The Journal of Molecular Diagnostics, Vol. 17, No. 4, July 2015

the Journal of Molecular Diagnostics
jmd.amjpathol.org

# Confirming Variants in Next-Generation Sequencing Panel Testing by Sanger Sequencing

Linnea M. Baudhuin,* Susan A. Lagerstedt,* Eric W. Klee,† Numrah Fadra,† Devin Oglesbee,* and Matthew J. Ferber*

*From the Department of Laboratory Medicine and Pathology* and the Division of Biomedical Statistics and Informatics,† Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota*

Accepted for publication March 24, 2015.

Address correspondence to Linnea M. Baudhuin, Ph.D., or Matthew J. Ferber, Ph.D., Department of Laboratory Medicine and Pathology, Mayo Clinic, 200 First St. SW, Rochester, MN 55905. E-mail: baudhuin.linnea@mayo.edu or ferber.matthew@mayo.edu.

Current clinical laboratory practice guidelines for next-generation sequencing (NGS) do not provide definitive guidance on confirming NGS variants. Sanger confirmation of NGS results can be inefficient, redundant, and expensive. We evaluated the accuracy of NGS-detected single-nucleotide variants (SNVs) and insertion/deletion variants (indels) and the necessity of NGS variant confirmation using four NGS target-capture gene panels covering 117 genes, 568 Kbp, and 77 patient DNA samples. Unique NGS-detected variants (1080 SNVs and 124 indels) underwent Sanger confirmation and/or were compared to data from the 1000 Genomes Project (1000G). Recurrent variants in unrelated samples resulted in 919 comparisons between NGS and Sanger, with 100% concordance. In a second comparison, 762 unique NGS results (736 SNVs, 26 indels) from seven 1000G samples were found to have 97.1% concordance with 1000G phase 1 data. Sanger sequencing and 1000G phase 3 data confirmed the accuracy of the NGS results for all 1000G phase 1 discrepancies. In all samples, the depth of coverage exceeded 100× in >99.7% of bases in the target regions. In conclusion, confirmatory analysis by Sanger sequencing of SNVs detected via capture-based NGS testing that meets appropriate quality thresholds is unnecessarily redundant. In contrast, Sanger sequencing for indels may be required for defining the correct genomic location, and Sanger may be used for quality-assurance purposes. *(J Mol Diagn 2015, 17: 456—461; http://dx.doi.org/10.1016/j.jmoldx.2015.03.004)*

Next-generation sequencing (NGS) is a powerful technology for interrogating the human genome and expediting clinical molecular testing for inherited conditions. Applications for examining the entire human genome, or more targeted approaches looking at panels of genes, including the majority of the exome, are possible by NGS.[1—3] Current clinical applications for the detection of inherited disorders often exploit a gene panel approach, with a focus on enhancing the clinical utility of laboratory testing.

NGS panel tests can evaluate multiple genes simultaneously, at a fraction of the cost of conventional Sanger sequencing, making NGS much more amenable to gene panel testing. Although Sanger sequencing is being replaced by NGS targeted panels, it has not disappeared but rather has taken on a different role in laboratory testing processes. Due to the variability in NGS capture technology, Sanger sequencing has been used for analyzing regions (usually exons) where NGS fails to achieve sufficient depth of coverage or to generate data of high enough quality. Sanger sequencing is also used for confirming NGS variants before they are clinically reported.

It is this second application of Sanger sequencing that is the focus of this study. Given the complexity of, and limited experience with, NGS in the clinical space, there is a certain level of discomfort in accepting NGS results without confirmation by another method often considered a gold standard of DNA sequencing. This unease is furthered by the paucity of published clinical regulatory guidelines addressing NGS quality metrics or the confirmation of clinically actionable NGS results. The relative lack of regulatory guidance is due in large part to the fact that NGS is a rapidly changing, highly complex technology with multiple platforms and applications, viewed through a kaleidoscope of

Supported by funds from the Department of Laboratory Medicine and Pathology, Mayo Clinic.

Disclosures: None declared.


http://dx.doi.org/10.1016/j.jmoldx.2015.03.004

**Table 1** Characteristics of NGS Multigene Panels and Variants Observed

| NGS panel | Size (Kbp) | Gene list | Patient DNA samples (*n*) | 1000G DNA samples (*n*)* | Unique NGS variants detected (*n*) | Unique NGS variants confirmed by Sanger (*n*) | NGS variant occurrences confirmed by Sanger (*n*) | Unique NGS variants confirmed by 1000G | Unique NGS variants discordant with 1000G |
|---|---|---|---|---|---|---|---|---|---|
| Hereditary colon cancer | 49 | *APC, AXIN2, BMPR1A, CDH1, CHEK2, MLH1, MLH3, MSH2, MSH6, MUTYH, PTEN, SMAD4, STK11, TP53* | 60 | | 190 | 66 | 223 | | |
| Hereditary arrhythmias | 78 | *AKAP9, ANK2, CACNA1C, CACNA2D1, CACNB2, CAV3, GPD1L, KCNE1, KCNE2, KCNE3, KCNH2, KCNJ2, KCNJ5, KCNJ8, KCNQ1, SCN1B, SCN3B, SCN4B, SCN5A, SNTA1* | 5 | 7 | 149 | 56 | 73 | 96 | 1 |
| Other cardiovascular-related genes | 116 | *ABCC6, ABCG5, ABCG8, ACTA2, ACVRL1, APOB, CBS, COL3A1, ENG, FBN1, FBN2, LDLR, LDLRAP1, LTBP2, MYH11, MYLK, PCSK9, SKI, SLC2A10, SMAD3, SMAD4, TGFB2, TGFBR1, TGFBR2* | 6 | 7 | 298 | 131 | 286 | 246 | 13 |
| Hereditary cardiomyopathies | 325 | *ABCC9, ACTC1, ACTN2, ANKRD1, BRAF, CASQ2, CAV3, CBL, CRYAB, CSRP3, CTF1, DES, DSC2, DSC2, DSG2, DSP, DTNA, GLA, HRAS, JUP, KRAS, LAMA4, LAMP2, LDB3, LMNA, MAP2K1, MAP2K2, MYBPC3, MYH6, MYH7, MYL2, MYL3, MYLK2, MYOZ2, MYPN, NEXN, NRAS, PKP2, PLN, PRKAG2, PTPN11, RAF1, RBM20, RYR2, SCN5A, SGCD, SHOC2, SOS1, TAZ, TCAP, TMEM43, TMPO, TNNC1, TNNI3, TNNT2, TPM1, TTN, TTR, VCL* | 6 | 7 | 567 | 190 | 337 | 420 | 7 |
| Total | 568 | 117 genes | 77 | 7 | 1204 | 443 | 919 | 762 | 21 |

*The same seven samples analyzed for 1000G data were used in each NGS panel.
1000G, 1000 Genome Project; NGS, next-generation sequencing.

unique bioinformatic pipelines. Recently published laboratory standards from the College of American Pathologists states, "[We] preferred to give laboratories performing NGS-based assays flexibility in determining when confirmatory testing should be performed, [and] how this testing is performed…."[4,p484] Thus, the onus of establishing and adhering to a set of metrics and best practices for ensuring the quality of NGS test results falls on the laboratory.

Recently, Strom et al[5] published a report that examined the performance of 144 clinical whole-exome sequencing variants and assessed the necessity of confirmatory testing. They concluded that the NGS variant calls were accurate and that their laboratory would discontinue routine Sanger confirmation of clinical exome sequencing variants meeting quality metric thresholds. In a similar context, we report here on the comparison of 1204 NGS variant occurrences (919 unique) with Sanger sequencing, arising from 117 target-captured genes. We also discuss our additional comparison of 762 NGS variants to data from the 1000 Genomes Project (1000G).

## Materials and Methods

### Samples

Seventy-seven deidentified patient DNA samples and seven DNA samples from 1000G, purchased from the Coriell Cell Repository (Camden, NJ; NA18517, NA07357, NA12003, NA18507, HG00421, HG00422, and HG01060) were included in our comparative study. Variant calls from 1000G were achieved by integrating diverse data from multiple technologies in several testing sites, including low-coverage, whole-genome and whole-exome data.[6] Data from 1000G phases 1 and 3 were downloaded (*http://browser.1000genomes.org/index.html*; phase 1 data accessed on May 14, 2014, and phase 3 data accessed on November 4, 2014). The data were filtered using the target regions, and the variants were translated into the correct strand variant calls.

### NGS Library Preparation and Sequencing

Data from multiple NGS targeted multigene panel tests covering 117 genes were used for this study (Table 1). The SureSelect Custom Target Enrichment System (Agilent Technologies, Inc., Santa Clara, CA) was designed for referencing sequence GRCh37/hg19 (February 2009) using Agilent eArray. This capture chemistry was used for library preparation and involved mechanical shearing (ultrasonication) of genomic DNA to 150- to 350-Bp fragments with an LE220 ultrasonicator (Covaris, Woburn, MA). Library preparation was automated on the Biomek FXp Laboratory Workstation (Beckman Coulter, Inc., Jersey City, NJ) to include: enzyme-mediated end-repair, adenine addition a-tailing, adapter oligonucleotide ligation, and enrichment of adapter-ligated fragments via limited-cycle PCR. After each of these process steps, the libraries were purified with