Supplemental Material

1. Search term:

(SNP[Text Word] OR ("polymorphism, genetic"[MeSH Terms] OR ("polymorphism"[All Fields] AND "genetic"[All Fields]) OR "genetic polymorphism"[All Fields] OR "polymorphism"[All Fields])) AND (allo-HSCT[All Fields] OR allo-HCT[All Fields] OR ("unrelated"[All Fields] AND ("donor"[All Fields] OR "donors"[All Fields]) AND ("transplant"[All Fields] OR "transplantation"[All Fields])) OR ("allogeneic"[All Fields] AND ("transplant"[All Fields] OR "transplantation"[All Fields])) OR ("hematopoietic"[All Fields] AND ("transplant"[All Fields] OR "transplantation"[All Fields]))) AND (("mortality"[Subheading] OR "mortality"[All Fields] OR "mortality"[MeSH Terms]) OR ("mortality"[Subheading] OR "mortality"[All Fields] OR "survival"[All Fields] OR "survival"[MeSH Terms]) OR (non[All Fields] AND ("recurrence"[MeSH Terms] OR "recurrence"[All Fields] OR "relapse"[All Fields])) OR non-relapse[All Fields]) AND English[Language]

Inclusion criteria:
1. non-HLA genes
2. survival after BMT as phenotype

Excluded:
1. Reviews
2. Working group studies
3. Non-English papers
4. SNPs not in build hg19
5. Haplotypes
6. CML or MM or lymphoma only papers
7. Autosomal only
8. Microsatellites

List of supplemental figures and tables:

1. Table S1 – Data collected from candidate gene papers
2. Table S2 – Counts of publications that studied respective SNPs categorized by validation or replication attempts in DISCOVeRY-BMT
3. Table S3 – Replication attempt of previous publications that met the replication criteria
4. Table S4 – Validation attempt of previous publications that met the validation criteria
5. Table S5 – Gene based analyses of all previously published candidate genes
6. Table S6 – RegulomeDB and eQTL annotation
7. Figure S1 – Validation attempts of previously reported significant candidate gene association studies in DISCOVeRY-BMT
8. Figure S2 – Quantile–quantile (QQ) plot of SNP based p-values in DISCOVeRY-BMT for all previously studied SNPs for association with post-BMT outcomes.
9. Figure S3 – Power calculations for the DISCOVeRY-BMT cohort.

Further Details for columns

Column	Description	Details
Gene	Candidate gene studied	*HGNC gene symbol*
Reference	Reference of the publication	
PMID	PMID of the publication	
Population	Reported race of the study population by the publication	
Disease	Type of disease observed in the study population	Mixed: 3 or more types of blood diseases including non-malignant
Graft	Type of donor	RD: Related Donor; URD: Unrelated Donor
N (Initial/Replication)	Sample size of the study and if reported the replication cohort size	D: donors only R: recipients only D-R: donor-recipient pair
SNP	rsID of the SNP studied	
Alleles (MAF)	Alleles and minor allele frequency as reported by the author or 1000	

	Genomes for the respective population if not reported	
Model	Genetic model used for analyses by the authors	A: Additive D: Dominant R: Recessive
Outcome	Outcome tested by the authors	OS
Genome	Association tested for donor, recipient or donor-recipient mismatched genetic variation	R: Recipient D: Donor M: Mismatch
Testing	Type of model used for analysis by the authors	M: Multivariate U: Univariate
Effect Size (95% CI)	Effect size determined by the authors	HR: Hazard ratio RR: Relative risk OR: Odds ratio NP: Not presented by the study
P.Value	P-value of the association as reported by the authors	NS: Not significant

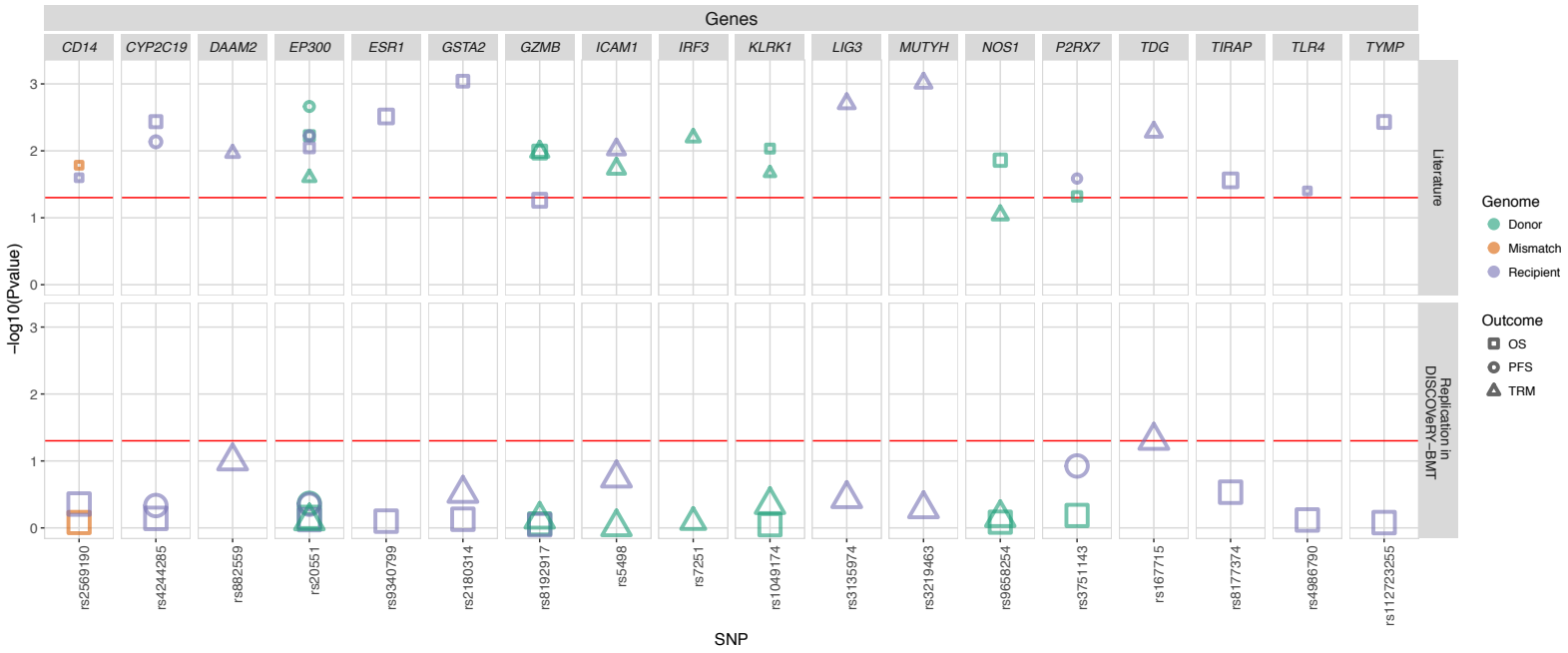


Figure S1. Validation attempts of previously reported significant candidate gene association studies in DISCOVeRY-BMT reported only once

Association p-values as reported once in previous literature (top) and validation attempts of these associations in DISCOVeRY-BMT cohort (bottom) are shown as data points. Horizontal panels indicate the genes in which the polymorphism is located or closest to as reported by the previous studies. Shapes represent associations with survival outcomes, OS (□), PFS (○) and TRM (△) and colors correspond to donor (green), recipient (purple) or donor-recipient mismatch (orange) polymorphisms. The size of the point represents the sample size of the study, with larger points reflecting a bigger sample size. Shown on x-axis are the polymorphisms from the literature reporting associations at $p<0.05$ with OS, PFS or TRM by one or more previously published studies and the y-axis is the $-\log_{10}$(p-value). The red line indicates $p=0.05$.

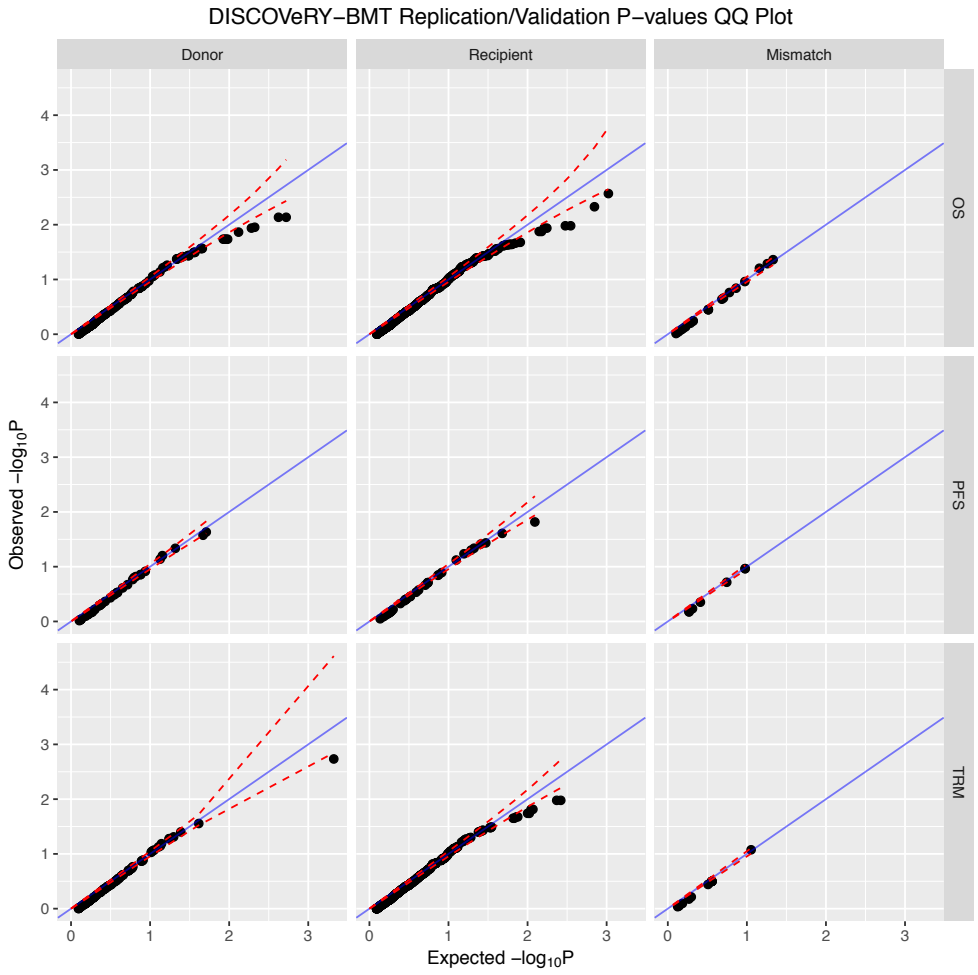
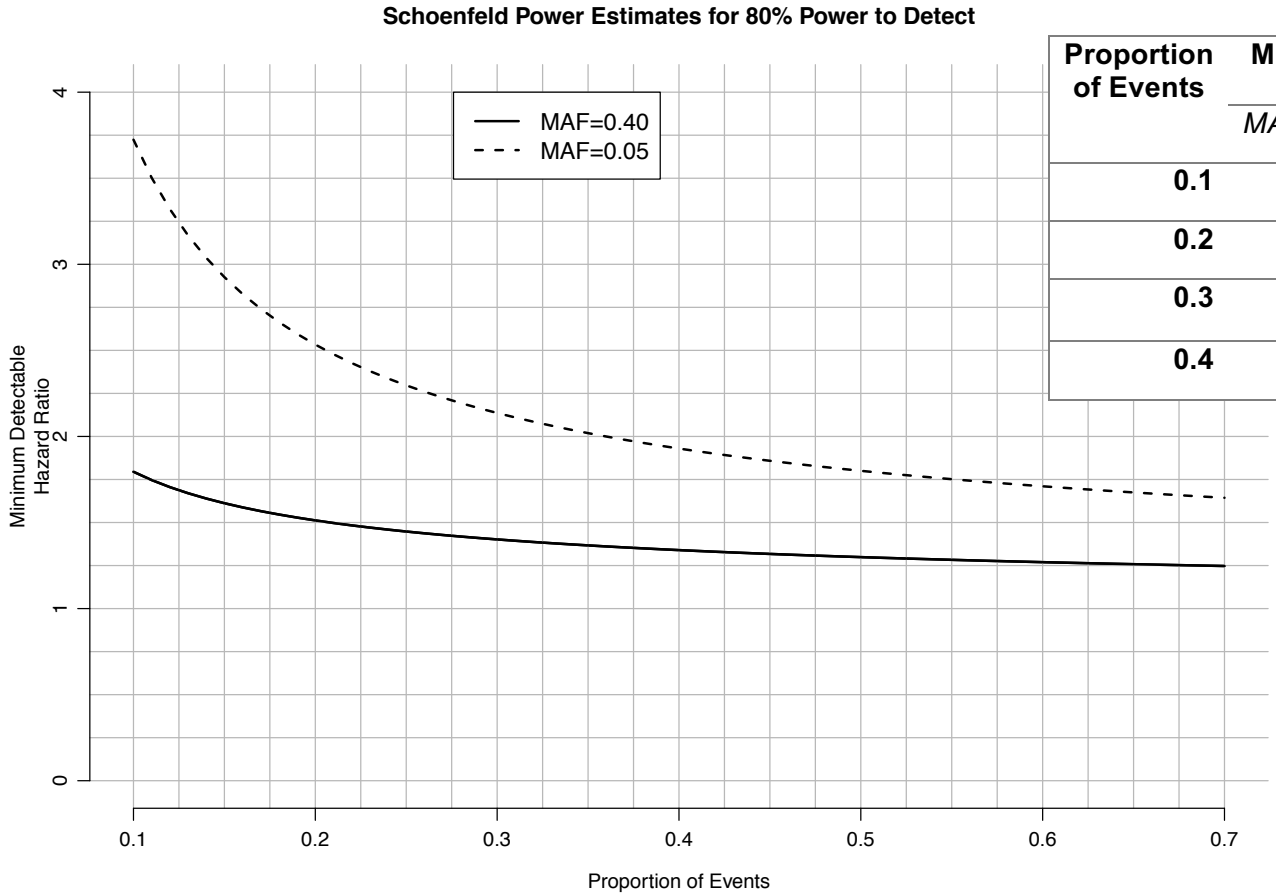


Figure S2 Quantile–quantile (QQ) plot of SNP p-values in DISCOVeRY-BMT for all previously studied SNPs. From left to right, vertical panels represent the donor, recipient, and mismatch genome, horizontal panel represents the survival outcome that was tested, from top to bottom overall survival (OS), progression free survival (PFS) and transplant related mortality (TRM). Dashed red lines indicate 95% confidence intervals.



Proportion of Events	Minimum Detectable Hazard Ratio	
	MAF=0.40	*MAF=0.05*
0.1	1.79	3.72
0.2	1.51	2.53
0.3	1.40	2.14
0.4	1.34	1.93

Figure S3. Power calculations for the DISCOVeRY-BMT cohort. Minimum detectable hazard ratios (y axis) for the proportion of events (y axis) ranging from 0.1 to 0.7 are shown for minor allele frequency of 0.4 (smooth line) and 0.05 (dashed line) respectively. Hazard ratio estimation was calculated with the assumption that the SNP-chip captures 85% of the genetic variability and for a sample size n=2883**.** The table shows minimum detectable hazard ratio given proportion of events (OS, TRM, DRM and PFS) 10% to 40%. Our study is well powered to detect all hazard ratios from the literature.