

iScience, Volume 27

Supplemental information

**A statistical approach for identifying single
nucleotide variants that affect
transcription factor binding**

Nina Baumgarten, Laura Rumpf, Thorsten Kessler, and Marcel H. Schulz

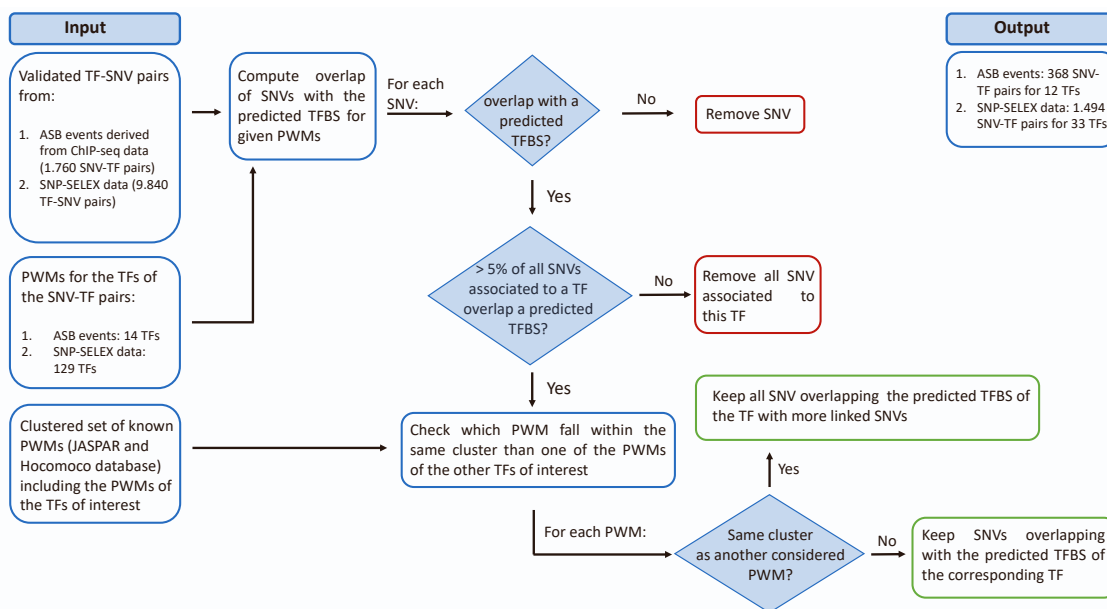


Figure S 1: Flowchart providing detailed description on how we filtered the SNVs from the data sets used to evaluate our statistical approach. TFs were only considered when published data was available, where it has been shown that the SNV affects the TFBS of the analysed TF either in the ASB-events or the SNP-SELEX data sets. Thus, we only use a small subset of all available PWMs to evaluate our approach and compare it to *atSNP*. Related to STAR Methods Section *Collecting allele-specific binding events* and STAR Methods Section *Collecting SNP-SELEX data*.

TF	PWM ID	source with version	species
TCF12	MA1648.1	JASPAR database (version 2022)	human
RUNX3	MA0684.2	JASPAR database (version 2022)	human
EGR1	MA0162.4	JASPAR database (version 2022)	human
PAX5	MA0014.3	JASPAR database (version 2022)	human
SRF	MA0083.3	JASPAR database (version 2022)	human
EBF1	MA0154.4	JASPAR database (version 2022)	human
NFYB	MA0502.2	JASPAR database (version 2022)	human
MEF2A	MA0052.4	JASPAR database (version 2022)	human
BATF	MA1634.1	JASPAR database (version 2022)	human
BCL11A	BC11A_HUMAN.H11MO.0.A	Hocomoco database (version 11)	human
BHLHE40	MA0464.2	JASPAR database (version 2022)	human
CTCF	MA0139.1	JASPAR database (version 2022)	human

Table S 1: TFs of the ASB-events we used to evaluate and compare our statistical approach. The corresponding PWMs of the motifs are provided in TRANSFAC and MEME format in our ZENODO repository (10.5281/zenodo.7588271). Related to STAR Methods *Collecting allele-specific binding events*.

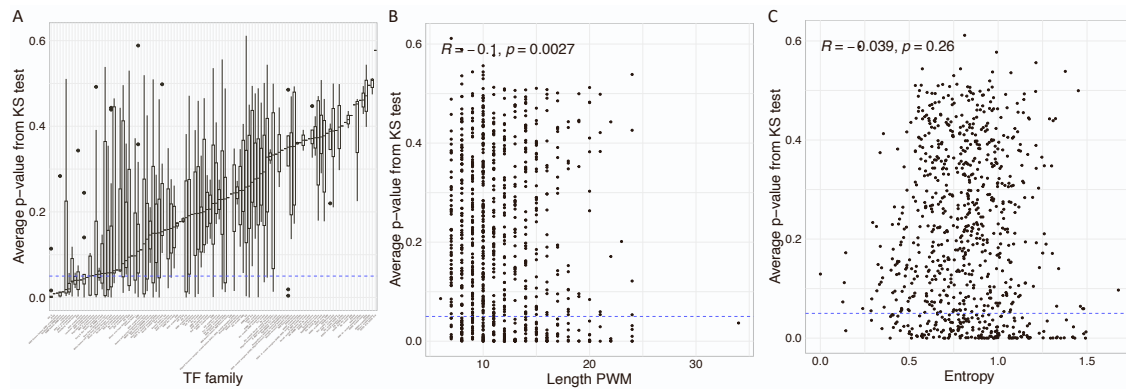


Figure S 2: Comparison of the average p -value of the Kolmogorov-Smirnov (KS) test to different features of the TF motifs. **A** Boxplot showing the average p -value in relation to the TF family. **B** and **C** Scatterplots that displays the length of the TF motif (**A**) and the overall entropy of the TF motif (**C**) in comparison to the average p -value of the KS test. The resulting Spearman's correlation coefficient and p -value are reported in the left upper corner. Related to STAR Methods .

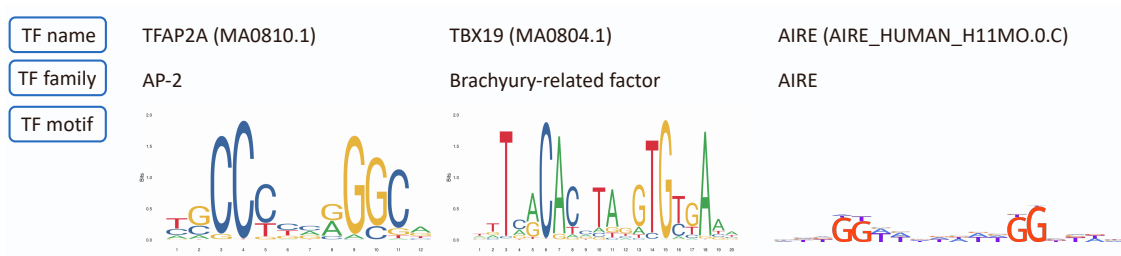


Figure S 3: Example of TF motifs with high entropy bases centered. Related to STAR Methods.

TF	PWM ID	source with version	species
OVOL1	MA1544.1	JASPAR database (version 2022)	human
PHOX2B	MA0681.2	JASPAR database (version 2022)	human
PAX7	MA0680.1	JASPAR database (version 2022)	human
CDX1	MA0878.3	JASPAR database (version 2022)	human
ZIC1	MA0696.1	JASPAR database (version 2022)	human
HSF1	MA0486.2	JASPAR database (version 2022)	human
ELF2	MA1483.2	JASPAR database (version 2022)	human
SNAI2	MA0745.2	JASPAR database (version 2022)	human
LMX1B	MA0703.2	JASPAR database (version 2022)	human
MEIS3	MA0775.1	JASPAR database (version 2022)	human
DBP	MA0639.1	JASPAR database (version 2022)	human
SCRT1	MA0743.2	JASPAR database (version 2022)	human
GCM1	MA0646.1	JASPAR database (version 2022)	human
YY1	MA0095.2	JASPAR database (version 2022)	human
ATF2	MA1632.1	JASPAR database (version 2022)	human
NR5A1	MA1540.1	JASPAR database (version 2022)	human
DMRTA1	MA1707.1	JASPAR database (version 2022)	human
GATA5	MA0766.2	JASPAR database (version 2022)	human
JDP2	MA0655.1	JASPAR database (version 2022)	human
NEUROG1	MA0623.2	JASPAR database (version 2022)	human
CLOCK	MA0819.2	JASPAR database (version 2022)	human
FOXB1	MA0845.1	JASPAR database (version 2022)	human
EGR1	MA0162.4	JASPAR database (version 2022)	human
IRF9	MA0653.1	JASPAR database (version 2022)	human
RFX2	MA0600.2	JASPAR database (version 2022)	human
NFIX	MA0671.1	JASPAR database (version 2022)	human
NRL	MA0842.2	JASPAR database (version 2022)	human
POU5F1	MA1115.1	JASPAR database (version 2022)	human
TFAP2B	MA0811.1	JASPAR database (version 2022)	human
MAF	MA1520.1	JASPAR database (version 2022)	human
NFATC3	MA0625.2	JASPAR database (version 2022)	human
POU4F3	MA0791.1	JASPAR database (version 2022)	human
ZBTB26	MA1579.1	JASPAR database (version 2022)	human

Table S 2: TFs of the SNP-SELEX dataset we used to evaluate and compare our statistical approach. The corresponding PWMs of the motifs are provided in TRANSFAC and MEME format in our ZENODO repository (10.5281/zenodo.7588271). Related to STAR Methods *Collecting SNP-SELEX data*

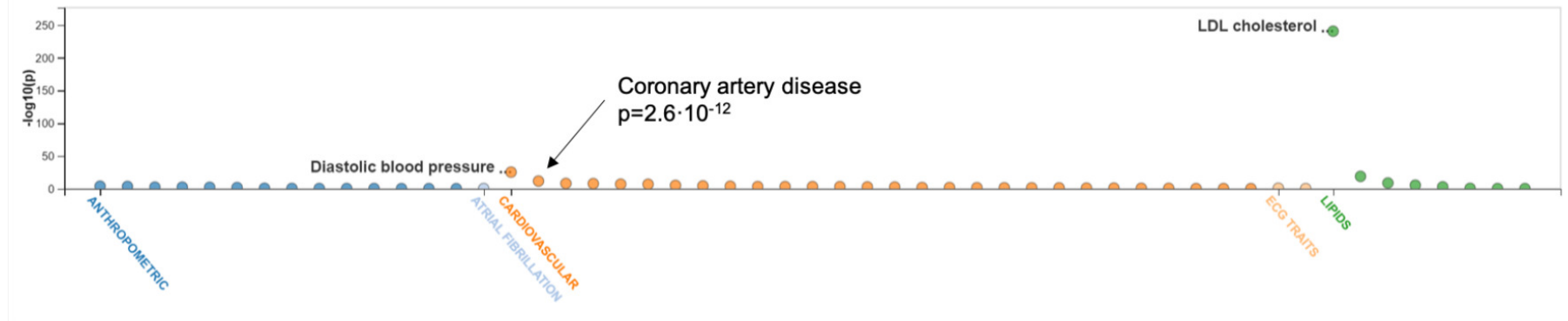
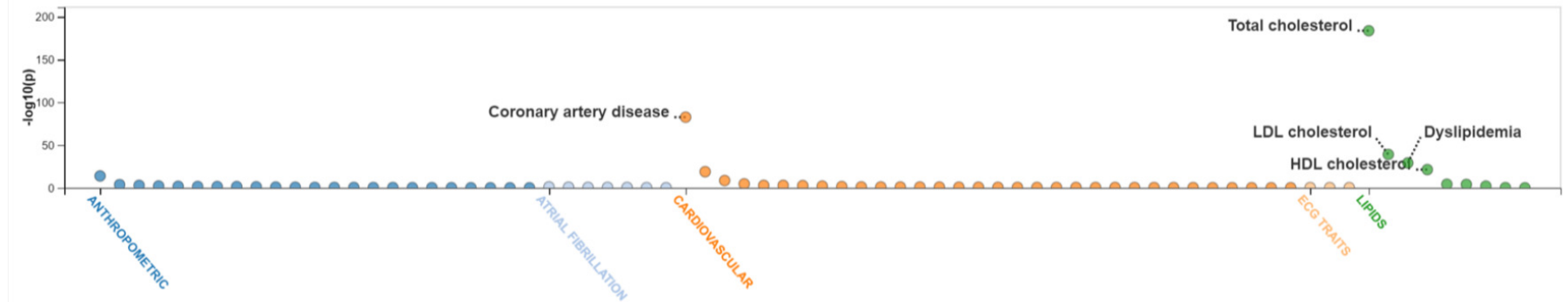
A**B**

Figure S 4: Phenome-wide associations of SNVs at predicted TF binding sites. **A** Rs582094 at the ABO CAD locus. **B** Rs629301 at the CELSR2-SORT1 CAD locus. Only significant interactions are shown ($p \leq 0.05$). Data extracted from Cardiovascular Disease Knowledge Portal (Accessed on 2021-09-01; <https://cvd.hugeamp.org/variant.html?variant=rs582094>; <https://cvd.hugeamp.org/variant.html?variant=rs629301>). Related to Result Section *Identification of TFs with altered binding sites induced by genetic variants mediating gene expression*.