Next Generation Sequencing with the SOLiD® System and Ion Torrent's PGM™ Sequencer: **Amplicon Resequencing of the *CFTR* gene using multiplexing**

Roland Achmann[1], Robert Greither[2], Rick Dixon[2], Alain Rico[2], Alexander Sartori[2], Franziska Stellmer[1], Susanne Zahn[1], Loukas Argyriou[1], Bernd Auber[1]*

[1]MVZ genteQ GmbH, Lornsenstr. 4, 22767 Hamburg, Germany; [2]Life Technologies GmbH, Frankfurter Str. 129B, 64293 Darmstadt, Germany

***corresponding author**

INTRODUCTION

Cystic Fibrosis (CF) is a common autosomal recessive disorder caused by mutations in the Cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. More than 1800 *CFTR* mutations are known and there is great mutational heterogeneity depending on ethnical and phenotypic background.

Commercial screening kits consider only a subset of relatively frequent mutations covering around 80% of mutations causing classic CF but only approximately 50 to 60% of mutant alleles in the *CFTR* genes of individuals with atypical forms of CF like congenital aplasia of the vas deferens (CBAVD).

Screening of the entire coding sequence of the gene would increase the sensitivity of mutation detection, but Sanger sequencing is still expensive. As an alternative, Next-Generation Sequencing is very cost effective, especially when samples are multiplexed.

We describe a pilot study comprising 20 samples (17 with known and 3 control samples from healthy individuals with unknown *CFTR* gene mutational status), using a Multiplicom multiplexing assay for the *CFTR* gene, sample barcoding and subsequent Next-Generation Sequencing (NGS) with the SOLiD™ 4 System. The result of NGS is compared with Sanger sequencing data. Further, we developed a protocol for Ion Torrent's semiconductor-based Personal Genome Machine (PGM™) System that enabled the sequencing of established (larger) amplicons despite the current system read-length of 100 bp and demonstrate its utility by sequencing of a selected sample. Finally, 20 samples were sequenced on the new 5500xl Genetic Analyzer.

EXPERIMENTAL DESIGN

The aim of this study was to develop NGS workflows for fast and cost efficient specific resequencing of the coding and flanking intron regions of the *CFTR* gene.

All samples selected for this study (see Material and Methods) were sequenced on SOLiD® 4 system and by capillary electrophoresis using a 3500xL Dx Genetic Analyzer. On SOLiD® 4, a set of 20 samples was barcoded and processed as a pool on one fourth of a slide (quadrant).

For comparison, 5 of these samples were sequenced as singletons on one slide octet each. In addition, we sequenced the same samples as a 20-plex with the new 5500xl Genetic Analyzer (analysis ongoing). In order to perform sequencing of the same *CFTR* amplicons on Ion Torrent PGM™ sequencer we developed a library protocol based on concatenation of the Multiplicom amplicons and shearing of fragments to a length of around 100 bp to meet the current system specifications. We selected one sample for the demonstration of this protocol. Sequencing of more samples will be conducted using sample barcoding.

Sample ID	Source	Allele 1	Allele 2	SOLiD 4 20-plex	SOLiD 4 single	5500 SOLiD 20-plex	PGM	Sanger
CF01	Coriell 07441	c.489+1G>T	c.2988+1G>A		–		–	
CF02	Coriell 07469	c.1521_1523delCTT	c.1657C>T		–		–	
CF03	Coriell 07552	c.1521_1523delCTT	c.1657C>T		–		–	
CF04	Coriell 11275	c.1521_1523delCTT	c.3528delC					
CF05	Coriell 11280	c.489+1G>T	c.579+1G>T					
CF06	Coriell 11282	c.254G>A	c.489+1G>T		–		–	
CF07	Coriell 11284	c.1521_1523delCTT	c.1679G>C		–		–	
CF08	Coriell 11290	c.489+1G>T	c.1364C>A		–		–	
CF09	Coriell 11472	c.3909C>G	c.4046G>A		–		–	
CF10	Coriell 11859	c.2657+5G>A	c.2657+5G>A		–		–	
CF11	Coriell 12785	c.1040G>C	c.1652G>A		–		–	
CF12	Coriell 12960	c.1000C>T	c.3368-2A>T *		–		–	
CF13	Coriell 13423	c.254G>A	c.3454G>C					
CF14	Coriell 13591	c.350G>A	c.1521_1523delCTT		–		–	
CF15	Coriell 13668	c.1521_1523delCTT	c.54-5940_273+10250del21kb					
CF16	Coriell 13800	c.1521_1523delCTT	c.1766+1G>A		–		–	
CF17	Coriell 13803	c.1521_1523delCTT	c.2051_2052delAAinsG					
CF22	control sample	normal *	normal *		–		–	
CF23	control sample	normal *	normal *		–		–	
CF24	control sample	normal *	normal *		–		–	
CF20	QAS ** sample	c.1521_1523delCTT	normal	–	–	–		

Table 1: Experimental study design and mutational status of Coriell and control samples: sample sequenced and fully analyzed; sample sequenced, analysis in progress; sample sequencing pending; SOLiD® 4 system data quality unsatisfactory (see 'Results' for details); sample with results from ALL platforms after finalisation of study available; * Mutational status previously unknown. Mutations were detected by NGS and Sanger sequencing. ** Sample from Quality Assessment Scheme.

MATERIALS AND METHODS

DNA Samples with known mutations (Table 1) were purchased from Coriell Cell Repositories (CF01 to CF17). Three samples (CF22, CF23, CF24) were obtained from healthy control individuals with unknown mutational status. For PGM sequencing, a sample heterozygous for a F508del mutation (c.1521_1523delCTT) was used. An overview of the experimental steps performed for Sanger Sequencing and NGS is shown in Figure 1.

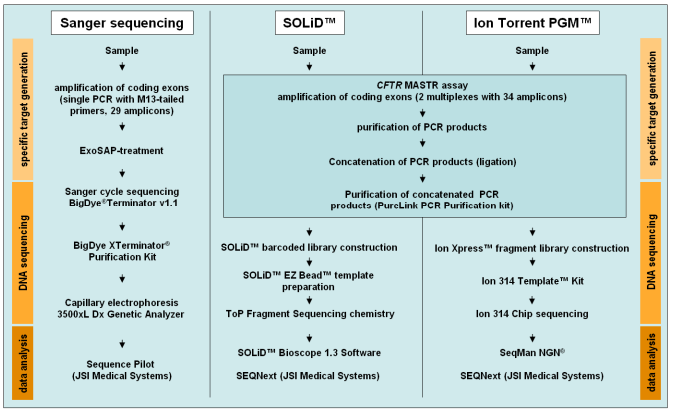

Fig. 1: Experimental steps for Sanger and Next-Generation Sequencing

Sanger Sequencing

CFTR coding exons (27) were amplified with M13-tailed target specific primers. For the large exon 14, three overlapping amplicons were prepared. Data analysis of sequence runs was restricted to coding exons ± 20 bp of flanking sequences. Reference sequence NM_000492.3.

Next-Generation Sequencing

Preceeding NGS amplicons were generated with the *CFTR* MASTR assay (Multiplicom). The assay enables amplification of 34 amplicons in 2 multiplex PCR reactions representing all coding sequences of the *CFTR* gene.

SOLiD® 4, 5500 Genetic Analyzer, Ion Torrent PGM™ sequencer: Reads were mapped to the hg18 reference.

RESULTS

SOLiD® 4: SOLiD® system sequencing of multiplexed barcoded samples with known and unknown mutational status (Table 1) revealed evaluable results for 17 out of 20 samples. Analysis of samples CF02, CF08 and CF12 was constrained by various differences in comparison to the expected occurrence of mutations as well as to Sanger sequencing data. We reviewed every experimental step and found that PCR product yield was insufficient for constructing high quality libraries. This resulted in low sequence coverage for these three samples.

For the remaining Coriell samples, all expected mutations were identified. The deletion c.54-5940_273+10250del21kb could not be directly detected because the genomic breakpoint of this mutation is not covered by the *CFTR* MASTR assay.

So far SOLiD® system results have been confirmed with Sanger sequencing for 12 out of 20 samples (see Table 1 and Figure 2). Comparison of single and multiplex sample analysis revealed identical results for all samples tested (CF04, CF05, CF13, CF15 and CF17, respectively). No mutations were found in control samples of three healthy individuals. Each of the multiplexed samples which were sequenced on a slide quadrant showed a minimal coverage of 10,000-fold.

Ion Torrent PGM: PGM™ sytstem sequencing of a single F508del heterozygous sample showed that amplicons generated with the *CFTR* MASTR assay can be used as template for PGM™ system sequencing by amplicon concatenation and subsequent shearing to the desired fragment length. Complete target coverage allowed to detect all variants as compared to Sanger sequencing (Figure 3).

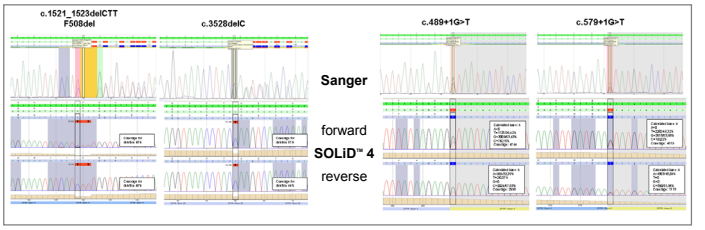

Fig. 2: Examples for detected *CFTR* mutations for samples CF04 and CF05 (mulitplex run). For SOLiD® system results are presented as pseudo-electropherograms (SeqNext, JSI Medical Systems). Note that coverage for deletion mutations is decreased to approximately 50%.

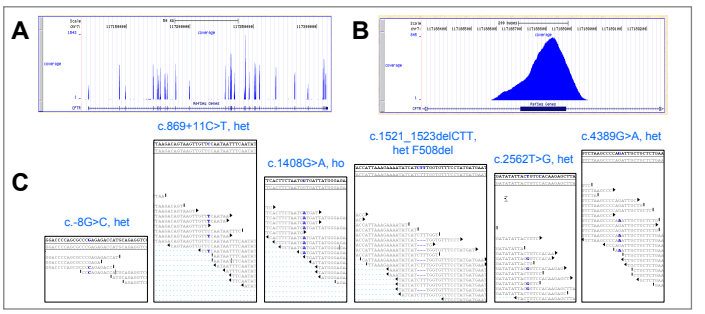

Fig. 3: Sample CF20. A) Coverage plot of PGM sequenced *CFTR* regions. B) Typical coverage of a single exon. C) Detected variants and mutations; het: heterozygous; ho: homozygous.

CONCLUSIONS and OUTLOOK

The pilot study shows that amplicon resequencing of the *CFTR* gene using multiplex amplification, pooling of barcoded samples and NGS can in principal be used for high-throughput mutation screening.

The excessive coverage per sample achieved in this study reveals a possible multiplexing of up to 96 samples on one octet of a slide on SOLiD® 4 or one lane on the 5500 Genetic Analyzer with still more than sufficient coverage. Moreover, on the 5500 Genetic Analyzer, sequencing can be conducted in less than half the time and more cost efficiently due to the flexible single lane operation. On PGM™ sequencer, the entire workflow can be processed in less than two working days.

The NGS workflow is well developed and all mutations were detected using the SOLiD® Bioscope™ 1.3 as well as the JSI SeqNext software modules. Further improvements of software analysis pipelines would help to allow for automated *CFTR* mutation detection, particularly in repetitive and homopolymer sequence regions (e.g. polyT-region of the *CFTR* gene).

ACKNOWLEDGEMENTS

We thank Silke Deckert, Christina Schmidt, Maren Schnelle and Shima Soltani for technical assistance. We are especially grateful to JSI Medical Systems for their support in data analysis. We also thank Dirk Goossens and Geoffrey Henno from Multiplicom for help with the *CFTR* MASTR assay.