

Benchmarking the performance of Reduced Representation Methylation Sequencing (RRMS)

To benchmark **RRMS** performance, we performed human **RRMS** on five replicates of a metastatic melanoma cell line and its normal pair for a male individual (COLO829/COLO829_BL) and a triple-negative breast cancer cell-line pair (HCC1395/HCC1935_BL), and compared the results with those obtained using Reduced Representation Bisulfite Sequencing (**RRBS**). **RRBS** is a method used to obtain genome-wide methylation analysis without the need to sequence the whole genome: CGIs are enriched through restriction digest and size selection followed by bisulfite conversion of methylated bases and next-generation sequencing. The bisulfite workflow is complex, time consuming and is also imprecise, as it does not specifically target any promoter region.

Library preparation and sequencing

Genomic DNA was extracted from human cells grown in culture using the QIAGEN Genra Puregene Cell Kit (for gDNA extraction from blood or tissues, please see the [Nanopore Community](#) for more details) and quantified using the Qubit fluorimeter. For **RRBS**, samples were sent to Diagenode for processing and sequencing. For **RRMS**, sample preparation and sequencing was performed in-house. Firstly, DNA was sheared using a Covaris g-TUBE (Fig. 1).

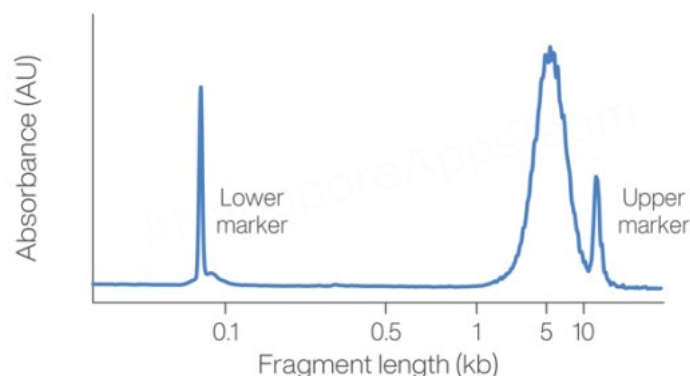


Figure 1. Fragment-length distribution of RRMS libraries after g-TUBE fragmentation.

Fragmented DNA was then prepared for sequencing using the Ligation Sequencing Kit, loaded on MinION flow cells and run on a GridION MK1. Off-target reads are ejected from the pore after approximately 1 second of sequencing and consequently have a read length of ~500 bp. On-target reads are accepted and sequence as normal, generating read lengths ~6 kb (as a result of the g-TUBE shearing, Fig. 2a). To optimise output, flow cells were washed twice, after 30 and 64 hours, and fresh library was re-loaded each time (Fig. 2b).

In this benchmarking experiment, each single sample was run on single MinION flow cell. However, it is now possible to multiplex up to 4 samples on a single PromethION flow cell and run on P2S or P24/48 platforms.

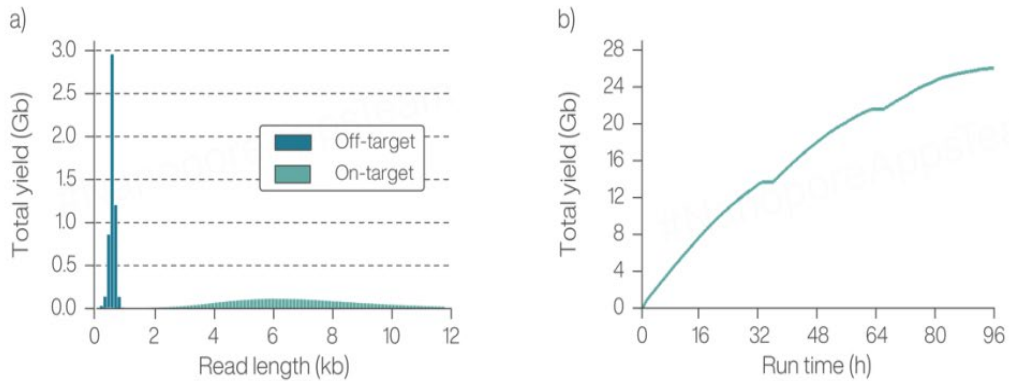


Figure 22. a) Read-length distribution and b) output from a typical RRMS run on a MinION flow cell

Results

RRMS generated ~35x average coverage across target regions and results in high confidence methylation calls for 7.3 - 8.5 million CpGs per sample, covering >90% of targeted promoters, CGIs, shores and shelves (NB if >50% of CpGs within a given feature were called then the feature is classified as being covered), and coverage across target regions is uniform and reproducible. RRBS, in comparison, yields 1.7 - 2.5 high-confidence calls per sample (Fig. 3).

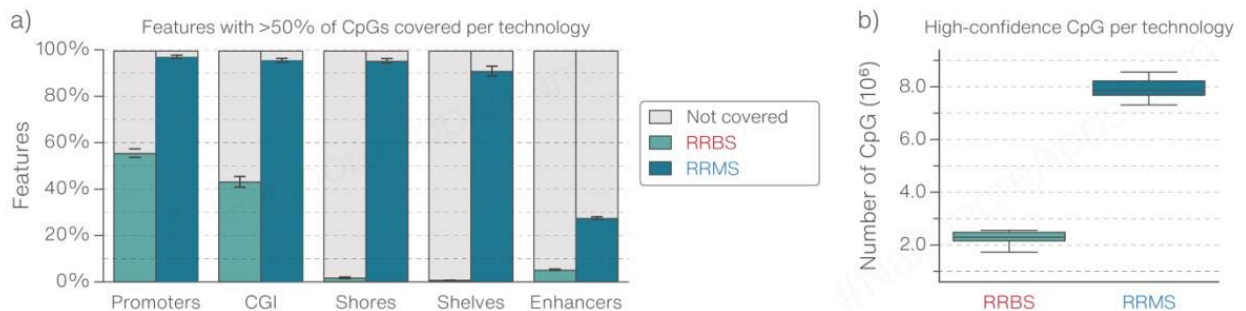


Figure 3.3 Typical performance of RRMS run. a) shows the average -fold coverage obtained for on-target ROI is >30x. b) shows the number of CpG sites where the methylation status could be determined with high confidence (i.e. >10 spanning read)

Methylation frequencies called by RRMS and RRBS are highly similar for CpGs covered by both technologies ($R > 0.967$) (Fig. 4).

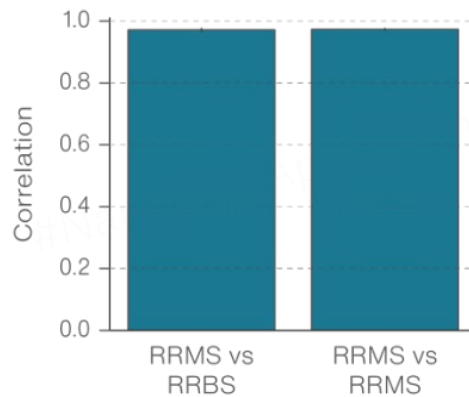


Figure 44. Comparison of CpG calls between RRMS and RRBS shows that where a CpG is covered by both technologies the methylation calls are highly correlated.

With **RRMS** we were able to detect ~62 Mb of differentially methylated regions (DMRs) between tumour and normal pairs, of which a high proportion overlapped with cancer census genes. In comparison **RRBS** yielded ~20 Mb of DMRs.

When investigating the DMRs further we identified that the de-novo methyltransferase DNMT3A promoter is clearly methylated in both tumours but not in their normal counterparts. These methyltransferases are essential for establishing and maintaining normal levels of methylation, and so their dysregulation can contribute to cancer development. Therefore, **RRMS** can provide key information for tumour characterisation as well as methylation status (Fig. 5).

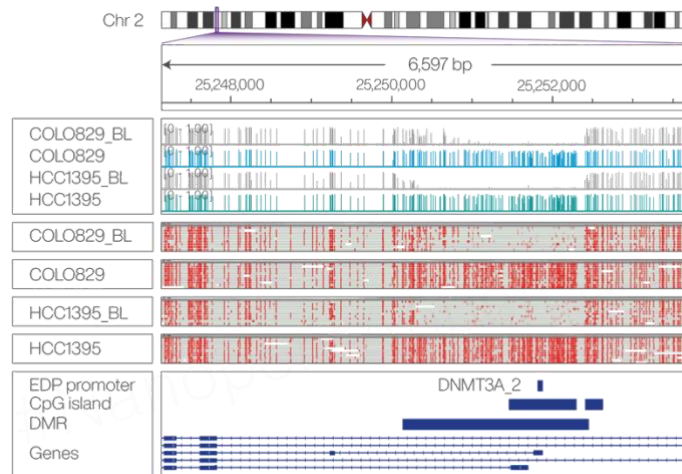


Figure 55 Differences in methylation between samples can be determined using RRMS. The methylation status of a 6.5 kb region of chromosome 2 is compared for two tumour/normal cell line pairs and clear differences in the methylation patterns can be observed.

Finally, as an additional benefit of **RRMS**, it is possible to call copy-numbers variants (CNV) across the whole genome without any additional sample preparation (i.e. you get CNV calling “for free”), by using the reads which are ejected by AS during the course of an **RRMS** run. We observe that the calls made using the ejected reads from **RRMS** are highly correlated to the calls made using short-read WGS and ONT WGS (Fig. 6).

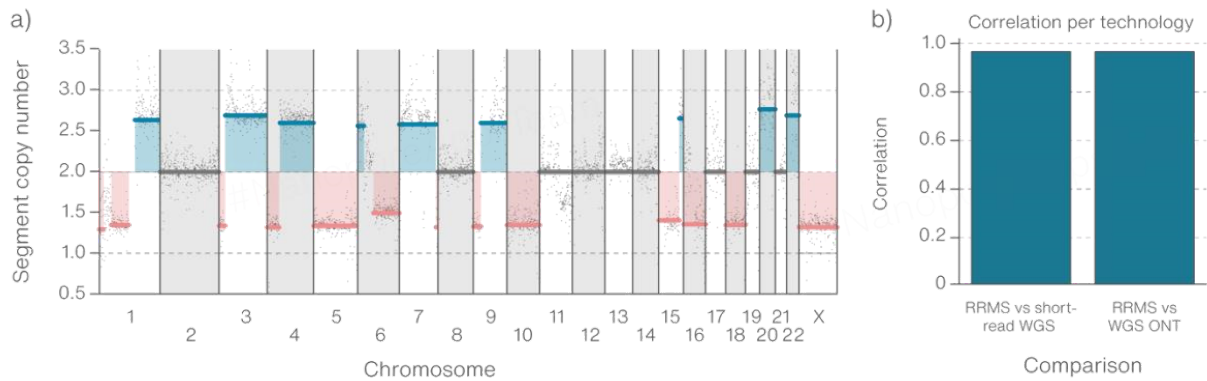


Figure 66 CNV calling using RRMS. a) The reads which are ejected during the course of an RRMS run can be used to call copy-number variation across the whole genome. b) The CNV calls made using RRMS are highly correlated to the calls made using short-read WGS and ONT WGS.

Conclusion

RRMS is a method that combines direct methylation detection and adaptive sampling to generate >7 million high-confidence CpG calls, covering all annotated CpG islands, shores, shelves and annotated promoters within the human genome, with minimal upfront library preparation. It enables detection of differentially methylated regions between samples and whole-genome CNV calling. In addition, it can be extended to include custom targets like marker SNPs or SVs. RRMS therefore provides key information with potential application to tumour characterisation and methylation status. Combined with its ease of use and ability to scale to a high number of samples, RRMS is well suited to investigating methylation differences in large cohorts, as well as providing deeper insights into the mechanisms behind diseases such as cancer and monitoring tumour progression.