

15  
MAY

SAINT-PETERSBURG  
15 – 18 MAY 2014

# SYSTEMS BIOLOGY WORKSHOP PROGRAMME



## WORKSHOP VENUE:

*ITMO University / Birgevaya line, bld. 14, Vasilyevsky Island.  
We'll meet you at 9:15 am near the turnstiles on the ground floor.*

## INTRODUCTION TO SEQUENCING. RNA-SEQ & GENE EXPRESSION

- 09:30 — 10:00 **Workshop opening**
- 10:00 — 11:00 **Module 1. Practical lessons from human genome race**
- Historic timeline and major player in sequencing: LMB/Wellcome Sanger Institute, Human Genome Project, Celera Genomics.
  - Practical lessons from genome race: paired-end sequencing, whole-genome shotgun sequencing etc.
- 11:15 — 12:15 **Module 1. Practical lessons from human genome race**
- Practice: Alignments to genome — Bowtie/TopHat and BWA.
  - Practice: Basic quality control — FastQC etc.
- 12:30 — 13:30 **Module 2. RNA structure and quantitation**
- RNA structure and regulation: promoters, 5'end, 3'end, splicing.
  - Major modification of RNA-sequencing — CAGE, 3'end and full-length RNA-seq — experimental and practical.
- 13:30 — 14:30 **Lunch**
- 14:30 — 15:30 **Module 2. RNA structure and quantitation**
- Practice: Transcriptome reconstruction — Cufflinks, Scripture etc.; transcriptome quantitation and differential expression — Rsem, DESeq, edgeR etc.
- 15:45 — 16:45 **Module 3. Gene expression analysis and biological hypothesis generation**
- Basic gene expression concepts.
  - Gene coexpression and transcriptional modules/clusters.
  - Pathway and gene set enrichment tests.
- 17:00 — 18:00 **Module 3. Gene expression analysis and biological hypothesis generation**
- Practice: Basic.
- 18:00 — ... **Networking**



## MEDICAL AND POPULATION GENETICS

- 10:00 — 11:00**    **Module 1. Genetic causes of a disease. DNA variants. HapMap & GWAS**
- Genetic background of diseases.
  - DNA sequence linkage to phenotype.
  - Genotyping and HapMap project.
  - Introduction to statistical methods. GWAS.
  - SNP array analysis with PLINK.
- 11:15 — 12:15**    **Module 1. Genetic causes of a disease. DNA variants. HapMap & GWAS**
- Practice: Public resources review. dbSNP, Genome Browser.
  - Practice: SNP array data processing pipeline structure: file formats, QC metrics. Case-control matching: PCA and other metrics.
  - Practice: Association tests and standard results presentation.
- 12:30 — 13:30**    **Module 2. Next generation sequencing. Exome and Genome sequencing**
- Motivation for rare DNA variation analysis.
  - Experimental methodology overview.
  - Data processing challenges and strategy. File formats. GATK data processing pipeline.
- 13:30 — 14:30**    **Lunch**
- 14:30 — 15:30**    **Module 2. Next generation sequencing. Exome and Genome sequencing**
- Practice: Raw sequencing data processing towards analysis ready form using GATK, snpEff, Picard and BWA.
  - Practice: DNA variant annotation tools.
  - Practice: Data QC metrics. Case-control matching approaches. GATK tools overview. FastQC. Plink-seq.
  - Practice: Exome sequencing analysis and standard results presentation.
- 15:45 — 16:45**    **Module 3. Data analysis interpretation and validation**
- Sanger sequencing validation of the key mutation findings.
  - Protein-protein interactions analysis and pathways discovery.
  - Functional validation approaches.
- 17:00 — 18:00**    **Module 3. Data analysis interpretation and validation**
- Practice: DAPPLE analysis.



## TRANSCRIPTIONAL AND EPIGENETIC REGULATION

- 10:00 — 11:00**    **Module 1. Basics of epigenetic regulation**
- Gene expression regulation: why study it?
  - DNA packaging: different levels of chromatin organization.
  - What regulates gene expression?
  - Definition and main mechanisms of epigenetic regulation: CpG methylation and histone modifications.
  - Model of transcriptional regulation.
- 11:15 — 12:15**    **Module 1. Basics of epigenetic regulation**
- Practice: Chromatin immunoprecipitation and ChIP-Seq: basic experimental workflow and bioinformatic perspective.
  - Practice: Wide and narrow chromatin marks and corresponding bioinformatic treatment.
- 12:30 — 13:30**    **Module 2. ENCODE - source of ChIP-Seq data**
- ENCODE project introduction.
  - Making sense of multiple ChIPSeq data: ENCODE experience, methods and standards.
  - Enhancers and Superenhancers.
  - Modern variations of ChIP-Seq, e.g. Chip-Exo.
  - Downstream analysis — MEME etc.
- 13:30 — 14:30**    **Lunch**
- 14:30 — 15:30**    **Module 2. ENCODE - source of chip-seq data**
- Practice: access to ENCODE data.
  - Practice: ChromHMM — machine learning based methods for analyzing multiple ChIP-Seq experiments (ChromHMM).
- 15:45 — 16:45**    **Module 3. RNA-based epigenetic mechanism**
- Types of RNA.
  - miRNA and their epigenetic role.
  - ENCODE prospective on RNA.
  - Epigenetics and medicine.
- 17:00 — 18:00**    **Module 3. RNA-based epigenetic mechanism**
- Practice: rediscovery of lincRNAs.

**18**  
MAY  
SAINT-PETERSBURG  
15 – 18 MAY 2014

# SYSTEMS BIOLOGY WORKSHOP PROGRAMME



## NETWORK BASED DATA INTEGRATION. INTRODUCTION INTO METABOLOMICS

**10:00 — 11:30**    **Module 1. Basics of metabolomics**

- Global metabolic networks. KEGG database.
- Basics of Flux Balance Analysis.

**11:45 — 12:45**    **Module 2. Network based data integration**

- Protein-Protein Interaction Networks.
- Basics of topological analysis and identification of enriched subnetworks.

**13:00 — 14:00**    **Module 2. Network based data integration**

- Practice: Cytoscape, BioNet.

**14:00 — 15:00**    **Lunch**

**18:00 — 21:00**    **Workshop closing. Farewell dinner.**

*The closing ceremony will take place on a boat cruising the rivers and canals of St. Petersburg. We meet at 5:50 pm at the pier in front of Tuchkov bridge (approximate address: Makarov Emb., 26).*

SEE YOU NEXT TIME!

# THANK YOU FOR BEING WITH US