Tin sinh học ứng dụng

(Applied Bioinformatics)

Phuc Loi Luu, PhD

Luu.p.loi@googlemail.com

Nov 02 2024

Content

- Class member introduction
- Curriculum of Applied Bioinformatics 2024
- Projects and important dates
- Evalutation for the course
- How my lecture work?
- Linux OS, Google Colab and Linux Command Lines

Class member introduction

EDUCATION AND PROFESSION

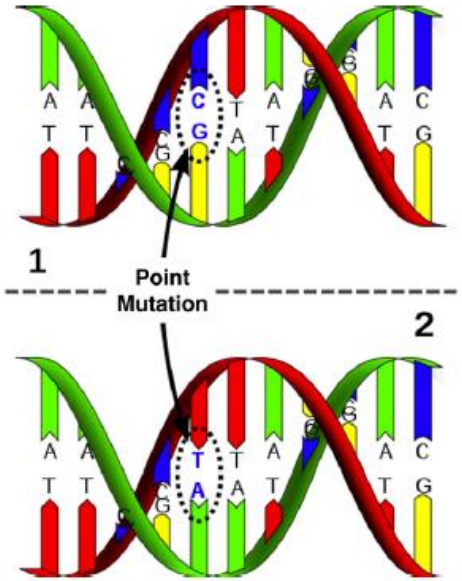
2014 – 2022	**Garvan Institute of Medical Research,** UNSW Sydney	*Bioinformatics researcher in Computational Epigenomics under the supervision of Prof. Susan Clark Conjoint Senior Lecturer*	Sydney, Australia
2011 - 2014	**Max Planck Institute for Molecular Biomedicine,** University of Muenster	*PhD student in stem cell computational biology under the supervision of Prof. Hans R. Schöler*	Muenster, Germany
2010 – 2011	**KIST–Europe**	*Data Scientist*	Saarbruecken, Germany
2008 – 2010	**Max Planck Institute for Informatics,** University of Saarland	*Bioinformatics Post-graduate student under the supervision of Prof. Thomas Langauer*	Saarbruecken, Germany
2005 – 2008	**Nong Lam University** - Ho Chi Minh City	*Lecturer*	Ho Chi Minh City, Vietnam
2000 – 2005	**University of Science -** Ho Chi Minh City National University	*Bioinformatics undergraduate student under the supervision of Prof. Ho Huynh Thuy Duong*	Ho Chi Minh City, Vietnam

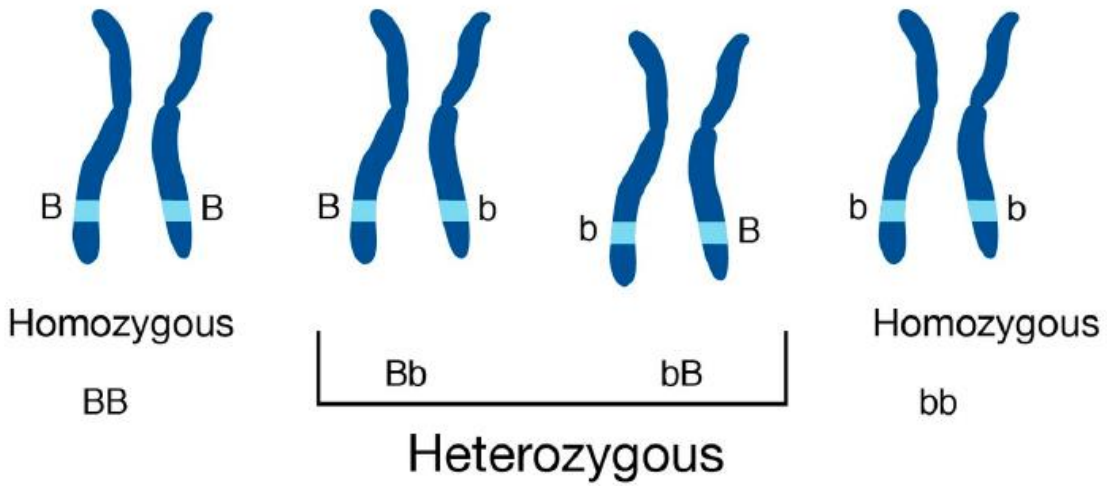
Curriculum of Applied Bioinformatics 2024

1. Linux OS, Google Colab and Linux Command Lines (02 Nov)
2. Basic Bash script and Awk (09 Nov)
3. Basic R (16 Nov)
4. Statistics with R (23 Nov)
5. Data visualization with R (30 Nov)
6. Review basic molecular biology and the human genome (09 Nov)
7. Introduction to Next Generation Sequencing (NGS), Gene annotation and the related databases (23 Nov)
8. Next Generation Sequencing (NGS) and its applications in clinic and agriculture (30 Nov)
9. Short read sequencing and upstream analysis for DNA-seq (7 Dec)
10. Short read sequencing and downstream analysis (variant calling and variant annotation) for DNA-seq (14 Dec)
11. A brief introduction to clinical genomics and precision medicine (21 Dec)
12. Long read sequencing introduction and upstream analysis for DNA-seq (28 Dec)
13. Long read sequencing and downstream analysis (denovo assembly and genome annotation) for DNA-seq (4 Jan 2025)
14. Presenting group projects (11 Jan 2025)

Human Genome Variation

- Humans genomes are >99% similar by sequence
- A typical human genome has ~5 million variants with 3-4 million single-nucleotide variants
- Humans are diploid

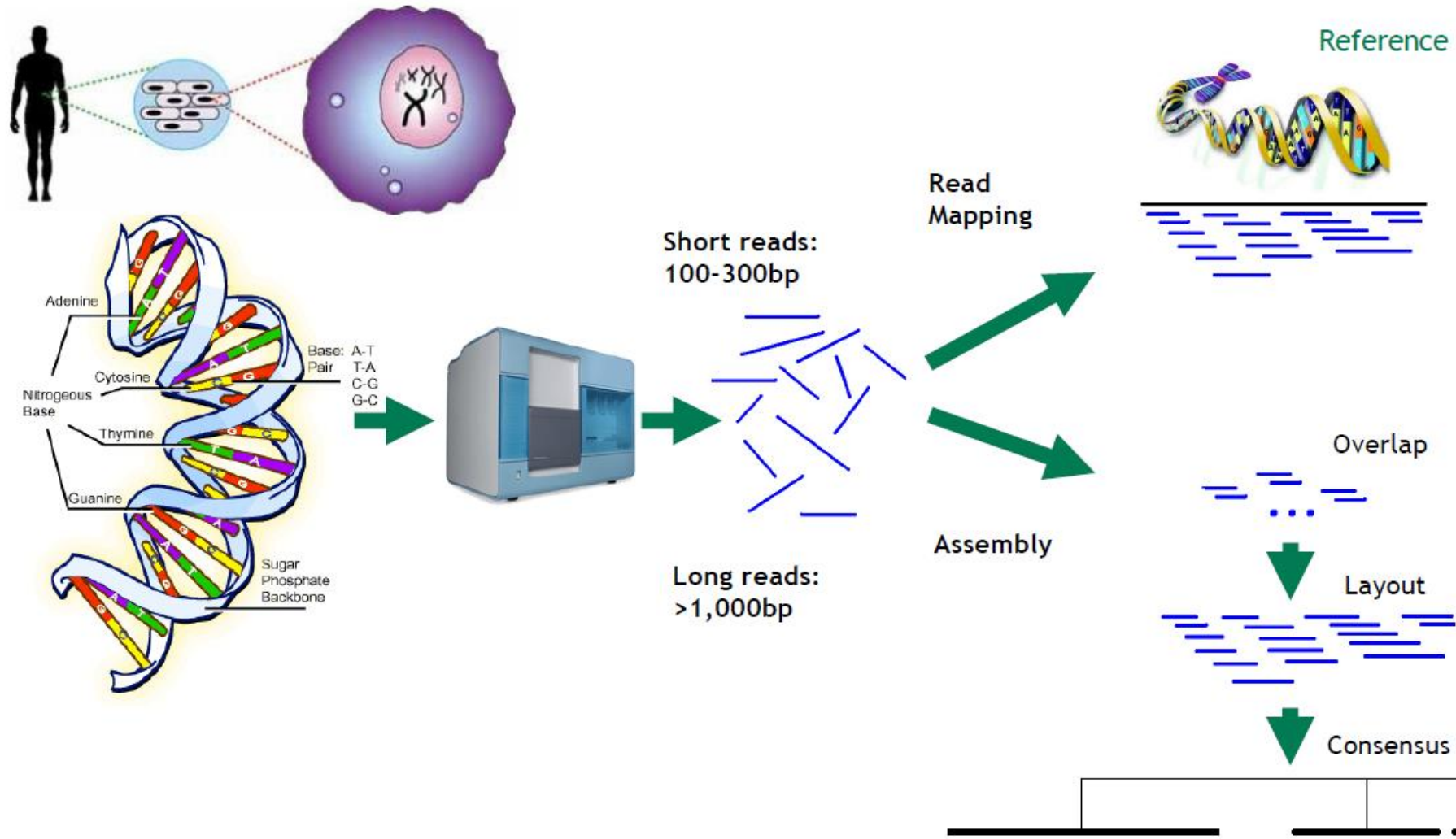




EMBL

Source: https://rosalind.info/media/point_mutation.png, https://en.wikipedia.org/wiki/Zygosity

Sequencing DNA

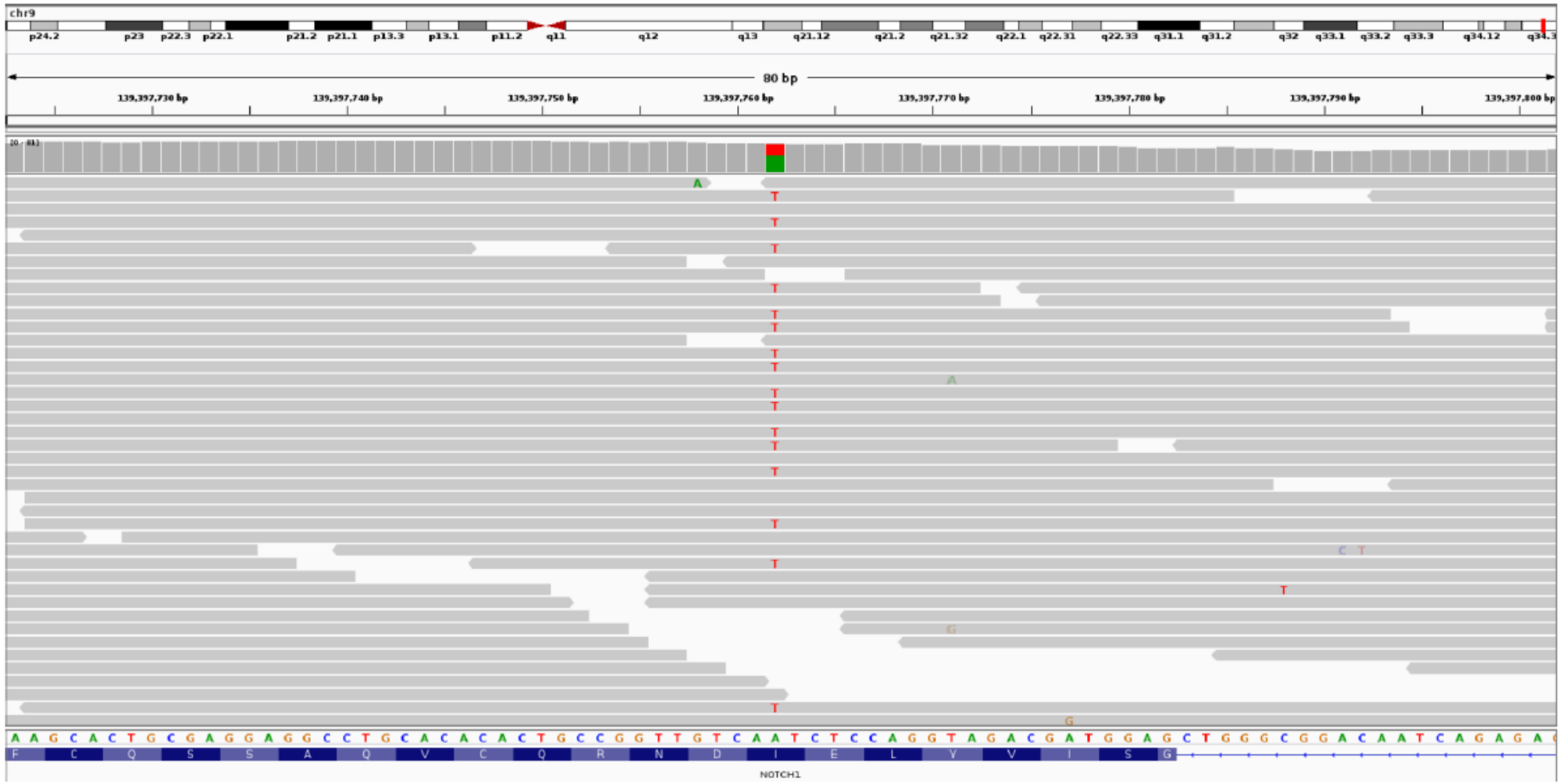


EMBL

Read Alignment

EMBL

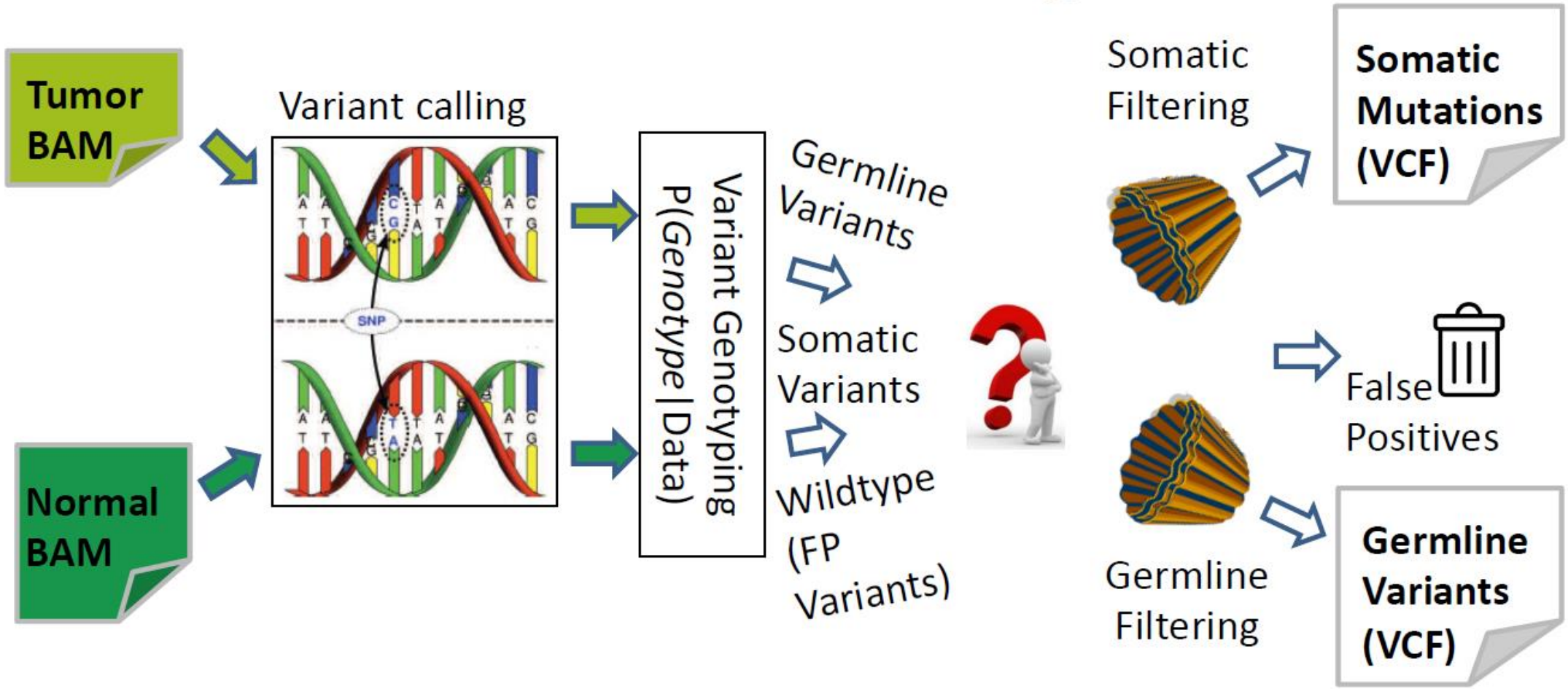
Alignment and variant viewers



': https://software.broadinstitute.org/software/igv/

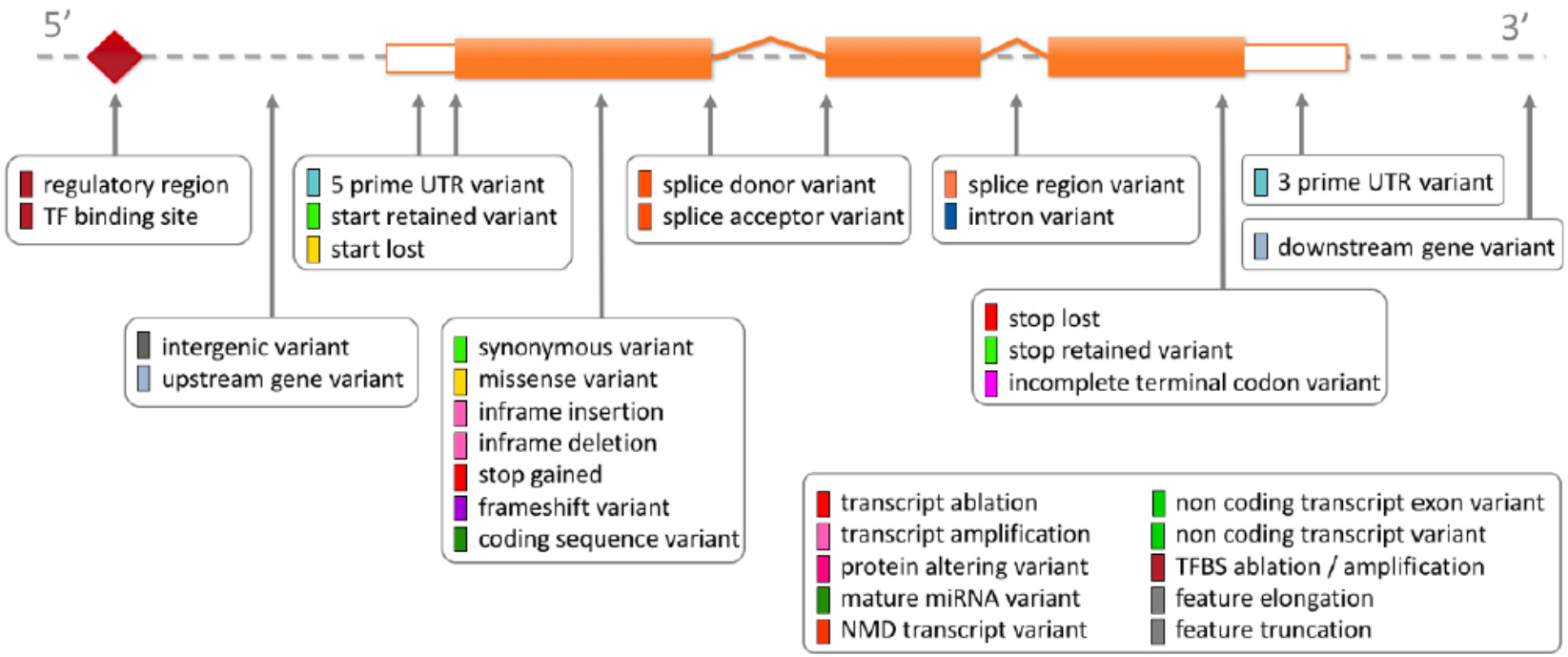
EMBL

Cancer Genomics Variant Calling



EMBL

Interpreting Genomic Variants



Popular Tools: VEP, Annovar, snpEff

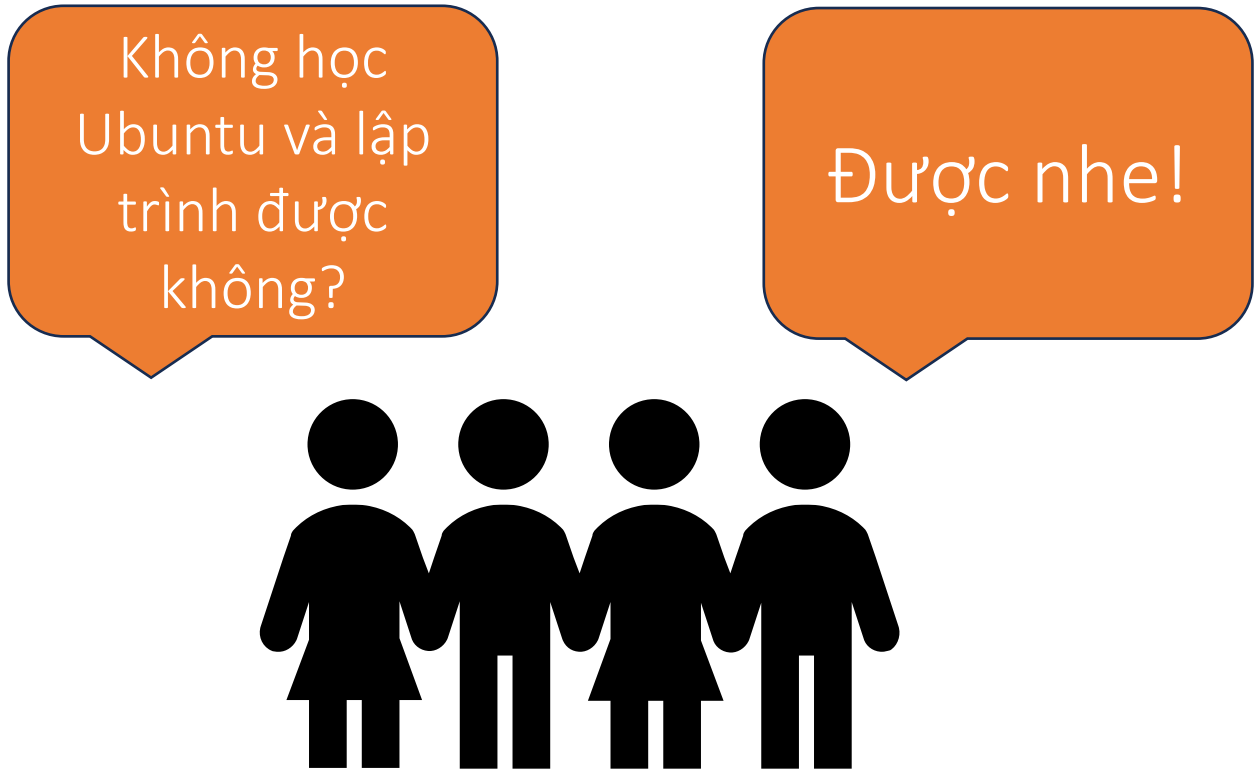
Source: https://m.ensembl.org/info/genome/variation/prediction/predicted_data.html

EMBL

Projects and important dates

No	Project	Aim	Requirement	Group
1	A breast cancer gene panel (BRCA1 and BRCA2)	Processing a dataset of 10 samples	- Programming bash - Human Moclecular biology - Clinical background	
2	A gene panel sequencing dataset of a cancer	Processing a dataset of 2 samples	- Programming bash - Human Moclecular biology - Clinical background	
3	Whole Exome sequencing (WES) dataset of a cancer	Processing WES and interpretation of the result	- Programming bash - Human Moclecular biology - Clinical background	
4	Whole Exome sequencing (WES) dataset of a genetic disease	Processing WES and interpretation of the result	- Programming bash - Human Moclecular biology - Clinical background	
5	Denovo assembly of a viral genome ONT/PacBio	Understanding processing and interpretation of the result	- Programming bash - Moclecular microbiology - Clinical background	
6	Variant calling and subtyping of bacteria Whole Genome Sequencing	Understanding the artical and interpretation of the result	- Programming bash - Moclecular microbiology - Clinical background	

Do we need to install and learn Ubuntu and programing?



Do we need to install and learn Ubuntu and programing?

Evalutation for the course

- Presentation date (final exam): 11 Jan 2025
- Max 30 min each group
- 5-10 questions each student
- Report (20) +

  Presentation (20) +

  Slide (10) +

  Questions for the presentation (25) +

  Questions for the all of lectures (25)

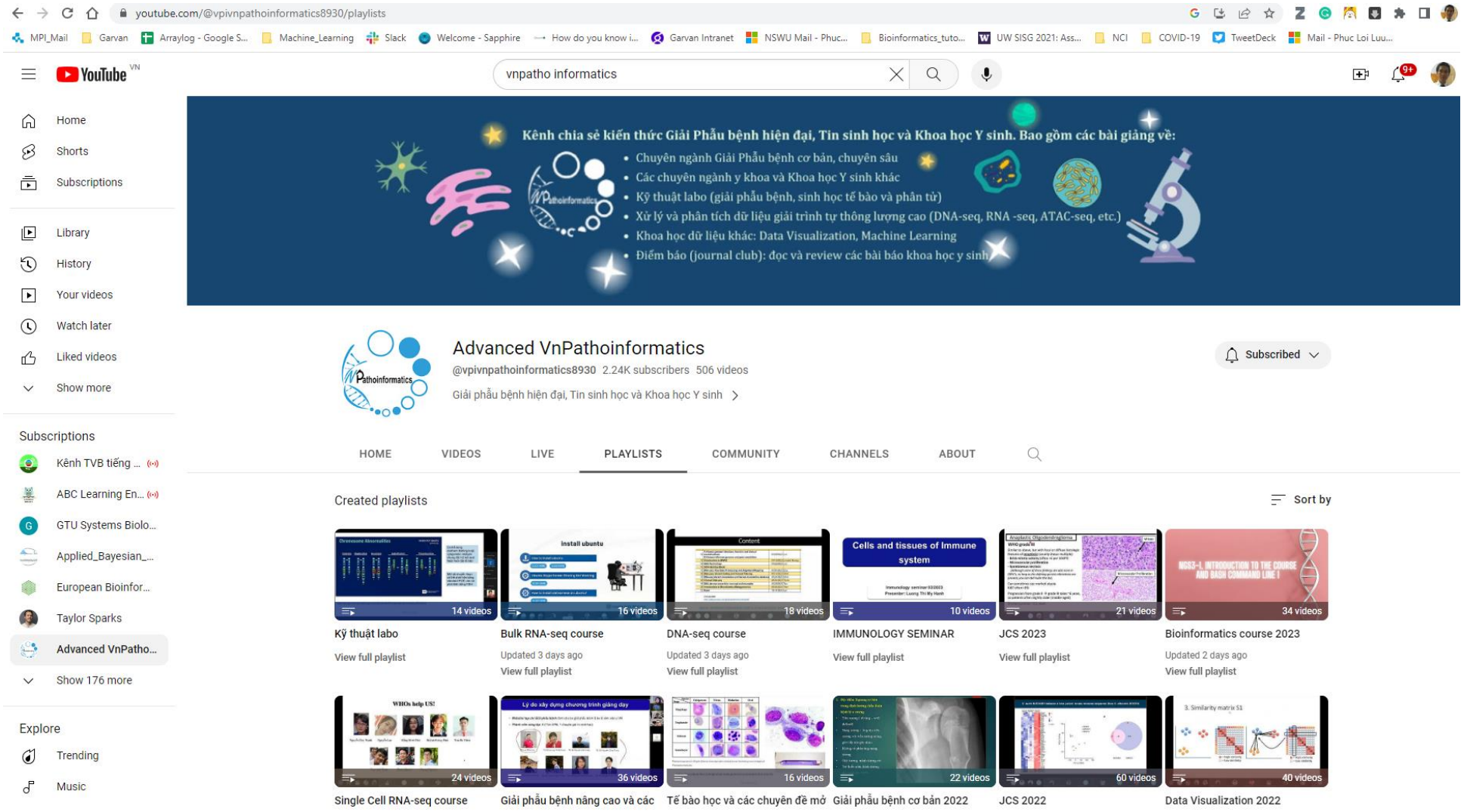
How does my lecture work?

- Start at 1:30 PM every Saturday from **02 Nov 2024** to **12 Jan 2025** (14 weeks)
- 15 min oral test at the begin of the lecture
- Time breaks in a lecture (**2:45-3:00 P**M and **4:00-4:15** PM)
- Lecture end at **5** PM
- No attendance checking

No attendance checking

No attendance checking

https://github.com/luuloi/AppliedBioinformatics2024/tree/main



https://www.youtube.com/@vpivnpathoinformatics8930/playlists

Thank you for your attention!