Human Genome Variation Group Rotation

Report

Authors: Dedan Githae, Szymon Gaweł, Daria Krzysztofik, Michał Kowalski, Agnieszka Pełesz, Joanna Rudnicka, Pola Wartalska, Elżbieta Buczek

Table of content

Scope of rotation

During rotation at Human Genome Variation Group at Małopolska Centre of Biotechnology, we ought to learn about the main scientific topics around which this group is focusing their work. After the introduction to main scopes, we were familiarized with methodology of currently conducted studies and also of those concluded ones. We learned that modern day genomics and genetics incorporates a vast amount of interdisciplinary techniques from biochemistry through bioinformatics and computer science to automated robotics for tasks that belong to core facility within the group.

Biochemistry in modern genomics

Biological and Biochemical techniques with which we had opportunity to get along were presented in form of laboratory practice. We learned how to prepare samples for detection and quantification of methylated DNA for cancer studies, gene expression studies, detection of genetic diseases and many more diagnostic biological aspects. We had also learned how to prepare samples for SNaP shot Multiplex for Single Nucleotide Polymorphisms Genotyping and how to perform such genotyping for age estimation from saliva. There was also an opportunity to get familiarized with DNA isolation for further age estimation from other tissues like sperm.

DNA methylation

The showcase of DNA Methylation detection was based on EZ DNA Methylation-Direct™ Kit from ZYMO RESEARCH. During this practice everyone had an opportunity to watch, learn and take a part in sample preparation following the protocol with supervision of the wet-lab scientists **[https://files.zymoresearch.com/protocols/_d5020_d5021_ez_dna_methylation-direct_kit.pdf]**. In order to study DNA methylation, we performed bisulfide conversion, which involves converting unmethylated cytosines to uracil, while methylated cytosines remain unchanged. This method is the gold standard for detection DNA methylation level, but it is not without drawbacks, as, for example, DNA after such conversion is heavily degraded. Also during sequencing, uracil is read as thymine, so a lot of information is lost, that is why new alternative methods are still being sought. During this practice we were able to educate ourselves not only in terms of laboratory techniques, but also in terms of understanding why such steps are needed to be performed and what will be the gain of information as the result of sample preparation and sequencing after bisulfite treatment. Material which we were converting with EZ DNA Methylation-Direct™ Kit were our own buccal swabs. Later on we were also introduced to Quality Control of such prepared samples to see whether they are good enough for sequencing.

SNaP shot system

SNaP shot System is a primer extension–based method developed for the analysis of single nucleotide polymorphisms (SNPs). Based on this method we can determine the changes in individual nucleotides. Using this method in the HGVG laboratory, we were able to find out that it could be useful for to identify age estimation based on blood, sperm and saliva samples. Often this method is used to assess epigenetic effects, to detect methylation. During the rotation, we were presented the entire method and the data interpretation process.

DNA isolation from tissues - Phenol-Chloroform method

During the exercises, we had the opportunity to become familiar with the isolation of DNA from tissues by the phenol-chloroform method. Its mechanism is based on the solubility of biomolecules, and uses phenol, chloroform, water and isoamyl alcohol. This method can also be used to obtain RNA using the appropriate pH (when obtaining DNA, the addition of a buffer with a pH of 8.0 is used, and acidified phenol for RNA isolation)

Each of these reagents has its use in the reaction, phenol denatures proteins, chloroform, like phenol, helps in denaturation of proteins, also dissolves lipids and acts as an additional factor facilitating the correct separation of the water and phenol phase. Isoamyl alcohol acts as an antifoam, preventing the mixture from emulsifying.

Since DNA is a large, polar molecule with a negative charge, after lysis of cells and denaturation with phenol, DNA is not able to dissolve in it, and therefore completely dissolves in water present in the mixture. The denatured proteins, in turn, will remain in phenol after centrifugation. Because phenol is more dense than water, phase separation occurs, the upper is water with dissolved DNA, and the lower is phenol with denatured proteins. This allows the water phase and, in effect, the DNA to be recovered.

The efficiency of this method, as well as the purity of the obtained nucleic acids is worse than in the case of advanced DNA isolation kits, but if you have a large number of tissues or cells, its simplicity is an undoubted advantage.

Bioinformatics and Computer Science

Next Generation Sequencing Data Analysis

During the lab presentation we were briefly introduced to bioinformatic pipelines used for NGS data analysis in Human Genome Variation Group. Although this rotation was prepared mostly to learn about scientific scopes and wet-lab techniques we had gained insight that most of the analysis is performed with usage of R programming language **[https://www.r-**

project.org/about.html] with usage of repository of bioinformatic pipelines, libraries and tools called Bioconductor **[https://bioconductor.org/]**.

Computer Science

Some solutions discovered or developed needs to be implemented as web services, libraries or standalone apps. Those solutions also are mostly developed in a way that will be from beginning to end in high standard or even state of the art methodology. We were introduced to one of the most curious examples of incorporation of machine learning techniques with genomics – the HIrisPlex-S System **[https://hirisplex.erasmusmc.nl/]** which was designed for prediction of skin, eye and hair color from Single Nucleotide Polymorphisms data. We learned that this system was developed in high standards with almost unbiased classes. The Multinomial Logistic Regression (MLR) algorithm which is the backend of the system was chosen with the Ockham Razor Principle for explainability and reproducibility. This allowed to obtain probabilities and decision borders for each class without One vs All problem with simultaneous preservation of simplicity and efficiency of system **[https://pubmed.ncbi.nlm.nih.gov/22917817/]**.

Automated robotics and laboratory equipment

Throughout 2 weeks of rotations we were able to use and observe many sophisticated laboratory accoutrements. We learned that core facility is equiped with state of the art robot (TECAN FREEDOM EVO 150) for automated pippeting for large quantities of material. There were also many sequenators for Next Generation Sequencing. Among them we had seen and used:

- Thermo Fisher Ion Torrent PGM **[https://www.thermofisher.com/order/catalog/product/4462921#/4462921]**,
- Thermo Fisher Ion Proton **[https://www.thermofisher.com/order/catalog/product/4476610#/4476610]**,
- Oxford Nanopore MinION **[https://nanoporetech.com/products/minion]**,
- Qiagen PyroMark Q48 **[https://www.qiagen.com/pl/products/discovery-and-translational-research/pyrosequencing/instruments/pyromark-q48-autoprep/#orderinginformation]** and
- Thermo Fisher ABI 3500.

For the QC of material there are Nanodrops, Thermo Fisher Qubit,Agilent Bioanalyzer 2100, Applied Biosystems QuantStudio 12k Flex. For preparation of samples we had also used Qiagen TissueLyser II, Bertin Instruments Tissue Homogenizer.

Practical applications of modern genomics

Forensics

The main goal of forensic genetics is to adapt medical knowledge and skills to the needs of courts, and the role is to combine genetic methods with the law. Forensic genetics can allow for: identification of human corpses and remains, identification of persons, including newborns swapped in a hospital, identification of swapped samples of biological material, determination of disputed paternity and motherhood or examination of polymorphic features inherited according to sex (mtDNA and features from the Y chromosome).

Earlier mentioned HIrisPlex system is a great example of how human genome variation studies can be used in forensic applications. Currently, we had also learned that Human Genome Variation Group is also a part of a new grant, focused around understanding and usage of soil metagenomes for forensic applications and also another one where the main topis is epigenome analysis for investigation purposes – searching and developing new methods increasing the ability to identify and detect DNA.

Genome variability studies

All sexually reproducing organisms have an individual combination of DNA sequences and a unique genotype. The smallest differences may concern individual methyl groups attached to cytosines, the largest ones include fragments with a length of millions of base pairs. During life, differences in the methylation pattern will also increase in DNA of monozygotic twins. Some of the differences that we observe in the DNA sequence affect the diversity of phenotypic traits. Nowadays, we know specific sites in the genome that we call biomarkers (we distinguish biomarkers of individual's aging rate, general health, age-related diseases and mortality / survival). Accurately described, with the help of molecular biology tools, they can inform us about the characteristics of the person from whom DNA was obtained, such as progressive appearance features (graying, hair loss, wrinkles). What is more, information about our lifestyle (smoking, playing sports or alcohol abuse) is also stored in our genome.

Epigenome studies

Epigenetics involves genetic control by factors other than an individual's DNA sequence. Epigenetic changes are an indispensable element of our lives. One such epigenetic factor is the DNA methylation process, which is determined, among others on the degree of activation of individual genes and the amount of synthesized proteins. For some time, science has focused more on the epigenome and its role in our lives. In the group HGVG of interest is based on age prediction methylation, which can be crucial in the context of forensic tests, but not only. Studying epigenetic aging markers can help with better understand how environmental factors affect people.

Conclusions/Summary

Laboratory rotation at Human Genome Variation Group at Małopolska Centre of Biotechnology has given an introduction into the human genetic and epigenetic research workflow. The 2-week laboratory course has provided insights into: theoretical background of human epigenomics, high-throughput DNA sequencing technologies use, bioinformatics and computer science application. We ought to learn about interdisciplinary studies on human genome variation that can define phenotypic features and the influence of biological aging on phenotype progression. Predictive DNA analysis, in forensics and anthropology, can give important information about age, appearance and even ancestry. What is important, methodology and data interpretation process were presented step-by-step, which gives an idea about the whole picture of presented research area. Therefore, we gained the experience not only in terms of laboratory practice, but also in understanding the purpose of human genome variation research.

References

America, S. (2003). Instruction Manual Instruction Manual. *International Business*, 1–183.

Macherey-Nagel. (2014). *Genomic DNA from Tissue, User manual, NucleoSpin®Tissue*. June, 40.

Number, F. C., Samples, D. N. A., Kit, S., Agencourt, C., Xp, A., Thermal, E., Module, C., Technologies, O. N., Sequencing, L., Ffpe, N., Repair, D. N. A., Promethion, M., Cell, F., Kit, P., Flp, E.-, Cell, F., Kit, P., Flp, E.-, End, N., & Module, A. (2019). *Before start checklist*. 1–6.

Qiagen. (2015). *PyroMark ® Q48 Advanced Reagents and PyroMark Q48* Advanced CpG Reagents. October.

Technologies, A. (2011). *Quick-Start Protocol Preparation of 6xHis Protein Ladder January 2011 Sample & Assay Technologies Sample & Assay Technologies*. 34705, 2–3.

Tga, A. G. G., Ttt, G. T. T., Tgg, A. T. T., Tta, T. A. T., Aa, A. G. A., Gag, G. G. A., Ggt, G. G. A., (n.d.). *Multiplex SNaPshot for Age Estimation Using Saliva Reagents Needed : Post-PCR Reaction Enzyme Purification of the PCR Product*. 1–4.

Walker, J. M. (n.d.). *Forensic DNA Typing Protocols IN Series Editor*.

Walsh, S., Liu, F., Ballantyne, K. N., Van Oven, M., Lao, O., & Kayser, M. (2011). IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Science International: Genetics*, *5*(3), 170–180. https://doi.org/10.1016/j.fsigen.2010.02.004

Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., Branicki, W., & Kayser, M. (2013). The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Science International: Genetics*, *7*(1), 98–115. https://doi.org/10.1016/j.fsigen.2012.07.005