

FAANG

Functional Annotation of Animal Genomes

Updates on Ongoing FAANG Activities

Elisabetta Giuffra

Animal Genetics Div., INRA, France



elisabetta.giuffra@juy.inra.fr

Contents

1. Pilot and FAANG-related projects
2. Main facts from FAANG Committees
3. Networking actions and Data Coordination Centre
4. New communities joining FAANG: the Salmonidae consortium (by Caird E. Rexroad, ARS-USDA)

Future ISAG-related plans

1. Pilot and FAANG-related projects



Premise:

FAANG Phase 1:

Obtain a biological reference for main farm species (i.e. with sufficiently annotated genomes); tissues from adult animals

FAANG Phase 2:

‘Expansion’: different developmental stages, different physiological conditions, additional species...

‘Genotype-to-phenotype’: new breeds/phenotyped populations (involving industry)

(Andersson et al. Genome Biol. 2015 Mar 25;16:57)



1. Pilot and FAANG-related projects

FAANG-pilot: few animals, several tissues/assays; connected to ongoing research

FAANG-related: several animals, usually less assays; part of ongoing research

Species	Leading Inst. & country		FAANG- FAANG-	
			pilot	rel.
pig, cattle, goat, chicken	INRA	France	X	x
pig, cattle, chicken	UC-Davis	US	X	x
horse	UC-Davis & Nebraska Univ.	US	X	x
cattle	Leibniz Inst. for Farm Anim. Biology	Germany	X	x
Sheep	Int. Sheep Genome Consortium & others	Australia & US	X	x
pig, chicken	Wageningen Univ.	The Netherlands	X	X
cattle	Dairy Futures Coop. Res. Centre, AgriBio & others	Australia	X	X
cattle, pig	Alberta & Guelph Univ. & others	Canada	x	X
cattle and zebu	Adelaide Univ.	Australia	x	X
sheep, buffalo	Roslin- Edinburgh Univ.	UK	x	X
cattle	Washington State Univ.	US	x	X
'Arctic Ark'	Natural Resources Institute	Finland		X

1. Pilot and FAANG-related projects



Fr-AgENCODE (coord. INRA)

[http://faang.org/bbsdb/PAG2016/Giuffra Epigen workshop.pdf](http://faang.org/bbsdb/PAG2016/Giuffra_Epigen_workshop.pdf)

P3038 (ATAC-seq, data collection completed, 4 species)

Hi-C almost completed; RNA-seq and small RNA-seq in progress;

Fr-AgENCODE related: Muret et al. “Long noncoding RNA repertoire in chicken liver and adipose tissue” – subm.

Functional Annotation of Regulatory Elements in Livestock Species (coord. UC-Davis)

[http://faang.org/bbsdb/PAG2016/Zhou Epigenetics2016.pdf](http://faang.org/bbsdb/PAG2016/Zhou_Epigenetics2016.pdf)

P1043

P2024 and P2025 (chicken)

1. Pilot and FAANG-related projects

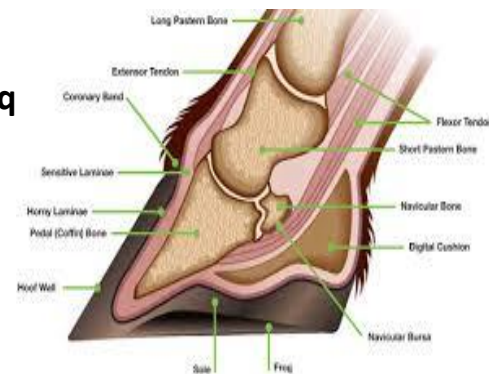
Updates since presentations at PAG XXIV:

http://faang.org/bbsdb/PAG2016/Eg_FAANG_associ_proj_2016.pdf

Species	Leading Inst. & country		FAANG	
			-pilot	-rel.
pig, cattle, goat, chicken	INRA	France	X	x
pig, cattle, chicken	UC-Davis	US	X	x
horse	UC-Davis & Nebraska Univ.	US	X	x
cattle	Leibniz Inst. for Farm Anim. Biology	Germany	X	x
sheep	Int. Sheep Genome Consortium & others	Australia & US	X	x
pig, chicken	Wageningen Univ.	The Netherlands	X	X
cattle	Dairy Futures Coop. Res. Centre, AgriBio & others	Australia	X	X
cattle, pig	Alberta & Guelph Univ. & others	Canada	x	X
cattle and zebu	Adelaide Univ.	Australia	x	X
sheep, buffalo	Roslin- Edinburgh Univ.	UK	x	X
cattle	Washington State Univ.	US	x	X
'Arctic Ark'	Natural Resources Institute	Finland		X

Equine Functional Annotation

- **Samples:** 2 healthy 5-yr old Thoroughbred mares
- **50 tissues to be collected**
 - July 28 and August 4, 2016
 - Full histology performed on each tissue
- **8 prioritized tissues for assays:** Skeletal muscle, liver, ovary, cerebral cortex, spleen, heart, lung, and laminae
- **Assays:**
 - **FUNDED:** WGS, RNA-Seq, small RNA-Seq, CHIP-seq (H3K4me3, H3K4me1, H3K27me3, H3K27ac)
 - **FUTURE PROPOSAL:** CTCF-sequencing, DNase-I hypersensitivity assays
- **Funding:**
 - Grayson Jockey Club Foundation (2016-2018)
 - USDA NRSP-8 National Animal Genome Research Program equine coordinator's funds



Sunday , July 24, 2016

14:00 Workshop: Horse genetics and genomics, Grand Ballroom A

Genetic variations associated with feed efficiency and methane yield in beef cattle

PI: **Graham Plastow, Univ. of Alberta** with Angela Canovas, Univ. of Guelph and others)

- **Collaboration with the lead groups in the FAANG consortium** (at UC Davis, INRA and the Roslin Institute) to develop a reference annotation for cattle.
- **Helping develop the sample and assay protocols** to expand the characterization of regulatory elements (and chromatin context) in beef cattle (**this will integrate Canadian efforts to expand the information available on different biological states and development stages**); and
- **Collecting samples from Canadian beef animals** to create reference information for feed efficiency and methane yield.

1. Pilot and FAANG-related projects

By Graham Plastow: plastow@ualberta.ca



Existing projects – now underway

Genome Canada Large Scale Applied Research Projects on:

- Dairy Feed Efficiency (PIs: Filippo Miglior, CDN & U of Guelph, and Paul Stothard, U of Alberta)
- Pig Disease Resilience (PIs: Mike Dyck (U of Alberta), John Harding (U of Saskatchewan), Bob Kemp (PigGen Canada))
- Samples and data for FAANG Phase 2
- Collaboration with Roslin Inst.



GenomeCanada

1. Pilot and FAANG-related projects



Species	Leading Inst. & country		FAANG- pilot	FAANG- rel.
pig, cattle, goat, chicken	INRA	France	X	x
pig, cattle, chicken	UC-Davis	US	X	x
horse	UC-Davis & Nebraska Univ.	US	X	x
cattle	Leibniz Inst. for Farm Anim. Biology	Germany	X	x
sheep	Int. Sheep Genome Consortium & others	Australia & US	X	x
pig, chicken	Wageningen Univ.	The Netherlands	X	X
cattle	Dairy Futures Coop. Res. Centre, AgriBio & others	Australia	X	X
cattle, pig	Alberta & Guelph Univ. & others	Canada	x	X
cattle and zebu	Adelaide Univ.	Australia	x	X
sheep, buffalo	Roslin- Edinburgh Univ.	UK	x	X
cattle	Washington State Univ.	US	x	X
'Arctic Ark'	Natural Resources Institute	Finland		X

1. Pilot and FAANG-related projects



Project funded by the German Research Council (DFG)



Aims ❖ **Identification of the repertoire of long noncoding RNAs (lncRNAs) in liver, skeletal muscle, intestine and rumen of phenotypically divergent male and female cattle**



Date of start **March 2016**

Methodology ❖ Deep metabolic characterization and phenotyping of animals from a dairy type x beef type cattle cross population

- ❖ Global transcriptome sequencing (RNA-seq) of liver, skeletal muscle, intestine, rumen
- ❖ Genome-wide mapping of lncRNA-interacting DNA-regions for selected lncRNAs potentially specific for metabolic types using ChIRP-seq (Chromatin Isolation by RNA Purification Sequencing)

Expected results

- ❖ Deeper functional annotation of the bovine genome and transcriptome
- ❖ lncRNAs representing central regulatory nodes, where changes in energy metabolism are sensed and adaptive gene expression is modulated specifically

Contact **Christa Kühn & Rosemarie Weikard**
Leibniz Institute for Farm Animal Biology (FBN)
Dummerstorf, Germany



kuehn@fbn-dummerstorf.de
weikard@fbn-dummerstorf.de

1. Pilot and FAANG-related projects



Participation in the FAANG Consortium

Date of start March 2016

Methodology ❖ **Global transcriptome sequencing** (RNA-seq) of >20 bovine tissues from 1 male and 1 female animal from a German Holstein x Charolais cross population with deep phenotypes

Female: 30 d in lactation

Male: Adult

Tissues: Liver, pancreas, lung, spleen, cardiac muscle, skeletal muscle, intestine, jejunum, ileum, duodenum, colon, rumen, kidney, visceral fat, subcutaneous fat, skin, pituitary gland, thyroid gland, adrenal cortex, cerebral cortex, cerebellum, hypothalamus, lymph node, uterus/testes, mammary gland...

Status: Libraries for category 1 tissues (liver, skeletal muscle, visceral fat, lung, spleen, uterus/testes) are ready for sequencing

❖ **Selection of a second method for genome-wide sequencing of the samples, e.g.,** CHIP-seq, ATAC-seq



Contact

Christa Kühn & Rosemarie Weikard
Leibniz Institute for Farm Animal Biology (FBN)
Dummerstorf, Germany



kuehn@fbn-dummerstorf.de
weikard@fbn-dummerstorf.de

Sheep annotation: Sample Collection

- Blood collected from donor ewe in 2015. Used for *de novo* PacBio reference assembly construction.
- Same donor ewe and sister used for extensive tissue collection in 2016 for FAANG assays.
- Samples were collected from 101 tissues
 - Hi-C data for scaffolding the genome
 - PacBio IsoSeq - ~12 tissues planned
 - Illumina RNASeq - ~65 tissues planned
 - ATAC-Seq – tissues collected
 - Other assays to annotate functional elements
 - Intended to meet FAANG guidelines
 - Most frozen within 45 min
 - Some brain samples within 1.5 hours
 - Three aliquots for most tissues
 - USDA MARC
 - USU
 - ATAC-Seq



The donor ewe Benz2616 and her sister. Photo Dave Forrester, USU.

30 people: Ovine FAANG participants Noelle Cockett, Alisha Massa, Brian Sayre, Michelle Mousel & Brenda Murdoch. USU pathologists and veterinarians: Tom Baldwin, Rusty Stott, Arnaud Van Wettere, Gordon Hullinger, Jaqueline LaRose, Holly Mason, and Kerry Rood. Codie Durfee processed DNA for ATAC-seq assays. Tracy Hadfield organized the collection.

FAANG planning: Kim Worley, Stephen White, Brian Dalrymple, James Kijas, Tim Smith, and Mike Heaton.

1. Pilot and FAANG-related projects

ACKNOWLEDGEMENTS



Baylor College of Medicine®

GIVING LIFE TO POSSIBLE

FAANG Collection: Tracy Hadfield, Tom Baldwin (USU), Rusty Stott (USU), Stephen White, Arnaud Van Wettere (USU), Gordon Hullinger (USU), Holly Mason (USU), Jaqualine LaRose (USU), Dave Forrester (USU), Angie Wersal (USU), Codie Durfee (ARS, ADRU), and Brenda Murdoch (Utah), and Brian Sayre (VSU), Corey Wareham (USU), Alisha Massa (WSU), Maria Herndon (WSU), Caylee Birge (ARS, ADRU), Michelle R. Mousel (ARS, ADRU), Sarah Behunin, Kara Thornton, Rachael Christianson, Nicole, Angie Robinson, Dallin Wengert, Kerry Rood, Erica Moscoso, Rickie Warr, Dustin Kinney, Abbey Benninghoff, Sumira Phatak, Kevin, Branden, Irena Polejaeva, Misha Regouski, Noelle Cockett.

Stephen White (USDA)
Brian Dalrymple (CSIRO)
James Kijas (CSIRO)
Timothy Smith (USDA)
Mike Heaton (USDA)
Kim Worley (BCM)
Noelle Cockett (USU)

USDA National Institute of Food and
Agriculture Agriculture and Food
Research Initiative 2013-67015-21228, 2013-
67015-21372 and
The International Sheep
Genome Consortium.



Qingchang Meng
Yi Han

Donna M. Muzny
Eric Boerwinkle
Richard A. Gibbs



NIH NHGRI Large-Scale
Sequencing Program (U54
HG003273)

1. Pilot and FAANG-related projects



WASHINGTON STATE
UNIVERSITY



Genome wide mapping of alternative polyadenylation sites in cattle

- Funded by: USDA/NIFA 2016-67015-24470.
- Start data: 01/01/2016
- **Goal:** To develop bovine alternative polyadenylation site (APS) resources and tools for: 1) supporting the FAANG initiative; 2) improving gene functional characterization; and 3) promoting whole genome selection in cattle.
- **Methodology:** WTTS-seq (whole transcriptome termini site sequencing).
- P1044 (see also Zhou X. et al. Genetics. 2016 Jun;203(2):683-97)
- **Contact: Dr. Zhihua Jiang**, Professor of Comparative Genomics, Department of Animal Sciences and Center for Reproductive Biology, Washington State University, Veterinary and Biomedicine Research Building, Room 151, Pullman, WA 99164-7620. Phone: 509-335 8761; Email: jiangz@wsu.edu

1. Pilot and FAANG-related projects



Functional annotation to understand how imprinting affects phenotypes in pure bred and hybrid cattle

- Adelaide lead - funded by the Davies Research Centre.
- **People involved: Stefan Hiendleder, Cindy Bottema, Tong Chen and John Williams**
- Overview: **Seventy-three** pure-bred Angus (**Bos taurus**) and Brahman (**Bos indicus**) and **reciprocal cross foetuses** were generated after standard oestrous cycle synchronization and recovered at Day 153 of gestation (Xiang et al. 2014). Tissues were collected and stored both in RNA later and as snap frozen. This material is now being used to examine **gene expression, small non-coding RNA, chromatin structure and epigenetic modifications**. Expression will be analysed for messenger RNA and long non-coding RNA by PACIFIC BIOSCIENCES sequencing to capture full length transcripts and reveal variant transcripts, while Illumina short read sequence will reveal expression levels, allele specific expression, ncRNA and small RNA. **DNA methylation** patterns will be explored using whole genome bisulphite treatment and Illumina 50bp short reads. **Chromatin structure** will be probed using the Assay for Transposase Active Chromatin (**ATAC-seq**, Buenrostro et al 2015). **Analyses of Liver and muscle samples is underway.**
- **Contacts:** stefan.hiendleder@adelaide.edu.au, cynthia.bottema@adelaide.edu.au, tong.chen@adelaide.edu.au, john.williams01@adelaide.edu.au,

1. Pilot and FAANG-related projects



Gene Expression Atlas Projects

- Sheep – Scottish Blackface x Texel
 - **Metadata for 2340 tissue specimens uploaded to Biosamples**
 - **72 total RNA (100M read depth)** Illumina TruSeq libraries for imminent upload to the European Nucleotide Archive (ENA)
- Buffalo – Mediterranean and Indian breeds
 - **Metadata for 913 tissue specimens uploaded to Biosamples**
 - **70 total RNA (100M read depth)** Illumina TruSeq libraries for imminent upload to the European Nucleotide Archive (ENA)

Contacts: emily.clark@roslin.ed.ac.uk (sheep) and rachel.young@roslin.ed.ac.uk (buffalo)

2. Main facts from FAANG Committees

Animals, Samples, Assays (ASA)

Elisabetta Giuffra and Huaijun Zhou

- Open access implemented to 16 (so far) samples' collection and assays' protocols: <ftp://ftp.faang.ebi.ac.uk/ftp/protocols/>
- A database and Sample Collection Schedules organized and set up (H. Zhou with Z. Hu, Iowa State University)
- Since April 2016: hosts the meetings of Working Group 2 of 'FAANG Europe' COST Action (http://www.cost.eu/COST_Actions/ca/CA15112) (E. Giuffra and Bo Thomsen)
- Work is ongoing for common use of cell lines in FAANG (next 2 slides)

2. Main facts from FAANG Committees

By Richard Crooijmans: richard.crooijmans@wur.nl

Cell lines at Wageningen University

- Porcine embryonic epithelial cell line IPEC-J2
 - WGS*
 - WGBS*
 - RNA-seq*
 - Histone marks* (H3K4me3, H3K27me3, H3K27Ac, H3K4Me1, CTCF)
 - Hi-C**
 - ATAC-seq**



Analyse with Bowtie2/ MACS2

Example: results H3K4me3



*= being analysed at the moment

** in preparation

2. Main facts from FAANG Committees

By Richard Crooijmans: richard.crooijmans@wur.nl

Cell lines at Wageningen University

- Chicken embryonic epithelial cell line

SL-29

- WGS*
- WGBS*
- RNA-seq*
- Histone marks (H3K4me3, H3K27me3, H3K27Ac, H3K4Me1, CTCF)
- Hi-C*
- ATAC-seq*

* in preparation



Metadata and Data Sharing Committee (M&DS)

Laura Clarke and Carl Schmidt

- Metadata standards have been defined
 - Sample
 - Experiment
 - Analysis

<http://www.github.com/faang/faang-metadata>

- Validation tools also exist

<http://www.ebi.ac.uk/vg/faang/validate>

- Full instructions can be found on

<http://tinyurl.com/submit-faang-data>

M&DS: FAANG Samples

- 819 FAANG sample records in BioSamples (<https://www.ebi.ac.uk/biosamples/>)
- 7 submitting centers
- 3 species
 - 105 chickens
 - 10 cows
 - 6 pigs
- 131 different tissues or cell types

Bioinformatics and Data Analysis Committee (B&DA)

James Reecy and Mick Watson

- RNA Analysis working group – Lel Eory
- Methylation working group – Ole Madsen
- Structural working group – Sylvain Foissac
- DNA binding working group – Pablo Ross

RNA Analysis

- Assays: RNA-seq, small RNA-seq, CAGE
- Current activities
 - RNA-seq: compare mappers (HISAT, STAR, TopHat, GSNAP) and assemblers (Cufflinks, Stringtie, Ensembl, Augustus)
 - CAGE: compare pipelines based on bwa + Paraclu/CAGEr
 - Expression: RSEM, kallisto
 - Functional categorisation: protein-coding, lncRNA, miRNA
- Pipelines
 - <https://www.ebi.ac.uk/seqdb/confluence/display/FAANG/RNA-Seq>

Methylation

- Current activities:
 - Testing alignment mappers and methylation callers.
- Pipelines:
 - <https://www.ebi.ac.uk/seqdb/confluence/display/FAANG/Bisulfite+Sequencing+%28BS%29+pipeline>
 - <https://github.com/FAANG/faang-methylation/tree/master/workflowbs>

2. Main facts from FAANG Committees



DNA Binding

<https://www.ebi.ac.uk/seqdb/confluence/display/FAANG/ChIP-Seq>

- **ChIP-Seq data acquisition standards**
 - Sequencing: Single end 50bp
 - Read coverage: Narrow peaks: 20 million; Broad peaks: 40 million uniquely mapped reads
 - Input from same sonication batch
 - Biological replicates (at least 2 – same sex)
- **Pipelines**
 - ENCODE ChIP-Seq pipeline: <https://github.com/ENCODE-DCC/chip-seq-pipeline>
 - UC Davis ChIP-Seq pipeline: <https://github.com/kernco/chipseq-pipeline>
 - EpiDB pipeline: <https://github.com/ercfrtz/epidb/blob/master/epidb.load.chipseq.pl>
- **Sample datasets**
 - Identified ChIP-seq samples passing all QC levels (by FASTQC) from the SRA/ENA public databases though EpiDB.
- **Current activities**
 - Testing pipelines using sample datasets

Structural

Assays

- DNase-seq, ATAC-seq, Hi-C
- Datasets
 - Reference datasets from published studies (human, mouse)
 - First data from FR-AgENCODER pilot project
 - (cow, pig, chicken, goat)
- Current activities
 - Gathering software and user experience
 - Testing and optimizing ATAC-seq and Hi-C pipelines

<https://www.ebi.ac.uk/seqdb/confluence/display/FAANG/Chromatin+Structure>

3. Networking actions and Data Coordination Centre

Leading Inst. & country		<i>Networking Actions</i>	DCC
Delaware Univ.	US	X	
Roslin- Edinburgh Univ.	UK	X	
EMBL-EBI	UK		X

International networking and cooperation



Animal Genome to Phenome Research Coordination Network (AG2P)
Carl Schmidt



Functional Annotation of Animal Genomes - European network (FAANG Europe)
Alan Archibald

http://www.cost.eu/COST_Actions/ca/CA15112
<http://ag2p.org>



FAANG Data Coordination Center

- Grant awarded by BBSRC
- EMBL-EBI and Roslin institute collaboration
- **Starts November 1st**
- **Will provide**
 - Submission and validation support
 - Data portal
 - Standard analysis results

EMBL-EBI can provide submission support now

Overall:

- ‘ENCODE-like’ funded projects would speed up obtaining the biological references of genomes of different species (in support of several research topics) – but...
- In this evolving scenario, means for global coordination represent a key need (interconnected Data Analysis Centers; expansion to additional groups and connection to related communities; etc.)

Fill the online questionnaire please!

4. New communities joining FAANG

Functional Annotation of All Salmonid Genomes (FAASG)

Caird E. Rexroad

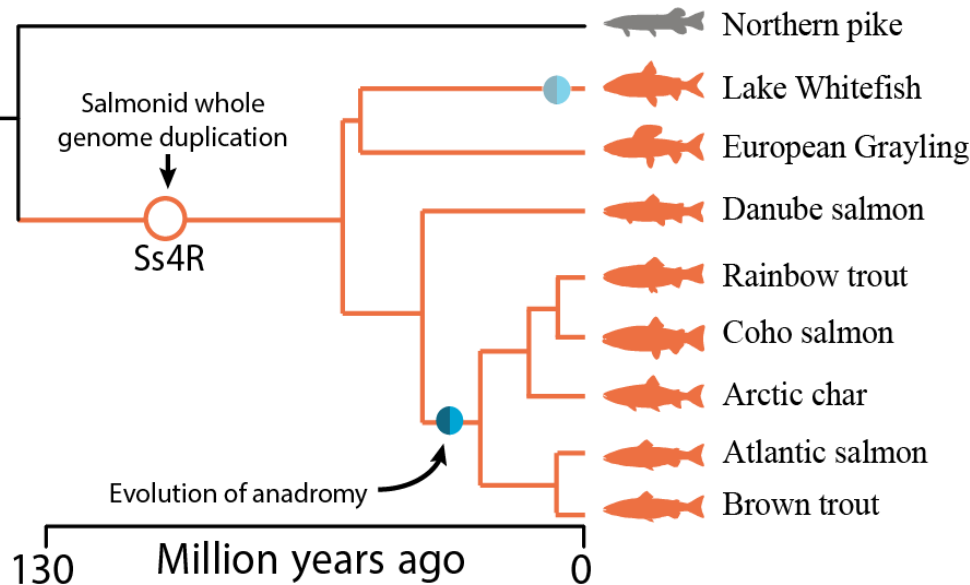
ARS-USDA

Functional Annotation of All Salmonid Genomes (FAASG)

Opportunities

Develop comparative data sets to understand how genome functions evolve after whole genome duplication events in vertebrates and how genomic variation and expression translates into phenotypes

Develop new approaches and resources to facilitate the creation of solutions that enhance sustainable salmonid fisheries, conservation and aquaculture production worldwide

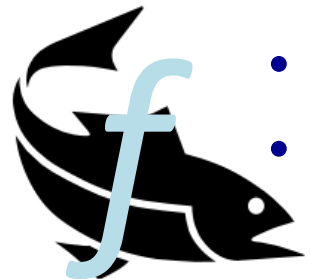


Approach: Coordinate the international salmonid community to acquire, standardize and share data for comprehensive mapping and characterization of the functional elements of salmonid genomes.



Toronto Workshop, June 13-14, 2016

- **Objective:** Discuss opportunities for an international collaboration afforded by the availability of reference salmon genomes
- **Format:** Participants from 19 institutions in ten countries including researchers and funding agency representatives
 - Set the scene 'state of salmonid research'
 - FAANG overview
 - Round Table: Towards a functional annotation of salmonids initiative (FAASG)
- **Next Steps:**
 - Workshop Summary Report and White Paper
 - Testing of core assays
 - Engagement with FAANG and salmonid community
 - Workshop on margins of PAG (2017)



www.faasg.org coming soon