H028

CEBPA Mutation Phasing Using Pacific Biosciences Circular Consensus Sequencing

Li Cai[1], Scott Parker[1], Kimberley Wagner[1], Lax Iyer[2], Ingrid Chen[2], Dan Wang[1], Cheryl Heiner[3], Roberto Lleras[3], Ian McLaughlin[3], George Yuan[3], Marcia Eisenberg[1], Brian Krueger[1], Anjen Chenn[1]

1.Center for Molecular Biology and Pathology, LabCorp, 2. Center for Bioinformatics, LabCorp, 3. Pacific Biosciences

Integrated ONCOLOGY
LabCorp Specialty Testing Group

Background

The CEBPA (CCAAT/enhancer binding protein alpha) gene encodes a transcription factor important for granulocyte differentiation. CEBPA mutations are found in 6-15% of de novo acute myeloid leukemia (AML) cases and in 15-18% of AML cases with normal karyotypes. CEBPA biallelic mutations, which are defined by the presence of a mutation on both alleles of the CEBPA gene, are generally associated with a favorable clinical outcome and lead to improved overall survival because these mutations usually lead to differentiation arrest in the absence of a wild type copy. Because CEBPA is a single exon gene with a coding sequence of 1077bp, long read sequencing is required for variant phasing. In this work, Pacific Biosciences' single molecule, real-time (SMRT) Circular Consensus sequencing (CCS) was used to phase the variants in 26 CEBPA positive samples with two or more mutations and 800 CEBPA negative samples or positive samples with one mutation.

Methods

26 CEBPA specimens with two or more mutations identified by Sanger sequencing were selected along with 800 CEBPA negative samples and positive samples with one mutation. Genomic DNA was purified from blood or bone marrow specimens. Amplification primers were designed with M13 tags to target a 1.2kb region of CEBPA that encompassed the coding sequence. Asymmetric barcodes were added onto the M13 tags after a second round of PCR before SMRT Bell ligation, amplicon pooling, and sequencing. CCS reads were generated with a read quality of >99.9% and a minimum read depth of 100x. CCS reads were aligned to the hg19 human genome reference using pbmm2 and phasing of variants was performed by a custom analysis pipeline. As a final step of the analysis, variants were visualized in the Integrative Genomics Viewer (IGV).

Results

Our results suggest that long read sequencing can phase variants in the CEBPA gene. Results of CCS reads were concordant with the Sanger sequencing results. No variants of significance were found in the negative samples.

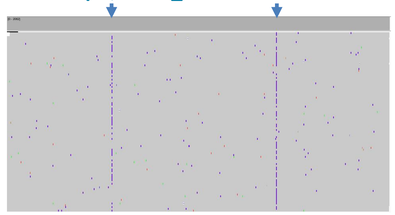
Positives Comparison to Sanger Calls

Sample	Sanger Call	PacBio Call	Cis or Trans	WT Frequency	PacBio Coverage
1	c.318dupT; c.939_940ins30	Y	Trans	0.37	1730
2	c.296_347del52; c.934_936dupCAG	Y	Trans	0.12	1022
3	c. 247C>T & c.988C>T	Y	Trans	0.15	1792
4	c.247delC; c.923_925dupTGG	Y	Cis	0.32	1488
5	c.250delG/c.958_959ins36	Y	Trans	0.55	1276
6	c.247delC; c.441C>G	Y	Trans	0.39	1656
7	c.250delG/c.958_959ins36	Y	Trans	0.48	1528
8	c.284_293delTGGGCCCCAC; c.950_952dupTGA	Y	Cis	0.06	1445
9	c.318dupT/c.939_940ins30	Y	Trans	0.37	1413
10	c.206delA; c.934_936dupCAG	Y	Trans	0.19	1678
11	c.247delC/c.923_925dupTGG	Y	Trans	0.32	1762
12	c.853_854insCG, c.971T>A	Y	Trans	0.01	570
13	c.68dupC 2. c.971T>C	Y	Trans	0.13	1903
14	c.187G>A;2.c.393C>G	Y	Trans	0.07	1593
15	c.324C>A 2.c.910_912dupAAG	Y	Cis	0.42	2291
16	c.533TdelT, c.93delC	Y	Trans	0.3	1125
17	**c.266dupA; c.912_913insTTG**	**Y**	**Cis**	**0.08**	**1993**
18	**c.945_946ins33, c.102_108delCCGGGGC**	**Y**	**Trans**	**0.03**	**787**
19	**c.68_78del11, c.906_920dup15**	**Y**	**Trans**	**0**	**1219**
20	c.68delC, c.924_925insAGGGGG	Y	Trans	0.76	1241
21	c.113dupG; c.929_934delCGCAGC	Y	Trans	0.07	1674
22	c.68delC, c.924_925insAGGGGG	Y	Trans	0.82	1714
23	c.209dupC; c.907_933dup27	Y	Cis	0.2	1105
24	c.68dupC, 2. c.937_939dupAAG	Y	Trans	0.15	1766
25	**c.158_164delGCGGCAT; c.937_939dupAAG**	**Y**	**Cis**	**0**	1637
26	c.180_183dupGTCC	Y	N/A	0.75	1597
27	c.383dupC	Y	N/A	0.8	2329
28	c.176dupA/c.470dupC & p.T60Dfs/p.L158Afs	Y	Trans	0.16	1621

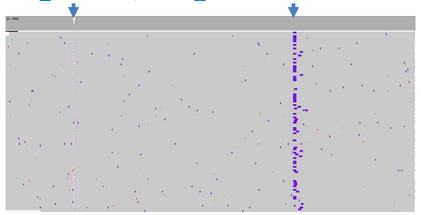
Table 1: 28 positive samples were compared to their Sanger result
The frequency of non-variant reads (WT) was calculated for all samples with 2 or more variants. This was done by phasing variants within the reads and calculating the frequency of WT reads in each sample.

IGV Inspection of Select Samples

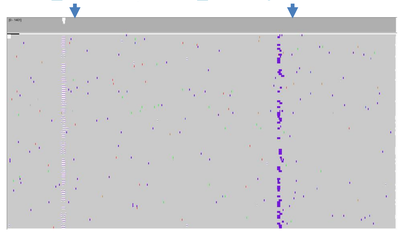
c.266dupA; c.912_913insTTG - Cis



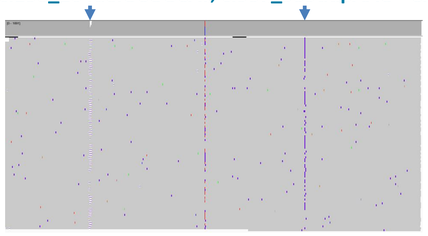
c.945_946ins33, c.102_108delCCGGGGC - Trans



c.68_78del11, c.906_920dup15 - Trans



c.158_164delGCGGCAT; c.937_939dupAAG - Cis



Conclusions

Pacific Bioscience CCS reads are a powerful new tool for phasing variants in critical clinical samples, such as those from patients suffering from AML that have variants in CEBPA. Very high multiplexing of target amplicons on a SMRT cell now enables us to elevate the standard of care provided to patients in the clinic by revealing the phase of these mutations in CEBPA and communicate to the patient whether these mutations afford them a prognostic advantage.

LabCorp