Hirschsprung disease consequent to mutations in the *RET* gene regulatory network

From Genome Function to Biomedical Insight:
ENCODE and Beyond
National Institutes of Health, Bethesda, MD
11 March, 2015

Aravinda Chakravarti, PhD
McKusick-Nathans Institute of Genetic Medicine
Johns Hopkins University School of Medicine
(Disclosure: SAB, Biogen Idec)

What do we need to understand in disease?

"Why me? Why this disease? Why now?"
Genetic Medicine: A logic of disease, Barton Childs (2003)

- We have been preoccupied with mapping disease genes and have built quite impressive catalogs of both putative disease genes and putative causal variants…the precise identities are usually elusive and when known its mechanisms opaque;

- ENCODE is providing a view of non-coding functions through cis-regulatory elements (CREs) but simple intersection of sequence variants with CREs is insufficient and incomplete;

- Understanding human disease should involve answering, in mechanistic and quantitative terms, the "Why me? Why this disease?" part…we are nowhere close to that ideal.

Complex disease genomics: Reasonable speculations

- Numerous genes, probably in the 100s to 1000s;

- Common variants explain at least 50% of the trait/disease variation;

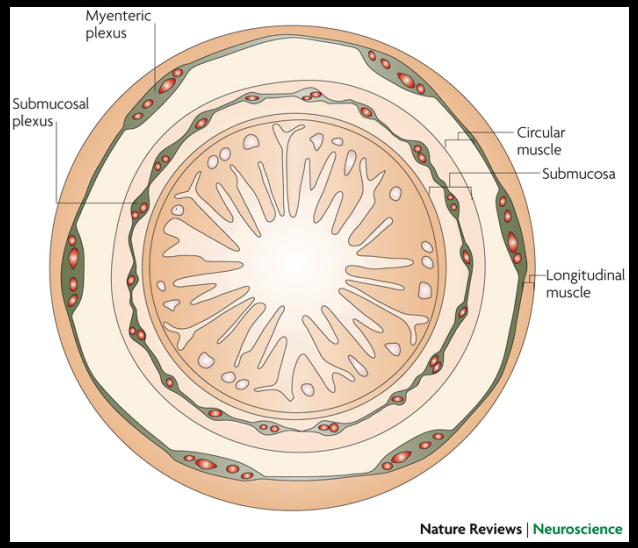
- Statistical power considerations will never allow us to individually map most common variants, let alone rare ones;

- Each locus may have >1 culprit gene but likely has many non-coding causal interacting variants since there are many CREs per gene and they act across time and space;

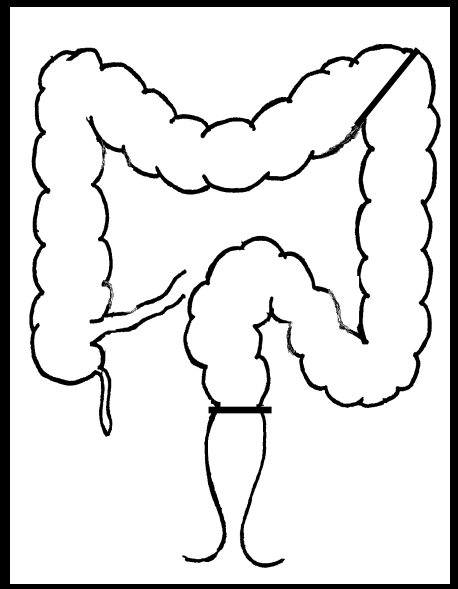
- Genomic CRE distributions suggest that the functional (physical) not genetic (recombination) haplotype should define the gene unit.

Hirschsprung disease as a model

Harald Hirschsprung (1830-1916)

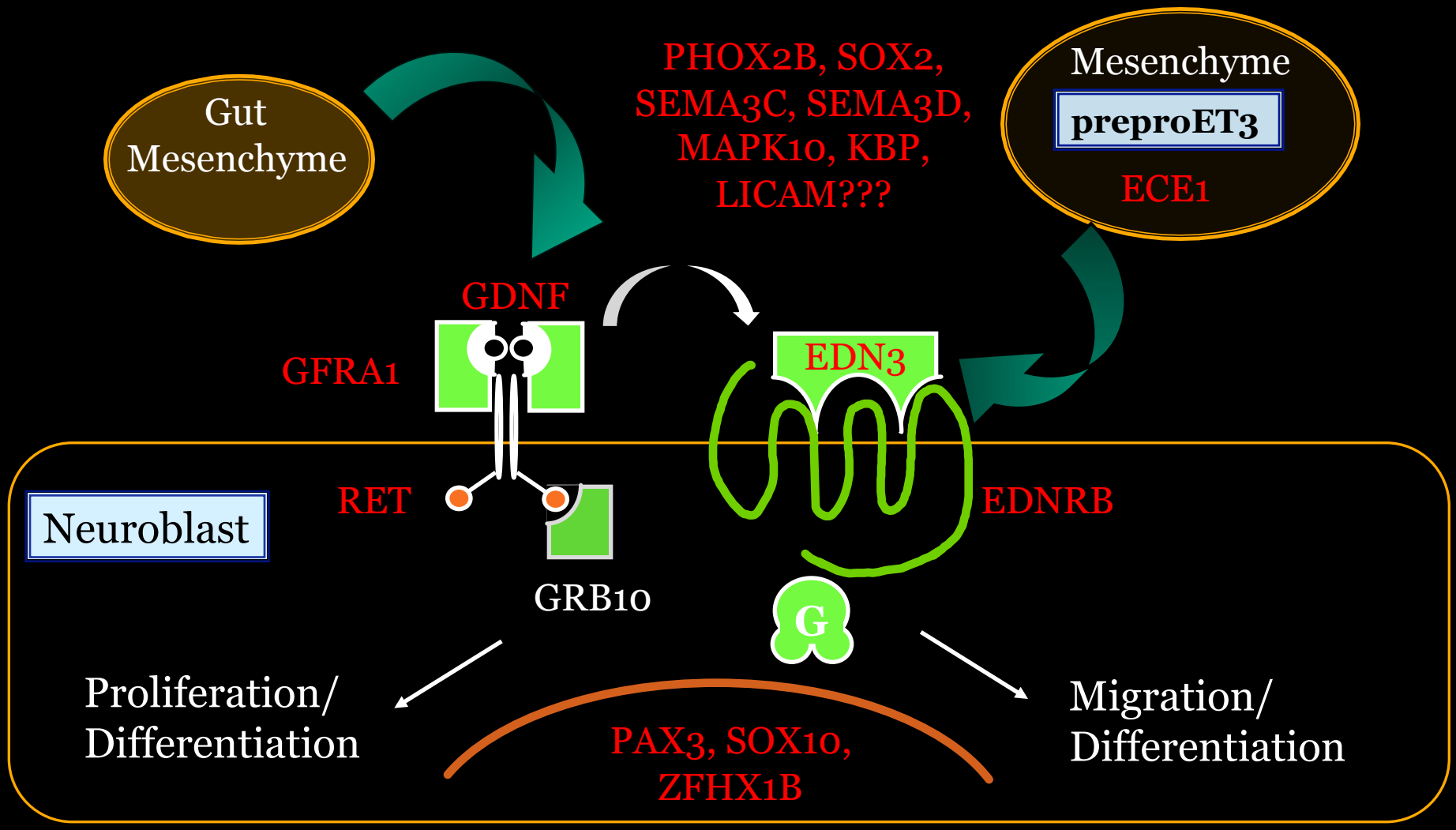


- absence of enteric ganglia
- 3 segmental forms: total colonic aganglionosis (5%), long (15%) & short (80%)
- 1/5,000 live-births in Europeans, 2X among Asians
- multifactorial inheritance with 4% recurrence risk to siblings
- 4:1 male: female affected
- 22 known syndromic associations (Down, Waardenberg-Shah, Mowat-Wilson, Bardet-Biedl, Goldberg-Shprintzen, etc.)



Lyonnet & Chakravarti, *MMBID*, 2001, 2008

The logic of HSCR: Disrupted cellular interaction between enteric neuroblasts and intestinal mesenchyme affecting proliferation, differentiation, migration and innervation



RET is the major gene for HSCR

- Deletions and LoF coding mutations in syndromic and severe (L-HSCR/TCA) forms;
- numerous rare, *de novo* and segregating low penetrance (<50%) coding variants in L-HSCR/TCA cases with partial or complete LoF;
- analogous LoF variants in other members of the pathway (*SOX10, GDNF, GFRA1*);
- polymorphic (25%) variant in a gut-specific, SOX10-bound *RET* intronic enhancer with *RET* LoF confers high risk (OR ~4), is associated with S-HSCR and other features and explains 20X the variance;
- one would assume that any regulatory variant that reduces *RET* expression would have similar effects on HSCR.

Emison et al *Nature* 2005, *Am J Hum Genet* 2010

Kapoor et al *Hum Mol Genet* 2015

Forward genetic screen for *RET* – HSCR SNP associations

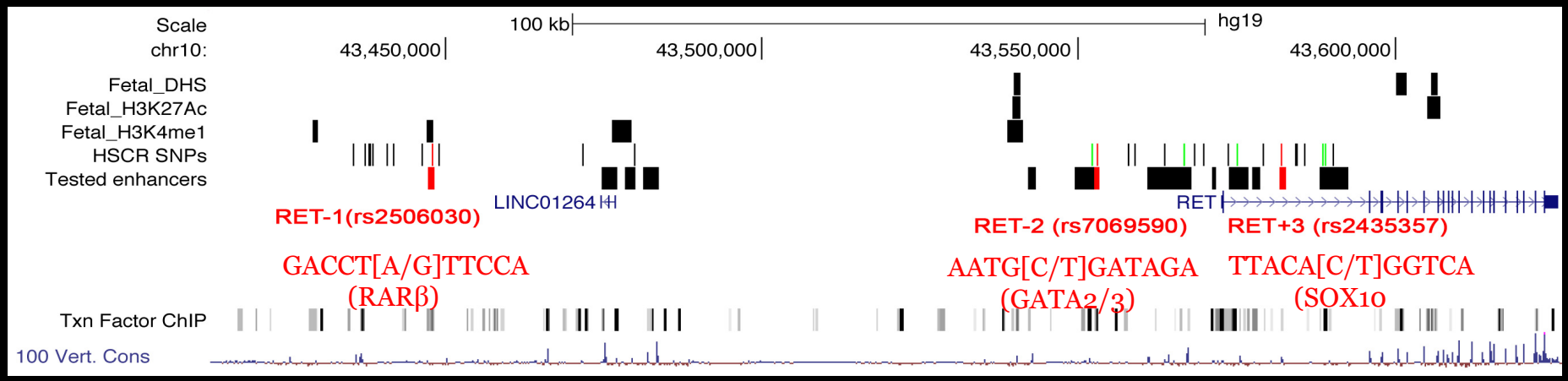
SNP	RET+3 (rs2435357)	RET-1 (rs2506030)	RET-2 (rs7069590)
Risk/non-risk allele	T/C	G/A	T/C
Risk allele frequency (EUR)	25%	39%	75%
Odds ratio	4.0 ($P=2.93 \times 10^{-41}$)	1.9 ($P=4.03 \times 10^{-11}$)	1.7 ($P=9.55 \times 10^{-6}$)
Distance from *RET*	+9.7 kb	-125 kb	-18 kb

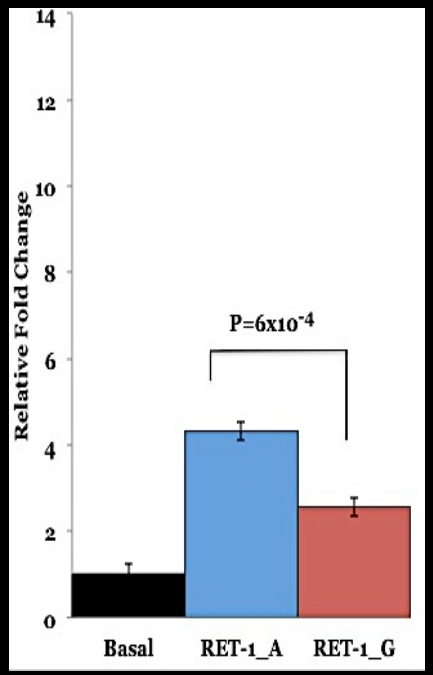
- 125 SNPs with MAF >10% at RET, 34 are HSCR associated ($P=5 \times 10^{-8}$), 8 lie within ChIP-seq peaks in SK-N-SH, all are enhancers (Grice enhancers), only 3 show allelic differences;
- Effect sizes are substantial marginally and combinatorially (odds ratios vary >30-fold by genotype);
- SNPs show significant levels of LD ($r^2 \sim 0.1$ but D' > 0.6);
- Is a major source of missing heritability.

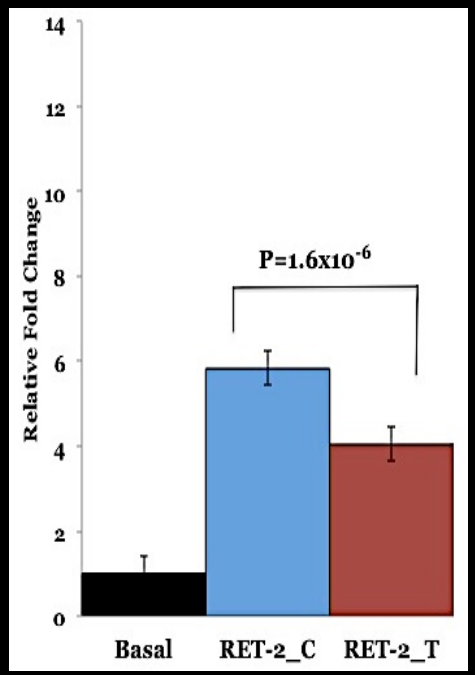
Ashish Kapoor

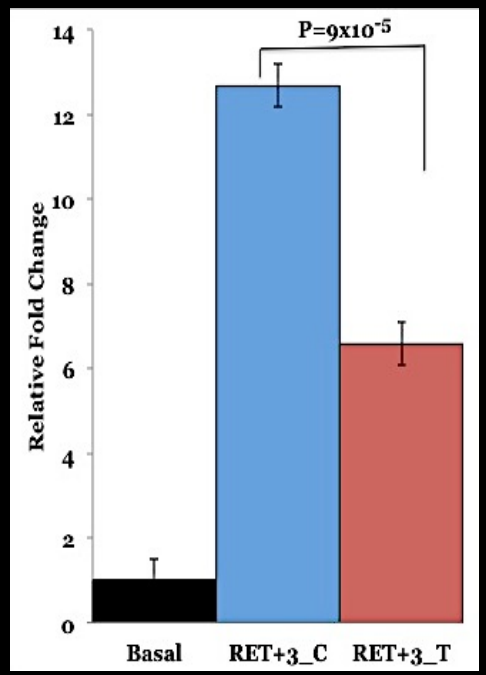
RET cis-regulatory elements with HSCR variants have enhancer activity

(luciferase assays in Neuro2A cells)

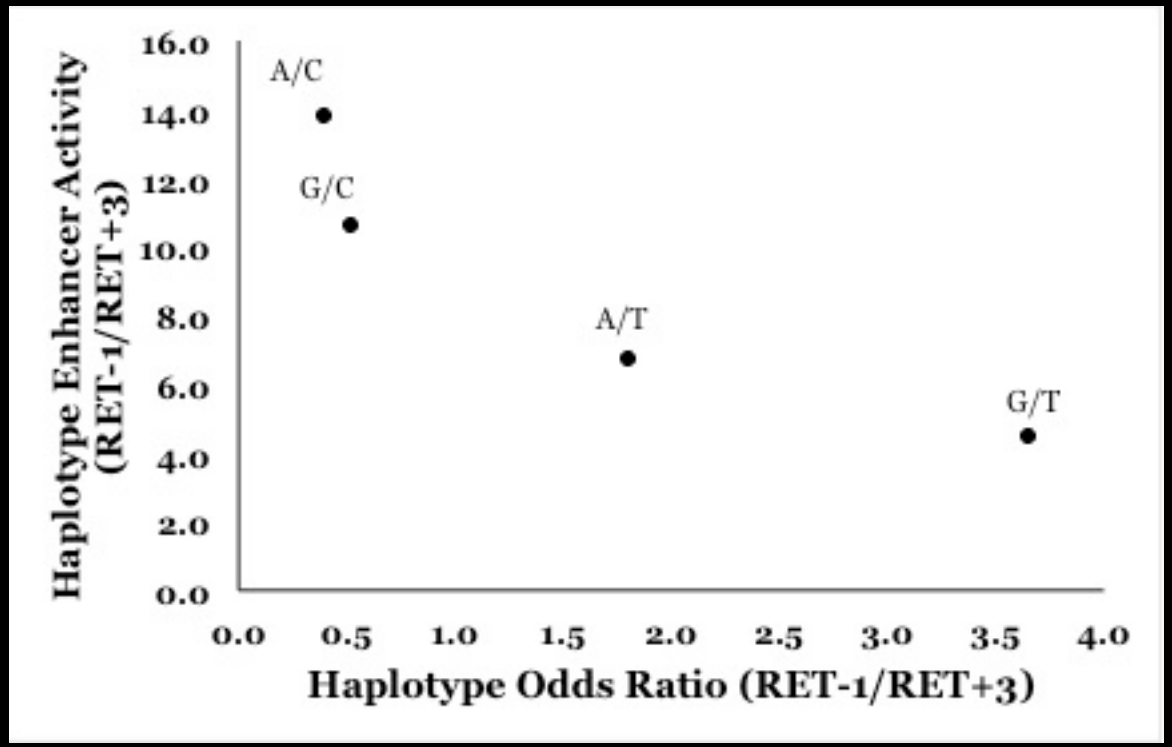








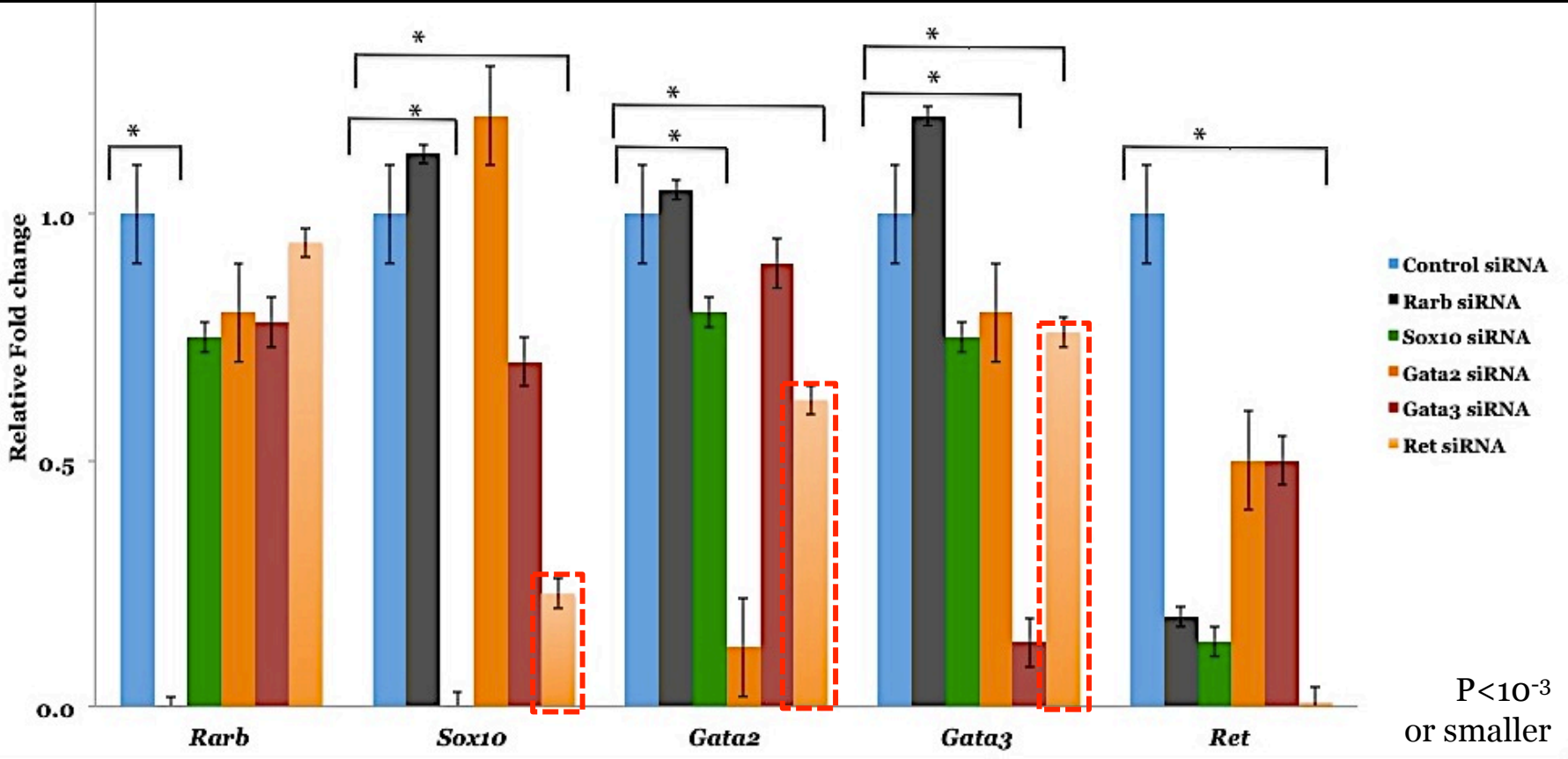
Synthetic haplotype enhancer activity is a surrogate* for disease odds ratio



* logarithmic effect

Rarb, Sox10, Gata2/3 regulate *Ret* transcription

(*in vitro* analysis in Neuro2A cells)



$P<10^{-3}$ or smaller

Interestingly, *Ret* knockdown decreases *Sox10, Gata2 and Gata3* (but not *Rarb*) implying positive feedback.

Variants in HSCR affect all components of RET signaling: production, activation & signal termination through its GRN…total effect depends on degree of LoF

PRODUCTION
(+)

Gata2,3 *Sox10* → Sox10 → *Ret* → Ret
Rarb Gata2,3
Rarb

Gdnf → Gdnf
Gfra1 → *Gfra1*

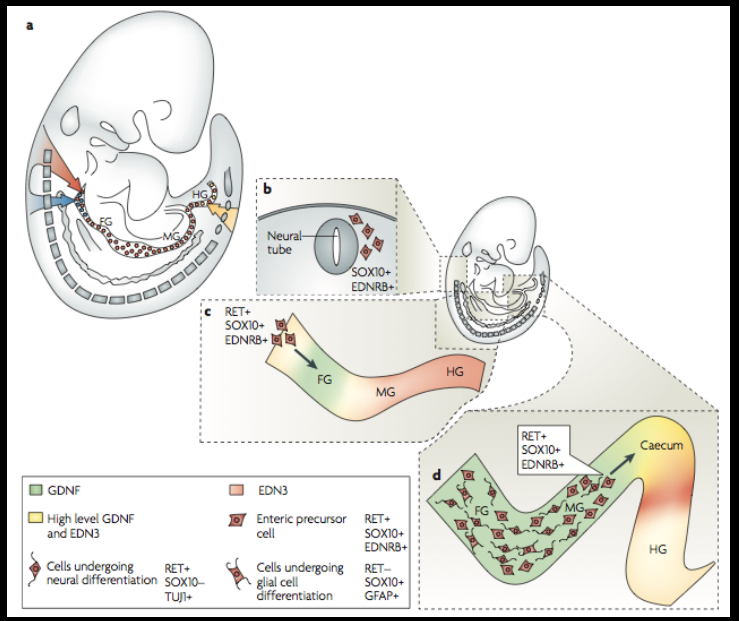
ACTIVATION
(-)

(Ret-Ret)Ⓟ

SIGNAL
TERMINATION (+)

*Cbl** → Cbl

~~(Ret-Ret)Ⓟ~~

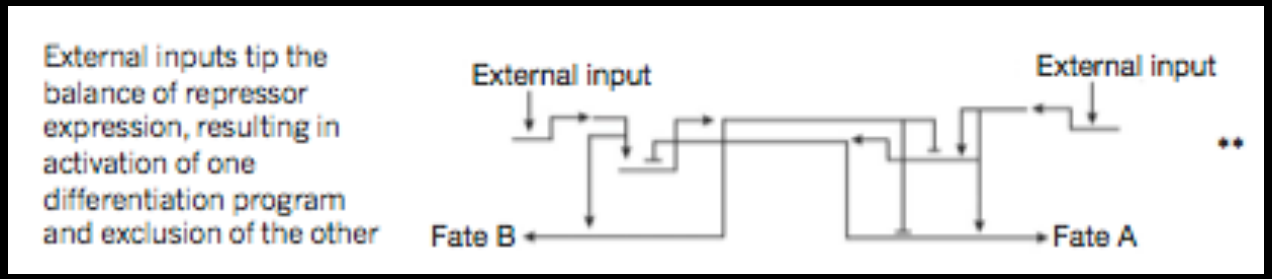


**Cbl* expression is reduced in Ret null vs. wildtype mice

Gene Regulatory Networks (GRN)

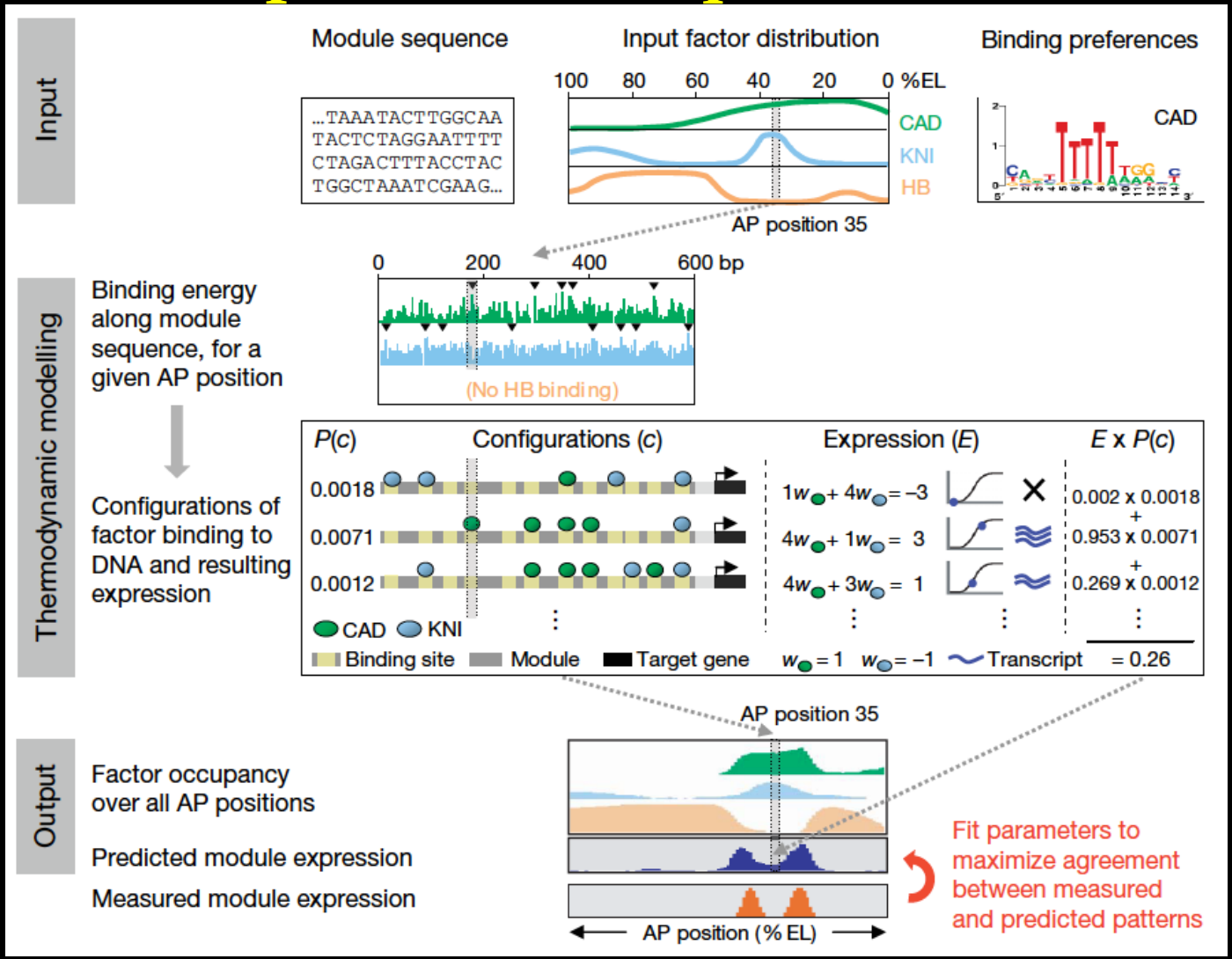
Since no gene acts alone, understanding the functional context of its expression is crucial to understanding, modeling and predicting its mutational consequences. The GRN is the fundamental unit to consider because it is:

(1) modular and "conserved";
(2) comes in a defined number of sub-circuit classes;
(3) provides systems-level explanations of developmental and physiological functions;
(4) mathematical modeling can provide quantitative explanations of processes that result from GRN function;
(5) provides a basis for understanding complex inheritance.



Eric H Davidson 2001
Uri Alon 2006
Eric H Davidson *Nature* 2010

A quantitative, mechanistic model for gene activity, trait values (quantitative) and penetrance (qualitative)



Segal et al
Nature 2008

Acknowledgements

Johns Hopkins University
Sumantra Chatterjee
Ashish Kapoor

Dongwon Lee
Tychele Turner
Maria Sosa
Courtney Berrios

Lawrence Berkeley Labs, Berkeley, CA
Jennifer Akiyama
Len Pennacchio
Eddy Rubin

Broad Institute, Cambridge, MA
Namrata Gupta
Stacey Gabriel

Funding: NICHD (MERIT)

ENCODE4 Questions

1) What gap does your proposed project (s) fill and why is it a high priority?
2) Why is this project appropriate for NHGRI vs other ICs?

Comprehensive CRE mapping, analysis and subsequent tests of genetic association for a quantitative trait…in other words, do some 'forward genetics' demonstration projects. Include modelers and quantitative scientists upfront during study design.

3) What new technological breakthroughs would be transformative?

Easy, high-efficiency perturbation (knockdown/deletion/over-expression) of individual TFs, co-factors and target genes with subsequent RNA-seq in a multiplexed manner;

4) What additional unbiased data generation efforts would facilitate these studies?

Specific perturbation experiments in a series of cell-types and cell lines…both to assess and refine the technologies and to recover the gene expression program and its hierarchy.

Create a developmental versus maintenance ENCODE data set to understand which regulatory programs are 'fixed' and which 'dynamic'.

5) Would this project benefit from a particular scale and/or organizational structure?

Consortia for large-scale data generation but individual labs for the genetics.