HUMAN TECHNOPOLE

NATIONAL FACILITY FOR GENOMICS

CALL FOR ACCESS

24-G-PILOT

Amendment I

The present amendment is aimed at *i.* revising the eligibility and admissibility criteria described in paragraph 4 of the call for Access (points 1 and 2) and *ii.* updating the service list (point 3).

1. List of eligible public research entities, given the peculiarity of the research topics covered:

   a) Area di Ricerca Scientifica e Tecnologica di Trieste - Area Science Park;
   b) Agenzia Spaziale Italiana - ASI;
   c) Consiglio Nazionale delle Ricerche - CNR;
   d) Istituto Italiano di Studi Germanici;
   e) Istituto Nazionale di Astrofisica - INAF;
   f) Istituto Nazionale di Alta Matematica "Francesco Severi" - INDAM;
   g) Istituto Nazionale di Fisica Nucleare - INFN;
   h) Istituto Nazionale di Geofisica e Vulcanologia - INGV;
   i) Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS;
   l) Istituto Nazionale di Ricerca Metrologica - INRIM;
   m) Museo Storico della Fisica e Centro Studi e Ricerche "Enrico Fermi";
   n) Stazione Zoologica "Anton Dohrn";
   o) Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione - INVALSI;
   p) Istituto Nazionale di Documentazione, Innovazione e Ricerca Educativa - INDIRE;
   q) Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria - CREA;
   r) Agenzia Nazionale per le Nuove Tecnologie, l'energia e lo Sviluppo Sostenibile - ENEA;
   s) Istituto per lo Sviluppo della Formazione Professionale dei Lavoratori - ISFOL (a decorrere dal 1° dicembre 2016 denominato Istituto nazionale per l'analisi delle politiche pubbliche - INAPP);
   t) Istituto Nazionale di Statistica - ISTAT;
   u) Istituto Superiore di Sanità - ISS;
   v) Istituto Superiore per la Protezione e la Ricerca Ambientale - ISPRA, ferme restando le disposizioni di cui alla legge 28 giugno 2016 n. 132;

   z) Istituto nazionale per l'assicurazione contro gli infortuni sul lavoro – INAIL.

2. The sentence *"Eligible Institutions are strongly encouraged to limit the number of applications to the very best 2/ NF, with at least 50% coming from Junior PIs."* Is revised as follows *"Eligible Institutions/ Institutes are strongly encouraged to limit the number of applications submitted during one evaluation window to the very best two for each NF, with at least 50% coming from Junior PIs."*

3. The service SID: G-001 – Whole Genome Sequencing (WGS) is now available also for animal studies.

   Service description is updated as follows:

   SID: G-001 – Whole Genome Sequencing (WGS)

   Services description:

Whole Genome Sequencing (WGS) is a comprehensive and high-throughput technique that enables the complete DNA sequence of an organism's entire genome. Whole Genome Sequencing is a powerful tool with applications in various fields, including genomics research, personalized medicine, and clinical diagnostics. It provides a comprehensive view of an organism's genetic makeup, enabling a deeper understanding of genetic variations, evolution, and the genetic basis of diseases.

Here is an overview of the whole genome sequencing process:

DNA Extraction: The first step involves extracting genomic DNA (gDNA) from the biological sample, which could be cells, tissues, or even an entire organism. The goal is to obtain a high-quality and pure DNA sample (this task must be undertaken by the Users).

Library Preparation: The extracted DNA is then fragmented into smaller, manageable pieces. Adapters are added to these fragments to allow the subsequent sequencing process.

Sequencing: The prepared DNA library is subjected to high-throughput sequencing technologies, such as next-generation sequencing (NGS) platforms. These technologies generate short DNA sequences, or reads, from the DNA fragments.

Bioinformatic analysis of data from WGS can consist of several steps depending on the biological question:

- Read Mapping: When a reference sequence for the organism being sequenced is available (common case for model organisms), the reads are mapped to the reference generating an alignment.
- Genome Assembly: The generated sequencing data can be computationally assembled to reconstruct the complete genome sequence. This process involves overlapping the short reads to create longer contiguous sequences, or contigs.
- Genome Annotation: The assembled genome is annotated to identify and characterize various genomic features, including genes, regulatory elements, and non-coding regions. Annotation helps in understanding the functional elements within the genome.
- Variant Calling: The sequenced reads are compared to the reference genome to identify variations, such as single nucleotide polymorphisms (SNPs), insertions, deletions, and structural variations. This information is crucial for studying genetic diversity, population genetics, and identifying potential disease-causing mutations.

Bioinformatic analysis of WGS data can be provided as a combined service by the NF for Data Handling and Analysis. SID: NF62.001

Library preparation protocol:

Libraries will be prepared by following the protocol:

Illumina DNA PCR-Free Library Prep Reference Guide (1000000086922)

Illumina DNA PCR-Free offers a unique combination of benefits from on-bead tagmentation and PCR-free chemistry steps. On-bead tagmentation supports bead-

based normalization, easy volume-based library pooling, and elimination of pre- and post-library quantification steps. The PCR-free workflow simplifies and reduces the overall workflow time while providing highly uniform coverage across repetitive or uneven genome regions. For sensitive applications such as human WGS, de novo assembly of microbial genomes, or tumour–normal variant calling, Illumina DNA PCR-Free delivers uniform coverage, and high-accuracy data.

Libraries sequencing and NGS coverage: Libraries will be sequenced using the NovaSeq 6000 system (Illumina) by generating 150 bp Paired End reads.

Next-generation sequencing (NGS) coverage describes the average number of reads that align to, or "cover," known reference bases. Sequencing coverage requirements vary by application, as noted below. At higher levels of coverage, each base is covered by a greater number of aligned sequences reads, so base calls can be made with a higher degree of confidence.

For WGS applied to population and rare disease studies an average human genome coverage of 20X per sample will be obtained.

For WGS applied to Cancer studies an average human genome coverage of 50X per sample will be obtained.

For WGS applied to Metagenomic studies the coverage per sample will be determined considering the type and complexity of the metagenome under study.

For WGS applied to Plant studies the coverage per sample will be determined considering the plant genome size and ploidy level.

For WGS applied to Animal studies the coverage per sample will be determined considering the animal genome size.

Within this call the NF for Genomics will provide:

WGS for Population and medical studies (coverage 20X) for a maximum number of 3000 samples. Projects with a sample size ranging from a minimum of 1000 to a maximum of 3000 samples will be accepted.

WGS for Rare diseases studies (coverage 20X) for a maximum number of 100 samples. Projects with a sample size ranging from a minimum of 20 to a maximum of 100 samples will be accepted.

WGS for Cancer studies (coverage 50X) for a maximum number of 100 samples. Projects with a sample size ranging from a minimum of 20 to a maximum of 100 samples will be accepted.

WGS for Metagenomics studies (coverage determined considering type and complexity of the metagenome) for a maximum number of 200 samples. Projects with a sample size ranging from a minimum of 50 to a maximum of 200 samples will be accepted.

WGS for Plants studies (coverage determined considering the plant genome size and ploidy level) for a maximum number of 20 samples. Projects with a sample size ranging from a minimum of 5 to a maximum of 20 samples will be accepted.

WGS for Animal studies for a maximum number of 100 samples. Projects with a sample size ranging from a minimum of 20 to a maximum of 100 samples will be accepted.

Technical requirements: gDNA samples should be provided in low-bind full-skirted PCR plates (i.e. Eppendorf twin.tec PCR plates) sealed using peelable adhesive PCR films. Samples should be ordered column-wise without leaving empty positions. gDNA samples should have a concentration of at least 15 ng/µl in 50 ul of nuclease-free ultrapure water. gDNA samples should be quantified by using a fluorometer (i.e Qubit/ Glomax). Quality of gDNA should be evaluated by Agilent Tape Station/Bioanalyzer, samples should have a DIN≥6 (DNA Integrity Number≥6). Purity of the gDNA samples should be assessed with a Spectrophotometer (i.e. Nanodrop; 260/280≥1,8 and 260/230≥1,8).

Results that will be delivered to the Users:

The NF for Genomics will deliver to the Users the following files for every sample sequenced:

- FASTQ files
- QC report
- Mapping metrics (if reference genome/transcriptome is available)

For projects in which the bioinformatic analysis of WGS data will be requested as a combined service from the NF for Data Handling and Analysis, the following files will be delivered to the Users:

- Alignment (BAM files)
- Mapping metrics
- Identified variants (VCF files)
- Analysis QC reports

Access modality available: Access to NF services.

Services available in combination with the NF for Data Handling and Analysis: SID: NF62.001 - Bioinformatic analysis.