

データ提供機関	ナショナルセンター・バイオバンクネットワーク (NCBN)
対象	健常者（難病やがんを罹患していない生活習慣病を持つ人を含む）(ICD10：Z006)：9,830名
Platform	Illumina [NovaSeq 6000]
ソース	各バイオバンクから選定した健常者の末梢血（または唾液）から抽出したDNA
ライブラリ作製方法（キット名）	TruSeq DNA PCR-Free HT Library Prep Kit
断片化の方法	超音波断片化
ライブラリ構築方法	Paired-end
リード長（除：バーコード、アダプタ、プライマー、リンカー）	150bp
インサートサイズ	550bp
クオリティコントロール方法	以下の条件で全ゲノムシーケンス解析を実施 - ライブラリサイズが400bp-750bpであることを確認 - QV30以上の塩基の割合が75%以上 - FASTQCによる重複リード除去後の総塩基数が900億塩基以上 アライメントとバリエーションコール後に以下のサンプルを解析から除外した - Depthおよびマップ率が異常値を示すサンプル - 性染色体のDepthが臨床情報の性別と矛盾するサンプル - KINGプログラムで2親等以内と判定されたサンプルのいずれか バリエーションコールの結果は以下のフィルタリングを実行した - GQ<20またはDP<11またはDP>64またはヘテロ接合体コールにおいてマイナーアレルの割合が25%未満のジェノタイプはno callに設定 - VQSRの結果をVCFのFILTERフィールドにセット - コール率95%未満のバリエーションにはFILTERにLowCRをセット - Hardy-Weinberg平衡検定のP値が10-6未満のバリエーションにはFILTERにHWEをセット
重複するリードの除去方法	MarkDuplicates (GATK4.1.0) 互換アルゴリズム (Parabricks 3.1.0 fq2bam)
マッピング方法	bwa mem (v0.7.15) 互換アルゴリズム (Parabricks 3.1.0 fq2bam)
マッピングの際のリファレンス配列	GRCh38 (+HLA+decoy)
平均カバー率 (Depth)	34.0 (常染色体)
変異検出方法	HaplotypeCaller (GATK4.1.0) 互換アルゴリズム (Parabricks 3.1.0 haplotypcaller)
総データ量	869TiB (FASTQ、CRAM、gVCF)
利用にあたっての制限事項	