Supplemental Methods

SNP associations were placed in context of previous GWAS using Phenoscanner, a variant-phenotype comprehensive database of large GWAS, which includes results from the NHGRI-EBI GWAS catalogue, the UK Biobank, NIH Genome-Wide Repository of Associations between SNPs and Phenotypes and publicly available summary statistics from more than 150 published GWAS.[1] Publicly available results for all phenotypes were filtered at $P < 5 \times 10^{-8}$ and the R statistical software package phenoscanner was used to download the data for all significant variants.[2]

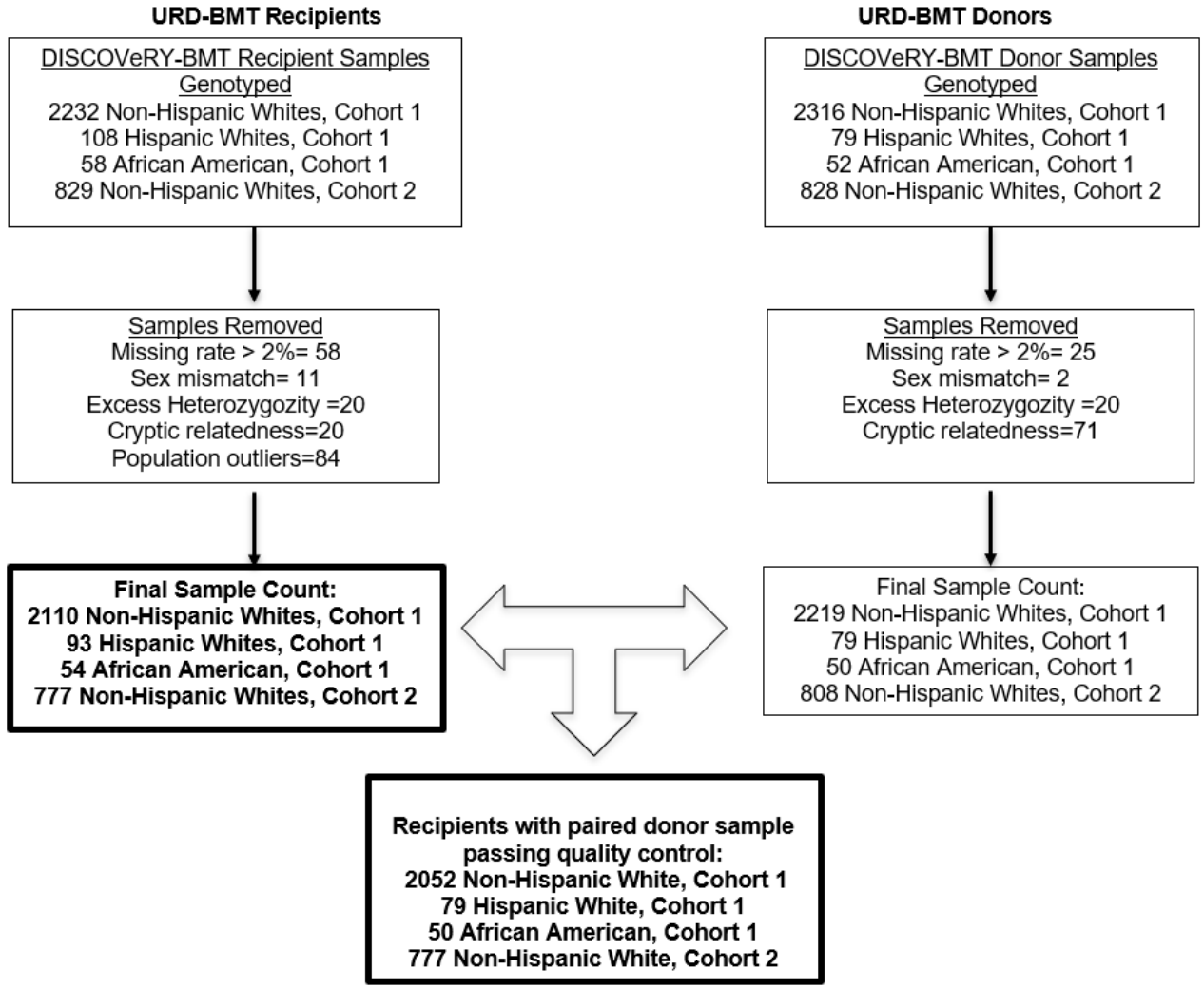
SNP data were integrated with Hi-C cell line data and visualized using Hi-C plots to determine significant regions of interaction.[3] Chromatin state data based on 25-state Imputation Based Chromatin State Model across 24 Blood, T-cell, HSC and B-cell lines was downloaded from the Roadmap Epigenomics project.[4] Figures including chromatin state information, eQTL associations, CADD scores, RegulomeDB scores and previous GWAS hits were constructed using the R Bioconductor package gviz and FUMA.[5-7]

All significant regions were further examined in the context of topologically associating domains (TADs).[8,9] TADs are self-interacting megabase-long genomic regions that contain interacting *cis*-regulatory elements and target genes; TADs often contain clusters of genes that interact with each other. TAD boundaries are the regions between TADs that restrict interactions of *cis*-regulatory sequences (such as enhancers) to target genes. Recent evidence shows that GWAS of genetic variation in TAD boundaries contributes to complex trait heritability, especially for immunologic, hematologic, and metabolic traits.[9] We used TAD maps across 37 diverse cell types to determine if associated loci were in TADs or TAD boundaries.[10] Lastly, Multi-marker Analysis of GenoMic Annotation (MAGMA) v8 as implemented in FUMA was used to perform genome-wide gene-level associations with OS, TRM and DRM.[11]

1. Kamat MA, Blackshaw JA, Young R, *et al.*, PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics* 2019; **35**: 4851-4853.
2. Staley, JR, Blackshaw J, Kamat MA, *et al.*, PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* 2016; **32**: 3207-3209.
3. Yu Y, Ouyang Y, Yao W. shinyCircos: an R/Shiny application for interactive creation of Circos plot. *Bioinformatics* 2018; **34:** 1229-1231.
4. Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature* 2015; **518:** 317-330.

5. Cairns J., Freire-Pritchett P, Wingett SW, *et al.*, CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol* 2016; **17:** 127.
6. Mifsud B, Tavares-Cadete F, Young AN, *et al.*, Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet* 2015; **47:** 598-606.
7. Spurrell CH, Dickel DE, Visel A. The Ties That Bind: Mapping the Dynamic Enhancer-Promoter Interactome. *Cell* 2016; **167:** 1163-1166.
8. Dixon JR, Gorkin DU, Ren B. Chromatin Domains: The Unit of Chromosome Organization. *Mol Cell* 2016; **62:** 668-680.
9. MacArthur E, Capra J. Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. Am J Hum Genet 2021; **108**: P269-283.
10. Dixon JR, Selvaraj S, Yue F, *et al.*, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012; **485:** 376-380.
11. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015; **11:** e1004219.

Supplemental Figure 1. Genotyping, Imputation and Quality Control
Sample sizes available for analysis following quality control



Anti-cancer drugs

Radiation

Pre-transplant AML, MDS and ALL patients

Healthy unrelated 8/8 HLA matched BMT donors, >18 y

Allogeneic BMT donor-recipient pairs

Cohort 1: 2111 non-hispanic EAs, 125 EA Hispanics, 54 African Americans, 34 Asian Americans

Cohort 2: 779 non-hispanic EAs, 62 Hispanic EAs, 25 African Americans, 17 Asian Americans

Genotyping and Phenotyping

Genotyping

OmniExpress

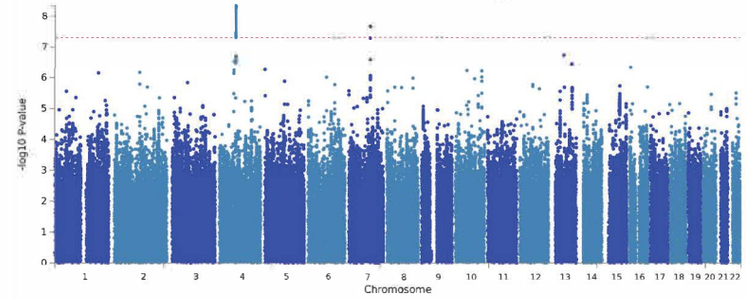
SNPs & SNVs

Survival data one year post-BMT (OS, DRM and TRM)

QC and Genotype Imputation w/HRC

Genome wide survival analysis

European Americans Cohorts 1 & 2 w/`gwasurvivr`: donor genome, recipient genome, donor-recipient allele mismatch

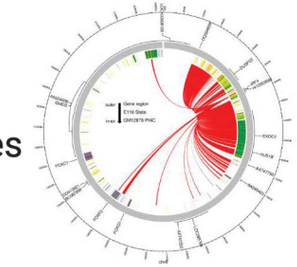


Meta-analysis of Cohorts 1 and 2 (~9M variants)

Fine-mapping, annotation & Bioinformatic analyses

-Fine mapping- Approx. Bayes Factor (w/ `Coloc`)

-Roadmap Epigenome Hi-C analyses

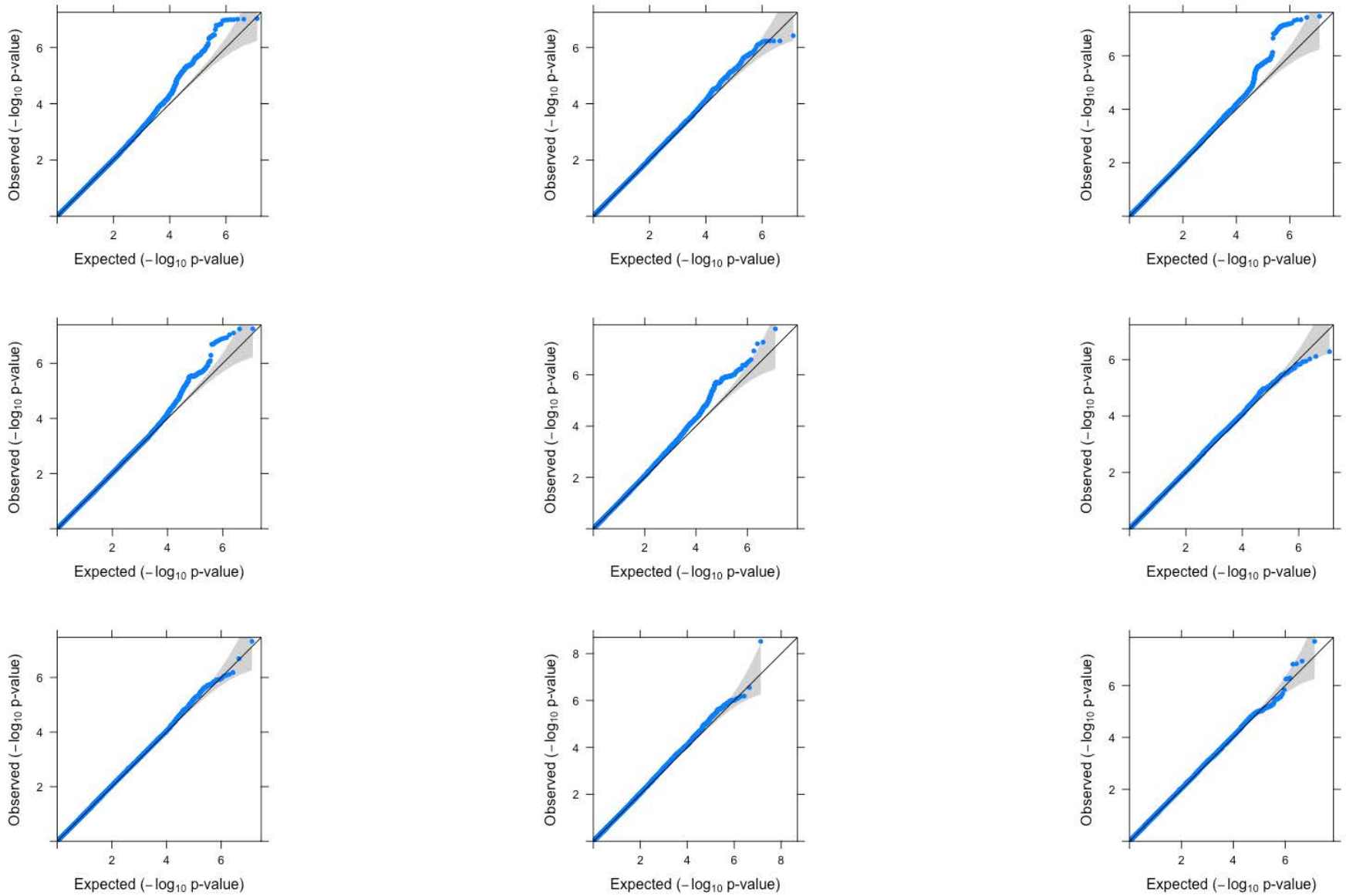


-gene-based analyses (MAGMA), CADD scores, regulomedb scores, eQTLs, pQTLs (additional FUMA annotation)

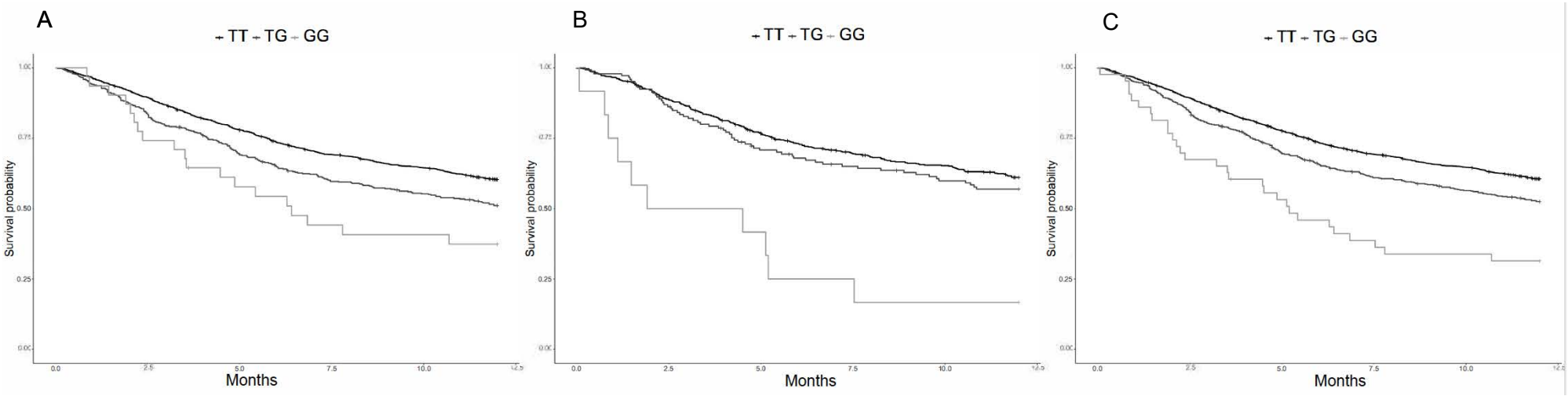
Supplemental Figure 2. Work flow and study design of DISCOVeRY-BMT study

created with Biorender

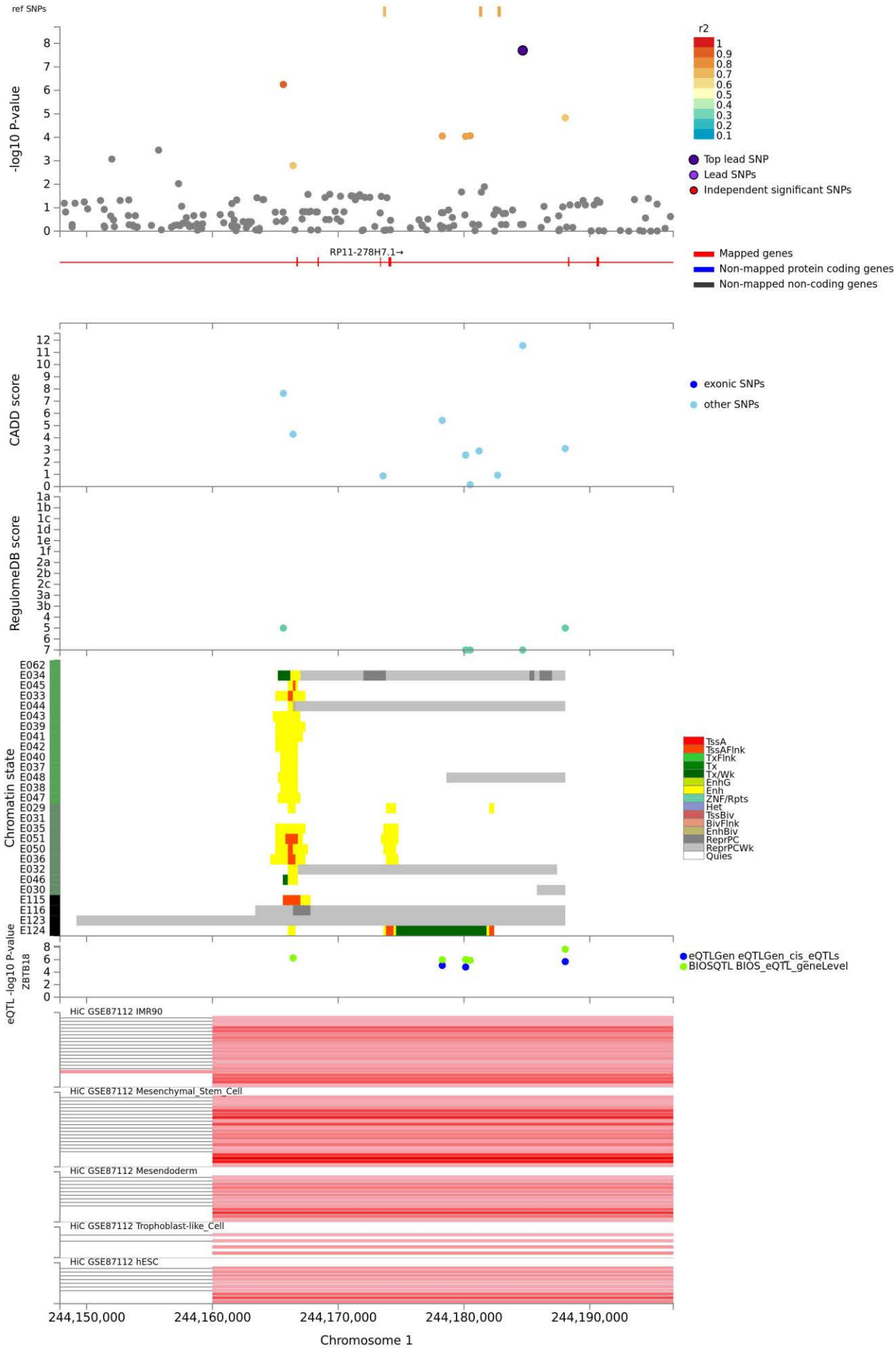
Supplemental Figure 3. Q-Q plots of SNP association by cohort 1, cohort 2 and overall survival
Observed p-values for all SNPs are plotted in grey against expected p-values for meta-analysis. From left to right, cohort 1, cohort 2 and meta-analysis. The top row are qqplots of p-values from recipient associations with overall survival, the middle row are qqplots of p-values for donor associations with OS, the bottom row are qqplots of p-values for donor-recipient mismatch associations with OS.



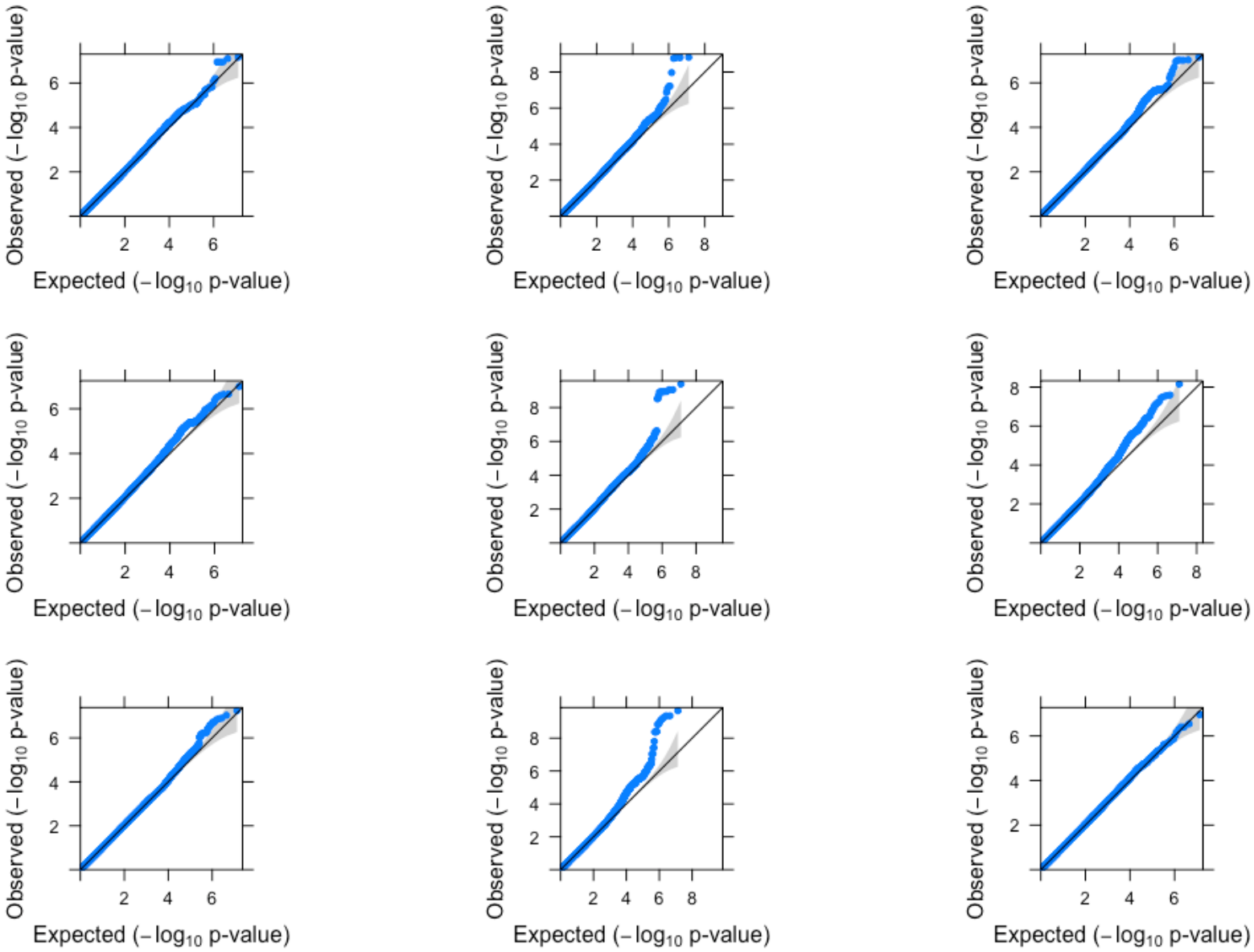
Supplemental Figure 4. One-year survival curves by genotype for rs9990017
A) Cohort 1 B) Cohort 2 and C) Cohorts 1 and 2 combined



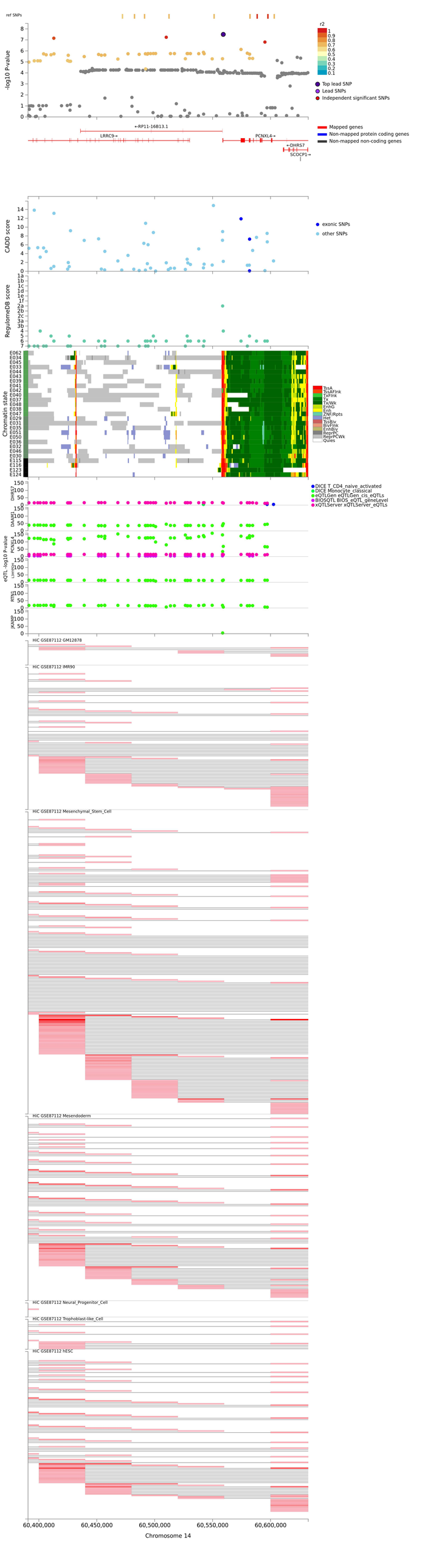
Supplemental Figure 5. OS regional association plot of allele mismatch in Chromosome 1
Regional plot of the chromosome 1 associated locus with, from the top, GWAS *P*-value (SNPs are colored based on r^2), CADD score, RequlomeDB score and eQTL *P*-value across multiple datasets. eQTLs are plotted per gene and colored based on tissue types. In the plots of CADD score, RegulomeDB score and eQTLs, SNPs which are not in LD with the lead SNP are colored gray. 15-core chromatin states for Blood and T-cells are blood are shown with the states color-coded and shown in the legend at the far right. Significant Hi-C data loops across 21 tissues from GSE87112 are shown at the bottom, these are re-processed data (output of Fit-Hi-C) from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87112


ref SNPs
-log10 P-value
r2
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Top lead SNP
Lead SNPs
Independent significant SNPs
RP11-278H7.1→
Mapped genes
Non-mapped protein coding genes
Non-mapped non-coding genes
CADD score
exonic SNPs
other SNPs
RegulomeDB score
Chromatin state
E062
E034
E045
E033
E044
E043
E039
E041
E042
E040
E037
E048
E038
E047
E029
E031
E035
E051
E050
E036
E032
E046
E030
E115
E116
E123
E124
TssA
TssAFlnk
TxFlnk
Tx
Tx/Wk
EnhG
Enh
ZNF/Rpts
Het
TssBiv
BivFlnk
EnhBiv
ReprPC
ReprPCWk
Quies
eQTL -log10 P-value
ZBTB18
eQTLGen eQTLGen_cis_eQTLs
BIOSQTL BIOS_eQTL_geneLevel
HiC GSE87112 IMR90
HiC GSE87112 Mesenchymal_Stem_Cell
HiC GSE87112 Mesendoderm
HiC GSE87112 Trophoblast-like_Cell
HiC GSE87112 hESC
244,150,000
244,160,000
244,170,000
244,180,000
244,190,000
Chromosome 1

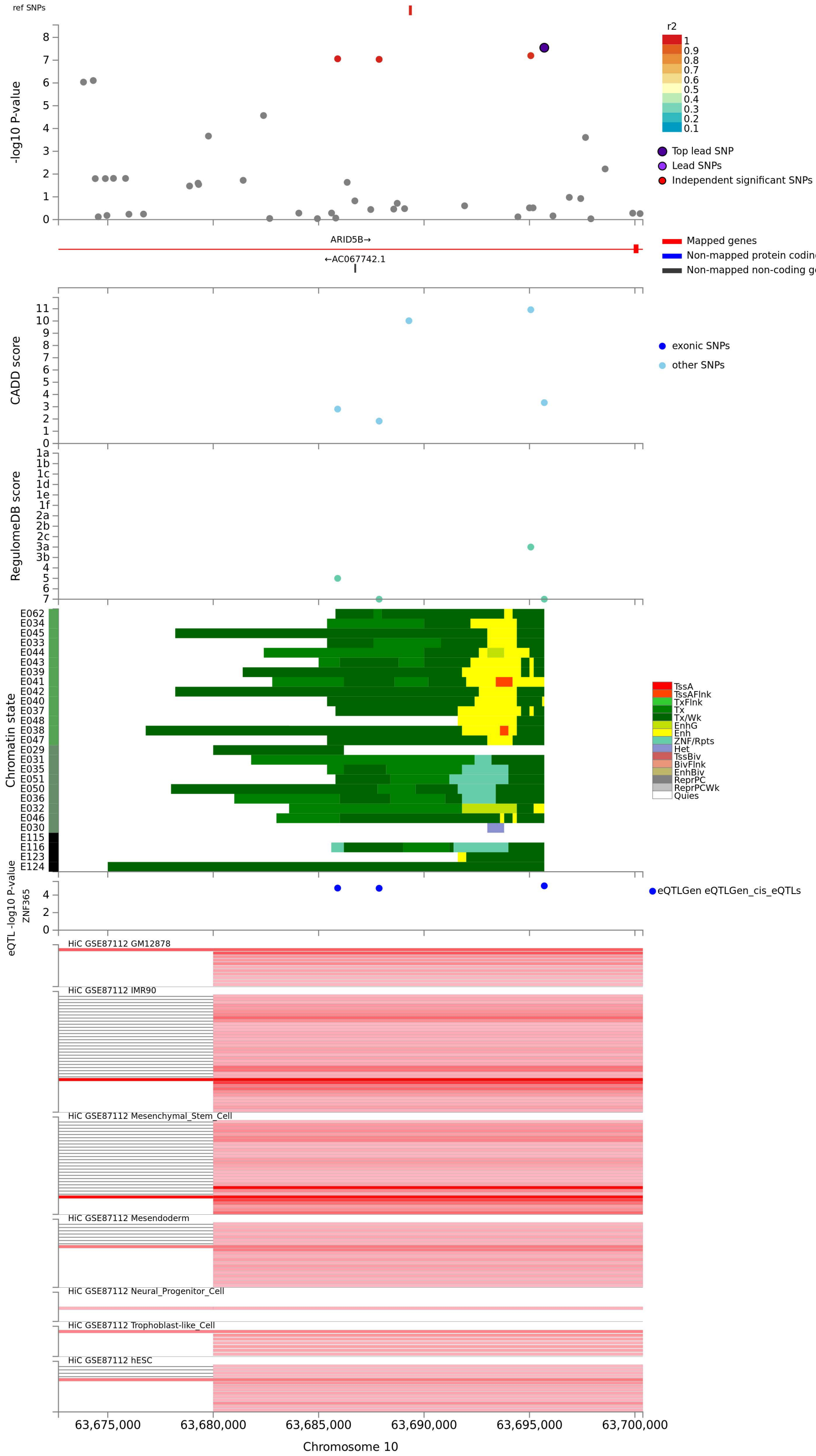
Supplemental Figure 6. Q-Q plot of donor associations with DRM
Observed p-values for all SNPs are plotted in grey against expected p-values for meta-analysis.
From left to right, cohort 1, cohort 2 and meta-analysis. The top row are qqplots of p-values from recipient associations with disease-related mortality (DRM), the middle row are qqplots of p-values for donor associations with DRM, the bottom row are qqplots of p-values for donor-recipient mismatch associations with DRM.



Supplemental Figure 7. DRM regional association plot of donor chromosome 14 *PCNXL4* region
Regional plot of the chromosome 14 associated locus with, from the top, GWAS *P*-value (SNPs are colored based on r^2), CADD score, RequlomeDB score and eQTL *P*-value across multiple datasets. eQTLs are plotted per gene and colored based on tissue types. In the plots of CADD score, RegulomeDB score and eQTLs, SNPs which are not in LD with the lead SNP are colored gray. 15-core chromatin states for Blood and T-cells are blood are shown with the states color-coded and shown in the legend at the far right. Significant Hi-C data loops across 21 tissues from GSE87112 are shown at the bottom, these are re-processed data (output of Fit-Hi-C) from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87112


ref SNPs
-log10 P-value
r2
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Top lead SNP
Lead SNPs
Independent significant SNPs
←RP11-16B13.1
LRRC9→
PCNXL4→
←DHRS7
SCOCP1→
Mapped genes
Non-mapped protein coding genes
Non-mapped non-coding genes
CADD score
exonic SNPs
other SNPs
RegulomeDB score
Chromatin state
E062
E034
E045
E033
E044
E043
E039
E041
E042
E040
E017
E048
E028
E047
E029
E031
E035
E051
E050
E036
E032
E046
E030
E118
E116
E123
E124
TssA
TssAFlnk
TxFlnk
Tx
TxWk
EnhG
Enh
ZNF/Rpts
Het
TssBiv
BivFlnk
EnhBiv
ReprPC
ReprPCWk
Quies
DHRS7
DAAM1
PCNXL4
LRRC9
RTN1
KIAA0...
eQTL -log10 P-value
DICE T_CD4_naive_activated
DICE Monocyte_classical
eQTLGen eQTLGen_cis_eQTLs
BIOSQTL BIOS_eQTL_geneLevel
xQTLServer xQTLServer_eQTLs
HiC GSE87112 GM12878
HiC GSE87112 IMR90
HiC GSE87112 Mesenchymal_Stem_Cell
HiC GSE87112 Mesendoderm
HiC GSE87112 Neural_Progenitor_Cell
HiC GSE87112 Trophoblast-like_Cell
HiC GSE87112 hESC
60,400,000
60,450,000
60,500,000
60,550,000
60,600,000
Chromosome 14

Supplemental Figure 8. DRM regional association plot of donor chromosome 10 *ARID5B* region
Regional plot of the chromosome 10 associated locus with, from the top, GWAS *P*-value (SNPs are colored based on r^2), CADD score, RequlomeDB score and eQTL *P*-value across multiple datasets. eQTLs are plotted per gene and colored based on tissue types. In the plots of CADD score, RegulomeDB score and eQTLs, SNPs which are not in LD with the lead SNP are colored gray. 15-core chromatin states for Blood and T-cells are blood are shown with the states color-coded and shown in the legend at the far right. Significant Hi-C data loops across 21 tissues from GSE87112 are shown at the bottom, these are re-processed data (output of Fit-Hi-C) from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87112


ref SNPs
-log10 P-value
r2
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Top lead SNP
Lead SNPs
Independent significant SNPs
ARID5B→
←AC067742.1
Mapped genes
Non-mapped protein coding genes
Non-mapped non-coding genes
CADD score
exonic SNPs
other SNPs
RegulomeDB score
Chromatin state
E062
E034
E045
E033
E044
E043
E039
E041
E042
E040
E037
E048
E038
E047
E029
E031
E035
E051
E050
E036
E032
E046
E030
E115
E116
E123
E124
TssA
TssAFlnk
TxFlnk
Tx
Tx/Wk
EnhG
Enh
ZNF/Rpts
Het
TssBiv
BivFlnk
EnhBiv
ReprPC
ReprPCWk
Quies
eQTL -log10 P-value
ZNF365
eQTLGen eQTLGen_cis_eQTLs
HiC GSE87112 GM12878
HiC GSE87112 IMR90
HiC GSE87112 Mesenchymal_Stem_Cell
HiC GSE87112 Mesendoderm
HiC GSE87112 Neural_Progenitor_Cell
HiC GSE87112 Trophoblast-like_Cell
HiC GSE87112 hESC
63,675,000
63,680,000
63,685,000
63,690,000
63,695,000
63,700,000
Chromosome 10

Supplemental Figure 9. DRM regional association plot of donor chromosome 10 *LINC02628* region
Regional plot of the chromosome 10 associated locus with, from the top, GWAS *P*-value (SNPs are colored based on r^2), CADD score, RequlomeDB score and eQTL *P*-value across multiple datasets. eQTLs are plotted per gene and colored based on tissue types. In the plots of CADD score, RegulomeDB score and eQTLs, SNPs which are not in LD with the lead SNP are colored gray. 15-core chromatin states for Blood and T-cells are blood are shown with the states color-coded and shown in the legend at the far right. Significant Hi-C data loops across 21 tissues from GSE87112 are shown at the bottom, these are re-processed data (output of Fit-Hi-C) from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87112

ref SNPs

-log10 P-value

r2
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1

Top lead SNP
Lead SNPs
Independent significant SNPs

←RP11-476F14.1

Mapped genes
Non-mapped protein coding genes
Non-mapped non-coding genes

CADD score

exonic SNPs
other SNPs

RegulomeDB score

Chromatin state

TssA
TssAFlnk
TxFlnk
Tx
Tx/Wk
EnhG
Enh
ZNF/Rpts
Het
TssBiv
BivFlnk
EnhBiv
ReprPC
ReprPCWk
Quies

HiC GSE87112 IMR90
HiC GSE87112 Mesenchymal_Stem_Cell
HiC GSE87112 Mesendoderm
HiC GSE87112 Neural_Progenitor_Cell
HiC GSE87112 Trophoblast-like_Cell
HiC GSE87112 hESC

33,876,000 33,877,000 33,878,000 33,879,000 33,880,000

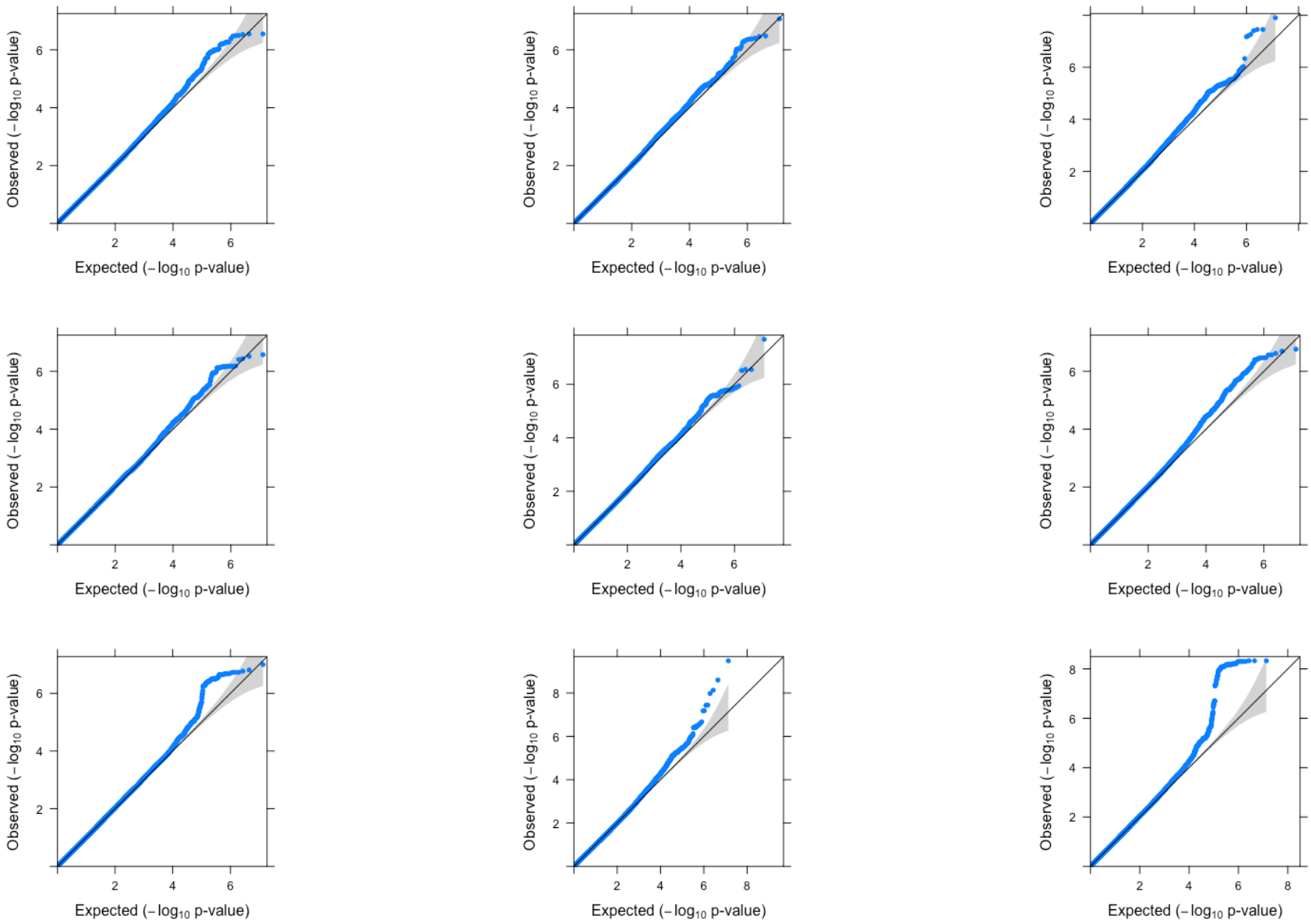
Chromosome 10

No eQTL of selected tissues exists in this region.

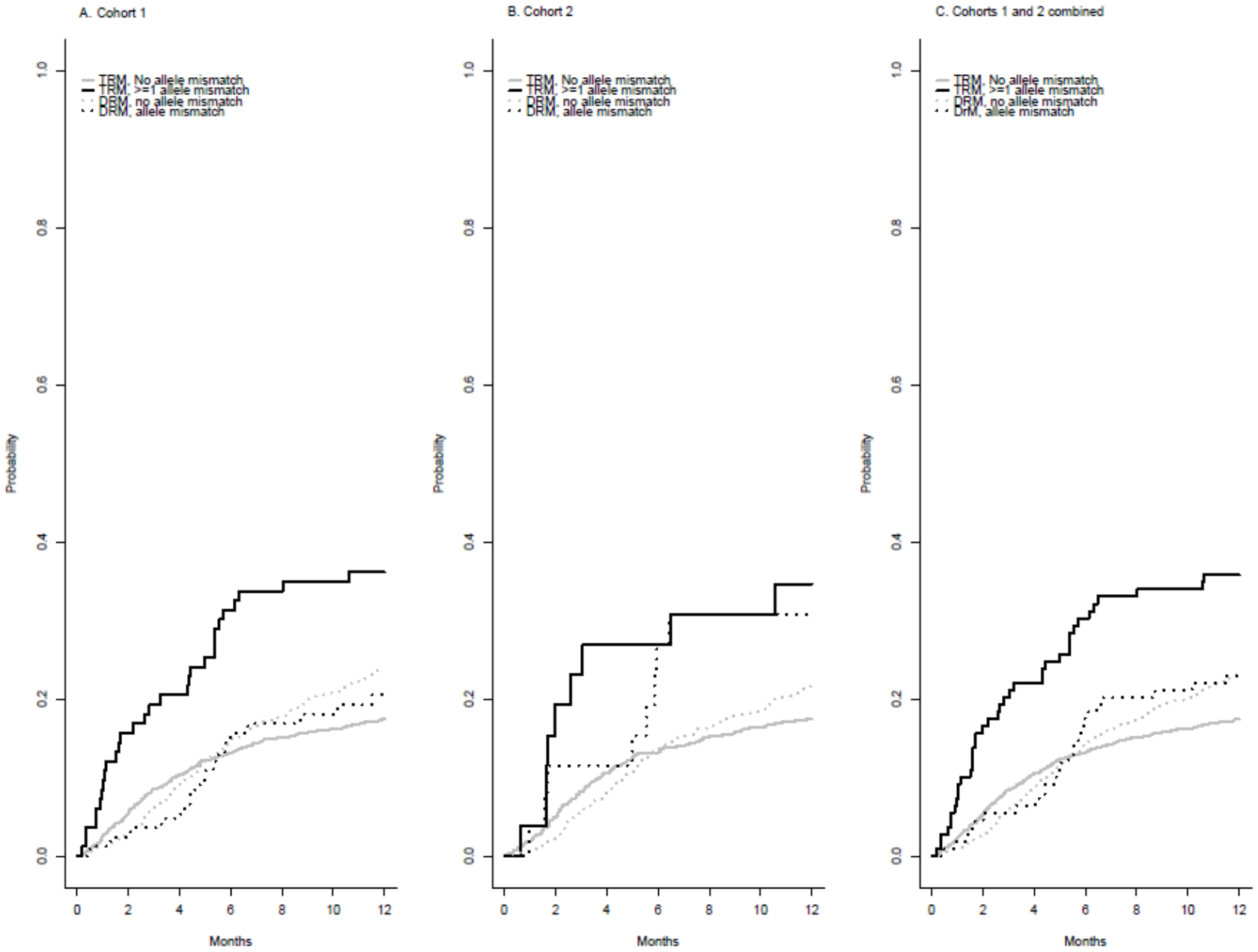
Supplemental Figure 10. DRM Regional association plot of donor chromosome 5 *CT49* region
Regional plot of the chromosome 5 associated locus with, from the top, GWAS *P*-value (SNPs are colored based on r^2), CADD score, RequlomeDB score and eQTL *P*-value across multiple datasets. eQTLs are plotted per gene and colored based on tissue types. In the plots of CADD score, RegulomeDB score and eQTLs, SNPs which are not in LD with the lead SNP are colored gray. 15-core chromatin states for Blood and T-cells are blood are shown with the states color-coded and shown in the legend at the far right. Significant Hi-C data loops across 21 tissues from GSE87112 are shown at the bottom, these are re-processed data (output of Fit-Hi-C) from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87112

No eQTL of selected tissues exists in this region.

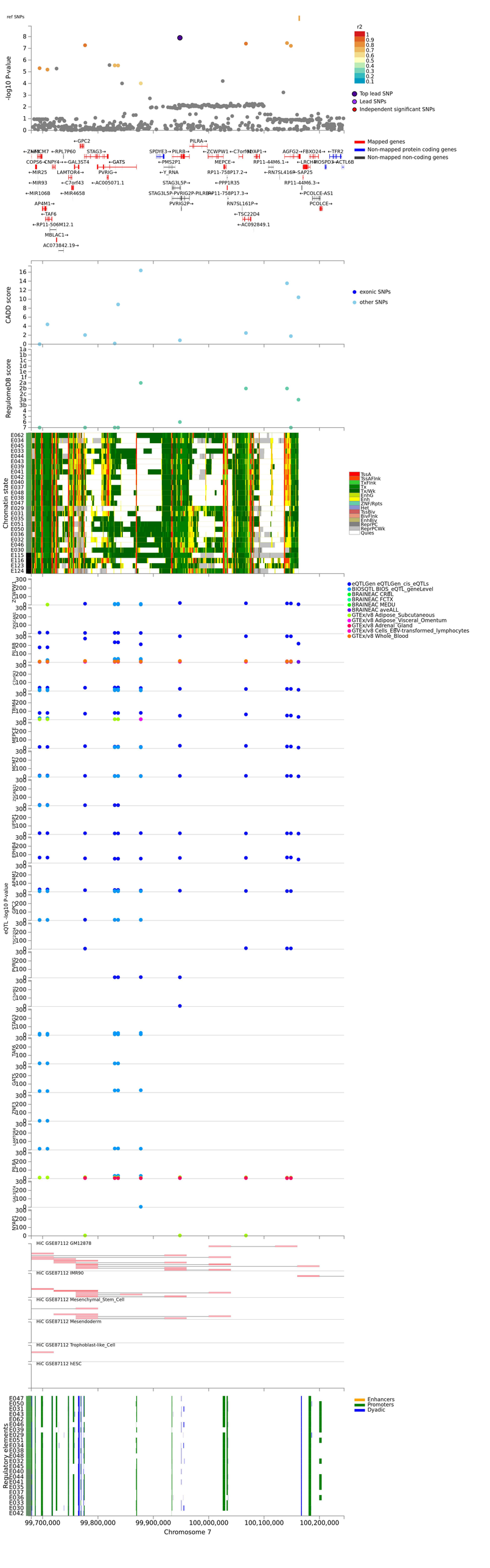
Supplemental Figure 11. Q-Q plots of SNP association by cohort 1, cohort 2 and TRM
Observed p-values for all SNPs are plotted in grey against expected p-values for meta-analysis. From left to right, cohort 1, cohort 2 and meta-analysis. The top row are qqplots of p-values from recipient associations with TRM, the middle row are qqplots of p-values for donor associations with TRM, the bottom row are qqplots of p-values for donor-recipient mismatch associations with TRM.



Supplemental Figure 12. Cumulative incidence of TRM, and the competing risk DRM, by allele mismatches at SNP rs16858805 for A) Cohort 1, B) Cohort 2, C) both Cohorts combined. Typed SNP rs16858805 is in perfect LD (r^2=1.0) with imputed SNP rs75868097 in chromosome 4.



Supplemental Figure 13. TRM regional association plot of recipient chromosome 7 *PILRB/PILRA* region
Zoomed in regional plot of the chromosome 7 associated locus with, from the top, GWAS *P*-value (SNPs are colored based on r^2), CADD score, RequlomeDB score and eQTL *P*-value across multiple datasets. eQTLs are plotted per gene and colored based on tissue types. In the plots of CADD score, RegulomeDB score and eQTLs, SNPs which are not in LD with the lead SNP are colored gray. 15-core chromatin states for Blood and T-cells are blood are shown with the states color-coded and shown in the legend at the far right. Significant Hi-C data loops across 21 tissues from GSE87112 are shown at the bottom, these are re-processed data (output of Fit-Hi-C) from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87112


ref SNPs
-log10 P-value
r2
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Top lead SNP
Lead SNPs
Independent significant SNPs
Mapped genes
Non-mapped protein coding genes
Non-mapped non-coding genes
CADD score
exonic SNPs
other SNPs
RegulomeDB score
Chromatin state
TssA
TssAFlnk
TxFlnk
Tx
TxWk
EnhG
Enh
ZNF/Rpts
Het
TssBiv
BivFlnk
EnhBiv
ReprPC
ReprPCWk
Quies
eQTL -log10 P-value
eQTLGen eQTLGen_cis_eQTLs
BIOSQTL BIOS_eQTL_geneLevel
BRAINEAC CRBL
BRAINEAC FCTX
BRAINEAC MEDU
BRAINEAC aveALL
GTEx/v8 Adipose_Subcutaneous
GTEx/v8 Adipose_Visceral_Omentum
GTEx/v8 Adrenal_Gland
GTEx/v8 Cells_EBV-transformed_lymphocytes
GTEx/v8 Whole_Blood
HiC GSE87112 GM12878
HiC GSE87112 IMR90
HiC GSE87112 Mesenchymal_Stem_Cell
HiC GSE87112 Mesendoderm
HiC GSE87112 Trophoblast-like_Cell
HiC GSE87112 hESC
Regulatory elements
Enhancers
Promoters
Dyadic
99,700,000
99,800,000
99,900,000
100,000,000
100,100,000
100,200,000
Chromosome 7