MCDB 4100
Experimental Design & CRISPR Mutagenesis in *Xenopus*

Prof. Michael Klymkowsky

michael.klymkowsky@colorado.edu

Dr. Bilge Birsoy

bilge.birsoy@colorado.edu

Office hours: TBD depending on your response

Location: TBD

CRISPR-Cas9 Mutagenesis in *Xenopus*

A gene

A mutation

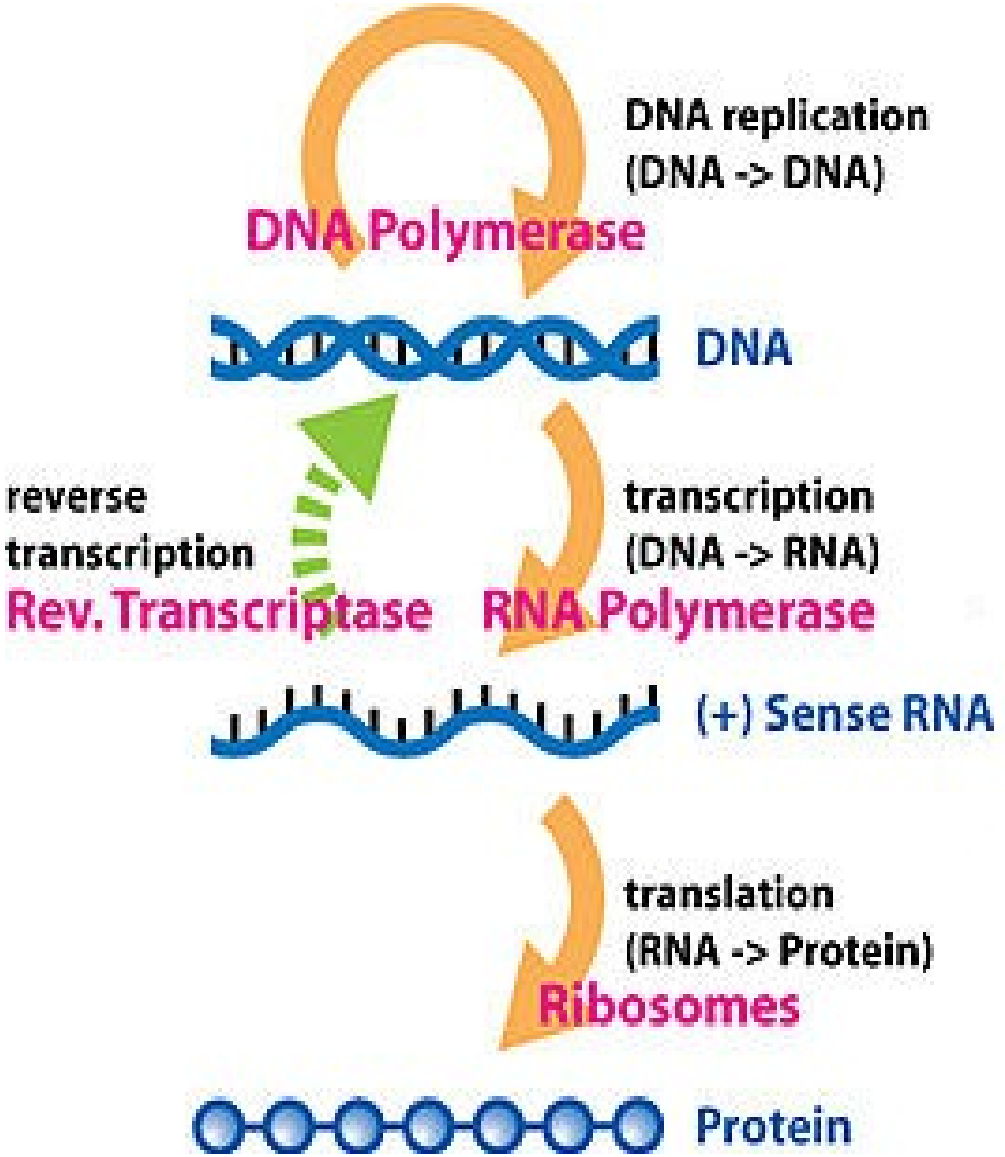
A gene is

basic physical and functional unit of heredity

Coding RNA for proteins

Non-coding RNA (genes for tRNA, rRNA, miR)

Central Dogma



Gene Structure Activity

Working as a group, draw a gene structure to include <u>all</u> the terms

Enhancer

Promoter

Transcription initiation site

Exon

Intron

5' UTR

3'UTR

PolyA signal

Transcription termination site

Translation initiation site

Translation termination site

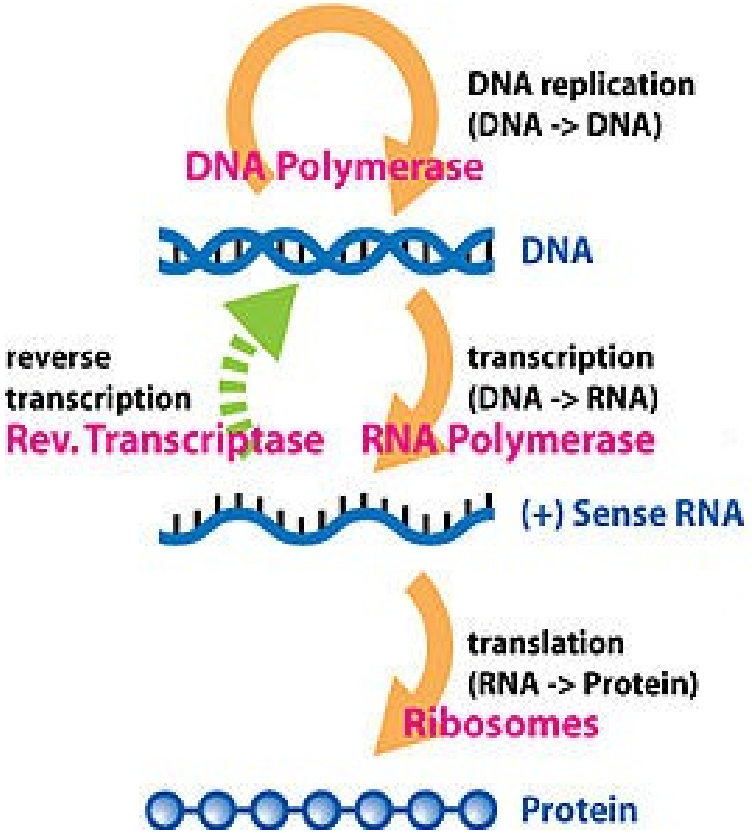
ORF (open reading frame)

Pre-mRNA vs mature mRNA activity

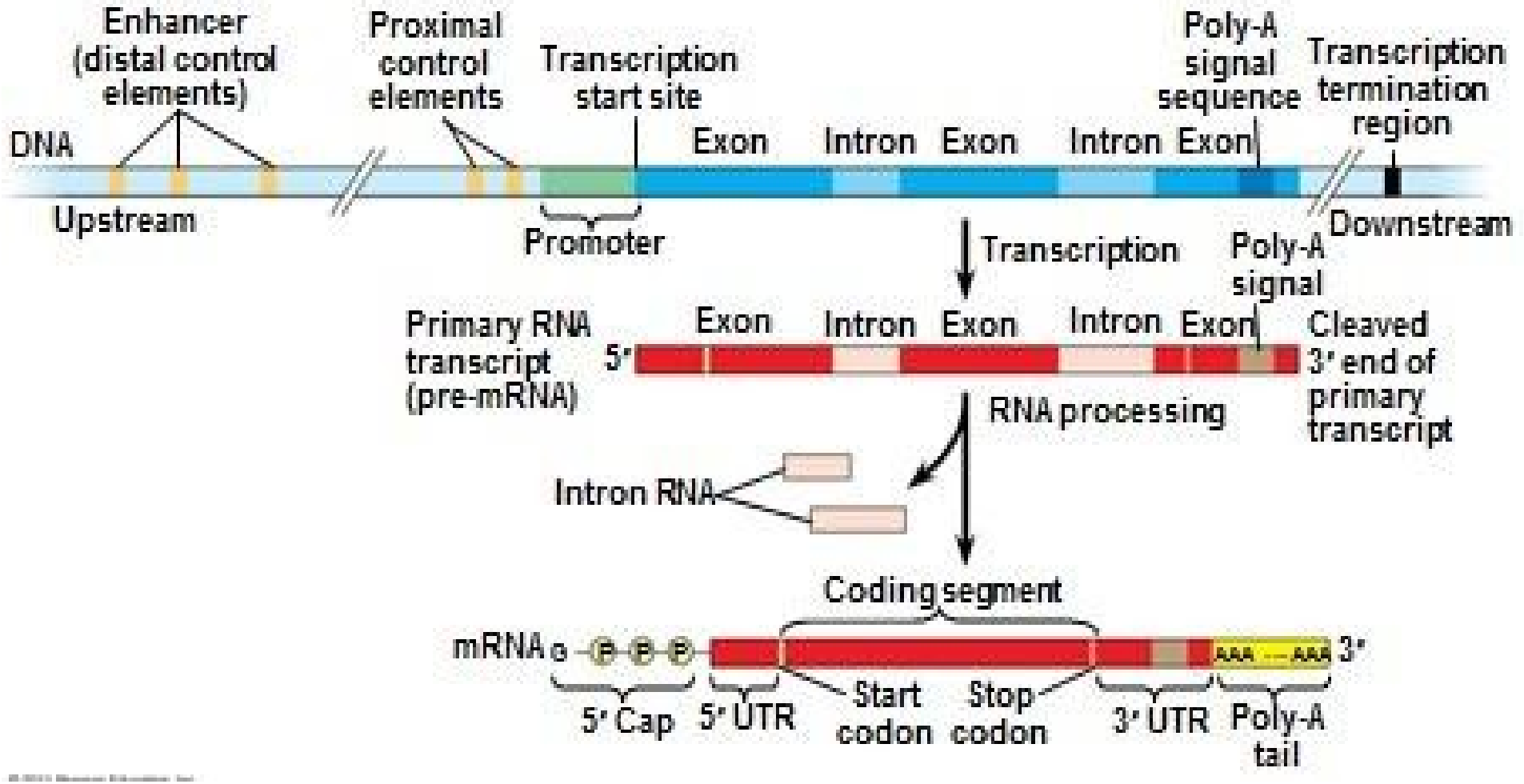
As a group, draw one pre-mRNA and its mature mRNA and compare the differences

Some Terms to Know

- gDNA: **g**enomic DNA
- cDNA: **c**omplementary DNA
- mRNA: **m**essanger RNA
- UTR: **U**n**T**ranslated **R**egion
- Probe: a molecule that specifically binds a nucleic acid or peptide sequence



Eukaryotic Gene Structure



Exon, UTR, ORF

Of the ~24,500 RefSeq from the human genome

~8,500 (35%) RefSet mRNAs have at least one intron within the 5' UTR (more than 90% have a single intron within the 5'UTR)

5% have one intron within the 3' UTR

Translation initiation is typically the first AUG following the 5'UTR but not always

Translation may start at the second AUG or later.

Translation may even start at a non-AUG codon (such as CUG or AUU, 54 genes in the human genome, which are still recognized by initiator tRNA that is bound to Met amino acid).

Translation Initiation Site is determined by many factors including 5'UTR structure, optimal context (Kozak sequence around the AUG)

Activity

a) 5'-**TATAAA**NNNN---AGC ATG GAT CAT----------TCA TGC TGA GAC---AATAAA---3'

b) 3'-**ATATTT**NNNN---TCG TAC CTA GTA----------AGT ACG ACT CTG---TTATTT----5'

- Which DNA strand is coding strand, which is template strand?

- Write down mature mRNA sequence.

- Write a short probe sequence to hybridize with endogenous mRNA

Activity

a) 5'-**TATAAA**NNNN---AGC ATG GAT CAT----------TCA TGC TGA GAC---AATAAA---3'

b) 3'-**ATATTT**NNNN---TCG TAC CTA GTA----------AGT ACG ACT CTG---TTATTT----5'

- Coding DNA strand (a) is the same sequence as mRNA (except for presence of T as opposed to U in RNA) and contains TATA box. Template strand (b) is the strand used for mRNA transcription.
- Write down mature mRNA sequence.

5' cap----AGC AUG GAU CAU----------UCA UGC UGA GAC----AAUAAA-----AAAAAAAA+

- Write a short specific RNA probe to hybridize with endogenous noggin mRNA.

3'---UCG UAC CUA GUA----------**AGU ACG ACU CUG**---UUAUUU----UUUUUUUU---5'

Allele, Mutation, Polymorphism

An allele is a variant/alternative form of a gene. For each gene, we inherit one allele from our mother and one from our father. For most genes (e.g. coding structural proteins), alleles are identical in DNA sequence. There are also many genes that show variations in alleles (e.g. blood groups, liver enzymes, eye color)

A mutation is a change in the DNA sequence. A mutation may be a change of one or multiple nucleotides (e.g. point mutation or large deletion) but it is a change that is rare (<1%) in the population.

A polymorphism is a DNA sequence variation (a mutation) that is observed in at least 1% of the population. Most polymorphisms have no affect on development or health. Humans are about 0.1% polymorphic (1 variation in 1000 nucleotides, this means 3 million variant sequences in 3 billion nucleotides of the human genome). 50-90% of human polymorphisms are single nucleotide polymorphisms (SNP), 5% of SNPs are in exons.

ExAC paper says 1 variant every 8 nucleotides in the exome (not to be confused with exon)

Dominant vs Recessive

Dominant allele or dominant mutation (not a dominant gene)

Recessive does not mean its frequency is less than the frequency of its dominant allele.

There can be many genetic variations within a given gene sequence and its two alleles.

Mutation that causes

complete loss of gene function (null) **Amorph**

partial loss of gene function **Hypomorph**

increased gene function **Hypermorph**

opposing gene function (dominant negative) **Antimorph**

a new gene function **Neomorph**

Mutations by

Small scale changes

Substitution: point mutation

Deletion: may alter splicing or cause a frame shift

Insertion: may alter splicing or cause a frame shift

Inversion: may change amino acid sequence or introduce a STOP codon

Large scale changes

Amplification

Deletion

Translocation

Mutations

Point mutation: change of a single nucleotide (a base substitution)

Silent: causes no change in amino acid function

Missense: changes one amino acid to another with a different structure and function

Nonsense: changes the codon from an amino acid coding codon to a STOP codon

Frame shift mutation: an insertion or deletion that alters the triplet grouping of nucleotides into codons and **shifts** the reading **frame** so that all the nucleotides downstrasm of the mutation will be improperly grouped and translated into a different protein sequence.

ExAC paper found small numbers of frameshift mutations compared with in-frame indels. What could explain this?

Activity

Work as a group at your table and write down the amino acid sequence for the given mutations using the codon table.

Second base of codon

First base of codon	U	C	A	G	Third base of codon
U	UUU Phenylalanine phe UUC	UCU Serine ser UCC UCA UCG	UAU Tyrosine tyr UAC	UGU Cysteine cys UGC	U C
	UUA Leucine leu UUG		UAA STOP codon UAG	UGA STOP codon UGG Tryptonphan trp	A G
C	CUU Leucine leu CUC CUA CUG	CCU Proline pro CCC CCA CCG	CAU Histidine his CAC	CGU Arginine arg CGC CGA CGG	U C
			CAA Glutamine gin CAG		A G
A	AUU Isoleucine ile AUC AUA	ACU Threonine thr ACC ACA ACG	AAU Asparagine asn AAC	AGU Serine ser AGC	U C
	AUG Methionine met (start codon)		AAA Lysine lys AAG	AGA Arginine arg AGG	A G
G	GUU Valine val GUC GUA GUG	GCU Alanine ala GCC GCA GCG	GAU Aspartic acid asp GAC	GGU Glycine gly GGC GGA GGG	U C
			GAA Glutamic acid glu GAG		A G

Part of tkv sequence

AUG GCC GCC GCC GCU UUG AGU GGC-----

M A A A A L S G -----

Write the amino acid sequence for

Deletion: What if the sequence GCU becomes GC?

Substitution:What would happen if UUG is mutated to UAG?

Inversion: What would happen if AGU is mutated to GAU?

Insertion: What would happen if a G is inserted after AUG such as AUG GGCC ?

Part of tkv sequence

AUG GCC GCC GCC GCU UUG AGU GGC-----

M A A A A L S G -----

Deletion: What if the sequence GCU becomes GC?

M A A A A STOP

Substitution:What would happen if UUG is mutated to UAG?

M A A A A STOP

Inversion: What would happen if AGU is mutated to GAU?

M A A A A L D G

Insertion: What would happen if a G is inserted after AUG such as AUG GGCC ? FRAME SHIFT

AUG GGC CGC CGC CGC UUU GAG UGG C

M G R R R F E W

For CRISPR-Cas9 targeting

Prefer a protein coding exon at the beginning of the sequence hoping to get a frameshift mutation

Avoid exon-intron junctions

Avoid within introns since these may contain other genes such as miRs

What makes a gene interesting to study?

Genes implicated in an interesting biological process or human disorder

Genes that are highly conserved between species (especially human vs others)

- Human perspective: Fix the 3Ds: death, disease, disability

Genes of unknown function

Practical limitations:

Gene must have an homolog or an ortholog in model organism of choice (*Xenopus laevis*)

Phenotype must be observable within the time frame of the project (semester)

Xenopus gene function must not have been published yet (DISCOVERY)

Genes that display early lethal phenotypes or housekeeping genes are clearly important genes but they are trickier to study and may require additional justification.

Interesting Biological Process

AmiGO

http://amigo.geneontology.org/amigo/

Gene Ontology (types, properties and interrelationships)

National Center for Biotec... | AmiGO 2: Welcome | ApE- A plasmid Editor

amigo.geneontology.org/amigo

ape plasmid

AmiGO 2 | Home | Search | Browse | Tools & Resources | Help | Feedback | About | AmiGO 1.8

AmiGO 2

More information on quick search

Quick search

Search

Search Templates

Use predefined **templates** to explore Gene Ontology data.

Go »

Advanced Search

Interactively **search** the Gene Ontology data for annotations, gene products, and terms using a powerful search syntax and filters.

Search

Browse the Ontology

Use the drill-down **browser** to view the ontology structure with annotation counts.

Go »

GOOSE

Use GOOSE to query the legacy GO database with **SQL**.

Go »

Term Enrichment Service

Statistics

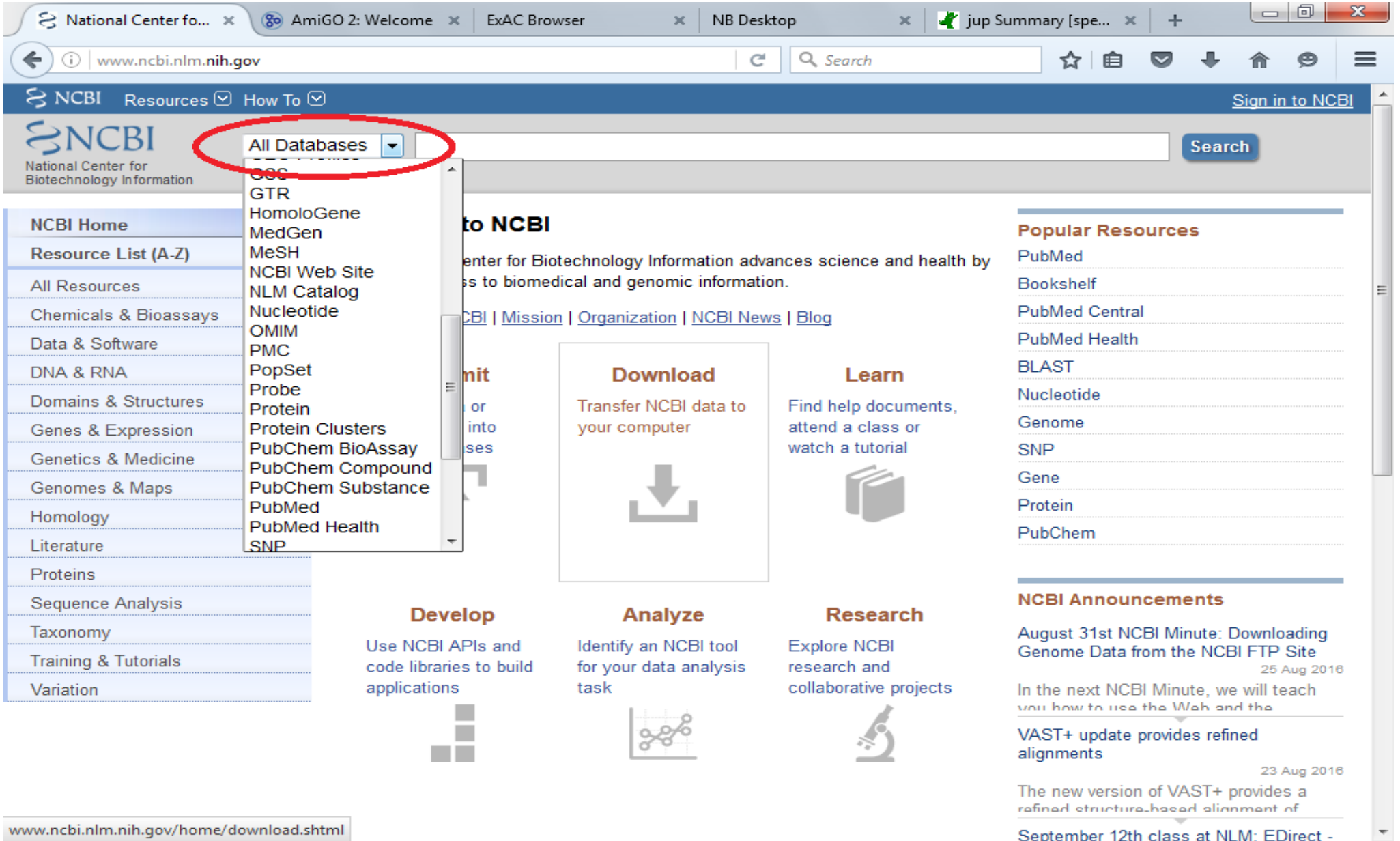
And Much More...

amigo.geneontology.org/amigo/landing

NCBI

http://www.ncbi.nlm.nih.gov/

Change the database and search for a disease or a gene



Xenbase

www.xenbase.org

To find info on

Xenopus gene expression profiles (developmental stage and tissues)

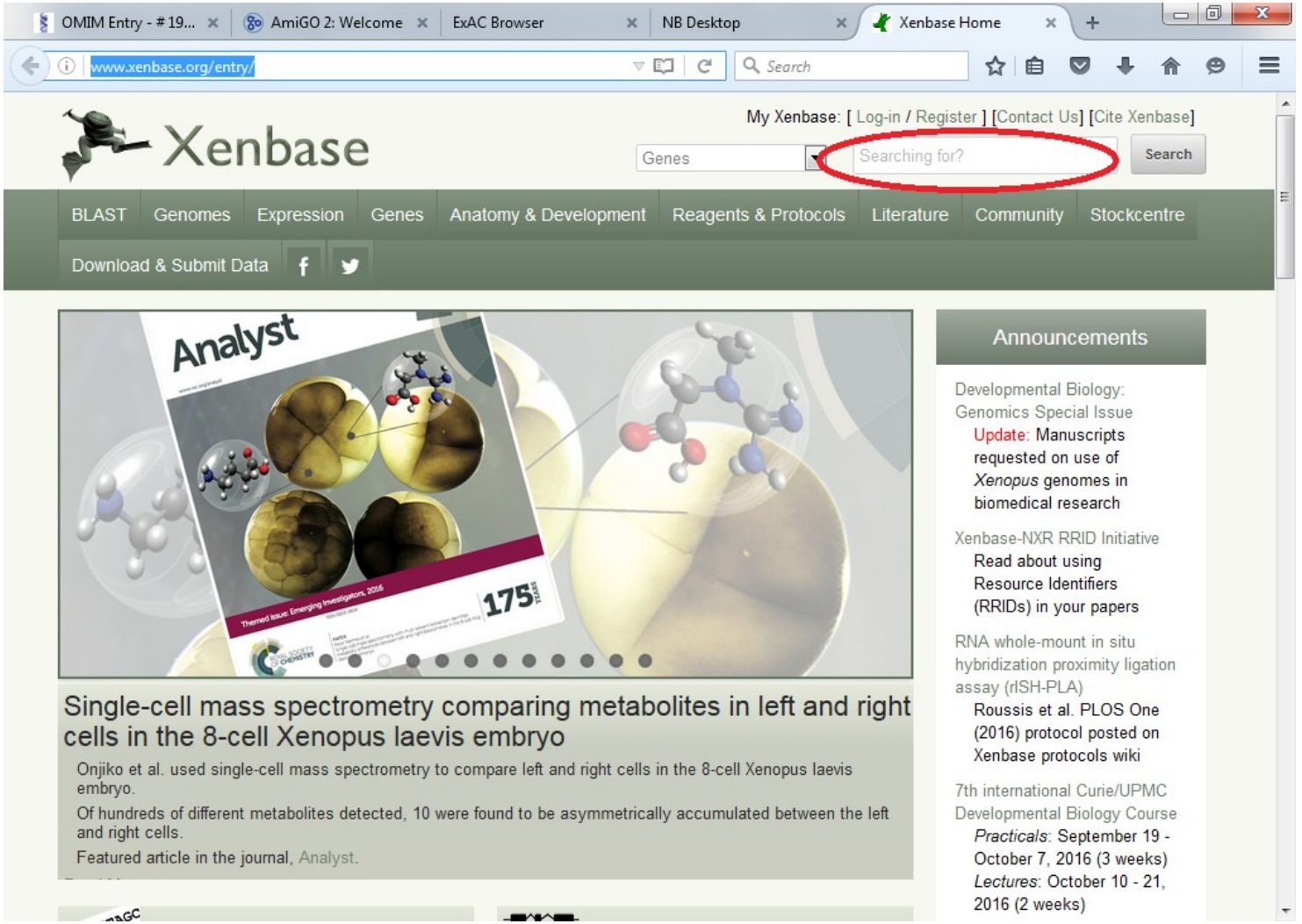
Xenopus loss-of-function or gain-of-function experiments

Xenopus experimental protocols

Xenopus developmental stages

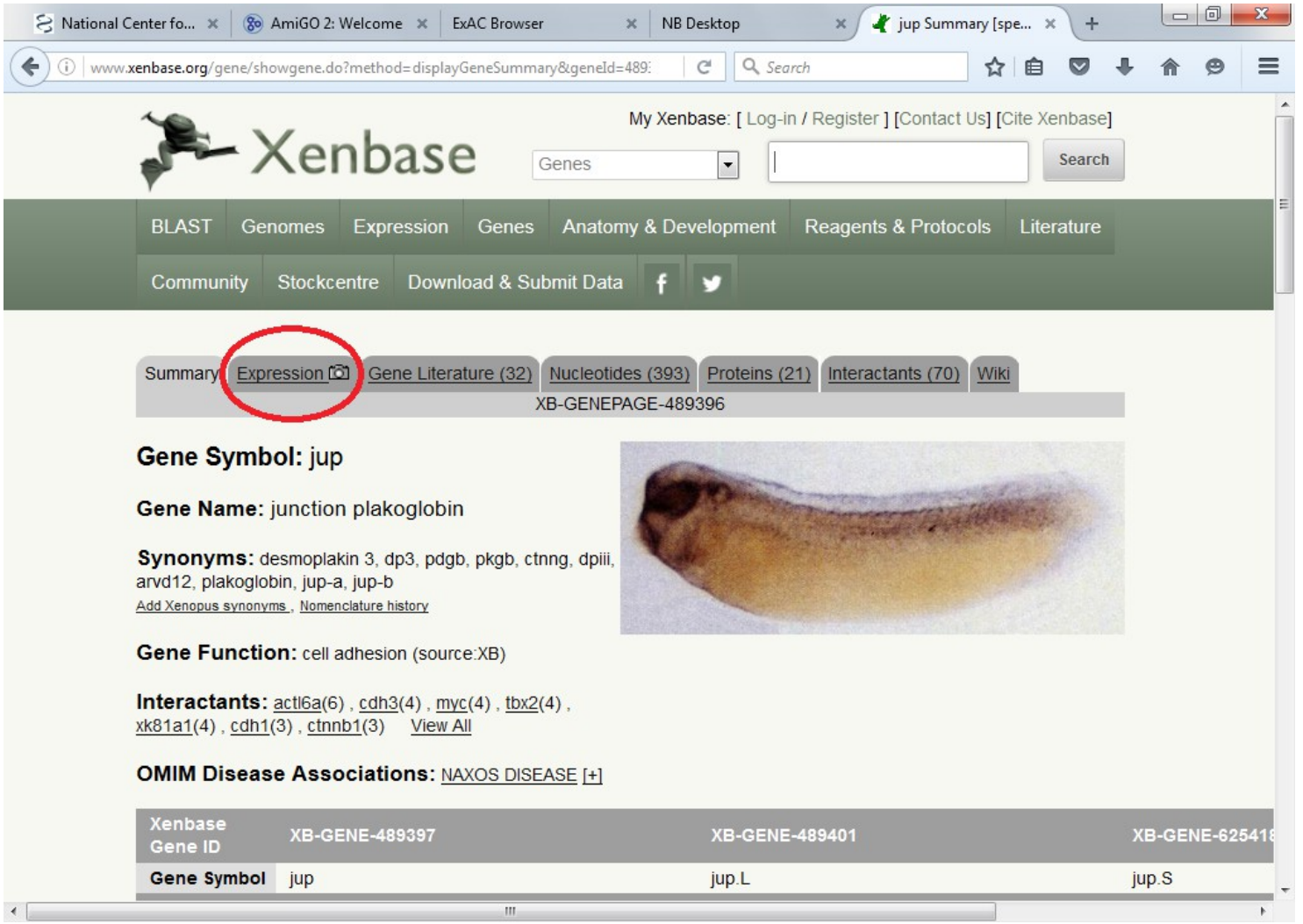
and much much more

Type in a gene name in the red oval



Page for gene “jup”

You may click on expression tab or



Or alternatively for gene expression info

Under “Expression”

Next to “Images”

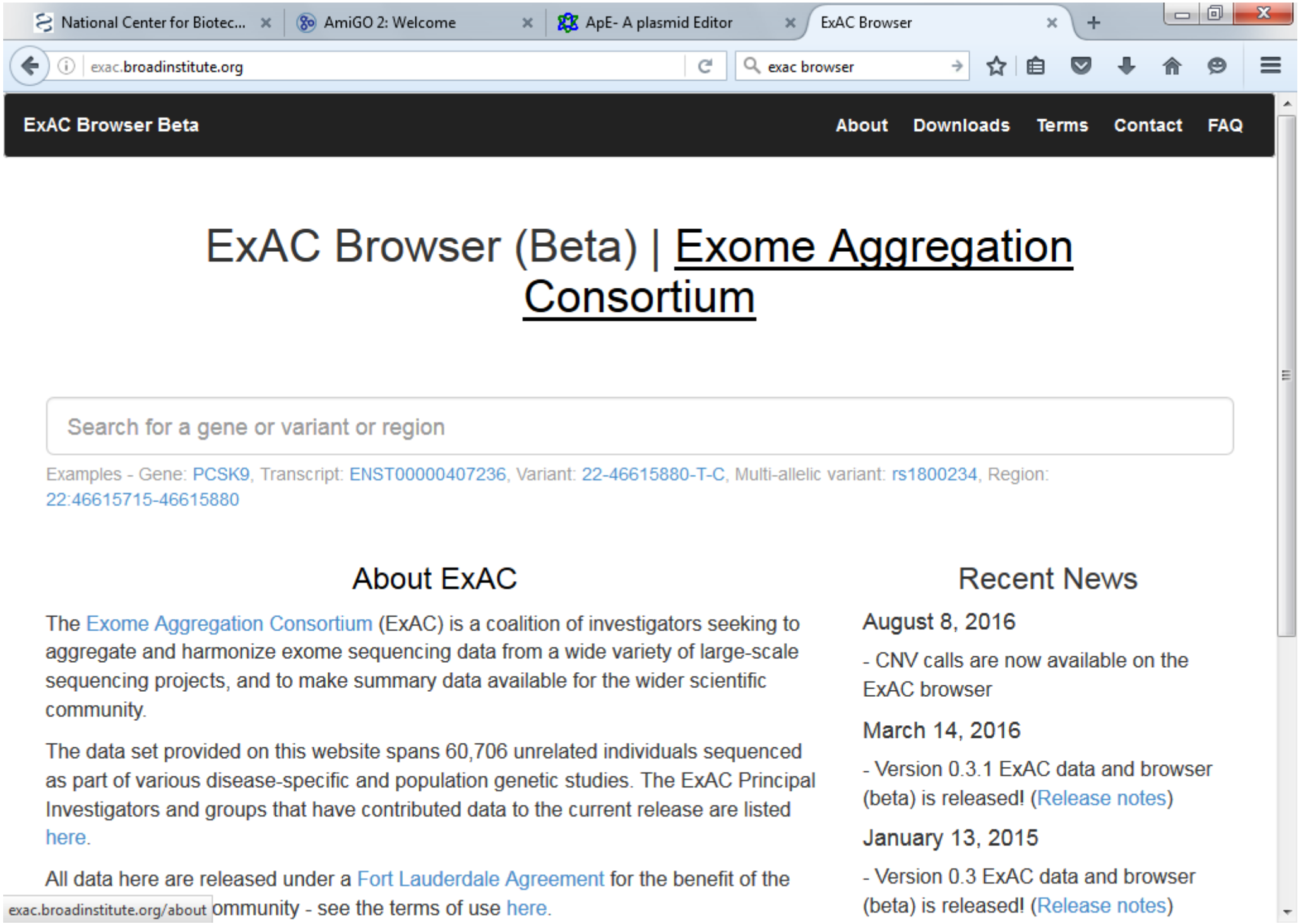
Click on the underlined images section

Links			
Genomic Synteny	Genomicus Metazome		
Expression	**tropicalis**	**laevis.L**	**laevis.S**
Ensembl	Ensembl		
UniGene	Str.5916	Xl.21696	Xl.2844
RNA-seq profile: Owens, Blitz et al. 2016	jup		
Images	2 expression image(s) for tropicalis and laevis		
Data Mining			
XenMine	jup	jup.L	jup.S
Phenotype	**tropicalis**	**laevis.L**	**laevis.S**

ExAC Browser

http://exac.broadinstitute.org/

Most reliable search term is gene accession number


National Center for Biotec...
AmiGO 2: Welcome
ApE- A plasmid Editor
ExAC Browser
exac.broadinstitute.org
exac browser
ExAC Browser Beta
About
Downloads
Terms
Contact
FAQ
ExAC Browser (Beta) | Exome Aggregation Consortium
Search for a gene or variant or region
Examples - Gene: PCSK9, Transcript: ENST00000407236, Variant: 22-46615880-T-C, Multi-allelic variant: rs1800234, Region: 22:46615715-46615880
About ExAC
The Exome Aggregation Consortium (ExAC) is a coalition of investigators seeking to aggregate and harmonize exome sequencing data from a wide variety of large-scale sequencing projects, and to make summary data available for the wider scientific community.
The data set provided on this website spans 60,706 unrelated individuals sequenced as part of various disease-specific and population genetic studies. The ExAC Principal Investigators and groups that have contributed data to the current release are listed here.
All data here are released under a Fort Lauderdale Agreement for the benefit of the
exac.broadinstitute.org/about
ommunity - see the terms of use here.
Recent News
August 8, 2016
- CNV calls are now available on the ExAC browser
March 14, 2016
- Version 0.3.1 ExAC data and browser (beta) is released! (Release notes)
January 13, 2015
- Version 0.3 ExAC data and browser (beta) is released! (Release notes)

Q: You identify an allele of a gene that produces a dominant disease phenotype. Suggest two distinct mechanisms that could produce this situation.

Q: You discover an dominant allele of a gene that produces a severe disease phenotype. Later you identify a small subset of people who have the disease-causing allele, but do not have the disease. Provide a plausible explanation for this situation?

Q: You identify a recessive allele that is present a high frequency in the population. When homozygous, the presence of this allele leads to early childhood death. Suggest a plausible mechanisms that could lead to this situation.

Q: You identify an allele of a gene that produces a recessive disease phenotype. Suggest two distinct mechanisms that could produce this situation.

Q: You compare the genomes of humans and other mammals. You discover that one set of sequences, conserved in other mammals, that have either been deleted or changed in humans. What process(es) could explain this observation?

For Tuesday

Re-do the genetics survey In light of today's discussion

Continue to search 1-2 genes that you may be interested in pursuing. Is there a human disease you are interested in? Is there a biological process you'd like to learn more about? Do you want to go after a PUF(Protein of Unknown Function)?

Search databases for interesting genes

For PubMed, NCBI (OMIM, Unigene, Blast), AmiGO, Xenbase, ExAC browser

Also download ApE on to your device

http://biologylabs.utah.edu/jorgensen/wayned/ape/

Login to Nota Bene: Read the ExAC paper

Highlight terms, concepts, methods you don't understand or want to discuss