Targeted Nanopore sequencing for MGMT promoter methylation evaluation

Preliminary data analysis

Skarphedinn Halldorsson

Compiled on 01 desember, 2023

0.1 Figures

0.1.1 MGMT Promoter CpG island and methods of evaluation

Table has been reformated for Latex

Table 1: Overview of samples included in this study

	DenStem	Radium	Rapid-CNS	WGS	Total
Astrocytoma	3	1	3	2	9
Astrocytoma HG	0	4	4	1	9
ATRT_SHH	0	0	1	0	1
Ganglioglioma	0	1	0	0	1
Glioblastoma	13	28	49	13	103
HPC	0	1	0	0	1
LGG_PA	0	0	4	0	4
Medulloblastoma	0	1	0	0	1
Meningioma	0	12	0	0	12
Metastasis	0	7	0	0	7
Oligodendroglioma	0	2	6	2	10
PCNSL	0	2	0	0	2
PXA	0	2	0	0	2
Unknown	0	3	0	0	3
Total	16	64	67	18	165

0.1.2 Overview of samples included in study

Also simplified version

0.2 Data acquisition

0.3 Direct comparison of Nanopore sequencing and Pyrosequencing

Comparison of individual CpGs as measured by nanopore and pyrosequencing

Group average as measured by nanopore and pyrosequencing

Use 4 CpG sites (methylation cutoff 10%) to classify all samples, compared to known status
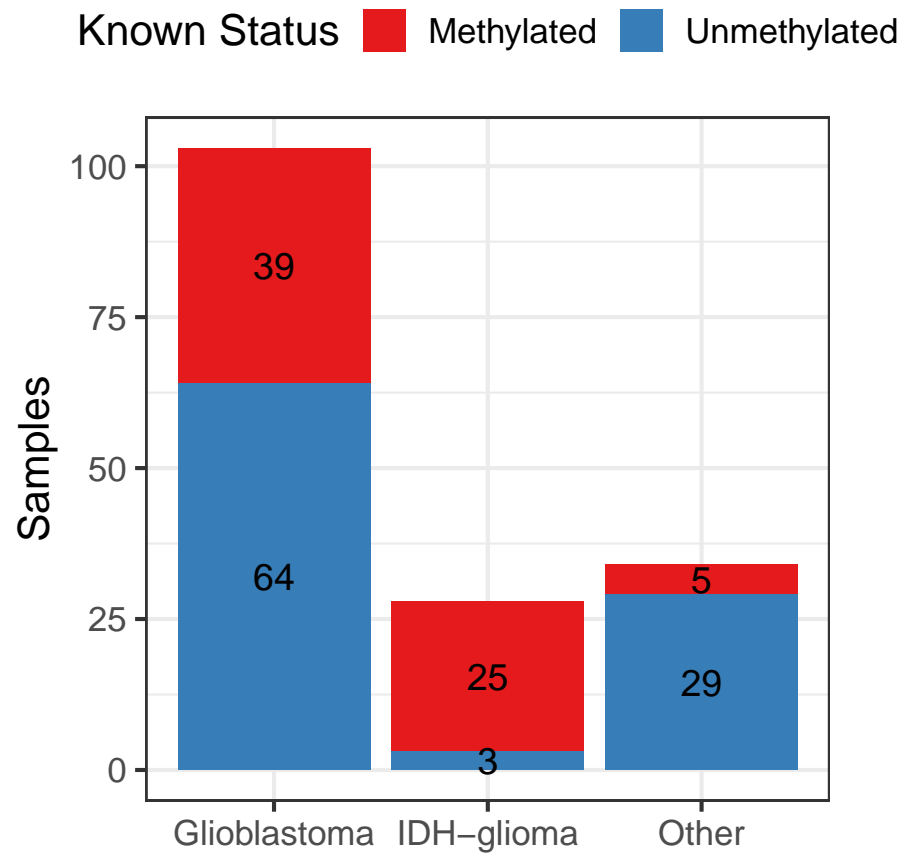
```
##
## Call:
```

Figure 1: Distribution of methylated versus unmethylated samples according to known status prior to nanopore sequencing
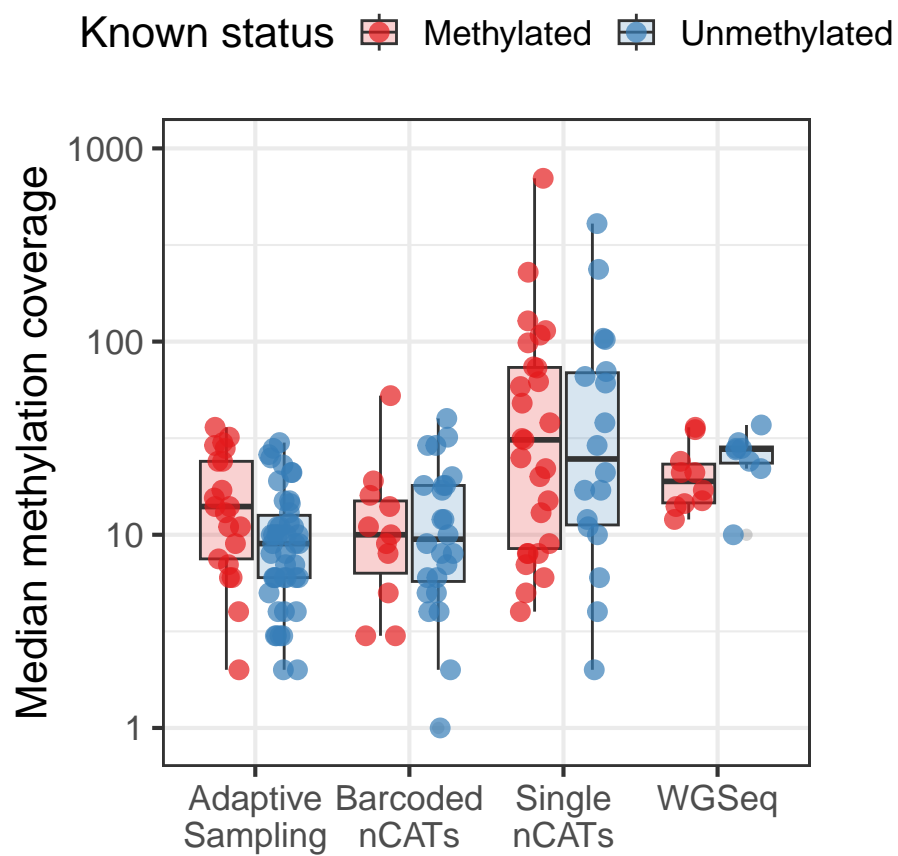
Figure 2: Sequencing depth of samples by method of acquisition, group median represented by red diamond. No bias in sequence depth observed between methylated and unmethylated samples.
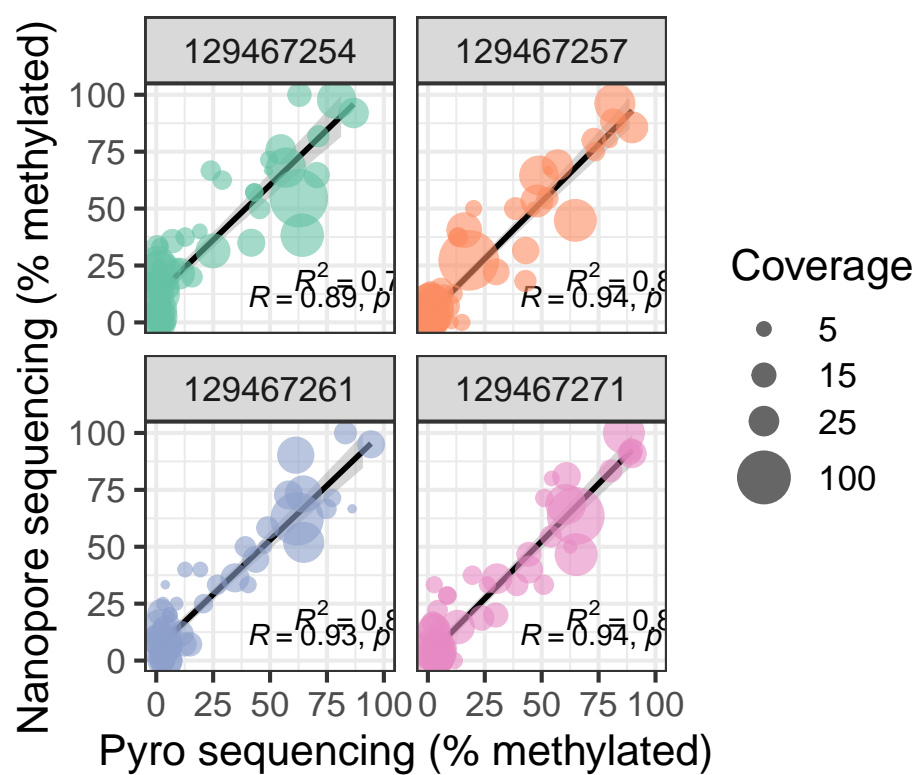
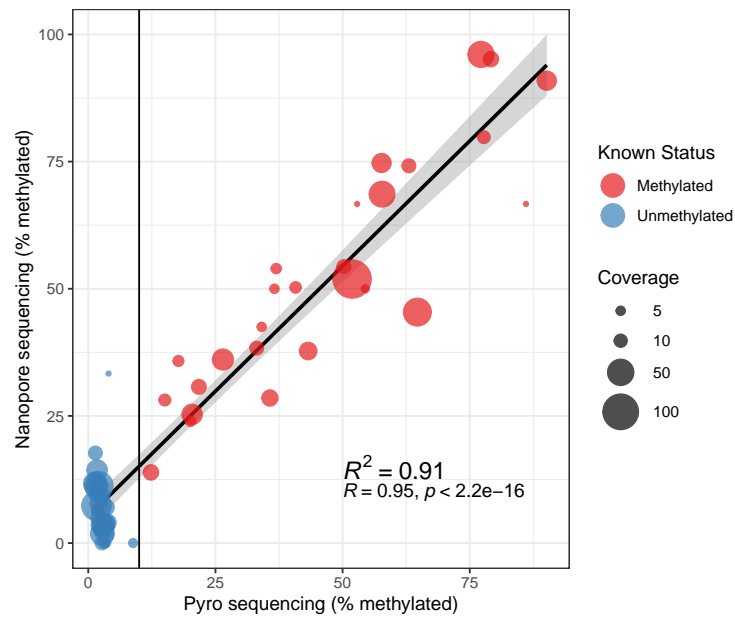Figure 3: Comparison of individual CpGs within exon 1

Figure 4: Comparison of nanopore sequencing and Pyrosequencing results of 4 CpGs in exon 1 of the MGMT promoter. Plotted values are average methylation of 4 CpGs. Black horizontal and vertical lines mark the 9 percent cut-off value between methylated and unmethylated samples, as determined by pyrosequencing.

```
## glm(formula = P ~ Average_Nano, family = "binomial", data = Train)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -5.59971    1.56989  -3.567 0.000361 ***
## Average_Nano  0.25579    0.07427   3.444 0.000573 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 84.915  on 61  degrees of freedom
## Residual deviance: 15.378  on 60  degrees of freedom
## AIC: 19.378
##
## Number of Fisher Scoring iterations: 8
```
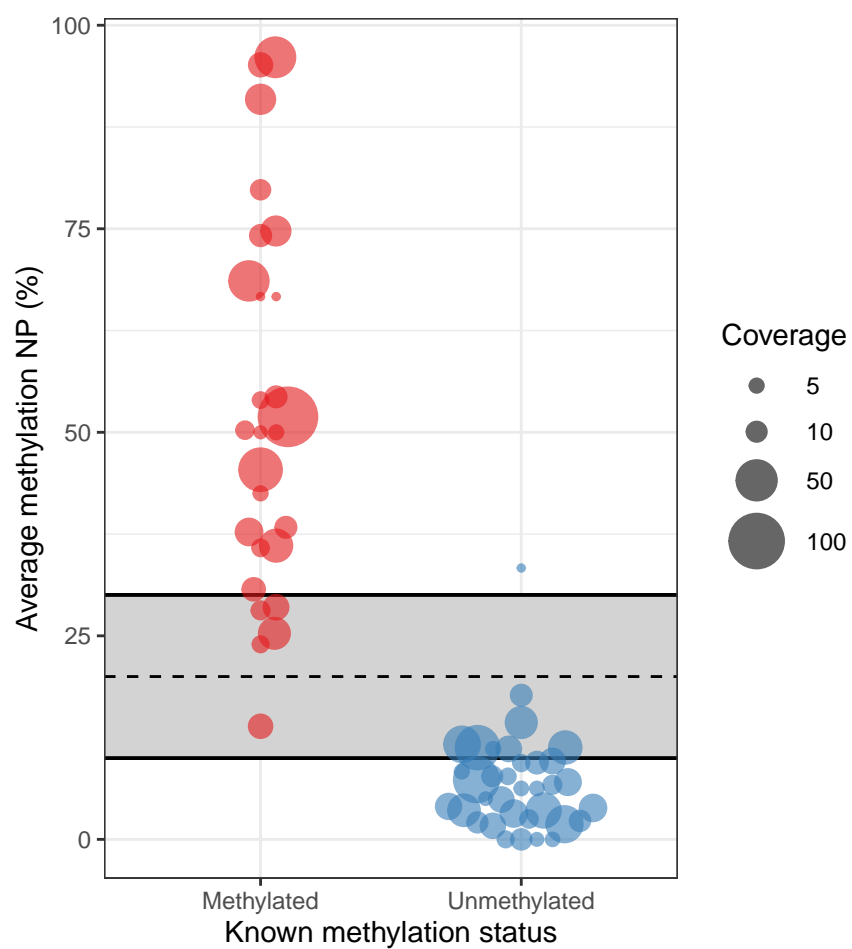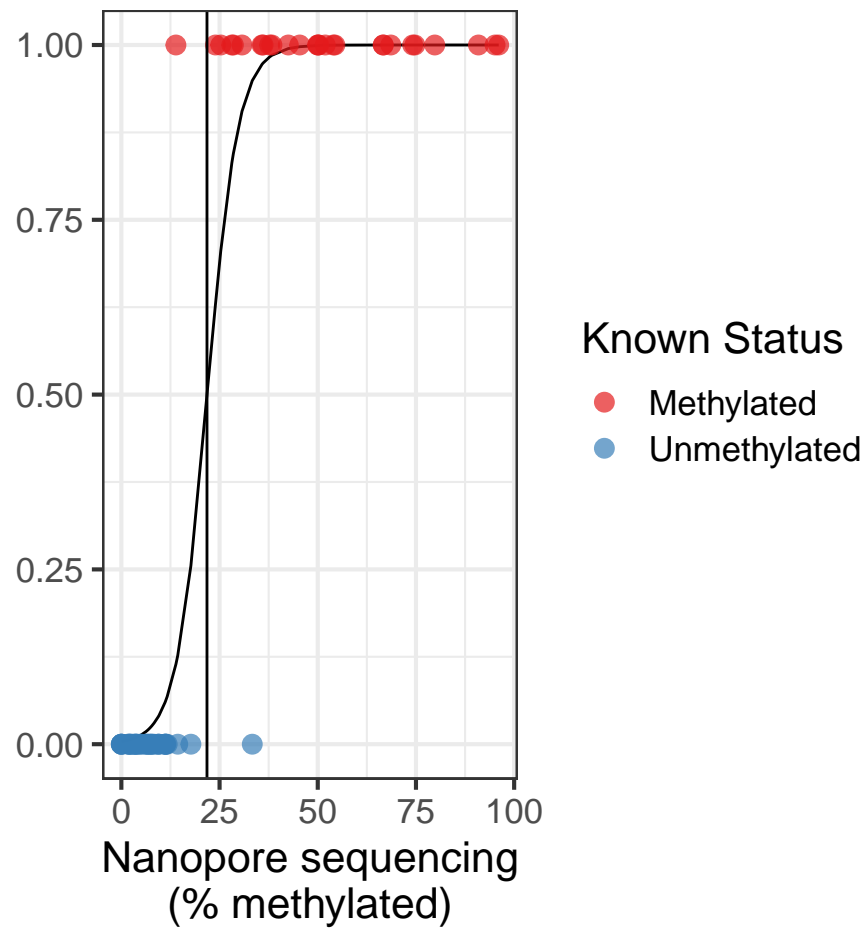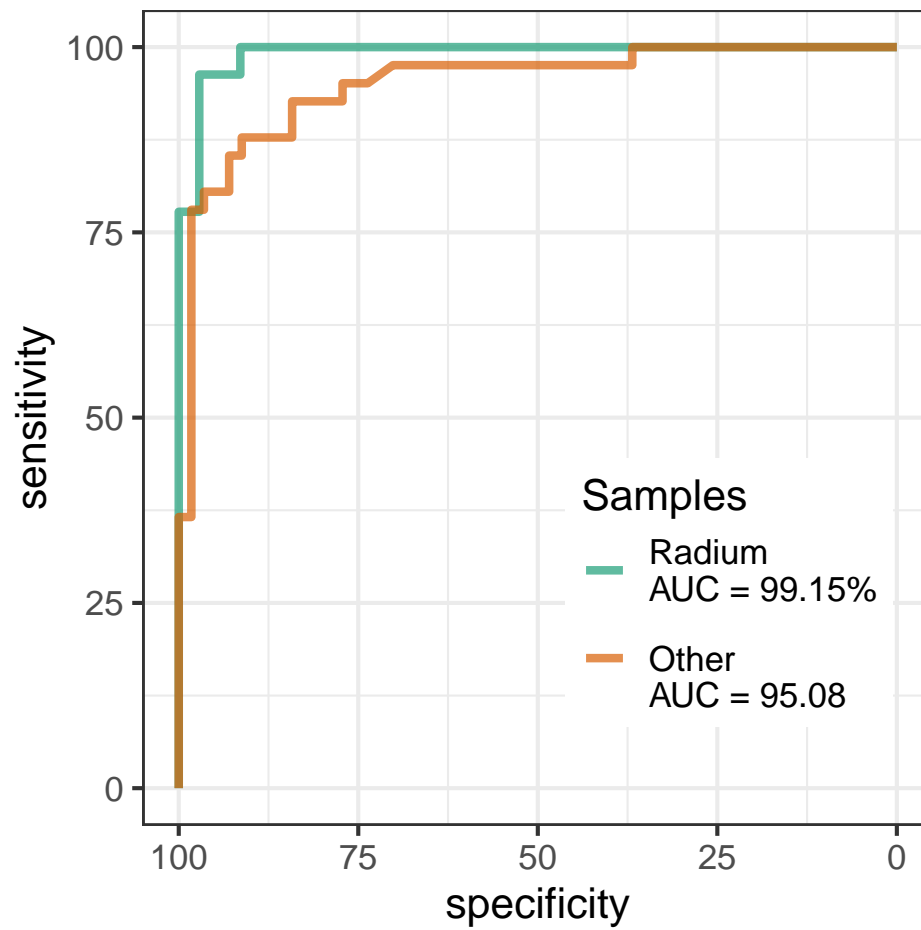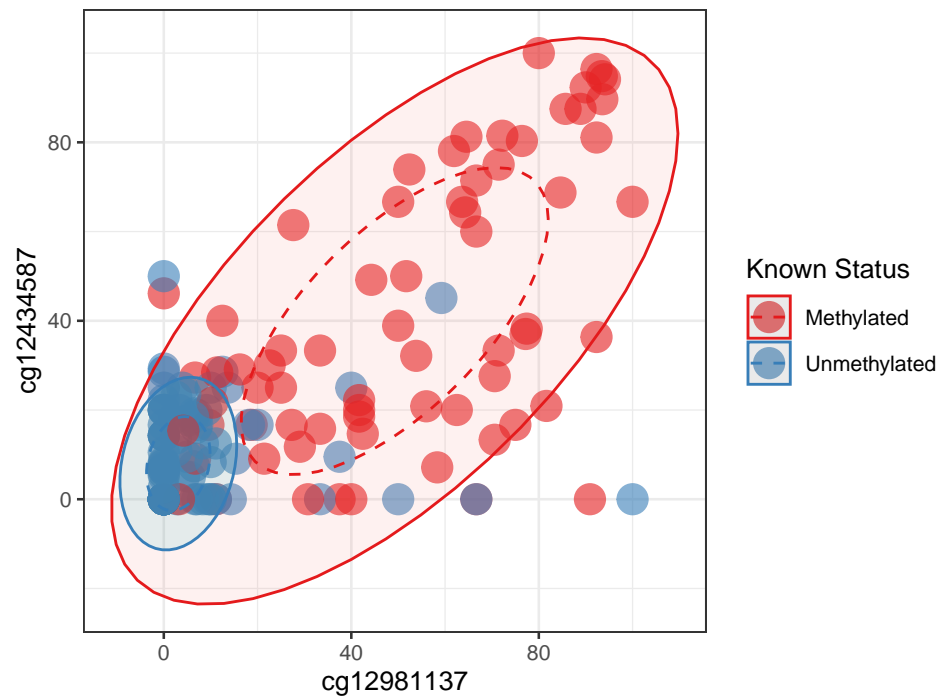
Figure 5: Average nanopore sequencing percentage of CpGs 76-79 agains the Pyrosequencing classification
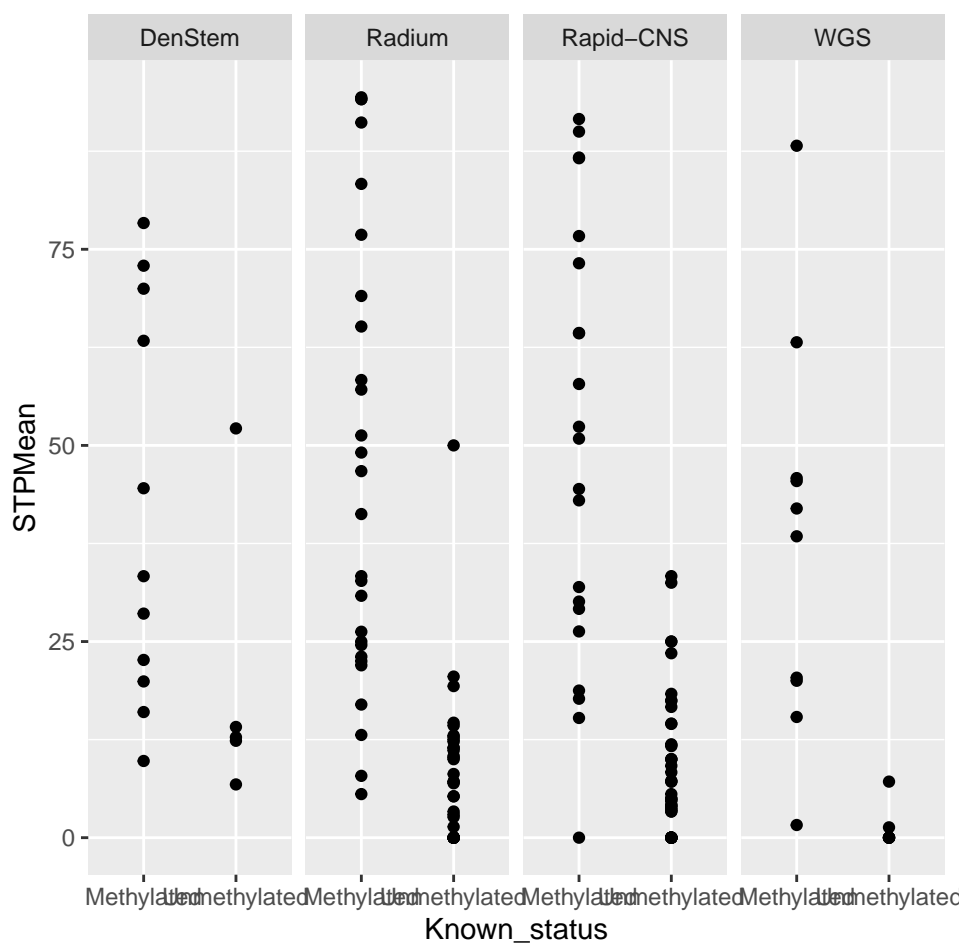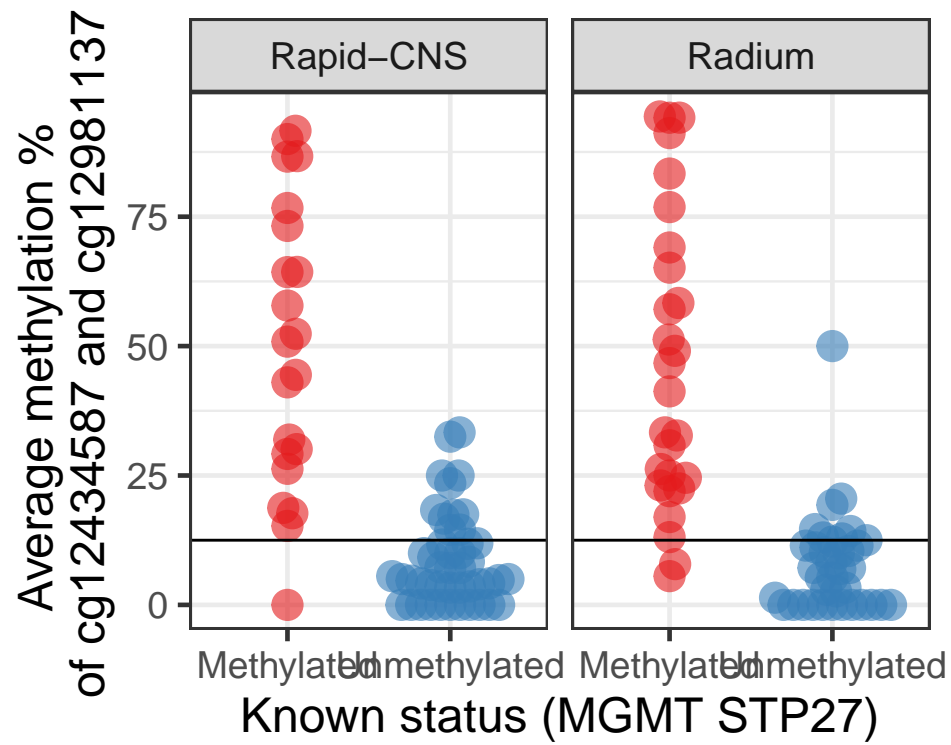
```
## Area under the curve: 99.15%
```
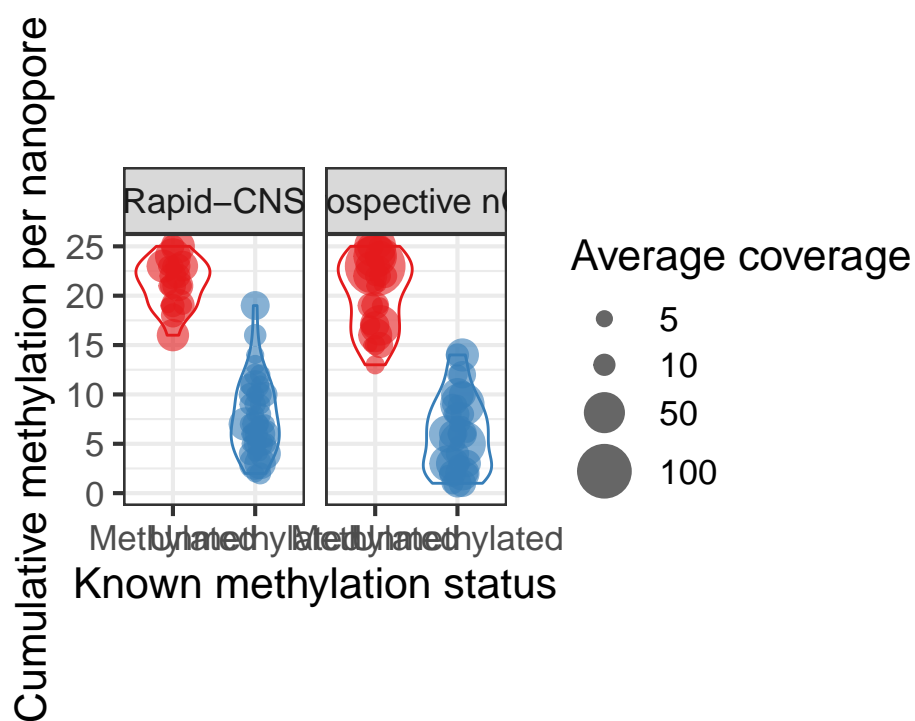
```
## Area under the curve: 95.08%
```

Figure 6: Cumulative methylation of MGMT promoter as analysed by nanopore sequencing agains the known classification of all samples
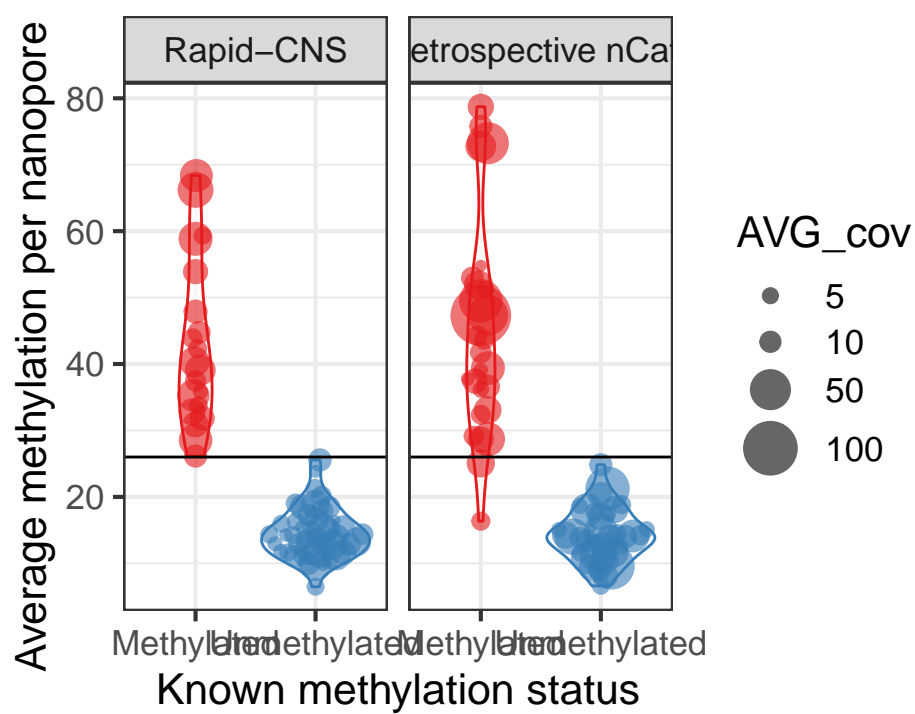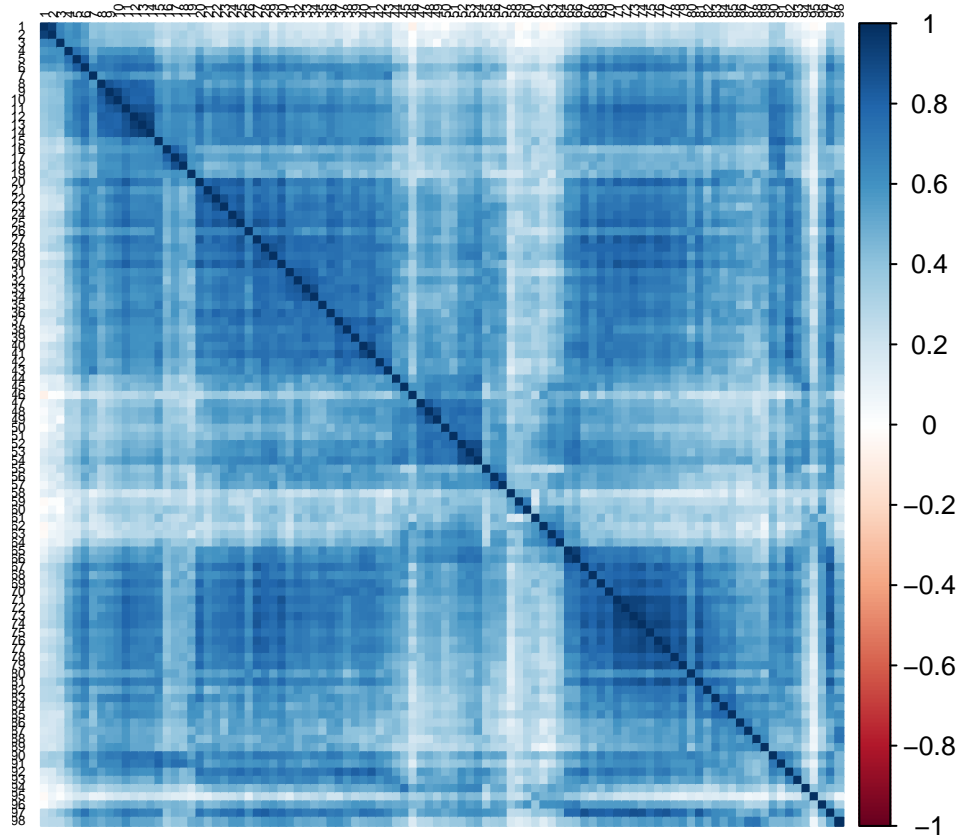
Figure 7: Cumulative methylation of MGMT promoter as analysed by nanopore sequencing agains the known classification of all samples

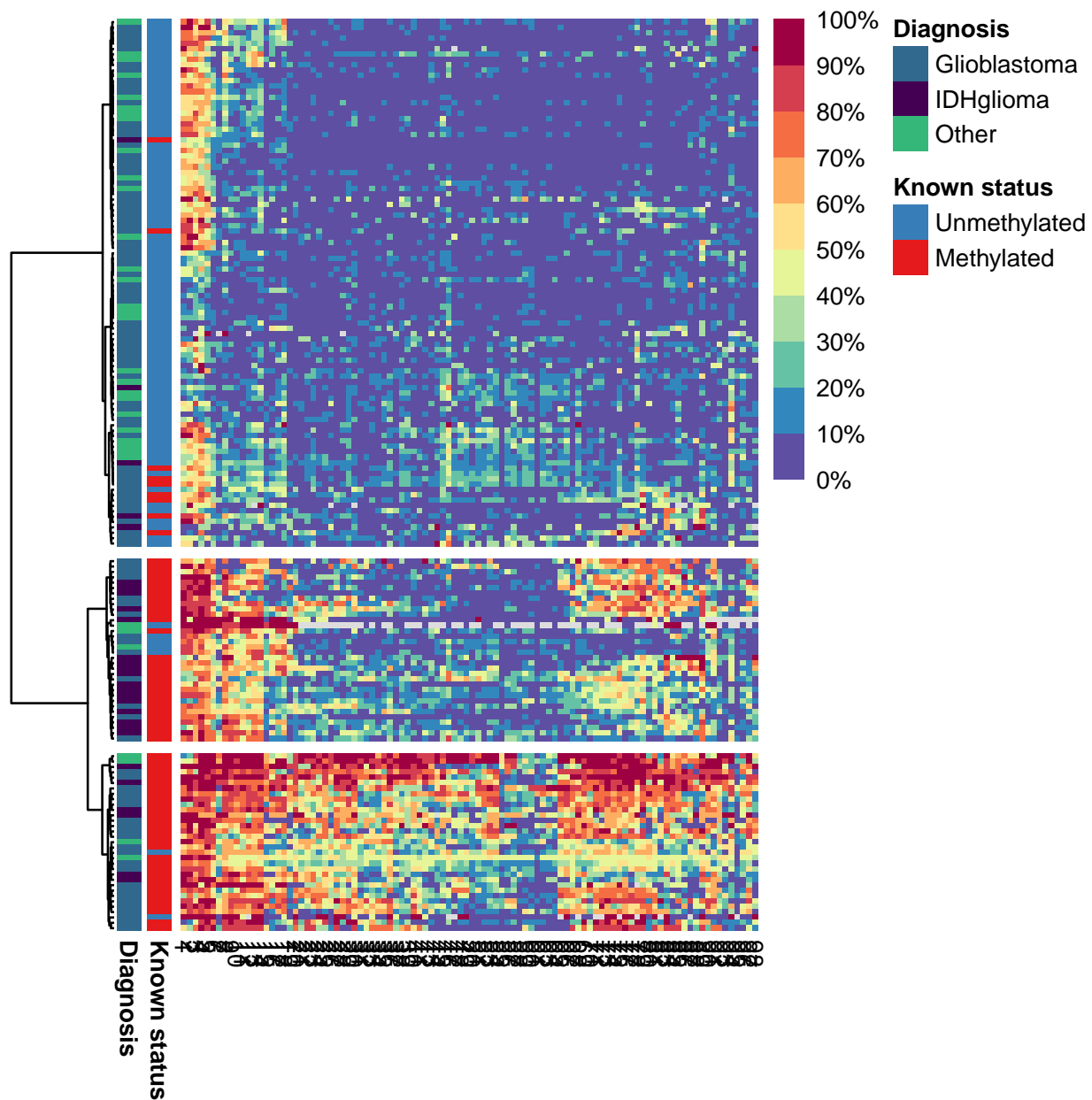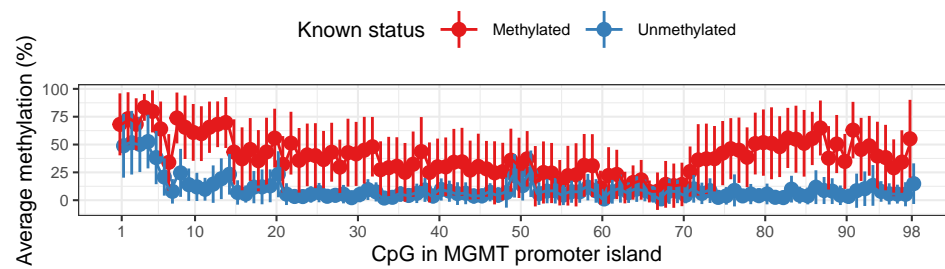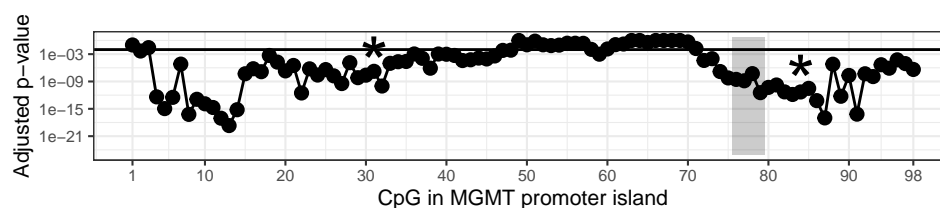## 0.4   Unsupervised clustering of nanopore results

Figure 8: Clustered heatmap of all samples based on nanopore sequencing of CpG island of the MGMT promoter

Unsupervised clustering of all the samples included in the study shows very clear separation of methylated and unmethylated samples (Figure 7). The samples previously defined as unmethylated have very low methylation throughout the CpG island apart from the first 5 CpGs that are often methylated. Samples previously defined as methylated have a gradient of methylation which tends to show highest methylation values towards either end of the CpG island. Only six samples samples do not cluster according to their previously determined methylation status. Four of these samples belong to the DenStem cohort that is not directly comparable, as previously mentioned. A single sample from the Radium cohort was classified as methylated by pyrosequencing but clusters with the unmethylated samples. This sample is interesting as it has robust methylation in the fist exon but very low methylation elsewhere in the CpG island.

We can also look at methylation patterns specifically in the glioblastoma samples (Table 1).

In Figure 8 we see that all of the Rapid-CNS GBM samples fall within their previously determined classes and only one of the Radium samples is "mis-classified". We also see there are clusters within both the methylated and unmethylated samples. These clusters can also be represented by mean methylation values within each cluster (Figure 9).

0.5 Survival analysis

{NEEDS UPDATE}Patient survival data is currently only available for 43 patients, including 37 GBM patients. Of these, 26 are primary GBMs. Comparison of survival of these 26 patients can be seen in figure 10. There is little difference between the results if patients are classified by pyrosequencing or nanopore sequencing, only two patients switch groups. We are expecting more data to strengthen these results. But they are promising: classification by nanopore sequencing is as good or better than pyrosequencing.
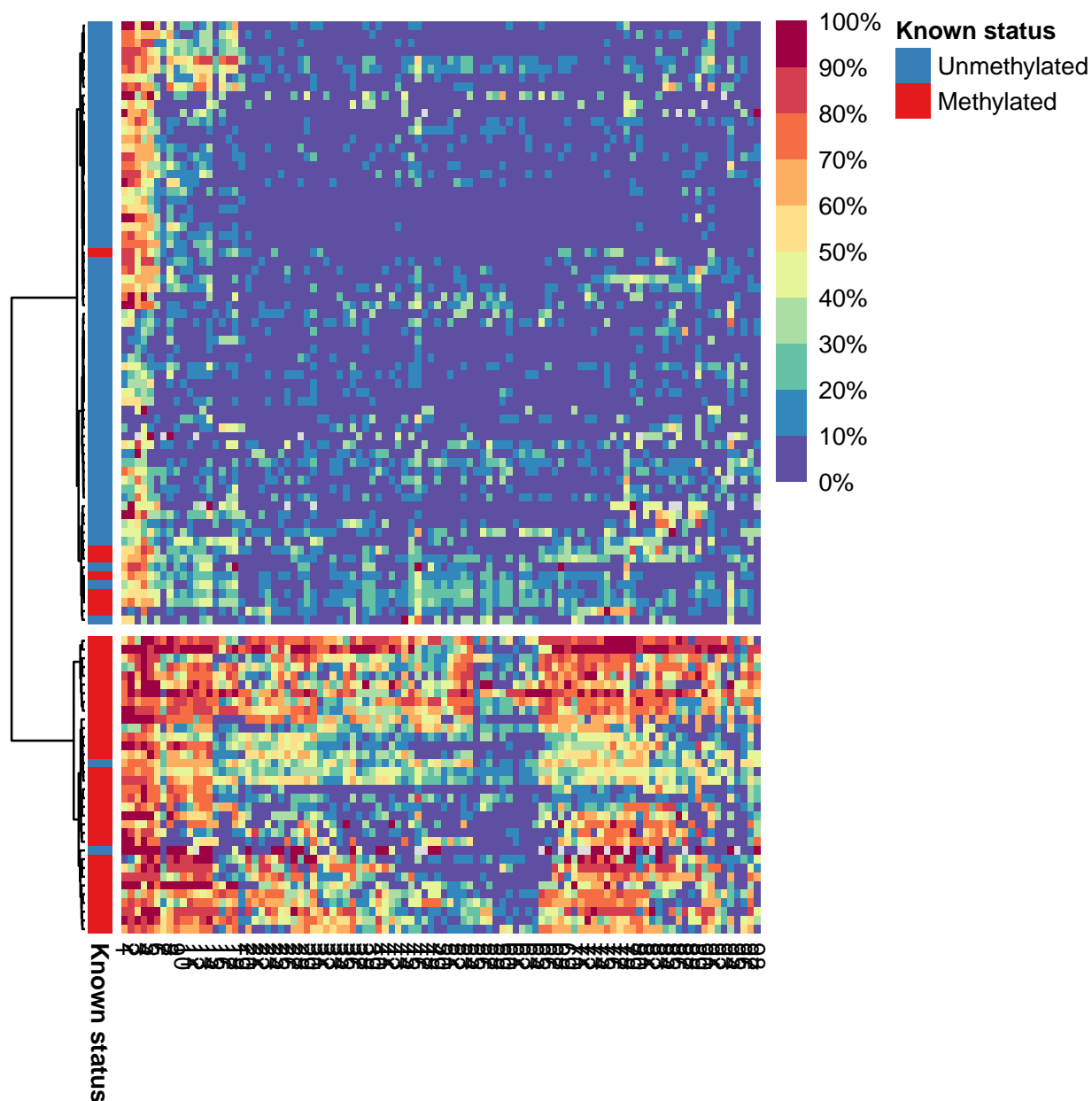
Figure 9: Heatmap showing unsupervised clustering of glioblastoma samples based on nanopore sequencing of the CpG island in the MGMT promoter
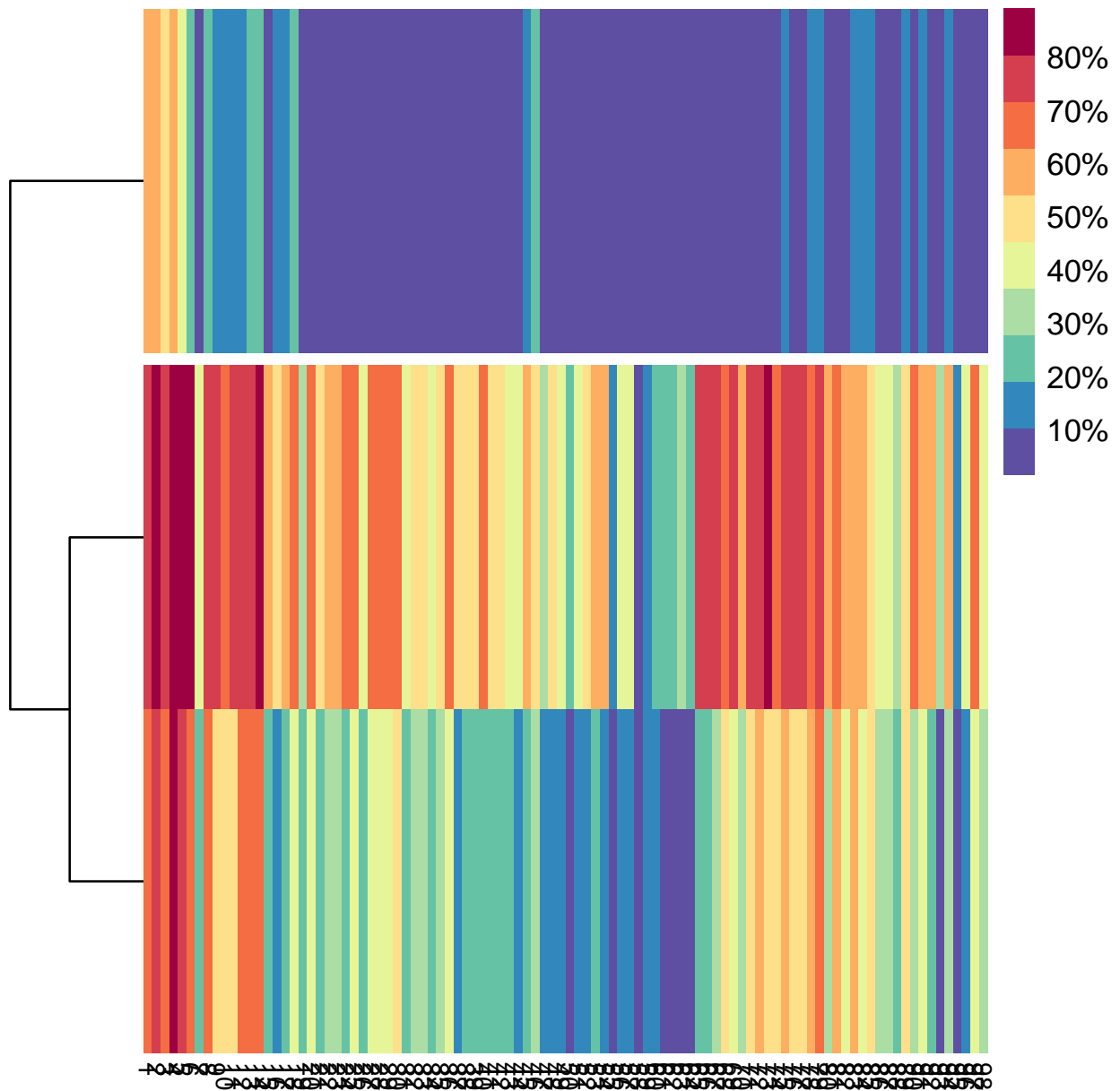
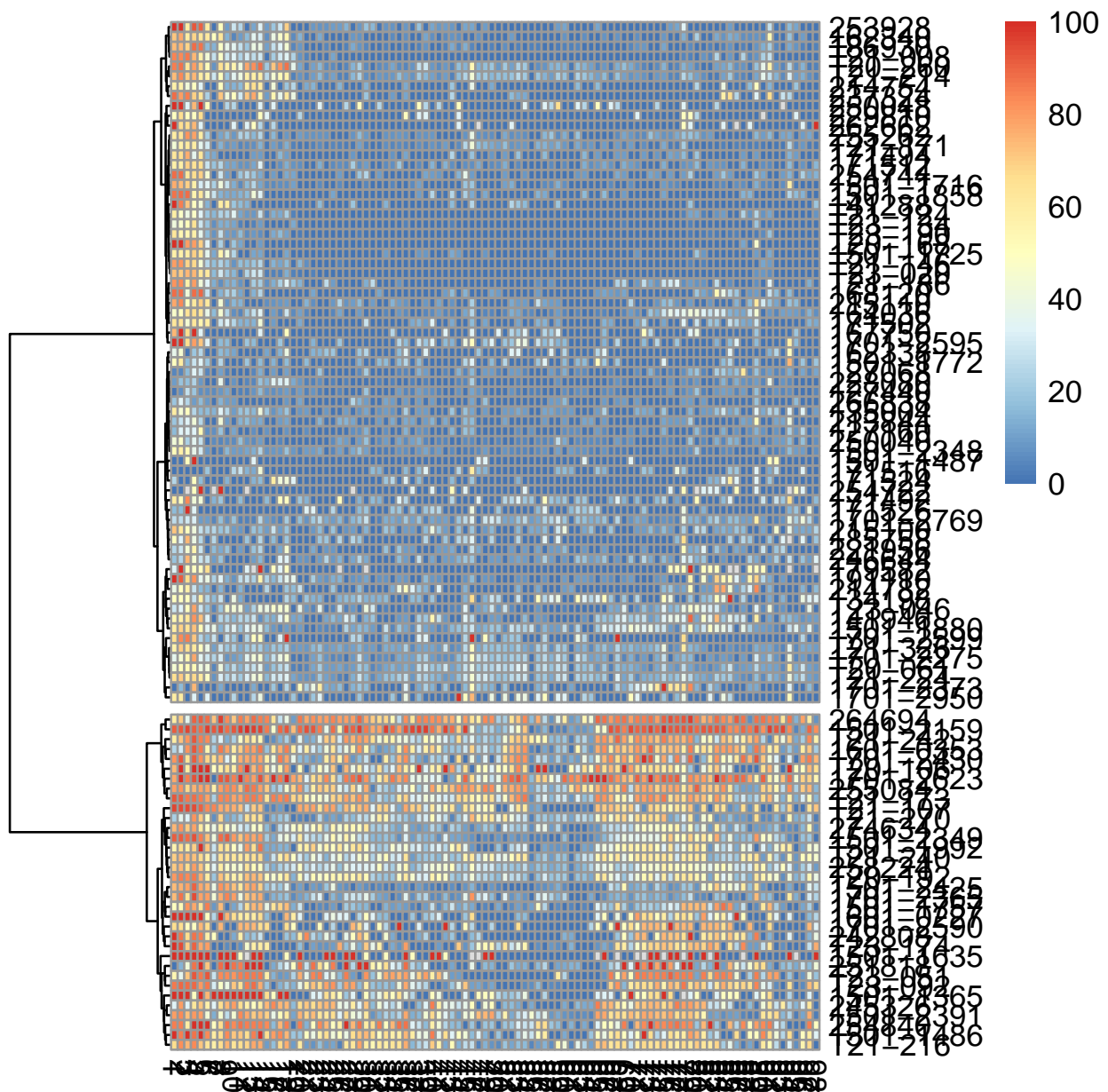Figure 10: K-means clustering of glioblastoma samples
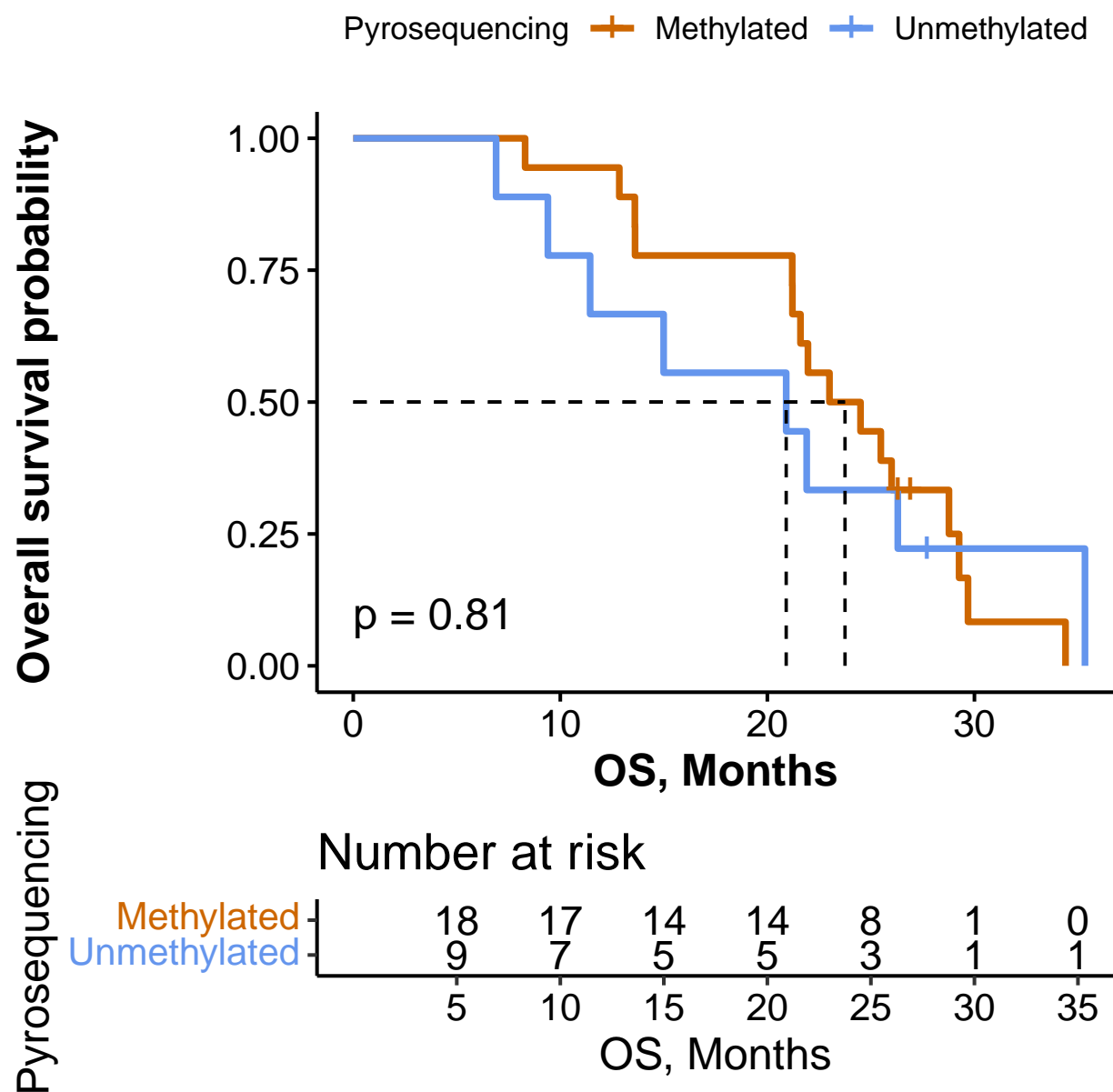
Figure 11: Kaplan-Meier Overall Survival curves
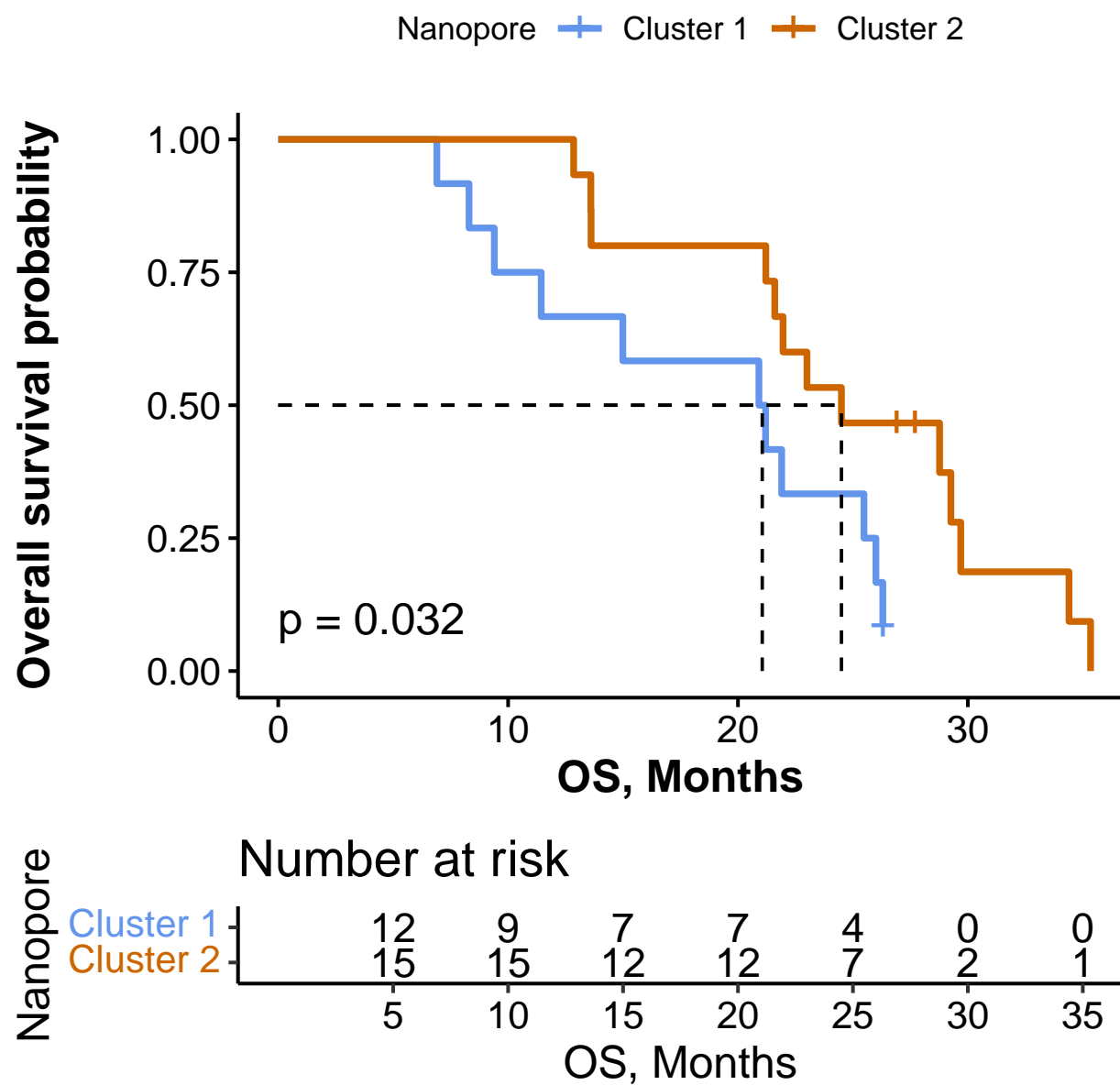
Figure 12: Kaplan-Meier Overall Survival curves

Figure 13: Kaplan-Meier Overall Survival curves

0.6   Discussion

We have a fairly extensive dataset of 153 samples, including 91 GBMs. I've not seen data with this many samples that looks at methylation within the whole CpG island of MGMT. We have the option of looking for CpG methylation further upstream and downstream than the 98 CpG presented here, at least 10 CpGs in both directions without losing any samples due to lack of coverage. By focusing on high depth samples we can look much further. I have not included that here but will be looking into it and will report if I find anything interesting.

We can conclude that nanopore sequencing of the MGMT promoter region performs as well or better than standard methods such as pyrosequencing. This is true for both cas9 targetted sequencing of the MGMT promoter and inclusion of the MGMT promoter into an adaptive sequencing panel. Distinct subgroups within both methylated and unmethylated samples are captured via nanopore sequencing, it will be very interesting to see if there is a difference in patient outcome between these clusters.