SureSelect Target Enrichment System for Sequencing on the Ion Torrent PGM

Application Note

Authors

Julien Tarabeux[1]

Jean Muller[1,2]

[1]Laboratoire de Diagnostic Génétique
Hôpitaux Universitaires de Strasbourg
Strasbourg, France.

[2]Laboratoire de Génétique Médicale
Institut de Génétique Médicale d'Alsace
Fédération de Médecine Translationnelle
de Strasbourg (FMTS)
Université de Strasbourg
Strasbourg, France

Abstract

SureSelect, a hybridization-based target enrichment solution, is sequencer agnostic and available for a variety of NGS sequencers. There have been few instances of its use with the Ion Torrent PGM (Personal Genome Machine) sequencer.

Introduction

To date, most clinical research laboratories are equipped with next-generation sequencers, a large proportion of which are from ThermoFischer Scientific. These systems include the Ion Torrent PGM, the Ion Torrent Proton and more recently the Ion S5 and Ion S5 XL. These sequencers offer remarkable performance in terms of speed, cost, read length and flexibility with multiple output modes (e.g., the PGM has 3 chips 314, 316 or 318). They are also well suited for clinical research laboratories requiring a high level of flexibility in sample throughput ranging from a couple to 100 a week.

In this Application Note, the SureSelect for Ion Proton protocol was adapted to the PGM sequencer. Because the PGM is able to produce longer reads than the Ion Proton, some modifications were made in order to improve the "per run sequencing capacity." The protocol was also shortened with the removal of a size selection step, which was found to be unnecessary for the PGM. Finally, samples were multiplexed before hybridization to reduce hands on time and per sample cost.

Agilent Technologies

Materials and Methods

Panel design

Baits were designed covering the exons (+/- 20bp into the intronic region) of a set of 58 genes implicated in the Bardet-Biedl Syndrome and related pathologies using the advanced mode of SureDesign, a free online design software (https://earray.chem.agilent.com/suredesign/). The total number of probes contained in the design was 16,730, spanning a total of 336.597 kilobase pairs (kbp). In total, >99% of our target was covered by the design. The <1% that was uncovered was located in the UTR due to the low complexity of this region. The main purpose of this design was to identify point mutations and indels within the coding sequences of 58 genes. The design report for the library provided by SureDesign recommended a minimum sequencing per sample of 67.4 million base pairs (Mbp) to achieve an average read depth of between 30-40X with at least 80% of the targeted bases covered at 20X. Based on this recommendation, we anticipated that we can multiplex a maximum of 8 samples, based on the average output of the Ion PGM 318 v2 chips in our laboratory.

Sample preparation

Twenty different samples were processed in this study. In all cases the last size selection step described in the SureSelect Target Enrichment for the Ion Proton™ Next-Generation Sequencing System protocol (version B0) was skipped. A total of 4 runs were performed in order to optimize and assess protocol modifications.

Four samples were processed strictly following the SureSelect Target Enrichment for the Ion Proton™ Next-Generation Sequencing System (Run1). Eight samples, including the four above, were processed following the Agilent protocol on the Ion Proton, with the exception that samples were pooled prior to capture (750ng in total, 93.75ng of each DNA, Run2). This was done to evaluate pre-capture pooling and determine the maximum number of samples to include in one run.

In order to take advantage of the PGM read length capacity, the following modifications were then evaluated to increase read length of the library. The enzymatic shearing time was decreased from 50 to 30 minutes. The Agencourt AMPure XP beads ratios in the first size selection step were modified to fit the increased size of the selected fragments. The first volume of the bead was adjusted to from 60.5µl to 38.5µl and the second volume was adjusted from 38.5µl to 24µl.

Twelve samples in total were processed using these modifications. Six samples were processed with one sample per capture (Run3 and Run4). Two samples were pooled together into one capture (Run4), and four samples were pooled together into one capture (Run4).

Sequencing and analysis

Samples were subjected to emulsion PCR using the Ion One Touch 2 system and the Ion PGM Hi-Q OT2 kit in 400 bp mode. Ion sphere particles (ISP) were enriched using the E/S module and were sequenced with an Ion PGM in a 400-bp configuration run. All samples were processed on the Ion 318 v2 chips (Table 1). Table 2 provides the description of the experiment and the resulting mapping statistics. Sequencing data was analyzed using the Torrent Suite Software v4.4.2.

Different final dilutions of the library were tested for sequencing (Table1). Including all of the modifications in the protocol, the total output was increased by 34% (516Mb in Run1 to 691Mb in Run4)

Results and Discussion

Performance of the unmodified protocol (with the exception of the last size-selection step) was first tested with 4 samples (Run1: samples 1- 4). The % on-target reads ranged from 21.66% to 32.68% (Table 2). Mean read length ranged from 107 to 124bp and mean depth of coverage was between 65X and 143X. Uniformity, defined as the percent of all target bases covered by at least 0.2x the average base depth coverage, ranged from 98.21% to 98.71%. When pre-capture pooling was performed with 8 samples (Run2), a decrease of both the mean depth coverage (ranging from 26.3X to 33X) and the % on-target reads (ranging from 15.23% to 22.95%) was observed versus Run1 as expected, whereas the uniformity (ranging from 96.8% to 97.71%) decreased only slightly.

In Run1, the total number of single nucleotide variant per sample (SNV) ranged from 206 to 274 (Table 3). In Run2, SNV ranged from 195 to 262 variants per sample. The increased number of samples per run as well as the additional pre-capture pooling step, did not significantly impact the number of variants detected per sample. Moreover, the concordance of the variants detected from sample 1 to sample 4 between Run1 and Run2 was above 86% on average (Table 3). This confirmed that the addition of pre-capture pooling to the SureSelect protocol did not affect the outcome on the PGM.

	Final dilution tested	Chip loading percentage	Polyclonal ISP*	Low Quality ISP*	Total number of bases
Run 1	22pM	64%	35%	4%	516Mb
Run 2	26pM	66%	40%	8%	453Mb
Run 3	16pM	73%	52%	12%	481Mb
Run 4	13pM	77%	40%	7%	691Mb

*ISP = Ion Sphere Particles

Table 1. Sequencing statistics

The modifications made to the protocol to increase read length were successful for the tested samples in Run3 and Run4, compared to the standard condition in Run1, as illustrated in Figure 1. Read length ranged from 126 to 184bp for Run3 and Run4 samples, an increase of 46% compared to Run1. Pre-capture pooling in Run2 and Run4 did not have a negative impact on read length, compared to Run1 (Table 2).

A significant increase in the percent of on-target reads for the samples in Run3 was observed, compared to Run1. Run3 samples (with no pre-capture pooling) generated 44.68% to 48.61% on-target reads, versus 21.66% to 32.68% for Run1 (Table 2). These results indicate that the protocol modifications made in Run3 and Run4 to the SureSelect Target Enrichment for the Ion Proton™ Next-Generation Sequencing System (see Sample preparation) doubled the % on-target for almost all Run3 samples. Unexpectedly, samples 13 and 14 of Run4 did not display similar results, as the % on-target reads with no pre-capture pooling were similar for Run4 (20.42% to 31.33%) and Run1.

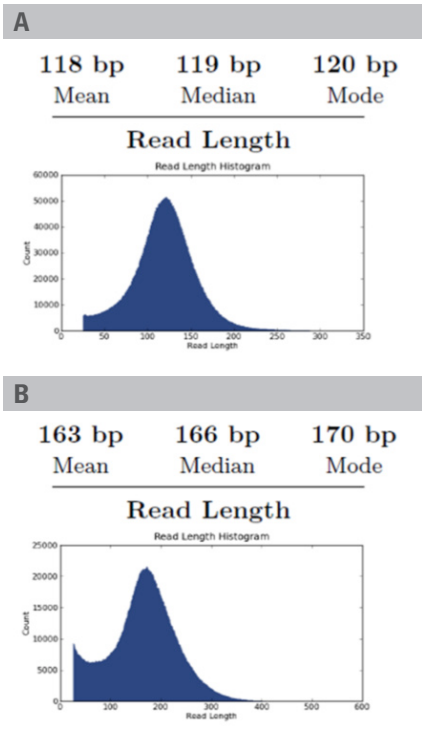

Figure 1. Read length comparison without modification of the protocol (A: Run1) versus with the modifications (B: Run3)

Additional investigation will be required (i.e. running the protocol modifications with a different panel) to verify the impact to % on-target reads seen in Run 3.

Downstream data analysis showed a total of 948 non-redundant variants observed in the 20 samples analyzed. Of these, 287 were found to be both homozygous and heterozygous, while only 608 were heterozygous only and 53 were homozygous only. On average, each sample carried 236 variants. As expected, the probe hybridization enrichment methods were highly reproducible across all samples. This was the case in spite of the modifications made to the protocol, especially the pre-capture pooling (Figure 2A). We were able to identify several pathogenic mutations in patients harboring either the BBS or Alström phenotype. For example, a heterozygous SNV and an indel in *BBS2* (Figure 2B) and a homozygous deletion in *ALMS1* (Figure 2C) were detected.

In general, results were in accordance with expectations. However, the lower % on-target was likely due to the small size of the panel and the specificities of the design. Variability was observed in the % on-target reads between experiments, but this might not be related to the changes in the protocol. For example, this variability was observed for samples that were prepared following the same conditions (samples 9 to 12 versus samples 13 and 14).

In order to increase the output per run, an attempt was made to increase read length by modifying the shearing incubation time and the first size selection step (Run3 and Run4). On average, read length was increased by 46%, and there may still be room for improvement.

#Run	#Sample	Pre-capture pooling	Mapped Reads	Mean Read Length	Mean Depth	On-Target Reads (%)	Uniformity (%)
Run 1	Sample 1	No	694808	124	78.9	32.68	32.68
	Sample 2		1353082	118	143.2	31.70	31.70
	Sample 3		1297881	123	121.7	27.24	27.24
	Sample 4		977973	107	64.77	21.66	21.66
Run 2	Sample 1	8 per capture	382797	141	22.95	29.90	97.38
	Sample 2		367635	122	21.74	26.23	96.98
	Sample 3		464840	116	18.73	29.47	97.24
	Sample 4		593011	121	18.43	32.94	96.83
	Sample 5		401155	106	17.48	26.81	97.71
	Sample 6		583542	121	16.69	29.65	96.80
	Sample 7		510595	124	15.23	26.32	97.27
	Sample 8		477518	105	16.24	26.82	97.58
Run 3	Sample 9	No	588150	170	123.4	46.94	98.74
	Sample 10		542746	180	111.8	44.68	97.65
	Sample 11		788342	164	166.4	48.61	97.81
	Sample 12		876960	152	172.3	48.21	98.18
Run 4	Sample 13	No	742529	126	58.59	21.53	97.74
	Sample 14		533543	142	43.24	20.78	98.67
	Sample 15	2 per capture	634152	178	59.83	20.42	98.54
	Sample 16		603691	143	51.25	21.53	98.46
	Sample 17	4 per capture	695757	169	78.96	25.51	98.59
	Sample 18		435126	141	46.49	27.54	97.17
	Sample 19		660116	145	82.22	31.33	98.62
	Sample 20		155313	184	18.18	24.81	97.88

Table 2. Description of the experiment and mapping statistics

Conclusions

The existing SureSelect Target Enrichment Ion Proton protocol has been successfully adapted to the PGM. In particular, the library size was optimized to benefit from the increased read length of the PGM sequencer. Modifications made to the shearing step and the first size selection step significantly increased average read length. Uniformity was similar for the modified protocol versus the standard one. Pre-capture pooling of the samples did not significantly impact the quality of the results, while increasing the throughput and ease of use of the workflow, as the total base pair output was increased 34% with these modifications. This modified protocol extends the benefits of SureSelect to yet another NGS sequencer, providing solutions for library preparation, target enrichment, and sample QC.

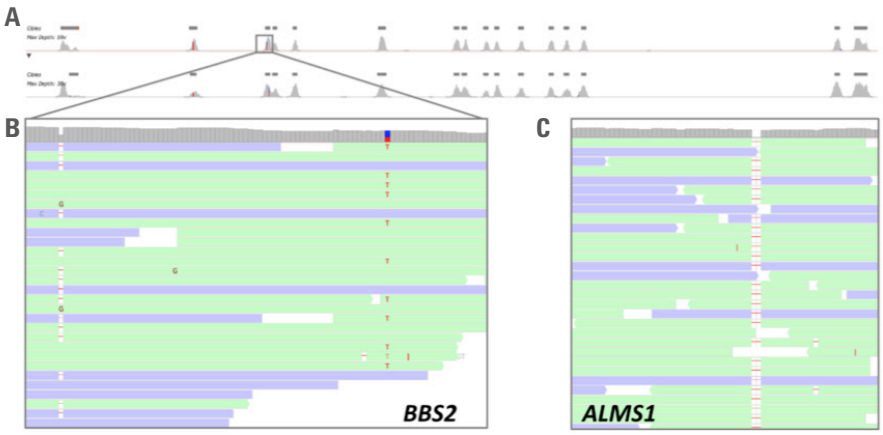

Figure 2. A. Sequence coverage for the gene *BBS2* in 2 samples. B. Example of 2 heterozygous pathogenic mutations in *BBS2* that are in trans (overlapping reads only harbor one of the 2 mutations). C. Homozygous deletions of 2 bases in the *ALMS1* gene. The alignments were visualized using Alamut Visual software (Interactive Biosoftware, Rouen, France).

#Sample	#SNV* Run1	#SNV Run2	Total Non-redundant SNV from both runs	#Common	#SNV from Run1	#SNV from Run2	%Common
Sample 1	229	218	243	198	15	30	81.5
Sample 2	274	257	281	245	7	29	87.2
Sample 3	266	262	275	245	10	20	89.1
Sample 4	206	195	213	186	5	22	87.3

*SNV= Total number of variants per sample

Table 3. Comparison of the samples analyzed in Run1 and Run2

Trusted Answers. Together.

LEARN MORE AT:
www.agilent.com/genomics/NGS

Find an Agilent customer center in your country:
www.agilent.com/genomics/contactus

U.S. and Canada
1-800-227-9770
agilent_inquiries@agilent.com

Europe
info_agilent@agilent.com

Asia Pacific
info_agilent@agilent.com

For Research Use Only. Not for use in diagnostic procedures.

PR7000-0297

Published in the USA, November 1, 2016
5991-7571EN

Agilent Technologies