Supporting information for:

The distinctive topology of age-associated epigenetic drift in the human interactome

James West [*,†,‡], **Martin Widschwendter** [§] **and Andrew E. Teschendorff** [*]

[*] Statistical Cancer Genomics, Paul O'Gorman Building, UCL Cancer Institute, University College London, 72 Huntley Street, LondonWC1E6BT, United Kingdom, [†] Department of Computer Science, University College London, London WC1E 6BT, United Kingdom, [‡] Centre for Mathematics and Physics in the Life Sciences and Experimental Biology, University College London, London WC1E 6BT, United Kingdom and [§] Institute for Women's Health, University College London, London WC1E 6BT, United Kingdom.

SUPPLEMENTARY MATERIALS AND METHODS:

DNA methylation data sets: The DNA methylation data sets used in this study are summarized in Table S1 and were generated using either the Illumina Infinium 27k platform [1] or the Illumina Infinium 450k platform [2]. Data sets were subjected to quality control procedures, including P-value histograms and singular value decompositions to check for potential confounders.

Definitions of transcription factors and kinases: The transcription factor and kinase gene classes used were obtained from the transcriptional regulation atlas [3] and the March 2012 Molecular Signatures database (MSigDB) entry ``Kinases''**.**

Overlap analysis: The significance of overlaps was checked using Fisher's exact test. For all gene classes other than GAMPs, the space of genes was taken to be all Entrez IDs in MSigDB. In the case of the GAMP class, all ~14,000 genes annotated on the Illumina Infinium 27k platform were taken as the reference space.

Copy Number Aberrations: The class of aging genes associated with copy number aberrations (including gain, loss, and neutral) was obtained from the 26,136 cancer-free controls analyzed in Laurie et al [4]. All DNA sources were used, including blood, buccal, and other tissues. The changes identified in the data were split into 5-year bins, and genes in any of these bins were merged together to form the class of copy-number–related aging genes. The median starting age for a bin was 70. Upon mapping to the PIN, the 9,039 genes identified restricted to 3,668 distinct Entrez IDs.

Age-Related Changes in Gene Expression: The genes with age-related changes in gene expression were taken from the gene expression resources in HAGR (http://genomics.senescence.info/). The data originated from a meta-analysis of age-related gene expression profiles using 27 datasets from mice, rats, and humans [5]. The results of this meta-analysis were mapped to their Entrez IDs, and those with a q-value (FDR) < 0.1 were selected, resulting in 55 genes, all of which mapped to the PIN. To obtain a larger number of genes (as used in Fig. 3 A and B), we relaxed the threshold to q<0:3, yielding 206 genes, of which 140 were annotated to the PIN.

Topological Impact Synergy Analysis: The main idea behind this analysis is to assess if two distinct classes of genes interact frequently in the context of the PIN. One way to assess this is to construct the largest connected component induced by combining the two classes of genes. Thus, the induced subnetwork is constructed from all edges in the PIN linking genes in the union set of the two separate classes. To assess the statistical

significance of the size of the induced largest connected component, 1,000 degree-matched collections of nodes were sampled independently from the PIN, and the sizes of their largest connected components were computed. It was found that a normal distribution typically fit the observed distribution of largest connected component sizes well, and P values were estimated using this normal model. This technique differs slightly from that of ref. 15, in which degree-conserved networks were sampled as control. One advantage is that sampling from the PIN preserves the finer topological structure than the degree distribution, however a disadvantage is that nodes with unique degree must occur in every sample. Around 26% of all distinct degrees were unique, primarily those of higher degree. Thus, our significance estimates are conservative.

SUPPLEMENTARY FIGURES AND TABLES:

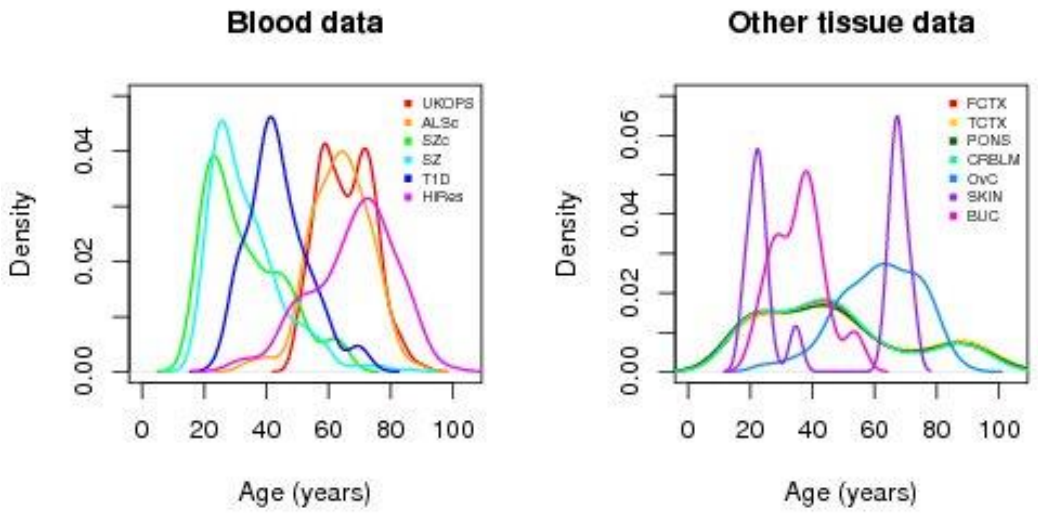

Fig.S1: Age-density distributions for all whole blood data sets used (left) and for the data sets measuring DNA methylation in all other tissue types (right panel). Datasets are described in detail in Table S1.

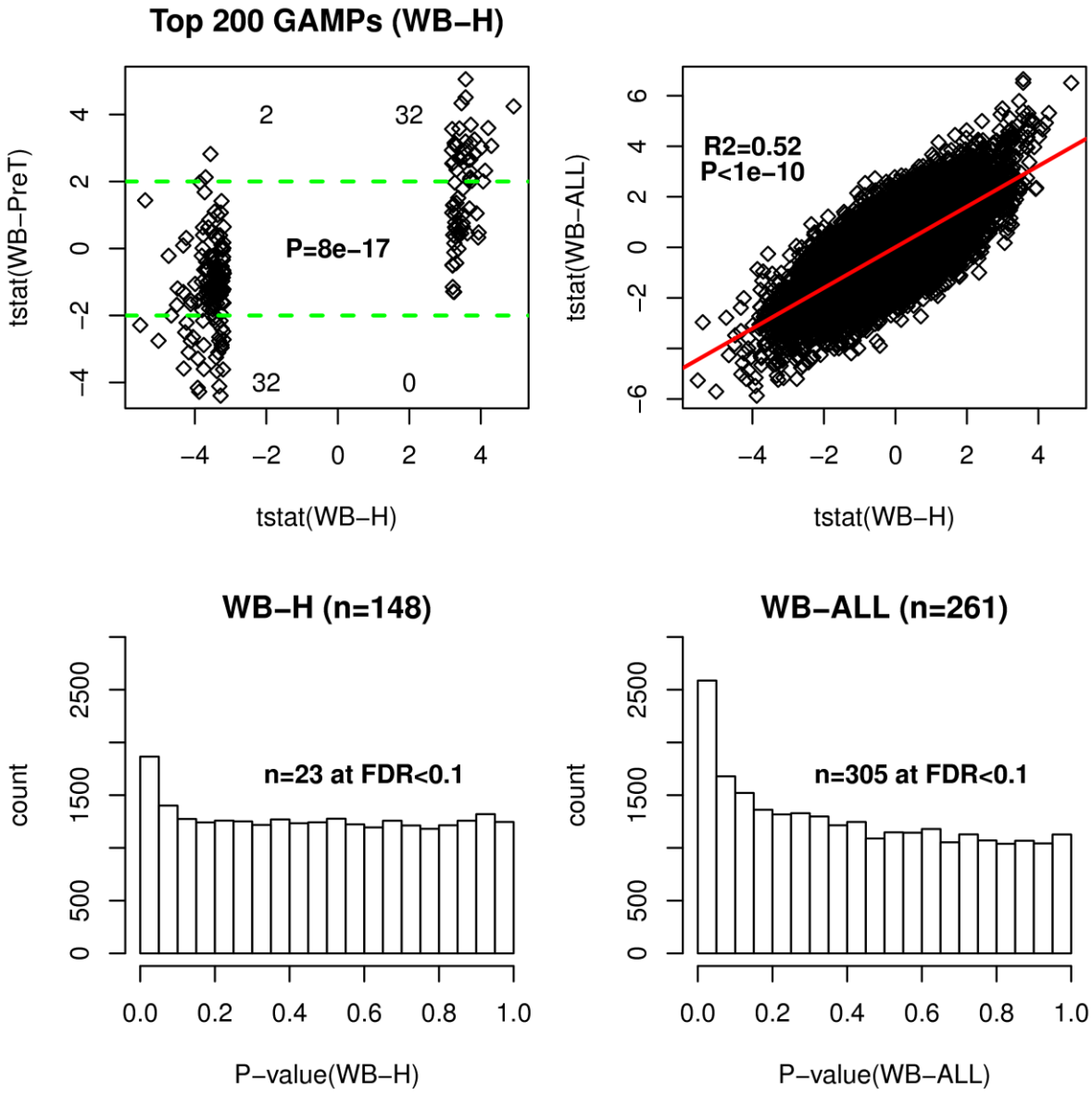

Fig.S2: Top left panel: Scatterplot of regression t-statistics of the top 200 GAMPs, obtained by correlating DNA methylation to age in the 148 whole blood samples from healthy postmenopausal women (x-axis, WB-H) against those obtained in the whole blood samples from pretreatment cases (n=113) (y-axis, WB-PreT). Agreement between statistics is shown using a Fisher's exact test, counting the number of GAMPs that pass statistical significance in the 113 pretreatment set. Top right panel: Scatterplot of regression t-statistics obtained by correlating DNA methylation to age in the 148 whole blood samples from healthy postmenopausal women (x-axis, WB-H) against those obtained from the combined 148+113 (261) whole blood set (y-axis, WB-ALL). R^2 and P-value from linear regression given. Lower panels depict the histogram of P-values derived from the case where only the 148 whole blood samples are used (left panel), and from the case where all 261 (148+113) whole blood samples are used (right). Note how the number of detected GAMPs at an FDR < 0.1 is improved by using all samples.

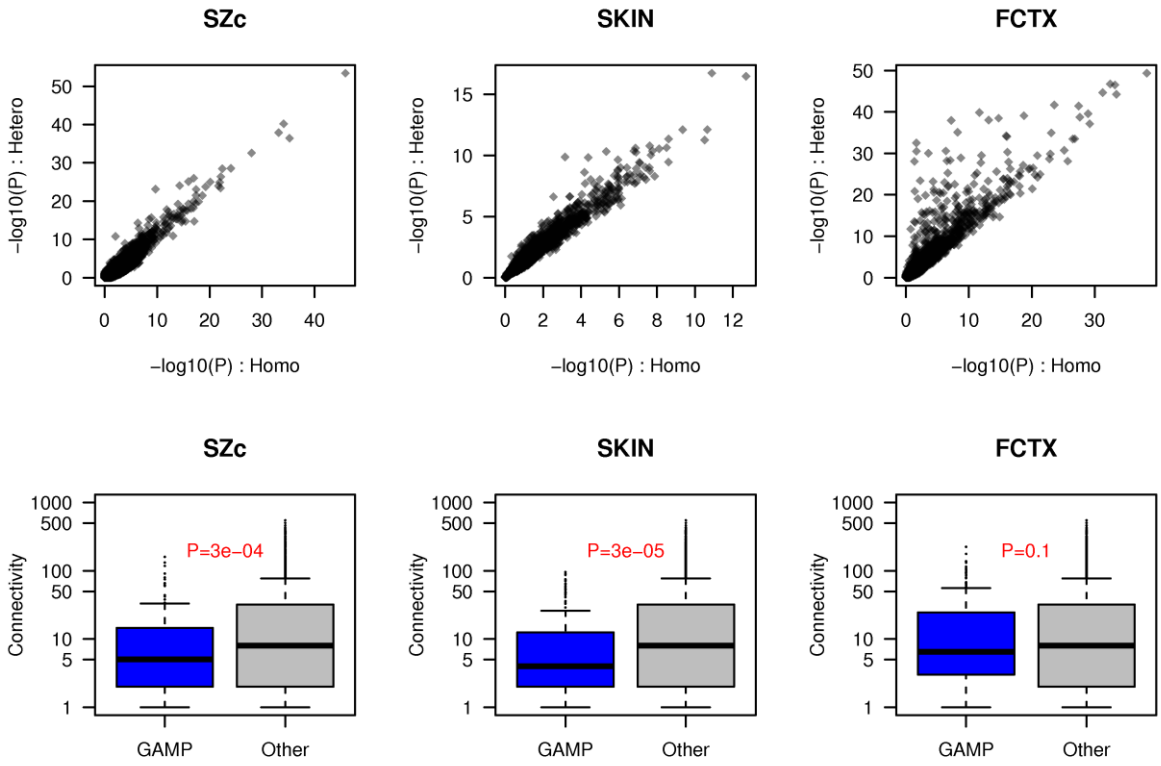

Fig.S3: Top panels represent scatterplots of $-\log_{10}$(P-values) from homoscedastic linear regressions (x-axis) against the corresponding $-\log_{10}$(P-values) obtained from heteroscedastic linear regressions (y-axis), as obtained from the *hett* R package *(www.r-project.org)* [6], for three different studies encompassing whole blood (SZc), skin (SKIN) and frontal cortex brain tissue (FCTX). Lower panels compare the connectivity of inferred GAMPs (from the heteroscedastic model) to non-GAMPs. Wilcoxon rank sum test P-values are given.

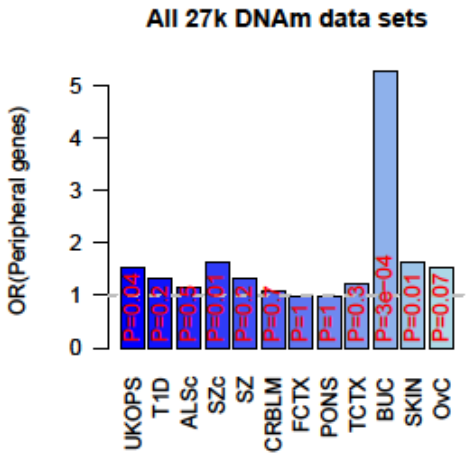

Fig.S4: Barplots of odds ratios (OR) for over-representation of peripheral genes (i.e. with centrality values between 1 and 2) among GAMPs. *P* values are from one-sided Fisher test.

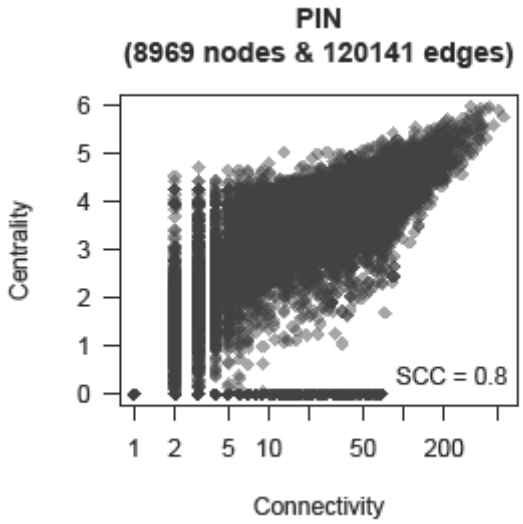

Fig.S5: Correlation between connectivity (i.e. degree) and centrality (i.e. $\log_{10}$(1+betweenness)) for the protein interaction network (PIN) used throughout. A Spearman correlation coefficient (SCC) of 0.8 was found.

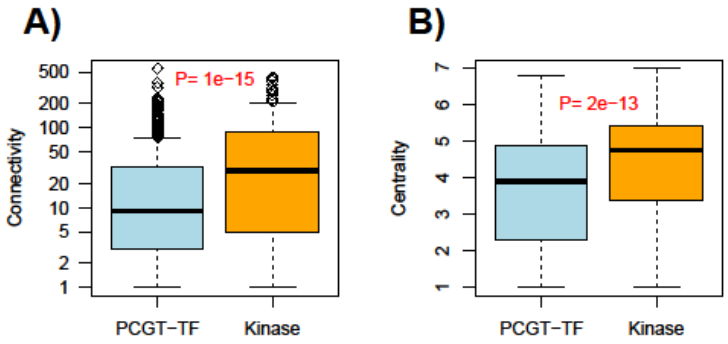

Fig.S6: Transcription factors have lower connectivity and betweenness centrality than kinases. **A)** Boxplots of connectivity distribution for Polycomb Group Targets (PCGTs) which are also transcription factors (PCGT-TF) and kinases. **B)** Boxplots of betweenness centrality distribution for PCGT-TF and kinases. Connectivity and centrality computed with respect to the PIN.

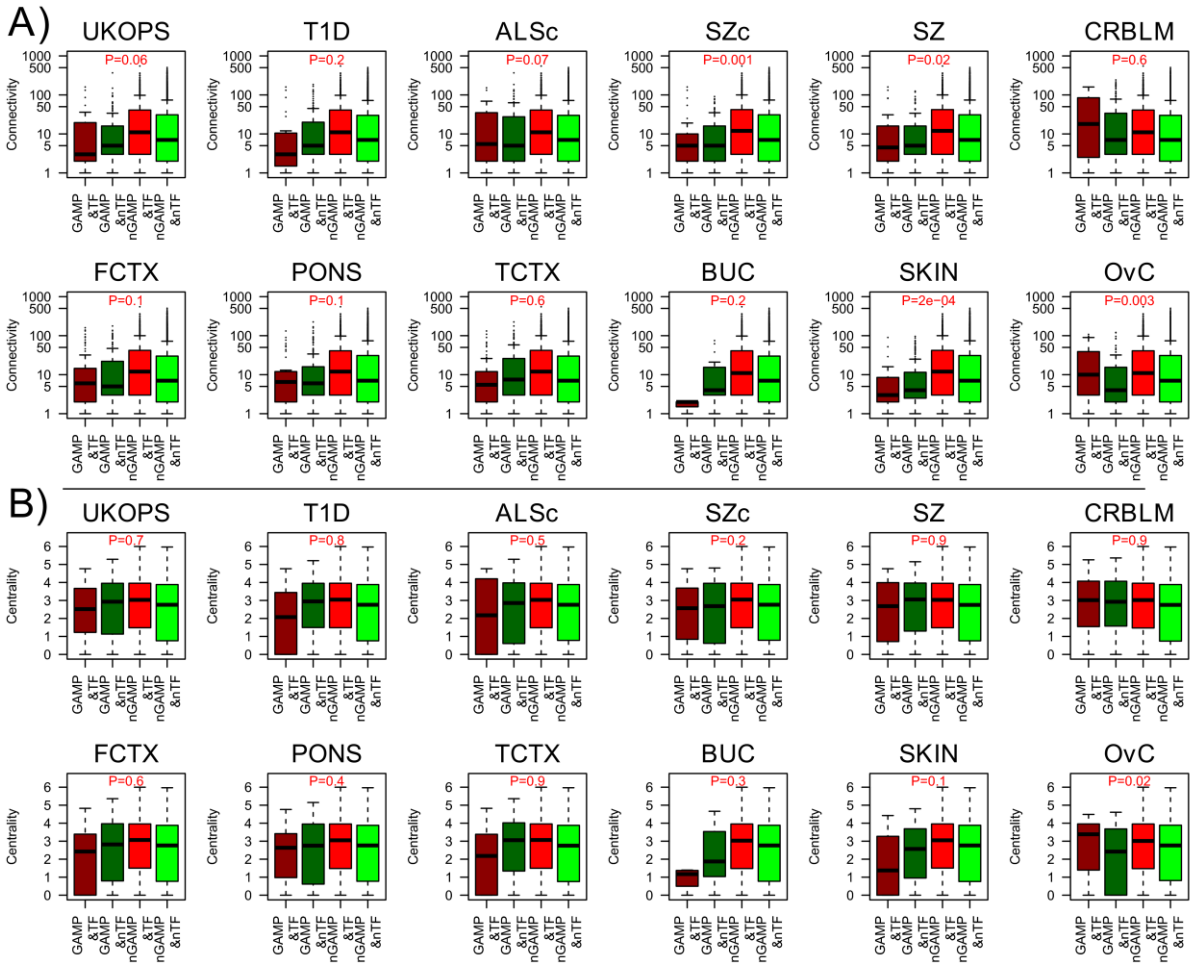

Fig.S7: Comparison of connectivity and centrality of genes in relation to GAMP and transcription factor status. **A)** Boxplots of connectivity distributions shown for all Infinium 27k datasets according to whether gene is a GAMP or not (nGAMP) and whether it is a transcription factor (TF) or not (nTF). *P* values shown are from a one-sided Wilcoxon rank sum test of non-transcription factor GAMPs (GAMP&nTF, dark green) against all non-GAMPs (nGAMP&TF combined with nGAMP&nTF, red and green). **B)** As A) but now for centrality (i.e. $\log_{10}$(1+betweenness)).

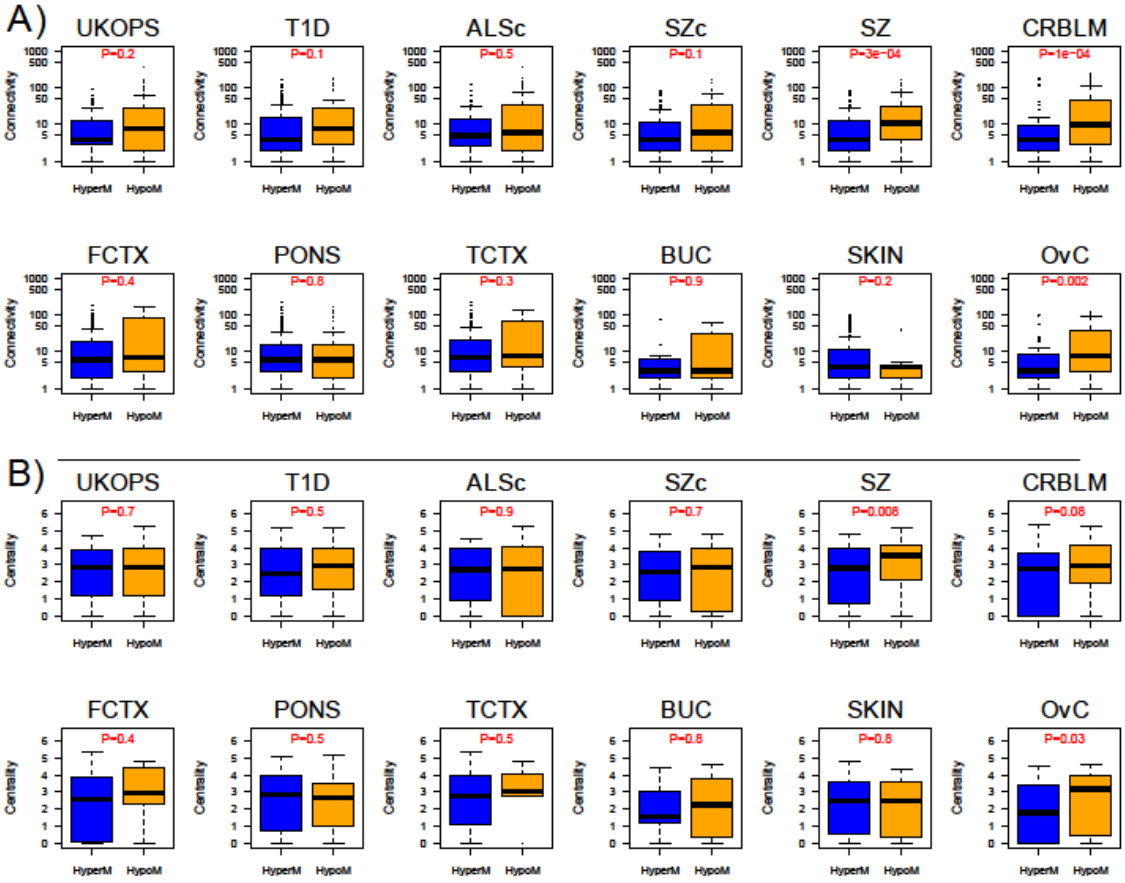

Fig.S8: Connectivity and centrality of GAMPs is largely independent of direction of age-associated methylation. **A)** Boxplots showing the connectivity distribution of GAMPs, according to whether GAMP is hypermethylated (HyperM) or hypomethylated (HyperM) with age. *P* values shown are from two-sided Wilcox test. **B)** As A) but for centrality.

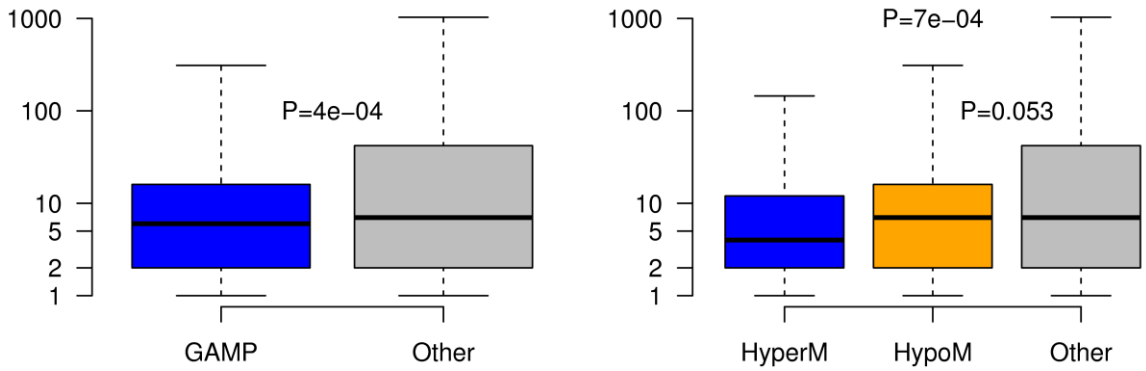

Fig.S9: Left panel: Connectivity of GAMPs vs that of non-GAMPs (Other) in the pediatric set (WB-PED). One sided Wilcoxon rank sum test P-value given. Right panel: As left panel, but separating out the GAMPs into those exhibiting age-associated hypermethylation (HyperM) vs. those exhibiting hypomethylation (HypoM). In this case we provide the one-sided Wilcoxon rank sum test P-values between HyperM and Other groups (middle P-value) and between HypoM and Other (P-value on far right).

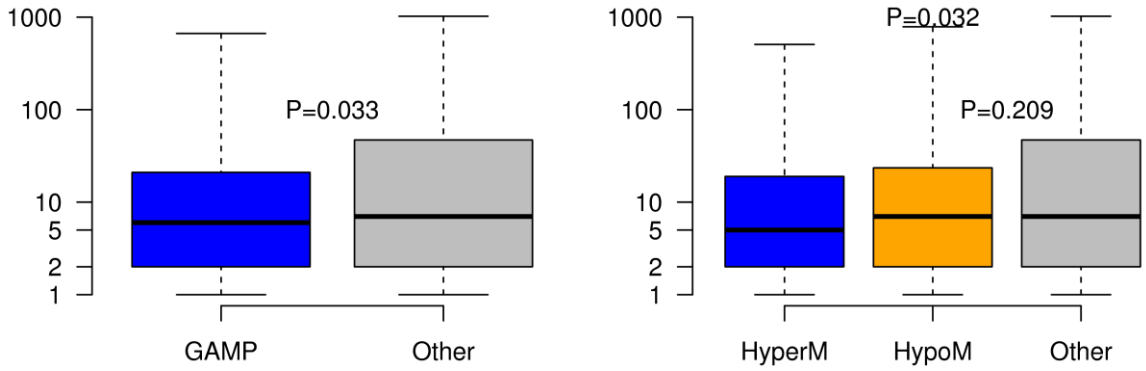

Fig.S10: Left panel: Connectivity of GAMPs vs that of non-GAMPs (Other) in the UKOPS set (ages>50 years), after removal of age-hypermethylated GAMPs from the WB-PED set (ages 3-17 years). One sided Wilcoxon rank sum test P-value given. Right panel: As left panel, but separating out the GAMPs into those exhibiting age-associated hypermethylation (HyperM) vs. those exhibiting hypomethylation (HypoM). In this case we provide the one-sided Wilcoxon rank sum test P-values between HyperM and Other groups (middle P-value) and between HypoM and Other (P-value on far right).

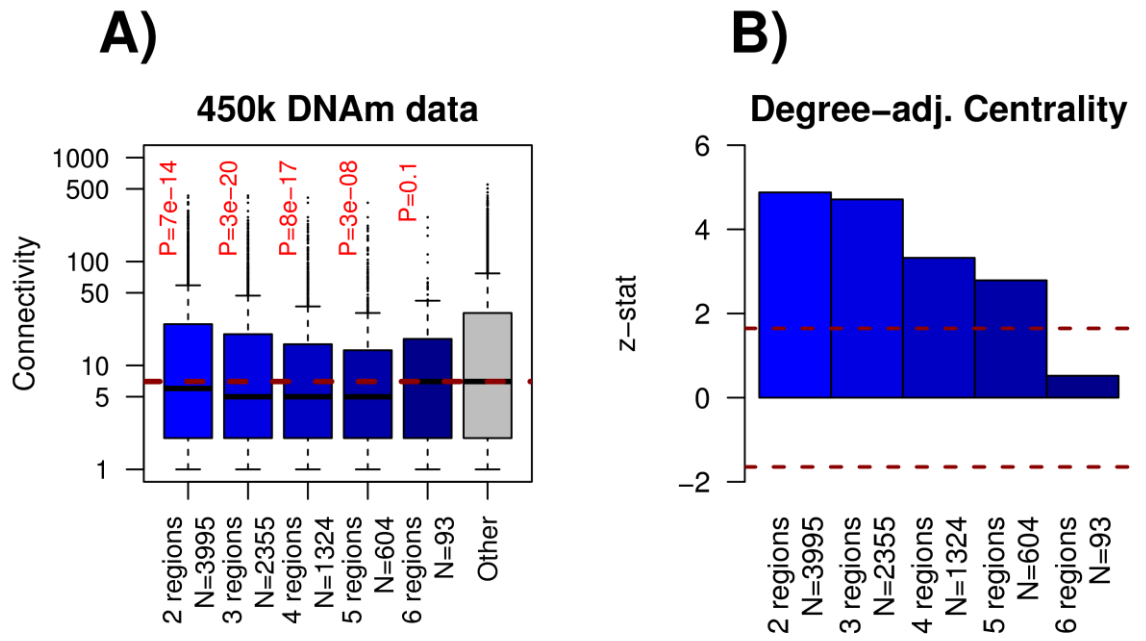

Fig.S11: Genes with age-associated methylation changes (GAMs) in multiple gene regions exhibit increasingly lower degree and degree adjusted centrality as the number of distinct affected regions increases (except for case of all 6 gene regions). **A)** Boxplots of connectivity distributions for genes found with between two and six distinct genomic regions with age-associated methylation at the 1% FDR level. *P* values are obtained from one-sided Wilcox tests. **B)** Barplots of *z*-statistics reflecting significance of the degree-adjusted centrality of GAMs, plotted according to the number of genomic regions exhibiting age-associated changes.

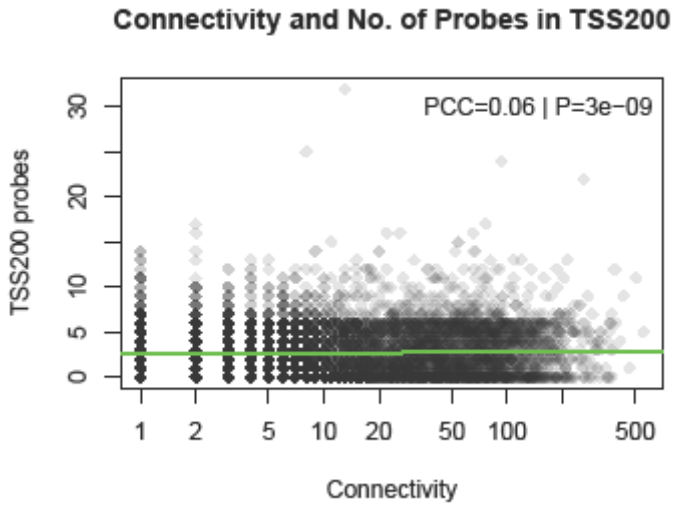

Fig.S12: Scatterplot of the number of probes in the TSS200 region (200 base pairs from the transcription start site) on the Infinium 450k platform plotted against the connectivity of each gene. Pearson correlation coefficient PCC is 0.06 and *P* value is from Pearson correlation test. Green line indicates linear regression line.

Dataset	Platform	Tissue	Number	Type	Ages	Identifier
ALSc	27k	Blood	92	92N	34 - 88	GSE41037 [7]
SZc	27k	Blood	273	273N	16 - 65	GSE41037 [7]
SZ	27k	Blood	293	293N*	17 - 86	GSE41037 [7]
UKOPS	27k	Blood	261	261N*	50 - 84	GSE19711 [8]
T1D	27k	Blood	187	187N*	24 - 74	GSE20067 [9]
FCTX	27k	F. Cortex	132	132N	16 - 101	GSE15745 [10]
TCTX	27k	T. Cortex	126	126N	15 - 101	GSE15745 [10]
PONS	27k	Pons	123	123N	15 - 101	GSE15745 [10]
CRBLM	27k	Cerebellum	111	11N	16 - 96	GSE15745 [10]
SKIN	27k	Skin	50	50N	18 - 72	E-MTAB-202 [11]
BUC	27k	Saliva	71	71N	21 - 55	GSE28746 [12]
OvC	27k	Ovary	177	177C	25 - 89	See: Methods [8]
WB-PED	27k	Blood	398	398N	3 - 17	GSE27097 [13]
HiRes	450k	Blood	255	255N	28 - 101	GSE40279 [14]

Table S1: Methylation data sets used. Columns list the abbreviation used for data set, the platform upon which the data was generated, tissue type, number of samples, cell-type or disease state, age-range and an identifier giving either a GEO code (GSE), an EMBL-EBI ArrayExpress code (E-MTAB) and a reference to a paper with further details. Phenotypes: N = from normal or healthy tissue, C = cancer. N*: normal tissue but taken from patients with a disease, causally unrelated to the tissue itself. In the case of UKOPS, 261 whole blood samples from 148 healthy women and 113 from women with ovarian cancer (before treatment). In the case of T1D, whole blood samples from 187 type1 diabetics. In the case of SZ, 293 whole blood samples from patients with schizophrenia.

Dataset	Hyper-GAMPs	nPCGT-TF	OR	P value
UKOPS	81	15	1.615604	0.093663
T1D	90	17	1.920535	0.021883
OvC	51	12	2.645174	0.007285
ALSc	58	10	1.999054	0.044623
SZc	1740	293	1.236172	0.010255
SZ	958	168	1.67136	9.36E-07
CRBLM	71	14	1.55523	0.138
FCTX	1810	333	1.695066	4.75E-10
PONS	3370	643	1.612639	6.17E-10
TCTX	2220	395	1.563526	3.35E-08
BUC	15	4	3.478564	0.047126
SKIN	839	152	1.744257	4.03E-07

Table S2 Transcription factor PCGTs are enriched amongst age-hypermethylated GAMPs. Columns give the dataset used, the total number of age-hypermethylated GAMPs (hyper-GAMPs) at 30% FDR, the overlap with transcription factors that are also PCGTs (nPCGT-TF), the odds ratio (OR) and associated *P* value from Fisher test.

378884, 54206, 9612, 128178, 443, 2006, 1453, 5143, 1010, 9247, 80270, 5260, 6656, 7425,
401563, 3787, 84467, 4725, 3097, 1454, 79745, 8482, 4055, 9557, 8326, 11166, 240, 26585,
151647, 6750, 5829, 51761, 27201, 3902, 26027, 6347, 9324, 3356, 2862, 84818, 10501, 84267,
3326, 80004, 5593, 85415, 166785, 2731, 84514, 5327, 11255, 81831, 64098, 4257, 2741, 59269,
55643, 22937, 6123, 80325, 939, 65268, 2974, 126792, 85453, 9770, 144165, 353500, 80714,
10365, 2891, 4920, 60312, 29984, 23348, 2039, 57447, 11269, 10814, 9308, 9586, 656, 4741,
6919, 114770, 10735, 282991, 221914, 1102, 51161, 84366, 10993, 401, 83874, 11000, 153, 10911,
5458, 6456, 11245, 3587, 116173, 5688, 162417, 2977, 26301, 10174, 27233, 65981, 2633, 26145,
6364, 26526, 7179, 257397, 9956, 10678, 84958, 56135, 1212, 2801, 85016, 284119, 51663, 55502,
10800, 81563, 10158, 10417, 284739, 64405, 197021, 2483, 3084, 284948, 9620, 919, 163255,
89874, 57369, 2992, 1203, 1907, 111, 247, 957, 8303, 91227, 92086, 283212, 6375, 7036, 388611,
1991, 2811, 10407, 915, 170302, 9887, 684, 1674, 170692, 6787, 53831, 221178, 670, 222008,
80235, 201780, 54207, 4897, 430, 124641, 53630, 55133, 1813, 1809, 157807, 9942, 252995, 6869,
284837, 6422, 4063, 83937, 51343, 8111, 6539, 120935, 64170, 64127, 170482, 6181, 7053, 7461,
92714, 8618, 1138, 381, 54209, 10381, 3748, 84765, 64321, 27022, 2848, 81615, 10040, 80380,
219770, 9404, 9982, 79035, 84969, 6749, 3741, 10000, 1735, 7730, 2535, 55888, 2743, 7462,
7545, 341883, 339488, 222546, 53820, 9421, 10974, 29947, 3746, 140691, 3108, 197257, 5649,
89858, 122970, 284451, 635, 3199, 93100, 171017, 908, 163589, 5995, 146225, 57057, 9351,
126393, 860, 5032, 353324, 50486, 3798, 51296, 1501, 54210, 9659, 7379, 81875, 83850, 1368,
130574, 3898, 377841, 158062, 27120, 22981, 846, 388677, 388021, 250, 7851, 1630, 2395, 10633,
148113, 1440, 22899, 51412, 6708, 2917, 29015, 4291, 2532, 26002, 162517, 64403, 5433, 6000,
57573, 781, 8660, 54659, 7472, 23305, 84645, 2922, 284459, 84179, 4978, 92305, 1908, 148423,
148523, 9050, 55800, 4885, 10842, 84518, 140885, 220, 4351, 23136, 152, 55258, 51673, 362,
284186, 400696, 2525, 53904, 3237, 148641, 56121, 8942, 8987, 6664, 5657, 81832, 79645, 22903,
118425, 143425, 2324, 91937, 3430, 28964, 285268, 255762, 55898, 10150, 66002, 115290, 55902,
7078, 5902, 55612, 285759, 92304, 389799, 6355, 2246, 3956, 881, 8859, 5753, 161753, 114904,
167465, 201294, 6939, 7804, 51435, 6332, 23604, 5655, 9159, 8882, 25803, 27124, 79816, 1562,
1999, 10288, 54494, 6715, 339403, 55646, 10103, 132884, 83482, 7416, 388324, 10659, 8435,
2620, 378708, 23576, 83546, 10382, 126208, 2038, 5352, 3931, 54677, 7027, 9785, 7710, 118672,
6704, 63898, 245711, 114960, 124220, 114, 5409, 4320, 57863, 23484, 3593, 8970, 4352, 1717,
403, 6538, 1902, 11329, 5020, 9355, 3687, 9290, 8331, 26040, 23321, 1511, 117156, 79370, 7044,
3362, 10277, 56169, 3562, 84220, 6348, 26212, 11093, 1400, 11331, 112755, 26133, 5308, 1318,
3983, 113730, 3024, 9816, 2001, 374655, 875, 9314, 80320, 143379, 275, 51378, 8630, 153572,
115703, 64577, 220388, 220004, 10346, 140733, 8641, 222643, 2668, 5179, 51085, 56098, 84504,
11061, 1825, 9037, 83690, 81035, 10687, 6564, 9480, 256364, 8854, 2056, 378108, 388759, 9935,
5923, 54345, 2919, 706, 9362, 8433, 256051, 93145, 60482, 3226, 5333, 147381, 5015, 80157,
57453, 30815, 2128, 6662, 79852, 7373, 54886, 1000, 4353, 9607, 8091, 348807, 168391, 63928,
57144, 284217, 89780, 55742, 9630, 7060, 51162, 91133, 1289, 115827, 3952, 54768, 5455, 6753,
377677, 1745, 4487, 116519, 116447, 9496, 204962, 78997, 9576, 7546, 56884, 6716, 133060,
1749, 3170, 80274, 57468, 1301, 6563, 64063, 124216, 9365, 84539, 3229, 1833, 349136, 27023,
2746, 8930, 22941, 83638, 9615, 56132, 7066, 4016, 3270, 160777, 51117, 55640, 6374, 8001,
1641, 343472, 57717, 56914, 2125, 283078, 30012, 8644, 2849, 57167, 285676, 1298, 23231,
84457, 202559, 54102, 27300, 8294, 4544, 249, 206338, 10361, 51427, 26167, 225689, 23779,
9555, 200132, 221409, 197, 6910, 21, 79442, 5239, 284348, 10398, 152789, 1816, 8490, 7771,
22796, 55786, 80830, 10655, 9464, 1278, 284111, 286128, 975, 2830, 326624, 2290, 7113, 1600,
89792, 112476, 4685, 3975, 25893, 30009, 6624, 5587, 22943, 4018, 6886, 201232, 4852, 139341,
6447, 5459, 6543, 144195, 25966, 7345, 56099, 4829, 671, 5141, 401498, 2834, 284406, 6344,
26476, 5746, 286046, 259232, 26047, 284656, 348980, 9933, 79891, 51006, 9706, 8814, 3394, 762,
8989, 11055, 57467, 56341, 338917, 4761, 1396, 7441, 84109, 128653, 30819, 51112, 5605, 23551,
3049, 7015, 245806, 285175, 1395, 2533, 55130, 79948, 93190, 8363, 6452, 25830, 92400, 2115,
121256, 4744, 85455, 130497, 9118, 1039, 222894, 1036, 1135, 51352, 5088, 5644, 3101, 3201,
26493, 6934, 27164, 51068, 9833, 399823, 116159, 3772, 163747, 26025, 10969, 340061, 9874,
6822, 4967, 11027, 138639, 10394, 51805, 23504, 116461, 136647, 9705, 23676, 10647, 256297,
345274, 5729, 23199, 51305, 285242, 11024, 56253, 54491, 11095, 4763, 3069, 148022, 10241,
317761, 10570, 859, 9350, 9371, 2898, 2191, 1109, 90390, 140689, 26118, 10457, 27136, 83931,
6928, 259295, 6423, 6898, 51142, 4204, 10566, 8799, 7314, 23461, 55303, 85021, 58473, 53822,
57415, 1325, 1815, 147183, 640, 4288, 4060, 1472, 11075, 2581, 4299, 84699, 58476, 3205, 6100,
64499, 2766, 8756, 6542, 8309, 84222, 563, 6677, 29926, 9491, 79843, 56124, 5136, 6812,
342574, 758, 79897, 11026, 10268, 1464, 2786, 28969, 51025, 5630, 63970, 1594, 84294, 145781,
23299, 64091, 11172, 8388, 9358, 2813, 5971, 148898, 84163, 6747, 4161, 2530, 57404, 51702,
8492, 427, 5858, 3238, 23524, 26292, 1435, 26529, 6014, 400661, 435, 84528, 10391, 6489, 1089,
55615, 3219, 11186, 1634, 5243, 123803, 11025, 27350, 51214, 2184

Table S3 Entrez IDs of all 855 GAMPs pooled across the non-brain tissue datasets.

Supplementary References:

1. Bibikova, M. et al. Genome-wide DNA methylation profiling using Infinium(R) assay. *Epigenomics* **1**, 177-200 (2009).
2. Dedeurwaerder, S. et al. Evaluation of the Infinium Methylation 450K technology. *Epigenomics* **3**, 771-784 (2011).
3. Ravasi, T. et al. An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* **140**, 744-752 (2010).
4. Laurie, C.C. et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nature genetics* **44**, 642-650 (2012).
5. de Magalhaes, J.P., Curado, J. & Church, G.M. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics* **25**, 875-881 (2009).
6. Taylor, J.D. & Verbyla, A.P. Joint modelling of the location and scale parameters of the

t-distribution. *Statistical Modelling* **4**, 91-112 (2004).
7. Horvath, S. et al. Aging effects on DNA methylation modules in human brain and blood tissue. *Genome biology* **13**, R97 (2012).
8. Teschendorff, A.E. et al. An epigenetic signature in peripheral blood predicts active ovarian cancer. *PloS one* **4**, e8274 (2009).
9. Bell, C.G. et al. Genome-wide DNA methylation analysis for diabetic nephropathy in type 1 diabetes mellitus. *BMC medical genomics* **3**, 33 (2010).
10. Gibbs, J.R. et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS genetics* **6**, e1000952 (2010).
11. Gronniger, E. et al. Aging and chronic sun exposure cause distinct epigenetic changes in human skin. *PLoS genetics* **6**, e1000971 (2010).
12. Bocklandt, S. et al. Epigenetic predictor of age. *PloS one* **6**, e14821 (2011).
13. Alisch, R.S. et al. Age-associated DNA methylation in pediatric populations. *Genome research* **22**, 623-632 (2012).
14. Hannum, G. et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular cell* **49**, 359-367 (2013).