データ提供機関	東北メディカル・メガバンク計画 (TMM)
対象	一般住民
Platform	Illumina [HiSeq2500, NovaSeq6000]
ソース	末梢血、または唾液から抽出された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)	HiSeq2500：162bp, 259bp NovaSeq6000：150bp
インサートサイズ	HiSeq2500：550bp NovaSeq6000：350bp
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クエンス解析を実施 - 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 サンプルごとのマッピング・バリアントコールを行った。その後の解析 (Joint Call) を行う前に以下のサンプルを除外した - マップ率が98%未満 - 常染色体の平均Depth が20x (HiSeq2500 162PE, NovaSeq6000 150PE) , 18x (HiSeq2500 259PE) 未満 - verifyBamID2プログラムによって推定されたコンタミ率が3%以上 - ゲノムから推定される性別が、調査票上の性別と異なるサンプル Joint Call結果はVQSRアルゴリズムによるフィルタリングを実施し、その結果をVCFファイルのFILTERフィールドにセットした
重複するリードの除去方法	MarkDuplicates (GATK 4.1.0.0)
マッピング方法	BWA mem (v0.7.15) または BWA-mem2 (v2.2)
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)	30 (常染色体)
変異検出方法	HaplotypeCaller (GATK 4.1.0.0)
総データ量	約730TiB (FASTQ, CRAM, gVCF)
利用にあたっての制限事項	