HUMAN TECHNOPOLE

NATIONAL FACILITY FOR GENOMICS

CALL FOR ACCESS

26-G-ROUND-3

Table of Contents

1. SUMMARY OF CHANGES TO THIS CALL FOR ACCESS 3

2. INTRODUCTION 3

3. TERMS AND DEFINITIONS 4

4. APPLICATION TYPE 6

5. ELIGIBILITY AND ADMISSIBILITY 6

6. APPLICATION 8

6.1 Application content and format 8

6.2 Accountability of Applicants and host institutions 11

7. APPLICATION SUBMISSION METHODS, CALL DEADLINE AND EVALUATION PERIODS 12

8. EVALUATION OF APPLICATIONS 13

9. RESUBMISSION OF PROPOSALS PREVIOUSLY EVALUATED AS "RESERVE LIST" OR "ACCESS NOT GRANTED" 16

9.1 Resubmission of a proposal evaluated as "Reserve list" in the previous round 16

9.2 Resubmission 17

10. PROJECT CONTINUATION APPLICATIONS 17

11. AFTER ACCESS HAS BEEN APPROVED 18

12. AFTER ACCESS HAS BEEN COMPLETED 18

12.1 Reporting and feedback 18

12.2 Ownership of results and research outputs 19

13. CONTACTS 19

14. REFERENCES 19

15. CHANGES TO THE CALL 19

ANNEX I: LETTER OF INSTITUTIONAL ENDORSEMENT template 20

ANNEX II 22

A. Project proposal formats and template 22

B. Biosketch template 23

ANNEX III: SERVICE LIST 24

1. SUMMARY OF CHANGES TO THIS CALL FOR ACCESS

Compared to previous rounds, the following changes have been introduced in this call:

- *Specific procedures for the **resubmission** of proposals already evaluated in previous calls, and for the submission of **project continuation** (formerly known as "follow-up"), i.e. research proposals that build on the findings and results of a previously approved Access, have been implemented in the application process (section 9 and section 10).*
- ***Changes in the available services for this call (Annex III).***
- *Rules for the "Responsible use of generative Artificial Intelligence tools in proposal preparation and evaluation" (link to the document).*
- *Eligibility and admissibility (section 5).*
- *Application content and **Accountability of Applicants and host institutions** (section 6).*

We recommend that the applicants read the present document and the guidelines (link) carefully before applying to the current call for Access.

2. 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 the NFs are available through regular open calls for Access that are published yearly on the HT website (link) and are entirely subsidized by HT through an indirect financing system 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 subsidized. This includes 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).

2.1 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 (Physical Access)**: 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 (Remote Access)**: 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. In this case, physical Access may be approved.
- **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. **Note that in the application, a motivation underlying the request for "Access with training" must be specified.**

3. TERMS AND DEFINITIONS

3.1 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.

3.1.1. Physical Access (or in-person Access)

"Physical Access" applies when a User accesses a NF or a service by being physically present at the selected NF. In order to obtain physical Access, the User must complete all necessary security and health, safety and environment (HSE) checks, complete all required training and be provided with all the necessary insurances. Physical Access is also required for on-site training.

3.1.2. Remote Access

"Remote Access" means Access to a NF or service without the User being physically present at the NF. Typically, NF staff perform the experimental protocols entailed by the service and return the Data and results produced to the User together with any resulting samples (when appropriate). If applicable, "remote Access" can also include remote analysis of the samples by the User after sending them to the NF.

IMPORTANT - Additional information and details about **Access** are reported in the NF Access rules available on the NFs dedicated webpage (link).

3.2 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.

3.3 Principal Investigator, Junior and Established PI

"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.

3.4 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 5.

3.5 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.

4. 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, substantiated by robust preliminary data and/or requesting a standard service based on existing, well validated experimental conditions and/or protocol(s).

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).

Additional details related to the application contents are provided in section 6.

5. ELIGIBILITY AND ADMISSIBILITY

PIs, as defined in section 3.3 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:
  - State funded public universities, listed under the following link.
  - Specialized superior graduate schools or institutions, listed under the following link.
  - Legally recognized non-public universities, listed under the following link.
  - 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 Formazione - INVALSI;
o) Istituto Nazionale di Documentazione, Innovazione e Ricerca Educativa - INDIRE;
p) Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria - CREA;
q) Agenzia Nazionale per le Nuove Tecnologie, l'energia e lo Sviluppo Sostenibile - ENEA;
r) 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);
s) Istituto Nazionale di Statistica - ISTAT;
t) Istituto Superiore di Sanità - ISS;
u) Istituto Superiore per la Protezione e la Ricerca Ambientale - ISPRA, ferme restando le disposizioni di cui alla legge 28 giugno 2016 n.132;
v) Istituto nazionale per l'assicurazione contro gli infortuni sul lavoro – INAIL.

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, host institution or elsewhere.

Applicants shall confirm the **economic and scientific feasibility** for the aspects of the project to be performed outside the NFs.

Applicants will need to certify that **samples/ data and relevant authorisations are available at the moment of application or no later than two (2) months** from receiving Access approval (refer to section 8). **If samples/ data and/ or relevant ethical and legal authorisation(s) for their use will not be provided within this time frame, the request for Access will be automatically rescinded and PI will need to reapply at a subsequent call**.

Applicants **cannot request Access for the same service** if an approved Access is ongoing (i.e., Access that has been granted in a previous call for Access and is not yet completed). 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** in the dedicated section of the application form.

A PI submitting an application to this call for Access **cannot request Access to other NFs** (i.e., cannot participate to other 26-ROUND-3 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 one before receiving notification of the results. If an application is erroneously submitted, this will be rejected at the administrative review stage.

Applications must be **written in English**, they must be **complete** (i.e., consist of all the requested elements and information) and **abide to all administrative and technical requirements** (e.g., proposal and/or CV format, mandatory declarations, technical requirements for the services, samples/ data availability, samples/ data requirements,

including but not limited to number of samples to be analysed, and research data management plan).

Incomplete applications or applications that do not meet the call requirements will be considered not admissible and will be rejected at the administrative review stage.

6. APPLICATION

6.1 Application content and format

All applications must be submitted through the online portal PICA (link), following carefully the guidelines (link) and filling all mandatory fields of the application form. Applicants should also consider the information provided in section 7 before initiating the submission process.

The main contents of the application form are:

1. **Applicant's general information**.
2. **Justification for requesting Access to the NF,** explaining why the project cannot be performed at the Applicant's institution.

   We recommend that the Applicant contacts the institute representative, i.e., the individual authorized to sign the Letter of Institutional Endorsement (LoE) to ensure that the most appropriate justification is provided and that it aligns with the one included in the LoE (point 6.c, below).

   If the justification provided by the Applicant does not match the one provided by the host institution through the LoE, the application will be rejected during the administrative review stage.

   The Applicant must choose the option that best applies and must provide further details supporting the choice:

   - The requested service/ technology is not available at the host institution.
   - The requested service/ technology cannot be performed at the host institution or elsewhere at an affordable cost.
   - The requested service/ technology is available at the host institution, but the necessary expertise is lacking.
   - The requested service/ technology is available at the host institution, but the service cannot be performed in a timeframe or scale compatible with the experimental requirements.

   NOTE: The Standing Independent Evaluation Committee (SIEC), in charge of the evaluation procedure, may reserve the option to contact the host Institution and its core facilities to confirm the justification provided. **Read carefully information in section 6.2**.

3. **Project Title, Abstract** and **Area of Research**, to be inserted in the dedicated section on the application form (Max 1500 characters including spaces).

   IMPORTANT - Do not share any confidential information: project title and abstract will be published on HT website as provided for by the Convenzione, Art 5, Comma 5 (for more information, the Convenzione is available at this link or as pdf).

4. **Project proposal**, to be uploaded in PDF format in the dedicated section on the application form, shall include the following (see also Annex II):

   - Title
   - Significance
   - Innovation
   - Approach, including aims, preliminary data in support of the proposed experiments, experimental design and anticipated results
   - Environment, including facilities and resources available to support the aspects of the project to be performed elsewhere (i.e., outside the NF).

   The **use of generative AI tools** in the preparation of proposals is governed by the rules and recommendations outlined in the "**Rules for the Responsible Use of Generative Artificial Intelligence in Proposal Preparation and Evaluation for Calls for Access to National Facilities**" (link to the document), which is an integral part of this call. Applicants are also required to declare any use of generative AI by responding to the relevant questions in the application form.

   Below are the **mandatory formats** for the proposal, also based on the type of application (see sections 4, 9 and 10 for additional details):

   a. **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 (e.g., ID123456_PROPOSAL_Rossi)

   b. **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 (e.g., ID123456_PROPOSAL_Rossi)

   c. For **Resubmissions** (section 9), a **Resubmission cover page** must be included in the proposal. The total length of the application is hence increased by one page with respect to the format illustrated above. Refer to section 9 for important details on Resubmissions.

   d. For **Project Continuation** (section 10), the **Prior Results and Research Advancement cover page** must be included in the proposal. The total length of the application is hence increased by one page with respect to the format illustrated above. Refer to section 10 for important details on Project continuation.

   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 as PDF file, shall be in English and produced by using the **NIH biosketch template available in Annex II or at this link.** For support in drafting the CV, you can still refer to the NIH website Biographical Sketch Common Form (nih.gov). Please note that having an eRA COMMONS USERNAME is NOT required.

   Biosketch mandatory format: max 4-5 pages, page format: A4, Font type: Arial, Font size: at least 11, Line spacing: single, Margins 2 cm side/ 1.5 bottom.

   Accepted file format: PDF. Max size: 30MB - Name the file as APPLICATION ID_CV_Surname (e.g., ID123456_CV_Rossi).

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 3.3).*
   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 (refer to section 8), in order to process samples and data through the NFs.*
   c. *Justification for requesting Access to the NF explaining why the project cannot be performed at the Applicant's institution. We recommend that the Applicant contacts the institute representative, i.e., the individual authorized to sign the Letter of Institutional Endorsement (LoE) to ensure that the most appropriate justification is provided and that it aligns with the one included in the LoE.*

      If the justification provided by the Applicant does not match the one provided by the host institution, the application will be rejected during the administrative review stage.

      The host institution must select ***only one option****, from the four listed below — please do not alter the statements:*

      - *The requested service/ technology is not available at the host institution.*
      - *The requested service/ technology cannot be performed at the host institution or elsewhere at an affordable cost.*
      - *The requested service/ technology is available at the host institution, but the necessary expertise is lacking.*
      - *The requested service/ technology is available at the host institution, but the service cannot be performed in a timeframe or scale compatible with the experimental requirements.*

      NOTE*: The SIEC, in charge of the evaluation procedure, may reserve the option to contact the host institution and its core facilities to confirm the justification provided.* **Read carefully information in section 6.2**.
   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, host 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* (link).

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

Name the file as APPLICATION ID_ENDORSEMENT_Surname (e.g., ID123456_ENDORSEMENT_Rossi).

IMPORTANT: *Do not modify, add or remove any part of the Letter of Endorsement and fill in all the required information. Please, make sure that the signature of the letter is done in compliance with rules and procedures of your host institution.*

7. **Technical information**, to be filled in in the dedicated section(s) of the application form, indicatively including:
    a. *Requested service(s), as described in Annex III of this call.*
    b. *Samples technical information.*
    c. *Additional data, useful for technical feasibility analysis (if applicable).*
    d. *Whether the entire sample/ data set is already available, or will be available no later than two (2) months from receiving Access approval (refer to section 8).* ***Please note that if samples/ data and/ or relevant ethical and legal authorisation(s) for their use will not be provided within this time frame, the request for Access will be automatically rescinded and PI will need to reapply at a subsequent call.***
    e. *Resources and expertise to receive and process the output – 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 form.

Information provided in points 1 to 6 (application contents) is used for the **eligibility and admissibility check.**

Information provided in point 7 (technical information of the application) 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 SIEC to assess its scientific merit, according to the evaluation process illustrated in section 8.

6.2 Accountability of Applicants and host institutions

The Applicant and the host institution are solely responsible for the accuracy, completeness, and truthfulness of all data and information provided in the application. By submitting the application, they declare that all statements and supporting documentation are true and correct, acknowledging the liability and the sanctions stated in art. 75 and 76 of the DPR 445/2000 for false affirmations and mendacious declarations, including the rejection of the application and/ or the loss of any granted benefits.

In case of Letters of Endorsement deemed not compliant or presenting inconsistencies in the submitted information (through the application), the User Access Office can conduct, on behalf of the SIEC, further verifications and/or to reject the application at the administrative review stage. This information can be also taken into consideration by the SIEC during the evaluation process and final approval decision.

Furthermore, acting on behalf of the SIEC, UAO may conduct **audits of participating institutions** to assess availability of any given or all services and ensure that these match the information submitted in their submitted applications.

In the event of violations and/or inaccurate or false declarations, the SIEC reserves the right to ban the Applicant and/or the host institution from participating in future Calls for Access for a period of up to twenty-four (24) months.

7. 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.

Application guidelines containing important information related to the submission procedure are available at this link.

This call for Access (Call ID: 26-G-ROUND-3) will open on the 1st of October 2026 (13:00 CET) and will close on the 31st of January 2027 (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 webpages (call for Access; services).

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

Samples/ data as well as relevant authorisation for their use, **shall ideally be available by when the application is submitted, but categorically not later than two (2) months after Access approval** (refer to section 8). When the project foresees the analysis of more than one batch of samples/ data, similarly, the first batch should be available when the application is submitted or not later than two (2) months after Access approval.

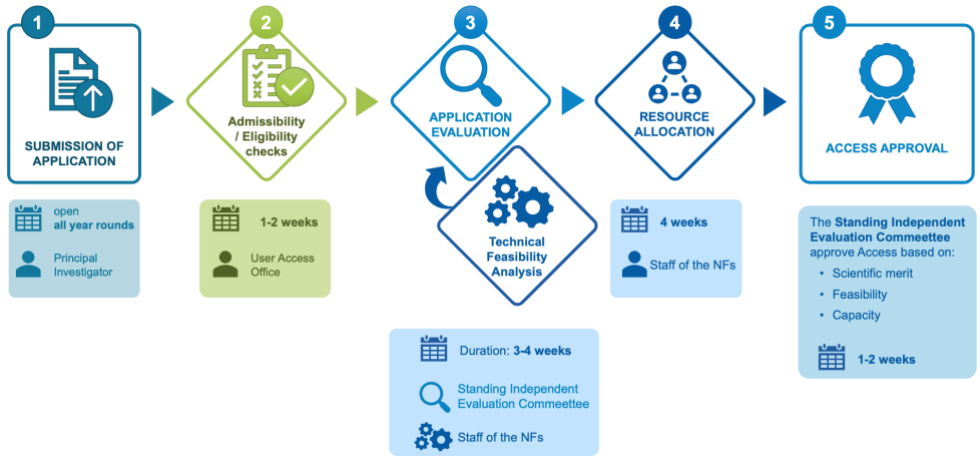
8. EVALUATION OF APPLICATIONS

Overview

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 two (2) SIEC members, who will act as Chairs, along with a variable number of appointed external Reviewers selected according to the number of submitted proposals and their corresponding areas of expertise.

Below is a scheme describing the evaluation steps and the **indicative timeline for the process**. Evaluation results will be communicated through the PICA portal within 8 to 12 weeks after the closing of the call for Access.



The table below reports the indicative timeline for this call.

Opening	Closure	Evaluation	Access Approval
01.10.2026	31.01.2027	February – April 2027	By April 2027

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.

8.1 Triage

Based on the number of applications, if the requested services exceed by a factor of two (2) the estimated capacity of the NF, a triage will be applied within each application category.

Triage criteria will include:

1. **Justification for requesting Access to the NF**: priority will be given to researchers who do not have direct Access to the service/ technology at their home institution. For the Triage, the Justification provided in the Letter of Endorsement (section 6, point 6.c) will be taken into consideration.
2. **Ongoing and previous support received by the NFs**: priority will be given to researchers who do not have ongoing Access to the same NF they are applying for, and/ or who have not been granted Access to it in the two calls preceding the current one *(e.g. in the call 26-G-ROUND-3 priority will be given to researchers who do not have any ongoing access to the NF for Genomics and/or who were not granted access in calls 26-G-ROUND-1 and 26-G-ROUND-2).*

Applicant career stage (Established and Junior PIs) will be considered during the triage phase to ensure alignment with the requirements of the evaluation procedure (section 8.2).

Should the number of requested services still exceed the allowable estimated limit after having applied the triage, as a tool of last resort, a lottery will be applied.

To ensure broader Access for all institutions across Italy, proposals submitted by a single Institution/Institute that are sent for evaluation should not exceed the 10% of the total for any given career-based category.

8.2 Evaluation procedure and criteria

The application is then sent to the Review Panel for assessing technical feasibility and scientific merit. A comprehensive analysis of the technical feasibility of the project, which is performed by the NF staff, is provided as supporting documentation (section 8.4).

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

The SIEC commits to allocating at least 50% of the available Access to applications from Junior PIs, while maintaining the Scientific merit as the primary criterion.

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.

8.3 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.

The use of **generative AI tools** in the review and evaluation of proposals is governed by the rules and recommendations outlined in the "Rules for the Responsible Use of Generative Artificial Intelligence in Proposal Preparation and Evaluation for Calls for Access to National Facilities" (link to the document), which is an integral part of this call. Reviewers are also required to declare any use of generative AI by responding to the relevant questions in the PICA portal.

8.4 Technical feasibility analysis

During the evaluation, the SIEC Chairs as well as the Reviewers will receive a report from the 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.

At this stage, the NF staff provides the SIEC Chairs with information on the resources needed (cost and time) to perform the proposed projects.

8.5 Evaluation results and Access approval

Applications with the highest scientific score that fulfil all technical requirements are approved for Access by the SIEC, subject to the capacity of the NF to host and execute the projects.

The SIEC commits to allocating at least 50% of the available Access to applications from Junior PIs, while maintaining the Scientific merit as the primary criterion, provided that a sufficient number of qualified proposals are received.

In case of comparable scores, for applicants that are at the same career level category, the SIEC Chairs will have the authority to rank the applications based on secondary parameters such as number of applications per Institution, previous Access, research outputs related to previous Access to the NFs (e.g., publications) originated by previous Access, budgetary considerations and geographical distribution.

Evaluation results – Access granted, Access conditionally granted, Reserve list, Access not granted, Rejected (not eligible or not feasible proposals), Excluded during triage – are communicated to the Applicant through the portal PICA.

A selected number of applications may be placed on a reserve list (i.e., waiting list in case of cancellations of Access granted projects). Applicants whose applications are placed on this list will receive additional information advising whether the project can be Access approved or should be resubmitted within the subsequent application window as resubmissions (see section 9).

9. RESUBMISSION OF PROPOSALS PREVIOUSLY EVALUATED AS "RESERVE LIST" OR "ACCESS NOT GRANTED"

9.1 Resubmission of a proposal evaluated as "Reserve list" in the previous round

Reserve list projects that are not accommodated in the round they have been evaluated for will be granted two options:

a. To be included in this new evaluation round, maintaining the same project and review score as their initial application. Here, the application cannot be modified and is not sent out again for evaluation.

To choose this option, the Applicant shall submit the request through the dedicated application call (Application for Reserve List Projects) accessible at this link, indicating the call and the ID of the reserve list project.

This option is provided only once, for the subsequent round only. An exception applies when the requested service is not available in the subsequent round. In such cases, the Reserve list application can be submitted to the first upcoming round of availability of the service *(e.g., the proposal has been submitted in the call 26-ROUND-1 and the evaluation outcome is "Reserve list"; the requested service is not available in the subsequent call 26-ROUND-2 but it is in the call 26-ROUND-3: the applicant can participate to the 26-Round-3 applying to the "Application for Reserve List Projects" call).*

b. To resubmit an updated version of the proposal (for example, including new preliminary data and/or taking into account Reviewer's comments), the Applicant should follow **the resubmission procedure described below** (section 9.2).

9.2 Resubmission

Applications that have been evaluated in a previous call, but that were not granted Access, or Reserve list projects without further Access being granted (Applicants that opt for option b above), can be submitted as a "**Resubmission**". Resubmission is not allowed for applications that were rejected (i.e., applications submitted to a previous round that were not eligible or not feasible) or excluded during triage in a previous evaluation round.

Of note, a **resubmission** is a proposal that has been **substantially improved** by adding new preliminary data, and/or by implementing Reviewers' suggestions and/or addressing concerns, etc. Applicants choosing the resubmission option shall include, in the proposal, the "Resubmission Cover page". **The Resubmission cover page must clearly describe how the proposal has been improved compared to the original version**.

IMPORTANT – The Resubmission Cover page must be included in the pdf proposal, as its cover page (Standard applications max 1+3 pages, PoC applications max 1+2 pages) and it must have the following format: *Max 1 page (Page format: A4, Font type: Arial, Font size: at least 11, Line spacing: single, Margins 2 cm side/ 1.5 bottom).*

Resubmissions that do not meet these requirements will be considered not admissible and the application will be rejected at the initial administrative review stage.

Resubmissions will enter the standard competitive evaluation procedure along with all proposals submitted to the current round (section 8).

Resubmissions are allowed only once, regardless of the round. A project proposal submitted multiple times (i.e., more than one) as a resubmission will be considered not eligible and will be rejected at the initial administrative review stage. Applicants can participate to future calls for Access submitting a new application.

10. PROJECT CONTINUATION APPLICATIONS

Applicants that would like to request Access for continuing a project that already benefited of an Access to the National Facilities, shall submit a "**Project continuation application**". These are research proposals that build on the findings and results of a previously approved Access, with the aim of further developing the research and advancing knowledge in the field.

The previously approved project has been conducted in any of the National Facilities, and it is ongoing or it has been completed (End of Access) at the time of the submission of the project continuation proposal. Please consider that, if the previously approved project is ongoing applicants cannot request **Access for the same service**. Before submitting an application requesting the same service of an ongoing Access, Applicant shall consult with the NF staff and confirm that the ongoing Access will be completed before the end of the evaluation round (refer also to section 5).

Applicants choosing to submit a project continuation application shall include, at the beginning of the proposal, the summary of **Prior Results and Research Advancement**. This summary must clearly outline the main achievements of the previously approved Access, citing any resulting publication (where applicable), and must explain how this project continuation proposal builds on and differs from the previously approved one, highlighting how it advances research knowledge.

IMPORTANT – The summary of "Prior Results and Research Advancement" must be included in the pdf proposal, as its cover page (Standard applications max 1+3 pages, PoC applications

max 1+2 pages) and it must have the following format: *Max 1 page (Page format: A4, Font type: Arial, Font size: at least 11, Line spacing: single, Margins 2 cm side/ 1.5 bottom).*

Project continuation applications that do not meet these requirements will be considered not admissible and the application will be rejected at the initial administrative review stage.

Project continuation applications will enter the standard competitive evaluation procedure along with all proposals submitted to the current round (section 8).

11. AFTER ACCESS HAS BEEN APPROVED

For requests where access is granted, Applicants accept the National Facilities Access rules. **Access rules** are detailed in the document: "NF Access Rules_Convenzione" (link) **which constitutes an integral part of this call**.

A kick-off meeting is organised after Access approval, in which 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/ data has(have) been provided, the NF User Access Office coordinates the signature of the required formal Agreements (e.g., Access Agreement, Collaboration Agreement, other), when required, and the project can commence.

Of note, no biological material and/or experimental support material may be transferred from Human Technopole to the PI or from the PI to HT prior to the project having officially started (refer to the Access rules for additional details).

If samples/ data and/ or relevant ethical and legal authorisation(s) for their use are not available within two (2) months after Access approval (refer to section 8), the request for Access will be automatically rescinded and PI will need to reapply at a subsequent call.

12. AFTER ACCESS HAS BEEN COMPLETED

12.1 Reporting and feedback

At the end of the activities carried out at the NF, and not later than three (3) months thereafter, if not differently agreed with the NF User Access Office, the Applicant must submit a short report to be published on the NFs website on the results obtained and the impact of the service on their research.

Moreover, a final report 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. The NF User Access Office will provide a template for the requested reports including the information required (activities performed, outcomes, impact on PI's research, plan for data sharing with scientific community, among others).

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. **Refer to the Access rules document (paragraph 15) for additional details.**

12.2 Ownership of results and research outputs

Research data obtained during Access shall be made available to the scientific community following the FAIR principles. It is requested that Applicants make raw data available to the scientific community by publishing them in an appropriate public repository. Applicant must inform the NF User Access Office (via email to national.facilities@fht.org) when and how the data are made public, providing a citable identifier.

In particular, the Applicant must communicate to the NF User Access Office (via email to national.facilities@fht.org) any **publication** acknowledging the NF and/or in which any member of the NF staff is listed as a co-author. Additional details about **ownership of the data, publication of the results and intellectual property rights** are reported in the Access rules document (paragraphs 13, 14).

13. CONTACTS

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

14. REFERENCES

NF Access Workflow_Convenzione (link)

NF Access Rules_Convenzione (link)

NF Access Agreement_Convenzione (link)

NF Application Guidelines_PICA portal (link)

15. 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 - please, do not modify, add or remove any section of the letter and fill in all the required information)

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 (*Title*)……., 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: **(*please select the one that applies: 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 Applicant's request to Access the National Facilities is justified for the following reason (***please mark the one that applies or best fits this application; only one choice is permitted***):

  ☐The requested service/ technology is not available at the Host Institution;

  ☐The requested service/ technology cannot be performed at the Host Institution or elsewhere at an affordable cost;

  ☐The requested service/ technology is available at the Host Institution, but the necessary expertise is lacking;

  ☐The requested service/ technology is available at the Host Institution, but the service cannot be performed in a timeframe or scale compatible with the experimental requirements.

- That relevant authorisations, declarations and accreditations from the competent authority(ies) have been obtained in order to process samples and data through Human Technopole OR that, if relevant authorisations, declarations and accreditations from the competent authority(ies) have not been obtained yet, they will be available before the

starting date of the project, and not later than 2 months after Access approval (refer to section 8 of the call text);

- That, if applicable, biological specimens have been obtained or will be 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 and/ or data were obtained or will be obtained from subjects who signed an 'informed consent', said informed consent allows or will allow that sequencing data and results are included in secure controlled Access databases and accessed/ used by authorized third parties;
- That, if applicable, copy of the relevant authorisations, declarations and accreditations will be provided at the moment of the application or not later than 2 months after Access approval (refer to section 8 of the call text);
- That, in case Physical Access to the National Facilities is requested, Applicant and/ or the team member(s) who will Access the National Facility have comprehensive insurance coverage for accidents and third-party liability, encompassing all their activities during their stay at HT and ensure that HT is recognized as third party. Name of the insurance company, insurance policy number and expiration date will be provided to HT before physical Access;
- That the information provided in the Application is true and accurate, being aware of the accountability implications set out in section 6.2 of the call text;
- That the above data are true, being aware of the liability and the penal sanctions stated in art. 75 and 76 of the DPR 445/2000 for false affirmations and mendacious declarations that will involve the loss of the granted benefits.

and is committed

To accept the terms and conditions to Access Human Technopole National Facilities as described in the National Facilities Access Rules (link) and, when applicable, the Access Agreement and its annexes (link).

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

Date ……

Name and Title …… ; ……

Email and Signature of **legal representative or delegated person (e.g., Head of Department)**

…… ; ……

ANNEX II

A. Project proposal formats and template

Mandatory proposal formats

***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* (e.g., ID123456_PROPOSAL_Rossi)

***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* (e.g., ID123456_PROPOSAL_Rossi)

Resubmission Cover page: *must be included in the pdf proposal, as its cover page (Standard applications max 1+3 pages, PoC applications max 1+2 pages) and it must have the following format: Max 1 page (Page format: A4, Font type: Arial, Font size: at least 11, Line spacing: single, Margins 2 cm side/ 1.5 bottom).*

Project Continuation Cover page: *must be included in the pdf proposal, as its cover page (Standard applications max 1+3 pages, PoC applications max 1+2 pages) and it must have the following format: Max 1 page (Page format: A4, Font type: Arial, Font size: at least 11, Line spacing: single, Margins 2 cm side/ 1.5 bottom).*

Proposal content:

- Resubmission cover page - applies to Resubmissions only (see section 9 for details)
- Project continuation cover page - applies to Project continuation applications only (see section 10 for details)
- Sections:
  1. TITLE
  2. SIGNIFICANCE
  3. INNOVATION
  4. APPROACH - including aims, preliminary data in support of the proposed experiments, experimental design and anticipated results
  5. ENVIRONMENT - including facilities and resources available to support the aspects of the project to be performed elsewhere (i.e., outside the NF)
  6. REFERENCES (Optional)

NOTE: The cover page (where required), together with Sections 1 through 6, constitutes the content of the proposal PDF to be uploaded as part of the application form.

B. Biosketch template

A **Biosketch template** can be downloaded from this link.

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors. Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME:

eRA COMMONS USER NAME (credential, e.g., agency login):

POSITION TITLE:

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY

A. Personal Statement

B. Positions, Scientific Appointments, and Honors

C. Contributions to Science

ANNEX III: SERVICE LIST

HUMAN TECHNOPOLE

NATIONAL FACILITY FOR GENOMICS

CALL FOR ACCESS

26-G-ROUND-3

Table of Contents

1. SERVICE LIST 27
SID: G-001 – Whole Genome Sequencing (WGS) 27
SID: G-002 – Whole Exome Sequencing (WES) 29
SID: G-003 – Amplicon sequencing for microbiome analysis (16S-ITS) 31
SID: G-004 – Methylation sequencing (Methyl-seq) 32
SID: G-005 - mRNA sequencing from standard and low input 34
SID: G-006 – totalRNA sequencing from standard and low input. 36
SID: G-007 - smallRNA sequencing 40
SID: G-008_G-008.1/G-012 - Single-cell 3'RNAsequencing or Single-cell gene Expression Flex (10X Genomics) 42
SID: G-009/012/014 – Single-cell Immune profiling-V(D)J (10X Genomics or BD Rhapsody) 47
SID: G-010/012/013 – Single-cell multiome ATAC + Gene expression (10X Genomics or BD Rhapsody) 51
SID: G-015/016_G-029/030 - Spatial Transcriptomics analysis from Fresh-Frozen, Fixed Frozen or FFPE tissues using Visium HD (10X Genomics) or StereoSeq (STOmics-MGI) 56
SID: G017/018 – GeoMx Digital Spatial Profiling from Fresh-Frozen, Fixed Frozen or FFPE tissues (Nanostring) 62
SID: G-019 - Nanopore gDNA sequencing (long reads or ultra long reads) 64
SID: G-020 - Nanopore small gDNA sequencing (long reads) 66
SID: G-021 - Nanopore Direct RNA Sequencing 68
SID: G-022 - Nanopore cDNA sequencing (bulk cDNA or single-cell cDNA from 10x Genomics protocol) (Human-Mouse) 70
SID: G-023 - Nanopore cell-free DNA sequencing (Human) 73
SID: G-024_G-031 - Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech) 75
SID: G-025 - Sequencing only with NextSeq 2000 (Illumina) 77
SID: G-027 – Suspension BLISS (sBLISS) 78
SID: G-028 – directCAGE 80
Appendix 1: Summary of technical requirements 82
Appendix 2: Description of the Data analysis services available in combination with the NF for Genomics services 88

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 focuses 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** specializes in multi-omics technologies. Its focus extends to providing cutting-edge services in single-cell and spatial multi-omics analysis, as well as longread sequencing. This Unit closely collaborates with the Tissue Processing Infrastructural Unit part of the NF for Light Imaging, specifically dedicated to implementing and providing 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 preprocessing 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 is a list of available services with a detailed description of the service and technical requirements.

Please note that one single service per call for Access can be selected. Access to further service can be requested by submitting a new application to a future call.

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

1. SERVICE LIST

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

Services description - Whole Genome Sequencing (WGS) is a comprehensive and high-throughput technique that enables to generate 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 subsequently sequenced using high-throughput next-generation sequencing (NGS) platforms. These technologies generate short DNA sequences, or reads, from DNA fragments.

Bioinformatic analysis: 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 for rare disease and cancer studies and SID: NF62.02.03 for Metagenomics/Microbiome Analysis.

Library preparation protocol - Libraries will be prepared by using the following protocols, depending on the quality of the DNA extractions:

For high-quality DNA extracts:

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.

For degraded or low concentration DNA extracts:

NEB Ultra II DNA Library Prep Kit for Illumina

The Ultra II DNA Library Prep Kit for Illumina with UMI ensures high yield preparation of high-quality libraries using a fast, streamlined, automatable workflow and enables the use of fewer PCR cycles while also improving GC coverage. Unique Dual Index (UMI) adaptors allow for accurate deduplication, identifying and removing PCR duplicates for improved variant calling and sensitivity, also from FFPE samples. The kit is also compatible with PCR-free workflows.

Libraries sequencing and NGS coverage: Libraries will be sequenced using the NovaSeq X Plus systems (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 sequenced reads, so base calls can be made with a higher degree of confidence.

For WGS applied to Cancer studies an average human genome coverage of 50X for germiline mutations and 100X for somatic mutations analyses will be obtained. For WGS applied to Medical Studies and Rare Diseases, an average human genome coverage of 20 X or 30X per sample will be obtained depending on the application.

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**	**Max samples if Data Analysis is requested**
G-001-A WGS for Population and medical studies (coverage 20X)	NOT AVAILABLE FOR THIS EVALUATION ROUND		
G-001-B WGS for Medical Studies and Rare Diseases (coverage 20X-30X)	20	200	200
G-001-C WGS for Cancer studies (coverage 50X and/or 100X)	20	100	100
G-001-D WGS Metagenomics studies (coverage determined considering type and complexity of the metagenome)	50	200	200
G-001-E WGS Plants studies (coverage determined considering the plant genome size and ploidy level)	5	20	20
G-001-F WGS for Animal studies	20	100	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:
    - For Illumina PCR Free: 15 ng/µl in 50 ul of nuclease free ultrapure water
    - For NEB Ultra II: 4 ng/µl in 60 ul of nuclease free ultrapure water
- gDNA samples should be quantified by using a fluorometer (i.e. Qubit/Glomax).
- For **Illumina PCR Free**: Quality of gDNA should be evaluated by Agilent Tape Station; samples should have a DIN (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 or NF62.02.03. Metagenomics/Microbiome Analysis

Access modality: Access to NF services.

Services available in combination with the NF for Data Handling and Analysis: NF62.02.01 – WGS analysis or NF62.02.03 - Metagenomics/Microbiome Analysis.

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 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.

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 subsequently sequenced using high-throughput next-generation sequencing (NGS) platforms. The sequencing generates short DNA sequences, or reads, from the exonic regions.

Bioinformatic analysis: 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 -

Libraries will be prepared by using the following protocol:

Enzymatic Fragmentation and Twist Universal Adapter System

Prepared libraries will be enriched for exonic regions by hybridization with the Twist Comprehensive Exome Panel: 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. The following protocol will be used for the target enrichment:

Twist Target Enrichment Fast Hybridization Protocol

Libraries sequencing and NGS coverage: Libraries will be sequenced using the NovaSeq X Plus 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**	**Max samples if Data Analysis is requested**
G-002-A Whole Exome Sequencing for Rare diseases studies	25	100	100
G-002-B Whole Exome Sequencing for Cancer studies	50	400	400

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, samples should have a DIN (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 and enrichment 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: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: NF62.02.02 - WES analysis.

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 the 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 subsequently sequenced using high-throughput next-generation sequencing (NGS) platforms. This step generates short DNA sequences (300bp) from the variable regions of the 16S rRNA and ITS genes.

Bioinformatic analysis: 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 - Metagenomics/Microbiome Analysis.

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 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 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	Max samples if Data Analysis is requested
G-003 Microbiome analysis 16S-ITS (commercially available Region panels)	100	1200	1200

- **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.
- 20 µl of High-quality DNA having a concentration ranging 0,1-1 ng/ul should be provided.
- Quality of DNA should be evaluated by Agilent Tape Station; samples should have a DIN (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 - Metagenomics/Microbiome Analysis.

Access modality: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis:NF62.02.03 - Metagenomics/Microbiome Analysis.

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 taken 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 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 subsequently sequenced using high-throughput next-generation sequencing (NGS) platforms. These technologies generate short DNA sequences called “reads” from DNA fragments.

Pre-processing data analysis: QC and summary statistics.

Bioinformatic analysis: Bioinformatic analysis can be provided as a combined service by the **National Facility for Data Handling and Analysis**. Please select: SID: NF62.02.04.

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 5hydromethylcytosine (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 X Plus systems (Illumina) by generating 100 bp Paired End reads. On average, 70 million read 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**	**Max samples if Data Analysis is requested**
G-004-A Methyl-seq for population studies on humans	100	500	Not Available
G-004-B Methyl-seq for studies on diseases (humans, organoids derived from human cells)	16	200	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, samples should have a DIN (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 and enrichment (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: NF62.02.04

Access modality: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: NF62.02.04 - Methyl-Seq 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, including 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:

RNA Extraction: 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: Following the capture of poly(A)-tailed mRNA molecules, 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, and the prepared library is amplified using polymerase chain reaction (PCR) to generate sufficient material for sequencing.

Sequencing: The prepared library is subsequently sequenced using high-throughput next-generation sequencing (NGS) platforms.

Bioinformatic analysis: 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® mRNA LP User Manual

This kit uses oligo(dT) priming to generate high-quality, full-length cDNA directly from multiple intact cells or total RNA. The SMART (Switching Mechanism at 5' End of RNA Template) technology employed by the SMART-Seq® mRNA LP 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 the NovaSeq X Plus systems (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**	**Max samples if Data Analysis is requested**
G-005-A mRNAseq in humans, animal, organoids studies	50	300	300
G-005-B mRNAseq in plants studies	50	200	200

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).

Technical requirements 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 200pg-100ng 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
- 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 NF62.01.01 - Bulk RNA-Seq analysis.

Access modality: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: NF62.01.01 – Bulk RNA-Seq analysis.

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 noncoding 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 brief 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).

cDNA synthesis: After ribosomal RNA depletion, 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, and the prepared library is amplified using polymerase chain reaction (PCR) to generate sufficient material for sequencing.

Sequencing: The prepared library is subsequently sequenced using high-throughput next-generation sequencing (NGS) platforms.

Bioinformatic analysis: 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 (RNA-seq analysis (coding transcripts) or SID: NF62.01.03 (lncRNA analysis).

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

Or

TruSeq Stranded Total RNA

Or

Watchmaker RNA Library Prep Kit with Polaris Depletion

Or

QIAseq Fast Select Epidemiology Kits

Illumina Total RNA Prep with Ribo-Zero Plus supports a broad range of RNA inputs. It's compatible with different sample types, including low-quality samples such as formalin-fixed paraffin-embedded (FFPE). The included Ribo-Zero Plus or Ribo-Zero Plus Microbiome Depletion kit 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.

The Illumina TruSeq Stranded Total RNA kit allows whole-transcriptome analysis of a broad range of RNA inputs. The optimization of fragmentation conditions allows to process degraded-RNA and formalin-fixed paraffin embedded (FFPE) samples. The kit can be combined with Ribo-Zero Gold that depletes samples of both cytoplasmic and mitochondrial rRNA or Ribo-Zero Plant that targets cytoplasmic and chloroplast rRNA.

Please check if your species of interest are included in the following table: Stranded Total RNA Species Compatibility Tables.

The Watchmaker RNA Library Prep Kit with Polaris Depletion enables rapid generation of stranded whole-transcriptome libraries by combining RNA library prep with rRNA and globin depletion in a single workflow.
It supports low-input and degraded samples (including FFPE), improving sensitivity and

transcript coverage across coding and non-coding RNA species. An engineered reverse transcriptase enhances cDNA synthesis efficiency, yielding high-complexity libraries suitable for diverse RNA-seq applications. The streamlined Polaris Depletion module depletes highly abundant rRNA in human, mouse, and rat samples, and human globin transcripts. For other species, dedicated ribodepletion probe pools can be applied (see https://www.sitoolsbiotech.com/products/ribopools/rrna-depletion#c5432)

QIAseq FastSelect Epidemiology Kits are developed to support dual-sequencing (dual-seq) and host–pathogen RNA-seq library preparation workflows. During library construction, these kits efficiently deplete bacterial 5S, 16S, and 23S ribosomal RNA together with cytoplasmic and mitochondrial ribosomal RNA from human, mouse, and rat samples.

The kits are compatible with input RNA quantities ranging from 100 ng to 1 µg while maintaining reliable performance across this range. They also perform effectively with a variety of RNA sample qualities, including low-quality, highly fragmented RNA as well as high-quality, intact full-length RNA.

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 RNAseq 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 the NovaSeq X Plus systems (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**	**Max samples if Data Analysis is requested**
G-006-A totRNAseq in humans, animal, organoids studies	50	300	300 (50 in case of lncRNA)
G-006-B totRNAseq in plants studies	20	200	200 (50 in case of lncRNA)
G-006-C totRNAseq in Bacteria (host/pathogen interaction, in vitro and ex-vivo)	20	100	100 (50 in case of lncRNA)

Technical requirements totalRNA sequencing from standard input:

For Illumina protocols:

- 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).

For Watchmaker protocols:

- 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 . For FFPE samples, DV200≥50%
- 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).

For QIAseq Fast Select Epidemiology Kits:

- 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 100ng-1000ng in a volume range of 20-30ul in nuclease-free ultrapure water.
- totalRNA should be DNase treated.
- 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).

Technical requirements 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 200pg-30ng in a total volume of maximum 15-20ul 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
- Mapping metrics (if reference genome/transcriptome is available)

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 NF62.01.01 - Bulk RNA-Seq analysis or NF62.01.03 - lncRNA analysis.

Access modality: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: NF62.01.01 - Bulk RNA-Seq analysis or NF62.01.03 - lncRNA analysis.

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 prepared library is subsequently sequenced using high-throughput next-generation sequencing (NGS) platforms.

Bioinformatic analysis: 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 or the NovaSeq X Plus systems (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**	**Max samples if Data Analysis is requested**
G-007 smallRNAseq in humans, animal, organoids samples	20	60	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 50ng-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)
- 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 (if reference genome/transcriptome is available)

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 NF62.01.02 – miRNA analysis.

Access modality: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: NF62.01.02 – miRNA analysis.

SID: G-008_G-008.1/G-012 - Single-cell 3'RNAsequencing or Single-cell gene Expression Flex (10X Genomics)

IMPORTANT – this service will not be offered in this call round.

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 with the Chromium X instrument. Each droplet contains a bead with a unique barcode. Typically, 10’000 cells per sample will be targeted.

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.

Bioinformatic analysis: 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:

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

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

GEMX Flex Gene Expression Reagent Kits for multiplex samples - CG000787

Libraries sequencing and NGS coverage: libraries will be sequenced using the NovaSeq X Plus systems (Illumina), generating read lengths as required by the Single-Cell 3' RNA Sequencing or Single-Cell Gene Expression Flex protocols (10x Genomics). Depending on the project's needs and sample size, sequencing of cDNAs generated with the 10x Genomics protocol using PromethION Nanopore flow cells may also be offered by the facility. 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**	**Max samples if Data Analysis is requested**
G-008/G-012- A Single-cell 3'RNA sequencing for studies on diseases (humans, animals, organoids, plants*)	NOT AVAILABLE FOR THIS EVALUATION ROUND		

G-008/012- B Single-cell 3'RNA sequencing for differentiation studies (humans, animals, organoids, plants*)	NOT AVAILABLE FOR THIS EVALUATION ROUND
G-008.1/G-012- C Single-cell gene Expression Flex (humans, animal models, organoids) (16-plex)	NOT AVAILABLE FOR THIS EVALUATION ROUND

* For plant samples, the feasibility of performing 3′ scRNA-seq on cells or nuclei will be evaluated according to the cell type and dimensions.

Technical requirements - Single-cell 3'RNAseq:

Accepted Sample Types

1. **Frozen single-cell suspensions**
   - Minimum input: 500,000 cells
   - Pre-freezing viability: ≥ 90%
   - Cell concentration:
     - 1–5 × 10^6 cells/mL (standard)
     - Up to 10 × 10^6 cells/mL (primary cells only)

2. **Organoids**

For organoid samples, only the single-nucleus workflow (snRNA-seq) will be performed.

- Freeze the whole organoids without prior dissociation
- Ensure that organoids are free of residual matrigel before cryopreservation
- After cryopreservation, if possible, perform a test dissociation to determine the expected cell yield and assess cell viability. Cell viability must be >90%.
- If the estimated total cell yield is <1 × 10^6 cells, pool multiple organoids to achieve the required input whenever possible.
- If pooling is not feasible, or if cell viability is <90%, contact the facility
- If a protocol for nuclei isolation from the organoid is available, perform a pilot nuclei isolation and provide a representative microscope image of the isolated nuclei to assess their quality and estimate the nuclei yield.
- For downstream applications, the isolation procedure should yield at least 50,000 nuclei at a final concentration of 5,000 nuclei/µL.

3. Snap-frozen tissue

For tissue samples, only the single-nucleus workflow (snRNA-seq) will be performed.

- minimum 50 mg – maximum 200 m
- Please, if possible, provide the RNA Integrity Number (RIN)RIN of the RNA extracted form the tissue

General Requirements

- A backup sample is required.
- Samples must be provided in sterile cryogenic storage vials (cryovials).
- Samples must be prepared according to:
  - Cell Preparation for Single Cell Protocols – CG00053
  - Isolation of Leukocytes, Bone Marrow and Peripheral Blood Mononuclear Cells for Single Cell RNA Sequencing- CG000392
  - After reviewing the protocol, if you have any remaining questions, please contact the Facility by email with your specific queries.
  - Pre-freezing viability must be ≥ 90%
  - If cells are sorted, viability and count must be measured after sorting and before freezing
  - Post-thaw viability testing is mandatory
  - Post-thaw viability results must be shared with the Facility before submissio
  - Cell viability must be documented with representative microscope images of the Bürker counting chamber acquired during the viability assessment and shared with the Facility before sample submission.
- **Only samples that meet both conditions will be processed:**
  - Post-thaw viability ≥ 75%
  - Low cellular debris

Sample Handling, Storage and Shipment

Proper cryopreservation, storage and shipment are critical to preserve sample integrity and ensure optimal sequencing results.

Cryopreservation

Cells

- Pre-freezing viability must be ≥ 90%.
- Recommended cell concentration:
  - 1–5 × 10^6 cells/mL (standard)
  - Up to 10 × 10^6 cells/mL (primary cells only)
- Freeze using a controlled cooling rate (~1 °C/min) with an appropriate cryoprotectant (e.g. 10% DMSO or commercial cryopreservation medium).
- Store samples overnight at −80 °C, then transfer to liquid nitrogen or a −150 °C freezer for long-term storage.
- Long-term storage at −80 °C is not recommended.

Tissue

- Tissue must be snap-frozen immediately after collection using liquid nitrogen (or LN_2-cooled isopentane).

- If immediate cryopreservation is not feasible, transfer the tissue to a validated tissue preservation buffer compatible with downstream single-cell/multiomic applications. Follow the manufacturer's recommendations for storage conditions and maximum preservation time before cryopreservation.
- Store in liquid nitrogen for long-term preservation; −80°C is suitable only for short-term storage.

Storage

- Long-term storage: < −130 °C (liquid nitrogen or equivalent).
- Minimize sample handling and avoid freeze–thaw cycles.
- Record and communicate:
    a) sample collection date;
    b) storage conditions and duration;
    c) any freeze–thaw events.

Shipment

- Ship samples frozen at all times using:
    - a nitrogen vapor shipper, or
    - dry ice (recommended for cryopreserved cells).
- Ensure sufficient dry ice to maintain samples frozen throughout transport.
- Ship samples in sterile cryovials only.

Technical requirements - RNA Flex:

Accepted Sample Types

1) Fixed single-cell/Nuclei suspensions:

- Minimum input: 100,000 fixed cells/nuclei with pre-fixation viability ≥90% and minimal debris. If the pre-fixation viability is below this threshold, or if fewer than 100,000 cells/nuclei are available, please contact the Facility before proceeding.
- Cell viability before fixation must be documented with representative microscope images of the Bürker counting chamber acquired during the viability assessment and shared with the Facility before sample submission.

Samples must be prepared following:

- Cell Preparation for Single Cell Protocols – CG00053
- Fixation of Cells & Nuclei for GEM-X Flex Gene Expression – CG000782
- Sample Preparation from FFPE Tissue Sections for GEM-X Flex Gene Expression - CG000784
- Blood Fixation and Cell Isolation for GEM-X Flex Gene Expression- CG000785

After reviewing the protocol, if you have any remaining questions, please contact the Facility by email with your specific queries.

2) Flash-frozen and fresh tissues:

Projects involving tissue types that are not listed among the validated tissues in the link below will be treated as proof-of-concept studies.

  - Single Cell Flex Tested Tissues

- minimum 50 mg – maximum 200 mg (assay will be performed on nuclei directly isolated from tissue and fixed by the facility)

Please, if possible, provide the RNA Integrity Number (RIN) of the RNA extracted from the tissue.

Samples must be prepared following:

- Nuclei Isolation from Cell Suspensions & Tissues for Single Cell RNA Sequencing - CG000124
- Tissue Fixation & Dissociation for GEM-X Flex Gene Expression - CG000783

After reviewing the protocol, if you have any remaining questions, please contact the Facility by email with your specific queries.

General Requirements

- A backup sample is required.
- Whenever possible, samples should be provided in 1.5 mL low-binding tubes or in sterile cryovials in case storage in liquid nitrogen is needed.
- Contact the facility to receive appropriate preparation workflow for dissociation and fixation of other sample types (organoids, FFPE blocks).

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 NF62.03.01 - scRNA-Seq analysis.

Access modality: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: NF62.03.01 - scRNA-Seq analysis.

Please be aware that the NF for Data Handling and Analysis can analyse **up to 32 samples per project.**

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

IMPORTANT – this service will not be offered in this call round.

Services description - Single-cell immune profiling 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 compartments (droplets or microwells) using microfluidics technology with the Chromium X or BD Rhapsody instrument. Each compartment contains a bead with a unique barcode. Typically, 10'000 cells per sample will be targeted.

Barcoding and cDNA Synthesis: Within each compartment, the cell's RNA is captured end 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: 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 projects' needs, the samples will be processed using the Chromium X or BD Rhapsody platforms, aiming to recover 10'000 cells per sample.

Libraries will be prepared by following the protocols:

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

Bd biosciences protocols, single-cell-multiomics

Libraries sequencing and NGS coverage: libraries will be sequenced NovaSeq X Plus systems (Illumina), generating read lengths as required by the Single-cell Immune profiling V(D)J protocol (10x Genomics or BD Rhapsody).

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**	**Max samples if Data Analysis is requested**
G-009/G-012/G-014 Single-cell Immune profiling-VDJ for studies on diseases (humans, animal models)	NOT AVAILABLE FOR THIS EVALUATION ROUND		

Technical requirements:

Accepted Sample Types

1. **Frozen single-cell suspensions**
   - Minimum input: 500,000 cells
   - Pre-freezing viability: ≥ 90%
   - Cell concentration:
     - 1–5 × 10^6 cells/mL (standard)
     - Up to 10 × 10^6 cells/mL (primary cells only)
2. **Organoids**

For organoid samples, only the single-nucleus workflow (snRNA-seq) is available. Consequently, TCR and BCR V(D)J library preparation cannot be performed.

- Freeze the whole organoids without prior dissociation
- Ensure that organoids are free of residual matrigel before cryopreservation
- After cryopreservation, if possible, perform a test dissociation to determine the expected cell yield and assess cell viability. Cell viability must be >90%.
- If the estimated total cell yield is <1x10^6 cells, pool multiple organoids to achieve the required input whenever possible.
- If pooling is not feasible, or if cell viability is <90%, contact the facility
- If a protocol for nuclei isolation from the organoid is available, perform a pilot nuclei isolation and provide a representative microscope image of the isolated nuclei to assess their quality and estimate the nuclei yield.
- For downstream applications, the isolation procedure should yield at least 50,000 nuclei at a final concentration of 5,000 nuclei/µL.

3. **Snap-frozen tissue**

For tissue samples, only the single-nucleus workflow (snRNA-seq) is available. Consequently, TCR and BCR V(D)J library preparation cannot be performed.

- minimum 50 mg – maximum 200 mg
- Please, if possible, provide the RNA Integrity Number (RIN)RIN of the RNA extracted from the tissue

General Requirements

- A backup sample is required.
- Samples must be provided in sterile cryogenic storage vials (cryovials).
- Samples must be prepared according to:
- Cell Preparation for Single Cell Protocols - CG00053
- Performance of Blood Collection Tubes and PBMC Isolation Methods for Single Cell Gene Expression - CG000411
- After reviewing the protocol, if you have any remaining questions, please contact the Facility by email with your specific queries.

- Pre-freezing viability must be ≥ 90%
- If cells are sorted, viability and count must be measured after sorting and before freezing
- Post-thaw viability testing is mandatory
- Post-thaw viability results must be shared with the Facility before submission
- Cell viability must be documented with representative microscope images of the Bürker counting chamber acquired during the viability assessment and shared with the Facility before sample submission.
- Only samples that meet both conditions will be processed:
    - Post-thaw viability ≥ 75%
    - Low cellular debris

Sample Handling, Storage and Shipment

Proper cryopreservation, storage and shipment are critical to preserve sample integrity and ensure optimal sequencing results.

Cryopreservation

Cells

- Pre-freezing viability must be ≥ 90%
- Recommended cell concentration:
    - 1–5 × 10^6 cells/mL (standard)
    - Up to 10 × 10^6 cells/mL (primary cells only)
- Freeze using a controlled cooling rate (~1 °C/min) with an appropriate cryoprotectant (e.g. 10% DMSO or commercial cryopreservation medium).
    - Store samples overnight at −80 °C, then transfer to liquid nitrogen or a −150 °C freezer for long-term storage.
    - Long-term storage at −80 °C is not recommended.

Tissues

- Tissue must be snap-frozen immediately after collection using liquid nitrogen (or LN_2-cooled isopentane).
- If immediate cryopreservation is not feasible, transfer the tissue to a validated tissue preservation buffer compatible with downstream single-cell/multiomic applications. Follow the manufacturer's recommendations for storage conditions and maximum preservation time before cryopreservation.
- Store in liquid nitrogen for long-term preservation; −80 °C is suitable only for short-term storage.

Storage

- Long-term storage: < −130 °C (liquid nitrogen or equivalent).

- Minimize sample handling and avoid freeze–thaw cycles.
- Record and communicate:
    a) sample collection date;
    b) storage conditions and duration;
    c) any freeze–thaw events.

Shipment

- Ship samples frozen at all times using:
    - a nitrogen vapor shipper, or
    - dry ice (recommended for cryopreserved cells).
- Ensure sufficient dry ice to maintain samples frozen throughout transport.
- Ship samples in sterile cryovials only.

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 NF62.03.03 - Single-cell immune profiling (VDJ).

Access modality: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: NF62.03.03 - Single-cell immune profiling (VDJ).

Please be aware that the NF for Data Handling and Analysis can analyse **up to 32 samples per project.**

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

IMPORTANT – this service will not be offered in this call round.

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. For some applications, these technologies can be further adapted to include other omics layers. Non-standard applications that are not officially supported by commercial kits, will be carefully evaluated by the Facility Staff upon request, as part of a "proof-of-concept" project, to determine their technical feasibility.

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 or BD Rhapsody platform typically involves several key steps:

Cell Isolation and Nuclei Preparation: Nuclei might be isolated from suspension of single cell, organoids or directly from frozen tissues as the chromatin accessibility information is primarly 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.

Nuclei compartmentalization: The isolated nuclei are compartmentalized into Gel Bead-in-Emulsion (GEM) droplets using the 10X Genomics Chromium platforms or in microwells using BD Rhapsody. In each compartment, beads contain 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 compartment, 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 the Novaseq X Plus platform. The sequencing reads contain information about both chromatin accessibility and gene expression in individual cells/nuclei.

QC and summary statistics

Bioinformatic analysis: 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:

GEM-X Epi Multiome User Guide - CG001689

BD biosciences protocols, single-cell-multiomics

Libraries sequencing and NGS coverage:

Depending on the projects' needs, the samples will be processed using the Chromium X or BD Rhapsody platforms, which allow to recover and analyze respectively a maximum of 10,000 nuclei per single sample without multiplexing.

Libraries will be sequenced using NovaSeq X Plus systems (Illumina), generating read lengths as required by the single-cell multiome ATAC + Gene expression protocol (10x Genomics or BD Rhapsody).

On average, 50,000 reads per cell/nucleus will be generated for Gene Expression (GEX) libraries and 25,000 reads per cell/nucleus for ATAC-seq libraries generated using the 10x Genomics Single Cell Multiome ATAC + Gene Expression protocol. For libraries generated using the BD Rhapsody™ Single-Cell Multiomics (WTA + ATAC) protocol, an average of 50,000 reads per cell/nucleus will be generated for both Whole Transcriptome Analysis (WTA) and ATAC-seq libraries.

Within this call the National Facility for Genomics will provide:

	Min samples/project	**Max samples/project**	**Max samples if Data Analysis is requested**
G-010/G-012/G-013- A Single-cell multiome ATAC + Gene expression for studies on diseases (humans, animals, organoids)	NOT AVAILABLE FOR THIS EVALUATION ROUND		

Technical requirements:

Accepted Sample Types

1) **Snap-frozen tissue**
   - Minimum 50mg – maximum 200 mg.
   - Please, if possible, provide the RNA Integrity Number (RIN) of the RNA extracted from the tissue

2) Frozen single-cell suspensions:
   - Minimum input: **1,000,000 cells**
   - **pre-freezing viability ≥ 90%**
     - Cell viability must be documented with representative microscope images of the Bürker counting chamber acquired during the viability assessment and shared with the Facility before sample submission.
     - A **backup sample must be provided**
     - Samples must be provided in **sterile cryogenic storage vials (cryovials)**
   - Samples must be prepared according to:
     - Cell Preparation for Single Cell Protocols - CG00053
     - Isolation of Leukocytes, Bone Marrow and Peripheral Blood Mononuclear Cells for Single Cell RNA Sequencing- CG000392
   - After reviewing the protocol, if you have any remaining questions, please contact the Facility by email with your specific queries

- If cells are sorted, viability and count **must be measured after sorting and before freezing**
- Post-thaw viability testing is mandatory
- Post-thaw viability results must be shared with the Facility before submission
- **Only samples that meet both conditions will be processed:**
  - Post-thaw viability ≥ 75%
  - Low cellular debris

3) Organoids:

- Freeze the whole organoids without prior dissociation
- Ensure that organoids are free of residual matrigel before cryopreservation
- After cryopreservation, if possible, perform a test dissociation to determine the expected cell yield and assess cell viability. Cell viability **must be >90%**.
- If the estimated total cell yield is **<1 × 10^6 cells**, pool multiple organoids to achieve the required input whenever possible.
- If pooling is not feasible, or if cell viability is **<90%**, contact the facility
- If a protocol for nuclei isolation from the organoid is available, perform a pilot nuclei isolation and provide a representative microscope image of the isolated nuclei to assess their quality and estimate the nuclei yield.
- For downstream applications, the isolation procedure should yield **at least 50,000 nuclei** at a final concentration of **5,000 nuclei/µL**.
- Minimum input: 1,000,000 cells per sample
- Contact the Facility before sample preparation to discuss storage and shipment procedures

Sample Handling, Storage and Shipment

Proper cryopreservation, storage and shipment are critical to preserve sample integrity and ensure optimal sequencing results.

Cryopreservation

Cells

- Pre-freezing viability must be ≥ 90%.
- Recommended cell concentration:
  - 1–5 × 10^6 cells/mL (standard)
  - Up to 10 × 10^6 cells/mL (primary cells only)
- Freeze using a controlled cooling rate (~1 °C/min) with an appropriate cryoprotectant (e.g. 10% DMSO or commercial cryopreservation medium).
  - Store samples overnight at −80 °C, then transfer to liquid nitrogen or a −150 °C freezer for long-term storage.
  - Long-term storage at −80 °C is not recommended.

Tissues

- Tissue must be snap-frozen immediately after collection using liquid nitrogen (or LN_2-cooled isopentane).
- If immediate cryopreservation is not feasible, transfer the tissue to a validated tissue preservation buffer compatible with downstream single-cell/multiomic applications. Follow the manufacturer's recommendations for storage conditions and maximum preservation time before cryopreservation.
- Store in liquid nitrogen for long-term preservation; −80 °C is suitable only for short-term storage.

Storage

- Long-term storage: < −130 °C (liquid nitrogen or equivalent).
- Minimize sample handling and avoid freeze–thaw cycles.
- Record and communicate:
    a) sample collection date;
    b) storage conditions and duration;
    c) any freeze–thaw events.

Shipment

- Ship samples frozen at all times using:
    - a nitrogen vapor shipper, or
    - dry ice (recommended for cryopreserved cells).
- Ensure sufficient dry ice to maintain samples frozen throughout transport.
- Ship samples in sterile cryovials only.

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

Access modality: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: NF62.03.04. - Single-cell multiome (ATAC + gene expression) analysis.

SID: G-015/016_G-029/030 - Spatial Transcriptomics analysis from Fresh-Frozen, Fixed Frozen or FFPE tissues using Visium HD (10X Genomics) or StereoSeq (STOmics-MGI)

IMPORTANT – this service will not be offered in this call round.

Services description - our spatial transcriptomics service portfolio provides an end-to-end solution for mapping gene expression within intact tissues using high-resolution, sequencing-based technologies. By combining morphological context with unbiased transcriptome profiling, we enable researchers to investigate tissue organization, cellular interactions, and spatially defined molecular programs with cellular and subcellular resolution.

The service supports multiple platforms, including 10x Genomics Visium HD and BGI/MGI STOmics Stereo-seq, allowing us to tailor experimental design and resolution to the biological question and tissue type.

10X Genomics spatial transcriptomics:

Visium HD Spatial Gene Expression (10X Genomics) allows 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.

Visium HD WT Panel Gene Expression:

Tissue Slide Preparation and Staining: The National Facility for Genomics supports Visium HD experiments exclusively with H&E-stained fresh-frozen tissue sections.

This workflow enables spatially resolved whole transcriptome profiling from fresh-frozen, fixed frozen, or FFPE human and mouse tissue sections. Following tissue preparation and H&E, probes are hybridized to RNA targets for sensitive detection of gene expression.

The Visium CytAssist system facilitates the spatial transfer of transcriptomic information onto capture areas. Libraries are then prepared and sequenced using high-throughput platforms.

Services available in combination with the National Facility for Data Handling and Analysis: Bioinformatic analysis of Visium Spatial data can be provided as a combined service by the National Facility for Data Handling and Analysis: NF62.03.05 - Spatial transcriptomics (10X Visium /StereoSeq platforms) analysis.

Please be aware that the combined data analysis service can be provided only when **a single sample** per area is processed.

Visium HD 3' Gene Expression:

This workflow enables spatially resolved transcriptomic profiling from fresh-frozen tissue sections. After tissue sectioning and mounting onto Visium slides, the RNA is captured in a spatial context through fixation, permeabilization, and reverse transcription. The resulting cDNA is then processed into sequencing libraries, which are sequenced using high-throughput platforms. Optional bioinformatic analysis can be provided upon request.

While these protocols have been standardized for human and mouse tissues by 10x Genomics, adaptations to other tissue types (e.g., plant tissues) and the addition of customized probes to capture specific transcripts (e.g., pathogen transcripts) will be carefully evaluated by the Facility Staff upon request, as part of "proof-of-concept" projects, to determine the technical feasibility.

Services available in combination with the National Facility for Data Handling and Analysis: Bioinformatic analysis of Visium Spatial data can be provided as a combined service by the National Facility for Data Handling and Analysis: NF62.03.05 - Spatial transcriptomics (10X Visium /StereoSeq platforms) analysis.

Please be aware that the combined data analysis service can be provided only when **a single sample** per area is processed.

STOmics-MGI spatial transcriptomics:

Stereo-seq (STOmics-MGI) (SpaTial Enhanced REsolution Omics-sequencing) is a sequencing-based spatial transcriptomics technology that enables the exploration of spatial biology with subcellular resolution (500 nm) and a centimeter-scale field of view (FOV) across different tissue types. Stereo-seq Transcriptomics is built on DNA Nanoball (DNB) technology, using DNB-patterned array chips loaded with spatially barcoded probes to capture and prime RNA directly from tissue sections *in situ*. The resulting spatially barcoded cDNA libraries are sequenced on MGI DNBSEQ platforms (DNBSEQ-T7), after which sequencing data are mapped back to their spatial coordinates to enable high-resolution gene expression mapping and visualization of the tissue architecture.

Stereo-seq V1.3:

Stereo-seq V1.3 is specifically designed for mRNA whole-transcriptome profiling, it is based on ultra-dense DNA nanoball (DNB) arrays carrying unique spatial barcodes that selectively capture polyadenylated mRNA molecules released from tissue sections after permeabilization, it generates high-resolution mRNA gene expression maps from subcellular to tissue-wide scale. Stereo-seq V1.3 is compatible only with fresh frozen tissues.

Services available in combination with the National Facility for Data Handling and Analysis: Bioinformatic analysis of Stereo-seq V1.3 data can be provided as a combined service by the National Facility for Data Handling and Analysis: NF62.03.05 - Spatial transcriptomics (10X Visium /StereoSeq platforms) analysis.

Please be aware that the combined data analysis service can be provided only when **a single sample** per area is processed.

Stereo-seq OMNI:

Stereo-seq OMNI is a spatial transcriptomics solution specifically optimized for FFPE tissues, enabling whole-transcriptome spatial profiling of diverse tissue types and archival samples. The technology uses ultra-dense DNA nanoball (DNB) arrays combined with a random-primed capture strategy, allowing the efficient capture of both coding and non-coding RNA species and fragmented RNA typically present in FFPE material. Stereo-seq OMNI enables the spatial investigation of host–pathogen interactions, allowing the simultaneous detection of host and microbial transcripts within their native tissue context.

Services available in combination with the National Facility for Data Handling and Analysis: Bioinformatic analysis of Stereo-seq OMNI data can be provided as a combined service by the National Facility for Data Handling and Analysis: NF62.03.05 - Spatial transcriptomics (10X Visium /StereoSeq platforms) analysis.

Please be aware that the combined data analysis service can be provided only when **a single sample** per area is processed.

Libraries Sequencing and NGS coverage:

Libraries will be sequenced using the NovaSeq X Plus systems (Illumina) for the 10x Genomics Visium HD and using DNBSEQ-T7 system (MGI Tech) for Stereo-seq generating read lengths and coverage accordingly with the protocols specification.

Within this call the National Facility for Genomics will provide:

	Min area*/project	Max area*/project	Max samples if Data Analysis is requested
G-015/016_G-029/030 -A - Spatial Transcriptomics analysis from Fresh-Frozen, Fixed Frozen or FFPE tissues using Visium HD (10X Genomics) or StereoSeq (STOmics-MGI) for disease studies (humans, animal models, organoids)	NOT AVAILABLE FOR THIS EVALUATION ROUND		
G-015/016_G-029/030 -B - Spatial Transcriptomics analysis from Fresh-Frozen, Fixed Frozen or FFPE tissues using Visium HD (10X Genomics) or StereoSeq (STOmics-MGI) for differentiation studies (humans, animal models, organoids)	NOT AVAILABLE FOR THIS EVALUATION ROUND		

* The functional area in Spatial Gene Expression assay refers to the capture area on the slide where spatially barcoded spots are located. This region is where tissue sections are placed, permeabilized, and RNA is captured for spatial transcriptomic analysis.

IMPORTANT:

The size of the Visium HD area is 6.5x6.5 mm.

The size of the Stereo-seq area is 10x10 mm.

Please note that only one tissue section may be placed within each functional area. In case of organoids, please contact the Facility in advance to discuss the experiumental design.**IMPORTANT***:*

For Visium HD only the H&E staining of the tissue is supported by the facility

For Stereo-seq only the fluorescent staining of the tissue using Qubit ssDNA Reagent is supported by the facility.

Technical requirements:

For Fresh Frozen and Fixed Frozen Visium HD and Fresh Frozen Visium HD 3':

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

Visium HD Fresh Frozen Tissue Preparation Handbook 2.0 - CG001677

Visium HD Fixed Frozen Tissue Preparation Handbook 2.0 - CG001678

Visium HD 3' Fresh Frozen Tissue Preparation Handbook - CG000804

Please be aware that NF Genomics will only accept samples prepared by the user following the mandatory preparation guidelines outlined below:

- **Read the guidance** on tissue freezing, embedding, and sectioning.
- **RNA Quality Assessment:** Before placing sections on slides, assess RNA quality of the tissue block by calculating the RNA Integrity Number (RIN) from RNA extracted from tissue sections.

Only samples with **RIN ≥ 4** will be processed for **Fresh Frozen Visium HD**.

Only samples with **DV200 ≥ 50%** will be processed for **Fixed Frozen Visium HD**.

Only samples with **RIN ≥ 7** will be processed for **Fresh Frozen Visium HD 3'**.

- Place the tissue section on **compatible blank slides** listed in the Visium HD Spatial Gene Expression Protocol Planner (CG000698) or the Visium HD 3' Spatial Gene Expression Protocol Planner (CG000803).
- 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. Refer to the Visium CytAssist Accessory Kit Quick Reference Card - CG000548
- The **recommended section thickness** for most tissue types is **10 µm**; however, sections between **10–20 µm** are compatible with the assay. Tissues with high fat content (e.g., breast tissue) may require sections closer to **20 µm**.
- The user must perform **H&E staining on a separate section** exclusively for tissue morphology assessment, as described in H&E Staining, mount the coverslip as described in Coverslip Mounting, and acquire images according to Imaging in the Tissue Preparation Handbook. Please share the **H&E image** with NF Genomics.
- If extra tissue slides are available, for **Visium HD**, optional **tissue morphology assessment via DAPI** on a separate section is recommended, as described in the Tissue Preparation Handbook.
- **Annotate the AOI (Area of Interest)** using a green marker on the back of the H&E Tissue Morphology Assessment slide. Please share the **AOI image** with NF Genomics.
- Read the **guidance on slide shipping** provided in the Tissue Preparation Handbook. Please note that **shipping slides may result in RNA degradation**.

For FFPE Visium HD

Tissues should be prepared according to the protocols:

Visium HD FFPE Tissue Preparation Handbook 2.0 - CG001676

Please be aware that NF Genomics will only accept samples prepared by the user following the mandatory preparation guidelines outlined below:

- **Read the guidance** on tissue handling, fixation, embedding, and sectioning. If the tissue has already been fixed and embedded, please provide NF Genomics with all relevant information regarding these steps.
- **RNA Quality Assessment:** before placing sections on slides, assess RNA quality of the tissue block by calculating the **DV200** from RNA extracted from tissue sections.

Only samples with **DV200 ≥ 30%** will be processed for **Visium HD FFPE**.

- Place the tissue section on **compatible blank slides** listed in the Visium HD Spatial Gene Expression Protocol Planner (CG000698)

- 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. Refer to the Visium CytAssist Accessory Kit Quick Reference Card - CG000548
- The **recommended section thickness** for most tissue types is **5 µm**. For tissue blocks with exposed tissue, discard the first few sections and begin collecting from subsequent sections.
- The user must perform **H&E staining on a separate section** exclusively for tissue morphology assessment, as described in H&E Staining. Mount the coverslip as described in Coverslip Mounting and acquire images as described in Imaging in the Tissue Preparation Handbook.

Please share the H&E stained tissue image with NF Genomics.

- If additional tissue slides are available, optional tissue morphology assessment via DAPI on a separate section is recommended for Visium HD, as described in the Tissue Preparation Handbook.
- Annotate the AOI (Area of Interest) using a green marker on the back of the H&E morphology slide. Please share the AOI image with NF Genomics.
- Read the guidance on slide shipping provided in the Tissue Preparation Handbook

For Fresh Frozen STOmics Stereo-seq, all protocols can be downloaded at the following link:

https://en.stomics.tech/resources/documents

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

STUM-SP001 – Sample Preparation Guide for Fresh Frozen Samples on Stereo-seq Chip Slides

NF Genomics will provide all chip slides and reagents necessary for tissue preparation.

Please note that **NF Genomics will only accept samples prepared by the user in full accordance with the mandatory preparation guidelines outlined below**:

- Carefully read the guidance on tissue embedding.
- **RNA Quality Assessment:** Before placing tissue sections on slides, assess RNA quality by calculating the RNA Integrity Number (RIN) from RNA extracted from tissue sections. Only samples with **RIN ≥ 4** will be processed for Fresh Frozen STOmics Stereo-seq.
- Place only the **region(s) of interest** of the tissue on the **Stereo-seq Chip P slides**. These slides will be sent to the facility to test permeabilization conditions for the specific tissue of interest following protocol STUM-PR002– Stereo-seq Permeabilization Set V1.1 for Chip-on-a-slide User Manual.
- Place only the **region(s) of interest** of the tissue on the **Stereo-seq Chip T slides** following protocol STUM-TT001 – Stereo-seq Transcriptomics Set V1.3 for Chip-on-a-slide User Manual.
- Stereo-seq Chip P slides and Stereo-seq Chip T slides are distinguished by a **laser-engraved label** at the end of each slide.

- The user must perform **H&E staining for tissue morphology assessment** and share the H&E image of the stained tissue section with NF Genomics.
- Carefully read the guidance on how to **store the Stereo-seq slides at the first STOPPING POINT** described in the user manual.

For FFPE tissue STOmics Stereo-seq OMNI.

All the protocols can be download at this link:

https://en.stomics.tech/resources/documents

Tissues should be prepared according to the protocol:

STUM-SP003-Sample-Preparation,-Sectioning,-and-Mounting-Guide-for-FFPE-Samples-on-Stereo-seq-Chip-Slides

Please note that **NF Genomics will only accept samples prepared by the user in full accordance with the mandatory preparation guidelines outlined below**:

- Carefully read the guidance on **tissue sectioning**.
- **RNA Quality Assessment:** Before placing sections on slides, assess RNA quality from the tissue block by calculating the **DV200** of RNA extracted from tissue sections. Only samples with **DV200 ≥ 30%** will be processed for **Stereo-seq OMNI**.
- Read the guidance on **preservation and transportation of paraffin sections**. Please collect and send to NF Genomics only the **region of interest**, ensuring that the tissue section size is **≤ 0.9 cm × 0.9 cm**.
- The user must perform **H&E staining for tissue morphology assessment** and share the H&E-stained tissue image with NF Genomics.
- If extra tissue slides are available, **optional DAPI morphology assessment** is recommended. Please share the DAPI-stained tissue image with NF Genomics.
- The **recommended section thickness** for most tissue types is **5 µm** (5 µm for regular tissue and **4 µm for high-fat-content tissue** to reduce the risk of section detachment during subsequent operations). For tissue blocks with exposed tissue, discard the first few sections and begin collecting from subsequent sections.

Provide NF Genomics with all relevant information regarding **tissue handling, fixation, and embedding**.

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: NF62.03.05.

Access modality: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: Bioinformatic analysis of Spatial transcriptomics (10X Visium/StereoSeq platforms)

data can be provided as a combined service by the National Facility for Data Handling and Analysis: NF62.03.05 - Spatial transcriptomics (10X Visium /StereoSeq platforms) analysis.

Please be aware that the combined data analysis service can be provided only when **a single sample** per area is processed.

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

IMPORTANT – this service will not be offered in this call round.

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).

Libraries Sequencing and NGS coverage:

Libraries will be sequenced using either the NextSeq 2000, NovaSeq 6000, or NovaSeq X Plus systems (Illumina), generating read depth and length as specified by the GeoMX protocol.

Within this call the National Facility for Genomics will provide:

	Min tissue slides /project	**Max tissue slides /project**	**Max samples if Data Analysis is requested**
G017/018 -A GeoMx Digital Spatial Profiling from Fresh-Frozen, Fixed Frozen or FFPE tissues for studies on diseases (humans, animal models, organoids)	NOT AVAILABLE FOR THIS EVALUATION ROUND		NA

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

Access modality: 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 mechanical, while for ultra-long reads, larger fragments are preserved through a transposase-based fragmentation.
- Repair Ends: Repair the DNA ends to ensure they have a consistent structure suitable for ligation.
- 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 optimized 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 analyzed 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 each sample will be sequenced in one PromethION flow cell.

For nanopore gDNA sequencing applied to Animal studies the number of flowcell needed to sequence each sample will be determined considering the animal genome size.

For nanopore gDNA sequencing applied to Plant studies the number of flowcell needed to sequence each sample will be determined considering the plant genome size and ploidy level.

Within this call the National Facility for Genomics will provide:

	Min flowcells/project**	**Max flowcells**/project**	**Max samples if Data Analysis is requested**
G-019-A - nanopore gDNA sequencing (Human) ***	8	Long reads: 32*** or Ultra long: 16***	NA
G-019-B - nanopore Ligation sequencing gDNA (Plants) ***	8	16 (long or ultra long***)	NA
G-019-C - nanopore Ligation sequencing gDNA (Animals) ***	8	16 (long or ultra long***)	NA

** The PromethION flowcell (Oxford Nanopore) is a single-use chip designed for long-read sequencing. The amount of Gb that one flowcell can produce depends on the sample quality, sample type, and library preparation.

*** Long reads: length of reads around 15 kb

Ultra long reads: length of reads around 70 kb

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 HS 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 2 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 (HAC basecalling)
- QC report

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 rapid 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 rapid barcoding kit includes reagents to tag each gDNA sample with a unique barcode. This is typically done by addition of barcode adapters to the gDNA by means of a transposase.

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:

Rapid Barcoding Kit 96 V14 (SQK-RBK114.96)

The Rapid Barcoding Sequencing Kit 96 V14 is a standalone kit providing 96 unique barcodes to enable PCR-free multiplexing of genomic DNA samples in a rapid and simplified workflow. Library preparation is based on a transposase-mediated approach, in which template DNA is simultaneously fragmented and tagged with barcode sequences in a single step. Barcoded samples are then pooled, purified, and sequencing adapters are added to the tagged ends prior to loading onto the flow cell.

The kit is optimised for speed and ease of use, while maintaining high sequencing performance (Q20+ accuracy) on the latest R10.4.1 nanopore flow cells.

Libraries sequencing and NGS coverage:

For nanopore small gDNA sequencing applied to Bacteria genomes the number of flowcell needed to sequence each 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 12 batches/pools of samples considering that every batch/pool of samples can be composed of a maximum of 96 samples (max batch 96).

	Min flowcells**/project	Max flowcells**/project	Max samples if Data Analysis is requested
G-020 - Nanopore small gDNA sequencing (long reads)	4	12	NA

** The PromethION flowcell (Oxford Nanopore) is a single-use chip designed for long-read sequencing. The amount of Gb that one flowcell can produce depends on the sample quality, sample type, and library preparation.

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 50 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 (HAC basecalling)
- QC report

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.

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 electrical signals into the corresponding RNA sequence.

Bioinformatic analysis: Bioinformatic analysis of direct RNAseq data can be provided as a combined service by the National Facility for Data Handling and Analysis.

Please select: SID: NF62.04.01.

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-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 flowcells/project**	**Max flowcells**/project**	**Max samples if Data Analysis is requested**
G021 - Nanopore Direct RNA Sequencing	10	20	20

** The PromethION flowcell (Oxford Nanopore) is a single-use chip designed for long-read sequencing. The amount of Gb that one flowcell can produce depends on the sample quality, sample type, and library preparation.

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 (HAC basecalling)
- 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 NF62.04.01 - Long-read RNA sequencing analysis.

Access modality: Access to NF services.

Services available in combination with the National Facility for Data Handling and Analysis: NF62.04.01 - Long-read RNA sequencing 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 (for bulk RNA sequencing): 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 (for bulk RNA and cDNA sequencing): 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 is performed by the facility for sequencing starting from bulk RNA, but Users can submit for sequencing already prepared double-stranded cDNA (from bulk or single-cell samples) Library Preparation (for both single-cell and bulk RNA sequencing): Prepare the cDNA for sequencing by adding sequencing adapters to the ends.

Sequencing for both single-cell and bulk RNA 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 optimization may be required.

The protocol uses a strand switching method to select 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:

Single-cell transcriptomics sequencing from 5' cDNA prepared with 10x Genomics using SQK-LSK114

Single-cell transcriptomics sequencing from 3' cDNA prepared with 10x Genomics using SQK-LSK114 and EXP-PCA001A primer pair compatible with cDNA generated via 10X GenomicsGene Expression protocols is used to amplify 10 ng of cDNA. For the 3' cDNA protocol the amplified cDNA using a biotinilated primer annealing to the 3' of the cDNA is captured with streptavidin beads to enrich full-length cDNA and remove RT artifacts due to

priming internal on the RNA: the captured cDNA is enriched by a second round of PCR amplification.. 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 optimized to achieve sequencing accuracies of over 99% (Q20+) with high output on the latest nanopore Flowcells R10.4.1.

Libraries sequencing and NGS coverage:

For nanopore bulk RNA/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 flowcells/project**	**Max flowcells**/project**	**Max samples if Data Analysis is requested**
G-022 - Nanopore cDNA sequencing (bulk cDNA or single-cell cDNA from 10x Genomics protocol) (Human-Mouse)	10	20	NA

** The PromethION flowcell (Oxford Nanopore) is a single-use chip designed for long-read sequencing. The amount of Gb that one flowcell can produce depends on the sample quality, sample type, and library preparation.

Technical requirements:

For bulk RNA sequencing:

- Provide either three aliquots containing 800 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 cDNA from bulk or single-cell technologies other than 10x Genomics:

- two aliquots containing 100 ng of cDNA in 50ul should be provided by the User. The cDNA should be double stranded
- cDNA samples should be quantified by using a fluorometer (i.e Qubit/ Glomax) and their quality evaluated by Agilent Tape Station/Bioanalyzer.

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 (HAC basecalling)
- QC report

Access modality: 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.

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 for simplex sequencing: DNA damage is repaired using the NEBNext FFPE DNA Repair v2 Module. Ends are then repaired and dA-tailed using the NEBNext End Repair/dA-tailing module before the sequencing adapters (supplied in the Ligation Sequencing Kit V14) are ligated onto the prepared ends. The kit is optimized 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 optimized to recover short DNA fragments.

Library preparation method steps for multiplex sequencing: The Native Barcoding Kit 96 V14 is a kit providing 96 unique barcodes to enable PCR-free multiplexing of cfDNA samples The library preparation method is similar to the Ligation Sequencing Kit protocol; cfDNA damage is repaired using the NEBNext FFPE DNA Repair v2 Module. Ends are then 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 (maximum pool size recommended is 12 samples). 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 cell-free DNA sequencing applied to Human samples each sample or pool of samples will be sequenced in one PromethION Flowcell.

Within this call the National Facility for Genomics will provide:

	Min flowcells/project**	**Max flowcells**/project**	**Max samples if Data Analysis is requested**
G-023 - Nanopore cell-free DNA sequencing (Human)	16	48	NA

** The PromethION flowcell (Oxford Nanopore) is a single-use chip designed for long-read sequencing. The amount of Gb that one flowcell can produce depends on the sample quality, sample type, and library preparation.

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 for the simplex protocol or >6ng in 13 µl for the multiplex protocol. 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 (HAC basecalling)
- QC report

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

SID: G-024_G-031 - Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech)

Services description - the National Facility for Genomics will provide sequencing of pools of libraries prepared by the Users with the sequencing platforms NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech)

Those platforms are commonly used for: Whole Genome Sequencing (WGS), Exome Sequencing (WES), Transcriptome Analysis (bulk RNA-Seq, single-cell RNAseq, Spatial transcriptomics), Metagenomics, Epigenomics studies, Population Genomics.

Libraries sequencing and NGS coverage:

For Sequencing only 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 Billion reads required.

Sequencing only for a total of 60 billion reads 300 cycles (generating 150bp PE reads) and for a total of 60 Billion reads 200 cycles (generating 100bp PE reads). Projects that will require a minimum of 10 to a maximum of 60 billion reads will be accepted.

Within this call the National Facility for Genomics will provide:

	Min Billion reads/project	Max Billion reads/project	Max samples if Data Analysis is requested

G-024_G-031 -A - Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech) (10 Billion reads 150bp PE reads)	10	60	To discuss with NF DHA
G-024_G-031-B - Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech) (10 Billion reads 100bp PE reads)	10	60	To discuss with NF DHA

Technical requirements

- Libraries should be provided already pooled in a Low Bind tube (i.e. 1,5 ml Eppendorf tubes).
- The Pool molarity will be indicated by the NF staff based on the sequencing platform used for the project
- 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.
- For MGI T7 sequencing, please refer to the index multiplexing protocol for the barcode combination selection.

At the following link, a tool for calculating how to prepare the pool of libraries for sequencing on Illumina Platforms 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 BulkRNA-Seq analysis
- NF62.01.02 miRNA analysis
- NF62.01.03 lncRNA analysis
- NF62.02.01 WGS analysis
- NF62.02.02 WES analysis
- NF62.02.03 Metagenomics/Microbiome analysis

  NF62.03.01 scRNA-seq analysis

  NF62.03.03 Single-Cell Immune Profiling – V(D)J

Access modality: 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 abovementioned Data Handling and Analysis additional services.

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 a minimum of 5 to a maximum of 10 P2/P3 Flowcells RUNs will be accepted.

	Min flowcell*/project**	**Max flowcells***/project**	**Max samples if Data Analysis is requested**
G-025-A - Sequencing only (NextSeq) (P3 Flowcells 300 cycles generating 150bp PE reads)	5	10	To discuss with NF DHA
G-025-B - Sequencing only (NextSeq) (P3 Flowcells 200 cycles generating 100bp PE reads)	5	10	To discuss with NF DHA
G-025-C - Sequencing only (NextSeq) (P3 Flowcells 600 cycles generating 300bp PE reads)	5	10	To discuss with NF DHA

*** One P3 Flowcell generates up to 1.2 Billions SE reads

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 on Illumina Platforms 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 BulkRNA-Seq analysis
- NF62.01.02 miRNA analysis
- NF62.01.03 lncRNA analysis
- NF62.02.01 WGS analysis
- NF62.02.02 WES analysis
- NF62.02.03 Metagenomics/Microbiome analysis

NF62.03.01 scRNA-seq analysis

NF62.03.03 Single-Cell Immune Profiling – V(D)J

Access modality: 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 abovementioned Data Handling and Analysis additional services.

SID: G-027 – Suspension BLISS (sBLISS)

Service description - Suspension BLISS (sBLISS – in-suspension Breaks Labeling In Situ and Sequencing) is a genome-wide method for identifying and mapping DNA double-strand breaks (DSBs) at nucleotide resolution (Bouwman et al., 2020).

DSBs are among the most critical forms of DNA damage and play key roles in both physiological and pathological processes, including replication stress, transcription-associated fragility, and the cellular response to genotoxic agents such as chemotherapeutics, radiomimetic compounds, genome-editing nucleases, or restriction endonucleases.

sBLISS generates quantitative profiles of both endogenous and induced DSBs in any cell type that can be brought into suspension, and has been successfully applied to cultured cells, organoids, and dissociated tissues. By enabling direct detection of DSBs, sBLISS supports research on DNA repair mechanisms, genome stability, and the effects of genotoxic compounds across diverse biological systems.

sBLISS Workflow overview:

Note: *The NF Genomics sBLISS protocol has been adapted from* Bouwman et al., 2020 *with some modifications.*

9. **Library preparation:** Standard sBLISS workflow requires an input of 1M fixed cells. Following nuclei permeabilization, DSB ends are blunted, ligated in situ with BLISS adapter containing UMI, sample barcode, T7 promoter binding site and a common sequence for amplification (RA5). After DNA extraction and fragmentation, library is enriched for BLISS adaptor-ligated DSB sites via in vitro transcription. The produced ssRNA is ligated to a second adapter ligation (RA3) and amplified with Illumina-compatible UDI primers to generate sequencing libraries.
10. ***Sequencing***: Libraries are sequenced on Novaseq X Plus instruments to generate a target of 50M 100 bp Paired End reads/sample.
11. ***Data processing***: Sequencing QC, summary statistics, genome-wide DSB mapping will be provided.

Request involving samples derived from sources other than human and mouse cell lines (e.g. organoids, isolated from tissues, ...) or low-input setups will be evaluated by the Facility Staff upon request, as part of a "**proof-of-concept**" projects, to determine their technical feasibility.

Within this call, the National Facility for Genomics provides:

	Min samples/project	**Max samples/project**	**Max samples if Data Analysis is requested**
G-027 – Suspension BLISS (sBLISS) for cell lines (human, mouse)	6	24	NA

Technical requirements:

- **Sample type**: 2 million cell suspension ($2*10^6$ cells/ml) in 1× DPBS (no $MgCl_2$ or $CaCl_2$). Cells must be fixed with methanol-free paraformaldehyde (final concentration 2%); DNase-free conditions. Cell concentration must have been assessed by vital counts using either manual or automated cell counter with vital dye.
- **Purity/quality**: Cell viability before fixation ≥ 80%; absence of debris or clumps.
- **Format**: Fixed cells should be provided in protein low bind tubes (i.e. 1,5 ml Eppendorf tubes)
- **Shipping**: fixed cells must be shipped at 4 °C. Do not freeze cells.

Please note: Project specific requirements for preparing and shipping cells must be discussed with NF personnel upon project approval.

Results that will be delivered to the Users:

- FASTQ files
- FASTQC report
- Mapping metrics
- List of detected DSBs

- Genome-wide DSB counts table across bin sizes (e.g. 10KB, 50KB, 100KB)
- DSB counts across gene, promoters, and eventual user-provided regions of interest

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

SID: G-028 – directCAGE

Service description - Direct Cap Analysis of Gene Expression (direct CAGE) is an high-throughput method which enables genome-wide profiling of transcription start sites (TSS) at single-base resolution, allowing precise mapping of promoter and enhancer initiation events, as well as quantitative measurement of gene expression across different cell types and differentiation states (Delobel, D et al.,2025)

This method is based on the selection of 5' ends of capped RNAs ('cap-trap', Carninci et al. 1996). CAGE tags are mapped to the reference genome to identify TSS and their related promoter regions. The number of CAGE tags mapped at specific locations in the genome provides quantification on promoter activities on a genome-wide scale.

This protocol can be used for the detection of differential gene expression between different cell types (tissue specific expression), different stages of cell differentiation (regulation of cell differentiation), determination of promoter and enhancer usage.

Workflow overview:

Note*: The NF Genomics directCAGE protocol has been adapted from* (Delobel, D et al.,2025) *for automation on liquid handling platforms.*

- ***Library preparation*:** Starting from 5 ug purified RNA, first strand cDNA synthesis is performed by random priming. On the cDNA/RNA hybrid molecules, the cap structure of the RNA is chemically modified (oxidized with sodium periodate) and biotinylated. After RNaseI treatment to select full-length cDNA, capped sequences are specifically captured using magnetic streptavidin beads (cap-trap reaction). Single-stranded cDNA is ligated to a 5'-end adapter, and then to the 3'-end adapter. After 2nd strand synthesis, final indexed dsDNA directCAGE libraries are pooled for sequencing.
- ***Sequencing***: DirectCAGE libraries are sequenced on Novaseq X Plus instrument to generate a target of 50M 100 bp Paired End reads/sample.
- ***Data processing***: The NF for Genomics will perform data pre-processing with an analysis pipeline that trims adapters and removes rRNA reads, then maps the remaining sequences to the genome. The 5′ ends of aligned reads are extracted to define CAGE transcription start sites (CTSS) which are then clustered to identify promoter-associated TSS regions (transcribed cis-regulatory elements, tCREs). The resulting CTSS tables and TSS clusters will be provided to the user for secondary data analysis (i.e., quantification, comparison between samples, and visualization of promoter activity), which will be the responsibility of the user.

Within this call, the National Facility for Genomics provides:

	Min samples/project	Max samples/project	Max samples if Data Analysis is requested
G-028 – direct CAGE on eukaryotic (human/mouse) RNA	8	48	NA

Technical requirements:

- **Sample type**: For each sample, we require a minimum of 5 µg (1 µg/µL) of purified RNA
- **Purity/quality**: RNA quality must be: A260/230 > 1.8, A260/280 > 1.8, RIN > 7 (assessed via TapeStation, Bioanalyzer, Fragment Analyzer, ...)
- **Format**: 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.
- **Shipping**: dry ice

Results that will be delivered to the Users:

- FASTQ files
- FASTQC report
- Alignment results (BAM)
- Matrix with the TSS and tCREs counts for all the samples
- CSV files with TSS and tCREs annotation

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

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)	M C v (
G-001 -Whole Genome Sequencing (WGS) Illumina PCR Free	Low-bind full-skirted PCR plates	2		15		50		6	1,8	1,8			
G-001 -Whole Genome Sequencing (WGS) Neb Ultra II	Low-bind full-skirted PCR plates	2		4		60		na	1,8	1,8			
G-002 - Whole Exome Sequencing (WES)	Low-bind full-skirted PCR plates	2		5		40		6	1,8	1,8			
G-003 - Amplicon sequencing for microbiome analysis (16S-ITS)	Low-bind full-skirted PCR plates	2		0,1		20		6	1,8	1,8			
G-004 - Methylation sequencing (Methyl-seq)	Low-bind full-skirted PCR plates	2		15		60		6	1,8	1,8			
G-005 - mRNA sequencing from standard input	Low-bind full-skirted PCR plates	2	500			20	50	7	1,8	1,8	x		
G-005 - mRNA sequencing from low input	Low-bind full-skirted PCR plates	2	0,2			15	20	7	1,8	1,8	x		

G-006 - totalRNA sequencing from standard input	Low-bind full-skirted PCR plates	2	500			20	50	4	1,8	1,8	x		
G-006 - totalRNA sequencing from low input	Low-bind full-skirted PCR plates	2	0,2			15	20	4	1,8	1,8	x		
G-007 - smallRNA sequencing	Low-bind full-skirted PCR plates	2	50			20	50	7	1,8	1,8	x		
G-008/G-012 - Single-cell 3'RNAsequencing (10X Genomics)	Sterile cryogenic storage vials	2										1	9
G-008.1/G-012 - Single-cell gene Expression Flex (10X Genomics)	Sterile cryogenic storage vials	2										0,1	9
G-009/012/014 – Single-cell Immune profiling-V(D)J (10X Genomics or BD Rhapsody)	Sterile cryogenic storage vials	2										1	9
G-010/012/013 – Single-cell multiome ATAC + Gene expression (10X Genomics or BD Rhapsody)	Sterile cryogenic storage vials	2										1	9
G-015/016 - Spatial Transcriptomics analysis from Fresh-Frozen tissues	See manifacturer's handbook for	2						4					

	slide compatibility												
G-015/016 - Spatial Transcriptomics analysis from Fixed Frozen tissues	See manifacturer's handbook for slide compatibility	2						50%					
G-015/016 - Spatial Transcriptomics analysis from Fresh Frozen Visium HD 3' tissues	See manifacturer's handbook for slide compatibility	2						7					
G-015/016 - Spatial Transcriptomics analysis from FFPE Visium HD tissues	See manifacturer's handbook for slide compatibility	2						30%					
G-029/030 - Spatial Transcriptomics analysis from Fresh Frozen STOmics Stereo-seq tissues	See manifacturer's handbook for slide compatibility	2						4					
G-029/030 - Spatial Transcriptomics analysis from FFPE tissue STOmics Stereo-seq OMNI tissues	See manifacturer's handbook for slide compatibility	2						30%					
G017/018 - GeoMx Digital Spatial Profiling	See manifacturer's	2											

from Fresh-Frozen, Fixed Frozen or FFPE tissues (Nanostring)	handbook for slide compatibility												
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		20		30		6	1,8	1,8			
G021 - Nanopore Direct RNA Sequencing	Low-bind 1.5 ml tubes	3		1200		10		8	1,8	1,8	x		
G022 - Nanopore cDNA sequencing (bulk cDNA) from RNA	Low-bind full-skirted PCR plates	3	200			50		8	1,8	1,8	x		
G022 - Nanopore cDNA sequencing (bulk cDNA) from cDNA	Low-bind full-skirted PCR plates	2	100			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) (simplex protocol)	Low-bind full-skirted PCR plates	2	30			50			1,8	1,8			
G023 - Nanopore cell-free DNA sequencing (Human) (multiplex protocol)	Low-bind full-skirted PCR plates	2	>6			13			1,8	1,8			
G-024_G-031 - Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech)	Low-bind 1.5 ml tubes	2			See manifacturer's guidelines	See manifacturer's guidelines							
G025 - Sequencing only with NextSeq 2000 (Illumina) (P3 Flowcells)	Low-bind 1.5 ml tubes	2			5	60							
G-027 – Suspension BLISS (sBLISS)	Protein Low-bind 1.5 ml tubes	1										2	8
G-028 – directCAGE	Low-bind full-skirted PCR plates	1	5000	1000				7					

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

Table 1: *Overview of the NF for Genomics services which can be combined with the Data Analysis services provided by the NF for Data Handling and Analysis (DHA).*

G-Service code	NF-Genomics Service Name	NF-Data-Service code	NF-DATA 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	Metagenomics/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-006	totalRNA from standard input	NF62.01.03	lncRNA analysis
G-007	Small RNA sequencing	NF62.01.02	miRNA analysis
G-008_G-008.1/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 or BD Rapsody)	NF62.03.03	Single-cell immune profiling (VDJ)
G-010/012/013	Single-cell multiome ATAC + Gene expression (10X Genomics or BD Rapsody)	NF62.03.04	Single-cell multiome (ATAC + gene expression)
G-015/016_G-029/030	Spatial Transcriptomics analysis from Fresh-Frozen, Fixed Frozen or FFPE tissues using Visium HD (10X Genomics) or StereoSeq (STOmics-MGI).	NF62.03.05	Spatial transcriptomics (10X Visium / StereoSeq platforms)
G-021	Nanopore Direct RNA Sequencing	NF62.04.01	Long-read RNA sequencing
G-024_G-031	Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech)	NF62.01.01	Bulk RNA-Seq analysis
G-024_G-031	Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech)	NF62.01.02	miRNA analysis

G-024_G-031	Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech)	NF62.01.03	lncRNA analysis
G-024_G-031	Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech)	NF62.02.01	WGS analysis
G-024_G-031	Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech)	NF62.02.02	WES analysis
G-024_G-031	Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech)	NF62.02.03	Metagenomics/Microbiome Analysis
G-024_G-031	Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech)	NF62.03.01	scRNA-Seq analysis
G-024_G-031	Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech)	NF62.03.03	Single-cell immune profiling (VDJ)
G-025	Sequencing only with NextSeq 2000 (Illumina)	NF62.01.01	Bulk RNA-Seq analysis
G-025	Sequencing only with NextSeq 2000 (Illumina)	NF62.01.02	miRNA analysis
G-025	Sequencing only with NextSeq 2000 (Illumina)	NF62.01.03	lncRNA 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	Metagenomics/Microbiome Analysis
G-025	Sequencing only with NextSeq 2000 (Illumina)	NF62.03.01	scRNA-Seq analysis
G-025	Sequencing only with NextSeq 2000 (Illumina)	NF62.03.03	Single-cell immune profiling (VDJ)

Please see Table 3.5 in APPENDIX 3 for the suggested maximum number of samples and comparison allowed for each service type. In case of significant deviations from these limits, please contact the National Facility before submitting your application.

A detailed description of DHA service is available in the following section.

Table of Contents

Data Handling and Analysis additional services 91
NF62.01.01 Bulk RNA-seq analysis 91
NF62.01.02 miRNA analysis 93
NF62.01.03 lncRNA analysis 96
NF62.02.01 WGS analysis 98
NF62.02.02 WES analysis 100
NF62.02.03 Metagenomics/Microbiome Analysis 103
NF62.02.04 Methyl-Seq analysis 105
NF62.03.01 scRNA-seq analysis 107
NF62.03.02 scATAC-seq analysis 110
NF62.03.03 Single-cell Immune profiling-V(D)J 111
NF62.03.04 Single-cell multiome (ATAC + gene expression) analysis 114
NF62.03.05 Spatial transcriptomics (10X Visum / StereoSeq platforms) 115
NF62.04.01 Long-read RNA sequencing 117

Appendix 3: Description of the Data analysis services available in combination with the NF for Genomics services 120
3.1 Metadata file example 120
3.2 Glossary of terms 121
3.3 Cited databases 122
3.4 Tools used 124
3.5 Maximum project dimensions 127

Data Handling and Analysis additional services

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), Multi-Dimensional Scaling (MDS), and sample-to-sample correlation analysis are performed to assess the variability structure of the dataset, the correlation structure among samples, and their possible relationship with sample 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.
9. **Automatic functional enrichment analysis**: For human and mouse datasets, an over-representation analysis (ORA) is performed to test the enrichment of the lists of differentially expressed genes against Gene Ontology and the main pathway collections (e.g. KEGG, Reactome, BioCarta, Hallmark). A pre-ranked Gene Set Enrichment Analysis

(GSEA) is also performed on the same pathway collections. For other species, the two analyses are performed only when suitable Gene Ontology annotations and/or pathway collections are available for the corresponding organism.

Advanced (optional) analysis steps include the following:

1. **Advanced pathway analysis of significant genes**: Advanced pathway analysis can be performed using industry standard tools. Significant differentially expressed genes are analyzed to identify enriched canonical pathways, upstream regulators, causal networks, biological functions, and disease associations. This analysis is available for human, mouse, and rat datasets. Support for other species will be discussed during the project kickoff meeting.
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 the example provided in Appendix 3.1. 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 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. BAM files for each sample.
2. Raw and normalized count matrices containing expression values for each gene in each sample.

3. Results from exploratory analysis: PCA and MDS scores and plots, table with sample-to-sample correlation values and the corresponding heatmap, plots showing the expression levels of control genes.
4. Tables of differentially expressed genes with statistical significance information, and associated diagnostic plots (volcano plots, histograms of nominal pvalues).
5. Results from functional enrichment: tables of enriched pathways with statistical significance information, GSEA enrichment curves.
6. 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.
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-005 - mRNA sequencing from standard and low input
- G-006 - totalRNA from standard input
- G-024_G-031 - Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech)
- 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. **miRNA identification**: Canonical and non-canonical miRNAs are identified. An interactive report is produced with an overview of all detected miRNAs. This service is available for organisms having a complete and well-annotated reference genome.
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. This service is available for organisms having a complete and well-annotated reference genome.

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).

For differential analysis, we recommend a design with at least two biological conditions with ≥3 replicates each to ensure adequate statistical analysis. 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 the example provided in Appendix 3.1. 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 small RNA 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.
5. BAM files for each sample.
6. Raw and normalized count matrices containing expression values for each miRNA in each sample.
7. Tables of differentially expressed miRNAs with statistical significance information.
1. 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.
8. 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.
- G-024_G-031 - Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech)
- 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.

NF62.01.03 lncRNA analysis

Service description

Long non-coding RNAs (lncRNAs) regulate gene expression without encoding proteins, and can affect transcription, mRNA stability and translation, protein activity, and cell signaling by recruiting chromatin-modifying complexes, influencing splicing, and acting as scaffolds for protein complexes. They play a key role in cell development and differentiation, through embryonic patterning and stem cell maintenance. The study of lncRNA expression can reveal previously overlooked regulatory layers and mechanisms, that play an important role in many highly relevant diseases (cancer, neurological disorders, cardiovascular and metabolic diseases).

LncRNA analysis processes raw sequencing reads through quality control, alignment and/or de novo transcriptome assembly, then performs lncRNA prediction using multiple classifiers, and differential expression analysis.

The standard analysis pipeline includes the following steps:

1. **Quality check** of the raw sequence data: Evaluating sequencing quality of raw reads, assessing nucleotide composition, 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. Removal of rRNA. Performing a second QC step on the trimmed reads.
3. **Mapping and Assembly**: High-quality reads are mapped to a reference genome, generating a BAM file.
4. **Filtering**: Transcripts matching known coding RNAs are discarded. The remaining transcripts are filtered by length and exon counts.
5. **Quantification**: The relative abundances of lncRNA species are computed from the counts of reads aligning to each transcript.
6. **Differential expression**: Identification of over/under-expressed lncRNAs following the provided experimental design.

Advanced (optional) analysis steps include the following:

1. **Classification**: lncRNA can be classified on the basis of their genomic location, their regulatory potential, and their putative function.
2. **Identification of novel lncRNAs:** Consensus of coding potential predictors, each transcript is assessed and novel transcripts with no coding potential outcome are kept.
3. **Co-expression analysis:** If mRNA data are available, a lncRNA-mRNA correlation can be performed by co-expression network analysis.

Access modality available

- Access to facility service

Requested inputs from Users

All fastq files associated to all the samples must be provided in gzip-compressed form, together with the corresponding md5 files (unless sequencing was performed by the National Facility for Genomics).

For differential analysis, we recommend a design with at least two biological conditions with ≥3 replicates each to ensure adequate statistical analysis. 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 the example provided in Appendix 3.1.

Unless the organism under study is human or a model organism, the User shall specify 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.

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. Complete lncRNA catalogue: Novel + known lncRNAs with multi-classifier confidence scores (as gtf and tsv files).
4. Expression quantification: Ready-to-use matrices for all transcripts (as tsv/csv files).
5. Differential expression: Statistically significant lncRNAs between conditions (tsv/csv files, plots).
6. Quality metrics: QC reports + lncRNA-specific statistics (length distribution, exon count, coding potential)
7. Complete analysis report as a customized MultiQC report.

Combined services

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

- G-006 – TotalRNA from standard input
- G-024_G-031 - Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech)
- 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.

NF62.02.01 WGS analysis

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 the example provided in Appendix 3.1.

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].

In the case of tumour studies, we strongly recommend providing both tumour and matched normal samples.

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.
5. BAM files for each sample.
6. 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.
1. 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.

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-001 - Whole Genome Sequencing
- G-024_G-031 - Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech)
- 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.

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

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 the example provided in Appendix 3.1.

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 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.
6. BAM files for each sample.
7. 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.
8. 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.
9. 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_G-031 - Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech)
- SID: 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.

---

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

NF62.02.03 Metagenomics/Microbiome Analysis

Service description

Metagenomic and microbiome analysis provides a comprehensive bioinformatics service for characterizing biological communities from high-throughput sequencing data. The service is designed to profile the composition and diversity of microbiomes across environmental, clinical, agricultural and host-associated samples.

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

For microbiome analysis, the expected input data are single-end or paired-end FASTQ files from Illumina from 16S rRNA (V3-V5 regions) or ITS amplicon sequencing. For shotgun metagenomic analysis, the expected input data are paired-end FASTQ files from whole-metagenome/genome sequencing reads. Host reference information may be provided for host-read removal.

The standard bioinformatics analysis focuses on:

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. Host or contaminant read removal will be performed when a suitable reference is provided.

   Abundance and relative abundance: Calculating abundance based on the computed ASVs/OTUs (if applicable) 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.
3. **Quality report:** Aggregating QC, preprocessing, classification and workflow execution metrics into summary reports.

Microbiome standard analysis also includes:

1. **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.
2. **Taxonomic classification**: Clustering reads into operational taxonomic units (OTUs) or ASVs, referring to the SILVA database for 16S and the UNITE database for ITS.
3. **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.

Shotgun metagenomics standard analysis also includes:

1. **Taxonomic profiling of sequencing reads:** Classifying metagenomic reads against databases using read-base taxonomic profilers. This step estimates the microbiomal

composition of each sample and generates taxonomic abundance profiles across supported taxonomic levels.

Advanced (optional) analysis steps include the following:

1. **Differential abundance analysis (DAA)**: Differential abundance analysis can be performed as a downstream statistical analysis when appropriate sample metadata, experimental groups and biological replication are available. For microbiome analysis, DAA can be applied to ASvs, OTUs or taxonomic abundance tables. For shotgun metagenomics, DAA can be applied to taxonomic abundance tables generated from read-basd profilers.

Microbiome specific advanced analysis:

1. **Alpha diversity rarefaction curves**: Produce rarefaction plots displaying alpha diversity indices that determine samples richness.
2. **Functional abundances**: Functional abundances are predicted based on marker gene sequences. Enzyme Classification numbers and KEGG orthologs will be predicted for each sample.

Advanced shotgun Metagenomics analysis:

1. **Metagenome assembly:** Assembling quality-filtered reads into contigs when the study design, sequencing depth and sample complexity support assembly-based analysis.
2. **Contig binning:** grouping assembles contigs into genome bins and refining bins to reconstruct metagenome-assembled genomes.
3. **Protein-coding gene prediction:** Predicting coding sequences from assembled contigs if data supports it.
4. **Viral sequence identification:** Identifying putative viral sequences from assembled contigs if data supports it.
5. **MAG quality assessment:** Assessing reconstructed genome quality, including completeness and contamination.

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 the example provided in Appendix Metadata file example.

If differential abundance analysis is requested, Users should provide the list of conditions to be compared. If host-removal is requested, Users should provide host reference information.

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. An example is provided in Appendix 3.1.

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-024_G-031 - Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech)
- SID: 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.

NF62.02.04 Methyl-Seq analysis

Service description

Methyl-Seq is a robust NGS technique that uses bisulfite sequencing to analyze DNA methylation patterns. Methylation refers to the addition of methyl groups to cytosines, primarily at CpG sites, which is a key epigenetic modification regulating gene expression. Methyl-Seq provides a cost-effective way to profile methylation at a single-nucleotide resolution and can be applied at a genome-wide scale, on predefined target regions (focusing on high-CpG-density regions, such as gene promoters or CpG islands), or on a size-selected fraction of the genome (Reduced-Representation Bisulfite Sequencing, RRBS).

Methylation analysis is widely employed in epigenomics research, developmental biology, and disease studies, particularly in cancer, to investigate DNA methylation changes potentially associated with gene regulation. The technique helps to identify differentially methylated regions (DMRs) that could influence gene expression, and it provides valuable insights into how epigenetic modifications affect cellular processes and contribute to disease development.

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

1. **Quality check of raw sequence data**: Evaluating raw read sequencing quality, assessing nucleotide composition of the reads, K-mer over-representation, duplication rate, and adapter content**.** Exclusion of reads showing evidence of incomplete conversion.
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 trimmed reads to a properly processed reference genome (to account for bisulfite conversion). If data were produced with a target sequencing approach, evaluating the coverage on targeted regions.
4. **Spike-in evaluation**: Spike-ins are genomic materials from external sources (e.g. synthetic sequences, cell lines or bacteria like E. Coli) added in low-concentration to the samples. Since their methylation status is known, they are used as controls to assess the effectiveness of the bisulfite conversion procedure.
5. **Methylation calling**: Inferring the methylation status of each cytosine via the proportion of bisulfite converted and unconverted reads. Reporting the methylation percentage of each cytosine in each sample in tabular format.
6. **Filtering and normalization**: Filtering data to include only cytosines covered by a minimum number of reads on a minimum number of samples and normalized according to the different read coverage distributions between samples. Producing descriptive statistics of coverage and methylation percentages.
7. **Summary statistics on methylation data**: Calculating correlations between each pair of samples based on their methylation profiles. Performing Principal Component Analysis (PCA) and unsupervised clustering to inspect the variability structure of the data and its possible relationship with samples characteristics.

Advanced (optional) analysis steps include the following:

1. **Differential methylation analysis**: Methylation percentages are compared between different groups of samples, using statistical models based on the experimental design (e.g. paired models, regression of covariates etc.). Depending on the sequencing approach, the size of the dataset and the aim of the study, differential methylation (DM) could be performed on the single cytosine positions or on aggregated genomics regions.
2. **Annotation of differentially methylated regions**: The differentially methylated cytosines/regions identified with DM analysis are annotated according to overlapping or proximal genomic features such as CpG islands, gene promoters or Transcription Start Sites (TSS).
3. **Functional enrichment of differentially methylated regions**: An over-representation analysis is performed to test the enrichment of the identified differentially methylated features against Gene Ontology and the main pathway collections (e.g. KEGG, Reactome, Biocarta, Hallmark, IPA).
4. **Methylation scores calculation on specific sites**: The methylation status of a predefined set of positions is evaluated and the average methylation score is calculated for each sample across all positions. This calculation is possible only with targeted sequencing, that ensures that the same set of positions is consistently covered among all samples.

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 the example provided in Appendix 3.1.

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 of 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 is preferred).

We recommend a minimum sequencing depth of 80 million reads per sample for whole-genome methylation analysis mammalian-sized genomes (this limit can be reduced in the case of smaller genomes), and a Q30 cutoff of 80%. For RRBS and targeted approaches, recommended coverage will be evaluated on a case-by-case basis.

Results

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

1. Trimmed and filtered fastq files for each sample.
2. BAM files for each sample.
3. Tables containing methylation percentages and supporting coverage for each cytosine in each sample.
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. Interactive plots (e.g. clustering dendrograms, PCA/MDS, histograms, heatmaps, etc. for standard analysis; volcano plots, Circos plots, etc. for advanced analyses) and tables included to the report will also be provided as separate files.
5. Pipeline and scripts used to perform the analysis.

Combined services

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

- G-004-B Methyl-seq for studies on diseases (humans, organoids derived from human cells).

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 concatenation**: Merging expression data from all sequenced samples into one single dataset on which the overall analysis is performed. Grouping of the samples into merged 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.
8. **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.

Advanced (optional) analysis steps include the following:

1. **“Advanced” cell type annotation**: custom cell type annotation based on user-provided signatures and large-language models.
2. **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.

3. **Pathway analysis and gene ontology analysis** of gene signatures coming from the advanced DGE model.
4. **Differential abundance analysis**: Testing whether the proportions of specific cell types vary across different types of samples.
5. **Cell-cell interactions**: Inspecting communication across different cell types through cell type-specific expression of signalling molecules such as ligands, receptors, and their downstream signalling pathways.
6. **Trajectory analysis**: When the experimental design allows it, inferring transcriptional changes related to developmental processes, cell proliferation or response to stimuli, via a pseudo-time trajectory.
7. **Copy-number inference**: Inferring large-scale copy number variations (CNVs) from single-cell RNA-seq gene expression profiles to identify genomic aberrations and distinguish malignant from non-malignant cell populations in tumour samples.

Requested inputs from Users

Raw files associated with all samples must be provided (FASTQ files are strongly preferred), 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 the example provided in Appendix 3.1.

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 (in h5ad format).
4. Complete reports (in HTML and publication-ready PDF formats) describing the quality of the data, all the analysis performed on the dataset, and their results. Interactive 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 (.h5mu) 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
- G-024_G-031 - Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech)
- 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.

NF62.03.02 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:** Aligning high-quality sequencing reads to the reference genome and generating a fragment file for each sample.
3. **Barcode counting and cell calling:** Identifying cell-associated barcodes and distinguishing true cells from background based on fragment count distributions.
4. **ATAC modality cell-level and sample-level QC metrics collection:** Identifying low-quality cells based on ATAC-specific quality metrics, including transcription start site (TSS) enrichment score, nucleosome signal, and the fraction of fragments in genomic blacklist regions. Sample-level quality is assessed using additional metrics such as the fraction of fragments in peaks (FRIP).

Advanced (optional) analysis steps include the following:

1. **Pseudobulk peak calling:** Calling accessible chromatin peaks on pseudobulk profiles generated by aggregating cells within each sample (or cluster), providing a robust consensus peak set for downstream analyses.
2. **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) to emphasize most informative features. Using the most variable features within each dataset for dimensionality reduction (e.g. SVD, PCA, UMAP).

3. **Differential accessibility analysis**: When the experimental design allows it, differential accessibility region (DAR) analysis is performed to detect differences in chromatin accessibility across sample conditions.
4. **Characterization of differentially accessible peaks:** this may include functional analysis of enriched peaks, motif enrichment, etc.

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 the example provided in Appendix 3.1.

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 peak-by-cell count matrices generated using a merged consensus peak set derived from sample-specific pseudobulk peak calling.
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 (in .h5mu format) 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*.

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). GEX data analysis will be performed as described under service **NF62.03.01** (scRNA-seq analysis).

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

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 labelled 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:

1. **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.
2. **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 the example provided in Appendix 3.1.

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 (in .h5mu format) 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)
- G-024_G-031 - Sequencing only with NovaSeq X Plus (Illumina), DNBSEQ-T7 (MGI Tech)
- 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.

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

IMPORTANT – this service will not be offered in this call round.

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 **NF62.03.01 scRNA-seq analysis**.
2. For the ATAC modality, refer to **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 raw files associated to all the samples must be provided (FASTQ files are strongly preferred), 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 the example provided in Appendix 3.1.

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%.

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.

1. BAM files for each sample and modality.
2. Count matrices containing expression values for each gene in each cell (in h5ad format).
3. Raw peak-by-cell count matrices generated using a merged consensus peak set derived from sample-specific pseudobulk peak calling.
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, volcano plots, PCA/MDS) and tables included to the report will also be provided as separate files.
5. Python objects (in .h5mu format) 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-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 / StereoSeq platforms)

IMPORTANT – this service will not be offered in this call round.

Services description

Spatial transcriptomics 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.

This service is available for data produced using the Visium Spatial Gene Expression solution from 10X Genomics or the StereoSeq platform from MGI.

The standard bioinformatics analysis for a spatial transcriptomics dataset comprises the following steps:

1. **Quality check of the original sequences**: Evaluating sequencing quality of raw reads, assessing the most relevant fastQ quality metrics.
2. **Spatial barcodes identification and extraction**: Identifying spatial barcodes associated with tissue-covered regions,and assigning sequencing reads to their corresponding spatial locations to generate a spatial gene expression matrix for downstream analyses.
3. **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 for each spatial unit..
4. **Quality control:** calculate spatial feature-level quality metrics (e.g. number of total reads, number of detected genes, percentage of reads associated to mitochondrial genes) to identify low -quality spatial units.

Advanced (optional) analysis steps include the following:

1. **Nuclei/cell and tissue-domains segmentation:** When high-resolution microscopy images are available, nuclei and cells are segmented to reconstruct cell-level profiles. Tissue images can also be segmented into biologically relevant domains to characterize tissue architecture and support downstream spatial analyses. Evaluation of biological relevance of segmentation.
2. **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. Depending on the data and segmentation quality, the analysis is performed on spatial units (e.g., spots or bins) or at the **reconstructed cell level**. Cluster assignments are then mapped back onto the tissue to visualize the spatial distribution of distinct cellular or transcriptional populations.
3. **Spot deconvolution**: Since spatial units (e.g., spots or bins) may contain multiple cells, their cellular composition can be inferred by deconvolving the mixed gene expression profiles using matched single-cell RNA-seq datasets . This can be achieved using a matched single-cell RNA-seq dataset generated under the same experimental conditions (required) or relying on publicly available data.
4. **Spatial gene mapping**: Identification of spatially variable genes or evaluation of the spatial expression of custom gene signatures.
5. **Comparison across conditions:** When supported by the experimental design, cell-level profiles are aggregated into pseudo-bulk samples for differential expression analysis between biological conditions using biological replicates, and functional analysis can be performed.

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 the example provided in Appendix 3.1.

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

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, Visium HD, or StereoSeq platforms 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/reconstructed 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. 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 - Spatial Transcriptomics analysis from Fresh-Frozen, Fixed Frozen or FFPE tissues using Visium HD (10X Genomics) or StereoSeq (STOmics-MGI). To access the combined services, please submit an application to the National Facility for Genomics requesting data analysis.

NF62.04.01 Long-read RNA sequencing

Services description

Long-read sequencing platforms allow sequencing full-length RNAs, enabling accurate analysis of alternative splicing, isoform abundance, and post-transcriptional modifications. When direct RNA sequencing is performed, this technology also allows for the detection of methylation and RNA modification events.

The analysis pipeline performs transcript assembly from direct RNA reads. The pipeline processes raw sequencing reads through quality control, alignment and pseudoalignment, transcript annotation, differential expression analysis, differential transcript usage and comprehensive reporting.

Standard analysis includes the following steps:

1. **Preprocessing of raw reads and quality control:** Evaluate raw read quality, read-length distributions, and per-read quality scores characteristic of long-read data.
2. **Trimming:** Trimming reads by base quality and length, removing adapters and technical residuals, and discarding excessively short or low-quality reads. A second QC pass on the trimmed reads confirms improvement before downstream steps; the direct-RNA path runs a dedicated QC tailored to native RNA reads.
3. **Alignment and/or assembly of transcripts:** Mapping long reads to the reference genome with a splice-aware strategy that accounts for full-length reads spanning multiple exons. Assembling transcript structures per sample directly from the spliced alignments and merging them into a unified set. Comparing assembled transcripts against the reference annotation to classify them as known or novel.
4. **Transcript quantification:** Counting reads at the gene level using a long-read aware mode that assigns each full-length read to its feature. Estimating transcript-level abundances by resolving reads across overlapping isoforms. Producing count and abundance matrices used as the input for all downstream analyses.

Reporting, including custom statistics and visualizations: Aggregating quality, alignment, and quantification metrics from every sample into a single consolidated report. Generating

genome-browser coverage tracks so alignments and assemblies can be inspected visually. Summarizing counts and pipeline statistics into tables and plots.

Advanced analysis includes the following options:

1. **Differential expression analysis:** Modeling gene-level counts across conditions and testing for genes whose expression changes between experimental groups. Estimating fold changes and applying multiple-testing correction to control false discoveries. Reporting significant genes with effect sizes and adjusted significance for interpretation.
2. **Differential transcript abundance analysis:** Testing whether the relative proportions of a gene's isoforms shift between conditions, independent of total gene expression. Detecting isoform-switching events where the dominant transcript of a gene changes between groups. Highlighting switches with potential functional consequences, exploiting the isoform resolution that long reads provide.
3. **Detection and quantification of RNA modifications:** Interrogating the raw electrical signal of native direct-RNA reads, where chemically modified bases perturb the current as the molecule passes through the pore. Calling modified positions and estimating their stoichiometry (the fraction of molecules carrying the modification at each site). Requires **raw signal data (FAST5/POD5)**; restricted to direct-RNA input.
4. **Detection and quantification of RNA methylation events:** Focusing on methylated modifications (notably N6-methyladenosine, m6A) as a subset of RNA modifications detectable from the same native-signal analysis. Reporting methylated sites, their per-site methylation level, and, if the design experiment is appropriate, differential methylation between experimental groups. Mapping the methylation landscape onto specific transcripts and positions to link it to post-transcriptional regulation. Requires **raw signal data (FAST5/POD5)**; restricted to direct-RNA input, with differential testing needing at least 2 conditions and ≥3 replicates each.

Access modality available

- Access to facility service

Requested inputs from Users

Sequence 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).

For differential analysis, we recommend a design with at least two biological conditions with ≥3 replicates each to ensure adequate statistical analysis. 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 input files. See the example provided in Appendix 3.1.

If detection and quantification of RNA modifications and/or RNA methylation events are requested, the User shall provide the **raw signal data (FAST5/POD5).**

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.

Results

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

1. Identified transcripts: Novel and known transcripts with multi-classifier confidence scores (gtf and tsv files);
2. Expression quantification: Ready-to-use matrices for all transcripts (tsv/csv files);
3. Differential gene expression: Statistically significant genes between conditions (tsv/csv files, plots);
4. Differential transcript usage: Statistically significant genes and transcripts between conditions (tsv/csv files, plots);
5. Detection and quantification of RNA modifications: Modified sites with per-site stoichiometry across transcript positions (tsv/csv files, plots);
6. Detection and quantification of RNA methylation events: Methylated sites with per-site level and, when the design allows, differential methylation between conditions (tsv/csv files, plots);
7. IGV tracks, allowing exploration of the assembled transcriptome using a genome browser;
8. Quality metrics: QC reports and transcript-specific statistics.

Combined services

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

G-021 – Nanopore Direct RNA Sequencing

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

Appendix 3: Description of the Data analysis services available in combination with the NF for Genomics services

3.1 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	Var 1	Var 2	Var …
control	sample 1	sample1_R1.fastq.gz	sample1_R2.fastq.gz			
control	sample 2	sample2_R1.fastq.gz	sample2_R2.fastq.gz			
treatment	sample 3	sample3_R1.fastq.gz	sample3_R2.fastq.gz			
treatment	sample 4	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

3.2 Glossary of terms

Section	Term	Definition
2.1	Signal-to-noise ratio (SNR)	Measure of how well the signal of interest coming from the sample can be distinguished from noise on the microscope detector.
2.1	Image restoration	Process of improving image quality by removing random noise, enhancing the signal-to-noise ratio (SNR).
2.1	Semantic and Instance Segmentation	Techniques to identify and separate distinct objects in an image, generating precise object boundaries (masks).
2.1	Morphometric Analysis	Measurement and analysis of shapes, sizes, and structures in biological images.
2.1	Single-Particle Analysis (SPA)	Computational technique for reconstructing 3D structures of macromolecules from 2D images obtained by cryo-electron microscopy.
2.1	Sub-tomogram Averaging (STA)	Method for enhancing resolution in 3D reconstructions of structures from cryo-electron tomography data by averaging multiple similar regions.
2.2	CTF	Contrast Transfer Function.
3.1	Splice-aware aligner	Tool that aligns RNA-seq reads, accounting for exon-exon junctions (e.g., STAR, HISAT2).
3.1	Pseudoalignment	Assigns reads to transcripts without full alignment (e.g., used by Kallisto, Salmon).
3.1	Q30	Quality score indicating a 1 in 1000 error rate in sequencing (99.9% accuracy).
3.55	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.
3.5	Alpha Diversity	Measure of species diversity within a single sample, considering species richness (number of species) and evenness (distribution of species).
3.5	Beta Diversity	Measure of species diversity between samples
3.66	Pseudo-bulk	Data aggregation technique where cell-level data are grouped to create virtual bulk samples for statistical analysis.
3.7	Peak calling	Identifying regions in the genome where sequencing reads are highly concentrated, indicating active or accessible DNA sites.
3.7	Blacklist region	Genomic regions known to produce unreliable or artifact signals in sequencing experiments, typically excluded from analysis to avoid misinterpretation of data.

3.10	Spot deconvolution	Estimating the proportions of different cell types within a single spatial transcriptomics data spot, as spots often contain multiple cells.

3.3 Cited databases

Section	Database name	Database description	URL
3.1	KEGG	Linking genes and proteins to metabolic and disease-related functions	https://www.kegg.jp/
3.1	Reactome	Detailed gene-protein relationships and cross-pathway interactions	https://reactome.org/
3.1	Biocarta	Molecular mechanisms underlying well-known biological pathways	https://maayanlab.cloud/Harmonizome/dataset/Biocarta+Pathways
3.1	Hallmark	Gene sets of high-level biological states or processes	https://www.gsea-msigdb.org/gsea/msigdb/
3.1	IPA	Manually curated pathways to predict causal relationships and identify regulatory networks	Provided as software
3.2	miRbase	Repository for miRNA sequences and annotations	https://www.mirbase.org/
3.2	DIANA-microT-CDS	Predicts miRNA targets in CDS and 3' UTRs.	https://dianalab.e-ce.uth.gr/microt_webserver/
3.2	MicroCosm	miRNA target identification and annotation.	https://tools4mirs.org/software/mirna_databases/microcosm-targets/
3.2	miRanda	miRNA target prediction using sequence analysis.	https://bioweb.pasteur.fr/packages/pack@miRanda@3.3a
3.2	miRDB	Predicts miRNA targets using machine learning.	http://www.mirdb.org/
3.2	PicTar	Conserved miRNA target prediction in animals.	https://pictar.mdc-berlin.de/
3.2	TargetScan	miRNA target prediction using conservation.	https://www.targetscan.org/
3.2	miRecords	Curated database of miRNA-target interactions.	http://c1.accurascience.com/miRecords/
3.2	miRTarBase	Experimentally validated miRNA targets.	https://mirtarbase.cuhk.edu.cn/ https://mirtarbase.cuhk.edu.cn/~miRTarBase/miRTa

			rBase_2019/php/index.php
3.2	TarBase	Manually curated miRNA-target interactions.	https://dianalab.e-ce.uth.gr/tarbasev9
3.3	Clinvar	Reports of human variants classified for diseases and drug responses.	https://www.ncbi.nlm.nih.gov/clinvar/
3.3	OMIM	Catalog of human genes and genetic disorders.	https://www.omim.org/
3.3	ACMG	Genetic variant classification guidelines.	https://www.acmg.net/
3.3	COSMIC	Catalog of somatic mutations in cancer.	https://cancer.sanger.ac.uk/cosmic
3.3	Civic	Clinical evidence for cancer variants.	https://civicdb.org/
3.3	OncoKB	Precision oncology knowledge base.	https://www.oncokb.org/
3.3	AMP	Molecular testing guidelines for cancer.	https://www.amp.org/
3.5	SILVA	rRNA sequences for taxonomy and phylogeny	https://www.arb-silva.de/
3.5	UNITE	Fungal ITS sequences for taxonomy	https://unite.ut.ee/

3.4 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
ImageJ	Image processing and quantification for microscopy data.	Spatial transcriptomics
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

3.5 Maximum project dimensions

Category	Service code	Service name	Max number of samples	Max number of comparisons
RNA	NF62.01.01	Bulk RNA-Seq analysis	no limit	10
	NF62.01.02	miRNA analysis	no limit	10
	NF62.01.03	lncRNA analysis	50	10
DNA	NF62.02.01	WGS analysis - population and medical studies	3000	10
	NF62.02.01	WGS analysis - rare diseases and cancer	200	10
	NF62.02.02	WES analysis	300	10
	NF62.02.03	Metagenomics/Microbiome Analysis	100	-
	NF62.02.04	Methyl-Seq analysis		
Single-cell / spatial	NF62.03.01	scRNA-Seq analysis	32	6
	NF62.03.02	scATAC-Seq analysis	32	6
	NF62.03.03	Single-cell immune profiling (VDJ)	32	6
	NF62.03.04	Single-cell multiome (ATAC + gene expression)	32	6
	NF62.03.05	Spatial transcriptomics	24*	3
Long reads	NF62.04.01	Long-read RNA sequencing		
* No more than 1 samples per slide				