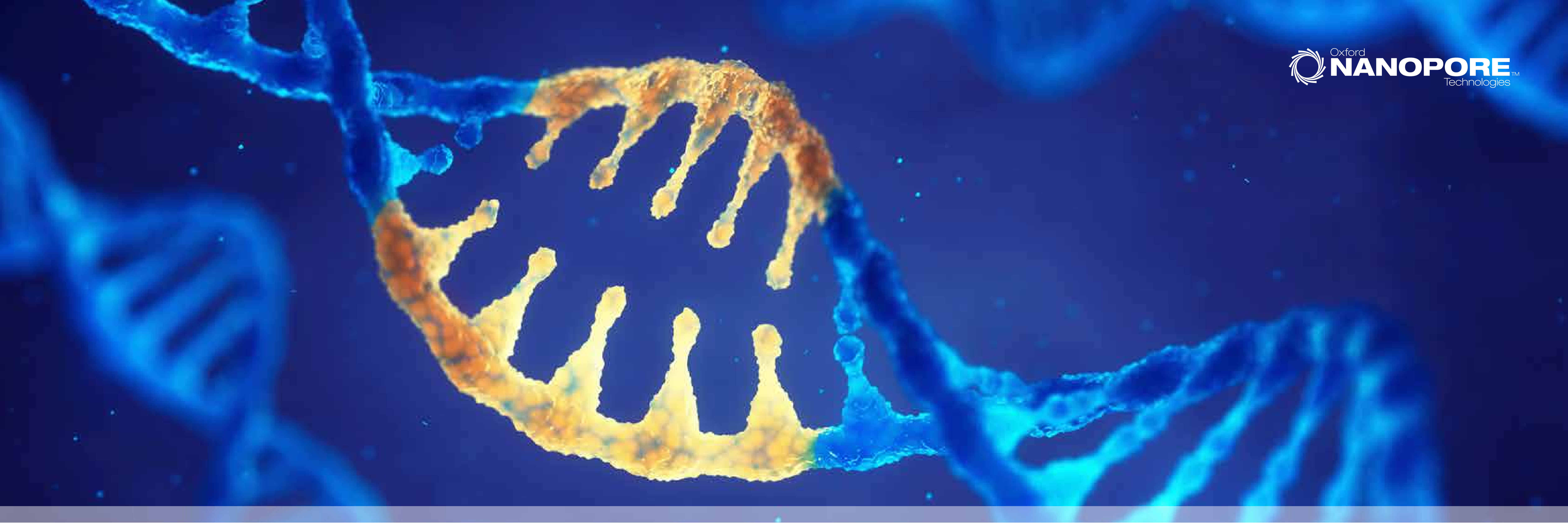


Accessible and robust base modification analysis with Modkit, the multi-tool for nanopore epigenetics

Leveraging DNA and RNA methylation data in your experiment shouldn't be hard — so we made a tool to make it easy. Modkit is open source and integrated into EPI2ME™ workflows

Contact: **art.rand@nanoporetech.com** More information at: **nanoporetech.github.io/modkit/**

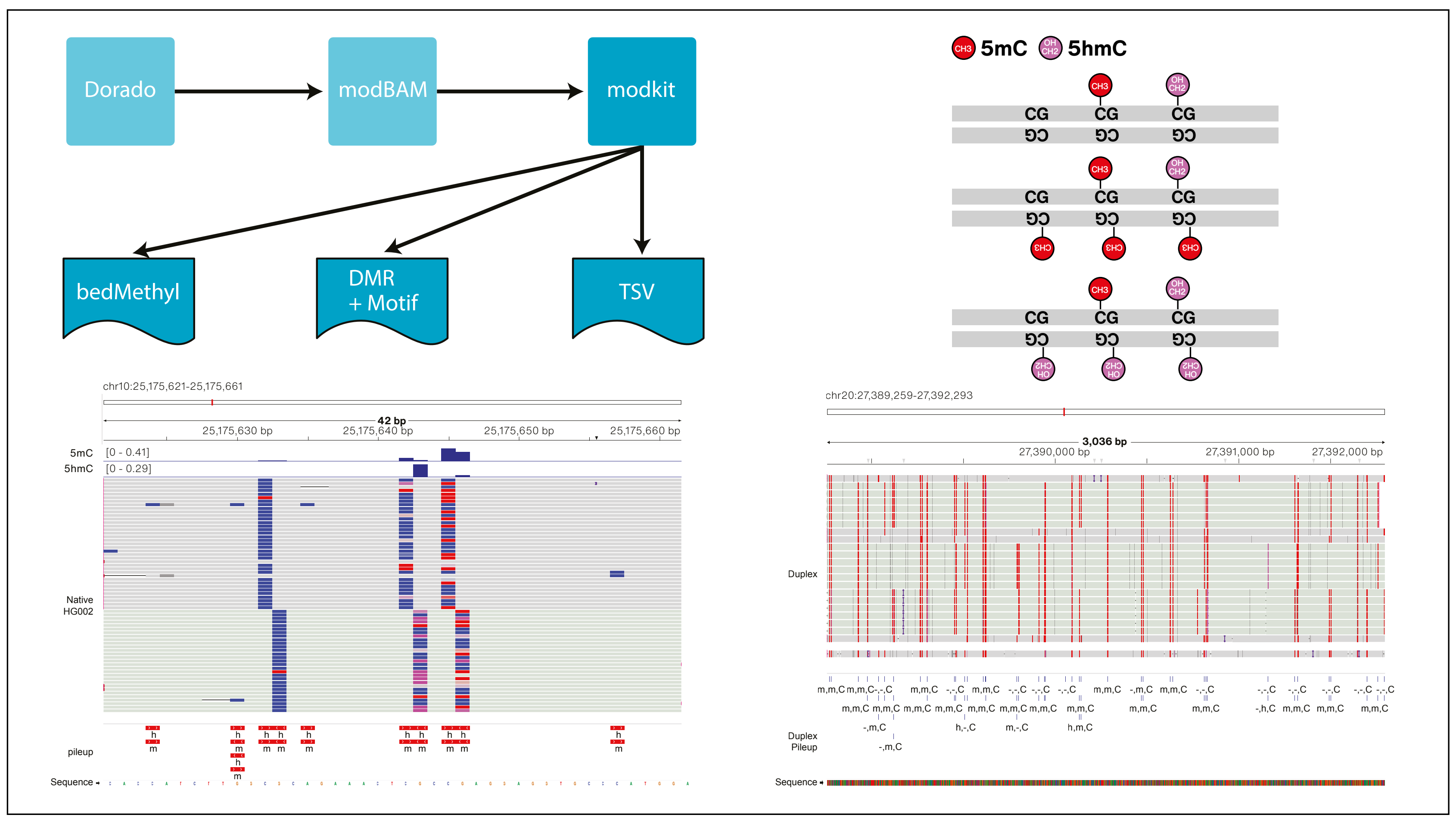


Fig. 1 Genome browser showing methylation tracks generated by Modkit

The first step in base modification steps: 'pileup' and 'extract'

Modified base detection is integrated into Oxford Nanopore basecalling software Dorado and MinKNOW™. The first step is to aggregate base modification counts across genomic or transcriptomic positions. This can be done by performing the 'pileup' command, or the duplex-enabled cousin 'pileup-hemi'. These tables, bedMethyl tables, are the input to many downstream steps and can be parsed with common analytics software. The 'extract' command will generate per-read tables allowing deeper investigation into single-molecule base modifications. The 'pileup-hemi' command will count the occurrences of double-stranded methylation patterns (hemi-methylation) when provided with duplex reads. Finally, 'stats' will collapse genomic windows or regions into summary statistics.

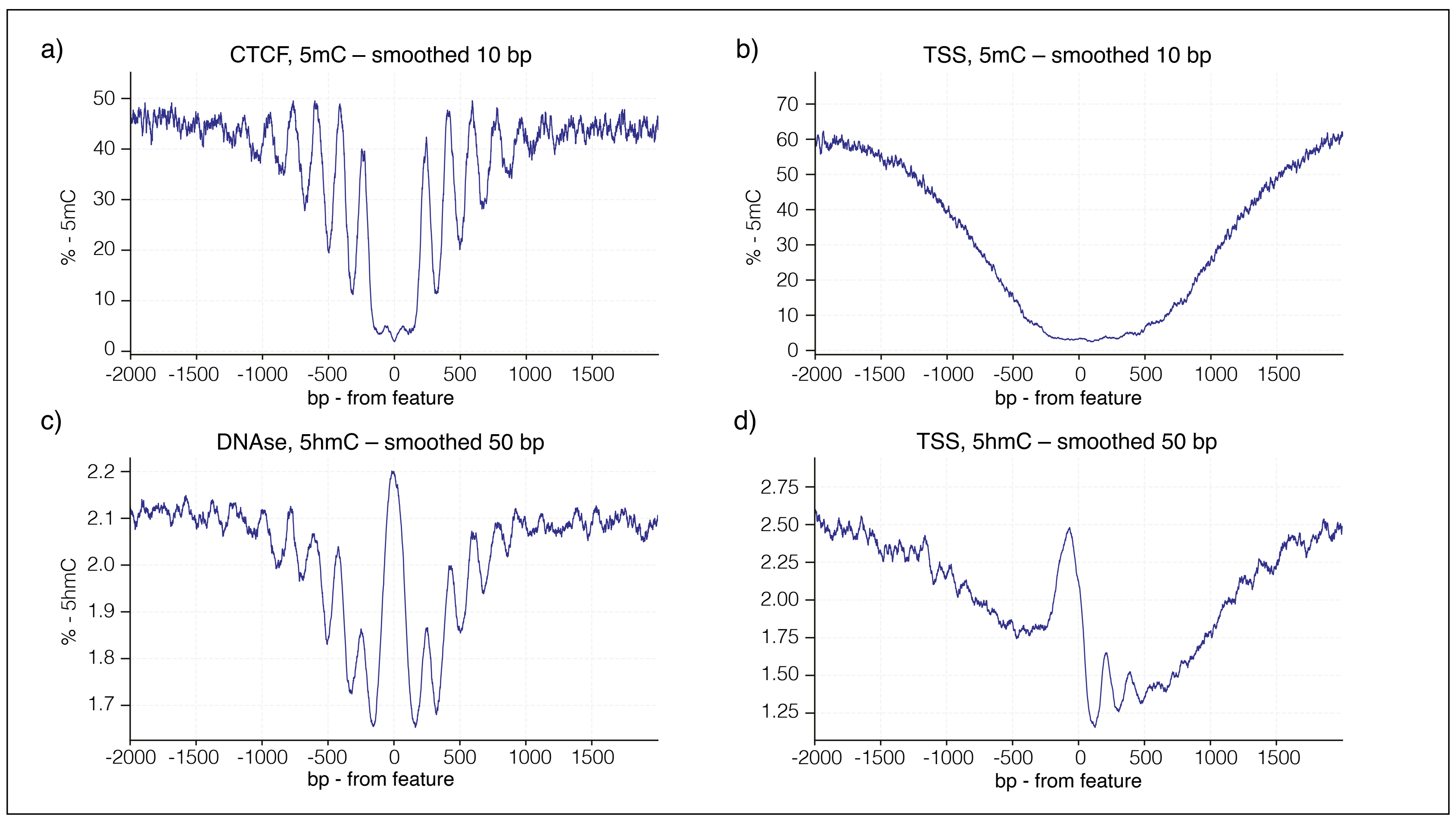


Fig. 2 Inspect modification patterns localised to genomic features of interest

Visualise and compare methylation patterns across genomic annotations with 'localize'

Changes in base modification patterns can be indicative of biological mechanisms. The 'localize' command quickly aggregates base modification rates localised to genomic features of interest. The tabulation is fast, speeding up iteration and exploratory data analysis. Aggregation is performed per base modification, allowing inspection of, for example, 5mC and 5hmC in the same sample. Example plots show 5mC patterns at CTCF binding locations (Fig. 2a) and transcription start sites (TSS) (Fig. 2b). Peaks of 5hmC can be seen at DNAse hypersensitivitysites (Fig. 2c) and TSS (Fig. 2d).

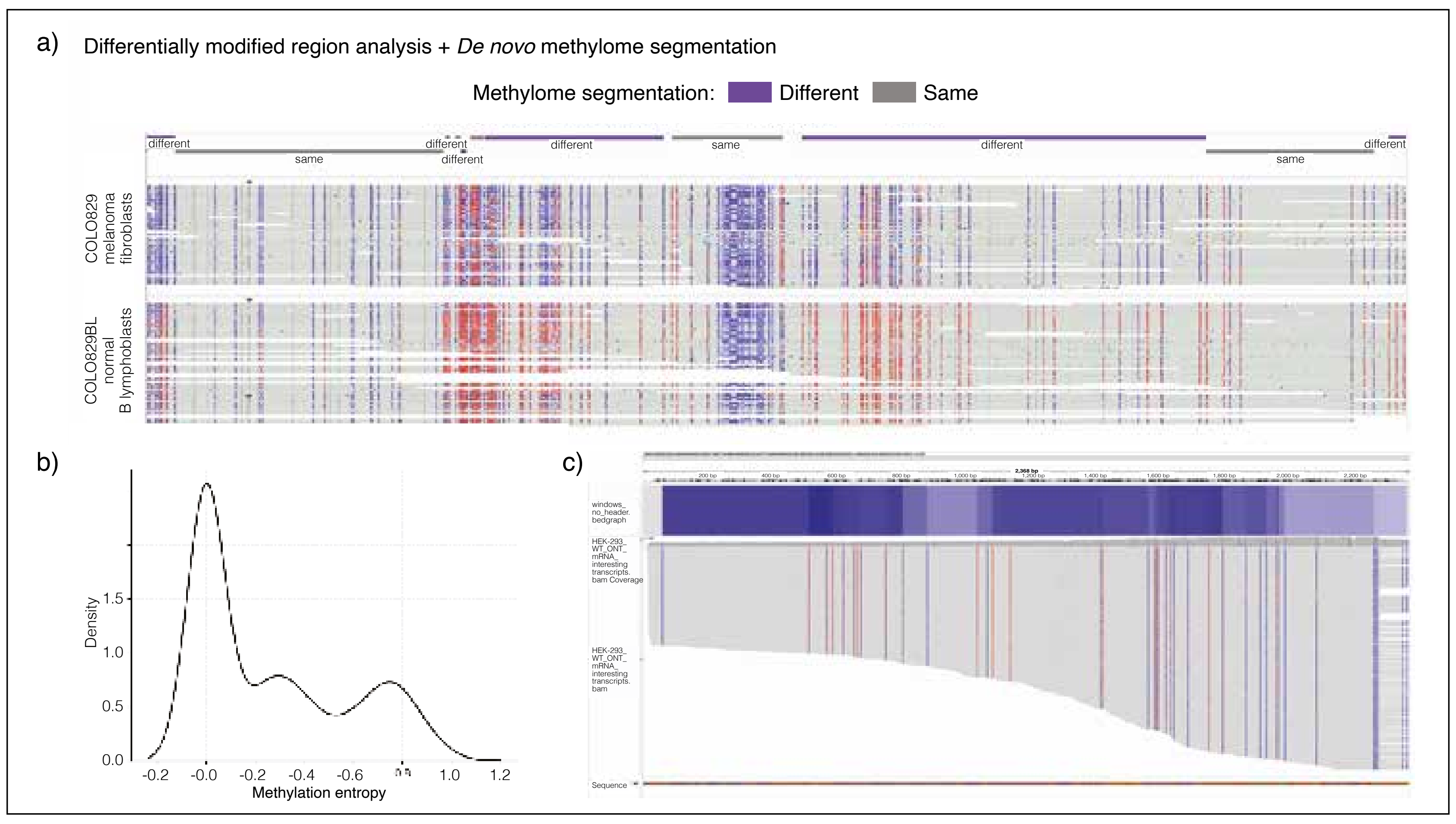
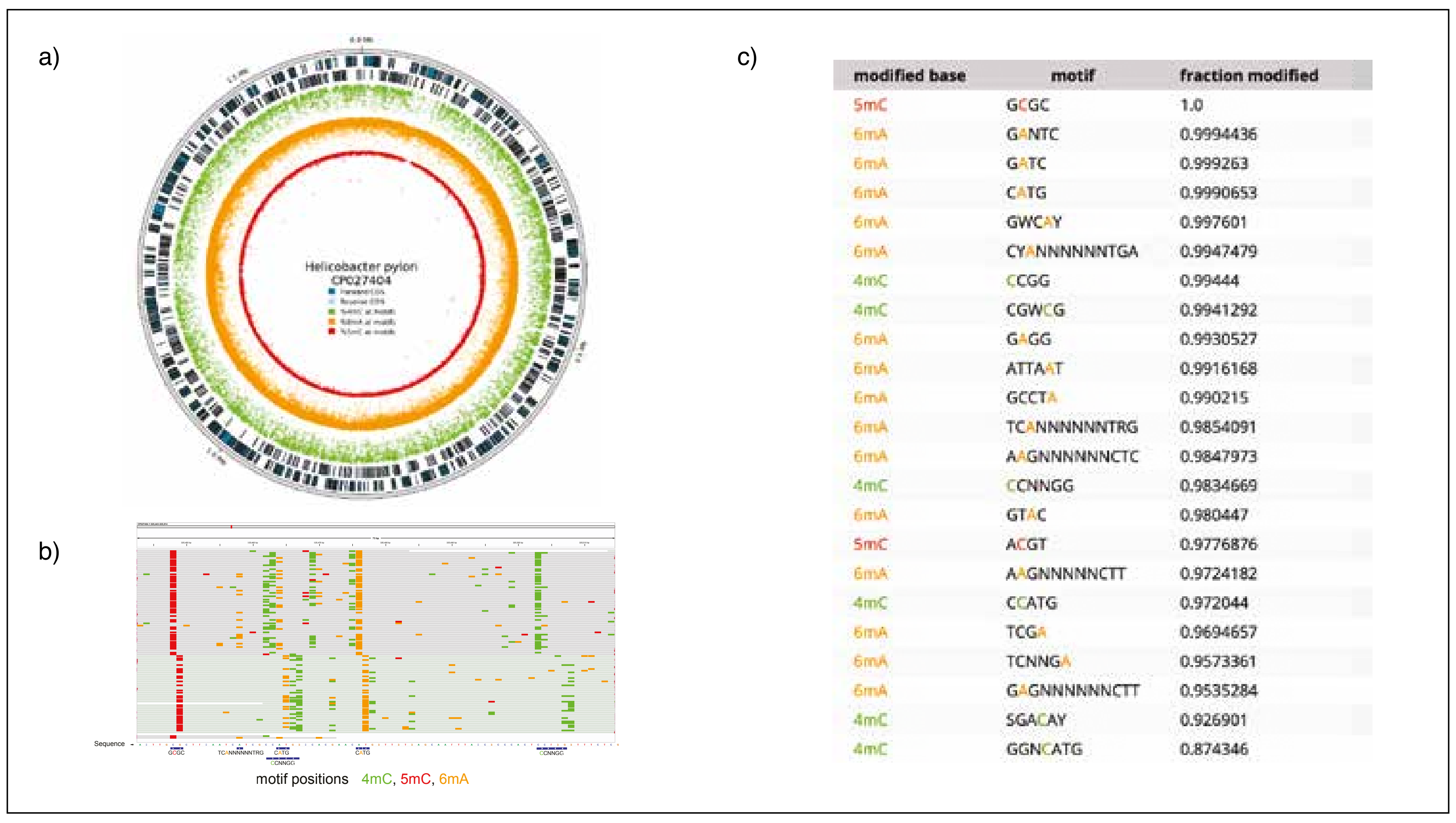


Fig. 3 Genome browser showing automatic detection of differentially methylated regions (DMRs), and high scoring CpG island

Explore differentially methylated regions with automatic segmentation, using 'dmr' and 'entropy'

The Modkit tool suite contains flexible and intuitive differential methylation analysis algorithms that leverage multiple modification types (e.g. 5mC and 5hmC). A bedMethyl table (generated by 'pileup') along with regions of interest (for example, CpG islands) can be quickly scanned for differential methylation, ranked, and triaged. Emphasis has been placed on making the output intuitive and enabling exploratory data analysis. Alternatively, *de novo* segmentation can be performed, as shown for the COLO829 tumor-normal research samples (Fig. 3a); differentially methylated regions are highlighted in purple. Direct RNA sequencing of HEK296 samples revealed a distribution of methylation entropy (Fig. 3b), with an example high-entropy transcript shown here (Fig. 3c).



modified base	motif	fraction modified
5mC	GCGC	1.0
6mA	GANTC	0.9994436
6mA	GATC	0.999263
6mA	CATG	0.9990653
6mA	GWCAY	0.997601
6mA	CYANNNNNNTGA	0.9947479
4mC	CCGG	0.99444
4mC	CGWCG	0.9941292
6mA	GAGG	0.9930527
6mA	ATTAAT	0.9916168
6mA	GCCTA	0.990215
6mA	TCANNNNNNTRG	0.9854091
6mA	AAGNNNNNNCTC	0.9847973
4mC	CCNNGG	0.9834669
6mA	GTAC	0.980447
5mC	ACGT	0.9776876
6mA	AAGNNNNNCTT	0.9724182
4mC	CCATG	0.972044
6mA	TCGA	0.9694657
6mA	TCNNGA	0.9573361
6mA	GAGNNNNNNCTT	0.9535284
4mC	SGACAY	0.926901
4mC	GGNCATG	0.874346

Fig. 4 Genome browser showing discovered methylation motifs and example output table

Automatically find sequence motifs enriched for methylation with 'find-motifs'

Methyltransferase enzymes will methylate specific sequence motifs. Nanopore sequencing can detect the most common prokaryotic modifications (4mC/5mC and 6mA) at the same time, allowing the discovery of adenine and cytosine methyltransferase motifs. In (Fig. 4a), we show a browser shot of *Helicobacter pylori* sequencing reads with 4mC, 5mC, and 6mA discovered motifs. In an example table from the same organism (Fig. 4b), 23 *de novo* discovered motifs are seen, with varying levels of modification and specificity.