Report from participation in the course:

Analysis of Next Generation Sequence Data Course For Complex and Mendelian Traits at Max Delbrück Center for Molecular Medicine,

Berlin, Germany, June 23-27, 2014,

by Paweł Gaj

The *Analysis of Next Generation Sequence Data Course For Complex and Mendelian Traits* course emphasized teaching of both theory and application of various methods to analyze next generation sequence (NGS) data.

Paweł Gaj and Piotr Stawiński have participated in the NGS course held Max Delbrück Center for Molecular Medicine, Berlin, Germany between June 23-27, 2014. Their participation was made possible in the framework of the T2.3 (know-how sharing and promotion) task of the BASTION project (FP7-REGPOT-2012-CT2012-316254-BASTION).

During the course attendees were learning how to design studies, call variants, analyze population- and family-based sequence data and evaluate variant functionality. Analysis of NGS data included performing complex trait rare variants, as well as (to certain extent) common variants association analysis for population-based and trio data. It also covered identifying variants for Mendelian traits.

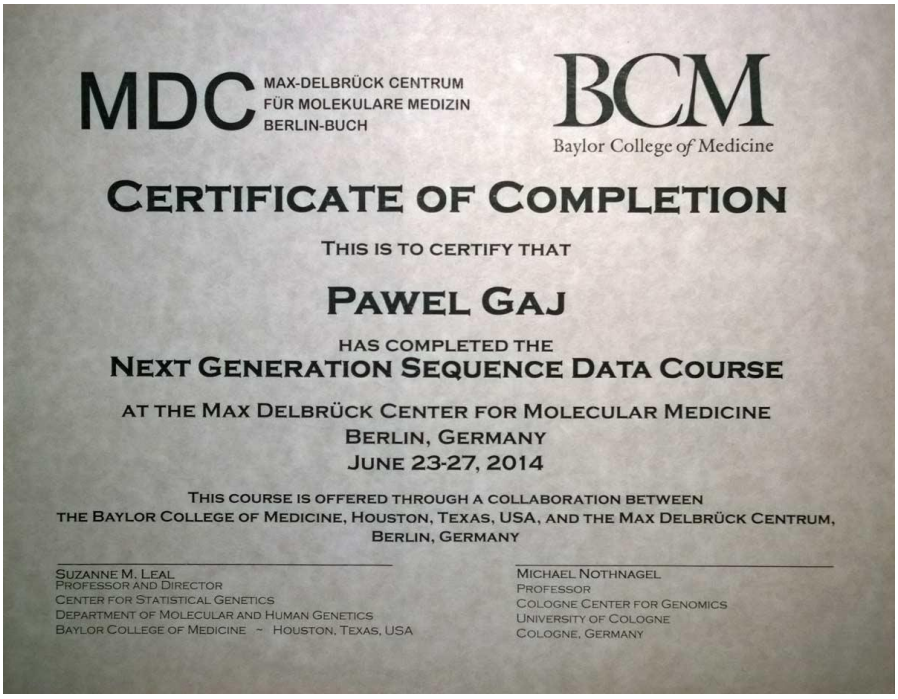
Specific topics of the course included: sequence alignment, calling variants, population genetics, data quality control, association testing, rare variant association methods, power estimation, identifying Mendelian variants, imputation, and evaluating variant functionality. Exercises were conducted using variety of computer programs including: GATK, Polyphen2, PSEQ, SEQPower, & VAT.

Personally, participation in the course has also given me a precious opportunity to discuss issues related to modeling complex interactions between genetic variants and various types of confounder variables related to both clinical sample annotations as well as population substructure of the studied cohorts.

In this context, particularly valuable was the top notch experience of the instructors (Michael Nothnagel and Suzanne Leal) in programming in commonly used R-project environment and other popular programming languages.

Just after completing the course I had a nice opportunity to use the newly acquired expertise in a project studding effects of genetic variation found in the genomic DNA of the CML and myeloma patients on the efficiency of Imatinib- and Lenalidomide-based treatments. Abstracts mentioning the obtained results of these projects were recently submitted for revision by the American Society of Hematoloty (ASH).

Instructors: Laurent Francioli; Suzanne Leal; Michael Nothnagel & Peter Robinson



MDC MAX-DELBRÜCK CENTRUM FÜR MOLEKULARE MEDIZIN BERLIN-BUCH

BCM Baylor College of Medicine

CERTIFICATE OF COMPLETION

THIS IS TO CERTIFY THAT

PAWEL GAJ

HAS COMPLETED THE

NEXT GENERATION SEQUENCE DATA COURSE

AT THE MAX DELBRÜCK CENTER FOR MOLECULAR MEDICINE
BERLIN, GERMANY
JUNE 23-27, 2014

THIS COURSE IS OFFERED THROUGH A COLLABORATION BETWEEN
THE BAYLOR COLLEGE OF MEDICINE, HOUSTON, TEXAS, USA, AND THE MAX DELBRÜCK CENTRUM,
BERLIN, GERMANY

SUZANNE M. LEAL
PROFESSOR AND DIRECTOR
CENTER FOR STATISTICAL GENETICS
DEPARTMENT OF MOLECULAR AND HUMAN GENETICS
BAYLOR COLLEGE OF MEDICINE ~ HOUSTON, TEXAS, USA

MICHAEL NOTHNAGEL
PROFESSOR
COLOGNE CENTER FOR GENOMICS
UNIVERSITY OF COLOGNE
COLOGNE, GERMANY

Course Schedule

Monday June 23rd:

Morning:

Lecture: Aligning Sequence Data, Calling single nucleotide variants (SNVs); Recalibration

Afternoon:

Lecture: Detection of Indels/Copy number variants (CNVs), VCF file format and annotation, Visualization of next generation sequence (NGS)Data
Computer exercises: FASTQC, Genome Analisis Toolkit (GATK) and Integrative Genome Viewer (IGV)

Tuesday June 24th:

Morning:

Lecture: Quality control for NGS data, Annotation of NGS data, Detecting variants for Mendelian traits, Computer exercises ANNOVAR, VEP

Afternoon:

Lecture: Detecting variants for Mendelian traits (continued), Poulation Genetics (drift, nutation drift & equilibrium, effective population size, selection, population substructure)

Pencil and Paper Exercise:
Calculation of population-based statistics e.g. Fst
Computer Exercises: GEMINI

Wednesday June 25th:

Morning:

Lecture: Association Analysis Testing Framework for qualitative and quantitative traits. Fixed effects and variance components analysis
Computer Exercises: R

Thursday June 26th:

Morning:

Lecture: Rare variant association methods for population based and trio data, controlling for covariates and population substructure/admixture
Computer Exercises: PSEQ, VAT

Afternoon:

Lecture: Power Analysis for rare variants, replicating rare variant associations
Computer Exercises: SEQPower

Friday June 27th:

Morning:

Lecture: Imputetion of rare variants and their analysis

Afternoon:

Lecture: Predicting functionality of Variants using bioinformatics tools
Computer Exercises: GERP, PhyloP, PhastCons, Polyphen2, SIFT, Provean, FATHMM, Mutation Assessor, Mutation Taster, CADD