

Identification of differing probes

Delta Beta Method

- Calculation of delta Beta for each probe for each Sample

$$|\beta_{EPIC\ v1} - \beta_{EPIC\ v2}| = \text{Delta Beta}$$

- Median of each probe is taken
- Standard deviation of All the Medians is taken (SD)
- Median of all the Medians is taken (MoM)
- Cutoff value is set as (MoM + 2*SD)
- This was done for the original 16 Samples and 11 Kaur et al Samples
- Probes above cutoff were compared.

Data Analysis Method/Pipeline for Beta-Values

Betas EPIC v1					
Probes	Sample 1	Sample 2	Sample 3	Sample 4	etc.
cg00001510	0.360868	0.406859	0.472296	0.206835	0.391828
cg00101235	0.644655	0.497198	0.846551	0.675296	0.804703
cg08786079	0.717317	0.803968	0.59922	0.755058	0.675481
cg12981137	0.080322	0.069465	0.099431	0.068025	0.271951
cg13699845	0.645325	0.206029	0.164735	0.577102	0.704511
cg15934644	0.678567	0.663951	0.890643	0.881358	0.915806
cg16781484	0.870962	0.903339	0.811486	0.863965	0.745836
cg16819999	0.646274	0.552081	0.782722	0.604336	0.752426
cg19968202	0.755919	0.282533	0.542365	0.712598	0.834492
cg23509869	0.341317	0.360958	0.494545	0.22539	0.526057
cg23729101	0.835866	0.708742	0.916817	0.779065	0.671998
cg24130641	0.88816	0.843137	0.835868	0.837237	0.833462

Betas EPIC v2					
Probes	Sample 1	Sample 2	Sample 3	Sample 4	etc.
cg00001510_BC21	0.305839	0.434309	0.466366	0.471151	0.406775
cg00101235_BC21	0.772716	0.776497	0.738311	0.659043	0.549244
cg08786079_BC21	0.676144	0.507173	0.774397	0.8125	0.631684
cg12981137_BC21	0.079351	0.107116	0.098179	0.121715	0.095359
cg13699845_TC21	0.298405	0.580583	0.712194	0.576507	0.603451
cg15934644_BC21	0.819637	0.881262	0.872367	0.828646	0.819998
cg16781484_TC21	0.427295	0.870992	0.85284	0.88366	0.890077
cg16819999_TC21	0.724759	0.736795	0.691121	0.627827	0.666185
cg19968202_BC21	0.728354	0.767227	0.793444	0.712416	0.699533
cg23509869_BC21	0.405667	0.503031	0.513358	0.681571	0.663598
cg23729101_BC21	0.856618	0.617512	0.819712	0.80921	0.85501
cg24130641_BC21	0.799122	0.916571	0.847515	0.860805	0.882471

	Sample 1	Sample 2	Sample 2
cg00001510	0.360868	0.305839	0.406859
cg00101235	0.644655	0.772716	0.497198
cg08786079	0.717317	0.676144	0.803968
cg12981137	0.080322	0.079351	0.069465
cg13699845	0.645325	0.298405	0.206029
cg15934644	0.678567	0.819637	0.663951
cg16781484	0.870962	0.427295	0.903339
cg16819999	0.646274	0.724759	0.552081
cg19968202	0.755919	0.728354	0.282533
cg23509869	0.341317	0.405667	0.360958
cg23729101	0.835866	0.856618	0.708742
cg24130641	0.88816	0.799122	0.843137

$$|\beta_{EPIC\ v1} - \beta_{EPIC\ v2}| = \text{Delta Beta}$$

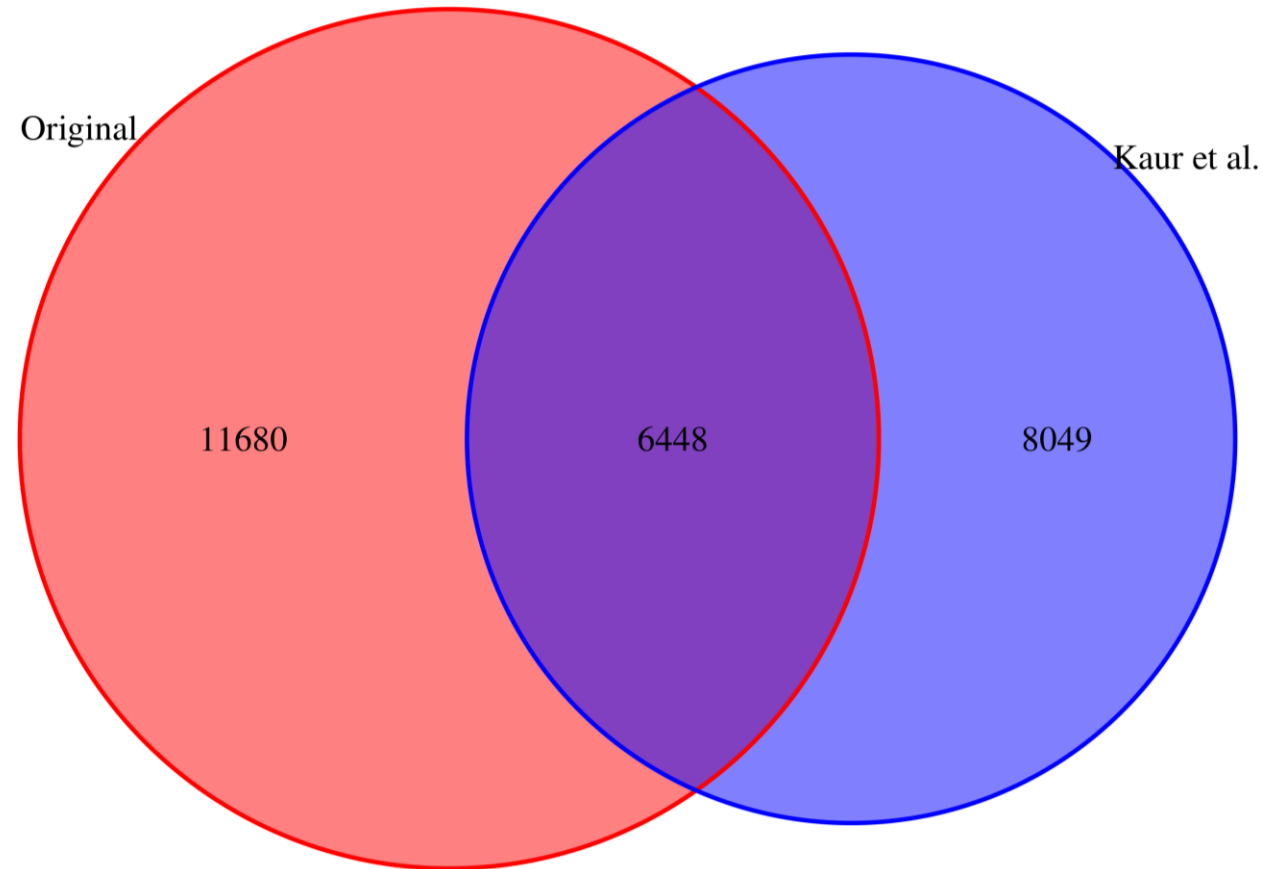
				Median
cg00001510	0.0550	0.0275	0.0059	0.0275
cg00101235	0.1281	0.2793	0.1082	0.1281
cg08786079	0.0412	0.2968	0.1752	0.1752
cg12981137	0.0010	0.0377	0.0013	0.0013
cg13699845	0.3469	0.3746	0.5475	0.3746
cg15934644	0.1411	0.2173	0.0183	0.1411
cg16781484	0.4437	0.0323	0.0414	0.0414
cg16819999	0.0785	0.1847	0.0916	0.0916
cg19968202	0.0276	0.4847	0.2511	0.2511
cg23509869	0.0644	0.1421	0.0188	0.0644
cg23729101	0.0208	0.0912	0.0971	0.0912
cg24130641	0.0890	0.0734	0.0116	0.0734

Cutoff
0.301392881

$$MoM + 2 * SD = Cutoff$$

Median of Medians (MoM)
0.0914
Standard Deviation (SD)
0.1050

Differing Probes by Delta-Beta-Values



Binarization Method

- If a Beta value is above or equal to 0.5 the value gets replaced with 1 (Methylated)
- If its below 0.5 it gets replaced with 0 (Unmethylated)
- This is done for each probe in each Sample and if the status of the probe differs(Methylated/Unmethylated) between EPIC v1 and EPIC v2 one is added to the counter of the probe
- In the end counter reflects how many times the EPIC v1 and EPIC v2 binarized methylation status differs for each probe
- Probes that differed at least two times are compared between the original 16 Samples and the Kaur et al Samples.

Data Analysis Method/Pipeline for Binarization

	Betas EPIC v1				
Probes	Sample 1	Sample 2	Sample 3	Sample 4	etc.
cg00001510	0.36087	0.40686	0.4723	0.20684	0.35183
cg00101235	0.64465	0.4972	0.84655	0.6753	0.8047
cg08786079	0.71732	0.80397	0.59922	0.75506	0.67548
cg12981137	0.08032	0.06946	0.09943	0.06802	0.27195
cg13699845	0.64532	0.20603	0.16474	0.5771	0.70451
cg15934644	0.67857	0.66395	0.89064	0.88136	0.91581
cg16781484	0.87096	0.90334	0.81149	0.86396	0.74584
cg16819999	0.64627	0.55208	0.78272	0.60434	0.75243
cg19968202	0.75592	0.28253	0.54237	0.7126	0.83449
cg23509869	0.34132	0.36096	0.49455	0.22539	0.52606
cg23729101	0.83587	0.70874	0.91682	0.77907	0.672
cg24130641	0.88816	0.84314	0.83587	0.83724	0.83346

	Betas EPIC v2				
Probes	Sample 1	Sample 2	Sample 3	Sample 4	etc.
cg00001510_BC21	0.30584	0.43431	0.46637	0.47115	0.40677
cg00101235_BC21	0.77272	0.7765	0.73831	0.65904	0.54924
cg08786079_BC21	0.67614	0.50717	0.7744	0.8125	0.63168
cg12981137_BC21	0.07935	0.10712	0.09818	0.12172	0.09536
cg13699845_TC21	0.29841	0.58058	0.71219	0.57651	0.60345
cg15934644_BC21	0.81964	0.88126	0.87237	0.82865	0.82
cg16781484_TC21	0.42729	0.87099	0.85284	0.88366	0.89008
cg16819999_TC21	0.72476	0.73679	0.69112	0.62783	0.66619
cg19968202_BC21	0.72835	0.76723	0.79344	0.71242	0.69953
cg23509869_BC21	0.40567	0.50303	0.51336	0.68157	0.6636
cg23729101_BC21	0.85662	0.61751	0.81971	0.80921	0.85501
cg24130641_BC21	0.79912	0.91657	0.84752	0.86081	0.88247

If the Beta-Value is above or equal to 0.5 the entry gets replaced with a 1 if its below 0.5 it gets replaced with a 0

	Betas EPIC v1				
Probes	Sample 1	Sample 2	Sample 3	Sample 4	etc.
cg00001510	0	0	0	0	0
cg00101235	1	0	1	1	1
cg08786079	1	1	1	1	1
cg12981137	0	0	0	0	0
cg13699845	1	0	0	1	1
cg15934644	1	1	1	1	1
cg16781484	1	1	1	1	1
cg16819999	1	1	1	1	1
cg19968202	1	0	1	1	1
cg23509869	0	0	0	0	1
cg23729101	1	1	1	1	1
cg24130641	1	1	1	1	1

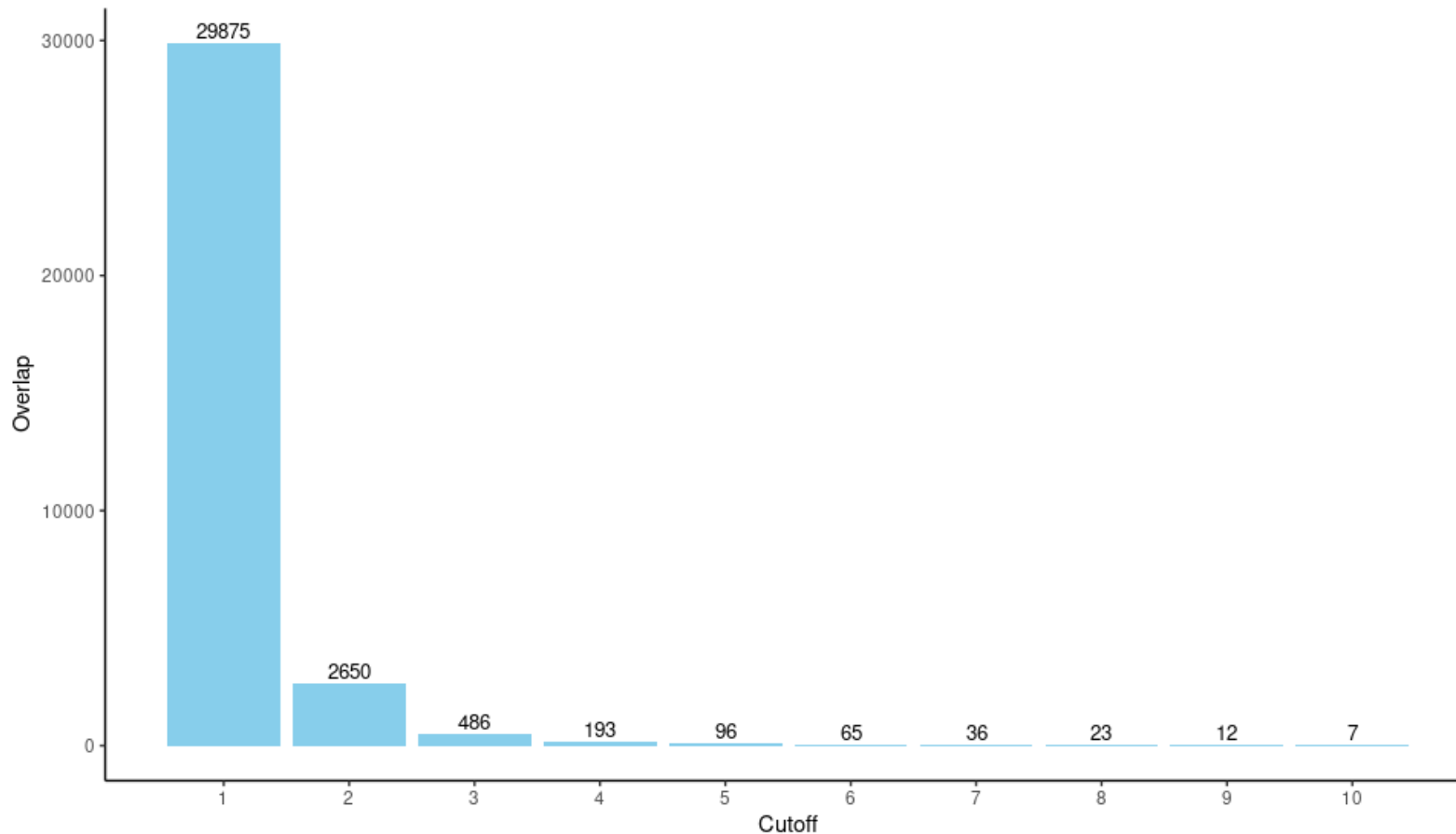
	Betas EPIC v2				
Probes	Sample 1	Sample 2	Sample 3	Sample 4	etc.
cg00001510_BC21	0	0	0	0	0
cg00101235_BC21	1	1	1	1	1
cg08786079_BC21	1	1	1	1	1
cg12981137_BC21	0	0	0	0	0
cg13699845_TC21	0	1	1	1	1
cg15934644_BC21	1	1	1	1	1
cg16781484_TC21	0	1	1	1	1
cg16819999_TC21	1	1	1	1	1
cg19968202_BC21	1	1	1	1	1
cg23509869_BC21	0	1	1	1	1
cg23729101_BC21	1	1	1	1	1
cg24130641_BC21	1	1	1	1	1

If a probe has the same value in EPIC v1 and EPIC v2 they are then replaced with a 0 if they are different they are replaced with a 1

	Betas EPIC v1					
Probes	Sample 1	Sample 2	Sample 3	Sample 4	etc.	Count
cg00001510	0	0	0	0	0	0
cg00101235	0	1	0	0	0	1
cg08786079	0	0	0	0	0	0
cg12981137	0	0	0	0	0	0
cg13699845	1	1	1	0	0	3
cg15934644	0	0	0	0	0	0
cg16781484	1	0	0	0	0	1
cg16819999	0	0	0	0	0	0
cg19968202	0	1	0	0	0	1
cg23509869	0	1	1	1	0	3
cg23729101	0	0	0	0	0	0
cg24130641	0	0	0	0	0	0

For every row/probe the sum/count of the row is calculated. A cutoff of 2 was decided for the following two Venn Diagrams

Bar-plot Overlap of differing probes for every cutoff



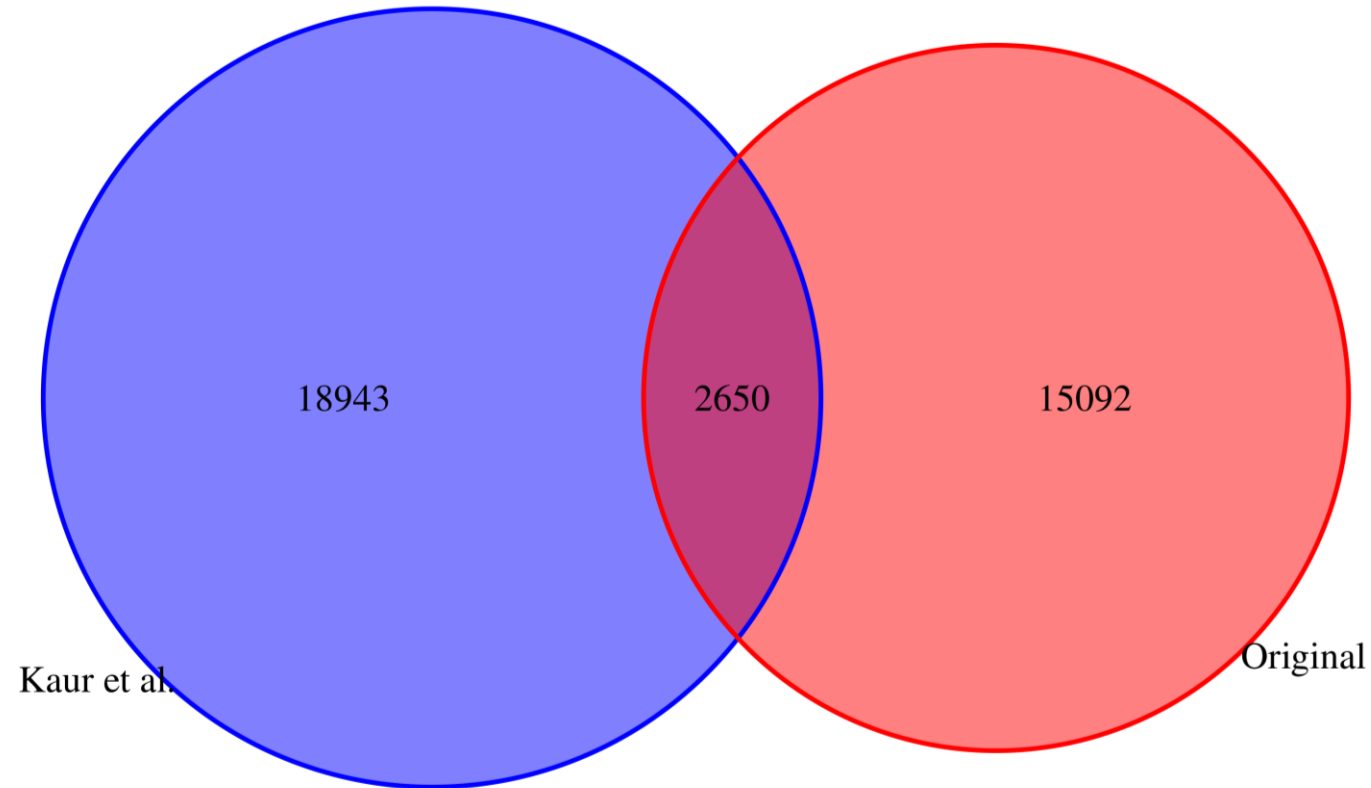
-The X axis is the Cutoff and the Y Axis is the number of CpG's/Probes that after binarization differed at least the amount of times as the cutoff in **both** the Original data and the Kaur et al data

-So for cutoff 1 the CpG differed in at least one sample (in both the Original 16 Samples and the 11 Kaur et al Samples)

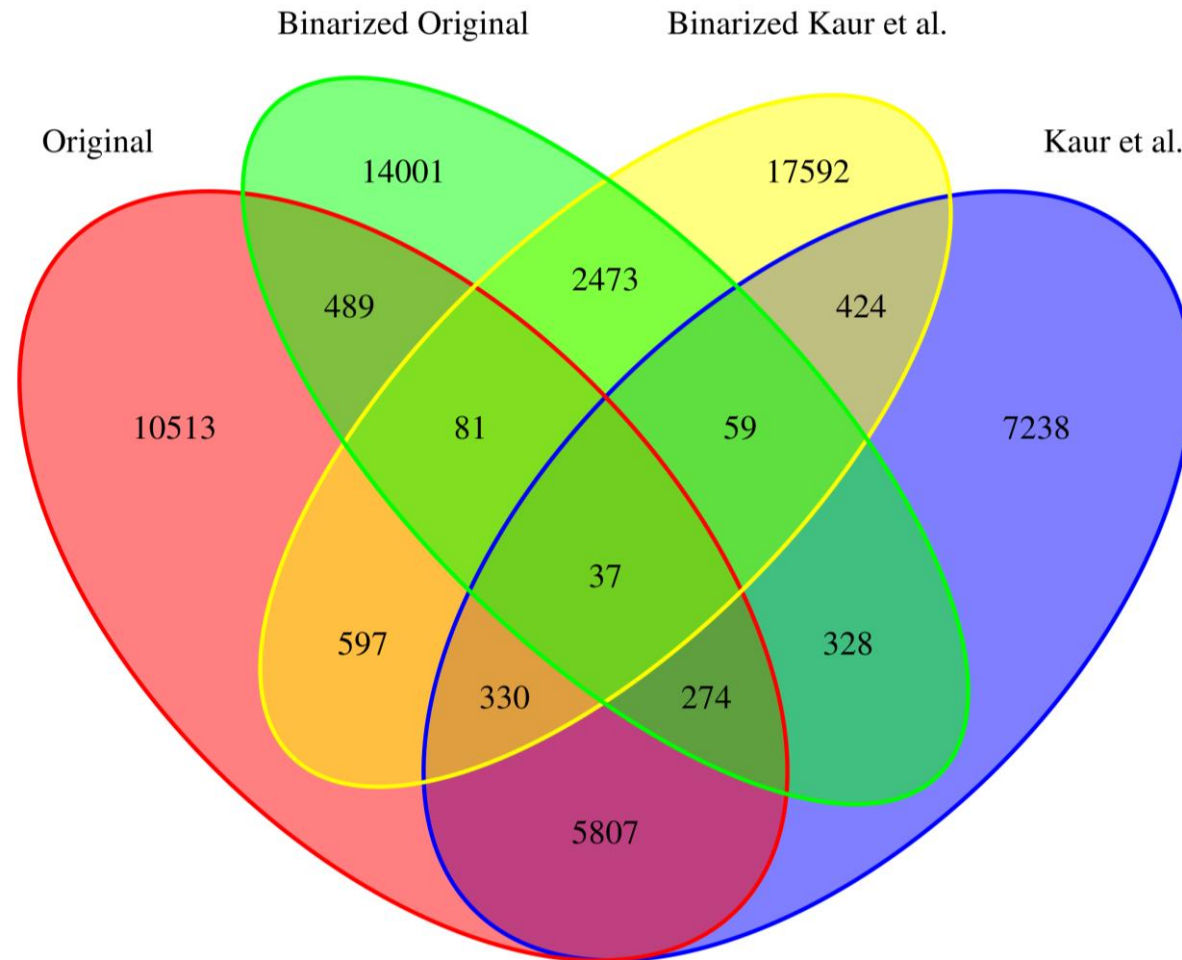
-At cutoff 2 in at least 2 Samples etc. (in both the Original 16 Samples and the 11 Kaur et al Samples)

-10 is the maximum because in the Kaur et al. Data-set none differed in more than 10 Samples (after binarization).

Differing Probes by binarized Beta-Values



Overlap of probes identified using the delta beta and binarization method



Trimerization Method

- If a Beta value is above or equal to 0.666 the value gets replaced with 1 (Methylated)
- If its below 0.666 and above or equal to 0.333 it gets replaced with 0.5 (Mixed)
- And if its below 0.333 it gets replaced with 0 (Unmethylated)
- This is done for each probe in each Sample and if the status of the probe differs(Methylated/Unmethylated) between EPIC v1 and EPIC v2 one is added to the counter of the probe
- In the end the counter reflects how many times the EPIC v1 and EPIC v2 trimerized methylation status differs for each probe
- Probes that differed at least two times are compared between the original 16 Samples and the Kaur et al Samples.

Data Analysis Method/Pipeline for Trimerization

Probes	Betas EPIC v1				
	Sample 1	Sample 2	Sample 3	Sample 4	etc.
cg00001510	0.360868	0.406859	0.472296	0.206835	0.351828
cg00101235	0.644655	0.497198	0.846551	0.675296	0.804703
cg08786079	0.717317	0.803968	0.59922	0.755058	0.675481
cg12981137	0.080322	0.069465	0.099431	0.068025	0.271951
cg13699845	0.645325	0.206029	0.164735	0.577102	0.704511
cg15934644	0.678567	0.663951	0.890643	0.881358	0.915806
cg16781484	0.870962	0.903339	0.811486	0.863965	0.745836
cg16819999	0.646274	0.552081	0.782722	0.604336	0.752426
cg19968202	0.755919	0.282533	0.542365	0.712598	0.834492
cg23509869	0.341317	0.360958	0.494545	0.22539	0.526057
cg23729101	0.835866	0.708742	0.916817	0.779065	0.671998
cg24130641	0.88816	0.843137	0.835868	0.837237	0.833462

Probes	Betas EPIC v2				
	Sample 1	Sample 2	Sample 3	Sample 4	etc.
cg00001510_BC21	0.305839	0.434309	0.466366	0.471151	0.406775
cg00101235_BC21	0.772716	0.776497	0.738311	0.659043	0.549244
cg08786079_BC21	0.676144	0.507173	0.774397	0.8125	0.631684
cg12981137_BC21	0.079351	0.107116	0.098179	0.121715	0.095359
cg13699845_TC21	0.298405	0.580583	0.712194	0.576507	0.603451
cg15934644_BC21	0.819637	0.881262	0.872367	0.828646	0.819998
cg16781484_TC21	0.427295	0.870992	0.85284	0.88366	0.890077
cg16819999_TC21	0.724759	0.736795	0.691121	0.627827	0.666185
cg19968202_BC21	0.728354	0.767227	0.793444	0.712416	0.699533
cg23509869_BC21	0.405667	0.503031	0.513358	0.681571	0.663598
cg23729101_BC21	0.856618	0.617512	0.819712	0.80921	0.85501
cg24130641_BC21	0.799122	0.916571	0.847515	0.860805	0.882471

If the Beta-Value is above or equal to 0.66 the entry gets replaced with a 1 if its below 0.666 and above or equal to 0.333 it gets replaced with a 0.5 and if its below 0.333 it gets replaced with 0

Probes	Betas EPIC v1				
	Sample 1	Sample 2	Sample 3	Sample 4	etc.
cg00001510	0.5	0.5	0.5	0	0.5
cg00101235	0.5	0.5	1	1	1
cg08786079	1	1	0.5	1	1
cg12981137	0	0	0	0	0
cg13699845	0.5	0	0	0.5	1
cg15934644	1	0.5	1	1	1
cg16781484	1	1	1	1	1
cg16819999	0.5	0.5	1	0.5	1
cg19968202	1	0	0.5	1	1
cg23509869	0.5	0.5	0.5	0	0.5
cg23729101	1	1	1	1	1
cg24130641	1	1	1	1	1

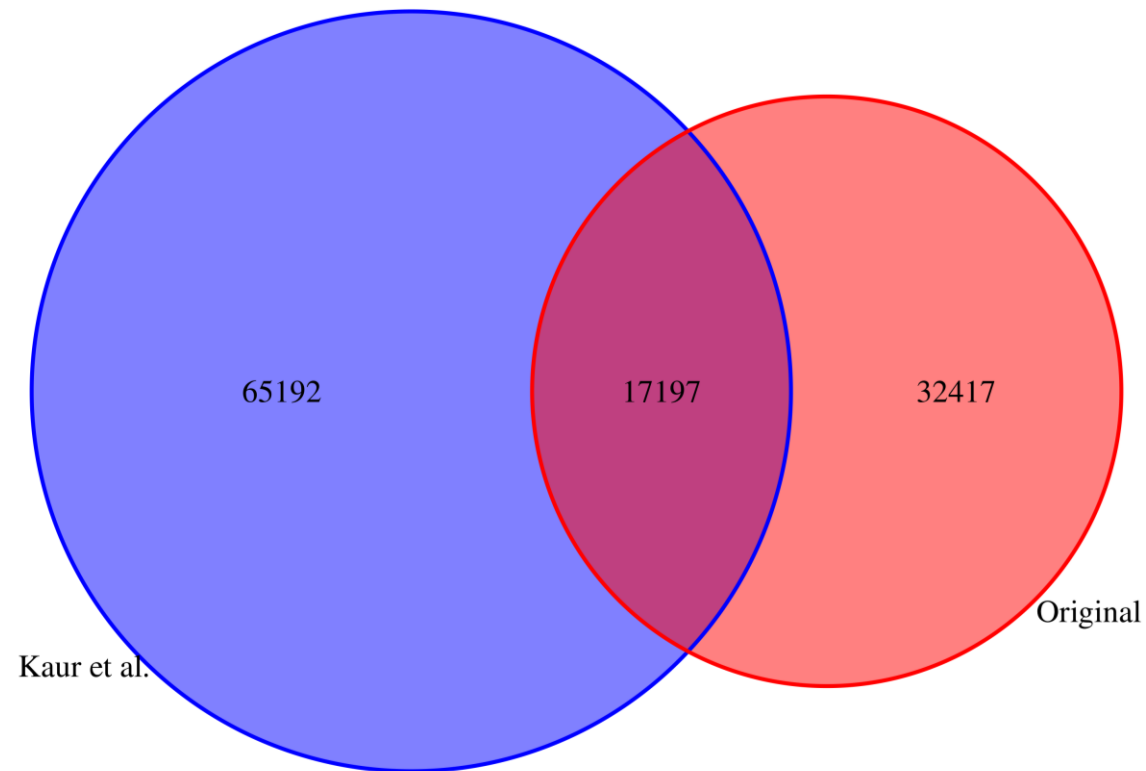
Probes	Betas EPIC v2				
	Sample 1	Sample 2	Sample 3	Sample 4	etc.
cg00001510_BC21	0	0.5	0.5	0.5	0.5
cg00101235_BC21	1	1	1	0.5	0.5
cg08786079_BC21	1	0.5	1	1	0.5
cg12981137_BC21	0	0	0	0	0
cg13699845_TC21	0	0.5	1	0.5	0.5
cg15934644_BC21	1	1	1	1	1
cg16781484_TC21	0.5	1	1	1	1
cg16819999_TC21	1	1	1	0.5	1
cg19968202_BC21	1	1	1	1	1
cg23509869_BC21	0.5	0.5	0.5	1	0.5
cg23729101_BC21	1	0.5	1	1	1
cg24130641_BC21	1	1	1	1	1

If a probe has the same value in EPIC v1 and EPIC v2 they are then replaced with a 0 if they are different they are replaced with a 1

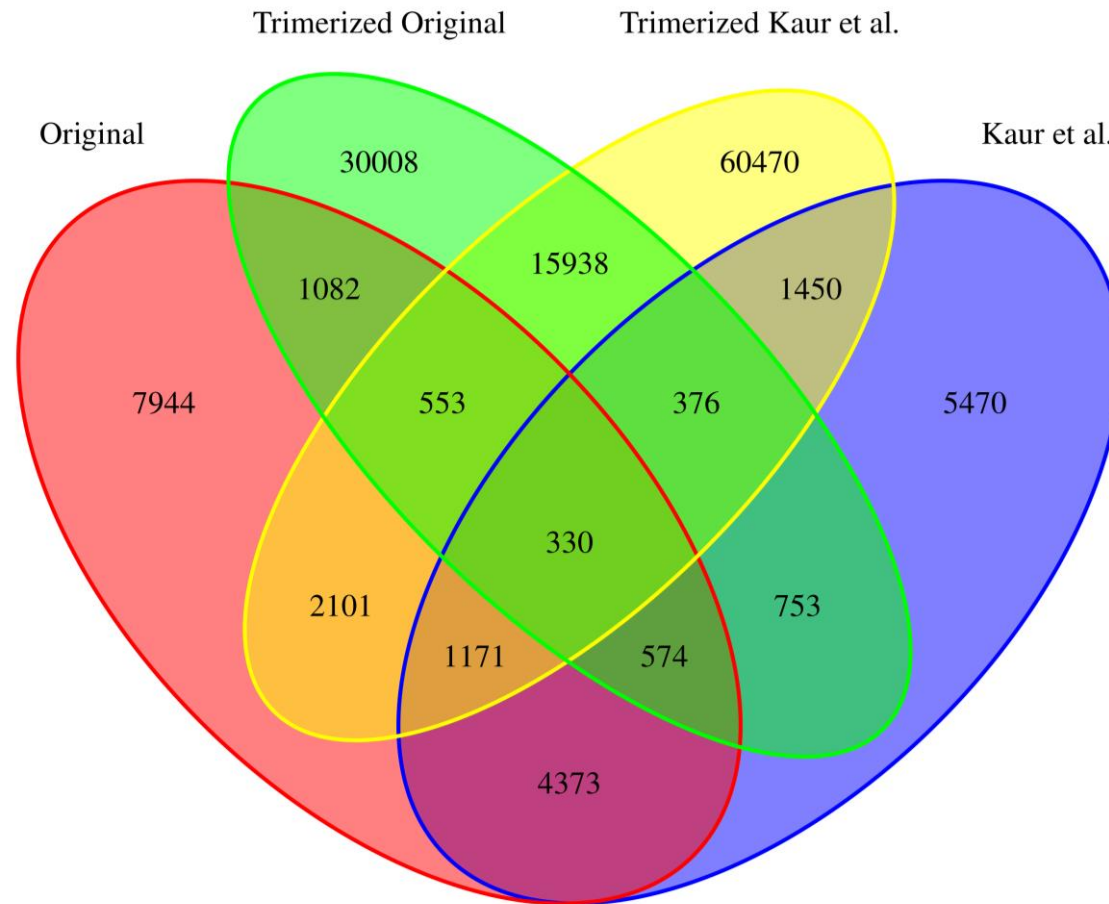
Probes	Betas EPIC v1					Count
	Sample 1	Sample 2	Sample 3	Sample 4	etc.	
cg000015	1	0	0	1	0	2
cg001012	1	1	0	1	1	4
cg087860	0	1	1	0	1	3
cg129811	0	0	0	0	0	0
cg136998	1	1	1	0	1	4
cg159346	0	1	0	0	0	1
cg167814	1	0	0	0	0	1
cg168199	1	1	0	0	0	2
cg199682	0	1	1	0	0	2
cg235098	0	0	0	1	0	1
cg237291	0	1	0	0	0	1
cg241306	0	0	0	0	0	0

For every row/probe the sum/count of the row is calculated. A cutoff of 2 was decided for the following two Venn Diagrams

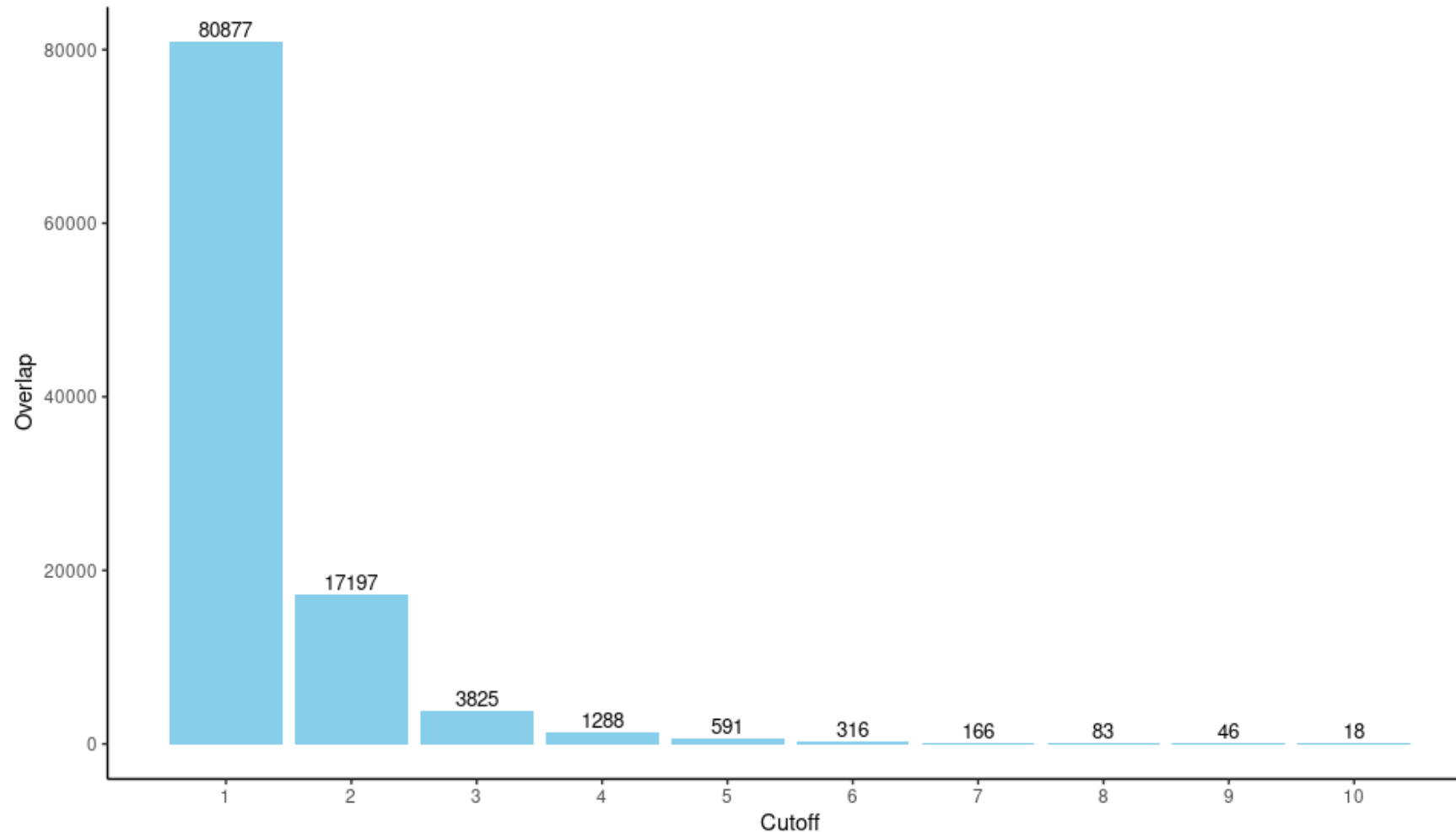
Differing Probes by trimerized Beta-Values



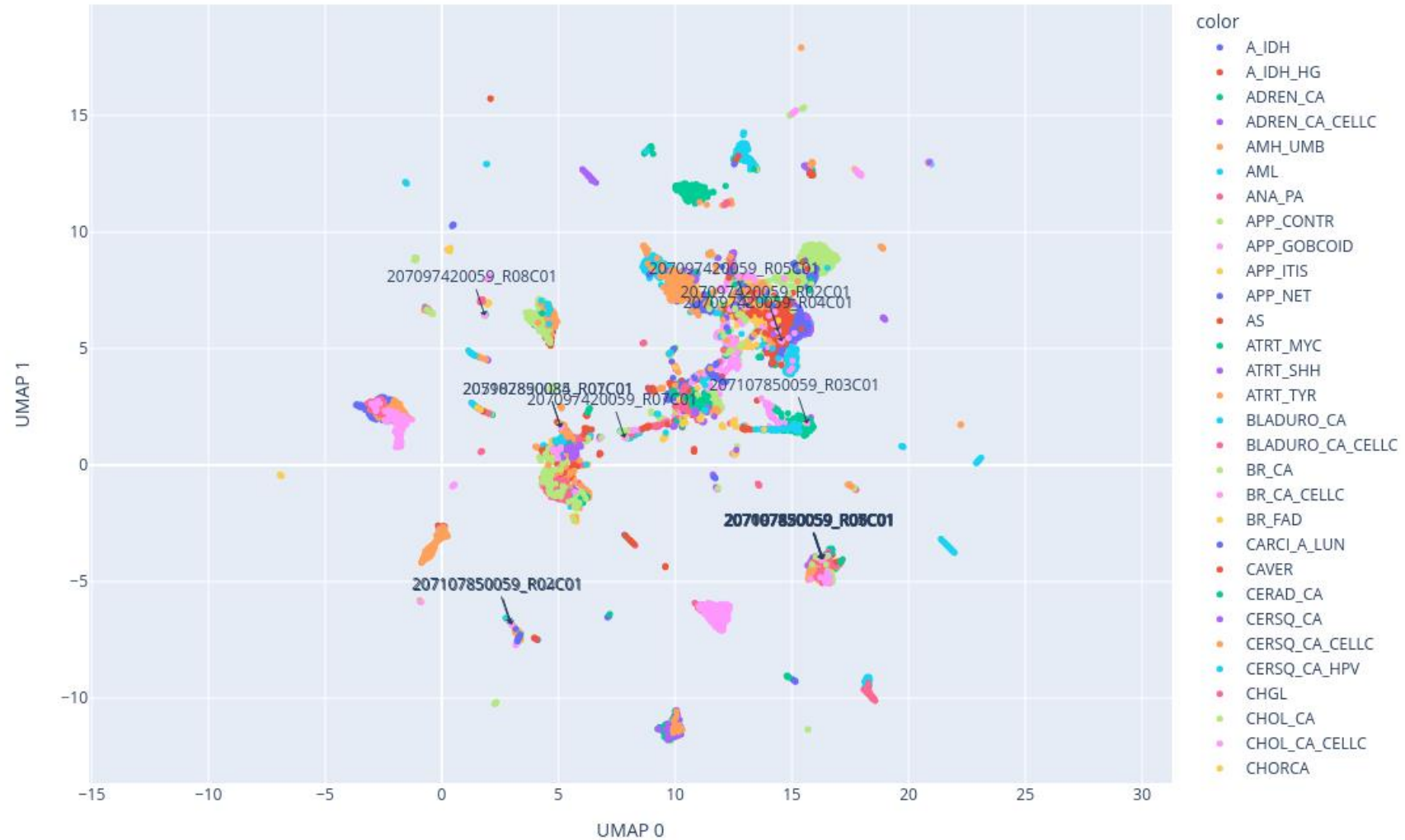
Overlap of probes identified using the delta beta and trimerization method



Bar-plot Overlap of differing probes for every cutoff after trimerization



Testing of Blacklists



Method for Blacklist Ranking

- All reference Samples + all 32 original samples were plotted on the Laptop provided by Jürgen with 4 different Blacklist 10 times each (40 Plots in total)
 - 32 original samples consist of 16 EPIC v1 and 16 EPIC v2 samples(the same DNA)
 - The 4 Blacklists are :
 - Original Blacklist (already included on the Laptop)
 - Delta-Beta Blacklist
 - Binarization Blacklist
 - Trimerization Blacklist
- Each of these Samples was classified as one MethClasses from the reference sample data-set using one of 4 Methods:
 - Nearest Sample
 - Nearest Cluster
 - Nearest Neighbors
 - Minimum Max Distance

Method for Blacklist Ranking

- A score for each individual UMAP was generated
- The score was generated by comparing the EPIC v1 and the EPIC v2 MethClass for each of the 16 Original Samples. If EPIC v1 and EPIC v2 had the same Methclass: 1 was added to the counter of the UMAP
 - The max score was 16 if all EPIC v1 and EPIC v2 classifications/MethClasses were the same for each sample
- The Median of all 10 UMAP scores for each Blacklist was calculated
- This Median was used to rank the Blacklists

Process for Every UMAP-Plot

(Repeated 10 Times for every Blacklist)

Run UMAP to get Coordinates

	UMAP 1			
	Sentrix ID	MethClass	X -Coordinates	Y-Coordinates
Reference samples	203049640041_R08C01	CTRL_BLOOD	15.250681	6.395354
	200928190151_R03C01	AFH	10.070036	2.317001
	201465950038_R03C01	EHE	9.717898	1.987853
	201465950038_R06C01	EHE	9.542107	1.758346
	201414140114_R08C01	EHE	9.688657	1.873135
	9533774023_R03C02	RMS_MYOD1	12.688134	7.406944
	200325530093_R04C02	RMS_MYOD1	12.571888	7.472862
	201869680016_R05C01	RMS_MYOD1	9.725392	1.987692
	202273260014_R03C01	RMS_MYOD1	9.815007	1.866416
	202292310098_R05C01	ESS_HG	5.783106	-3.241021
	202250800134_R08C01	NFA	9.427515	1.461566
	200848860106_R02C01	NFA	9.428597	1.497979
	200848860106_R07C01	NFA	9.402017	1.494115
Unclassified Samples	205705530027_R06C01	NaN	22.144543	4.130687
	205707890147_R04C01	NaN	12.844316	8.475537
	205751550059_R06C01	NaN	12.082621	6.15128
	205756360010_R01C01	NaN	12.724951	7.812495
	205800600144_R01C01	NaN	0.784669	7.90254
	205800600144_R03C01	NaN	22.174519	3.799611
	205800600144_R04C01	NaN	5.560182	9.801536
	205800610010_R01C01	NaN	22.299681	3.952282
	205800610132_R07C01	NaN	7.519009	-0.090841
	205800610140_R04C01	NaN	4.058029	0.271204
	205800610140_R07C01	NaN	0.822111	7.863798
	205800610140_R08C01	NaN	22.162395	4.111175
	205809360101_R05C01	NaN	22.312954	4.318967
	205921770008_R02C01	NaN	21.891876	3.784404
	205982890034_R02C01	NaN	22.356445	4.414614
	205982890034_R07C01	NaN	5.814699	0.991733

Nearest Neighbor

(or other
classification
method)

	UMAP 1			
	Sentrix ID	MethClass	X -Coordinates	Y-Coordinates
Reference samples	203049640041_R08C01	CTRL_BLOOD	15.250681	6.395354
	200928190151_R03C01	AFH	10.070036	2.317001
	201465950038_R03C01	EHE	9.717898	1.987853
	201465950038_R06C01	EHE	9.542107	1.758346
	201414140114_R08C01	EHE	9.688657	1.873135
	9533774023_R03C02	RMS_MYOD1	12.688134	7.406944
	200325530093_R04C02	RMS_MYOD1	12.571888	7.472862
	201869680016_R05C01	RMS_MYOD1	9.725392	1.987692
	202273260014_R03C01	RMS_MYOD1	9.815007	1.866416
	202292310098_R05C01	ESS_HG	5.783106	-3.241021
	202250800134_R08C01	NFA	9.427515	1.461566
	200848860106_R02C01	NFA	9.428597	1.497979
	200848860106_R07C01	NFA	9.402017	1.494115
Unclassified Samples	205705530027_R06C01	MNG_INT-A	22.144543	4.130687
	205707890147_R04C01	HNSQ_CA	12.844316	8.475537
	205751550059_R06C01	NSCLC_AD	12.082621	6.15128
	205756360010_R01C01	NSCLC_SC	12.724951	7.812495
	205800600144_R01C01	PITAD_FSH_LH	0.784669	7.90254
	205800600144_R03C01	MNG_BEN-2	22.174519	3.799611
	205800600144_R04C01	ENDOM_CA	5.560182	9.801536
	205800610010_R01C01	MNG_BEN-2	22.299681	3.952282
	205800610132_R07C01	SCHW	7.519009	-0.090841
	205800610140_R04C01	CN	4.058029	0.271204
	205800610140_R07C01	PITAD_FSH_LH	0.822111	7.863798
	205800610140_R08C01	MNG_INT-A	22.162395	4.111175
	205809360101_R05C01	MNG_INT-A	22.312954	4.318967
	205921770008_R02C01	MNG_BEN-2	21.891876	3.784404
	205982890034_R02C01	MNG	22.356445	4.414614
	205982890034_R07C01	LGG_PA_PF	5.814699	0.991733

Comparison of EPIC
v1 and EPIC v2

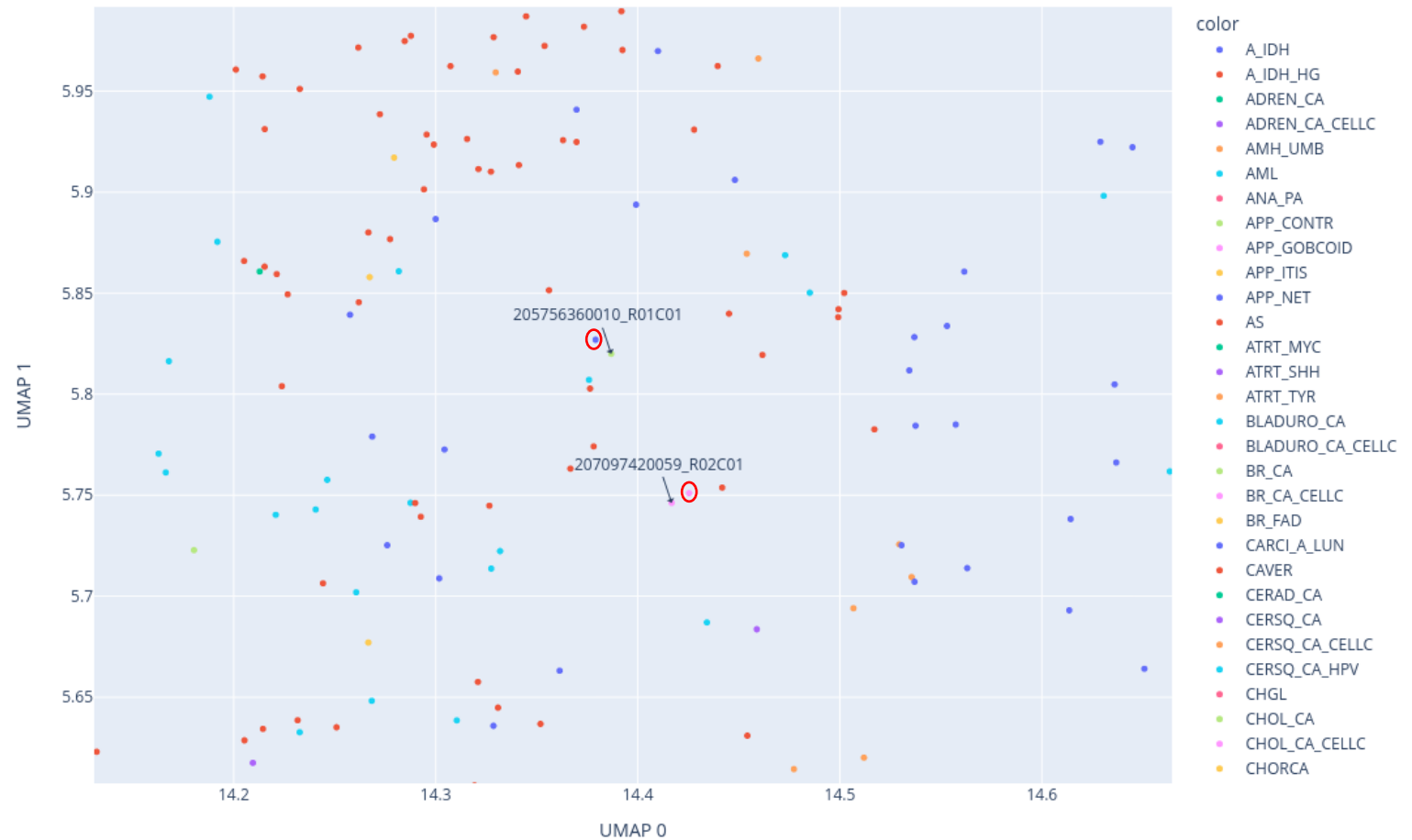
	EPIC v1	EPIC v2	Comaprison
Sample 1	MNG_INT-A	MNG_INT-A	1
Sample 2	HNSQ_CA	HNSQ_CA	1
Sample 3	NSCLC_AD	NSCLC_SC	0
Sample 4	NSCLC_SC	NSCLC_AD	0
Sample 5	PITAD_FSH_LH	PITAD_FSH_LH	1
Sample 6	MNG_BEN-2	MNG_INT-A	0
Sample 7	ENDOM_CA	ENDOM_CA	0
Sample 8	MNG_BEN-2	MNG_BEN-2	1
Sample 9	SCHW	SCHW	1
Sample 10	CN	CN	1
Sample 11	PITAD_FSH_LH	PITAD_FSH_LH	1
Sample 12	MNG_INT-A	MNG	0
Sample 13	MNG_INT-A	MNG_INT-A	1
Sample 14	MNG_BEN-2	MNG_BEN-2	1
Sample 15	MNG	MNG_INT-A	0
Sample 16	LGG_PA_PF	LGG_PA_PF	1

10

Score

Nearest Sample Method

- For each of the 16 v1 and 16 v2 Samples the closest Reference-Sample was determined using the Euclidian distance.
- The MethClass of this closest Reference-Sample was then assigned to the original Sample

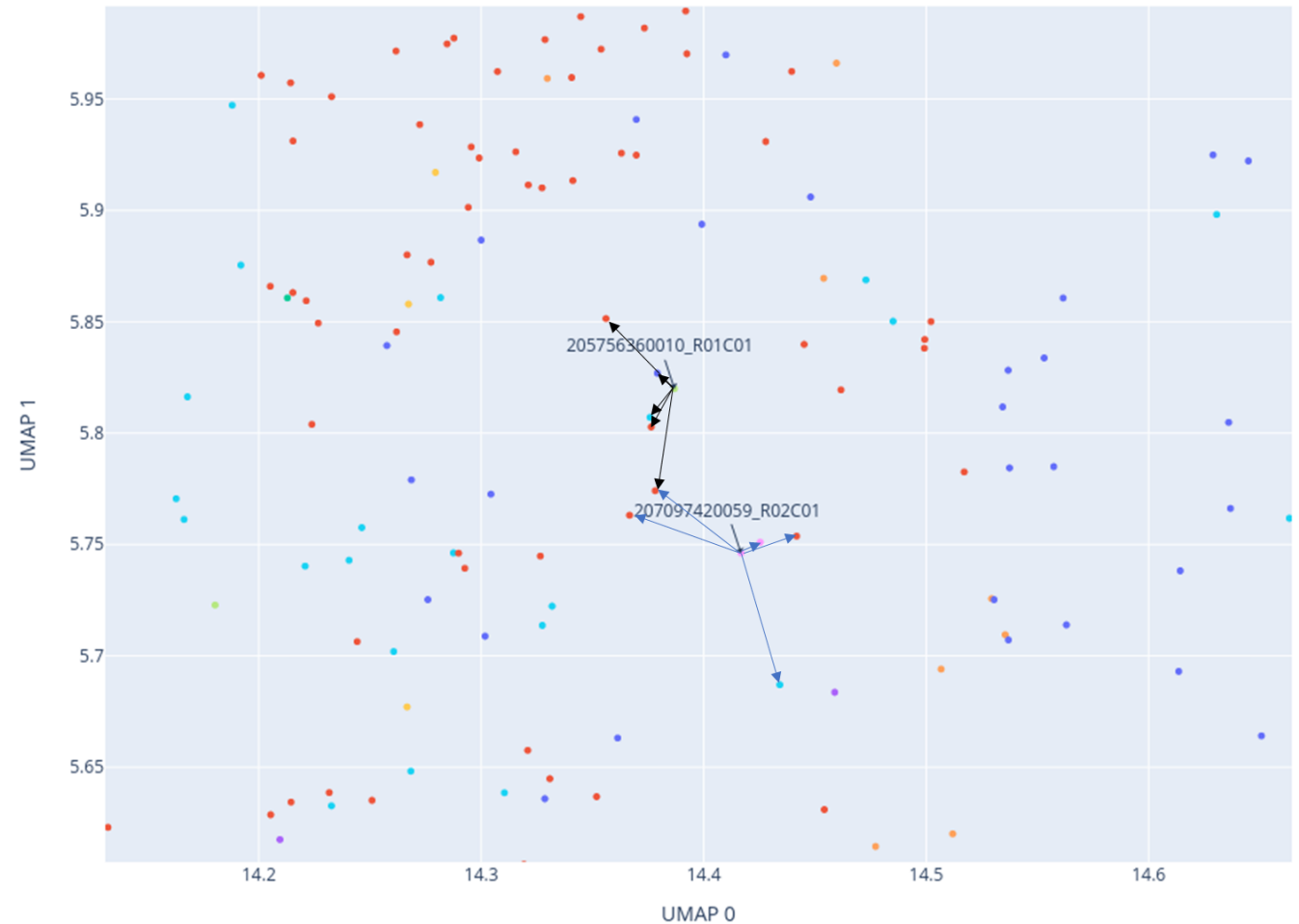


Nearest Cluster Method

- For all MethClasses the Center of all the reference Sample(Centroid) is determined (The mean of all coordinates from all reference samples)
- For each of the 16 v1 and 16 v2 Samples the closest centroid was determined using the Euclidian distance.
- The MethClass of the Centroid was set as the new MethClass of the original sample.

Nearest Neighbors Method

- For each of the 16 v1 and 16 v2 Samples the 5 closest reference Samples were determined using the Euclidian distance.
- The MethClass that Occurs the most in the 5 closest reference Samples is then set as the MethClass of the Sample
 - If two MethClasses occur the same amount the one closer to the original Sample is set as the new MethClass

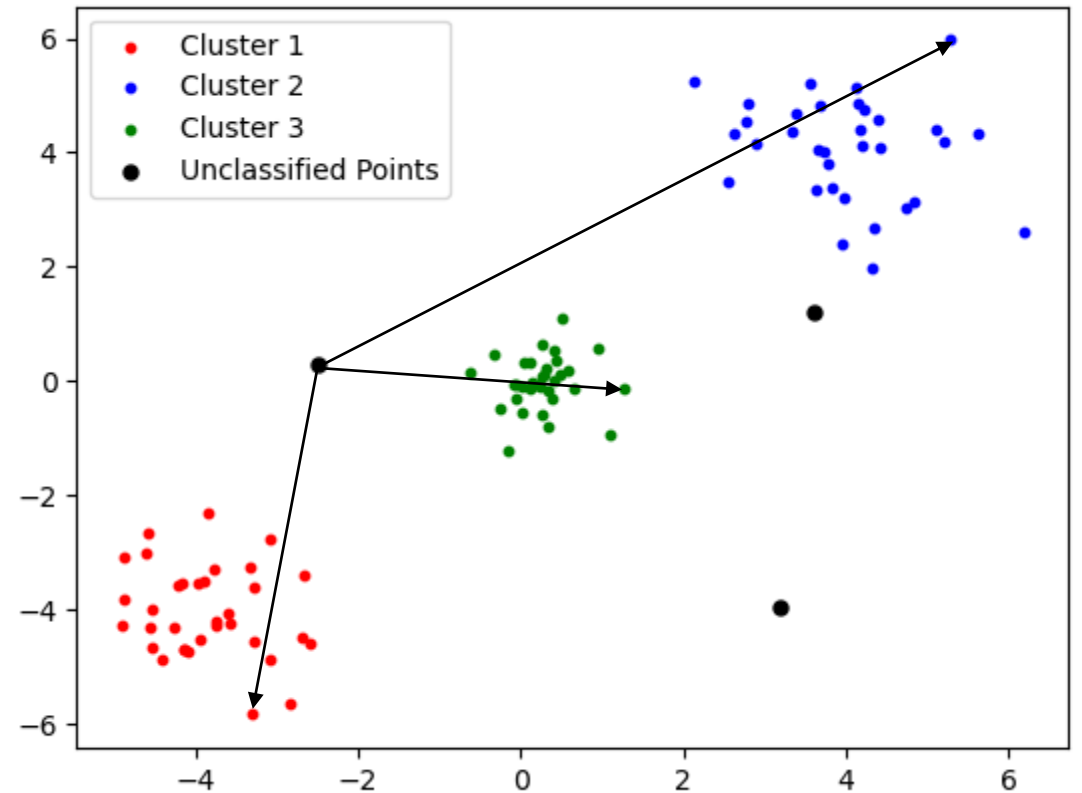


Min Max Distance

- For every cluster identify the point furthest away from the unidentified point(Euclidian Distance).

-From the list of furthest away points, the closest one is picked.

The Class from the picked point is assigned to the unclassified point.



Min Max Distance

-Calculate Euclid.
Distance of the
Unclassified point to
all other points.

Euclidian Distance		
Cluster 1	Cluster 2	Cluster 3
5.500222	6.18516	6.83762
4.099966	7.027124	2.500917
5.2704	7.257649	2.875911
5.049792	8.189346	3.786803
4.6604	8.97299	2.942074
4.899248	7.665886	2.644498
4.052056	6.764862	3.065657
4.662088	6.506137	2.633319
5.289847	7.243065	2.537594
4.745413	8.614204	2.600498
3.699969	7.865066	2.758617
5.947023	7.732657	2.896846
4.137851	7.095092	1.852154
3.133099	7.107012	2.757107
4.764113	7.877985	2.757241
4.900794	6.989413	2.522714
2.943475	7.84075	3.109073
5.258081	8.121872	2.725553
3.910841	8.067523	2.919443
3.805232	7.196975	3.021574
5.2045	6.773024	2.360353
4.174018	7.178317	2.982449
4.757008	8.649471	2.785992
4.783124	9.613612	3.770584
4.492519	7.79754	3.179002
4.234317	6.640645	2.898611

- And for every
Cluster find the
point furthest away
from the
Unclassified point

Euclidian Distance		
Cluster 1	Cluster 2	Cluster 3
5.500222	6.18516	6.83762
4.099966	7.027124	2.500917
5.2704	7.257649	2.875911
5.049792	8.189346	3.786803
4.6604	8.97299	2.942074
4.899248	7.665886	2.644498
4.052056	6.764862	3.065657
4.662088	6.506137	2.633319
5.289847	7.243065	2.537594
4.745413	8.614204	2.600498
3.699969	7.865066	2.758617
5.94702	7.732657	2.896846
4.137851	7.095092	1.852154
3.133099	7.107012	2.757107
4.764113	7.877985	2.757241
4.900794	6.989413	2.522714
2.943475	7.84075	3.109073
5.258081	8.121872	2.725553
3.910841	8.067523	2.919443
3.805232	7.196975	3.021574
5.2045	6.773024	2.360353
4.174018	7.178317	2.982449
4.757008	8.649471	2.785992
4.783124	9.61361	3.770584
4.492519	7.79754	3.179002
4.234317	6.640645	2.898611

- Find the smallest
of the max distance
points

Cluster 1	5.947023
Cluster 2	9.613612
Cluster 3	6.83762

-Classify
unclassified point

Cluster 1
(2.48372122, 0.2888025)

Unclassified point
(2.48372122, 0.2888025)

Results for Nearest Sample

Original-Blacklist	
UMAP	Score
1	8
2	8
3	12
4	8
5	9
6	12
7	9
8	9
9	8
10	6
Median	8.5

Delta-Beta-Blacklist	
UMAP	Score
1	12
2	8
3	9
4	10
5	11
6	9
7	12
8	7
9	9
10	11
Median	9.5

Binarization-Blacklist	
UMAP	Score
1	13
2	6
3	9
4	10
5	12
6	8
7	10
8	8
9	11
10	8
Median	9.5

Trimerization-Blacklist	
UMAP	Score
1	5
2	9
3	12
4	9
5	10
6	8
7	11
8	9
9	7
10	13
Median	9.0

Result for Nearest Cluster

Original-Blacklist	
UMAP	Score
1	10
2	13
3	11
4	14
5	13
6	14
7	13
8	9
9	14
10	10
Median	13

Delta-Beta-Blacklist	
UMAP	Score
1	13
2	12
3	10
4	11
5	13
6	12
7	14
8	7
9	14
10	14
Median	12.5

Binarization-Blacklist	
UMAP	Score
1	12
2	12
3	10
4	14
5	11
6	15
7	15
8	12
9	14
10	9
Median	12.0

Trimerization-Blacklist	
UMAP	Score
1	11
2	15
3	8
4	11
5	12
6	12
7	15
8	11
9	14
10	13
Median	12.0

Results for Nearest Neighbor (N=15)

Original-Blacklist	
UMAP	Score
1	13
2	14
3	11
4	11
5	11
6	13
7	12
8	12
9	12
10	13
Median	12

Delta-Beta-Blacklist	
UMAP	Score
1	12
2	13
3	13
4	13
5	12
6	12
7	13
8	11
9	11
10	13
Median	12.5

Binarization-Blacklist	
UMAP	Score
1	11
2	12
3	14
4	11
5	12
6	15
7	14
8	12
9	12
10	12
Median	12

Trimerization-Blacklist	
UMAP	Score
1	10
2	13
3	10
4	12
5	11
6	14
7	13
8	11
9	12
10	11
Median	11.5

Results for Min Max Distance

Original-Blacklist	
UMAP	Score
1	14
2	15
3	15
4	14
5	16
6	14
7	13
8	15
9	15
10	14
Median	14.5

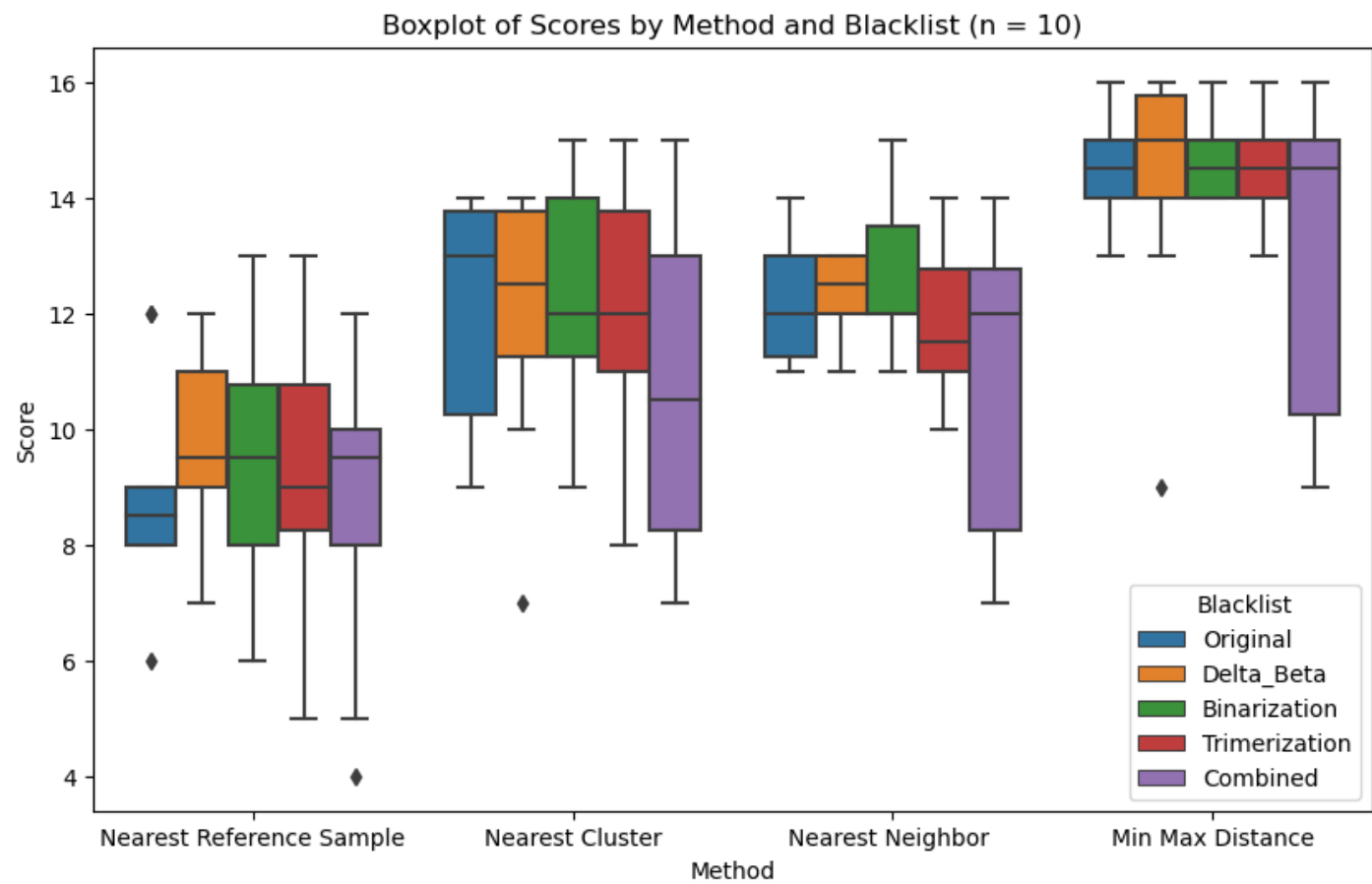
Delta-Beta-Blacklist	
UMAP	Score
1	15
2	15
3	15
4	13
5	16
6	16
7	14
8	9
9	14
10	16
Median	15

Binarization-Blacklist	
UMAP	Score
1	15
2	14
3	15
4	16
5	14
6	14
7	15
8	16
9	14
10	14
Median	14.5

Trimerization-Blacklist	
UMAP	Score
1	14
2	16
3	15
4	14
5	16
6	14
7	14
8	13
9	15
10	15
Median	14.5

Results (Median)

Method	Original	Delat_Beta	Binarization	Trimerization	Combined
Nearest Sample	8.5	9.5	9.5	9.0	9.5
Nearest Cluster	13.0	12.5	12.0	12.0	10.5
Nearest Neighbor	12.0	12.5	12.0	11.5	12.0
Min Max Distance	14.5	15.0	14.5	14.5	14.5



HM450K – EPIC v1

Delta-Beta Method

3 Papers I found that also used the delta-beta method

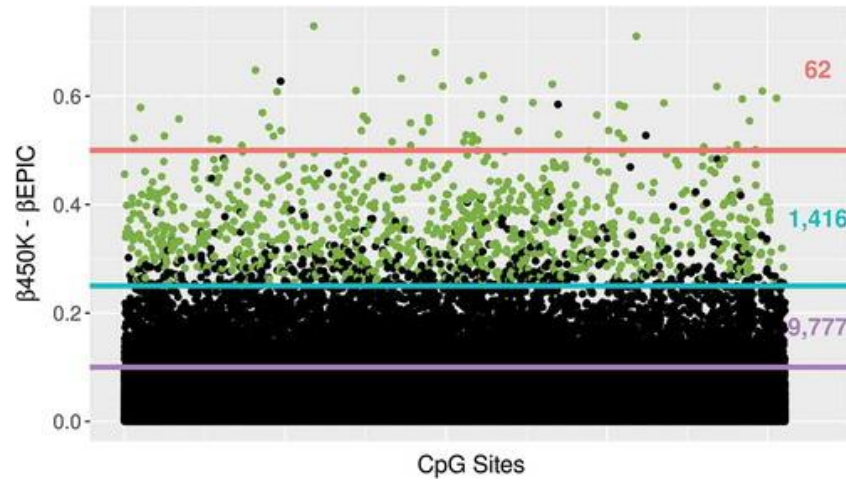
(to find differing probes)

- Comparison of DNA methylation measured by Illumina 450K and EPIC BeadChips in blood of newborns and 14-year-old children
 - (Solomon et al., 2018)
- Comparison of Illumina 450K and EPIC arrays in placental DNA methylation
 - (Fernandez-Jimenez et al., 2019)
- Correlation of Infinium HumanMethylation450K and MethylationEPIC BeadChip arrays in cartilage
 - (Cheung et al., 2020)

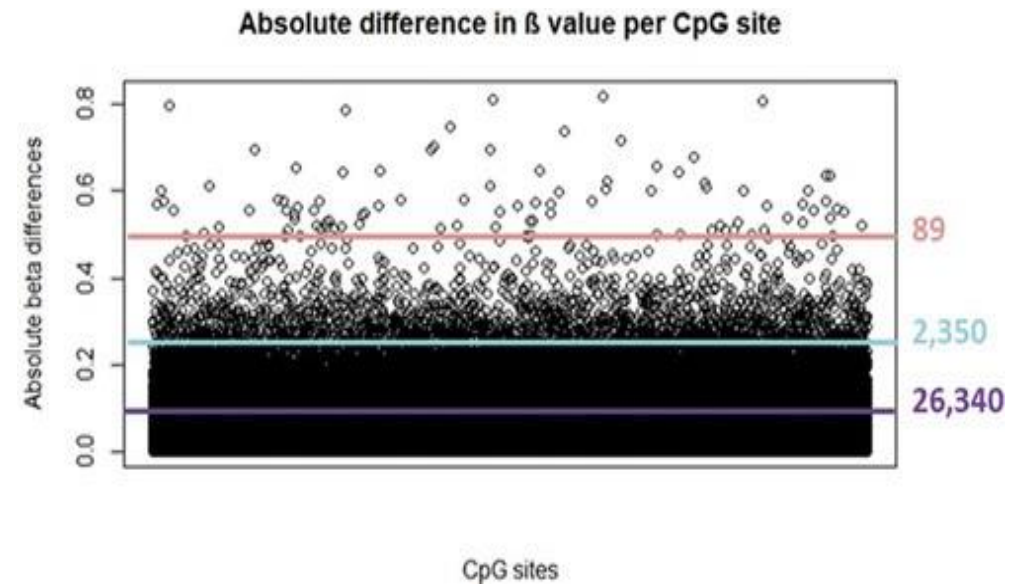
Amount of probes

- From all Papers 32'379 Unique Probes were flagged and published in the supplemental material
 - In Solomon et al. only probes with a delta beta value(DBV) >0.5 were published
 - Fernandez-Jimenez et al. published all probes with a DBV > 0.1
 - Cheung et al. published all probes with a beta difference $>10\%$
- 26'473 of those are in the UMAP Index
 - 3562 of those are already Blacklisted in the original Blacklist
 - 4403 are in the Blacklist I made
- 18'508 are not Blacklisted

Visualization

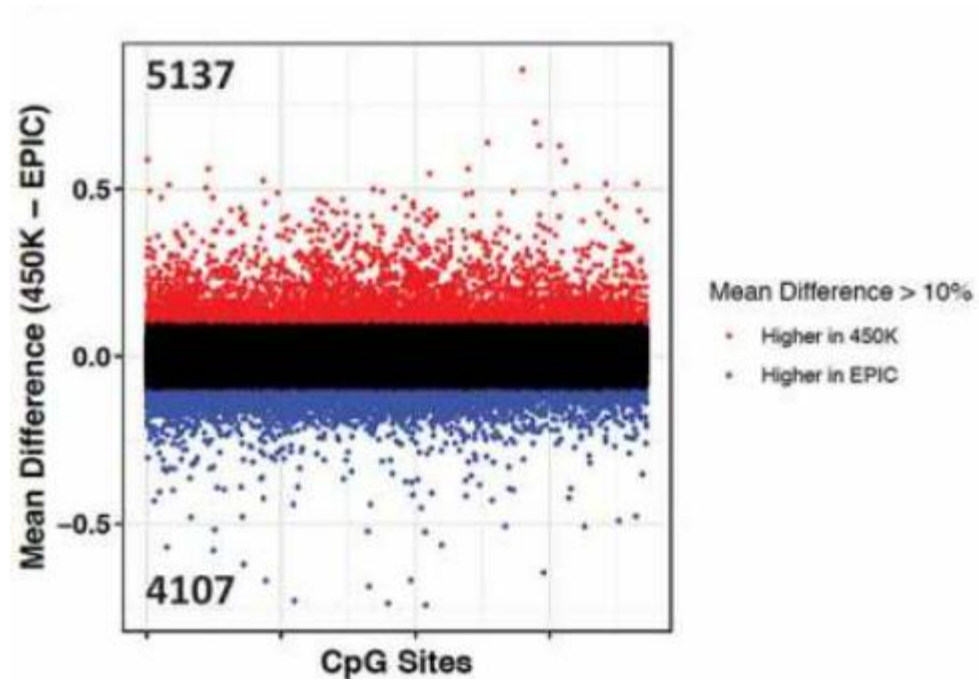


(Solomon et al., 2018)

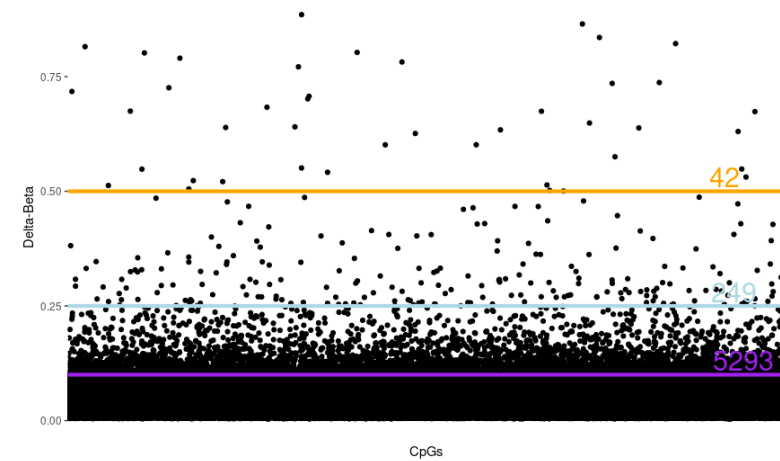


(Fernandez-Jimenez et al., 2019)

Visualization



(Cheung et al., 2020)



My data (original 16 Samples)

Note

- All three papers say that the probes are not necessarily low variance probes.

β -EPICv1 – β -EPICv2

- For each CpG the Beta value of EPICv2 was subtracted from the Beta value of EPIC v1
- The Average of these Values was plotted on the graph on the right

