DeepWAS: Directly integrating regulatory information into GWAS using deep learning supports master regulator MEF2C as risk factor for major depressive disorder

Gökcen Eraslan, Janine Arloth, Jade Martins, Stella Iurato, Darina Czamara, Elisabeth B. Binder, Fabian J. Theis, Nikola S. Mueller

Supplementary materials

Content:

Supplementary Figures

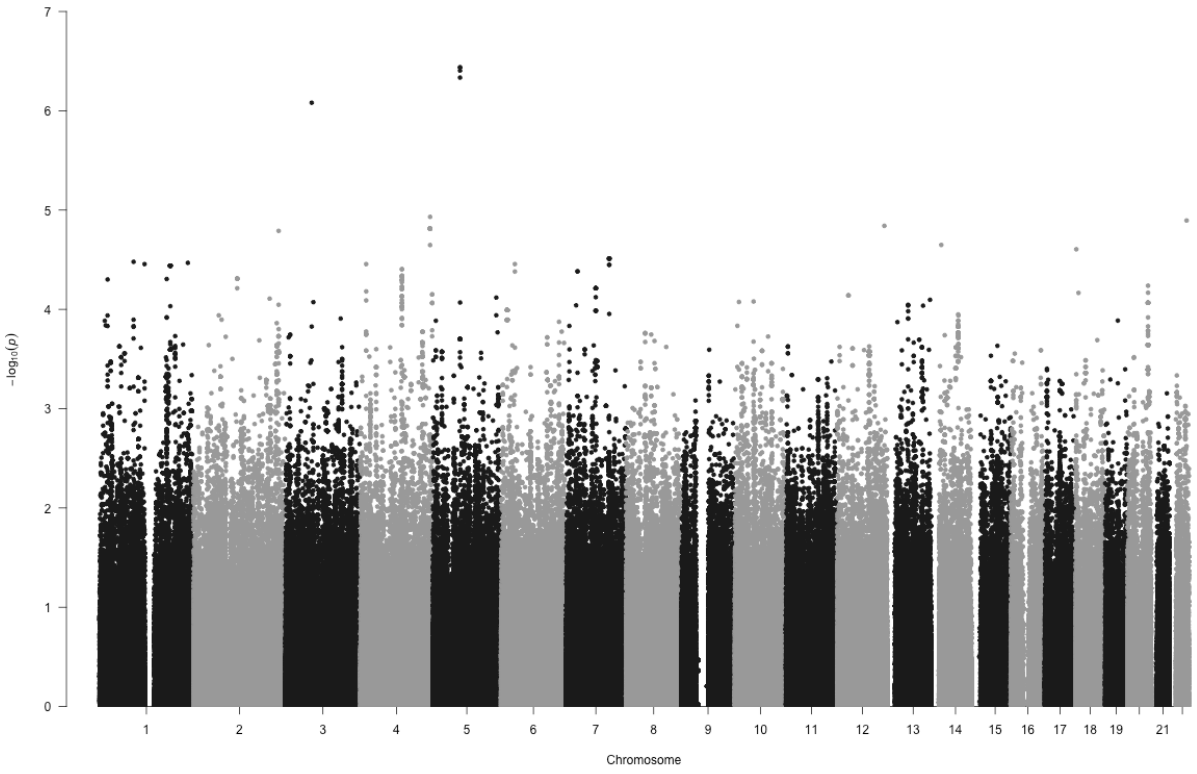

Figure S1: Genome-wide representation of MDD associations in the GSK data set.

Manhattan plot shows the strength of evidence for association (p value) in the recMDD cohort. Each variant is shown as a dot, with alternating shades according to chromosome.

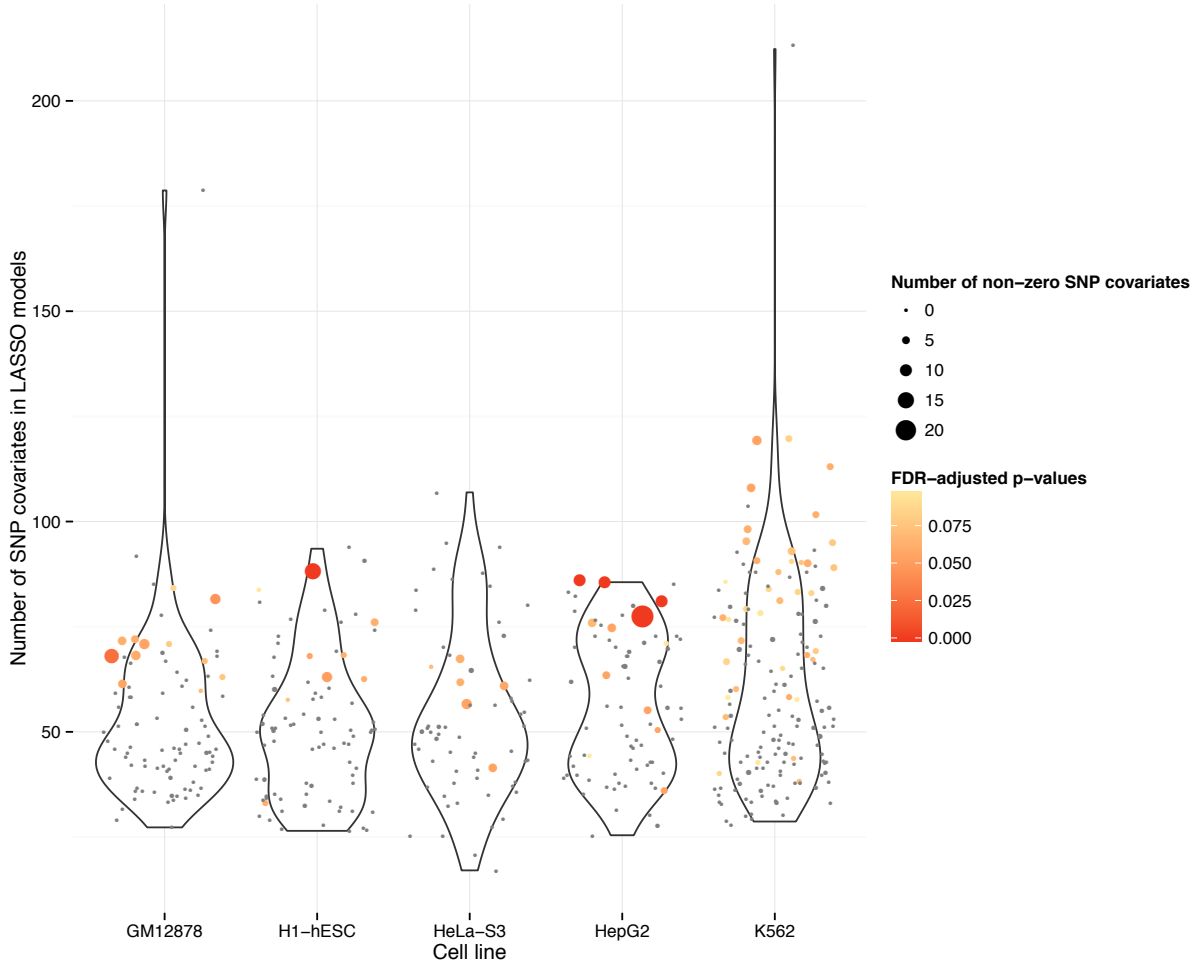

Figure S2: Number of SNP features in LASSO models for 5 cell lines

Each point represents a LASSO model for 5 cell lines and various regulatory elements. Color code shows the significance of each model whereas y-axis shows the number of SNP features in each model. The size of the points represents number of SNPs that are selected in LASSO models i.e. SNP predictors with non-zero regression coefficients.

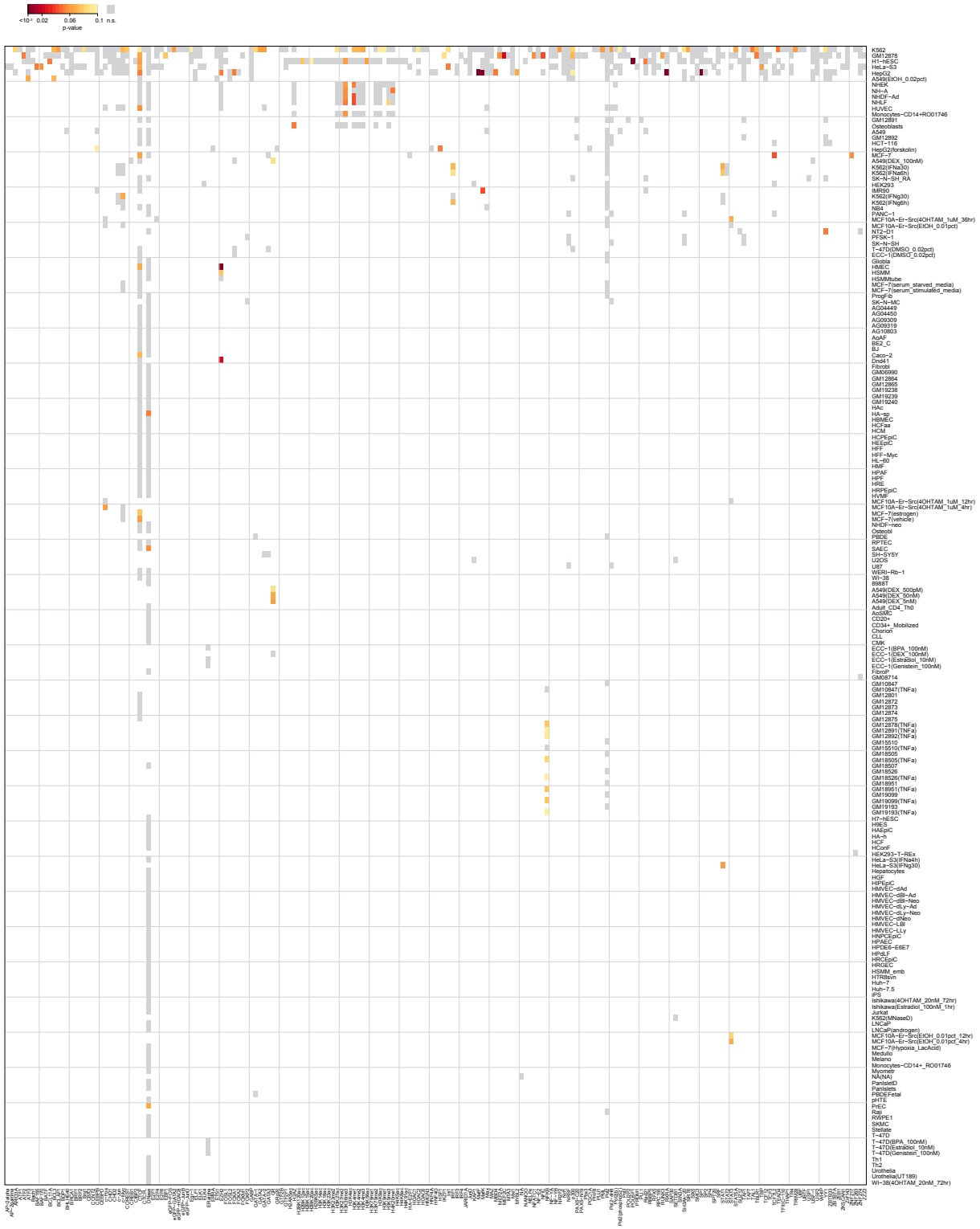

Figure S3: FDR-corrected p-values of LASSO models of all functional units.

Figure S4: Enrichment of regulatory SNPs in promoters and enhancers identified by 15-state ChromHMM model in various tissues.

x-axis and y-axis represent Roadmap epigenome IDs and FDR-adjusted p-values, respectively. 0.1 FDR cutoff is shown as a horizontal line. Color coding indicates regulatory elements.

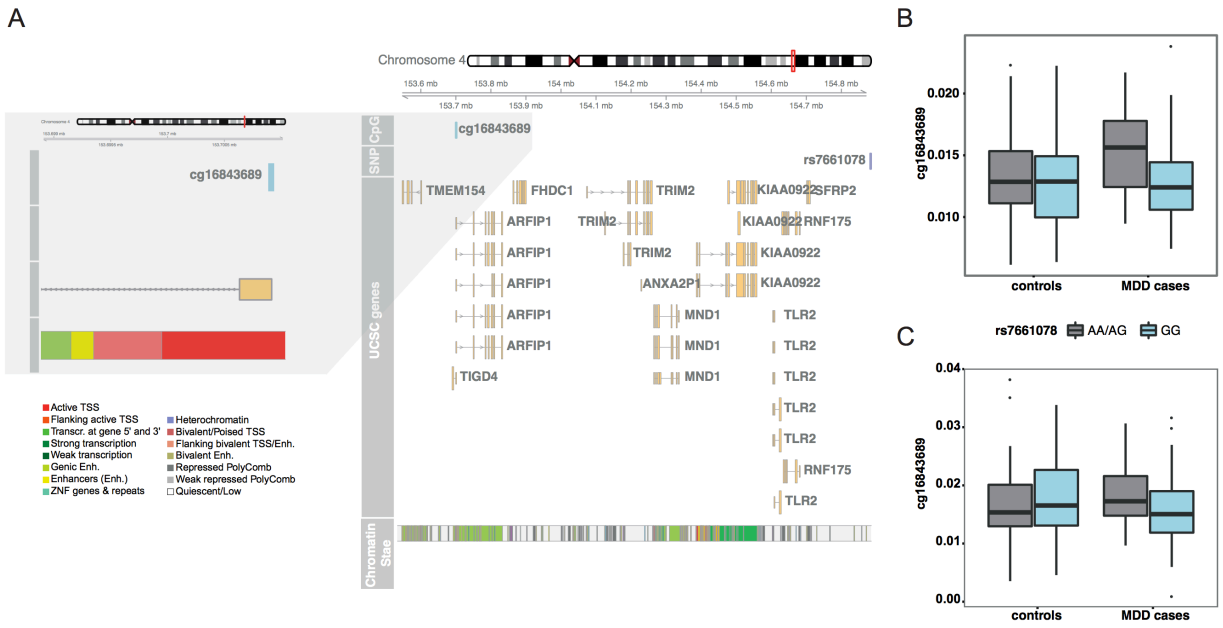

Figure S5: Disease specific meQTL results for the blood deepSNP rs7661078

(A) Overview of the rs7661078 meQTL locus on chromosome 5. Top panel, ideogram for chromosome 4. A red box isolates the region shown (enlarged) in the bottom panels. Bottom panels: Location of CpG cg16843689 that is significantly associated with rs7661078, SNP position, genes nearby (based on USCS knownGene annotation) and annotation of chromatin states (based on GM12878 15-state core ChromHMM model). **(B)** Association of rs7661078 (dominant SNP model) and cg16843689 stratified by the disease status within the recMDD cohort. **(C)** Validation of this case specific meQTL in the MPIP cohort. Y-axis in B and C shows the DNAm levels (beta values) and the x-axis represents the disease-status.

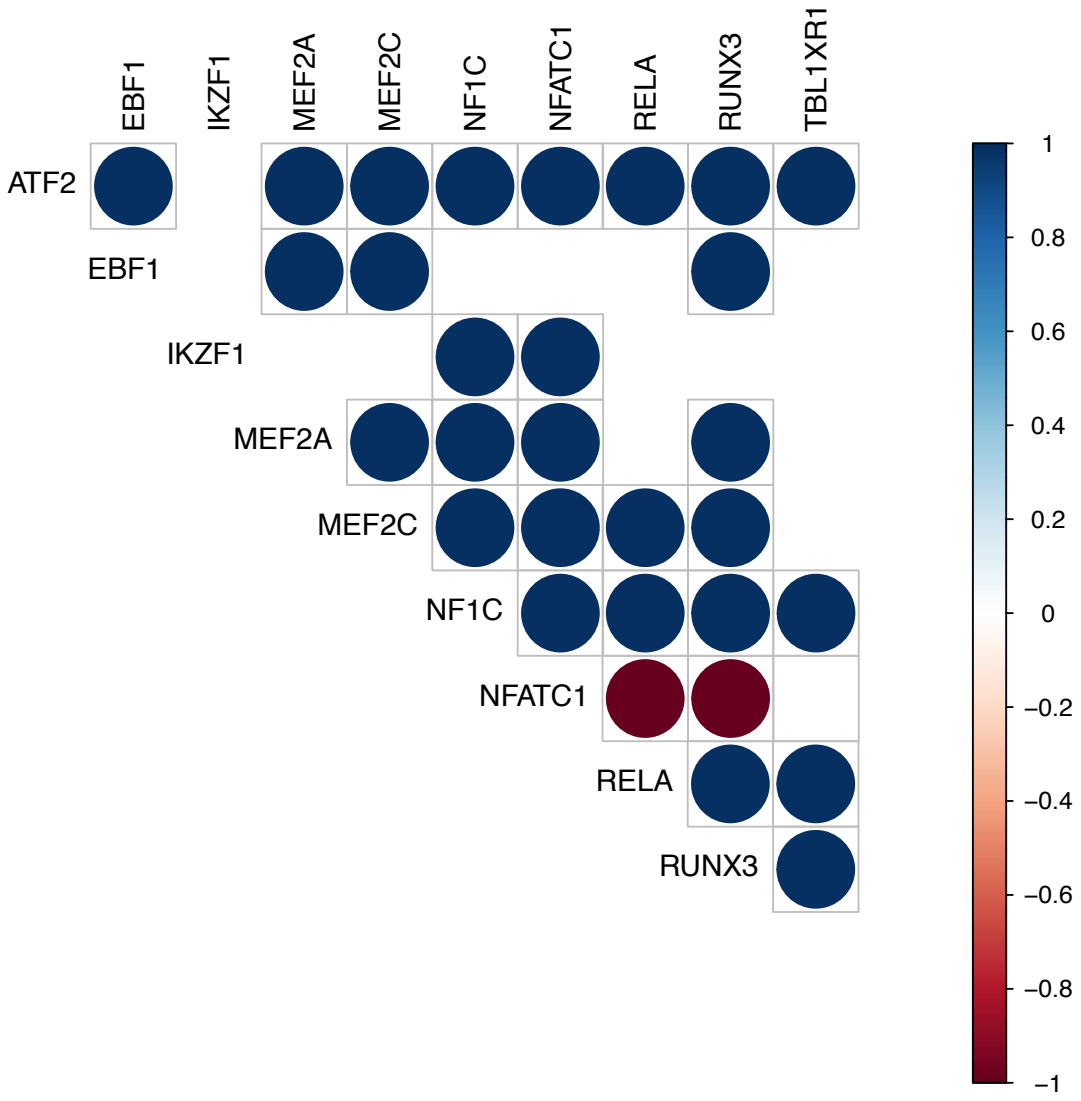

Figure S6: TF-TF associations from ReMap project

Overlap between ChIP-seq peaks of various transcription factors. ReMap annotation tool was used with 90% minimum overlap constraint on both BED files. e-values provided by the annotation tool of ReMap is binarized using a cutoff of 10^{-3}. Note that; NF1C, RELA and TBLXR1 in ReMap dataset correspond to the transcription factors NFIC, NFKB and TBL1 in our study.

Supplementary Tables and Table Descriptions

CpG	SNP	Chr	SNP Position	SNP Location	Genes nearby SNP	CpG Position	CpG Location	Genes nearby CpG	P value	Validation
cg02157078	rs486755	1	114884179	intergenic	SYT6,TRIM33	114888769	intergenic	SYT6,TRIM33	0.00176	yes
cg08169020	rs6573943	14	70699464	intergenic	SLC8A3,ADAM21P1	69256840	exonic	ZFP36L1	0.0266	no
cg09535027	rs7939217	11	85280702	intronic	DLG2	86432027	intergenic	ME3,PRSS23	0.031	same direction
cg11247129	rs2150175	6	97269213	intronic	GPR63	97457432	intronic	KLHL32	0.0397	no
cg16843689	rs7661078	4	154883600	intergenic	SFRP2,DCHS2	153700888	UTR5	TIGD4	0.00173	yes

Table S1: Significant blood deepSNPs associated with MDD-specific methylation changes

Significant blood deepSNPs representing a case-specific meQTL (recMDD cohort). P values are corrected according to Bonferroni for the number of tested CpGs per SNP.

- CpG: CpG identitidfier (Illumina 450k)
- SNP: SNP identifier
- Chr: Chromosome number
- SNP/CpG position: Genomic position of SNP/CpG in hg19 assembly
- SNP/CpG location: The most important SNP/CpG location according to ANNOVAR is listed
- Genes nearby SNP/CpG: If the SNP is exonic/intronic/ncRNA, the name of the gene is listed, if not, the two neighboring genes are listed.
- P value: Bonferroni corrected p values
- Validation: Validation of association in MPIP cohort.

CpG	Probe Id	Chr	CpG Position	Genes nearby CpG	CpG location	P start	P end	P gene	P value
cg19235974	ILMN_1696317	12	1063148	RAD52,ERC1	intergenic	1901374	1901423	CACNA2D4	0.048
cg19235974	ILMN_1859030	12	1063148	RAD52,ERC1	intergenic	2329963	2330012	CACNA1C	0.05
cg21290162	ILMN_2228180	8	9741911	TNKS,LINC00599	intergenic	10286154	10286205	MSRA	0.031

Table S2: Association of CpG methylation of the deepSNP-CpG pairs and gene expression

Table lists CpG methylation sites associated to MDD deepSNPs on matched gene expression levels from the MPIP cohort (eQTMs). P values are corrected according to Bonferroni for the number of tested transcripts per CpG.

- CpG: CpG identitidfier (Illumina 450k)
- Probe Id: gene expression probe identifier (Human HT-12 v3 and v4)
- Chr: Chromosome number
- CpG position: Genomic position of CpG in hg19 assembly
- CpG location: The most important CpG location according to ANNOVAR is listed
- Genes nearby CpG: If the SNP is exonic/intronic/ncRNA, the name of the gene is listed, if not, the two neighboring genes are listed.
- P start/P end: transcript position in hg19 assembly
- P gene: transcript gene
- P value: Bonferroni corrected p values

Probe Id	SNP	Chr	SNP Position	SNP Location	Genes nearby SNP	P start	P end	P gene	P value
ILMN_2228180	rs12541159	8	9769098	intergenic	MIR124-1,MSRA	10286154	10286205	MSRA	0.0225
ILMN_1765332	rs1868881	11	57901176	intronic	OR9Q1	57296017	57296066	TIMM10	0.00276
ILMN_1794825	rs4646797	17	19560368	intronic	ALDH3A2	19580356	19580406	ALDH3A2	1.57e-06

Table S3: Significant blood deepSNPs representing an eQTL

Significant blood deepSNPs representing an eQTL in th MPIP cohort. P values are corrected according to Bonferroni for the number of tested transcripts per SNP.

- Probe Id: gene expression probe identifier (Human HT-12 v3 and v4)
- SNP: SNP identifier
- Chr: Chromosome number
- SNPposition: Genomic position of SNP in hg19 assembly
- SNP location: The most important SNP location according to ANNOVAR is listed
- Genes nearby SNP: If the SNP is exonic/intronic/ncRNA, the name of the gene is listed, if not, the two neighboring genes are listed.
- P start/P end: transcript position in hg19 assembly
- P gene: transcript gene
- P value: Bonferroni corrected p values