

1 **Deconvolution of DNA methylation identifies differentially methylated gene regions**  
2 **on 1p36 across breast cancer subtypes**

3

4 Alexander J. Titus<sup>1,a,d</sup>, Gregory P. Way<sup>1,b</sup>, Kevin C. Johnson<sup>c,d</sup>, and Brock C.  
5 Christensen<sup>\*d,e,f</sup>

6

7 <sup>1</sup>Authors contributed equally

8 <sup>a</sup> Program in Quantitative Biomedical Sciences, Geisel School of Medicine at Dartmouth,  
9 Hanover, NH 03755

10 <sup>b</sup> Genomics and Computational Biology Graduate Program, University of Pennsylvania,  
11 Philadelphia, PA 19104

12 <sup>c</sup> The Jackson Laboratory for Genomic Medicine, Farmington, CT 06032, USA

13 <sup>d</sup>Department of Epidemiology,

14 <sup>e</sup> Department of Molecular and Systems Biology

15 <sup>f</sup> Department of Community and Family Medicine, Geisel School of Medicine at  
16 Dartmouth, Hanover, NH 03755

17

18 CORRESPONDING AUTHOR:

19 \* 1 Medical Center Drive, Williamson Level 6, HB7650, Lebanon, NH 03766; Phone:  
20 603-650-1828; Fax: 603-650-1840 Brock.Christensen@Dartmouth.edu

21

22 AUTHOR EMAILS:

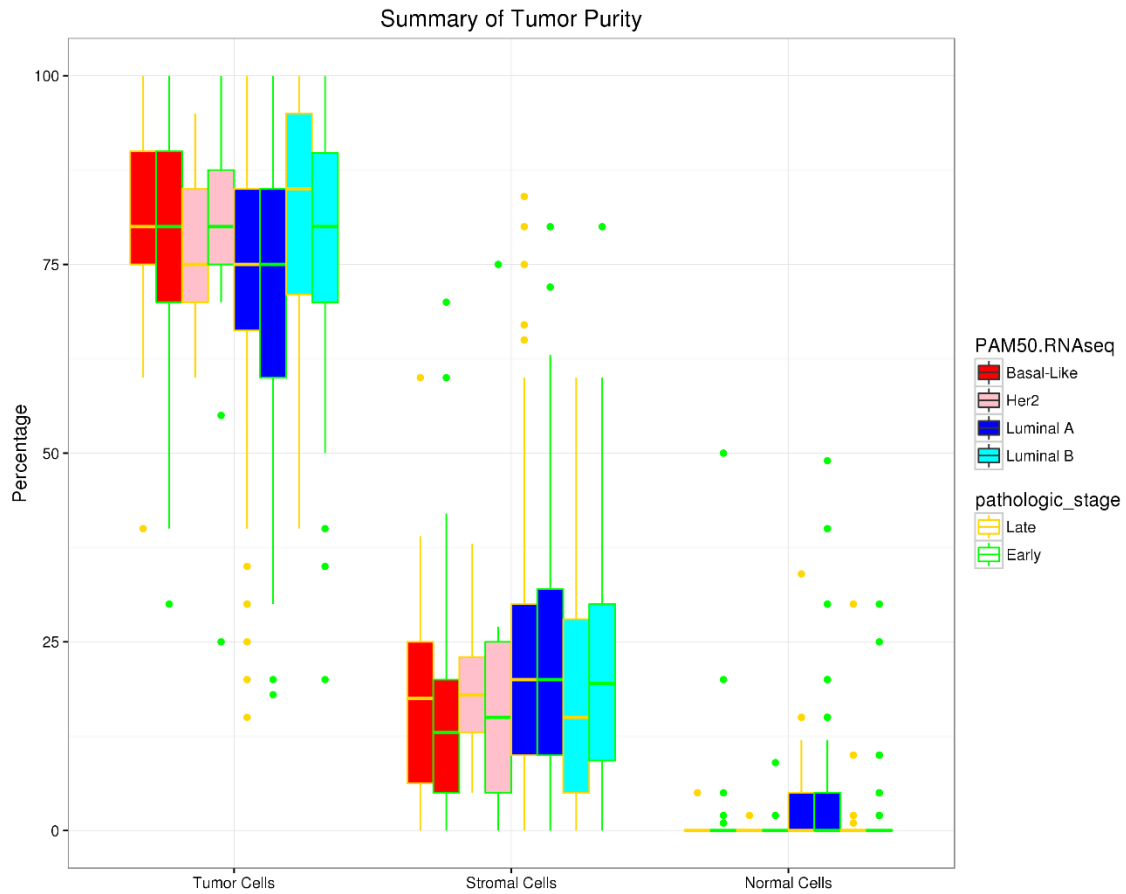
23 AT: Alexander.J.Titus.gr@dartmouth.edu

24 GW: GregWay@upenn.edu

25 KJ: Kevin.C.Johnson@jax.org

26 BC: Brock.Christensen@dartmouth.edu

27 SUPPLEMENTAL FIGURES:

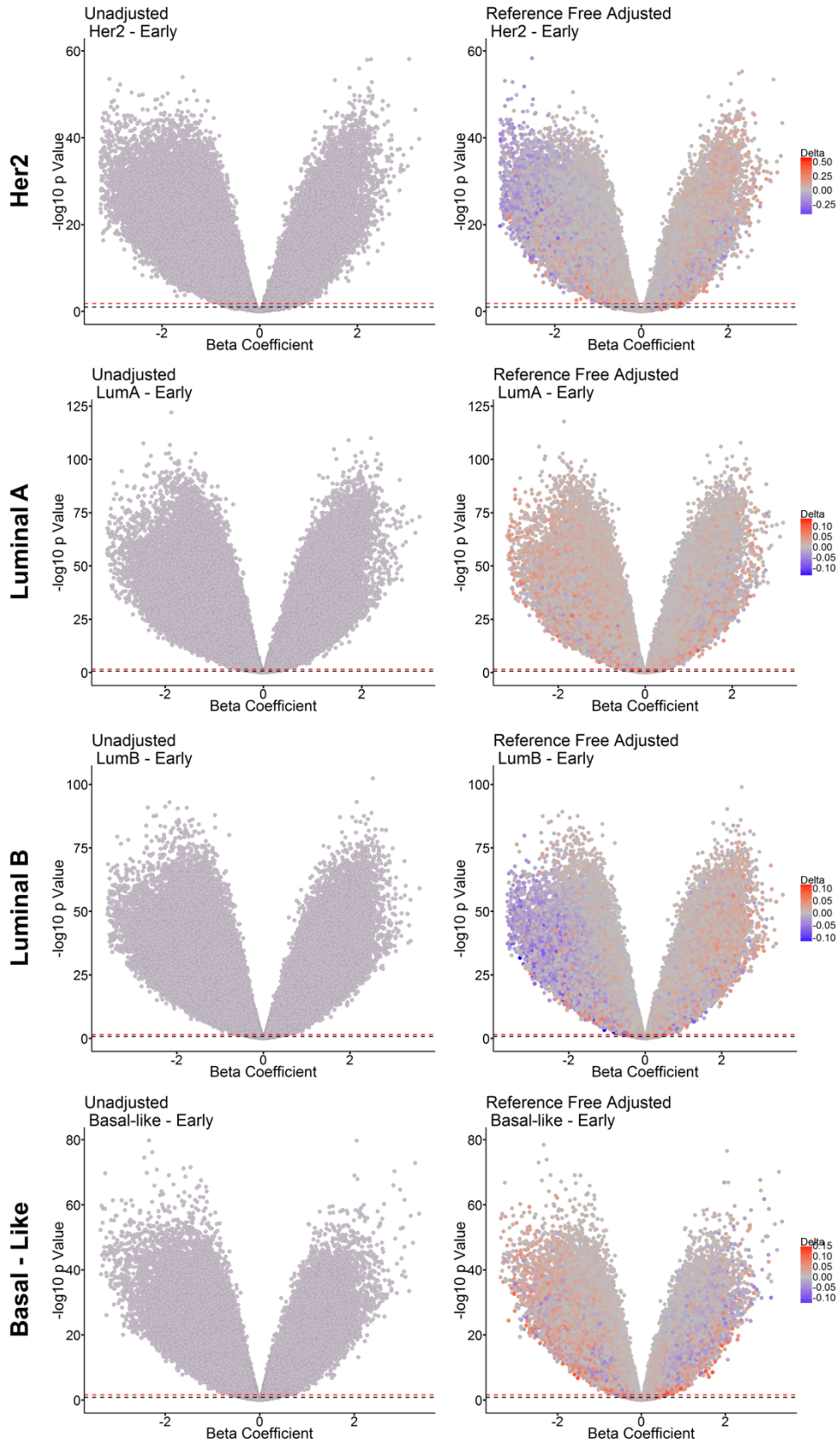


28

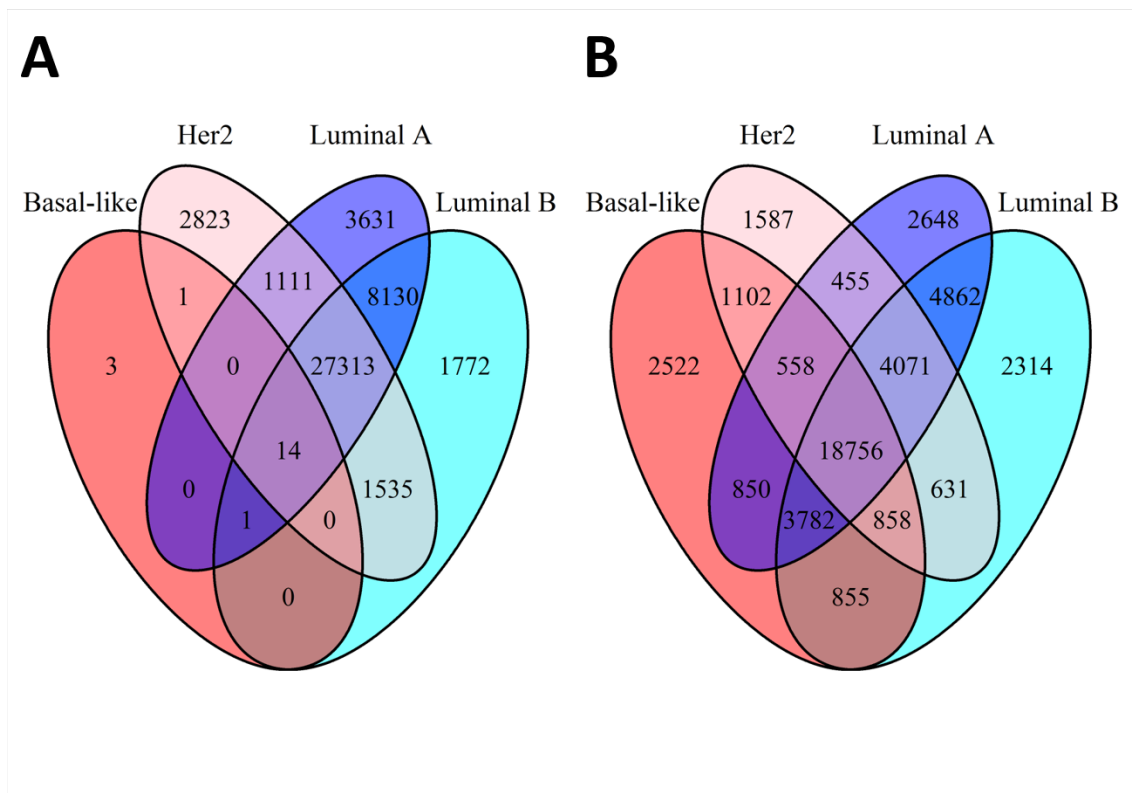
29 *Supplementary Figure S1. Box plots show the distribution of tumor purity across all*  
30 *subtypes for both early and late stages of the TCGA dataset. The measurements estimated*  
31 *by TCGA are based on histology slides and indicate the estimated distribution of the*  
32 *number of tumor cells, stromal cells, and normal cells in each sample. See the NCI CDE*  
33 *Browser for more details.*

34

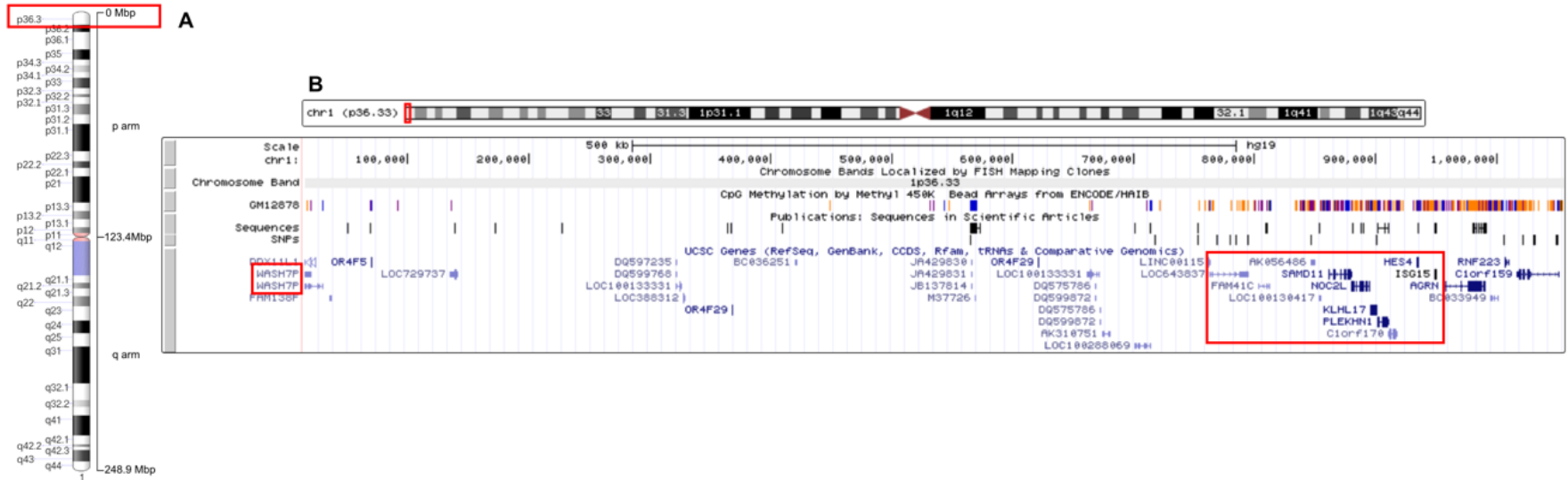
Early stage



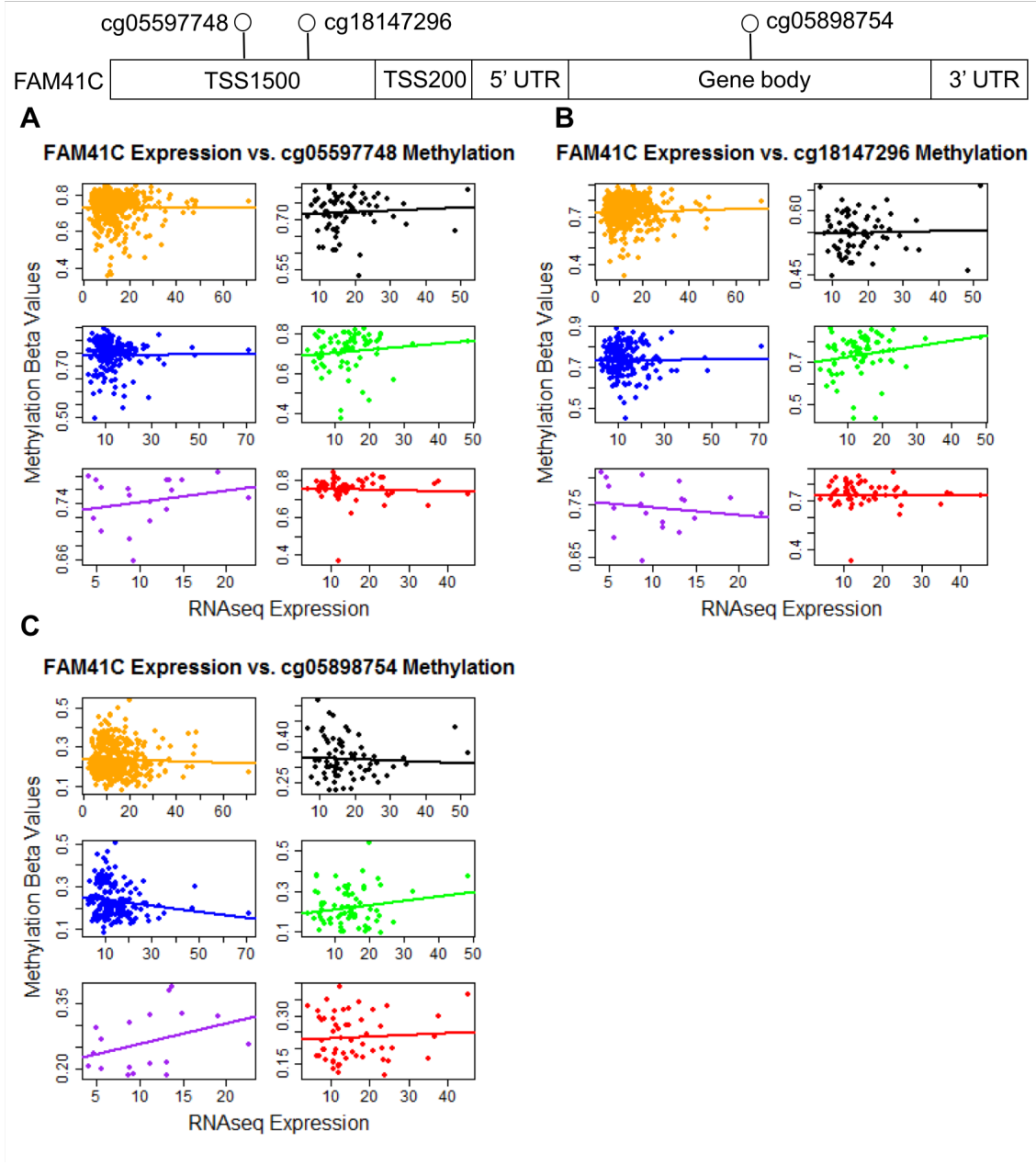
36 **Supplementary Figure S2.** Volcano plots from the early-stage models. The left most  
 37 panel in each model indicates unadjusted *P* values and the right panel indicates RefFree  
 38 adjusted *P* values. Each point represents a CpG considered in the model and the color of  
 39 the points represents the change in the beta coefficient following adjustment (delta  
 40 value). The red lines indicate a *Q* value cutoff of 0.01 and the black lines indicate a *Q*  
 41 value cutoff of 0.05.  
 42



43  
 44 **Supplementary Figure S3.** Venn diagram depicting overlapping Illumina annotation file  
 45 UCSC regions between (A) early and (B) late stage tumors stratified by subtype. The  
 46 regions consist of mappings relative to CpG island definitions (e.g. <Gene Name>  
 47 *N\_Shore*).

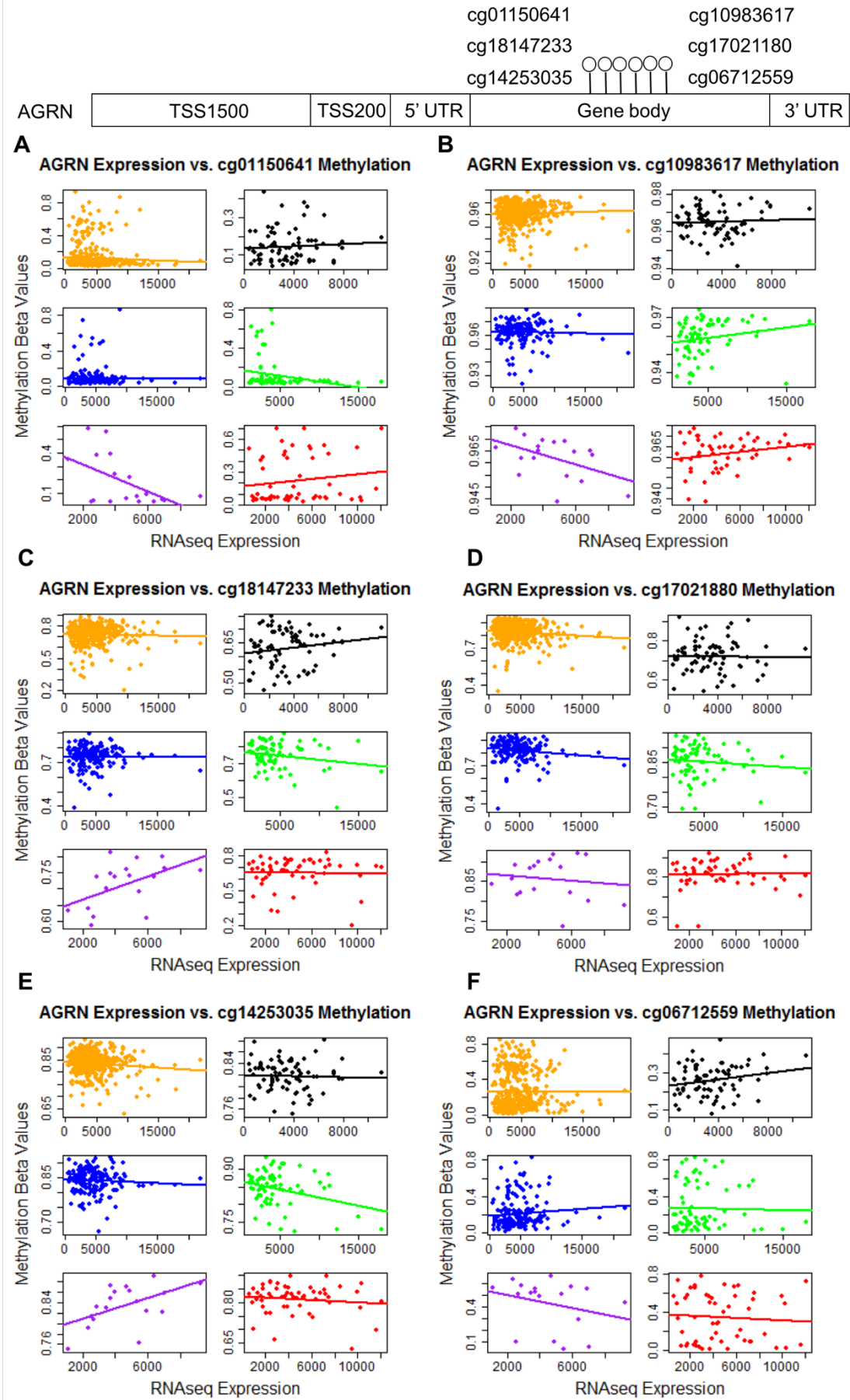


50 *Supplementary Figure S4. Diagram of chromosome 1. (A) The entire chromosome 1 with regions annotated. (B) A zoomed in view of*  
 51 *chromosome 1p36.3 with each identified gene annotated on a track and highlighted in red boxes indicating a gene cluster between*  
 52 *base pairs 868,071 - 1,056,116.*

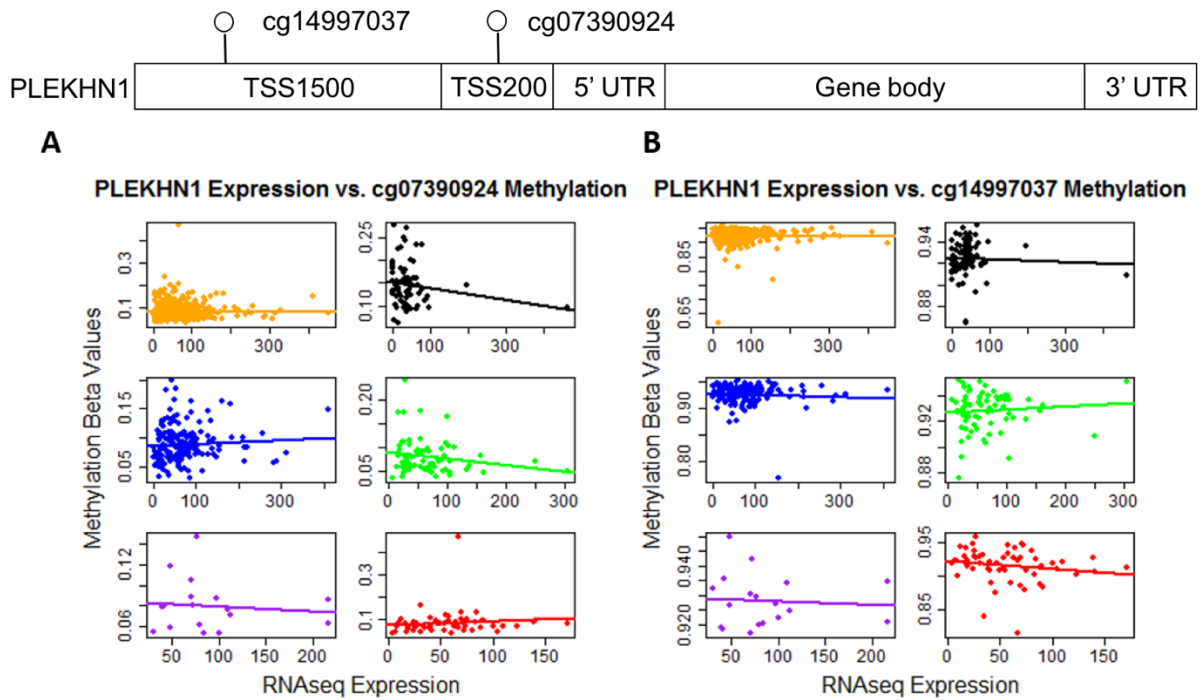


53

54 **Supplementary Figure S5.** *The relationship between differentially methylated CpG sites*  
 55 *and FAM41C gene expression in early stage tumors and normal tissue with matched*  
 56 *RNAseq samples stratified by PAM50 subtype. All tumors (orange), all normal tissue*  
 57 *(black), Luminal A (blue), Luminal B (green), Her2 (purple), and Basal-like (red) are*  
 58 *given in the different facets of the figure. Above the figure is a diagram of which DMGR*  
 59 *the CpG sites are located within.*

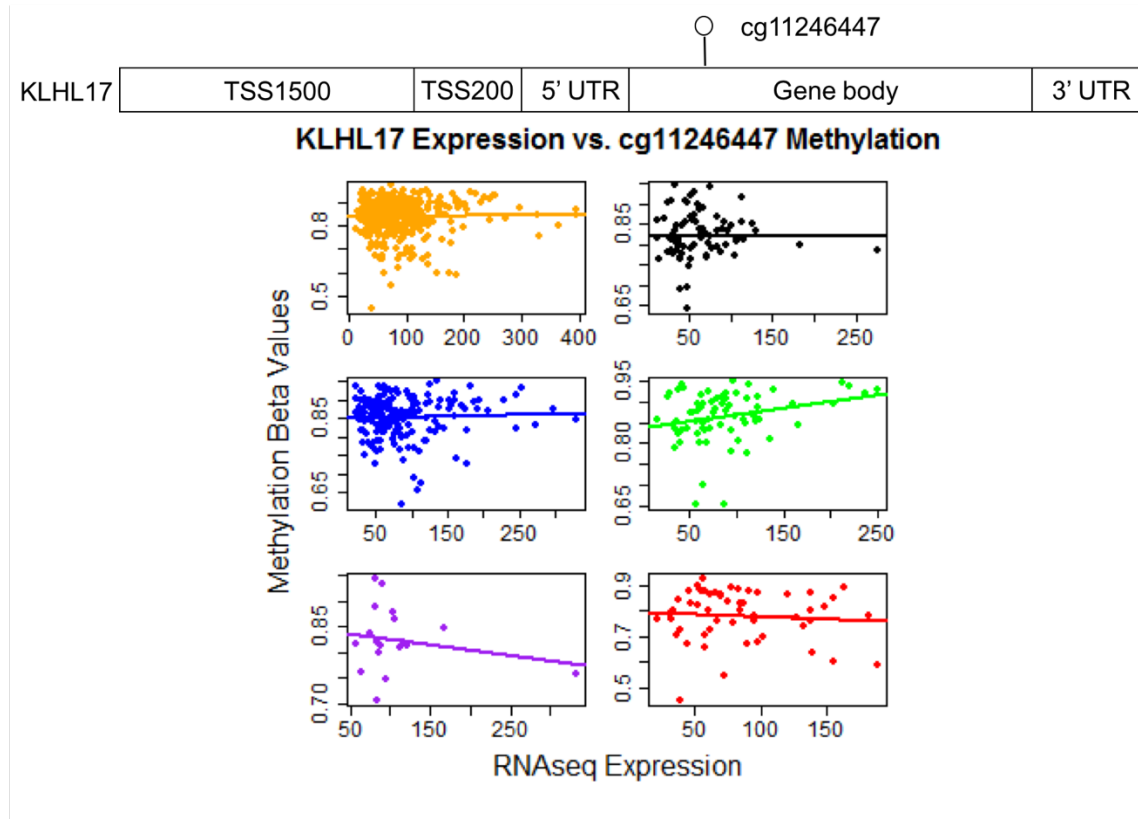


61 **Supplementary Figure S6.** The relationship between differentially methylated CpG sites  
 62 and *AGRN* gene expression in early stage tumors and normal tissue with matched  
 63 RNAseq samples stratified by PAM50 subtype. All tumors (orange), all normal tissue  
 64 (black), Luminal A (blue), Luminal B (green), Her2 (purple), and Basal-like (red) are  
 65 given in the different facets of the figure. Above the figure is a diagram of which DMGR  
 66 the CpG sites are located within.  
 67



68  
 69 **Supplementary Figure S7.** The relationship between differentially methylated CpG sites  
 70 and *PLEKHN1* gene expression in early stage tumors and normal tissue with matched  
 71 RNAseq samples stratified by PAM50 subtype. All tumors (orange), all normal tissue  
 72 (black), Luminal A (blue), Luminal B (green), Her2 (purple), and Basal-like (red) are  
 73 given in the different facets of the figure. Above the figure is a diagram of which DMGR  
 74 the CpG sites are located within.

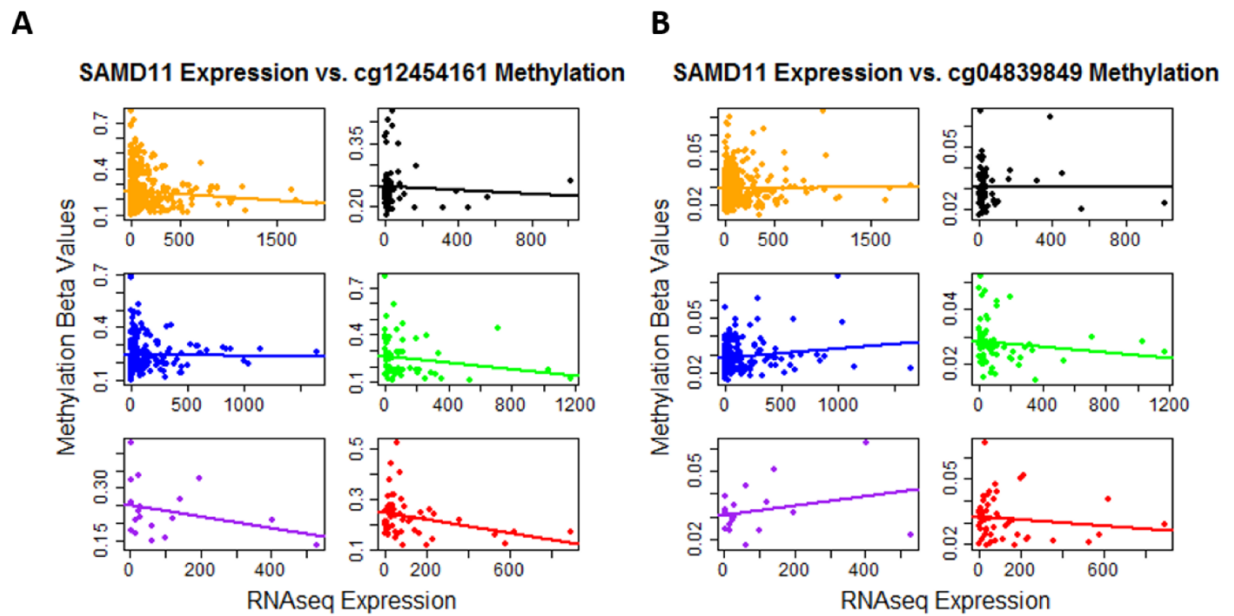




76

77 **Supplementary Figure S8.** The relationship between differentially methylated CpG sites  
 78 and KLHL17 gene expression in early stage tumors and normal tissue with matched  
 79 RNAseq samples stratified by PAM50 subtype. All tumors (orange), all normal tissue  
 80 (black), Luminal A (blue), Luminal B (green), Her2 (purple), and Basal-like (red) are  
 81 given in the different facets of the figure. Above the figure is a diagram of which DMGR  
 82 the CpG sites are located within.

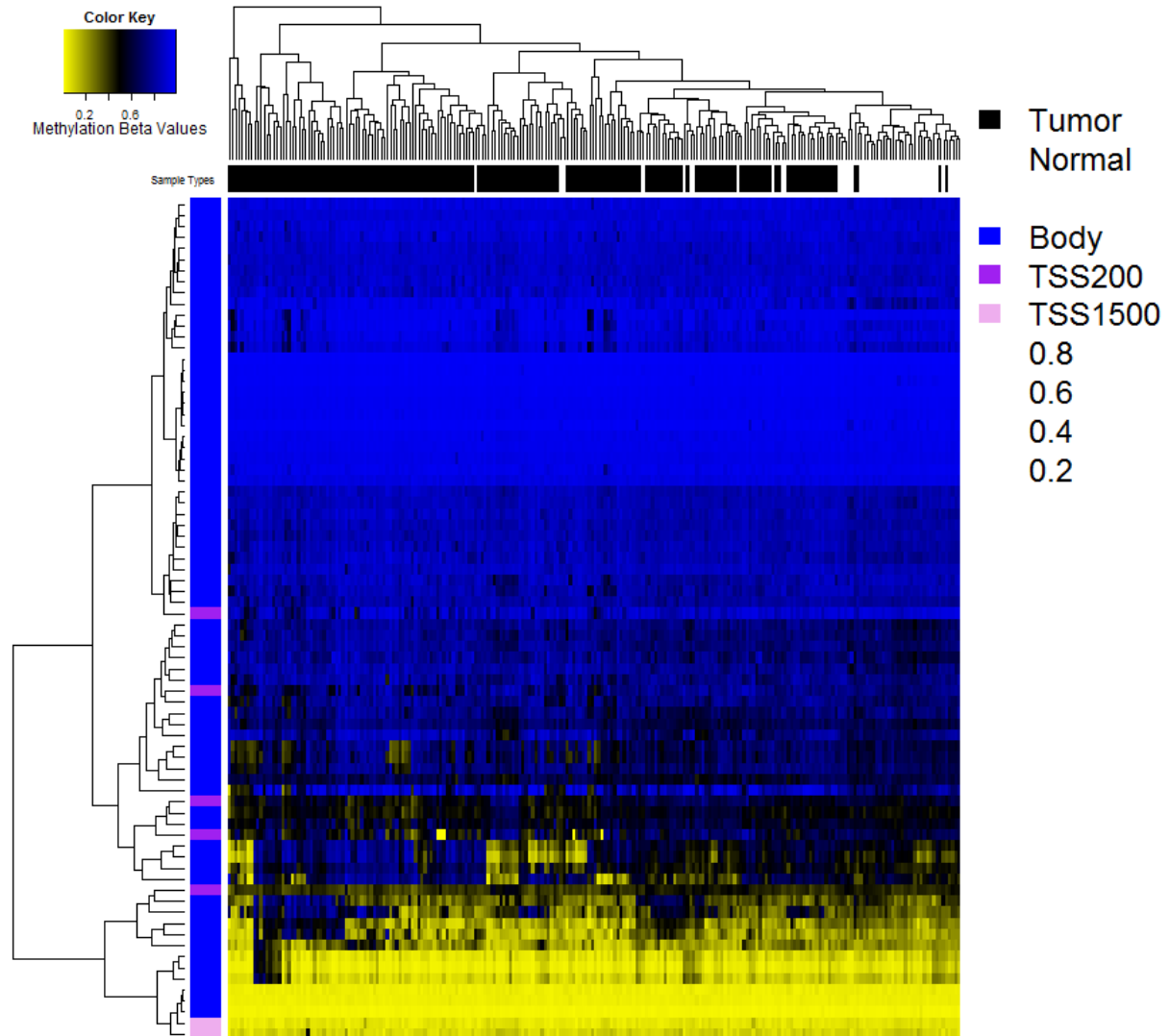
83



84

85 **Supplementary Figure S9.** *The relationship between differentially methylated CpG sites*  
 86 *and SAMD11 gene expression in early stage tumors and normal tissue with matched*  
 87 *RNaseq samples stratified by PAM50 subtype. All tumors (orange), all normal tissue*  
 88 *(black), Luminal A (blue), Luminal B (green), Her2 (purple), and Basal-like (red) are*  
 89 *given in the different facets of the figure. Above the figure is a diagram of which DMGR*  
 90 *the CpG sites are located within.*

91



92

93 **Supplementary Figure S10.** Results from the validation set (Fleischer et al 2014;  
 94 GSE60185). Validation set raw (unadjusted) beta value heatmap of the significantly  
 95 differentially methylated CpG sites in the common early stage differentially methylated  
 96 gene regions (DMGRs) identified in the initial analysis. The genomic context is given in  
 97 the vertical color bar (blue = gene body, dark pink = TSS200, light pink = TSS1500) and  
 98 tumor vs. normal status is given in the horizontal color bar (black = tumor, white =  
 99 normal tissue). In the heatmap, yellow indicates low methylation and blue indicates high  
 100 methylation.

101 SUPPLEMENTAL TABLES:

102 Due to size limitations of this document and the size of the supplemental tables available

103 for this manuscript, supplemental tables may be found at the following DOI link:

104 DOI: [10.5281/zenodo.400247](https://doi.org/10.5281/zenodo.400247)