Tuesday, October 14, 10:00 am – 12:00 pm

Nextflow Run: Getting Started with Nextflow for Bioinformatics

Workshop Description

Nextflow is a powerful and flexible open-source workflow management system that simplifies the development, execution, and scalability of data-driven computational pipelines. It is widely used in bioinformatics and other scientific fields to automate complex analyses, making it easier to manage and reproduce large-scale data analysis workflows, using containerized software to maximize portability and reproducibility. Complementing this system is the nf-core project, a collaborative effort to develop and maintain high-quality open-source bioinformatics pipelines for a wide range of research applications.

This training workshop is intended as a "getting started" course for students and early-career researchers who are completely new to Nextflow and nf-core. It aims to equip participants with foundational knowledge and skills in three key areas: (1) understanding the logic of how data analysis workflows are constructed using Nextflow, (2) configuring and executing workflows effectively on any platform (HPC or cloud), and (3) utilizing existing nf-core tools and pipelines to increase standardization and reduce duplication of effort for common types of analyses.

Participants will be guided through hands-on, goal-oriented exercises that will allow them to practice the following skills:

- Understand how core components of the Nextflow language are used to manage input data and construct multi-step workflows that support common dataflow patterns
- Launch Nextflow workflows locally, navigate output directories to access results, interpret log outputs for insights into workflow execution, and troubleshoot basic issues that may arise during workflow execution.
- Find and utilize nf-core pipelines that are relevant to their work.

By the end of the workshop, participants will be well-prepared for tackling the next steps in their journey to develop and apply reproducible workflows for their scientific computing needs. Additional study-at-home materials will be provided for them to continue learning and developing their skills further.

Schedule

- **Introduction to bioinformatics workflows:** A general introduction to why workflows are important and what are the main concepts and mechanisms involved.
- **Reading and running Nextflow:** Interactive exercises focused on interpreting and running a set of Nextflow workflows of increasing complexity.
- **Configuration:** Interactive exercises demonstrating how to use Nextflow configuration options in order to run the same workflows on different computational platforms without needing to modify any code.
- **Finding and using nf-core pipelines:** Interactive exercises demonstrating how to find relevant nf-core pipelines and try them out.
- **Next steps and Q&A:** Discussion of educational resources that participants can use to continue developing their Nextflow skills, specifics depending on audience interests.

Learning Objectives:

- Investigate the logic of how data analysis workflows are constructed using Nextflow
- Practice launching and managing workflows efficiently
- Apply configuration options to run Nextflow workflows on any computational platform
- Identify and evaluate existing pipelines from the nf-core project

Tuesday, October 14, 10:00 am – 12:00 pm

Assessing Genetic Variants for Antisense Oligonucleotide Therapy Amenability: Practical Training Workshop and Discussion

Workshop Description

Of the around 7,000 known rare diseases worldwide, disease-modifying treatments are available for fewer than 5%, leaving millions of individuals without specialized therapeutic strategies. In recent years, antisense oligonucleotides (ASOs) have shown promise as individualized genetic interventions for rare genetic diseases. However, until recently, there was no consensus on which disease-causing DNA variants are suitable candidates for this type of genetic therapy. The Patient Identification Working Group of the N=1 Collaborative (N1C), alongside an international group of volunteer assessors, has developed and piloted consensus guidelines for assessing the eligibility of pathogenic DNA variants for ASO treatments in 2024. In this workshop, we will introduce the N1C VARIANT (Variant Assessments towards Eligibility for Antisense Oligonucleotide Treatment) guidelines and instruct the participants on how to use them in assessing their patient cases/cohorts.

The workshop will introduce ASO therapies as a possibility to treat individuals suffering from rare genetic disorders and provide instructions on how to assess pathogenic, disease-causing variants for the three currently best-established ASO treatment approaches: splice correction, exon skipping, and downregulation of RNA transcripts. We will explain how a variant is classified as either “eligible”, “likely eligible”, “unlikely eligible”, “not eligible”, or “unable to assess”; and what these classifications mean in regard to prioritizing individuals for potential therapeutic intervention. We will further review key considerations for assessment for upregulation of transcripts from the wildtype allele, an emerging ASO therapeutic strategy. As part of the workshop, we will demonstrate tools and training material to enable clinicians and researchers to use these guidelines for their own eligibility assessments.

Overall, we will provide the medical genetics community with guidance on identifying suitable candidates for variant-specific ASO-based therapies and discuss the possibility of integrating such assessments into routine clinical practice.

Schedule

- **Introduction to ASO treatments:** We will provide an overview of the currently available ASO technologies, including their possibilities and limitations.
- **Introduction to the N1C VARIANT guidelines:** We will introduce the participants to the N1C VARIANT guidelines, how they were established, and available training material.
- **Live examples and interactive tutorial: 4 examples + Q&A:** Together with the participants, we will perform 4 assessments live, using examples that have not yet been released in our training materials. Throughout the assessment process, we will be taking questions from the audience, allowing for a dynamic discussion.
- **Flipped classroom:** We will discuss with the participants how our guidelines can be implemented into day-to-day clinical practice and what potential tools need to look like for a user-friendly variant assessment.

Learning Objectives:

- Define and summarize the possibilities and limitations of ASO therapies
- Examine rare disease variants and determine their suitability for N=1 ASO therapies using the N1C VARIANT Guidelines
- Utilize publicly available tools and resources in variant assessment
- Identify emerging therapeutic strategies and their applications to rare disease patients

Tuesday, October 14, 10:00 am – 12:00 pm

The Human Pangenome: Data, Tools, and Workflows

Workshop Description

The human pangenome reference will be an important foundation for identifying and predicting the functional outcomes of variants across diverse populations. Researchers and clinicians conducting genomics research will need to know how to access pangenome data and resources, and how to integrate them with their own genomics datasets. The new pangenome reference will increase the genomic search-space and highlight population-specific variations that will empower identification of novel disease-risk alleles and previously unobserved rare variants. This will facilitate improved, more inclusive study designs and statistical models for both common and rare disease studies, leading to better understanding of both population-specific and shared variants for disease.

Beyond using pangenome data and tools, researchers must understand how enhanced representation of global genomic variation may shift conceptual models of how that variation is distributed within and among populations. This will have downstream implications for methods and frameworks to analyze both short- and long-read genomic datasets, implement clinical algorithms, and conduct gene and variation curation or annotation across genomics research and clinical genetics practice.

Though many in the field acknowledge the need to upgrade the human reference genome, the practical, financial, and logistical challenges of adopting a new graph-based data structure are steep. HPRC investigators acknowledge these challenges and seek to address them head-on by conducting workshops such as this proposed session to facilitate widespread adoption of the human pangenome reference.

Schedule

- **Topic overview and lecture:** We will introduce participants to the new human pangenome reference, how samples were chosen to maximize variation, and the technologies used to build the pangenome.
- **Activity instruction and tutorial:** We will provide hands-on instruction on how to access that pangenome resource, how to build a pangenome graph, how to conduct variant calling and short- and long-read mapping to the pangenome.
- **Q&A:** We will provide the audience an opportunity to ask questions about all aspects of the new human pangenome reference

Learning Objectives:

- Summarize the new human pangenome reference resource, including methods of sample selection to maximize genetic variation.
- Summarize the new human pangenome reference resource, including methods of sample selection to maximize genetic variation.
- Illustrate how to build pangenome graphs.
- Illustrate how to conduct variant analyses using the pangenome and to utilize tools for mapping their own short- and long-read data to the pangenome.

Tuesday, October 14, 2:00 pm – 4:00 pm

Pathway Analysis for All: Hands-On Training with Reactome

Workshop Description

The growth of high-dimensional single-cell 'omics data has created a need for computational analysis to guide interpretation and translate findings into biomedical insights. Biological pathway analysis methods, like Reactome, bridge this gap. Reactome is a curated, peer-reviewed knowledgebase offering high-quality pathway annotations across physiological and pathological processes.

In this workshop, we'll introduce Reactome's website and tools through guided presentations and hands-on exercises, using either a public dataset or participants' own data. We will provide a hands-on introduction to the Reactome pathway database and its suite of analysis tools. Through a combination of short guided presentations and interactive exercises, participants will learn how to leverage Reactome for pathway analysis and biological interpretation.

To accommodate different levels of experience and data availability, we offer two options for hands-on exercises. Participants can either:

- Use a publicly available dataset, ideal for those who do not have their own primary data.
- Analyze their own normalized dataset, providing a customized learning experience with their specific research questions in mind.

Detailed step-by-step instructions will be provided for both approaches, ensuring that participants can confidently navigate the analysis process.

Schedule

- **Short presentations**: Each section begins with a concise introduction to key concepts and tool functionalities.
- **Hands-on exercises**: Participants will apply the concepts in guided activities using ReactomeGSA.
- **Pair-and-share discussions**: After each exercise, small group discussions will help reinforce learning and encourage knowledge exchange.
- **Live Q&A via Slido**: We will collect and address audience questions
- **Next Steps and Community Involvement**: To conclude the session, we will highlight additional resources for further learning, including documentation, tutorials, and community forums.

By the end of this tutorial, participants will have gained practical skills in pathway analysis, developed confidence in using Reactome's tools, and explored how AI-driven support can enhance their research experience.

Learning Objectives:

- Analyze how pathway analysis methods can be applied to interpret high-dimensional 'omics data.
- Use Reactome's pathway analysis tools, including ReactomeGSA, to explore biological pathways.
- Explore the React-to-Me AI chatbot, which supports biological interpretation through interactive and multi-language assistance.
- Explore and navigate Reactome's pathway diagrams.

Tuesday, October 14, 2:00 pm – 4:00 pm

The Michigan Imputation Server: Data Preparation, Genotype Imputation, and Data Analysis

Workshop Description

Genotype imputation is a key component of modern genetic association studies. The Michigan Imputation Server (MIS) has helped >13,000 researchers from around the world impute >110 million genomes. This interactive workshop is designed for anyone interested in learning how to impute genotypes and how to use the imputed genotypes, avoiding common pitfalls. We encourage participants to ask specific questions about their own projects.

This interactive workshop is aimed at anyone interested in learning how to impute genotypes and how to use the imputed genotypes, highlighting the latest reference panels, including the multi-ancestry panel from the TOPMed programme and a specialised HLA panel. A brief overview of imputation and the server will be followed by demonstrations and exercises, including 1) quality control and preparation of genetic data for use on the MIS, with a special focus on multiple ancestry, chromosome X and the HLA region; 2) tracking runs and using the application program interface for larger jobs; 3) downloading data from the MIS and preparing data for genetic analysis; 4) performing a GWAS using imputed data (including the HLA region) and interpreting the results, taking into account the imputation quality; 5) using the additional features, such as the calculation of polygenic risk scores. Finally, we will also discuss the MIS-based TOPMed and Munich Imputation Server, taking into account GDPR-related challenges. We encourage participants to ask specific questions about their own projects. Workshop materials, including slides and example datasets, will be made available prior to the workshop and will remain online on the MIS website. We expect that this workshop will enable participants to generate high quality imputed datasets and analyse them effectively.

Schedule

- **Review of Core Concepts**: Essential imputation theory, the Michigan Imputation Server, reference panel evolution, and MIS architecture overview.
- **Demo + Interactive Exercises**:
    - Pre-Imputation QC: Data formatting, VCF file validation, batch submission strategies
    - Your first genotype imputation using the MIS
    - Post-Imputation Analysis: Quality metric interpretation (R^2/INFO scores), dosage conversion, and GWAS pipeline setup with TOPMed-imputed data. Includes HLA region analysis nuances.
    - Advanced Features: Parallel implementation of polygenic risk scoring, imputation quality browser diagnostics, and automated workflow optimization using Imputation Bot.
- **Ecosystem Overview**: Presentations of the TOPMed and Munich Imputation Servers, including data privacy considerations and panel selection strategies.
- **Guided Q&A**: Structured troubleshooting session where participants discuss their genotype imputation needs or challenges.

Learning Objectives:

- Perform quality control and prepare genotype data for imputation.
- Use the imputation server, understand the available parameters, choose the appropriate reference panel, and interpret (error) messages.
- Perform quality control on imputed genotypes, interpret quality metrics, use the imputed genotypes for GWAS, PRS, or meta-imputation.
- Conduct a GWAS analysis using imputed genotypes (with special focus on the HLA region and chromosome X).