Human Genomics Analysis Interface for FaceBase

Mary L. Marazita

FaceBase Scientific Meeting

January 8-9, 2015

Project ScientificTeam

- Mary Marazita (PI)
- Eleanor Feingold (co-I)
- Harry Hochheiser (co-I)
- Elizabeth Leslie (post-doc)

Human Genomics Analysis Interface

- Goal 1: software interface to explore human craniofacial genomics databases
- Goal 2: identify pertinent genomics data to incorporate
- Goal 3: create analysis results datasets to make available on FaceBase
- NOTE: individual level data will not be distributed

Year 1, focus on Goals 1 and 2

- Develop a software interface that will enable FB users to apply human genetics analysis software (e.g. PLINK) to human genomics data from craniofacial research, with access to these tools through the FB Hub
- Identify data to include

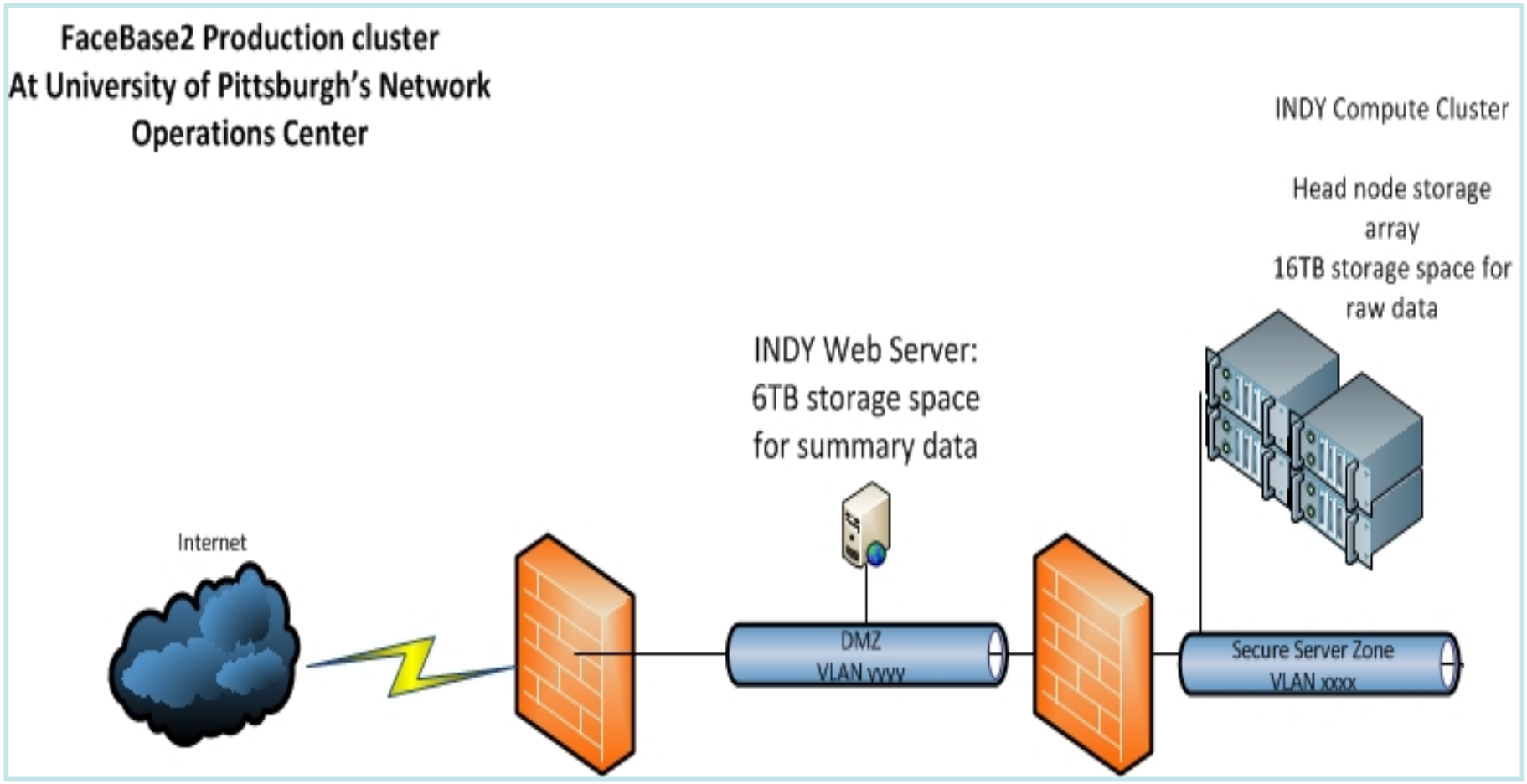
Notes

- the data for analysis will be located on a secure location at Pitt
- no individual level data will be available to users through FB (but can be obtained)
- purpose of these tools:
  - allow FB users to explore genomics data, or ask initial research questions
  - And/or to decide whether to download the underlying individual data from dbGaP

GOAL 1 milestones, Year 1

- Develop secure network infrastructure

Network Infrastructure



Infrastructure notes

- Web server (for users) contains only summary data files and images
  - User access only to visualizations or results datasets premade or created dynamically
- Analysis (raw files): head node in secure server zone (FISMA compliant) that only designated network workstations have access to
  - Necessary analysis software is installed and executed here

GOAL 1 milestones, Year 1 (cont)

- Develop secure network infrastructure

- Develop draft web page
- Implement PLINK (other packages in later years)
- Format available genomics data
  - Started with Beaty GWAS dataset

GOAL 1 milestones, Year 1 (cont)

- Create results databases with PLINK
- Develop displays of results (eg Manhattan plots, LocusZoom)

DEMO

- Beaty data

Goal 2: Analysis Datasets-status

- Beaty/Murray/Marazita (OFC: GWAS, sequencing) ←FB1
- Marazita (OFC: Guatemala GWAS)
- Spritz (facial variation in Tanzanians: GWAS)←FB1
- Weinberg/Marazita/Spritz (facial variation in Caucasians: GWAS)←FB1
- Murray/Marazita/Leslie (OFC: targeted seq)
- Marazita/Murray et al (OFC: GWAS)

Year 1 milestones to do

- Refine analysis tool
  - Acknowledgements, analysis details, links to dbGaP, crosslink with other FB projects on Hub
- Implement custom calculations
  - E.g. allow users to specify SNPs or chromosomal regions
- Better annotate/tag datasets for custom use and crosslink with FB Hub

Year 1, with FB Hub and SC

- Design, annotation, metadata
  - Before we finalize
- Implementation plan with FB Hub
  - Crosslink with other FB projects on Hub
  - Eg TDFN/Tanzania, but also eg animal researchers who want to see human results
- Data: embargo issues
- User access—registered only?