Functional Annotation of ANimal Genomes (FAANG)--
Global FAANG Consortium

Organic growth of FAANG during 2014-2019
Current FAANG contributors= 432

2014 membership- GB authors

March 2019 membership
Outline – Introduction to FAANG Consortium

1. Why FAANG?
2. ENCODE results as a FAANG goal
3. Brief history and goals
4. FAANG Committees and activities
5. Summary of current FAANG projects
6. FAANG data available→ getting involved
Why is FAANG important?

- Understanding the genotype to phenotype link:
- Providing information to accelerate genetic improvement
- Improving fundamental understanding of biology
- We need to *annotate the function* of the genome of our animals
What is functional annotation?

- We have the DNA sequence for the genome of our major domesticated animals...

This is not English!

What is important information and what isn’t??
Annotation can help:

- Find what part codes for proteins
- Understand effects of genetic variation on these proteins
Why is functional annotation important?

Is trait-associated genetic variation located (enriched) in genes?

Yes!

But also enriched in regions flanking/outside known coding regions!

GWAS SNPs significantly enriched in protein-coding regions.

The genomic signature of trait-associated variants

Alida S D Kindt, Pau Navarro, Colin A M Semple and Chris S Haley

BMC Genomics 2013 14:108
Shown in livestock as well: Milk trait-associated SNPs enriched in predicted regulatory regions

Many sig SNPs in bovine enhancer regions predicted from mapping ENCODE to bovine genome

Fat yield

Milk yield

protein yield

Much fewer sig SNPs in random sampling

DOI 10.1186/s12711-017-0331-4
Why is functional annotation important?

- SNPs associated with disease/traits are often outside genes i.e., Maurano et al. 2012 (human), Wang et al. 2017 (bovine)
- This says the function of genomes are not just in proteins...
- So we must understand function of “non-genic” regions of genome!

- But the genome is mostly “empty”- only 1.5% is protein-coding- which is the easiest parts to find and annotate...
How do we figure out what parts are important then?

**Protein-coding genes**
- Regulatory region - “ON” in skeletal muscle
- Regulatory region - “ON” in mammary gland
- Regulation - ???

SNPs associated with muscle growth

---

**FAANG**
Functional Annotation of Animal Genomes

**Iowa State University**
Of Science and Technology
Why is functional annotation important?

• If we want to be able to better predict trait value from understanding the importance of associated SNPs, we need to find all the functional parts of animal genomes

• “Predicting phenotype from genotype”…

• How to get there?
Genome functional annotation described by ENCODE

To understand function:

a) what part is expressed?

b) what part controls this expression (specific tissue, response to infection, etc):

RNA levels and location

Biochemical assays of Chromatin structure
Roadmap Epigenome:
Extending ENCODE to tissues → many biological states

More biological states increases power to identify specific functional regions
Collecting all these data on the same samples allowed **predictive models for genome function** to be developed.

- Predicts **Chromatin State**: OPEN or CLOSED
- Created a **functional map** of genome segments
Success in ENCODE and Roadmap required:

- High quality reference genome sequence
- Common infrastructure providing
  - Biological resources
  - Bioinformatics tools
  - Databases
- Effective coordination and communication

FAANG needs all these to succeed!
Brief FAANG history

• Workshop in January 2014
• Organizational meeting in Jan 2015 in San Diego
• White paper published in 2015—“Highly accessed”
  • > 9,400 accesses as of March 2019, 82 citations
• 2015 Workshop to kick off FAANG

Giuffra, Tuggle, et al., Annual Reviews in Animal Biosciences
February 2019
Gathering On FAANG Workshop October 7-8, 2015
National Academy of Sciences Building
Washington, DC

- 100 attendees
- Plenary talks and Breakout sessions
- 6 funding agencies presented talks
- Meeting Report Published OA
- Set in motion organizing calls for funding for FAANG projects in US and Europe
- RFP in 2017 at USDA, 2018 in EU
FAANG is global - >400 members across six continents

- FAANG will globally coordinate functional annotation across species for efficiency, comparisons

FAANG Committees have conference calls on-line periodically

FAANG members can self-select membership in these committees

Anyone can join FAANG!
Main current activities

- Establish set of Core assays, begin to develop ENCODE-type functional data
- Develop tissue description, storage and sharing protocols
- Develop computational tools to analyze data
- Develop bioinformatics infrastructure
- Develop communication mechanisms

Reviewed in *Annual Reviews in Animal Biosciences February 2019*
Giuffra, Tuggle, et al.
Main current activities

- **Phase I**
  - Emphasis on establishing reference datasets for healthy adults
  - Few biological replicates, no treatments
  - Validate metadata toolboxes and analytical pipelines

- **Phase II**
  - Expand biological states- treatments, developmental stage, genetic variation
Core assays needed for chromatin state modeling

• RNAseq (stranded, deep) analysis of many tissues from the same 2-4 individuals
• Chromatin accessibility/architecture assays on these tissues (DNase I or ATAC-seq)
• Histone modification marks
  • H3K4 monomethylation
  • H3K4 trimethylation
  • H3K27acetylation
  • H3K27 trimethylation
Minimum information for finding functional regions in livestock genomes
Core assays needed for modeling

- RNAseq (stranded, deep) analysis of as many tissues from the same 2-4 individuals
- Chromatin accessibility/architecture assays on these tissues (DNase I or ATAC-seq)
- Histone modification marks
  - H3K4 monomethylation
  - H3K4 trimethylation
  - H3K27 acetylation
  - H3K27 trimethylation

Additional assays to determine genome function

- DNA methylation
- Regulatory factor protein binding (CTCF)
- Chromatin interaction (Hi-C)
## FAANG projects- an early list

<table>
<thead>
<tr>
<th>Species</th>
<th>Leading Inst. &amp; country</th>
<th>FAANG-FAANG-</th>
</tr>
</thead>
<tbody>
<tr>
<td>pig, cattle, goat,</td>
<td>INRA</td>
<td>pilot rel.</td>
</tr>
<tr>
<td>chicken</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pig, cattle, chicken</td>
<td>UC-Davis</td>
<td></td>
</tr>
<tr>
<td>horse</td>
<td>UC-Davis &amp; Nebraska Univ.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leibniz Inst. for Farm Anim. Biology</td>
<td></td>
</tr>
<tr>
<td>cattle</td>
<td>Int. Sheep Genome Consortium &amp; others</td>
<td></td>
</tr>
<tr>
<td>Sheep</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pig, chicken</td>
<td>Wageningen Univ.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dairy Futures Coop. Res. Centre, AgriBio &amp; others</td>
<td></td>
</tr>
<tr>
<td>cattle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cattle, pig</td>
<td>Alberta &amp; Guelph Univ. &amp; others</td>
<td></td>
</tr>
<tr>
<td>cattle and zebu</td>
<td>Adelaide Univ.</td>
<td></td>
</tr>
<tr>
<td>sheep, buffalo</td>
<td>Roslin- Edinburgh Univ.</td>
<td></td>
</tr>
<tr>
<td>cattle</td>
<td>Washington State Univ.</td>
<td></td>
</tr>
<tr>
<td>‘Arctic Ark’</td>
<td>Natural Resources Institute</td>
<td></td>
</tr>
</tbody>
</table>

FAANG-FAANG projects:
- Sylvain Foissac talk today
- Jessica Peterson talk today
- Brenda Murdoch talk today
Progress on UCD FAANG Pilot project

<table>
<thead>
<tr>
<th>Method</th>
<th>Data Analysis</th>
<th>Data Analysis</th>
<th>Data Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA-seq</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNase/ATAC-seq</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3K4me3</td>
<td>Data Analysis</td>
<td>Data Analysis</td>
<td>Chromatin Sheared</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>Data Analysis</td>
<td>Data Analysis</td>
<td>Chromatin Sheared</td>
</tr>
<tr>
<td>H3K4me1</td>
<td>Data Analysis</td>
<td>Chromatin Sheared</td>
<td>Data Analysis</td>
</tr>
<tr>
<td>H3K27ac</td>
<td>Data Analysis</td>
<td>Chromatin Sheared</td>
<td>Data Analysis</td>
</tr>
<tr>
<td>CTCF</td>
<td>Data Analysis</td>
<td>Data Analysis</td>
<td>Chromatin Sheared</td>
</tr>
<tr>
<td>RRBS-seq</td>
<td>Data Analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WTTS-seq</td>
<td>Data Analysis</td>
<td>Tissues Collected</td>
<td>Tissues Collected</td>
</tr>
</tbody>
</table>
Example data for UCD pilot FAANG

Kern, Zhou, et al., unpublished data
Future of livestock functional genomics

- **New** $2.5 million Porcine FAANG project funded by NIFA-AFRI
  - *Headquartered at ISU*
- One of three funded (bovine, chicken) in 2018
- **Overall goal**: catalog functional elements in the porcine genome for many biological states, create Chromatin state map
- **Aim 1 Adult tissues**
  - extend FAANG pilot of 8 -> 25 tissues
- **Aim 2 Fetal tissues**
  - Correlate allele-specific expression and chromatin modification
  - Parent-of-original effects (reciprocal crosses Meishan x WhiteCross)
- **Aim 3 Immune System**
  - Circulating Blood Cells- flow-sorted into functional types
  - *Single cell analysis of blood PMBC and immune tissues*
Getting Involved: Resources

- FAANG website www.faang.org
- FAANG Publications
- FAANG meetings (talk ppts available)
Getting Involved: Resources

- **FAANG data portal**: [http://data.faang.org](http://data.faang.org)

Access to FAANG data limited to members → agree to share data and not publish first on others' data.
**FAANG data publicly available**  
*(circa October 2018)*

http://data.faang.org

Interested in submitting your molecular data to FAANG?

We want your data!

Plans to set standards for data to be included in meta-analyses.  
**RNAseq (ENCODE):**  
- stranded  
- >30M aligned reads

*Figure 4*  
Registered specimens in the Functional Annotation of Animal Genomes (FAANG) data portal by species (as of October 2018) data are available from http://data.faang.org/home.
Conclusions and Outlook

• > $40 Million for FAANG world-wide and more to come
  • ~$7.5M from USDA-NIFA-AFRI ($6 M started in 2018)
  • ~$20M from EC H2020 (to start in 2019)
  • ~$5M from Australia
  • ~$7M from Canada

• Next five-ten years will see an explosion in functional data in the cattle, pig, chicken, sheep and horse genomes

• Realizing a Predictive Biology- increase usefulness to society and industry
Institutions supporting FAANG

Join FAANG (it’s free!)
Visit: www.faang.org
Email us: faang@iastate.org