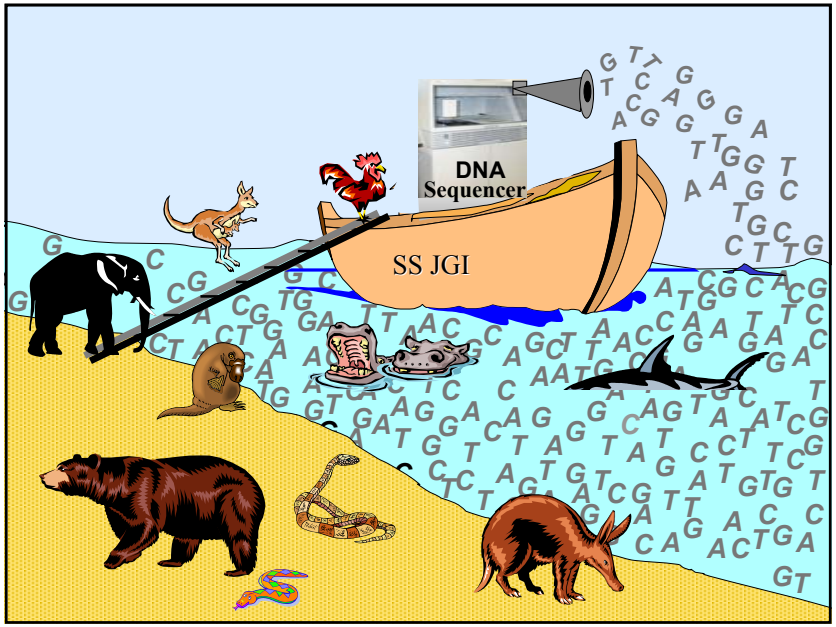


Computational Tools for Biologists to Draw Insights from Comparative Sequence Data

***

Tales of How Comparative Sequence Data Led to Biological Insights

VISUALIZATION TOOLS FOR ALIGNMENTS

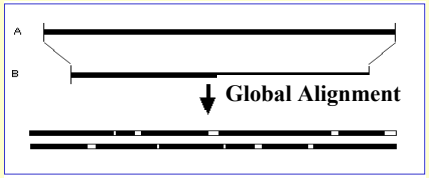
VISTA

VISTA is an integrated system for global sequence alignment and visualization,

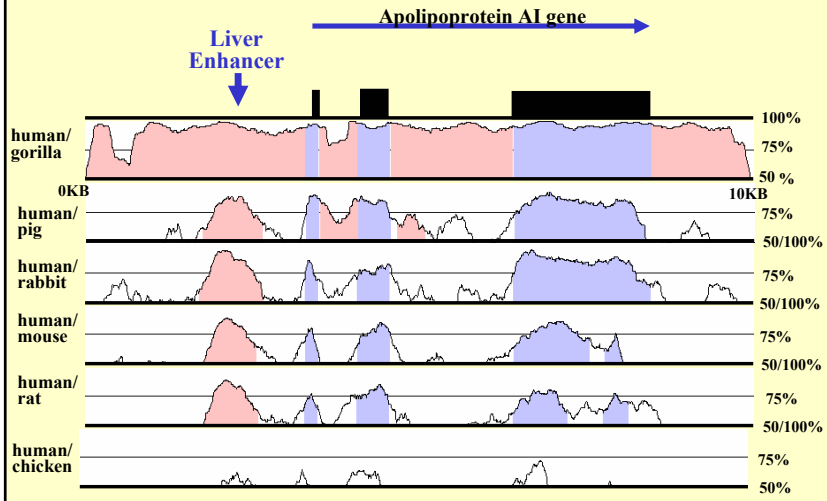
- VISTA – Standard VISTA for 2 way sequence comparisons
- rVISTA – regulatory VISTA
- aVISTA – audio VISTA

Godzilla: Whole Vertebrate Genome Comparisons

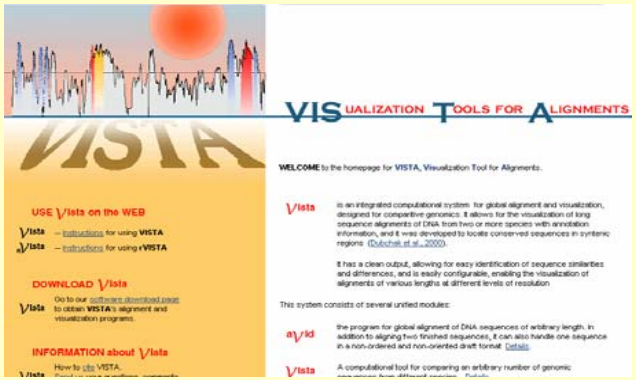
AVID - The Alignment Engine Behind VISTA



VISTA 2 Way Sequence Comparison



http://www-gsd.lbl.gov/vista



Since June 2000 > 5500 request for sequence analysis and > 320 copies of software distributed

A pipeline for real-time comparative analysis of human and vertebrate genomes

GODZILLA

http://pipeline.lbl.gov

- Comparative analysis of whole vertebrate Genomes
- Annotation and Visualization of elements conserved in two genomes

A pipeline for real-time comparative analysis of human and vertebrate genomes

GODZILLA

Input data (pipeline)

Human Genome – GoldenPath Assembly: genome.ucsc.edu

RefSeq gene annotation -- UCSC
Ensemble gene annotation -- Ensemble

2nd Vertebrate Genome:

Mouse:

Fugu

Rat

A PIPELINE FOR REAL-TIME COMPARATIVE ANALYSIS OF THE HUMAN AND MOUSE GENOMES

GODZILLA

Lior Pachter
Ivan Ovcharenko
Inna Dubchak

http://pipeline.lbl.gov/

ATGTCCA
TATATAT
GTCTTCT

Current Mouse Contigs and Supercontigs

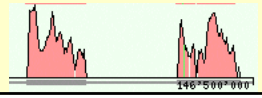
ATGTCCA
NNNNNNN
GTCTTCT

Mask repetitive elements *RepeatMasker*

Localize homologous sequence in the Human Genome (*BLAT w/postprocessing)*

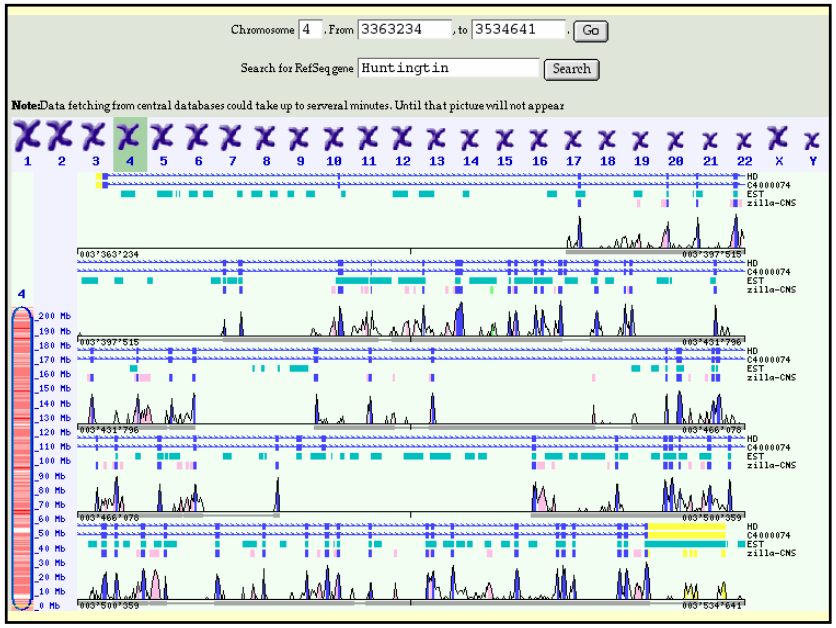
```
                10
human     GCTACTCGGG
          || ||| |||
mouse     GC-ACTTGGG
              13240
```

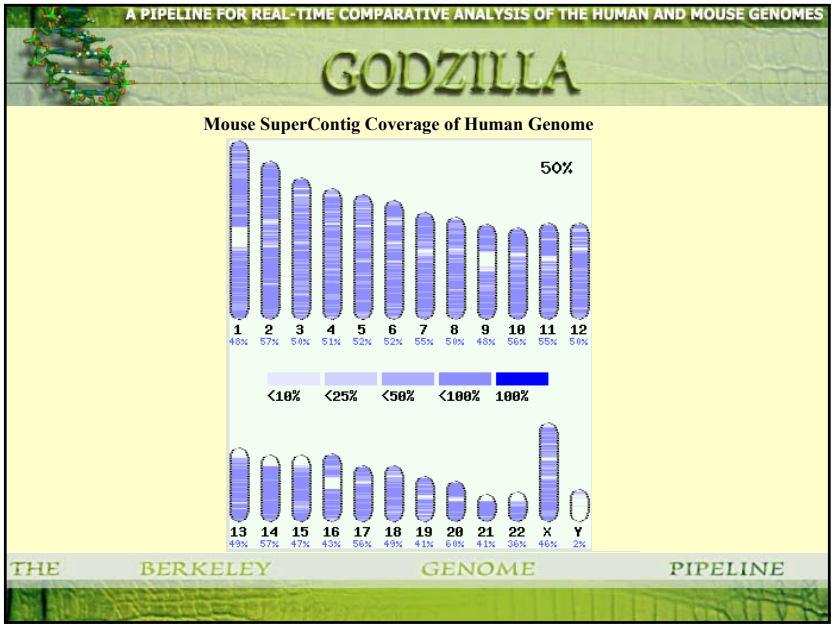
Align Human and Mouse sequences (*AVID)*

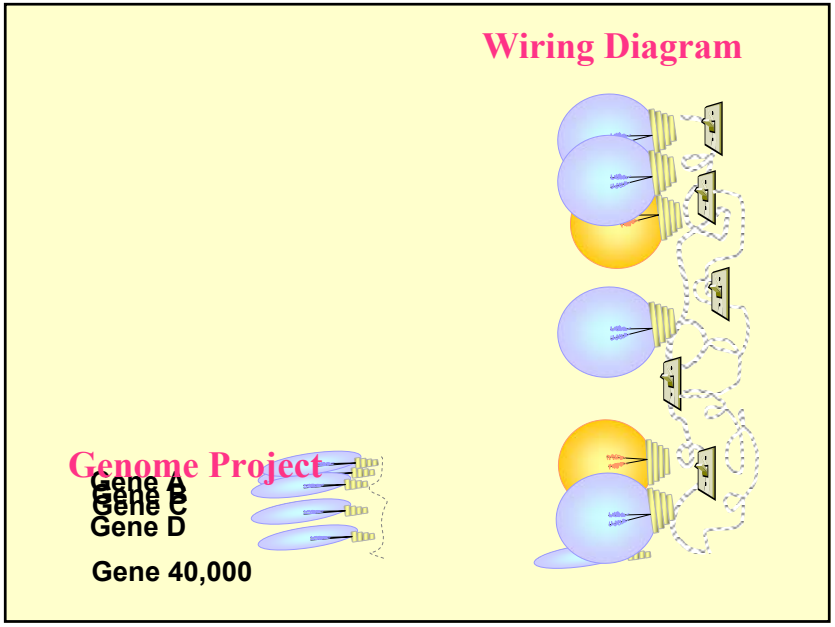


Build VISTA plots

Running the pipeline…

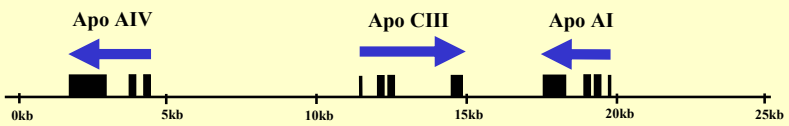

Chromosome 4 , From 3363234 , to 3534641 . Go
Search for RefSeq gene Huntingtin Search
Note: Data fetching from central databases could take up to serveral minutes. Until that picture will not appear
HD
C4000074
EST
zilla-CNS


A PIPELINE FOR REAL-TIME COMPARATIVE ANALYSIS OF THE HUMAN AND MOUSE GENOMES
GODZILLA
Mouse SuperContig Coverage of Human Genome
50%
<10% <25% <50% <100% 100%
THE BERKELEY GENOME PIPELINE

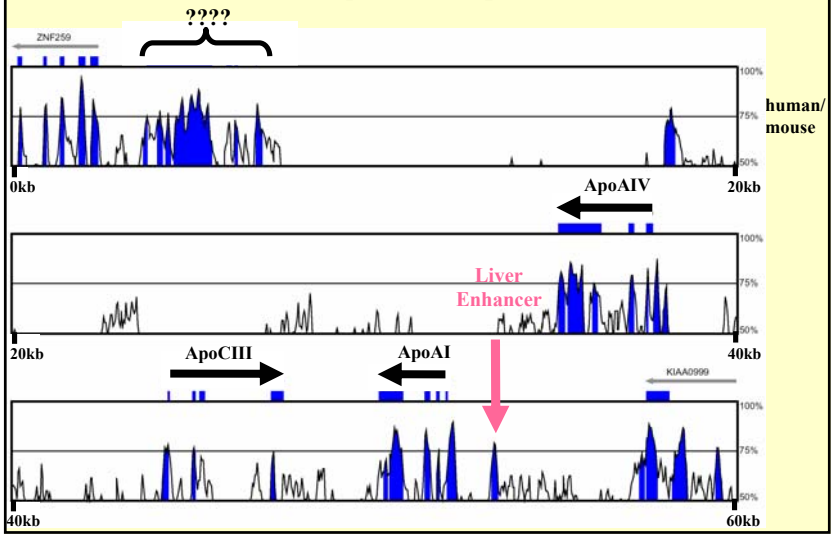


Wayward Discovery of a New Apolipoprotein Gene

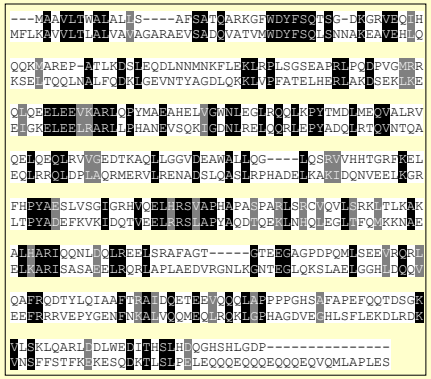
Human Chromosome 11q23 Apolipoprotein Gene Cluster



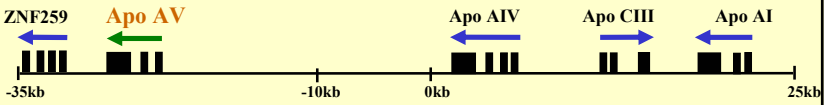
Human/Mouse Apolipoprotein Gene Cluster Sequence Comparison



Predicted Protein Sequence Has Homology to ApoAIV

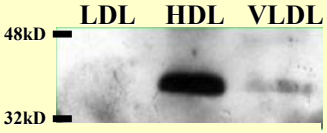


Identity: 26%
Similarity: 45%



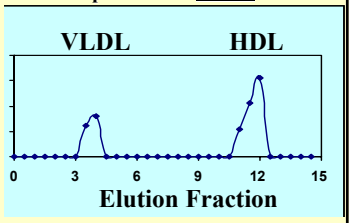
ApoAV associates with HDL and VLDL

Western Blot of Human Plasma

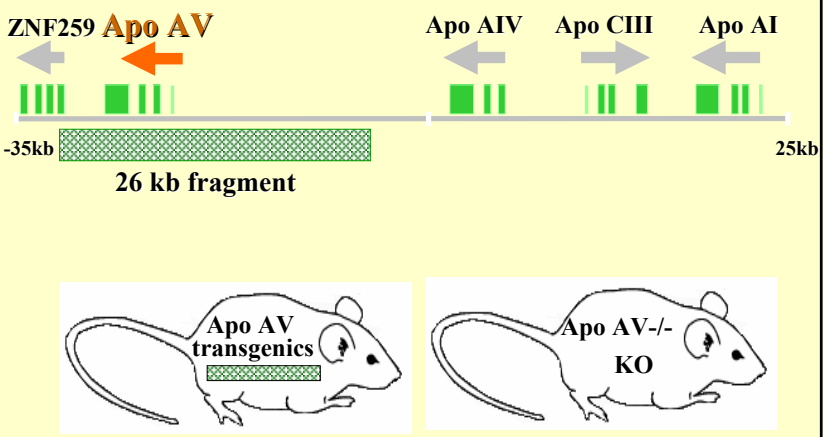


Anti-ApoAV Antibodies

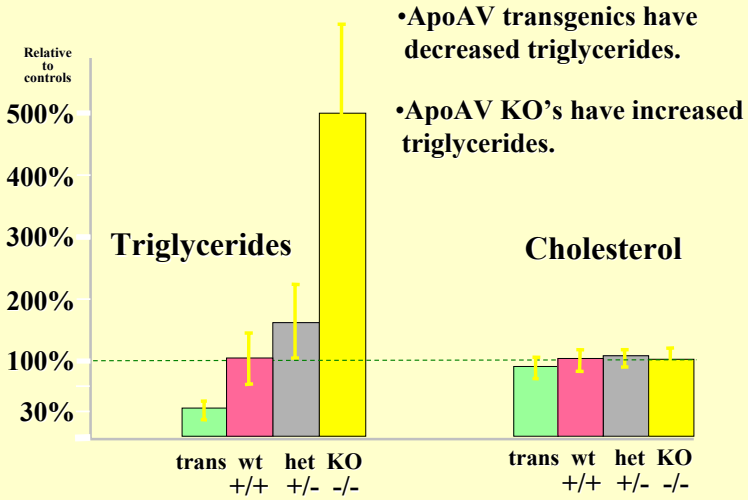
FPLC Separation of Mouse Plasma



Human ApoAV Transgenic and Knock Out Mice



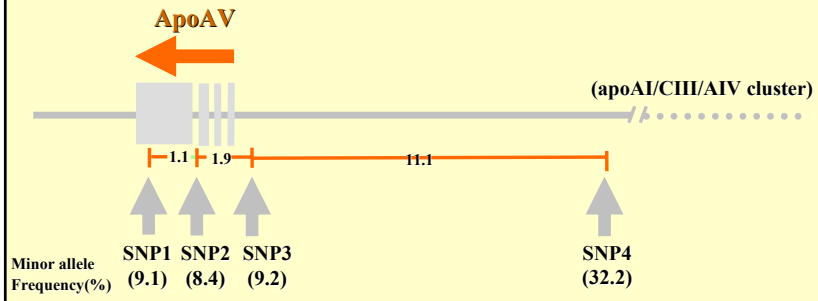
Lipid Levels in Mice with High and Low ApoAV



- ApoAV transgenics have decreased triglycerides.
- ApoAV KO's have increased triglycerides.

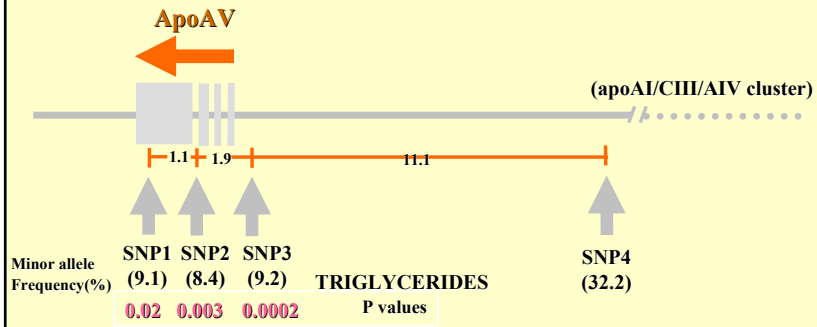
The Role ApoAV in Humans?

Association Study I: ApoAV Polymorphisms and Plasma Parameters



Genotyped 500 normal individuals phenotyped for plasma:
-Triglycerides
-IDL, LDL, HDL, VLDL Mass
-HDL, LDL Cholesterol
-ApoAI, ApoB

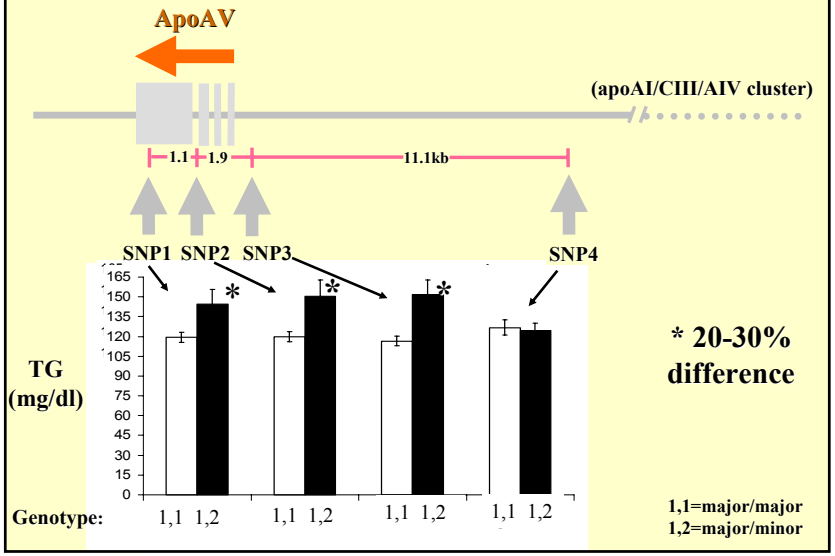
Association Study I: ApoAV Polymorphisms and Plasma Parameters



Genotyped 500 normal individuals phenotyped for plasma:

- **Triglycerides***
- **IDL, LDL, HDL, VLDL Mass***
- **HDL, LDL Cholesterol**
- **ApoAI, ApoB**

Effect of ApoAV SNPs on Triglyceride Levels

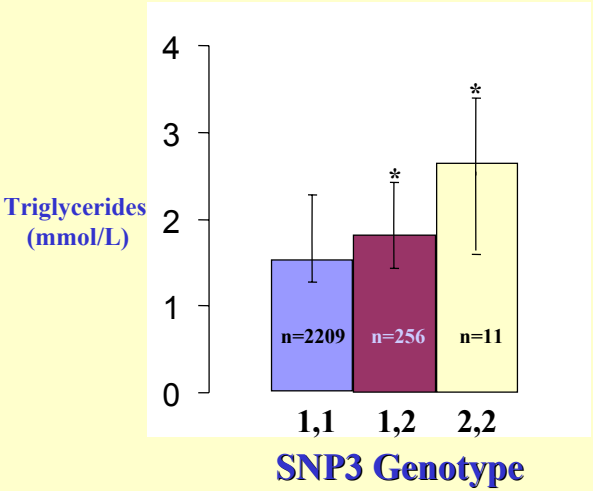


*** 20-30% difference**

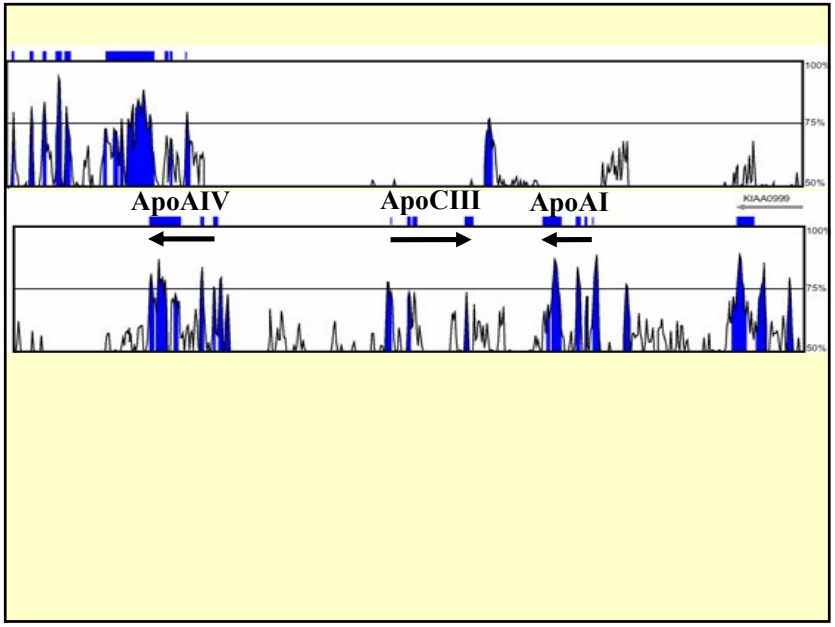
Minor ApoAV haplotype in human is associated with increased plasma triglycerides

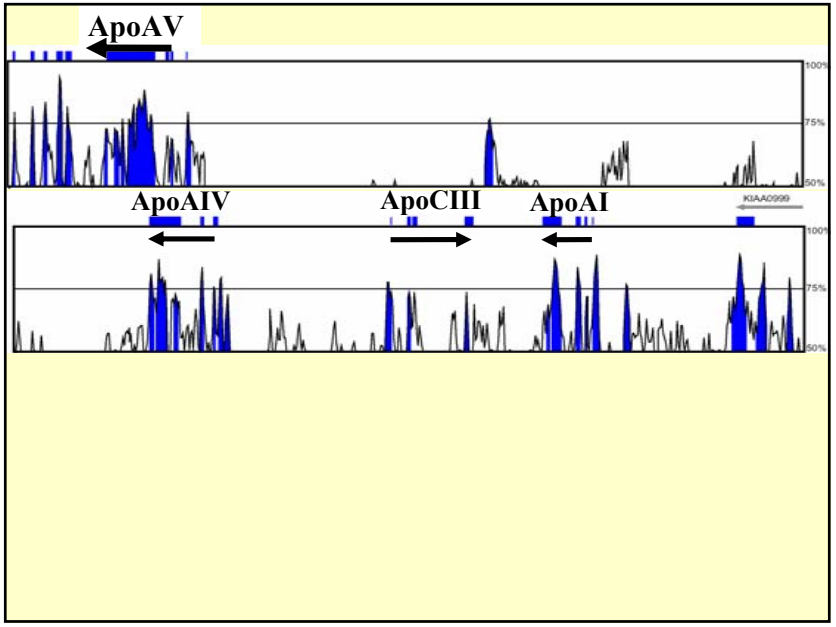
Association Studies….. Is this finding reproducible????

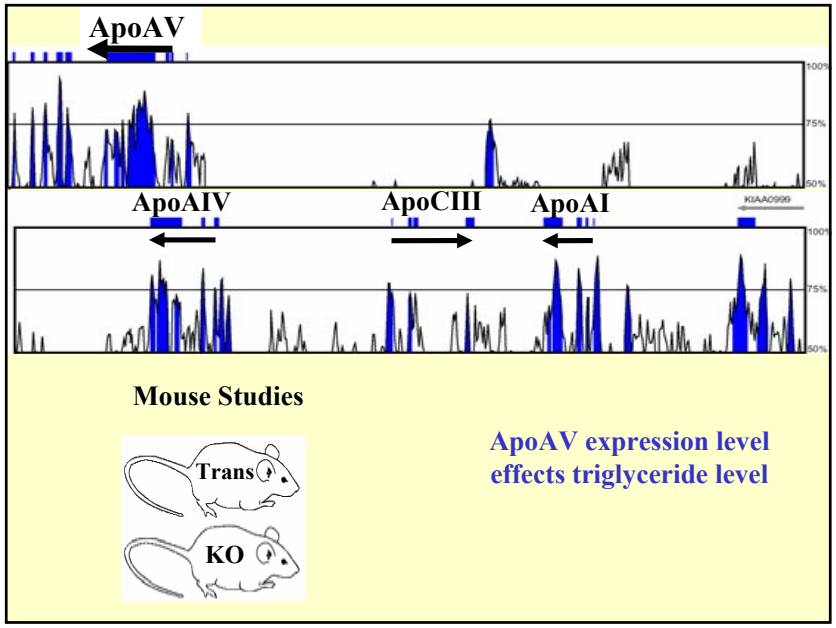
Association Study III: 2200 Male English Civil Servants

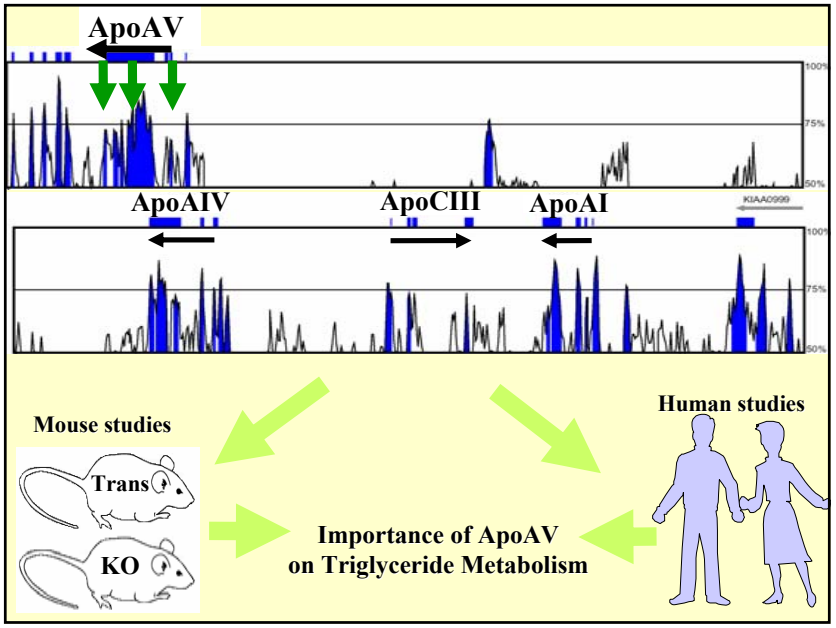


*** P<0.01**


ApoAIV
ApoCIII
ApoAI
KIAA0999
100%
75%
50%

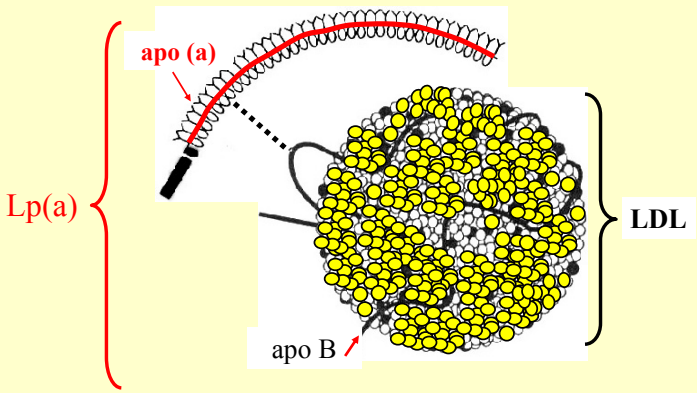

ApoAV
ApoAIV
ApoCIII
ApoAI
KIAA0999
100%
75%
50%


ApoAV
ApoAIV
ApoCIII
ApoAI
KIAA0999
100%
75%
50%
Mouse Studies
Trans
KO
ApoAV expression level
effects triglyceride level


ApoAV
ApoAIV
ApoCIII
ApoAI
KIAA0999
100%
75%
50%
Mouse studies
Trans
KO
Human studies
Importance of ApoAV
on Triglyceride Metabolism

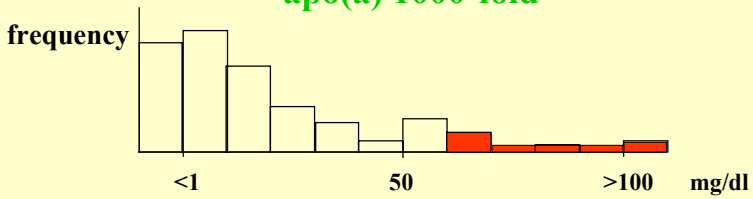
How a Room Full of Primates are Better Than One for Finding Primate Specific Regulatory Elements

Apolipoprotein(a)



Apo(a) Plasma Levels are Highly Variable

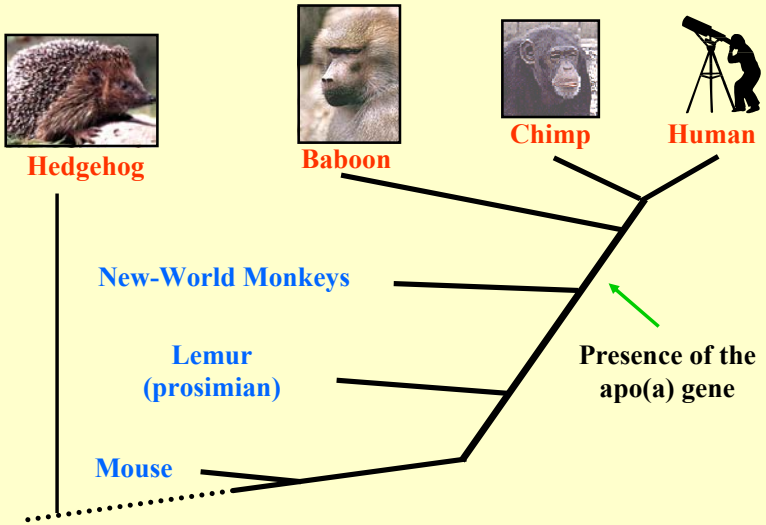
* apo(a) 1000-fold



* standard protein ~10-fold

~ 50% of apo(a) variability in levels is attributable to the regulation of the transcription of the gene

Apo(a) : Limited Distribution Among Mammals

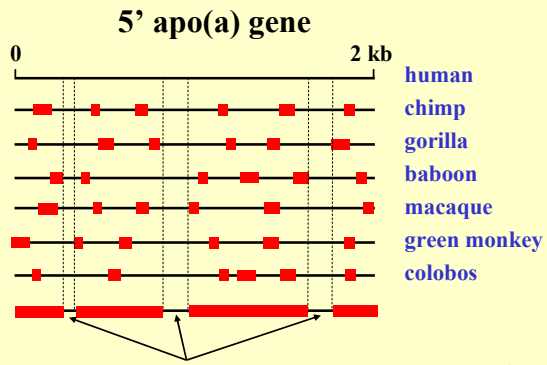


Strategy to identify regulatory elements of a recently evolved gene:

Phylogenetic Shadowing

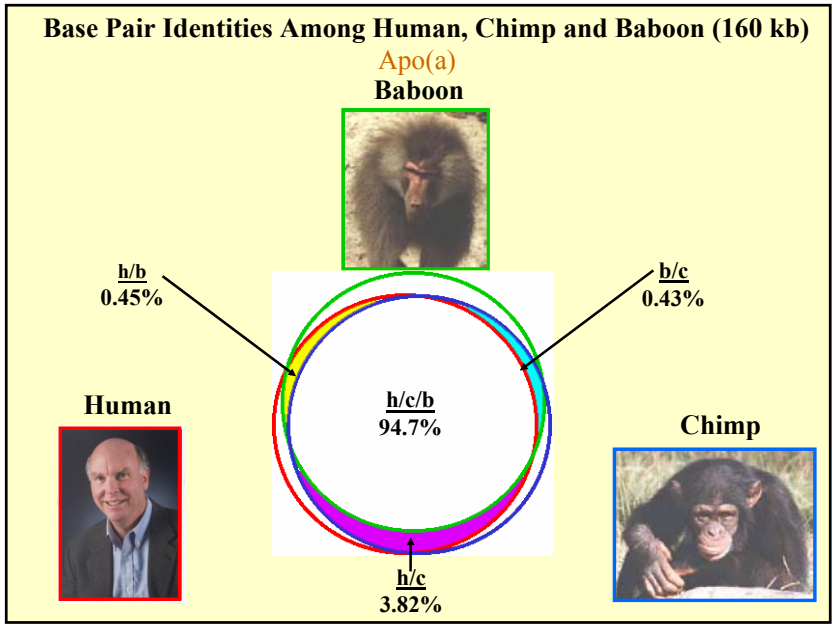
sequence conservation across multiple close species as a means to increase evolutionary distance

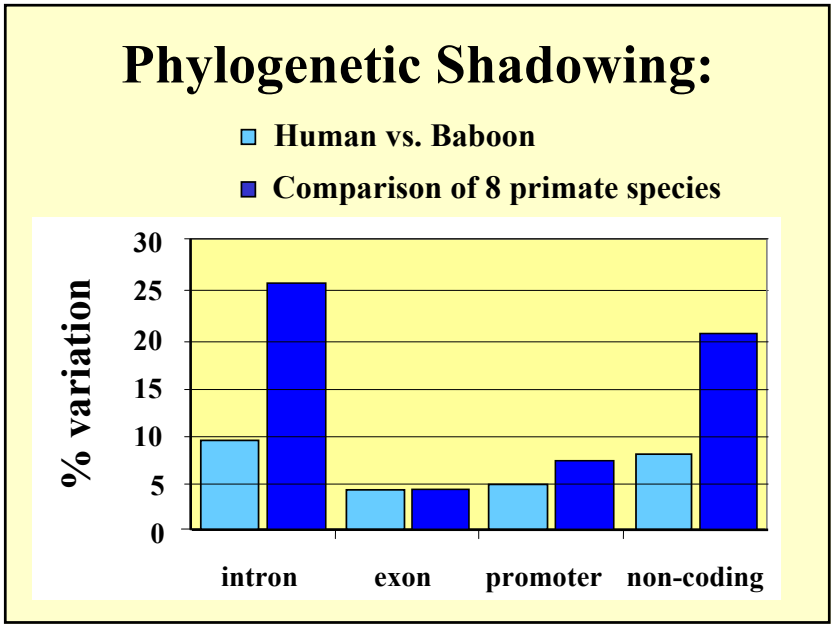
Phylogenetic Shadowing:



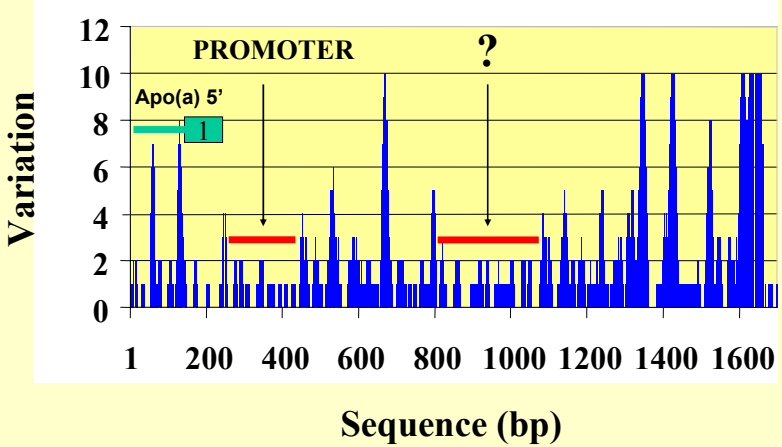
Sequence elements conserved across all species: candidate regulatory element

: nucleotide difference with at least one species

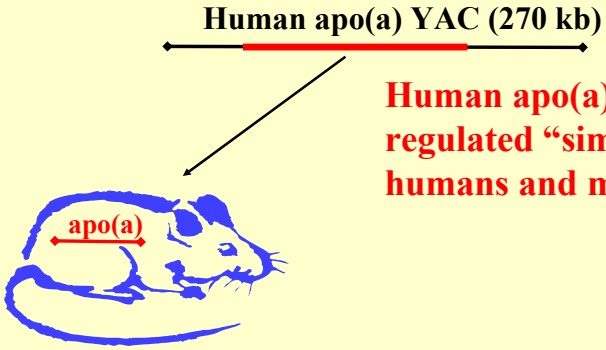

Base Pair Identities Among Human, Chimp and Baboon (160 kb)
Apo(a)
Baboon
h/b
0.45%
b/c
0.43%
h/c/b
94.7%
Human
Chimp
h/c
3.82%


Phylogenetic Shadowing:
Human vs. Baboon
Comparison of 8 primate species
% variation
30
25
20
15
10
5
0
intron
exon
promoter
non-coding

Sequence Conserved Across 8 Primate Species

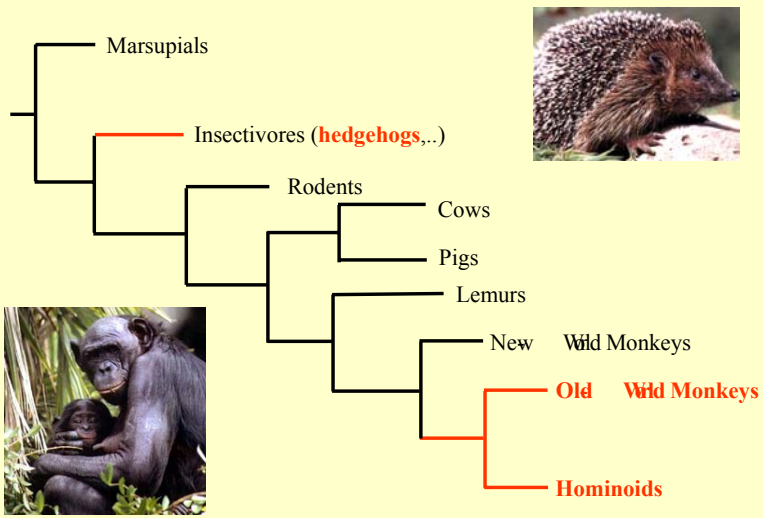


How Do We Test In Vivo Regulatory Elements of a Gene not Present in Mice?



Human apo(a) is regulated "similarly" in humans and mice:

apo(a) present in some primates but...



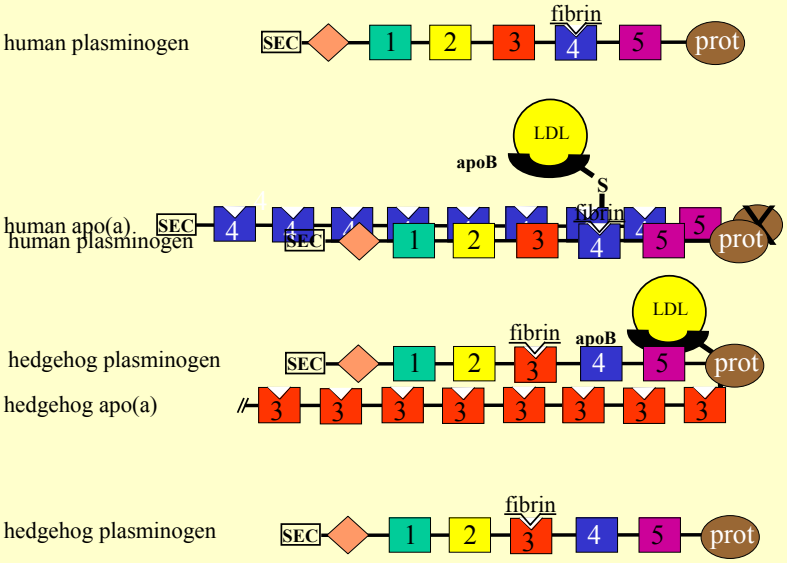
The hedgehog and primate apo(a) gene an example of a) gene loss or b) convergent evolution ?

Gene Loss: When species had a gene but lost it via deletions or inactivation

Convergent evolution: When genes evolve independently to the same function.

Genomic BAC library

Sequence hedgehog apo(a) gene



Evolution acts in mysterious ways

Sequence data from many species may reveal some of its motives

LBNL

Jan-Fang Cheng
Len Pennacchio
Dario Boffeli
Jody Schwartz
Ivan Ovcharenko
Alex Poliakov

Inna Dubchak –
Lior Pachter – UCB math / LBNL

http://pipeline.lbl.gov

Ronald Krauss

http://www-gsd.lbl.gov/

UT-SW
Jonathan Cohen
Helen Hobbs
Jaroslav Hubacek

MCW
Michael Olivier

Pasteur Institute-Lille
Jamila Fruchart
Jean-Charles Fruchart