SUPPLEMENTARY FILE 2: SIMULATION STUDIES

In this document, we first describe the correlatoin preserving simulation scheme in details. We then provide detailed simulation results, from an extensive simulation study.

1. Simulating methylation data

We adapt the strategy of Gaile et al (2007) to generate (spatial) correlation-preserved methylation data. More specifically, we consider the comparison of methylation β values for two subgroups (e.g., low and high exposure). Simulated datasets consist of methylation assay β values for $n = 40$ samples from each subgroup. Assignment of methylation β values to the samples is designed to preserve the true correlation structure within regional blocks of dense array coverage (described in greater detail below).

1.1. **Description of Data.** The dataset on which the simulations were based consists of batch-corrected methylation β values for $N = 539$ breast invasive adenocarcinoma samples obtained from The Cancer Genome Atlas (TCGA); specific dataset descriptions are provided in the Appendix of this supplement. These 539 samples were assayed with the Illumina 450K array and provide estimates of 485577 β methylation values located across the genome. In the simulations, to be described more thoroughly later, we first performed an initial study using loci from chromosome 1. We used these simulations to find optimal parameters, and then performed simulations on the entire Epigenome data.

After ordering the methylation values in each chromosome by location, we partition each chromosome into regional blocks by defining breakpoints in areas of low array/site coverage. Specifically, breakpoints were formed between adjacent sites greater than 10K base-pairs apart. Any singleton sites were subsumed into the closest block (in terms of base-pair distance) with two or more members. Thus, contiguous portions of each chromosome that are densely assayed will be grouped within the same regional block so that the correlation in these dense areas will be preserved.

1.2. **Selectively Weighted Observations.** Each of the simulated datasets was constructed by preferentially re-sampling pre-specified regional blocks based on methylation values at the locations of M putative ”targets” ($M = 5$ for chromosome 1 simulations and $M = 10$ for the Epigenome simulations). The sites selected to serve as “targets” satisfy three properties: (1) evidence of a substantial variability in methylation β values in the original 539 samples, (2) evidence of substantial correlations with neighboring sites ($cor > .5$ for at least two neighbors) within the same regional block across the original 539 samples, and (3) none of the M targets could be located within the same regional block. Note that although the preferential re-sampling to be described below is focused on single targets, due to the correlation structure among neighboring sites, the entire correlated "cluster" will be associated with exposure.

Let $\mathbf{Y}_{(IxN)}$ represent the data matrix of the β values, where the Y_{ij}^{th} element is the methylation β value of the i^{th} site for the j^{th} subject, $i = 1, \ldots, I$, $j = 1, \ldots, N$. We derive from $\mathbf{Y}_{(IxN)}$ the simulated data matrix $\mathbf{W}_{(Ix2n)}$ methylation β values, where the rows again correspond to the I sites, but the columns are such that the first n correspond to subjects from subgroup H (representing high exposure) and the second n columns correspond to subjects from subgroup L (simulating low exposure). Let $\gamma(m)$ be the function that maps the m^{th} target, $m = 1, \ldots, M$, to its site index i, and similarly let $\eta(m)$ be the function that maps the m^{th} target to its regional index ℓ, $\ell = 1, \ldots, L$. Thus $\mathbf{Y}_{\gamma(m)j}$ is the m^{th} target methylation β value for the j^{th} subject.

For a prespecified L, n, M, and selection weights w_m (to be described below), we propose the algorithm below to create simulated methylation datasets which preserve intra-region correlations.

1.3. **Algorithm.** Let $J = 1, \ldots, N$, the set of all subject indices. For the mth regional block containing a target, $\gamma(m)$, $m = 1, \ldots, M$:

(1) Define sampling weights for each subject and create index groups
  - Obtain $r_{H,j} : r_{H,j} = rank(\mathbf{Y}_{\gamma(m)j})/(N+1), j \in J$, where ranks are evaluated with respect to $\{\mathbf{Y}_{\gamma(m)j} | j \in J\}$.
  - Sample n indices without replacement from J according to weights $P_j^{H,m} = (1 - r_{H,j})^{w_m}$ to form index set $H, H \subset J$ and $|H| = n$.
  - Define $J_L = J \cap H^c$, the set of all indices not contained in set H, and obtain $r_{L,(j)}$: $r_{L,(j)} = rank(\mathbf{Y}_{\gamma(m)(j)})/[(N-n)+1]$, $(j) \in J_L$, where ranks are evaluated with respect to $\{\mathbf{Y}_{\gamma(m)(j)} | (j) \in J_L\}$.
  - Sample n indices without replacement from J_L according to weights $P_{(j)}^{L,m} = (r_{L,(j)})^{w_m}$ to form index set $L, L \subset J_L$ and $|L| = n$.

(2) Assign regional blocks $\eta(m)$ to the simulated subgroups of size n
  - Assign the n regional blocks $\eta(m)$ indexed by set H to the simulated subgroup H.
  - Assign the n regional blocks $\eta(m)$ indexed by set L to the simulated subgroup L.

For each of the remaining regional blocks, randomly sample $2n$ of the corresponding region without replacement from J; assign n to subgroup H and n to subgroup L.

Figure 1 demonstrates the sampling procedure in a sequence of three regional blocks. Note that the pseudocode samples each regional block at a time, while the figure shows the process as if it is in parallel.

1.4. **Selection Weights.** Selection weights, w_m , were tuned such that the average value (across 1000 simulations) of single-site statistics matched typical "high signal" sites. We specifically targeted Wald statistics such that the mean (across 1000 simulations) single-site (unadjusted) p-values were approximately 0.001.

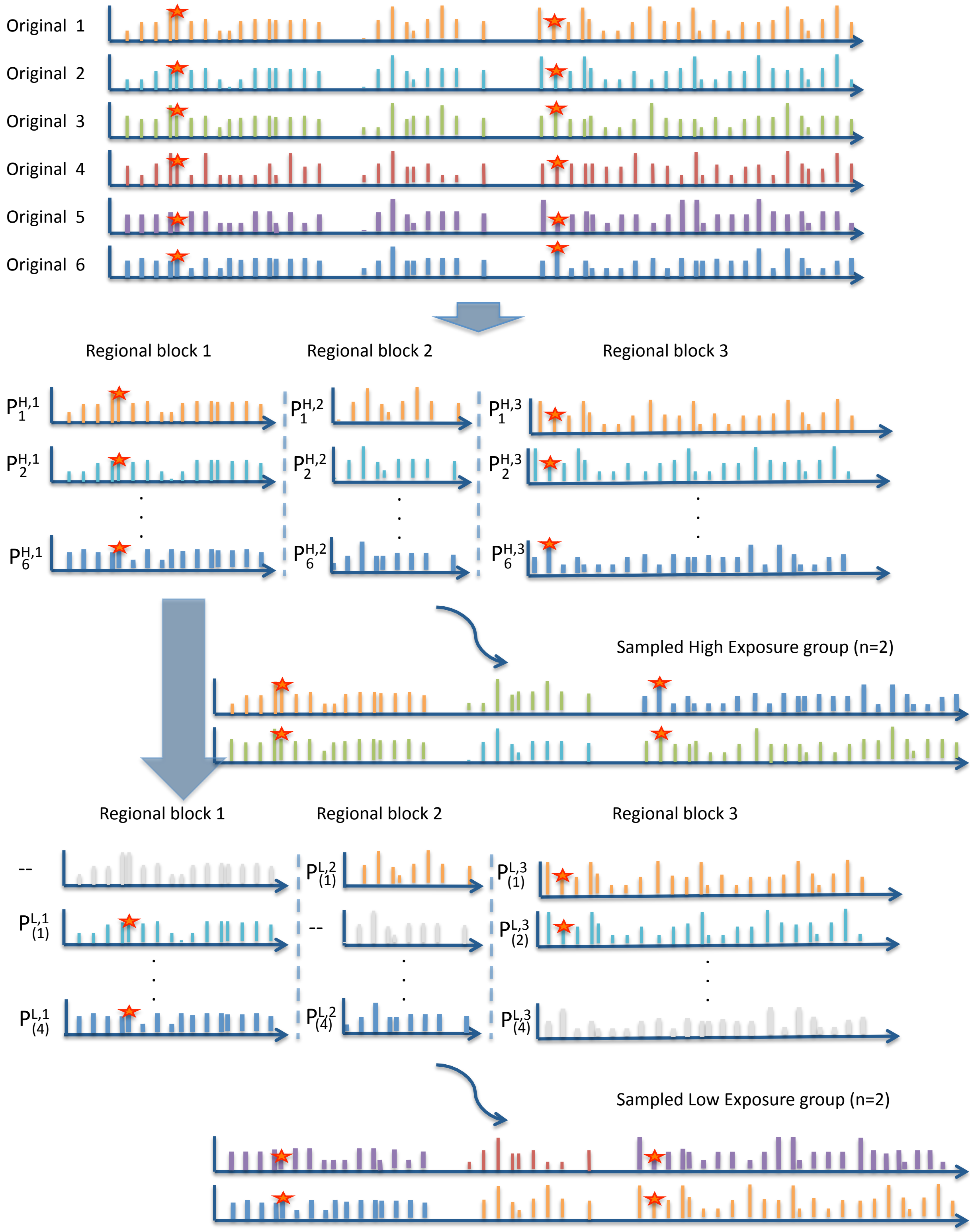


FIGURE 1. Preferential resampling of regional blocks 1,2, and 3 into high (H) and low (L) exposure groups according to calculated sampling probabilities $P_j^{H,m}$ and then $P_{(j)}^{L,m}$. The red stars represent the $M = 2$ targets in this example.

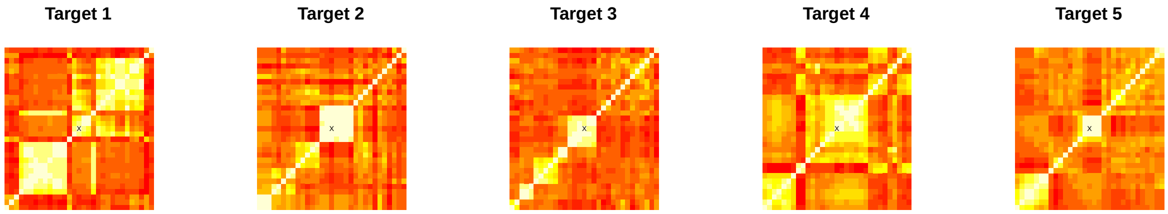


FIGURE 2. Pearson correlation heatmaps from the N = 539 breast invasive adenocarcinoma samples for sites in the neighborhood (± 15 sites) of the five targets selected for chromosome 1 simulation. The X marks the target.

2. CHROMOSOME 1 SIMULATIONS

We first performed simulations using chromosome 1 data. There were 30796 methylation sites after quality control. There were 2861 regional blocks. We first carried out an initial set of simulations with all combinations of the following analysis parameters: distance threshold $\bar{\mathcal{D}} = 0.25$, Pearson and Spearman correlations, distance type *single, average* and *complete*, with and without d_{bp}-merge (merging sites up to 999 base pairs away from each other) and with and without base-pair distance restriction for merging $\bar{d}_{bp} = 1000$. After seeing that either the *complete* or *average* distance types are more performant than the *single*, and that restricting by base-pair distance $\bar{d}_{bp} = 1000$ is beneficial, we performed sensitivity analysis in which we varied the distances $\bar{\mathcal{D}}$, for the combinations of Pearson and Spearman correlations, *complete* and *average* distances types, and with and without d_{bp}-merge. We first describe the initial chromosome 1 simulations results, and then the sensitivity analysis results.

2.0.1. *Bump Hunting settings.* All defaults settings were used in `dmrFind` to implement Bump Hunting "default", with the exception of (1) *maxGap*= 700 for `clusterMaker`, so it would pick up all clusters as a possible clusters (default setting did not allow this) (2) no batch adjustment as the data were previously batch-adjusted. Notably, the default choice for option *sortBy* is "area.raw", which influences how the qvalue is computed (see below). Method "adjusted" uses sortBy = "max".

Figure 2 provides the (Pearson) correlation heatmaps for neighboring regions of the targets selected for simulation.

2.1. **Initial chromosome 1 simulations.** We simulated 100 data sets. The results, averaged over the 100 simulations, are reported in Tables 1, and 2. Table 1 compares the results of analyses based on the different clustering parameters, and the Bump Hunting's results. For each clustering method we report:

- Cluster = number of clusters identified (not necessarily differentially methylated) with 3 or more members. Note this quantity is not reported for Bump Hunting. Bump

Hunting does report a list of potential DMRs, but they are selected according to having a possible exposure effect, while in Aclust they are independent of exposure effect.

- Membj = number of sites in cluster containing jth target (j=1,..5) when the cluster containing the target was identified.
- TPRj = indicator of whether cluster containing target j (j=1,..5) had Benjamini-Hochberg (BH) adjusted P-value < 0.05 for Aclust or q-value ¡0.05 for Bump Hunting.
- TPR = indicator of whether all clusters containing each target had adjusted P-value < 0.05 (Aclust) or q-value¡0.05 (Bump Hunting).
- FP = number of clusters with adjusted P-value < 0.05 (Aclust) or q-value¡0.05 (Bump Hunting) that do not contain any of the 5 targets
- FPR = indicator of whether at least 1 cluster had adjusted p-value < 0.05 and the cluster did not contain the target.
- Time = "elapsed time" in seconds (third entry from R proc.time() output) using Odysessy cluster. For Bump Hunting, the timing is dependent on the number of iterations used in the q-value computation. We used 250 iterations to compute this q-value.

We also report single-site analysis results. Comparing a regional-based analysis and a single site analysis is difficult in that (1) from a regional-based perspective, each of the sites in a cluster is a valid site to detect, and (2), the clusters, in this form of simulations, are not clearly defined, as is seen by the fact that different parameter settings yield slightly different cluster detection (e.g. a cluster may have 4 sites when using a *complete* type, but 5 sites when *single* is used). Therefore, the single-site analysis results *appear* to be different if judged by the results of different clustering methods. For instance, continuing the previous example, if only 4 sites were detected as belonging to a simulated differentially methylated cluster, and the single-site analysis detected a 5th site, adjacent to this cluster, it would be judged as a false positive detection. But if one uses the clustering results of a method that determined that this 5th site in fact belongs to this cluster, it wouldn't be judged as a false positive. For the sake of being comprehensive, we report the single-site results as judged by all clustering implementations in Table **??**. To summarize these results, the regional-based analyses are more powerful than the single-site analysis.

	Cluster	Memb1	TPR1	Memb2	TPR2	Memb3	TPR3	Memb4	TPR4	Memb5	TPR5	TPR	FP	Time
Pearson Correlation														
A-clustering; $\bar{\mathcal{D}}$=0.25														
Single														
d+Aclust	1935.38	6.00	0.56	7.02	0.70	6.00	0.61	8.66	0.71	4.00	0.67	0.17	0.29	233.66
Aclust	1441.02	3.94	0.66	7.00	0.71	6.00	0.65	8.63	0.74	4.00	0.73	0.23	0.45	43.41
d+Aclust(1000)	1888.57	6.00	0.56	7.02	0.70	6.00	0.61	8.66	0.71	4.00	0.67	0.17	0.24	194.05
Aclust(1000)	1367.53	3.94	0.66	7.00	0.72	6.00	0.65	8.63	0.74	4.00	0.74	0.24	0.43	116.96
Average														
d+Aclust	1941.86	6.00	0.56	7.02	0.70	6.00	0.61	8.66	0.71	4.00	0.67	0.17	0.25	180.25
Aclust	1312.47	3.41	0.71	7.00	0.72	5.94	0.67	7.95	0.77	4.00	0.75	0.26	0.46	101.07
d+Aclust(1000)	1889.73	6.00	0.56	7.02	0.70	6.00	0.61	8.66	0.71	4.00	0.67	0.17	0.23	438.24
Aclust(1000)	1241.28	3.41	0.71	7.00	0.72	5.94	0.67	7.95	0.77	4.00	0.76	0.27	0.44	273.36
Complete														
d+Aclust	1938.84	6.00	0.56	7.02	0.70	6.00	0.60	8.66	0.71	4.00	0.67	0.17	0.25	192.30
Aclust	1118.18	2.54	0.27	7.00	0.73	5.58	0.68	7.42	0.76	4.00	0.73	0.14	0.42	95.72
d+Aclust(1000)	1888.17	6.00	0.56	7.02	0.70	6.00	0.61	8.66	0.71	4.00	0.67	0.17	0.22	257.34
Aclust(1000)	1057.23	2.54	0.27	7.00	0.73	5.58	0.68	7.42	0.76	4.00	0.74	0.14	0.40	131.09
Spearman Correlation														
A-clustering; $\bar{\mathcal{D}}$=0.25														
Single														
d+Aclust	1096.64	5.99	0.64	7.04	0.73	6.00	0.69	8.82	0.75	4.00	0.76	0.25	0.51	285.90
Aclust	912.01	3.92	0.66	7.00	0.74	6.00	0.71	8.80	0.79	4.00	0.78	0.24	0.58	113.46
d+Aclust(1000)	1063.18	5.99	0.64	7.04	0.73	6.00	0.69	8.82	0.76	4.00	0.76	0.25	0.51	495.45
Aclust(1000)	863.61	3.92	0.66	7.00	0.75	6.00	0.71	8.80	0.79	4.00	0.78	0.25	0.57	243.45
Average														
d+Aclust	1098.25	5.99	0.64	7.04	0.73	6.00	0.68	8.82	0.75	4.00	0.76	0.24	0.49	266.90
Aclust	814.18	3.72	0.45	7.00	0.75	5.79	0.73	8.25	0.83	3.99	0.78	0.19	0.56	95.59
d+Aclust(1000)	1063.11	5.99	0.64	7.04	0.73	6.00	0.69	8.82	0.76	4.00	0.76	0.25	0.50	441.24
Aclust(1000)	769.87	3.72	0.45	7.00	0.77	5.79	0.73	8.25	0.83	3.99	0.79	0.19	0.54	205.87
Complete														
d+Aclust	1094.61	5.99	0.64	7.04	0.73	6.00	0.68	8.82	0.75	4.00	0.76	0.24	0.49	262.80
Aclust	681.27	2.24	0.12	6.99	0.78	5.09	0.75	7.28	0.83	3.99	0.79	0.07	0.51	85.22
d+Aclust(1000)	1061.98	5.99	0.64	7.04	0.73	6.00	0.69	8.82	0.76	4.00	0.76	0.25	0.50	262.26
Aclust(1000)	644.39	2.24	0.12	6.99	0.78	5.09	0.76	7.28	0.83	3.99	0.80	0.07	0.47	83.90
Bump Hunting														
default	–	4.49	0.00	7.00	0.06	5.99	0.00	9.49	0.31	4.00	0.00	0.00	0.14	785.08
adjusted	–	4.49	0.59	7.00	0.69	5.99	0.62	9.49	0.76	4.00	0.42	0.11	0.23	789.48

Table 1. Clustering Results in the Chromosome 1 Simulation Study. All basic A-clustering methods are noted by Aclust. The addition of (1000) in parenthesis, as in Aclust(1000) and d+Aclust(1000) means that 1000-$\bar{d}_{bp}$ restriction was posed so that two adjacent sites could not be merged if there were at least 1000bp between them. A prefix of d+, as in d+Aclust and d+Aclust(1000) stands for 999 d_{bp}-merge initiation. Cluster is the number of clusters of at least 3 sites identified by Aclust. Memb$m$ is the mean number of sites identified in the m cluster. TPRm is the proportion of simulations in which cluster m was determined to be significantly associated with exposure after FDR correction. TPR is the proportion of simulations in which all 5 sites were significantly detected as DMRs. FP is the mean number of falsely detected clusters. Time is the mean analysis computation time in seconds.

	fuzz1	fuzz2	fuzz3	fuzz4	fuzz5	1inallclust	FPR
Pearson							
d+Aclust BH	0.66	0.68	0.60	0.74	0.64	0.20	0.70
d+Aclust BY	0.47	0.48	0.33	0.50	0.38	0.02	0.09
Aclust BH	0.62	0.68	0.60	0.74	0.64	0.19	0.77-0.84
Aclust BY	0.47	0.48	0.33	0.50	0.38	0.02	0.31-0.37
Spearman							
d+Aclust BH	0.66	0.68	0.60	0.74	0.64	0.20	0.70
d+Aclust BY	0.47	0.48	0.33	0.50	0.38	0.02	0.10
Aclust BH	0.62	0.68	0.60	0.74	0.64	0.19	0.78-0.84
Aclust BY	0.45-0.47	0.48	0.33	0.50	0.38	0.02	0.31-0.44

Table 2. Single site results *as judged by* Pearson or Spearman correlation based clustering implementations. The results are summarized across different implementations due to similarity in results, so that each row represent range of values for each of *single, average* and *complete* types, and with or without max.dist 1000-$\bar{d}_{bp}$ restriction. $\bar{\mathcal{D}}$=0.25 always. BH and BY are Benjamini-Hochberg and Benjamini-Yekutieli. fuzz$m$ is the proportion of simulations in which at least one site in cluster m had adjusted p-value < 0.05. 1inallclust is the proportion of simulations in which sites were detected in all 5 clusters. FPR is the proportion of simulations in which at least 1 site was detected that is not in one of the 5 clusters.

2.2. **Sensitivity analysis results.** In the sensitivity analysis, we still compared the Spearman and Pearson correlations, with and without d_{bp}-merge initiation, but dropped the *single* clustering type and always restricted the base pair distance for merging $\bar{d}_{bp}$ (max.dist in the `Aclust` R package). The results are slightly complicated by the fact that although the number to detect is very well defined (5 clusters), and thus the number of false positive detections and false negatives are well defined, the RATES are variable as the total number of identified clusters is not fixed across simulations. For instance, usually, if the number of clusters was predefined in advance, false positive rate will be the number of cluster detections that are not in fact DMRs, out of the total number of non-DMR clusters. But here, each clustering methods discovers a slightly different number of clusters, and there is no "correct" number. Therefore, we considered two measures of total error, namely TE and TE*, and eventually based our "winning" method on a measure denoted by TE*. **Based on this sensitivity analysis, we chose the most appropriate parameter settings for the clustering algorithm as Spearman correlation,** *average* **type,** $\bar{\mathcal{D}} = 0.2$**, and with** d_{bp}**-merge initiation.**

Complete report of analyses is next in Tables 4 and 5. For each clustering settings, we report the mean across 100 simulations of:

- FPR (false positive rate). In each simulation, the false positive rate is calculated as the proportion of true non-DMR clusters that were detected as DMRs. ($B/(A+B)$ in Table 3).
- FNR (false negative rate). In each simulation, the false negative rate is the proportion of true DMR clusters that were determined to be non-differentially methylated.($C/(C+D)$ in Table 3).
- TE (total error). The proportion of clusters with wrong classification (i.e. the number of clusters wrongly labeled as DMRs or non-DMRs, out of the total number of clusters). ($(B+C)/(A+B+C+D)$ in Table 3).
- TE* (another measure of total error). FPR + FNR. ($B/(A+B)+C/(C+D)$ in Table 3).

Table 4 provides the sensitivity analysis results for all parameters combinations WITH d_{bp}-merge initiation, and Table 5 provides the sensitivity analysis results for all parameters combinations WITHOUT d_{bp}-merge initiation. Finally, Table 6 provides a detailed simulation results for the "best" scenarios, corresponding to "best" $\bar{\mathcal{D}}$ in each one of combinations of Pearson/Spearman correlation, *average*/*complete* type, and with/out d_{bp}-merge initiation.

	Observed	
Truth	0	1
0	A	B
1	C	D

Table 3. Classification table to clarify the sensitivity analysis results. Let Truth 0/1 is an indicator of whether a cluster contains a target, and Observed 0/1 is an indicator of whether a cluster was deemed differentially methylated based upon adjusted p-value < 0.05. A-D are counts of cross-classified clusters within a single simulated data set. The measures reported for the sensitivity analysis (averaged over 100 simulations) are FPR=$B/(A+B)$, FNR=$C/(C+D)$, TE=$(B+C)/(A+B+C+D)$, and TE*=$B/(A+B)+C/(C+D)$.

Average	Pearson				Spearman			
dist	FPR	FNR	TE	TE*	FPR	FNR	TE	TE*
0.05	0.00052	0.75600	0.02015	0.75652	0.00514	0.80000	0.07907	0.80514
0.10	0.00043	0.34600	0.00331	0.34643	0.00196	0.37200	0.00979	0.37396
0.15	0.00029	0.28000	0.00160	0.28029	0.00091	0.27000	0.00364	0.27091
0.20	0.00019	0.31000	0.00122	0.31019	0.00057	0.24800	0.00218	0.24857
0.25	0.00012	0.35000	0.00105	0.35012	0.00047	0.28400	0.00181	0.28447
0.30	0.00013	0.36400	0.00096	0.36413	0.00032	0.31400	0.00147	0.31432
0.35	0.00013	0.36800	0.00089	0.36813	0.00026	0.33200	0.00123	0.33226
0.40	0.00014	0.37600	0.00087	0.37614	0.00019	0.35200	0.00104	0.35219
0.45	0.00013	0.38000	0.00084	0.38013	0.00016	0.37200	0.00095	0.37216
0.50	0.00013	0.38400	0.00083	0.38413	0.00014	0.38000	0.00087	0.38014
Complete	Pearson				Spearman			
dist	FPR	FNR	TE	TE*	FPR	FNR	TE	TE*
0.05	0.00052	0.75600	0.02015	0.75652	0.00514	0.80000	0.07907	0.80514
0.10	0.00043	0.34600	0.00331	0.34643	0.00196	0.37200	0.00980	0.37396
0.15	0.00029	0.28000	0.00160	0.28029	0.00091	0.27000	0.00364	0.27091
0.20	0.00019	0.31000	0.00122	0.31019	0.00058	0.24800	0.00218	0.24858
0.25	0.00012	0.35000	0.00104	0.35012	0.00047	0.28400	0.00181	0.28447
0.30	0.00013	0.36400	0.00096	0.36413	0.00032	0.31400	0.00146	0.31432
0.35	0.00013	0.36800	0.00089	0.36813	0.00025	0.33200	0.00123	0.33225
0.40	0.00014	0.37600	0.00087	0.37614	0.00019	0.35200	0.00105	0.35219
0.45	0.00013	0.38000	0.00084	0.38013	0.00016	0.37200	0.00095	0.37216
0.50	0.00013	0.38400	0.00083	0.38413	0.00014	0.38000	0.00087	0.38014

Table 4. d+Aclust Sensitivity with type$\in$ $\{$*average, complete*$\}$ and correlation$\in$ {Pearson, Spearman }; max.dist=1000.

Average	Pearson				Spearman			
dist	FPR	FNR	TE	TE*	FPR	FNR	TE	TE*
0.05	0.01073	0.91200	0.06835	0.92273	0.03197	0.98200	0.22633	1.01397
0.10	0.00140	0.48800	0.01002	0.48940	0.00469	0.49600	0.02412	0.50069
0.15	0.00032	0.35000	0.00339	0.35032	0.00100	0.33600	0.00653	0.33700
0.20	0.00035	0.30000	0.00201	0.30035	0.00084	0.32800	0.00394	0.32884
0.25	0.00036	0.27400	0.00146	0.27436	0.00071	0.28600	0.00256	0.28671
0.30	0.00026	0.30000	0.00122	0.30026	0.00059	0.27200	0.00193	0.27259
0.35	0.00022	0.32400	0.00110	0.32422	0.00044	0.29200	0.00159	0.29244
0.40	0.00019	0.34600	0.00102	0.34619	0.00031	0.30600	0.00132	0.30631
0.45	0.00017	0.35600	0.00095	0.35617	0.00030	0.32600	0.00122	0.32630
0.50	0.00015	0.36400	0.00089	0.36415	0.00022	0.34000	0.00106	0.34022
Complete	Pearson				Spearman			
dist	FPR	FNR	TE	TE*	FPR	FNR	TE	TE*
0.05	0.00881	0.96400	0.08622	0.97281	0.02767	0.99400	0.29170	1.02167
0.10	0.00303	0.58600	0.01625	0.58903	0.01009	0.65600	0.04387	0.66609
0.15	0.00076	0.43600	0.00548	0.43676	0.00166	0.38400	0.00966	0.38566
0.20	0.00047	0.38000	0.00301	0.38047	0.00093	0.36000	0.00511	0.36093
0.25	0.00038	0.36400	0.00210	0.36438	0.00074	0.34200	0.00338	0.34274
0.30	0.00034	0.30400	0.00145	0.30434	0.00068	0.31000	0.00244	0.31068
0.35	0.00027	0.30000	0.00118	0.30027	0.00054	0.27200	0.00176	0.27254
0.40	0.00023	0.32600	0.00108	0.32623	0.00042	0.28400	0.00147	0.28442
0.45	0.00021	0.34800	0.00102	0.34821	0.00033	0.31200	0.00131	0.31233
0.50	0.00017	0.35800	0.00093	0.35817	0.00025	0.33000	0.00115	0.33025

TABLE 5. Aclust Sensitivity with type∈ {*average, complete*} and correlation∈ {Pearson, Spearman }; max.dist=1000.

	Cluster	Memb1	TPR1	Memb2	TPR2	Memb3	TPR3	Memb4	TPR4	Memb5	TPR5	TPR	FP	FPR	Time
A-clustering															
Pearson Correlation															
Average															
d+Aclust(0.15)	1068.86	5.41	0.61	7.00	0.73	5.89	0.70	6.76	0.80	3.95	0.76	0.26	0.31	0.25	143.98
Aclust(0.25)	1241.28	3.41	0.71	7.00	0.72	5.94	0.67	7.95	0.77	4.00	0.76	0.27	0.44	0.30	273.36
Complete															
d+Aclust(0.15)	1068.67	5.41	0.61	7.00	0.73	5.89	0.70	6.76	0.80	3.95	0.76	0.26	0.31	0.25	137.10
Aclust(0.35)	1656.38	3.25	0.70	7.00	0.71	5.96	0.64	8.13	0.75	4.22	0.70	0.22	0.45	0.31	138.95
Spearson Correlation															
Average															
d+Aclust(0.20)	771.60	5.91	0.65	7.00	0.76	6.00	0.73	8.06	0.82	3.98	0.80	0.32	0.44	0.37	260.21
Aclust(0.30)	1018.20	3.79	0.67	7.00	0.73	5.95	0.69	8.78	0.77	4.00	0.78	0.24	0.60	0.40	132.14
Complete															
d+Aclust(0.20)	771.21	5.89	0.65	7.00	0.76	5.92	0.73	8.06	0.82	3.98	0.80	0.32	0.44	0.37	299.06
Aclust(0.35)	1113.85	3.23	0.66	7.00	0.73	5.82	0.70	8.32	0.78	4.15	0.77	0.24	0.60	0.41	115.35
Bump Hunting															
default	–	4.49	0.00	7.00	0.06	5.99	0.00	9.49	0.31	4.00	0.00	0.00	0.14	0.12	785.08
adjusted	–	4.49	0.59	7.00	0.69	5.99	0.62	9.49	0.76	4.00	0.42	0.11	0.23	0.19	789.48

TABLE 6. Clustering Results for best $\bar{\mathcal{D}}$ (written in parenthesis) based on TE*. All basic A-clustering methods are noted by Aclust. A prefix of d+, as in d+Aclust(0.30) stands for 999 d_{bp}-merge initiation. 1000-$\bar{d}_{bp}$ restriction was always used. Type is either *average* or *complete*.

3. EWAS simulations

In this set of simulations, we chose $M = 10$ "targets", none of them overlapping with those used for chromosome 1 simulations, to allow for more scenarios. Pearson and Spearman correlation in the neighborhood of targets are depicted in Figure 3. We compared "the best" parameters settings from chromosome 1 simulations with Bump Hunting. (The "best" settings: Spearman correlation, *average* clustering type, $\bar{\mathcal{D}} = 0.2$, max.dist/$\bar{d}_{bp} = 1000$, and 999-d_{bp}-merge initiation). As in the other simulations, the minimum cluster size was set to 3.

Table 7 gives a brief summary of the simulations with only two reported measures: the average number of TP (true positive) detections across 100 simulations (here the maximum number is 10) and the average number of FP (false positive) detection, or average number of clusters detected that are not in fact DMRs. Table 8 gives the detailed clustering results. For each cluster we report the cluster size (Memb), and TPR (true positive rate). Here TPR is the mean number of simulations in which the cluster was detected as a DMR. Finally, Table 9 provides simple single site analysis results. TPRm provides the proportion of simulations in which the mth *target* (rather than any site in the mth *cluster*, as in the chromosome 1 simulations!) had p-value≤ 0.05 after adjustment, nTP is the mean number of true targets detected in the each of the simulations, and nFP is the mean number of falsely detected sites (i.e. sites that are not the targets).

	TP	FP
d+Aclust	6.93	2.44
Bump Hunting (adjusted)	4.52	1.48

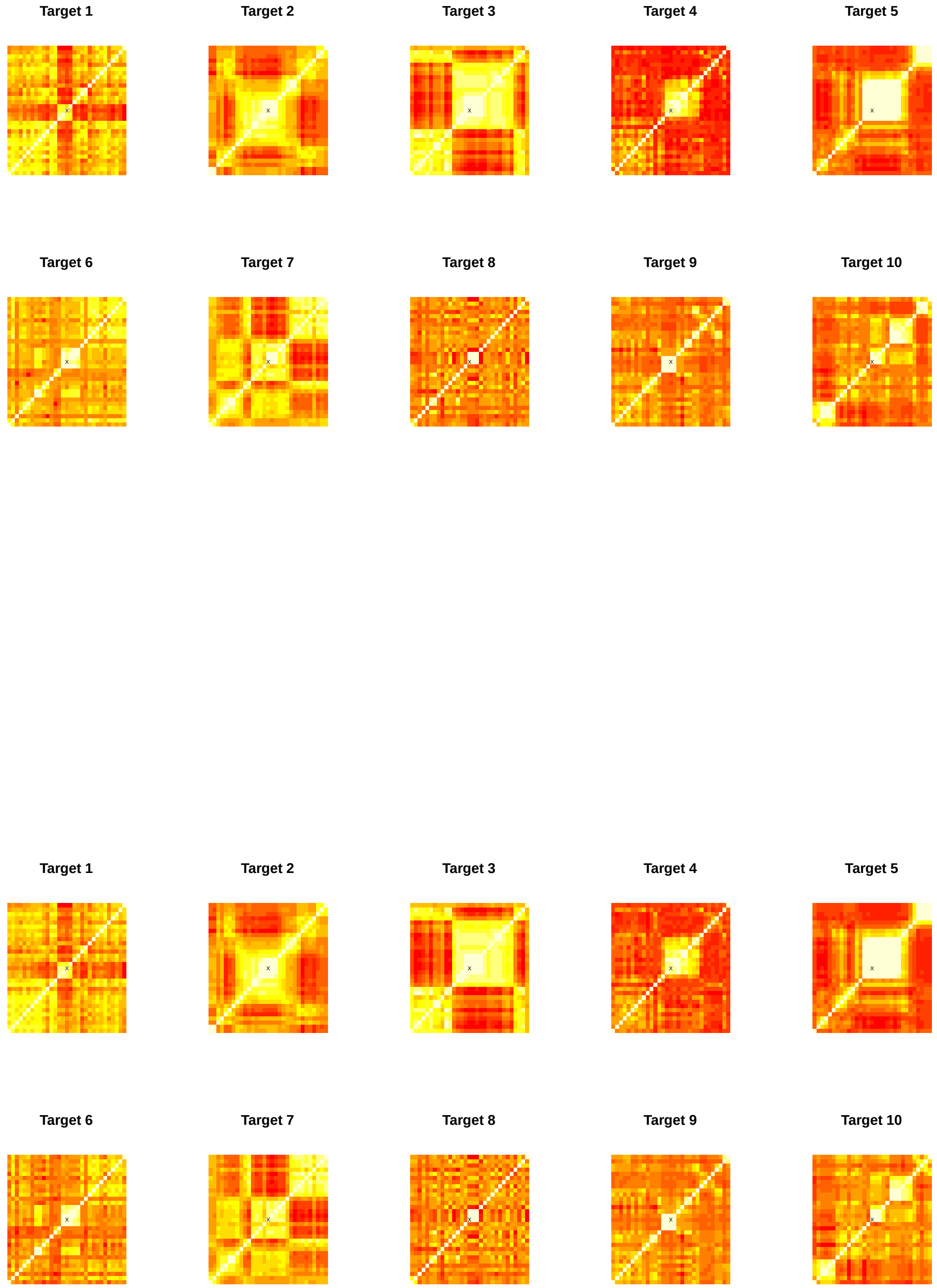
Table 7. Summarized EWAS Clustering Results. Here, d+Aclust uses Spearman correlation, $\bar{\mathcal{D}} = 0.2$, type *average*, max.dist $\bar{d}_{bp} = 1000$. TP= average number of true positive clusters across 100 simulations (ideal value is 10); FP= average number of false positive clusters across 100 simulations.

	Cluster	Memb1 TPR1	Memb2 TPR2	Memb3 TPR3	Memb4 TPR4	Memb5 TPR5	Memb6 TPR6
A-clustering; Spearman, $\bar{\mathcal{D}} = 0.2$, type *average*, $\bar{d}_{bp} = 1000$							
d+Aclust	7270.95	3.56 0.68	7.44 0.66	9.71 0.68	6.23 0.70	10.02 0.67	4.92 0.71
Bump Hunting (adjusted)	303.47	NA 0.00	12.18 0.52	14.47 0.54	6.59 0.62	10.72 0.57	4.28 0.56

	Memb7 TPR7	Memb8 TPR8	Memb9 TPR9	Memb10 TPR10	FP	Time
A-clustering; Spearman, $\bar{\mathcal{D}} = 0.2$, type *average*, $\bar{d}_{bp} = 1000$						
d+Aclust	4.55 0.64	3.00 0.78	4.00 0.68	3.22 0.73	2.44	3774.75
Bump Hunting (adjusted)	9.05 0.54	NA 0.00	4.00 0.55	4.69 0.62	1.48	8924.05

Table 8. Clustering Results. For each of the 10 clusters, we report the number of members identified by each of the methods, and the rate of detection of this clusters. Time is the elapsed computation time (in seconds).

Figure 3. Pearson (top) Spearman (bottom) Correlation heatmaps from the N = 539 breast invasive adenocarcinoma samples for sites in the neighborhood (± 15 sites) of the ten targets selected for EWAS simulation. The X marks the target.



	TPR1	TPR2	TPR3	TPR4	TPR5	TPR6	TPR7	TPR8	TPR9	TPR10	nTP	nFP
BH	0.56	0.51	0.49	0.62	0.49	0.63	0.55	0.57	0.55	0.56	5.53	24.98
BY	0.36	0.35	0.32	0.43	0.37	0.42	0.37	0.34	0.33	0.38	3.67	10.78

Table 9. Single Site Results. TPRm is the proportion of simulations in which target m was statistically significant after FDR correction. nTP is the mean number of targest with true positive detection. nFP is the mean number of false detections (sites that are not the targets). BH and BY are Benjamini-Hochberg and Benjamini-Yekutieli.

Appendix

Datasets for simulation were obtained from http://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp. We used samples from the following data files:

```
jhu-usc.edu_BRCA.HumanMethylation450.Level_3.1.1.0.tar.gz
jhu-usc.edu_BRCA.HumanMethylation450.Level_3.2.1.0.tar.gz
jhu-usc.edu_BRCA.HumanMethylation450.Level_3.3.1.0.tar.gz
jhu-usc.edu_BRCA.HumanMethylation450.Level_3.4.1.0.tar.gz
jhu-usc.edu_BRCA.HumanMethylation450.Level_3.5.1.0.tar.gz
jhu-usc.edu_BRCA.HumanMethylation450.Level_3.6.1.0.tar.gz
jhu-usc.edu_BRCA.HumanMethylation450.Level_3.7.1.0.tar.gz
jhu-usc.edu_BRCA.HumanMethylation450.Level_3.8.1.0.tar.gz
jhu-usc.edu_BRCA.HumanMethylation450.Level_3.9.1.0.tar.gz
jhu-usc.edu_BRCA.HumanMethylation450.Level_3.10.1.0.tar.gz
jhu-usc.edu_BRCA.HumanMethylation450.Level_3.11.1.0.tar.gz
jhu-usc.edu_BRCA.HumanMethylation450.Level_3.12.1.0.tar.gz
jhu-usc.edu_BRCA.HumanMethylation450.Level_3.13.1.0.tar.gz
jhu-usc.edu_BRCA.HumanMethylation450.Level_3.14.1.0.tar.gz
jhu-usc.edu_BRCA.HumanMethylation450.Level_3.15.1.0.tar.gz
jhu-usc.edu_BRCA.HumanMethylation450.Level_3.16.1.0.tar.gz
```