

Summary

UK Biobank requires a sequencing provider that can offer Whole Genome Sequencing (WGS) services that meet the requirements outlined in this specification.

The key parameters for any bidder are as follows:

- There is up to £30m funding available for a vanguard phase of whole genome sequencing;
- The scope of the tender is for sequencing (and certain named ancillary) services only;
- The sequencing for this vanguard phase should ideally be completed by the end of 2019 (see section 1.2.1 for preferred timetable).

The scope of this tender solely relates to the sequencing of samples to produce read-level data and associated quality metrics. UK Biobank recognises the separate need for a substantial informatics capability that will turn these output data into research ready artefacts (including the genotyping of Single Nucleotide Variants (SNVs), insertion deletion events (indels) and Structural Variants (SVs)) and the data storage and dissemination of these to the external research community.

This specification section sets out the aggregate criteria on which all bids will be assessed and marked. There are four categories by which bids will be assessed. These are set out in section 1.5 of Volume 1 along with the weighting which will be given to the responses.

Bidders are specifically asked to consider their responses in relation to a) their ability to deliver the specification in the manner required and b) their ability to deliver the specification to the required quality standards.

What UK Biobank will provide in terms of samples

UK Biobank extracted DNA from an 850 µL buffy coat aliquot derived from 10 mL of whole blood collected into an Ethylenediaminetetraacetic acid (EDTA) vacutainer during baseline sample collection. Samples are currently in long term storage at -80°C stored as ~425 µL aliquots within the UK Biobank automated biorepository.

UK Biobank will provide extracted, high-quality genomic DNA quantified by ultraviolet-visible spectroscopy (UV/VIS) method using the Trinean DropSense™ 96¹ and normalised to the sequencing provider's specified concentration in labware provided by the selected provider (and agreed by UK Biobank). Labware and barcodes provided by the supplier must be compatible with UK Biobank's TECAN automated liquid handlers (SBS footprint labware). A sample manifest will be provided by UK Biobank with each shipment of samples containing a pseudo participant identifier, plate well position (or tube ID if preferred), plate ID and UV/VIS (Trinean) concentration.

¹ UK Biobank DNA samples have been quantified by UV/VIS measured on the Trinean DropSense™ 96. This method has been shown to correlate well (R = 0.85) with concentrations obtained by Affymetrix using the Picogreen method on these samples - Welsh, S., et al., *Comparison of DNA quantification methodology used in the DNA extraction protocol for the UK Biobank cohort*. BMC Genomics, 2017. **18**(1): p. 26.

1. Technical Requirements

1.1 What bidders will be expected to provide in terms of shipping and storage of samples

- Barcoded labware for formatting of the samples at UK Biobank. The plate type must be specified in the supplier response. UK Biobank will ship samples in the supplier labware to the sequencing provider. The bidder must describe both the barcode format and orientation (i.e. long or short side of plate) in addition to the barcode size (including any spacing or other requirements);
- Shipping services for collection of samples from UK Biobank premises (UK location) including organising shipments with their chosen courier and provision of shipping consumables (as per UN3373 guidelines) in shipments of no more than 5,000 samples per shipment and no more than 5,000 samples per month;
- Interim storage of UK Biobank samples prior to processing (the provider should have a demonstrably robust procedure for sample receipt, sample storage and access control; this should include routine temperature monitoring and back up procedures should storage equipment fail).

1.2 What bidders will be expected to provide in terms of sequencing and generation of data

1.2.1 Target throughput:

The sequencing provider must commit to meeting target throughput during this vanguard phase. These targets will be jointly agreed, but UK Biobank's expectation is that this will involve a minimum of:

- 20% of genomes to be sequenced within the first 6 months from the date of first shipment of samples from UK Biobank (which is intended to be the start of calendar Q3 of 2018);
- 100% of genomes to be sequenced within (preferably) 18 months from the date of first shipment of samples from UK Biobank.

This should not be taken to rule out bids where the timeframe is longer than 18 months (but before the end of 2020) and further, bidders are invited to specify the cost implications (if any) of an 18 month (as opposed to a longer) timeframe.

1.2.2 Read-level and quality control data:

Data format for read-level data:

- Widely-supported open file format with at least 2 publicly available implementations (e.g. CRAM v3)
- Data density less than 1 byte/base
- Total data volume per genome <100GB
- Includes unmapped reads
- Duplicate reads marked
- Reads sorted by start coordinate
- Index file provided, if relevant
- MD5 checksum provided
- Compatible with informatics tool chain described below.

Quality control data:

- All calculations of data quantity and coverage to be excluding duplicate reads, adaptors, overlapping bases from reads from the same fragment, soft-clipped bases;
- Minimum read-length >100bp (to ensure downstream structural variation call quality);
- Minimum of 85Gbp of sequencing data per sample (equating to 30X WGS coverage);
- At least 95% of the genome covered to >15X;
- If using sequencing technology that generates read-pairs from the same DNA fragment, the proportion of mapped read-pairs with appropriate orientation and separation should be >95% of the total mapped reads.

Sample Quality Control:

- Contamination level determined to be low (Freemix <1%);
- Sample identity verified against UK Biobank genotype data, discordance at non-reference genotypes <2%.

1.2.3 Existence of an informatics tool chain:

- Demonstrable existence of an informatics tool chain for generating SNV and indel calls (with preference for open source tools):
 - *of the required quality:* SNV recall and precision >99.9%, indel recall and precision >99% on precision FDA standard genome [comparison data](#);
 - *that is cost efficient:* <£10 per genome to perform data processing from unmapped reads to SNV and indel genotypes;
 - *that is sufficiently scalable:* capable of handling 100,000's genomes per year;
 - *is flexibly deployable:* i.e. within an academic HPC and/or in public clouds.

1.3 *What bidders will be expected to provide in terms of provision and transfer of data*

The deliverable data comprises:

- Read-level data per sample (aligned map file, index file and MD5 checksum);
- Quality control data (both data quality control and sample quality control).

The deliverable data must be transiently stored by the supplier in a secure network system at such standard as would be reasonably expected for the storage of valuable, sensitive, clinical and/or confidential information.

The sequencing provider should have sufficient transient storage capacity to securely store data for up to 5,000 whole genomes through the course of the vanguard project which would also allow time for UK Biobank to put in place storage and informatics to receive these data at UK Biobank (or its nominated third party).

The sequencing provider will need to establish a secure transfer mechanism that will meet the necessary throughput requirements (>2Gbit/s) and be responsible for any data egress charges from their sequencing facility.

Once data have been transferred and receipt confirmed by UK Biobank, upon instruction by UK Biobank, the sequencing provider shall delete all remaining copies of transferred data.

1.4 What bidders will be expected to provide in terms of quality management processes

The sequencing provider must be able to describe their quality management approach (including any internationally recognised quality accreditations) to the entire in-scope sequencing process (from shipping to the provision of data) including validation, procedure, training of staff and data acquisition quality assurance and any subsequent downstream data processing. Consideration should be given to whether there are distinct quality requirements between data for research as compared to clinical purposes.

The sequencing provider must have processes, systems and infrastructure in place to mitigate the relevant risks to the vanguard project, including the following specific risks:

- Prevention of loss/degradation of stored batches of DNA dispatched from UK Biobank;
- Prevention of cross-contamination of stored samples prior to processing;
- Prevention of unauthorised access to the stored batches of DNA;
- Systems to maintain a robust data trail from DNA sample to data (read-level data, data quality control, and sample quality control data).

The sequencing provider must have processes in place to monitor and control reagent batch quality plus details of any processes required and in place to ensure that changes of reagent batch / lot do not negatively impact data quality.

The sequencing provider must have mechanisms in place to ensure continued supply of services/consumables and reagents (for example, if a batch of reagent fails quality control, how will the sequencing service be continued).

1.5 What bidders will be expected to provide in terms of project governance

UK Biobank will dedicate one or more project manager(s) to this project. They will liaise closely with the sequencing provider to maintain an efficient flow of samples and data and to develop and report appropriate project metrics. They will be the principal point of contact for day-to-day management of the project and for planning and issue resolution. The sequencing provider must provide an equivalent project manager(s) to act as point of contact for managing their responsibilities within the project.

UK Biobank has established an expert working group to guide the approach to whole genome sequencing analysis and dissemination of these output data to the research community. UK Biobank would require that a named representative from the successful bidder for the provision of WGS sequencing services be available to join the expert working group and partner with UK Biobank, so that this group can maximise the efficiency and process flows across all aspects of the WGS programme.

2. Price

Prices given should be total prices per genome for the provision of all services outlined within this invitation to tender and inclusive of:

- Provision of consumables into which DNA samples can be reformatted by UK Biobank;
- Collection and shipping of DNA samples from UK Biobank as outlined in section 1.1;
- Whole genome sequencing of the DNA samples and generation of deliverable data as outlined in section 1.2;
- Holding the data output files resulting from the sequencing and provision of those data to UK Biobank (or its nominated third party) in the format outlined within section 1.3.

Prices given should reflect sequential numbers of UK Biobank DNA samples at 10,000 sample intervals up to the maximum number of samples they are prepared to sequence (which does not necessarily have to be an exact multiple of 10,000) for the funds available. UK Biobank proposes to make appropriate pro-rata payments to suppliers that reflect acceptance of WGS data of appropriate quality at regular intervals through the term of the agreement.

All prices submitted must be quoted in Pounds Sterling with a fixed price for the duration of the Contract.

Bidders are also asked for their (non-binding) views on the probable factors which may influence the pricing and the general direction of such pricing for whole genome sequencing over the next 5 years.

3. Scale

UK Biobank is seeking a partner with proven capability to operate at a scale that can deliver this vanguard phase and inform the specification of the main programme of sequencing all 500,000 UK Biobank participants (subject to confirmation of funding).

Bidders are asked to address whether their operations have the necessary scale to undertake whole genome sequencing for this vanguard phase, and address whether they have the ability to scale for the rest of the UK Biobank participants within the next 5 years.

UK Biobank's expectation is to minimise the number of technology changes, both during the vanguard phase and into the main phase of the programme. Bidders are asked to address the ability of whole genome sequence data generated from their platforms to work interoperably with equivalent data generated from other platforms.

4. Collaboration

4.1 Collaboration during the vanguard phase of sequencing

UK Biobank realises that the undertaking of this vanguard phase of WGS of the UK Biobank participant samples is a significant undertaking. It is fully anticipated that the WGS protocol and processes may need to develop over the course of the project in order to maximise the value of the outputs to the research community. As a result, UK Biobank is seeking to enter into a relationship with the provider of WGS which is more collaborative as compared to that of a typical customer/supplier arrangement.

Bidders should consider how they can work with UK Biobank for the benefit of this project; for example how they would support decision-making process in respect of the WGS itself and develop the downstream data formats and the informatics platforms and processes.

Consideration should be given to how UK Biobank expects that the vanguard phase will require a collaborative approach with the sequencing provider, where UK Biobank and the provider would work together to determine the optimum specifications and outputs (balancing cost, quality and time) for sequencing the whole cohort.

4.2 Establishing research capacity to support the main phase of sequencing

The present tender is for a vanguard (i.e. pilot) phase of WGS for UK Biobank, and it is intended that the remaining participants will be sequenced as part of a main phase (subject to funding). This is one of the key objectives (in the field of genetic sequencing) articulated in the UK Life Sciences Strategy [[available here](#)].

Bidders are asked to consider how their involvement with UK Biobank will contribute to the delivery of the genetic goals set out in the Life Sciences Strategy (for example, in the areas of collaborative working between different sequencing initiatives, technological enhancements and skills development) and serve to increase related research and development spending generally.