**Supplemental Digital Content 1**

Review of the methods employed for analysis and statistical calculations of the pooled data. 1-12

Stage 1 - QC filtering of the frequency data in the pooled samples

Single nucleotide polymorphisms (SNPs) from each pool were prioritized for individual genotyping based on several different criteria outlined below. For all methods we excluded rare SNPs (true allele frequencies less than 5% in either the Caucasian population or predicted in the control pool sample) and the 5% of SNPs showing the highest variability as indicated by the size of the standard deviation among measures from the replicate assays. In addition, we excluded all markers used for the estimation of copy number variations in the genome. These filtering methods resulted in reduction of the SNPs available for statistical analysis from 620,901 (initial number of interrogated SNPs for each sample) to 447,283 SNPs.

Stage 2 - Ranking by test statistics

The test statistic was based on p-values estimated using the following test (Tcomb) which combines experimental and sampling errors:



This statistic combines:

1. chi-square statistic T1for testing differences between two proportions (allele frequencies) in cases and controls accounting for the sampling variance:

|  |  |
| --- | --- |
| A description... |  |

where is the mean of the allele frequencies over *nk* pool replicates, is the binomial sampling variance and *Nk* is number of controls and cases respectively (*k* = 1,2).

b) *Z-*statistics for testing the difference in mean allele frequencies between cases and controls:

|  |  |
| --- | --- |
| A description... |  |

Where is the square of the standard error due to experimental error. The use of this statistic (which follows an approximate normal distribution with one degree of freedom under the null hypothesis of equal allele frequencies for cases and controls) takes into account the two single available sources of error: sampling error and experimental error.

Whereas stage 2 analyzed markers independently, a multimarker approach may provide for smoothing of measurement noise and identify loci with lower odds ratio by leveraging linkage disequilibrium between adjacent SNPs. A sliding-window approach of mean test statistic values (as delineated in stage 2) was implemented for four different window sizes (5, 10, 20 and 30 consecutive SNPs) across all windows throughout the genome. The results of sliding-window analysis were compared with the results of the single SNP analysis of markers from stage 2. The sliding-window analysis also served as both a validation of the highest ranking SNPs selected from stage 2 (i.e. high ranking SNPs from stage 2 should be reflected in the high rank of different sized windows around its locus), as well as a tool for selecting additional SNPs with relative lower ranking in relation to the selected arbitrary threshold value of statistical significance (10-7 < p < 10-6).

References

1. Abraham R, Moskvina V, Sims R, Hollingworth P, Morgan A, Georgieva L, Dowzell K,

Cichon S, Hillmer AM, O'Donovan MC, Williams J, Owen MJ, Kirov G: A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. BMC Med Genomics 2008; 1: 44

2. Brown KM, Macgregor S, Montgomery GW, Craig DW, Zhao ZZ, Iyadurai K, Henders AK, Homer N, Campbell MJ, Stark M, Thomas S, Schmid H, Holland EA, Gillanders EM, Duffy DL, Maskiell JA, Jetann J, Ferguson M, Stephan DA, Cust AE, Whiteman D, Green A, Olsson H, Puig S, Ghiorzo P, Hansson J, Demenais F, Goldstein AM, Gruis NA, Elder DE, Bishop JN, Kefford RF, Giles GG, Armstrong BK, Aitken JF, Hopper JL, Martin NG, Trent JM, Mann GJ, Hayward NK: Common sequence variants on 20q11.22

confer melanoma susceptibility. Nat Genet 2008; 40: 838-40

3. Craig DW, Huentelman MJ, Hu-Lince D, Zismann VL, Kruer MC, Lee AM, Puffenberger EG, Pearson JM, Stephan DA: Identification of disease causing loci using an array-based genotyping approach on pooled DNA. BMC Genomics 2005; 6: 138

4. Craig DW, Millis MP, DiStefano JK: Genome-wide SNP genotyping study using pooled DNA to identify candidate markers mediating susceptibility to end-stage renal disease attributed to Type 1 diabetes. Diabet Med 2009; 26: 1090-8

5. Hanson RL, Craig DW, Millis MP, Yeatts KA, Kobes S, Pearson JV, Lee AM, Knowler WC, Nelson RG, Wolford JK: Identification of PVT1 as a candidate gene for end-stage renal disease in type 2 diabetes using a pooling-based genome-wide single nucleotide polymorphism association study. Diabetes 2007; 56: 975-83

6. Kirov G, Zaharieva I, Georgieva L, Moskvina V, Nikolov I, Cichon S, Hillmer A, Toncheva D, Owen MJ, O'Donovan MC: A genome-wide association study in 574 schizophrenia trios using DNA pooling. Mol Psychiatry 2009; 14: 796-803

7. Macgregor S, Visscher PM, Montgomery G: Analysis of pooled DNA samples on high

density arrays without prior knowledge of differential hybridization rates. Nucleic Acids Res 2006; 34: e55

8. Macgregor S, Zhao ZZ, Henders A, Nicholas MG, Montgomery GW, Visscher PM: Highly cost-efficient genome-wide association studies using DNA pools and dense SNP arrays. Nucleic Acids Res 2008; 36: e35

9. Moskvina V, Norton N, Williams N, Holmans P, Owen M, O'Donovan M: Streamlined

analysis of pooled genotype data in SNP-based association studies. Genet Epidemiol 2005; 28: 273-82

10. Pearson JV, Huentelman MJ, Halperin RF, Tembe WD, Melquist S, Homer N, Brun M, Szelinger S, Coon KD, Zismann VL, Webster JA, Beach T, Sando SB, Aasly JO, Heun R, Jessen F, Kolsch H, Tsolaki M, Daniilidou M, Reiman EM, Papassotiropoulos A, Hutton ML, Stephan DA, Craig DW: Identification of the genetic basis for complex disorders by use of pooling-based genomewide singlenucleotide-polymorphism association studies. Am J Hum Genet 2007; 80: 126-39

11. Steer S, Abkevich V, Gutin A, Cordell HJ, Gendall KL, Merriman ME, Rodger RA, Rowley KA, Chapman P, Gow P, Harrison AA, Highton J, Jones PB, O'Donnell J, Stamp L, Fitzgerald L, Iliev D, Kouzmine A, Tran T, Skolnick MH, Timms KM, Lanchbury JS, Merriman TR: Genomic DNA pooling for whole-genome association scans in complex disease: empirical demonstration of efficacy in rheumatoid arthritis. Genes Immun 2007; 8: 57-68

12. Sebastiani P, Zhao Z, Abad-Grau MM, Riva A, Hartley SW, Sedgewick AE, Doria A,

Montano M, Melista E, Terry D, Perls TT, Steinberg MH, Baldwin CT: A hierarchical and modular approach to the discovery of robust associations in genome-wide association studies from pooled DNA samples. BMC Genet 2008; 9: 6