HUMAN TECHNOPOLE

NATIONAL FACILITY FOR GENOMICS

CALL FOR ACCESS

25-G-ROUND1

Table of Contents

1. INTRODUCTION

The Access of Researchers affiliated with Universities, *Istituti di Ricovero e Cura a Carattere Scientifico* (IRCCS), and Public Research Entities to Fondazione Human Technopole (HT) National Facilities (NFs) is regulated by the NF Access rules available on the NFs dedicated webpage (link).

Services offered by NFs are available through regular open calls for Access that are published yearly on the HT website (link) and are free of charge for the project (or aspects of the project) approved for Access.

The open call for Access is aimed at supporting Access to the technologies offered by the NFs and it is not meant to provide direct funding to the Applicant. The costs for the activities to be performed at the NFs will be fully covered, including shipment of relevant material from and to the Applicant's laboratory as well as travel and accommodation for the Applicant and/ or Applicant's team member(s) (User) while accessing the NF. Project-related costs (personnel, consumables, and other costs) at the Applicant's laboratory are not funded.

The User Access workflow comprises different steps, spanning from the initial submission of the application to evaluation and Access approval, Access to the performance of the service(s) and Access conclusion. A detailed description of the workflow is available on the NFs dedicated webpage (link).

1.1 Access modalities

Three different Access modalities can be requested. Their availability will vary, based on the service specifics of each NF:

- **"Simple" Access to NF or individual instruments thereof**: This modality is intended for Users involved in projects requiring technologies that are available at the NF for **direct Access by User**. This Access modality requires prior expertise with the technology of interest. After an initial introductory training aimed at defining the level of expertise of the User, **the use of the instrument with limited supervision by NF staff is authorised**. For defined NFs/ instruments/ services this Access modality may be restricted or not available.
- **Access to NF services**: This procedure entails the provision of **services performed by NF staff on behalf of the User**. NF services may include both standard services as well as, when foreseen by the technology development specifics of each NF, bespoke services conceived and discussed with the User. To allow the NF staff to best align the experimental activity to the research objective, the User may be invited, if needed, to assist the NF staff while performing the project or aspects of it.
- **Access to NF services including training**: This procedure **entails training by NF staff** to provide Users, in addition to or alternatively to the services described in the previous modality, with training courses and/or programs, aimed at transferring the expertise necessary for the independent use of the specific technology. In this case, technical and/or experimental activities are conducted with the active participation of the User. Training can be provided by NF staff while performing the service(s) or in a dedicated session. This type of Access is also aimed at researchers who want to acquire expertise for subsequent independent use of a specific technology in other laboratories.

2. TERMS AND DEFINITIONS

2.1 Access

"Access" refers to the authorised use of the NF and of the services offered. Such Access can be granted for sample preparation, set-up, execution and dismantling of experiments, education and training, expert support and analytical services, among others. Access to the NFs includes all infrastructural, logistical, technical and scientific support (including training) that is necessary to perform the aspects of the project approved for Access.

2.2 Researcher

"Researcher" is a professional engaged in the conception or creation of scientific knowledge. They conduct research and improve or develop concepts, theories, models, techniques, instrumentation, software or operational methods.

2.3 Principal Investigator

"Principal Investigator" (PI) is the Researcher affiliated with an eligible Institution with the role of independent Group Leader, who is responsible for coordinating the research activities conducted within the framework of the submitted project.

The PI shall hold a primary appointment as Group Leader at an eligible Institution, with the following requisites:

- Coordinate an independent research team.
- Have a supervisory role towards junior and/ or senior Researchers.
- Their Group has an autonomous budget sufficient to cover their current research expenses.
- Be the recipient of independent research funding as PI or co-PI.

Junior PI: Up to 6 years from their first appointment in an independent Group Leader position.

The period specified above may be extended beyond 6 years in the event of adequately documented career breaks, occurring before the submission of the application and resulting from:

i. Maternity leave: The time limit is increased by 18 months for each child born after their first appointment in an independent group leader position; if the Applicant is able to document a longer total maternity leave, the period of eligibility will be extended by a period equal to the documented leave, taken before the submission of the application. Maternity status must be documented by submitting the birth certificate of the child or children.

ii. Paternity leave: The time limit is increased by the actual amount of paternity leave taken before the application submission deadline for each child born after their first appointment in an independent group leader position. Paternity status must be documented by submitting the birth certificate of the child or children.

iii. Long-term illness of more than 90 days, or national service: The time limit is increased, for each eligible event occurring after their first appointment in an independent group leader position, by the actual amount of leave from which the Applicant has benefited prior to the application submission deadline.

Established PI: More than 6 years from their first appointment in an independent group leader position.

2.4 Applicant

"Applicant" is the Principal Investigator who applies to a NF open call for Access and who is responsible for the submitted project. They can be of any nationality and must be affiliated with an eligible Italian Institution, as detailed in section 4.

2.5 User

A "User" is intended as a Researcher affiliated with an eligible Institution who accesses, physically or remotely, the NFs to perform the approved activities or to support the National Facility staff while performing the approved service.

If requested by the Applicant, the User of the NF can also be a separate member of their research team.

3. APPLICATION TYPE

Applicants shall select the type of application they want to submit, choosing between two options:

a. **Standard** application for projects that are technically mature.

b. **Proof-of-concept** application for:

i. Projects with high scientific potential but with insufficient technical maturity or preliminary data.

ii. Projects aimed at setting up the experimental conditions required for a standard project, including methods or technology development projects.

iii. Time-limited Access projects (e.g., to acquire data to complete a manuscript, or preliminary data needed for a grant application, or single microscopy session).

4. ELIGIBILITY AND ADMISSIBILITY

PIs, as defined in section 2.3 of this call, affiliated with an eligible Institution are eligible to apply. The Applicant's role as a PI shall be confirmed by their Institution in a mandatory letter of Institutional endorsement (Template available in Annex I).

Applications from Researchers who are not independent should be submitted by their Group Leader. Applicants are strongly encouraged to support NF Access by young Researchers (R1 and R2 profiles of the European Framework for Research Careers, link) who are part of their group. In this case, the Applicant shall indicate in the application form that the NF User is a member of their group, specifying User's career stage.

Below are the links to the relevant lists of eligible Institutions:

Universities: This category includes Institutions recognized by the Ministry of University and Research (link). In detail:

i. State funded public universities, listed under the following link.

ii. Specialized superior graduate schools or Institutions, listed under the following link.

iii. Legally recognized non-public universities, listed under the following link.

iv. On-line universities, listed under the following link.

Istituti di Ricerca e Cura a Carattere Scientifico (IRCCS): this category includes Institutions recognized by the Ministry of Health and listed at the following link.

Public research entities: this category includes:

a) Institutions recognized by the Ministry of University and Research and listed at the following link.
b) Area di Ricerca Scientifica e Tecnologica di Trieste - Area Science Park;
c) Agenzia Spaziale Italiana - ASI;
d) Consiglio Nazionale delle Ricerche - CNR;
e) Istituto Italiano di Studi Germanici;
f) Istituto Nazionale di Astrofisica - INAF;
g) Istituto Nazionale di Alta Matematica "Francesco Severi" - INDAM;
h) Istituto Nazionale di Fisica Nucleare - INFN;
i) Istituto Nazionale di Geofisica e Vulcanologia - INGV;
j) Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS;
k) Istituto Nazionale di Ricerca Metrologica - INRIM;
l) Museo Storico della Fisica e Centro Studi e Ricerche "Enrico Fermi";
m) Stazione Zoologica "Anton Dohrn";
n) Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di
o) 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;
w) Istituto nazionale per l'assicurazione contro gli infortuni sul lavoro – INAIL.

Eligible Institutions/ Institutes are strongly encouraged to limit the number of applications submitted to this call for Access to the very best two, with at least 50% coming from Junior PIs.

Such indication does not represent an eligibility criterion but rather a guideline aimed at ensuring the widest distribution of Access among Institutions in the Country.

Applicants shall declare that they have not received funding to perform the submitted project (limited to the aspects included for Access to the NF) in their own laboratory, home Institution or elsewhere. Applicants shall confirm the economic and scientific feasibility for the aspects of the project to be performed outside the NFs.

Applicants cannot request Access for the same service if an approved Access is ongoing. Before submitting a new application for the same service, Applicant shall consult with the NF staff and confirm that the ongoing Access will be completed before the end of the next evaluation round. A clear motivation for the request must be provided.

A PI submitting an application to this call for Access cannot request access to other NFs (i.e., cannot participate to other 2025 - ROUND 1 calls for Access). If more than one application is submitted, **ALL will be rejected** during administrative review. Applicants who have an application under evaluation are not allowed to submit another application before receiving notification of the results.

Applications must be written in English and must be complete (i.e., consist of all the requested elements and information) and respect all administrative and technical requirements (e.g., proposal or CV format, mandatory declarations, technical requirements of the services, sample availability, sample requirements, including number of samples to be analysed). Incomplete applications or applications that do not meet the requirements will be considered not admissible and will be rejected at the administrative review stage.

5. APPLICATION CONTENT AND FORMAT

The application, to be submitted through the online portal PICA (link) consists of six components:

1. **Applicant's general information**.
2. **Justification for requesting Access to the NF**.
3. **Abstract** to be inserted in the dedicated section on the application portal (Max 1500 characters including spaces).
4. **Project proposal**, to be uploaded in PDF format in the dedicated section on the application portal, shall include the following sections:
    a. *Title.*
    b. *Significance.*
    c. *Innovation.*
    d. *Approach, including aims, preliminary data in support of the proposed experiments, experimental design and anticipated results.*
    e. *Environment, including facilities and resources available to support the aspects of the project to be performed elsewhere (i.e., outside the NF).*

    Below, the mandatory format for the proposal:

    Standard application: Max 3 pages (Page format: A4, Font type: Arial, Font size: at least 11, Line spacing: single, Margins 2 cm side/ 1.5 bottom) figures included, references excluded. Accepted file formats: PDF. Max size: 30MB - Name the file as APPLICATION ID_PROPOSAL_Surname

    Proof-of-Concept application: Max 2 pages (Page format: A4, Font type: Arial, Font size: at least 11, Line spacing: single, Margins 2 cm side/ 1.5 bottom) figures included, references excluded. Accepted file formats: PDF. Max size: 30MB - Name the file as APPLICATION ID_PROPOSAL_Surname

Proposal template is available in Annex II of this call.

Applications that do not meet the format requirements will be considered not admissible and will be rejected at the initial administrative review stage.

5. **Applicant's CV in NIH biosketch format**. The CV, to be uploaded in PDF, shall be drafted in English, using the template available at this link and following the mandatory format: max 4 pages, page format: A4, Font type: Arial, Font size: at least 11, Line spacing: single, Margins 2 cm side/ 1.5 bottom. For support in drafting the CV, please refer to NIH website: Create Biosketches | NIAID: National Institute of Allergy and Infectious Diseases (nih.gov).

   Applications that do not meet the format requirements will be considered not admissible and will be rejected at the administrative review stage.

6. **Letter of Institutional Endorsement**, addressing the following points:

   a. *Confirmation of the Applicant's role at their Institution, and their eligibility under the category of PI (see section 2.3).*

   b. *Confirmation that relevant authorisations, declarations and accreditation from the competent authority(ies) have been obtained or will be obtained no later than two (2) months after Access approval, in order to process samples and data through the NFs.*

   c. *Justification of the request for Access – including a statement on why the project cannot be performed at the Applicant's Institution.*

   d. *Confirmation that the Applicant has not received funding for performing the submitted project, for the aspects to be performed at the NFs, in their own laboratory, home Institution, or elsewhere.*

   e. *Confirmation of the project's economic and scientific feasibility for the aspects to be performed at the host Institution.*

   f. *Acceptance of NF Access Rules.*

   The Letter of Institutional Endorsement, to be uploaded in PDF or p7m in the dedicated section on the application portal, shall be drafted using the facsimile available as Annex I of this call.

7. **Technical information**, to be filled in in the dedicated section(s) of the application portal, indicatively including:

   a. *Requested service(s), as described in Annex III of this call.*

   b. *Sample technical information.*

   c. *Requested preliminary data for technical feasibility analysis (if applicable).*

   d. *Whether the entire sample set is already available (otherwise indicate the date of availability of the entire sample set).* ***It is mandatory that samples and relevant authorisations are available at the moment of application or no later than two (2) months from receiving Access approval****.*

   e. *Resources and expertise to receive and process the products – data (e.g. Cryo-EM micrographs) or reagents (e.g. human iPSCs) – generated by the NF.*

   f. *Research data management plan and bioinformatics support for data analysis, specifying (mandatory when the project output includes research data - e.g.,*

genomics or proteomics data, bioimages from microscopy services, among other):

i. *How the bioinformatics analysis of the data generated by the NF will be performed (if such analysis is not provided by the NF for Data Handling and Analysis).*

ii. *How the data generated by the NF will be handled during and after the end of the project.*

iii. *Whether and how the data will be shared/ made Open Access.*

iv. *How data will be curated and preserved, including after the end of the project.*

Details and format of the technical information to be provided are available in the dedicated section of the application portal.

Information provided in sections 1 and 6 are used for the eligibility and admissibility check.

Information provided in section 7 is used for assessing the technical feasibility of the aspects of the project to be performed at the NF.

The entire application is evaluated by the Standing Independent Evaluation Committee (SIEC) to assess its scientific merit.

6. APPLICATION SUBMISSION METHODS, CALL DEADLINE AND EVALUATION PERIODS

Applications shall be submitted exclusively through the application portal PICA managed by CINECA and accessible at this link, according to the indicated terms and methods.

This call for Access (Call ID: 25-G-ROUND 1) will open on the 15th of February 2025 (13:00 CET) and will close on the 31st of May 2025 (13:00 CET).

A comprehensive list of services, available equipment and the technical requirements for Access as well as terms and conditions are available on the dedicated NFs webpage (link).

The complete list of offered services and technical requirements are available in the Annex III of this call.

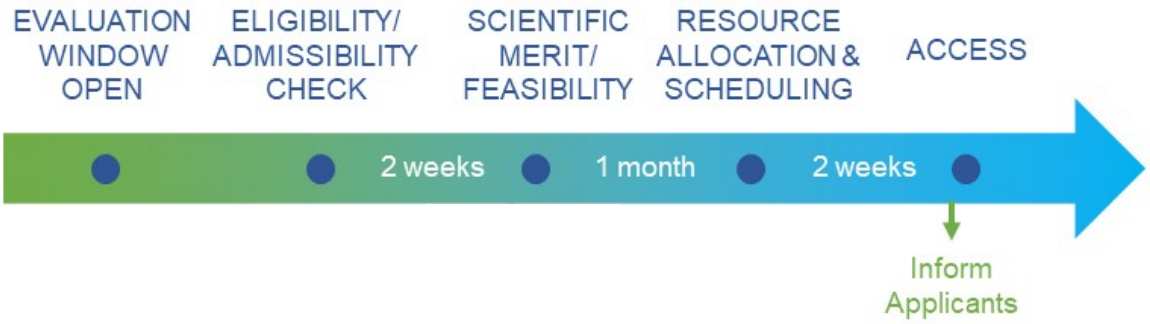
Samples as well as relevant authorisation for their use, shall be available at the moment of submitting the application or not later than two (2) months after Access approval. When the project foresees the analysis of more than one batch of samples, the first batch shall be available when the application is submitted or not later than two (2) months after Access approval.

7. EVALUATION OF APPLICATION

The evaluation procedure is conducted by the SIEC that is supported by a Panel of independent external Reviewers (Review Panel) selected by the SIEC on the basis of their scientific expertise.

Each Review Panel is composed of 2 SIEC members, who will act as Chairs, plus 10 appointed external Reviewers, with the relevant expertise.

Below is a scheme describing the evaluation steps and timeline.



There are four application categories that are evaluated and ranked separately:

- Junior PI – Standard application
- Established PI – Standard application
- Junior PI – Proof of Concept application
- Established PI – Proof of Concept application

The NF User Access Office first performs an administrative review of the application to ensure that all the requested components have been provided, and that all eligibility criteria have been met. Incomplete applications or applications that do not meet all the requirements will be considered not admissible and will be rejected at the administrative review stage.

The application is then sent to the Review Panel for assessing scientific merit and technical feasibility.

If the number of applications exceeds by a factor of 4 the estimated capacity of the NF, a triage will be applied within each application category by the relevant Review Panel.

Triage criteria will include:

a. Justification for requesting Access to the NF.
b. Field-Weighted Citation Impact (FWCI).
c. Track record in securing research funding.

The application will remain confidential throughout the entire evaluation process. Reviewers will be asked to declare that they do not have any conflict of interest, and they will be bound by a Confidentiality Agreement.

The application will be individually evaluated by three Reviewers who are part of the relevant Review Panel.

Proposals will be evaluated and ranked based on their average score, within each category.

An on-line meeting of the Review Panel may be requested by the Chairs if deemed necessary (for example to discuss proposals with highly discrepant scores).

At least 50% of the available Access will be allocated to applications from the two Junior PI categories.

7.1 Evaluation criteria

The scientific merit of the project is assessed based on the following criteria:

- **Significance**: Overall scientific merit of the proposed research. If all the experiments proposed are successful, how will the resulting knowledge advance the field?
- **Innovation**: Degree of innovation (conceptual and/ or technological), and ambition of the proposed study compared to the state-of-the-art in the relevant field.
- **Approach**: Appropriateness of proposed methodology, preliminary data in support of proposed experiments, and project feasibility.
- **Environment**: Facilities and resources available to support the aspects of the project to be performed elsewhere (i.e., outside the NF).
- **Justification for requesting Access to the NF**: Explanation on why the service cannot be performed at the host Institution, at a cost which is deemed affordable for the applicant.
- **Applicant**: PI's scientific background and expertise.

7.2 Scoring system

A numeric score between 1 (exceptional) and 9 (poor) is provided for each of the six evaluation criteria. Moreover, an overall project score including a short descriptive comment is provided as feedback to the Applicant.

- **HIGH**:
  - **Score 1 (Outstanding)** – The proposal successfully addresses all relevant aspects of the criterion. There are no weaknesses.
  - **Score 2-3 (Excellent - Very Good)** – The proposal addresses the criterion exceptionally well, aside from a small number of minor weaknesses.
- **MEDIUM**:
  - **Score 4-6 (Very good - Good)** – The proposal addresses the criterion well, but a number of weaknesses are present.
- **LOW**:
  - **Score 7-8 (Fair - Poor)** – The proposal broadly addresses the criterion, but there are significant weaknesses.
  - **Score 9 (Poor)** – The criterion is inadequately addressed, or there are serious inherent weaknesses.

7.3 Technical feasibility analysis

During the evaluation, the relevant experts from SIEC will receive a report from NF staff who will perform a comprehensive analysis of the proposed project's technical feasibility. Technical feasibility also includes an evaluation of the fulfilment of the technical requirements in terms of capacity to receive and process the research data generated by the NF, as described in the research data management plan. This latter evaluation is performed in consultation with the NF for Data Handling and Analysis.

Based on the technical maturity of the project, the application can be assessed as Feasible/ Not Feasible/ Proof-of-Concept study required.

7.4 Evaluation results and Access approval

NF staff provides the SIEC with information on the resources needed (cost and time) to perform the highest ranked projects. Applications with the highest scientific score that fulfil all technical requirements are approved for Access by the SIEC, based on the capacity of the NF. NF staff schedules Access. A selected number of applications may be placed on a waiting list (in case of cancellations).

Evaluation results – Access granted, Access conditionally granted, Access waitlisted, Access not granted – are communicated to the Applicant through the Access portal.

Applicants whose applications are placed on the waiting list will receive additional information advising whether the project can be Access approved or should be resubmitted within the subsequent application window.

7. AFTER ACCESS HAS BEEN APPROVED

After Access approval, a kick-off meeting is organised and the Applicant is invited to meet NF staff to discuss the experimental design of the project and to finalize the project plan.

Once the project plan has been agreed and the relevant ethical and legal authorisation(s) for the use of the samples has(have) been provided, the NF User Access Office coordinates the signature of the required formal Agreements (e.g., Access Agreement, Collaboration Agreement, other) and the project can commence.

8. AFTER ACCESS HAS BEEN COMPLETED

At the end of the activities carried out at the NF, and not later than 3 months thereafter, if not differently agreed with the NF User Access Office, the Applicant must submit a short report on the results obtained and the impact of the service on their research. Moreover, a final report to be published on the NFs website and describing the impact of the Access to the NF on the research project for which the service has been requested, shall be provided upon publication of the relevant results. Applicants who will not be able to demonstrate the consistency and relevance of the activities carried out at the NF with the research project for which Access was requested will be considered not eligible to participate in the subsequent calls for Access.

Moreover, the Applicant will be asked to fill in a brief, mandatory survey regarding their experience, providing feedback and suggestions for further service improvement.

The Applicant must communicate to the NF User Access Office (via email to national.facilities@fht.org) any publication acknowledging the NF.

Research data obtained during Access shall be made available to the scientific community following the FAIR principles. Applicant must inform the NF User Access Office (via email to national.facilities@fht.org) when and how the data are made public.

9. CONTACTS

Requests for information and/or clarifications concerning the application procedure may be sent to the dedicated e-mail address national.facilities@fht.org, indicating the call ID in the subject line.

11. REFERENCES

NF Access Workflow_Convenzione (link)

NF Access Rules_Convenzione (link)

NF Access Agreement_Convenzione (link)

11. CHANGES TO THE CALL

Any changes or additions to this notice will be communicated through publication on the NFs website (link).

ANNEX I: LETTER OF INSTITUTIONAL ENDORSEMENT TEMPLATE

(Print on paper bearing the official letterhead of the host Institution)

Endorsement letter of the host Institution

To whom it may concern:

I, the undersigned, …….. (*name of legal representative or special attorney*), born in ……. (*city*) on ……..(*date*), as legal representative *(or special attorney, by means of special power of attorney identified by ……..*) and on behalf of ……(*name of the host Institution*), legal residence in *(referred to the host Institution*) ……(*city*), address …….., regarding the project ID (*refer to the ID allocated to the application on the PICA portal*)…….., presented by ……(*Applicants's first name and surname*), as Principal Investigator on the call for Access to Human Technopole National Facilities…..(ID *of the call*),

Declare

- That the host Institution is among those eligible to participate in the call for Access as it belongs to the following eligible category: (select among University, IRCSS, Public Research Entities);
- That the Applicant, Dr ……. (*Applicant's first name and surname*) is an independent group leader (Principal Investigator) affiliated with a primary appointment at the host Institution and that they meet the eligibility criteria as indicated in the call;
- That the Applicant has not received funding for performing elsewhere, the aspects of the project for which they are seeking here support from or Access to Human Technopole National Facilities;
- That the services requested here cannot be performed by the Applicant at the host Institution, at a cost which is deemed affordable for them;
- That relevant authorisations, declarations and accreditation from the competent authority(ies) have been obtained or will be obtained within two (2) months after the approval of the Access in order to process samples and data through Human Technopole;
- That, if applicable, biological specimens have been obtained with the corresponding approval of the Bioethics Committee and appropriately signed 'informed consent', both for their collection and their use, including conservation, manipulation, derivation and processing to be carried out by Human Technopole National Facilities;
- That, if samples were obtained from subjects who signed an 'informed consent', said informed consent allows that sequencing data and results are included in secure controlled Access databases and accessed/ used by authorised third parties;

and is committed

- To accept the terms and conditions to Access Human Technopole National Facilities as described in the National Facilities Access rules (link);
- To sign the Access Agreement should the project be approved (link)

For the host Institution (Applicant legal entity/beneficiary):

Date ……..

Name and Title …….. ; ……..

Email and Signature of legal representative or delegated person

…….. ; ……..

ANNEX II: PROJECT PROPOSAL TEMPLATE

Mandatory proposal format

Standard application*: Max 3 pages (Page format: A4, Font type: Arial, Font size: at least 11, Line spacing: single, Margins 2 cm side/ 1.5 bottom) figures included, references excluded. Accepted file formats: PDF. Max size: 30MB - Name the file as APPLICATION ID_PROPOSAL_Surname*

Proof-of-Concept application*: Max 2 pages (Page format: A4, Font type: Arial, Font size: at least 11, Line spacing: single, Margins 2 cm side/ 1.5 bottom) figures included, references excluded. Accepted file formats: PDF. Max size: 30MB - Name the file as APPLICATION ID_PROPOSAL_Surname*

PLEASE REMOVE THE INFORMATION ABOVE BEFORE SUBMITTING

Proposal content:

1. TITLE

2. SIGNIFICANCE

3. INNOVATION

4. APPROACH

5. ENVIRONMENT

6. REFERENCES (Optional)

ANNEX III: NATIONAL FACILITY FOR GENOMICS: SERVICE LIST

HUMAN TECHNOPOLE

NATIONAL FACILITY FOR GENOMICS

CALL FOR ACCESS

25-G-ROUND1

SERVICE LIST

Table of contents

1. INTRODUCTION

The NF for Genomics provides state-of-the-art and innovative services in genomics. Its core mission is to implement robust experimental and analytical workflows to probe all major domains of genomic exploration, including but not limited to the analysis of DNA, RNA, chromatin, and other markers of epigenetic and regulatory activity. These techniques can be applied to diverse areas of biology, with resolution spanning to whole organisms, to tissues or individual cells. Overall, the NF for Genomics aims to empower research in all domains of genomics for the Italian scientific community at large.

The NF for Genomics includes four Infrastructural Units (IU):

UI1 High Throughput Sequencing mainly focused on providing state-of-the-art high-throughput sequencing services. This dedicated unit specializes in delivering top-tier genomics, transcriptomics, and epigenomics analyses, employing the latest protocols for sample processing and sequencing.

UI2 Multi-Omics Technologies specialized in multi-omics technologies. Its focus extends to providing cutting-edge services in single-cell and spatial multi-omics analysis, as well as long-read sequencing. This Unit closely collaborates with the Tissue Processing Infrastructural Unit part of the NF for Light Imaging, specifically dedicated to implementing spatial transcriptomics protocols.

UI3 Computational Genomics is the computational core of the NF for Genomics, the team is dedicated to developing, implementing, and maintaining automated pipelines for the pre-processing and primary data analysis of sequencing data. This unit closely collaborates with the NF for Data Handling and Analysis.

UI4 Technology Development stays at the forefront of innovation, this dedicated team is committed to methods and technology development. The highly qualified staff is continuously engaged in the evaluation of new technologies and instruments, considering them for acquisition into our NF. This dedicated unit specializes in the optimization and scale-up of custom experimental protocols, offering these services to the broader scientific community.

Below a list of available services with a detailed description of the service and technical requirements.

Appendix I below summarized the technical requirements for each available service.

2. SERVICE LIST

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. WGS 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 (see technical requirements below).

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.

QC and summary statistics

Bioinformatic analysis of WGS data can be provided as a combined service by the National Facility for Data Handling and Analysis. Please select: SID: NF62.02.01

Library preparation protocol:

Libraries will be prepared by following the protocol:

Illumina DNA PCR-Free Library Prep

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.

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:

	Min samples/project	**Max samples/project**
G-001-A WGS for Population and medical studies (coverage 20X)	1000	3000
G-001-B WGS for Rare diseases studies (coverage 20X)	20	100
G-001-C WGS for Cancer studies (coverage 50X or 100X)	20	100
G-001-D WGS for Metagenomics studies (coverage determined considering type and complexity of the metagenome)	50	200
G-001-E WGS for Plants studies (coverage determined considering the plant genome size and ploidy level)	5	20
G-001-F WGS for Animal studies	20	100

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 the users requesting data analysis as a combined service, the **National Facility for Data Handling and Analysis** will also provide the files described in the SID: NF62.02.01

Access modality available: Access to NF services.

Services available in combination with the NF for Data Handling and Analysis: Please select: SID: NF62.02.01

SID: G-002 – Whole Exome Sequencing (WES)

Services description:

Whole Exome Sequencing (WES) is a targeted sequencing approach that focuses on sequencing the protein-coding regions of the genome, known as the exome. The exome comprises the exons, which are the coding regions of genes, and represents only a small fraction (about 1-2%) of the entire genome. Despite this, the exome contains most known disease-causing mutations, making WES a cost-effective alternative to Whole Genome Sequencing (WGS) for many applications. WES is widely used in both clinical and research settings. In clinical genetics, it is employed for diagnosing genetic disorders, identifying causative mutations, and understanding the genetic basis of rare diseases. In research, WES is valuable for studying the genetics of complex traits and diseases.

Here's an overview of the Whole Exome Sequencing process:

DNA Extraction: Like WGS, the process begins with the extraction of genomic DNA from the biological sample of interest, such as cells or tissues. This step must be taken by users (see technical requirements below).

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

Exome Capture: The next step involves selectively capturing and enriching the DNA fragments corresponding to the exonic regions. This is typically done using target-specific probes or baits designed to hybridize with and capture the exonic sequences.

Sequencing: The prepared exome library is subjected to high-throughput sequencing technologies, such as next-generation sequencing (NGS). The sequencing generates short DNA sequences, or reads, from the exonic regions.

QC and summary statistics

Bioinformatic analysis of WES data can be provided as a combined service by the **National Facility for Data Handling and Analysis**. Please select: SID: NF62.02.02

Library preparation protocol:

The National Facility for Genomics uses the Twist Comprehensive Exome Panel for WES analysis; this panel offers coverage of greater than 99% of protein coding genes. The panel's superior performance provides the optimal exome sequencing solution, while focusing on the most accurate curated subset—CCDS database. It has been created to include expanded content of RefSeq and GENCODE databases. The panel targets in total 36.8 Mb with a design size of only 41.2 Mb covering more than 99% of protein coding genes.

Enzymatic Fragmentation and Twist Universal Adapter System

Twist Target Enrichment Fast Hybridization Protocol

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

For WES an average exome coverage of 50X or 100X per sample will be obtained depending on the application field.

Within this call the National Facility for Genomics will provide:

	Min samples/project	**Max samples/project**
G-002-A Whole Exome Sequencing for Rare diseases studies	25	100
G-002-B Whole Exome Sequencing for Cancer studies	50	700

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 5 ng/µl in 40 µl 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 National Facility 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 the users requesting data analysis as a combined service, the **National Facility for Data Handling and Analysis** will also provide the files described in the SID: NF62.02.02

Access modality available: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: Please select: SID: NF62.02.02.

SID: G-003 – Amplicon sequencing for microbiome analysis (16S-ITS)

Services description:

Microbiome analysis using 16S and ITS amplicon sequencing is a widely used technique to study the composition and diversity of microbial communities, particularly bacteria and fungi. The 16S ribosomal RNA (rRNA) gene is a molecular marker found in the genomes of bacteria and archaea, and its variable regions are commonly used for taxonomic classification, while ITS is used to profile fungal communities. Microbiome analysis using 16S and ITS amplicon sequencing is valuable in a range of fields, including environmental science, human health, agriculture, and more. It provides a cost-effective way to characterize microbial communities and understand their roles in various ecosystems or host-associated environments.

Here's an overview of the process:

Sample Collection: Microbiome analysis typically begins with the collection of samples from the environment of interest, such as soil, water, or biological samples like feces, saliva, or skin swabs (this step must be taken by users).

DNA Extraction: The next step involves extracting DNA from the collected samples. This DNA will contain the 16S rRNA gene from the microbial organisms present in the sample. This step must be taken by users (see technical requirements below).

PCR Amplification: Polymerase Chain Reaction (PCR) is used to selectively amplify the variable regions of the 16S rRNA gene. Primers designed to bind to conserved regions flanking the variable regions are used in this process. The choice of primers can influence the taxonomic resolution and coverage of the analysis.

Library Preparation: The PCR-amplified DNA is then converted into a sequencing library. Adapters are added during the amplification step to enable high-throughput sequencing.

High-Throughput Sequencing: The prepared library is subjected to high-throughput sequencing, commonly using next-generation sequencing (NGS) platforms. This step generates short DNA sequences (300bp) from the variable regions of the 16S rRNA and ITS genes.

QC and summary statistics

Bioinformatic analysis of 16S and ITS amplicon can be provided as a combined service by the **National Facility for Data Handling and Analysis**. Please select: SID: NF62.02.03

Library preparation protocol:

Libraries will be prepared by following the protocol:

QIAseq 16S/ITS Region Panels

QIAseq 16S/ITS Panel Handbook

QIAseq 16S/ITS Panels are used to perform a robust profiling of bacterial and fungal communities. The panels have been developed for sequencing 16S rRNA and ITS regions on Illumina platforms. QIAseq 16S/ITS Panels use "phased primers" to increase the quality of reads and base calling, and eliminate the need for PhiX spike-in. In addition, QIAseq 16S/ITS Panels incorporate low-bioburden reagents to decrease background contamination.

QIAseq 16S/ITS Panels can be configured to target different 16S variable regions and/or ITS according to different pools of primers,

Libraries sequencing and NGS coverage: Libraries will be sequenced using either the MiSeq or the NextSeq 2000 system (Illumina) by generating 300 bp Paired End reads.

The sequencing throughput per sample will be determined considering the type and complexity of the microbiome under study.

Within this call the National Facility for Genomics will provide:

	Min samples/project	**Max samples/project**
G-003-A Microbiome analysis 16S-ITS (commercially available Region panels)	100	1200

Technical requirements

- DNA samples should be provided in 96 full-skirted PCR plates.
- 20 µl of High-quality DNA having a concentration ranging from 0,1-1 ng/ul should be provided.
- Quality of DNA should be evaluated by Agilent Tape Station/Bioanalyzer, samples should have a DIN≥6 (DNA Integrity Number≥6).
- Purity of the DNA samples should be assessed with a Spectrophotometer (i.e. Nanodrop; 260/280≥1,8 and 260/230≥1,8).
- DNA samples should be quantified by using a fluorometer (i.e Qubit/ Glomax).

Results that will be delivered to the users:

The National Facility for Genomics will deliver to the users the following files for every sample sequenced:

- FASTQ files
- QC report

For the users requesting data analysis as a combined service, the **National Facility for Data Handling and Analysis** will also provide the files described in the SID: NF62.02.03

Access modality available: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: Please select: SID: NF62.02.03.

SID: G-004 – Methylation sequencing (Methyl-seq)

Service description:

Methylation sequencing (Methyl-seq) is a technique used to study DNA methylation, a key epigenetic modification where methyl groups are added to DNA molecules, typically at cytosine bases in CpG sites. This method allows researchers to identify and quantify methylation patterns across the genome, providing insights into gene regulation, cellular differentiation, development, and disease processes such as cancer, where abnormal methylation often occurs.

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 (see technical requirements below).

Library Preparation: The extracted DNA is then fragmented into smaller, manageable pieces. Adapters are added to these fragments to allow the subsequent sequencing process. This is followed by two sets of enzymatic conversion steps to differentiate unmethylated cytosines from 5mC/5hmC.

Methylome capture: This step involves selectively capturing and enriching the DNA fragments corresponding to the regions rich in CpG sites. This is typically done using target-specific probes or baits designed to hybridize with and capture CpG regions.

Sequencing: The enriched Methyl-seq libraries are subjected to high-throughput sequencing technologies, such as next-generation sequencing (NGS) platforms. These technologies generate short DNA sequences called “reads” from the DNA fragments.

Pre-processing data analysis: QC and summary statistics.

Data Analysis: The generated sequencing data can be analysed bioinformatically to generate genome-wide methylation profiles, identifying methylated and unmethylated regions of DNA with base-pair resolution (users are responsible for data analysis).

Library preparation protocol:

The National Facility for Genomics uses the Twist NGS Methylation Detection System, which includes the NEBNext Enzymatic Methyl-seq Library Preparation Protocol together with the Twist Targeted Methylation Sequencing Protocol. The NEBNext Enzymatic Methyl-seq Library Preparation Protocol converts methylated genomic DNA into double-stranded, adapter-ligated DNA libraries. It uses a two-step enzymatic conversion process to distinguish between unmethylated and methylated cytosines: TET2 oxidizes 5-methylcytosine (5mC) and 5-hydromethylcytosine (5hmC) sites in the first step, protecting them from enzymatic deamination by APOBEC in the second step. Because unmethylated cytosines are converted to thymines, sequenced cytosines in the resulting library represent 5mC or 5hmC sites. In the

following Twist Targeted Methylation Sequencing Protocol, the prepared libraries are enriched for biologically relevant methylation markers in a hybrid capture step with the Twist Human Methylome Panel, which targets 3.98M CpG sites through 123 Mb of genomic content.

Twist NGS Methylation Detection System

Twist Human Methylome Panel

Libraries sequencing and NGS coverage:

Libraries will be sequenced using the NovaSeq 6000 system (Illumina) by generating 100 bp Paired End reads. On average, 70 million reads pairs per sample will be generated corresponding to a raw coverage of 100X.

Within this call the National Facility for Genomics will provide:

	Min samples/project	**Max samples/project**
G-004-A Methyl-seq for population studies on humans	200	800
G-004-B Methyl-seq for studies on diseases (humans, organoids derived from human cells)	16	200

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 at least 60 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)

Access modality available: Access to NF services (users are responsible for data analysis).

SID: G-005 - mRNA sequencing from standard and low input

Services description:

mRNA sequencing is a powerful molecular biology technique used to analyse the transcriptome of a biological sample. The transcriptome refers to the complete set of RNA molecules, in particular messenger RNA (mRNA), in a cell or tissue. mRNA sequencing is widely used in genomics research, functional genomics, and clinical studies to understand gene expression patterns, identify novel transcripts, and investigate how gene expression varies under different conditions.

Here is a brief description of the mRNA sequencing process:

Isolation of RNA: The first step involves extracting RNA from the biological sample, such as cells or tissues. This can be done using various methods to ensure the preservation of the RNA molecules. This task must be undertaken by the Users (see technical requirements below).

cDNA synthesis: Complementary DNA (cDNA) is synthesized from the fragmented RNA using reverse transcription. This step converts RNA into a complementary DNA strand, creating a library of cDNA molecules.

Library preparation: The cDNA library is then prepared for sequencing. Adapters are added to the cDNA fragments, allowing them to be sequenced efficiently.

Sequencing: The prepared library is subjected to high-throughput sequencing techniques, such as next-generation sequencing (NGS). This step generates short sequences, or reads, from the cDNA fragments.

QC and summary statistics

Bioinformatic analysis of mRNAseq data can be provided as a combined service by the National Facility for Data Handling and Analysis. Please select: SID: NF62.01.01.

Library preparation protocol:

Standard input libraries will be prepared by following the protocol:

Illumina Stranded mRNA - Standard Input

Illumina Stranded mRNA Prep enables precise measurement of strand orientation, uniform coverage, and high-confidence discovery of features such as novel isoforms, gene fusions, and allele-specific expression. Illumina Stranded mRNA Prep is optimized to provide good polyA capture efficiency and coverage uniformity. It minimizes the required sequencing depth for accurate, unbiased detection of the coding transcriptome.

Low input libraries will be prepared by following the protocol:

SMART-Seq v4 PLUS Kit User Manual

This kit uses oligo(dT) priming to generate high-quality, full-length cDNA directly from multiple intact cells or total RNA. In addition to the cDNA synthesis kit, the SSv4 PLUS kit also includes a library preparation kit and a single-use unique dual index (UDI) plate to generate Illumina-compatible sequencing libraries.

The SMART (Switching Mechanism at 5' End of RNA Template) technology employed by the SMART-Seq v4 kit provides full-length transcript information, enabling analysis of transcript

isoforms, gene fusions, point mutations, etc. Additionally, it incorporates locked nucleic acid (LNA) technology in the SMART-Seq v4 Oligo for more efficient template switching. This allows for the identification of higher numbers of genes relative to other methods, high reproducibility, even gene-body coverage, and an accurate representation of GC-rich transcripts.

Libraries sequencing and NGS coverage: Libraries will be sequenced using either the the NextSeq 2000 system or the NovaSeq 6000 system (Illumina) by generating 100 bp Paired End reads.

On average 40 million reads pairs (20 million clusters 100bp PE) per sample will be generated for species with reference genomes and 80 million reads pairs (40 million clusters 100bp PE) will be generated for species without reference genomes.

Within this call the National Facility for Genomics will provide:

	Min samples/project	**Max samples/project**
G-005-A mRNAseq for studies on diseases (humans, animal models, organoids)	50	300
G-005-B mRNAseq for differentiation studies (humans, animal models, organoids)	50	200
G-005-C mRNAseq in plants studies	50	100

Technical requirements

mRNA sequencing from standard input:

- totalRNA samples should be provided in LoBind full skirted PCR plate (i.e. Eppendorf twin.tec PCR plates) properly sealed using peelable adhesive PCR films.
- Samples should be ordered column-wise without empty wells.
- The totalRNA amount should be between 500ng-1000ng in a volume range of 20-50ul in nuclease-free ultrapure water.
- totalRNA should be DNase treated and the RIN≥7 (RNA Integrity Number≥7), the quality of totalRNAs should be evaluated by Agilent Tape Station/Bioanalyzer.
- Purity of the totalRNAs should be assessed with a Spectrophotometer (i.e. Nanodrop; 260/280≥1,8 and 260/230≥1,8).
- totalRNA samples should be quantified by using a fluorometer (i.e Qubit/Glomax).

mRNA sequencing from low input:

- RNA samples should be provided in LoBind full skirted PCR plate (i.e. Eppendorf twin.tec PCR plates) properly sealed using peelable adhesive PCR films. Samples should be ordered column-wise without empty wells.

- The RNA amount should be between 50pg-10ng in a volume range of 15-20 ul in nuclease-free ultrapure water.
- Total RNA should be DNase treated and the RIN≥7 (RNA Integrity Number≥7),
- Quality of totalRNAs should be evaluated by Agilent Tape Station/Bioanalyzer.
- Purity of the totalRNAs should be assessed with a Spectrophotometer (i.e. Nanodrop; 260/280≥1,8 and 260/230≥1,8).
- totalRNA samples should be quantified by using a fluorometer (i.e Qubit/Glomax).

Results that will be delivered to the users:

The **National Facility for Genomics** will deliver to the users the following files for every sample sequenced:

- FASTQ files
- QC report

For the users requesting data analysis as a combined service, the **National Facility for Data Handling and Analysis** will also provide the files described in the SID: NF62.01.01.

Access modality available: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: Please select: SID: NF62.01.01.

SID: G-006 - totalRNA sequencing from standard and low input.

Services description:

Total RNA sequencing is a powerful and widely used molecular biology technique that aims to analyse and quantify the entire transcriptome of a biological sample. The transcriptome represents the complete set of RNA molecules, including messenger RNA (mRNA), and non-coding RNAs, present in a cell or tissue. Total RNA sequencing provides a comprehensive view of the transcriptome, allowing researchers to gain insights into gene expression patterns, identify novel transcripts, and understand the regulatory mechanisms underlying various biological processes.

Here is a step-by-step description of the Total RNA sequencing process:

RNA Extraction: Total RNA is isolated from the biological sample of interest, such as cells or tissues. This extraction process is crucial to obtain a representative snapshot of the RNA present in the sample. This task must be undertaken by the Users (see technical requirements below).

Library Preparation: After ribosomal RNA depletion, the RNA is converted into a complementary DNA (cDNA) library through a process called reverse transcription. This step involves the use of reverse transcriptase to synthesize cDNA from the RNA template.

Fragmentation and Adaptor Ligation: The cDNA is then fragmented, and sequencing adaptors are added at the ends of the fragments. Adaptors contain sequences necessary for the subsequent steps of the sequencing process.

Library Amplification: The prepared library is amplified using polymerase chain reaction (PCR) to generate sufficient material for sequencing.

Sequencing: The amplified cDNA library is then subjected to high-throughput sequencing, with NGS sequencing platforms.

QC and summary statistics

Bioinformatic analysis of totalRNAseq data can be provided as a combined service by the National Facility for Data Handling and Analysis. Please select SID: NF62.01.01.

Library preparation protocol:

Libraries will be prepared from standard input by following the protocol:

Stranded Total RNA Prep, Ligation kit with Ribo-Zero Plus or Ribo-Zero Plus Microbiome

Illumina Total RNA Prep with Ribo-Zero Plus supports a broad range of RNA inputs. It's compatible with various sample types, including formalin-fixed paraffin-embedded (FFPE) and other low-quality samples. The included Ribo-Zero Plus or Ribo-Zero Plus Microbiome removes abundant RNA from multiple species, including human, mouse, rat, bacteria, and epidemiology samples or complex microbial samples, including stool samples, for meta-transcriptomic studies.

Libraries will be prepared from low input by following the protocol:

SMART-Seq® Total RNA Single Cell (ZapR™ Mammalian) User Manual

SMART-Seq Total RNA-Seq Single Cell (ZapR Mammalian) generates strand-specific RNA-seq libraries for Illumina sequencing from ultra low input of purified total RNA. The kit was specifically designed to deliver highly sensitive and reproducible data from ultra low input of total RNA while keeping the workflow short and user friendly. The kit does not require additional rRNA removal methods or kits and produces sequencing libraries that retain strand-of-origin information.

Libraries sequencing and NGS coverage:

Libraries will be sequenced using either the NextSeq 2000 system or the NovaSeq 6000 system (Illumina) by generating 100 bp Paired End reads.

On average 80 million reads pairs (40 million clusters 100bp PE) per sample will be generated for species with reference genomes and 160 million reads pairs (80 million clusters 100bp PE) will be generated for species without reference genomes or for meta-transcriptomics studies.

Within this call the National Facility for Genomics will provide:

	Min samples/project	**Max samples/project**
G-006-A totRNAseq for studies on diseases (humans, animal models, organoids)	50	200

G-006-B totRNAseq for differentiation studies (humans, animal models, organoids)	50	200
G-006-C totRNAseq in plants studies	20	100
G-006-D totRNAseq in infection studies (host/pathogen interaction, in vitro and ex-vivo)	20	100

Technical requirements

totalRNA sequencing from standard input:

- totalRNA samples should be provided in LoBind full skirted PCR plate (i.e. Eppendorf twin.tec PCR plates) properly sealed using peelable adhesive PCR films.
- Samples should be ordered column-wise without empty wells.
- The totalRNA amount should be between 500ng-1000ng in a volume range of 20-50ul in nuclease-free ultrapure water.
- totalRNA should be DNase treated and the RIN≥4 (RNA Integrity Number≥4)
- Quality of totalRNAs should be evaluated by Agilent Tape Station/Bioanalyzer.
- Purity of the totalRNAs should be assessed with a Spectrophotometer (i.e., Nanodrop; 260/280≥1,8 and 260/230≥1,8).
- totalRNA samples should be quantified by using a fluorometer (i.e Qubit/ Glomax).

totalRNA sequencing from low input:

- totalRNA samples should be provided in LoBind full skirted PCR plate (i.e. Eppendorf twin.tec PCR plates) properly sealed using peelable adhesive PCR films. Samples should be ordered column-wise without empty wells.
- The totalRNA amount should be between 50pg-10ng in a total volume of maximum 7ul in nuclease-free ultrapure water.
- totalRNA should be DNase treated and the RIN≥4 (RNA Integrity Number≥4), the quality of totalRNAs should be evaluated by Agilent Tape Station/Bioanalyzer.
- Purity of the totalRNAs should be assessed with a Spectrophotometer (i.e., Nanodrop; 260/280≥1,8 and 260/230≥1,8).
- totalRNA samples should be quantified by using a fluorometer (i.e Qubit/ Glomax).

Results that will be delivered to the users:

The National Facility for Genomics will deliver to the users the following files for every sample sequenced:

- FASTQ files
- QC report

For the users requesting data analysis as a combine service, the **National Facility for Data Handling and Analysis** will also provide the files described in the SID: NF62.01.01.

Access modality available: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: Please select SID: NF62.01.01.

SID: G-007 - smallRNA sequencing

Services description:

Small RNA sequencing is a specialized technique designed to analyse and profile small RNA molecules present in a biological sample. Small RNAs are short RNA molecules, typically ranging from 18 to 30 nucleotides in length, and they play essential roles in various cellular processes, including gene regulation, RNA silencing, and post-transcriptional control. Small RNA sequencing is widely used to study the expression profiles of miRNAs and other small RNAs, providing valuable insights into their roles in gene regulation, development, and disease.

Here is a step-by-step description of small RNA sequencing:

RNA Extraction: Like total RNA sequencing, the process begins with the extraction of RNA from the biological sample. However, in small RNA sequencing, specific methods can be employed to selectively enrich small RNA molecules. This task must be undertaken by the Users (see technical requirements below).

Size Selection: The extracted RNA can be subjected to size selection to isolate the small RNA fraction, but also totalRNA can be used as input.

Adapter Ligation: After smRNAs poly-adenilation and retro transcription adapters are ligated to both ends of resulting cDNAs and serve as primers during the library preparation steps.

Library Amplification: The cDNA library is then amplified using PCR. This step adds the necessary sequences for subsequent sequencing steps

Size Selection and Purification: The amplified library undergoes a size selection step to remove unwanted fragments and purify the small RNA-containing fraction.

Sequencing: The purified small RNA library is subjected to high-throughput sequencing, using NGS platforms.

QC and summary statistics

Bioinformatic analysis of miRNA data can be provided as a combined service by the National Facility for Data Handling and Analysis. Please select SID: NF62.01.02.

Data Analysis: small RNA sequencing allows to identify and quantify different types of small RNAs, such as microRNAs (miRNAs), small interfering RNAs (siRNAs), and piwi-interacting RNAs (piRNAs). This involves aligning the sequences to a properly annotated reference genome or small RNA databases. Currently the **NF for Data Handling and Analysis** can provide bioinformatic support for the analysis of miRNAs (20-25 nucleotides) only.

Library preparation protocol:

Libraries will be prepared by following the protocol:

SMARTer smRNA-Seq Kit

The SMARTer smRNA-Seq Kit is used to generate small RNA-seq libraries for sequencing on Illumina platforms. This kit works directly with total RNA or enriched small RNA (including microRNA). It incorporates features of the SMARTer Stranded RNA-Seq kits, including the SMART (Switching Mechanism at the 5' end of RNA Template) technology, and primers that include locked nucleic acids (LNAs). This kit allows users to analyze a wide range of smRNA species and generate sequencing libraries of considerable complexity. Illumina adapters and index sequences are incorporated in a ligation-free manner during library amplification, ensuring that diverse smRNA species are represented with minimal bias.

Libraries sequencing and NGS coverage:

Libraries will be sequenced using either the NextSeq 2000 system or the NovaSeq 6000 system (Illumina) by generating 100 bp Paired End reads.

On average 20 million reads pairs (10 million clusters 50bp PE) per sample will be generated.

Within this call the National Facility for Genomics will provide:

	Min samples/project	**Max samples/project**
G-007-A smallRNAseq in humans, animal models, organoids samples	20	60

Technical requirements

- totalRNA samples should be provided in LoBind full skirted PCR plate (i.e. Eppendorf twin.tec PCR plates) properly sealed using peelable adhesive PCR films.
- Samples should be ordered column-wise.
- The totalRNA amount should be between 100ng-1000ng in a volume range of 20-50ul in nuclease-free ultrapure water.
- RNA should be DNase treated and the RIN≥7(RNA Integrity Number≥7)
- Quality of totalRNAs should be evaluated by Agilent Tape Station/Bioanalyzer.
- Purity of the totalRNAs should be assessed with a Spectrophotometer (i.e. Nanodrop; 260/280≥1,8 and 260/230≥1,8).
- totalRNA samples should be quantified by using a fluorometer (i.e Qubit/ Glomax).

Results that will be delivered to the users

The National Facility for Genomics will deliver to the users the following files for every sample sequenced:

- FASTQ files
- QC report
- Mapping metrics

For the users requesting miRNA data analysis as a combined service, the **National Facility for Data Handling and Analysis** will also provide the files described in the SID: NF62.01.02.

Access modality available: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: (Please select SID: NF62.01.02).

SID: G-008/G-012 - Single-cell 3'RNAsequencing or Single-cell gene Expression Flex (10X Genomics)

Services description:

Single-cell 3'RNAsequencing

Single-cell 3' RNA sequencing with 10x Genomics technology is a powerful method for studying gene expression at the single-cell level. It enables the profiling of thousands to tens of thousands of single cells in parallel. It captures heterogeneity within cell populations, allowing the identification of rare cell types and subpopulations. It reveals differences in gene expression between individual cells, providing a more nuanced understanding of cellular diversity.

Here's an overview of the process:

Cell Capture: Single cells are isolated into individual droplets using microfluidics technology. Each droplet contains a bead with a unique barcode.

Barcoding and cDNA Synthesis: Within each droplet, the cell's RNA is captured at the 3' end and is barcoded using both a barcode associated to each cell and unique molecular identifiers (UMIs) that will allow transcripts counting and normalization. This step ensures that each transcript originating from the same cell gets the same cell barcode. cDNA (complementary DNA) is synthesized from the barcoded RNA templates

Library Preparation: The cDNA is then amplified, and Illumina sequencing adapters are added, preparing the library for sequencing.

Pooling and Sequencing: All the individually barcoded and amplified cDNAs from the different cells are pooled together. The pooled library is sequenced on a NGS platform.

QC and summary statistics

Bioinformatic analysis of scRNA data can be provided as a combined service by the National Facility for Data Handling and Analysis. Please select SID: NF62.03.01.

Single-cell gene Expression Flex:

The Fixed RNA Profiling assay, also called the Single Cell Gene Expression Flex assay by 10X Genomics, is a way to prepare single cell RNA-seq libraries from formaldehyde-fixed cells and tissues as well as FFPE tissue blocks (please consider that from FFPE tissues the isolated population primarily consists of nuclei suspension, thus the RNA-seq analysis is to be intended at nuclear level). This process allows researchers to lock in the biological state of their samples at the time of fixation and store the fixed cell suspension for at least 6 months at -80°C.

Because the fixation process is expected to lead to a certain amount of RNA degradation, the fixed RNA assay uses pairs of sequence-specific probes to bind transcripts for approximately 18,000 genes for human and about 19.000 genes for mouse. These probe pairs bind adjacent sites on their target transcripts, are ligated together, and then become the substrate for library construction.

Sample multiplexing allows users to label discrete cell populations with a sample-specific barcode so that these samples can be pooled into a single GEM droplet reaction, reducing the number of reagent units needed for the experiment and potentially lowering the overall cost. Multiplexing in the Flex assay is built in by the inclusion of sample barcodes in the transcript-detecting probes. The 64-reaction multiplex kit, that will be used by the National Facility for Genomics to provide this service, includes four units of gel beads and 16 sets of probes with unique barcodes, allowing to pool up to 16 samples and to recover and analyse potentially 320.000 cells per GEM reaction.

This type of analysis can be performed only using the Chromium X platform.

Library preparation protocols:

In the Single-cell 3' RNA sequencing approach depending on the number of cells the user wishes to analyse; the samples will be processed using the Chromium X platform which allows to recover and analyse a maximum of 20.000 cells per single sample.

Here below the list of protocols that will be used by the National Facility staff to perform Single-cell 3'RNAsequencing or Single-cell gene Expression Flex experiments:

Chromium GEM-X Single Cell 3' Reagent Kits v4 - CG000731

GEM-X Flex Gene Expression Reagent Kits - CG000786

Chromium Fixed RNA Profiling Reagent Kits for Multiplexed Samples User Guide and Pooling CG000527/CG000565

Libraries sequencing and NGS coverage:

Libraries will be sequenced using either the NextSeq 2000 system or the NovaSeq 6000 system (Illumina) by generating reads length as required by the Single-cell 3'RNAsequencing or the Single-cell gene Expression Flex protocol (10x Genomics).

On average, 50.000 reads per cell will be generated for sequencing GEX libraries prepared using the Single-Cell 3' RNA Sequencing Protocol, and approximately 20.000 reads per cell for GEX libraries prepared using the Single-Cell Gene Expression Flex Protocol.

Within this call the National Facility for Genomics will provide:

	Min samples/project	**Max samples/project**
G-008/G-012 - A Single-cell 3'RNAsequencing for studies on diseases (humans, animal models, organoids)	8	32
G-008/G-012 - B Single-cell 3'RNAsequencing for differentiation studies (humans, animal models, organoids)	8	32

G-008/G-012- C Single-cell gene Expression Flex (humans, animal models, organoids) (16-plex)	32	64

Technical requirements

Single-cell 3'RNAseq: Frozen Cells should be provided in low bind tubes (i.e., 1,5 ml Eppendorf tubes) providing an input of a minimum of 500k frozen cells per sample.

Cell Preparation for Single Cell Protocols - CG00053

Cell viability before freezing should be ≥80%. Only samples that will have a cell viability ≥70% and a low level of cellular debris after thawing will be processed.

RNA Flex: Fixed cells should be provided in low bind tubes (i.e. 1,5 ml Eppendorf tubes). Cells fixation should be performed following the protocols listed below:

Blood Fixation and Cell Isolation for Chromium Fixed RNA Profiling - CG000721

Fixation of Cells & Nuclei for GEM-X Flex Gene Expression - CG000782

Tissue Fixation & Dissociation for GEM-X Flex Gene Expression - CG000783

Sample Preparation from FFPE Tissue Sections for GEM-X Flex Gene Expression - CG000784

Cell viability before fixation should be ≥80%. Two aliquots containing 25k to 500k fixed cells are required per sample (recommended to default to 300,000 fixed cells/nuclei)

Results that will be delivered to the users:

The National Facility for Genomics will deliver to the users the following files for every sample sequenced:

- FASTQ files
- QC report
- Mapping metrics

For the users requesting data analysis as a combined service, the **National Facility for Data Handling and Analysis** will also provide the files described in the SID: NF62.03.01

Access modality available: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: Please select SID: NF62.03.01.

SID: G-009/012/014 – Single-cell Immune profiling-V(D)J (10X Genomics)

Services description:

Single-cell immune profiling with 10x Genomics technology is a powerful method for 5' RNA sequencing at the single-cell level and to profile at the same time the T-cells and/or B-cells receptors at single cell level by sequencing the V(D)J regions.

Here is an overview of the process:

Cell Capture: Single cells are isolated into individual droplets using microfluidics technology. Each droplet contains a bead with a unique barcode.

Barcoding and cDNA Synthesis: Within each droplet, the cell's RNA is captured at the 5' end and is barcoded using both a barcode associated to each cell and unique molecular identifiers (UMIs) that will allow transcripts counting and normalization. This step ensures that each transcript originating from the same cell gets the same cell barcode. cDNA (complementary DNA) is synthesized from the barcoded RNA templates.

Library Preparation: Three libraries can be potentially obtained from the cDNA synthesized from the barcoded RNA templates. The GEX library, for Gene Expression profiling, resulting from cDNA amplification and addition of Illumina sequencing adapters. Libraries for V(D)J Gene Profiling deriving from targeted cDNA amplification of the V(D)J regions specific of T cell receptors (TCRs) and B cell receptors (BCRs).

Sequencing: The pooled libraries are sequenced on a high-throughput NGS platform.

QC and summary statistic

Bioinformatic analysis of single cell Immune Profiling (VDJ) data can be provided as a combined service by the National Facility for Data Handling and Analysis. Please select SID: NF62.03.03.

Library preparation protocol:

Depending on the number of cells the user wishes to analyse, the samples will be processed using the Chromium X platform which allows to recover and analyse a maximum of 20,000 cells per single sample.

Libraries will be prepared by following the protocols:

Chromium GEM-X Single Cell 5'-V(D)J Reagent Kits v3 - CG000733

Libraries sequencing and NGS coverage:

Libraries will be sequenced using either the NextSeq 2000 system or the NovaSeq 6000 system (Illumina) by generating reads length as required by the Single-cell Immune profiling-V(D)J protocol (10x Genomics)

On average 50.000 reads per cell will be generated for sequencing GEX libraries and at least 5.000 reads per cell for sequencing TCR and BCR libraries obtained with the Single-cell Immune profiling-V(D)J protocol.

Within this call the National Facility for Genomics will provide:

	Min samples/project	**Max samples/project**
G-009/012/014 - A Single-cell Immune profiling-VDJ for studies on diseases (humans, animal models)	16	64

Technical requirements

- Frozen Cells should be provided in low bind tubes (i.e. 1,5 ml Eppendorf tubes).
- Provide an input of a minimum of 500k frozen cells per sample.
- Cells should be prepared following the protocol:

  Cell Preparation for Single Cell Protocols - CG00053
- Cell viability before freezing should be ≥80%. Only samples that will have a cell viability ≥70% and a low level of cellular debris after thawing will be processed.

Results that will be delivered to the users:

The National Facility for Genomics will deliver to the users the following files for every sample sequenced:

- FASTQ files
- QC report
- Mapping metrics

For the users requesting data analysis as a combined service, the **National Facility for Data Handling and Analysis** will also provide the files described in the SID: NF62.03.03

Access modality available: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: Please select SID: NF62.03.03.

SID: G-010/012/013 – Single-cell multiome ATAC + Gene expression (10X Genomics)

Services description:

Single-cell multiome ATAC + Gene Expression is a cutting-edge technology that enables the simultaneous profiling of chromatin accessibility and gene expression at the single-cell level, providing a comprehensive view of the molecular landscape within individual cells. This technology is widely used in various biological research areas, including understanding cellular diversity in tissues, identifying cell types and states, deciphering regulatory networks, and gaining insights into how chromatin accessibility relates to gene expression at the single-cell level.

Here's an overview of the key steps:

Single-Cell Resolution and Multiome analysis: The technology captures information at the level of individual cells/nuclei, combining two crucial molecular profiles – ATAC-seq (Assay for Transposase-Accessible Chromatin with high-throughput sequencing) for chromatin accessibility and RNA-seq (or gene expression profiling) for understanding the transcriptional activity of each cell/nucleus.

ATAC-seq: This part of the technology focuses on assessing the accessibility of chromatin, providing insights into the regions of the genome that are open and accessible for transcription factors and other regulatory elements.

RNA-seq: Concurrently, the gene expression analysis captures the messenger RNA transcripts present in each nucleus, shedding light on the active genes and their expression levels.

The experimental workflow for Single-cell multiome ATAC + Gene Expression using the 10X Genomics platform typically involves several key steps:

Cell Isolation and Nuclei Preparation: Begin by isolating the cells of interest. This could be from tissues or cell cultures. Prepare a single-cell suspension to ensure individual cells can be processed independently (this step must be taken by users). Extract the nuclei from the cells, as the chromatin accessibility information is primarily derived from the open regions in the nuclei.

Transposase Reaction (ATAC-seq): Add a transposase enzyme to the nuclei suspension. The transposase fragments the chromatin and adds sequencing adapters simultaneously. These sequencing adapters are essential for downstream library preparation.

Chromium Gel Bead-in-Emulsion (GEM) Formation: The isolated nuclei are encapsulated into Gel Bead-in-Emulsion (GEM) droplets using the 10X Genomics Chromium platforms. Each GEM contains a unique barcode that will be associated with both the chromatin accessibility and gene expression data from a specific nucleus.

Reverse Transcription: In the same GEMs, reverse transcription is performed to convert nuclear RNA into complementary DNA (cDNA).

Library Preparation: Following the transposase reaction and reverse transcription, PCR amplification is performed to create the final sequencing libraries (ATAC and GEX libraries). Sequencing adapters are added during this step to allow sequencing on NGS platforms.

High-Throughput Sequencing: Sequence the prepared libraries using high-throughput sequencing platforms. The sequencing reads contain information about both chromatin accessibility and gene expression in individual cells/nuclei.

QC and summary statistics

Bioinformatic analysis of sc multiome ATAC + GEX data can be provided as a combined service by the National Facility for Data Handling and Analysis. Please select SID: NF62.03.04.

Library preparation protocol:

Libraries will be prepared by following the protocol:

Chromium Next GEM Single Cell Multiome ATAC + Gene Expression - CG000338

Libraries sequencing and NGS coverage:

Libraries will be sequenced using either the NextSeq 2000 system or the NovaSeq 6000 system (Illumina) by generating reads length as required by the single-cell multiome ATAC + Gene expression protocol (10x Genomics).

On average 50.000 reads per cell/nucleus will be generated for sequencing GEX libraries and 25.000 reads per cell/nucleus for sequencing ATACseq libraries obtained with the single-cell multiome ATAC + Gene expression protocol.

Within this call the National Facility for Genomics will provide:

	Min samples/project	**Max samples/project**
G-010/012/013 - A Single-cell multiome ATAC + Gene expression for studies on diseases (humans, animal models, organoids)	8	16
G-010/012/013 - B Single-cell multiome ATAC + Gene expression for differentiation studies (humans, animal models, organoids)	8	16

Technical requirements

- Frozen cells should be provided in low bind tubes (i.e. 1,5 ml Eppendorf tubes). Provide an input of a minimum of 500k frozen cells. Cells should be prepared following the protocol:

Cell Preparation for Single Cell Protocols - CG00053

- Cell viability before freezing should be ≥80%.
- Nuclei will be prepared by the National Facility staff following the protocol:

Nuclei Isolation for Single Cell Multiome ATAC + Gene Expression Sequencing - CG000365

Results that will be delivered to the users:

The National Facility for Genomics will deliver to the users the following files for every sample sequenced:

- FASTQ files
- QC report
- Mapping metrics

For the users requesting data analysis as a combined service, the **National Facility for Data Handling and Analysis** will also provide the files described in the SID: NF62.03.04.

Access modality available: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: Please select SID: NF62.03.04.

SID: G-011/013 - Single-cell ATAC sequencing (10X Genomics)

Services description:

Single-cell ATAC sequencing with the 10X Genomics platform involves profiling the chromatin accessibility of individual cells/nuclei at a high resolution. The assay, commonly known as scATAC-seq (Single-cell Assay for Transposase-Accessible Chromatin sequencing), utilizes the 10X Genomics Chromium system to barcode and index individual nuclei, it provides valuable insights into the epigenomic landscape of individual cells, allowing researchers to study cellular heterogeneity and regulatory processes at high resolution.

Here's a general overview of the protocol:

Cell Isolation and Nuclei Extraction: Start by isolating the cells of interest. This could be from tissues or cell cultures (this step must be taken by users). Extract the nuclei from the cells, as the chromatin accessibility information is primarily derived from the open regions in the nuclei.

Transposase Reaction: Add a transposase enzyme to the nuclei suspension. The transposase simultaneously fragments the chromatin and adds sequencing adapters. The sequencing adapters are important for later steps in library preparation.

Chromium Gel Bead-in-Emulsion (GEM) Formation: The isolated nuclei are then encapsulated into Gel Bead-in-Emulsion (GEM) droplets using the 10X Genomics Chromium platforms. Each GEM contains a unique barcode that will be associated with the chromatin accessibility data from a specific nucleus.

Library Preparation: Following the transposase reaction and nuclei encapsulation in GEM beads, PCR amplification is performed to create the final sequencing libraries. for accurate quantification and removal of PCR duplicates during data analysis.

High-Throughput Sequencing: Sequencing is performed on the prepared libraries using high-throughput sequencing platforms. The sequencing reads contain information about the open chromatin regions in individual cells/nuclei.

QC and summary statistics

Bioinformatic analysis of scATAC data can be provided as a combined service by the National Facility for Data Handling and Analysis. Please select SID: NF62.03.02.

Library preparation protocol:

Libraries will be prepared by following the protocol:

Chromium Next GEM Single Cell ATAC - CG000496

Libraries sequencing and NGS coverage:

Libraries will be sequenced using either the NextSeq 2000 system or the NovaSeq 6000 system (Illumina) by generating reads length as required by the Single-cell ATAC sequencing protocol (10x Genomics).

On average 25.000 reads per cell/nucleus will be generated for sequencing ATACseq libraries obtained with the single-cell multiome ATAC + Gene expression protocol.

Within this call the National Facility for Genomics will provide:

	Min samples/project	**Max samples/project**
G-011/013 -A Single-cell ATAC sequencing for pathologies studies (humans, animal models, organoids)	8	16
G-011/013 -B Single-cell ATAC sequencing for pathologies studies (humans, animal models, organoids)	8	16

Technical requirements

- Frozen cells should be provided in low bind tubes (i.e. 1,5 ml Eppendorf tubes). Provide an input of a minimum of 500k frozen cells. Cells should be prepared following the protocol:

Cell Preparation for Single Cell Protocols - CG00053

- Cell viability before freezing should be ≥80%.

- Nuclei will be prepared by the National Facility staff following the protocol:

Nuclei Isolation for Single Cell Multiome ATAC + Gene Expression Sequencing - CG000365

Chromium Nuclei Isolation Kit Sample Prep User Guide - CG000505

Results that will be delivered to the users:

The National Facility for Genomics will deliver to the users the following files for every sample sequenced:

- FASTQ files
- QC report
- Mapping metrics

For the users requesting data analysis as a combined service, the **National Facility for Data Handling and Analysis** will also provide the files described in the SID: NF62.03.02.

Access modality available: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: Please select SID: NF62.03.02.

SID: G-015/016 - Visium HD Spatial gene expression from Fresh-Frozen, Fixed Frozen or FFPE tissues (10X Genomics). Visium Fresh-Frozen (Direct placement)

Services description:

Visium HD Spatial Gene Expression (10X Genomics) and the Visium Spatial Gene Expression for Fresh Frozen (Direct placement) allow the spatial profiling of gene expression within intact tissue sections. The protocols allow for the analysis of gene expression while preserving the spatial context of cells within a tissue sample.

Here is a general overview of Visium HD Spatial Gene Expression protocols for Fresh-Frozen (FF), Formalin Fixed Paraffin Embedded (FFPE) and Fixed frozen (FxF) tissues and Visium Spatial Gene Expression for Fresh Frozen (Direct placement)

Visium HD Spatial Gene Expression:

Tissue Slide Preparation and Staining: Visium HD is compatible with H&E or IF-stained fresh frozen, fixed frozen, and FFPE tissue sections from human and mouse samples. Tissue slides are

generated via tissue section, deparaffinization (for FFPE tissue), staining, and imaging on a glass slide prior to using the Visium CytAssist

Probe hybridization: Pre-designed probe pairs are hybridized to the RNA target for highly specific and sensitive detection of the whole transcriptome (18536 genes targeted by 54018 probes in Human and 19405 genes targeted by 55538 probes in mouse)

CytAssist mounting: After probe hybridization, two standard glass slides and a two–Capture Area Visium CytAssist gene expression slide are placed in the CytAssist instrument so that the tissue sections on the standard slides can be aligned on top of the two Visium Capture Areas followed by permeabilization of the tissue and transfer of transcriptomic probes to the Visium slide.

Library Preparation: The probes are extended, the sample is eluted and transferred to a new tube to initiate the process of constructing a gene expression library

High-Throughput Sequencing: Sequence the prepared libraries using high-throughput NGS platforms.

QC and summary statistics

Bioinformatic analysis of Visium Spatial data can be provided as a combined service by the National Facility for Data Handling and Analysis. Please select SID: NF62.03.05.

Visium Spatial Gene Expression for Fresh Frozen (Direct placement):

Tissue Sectioning: Obtain a fresh-frozen tissue sample included in OCT. Section the tissue into thin slices, typically in the range of 10-20 µm, using a cryostat (this step must be taken by the users).

Tissue Mounting: Place the tissue sections onto a Visium Spatial Gene Expression slide (this step must be taken by the users, Visium slides will be provided by the facility). Ensure proper mounting and orientation to maintain the spatial information.

Tissue Fixation: Fix the tissue sections on the slide to preserve the spatial structure and prevent RNA degradation.

Permeabilization: The tissue sections are then permeabilized, and the mRNA molecules within cells are captured by the poly(dT) sequence on the slide surface

Reverse Transcription: Capture the spatially barcoded mRNA using reverse transcription, converting RNA into cDNA.

Library Preparation: Amplify and tag the cDNA with sample-specific barcodes during library preparation. Add sequencing adapters for subsequent high-throughput sequencing.

High-Throughput Sequencing: Sequence the prepared libraries using high-throughput NGS platforms.

QC and summary statistics

Bioinformatic analysis of Visium Spatial data can be provided as a combined service by the National Facility for Data Handling and Analysis. Please select SID: NF62.03.05.

Library preparation protocol:

Histological slides containing tissue sections derived from formalin fixed & paraffin embedded (FFPE), fresh frozen (FF), or fixed frozen (FxF) tissue samples will be processed by using the Visium CytAssist instrument. Libraries from FF, FxF and FFPE tissue will be prepared following the protocol:

Visium CytAssist Spatial Gene Expression Reagent Kits User Guide - CG000495

Visium HD Spatial Gene Expression - CG000685

Libraries sequencing and NGS coverage:

Libraries will be sequenced using either the NextSeq 2000 system or the NovaSeq 6000 system (Illumina) by generating reads length as required by the Visium Spatial gene expression from FF, FxF or FFPE tissues protocol (10x Genomics).

On average 50.000 reads per spot will be generated for sequencing libraries obtained with the Visium Spatial gene expression from FF (Direct placement) tissue protocol. The sequencing depth for Visium HD libraries will be calculated as indicated in the Visium HD Spatial Gene Expression - CG000685.

Within this call the National Facility for Genomics will provide:

	Min samples/project	Max samples/project
G-015/016 - A Visium HD Spatial gene expression from Fresh-Frozen, Fixed Frozen or FFPE tissues. Visium Fresh-Frozen (Direct placement) for studies on diseases (humans, animal models, organoids)	8	24
G-015/016 -B Visium HD Spatial gene expression from Fresh-Frozen, Fixed Frozen or FFPE tissues. Visium Fresh-Frozen (Direct placement) for differentiation studies (humans, animal models, organoids)	8	24

Technical requirements

For Fresh Frozen Visium:

Tissues should be prepared (frozen and embedded in OCT) according to the protocols:

Visium HD Fresh Frozen Tissue Preparation Handbook - CG000763

Section Thickness

Recommended section thickness for most tissue types is 10 µm, but tissues from 10−20 µm are compatible with the assay. Tissues with higher fat content (e.g., breast tissue) may require sections closer to 20 µm.

Before section placement, draw an outline of the allowable area on the back of the blank slide to ensure compatibility with the Visium CytAssist instrument and Tissue Slide Cassette.

Consult the: Visium CytAssist Accessory Kit Quick Reference Card - CG000548.

RNA quality assessment is recommended: For each sample/tissue the quality of RNA extracted from the tissue should be evaluated following the “RNA quality assessment section” of the same protocol and by using the Tape Station/Bioanalyzer.

Only samples having RIN≥4 will be processed for Visium HD approaches.

For Fresh Frozen Visium Direct Placement

Visium Fresh-Frozen (Direct placement) Tissues, previously frozen and embedded in OCT according to the protocol:

Visium Spatial Protocols – Tissue Preparation Guide - CG000240

should be placed by the User on the Visium slides/capture areas, that will be provided by the NF for Genomics. For correctly placing the tissue slices on the Visium slides/areas and storing them please follow the same protocol (section 2.3). For each sample/tissue the quality of RNA extracted from the tissue should be evaluated following the "RNA quality assessment section" of the same protocol and by using the Tape Station/Bioanalyzer. Only samples having a RIN≥7 will be processed.

For FFPE Visium: Tissues should be prepared according to the protocols:

Visium HD FFPE Tissue Preparation Handbook - CG000684

Section Thickness

Recommended section thickness is 3–10 µm. Most sections were cut at 5 µm

Before section placement, draw an outline of the allowable area on the back of the blank slide to ensure compatibility with the Visium CytAssist instrument and Tissue Slide Cassette.

Consult the: Visium CytAssist Accessory Kit Quick Reference Card - CG000548.

RNA quality assessment is recommended: For each sample/tissue the quality of RNA extracted from the tissue should be evaluated following the "RNA quality assessment section" of the same protocol and by using the Tape Station/Bioanalyzer.

Only samples having a DV200≥30% will be processed.

For Fixed Frozen Visium HD: Tissues should be prepared according to the protocols:

Visium HD Fixed Frozen Tissue Preparation Handbook - CG000764

Section Thickness

Recommended section thickness for most tissue types is 10 µm, but tissues from 10−20 µm are compatible with the assay.

Sections outside of that range may result in reduced performance.

Before section placement, draw an outline of the allowable area on the back of the blank slide to ensure compatibility with the Visium CytAssist instrument and Tissue Slide Cassette.

Consult the: Visium CytAssist Accessory Kit Quick Reference Card - CG000548.

RNA quality assessment is recommended: For each sample/tissue the quality of RNA extracted from the tissue should be evaluated following the "RNA quality assessment section" of the same protocol and by using the Tape Station/Bioanalyzer.

Only samples having a DV200≥50% will be processed.

Results that will be delivered to the users:

The National Facility for Genomics will deliver to the users the following files for every sample sequenced:

- FASTQ files
- QC report
- Mapping metrics

For the users requesting data analysis as a combined service, the **National Facility for Data Handling and Analysis** will also provide the files described in the SID: NF62.03.05.

Access modality available: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: Please select SID: NF62.03.05.

SID: G017/018 – GeoMx Digital Spatial Profiling from Fresh-Frozen, Fixed Frozen or FFPE tissues (Nanostring)

Service description:

The GeoMx Digital Spatial Profiling service provides cutting-edge technology for spatially resolved gene expression and protein analysis. This service supports fresh-frozen (FF), fixed-frozen and formalin-fixed, paraffin-embedded (FFPE) tissue samples, enabling researchers to explore spatial biology in diverse sample types.

Tissue Preparation and Sectioning:

Fresh Frozen tissues or Fixed Frozen tissues should be embedded in OCT compound before sectioning. Sections should be cut at 5–10 µm thickness on a calibrated cryostat and mounted immediately on the slide, without scratches or folds (this step must be taken by the users).

FFPE tissues should be cut at 5 µm thickness on a calibrated microtome and mounted on the slide immediately, without scratches or folds (this step must be taken by the users).

Tissue Mounting Tissue sections must be placed onto positively charged slides inside the correct area (this step must be taken by the users).

Tissue Fixation for Fresh Frozen tissues: Tissue sections must be fixed on the slide to preserve the spatial structure and prevent RNA degradation (this step will be taken by the Tissue Processing IU2 of the Imaging National Facility).

Tissue Deparaffinization and Rehydration for FFPE tissues: Paraffin is removed from the FFPE tissue sections using a deparaffinization process. This step is crucial to expose the RNA for downstream processing, and rehydration of the tissue sections to restore their natural state.

Antigen Retrieval: Antigen retrieval is performed to improve the accessibility of RNA for subsequent steps. This step is critical for fixed tissues, where formalin fixation can crosslink and modify nucleic acids.

Permeabilization: The fixed tissue sections are permeabilized to allow the probes to access the cellular RNA.

Tissue Hybridization and Staining: Tissue sections are hybridized with the proper probe panel Whole Transcriptome Atlas (WTA) overnight then staining is performed with Morphology markers (kit contains a nuclear stain and two fluorescently labelled antibodies against specific biological targets PanCK and CD45).

GeoMx Run and Regions of Interest (ROIs) Selection: After loading the slide into the GeoMx system tissue's images are obtained for the selection and acquisition of the ROIs. ROIs can be segmented into discrete compartments or areas of illumination (AOI). The minimum ROI size tested at NanoString is 32µm. To reach the limit of detection in RNA analysis it is recommended to capture at least 200 cells (this step must be taken by the users connecting from remote to the instrument).

ROIs collection: Oligos from selected ROIs are collected into a 96 well plate up to a 96 AOI (Areas of interest).

Library Preparation: Amplify and tag the oligos with ROI-specific barcodes during library preparation, adding sequencing adapters for subsequent high-throughput sequencing.

High-Throughput Sequencing: Sequence the prepared libraries using high-throughput NGS platforms.

Data Analysis: Use specialized bioinformatics tools to analyze the sequencing data. Align reads to the reference genome and quantify gene expression while retaining spatial information.

Spatial Mapping: Map the gene expression data back to the spatial locations on the tissue section using the spatial barcodes and dedicated software (users are responsible for data analysis).

Within this call the National Facility for Genomics will provide:

	Min tissue slides /project	**Max tissue slides /project**
G017/018 -A GeoMx Digital Spatial Profiling from Fresh-Frozen, Fixed Frozen or FFPE tissues for	6	40

studies on diseases (humans, animal models, organoids)		

Technical requirements

For Fresh Frozen and Fixed Frozen tissues:

Tissues should be prepared according to the guidelines listed in the protocol:

GeoMX DSP Manual Slide Preparation

Tissue sections must be placed onto positively charged slides (NanoString recommends SuperFrost™ Plus slides (for manual slide preparation) or Apex BOND slides) inside the correct area (this step must be taken by the users).

Tissue sections must be placed in the Scan Area in the center of the slide and be no larger than 35.3 mm long by 14.1 mm wide. If mounting multiple sections per slide, ensure that tissues are at least 2–3 mm apart and still fit within the Scan Area. Ensure proper mounting and orientation to maintain the spatial information. Slides can be stored at -80°C for several weeks before use.

For FFPE tissues:

Tissue should be prepared according to the guidelines listed in the protocol:

GeoMX DSP Manual Slide Preparation

Place the tissue sections onto positively charged slides (*NanoString recommends SuperFrost™ Plus slides)* inside the correct area (this step must be taken by the users). Tissue sections must be placed in the Scan Area in the center of the slide and be no larger than 35.3 mm long by 14.1 mm wide. If mounting multiple sections per slide, ensure that tissues are at least 2–3 mm apart and still fit within the Scan Area. Ensure proper mounting and orientation to maintain the spatial information. Slides stored in a dessicator (or in a sealed container with a dessicant pouch) at 4°C yield quality results for up to 3 months. The quality of results is tissue and block dependent and should be tested empirically.

Results that will be delivered to the users:

The National Facility for Genomics will deliver to the users the following files for every sample sequenced:

- FASTQ files

- QC report
- Mapping metrics

Access modality available: Access to NF services (users are responsible for data analysis).

SID: G-019 - Nanopore gDNA sequencing (long reads or ultra long reads)

Services description:

Nanopore sequencing is a next-generation sequencing technology that uses nanopores to directly sequence DNA molecules and to analyse DNA bases modifications. This method is known for producing long reads, and in some cases, ultra-long reads, making it valuable for various genomic applications.

Here is a general overview of the Nanopore gDNA (genomic DNA) sequencing protocol for long reads or ultra-long reads:

DNA Extraction: Start with the extraction of high-quality genomic DNA from the biological sample of interest. Different DNA extraction methods can be used based on the sample type.

Library Preparation steps:

- Fragmentation: Fragment the genomic DNA to the desired size. For long reads, the fragmentation is typically minimal, while for ultra-long reads, larger fragments may be preserved so no fragmentation will be performed.
- Repair Ends: Repair the DNA ends to ensure they have a consistent structure suitable for sequencing.
- Adapter Ligation: Attach sequencing adapters to the DNA fragments. These adapters contain nanopore-specific sequences that facilitate the capture and reading of the DNA sequence.
- Loading the Sequencing Device: Prepare the nanopore sequencing device according to the manufacturer's instructions. This may involve priming the flow cell, loading the prepared library, and initiating the sequencing run.

Sequencing: In nanopore sequencing, a single DNA strand passes through a nanopore, and as the DNA moves through the pore, the changes in electrical current are measured. The sequencing instrument records these changes in current, allowing for the identification of individual bases in the DNA sequence.

Data Collection: During the sequencing run, data is continuously collected in the form of raw electrical signals. The raw signals are then base-called to convert the electrical signal data into the corresponding DNA sequence.

Data Analysis: Process the base-called data through bioinformatics pipelines to correct errors, filter out low-quality reads, and assemble the long or ultra-long reads (users are responsible for data analysis).

Library preparation protocol:

Libraries will be prepared by following the protocols:

Ligation Sequencing Kit V14 (SQK-LSK114) for long reads

Ligation sequencing DNA V14 (SQK-LSK114)

DNA ends are repaired and dA-tailed using the NEBNext End Repair/dA-tailing module before the sequencing adapters, supplied in the kit, are ligated onto the prepared ends. The kit is optimised to achieve sequencing accuracies of over 99% (Q20+) with high output on the latest nanopore Flowcells R10.4.1. The Ligation Sequencing Kit V14 is compatible with upstream processes such as target enrichment by sequence capture, whole genome amplification, and size selection (for enrichment of specified fragment lengths). PCR- and WGA-free workflows remove amplification bias and retain base modification information, which can be analysed using bioinformatic tools supported by Oxford Nanopore.

Ultra-Long DNA Sequencing Kit V14 (SQK-ULK114) for ultra long reads

Ultra-Long DNA Sequencing Kit V14 (SQK-ULK114)

The Ultra-Long DNA Sequencing Kit V14 offers a means of preparing ultra-high molecular weight (uHMW) DNA for sequencing, which has shown to give N50s >50 kb and reads up to 4+ Mb. The kit is based on transposase chemistry: the transposase simultaneously cleaves template molecules and attaches tags to the cleaved ends. Rapid sequencing adapters are then added to the tagged ends. The last step is an overnight elution of the DNA library. This kit has been updated to use the newest Kit 14 chemistry which includes improved modal raw read sequencing accuracies with higher output on the latest nanopore Flowcells R10.4.1.

Libraries sequencing and NGS coverage:

For nanopore gDNA sequencing applied to Human studies a coverage of 20X per sample will be obtained.

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

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

Within this call the National Facility for Genomics will provide:

	Min samples/project	**Max samples/project**
G-019-A - nanopore gDNA sequencing (Human)	8	32
G-019-B - nanopore Ligation sequencing gDNA (Plants)	5	12
G-019-C - nanopore Ligation sequencing gDNA (Animals)	5	10

Technical requirements

For long-reads generation:

- gDNA Samples should be provided in low bind tubes (i.e. 1,5 ml Eppendorf tubes).
- At least 1 µg in 50 µl of gDNA should be provided
- DNA quantity evaluated by fluorimeter; i.e. Qubit dsDNA BR Assay Kit
- gDNA should be of high molecular weight (DIN≥6 obtained from Tape Station Agilent/Bioanalyzer), and clear of contaminant.
- Purity of the gDNA should be assessed with a Spectrophotometer (i.e. Nanodrop; 260/280≥1,8 and 260/230≥1,8).
- gDNA samples should be quantified by using a fluorometer (i.e Qubit/ Glomax).

For ultra long-reads generation: At least 3 pellets of 6 million each of frozen cells per sample should be provided. Cell pellets should be provided in low bind tubes (i.e. 1,5 ml Eppendorf tubes).

Results that will be delivered to the users:

The National Facility for Genomics will deliver to the users the following files for every sample sequenced:

- Pod5 and FASTQ files
- QC report
- Mapping metrics (if reference genome/transcriptome is available)

Access modality available: Access to NF services (users are responsible for data analysis).

SID: G-020 - Nanopore small gDNA sequencing (long reads)

Services description:

Nanopore sequencing of small gDNAs will be performed with the native barcoding kit that refers to a set of reagents and protocols designed to enable the simultaneous sequencing of multiple samples by adding unique barcodes to each sample before sequencing. This approach is particularly useful for studying small bacteria genomes (gDNA) as it allows for high-throughput sequencing and analysis of multiple samples using a single flowcell.

This approach provides a cost-effective and efficient way to sequence multiple small bacteria genomes simultaneously, making it a valuable tool in microbiome studies, environmental monitoring, and other applications in microbial genomics.

Here is a step-by-step description of the process:

DNA Extraction: Begin with the extraction of genomic DNA (gDNA) from small bacteria samples using standard DNA extraction methods. This step must be taken by users (see technical requirements below).

Library Preparation: The native barcoding kit includes reagents to tag each gDNA sample with a unique barcode. This is typically done by ligating barcode adapters to the gDNA fragments.

Barcoding: Introduce sample-specific barcodes during library preparation. These barcodes serve as unique molecular identifiers, allowing for the identification and demultiplexing of the sequences from different samples in the subsequent analysis.

Loading the Sequencing Device: Load the barcoded libraries onto a nanopore sequencing platform and Flowcell.

Sequencing: The nanopore sequencer reads individual DNA strands as they pass through a nanopore, generating long reads with real-time sequencing information.

Base Calling and Analysis: The raw electrical signals generated by the nanopore sequencer are translated into DNA base calls through base-calling algorithms.

Data Analysis: Use bioinformatics tools to analyse the sequencing data, including aligning the reads to a reference genome, variant calling, and identifying unique features of the small bacteria genomes (users are responsible for data analysis).

Library preparation protocol:

Libraries will be prepared by following the protocol:

Ligation sequencing gDNA - Native Barcoding (SQK-NBD114.96)

The Native Barcoding Kit 96 V14 is a kit providing 96 unique barcodes to enable PCR-free multiplexing of dsDNA samples such as gDNA and amplicons. The library preparation method is similar to the Ligation Sequencing Kit protocol; gDNA is repaired and dA-tailed using the NEBNext End Repair/dA-tailing module, and then a unique dT-tailed barcode adapter is ligated on the dA-tailed template. Barcoded samples are then pooled together. Each barcode adapter also has a cohesive end, and this is used as a hook to ligate to the supplied sequencing adapter. The kit is optimised to achieve modal sequencing accuracies of over 99% (Q20+) with high output on the latest nanopore Flowcells R10.4.1.

Libraries sequencing and NGS coverage:

For nanopore smallgDNA sequencing applied to Bacteria genomes the coverage per sample will be determined considering the Bacteria genome size and the number of bacteria genomes loaded in a single flowcell.

Within this call the National Facility for Genomics will provide:

nanopore small gDNA seq (for Bacteria samples) for a maximum of 24 batches/pools of samples considering that every batch/pool of samples can be composed of a maximum of 96 samples (max batch 96).

	Min samples/project	**Max samples/project**
G-020 -A - nanopore small gDNA seq - native barcoding (Bacteria) (max batch 96)	4	24

Technical requirements

- gDNA Samples should be provided in Low Bind full skirted PCR plates (i.e. Eppendorf twin.tec PCR plates).
- At least 1 µg in 15 µl of gDNA should be provided.
- gDNA should be of high molecular weight (DIN≥6 obtained from Agilent Tape Station/Bionalyzer), and clear of contaminant.
- Purity of the gDNA should be assessed with a Spectrophotometer (i.e. Nanodrop; 260/280≥1,8 and 260/230≥1,8).
- gDNA samples should be quantified by using a fluorometer (i.e Qubit/ Glomax).

Results that will be delivered to the users:

The National Facility for Genomics will deliver to the users the following files for every sample sequenced:

- Pod5 and FASTQ files
- QC report
- Mapping metrics (if reference genome/transcriptome is available).

Access modality available: Access to NF services (users are responsible for data analysis).

SID: G-021 - Nanopore Direct RNA Sequencing

Services description:

Direct RNA Sequencing with Nanopore technology is a cutting-edge method for sequencing RNA molecules without the need for conversion to complementary DNA (cDNA) as required in traditional RNA sequencing methods. The protocol involves the direct sequencing of RNA strands through nanopores, allowing for the real-time detection of RNA sequences and RNA bases modifications. Direct RNA Sequencing with Nanopore technology offers the advantage of studying RNA molecules in their native state, providing valuable insights into RNA processing, alternative splicing, and modifications without the need for reverse transcription. It is particularly valuable for capturing the full complexity of the transcriptome.

Here's a general overview of the Direct RNA Sequencing protocol using Nanopore technology:

RNA Extraction: Begin by isolating high-quality RNA from the biological sample of interest. This could be total RNA or specific RNA fractions, depending on the experimental goals (this step must be taken by users).

Library Preparation steps:

Adapter Ligation: Attach sequencing adapters to the RNA fragments. These adapters contain nanopore-specific sequences necessary for capturing and sequencing the RNA.

Loading the Sequencing Device: Prime the nanopore sequencing device according to the manufacturer's instructions. Load the prepared RNA library onto the flow cell, initiating the sequencing run.

Sequencing: As RNA strands pass through the nanopore, changes in electrical current are measured in real-time. The changes in current are characteristic of the specific RNA bases passing through the nanopore, allowing for base-calling and the reconstruction of the RNA sequence.

Data Collection: Raw electrical signals are continuously recorded during the sequencing run. Base-calling algorithms are applied to convert the electrical signals into the corresponding RNA sequence.

Data Analysis: Process the base-called data through bioinformatics pipelines to correct errors and filter out low-quality reads. Align the reads to a reference genome or transcriptome, if available, to identify known transcripts (users are responsible for data analysis).

Library preparation protocol:

Libraries will be prepared by following the protocol:

Direct RNA sequencing (SQK-RNA004)

The Direct RNA Sequencing Kit (SQK-RNA004) is used to prepare and sequence native RNA without conversion to cDNA. Inputs include poly(A)-tailed RNA or total RNA, such as eukaryotic mRNA and viral RNA.

This kit upgrade includes increased sequencing output and improved modal raw read accuracy on the new RNA flow cell (FLO-MIN004RA and FLO-PRO004RA).

Libraries sequencing and NGS coverage:

For nanopore Direct RNA Sequencing applied to Human samples each sample will be sequenced in one PromethION Flowcell.

Within this call the National Facility for Genomics will provide:

nanopore Direct RNA Sequencing (for Human samples) for a maximum number of 20 samples. Projects with a sample size ranging from a minimum of 10 to a maximum of 20 samples will be accepted.

	Min samples/project	**Max samples/project**
G-021 -A - nanopore Direct RNA Sequencing (Human)	10	20

Technical requirements

- RNA Samples should be provided in low bind tubes (i.e. 1,5 ml Eppendorf tubes).

- Three aliquots containing 1 µg of total RNA in 10ul per sample should be provided by the user.
- Total RNA should be DNase treated and the RIN≥8 (quality of RNAs evaluated by Agilent Tape Station/Bioanalyzer).
- Purity of the total RNAs should be assessed with a Spectrophotometer (i.e. Nanodrop; 260/280≥1,8 and 260/230≥1,8).
- Total RNA samples should be quantified by using a fluorometer (i.e Qubit/ Glomax).

Results that will be delivered to the users:

The National Facility for Genomics will deliver to the users the following files for every sample sequenced:

- Pod5 and FASTQ files
- QC report
- Mapping metrics (if reference genome/transcriptome is available).

Access modality available: Access to NF services (users are responsible for data analysis).

SID: G-022 - Nanopore cDNA sequencing (bulk cDNA or single-cell cDNA from 10x Genomics protocol) (Human-Mouse)

Services description:

Nanopore cDNA sequencing is a powerful technique that combines the benefits of nanopore sequencing technology with the study of complementary DNA (cDNA), which represents the transcribed RNA in a biological sample. This method allows researchers to investigate gene expression, alternative splicing, and other aspects of RNA biology with long-read sequencing capabilities. Nanopore cDNA sequencing offers several advantages, including the ability to generate long reads that span entire transcripts. This makes it particularly valuable for studying complex transcriptomes, characterizing novel isoforms, and exploring the dynamics of gene expression in various biological contexts.

Here is a step-by-step description of the process:

RNA Extraction: Begin by extracting total RNA from the biological sample of interest. This can be done using standard RNA extraction protocols, ensuring the preservation of RNA integrity. This step must be taken by users (see technical requirements below).

cDNA Synthesis: Reverse transcribe the RNA into complementary DNA (cDNA) using reverse transcription enzymes and primers. This step converts the RNA into a stable form of DNA that can be sequenced (this step can be taken by users or asked to the facility).

Library Preparation: Prepare the cDNA for sequencing by adding sequencing adapters to the ends. This usually involves ligating adapters with unique barcodes to distinguish different samples or conditions.

Sequencing: Load the prepared cDNA libraries onto a nanopore sequencing ONT flowcell and platform. The nanopore sequencer reads individual cDNA strands as they pass through a nanopore, producing long-read sequences with real-time sequencing information.

Base Calling and Analysis: Translate the raw electrical signals generated by the nanopore sequencer into DNA base calls using base-calling algorithms.

Use bioinformatics tools to align the cDNA reads to a reference genome or transcriptome, identify gene isoforms, quantify gene expression levels, and analyze alternative splicing events (users are responsible for data analysis).

Library preparation protocol:

For bulk cDNA sequencing **l**ibraries will be prepared by following the protocol:

cDNA-PCR Sequencing V14 (SQK-PCS114)

The PCR-cDNA Sequencing Kit is used to prepare cDNA for nanopore sequencing from an input of as low as 4 ng poly(A)+ RNA. When poly(A)+ enriched RNA is not available it is possible to use 200 ng of total RNA, but additional optimisation may be required.

The protocol uses a strand switching method to select for full length transcripts, allowing the identification of splice variants, with the incorporation of unique molecular identifiers (UMI) during this step. Taking full-length poly(A)+ RNA, complementary strand synthesis and strand switching are performed using kit-supplied oligonucleotides. Double-stranded cDNA is then generated by PCR amplification using primers that contain 5' tags which facilitate the ligase-free attachment of Rapid Sequencing Adapters.

The PCR-cDNA Sequencing Kit also includes a new cDNA RT adapter and RT primer to prime cDNA synthesis from the end of a transcript to reduce overlaps during the reverse transcription step and to allow users to measure polyA+ tail lengths.

For single-cell cDNA from 10x Genomics prep libraries will be prepared by following the protocol:

cDNA-PCR Sequencing V14 (SQK-PCS114)

A primer pair compatible with the PCR-cDNA Sequencing Kit and with a biotin on the primer annealing to the 3' end of the cDNA, is used to amplify 10 ng of cDNA generated with 10X Genomics Gene Expression protocols. The amplified cDNA is captured with streptavidin beads to enrich for full-length cDNA and remove RT artifact due to priming internal on the RNA. The captured cDNA is enriched by PCR amplification using primers that contain 5' tags which facilitate the ligase-free attachment of Rapid Adapter T (RAP T).

Libraries sequencing and NGS coverage:

For nanopore bulk cDNA or single-cell cDNA sequencing (from 10x Genomics protocols) applied to Human/mouse samples each sample will be sequenced in one PromethION Flowcell.

Within this call the National Facility for Genomics will provide:

nanopore bulk cDNA or single-cell cDNA Sequencing (for Human samples) for a maximum number of 20 samples. Projects with a sample size ranging from a minimum of 10 to a maximum of 20 samples will be accepted.

	Min samples/project	Max samples/project
G-022-Nanopore cDNA sequencing (bulk cDNA or single-cell cDNA from 10x Genomics protocol) (Human-Mouse)	10	20

Technical requirements

For bulk cDNA sequencing:

- Provide either three aliquots containing 500 ng total RNA in 10ul or three aliquots containing 200ng cDNA in 50ul.
- Total RNA should be DNase treated and the RIN≥8 (quality of RNAs evaluated by Agilent Tape Station/Bioanalyzer).
- Purity of the total RNAs should be assessed with a Spectrophotometer (i.e. Nanodrop; 260/280≥1,8 and 260/230≥1,8).
- Both total RNA and cDNA samples should be quantified by using a fluorometer (i.e Qubit/ Glomax).

For single-cell cDNA from 10x Genomics prep:

- two aliquots containing 10 ng of cDNA in 25ul should be provided by the user. The cDNA should be generated by using either the

Chromium Single Cell 3' Reagent Kits User Guide (v3.1 Chemistry Dual Index) - CG000315

Chromium GEM-X Single Cell 3' Reagent Kits v4 - CG000731

Or

Chromium Single Cell 5' Reagent Kits User Guide (v2 Chemistry Dual Index) - CG000331

Chromium GEM-X Single Cell 5'-V(D)J Reagent Kits v3 - CG000733

cDNA samples should be quantified by using a fluorometer (i.e Qubit/ Glomax) and their quality evaluated by Agilent Tape Station/Bioanalyzer.

Results that will be delivered to the users:

The National Facility for Genomics will deliver to the users the following files for every sample sequenced:

- Pod5 and FASTQ files
- QC report
- Mapping metrics (if reference genome/transcriptome is available).

Access modality available: Access to NF services (users are responsible for data analysis).

SID: G-023 - Nanopore cell-free DNA sequencing (Human)

Services description:

Cell-free DNA (cfDNA) sequencing with Nanopore technology represents a revolutionary approach to interrogate the genetic information present in circulating DNA, often extracted from blood plasma. Unlike traditional sequencing methods, this protocol allows for the direct sequencing of cfDNA without the need for intermediate steps such as PCR amplification or conversion and allows the analysis in real time of the methylation status of circulating cfDNA. Its applications extend to clinical diagnostics, providing valuable information for personalized medicine and disease monitoring such as liquid biopsy for cancer detection, monitoring treatment response, and identifying minimal residual disease.

Here is a general overview of the Cell-free DNA Sequencing with Nanopore Protocol:

cfDNA Extraction: Begin by isolating cell-free DNA from a biological fluid, such as blood plasma, using optimized extraction methods. This step must be taken by users (see technical requirements below).

Library Preparation: preparing the library for nanopore sequencing ligating sequencing adapters.

Loading the Sequencing Device: Load the prepared cfDNA library onto the Nanopore sequencing device to initiate the sequencing run.

Sequencing: As cfDNA molecules pass through the nanopore, real-time changes in electrical current enable the identification of DNA sequences and methylation status.

Data Collection: Record raw electrical signals continuously during the sequencing run, followed by super accurate base-calling to convert signals into DNA sequences.

Data Analysis: Process and analyse base-called data, aligning sequences to a reference genome or performing de novo assembly.

Variant Calling and Copy Number Analysis: Identify genetic variants and conduct copy number analysis to detect amplifications or deletions in cfDNA-covered genomic regions or evaluate the methylations status of the cfDNA fragments (users are responsible for data analysis).

Library preparation protocol:

Libraries will be prepared by following the protocol:

Ligation sequencing V14 — Human cfDNA singleplex (SQK-LSK114)

Ligation sequencing V14 — Human cfDNA multiplex (SQK-NBD114.24)

Library preparation method steps: DNA ends are repaired and dA-tailed using the NEBNext End Repair/dA-tailing module before the sequencing adapters (supplied in the kit) are ligated onto the prepared ends. The kit is optimised to achieve sequencing accuracies of over 99% (Q20+) with high output on the latest nanopore Flowcells R10.4.1. The protocol uses a modified version of the long-reads protocol optimised to recover short DNA fragments.

Libraries sequencing and NGS coverage:

For nanopore cell-free DNA sequencing applied to Human samples each sample will be sequenced in one PromethION Flowcell.

Within this call the National Facility for Genomics will provide:

	Min samples/project	**Max samples/project**
G-023-Nanopore cell-free DNA sequencing (Human)	20	60

Technical requirements

- cfDNA samples should be provided in Low Bind full skirted PCR plates (i.e. Eppendorf twin.tec PCR plates).
- Provide an input of at least 30ng in 50 µl. DNA should be clear of contaminant.
- Purity of the DNA should be assessed with a Nanodrop 2000 Spectrophotometer (i.e. Nanodrop; 260/280≥1,8 and 260/230≥1,8).
- cfDNA samples should be quantified by using of a fluorometer (i.e Qubit/ Glomax) and their quality evaluated by Agilent Tape Station/Bioanalyzer

Results that will be delivered to the users:

The National Facility for Genomics will deliver to the users the following files for every sample sequenced:

- Pod5 and FASTQ files
- QC report
- Mapping metrics (if reference genome/transcriptome is available).

Access modality available: Access to NF services (users are responsible for data analysis).

SID: G-024 - Sequencing only with NovaSeq 6000 (Illumina)

Services description:

The National Facility for Genomics will provide sequencing of pools of libraries prepared by the users with the NovaSeq6000 sequencing platform (Illumina).

The NovaSeq 6000 is a high-throughput sequencing platform developed by Illumina. It's designed to handle a wide range of applications and large-scale genomic projects.

NovaSeq 6000 is commonly used for: Whole Genome Sequencing (WGS), Exome Sequencing (WES), Transcriptome Analysis (bulk RNA-Seq or single-cell RNAseq), Metagenomics, Epigenomics studies, Population Genomics.

Libraries sequencing and NGS coverage:

For Sequencing only with NovaSeq 6000 (Illumina) the total amount of reads that will be generated per sample or the coverage per genome will be determined depending on the projects' needs and on the number of samples that will fit within the number of S4 Flowcells required.

Within this call the National Facility for Genomics will provide:

Sequencing only with NovaSeq 6000 (Illumina) for a total of 6 S4 Flowcells 300 cycles (generating 150bp PE reads) and for a total of 6 S4 Flowcells 200 cycles (generating 100bp PE reads). Projects that will require from a minimum of 1 to a maximum of 6 S4 Flowcells RUNs will be accepted.

	Min flowcell/project	**Max flowcell/project**
G-024-A - Sequencing only (NovaSeq) (S4 Flowcells 300 cycles generating 150bp PE reads)	1	6
G-024-B - Sequencing only (NovaSeq) S4 Flowcells 200 cycles generating 100bp PE reads)	1	6

Technical requirements

- Libraries should be provided already pooled in a Low Bind tube (i.e. 1,5 ml Eppendorf tubes).
- The Pool molarity should be at least 5nM in a volume of 200ul in either nuclease-free water, Illumina RSB buffer, or 10mM Tris HCl pH 8.5.
- The pool should be quantified by a fluorometer (i.e Qubit/ Glomax) and its quality evaluated by Agilent Tape Station/Bioanalyzer.
- Only unique dual indexed libraries will be accepted.

At the following link, a tool for calculating how to prepare the pool of libraries for sequencing is available: **Link to Pooling Calculator**

Results that will be delivered to the users:

The National Facility for Genomics will deliver to the users the following files for every sample sequenced:

- FASTQ files
- QC report

For the users requesting data analysis as a combined service, the **National Facility for Data Handling and Analysis** will also provide the files described in:

- NF62.01.01 Bulk RNA-Seq analysis
- NF62.01.02 miRNA analysis
- NF62.02.01 WGS analysis
- NF62.02.02 WES analysis
- NF62.02.03 Microbiome analysis

Access modality available: Access to NF services. **Service available in combination with the National Facility for Data Handling and Analysis**. To identify the correct service, please check the Data Handling and Analysis Call for Access.

SID: G-025 - Sequencing only with NextSeq 2000 (Illumina)

Services description:

The National Facility for Genomics will provide sequencing of pools of libraries prepared by the users with the NextSeq2000 sequencing platform (Illumina).

The NextSeq 2000 is a next-generation sequencing platform developed by Illumina. It is designed to offer high-throughput sequencing with flexibility for various applications.

The NextSeq 2000 is commonly used for: Whole Genome Sequencing (WGS), RNA Sequencing (RNA-Seq and single-cell RNA-seq), Targeted Sequencing (amplicon sequencing), Exome Sequencing (WES), Metagenomics, ChIP-Seq and Epigenomics.

Libraries sequencing and NGS coverage:

For Sequencing only with NextSeq 2000 (Illumina) the total amount of reads that will be generated per sample or the coverage per genome will be determined depending on the projects' needs and on the number of samples that will fit within the number of P3 Flowcells required.

Within this call the National Facility for Genomics will provide:

Sequencing only with NextSeq 2000 (Illumina) for a total of 10 P3 Flowcells 300 cycles (generating 150bp PE reads), for a total of 10 P3 Flowcells 200 cycles (generating 100bp PE reads) and for a total of 10 P2 Flowcells 600 cycles (generating 300bp PE reads). Projects that will require from a minimum of 5 to a maximum of 10 P2/P3 Flowcells RUNs will be accepted.

	Min flowcell/project	**Max flowcell/project**
G-025-A - Sequencing only (NextSeq) (P3 Flowcells 300 cycles generating 150bp PE reads)	5	10
G-025-B - Sequencing only (NextSeq) (P3 Flowcells 200 cycles generating 100bp PE reads)	5	10
G-025-C - Sequencing only (NextSeq) (P2 Flowcells 600 cycles generating 300bp PE reads)	5	10

Technical requirements

- Libraries should be provided already pooled in a Low Bind tube (i.e. 1,5 ml Eppendorf tubes).
- The Pool molarity should be at least 5nM in a volume of 60ul in either nuclease-free water, Illumina RSB buffer, or 10mM Tris HCl pH 8.5.
- The pool should be quantified by a fluorometer (i.e Qubit/ Glomax) and its quality evaluated by Agilent Tape Station/Bioanalyzer.
- Only dual indexed libraries will be accepted.

At the following link, a tool for calculating how to prepare the pool of libraries for sequencing is available: **Link to Pooling Calculator**

Results that will be delivered to the users:

The National Facility for Genomics will deliver to the users the following files for every sample sequenced:

- FASTQ files
- QC report

For the users requesting data analysis as a combined service, the **National Facility for Data Handling and Analysis** will also provide the files described in:

- NF62.01.01 Bulk RNA-Seq analysis
- NF62.01.02 miRNA analysis
- NF62.02.01 WGS analysis
- NF62.02.02 WES analysis
- NF62.02.03 Microbiome analysis

Access modality available: Access to NF services. **Service available in combination with the National Facility for Data Handling and Analysis**. To identify the correct service, please check the Data Handling and Analysis Call for Access.

SID: G-026 - Sequencing only with MiSeq (Illumina)

Services description:

The National Facility for Genomics will provide sequencing of pools of libraries prepared by the users with the MiSeq sequencing platform (Illumina).

the MiSeq is a benchtop next-generation sequencing platform developed by Illumina. It's designed for smaller-scale sequencing projects and offers flexibility for various applications.

MiSeq is commonly used for: Targeted Sequencing (amplicon sequencing), Small Genomes Sequencing, Metagenomics, 16S rRNA Sequencing, Small RNA-Seq, Viral genomes Sequencing.

Libraries sequencing and NGS coverage:

For Sequencing only with MiSeq (Illumina) the total amount of reads that will be generated per sample or the coverage per genome will be determined depending on the projects' needs and on the number of samples that will fit within the number of Flowcells required.

Within this call the National Facility for Genomics will provide:

Sequencing only with MiSeq (Illumina) for a total of 30 MiSeq v3 (600 cycle) Flowcells (generating 300bp PE reads) Projects that will require from a minimum of 10 to a maximum of 30 MiSeq v3 (600 cycle) Flowcells RUNs will be accepted.

	Min flowcell/project	**Max flowcell/project**
G-026 - Sequencing only with MiSeq (Illumina) (MiSeq v3 600 cycles Flowcells)	5	30

Technical requirements

- Libraries should be provided already pooled in a Low Bind tube (i.e. 1,5 ml Eppendorf tubes).
- The Pool molarity should be at least 5nM in a volume of 60ul in either nuclease-free water, Illumina RSB buffer, or 10mM Tris HCl pH 8.5.
- The pool should be quantified by a fluorometer (i.e Qubit/ Glomax) and its quality evaluated by Agilent Tape Station/Bioanalyzer.
- Only dual indexed libraries will be accepted.

At the following link, a tool for calculating how to prepare the pool of libraries for sequencing is available:

Link to Pooling Calculator

Results that will be delivered to the users:

The National Facility for Genomics will deliver to the users the following files for every sample sequenced:

- FASTQ files
- QC report

For the users requesting data analysis as a combined service, the **National Facility for Data Handling and Analysis** will also provide the files described in:

- NF62.02.03 Microbiome analysis

Access modality available: Access to NF services. **Service available in combination with the National Facility for Data Handling and Analysis**. To identify the correct service, please check the Data Handling and Analysis Call for Access.

Appendix 1: Summary of technical requirements

Service ID (SID)	Tube/Support type	N° of aliquots/ samples	Min Mass (ng)	Min Conc (ng/µl)	Min Conc (nM)	Min Vol (ul)	Max Vol (ul)	Min DIN/RIN/ DV200	Min 260/280	Min 260/230	DNAse treat	Min n° of Cells (M)	Min Cell viability (%)
G001 -Whole Genome Sequencing (WGS)	Low-bind full-skirted PCR plates	2		15		50		6	1.8	1.8			
G002 - Whole Exome Sequencing (WES)	Low-bind full-skirted PCR plates	2		5		40		6	1.8	1.8			
G003 - Amplicon sequencing for microbiome analysis (16S-ITS)	Low-bind full-skirted PCR plates	2		0.1		20		6	1.8	1.8			
G004 - Methylation sequencing (Methyl-seq)	Low-bind full-skirted PCR plates	2		15		60		6	1.8	1.8			
G005 - mRNA sequencing from standard input	Low-bind full-skirted PCR plates	2	500			20	50	7	1.8	1.8	x		
G005 - mRNA sequencing from low input	Low-bind full-skirted PCR plates	2	0.05			15	20	7	1.8	1.8	x		
G006 - totalRNA sequencing from standard input	Low-bind full-skirted PCR plates	2	500			20	50	4	1.8	1.8	x		
G006 - totalRNA sequencing from low input	Low-bind full-skirted PCR plates	2	0.05				7	4	1.8	1.8	x		
G007 - smallRNA sequencing	Low-bind full-skirted PCR plates	2	100			20	50	7	1.8	1.8	x		
G008-G012 - Single-cell 3'RNAsequencing	Low-bind 1.5 ml tubes	2										0.5	80
G008-G012 - Single-cell gene Expression Flex	Low-bind 1.5 ml tubes	2										0.3	80
G009-G012-G014 -Single-cell Immune profiling-VDJ (10X Genomics)	Low-bind 1.5 ml tubes	2										0.5	80
G010-G012-G013 -Single-cell multiome ATAC + Gene expression (10X Genomics)	Low-bind 1.5 ml tubes	2										0.5	80
G011-G013 - Single-cell ATAC sequencing (10X Genomics)	Low-bind 1.5 ml tubes	2										0.5	80
G015-G016 - Visium HD Spatial gene expression from Fresh-Frozen (10X Genomics)	Fisherbrand superfrost plus slides	2						4					
G015-G016 - Visium HD Spatial gene expression from Fixed Frozen (10X Genomics)	Fisherbrand superfrost plus slides	2						50%					
G015-G016 - Visium HD Spatial gene expression from FFPE tissues (10X Genomics)	Fisherbrand superfrost plus slides	2						30%					
G015-G016 - Visium Fresh-Frozen (Direct Placement)	Visium Slides (provided by NF)	1						7					
G017-018 - GeoMx Digital Spatial Profiling from Fresh-Frozen tissues (Nanostring)	SuperFrost™ Plus slides	2											
G017-018 - GeoMx Digital Spatial Profiling from FFPE tissues (Nanostring)	SuperFrost™ Plus slides	2											
G019 - Nanopore gDNA sequencing (long reads)	Low-bind 1.5 ml tubes	2		1000			50	6	1.8	1.8			
G019 - Nanopore gDNA sequencing (ultra long reads)	Low-bind 1.5 ml tubes	3										6 each	
G020 - Nanopore small gDNA sequencing (long reads)	Low-bind full-skirted PCR plates	2		1000			15	6	1.8	1.8			
G021 - Nanopore Direct RNA Sequencing	Low-bind 1.5 ml tubes	2		1000			10	8	1.8	1.8	x		
G022 - Nanopore cDNA sequencing (bulk cDNA) from RNA	Low-bind full-skirted PCR plates	3	500				10	8	1.8	1.8	x		
G022 - Nanopore cDNA sequencing (bulk cDNA) from cDNA	Low-bind full-skirted PCR plates	3	200				50		1.8	1.8			
G022 - Nanopore cDNA sequencing (single-cell cDNA from 10x Genomics protocol)(Human-Mouse)	Low-bind full-skirted PCR plates	2	10				25		1.8	1.8			
G023 - Nanopore cell-free DNA sequencing (Human)	Low-bind full-skirted PCR plates	2	30				50		1.8	1.8			
G024 - Sequencing only with NovaSeq 6000 (Illumina) (S4 Flowcells)	Low-bind 1.5 ml tubes	2			5	200							
G025 - Sequencing only with NextSeq 2000 (Illumina) (P3 Flowcells)	Low-bind 1.5 ml tubes	2			5	60							
G026 - Sequencing only with MiSeq (Illumina) (MiSeq v3 600 cycles Flowcells)	Low-bind 1.5 ml tubes	2			5	60							

Appendix 2: Description of the Data analysis services available in combination with the NF for Genomics services

Table 1: *Overview of the NF-Genomics Services which can be combined with the data analysis services provided by the NF for Data Handling and Analysis*

Service code	Service Name	Data Analysis Service code	Data Analysis Service Name
G-001	Whole Genome Sequencing	NF62.02.01	WGS analysis
G-002	Whole Exome Sequencing	NF62.02.02	WES analysis
G-003	Amplicon sequencing for microbiome analysis (16S-ITS)	NF62.02.03	Microbiome analysis
G-005	mRNA sequencing from standard and low input	NF62.01.01	Bulk RNA-Seq analysis
G-006	totalRNA from standard input	NF62.01.01	Bulk RNA-Seq analysis
G-007	Small RNA sequencing	NF62.01.02	miRNA analysis
G-008/G-012	Single-cell 3'RNAsequencing or Single-cell gene Expression Flex	NF62.03.01	scRNA-Seq analysis
G-009/012/014	Single-cell Immune profiling-V(D)J (10X Genomics)	NF62.03.03	Single-cell immune profiling (VDJ)
G-010/012/013	Single-cell multiome ATAC + Gene expression (10X Genomics)	NF62.03.04	Single-cell multiome (ATAC + gene expression)
G-011/G-013	Single-cell ATAC sequencing (10X Genomics)	NF62.03.02	scATAC-Seq analysis
G-015/016	Visium Spatial gene expression from Fresh-Frozen or FFPE tissues (10X Genomics)	NF62.03.05	Spatial transcriptomics (10X Visium platform)

G-024	Sequencing only with NovaSeq 6000 (Illumina)	NF62.02.01	WGS analysis
G-024	Sequencing only with NovaSeq 6000 (Illumina)	NF62.02.02	WES analysis
G-024	Sequencing only with NovaSeq 6000 (Illumina)	NF62.02.03	Microbiome analysis
G-024	Sequencing only with NovaSeq 6000 (Illumina)	NF62.01.01	Bulk RNA-Seq analysis
G-025	Sequencing only with NextSeq 2000 (Illumina)	NF62.02.01	WGS analysis
G-025	Sequencing only with NextSeq 2000 (Illumina)	NF62.02.02	WES analysis
G-025	Sequencing only with NextSeq 2000 (Illumina)	NF62.02.03	Microbiome analysis
G-025	Sequencing only with NextSeq 2000 (Illumina)	NF62.01.01	Bulk RNA-Seq analysis
G-026	Sequencing only with MiSeq (Illumina)	NF62.02.03	Microbiome analysis

A detailed description of each data analysis service offered is available below.

Table of contents

NF62.01.01 Bulk RNA-seq analysis

Service description

RNA sequencing is a powerful molecular biology technique used to analyse the transcriptome of a biological sample. The transcriptome refers to the complete set of RNA molecules, in particular messenger RNA (for mRNA sequencing) and/or non-coding RNAs (for total RNA sequencing), in a cell or tissue.

RNA sequencing is widely used in genomics research, functional genomics, and clinical studies to understand gene expression patterns, identify novel transcripts, and investigate how gene expression varies under different conditions. In addition, total RNA sequencing provides an insight also on the regulatory mechanisms underlying various biological processes.

The standard bioinformatics analysis for RNA-seq datasets comprises the following steps:

1. **Quality check of the raw sequence data**: Evaluating raw read sequencing quality, assessing nucleotide composition of the reads, K-mer over-representation, duplication rate, and adapter content.
2. **Trimming**: Reads are trimmed according to base quality, and reads having low average quality, as well as reads that are too short, are excluded from analysis. This step also trims adapters and other technical residuals. Quality control is performed again on the trimmed reads.
3. **Mapping to the reference genome**: The surviving good-quality reads are mapped to the reference genome using a splice-aware aligner (e.g. STAR). In parallel, pseudoalignment is performed too.
4. **Quantification of gene expression**: Uniquely mapped reads are assigned to the corresponding genomic features (i.e. exons, transcripts or genes). A count matrix is produced that summarizes the inferred expression level for each gene in each sample.
5. **Quality metrics collection**: Quality metrics from sequencing, trimming, alignment and quantification are collected and summarized in a complete, interactive report.
6. **Normalization and filtering**: Expression levels are normalized to account for the different library size across samples and/or the lengths of different genes. Non-expressed genes are filtered out.
7. **Exploratory analysis on expression data**: Principal Component Analysis (PCA) and Multi-Dimensional Scaling (MDS) are performed to inspect the variability structure of the data and its possible relationship with samples characteristics. If applicable and necessary, batch correction is performed using regression models. The expression of selected housekeeping genes can be evaluated across all samples, as well as the expression of gender-specific genes (for human datasets) and project-specific genes (if applicable, e.g. knocked out genes, tissue markers etc.).
8. **Differential expression analysis**: Expression levels are compared between different groups of samples, using statistical models based on the experimental design (e.g. paired models, regression of covariates etc.). Differentially expressed transcripts are identified by setting cutoffs on the obtained p-values and log2FoldChanges.

Advanced (optional) analysis steps include the following:

1. **Functional enrichment and pathway analysis of significant genes**: An over-representation analysis is performed to test the enrichment of the list of differentially

expressed genes against Gene Ontology and the main pathway collections (e.g. KEGG, Reactome, Biocarta, Hallmark, IPA).

2. **Alternative Splicing Analysis**: Mapped reads are analyzed to identify splicing isoforms and novel splice variants. Observed alternative splicing events are summarized and annotated.
3. **Identification of gene fusions events**: Gene fusions, resulting from the joining of two separate genes, have been found in various tumor types, leading to the overexpression and constitutive activation of genes not normally expressed.
4. **Variant calling**: RNA-seq datasets can be analyzed to identify variants in coding regions. Although an exact assessment of frequencies is not possible, this analysis may identify variants with a high potential for functional effects.

Access modality available

- Access to facility service

Requested inputs from users

All fastq files associated to all the samples must be provided, together with the corresponding md5 files (unless sequencing is performed by the National Facility for Genomics).

The user shall provide a table listing all biological and experimental conditions in the study and all samples belonging to each condition, ensuring that the sample names exactly match the names of the provided fastq files. For differential gene expression analysis, users should make sure to specify the conditions to be compared (see Metadata file example). The organism under study is human or a model organism, the user shall provide a reference to the annotated reference genome to be used for analysis.

Technical requirements

The libraries for all the samples must have been prepared using the same preparation kit, and sequencing must have been performed with the same pairing modality for all the samples (paired-end/single-end), ideally in the same sequencing run. We recommend a minimum sequencing depth of 30 million reads per sample for mammalian-sized genomes (this limit can be reduced in the case of smaller genomes), and a Q30 cutoff of 80%.

Results

The National Facility for Data Handling and Analysis will deliver the following to the users:

1. Trimmed and filtered fastq files for each sample.
2. BAM files for each sample.
3. Raw and normalized count matrices containing expression values for each gene in each sample.
4. Tables of differentially expressed genes with statistical significance information.
5. Complete reports (in interactive HTML and publication-ready PDF formats) describing the quality of the data, all the analysis performed on the dataset, and their results. Plots (e.g. heatmaps, volcano plots, dotplots, PCA/MDS) and tables included to the report will also be provided as separate files.
6. Pipeline and scripts used to perform the analysis, if applicable.

The facility will also assist the user in submitting raw data to public repositories, as stipulated in the *National Facilities Access Rules*.

Combined services

This service can be combined with the following services offered by the National Facility for Genomics:

G-005 - mRNA sequencing from standard and low input

G-006 - totalRNA from standard input

G-024 - Sequencing only with NovaSeq 6000 (Illumina)

G-025 - Sequencing only with NextSeq 2000 (Illumina)

To access the combined services, please submit an application to the National Facility for Genomics requesting data analysis.

NF62.01.02 miRNA analysis

Service description

Small RNA sequencing is a specialized technique designed to analyze and profile small RNA molecules present in a biological sample. It is widely used to study the expression profiles of miRNAs and other small RNAs, providing valuable insights into their roles in gene regulation, development, and disease. Small RNAs are polymeric ribonucleic acid molecules with a length lower than 200 nucleotides, comprising microRNA (miRNA), PIWI-interacting RNA (piRNA), small interfering RNA (siRNA), and tRNA-derived small RNA (tsRNA).

miRNAs are the most studied type of small RNAs, constituted by 20 to 25 nucleotides. They participate in several processes and can regulate gene expression at a posttranscriptional level. miRNAs can also act as transcription factors by binding the seed sequence within 3'UTR of target genes, leading to a variety of cell activities at different levels.

The standard bioinformatics analysis for miRNA datasets comprises the following steps:

1. **Quality check of the raw sequence data**: Sequencing quality of the raw reads is evaluated. It assesses the data quality distribution across reads, per-base content, and adapter contamination.
2. **UMI extraction and trimming**: Low-quality reads, UMI sequences and adapter contamination will be removed and excluded from the analysis. QC is performed again on the trimmed reads.
3. **Filtering for miRNA**: Filtering reads according to length and assessing their nature with respect to other types of small RNAs.
4. **Mapping**: Trimmed reads will be mapped against the reference genome, and mature miRNAs and precursors (hairpins) will be obtained from miRBase.

5. **Expression quantification**: Uniquely mapped reads are assigned to the corresponding features (mature miRNAs and miRNA precursors (hairpins)). A counts matrix is produced that summarizes the inferred expression level for each known miRNA in each sample.
6. **Quality metrics collection**: Quality metrics from sequencing, trimming, alignment and quantification are collected and summarized in a complete, interactive report.
7. **Normalization and filtering**: Expression levels are normalized to account for the different library sizes across samples. Non-expressed miRNAs are filtered out.
8. **Exploratory analysis on expression data**: Principal Component Analysis (PCA) and Multi-Dimensional Scaling (MDS) are performed to inspect the variability structure of the data and its possible relationship with sample characteristics. If applicable and necessary, batch correction is performed using regression models. The expression of selected project-specific targets (if applicable, e.g. knocked-out genes, tissue markers, etc.) is evaluated across all samples.
9. **Differential expression analysis**: Expression levels are compared between different groups of samples using statistical models based on the experimental design (e.g. paired models, regression of covariates, etc.). Differentially expressed miRNAs are identified by setting cutoffs on the obtained p-values and log2FoldChanges.

Advanced (optional) analysis steps include the following:

1. **Known and novel miRNA identification**: Canonical and non-canonical miRNAs are identified. An interactive report is produced with an overview of all detected miRNAs.
2. **Isomir identification**: BAM files are parsed, and a mirGFF3 file is created with the information about miRNAs and isomirs. Results will indicate unique isomirs for each miRNA, isomir sequences highlighting canonical sequences, and additions/deletions at 5' or 3' ends. Count matrices summarize total isomirs detected, reference sequence (miRBase) and number of miRNAs detected overall, and after filtering for the isomirs present in all samples.
3. **miRNA-targets identification**: miRNA-targets are obtained from external databases containing predicted (DIANA-microT-CDS, MicroCosm, miRanda, miRDB, PicTar, and TargetScan) or experimentally validated (miRecords, miRTarBase, and TarBase) miRNA-target interactions.
4. **Functional enrichment and pathway analysis of significant genes**: An over-representation analysis is performed to test the enrichment of the list of differentially expressed miRNAs and/or their target genes.

Access modality available

- Access to facility service

Requested inputs from users

All fastq files associated to all the samples must be provided, together with the corresponding md5 files (unless sequencing is performed by the National Facility for Genomics).

The user shall provide a table listing all biological and experimental conditions in the study and all samples belonging to each condition, ensuring that the sample names exactly match the names of the provided fastq files. For differential gene expression analysis, users should

make sure to specify the conditions to be compared (see Metadata file example). Unless the organism under study is human or a model organism, the user shall provide a reference to the annotated reference genome to be used for analysis.

Technical requirements

The libraries for all the samples must have been prepared using the same preparation kit, and sequencing must have been performed with the same pairing modality for all the samples (paired-end or single-end), ideally in the same sequencing run. We recommend a minimum sequencing depth of 5 million reads per sample for mammalian-sized genomes (this limit can be reduced in the case of smaller genomes), and a Q30 cutoff of 80%.

Results

The National Facility for Data Handling and Analysis will deliver the following to the users:

1. Trimmed and filtered fastq files for each sample.
2. BAM files for each sample.
3. Raw and normalized count matrices containing expression values for each miRNA in each sample.
4. Tables of differentially expressed miRNAs with statistical significance information.
5. Complete reports (in interactive HTML and publication-ready PDF formats) describing the quality of the data, all the analysis performed on the dataset, and their results. Plots (e.g. heatmaps, volcano plots, PCA/MDS) and tables included to the report will also be provided as separate files.
6. Pipeline and scripts used to perform the analysis, if applicable.

The facility will also assist the user in submitting raw data to public repositories, as stipulated in the *National Facilities Access Rules*.

Combined services

This service can be combined with the following services offered by the National Facility for Genomics:

G-007 – Small RNA sequencing.

To access the combined services, please submit an application to the National Facility for Genomics requesting data analysis.

NF62.02.01 WGS analysis

Service description

DNA sequencing is critical for genetic research, evolutionary studies, and personalized medicine, where it helps to uncover the genetic basis of diseases, track hereditary conditions, and guide targeted therapies. It provides a detailed understanding of an organism's complete genetic makeup, offering insights into complex biological processes and evolutionary relationships.

Whole-Genome Sequencing (WGS) involves sequencing the entire genome, including both coding and non-coding regions. WGS provides the most comprehensive view of an organism's genetic information, as the focus is not only on identifying genetic variants (e.g., single nucleotide variants, insertions, deletions, copy-number variation), but also on identifying rare variants, structural variations, and novel mutations in both coding and non-coding regions. Different algorithms will be applied for germline or somatic samples – the former algorithms are designed to identify inherited variants present in all cells, whereas the latter algorithms focus on detecting mutations acquired in specific tissues which are present only in a subset of cells thus requiring specialized methods to account for tissue purity and heterogeneity.

The standard bioinformatics analysis for WGS projects comprises the following steps:

1. **Quality check of the raw sequence data**: Evaluating raw read sequencing quality, assessing nucleotide composition of the reads, K-mer over-representation, duplication rate, and adapter content.
2. **Trimming**: Trimming reads according to base quality, excluding excessively short reads and those with low average quality from the analysis. Trimming adapters and other technical residuals. Performing a second QC step on the trimmed reads.
3. **Mapping to the reference genome**: Aligning high-quality reads to a reference genome, generating a BAM file. Removing PCR duplicates post-alignment to reduce bias, applying quality score recalibration to correct sequencing errors. Indel realignment may also be performed to refine alignment accuracy around insertion-deletions, ensuring reliable variant calling in downstream analysis.
4. **Variant calling**: Variant calling is the process by which algorithms scan the aligned reads for deviations from the reference genome, marking potential variant sites. Depending on the scientific purpose of the analysis we will identify different variant families:
    a. **Single Nucleotide Polymorphisms** (SNPs) are single-base changes in the DNA sequence, which may or may not affect gene function.
    b. **Insertion–deletion mutations** refer to small insertions or deletions (of less than 50 bases) in the genome.
    c. **Copy Number Variants** (CNVs) are structural variations in the genome, typically spanning kilobases to megabases, where segments of DNA are either duplicated or deleted.
    d. **Structural Variants** (SVs) are large-scale changes in the genome structure, such as inversions, translocations, duplications, or large insertions/deletions (more than 50 bases).

Advanced (optional) analysis steps include the following:

1. **Annotation and gene-level interpretation**: For all the different classes of standard analysis (SNPs, indels, CNV and SV) we will provide basic information on the genes and regulatory elements affected by the variation, which can reveal potential disease associations (functional impact, consequence on protein, pathogenicity predictors, population frequency data).
2. **Disease association analysis**: Based on the experiment we can provide specific annotations for germline (e.g. Clinvar, ACMG) or somatic variants (e.g. COSMIC, Civic, OncoKB, AMP).

3. **Trio analysis**: Identifies variants by inheritance patterns: de novo, autosomal recessive, autosomal dominant, or compound heterozygosity.
4. **Cancer-specific analysis**: Identification of tumor-Specific Signature Analysis; Actionable Mutation Identification (e.g., KRAS, BRCA1/2, BRAF, EGFR); Tumor Mutational Burden (TMB); Microsatellite Instability (MSI) and Mismatch Repair (MMR) Deficiency.
5. **Differential analysis**: Based on the experimental design, we can apply different statistical analyses to interpret genomic differences between the variants of the groups under examination

Access modality available

- Access to facility service

Requested inputs from users

All fastq files associated to all the samples must be provided, together with the corresponding md5 files (unless sequencing is performed by the National Facility for Genomics).

The user shall provide a table listing all biological and experimental conditions in the study and all samples belonging to each condition, ensuring that the sample names exactly match the names of the provided fastq files (see Metadata file example).

Unless the organism under study is human or a model organism, the user shall provide a reference to the annotated reference genome to be used for analysis.

Technical requirements

The libraries for all the samples must have been prepared using the same preparation kit, and sequencing must have been performed with the same pairing modality for all the samples (paired-end/single-end), ideally in the same sequencing run. We recommend a minimum average coverage of 30X (for germline variants) or 100X (for somatic variants), and a Q30 cutoff of 80%.

Somatic Variant Calling requires a panel of normal (PON) to perform the analysis. GATK recommends aiming for a minimum of 40 samples to create a PON[1].

For **CNV analysis** we recommend a reference set of at least 20 samples to ensure adequate representation of natural variation. It is best to include samples that are as similar as possible to the cases you are analyzing in terms of tissue type and other relevant characteristics.

Results

The National Facility for Data Handling and Analysis will deliver the following to the users:

1. Trimmed and filtered fastq files for each sample.
2. BAM files for each sample.
3. Raw VCF files for all samples, In case of advanced analysis, we will also provide filtered and annotated VCF files for all samples, and genotypes in tabular format.
4. Complete reports (in interactive HTML and publication-ready PDF formats) describing the quality of the data, all the analysis performed on the dataset, and their results. Plots (e.g.

[1] https://gatk.broadinstitute.org/hc/en-us/articles/360035890631-Panel-of-Normals-PON

heatmaps, Circos plots, Manhattan plots) and tables included to the report will also be provided as separate files.

5. Pipeline and scripts used to perform the analysis, if applicable.

The facility will also assist the user in submitting raw data to public repositories, as stipulated in the *National Facilities Access Rules*.

Combined services

This service can be combined with the following services offered by the National Facility for Genomics:

G-001 - Whole Genome Sequencing

G-024 - Sequencing only with NovaSeq 6000 (Illumina)

G-025 - Sequencing only with NextSeq 2000 (Illumina)

To access the combined services, please submit an application to the National Facility for Genomics requesting data analysis.

NF62.02.02 WES analysis

Service description

Exome sequencing is an application of DNA sequencing (see NF62.02.01) that focuses on preferentially sequencing the exons, or protein-coding regions, which make up about 1-2% of the genome and are more likely to harbor disease-causing mutations. It is used to study genetic variations that affect protein function, thus particularly in disease research. The focus is on identifying genetic variants (e.g., single nucleotide variants, insertions, deletions, copy-number variation). Different algorithms will be applied for germline or somatic samples – the former algorithms are designed to identify inherited variants present in all cells, whereas the latter algorithms require specialized methods to account for tissue purity and heterogeneity to detect mutations acquired in specific tissues which are present only in a subset of cells.

The standard bioinformatics analysis for WES projects comprises the following steps:

1. **Quality check of the raw sequence data**: Evaluating raw read sequencing quality, assessing nucleotide composition of the reads, K-mer over-representation, duplication rate, and adapter content.
2. **Trimming**: Trimming reads according to base quality, excluding excessively short reads and those with low average quality from the analysis. Trimming adapters and other technical residuals. Performing a second QC step on the trimmed reads.
3. **Mapping to the reference genome**: Aligning high-quality reads to a reference genome, generating a BAM file. Removing PCR duplicates post-alignment to reduce bias, applying quality score recalibration to correct sequencing errors.
4. **Variant calling**: Variant calling is the process by which algorithms scan the aligned reads for deviations from the reference genome, marking potential variant sites. Depending on the scientific purpose of the analysis we will identify different variant classes:

a. **Single Nucleotide Polymorphisms** (SNPs) are single-base changes in the DNA sequence, which may or may not affect gene function.

b. **Insertion–deletion mutations** refer to small insertions or deletions (of less than 50 bases) in the genome.

c. **Copy Number Variants** (CNVs) are structural variations in the genome, typically spanning kilobases to megabases, where segments of DNA are duplicated or deleted.

Advanced (optional) analysis steps include the following:

1. **Annotation and gene-level interpretation**: For all the different classes of standard analysis (SNPs, indels, and CNVs) we will provide basic information on the genes and regulatory elements affected by the variation, which can reveal potential disease associations (functional impact, consequence on protein, pathogenicity predictors, population frequency data).
2. **Disease association analysis**: Based on the experiment we can provide specific annotations for **germline** (e.g. Clinvar, OMIM, ACMG) or **somatic** variants (e.g. COSMIC, Civic, OncoKB, AMP).
3. **Trio analysis**: Identifies variants by inheritance patterns: de novo, autosomal recessive, autosomal dominant, or compound heterozygosity.
4. **Cancer-specific analysis**: Tumor-Specific Signature Analysis; Actionable Mutation Identification (e.g., KRAS, BRCA1/2, BRAF, EGFR); Tumor Mutational Burden (TMB); Microsatellite Instability (MSI) and Mismatch Repair (MMR) Deficiency.
5. **Differential Analysis**: Based on the experimental design, we can apply different statistical analyses to interpret genomic differences between the variants of the groups under examination.

Access modality available

- Access to facility service

Requested inputs from users

All fastq files associated to all the samples must be provided, together with the corresponding md5 files (unless sequencing is performed by the National Facility for Genomics).

The user shall provide a table listing all biological and experimental conditions in the study and all samples belonging to each condition, ensuring that the sample names exactly match the names of the provided fastq files (see Metadata file example).

Unless the organism under study is human or a model organism, the user shall provide a reference to the annotated reference genome to be used for analysis.

Technical requirements

The libraries for all the samples must have been prepared using the same preparation kit, and sequencing must have been performed with the same pairing modality for all the samples (paired-end/single-end), ideally in the same sequencing run. We recommend a minimum average coverage of 30X (for germline variants) or 100X (for somatic variants), and a Q30 cutoff of 80%.

Somatic Variant Calling requires a panel of normal (PON) to perform the analysis. GATK recommends aiming for a minimum of 40 samples to create a PON[2].

For **CNV analysis** we recommend a reference set of at least 20 samples to ensure adequate representation of natural variation. It is best to include samples that are as similar as possible to the cases analyzed in terms of tissue type and other relevant characteristics.

Results

The National Facility for Data Handling and Analysis will deliver the following to the users:

1. Trimmed and filtered fastq files for each sample.
2. BAM files for each sample.
3. Raw VCF files for all samples; in case of advanced analysis, we will also provide filtered and annotated VCF files for all samples, and genotypes in tabular format.
4. Complete reports (in interactive HTML and publication-ready PDF formats) describing the quality of the data, all the analysis performed on the dataset, and their results. Plots (e.g. heatmaps, Circos plots, Manhattan plots) and tables included to the report will also be provided as separate files.
5. Pipeline and scripts used to perform the analysis, if applicable.

The facility will also assist the user in submitting raw data to public repositories, as stipulated in the *National Facilities Access Rules*.

Combined services

This service can be combined with the following services offered by the National Facility for Genomics:

G-002 - Whole Exome Sequencing

G-024 - Sequencing only with NovaSeq 6000 (Illumina)

G-025 - Sequencing only with NextSeq 2000 (Illumina)

To access the combined services, please submit an application to the National Facility for Genomics requesting data analysis.

NF62.02.03 Microbiome Analysis

Service description

Microbiome analysis using 16S and ITS amplicon sequencing is a widely used technique to study the composition and diversity of microbial communities, particularly bacteria and fungi. The 16S ribosomal RNA (rRNA) gene is a molecular marker found in the genomes of bacteria and archaea, and its variable regions are commonly used for taxonomic classification, while ITS is used to profile fungal communities.

Microbiome analysis using 16S and ITS amplicon sequencing is valuable in a range of fields, including environmental science, human health, and agriculture. It provides a cost-effective

[2] https://gatk.broadinstitute.org/hc/en-us/articles/360035890631-Panel-of-Normals-PON

way to characterize microbial communities and understand their roles in various ecosystems or host-associated environments.

The standard bioinformatics analysis for microbiome datasets of variable regions of the 16S rRNA (V3-V5 regions) or ITS comprises the following steps:

1. **Quality check of the raw sequence data**: Evaluating sequencing quality of raw reads, assessing nucleotide composition of the reads, K-mer over-representation, duplication rate, and adapter content.
2. **Trimming**: Trimming reads according to base quality, excluding excessively short reads and those with low average quality from the analysis. Trimming adapters and other technical residuals. Performing a second QC step on the trimmed reads.
3. **Amplicon Sequence Variants (ASVs) inference**: Inferring ASVs from amplicon data by computing an error model on the sequencing reads. Dereplicating sequences via quality filtering, denoising, read pair merging (for paired end Illumina reads only) and PCR chimera removal. Removing mitochondrial and chloroplast sequences in order to focus exclusively on the microbial community.
4. **Taxonomic classification**: Clustering reads into operational taxonomic units (OTUs) or ASVs, referring to the SILVA database for 16S and the UNITE database for ITS.
5. **Abundance and relative abundance**: Calculating abundance based on the computed ASVs and taxonomic classification. Calculating relative abundance based on TSS (Total Sum Scaling normalization) for several taxonomic levels for each sample and reporting in tabular format.
6. **Diversity and Community Analysis (Alpha and Beta diversity)**: Assessing richness, evenness, and composition of the microbial communities using the alpha diversity (within-sample) and beta diversity (between-sample) measures.

Advanced (optional) analysis steps include the following:

1. **Differential abundance**: Differential abundance analysis identifies relative abundance from microbial features across sample groups using ANCOM statistical framework.
2. **Alpha diversity rarefaction curves**: Produce rarefaction plots displaying alpha diversity indices that determine samples richness.
3. **Functional abundances**: Functional abundances are predicted based on marker gene sequences. Enzyme Classification numbers and KEGG orthologs will be predicted for each sample.

Access modality available

- Access to facility services

Requested inputs from users

All fastq files associated to all the samples must be provided, together with the corresponding md5 files (unless sequencing is performed by the National Facility for Genomics).

The user shall provide a table listing all biological and experimental conditions in the study and all samples belonging to each condition, ensuring that the sample names exactly match the names of the provided fastq files (see Metadata file example).

If differential abundance analysis is requested, users should provide the list of conditions to be compared.

Technical requirements

All FASTQ files associated with all the samples must be provided, including the sequences of the amplicons, together with the corresponding md5 checksum files (unless sequencing is performed by the National Facility for Genomics).

The libraries for all the samples must have been prepared using the same preparation kit, and sequencing must have been performed with the same pairing modality for all the samples (paired-end or single-end), ideally in the same sequencing run. We recommend a minimum sequencing depth of 5 million reads per sample, and a Q30 cutoff of 80%.

The user shall provide a table listing all biological conditions in the experiment and all samples belonging to each condition, ensuring that the sample names exactly match the names of the provided fastq files (see Metadata file example).

Results

The National Facility for Data Handling and Analysis will deliver the following to the users:

1. Abundance, taxonomic, and ASV tables for each sample.
2. Complete reports (in interactive HTML and publication-ready PDF formats) describing the quality of the data, all the analysis performed on the dataset, and their results. Plots (*e.g.* taxonomic abundance barplots, phylogenetic trees, *etc.*) and tables included to the report will also be provided as separate files.
3. Phyloseq R object and scripts used to perform the analysis, if applicable.

The facility will also assist the user in submitting raw data to public repositories, as stipulated in the *National Facilities Access Rules*.

Combined services

This service can be combined with the following services offered by the National Facility for Genomics:

G-003 - Amplicon sequencing for microbiome analysis (16S-ITS)

G-026 - Sequencing only with MiSeq (Illumina)

G-025 - Sequencing only with NextSeq 2000 (Illumina)

G-024 - Sequencing only with NovaSeq 6000 (Illumina)

To access the combined services, please submit an application to the National Facility for Genomics requesting data analysis.

NF62.03.01 scRNA-seq analysis

Service description

Single-cell RNA sequencing (scRNA-seq) is a technique used to analyse gene expression at the individual cell level, making it possible to resolve cellular heterogeneity within a biological sample. Unlike bulk RNA sequencing, which averages gene expression across many cells, scRNA-seq enables the identification of distinct cell types, states, and subpopulations.

This approach is crucial for understanding complex tissues, developmental processes, and disease progression, as it reveals how gene expression varies from cell to cell, and it is widely applied in biomedical research to advance personalized medicine, immunology, cancer research, and tissue regeneration studies.

The standard bioinformatics analysis for a scRNA-seq dataset comprises the following steps:

1. **Quality check of the raw sequence data**: Evaluating raw read sequencing quality, assessing nucleotide composition of the reads, K-mer over-representation, duplication rate, and adapter content.
2. **Cell barcode identification and extraction**: Identifying unique cell barcodes corresponding to true cells using the number of associated transcripts as a proxy for their vitality. Extracting transcripts associated to these cells from sequencing reads.
3. **Mapping to the reference genome**: Aligning reads to a reference genome, generating a BAM file. Quantifying the number of reads mapped to each gene in order to measure gene expression per cell.
4. **Doublet detection and quality filtering on individual samples**: Identification of potential doublets (i.e. two or more cells mistakenly captured as one) based on the number and consistency of their expressed genes. Marking corresponding barcodes as potentially derived from multiplets, without initially excluding from the analysis. Applying other quality filters to exclude low-quality, stressed, or damaged cells.
5. **Dataset integration**: Integrating expression data from all sequenced samples into one single dataset on which the overall analysis is performed. Grouping of the samples into integrated datasets will depend on the experimental design and project requirements (e.g. in the case of samples derived from different species and/or tissues, etc.).
6. **Normalization and batch correction**: Normalizing expression values according to different library sizes and subsequent scaling. Identification of most variable genes within each dataset for use in subsequent steps. Evaluation and correction of "batch effect" variability related to the different samples of origin via data harmonization algorithms. Assessment of other potential sources of intrinsic variability, such as the cell cycle.
7. **Analysis of cell populations within the integrated datasets**: Analysis of the cellular composition of each integrated dataset using a standard workflow based on dimensionality reduction techniques (e.g., PCA, UMAP or t-SNE) and clustering algorithms to identify distinct groups of cells. Performing differential gene expression analysis (DGE) between these groups to detect marker genes specifically expressed by certain populations. These marker genes could be used to infer the identity of each cell type. If samples from different conditions were pooled together, a DGE could also be performed to compare the expression profiles of cell populations across conditions.

Advanced (optional) analysis steps include the following:

1. **Automatic cell type annotation**: Inferring the identity of each cell population using automated tools based on the lists of previously identified marker genes and known cell type-specific signatures.
1. **Advanced DGE models using pseudo-bulk**: In case of complex design, application of pseudo-bulk approaches to compare the expression profiles of specific cell populations across different conditions, while adjusting for biological and technical variables.
2. **Differential abundance analysis**: Testing whether the proportions of specific cell types vary across different types of samples.
3. **Variational autoencoders**: Employing these advanced analytical tools for various purposes, such as cleaning up noisy data, filling in missing information, combining datasets, or transferring labels between datasets.
4. **Cell-cell interactions**: Inspecting communication across different cell types through cell type-specific expression of signaling molecules such as ligands, receptors, and their downstream signaling pathways.
5. **Trajectory analysis**: Inferring transcriptional changes related to developmental processes, cell proliferation or response to simuli, via a pseudotime trajectory.

Requested inputs from users

All fastq files associated to all the samples must be provided, together with the corresponding md5 files (unless sequencing is performed by the National Facility for Genomics).

The user shall provide a table listing all biological and experimental conditions in the study and all samples belonging to each condition, ensuring that the sample names exactly match the names of the provided fastq files (see Metadata file example).

Unless the organism under study is human or a model organism, the user shall provide a reference to the annotated reference genome to be used for analysis.

Access modality available

- Access to facility service
- Access to facility service including training

Technical requirements

The libraries for all the samples must have been prepared using the same kit and barcoding strategy, ideally in the same sequencing run. We recommend at least 1000 cells per sample and a minimum of 50.000 reads per cell, with a Q30 cutoff of 80%.

Results

The National Facility for Data Handling and Analysis will deliver to the users the following files:

1. Raw and filtered fastq files for each sample.
2. BAM files for each sample.
3. Count matrices containing expression values for each gene in each cell.
4. Complete reports (in interactive HTML and publication-ready PDF formats) describing the quality of the data, all the analysis performed on the dataset, and their results. Plots (e.g. UMAP, dotplots, volcano plots, feature plots, etc.) and tables included to the report will also be provided as separate files.

5. Python objects (.h5ad) containing the processed data.
6. Pipeline and scripts used to perform the analysis, if applicable.

The facility will also assist the user in submitting raw data to public repositories, as stipulated in the *National Facilities Access Rules*.

Combined services

This service can be combined with the following services offered by the National Facility for Genomics:

G-008/G-012 - Single-cell 3'RNAsequencing or Single-cell gene Expression Flex

To access the combined services, please submit an application to the National Facility for Genomics requesting data analysis.

NF62.03.02 scATAC-seq analysis

Service description

Single-cell ATAC sequencing (scATAC-seq) is a powerful molecular biology technique used to profile the chromatin accessibility of individual cells/nuclei at a high resolution. Chromatin accessibility refers to the degree to which DNA within chromatin is accessible by cellular machinery, particularly those parts involved in transcription, such as transcription factors and RNA polymerase.

Unlike bulk ATAC sequencing, which cannot determine the chromatin states of individual subpopulations of cells within a sample, scATAC-seq is widely used to provide valuable insights into chromatin accessibility, transcription factor binding, epigenetic modifications, and gene regulation. This technology is particularly useful in studying various processes and biological mechanisms including developmental processes, tumorigenesis, and immunological memory establishment.

The standard bioinformatics analysis for scATAC-seq datasets comprises the following steps:

1. **Quality check of the raw sequence data**: Evaluating sequencing quality of raw reads, assessing nucleotide composition of the reads, K-mer over-representation, duplication rate, and adapter content.
2. **Mapping to the reference genome and peak calling**: Aligning good quality sequencing reads to the reference genome. Quantifying the number of fragments mapped to coding and non-coding regions (i.e promoters, enhancers) to identify accessible chromatin peaks.
3. **Barcode counting**: Identifying cell barcodes corresponding to true cells using the number of fragments overlapping peaks.
4. **Quantification of chromatin accessibility**: Summarizing the inferred chromatin accessibility level for each peak in each cell in a count matrix.
5. **Cell-level and sample-level QC metrics collection**: Identifying low quality cells based on several metrics including transcription start site (TSS) enrichment score, nucleosome signal, and the ratio of fragments in genomic blacklist regions. Evaluate sample-level quality through other quality filters such as the fraction of fragments in peak (FRIP).
6. **Normalization and dimensionality reduction**: Normalizing peak-cell matrices according to different library sizes and/or across peaks (e.g. frequency-inverse document frequency

(TF-IDF) normalization) in order to emphasize most informative features. Using the most variable features within each dataset for dimensionality reduction (e.g. SVD, PCA, UMAP).

Advanced (optional) analysis steps include the following:

1. **Differential accessibility analysis**: Differential accessibility region (DAR) analysis is performed to detect differences in chromatin accessibility across sample conditions.

Access modality available

- Access to facility service
- Access to facility service including training

Requested inputs from users

All fastq files associated to all the samples must be provided, together with the corresponding md5 files (unless sequencing is performed by the National Facility for Genomics).

The user shall provide a table listing all biological and experimental conditions in the study and all samples belonging to each condition, ensuring that the sample names exactly match the names of the provided fastq files (see Metadata file example).

Unless the organism under study is human or a model organism, the user shall provide a reference to the annotated reference genome to be used for analysis.

Technical requirements

The libraries for all the samples must have been prepared using the same kit and barcoding strategy, ideally in the same sequencing run. We recommend at least 1000 cells per sample and a minimum of 50.000 reads per cell, with a Q30 cutoff of 80%.

Results

The National Facility for Data Handling and Analysis will deliver the following to the users:

1. Raw and filtered fastq files for each sample (if not already available).
2. BAM files for each sample.
3. Raw and normalized peak-by-cell matrices containing peaks for each region of the genome in each cell.
4. Complete reports (in interactive HTML and publication-ready PDF formats) describing the quality of the data, all the analysis performed on the dataset, and their results. Plots (e.g. motif enrichment plots, etc.) and tables included to the report will also be provided as separate files.
5. Python objects (.h5ad) containing the processed data.
6. Pipeline and scripts used to perform the analysis, if applicable.

The facility will also assist the user in submitting raw data to public repositories, as stipulated in the *National Facilities Access Rules*.

Combined services

This service can be combined with the following services offered by the National Facility for Genomics:

G-011/G-013 - Single-cell ATAC sequencing (10X Genomics)

To access the combined services, please submit an application to the National Facility for Genomics requesting data analysis.

NF62.03.03 Single-cell Immune profiling-V(D)J

Service description

Single-cell immune profiling-V(D)J is a powerful molecular biology technique used to profile both 5' gene expression and T-cell and/or B-cell receptors of individual cells at a high resolution allowing the characterization of cellular heterogeneity and clonal expansion within a biological sample.

Unlike bulk RNA and T/B-cell receptor (TCR/BCR) sequencing, which allow to study gene expression and TCR/BCR repertoires across many cells, single-cell immune profiling-V(D)J enables the identification of distinct cell types, states, and subpopulations both in terms of transcriptional profile (GEX data) and TCR/BCR repertoires (V(D)J data). This approach is crucial for understanding complex tissues, developmental progression, tumorigenesis, and tracking clonal expansion and immune responses. It is widely applied in biomedical research to advance personalized medicine, immunology, cancer immunotherapy, autoimmune disease and infection disease.

The Single-cell Immune profiling-V(D)J datasets include two modalities: gene expression (GEX) and TCR/BCR (V(D)J).

The standard bioinformatics analysis for Single-cell Immune profiling-V(D)J datasets comprises the following steps:

Regarding the GEX data analysis:

1. **Quality check of the original sequences**: Evaluating raw read sequencing quality, assessing nucleotide composition of the reads, K-mer over-representation, duplication rate, and adapter content.
2. **Cell barcodes identification and extraction**: Identifying unique cell barcodes corresponding to true cells using the number of associated transcripts as a proxy for their vitality. Extracting transcripts associated to these cells from sequencing reads.
3. **Mapping to the reference genome and quantification of gene expression**: Aligning sequencing reads to the reference genome. Quantifying the number of reads mapped to each gene as a proxy for gene expression per cell.
4. **Doublet detection and quality filtering on individual samples**: Identification of potential doublets (i.e. two or more cells mistakenly captured as one) based on the number and consistency of their expressed genes. Marking corresponding barcodes as potentially derived from multiplets, without initially excluding from the analysis. Applying other quality filters to exclude low-quality, stressed, or damaged cells.

5. **Dataset integration**: Integrating expression data from all sequenced samples into one single dataset on which the overall analysis is performed. Grouping of the samples into integrated datasets will depend on the experimental design and project requirements, (e.g. in the case of samples derived from different species and/or tissues, etc.).
6. **Normalization and batch correction**: Normalizing expression values according to different library sizes and subsequent scaling. Identifying the most variable genes within each dataset are identified for further analysis steps. Evaluating and correcting "batch effect" variability related to the different samples of origin using data harmonization algorithms. Assessing other potential sources of intrinsic variability, such as cell cycle stage.
7. **Analysis of cell populations within integrated datasets**: Analyzing the cellular composition of each integrated dataset using a standard workflow based on dimensionality reduction techniques (e.g., PCA, UMAP or t-SNE) and clustering algorithms to identify distinct groups of cells. Detecting marker genes specifically expressed by certain populations via differential gene expression analysis (DGE) between these groups. Inferring cell type identity via marker genes uncovered in groups. If samples from different conditions were pooled together, performing a DGE to compare the expression profiles of cell populations also across conditions.

Regarding the V(D)J data analysis:

1. **Quality check of the original sequences**: Evaluating raw read sequencing quality, assessing nucleotide composition of the reads, K-mer over-representation, duplication rate, and adapter content.
2. **Mapping to the V(D)J reference transcriptome**: Aligning sequencing reads to the reference genome. Quantifying the number of reads mapped to constant and complementarity determining regions (CDRs) of TCR/BCR.
3. **Contig assembly and annotation**: Assembling reads into longer contigs to reconstruct the full TCR/BCR sequence. Annotating contigs by aligning them to V, D and J segments, and by identifying the CD3R sequences.
4. **T and B cell barcode identification and extraction**: Selecting unique cell barcodes corresponding to productive and confident contigs, indeed only T and B cells produce fully rearranged transcripts that contain both a V and a C segments.
5. **Clonotype generation**: Cells with minimal CDR3 sequence mutations are labeled as belonging to the same clonotype by assigning them a unique clonotype ID.
6. **Mapping of clonotypes**: Mapping the identified clonotypes onto dimensionality reduced space generated from GEX modality (e.g., PCA, UMAP or t-SNE) to facilitate the characterization of their transcriptional profile and clustering within immune cell populations.

Advanced (optional) analysis steps regarding the V(D)J data analysis include the following:

7. **Identification of expanded clones**: Identified clonotypes with the same clonotype ID are grouped together to define clonal cells within the same subjects and/or across multiple subjects. Clone size is measured by the number of cells sharing the same clonotype.

8. **Clonotypes characterization**: Identified clonotypes are characterized within the same subjects and/or across multiple subjects based on V and J segment usage vectors, CDR3 length, repertoire overlap and diversity.

Access modality available

- Access to facility service
- Access to facility service including training

Requested inputs from users

All fastq files associated to all the samples must be provided, together with the corresponding md5 files (unless sequencing is performed by the National Facility for Genomics).

The user shall provide a table listing all biological and experimental conditions in the study and all samples belonging to each condition, ensuring that the sample names exactly match the names of the provided fastq files (see Metadata file example).

Unless the organism under study is human or a model organism, the user shall provide a reference to the annotated reference genome to be used for analysis.

Technical requirements

The libraries for all the samples must have been prepared using the same kit and barcoding strategy, ideally in the same sequencing run. We recommend at least 1000 cells per sample and a minimum of 50.000 reads per cell, with a Q30 cutoff of 80% for GEX libraries and at a minimum of 5.000 reads per cell, with a Q30 cutoff of 80% for V(D)J libraries.

Results

The National Facility for Data Handling and Analysis will deliver the following results to the users:

1. Raw and filtered fastq files for each sample and modality.
2. BAM files for each sample and modality.
3. Count matrices containing expression values for each gene in each cell (in .h5ad).
4. Tables containing high-level description of each clonotype for each cell (in .csv).
5. Complete reports (in interactive HTML and publication-ready PDF formats) describing the quality of the data, all the analysis performed on the dataset, and their results. Plots (e.g. alluvial plot, Circos plots, etc.) and tables included to the report will also be provided as separate files.
6. Python objects (.h5ad) containing the processed data.
7. Pipeline and scripts used to perform the analysis, if applicable.

The facility will also assist the user in submitting raw data to public repositories, as stipulated in the *National Facilities Access Rules*.

Combined services

This service can be combined with the following services offered by the National Facility for Genomics:

G-009/012/014 - Single-cell Immune profiling-V(D)J (10X Genomics)

To access the combined services, please submit an application to the National Facility for Genomics requesting data analysis.

NF62.03.04 Single-cell multiome (ATAC + gene expression)

Service description

Single-cell multiome sequencing (scRNA-seq + scATAC-seq) is molecular biology technique used to analyze both gene expression and chromatin accessibility of individual cells/nuclei at a high resolution, allowing the resolution of cellular heterogeneity within a biological sample.

Unlike bulk RNA and ATAC sequencing, which averages gene expression and chromatin accessibility across many cells, multiome sequencing enables the identification of distinct cell types, states, and subpopulations both in terms of transcriptional and epigenetic profiles. This technology is particularly useful in studying various processes and biological mechanisms including developmental processes, tumorigenesis, and immunological memory establishment. It is widely applied in biomedical research to advance personalized medicine, immunology, cancer research, and tissue regeneration studies.

Multiome datasets include two different assays: gene expression (GEX) and chromatin accessibility (ATAC).

The standard bioinformatics analysis for multiome datasets comprises the following steps:

1. For the GEX modality, refer to Section 0 NF62.03.01 scRNA-seq analysis.
2. For the ATAC modality, refer to Section 0 NF62.03.02 scATAC-seq analysis.

Advanced (optional) analysis steps on single-cell multiome data include the following:

1. **Dataset integration**: Integrating data from both modalities into a single dataset on which the overall analysis is performed. Grouping of the two modalities into integrated datasets will depend on the experimental design and project requirements.

Access modality available

- Access to facility service
- Access to facility service including training

Requested inputs from users

All fastq files associated to all the samples must be provided, together with the corresponding md5 files (unless sequencing is performed by the National Facility for Genomics).

The user shall provide a table listing all biological and experimental conditions in the study and all samples belonging to each condition, ensuring that the sample names exactly match the names of the provided fastq files (see Metadata file example).

Unless the organism under study is human or a model organism, the user shall provide a reference to the annotated reference genome to be used for analysis.

Technical requirements

The libraries for all the samples must have been prepared using the same kit and barcoding strategy, ideally in the same sequencing run. We recommend at least 1000 cells per sample and a minimum of 50.000 reads per cell for both modalities, with a Q30 cutoff of 80%.

Results

The National Facility for Data Handling and Analysis will deliver to the users the following files:

1. Raw and filtered fastq files for each sample and modality.
2. BAM files for each sample and modality.
3. Count matrices containing expression values for each gene in each cell.
4. Raw and normalized peak-by-cell matrices containing peaks for each region of the genome in each cell.
5. Complete reports (in interactive HTML and publication-ready PDF formats) describing the quality of the data, all the analysis performed on the dataset, and their results. Plots (e.g. heatmaps, volcano plots, PCA/MDS) and tables included to the report will also be provided as separate files.
6. Python objects (.h5ad) containing the processed data.
7. Pipeline and scripts used to perform the analysis, if applicable.

The facility will also assist the user in submitting raw data to public repositories, as stipulated in the *National Facilities Access Rules*.

Combined services

This service can be combined with the following services offered by the National Facility for Genomics:

G-010/012/013 – Single-cell multiome ATAC + Gene expression (10X Genomics)

To access the combined services, please submit an application to the National Facility for Genomics requesting data analysis.

NF62.03.05 Spatial transcriptomics (10X Visum platform)

Services description

The Visium Spatial Gene Expression solution from 10X Genomics enables spatial profiling of gene expression within intact tissue sections. This protocol allows for the analysis of gene expression while preserving the spatial context of cells within a tissue sample. It provides valuable insights into spatially distinct gene expression patterns and cell-type localization, facilitating a deeper understanding of tissue organization, disease progression, and cellular microenvironments.

The standard bioinformatics analysis for a Visium Spatial Gene Expression dataset comprises the following steps:

1. **Quality check of the original sequences**: Evaluating sequencing quality of raw reads, assessing nucleotide composition of the reads, K-mer over-representation, duplication rate, and adapter content.
2. **Image segmentation**: Processing and segmenting the high-resolution images of tissue slices to consider only Visium spots covered by tissue. Only the reads derived from those spots are retained in the analysis.

3. **Cell barcodes identification and extraction**: Identifying cell barcodes associated to the tissue-covered spots, extracting transcripts associated to these cells from sequencing reads to undergo subsequent analysis.
4. **Mapping to the reference genome and quantification of gene expression**: Aligning reads to a reference genome, generating a BAM file. Quantifying the number of reads mapped to each gene in order to measure gene expression per cell.
5. **Normalization**: Normalizing expression data to account for technical artifacts while preserving biological variance, such as heterogeneous cell density across various parts of the tissue. The most variable genes are identified and used for further steps of the analysis.
6. **Dimensionality reduction and clustering analysis**: Applying the standard single-cell analysis workflow based on dimensionality reduction techniques (e.g., PCA and UMAP) and unsupervised clustering algorithms to each sample to identify distinct groups of spots. Performing differential gene expression analysis (DGE) between these groups to detect marker genes specifically expressed in certain clusters. Coloring the tissue image according to the clusters assigned to each spot, thus highlighting the spatial distribution of the different groups of spots within the slide.

Advanced (optional) analysis steps include the following:

1. **Multi-samples integration**: If multiple samples are processed, their expression data could be integrated into one single gene expression dataset on which the standard single-cell analysis workflow is applied. In this case, dimensionality reduction and unsupervised clustering are performed on this integrated dataset, and the obtained clusters are projected to each individual tissue slide.
2. **Spot deconvolution and/or signatures evaluation**: Given than each spot of the 10X Visium platform usually embeds more than one single cell, the cell type proportions composing each spot could be inferred by disentangling its mixed gene expression signals. This could be done using a matched single-cell RNA-seq dataset produced in the same experimental conditions (recommended) or relying on publicly available data. Expression of specific genes and signatures could also be evaluated and the spatial distribution of the corresponding scores is correlated with the identified clusters.

Access modality available

- Access to facility service
- Access to facility service including training

Technical requirements

All fastq files associated to all the samples must be provided, together with the corresponding md5 files (unless sequencing is performed by the National Facility for Genomics).

The user shall provide a table listing all biological and experimental conditions in the study and all samples belonging to each condition, ensuring that the sample names exactly match the names of the provided fastq files (see Metadata file example).

Unless the organism under study is human or a model organism, the user shall provide a reference to the annotated reference genome to be used for analysis.

High-resolution images of each considered tissue slice must also be provided in .tiff format.

The preparation of all the samples must have been performed using the 10X Visium or Visium HD platform starting from Fresh Frozen or FFPE tissues.

Results

The National Facility for Data Handling and Analysis will deliver to the users the following files:

1. fastq files for each sample.
2. BAM files for each sample.
3. Count matrix containing expression values for each gene in each spot.
4. Complete reports (in interactive HTML and publication-ready PDF formats) describing the quality of the data, all the analysis performed on the dataset, and their results. Plots (e.g. spatial feature plots, UMAP, violin plots) and tables included to the report will also be provided as separate files.
5. Pipeline and scripts used to perform the analysis.

The facility will also assist the user in submitting raw data to public repositories, as stipulated in the *National Facilities Access Rules*.

Combined services

This service can be combined with the following services offered by the National Facility for Genomics:

G-015/016 – Visium Spatial gene expression from Fresh-Frozen or FFPE tissues (10X Genomics)

To access the combined services, please submit an application to the National Facility for Genomics requesting data analysis.

Metadata file example

Metadata about sequenced samples should be provided in a table (in tab-delimited or Excel format) with the following structure:

Condition	Sample	FASTQ1	FASTQ2	Var1	Var2	Var…
control	sample1	sample1_R1.fastq.gz	sample1_R2.fastq.gz			
control	sample2	sample2_R1.fastq.gz	sample2_R2.fastq.gz			
treatment	sample3	sample3_R1.fastq.gz	sample3_R2.fastq.gz			
treatment	sample4	sample4_R1.fastq.gz	sample4_R2.fastq.gz			

- The first three columns are required and should be named Condition, Sample, and FASTQ1 respectively.

- The fourth column can be omitted in the case of single-end sequencing. If present, it should be named FASTQ2.
- Condition names and sample names should only contain letters, digits, and the underscore character. Please do not include spaces, symbols, or special characters.
- Additional variables associated with each sample can be added to the table, and will be included in the final reports.

For metagenomic projects, please add the primers used to amplify the target regions, as in the following example:

Condition	Sample	FASTQ1	Forward Primer	Reverse Primer
control	sample1	sample1_R1.fastq.gz	GTGYCAGCMGCCGCGGTAA	GGACTACNVGGGTWTCTAAT
control	sample2	sample2_R1.fastq.gz	GTGYCAGCMGCCGCGGTAA	GGACTACNVGGGTWTCTAAT
treatment	sample3	sample3_R1.fastq.gz	GTGYCAGCMGCCGCGGTAA	GGACTACNVGGGTWTCTAAT
treatment	sample4	sample4_R1.fastq.gz	GTGYCAGCMGCCGCGGTAA	GGACTACNVGGGTWTCTAAT

Glossary of terms

Term	Definition
Splice-aware aligner	Tool that aligns RNA-seq reads, accounting for exon-exon junctions (e.g., STAR, HISAT2).
Pseudoalignment	Assigns reads to transcripts without full alignment (e.g., used by Kallisto, Salmon).
Q30	Quality score indicating a 1 in 1000 error rate in sequencing (99.9% accuracy).
Amplicon Sequence Variants (ASVs)	Unique single-nucleotide precision DNA sequences from amplicon data (e.g., 16S rRNA), alternative to traditional OTUs for microbial diversity analysis.
Alpha Diversity	Measure of species diversity within a single sample, considering species richness (number of species) and evenness (distribution of species).
Beta Diversity	Measure of species diversity between samples
Pseudo-bulk	Data aggregation technique where cell-level data are grouped to create virtual bulk samples for statistical analysis.

Peak calling	Identifying regions in the genome where sequencing reads are highly concentrated, indicating active or accessible DNA sites.
Blacklist region	Genomic regions known to produce unreliable or artifact signals in sequencing experiments, typically excluded from analysis to avoid misinterpretation of data.
Spot deconvolution	Estimating the proportions of different cell types within a single spatial transcriptomics data spot, as spots often contain multiple cells.

Cited databases

Database name	Database description	URL
KEGG	Linking genes and proteins to metabolic and disease-related functions	https://www.kegg.jp/
Reactome	Detailed gene-protein relationships and cross-pathway interactions	https://reactome.org/
Biocarta	Molecular mechanisms underlying well-known biological pathways	https://maayanlab.cloud/Harmonizome/dataset/Biocarta+Pathways
Hallmark	Gene sets of high-level biological states or processes	https://www.gsea-msigdb.org/gsea/msigdb/
IPA	Manually curated pathways to predict causal relationships and identify regulatory networks	Provided as software
miRbase	Repository for miRNA sequences and annotations	https://www.mirbase.org/
DIANA-microT-CDS	Predicts miRNA targets in CDS and 3' UTRs.	https://dianalab.e-ce.uth.gr/microt_webserver/
MicroCosm	miRNA target identification and annotation.	https://tools4mirs.org/software/mirna_databases/microcosm-targets/
miRanda	miRNA target prediction using sequence analysis.	https://bioweb.pasteur.fr/packages/pack@miRanda@3.3a
miRDB	Predicts miRNA targets using machine learning.	http://www.mirdb.org/
PicTar	Conserved miRNA target prediction in animals.	https://pictar.mdc-berlin.de/

TargetScan	miRNA target prediction using conservation.	https://www.targetscan.org/
miRecords	Curated database of miRNA-target interactions.	http://c1.accurascience.com/miRecords/
miRTarBase	Experimentally validated miRNA targets.	https://mirtarbase.cuhk.edu.cn/~miRTarBase/miRTarBase_2019/php/index.php
TarBase	Manually curated miRNA-target interactions.	https://dianalab.e-ce.uth.gr/tarbasev9
Clinvar	Reports of human variants classified for diseases and drug responses.	https://www.ncbi.nlm.nih.gov/clinvar/
OMIM	Catalog of human genes and genetic disorders.	https://www.omim.org/
ACMG	Genetic variant classification guidelines.	https://www.acmg.net/
COSMIC	Catalog of somatic mutations in cancer.	https://cancer.sanger.ac.uk/cosmic
Civic	Clinical evidence for cancer variants.	https://civicdb.org/
OncoKB	Precision oncology knowledge base.	https://www.oncokb.org/
AMP	Molecular testing guidelines for cancer.	https://www.amp.org/
SILVA	rRNA sequences for taxonomy and phylogeny	https://www.arb-silva.de/
UNITE	Fungal ITS sequences for taxonomy	https://unite.ut.ee/

Tools used

The following table shows examples of software tools that may be used in our analysis pipelines. The specific tools used in a project will be listed in the final analysis report.

Tool	Purpose	Service
FastQC	Quality control of sequencing reads, checking for adapter content and base quality.	Bulk RNA-Seq, scRNA-Seq, scATAC-Seq, WGS, WES, ChIP-Seq, Methyl-Seq, Microbiome analysis, miRNA analysis
MultiQC	Aggregating QC metrics for different pipeline steps.	Bulk RNA-Seq, scRNA-Seq, scATAC-Seq, WGS, WES, ChIP-Seq, Methyl-Seq, Microbiome analysis, miRNA analysis
TrimGalore!	Trimming low-quality bases and adapters from sequencing reads.	Bulk RNA-Seq, scRNA-Seq, scATAC-Seq, WGS, WES, ChIP-Seq, Methyl-Seq
FastP	Trimming low-quality bases and adapters from sequencing reads.	WGS, WES, miRNA analysis
Cutadapt	Trimming low-quality bases and adapters from sequencing reads.	Microbiome analysis
SAMtools	Manipulating and analyzing BAM/CRAM files from sequencing data.	All sequencing analyses (general-purpose tool)
BEDTools	Tools to analyze BAM/BED files from sequencing data.	ChIP-Seq, scATAC-Seq, WGS, WES, Methyl-Seq
Picard	Tools for BAM file manipulation and quality assessment.	WGS, WES, Bulk RNA-Seq
Bowtie2	Alignment of short reads to a reference genome.	WGS, WES, ChIP-Seq, scATAC-Seq
Bowtie	Alignment of short reads to mature miRNAs and miRNA precursors (hairpins).	mirRNA analysis
BWA-MEM2	Alignment of short reads to a reference genome.	WGS, WES, ChIP-Seq, scATAC-Seq
STAR	Splice-aware alignment of RNA-seq reads to the reference genome.	Bulk RNA-Seq, scRNA-Seq
Salmon	Pseudoalignment and quantification of transcript abundance.	Bulk RNA-Seq, scRNA-Seq

DESeq2	Differential gene expression analysis for RNA-seq count data.	Bulk RNA-Seq, scRNA-Seq, miRNA analysis
edgeR	Differential expression and count-based RNA-seq analysis.	Bulk RNA-Seq, scRNA-Seq, miRNA analysis
enrichR	Functional enrichment of GO and pathway data.	Bulk RNA-Seq, scRNA-Seq, ChIP-Seq, scATAC-Seq
ClusterProfiler	Statistical analysis of GO and pathway data.	Bulk RNA-Seq, scRNA-Seq, ChIP-Seq, scATAC-Seq
CellRanger	Preprocessing, alignment, and quantification of RNA-seq, ATAC-seq, and TCR-seq at the single-cell level.	scRNA-Seq, scATAC-Seq, Single-cell immune profiling, Single-cell multiome
Scanpy	Analysis and visualization of single-cell and spatial RNA-seq data.	scRNA-Seq, Spatial transcriptomics
muon	Analysis and filtering of single-cell ATAC-seq data for visualization of quality metrics.	scATAC-Seq, Single-cell multiome
Scrublet	Doublet detection in single-cell sequencing.	scRNA-Seq, scATAC-Seq, Single-cell multiome
DoubletFinder	Doublet detection in single-cell sequencing.	scRNA-Seq, scATAC-Seq, Single-cell multiome
SCVI	Integrating multiple samples, layers, and modes in single-cell data.	scRNA-Seq, scATAC-Seq, Single-cell multiome
Harmony	Integration of multiple samples in single-cell datasets.	scRNA-Seq, scATAC-Seq, Single-cell multiome
Space Ranger	Preprocessing, alignment, and quantification of spatially resolved RNA-seq data.	Spatial transcriptomics
Loupe Browser	Manual segmentation of spatial transcriptomics images and data visualization.	Spatial transcriptomics
HaplotypeCaller	Germline variant calling (SNP/indel).	WGS, WES
Mutect2	Somatic variant calling (SNP/indel).	WGS, WES
Strelka	Somatic variant calling (SNP/indel).	WGS, WES
Manta	Structural variant detection.	WGS, WES
DeepVariant	Germline variant calling (SNP/indel).	WGS, WES

MACS2	Peak calling for ChIP-seq, ATAC-seq, and single-cell ATAC-seq data.	ChIP-Seq, scATAC-Seq
SEACR	Peak calling for low-background assays like CUT&RUN.	CUT&RUN
HOMER	Motif discovery and annotation of regulatory regions in genomic datasets.	ChIP-Seq, ATAC-Seq, scATAC-Seq
Scirpy	Analysis and visualization of TCR/BCR data for immune profiling at the single-cell level.	Single-cell immune profiling (VDJ)
IGV	Interactive visualization of genomic data.	WGS, WES, Bulk RNA-Seq, scRNA-Seq, scATAC-Seq, ChIP-Seq, Methyl-Seq, Spatial transcriptomics
DADA2	Microbiome data analysis and Amplicon Sequence Variant (ASV) inference from microbiome data.	Microbiome analysis
QIIME2	Microbiome data analysis, including taxonomic classification and diversity analysis.	Microbiome analysis
Phyloseq	R object for working microbiome data.	Microbiome analysis
ANCOM	Analysis of composition of microbiomes.	Microbiome analysis
PICRUSt2	Phylogenetic Investigation of Communities by Reconstruction of Unobserved States.	Microbiome analysis
miRDeep2	Identification of novel and known miRNAs.	miRNA analysis
miRTrace	Quality control for small RNA-seq data.	miRNA analysis