Shruti Bhide[1], Holly LaDuca[1], Kelly Gonzalez[1], Layla Shahmirzadi[1], Sha Tang[1], Soren Fischbach[1], Hong Lu[1], Hsiao-Mei Lu[1], Xiang Li[1]
1) Ambry Genetics, Aliso Viejo, CA

Mutation Detection Rate of Exome Sequencing Using Next Generation Sequencing Panel Results as a Reference

BACKGROUND

- The decision to pursue next generation sequencing (NGS) panels vs. exome sequencing can be complicated by differences in coverage and the number of genes analyzed for the two testing methods (Wooderchack-Donahue, 2012).
- Complete coverage of specified regions is achieved on NGS panels at our laboratory, as follow -up Sanger sequencing is performed for any regions with insufficient coverage to make a reliable variant call. In comparison, complete gene coverage on exome sequencing in our laboratory is roughly ~ 90%.
- Factors such as GC-rich and repetitive regions result in difficulty capturing the entire exome (Clark, 2011; Asan, 2011; Majewski, 2011). In addition, exome sequencing has limitations in the detection of insertions and deletions, trinucleotide repeats, and copy number variations (O'Daniel, 2012).
- While NGS panels are currently superior to exome sequencing in coverage, exome sequencing carries the advantages of analyzing a greater number of genes including novel disease genes and incorporating samples from affected and unaffected family members into the analysis.

Table I. Next Generation Sequencing (NGS) Panels Included in Exome Sequencing Comparison

NGS Panel Name	Genes Included On Panel
PanCardio	ABCC9, ACTC1, ACTN2, AKAP9, ANK2, ANKRD1, BAG3, CACNA1C, CACNA2D1, CACNB2, CALR3, CASQ2, CAV3, CRYAB, CSRP3, DES, DMD, DSC2, DSG2, DSP, EMD, EYA4, FXN, GATA4, GLA, GPD1L, ILK, JAG1, JPH2, JUP, KCNE1, KCNE2, KCNE3, KCNH2, KCNJ2, KCNJ8, KCNQ1, LAMP2, LDB3/ZASP, LMNA, MYBPC3, MYH6, MYH7, MYL2, MYL3, MYOM1, MYOZ2, MYPN, NEBL, NEXN, NKX2.5, PDLIM3, PKP2, PLN, PRKAG2, PTPN11, RAF1, RBM20, RYR2, SCN1B, SCN3B, SCN4B, SCN5A, SGCD, SNTA1, TAZ, TBX1, TBX5, TCAP, TMEM43, TMPO, TNNC1, TNNI3, TNNT2, TPM1, TTN, TTR, TXNRD2, VCL
Cardiomyopathy	ABCC9, ACTC1, ACTN2, ANKRD1, BAG3, CALR3, CAV3, CRYAB, CSRP3, DES, DMD, DSC2, DSG2, DSP, EMD, EYA4, FXN, GLA, ILK, JPH2, JUP, LAMP2, LDB3/ZASP, LMNA, MYBPC3, MYH6, MYH7, MYL2, MYL3, MYOM1, MYOZ2, MYPN, NEBL, NEXN, PDLIM3, PKP2, PLN, PRKAG2, PTPN11, RAF1, RBM20, RYR2, SCN5A, SGCD, TAZ, TCAP, TMEM43, TMPO, TNNC1, TNNI3, TNNT2, TPM1, TTN, TTR, TXNRD2, VCL
Hypertrophic Cardiomyopathy (HCM)	ACTC1, ACTN2, ANKRD1, CALR3, CAV3, CSRP3, DES, FXN, GLA, JPH2, LAMP2, MYBPC3, MYH6, MYH7, MYL2, MYL3, MYOM1, MYOZ2, NEXN, PLN, PRKAG2, PTPN11, RAF1, TCAP, TNNC1, TNNI3, TNNT2, TPM1, TTN, TTR, VCL
Dilated Cardiomyopathy (DCM)	ABCC9, ACTC1, ACTN2, ANKRD1, BAG3, CRYAB, CSRP3, DES, DMD, EMD, EYA4, ILK, LAMP2, LDB3/ZASP, LMNA, MYBPC3, MYH6, MYH7, MYPN, NEBL, NEXN, PDLIM3, PLN, RBM20, SCN5A, SGCD, TAZ, TCAP, TMPO, TNNC1, TNNI3, TNNT2, TPM1, TTN, TTR, TXNRD2, VCL
Arrhythmia	AKAP9, ANK2, CACNA1C, CACNA2D1, CACNB2, CASQ2, CAV3, DES, DSC2, DSG2, DSP, GPD1L, JUP, KCNE1, KCNE2, KCNE3, KCNH2, KCNJ2, KCNJ8, KCNQ1, LMNA, PKP2, RYR2, SCN1B, SCN3B, SCN4B, SCN5A, SNTA1, TMEM43
LongQT Syndrome	AKAP9, ANK2, CACNA1C, CAV3, KCNE1, KCNE2, KCNH2, KCNJ2, KCNQ1, SCN4B, SCN5A, SNTA1
Brugada Syndrome	CACNA1C, CACNA2D1, CACNB2, GPD1L, KCNE3, KCNJ8, SCN1B, SCN3B, SCN5A
Marfan, Aneurysm, and Related Disorders	ACTA2, CBS, FBN1, FBN2, MYH11, COL3A1, SLC2A10, SMAD3, TGFBR1, TGFBR2
X-Linked Intellectual Disability (XLID)	ABCD1, ACSL4/FACL4, AGTR2, AP1S2, ARHGEF6, ARHGEF9, ARX, ATP6AP2, ATP7A, ATRX/XNP/XH2, BCOR, BRWD3, CASK, CDKL5, CUL4B, DCX, DKC1, DLG3, FANCB, FGD1, FLNA/FLN1, FMR1, FTSJ1, GDI1, GJB1/CMTX1, GK, GPC3, GRIA3, HCCS, HPRT, HSD17B10/HADH2, HUWE1, IDS, IL1RAPL1, KDM5C/JARID1C/SMCX, KIAA2022, L1CAM, LAMP2, MAOA, MECP2, MED12/HOPA, MID1, MTM1, NDP, NDUFA1, NHS, NLGN3, NLGN4/NLGN4X, NXF5, OCRL, OFD1, OPHN1, OTC, PAK3, PDHA1, PGK1, PHF6, PHF8, PLP1, PORCN, PQBP1, RPL10, PRPS1, RPS6KA3/RSK2, SHROOM4/KIAA1202, SLC9A6, SLC16A2/MCT8, SMC1A/SMC1L1, SMS, SOX3, SRPX2, SYN1, SYP, TIMM8A, TSPAN7/TM4SF2, UBE2A, ZDHHC9, ZNF41, ZNF81, ZNF674, ZNF711
Primary Ciliary Dyskinesia (PCD)	DNAAF1/ LRRC50, DNAAF2/ c14orf104, DNAH5, DNAH11, DNAI1, DNAI2, RSPH4A, RSPH9, TXNDC3, OFD1, RPGR, CFTR
Hereditary Cancer Panels: BreastNext	ATM, BARD1, BRCA1, BRCA2, BRIP1, CDH1, CHEK2, MRE11A, MUTYH, NBN, PALB2, PTEN, RAD50, RAD51C, STK11, TP53
OvaNext	ATM, BRIP1, BRCA1, BRCA2, CDH1, CHEK2, EPCAM, MLH1, MSH2, MSH6, MUTYH, NBN, PALB2, PMS2, PTEN, STK11, TP53
ColoNext	APC, BMPR1A, CDH1, CHEK2, EPCAM, MLH1, MSH2, MSH6, MUTYH, PMS2, PTEN, SMAD4, STK11, TP53
CancerNext	APC, ATM, BARD1, BRCA1, BRCA2, BRIP1, BMPR1A, CDH1, CHEK2, EPCAM, MLH1, MRE11A, MSH2, MSH6, MUTYH, NBN, PALB2, PMS2, PTEN, RAD50, RAD51C, SMAD4, STK11, TP53

Table II. Sequence Mutations Identified on NGS Panels Likely to Have Been Detected on Exome Sequencing

NGS Panel	Sequence Mutations Identified on Panel N	Number of Sequence Mutations that Would Have Likely Been Detected on Exome Sequencing[a] N (%)
PanCardio	6	6 (100)
Cardiomyopathy	7	6 (85.7)
HCM	9	8 (88.9)
DCM	5	5 (100)
Arrhythmia	3	3 (100)
LongQT	2	2 (100)
Brugada	0	n/a
Marfan	31	31 (100)
XLID	10	9 (90.0)
PCD	42	40 (95.2)
BreastNext	54	54 (100)
OvaNext	19	19 (100)
ColoNext	46	45 (97.8)
CancerNext	36	36 (100)
Total	270	264 (97.8)

* a Coverage >10x

METHODS

- Our internal database was reviewed for pathogenic and suspected pathogenic sequence mutations reported on 14 different NGS panels offered by our laboratory (Table I). The following types of sequence mutations were included in our search: missense/nonsense, splicing, and small insertions/deletions (≤20 bp) (Stenson, 2009). The average coverage of the respective nucleotide positions was then calculated from our exome sequencing database. Data from twenty-six exomes were included in this analysis. For an alteration to be detected on exome sequencing, the nucleotide position must have a minimum coverage of 10x, based on our laboratory's clinical test specificity. Therefore, we analyzed exome base pair coverage at nucleotide positions corresponding to the mutations detected from NGS panels to determine whether each alteration would have likely been covered.

Next Generation Sequencing (NGS) Panels

- Genomic deoxyribonucleic acid (gDNA) was isolated from the patient's specimen (blood or saliva) using a standardized kit and quantified by agarose gel electrophoresis (Qiagen, Valencia, CA and DNAgenotek, Kanata, Ontario, Canada for blood and saliva, respectively). Sequence enrichment was carried out by incorporating the gDNA into microdroplets along with primer pairs designed to the target gene coding exons followed by polymerase chain reaction (PCR) and next-generation sequencing (RainDance Technologies, Billerica, MA and Illumina, San Diego, CA, respectively). A secondary sequencing method was performed for any regions with insufficient read depth coverage for reliable heterozygous variant detection. Variant calls other than polymorphisms were verified by sequencing in sense and antisense directions. For the NGS hereditary cancer panels only, gene copy number analysis was performed for all genes on the panels via a targeted chromosomal microarray (Aglient, Santa Clara, CA).
- Variant classification was based on thorough assessment and review of available evidence (e.g. population frequency information, published case reports and functional studies, internal co-occurrence and co-segregation data, evolutionary conservation, and *in silico* predictions).

Exome Sequencing

- Genomic deoxyribonucleic acid (gDNA) was isolated from whole blood from patient samples provided to the laboratory for diagnostic exome sequencing (DES). Samples were prepared using the SureSelect Target Enrichment System (Agilent Technologies, Santa Clara, CA) and sequenced using paired-end, 100-cycle chemistry on the Illumina HiSeq 2000 (Illumina, San Diego, CA). Approximately 90% of bases have basecall quality scores >Q30 with our laboratory's established run conditions for exome sequencing.

Table III. Sequence Mutations Identified on NGS Panels that Would Have Escaped Detection on Exome Sequencing

Gene (NGS Panel)	Mutation	Average Coverage of Nucleotide Position	Average Coverage of Exon	% GC Content of Exon	Reason for Poor Coverage
MSH2 (ColoNext)	c.97A>C	7.35x	7.33x	66.82	EXON 1/GC-rich
RPGR (PCD)	c.28+1G>A	7.56x	7.22x	64.29	1 nucleotide after EXON 1/GC-rich
CFTR (PCD)	c.3717+12191C>T	1.43x	n/a	n/a	Deep Intronic
SYN1 (XLID)	c.95DEL	6.12x	10.48x	72.68	EXON1/GC-rich
LMNA (Cardiomyopathy)	c.1622G>A	3.48x	3.06x	64.13%	GC-rich
MYBPC3 (HCM)	c.2727C>A	9.81x	14.16x	67.41%	GC-rich

RESULTS/ DISCUSSION

- In total, 270 pathogenic and suspected pathogenic sequence mutations have been reported on NGS panels at our laboratory (Table II). 97.8% (n=264) of the respective nucleotide positions yielded coverage of 10x or higher on exome sequencing, with an average coverage of 129.6x (min= 14.7, max= 405.9). The percentage of mutations that would likely have been identified on exome sequencing varied by testing panel, ranging from 85.7% to 100% (Table II).
- Respective nucleotide positions for the remaining six mutations yielded coverage under 10x on exome sequencing. Therefore, these mutations would not have been detected on exome sequencing. Likely explanations for poor coverage of these nucleotide positions are listed in Table III. One of the nucleotide positions was located deep in an intron, and introns are not targeted on exome sequencing. Three of the nucleotide positions were located in or adjacent to the first exon of the respective genes. First exons can be difficult to capture, as they tend to have a higher GC content than other exons (Kalari, 2006; Asan, 2012). GC-content for all three corresponding exons exceeded 60%, which is elevated compared to the average genome-wide GC content of 41% (Lander, 2001). The remaining two nucleotide positions were located within exons with high GC-content as well.
- For the four hereditary cancer panels offered by our laboratory, deletion/duplication analysis is performed for all genes on the panels via a targeted chromosomal microarray and MLPA for *PMS2* when indicated (due to pseudogene interference); therefore, our internal database was also reviewed for pathogenic gross deletions/duplications identified on the four hereditary cancer NGS panels. In our dataset, a total of 13 gross deletions/duplications were detected on hereditary cancer NGS panels that would not likely have been detected on exome sequencing.
- Gross deletion/duplication analysis is not included in the other NGS panels used in this analysis.
- Considering these gross deletions/duplications, 93.3% (264/283) of the mutations identified on our panels would likely have been detected on exome sequencing.

CONCLUSIONS

- These data suggest that while expected coverage is higher on NGS panels than exome sequencing, the majority of pathogenic and suspected pathogenic sequence mutations reported by our laboratory to date on NGS panels would have likely been identified on exome sequencing as well.
- For clinicians considering an exome sequencing approach for hereditary cancer susceptibility, ordering a chromosomal microarray targeted to cancer susceptibility genes either concurrently or reflexively would result in an increased sensitivity.
- This information may aid clinicians in deciding on the most appropriate testing strategy for patients.

REFERENCES

1. Wooderchak-Donahue, W.L., et al., *A direct comparison of next generation sequencing enrichment methods using an aortopathy gene panel- clinical diagnostics perspective.* BMC Med Genomics, 2012. **5**: p. 50.
2. Clark, M.J., et al., *Performance comparison of exome DNA sequencing technologies.* Nat Biotechnol, 2011. **29**(10): p. 908-14.
3. Asan, et al., *Comprehensive comparison of three commercial human whole-exome capture platforms.* Genome Biol, 2011. **12**(9): p. R95.
4. Majewski, J., et al., *What can exome sequencing do for you?* J Med Genet, 2011. **48**(9): p. 580-9.
5. O'Daniel, J.M. and K. Lee, *Whole-genome and whole-exome sequencing in hereditary cancer: impact on genetic testing and counseling.* Cancer J, 2012. **18**(4): p. 287-92.
6. Stenson, P.D., et al., *The Human Gene Mutation Database: 2008 update.* Genome Med, 2009. **1**(1): p. 13.
7. Kalari, K.R., et al., *First exons and introns--a survey of GC content and gene structure in the human genome.* In Silico Biol, 2006. **6**(3): p. 237-42.
8. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.