Experimental Data Subgroup

Len Pennacchio
Jay Shendure
John Stamatoyannopoulos
Wendy Winckler

Experimental Data Subgroup

- Goal: evaluate methods by which investigators can query whether candidate variants have a biological effect

Motivation for Functional Analysis

- **GWAS** peak → causal variant(s)
- Clinical genetics → Functional consequences of **variants of unknown significance**
- "Functionalizing" poorly characterized genes of interest, *i.e.* **developing functional assays**

Need generic, accessible, high-throughput **methods** and **resources** to facilitate the functional analysis of both **coding** and **regulatory** variation

Challenges

- Spectrum of experimental methods exist
- How to select the most appropriate method?
  - Type of variant
  - Context
  - Access to samples, reagents, and techniques
  - Throughput, time, and cost
  - Weight of evidence
- Caution required in interpretation—for both negative and positive results

Selected Experimental Methods

Techniques	Variant class	Experiment type	High-throughput?	Evidence
mouse or zebrafish knockin	Any	in vivo	no	strong
genome editing	Any	in vivo, in vitro	yes	strong
mouse or zebrafish knockout	LoF allele	in vivo	no	strong
cell culture shRNA knockdown	LoF allele	in vitro	yes	suggestive
cDNA complementation	LoF allele	in vitro	no	suggestive
splicing assay	Splicing	ex vivo, in vitro, in vivo	yes	strong
protein-specific biochemical or cellular assays	Protein-altering alleles	in vitro	no	suggestive
correlation with expression	Regulatory	ex vivo, in vitro, in vivo	yes	suggestive
reporter construct	Regulatory	in vitro	yes	suggestive

Regulatory variants

Functional regulatory variation: Levels of evidence framework

Level 1: *in vivo* evidence from *in situ* models

1a *In situ* / whole locus model of strongly genetically implicated variant that precisely recapitulates the phenotype at the organismal level

1b *In situ* genome modification *(e.g., genome editing / knock-in/out)*

1c Whole-locus transgenic lines *(e.g., YAC, BAC; single copy)*

1d *In situ* measurement of gain/loss of regulatory protein binding directly coupled to *in vivo* gene product phenotype

1e *In situ* gain/loss of regulatory protein not coupled to gene product

Level 2: Evidence from artificial/condensed construct models

2a Standard transgenic animal

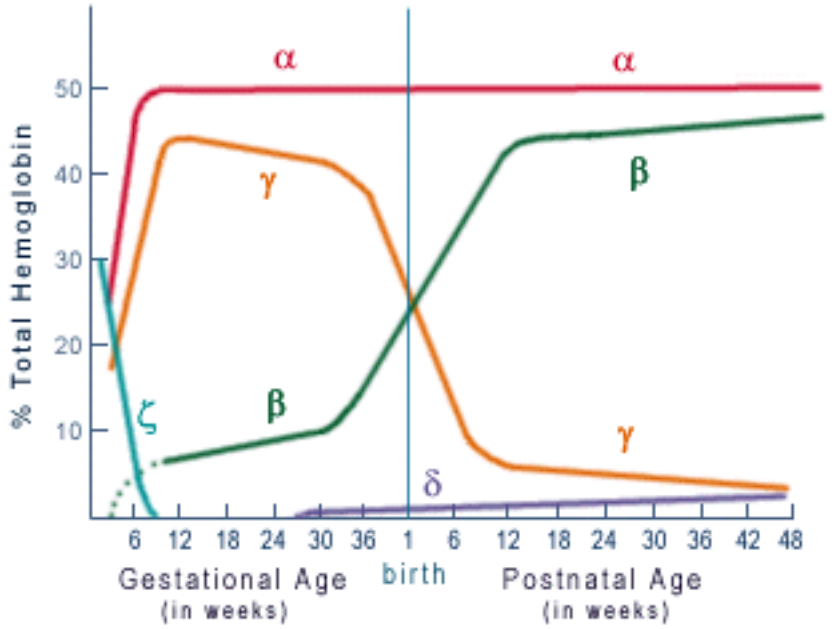
2b Stable transfection (integrated into genome)

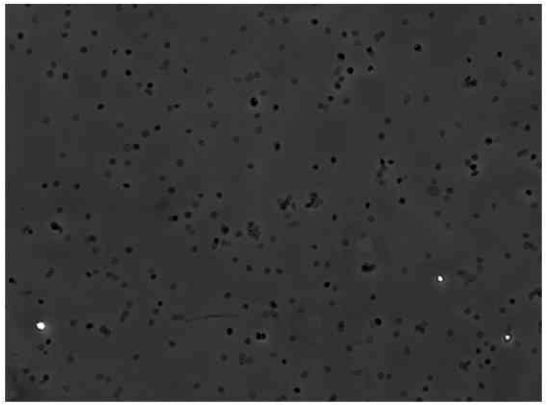
2c Transient transfection (ex-genomic)

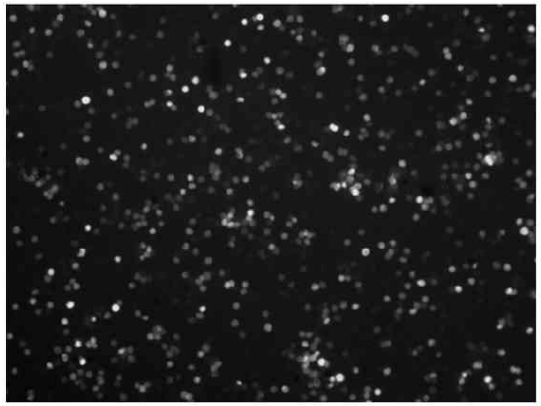
Level 3: Non-cellular assays (e.g., gel shifts)

Example: Level 1a

Hereditary Persistence of Fetal Hemoglobin (HPFH)







Scientific trajectory of variant characterization

1985

1992

1995

Scientific trajectory of variant characterization

1985

1992

1995

30. Shafit-Azgardo, B., Maio, J. & Brown, F. *Nucleic Acids Res.* **10**, 3175-3193 (1982).
31. Hohn, B. & Collins, J. *Gene* **11**, 291-298 (1980).
32. Ish-Horowicz, D. & Burke, J. *Nucleic Acids Res.* **9**, 2989-2998 (1981).
33. Grosveld, F. G., Dahl, H., deBoer, E. & Flavell, R. *Gene* **13**, 227-237 (1981).
34. Barsh, G., Seeburg, P. & Gelinas, R. *Nucleic Acids Res.* **11**, 3939-3958 (1983).
35. Norrander, J., Kempe, T. & Messing, J. *Gene* **26**, 101-106 (1983).
36. Henikoff, S. *Gene* **28**, 351-359 (1984).
37. Sanger, F., Nicklen, S. & Coulson, A. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).

A point mutation in the $^{A}\gamma$-globin gene promoter in Greek hereditary persistence of fetal haemoglobin

Francis S. Collins*, James E. Metherall, Minoru Yamakawa, Julian Pan, Sherman M. Weissman & Bernard G. Forget†

Departments of Human Genetics, Internal Medicine, and Molecular Biophysics and Biochemistry, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA

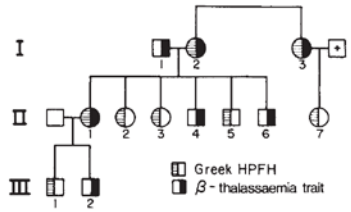
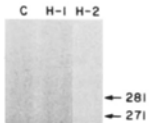


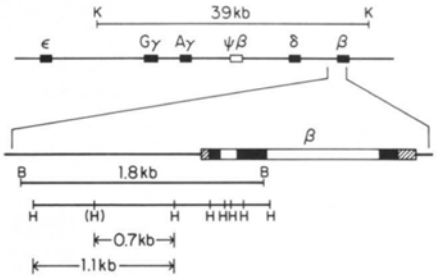
Fig. 1 Pedigree of the Greek HPFH family, originally described in ref. 4. Individual I-3, who is doubly heterozygous for Greek HPFH and β thalassaemia, was the source of DNA for cosmid cloning.

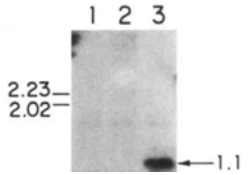


11. Bunch, C., Wood, W. G., Weatherall, D. J., Robinson, J. S. & Corp, M. J. *Br. J. Haemat.* **49**, 325-336 (1981).
12. Wood, W. G. & Bunch, C. in *Globin Gene Expression and Hematopoietic Differentiation* (eds Stamatoyannopoulos, G. & Nienhuis, A.) 511-521 (Liss, New York, 1983).
13. Weatherall, D. J., Clegg, J. B. & Wood, W. G. *Lancet* ii, 660-663 (1976).
14. Fraser, I. D. & Raper, A. B. *Archs Dis. Childh.* **37**, 289-296 (1962).
15. Dan, M. & Hagiwara, A. *Expl Cell Res.* **46**, 596-598 (1967).
16. Kamada, M. *Yokohama med. Bull.* **20**, 127-135 (1969).
17. Kidoguchi, K., Ogawa, M., Karam, J. D., McNeil, J. S. & Fitch, M. S. *Blood* **53**, 519-522 (1979).
18. Comi, P. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 362-365 (1980).
19. Wood, W. G. in *Biochemical Development of the Fetus and Neonate* (ed. Jones, C. T.) 127-162 (Elsevier, Amsterdam, 1982).
20. Keating, A. *et al. Nature* **298**, 280-283 (1982).
21. Potter, C. G., Rowell, A. C. & Weatherall, D. J. *Clin. Lab. Haemat.* **3**, 245-255 (1981).
22. Widmer, H. J., Hosbach, H. A. & Weber, R. *Devl Biol.* **99**, 50-60 (1983).
23. Kliehauer, E., Braun, H. & Betke, K. *Nature* **35**, 635-636 (1957).
24. Wood, W. G. *et al. Nature* **264**, 799-801 (1976).

G to A substitution in the distal CCAAT box of the $^{A}\gamma$-globin gene in Greek hereditary persistence of fetal haemoglobin

Richard Gelinas*, Brian Endlich*, Carla Pfeiffer*, Mayumi Yagi† & George Stamatoyannopoulos†





Scientific trajectory of variant characterization

1985

1992

1995

Received 27 May; accepted 26 June 1992.

1. Guyader, M. *et al. Nature* **326**, 662–669 (1987).
2. Zagury, J. F. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 5941–5945 (1988).
3. Franchini, G. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 2433–2437 (1989).
4. Kumar, P. *et al. J. Virol.* **64**, 890–901 (1990).
5. Hasegawa, A. *et al. AIDS Res. hum. Retrovir.* **5**, 593–604 (1989).
6. Kirchhoff, F., Jentsch, K. D., Stuke, A., Mous, J. & Hunsmann, G. *AIDS* **4**, 847–857 (1990).
7. Dietrich, U. *et al. Nature* **342**, 948–950 (1989).

Champaign, Illinois, 1991).

ACKNOWLEDGEMENTS. This paper is dedicated to the memory of B.M.G. We thank G. Myers and K. MacInnes for assistance with phylogenetic analyses; the Irish National Centre for Bioinformatics for their facilities; J. Hoxie for independent attempts at cultivating blood samples from subject 2238; Serologicals, Inc. (Atlanta, GA) for blood specimens from subject 7312A; R. Desrosiers, P. Fultz and D. Ho for discussion; D. Decker and M. Mixon for technical assistance; and C. Davis and A. J. Nicholson for manuscript preparation. This work was supported by grants from the NIH, the US Army Medical Research Acquisition Activity, the Life and Health Insurance Medical Research Fund, and the Birmingham Center for AIDS Research. G.M.S. is a PEW Scholar in the Biomedical Sciences.

A single point mutation is the cause of the Greek form of hereditary persistence of fetal haemoglobin

Meera Berry, Frank Grosveld & Niall Dillon

Laboratory of Gene Structure and Expression, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

IN normal humans the fetal stage-specific γ-globin genes are silenced after birth and not expressed in the adult. Exceptions are seen in cases of hereditary persistence of fetal haemoglobin (HPFH). These are clinically important because the elevated levels

to establish a large number of bred lines. When the wild-type γβ minilocus was introduced into fertilized mouse eggs, five transgenic mice were obtained. Southern blots showed that two of the founders were mosaic (31 and 36) and that all contained the intact γβ minilocus, albeit at different copy numbers (Table 1, and data not shown). S1 nuclease protection analysis showed that the γ-globin gene expression was suppressed in adult mice (Fig. 1*a*, *b*). In contrast, the human β-globin gene was expressed at this stage at levels comparable to those observed for the mouse β-maj-globin genes[5] (Fig. 1*b*; Table 1). The suppression of the wild-type γ-globin gene is in agreement with results obtained when a minilocus containing only the γ-globin gene is introduced into mice[4]. Repeated phlebotomy increases the number of reticulocytes, but even under those conditions the γ-globin gene remains suppressed (Fig. 1*b*). When the −117 mutant γβ minilocus was introduced into mice, nine transgenic mice were obtained and Southern blots showed that they con-

Scientific trajectory of variant characterization

1985

1992

1995

Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 5655–5659, June 1995
Developmental Biology

Use of yeast artificial chromosomes (YACs) in studies of mammalian development: Production of β-globin locus YAC mice carrying human globin developmental mutants

(developmental regulation/transgenic mice/hereditary persistence of fetal hemoglobin/δβ-thalassemia)

KENNETH R. PETERSON*†, QI LIANG LI*, CHRISTOPHER H. CLEGG*‡, TATSUO FURUKAWA*, PATRICK A. NAVAS*, ELIZABETH J. NORTON*, TYLER G. KIMBROUGH*, AND GEORGE STAMATOYANNOPOULOS*§

*Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA 98195; ‡Bristol-Myers Squibb Pharmaceutical Research Institute, Seattle, WA 98121; and §Department of Genetics, University of Washington, Seattle, WA 98195

Communicated by Stanley M. Gartler, University of Washington, Seattle, WA, March 8, 1995

ABSTRACT **To test whether yeast artificial chromosomes (YACs) can be used in the investigation of mammalian development, we analyzed the phenotypes of transgenic mice carrying two types of β-globin locus YAC developmental mutants: (*i*) mice carrying a G → A transition at position −117 of the $^{A}\gamma$ gene, which is responsible for the Greek $^{A}\gamma$ form of**

developmental regulation of gene expression in transgenic mice (1). Our data show that the genes of the β-globin locus YAC (β-YAC) are correctly regulated during development in the mouse (1), thus demonstrating the usefulness of the YAC/transgenic mouse system.

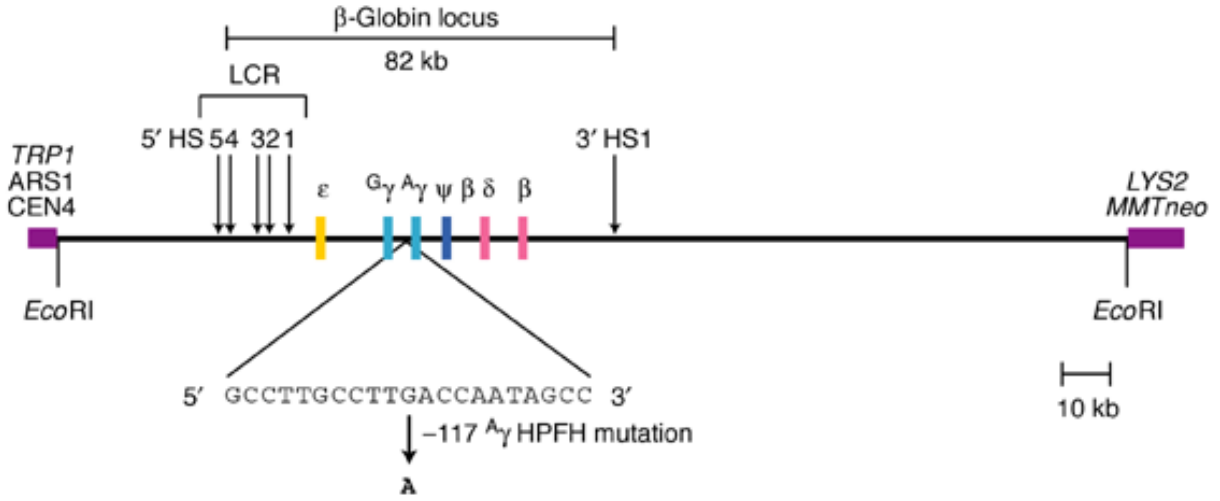
In this work, we test whether YACs can be used for the

Scientific trajectory of variant characterization

1985

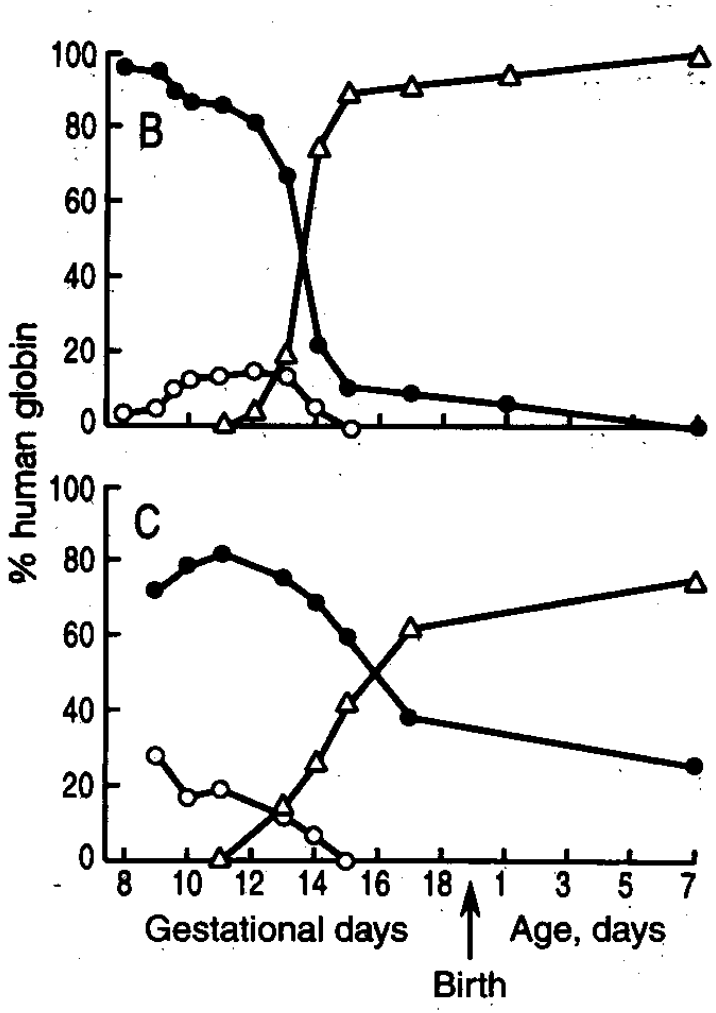
1992

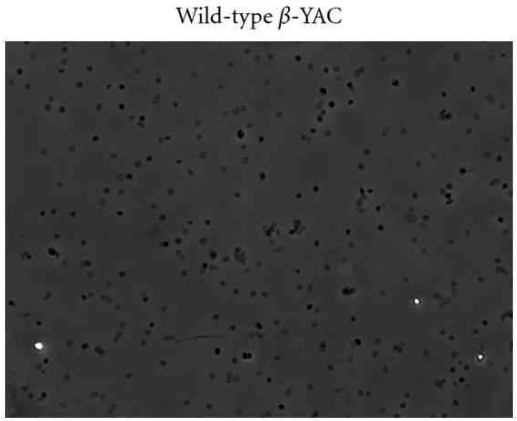
1995

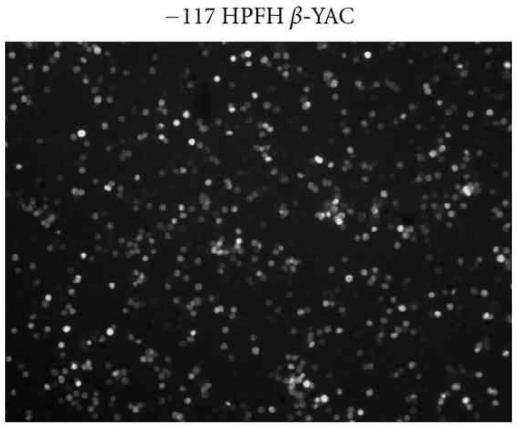


Structure of a human β-globin locus YAC

A single point mutation in a 273kb single copy YAC, functionally profiled across development







Example: Level 1d

Alpha thalassemia

A Regulatory SNP Causes a Human Genetic Disease by Creating a New Transcriptional Promoter

Marco De Gobbi,[1*] Vip Viprakasit,[2*] Jim R. Hughes,[1] Chris Fisher,[1] Veronica J. Buckle,[1] Helena Ayyub,[1] Richard J. Gibbons,[1] Douglas Vernimmen,[1] Yuko Yoshinaga,[3] Pieter de Jong,[3] Jan-Fang Cheng,[4] Edward M. Rubin,[4] William G. Wood,[1] Don Bowden,[5] Douglas R. Higgs[1]‡

We describe a pathogenetic mechanism underlying a variant form of the inherited blood disorder α thalassemia. Association studies of affected individuals from Melanesia localized the disease trait to the telomeric region of human chromosome 16, which includes the α-globin gene cluster, but no molecular defects were detected by conventional approaches. After resequencing and using a combination of chromatin immunoprecipitation and expression analysis on a tiled oligonucleotide array, we identified a gain-of-function regulatory single-nucleotide polymorphism (rSNP) in a nongenic region between the α-globin genes and their upstream regulatory elements. The rSNP creates a new promoterlike element that interferes with normal activation of all downstream α-like globin genes. Thus, our work illustrates a strategy for distinguishing between neutral and functionally important rSNPs, and it also identifies a pathogenetic mechanism that could potentially underlie other genetic diseases.

The human α-globin cluster, located at the telomeric region of chromosome 16 (16p13.3), includes an embryonic gene (ζ), two minor α-like genes [α^D (also called μ) and θ], two α genes (α2 and α1), and two pseudogenes (ψα1 and ψζ) (*1*, *2*). previously described molecular defects could be found. The pattern of inheritance suggested that individuals with HbH disease are homozygotes for a codominant defect, referred to here as $(\alpha\alpha)^T$, causing α thalassemia with a predicted genotype of $(\alpha\alpha)^T/(\alpha\alpha)^T$ (table S1). linkage to a variable number of tandem repeats (VNTR) (*6*) located ~8.5 kb from the α-globin genes (Fig. 1), we found that all individuals with the $(\alpha\alpha)^T$ mutation shared a common VNTR allele (fig. S1), demonstrating that this is a cis-linked defect. Further association studies, using known SNPs, showed that the $(\alpha\alpha)^T$ haplotype extends from the 16p telomere, with loss of association immediately downstream of the α-globin cluster (coordinate 168,467 in Fig. 1) defining the centromeric border of the region containing the cis-acting mutation. We estimated that the frequency of the $(\alpha\alpha)^T$ defect in the island population is ~0.04 (fig. S1).

We therefore resequenced the $(\alpha\alpha)^T$ haplotype by isolating bacterial artificial chromosomes (BACs) from a library constructed from the peripheral blood DNA of patient L with the Melanesian type of HbH disease [$(\alpha\alpha)^T/(\alpha\alpha)^T$]. BACs spanning the α-globin cluster and the surrounding ~213 kb of DNA (coordinates 21,059 to 234,236) were sequenced (DQ431198), and we identified 283 SNPs and/or sequence differences (Fig. 1) by comparison with the current wild-type sequence (National Center for Biotechnology Information database build 35, coordinates 1 to 223478), consistent with estimates of the frequency of SNPs throughout the genome (*7*). This now presented a sit-

Example 2:
Mouse Knock-ins
Mouse Site-Specific Integration

KNOCK-IN MOUSE STUDIES: Introduce Human Mutation into Mouse Gene

Examples

CAG Repeat Expansions Introduced into:

1) Huntingtin gene- Short Repeat: Nuclear Inclusion Body Formation in Striatal Neurons

Wheeler et al. (2000) Long glutamine tracts cause nuclear localization of a novel form of huntingtin in medium spiny striatal neurons in HdhQ92 and HdhQ111 knock-in mice. Hum. Mol. Genet. 9, 503–513.

2) Huntingtin gene- Long Repeat: Neurological Abnormalities

Lin et al. (2001) Neurological abnormalities in a knock-in mouse model of Huntington's disease. Hum. Mol. Genet. 10, 137–144.

3) Spinocerebellar Ataxia Type 1 Gene- Motor Coordination Defects

Lorenzetti D., Watase K., Xu B., et al. (2000) Repeat instability and motor incoordination in mice with a targeted expanded CAG repeat in the Sca1 locus. Hum. Mol. Genet. 9, 779–785.

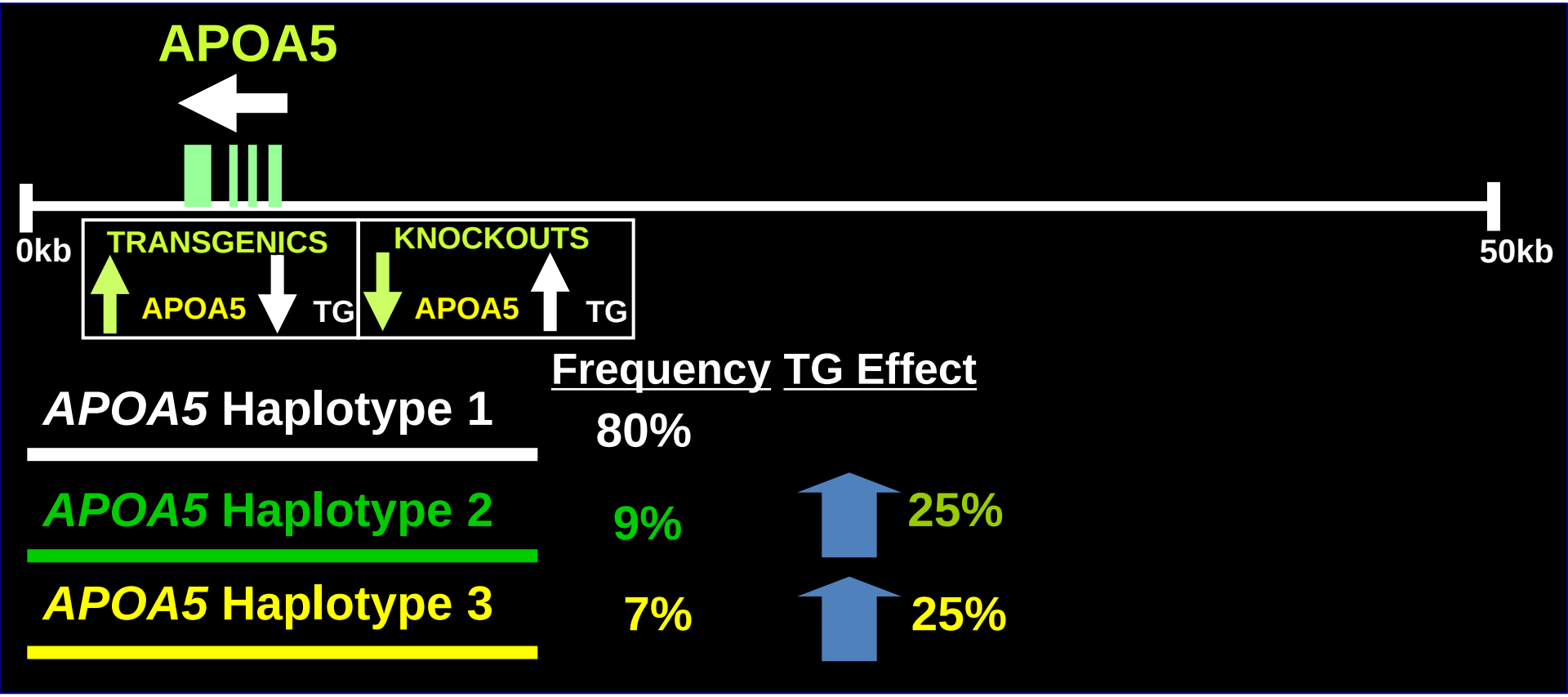
Point Mutation Introduced into:

1) Presenilin-1 Gene- Single Amino Acid Change Causes Hippocampus Neuron Sensitivities

Guo Q., Fu W., Sopher B. L., et al. (1999) Increased vulnerability of hippocampal neurons to excitotoxic necrosis in presenilin-1 mutant knock-in mice. Nat. Med. 5, 101–106.

Mouse Site-Specific Integration:

Reproducible Association between Human *APOA5* Common Variation and Plasma Triglyceride Levels

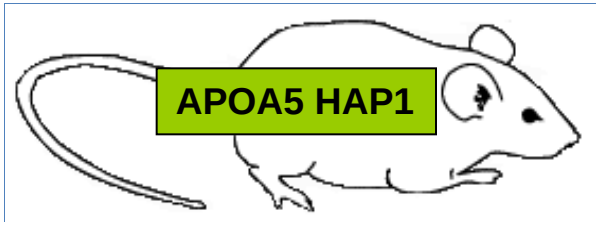


Do these Haplotypes Affect *APOA5* Gene Product Levels *In Vivo*?

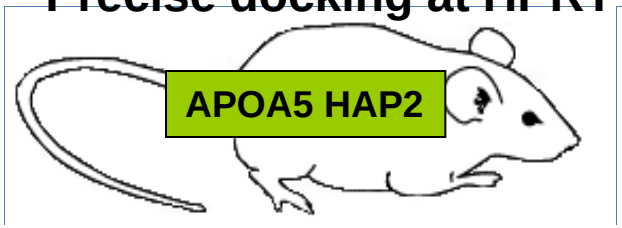
Pennacchio et al, 2001 Science

Generation of Site-Specific Single-Integrant Haplotype Transgenes

Precise docking at HPRT

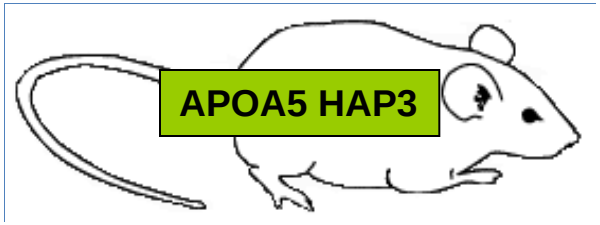


Common



Noncoding, 5UTR (Kozak)

(7 change)

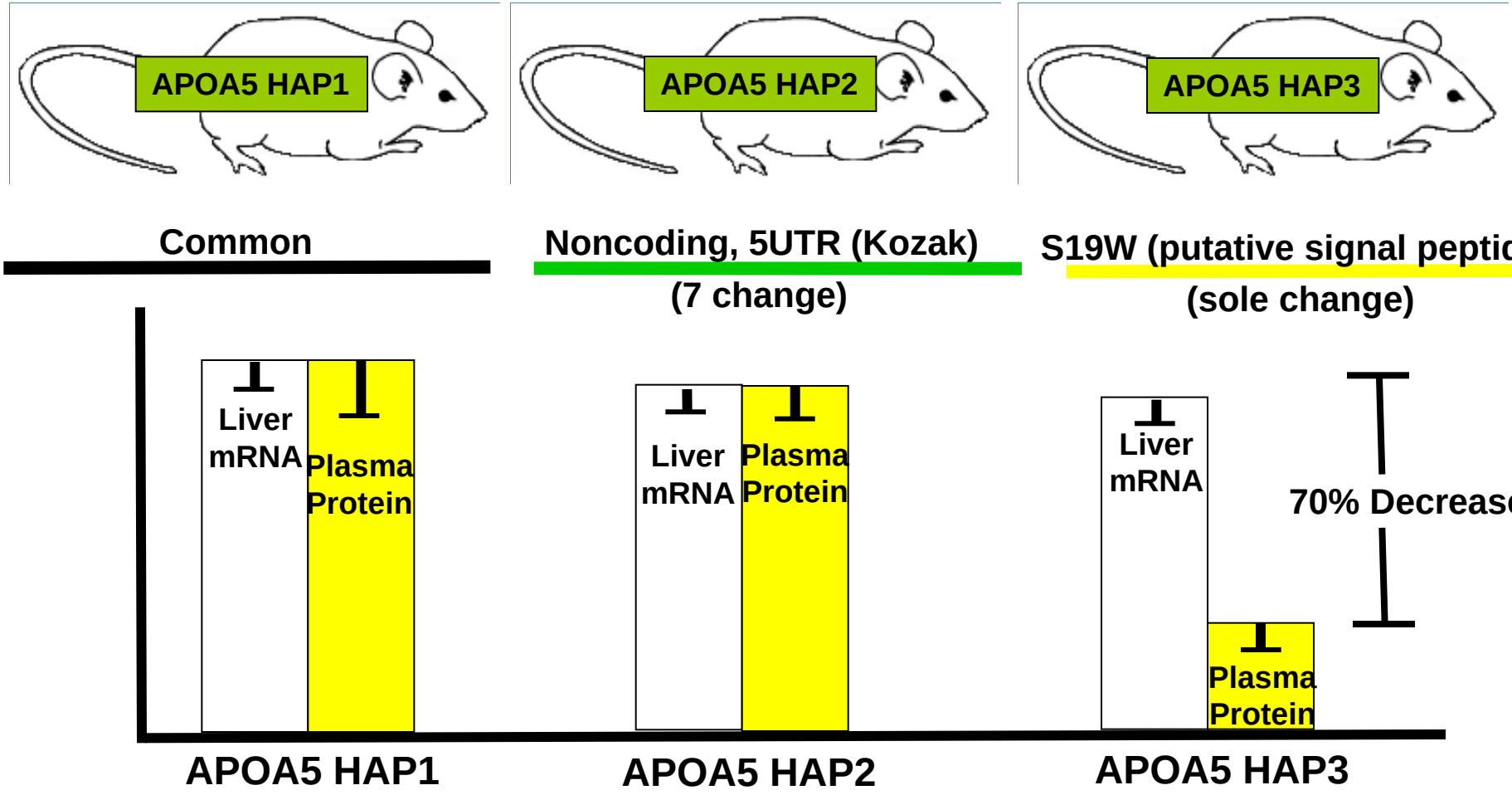


S19W (putative signal peptide)

(sole change)

Compare APOA5:
mRNA Levels in Liver
Protein Levels in Plasma

Generation of Site-Specific Single-Integrant Haplotype Transgenes



S19W is likely responsible for TG Association

Ahituv et al. 2007. In Vivo Characterization of Human APOA5 Haplotypes. Genomics, 90(6):674-9.

Example 3:

9p21 and Coronary Artery Disease

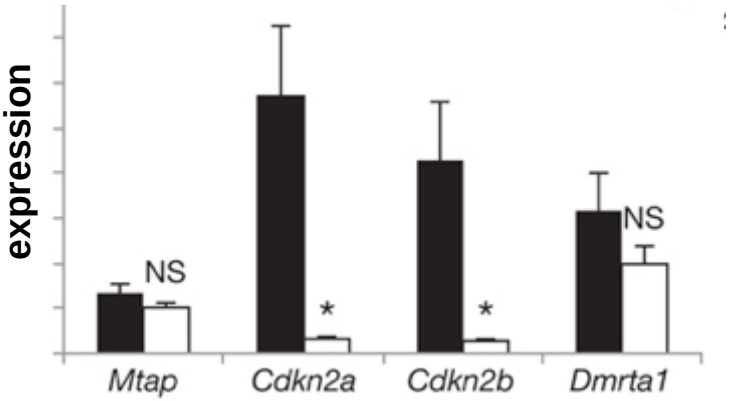
9p21 common risk variant:

increases CAD risk by 30%

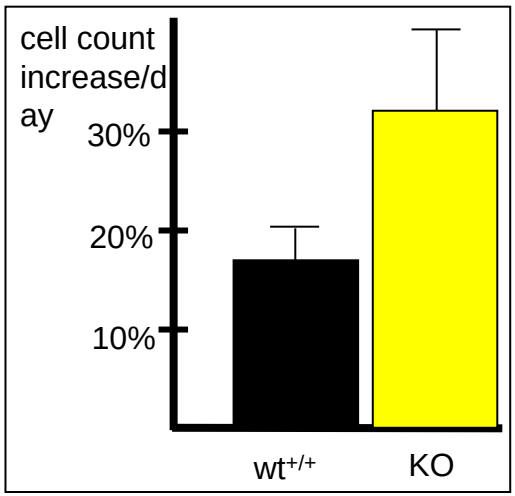
>20% of population homozygous

(2007, J Cohen and K Stefansson Labs)

70kb non-coding region knocked out in mice



Dysregulation of Cdkn Genes in the Heart

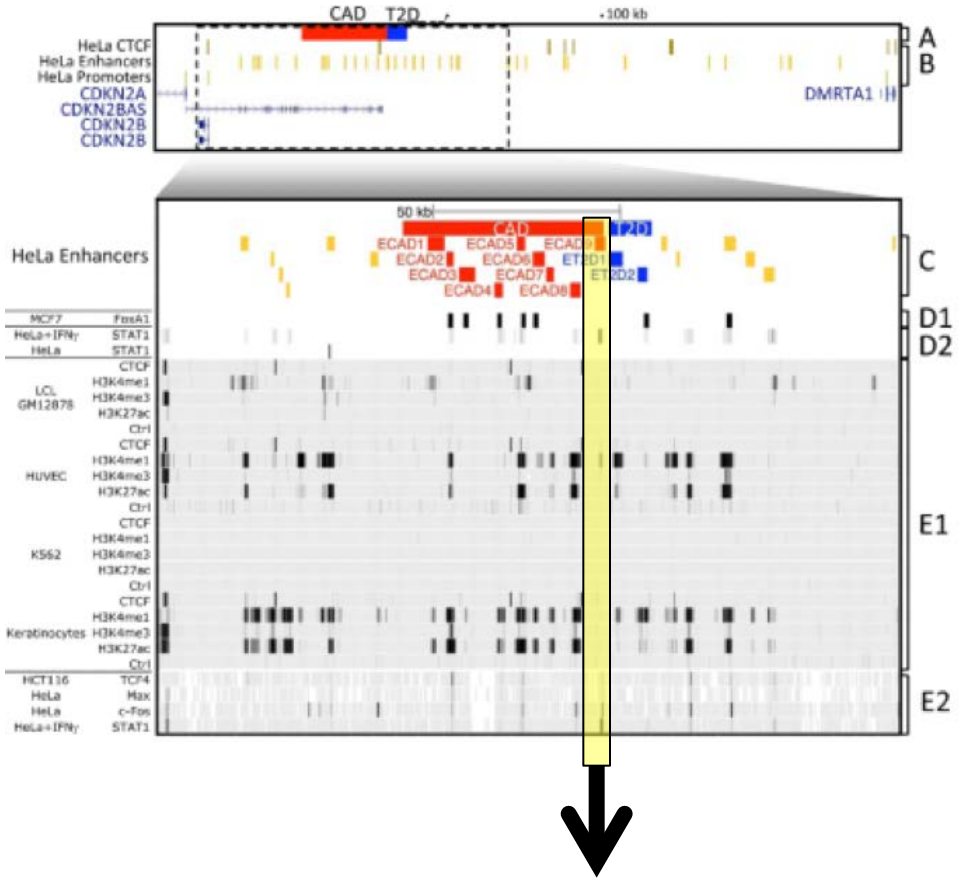


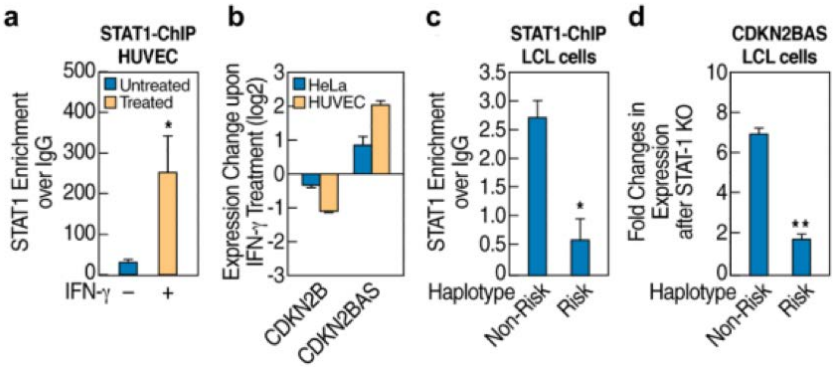
Increased Aortic Smooth Muscle Cell proliferation

Supports CAD risk interval harbors distant-acting regulatory function(s)

Visel et al. *Nature* 2010

In Vitro Studies: Empowered by ENCODE





This Element:
1) binds STAT1 *in vitro*
2) binding lowers CDNK2B RNA *in vitro*
3) STAT1 occupancy is less in Risk CAD LCL

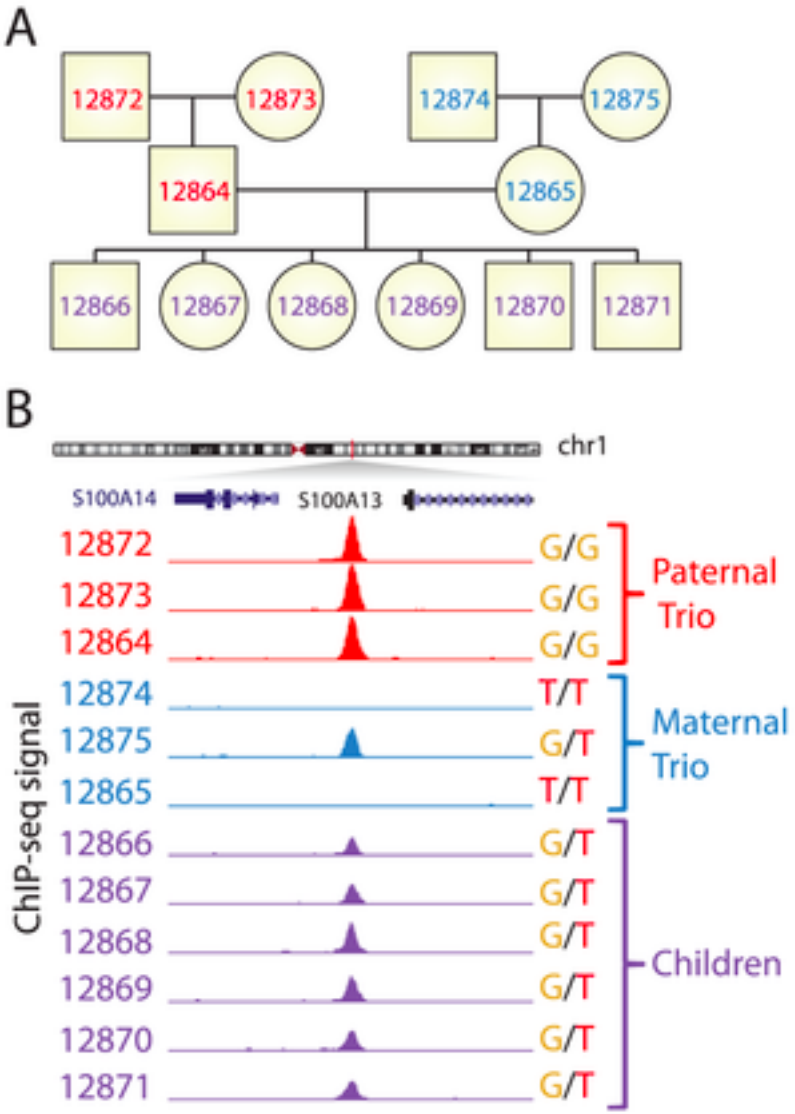
One Third of Variants found in this Element
Several Effect Putative STAT1 Binding Site

9p21 DNA variants associated with Coronary Artery Disease impair IFNγ signaling response
Kelly Frazer et al. Nature 2011

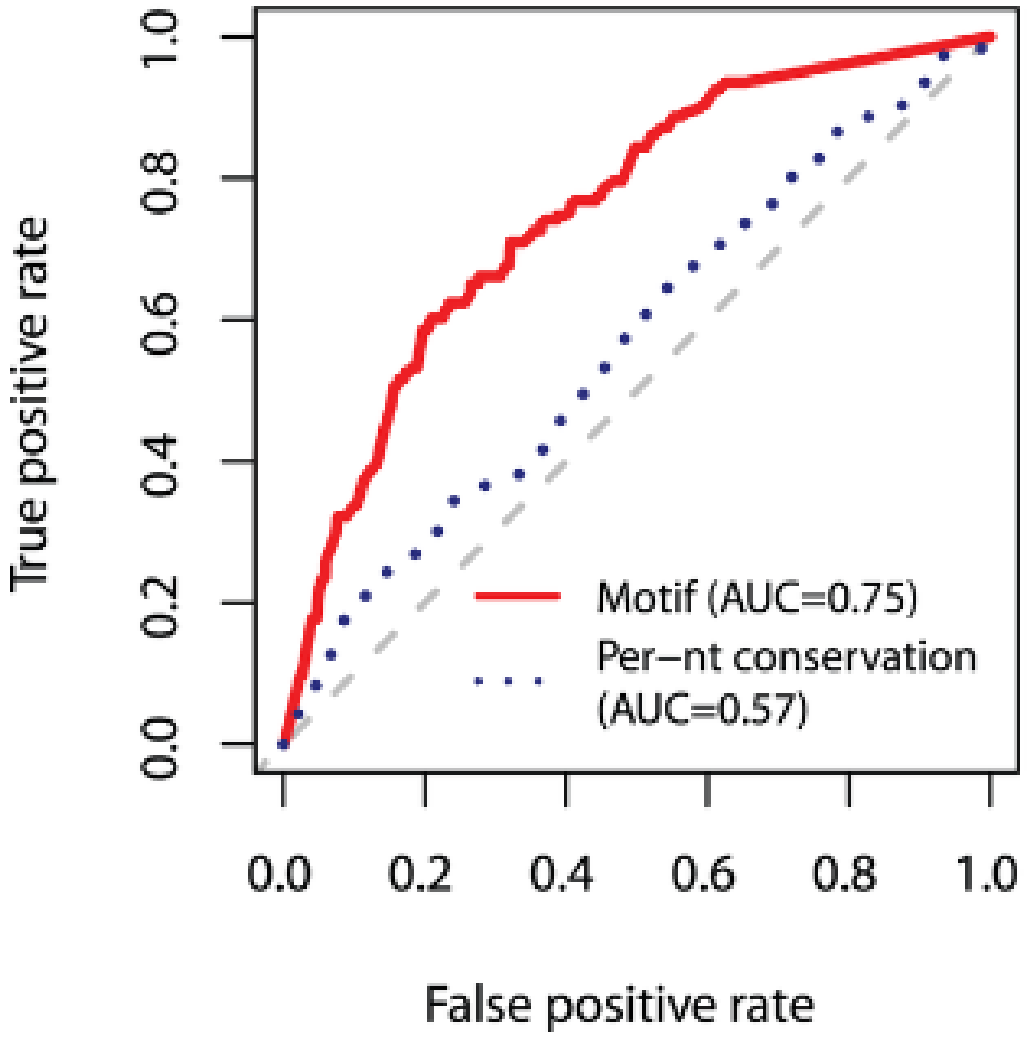
Future of Experimental Data

- **1000+ GWAS peaks → causal variant(s)**
- **Clinical genetics → Functional consequences of variants of unknown significance**
- **Facilitating genetics → biology**
- High-throughput or massively parallel methods for assessing the functional consequences of **observed** and **potential** variation

Functional assessment of **observed** regulatory variation *in situ allelic occupancy*

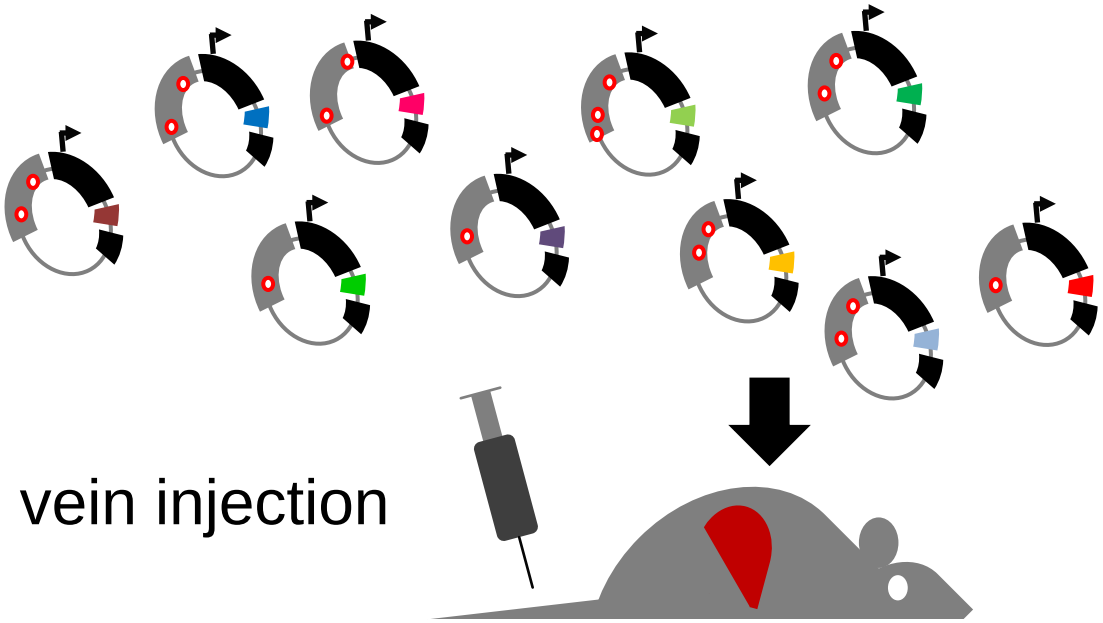


Conservation is an imperfect guide to regulatory function



ENCODE

Massively parallel functional assessment of **potential** regulatory variants



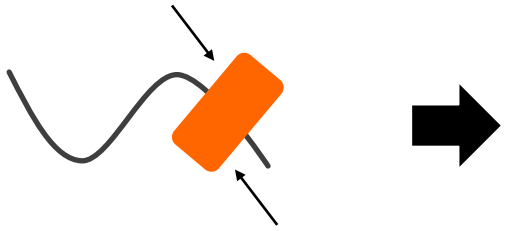
Complex library of tagged enhancer haplotypes

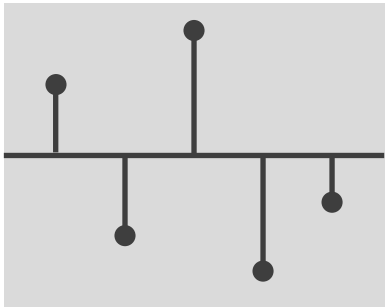
Tail vein injection

RNA from liver

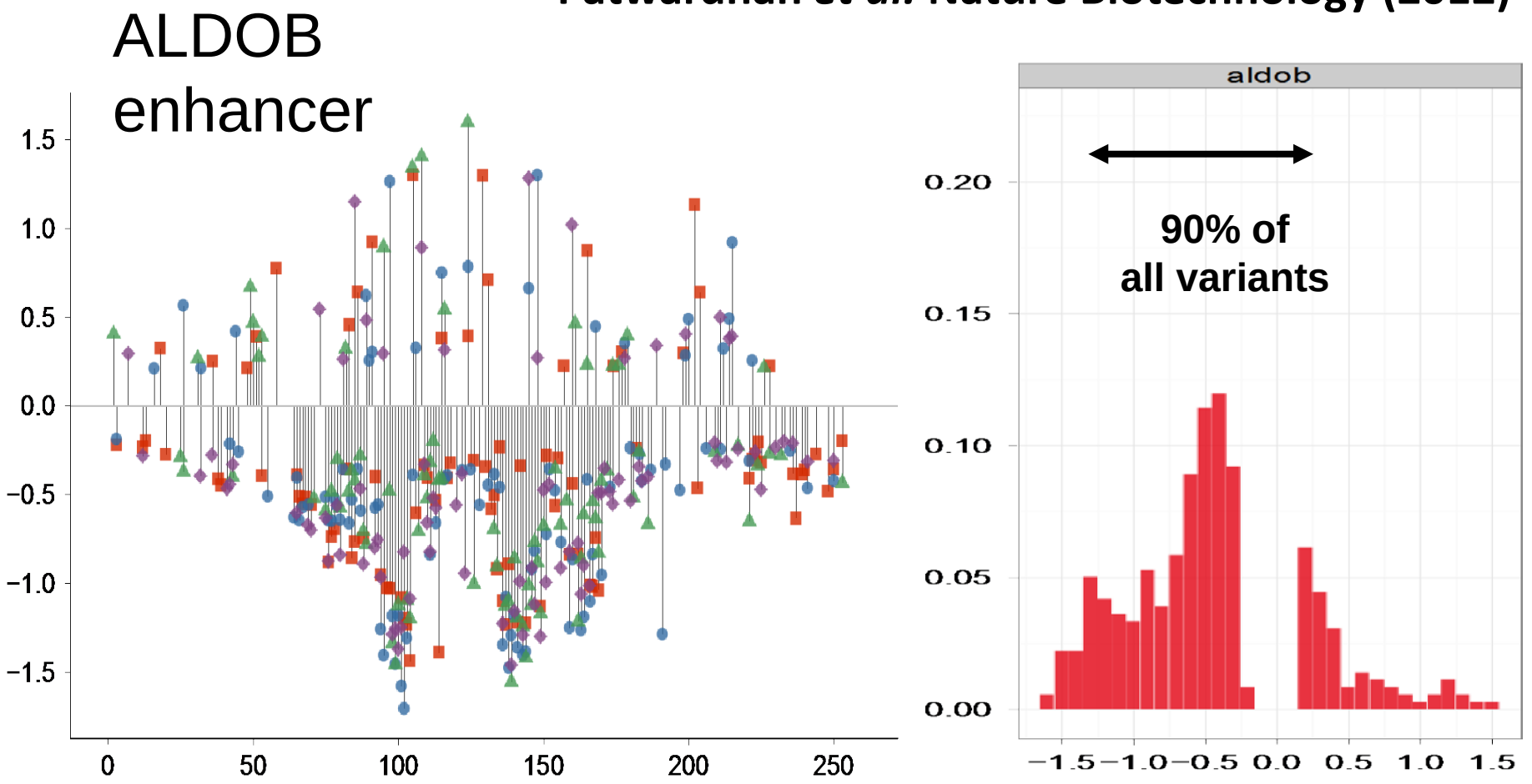
Activity Profile

PCR and sequencing of transcribed tags





Patwardhan *et al.* Nature Biotechnology (2012)



- **All possible mutations** assayed in one experiment
- **Distribution of effect sizes** for regulatory mutations (*i.e.* establishing null distribution)

Key points

- Experimental data can be very useful!
  - Identifying causal gene / variant(s)
  - Variants of (unknown → known) significance
  - Genetics → biological understanding
- Subjective exercise: no experiment is perfect
  - Demonstrating experimental effect ≠ causation
  - Failure to show effect ≠ non-causation
  - Multiple lines of evidence better
- Need for more high-throughput approaches

Discussion Questions

1. Feedback on accuracy, completeness and organization of experimental methods table?

2. Feedback on proposed levels of evidence?

1. How should experimental data be weighted relative to genetic analysis?

2. How should editors and reviewers be guided to think about experimental data in the context of manuscripts?

Selected Experimental Methods

Techniques	Variant class	Experiment type	High-throughput?	Evidence
mouse or zebrafish knockin	Any	in vivo	no	strong
genome editing	Any	in vivo, in vitro	yes	strong
mouse or zebrafish knockout	LoF allele	in vivo	no	strong
cell culture shRNA knockdown	LoF allele	in vitro	yes	suggestive
cDNA complementation	LoF allele	in vitro	no	suggestive
splicing assay	Splicing	ex vivo, in vitro, in vivo	yes	strong
protein-specific biochemical or cellular assays	Protein-altering alleles	in vitro	no	suggestive
correlation with expression	Regulatory	ex vivo, in vitro, in vivo	yes	suggestive
reporter construct	Regulatory	in vitro	yes	suggestive

Functional regulatory variation: Levels of evidence framework

Level 1: *in vivo* evidence from *in situ* models

1a *In situ* / whole locus model of strongly genetically implicated variant that precisely recapitulates the phenotype at the organismal level

1b *In situ* genome modification *(e.g., genome editing / knock-in/out)*

1c Whole-locus transgenic lines *(e.g., YAC, BAC; single copy)*

1d *In situ* measurement of gain/loss of regulatory protein binding directly coupled to *in vivo* gene product phenotype

1e *In situ* gain/loss of regulatory protein not coupled to gene product

Level 2: Evidence from artificial/condensed construct models

2a Standard transgenic animal

2b Stable transfection (integrated into genome)

2c Transient transfection (ex-genomic)

Level 3: Non-cellular assays (e.g., gel shifts)