

newsletter

Technologies, Applications, and Access to Support

Next Generation Sequencing at the Functional Genomics Center Zurich

OVERVIEW

1
NGS AT THE FGCZ
Technologies and organization

2
NGS APPLICATIONS AND TECHNOLOGIES
Applications in sequencing

3-4
BIOINFORMATICS OF NGS
Analysis workflows and support

Next Generation Sequencing (NGS) has become one of the major technologies at the FGCZ. The impressive throughput and read lengths of the current high-end systems enable research groups to *de novo* sequence genomes of small to complex organisms. They allow for the analysis of genome alterations, gene expression and DNA modifications, as well as a constantly growing number of other applications.

To account for the vast diversity of life science research at the ETH Zurich and the University of Zurich, the FGCZ has continuously extended its portfolio of NGS technologies and applications. The emphasis of this expansion has not only been on capacities but also on capabilities, leading to the establishment of a comprehensive range of NGS technologies. Today, short read technologies (Illumina HiSeq) for the generation of very large numbers of short to medium length reads, medium length read technologies (Illumina MiSeq) for the generation of medium to large numbers of longer reads, plus technologies focusing on the very rapid generation of sequencing data (Ion Torrent PGM and IonProton) and very long reads and novel applications at the single molecule level (Pacific Biosciences RS).

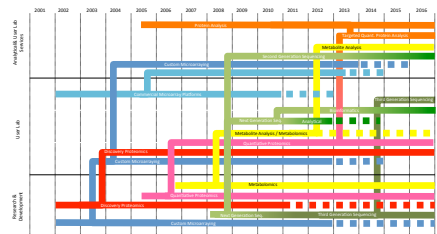


The on site availability of these systems allows us to flexibly combine the different technologies within or between projects and opens up combinations of possibilities that ensure suitable analytical strategies tailored to the research project's needs. As a consequence of this flexibility and the number of technologies available,

NGS projects need careful planning and consideration at multiple levels, including sample availability and quality, robustness of methods and protocols, processing and analysis of data, as well as financial and time constraints.

This complexity emphasizes the need for a close collaboration of ETH and UZH research groups with analytical and bioinformatics experts of the FGCZ, which is why access to the FGCZ NGS platform is provided through the User Lab. In parallel to the technology expansion, significant efforts have been undertaken in increasing the User Lab staff working in the technologies and bioinformatics sections. As a result, optimized analytical protocols and data analysis workflows have been established that lead to significantly shortened turnover times from sample generation to data interpretation.

Technologies and support modes at the FGCZ



In the following, we briefly describe the available technologies (from most recent to most established) including supported applications. More information on the setup of NGS and access via the FGCZ User Lab and User Lab Services can be found on the FGCZ website at www.fgcz.ch. For further or more specific questions, please contact us at sequencing@fgcz.ethz.ch

NGS@FGCZ Applications and Technologies

De novo sequencing of genomes and metagenomes

Emphasis on longer reads, medium to high throughput
 *** Pacific Biosciences RS
 ** Illumina HiSeq (** in combination with PacBio)
 ** Illumina MiSeq, Ion Torrent PGM (for smaller genomes)

De novo sequencing of transcriptomes

Emphasis on medium to longer reads, high to very high throughput
 *** Illumina HiSeq / MiSeq
 ** PacBio RS (** in combination with Illumina)
 ** Illumina MiSeq, Ion Torrent PGM

Genome-wide SNP discovery and variant detection

Emphasis on high to ultra-high throughput, short to medium read lengths
 *** Illumina HiSeq (max. throughput)
 *** Illumina MiSeq (max. flexibility)
 ** Ion Torrent PGM (max. speed)

Transcriptome analysis

Emphasis on high to ultra-high throughput, short to medium read lengths
 *** Illumina HiSeq (max. throughput)
 ** Illumina MiSeq & Ion Torrent PGM (max. speed)

Small RNAs

Emphasis on high to ultra-high throughput, short read lengths
 *** Illumina HiSeq & MiSeq
 ** Ion Torrent PGM (max. speed)

ChIP-seq

Emphasis on high to ultra-high throughput, short to medium read lengths
 *** Illumina HiSeq (max. throughput)
 ** Illumina MiSeq & Ion Torrent PGM (max. speed)

Amplicon sequencing

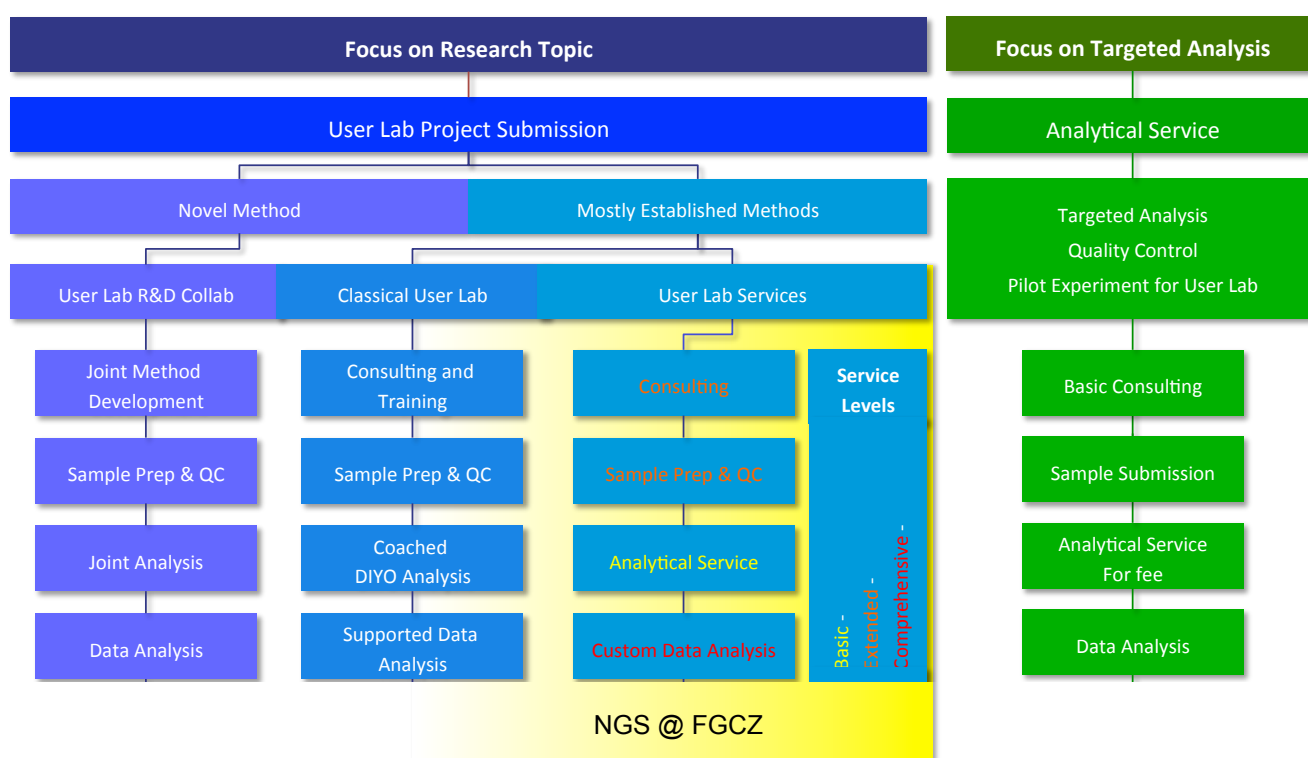
Emphasis on longer reads, medium to high throughput
 ** Illumina HiSeq & MiSeq
 ** Ion Torrent PGM

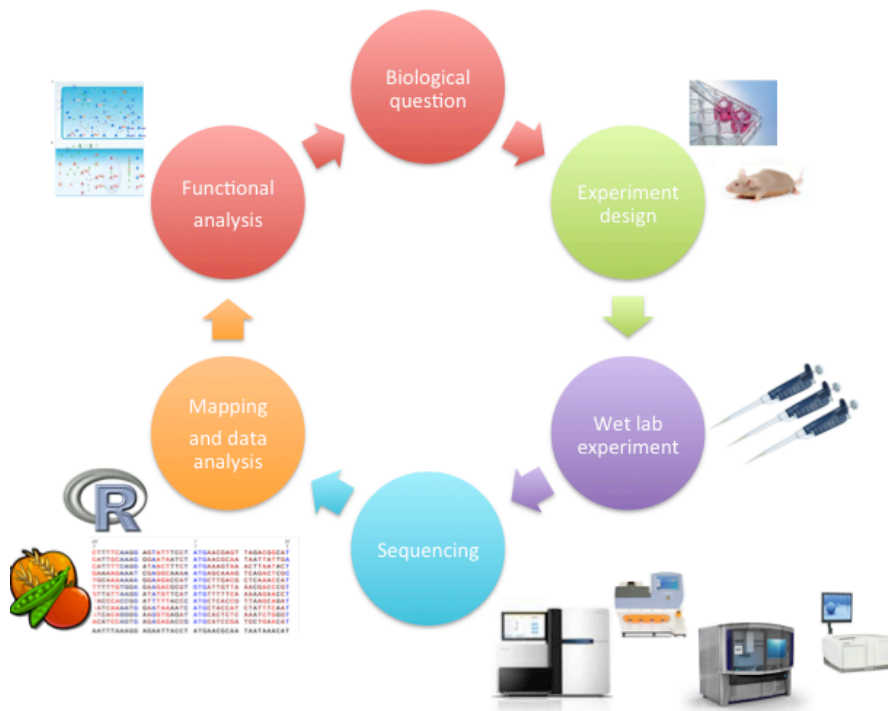
DNA Methylation analysis

Emphasis on flexibility and protocols
 *** Pacific Biosciences RS
 *** Illumina HiSeq & MiSeq

The list of applications and suitable technologies is not exhaustive nor does it exclude the use of a specific analytical technology for applications mentioned. The purpose of the list is to provide an initial overview that is always refined based on individual needs during the setup phase of an FGCZ project. Technologies of equal suitability are listed in alphabetical order.

NGS@FGCZ Project-based Services and Support





DATA ANALYSIS WORKFLOWS IN SHORT

De novo assembly
Mapping to the reference genome
SNP and INDEL detection
Digital expression
ChIP-seq region detection

Turning Sequencing Reads into Knowledge

NGS Bioinformatics

NGS technology and the resulting data are uniquely suited for answering a wide range of biological questions by overcoming limitations that so far existed using classical sequencing or alternative microarray approaches. Due to the massive amount of data involved, the management and analysis of the data requires dedicated software and high-performance and capacity computing resources. The FGCZ has developed and productively implemented data processing pipelines for standard data analysis workflows, which can be readily applied to analyze the sequencing data generated at the center. In addition to standard tools and workflows, the bioinformatics team provides access to its expertise and the optional development of customized solutions via the FGCZ User Lab. As for the experimental analysis part, the analysis and interpretation of the resulting data relies on the close interaction of FGCZ staff with the users that is essential to generate scientifically relevant results.

Standard Data Analysis Support

As an integral part of every sequencing project, the FGCZ provides experimental design consultation during the setup phase of a project in conjunction with the discussion for selecting the most suitable analytical platform. The actual data analysis support following data production in a given project then significantly depends on the type of the study:

For projects with a reference genome, the standard support includes:

- files of raw reads
- read alignment to the reference(s)
- initial secondary analysis (e.g. SNP calling, expression quantitation, peak finding, and so on)
- QC Report

For *de novo* sequencing projects, the standard support includes:

- files of raw reads
- assembly of the raw reads with vendor assembler using default parameters

The currently available standard data analysis workflows cover popular NGS applications, such as:

- Re-sequencing, SNP discovery and variant detection
- Transcriptome analysis and expression quantitation
- Regulome analysis (small RNA, transcription factor binding, histone modification, methylation)

Standard data analysis and support efforts in the form of a limited number of consulting hours are part of the User Lab Services and are free of charge. Analysis pipelines and consulting for additional applications can be added, based on demand and available resources.

Customized Data Analysis Service

A significant number of research projects using the large flexibility and many options of NGS show a distinct need for customized support or even the development of new data analysis procedures or tools. Depending on the availability of resources and subject to project-specific agreements, the FGCZ collaborates on non-standard data analysis, like for example, the discovery and *in silico* verification of new miRNAs.

Training and Support

In addition to project-specific data analysis, the FGCZ provides training and education at the conceptual and concrete software usage levels. Courses and tutorials are announced on the FGCZ [Genomics Bioinformatics website](#).



SOFTWARE TOOLS AND ACCESS TO RESOURCES

Software available:

NGS reads mapping and assembly

Tag counting, quantification, profile pattern discovery

Sequence analysis and annotation

Databases available:

Primary sequence database

Protein sequence database

Genome database

Specialized Hardware available

Access to Computing and Software

Resources for Users

Access to and organization of data

All NGS data is automatically stored in the FGCZ B-Fabric system and accessible via web interface. Sample meta-information and project information are available through the same interface to all members of the respective User Lab Services project.

Access to Computing Resources

While the high performance computing infrastructure for processing and basic analysis of NGS data is used exclusively

within the established pipelines and by FGCZ bioinformatics staff, the center provides users access to a dedicated high-performance linux computer at the FGCZ for NGS data analysis. This way, users have access to a wide range of NGS data analysis software tools and databases. Computationally-intensive data analysis tasks can be scheduled by the FGCZ team on the FGCZ computing cluster.

Bioinformatics Tools and Databases

More than one hundred bioinformatics software packages and many standard life science databases are implemented and maintained at the FGCZ.

FURTHER INFORMATION

SEQUENCING CONTACT

Via eMail to sequencing@fgcz.ethz.ch or sequencing@fgcz.uzh.ch. FGCZ staff will initiate a meeting to discuss options and workflows, issue quotes, and advice on all aspects from study design to analysis to data interpretation.

BIOINFORMATICS SUPPORT

Bioinformatics consulting and support is an integral part of User Lab projects and User Lab Services. As resources are limited, discussion about the level of support and training offered to users is part of the project setup process.

USER LAB AND SERVICES

More information on the general setup of the FGCZ and its access modes can be found at www.fgcz.ch

Sources: Pictures of instruments, consumables and workflows have been retrieved from vendor sources. Text is in part based on information materials of the vendors.

Disclaimer: The technologies mentioned may be well used for applications mentioned or non-mentioned also in the absence of an FGCZ recommendation: due to the large number of protocols and applications possible, the FGCZ can only support a limited number of protocols and methods per platform and therefore limits the number of standard recommendations. Before excluding options, please consult with us.



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich



University of
Zurich^{UZH}