Hypothesis-free identification of modulators of genetic risk factors

**Supplementary Tables**

Table S1. Number of samples before and after QC

Table S2. Replication of BBMRI and Geuvadis eQTLs

Table S3. Total number of primary cis-eQTLs

Table S4. GWAVA functional annotation of eSNPs (a subset)

Table S5. Enhancer enrichment (a subset)

Table S6. eQTLs associated with diseases and complex traits

Table S7. Top 100 proxy genes and corresponding eQTLs for the top 10 interaction modules

Table S8. Pathway enrichment analysis results for the cell type specific eQTL genes of the top 10 interaction modules for gene, exon and exon ratio level analysis

Table S9. TF enrichment analysis of significant context-specific eQTLs

Table S10. Correlation of eQTL interaction z-scores with cell-type-specific interaction analysis

Table S11. Overlap of eQTLs significantly interacting with identified modules and those significantly interacting with measured cell counts.

Table S12. Replication of interactions in Geuvadis

Table S13. GWAS hits for eQTLs significantly interacting with top 10 modules

Table S14. eQTL enrichment analysis results

Table S15. Enrichment of covariate genes in stimulation-specific DE genes

Table S16. Replication of interactions not falling into the top 10 modules in Geuvadis

Table S17. General mapping statistics

Table S18. The set of trait/disease-associated variants used for eQTL annotation

# Table S1. Number of samples before and after QC

|  |  |  |  |
| --- | --- | --- | --- |
| **Biobank** | **Original number of samples** | **Number of samples after QC** | **Number of samples with genotypes** |
| **LLD** | 630 | 628 | 626 |
| **CODAM** | 191 | 188 | 184 |
| **RS** | 658 | 652 | 652 |
| **LLS** | 697 | 664 | 654 |
| **Total** | **2,176** | **2,132** | **2,116** |

# Table S2. Replication of BBMRI and Geuvadis eQTLs

To test the overlap between two datasets, we took significant eQTLs (top SNP per gene) from one dataset (column “# eQTLs”) and tested these eQTLs in the other dataset (the number of tested eQTLs is given in the “tested” column). We then checked how many eQTLs out of all tested were significant (FDR < 0.05) in the other dataset (column “replicated” and “%replicated”) and for how many of the significant eQTLs the allelic direction was opposite.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **# eQTLs** | **Tested** | **Replicated** | **% Replicated** | **% Opposite allelic directions** |
| **Westra et al. --> BIOS** |
| gene-level | 764,497 | 716,053 | 598,075 | 84% | 10% |
| exon-level | 590,673 | 581,847 | 475,121 | 82% | 6% |
| **BIOS --> Geuvadis** |
| gene-level | 23,060 | 19,946 | 7,630 | 38% | 10% |
| exon-level | 105,207 | 102,824 | 35,307 | 34% | 8% |
| exon ratio level | 37,713 | 36,552 | 9,066 | 25% | 3% |
| polyA ratio level | 3,449 | 3,229 | 1,161 | 36% | 2% |
| **Geuvadis --> BIOS** |
| gene-level | 8,905 | 8,459 | 6,505 | 77% | 12% |
| exon-level | 43,062 | 39,237 | 28,038 | 71% | 11% |
| exon ratio level | 7,872 | 7,473 | 5,650 | 76% | 3% |
| polyA ratio level | 913 | 826 | 698 | 85% | 2% |

# Table S3. Total number of primary cis-eQTLs

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Features tested | SNPs tested | Unique *cis*-regulated features | Unique genes with *cis*-regulated features | Independent eQTLs |
| Gene | 57,220 | 5,913,484 | 23,060 | 23,060 | 44,720 |
| Exon | 282,474 | 5,952,925 | 105,207 | 21,888 | 171,904 |
| Exon ratio | 275,779 | 5,939,431 | 108,145 | 9,777 | 46,863 |
| PolyA ratio | 27,409 | 3,820,471 | 10,137 | 2,322 | 4,002 |

# Table S4. GWAVA functional annotation of eSNPs (a subset)

The table contains the average scores (for continuous scores) or counts (for count data) of GWAVA annotation for background SNPs (SNPs not in LD with the eSNP, matched by MAF and position, for details see Methods section), for real eQTL SNPs (top SNP per gene in gene level eQTL mapping) and a p-value showing the significance of the difference between the two SNP groups. In case of count data we tested the difference using Fisher’s exact test, in case of continuous data we used Wilcoxon rank-sum test.

The full table supplied separately

|  |  |  |  |
| --- | --- | --- | --- |
| Annotation | Background SNPs | eSNPs | p-value(Fisher or Wilcox) |
| TSS score | 0.2419 | 0.2543 | 1.62 X 10-10 |
| DNase I | 5,834 | 7,783 | 5.74 X 10-26 |
| FAIRE | 3,973 | 5,499 | 2.69 X 10-24 |
| H3K27ac | 5,767 | 7,526 | 1.42 X 10-18 |
| H3K27me3 | 5,629 | 6,225 | 8.01 X 10-03 |
| H3K36me3 | 6,093 | 7,691 | 3.00 X 10-11 |
| H3K4me1 | 8,381 | 10,323 | 2.53 X 10-10 |
| H3K4me2 | 5,906 | 7,606 | 1.58 X 10-15 |
| H3K4me3 | 4,694 | 6,230 | 4.76 X 10-18 |
| H3K79me2 | 3,876 | 5,294 | 2.46 X 10-20 |
| H3K9ac | 5,482 | 7,077 | 1.33 X 10-14 |
| H4K20me1 | 1,943 | 2,526 | 2.07 X 10-05 |
| bound\_motifs | 1,051 | 1,351 | 6.04 X 10-03 |

# Table S5. Enhancer enrichment (a subset)

The table contains the overlap of background SNPs (SNPs not in LD with the eSNP, matched by MAF and position, for details see Methods section) and real eSNPs with enhancers identified by the FANTOM5 project (the bed files are provided on this site: <http://enhancer.binf.ku.dk/presets/>). To test the difference between background and real SNPs we used Wilcoxon rank-sum test.

The full table supplied separately

|  |  |  |  |
| --- | --- | --- | --- |
|  | Background SNPs | eSNPs | p-value |
| Robust enhancers | 150 | 264 | 5.40 X 10-12 |
| Permissive enhancers | 158 | 270 | 1.73 X 10-11 |
| Monocyte DE enhancers | 51 | 129 | 1.87 X 10-11 |
| Neutrophil DE enhancers | 10 | 42 | 8.74 X 10-07 |
| Blood DE enhancers | 12 | 42 | 6.58 X 10-06 |
| T cell DE enhancers | 32 | 69 | 2.10 X 10-05 |
| B lymphocyte DE enhancers | 6 | 19 | 0.004 |
| Macrophage DE enhancers | 16 | 29 | 0.023 |
| Basophil DE enhancers | 5 | 13 | 0.035 |

# Table S6. eQTLs associated with diseases and complex traits

Supplied separately

# Table S7. Top 100 proxy genes and corresponding eQTLs for the top 10 interaction modules

Supplied separately

# Table S8. Pathway enrichment analysis results for the cell type specific eQTL genes of the top 10 interaction modules for gene, exon and exon ratio level analysis

Supplied separately

# Table S9. TF enrichment analysis of significant context-specific eQTLs

Supplied separately

# Table S10. Correlation of eQTL interaction z-scores with cell-type-specific interaction analysis

We ran an interaction analysis with each of the 5 measured cell percentages in a similar way as using proxy genes. For each of the top 10 modules (columns) the interaction z-scores across all eQTLs was correlated with the corresponding eQTL z-scores from the interaction with cell percentages (rows). Correlation was calculated using Spearman's rank correlation.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Actual** **cell type** | **Neutro-****phils 1** | **CD4+ T-cells** | **NK cells /CD8+ T-Cells** | **Erythrocytes** | **Monocytes /****Macrophages** | **Unkown** | **Type 1 inter-feron** | **Neutro-phils 2** | **B-cells** | **Eosino-phils** |
| Neutrophils | 0.80 | -0.11 | -0.17 | 0.05 | -0.02 | 0.03 | 0.02 | 0.17 | -0.09 | -0.12 |
| Lymphocytes | -0.79 | 0.17 | 0.28 | -0.01 | -0.06 | 0.01 | 0.02 | -0.13 | 0.14 | 0.01 |
| Eosinophils | -0.13 | 0.04 | -0.06 | -0.09 | 0.04 | -0.07 | -0.09 | -0.19 | -0.04 | 0.79 |
| Basophils | -0.15 | 0.01 | -0.03 | -0.07 | 0.01 | 0.00 | -0.10 | -0.16 | -0.07 | 0.15 |
| Monocytes | -0.14 | -0.19 | -0.23 | -0.10 | 0.27 | -0.09 | -0.10 | -0.02 | 0.00 | 0.03 |

# Table S11. Overlap of eQTLs significantly interacting with identified modules and those significantly interacting with measured cell counts.

In order to validate the identified cell-type-dependent eQTLs, we ran a similar interaction analysis testing measured cell type percentages as covariates. We calculated the overlap between the eQTLs significantly interacting with the 10 modules described in the paper (rows) and the eQTLs significantly interacting with measured cell counts (columns). The numbers represent the overlap in eQTLs, the subsequent number in brackets shows the percentage of overlapping eQTLs acting in the same interaction direction (increased covariate expression is associated with a stronger eQTL effect and vice versa).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Neutrophils** | **Lymphocytes**  | **Eosinophils** | **Basophils** | **Monocytes** |
| **Neutrophils 1** | 590 (99%) | 579 (1%) | 30 (10%) | 8 (0%) | 34 (18%) |
| **CD4+ T-cells** | 52 (42%) | 59 (61%) | 11 (55%) | 4 (25%) | 11 (0%) |
| **NK/CD8+ T-cell** | 62 (5%) | 75 (97%) | 6 (50%) | 1 (0%) | 11 (0%) |
| **Erythrocytes** | 9 (33%) | 11 (64%) | 2 (0%) | 1 (0%) | 1 (0%) |
| **Monocytes / Macrophages** | 31 (29%) | 37 (59%) | 5 (100%) | 1 (100%) | 20 (100%) |
| **Module 6** | 32 (59%) | 32 (47%) | 3 (67%) | 0 (0%) | 8 (62%) |
| **IFN1** | 16 (50%) | 16 (14%) | 5 (0%) | 0 (0%) | 6 (0%) |
| **Neutrophils 2** | 30 (100%) | 27 (0%) | 13 (0%) | 1 (0%) | 5 (0%) |
| **B-cells** | 9 (0%) | 11 (100%) | 2 (0%) | 0 (0%) | 1 (100%) |
| **Eosinophils** | 13 (0%) | 10 (50%) | 80 (99%) | 7 (100%) | 2 (100%) |

# Table S12. Replication of interactions in Geuvadis

We checked the replication rate and percentage of opposite allelic directions for eQTLs significantly positively interacting (eQTL effect increases when proxy gene expression increases) with neutrophil-specific modules (modules 1 and 8), significantly negatively interacting with neutrophil-specific modules (row Neg. neutrophils), and so on for lymphoid-specific modules (modules 2, 3 and 9) and other modules. We used Fisher exact test to calculate the significance of the difference in the number of replicated vs non-replicated eQTLs and in the number of eQTLs replicating with the same vs opposite allelic direction. As a background we used all eQTLs not significant for the module, but significant in any other modules.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Module** | **Tested significant interaction** | **Replicated** | **Replicated (opposite direction)** | **Fisher p-value replication** | **Fisher p-value direction** |
| Neutrophils | 758 | 39.31% | 27.18% | 0.3044 | 0.0002 |
| Neg. neutrophils | 181 | 50.28% | 15.38% | 0.0071 | 0.2689 |
| Lymphoid | 294 | 46.26% | 16.18% | 0.0465 | 0.2424 |
| Neg. lymphoid | 346 | 41.33% | 14.69% | 0.8571 | 0.0852 |
| Other modules | 534 | 39.70% | 16.51% | 0.5727 | 0.1635 |

# Table S13. GWAS hits for eQTLs significantly interacting with top 10 modules

Supplied separately

# Table S14. eQTL enrichment analysis results

Supplied separately

# Table S15. Enrichment of covariate genes in stimulation-specific DE genes

Enrichment analysis was performed on covariate genes from each module (module 7 is split based on the sign of covariate gene correlation with *SP140*) for four stimulation-specific DE genes identified in Caliskan et al.1 (Rhinovirus) and Fairfax et al.2 (Interferon gamma (IFNγ), lipopolisaccaride after 2 hours (LPS2), lipopolisaccaride after 24 hours (LPS24)). To determine the significance Fisher exact test was used (significant enrichment is shown in blue).

|  |  |  |  |
| --- | --- | --- | --- |
| **Module** | **Module name** | **Function** | **Enrichment for DE genes (Fisher p-values)** |
|  | **Rhinovirus** | **IFNγ** | **LPS2** | **LPS24** |
| 1 | Neutrophils 1 | Detection of bacterium | 0.81 | 0.09 | 0.01 | 8.38E-05 |
| 2 | CD4+ T-Cells | T cell selection | 0.82 | 0.75 | 0.21 | 0.27 |
| 3 | NK cells/CD8+ T-Cells | Cellular defense response | 0.84 | 0.65 | 0.59 | 0.43 |
| 4 | Erythrocytes | Hemoglobin metabolic process | 0.74 | 0.57 | 0.71 | 0.57 |
| 5 | Monocytes/Macrophages | Defense response to virus | 5.84E-11 | 6.31E-08 | 5.87E-06 | 0.51 |
| 6 | Unknown | Unknown | 9.40E-06 | 1.14E-07 | 0.08 | 0.26 |
| 7+ | Interferon response | Type 1 interferon response | 1.14E-09 | 3.33E-05 | 0.94 | 0.90 |
| 7- | Anti-bacterial  | Anti-bacterial  | 0.99 | 8.72E-4 | 3.23E-3 | 0.02 |
| 8 | Neutrophils 2 | Detection of bacterium | 0.22 | 0.02 | 0.62 | 5.16E-4 |
| 9 | B-cells | B cell receptor signaling pathway | 0.78 | 0.05 | 0.22 | 0.43 |
| 10 | Eosinophils | Regulation of myeloid leukocyte mediated immunity | 0.31 | 2.45E-4 | 0.19 | 0.01 |

1. Çalışkan, M., Baker, S. W., Gilad, Y. & Ober, C. Host genetic variation influences gene expression response to rhinovirus infection. PLoS Genet. 11, e1005111 (2015).
2. Fairfax, B. P. et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science 343, 1246949 (2014).

# Table S16. Replication of interactions not falling into the top 10 modules in Geuvadis

Supplied separately

# Table S17. General mapping statistics

|  |  |
| --- | --- |
| **Sequence Characteristics** | **Median (interquartile range)** |
| Number of reads passing QC (mln\*) | 39.5 (35.14-43.97) |
| Percentage of aligned reads | 92.16 (91.41-92.63) |
| Percentage of aligned reads mapping to annotated exons | 88.66 (87.54-89.66) |
| Number of genes detected (>0 count) per sample | 30,920 (30,200-31,610) |
| GC percentage\*\* | 52.13 (51.58-52.74) |

\*mln, million

\*\*GC percentage was estimated on raw fastq files.

# Table S18. The set of trait/disease-associated variants used for eQTL annotation

Supplied separately