Supplementary Data

Supplementary Table 1. Gene list and sizes of genes tiled on the custom microarray

Supplementary Table 2. Genomic locations of 181 MVPs within the 17 candidate breast cancer susceptibility genes and 38 peripherally targeted genes.

Supplementary Table 3. Primers used in this study.

Supplementary Table 4. Cell lines used in this study.

Supplementary Figure 1. Methylation microarray analysis of seventeen genes reveals methylation variability in repetitive elements. A) Schematic of each of the genes tiled on the custom microarray. **B)** Example of an inter-individual methylation variable position (MVP) (black bar above) within the gene-body of the ATM gene showing an association with the repetitive element (grey bars below).

Supplementary Figure 2. Methylation microarray data. A) Methylation microarray data as presented on the UCSC genome web browser via custom .WIG and .BED files online. B) Schematic of Methylation microarray data as presented on the UCSC genome web browser. Each MVP is represented by a vertical pink bar, genes are coloured dark blue (no intragenic MVPs), light blue (very few intragenic MVPs) or pink (containing numerous intragenic MVPs).

Supplementary Figure 3. Supporting statistical analysis of ATM mvp2b in peripheral blood DNA. Methylation values for cases and controls were randomly separated into two groups and the Wilcoxon signed rank test was used to calculate the

test statistic, W. Random sampling was performed 1000 times (median simulated W = 17935) compared to the observed W=14605. The actual P-value was calculated as p=0.002.

Supplementary Figure 4. Correlation between age and methylation for the ATM mvp2b. A). A) Distribution of ages in cases and controls. Methylation was plotted against age at blood draw for controls (B) and cases (C) or age at first (D) or second (E) diagnosis for cases.

Supplementary Figure 5. Methylation analysis of ATM mvp2b in blood cell fractions reveals no cell-type specific methylation. B-cells, T-cells and monocytes were isolated from the peripheral blood mononuclear cells (PBMCs) from two unaffected controls and pyrosequencing based methylation analysis of ATM mvp2b was performed. Methylation percentages represent an average of four CpG dinucleotides. As a positive control we have assayed a region, chr13:76462565-76463564, that was previously described as differentially methylated between B cells and T cells, shown as the theoretical value (+) (21).

Supplementary Table 1

Target Gene	Chr	UCSC start	UCSC end	strand	TSS
BRCA1	chr17	38449843	38530934	-1	38530934
BRCA2	chr13	31787616	31871805	1	31787616
CHEK2	chr22	27408284	27462376	-1	27462376
ER-a	chr6	152220799	152516520	1	152220799
SFN	chr1	26873774	26875089	1	26873774
CDKN2A	chr9	21992902	21999312	-1	21999312
HIC1	chr17	1905012	1909070	1	1905012
ATM	chr11	107599052	107741966	1	107599052
GSTP1	chr11	67107861	67110699	1	67107861
CDH13	chr16	81218078	82387698	1	81218078
PGR	chr11	100414312	100506465	-1	100506465
SFRP1	chr8	41238635	41286137	-1	41286137
HSD17B4	chr5	118816123	118905923	1	118816123
P53	chr17	7512463	7531642	-1	7531642
MLH1	chr3	37009982	37067341	1	37009982
CDH1	chr16	67328755	67424940	1	67328755
RAR-b	chr3	25517675	25614424	1	25517675
ZDHHC18	chr1	26837343	26866350	1	26837343
ATPBD1B	chr1	26890015	26901011	-1	26901011
GPATCH3	chr1	26901122	26911099	-1	26911099
NR0B2	chr1	27110566	27113047	-1	27113047
RPL27	chr17	38403972	38408497	1	38403972
IFI35	chr17	38412351	38419999	1	38412351
VAT1	chr17	38420148	38427985	-1	38427985
RND2	chr17	38430891	38434223	1	38430891
N4BP2L1	chr13	31872861	31900315	-1	31900315
CDKN2B	chr9	21992903	21999312	-1	21999312
HSBP1	chr16	82399094	82404095	1	82399094
HSCB	chr22	27462597	27478048	1	27462597
CCDC117	chr22	27493261	27509832	1	27493261
C11orf72	chr11	67126927	67130753	-1	67130753
NDUFV1	chr11	67130983	67136581	1	67130983
NUDT8	chr11	67151986	67153977	-1	67153977
DPH1	chr17	1880181	1893474	1	1880181
OVCA2	chr17	1892092	1893474	1	1892092
EPM2AIP1	chr3	37004233	37009799	-1	37009799
SAT2	chr17	7470281	7471889	-1	7471889
SHBG	chr17	7474216	7477395	1	7474216
ATP1B2	chr17	7494979	7501814	1	7494979
WDR79	chr17	7532520	7547544	1	7532520
EFNB3	chr17	7549245	7555416	1	7549245
NBR2	chr17	38531153	38550649	1	38531153
TMEM133	chr11	100368021	100369875	1	100368021

NPAT_5_partial	chr11	107533330	107598507	-1	107598507
C110rf65_3_partial	chr11	107758938	107843468	-1	107843468
NBR1_5_partial	chr17	38576772	38719232	1	38576772
FRY_3_partial	chr13	31503437	31768775	1	31503437
CDH3_3_partial	chr16	67236277	67290442	1	67236277
SYNE1_3_partial	chr6	152535029	153050648	-1	153050648
TBX10_3_partial	chr11	67155351	67163607	-1	67163607
RTN4RL1_5_partial	chr17	1784723	1874928	-1	1874928
SMG6_3_partial	chr17	1909883	2153819	-1	2153819
LRRFIP2_3_partial	chr3	37069503	37191086	-1	37191086
DNAH2_5_partial	chr17	7563764	7677783	1	7563764
TOP2B_3_partial	chr3	25614479	25680792	-1	25680792

Supplementary Table2

	chr	start	end	target gene	MATscore	length	actual gene
1	chr1	26909736	26910469	SFN	0.6391	733	GPATCH3
2	chr11	107594559	107595271	ATM	0.68907	712	NPAT_5_partial
3	chr11	107601991	107602934	ATM	0.69405	943	ATM
4	chr11	107612293	107613467	ATM	0.6582	1174	ATM
5	chr11	107654522	107655192	ATM	0.53761	670	ATM
6	chr11	107745337	107745986	ATM	0.54693	649	intergenic
7	chr11	107758230	107758921	ATM	0.52307	691	intergenic
8	chr11	107758922	107759593	ATM	0.52944	671	intergenic
9	chr11	100475121	100475833	PGR	0.55163	712	PGR
10	chr11	100555358	100556194	PGR	0.65253	836	intergenic
11	chr13	31771377	31772068	BRCA2	0.57653	691	intergenic
12	chr16	67335382	67336343	CDH1	0.72892	961	CDH1
13	chr16	67359350	67360209	CDH1	0.61284	859	CDH1
14	chr16	67360210	67360552	CDH1	0.78463	342	CDH1
15	chr16	67394044	67394735	CDH1	0.57967	691	CDH1
16	chr16	67394800	67395449	CDH1	0.67115	649	CDH1
17	chr16	67397554	67398245	CDH1	0.75929	691	CDH1
18	chr16	67413295	67414028	CDH1	0.64726	733	CDH1
19	chr16	67418405	67419348	CDH1	0.80531	943	CDH1
20	chr16	67438492	67439414	CDH1	0.71949	922	intergenic
21	chr16	67332491	67333371	CDH1	-0.95603	880	CDH1
22	chr16	67380436	67381232	CDH1	-0.72487	796	CDH1
23	chr16	67383627	67384823	CDH1	-0.81045	1196	CDH1
24	chr16	81171015	81172441	CDH13	0.9447	1426	intergenic
25	chr16	81190036	81190832	CDH13	0.61447	796	intergenic
26	chr16	81213746	81214689	CDH13	0.63987	943	intergenic
27	chr16	81230999	81231753	CDH13	0.56795	754	CDH13
28	chr16	81237300	81238425	CDH13	1.14279	1125	CDH13
29	chr16	81238426	81238737	CDH13	1.15402	311	CDH13
30	chr16	81240150	81241114	CDH13	0.90104	964	CDH13
31	chr16	81245778	81246511	CDH13	0.66185	733	CDH13
32	chr16	81246512	81247110	CDH13	1.09515	598	CDH13
33	chr16	81291361	81292346	CDH13	0.81174	985	CDH13
34	chr16	81292347	81292823	CDH13	0.52752	476	CDH13
35	chr16	81307400	81308637	CDH13	0.6442	1237	CDH13
36	chr16	81308649	81309403	CDH13	0.59198	754	CDH13
37	chr16	81309404	81309790	CDH13	0.57524	386	CDH13
38	chr16	81327441	81328615	CDH13	0.62109	1174	CDH13
39	chr16	81346881	81347761	CDH13	0.77859	880	CDH13
40	chr16	81347762	81348977	CDH13	0.85999	1215	CDH13
41	chr16	81370285	81371375	CDH13	0.68203	1090	CDH13
42	chr16	81379500	81380506	CDH13	0.60971	1006	CDH13
43	chr16	81402789	81403648	CDH13	0.58505	859	CDH13

44	chr16	81404511	81405503	CDH13	0.89972	992	CDH13
45	chr16	81421622	81422355	CDH13	0.51949	733	CDH13
46	chr16	81511390	81512102	CDH13	0.52129	712	CDH13
47	chr16	81521899	81522926	CDH13	1.46517	1027	CDH13
48	chr16	81574218	81574930	CDH13	0.56106	712	CDH13
49	chr16	81592756	81593405	CDH13	0.55266	649	CDH13
50	chr16	81613676	81614535	CDH13	0.6187	859	CDH13
51	chr16	81618814	81619463	CDH13	0.51961	649	CDH13
52	chr16	81675234	81675946	CDH13	0.64096	712	CDH13
53	chr16	81706279	81706928	CDH13	0.68946	649	CDH13
54	chr16	81712337	81713028	CDH13	0.84573	691	CDH13
55	chr16	81713029	81713321	CDH13	0.52775	292	CDH13
56	chr16	81717512	81718245	CDH13	0.66073	733	CDH13
57	chr16	81725614	81726537	CDH13	0.535	923	CDH13
58	chr16	81741737	81742659	CDH13	0.53224	922	CDH13
59	chr16	81746163	81747232	CDH13	1.13418	1069	CDH13
60	chr16	81749908	81750790	CDH13	0.60211	882	CDH13
61	chr16	81758328	81759172	CDH13	0.60512	844	CDH13
62	chr16	81767510	81768823	CDH13	0.59478	1313	CDH13
63	chr16	81830813	81831651	CDH13	0.54377	838	CDH13
64	chr16	81839300	81840327	CDH13	0.74069	1027	CDH13
65	chr16	81852296	81852966	CDH13	0.57449	670	CDH13
66	chr16	81857882	81858531	CDH13	0.50799	649	CDH13
67	chr16	81860439	81861193	CDH13	0.56374	754	CDH13
68	chr16	81865967	81867036	CDH13	0.99826	1069	CDH13
69	chr16	81874503	81875173	CDH13	0.52222	670	CDH13
70	chr16	81882910	81883580	CDH13	0.52818	670	CDH13
71	chr16	81903316	81904217	CDH13	0.86203	901	CDH13
72	chr16	81968022	81968881	CDH13	0.57793	859	CDH13
73	chr16	81973090	81973739	CDH13	0.62838	649	CDH13
74	chr16	82000215	82001305	CDH13	0.69711	1090	CDH13
75	chr16	82008170	82008882	CDH13	0.53256	712	CDH13
76	chr16	82028923	82030412	CDH13	0.65726	1489	CDH13
77	chr16	82063669	82064654	CDH13	0.93155	985	CDH13
78	chr16	82100406	82101055	CDH13	0.52739	649	CDH13
79	chr16	82107738	82108681	CDH13	0.5356	943	CDH13
80	chr16	82121918	82122735	CDH13	0.53811	817	CDH13
81	chr16	82141584	82142443	CDH13	0.89687	859	CDH13
82	chr16	82145779	82146596	CDH13	0.56461	817	CDH13
83	chr16	82147406	82148349	CDH13	0.90306	943	CDH13
84	chr16	82151487	82152220	CDH13	0.68513	733	CDH13
85	chr16	82157700	82158412	CDH13	0.74607	712	CDH13
86	chr16	82191369	82192123	CDH13	0.54661	754	CDH13
87	chr16	82196563	82197758	CDH13	0.66808	1195	CDH13
88	chr16	82212593	82213494	CDH13	1.00769	901	CDH13
89	chr16	82213495	82213869	CDH13	0.78454	374	CDH13

90	chr16	82234401	82235512	CDH13	0.57935	1111	CDH13
91	chr16	82236703	82237436	CDH13	0.60115	733	CDH13
92	chr16	82247811	82248460	CDH13	0.51356	649	CDH13
93	chr16	82248609	82250875	CDH13	1.54592	2266	CDH13
94	chr16	82258932	82259602	CDH13	0.55081	670	CDH13
95	chr16	82270576	82271393	CDH13	0.63772	817	CDH13
96	chr16	82280446	82281221	CDH13	0.60785	775	CDH13
97	chr16	82284610	82285385	CDH13	0.60021	775	CDH13
98	chr16	82300400	82301406	CDH13	0.91352	1006	CDH13
99	chr16	82371200	82371912	CDH13	0.57725	712	CDH13
100	chr16	81272853	81274366	CDH13	-1.1291	1513	CDH13
101	chr16	81649425	81650389	CDH13	-0.945	964	CDH13
102	chr16	81651615	81652478	CDH13	-0.74152	863	CDH13
103	chr16	81677277	81678304	CDH13	-1.07955	1027	CDH13
104	chr16	81686967	81687826	CDH13	-1.15064	859	CDH13
105	chr16	81691316	81692397	CDH13	-0.88957	1081	CDH13
106	chr16	81700258	81701117	CDH13	-0.68336	859	CDH13
107	chr16	81819741	81820390	CDH13	-0.67631	649	CDH13
108	chr16	82037431	82038447	CDH13	-0.92777	1016	CDH13
109	chr16	82317660	82318393	CDH13	-1.30188	733	CDH13
110	chr17	38404559	38405698	BRCA1	1.10476	1139	RPL27
111	chr17	38550869	38551581	BRCA1	0.61102	712	intergenic
112	chr17	38424335	38424984	BRCA1	-0.83484	649	VAT1
113	chr17	1863316	1864175	HIC1	0.81268	859	RTN4RL1_5_partial
114	chr17	1913975	1914813	HIC1	0.63173	838	SMG6_3_partial
115	chr17	1932874	1933880	HIC1	1.00477	1006	SMG6_3_partial
116	chr17	1899080	1899771	HIC1	-0.7072	691	intergenic
117	chr17	1919368	1920059	HIC1	-0.87779	691	SMG6_3_partial
118	chr17	7486252	7487279	P53	0.62512	1027	intergenic
119	chr17	7490480	7491150	P53	0.54259	670	intergenic
120	chr17	7493735	7494384	P53	0.6605	649	intergenic
121	chr17	7564243	7565333	P53	0.80339	1090	DNAH2_5_partial
122	chr17	7577520	7578610	P53	0.85223	1090	DNAH2_5_partial
123	chr22	27403259	27403950	CHEK2	0.51799	691	intergenic
124	chr22	27406426	27407369	CHEK2	0.56974	943	intergenic
125	chr22	27413353	27414002	CHEK2	0.51089	649	CHEK2
126	chr22	27414117	27415291	CHEK2	1.43113	1174	CHEK2
127	chr22	27415292	27416196	CHEK2	0.7189	904	CHEK2
128	chr22	27416197	27417022	CHEK2	0.93471	825	CHEK2
129	chr22	27480026	27481475	CHEK2	1.03664	1449	intergenic
130	chr22	27500495	27501228	CHEK2	0.5895	733	CCDC117
131	chr22	27407354	27408234	CHEK2	-0.8465	880	intergenic
132	chr3	37004752	37005695	MLH1	0.66433	943	EPM2AIP1
133	chr3	37005696	37006436	MLH1	1.07574	740	EPM2AIP1
134	chr3	37010352	37011001	MLH1	0.50958	649	MLH1
135	chr3	37038497	37039209	MLH1	0.61145	712	MLH1

136	chr3	37056240	37057146	MLH1	0.55585	906	MLH1
137	chr3	25504923	25507378	RAR_b	1.36698	2455	intergenic
138	chr3	25526101	25526834	RAR_b	0.54721	733	RAR-b
139	chr3	25526835	25527116	RAR_b	0.56921	281	RAR-b
140	chr3	25542754	25543424	RAR_b	0.54948	670	RAR-b
141	chr3	25552868	25553538	RAR_b	0.53839	670	RAR-b
142	chr3	25557038	25557792	RAR_b	0.532	754	RAR-b
143	chr3	25567870	25568897	RAR_b	0.59224	1027	RAR-b
144	chr3	25598174	25598823	RAR_b	0.54571	649	RAR-b
145	chr5	118773262	118774289	HSD17B4	1.04224	1027	intergenic
146	chr5	118796859	118797592	HSD17B4	0.75628	733	intergenic
147	chr5	118799403	118800094	HSD17B4	0.64587	691	intergenic
148	chr5	118813662	118814374	HSD17B4	1.30282	712	intergenic
149	chr5	118864138	118864787	HSD17B4	0.54867	649	HSD17B4
150	chr5	118926380	118928300	HSD17B4	0.95436	1920	intergenic
151	chr5	118934345	118935057	HSD17B4	0.58691	712	intergenic
152	chr6	152177717	152178492	ER_a	0.64377	775	intergenic
153	chr6	152198563	152199233	ER_a	0.54072	670	intergenic
154	chr6	152216306	152217270	ER_a	0.83841	964	intergenic
155	chr6	152222440	152223775	ER_a	0.571	1335	ER-a
156	chr6	152228283	152229183	ER_a	0.57668	900	ER-a
157	chr6	152259586	152260823	ER_a	1.03093	1237	ER-a
158	chr6	152371123	152371793	ER_a	0.5933	670	ER-a
159	chr6	152425284	152425933	ER_a	0.51375	649	ER-a
160	chr6	152470867	152472153	ER_a	0.87204	1286	ER-a
161	chr6	152508227	152509226	ER_a	0.82529	999	ER-a
162	chr6	152509227	152509695	ER_a	0.61305	468	ER-a
163	chr6	152564315	152565363	ER_a	-0.76654	1048	SYNE1_3_partial
164	chr8	41211786	41213044	SFRP1	0.72881	1258	intergenic
165	chr8	41233829	41234520	SFRP1	0.64138	691	intergenic
166	chr8	41234968	41236415	SFRP1	1.00286	1447	intergenic
167	chr8	41242208	41242878	SFRP1	0.5114	670	SFRP1
168	chr8	41255315	41256279	SFRP1	0.60025	964	SFRP1
169	chr8	41304162	41304874	SFRP1	0.58537	712	intergenic
170	chr8	41276893	41277647	SFRP1	-0.70858	754	SFRP1
171	chr8	41307521	41308359	SFRP1	-0.74345	838	intergenic
172	chr9	21944709	21945442	CDKN2A	0.55384	733	intergenic
173	chr9	21968348	21969060	CDKN2A	0.55985	712	CDKN2A
174	chr9	21973761	21974809	CDKN2A	0.54776	1048	CDKN2A
175	chr9	21978469	21979265	CDKN2A	0.75612	796	CDKN2A
176	chr9	21990964	21991655	CDKN2A	0.52609	691	intergenic
177	chr9	21997587	21998698	CDKN2A	0.60268	1111	CDKN2B
178	chr9	22002380	22003218	CDKN2A	0.75328	838	intergenic
179	chr9	22046066	22046820	CDKN2A	0.54442	754	intergenic
180	chr9	21950639	21951309	CDKN2A	-0.7734	670	intergenic
181	chr9	21983004	21983737	CDKN2A	-0.68403	733	CDKN2A

Supplementary Table 3 (primers)

Supplementary Table 3

iiMVP	Fragment	Oligos Name	Sequence
		UNIVERSAL TAG	gacGGGACACCGCTGATCGTTTA
		Biotynilated universal primer	BIOTAG-GGGACACCGCTGATCGTTTA
PCR Primers			
ATM-mvp1	P1-1	ATM-P1-1F	TgaaaatTTaaatatgatgaaa
		ATM-P1-1R-UNITAG	gacGGGACACCGCTGATCGTTTActcctAcctcaAcctctAaa
	P1-2	(outer)ATM-P1-2FB	TtTttgaTTtgttaggggtgtg
		(outer)ATM-P1-2RB	TCAAAAATTCTAAAAATCCCCAAA
		(inner)ATM-P1-2F	ggagggTaggaggaagTagT
		(inner)ATM-P1-2R	gacGGGACACCGCTGATCGTTTATTACCTTCTATATAAAACAACATACAA
ATM-mvp2	P2-1	(outer)ATM-2-1F1	agttgagatggaagtggtaga
		(outer)ATM-P2-1R1	ctttttattactctAaAaccaaaA
		(inner)ATM-P-2-1F2	aagTTtttggTtTagTaaaatga
		(inner)ATM-P-2-1R2	gacGGGACACCGCTGATCGTTTAcattaacaataaactatt
	P2-2	(outer)ATM-P2-2F3	gagtgTttttaaatagttTattgtTaatg
		(outer)ATM-P2-2R3	AACACAATAATTTTTCTTAACATTTCC
		(inner)ATM-P2-2F4	ttggtTtTagagtaataaaaagta
		(inner)ATM-P2-2R4B	gacGGGACACCGCTGATCGTTTACCAACATAATAAAACCCTATC
ATM-mvp3	P3-1	(outer)ATM_P3_1F	aaagtgaaagTaggTagggtgT
		(outer)ATM-P3-1R	gacGGGACACCGCTGATCGTTTAcccttAaactttcaactctctt
ATM-mvp4	P4-1	(outer)ATM_P4_1F	gttgaattttaaggattgaggt
		(outer)ATM-P4-1R	gacGGGACACCGCTGATCGTTTAttcaattcccacaaaaactaa
	(P4-1 a)	(inner)ATM P4-1 F2	AGATTGAATTGAATGTAATAAGG
		(inner)ATM P4-1 R2	gacGGGACACCGCTGATCGTTTACCATAAATCTAAAACCCTTACCAA
	(P4-1 b)	(inner)ATM P4-1 F3	GGGTTTTAGATTTATGGTAGAAGG
		(inner)ATM P4-1 R3	gacGGGACACCGCTGATCGTTTAAAAAACCTAACATAACCTTCC
	P4-2	ATM-P4-2F	gacGGGACACCGCTGATCGTTTAAGGTGTATTGATTAATTTT
		ATM-P4-2R	AaacaAcactAaactcaaact

ATM-mvp5	P5-1	ATM-P5-1F	gacGGGACACCGCTGATCGTTTAGTAGTGAACTACACTGTTTGAT
		ATM-P5-1R	ttActAAattccttcaaatc
	P5-2	(outer)ATM-P5-2F	GAAGTGGTTATGATTTTGTT
		(outer)ATM-P5-2R	aAAttAAttAcacaccactAAaaa
		(inner)ATM P5-2 F2	gacGGGACACCGCTGATCGTTTATTTTTGTTATGATATAAGTGTGTGTTT
		(inner)ATM P5-2 R2	TTACACACCACTAAAAATATATTTT
ATM-mvp6	P6-1	ATM_P6_1F	gtgtggTTtTtTagaTtagtt
		ATM-P6-1R-UNITAG	gacGGGACACCGCTGATCGTTTAAcctccaaAatAAtacacttA
	P6-2	ATM-P6-2F-UNITAG	gacGGGACACCGCTGATCGTTTAaggaTTtgggttttaggtT
		ATM_P6_2R	cactAtcacatttaactcttAcaAAa
ATM-mvp7	P7-1	ATM-P7-1F	TTTGGGTTTTAGGTTTGATT
		ATM-P7-1R-UNITAG	gacGGGACACCGCTGATCGTTTACAAAACACCAACTAAATACCA
ATM-CGI	ATM-CGI	ATM-cgi-F1	gagggtgggtgagagTTT
		ATM-cgi-R1	ctAcccaACRcccacttct
		ATM-cgi-F2	GTTGGTTATTGGTGGATATGG
		ATM-cgi-R2	TCAAAACACTACCCCAAAACATT
Sequencing Primers			
ATM-mvp1	P1-1	ATM-P1-1pos1,_2	AATTTTAGTATTTTGGGAG
		ATM-P1-1pos3_ 4_ 5	gtTtTtaTtaaaaataTaaaaatta
	P1-2	ATM-P1-2SEQ pos 1 2 3	gattattattttttattagataag
		ATM-P1-2SEQ pos 4 5	gttttagttgtaagttgttgg
ATM-mvp2	P2-1	ATM-P2-1-SEQpos1	aagCCtttggCtCagCaaaatga
		ATM-P2-1-SEQpos2 3 4 5 6	gtCaCCaggCtggagtgCag
	P2-2	ATM-P2-2POS1_2	AGTAATAAAAAGTAAATATATTT
		ATM-P2-2pos_3	TTTAGGTTGGAGTATAGTGG
		ATM-P2-2pos_4	TTTTGTTTTTTAGATTTAAG
		ATM-P2-2pos_5,6,7,8	TTTTGAGTAGTTGGGATTAT
ATM-mvp3	P3-1	ATM-P3-1POS1_2	AGGTAGGGTGTAGTGGTT
		ATM-P3-1SEQpos3	ATTTTGTTTTTAAAAAAAT
ATM-mvp4	P4-1	ATM-P4-1-SEQpos2	AAGAGGTTTATTTGGTATAT
		ATM-P4-1-SEQpos3	GAGGAAGTAAGGGTGGA
	P4-2	ATM-P4-2-SEQpos1	CTAACATTTAATCACTCTAA
		ATM-P4-2pos_2	AAAAAATAAACTATTCTCTA

ATM-mvp5	P5-1	ATM-P5-1-SEQpos1	TGGAAACACTAACCAAAA
		ATM-P5-1-SEQpos 2	TGTTAAGTTTTAATTTTTA
	P5-2	ATM-P5-2pos_2_3_4	CCACTAAAAATATATTT
		ATM-P5-2-SEQpos1	ATAACTCACCATTAATAATAT
ATM-mvp6	P6-1	ATM-P6-1pos1_unique	TTAGTTTTTAAAAGGTATT
	P6-2	ATM-P6-2pos1_unique	CAAAAATATAATAAAATAAATAC
ATM-mvp7	P7-1	ATM-P7-1-SEQunique	TTTGTAAAATGGGGATAATA
ATM-CGI	ATM-CGI	ATM-cgi-S1	TGGTTATTGGTGGATATG

Supplementary Table 4 (Cell lines)

Cell line	Tumour type	Cell type
1321 NI	Human Brain Astrocytoma	Brain
SK-PN-DW	Human malignant primitive neuroectodermal tumor	Brain
IMR-32	Human neuroblastoma	Brain
SK-N-FI	Human neuroblastoma metastatic	Brain
SK-N-SH	Human neuroblastoma metastatic	Brain
SK-N-AS	Human neuroblastoma metastatic	Brain
SK-N-DZ	Human neuroblastoma metastatic	Brain
MCF7	Human breast adenocarcinoma	Breast
MDA MB-231	Human breast adenocarcinoma	Breast
SK-BR3	Human breast adenocarcinoma metastatic	Breast
BT549	Human breast ductal carcinoma	Breast
T47D	Human breast ductal carcinoma	Breast
HELA	Human Epitheloid Carcinoma of the Cervix	Cervix
Caco 2.2	Human Colonic Adenocarcinoma	Colon
T 84	Human Colonic Adenocarcinoma	Colon
SW480	Human Colonic Adenocarcinoma	Colon
DLD1	Human Colonic Adenocarcinoma	Colon
SW 480	Human Colonic Adenocarcinoma	Colon
HEP 2	Human Larynx Carcinoma	Head and Neck
ECV 304	HUVEC	HUVEC
293 gp	Human Kidney (HEK-293) + GagPol	Kidney
293T	Human Kidney (HEK-293) + SV40 T-antigen	Kidney
T2	B lymphoblastic cell line)	LCL
K 562	Human Leukaemia	Leukaemia
JURKAT	Human Leukaemic T Cell	Leukaemia
MOLT-4	Human Lymphoblastic Leukemia	Leukaemia
THP-1	Human Monocyte	Leukaemia
HL-60	Human Promyelocytic Leukaemia	Leukaemia
HEP3B	Human Hepatocellular Carcinoma	Liver
HEP G2	Human Hepatocellular Carcinoma	Liver
A 549	Human Lung Carcinoma	Lung
BC3.14	Human B-cell lymphoma (KSHV pos)	Lymphoma
DG75	Human burkitts lymphoma (EBV neg)	Lymphoma
Ramos	Human burkitts lymphoma (EBV neg)	Lymphoma
Raji	Human burkitts lymphoma (EBV pos)	Lymphoma
P3HR1	Human burkitt's lymphoma (EBV pos)	Lymphoma
BCP-1	Human lymphoma (KSHV pos, EBV neg)	Lymphoma

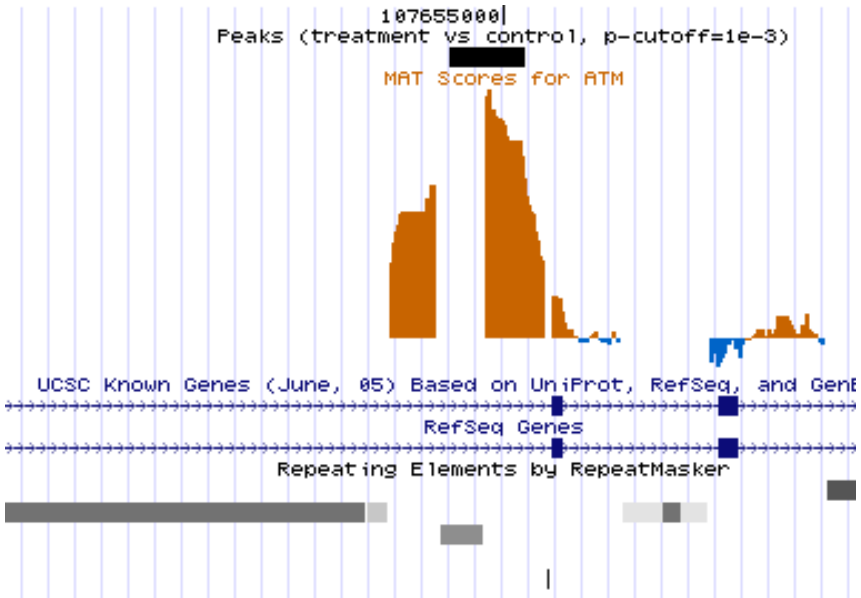
U 937	Human Monocytic Lymphoma	Lymphoma
BCBL-1	Human primary effusion lymphoma (KSHV pos)	Lymphoma
HBL-6	Human primary effusion lymphoma (KSHV pos?)	Lymphoma
A 375	Human Melanoma	Melanoma
MSC	mesenchymal stem cells	MSC
HOB	Human osteoblasts	Osteoblasts
SK-OV-3	Human ovary adenocarcinoma	Ovary
Capan-1	Human Pancreatic Tumour	Pancreatic
BeWo	Human placenta choriocarcinoma	Placenta
PC3	Human Prostate carcinoma	Prostate
SW872	Human liposarcoma	Sarcoma
HOS	Human osteosarcoma	Sarcoma
SAOS2	Human osteosarcoma	Sarcoma
U2OS	Human osteosarcoma	Sarcoma
G 292	Human Osteosarcoma	Sarcoma
RD	Human rhabdomyosarcoma	Sarcoma
A204	Human rhabdomyosarcoma	Sarcoma
SJRH30	Human rhabdomyosarcoma metastatic	Sarcoma
SW982	Human synovial sarcoma	Sarcoma
SK-LMS-1	Human vulva leiomyosarcoma	Sarcoma
SW1353	Human chondrosarcoma	Sarcoma
SK-VT-1	Human Ewing sarcoma	Sarcoma
RH1	Human Ewing sarcoma	Sarcoma
HT1080	Human fibrosarcoma	Sarcoma
SW684	Human fibrosarcoma	Sarcoma
A 431	Human Epidermoid Carcinoma	skin

A

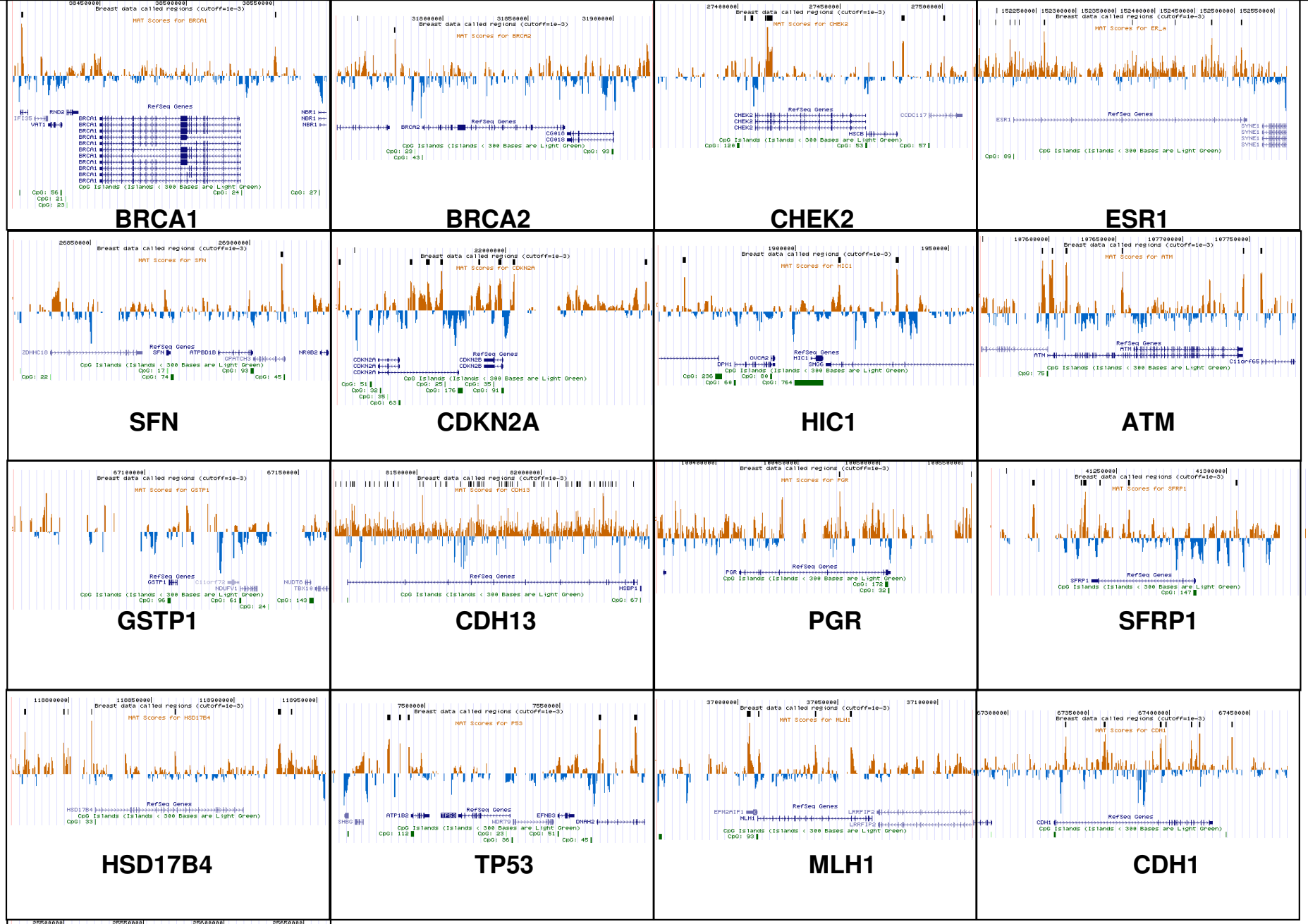
4Mb tiled sequence

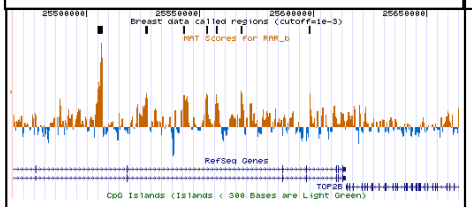
B

ATM mvp4



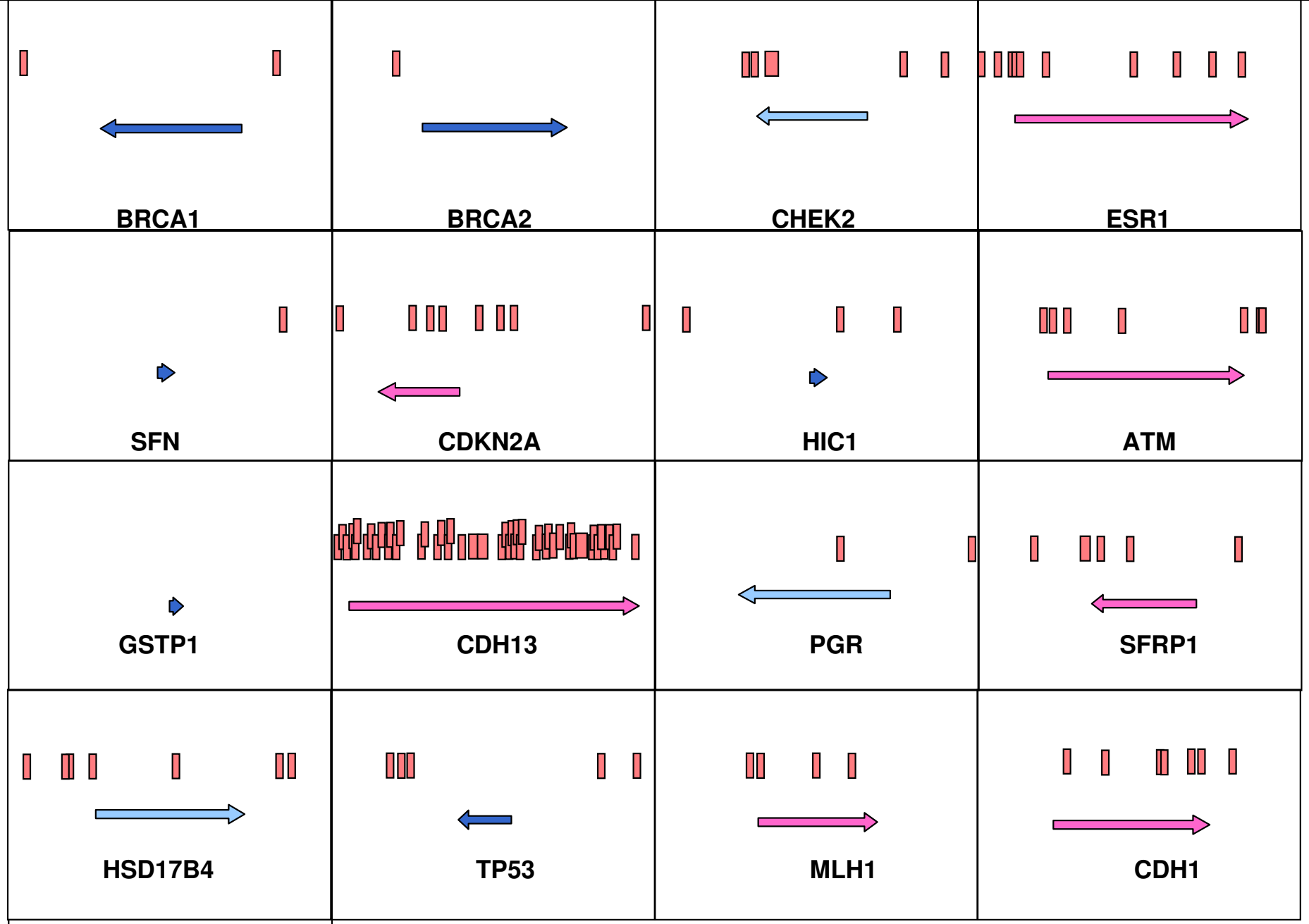
Supplementary Figure 1.

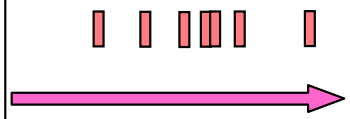




Gene lengths
not to scale

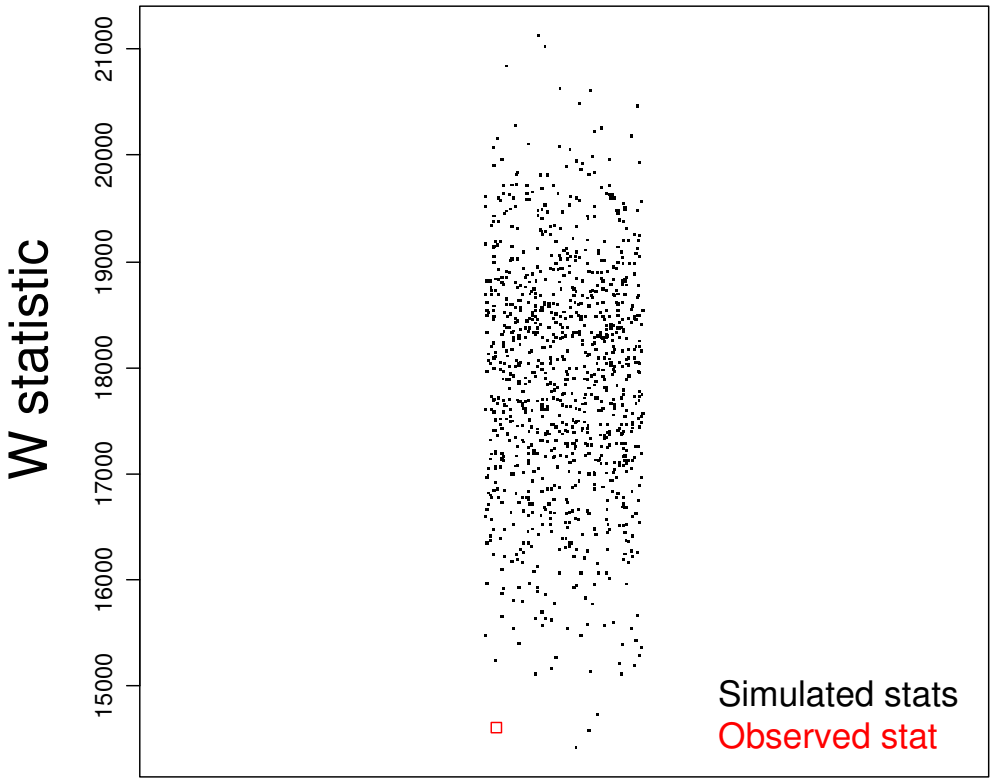
Supplementary Figure 2A.





Gene lengths
not to scale

Supplementary Figure 2B.

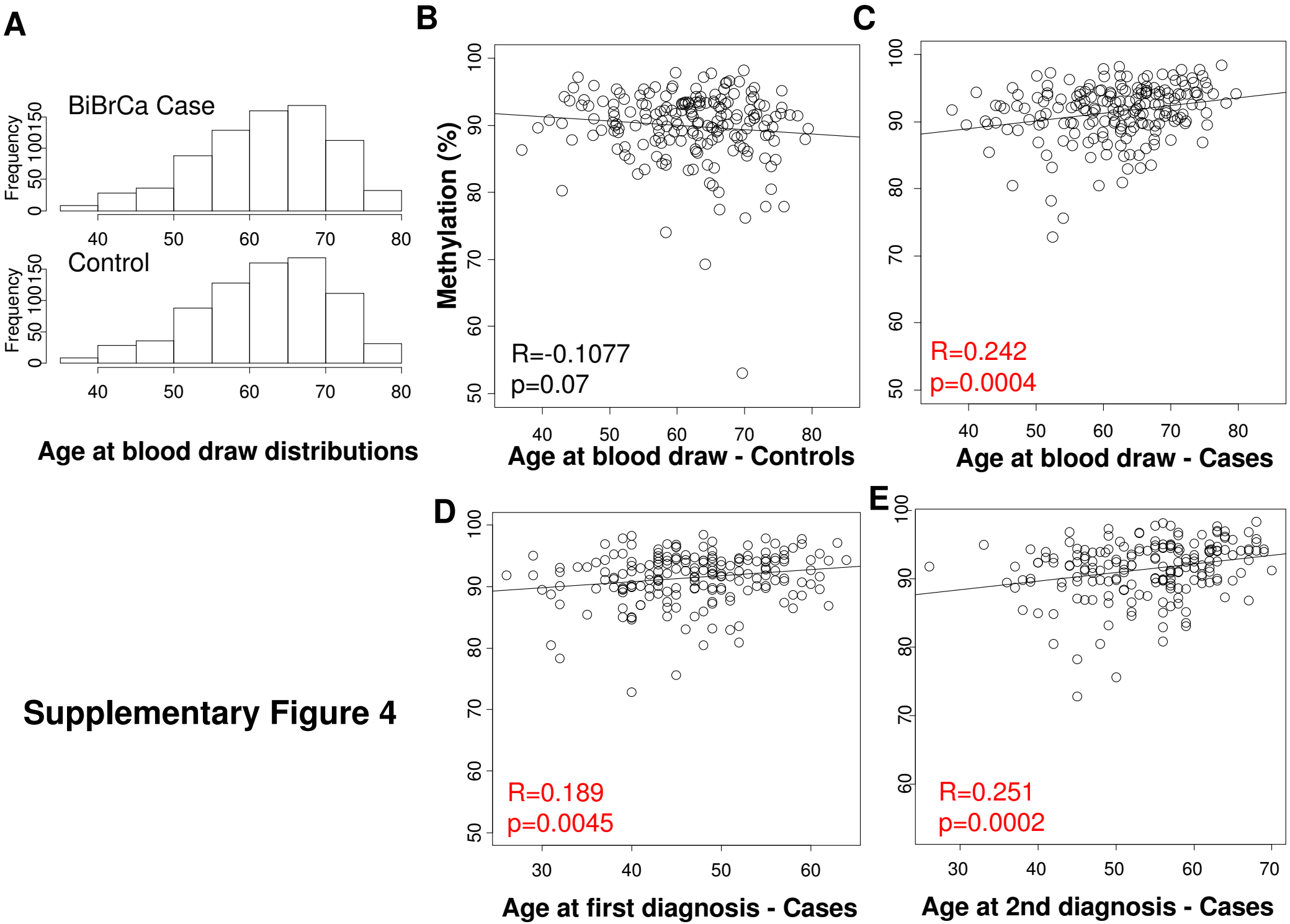


$$P = 1 - \frac{(\#Wsimulated > Wobserved)}{1000}$$

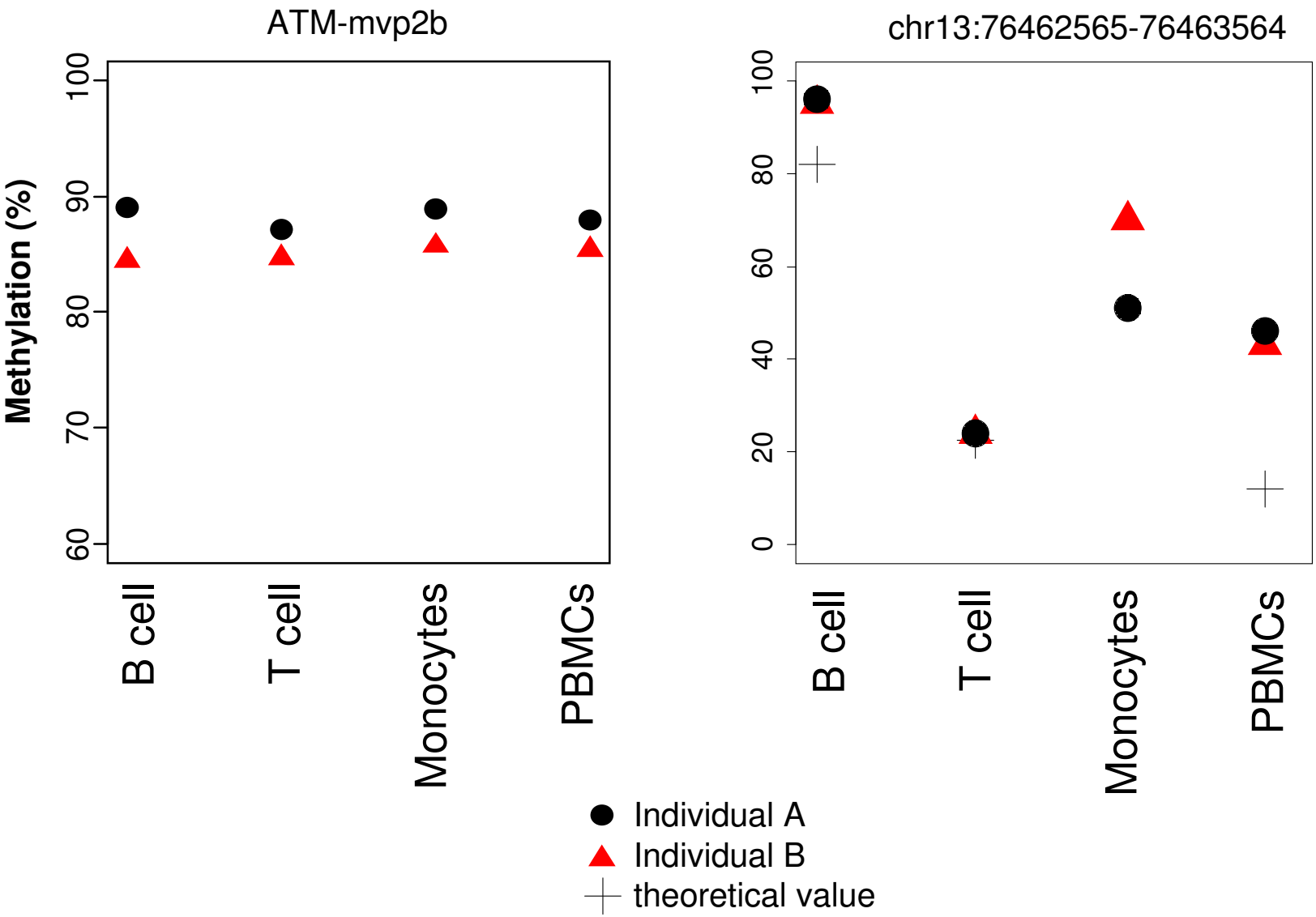
P= 1-(998/1000)

P=0.002

Supplementary Figure 3



Supplementary Figure 4



Supplementary Figure 5