Development of Genomic Resource Modules for a Diverse Workforce

Alejandro Ochoa[1,2], Yuncheng Duan[3], Revathy Venukuttan[4], Shannon Clarke[1,2], Timothy Reddy[1,2,4]

[1]Duke Center for Statistical Genetics and Genomics, [2]Department of Biostatistics and Bioinformatics, [3]Department of Biology, and
[4]Center for Advanced Genomic Technologies, Duke University, Durham, NC 27705, USA

Motivation

The genomics workforce lacks diversity, and does not represent the US population. Building a diverse genomics workforce has enormous potential to improve research by fostering new ideas and approaches, and better representing the interests and motivations of the US population. One barrier to entry for genetics and genomics research is the high cost of data generation; and reducing or eliminating that barrier would enable more individuals from more diverse labs and institutions to contribute to genetics and genomics research.

Since the sequencing of the human genome, there has been a massive expansion in the amount of freely available genetics and genomics data. Making use of those datasets – for example ENCODE, GTEx, gnomAD, and genetic association results – has the potential to dramatically lower the cost of genetics and genomics research.

Goal

To contribute to building a diverse genetics and genomics workforce. Specifically, we aim to

- enable researchers from diverse labs
- to use public genetic and genomics resources
- to advance investigation of their research questions.

Examples:

- Predicting the effects of genetic variants on gene regulation
- Predicting how changes in gene regulation contribute to phenotypes and disease.

Target Audience, Partnerships

- Primary audience: researchers from greater Raleigh/Durham interested in genetics and genomics, wanting to use existing data.
- Focus on individuals from historically marginalized communities
  - Duke BioCoRE program for increasing diversity in the broad sense
  - Local Historically Black Colleges and Universities (HBCUs): NCCU, NC A&T (Fig. 1).
- Students with limited computation or programming experience.

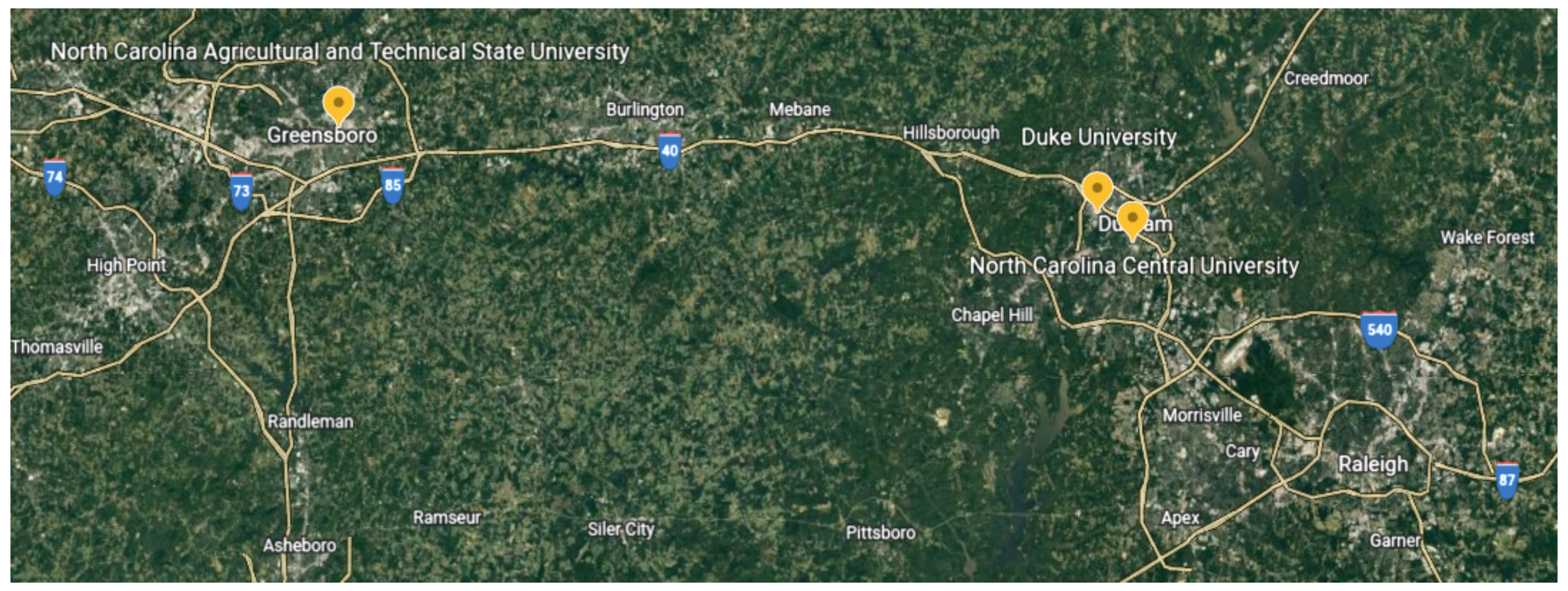

Figure 1:**Location of Duke and partner Historical Black Universities.**

Mode of Delivery

- Taught annually, in person or virtually
- Built upon Data Carpentry platform (Fig. 2)
  - Day divided into focused modules
  - Substantial time on hands-on practice
  - Common thread across workshop: in our case focus on two loci relevant to genetic disease in African ancestry (BCL11a and APOL1).
  - Students to lead workshop have taken "Data Carpentry for Genomics" course at Duke, taught by Hilmar Lapp and others.

Figure 2:**datacarpentry.org**

Key Components

- Highlight career pathways and partnerships with computational disciplines
- Including clinical and basic science endpoints, as well as individuals at various levels of education
- Identify mentors and initiative contributors that identify as individuals from historically marginalized communities
- Develop "next steps" including opportunities for ongoing engagement with network of researchers, internships, rotations.

Communication Strategy

- Partnering with NCCU/Duke Communication Summer Internship program to recruit interns.
- Primary goal: identify and build networks to reach key audiences and support the development of the genomic workforce, including scientific communication efforts.
- Tasks: developing a communication and networking plan, communication materials, for genomic modules project.
- Mentorship for intern by departmental communications team and the research program leadership.

Acknowledgments

NIH/NHGRI RM1HG011123, PIs Allen, Crawford, Gersbach, Kibbe, Reddy, Shashi, Wray.

Working Agenda

- Overarching focus on BCL11a enhancer and APOL1 locus
- Online — Overview of Human Genome — 1h
  - Goal for prework: students know how to get commonly used datasets for a locus of interest
  - Intro to Genome Browsers: led by Revathy Venukuttan
  - Gene Structure: led by Bill Majoros
  - Gene Expression Tracks (loading data live demo): led by Revathy Venukuttan
  - Group Exercise: led by Revathy Venukuttan
- Day 1 — Genetic Variation and Association — 3h
  - Genetic Association for Common Disease; How Variants Arise, Drift, etc.: led by Alex Ochoa
  - Group Exercise (processing association study data, visualize population structure): led by Yuncheng Duan
  - Consequences of Genetic Variation: led by Bill Majoros
  - Group Exercise (get VCF files and do simple regression): led by Apoorva Iyengar
- Day 2 — Genomic Analyses, Clinical Application and Motivation — 3h
  - Analyses Using High-Throughput Sequencing Data: led by Tim Reddy
  - Group exercise: led by Apoorva Iyengar
  - Clinical interpretation and current uses/room for growth: led by Makenzie Beaman
  - Bridging data generation, analyses, and clinical interpretation: led by Allison Ashley-Koch, Beth Hauser
  - Clinical collaborations and other areas to pursue investigation: led by Rasheed Gbadegesin, Gentzon Hall

Conclusions

- We are developing a program that highlights the role of bioinformatics in clinical research, exposing the field and opportunities.
- After workshop, connect people to internships to apply knowledge in real life applications, in their home institution.
- Long term plan to build partnerships with HBCUs, track alumni of program.
- Recruit students to Duke PhD programs, including Computational Biology and Bioinformatics (CBB) and the University Program in Genetics and Genomics (UPGG).
- Highlight endpoints other than PhD: Master's, industry, non-academic medicine.

DrAlexOchoa
🏠 ochoalab.github.io
✉ alejandro.ochoa@duke.edu