AJHG, Volume *98*

Supplemental Information

A Burden of Rare Variants Associated with Extremes of Gene Expression in Human Peripheral Blood

Jing Zhao, Idowu Akinsanmi, Dalia Arafat, T.J. Cradick, Ciaran M. Lee, Samridhi Banskota, Urko M. Marigorta, Gang Bao, and Greg Gibson

The American Journal of Human Genetics

Supplemental Data

A Burden of Rare Variants Associated with Extremes of Gene Expression in Human Peripheral Blood

Jing Zhao, Idowu Akinsanmi, Dalia Arafat, T.J. Cradick, Ciaran M. Lee, Samridhi Banskota, Urko M. Marigorta, Gang Bao, and Greg Gibson

Zhao *et al*, 2015. **Figure S1**

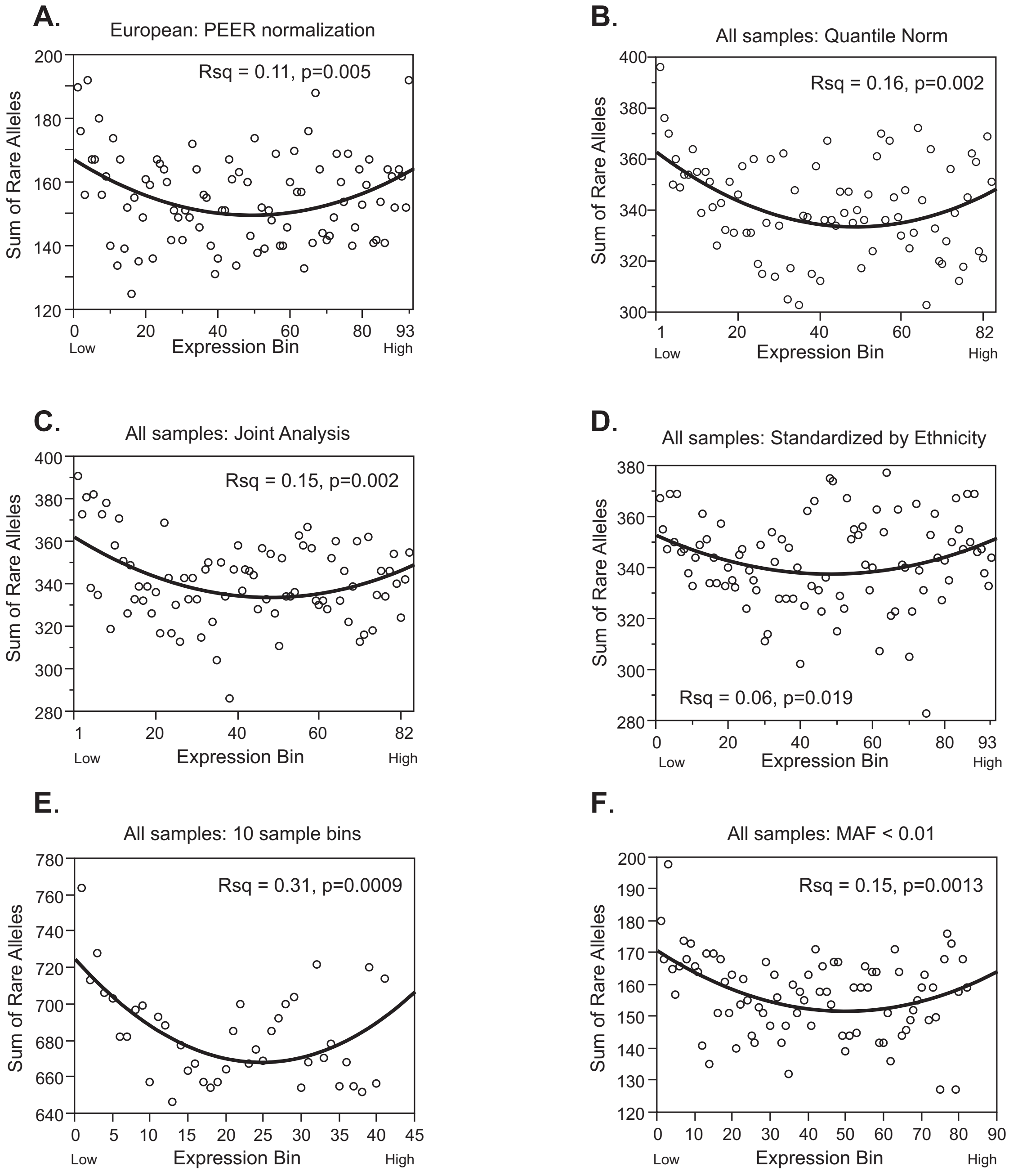

Figure S1. Quadratic regression plots with alternate normalization, bin size or MAF. As in Figure 1, each plot shows the quadratic regression fit of the sum of rare variants (A-E: MAF < 0.05; F, MAF < 0.01) in equal-sized expression bins from low to high (A-D,F: 5 individuals; E: 10 individuals) . Rsq and p-values refer to the full model R-squared and p-value from a quadratic regression. (A) European only analysis with PEER normalization; (B) Quantile of all samples, no adjustment; (C) Joint analysis of z-scores of all genes; (D) SNM normalization of all samples after standardization by ancestry; (E) as in Figure 1A but with bin size of 10; (F) All samples SNM but with MAF<0.01.

Zhao *et al*, 2015. **Figure S2**

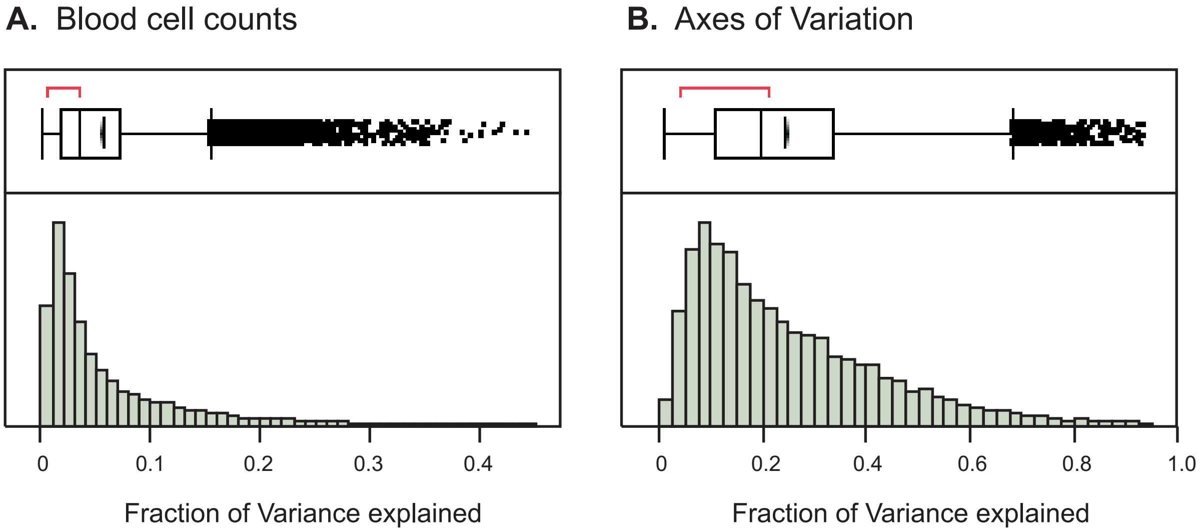

Figure S2. Variance explained by Blood Counts and Axes of Variation in CHDWB cohort. Histograms show the proportion of genes with the indicated fraction of variance explained by blood counts (lymphocytes, neutrophils, monocytes, red blood cells, platelets) and seven common Axes of variation respectively in (A) and (B), as the R-squared for the full model in a multiple regression. Boxes show the mean and inter-quartile range (25th to 75th percentile), and whiskers show 1.5X the respective IQR with outlier points shown as small squares.

Zhao *et al*, 2015. **Figure S3**

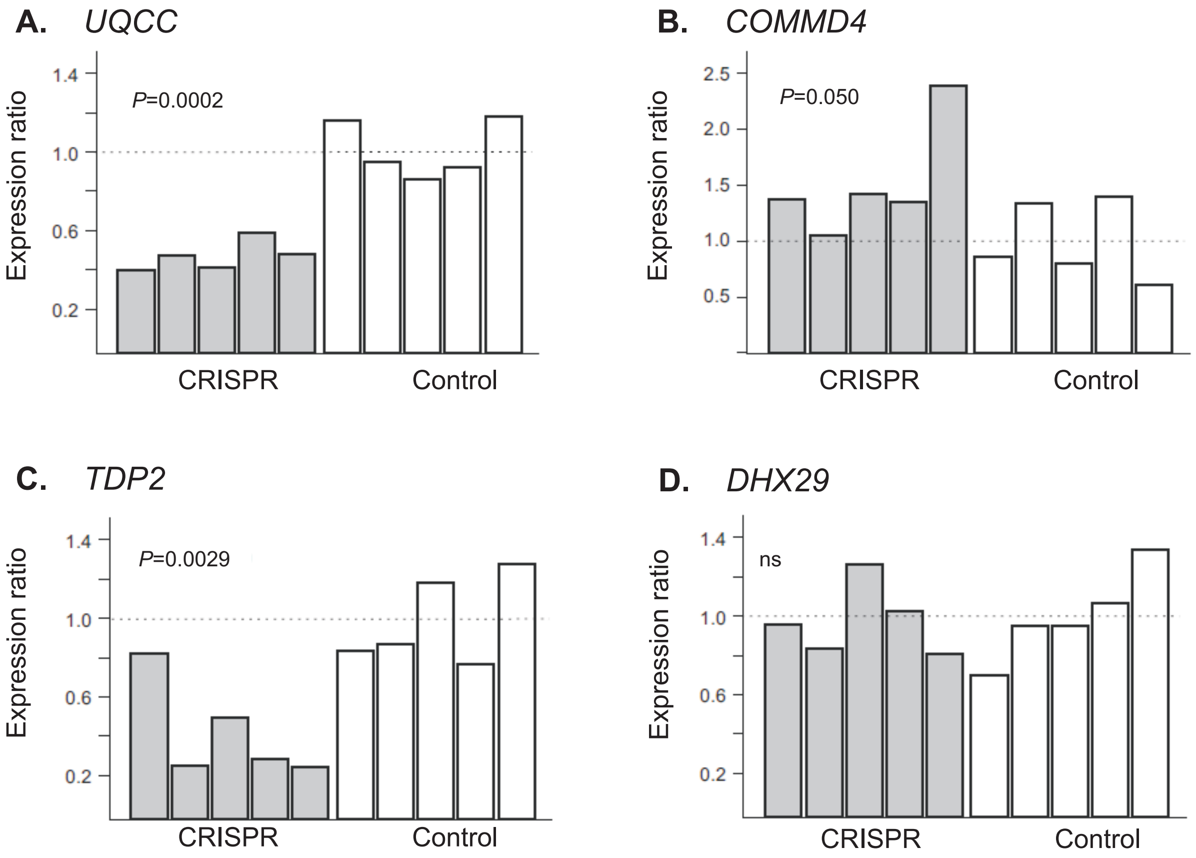

Figure S3. CRISPR / Cas9 mutagenesis validation of rare SNP regulatory effects. Average relative expression measures for the indicated transcript for five CRISPR mutagenized K562 10-cell clones in the same gene (gray) and in a different negative control gene (white). Dotted line represents mean expression of the control clones, to which the mutant clones are compared. See Supplementary methods for details of analysis.

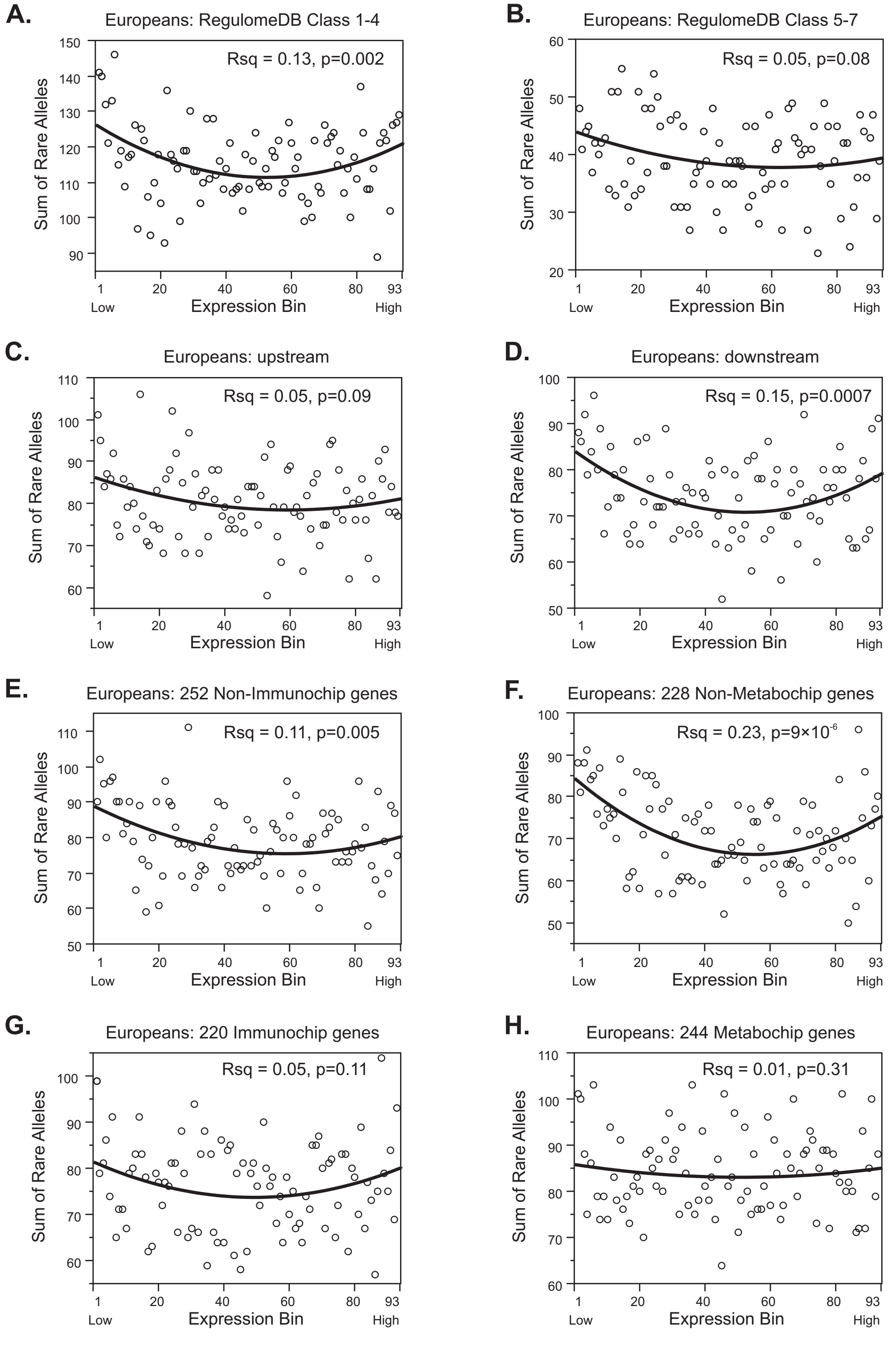

Figure S4. Quadratic regression plots for select sub-sets of SNPs or genes Rsq and p-values refer to the full model R-squared and p-value from a quadratic regression. See Suppl. Table 1 for significance of linear and quadratic terms separately.

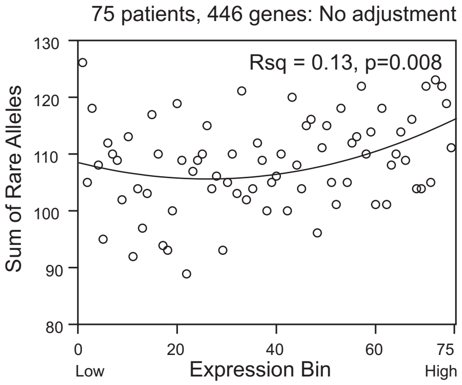

Figure S5. Quadratic regression plots for Replication dataset with 446 genes only, namely the same genes as in the CHDWB that are also present in the RNASeq dataset. Figure 1C shows an expanded replication dataset

Table S1 Comparisons of rare variant burden in subsets of SNPs and Genes

Comparison		First set				Second set			
		AvgCt[1]	Model[2]	Linear[3]	Quad[4]	AvgCt[1]	Model[2]	Linear[3]	Quad[4]
High polymorphism (242)	vs Low polymorphism (230)	104.9	14.4**	0.032	0.002	50.1	0.03^{ns}	0.50	0.12
High expression (236)	vs Low expression (236)	74.9	10.7*	0.35	0.002	80.2	14.5**	0.0008	0.08
Metabochip (244)	vs non-Metabochip (228)	83.9	0.7^{ns}	0.81	0.45	71.1	22.7***	0.006	<0.0001
Immunochip (220)	vs non-Immunochip (252)	76.1	4.8^{ns}	0.72	0.04	78.9	11.2*	0.021	0.017
With eQTL (207)	vs No eQTL (265)	68.7	15.3*	0.027	0.001	86.4	4.9^{ns}	0.33	0.06
Upstream (472)	vs Downstream (472)	80.5	5.1^{ns}	0.12	0.12	74.5	15.0*	0.12	0.0004
RegulomeDB 1-4 (472)	vs RegulomeDB 5-7 (472)	115.7	12.7*	0.16	0.001	39.4	5.5^{ns}	0.08	0.15

[1] Average count of rare variants in each bin in the dataset

[2] Model R-squared, with significance: * 0.01<p<0.001; ** 0.001<p<0.0001; *** p<0.0001

[3] p-value for linear term in the model

[4] p-value for quadratic term in the model

Table S2. Archaic Haplotypes detected in Caucasians

Gene 1: LILRB2. Two individuals with haplotypes introgressed from Archaics

Marker	Chr	Pos (hg19)	Allele 1	Allele 0	MAF	Altai Neanderthal sequence	Denisovan sequence	Caucasian individuals with at least 1 non-HuRef site																					
rs118126757	19	54784069	G	C	0.00244									1		1	1												
rs147377983	19	54784404	T	C	0.00488																								
chr19:54784451	19	54784451	A	T	0.00122																								
rs191476279	19	54784500	T	C	0.00488																								
rs140128507	19	54784516	A	G	0.00122	1			1																				1
rs4806734	19	54784587	T	C	0.00366																								
rs188025641	19	54784621	T	C	0.00122					1					1						1								
chr19:54785045	19	54785045	G	A	0.00488															1									
rs146571067	19	54785085	G	C	0.00122	1	1		1																				1
chr19:54785088	19	54785088	G	C	0.00122																								
chr19:54785089	19	54785089	T	C	0.00366																								
rs112796768	19	54785092	G	C	0.00488						1																		
chr19:54785135	19	54785135	C	A	0.00488	1			1																				1
rs188272820	19	54785143	T	C	0.00488	1			1																				1
rs182465889	19	54785150	A	G	0.00122	1			1																				1
chr19:54785189	19	54785189	A	C	0.00488																			1					
rs149524204	19	54785217	A	G	0.00122	1	1		1																				1
chr19:54785227	19	54785227	A	G	0.00488																					1			
rs117972222	19	54785329	C	T	0.00976	1	1		1																				1
rs191923839	19	54785346	T	C	0.00488			1				1	1					1							2		1		
rs117542982	19	54785534	G	A	0.00122	1			1																				1
chr19:54785535	19	54785535	C	T	0.00488																	1							
rs117127854	19	54785610	T	C	0.00122	1			1																				1
chr19:54785651	19	54785651	T	C	0.00122														1										
chr19:54785721	19	54785721	A	C	0.00244																		1						
rs113145131	19	54785835	T	C	0.01098																								
rs113412095	19	54785863	A	C	0.00122																1								

Gene 2: TGF. Five individuals with haplotypes introgressed from Archaics

Marker	Chr	Pos (hg19)	Allele 1	Allele 0	MAF	Altai Neanderthal sequence	Denisovan sequence	Caucasian individuals with at least 1 non-HuRef site																																								
rs116650206	3	100427151	A	T	0.00976																																											
rs116622528	3	100427159	T	C	0.00854	1						1		1																1			1				1											
rs114652332	3	100427191	T	C	0.00244																																											
rs62274023	3	100427207	C	T	0.0061											1	1												1																	1		
rs9853377	3	100427253	T	C	0.03293	1						1		1																1			1				1											
rs146889339	3	100427283	A	C	0.00122																																											
chr3:100427416	3	100427416	T	C	0.00122																																							1				
rs140713495	3	100427524	G	C	0.00366																																											
rs967308	3	100427676	C	T	0.03293	1						1		1																1			1				1											
rs145846729	3	100427729	T	C	0.00122													1																														
rs967309	3	100427738	T	C	0.03293	1						1		1																1			1				1											
chr3:100427772	3	100427772	T	G	0.00122																					1																						
chr3:100427953	3	100427953	C	T	0.00244					1	1																																					
rs75268146	3	100428061	C	T	0.03415	1						1		1																1			1				1											
rs28364602	3	100428066	G	A	0.00976																																											
rs12493765	3	100428072	A	C	0.00976										1						1					1	1							1														
rs141024756	3	100428195	C	G	0.00854	1						1		1																1			1				1											
rs190536612	3	100428205	G	A	0.00854	1						1		1																1			1				1											
rs112759386	3	100428243	T	C	0.0061																			1																								
chr3:100428245	3	100428245	A	T	0.00122																																								1			
rs72919417	3	100428286	T	G	0.02561	1						1		1																1			1				1											
rs111540048	3	100428298	A	G	0.0122																																											
rs114124529	3	100428356	G	A	0.02561			1	1				1						1	1		1	1		1			1			1	1			1	1		1	1	1	1	1	1				1	1

Gene 3: DLAT. Nine individuals with haplotypes introgressed from Archaics

Marker	Chr	Pos (hg19)	Allele 1	Allele 0	MAF	Altai Neanderthal sequence	Denisovan sequence	Caucasian individuals with at least 1 non-HuRef site																								
rs12099187	11	111894560	T	C	0.00488																											
rs151035291	11	111894896	A	G	0.00732																											
rs78290067	11	111894991	C	T	0.03049										1																	
chr11:111895066	11	111895066	C	G	0.00122																											
chr11:111895067	11	111895067	C	G	0.00244																			1							1	
chr11:111895102	11	111895102	G	T	0.00122							1																				
rs73561950	11	111895193	T	G	0.03415	1	1		1	1	1						1			1			1		1				1			1
rs190887558	11	111895550	A	C	0.00488																											
rs114863504	11	111895560	A	G	0.00732																	1										
rs78298568	11	111895599	T	C	0.0439	1	1		1	1	1						1			1		1	1		1				1			1
chr11:111895636	11	111895636	A	G	0.00122													1														
chr11:111895704	11	111895704	A	C	0.00122																											
chr11:111895840	11	111895840	T	G	0.00244								1																	1		
chr11:111895960	11	111895960	G	C	0.00122																							1				
chr11:111895985	11	111895985	G	C	0.00366									1												1					1	
chr11:111895994	11	111895994	T	C	0.00122											1																
chr11:111896040	11	111896040	A	G	0.00732																	1										
rs115067052	11	111896041	G	C	0.00732																	1										
chr11:111896072	11	111896072	C	A	0.00244			1																								
chr11:111896107	11	111896107	C	G	0.00122																						1					
rs150145390	11	111896242	A	G	0.00122																											
rs61757217	11	111896251	C	G	0.01098	1			1	1	1						1			1			1		1				1			1
chr11:111896339	11	111896339	A	T	0.00122																1											
chr11:111896404	11	111896404	T	C	0.00122														1													
rs201934276	11	111896514	C	T	0.00122																	1										
rs137879141	11	111896524	G	A	0.0122										1																	
rs192971364	11	111896531	A	G	0.00122																	1										

Gene 4: NOP10. Eight individuals with haplotypes introgressed from Archaics

Marker	Chr	Pos (hg19)	Allele 1	Allele 0	MAF	Altai Neanderthal sequence	Denisovan sequence	Caucasian individuals with at least 1 non-HuRef site																					
rs146293018	15	34634434	C	T	0.0122	1	1	1						1					1		1			1	1		1		1
chr15:34634473	15	34634473	C	T	0.00122										1														
rs113680082	15	34634512	C	G	0.00244																								
rs144573482	15	34634727	T	C	0.00488																								
rs148453358	15	34634853	G	T	0.0122	1	1	1						1					1		1			1	1		1		1
chr15:34634987	15	34634987	G	A	0.00122																								
rs146261631	15	34635241	G	C	0.00854			1	1				1			1				1									
chr15:34635348	15	34635348	A	G	0.00122																								
rs149606664	15	34635380	C	T	0.01341	1	1	1					1	1					1		1			1	1		1		1
rs139502071	15	34635383	T	C	0.0122	1		1						1					1		1			1	1		1		1
rs76537972	15	34635391	C	T	0.00366						1	1																1	
rs76029512	15	34635393	A	G	0.00732												1	1				1	1			1			
rs186487749	15	34635399	G	C	0.00854			1	1				1			1				1									
rs75061777	15	34635597	A	G	0.00366																								
rs114643475	15	34635646	G	A	0.00122																								
rs60281344	15	34635684	T	A	0.00366					1																			

Gene 5: MRPL53. One individual with haplotypes introgressed from Archaics

Marker	Chr	Pos (hg19)	Allele 1	Allele 0	MAF	Altai Neanderthal sequence	Denisovan sequence	Caucasian individuals with at least 1 non-HuRef site																														
rs112896280	2	74699066	C	G	0.00488							2																										
chr2:74699141	2	74699141	A	G	0.00854									1						1	1								1				1						1	1
chr2:74699184	2	74699184	T	C	0.00122																						|1|											
rs148007344	2	74699584	C	T	0.00488	1	1																													|1|		
rs78834087	2	74699595	T	G	0.01463						1				1	1					1		1		1				1			1	1			1		
rs201235080	2	74699675	T	C	0.00122																																			
rs200975054	2	74699680	A	C	0.00122																															|1|		
rs141704877	2	74699715	C	T	0.0122			1	1	1			1				1	1										1		1				1				
rs77266305	2	74699922	C	T	0.00488	1	1																													|1|		
chr2:74700003	2	74700003	T	C	0.00122																|1|																	
rs183078378	2	74700031	C	T	0.00366																																			
chr2:74700040	2	74700040	A	G	0.00122																																			
rs13403485	2	74700190	C	A	0.01341																																			
chr2:74700213	2	74700213	G	A	0.00122																																			
rs139465854	2	74700214	A	G	0.00244																																			
rs141818840	2	74700242	A	T	0.00244																																			
rs17009955	2	74700451	C	A	0.0122	1	1					2																									1			
rs183406812	2	74700504	C	G	0.00244																																			
chr2:74700609	2	74700609	A	G	0.00122																															|1|		
chr2:74700656	2	74700656	T	C	0.00244																																			
chr2:74700659	2	74700659	A	G	0.00122																		|1|															
chr2:74700754	2	74700754	A	G	0.00122																																			

Gene 6: ALDH3B1. One individual with haplotypes introgressed from Archaics

Marker	Chr	Pos (hg19)	Allele 1	Allele 0	MAF	Altai Neanderthal sequence	Denisovan sequence	Caucasian individuals with at least 1 non-HuRef site																																																				
chr11:67775057	11	67775057	A	G	0.00244																								|1|							|1|																							
chr11:67775089	11	67775089	T	C	0.00122																																																							|1
chr11:67775337	11	67775337	A	C	0.00122															1																																								
chr11:67775397	11	67775397	A	C	0.00122																																									1												
rs78790062	11	67775439	G	C	0.00366																																																							
chr11:67775464	11	67775464	C	A	0.00122										1																																													
rs308340	11	67775505	A	G	0.02317																																																							
chr11:67775709	11	67775709	C	G	0.00122																																				|1|																		
chr11:67775821	11	67775821	T	C	0.00122																																																							
rs75423674	11	67775822	A	G	0.02317	1	1		|1|															|1|																																				
chr11:67775842	11	67775842	A	C	0.00122									1																																														
chr11:67776124	11	67776124	A	G	0.00122												1																																											
chr11:67776137	11	67776137	C	T	0.00122																																									1														
chr11:67776152	11	67776152	T	C	0.00122																																																							
chr11:67776153	11	67776153	A	G	0.00122																																																							
chr11:67776279	11	67776279	A	G	0.00122																																																							
rs117602660	11	67776283	T	C	0.01707		1	1	1								1										1				1									1			1		1					1			1		1				1	1	
rs308337	11	67776292	A	C	0.03659								1																					1																											
chr11:67776311	11	67776311	A	G	0.00122																																												1									
rs308336	11	67776320	C	T	0.03659								1																					1																											
rs184759902	11	67776353	T	C	0.00244				|1|															|1|																																				
rs3763941	11	67776362	T	C	0.0439						1	1				1		1	1			1	1	1	1				1	1		1		1	1	2		2		1			1		1	1	1			1			1		1	1	1			1
rs140148305	11	67776387	T	C	0.00854																																																							
chr11:67776410	11	67776410	A	G	0.00122														1																																									
rs557098	11	67776465	C	T	0.0378								1																					1																											
rs116715743	11	67776488	G	A	0.00122																																																							
rs149823943	11	67776579	T	G	0.00488																																																							
chr11:67776620	11	67776620	T	C	0.00122																																																							
rs7113328	11	67776664	A	G	0.00366				|1|															|1|																																				
rs308335	11	67776686	G	A	0.03659								1																					1																											
rs308334	11	67776852	C	G	0.03659								1																					1																											
rs189473290	11	67776917	T	G	0.00244	1			|1|															|1|																																				
rs147659398	11	67776924	T	G	0.0061																																																							
rs308333	11	67777026	A	G	0.03659								1																					1																											
chr11:67777046	11	67777046	T	C	0.00122																																																							

Gene 7: C7orf25. 22 individuals with haplotypes introgressed from Archaics

Marker	Chr	Pos (hg19)	Allele 1	Allele 0	MAF	Altai Neanderthal sequence	Denisovan sequence	Caucasian individuals with at least 1 non-HuRef site																								
chr7:42950799	7	42950799	C	T	0.00122																											
chr7:42951067	7	42951067	C	G	0.00122																											
rs148046461	7	42951174	A	G	0.00244																											
rs28372722	7	42951646	T	C	0.03171	1		1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1		1	1	1	2	1	1	1
chr7:42951652	7	42951652	A	G	0.00244																											
chr7:42951824	7	42951824	G	T	0.00244																											
chr7:42951879	7	42951879	T	G	0.00122																											
rs182726646	7	42951898	T	C	0.00122																			1								
rs7786568	7	42952016	G	C	0.00976																											
rs28372721	7	42952186	C	G	0.03171	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1		1	1	1	2	1	1	1
chr7:42952282	7	42952282	G	A	0.00122													1														
rs184091995	7	42952346	T	C	0.00122															1												

Gene 8: RBBP4. 23 individuals with haplotypes introgressed from Archaics

Marker	Chr	Pos (hg19)	Allele 1	Allele 0	MAF	Altai Neanderthal sequence	Denisovan sequence	Caucasian individuals with at least 1 non-HuRef site																																																								
rs145261649	1	33116103	A	C	0.00122																						1																																										
rs138735449	1	33116168	T	C	0.00122																																																															
rs41265851	1	33116206	T	A	0.02317					1						1				1	1		1	1												1					1		1	1		1		1		1																	
rs113601952	1	33116359	G	A	0.00976				1				1															1					1														1					1		1													
rs188230717	1	33116546	T	C	0.00122																																																														
chr1:33116575	1	33116575	A	G	0.00122																																1																														
chr1:33116635	1	33116635	T	C	0.00122																																																														
chr1:33116811	1	33116811	G	A	0.00122																																																						1							
chr1:33116852	1	33116852	C	T	0.01829										1				1						1				1	1				1		1				1					1							1						1								1	
chr1:33116853	1	33116853	T	C	0.00122																																										1																			
chr1:33116961	1	33116961	A	G	0.00244																																																															
chr1:33117052	1	33117052	A	G	0.00122														1																																																
rs583114	1	33117071	T	G	0.02683																																																															
rs12407673	1	33117092	T	C	0.03415	1		1			1			1			1	1				1			1			1	1		1		1					1			1							1			1	1		1			1	1		1	1		1		1
chr1:33117099	1	33117099	T	C	0.00244							1																										1																														
rs12407680	1	33117137	G	C	0.03415	1	1	1			1			1			1	1				1			1			1	1		1		1					1			1							1			1	1		1			1	1		1	1		1		1
rs664567	1	33117231	G	A	0.02683																																																															
chr1:33117343	1	33117343	T	C	0.00122																																																															
rs146406493	1	33117429	A	G	0.0061																																																															
chr1:33117683	1	33117683	A	G	0.00122																																																															
rs182390022	1	33117735	T	C	0.00122																																																								1			
rs186772311	1	33117743	G	A	0.00244																																																																

Table S3. Validation of SNP effects by CRISPR/Cas9

Gene	SNP	Estimated effect[1]	Change in ddPCR[2]	p-value	N clones Δ[3]
UQCC	chr20:33999719	0.17X	0.48X	0.0002	5 of 5
TDP2	chr6:24667167	0.11X	0.46X	0.0029	4 of 5
COMMD4	rs182080358	12.5X	1.53X	0.050	4 of 5
DHX29	chr5:54603837	12.9X	0.98X	0.43	1 of 5

[1] Inferred from difference in log2 fluorescence intensity between the individual with the mutation and all other 409 samples

[2] Average change in ratio of drop digital PCR counts relative to two reference genes normalized to unity

[3] Number of the 5 clones deviant in the expected direction (see Suppl. Fig. 6).

APPENDIX: R code

//Code for PEER Normalization

Input is 279 Caucasian samples with 11,056 genes, Quantile normalized with SAS/JMP

//std for genes within batches(279, 11056G), get lists of std by batch

```
res <- apply(qnm_edata, 2, tapply, qnm_edata$Batch, scale)
```

  //combine each gene, then combine all genes

```
std=matrix(rep(0),279,11056)
for(i in 1:11056){
tmp=do.call("rbind",res[[i]])
std[,i]=tmp[order(row.names(tmp)),]}
colnames(std)=colnames(qnm_edata)[2,11057]
rownames(std)=rownames(qnm_edata)
```

//ComBat fit batch

```
library(SVA)
edata=t(std)
batch=pheno$Batch
modcombat = model.matrix(~1, data=pheno)
combat_edata = ComBat(dat=edata, batch=batch, mod=modcombat, par.prior=TRUE, prior.plots=FALSE)
```

//transpose data

```
edata_ComBat=as.matrix(combat_edata)
transpose_ComBat_edata=t(edata_ComBat)
ComBat_edata=as.data.frame(transpose_ComBat_edata)
```

//Fit Age and Sex

```
rr=matrix(rep(0),279,11056)
for(i in 1:11056){
 ff=lm(ComBat_edata[,i]~pheno$Gender+pheno$Age)
 rr[,i]=ff$res
 }
colnames(rr)=colnames(transpose_ComBat_edata)
rownames(rr)=rownames(transpose_ComBat_edata)
```

//Fit Axes (analysis not reported in paper, for comparison with SVA)

```
exp=rr
res=matrix(rep(0),279,11056)
```

```
Pmin=rep(1,11056)
Cmin=rep(1,11056)

for(i in 1:11056){
        Pmin[i]=1
        Cmin[i]=1
        for(j in 1:7){
                fit=lm(exp[,i]~Axes[,j])
                if(summary(fit)$coef[2,4]<Pmin[i]){
                        Cmin[i]=j
                        Pmin[i]=summary(fit)$coef[2,4]
                        res[,i]=fit$res}
                }
        }
Bind=cbind(Cmin,Pmin)
colnames(res)=colnames(transpose_ComBat_edata)
rownames(res)=rownames(transpose_ComBat_edata)
write.csv(Bind,file="279CAU_qnm_std_ComBat_fitAgeSex_fitAxis_coef.csv")
write.csv(res,file="279CAU_qnm_std_ComBat_fitAgeSex_fitAxis.csv")
```

```
//PEER -- NK20

library(peer)
expr=res
model = PEER()
PEER_setPhenoMean(model,as.matrix(expr))
PEER_setNk(model,20)
PEER_update(model)
factors = PEER_getX(model)
weights = PEER_getW(model)
precision = PEER_getAlpha(model)
residuals = PEER_getResiduals(model)
pdf('PEER_Nk20_Model.pdf')
PEER_plotModel(model)
dev.off()
pdf('PEER_Nk20_Precision.pdf')
plot(precision)
dev.off()
```

//SNM normalization (on 546 samples, 14111 probes)

//Take mean of probes for each gene (on 546 samples, 14111 probes) using apply function in R

//Remove duplicate sample GG2_0043 and American Indian sample(GG1-000149) --> 544sample_SNM.csv

//Read tab-delimited files for expression, biological, adjustment and samples

```
chdwb.snm = read.csv("C:/Documents/544sample_SNM.csv", header=T, row.names=1)
chdwb.bio = read.csv("C:/Documents/chdwb_ageBethnR_bio.csv", header=T, row.names=1)
chdwb.adj = read.csv("C:/Documents/chdwb_ageBethnR_adj.csv", header=T, row.names=1)
chdwb.int = read.csv("C:/Documents/chdwb_ageBethnR_int.csv", header=T)
```

//Create model matrices and run SNM

```
int.var = chdwb.int
int.var$Array = as.factor(int.var$Array)
adj.var = model.matrix(~.,chdwb.adj)
bio.var = model.matrix(~.,chdwb.bio)
raw.data = as.matrix(chdwb.snm)
snmR.chdwb = snm(raw.data,bio.var,adj.var,int.var,rm.adj=TRUE,num.iter=10)
write.table(snmR.chdwb$norm.dat, file = "C:/Documents/544sample_SNM_stdBatch.csv", sep=",",
col.names=NA)
```

//For SVA analysis, this step was included to use a linear model to fit age and gender to 472 genes and 411 samples among 544 which have genotyping information

```
        //input "544sample_SNM" to R as "exp"
        //input phenotype file to R as "pheno"
rr=matrix(rep(0),411,472)
for(i in 1:472){
 ff=lm(exp[,i]~pheno$Gender+pheno$Age)
 rr[,i]=ff$res
 }
write.table(rr, file = "C:/Documents/544sample_SVA_stdBatch_agegender.csv", sep=",", col.names=NA)
```

//Fit Independent Common eQTL (within 1kb upstream and gene region) for 411 samples with genotyping information

//Prepare eQTL files with row as samples, column as each common eSNP (order by coordinates) --> input to R as "gen"

//Prepare expression file with 207 genes with common cis-eQTLs (order by coordinates)
        --> input to R as "exp"

//Prepare another expression file with the left 265 genes without cis-eQTLs

//Prepare a file with two columns, the first column is gene name, the second column is the number of common eQTLs.   --> input to R as "count"

```
j=1
res=exp
Pmin=rep(1,207)
Cmin=rep(1,207)
for(i in 1:207){
        Cend=j+count[i,2]-1
        for (m in j:Cend){
                fit=lm(exp[,i]~gen[,m])
                if(summary(fit)$coef[2,4]<Pmin[i]&&summary(fit)$coef[2,4]<0.05){
                Cmin[i]=m
                Pmin[i]=summary(fit)$coef[2,4]
                res[,i]=fit$res}
                }
        exp[,i]=res[,i]
        j=Cend+1}
        //Check Pmin, if there is at least one Pmin<0.05, then re-do the whole steps, until all Pmins>=0.05 for all genes
        //Output res, combine with the 265 gene expression.
```

//385 samples who have expression levels.were included in 411 samples   --> 385sample_SNM_stdBatch_fitEqtl

//**Fit the most significant Axis** to 385 samples

//input "385sample_SNM_stdBatch_fitEqtl" to R as "exp"

//Prepare axis file with 7 axes   --> input to R as "Axes"

```
res=matrix(rep(0),385,472)
Pmin=rep(1,472)
Cmin=rep(1,472)
for(i in 1:472){
        Pmin[i]=1
        Cmin[i]=1
        for(j in 1:7){
                fit=lm(exp[,i]~Axes[,j])
                if(summary(fit)$coef[2,4]<Pmin[i]){
                        Cmin[i]=j
                        Pmin[i]=summary(fit)$coef[2,4]
                        res[,i]=fit$res}
                }
        }
```

//res --> expression data for 385 samples after normalization and adjustment

//279 are Caucasians among 385 samples. --> 279CAU_SNM_stdBatch_fitEqtlAxis

//**Rare variant association test**

//input "279CAU_SNM_stdBatch_fitEqtlAxis" as "exp"

//input rare variant counts per gene per sample as "snp"

//the below code split genes to 93 bins with 3 genes in each bin for each sample

```
count=c(rep(0,93))
for(i in 1:472){
Rk=rank(exp[i,],ties.method="first")
        for(j in 1:279){
                zz=as.integer((Rk[j]+2)/3)
                if(snp[i,j]!=0){
                count[zz]=count[zz]+snp[i,j]}
        j=j+1}
i=i+1}
perc=1:93
```

//"count" --> rare variant counts in each bin

//"perc" --> bin number (expression bin from lowest to highest)

//**Perform quadratic test on count and perc**

SAS/JMP -- Analyze -- Fit Y by X -- Fit polynomial – quadratic