

In vivo
**characterization of
evolutionary
conserved
noncoding
sequences**

**Nadav Ahituv
UCSF**

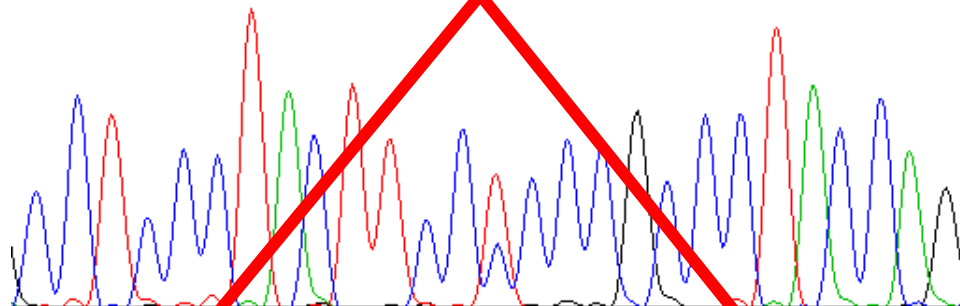
AGGAAGGGAAAGCGCAAGAGAGAGCGCACACGCGCACACCCGCGCGCGCACTCGCGCACGGAC
CCGCACGGGGACAGCTCGGAAGTCATCAGTTCCATGGCGAATGCTGCTGCTGGCGAGATGTCTG
CTGCTAGTCCTCGTCTCCTCGCTGCGGTATGCTCGGGACTGGCGTGCGGACCGGGCAGGGGGTTC
GGGAAGAGGAGGCACCCCAAAAAGCTGACCCCTTAGCCTACAAGCAGTTTATCCCAATGTGGCC
GAGAAGACCCTAGGCGCCAGCGGAAGGTATGAAGGGAAGATCTCCAGAACTCCGAGCGATTTAA
GGAACTCACCCCAATTACAACCCCGACATCATATTTAAGGATGAAGAAAACACCGGAGCGGACA
GGCTGATGACTCAGTAGGAACCCAGCGCCGGGGCGTGGAATGTGTGGCTTTCCAGGGGGTTACGA
GAAGCCGAACACTTCCAGACTTAACTCTGTTTGGCTCTTCGGGCAGATGAAGGTGATTTACCCGCTC
CTCCCCACCCACCTGCCCGCCCCCATTCTTCTCTTCTGGAGGAGAATGGAGGTCAAGGGTCC
AGCTGGAGAAGTTAGGGTGTGGTGGGGGTGAGGACGGTAACAGACGTGGTTCATTATGGCCTGATT
TGATGAGTCTTGCTACAATGGCCTTCCCCATCCTACCTCTGCCTGGCTTGTAAGTTGGGGAGACCTT
CACTTTGGGGGCGTCGGCCCTTCCAGTCAGGAGTGGAATGGAAGGAGAGGCTGGGAATCCCCC
TCCCACAAACATGAAGTGGTCTCCTGGTACTGTACGAACGAACGAACGTAGCCTTGGGCTTGGAGC
TCAGAGCCCCACGTTTCCCGTTGCCTCTGTGGTTTTCTTCCACCACTACCCCCACCTGCACCTC
CCCACCAAAGAATTCTCAACTGGAAAAGCCAGGAGGCGGTTCTGACAAAAGGCAGGGGGCTCCAGG
GGAGACTCCCCGTCCCTGGGTGGCTGGCTGTATCGCAGAGCTGGCTTTGCGATTGCGTGTCCGCA
ATTGTGCCCATCAGAGTGTGAATGTATTGATATTTCTTTAAGGATGCTCTTTCGTTCTTCCAAGCCCG
AGGTACCTTAGGGGAGGGACTTAGAACTTATTGGCATTGCATCACTTTAGTTTTCAACCTGCTTGCA
TAAGAATTAAGAGCGAATAAATATTAGTGTGGGGGGAGGGGAAGCTAAGCAAAATATGAATTCCTC
TCTCTCTCCCACCTCCTTTGAGATTTCTGAGCTGCCAATCTCCCAGCCAATTCTAGACTTTCTGAAA
CTCCATGCACGTATAACTGAAGCCAGAAATGGGTTTCCTTGCAAATATAGGTCAACATCCTTTTTTATT
GCCCTATTAATAATTCAAGTCCTACCTTTAGGGCTAGGTGCGTACAGCGGCTGATGGAGTGGCGC
TGGTGGGGCGCAAGTGCAGGGGGAGGGTACTGACGGCAGAGAGAGAGGAGCTACCTCCGTGCCG
CCCTGCTTCCCGACCCGATTCCAGGCTTGCTTGAGGCCGAGAAAGGCGAGGGGGCAGGCAAGGTA
GCCTGCTCCAGCTGTCGGAAGGGAGAGGAATGGGAAATGGTCCTGATTTCTTGCTCTCCCTCATC
TGCTCCCGACCACCTTAAATCTGGACCGCGAGTGTGGACGCGCGCGCCAGTGCCAGACAGCAGCG
CGATCCACAATTA ACTCTGCACGGGCCATGGGGTGCCCGTTGCGTGCAGCTGGCTGGAGGGAGTTC
TCCGGCTAGCCCGAGGCGCCATCCCTCTCGTCACCTCACTCCCCGCGGAGGAGGGGCCTTGCCA
GGGTCCCTCGGAACCCGAGAGGAGGGAGGCACTGCGGAGAGAGCGGGCGGGGGCGTGGATAACC
GAGGTCCAGAGCCAGAGTGGGTGAGCTTCTGACCTGCTCTGCGGGAGGCCAATACCGCAGAAGG
GGTCTGGGCTCGCACACCTTCCAGGGCTTGAGCCTTGACGCCCTGCTGCAATAACTACCCGTGA

AGGAAGGGAAAGCGCAAGAGAGAGCGCACACGCGCACACACCCGCGCGCGCACTCGCGCACGGAC
CCGCACGGGGACAGCTCGGAAGTCATCAGTTCCATGGGCGAATGCTGCTGCTGGCGAGATGTCTG
CTGCTAGTCCTCGTCTCCTCGCTGCGGTATGCTCGGGACTGGCGTGCGGACCGGGGCAGGGGGTTC
GGGAAGAGGAGGCACCCCAAAAAGCTGACCCCTTTAGCCTACAAGCAGTTTATCCCCAATGTGGCC
GAGAAGACCCTAGGCGCCAGCGGAAGGTATGAAGGGAAGATCTCCAGAAACTCCGAGCGATTTAA
GGAACTCACCCCAATTACAACCCCGACATCATATTTAAGGATGAAGAAAACACCGGAGCGGACA
GGCTGATGACTCAGTAGGAACCCAGCGCCGGGGCGTGGAATGTGTGGCTTTCCAGGGGGTTACGA
GAAGCCGAACACTTCCAGACTTAACTCTGTTTGGCTCTTCGGGCAGATGAAGGTGATTTACCCGCTC
CTTCCCACCCACCTGCCCGCCCCCATTCTTCTCTTCTGGAGGAGAATGGAGGTCAAGGGTCC
AGCTGGAGAAGTTAGGGTGTGGTGGGGGTGAGGACGGTAACAGACGTGGTTCATTATGGCCTGATT
TGATGAGTCTTGCTACAATGGCCTTCCCCATCCTACCTCTGCCTGGCTTGTAAGTTGGGGAGACCTT
CACTTTGGGGGCGTCGGCCCTTTCCAGTCAGGAGTGGAATGGAAGGAGAGGCTGGGAATCCCCC
TCCCACAAACATGAAGTGGTCTCCTGGTACTGTACGAACGAACGAACGTAGCCTTGGGCTTGGAGC
TCAGAGCCCCACGTTTCCCGTTGCTCTCTGTTTCTTCCACCCTACCCCCACCCTGCACCTC
CCCACCAAAGAATTCTCAACTGGCAAGCAAGGCGCTTCTGACAAAAGGCAGGGGGCTCCAGG
GGAGACTCCCCGTCCCTGGGTCCCGGCTATCGCAGAGCTGGCTTTGCGATTGCGTGTCCGCA
ATTGTGCCCATCAGAGTGTGAATCTAATGTAATCTTAAAGATGCTCTTTCGTTCTTCCAAGCCCG
AGGTACCTTAGGGGAGGGACTTAGAACTTATTGGCATTGCATCACTTTAGTTTTCAACCTGCTTGCA
TAAGAATTAAGAGCGAATAAATATTAGTGTGGGGGGAGGGGAAGCTAAGCAAAATATGAATTCCTC
TCTCTCTCCCACCTCCTTTGAGATTTCTGAGCTGCCAATCTCCCAGCCAATTCTAGACTTTCTGAAA
CTCCATGCACGTATAACTGAAGCCAGAAATGGGTTTCCTTGCAAATATAGGTCAACATCCTTTTTATT
GCCCTATTAATAATTCAAGTCCTACCTTTAGGGCTAGGTGCGTACAGCGGCTGATGGAGTGGCGC
TGGTGGGGCGCAAGTGCAGGGGGAGGGTACTGACGGCAGAGAGAGAGGAGCTACCTCCGTGCCG
CCCTGCTTCCCGACCCGATTCCCAGGCTTGCTTGAGGCCGAGAAAGGCGAGGGGGCAGGCAAGGTA
GCCTGCTCCAGCTGTCGGAAGGGAGAGGAATGGGAAATGGTCCTGATTTCTTGCTCTCCCTCATC
TGCTCCCGACCACCTTAAATCTGGACCGCGAGTGTGGACGCGCGCGCCAGTGCCAGACAGCAGCG
CGATCCACAATTA ACTCTGCACGGGCCATGGGGTGCCCGTTGCGTGCAGCTGGCTGGAGGGAGTTC
TCCGGCTAGCCCGAGGCGCCATCCTCTCGTCACCCTCACTCCCCGCGGAGGAGGGGCCTTGCCA
GGGTCCCTCGGAACCCGAGAGGAGGGAGGCACTGCGGAGAGAGCGGCGGGGGCGTGATAACC
GAGGTCCCAGAGCCAGAGTGGGTGAGCTTCTGACCTGCTCTGCGGGAGGCCAATACCGCAGAAGG
GGTCTGGGCTCGCACACCTTCCCAGGGCTTGAGCCTTGACGCCCTGCTGCAAAAAAAAAAAAAAAAAA

98%

Second base

		Second base							
		U	C	A	G				
First base	U	UUU } Phenylalanine F UUC } UUU } Leucine L UUG }	UCU } Serine S UCC } UCA } UCG }	UAU } Tyrosine Y UAC } UAA } Stop codon UAG } Stop codon	UGU } Cysteine C UGC } UGA } Stop codon UGG } Tryptophan W	U	C	A	G
	C	CUU } Leucine L CUC } CUA } CUG }	CCU } Proline P CCC } CCA } CCG }	CAU } Histidine H CAC } CAA } Glutamine Q CAG }	CGU } Arginine R CGC } CGA } CGG }	U	C	A	G
	A	AUU } Isoleucine I AUC } AUA } AUG } Methionine start codon M	ACU } Threonine T ACC } ACA } ACG }	AAU } Asparagine N AAC } AAA } Lysine K AAG }	AGU } Serine S AGC } AGA } Arginine R AGG }	U	C	A	G
	G	GUU } Valine V GUC } GUA } GUG }	GCU } Alanine A GCC } GCA } GCG }	GAU } Aspartic acid D GAC } GAA } Glutamic acid E GAG }	GGU } Glycine G GGC } GGA } GGG }	U	C	A	G
						Third base			



CCTCCCTACTTCC^CCCGCCCTACCAG
T

ProProTyrPheProProProTyrGln



Leu

Noncoding regulatory sequences

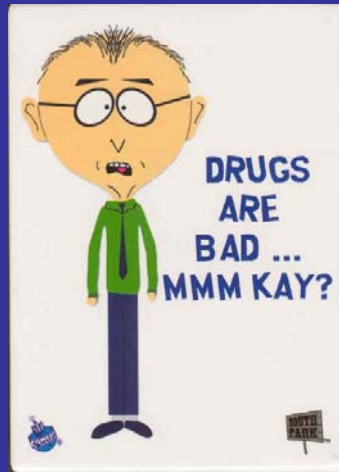
nucleotide
changes

Differences in gene expression

Human variation



Pharmacogenetics

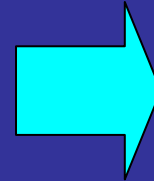


Human disease



98% of the genome

AGGAAGGGAAAGCGCAAGAGAGAGCGCACA
 CGCACACACCCGCCGCGCGCACTCGCGCAC
 GGACCCGCACGGGGACAGCTCGGAAGTCAT
 CAGTTCATGGGCGAATGCTGCTGCTGGCGA
 GATGTCTGCTGCTAGTCCTCGTCTCCTCGCTG

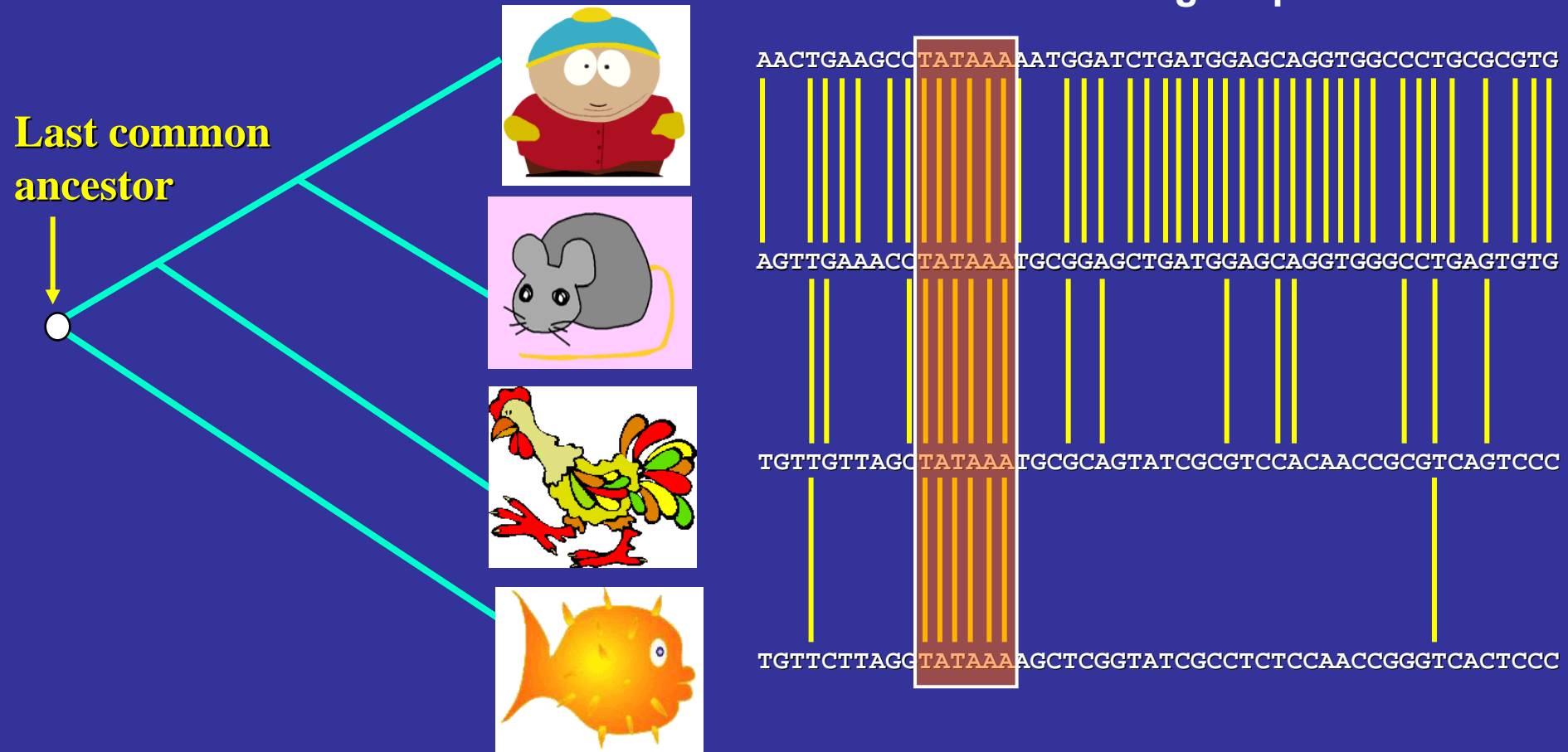


Comparative Genomics

NO CODE

		Second base						
		U	C	A	G			
First base	U	UUU } Phenyl- alanine F UUC } UUA } Leucine L UUG }	UCU } Serine S UCC } UCA } UCG }	UAU } Tyrosine Y UAC } UAA } Stop codon UAG } Stop codon	UGU } Cysteine C UGC } UGA } Stop codon UGG } Tryptophan W	U	C	
	C	CUU } Leucine L CUC } CUA } CUG }	CCU } Proline P CCC } CCA } CCG }	CAU } Histidine H CAC } CAA } Glutamine Q CAG }	CGU } Arginine R CGC } CGA } CGG }	C	A	
	A	AUU } Isoleucine I AUC } AUA } AUG M } Methionine start codon	ACU } Threonine T ACC } ACA } ACG }	AAU } Asparagine N AAC } AAA } Lysine K AAG }	AGU } Serine S AGC } AGA } Arginine R AGG }	A	G	
	G	GUU } Valine V GUC } GUA } GUG }	GCU } Alanine A GCC } GCA } GCG }	GAU } Aspartic acid D GAC } GAA } Glutamic acid E GAG }	GGU } Glycine G GGC } GGA } GGG }	G		
					U	C	A	G

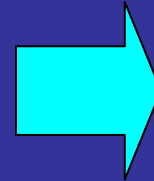
Evolutionary conservation can find regulatory elements



**Evolutionarily Conserved Sequences
are Functionally Important.**

98% of the genome

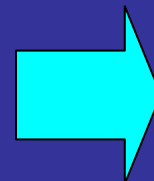
AGGAAGGGAAAGCGCAAGAGAGAGCGCACA
 CGCACACCCGCCGCGCGCACTCGCGCAC
 GGACCCGCACGGGGACAGCTCGGAAGTCAT
 CAGTTCATGGGCGAATGCTGCTGCTGGCGA
 GATGTCTGCTGCTAGTCCTCGTCTCCTCGCTG



Comparative Genomics

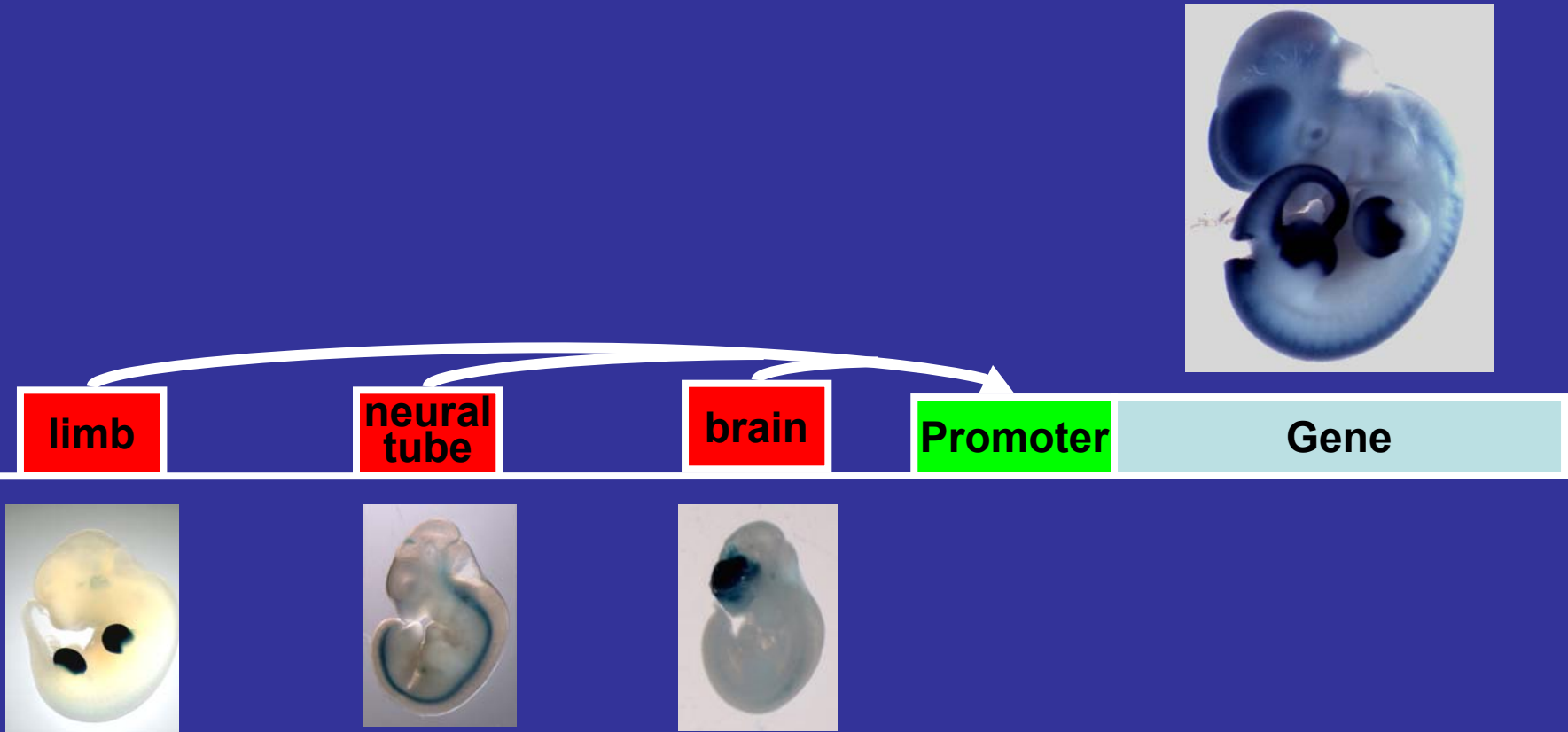
NO CODE

		Second base					
		U	C	A	G		
First base	U	UUU } Phenyl-alanine F UUC } UUA } Leucine L UUG }	UCU } Serine S UCC } UCA } UCG }	UAU } Tyrosine Y UAC } UAA } Stop codon UAG } Stop codon	UGU } Cysteine C UGC } UGA } Stop codon UGG } Tryptophan W	U	C
	C	CUU } Leucine L CUC } CUA } CUG }	CCU } Proline P CCC } CCA } CCG }	CAU } Histidine H CAC } CAA } Glutamine Q CAG }	CGU } Arginine R CGC } CGA } CGG }	C	A
	A	AUU } Isoleucine I AUC } AUA } AUG } Methionine start codon M	ACU } Threonine T ACC } ACA } ACG }	AAU } Asparagine N AAC } AAA } Lysine K AAG }	AGU } Serine S AGC } AGA } Arginine R AGG }	A	G
	G	GUU } Valine V GUC } GUA } GUG }	GCU } Alanine A GCC } GCA } GCG }	GAU } Aspartic acid D GAC } GAA } Glutamic acid E GAG }	G	U	C



High-Throughput Characterization

Distant cis-regulatory element: Enhancer



Extremely conserved non-coding sequences tested

1. Human-Fish (N=3,124)

70% & ≥ 100 bp Human-Fugu



AACTGAAGCCTATAAAAATGGATCTGA



AGTTGAAACCTATAAATGCGGAGCTGA



TGTTGTTAGCTATAAATGCGCAGTATC



TGTTCTTAGCTATAAAGCTCGGTATC

2. Ultraconserved (N=256)

100% & ≥ 200 bp Human-Mouse-Rat



AACTGAAGCCTATAAAAATGGATCTGA



AACTGAAGCCTATAAAAATGGATCTGA



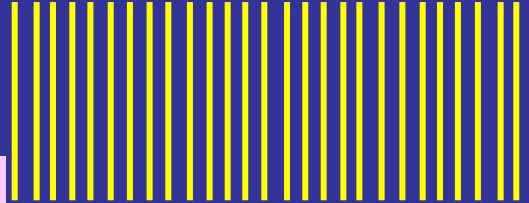
AACTGAAGCCTATAAAAATGGATCTGA

~50% enhancer positive success rate

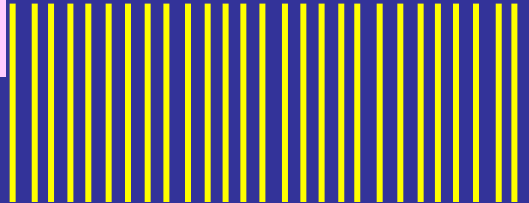
Mouse Enhancer Assay



AACTGAAGCCTATAAAAAATGGATCTGA

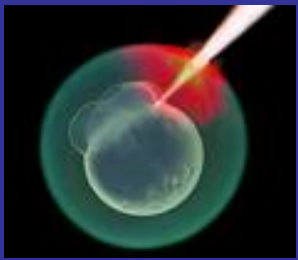


AACTGAAGCCTATAAAAAATGGATCTGA



AACTGAAGCCTATAAAAAATGGATCTGA

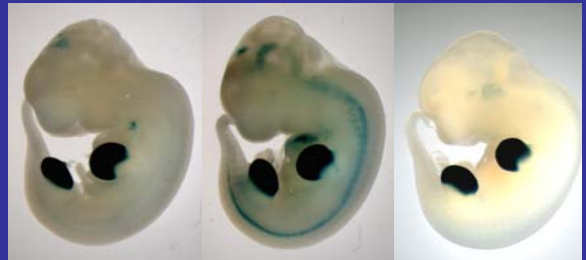
Zebrafish



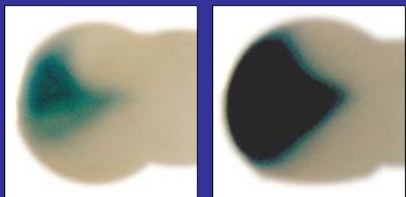
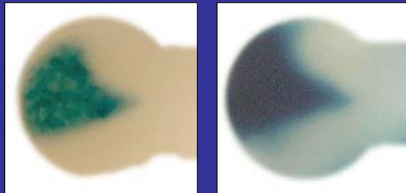
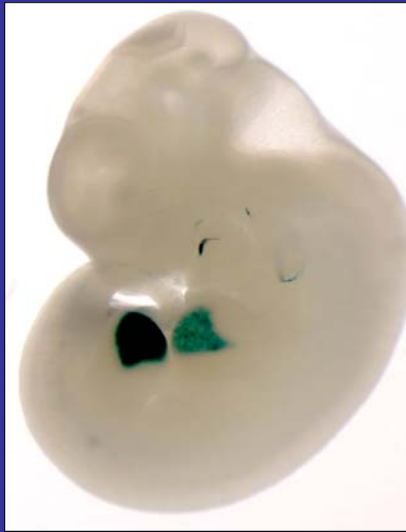
Injection into mouse fertilized eggs



Harvest E11.5 embryos, LacZ stain

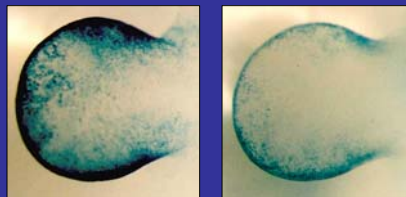
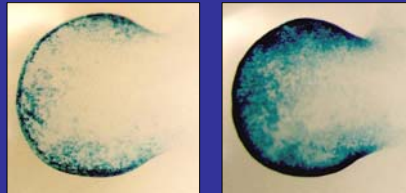
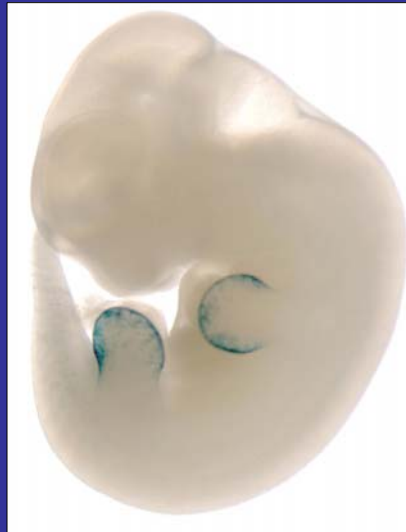


Tissue-Specific Expression is Highly Reproducible and Specific



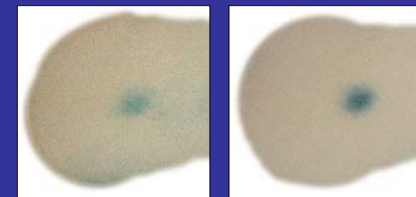
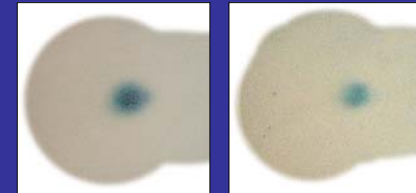
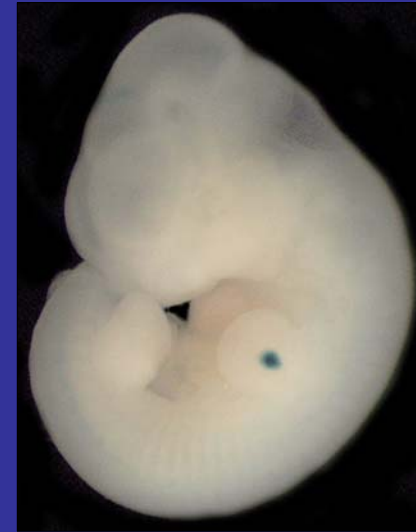
4/4

(triangular distal wedge)



9/10

(apical ectodermal ridge)

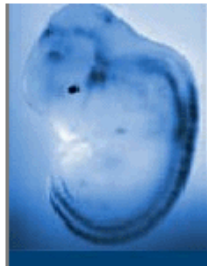


10/14

(forelimb-specific, single medial dorsal spot)

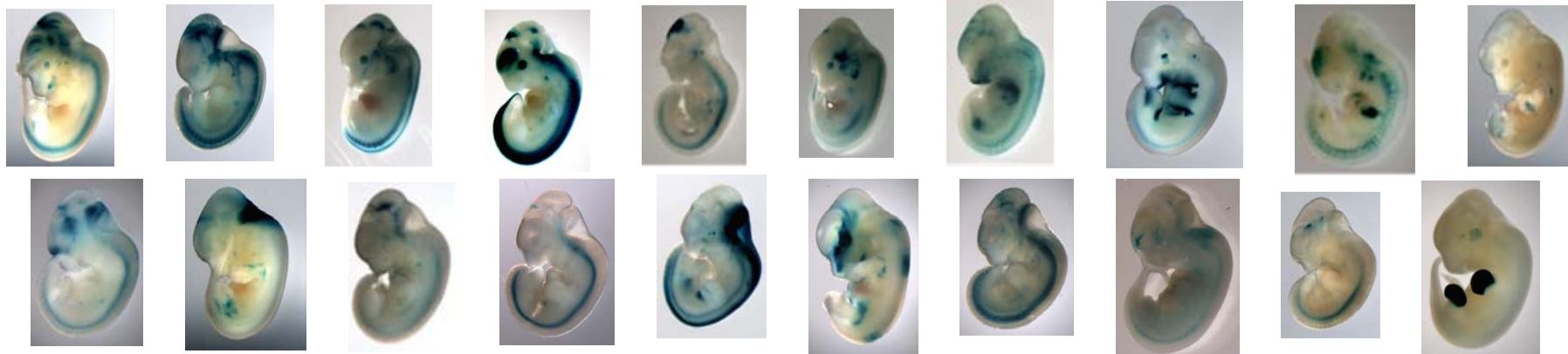
The Jewish Enhancer



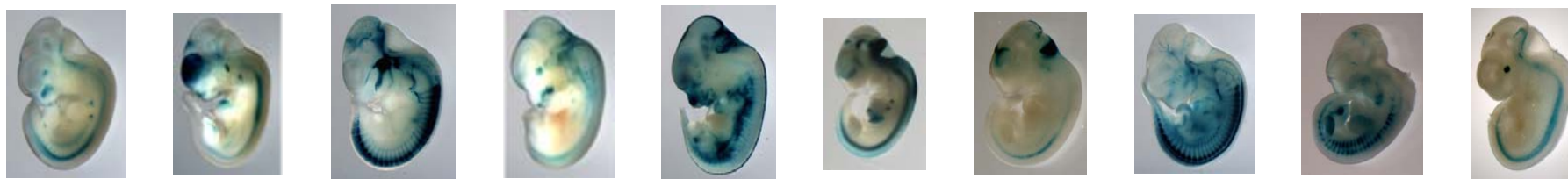


VISTA Enhancer Browser

whole genome enhancer browser



<http://enhancer.ibl.gov>



346 human enhancers out of 849 tested



VISTA Enhancer Browser

whole genome enhancer browser

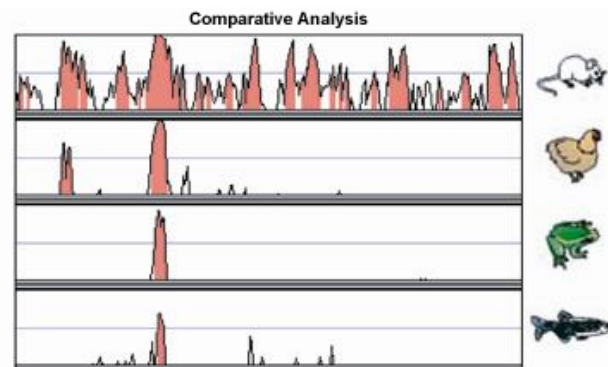


The **VISTA Enhancer Browser** is a central resource for experimentally validated human noncoding fragments with gene enhancer activity as assessed in transgenic mice. Most of these noncoding elements were selected for testing based on their extreme conservation with other vertebrates. The results of this enhancer screen are provided through this publicly available website.

This program is located at [Lawrence Berkeley National Laboratory](#). See [Handbook](#) for additional details on this work or visit the [Experimental Results](#) to view data.

The browser also features relevant results by external contributors and a large collection of additional [genome-wide conserved noncoding elements](#) which are candidate enhancer sequences. We invite external groups to [submit computational predictions](#) of developmental enhancers.

As of **5/26/2006** the database contains information on **255** in vivo tested elements - **106** elements with enhancer activity.



Keyword Search

Examples: gene, accession number, locus link, genomic position

[Expression Pattern Search](#)

[Advanced Search](#)



Mouse Egg Microinjection



E11.5 Reporter Staining



VISTA Enhancer Browser

whole genome enhancer browser



Home Browser Handbook and Methods Text Search for Expression Pattern Advanced Search Experimental Data Computational Dataset Participants and Contact

Id	Location	Bracketing Genes	Expression	Conservation			
62	chr16:51088042-51089472	SALL1-CHD9					
63	chr16:51061335-51062427	SALL1-CHD9					
64	chr16:50993756-50994877	SALL1-CHD9					
65	chr16:50515789-50517189	SALL1-CHD9					
66	chr16:50375933	SALL1-CHD9					
68	chr16:5029540	SALL1-CHD9					
69	chr16:5029544	SALL1-CHD9					
70	chr16:5029515	SALL1-CHD9					
71	chr16:50228682-50229540	SALL1-CHD9					
72	chr16:50215311-50216217	SALL1-CHD9					
73	chr16:50134226-50135602	SALL1-CHD9					
74	chr16:50117697-50118426	SALL1-CHD9					
75	chr16:49946796-49948220	SALL1-CHD9					
76	chr16:49788622-49789863	SALL1-CHD9					

chromosomal location

neighboring genes

**result of assay
blue=positive
white=negative**

conserved species

ultra

Text Search for Expression Patterns

Advanced Search

Experimental Data

Participants and Contact

Home

ULTRA



VISTA Enhancer Browser

whole genome enhancer browser

Home Browser Handbook and Methods Text Search for Expression Pattern Advanced Search Experimental Data Computational Dataset Participants and Contact

Expression pattern

- branchial arch
- forebrain
- inner ear
- midbrain (mesencephalon)
- somite
- cranial nerve
- genital tubercle
- limb
- neural tube
- trigeminal V (ganglion, cranial)
- dorsal root ganglion
- heart
- melanocytes
- nose
- No pattern
- eye
- hindbrain (rhombencephalon)
- mesenchyme derived from neural crest
- notochord

Home

Text Search for Expression

Id	Location	Bracketing Genes	Expression	Conservation
1	chr16:84987588-84988227	IRF8-FOXF1		
27	chr16:53636738-53637930	IRX5-IRX6		
72	chr16:50215311-50216217	SALL1-CHD9		
108	chr16:70962034-70965092	PMFBP1-ATBF1		
126	chr13:70241610-70242869	KLHL1-DACH1		
142	chr13:71979448-71981755	DACH1-FLJ22624		
174	chr1:87533642-87535103	LMO4-PKN2		ULTRA
200	chr1:50877216-50878807	FAF1(intragenic)		ULTRA
213	chr3:148532234-148533798	PLSCR1-ZIC4		ULTRA
243	chr2:174811051-174812971	PTD004(intragenic)		ULTRA
258	chr3:181919129-181920660	TTC14-FXR1		ULTRA
302	chr8:37368752-37370340	FKSG2-FLJ14299		ULTRA
305	chr9:13929547-13930363	MPDZ-NFIB		ULTRA

Advanced Search

Experimental Data

Participants and Contact

Flanking genes: [SALL1-CHD9](#)

The patterns are strongly reproducible.

Expression Pattern

limbs

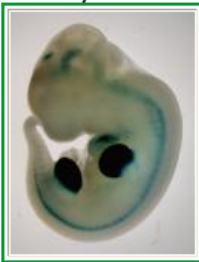
Embryo 1



Embryo 2



Embryo 4



Fasta sequence

```

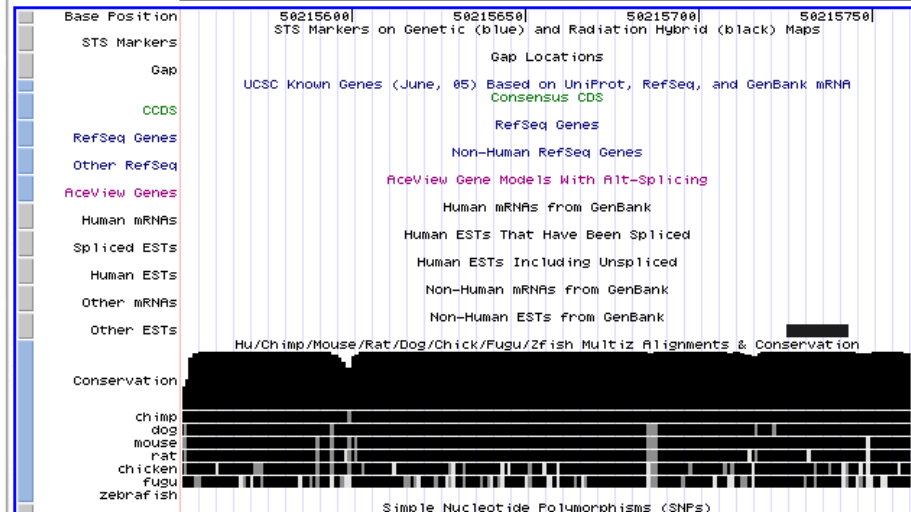
ttgctgaattctgaaagaggcaggagtgcacaaaaagagatcgagctgttttttctg
aaggcaatttattggactaccacagggatgtttatcttaactagaggtcatgagaca
caaaggccactaatgagtgaatttggctggctgggggagtcagggtcttctcactctt
ccaagtcttcttagactgaaaaaaaaaaaaagggaaacaaacaaaaacaaaaaatg
gtccctgtaaatttatgactgcagtgcaaaagtcttctatctaaattgcagtaaat
ttctaatgaagtctggaagaagccatttgaatggctcggggcagagattaaagatag
cctgttctagttagctggtcattagagagatctaacatgcatgggaaatggttctgt
tctaagtgtttaaaaccaagtttaggaaaccccggttctgccggtcatttggagac
aacagtgccaaccggtcagacttttcagatggccaagtgtcaccttctcagctaca
tttagttcgctgcagcagcctctaccagggagagtggttctcaggggtcctg
gcaggagtagatagctccacagagcaggagatctggacggggcaccactgaaagctgt
ccttgcctctgttgaatgcattgggattgagtgattcatgggatgagtcacaaacct
caggagaatctccagtcacagcctagatcatctatctagaaacagactgaggggact
ggtgagaaaatggtggataaactccctggttgggagggacagggtagctgatgggag
gaaattctttggctacctgatgtgctgtgcattaaaatgggattt
    
```

Primer

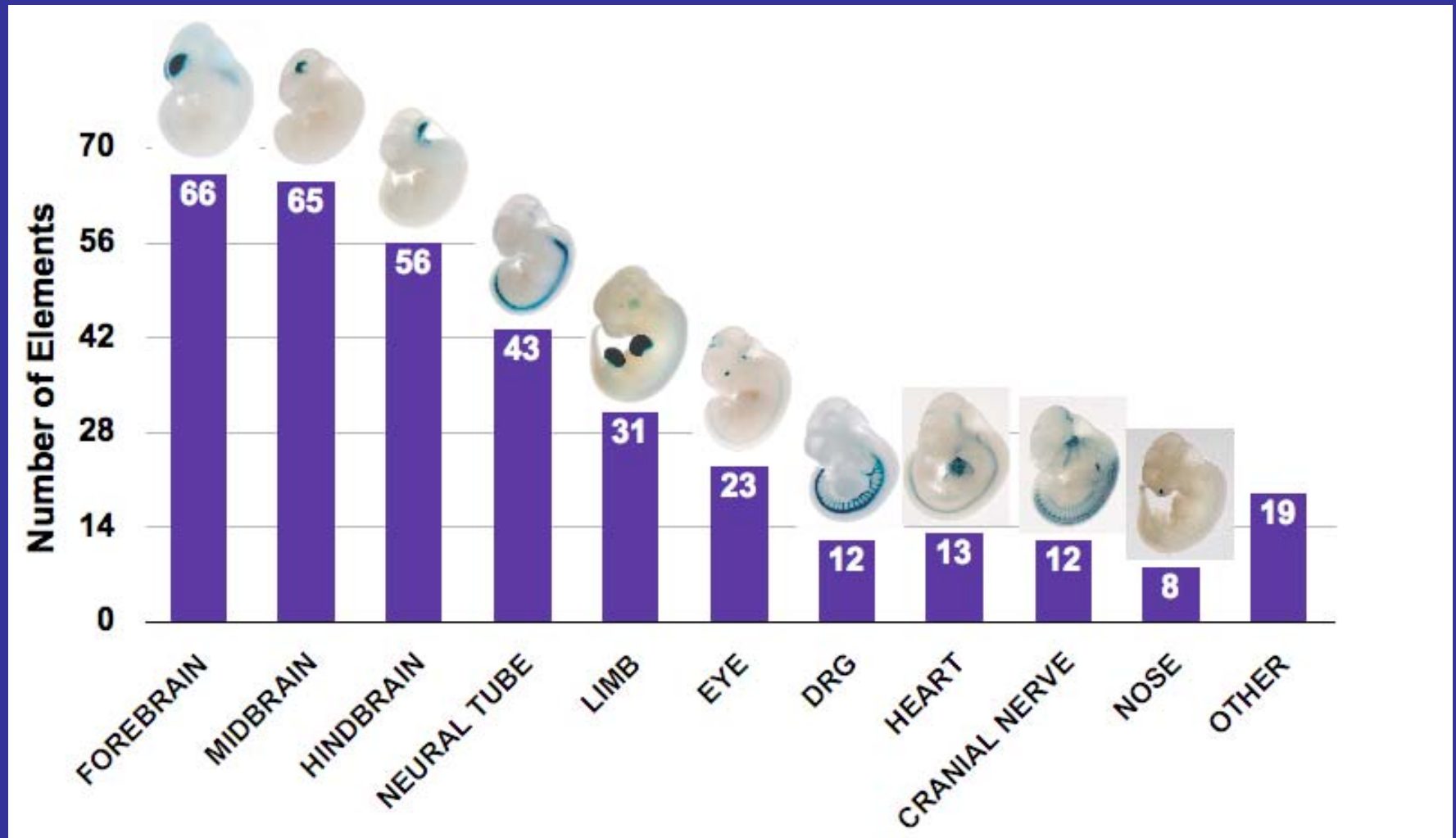
(+)ctgctgaattctgaaagaggca

(-)ctgggacttcttgggttgg

Track 95
 Mouse May 2004
 chr8 (+)
 88769439-88769646



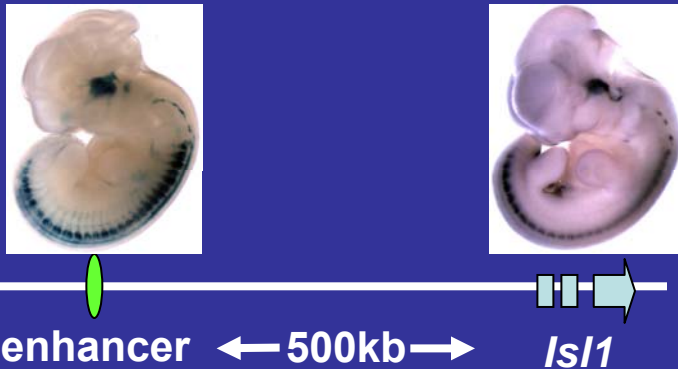
Bins of Enhancer Activities



Applications

1. Gene/Genome annotation
2. Reagent to drive tissue specific expression
3. EvoDevo
4. Exploring the regulatory code
5. Role of regulatory sequences in human disease

1.



2.



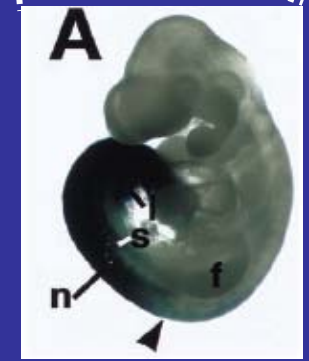
3. EvoDevo

Hoxc9

Early
Enhancer

Hoxc8

Consensus	82	97	112
Human	GCG-ACGTGGCTAGACGCTCTGGGCTTAAATGCTTTATGGTTA		
Mouse	"G"GC"		
Armadillo	"AG"GC"		
Bat	"G"GC"		
Cow	"G"GC"		
Hippopotamus	"G"GC"		
Ocelot	"G"GC"		
Wallaby	T"GG"GC"		
Elephant	"G"GC"		
Camel	"G"GC"		
Llama	"G"GC"		
Giraffe	"AG"GCC"		
<u>Toothed Whales:</u>			
Northern Bottlenose			
Common Porpoise			
Narwhal			
Beluga			
Whitebeak Dolphin			
Indus River Dolphin			
Antillean Beaked			
North Sea Beaked			
La Plata Dolphin			
Baird's Beaked Dol.			
<u>Sperm Whales:</u>			
Great Sperm Whale	"G"	A	G
Pygmy Sperm Whale	"G"	A	G
<u>Baleen Whales:</u>			
Fin Whale			
Sei Whale			
Minke Whale			
Bowhead Whale			
Humpback Whale			

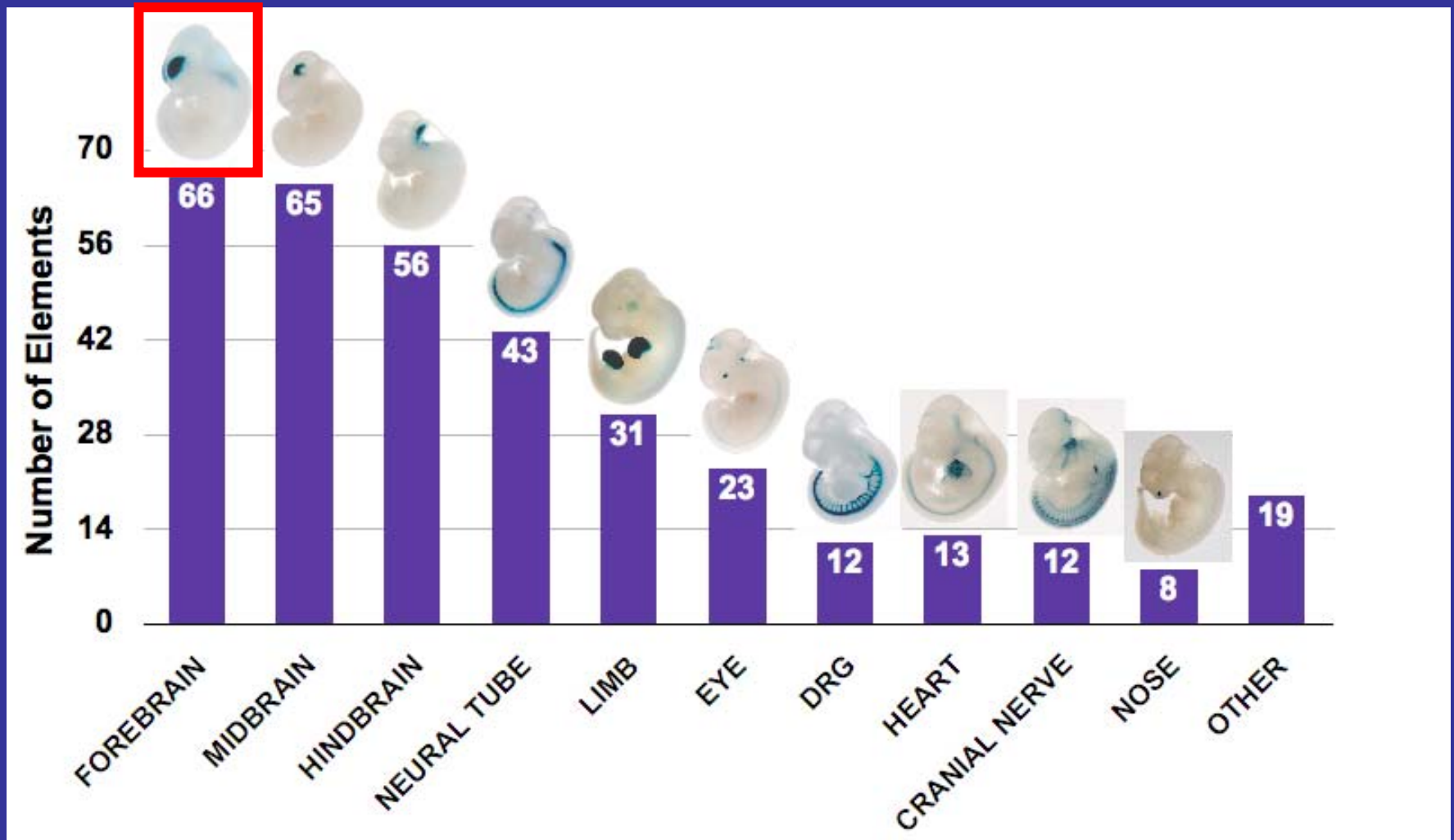


Mouse



Baleen whale

4. Exploring the regulatory code

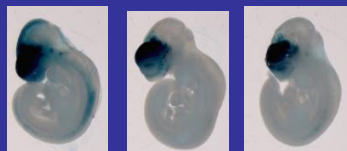


Forebrain Predictions

• CNS injected= 5%

Training Set

Element 1



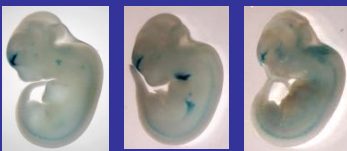
Element 2



Element 3



Element 4



3-fold

Prediction=17% (4/23)

Predictions

Element 1



Element 2



Element 3



Element 4



Over-representation

6 different motifs

ATTAA CANNGGC
GATTA TTNNAAA
TANNTGA TTNNTTT

Applications

1. Gene/Genome annotation
2. Reagent to drive tissue specific expression
3. EvoDevo
4. Exploring the regulatory code
5. Role of regulatory sequences in human disease

Human phenotypes studied in the lab

- Limb malformations



- Obesity

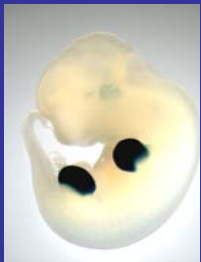


- Pharmacogenomics



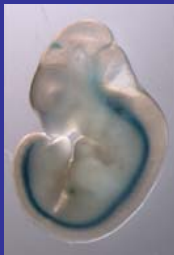
LIMB MALFORMATIONS

- 1 in 500 births.
- Appear in numerous syndromes



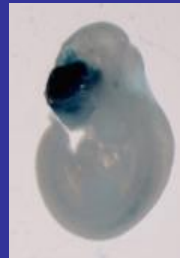
MUTATIONS

- Limb



MUTATIONS

- Neural tube



MUTATIONS

- Brain

MUTATIONS

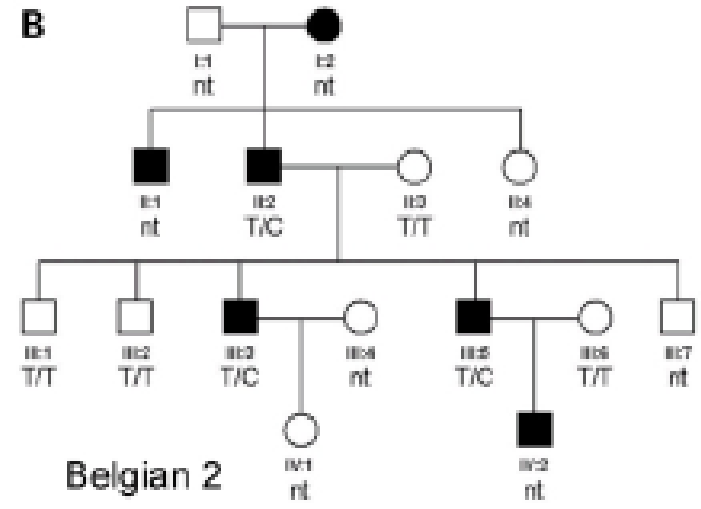
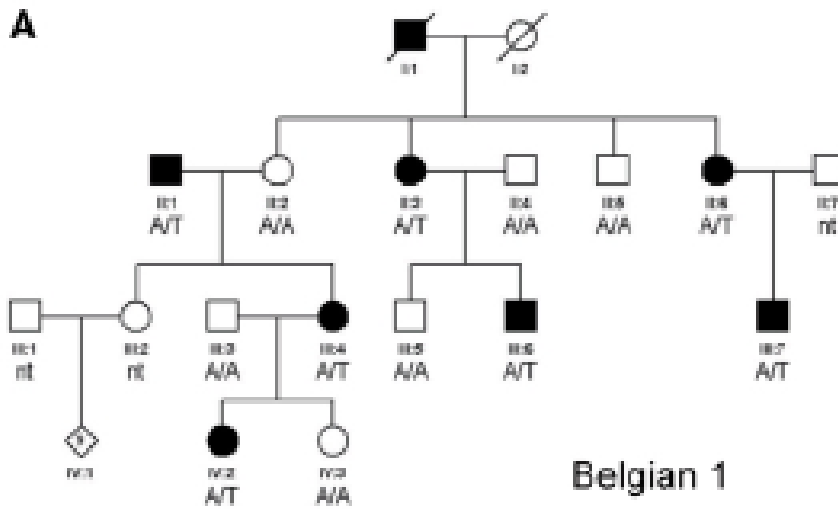
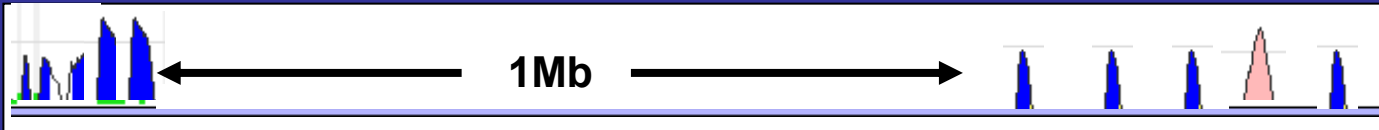
- Brain
- Neural tube
- Limb

- Well characterized and good human-mouse correlations

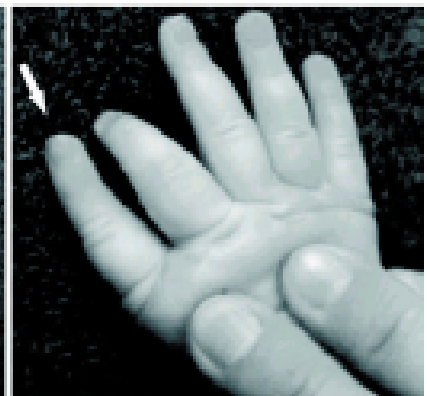
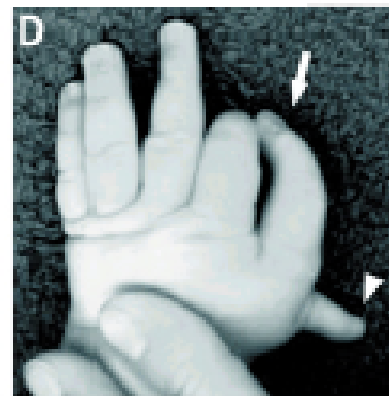
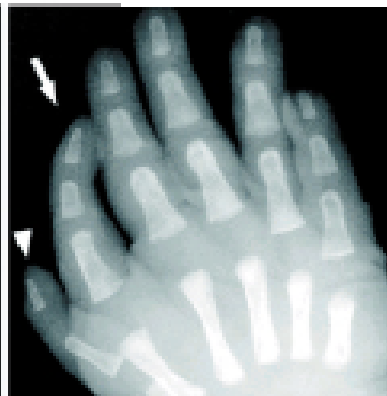
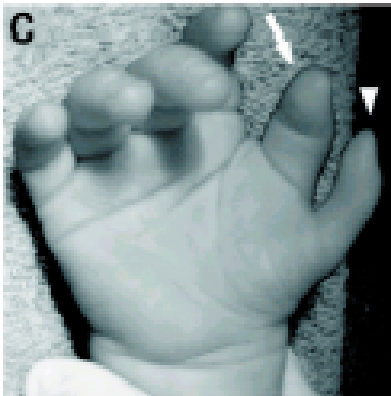
Limb enhancer with substantial evidence for involvement in human disease



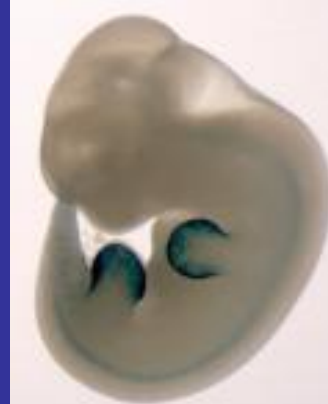
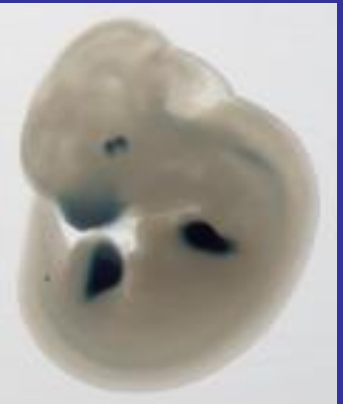
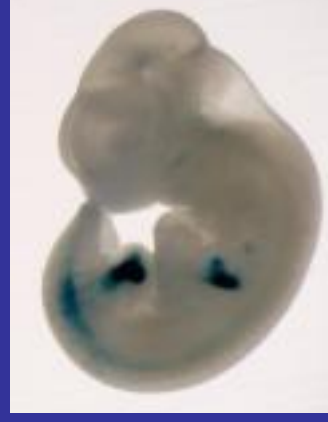
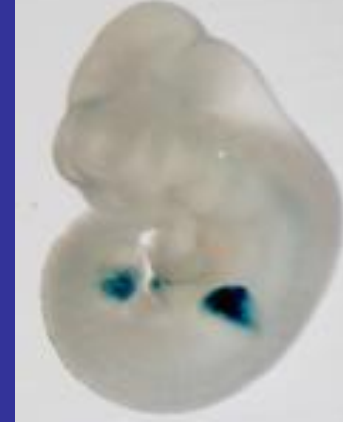
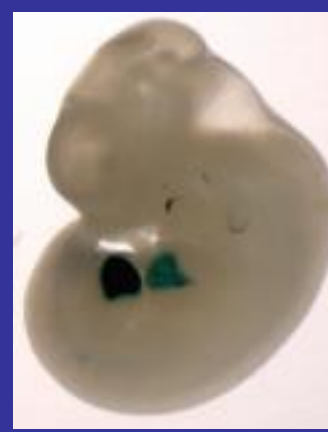
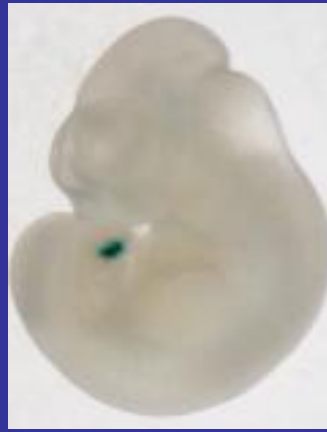
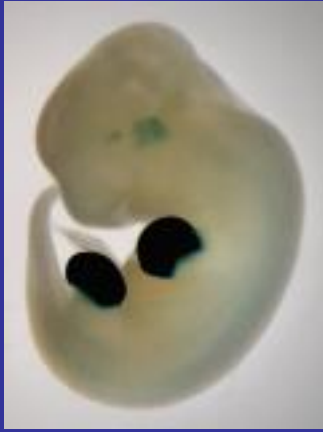
Human-Fugu



Lettice et al. *HMG* 2003 12: 1725-35



>40 limb enhancers discovered



Limb malformations outline

- Mutation analysis in humans



Find Additional Limb Enhancers

- Computational approaches

- Comparative genomics around limb associated genes.
- Limb specific signatures.

- Limb enhancer assays

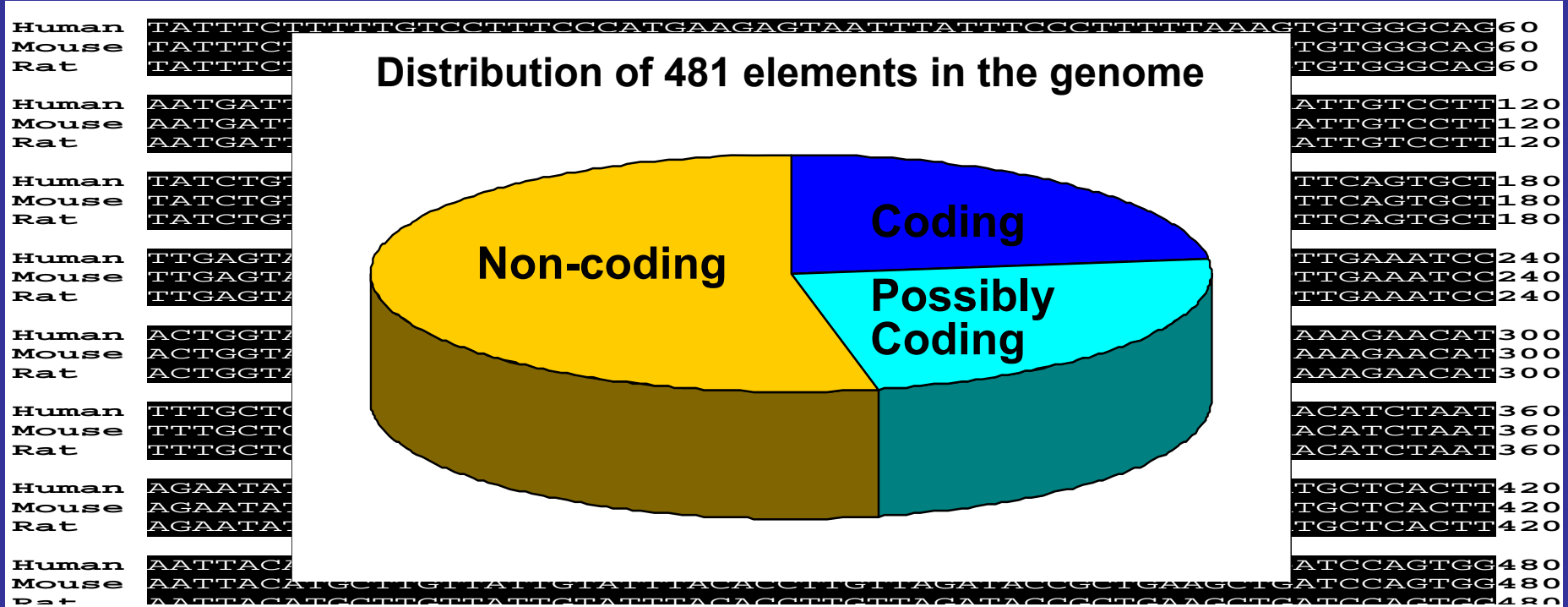


Ultraconserved elements

Mt. Everest



Ultraconserved elements (100% identity and ≥ 200 bp human-mouse-rat)



- **Significantly less SNPs within these elements**

(Drake J.A. et al. *Nature Genetics* 2006, 38: 223-7

Katzman et al. *Science* 2007, 317: 915)

- **Significantly depleted in segmental duplications and copy number variants**

(Derti A. *Nature Genetics* 2006, 38:1216-20)

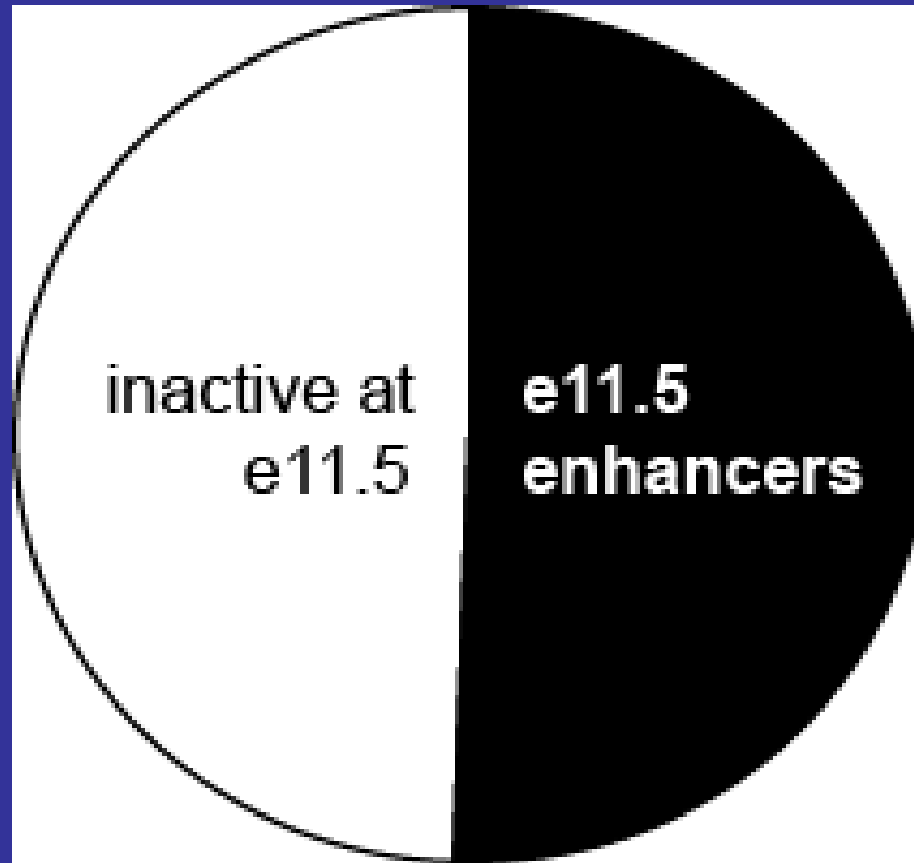


VISTA Enhancer Browser

whole genome enhancer browser



[Home](#) [Browser Handbook and Methods](#) [Experimental Data](#) [Computational Dataset](#) [Participants and Contact](#)



~50% function as enhancers in mice at E11.5

Ultraconserved elements (100% identity and ≥ 200 bp human-mouse-rat)

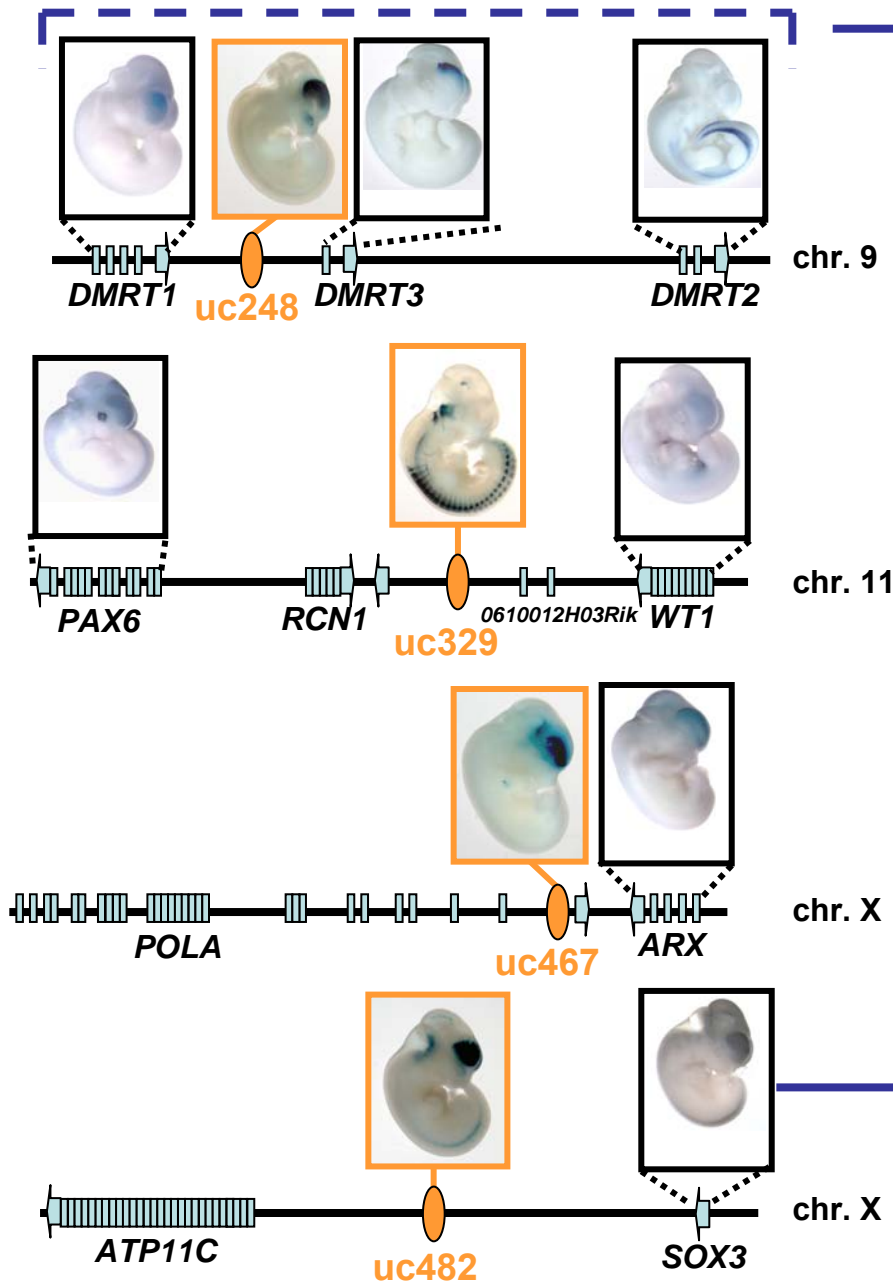
Human	TATTTCTTTTTTGTCTTTCCCATGAAGAGTAATTTATTTCCCTTTTTTAAAGTGTGGGCAG	60
Mouse	TATTTCTTTTTTGTCTTTCCCATGAAGAGTAATTTATTTCCCTTTTTTAAAGTGTGGGCAG	60
Rat	TATTTCTTTTTTGTCTTTCCCATGAAGAGTAATTTATTTCCCTTTTTTAAAGTGTGGGCAG	60
Human	AATGATTACTAGCGCACTGTAATGTAATTGTGTTGCATTGATAAAAATAAAAATTGTCCTT	120
Mouse	AATGATTACTAGCGCACTGTAATGTAATTGTGTTGCATTGATAAAAATAAAAATTGTCCTT	120
Rat	AATGATTACTAGCGCACTGTAATGTAATTGTGTTGCATTGATAAAAATAAAAATTGTCCTT	120
Human	TATCTGTGTCCTGATAATGTTCAATTTACAGGCTGGCTTCTCTGCCACCTCTTCAGTGCT	180
Mouse	TATCTGTGTCCTGATAATGTTCAATTTACAGGCTGGCTTCTCTGCCACCTCTTCAGTGCT	180
Rat	TATCTGTGTCCTGATAATGTTCAATTTACAGGCTGGCTTCTCTGCCACCTCTTCAGTGCT	180
Human	TTGAGTATTCCAGTTCTCCTCCCTTCCCTGCTCTGAGAGCGCACAGAACTGTTTGAAATCC	240
Mouse	TTGAGTATTCCAGTTCTCCTCCCTTCCCTGCTCTGAGAGCGCACAGAACTGTTTGAAATCC	240
Rat	TTGAGTATTCCAGTTCTCCTCCCTTCCCTGCTCTGAGAGCGCACAGAACTGTTTGAAATCC	240
Hu		00
Mc		00
Ra		00
Hu		60
Mc		60
Ra		60
Hu		20
Mc		20
Ra		20
Hu		80
Mouse	AATTACATGCTTGTATTATTGATTACACCTTGTTAGATACCGCTGAAGCTGATCCAGTGG	480
Rat	AATTACATGCTTGTATTATTGATTACACCTTGTTAGATACCGCTGAAGCTGATCCAGTGG	480
Human	CTGGCCGGGAATTGGAAGCGTCTGTTCATGGGGCAGTTGGAGCGCGTTTTTGTAGGAAATGC	540
Mouse	CTGGCCGGGAATTGGAAGCGTCTGTTCATGGGGCAGTTGGAGCGCGTTTTTGTAGGAAATGC	540
Rat	CTGGCCGGGAATTGGAAGCGTCTGTTCATGGGGCAGTTGGAGCGCGTTTTTGTAGGAAATGC	540
Human	TATTTATTTTAAATGCTCCACCTGCTGGGAGCCGAGGTTAGTCAGCAGCACTGAGATGAA	600
Mouse	TATTTATTTTAAATGCTCCACCTGCTGGGAGCCGAGGTTAGTCAGCAGCACTGAGATGAA	600
Rat	TATTTATTTTAAATGCTCCACCTGCTGGGAGCCGAGGTTAGTCAGCAGCACTGAGATGAA	600
Human	TTGGGAAACGGGGTGTAAAAAGAAATAATGTGCTTCTGACAGGCTCCCGTGGCTTTTAAAGT	660
Mouse	TTGGGAAACGGGGTGTAAAAAGAAATAATGTGCTTCTGACAGGCTCCCGTGGCTTTTAAAGT	660
Rat	TTGGGAAACGGGGTGTAAAAAGAAATAATGTGCTTCTGACAGGCTCCCGTGGCTTTTAAAGT	660
Human	TGCTCTCATTTCAGCCACTTTCACAAAAAATTATTTTATTCCATCTCTCAGTGATGATGACA	720
Mouse	TGCTCTCATTTCAGCCACTTTCACAAAAAATTATTTTATTCCATCTCTCAGTGATGATGACA	720
Rat	TGCTCTCATTTCAGCCACTTTCACAAAAAATTATTTTATTCCATCTCTCAGTGATGATGACA	720
Human	TGATTGCTTTT	731
Mouse	TGATTGCTTTT	731
Rat	TGATTGCTTTT	731

Extreme Sequence Constraint

=

Extreme Functional Constraint?

Increasing the likelihood of observing a phenotype



Heterozygous deletion of region in humans: XY females

Selected elements were:

- Positive in our enhancer assay
- Near genes that when mutated lead to either a lethal or sexual development phenotype
- Near genes that exhibit dosage sensitive abnormalities

Mouse KO: Sexual and pituitary development abnormalities

Variable size

uc248 (222bp)

Human AGCCCCCTGTCTCTGAAAGGCTTTCTGCTGGGATCCCATTATGTGCTTGAATAAACCCCTTT60
 Mouse AGCCCCCTGTCTCTGAAAGGCTTTCTGCTGGGATCCCATTATGTGCTTGAATAAACCCCTTT60
 Rat AGCCCCCTGTCTCTGAAAGGCTTTCTGCTGGGATCCCATTATGTGCTTGAATAAACCCCTTT60

Human CTGCAAAACAGAAATGGGACTCGGGGTTGTCAGGTGTTGGGTTACAATAGACTTTGAGGCCA120
 Mouse CTGCAAAACAGAAATGGGACTCGGGGTTGTCAGGTGTTGGGTTACAATAGACTTTGAGGCCA120
 Rat CTGCAAAACAGAAATGGGACTCGGGGTTGTCAGGTGTTGGGTTACAATAGACTTTGAGGCCA120

Human GGGGACATTTTACCAACATAAAATACAAACCTGTCCCTATAGGGAGATGTTGTCAAATAC180
 Mouse GGGGACATTTTACCAACATAAAATACAAACCTGTCCCTATAGGGAGATGTTGTCAAATAC180
 Rat GGGGACATTTTACCAACATAAAATACAAACCTGTCCCTATAGGGAGATGTTGTCAAATAC180

Human GGGATATGGAAAACATTCCCATCAAATATGAGAAAAGGATTA222
 Mouse GGGATATGGAAAACATTCCCATCAAATATGAGAAAAGGATTA222
 Rat GGGATATGGAAAACATTCCCATCAAATATGAGAAAAGGATTA222

uc467 (731bp)

Human TATTTCCTTTTGTCCCTTCCCATGAAGAGTAATTTATTCCCTTTTAAAGTGTGGGCAG60
 Mouse TATTTCCTTTTGTCCCTTCCCATGAAGAGTAATTTATTCCCTTTTAAAGTGTGGGCAG60
 Rat TATTTCCTTTTGTCCCTTCCCATGAAGAGTAATTTATTCCCTTTTAAAGTGTGGGCAG60

Human AATGATTACTAGGCCACTGTAATGTAAATGTGTTCGATTGATAAAATAAAAATTTGCCTT120
 Mouse AATGATTACTAGGCCACTGTAATGTAAATGTGTTCGATTGATAAAATAAAAATTTGCCTT120
 Rat AATGATTACTAGGCCACTGTAATGTAAATGTGTTCGATTGATAAAATAAAAATTTGCCTT120

Human TATCTGTCTCCTGATAATGTTCAATTTACAGGCTGGCTTCTCTGCCACCTCTTCAGTGCT180
 Mouse TATCTGTCTCCTGATAATGTTCAATTTACAGGCTGGCTTCTCTGCCACCTCTTCAGTGCT180
 Rat TATCTGTCTCCTGATAATGTTCAATTTACAGGCTGGCTTCTCTGCCACCTCTTCAGTGCT180

Human TTGAGTATTCAGAGTTTCTCCTCCCTTCTGCTCTGAGAGCGCAGACAAGTGTGTAATTC240
 Mouse TTGAGTATTCAGAGTTTCTCCTCCCTTCTGCTCTGAGAGCGCAGACAAGTGTGTAATTC240
 Rat TTGAGTATTCAGAGTTTCTCCTCCCTTCTGCTCTGAGAGCGCAGACAAGTGTGTAATTC240

Human ACTGGTACAATTTGCAAAATCAATTTTCAATTCCTGCAATTTATGCTCGCACAAAGAACAT300
 Mouse ACTGGTACAATTTGCAAAATCAATTTTCAATTCCTGCAATTTATGCTCGCACAAAGAACAT300
 Rat ACTGGTACAATTTGCAAAATCAATTTTCAATTCCTGCAATTTATGCTCGCACAAAGAACAT300

Human TTTGCTGTCTGAATGATGATTAATAAATACAGCTATTCCAGCTGCCGTGATAACATCTAA360
 Mouse TTTGCTGTCTGAATGATGATTAATAAATACAGCTATTCCAGCTGCCGTGATAACATCTAA360
 Rat TTTGCTGTCTGAATGATGATTAATAAATACAGCTATTCCAGCTGCCGTGATAACATCTAA360

Human AGAATATTCATAAGCCCAAAATGGAATGAATTTATCCATTAACTTCATCATGCTCACT420
 Mouse AGAATATTCATAAGCCCAAAATGGAATGAATTTATCCATTAACTTCATCATGCTCACT420
 Rat AGAATATTCATAAGCCCAAAATGGAATGAATTTATCCATTAACTTCATCATGCTCACT420

Human AATTACATGCTTGTATTGTATTACACCTTGTAGATACCGCTGAAGCTGATCCAGTGG480
 Mouse AATTACATGCTTGTATTGTATTACACCTTGTAGATACCGCTGAAGCTGATCCAGTGG480
 Rat AATTACATGCTTGTATTGTATTACACCTTGTAGATACCGCTGAAGCTGATCCAGTGG480

Human CTGGCCGGGAATGGAAGCGTCTGTCTATGGGGCAGTTGGAGCGCTTTTGTAGGAAATGG540
 Mouse CTGGCCGGGAATGGAAGCGTCTGTCTATGGGGCAGTTGGAGCGCTTTTGTAGGAAATGG540
 Rat CTGGCCGGGAATGGAAGCGTCTGTCTATGGGGCAGTTGGAGCGCTTTTGTAGGAAATGG540

Human TATTTATTTTAAATGCTCCACCTGCTGGAGCCGAGGTTAGTCAGCAGCTGAGATGAA600
 Mouse TATTTATTTTAAATGCTCCACCTGCTGGAGCCGAGGTTAGTCAGCAGCTGAGATGAA600
 Rat TATTTATTTTAAATGCTCCACCTGCTGGAGCCGAGGTTAGTCAGCAGCTGAGATGAA600

Human TTGGGAAACGGGGTGTAAAAAGAAATAATGTGCTTCTGACAGGCTCCGTGGCTTTTAAAGT560
 Mouse TTGGGAAACGGGGTGTAAAAAGAAATAATGTGCTTCTGACAGGCTCCGTGGCTTTTAAAGT560
 Rat TTGGGAAACGGGGTGTAAAAAGAAATAATGTGCTTCTGACAGGCTCCGTGGCTTTTAAAGT560

Human TGCTCTCATTGAGCCACTTCCAAAAAATTTATTTTATCCATCTCTCAGTGATGATGACA720
 Mouse TGCTCTCATTGAGCCACTTCCAAAAAATTTATTTTATCCATCTCTCAGTGATGATGACA720
 Rat TGCTCTCATTGAGCCACTTCCAAAAAATTTATTTTATCCATCTCTCAGTGATGATGACA720

Human TGATTGCTTTT731
 Mouse TGATTGCTTTT731
 Rat TGATTGCTTTT731

uc329 (307bp)

Human ATGCACTTGTGCGCCCTTCTTATTACATCATGAATTTTGATAAGCCCATTAGCCAGCCAGG60
 Mouse ATGCACTTGTGCGCCCTTCTTATTACATCATGAATTTTGATAAGCCCATTAGCCAGCCAGG60
 Rat ATGCACTTGTGCGCCCTTCTTATTACATCATGAATTTTGATAAGCCCATTAGCCAGCCAGG60

Human CAGCCAGGAGGAGGATGTCTTTTCTCACTTGATGACTACACAATGAGGAGTAATATATTTT120
 Mouse CAGCCAGGAGGAGGATGTCTTTTCTCACTTGATGACTACACAATGAGGAGTAATATATTTT120
 Rat CAGCCAGGAGGAGGATGTCTTTTCTCACTTGATGACTACACAATGAGGAGTAATATATTTT120

Human AATTTAAACACCATAAAACCTATTGAGTGCCTAATTAACCTTTAACCGCTCTTTAGGGGAATG180
 Mouse AATTTAAACACCATAAAACCTATTGAGTGCCTAATTAACCTTTAACCGCTCTTTAGGGGAATG180
 Rat AATTTAAACACCATAAAACCTATTGAGTGCCTAATTAACCTTTAACCGCTCTTTAGGGGAATG180

Human AAGCCAAATTTGAGGGGGTCAAAGGAGTGTTCATATCAAACAGTTTTACATAGTCAA240
 Mouse AAGCCAAATTTGAGGGGGTCAAAGGAGTGTTCATATCAAACAGTTTTACATAGTCAA240
 Rat AAGCCAAATTTGAGGGGGTCAAAGGAGTGTTCATATCAAACAGTTTTACATAGTCAA240

Human TCCATTCACTCGGCCACAGCTACTGATTAACACACAAAAATTAACATGTTAGTAAAAAT300
 Mouse TCCATTCACTCGGCCACAGCTACTGATTAACACACAAAAATTAACATGTTAGTAAAAAT300
 Rat TCCATTCACTCGGCCACAGCTACTGATTAACACACAAAAATTAACATGTTAGTAAAAAT300

Human CATATTC307
 Mouse CATATTC307
 Rat CATATTC307

uc482(295bp)

Human CGCATTAATGATTTTCAGTAAAGGAATGTGAGCATTAAATAGAAATTTGTTAGCAGATAC60
 Mouse CGCATTAATGATTTTCAGTAAAGGAATGTGAGCATTAAATAGAAATTTGTTAGCAGATAC60
 Rat CGCATTAATGATTTTCAGTAAAGGAATGTGAGCATTAAATAGAAATTTGTTAGCAGATAC60

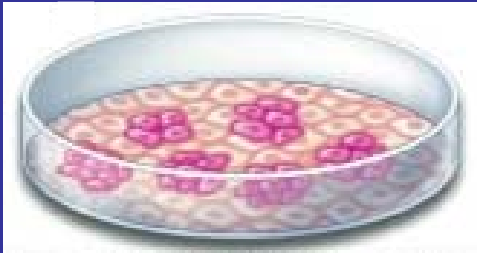
Human CAAGGGGAAAAGTTTCTTCAACAACACTATAGGAGGTGTCATTGTACAAAGACTTTTTT120
 Mouse CAAGGGGAAAAGTTTCTTCAACAACACTATAGGAGGTGTCATTGTACAAAGACTTTTTT120
 Rat CAAGGGGAAAAGTTTCTTCAACAACACTATAGGAGGTGTCATTGTACAAAGACTTTTTT120

Human GTTTTCTTTCCCCATTACATATCACTGCAATATGTAAGGTAGTGATTCTTTTCATTAA180
 Mouse GTTTTCTTTCCCCATTACATATCACTGCAATATGTAAGGTAGTGATTCTTTTCATTAA180
 Rat GTTTTCTTTCCCCATTACATATCACTGCAATATGTAAGGTAGTGATTCTTTTCATTAA180

Human ATTAAGTGAAAAGCTACTAAATCTGTTTTATTCTCCAAAAAGAAATAGGTCCTTAACAC240
 Mouse ATTAAGTGAAAAGCTACTAAATCTGTTTTATTCTCCAAAAAGAAATAGGTCCTTAACAC240
 Rat ATTAAGTGAAAAGCTACTAAATCTGTTTTATTCTCCAAAAAGAAATAGGTCCTTAACAC240

Human TTATCTTTTAACTTCTTTCAATTTCTTAAGGTTGGTGTGTTACTTTTGTCTTTCAGTC295
 Mouse TTATCTTTTAACTTCTTTCAATTTCTTAAGGTTGGTGTGTTACTTTTGTCTTTCAGTC295
 Rat TTATCTTTTAACTTCTTTCAATTTCTTAAGGTTGGTGTGTTACTTTTGTCTTTCAGTC295

Generation of non-coding ultraconserved mouse knockouts



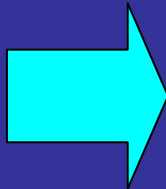
ES cells



Blastocyst injections




Chimeras



Heterozygous KOs

uc467 (731bp)

Human	TATTTCTTTTTTGTCCCTTCCCATGAAGAGTAATTTAT	TCCCTTTTTTAAAGTGTGGCCAC	60
Mouse	TATTTCTTTTTTGTCCCTTCCCATGAAGAGTAATTTAT	TCCCTTTTTTAAAGTGTGGCCAC	60
Rat	TATTTCTTTTTTGTCCCTTCCCATGAAGAGTAATTTAT	TCCCTTTTTTAAAGTGTGGCCAC	60
Human	AATGATTAACCTAGCCCACTGTAAATGTAATTTGCTGCA	TTGATAAAAAATAAAAAATGTCCCT	120
Mouse	AATGATTAACCTAGCCCACTGTAAATGTAATTTGCTGCA	TTGATAAAAAATAAAAAATGTCCCT	120
Rat	AATGATTAACCTAGCCCACTGTAAATGTAATTTGCTGCA	TTGATAAAAAATAAAAAATGTCCCT	120
Human	TATCTGTGTCTCTGAAAATGTTCAATTTACAGGGCTG	CTTCTCTGCCACCTCTTCAGTGC	180
Mouse	TATCTGTGTCTCTGAAAATGTTCAATTTACAGGGCTG	CTTCTCTGCCACCTCTTCAGTGC	180
Rat	TATCTGTGTCTCTGAAAATGTTCAATTTACAGGGCTG	CTTCTCTGCCACCTCTTCAGTGC	180
Human	TTGAGTATTCACAGTTCCTCTCCCTTCCCTGCTCT	GCTTCTCTGCCACCTCTTCAGTGC	180
Mouse	TTGAGTATTCACAGTTCCTCTCCCTTCCCTGCTCT	GCTTCTCTGCCACCTCTTCAGTGC	180
Rat	TTGAGTATTCACAGTTCCTCTCCCTTCCCTGCTCT	GCTTCTCTGCCACCTCTTCAGTGC	180
Human	ACTGCTACAATTTGTCAAAATCAATTTATTCATTC	TAGAGCCGACAGAACTGTTTGAATCC	240
Mouse	ACTGCTACAATTTGTCAAAATCAATTTATTCATTC	TAGAGCCGACAGAACTGTTTGAATCC	240
Rat	ACTGCTACAATTTGTCAAAATCAATTTATTCATTC	TAGAGCCGACAGAACTGTTTGAATCC	240
Human	TTTGCTGGTCTGAAATGATGATTAATAAACA	CTGCCAATTATGCTCCGACAAAGACAT	300
Mouse	TTTGCTGGTCTGAAATGATGATTAATAAACA	CTGCCAATTATGCTCCGACAAAGACAT	300
Rat	TTTGCTGGTCTGAAATGATGATTAATAAACA	CTGCCAATTATGCTCCGACAAAGACAT	300
Human	AGAATATTCATAAAGCCCAAATGGAATGAA	ATCTCCATTAACCTTCATCATGCTCAC	360
Mouse	AGAATATTCATAAAGCCCAAATGGAATGAA	ATCTCCATTAACCTTCATCATGCTCAC	360
Rat	AGAATATTCATAAAGCCCAAATGGAATGAA	ATCTCCATTAACCTTCATCATGCTCAC	360
Human	AATTACATGCTTTGTTATTGTTATTT	TTAGATACCCGCTGAAAGCTGATCCAGT	420
Mouse	AATTACATGCTTTGTTATTGTTATTT	TTAGATACCCGCTGAAAGCTGATCCAGT	420
Rat	AATTACATGCTTTGTTATTGTTATTT	TTAGATACCCGCTGAAAGCTGATCCAGT	420
Human	CTGGCCGGGAATTGGAAGCGTCT	GTTCGGAGCCGCTTTTGTAGGAAA	480
Mouse	CTGGCCGGGAATTGGAAGCGTCT	GTTCGGAGCCGCTTTTGTAGGAAA	480
Rat	CTGGCCGGGAATTGGAAGCGTCT	GTTCGGAGCCGCTTTTGTAGGAAA	480
Human	TATTTATTTTAAATGCTCCAC	AGGGTTAGTCAGCAGCAGCTGAGAT	540
Mouse	TATTTATTTTAAATGCTCCAC	AGGGTTAGTCAGCAGCAGCTGAGAT	540
Rat	TATTTATTTTAAATGCTCCAC	AGGGTTAGTCAGCAGCAGCTGAGAT	540
Human	TTGGGAAACCGGGTGTAAAA	CGAGGTTAGTCAGCAGCAGCTGAGAT	600
Mouse	TTGGGAAACCGGGTGTAAAA	CGAGGTTAGTCAGCAGCAGCTGAGAT	600
Rat	TTGGGAAACCGGGTGTAAAA	CGAGGTTAGTCAGCAGCAGCTGAGAT	600
Human	TGCTCTCATTACAGCCACTTC	CGCTTCTGACAGGCTCCCGTGGCTTTT	660
Mouse	TGCTCTCATTACAGCCACTTC	CGCTTCTGACAGGCTCCCGTGGCTTTT	660
Rat	TGCTCTCATTACAGCCACTTC	CGCTTCTGACAGGCTCCCGTGGCTTTT	660
Human	TGATTTGCTTTT731	TATTTTATTTCCATCTCTCAGTGTATGAT	720
Mouse	TGATTTGCTTTT731	TATTTTATTTCCATCTCTCAGTGTATGAT	720
Rat	TGATTTGCTTTT731	TATTTTATTTCCATCTCTCAGTGTATGAT	720

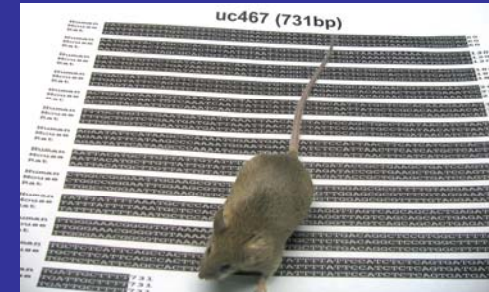


NO APPARENT PHENOTYPE

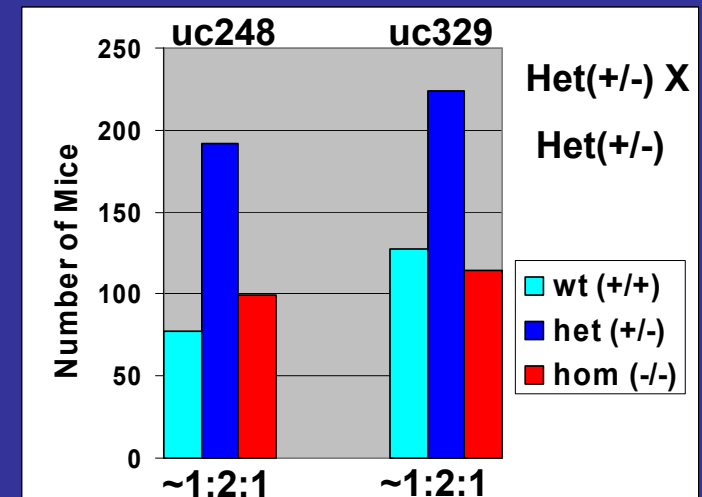


General Phenotyping Reveals No Apparent Abnormalities

- Longevity: up to 6-12 months



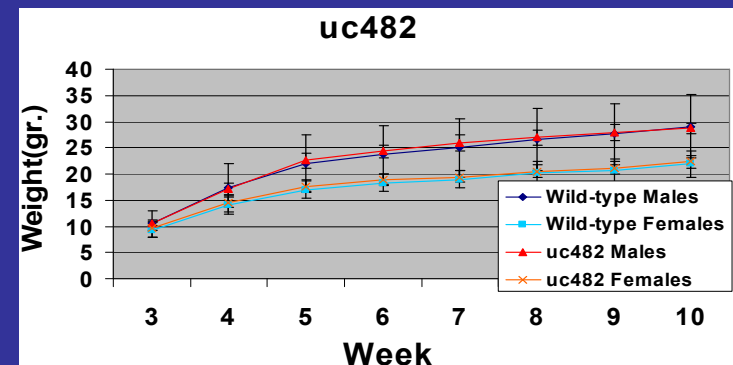
- Fertility: No reduction in Homozygous KO numbers



- Body weight measurements

- Pathology

- Clinical Chemistry



Gene Specific Phenotyping Reveals No Apparent Abnormalities



• Dmrt1 KO: -Male sexual development (*Genes & Dev* 2000, 14:2587-95)

• Dmrt3 KO: -Death from starvation due to dental malocclusions
-Male sexual development



(N. Ahituv, unpublished)

uc248 KO PHENOTYPING
-Male sexual development
-Dental

SUMMARY



- Using comparative genomics we can find distant regulatory elements.
- We can do functional assays to test their function.
- Distant regulatory elements are candidates for human disease and other phenotypes.
- Future work will link nucleotide changes within these elements to human disease and other phenotypes.



Ahituv Lab

Ramon Birnbaum

Mee J. Kim

Qiang Li

Aziza Nelson

Loan Nguyen

Julia VanderMeer

UCSF

Kathleen M. Giacomini

Ophir Klein

Deanna L. Kroetz

Christian Vaisse

Berkeley Lab & JGI

Len A. Pennacchio

Edward M. Rubin

Stanford

Gill Bejerano