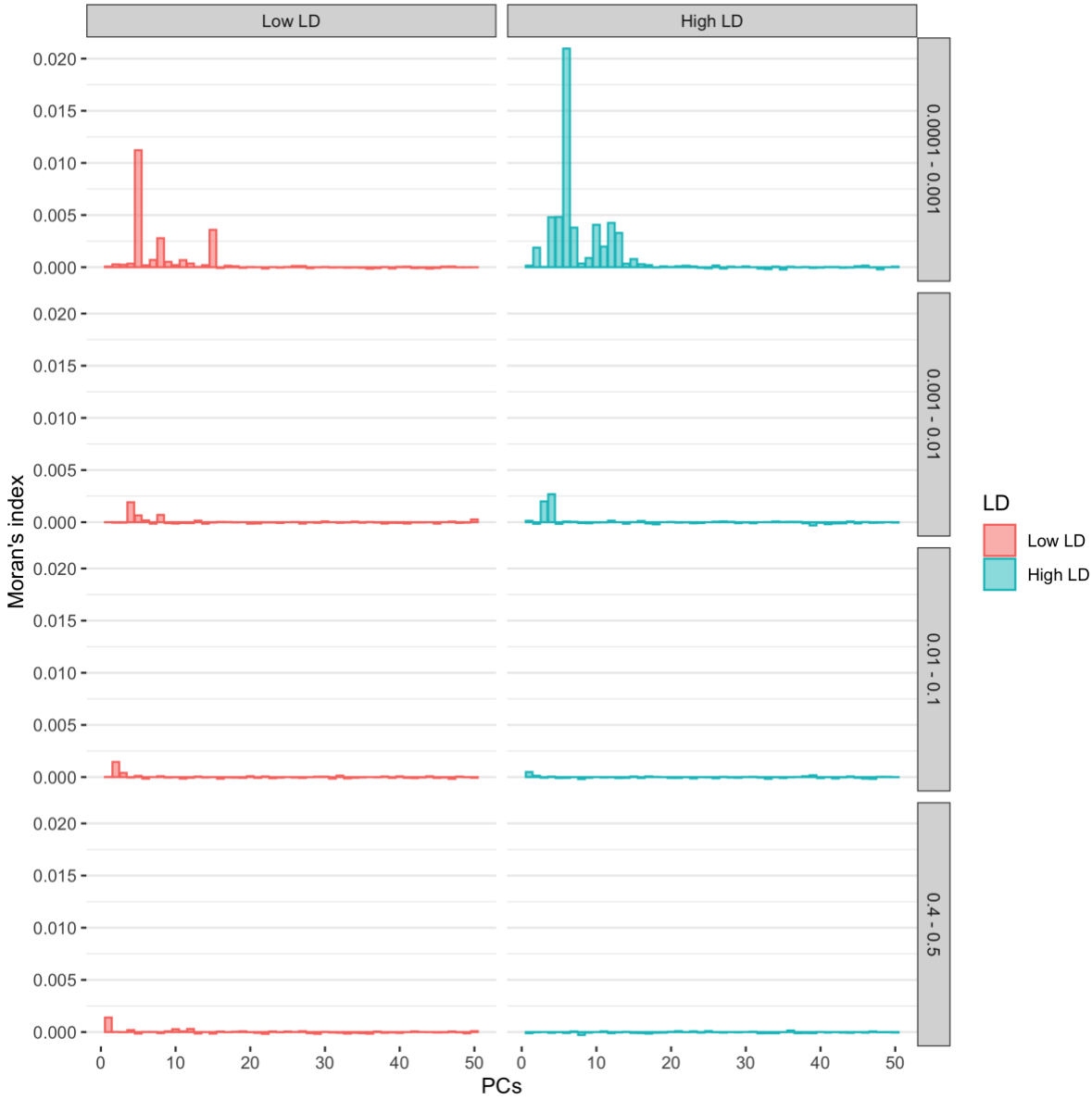


Supplementary Figure 23: Adjusted R^2 from fitting 150 PCs computed from LD-pruned SNPs for different MAF and LD bins on odd and even chromosomes GRMs. The mean R^2 from fitting a PC from one set on all the other set PCs is shown. As we do not expect Inter-chromosomal correlations under random mating, observing a $R^2 > 0$ between sets of chromosomes most likely show some population stratification. (A) Mean R^2 for the N=35,867 UKB Exome samples. (B) Mean $R^2$for the N=25,465 TOPMed samples. Vertical lines show the threshold determining the best number of PCs to fit to account for population stratification. The thresholds for each MAF/LD bin was computed from a segmented regression model. From this analysis, the number of PCs to fit for each bin to correct for population stratification is respectively 15, 7, 6, 5, 5, 5, 5, 0 for a total of 48 PCs.

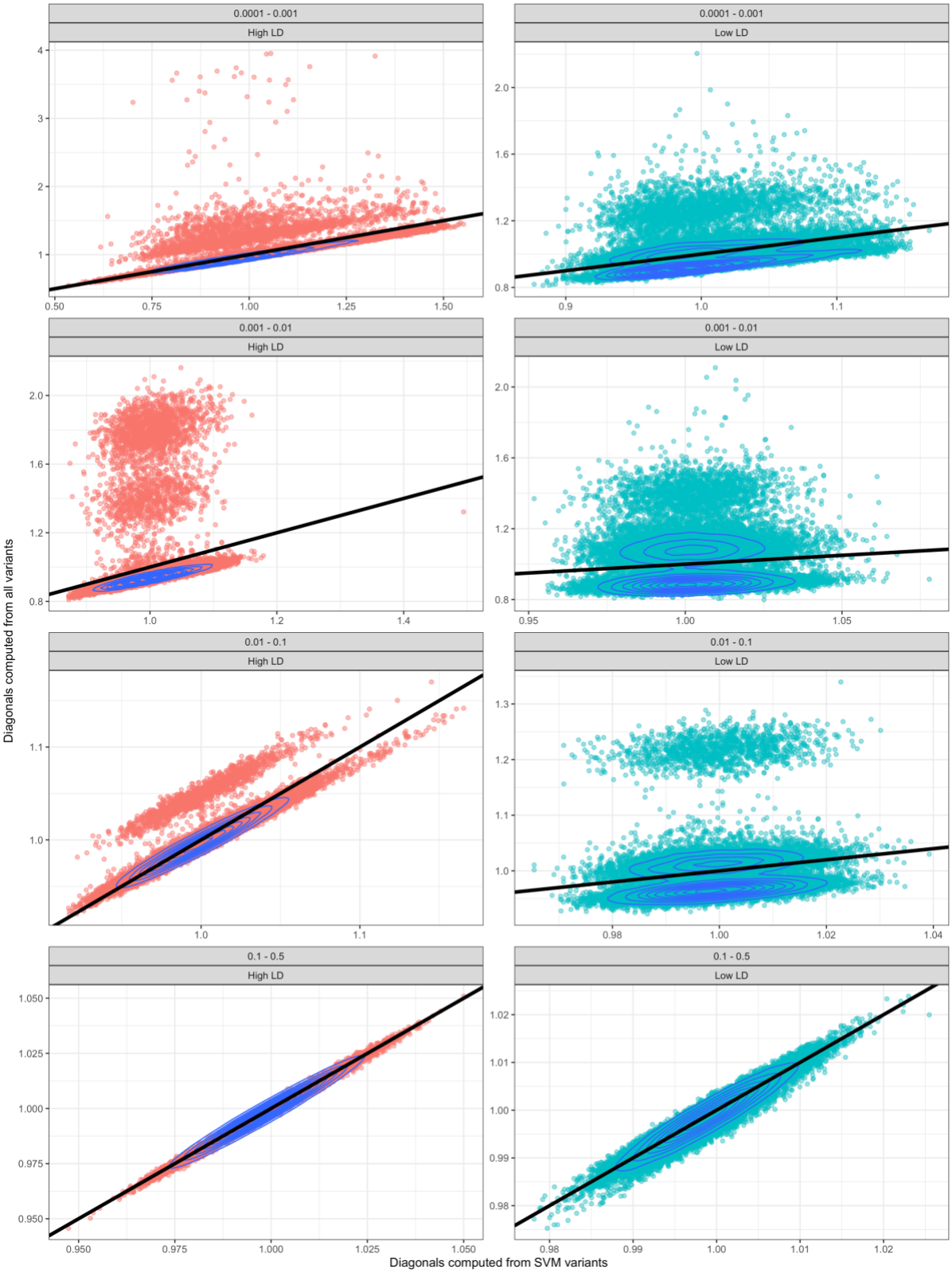


Supplementary Figure 24: Product of $PC_{i=1,100} = PC^{i}_{Odd} * PC^{i}_{Even}$. for each PC for each sample, further transformed by centering and applying a RINT transformation to smoothen the effect of outliers. For each individual (N=35,867), this product of odd-even chromosomes PCs was plotted according to their East and North birth coordinates.(A) PC5 on rare variants (0.0001 < MAF < 0.001) in low LD. (B) PC100 on rare variants (low LD). (C) PC5 on common variants (0.4 <

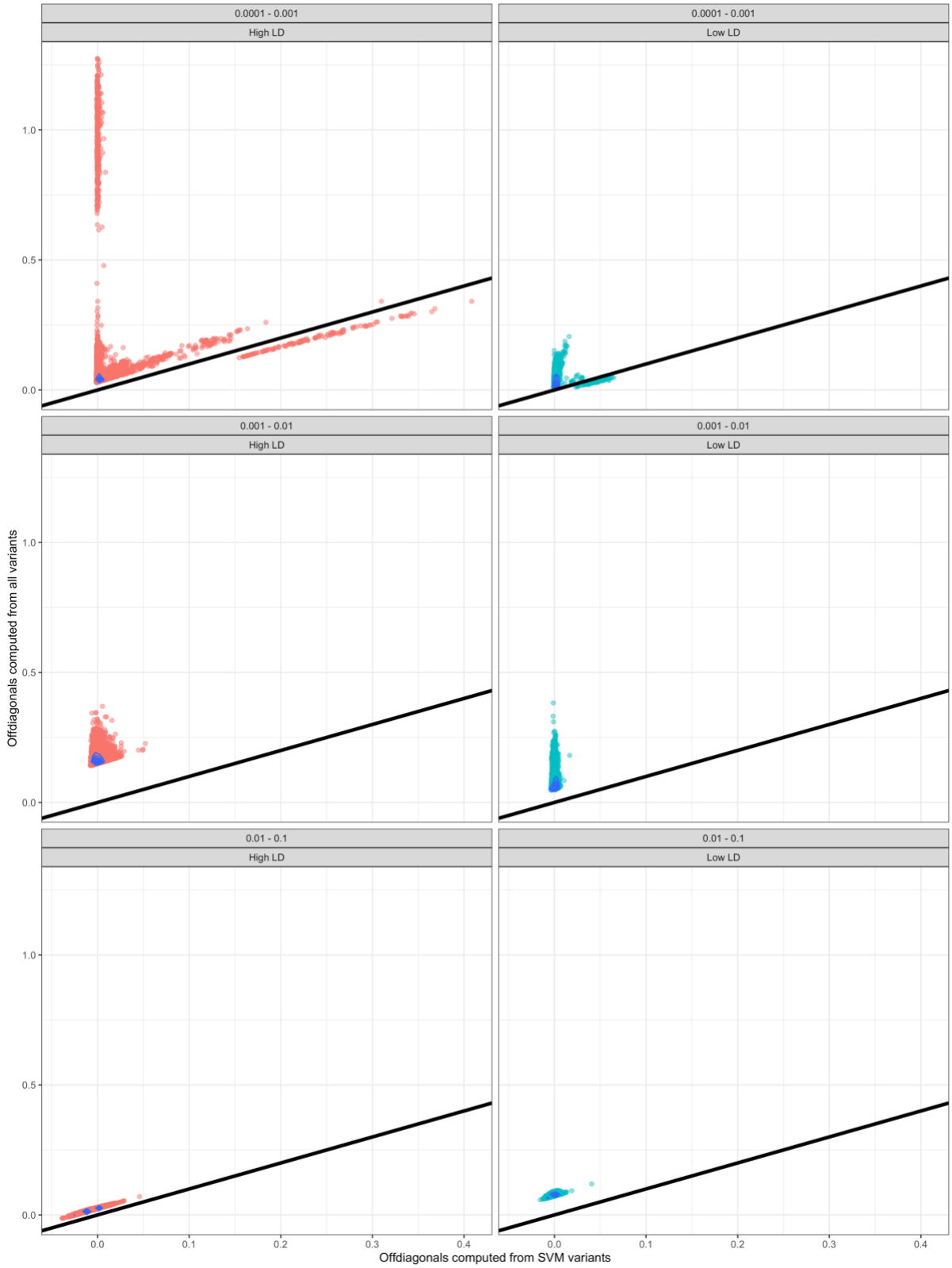
MAF<0.5) . (D) PC100 on common variants. We can see a larger stratification for (A) potentially indicating a north-south stratification.



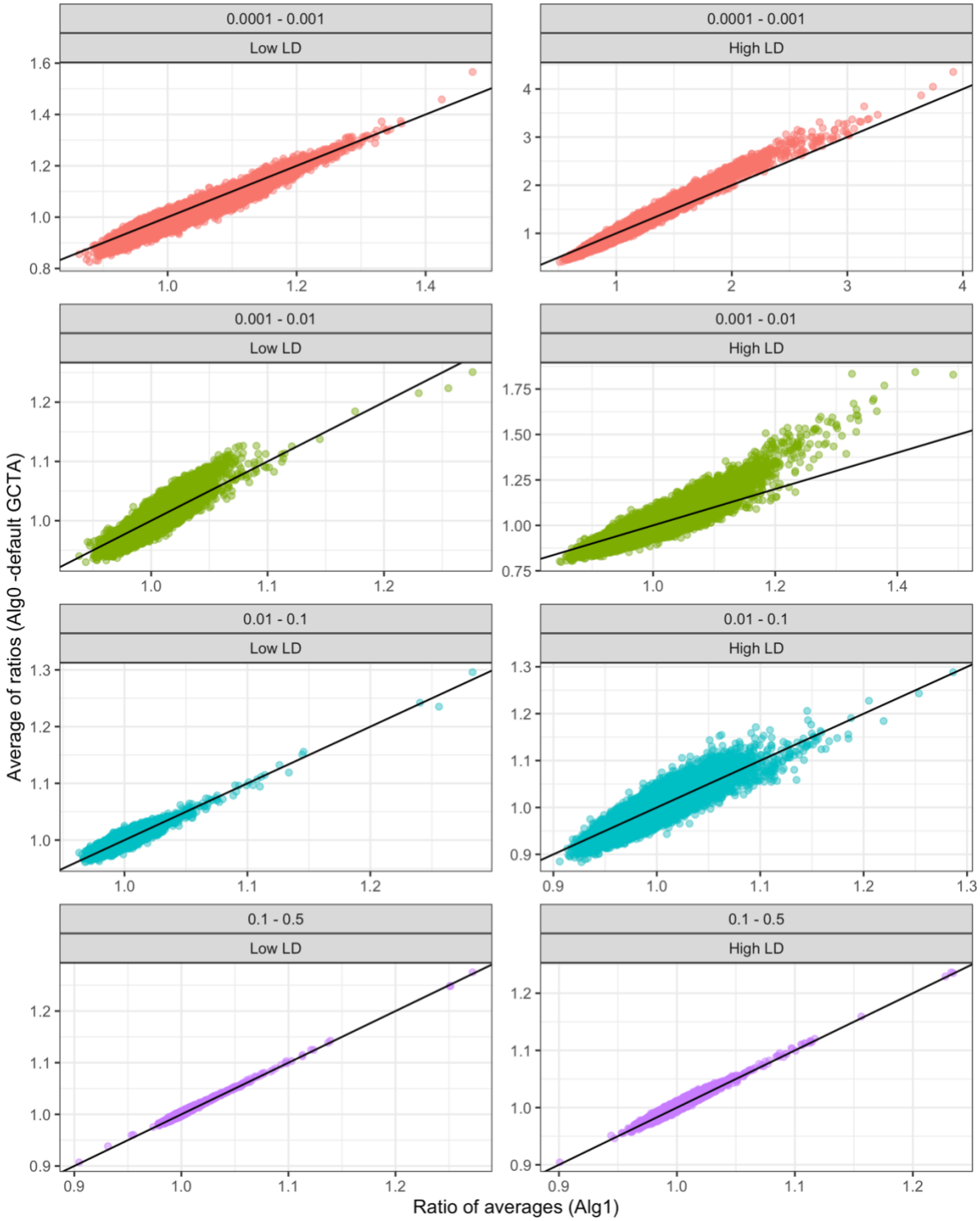
Supplementary Figure 25: Moran's I auto-correlation index calculated of UKB-WES samples (N=35,867). This index was calculated across all the samples for each of the first 50s on 3 rare MAF bins and one common MAF bin partitioned according to SNP-based LD. The product of $PC_{i=1,50} = PC^{i}_{Odd} * PC^{i}_{Even}$ was used to compute Moran's I index. We can observe a larger spatial correlation for the first PCs of the rare variants bins.



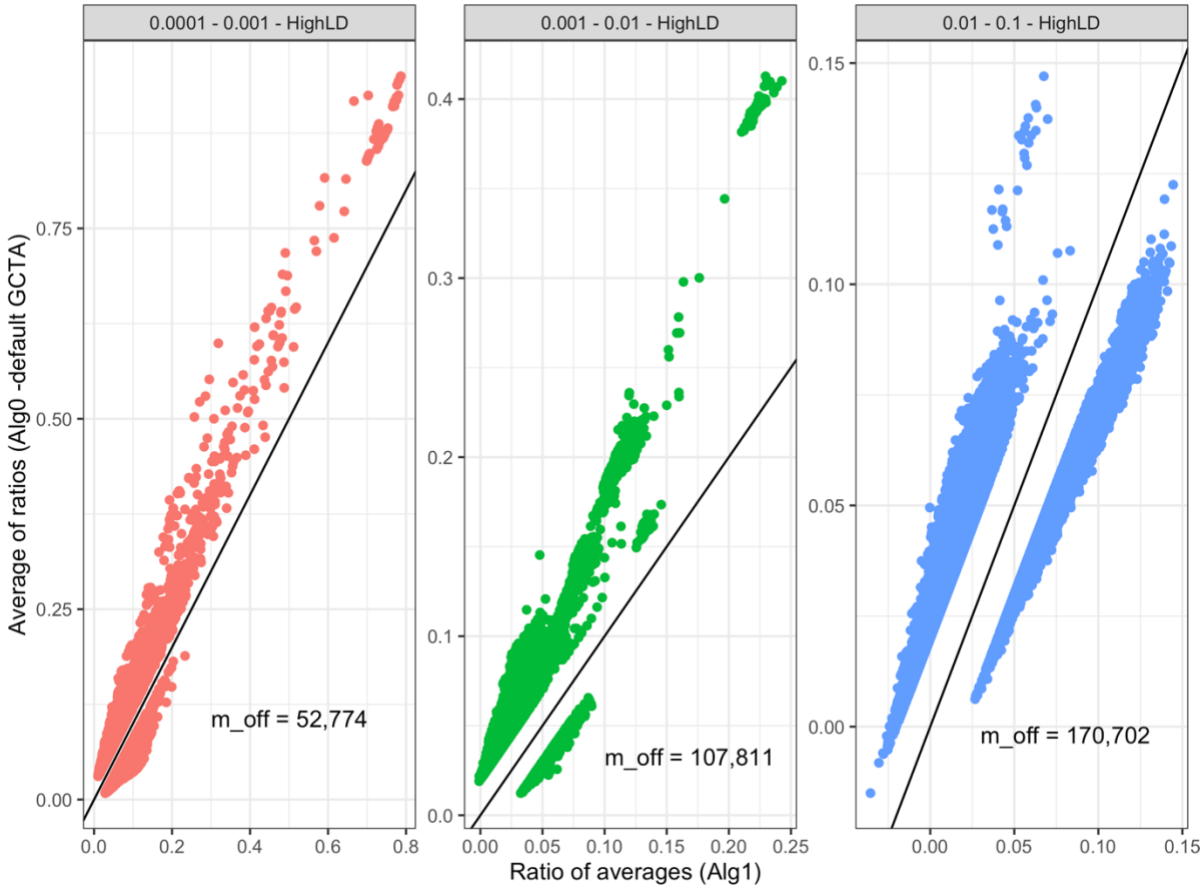
Supplementary Figure 26: GRM diagonals comparison between GRM computed from the high quality variants (SVM) or all genotyped variants. GRMs were computed using the ratio of average method. There are N=25,465 TOPMed samples per bin.



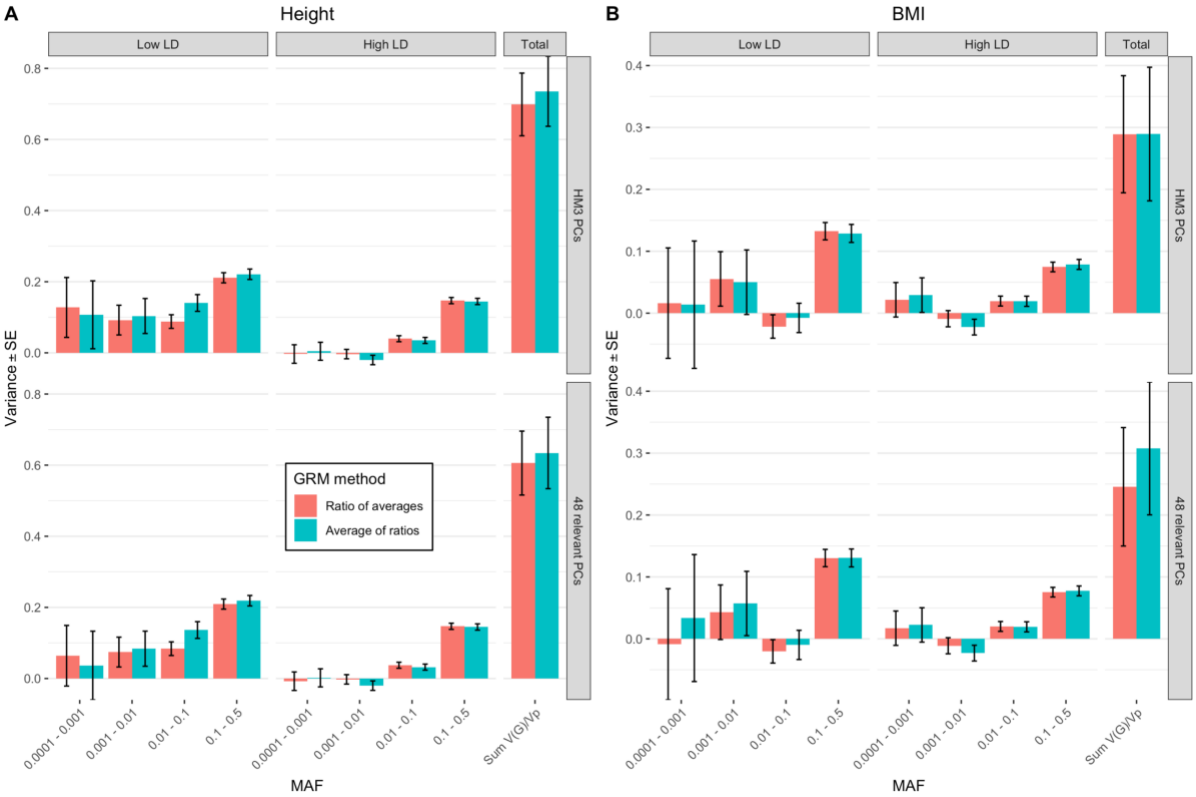
Supplementary Figure 27: GRM off-diagonals comparison between GRM computed from the high quality variants (SVM) or all genotyped variants from N=25,465 samples. GRM were computed using the ratio of average method. Only the largest 20,000 differences are shown for each bin with a minimum threshold of 1%. Common variants bins (0.1 < MAF < 0.5) did not show differences between the two methods larger than 1%.



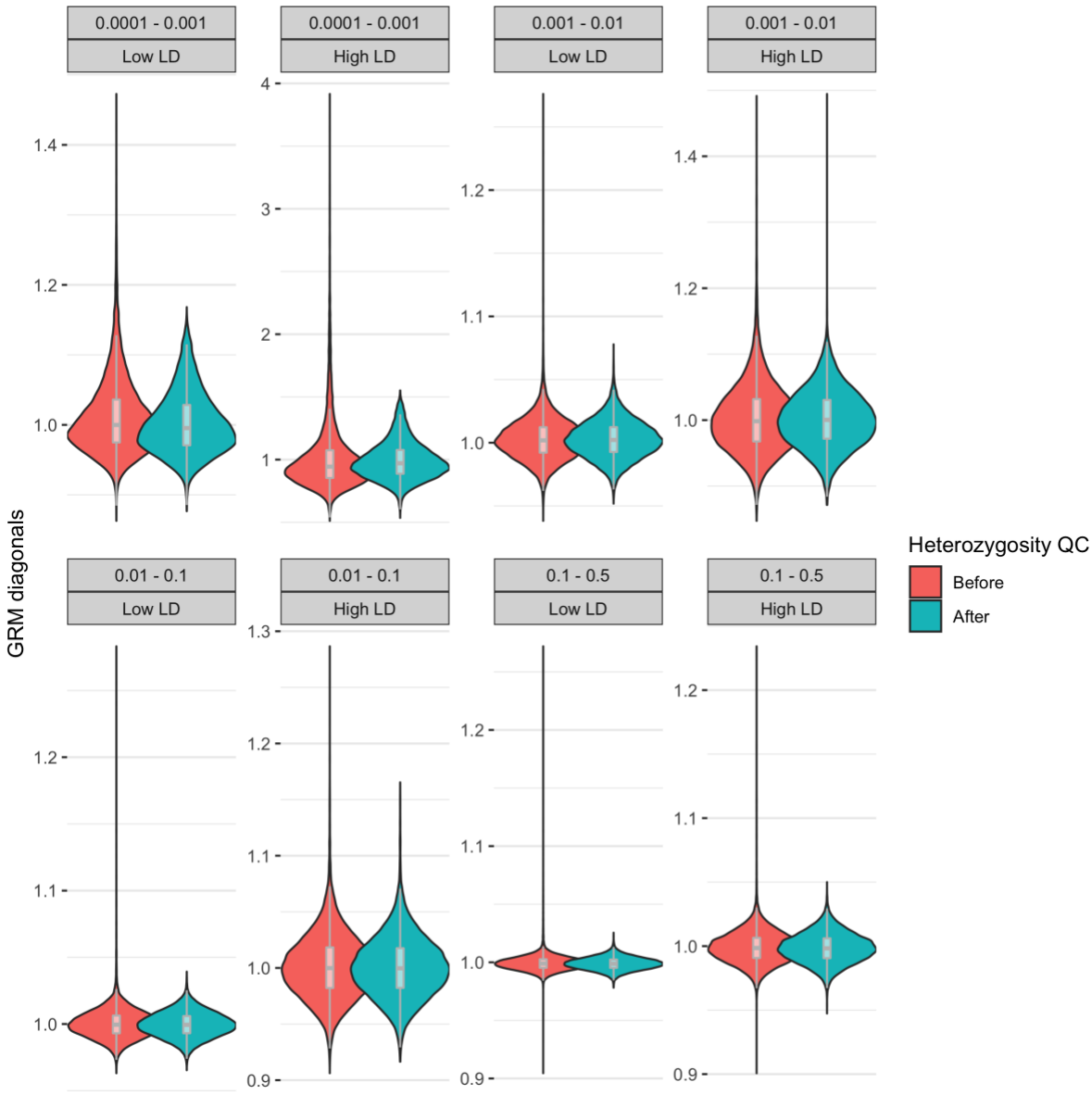
Supplementary Figure 28: Difference in diagonal elements for the ratio of averages and average of ratios GRM estimators. Diagonal for each MAF/LD bins are plotted using N=28,755 unrelated samples of European ancestry.



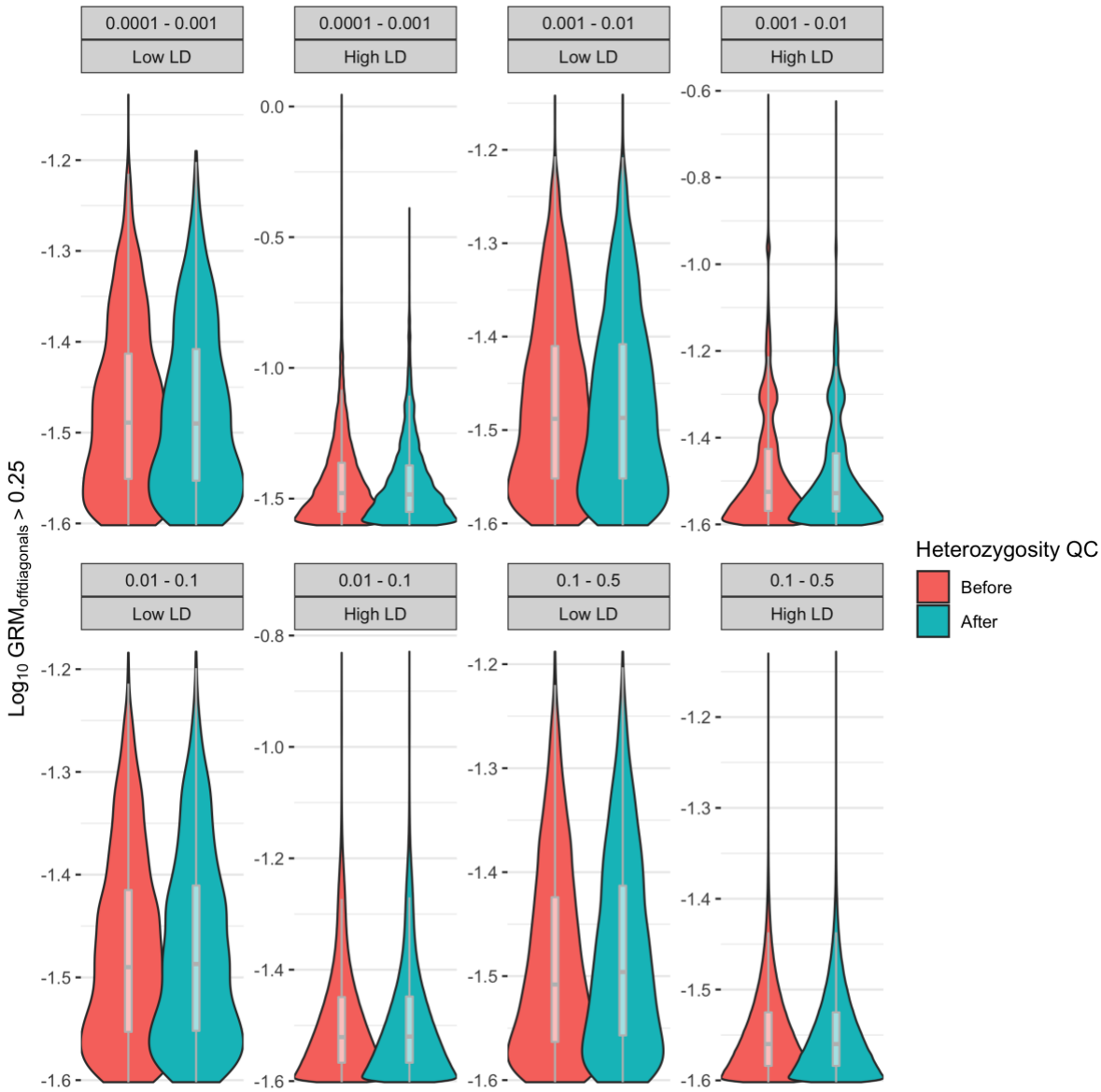
Supplementary Figure 29: Impact of the GRM estimator on off-diagonal elements using N=28,755 unrelated samples of European ancestry, only pairs with a relatedness difference larger than 2% are shown. Total number of diverging pairs is indicated within each window. The total number of pairs in the GRM ~413M. Low LD bins and common ones (0.1 < MAF < 0.5) did not present any difference larger than 2%.



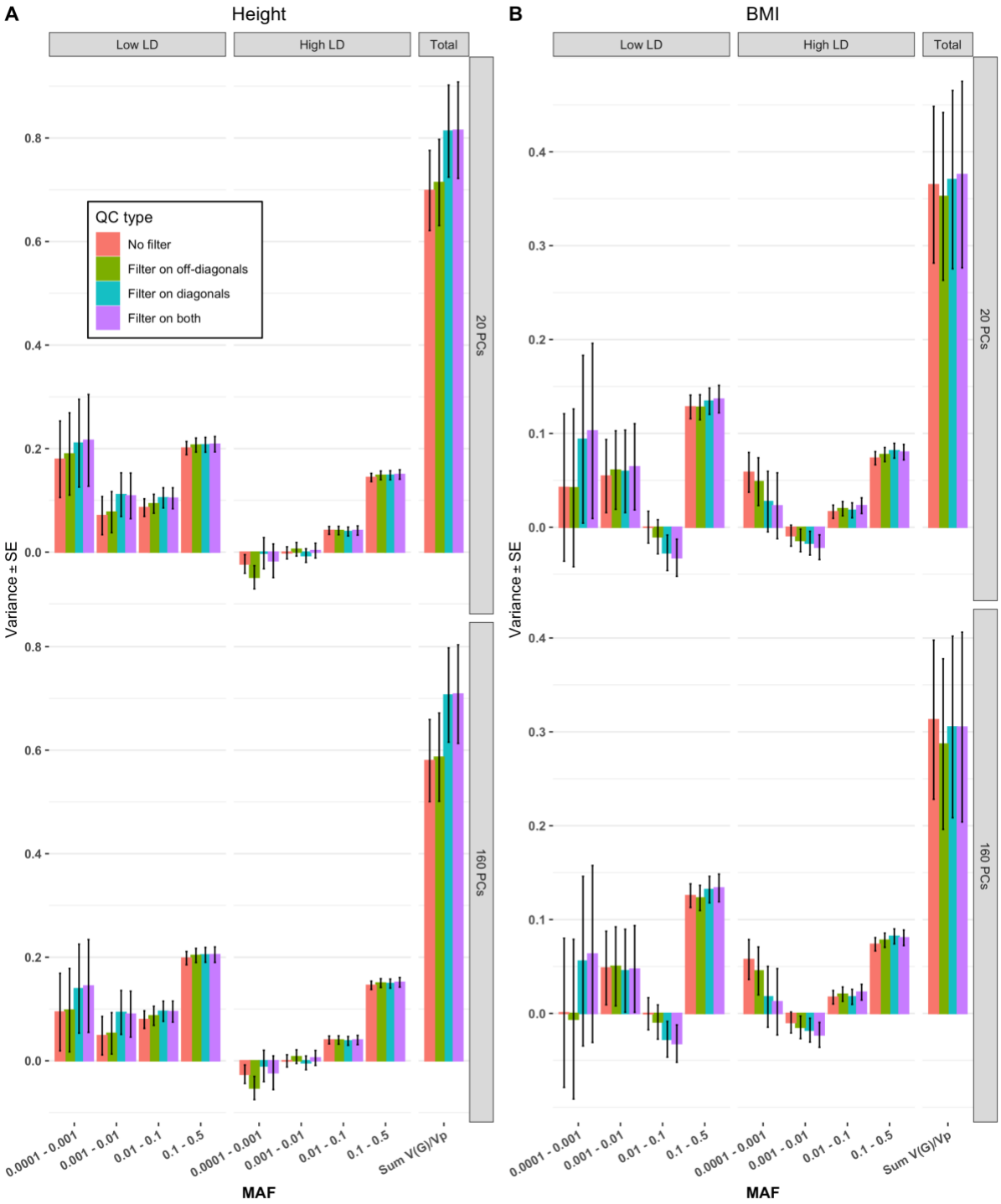
Supplementary Figure 30: GREML-LDMS using WGS (~33.7M variants, N=25,465) estimates using either the average over loci of ratios (blue) or the ratio over averages of loci (red) methods to calculate the GRM values. (A) Estimates for height fitting 8 bins: 0.70 - 0.74 (SE 0.09 - 0.10) fitting 20 HM3 PCs and 0.61 – 0.63 (SE 0.09 – 0.10) fitting 48 PCs. (B) Estimates for BMI fitting 8 bins: 0.29 - 0.29 (SE 0.09 - 0.11) fitting 20 HM3 PCs and 0.25 – 0.31 (SE 0.10 – 0.11) fitting 48 PCs.



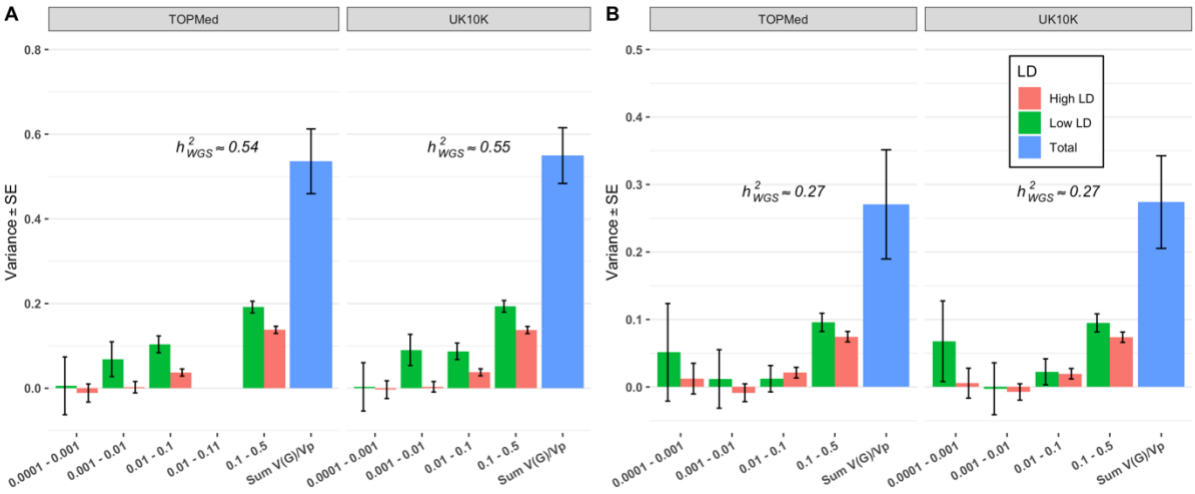
*Supplementary Figure 31: Distribution of diagonals elements for each MAF/LD GRM before and after filtering on sample bin heterozygosity (from N=28,755 (top) to N=25,465 (bottom) samples) Lower and upper boxplot hinges correspond to the first and third quartiles and whiskers extend to +/- 1.5 * inter-quartile range.*



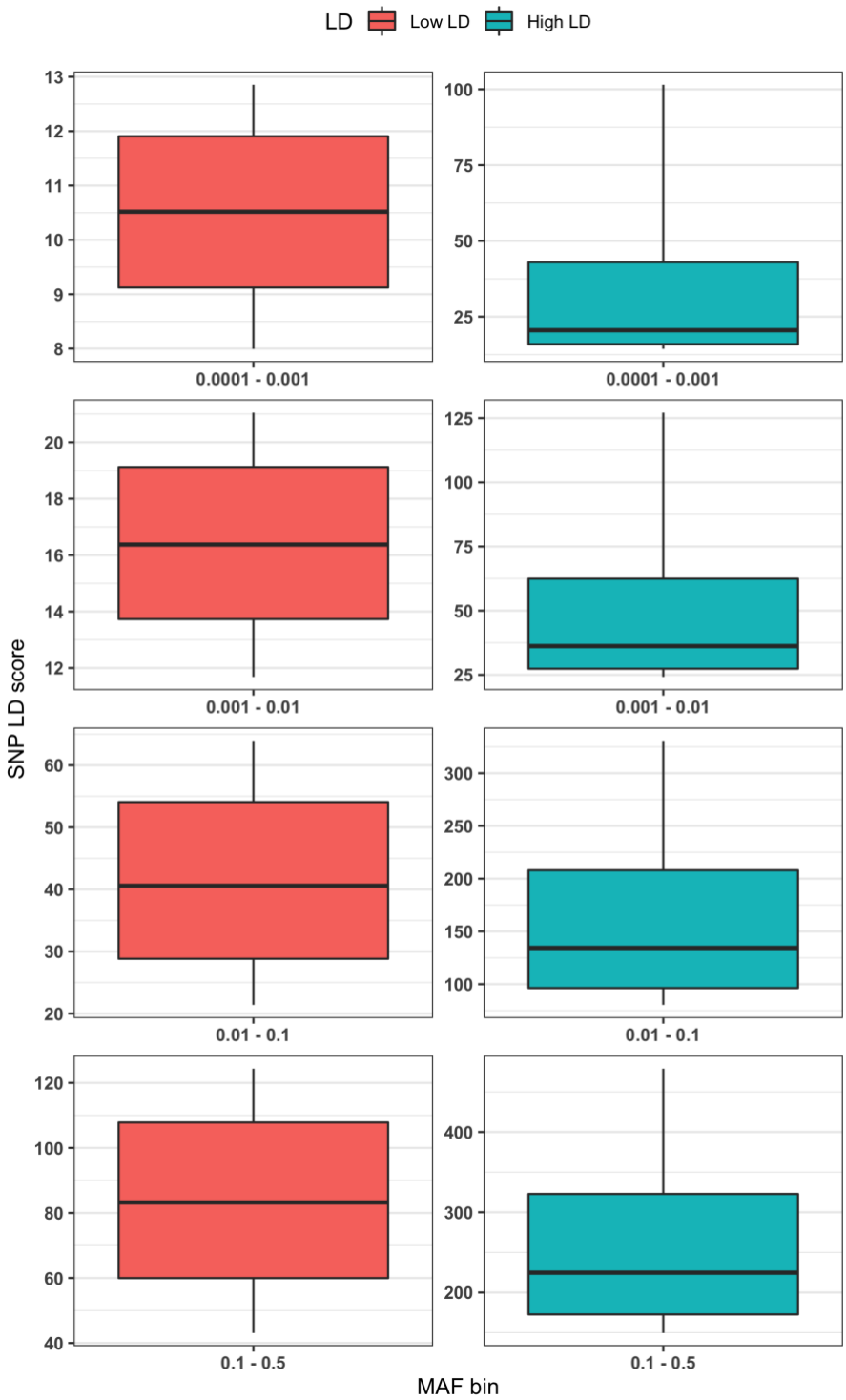
*Supplementary Figure 32: Distribution of off-diagonals elements for each MAF/LD GRM before and after filtering on sample bin heterozygosity (from N=28,755 (top) to N=25,465 (bottom) samples). Lower and upper boxplot hinges correspond to the first and third quartiles and whiskers extend to +/- 1.5 * inter-quartile range.*



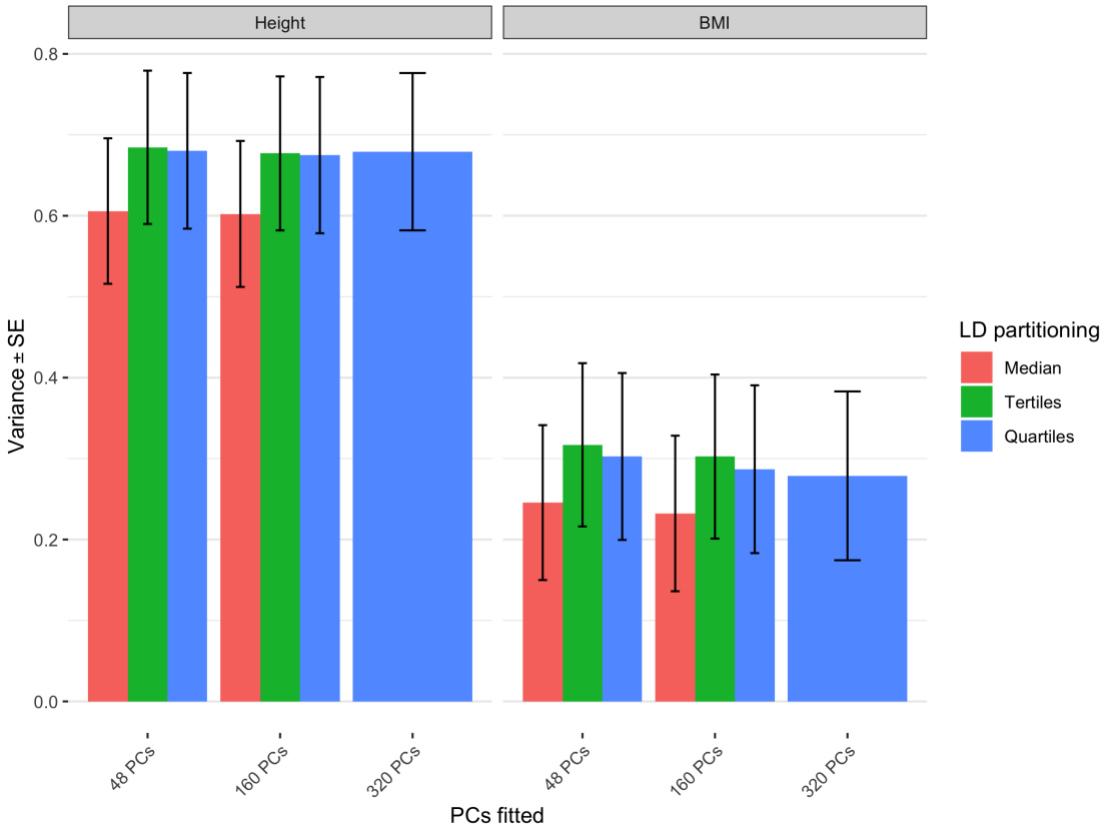
Supplementary Figure 33: GREML-LDMS estimates by removing samples from the full GRM of Europeans (before QC on sample heterozygosity, N=28,755). Samples were removed either based on their off-diagonal values (>0.1) across all GRMs (all residual relatedness coming from the rare variants GRMs) or their diagonal values (< 0.7 or > 1.3) or both (removing samples based on the diagonals first then off-diagonals). (A) GREML-LDMS estimates for height fitting either 20 HM3 PCs or 160 PCs computed from all WGS bins. Estimates without any filtering 0.58 – 0.70 (SE 0.08), estimates when filtering on off-diagonals 0.59 – 0.71 (SE 0.08 – 0.09), estimates when filtering on diagonals 0.71 – 0.81 (SE 0.09), estimates when filtering on both 0.71 – 0.82 (SE 0.09 – 0.10). (B) GREML-LDMS estimates for BMI without filtering 0.31 – 0.36 (SE 0.08), estimates when filtering on off-diagonals 0.29 – 0.35 (SE 0.09), estimates when filtering on diagonals 0.31 – 0.37 (SE 0.09 – 0.10), estimates when filtering on both 0.31 – 0.38 (SE 0.10).



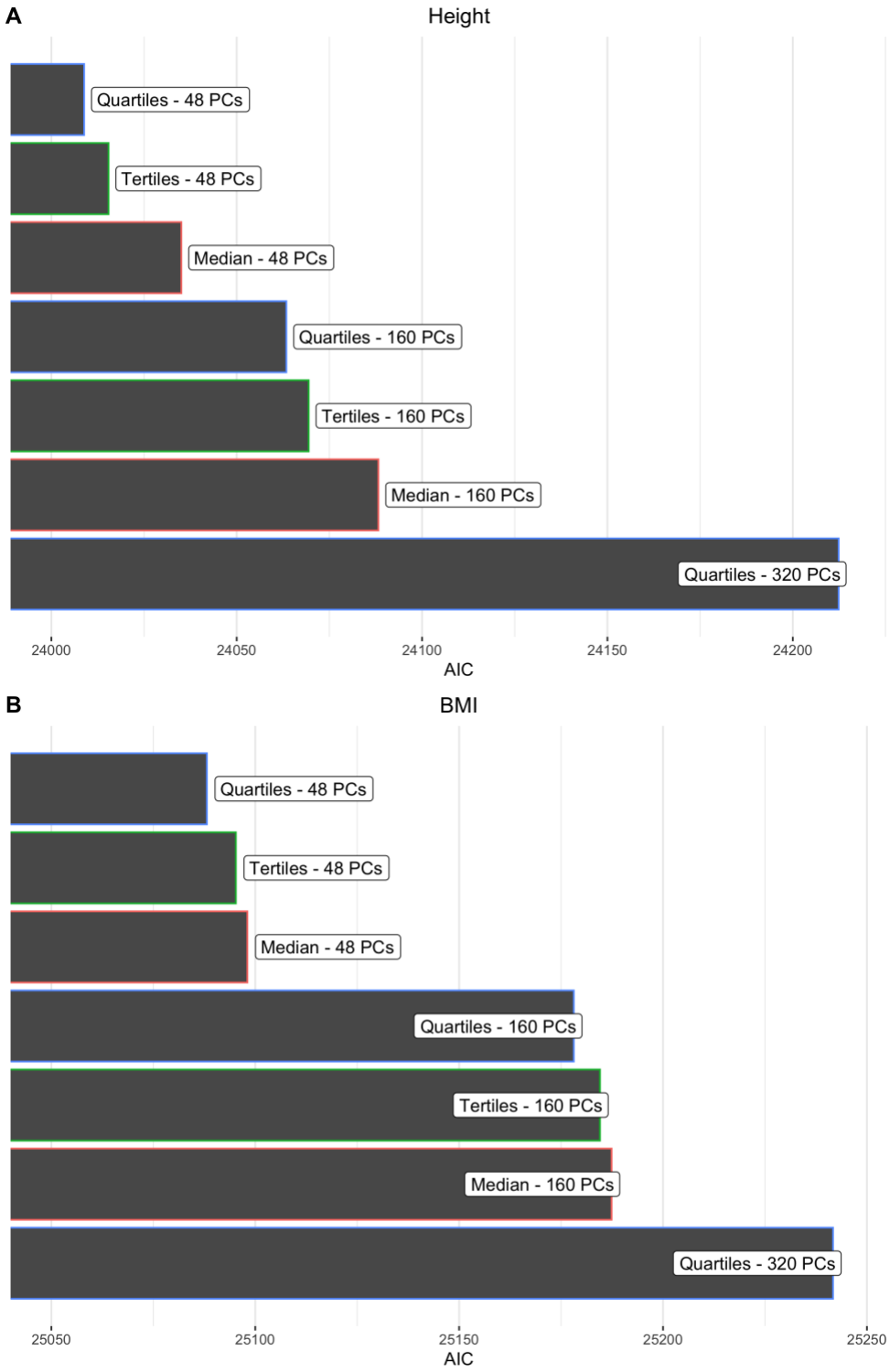
Supplementary Figure 34: GREML-LDMS estimates from variants in common (~20.0M variants, N=25,465) between TOPMed and UK10K datasets stratified in 8 bins according to variants MAF and LD properties using either TOPMed or UK10K LD and MAF reference, corrected for 160 PCs from WGS independent variants. (A) Estimates of h^2_{WGS} for height (~0.54 – 0.55 (SE 0.07 – 0.08)) or (B) BMI (~0.27 (SE 0.07 – 0.08)) are similar and independent of the LD and MAF reference. The number of variants in each of the 7 MAF bins (twice the number in each LD bin) is, from the lowest to highest MAF bins, 9.1.M, 4.5M, 3.3M and 3.1M, respectively.



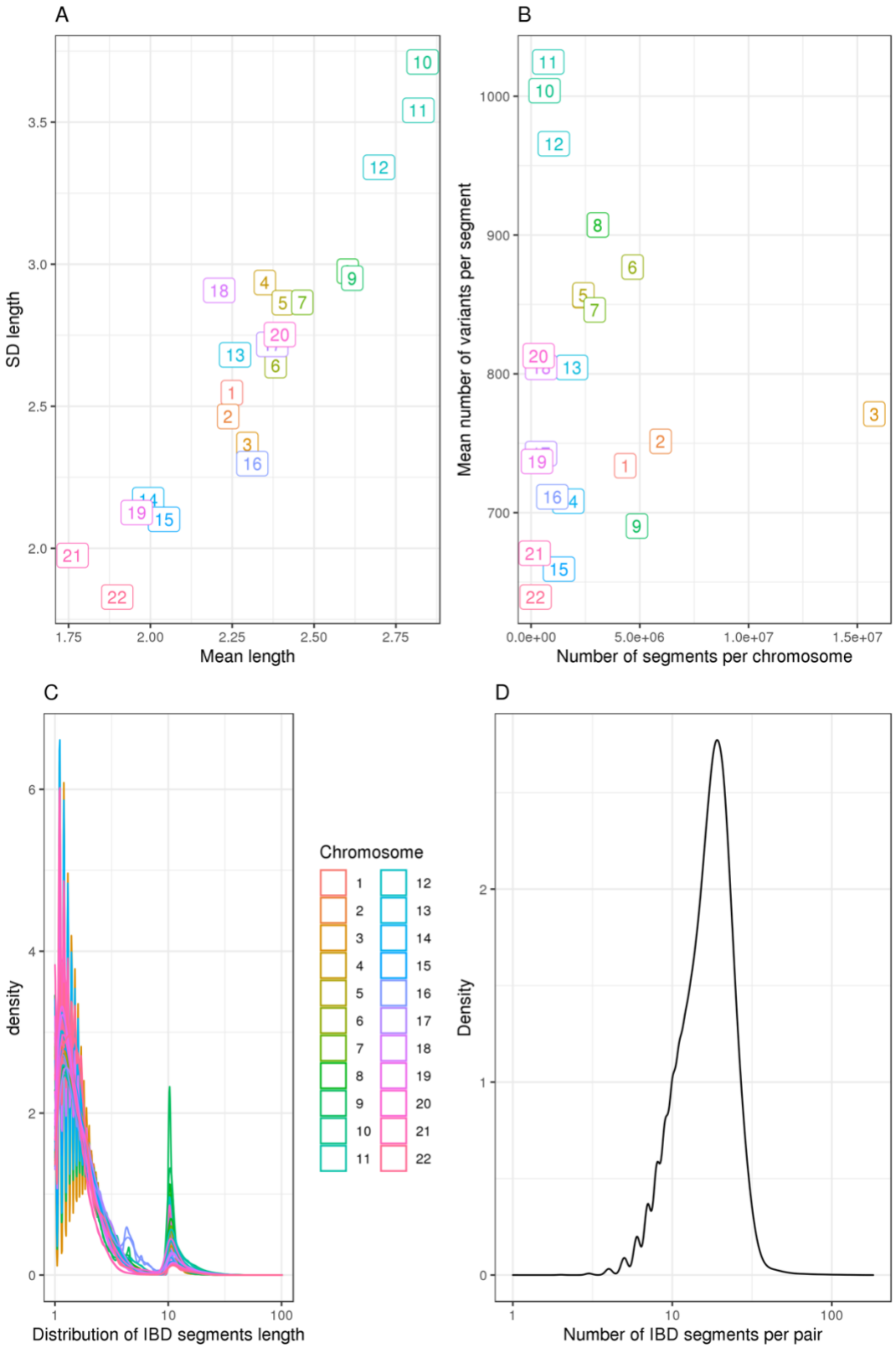
*Supplementary Figure 35: Distribution of SNP LD score for each MAF/LD bin for N=25,465 samples. For more clarity, boxplot outliers are not shown here. Most of the extreme outliers are from rare SNPs in very high LD bins. Lower and upper hinges correspond to the first and third quartiles and whiskers extend to +/- 1.5 * inter-quartile range.*



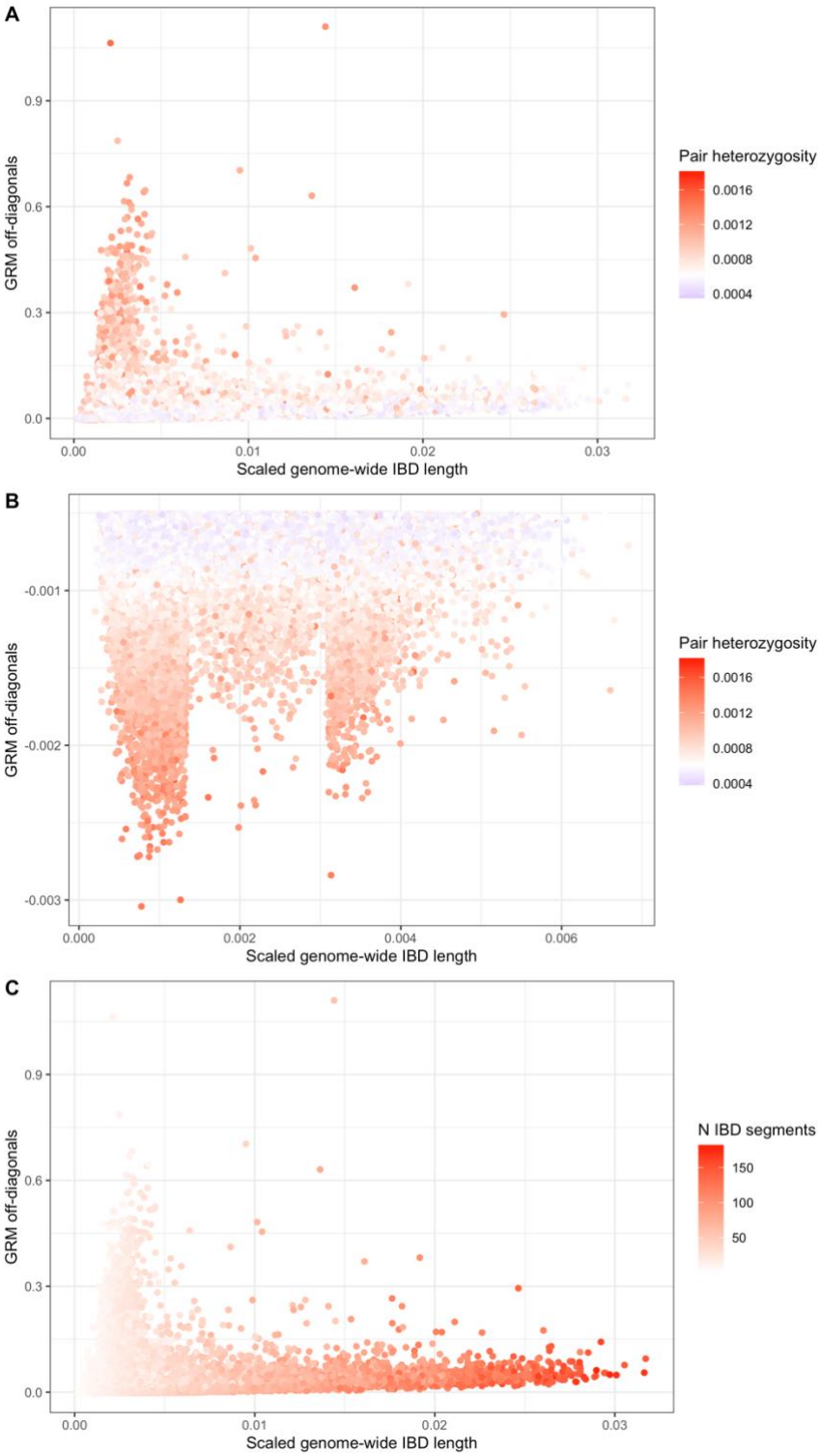
Supplementary Figure 36: Summary of the impact of different LD grouping strategies on total heritability estimates for N=25,465 samples using 2, 3 or 4 LD grouping for each MAF bin correcting by 48/160/320 PCs computed from WGS independent SNPs. Estimates for height 0.60 – 0.68 (SE 0.09 – 0.10), estimates for BMI 0.23 – 0.32 (SE 0.10).



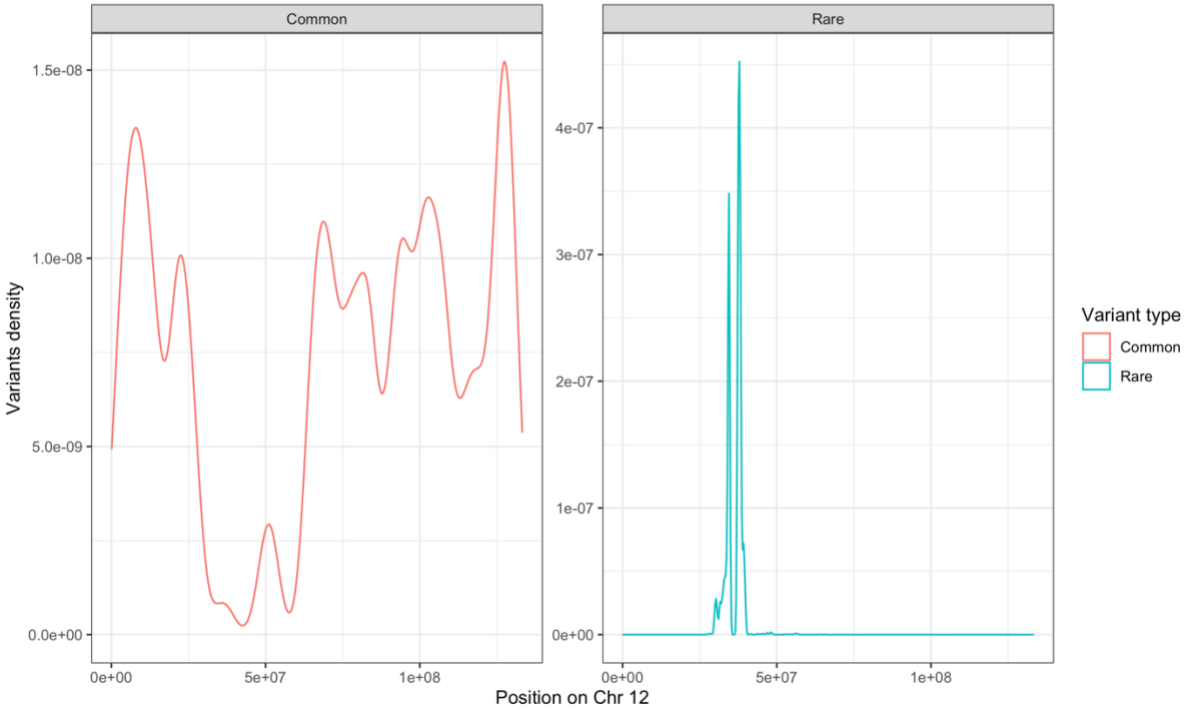
Supplementary Figure 37: AIC fitting different models, with either 2, 3 or 4 LD bins and 40, 160 or 320 PCs for height (A) and BMI (B), using N=25,465 samples. The lower AIC indicates a better fitting model.



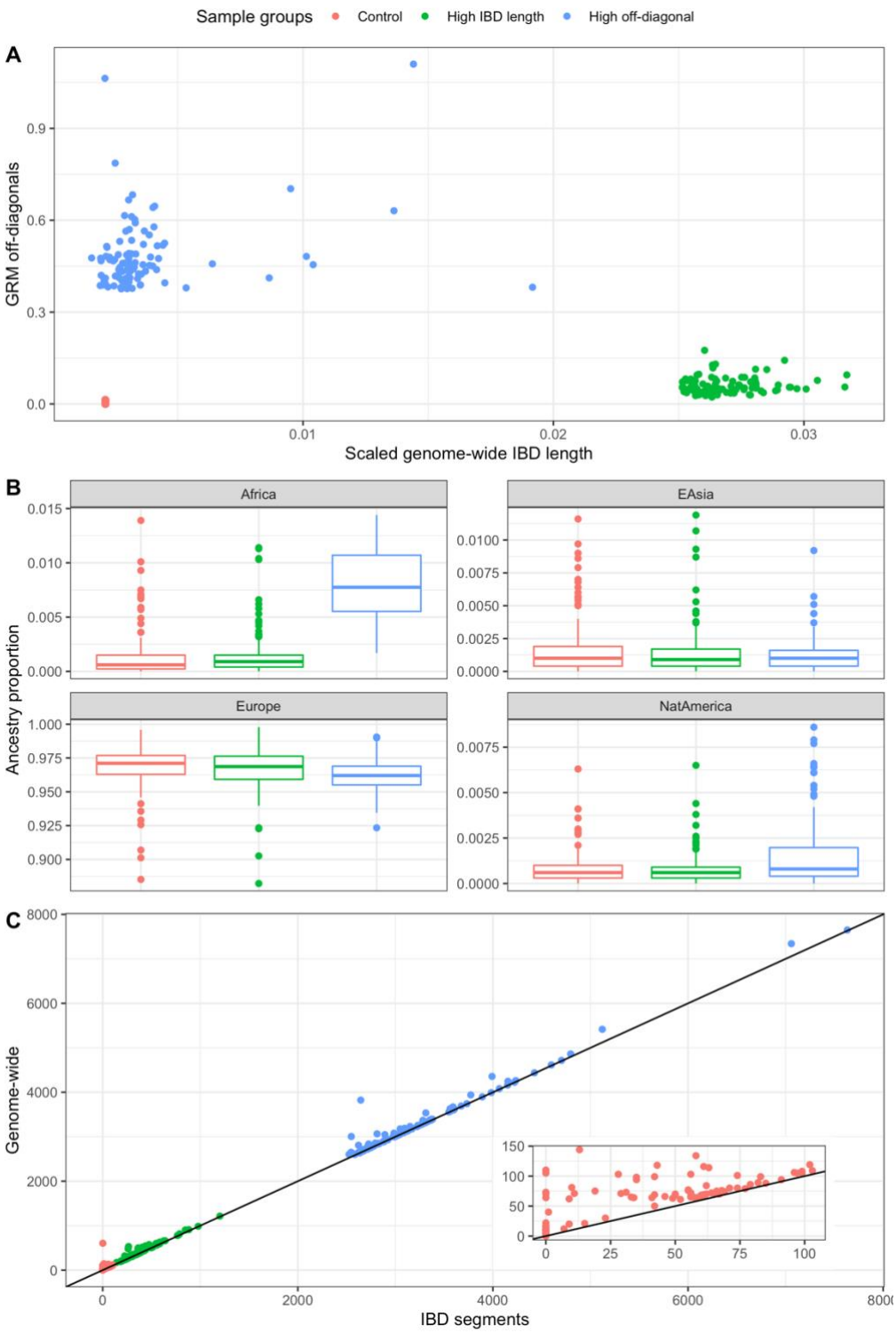
Supplementary Figure 38: Information on IBD segments distribution among N=28,755 unrelated Europeans. (A) Mean segment length per chromosome and length SD. (B) Number of segments and mean number of variants per segment. (C) Distribution (log10 scale) of the length of IBD segments per chromosome. (D) Distribution of the number of segments per unique sample pair sharing a segment IBD (log10 scale).



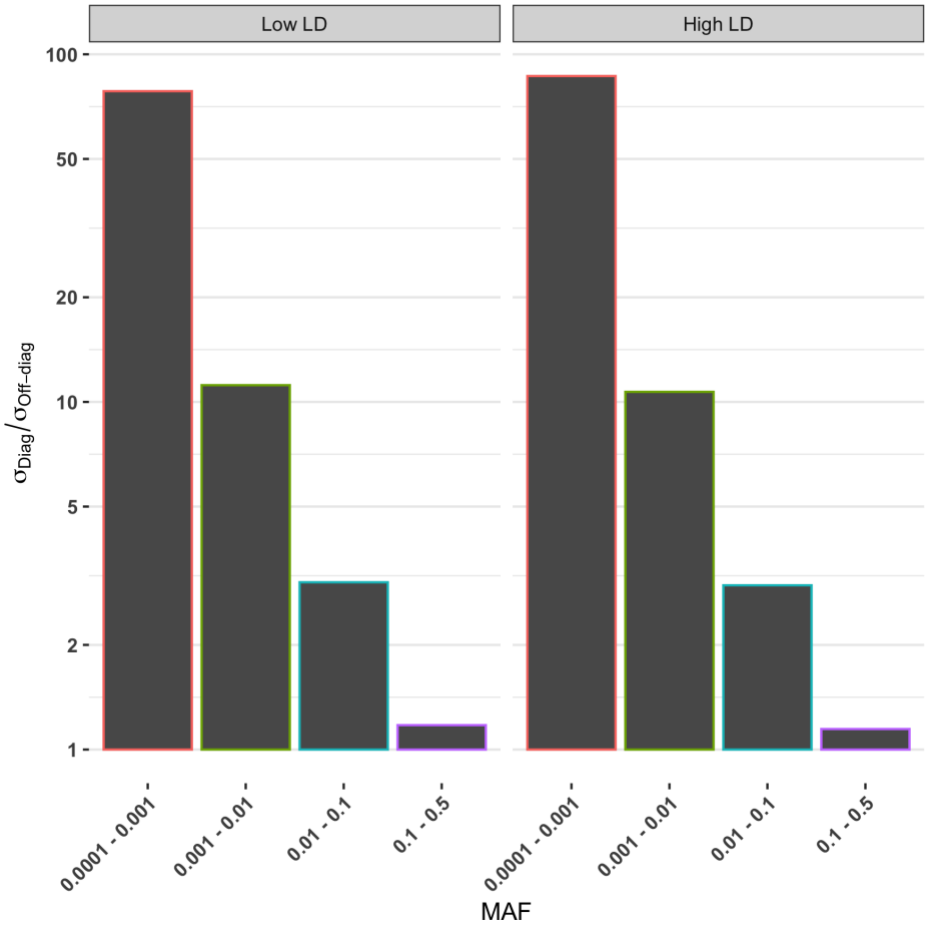
Supplementary Figure 39: Investigation on the relationship between genome-wide length shared IBD and GRM off-diagonal elements computed from N=28,755 for variants 0.0001 < MAF < 0.001 in high LD. (A) Each sample pair is coloured by their bin specific mean individual heterozygosity rate. (B) Similar to (A), but showing only pairs with low relatedness value. (C) Each sample pair is coloured by the number segments shared IBD.



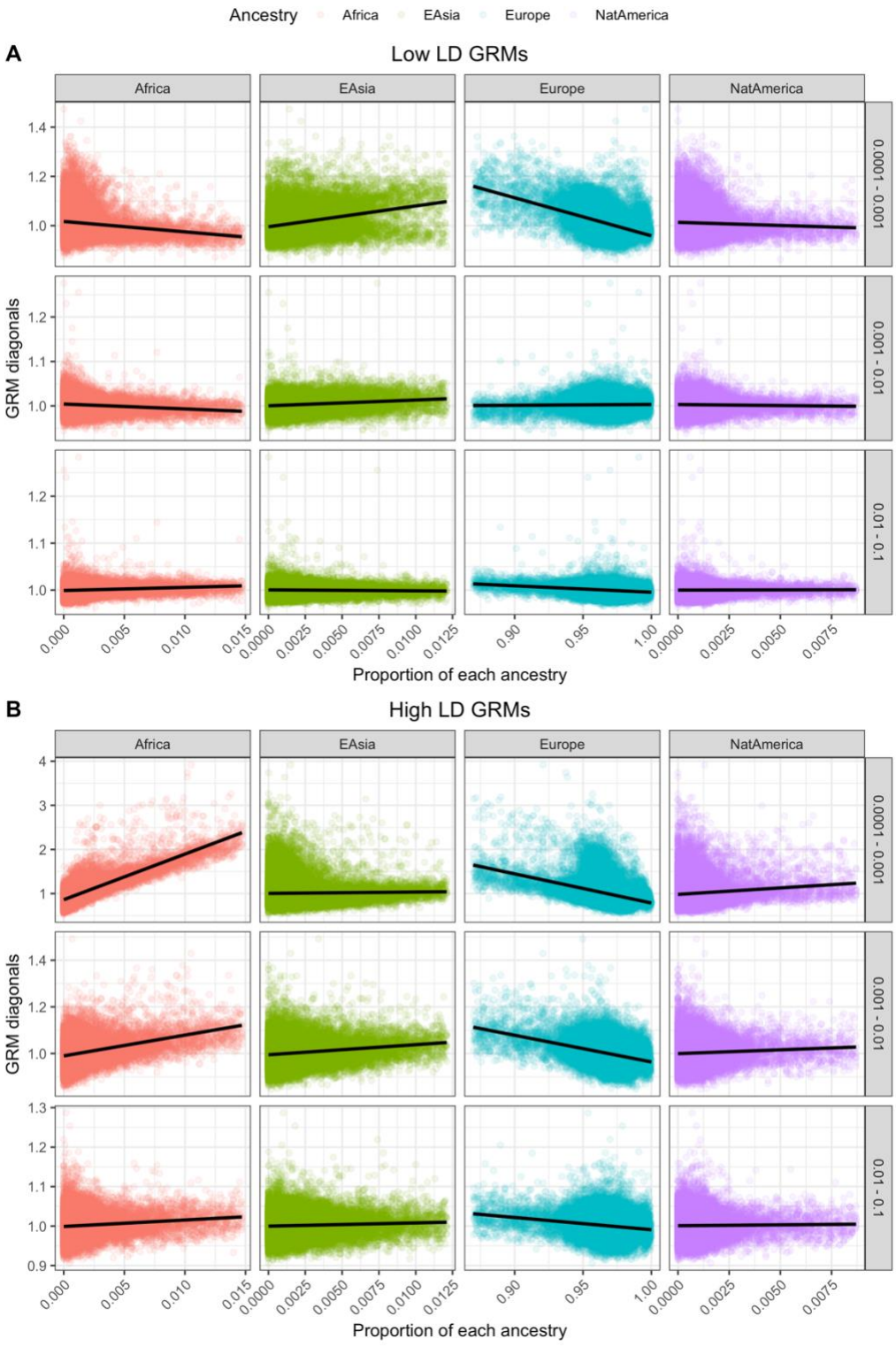
Supplementary Figure 40: Variant density for a pair of samples showing extreme relatedness in the off-diagonal of the GRM from variants of 0.0001 < MAF < 0.001 in high LD. Common variants (0.1 < MAF < 0.5) distribution on chromosome 12 (left panel) filtering out pair-specific heterozygotes variants. Rare variants distribution (right panel) with a MAC 2 threshold for the pair considered. Note that over 99% of shared heterozygotes rare variants were located on chromosome 12. The lack of oppositive homozygotes variants in common SNPs, at the same location of the pair specific rare heterozygotes variants suggests a region shared IBD.



*Supplementary Figure 41: Investigating the characteristics of extreme GRM pairs of N=28,755 samples. (A) Top 100 pairs with either the largest proportion of genome shared IBD (green), the largest off-diagonal values for rare variants in high LD (blue) or 100 pairs selected near the median (red). Most of the high off-diagonal values do not present large segments shared IBD. (B) Ancestry proportion of samples within each group. The highly related samples show an increase of genome-wide African ancestry compared to the two other groups. Lower and upper boxplot hinges correspond to the first and third quartiles and whiskers extend to +/- 1.5 * inter-quartile range. (C) Origin of the pair-specific heterozygotes variants considering only the very rare variants in high LD, either coming from the segments shared IBD by each pair or coming from the shared SNPs genome-wide. Embedded plot shows the control pairs only.*



Supplementary Figure 42: Ratio of GRM diagonal elements variance over GRM off-diagonal elements variance, for each MAF and LD bin using N=25,465 unrelated Europeans samples. GRMs were computed using the ratio of average method.



Supplementary Figure 43: Proportion ancestry and diagonal values for each MAF/LD GRM using N=28,755 unrelated Europeans. GRMs were computed using the ratio of averages method (Van Raden GRM estimator). (A) For GRMs from low LD variants. (B) For GRMs from high LD variants.