

静岡多目的コホート研究事業ヒト全ゲノムシーケンス解析業務委託仕様書（案）

公立大学法人静岡社会健康医学大学院大学（以下「甲」という。）は、〇〇〇〇〇〇（以下「乙」という。）に対して、以下のとおり、静岡多目的コホート研究事業ヒト全ゲノムシーケンス解析業務委託する。

1 目的

個人の体質に応じた最適な疾病予防方法の検討や疾病リスク因子を明らかにするため、静岡多目的コホート研究事業で収集したヒト末梢血由来 DNA を用いた全ゲノムシーケンスを行う。

2 業務内容と条件

2.1 乙は、甲が提供するヒト末梢血由来 DNA（依頼件数 1,500 検体～2,500 検体）を用いて、二色蛍光検出の次世代シーケンサーNovaSeq6000 または NovaSeq X システムを用いた全ゲノムシーケンス解析を実施し、その結果を甲に報告するものとする。

2.2 責任者の選定

2.2.1 本契約締結後速やかに、甲乙双方から、本件業務の業務責任者を選定し、その相手方に報告しなければならない。当該責任者を変更するときも同様とする。

2.2.2 甲乙双方の協議、甲からの指示等については、全て前項の業務責任者を通じて行うものとする。

2.3 残余検体の扱い

2.3.1 乙は、本件業務の完了後、検体のうち残余検体については、甲に返還する。

2.3.2 前項の返還費用は甲が負担する。

2.4 資料等の提供および返還

2.4.1 乙は、甲から提供された資料等を善良な管理者の注意義務をもって使用および保管するものとする。また、乙は、本件業務以外の目的に使用してはならない。

2.4.2 乙は、本件業務が完了したとき、本契約が解除等により終了したときまたは甲が要求したときは、甲より受領した資料等を、甲の指示に従い、速やかに甲に返還する。

3 提供物・提供時期

3.1 検体は、契約締結後速やかにすべての検体を提供する。

3.2 検体はチューブもしくは 96 ウェルプレートに整列化した DNA 溶液とする。96 ウェルプレートでの提供を希望する場合、送付用の 96 ウェルプレートおよびキャップは乙から提供すること。

3.3 提供検体は、検体あたり 1.65 µg 以上 (30ng/µL) とする。これを下回る場合の対応については、都度協議すること。

3.4 委託者からは、以下の情報を Excel ファイルまたはテキストファイルにまとめ、メール添付または記録メディアにて提供する。

プレート提供時

96 ウェルプレート ID、ウェル位置、検体 ID、検体濃度 (ng/µL) (吸光法又は蛍光法によ

る DNA 定量結果)、液量 (μL)

チューブ提供時

ラック ID、ラック位置、チューブ ID、検体 ID、検体濃度 (ng/μL) (吸光法又は蛍光法による DNA 定量結果)、液量 (μL)

4 技術要件 (ゲノムシーケンス)

- 4.1 DNA サンプルの品質を Qubit (サーモフィッシャーサイエンティフィック社) あるいは同等の方法による濃度測定により確認すること。濃度が提供基準値に満たない場合は、協議の上、対応方法を相談すること。
- 4.2 ゲノム DNA を数百 bp に断片化して、96well プレート単位で処理できるラボラトリーオートメーション装置 (自動分注器) を用いて 550 bp インサートの DNA ライブラリ作製をすること。DNA 量が条件に満たない場合は、事前に協議の上、350bp プロトコルでの実施を検討すること。
- 4.3 DNA Shearing システム (Covaris 社) を用いた超音波法で DNA を断片化すること。
- 4.4 ライブラリ作製には TruSeq DNA PCR-Free HT Library Prep Kit (イルミナ社) を用いて Index Hopping が生じない対策を施していること。
- 4.5 解析において、検体間の品質の差を最小限にするために自動分注装置を用いる等の対策を講じること。
- 4.6 DNA ライブラリの品質検定はイルミナ社シーケンサーMiSeq system または iSeq system を用いて取得したシーケンスリードを用いてインサートサイズ、および DNA 濃度を測定し、DNA 濃度に基づき検体間のデータ出力量を補正すること。ライブラリのインサートサイズは、550bp プロトコルの場合は 400~750 bp であることを確認すること。350bp プロトコルの場合は 200~500bp であることを確認すること。ライブラリの DNA 濃度は 2.1 nM 以上あることを確認すること。この基準を満たさない場合は、残余検体を用いて再作製すること。再作製が難しい場合は、甲からの追加検体の提供等の対応方法を協議すること。
- 4.7 作製したライブラリは、イルミナ社シーケンサーNovaSeq6000 または NovaSeq X Plus を用いて、150base 長ペアエンドシーケンスを行い、シーケンサー付属のソフトウェアにより塩基配列 (リード配列) を得ること。
- 4.8 1 サンプルあたり複数レーンでシーケンスを行ってデータを取得すること。
- 4.9 得られたシーケンスデータは、レーンごとに精度情報 (QV30) を算出し、QV30 以上の塩基割合が 75%以上であることを確認すること。75%未満の場合は、協議の上、そのシーケンスデータをすべて廃棄する、または一部利用することを決定すること。
- 4.10 各検体のシーケンス量は fastp プログラムで算出された Duplicated read を除いて検体当たりのデータ量 (Gbase)、重複リード (Duplicated Read) の割合を算出すること。検体当たり、重複リードを除き 900 億塩基 (90G 塩基) 以上のデータを取得すること。

5 技術要件 (バイオインフォマティクス解析)

5.1 FASTQ データの作成

シーケンスデータは検体ごとに、ペアドエンドの (2 ファイル 1 組の) fastq ファイル (gz 圧

縮)を適切な命名規則で作成すること。

5.2 各多型のジェノタイピング (SNP/InsDel・CNV・SV・HLA)

全ての検体の FASTQ データを対象に、Illumina DRAGEN Bio-IT Platform を用いて以下の解析を行い、適切な命名規則でファイルを作成すること。この際、参照配列は GRCh38 を用いること。本委託業務終了後に新たに全ゲノムシーケンスを委託した場合、本委託業務で得られた全ての検体の FASTQ データを含めて以下の解析を行うこと。

- ・ SNP/InDel 解析 (cram/gVCF)
- ・ CNV 解析 (tsv)
- ・ SV 解析 (vcf)
- ・ HLA 解析 (json)

参考)

```
> dragen --mode mapping ¥
--ref-dir /path/to/ref-dir ¥
--fastq-list fastqlist.csv ¥
--output-directory /path/to/output_dir ¥
--output-file-prefix sample ¥
--output-format cram ¥
--enable-duplicate-marking true ¥
--enable-variant-caller true ¥
--generate-gvcf true ¥
--cnv-enable true ¥
--enable-sv true ¥
--enable-hla true
```

5.3 SNP/InDel の Joint Calling とアノテーション

全ての gVCF ファイルと用いて Iterative gVCF Genotyper analysis (Joint Calling --mode joint-genotyping) を実施し、VCF ファイルを作成すること。得られたすべての多型に対し、Variant Effect Predictor を用いてアノテーションを付与した VCF ファイルを別途作成すること。本委託業務終了後に新たに全ゲノムシーケンスを委託した場合、本委託業務の全 gVCF ファイルと、新規委託業務の gVCF ファイルを統合し、同様の解析を行うこと。

5.4 納品とファイル保存期間

4.1～4.3 において下線で指定されたファイルは、指定するサンプルコードを含む適切な命名規則に従ったファイル名にて、メタデータ (解析コマンド、ログ等) とともに納品すること。SNP/InDel の gVCF ファイル (*.g.vcf.gz) は、Joint Calling で再利用される。そのため、委託者からの削除指示があるまで、乙における保存期間とすること。これ以外のファイルは委託元への納品完了から 6 ヶ月を乙における保存期間とすること。

6 技術要件 (その他)

6.1 ライブラリ作製、シーケンスについて異常が生じた場合は作業進行、再作業について委託者と

協議すること。

- 6.2 検体の取違を防止するため、検体管理システムを用いた検体・データ管理システムに基づいて業務を実施できること。
- 6.3 ラボラトリーオートメーション装置（自動分注器）、イルミナ社シーケンサーNovaSeq6000 または NovaSeq X Plus を国内同一施設に複数台設置し、遅延なく業務を行える体制を備えていること。
- 6.4 解析情報の拡散リスクを減少するため、また均一な品質のデータを得るために、全作業内容を同一施設内で実施できること（他の事業者には再委託しないこと）。
- 6.5 輸送時の供与物の品質劣化を抑えるために、解析時にトラブル等が認められた場合迅速に対応策を決定するため、解析施設は日本国内にあること。
- 6.6 作業が適切に行われるために、CAP-LAP、ISO13485 および衛生検査所の承認を持つ施設であること。
- 6.7 解析完了後、委託者と残余検体（供与サンプル、作製ライブラリ）の返送について協議すること。甲に返送が必要な場合における返送費用は、乙の負担とする。
- 6.8 データストレージ装置にアクセス制限ならびにファイル変更履歴確認システムを備えていること。また解析サーバーは国外からアクセスできないものを使用すること。
- 6.9 委託業務の実施状況に関して問い合わせを行った場合、書面又は口頭による報告を、2 営業日以内に行うことができること。更に、要望に応じて査察対応が可能であること。
- 6.10 NovaSeq6000 または NovaSeq X Plus を使用したヒト全ゲノムシーケンス解析について、年間 4,000 検体以上の解析実績があること。また、国立研究開発法人日本医療研究開発機構の実施する研究事業からの委託業務として、ヒト全ゲノムシーケンス解析の受託実績が年間 3,000 検体以上あること。
- 6.11 全ゲノムシーケンスデータ取得について、外部精度評価を受けていること。

7 納品場所

納 期 乙は、止むを得ない事情により、契約終了日までの納入が困難であることが見込まれる場合は、少なくとも3か月前までに書面により甲に報告の上、協議すること

納品場所 静岡社会健康医学大学院大学

8 成果品

以下の情報を、2つのハードディスクにミラーコピーしたものを納品すること。

- 1) 解析報告書
- 2) Fastq / BAM / VCF ファイル
- 3) リード配列の品質情報
- 4) 検体情報

9 定めのない事項

本契約及び仕様に定めのない事項は、静岡県業務委託契約約款に拠り処理する。