Report

Analytic Validation

Blueprint Genetics

Bluepring Genetics Analytic Validation Report

Executive summary

The analytic validity of Blueprint Genetics' genetic tests has been demonstrated for single nucleotide polymorphisms (SNPs) and INDELS. We established SNP analytic validity in a study using a standard reference sample with high-quality variant calls. To validate for INDELs, we used a cohort of reference samples from the 1000 Genomes Project. The analytic validation results show that Blueprint Genetics' NGS assays achieve, on average, 0.992 sensitivity, 1.000 specificity, 0.992 positive predictive value, 1.000 accuracy, and 0.995 reproducibility for detection of SNPs. For detection of INDELs of up to 18 bases, our assays achieve >0.92 sensitivity.

Background

Blueprint Genetics is a diagnostic laboratory that uses a targeted Next-Generation Sequencing (NGS) approach called Oligonucleotide-Selective Sequencing (OS-Seq) (1) to test DNA variations in inherited diseases. At the time of this study, Blueprint Genetics' test offering included 18 gene panels. Tests were analyzed using two assays and protocols, named BpG-1.1 (Table 1) and BpG-2.1 (Table 2), depending on which test is ordered for the patient.

Analytic validation confirms that specific tests are suitable for their intended use. Several regulations, including the Clinical Laboratory Improvement Act of 1988 (CLIA), as well as various quality standards for laboratories require validation of analytical methods before and during routine use. While there are no specific regulations addressing the methods used to validate genetic tests, the American College of Medical Genetics (ACMG) has established some guidelines (2). According to the ACMG guidelines, laboratories using NGS should validate their methods against reference samples for which Sanger sequencing data or other high-quality and independent data exists. Blueprint Genetics has performed the analytic validation studies according to the ACMG guidelines and in accordance with the requirements of CLIA and ISO 15189:2012 quality standard for medical laboratories.

Table 1. Tests in the BpG-1.1 assay.

Tests	Genes	Exons	Bases
Heart	133	2,459	489,692
Pan Cardiomyopathy	103	2,018	408,871
Core Cardiomyopathy	72	1,439	304,264
Noonan	12	139	22,479
Arrhythmia	47	783	147,353
Long QT	16	251	55,825
Short QT	6	142	22,201
Brugada syndrome	18	293	57,420
Catecholaminergic Polymorphic VT	6	213	35,695
Becker and Duchenne Muscular Dystrophy	1	81	13,471
Emery-Dreifuss Muscular Dystrophy	7	616	174,250
GM1 Gangliosidosis	1	16	2,306
All panels in BpG-1.1 assay*	136	2,486	495,567

*Total of unique genes, exons, and bases. Some test panels have overlap.

Table 2. Tests in the BpG-2.1 assay.

Panel	Genes	Exons	Bases
Aorta	18	498	75,998
Pulmonary Artery Hypertension	7	59	12,638
Marfan	9	228	35,711
Nephrotic Syndrome	9	186	33,427
Periodic Fever	9	95	18,581
Hyperlipidemia	12	107	28,330
All panels in BpG-2.1 assay*	60	919	159,198

*Total of unique genes, exons, and bases. Some test panels have overlap.

Reference samples

Blueprint Genetics used a gold standard reference sample NA12878 (Coriell Cell Repositories, NJ) for analytic validation of SNPs. The National Institute of Standards and Technology (NIST) has produced a highly confident set of SNP genotypes for NA12878 by integrating several whole genome sequencing datasets generated using different sequencing platforms (3). For the validation of INDELs, we used a cohort of reference samples from the 1000 Genomes Project.

In this analytic validation study, 82 replicates of NA12878 (Table 3) and 43 reference samples from the 1000 Genomes Project (Table 4) were analyzed using Blueprint Genetics' NGS assays. Each sample was processed independently and according to documented procedures, BpG-QA-0006 for validation of SNPs and BpG-QA-0007 for validation of INDELs.

Table 3. Reference samples used in analytic validation of SNPs and INDELs.

Cohort	Assay	Sequenced to date	Analytes
NA12878 replicates	BpG-1.1	61	191 SNPs
NA12878 replicates	BpG-2.1	21	95 SNPs
1000 Genomes	BpG-1.1	32	135 INDELs
1000 Genomes	BpG-2.1	40	70 INDELs

Sequencing depth and coverage evaluation

Sequencing depth and coverage are the primary means of evaluating the performance of targeted NGS assays. Sequencing depth refers to the number of times a specific nucleotide is read during sequencing and is often reported as an average for targeted genomic regions. Coverage denotes the breadth of coverage over targeted genomic regions, defined as the percentage of target bases that are sequenced at a given sequencing depth. Previous research shows that even in cases of calling heterozygous variants, more than 13 overlapping sequence reads is rarely required. Thus, a sequencing depth of 15x is generally considered adequate for making confident calls on genetic variants (4). A sequencing depth threshold of 15x is also supported by our in-house analytic validation of variant calling (see Figure 1).

To evaluate the average sequencing depth of Blueprint Genetics' NGS assays, the number of overlapping sequence reads was collected at each nucleotide position across the cumulative regions targeted by the BpG-1.1 assay (495,567 base-pairs long) and the BpG-2 assay (179,487 base-pairs long). The average was then calculated using the median, which is less biased by outliers compared to the mean.

Table 4. Reference samples from the 1000 Genomes Project used for INDEL validation.

Sample	BpG-1.1	BpG-2.1
HG00134	yes	yes
HG00154	yes	no
HG00329	yes	yes
HG00332	yes	yes
HG00356	yes	yes
HG00359	yes	no
HG00361	yes	yes
HG00362	yes	yes
HG00449	yes	yes
HG00457	yes	yes
HG00543	yes	yes
HG00584	yes	yes
HG00625	yes	yes
HG00653	yes	yes
HG01198	no	yes
HG01353	yes	yes
HG01366	yes	yes
NA18530	yes	yes
NA18550	yes	yes
NA18609	no	yes
NA18631	yes	yes
NA18975	yes	yes
NA18984	yes	yes
NA18988	yes	yes
NA19036	yes	yes
NA19070	no	yes
NA19093	no	yes
NA19096	yes	yes
NA19113	yes	yes
NA19130	yes	yes
NA19150	no	yes
NA19225	no	yes
NA19332	yes	yes
NA19375	no	yes
NA19436	no	yes
NA19818	no	yes
NA19920	no	yes
NA20340	no	yes
NA20359	yes	yes
NA20414	yes	yes
NA20506	yes	yes
NA20530	yes	no
NA20533	yes	yes

a. BpG-1.1 assay

b. BpG-2.1 assay

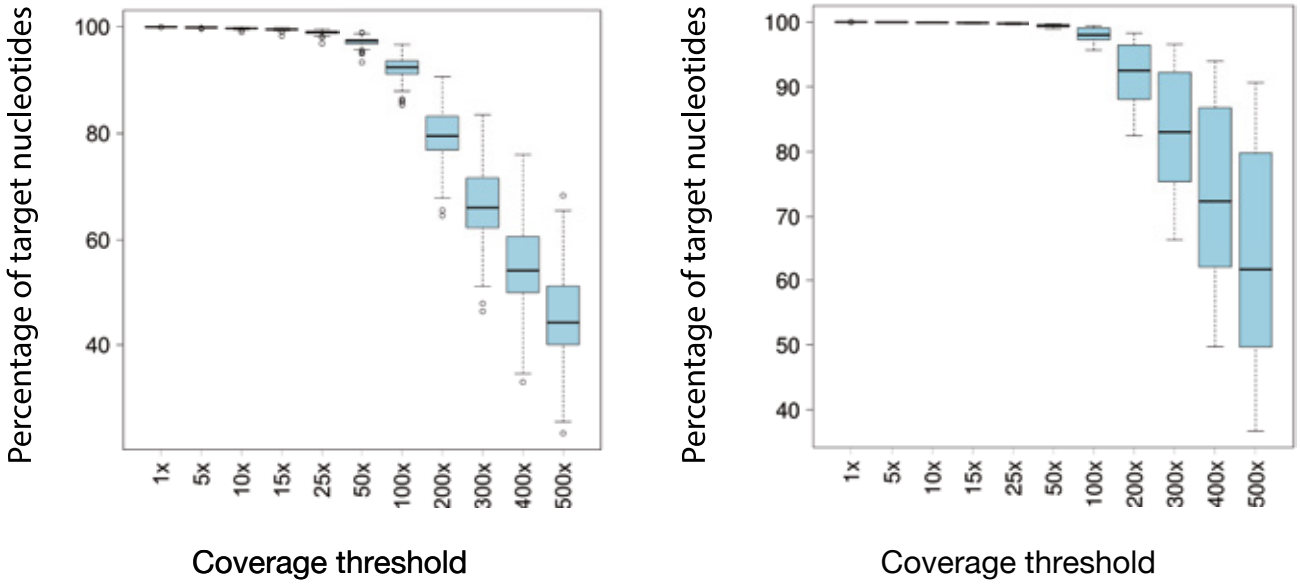

Figure 1. Percentage of target nucleotides that are above various sequence coverage thresholds (1x – 500x) in a) BpG-1.1 assay and b) BpG-2.1 assay. The number of reference samples independently analyzed for BpG-1.1 and BpG-2.1 assays were 61 and 21, respectively.

Table 5. Sequencing depth and coverage performance of Blueprint Genetics' NGS tests.

Test	Median sequencing depth	Relative speed	% of nucleotides >15x
Heart	441	1.23	99.49
Pan cardiomyopathy	423	1.23	99.45
Core cardiomyopathy	470.5	1.18	99.57
Noonan	395	1.25	99.13
Arrhythmia	693	1.07	99.83
Long QT	624.5	1.09	99.90
Short QT	461.5	1.06	99.65
Brugada syndrome	634.5	1.11	99.78
Catecholaminergic Polymorphic VT	725	1.10	100.00
Becker and Duchenne Muscular Dystrophy	294	1.20	99.42
Emery-Dreifuss Muscular Dystrophy	413.5	1.13	99.58
GM1 Gangliosidosis	328.25	1.12	100.00
BpG-1.1 assay	437	1.23	99.50
Aorta	757	0.89	100.00
Pulmonary Artery Hypertension (PAH)	811	0.87	100.00
Marfan	709	0.82	100.00
Nephrotic Syndrome	443	0.88	99.33
Periodic Fever	687	0.76	100.00
BpG-2.1 assay	618	0.91	99.86

The median of the average sequencing depths across the 61 reference samples analyzed with the BpG-1.1 assay was 437. The smaller BpG-2.1 assay performed at an even higher level; the median of average sequencing depth across the 21 reference samples was 618.

To demonstrate the sequencing coverage performance, we calculated the percentage of nucleotides in the target regions exceeding given sequencing depths (1x – 500x). As illustrated in Figure 1a, the BpG-1.1 assay delivered a very high (>99) percentage of target nucleotides with depth above 15x, the accepted level at which confident genotype calls can be made. For sequence coverage, the BpG-2.1 assay yielded percentages in excess of 99 all the way up to 50x coverage (Figure 1b) whereas a 15x coverage threshold resulted in close to 100% coverage of target nucleotides. At depths >100x, >98% of target nucleotides were covered.

The sequencing depth and coverage performance of individual tests included in the BpG-1.1 and BpG-2.1 assays were also assessed (Table 5). Depending on the test, the median average depth across the control samples ranged between 294 and 811. Sample to sample variability was evaluated by calculating the relative spread, which relates the interquartile range (75%–25%) of coverage to the median coverage. The relative spread varied from 1.06 to 1.25 in the BpG-1.1 assay and 0.76 to 0.89 in the BpG-2.1 assay. While these values do not imply a high degree of homogeneity, the variation occurs because the number of samples in a sequencing run varies from one run to another. Some runs do not contain a full set of samples, resulting in a higher than normal coverage, and thus increasing the relative spread. The percentage of targeted nucleotides covered to a depth of at least 15x varied between 99.13% and 100%.

Methods used in comparison study

NA12878. SNP detection using Blueprint Genetics' NGS assays was analytically validated using a "gold standard" reference sample NA12878. To establish a set of highly confident SNP and INDEL calls for NA12878, 12 datasets generated using 5 different sequencing platforms have been produced. The integrated dataset for NA12878 contains a bed file that allows exclusion of regions and variant locations that are uncertain due to low coverage, genotypes called in <3 datasets, locations with unresolved discordant genotypes, locations where most datasets have evidence of bias (systematic sequencing errors, local alignment problems, mapping problems, or abnormal allele balance), variants inside possible deletions, known segmental duplications, and structural variants reported in dbVar for NA12878. In all, quality filtering excludes ~15% of the non-N bases in the GRCh37 reference assembly. VCF and BED files for NA12878, as well as a README.NIST with up-to-date information were downloaded from the NIST website (5). The complete NIST dataset for NA12878 was used in the comparison and no additional filtering has been applied in the analysis.

1000 Genomes. Blueprint Genetics' NGS assays target too few of the INDELs identified in NA12878 to allow comprehensive analytical validation using the gold standard reference sample. Instead, we collected a cohort of samples from the 1000 Genomes Project (1000GP) (6), to yield a sufficient number of INDELs for rigorous validation. The 1000GP is an international effort to establish a catalog of human genome variation. In all, genomes of 1,092 individuals have been sequenced to date. However, whole genome sequencing data provided by the 1000GP has limitations related to low quality. Typically, whole genome sequencing achieves an average depth of 20x, with only 80% of the genome reaching a 10x sequencing depth (7). Therefore, 1000GP data does not meet the quality requirements set out for clinical NGS approaches. To generate a high-quality collection of INDELs from the 43 analyzed reference samples, we omitted low quality and ambiguous variants in the 1000GP data by discarding variants in regions with low sequencing depth and selecting variants that were shared by at least five individuals. This approach allows accurate estimation of true positive and false negative rates, but does not allow their confident assessment. Hence, specificity, accuracy, and positive predictive values were not calculated for this reference set.

Analytic range

We defined the analytic range by collecting all true positive variants from NA12878 and 1000GP samples. The collection of detected variants included SNPs and INDELs with lengths up to 18 bases (Table 6). The analytical target regions of Blueprint Genetics' NGS assays include all coding exons and flanking 8-base intronic regions. Exon definitions for the target regions were derived from consensus coding sequence (CCDS) project (8) and included all exons from all transcripts. Additional non-coding regions with known pathogenic mutations reported in the literature have also been included in the targets. For the *KCNK3* gene included in the PAH test, the target region omits exon 1 due to technical limitations related to the high degree of repetitive sequences combined with high GC-percentage in the flanking region, as well as in the exon itself.

Table 6. Analytic range of Blueprint Genetics' NGS assays.

Variants	Length	Validated variants
SNPs	1 nucleotide	286
INDELs	1–5 nucleotides	181
INDELs	6–10 nucleotides	24
INDELs	18 nucleotides	1

Results

The analytic validity of Blueprint Genetics' NGS assays is high, as demonstrated by the agreement of data between our test results and the gold standard references (Tables 7 and 8).

The SNP validation analysis revealed very high levels for all the performance metrics across both assays. Sensitivity scored from 0.991 in the BpG-1.1 assay to 1.000 for the BpG-2.1 assay (0.992 combined), whereas the positive predictive value ranged from 0.991 in BpG-1.1 to 0.998 for BpG-2.1 (0.992 combined). Both specificity and accuracy were calculated to 1.000 across the assays, whereas reproducibility was found to be 0.997 and 1.000, in BpG-1.1 and BpG-2.1, respectively.

The INDEL validation analysis showed high sensitivity to detect the 206 INDELs analyzed across the 43 samples, with 195 positive identifications and 11 false negatives. That resulted in sensitivity scores of 0.956 and 0.929 for BpG-1.1 and BpG-2.1 assays, respectively. The longest INDEL included in the set of reference samples, an 18 base-pair deletion, was correctly identified, validating that the assays are suitable for detecting even longer INDELS.

Table 7. Analytic validation of SNPs.

Assay	BpG-1.1	BpG-2.1	All
True positives	11,545	1,995	13,540
False negatives	106	0	106
False positives	108	5	113
True negatives	25,564,561	3,286,439	28,851,000
Sensitivity	0.991	1.000	0.992
Specificity	1.000	1.000	1.000
PPV	0.991	0.998	0.992
Accuracy	1.000	1.000	1.000
Reproducibility (95% CI)	0.993 (0.989–0.999)	0.998 (0.995–1.000)	0.995 (0.992–0.998)

Table 8. Analytic validation of INDELs.*

Assay	Length	True positives	False negatives	Sensitivity
BpG-1.1	1–5	108	6	0.947
BpG-2.1	1–5	62	5	0.925
BpG-1.1	6–18	22	1	0.957
BpG-2.1	6–18	3	0	1.000

* This table was revised from the Dec 11, 2014 version to include INDELs up to 18 bases.

Conclusions

- Results obtained using the Blueprint Genetics' NGS assays were confirmed by analytic validation of independent data from golden standard reference samples.
- Blueprint Genetics' tests show high sensitivity, specificity, positive predictive value, accuracy, and reproducibility to detect SNPs. In addition, detection of INDELs was validated at high sensitivity.
- Our conclusion is that Blueprint Genetics' NGS assays are suitable for detecting SNPs and INDELs from clinical samples.

References

1. Myllykangas, S., Buenrostro, J.D., Natsoulis, G., Bell, J.M., and Ji, H.P. Efficient targeted resequencing of human germline and cancer genomes by oligonucleotide-selective sequencing. Nat Biotech 2011;11:1024-7.
2. Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, et. al. ACMG Clinical Laboratory Standards for Next-Generation Sequencing. Genet Med 2013;15:733-47.
3. Zook JM, Chapman B, Wang J, Mittelman D, Hofmann O, Hide W, Salit M. Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. Nat Biotech 2014;32:246-51.
4. Meynert AM, Bicknell LS, Hurles ME, Jackson AP, Taylor MS. Quantifying single nucleotide variant detection sensitivity in exome sequencing. BMC Bioinformatics 2013;14:195. ftp://ftp-trace.ncbi.nih.gov/giab/ftp/data/NA12878/variant_calls/NIST.
5. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56-65.
6. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet 2014;15:121-32.
7. Pruitt KD, Harrow J, Harte RA, Wallin C, Diekhans M, Maglott DR, et al. The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. Genome Res 2009;19:1316-23.

Blueprint Genetics

Blueprint Genetics Customer Support
support@blueprintgenetics.com
Phone: +358 40 2511 372
Fax: +358 9 8565 7177

www.blueprintgenetics.com