

エクソームシーケンスにおける 最適 DNA インプット量の検証

はじめに

希少疾患において、エクソームシーケンスは臨床表現型の原因となる遺伝的な異常を同定する効果的な手法として幅広く利用されている。浜松医科大学 医化学講座 才津浩智教授ご協力のもと、イルミナ改良版エクソームキット TruSeq Exome Library Prep Kit を用いて、ヒト臨床検体サンプルからのエクソーム解析を行った。ゲノム DNA インプット量について、プロトコル推奨量に加え複数条件で検討を行い、解析結果への影響を検証した。

実験方法

病変部位あるいは血液から取得したサンプルを 4 検体分用意し、ゲノム DNA を Proteinase K-SDS 溶解による標準法および QuickGene-610L を用いて抽出し、Qubit Fluorometer で濃度を測定した。インプット DNA 量を 100ng、1,000ng、3,000ng に振り、TruSeq Exome Library Prep Kit (カタログ番号 FC-150-1001 および FC-150-1003) 推奨プロトコル (Document #15059911 v01) に沿ってプレプール法でエクソームライブラリーを作製した。ターゲット濃縮・キャプチャー時にプールを 2 つ準備した。プール A は 3 検体に前述の 3 条件のインプット DNA 量を準備し、合計 9 種類のライブラリーをプールした。プール B は 1 検体に前述の 3 条件のインプット DNA 量を準備し、合計 3 種類のライブラリーをプールした。プール A は NextSeq 500/550 High Output Kit v2 (150 Cycles) (カタログ番号 FC-404-2002)、プール B は NextSeq 500/550 Mid Output Kit v2 (150 Cycles) (カタログ番号 FC-404-2001) を用いて NextSeq 500 システムで 75bp × 2 の条件でシーケンスを実施し、NCS v2.1.2 で bcl ファイル、bcl2fastq v2.18 で FASTQ を取得した。その後の解析は、bwa-mem によるアライメント、Picard による重複リードの除去、GATK によるバリエーションの検出、AnnoVar によるアノテーション付与、アレール頻度によるフィルタリングを行い、候補バリエーションを抽出した (図 1)。

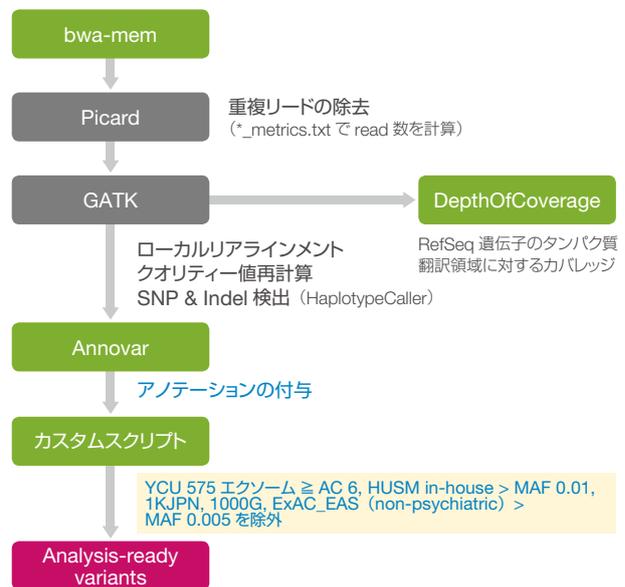


図 1: 変異検出データ解析フロー

結果

NextSeq 500 システムによるシーケンス結果は、プール A およびプール B いずれの場合でも、パスフィルター 92% 以上、Q30 も 92% 以上と高品質な結果が得られた。

インプット DNA 量の差によらず検出されたリード数はほぼ等量で、100ng の条件でも高品質な結果が得られた。一方で、デュプリケート率は 100ng においても 10% 以下と良好な結果であったが、インプット量を増やすことで改善が見られた (図 2)。

(a)

	インプット量	Reads#	Aligned	%Aligned	%Duplication
プール A-1	100 ng	109,488,588	109,125,425	99.7%	8.54
	1000 ng	107,439,312	107,066,508	99.7%	6.00
	3000 ng	96,190,008	95,826,822	98.2%	5.72
プール A-2	100 ng	158,818,162	158,320,274	99.7%	6.72
	1000 ng	110,442,592	109,974,497	99.6%	6.57
	3000 ng	113,436,260	112,952,866	99.6%	5.85
プール A-3	100 ng	99,223,884	98,900,361	99.7%	8.05
	1000 ng	101,851,406	101,464,632	99.6%	5.96
	3000 ng	106,849,054	106,390,601	99.6%	5.87
プール B-1	100 ng	92,479,326	92,120,037	99.6%	6.25
	1000 ng	128,319,990	127,777,305	99.6%	4.45
	3000 ng	114,245,394	113,612,083	99.4%	4.22

(b)

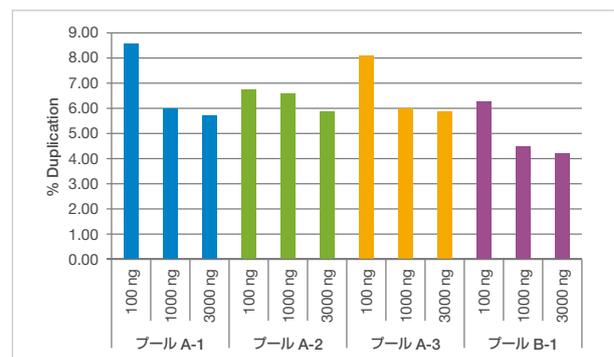


図 2: エクソーム解析結果 プール A-1~3 の 9 検体、プール B-1 の 3 検体はそれぞれまとめて濃縮反応を行った。

(a) インプット量に関わらず得られたリード数に差は見られなかった。(b) デュプリケート率は、インプット量を増やすことによって改善が見られた。

20x カバレッジを得た領域の対象領域（コーディングエクソン 33.5Mb）に対する割合について比較すると、いずれのプール・サンプルにおいても 95% 以上を達成した。インプット量を 3,000ng に増やしても結果に差がみられなかったため、インプット量は 100ng~1,000ng の間で十分と考えられる（図 3）。

最後に、フィルタリング後に検出された稀な変異の数について各条件で比較を行った。各条件で高い一致率が認められた（図 4）。

まとめ

TruSeq Exome Library Prep を用いてわずか 100ng という少量のインプット量でも高品質ライブラリーを作製できた。さらには、100ng インプット量においても検出された変異に差は見られず、100ng インプット量での性能に問題はないと考えられる。インプット量を 1,000ng に増やすことによって、デュプリケート率を減らせることが示された。

(a)

	インプット量	データ出力 (Gb)	平均深度	20x 以上のカバレッジ率
プール A-1	100 ng	8.2	104.9	96.4
	1000 ng	8.1	103.2	96.7
	3000 ng	7.2	90.4	95.6
プール A-2	100 ng	11.9	149.9	98.1
	1000 ng	8.3	104.6	97.1
	3000 ng	8.5	106.7	97.0
プール A-3	100 ng	7.4	95.6	95.8
	1000 ng	7.6	99.9	96.5
	3000 ng	8.0	100.9	97.0
プール B-1	100 ng	6.9	84.1	94.0
	1000 ng	9.6	115.9	96.8
	3000 ng	8.6	106.3	96.1

(b)

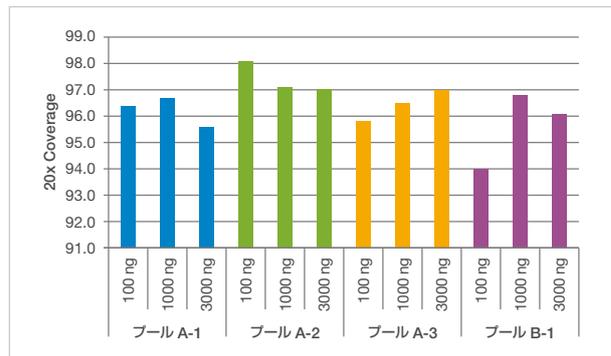


図 3: 各インプット量に対するデータ出力とカバレッジ

(a) いずれのインプット量においてもデータ出力量、平均深度に差は見られなかった。20x カバレッジは対象領域の 95% 以上と製品仕様 (90%) を超える結果が得られた。(b) 20x カバレッジは、インプット量 1,000ng でプラトーに達している。

	インプット量	vs 100ng	vs 1000ng	vs 3000ng
プール A-1	100 ng		470 (94.4%)	468 (94.0%)
	1000 ng	469 (93.0%)		474 (94.0%)
	3000 ng	468 (92.4%)	475 (93.9%)	
プール A-2	100 ng		405 (89.4%)	412 (90.9%)
	1000 ng	406 (91.8%)		412 (93.2%)
	3000 ng	412 (90.7%)	413 (91.0%)	
プール A-3	100 ng		371 (89.0%)	390 (93.5%)
	1000 ng	376 (90.2%)		378 (90.6%)
	3000 ng	391 (90.5%)	377 (87.3%)	
プール B-1	100 ng		424 (93.0%)	418 (91.7%)
	1000 ng	429 (90.7%)		436 (92.2%)
	3000 ng	421 (91.1%)	434 (94.0%)	

図 4: 各インプット量におけるコールされた変異数の一致率比較

お客様の声

プレプールによるばらつきがあるのではないかと懸念していましたが、ばらつきはほとんど見られませんでした。多くのサンプルを処理する必要がある場合には、プレプールによるコストメリットがあると思います。異なるインプット量でも再現性の高い結果が得られており、100ng という少量からでも十分高品質な結果が得られました。Biopsy サンプルなどゲノム量が限られるようなサンプルの場合には少量インプットからスタートできることは大きなメリットだと感じます。1,000ng あればより高カバレッジな結果が得られると思います。20x カバレッジについても 95% 前後と良好でした。NextSeq は、短時間で全エクソーム解析が出来、小回りの利く使い勝手のよいシーケンサーだと感じています。

謝辞：製品性能検証にあたり浜松医科大学 医化学講座 才津浩智教授に多大なるご協力を賜りました。ここにあらためて感謝の意を表します。

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
jp.illumina.com

www.facebook.com/illuminakk

代理店

本製品の使用目的は研究に限定されます。 販売条件：jp.illumina.com/tc

Pub. No. 5015-170509-02

© 2017 Illumina, Inc. All rights reserved.

Illumina, BaseSpace, BeadArray, BeadXpress, cBot, CSeq, DASL, Design Studio, GALx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, NextSeq, NovaSeq, 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。予告なしに仕様および希望販売価格を変更する場合があります。