Sifting Comparative Sequence:
Lipid Metabolism Genes and Regulation

Outline

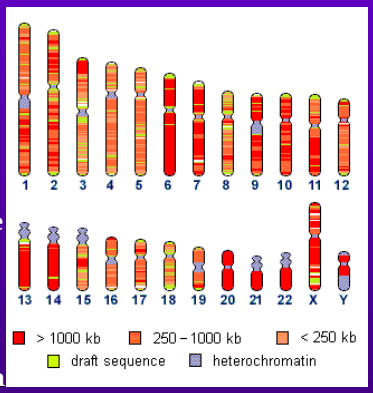
Computational Tools and Databases
- -VISTA
- -Cardiovascular Gene Resource
  - -Examples
- -Pipeline (Godzilla)
  - -Human/Mouse Genome Comparison

Identification of a Novel Gene (ApoAV)
- -Functional Characterization

Background

-Majority of human genomic sequence is available.
- -63% Finished
- -34% Draft

-Mouse/Rat genomic sequence is also available.



Challenge:

Raw Sequence----> Biological Function

Raw Sequence----> Biological Function outside of coding regions

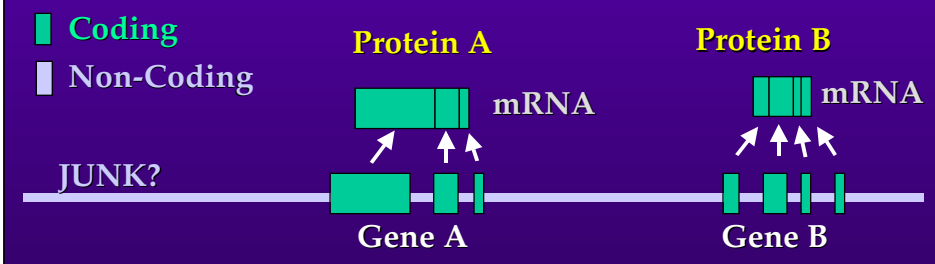
In general, functionally important sequences are conserved.

Hypothesis: Conserved sequences are functionally important.

Categories of DNA

Coding ~5%

Non-coding ~95%



VISUALIZATION TOOLS FOR ALIGNMENTS

VISTA is an integrated system for global sequence alignment and visualization, designed for comparative genomic analysis.

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

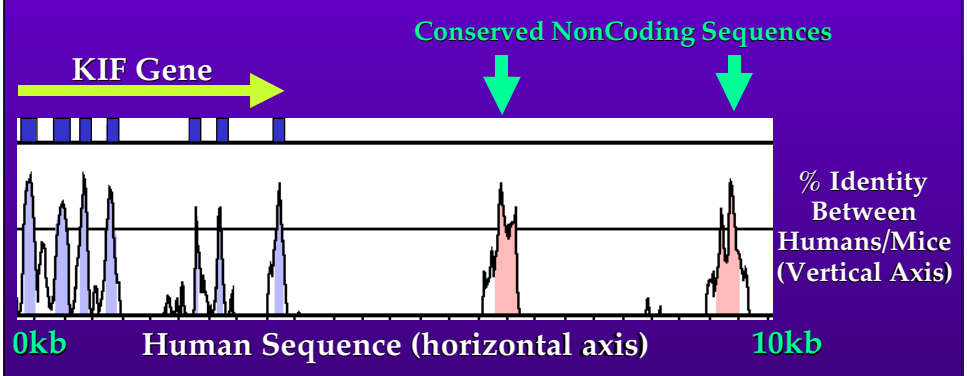
AVID – The Alignment Engine Behind VISTA
(Lior Pachter)

- Very fast global alignment of megabases of sequence.
- Provides details about ordered and oriented contigs, and accurate placement in the finished sequence.
- Full integration with repeat masking.

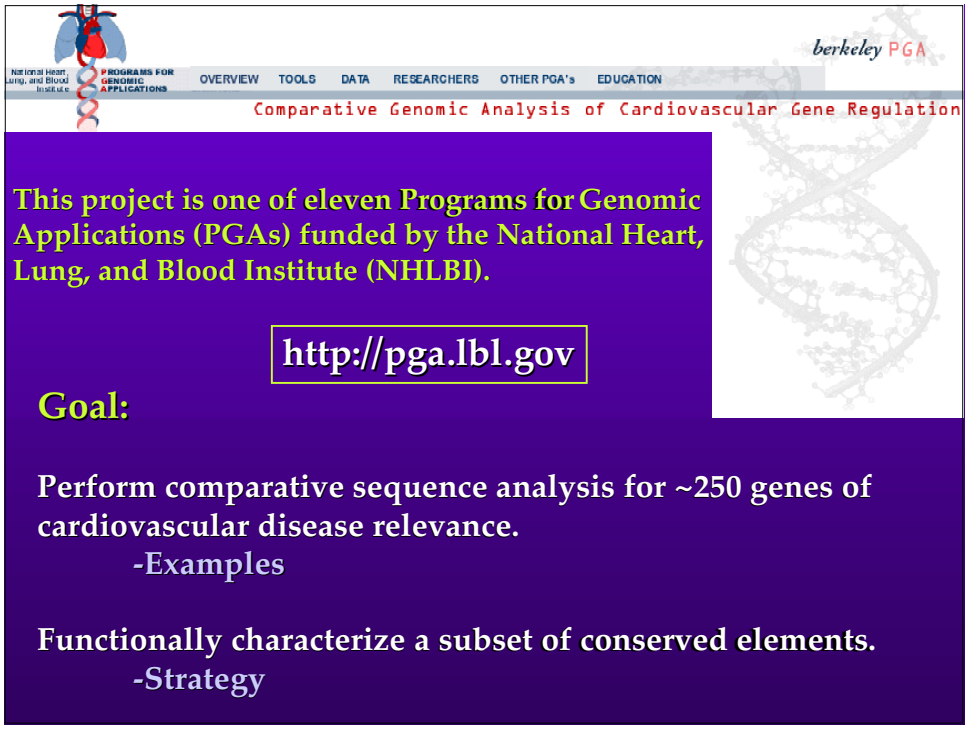
Visualization

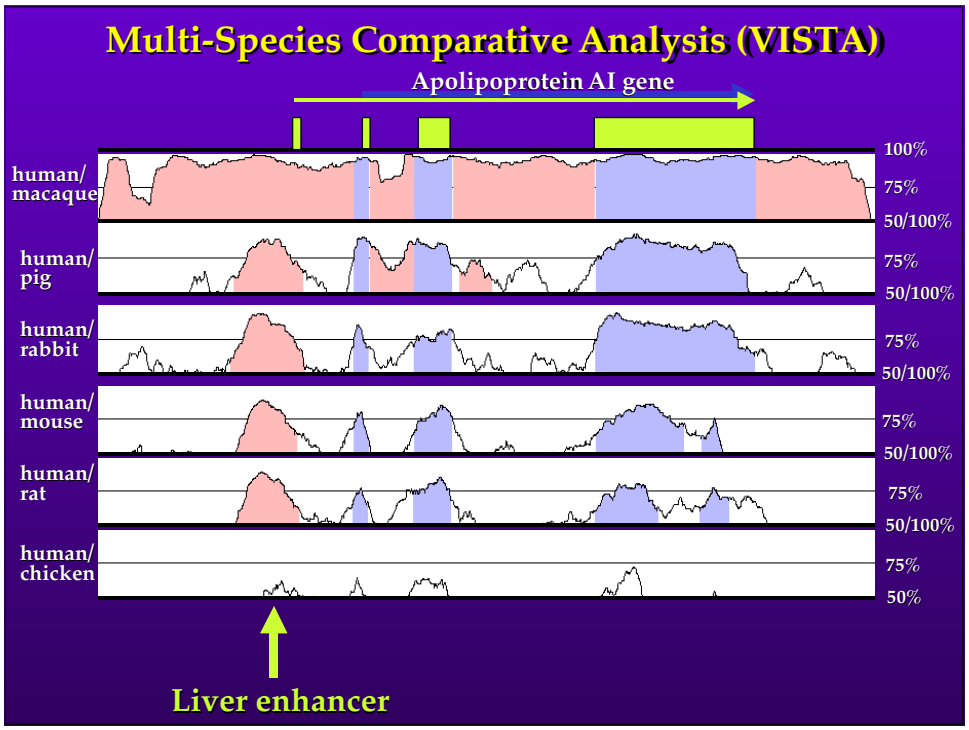
```
tggtaacattcaaattatg-----ttctcaaagtgagcatgaca-actttttccatgg
|| | ||||  |  |  ||     || | | |    |||||| |  ||   |   | ||
tgatgacatctatttgctgtttccttttagaaactgcatgagagcctggctagtaggg
```

VISTA Plot
(VISual Tool for Alignment)



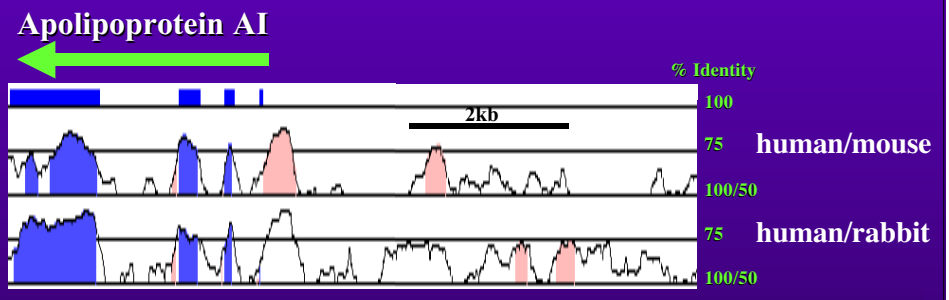
http://www-gsd.lbl.gov/vista


berkeley PGA
PROGRAMS FOR GENOMIC APPLICATIONS
OVERVIEW TOOLS DATA RESEARCHERS OTHER PGA's EDUCATION
Comparative Genomic Analysis of Cardiovascular Gene Regulation
This project is one of eleven Programs for Genomic Applications (PGAs) funded by the National Heart, Lung, and Blood Institute (NHLBI).
http://pga.lbl.gov
Goal:
Perform comparative sequence analysis for ~250 genes of cardiovascular disease relevance.
-Examples
Functionally characterize a subset of conserved elements.
-Strategy


Multi-Species Comparative Analysis (VISTA)
Apolipoprotein AI gene
human/ macaque
human/ pig
human/ rabbit
human/ mouse
human/ rat
human/ chicken
100%
75%
50/100%
75%
50/100%
75%
50/100%
75%
50/100%
75%
50/100%
75%
50%
Liver enhancer

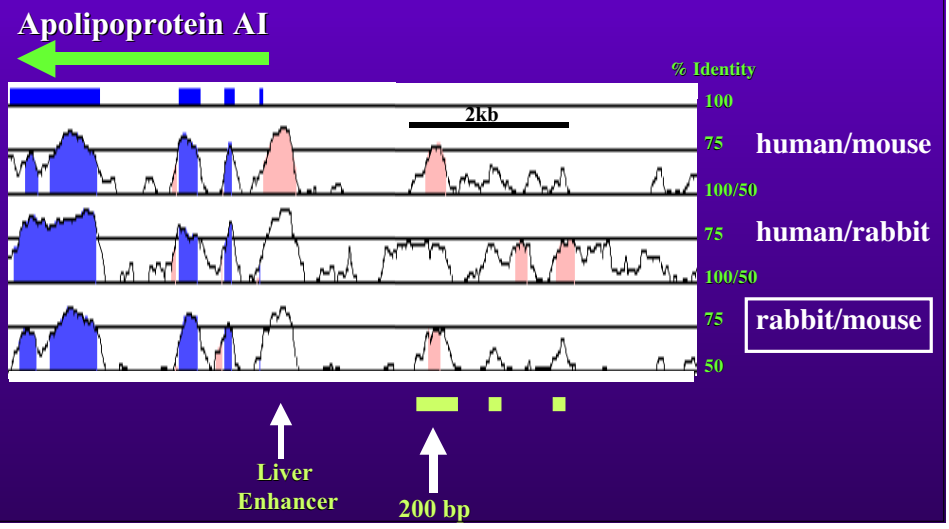
Three Species Sequence Comparison (Humans, Mice and Rabbits)

Aiding in the refinement of potential regulatory sequences



Three Species Sequence Comparison (Humans, Mice and Rabbits)

Aiding in the refinement of potential regulatory sequences



Power of Deep Sequence Alignments

3-way Comparison

Hepatic Site C | CCAAT-Box

```
Mouse  NNNNAGCCTCAGGAACAGAGCTGATCCTTGAACTCT-AAGTTCCACATCGCCAGCAAAAG
Rabbit NNNN-GCCCTAGGCAGCGAGCTGATCCTTGAACTCT-AAGTTCCACATGGCCAGCACCAG
Human  NNNNAGTCCCAGGCAGCAGAGCTGATCCTTGAACTCTTAAGTTCCACATTGCCAGCACCAG

Mouse  TAAGCAGTGGCAGGGCCAG-GCTGAGCTTATCAGTCTCCCAGCCCAGCCCCTGCCCAGAC
Rabbit GGAGCAGTGACTAGGCCCA-GCTGGGCTTATCAGCCTCACAGCCCAGCCCCTGCCTGGAG
Human  TGAGCAGCAAGAGGGCCAGGGCTGCGCTTATCAGCCTCCCAGCCCAGACCCTGCCTGGAG
```

TATA-Box

```
Mouse  ACATATATAGACCAGCGAAGAAGAGCTGGACAGCC-
Rabbit ACATAAATAGCCAGCGGCCA---GCTGGCCGCAGG
Human  ACATAAATAGCCCTCCAAGA---GCTGGCTGC---
```

4-way Comparison

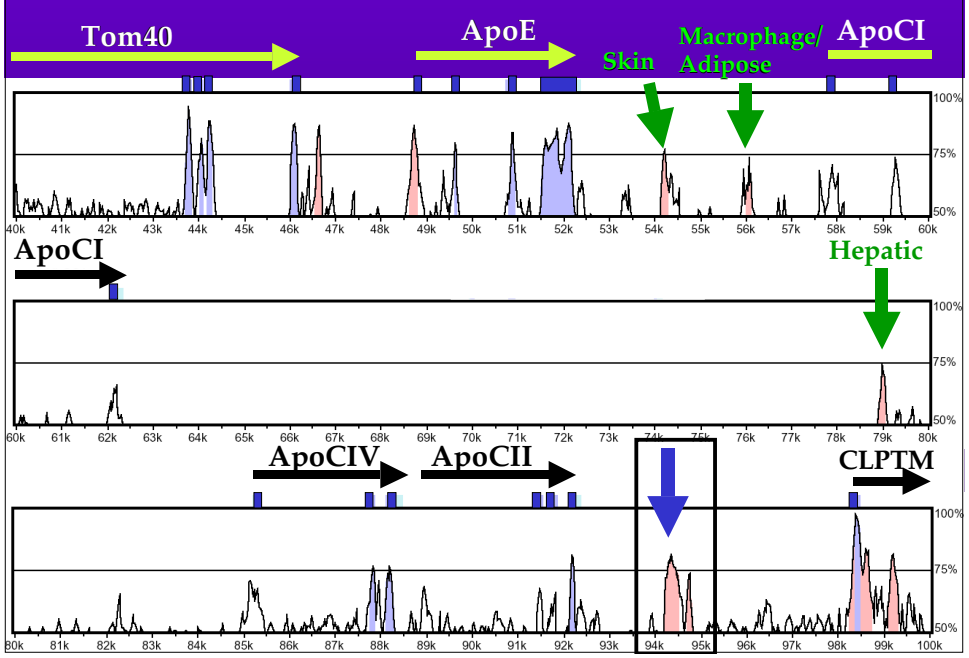
Hepatic Site C | CCAAT-Box

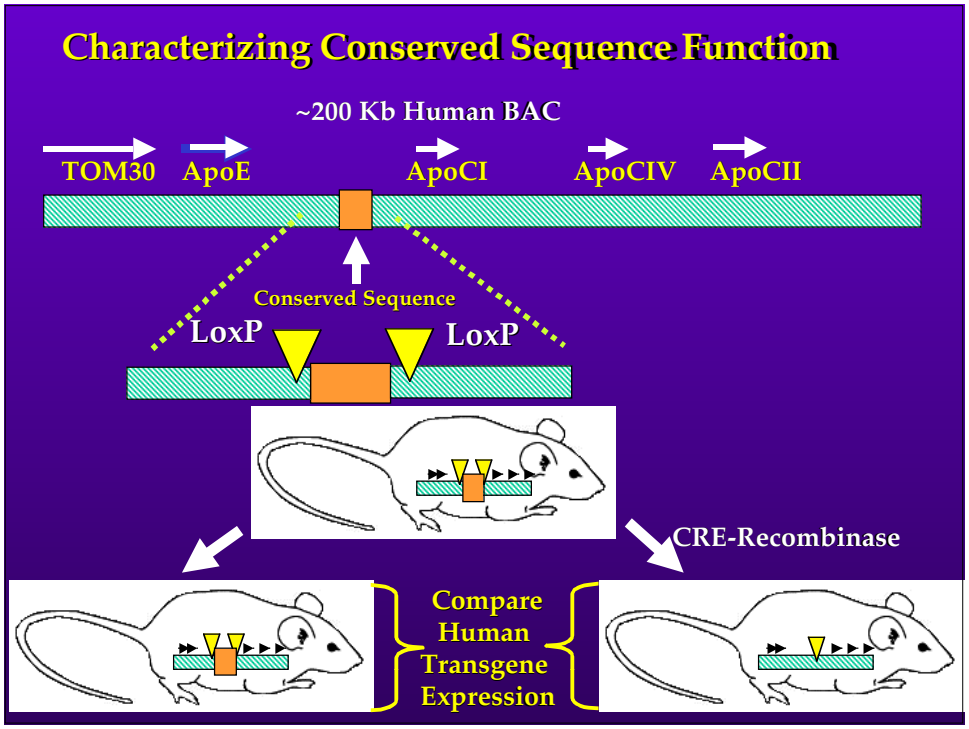
```
Mouse   AGCCTCAGGAACA-GAGC-TGATCCTTGAACTCT-AAGTTCCACATCGCCAGCAAAAGTA
Rabbit  -GCCCTAGGGAGG-GAGC-TGATCCTTGAACTCT-AAGTTCCACATGGCCAGCGACCAGGG
Human   AGTCCCAGGGACA-GAGC-TGATCCTTGAACTCTTAAGTTCCACATTGCCAGCGACCAGTG
Chicken CTCTCCCGGGGCGTGGGACGATCCTTGAACTCT-AGGCCCACATCGCCCCGCCCGCGGA

Mouse   AGCAGTGGCAGGGC--CAG-GCTGAGCTTATCAGTCTCCCAGCCCAGCCCCTGCCCACAC
Rabbit  AGCAGTGACTAGGC--CCA-GCTGGGCTTATCAGCCTCAGAGCCCAGCCCCTGCCTGGAG
Human   AGCAGCAACAGGGC--CAGGGCTGGGCTTATCAGCCTCCCAGCCCAGACCCTGCTGCAG
Chicken GTGATTTCTTGGGCTGGGCGCTG-GCTTATCTGTGCGGAACT--CCCCCTG-TG---
```

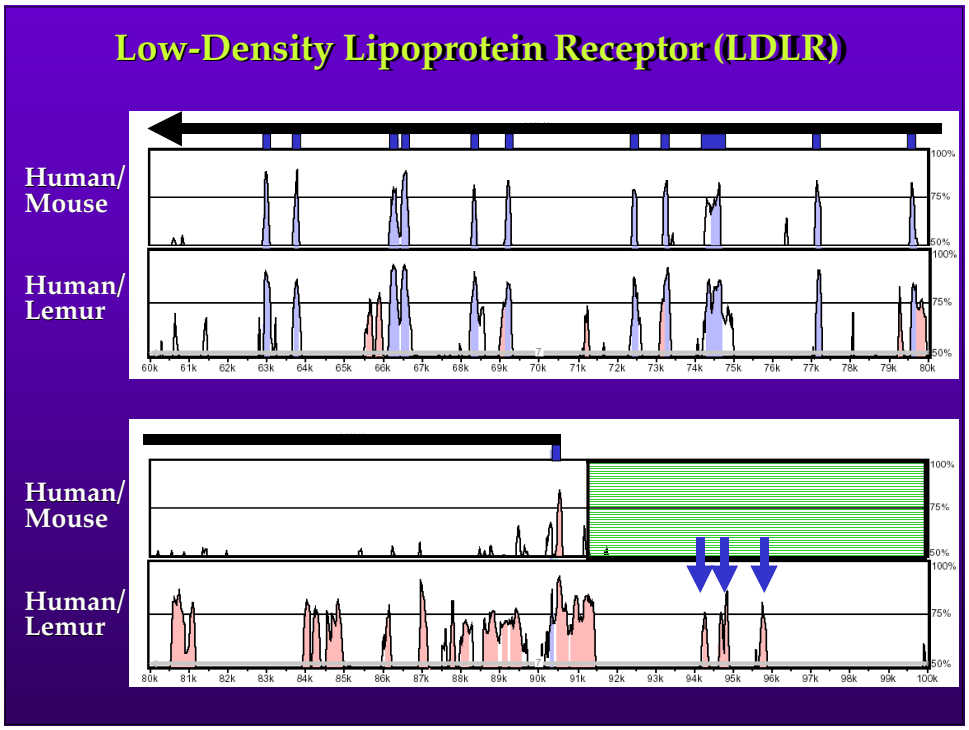
TATA-Box

```
Mouse   ACATATATAGACCAGCGAAGAAGAGCTGGACAGCC-
Rabbit  ACATAAATAGCCAGCGGCCA---GCTGGCCGCAGG
Human   ACATAAATAGCCCTCCAAGA---GCTGGCTGC---
Chicken -CATAAATAGCGGCGCCGGGA---ACGGGCTCAC-
```

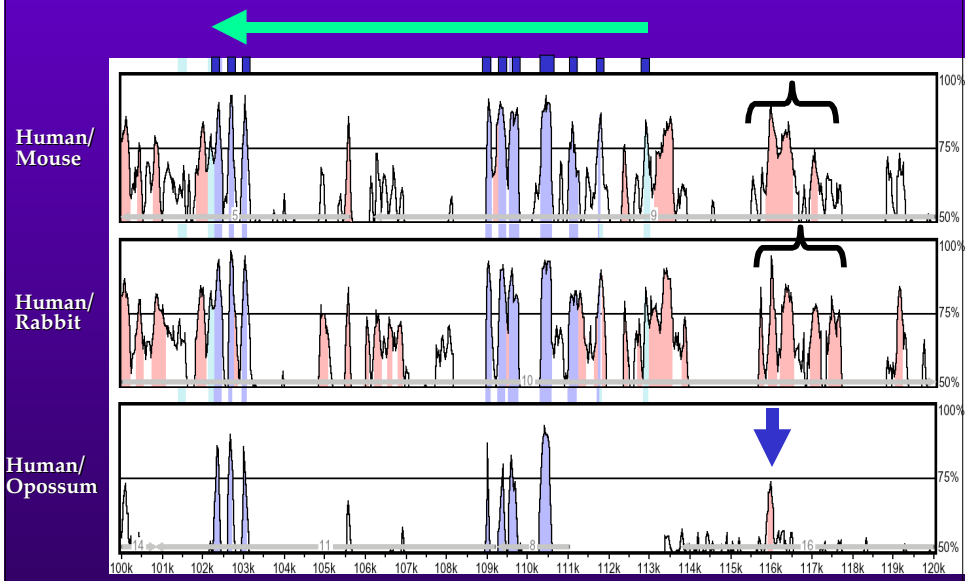
VISTA Plot of Human/Mouse Apo Gene Cluster on chr 19




Characterizing Conserved Sequence Function
~200 Kb Human BAC
TOM30
ApoE
ApoCI
ApoCIV
ApoCII
Conserved Sequence
LoxP
LoxP
CRE-Recombinase
Compare
Human
Transgene
Expression


Low-Density Lipoprotein Receptor (LDLR)
Human/
Mouse
Human/
Lemur
Human/
Mouse
Human/
Lemur

Nuclear Hormone Receptor:LXR-Alpha



VISUALIZATION TOOLS FOR ALIGNMENTS

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

mVISTA: main VISTA
-standard comparative sequence plots

rVISTA: regulatory VISTA
-conserved transcription factor binding sites

Jan 2002: 710 users
320 copies have been distributed

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

GODZILLA

Ivan Ovcharenko
Lior Pachter
Inna Dubchak

http://pipeline.lbl.gov/

ATGTCCA
TATATAT
GTCTTCT

Current Mouse Contigs and Supercontigs

ATGTCCA
NNNNNNN
GTCTTCT

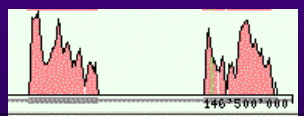
Mask repetitive elements *RepeatMasker*

Hs1:1001-2231

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

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

Align Human and Mouse sequences (*AVID*)

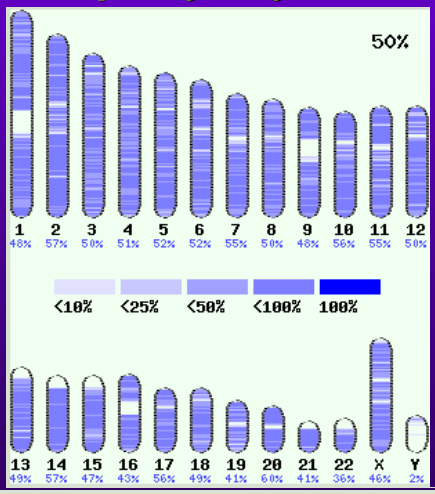


Build VISTA plots

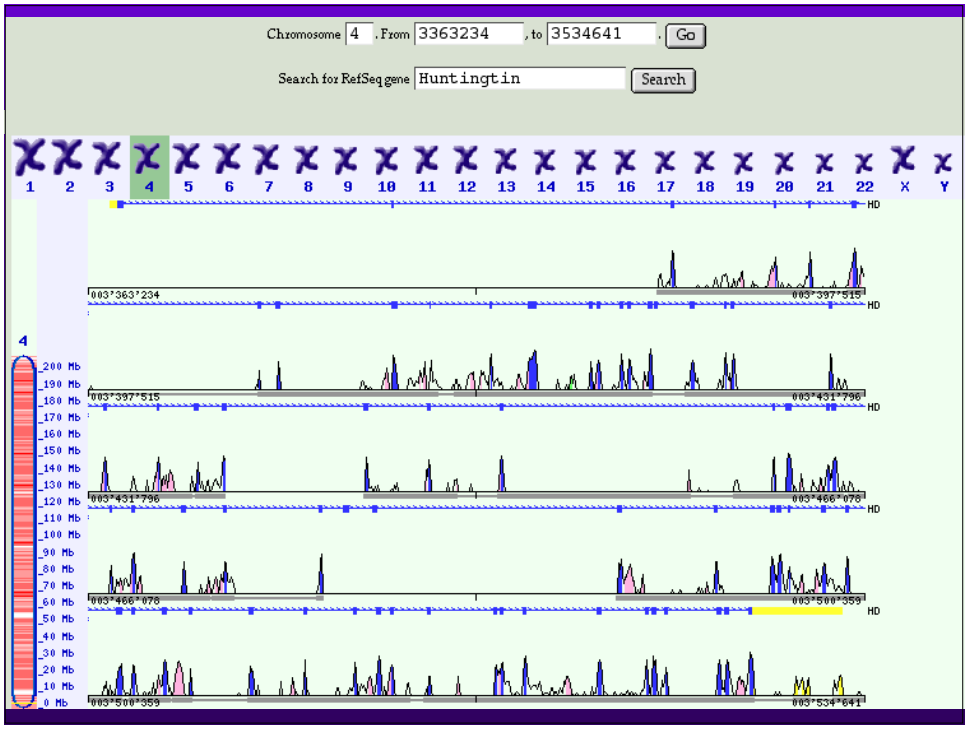
A PIPELINE FOR REAL-TIME COMPARATIVE ANALYSIS OF THE HUMAN AND MOUSE GENOMES

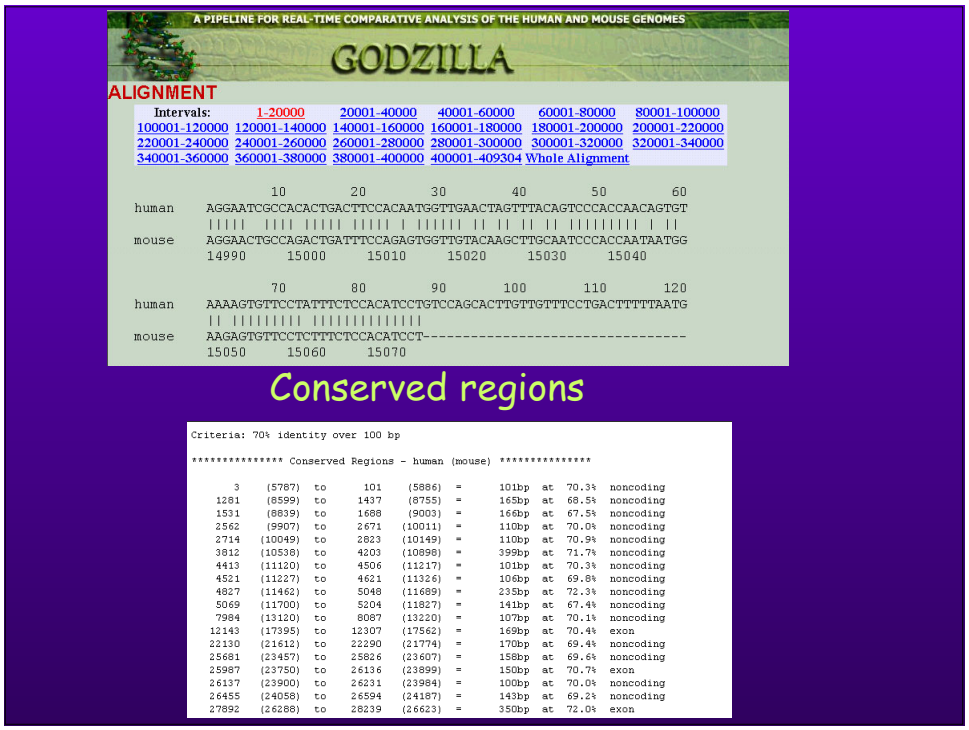
GODZILLA

Mouse SuperContig Coverage of Human Genome



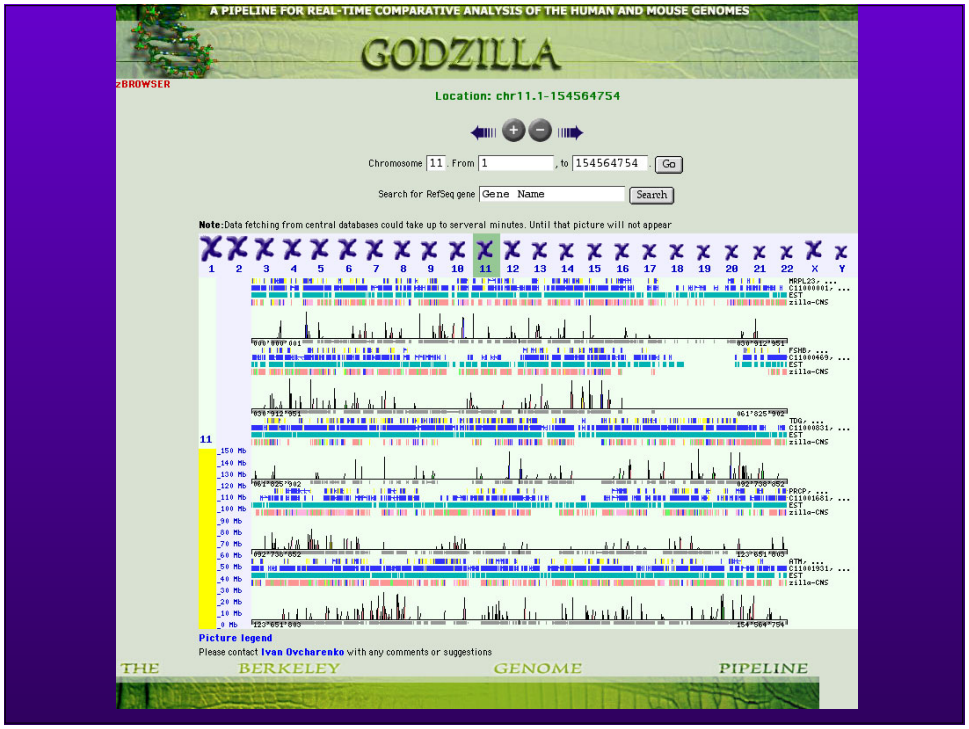
THE BERKELEY GENOME PIPELINE

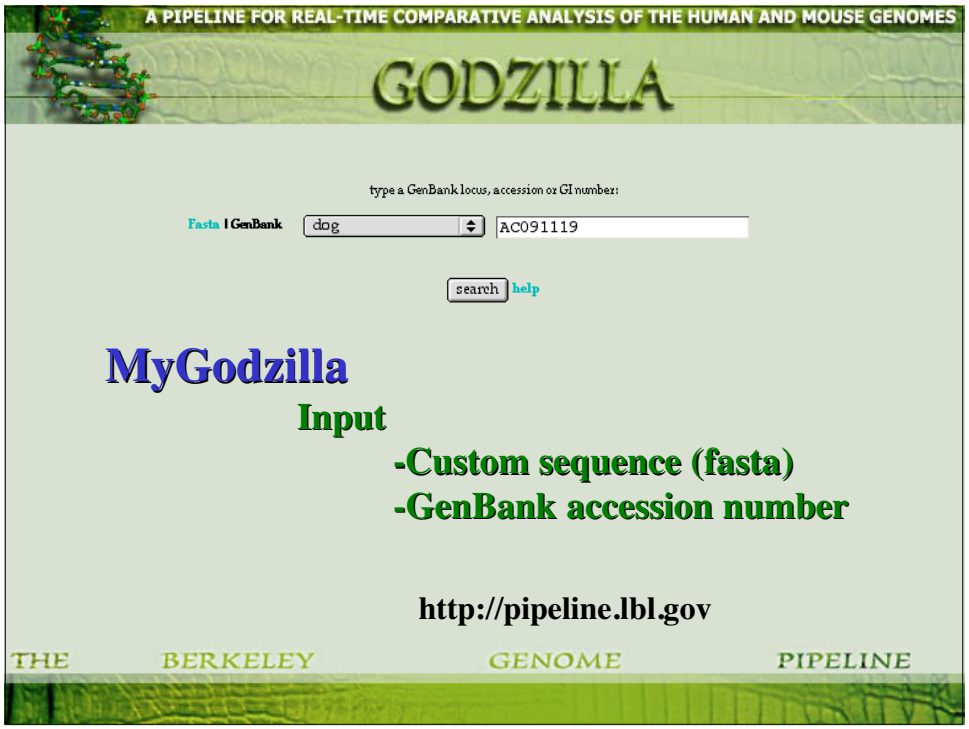


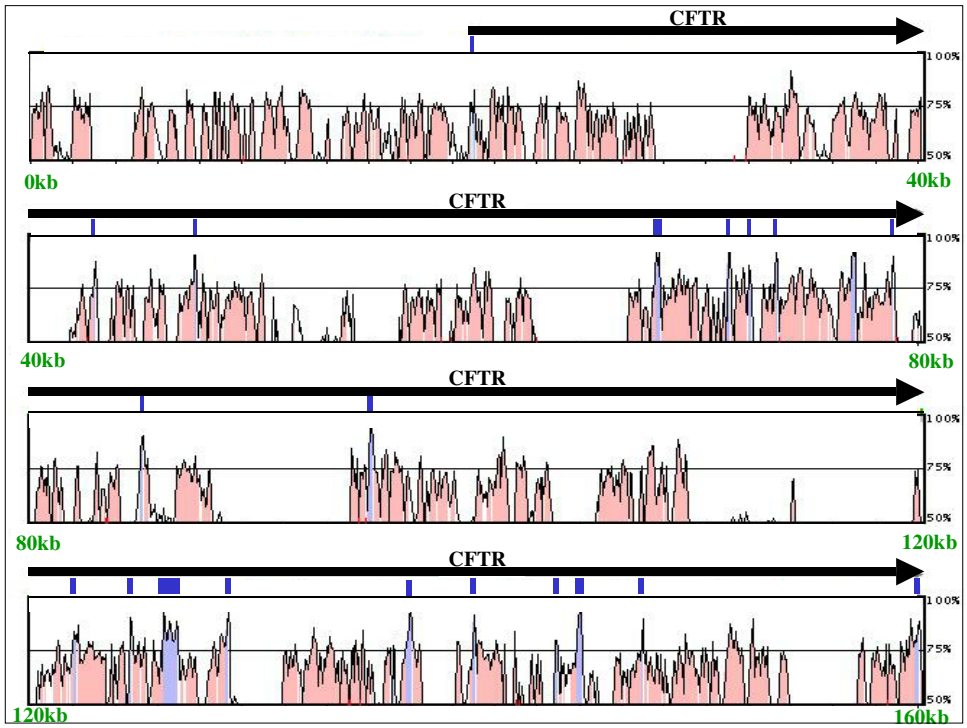


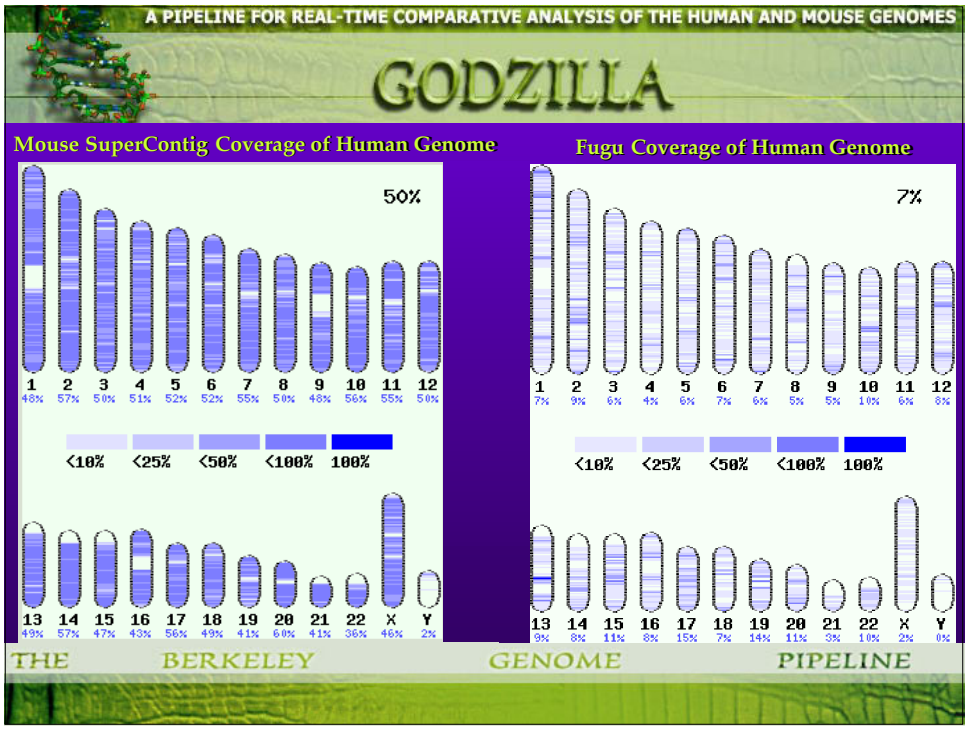
Conserved regions

```
Criteria: 70% identity over 100 bp

*************** Conserved Regions - human (mouse) ***************

       3    (5787)  to     101    (5886)  =     101bp  at  70.3%  noncoding
    1281    (8599)  to    1437    (8755)  =     165bp  at  68.5%  noncoding
    1531    (8839)  to    1688    (9003)  =     166bp  at  67.5%  noncoding
    2562    (9907)  to    2671   (10011)  =     110bp  at  70.0%  noncoding
    2714   (10049)  to    2823   (10149)  =     110bp  at  70.9%  noncoding
    3812   (10538)  to    4203   (10898)  =     399bp  at  71.7%  noncoding
    4413   (11120)  to    4506   (11217)  =     101bp  at  70.3%  noncoding
    4521   (11227)  to    4621   (11326)  =     106bp  at  69.8%  noncoding
    4827   (11462)  to    5048   (11689)  =     235bp  at  72.3%  noncoding
    5069   (11700)  to    5204   (11827)  =     141bp  at  67.4%  noncoding
    7984   (13120)  to    8087   (13220)  =     107bp  at  70.1%  noncoding
   12143   (17395)  to   12307   (17562)  =     169bp  at  70.4%  exon
   22130   (21612)  to   22290   (21774)  =     170bp  at  69.4%  noncoding
   25681   (23457)  to   25826   (23607)  =     158bp  at  69.6%  noncoding
   25987   (23750)  to   26136   (23899)  =     150bp  at  70.7%  exon
   26137   (23900)  to   26231   (23984)  =     100bp  at  70.0%  noncoding
   26455   (24058)  to   26594   (24187)  =     143bp  at  69.2%  noncoding
   27892   (26288)  to   28239   (26623)  =     350bp  at  72.0%  exon
```

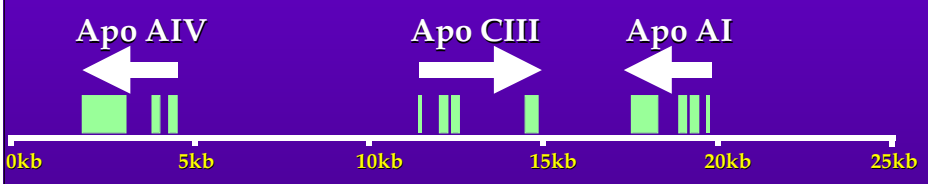

A PIPELINE FOR REAL-TIME COMPARATIVE ANALYSIS OF THE HUMAN AND MOUSE GENOMES
GODZILLA
Location: chr11.1-154564754
Chromosome 11 From 1 to 154564754 Go
Search for RefSeq gene Gene Name Search
Note: Data fetching from central databases could take up to serveral minutes. Until that picture will not appear
Picture legend
Please contact Ivan Ovcharenko with any comments or suggestions
THE BERKELEY GENOME PIPELINE


A PIPELINE FOR REAL-TIME COMPARATIVE ANALYSIS OF THE HUMAN AND MOUSE GENOMES
GODZILLA
type a GenBank locus, accession or GI number:
Fasta | GenBank
dog
AC091119
search
help
MyGodzilla
Input
-Custom sequence (fasta)
-GenBank accession number
http://pipeline.lbl.gov
THE BERKELEY GENOME PIPELINE

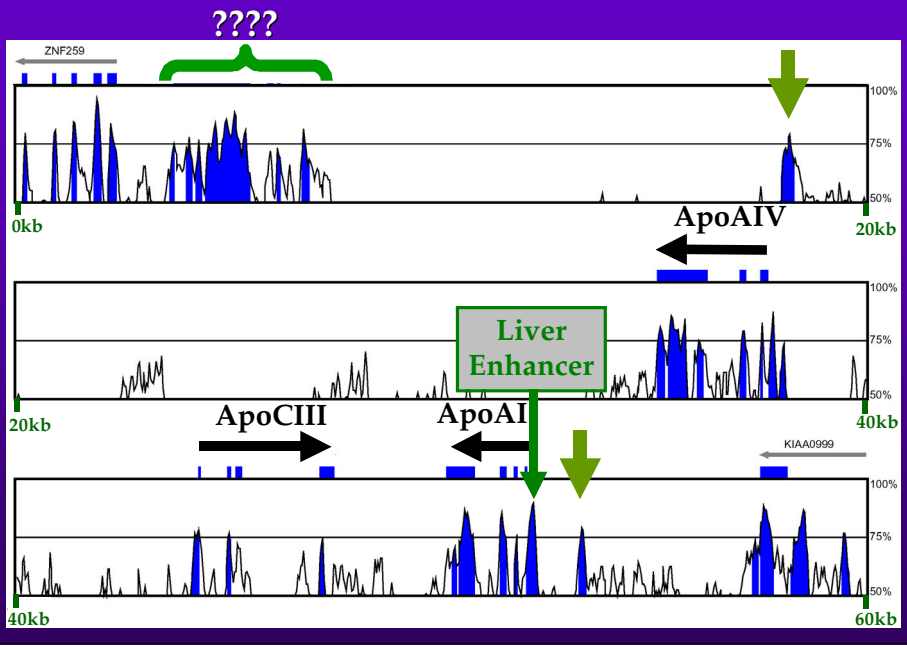

CFTR
100%
75%
50%
0kb
40kb
CFTR
100%
75%
50%
40kb
80kb
CFTR
100%
75%
50%
80kb
120kb
CFTR
100%
75%
50%
120kb
160kb


A PIPELINE FOR REAL-TIME COMPARATIVE ANALYSIS OF THE HUMAN AND MOUSE GENOMES
GODZILLA
Mouse SuperContig Coverage of Human Genome
50%
1 2 3 4 5 6 7 8 9 10 11 12
48% 57% 50% 51% 52% 52% 55% 50% 48% 56% 55% 50%
<10% <25% <50% <100% 100%
13 14 15 16 17 18 19 20 21 22 X Y
49% 57% 47% 43% 56% 49% 41% 60% 41% 36% 46% 2%
Fugu Coverage of Human Genome
7%
1 2 3 4 5 6 7 8 9 10 11 12
7% 9% 6% 4% 6% 7% 6% 5% 5% 10% 6% 8%
<10% <25% <50% <100% 100%
13 14 15 16 17 18 19 20 21 22 X Y
9% 8% 11% 8% 15% 7% 14% 11% 3% 10% 2% 0%
THE BERKELEY GENOME PIPELINE

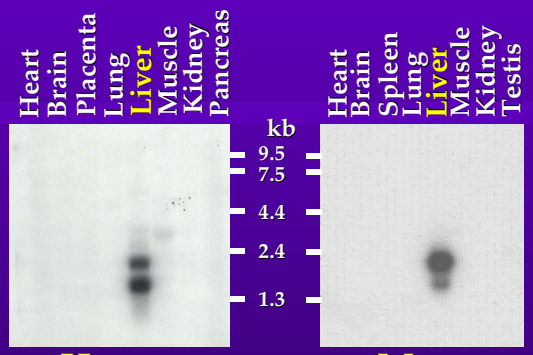
Human Chromosome 11q23 Apolipoprotein Gene Cluster



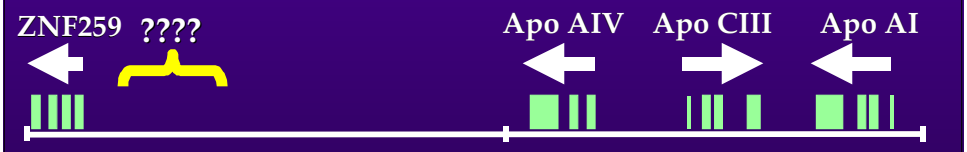
Human/Mouse Apolipoprotein Gene Cluster Sequence Comparison



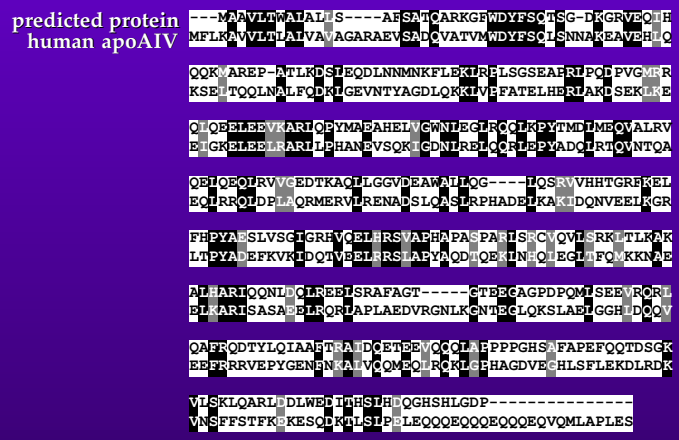
Northern Blot Analysis of Conserved Sequence



Liver transcripts detected in both human and mouse.

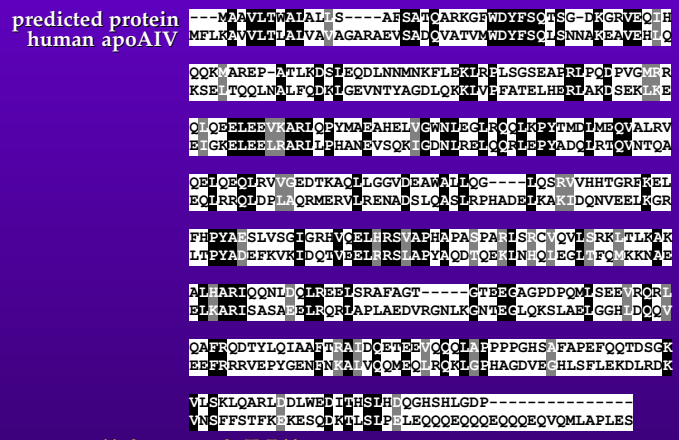


Predicted protein has homology to ApoAIV



Identity: 26%
Similarity: 45%

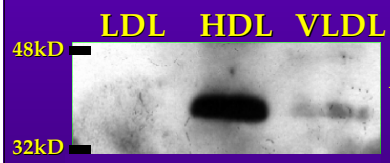
Predicted protein has homology to ApoAIV



Identity: 26%
Similarity: 45%

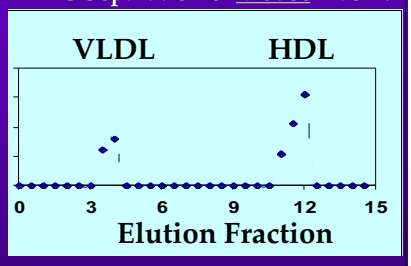
ApoAV associates with HDL and VLDL

Western Blot of Human Plasma

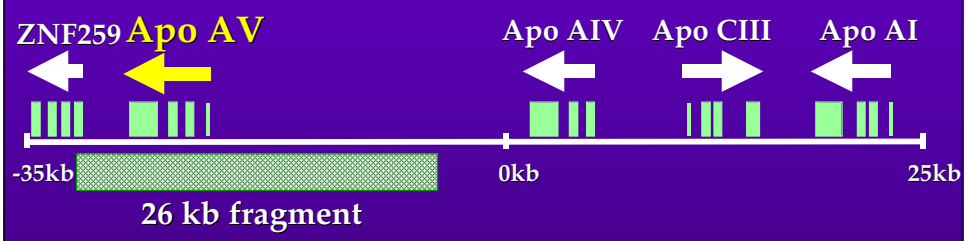


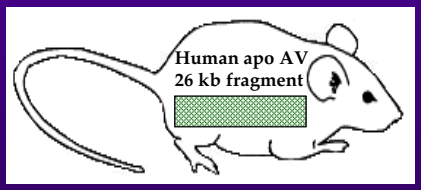
Anti-ApoAV Antibodies

FPLC Separation of Mouse Plasma

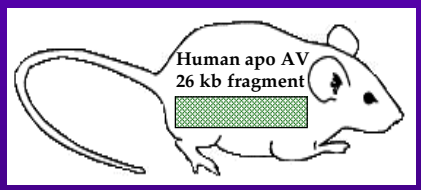


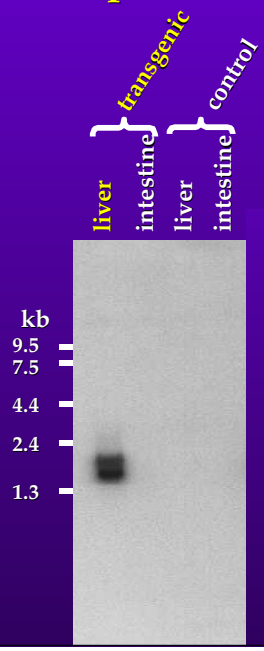
Generation of human ApoAV transgenic mice



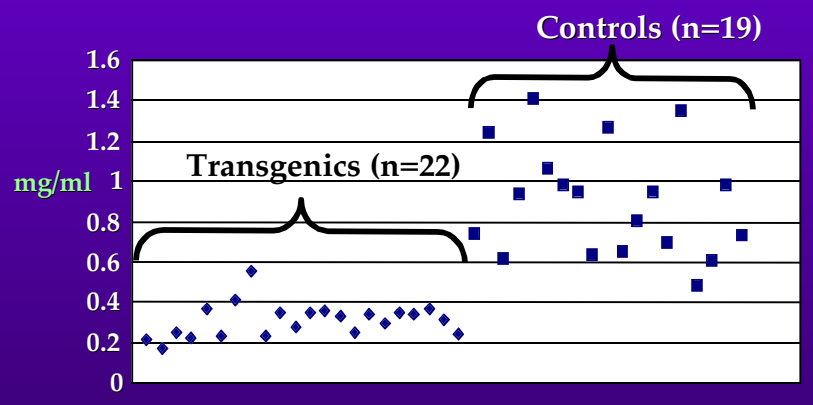


Liver expression of human ApoAV transgene





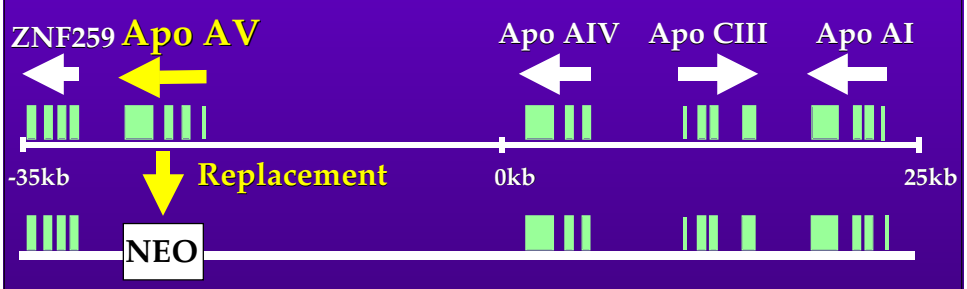
ApoAV transgenics have decreased plasma triglycerides

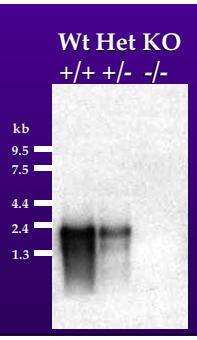


Average (SD)

Transgenics 0.309(0.084)
Controls 0.902(0.272)
$p<0.001$

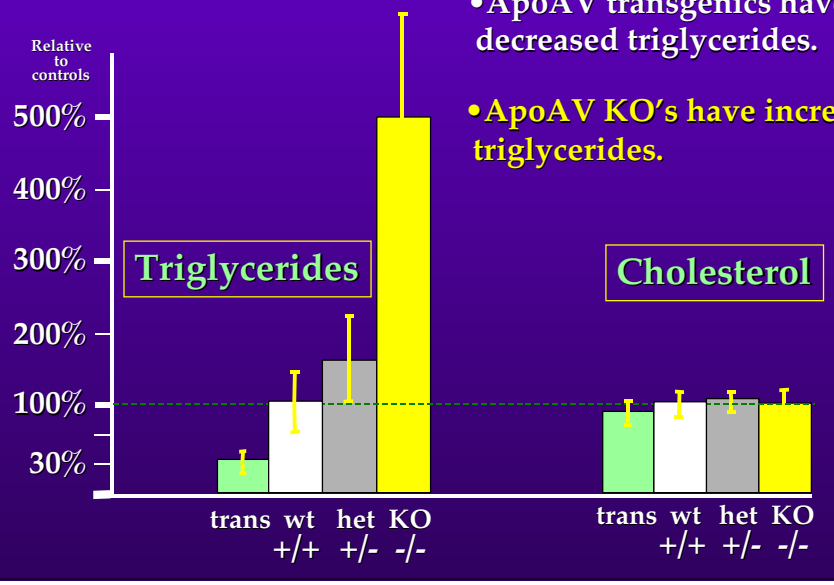
Generation of ApoAV knockout mice





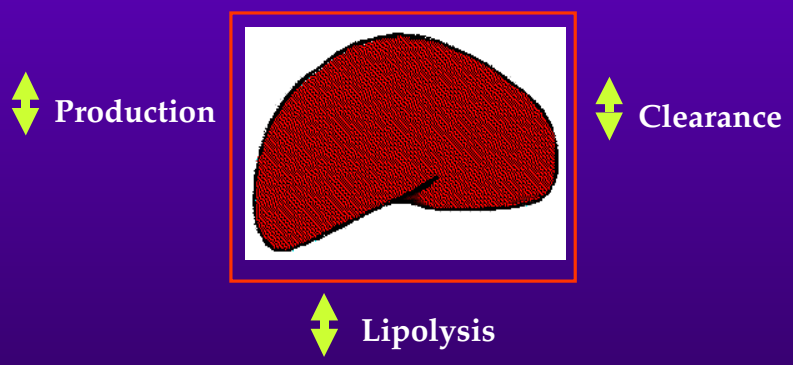
Homozygous deletion mice lack apoAV transcript (liver).

ApoAV Transgenic and Knockout Plasma Levels



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

Mechanisms for Altered Plasma Triglycerides



Summary I: ApoAV

- A new apolipoprotein belonging to the ApoAI/CIII/AIV gene cluster.
- Expressed in the liver & associates with HDL/VLDL.
- An important modulator of triglycerides (TG) in mice.

↑ ApoAV ↓ TG ↓ ApoAV ↑ TG

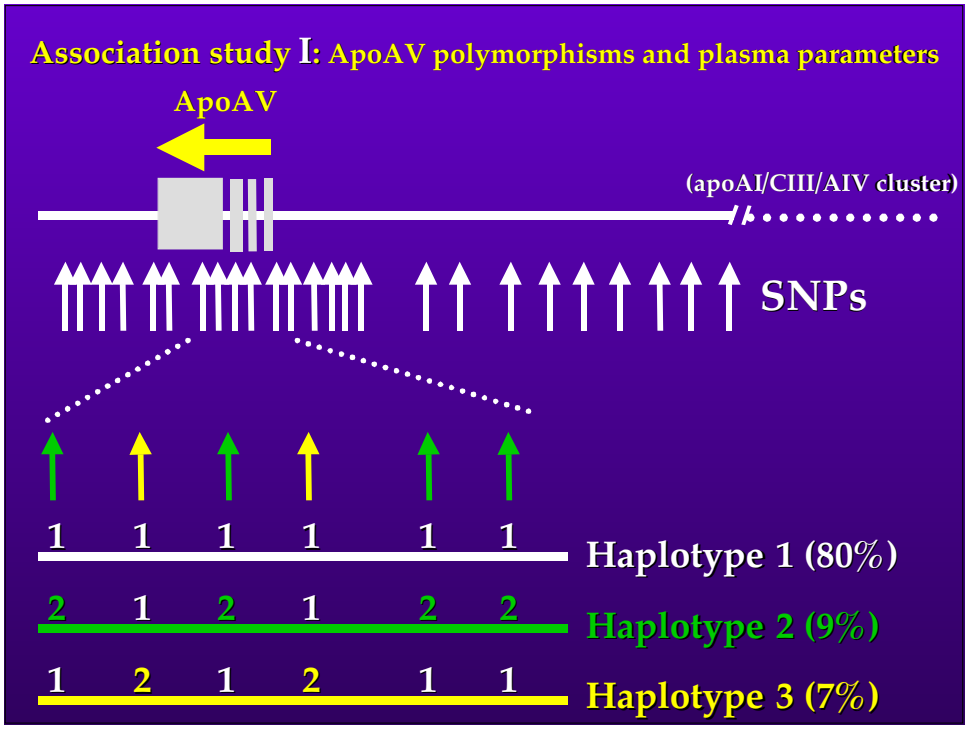
Is ApoAV involved in human biology/disease?

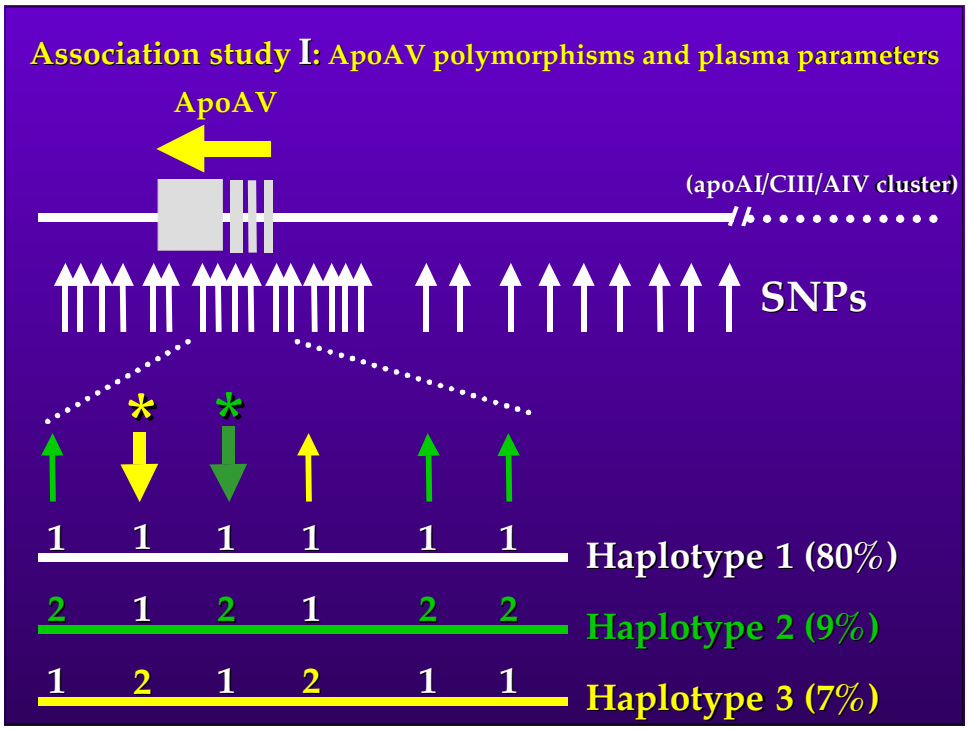
Association study I: ApoAV polymorphisms and plasma parameters

Berkeley Population

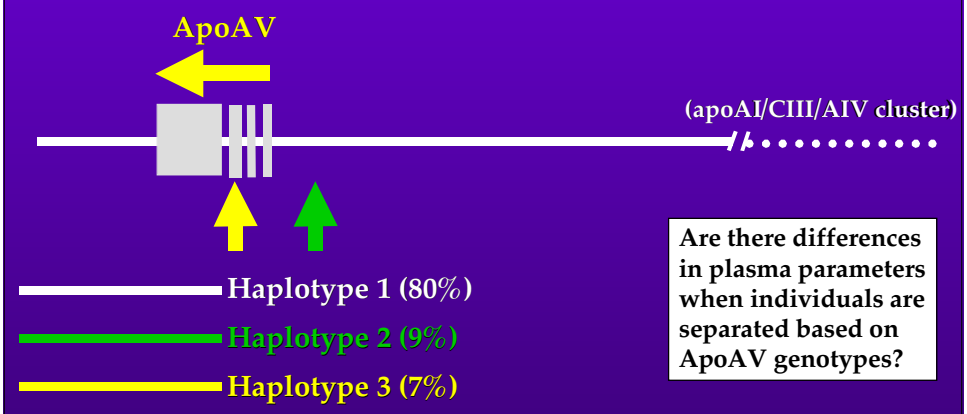
500 normal individuals phenotyped for plasma:

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


Association study I: ApoAV polymorphisms and plasma parameters
ApoAV
(apoAI/CIII/AIV cluster)
SNPs
1 1 1 1 1 1
Haplotype 1 (80%)
2 1 2 1 2 2
Haplotype 2 (9%)
1 2 1 2 1 1
Haplotype 3 (7%)


Association study I: ApoAV polymorphisms and plasma parameters
ApoAV
(apoAI/CIII/AIV cluster)
SNPs
*
*
1 1 1 1 1 1
Haplotype 1 (80%)
2 1 2 1 2 2
Haplotype 2 (9%)
1 2 1 2 1 1
Haplotype 3 (7%)

Association study I: ApoAV polymorphisms and plasma parameters

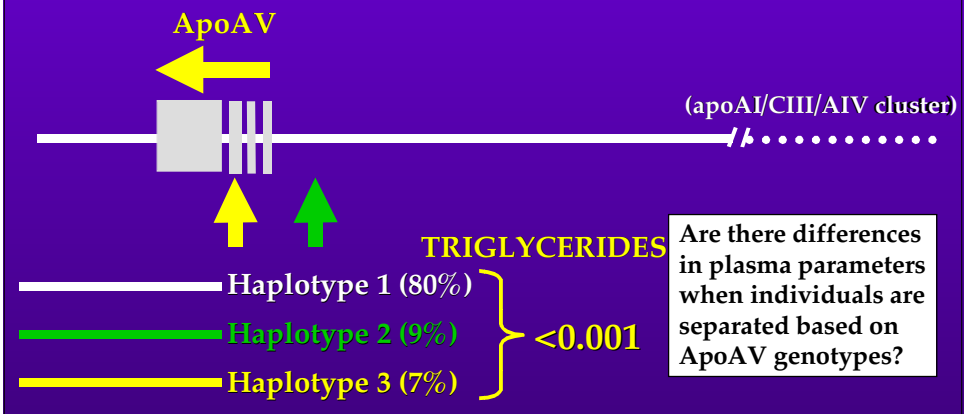


Are there differences in plasma parameters when individuals are separated based on ApoAV genotypes?

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

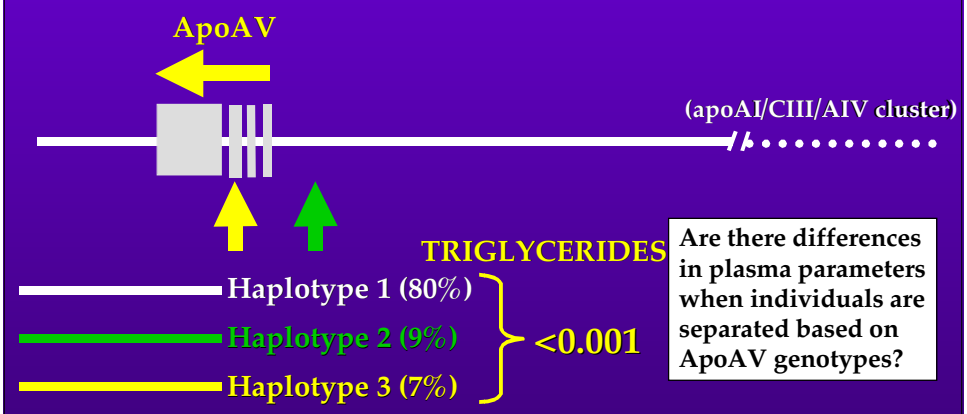


Are there differences in plasma parameters when individuals are separated based on ApoAV genotypes?

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



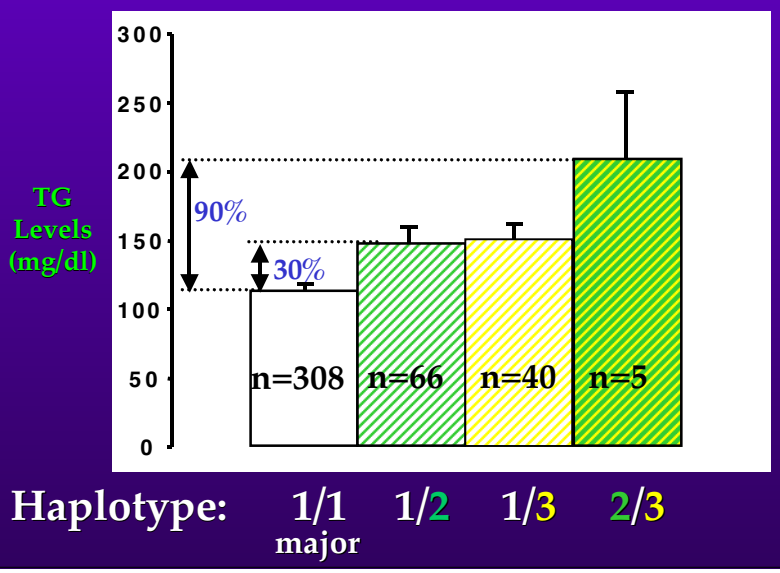
Are there differences in plasma parameters when individuals are separated based on ApoAV genotypes?

Genotyped 500 normal individuals phenotyped for plasma:

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

What is the amount of this difference in triglyceride levels?

Association between ApoAV and Triglyceride Levels



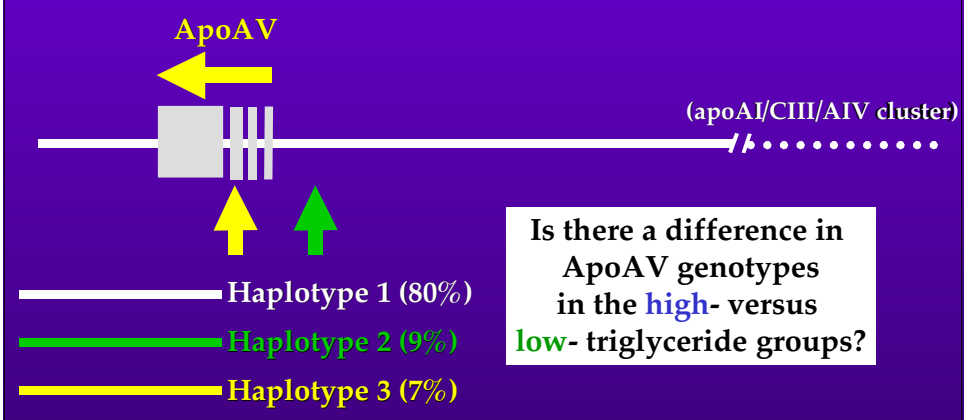
Summary II

- Two independent minor haplotypes in human ApoAV are associated with increased plasma triglycerides.
  - The minor haploype frequencies are ~10% in Caucasians.
  - Individuals with one copy of either of these minor haplotypes have ~30% more plasma triglycerides.
    - Compound heterozygotes have ~90% more triglycerides.

Association Studies

Is this finding reproducible????

Association study II: ApoAV polymorphisms and plasma parameters

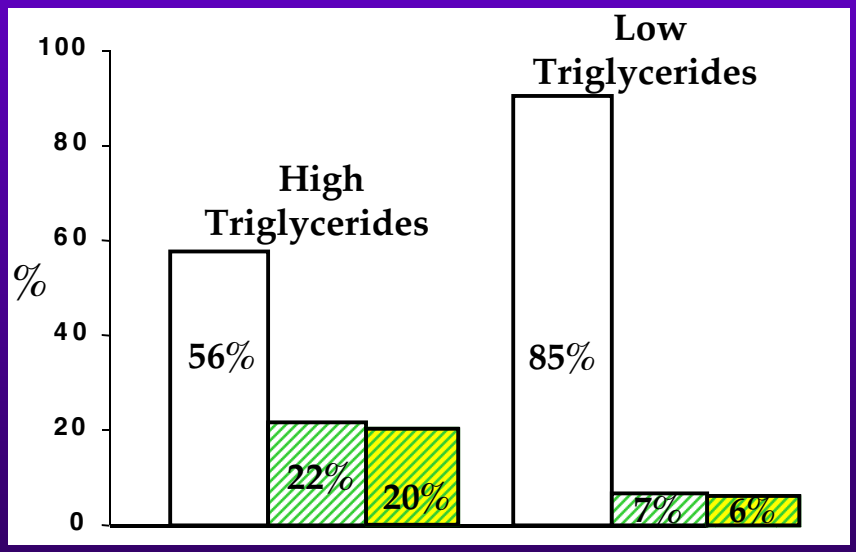


Is there a difference in ApoAV genotypes in the high- versus low- triglyceride groups?

Genotyped 460 individuals stratified based on triglyceride levels.

High Triglycerides: Avg 340 mg/dl
Low Triglycerides: Avg 50 mg/dl

Association study II: ApoAV polymorphisms and plasma parameters

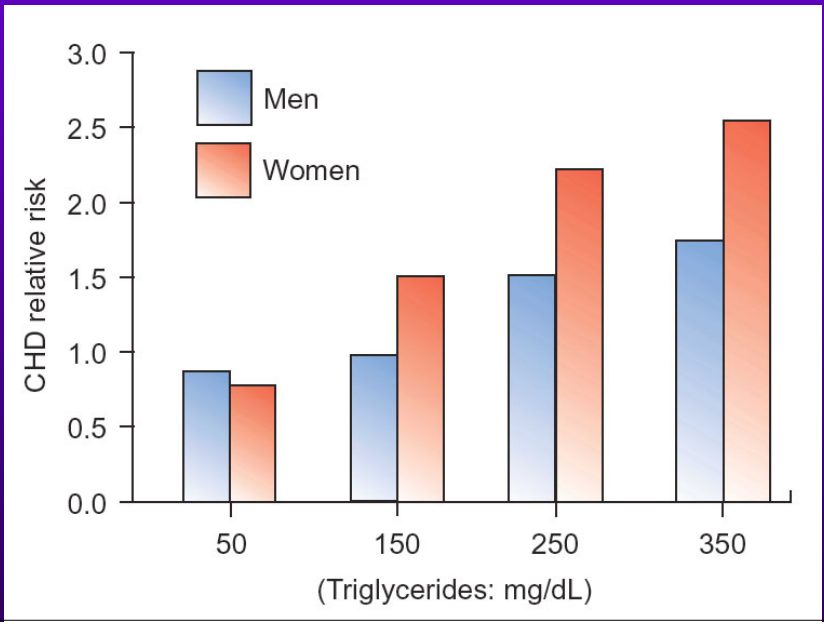


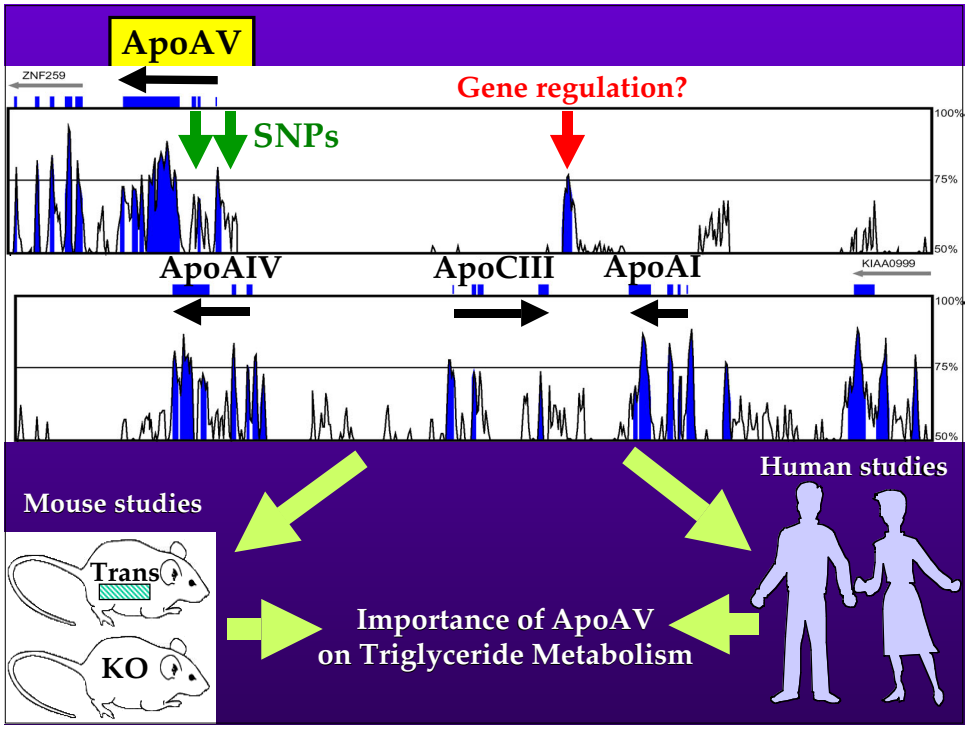
ApoA5 and Triglyceride Levels

An example of common human variation contributing to a quantitative phenotype

Ethnicity:	Carriers of Minor Haplotype 2 and/or 3:
Caucasian	24%
African American	36%
Hispanic	51%

↑ Triglycerides ↑ Heart Disease





Acknowledgements

LBNL
Edward Rubin
Nadine Barouhk
Elaine Gong
Jennifer Akiyama
Kathryn Houston
Keith Lewis
Willow Dean
Jan-Fang Cheng
Inna Dubchak
Lior Pachter
Jody Schwartz
Veena Afzal
Xinli Yang
Ronald Krauss
Patricia Blanche
Laura Holl
Joseph Orr

UT-SW
Jonathan Cohen
Helen Hobbs
Jaroslav Hubacek

Rayne Institute
Philippa Talmud
Steve Humphries

Pasteur Institute-Lille
Jamila Fruchart
Jean-Charles Fruchart

MCW
Michael Olivier

NIH/NHLBI
http://pga.lbl.gov

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