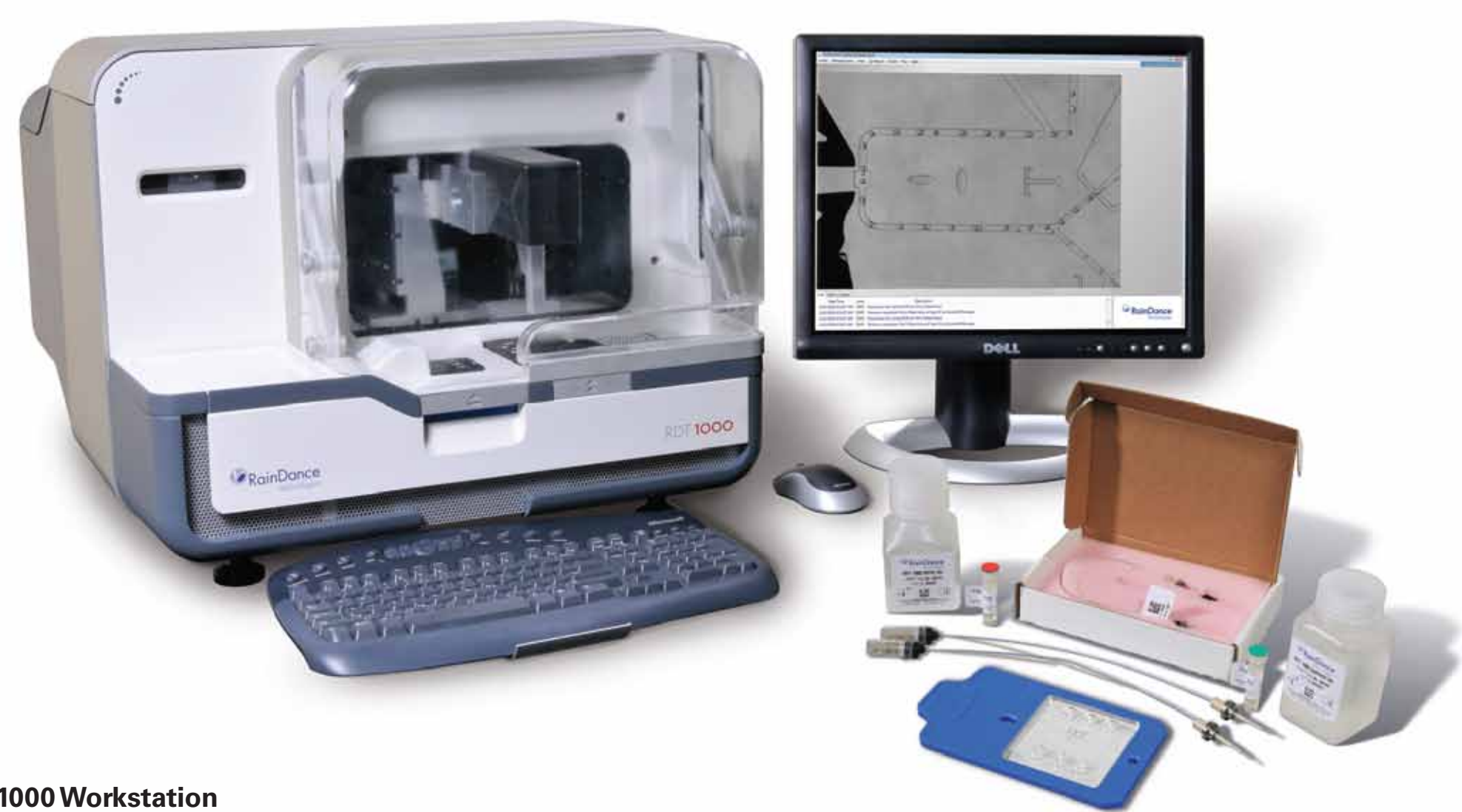


# CpG Methylation Analysis from Targeted Sequencing of Bisulfite Converted DNA

Jason Warner<sup>1</sup>, Steve Kotsopoulos<sup>1</sup>, Michael Samuels<sup>1</sup>, Jochen Hampe<sup>2</sup>, Alexander Herrmann<sup>2</sup>, Kathryn Stemshorn<sup>2</sup>, Karol Szafranski<sup>2</sup>, Peter Nuernberg<sup>2</sup>, Darren Link<sup>1</sup>, Jeff Olson<sup>1</sup>, <sup>1</sup>RainDance Technologies, Inc., Lexington, MA, <sup>2</sup>University of Kiel, Kiel, Germany.

## OVERVIEW



RDT 1000 Workstation

RainDance Technologies (RDT) has extended the capabilities of the RDT 1000 to target genomic regions of bisulfite converted DNA. This approach used in conjunction with high-throughput sequencing enables researchers to measure the methylation status of targeted regions of the genome with complete sequence coverage, specificity, and uniformity.

Recent advances in high-throughput sequencing technologies have enabled the analysis of differential methylation patterns at a genome-wide scale. These genome-wide approaches have enabled the discovery of epigenetic variations associated with disease progression, including cancer. To validate these candidate regions with a sufficient number of samples, a targeted approach is required to maximize the breadth and the depth of coverage for these regions.

Targeted bisulfite sequencing data provides a unique set of challenges. The majority of bisulfite-treated DNA is devoid of cytosines, which creates difficulties when mapping the short sequencing reads from Next-Generation Sequencing platforms. Due to the nature of a bisulfite converted genome, the sense and antisense strands are no longer complementary following bisulfite conversion, which further increases the size and complexity of the sequence space.

RainDance has developed a Targeted Bisulfite Sequencing Analysis Pipeline designed to address these challenges by supporting standard genomic sequence assemblies of bisulfite-treated DNA. Our approach utilizes any standard FASTQ format that is pre-processed using a custom RainDance Perl script. The processed FASTQ file is inputted into the CLC Genomics Workbench from CLC bio for sequence alignment to produce a standard SAM assembly. The SAM file is then further processed with another custom RainDance Perl script to produce the final data output.

The RainDance Targeted Bisulfite Sequencing Analysis Pipeline was demonstrated to enable the quantitative analysis of CpG methylation regions with high completeness (C1 of 97.3%; AVoC of 313), specificity (76.8% of mapped reads were on target) and uniformity (base coverage at 0.2x of AVoC was 75.9%) of alignment of sequencing reads.

## Targeted Bisulfite Resequencing

RainDance Technologies has developed an enrichment platform for targeted resequencing that leverages the sensitivity and the specificity of PCR to target genomic regions from either thousands of exons or large contiguous loci. The Sequence Enrichment application utilizes a novel microdroplet-based format to rapidly and reproducibly generate more than 1 million independent PCR reactions. This poster outlines a strategy for the analysis of targeted bisulfite sequencing data to maximize the specificity and accuracy of the alignment of sequencing reads for quantitative analysis of CpG methylation.

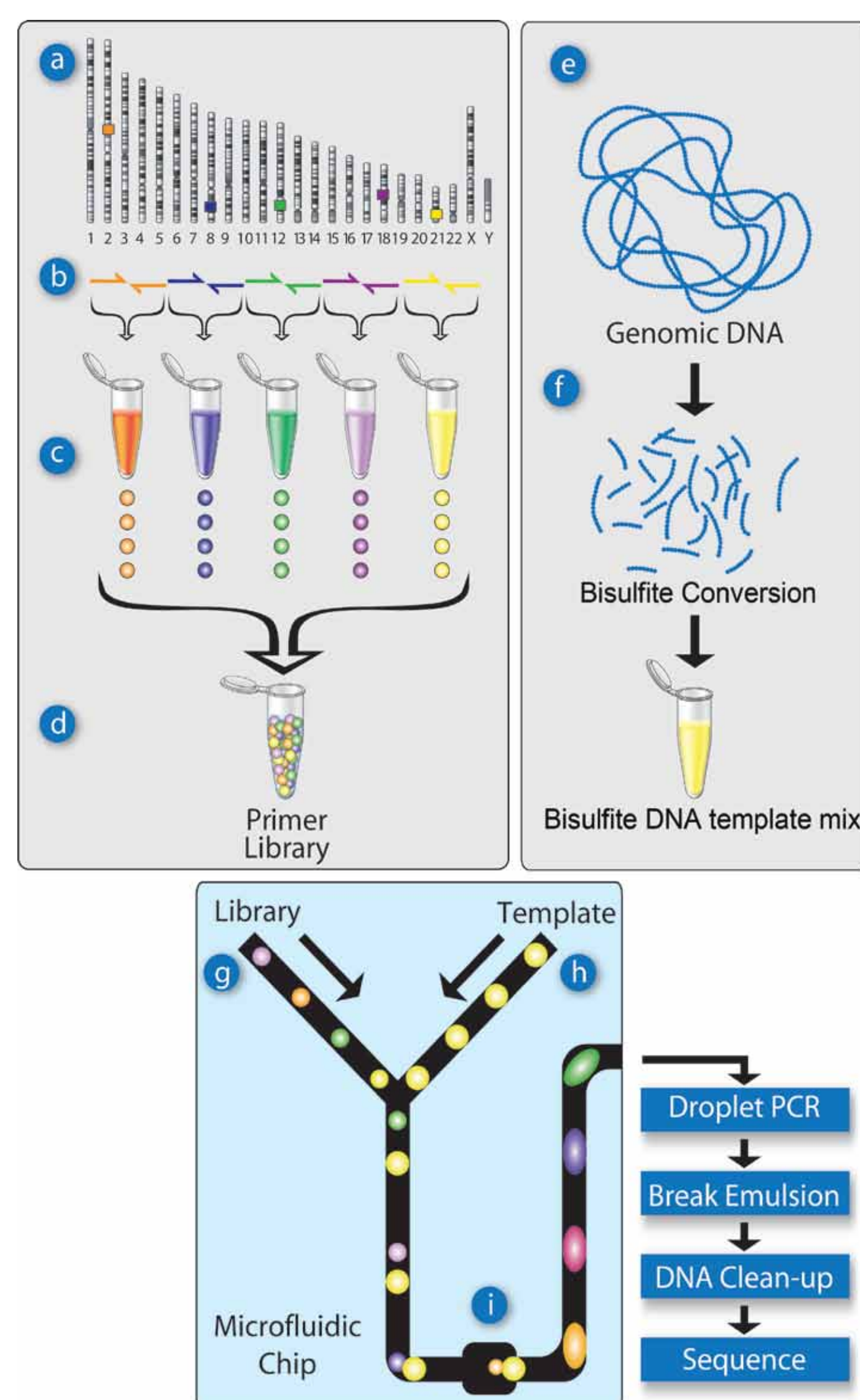
### RDT 1000 Targeted Resequencing Workflow<sup>1</sup>

#### Primer Library Generation

- 1 Identify targeted CpG regions of interest in the genome.
- 2 Design and synthesize forward and reverse primer pairs for each target, using an in silico bisulfite converted sense or antisense sequence as a template.
- 3 Generation of primer pair droplets. A microfluidic chip is used to encapsulate the aqueous PCR primers in inert fluorinated carrier oil with a block-copolymer surfactant to generate the equivalent of a picoliter scale test tube compatible with standard molecular biology.
- 4 Primer pair droplets are mixed together so that each library element has an equal representation.

#### Bisulfite-Treated Genomic DNA Template Mix Preparation

- 5 Genomic DNA is treated with sodium bisulfite to convert non-methylated cytosines to uracil while methylated cytosines remain unchanged.
- 6 The bisulfite-treated DNA is purified and mixed together with all of the components of the PCR reaction except the PCR primers.



#### Primer-template merge and PCR

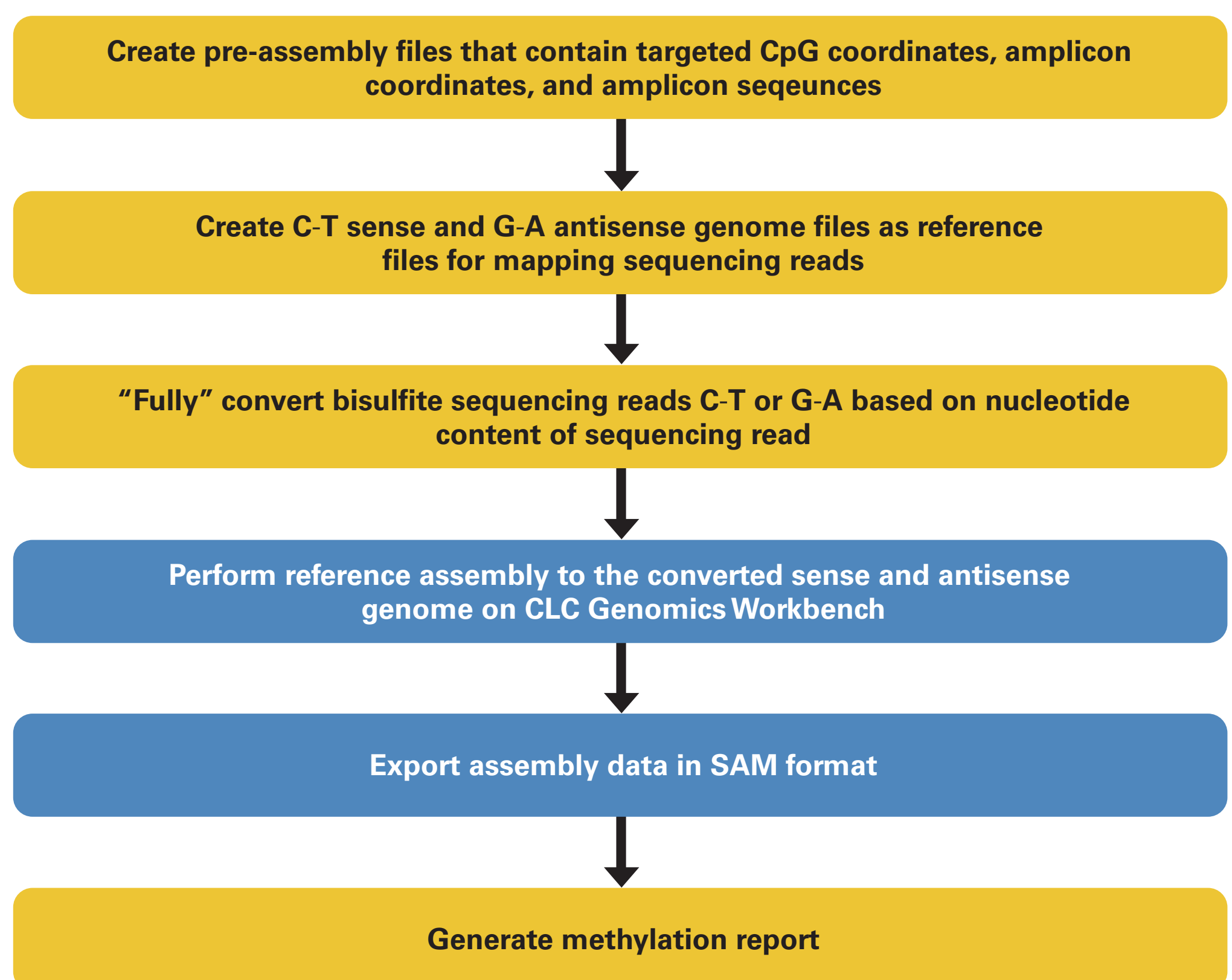
- 1 Primer Library droplets (8pL) are dispensed to the microfluidic chip.
- 2 Bisulfite-treated Genomic DNA Template is delivered as an aqueous solution and template droplets (18pL) are formed within the microfluidic chip. The primer pair droplets and template droplets are then paired together in a 1:1 ratio.
- 3 Paired droplets flow through the channel of the microfluidic chip to pass through a merge area where an electric field induces the two discrete droplets to coalesce into a single PCR droplet (26 pL). Approximately 1 million PCR droplets are collected into a single 0.2 ml PCR tube.
- 4 The collection of PCR droplets is processed in a standard thermal cycler for targeted amplification, followed by breaking the emulsion of PCR droplets to release the PCR amplicons into solution for purification and next-generation sequencing.

## Analysis Workflow

### Workflow Overview

Two of the main challenges of mapping bisulfite sequencing reads are: 1) sequenced bisulfite DNA contains sense and antisense sequences that are no longer complementary, meaning they will map to reference DNA differently. 2) Both the sense and antisense strands are mostly devoid of cytosines, which mean they will not map without mismatches to a standard reference genome.

Our solution to these challenges is outlined below and similar to the approach used by Deng, et al.<sup>2</sup> Analysis steps using CLC Genomics Workbench from CLC bio are shown in blue and analysis steps using RainDance Perl scripts are shown in yellow.



### Sequencing Statistics

Sequencing Platform ..... 454  
Average Read Length ..... 317 nt  
Target Size ..... 292 Kb

Total Reads	Mapped Reads	Target Reads (Specificity)
57,362	44,993	78.4%
	34,637	76.8%

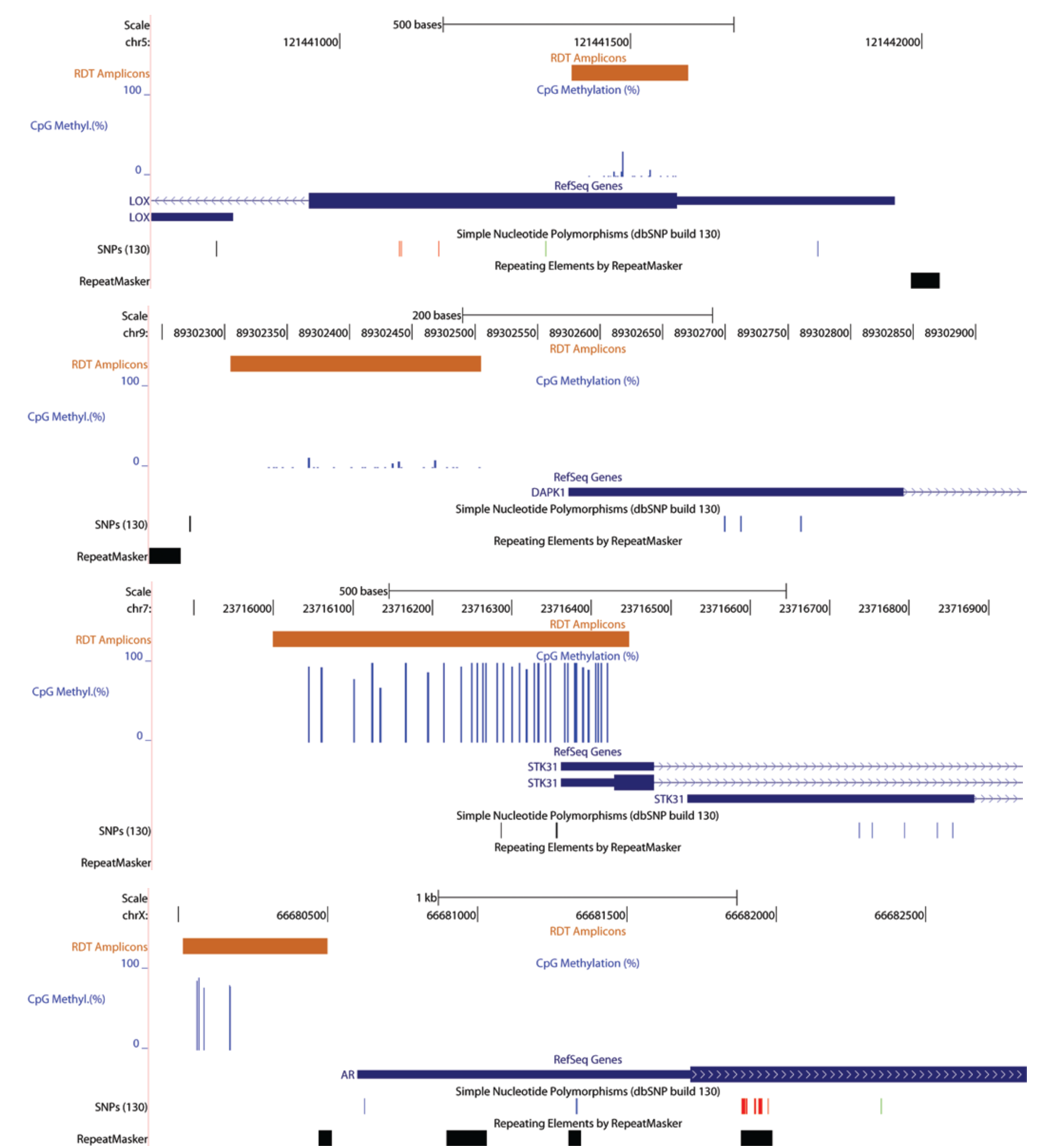
  

Bisulfite Conversion	CpG AD0C	C1	C20	C100	Base Coverage above 2x
>98%	313	97.3%	88.9%	69.2%	75.9%

The RDT Targeted Bisulfite Resequencing Workflow and Analysis pipeline yielded very high completeness (C1 of 97.3%; AVoC of 313), specificity (76.8% of mapped reads were on target), and uniformity (base coverage at 0.2x of AVoC was 75.9%).

## High Resolution CpG Methylation

Shown below are a few examples of typical methylation patterns observed around transcriptional start sites using the RDT Methyl-Seq workflow. RDT amplicon targets are shown in orange and CpG methylation percentage is shown as a blue bar graph. Note that a high or low methylation pattern is consistent in each transcriptional region.



## Conclusions

Our strategy for analysis of targeted bisulfite sequencing data maximized the specificity and accuracy of the alignment of sequencing reads to enable quantitative analysis of CpG methylation. The RDT Targeted Bisulfite Resequencing Workflow and Analysis pipeline yielded very high completeness (C1 of 97.3%; AVoC of 313), specificity (76.8% of mapped reads were on target), and uniformity (base coverage at 0.2x of AVoC was 75.9%).

The statistical power of the microdroplet-based PCR format will enable highly quantitative measurement of CpG methylation using targeted resequencing of bisulfite-converted template DNA.

## References

- <sup>1</sup>Tewhey R, Warner JB, Nakano M, Libby B, Medkova M, et al. (2009) *Microdroplet-based PCR enrichment for large-scale targeted sequencing*, *Nature Biotechnology* 27(11):1025-1031.
- <sup>2</sup>Deng J, et al. (2009) *Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming*, *Nature Biotechnology* 27(4):353-360.