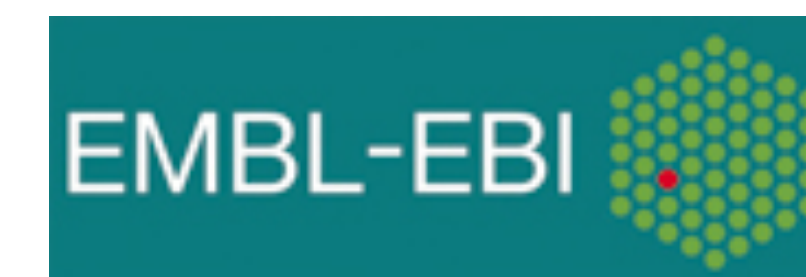


# The FAANG project's commitment to data standards, annotation and sharing

Ian Streeter, David Richardson, Laura Clarke, Paul Flicek, and the FAANG consortium  
European Molecular Biology Laboratory, European Bioinformatics Institute

email: [faang@iastate.edu](mailto:faang@iastate.edu)  
website: [www.faang.org](http://www.faang.org)  
twitter: @faangomics

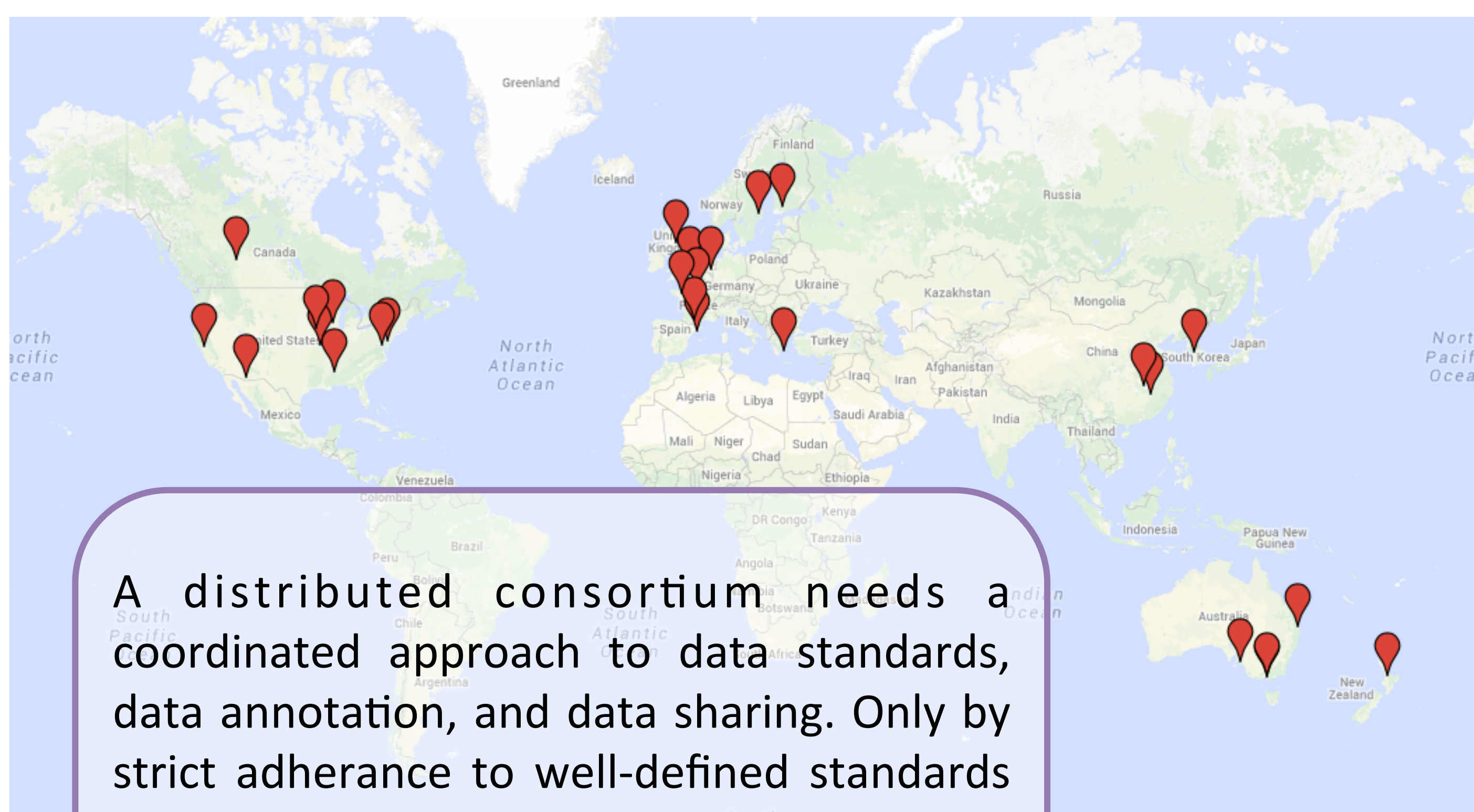


## From genotype to phenotype

The **F**unctional **A**nnotation of **A**nimal **G**enomes (FAANG) Consortium has one grand aim: to identify all functional elements in animal genomes.

Understanding this link between genotype and phenotype is of clear importance, motivated by the value of animals as food sources, models for human health, and as key ecological actors.

## An international collaboration



A distributed consortium needs a coordinated approach to data standards, data annotation, and data sharing. Only by strict adherence to well-defined standards will FAANG produce uniform datasets that are useful to all genomics research communities.

## Standards across borders

### Experimental Standards

FAANG has defined a set of core assay types:

- RNA-Seq
- ChIP-Seq
- Chromatin Accessibility

FAANG members will generate these core data for all samples using defined experimental protocols

### Metadata Standards

FAANG has defined metadata standards for annotating *experiments*, *samples* and *analyses* with standard attributes

These are the key to making data usable and understandable by the genomics community

Ontologies are used for specificity in data attributes, and to enable advanced data mining

<http://tinyurl.com/faanggit>



FAANG's metadata standards git repo

### Bioinformatics standards

FAANG is ensuring all data analysis products are equivalent and comparable with each other

- Uniform analysis pipelines
- Benchmarking of different pipelines
- Standard reference datasets
- Minimum aligned coverage
- Standard normalization methods
- Standard file formats

## It's good to share

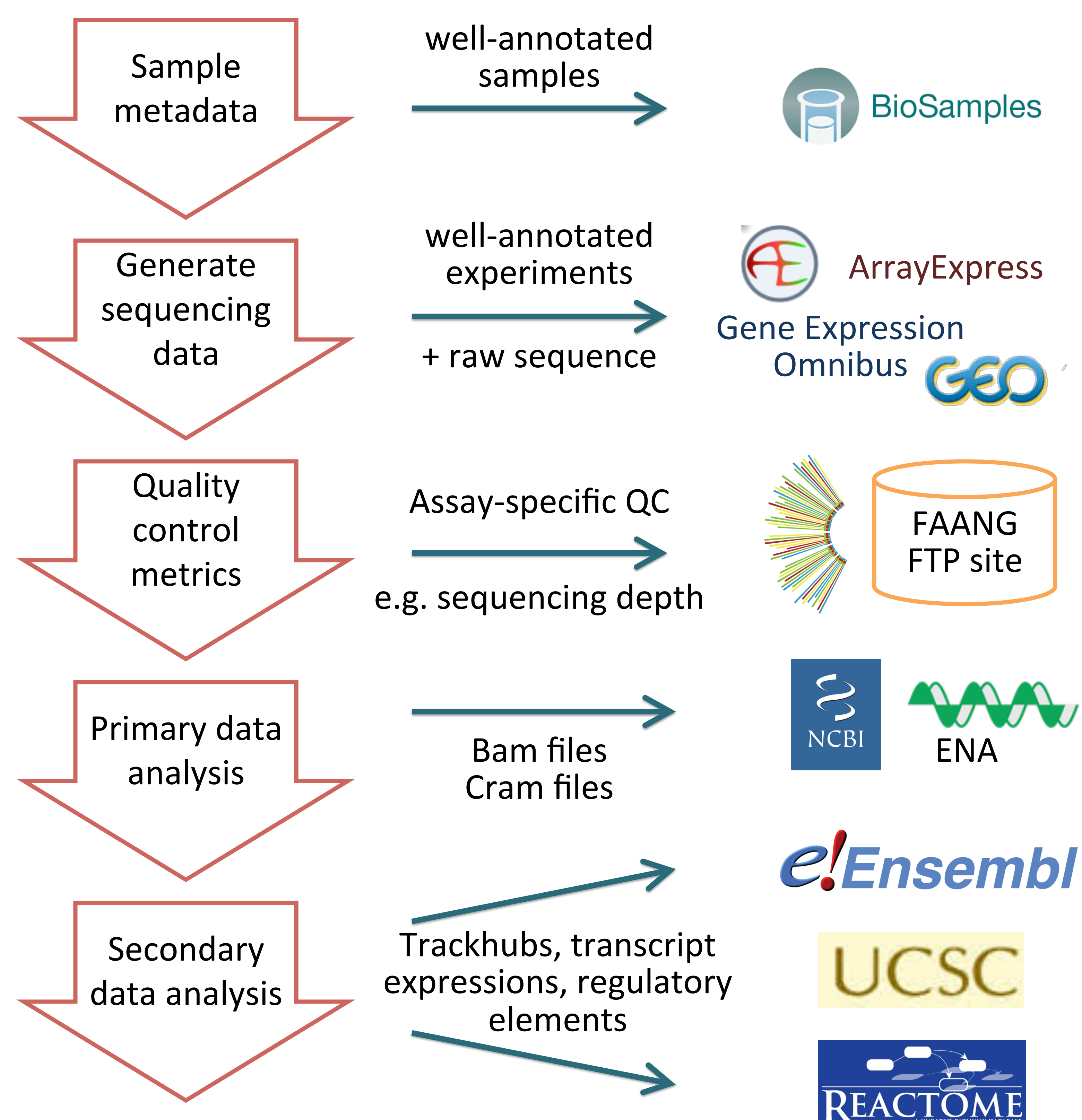
- FAANG is committed to sharing data rapidly, before publication. All members have agreed to the Toronto Data Release workshop guidelines for pre-publication data release<sup>1</sup>.
- FAANG also commit to making regular public releases of primary and secondary analysis results to provide access to the wider community.
- FAANG is building a portal and tools to aid discoverability and enable interactive analysis and comparison of the data.

Experiment	Source	Description	Name	Sex	Bisulfite-Seq	DNAse-Seq	RNA-Seq	ChIP Input	H3K4me3	H3K4me1	H3K9me3	H3K27ac	H3K27me3
ChIP Input													
H3K27me3	Venous blood	CD14-positive, CD16-negative classical monocyte	C0010K	Female									
H3K4me1	Venous blood	mature eosinophil	S00BKK	Female									
H3K4me3	Venous blood	CD4-positive, alpha-beta T cell	S007DD	Female									
H3K9me3	Cell Line	Mantle cell lymphoma	Z-138	Male									
Cell type	Cell Line	Germinal Center B-Cell-Like Diffuse Large B-Cell Lymphoma	KARPAS-422	Female									
CD14-positive, CD16-negative classical monocyte	Cord blood	CD14-positive, CD16-negative classical monocyte	S000RD	Male									
mature neutrophil													
alternatively activated macrophage													
CD4-positive, alpha-beta T cell													

## Want to join in?

Send an email to [faang@iastate.edu](mailto:faang@iastate.edu) to register interest in FAANG's activities

## Getting all data to the public domain



### References:

1. Toronto International Data Release Workshop Authors. Prepublication data sharing. Nature. 2009; 461:168-70.