Research Core Unit Genomics (RCUG)

Hannover Medical School, Carl-Neuberg Str.1, 30625 Hannover,
http://www.mh-hannover.de/genomics.html

Introduction and overall organization

The RCUG is a research-driven service unit. Since its formation in January 2017, the RCUG is open to all departments at the MHH and selected external institutions for the realization of sequencing- and microarray-based projects. The extensive expertise of the personnel (six permanent employees) in the fields of planning, technical execution, analysis and publication of genome- and transcriptome-wide studies has been acquired during many years of research in their respective fields. To date, (April 2022) more than 160 groups from more than 55 different MHH departments have employed the services of the RCUG or its predecessors. The offered services are dedicated to all methodical aspects of data generation, raw data processing and quality control, supplemented by comprehensive overall support (project planning, analysis strategy, analysis workshops). Co-authorships are not a mandatory prerequisite for collaboration unless extensive methodological or analytical modifications are required or otherwise agreed.

Organizational concept

The head of the RCUG (Dr. Oliver Dittrich-Breiholz) reports directly to the President of the MHH and is in turn the immediate superior of the permanent RCUG staff. He is responsible for the cost center and budget compliance, makes all essential decisions regarding the 'operative core business' and is the primary contact and responsible person for project inquiries.
The medium to long-term strategic direction of the RCUG is discussed and further developed in a steering group in close coordination with the President. The steering group meets two to four times a year and focuses on the following content:

- Strategic discussions on prioritization and medium-term orientation of the RCUG

- Increased synergies of the RCUG to large and joint projects
- Organization of required resources (personnel, infrastructural, financial)
- Coordination when using 'decentralized sequencers' and organization when integrating 'associated employees'
- Coordination and integration with bioinformatics and IT infrastructure at the MHH

The RCUG steering group is currently composed as follows:

- Prof. Dr. Meike Stiesch
- Prof. Dr. Helge Frieling
- Prof. Dr. Burkhard Tümmler
- Dr. Lutz Wiehlmann
- Dr. Oliver Dittrich-Breiholz

The main task of the RCUG is to implement projects for research groups in the context of technology provision and development, rather than running its own research projects.
However, the RCUG also directly contributes to third-party funded large projects, such as the Lower Saxony 'Big Data Initiative' or the Cluster of Excellence 'RESIST'.

Facilities

The RCUG is equipped with state of the art Illumina (**NextSeq 550**) sequencing technology. A **Sequel** machine for long-read analysis (Pacific Biosciences), one additional **NextSeq 500**, and three **MiSeqs** constitute so called 'decentralized sequencers'. These sequencers are operated by other groups at the MHH but are also available for use by the RCUG. A close cooperation in-house furthermore enables access to a production scale Illumina sequencer (**NovaSeq6000**). A **Chromium™ Single Cell Controller** (10x Genomics) located at the central sorter facility of the MHH is available for single cell RNA-seq experiments. Two decentralized **pipetting robots** are also on premises for large projects. Two **2100 Bioanalyzers** (Agilent Technologies), two **Qubit fluorometers** (Thermo), a **BluePippin system** (Sage science) and a **Nanodrop** spectrophotometer (Thermo) are present for quantification and fragment length analysis of generated cDNA libraries for sequencing applications. The RCUG is additionally equipped with an **Agilent microarray scanner** G2565C which allows automated scanning of up to 48 slides per run.

Personnel

The core personnel of the RCUG consist of two scientists who are in charge of overall coordination, experimental design, quality control and technical supervision. Two bioinformaticians are responsible for data management processes, data analysis and further development of data analysis strategies. Two technicians execute all wet chemistry work and maintain our laboratories.

Genomics services

- ***DNA QC:*** Measurement of DNA concentration, purity and integrity (Qubit and Bioanalyzer)
- ***Eukaryotic WGS:*** External sequencing to access most competitive pricing
- ***Prokaryotic WGS:*** Two platforms: Illumina (high throughput, short reads) and PacBio (low throughput, long reads)
- ***DNA fragmentation:*** The Covaris platform allows for the defined, sequence independent fragmentation of DNA for various applications
- ***Exomes:*** Standard kits from Agilent, Illumina and IDT on Illumina platforms
- ***Target Enrichment:*** Standard (and custom) kits from Agilent, Illumina and NEB
- ***Amplicons:*** Sequencing of specific loci in large cohorts
- ***Metagenomics:*** Microbial metagenome analysis (bacteria, viruses, fungi) of DNA samples (not limited to a specific habitat, disease or host)
- ***ChIP-Seq:*** The Covaris system allows for milder enrichment conditions and less

background.

- ***Epigenomics:*** Targeted sequencing of bisulfite converted eukaryotic DNA at the MHH. Whole epigenomes are available as external sequencing service. The PacBio platform allows sequencing of prokaryotic epigenomes, while Oxford Nanopore technology* is currently being tested for eukaryotic methylomes.

  * respective data are generated with MinION devices, owned by commissioning PIs

Transcriptomics services

- ***RNA QC:*** Measurement of RNA concentration, purity and integrity (Nanodrop, Qubit and Bioanalyzer)
- ***mRNA-Microarrays:*** Microarray-based mRNA expression profiling
- ***RNA-Seq:*** RNA-Sequencing after ribosomal RNA depletion *(also for partially degraded RNA)*
- ***mRNA-Seq:*** RNA-Sequencing after mRNA enrichment
- ***smallRNA-Seq:*** RNA-Sequencing of small RNA species
- ***microRNA-Seq:*** RNA-Sequencing of mature microRNA fraction of small RNA
- ***single cell RNA-Seq:*** single cell mRNA-sequencing (10x Genomics workflow), including Cite-seq, Cell hashing, B and T cell repertoire monitoring

General Services

- Consultation: study design and data analysis
- Integration of data from public databases
- Support for integration of generated data into public databases
- Support for research proposals
- Provision of computing power
- Provision of MHH Galaxy instance

Custom bioinformatics services *

- Long read analyses for bacterial whole genomes, metagenomics, long amplicons
- Transcriptome alignment and quantification (RNA-Seq)
- Microarray analysis (mRNA expression)
- Mapping of transcriptional start sites (RNA-Seq)
- De novo Assembly (Genome, Transcriptome)
- Re-sequencing and SNP detection
- Microbial metagenomics
- Evaluation of epigenomic data
- Data visualisation in genome browsers
- Provision of commercial data analysis programs

  * some bioinformatics services are not standard techniques and can only be performed with the appropriate capacity and separate agreement

Teaching

- Bioinformatics and data analysis courses (Galaxy platform, SLURM & HPC)
- Wet lab protocols (transfer of knowledge to working groups)
- Practical training courses and electives for students of Biomedicine, Biochemistry, Biomedical Data Sciences

Focus: Setup of the HPC-seq cluster at the MHH

Research employing high-throughput sequencing methodologies to perform epigenetics, large scale RNA-seq, whole exome or whole genome analysis can easily generate terabytes of data per week. Such data cannot be managed and stored long term for multiple decades (as required by regulations) on local Windows PCs. The HPC-seq cluster provides significant compute power (currently 1400 cores in more than 13 nodes, Ubuntu Linux, SLURM job scheduler) to allow large workloads to be efficiently processed in a timely fashion, and furthermore provides the backbone for resource-intensive graphical applications such as Galaxy and SMRT Portal. In particular, Galaxy is a user-friendly platform which allows scientists to rapidly and easily submit analysis jobs to the performant cluster nodes via a web interface before subsequently analyzing results within a web browser on their local PC. A three tiered storage system has been implemented, with jobs transiently stored on a fast 24TB SSD shelf, with important results saved into mid-term redundant NetApp shelves. Long-term backup is realized by a large tape drive system with redundant storage of 10 TB tapes. An interdisciplinary team from the Center for Information Management and the RCUG take care of administration, development, maintenance, software installation, testing and cluster upgrades. The cluster generally processes over 5000 analysis jobs every month.
The current usage concept provides graduated usage privileges regarding overall access, provided processing power, and queue management. Individual usage rights are assigned according to relative contribution to the HPC-seq cluster. At present, the HPC-seq hardware components are predominantly financed by third party funds, whereas central infrastructure is contributed by the MHH ZIMt. Regular training courses will be run for users not only in cluster usage, but in data analysis with our local MHH installation of Galaxy.

Focus: Comprehensive support for data analyses and visualization

One of the main objectives of the RCUG is to advise clients and cooperation partners on the best way to extract meaningful information from generated raw data. Hence, a huge variety of open source bioinformatics software is hosted on the HPC-seq cluster, which has been designed and configured by RCUG scientists together with ZIMt and many other departments. Software is available via Bioconda and Gitlab and documented with examples on an open Wiki website. A number of Galaxy flavours are hosted for point and click execution of NGS workflows. Visualization is also available via the web-based JBrowse or commercial software (below) and further data analysis via the web-based RStudio environment. The RCUG also provides access to several user-friendly commercial software applications bookable by Microsoft Outlook calendars. We currently offer Qlucore Omics Explorer v3.7, Agilent GeneSpring v13.1.1, Qiagen Ingenuity Pathway Analysis (IPA) and Avadis StrandNGS v4.0. Regular practical training courses are provided. These courses cover RNA-seq, DNA-seq, mutation analysis and, in future, long read analysis.