

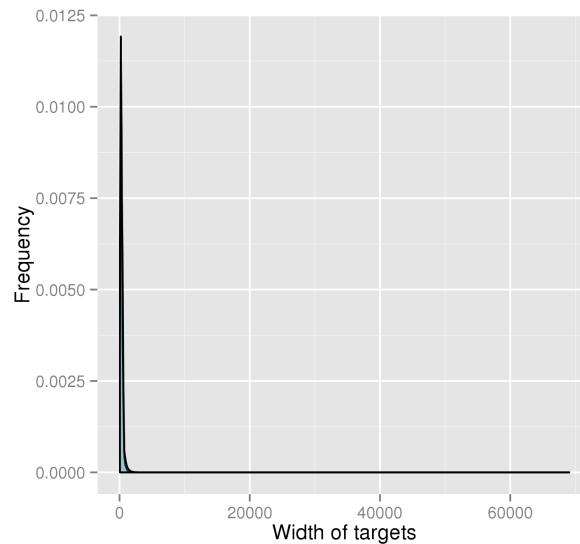
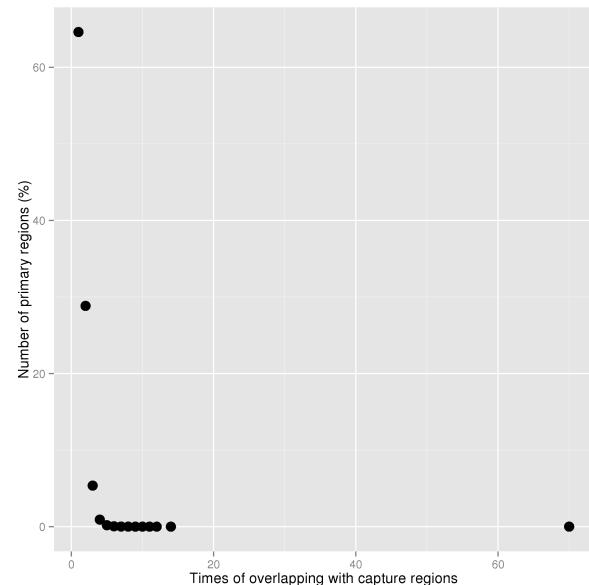
A first report on EpiScope project

Bis capture-seq data: quality control recap

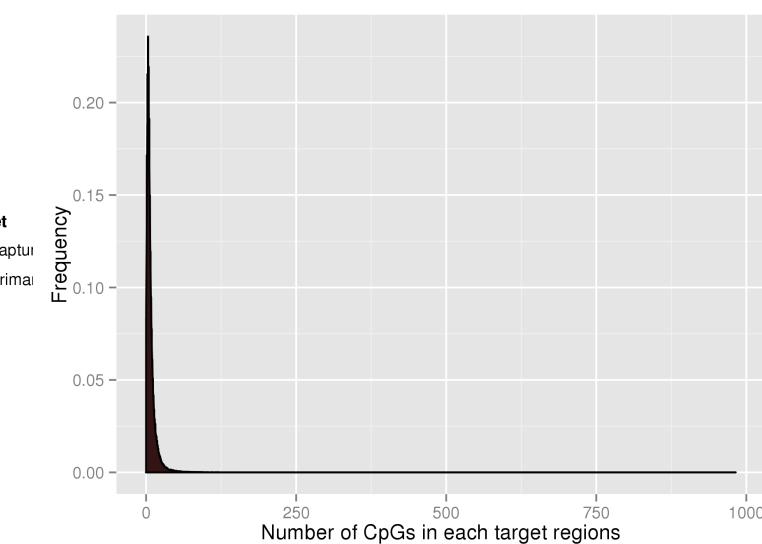
2. BisCaptureSeqQC: a quality control software package for bisulfite treatment and targeted enrichment sequencing

1. What do we capture?

	nTarget	Width.of.targets.bp.	nCpG	Internal.Overlap
<i>capture</i>	143,792	38,339,905	809,747	0
<i>primary</i>	100,641	35,674,231	774,294	0



	X0.	X25.	X50.	X75.	X100.
<i>capture</i>	55	136	248	319	7487
<i>primary</i>	200	200	200	400	69060

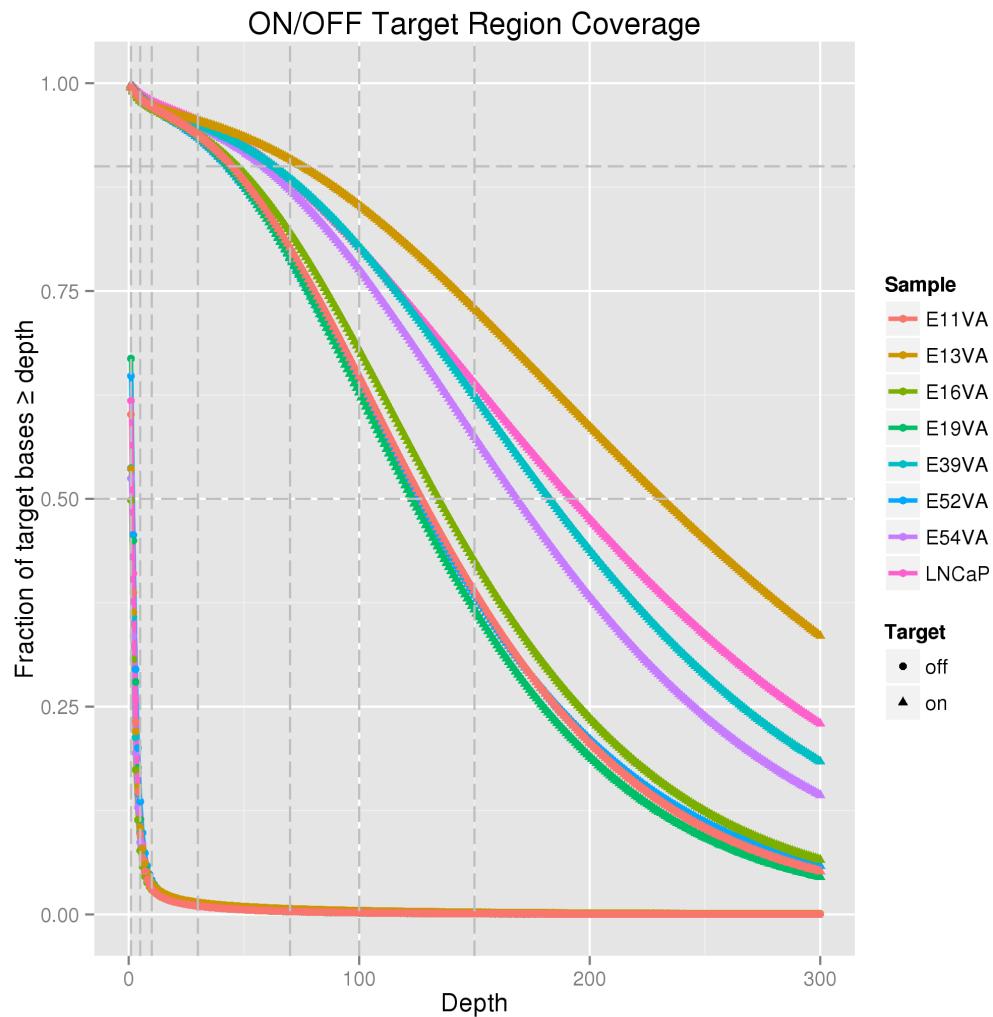


Number.of.CpGs	0	3	5	9	982

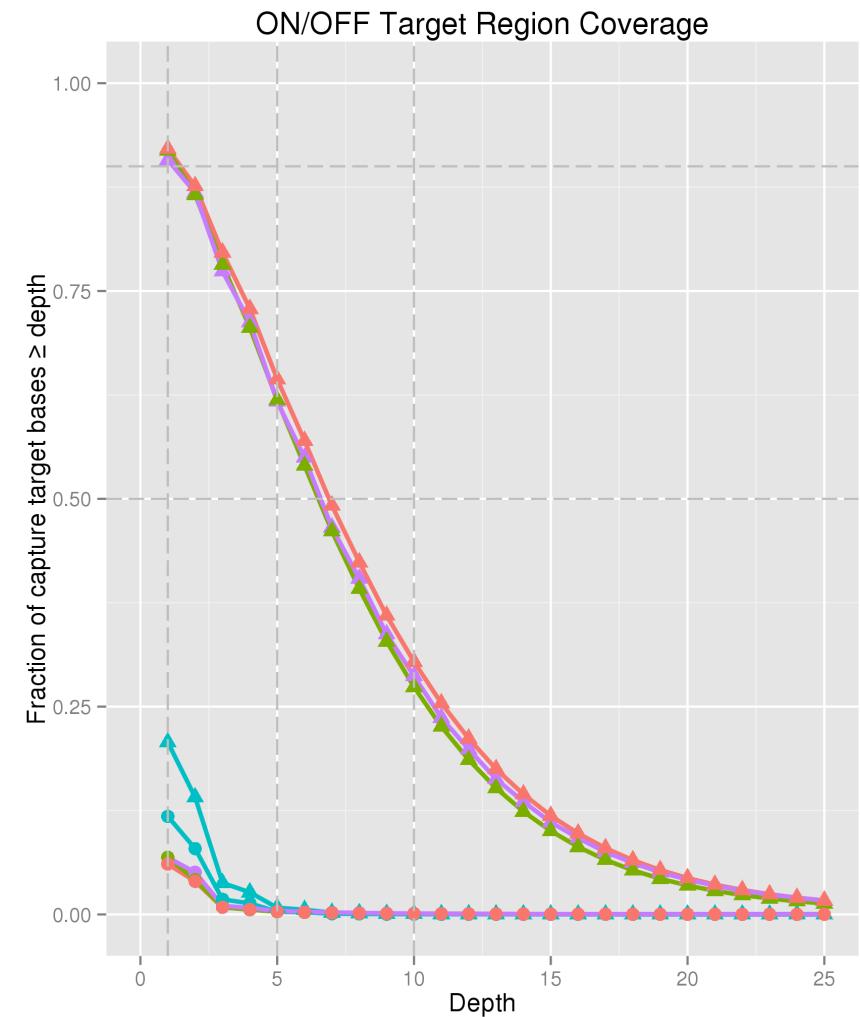
2. BisCaptureSeqQC

2. Coverage of bases

HiSeq2500



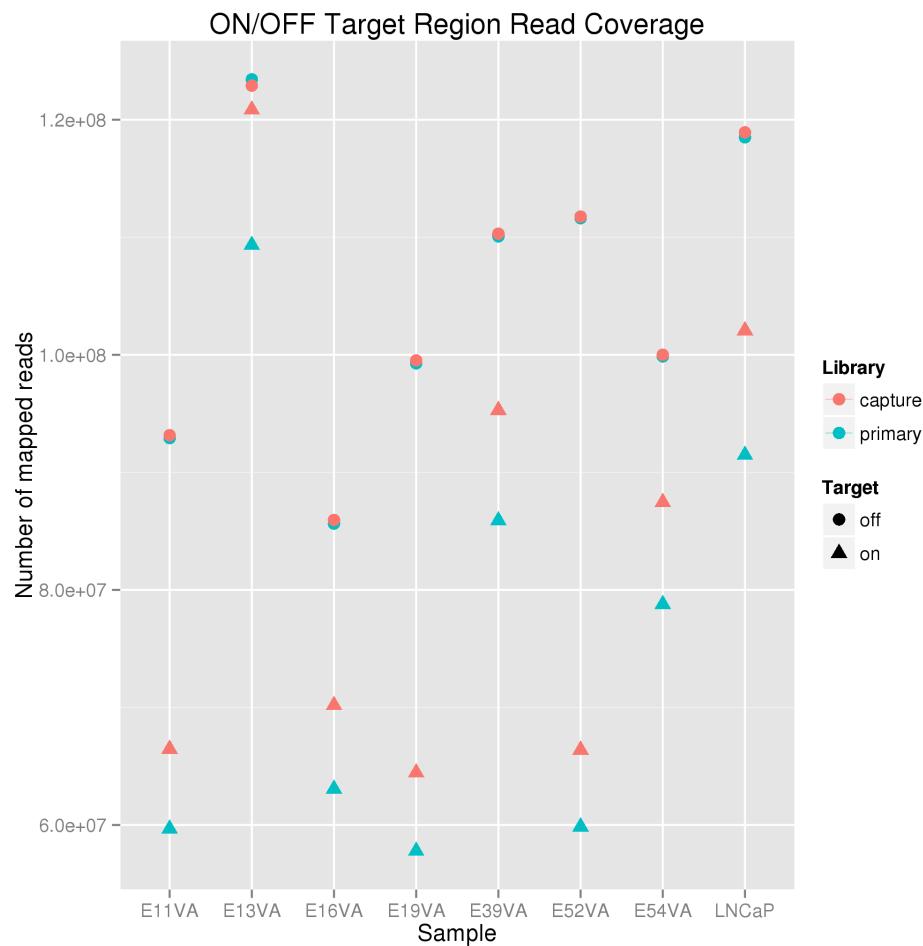
MiSeq



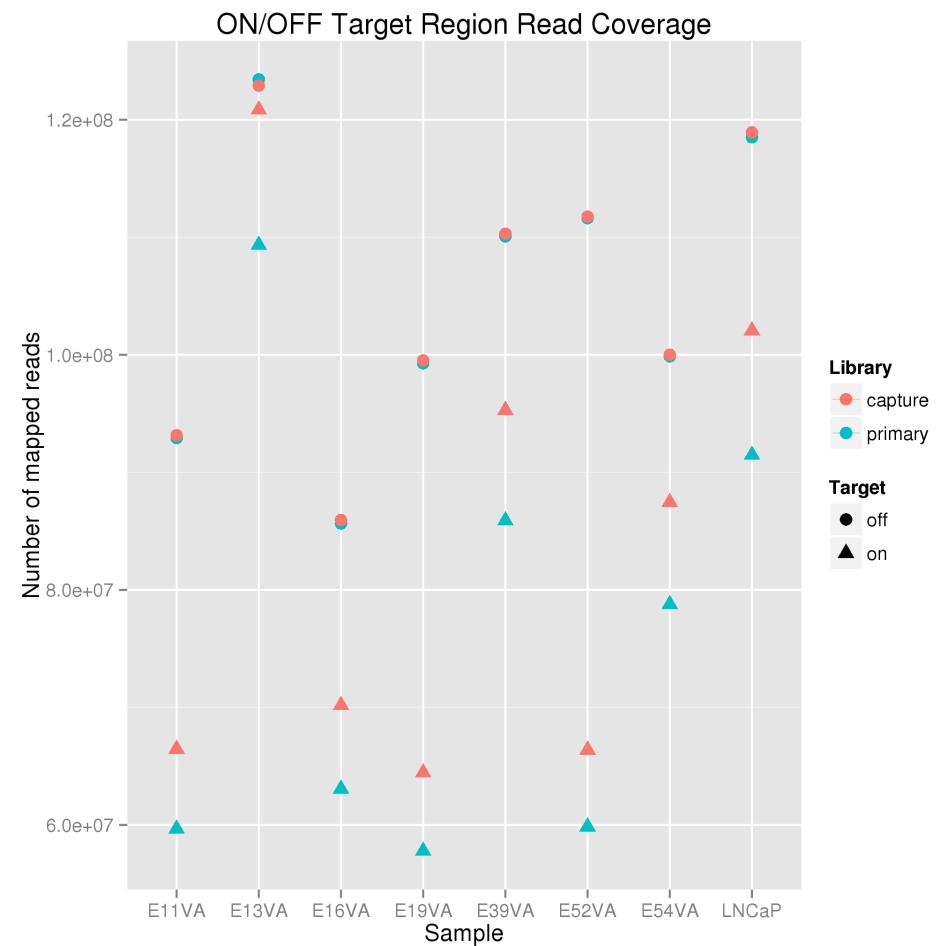
2. BisCaptureSeqQC

3. Coverage of reads

HiSeq2500



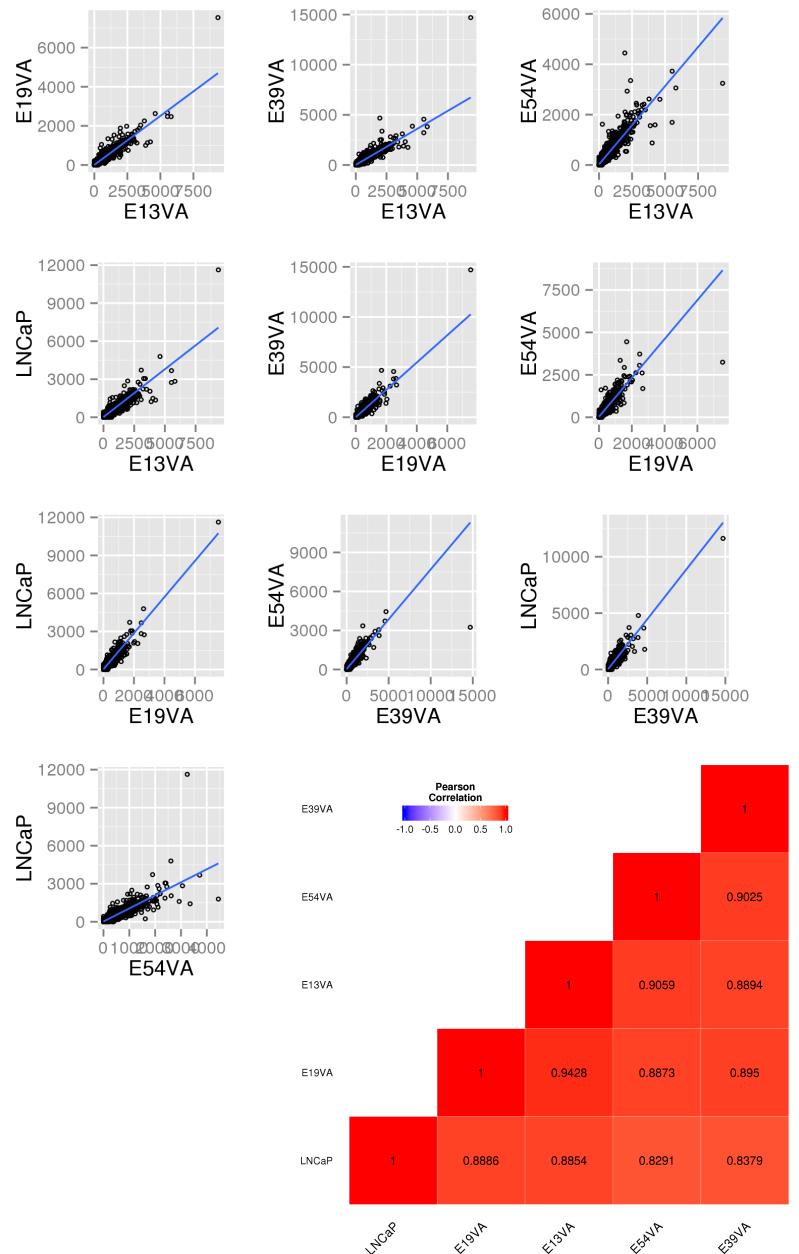
MiSeq



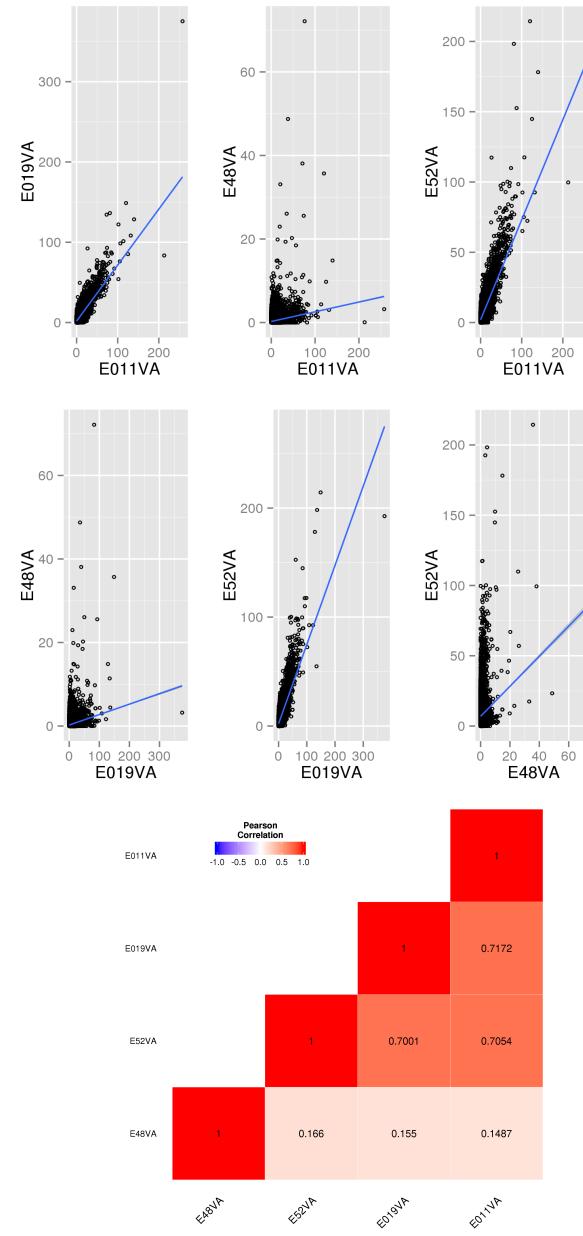
2. BisCaptureSeqQC

3. Is coverage correlation among samples?

HiSeq2500



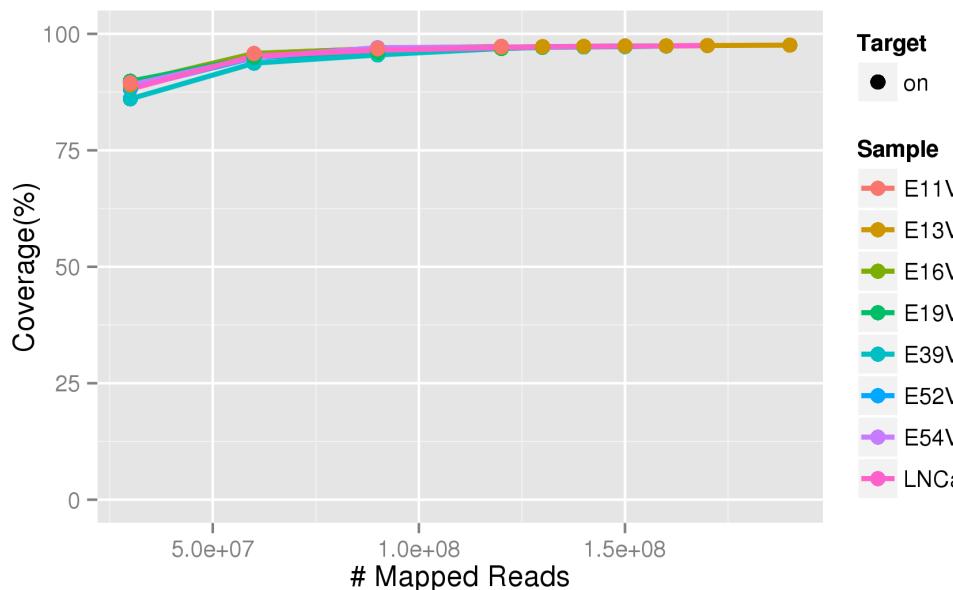
MiSeq



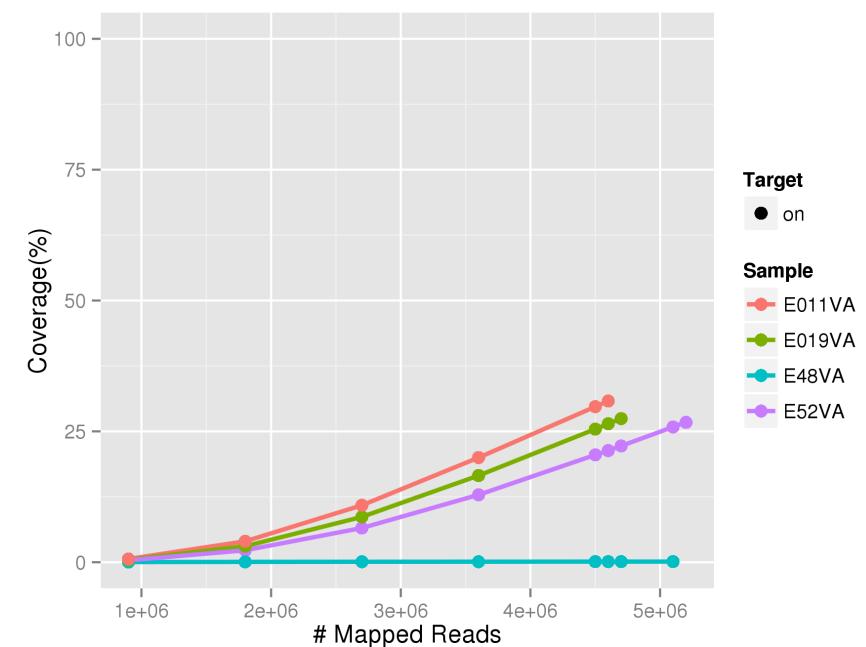
2. BisCaptureSeqQC

4. Do we need to run the sequencing more rounds?

HiSeq2500



MiSeq



2. BisCaptureSeqQC

5. Which targets did we miss?

HiSeq2500

1	Target	Sample
2	chr1:22580560-22580760	E13VA
3	chr1:110234372-110234572	E16VA
4	chr1:110234372-110234572	E52VA
5	chr1:110234372-110234572	E54VA
6	chr1:152562825-152563025	E11VA
7	chr1:152562825-152563025	E13VA
8	chr1:152562825-152563025	E19VA
9	chr1:152562825-152563025	E39VA
10	chr1:152562825-152563025	E54VA
11	chr1:152580068-152580268	E13VA
12	chr1:152580068-152580268	E39VA
13	chr1:152580068-152580268	E54VA
14	chr2:47734946-47735146	LNCaP
15	chr2:47761344-47761744	LNCaP
16	chr2:47952348-47952848	LNCaP
17	chr5:505419-505619	E16VA
18	chr5:60133072-60133272	E19VA
19	chr5:180417354-180417554	E13VA
20	chr6:32528828-32529328	E54VA

MiSeq

1	Target	Sample
2	chr1:237595-237995	E39VA
3	chr1:237595-237995	E52VA
4	chr1:237595-237995	E54VA
5	chr1:237595-237995	LNCaP
6	chr1:248976-249176	E16VA
7	chr1:248976-249176	E39VA
8	chr1:248976-249176	E52VA
9	chr1:248976-249176	E54VA
10	chr1:248976-249176	LNCaP
11	chr1:568750-569150	E39VA
12	chr1:568750-569150	E52VA
13	chr1:568750-569150	E54VA
14	chr1:568750-569150	LNCaP
15	chr1:569721-570121	E39VA
16	chr1:569721-570121	E52VA
17	chr1:569721-570121	E54VA
18	chr1:569721-570121	LNCaP
19	chr1:713076-713576	E16VA
20	chr1:713076-713576	E39VA
21	chr1:713076-713576	E52VA

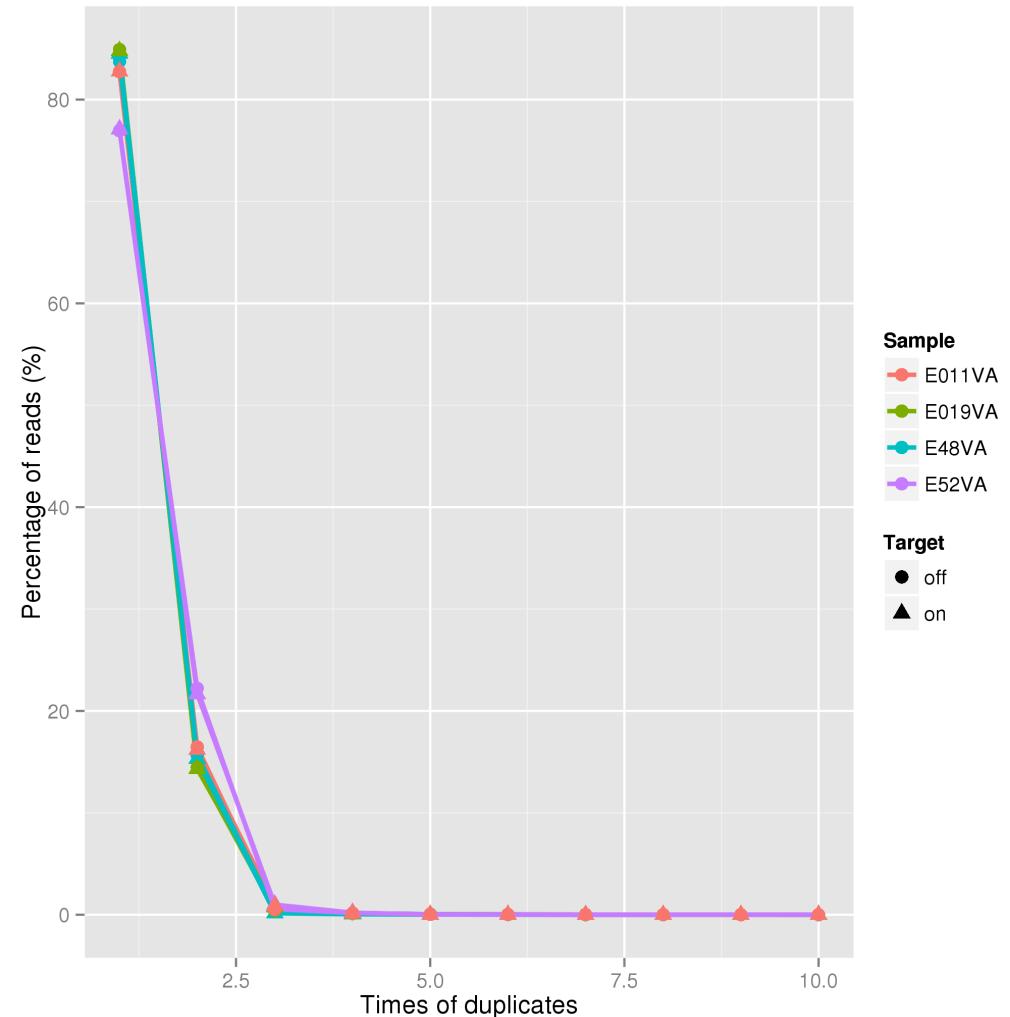
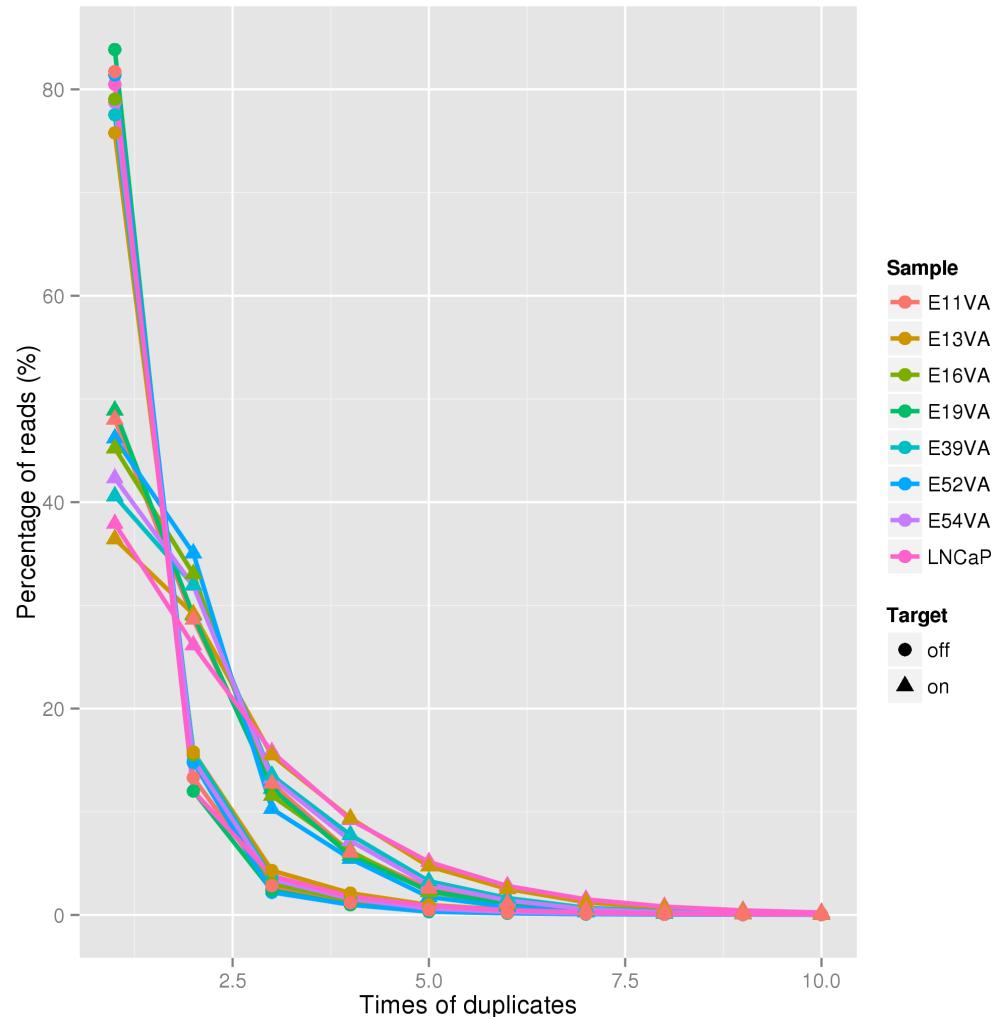
...

2. BisCaptureSeqQC

6. How is the duplication?

HiSeq2500

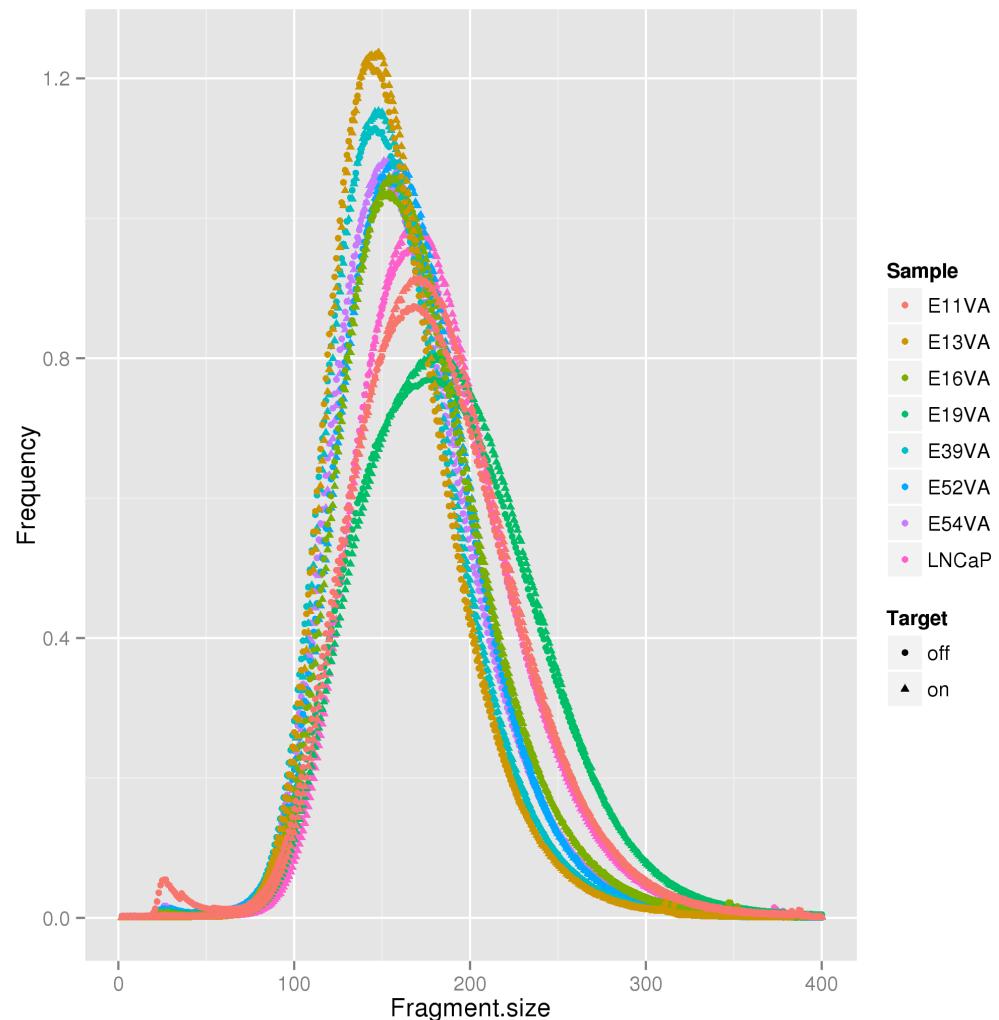
MiSeq



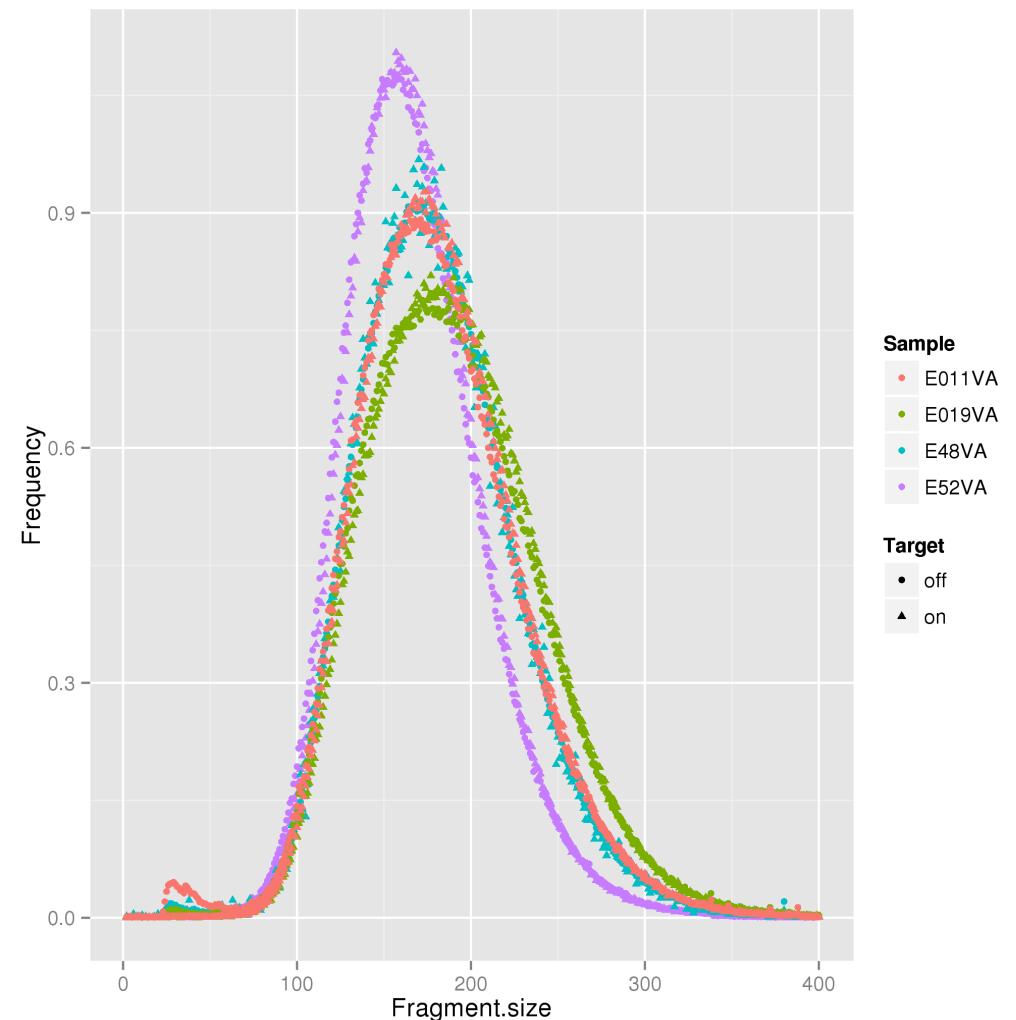
2. BisCaptureSeqQC

7. How is the fragment size?

HiSeq2500



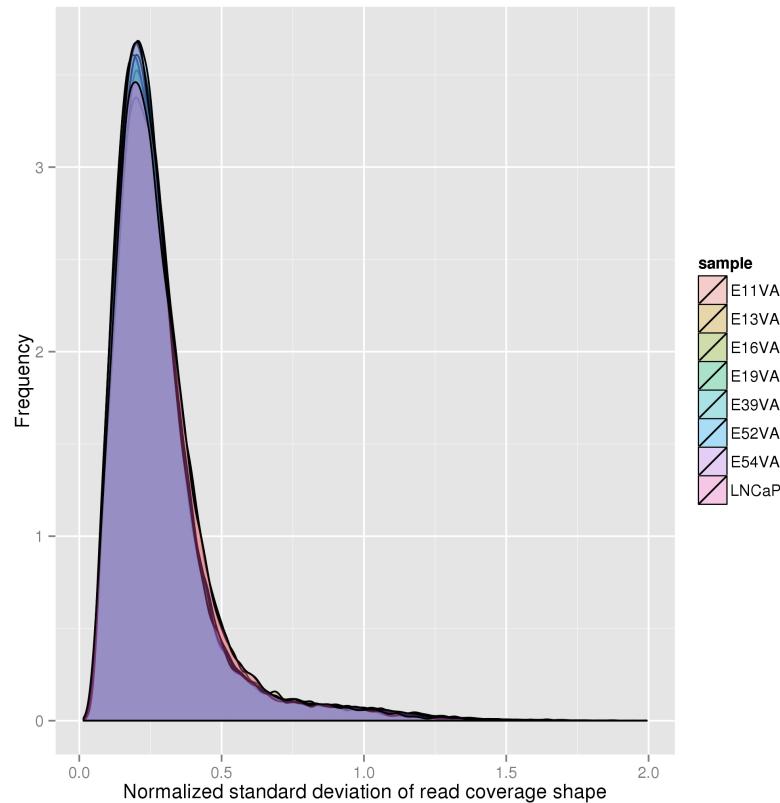
MiSeq



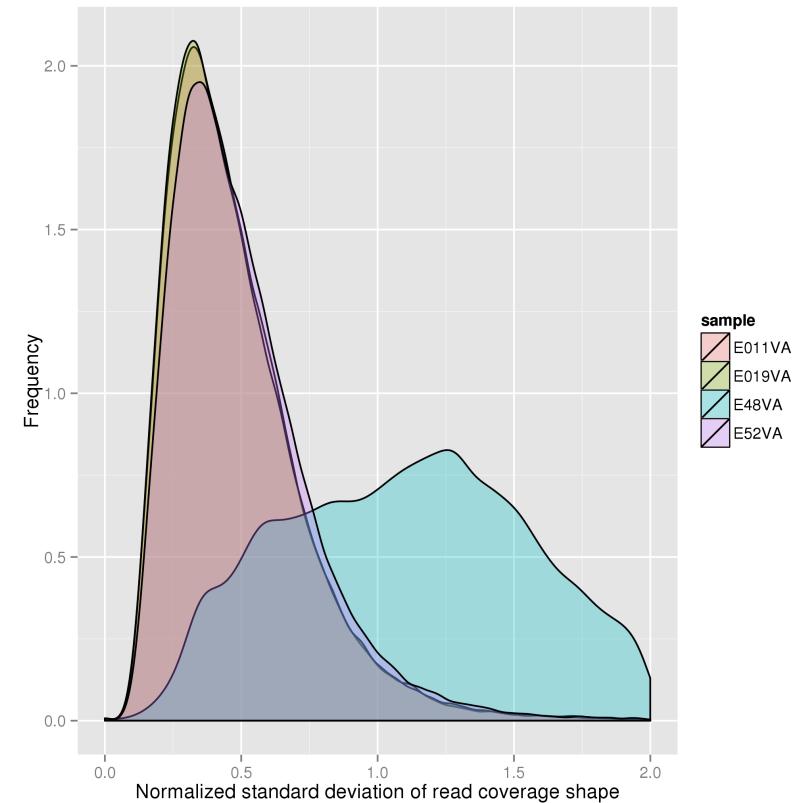
BisCaptureSeqQC

7. What is the coverage shape?

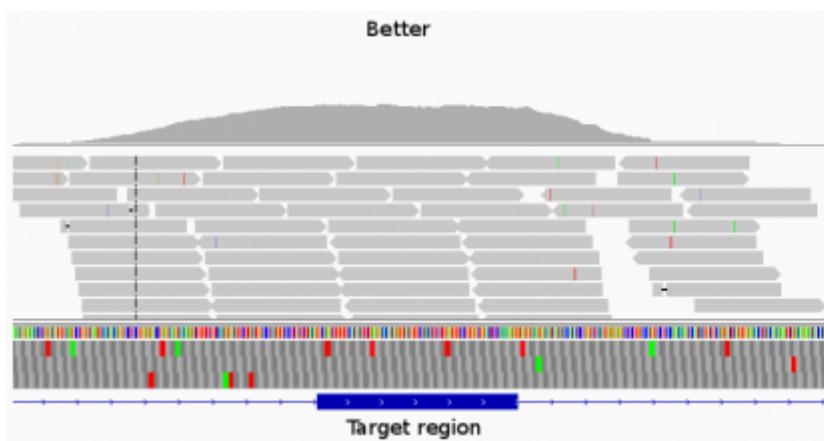
HiSeq2500



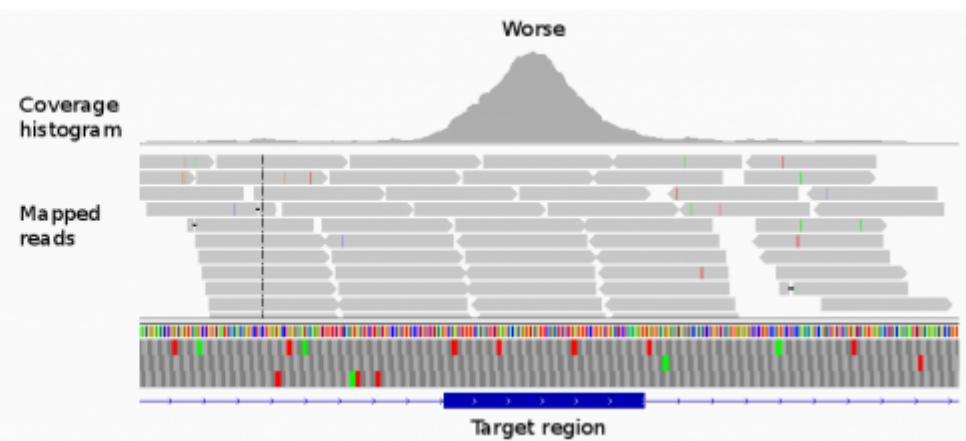
MiSeq



Better

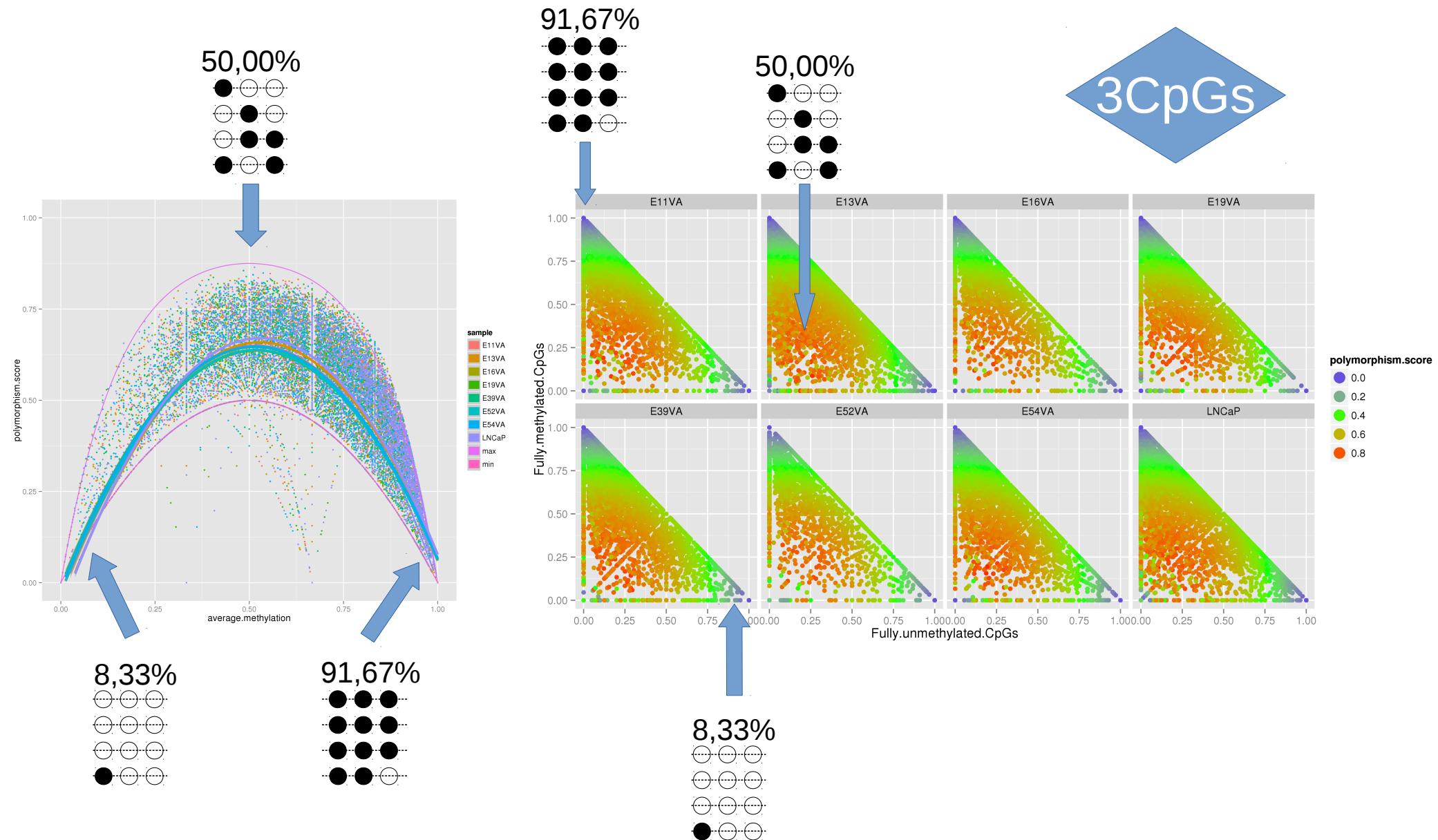


Worse



2. BisCaptureSeqQC

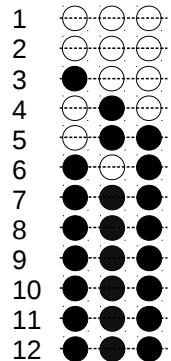
8. Does BiscaptureSeq capture the heterogeneity of DNA methylation patterns?



Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. Landan G, Cohen NM, Mukamel Z, Bar A, Molchadsky A, Brosh R, Horn-Saban S, Zalcenstein DA, Goldfinger N, Zundelevich A, Gal-Yam EN, Rotter V, Tanay A. *Nature Genetics* (2012)

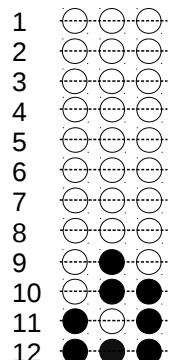
Explaining the previous right figure

67,68%

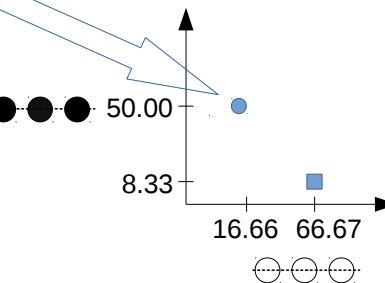


2 \rightarrow 2/12 = 16,66%
1 \rightarrow 1/12 = 8,33%
1 \rightarrow 1/12 = 8,33%
1 \rightarrow 1/12 = 8,33%
1 \rightarrow 2/12 = 8,33%
6 \rightarrow 6/12 = 50,00%

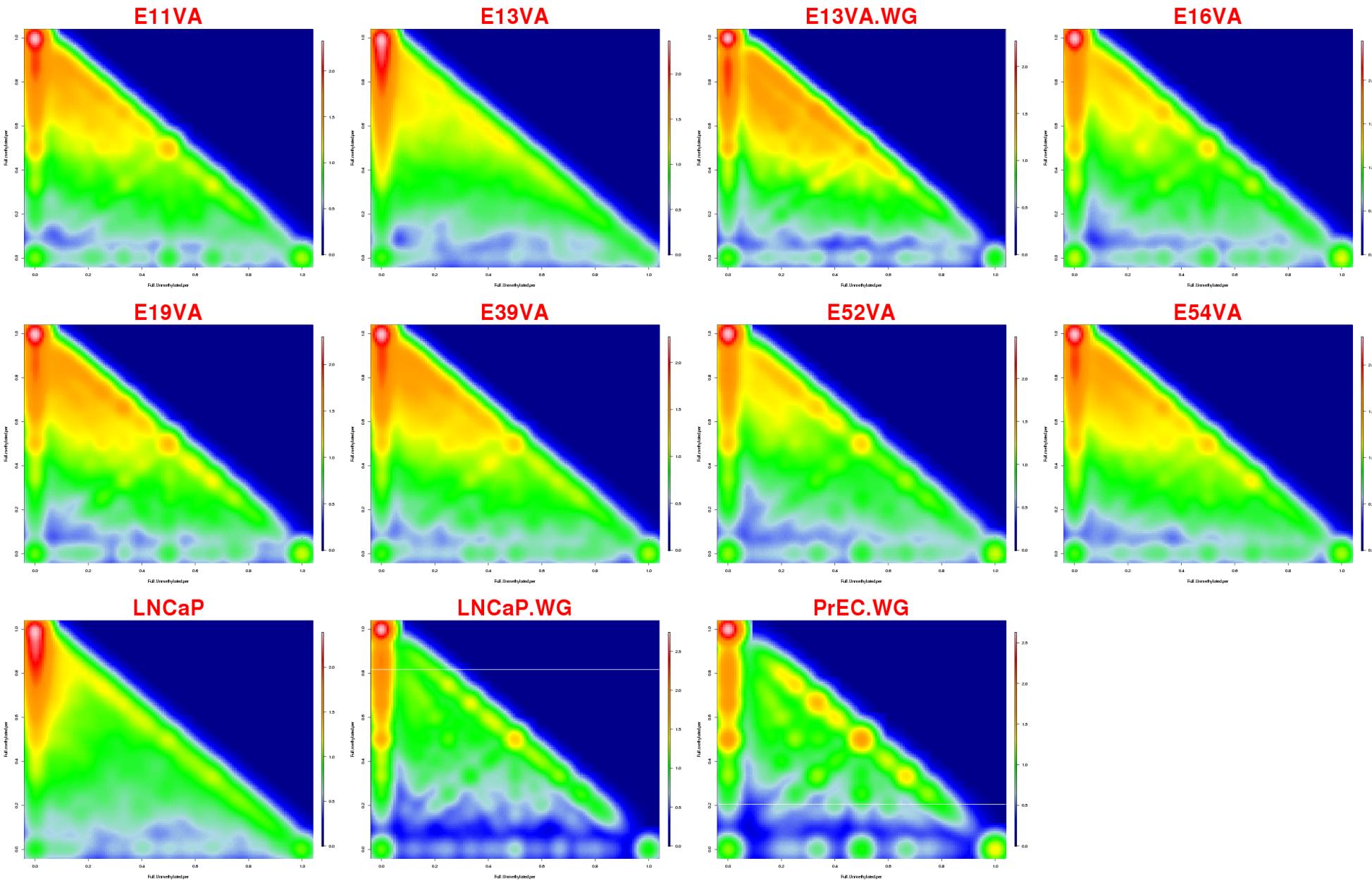
16.67%



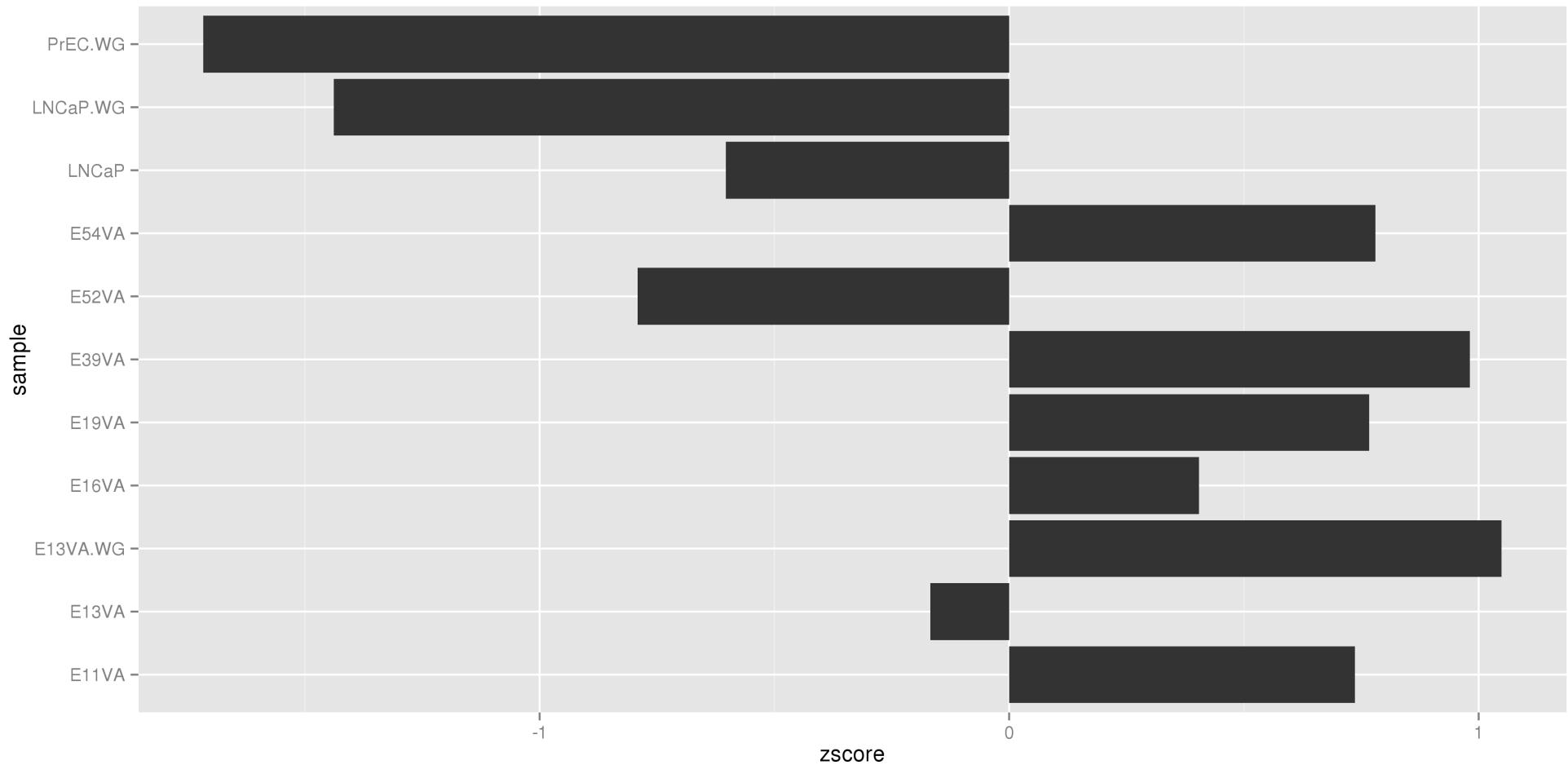
8 \rightarrow 2/12 = 66,67%
1 \rightarrow 1/12 = 8,33%
1 \rightarrow 1/12 = 8,33%
1 \rightarrow 1/12 = 8,33%
1 \rightarrow 1/12 = 8,33%



Same as the previous right figure but colour by density of points



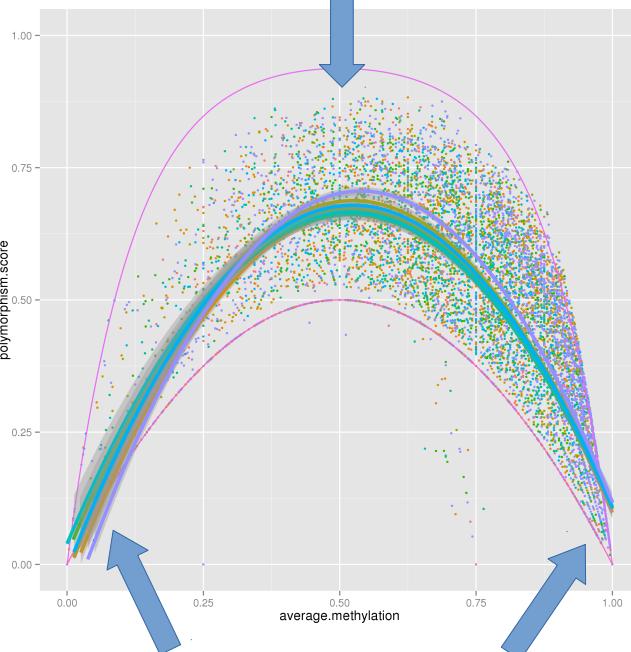
Fraction of loci with a disordered methylation state for samples, shown in terms of standard deviation from the average methylation in the sample



2. BisCaptureSeqQC

8. Does BiscaptureSeq capture the heterogeneity of DNA methylation patterns?

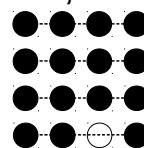
56,25%



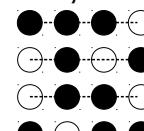
6,25%

93,75%

93,75%



56,25%



4CpGs

sample

E11VA

E13VA

E16VA

E19VA

E39VA

E52VA

E54VA

LNCaP

max

min

Fully.methylated.CpGs

E11VA

E13VA

E16VA

E19VA

E39VA

E52VA

E54VA

LNCaP

max

min

Fully.unmethylated.CpGs

E11VA

E13VA

E16VA

E19VA

E39VA

E52VA

E54VA

LNCaP

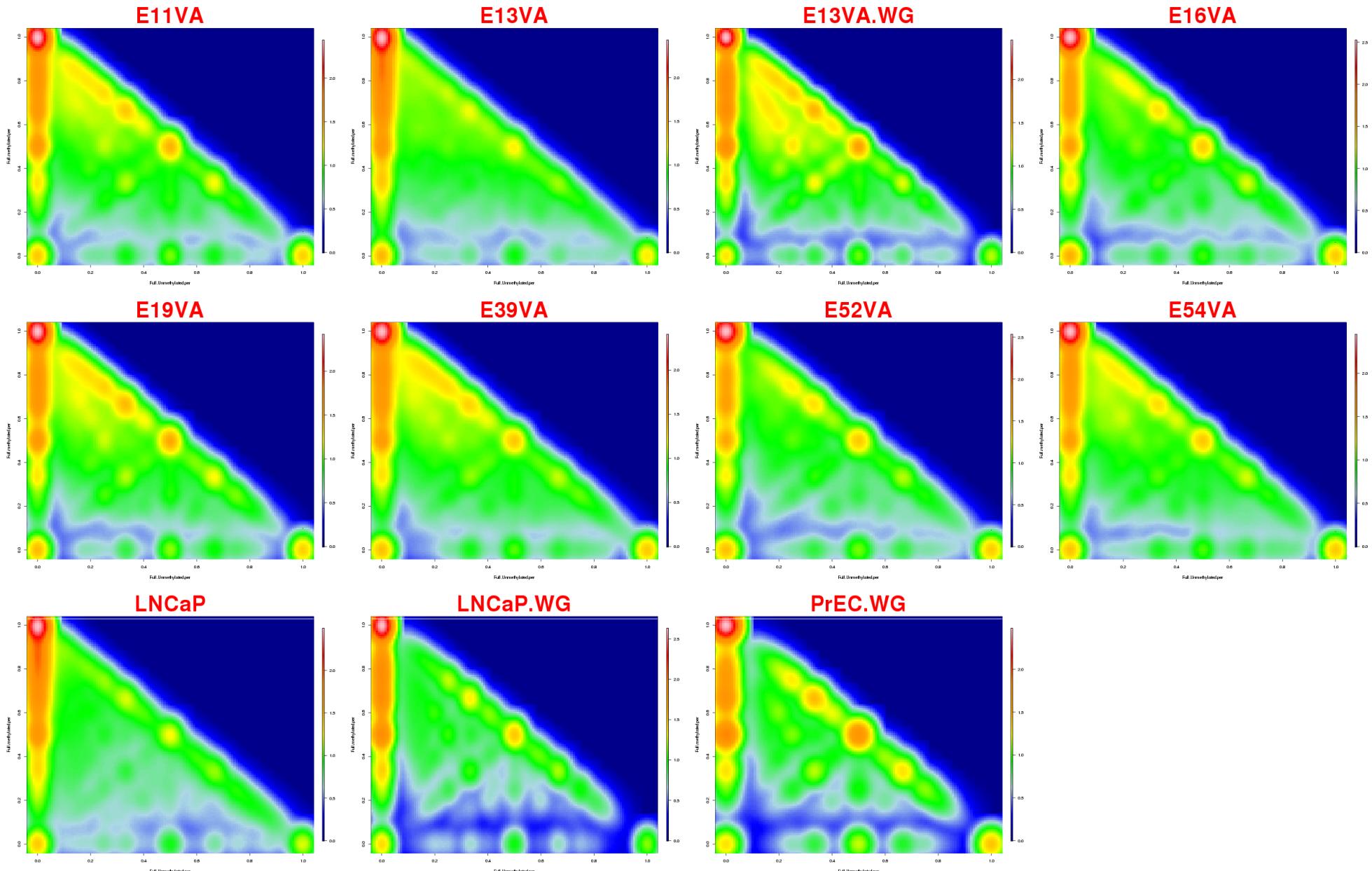
max

min

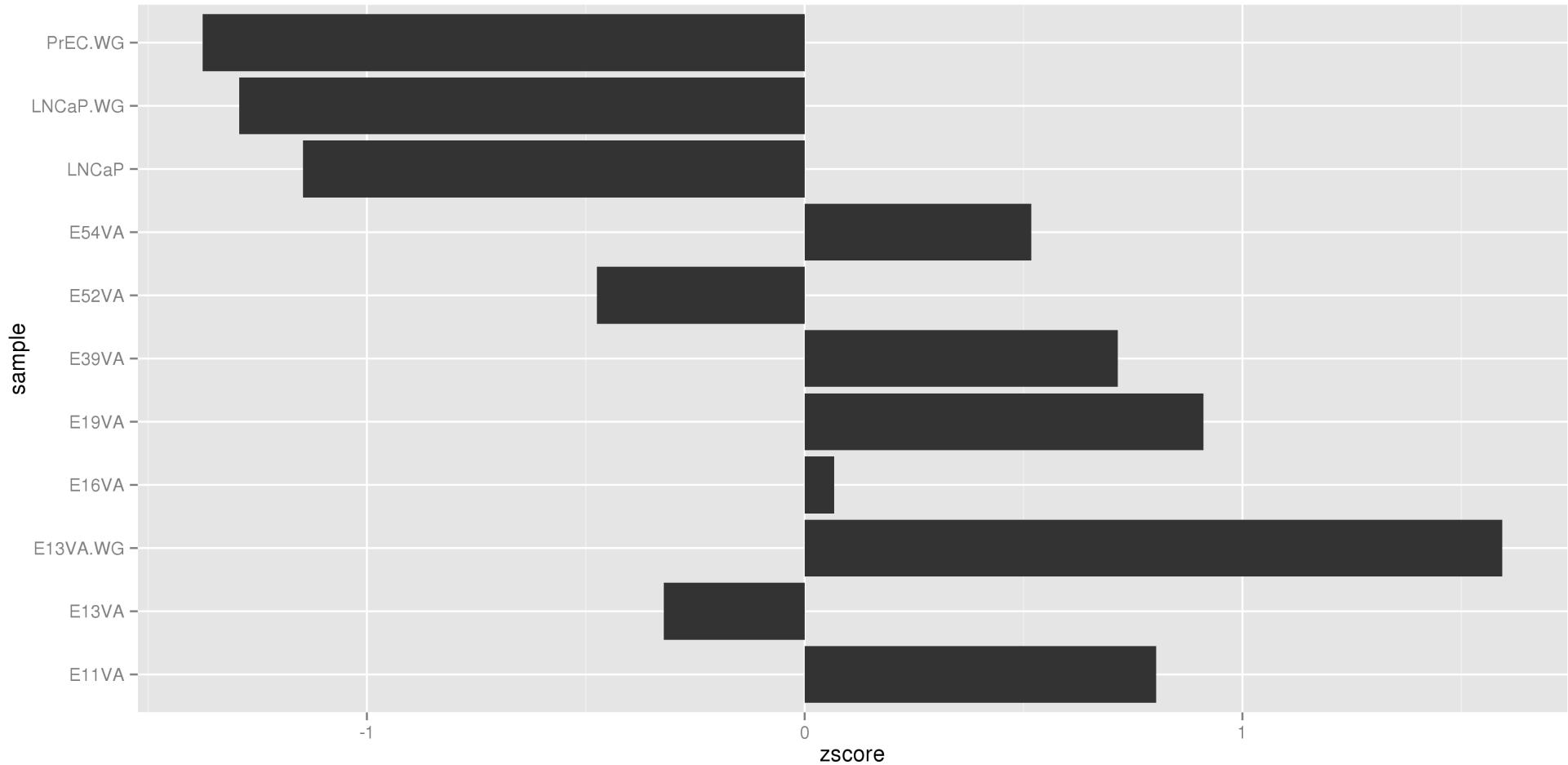
Fully.unmethylated.CpGs

6,25%

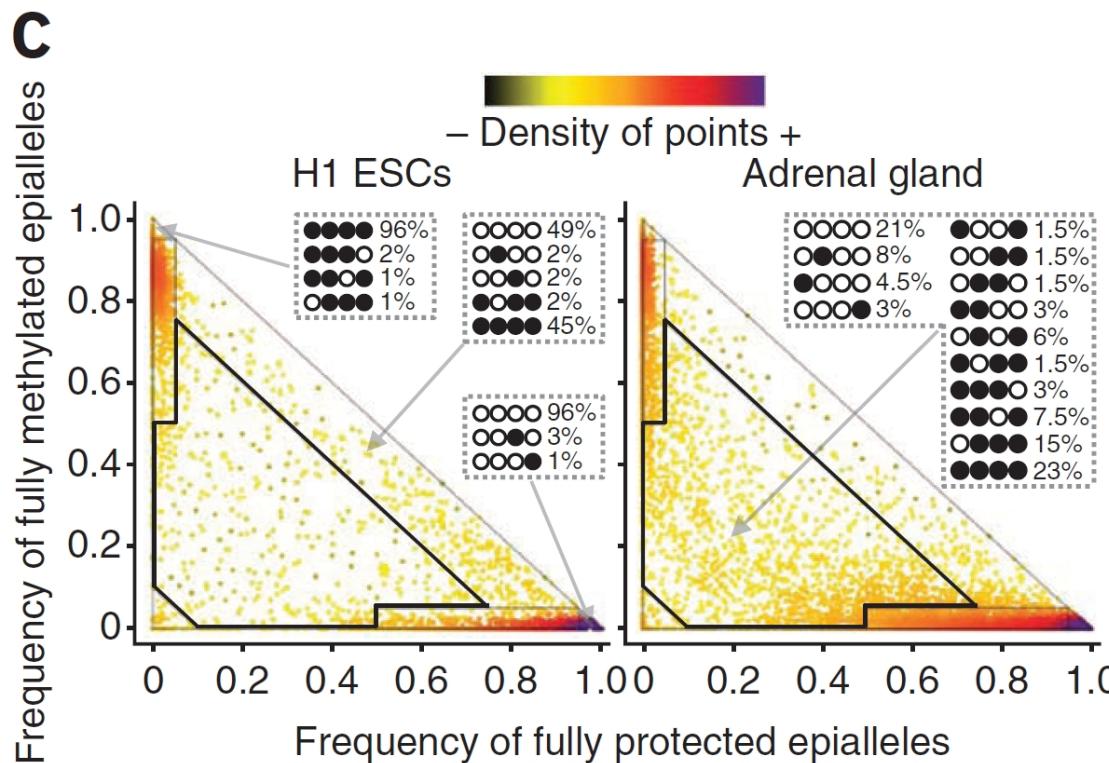
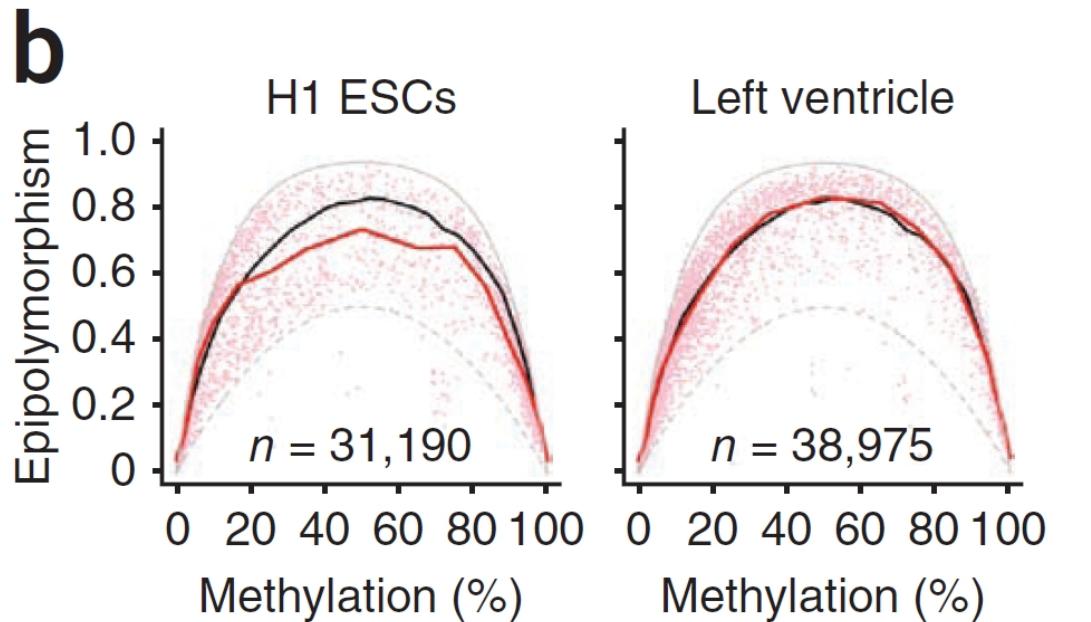
Same as the previous right figure but colour by density of points



Fraction of loci with a disordered methylation state for samples, shown in terms of standard deviation from the average methylation in the sample



Max-Min and neither uniform nor bimodal



2. BisCaptureSeqQC

Discussion

BisCaptureSeq

- + mapped reads on target > 40%
- + base coverage ~ 100
- + Hiseq2500 run makes saturated read coverage, no more run needed
- + DNA methylation patterns are fully captured (still need to improve the score)

The second report on EpiScope project

Methylation profiles
in the bis capture-seq data for some of the key
regions of interest
(especially in **PITX2** and **NKX2-5** genes)

Content

1. Does BiscaptureSeq give the same DNA methylation profile as WGBS at targeted regions?
2. How are DNA methylation status of some important genes such as PITX2 and NKX2-5?

1. DNA methylation profile comparison: BiscaptureSeq vs WGBS at targeted regions

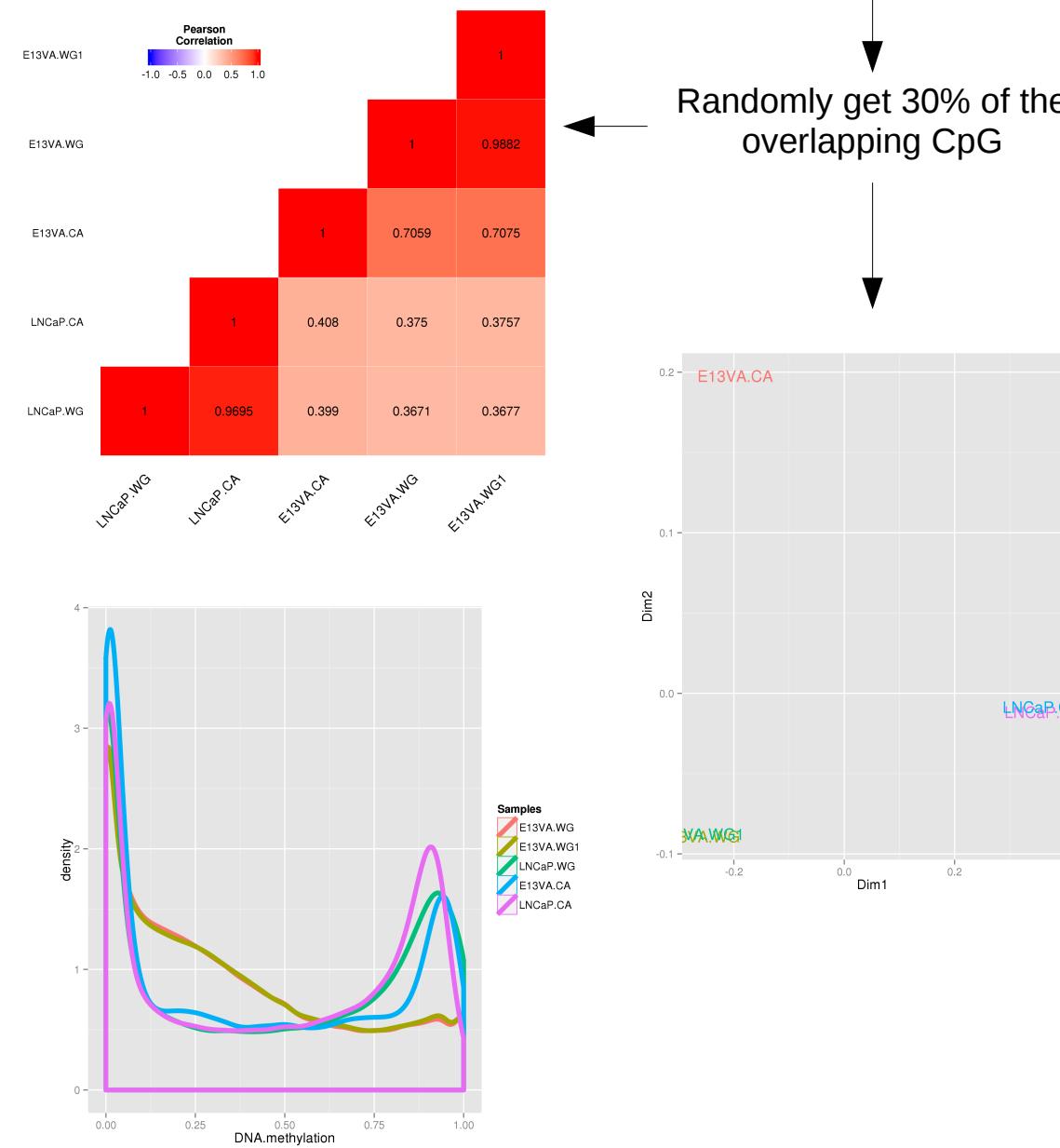
	Sample	Type	Coverage	location
1*	E13VA.WG	WGBS	21	Paper_2
2*	E13VA.WG1	WGBS	21	EpiSCOPE
3	LNCaP.WG	WGBS	26	prostate
4	E13VA.CA	BiscaptureSeq	100	EpiSCOPE (15.09.2015)
5	LNCaP.CA	BiscaptureSeq	100	EpiSCOPE (15.09.2015)

*) We aligned the reads of E13VA with two different aligners, so I put both of them here just want to make sure they are consistent with BisCaptureSeq.

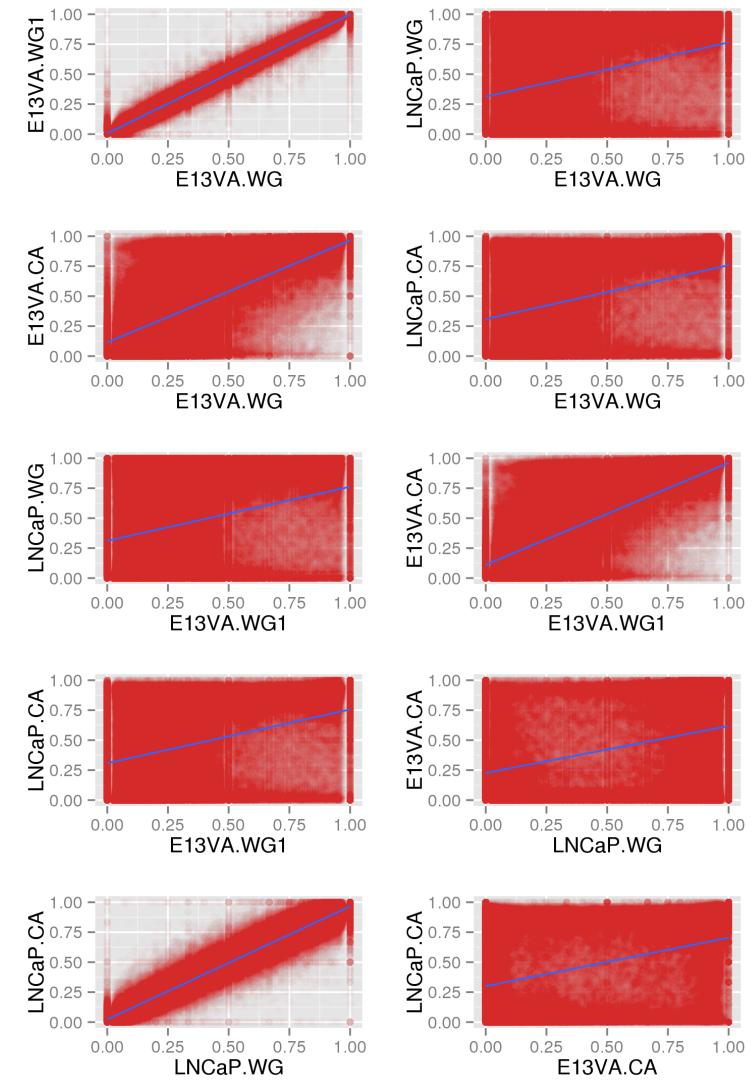
1. DNA methylation profile comparison: BiscaptureSeq vs WGBS at targeted regions

Overlapping among

CpG sites in primary targets \cap **BiscaptureSeq** DNA methylation profile \cap **WGBS** profile



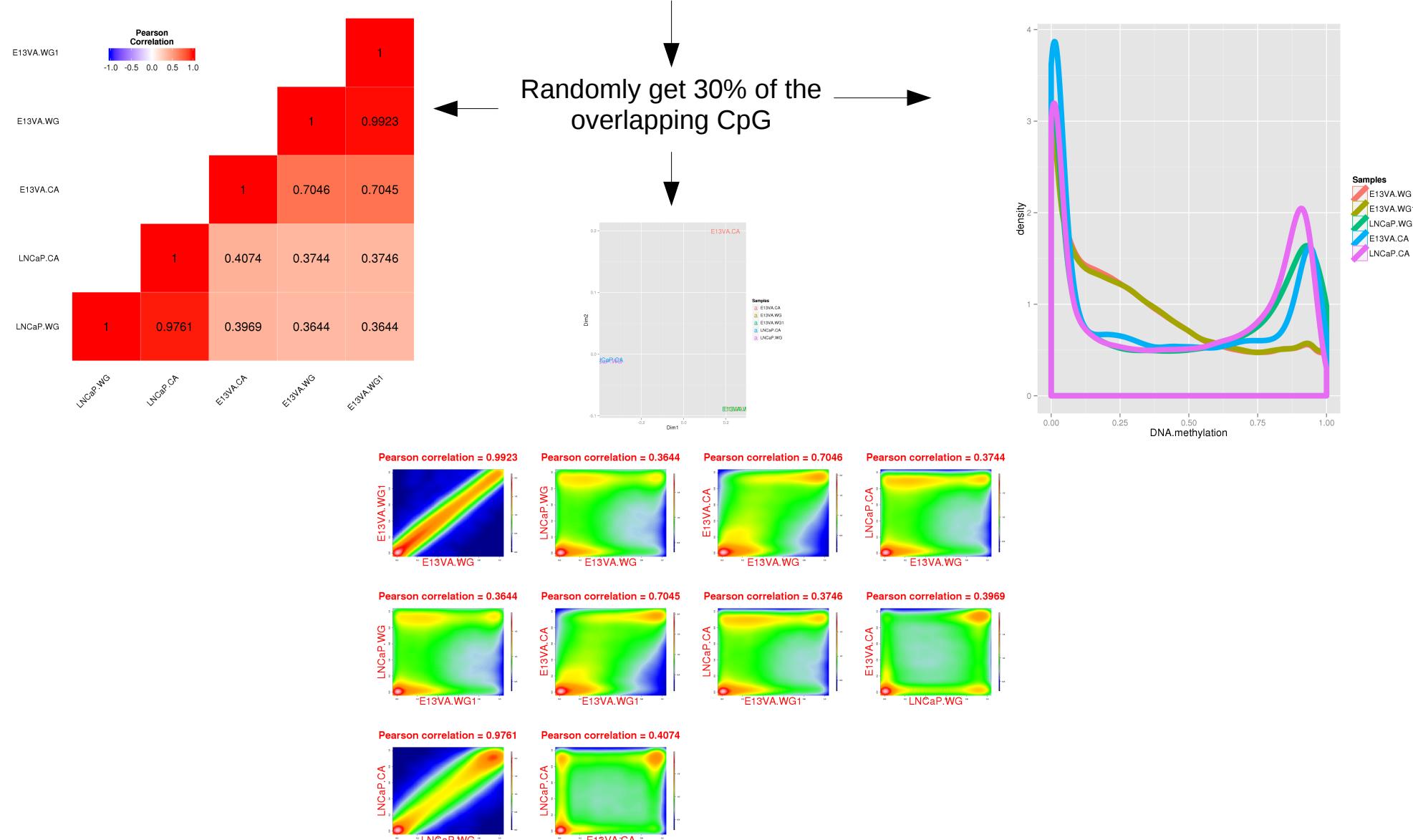
Randomly get 30% of the overlapping CpG



1. DNA methylation profile comparison: BiscaptureSeq vs WGBS at targeted regions

Overlapping among

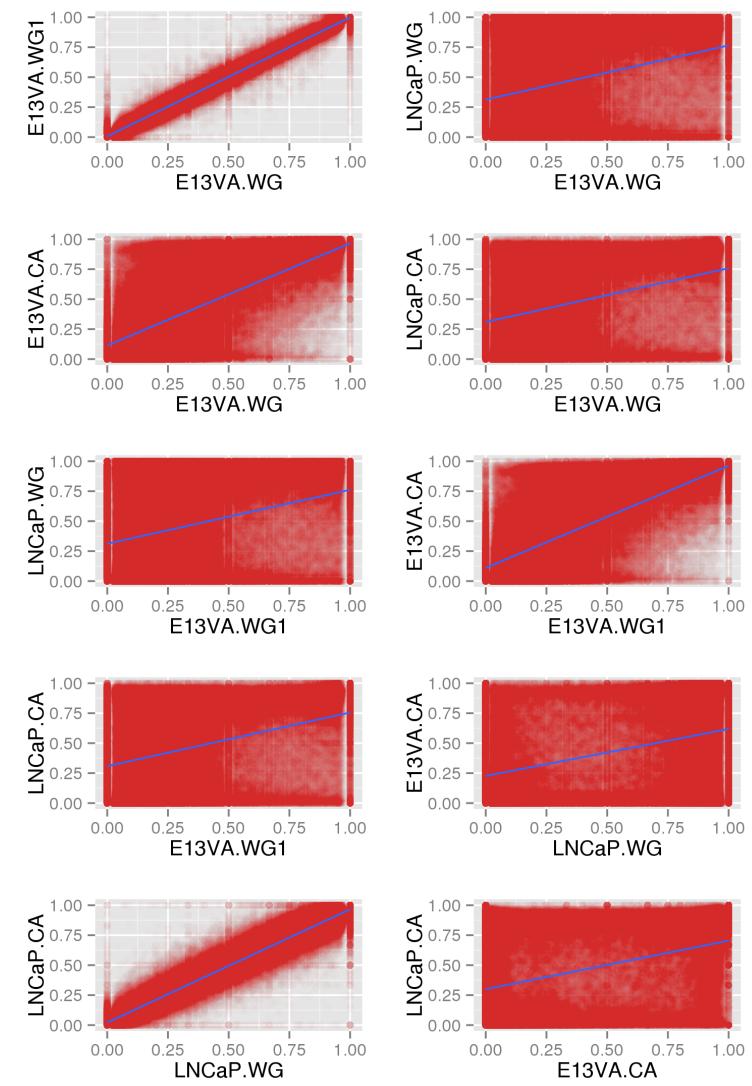
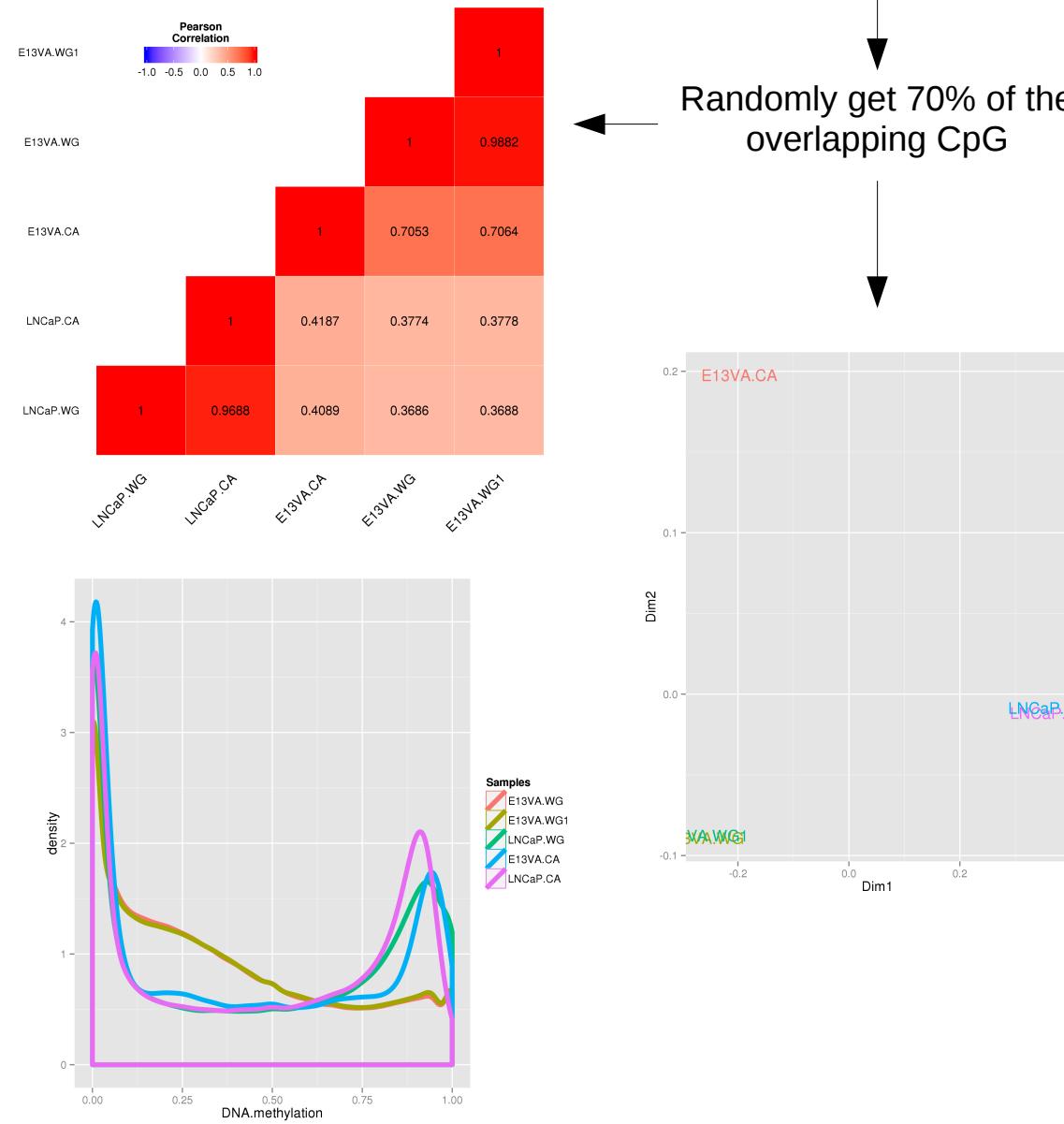
CpG sites in primary targets \cap **BiscaptureSeq** DNA methylation profile \cap **WGBS** profile with **Coverage ≥ 10**



1. DNA methylation profile comparison: BiscaptureSeq vs WGBS at targeted regions

Overlapping among

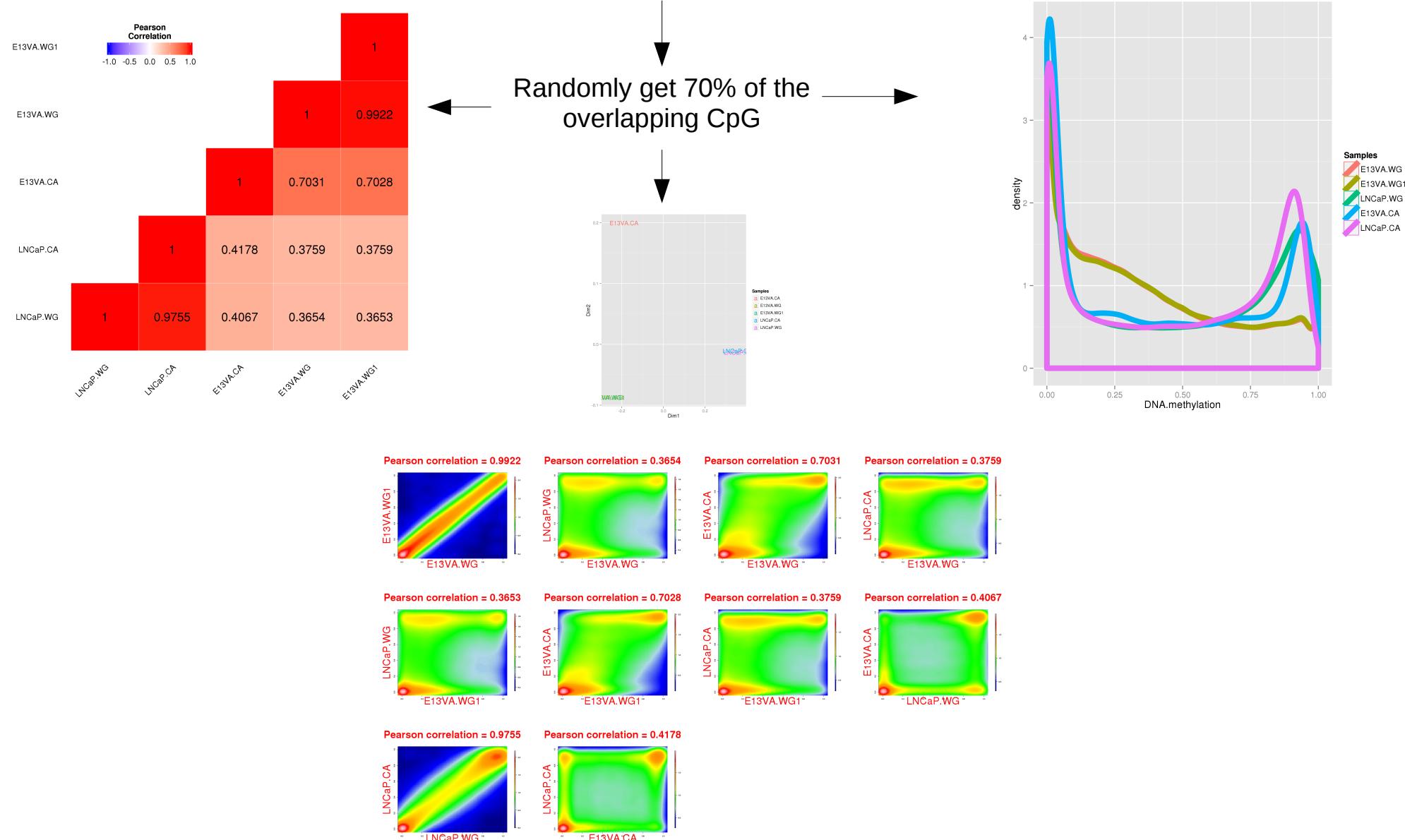
CpG sites in primary targets \cap **BiscaptureSeq** DNA methylation profile \cap **WGBS** profile



1. DNA methylation profile comparison: BiscaptureSeq vs WGBS at targeted regions

Overlapping among

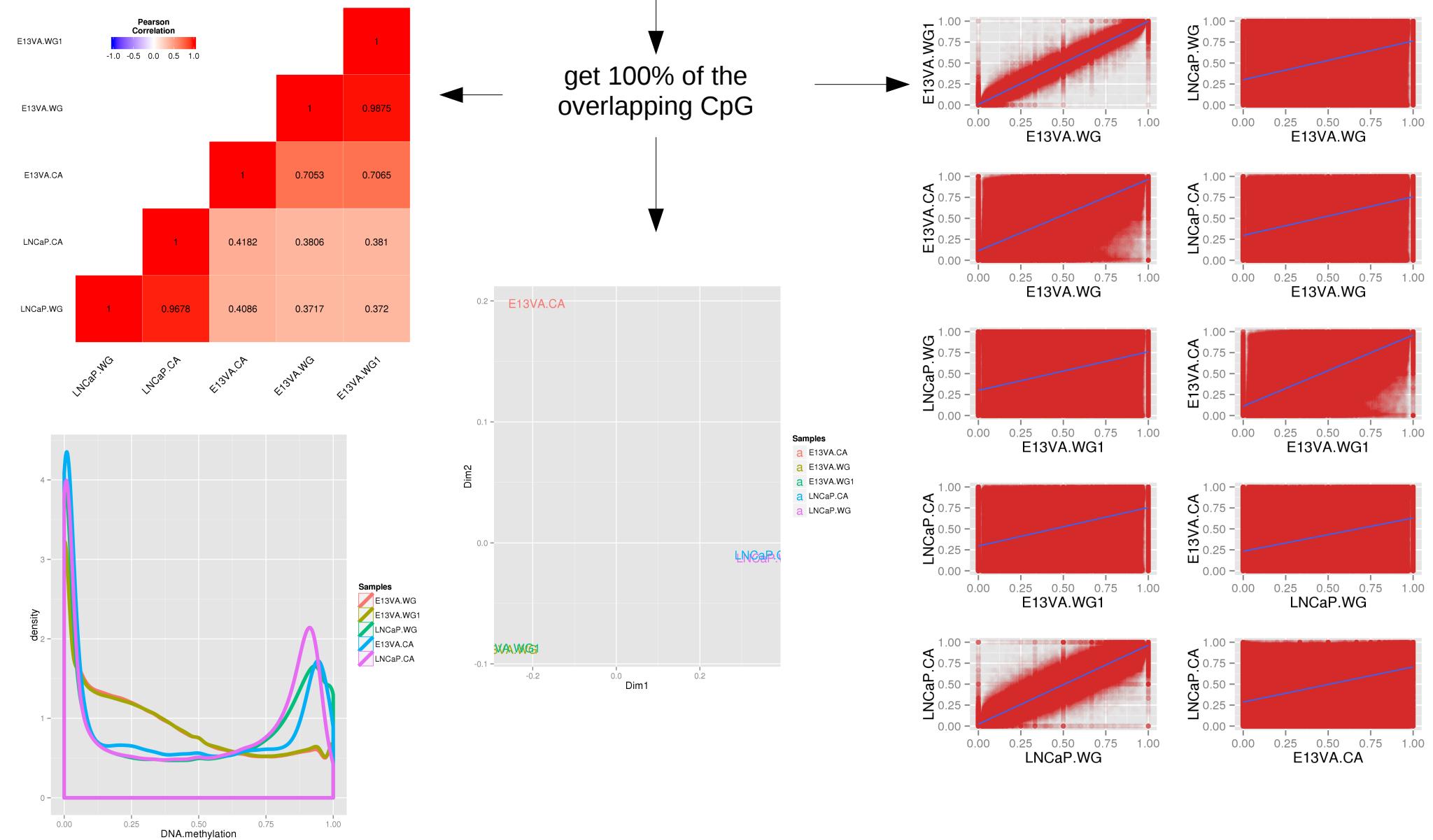
CpG sites in primary targets \cap **BiscaptureSeq** DNA methylation profile \cap **WGBS** profile with **Coverage ≥ 10**



1. DNA methylation profile comparison: BiscaptureSeq vs WGBS at targeted regions

Overlapping among

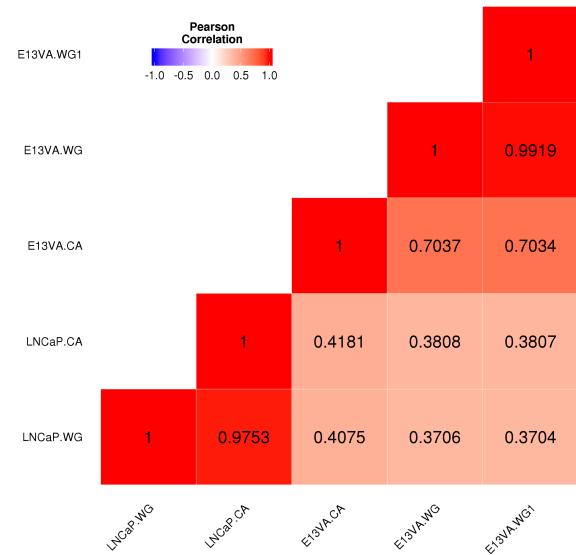
CpG sites in primary targets \cap **BiscaptureSeq** DNA methylation profile \cap **WGBS** profile



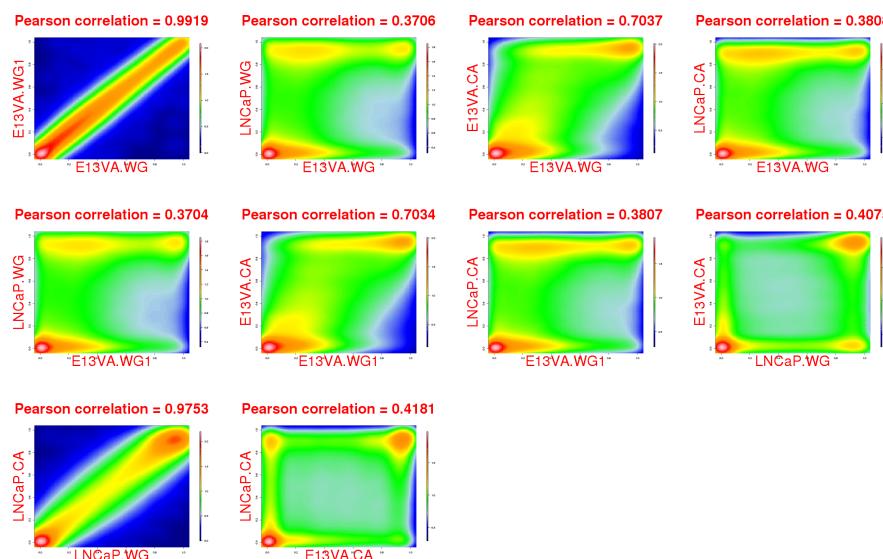
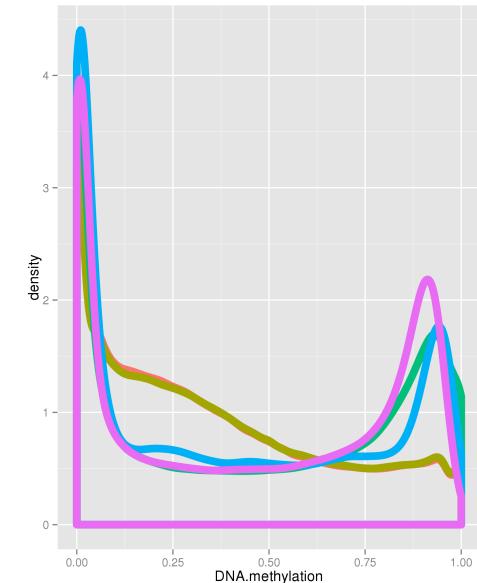
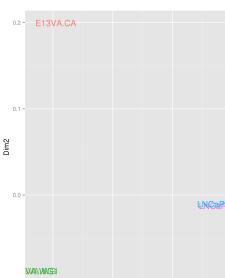
1. DNA methylation profile comparison: BiscaptureSeq vs WGBS at targeted regions

Overlapping among

CpG sites in primary targets \cap **BiscaptureSeq** DNA methylation profile \cap **WGBS** profile with **Coverage ≥ 10**



Randomly get 70% of the overlapping CpG



1. DNA methylation profile comparison: BiscaptureSeq vs WGBS at targeted regions

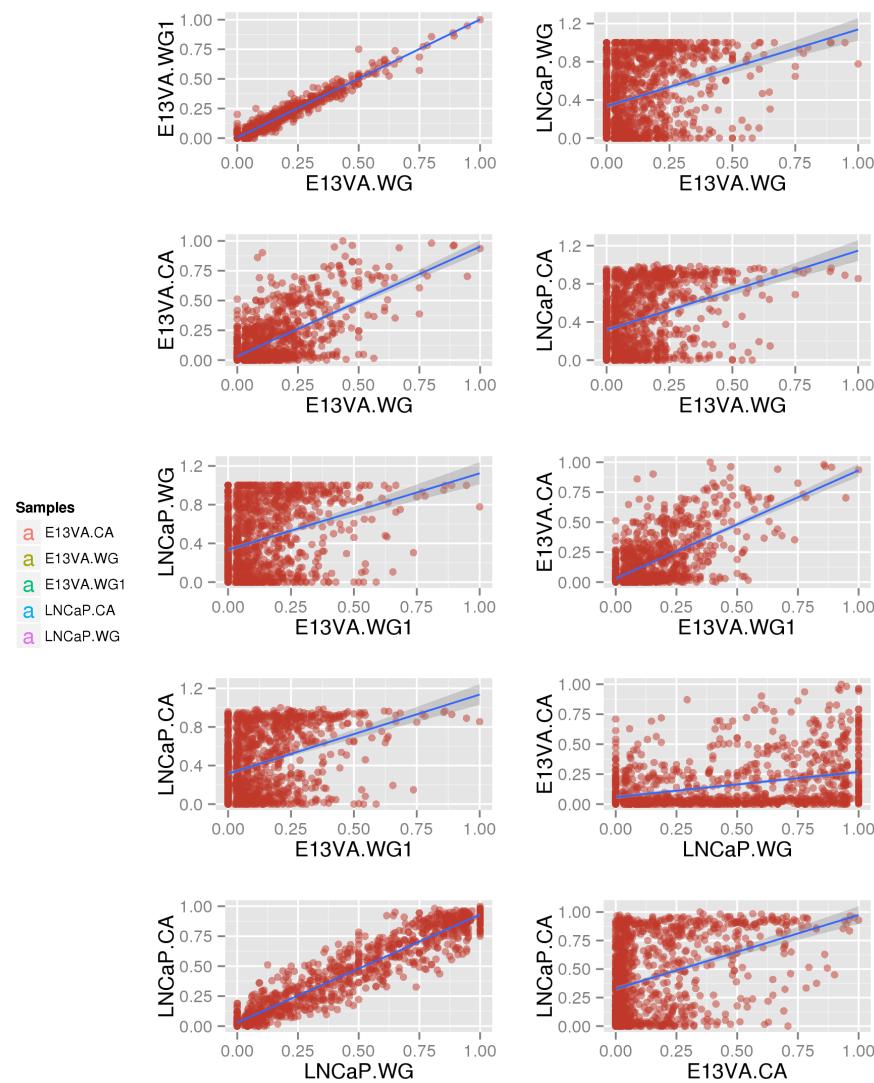
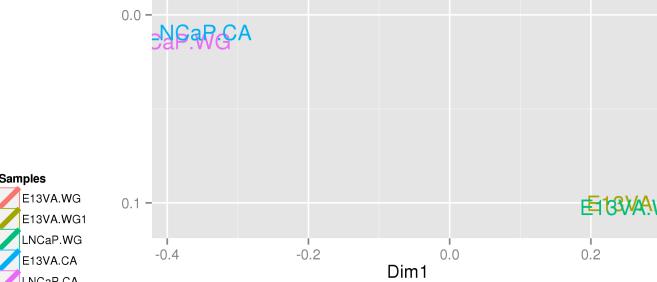
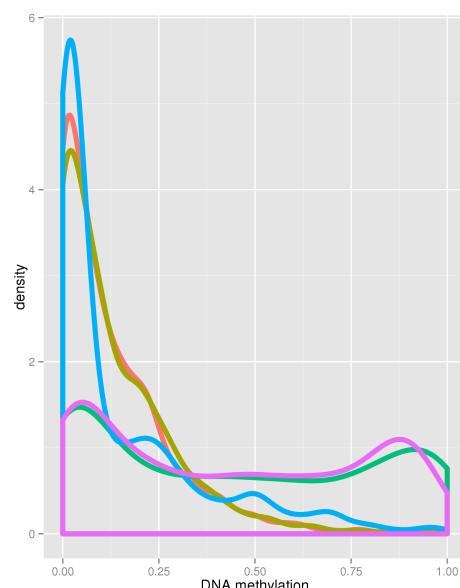
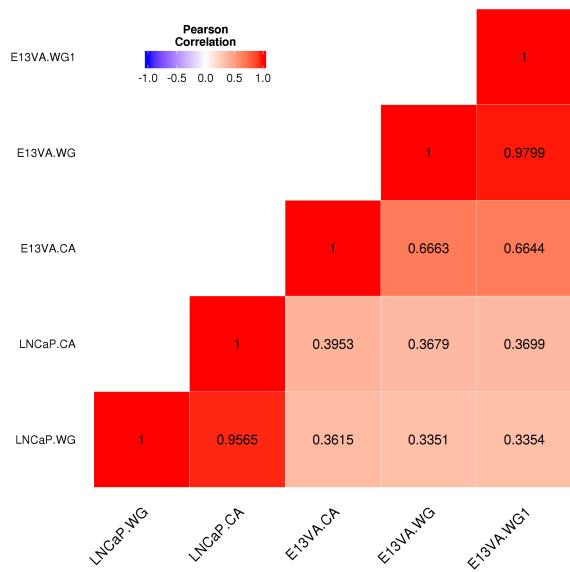
Discussion

- Correlation between DNA methylation profile of BiscaptureSeq vs WGBS is quite high (~0.8) in these samples
 - > BiscaptureSeq give the high similar DNA methylation profile as WGBS at targeted regions
- Correlation between DNA methylation profile of BiscaptureSeq vs WGBS in LNCaP is higher than in E13VA. It can be explained:
 - LNCaP WGBS is very high coverage ~26 while E13VA is ~21
 - These target regions are DMRs in E13VA, but not necessary in LNCaP, and we also know that DMRs are very polymorphism in DNA methylation patterns (*). Therefore, the fluctuation of DNA methylation in these DMRs are occurred.
 - LNCaP, a cell line, then it should be less heterogeneity comparing to primary sample as E13VA. Moreover, as Steve hypothesise that these tissues are mixed of different cell types.

(*) Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. Landan G, Cohen NM, Mukamel Z, Bar A, Molchadsky A, Brosh R, Horn-Saban S, Zalcenstein DA, Goldfinger N, Zundelevich A, Gal-Yam EN, Rotter V, Tanay A. *Nature Genetics* (2012)

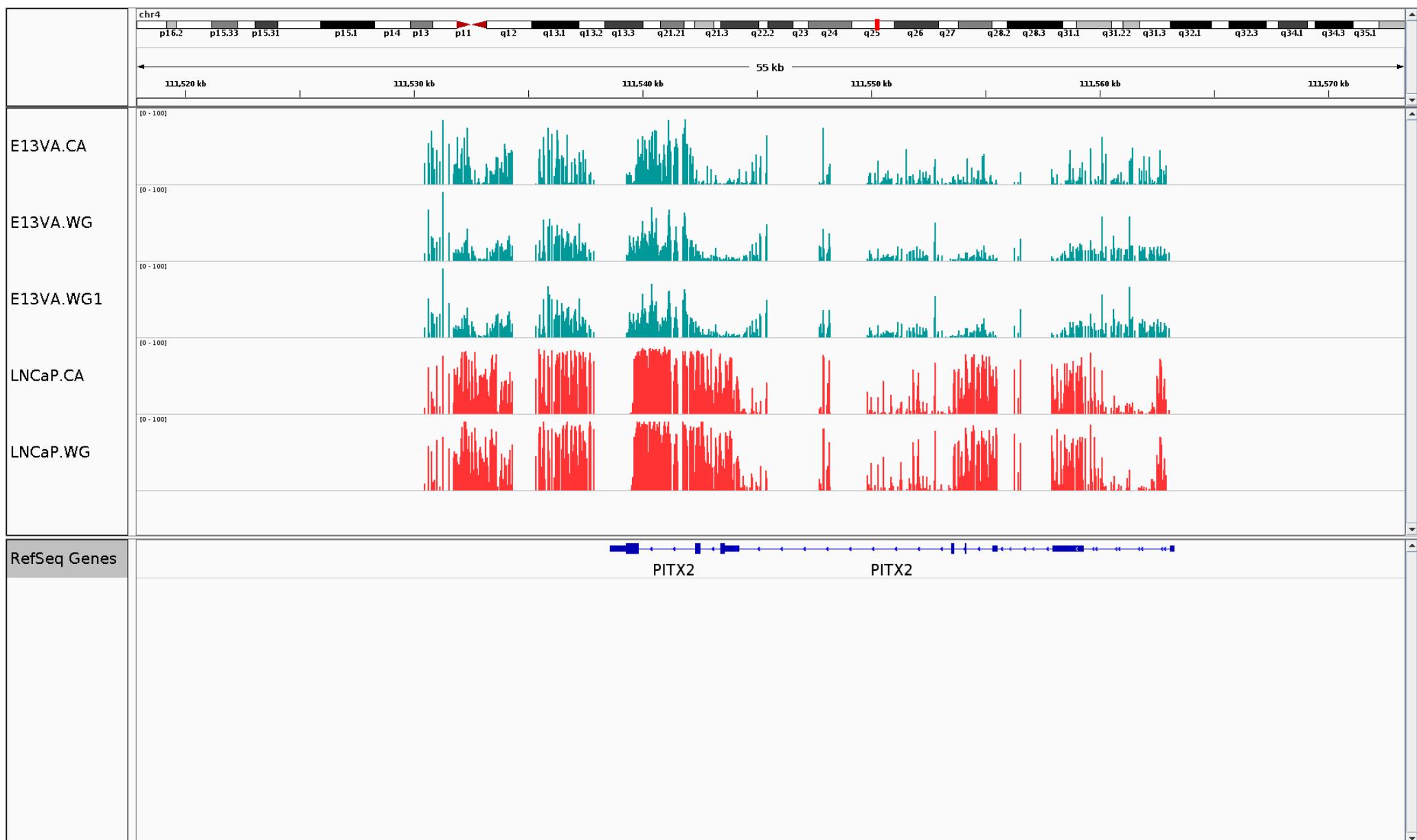
2. How are DNA methylation status of some important genes such as PITX2 and NKX2-5?

DNA methylation status of PITX2 in BisCaptureSeq and WGBS



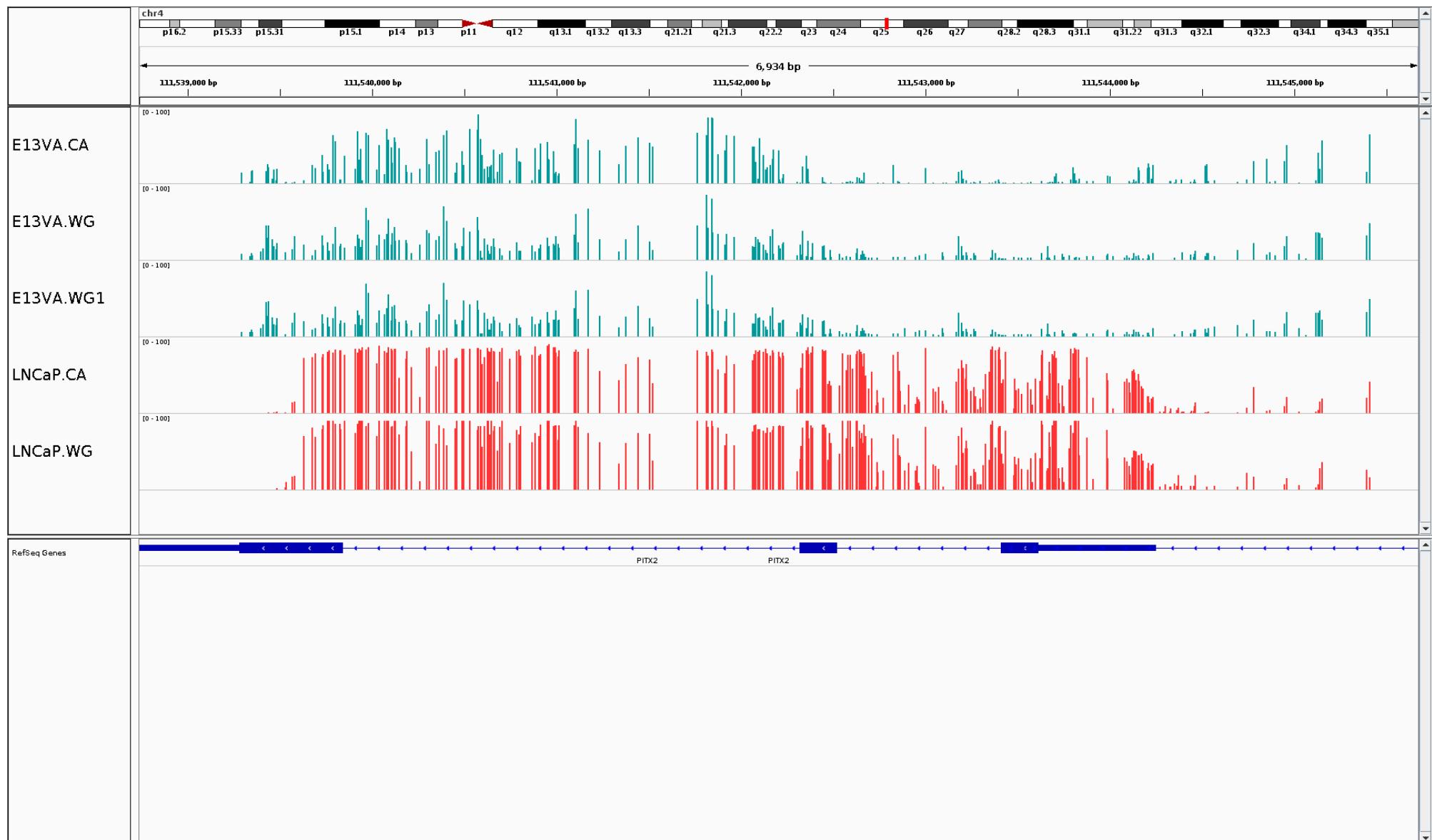
2. How are DNA methylation status of some important genes such as PITX2 and NKX2-5?

DNA methylation status of PITX2 in BisCaptureSeq and WGBS



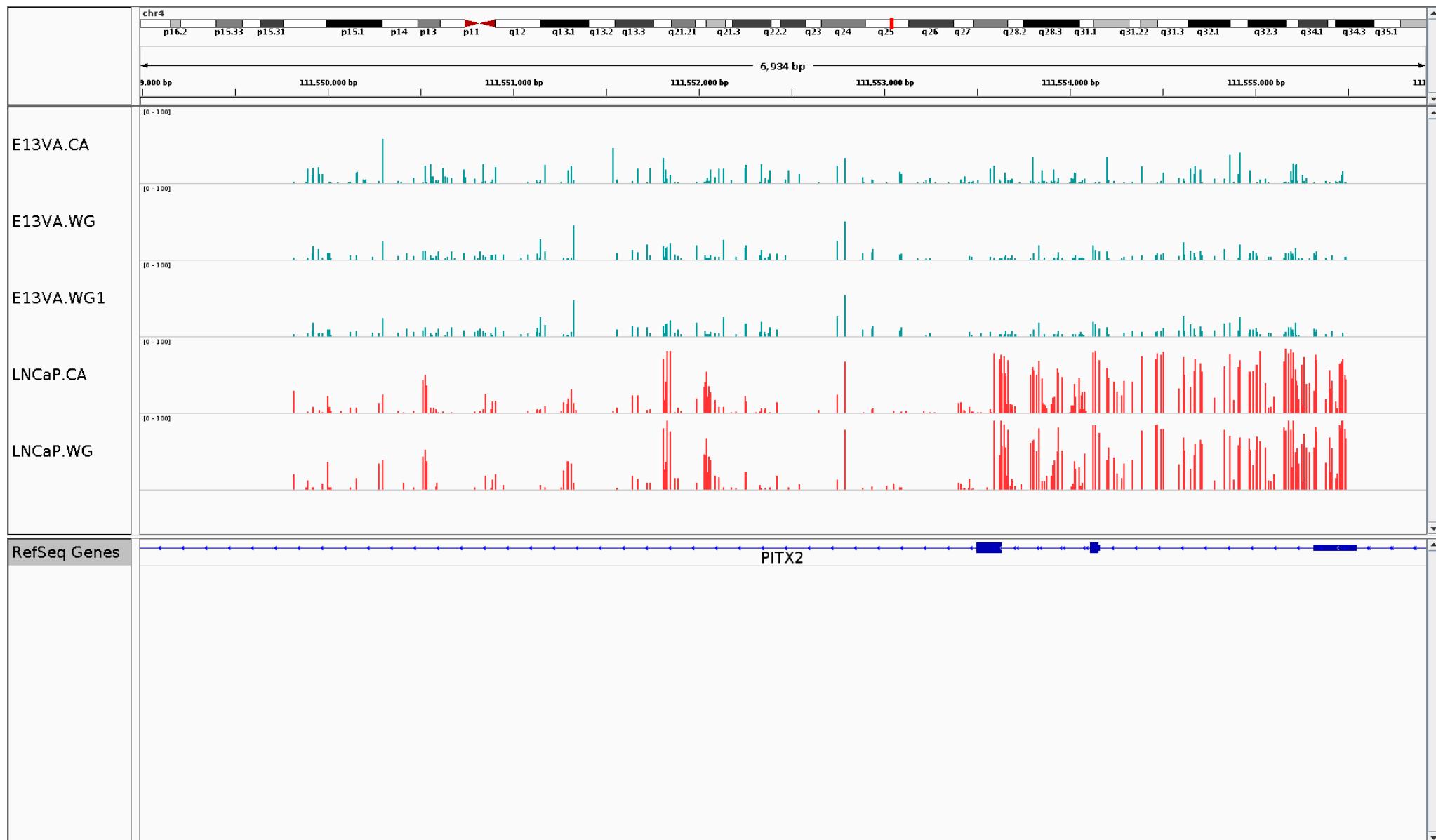
2. How are DNA methylation status of some important genes such as PITX2 and NKX2-5?

DNA methylation status of PITX2 in BisCaptureSeq and WGBS



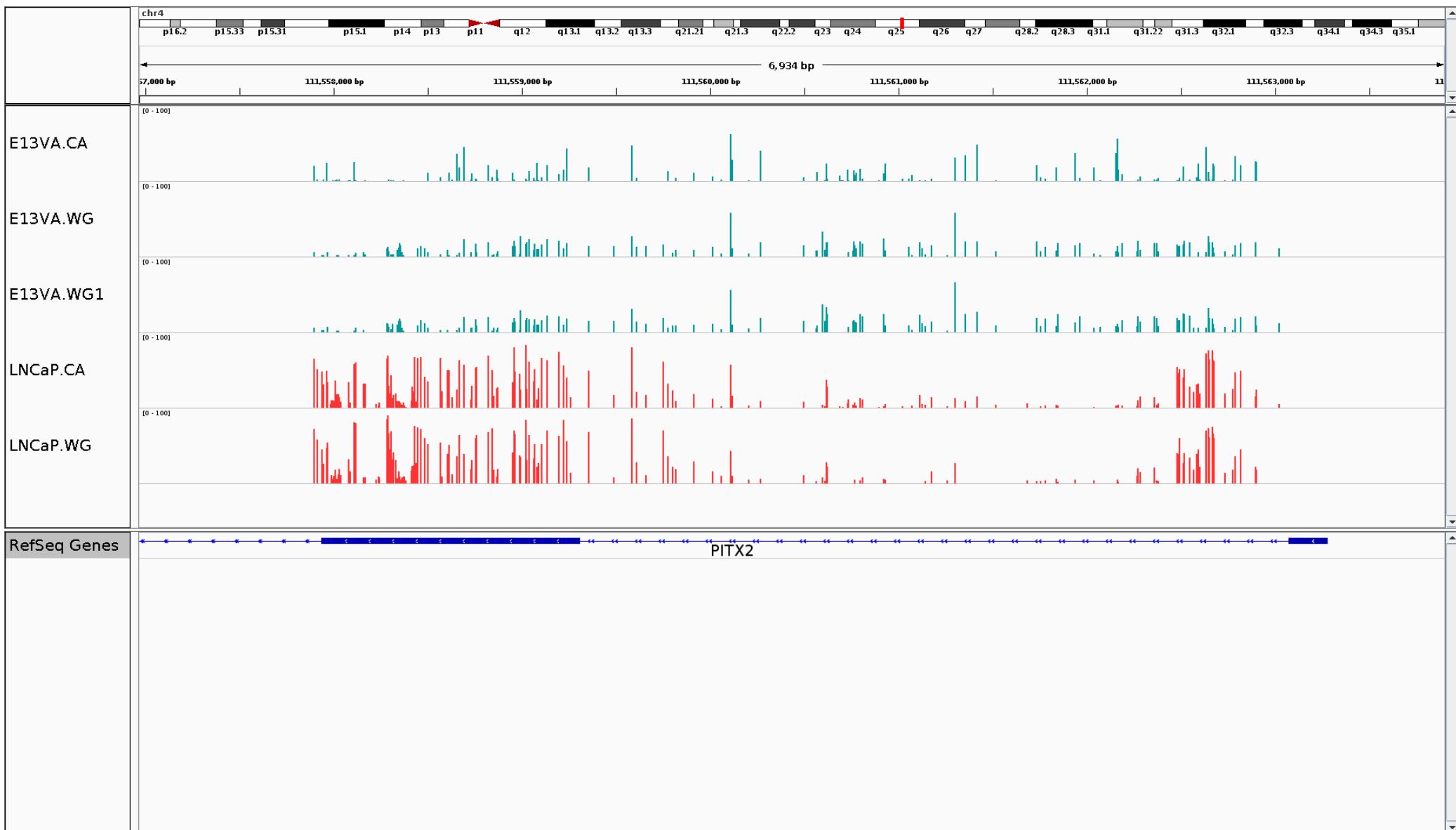
2. How are DNA methylation status of some important genes such as PITX2 and NKX2-5?

DNA methylation status of PITX2 in BisCaptureSeq and WGBS



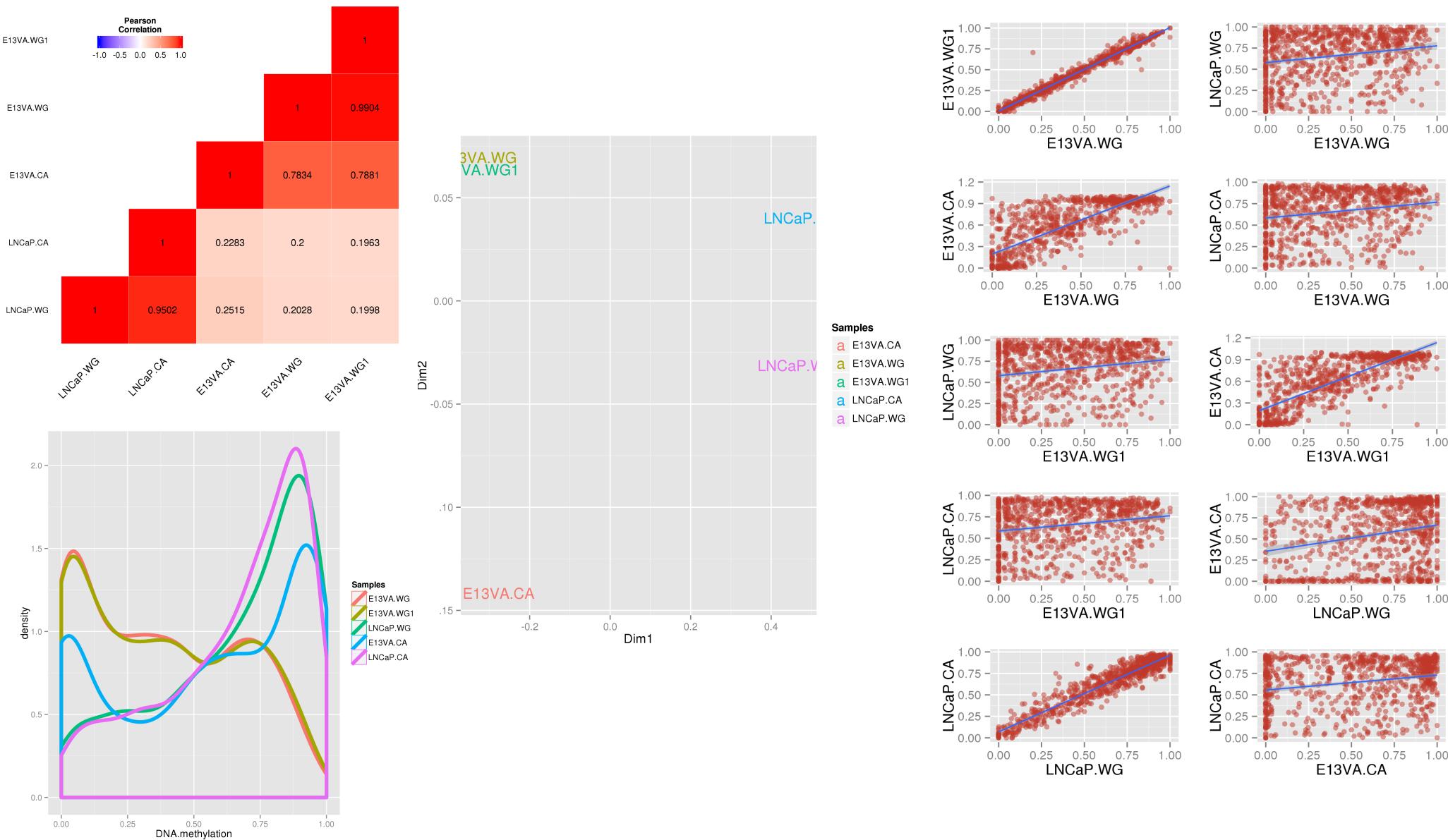
2. How are DNA methylation status of some important genes such as PITX2 and NKX2-5?

DNA methylation status of PITX2 in BisCaptureSeq and WGBS



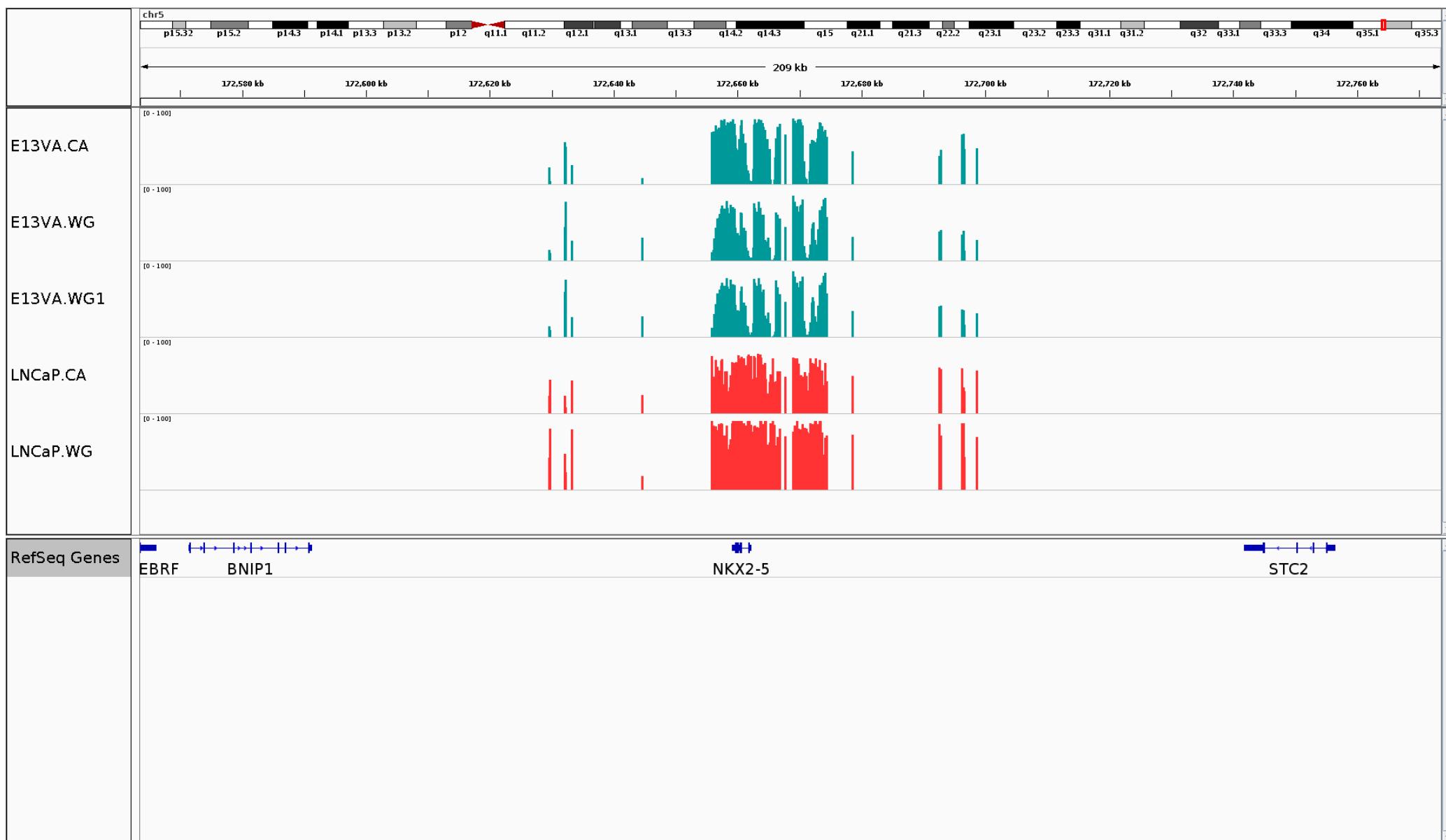
2. How are DNA methylation status of some important genes such as PITX2 and NKX2-5?

DNA methylation status of NKX2-5 in BisCaptureSeq and WGBS



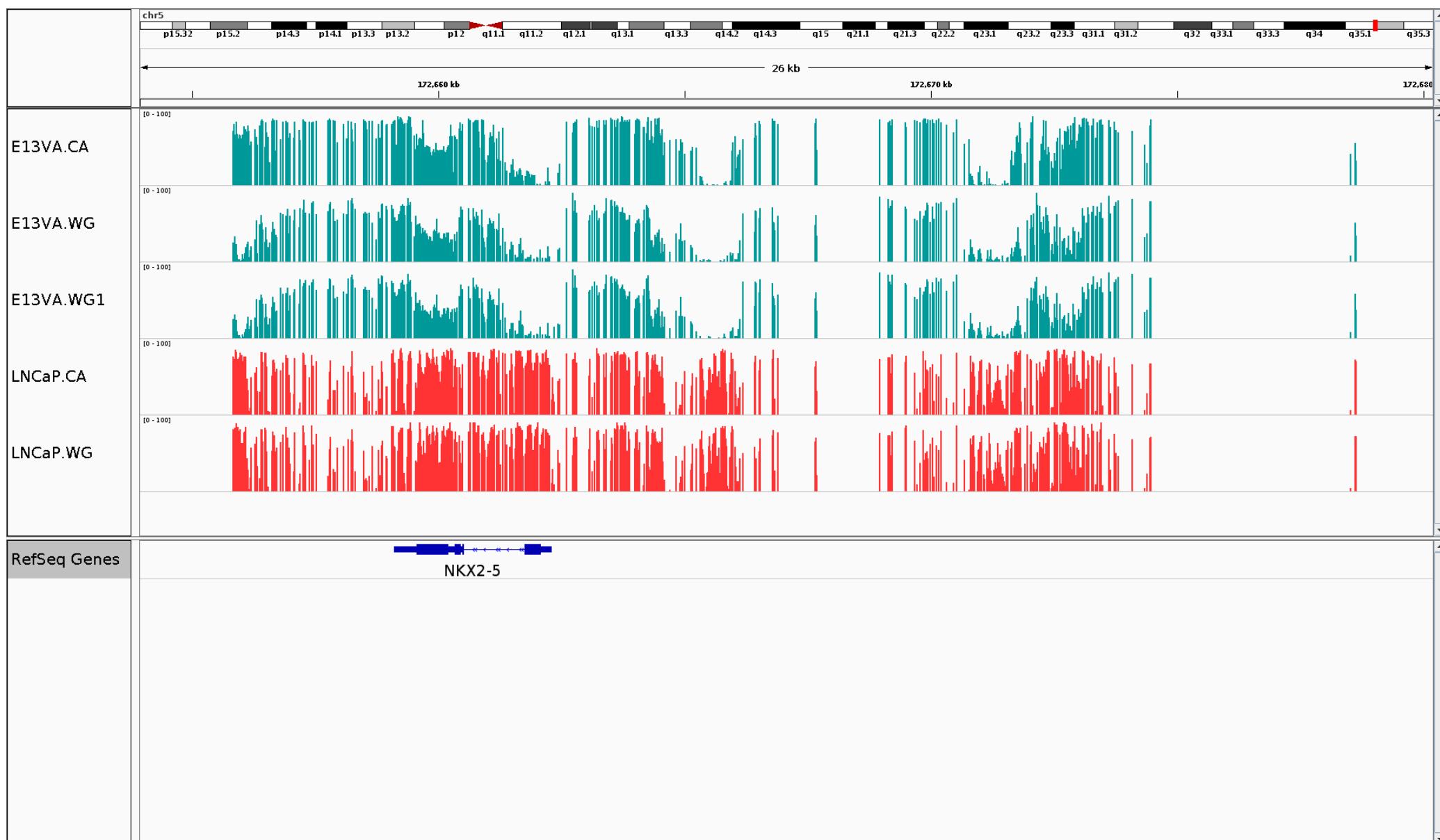
2. How are DNA methylation status of some important genes such as PITX2 and NKX2-5?

DNA methylation status of NKX2-5 in BisCaptureSeq and WGBS



2. How are DNA methylation status of some important genes such as PITX2 and NKX2-5?

DNA methylation status of NKX2-5 in BisCaptureSeq and WGBS



2. How are DNA methylation status of some important genes such as PITX2 and NKX2-5?

Discussion

- Correlation between DNA methylation profile of PITX2 and NKX2-5 in BiscaptureSeq vs WGBS is as high as whole profile in these samples respectively
--> Again, BiscaptureSeq give the high similar individual gene DNA methylation profile as WGBS at targeted regions
- The IGV visualization gives my impression that DNA methylation profile of PITX2 and NKX2-5 resemble very well in both BiscaptureSeq vs WGBS. I noticed that the difference between BiscaptureSeq vs WGBS in E13VA is in the same trend, BiscaptureSeq gives a strong signal. In this case, I would say one single number such as Pearson correlation coefficient fails to capture the picture of DNA methylation profile.

Conclusion

BisCaptureSeq produces DNA methylation profiles as similar as WGBS
DNA methylation patterns needed further investigated