Sequencing Panel for 4813 Genes with Known Associated Clinical Phenotypes

TruSight™ One Sequencing Panel provides high depth of coverage for accurate variant calling

Introduction

Clinical research laboratories often make use of several molecular profiling assays when studying a particular disease. Each assay targets a single gene and provides a limited amount of information; the cost and time to results can be significant for multiple assays. Next-generation sequencing (NGS) offers a faster and more cost-effective method for clinical genomics research.

To assist with this challenge, Illumina introduces the TruSight One Sequencing Panel for genomic analysis of the coding regions of 4813 genes with associated clinical phenotypes. The panel achieves high depth of coverage (> 20×) for multiple samples when sequenced on a MiniSeq™, MiSeq®, NextSeq®, or HiSeq® System. Researchers can choose to analyze all genes on the panel or focus on a specific subset relevant to the disease of interest. TruSight One is provided with VariantStudio software for analysis, classification, and reporting of genomic variants.

This technical note describes the TruSight One Sequencing Panel performance, how gene content was chosen, and the identity of target exons not covered by the panel or eliminated due to a high prevalence of incorrect variant calls.

TruSight One Sequencing Panel Performance

While probe-capture methods are powerful in their ability to enrich for and sequencemultiple genomic targets, they are often unable to achieve high levels of coverage uniformity. Due to the variable affinity of each probe for its target region, some regions register overly high amounts of coverage, while others do not achieve enough coverage to call variants accurately.

To improve the coverage uniformity of the TruSight One Sequencing Panel, Illumina employed 2 different methods during its design. Regions underrepresented relative to the entire panel were supplemented with additional probes to increase their coverage. These probes were added to the initial pool and their performance was verified. This process was repeated 3 times to make sure that every region captured by this method was adequately covered. Similarly, overrepresented regions (more than 2 times the mean coverage) were identified, and nonbiotinylated oligos consisting of the reverse complement of the probe sequence were added to the pool. These nonbiotinylated probes blocked the efficiency of the sense probes, decreasing representation of these regions and increasing representation of other regions (on average, this increase was ~ 5× per region). For example, a region covered at 15× is now covered at 20×. The result is increased uniformity of the TruSight One panel to a level where > 97% of the regions are covered at > 0.2× of the mean, which is the highest uniformity demonstrated for a probe-based enrichment panel.

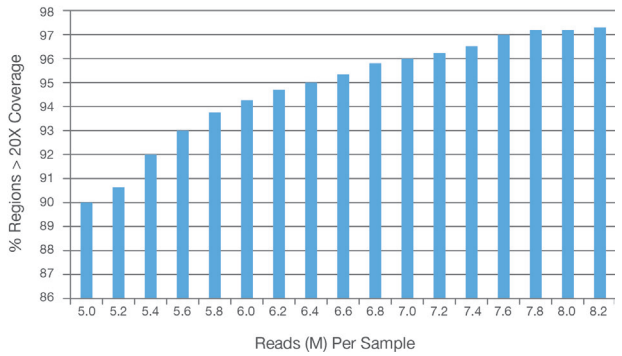

Figure 1: Reads vs. % Regions Covered—Depiction of the number of reads per sample vs. the percentage of regions on the TruSight One Sequencing Panel that achieve over 20× coverage.

Figure 1 depicts the percentage of regions in the TruSight One Sequencing Panel covered at 20× depth or higher, given a certain number of reads per sample. The high percentage of regions covered, even at a low number of reads, is a direct result of the high uniformity of this panel. For example, at 6.4 M reads, 95% of the regions are covered at 20× or higher; thus, at only 77% of the capacity of a MiSeq System (using the v3 chemistry and software), 3 samples can be sequenced at this high depth of coverage. Presumably, in the case of a child with a rare genetic disease, these 3 samples can be a trio (father, mother, and affected child). When used with trio subtraction techniques that can be performed with VariantStudio 2.0 software, this analysis greatly enhances the chances of identifying the causative variant.

Genes Targeted

When designing the TruSight One Sequencing Panel, Illumina attempted to include all genes with known associated clinical phenotypes. This design enables clinical research laboratories to use the panel as the basis for smaller sequencing panels that comprise a comprehensive portfolio of sequencing assays. As a baseline, Illumina included all genes in the TruSight Exome Content Set[1] (2761 genes, Figure 2). Genes not present in this set, but listed in the Human Gene Mutation Database (HGMD)[2] were added, while 51 genes not in the Online Mendelian Inheritance of Man (OMIM)[3] were eliminated, resulting in an additional 1966 genes. Next, genes not already in the panel but present in GeneTests.org[4] (69 genes) were added. Finally, Illumina added 17 genes not already in the panel that were included in other TruSight Sequencing Panels, such as TruSight Cancer and TruSight Tumor.[5]

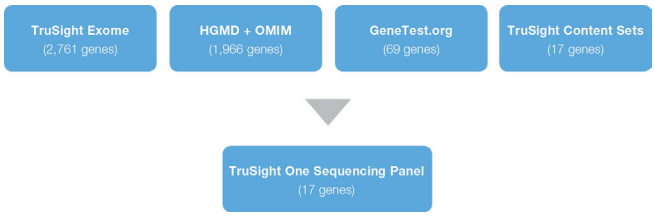

Figure 2: Design of the TruSight One Sequencing Panel

Identity of Target Regions Not Covered

Regions of the TruSight One Sequencing Panel shown in Tables 1 and 2 have exhibited very low depth of coverage through repeated sequencing runs (< 0.05× of the mean depth of coverage). Of the 62,000 regions covered by TruSight One, these 104 regions meet low performance thresholds. The 10 regions with entries in HGMD are highlighted in Table 2.

Table 1: TruSight One Targets with Low Depth of Coverage

Gene	Chromosome	Start	Stop
ABCC1	chr16	16043609	16043656
ACVR1B	chr12	52345528	52345618
ADAMTS2	chr5	178772190	178772329
AGO2	chr8	141645583	141645605
ANKRD11	chr16	89556652	89556969
ASMT	chrX	1711695	1711907
ASMT	chrX	1751596	1751680
BRSK2	chr11	1411515	1411605
BTBD9	chr6	38545376	38545495
C10orf11	chr10	78316967	78317043
C4BPA	chr1	207288761	207288860
C5orf42	chr5	37115062	37115151
C5orf42	chr5	37158326	37158447
C5orf42	chr5	37186388	37186496
C5orf42	chr5	37238959	37239062
CACNA2D3	chr3	54156741	54156862
CHD1L	chr1	146714354	146714480
CNKSR2	chrX	21508577	21508696
CNKSR2	chrX	21581356	21581570
COL4A6	chrX	107408600	107408716
CRLF1	chr19	18717351	18717466
CYB5R3	chr22	43045300	43045321
DCTD	chr4	183838441	183838466
DIAPH2	chrX	95940058	95940189
DLG3	chrX	69672471	69672605
EHMT1	chr9	140513480	140513501
EPHB2	chr1	23037475	23037536
ERBB2	chr17	37876040	37876087
ERMAP	chr1	43302829	43302861
ESR2	chr14	64716264	64716397
ETS1	chr11	128442957	128443025

Identity of Eliminated Regions

To verify the accuracy of base calls made using the TruSight One Sequencing Panel, base calls made from TruSight One libraries were compared to those of the "platinum genomes." Platinum genomes are a set of gold standard whole-genome sequencing data produced by Illumina6. While > 99.9% of the base calls were equivalent between the 2 libraries, there were a few exons from TruSight One libraries that harbored a high number of variants not seen in the platinum genomes data set. Upon further examination of these regions, Illumina found that many of them exhibited homologous sequences in other parts of the genome, either due to repeats or pseudogenes. These sequences could potentially cause nonspecific enrichment, misalignment, and incorrect variant calls. Though these exons do comprise gene content that is sequenced using the TruSight One Sequencing Panel, they have been removed from the TruSight One manifest. This means that they will not appear in variant call (.vcf) files, as the variant calls from them are suspect and thus would likely result in false positives. The identities of these exons are listed in Table 3.

Table 1: TruSight One Targets with Low Depth of Coverage

Gene	Chromosome	Start	Stop
EVI5	chr1	93257901	93257951
EXOC4	chr7	132937858	132937943
EXOC4	chr7	133002038	133002144
EXOC4	chr7	133314798	133314894
EXOC4	chr7	133682245	133682386
EZH2	chr7	148544274	148544390
FAAH2	chrX	57337026	57337162
FBXO18	chr10	5932309	5932309
FRMPD4	chrX	12157091	12157131
GABRD	chr1	1950863	1950930
GIT1	chr17	27916345	27916396
GNAQ	chr9	80409379	80409508
GRIK3	chr1	37267455	37267646
HS1BP3	chr2	20850792	20850823
HYDIN	chr16	70893971	70894109
HYDIN	chr16	71009039	71009172
IGF2R	chr6	160390279	160390427
JAG2	chr14	105634691	105634757
LLGL1	chr17	18128997	18129077
MAP2K4	chr17	11924204	11924318
MFI2	chr3	196730246	196730322
MIR2861	chr9	130548196	130548286
MIR502	chrX	49779205	49779291
MIR934	chrX	135633036	135633119
MUC1	chr1	155161502	155161530
NFIX	chr19	13106651	13106678
OPN4	chr10	88416952	88416984
OPN4	chr10	88417789	88417922
OPN4	chr10	88421038	88421145
ORMDL3	chr17	38078806	38078938
OSR1	chr2	19552039	19552171

Table 1: TruSight One Targets with Low Depth of Coverage

Gene	Chromosome	Start	Stop
OVGP1	chr1	111962232	111962348
P2RX4	chr12	121654758	121654805
P2RX4	chr12	121654937	121655084
P2RX4	chr12	121670414	121670479
P2RX7	chr12	121614950	121615015
PDPK1	chr16	2588114	2588137
PIGA	chrX	15353623	15353635
PTPRJ	chr11	48002465	48002560
RAC1	chr7	6414367	6414401
REPS2	chrX	16964985	16965257
REPS2	chrX	17065470	17065605
REPS2	chrX	17070505	17070568
REPS2	chrX	17095394	17095530
REPS2	chrX	17165536	17165601
RPA1	chr17	1733388	1733420
SCN1B	chr19	35521724	35521764
SDC3	chr1	31381296	31381433
SETD8	chr12	123873980	123874101
SH3BP2	chr4	2819950	2820117
SNTG2	chr2	946683	946754
STK39	chr2	169103738	169103945
TCF7	chr5	133450598	133450846
TGFBR1	chr9	101867487	101867584
TMEM8A	chr16	431701	431821
TRIO	chr5	14143835	14143991
TSEN54	chr17	73512641	73512697
TTTY13	chrY	23745485	23745548
ZNF350	chr19	52471831	52471926
ZNF365	chr10	64219500	64219556
ZNF385B	chr2	180309603	180309704
ZNRF1	chr16	75127470	75127565
ZNRF1	chr16	75140380	75140434

Conclusions

The TruSight One Sequencing Panel achieves high depth of coverage for the coding regions of > 4800 genes that are associated with clinically relevant phenotypes. The panel was designed to improve coverage of underrepresented regions while decreasing coverage of overrepresented regions. This unique design provides the highest uniformity ever achieved in a probe-based panel: over 97% of the regions are covered at > 0.2× of the mean. The TruSight One Sequencing Panel will provide an essential tool for clinical research into the genetic basis of various human diseases.

Learn More

To learn more about sequencing panels for clinical research, visit www.illumina.com/trusightone.

References

1. TruSight Exome Content Set Product Files. Illumina. support.illumina.com/downloads/trusight_exome_product_files.html. Accessed December 17, 2015.
2. The Human Gene Mutation Database (HGMD). http://www.hgmd.cf.ac.uk/ac/index.php. Accessed December 17, 2015.
3. OMIM - Online Mendelian Inheritance in Man. www.omim.org. Updated December 16, 2015. Accessed December 17, 2015.
4. GeneTests. www.genetests.org. Accessed December 17, 2015.
5. TruSight One Sequencing Panel | Most commonly ordered molecular assays. Illumina. www.illumina.com/products/trusight-one-sequencing-panel.html. Accessed December 17, 2015.

Table 2: TruSight One Targets with Low Depth of Coverage Referenced in HGMD

Gene	Chromosome	Start	Stop	HGMD Ref	Variant Type	Association
COX14	chr12	50513826	50514000	CM120176	SNV	Lactic acidosis, fatal neonatal
EVC	chr4	5713107	5713281	CM098918	SNV	Ellis-van Creveld syndrome
LRP5	chr11	68080182	68080273	CM014722	SNV	Osteoporosis-pseudoglioma syndrome
MATN3	chr2	20212169	20212392	CM052250	SNV	Multiple epiphyseal dysplasia
P2RX4	chr12	121670217	121670310	CM1111111	SNV	Increased pulse pressure
P2RX7	chr12	121592588	121592756	CM103405	SNV	Altered function
P2RX7	chr12	121613191	121613281	CM043948	SNV	Altered function
PKD1	chr16	2185475	2185690	CD076869	Deletion	Polycystic kidney disease
SEPN1	chr1	26126721	26126904	CM022835	SNV	Multiminicore disease
ZPBP2	chr17	38028523	38028741	CM1111300	SNV	Asthma

Table 3: Regions Eliminated from TruSight One

Gene	Chromosome	Start	Stop
ATXN8OS	chr13	70713372	70713885
CSGALNACT1	chr8	19261671	19263580
CYP2D6	chr22	42523448	42523636
CYP2D6	chr22	42525739	42525911
DSPP	chr4	88534936	88537720
FANCD2	chr3	10089600	10089735
FBXL6	chr8	145579960	145580191
FLG	chr1	152275175	152287223
FOXD4	chr9	116799	118119
FRG1	chr4	190876192	190876306
HBG1	chr11	5269588	5269717
HLA-A	chr6	29911045	29911320
HLA-B	chr6	31323943	31324219
HLA-B	chr6	31324464	31324734
HLA-C	chr6	31238850	31239125
HLA-C	chr6	31239376	31239645
HLA-C	chr6	31237987	31238262
HLA-DQA1	chr6	32610386	32610541
HLA-DQA1	chr6	32609748	32610030
HLA-DQB1	chr6	32629744	32630025
HLA-DQB1	chr6	32634276	32634384
HLA-DRB1	chr6	32549334	32549615
HLA-DRB1	chr6	32557420	32557519
HLA-DRB5	chr6	32497902	32498001
HS6ST1	chr2	129075610	129076137
HYDIN	chr16	70954494	70955120
HYDIN	chr16	70902473	70902691
HYDIN	chr16	71163542	71163726
IGHG2	chr14	106109389	106109812
KCNJ18	chr17	21318654	21319956
KIR2DL3	chr19	55255243	55255536
KRT6B	chr12	52845322	52845862
KRT6C	chr12	52866981	52867521
KRT81	chr12	52681379	52681505
MAP2K3	chr17	21204186	21204305

Table 3: Regions Eliminated from TruSight One

Gene	Chromosome	Start	Stop
MAP2K3	chr17	21215454	21215593
MAP2K3	chr17	21203857	21203970
MAP2K3	chr17	21202190	21202238
MAP2K3	chr17	21217459	21217539
MASP1	chr3	186951869	186954355
MLL3	chr7	151944987	151945705
MUC2	chr11	1092080	1093668
MUC3A	chr7	100551549	100552774
MUC3A	chr7	100608728	100608891
MUC3A	chr7	100552880	100553066
MUC3A	chr7	100607745	100607894
MUC4	chr3	195505661	195518368
MUC5B	chr11	1262081	1272973
MUC6	chr11	1015762	1018770
NBPF1	chr1	16890441	16890681
NT5C3	chr7	33053741	33054443
PHF2	chr9	96438876	96439245
PLIN4	chr19	4510458	4513716
PRB1	chr12	11506040	11506936
PRB3	chr12	11420458	11421082
PRB4	chr12	11461172	11461816
PRDM2	chr1	14104913	14109326
PRG4	chr1	186275449	186278272
PRSS2	chr7	142479907	142481375
PRSS3P2	chr7	142479908	142480068
PTCSC3	chr14	36604916	36605563
RP1L1	chr8	10464404	10470856
SEC63	chr6	108243000	108243113
TAS2R31	chr12	11183007	11183934
TBP	chr6	170870879	170871321
TNXB	chr6	31977307	31977404
TPTE	chr21	10910307	10910399
UNC93A	chr6	167728675	167728937
ZAN	chr7	100349335	100350852

Illumina • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

illumina®