Center for Public Health Genomics: Purpose, Goals and Progress

Endometrial Biology Genomics Working Group
August 21, 2014

Carl D. Langefeld, Ph.D.
Director, Center for Public Health Genomics
Professor, Department of Biostatistical Sciences

Laurie P. Russell, M.S.
Administrative Manager, Center for Public Health Genomics
Project Manager, Department of Biostatistical Sciences

Center for Public Health Genomics
Wake Forest™ School of Medicine

Purpose

- The Wake Forest School of Medicine's strategic plan identifies several target areas of research and all of these areas have a strong genomic component – albeit, some underdeveloped.
- The purpose of the Center for Public Health Genomics is to encourage and facilitate research on the genomics of diseases and traits of significant personal and public health burden, including institutional strategic areas.
- The Center's focus is primarily on the quantitative and population-based aspects of genomics research and is complementary and synergistic to other genomics and institutional efforts.

Motivation

- The very rapid pace of discovery and evolution of the field of public health genomics has been technology driven.
  - The era of genome-wide association studies is past its prime.
  - The field is transitioning to whole genome sequencing technologies for genotypes, structural variation, epigenetics and gene expression profiling.
  - The release of the Encyclopedia of DNA Elements (ENCODE) Project has provided paradigm-shifting information that has altered our view of the genome and the impact of regulatory elements.
  - The coming era of integrative genomics is here.
- With this rapid pace of change comes an opportunity for fundamental discoveries with translational impact.
- Wake Forest needs to keep pace with these changes or it will not be as competitive in many of the institutional strategic plan areas.
  - To be competitive, Dean Abraham notes that we need to be preeminent in key areas
  - To enable researchers across the entire Medical Center to reach that target, we need to be preeminent in the quantitative aspects of genomics, pharmacogenomics (e.g., pharmaco-epigenomics) and personalized medicine.

Intent

- The Center will contribute to the institutional mission by connecting researchers, encouraging multi-disciplinary collaborations and developing statistical and bioinformatic infrastructure for “Next Generation” genomic / omic questions.
- Output from this Center is intended to be an institutional resource with equal access and inclusion for anyone, regardless of being a formal member.

Some of the Previous Contributions

1. The center invested $155,807 in DEAC computing (compute nodes and storage)
   - Augmented by >$94,000 from individual grants from Dr. Langefeld.
   - Equally available to all faculty, staff and students at both the Medical Center and Reynolda campuses.
   - Modest investments in novel computing technologies.
2. Develop a number of statistical genetics and bioinformatics software (SNPDoc, Dandelion, SNPGWA, QSNPGWA, SNPLash, Dprime), in response to needs of researchers from the Center for Genomics and Personalized Medicine Research and Diabetes Center, Primate Center, cardiology, Regenerative Medicine and in anticipation of emerging technologies (e.g., genome-wide association studies, epigenetic studies).
3. Cost-shared on external software (Ingenuity) with the library and specialty software (e.g., Cancer Center, Diabetes Center).
4. Wei Wang (Bioinformatician), Paula Ramos (Post-doc) and Carl Langefeld visited several departments and centers to teach/facilitate the use of bioinformatic and statistical genetics software.
5. Helped other investigators download publicly available genetic data on dbGaP and provided guidance on how to integrate and analyze.
6. Sponsored and managed the Statistical Genetics, Genetic Epidemiology and Bioinformatics Seminar Series Club.
7. Co-sponsored and financially contributed to the annual Statistical Genetics conference held in the research triangle.

The Center's Three Specific Aims.

- Aim 1: Develop synergies and collaborations through sponsoring
  - Open working groups whose efforts center on the intersection of public health, genomics and a content area (e.g., ethics and policy, cancer, hypertension, obesity, diabetes and metabolism, cardiovascular disease, personalized medicine, aging/cognition).
  - Statistical Genetics, Genetic Epidemiology and Bioinformatics Seminar Series and the annual Statistical Genetics conference held in the research triangle.
- Aim 2: Sponsor two $20,000 Pilot Project RFAs (first and third year), where the intent is to develop material (e.g., data, theory) for subsequent NIH grant proposals.
  - Would like to work with other centers to co-sponsor projects.
- Aim 3 Develop statistical and bioinformatic software and analysis pipelines for sequence data, expression quantitative trait locus analyses, epigenetic data and summarization tools.
  - These tools should reduce time and cost of analysis while improving the quality of base analyses, provide current "best practices" algorithms and insert novel analytic methods.
  - Importantly, the tools will also make genomic grant applications more competitive as it will provide documentable evidence of expertise, experience, institutional support, strong research environment and analytic innovation.

Statistical Genetics, Genetic Epidemiology and Bioinformatics e-Seminar Series

- Continue to sponsor and manage the seminar series
- Continue the GoToMeeting format
  - Three Wake campuses
  - External – UNC, NC State, Duke, Medical University of South Carolina, University of Colorado School of Public Health, National Jewish Health, Central Michigan University.
- Recent past speakers include:
  - Matt Nelson (GSK), Beth Hauser (Duke, Director of Genomics Center), Ethan Lange and Leslie Lange (UNC)
- Are there researchers here at Wake or at neighboring institutions that you would like to hear a talk consistent with a public health genomics theme?

Co-Sponsorship of Statistical Genetics Conference Held in Research Triangle

2013 Triangle Statistical Genetics Conference
Monday, October 7, 8:30-4:00
Thomas Center, Duke University, Durham

8:30 **Check in and continental breakfast**

9:00 ***Welcome***
Andrew Allen, Associate Professor, Biostatistics and Bioinformatics, Duke

Session 1: Chaired by Andrew Allen
9:05 **Tim Reddy**, Assistant Professor, Biostatistics & Bioinformatics, Duke
Genetics and genomics of gene expression
9:30 **Yun Li**, Assistant Professor, Biostatistics, UNC
The 1000 Genomes Project
9:55 **Fred Wright**, Professor, Statistics and Biological Sciences, NCSU
The Genotype Tissue Expression Project
10:20 **David Aylor**, Assistant Professor, Biological Sciences, NCSU
Experimental designs for mammalian systems genetics

10:45 ***Break, Poster Session*** (Kirby Reading Room)

Session 2: Chaired by Langefeld, Professor, Biostatistical Sciences, WFU
11:15 **Paula Ramos**, Medical University of South Carolina
Identification of regions under positive selection in the Gullah African American population of South Carolina
11:30 **Yu Jiang**, Computational Biology & Bioinformatics, Duke
Utilizing population controls in rare-variant case-parent association tests
11:45 **Jin Szatkiewicz**, Research Assistant Professor, Genetics, UNC
Improving detection of copy number variation by simultaneous bias correction and read-depth segmentation
12:00 **Clementina Davenport**, PhD Candidate, NCSU
Joint modeling of gene expression and copy number variation data using functional regression models

12:15 ***Lunch (provided)***

Special thanks to our meeting sponsors

Keynote Speaker: Chaired by Matt Nelson, Principal Scientist, GSK
1:15 **David Goldstein**, Distinguished Professor, School of Medicine, Duke
Identifying pathogenic mutations in patients with neurodevelopmental diseases

Session 3: Chaired by Yun Li, Assistant Professor, Biostatistics, UNC
2:00 **Ethan Lange**, Associate Professor, Genetics & Biostatistics, UNC
Identification of the first high-penetrant prostate cancer gene using next generation sequencing: Lessons learned
2:15 **Kouros Owzar**, Associate Professor, Biostatistics & Bioinformatics, Duke
Drug Induced Toxicity Phenotypes in Cancer Pharmacogenomics
2:30 **Zhao-Bang Zeng**, Distinguished Professor, Bioinformatics Research Center, Statistics & Genetics, NCSU
Study of interaction of quantitative trait loci

2:45 ***Afternoon break***

Session 4: Chaired by Jung-Ying Tzeng, Associate Professor, Statistics, NCSU
3:00 **Matt Nelson**, Principal Investigator, GSK
Impact of genetic association on drug discovery success
3:15 **Karen Mohlke**, Associate Professor, Genetics, UNC
Low frequency variants for insulin processing and secretion
3:30 **Hua Zhou**, Assistant Professor, Statistics, NCSU
Fast Genome-Wide QTL Association Mapping on General Pedigrees with an Application to Longitudinal HDL Data
3:45 **David Reif**, Associate Professor, Bioinformatics Research Center & Biological Sciences, NCSU
High-throughput characterization of behavioral and developmental zebrafish phenotypes in response to chemical exposure: Setting the stage for GxE studies

4:00 ***Concluding Remarks***
Fred Wright, Professor, Statistics and Biological Sciences, NCSU

Conference Organizing Committee
Matt Nelson (Chair, GSK)
Andrew Allen (Duke)
Carl Langefeld (WFU)
Yun Li (UNC)
Jung-Ying Tzeng (NCSU)
Fred Wright (NCSU)

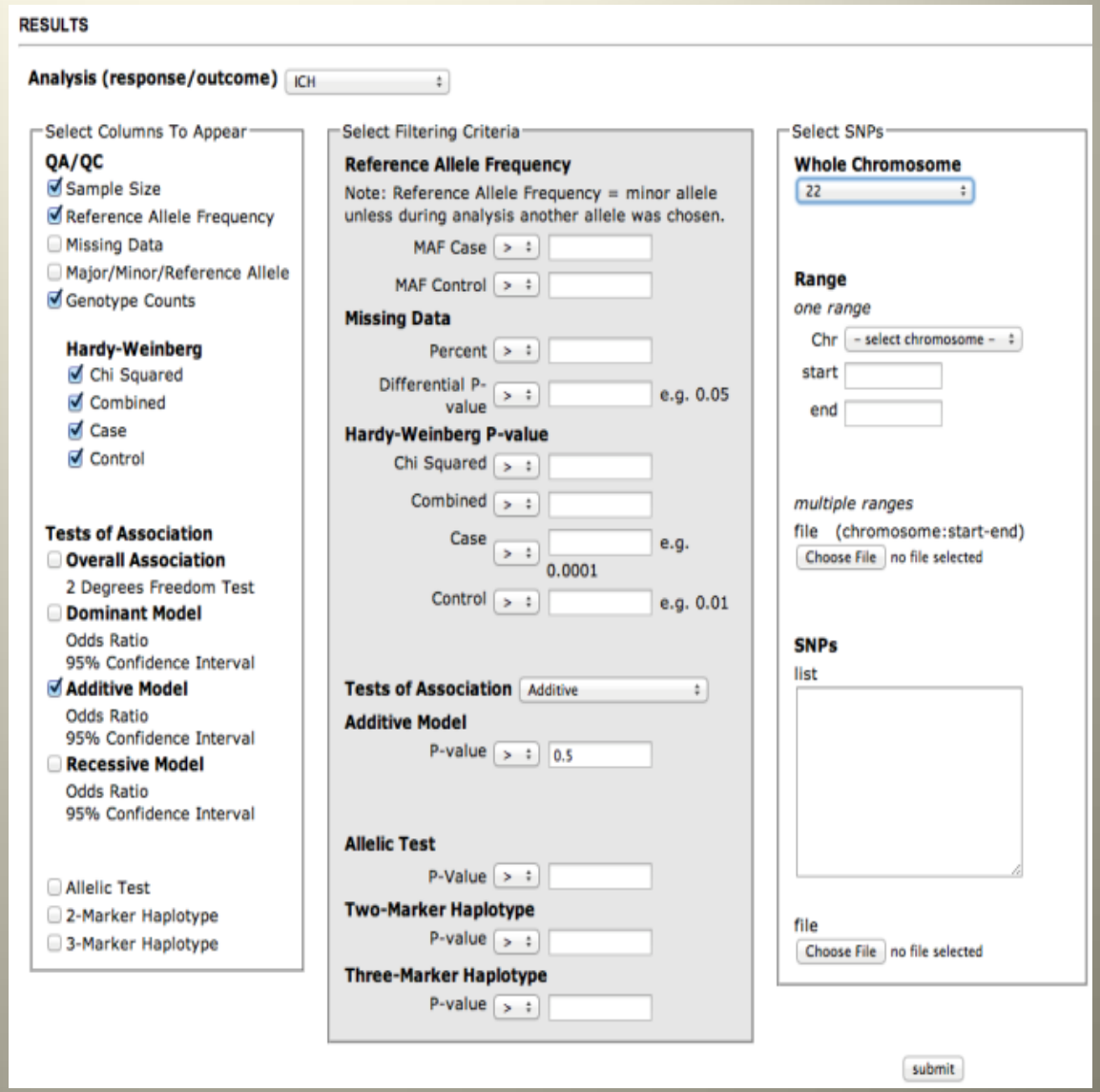
Sequencing Technologies

- David McWilliams, Jeff Chou, Greg Hawkins, and Carl Langefeld
  - DNA-seq analysis pipeline (Life Technologies, SOLiD)
  - Evaluating several aligners and genotype callers
- Developing sequencing pipeline for the Ion Torrent (Life Technologies) for cancer patients
    - Pat Koty, David McWilliams, Jeff Chou, Carl Langefeld
    - Goal: sequence normal and tumor DNA for key polymorphisms to inform personalized medicine
    - Could also result in a potential research biobank
    - Will expand to other diseases once "proof of concept" is complete – easy expansion if deemed helpful for patient care and is cost effective
- Jeff Chou, David McWilliams, Lance Miller and Carl Langefeld
  - Launching effort for RNA-seq analysis pipeline
  - Similar goals to the DNA sequencing above
- John Beal has been making progress on updating bioinformatic software such as SNPDoc
  - Updating for functional implications, integrating gene expression data
  - Design for sister program for NextGen sequencing technologies

Results Retrieval System

John Beal, Carl Langefeld, Mindy Marion, Julie Ziegler

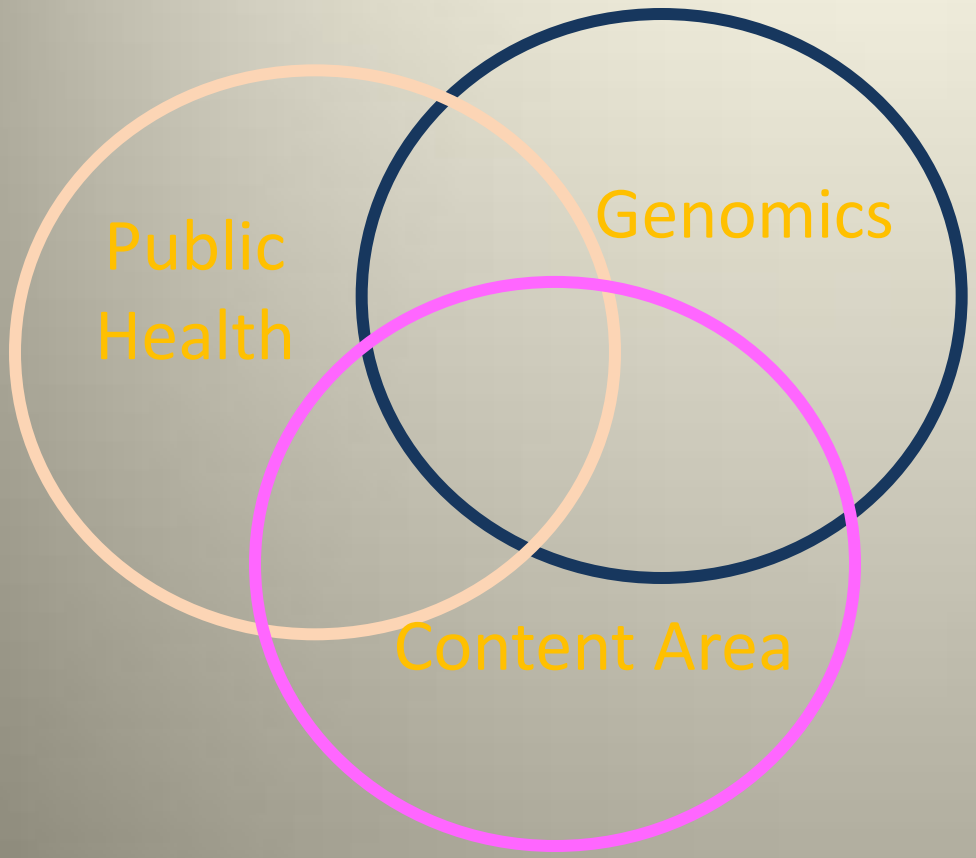
- As the volume of genomic data increases the ability to place results in an investigator-friendly retrieval system is important.
- With programming from John Beal, we have developed a prototype Results Retrieval System
  - Stored in SQL Server database
  - Results returned in an excel spreadsheet
- Goal: using a secure server or potentially on individual machines a system that allows quick look ups of data
  - Saves investigator time
  - Saves analyst time
  - Bit of a “storage hog”



Center for Public Health Genomics Pilot Grants

- Between the fall and spring RFA releases, we received six applications
  - Mini-NIH review process
  - Doctoral students and post-doctoral fellows allowed to review grants and observe process but their scores did not count in the final ranking
- Thank you to Jasmin Divers and Laurie Russell for shepherding the reviewing of the applications
- Thank you to the faculty who reviewed the grants
  - Nick Pajewski, Greg Hawkins, Fang-Chi Hsu, and Jasmin Divers (Chair)
- One grant awarded
  - *Genetic investigations into adiponectin as a biologic mediator for cardio-metabolic health in African Americans*
  - PI's: Nichole (Palmer) Allred and Swapan Das

Working Group Idea



The purpose of the Working Group is to create an environment to bring researchers together from across departments and institutions to exchange ideas, connect with other researchers, develop collaborations and identify future opportunities within public health and genomics.

Working Groups

- Develop regular meeting Working Groups to advance the exchange of ideas, connecting researchers, increase collaborations and identify research opportunities
  - Institution wide (both campuses)
  - UNC, UNCG, WSSU, Glaxo-Smith-Kline, others as technology permit

- Hope to develop about 6 groups that meet twice a year for a couple of hours for presentations, interaction, discussion
- Goal
  - Provide contacts and mentoring for junior faculty, postdocs
  - Spawn ideas and teams for R01, P01, etc.

Current Members

Faculty

- Barry Freedman
- Fang-Chi Hsu
- Greg Hawkins
- Tim Howard
- Lijun Ma
- Lynne Wagenknecht
- Maggie Ng
- Nichole Allred
- Nick Pajewski
- Nancy King
- Pete Santago
- Lance Miller
- Steve Walker
- Carl Langefeld

Staff

- Laurie Russell
- David McWilliams
- Jeff Chou
- Adrienne Williams
- Julie Ziegler
- Leora Henkin
- Lingyi Liu
- Neeraj Sharma
- Mark Brown
- Mary Comeau
- Mindy Marion

Students and post-doctoral fellows

- Lindsay Reynolds
- Hannah Ainsworth
- Satria Sajuthi
- Chuan Gao

Genomic Analytic Expertise

- Jeff Chou, PhD
  - Bioinformatician, Public Health Genomics and Department of Biostatistical Sciences
  - Many years of gene expression analysis
  - Multiple programming languages (JAVA, R, SAS, C++)
  - Currently working on sequence analysis pipeline
- David McWilliams, PhD
  - Bioinformatician , Public Health Genomics and Department of Biostatistical Sciences
  - Chemistry and Proteomics background
  - Multiple programming languages (R, PERL)
  - Currently working on sequencing analysis pipeline
- Carl Langefeld, PhD
  - Biostatistician, Professor in Department of Biostatistical Sciences
  - Director, Center for Public Health Genomics
  - Genome-wide association, epigenetics
  - Statistical methods for gene-gene, gene-environment interactions, classic statistical methods and machine learning methods
  - Currently supervising development of multiple analysis pipelines
- Jasmin Divers, PhD
  - Biostatistician, Professor in Department of Biostatistical Sciences
  - Genetic association
  - APOL1 and ESRD major focus
- Nicholas Pajewski, PhD
  - Biostatistician, Professor in Department of Biostatistical Sciences
  - Genetic analysis
- Others
  - Miranda Marion, MA, Biostatistician
  - Adrienne Williams, MA, Biostatistician
  - Laurie Russell, MS, Project Manager

Open Discussion on Potential Areas of Collaboration

Thank you for your time and interest.

Carl D. Langefeld: clangefe@wakehealth.edu
Laurie P. Russell: lrussell@wakehealth.edu

Discussion

Thank you for your time and interest.

Carl D. Langefeld: clangefe@wakehealth.edu
Laurie P. Russell: lrussell@wakehealth.edu