Duc-MINH NGUYEN-Le, MD

ducminhnguyenle30091996@gmail.com

https://github.com/ducminhnguyenle | https://www.linkedin.com/in/minh-nguyen-bio/

+84 0909724576 | Ward 11, Binh Thanh district, Ho Chi Minh city, Viet Nam.

SUMMARY

- A passionate bioinformatician, applying technology to improve human health. Previously a practicing pathologist and internal medicine physician. A genomic storyteller through data, visualization, and human interactions.
- Highly skilled in data analysis, I leverage statistical methods and AI tools to derive actionable insights, showcasing proficiency in employing data visualization techniques for effective communication. With strong expertise in handling large datasets and implementing advanced analytical models, I am a detail-oriented problem solver with a keen ability to identify patterns and trends. My commitment to continuous learning drives my adaptability in the evolving field of data analysis and bioinformatics.
- In health-focused genomic data analysis, I possess a basic yet solid foundation in statistical learning, AI, and molecular modeling. My specialization lies in translating complex healthcare datasets into actionable insights and optimizing processes. Eager to contribute to the healthcare sector, I am committed to further enhancing my skills, aiming to apply them to real-world challenges for improved outcomes. This dual proficiency positions me to bring valuable contributions to a dynamic data analysis role within the health domain.

EXPERIENCE

- **Institute for Applied Research in Health Sciences and Aging (ARiHA) - Thong Nhat hospital**

  Research scientist, HCMC Vietnam*, Mar 2025 - Present*

- **Genomics group at Department of Biomedical and Data Science - Tam Anh Research Institute (TAMRI)**

Bioinformatics Specialist, HCMC Vietnam, *Apr 2024 - Dec 2024*

- Contributed to clinical trial data analysis with a focus on genomics and metagenomics applications, leveraging insights to support research outcomes and advance precision medicine
- Designed and implemented a bioinformatics workflow tailored for Preimplantation Genetic Testing for Aneuploidy (PGT-A), optimizing accuracy and efficiency in detecting chromosomal abnormalities
- Implemented a specialized bioinformatics workflow for analyzing whole genome sequencing data in *Hepatitis D virus* and *Streptococcus pneumoniae* projects, enabling precise genomic insights and supporting targeted research outcomes
- Certificate for The Good Clinical Practice and Ethics Training (ICH-GCP) & Safety Reporting in Clinical trials training
- Actively participated in the testing and validation of electronic health record (EHR) capture and project management web applications, providing detailed feedback to optimize functionality, user experience, and data reliability

- **DataXight**

Clinical Genomic Informatician, HCMC Vietnam, *Dec 2023 - Apr 2024*

- Focused on the use of information systems and optimizing value stream mapping processes including biobank, pathology quality control of collected samples, liquid biopsy and cell culturing procedures and bioinformatics analysis workflow for NGS
- Optimized and implemented clinical network workflow to enhance transparency, reproducibility and productivity

- **Pacific Informatics - Zymo Research**

Bioinformatician, HCMC Vietnam*, Mar 2023 - Dec 2023*

- Designed, maintained and developed pipelines using Nextflow to improve the automation and end-to-end workflow time
- Communicated scientific and clinical results effectively with customers
- Led DNA high-throughput-sequencing upstream and downstream analysis projects, including WGS, WES and gene panels in application of rare genetic diseases, cancer

diseases

- Designed and developed data curation and engineering pipelines for large bioinformatic pipelines
- Designed and implemented bioinformatics pipelines using Nextflow to analyze genomic, transcriptomic and epigenetic datasets on AWS
- Benchmarking and statistical analysis of various bioinformatic tools

- **Gia Dinh People ‘s Hospital**

  Pathologist & Internal Medicine Physician, HCMC*, Feb 2021 – Oct 2022*

  - Gia Dinh People ‘s Hospital is a first-class general hospital and medical training facility, serving Ho Chi Minh city, a city of 9M denizens.
  - Harmonized clinical data with NGS data to interpret rare variants utilizing maternal and paternal WES data

EDUCATION & CERTIFICATIONS

- **Doctor of Medicine** – Pham Ngoc Thach University of Medicine (Sep 2014 – Dec 2020)
  - Graduated as a general physician with the second-highest score in the 2014-2020 class
  - Studied and practiced in pathology and internal medicine in Gia Dinh People ‘s hospital from 2020 to October 2022. This facility is a first - class general hospital and medical training site for the Ho Chi Minh City University of Medicine and Pharmacy (UMP) and Pham Ngoc Thach University of Medicine (PNTU)
- **Quality Control and Management Training Certification** - ISO 15189 criteria for Clinical Laboratory - Roche Vietnam and Green ISO company
- Certificate for **The Good Clinical Practice and Ethics Training (ICH-GCP) & Safety Reporting in Clinical trials** training
- Certificate of achievement of Swiss Institute of Bioinformatics (SIB): **Ensuring More Accurate, Generalisable, and Interpretable Machine Learning Models for Bioinformatics**

SKILLS

- **Programming language and environment**

- Proficiency in R, Python, Bash scripting
- Proficiency in Linux OS environment
- Basic understanding of AWS cloud environment
- Experience with Git & GitHub, Nextflow
- Familiarity with Docker containerization

- **Data analysis**
  - Proficient in using R packages such as ggplot2, tidyverse for data visualization and statistical analysis
  - Proficient in Python for statistical analysis, data processing and data visualization utilizing packages such as Pandas, NumPy, SciPy, and matplotlib
  - Experienced in data cleaning and preprocessing using Python libraries and R packages
  - Data curation for genomics, transcriptomics and epigenomics data
- **Statistical and machine learning application**:
  - Supervised learning techniques, including regression and classification models
  - Unsupervised learning techniques, including clustering, dimensionality reduction such as principal component analysis (PCA) and linear discriminant analysis (LDA)
  - Feature selection, including Multinomial Regression, Feature importances – Random-forest, partial least squares (PLS)
- **Bioinformatic data analysis**
  - Experienced in DNA sequencing upstream and downstream analysis, including Whole Genome Sequencing (WGS), Whole Exome Sequencing (WES), and Targeted Gene Panels
  - Proficient in Bulk RNA sequencing analysis using NGS, with abilities in identifying gene expressions and conducting enrichment pathway analysis
  - Familiar with publicly available databases such as NCBI, Ensembl, EMBL-EBI.
  - Proficient in designing and implementing bioinformatics workflows using Nextflow, with a basic understanding of pipeline architecture, efficient data handling, and error management
  - Proficient in DNA methylation analysis including:

- ❖ Reduced-representation bisulfite sequencing (RRBS-seq)
- ❖ Whole genome bisulfite sequencing (WGBS-seq)
- ❖ Targeted methylation region analysis (MethylCheck)
- ❖ Read level methylation analysis (CpG methylation patterns)

PERSONAL/COLLABORATIVE PROJECTS:

- **Clinical DNA Sequencing Analysis in Cancer Targeted Gene Panels:**
  - Gained hands-on experience in DNA sequencing data preprocessing
  - Performed variant calling, specifically identifying Single Nucleotide Polymorphisms (SNPs) and Copy Number Variations (CNVs)
  - Annotated variants using the ClinVar and ClinGen databases from NCBI
- **Quality Control Assessment and Optimization:**
  - Utilizing Nextflow bioinformatics pipelines to perform quality control for the MGI and GeneMind sequencing platform
- **Microcephaly Rare Disease analysis:**
  - Collaborating with Obstetricians and Genetic Physicians in TuDu's hospital to investigate and analyze a sophisticated Microcephaly disease
  - Hands-on experience in combining clinical records and high-throughput sequencing data to interpret rare variants utilizing maternal and paternal WES data
- **Developmental and Epileptic Encephalopathy analysis**:
  - Collaborating with pediatric neurologist in Children 2's hospital to perform a whole exome sequencing analysis to investigate and annotate variants related to epilepsy disease
- **Loss of 5-hydroxymethylcytosine and presence of NELL1 mutations are epigenetic and genetic hallmarks of granular cell tumors (GCT):**
  - Investigating the association between the number of gene SNVs vs 5hmC in 9 genes which found out from the previous studies in the literature (FAT3, ATP6AP1, ATP6AP2, ATP6V0C, BRD7, ALS2CL, TPRXL, LMAN1 and KCNN3)
  - Investigating the association between the number of gene SNVs vs 5hmC in whole exome sequencing (WES) datasets

- **A pathway-based genetic score for Oxidative Stress ($gPFS^{ox}$) and Inflammation ($gPFS^{in}$): Exploring the application of polygenic risk score (PRS) across diverse populations**
    - Utilizing 1000 Genomes Project public data for cross-continental allele frequency comparisons
    - Performing statistical analysis to identify a bell curve shaped-like gPFSox distribution among 6 different populations, including Vietnameses citizens

TEACHING EXPERIENCE:

- **Applied bioinformatics course 2023 and 2024 for Genetics Master Program at Genetics Department, University of Science - VNUHCM**
    - Teaching assistance for applied genetic bioinformatics and downstream DNA-sequencing analysis
    - Link:
        - https://github.com/luuloi/AppliedBioinformatics2023
        - https://github.com/luuloi/AppliedBioinformatics2024
- **UK clinical biomedical databank and genome-wide association studies (UK biobank) at University of Medicine and Pharmacy at Ho Chi Minh city**
    - Teaching assistance for statistical and bioinformatic analysis in GWAS case-control studies
    - Link: https://github.com/luuloi/GWAS_Introduction_2023
- **Next generation sequencing data analysis and clinical applications - Center for Bioscience and Biotechnology**
    - Teaching assistance for Bioinformatics Analysis: Research and Application of Next Generation Sequencing (NGS) in Breast Cancer
    - Link: https://github.com/precigene-team/CBBNGS-2024
- **Publicly Accessible Online Courses on YouTube by the Advanced VnPathoinformatics Group**
    - Contributing lectures and hands-on practical tutorials mainly on: Introduction to bioinformatics analysis for Next Generation Sequencing, DNA sequencing 2023, Bulk-RNA sequencing 2023 and Microbial Genome & Microbiome Analysis 2024 courses

- Link: https://www.youtube.com/@vpivnpathoinformatics8930

CURRENT PROJECTS/MANUSCRIPTS:

1. Profiling genetics and variant annotation of 500 congenital deafness, or congenital hearing loss using Whole Exome Sequencing (WES) with next generation sequencing (NGS) and Sanger sequencing
2. Profiling genetics and variant annotation of 12 epilepsy children using WES and Sanger sequencing
3. Profiling and re-classifying *BRCA1* and *BRCA2* variants of 124 Vietnamese breast and ovarian cancer patients using NGS
4. Single cell RNA-seq analysis of 12 Osteosarcoma (primary tumor, pre and post treatment, and metastasis)
5. Antimicrobial resistance prediction by clinical metagenomics in sepsis patients
6. Profiling soil microbiome in Lam Dong province using shotgun metagenomics
7. Profiling shrimp gut microbiome at aquaculture farm using 16S rDNA metagenomics
8. Developing bioinformatic algorithm for Noninvasive prenatal testing (NIPT)
9. Developing bioinformatic algorithm for Preimplantation Genetic Testing for aneuploidy (PGT-A)
10. Developing a Deep Learning Model for Classifying Lymphoid Lesions of the Digestive Tract on Hematoxylin-Eosin Stained Histopathological Image
11. An integrative analysis of $gPFS^{OX}$ scores across diverse populations

PUBLICATIONS

Phan-Canh T, Nguyen-Le D, Luu P, Khunweeraphong N, Kuchler K.0.Rapid in vitro evolution of flucytosine resistance in Candida auris.
mSphere0:e00977-24.https://doi.org/10.1128/msphere.00977-24

REFERENCES

LUU PHUC LOI, PhD
Head of Scientific Research Office
Institute for Applied Research in Health Sciences and Aging (ARiHA), Thong Nhat Hospital
Email: loilp@bvtn.org.vn