

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.

Tumor Information	Pancreatic Cyst
Tumor Type	Pancreatic Cyst
% Tumor	Colon Cancer (CRC)
Mutation Frequency Threshold	Breast Cancer
Additional Information for AML	Lung Cancer
Baseline?	AML
	Pancreatic Cyst
	Malignant Melanoma
	Thyroid Cancer
	GIST

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.

NGS Number	Analysis By	High stringency	Low stringency
REFSEQ	AC	>2% frequency, >100x coverage	>1% frequency, >300x coverage
Gene	Chrom	Position	Variant
BRCA1	chr2	25188244	C>T
BRCA2	chr13	34543324	G>A
CDKN2A	chr12	251971185	T>C
GNAS	chr20	57484421	G>A
KRAS	chr12	25368282	G>T

Text Color Key
Green: Germline SNP
Blue: Sequencing Artifact
Red: Less than 5% mutation frequency threshold
Orange: Less than 500X coverage
Pink: Less than 500X coverage and 5% mutation frequency threshold

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

Ion Torrent Next Generation Sequencing Report

Patient Name	Last, First	Analysis By	Rebecca
DMO #	0015-XXX	Analysis Date	10/28/2015
Cell Block/Thin Prep #	34	PGM Run #	500
Patient Sex	Female	Barcode #	100995X_003
Patient Date of Birth	10/28/1920	PGM Instrument	PGM2
Patient Age	95	Tumor Type	Pancreatic Cyst
Medical Record #	123456789	Tumor %	10-20%
Chip ID #	10-11-15-1	Analysis Threshold	5%
Control Name	Supercontrol	AML Baseline?	NA
Control Lot #	011415	Transplant Status	NA

QC Review: _____ Date: _____

Pre-Review: POST-REVIEW: MEDITECH: MEDITECH FAILURE:

Detected Mutations

Common Findings: GNAS (201R>HR), KRAS (12G>VG), MET (168E>DE) mutations were detected by Variant Caller and NextGene.

Variant Caller: See common findings.

NextGene: See common findings.

*See next page for detailed mutation report.

Manual review of clinically relevant genes

Gene	Chromosome	Location of ROI	Exon	Codons covered	Coverage (depth) x	Wildtype or Mutant?
BRCA1	17	43045313-4	11	588-601	791	Wildtype
BRCA2	13	34543324-3	15	598-612	1179	Wildtype
CDKN2A	12	251971185	2	58	5413	Wildtype
CDKN2A	12	251971130	2	80	391	Wildtype
GNAS	20	57484421	8	201	878	MUTANT
GNAS	20	57484596	9	227	2720	Wildtype
KRAS	12	25368282	2	12-13	2659	MUTANT

Place Patient Barcode Here (From LIS system)

Sample ID HID Key: F-TR1ASAW

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

MD15-xxx Last, First

GNAS and KRAS MUTATIONS WERE DETECTED (see NOTE and POSITIVE - Interpretation).

NEGATIVE - BRAF, CTNNB1, KRAS, PIK3CA, UBE1 mutations were NOT detected.

NEGATIVE - detected.

NOTE: GNAS mutations are associated with Intraductal Papillary Mucinous Neoplasm (IPMN) and a PIK3CA mutation is most often observed in Intraductal Tubulopapillary Neoplasm (ITPN).

INTERPRETATION

GENE #1: GNAS (GNAS complex locus)

MUTATION: p.R201H (c.602D>A)

TYPE: Missense (Substitution Mutation)

TYPE: Missense (Substitution Mutation)

CSMID #: COSM27895

Oncogene, Activating mutation

- stimulatory G-protein alpha subunit

- triggers signal transduction pathway activating adenyl cyclase and production of hormones

FUNCTION: that regulate endocrine glands (e.g. thyroid)

High frequency 'hotspot' somatic mutation found in a variety of tumors. p.R201H is most commonly seen in Appendiceal carcinoma, Fibrous dysplasia, Pancreatic cyst (IPMN) and Colorectal carcinoma.

No 'FDA approved' therapy targeting this THERAPY: mutation for this tumor type at this time.

POSITIVE - GNAS, KRAS MUTATIONS WERE DETECTED (see NOTE and Interpretation).

NEGATIVE - BRAF, CTNNB1, KRAS, PIK3CA, UBE1 mutations were NOT detected.

NOTE: The combination of KRAS and GNAS mutations is most often observed in Intraductal Papillary Mucinous Neoplasm (IPMN).

INTERPRETATION

GENE #1: GNAS (GNAS complex locus)

MUTATION: p.R201H (c.602D>A)

TYPE: Missense (Substitution Mutation)

CSMID #: COSM27895

FUNCTION: Oncogene, Activating mutation

- stimulatory G-protein alpha subunit

- triggers signal transduction pathway activating adenyl cyclase and production of hormones

DIAGNOSIS: High frequency 'hotspot' somatic mutation found in a variety of tumors. p.R201H is most commonly seen in Appendiceal carcinoma, Fibrous dysplasia, Pancreatic cyst (IPMN) and Colorectal carcinoma.

THERAPY: No 'FDA approved' therapy targeting this mutation for this tumor type at this time.

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.