Online methods

Contents

1 Discovery data

1.1 Data description

The Brisbane Systems Genetics Study (BSGS) comprises 846 individuals of European descent from 274 independent families.[1] DNA samples from each individual were genotyped on the Illumina 610-Quad Beadchip by the Scientific Services Division at deCODE Genetics Iceland. Full details of genotyping procedures are given in Medland et al.[2] Standard quality control (QC) filters were applied and the remaining 528,509 autosomal SNPs were carried forward for further analysis.

Gene expression profiles were generated from peripheral blood collected with PAXgene TM tubes (QIAGEN, Valencia, CA) using Illumina HT12-v4.0 bead arrays. The Illumina HT-12 v4.0 chip contains 47,323 probes, although some probes are not assigned to RefSeq genes. We removed any probes that did not match the following criteria: contained a SNP within the probe sequence with MAF > 0.05 within 1000 genomes data; did not map to a listed RefSeq gene; were not significantly expressed (based on a detection p-value < 0.05) in at least 90% of samples. After this stringent QC 7339 probes remained for 2D-eQTL mapping.

1.2 Normalisation

Gene expression profiles were normalised and adjusted for batch and polygenic effects. Profiles were first adjusted for raw background expression in each sample. Expression levels were then adjusted using quantile and $\log_2$ transformation to standardise distributions between samples. Batch and polygenic effects were adjusted using the linear model

$$y = \mu + \beta_1 c + \beta_2 p + \beta_3 s + \beta_4 a + g + e \tag{1}$$

where μ is the population mean expression levels, c, p, s and a are vectors of chip, chip position, sex and generation respectively, fitted as fixed effects; and g is a random additive polygenic effect with a variance covariance matrix

$$G_{jk} = \begin{cases} \sigma_a^2 & j = k \\ 2\phi_{jk}\sigma_a^2 & j \neq k \end{cases} \tag{2}$$

The parameter σ_a^2 is the variance component for additive background genetic. Here, we are using family based pedigree information rather than SNP based IBD to account for relationships between individuals and so ϕ_{jk} is the kinship coefficient between individuals j and k. The residual, e, from equation 1 is assumed to follow a multivariate normal distribution with a mean of zero. Residuals were normalised by rank transformation and used as the adjusted phenotype for the pairwise epistasis scan to remove any skewness and avoid results being driven by outliers. The GenABEL package for R was used to perform the normalisation.[3]

2 Exhaustive 2D-eQTL analysis

2.1 Two stage search

We used epiGPU[4] software to perform an exhaustive scan for pairwise interactions, such that each SNP is tested against all other SNPs for statistical association with the expression values for each of the 7339 probes. This uses the massively parallel computational architecture of graphical processing units (GPUs) to speed up the exhaustive search. For each SNP pair there are 9 possible genotype classes. We treat each genotype class as a fixed effect and fit an 8 *d.f.* F-test to test the following hypotheses:

$$H_0 : \sum_{i=1}^{3}\sum_{j=1}^{3}(\bar{x}_{ij} - \mu)^2 = 0; \tag{3}$$

$$H_1 : \sum_{i=1}^{3}\sum_{j=1}^{3}(\bar{x}_{ij} - \mu)^2 > 0; \tag{4}$$

where μ is the mean expression level and x_{ij} is the pairwise genotype class mean for genotype i at SNP 1 and genotype j at SNP 2. This type of test does not parameterize for specific types of epistasis, rather it tests for the joint genetic effects at two loci. This has been demonstrated to be statistically more efficient when searching for a wide range of epistatic patterns, although will also include any marginal effects of SNPs which must be dealt with post-hoc.[5]

2.1.1 Stage 1

The complete exhaustive scan for 7339 probes comprises 1.03×10^{15} F-tests. We used permutation analysis to estimate an appropriate significance threshold for the study. To do this we performed a further 1600 exhaustive 2D scans on permuted phenotypes to generate a null distribution of the extreme p-values expected to be obtained from this number of multiple tests given the correlation structure between the SNPs. We took the most extreme p-value from each of the 1600 scans and set the 5% FWER to be the 95% most extreme of these p-values, $T_* = 2.13 \times 10^{-12}$. The effective number of tests in one 2D scan being performed is therefore $N_* = 0.05/T_* \approx 2.33 \times 10^{10}$. To correct for the testing of multiple traits we established an experiment wide threshold of $T_e = 0.05/(N_* \times 7339) = 2.91 \times 10^{-16}$. This is likely to be conservative as it assumes independence between probes.

Filtering We used two approaches to filter SNPs from stage 1 to be tested for significant interaction effects in stage 2.

Filter 1 After keeping SNP pairs that surpassed the 2.91×10^{-16} threshold in stage 1 only SNP pairs with at least 5 data points in all 9 genotype classes were kept. We then calculated the LD between interacting SNPs (amongst

unrelated individuals within the discovery sample and also from 1000 genomes data) and removed any pairs with $r^2 > 0.1$ or $D'^2 > 0.1$ to avoid the inclusion of haplotype effects and to increase the accuracy of genetic variance decomposition. If multiple SNP pairs were present on the same chromosomes for a particular expression trait then only the sentinel SNP pair was retained, *i.e.* if a probe had multiple SNP pairs that were on chromosomes one and two then only the SNP pair with the most significant p-value was retained. At this stage 6404 filtered SNP pairs remained.

Filter 2 We also performed a second filtering screen applied to the list of SNP pairs from stage 1 that was identical to filter 1 but an additional step was included where any SNPs that had previously been shown to have a significant additive or dominant effect ($p < 1.29 \times 10^{-11}$) were removed,[6] creating a second set of 4751 unique filtered SNP pairs.

2.1.2 Stage 2

To ensure that interacting SNPs were driven by epistasis and not marginal effects we performed a nested ANOVA on each pair in the filtered set to test if the interaction terms were significant. We did this by contrasting the full genetic model (8 *d.f.*) against the reduced marginal effects model which included the additive and dominance terms at both SNPs (4 *d.f.*). Thus, a 4 *d.f.* F-test was performed on the residual genetic variation, representing the contribution of epistatic variance. Significance of epistasis was determined using a Bonferroni threshold of $0.05/(6404 + 4751) = 4.48 \times 10^{-6}$. This resulted in 406 and 95 SNP pairs with significant interaction terms from filters 1 and 2, respectively.

2.2 Type 1 error rate

Using a Bonferroni correction of 0.05 in the second stage of the two stage discovery scan implies a type 1 error rate of $\alpha = 0.05$. However, this could be underestimated because the number tests performed in the second stage depends on the number of tests in the first stage, and this depends on statistical power and model choice. We performed simulations to estimate the type 1 error rate of this study design.

We assumed a null model where there was one true additive effect and 7 other terms with no effect. To simulate a test statistic we simulated 8 z-scores, $z_1 \sim N(\sqrt{NCP}, 1)$ and $z_{2..8} \sim N(0, 1)$. Thus $z_{full} = \sum_{i=1}^{8} z_i \sim \chi^2_8$ (representing the 8 d.f. test) and $z_{int} = \sum_{i=5}^{8} z_i \sim \chi^2_4$ (representing the 4 d.f. test where the null hypothesis of no epistasis is true). For a particular value of NCP we simulated 100,000 z values, and calculated the p_{full}-value for the z_{full} test statistic. The n_{int} test statistics with $p_{full} < 2.31 \times 10^{-16}$ were kept for the second stage, where the type 1 error rate of stage 2 was calculated as the proportion of $p_{int} < 0.05/n_{int}$. The power at stage 1 was calculated as $n_{int}/100,000$. This procedure was performed for a range of NCP parameters that represented power ranging from ~ 0 to ~ 1.

2.3 Population stratification

We ruled out population stratification as a possible cause of inflated test statistics. To test for cryptic relatedness driving the interaction terms we tested for increased LD among the SNPs.[7] We calculated the mean of the off-diagonal elements of the correlation matrix of all unique SNPs from the 501 interactions (731 SNPs) using only unrelated individuals, $\bar{r^2} = 0.0039$. This is not significantly different from the null hypothesis of zero (sampling error $= 1/n_{\text{unrelated}} = 0.0039$).

2.4 Probe mapping

To avoid possibility that epistatic signals might arise due to expression probes hybridising in multiple locations we verified that probe sequences for genes with significant interactions mapped to only a single location. As an initial verification we performed a BLAST search of the full probe sequence against 1000 genomes phase 1 version 3 human genome reference and ensured that only one genomic location aligned significantly ($p < 0.05$). As a second step, to mitigate the possibility of weak hybridisation elsewhere in the genome we divided the probe sequence into three sections (1-25bp, 13-37bp, 26-50bp) and performed a BLAST search of these probe sequence fragments. No probe sequemces or probe sequence fragments mapped to positions other than the single expected genomic target ($p < 0.05$).

3 Replication

3.1 Data description

We attempted replication of the 501 significant interactions from the discovery set using three independent cohorts; Fehrmann, EGCUT, and CHDWB. It was required that LD $r^2 < 0.1$ and $D'^2 < 0.1$ between interacting SNPs (as measured in the replication sample directly), and all nine genotype classes had at least 5 individuals present in order to proceed with statistical testing for replication in both datasets. We also excluded any putative SNPs that had discordant allele frequencies in any of the datasets. Details of the cohorts are as follows.

Fehrmann: $n = 1240$ The Fehrmann dataset[8] consists of peripheral blood samples of 1240 unrelated individuals from the United Kingdom and the Netherlands. Some of these individuals are patients, while others are healthy controls. Individuals were genotyped using the Illumina HumanHap300, Illumina HumanHap370CNV, and Illumina 610 Quad platforms. RNA levels were quantified using the Illumina HT-12 V3.0 platform.

EGCUT: $n = 891$ The Estonian Genome Center of the University of Tartu (EGCUT) study[9] consists of peripheral blood samples of 891 unrelated individ-

uals from Estonia. They were genotyped using the Illumina HumanHap370CNV platform. RNA levels were quantified using the Illumina HT-12 V3.0 platform.

CDHWB: $n = 139$ The Center for Health Discovery and Well Being (CDHWB) Study[10] is a population based cohort consisting of 139 individuals of European descent collected in Atlanta USA. Gene expression profiles were generated with Illumina HT-12 V3.0 arrays from peripheral blood collected from Tempus tubes that preserve RNA. Whole genome genotypes were measured using Illumina OmniQuad arrays. Due to the small sample size, most SNP pairs did not pass filtering in this dataset (20 SNP pairs remained) and so we have excluded it from the rest of the analysis.

3.2 Meta Analysis

The 4 *d.f.* interaction p-values for each independent replication dataset were calculated using the same statistical test as was performed in the discovery dataset. We then took the interaction p-values from EGCUT and Fehrmann and calculated a joint p-value using Fisher's method of combining p-values for a meta analysis as $-2\ln p_1 - 2\ln p_2 \sim \chi^2_{4d.f.}$. As in the discovery analysis, all gene expression levels were normalised using rank transformation to avoid skew or outliers in the distribution.[11]

3.3 Concordance of direction of effects

We used four methods to calculate the concordance of the direction of effects between the discovery and replication datasets.

Test 1 Is the most significant epistatic effect in the discovery set in the same direction as the same epistatic effect in the replication sets? We decomposed the genetic variance into 8 orthogonal effects, four of which are epistatic ($A \times A$, $A \times D$, $D \times A$, $D \times D$). The sign of the epistatic effect that had the largest variance in the discovery was recorded, and then was compared to the same epistatic effect in the two replication datasets (regardless of whether or not the same epistatic effect was the largest in the replication datasets). The probability of the sign being the same in one dataset is 1/2. The probability of the sign being the same in two is 1/4.

Test 2 Is the most significant epistatic effect in the discovery the same as the largest epistatic effect in the replication set with the sign being concordant. As in Test 1, but this time we required that the largest effect was the same in the discovery and the replication, and that they had the same sign (*e.g.* if the largest effect in the discovery is $A \times A$, with a positive effect, then concordance is achieved if the same is true in the replication). The probability of one replication dataset being concordant by chance is 1/8, and concordance in both is 1/64.

Test 3 Do the epistatic effects that are significant at nominal $p < 0.05$ in the discovery have the same direction of effect as in the replication? Here we count all the epistatic variance components in the discovery that have $p < 0.05$ (1133 amongst the 434 discovery SNP pairs, *i.e.* each SNP pair has at least 1 and at most 4 significant epistatic variance components). Then we compare the direction of the effect in the replication dataset. The probability of the sign being the same in one dataset for any one significant effect is 1/2. The probability of the sign being the same in two is 1/4.

Test 4 If we count how many of the 4 epistatic effects are concordant between the discovery and replication data for each interaction then is this significant from what we expect by chance? There can be either 0, 1, 2, 3 or 4 concordant signs at each interaction, each with expectation of $p = 1/16, 4/16, 6/16, 4/16, 1/16$ under the null, respectively. Observed counts are multinomially distributed, and we tested if the observed proportions were statistically different from the expected proportions using an approximation of the multinomial test.[12]

The probability of observing the number of concordant signs in tests 1-3 is calculated using a binomial test. All variance decompositions were calculated using the NOIA method.[13]

4 Effects of LD on detection and replication

The power to detect genetic effects, when the observed markers are in LD with the causal variants, is proportional to r^x. For additive effects $x = 2$, but for non-additive effects x is larger, *i.e.* $x = 4$ for dominance or $A \times A$, $x = 6$ for $A \times D$ or $D \times A$, and $x = 8$ for $D \times D$. Many biologically realistic GP maps may be comprised of all 8 variance components.[5]

This is important for both detection and for replication of epistasis. For detection, if the epistatic effect includes the $D \times D$ term then if the two causal variants are tagged by observed markers that are each in LD $r = 0.9$, then if the true variance is V_t then the observed variance V_o at the markers will be $0.9^8 V_t = 0.43 V_t$. Therefore, it is important to consider the sampling variation of $\hat{r}^x$ in a sample given some true population value of r.

4.1 Simulation 1

For some values of fixed population parameters, p_1 (minor allele frequency at observed marker), q_1 (minor allele frequency at causal variant), and r (LD between marker and causal variant), the expected haplotype frequencies are

$$h_{11} = r\sqrt{p_1 q_1 p_2 q_2} + p_1 q_1 \tag{5}$$
$$h_{12} = p_1 q_2 - r\sqrt{p_1 q_1 p_2 q_2} \tag{6}$$
$$h_{21} = p_2 q_1 - r\sqrt{p_1 q_1 p_2 q_2} \tag{7}$$
$$h_{22} = r\sqrt{p_1 q_1 p_2 q_2} + p_2 q_2 \tag{8}$$

where $p_2 = 1 - p_1$ and $q_2 = 1 - q_1$. For a range of population parameters we randomly sampled $2n$ haplotypes where the expected haplotype frequencies were $h_{11}, h_{12}, h_{21}, h_{22}$. From the sample haplotype frequencies we then calculated sample estimates of $\hat{r}$ where

$$\hat{r} = \frac{\hat{h}_{11} - \hat{p}_1\hat{q}_1}{\sqrt{\hat{p}_1\hat{q}_1\hat{p}_2\hat{q}_2}} \tag{9}$$

For each value of combination of the parameters p_1, q_1, r, n 1000 simulations were performed and the sampling mean and sampling standard deviation of $\hat{r}, \hat{r}^2, \hat{r}^4, \hat{r}^6, \hat{r}^8$ were recorded. It was observed that sampling variance increases for increasing x in $\hat{r}^x$.

4.2 Simulation 2

We assume that the discovery SNP pairs are ascertained (from a very large number of tests) have high $\hat{r}$ between observed SNPs and causal variants because otherwise power of detection would be low. We can hypothesis that the distribution of $\hat{r}$ in this ascertained sample will be a mixture of r that is high and r that is lower but with ascertained higher values from sampling. Therefore, we would expect those with truly high r to have a higher replication rate in independent datasets, and those with ascertained high $\hat{r}$ to have lower replication because resampling is unlikely to result in the same extreme ascertainment. To obtain empirical estimates of $\hat{r}$ in discovery and replication datasets we conducted the following simulation.

1. Using 1000 genomes data (phase 1, version 3, 379 European samples) we selected the 528,509 "markers" used in the original discovery analysis, plus 100,000 randomly chosen "causal variants" (CVs) with minor allele frequence > 0.05.

2. The 379 individuals were split into discovery (190) and replication (189) sets.

3. For each CV the marker with the maximum $\hat{r}^2_D$ from the marker panel was recorded in the discovery set. This marker was known as the "discovery marker" (DM).

4. The $\hat{r}^2_R$ for each CV/DM pair was then calculated in the replication set where the discovery LD was ascertained to be high, such that $\hat{r}^2_D > 0.9$.

We observed that there was an average decrease in $\hat{r}^x_R$ relative to $\hat{r}^x_D$, and that this decrease was larger with increasing x. We observed that $(\hat{r}^2_R - \hat{r}^2_D)/\hat{r}^2_D = 0.029$ whereas $(\hat{r}^8_R - \hat{r}^8_D)/\hat{r}^8_D = 0.092$. The average drop in in replication $\hat{r}^8$ was 3 times higher than the drop in $\hat{r}^2$.

4.3 Interpretation

Simulation 1 shows that sampling variance of r^x increases as x increases. Detection of epistatis is highly dependent upon high $\hat{r}$. Amongst the discovery SNPs there will be a mixture of interactions where observed SNPs are either in true high LD with causal variants, or will have highly inflated sample $\hat{r}^x$ compared to the population r^x. Simulation 2 shows that as x gets larger, the average decrease in $\hat{r}^x$ between discovery and replication becomes larger, likely to be a result of ascertained high $\hat{r}$ in the discovery and increased sampling variance with increasing x in the replication. These results demonstrate that if all else is equal, the impact of sampling variance of r alone will reduce the replication rate of epistatic effects compared to additive effects.

5 Additive and non-additive variance estimation

5.1 Fixed effects

To compare the relative contribution to the phenotypic variance of gene expression levels between additive and epistatic effects we are constrained by the problem that non-additive variance components for a phenotype cannot be calculated directly. Here, we only have SNP pairs that exceed a threshold of $p < 2.91 \times 10^{-16} = T_e$. A strong conclusion cannot be made about the genome-wide variance contribution, but we can compare the variance explained by SNP effects at this threshold for additive scans and epistatic scans.

In Powell *et al* 2012[1] an expression quantitative trait locus (eQTL) study was performed searching for additive effects in the same BSGS dataset as was used for the discovery here. Using the threshold T_e for the additive eQTL study, 453 of the 7339 probes analysed here had at least one significant additive effect. Assuming that the phenotypic variance for each of the probes is normalised to 1, the total phenotypic variance of all 7339 explained by the significant additive effects was 1.73%.

Following the same procedure, at the threshold T_e there were 238 gene expression probes with at least one significant pairwise epistatic interaction out of the 7339 tested. In total the proportion of the phenotypic variance explained by the epistatic effects at these SNP pairs was 0.25%.

5.2 Pedigree estimates

The gene expression levels for MBNL1, TMEM149, NAPRT1, TRAPPC5 and CAST are influenced by large *cis-trans* epistatic networks (eight interactions or more). Though it is not possible to orthogonally estimate the non-additive genetic variance for non-clonal populations, an approximation of a component of non-additive variance can be estimated using pedigree information. The BSGS data is comprised of some related individuals and standard quantitative genetic analysis was used to calculate the additive and dominance variance components for each gene expression phenotype in Powell *et al* 2013.[6] The dominance effect

is likely to capture additive × additive genetic variance plus some fraction of other epistatic variance components. We found that the aforementioned genes had dominance variance component estimates within the top 5% of all 17,994 gene expression probes that were analysed in Powell *et al* 2013.

6 Functional enrichment analysis

6.1 Tissue specific transcriptionally active regions

We employed a recently published method (`http://www.broadinstitute.org/mpg/epigwas/`)[14] that tests for cell-type-specific enrichment of active chromatin, measured through H3K4me3 chromatin marks[15] in regions surrounding the 731 SNPs that comprise the 501 discovery interactions. The exact method used to perform this analysis has been described previously.[16] Briefly, we tested the hypothesis that the 731 SNPs were more likely to be in transcriptionally active regions (as measured by chromatin marks) than a random set of SNPs selected from the same SNP chip. This hypothesis was tested for 34 cell types across four broad tissue types (haematopoietic, gastrointenstinal, musculoskeletal and endocrine, and brain).

6.2 Chromosome interactions

It has been shown[17] that different regions on different chromosomes or within chromosomes spatially colocalise within the cell. We shall refer to the colocalisation of two chromosome regions as a chromosome interaction. A map of pairwise chromosome interactions for K562 blood cell lines was recently produced,[18] and we hypothesised that part of the underlying biological mechanism behind some of the 501 epistatic interactions may arise from chromosome interactions. We found that 44 of the putative epistatic interactions were amongst SNPs that were within 5Mb of known chromosome interactions. This means that SNP A was no more than 2.5Mb from the focal point of the chromosome interaction on chromosome A, and SNP B was no more than 2.5Mb from the focal point on chromosome B.

We performed simulations to test how extreme the observation of 44 epistatic interactions overlapping with chromosome interactions is compared to chance. Chromosome interactions fall within functional genomic regions,[17,18] and the SNPs in our epistatic interactions are enriched for functional genomic regions. Therefore, we designed the simulations to ensure that the null distribution was of chromosome interactions between SNPs enriched for functional genomic regions but with no known epistatic interactions. To do this we used the 731 SNPs that form the 501 putative epistatic interactions and randomly shuffled them to create new sets of 501 pairs, disallowing any SNP combinations that were in the original set. Therefore, each new random set was enriched for functional regions but had no genetic interactions. We scanned the map of chromosome interactions for overlaps with the new sets and then repeated the random shuffling

process. We performed 1,000 such permutations to generate a null distribution of chromosome interaction overlaps.

We repeated this process, searching for overlaps within 1Mb, 250kb, and 10kb.

6.3 SNP colocalisation with genomic features

We tested for enrichment of genomic features for the 687 IndexSNPs that comprise the 434 epistatic interactions with data present in discovery and replication datasets. For each of the 687 IndexSNPs we calculated LD with all regional SNPs within a radius of 0.5Mb and kept all regional SNPs with LD $r^2 > 0.8$. We then cross-referenced the remaining regional SNPs with the annotated chromatin structure reference[19]) querying whether the regional SNPs fell in Predicted promoter region including TSS (TSS), Predicted promoter flanking region (PF), Predicted enhancer (E), Predicted weak enhancer or open chromatin cis regulatory element (WE), CTCF enriched element (CTCF), Predicted transcribed region (T), or Predicted Repressed or Low Activity region (R) positions. Therefore a particular IndexSNP might cover multiple genomic features through LD.

We then performed the whole querying process for each of the 528,509 SNPs present in the SNP chip used in the scan, and used the results from this second analysis to establish a null distribution for the expected proportion of SNPs for each genomic feature. We calculated p-values for enrichment of each of the seven genomic features independently, and for *cis*- and *trans*-SNPs separately, using a binomial test. For each genomic feature we used the expected proportion of SNPs as the expected probability of "success" (p). Here, a success is defined as an IndexSNP residing in a region that includes the genomic feature. The observed number of successes for each IndexSNP (k) out of the total count of IndexSNPs (n) was then modelled as $\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

6.4 Transcription factor enrichment

To test for enrichment of transcription factor binding sites (TFBS) we followed a procedure similar to that described in Section 6.3. For each of the 687 IndexSNPs we extracted regional SNPs as previously described. We then used the PWMEnrich package in Bioconductor (`http://www.bioconductor.org/packages/2.12/bioc/html/PWMEnrich.html`) to identify which TFBSs each of the regional SNPs for one IndexSNP falls in (within a radius of 250bp). Thus, the number of occurrences of a particular TFBS was counted for each IndexSNP. We used the "Threshold-free affinity" method for identifying TFBSs.[20]

We constructed a null distribution of expected TFBS occurrences based on the same null hypothesis as described in Section 6.3 - the probability of an IndexSNP covering a particular TFBS is identical to any of the 528,509 SNPs in the discovery SNP chip. To do this, we performed the same procedure for each SNP in the discovery SNP chip as was performed for each IndexSNP to obtain an expected probability of covering a particular TFBS. We then tested the

IndexSNPs for enrichment of each TFBS independently, and for *cis*- and *trans*-SNPs separately. p-values were obtained using Z-scores, calculated by using a normal approximation to the sum of binomial random variables representing motif hits along the sequence.[21]

6.5 Defining previously identified SNP associations

The discovery dataset (BSGS) had previously been analysed for additive and dominant marginal effects for all gene expression levels.[1,6] To define SNPs that had been previously detected to have effects for a particular gene expression level we used a significance threshold accounting for multiple testing across SNPs and expression probes, $T_m = 0.05/(528509 \times 7339) = 1.29 \times 10^{-11}$. From this, we found that only nine of the 501 discovery interactions had known main effects, 64 were between SNPs that had no known marginal effects, and 439 were between a SNP with a known marginal effect and a SNP with no known marginal effect.

7 References

References

[1] Powell, J. E. *et al.* The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. *PloS one* **7**, e35430 (2012).

[2] Medland, S. E. *et al.* Common variants in the trichohyalin gene are associated with straight hair in Europeans. *American journal of human genetics* **85**, 750–5 (2009).

[3] Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics (Oxford, England)* **23**, 1294–6 (2007).

[4] Hemani, G., Theocharidis, A., Wei, W. & Haley, C. EpiGPU: exhaustive pairwise epistasis scans parallelized on consumer level graphics cards. *Bioinformatics (Oxford, England)* **27**, 1462–5 (2011).

[5] Hemani, G., Knott, S. & Haley, C. An Evolutionary Perspective on Epistasis and the Missing Heritability. *PLoS Genetics* **9**, e1003295 (2013).

[6] Powell, J. E. *et al.* Congruence of Additive and Non-Additive Effects on Gene Expression Estimated from Pedigree and SNP Data. *PLoS Genetics* **9**, e1003502 (2013).

[7] Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics* **43**, 519–525 (2011).

[8] Fehrmann, R. S. N. *et al.* Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS genetics* **7**, e1002197 (2011).

[9] Metspalu, A. The Estonian Genome Project. *Drug Development Research* **62**, 97–101 (2004).

[10] Preininger, M. *et al.* Blood-informative transcripts define nine common axes of peripheral blood gene expression. *PLoS genetics* **9**, e1003362 (2013).

[11] Westra, H.-J. *et al.* MixupMapper: correcting sample mix-ups in genome-wide datasets increases power to detect small genetic effects. *Bioinformatics (Oxford, England)* **27**, 2104–11 (2011).

[12] Williams, D. A. Improved likelihood ratio tests for complete contingency tables. *Biometrika* **63**, 33–37 (1976).

[13] Alvarez-Castro, J., Le Rouzic, A., Carlborg, O., Álvarez Castro, J. M. & Carlborg, O. How to perform meaningful estimates of genetic effects. *PLoS Genetics* **4**, e1000062 (2008).

[14] Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nature genetics* **45**, 124–30 (2013).

[15] Koch, C. M. *et al.* The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome research* **17**, 691–707 (2007).

[16] Rietveld, C. A. *et al.* GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment. *Science* (2013).

[17] Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)* **326**, 289–93 (2009).

[18] Lan, X. *et al.* Integration of Hi-C and ChIP-seq data reveals distinct types of chromatin linkages. *Nucleic acids research* **40**, 7690–704 (2012).

[19] Hoffman, M., Buske, O., Wang, J. & Weng, Z. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nature Methods* **9**, 473–476 (2012).

[20] Stormo, G. DNA binding sites: representation and discovery. *Bioinformatics* **16**, 16–23 (2000).

[21] Ho Sui, S. J. *et al.* oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes. *Nucleic acids research* **33**, 3154–64 (2005).