# Update on the Functional Annotation of ANimal Genomes (FAANG) initiative

ELISABETTA GIUFFRA (GABI-GIS)

*Science Animale Paris Saclay  (SAPS)*
Jouy-en-Josas, 10 July 2015

# TALK

❖ **Introduction**

  ▪ **Framework of reference: ENCODE projects**

❖ **Conception and start of FAANG**

  ▪ **White paper**

  ▪ **Fr-AgENCODE project (SelGen metaprogramme)**

❖ **FAANG ongoing actions**

❖ **What's next**

❖ **Common Discussion**

# TALK

- ❖ **Introduction**

  - ▪ **Framework of reference: ENCODE projects**

- ❖ **Conception and start of FAANG**

  - ▪ **White paper**

  - ▪ **Fr-AgENCODE project (SelGen metaprogramme)**

- ❖ **FAANG ongoing actions**

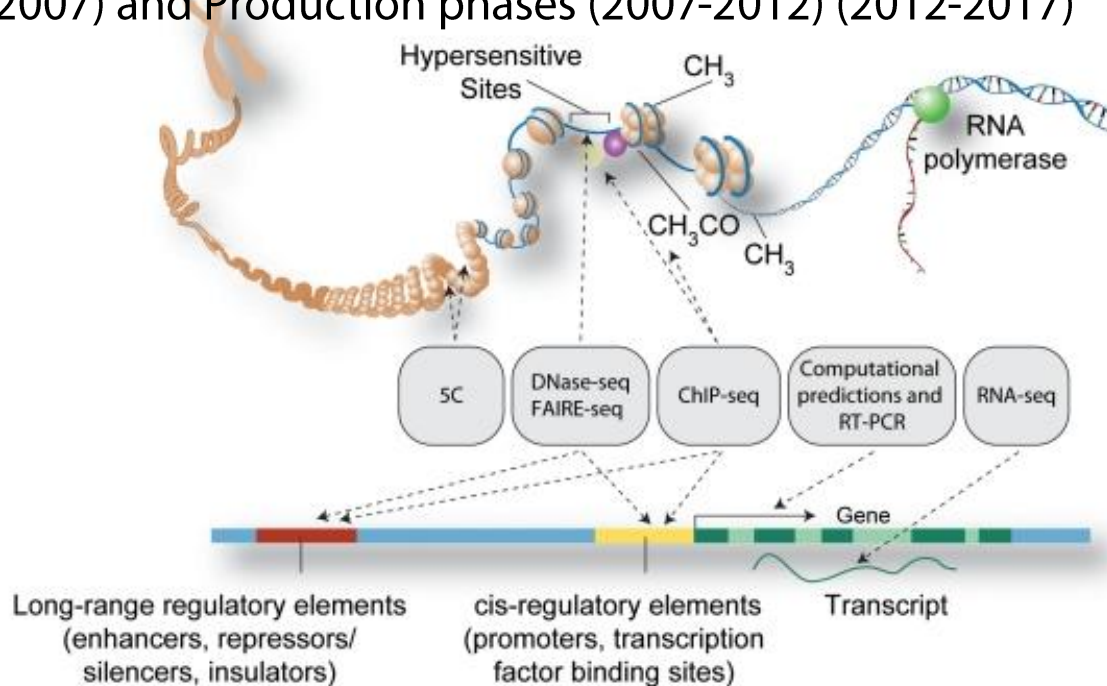- ❖ **What's next**

- ❖ **Common Discussion**

# ENCODE projects (human, mouse, drosophila, etc.)

**ENCODE *human*:**

❖ Produced **exaustive catalogues of regulatory genomic elements** in the human genome sequence

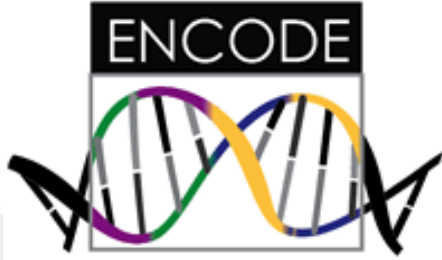❖ Pilot (2003-2007) and Production phases (2007-2012) (2012-2017)



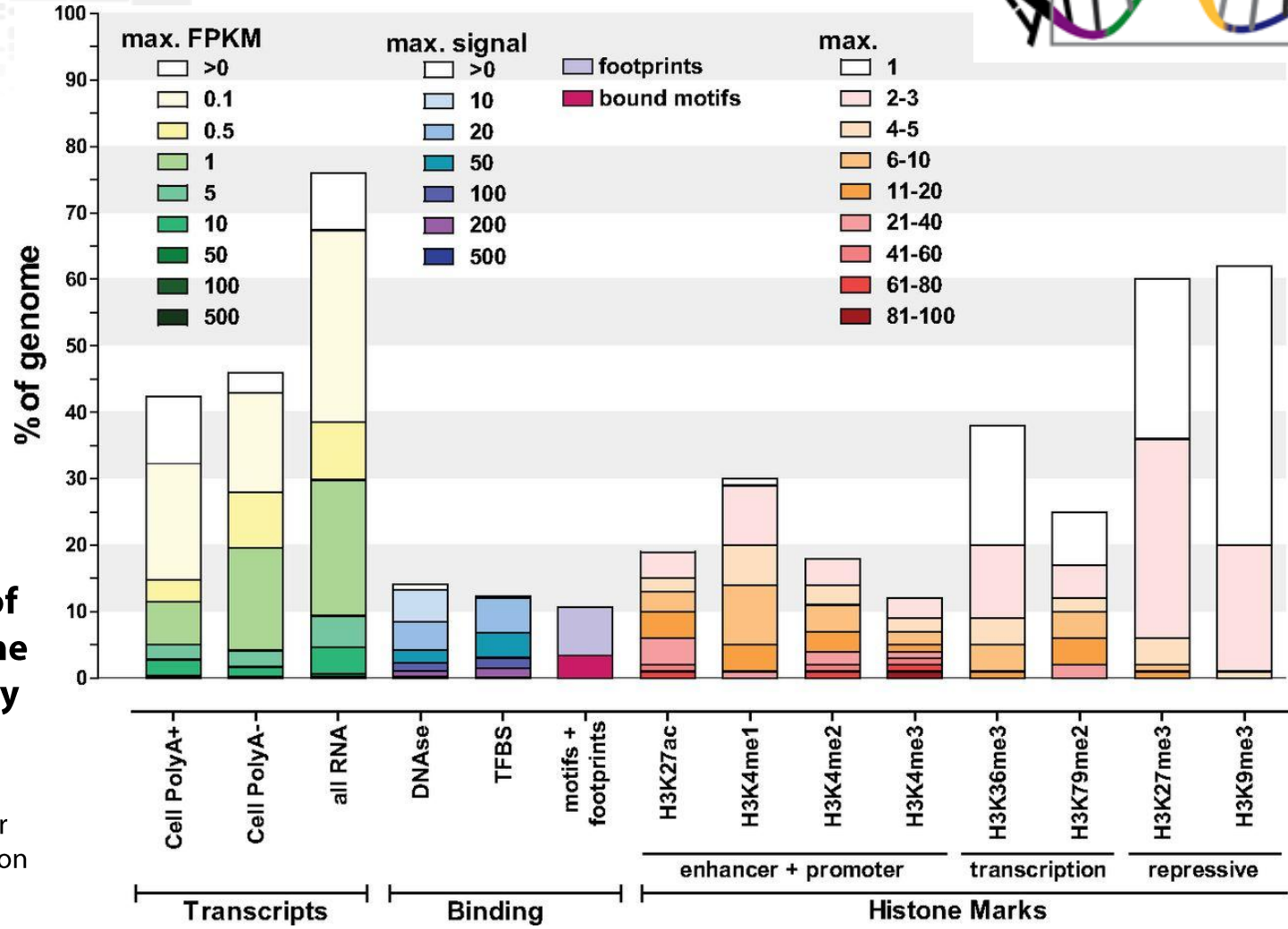*Modified from PLoS Biol 9-e1001046, 2011 & Science 306:636, 2004*

*32 institutes, 440 scientists, 1640 datasets from 147 different cell types, 1600 experiments, 12,000 files analyzed, 15TB of disk, over $300M in total (NHGRI)…*

# ENCODE projects: human

**Pervasive activity over an unexpectedly large fraction of the human genome**

**Kellis et al. PNAS 2014 ,vol. 111 (Fig. 2 Summary of the coverage of the human genome by ENCODE data)**

FPKM: read fragments per kilobase of exon per million reads

# ENCODE projects: human

## Key findings

- Much of the genome lies close to a "presumed" regulatory event
- ~ 400,000 'enhancer-like' and ~70,000 'promoter-like' regions
- Promoter functionality can explain most of the variation in RNA expression
- Many non-coding variants in individual genome sequences lie in ENCODE-annotated functional regions
- SNPs associated with disease are enriched within non-coding functional elements

❖ **The major contribution of ENCODE (to date) has been high-resolution, highly-reproducible maps of DNA segments with biochemical signatures associated with diverse molecular functions**

❖ **Emerging genome editing methods should considerably increase the throughput and resolution with which these candidate elements can be evaluated by genetic criteria**
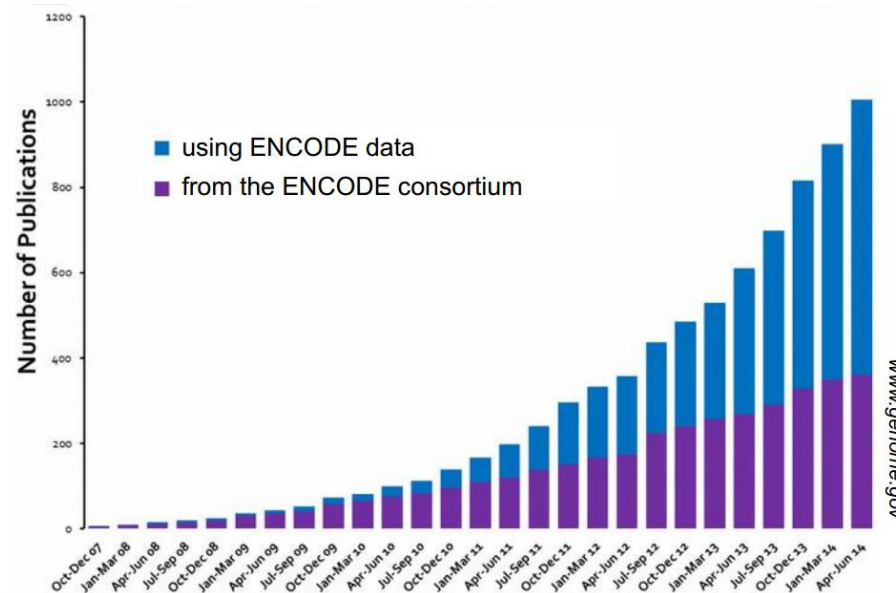
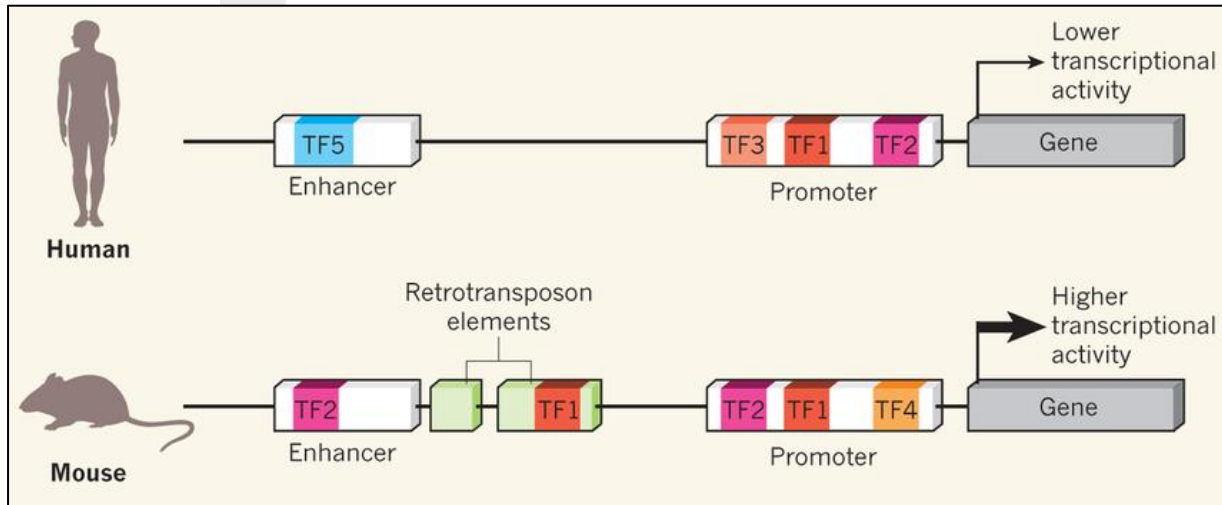*(Kellis et al. 2014, PNAS vol. 111)*

# Lessons from ENCODE projects

- Well established **consortium rules and policies (followed by other consortia)**

- Established technologies (= *now, lower costs*) and **experimental and data standards of reference**

- Demonstrated the **value of freely available reference datasets**, and **of integrating biochemical data alongside other evidence**

# Much conservation exists, but…



**Mouse ENCODE Nature, Nov. 2014 (4 papers*)**

*From: Carninci, P. 2014 Nature 515*

*(\*) Yue, F. et al.; Stergachis, A. B. et al.; Cheng, Y. et al.; Pope, B. D. et al.*

❖**The expression profiles of several genes within distinct biological pathways and the cis- regulatory landscape have diverged between human and mice**
❖ **Species-specific candidate regulatory sequences are significantly enriched for particular classes of repetitive DNA elements.**

❖**Understanding the phenotypes of interest in the authentic biological context requires organism specific information**
❖**How genetic variation affects the regulatory lanscape?**

# TALK

❖ **Introduction**

   ▪ **Framework of reference: ENCODE projects**

❖ **Conception and start of FAANG**

   ▪ **White paper**

   ▪ **Fr-AgENCODE project (SelGen metaprogramme)**
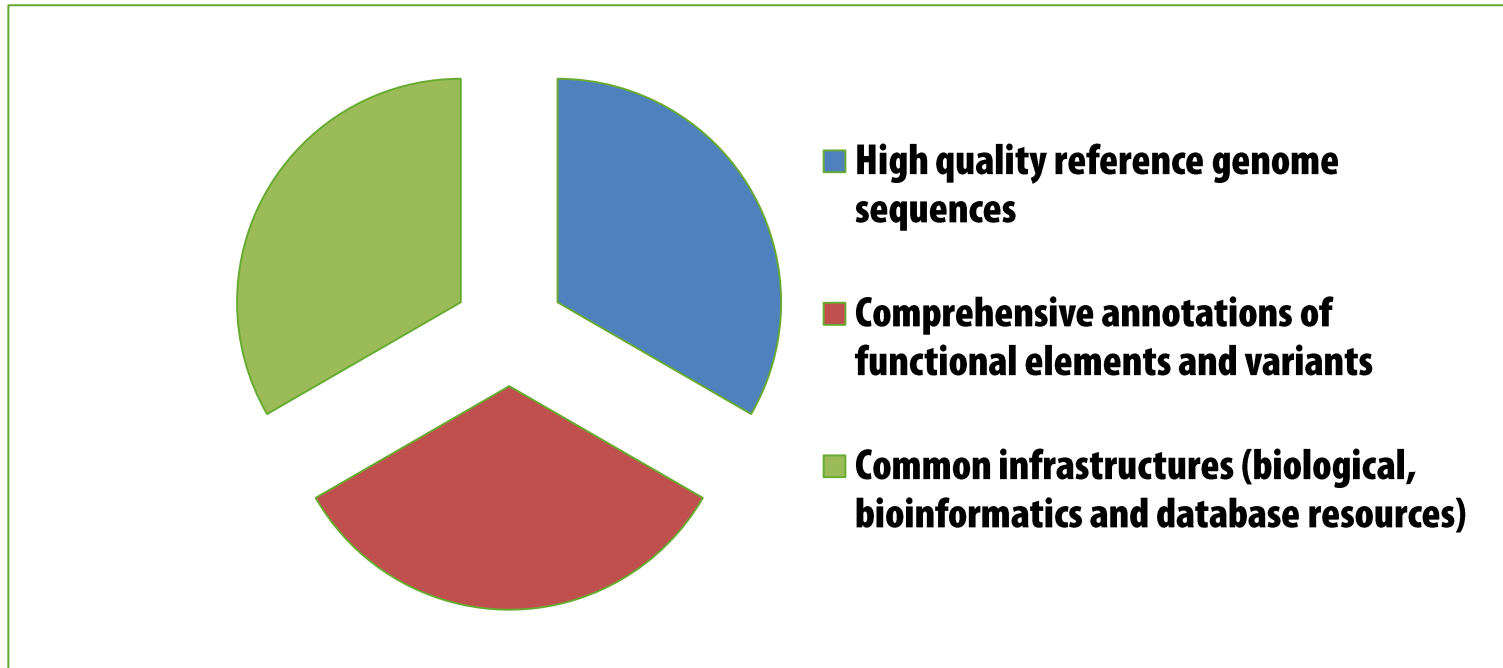
❖ **FAANG ongoing actions**

❖ **What's next**

❖ **Common Discussion**

# Why FAANG for Domesticated Animals?

**Understanding the genotype to phenotype link**
*(Fundamental research AND Genomic selection)*



- High quality reference genome sequences
- Comprehensive annotations of functional elements and variants
- Common infrastructures (biological, bioinformatics and database resources)

## Challenges:
❖ Many species and groups with limited resources
❖ Variable quality of existing reference genomes and annotation

# Conception and start


FAANG
Functional Annotation of Animal Genomes

**Alan L. Archibald** (Univ. Edinburgh, UK): **Proposing the case for a farm animal ENCODE**

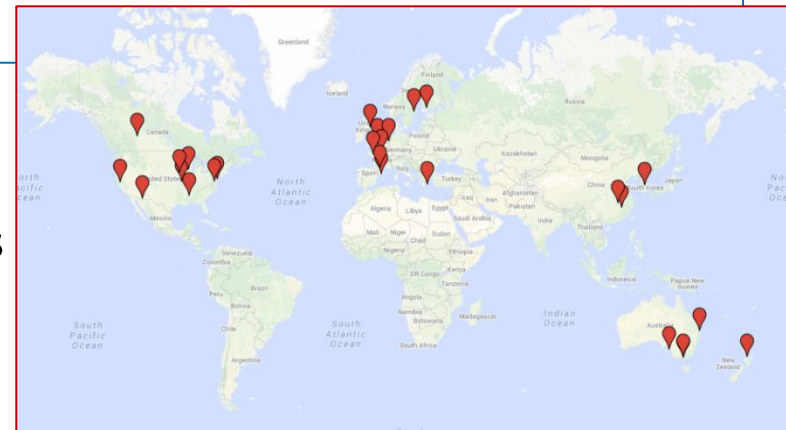**EU-US Animal Biotechnology Working Group (ABWG)**

*ABWG Workshops:*
**Washington DC, Nov 2011:** Strategic Priorities to Realize the Promise of Genomics to Animal Health, Well-being & Production.
**Hinxton, UK, Sept 2012:** A shared international cyber-infrastructure dedicated to genomic and phenomic animal data.
**S. Diego (PAG XXII), Jan. 2014:** Agencode workshop - to develop a plan for a coordinated ENCODE-like project for food animals (AgEncode).

❖**Preparative phase (since 2011) led by ABWG with a small group of scientists from 12 countries**

June 2014:  ~~AgENCODE~~  ⟹  FAANG!

# The FAANG white paper

**Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project**

- *Genome Biology 2015, 16:57 doi:10.1186/s13059-015-0622-4*

- Corresponding authors: E. Giuffra, C.K. Tuggle

- Open Access Paper

- > 3000 views since March 25th



❖*About:*

***Biological targets and resources*** *(aim of first phase: build a biological reference of as many as possible* <u>tissues/primary cells</u>*)*
***FAANG data types***
***Common data infrastructure:*** *common protocols and standard operating procedures (ENCODE, IHEC guidelines):* <u>*data sharing and policies*</u>*, increase power for meta-analyses*
<u>***Next stage:***</u> ***expansion of covered species and diversity within and between species***

# FAANG data types (Box 1)

**FAANG**
Functional Annotation of Animal Genomes

## Core Assays

### 1. Transcribed loci:
❖**RNA-seq** (exhaustive catalogues of gene expression; starting point for improving genome annotation)

### 2. Chromatin Accessibility and Architecture:
❖DNAse hypersensitivity **(DNase I-seq)** and possibly **ATAC-seq** (Transposase-Accessible Chromatin with high-throughput sequencing)

### 3. Histone modification marks _(updated to 6: H3K9me3 and H3K36me3)_
❖ **H3K4me3** (promoters of active genes and transcription starts)
❖ **H3K27me3** (genes that have been silenced through regional modification)
❖ **H3K27Ac** (active regulatory elements)
❖ **H3K4Me1(**regulatory elements associated with enhancers and other distal elements, also enriched downstream of transcription start sites).

## Additional Assays
❖**DNA methylation**
❖**Chromosome conformation capture** (by **Hi-C**: genome-wide interactions in 3D)
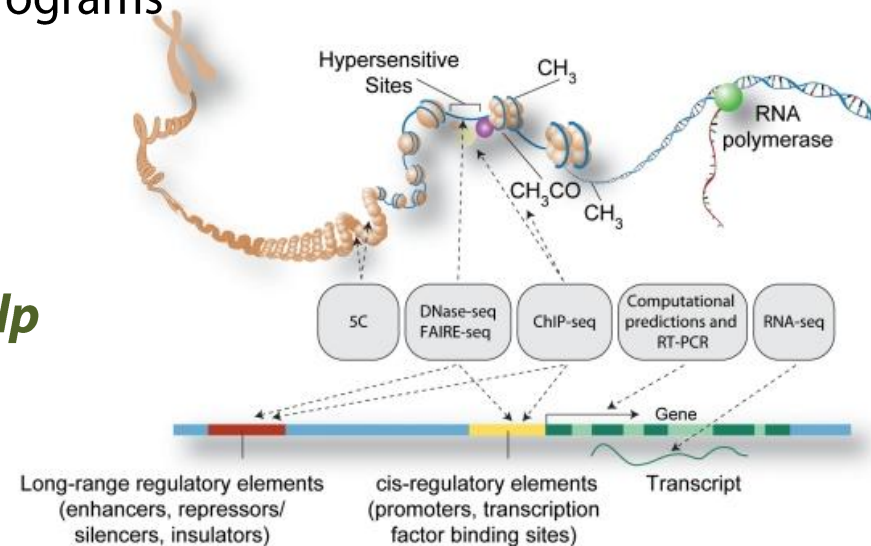
# The satellite FAANG project Fr-AgENCODE



- ❖ **A multi-species FAANG project** (cattle, pig, chicken and goat)
- ❖ **Federation of leading GA Div. Units** (with other INRA Units/ Divisions/ French institutes)
  into a durable network
- ❖  **Scalable** and linked to parallel research programs

**SelGen metaprogramme (INRA)**
AMI 2014; Axe 3: « Data production, storage and integration »
*Functional annotation of genomes can help improving genomic selection approaches*



➢*Internationalisation and visibility of the SelGen metaprogramme*
➢*Empowerment of activities of the GA Division in various large genome sequencing consortia*
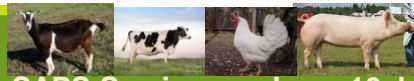
# FR-AgENCODE: an open network

*Currently: 58 scientists, FR, NL, UK (INRA: 9 units, 5 Divisions)*

FAANG
Functional Annotation of Animal Genomes

| | | | |
|---|---|---|---|
| **A** | **GABI (Coord.)** | **INRA, GA Div**. | **Elisabetta Giuffra**, Michele Tixier-Boichard, Marie-Helene Pinard |
| **A** | **VIM** | INRA, SA Div. | Nicolas Bertho |
| **B** | **PEGASE** | **INRA, GA Div.** | Sandrine Lagarrigue |
| **B** | **IGDR-CRNS-Univ.Rennes1** | CNRS, Université Rennes 1, IGDR, Institut de Génétique et Développement de Rennes | Thomas Derrien |
| **C** | **GENPHYSE** | **INRA, GA Div.** | **Sylvain Foissac,** Hervé Acloque, Stephane Fabre |
| **C** | **MIAT** | INRA, MIA Div. | Christophe Klopp, Christine Gaspin |
| **C** | **GenoToul** | **INRA, GA Div.** | Diane Esquerré |
| **D** | **URA** | INRA, PHASE Div. | Joen Gautron |
| **D** | **PRC** | INRA, PHASE Div. | Xavier Druart |
| **D** | **ISP** | INRA, SA Div. | Pascale Quéré |
| **E** | **WUR** | Animal Breeding and Genomics Centre - Wageningen Univ. | Ole Madsen, Martien Groenen |
| **F** | **RI** | The Roslin Institute and Royal (Dick) School of Veterinary Studies - Univ. Edinburgh | Alan Archibald |
| **G** | **EMBL-EBI** | European Molecular Biology Laboratory - European Bioinformatics Institute | Laura Clarke, Paul Flicek |

★ **Experimental animal facilities**

INRA
SCIENCE & IMPACT

CRB ANIM

Centres de Ressources Biologiques pour les Animaux Domestiques

SAPS Seminars – Jouy, 10 July 2015
Elisabetta.giuffra@jouy.inra.fr

www.faang.org; @faangomics

# Three Pilot Projects: FR-AgENCODE



## Biological targets and resources

**Jouy-en-Josas (coll. Tours)**
**GABI, VIM**
**Host immunity: immunocompetence; host-pathogen interactions, congenital diseases; host- microbiome interactions**
*Expression genomics, functional genomics (e.g. miRNAs), statistical modelling*

**Rennes**
**PEGASE** (coll. **IGDR-CRNS-Univ.Rennes1**)
**Energetic homeostasis and lipid metabolisms**
*RNA editing, Allele Specific Expression, gene and lncRNA modeling*

**Toulouse**
**GENPHYSE, MIAT+ GenoToul platform** *(France Génomique)*
**Animal production and reproductive traits +**
**Genomic data production and Bioinformatics**
*Genome methylation, genome sequencing analyses*

# FR-AgENCODE

**Collection and adequate storage of ALL tissues**
(Four biological replicates / 4 species)

**Target tissues of choice:**
- Liver (*hub*)
- Two primary lymphoid cells *(targeting blood cells)*

**FAANG assays of choice**
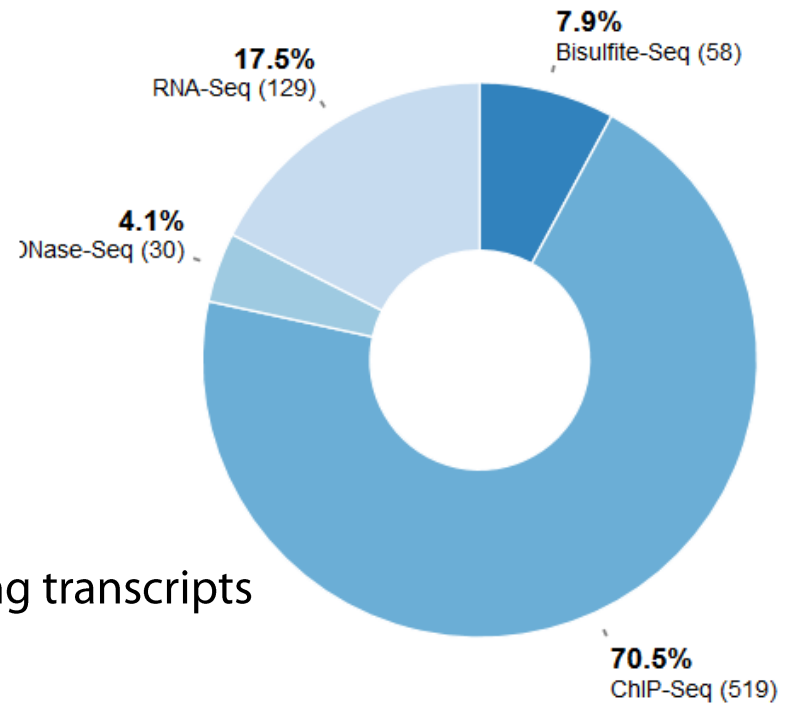1. **Whole Transcriptome** (RNA-seq, small RNA-seq)
- Improve annotation of coding genes and non coding transcripts

**2. Genome-wide chromatin interactome (Hi-C):**
- Intra-nuclear DNA-DNA interactions and comparative analysis: orthologous and species specific DNA-DNA interactions for each tissue/species.

**3. Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq):**
- Chromatin accessibility vs. on gene expression
- Complement/backup to Hi-C: explore the links between chromatin accessibility and organization



BL**UE**PRINT epigenome

- 7.9% Bisulfite-Seq (58)
- 17.5% RNA-Seq (129)
- 4.1% DNase-Seq (30)
- 70.5% ChIP-Seq (519)

# FR-AgENCODE: Hi-C

FAANG
Functional Annotation of Animal Genomes

❖**Hi-C data are analyzed to generate** *interaction maps* **highlighting trans- and cis interactions between genomic loci** (resolution 25 Kb: Chromatin compartments and subcompartments; 5kb and under: chromatin loops, promoters/ enhancers detection)
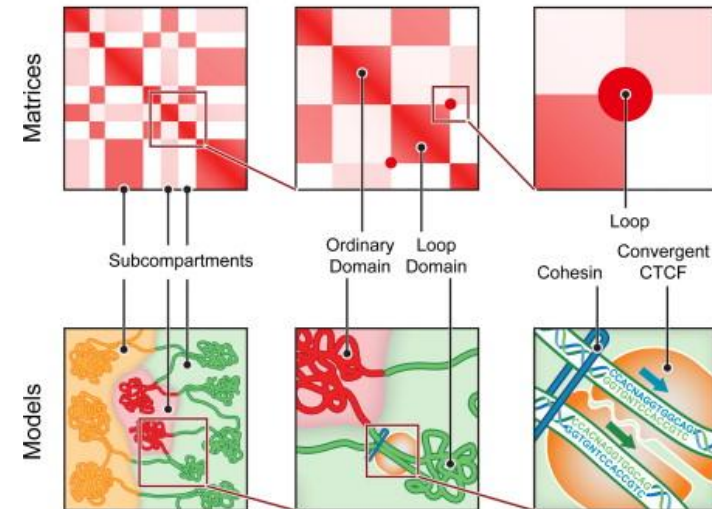
❖**Nuclear compartments and subcompartments defined using Hi-C data correspond to specific combination of histone marks**

= *a virtually complete set of annotations!*

*Challenges:*
▪*Bioinformatic analysis: complex, but in full development (Toulouse) on preliminary datasets (cell line and mouse/pig liver)*
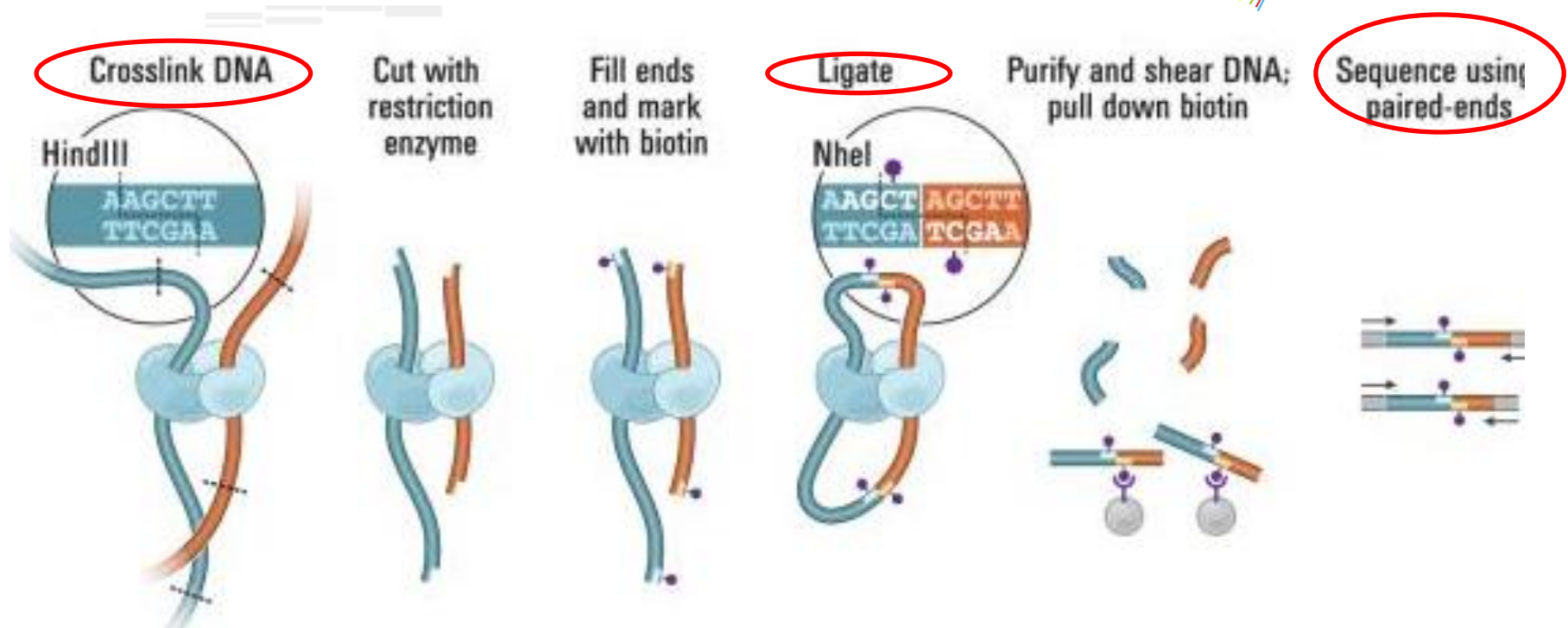
Hi-C Matrices and Models



Matrices

Subcompartments | Ordinary Domain | Loop Domain | Loop Convergent CTCF

Models

Cohesin

Suhas et al. **A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping** Cell, Volume 159, Issue 7, 2014

# FR-AgENCODE: Hi-C

Crosslink DNA
HindIII

Cut with restriction enzyme

Fill ends and mark with biotin

Ligate
NheI

Purify and shear DNA; pull down biotin

Sequence using paired-ends

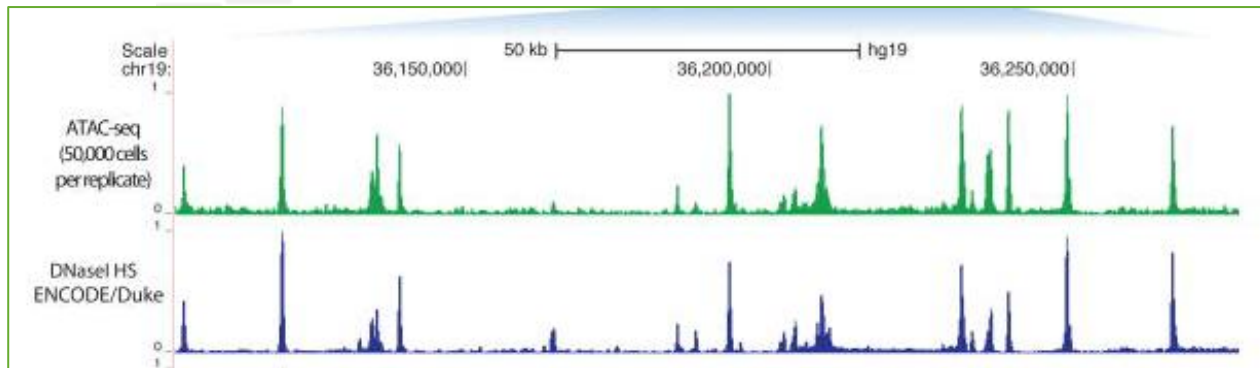**Advantages:**

**NOVELTY and "federation"**

**Challenges:**

▪**Requires fresh or _adequately stored_ tissues/cells**

▪**Procedures for samples preparation have been developed and implemented at Sampling Sites for a subset of tissues (main targets: liver and spleen)**
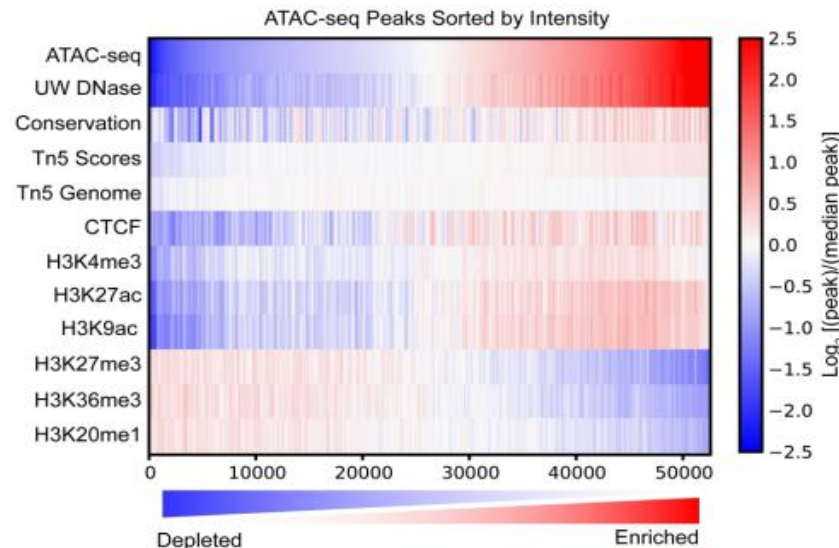
# Fr-AgENCODE: ATAC-seq

**FAANG**
Functional Annotation of Animal Genomes

**ATA-seq correlates with various features of the genome:**



**Buenrostro et al. A comparison of ATAC-seq to other open chromatin assays at a locus in GM12878 lymphoblastoid cells displaying high concordance. Nat Methods. 2013**
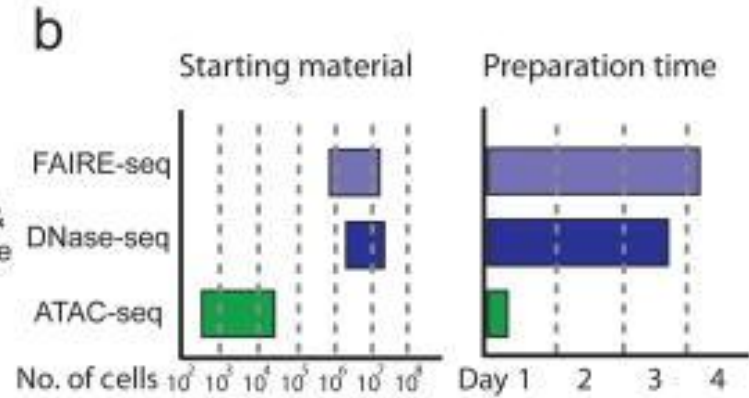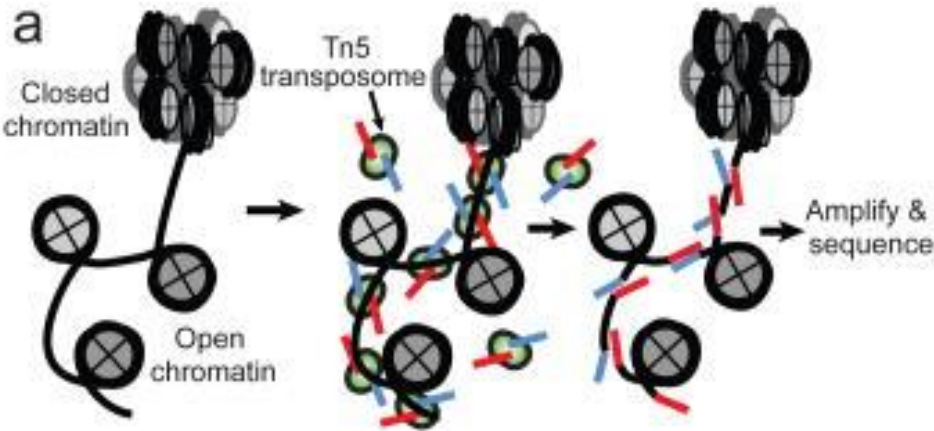
❖Peak intensity most strongly correlated with DNase hypersensitivity and CTCF

❖Correlated with histone marks associated with active chromatin and anti-correlated with histone marks associated with inactive chromatin and gene bodies



**Buenrostro et al Nat Methods. 2013, Curr Protoc Mol Biol. 2015**
**Recent: Single-cell chromatin accessibility! (Buenrostro et al. Nature 2015)**

# ATAC-seq= Transposase-Accessible Chromatin with high-throughput sequencing

*Advantages:*

**Novelty, much easier than DNAse footprinting**

**About 50000 cells (vs. 'millions for Hi-C)**

**Very easy process to prepare libraries (some already produced for the GM12878 cells)**

*Challenges:*

▪**Requires fresh tissues/cells**

▪**First stage of sample preparation (transposition) have been developed and implemented at Sampling Sites for a subset of tissues (main targets: liver and spleen)**

# Fr-AgENCODE – year 1 in progress

**FAANG**
Functional Annotation of Animal Genomes

❖ - Kick Off: 6 February (Paris)

- Hi-C set up completed; ATAC-seq in progress (only chicken and pigs)

- Cattle and chicken sampling completed (July)

❖ Active participation of several partners to all FAANG Committees: Steering, Communication, Animals/Samples/Assays (ASA), Bioinformatics and Analysis  (B&DA), Metadata and Data Sharing (M&DS)

❖ *New personnel joining:*

Kylie Munyard- Academic Study program (1 year), Curtin Univ., AU  (6 months, Toulouse)

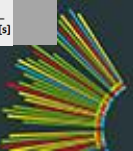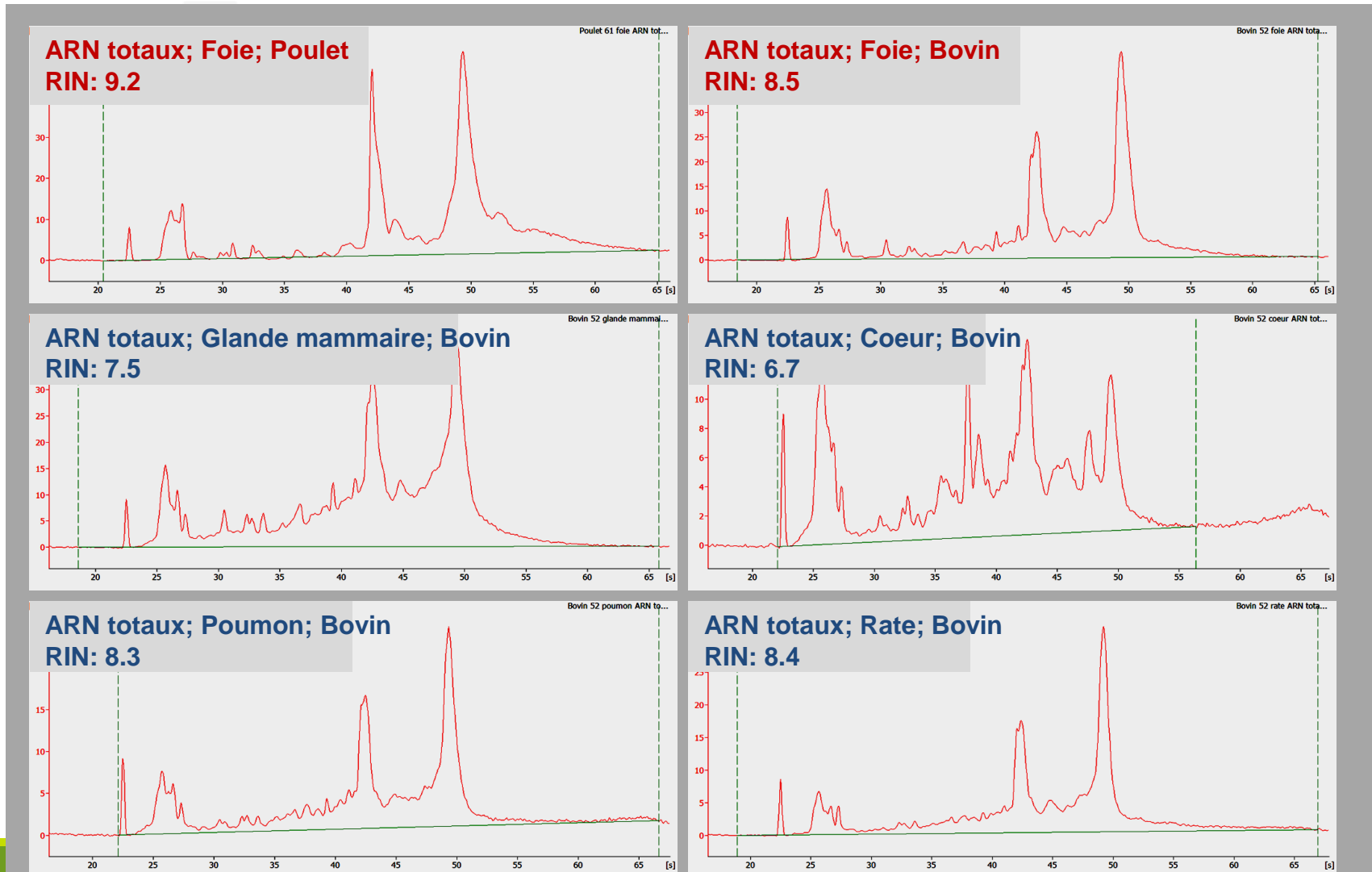Sarah Djebali- Agreenskill+ postdoctoral applicant (3 years, Toulouse)

*Expected 2015:*

○ **Samples collection completed (4 species)**

○ **Transcriptome + Hi-C profiles of liver in 4 species completed**

○ **Start of bioinformatic analyses**

# Sampling (WP1): update July 2015

- **Cattle and chicken sampling completed (with good surprises!)**

# TALK

- ❖ **Introduction**

  - ▪ **Framework of reference: ENCODE projects**

- ❖ **Conception and start of FAANG**

  - ▪ **White paper**

  - ▪ **Fr-AgENCODE project (SelGen metaprogramme)**

- ❖ **FAANG ongoing actions**

- ❖ **What's next**

- ❖ **Common Discussion**

# The FAANG Committes



## Working groups

🚩 **Steering Committee**

🏴 Animals, Samples and Assays (ASA)

🏴 Bioinformatics and Data Analysis (B&DA)

🏴 Communication (COM)

🏴 Metadata and Data Sharing (M&DS)

**OPEN participation to all FAANG members**

Operate by monthly calls and will provide soon web access to all materials (minutes, protocols, infos…)
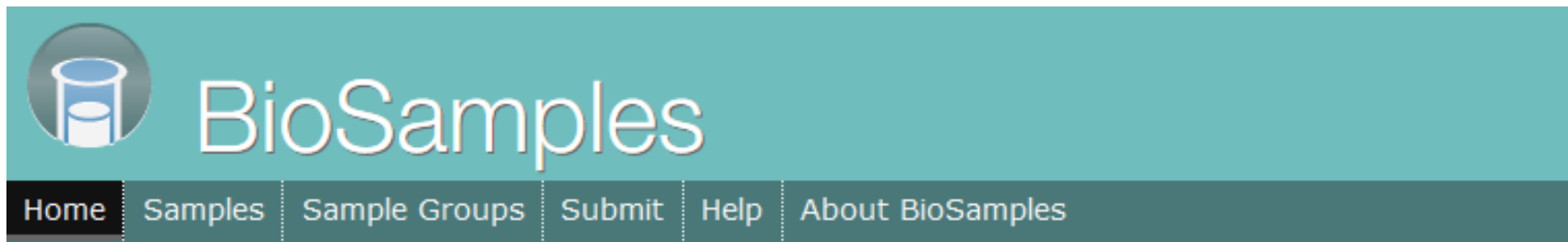
**INRA participation:**

- Steering: E.Giuffra, G. Tosser-Klopp

- Communication: G. Tosser-Klopp

- ASA: E.Giuffra ( co-leader), H. Acloque

- B&DA: S. Foissac (leading the « Analysis-Structural » group), S. lagarrigue, C. Klopp…

- M&DS: G. Tosser-Klopp

# Animals, Samples and Assays Working Committee (ASA)



- ❖ **Organizing contributions in terms of animals (breeds, populations…), sample collections (suited for as many as possible assays), shared protocols for the FAANG core assays**

- ❖ **Work by groups of discussion, e.g. on RNAs (standards to be proposed soon), and chromatin assays**

- ❖ **<u>Interacting with M&DS for METADATA of current and future collections</u>**

- ❖ **A paper is planned by ASA on common FAANG sample resources: M. Tixier-Boichard invited to organize and act as corresponding author**



## BioSamples - database of sample descriptions

# ASA: ruminants – July 2015

| Genetic line | Development stage | Tissues/Cells | Institute |
|---|---|---|---|
| **Cattle** | | | |
| Line 1 hereford | 14 months | all tissues | UC Davis |
| Holstein, 2 M and 2 F | Adult | All tissues + sorted blood cells | INRA |
| Holstein | Adult | PBMCs | INRA |
| Holstein (female) | 5th lactation, pregnant | all tissues | DEDJTR |
| Holstein (female) | 7th lactation, pregnant | all tissues | DEDJTR |
| Crossbred Cattle | | all tissues | UOA |
| | | | |
| **Sheep** | | | |
| Rambouillet Ewe | adult | all tissues | USDA,BCM,CSIRO,USU |
| | | | |
| **Goat** | | | |
| Alpine dairy, 2 M and 2 F | Adult | All tissues + sorted blood cells | INRA |
| Alpine dairy | 4 weeks | PBMCs | INRA |

# ASA: monogastrics – July 2015

| Genetic line | Development stage | Tissues/Cells | Institute |
|---|---|---|---|
| **Pig** | | | |
| Yorkshire | 6 months | all tissues | UC Davis |
| Large White, 2 M and 2 F | Adult | All tissues + sorted blood cells | INRA |
| Large White | 4 weeks | PBMCs | INRA |
| Yorkshire | 4 weeks | whole blood | ISU-NADC |
| Duroc | adult, male | all tissues | WUR |
| Pietrain | adult, male | all tissues | WUR |
| Large White | adult, male | all tissues | WUR |
| LW*LR challenged | 21d, 42d, 56d | all tissues | UOA |
| D*(LW*LR) challenged | 21 dp infection | reproductive tissues | UOA |
| Bama Xiang Pig | E55, 145d | all tissues | JXAU |
| Large White | 145d | all tissues | JXAU |
| | | | |
| **Horse** | | | |
| Thoroughbred | 3-6 years | all tissues | UCD/UN |
| Thoroughbred | Embryo at 34 d | whole embryo | Univ. KY |
| Thoroughbred | 23 months | all tissues | Univ. KY |
| Thoroughbred | foal | Tendon (superficial digital flexor) | Univ. KY |
| Thoroughbred | full term | Placental villous | Univ. KY |
| Freisian | 3-6 years | Cornea | UC Davis |

# ASA: birds – July 2015

| Genetic line | Development stage | Tissues/Cells | Institute |
|---|---|---|---|
| **Chicken** | | | |
| F1 (line 6X7) ADOL | 20 weeks | all tissues | UC Davis |
| F1 (line 6X7) ADOL | 2 weeks | CD4 T lymphocytes/spleen | ADOL/USDA |
| F1 (line 6X7) ADOL | 8 weeks | CD4 T lymphocytes/spleen | ADOL/USDA |
| White Leghorn, 2 M and 2 F | 42 weeks | All tissue+ sorted blood cells | INRA |
| F1 (broiler x layer) | 20 weeks | all tissues | ROSLIN |

## Gathering On Functional Annotation of ANimal Genomes (GO-FAANG Workshop)

The GO-FAANG workshop is scheduled for **October 7-8, 2015**. We will convene at the US National Academy of Sciences Building on the **National Mall, 2101 Constitution Ave NW, Washington, DC 20418**.
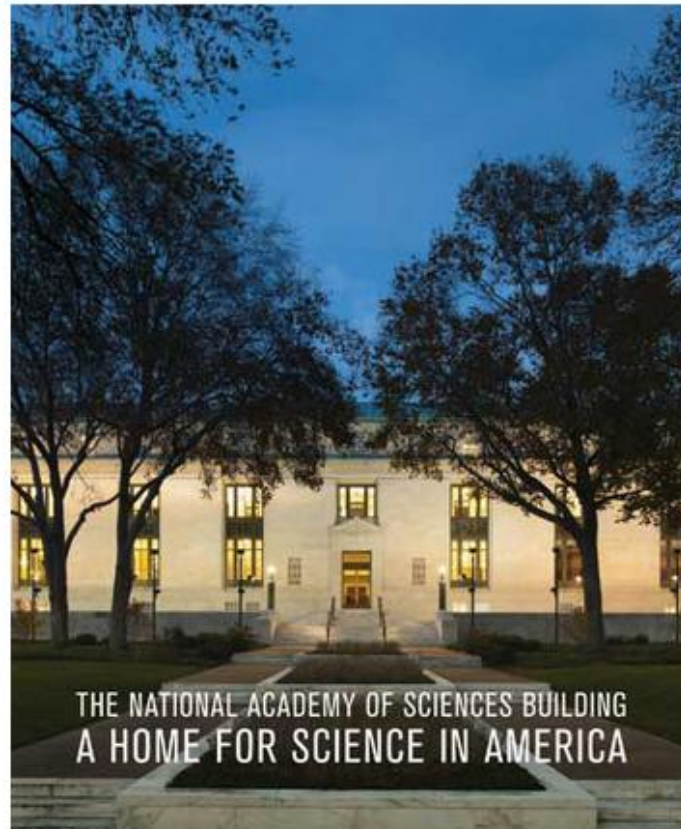
The purpose of this meeting is to exchange information and develop plans for organizing the FAANG effort world-wide.

Several Plenary speakers will provide inspiration for attendees to look forward to the value of functional annotation. The majority of the meeting time will be devoted to small and large group discussion of priorities and needs of the FAANG project.

Expected to attend: Genome scientists, industry and funding agency representatives from the US, Europe, Asia and Australia.

- Preliminary program information
- Venue information
- Hotel information
- Application to attend
  (The first pre-registration period was June 1-July 1; After July 1, you are still welcome to pre-register but will be put on a waiting list and contacted if space becomes available)

THE NATIONAL ACADEMY OF SCIENCES BUILDING
A HOME FOR SCIENCE IN AMERICA

# TALK

❖ **Introduction**

- ▪ **Framework of reference: ENCODE projects**

❖ **Conception and start of FAANG**

- ▪ **White paper**

- ▪ **Fr-AgENCODE project (SelGen metaprogramme)**

❖ **FAANG ongoing actions**

❖ **What's next**

❖ **Common Discussion**

# What's next? FRANCE

❖ **Reinforce/Expand Fr-AgENCODE (additional tissues and assays)**

Functional Annotation of Animal Genomes (FAANG) in 4 livestock species (<u>FAANG4</u>) – Subm. to *France Génomique* infrastructure as « large scale sequencing project » (June 2015) – <u>new partner: BDR (I. Hue)</u>

❖ **Reinforce interactions with research teams**

  ❖ **Other Divisions of INRA**

  ❖ **Other institutes for highly related methodological domains**

   *e.g.* large French infrastructures (France Genomique: <u>www.france-genomique.org</u>; and Institut Français de Bioinformatique: <u>http://www.france-bioinformatique.fr/</u>).

# What's next? everywhere

❖ **New grant applications (national, European and worldwide)**

**EU:**

- – Trans-domain COST Action submitted (outcome: Nov. 2015)

- – **Next call for EU infrastructures (MRSEI call, ANR) – the path towards an International Research Consortium**

- – Expected: support of several **specific** research projects relying on the knowledge generated by FAANG

❖ **Finalisation of FAANG common strategy:** *Gathering On Functional Annotation of ANimal Genomes (GO-FAANG Workshop) -* Washington DC, 7-8 Oct. 2015

# A thank to the (current) active participants to Fr-AgENCODE

**Management:** <u>Sylvain Foissac (co-coord)</u>, Sandrine Lagarrigue, Marie-Helene Pinard

**WP1 (sampling):** <u>Michele Tixier-Boichard, Stephane Fabre</u>

Pascale Queré, Francoise Drouet, Silvia Vincent Nailleau, Fany Blanc *(blood cells); all tissues:* Sophie Pollet, Adeline Goubil, Hervé Acloque, Cecile Berry, Joel Gautron, Xavier Druart, <u>and all (*many*) participants to *sampling* not listed here</u> *+ FAANG ASA and M&DS Committees*

**WP2 (assays)**: <u>Hervé Acloque, Diane Esquerré</u>, Sophie Pollet, Adeline Goubil *+ FAANG ASA Committee*

**WP3 (analysis:** <u>Sylvain Foissac, Christophe Klopp</u> *+ FAANG B&DA*

# Merci pour votre attention
## (Join FAANG!)

## - COMMON DISCUSSION -