The American Journal of Human Genetics, Volume *98*

Supplemental Data

Integrative Multi-omic Analysis of Human Platelet eQTLs Reveals Alternative Start Site in Mitofusin 2

Lukas M. Simon, Edward S. Chen, Leonard C. Edelstein, Xianguo Kong, Seema Bhatlekar, Isidore Rigoutsos, Paul F. Bray, and Chad A. Shaw

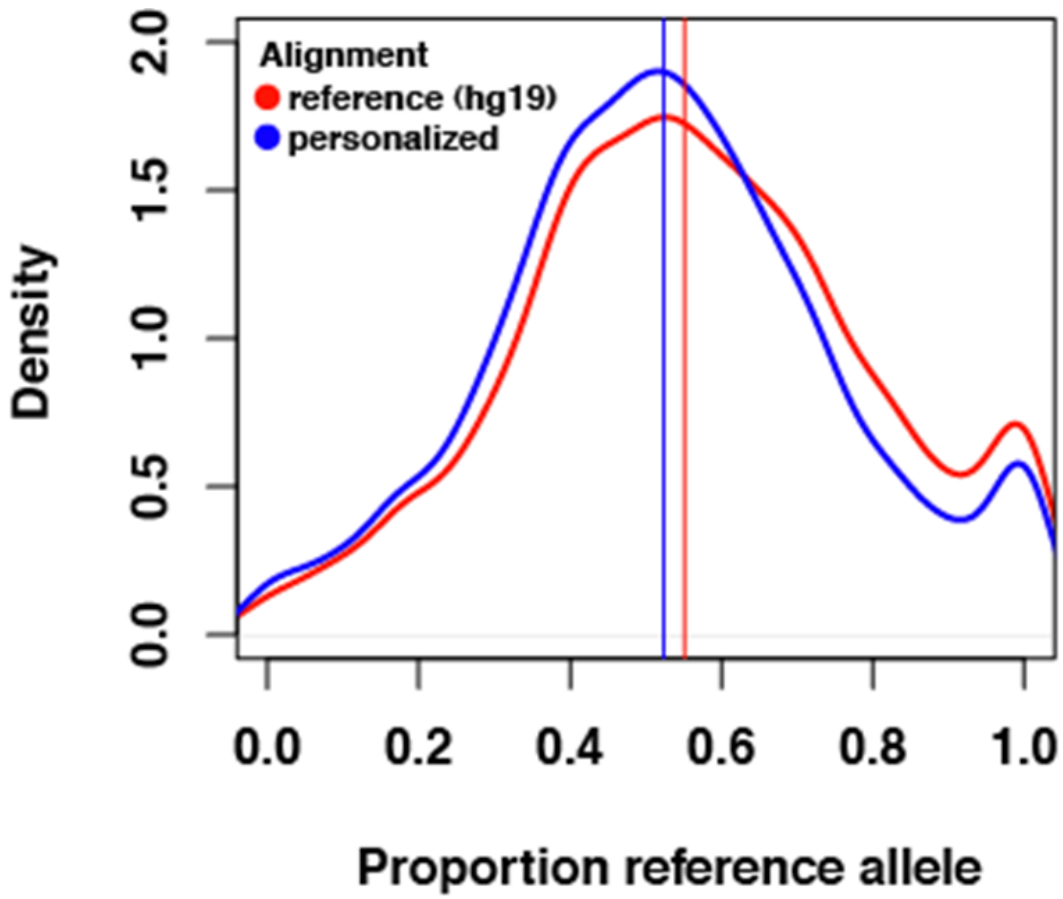
Alignment
reference (hg19)
personalized
Density
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
Proportion reference allele

Figure S1. Decrease of reference allele bias through personalized genome alignment. The blue and the red curves represent density estimates of the proportion of reference alleles based on data from 2764 heterozygous coding sites with 10 or more reads (not restricted to eGenes) when aligning RNA-seq reads to the hg19 reference genome or personalized genomes, respectively. Vertical lines indicate corresponding medians. Reference allele bias is significantly lower when aligning to personalized genomes ($P<1e-5$, KS-test).

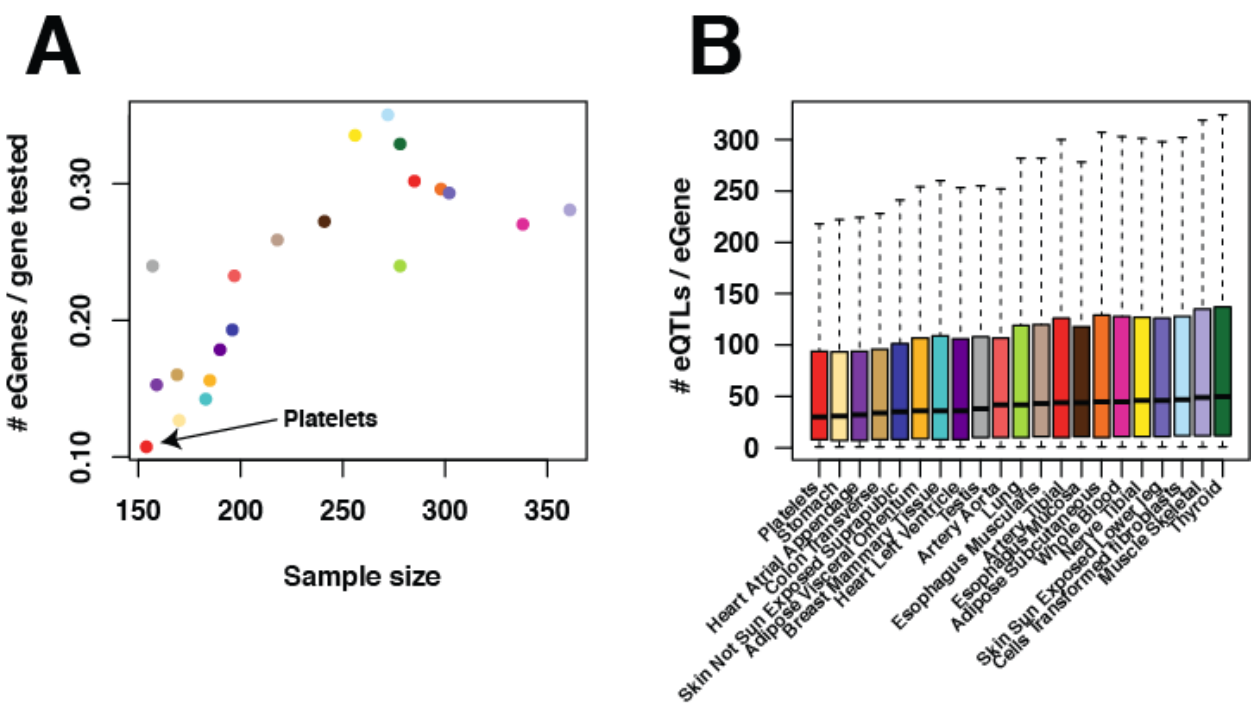
A
eGenes / gene tested
0.10
0.20
0.30
Platelets
150
200
250
300
350
Sample size
B
eQTLs / eGene
0
50
100
150
200
250
300
Platelets
Stomach
Heart Atrial Appendage
Colon Transverse
Skin Not Sun Exposed Suprapubic
Adipose Visceral Omentum
Breast Mammary Tissue
Heart Left Ventricle
Testis
Artery Aorta
Lung
Esophagus Muscularis
Artery Tibial
Esophagus Mucosa
Adipose Subcutaneous
Whole Blood
Nerve Tibial
Skin Sun Exposed Lower leg
Cells Transformed fibroblasts
Muscle Skeletal
Thyroid

Figure S2. Comparison of eGene per genes tested and eQTLs per eGene rates across tissues. Panel A depicts the number of eGenes per genes tested and sample size for all tissues on the Y and X axis, respectively. Given the sample size of our PRAX1 cohort, the rate of eGenes per gene tested is comparable to the GTEx data. Boxplot in panel B shows the distribution of eQTLs per eGene across tissues. The box represents the interquartile range, the horizontal line in the box is the median and the whiskers represent 1.5 times the interquartile range. For both panels, coloring scheme represents tissue.

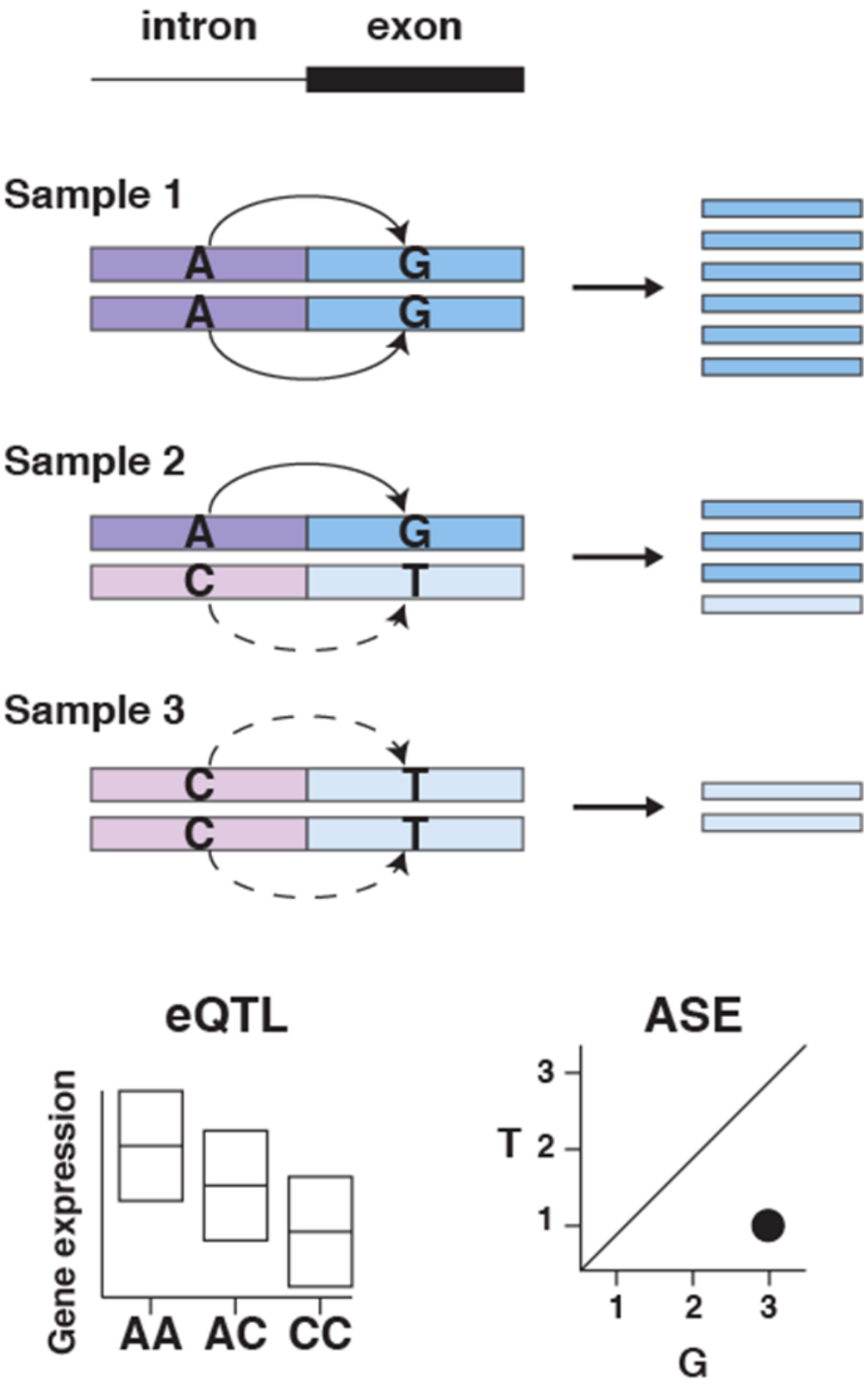
intron
exon
Sample 1
A
G
A
G
Sample 2
A
G
C
T
Sample 3
C
T
C
T
eQTL
Gene expression
AA
AC
CC
ASE
T
3
2
1
1
2
3
G

Figure S3. Schematic of the relationship between eQTLs and ASE events. Non-coding eQTL is in strong linkage disequilibrium with a coding variant. Samples 1, 2 and 3 represent all of the genotypic variation at the eQTL. RNA-seq reads from the exonic region contain information on expression level as measured by the total number of reads and allele content as illustrated by the horizontal bars in number and color, respectively. The ‘A’ allele of the eQTL is associated with increased transcription and correspondingly samples 1, 2 and 3 show high, intermediate and low expression levels, respectively. Bottom left, eQTL boxplot cartoon depicts this association between expression levels and genotype across samples. ASE information can be found in heterozygous samples by observing the allelic counts in the RNA-seq reads. Bottom right, ASE scatter plot shows the allelic counts of the linked coding variant in sample 2. RNA-seq reads contain a higher proportion of the ‘G’ allele, which corresponds to the ‘A’ allele of the eQTL, thereby validating the eQTL association.

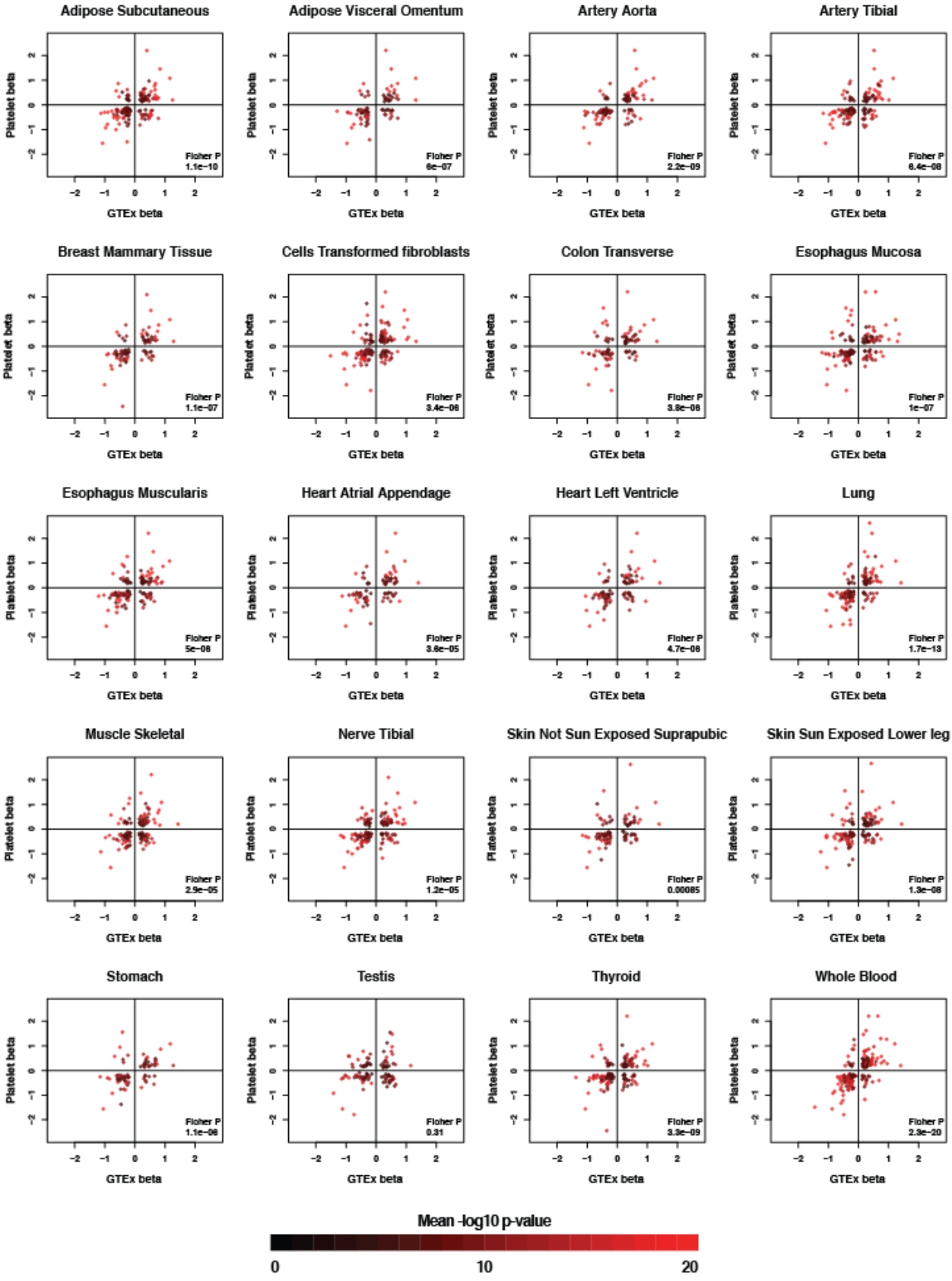
Adipose Subcutaneous
Fisher P 1.1e−10
Adipose Visceral Omentum
Fisher P 6e−07
Artery Aorta
Fisher P 2.2e−09
Artery Tibial
Fisher P 6.4e−08
Breast Mammary Tissue
Fisher P 1.1e−07
Cells Transformed fibroblasts
Fisher P 3.4e−06
Colon Transverse
Fisher P 3.8e−08
Esophagus Mucosa
Fisher P 1e−07
Esophagus Muscularis
Fisher P 5e−08
Heart Atrial Appendage
Fisher P 3.8e−05
Heart Left Ventricle
Fisher P 4.7e−08
Lung
Fisher P 1.7e−13
Muscle Skeletal
Fisher P 2.9e−05
Nerve Tibial
Fisher P 1.2e−05
Skin Not Sun Exposed Suprapubic
Fisher P 0.00085
Skin Sun Exposed Lower leg
Fisher P 1.3e−08
Stomach
Fisher P 1.1e−06
Testis
Fisher P 0.31
Thyroid
Fisher P 3.3e−09
Whole Blood
Fisher P 2.3e−20
Platelet beta
GTEx beta
Mean -log10 p-value
0
10
20

Figure S4. Concordance between platelet eQTL effect sizes and GTEx tissues. Each plot represents the comparison between platelets and one of the 20 GTEx tissues analyzed. For all plots, X and Y axes correspond to the GTEx tissue and platelet effect size, respectively. Each point represents a variant-gene eQTL association. Points are colored by their mean –log10 p-value in platelets and GTEx tissue. P-value on bottom right corners indicate Fisher's exact test p-value for the concordance between the sign of the effect size in platelets and GTEx tissue of comparison.

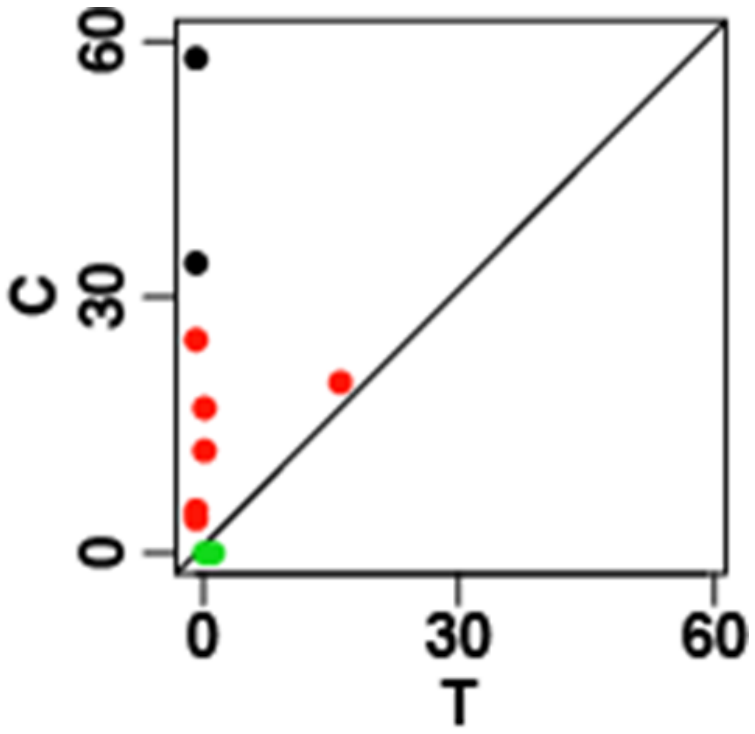
60
30
0
C
0
30
60
T

Figure S5. ASE validation of *MFN2* eQTL. Allelic counts extracted from Londin et al RNA-seq data at the genetic variant rs1474868. X and Y axes represent the count of the ‘T’ and ‘C’ alleles across 10 PRAX1 samples, respectively. Black, red and green colors indicate ‘CC’, ‘CT’ and ‘TT’ genotype, respectively. Most points fall above the diagonal line indicating that the RNA-seq reads contained a higher proportion of the ‘C’ compared to the ‘T’ allele.

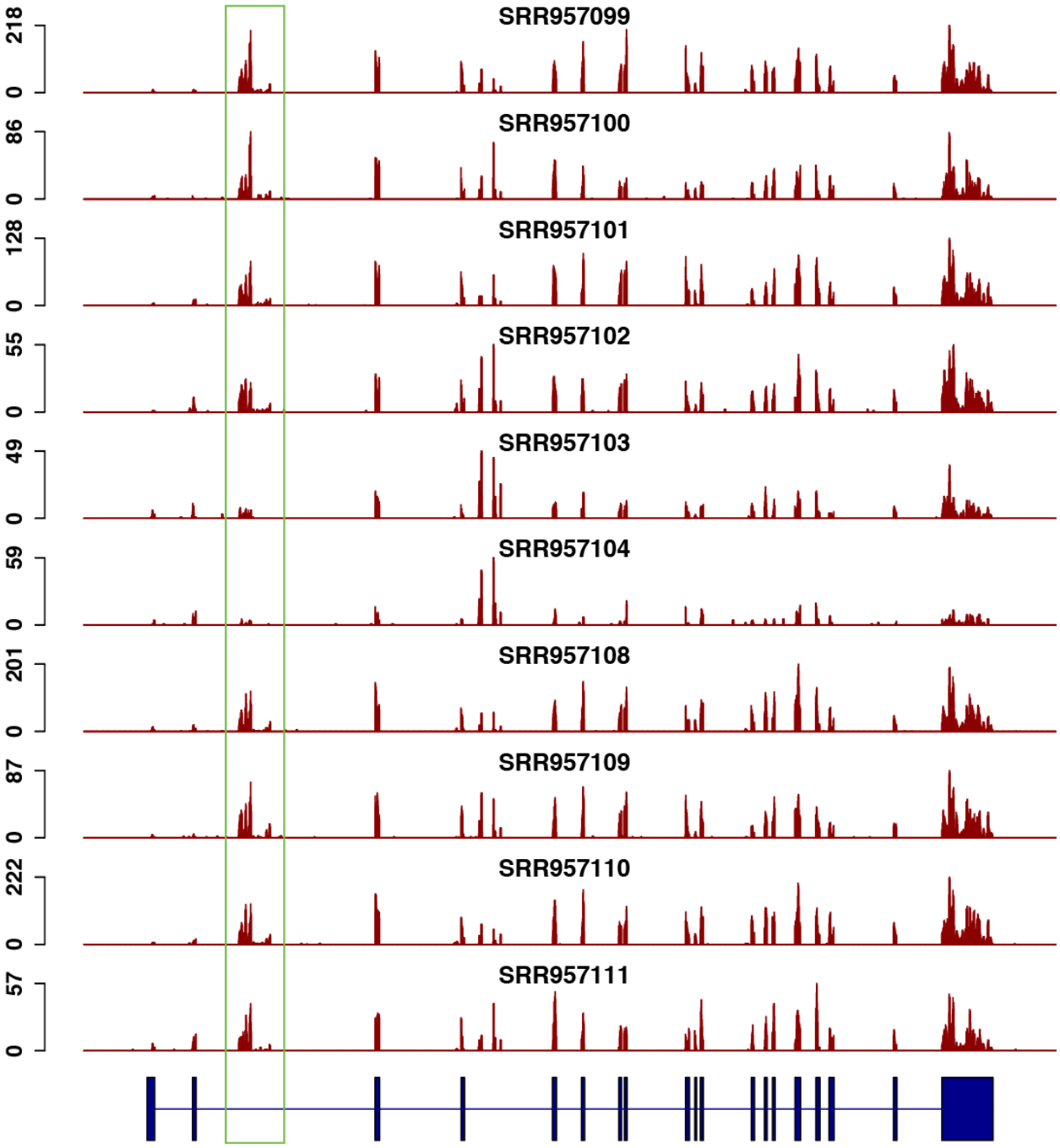
SRR957099
218
0
SRR957100
86
0
SRR957101
128
0
SRR957102
55
0
SRR957103
49
0
SRR957104
59
0
SRR957108
201
0
SRR957109
87
0
SRR957110
222
0
SRR957111
57
0

Figure S6. RNA-seq read coverage across *MFN2* for all PRAX1 RNA-seq samples. Red vertical bars indicate RNA-seq read density across 10 PRAX1 RNA-seq samples. Blue rectangles at the bottom represent the *MFN2* gene body. Green rectangle highlights a number of reads mapping into the second intron representing exon 2b.

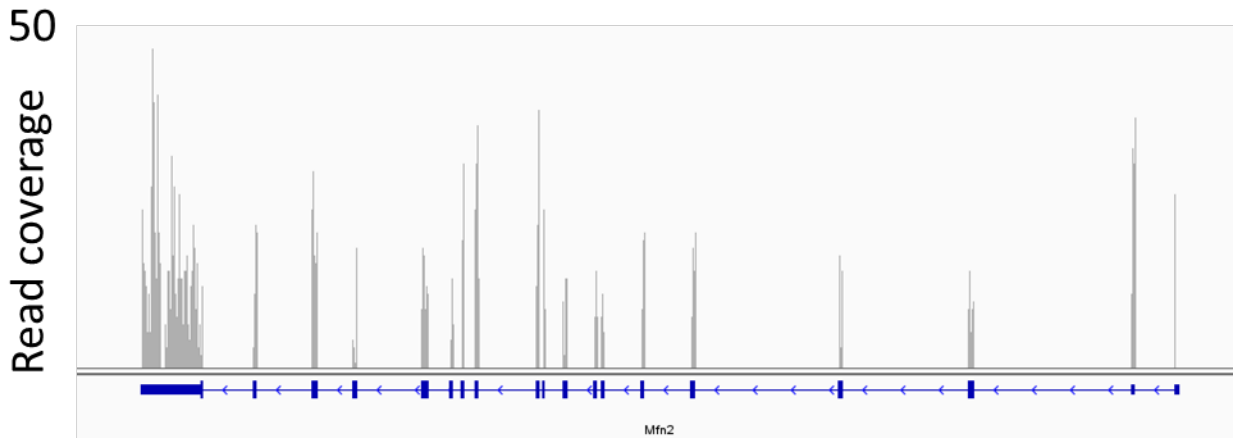
50
Read coverage
Mfn2

Figure S7. RNA-seq read coverage across *Mfn2* in mouse sample. Mouse RNA-seq data was taken from the Rowley et al data. Y and X axes correspond to read coverage and genomic coordinates. Blue rectangles below plot illustrate *Mfn2* gene body. There are no RNA-seq reads mapping into the second intron.

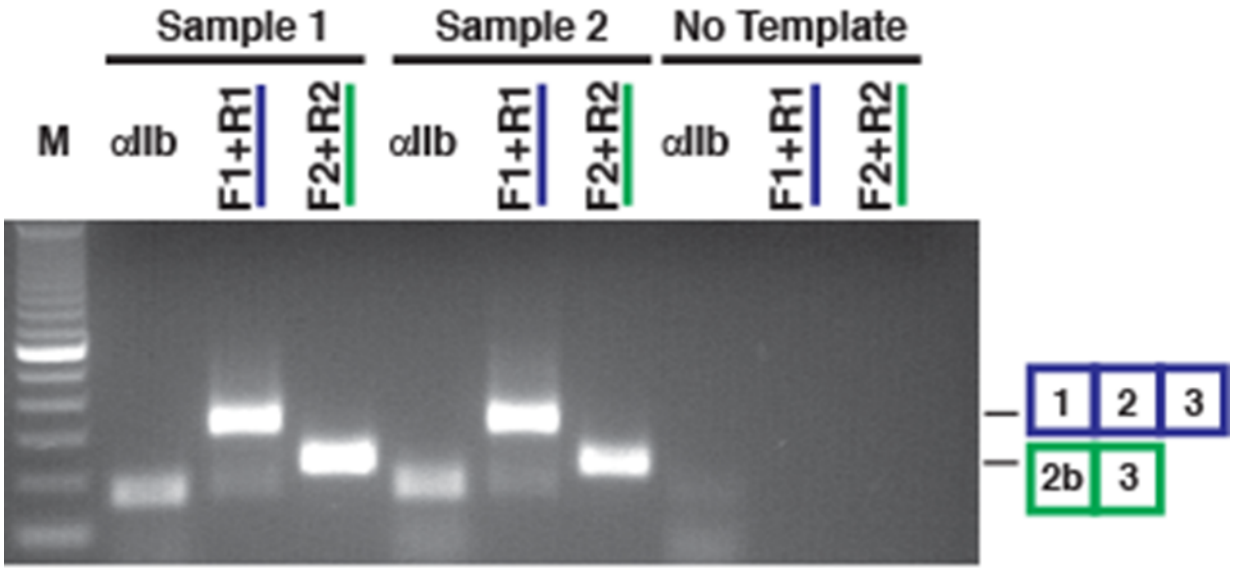
Sample 1
Sample 2
No Template
M
αIIb
F1+R1
F2+R2
αIIb
F1+R1
F2+R2
αIIb
F1+R1
F2+R2
1
2
3
2b
3

Figure S8. RT-PCR validation of exon 2b presence in human platelets. Non-quantitative RT-PCR was performed on LDP RNA using primers designed to amplify *MFN2* (spanning exons 1-2-3) or exon 2b containing *MFN2* (spanning exons 2b-3). Both reactions produced the expected sized products indicating the presence of exon 2b in platelet RNA. Integrin αIIb RNA was used as a control for RNA quality.

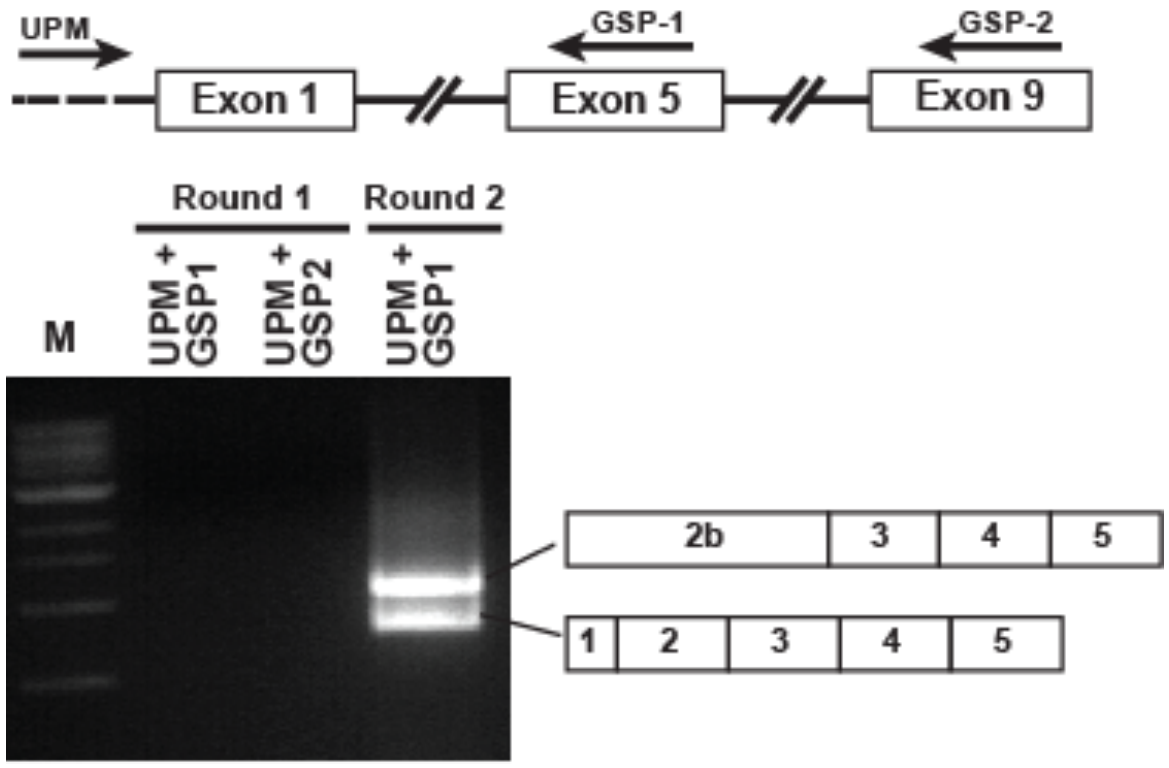
UPM
GSP-1
GSP-2
Exon 1
Exon 5
Exon 9
Round 1
Round 2
M
UPM + GSP1
UPM + GSP2
UPM + GSP1
2b
3
4
5
1
2
3
4
5

Figure S9. 5’ RACE assay validation of exon 2b as alternative start site. The 5’ end of *MFN2* transcripts in platelets were identified using 5’-RACE. Nested PCR using two gene specific primers (GSP-1 and GSP-2) and a universal primer mix (UPM) which anneals to sequence which is added at the 5’ end (dashed line). Sequencing of the two bands that resulted from this reaction indicated that platelets contain two *MFN2* isoforms with two different starting exons, 1 and 2b.

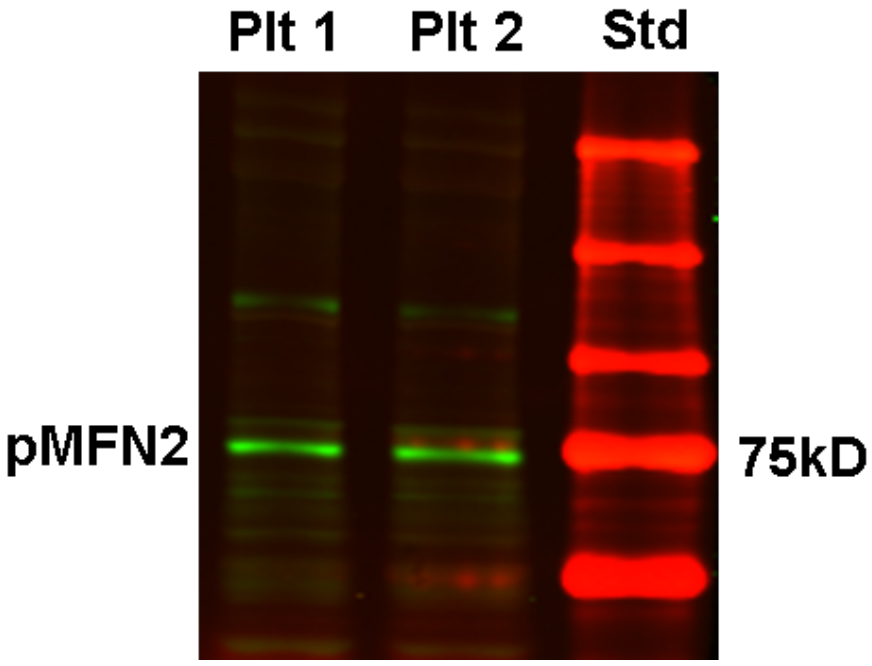
Plt 1
Plt 2
Std
pMFN2
75kD

Figure S10. pMFN2 western blot. Platelet lysates from 2 different donor separated by SDS 7% polyacrylamide get electrophoresis, transferred and probed with affinity purified anti-MFN2 rabbit polyclonal antisera (Sigma #M6319 N-terminal). Molecular weight standards are in rightmost lane.

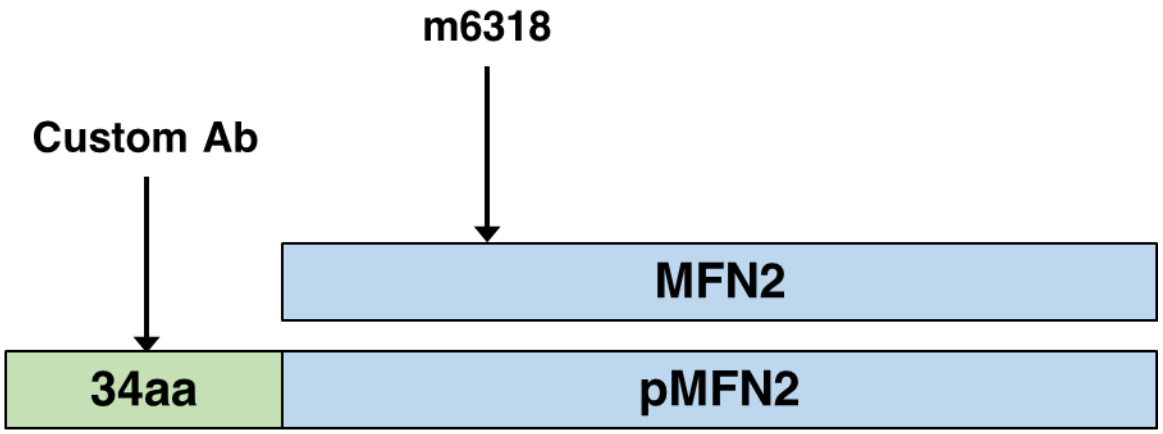
m6318
Custom Ab
MFN2
34aa
pMFN2

Figure S11. Predicted platelet pMFN2 schematic. Cartoon is comparing MFN2 and pMFN2. The blue rectangles represent shared amino acid sequence. Inclusion of exon 2b is predicted to add 34 amino acids (green rectangle) to the N-terminal of MFN2. The custom Ab arrow points to the pMFN2 specific sequence targeted by a custom antibody generated by the Bray lab. The m6318 arrow points to the non-specific target binding site for the Sigma #M6319 antibody.