Summer Research Mentorship Proposal

A Bioinformatic Re-analysis of Neutrophil Extracellular Trap (NET) and Nitric-Oxide Pathway Gene Expression in Inflammatory Disease

Student	Nandika Trivedi, Rising Senior, Milpitas High School (MUSD), CA
Mentor	Prof. (Dr.) Madhu Dikshit, Emeritus / former Director, CSIR–Central Drug Research Institute (CDRI), Lucknow, India
Format	Remote (online) mentorship
Duration	8 weeks (extendable)
Estimated effort	~10–15 hours per week
Project type	Computational re-analysis of publicly available transcriptomic data

---

1. Summary

This project is a guided, computational research experience in which the student re-analyzes openly available gene-expression datasets to ask which **neutrophil extracellular trap (NET)-associated genes** and **nitric-oxide (NO) / nitric-oxide-synthase (NOS) pathway genes** are differentially expressed in an inflammatory disease relative to healthy controls, and how those genes cluster into biological pathways. The work bridges genetics, immunology, and computational analysis, and is designed for a high-school student in eight weeks using web-based bioinformatics tools and introductory Python/R.

The topic sits directly within Prof. Dikshit's research program on how nitric oxide regulates neutrophil functions including NETosis, oxidative burst, and microbicidal activity. The student's task is not to generate new laboratory data but to mine and interpret existing public data through the specific lens of Prof. Dikshit's NO–neutrophil framework as a genuinely informative exercise.

2. Background and rationale

Neutrophils are front-line cells of the innate immune system. Under stress they release **neutrophil extracellular traps (NETs)** - web-like structures of DNA and antimicrobial proteins that trap pathogens but, in excess, drive tissue damage in conditions such as sepsis, thrombosis, and inflammatory bowel disease. Prof. Dikshit's body of work

established how **nitric oxide** modulates neutrophil differentiation, survival, free-radical generation, and NETosis, making the NO–NET axis a coherent and well-characterized biological theme.

In parallel, the research community has increasingly mined public transcriptomic repositories (e.g., NCBI GEO) to identify NET-related and inflammation-related gene signatures in disease, an approach demonstrated in recent studies of Crohn's disease and sepsis. This project adapts that established method and adds a distinctive angle aligned with the mentor's expertise: examining the NO/NOS pathway genes alongside the NET gene set, rather than NET genes alone.

The project aligns with the student's interest in interdisciplinary biology/genetics and technology, builds on her demonstrated strengths in biology, and applies the computer-science skills she is actively developing. The analysis is largely tool driven.

3. Research question and objectives

Primary question: Which NET-associated and NO/NOS-pathway genes are significantly differentially expressed in [target disease] versus healthy controls, and what biological pathways do these genes represent?

Objectives:

1. Select and quality-check one (or two) public gene-expression dataset(s) appropriate to the question.
2. Curate, from the literature, a defined NET-related gene set and a NO/NOS-pathway gene set.
3. Perform differential expression analysis (disease vs. control).
4. Identify which NET and NO genes are significantly dysregulated.
5. Characterize the biology via pathway-enrichment and protein–protein interaction analysis.
6. Synthesize the findings into a written report and a short presentation, interpreted through the NO–neutrophil framework.

Candidate diseases (to be finalized with the mentor): sepsis, deep vein thrombosis / thromboinflammation, or inflammatory bowel disease (Crohn's). Each has well-characterized public datasets and a documented neutrophil/NET role.

4. Scope: what this project is and is not

- **It is** a reproducible, literature-grounded computational re-analysis producing original figures and an interpretation.
- **It is not** the generation of new experimental/laboratory data, nor a claim of novel biological discovery. Findings are framed as a student re-analysis that re-examines public data through a specific lens.
- All data used are **publicly available and de-identified**; no human-subjects approval is required, and no private or sensitive data are handled.

Setting these expectations protects the integrity of the work and is itself good scientific practice the student will learn.

---

5. Data and methods (overview)

Component	Resource / approach
Source data	NCBI Gene Expression Omnibus (GEO) — public microarray or RNA-seq datasets
Differential expression	GEO2R (runs limma) and/or limma/DESeq2 in R; optional Python (PyDESeq2/scipy)
Gene-list curation	NET and NO/NOS gene sets compiled from peer-reviewed literature
Data wrangling	Python pandas or R dplyr (filtering, joins, ID mapping)
Visualization	Volcano plot, heatmap, enrichment bar charts (matplotlib/seaborn or ggplot2)
Functional enrichment	Enrichr, DAVID (GO / KEGG pathways)
Network analysis	STRING (protein–protein interaction network)
Reporting	Structured written report + slide deck; optional illustrated graphic abstract

A recommended practical approach is a **hybrid**: use the mature R/GEO2R tooling for the differential-expression step, then move the results into Python for filtering, visualization, and interpretation. The detailed skills document accompanies this proposal.

6. Timeline (8 weeks)

Week	Focus	Key activities	Deliverable	Mentor checkpoint
1	Orientation & literature	Read 3–4 core papers (incl. Dikshit's 2024 NO–neutrophil review); define the precise question; shortlist candidate datasets	One-page scoping note + annotated reading list	Kickoff + question sign-off
2	Data & gene lists	Finalize dataset(s) with QC criteria; learn the data structure; curate NET and NO/NOS gene lists	Locked dataset + curated gene-list spreadsheet	
3	Differential expression	Run DE analysis (GEO2R / limma); produce results table; build volcano plot	DE results table + volcano plot	Review DE results
4	Focus on target genes	Intersect DE results with NET and NO gene sets; resolve gene-ID mapping; build heatmap	Filtered target-gene table + heatmap	
5	Functional analysis	GO/KEGG enrichment (Enrichr/DAVID); PPI network (STRING); interpret pathways	Enrichment figures + network + interpretation notes	Review biology
6	Synthesis	Assemble the biological story; connect to NO–NETosis framework; note limitations	Draft results + discussion	

Week	Focus	Key activities	Deliverable	Mentor checkpoint
7	Reporting	Write full structured report; polish figures; begin graphic abstract (optional)	Full draft report	Feedback on draft
8	Finalize & present	Incorporate feedback; finalize report; build short slide deck/poster; plan next steps	Final report + presentation + next-steps plan	Final review

---

7. Deliverables

1. A structured final report (Introduction, Methods, Results, Discussion, Limitations, References).
2. A figure set: volcano plot, heatmap, enrichment chart, and PPI network.
3. A curated, documented gene-list spreadsheet.
4. A reproducible analysis notebook (Colab/RStudio) with commented code.
5. A short presentation (slides or poster).
6. *(Optional)* An illustrated graphic abstract communicating the findings visually.

8. Success criteria

The project is successful if the student can, by Week 8:

- Independently explain what differential expression and pathway enrichment mean and interpret her own results.
- Produce clean, correctly labeled figures from real data.
- Articulate, in her own words, what the analysis suggests about NET/NO genes in the chosen disease — including its limitations.
- Present the work coherently to a non-specialist.

The standard is **understanding and reproducibility**, not novel discovery.

9. Stretch goals (optional, time permitting)

- Compare two independent datasets to test whether findings are consistent.
- Add unsupervised clustering of samples by NET/NO gene expression.
- Produce the analysis as a clean, shareable, fully reproducible notebook.
- Develop the illustrated graphic abstract into a polished science-communication piece.
- Prepare the work for a student research symposium, science fair, or preprint, subject to the mentor's guidance.

10. Data ethics and integrity

All datasets are publicly available and de-identified; the project involves no human subjects, no private data, and no laboratory procedures. The student will cite all data sources and prior studies, will clearly distinguish her re-analysis from established findings, and will not overstate the novelty or clinical relevance of results.

11. Contact

Student	Nandika Trivedi, nandikat.30@gmail.com
Mentor	Prof. Madhu Dikshit, CSIR–CDRI
Proposed start date	06/15/2026
Proposed end date	07/31/2026

This proposal is a starting point for discussion; the research question, target disease, and dataset will be finalized jointly with the mentor in Week 1.