illumina®

既知の臨床的な表現型に関連する4,813遺伝子のシーケンスパネル

TruSight™ Oneシーケンスパネルで高いカバレッジ深度を提供し、精度の高い変異コールを可能に

はじめに

臨床研究では特定の疾病を研究する際、しばしば数種類の分子プロファイリングアッセイが行われます。それぞれのアッセイは単一遺伝子をターゲットとしており、得られる情報には限りがあります。また複数にわたるアッセイを行う必要があるので、そのコストや時間も見過ごせません。近年登場した次世代シーケンサーは、臨床ゲノム研究において、より速く、そしてコスト効果が高い手法を提供します。

イルミナのTruSight Oneシーケンスパネルは、臨床的な表現型に関連する4,813遺伝子のコード領域のゲノム解析を行うためにデザインされています。このパネルでは、MiSeq®もしくはHiSeq®システムを用いることを想定しており、複数のサンプルに対して高いカバレッジ深度（20x以上）を達成できます。またパネルに含まれる全遺伝子を解析するか、興味のある特定の疾病に関与する遺伝子サブセットのみにフォーカスを絞るかの選択も可能です。またTruSight OneシーケンスパネルとVariantStudioソフトウェアを用いて、ゲノム変異の解析や分類、レポート作成などを行えます。

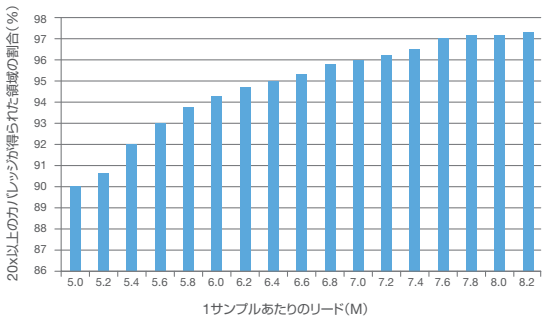
このテクニカルノートでは、パネルのパフォーマンスや、遺伝子コンテンツの選定方法、そしてパネルに含まれていないターゲットエクソンや不正確な変異コールをもたらす可能性があるため除外されたターゲットエクソンなどの詳細をまとめました。

TruSight Oneシーケンスパネルのパフォーマンス

プローブキャプチャー法は、数多いゲノムターゲットの濃縮やシーケンスにおいて非常に優れた手法です。しかし、高い水準のカバレッジ均一性を実現できないこともあります。それは、それぞれのプローブが持つターゲット領域への親和性が一定でないため、領域によってカバレッジが高くなりすぎることや、逆に低すぎて変異をコールできないことがあるためです。

カバレッジ均一性の改善に向け、イルミナではTruSight Oneシーケンスパネルをデザインする際に2つの手法を導入しました。まずパネル全体と比較してキャプチャーされにくい領域には、そのカバレッジを増加させるため、プローブを追加しました。この追加プローブは最初に作製したプローブセットに追加し、そのパフォーマンスの検証を行いました。このプロセスを3回繰り返すことで、この手法によってキャプチャーされるべき全領域が、十分にカバーされることを確認しています。同様に、シーケンスの読み取りが過剰にでてしまう領域（平均カバレッジの2倍を超える領域）を特定し、そのプローブに対して相補的なビオチン非標識オリゴヌクレオチドをプローブプールに添加しました。このビオチン非標識プローブは、センス側のプローブの効率を著しく下げるため、この領域が過剰にキャプチャーされにくくなり、他の領域のシーケンスカバレッジが改良されました（平均で1領域あたり約5xの増加が見られました）。例えば、従来15xのカバレッジを示した領域は、20xのカバレッジを示すようになります。その結果、全体として、キャプチャーされにくかった領域のカバレッジを増加させ、シーケンスの読み取りが過剰にでてしまう領域のカバレッジを減少させることで、TruSight Oneシーケンスパネルの均一性が向上し、97%超の領域で平均の0.2x以上のカバレッジを実現します。これはプローブベースの濃縮パネルでは最も高い均一性です。

図1：リード数vsカバレッジ深度の高い領域の割合（%）



TruSight Oneシーケンスパネルにおいて、1サンプルあたりのリード数と、20x以上のカバレッジが得られた領域の割合を比較しています。

図1はTruSight Oneシーケンスパネルで20x以上のカバレッジ深度が得られた領域の割合（%）と、サンプルあたりのリード数との比較を示しています。リード数が低いにも関わらずカバレッジ深度の高い領域の割合が高かったことから、このパネルの高い均一性が示されています。例えば、サンプルあたり6.4Mリードのランの場合、95%の領域が20x以上のカバレッジを示しています。したがって、MiSeqシステムを用いこの高いカバレッジで3サンプルを解析する場合（新しいv3のケミストリーとソフェアを使用）、1ランの総リード数の77%で達成することが可能です。例えば、子供が稀な遺伝子病を患っている場合、3サンプルをトリオ（病気を患う子供とその父母）として扱うことができます。そして VariantStudio 2.0ソフトウェアのトリオの差分法（Subtraction Techniques）と組み合わせることで、原因となる変異を同定する可能性が大幅に高まります。

表1：TruSight Oneシーケンスパネルでカバレッジ深度が低かったターゲット

遺伝子	染色体	スタート	ストップ	HGMD Ref	変異タイプ	関連性
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			
COX14	chr12	50513826	50514000	CM120176	SNV	Lactic acidosis, fatal neonatal
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			
EVC	chr4	5713107	5713281	CM098918	SNV	Ellis-van Creveld syndrome
EVI5	chr1	93257901	93257951			
EXOC4	chr7	132937858	132937943			
EXOC4	chr7	133002038	133002144			
EXOC4	chr7	133314798	133314894			
EXOC4	chr7	133682245	133682386			

遺伝子	染色体	スタート	ストップ	HGMD Ref	変異タイプ	関連性
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			
LRP5	chr11	68080182	68080273	CM014722	SNV	Osteoporosis-pseudoglioma syndrome
MAP2K4	chr17	11924204	11924318			
MATN3	chr2	20212169	20212392	CM052250	SNV	Multiple epiphyseal dysplasia
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			
OVGP1	chr1	111962232	111962348			
P2RX4	chr12	121654758	121654805			
P2RX4	chr12	121654937	121655084			
P2RX4	chr12	121670217	121670310	CM1111111	SNV	Increased pulse pressure
P2RX4	chr12	121670414	121670479			
P2RX7	chr12	121592588	121592756	CM103405	SNV	Altered function
P2RX7	chr12	121613191	121613281	CM043948	SNV	Altered function
P2RX7	chr12	121614950	121615015			
PDPK1	chr16	2588114	2588137			
PIGA	chrX	15353623	15353635			

遺伝子	染色体	スタート	ストップ	HGMD Ref	変異タイプ	関連性
PKD1	chr16	2185475	2185690	CD076869	Deletion	Polycystic kidney disease
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			
SEPN1	chr1	26126721	26126904	CM022835	SNV	Multiminicore disease
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			
ZPBP2	chr17	38028523	38028741	CM1111300	SNV	Asthma

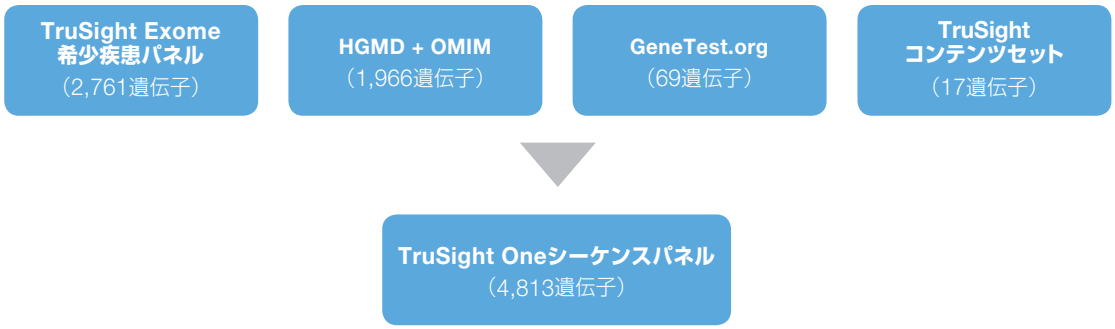
ターゲットとする遺伝子

TruSight Oneシーケンスパネルをデザインするにあたり、イルミナでは既知の臨床的な表現型と関連するすべての遺伝子を網羅することを目標としました。これによって、臨床研究を行う際、多数の小さいシーケンスパネルのベースとしてこのTruSight Oneを使用することができ、包括的なシーケンスアッセイのポートフォーリオとして利用できるようになります。初めに、TruSight Exome希少疾患パネル[1]（2,761遺伝子、図2）に含まれるすべての遺伝子を、このパネルにも含めました。またTruSight Exome希少疾患パネルにはなく、ヒト遺伝子変異データベース（HGMD）[2]に登録されているすべての遺伝子を追加した一方、オンライン・ヒトメンデル遺伝形質（OMIM）[3]に登録されていない51の遺伝子を除外しました。その結果1,966遺伝子が追加されました。次に、このパネルにはなく、GeneTests.org[4]に登録されている遺伝子（69遺伝子）を追加しました。最後に、このパネルにはなく、TruSight癌パネルやTruSight腫瘍パネル[5]など、他のTruSightシーケンスパネルに含まれている17の遺伝子を追加しました。

カバレッジの低いターゲット領域の詳細

表1は、TruSight Oneシーケンスパネルにおいて、シーケンスを繰り返してもカバレッジ深度が非常に低い領域を示します（平均カバレッジ深度の0.05x未満）。またTruSight Oneシーケンスパネルでカバーされる6万2,000領域のうち、104領域が低パフォーマンスの閾値に該当しました。その中で、HGMDに登録されている10領域については、表1でハイライト表示しています。

図2：TruSight Oneシーケンスパネルのデザイン



除外された領域の詳細

TruSight Oneシーケンスパネルを使用した際のベースコールの精度を検証するために、TruSight Oneシーケンスパネルのライブラリーによるベースコールを、「プラチナゲノム」でのベースコールと比較しました。プラチナゲノムとは、イルミナが独自に作成した、ゴールドスタンダードの全ゲノムシーケンスデータセットのことです[6]。ベースコールの99.9%以上は、2セットのライブラリー間で同等でした。一方、TruSight Oneシーケンスパネルのライブラリーの数個のエクソンには、プラチナゲノムデータセットでは見られない変異が数多く含まれていました。これらの領域についてさらに調査を進めたところ、その多くが、反復配列や偽遺伝子など、ゲノム上の他の部位にある相同配列であることがわかりました。これらの配列は非特異的な濃縮やミスアラインメント、さらには誤った変異コールにつながる恐れがあります。これらのエクソンは、TruSight Oneシーケンスパネルを用いてシーケンスされた遺伝子コンテンツであることに間違いありませんが、TruSightシーケンスパネルからは除外されました。実際、これらのエクソンからの変異コールは信頼性に欠け、偽陽性を生じる可能性が高いため、これらのエクソンが変異コール（.vcf）ファイルに現れることはありません。表2に除外されたエクソンを示します。

まとめ

TruSight Oneシーケンスパネルを使用することで、臨床的な表現型に関与する4,800を超える遺伝子のコード領域を高いカバレッジ深度でシーケンスできるようになりました。このパネルは、キャプチャーされにく領域のカバレッジを増加させ、過剰にシーケンスの読み取りがでてしまう領域のカバレッジを減少させることで、全体のカバレッジを改善するようデザインされています。この独自のデザインにより、プローブベースパネルでは最高の均一性を達成し、97%以上の領域で平均の0.2x以上のカバレッジを実現します。TruSight Oneシーケンスパネルは、臨床研究におけるさまざまなヒト疾患の遺伝子ベースとして、なくてはならないツールとなります。

詳しい情報

臨床研究のためのシーケンスパネルの詳細については、www.illuminakk.co.jp/clinical/translational_genomics/product.ilmn をご覧ください。

参考文献

1. TruSight Exome (www.illumina.com/clinical/translational_genomics/product/trusight_exome.ilmn)
2. Human Genome Mutation Database (www.hgmd.org)
3. Online Mendelian Inheritance in Man (www.omim.org)
4. GeneTests (www.genetests.org)
5. TruSight Products (www.illumina.com/clinical/translational_genomics/product.ilmn)
6. TruSight One (www.illumina.com/trusightone)

表2：TruSight Oneシーケンスパネルから除外された領域

遺伝子	染色体	スタート	ストップ
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

遺伝子	染色体	スタート	ストップ
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

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATT
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAA
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCGATTAAGAG
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAAT
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATT
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAAC
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAA

イルミナ株式会社

〒108-0014
東京都港区芝5-36-7　三田ベルジュビル22階
Tel（03）4578-2800　Fax（03）4578-2810
www.illuminakk.co.jp

代理店

本製品の使用目的は研究に限定されます。

© 2014 Illumina, Inc. All rights reserved.

Illumina, illuminaDx, BaseSpace, BeadArray, BeadXpress, cBot, CSPro, DASL, DesignStudio, Eco, GAIIx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, NuPCR, SeqMonitor, Solexa, TruSeq, TruSight, VeraCode, the pumpkin orange color, the Genetic Energy streaming bases design は Illumina, Inc の商標または登録商標です。
その他の会社名や商品名は、各社の商標または登録商標です。予告なしに仕様を変更する場合があります。

Pub. No. 0676-2013-J016 01SEP2014