Epigenome Characterization of Human Genomes using the PacBio Platform

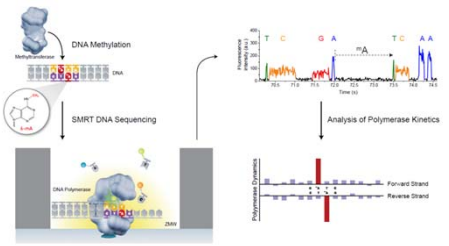
Jonas Korlach[1], Yuta Suzuki[2], Marie Classon[3], Vasantharajan Janakiraman[3], Eric W. Stawiski[3], Matthew Boitano[1], Jason Chin[1], Khai Luong[1], Stephen W. Turner[1], Steffen Durinck[3], Somasekar Seshagiri[3] and Shinichi Morishita[2]
[1]Pacific Biosciences Inc., Menlo Park, CA 94025, [2]University of Tokyo, Tokyo, Japan, and [3]Molecular 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.

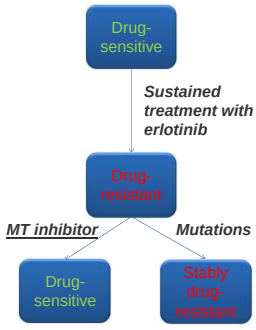
In contrast, Single Molecule, Real-Time (SMRT) DNA Sequencing is unique in that it can detect DNA base modifications as part of the sequencing process. It can thereby leverage the long read lengths and lack of GC bias for more comprehensive views of the human epigenome. I will highlight several examples 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

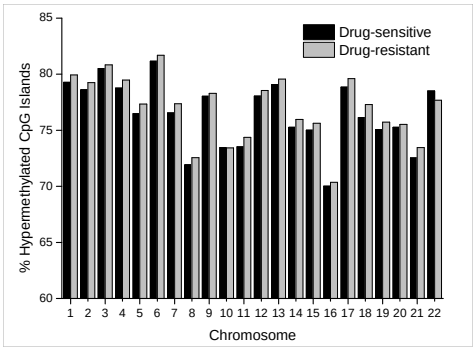
- PC-9 lung cancer cell line, studying drug susceptibility
- Methyltransferase (MT) inhibitor reverts cells to drug-sensitive



- Utilized PacBio Sequencing for:
  - Generate de *novo* cancer genome assembly
  - Characterize cancer genome structural variation
  - Detect gene fusions
  - **Characterize genome-wide methylome**

Cancer Epigenome

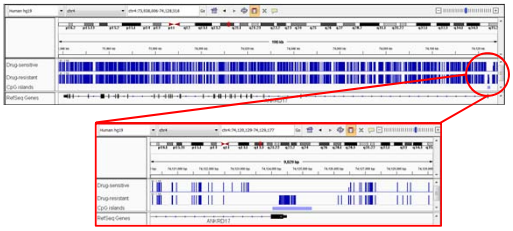
Differential methylation status of CpG islands inferred genome-wide from PacBio Sequencing data, algorithm at https://github.com/hacone/AgIn



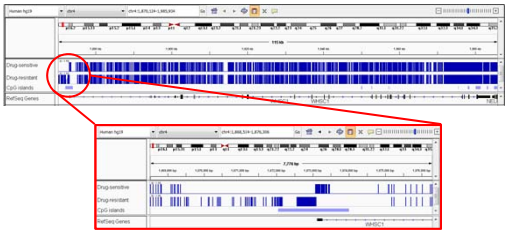
More hypermethylated CpG islands in drug-resistant sample

Examples:

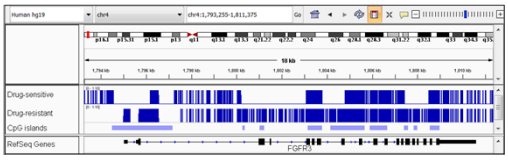
Chr. 4: ANKRD17 (already implicated in breast cancer)



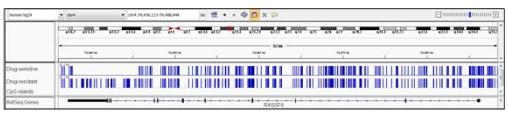
Chr. 4: WHSC1 (already implicated in myelomas)



Chr. 4: FGFR3 (fibroblast growth factor 3)

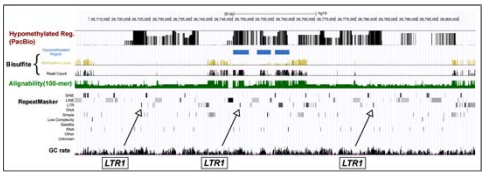


Chr. 4: RASSF6 (tumor suppressor gene)

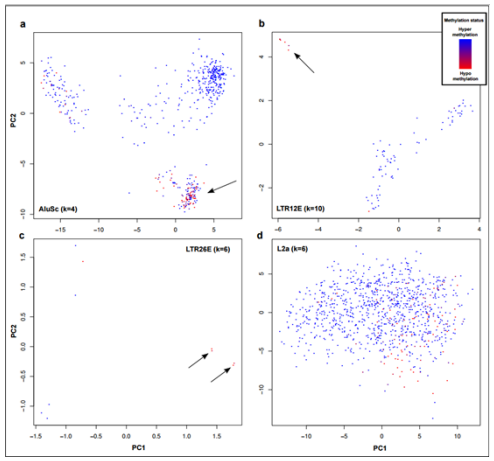


Resolving Repetitive Regions[1]

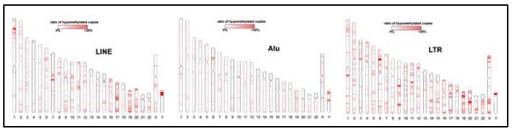
Example of hypomethylated repeat occurrences (chr. 6):



PCA analysis for 4 selected classes of repetitive elements:



Genome-wide distribution of hypomethylated repetitive elements:



Summary of methylation status on repetitive elements:

Class	With >9 CpGs (A)	Covered (B)	B/A	Covered with >5x (C)	C/A	Hypomethylated (D)	D/C
LINE/L1	50795	50127	98.7%	45379	89.3%	356	0.8%
LINE/L2	4977	4961	99.7%	4637	93.2%	244	5.3%
LINE/CR1	178	178	100.0%	165	92.7%	5	3.0%
LINE/RTE-X	65	64	98.5%	60	92.3%	1	1.7%
SINE/Alu	238701	235527	98.7%	214341	89.8%	2282	1.1%
SINE/MIR	374	371	99.2%	343	91.7%	169	49.3%
LTR/ERV1	19638	19354	98.6%	17739	90.3%	348	2.0%
LTR/ERVK	5175	5079	98.1%	4603	88.9%	87	1.9%
LTR/ERVL	4395	4350	99.0%	3991	90.8%	82	2.1%
LTR/ERVL-MaLR	4366	4327	99.1%	3933	90.1%	69	1.8%
LTR/Gypsy	108	104	96.3%	89	82.4%	9	10.1%
Retroposon/SVA	2906	2796	96.2%	2427	83.5%	3	0.1%
DNA/hAT-Blackjack	83	83	100.0%	75	90.4%	2	2.7%
DNA/hAT-Charlie	1460	1452	99.5%	1342	91.9%	55	4.1%
DNA/hAT-Tip100	326	322	98.8%	305	93.6%	19	6.2%
DNA/MULE-MuDR	92	92	100.0%	89	96.7%	2	2.2%
DNA/PiggyBac	57	55	96.5%	52	91.2%	1	1.9%
DNA/TcMar-Mariner	384	384	100.0%	360	93.8%	1	0.3%
DNA/TcMar-Tigger	2821	2801	99.3%	2649	93.9%	43	1.6%
rRNA	68	66	97.1%	66	97.1%	8	12.1%
Simple_repeat	6256	6191	99.0%	5434	86.9%	3849	70.8%
Low_complexity	1068	1064	99.6%	942	88.2%	789	83.8%

References

[1]Suzuki *et al.* (2015) Landscape of CpG methylation of individual repetitive elements.
http://biorxiv.org/content/early/2015/04/24/018531

Acknowledgements

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