Research Core Unit Genomics (RCUG)

Hannover Medical School, Carl-Neuberg Str.1, 30625 Hannover, 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 NGS- 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, (June 2018) 140 groups from 47 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.

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**, three **MiSeqs**, a number of **MinION**s and **Ion Torrent** family devices 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 with the "Helmholtz-Centre for Infection Research" (HZI) 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 will be available from 2019 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.

Offered Services - Genomics

- ***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 (viruses, bacteria, 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 applications:*** 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 Oxord Nanopore is currently being tested for Eukaryotic methylomes.

Offered Services – Transcriptomics

- ***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 smallRNA
- ***single cell RNA-Seq:*** *(will be established and offered in 2019)*
- ***mapping of transcriptional start sites:*** *(will be established and offered in 2018)*

General Services

- Consultations: study design and data analysis
- Integration of data from public databases
- Support for integration of generated data into public databases
- Provision of computing power (either together with ZIMt, or on own workstations)

Custom bioinformatics services

- 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

Teaching

- Bioinformatics and data analysis courses (Galaxy platform, SLURM & HPC)
- Wet lab protocols (transfer of knowledge to working groups)
- Practical training courses (2 days) and electives for students of Biomedicine (HBRS) and Biochemistry

Focus: Setup of a HPC 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 MHH HPC cluster provides significant compute power (currently 600 cores in 13 nodes, a total of 5 TB RAM, 300TB storage, 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. In total, an interdisciplinary team of the Center for Information Management and the RCUG take care of administration, development, maintenance, software installation, testing and upgrades of the cluster. At present (October 2018), over 299,000 analysis jobs have been submitted to the HPC cluster.
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 Cluster. At present, the HPC 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 MHH HPC 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 v3.4, Agilent GeneSpring v13.1.1, Qiagen Ingenuity Pathway Analysis (IPA) and Avadis StrandNGS v3.2. Regular practical training courses are provided in the dedicated course room of the RCUG (8 dual Windows/Linux workstations). These courses cover RNA-seq, DNA-seq and mutation analysis and, in future, long read analysis.