

DNA methylation analysis using Nanopore sequencing



Daria Meyer

TBI Winterseminar
14.02.2023

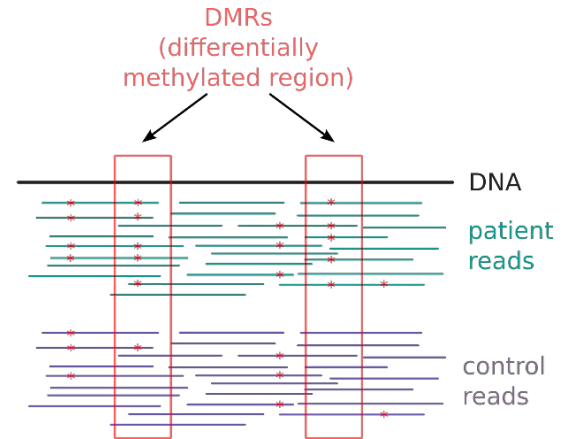
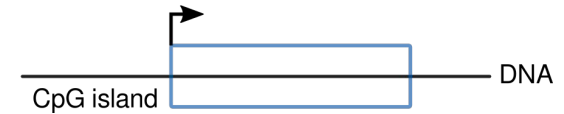
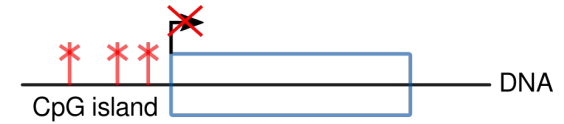
Detection of 5mC DNA methylation in human CpG islands

Assumption:

- Cancer specific DNA methylation (5mC) exist, especially within CpG islands

Goal:

- Detect these regions to create a PCR screening test based on these DNA methylation marker regions



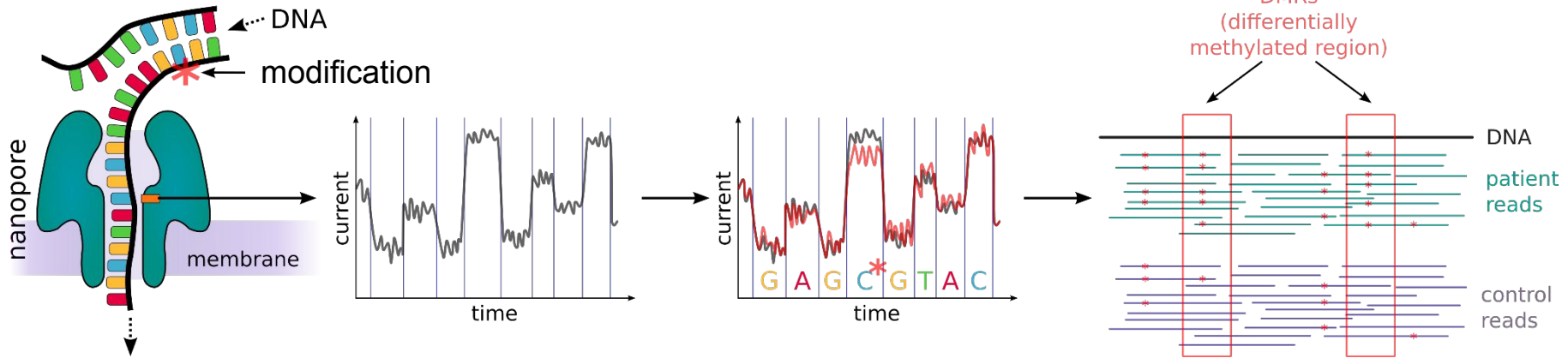
Detection of 5mC DNA methylation in human CpG islands

Assumption:

- Cancer specific DNA methylation (5mC) exist, especially within CpG islands

Goal:

- Detect these regions to create a PCR screening test based on these DNA methylation marker regions



Adaptive sampling

Some numbers:

- 3.1 Gb human genome
- ~20 Gb Output per Sequencing run on a MinION flow cell
- → 6-7 X coverage on the human genome

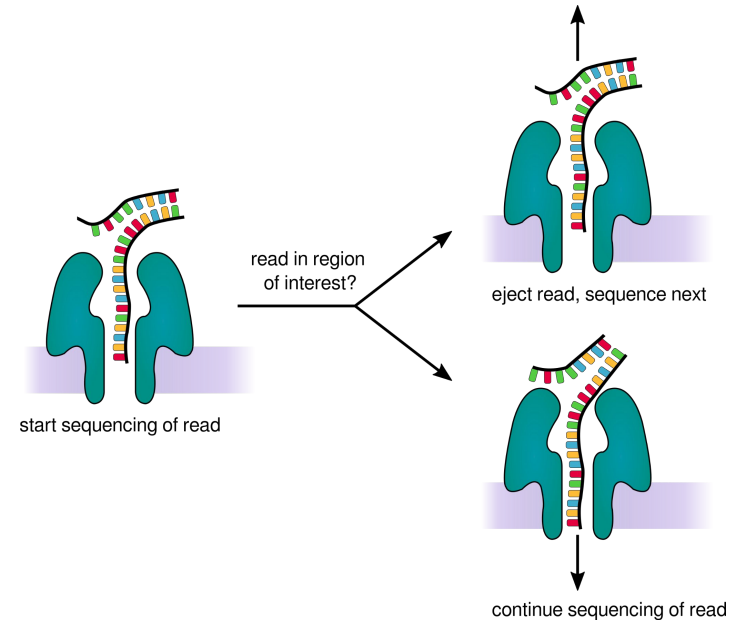
CpG islands:

- 31,448 CpG islands in the human genome
- 777 nt mean CpG island length

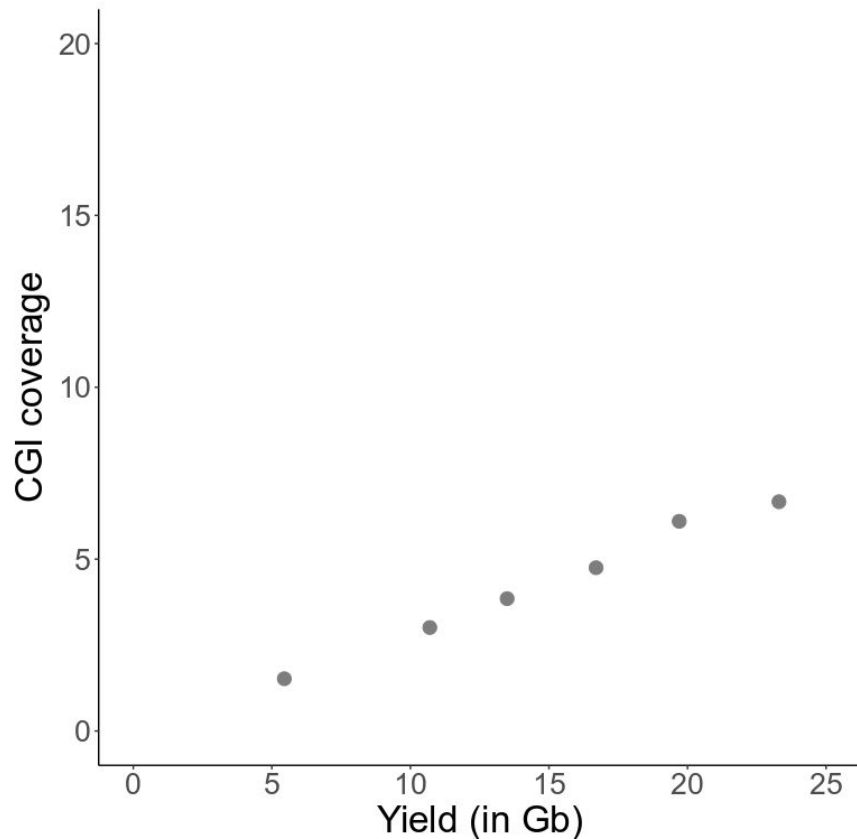


Adaptive Sampling:

- only DNA mapping to region of interest is sequenced further
- DNA is sequenced with 400 bp/s



Adaptive sampling



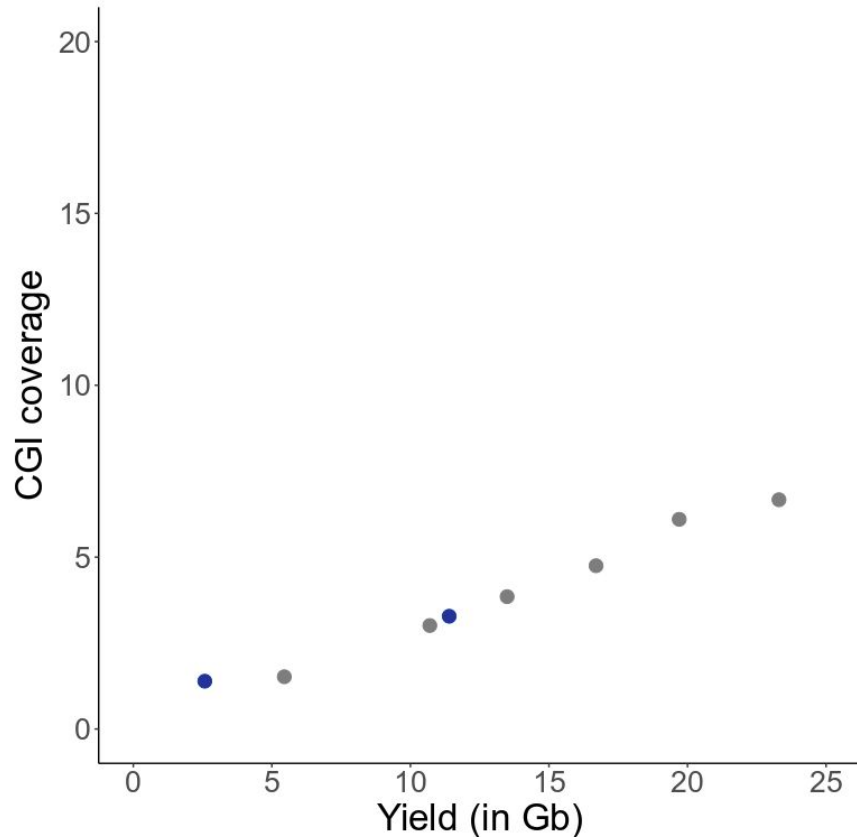
without adaptive sampling:

- CGI coverage correlates with yield

adaptive

● NA

Adaptive sampling



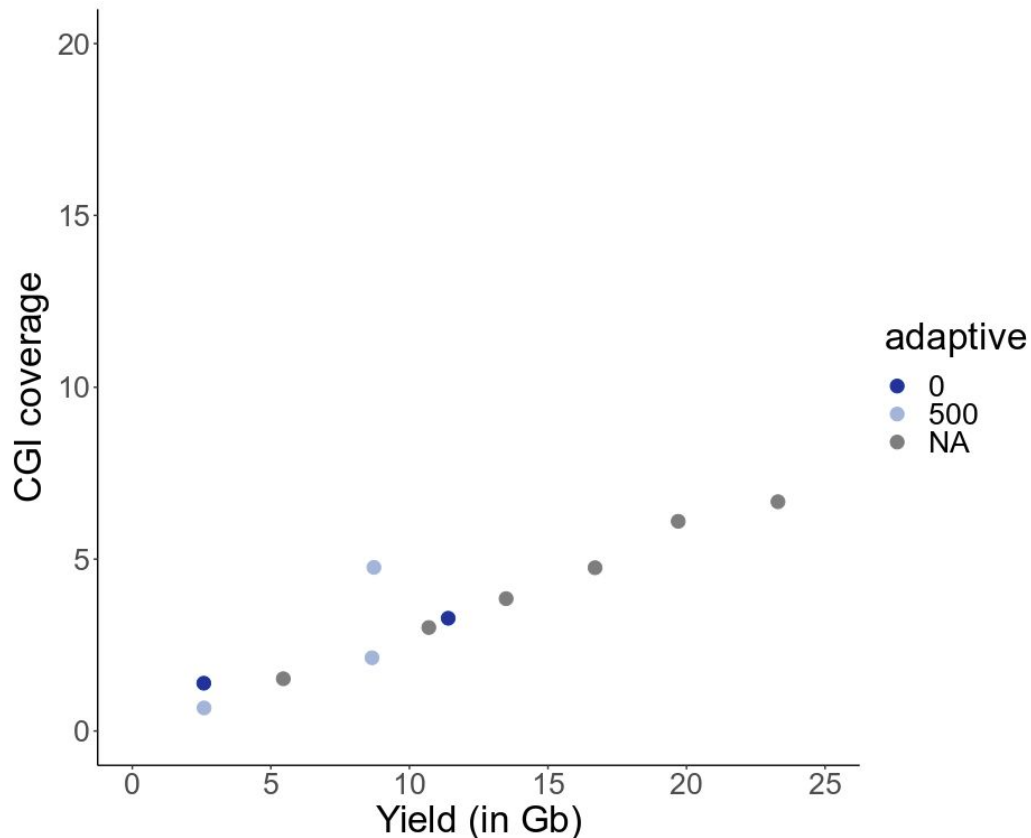
without adaptive sampling:

- CGI coverage correlates with yield

enrichment for CpG islands:

- no change

Adaptive sampling



without adaptive sampling:

- CGI coverage correlates with yield

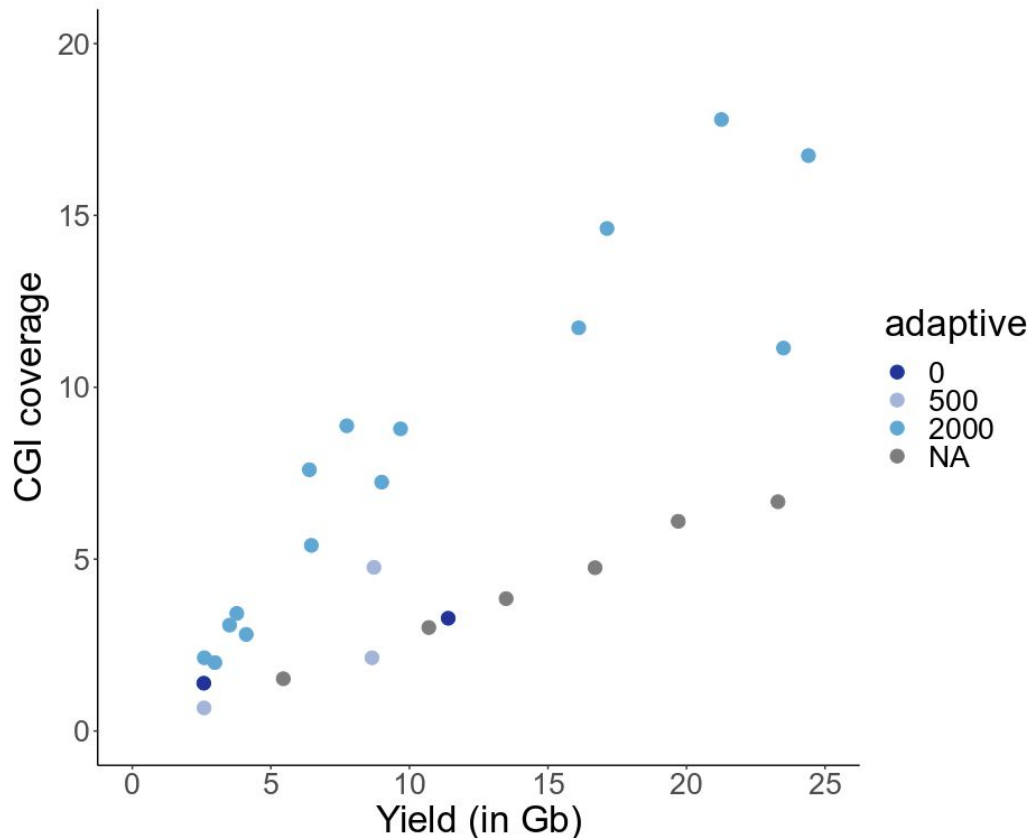
enrichment for CpG islands:

- no change

enrichment for CpG islands + 500nt extension:

- one (outlier?) with higher coverage

Adaptive sampling



without adaptive sampling:

- CGI coverage correlates with yield

enrichment for CpG islands:

- no change

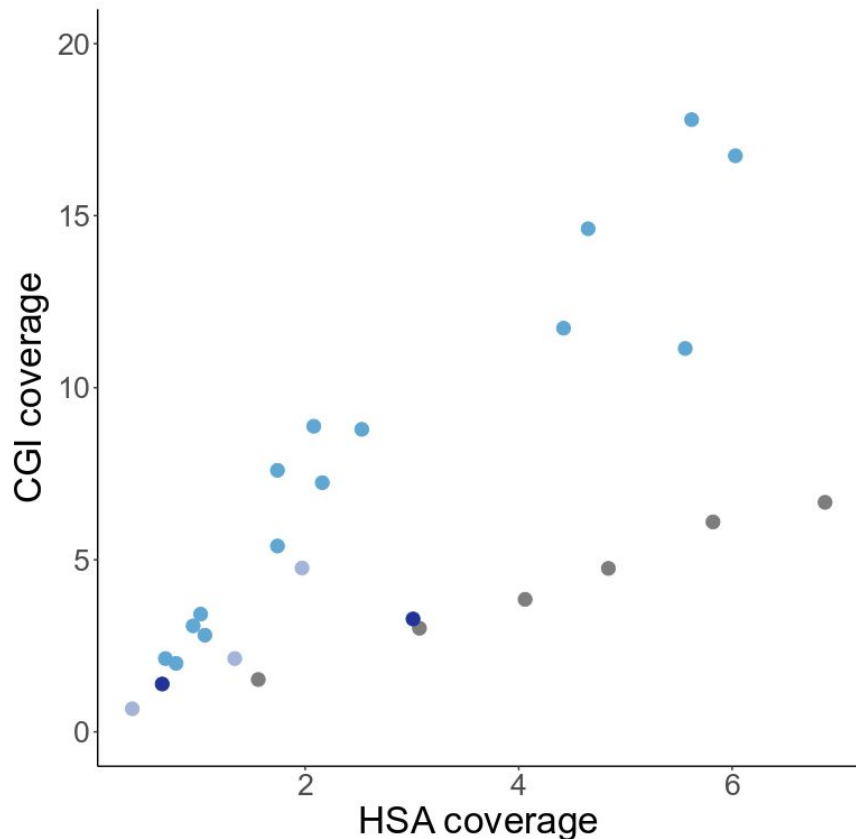
enrichment for CpG islands + 500nt extension:

- one (outlier?) with higher coverage

enrichment for CpG islands + 2000nt extension:

→ **increased coverage on CpG islands**

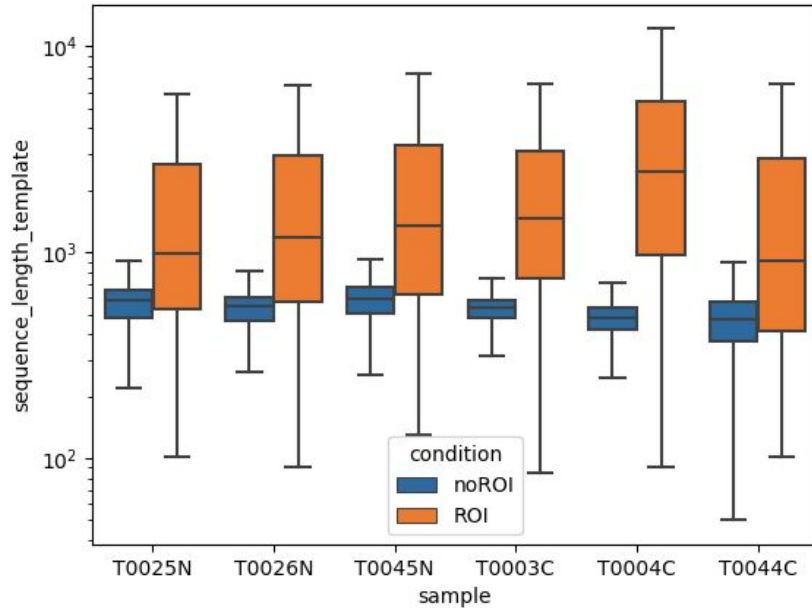
Adaptive sampling



- increased coverage on the region of interest
- important to extend the region of interest

- higher coverage on CpG islands than on the human genome.

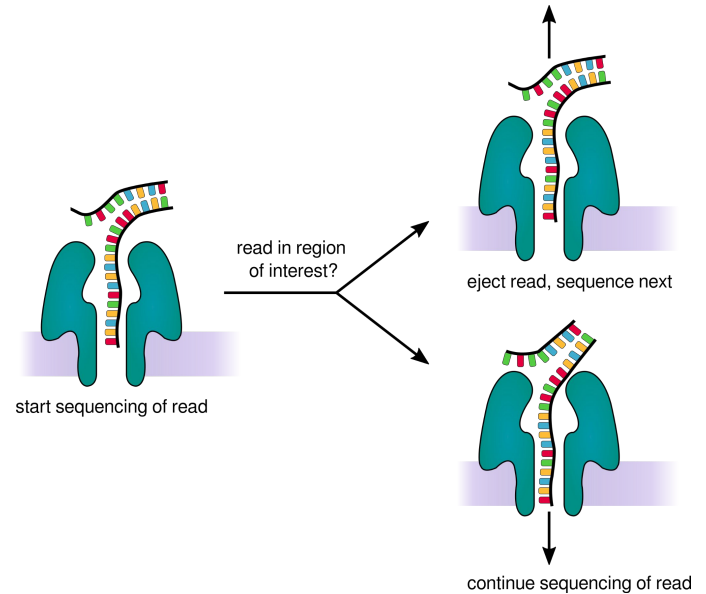
Adaptive sampling



→ shorter reads in general, long reads in CpG islands

Adaptive Sampling:

- only DNA mapping to region of interest is sequenced further
- DNA is sequenced with 400 bp/s

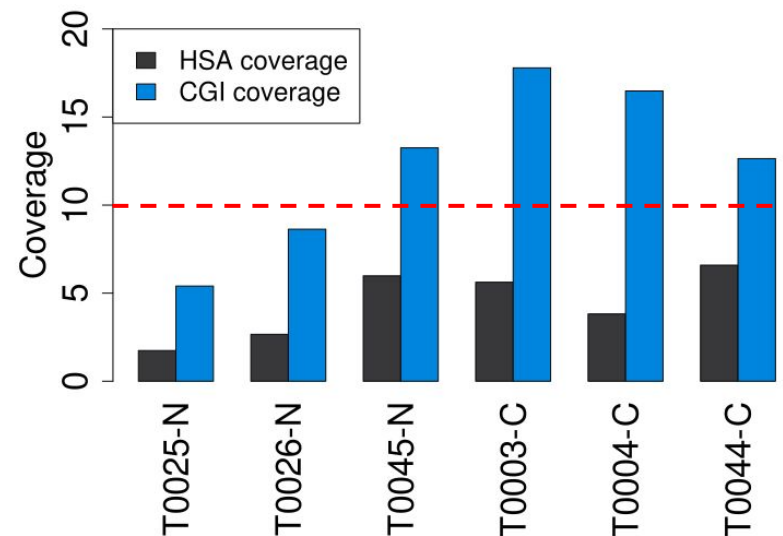


Our patient cohort

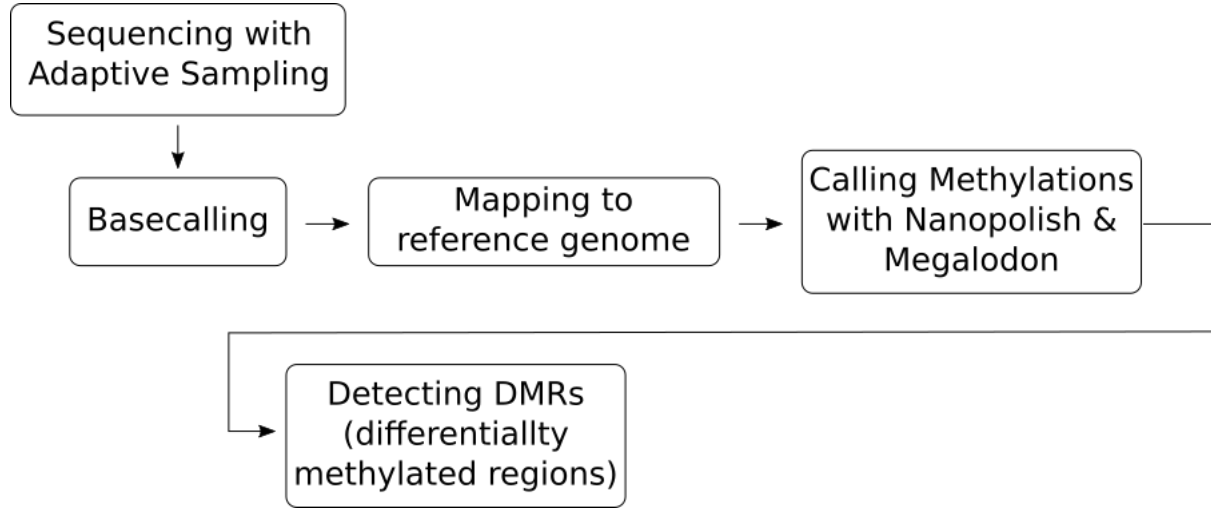
Control = Snorer

Patients = Head Neck Cancer Patients

	Sample	Sex	Age	HPV status	Localisation
Control	T0025N	m	59 y	neg	Uvula
	T0026N	m	58 y	unknown	Uvula
	T0045N	m	22 y	neg	Uvula
Tumor	T0003C	m	55 y	neg	Hypopharynx
	T0004C	m	68 y	neg	Hypopharynx
	T0044C	m	53 y	unknown	Oral cavity



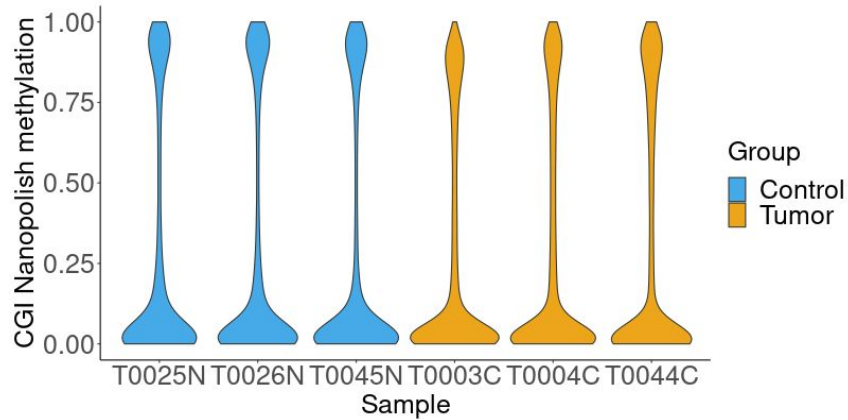
Sequencing & Methylation Calling workflow



→ Similar workflow described by ONT as RRMS (Reduced-Representation Methylation Sequencing)¹

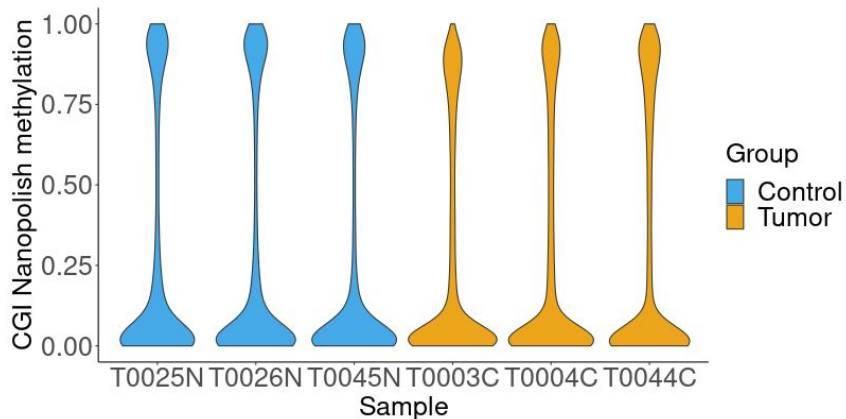
¹<https://nanoporetech.com/sites/default/files/s3/posters/lc2022/RRBS%20v1.1%20digital.pdf>

Differences between Control and Tumor

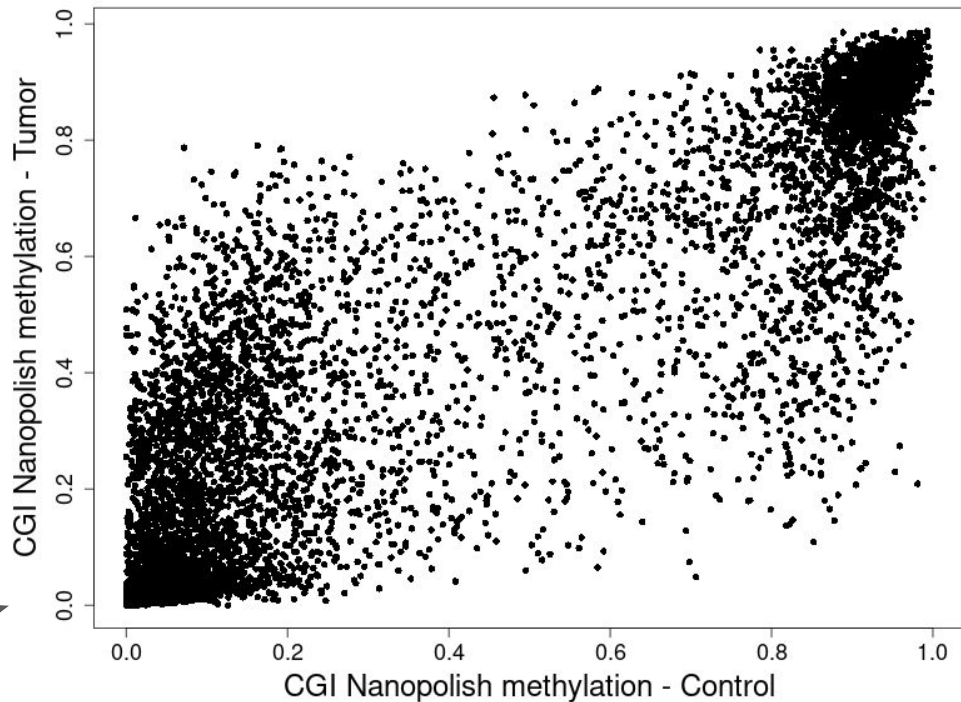


→ no differences between Control and Tumor samples in overall CGI methylation

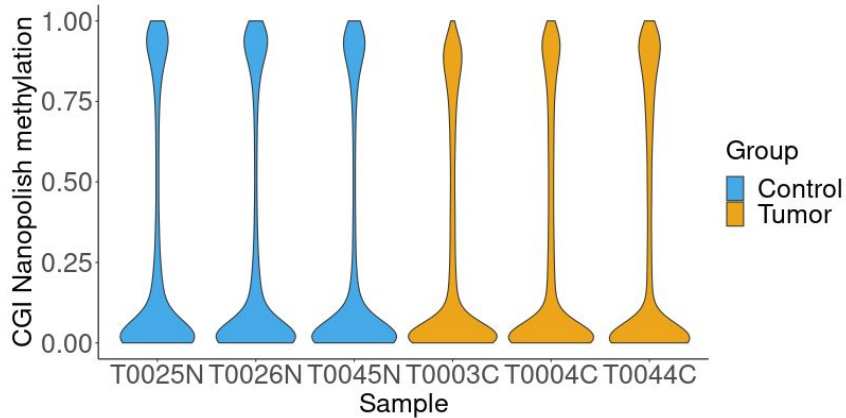
Differences between Control and Tumor



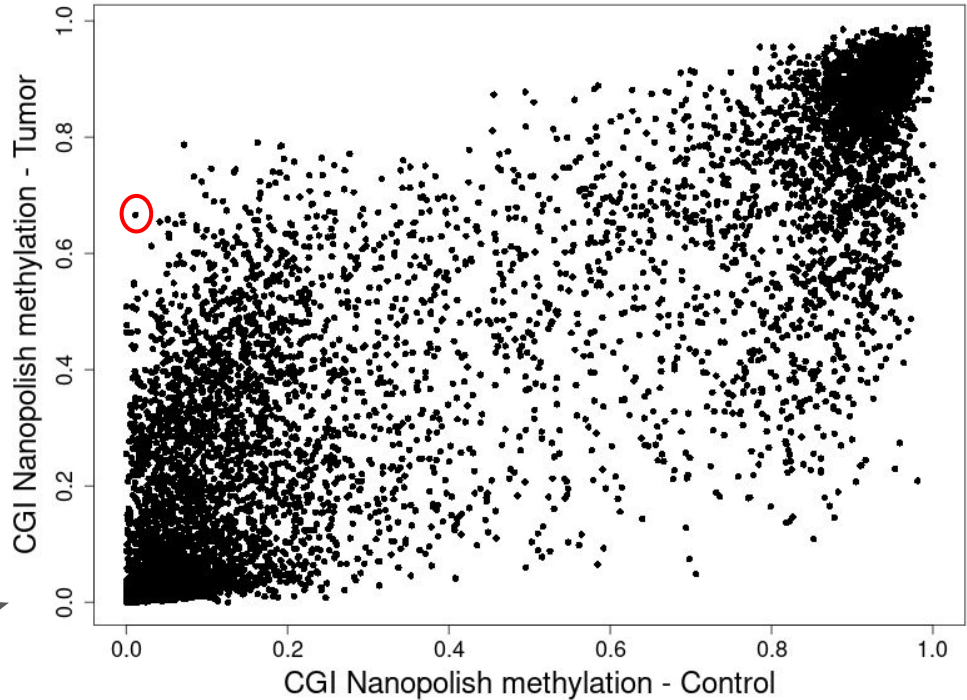
- no differences between Control and Tumor samples in overall CGI methylation
- differences between Control and Tumor at single CpG islands



Differences between Control and Tumor



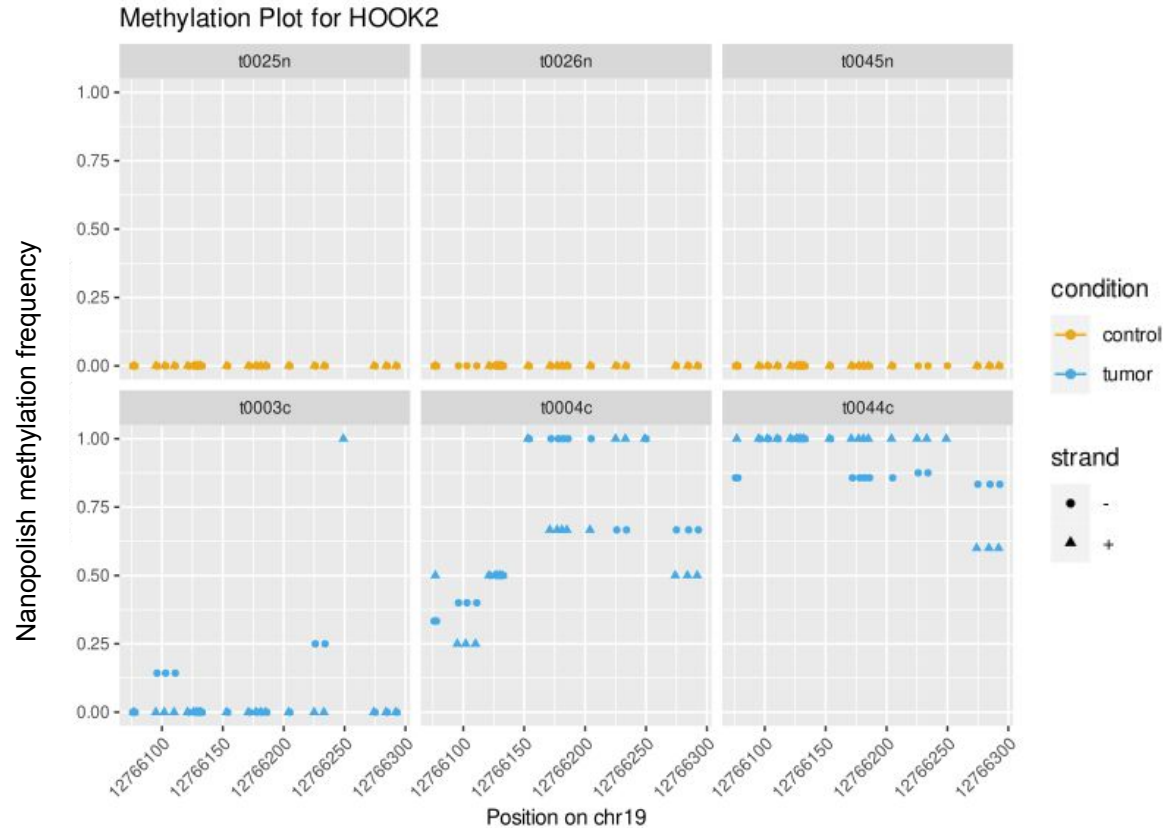
- no differences between Control and Tumor samples in overall CGI methylation
- differences between Control and Tumor at single CpG islands



Per nucleotide methylation frequency

Methylation frequency per nucleotide

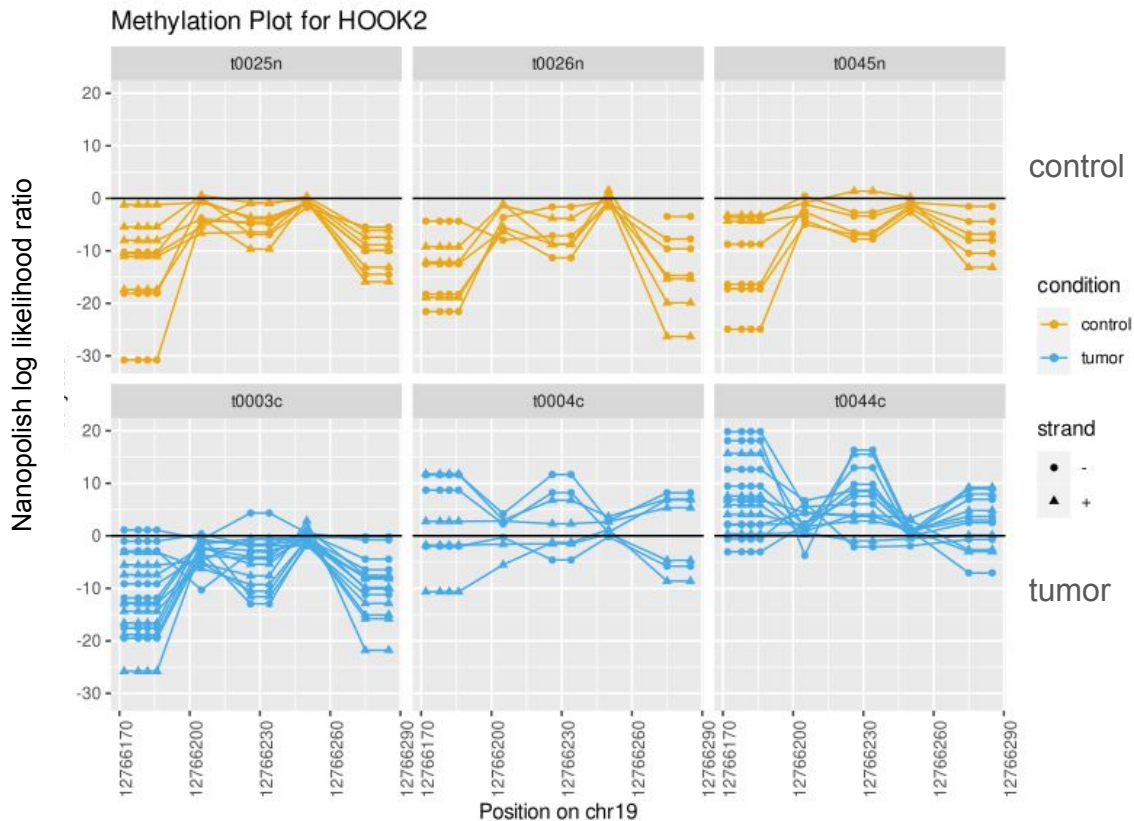
- Differences in methylations within CpG islands
- Average per CpG island might not be a good measurement



Prediction of new marker regions

Methylation frequency per nucleotide per read

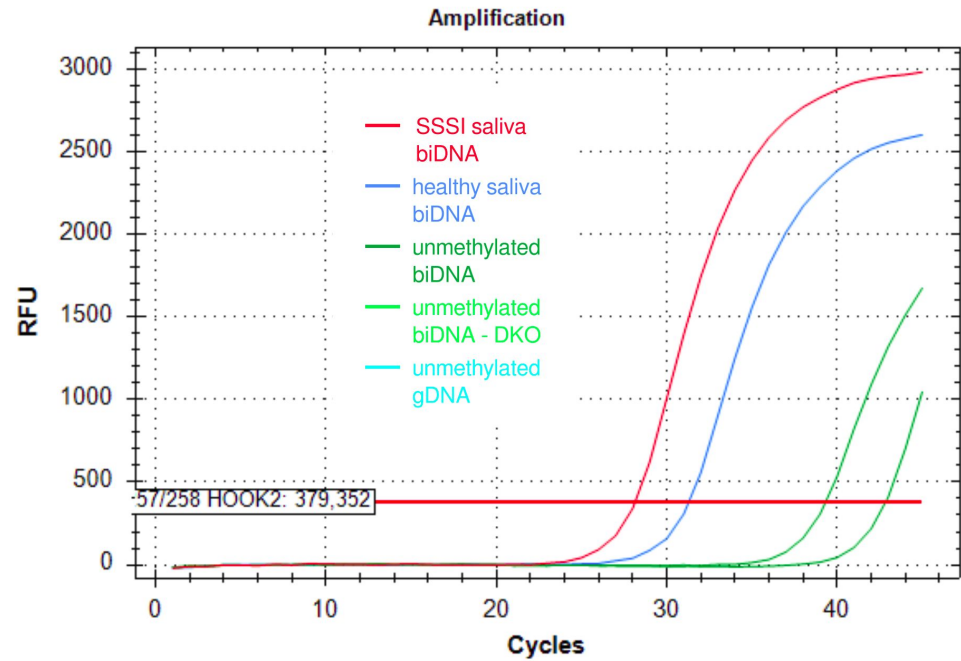
- Subregions within CpG islands show stronger differences in methylation
 - Differences between control and tumor tissue can be seen
- Average per CpG island might not be a good measurement



Validation with PCR assay (Lab)

- design PCR primer
- test with methylated & unmethylated DNA
- test with saliva from healthy individuals

- region is methylated in healthy control saliva



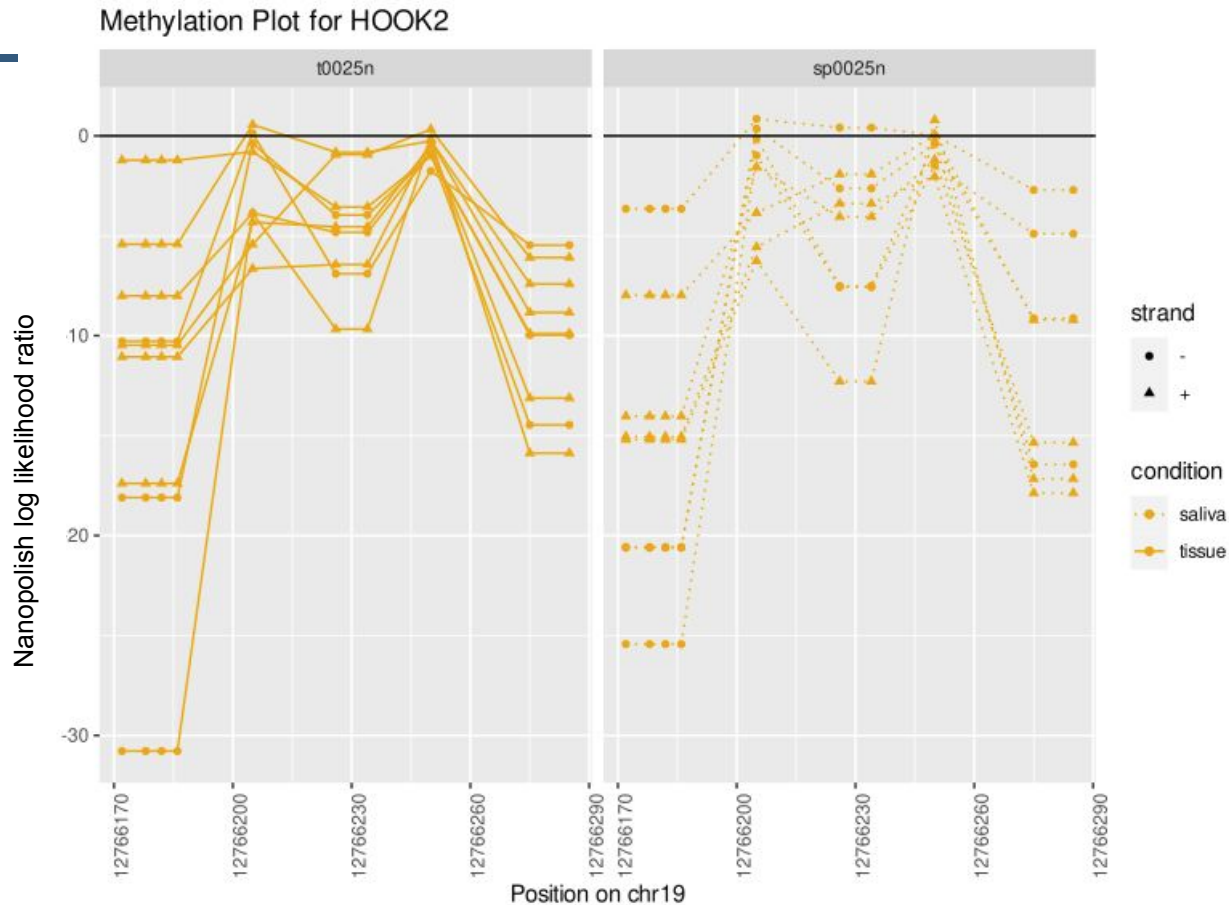
Target	SSSI saliva biDNA		healthy saliva biDNA		unmethylated biDNA		unmethylated biDNA - DKO		unmethylated gDNA	
	Cq	Tm	Cq	Tm	Cq	Tm	Cq	Tm	Cq	Tm
HOOK2	28.14	80.00	31.20	79.50	39.32	81.50		None		None
	28.11	80.00	31.28	80.00		None		None		None
	28.18	80.00	31.00	79.50	42.84	None		None		None

Comparison with saliva

sequenced saliva for one control sample
saliva is unmethylated here

- not in every patient methylated?
- Coverage too low?
- is our Nanopore sequencing approach sensitive enough?

- sequence saliva from tumor patients



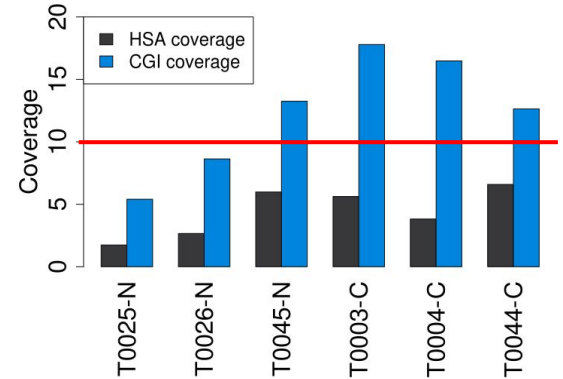
Outlook

Biological:

- increase coverage for control samples
- stick to one kind of head neck cancer (oral cancer)
- sequence saliva from tumor patients

Computational:

- Try the new Remora model for methylation calling
- Predict DNA methylation marker regions on CpG dinucleotide level



Thank you for your attention.



Special thanks to:

Manja Marz

Martina Schmitz
Alfred Hansel



Funded by Thüringen-Stipendium

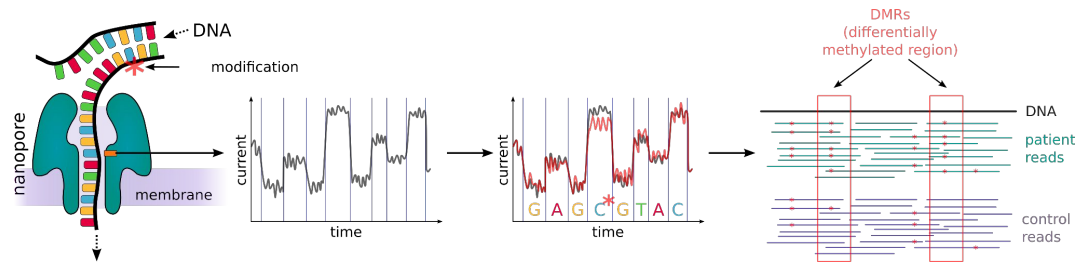
Outlook

- Try the new Remora model for methylation calling
- Predict DNA methylation marker regions on CpG dinucleotide level
- Compare different methods for the prediction
 - Detecting DMRs using existing BS-Seq tools (e.g. DSS)
 - Statistical tests
 - Sliding window: one condition completely unmethylated

Backup-Slides

Methylation Calling Algorithms

- Different methylation caller exist
- On the context of CpG islands and high GC-content especially Nanopolish and Megalodon are described as well performing methylation callers¹



¹Liu, Y., et al. DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. *Genome Biol* **22**, 295 (2021).

²Simpson, Jared T., et al. "Detecting DNA cytosine methylation using nanopore sequencing." *nature methods* 14.4 (2017): 407-410.

³Oxford Nanopore Technologies, <https://github.com/nanoporetech/megalodon>.

Methylation Calling Algorithms

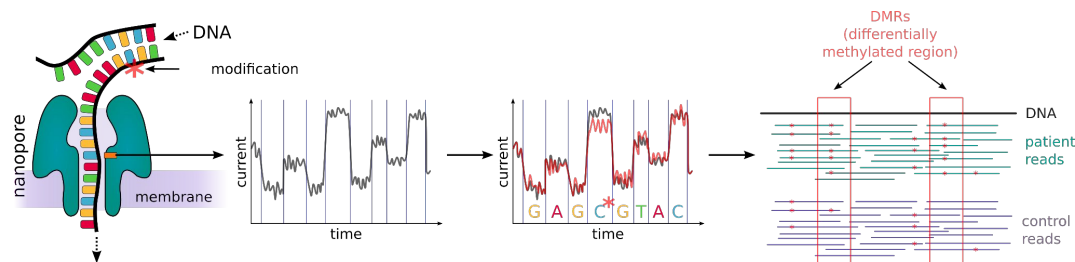
- Different methylation caller exist
- On the context of CpG islands and high GC-content especially Nanopolish and Megalodon are described as well performing methylation callers¹

Nanopolish²:

- HMM approach
- Returns log-likelihood per read and position

Megalodon³:

- Recurrent neural network
- Returns a score per position and read



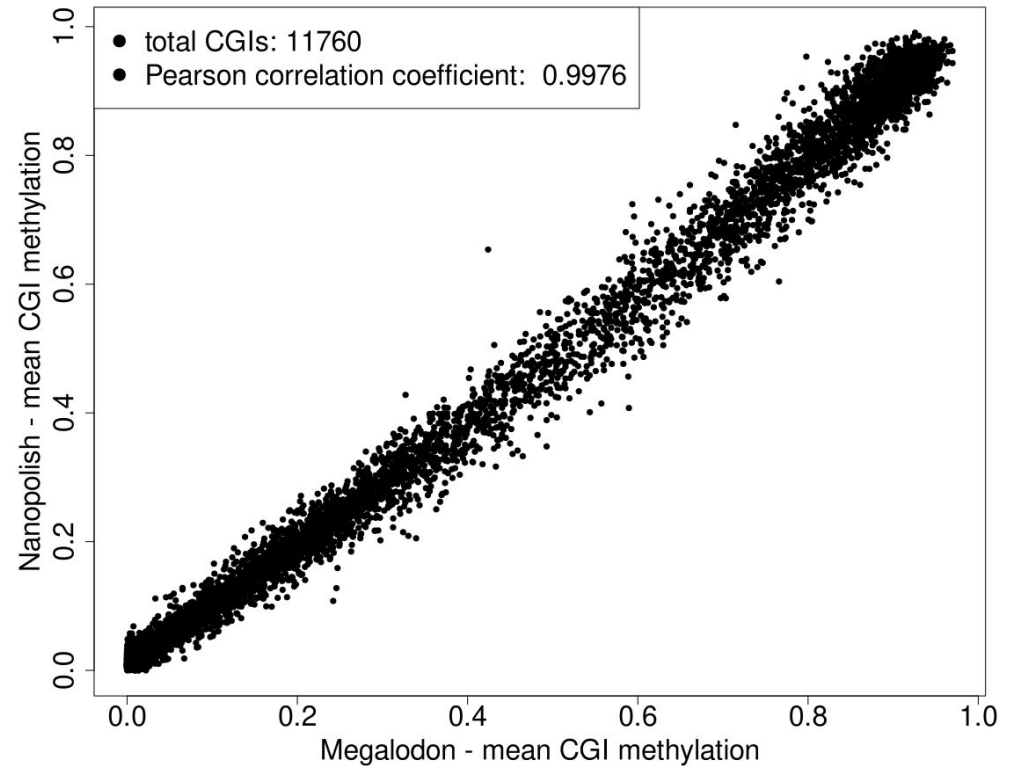
¹Liu, Y., et al. DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. *Genome Biol* **22**, 295 (2021).

²Simpson, Jared T., et al. "Detecting DNA cytosine methylation using nanopore sequencing." *nature methods* 14.4 (2017): 407-410.

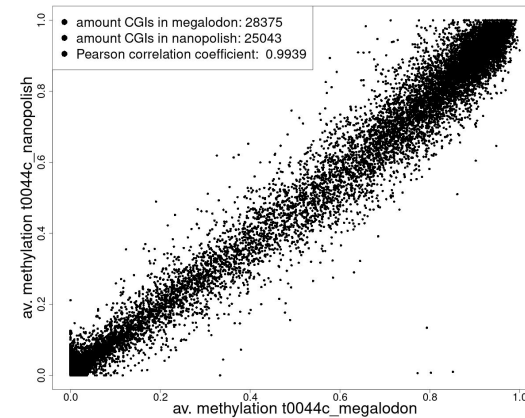
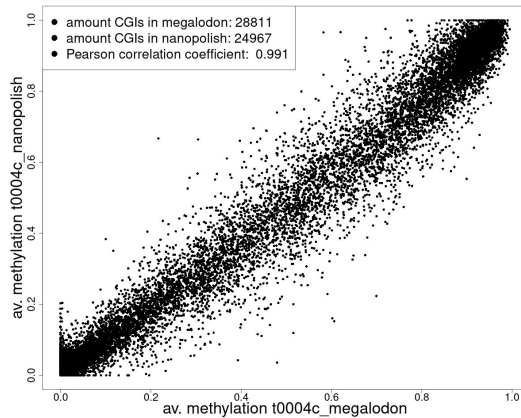
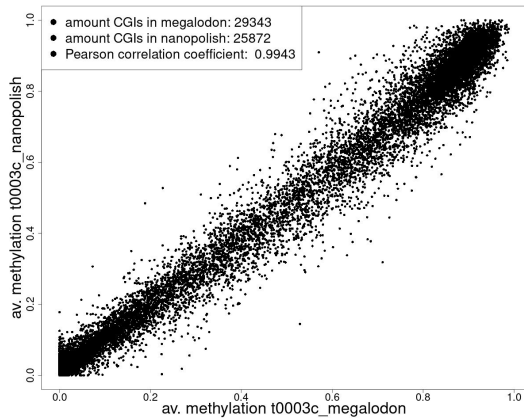
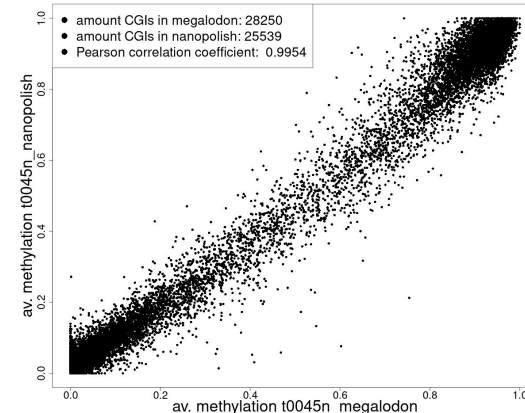
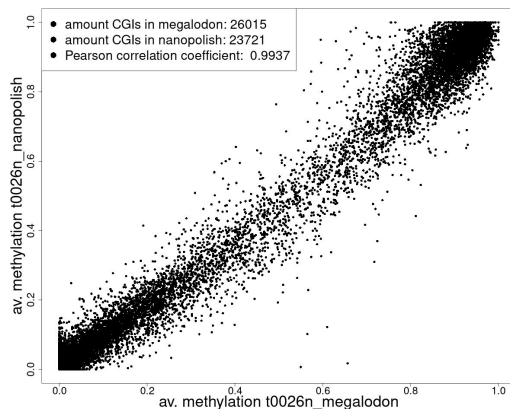
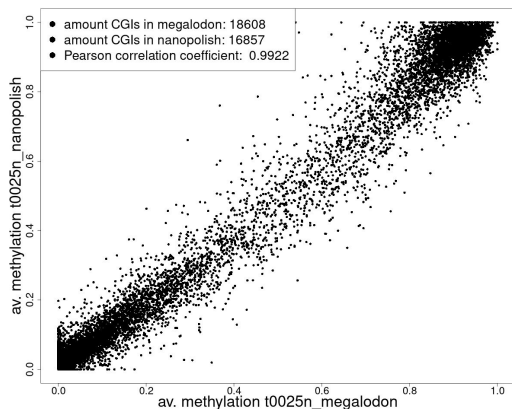
³Oxford Nanopore Technologies, <https://github.com/nanoporetech/megalodon>.

Differences between Nanopolish and Megalodon

- High correlation between the average per CpG island methylation between Nanopolish and Megalodon
- Few outliers exist

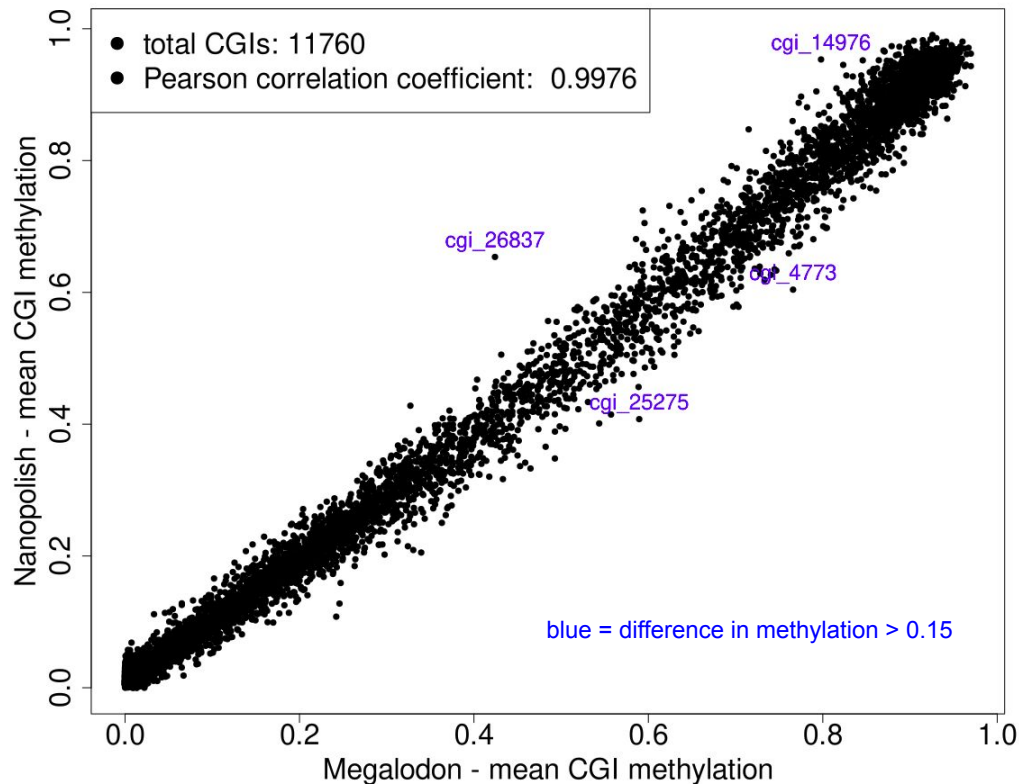


Differences between Nanopolish and Megalodon

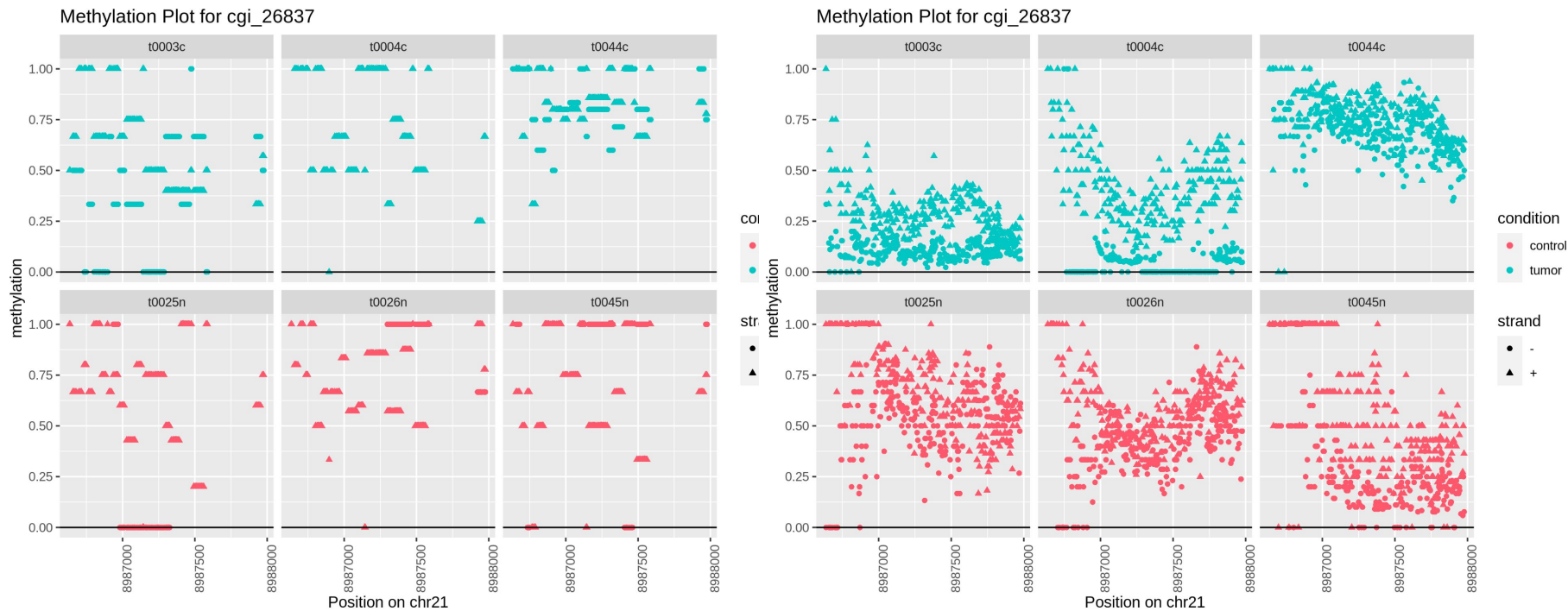


Differences between Nanopolish and Megalodon

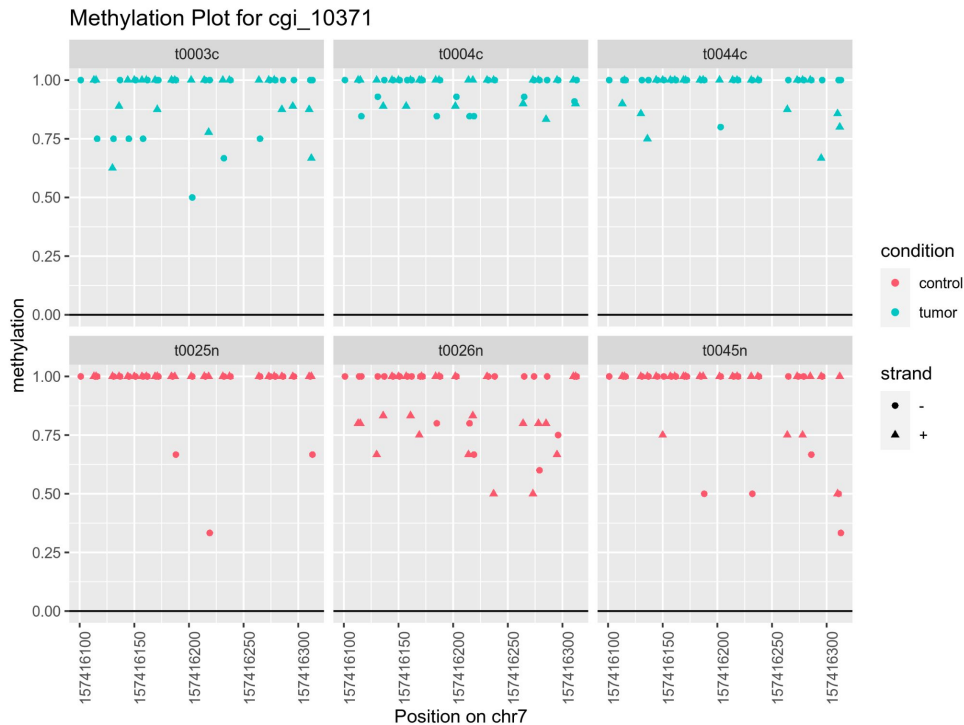
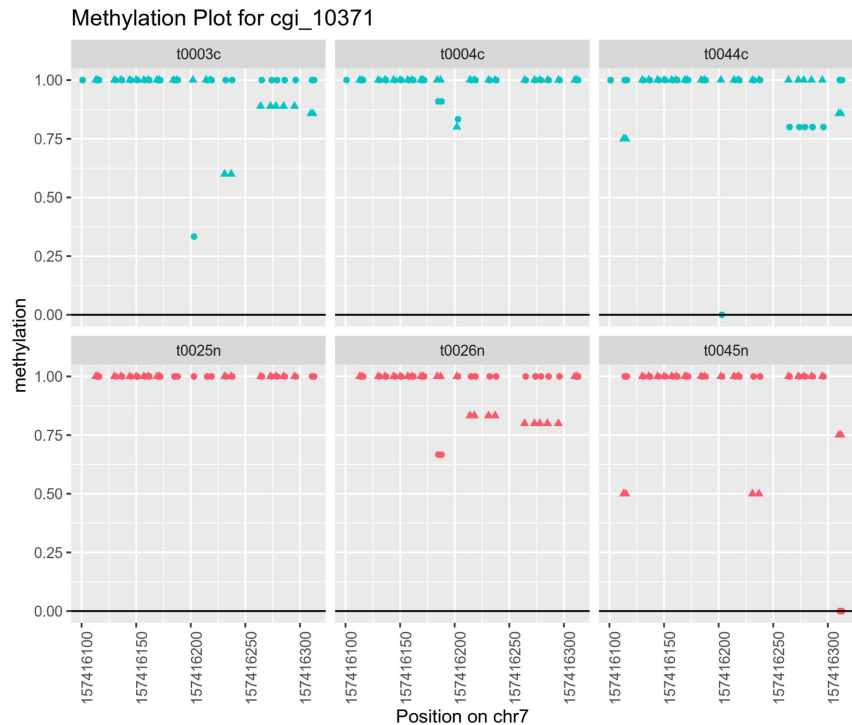
- High correlation between the average per CpG island methylation between Nanopolish and Megalodon
- Few outliers exist



Differences between Nanopolish and Megalodon



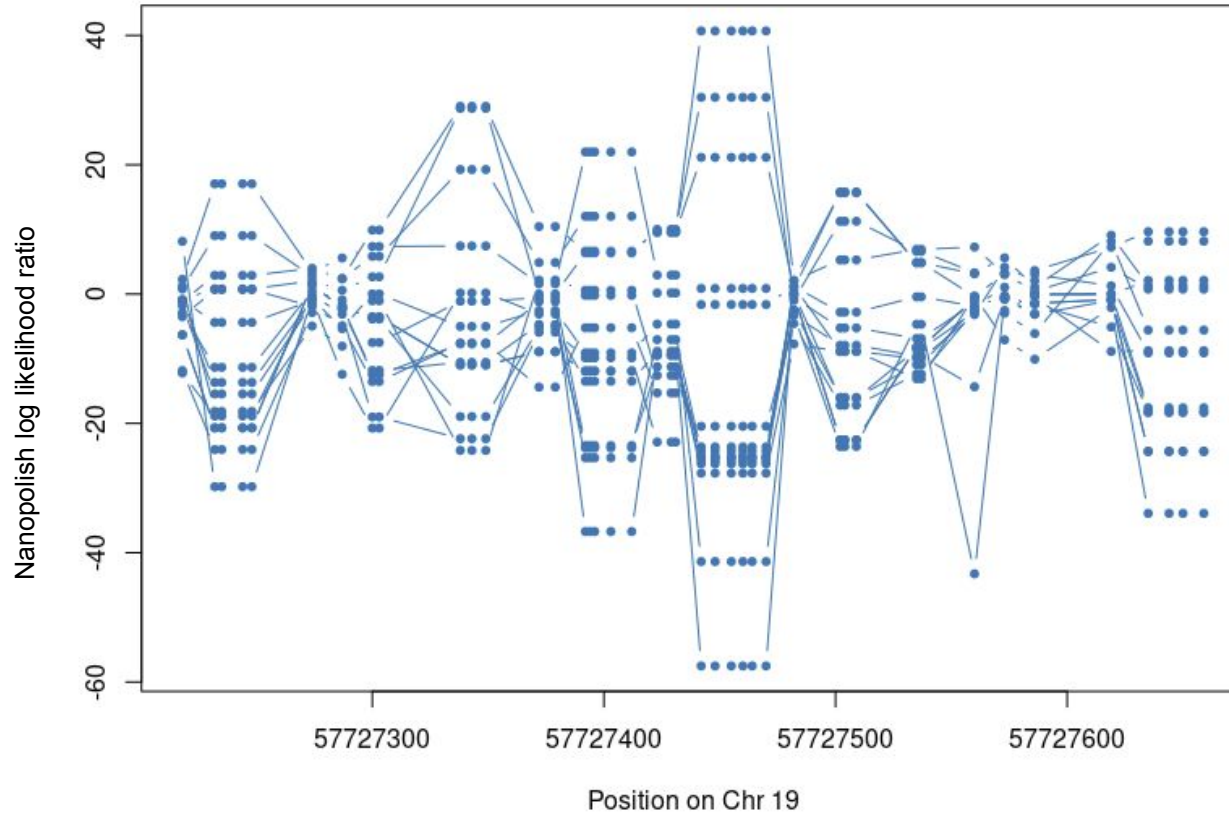
Differences between Nanopolish and Megalodon



Possible Modifications to call with Nanopore Sequencing

- Based on the modification calling tools, not the sequencing itself
- 4mC, 5mC, 5hmC, 6mA
- possible to train on your own data for specific modifications

Methylation frequency per nucleotide per read



Methylation-calling tools (not complete)

Statistical Tests:

- NanoMed → 5mC
- Tombo → 4mC, 5mC, 6mA

HMMs:

- Nanopolish → 5mC
- signalAlign → 5mC, 5hmC, 6mA

Neural Network:

- mCaller → 6mA
- DeepSignal → 5mC, 6mA
- Guppy → 5mC, 6mA
- Megalodon → 5mC, 6mA
- methBert → 5mC, 6mA

Machine Learning:

- METEORE → 5mC, 6mA

Liu, Y., Rosikiewicz, W., Pan, Z. *et al.* DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. *Genome Biol* **22**, 295 (2021).
<https://doi.org/10.1186/s13059-021-02510-z>

Methylation caller

- many different methylation caller exist
- different paper comparing them exist. Differences based on the DNA-context
- most popular: Nanopolish, Megalodon, Guppy, DeepSignal
- here we compare Nanopolish and Megalodon in the context of CpG islands

Nanopolish:

- HMM approach
- returns log-likelihood per read and position

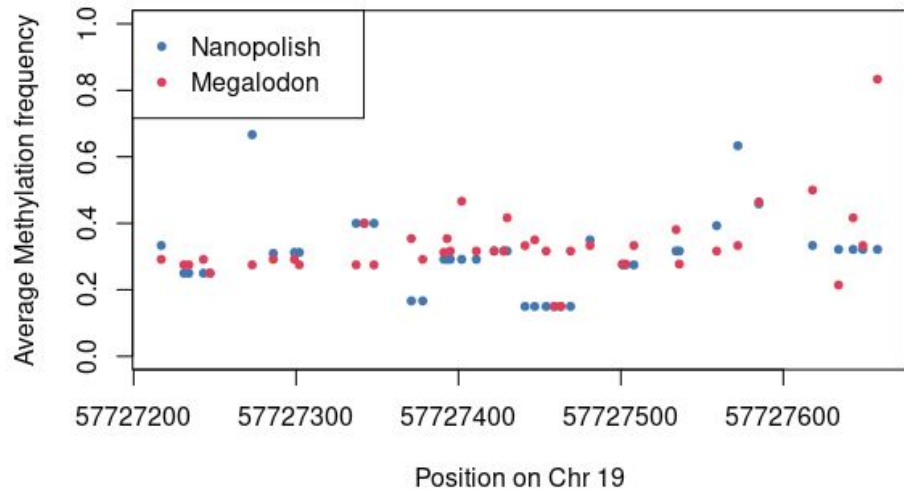
Megalodon:

- recurrent neural network
- returns a score per position and read

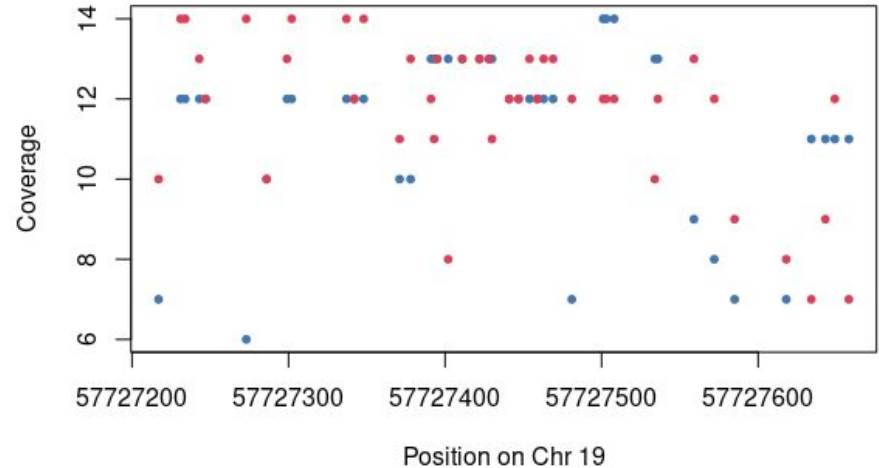
Liu, Y., Rosikiewicz, W., Pan, Z. *et al.* DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. *Genome Biol* **22**, 295 (2021).
<https://doi.org/10.1186/s13059-021-02510-z>

Per CpG methylation within a CpG island

Methylation for CpG island cgi25860 containing ZNF671

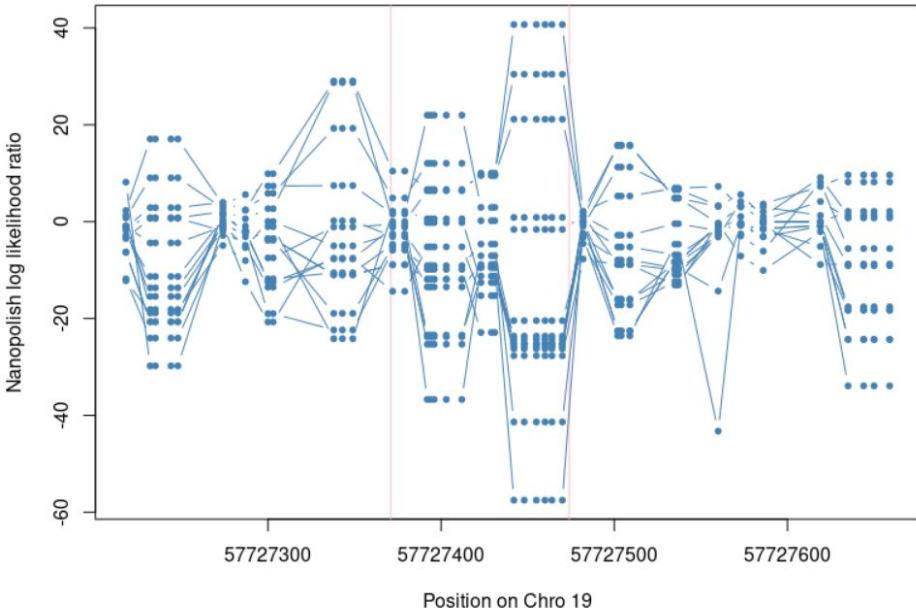


Coverage for CpG island cgi25860 containing ZNF671

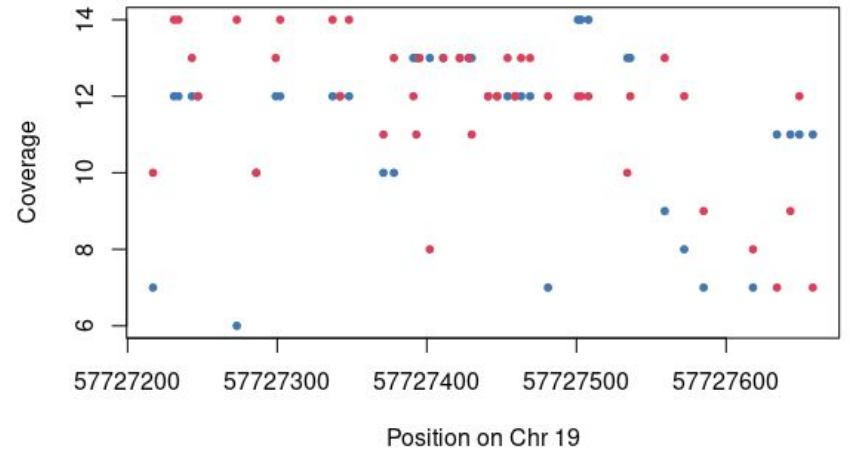


Per CpG methylation within a CpG island

CpG island cgi25860 containing ZNF671

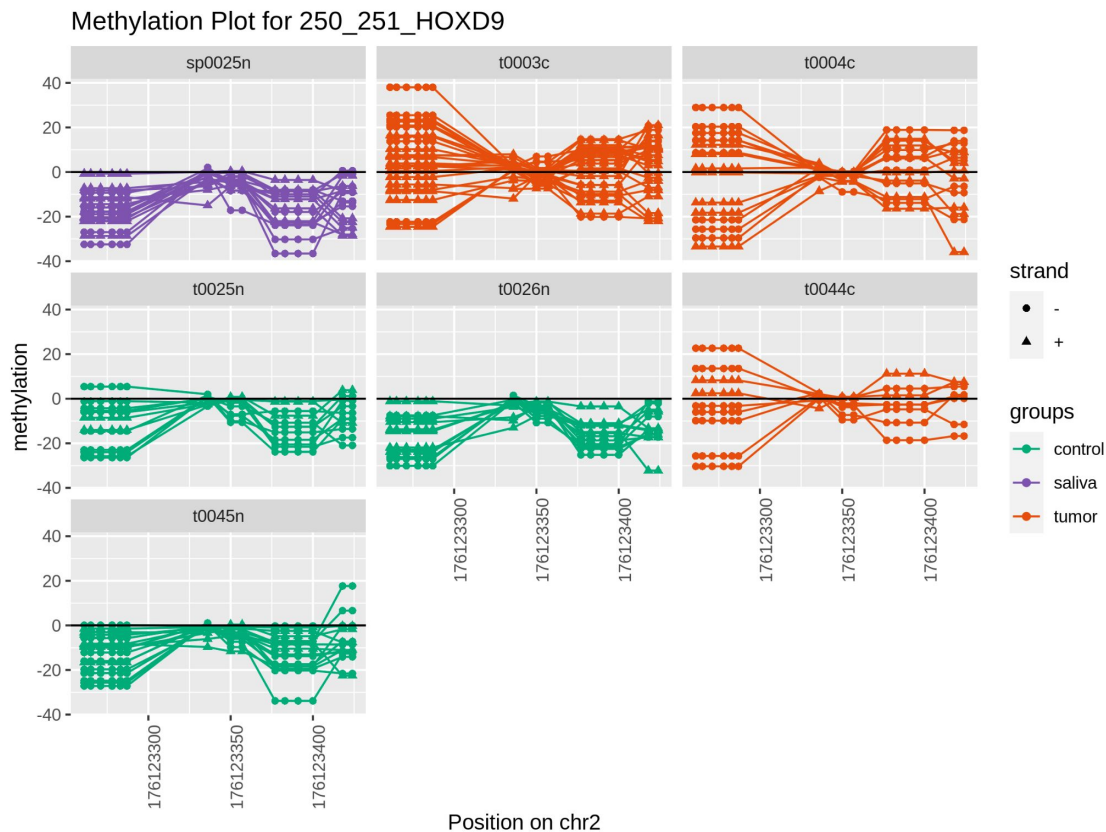


Coverage for CpG island cgi25860 containing ZNF671



Comparison with saliva

- HOXD9 has already been used before as marker
- results are consistent with PCR data



Comparison with PCR assay

Kontrollen	Diagnose	Lokalisation	Probennr.	ZIC1	ZNF833	ACTB	PAX6-1	HOXA9	ACTB	ZNF671	ACTB
T-0025-N	Kontrolle	Uvula	T-0025-N	32,90	35,19	25,35	31,48	29,00	25,41	34,43	25,31
SP0025	Kontrolle	Uvula	SP0025	0,00	0,00	26,25	0,00	34,89	26,05	38,66	26,01

Paar	ΔCq (Marker-ACTB)					$\Delta Cut-off$ ZIC1	$\Delta Cut-off$ ZNF833	$\Delta Cut-off$ PAX6-1	$\Delta Cut-off$ HOXA9	$\Delta Cut-off$ ZNF671	Σ 3 aus 5 positiv	HNOPOSITIVITÄT
	ΔCq ZIC1	ΔCq ZNF833	ΔCq PAX6-1	ΔCq HOXA9	ΔCq ZNF671							
T-0025-N	7,55	9,84	6,07	3,59	9,12	0	0	0	0	0	0	0
SP0025	keine Angabe	keine Angabe	keine Angabe	8,84	12,65	0	0	0	0	0	0	0
						G=SP	G=SP	G=SP	G=SP	G=SP	G=SP	G=SP