HUMAN TECHNOPOLE
NATIONAL FACILITY FOR
DATA HANDLING AND ANALYSIS
CALL FOR ACCESS
26-DHA-ROUND-3

Table of Contents

READ BEFORE APPLYNG TO THIS CALL **3**
1. SUMMARY OF CHANGES TO THIS CALL FOR ACCESS **3**
2. INTRODUCTION **3**
2.1 Access modalities 4
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 12
7. APPLICATION SUBMISSION METHODS, CALL DEADLINE AND EVALUATION PERIODS **12**
8. EVALUATION OF APPLICATIONS **12**
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 16
10. PROJECT CONTINUATION APPLICATIONS **17**
11. AFTER ACCESS HAS BEEN APPROVED **17**
12. AFTER ACCESS HAS BEEN COMPLETED **18**
12.1 Reporting and feedback 18
12.2 Ownership of results and research outputs 18
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**

READ BEFORE APPLYNG TO THIS CALL

The **call for Access to the National Facility (NF) for Data Handling and Analysis (DHA)** is intended exclusively for the **analysis of data already generated – at the NF for Genomics or elsewhere - by the Applicant** submitting the proposal. To know more about DHA services, read Annex III of this call and consult the "Available services" webpage.

Conversely, DHA services can be combined with other services within a single submission applying to the specific call for Access to the NF that will generate your data. Below is an example:

- Access to Genomics services (link) cannot be requested through the call for Access to the NF for DHA: Applicant shall apply to the dedicated call for Access to the NF for Genomics (open calls), using the option provided to request data analysis as additional service, if necessary.
- The same applies to the services available through the other calls for Access (link).

Applications requesting data generation submitted to call for Access to the NF for Data Handling and Analysis will be considered not admissible and will be rejected at the administrative review stage.

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-DHA-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.

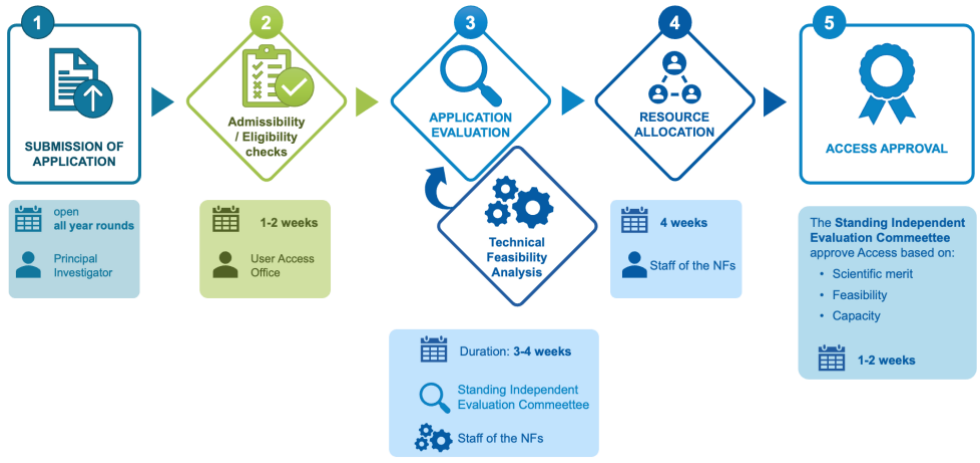
8. EVALUATION OF APPLICATIONS

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 three (3) 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-DHA-ROUND-3 priority will be given to researchers who do not have any ongoing Access to the NF for DHA and/ or who have not been granted Access in the calls 26-DHA-ROUND-1 and 26-DHA-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 (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

DATA HANDLING AND ANALYSIS

CALL FOR ACCESS

26-DHA-ROUND-3

Table of Contents

1. INTRODUCTION 26
2. SERVICE LIST 27
2.1 IU1 service list 27
NF61.01.01 Light Microscopy Analysis 27
NF61.02.01 Cryo-Electron Microscopy Analysis 29
NF61.03.01 Volumetric EM Analysis 30
2.2 IU2 service list 32
NF62.01.01 Bulk RNA-seq analysis 32
NF62.01.02 miRNA analysis 35
NF62.01.03 lncRNA analysis 38
NF62.02.01 WGS analysis 40
NF62.02.02 WES analysis 42
NF62.02.03 Metagenomics/ Microbiome Analysis 45
NF62.02.04 Methyl-Seq analysis 48
NF62.03.01 scRNA-seq analysis 50
NF62.03.02 scATAC-seq analysis 53
NF62.03.03 Single-cell Immune profiling-V(D)J 55
NF62.03.04 Single-cell multiome (ATAC + gene expression) 57
NF62.03.05 Spatial transcriptomics (10X Visium / SteroSeq platforms) 59
NF62.04.01 Long-read RNA sequencing 61
2.3 IU3 service list 65
NF63.01.01 – Web Application and Web Service Development 65
NF63.02.01 – Software Containerization 67
3. APPENDIX 68
3.1 Metadata file example 68
3.2 Glossary of terms 69
3.3 Cited databases 70
3.4 Tools used 72
3.5 Maximum project dimensions 76

1. INTRODUCTION

The mission of the National Facility for Data Handling and Analysis (NF-DATA) at Human Technopole (HT) is to support the national research community by providing state-of-the-art analysis of biological datasets generated by high-throughput genomic and imaging technologies. The main objective of this National Facility is to provide bioinformatics and bioimage analysis expertise for the interpretation of complex, large-scale biomedical datasets.

The National Facility for Data Handling and Analysis includes three Infrastructural Units:

Bioimage Analysis (IU1) - This unit provides high-quality image analysis solutions, including Quality Control (QC), denoising and image restoration, segmentation, and basic quantification for imaging data. It also develops and maintains new open-source image analysis tools that are released to the broader community and included in the service portfolio.

Omics Analysis (IU2) - this unit performs the analysis of “omics” data generated by other Facilities or by external sources. Basic services include QC, alignment to reference genomes, basic quantification and/or variant calling. In-depth analysis can be provided on a project-specific basis.

Technology Development, DevOps and Web development (IU3) - this unit focuses on providing installable and containerised versions of analysis tools and pipelines, as well as creating and maintaining User-friendly WebApps providing services to the scientific community.

All three units of the National Facility for Data Handling and Analysis engage in technology development, creating re-usable research software for use within the Facility, and enabling the deployment of novel resources that Users will be able to “take home” to their own institutions. Technology development ensures that the Facility remains state-of-the-art and adapts organically to the needs of the research community.

Another main pillar of the Facility is the provision of training for the Users through workshops and in-depth courses, allowing Users to take the acquired knowledge back to their home institutions and creating national awareness of our service portfolio.

Finally, the National Facility for Data Handling and Analysis supports Users in the management of the data produced by the National Facilities, assisting them with identifying proper storage for the data and the most effective data transfer methods.

The National Facility for Data Handling and Analysis is supported by a large Data Centre and Scientific Computing infrastructure initially composed of an HPC system (over 100 compute nodes, 30 GPU nodes, 20 PB of storage space) combined with access to cloud-based resources.

NF-DATA services can be accessed under two modalities:

- Access to facility service: the facility will perform the analysis autonomously and will deliver data, results, and a full analysis report to the Users. This is the standard modality, available for all services.

- Access to facility service including training: the facility will host up to two members of the User's group, and analysis will be performed collaboratively, allowing hosted scientists will learn how to perform it. This modality is only available for a subset of services.

2. SERVICE LIST

2.1 IU1 service list

IU1 offers the following services:

Category	Service code	Service name
Light Microscopy	NF61.01.01	Light microscopy analysis
Cryo-Electron Microscopy	NF61.02.01	Cryo-electron microscopy analysis
Volume-Electron Microscopy	NF61.03.01	Volume electron microscopy analysis

A detailed description of each IU1 service is found in the remainder of this section.

NF61.01.01 Light Microscopy Analysis

Service description

Light microscopy analysis encompasses the analysis of data generated by any light microscopy modality (ie, brightfield, phase contrast, widefield epi-fluorescence, confocal, lightsheet, etc) and across any sample type.

The services we provide include, but are not necessarily limited to, the following use-cases:

- **Image restoration and denoising**: Removal of pixel-independent noise from images to increase signal-to-noise ratio (SNR).
- **Semantic and Instance segmentation**: Identification and segmentation of objects in an image, generation of image masks.
- **Quantitative Image Analysis**: Quantification of intensity levels in images or segmented objects.
- **Morphometric Analysis**: Analysis of shape and morphology of segmented objects.
- **Custom pipeline developmen**t: Construction of an analysis pipeline combining two or more individual steps.

While these are examples of the services we can provide, we anticipate that most projects will require some combination of tools and services and so we will work with Users to craft pipelines that fulfil their analysis needs, as well as provide training and

support in their future use. Our ethos is to work openly and transparently with our Users in the spirit of scientific collaboration. During the application phase, it will only be necessary to describe the data and the desired form of the analysis result; the precise details of the analysis will be discussed with the Users upon selection of the project.

Access modality available

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

Requested inputs from Users

For this service, we require a detailed project description outlining the analysis goals, and the data to be analyzed. A full analysis plan will be developed in collaboration with successful applicants as the project proceeds. A detailed list of required information (ie number of images, resolution) can be found in the application. Upon successful application, Users will be requested to transfer all raw data to Human Technopole for analysis.

Technical requirements

Applicants must ensure that the dataset is available at sufficient quantity and quality (for example resolution and signal-to-noise ratio) before the closing date of the application period. This will be assessed on example data submitted during the application phase. Applicants are responsible for uploading their image data to Human Technopole file servers at the initiation of the project, and for downloading the final results at the conclusion of the project.

Results

Upon successful completion of the selected project, results will be delivered in a format of the Users' choosing and depending on the project needs. In addition, we will provide whatever software, code, and support is required for the User to reproduce the analysis at their home institute. The form will depend on the specifics of the project and the needs of the Users, but we anticipate delivery in the form of Python scripts and/or ImageJ macros. To reduce the burden of Access for our Users, we will use open-source software tools during the NF projects.

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 Light Imaging:

NF50 – All services

NF55.006 - Ion imaging assisted experiment

NF52.07.07 - Zeiss Axioscan Z.1 automated slide scanner

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

NF61.02.01 Cryo-Electron Microscopy Analysis

Service description

Cryo-Electron microscopy analysis encompasses the analysis of cryo-electron microscopy data, both single particle and tomographic reconstruction. This service includes, but is not necessarily limited to, the following use-cases:

- **Single-particle analysis (SPA)**: Development of image processing pipelines for the reconstruction of single particle 3D density maps, starting from cryoEM raw datasets or pre-processed micrographs/particles. Map validation.
- **Atomic Model Building**: De novo model building from reconstructed 3D density maps, fitting of existing atomic structures and refining of atomic models. Model validation.
- **Analysis of Flexibility and Heterogeneity**: Development of image processing pipelines for local reconstruction and refinement of flexible regions and evaluation of the conformational heterogeneity landscape of the macromolecules.
- **Tomography reconstruction**: Development of image processing pipelines for the reconstruction and analysis of tomograms, starting from tilt-series containing fiducial markers or fiducial less. Segmentation of the tomograms and sub-tomogram averaging (STA).
- **Custom pipeline development**: Construction of a computational pipeline combining two or more individual steps.

While these are examples of the services we can provide, we anticipate that most projects will require some combination of tools and services and so we will work with successful Applicants to craft pipelines that fulfil their analysis needs, as well as provide training and support in their future use. Our ethos is to work openly and transparently with our Users in the spirit of scientific collaboration. During the application phase, it will only be necessary to describe the data and the desired form of the analysis result; the precise details of the analysis will be discussed with the Applicants upon selection of the project.

Access modality available

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

Requested inputs from Users

For feasibility assessment, uploading a set of at least 10 movies/micrographs as part of the application process is required. It is important that these images accurately reflect the diversity of the data in the dataset (i.e., not the best set of possible images). Please ensure to include any applicable metadata. If CTF estimation is available, please include this as part of the application. Upon successful application, Users will be requested to transfer all raw data to Human Technopole for analysis.

Technical requirements

Applicants must ensure that the dataset is available at sufficient quantity and quality (for example resolution, contrast and signal-to-noise ratio) before the closing date of the application period. This will be assessed on example data submitted during the

application phase. Applicants are responsible for uploading their image data to Human Technopole file servers at the initiation of the project, and for downloading the final results at the conclusion of the project.

Results

Upon successful completion of the selected project, results will be delivered in a format of the Users' choosing and depending on the project needs (typically .mrc or .pdb, but other formats or intermediate files may be delivered depending on User preferences). In addition, we will provide whatever software, code, and support is required for the User to reproduce the analysis at their home institute.

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 Structural Biology:

SB-IU1-C – High-resolution Cryo-TEM Imaging

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

NF61.03.01 Volumetric EM Analysis

Service description

The Volumetric EM service provides segmentation and analysis of structures in serial EM section data. This service is restricted to segmentation and analysis of a finite number of structures per image (ie, mitochondria, organelles, vesicles, other similar objects of study). The proposal must include specific examples of the structure(s) of interest in order to judge feasibility. This service includes, but is not necessarily limited to, the following use-cases:

- **Data pre-processing**: de-streaking, alignment, contrast adjustment.
- **Data annotation**: generating dense labels covering structures of interest for the purpose of training AI algorithms.
- **Model Development**: Training and deployment of AI segmentation algorithms specific to the research question.
- **Downstream Analysis**: Analysis of segmented structures, morphology, number, distribution.

While these are examples of the services we can provide, we anticipate that most projects will require some combination of tools and services and so we will work with successful Applicants to craft pipelines that fulfil their analysis needs, as well as provide training and support in their future use. Our ethos is to work openly and transparently with our Users in the spirit of scientific collaboration. During the application phase, it will only be necessary to describe the data and the desired form of the analysis result; the precise details of the analysis will be discussed with the Applicants upon selection of the project.

Access modality available

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

Requested inputs from Users

For this service, we require a detailed project description outlining the analysis goals, and the data to be analyzed. A full analysis plan will be developed in collaboration with successful applicants as the project proceeds. A detailed list of required information (ie number of images, image dimensions, pixel size) can be found in the application. Upon successful application, Users will be requested to transfer all raw data to Human Technopole for analysis.

Technical requirements

Applicants must ensure that the dataset is available at sufficient quantity and quality (for example resolution and signal-to-noise ratio) before the closing date of the application period. This will be assessed on example data submitted during the application phase. Applicants are responsible for uploading their image data to Human Technopole file servers at the initiation of the project, and for downloading the final results at the conclusion of the project.

Results

Upon successful completion of the selected project, results will be delivered in a format of the Users' choosing and depending on the project needs. In addition, we will provide whatever software, code, models, and support is required for the User to reproduce the analysis at their home institute. The form will depend on the specifics of the project and the needs of the Users. To reduce the burden of access for our Users, we will use open-source software tools during the NF projects.

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 Structural Biology:

SB-IU1-D – Volume Electron Microscopy

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

2.2 IU2 service list

IU2 offers the following services:

Category	Service code	Service name
RNA	NF62.01.01	Bulk RNA-Seq analysis
	NF62.01.02	miRNA analysis
	NF62.01.03	lncRNA analysis
DNA	NF62.02.01	WGS analysis
	NF62.02.02	WES analysis
	NF62.02.03	Microbiome analysis
	NF62.02.04	Methyl-Seq analysis
Single-cell / spatial	NF62.03.01	scRNA-Seq analysis
	NF62.03.02	scATAC-Seq analysis
	NF62.03.03	Single-cell immune profiling (VDJ)
	NF62.03.04	Single-cell multiome (ATAC + gene expression)
	NF62.03.05	Spatial transcriptomics (10X Visium / StereoSeq platforms)
Long reads	NF62.04.01	Long-read RNA sequencing

A detailed description of each IU2 service is found in the remainder of this section.

Please see Table 3.5 in APPENDIX 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.

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 - Sequencing only with NovaSeq 6000 (Illumina)

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

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

NF62.01.02 miRNA analysis

Service description

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

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

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

1. **Quality check of the raw sequence data**: Sequencing quality of the raw reads is evaluated. It assesses the data quality distribution across reads, per-base content, and adapter contamination.
2. **UMI extraction and trimming**: Low-quality reads, UMI sequences and adapter contamination will be removed and excluded from the analysis. QC is performed again on the trimmed reads.
3. **Filtering for miRNA**: Filtering reads according to length and assessing their nature with respect to other types of small RNAs.
4. **Mapping**: Trimmed reads will be mapped against the reference genome, and mature miRNAs and precursors (hairpins) will be obtained from miRBase.
5. **Expression quantification**: Uniquely mapped reads are assigned to the corresponding features (mature miRNAs and miRNA precursors (hairpins)). A counts matrix is produced that summarizes the inferred expression level for each known miRNA in each sample.
6. **Quality metrics collection**: Quality metrics from sequencing, trimming, alignment and quantification are collected and summarized in a complete, interactive report.
7. **Normalization and filtering**: Expression levels are normalized to account for the different library sizes across samples. Non-expressed miRNAs are filtered out.
8. **Exploratory analysis on expression data**: Principal Component Analysis (PCA) and Multi-Dimensional Scaling (MDS) are performed to inspect the variability structure of the data and its possible relationship with sample characteristics. If applicable and necessary, batch correction is performed using regression models. The expression of selected project-specific targets (if applicable, e.g. knocked-out genes, tissue markers, etc.) is evaluated across all samples.
9. **Differential expression analysis**: Expression levels are compared between different groups of samples using statistical models based on the experimental design (e.g. paired models, regression of covariates, etc.). Differentially expressed miRNAs are identified by setting cutoffs on the obtained p-values and log2FoldChanges.

Advanced (optional) analysis steps include the following:

1. **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 0. 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.
2. BAM files for each sample.
3. Raw and normalized count matrices containing expression values for each miRNA in each sample.
4. Tables of differentially expressed miRNAs with statistical significance information.
5. Complete reports (in interactive HTML and publication-ready PDF formats) describing the quality of the data, all the analysis performed on the dataset, and their results. Plots (e.g. heatmaps, volcano plots, PCA/MDS) and tables included to the report will also be provided as separate files.
6. Pipeline and scripts used to perform the analysis, if applicable.

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

Combined services

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

G-007 – Small RNA sequencing.

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.

NF62.01.03 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)
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.

NF62.02.01 WGS analysis

Service description

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

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

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

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

d. **Structural Variants** (SVs) are large-scale changes in the genome structure, such as inversions, translocations, duplications, or large insertions/deletions (more than 50 bases).

Advanced (optional) analysis steps include the following:

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

Access modality available

- Access to facility service

Requested inputs from Users

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

The User shall provide a table listing all biological and experimental conditions in the study and all samples belonging to each condition, ensuring that the sample names exactly match the names of the provided fastq files. See the example provided in Appendix 0.

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

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

Combined services

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

G-001 - Whole Genome Sequencing

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

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

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

NF62.02.02 WES analysis

Service description

Exome sequencing is an application of DNA sequencing (see NF62.02.01) that focuses on preferentially sequencing the exons, or protein-coding regions, which make up about 1-2% of the genome and are more likely to harbor disease-causing mutations. It is used to study genetic variations that affect protein function, thus particularly in disease research. The focus is on identifying genetic variants (e.g.,

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

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 0.

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

Technical requirements

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

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

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

Results

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

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

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

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

Combined services

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

G-002 - Whole Exome Sequencing

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

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

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

NF62.02.03 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-based 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:** Assemblying 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 geneomes.
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 3.1 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 0.

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-025 - Sequencing only with NextSeq 2000 (Illumina)

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

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

NF62.02.04 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 0.

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)

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%.

Results:

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

1. Raw and filtered fastq files for each sample and modality.
2. BAM files for each sample and modality.

3. Count matrices containing expression values for each gene in each cell (in h5ad format).
4. Raw peak-by-cell count matrices generated using a merged consensus peak set derived from sample-specific pseudobulk peak calling.
5. Complete reports (in interactive HTML and publication-ready PDF formats) describing the quality of the data, all the analysis performed on the dataset, and their results. Plots (e.g. heatmaps, volcano plots, PCA/MDS) and tables included to the report will also be provided as separate files.
6. Python objects (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-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 Visium / SteroSeq 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

Service 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.

2.3 IU3 service list

IU3 offers the following services:

Category	Service code	Service name
Scientific software development	NF63.01.01	Web application and web service development
Scientific software support	NF63.02.01	Software containerization

A detailed description of each IU3 service is found in the remainder of this section.

NF63.01.01 – Web Application and Web Service Development

Service description

This service allows for the creation of web applications and web services that are of interest to the scientific community. Web based applications often represent better solutions compared to desktop applications, the latter entailing manual software installation, software updates, operating systems compatibility, and strict system requirements. Many of these issues are tamed or completely solved by the adoption of client-server architectures where Users can access services hosted on a remote machine. This includes both full-fledged web applications, available through a web browser interface, as well as lower-level services such as APIs.

Example of common applications are web portals with interactive charts to inspect data stored on databases, or application wrappers around bioinformatic packages that enable scientists with limited computer programming skills to use them through easy-to-use interfaces.

The service includes three main areas, which can be combined to achieve the goals of the proposed application:

1. Frontend application development
2. Backend application development
3. Data layer design and implementation

Area 1 includes the design of the User side of the application, which includes the GUI (graphical User interface) and the optional data retrieval logic for getting data from an external resource, such as the application backend or a third-party API. The final output is a web page running within a web browser, which can involve a greatly varying amount of application logic, ranging from plain static web pages (e.g. promotional websites and portfolios) to complex applications that resemble standard desktop applications.

Our unit provides the expertise to build websites and web applications taking advantage of technologies such as Single Page Applications (SPA), Server Side Rendered Applications (SSR), and Static Sites Generators (SSG), taking care of the design of the User interface, responsive layouts to fit multiple screen sizes, SEO friendliness and accessibility. Main technologies and languages in this area include

dynamic HTML, CSS, JS as well as very popular frameworks such as Vue.js and Astro.js.

Area 2 is focused on everything running on the server side, called the backend. Backend services span a plethora of different use cases, fulfilling many different needs. These services provide means to store, distribute, and process data, execute code, authenticating Users (in combination with other actions), and more. Applicants can request the development of backend services taking advantage of state-of-the-art technologies, best practices and frameworks.

Area 3 is directly linked to area 2 and deals with how to model, store and retrieve data. This involves data storage strategies, as well as relational and non-relational databases. An important activity is relational database design, where databases are engineered to optimize queries, ensure data normalization and provide robust relational constraints.

Such services can be combined to build web applications that wrap already existing bioinformatic packages developed by the applicants, for which a graphical User interface and simplified usage are required.

The service also includes support for the delivery of the software, providing case by case solutions for correct SW deployment, code and assets share, and access, taking into account aspects such as licensing, authentication and authorization.

Requested inputs from Users

Application proposals will be evaluated according to innovativeness and scientific impact, measured in terms of potential audience, number of Users, and society benefits. Applicants are encouraged to conduct a survey on already existing solutions, either free or paid, that address similar issues. If similar resources already exist, the applicant should clarify what is the added value and innovation component of their proposal.

Users must provide a comprehensive description of the application, accurately describing the desired inputs, outputs and desirable controls. In case of User interfaces, Users are encouraged to provide any useful information to ease the GUI design phase such as User interface sketches, User-session workflow, etc.

Data, biological algorithms and specific code implementations / packages which need to be encapsulated in the application must be provided, in compliance with legal licensing assessment performed during the evaluation phase of the proposal. Every file and dataset uploaded by the Users will be treated according to standard GDPR legislation.

Users may also request support in deploying the application on an appropriate hosting infrastructure. Hosting on Human Technopole servers is not currently offered.

Access modality available

- Access to facility service

Technical requirements

For this service, there are no strict technical requirements to be fulfilled. On the other hand, the National Facility reserves the right to select the most appropriate technology for implementing the applications, using industry standard and well-established open-source technologies such as the ones listed below.

- Frontend: available frameworks are Vue.js, Nuxt.js, Astro.js. Styling is provided by custom developed CSS or prebuilt systems such as Material Design or utility libraries like Tailwind.
- Backend: available languages are python and JS / TS (Node.js). Available backend frameworks are FastAPI (Python), Express.js (Node.js), Fastify (Node.js).
- Data layer: file storage through S3 buckets / local storage. Relational Databases: MySQL, MariaDB, Postgres. Non-relational databases: MongoDB, neo4j.

Specific technological needs must be documented and justified in the application proposal.

Results

The Users will receive the requested software application and all related source code, assets, containerized images, and other artifacts. The application will be released under an appropriate open-source license. Developed code will be shared through dedicated remote repositories (such as GitHub or GitLab). Distributable artifacts, such as code bundles and docker images, will be shared in dedicated indexes and registries. Full documentation of code and data models will be provided as well.

Implemented services will be provided in containerized format through docker images, hosted on online accessible registries. The National Facility may also be able to assist Users in the deployment of the application on a dedicated infrastructure (either cloud-based or on-premises) owned or controlled by the applicants. Feasibility of this step will be discussed with the applicants during the technical evaluation phase.

Combined services

This service may be combined with all analysis services offered by the National Facility for Data Handling and Analysis.

NF63.02.01 – Software Containerization

IMPORTANT – this service is currently permanently deactivated.

3. APPENDIX

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

- The first three columns are required and should be named Condition, Sample, and FASTQ1 respectively.
- The fourth column can be omitted in the case of single-end sequencing. If present, it should be named FASTQ2.
- Condition names and sample names should only contain letters, digits, and the underscore character. Please do not include spaces, symbols, or special characters.
- Additional variables associated with each sample can be added to the table, and will be included in the final reports.

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

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

3.2 Glossary of terms

Section	Term	Definition
0	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.
0	Image restoration	Process of improving image quality by removing random noise, enhancing the signal-to-noise ratio (SNR).
0	Semantic and Instance Segmentation	Techniques to identify and separate distinct objects in an image, generating precise object boundaries (masks).
0	Morphometric Analysis	Measurement and analysis of shapes, sizes, and structures in biological images.
0	Single-Particle Analysis (SPA)	Computational technique for reconstructing 3D structures of macromolecules from 2D images obtained by cryo-electron microscopy.
0	Sub-tomogram Averaging (STA)	Method for enhancing resolution in 3D reconstructions of structures from cryo-electron tomography data by averaging multiple similar regions.
0	CTF	Contrast Transfer Function.
0	Splice-aware aligner	Tool that aligns RNA-seq reads, accounting for exon-exon junctions (e.g., STAR, HISAT2).
0	Pseudoalignment	Assigns reads to transcripts without full alignment (e.g., used by Kallisto, Salmon).
0	Q30	Quality score indicating a 1 in 1000 error rate in sequencing (99.9% accuracy).
0**5**	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.
0	Alpha Diversity	Measure of species diversity within a single sample, considering species richness (number of species) and evenness (distribution of species).
0	Beta Diversity	Measure of species diversity between samples
06	Pseudo-bulk	Data aggregation technique where cell-level data are grouped to create virtual bulk samples for statistical analysis.

0	Peak calling	Identifying regions in the genome where sequencing reads are highly concentrated, indicating active or accessible DNA sites.
0	Blacklist region	Genomic regions known to produce unreliable or artifact signals in sequencing experiments, typically excluded from analysis to avoid misinterpretation of data.
0	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
0	KEGG	Linking genes and proteins to metabolic and disease-related functions	https://www.kegg.jp/
0	Reactome	Detailed gene-protein relationships and cross-pathway interactions	https://reactome.org/
0	Biocarta	Molecular mechanisms underlying well-known biological pathways	https://maayanlab.cloud/Harmonizome/dataset/Biocarta+Pathways
0	Hallmark	Gene sets of high-level biological states or processes	https://www.gsea-msigdb.org/gsea/msigdb/
0	IPA	Manually curated pathways to predict causal relationships and identify regulatory networks	Provided as software
0	miRbase	Repository for miRNA sequences and annotations	https://www.mirbase.org/
0	DIANA-microT-CDS	Predicts miRNA targets in CDS and 3' UTRs.	https://dianalab.e-ce.uth.gr/microt_webserver/
0	MicroCosm	miRNA target identification and annotation.	https://tools4mirs.org/software/mirna_databases/microcosm-targets/
0	miRanda	miRNA target prediction using sequence analysis.	https://bioweb.pasteur.fr/packages/pack@miRanda@3.3a

0	miRDB	Predicts miRNA targets using machine learning.	http://www.mirdb.org/
0	PicTar	Conserved miRNA target prediction in animals.	https://pictar.mdc-berlin.de/
0	TargetScan	miRNA target prediction using conservation.	https://www.targetscan.org/
0	miRecords	Curated database of miRNA-target interactions.	http://c1.accurascience.com/miRecords/
0	miRTarBase	Experimentally validated miRNA targets.	https://mirtarbase.cuhk.edu.cn/ https://mirtarbase.cuhk.edu.cn/~miRTarBase/miRTarBase_2019/php/index.php
0	TarBase	Manually curated miRNA-target interactions.	https://dianalab.e-ce.uth.gr/tarbasev9
0	Clinvar	Reports of human variants classified for diseases and drug responses.	https://www.ncbi.nlm.nih.gov/clinvar/
0	OMIM	Catalog of human genes and genetic disorders.	https://www.omim.org/
0	ACMG	Genetic variant classification guidelines.	https://www.acmg.net/
0	COSMIC	Catalog of somatic mutations in cancer.	https://cancer.sanger.ac.uk/cosmic
0	Civic	Clinical evidence for cancer variants.	https://civicdb.org/
0	OncoKB	Precision oncology knowledge base.	https://www.oncokb.org/
0	AMP	Molecular testing guidelines for cancer.	https://www.amp.org/
0	SILVA	rRNA sequences for taxonomy and phylogeny	https://www.arb-silva.de/
0	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	NOT AVAILABLE FOR THIS EVALUATION ROUND	
	NF62.02.01	WGS analysis - rare diseases and cancer	200	10
	NF62.02.01	WGS analysis – plant studies	20	10
	NF62.02.01	WGS analysis – animal studies	100	10
	NF62.02.02	WES analysis	400	10
	NF62.02.03	Microbiome/Metagenomics analysis	200	-
	NF62.02.04	Methyl-Seq analysis	200	
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)	NOT AVAILABLE FOR THIS EVALUATION ROUND	
	NF62.03.05	Spatial transcriptomics	NOT AVAILABLE FOR THIS EVALUATION ROUND	
Long reads	NF62.04.01	Long-read RNA sequencing	20	5