Cost Effective Bioinformatics Pipeline for NGS Reporting of Clinically Relevant Cancer Genes Using the Ampliseq Cancer Hotspot Panel v2

Rebecca Stolarczyk, Keith Tomaszewicz, Xiuling Meng, Bruce A Woda, Lloyd M. Hutchinson, Ediz F. Cosar
Department of Pathology, University of Massachusetts Medical School, Worcester, MA, USA

BACKGROUND

Laboratories with limited resources may lack an adequate bioinformatics pipeline to analyze and report NGS results. Our study aimed to develop an affordable solution for reviewing clinical NGS data.

MATERIALS AND METHODS

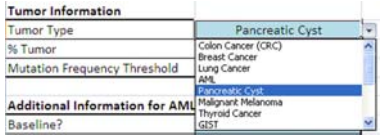
Variant caller (VCF) files were generated by the Variant Caller (VC) plug-in provided with the Ion PGM sequencer (high stringency) and 3rd party software, NextGene (NG) by Softgenetics (low stringency). VCF files, patient demographics, and NGS QC metrics were transferred to software developed in Excel Visual Basic. Analysis consists of two automated macros, prompted by user input. The first script checks rolling databases to annotate and sort raw data based on various filters, including mutation frequency (<1%), coverage (<500x), known SNPs, and sequencing artifacts. VCF files from VC and NG are compared and common variants are identified. A manual review of nucleotide sequences is performed for a predefined subset of genes based on tumor type. Technologists evaluate the annotated data and verify the mutations to report. The second script then compiles a comprehensive report based on mutational findings and patient demographics. To ensure technical accuracy, 131 known positive cases were analyzed using the bioinformatics pipeline and clinically relevant findings were validated via Sanger Sequencing. Cases consisted of pancreatic neoplasm (34%), lung carcinoma (15%), tumor cell line (15%), desmoids tumors (10%), myeloid neoplasm (9%), melanoma (7%) colon carcinoma (4%) and GIST (3%).

NGS Bioinformatics Workflow

Variant Caller Raw Data
VCF is obtained from Variant Caller (high stringency)

Nextgene Raw Data
VCF is obtained from NextGene (low stringency)

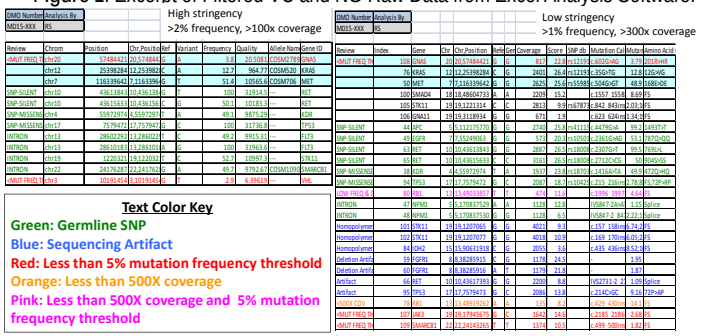
Excel NGS Analysis Software
Raw data is imported into Excel NGS Analysis Software.



NGS Bioinformatics Workflow (Continued)

Pre-Review
Software checks rolling databases to annotate and sort raw data

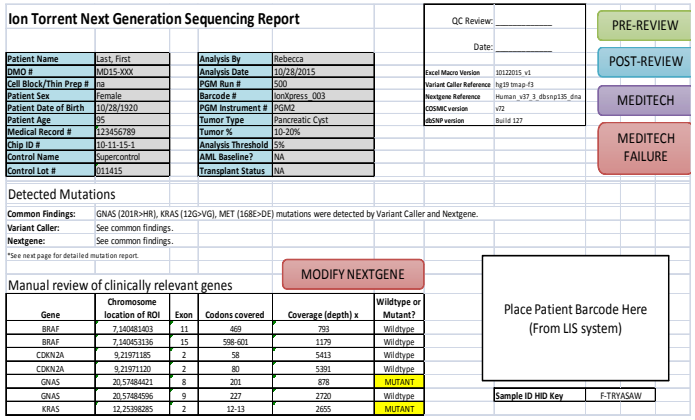
Figure 1: Excerpt of Filtered VC and NG Raw Data from Excel Analysis Software.



Mutation Review
Technologist review of sequence traces to confirm bio-informatic calls (Cells in blue above highlight variants common to NG and VC)

Post-Review
Software automatically generates a comprehensive final report from mutational findings and patient demographics.

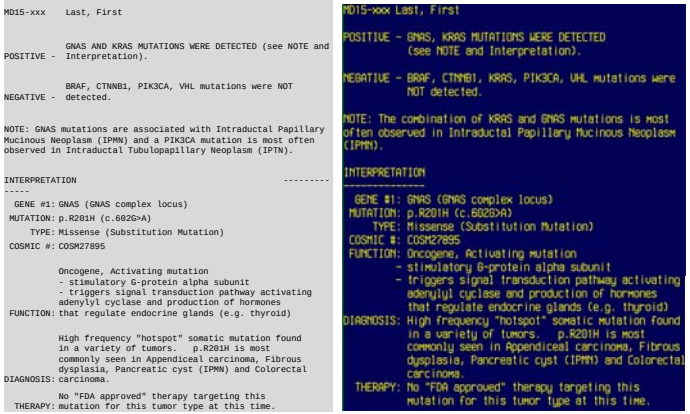
Figure 2: Excerpt of Final Report Generated by Excel Analysis Software



NGS Bioinformatics Workflow (Continued)

Laboratory Information System (LIS) Report
Software automatically compiles formatted LIS Reports

Figure 3: Report From Software Can Be Copied Directly Into LIS Program



RESULTS

- All clinically relevant mutations (100%) were detected by the NGS bioinformatics pipeline and by Sanger Sequencing.
- On average, VC made 22 calls per case, whereas NG made 86 calls.
- Of the total variants per case, an average of 24% were identified as common SNPs, 8% as sequencing artifacts, 42% as low frequency artifacts (<1% allele frequency), and 2% as variants with low coverage (<500x), ultimately filtering approximately 77% of the raw data.
- An average of 3 variants per case was common to both VC and NG.
- On average 2 somatic mutations per case (67%) and 1 rare hereditary germline variant (33%) were identified, accounting for 9% of total VC findings and 2% of total NG findings.

DISCUSSION/CONCLUSION

- This pipeline is an affordable and accessible alternative to commercially available products or bioinformatics team.
- The combination of two software packages into one pipeline reduces the likelihood of error and ensures accurate reporting.
- Automation of processes such as data sorting and report compilation reduces resources needed to analyze clinical samples.