Highly sensitive, non-invasive detection of colorectal cancer mutations using single molecule, third generation sequencing

Giancarlo Russo[1], Andrea Patrignani[1], Lucy Poveda[1], Frederic Hoehn[2], Bettina Scholtka[3], Ralph Schlapbach[1], Alex M. Garvin[2]

1. Functional Genomics Center Zurich, ETH-University of Zurich, Switzerland; 2. Droplet Diagnostics SAS, France; 3. Institute of Nutritional Science, University of Potsdam, Germany

Background :

- Colorectal cancer (CRC) is a major cause of cancer mortality.
- Unlike methylation, which occurs to some extent in normal tissue, mutations in the oncogenes and tumor suppressor genes that drive CRC are highly specific for tumor tissue.
- Many genes can be mutated in CRC, although the APC, KRAS, and P53 genes are mutated at far higher frequencies than other genes.
- Mutations in patient samples are unknown and can be found at many locations in these genes, therefore a scanning method for mutation detection, such as sequencing, is needed for high sensitivity.

Objectives :

- To determine the sensitivity and specificity of single molecule, third-generation sequencing for an assay that detects rare mutations in the genes that drive CRC

Methods :

- PacBio RS is a single molecule, 3rd generation sequencing technology; it generates highly accurate reads on short amplified fragments using circular consensus sequencing (CCS) to correct random errors **(Fig. 1).**

Experiment 1 :

- Assay consisting of 15 amplicons covering regions of 5 genes mutated in CRC (total test sequence is about 5,000 bp).
- Simulated stool sample obtained by mixing 97% wildtype DNA with 3% DLD1 cell line DNA (mutated at three known loci); heterozygous mutants represented at the 1.5% level.
- 15 amplicons covering critical regions of 5 genes known to be mutated in CRC generated using the wildtype and the mixed DLD1/wildtype DNA sample were sequenced using the PacBio RS, Illumina MiSeq and IonTorrent PGM platforms.

Experiment 2 :

- Genomic DNA extracted from tumor tissue and matched, adjacent normal tissue (colon) in a 37 y/o patient diagnosed with adenocarcinoma; the mutations and their locations were unknown.
- Samples were sequenced only on PacBio RS (1 SMRTcell per sample).

Experiment 3 :

- Genomic DNA was isolated from the stool of two patients taken after they were diagnosed with CRC by colonoscopy and prior to surgical removal of the tumor
- After the excision of the cancerous neoplasm, DNA was extracted from the tissue itself to confirm any mutations found in the DNA isolated from the stool samples
- Assay consisted of 4 regions on the APC gene

Results :

Experiment 1:

- PacBio RS **(Fig. 2)** shows 100 % sensitivity and specificity and average Phred score above 60.
- Illumina MiSeq and IonTorrent PGM **(Fig. 2)**, due to their systematic biases and lower base quality, miss one mutation and report many false positives.

Experiment 2:

- Two mutations were identified in the sample and confirmed by Sanger sequencing
- A germline, 5 base deletion at codon 1309 of gene APC (P < 1e-16); this is the most frequent mutation in APC (http://www.umd.be/APC/) **(Fig. 3)**
- A somatic G>A substitution at codon 237 of P53 present at ca. 13% in the tumor (P < 1e-16); this missense mutation (Arg > His) is also a known hotspot for CRC. (http://p53.free.fr/Database/p53_mutation_HB.html) **(Fig. 3)**

Experiment 3:

- In both patients, one known mutation was identified in the stool sample and confirmed in the tumor tissue. In one sample, an additional mutation is found in stool **(Tables 1,2)**

Conclusions :

- We performed three experiments employing third generation, single molecule sequencing to detect low frequency CRC mutations in a simulated stool sample, in tumor tissue and in real stool sample
- Simulated stool sample: with no systematic bias and a much higher raw base-calling CCS quality, PacBio RS shows 100 % sensitivity and specificity; Illumina MiSeq and IonTorrent PGM are instead characterized by their background noise in base calling and are not adequate to detect mutant DNA at the 1.5% level, which is often encountered in stool samples from CRC patients.
- Tumor tissue: two mutations were identified well beyond and both are known hotspots associated with CRC. In particular, the 5-bases, AAAGA deletion on gene APC at codon 1309 is found in both normal and tumor tissue; such a germline mutation was to be expected given the young age (37) of the patient.
- Real stool sample: we identified mutations corresponding to hotspots in the stools and confirmed them in the tumor tissue.

Fig. 1: The same template is read many times over and errors are corrected.

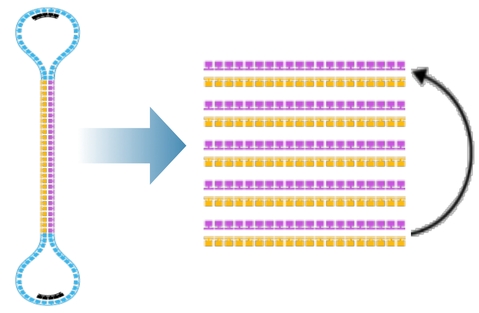

Fig. 2: With extremely high CCS-reads quality, the results obtained using PacBio are 100% specific and sensitive. On the contrary, both IonTorrent PGM and Illumina MiSeq report several false positive and miss to identify one variant

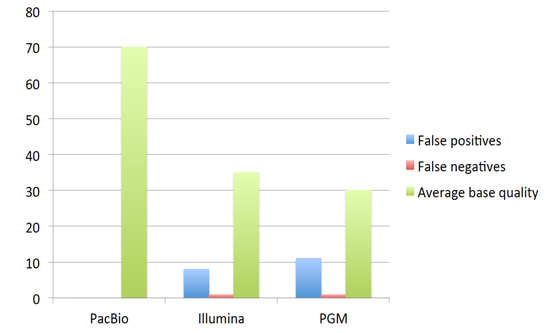

Fig. 3

LEFT COLUMN: This 5bp deletion represents the most frequently mutated codon in the APC gene. The DNA samples extracted form both the tumor and the healthy, adjacent tissues were sequenced on one SMRT cell. By means of third generation sequencing, the deletion was identified with frequencies of 48.22 % and 68.61 % in the normal and the tumor (A) tissues, respectively (P < 1e-256 in both cases, Fisher's exact test and confirmed by Sanger sequencing (B,C)

RIGHT COLUMN: This mutation represents the third most frequently mutated codon in the *TP53* gene. The mutation in the tumor sample is a heterozygous substitution which occurs at a frequency of 12.96 % (D). Despite this rather low cellularity, the total absence of background noise allows for an easy detection of the green A signal in the Sanger sequencing track (F), which is instead clearly absent in the healthy tissue (E).

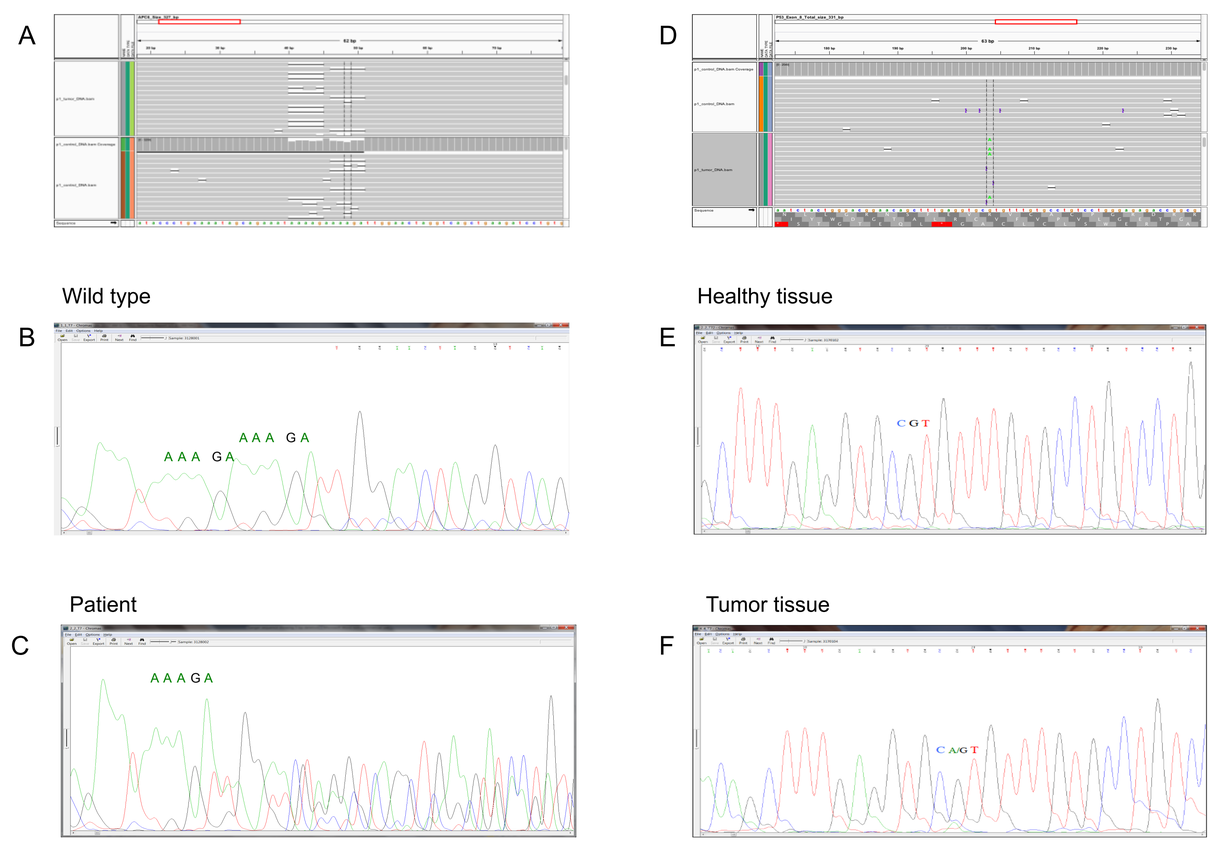

Table 1 – Mutation found in the DNA stool sample extracted from patient 1. The frequencies of mutated DNA in the stool samples is low; however, the absence of sequencing errors in the wild type allowed these observations to be confidently reported as significantly different from the wild type even after a stringent Bonferroni correction.

	Tumor	Stool
Gene	APC6	APC6
Codon	1309	1309
Reference allele	T	T
Alternative allele (AA)	-AAAAG	-AAAAG
Coverage in the wild type (WT)	4658	4658
AA frequency in the WT (%)	0.00	0.00
Coverage in the patient sample	3897	4550
AA frequency in the patient sample (%)	82.22	0.57
P-value	0.00E+000	1.22E-008
Adjusted P-value	0.00E+000	3.66E-006

Table 2 - The additional mutation found in patient 2 is the eighth most frequently mutated codon in CRC and it is extremely significant in our stool sample. Therefore, despite the fact that this polymorphism was not observed in the DNA from the tumor tissue, it is unlikely that represents a false positive, as there are no other positives reported, even at lower frequencies. We believe that this mutation might come from a secondary neoplastic formation which perhaps was not fully developed yet at the time of the colonoscopy and therefore was not resected.

	Tumor	Stool	
Gene	APC	APC	APC
Codon	1491	1491	1556
Reference allele	A	A	G
Alternative allele (AA)	-T	-T	+A
Coverage in the wild type (WT)	7372	7372	7359
AA frequency in the WT (%)	0.00	0.00	0.04
Coverage in the patient sample	8005	8005	7984
AA frequency in the patient sample (%)	19.05	0.37	1.32
P-value	0.00E+000	3.04E-009	2.72E-026
Adjusted P-value	0.00E+000	3.43E-007	8.16E-024

Contact: giancarlo.russo@fgcz.ethz.ch

ETH Eidgenössische Technische Hochschule Zürich Swiss Federal Institute of Technology Zurich

Universität Zürich[UZH]

f·g·c·z

Droplet Diagnostics

Universität Potsdam