Epigenome Characterization of Human Genomes using the PacBio Platform

Jonas Korlach¹, Yuta Suzuki², Marie Classon³, Vasantharajan Janakiraman³, Eric W. Stawiski³, Matthew Boitano¹, Jason Chin¹, Khai Luong¹, Stephen W. Turner¹, Steffen Durinck³, Somasekar Seshagiri³ and Shinichi Morishita²

¹Pacific Biosciences Inc., Menlo Park, CA 94025, ²University of Tokyo, Tokyo, Japan, and ³Molecular Biology, Bioinformatics, and Department of Oncology, Genentech Inc., South San Francisco, CA 94080

Abstract

In addition to the genome and transcriptome, epigenetic information is essential to understand biological processes and their regulation, and their misregulation underlying disease. Traditionally, epigenetic DNA modifications are detected using upfront sample preparation steps such as bisulfite conversion, followed by sequencing. Bisulfite sequencing has provided a wealth of knowledge about human epigenetics, however it does not access the entire genome due to limitations in read length and GC-bias of the sequencing technologies used.

In contrast, Single Molecule, Real-Time (SMRT) DNA Sequencing is unique in that it can detect DNA base modifications as part of the sequencing process. It can thereby leverage the long read lengths and lack of GC bias for more comprehensive views of the human epigenome. I will highlight several examples of this capability towards the generation of new biological insights, including the resolution of methylation states in repetitive and GC-rich regions of the genome, and large-scale changes in the methylation status across a cancer genome as a function of drug sensitivity.

Background

Cancer Epigenome Study Design

• PC-9 lung cancer cell line, studying drug susceptibility
  • Methyltransferase (MT) inhibitor reverts cells to drug-sensitive
  • Utilized PacBio Sequencing for:
    • Generate de novo cancer genome assembly
    • Characterize cancer genome structural variation
    • Detect gene fusions
    • Characterize genome-wide methylome

Differential methylation status of CpG islands inferred genome-wide from PacBio Sequencing data, algorithm at https://github.com/hacone/AgIn

More hypermethylated CpG islands in drug-resistant sample

Examples:

Chr. 4: ANKRD17 (already implicated in breast cancer)

Chr. 4: FGFR3 (fibroblast growth factor 3)

Chr. 4: WHSC1 (already implicated in myelomas)

Chr. 4: RASSF6 (tumor suppressor gene)

References


Acknowledgements

The authors would like to thank everyone who helped generate data for the poster.