The **F**unctional **A**nnotation of **An**imal **G**enomes Data Standards, Annotation and Sharing

FAANG
Functional Annotation of Animal Genomes

Laura Clarke[1], Alan L. Archibald[2], Paul Flicek[1], David Burt[2], David Hume[2], Douglas Vernimmen[2], Mick Watson[2], Mario Caccamo[3], Matt Clark[3], Federica Di-Palma[3], David Swarbreck[3] and The FAANG Consortium, 1:European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, United Kingdom, 2:The Roslin Institute and R(D)SVS, University of Edinburgh, Midlothian, United Kingdom, 3:The Genome Analysis Centre, Norwich, United Kingdom

High Quality Functional annotation of Animal Genomes

The **F**unctional **A**nnotation of **AN**imal **G**enomes (FAANG) Consortium aims to produce high quality functional annotation of animal genomes. To meet these aims FAANG will specify a set of core assays to be conducted on all samples to provide a reference set of functional annotation. FAANG will support standard lab protocol exchange and defined standards both experimental and bioinformatics.

FAANG is an international collaboration

Current FAANG members are based in laboratories around the world. FAANG is accepting new members. Please email faang@iastate.edu to register interest in joining.

Standards

FAANG members will work together to define the standards for the experiments conducted, the metadata associated with the experiments and the bioinformatics methods and metrics associated with the data.

Experimental Standards
FAANG members will share their experimental protocols and produced data for a set of core assay types:
- RNA-Seq
- ChIP-Seq
- Chromatin Accessibility

Metadata Standards
By annotating all experiments, samples and analyses with standard attributes, all FAANG members and the community will be able to understand the experiments and analysis conducted and the samples that have been assayed. Ontologies will be used to ensure specificity in attributes.
- Sample
- Experiment
- Analysis

Bioinformatics standards
Bioinformatics standards will ensure all FAANG data 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

Data Flow

Members of the FAANG consortium will follow a standard data flow pattern. The data producers will run experiments and apply standard data quality controls (QC) and submit the data to the public archives at EMBL-EBI and NCBI. The Data Coordination Centre (DCC) will retrieve the submitted data and provide it and standard analysis products to the Data Analysis Groups (DAG). The DCC will also create standard QC metrics and track project metadata. The DCC and DAG will work together to provide primary and secondary analysis products to the community and build data access tools which both facilitate discovery of the analysis results, allow attachment of the data to genome browsers through track hubs and interactive analysis and comparison of the data.
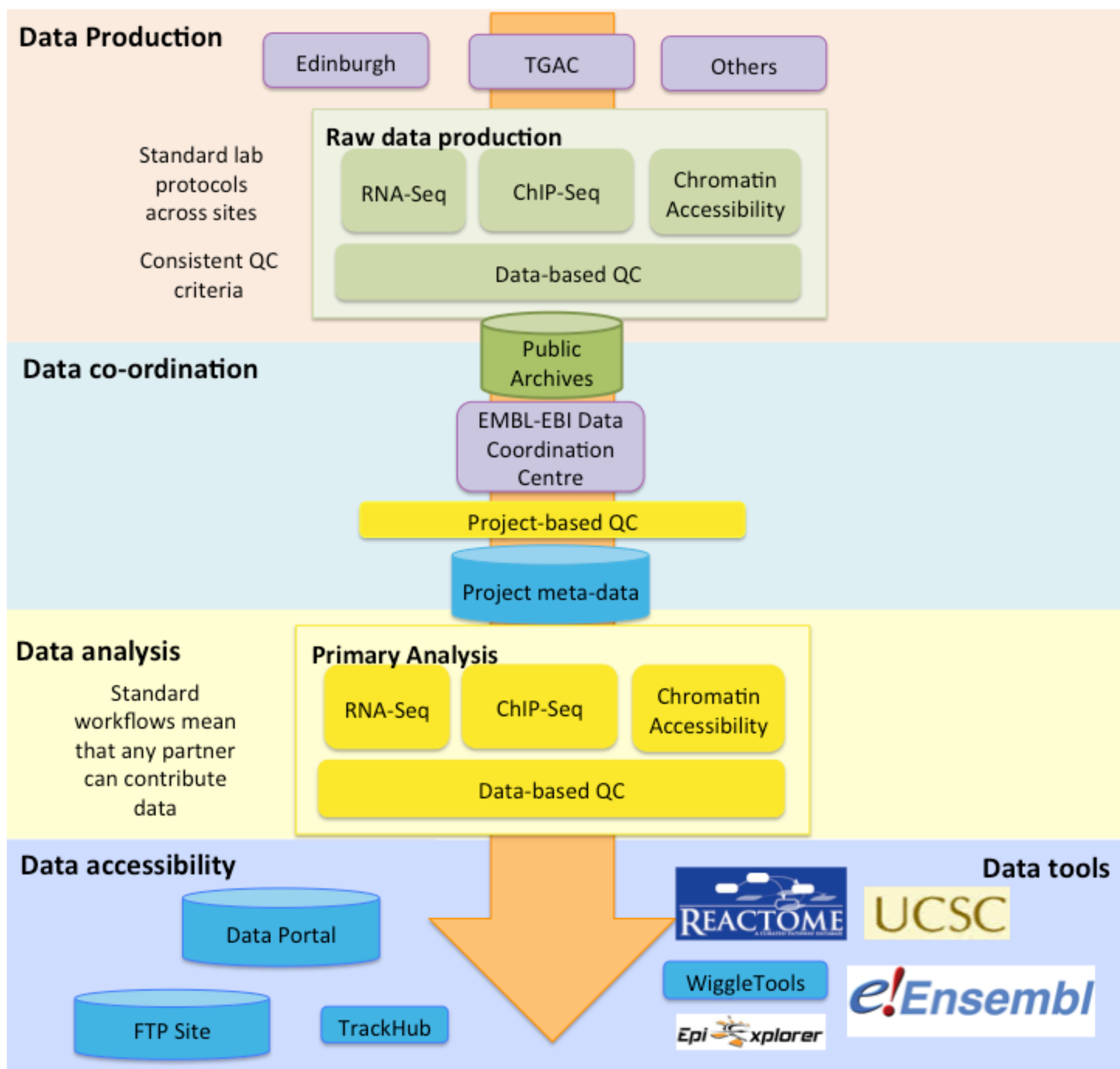
Data Sharing

All FAANG members commit to following the Toronto Data Release workshop guidelines for pre-publication data release[1]. This will be facilitated using the public archives at EMBL-EBI and NCBI. The FAANG DCC will also provide an internal dropbox to facilitate within-consortium results sharing. FAANG also commit to making regular public releases of primary and secondary analysis results to provide access to the wider community. FAANG will also build a portal and tools to aid discoverability and enable interactive analysis and comparison of the data.

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

email: faang@iastate.edu
website: www.faang.org
twitter: @faangomics

TGAC The Genome Analysis Centre™
ROSLIN
EMBL-EBI