SeqNext: Evaluation of data analysis software for next generation sequencing in a clinical laboratory

430

Jakub Patrick Sram, Harry Gao, Dongqing Gu, Wengang Chen, Juan-Sebastian Saldivar, Lawrence Weiss

City of Hope, Duarte, CA 91010

Abstract

Next generation sequencing is revolutionizing clinical molecular diagnostic testing. Many CLIA laboratories are adapting this innovative sequencing technology to develop new clinical tests to answer unmet needs, such as sequencing large gene panels in a cost efficient manner. As next generation sequencing replaces capillary sequencing in the clinical space, the cost per base becomes cheaper and cheaper, but the data analysis bottleneck still remains. Present medical geneticists and technologists need user-friendly software that will help them manage data, analyze for and find unknown mutations and report results produced by the next-gen sequencers. Unfortunately, the number of software products on the current market that can perform such a task is extremely limited.

Our objective was to thoroughly evaluate, analyze and validate the new software, SeqNext, developed by JSI Medical Systems. We used the Illumina GA IIx to sequence 12 p53 patient samples with 83 known alterations that were previously sequenced by an ABI 3730 capillary sequencer. The 83 mutations in the p53 patient samples included: 48 bp, 16 bp, 2bp and 1bp deletions; a 7bp insertion; indels (ins 5bp del 1 bp) and single base substitutions. SeqNext correctly detected 100% of the 83 mutations in the p53 samples with one optimized analysis setting.

Complex mutations such as two overlapping deletions or insertions are exceptionally difficult to detect by capillary sequencing. SeqNext detected all sequence variants in a mixture of three p53 samples, including two overlapping insertions (7bp and 4bp). In addition, we evaluated competitive software commercially available on the market but we were not able to detect the more complex mutations. We found SeqNext to be a user-friendly software with the potential to significantly reduce data analysis time without sacrificing accuracy in detecting unknown mutations.

Work flow

- Long PCR from exon 2-10 of *p53*
- DNA shearing
- Adding adapter and barcodes
- Mix samples at about equal ratio
- GAIIx sequencing
- Sort data by barcode
- Mutation analysis with SeqNext software
- Compare with capillary sequence data

Limitations For Capillary Sequencing

- High background
- Difficult to analyze heterozygous deletions, insertions and indels
- Much more difficult to analyze complex mutations
  - two overlapping deletions or insertions

del.T/ins.CCCAA (capillary)

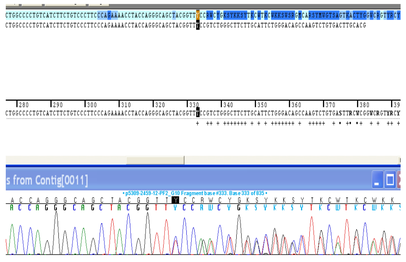


Fig 1. Example of a complex hetorozygous mutation sequenced by CE where a T base is deleted and CCCAA bases are inserted

Ins7bp & Ins4bp

Fig. 2. A mixture of three p53 samples was sequenced by GAIIX and SeqNext detected correctly two heterozygous insertions of 7 bp and 4 bp, CGGTTTC and CCCA, respectively.

48bp & 16bp overlapping deletions

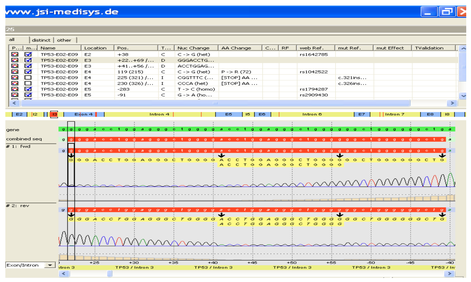


Fig. 3. In a mixture of three p53 samples, SeqNext was able to detect two large hetorozygous deletions of 48bp and 16bp at the same sequence position.

raw data view: 48bp deletion

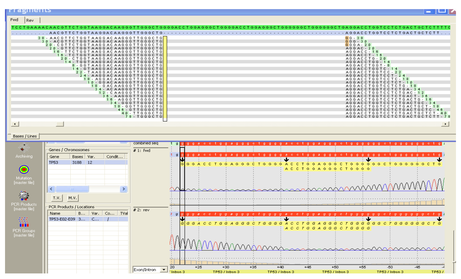


Fig. 4. 48bp deletion visualized by aligning reads to the p53 reference sequence.

raw data view: 16bp deletion

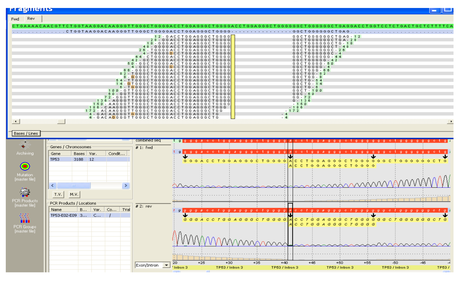


Fig. 5. 16bp deletion visualized by aligning reads to the p53 reference sequence.

T>G point mutation

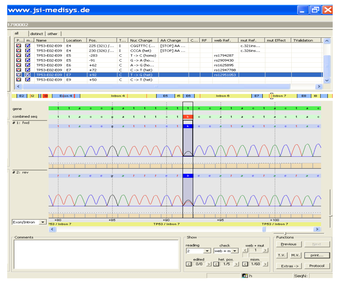


Fig. 6. Heterozygous T>G mutation visualized by SeqNext

T>G point mutation raw data

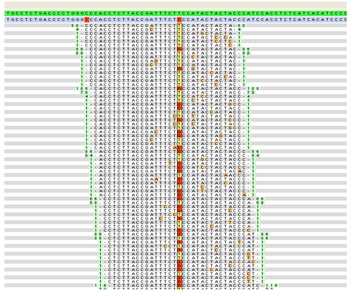
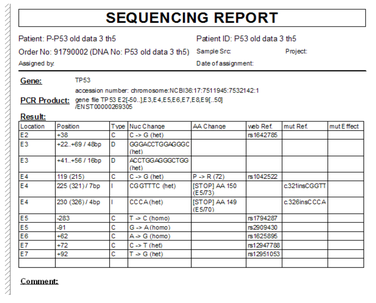


Fig. 7 . Heterozygous T>G mutation visualized by SeqNext by looking at the actual reads

Sample Report



SEQUENCING REPORT

Patient: P-P53 old data 3 th5 — Patient ID: P53 old data 3 th5

Order No: 91790002 (DNA No: P53 old data 3 th5) — Sample Src: — Project:

Assigned by: — Date of assignment:

Gene: TP53

accession number: chromosome:NCBI36:17:7511945:7532142:1

PCR Product: gene file TP53 E2[+50..]E3,E4,E5,E6,E7,E8,E9[..50] /ENST00000269305

Result:

Location	Position	Type	Nuc Change	AA Change	web Ref	mut Ref	mut Effect
E2	+38	C	C -> G (het)		rs1642785		
E3	+22..+69 / 48bp	D	GGGACCTGCAGGGC (het)				
E3	+41..+56 / 16bp	D	ACCTGCAGGGCTGG (het)				
E4	119 (215)	C	C -> G (het)	P -> R (72)	rs1042522		
E4	225 (321) / 7bp	I	CGGTTTC (het)	(STOP) AA 150 (E5/3)			c.321insCGGTT
E4	230 (326) / 4bp	I	CCCA (het)	(STOP) AA 149 (E5/3)			c.326insCCCA
E5	-283	C	T -> C (homo)		rs1794287		
E6	-91	C	G -> A (homo)		rs2909430		
E6	+62	C	A -> G (homo)		rs1625895		
E7	+72	C	C -> T (het)		rs12647786		
E7	+92	C	T -> G (het)		rs12951053		

Comment:

Fig.8. SeqNext can print a clinically relevant report

Low level mosaic mutation

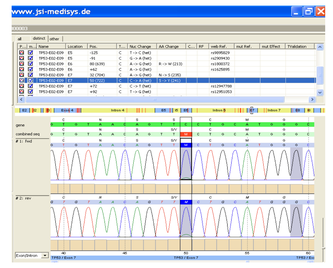


Fig.9. Low level mosaic mutations can be detected by GAIIx sequencing and SeqNext software.

Low level mosaic mutation raw data

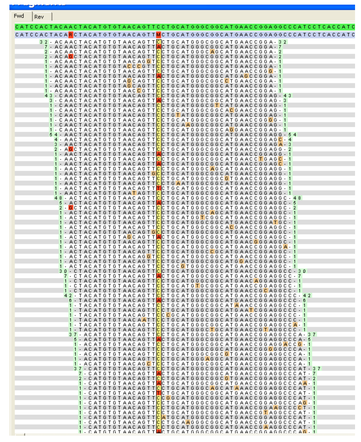


Fig.10. Actual reads visualized by SeqNext for the mosaic C>A change

Conclusions

- SeqNext correctly detected all complex mutations under one setting.
- We found SeqNext to be a user-friendly software
- Using GAIIx sequencing and SeqNext, low level mosaic mutations that otherwise could not be detected by Sanger sequencing can now be detected