Nobel Girmay

Data Scientist | Genomic Diagnostics | Clinical & Product Analytics

nobel-girmay.tech ♦ github.com/ngirmay

Professional Summary

Data Scientist with 5+ years of experience driving analytical decisions for FDA-regulated genomic diagnostics. Led high-stakes analytical validation studies, built scalable clinical data pipelines, and partnered cross-functionally across R&D, product, and regulatory teams. Strong background in statistics, NGS assay performance, and large-scale clinical datasets.

Skills

NGS: cfDNA/ctDNA pipelines, variant calling, TMB, MSI, CNV, Clinical reporting analytics

Programming Languages & Tools: R (tidyverse, ggplot2), Python (pandas, NumPy, scikit-learn), SQL, JupyterLab

Data & Systems: Bash scripting, Linux environments, on-premises compute clusters, AWS cloud services (EC2, S3, SageMaker)

Statistics: Hypothesis testing, regression modeling, variance component analysis, linearity assessment

Analytics & Product: Exploratory analysis, experiment design, cohort selection, product metrics

Operations & Collaboration: GitHub/Bitbucket, JIRA, Confluence, Agile workflows, cross-functional product development

Work History

Labcorp Integrated Oncology (Previously PGDx) – Baltimore, MD

Data Scientist II *09/2023 - Present*
Associate Data Scientist *06/2022 - 09/2023*
Bioinformatic Data Analyst *02/2021 - 06/2022*

- Designed QC metrics, analysis, and acceptance criteria for industry's first kitted pan-solid tumor liquid-biopsy test (*elio™ Plasma Focus Dx*)
- Owned 11 analytical verification studies (specificity, sensitivity, accuracy, precision, etc.) for FDA De Novo liquid biopsy submission, directly contributing to regulatory clearance and accelerated submission timelines
- Co-developed an internal sample database unifying metadata, QC flags, reagent lots, and procurement data, enabling cross-study analytics and improving operational decision-making
- Partnered with ML engineers to evaluate feature behavior, QC thresholds, and performance metrics for MRD signal detection models (*Labcorp® Plasma Detect™*)
- Led multiple large-scale regression, stress, and equivalency analyses across 4,000+ clinical samples to assess assay stability, detect performance drift, and inform pipeline release decisions
- Managed end-to-end analysis pipelines for multi-terabyte WGS-MRD datasets across multiple S3 buckets; created ETL workflows integrating variant calls, QC metrics, and run-level metadata
- Performed discrepancy investigations, statistical deep dives, and pipeline-change recommendations across assay versions and engineering releases, preventing recurring QC failures and reducing repeat sequencing events
- Developed statistical pipelines (probit, regression, PPA/NPA, APA/ANA) and distributional analyses for high-stakes regulatory studies
- Supported internal data processing for a clinical trial assessing potential benefit of tailoring immunotherapy treatment based on ctDNA molecular response in non-small cell lung cancer (NCT04093167)
- Assisted reimbursement (MolDx) submissions with curated sample-level analytics, annotations, and discrepancy summaries
- Partnered with R&D, engineering, regulatory, project management, and operations to meet system requirements and support customer-facing assay interpretation

MilliporeSigma, BioReliance – Rockville, MD

Associate Scientist *06/2020 - 02/2021*

- Conducted GMP/GLP-compliant mycoplasma detection assays within BSL-2 environment; maintained sterile cell-culture workflows and assay integrity

Education

Bachelor of Science: Biology, Salisbury University

Scientific Contributions:

- *Analytical Validation of PGDx elio™ Plasma Focus Dx to facilitate precision oncology through decentralized cell-free DNA solid tumor profiling*
- *Laboratory Automation of PGDx elio™ Tissue Complete on the Biomek NGeniuS System Enables Accurate and Reproducible Comprehensive Genomic Profiling*
- *Deep dive into the performance of structural variants reported by PGDx elio plasma focus Dx*