Atlas of imprinted and allele-specific DNA methylation in the human body

Jonathan Rosenski[1], Ayelet Peretz[2], Judith Magenheim[2], Netanel Loyfer[1], Ruth Shemer[2], Benjamin Glaser[3], Yuval Dor[2,4,*], and Tommy Kaplan[1,2,4,*]

Supplemental Information

Data S1. Genome-wide list of bimodal methylation regions by sample

Data S2. Known imprinted DMR coordinates, with regions of biallelic methylation masked

Data S3. List of SNPs associated with cell-type-specific ASMs

Data S4. Genome-wide list of SNPs associated with parental ASMs

Data S5. Novel parental-ASM participant information

Data S6. Novel parental-ASM locus and primer information

Data S7. Summary data of trio data families genotyping and methylation status.

Data S8. Transcription factor enrichment analysis (ReMap)

Data S9. Chromatin states enrichment analysis (chromHMM)

Data S10. All imprinted genes with parental ASM regions and known imprinted DMRs

Data S11. Genes exhibiting tissue-specific imprinting, tissue-specific bimodality

Data S12. Fold-change biased genes vs expressed genes in relevant tissue

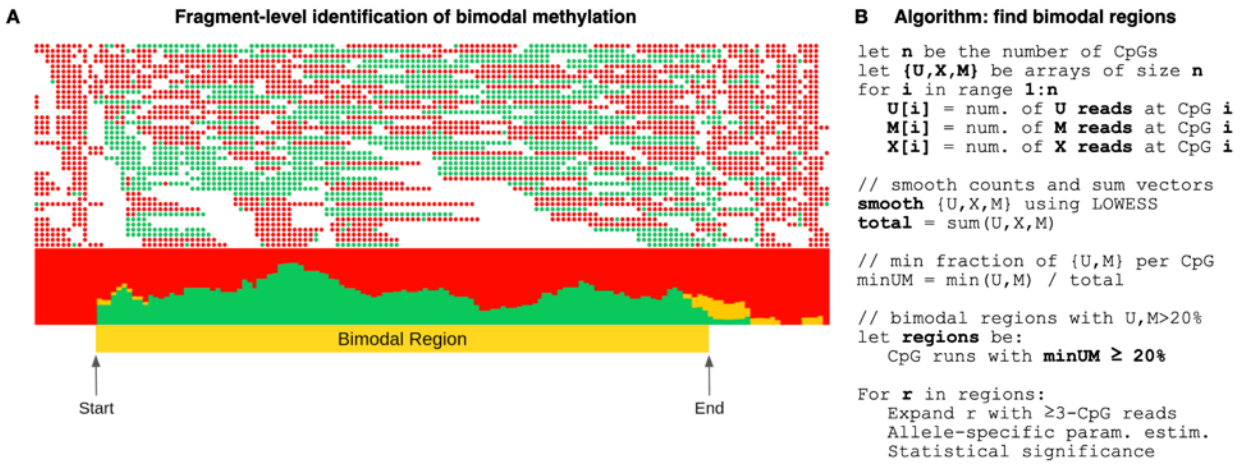
Data S13. Genes found with significant associations by Hofmeister et al.

Data S14. PanelApp Genomic imprinting panel associations with parental ASM

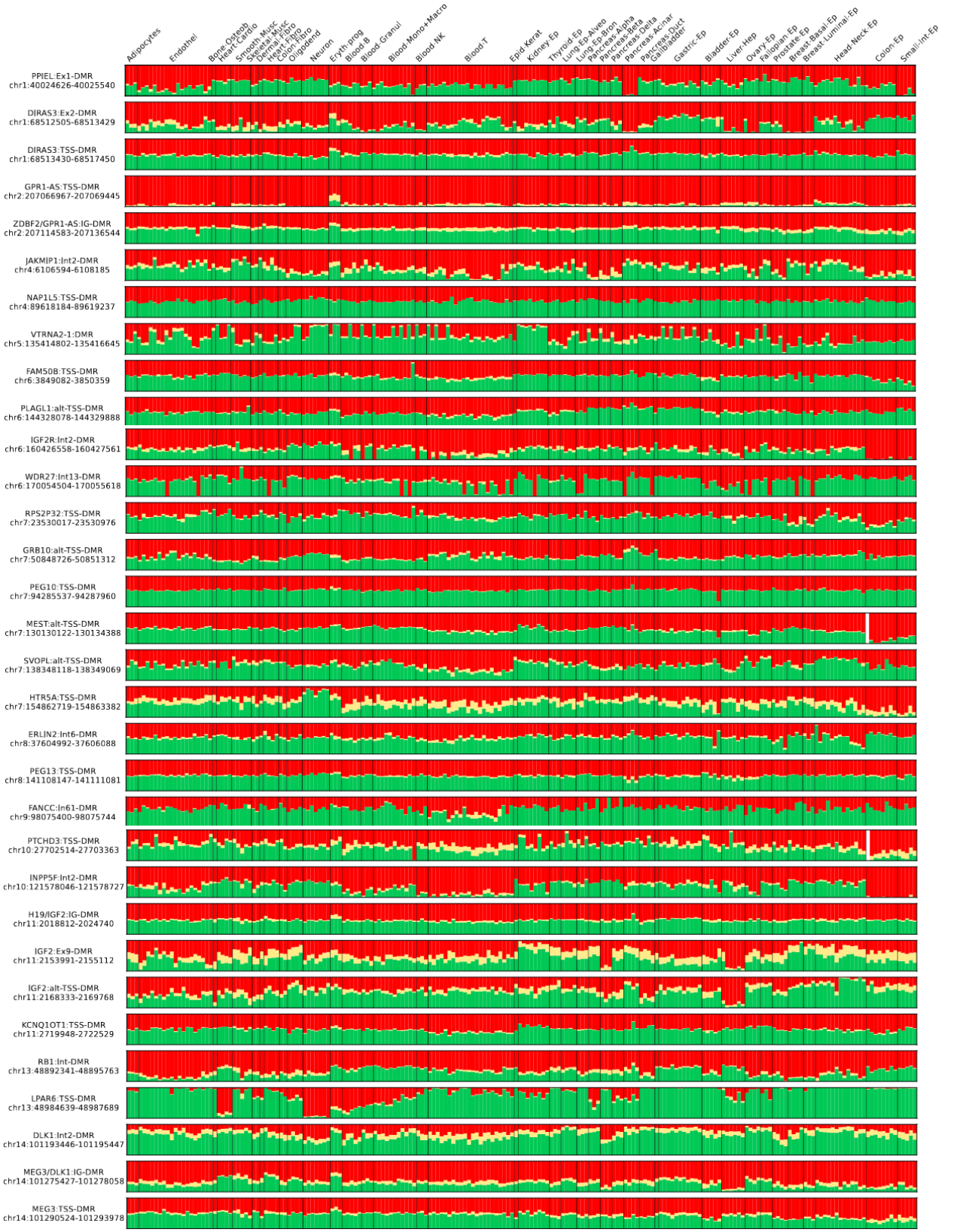
Data S15. Allelically biased genes associated with phenotypes and disease

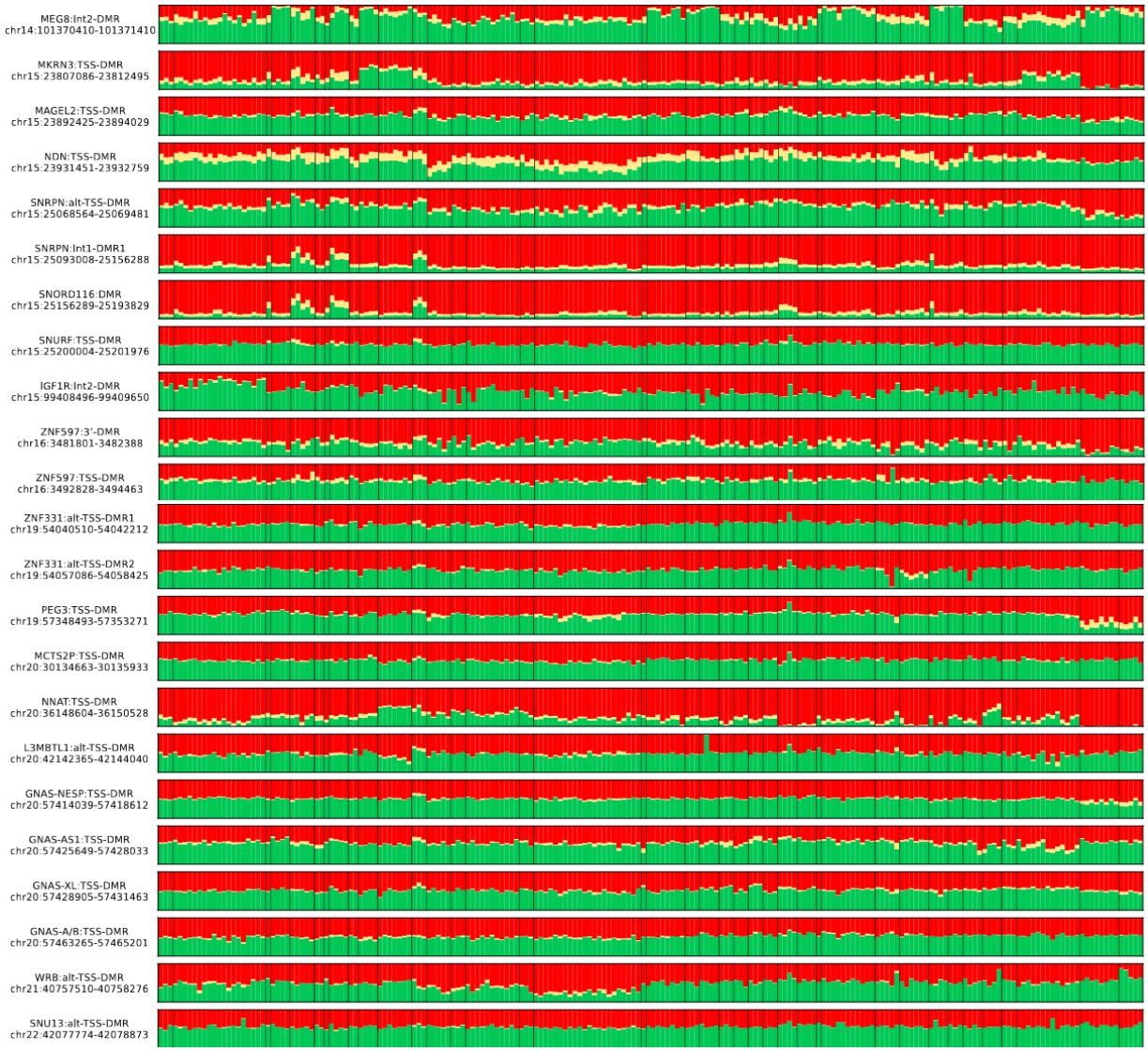
Data S16. Tissue-specific escape from imprinting near tissue-specific unmethylated regions

Data S17. Raw sequencing counts, per trio participant, split by variant allele

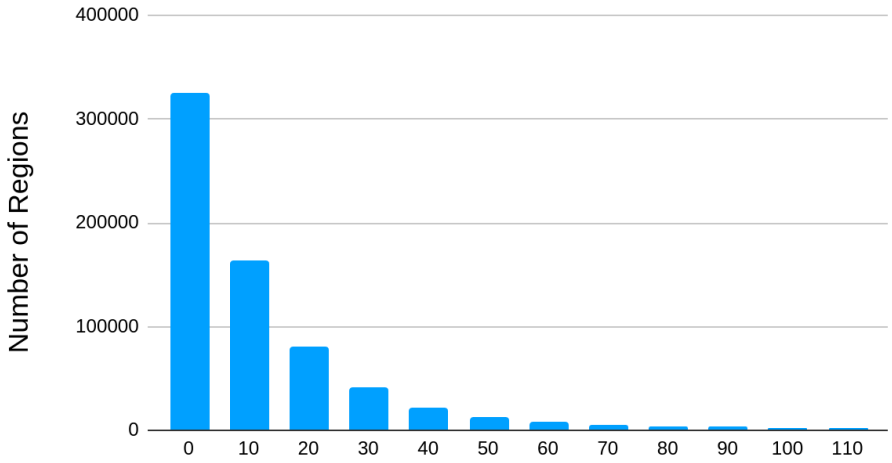


Supplemental Figure 1. Algorithm for the identification of bimodal regions. (A) Sequenced reads are presented as red/green bead strings (for methylation and unmethylated CpGs), separated by white spaces. Reads are classified as **U** (<35% methylated CpGs per fragment), **X** (not **U** or **M**), or **M** (>65% methylated CpGs, per fragment). The figure below shows the cumulative proportion of U/X/M reads across each CpG, with arrows marking where the min(U,M) fragment proportion is above 20%. We classify a region as bimodal if the proportion of both **U** and **M** reads is ≥ 20% across consecutive CpGs (Methods). (**B**) Pseudo-code of the efficient linear-time algorithm which identifies the bimodal regions. The algorithm works as visualized in **A**. The algorithm begins by transforming a WGBS bam file into a 3-dimensional array whereby each entry of the array represents a CpG site. At each index of the array we store how many *U*/*X*/*M* reads intersect this CpG site. We apply LOWESS smoothing on each array individually and we find those regions of contiguous CpG site where the proportion of **U** and **M** reads intersecting that site are at least 20%.

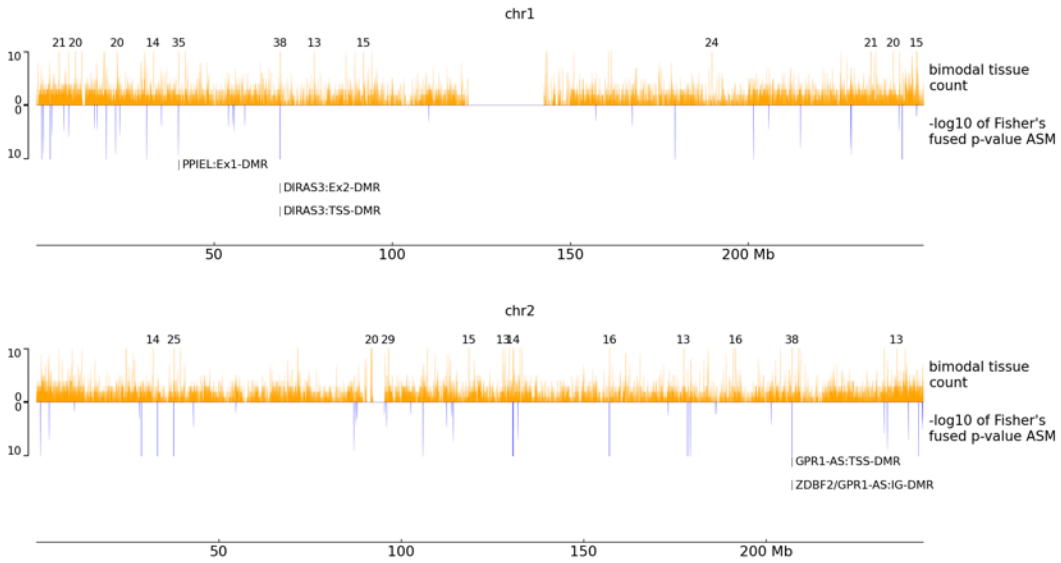

Adipocytes
Endothel
Bone-Osteob
Heart-Cardio
Smooth-Musc
Skeletal-Musc
Dermal-Fibro
Heart-Fibro
Colon-Fibro
Oligodend
Neuron
Eryth-prog
Blood-B
Blood-Granul
Blood-Mono+Macro
Blood-NK
Blood-T
Epid-Kerat
Kidney-Ep
Thyroid-Ep
Lung-Ep-Alveo
Lung-Ep-Bron
Pancreas-Alpha
Pancreas-Beta
Pancreas-Delta
Pancreas-Acinar
Pancreas-Duct
Gallbladder
Gastric-Ep
Bladder-Ep
Liver-Hep
Ovary-Ep
Fallopian-Ep
Prostate-Ep
Breast-Basal-Ep
Breast-Luminal-Ep
Head Neck-Ep
Colon-Ep
Small-Int-Ep
PPIEL:Ex1-DMR
chr1:40024626-40025540
DIRAS3:Ex2-DMR
chr1:68512505-68513429
DIRAS3:TSS-DMR
chr1:68513430-68517450
GPR1-AS:TSS-DMR
chr2:207066967-207069445
ZDBF2/GPR1-AS:IG-DMR
chr2:207114583-207136544
JAKMIP1:Int2-DMR
chr4:6106594-6108185
NAP1L5:TSS-DMR
chr4:89618184-89619237
VTRNA2-1:DMR
chr5:135414802-135416645
FAM50B:TSS-DMR
chr6:3849082-3850359
PLAGL1:alt-TSS-DMR
chr6:144328078-144329888
IGF2R:Int2-DMR
chr6:160426558-160427561
WDR27:Int13-DMR
chr6:170054504-170055618
RPS2P32:TSS-DMR
chr7:23530017-23530976
GRB10:alt-TSS-DMR
chr7:50848726-50851312
PEG10:TSS-DMR
chr7:94285537-94287960
MEST:alt-TSS-DMR
chr7:130130122-130134388
SVOPL:alt-TSS-DMR
chr7:138348118-138349069
HTR5A:TSS-DMR
chr7:154862719-154863382
ERLIN2:Int6-DMR
chr8:37604992-37606088
PEG13:TSS-DMR
chr8:141108147-141111081
FANCC:In61-DMR
chr9:98075400-98075744
PTCHD3:TSS-DMR
chr10:27702514-27703363
INPP5F:Int2-DMR
chr10:121578046-121578727
H19/IGF2:IG-DMR
chr11:2018812-2024740
IGF2:Ex9-DMR
chr11:2153991-2155112
IGF2:alt-TSS-DMR
chr11:2168333-2169768
KCNQ1OT1:TSS-DMR
chr11:2719948-2722529
RB1:Int-DMR
chr13:48892341-48895763
LPAR6:TSS-DMR
chr13:48984639-48987689
DLK1:Int2-DMR
chr14:101193446-101195447
MEG3/DLK1:IG-DMR
chr14:101275427-101278058
MEG3:TSS-DMR
chr14:101290524-101293978

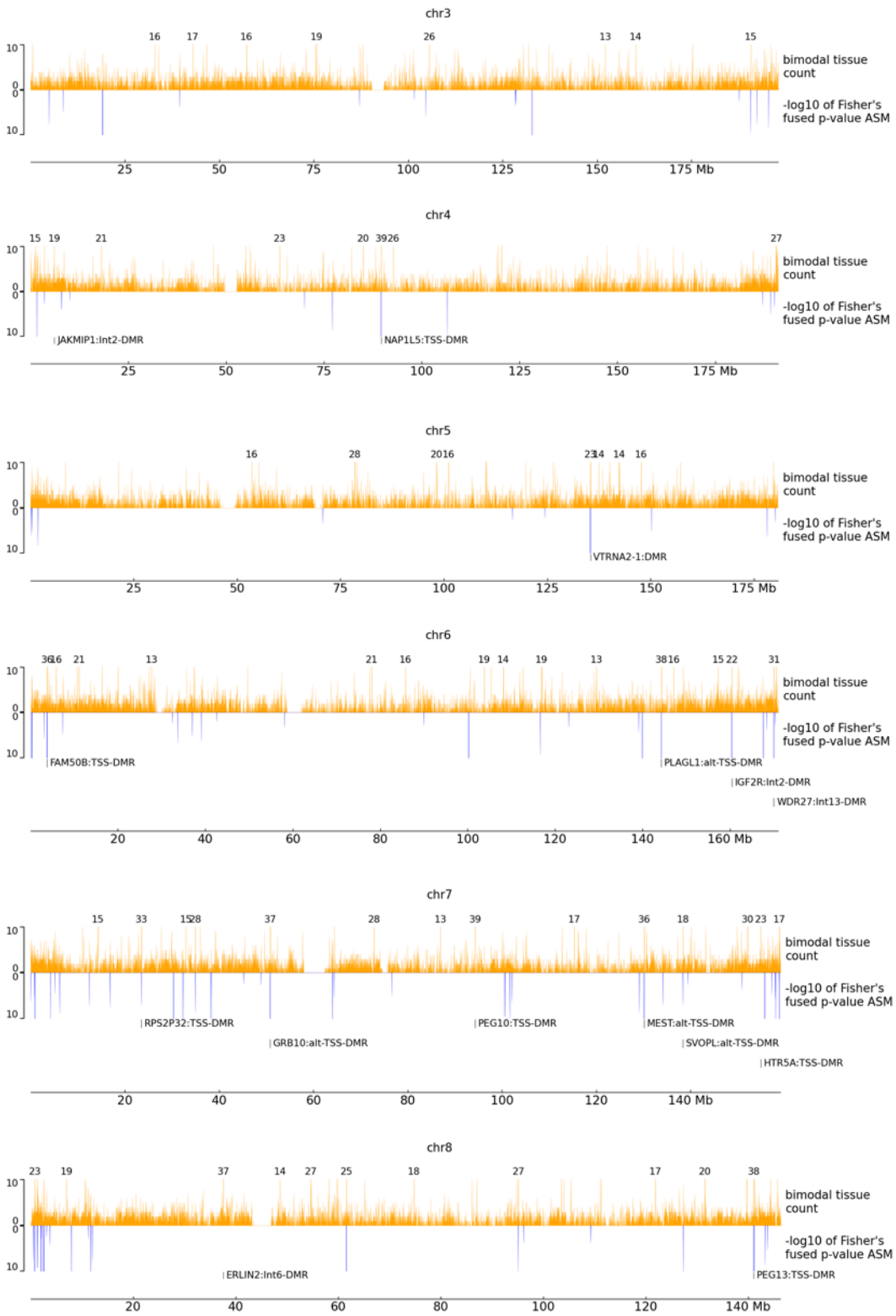


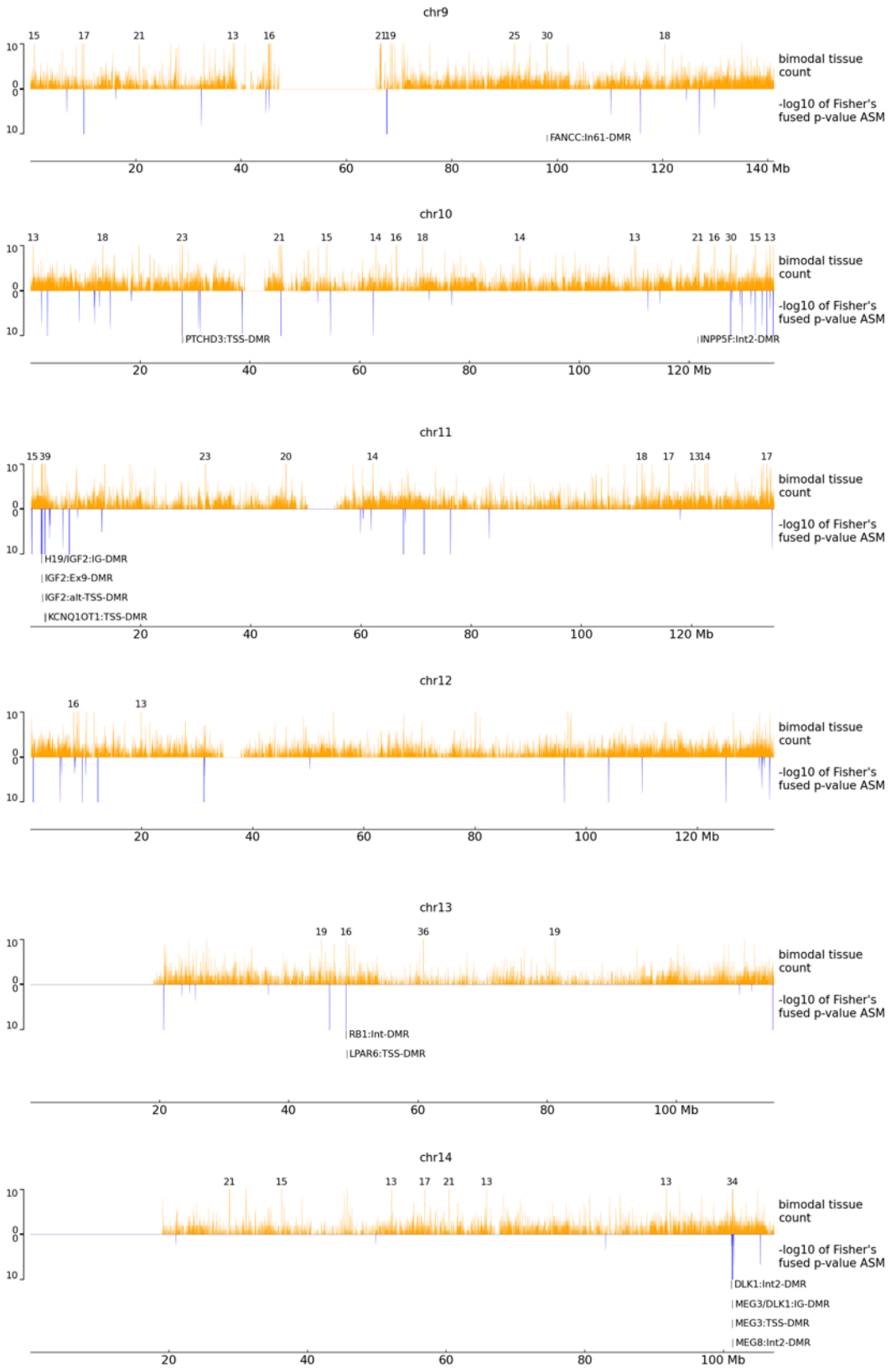
Supplemental Figure 2. Bimodality plot of known imprinted DMRs. For each known imprinted DMR (hg19 coordinates), we plot the percent of U/X/M sequenced fragments (≥3 CpGs) across 202 samples (columns). Fully bimodal patterns are visualized as 50% unmethylated fragments (green) and 50% methylated fragments (red). Sequenced fragments with >35% unmethylated CpGs and >35% methylated CpGs (e.g. a fragment of 4 CpGs, with exactly two methylated CpGs) are classified as mixed, and their empirical percentage in a give sample/region is shown in yellow.

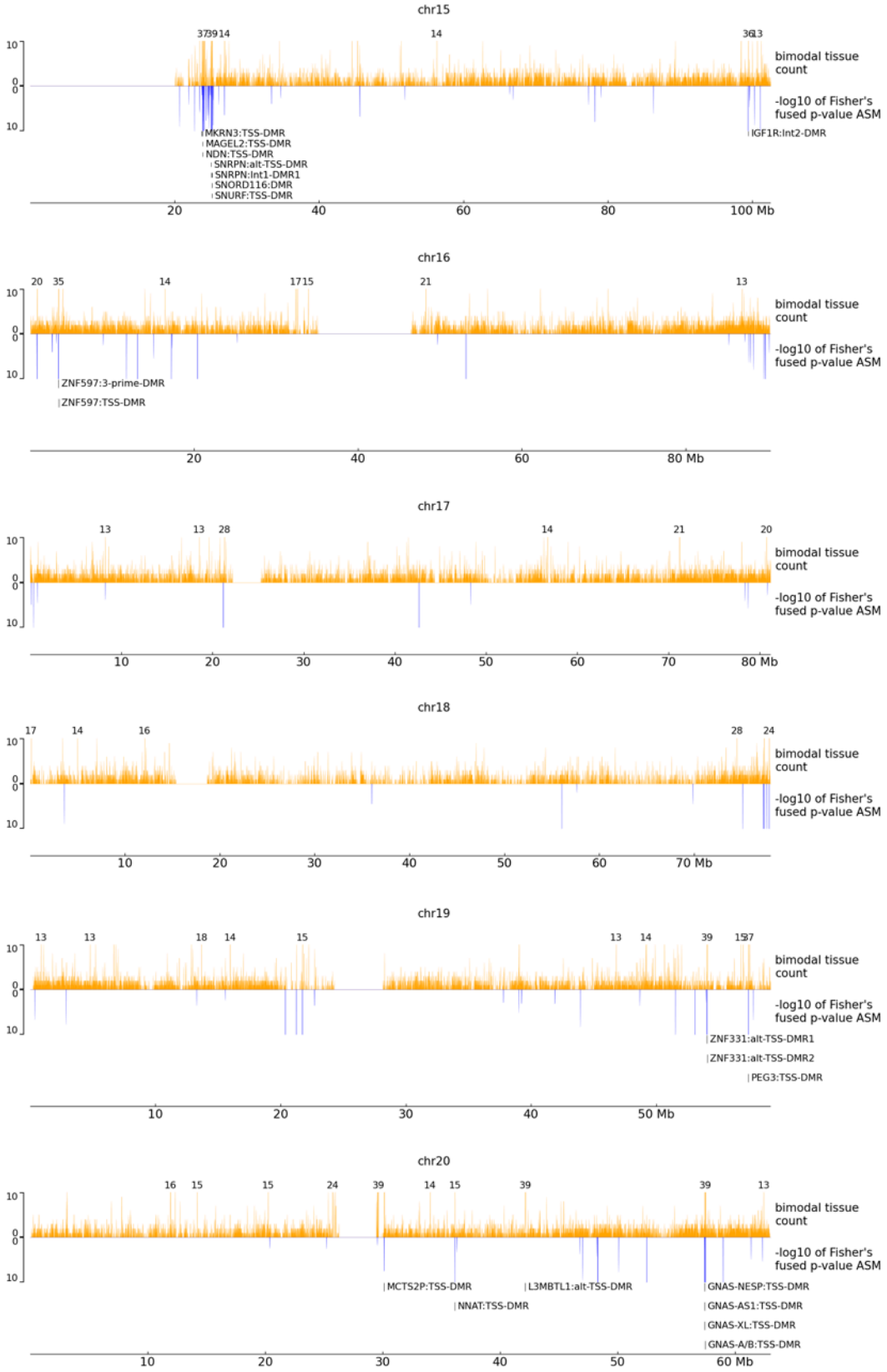


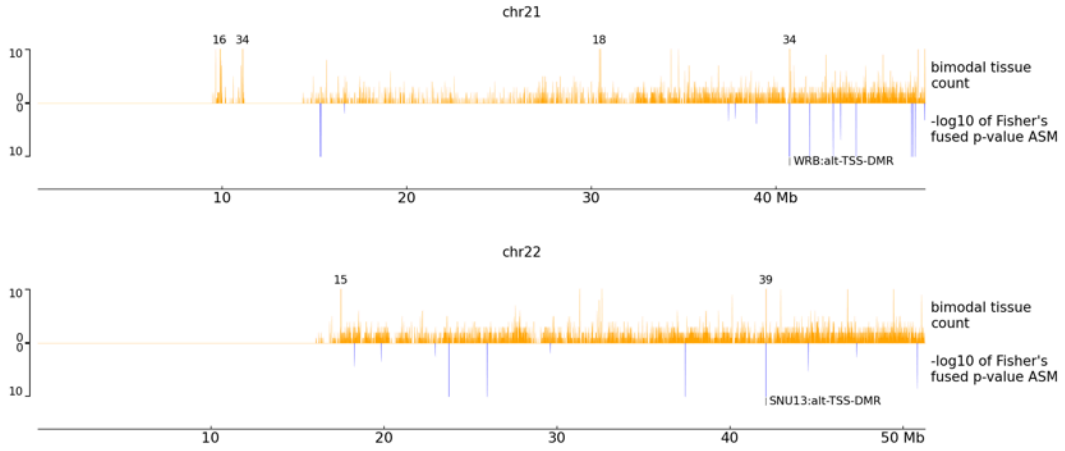
Supplemental Figure 3. Number of regions vs number of samples exhibiting bimodal methylation. Bar chart showing the number of regions (y-axis) which are bimodal in at least the number of samples specified in the x-axis.



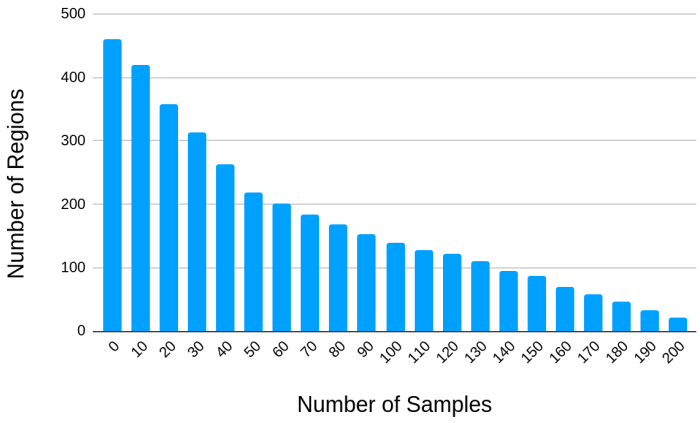

chr3
bimodal tissue count
-log10 of Fisher's fused p-value ASM
25
50
75
100
125
150
175 Mb
chr4
JAKMIP1:Int2-DMR
NAP1L5:TSS-DMR
chr5
VTRNA2-1:DMR
chr6
FAM50B:TSS-DMR
PLAGL1:alt-TSS-DMR
IGF2R:Int2-DMR
WDR27:Int13-DMR
20
40
60
80
100
120
140
160 Mb
chr7
RPS2P32:TSS-DMR
GRB10:alt-TSS-DMR
PEG10:TSS-DMR
MEST:alt-TSS-DMR
SVOPL:alt-TSS-DMR
HTR5A:TSS-DMR
140 Mb
chr8
ERLIN2:Int6-DMR
PEG13:TSS-DMR


chr9
15 17 21 13 16 21 19 25 30 18
FANCC:In61-DMR
20 40 60 80 100 120 140 Mb
chr10
13 18 23 21 15 14 16 18 14 13 21 16 30 15 13
PTCHD3:TSS-DMR
INPP5F:Int2-DMR
20 40 60 80 100 120 Mb
chr11
15 39 23 20 14 18 17 13 14 17
H19/IGF2:IG-DMR
IGF2:Ex9-DMR
IGF2:alt-TSS-DMR
KCNQ1OT1:TSS-DMR
20 40 60 80 100 120 Mb
chr12
16 13
20 40 60 80 100 120 Mb
chr13
19 16 36 19
RB1:Int-DMR
LPAR6:TSS-DMR
20 40 60 80 100 Mb
chr14
21 15 13 17 21 13 13 34
DLK1:Int2-DMR
MEG3/DLK1:IG-DMR
MEG3:TSS-DMR
MEG8:Int2-DMR
20 40 60 80 100 Mb
bimodal tissue count
-log10 of Fisher's fused p-value ASM

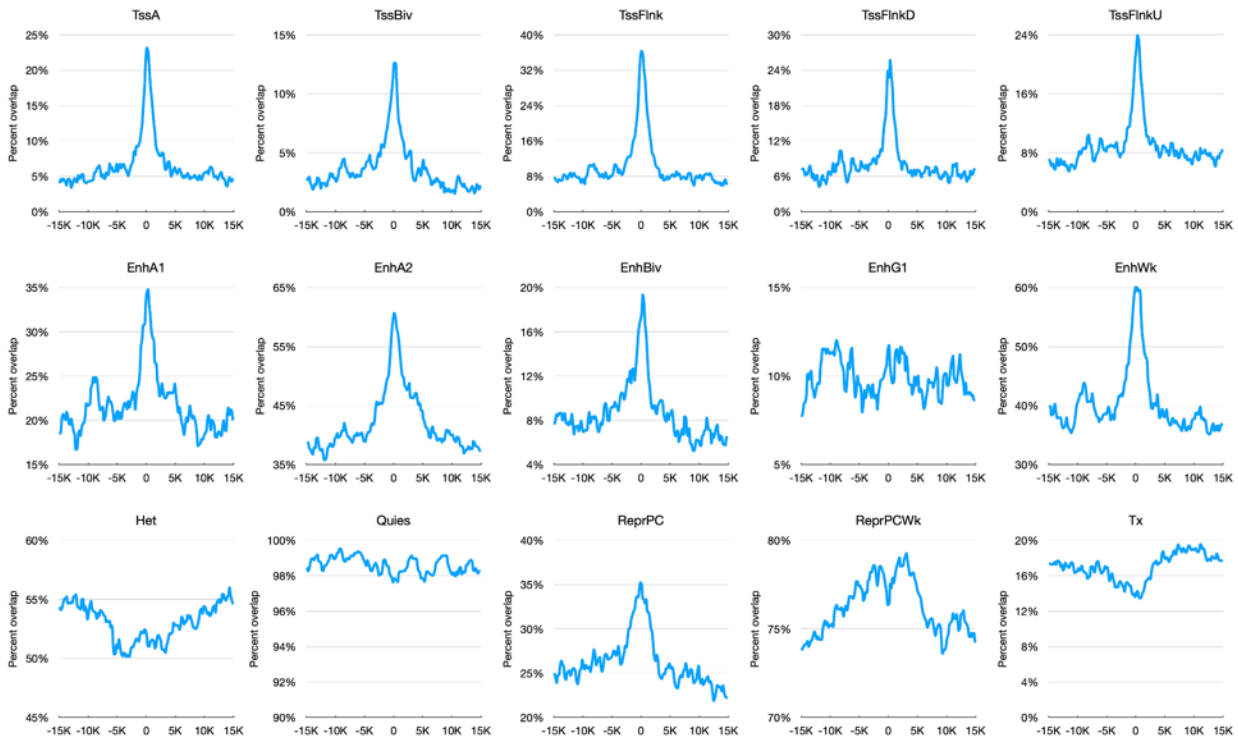

chr15
bimodal tissue count
-log10 of Fisher's fused p-value ASM
MKRN3:TSS-DMR
MAGEL2:TSS-DMR
NDN:TSS-DMR
SNRPN:alt-TSS-DMR
SNRPN:Int1-DMR1
SNORD116:DMR
SNURF:TSS-DMR
IGF1R:Int2-DMR
20
40
60
80
100 Mb
chr16
ZNF597:3-prime-DMR
ZNF597:TSS-DMR
80 Mb
chr17
70 Mb
chr18
chr19
ZNF331:alt-TSS-DMR1
ZNF331:alt-TSS-DMR2
PEG3:TSS-DMR
50 Mb
chr20
MCTS2P:TSS-DMR
NNAT:TSS-DMR
L3MBTL1:alt-TSS-DMR
GNAS-NESP:TSS-DMR
GNAS-AS1:TSS-DMR
GNAS-XL:TSS-DMR
GNAS-A/B:TSS-DMR
60 Mb



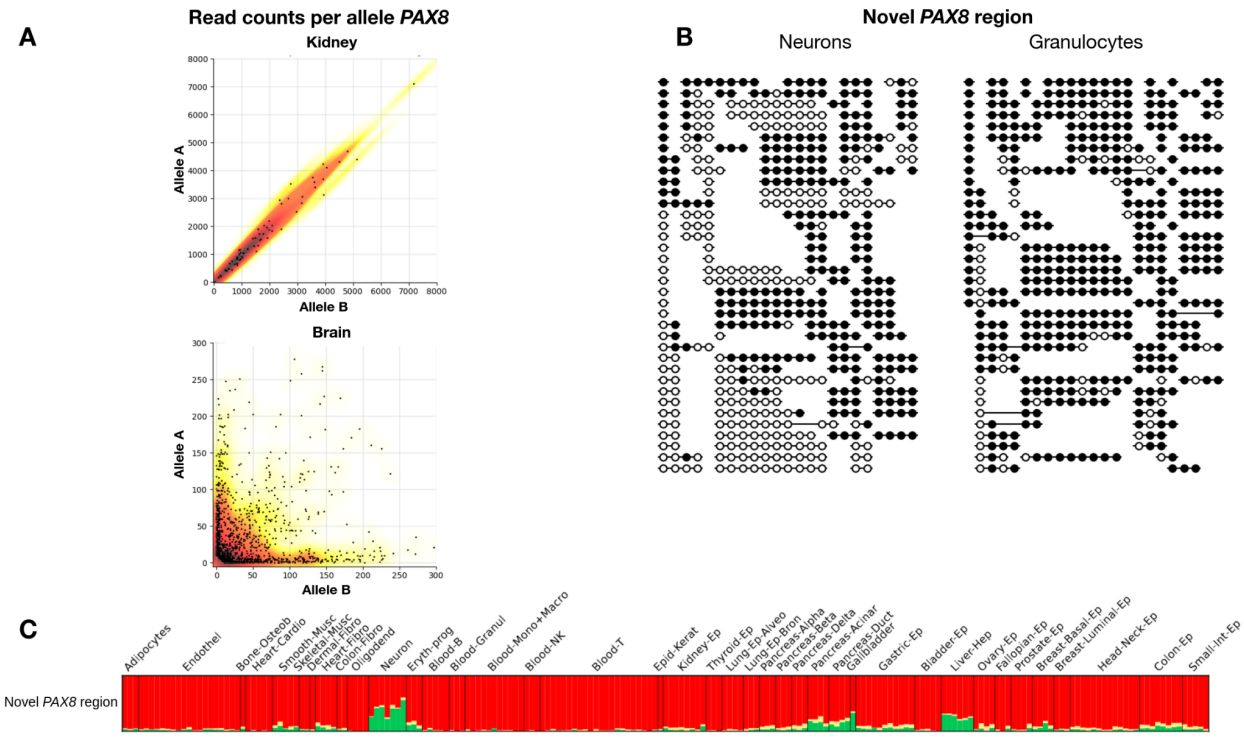
Supplemental Figure 4. Distribution of bimodal regions and ASM across the genome.
In orange, a track showing the number of tissues that are bimodal across the genome. The heights exceeding the threshold are shown above the bars. In blue the negative log10 Fisher's fused p-value of ASM across the genome is shown. Known imprinted DMRs are indicated below each chromosome.



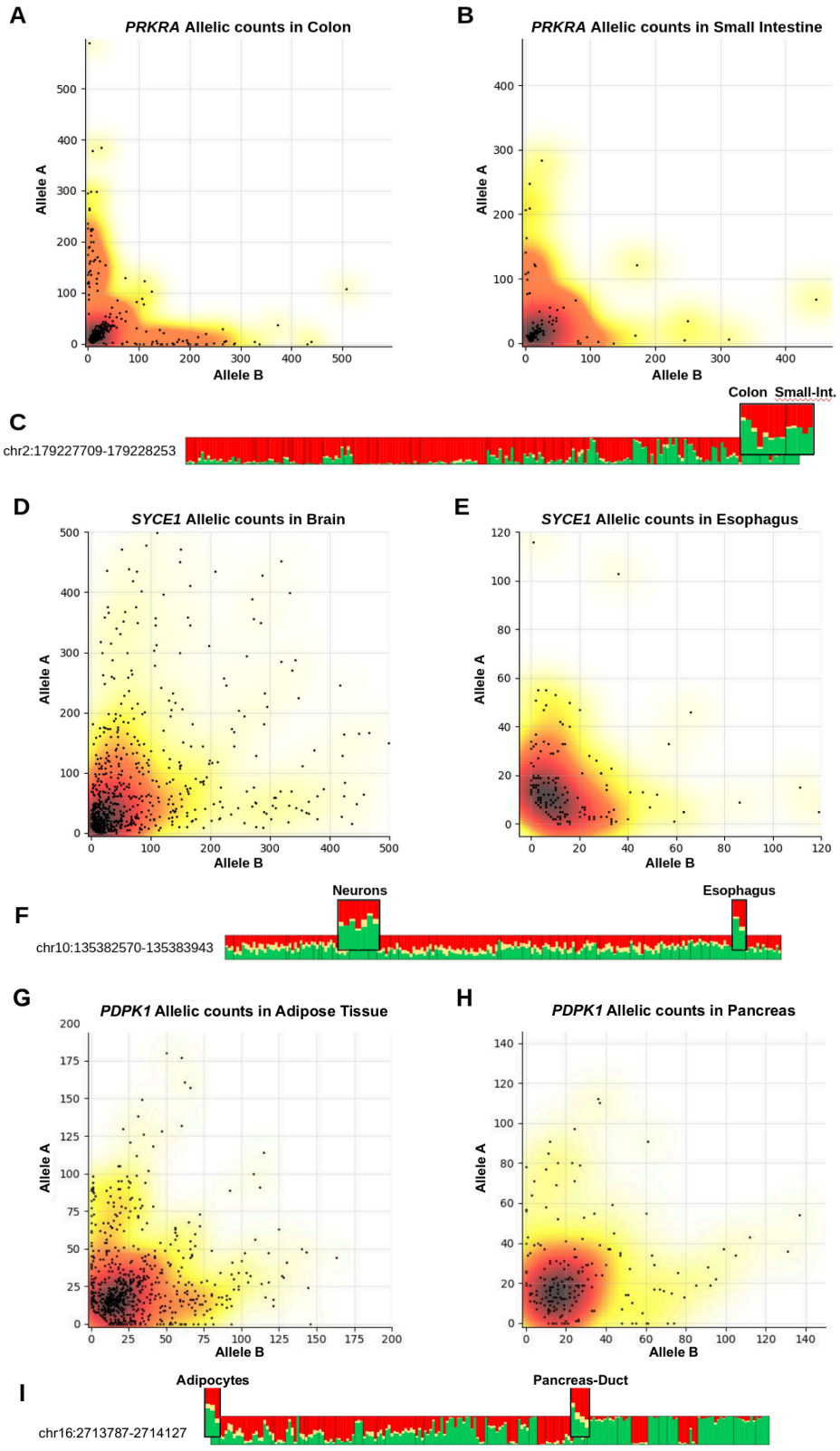
Supplemental Figure 5. Distributional properties of parental ASM regions.
Bar chart showing the number of parental-ASM-regions identified (y-axis) which are bimodal in at least the number of samples specified in the number of samples (x-axis).



Supplemental Figure 6. Local enrichment of parental ASM regions for various chromHMM annotations. Average percent of intersecting nucleotides between parental-ASM regions and various chromHMM state annotations.



Supplemental Figure 7. Tissue-specific allelic expression of the known imprinted gene *PAX8* is accompanied by matched bimodal methylation. (A) Allele-specific read counts from GTEx of the known imprinted gene *PAX8* show imprinted expression in kidney samples (top) and brain samples (bottom). Kidney samples exhibit biallelic expression (along the diagonal indicating both alleles are highly expressed) while brain samples exhibit allele-specific expression (either allele A along the y-axis or allele B along the x-axis). **(B)** Bimodal DNA methylation in purified cortical neurons (left) but not in other samples (e.g. granulocytes, right) at a bimodal region (chr2:113953706-113955952), 17Kb upstream of *PAX8*). **(C)** Fragment-level methylation analysis exhibits brain-specific bimodality.



Supplemental Figure 8. Density plots for allelic read counts of genes near parental-ASM regions Allele-specific read counts (GTEx) for genes near parental ASM regions show many samples of allele-specific expression of either the A (X-axis) or B (Y-axis) alleles. **(A)** *PRKRA* allelic expression in

Colon with top ranked eQTL at rs2288329 (chr2:179308502, hg19) FDR-corrected p-value ≤0.13 (colon) and 0.33 (small intestine). Parental ASM identified 60Kb downstream, chr2:179227286-179228253. **(D-F)** *SYCE1* allelic expression in the brain (top ranked eQTL rs1536826, chr10:135357239), p-value≤1e-3. Parental ASM identified at chr10:135382570-135385776, as well as 20Kb downstream at chr10:135340445-135344281. **(G-I)** *PDPK1* allelic expression in adipose tissue and pancreas (top ranked eQTL rs61608337, chr16:2538109) with FDR corrected p-value of 9e-3 (adipose) and p≤0.015 (pancreas). Parental ASM identified at chr16:2713787-2714127 (65Kb upstream).