GAW17 Mini-Exome Simulation: The "Answers"

Knowledge about biological pathways and statistical predictions regarding the potential deleteriousness of coding variants were used in designing the simulation model. A subset of genes in particular pathways that had sequence data available in the 1000 Genomes Project were chosen and effect sizes for coding variants within those genes were assigned using PolyPhen and SIFT predictions of the likelihood that the variant would be deleterious. Genes influencing Q1 come primary from the VEGF pathway, those influencing Q2 are primarily related to cardiovascular disease risk and inflammation, and those influencing latent disease liability also come primarily from VEGF (a different section of the pathway from the genes selected for Q1). Information on predicted deleteriousness was used to select functional variants. The functional variants include both rare and common alleles and a range of effect sizes, with most having small effects but a few having large effects that should be reliably detectable in most replicates of the data set. Some genes contain a single functional variant and others contain many. Population origin of the 1000 Genomes participants was not used in the phenotype simulations.

Quantitative risk factors Q1, Q2, and Q4 were simulated as normally distributed phenotypes. Disease was simulated using a liability threshold model and the top 30% of the distribution was declared affected. All SNP effects are additive on the quantitative trait or liability scale, with each copy of the minor allele increasing or decreasing the mean trait value by an equal amount. GxE effects were simulated for Q1.

Quantitative risk factor Q1 is influenced by 39 SNPs in 9 genes (see table 1). There are 1 – 11 functional variants per gene. Their minor allele frequencies (MAFs) in the 1000 Genomes data range from 0.07% (i.e. a single copy of the minor allele) to 16.5%. In all cases, the minor allele is associated with higher mean Q1; the beta column in the table provides the displacement in mean levels of Q1 for each copy of the minor allele. Q1 also has a residual heritability of 0.44, due to variants at loci not included in the current sequence data set. The residual genetic component of Q1 is correlated with the residual genetic components of Q2 and latent liability. There are also weaker environmental correlations between Q1 and Q2 and latent liability. Values of Q1 are higher in smokers and there is GxSmoking interaction for the effects of variants in the KDR gene on Q1. Effects of the KDR variants are 50% higher in smokers than in non-smokers. (Note that for KDR the effect sizes given in table 1 are those for non-smokers.) Q1 also increases with age.

Q2 is influenced by 72 SNPs in 13 genes (see table 2). There are 1 – 13 functional variants per gene. MAFs range from 0.07% to 17.07%. In all cases, the minor allele is associated with higher mean Q2. Q2 has a residual heritability of 0.29. The residual genetic component of Q2 is correlated with Q1 and with latent liability. There are also weaker environmental correlations between Q2 and Q1 and latent liability. Q2 is not influenced by age, sex, or smoking.

Q4 has a heritability of 0.70, but none of this genetic component is due to genes in this sequencing set. (i.e. it is not influenced by any of the genotyped exonic SNPs) Q4 is lower in smokers, decreases with age, and is lower in females. Q4 is protective – individuals with higher levels of Q4 have lower risk of disease.

2000.

A normally distributed latent liability trait (not included in the phenotype data distributed) was simulated that was influenced by 51 SNPs in 15 genes with 1 – 24 functional variants per gene (see table 3). MAFs of these variants range from 0.07% to 25.8%. In all cases, the minor allele is associated with higher mean liability. This latent liability trait is also higher in smokers and increases with age. Disease risk was a function of this latent liability, Q1, Q2, and Q4:

Liability to disease = latent liability + Q1 + Q2 – Q4

Using this formula, a quantitative liability score was calculated for each individual and the top 30% of the distribution was declared affected. The effect sizes in table 3 are for liability to disease.

TABLE 1: Effects on Q1

GENE	SNP	MAF	BETA
ARNT	C1S6533	0.011478	0.56190
ARNT	C1S6537	0.000717	0.64454
ARNT	C1S6540	0.001435	0.24129
ARNT	C1S6542	0.002152	0.46026
ARNT	C1S6561	0.000717	0.65721
ELAVL4	C1S3181	0.000717	0.76911
ELAVL4	C1S3182	0.000717	0.30432
FLT1	C13S320	0.001435	0.19605
FLT1	C13S399	0.000717	0.39602
FLT1	C13S431	0.017217	0.74136
FLT1	C13S479	0.000717	0.75946
FLT1	C13S505	0.000717	0.44850
FLT1	C13S514	0.000717	0.56643
FLT1	C13S522	0.027977	0.61830
FLT1	C13S523	0.066714	0.64997
FLT1	C13S524	0.004304	0.62223
FLT1	C13S547	0.000717	0.52601
FLT1	C13S567	0.000717	0.17493
FLT4	C5S5133	0.001435	0.15986
FLT4	C5S5156	0.000717	0.43010
HIF1A	C14S1718	0.000717	0.15382
HIF1A	C14S1729	0.002152	0.28532
HIF1A	C14S1734	0.012195	0.21203
HIF1A	C14S1736	0.000717	0.21716
HIF3A	C19S4799	0.000717	0.28351
HIF3A	C19S4815	0.000717	0.53114
HIF3A	C19S4831	0.000717	0.29287
KDR	C4S1861	0.002152	0.56311
KDR	C4S1873	0.000717	0.58301
KDR	C4S1874	0.000717	0.47262

KDR	C4S1877	0.000717	1.07706
KDR	C4S1878	0.164993	0.13573
KDR	C4S1879	0.000717	0.61830
KDR	C4S1884	0.020803	0.29558
KDR	C4S1887	0.000717	0.29558
KDR	C4S1889	0.000717	0.94133
KDR	C4S1890	0.002152	0.42407
VEGFA	C6S2981	0.002152	1.20645
VEGFC	C4S4935	0.000717	1.35726

TABLE 2: Effects on Q2

GENE	SNP	MAF	BETA
BCHE	C3S4834	0.000717	0.24092
BCHE	C3S4836	0.000717	0.23749
BCHE	C3S4856	0.000717	0.22027
BCHE	C3S4859	0.002152	0.59302
BCHE	C3S4860	0.000717	0.25057
BCHE	C3S4862	0.000717	1.01672
BCHE	C3S4867	0.000717	0.65326
BCHE	C3S4869	0.000717	1.01569
BCHE	C3S4873	0.002869	0.59096
BCHE	C3S4874	0.000717	1.00570
BCHE	C3S4875	0.000717	1.09484
BCHE	C3S4876	0.000717	0.75583
BCHE	C3S4880	0.001435	0.20651
GCKR	C2S354	0.012195	0.37757
INSIG1	C7S5132	0.000717	0.19962
INSIG1	C7S5133	0.000717	0.19618
INSIG1	C7S5144	0.000717	0.19275
LPL	C8S442	0.015782	0.49459
LPL	C8S476	0.000717	0.63365
LPL	C8S530	0.001435	0.72864
PDGFD	C11S5292	0.008608	0.58270
PDGFD	C11S5299	0.000717	0.82157
PDGFD	C11S5301	0.000717	0.87904
PDGFD	C11S5302	0.001435	0.81502
PLAT	C8S1741	0.003587	0.68079
PLAT	C8S1742	0.000717	0.84910
PLAT	C8S1758	0.001435	0.92516
PLAT	C8S1770	0.000717	0.62916
PLAT	C8S1772	0.001435	0.26296
PLAT	C8S1773	0.001435	0.55792
PLAT	C8S1799	0.005739	0.20651

PLAT	C8S1811	0.001435	0.13767
RARB	C3S635	0.000717	0.70936
RARB	C3S679	0.005022	0.63502
SIRT1	C10S3048	0.002152	0.83224
SIRT1	C10S3050	0.002152	0.97060
SIRT1	C10S3058	0.000717	0.36621
SIRT1	C10S3092	0.000717	0.43608
SIRT1	C10S3093	0.000717	0.53520
SIRT1	C10S3107	0.000717	0.93549
SIRT1	C10S3108	0.000717	0.53280
SIRT1	C10S3109	0.000717	0.51421
SIRT1	C10S3110	0.002152	0.10326
SREBF1	C17S1007	0.002152	0.53073
SREBF1	C17S1009	0.000717	0.64568
SREBF1	C17S1024	0.004304	0.45329
SREBF1	C17S1030	0.000717	0.80366
SREBF1	C17S1043	0.004304	0.49941
SREBF1	C17S1045	0.003587	0.33524
SREBF1	C17S1046	0.002869	0.62779
SREBF1	C17S1048	0.001435	0.28739
SREBF1	C17S1055	0.001435	0.87767
SREBF1	C17S1056	0.000717	0.51524
VLDLR	C9S367	0.000717	0.58476
VLDLR	C9S376	0.002869	0.53280
VLDLR	C9S377	0.001435	1.21565
VLDLR	C9S391	0.000717	0.52694
VLDLR	C9S430	0.000717	0.55551
VLDLR	C9S443	0.001435	0.62642
VLDLR	C9S444	0.001435	0.86528
VLDLR	C9S497	0.000717	0.65808
VNN1	C6S5378	0.005739	0.45811
VNN1	C6S5380	0.170732	0.24437
VNN3	C6S5412	0.000717	0.64431
VNN3	C6S5426	0.032999	0.10326
VNN3	C6S5439	0.000717	0.10326
VNN3	C6S5441	0.098278	0.27053
VNN3	C6S5446	0.000717	0.48014
VNN3	C6S5448	0.000717	0.54036
VNN3	C6S5449	0.010043	0.66909
VWF	C12S181	0.000717	0.74757
VWF	C12S211	0.005739	0.33661

TABLE 3: Effects on disease liability

GENE	SNP	MAF	BETA
AKT3	C1S11396	0.000717	0.16954
BCL2L11	C2S2286	0.000717	0.16019
BCL2L11	C2S2288	0.002869	0.24659
BCL2L11	C2S2307	0.000717	0.27778
ELAVL4	C1S3181	0.000717	0.30946
ELAVL4	C1S3182	0.000717	0.12245
HSP90AA1	C14S3630	0.000717	0.04976
HSP90AA1	C14S3695	0.000717	0.04854
HSP90AA1	C14S3704	0.003587	0.04005
HSP90AA1	C14S3706	0.258250	0.03944
NRAS	C1S5748	0.000717	0.18131
PIK3C2B	C1S9164	0.001435	0.06638
PIK3C2B	C1S9165	0.000717	0.07803
PIK3C2B	C1S9172	0.004304	0.23580
PIK3C2B	C1S9173	0.001435	0.04005
PIK3C2B	C1S9174	0.000717	0.26577
PIK3C2B	C1S9189	0.006456	0.19102
PIK3C2B	C1S9200	0.000717	0.28725
PIK3C2B	C1S9222	0.000717	0.21480
PIK3C2B	C1S9250	0.001435	0.13106
PIK3C2B	C1S9266	0.002869	0.10655
PIK3C2B	C1S9267	0.002152	0.21371
PIK3C2B	C1S9306	0.000717	0.10546
PIK3C2B	C1S9320	0.000717	0.26383
PIK3C2B	C1S9333	0.000717	0.30242
PIK3C2B	C1S9346	0.000717	0.15922
PIK3C2B	C1S9373	0.000717	0.19927
PIK3C2B	C1S9391	0.000717	0.27621
PIK3C2B	C1S9423	0.000717	0.27354
PIK3C2B	C1S9432	0.010760	0.20339
PIK3C2B	C1S9445	0.000717	0.23240
PIK3C2B	C1S9446	0.000717	0.21698
PIK3C2B	C1S9449	0.000717	0.25048
PIK3C2B	C1S9455	0.002869	0.22669
PIK3C2B	C1S9457	0.000717	0.24647
PIK3C3	C18S2475	0.000717	0.34817
PIK3C3	C18S2492	0.017217	0.23858
PIK3R3	C1S2919	0.000717	0.17172
PRKCA	C17S4578	0.166428	0.17038
PRKCA	C17S4581	0.000717	0.03884
PRKCB1	C16S1894	0.000717	0.23956
PTK2	C8S4825	0.000717	0.03762

PTK2	C8S4839	0.000717	0.03641
PTK2B	C8S886	0.000717	0.20764
PTK2B	C8S900	0.001435	0.03034
PTK2B	C8S909	0.001435	0.18215
RRAS	C19S4929	0.001435	0.15266
RRAS	C19S4937	0.001435	0.20958
SHC1	C1S7061	0.006456	0.09514
SOS2	C14S1381	0.000717	0.23895
SOS2	C14S1382	0.003587	0.28058