NHGRI Bioinformatics Program

Vivien Bonazzi

Putting the Byte into Biology

Bioinformatics

The application of computing technologies to biological information

- Development of new methods and algorithms
- Use of existing methods
- Development of LIMS
- Processing of biomedical data
- Development of analysis pipelines and environments
- Curation, collation, and representation of biomedical data

NHGRI Bioinformatics Program *(version 0.1)*

- Bioinformatics Grants Portfolio
- Other Bioinformatics Staff Activities
  - Bioinformatics Consulting Services
  - Trans NIH Bioinformatics Activities

NHGRI Bioinformatics Program *(version 0.1)*

- **Bioinformatics Grants Portfolio**

- Other Bioinformatics Staff Activities
  - Bioinformatics Consulting Services
  - Trans NIH Bioinformatics Activities

Bioinformatics Program

- Managed by many NHGRI Program Directors
  - Vivien Bonazzi
  - Lisa Brooks
  - Mike Pazin
  - Erin Ramos
  - Adam Felsenfeld
  - Heidi Sofia
  - Elise Feingold
  - Lucia Hindorff
  - Jeff Schloss
  - Jeff Struewing
  - Anastasia Wise
  - Tina Gatlin
  - Bettie Graham

Types of Bioinformatics Grants

- **Primary Bioinformatics**
  - Model Organism Databases (MODs)
  - Enabling Bioinformatics Technologies
  - Variation Analysis and Association
  - Grants related to other biological areas
    - Gene Regulation, NGS Data Analysis, Genome Annotation, Clinical Informatics, Gene Expression, Networks, Pathways and Systems, Biomedical Ontologies
  - DACC (Data Analysis and Coordination Centers)

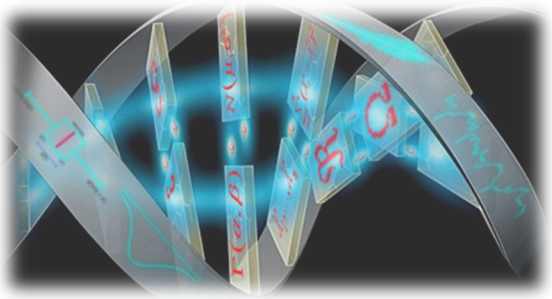
- **Component Bioinformatics**
  - Where the computational component is NOT the major focus of the grant

Types of Bioinformatics Grants

- **Primary Bioinformatics**
  - **Model Organism Databases (MODs)**
  - **Enabling Bioinformatics Technologies**
  - Variation Analysis and Association
  - Grants related to other biological areas
    - Gene Regulation, NGS Data Analysis, Genome Annotation, Clinical Informatics, Gene Expression, Networks, Pathways and Systems, Biomedical Ontologies
  - DACC (Data Analysis and Coordination Centers)
- Component Bioinformatics
  - Where the computational component is NOT the major focus of the grant

Types of Bioinformatics Grants

- Model Organism Databases (MODs)
- Enabling Bioinformatics Technologies
- Large Scale Informatics Grants
- U41, P41, U01, P01



Model Organism Databases (MODs)

- Highly curated, integrated data set of genes, proteins, variants, alleles, phenotypes, genetic disorders, etc
- Provide a common nomenclature for gene symbols, anatomic terms
- Search, analysis and viz tools to explore data

ZFIN

FlyBase

MGI

WormBase

RGD

OMIM®
Online Mendelian Inheritance in Man®

SGD Saccharomyces Genome Database

PO: Vivien Bonazzi

Enabling Bioinformatics Technologies

Enabling easier access, use, integration, visualization, or management of genomic data and tools

POs
Vivien Bonazzi
Lisa Brooks
Adam Felsenfeld
Lucia Hindorff
Erin Ramos
Heidi Sofia

Types of Bioinformatics Grants

- **Primary Bioinformatics**
  - Model Organism Databases (MODs)
  - Enabling Bioinformatics Technologies
  - **Variation Analysis and Association**
  - Grants related to other biological areas
    - Gene Regulation, NGS Data Analysis, Genome Annotation, Clinical Informatics, Gene Expression, Networks, Pathways and Systems, Biomedical Ontologies
  - DACC (Data Analysis and Coordination Centers)
- Component Bioinformatics
  - Where the computational component is NOT the major focus of the grant

Variation Analysis and Association

Development of methods for analyzing genetic variation and association studies

- Discovery of genetic variation
- Association and linkage studies
- Population genetic analyses
- Admixture analyses
- Human genetic variation databases

POs
Lisa Brooks
Erin Ramos

Types of Bioinformatics Grants

- **Primary Bioinformatics**
  - Model Organism Databases (MODs)
  - Enabling Bioinformatics Technologies
  - Variation Analysis and Association
  - **Grants related to other biological areas**
    - **Gene Regulation, NGS Data Analysis, Genome Annotation, Clinical Informatics, Gene Expression, Networks, Pathways and Systems, Biomedical Ontologies**
  - DACC (Data Analysis and Coordination Centers)
- Component Bioinformatics
  - Where the computational component is NOT the major focus of the grant

Grants: other biological areas

- Gene Regulation
- NGS Data Analysis
- Genome Annotation
- Gene Expression
- Networks, Pathways and Systems
- Biomedical Ontologies
- Bioinformatics Training
- Mostly Smaller Scale grants:
  - R01, R21, R25, R42,43/44, K25, K99
  - Some U01

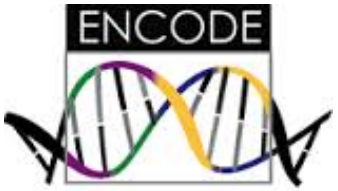
POs
Vivien Bonazzi
Lisa Brooks
Elise Feingold
Adam Felsenfeld
Tina Gatlin
Mike Pazin
Jeff Schloss
Heidi Sofia
Jeff Struewing

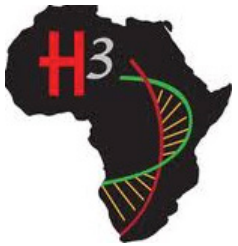
Types of Bioinformatics Grants

- **Primary Bioinformatics**
  - Model Organism Databases (MODs)
  - Enabling Bioinformatics Technologies
  - Variation Analysis and Association
  - Grants related to other biological areas
    - Gene Regulation, NGS Data Analysis, Genome Annotation, Clinical Informatics, Gene Expression, Networks, Pathways and Systems, Biomedical Ontologies
  - **DACC (Data Analysis and Coordination Centers)**
- Component Bioinformatics
  - Where the computational component is NOT the major focus of the grant

DACCs
Data Analysis and Coordination Centers

- Provide bioinformatics support services to specific projects
  - Development of data pipelines
  - Development of metadata standards
  - Storage of project data raw and derived
  - Deposition of data to NCBI or EBI
  - Provision of a web site for easy access to data, SOPs, lab protocols, tools etc.
  - Significant portion of staff time
- NHGRI provides total or partial funding

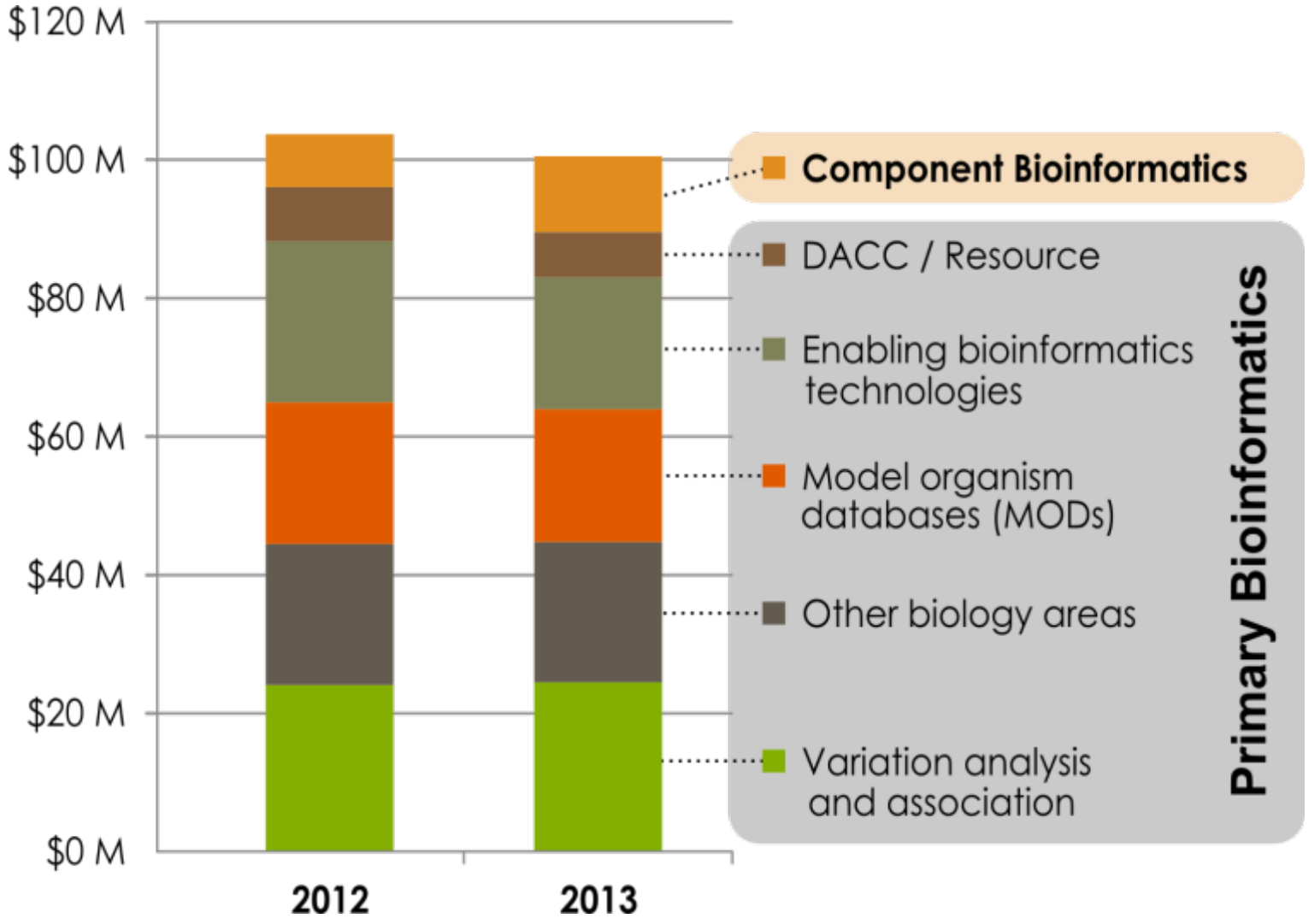




Grant Funding totals

	2012	2013
Primary Bioinformatics		
Variation analysis and association	$24.1 M	$24.5 M
Enabling Bioinformatics Technologies	$23.3 M	$19.1 M
MODs	$20.5 M	$19.2 M
Other biology categories	$20.4 M	$20.3 M
DACC / Resource	$7.9 M	$6.5 M
Subtotal for Primary Bioinformatics	$96.1 M	$89.6 M
Component Bioinformatics	$7.6 M	$11.0 M
Total	$103.7 M	$100.6 M

Grant Funding totals

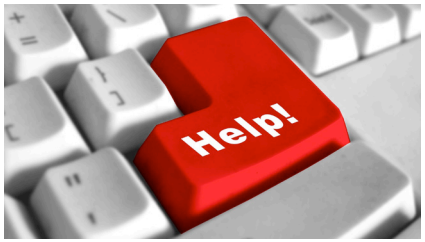


NHGRI Bioinformatics Program

- Bioinformatics Grants Portfolio

- **Other Bioinformatics Staff Activities**
  - Bioinformatics Consulting Services
  - *Trans NIH Bioinformatics Activities*

Bioinformatics Consulting Services to NHGRI staff



- Technical advice on grants that include computing components
- Provide an understanding and connectivity with NCBI/EBI resources and people
- Data submission to SRA/dbGAP (NCBI)
- Centers for Mendelian Genomics (CMG)
  - data metadata structure and submission to NCBI

NHGRI Bioinformatics Program

- Bioinformatics Grants Portfolio
- **Other Bioinformatics Staff Activities**
  - Bioinformatics Services
  - **Trans NIH Bioinformatics Activities**
    - **Interactions with NCBI**
    - **Projects with other ICs**
    - **BISTI**
    - **NIH – NSF Big Data initiative**
    - **BD2K**

Trans NIH Bioinformatics Activities

- NCBI
  - Tracking sequence data and metadata deposition from NHGRI sequencing programs
  - BioProject pages for various NHGRI programs
  - NCBI data/tools **on** the cloud (Amazon and Google)
  - **ClinVar /dbSNP**

Trans NIH Bioinformatics Activities

- BISTI Biomedical Information Science and Technology Initiative
  Your source for biomedical informatics at NIH

  - Consortium of representatives from each of the NIH institutes and centers.
  - Focuses on biomedical computing issues at NIH
  - Developed broad-based FOAs for computational biology -R01, R21, R42, R43, R44

Trans NIH Bioinformatics Activities

- NIH – NSF Big Data Initiative 2012

- Joint solicitation between NIH & NSF
  - NCI, NIGMS, NHGRI, NIBIB, NLM, NIDA, NINDS
  - Core Techniques and Technologies for Advancing Big Data Science & Engineering
- Develop and evaluate core technologies and tools that take advantage of available collections of large data sets to accelerate progress in science, biomedical research, and engineering.

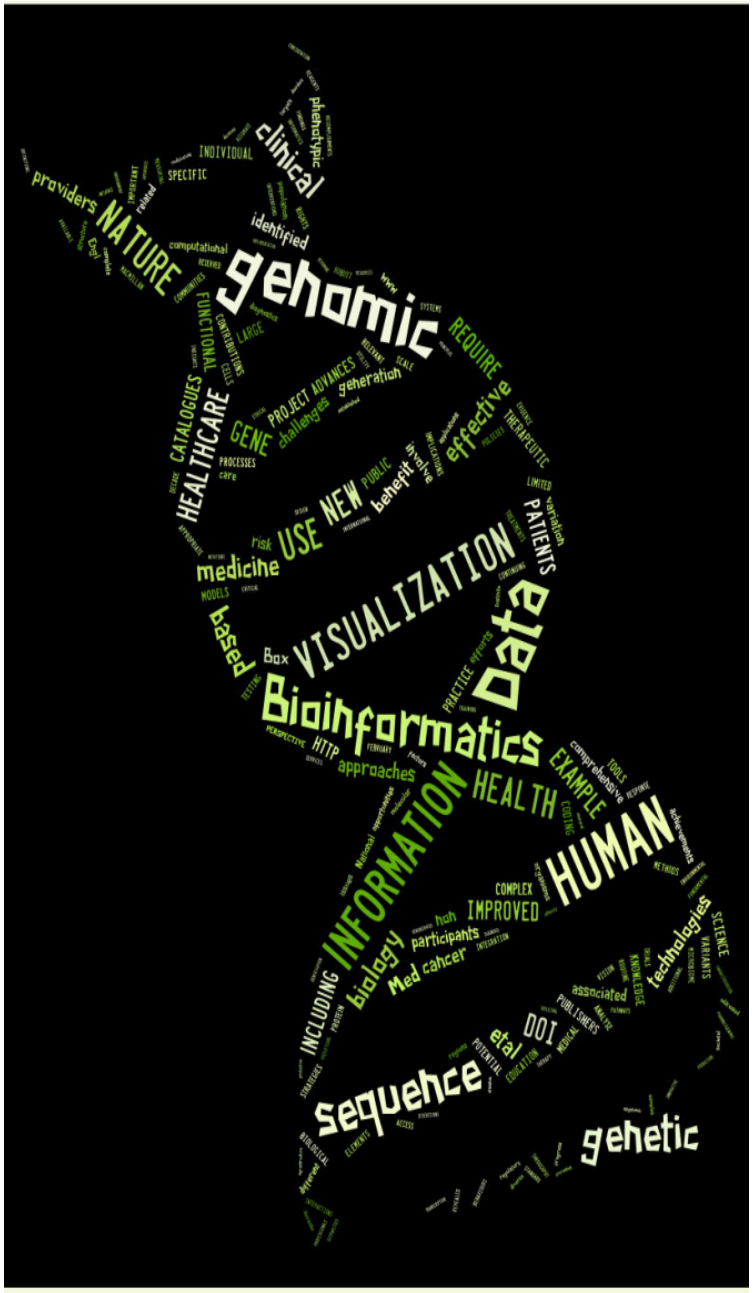
Trans NIH Bioinformatics Activities

- BD2K
  - Big Data to Knowledge Initiative
    - Data sharing and standards
    - Software methods and systems for analysis
    - Training
    - Centers of Excellence for Big Data Analysis
  - **Eric Green**, Mark Guyer, Vivien Bonazzi, Lisa Brooks, Ajay **Pillai**, Heidi Sofia, Leigh Finnegan Chris Wellington

Acknowledgements

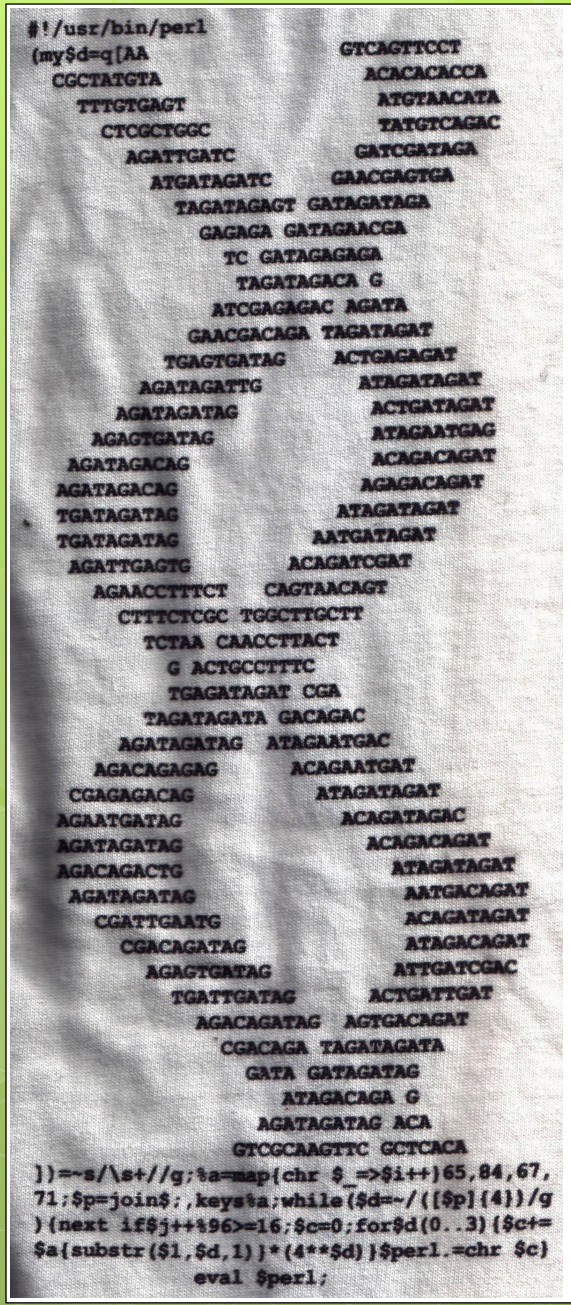
Bioinformatics Advisory Group to NACHGR

- A small group of Council members with bioinformatics experience, and skills from development to end user
- Provides advice on bioinformatics issues that **affect** the NHGRI Bioinformatics program
  - Carlos **Bustamante**
  - Jill Mesirov
  - Bob Nussbaum
  - Lucila Ohno-Machado



Acknowledgements

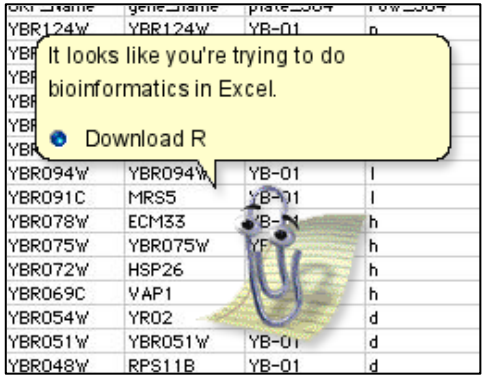
Lisa Brooks
Elise Feingold
Adam Felsenfeld
Zivile Goldner
Peter Good
Lucia Hindorff
Mike Pazin
Erin Ramos
Jeff Schloss
Mike Smith
Heidi Sofia

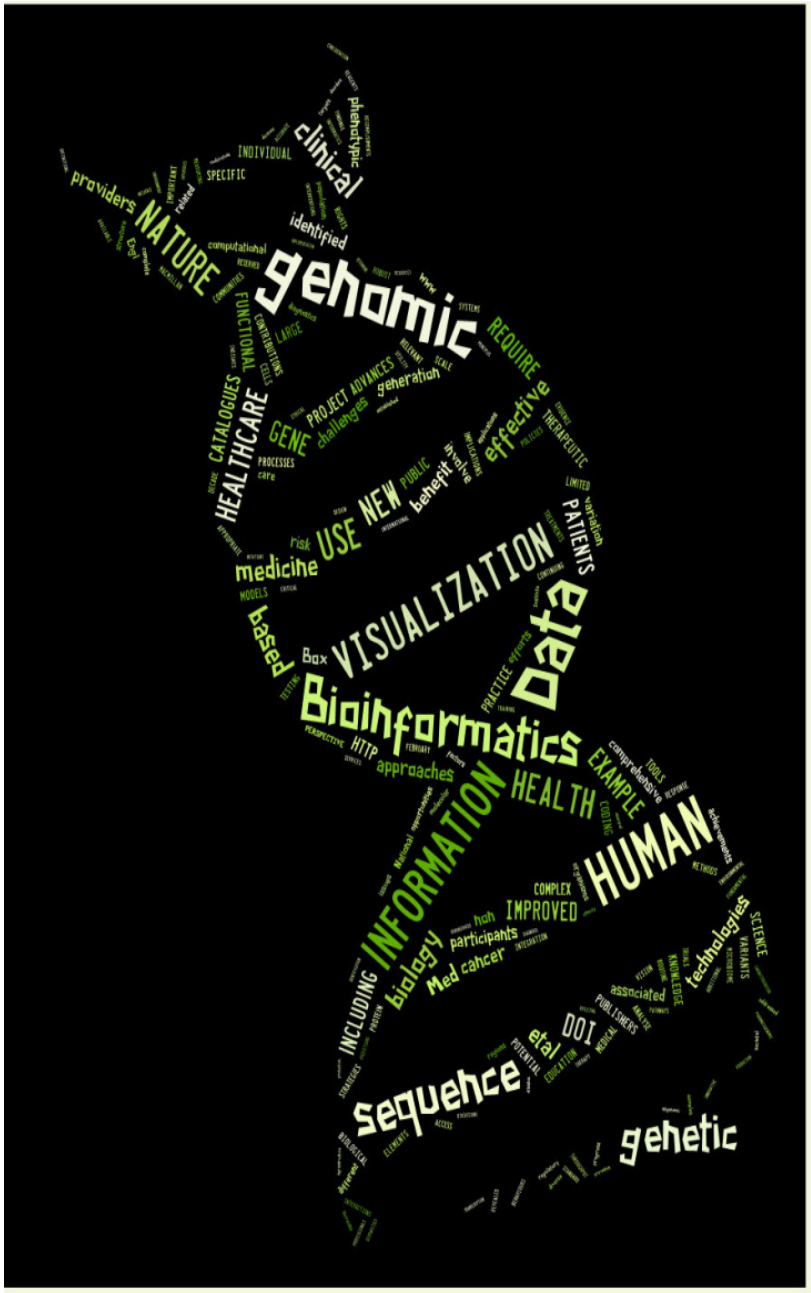


Acknowledgements

A special thanks to:

Chris Wellington




clinical
phenotypic
INDIVIDUAL
SPECIFIC
providers
NATURE
identified
computational
genomic
REQUIRE
FUNCTIONAL
CATALOGUES
HEALTHCARE
GENE
PROJECT
ADVANCES
challenges
generation
effective
THERAPEUTIC
PUBLIC
benefit
NEW
USE
risk
medicine
VISUALIZATION
PATIENTS
Data
based
Box
Bioinformatics
HTTP
approaches
EXAMPLE
HEALTH
INFORMATION
HUMAN
COMPLEX
IMPROVED
participants
Med cancer
biology
INCLUDING
technologies
SCIENCE
associated
DOI
et al
sequence
genetic

END