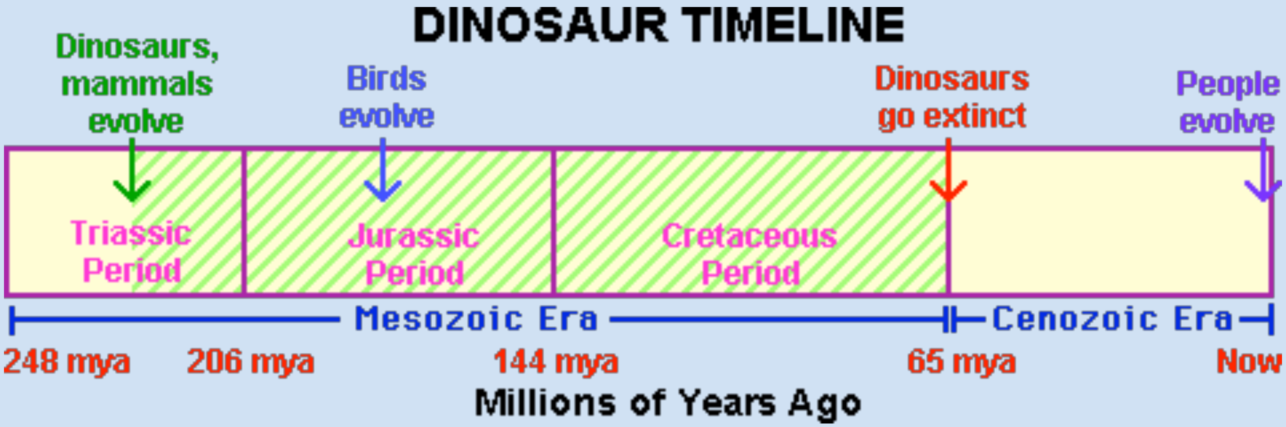
Human Genetic Variation and the Genotype-Phenotype Problem

Gholson Lyon, M.D. Ph.D.

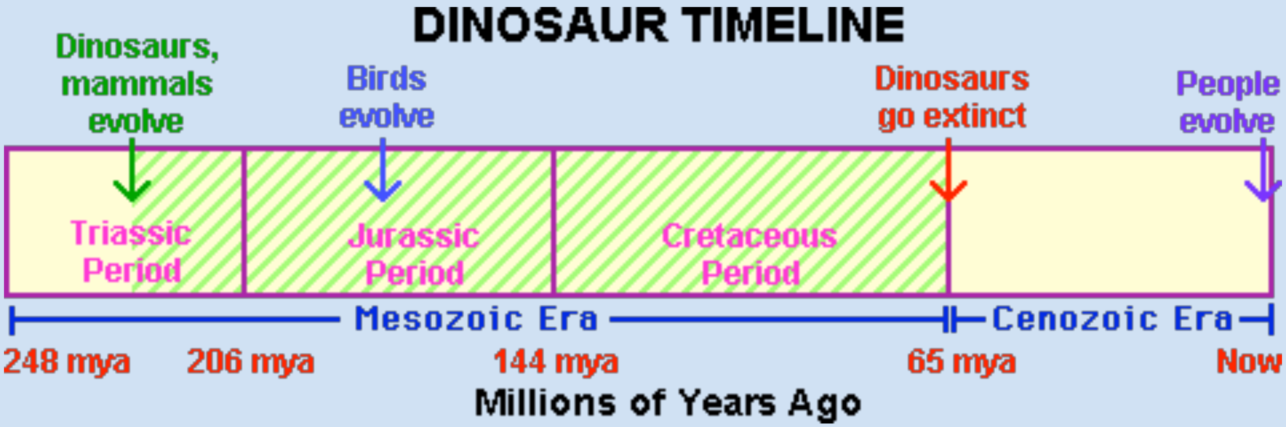
CSH
STANLEY INSTITUTE FOR
COGNITIVE GENOMICS
COLD SPRING HARBOR LABORATORY

UFBR
UTAH FOUNDATION FOR
BIOMEDICAL
RESEARCH

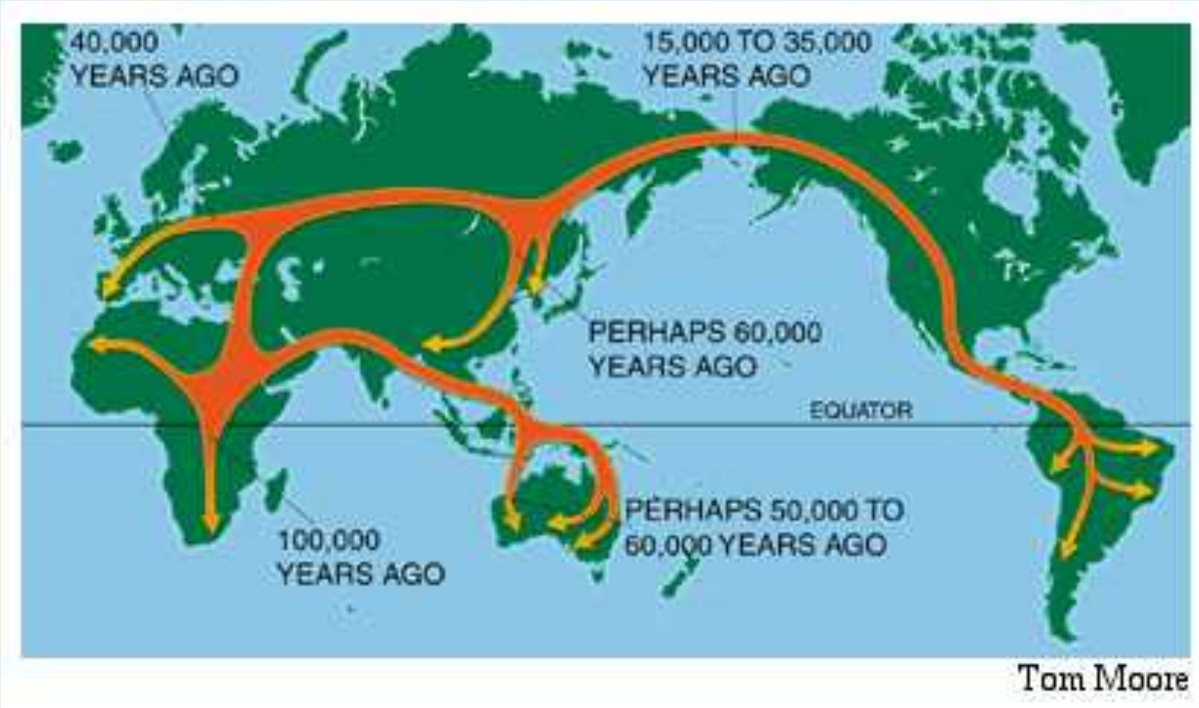

The Earth is 4.5 Billion Years Old
User: Fred Zabrodnik

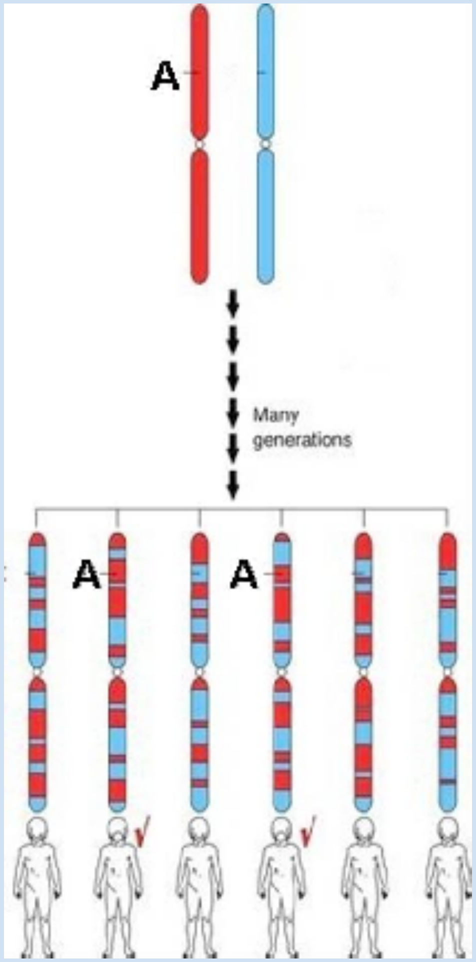

DINOSAUR TIMELINE
Dinosaurs, mammals evolve
Birds evolve
Dinosaurs go extinct
People evolve
Triassic Period
Jurassic Period
Cretaceous Period
Mesozoic Era
Cenozoic Era
248 mya
206 mya
144 mya
65 mya
Now
Millions of Years Ago


The Earth
is 4.5 Billion
Years Old

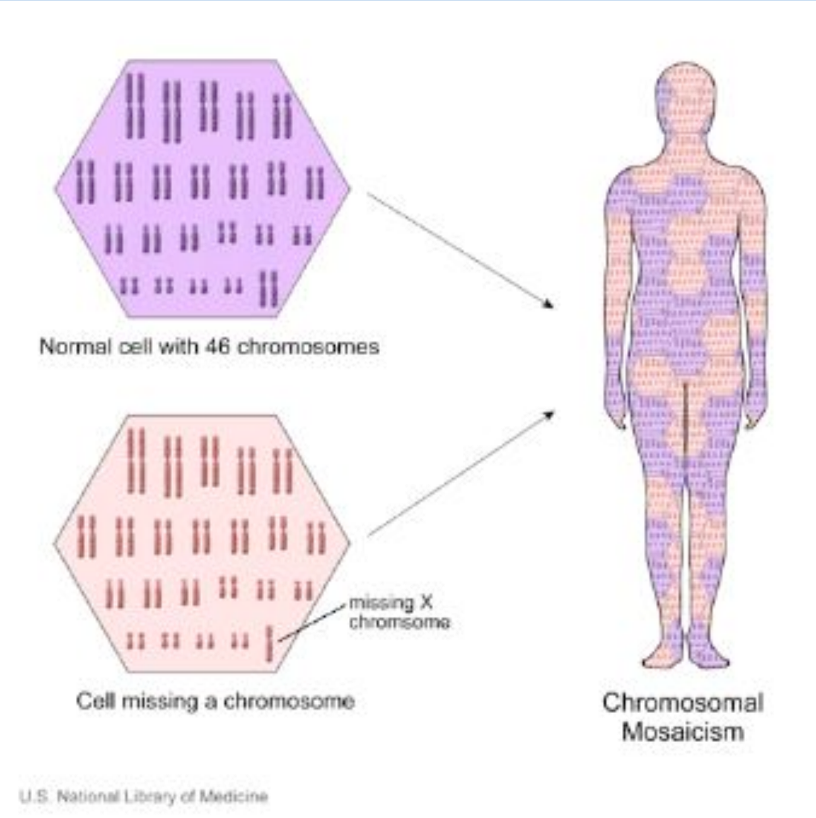

DINOSAUR TIMELINE
Dinosaurs, mammals evolve
Birds evolve
Dinosaurs go extinct
People evolve
Triassic Period
Jurassic Period
Cretaceous Period
Mesozoic Era
Cenozoic Era
248 mya
206 mya
144 mya
65 mya
Now
Millions of Years Ago


Gulf of Mexico
Campeche Escarpment
Outer rim
of buried
impact crater
Yucatán
Peninsula





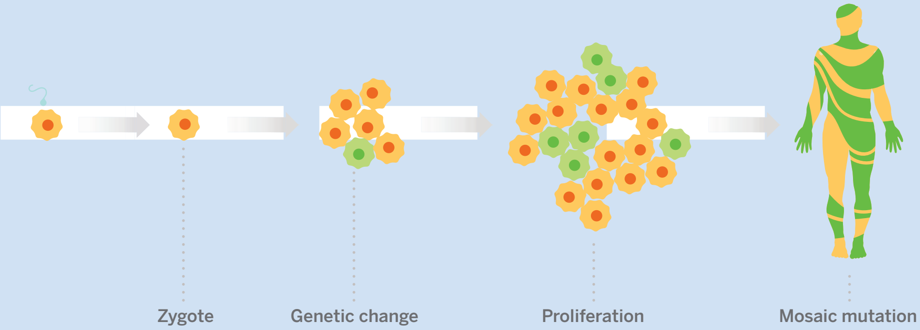
Population Stratification is from the migration patterns of haplotypes throughout human history



U.S. National Library of Medicine

Building a human mosaic

Depending on when and where in embryonic development a mutation occurs, a subset of adult cells will harbor the genetic change.



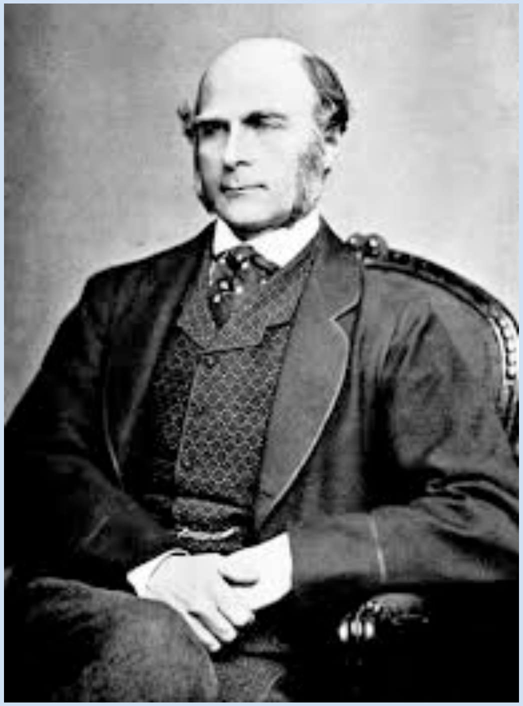
CLINICAL TESTING

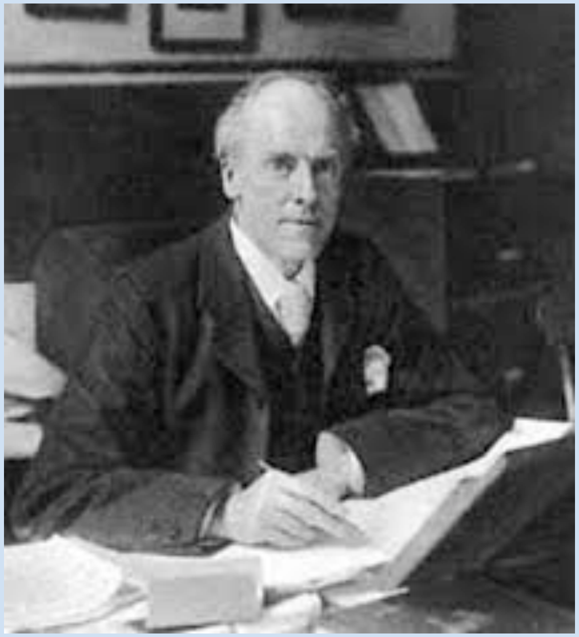
Harmful mutations can fly under the radar

With more sensitive genetic tests, researchers are hunting the roots of disease in the human “mosaic”


Charles Darwin
1809-1882


Gregor Mendel
1822-1884


Frances Galton
1822-1911



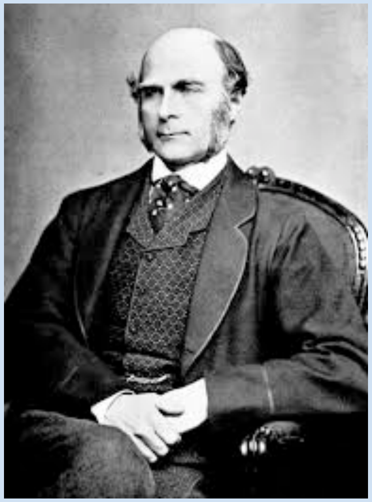
Karl Pearson
1857-1936

Walter Frank Raphael Weldon
1860-1906

William Bateson
1861-1926

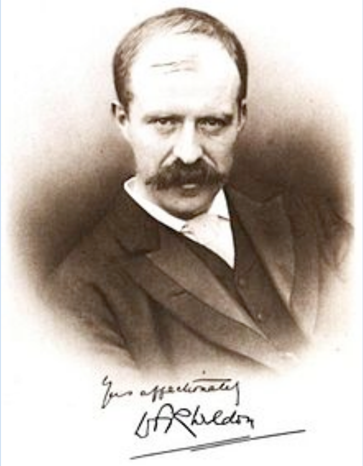
Charles Darwin
1809-1882

Gregor Mendel
1822-1884



Frances Galton
1822-1911

Karl Pearson
1857-1936



Walter Frank Raphael
Weldon
1860-1906

William
Bateson
1861-1926

Beyond our Kuhnian inheritance

A recent lecture by Prof Greg Radick questions our scientific inheritance, through textbook histories of genetics and Thomas Kuhn's legacy

http://www.guardian.co.uk/science/the-h-word/2012/aug/28/thomas-kuhn

Walter Frank Raphael Weldon

Vs.

William Bateson

Forthcoming by Greg Radick. Scholarly edition of W. F. R. Weldon's Theory of Inheritance (1904-1905), coedited with Annie Jamieson.

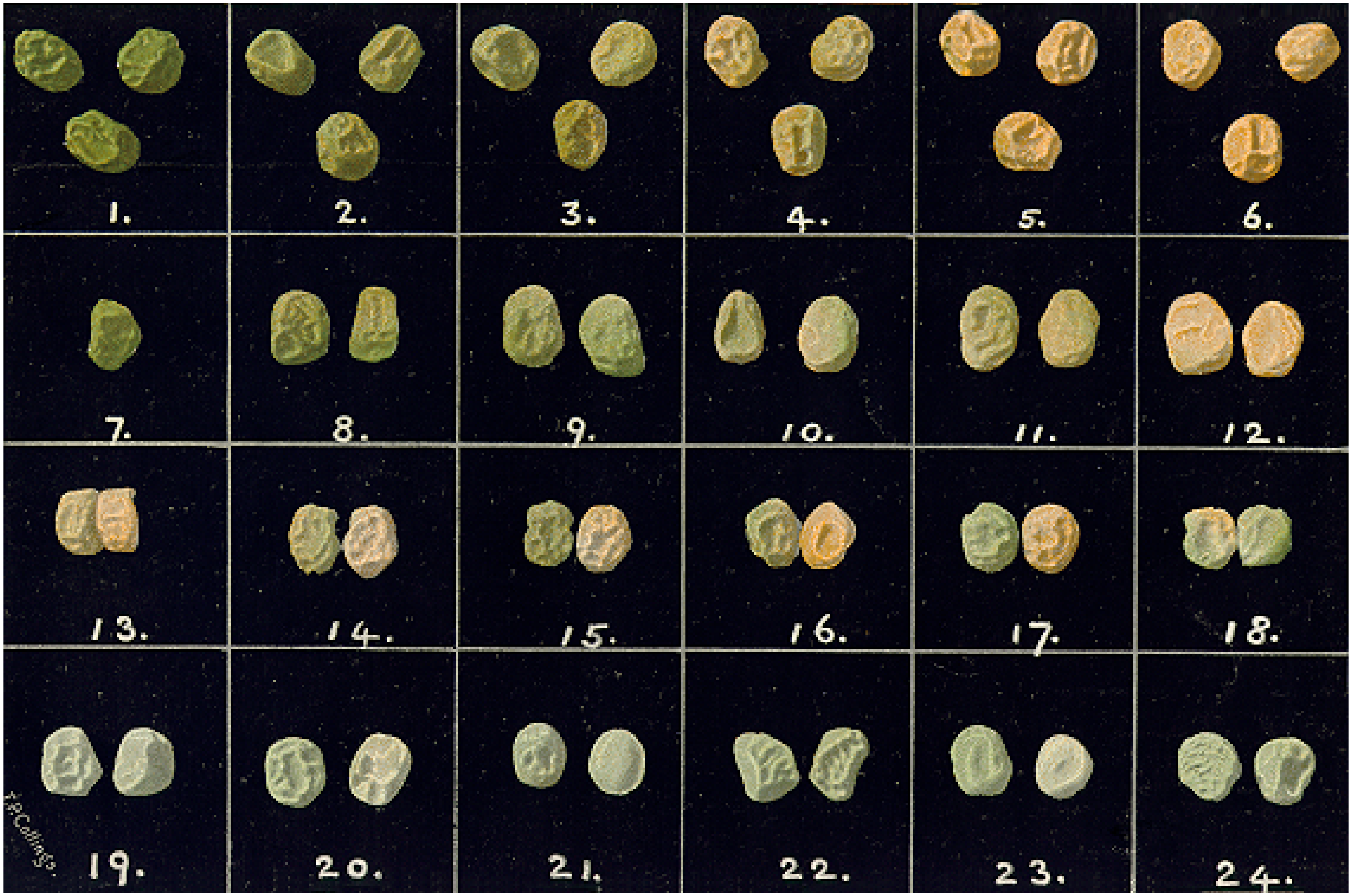


Plate I.

Weldon, W. F. R. 1902. Mendel's laws of alternative inheritance in peas. *Biometrika*, 1:228-254.

- *"The fundamental mistake which vitiates all work based upon Mendel's method is the neglect of ancestry, and the attempt to regard the whole effect upon offspring, produced by a particular parent, as due to the existence in the parent of particular structural characters; while the contradictory results obtained by those who have observed the offspring of parents apparently identical in certain characters show clearly enough that not only the parents themselves, but their race, that is their ancestry, must be taken into account before the result of pairing them can be predicted" – Walter Frank Raphael Weldon (Weldon, 1902).*

Walter Frank Raphael Weldon 1860–1906

A Memoir

Karl Pearson

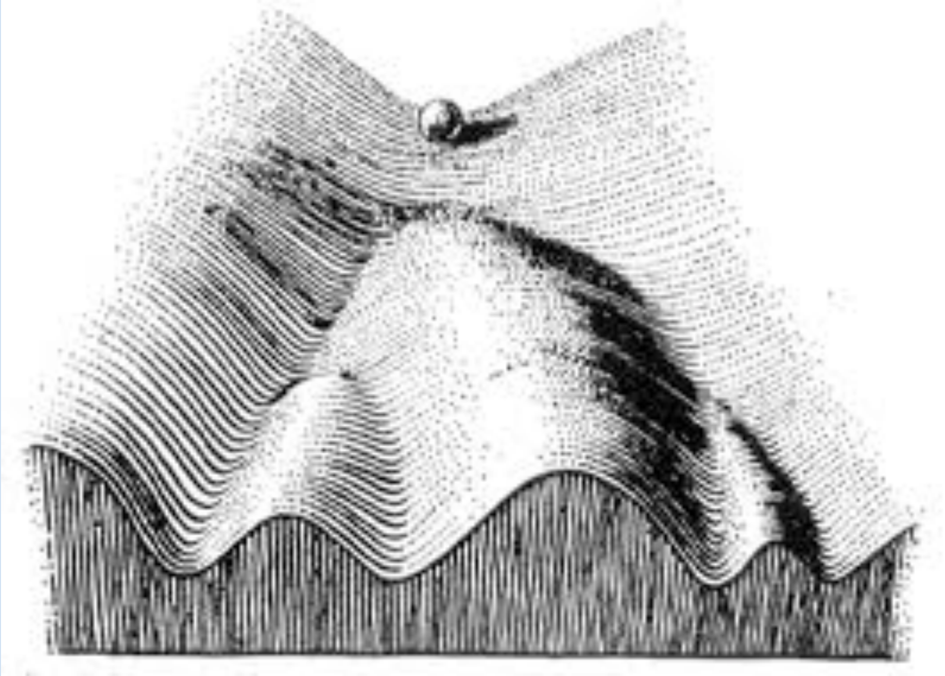
CAMBRIDGE

“Biological Indeterminacy”

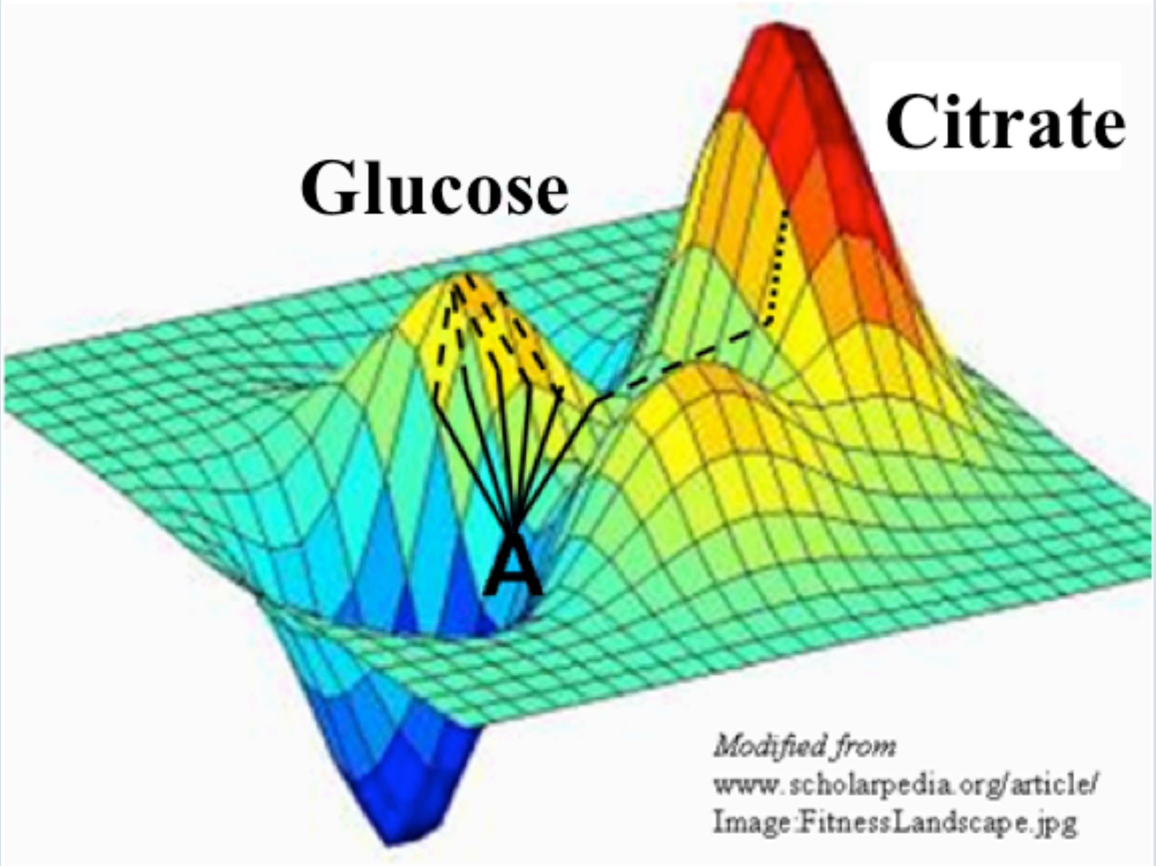
- Bateson became famous as the outspoken Mendelian antagonist of Walter Raphael Weldon, his former teacher, and Karl Pearson who led the biometric school of thinking. This concerned the debate over saltationism versus gradualism (Darwin had been a gradualist, but Bateson was a saltationist). Later, Ronald Fisher and J.B.S. Haldane showed that discrete mutations were compatible with gradual evolution: see the modern evolutionary synthesis.

Biological Indeterminacy. Greenspan RJ. Sci Eng Ethics. 2012 Jul 3

Biological Robustness.



The canalisation metaphor suggests that phenotypes are very robust to small perturbations, for which development does not exit the canal, and rapidly returns back down, with little effect on the final outcome of development. But perturbations whose magnitude exceeds a certain threshold will break out of the canal, moving the developmental process into uncharted territory. Strong robustness up to a limit, with little robustness beyond, is a pattern that could increase evolvability in a fluctuating environment.



E. coli adapting to low glucose conditions, in the context of media containing citrate.
– Richard Lemski experiment
"Finally, novel functions often emerge in rudimentary forms that must be refined to exploit the ecological opportunities. This three-step process — in which potentiation makes a trait possible, actualization makes the trait manifest, and refinement makes it effective — is probably typical of many new functions." - Lemski

Genotype ≠ Phenotype

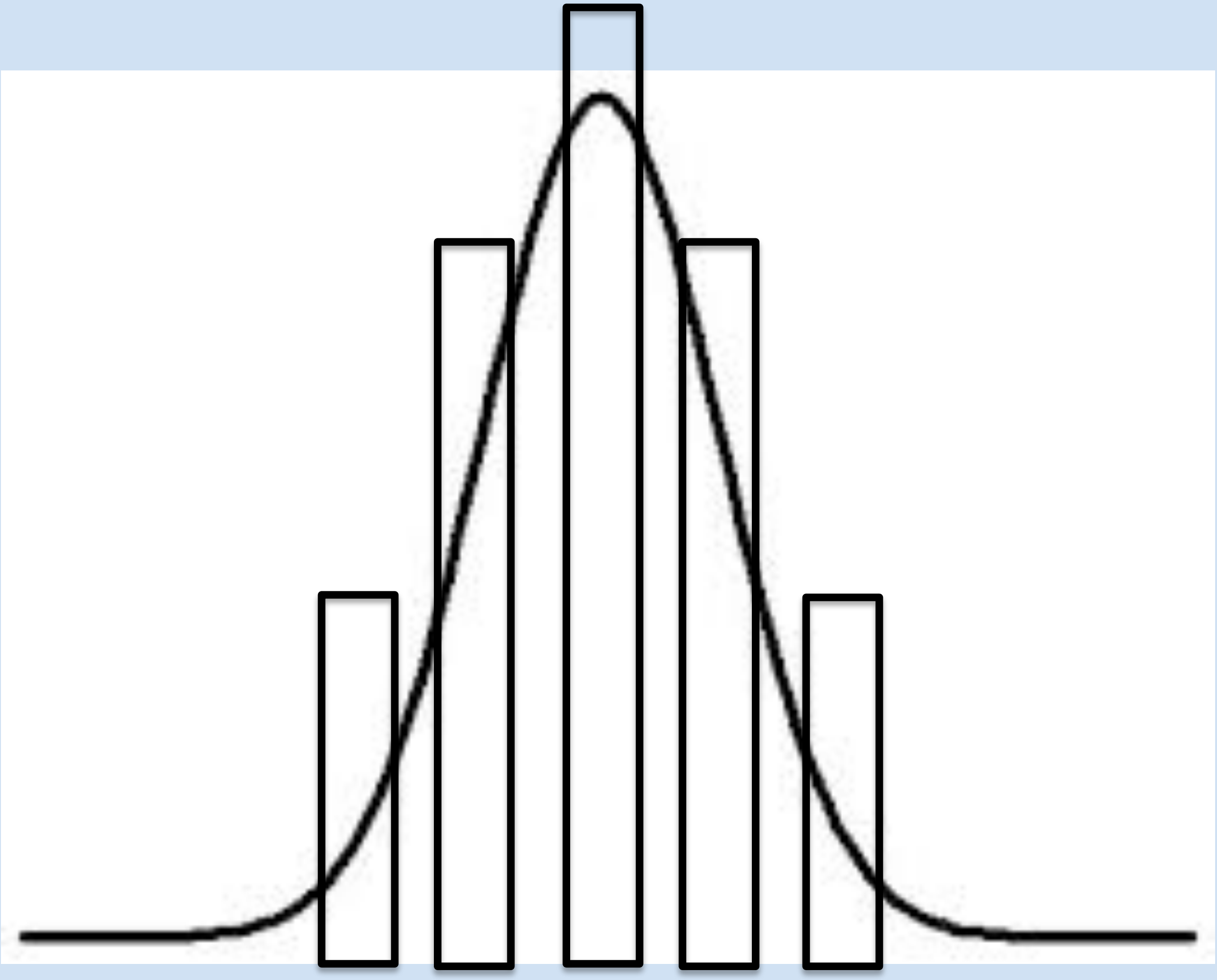
Environment matters!

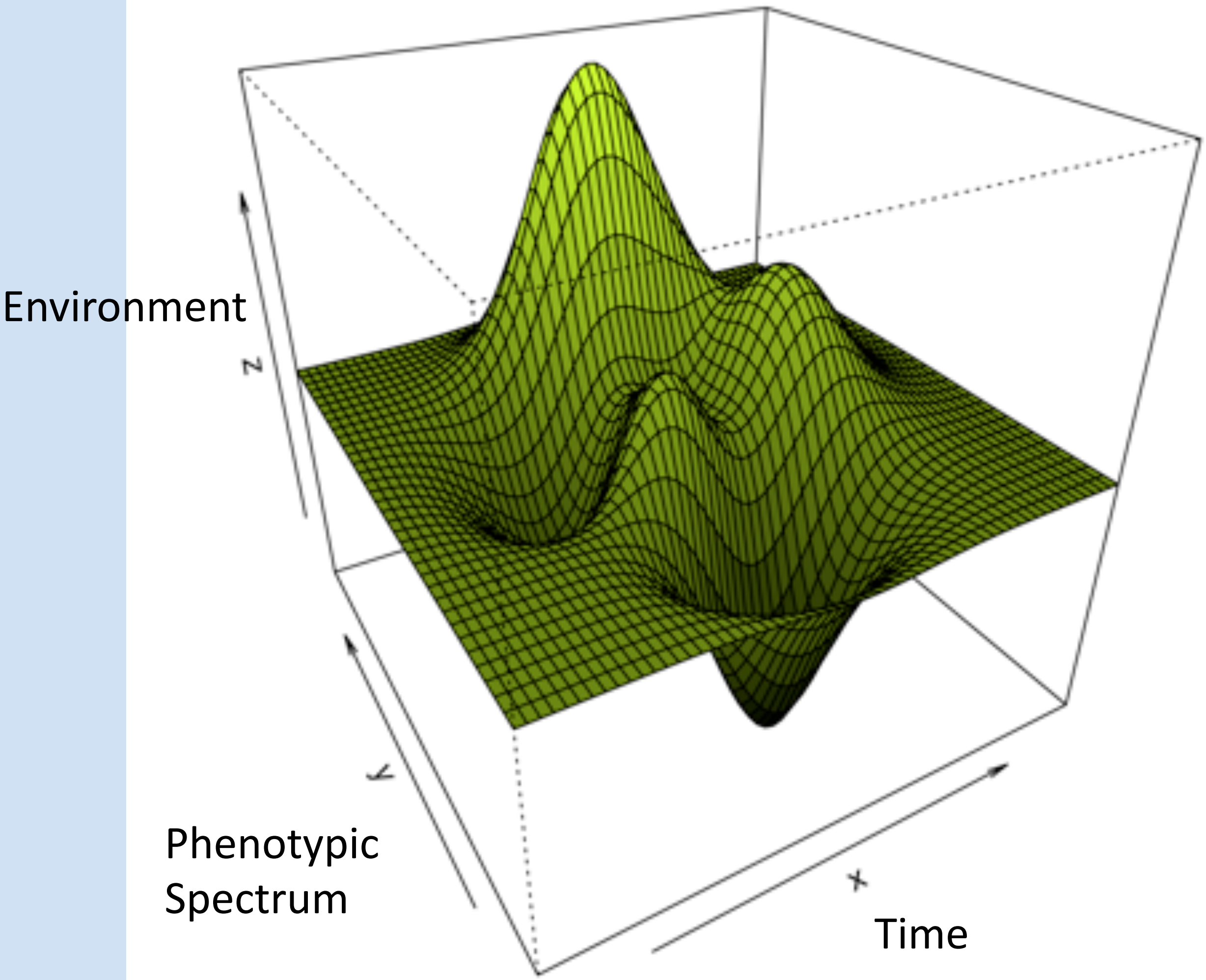
Ancestry matters!

Genomic background matters!

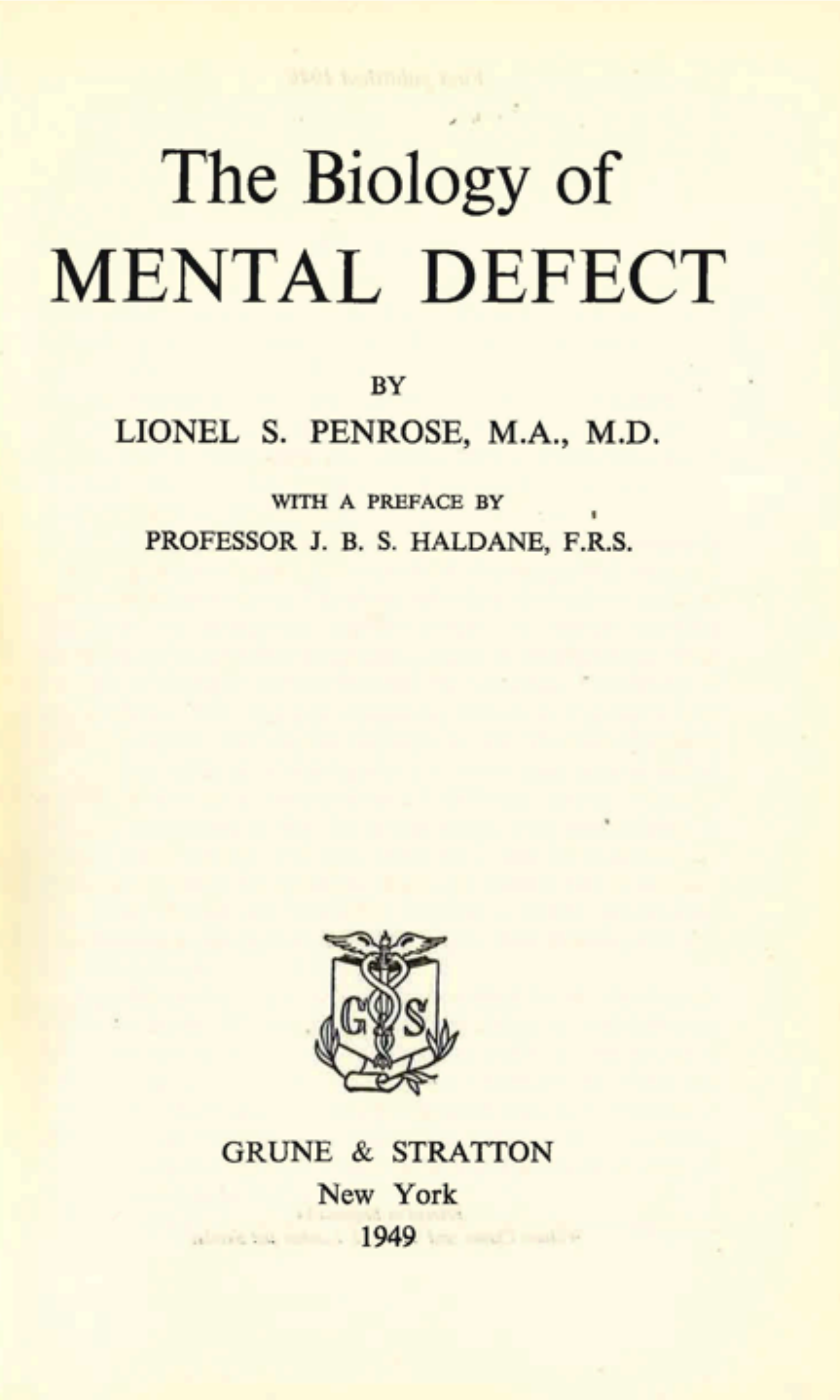
Longitudinal course matters!

Categorical Thinking Misses Complexity





A conceptual model of genotype-phenotype correlations. The *y* plane represents a phenotypic spectrum, the *x* plane represents the canalized progression of development through time, and the *z* plane represents environmental fluctuations.

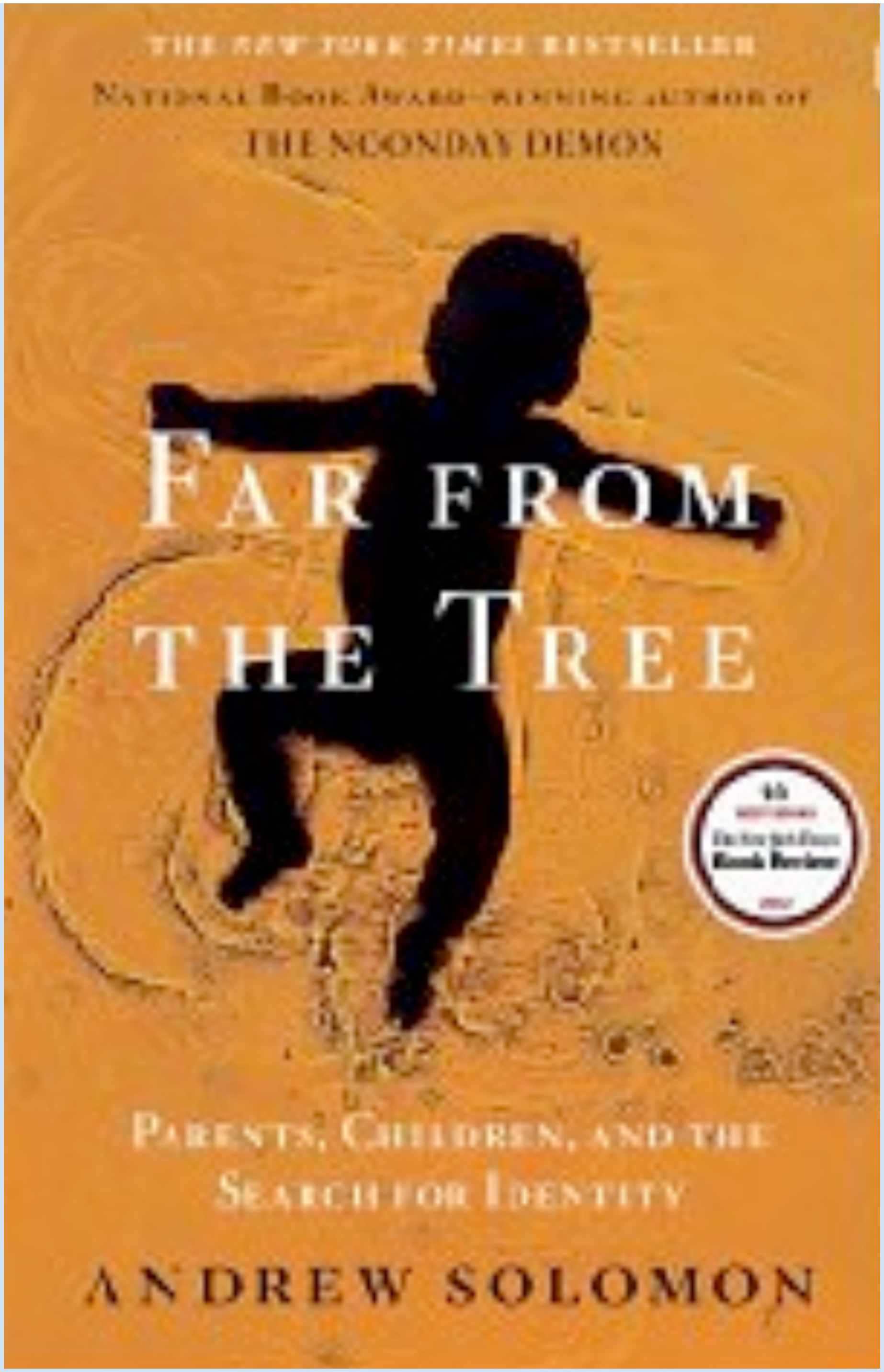


The Biology of
MENTAL DEFECT

BY
LIONEL S. PENROSE, M.A., M.D.

WITH A PREFACE BY
PROFESSOR J. B. S. HALDANE, F.R.S.

GRUNE & STRATTON
New York
1949



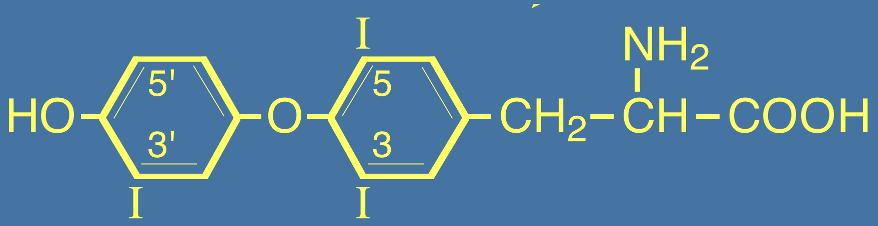
THE NEW YORK TIMES BESTSELLER
NATIONAL BOOK AWARD–WINNING AUTHOR OF
THE NOONDAY DEMON

FAR FROM THE TREE

PARENTS, CHILDREN, AND THE
SEARCH FOR IDENTITY

ANDREW SOLOMON

Cretinism – lack of iodine in the diet, leading to thyroid hormone deficiency.



Thyroid Hormone

Down Syndrome

Down Syndrome

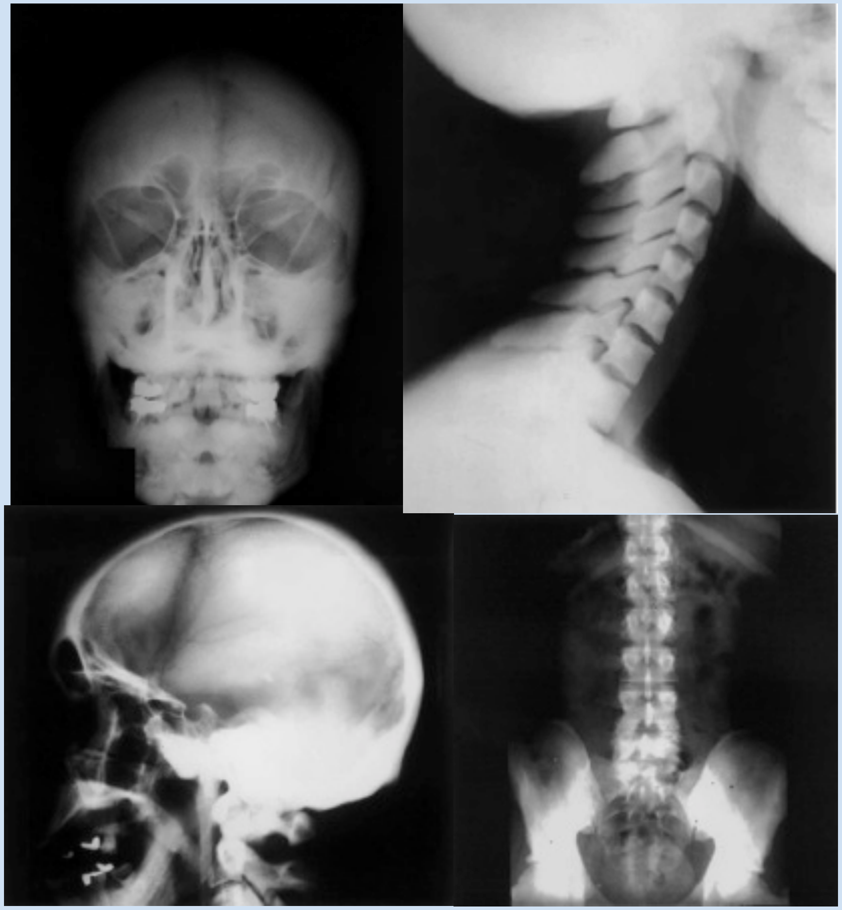
Christopher Joseph "Chris" Burke (born August 26, 1965) is an American actor and folk singer, who lives with Down syndrome, who has become best known for his character Charles "Corky" Thacher on the television series Life Goes On.

And there are people with Mosaic Down Syndrome, who are much less affected.

Velocardiofacial (22q11.2) Syndrome

“Superpower” mutations???

Myostatin mutation
Exon 2 allele P198A



LRP5 mutation
D111Y, G171R, A214T, A214V, A242T, and T253I

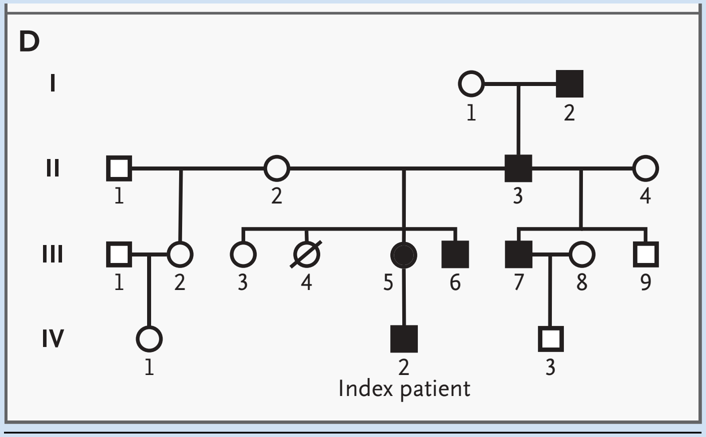
**Thanks to George Church for discussions on this.

Myostatin Mutation Associated with Gross Muscle Hypertrophy in a Child

Markus Schuelke, M.D., Kathryn R. Wagner, M.D., Ph.D., Leslie E. Stolz, Ph.D.,
Christoph Hübner, M.D., Thomas Riebel, M.D., Wolfgang Kömen, M.D.,
Thomas Braun, M.D., Ph.D., James F. Tobin, Ph.D., and Se-Jin Lee, M.D., Ph.D.

N ENGL J MED 350;26 WWW.NEJM.ORG JUNE 24, 2004


7 Months



Liam is homozygous for the mutation.

Another example: Liam Hoekstra, known as the world's strongest toddler at age 3, has a condition called myostatin-related muscle hypertrophy which results in increased muscle mass and reduced body fat. Myostatin-related muscle hypertrophy, or muscle enlargement, is an extremely rare genetic condition. – How rare???

http://videos.disabled-world.com/video/159/liam-hoekstra-strongest-boy-in

Belgian Blue is a breed of beef cattle from Belgium. The Belgian Blue has a natural mutation in the myostatin gene which codes for the protein, myostatin.

http://en.wikipedia.org/wiki/Belgian_Blue

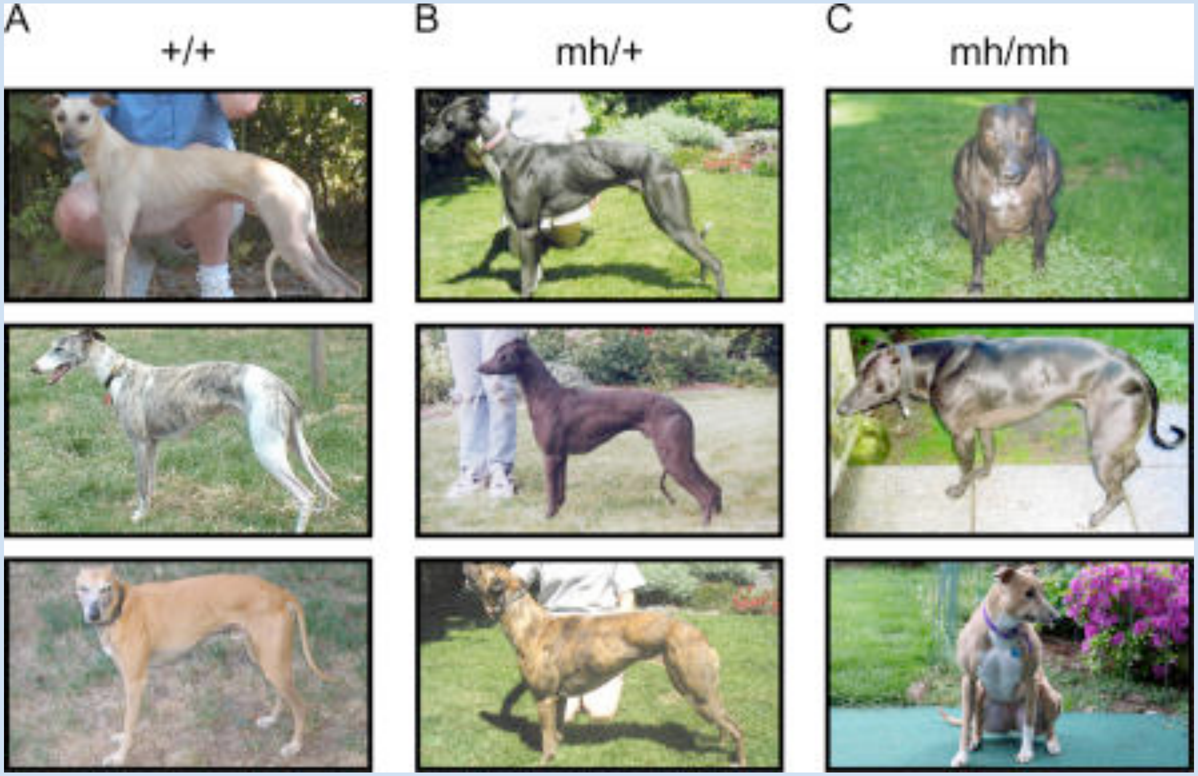
A Mutation in the Myostatin Gene Increases Muscle Mass and Enhances Racing Performance in Heterozygote Dogs

Dana S. Mosher[1], Pascale Quignon[1], Carlos D. Bustamante[2], Nathan B. Sutter[1], Cathryn S. Mellersh[3], Heidi G. Parker[1], Elaine A. Ostrander[1*]

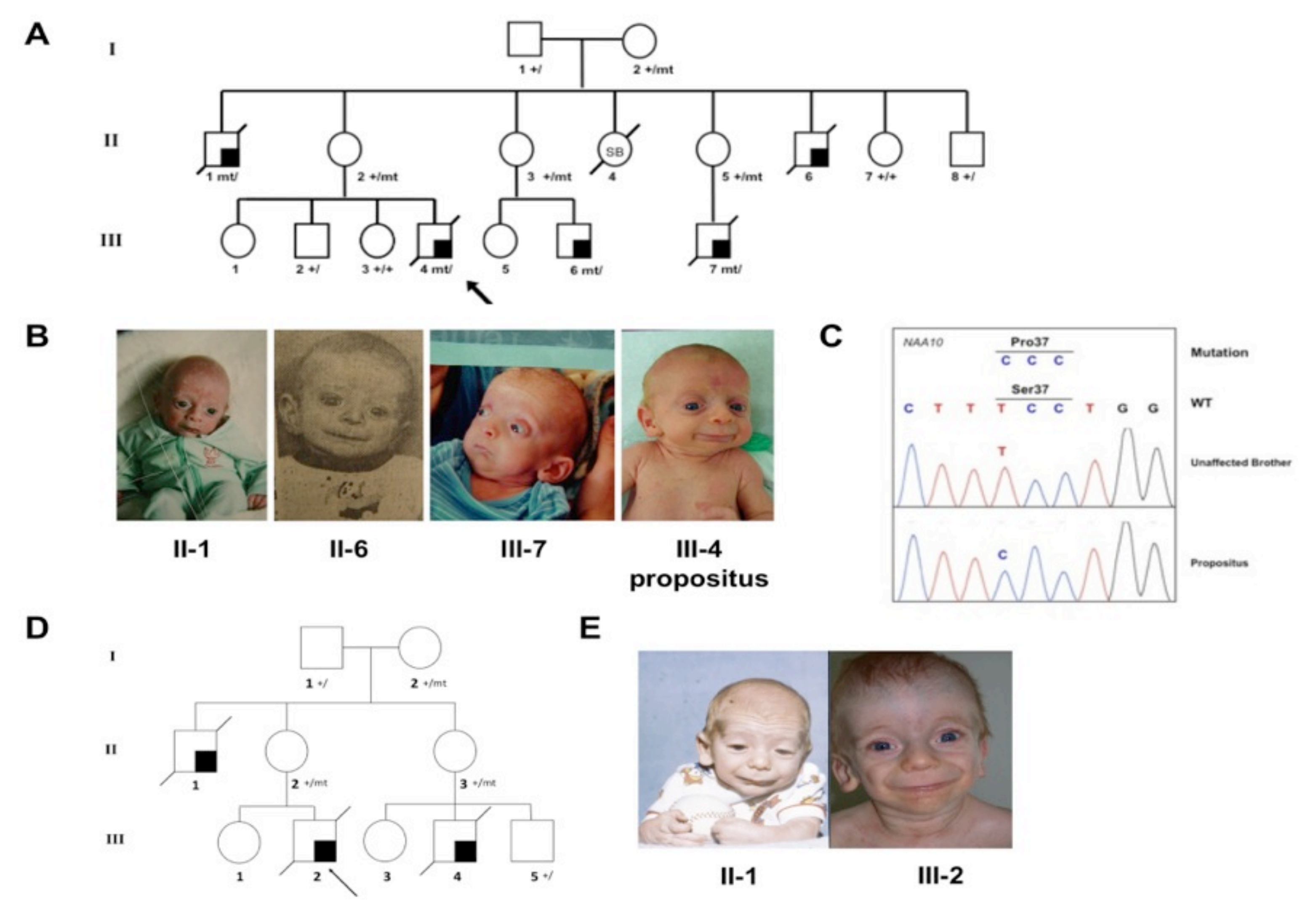
1 National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, United States of America, **2** Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York, United States of America, **3** Animal Health Trust, Center for Preventive Medicine, Newmarket, United Kingdom

PLoS Genetics | www.plosgenetics.org

May 2007 | Volume 3 | Issue 5 | e79



Ogden Syndrome



We found the SAME mutation in two unrelated families, with a very similar phenotype in both families, helping prove that this genotype contributes to the phenotype observed.

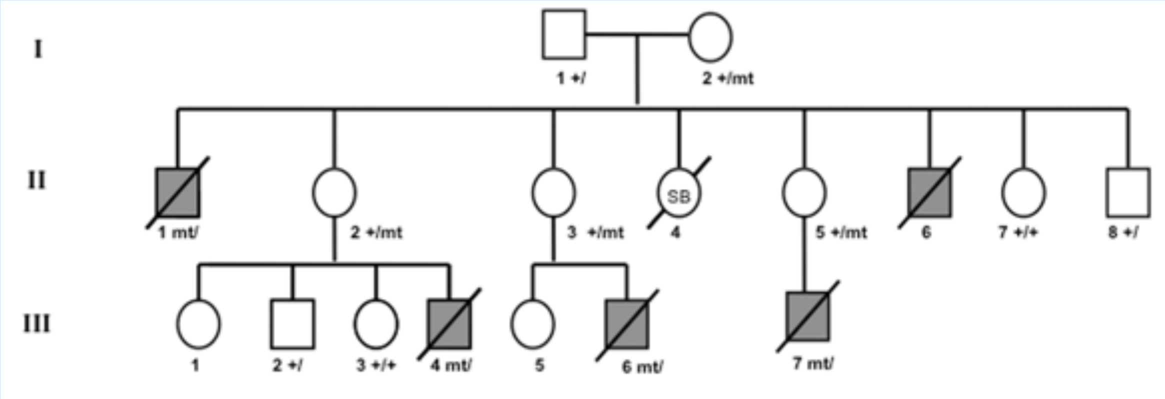
This is the first boy in the late 1970’s.

First boy. Called “a little old man” by the family. Died around ~1 year of age, from cardiac arrhythmias.

This is the “Proband” photograph presented at Case Conference.

prominence of eyes, down-sloping palpebral fissures, thickened eyelids, large ears, beaking of nose, flared nares, hypoplastic nasal alae, short columella, protruding upper lip, micro-retrognathia

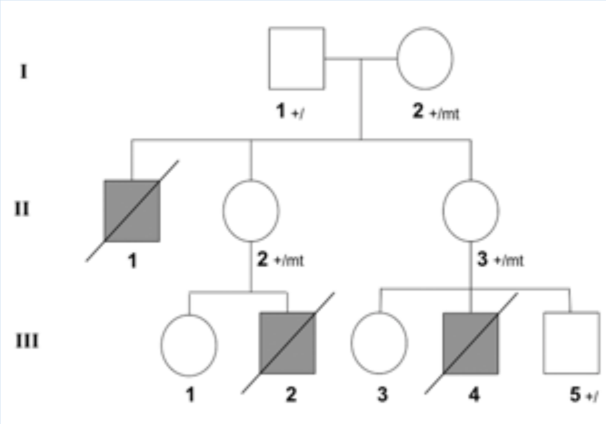
A



B

II-1 II-6 III-4 III-6 III-7

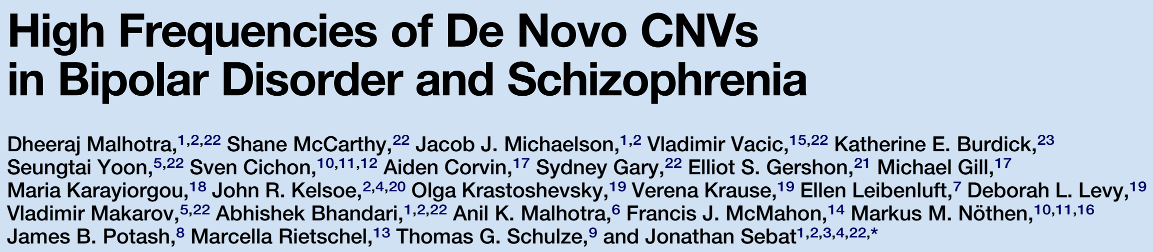
C



D

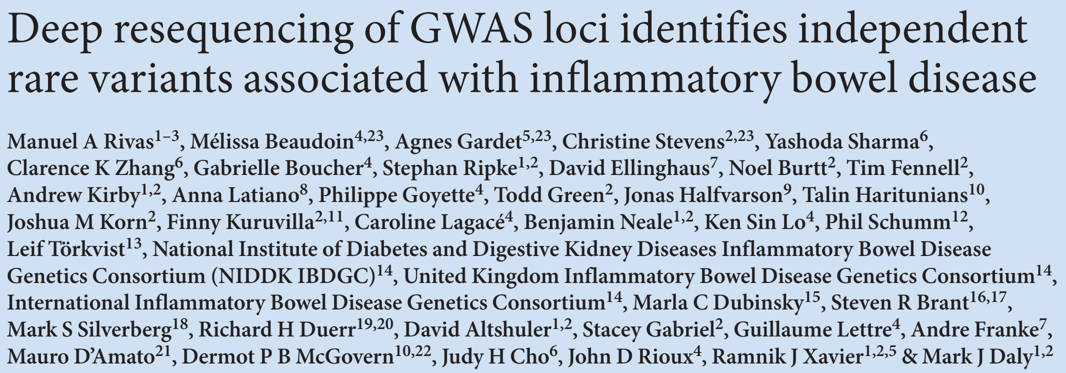
II-1 III-2

Rare Variants – CNVs, SNVs, indels, etc… in Rare AND Common diseases



High Frequencies of De Novo CNVs in Bipolar Disorder and Schizophrenia

Dheeraj Malhotra,[1,2,22] Shane McCarthy,[22] Jacob J. Michaelson,[1,2] Vladimir Vacic,[15,22] Katherine E. Burdick,[23]
Seungtai Yoon,[5,22] Sven Cichon,[10,11,12] Aiden Corvin,[17] Sydney Gary,[22] Elliot S. Gershon,[21] Michael Gill,[17]
Maria Karayiorgou,[18] John R. Kelsoe,[2,4,20] Olga Krastoshevsky,[19] Verena Krause,[19] Ellen Leibenluft,[7] Deborah L. Levy,[19]
Vladimir Makarov,[5,22] Abhishek Bhandari,[1,2,22] Anil K. Malhotra,[6] Francis J. McMahon,[14] Markus M. Nöthen,[10,11,16]
James B. Potash,[8] Marcella Rietschel,[13] Thomas G. Schulze,[9] and Jonathan Sebat[1,2,3,4,22,*]



Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease

Manuel A Rivas[1–3], Mélissa Beaudoin[4,23], Agnes Gardet[5,23], Christine Stevens[2,23], Yashoda Sharma[6],
Clarence K Zhang[6], Gabrielle Boucher[4], Stephan Ripke[1,2], David Ellinghaus[7], Noel Burtt[2], Tim Fennell[2],
Andrew Kirby[1,2], Anna Latiano[8], Philippe Goyette[4], Todd Green[2], Jonas Halfvarson[9], Talin Haritunians[10],
Joshua M Korn[2], Finny Kuruvilla[2,11], Caroline Lagacé[4], Benjamin Neale[1,2], Ken Sin Lo[4], Phil Schumm[12],
Leif Törkvist[13], National Institute of Diabetes and Digestive Kidney Diseases Inflammatory Bowel Disease
Genetics Consortium (NIDDK IBDGC)[14], United Kingdom Inflammatory Bowel Disease Genetics Consortium[14],
International Inflammatory Bowel Disease Genetics Consortium[14], Marla C Dubinsky[15], Steven R Brant[16,17],
Mark S Silverberg[18], Richard H Duerr[19,20], David Altshuler[1,2], Stacey Gabriel[2], Guillaume Lettre[4], Andre Franke[7],
Mauro D'Amato[21], Dermot P B McGovern[10,22], Judy H Cho[6], John D Rioux[4], Ramnik J Xavier[1,2,5] & Mark J Daly[1,2]

Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes

Jacob A. Tennessen,[1]* Abigail W. Bigham,[2]*† Timothy D. O'Connor,[1]* Wenqing Fu,[1] Eimear E. Kenny,[3] Simon Gravel,[3] Sean McGee,[1] Ron Do,[4,5] Xiaoming Liu,[6] Goo Jun,[7] Hyun Min Kang,[7] Daniel Jordan,[8] Suzanne M. Leal,[9] Stacey Gabriel,[4] Mark J. Rieder,[1] Goncalo Abecasis,[7] David Altshuler,[4] Deborah A. Nickerson,[1] Eric Boerwinkle,[6,10] Shamil Sunyaev,[4,8] Carlos D. Bustamante,[3] Michael J. Bamshad,[1,2]‡ Joshua M. Akey,[1]‡ Broad GO, Seattle GO, on behalf of the NHLBI Exome Sequencing Project

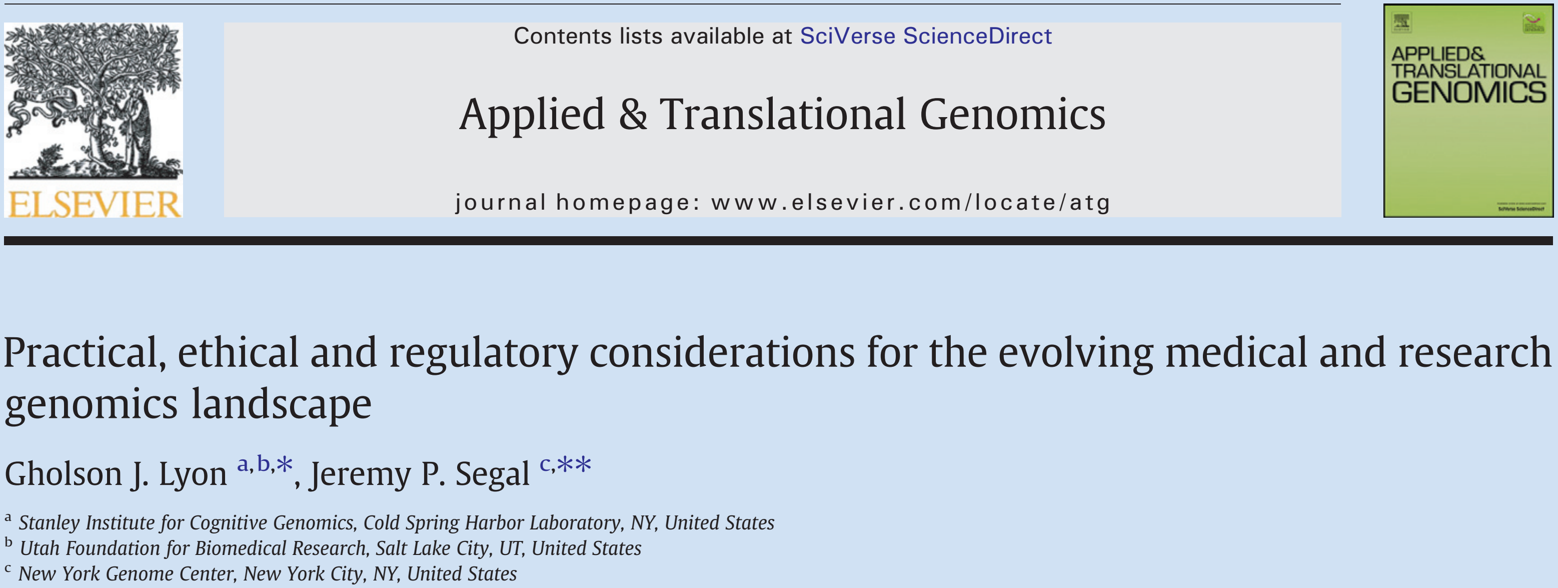
Lyon and Wang *Genome Medicine* 2012, 4:58
http://genomemedicine.com/content/4/7/58

Genome **Medicine**

REVIEW

Identifying disease mutations in genomic medicine settings: current challenges and how to accelerate progress

Gholson J Lyon[*1,2] and Kai Wang[*2,3]



Contents lists available at SciVerse ScienceDirect

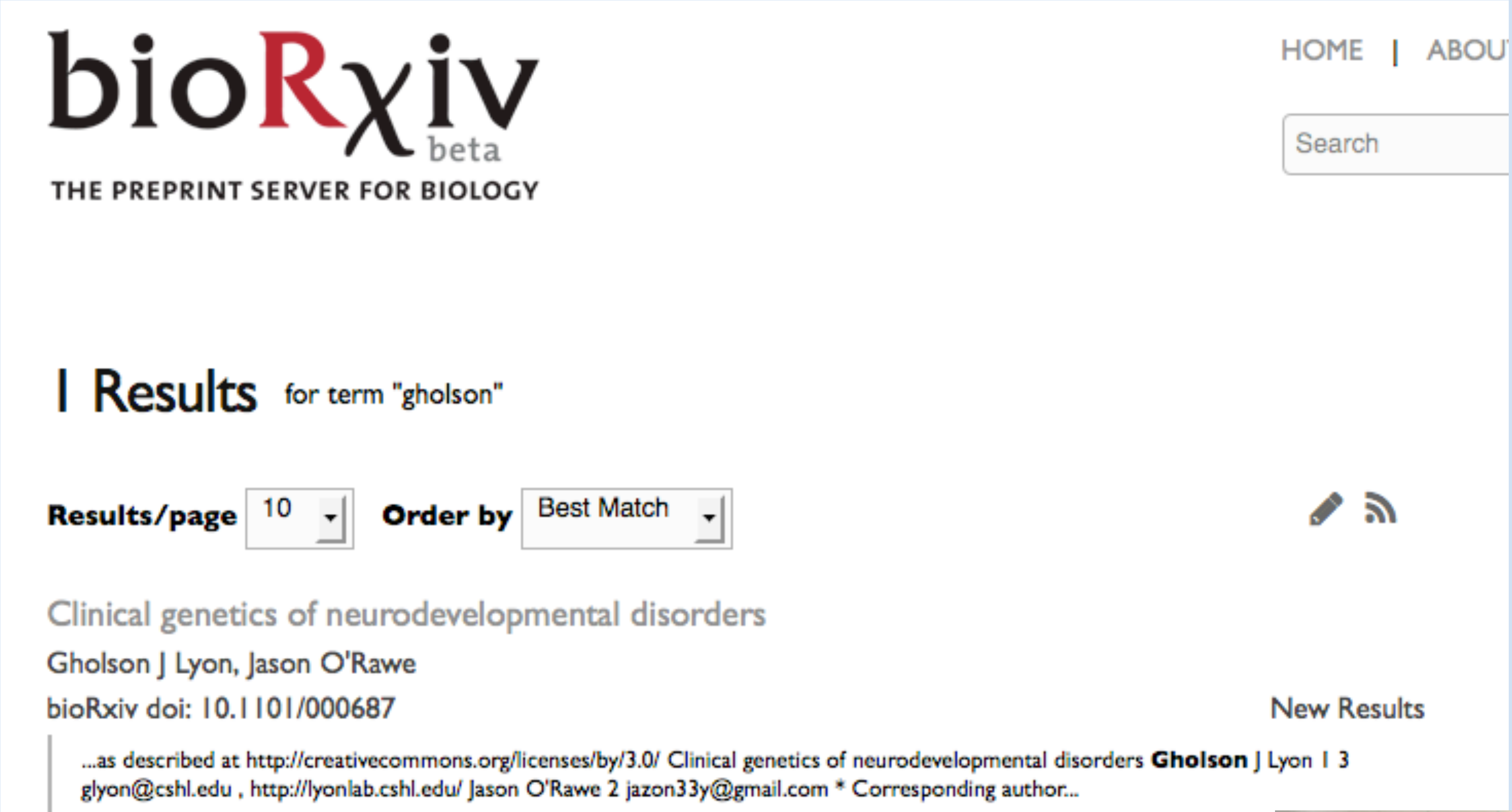
Applied & Translational Genomics

journal homepage: www.elsevier.com/locate/atg

Practical, ethical and regulatory considerations for the evolving medical and research genomics landscape

Gholson J. Lyon [a,b,*], Jeremy P. Segal [c,**]

[a] *Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, NY, United States*
[b] *Utah Foundation for Biomedical Research, Salt Lake City, UT, United States*
[c] *New York Genome Center, New York City, NY, United States*



1 Results for term "gholson"

Results/page 10 **Order by** Best Match

Clinical genetics of neurodevelopmental disorders

Gholson J Lyon, Jason O'Rawe

bioRxiv doi: 10.1101/000687

New Results

> ...as described at http://creativecommons.org/licenses/by/3.0/ Clinical genetics of neurodevelopmental disorders **Gholson** J Lyon 1 3 glyon@cshl.edu , http://lyonlab.cshl.edu/ Jason O'Rawe 2 jazon33y@gmail.com * Corresponding author...

- Seguin E. 1866, - “our incomplete studies do not permit actual classification; but it is better to leave things by themselves rather than to force them into classes which have their foundation only on paper”- from Idiocy and its treatment by the physiological method.

- Walter Frank Raphael Weldon 1902 – “the accumulation of records, in which results are massed together in ill-defined categories of variable and uncertain extent, can only result in harm”.

OBSERVATIONS ON AN ETHNIC CLASSIFICATION OF IDIOTS *

J. LANGDON H. DOWN M.D., *London*

London Hospital Clinical Lecture Report. 3, 259-262, 1866.

“Those who have given any attention to congenital mental lesions, must have been frequently puzzled how to arrange, in any satisfactory way, the different classes of this defect which may have come under their observation. Nor will the difficulty be lessened by an appeal to what has been written on the subject. The systems of classification are generally so vague and artificial, that, not only do they assist but feebly, in any mental arrangement of the phenomena represented, but they completely fail in exerting any practical influence on the subject.”

The Biology of
MENTAL DEFECT

BY

LIONEL S. PENROSE, M.A., M.D.

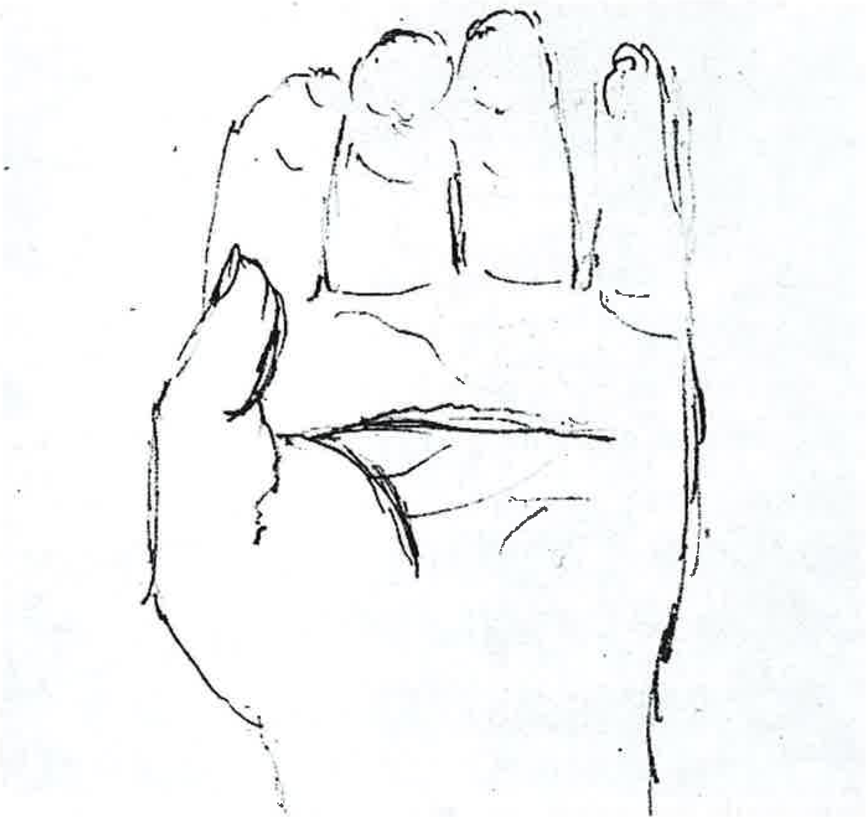
WITH A PREFACE BY

PROFESSOR J. B. S. HALDANE, F.R.S.

GRUNE & STRATTON
New York
1949

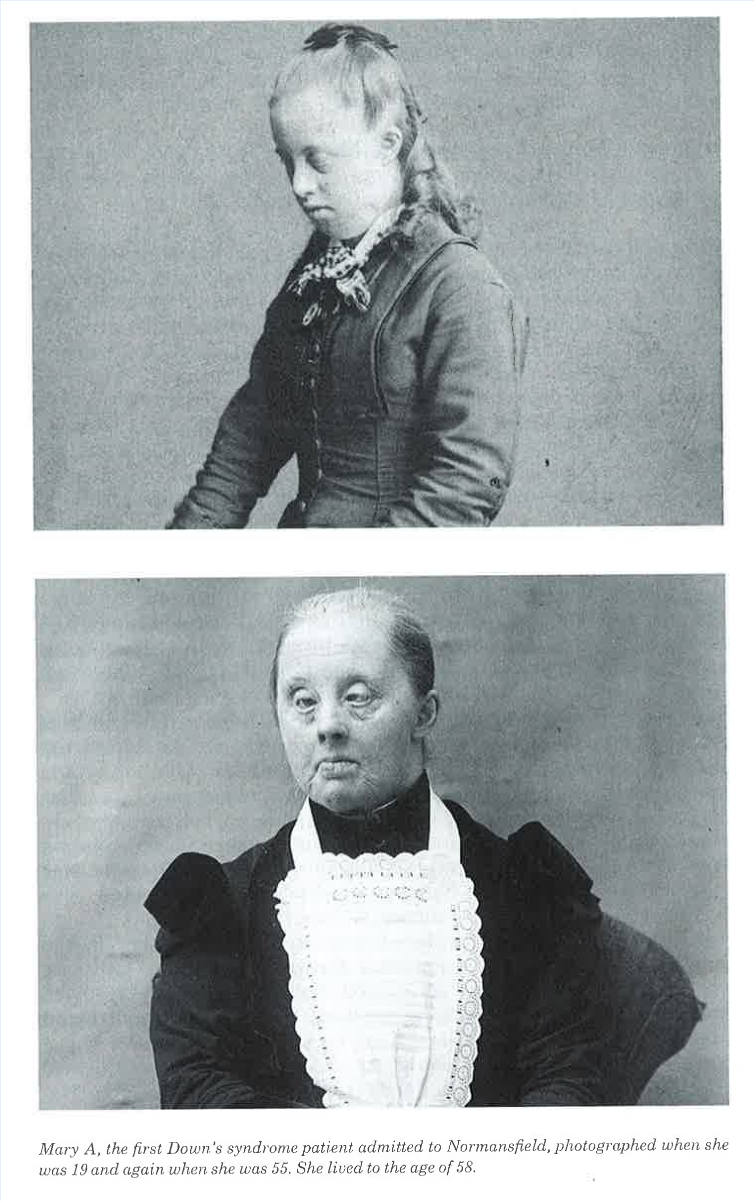
Plate VII—Mongolism in two imbecile brothers aged 10 (Colchester Survey, 1938, Case No. 750) and 5 years, with a normal child aged 2½ years.

As compared with the normal child, the younger mongoloid is seen to have a small head, decreased stature and dysplastic features. The characteristic fold of skin covering the inner canthus of each eye (epicanthic fold) was clearly marked in this case.



Reginald Langdon Down was the first to describe the pattern of creases in the palm in Down's syndrome patients. He drew this sketch in 1908.

Published in “Biology of Mental Defect”, by Lionel Penrose, 1949
And “John Langdon Down: A Caring Pioneer”, by O Conor Ward, 1998.



Mary A, the first Down's syndrome patient admitted to Normansfield, photographed when she was 19 and again when she was 55. She lived to the age of 58.

Florence T, a Down's syndrome patient at Normansfield. Photographed in 1886 when she was seven and again in 1899 aged 20.

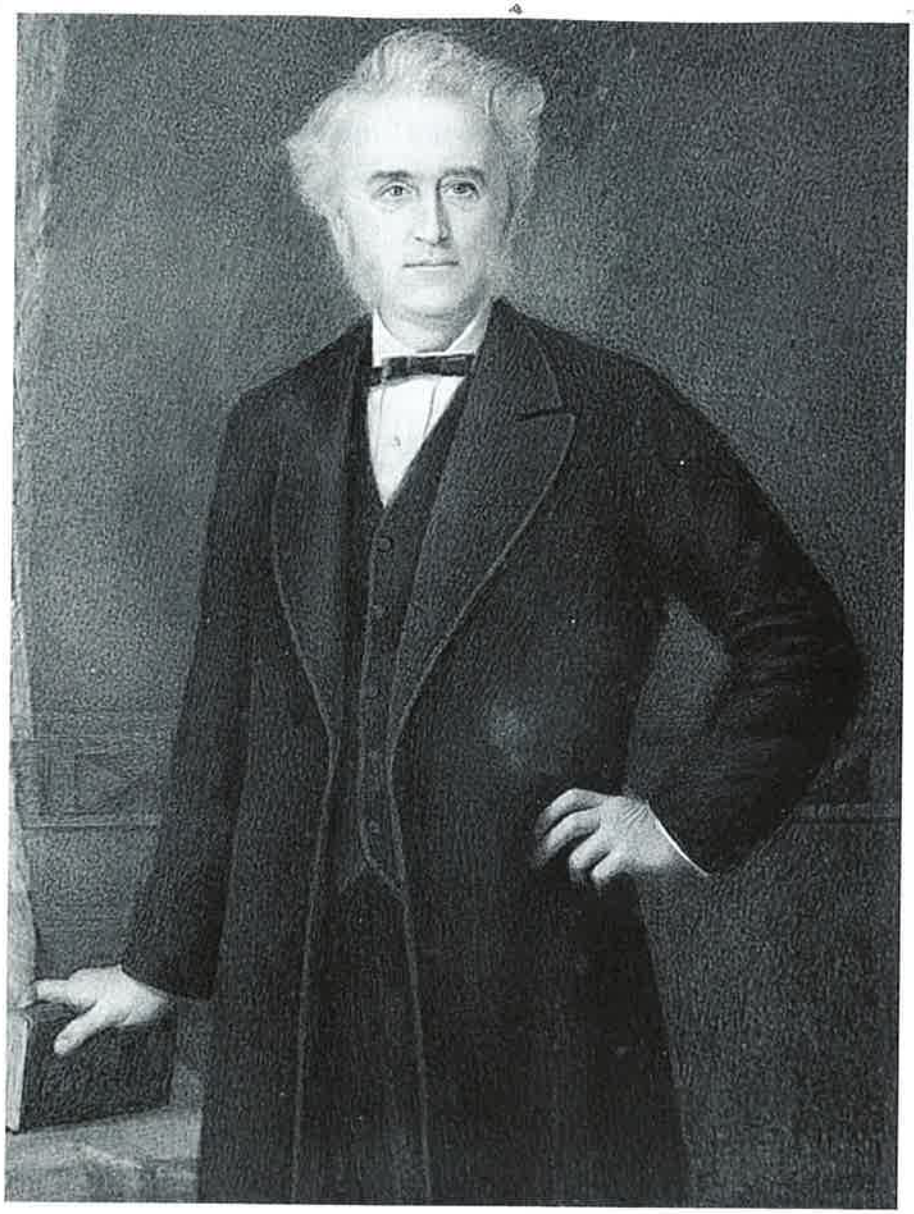
Published in “John Langdon Down: A Caring Pioneer”, by O Conor Ward, 1998.

Langdon Down began to take clinical photographs in 1862. His first photograph of an Earlswood resident with Down's syndrome was this unnamed girl in the 1865 series. She was probably the first ever Down's syndrome patient to be photographed.

Published in "John Langdon Down: A Caring Pioneer", by O Conor Ward, 1998.

Four Down's syndrome patients. Part of the Earlswood series, photographed in 1865.

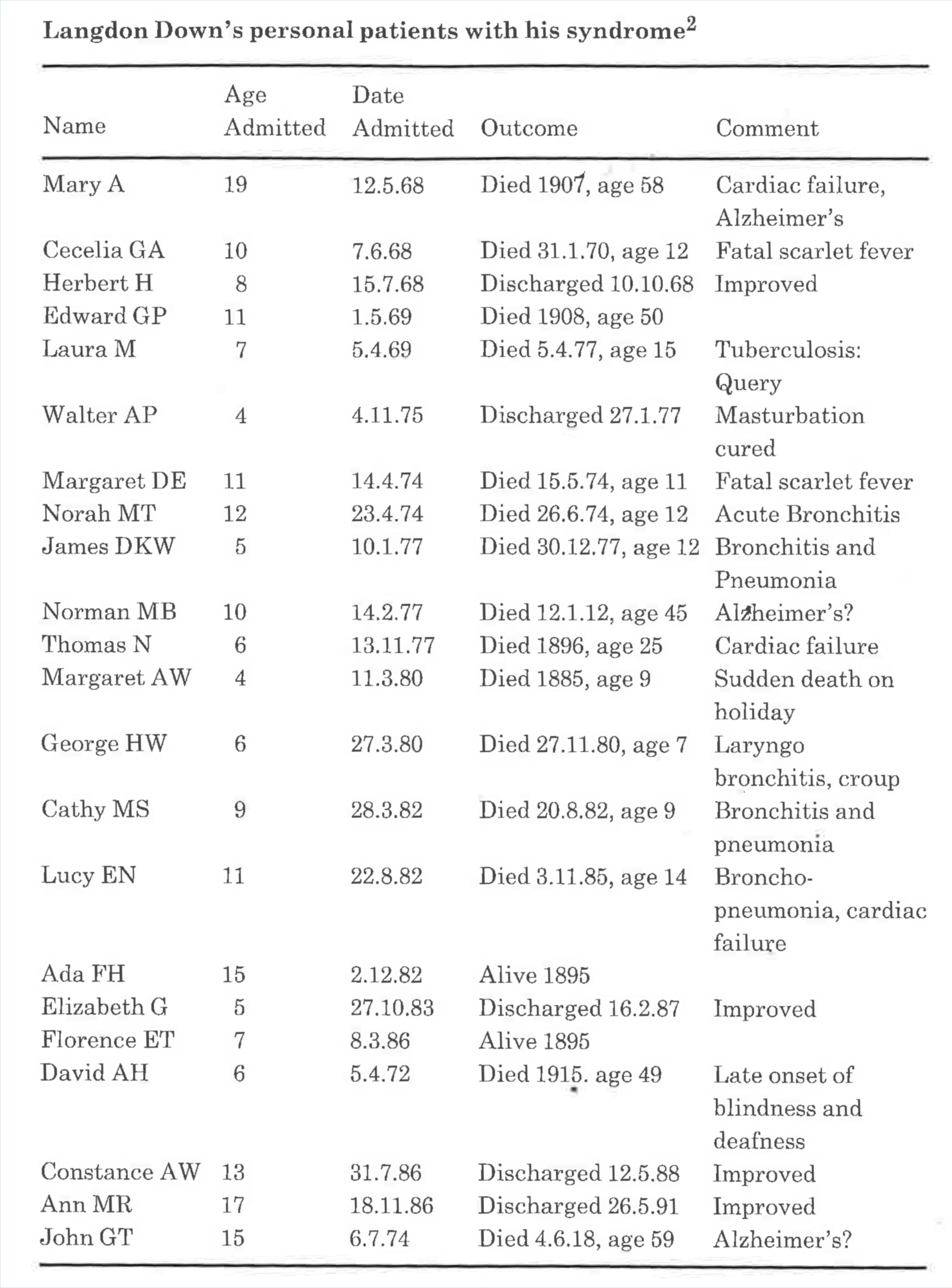
Langdon Down in court dress, 1887, when he gave the welcome address to the Prince and Princess of Wales at the opening of the London Hospital Nursing School.



Portrait of Langdon Down, painted by Sydney Hodges in 1883.

Dr Reginald Langdon Down with his daughters Stella and Elspie. Stella married Russell Brain and became Lady Brain. Elspie was an artist. The only son was John, who had Down's syndrome.

Dr Percival Langdon Down with his wife and children. His son Norman, was to be the last Langdon Down superintendent of Normansfield, ending a family connection that had lasted for 102 years. The elder daughter, Molly, was also a doctor and worked in Normansfield.



Langdon Down's personal patients with his syndrome[2]

Name	Age Admitted	Date Admitted	Outcome	Comment
Mary A	19	12.5.68	Died 1907, age 58	Cardiac failure, Alzheimer's
Cecelia GA	10	7.6.68	Died 31.1.70, age 12	Fatal scarlet fever
Herbert H	8	15.7.68	Discharged 10.10.68	Improved
Edward GP	11	1.5.69	Died 1908, age 50	
Laura M	7	5.4.69	Died 5.4.77, age 15	Tuberculosis: Query
Walter AP	4	4.11.75	Discharged 27.1.77	Masturbation cured
Margaret DE	11	14.4.74	Died 15.5.74, age 11	Fatal scarlet fever
Norah MT	12	23.4.74	Died 26.6.74, age 12	Acute Bronchitis
James DKW	5	10.1.77	Died 30.12.77, age 12	Bronchitis and Pneumonia
Norman MB	10	14.2.77	Died 12.1.12, age 45	Alzheimer's?
Thomas N	6	13.11.77	Died 1896, age 25	Cardiac failure
Margaret AW	4	11.3.80	Died 1885, age 9	Sudden death on holiday
George HW	6	27.3.80	Died 27.11.80, age 7	Laryngo bronchitis, croup
Cathy MS	9	28.3.82	Died 20.8.82, age 9	Bronchitis and pneumonia
Lucy EN	11	22.8.82	Died 3.11.85, age 14	Broncho-pneumonia, cardiac failure
Ada FH	15	2.12.82	Alive 1895	
Elizabeth G	5	27.10.83	Discharged 16.2.87	Improved
Florence ET	7	8.3.86	Alive 1895	
David AH	6	5.4.72	Died 1915. age 49	Late onset of blindness and deafness
Constance AW	13	31.7.86	Discharged 12.5.88	Improved
Ann MR	17	18.11.86	Discharged 26.5.91	Improved
John GT	15	6.7.74	Died 4.6.18, age 59	Alzheimer's?

A Genotype-First Approach to Defining the Subtypes of a Complex Disease

Holly A. Stessman,[1] Raphael Bernier,[2] and Evan E. Eichler[1,3,*]
[1]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA
[2]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA 98195, USA
[3]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA
*Correspondence: eee@gs.washington.edu
http://dx.doi.org/10.1016/j.cell.2014.02.002

Medical genetics typically entails the detailed characterization of a patient's phenotypes followed by genotyping to discover the responsible gene or mutation. Here, we propose that the systematic discovery of genetic variants associated with complex diseases such as autism are progressing to a point where a reverse strategy may be fruitful in assigning the pathogenic effects of many different genes and in determining whether particular genotypes manifest as clinically recognizable phenotypes. This "genotype-first" approach for complex disease necessitates the development of large, highly integrated networks of researchers, clinicians, and patient families, with the promise of improved therapies for subsets of patients.

Prioritization of neurodevelopmental disease genes by discovery of new mutations

Alexander Hoischen[1], Niklas Krumm[2] & Evan E Eichler[2,3]

Advances in genome sequencing technologies have begun to revolutionize neurogenetics, allowing the full spectrum of genetic variation to be better understood in relation to disease. Exome sequencing of hundreds to thousands of samples from patients with autism spectrum disorder, intellectual disability, epilepsy and schizophrenia provides strong evidence of the importance of *de novo* and gene-disruptive events. There are now several hundred new candidate genes and targeted resequencing technologies that allow screening of dozens of genes in tens of thousands of individuals with high specificity and sensitivity. The decision of which genes to pursue depends on many factors, including recurrence, previous evidence of overlap with pathogenic copy number variants, the position of the mutation in the protein, the mutational burden among healthy individuals and membership of the candidate gene in disease-implicated protein networks. We discuss these emerging criteria for gene prioritization and the potential impact on the field of neuroscience.

Table 4 Recurrent identical *de novo* mutations in 6 genes identified in 11 exome studies with different neurodevelopmental phenotypes

Gene	Coding effect	Mutation (genomic DNA level)	Mutation (cDNA level)	Mutation (protein level)	Study	Disorder
ALG13	Missense	ChrX(GRCh37):g.110928268A>G	NM_001099922.2:c.320A>G	p.Asn107Ser	de Ligt *et al.*[1]	ID
ALG13	Missense	ChrX(GRCh37):g.110928268A>G	NM_001099922.2:c.320A>G	p.Asn107Ser	Allen *et al.*[11]	EE
ALG13	Missense	ChrX(GRCh37):g.110928268A>G	NM_001099922.2:c.320A>G	p.Asn107Ser	Allen *et al.*[11]	EE
KCNQ3	Missense	Chr8(GRCh37):g.133192493G>A	NM_001204824.1:c.328C>T	p.Arg110Cys	Rauch *et al.*[2]	ID
KCNQ3	Missense	Chr8(GRCh37):g.133192493G>A	NM_001204824.1:c.328C>T	p.Arg110Cys	Allen *et al.*[11]	EE
SCN1A	Splice donor	LRG_8:g.24003G>A	NM_006920.4:c.602+1G>A	p.?	Allen *et al.*[11]	EE
SCN1A	Splice donor	LRG_8:g.24003G>A	NM_006920.4:c.602+1G>A	p.?	Allen *et al.*[11]	EE
CUX2	Missense	Chr12(GRCh37):g.111748354G>A	NM_015267.3:c.1768G>A	p.Glu590Lys	Rauch *et al.*[2]	ID
CUX2	Missense	Chr12(GRCh37):g.111748354G>A	NM_015267.3:c.1768G>A	p.Glu590Lys	Allen *et al.*[11]	EE
SCN2A	Missense	Chr2(GRCh37):g.166198975G>A	NM_021007.2:c.2558G>A	p.Arg853Gln	Allen *et al.*[11]	EE
SCN2A	Missense	Chr2(GRCh37):g.166198975G>A	NM_021007.2:c.2558G>A	p.Arg853Gln	Allen *et al.*[11]	EE
DUSP15	Missense	Chr20(GRCh37):g.30450489G>A	NM_080611.2:c.320C>T	p.Thr107Met	Neale *et al.*[7]	ASD
DUSP15	Missense	Chr20(GRCh37):g.30450489G>A	NM_080611.2:c.320C>T	p.Thr107Met	Fromer *et al.*[10]	SCZ

EE, epileptic encephalopathies; ASD, autism spectrum disorder; ID, intellectual disability; SCZ, schizophrenia.

Mutations as “Difference Makers”

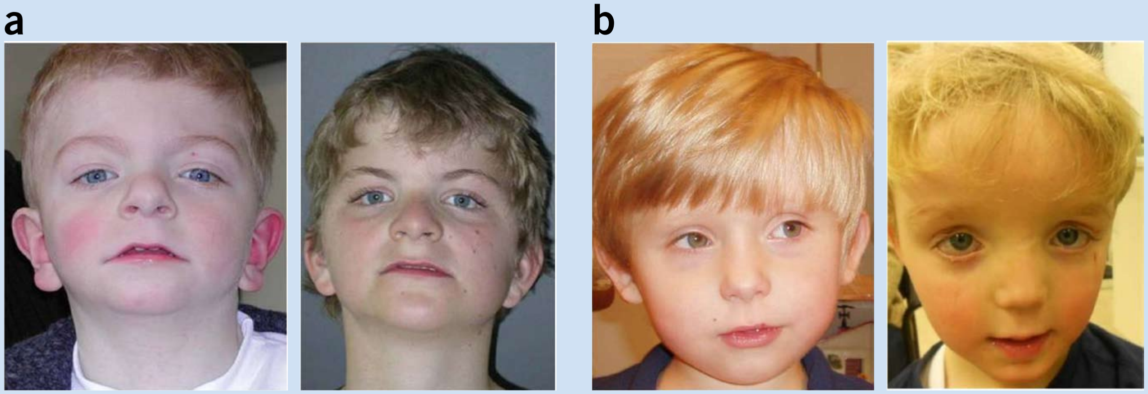


Figure 3 Phenotypic similarity of two patients with identical *PACS1 de novo* mutations and two patients with similar *ADNP* mutations. (**a**) These two unrelated patients show identical *de novo* point mutations (c.607C>T; p.Arg203Trp) in *PACS1* (RefSeq NM_018026.3)[53]. The striking similarity in phenotype includes low anterior hairline, highly arched eyebrows, synophrys, hypertelorism with downslanted palpebral fissures, long eyelashes, a bulbous nasal tip, a flat philtrum with a thin upper lip, downturned corners of the mouth and low-set ears. Reprinted from ref. 53, Copyright (2012), with permission from The American Society of Human Genetics. (**b**) These two unrelated patients both show LoF mutations in *ADNP* (c.2496_2499delTAAA; p.Asp832Lysfs*80 and c.2157C>G; p.Tyr719*)[44] resulting in a new SWI-SNF–related autism syndrome. Patients present with clinical similarities, including a prominent forehead, a thin upper lip and a broad nasal bridge. Reprinted from ref. 44.

Refining analyses of copy number variation identifies specific genes associated with developmental delay

Bradley P Coe[1], Kali Witherspoon[1], Jill A Rosenfeld[2], Bregje W M van Bon[3,4], Anneke T Vulto-van Silfhout[3], Paolo Bosco[5], Kathryn L Friend[4], Carl Baker[1], Serafino Buono[5], Lisenka E L M Vissers[3], Janneke H Schuurs-Hoeijmakers[3], Alex Hoischen[3], Rolph Pfundt[3], Nik Krumm[1], Gemma L Carvill[6], Deana Li[7], David Amaral[7], Natasha Brown[8,9], Paul J Lockhart[8,10], Ingrid E Scheffer[11], Antonino Alberti[5], Marie Shaw[4], Rosa Pettinato[5], Raymond Tervo[12], Nicole de Leeuw[3], Margot R F Reijnders[3], Beth S Torchia[2], Hilde Peeters[13,14], Elizabeth Thompson[4,15], Brian J O'Roak[1,18], Marco Fichera[5,18], Jayne Y Hehir-Kwa[3], Jay Shendure[1], Heather C Mefford[6], Eric Haan[4,15], Jozef Gécz[4,16], Bert B A de Vries[3], Corrado Romano[5] & Evan E Eichler[1,17]

Copy number variants (CNVs) are associated with many neurocognitive disorders; however, these events are typically large, and the underlying causative genes are unclear. We created an expanded CNV morbidity map from 29,085 children with developmental delay in comparison to 19,584 healthy controls, identifying 70 significant CNVs. We resequenced 26 candidate genes in 4,716 additional cases with developmental delay or autism and 2,193 controls. An integrated analysis of CNV and single-nucleotide variant (SNV) data pinpointed 10 genes enriched for putative loss of function. Follow-up of a subset of affected individuals identified new clinical subtypes of pediatric disease and the genes responsible for disease-associated CNVs. These genetic changes include haploinsufficiency of *SETBP1* associated with intellectual disability and loss of expressive language and truncations of *ZMYND11* in individuals with autism, aggression and complex neuropsychiatric features. This combined CNV and SNV approach facilitates the rapid discovery of new syndromes and genes involved in neuropsychiatric disease despite extensive genetic heterogeneity.

Table 2 Intersection of CNV and exome data

		Exome data			Array CGH			
Gene	Isoform	1,879 published cases LoF	1,879 published cases *de novo* LoF (ESP average read depth >20, Dustmasked)	6,500 ESP LoF (ESP average read depth >20, Dustmasked)	Signature deletions (*n* = 29,085)	Control deletions (*n* = 19,584)	Combined LoF *P* value	Combined LoF *q* value[e]
ANK2[a]	NM_020977.3[b]	1	1	0	5	0	0.0171	0.169
ARHGAP5	NM_001030055.1	1	1	0	7	0	0.0061	0.0833
BCL11A	NM_022893.3	1	0	0	4	0	0.0286	0.244
CAPRIN1	NM_005898.4	1	1	0	4	0	0.0286	0.244
CARKD	NM_001242881.1[c]	1	1	0	12	4	0.0363	0.28
CHD2[a]	NM_001271.3	3	3	0	0	0	0.0113	0.127
CHD8[a]	NM_001170629.1	3	3	0	2	0	0.00402	0.0703
CSDE1	NM_001130523.2	1	1	0	3	0	0.0479	0.311
CUL3[a]	NM_003590.4	2	2	0	5	0	0.00383	0.0703
DLL1	NM_005618.3	1	0	0	32	1	2.17×10^{-7}	2.68×10^{-5}
DYRK1A[a]	NM_001396.3	2	2	0	11	0	1.74×10^{-4}	8.60×10^{-3}
FAM8A1	NM_016255.2	1	1	0	5	0	0.0171	0.169
FOXP1[a]	NM_001244810.1	1	1	0	4	0	0.0286	0.244
GRIN2B[a]	NM_000834.3	3	3	0	2	0	0.00402	0.0703
GTPBP4	NM_012341.2	1	1	0	3	0	0.0479	0.311
LTN1	NM_015565.2	1	1	0	6	0	0.0102	0.12
MBD5[a]	NM_018328.4	1	1	0	16	6	0.0343	0.273
MYT1L	NM_015025.2	1	1	0	8	0	0.00365	0.0703
NAA15	NM_057175.3	2	2	0	5	3	0.0296	0.244
NCKAP1	NM_205842.1	2	2	0	7	0	0.00137	0.0564
NFIA	NM_001134673.3	1	1	0	3	0	0.0479	0.311
NRXN1[a]	NM_001135659.1	1	1	0	30	9	0.00427	0.0703
NTM	NM_001144058.1	1	1	0	40	0	2.53×10^{-10}	6.25×10^{-8}
PCOLCE	NM_002593.3	1	1	0	7	0	0.0061	0.0833
PHF2	NM_005392.3	1	1	0	4	0	0.0286	0.244
RAB2A	NM_002865.2	1	1	0	3	0	0.0479	0.311
SCN1A[a]	NM_001165963.1	4	4	0	10	1	7.36×10^{-5}	4.55×10^{-3}
SCN2A[a]	NM_021007.2	6	5	0	10	0	7.34×10^{-7}	6.04×10^{-5}
SLC6A1	NM_003042.3	1	1	0	6	0	0.0102	0.12
SRM	NM_003132.2	1	1	0	9	0	0.00218	0.0703
STXBP1[a]	NM_003165.3	2	2	0	4	0	0.00641	0.0833
SUV420H1	NM_016028.4[d]	1	1	0	3	0	0.0479	0.31135
SYNGAP1[a]	NM_006772.2	4	4	0	0	1	0.00252	0.0703
TBR1	NM_006593.2	2	2	0	7	1	0.00522	0.0806
UBN2	NM_173569.3	1	1	0	5	0	0.0171	0.169
WAC	NM_016628.4	1	1	0	3	0	0.0479	0.31135
WDFY3	NM_014991.4	1	1	0	8	0	0.00365	0.0703
ZMYND11	NM_006624.5	1	1	0	8	0	0.00365	0.0703

LOF, loss of function.

[a] Disease gene in OMIM. [b] Variant 2; this is the major form of ankyrin in the adult brain. [c] Variant 2; this isoform and variants 3 and 4 are shorter than variant 1. [d] Variant 2; this isoform is shorter and has a distinct C terminus in comparison to isoform 1. [e] Please see the **Supplementary Note** for discussion of the *q* values shown.

Table 4 Brief phenotypic description of cases with *SETBP1* loss-of-function variants

Case	Age at examination	Sex	Alteration	Inheritance	Cognitive	Hyperactive or ADHD	Social difficulties	Other behavioral difficulties	Speech delay	Motor delay	Facial dysmorphism	Seizures or EEG abnormalities
DNA03-00335	14 years	M	p.Ile822Tyrfs*13	*De novo*	Normal IQ			+	+	+	+	
DNA-008897	73 years	M	p.Leu411Glyfs*6		Profound ID		+	+	+	+	+	
Troina 1274	19 years	M	p.Trp532*	*De novo*	Severe ID			+	+	+	+	–
Troina 1512	17 years	M	p.Ser1011*	*De novo*	Mild ID	+ (3y 8m)	+		+	+	+	–
Troina 3097	34 years	F	p.Arg143Valfs*64		Severe ID				+	+	+	+
DNA11-21308Z	36 years	F	p.Arg625*		Mild to moderate ID	+	+	+	+	+	+	
DNA11-19324Z	9 years	F	p.Arg626*		2- to 2.5-year delay at 9 years old				+	–	+	–
DNA08-08272	9 years	M	p.Gly15Argfs*47		Mild ID	+		+	+	+	+	+
Rauch *et al.*	13 years	F	p.Lys592*		Mild ID	+	+		+	–	+	
9886269	5 years	M	Deletion	*De novo*	Global delay	+			+	+	+	+
Marseglia *et al.*	15 years	M	Deletion	*De novo*	Mild ID	+	+	+	+	+	+	+
Filges *et al.* pt. 1	7 years	M	Deletion	*De novo*	Moderate ID				+	+	+	+
Filges *et al.* pt. 2	4 years	M	Deletion	*De novo*					+	+	+	

ID, intellectual disability; EEG, electroencephalogram; M, male; F, female.

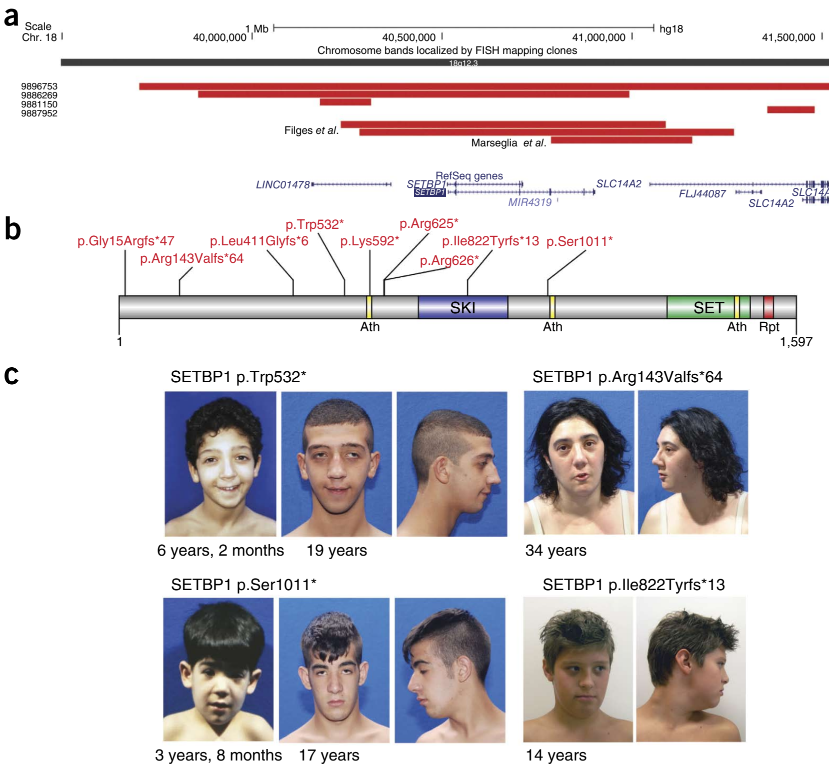


Figure 1 Truncating *SETBP1* mutations and associated phenotypes. (**a**) CNV data define a focal CNV region around *SETBP1*. Combining a focal *de novo* deletion observed in our study (9886269) with CNVs from Filges *et al.*[41] and Marseglia *et al.*[42] (red bars) highlights minimal common regions, including *SETBP1* and *LINC01478*. (**b**) Targeted resequencing identified eight truncating variants in SETBP1 and none in controls. Integration of published exome data identified one additional case and no truncating events in controls. Ath, AT hook; SKI, SKI-homologous region; SET, SET-binding domain; Rpt, repeat. (**c**) Phenotypic assessment (summarized in **Table 4**) identified a recognizable phenotype, including IQ deficits ranging from mild to severe, impaired speech and distinctive facial features. See the **Supplementary Note** for additional photographs of affected individuals and clinical descriptions. We obtained informed consent to publish the photographs.

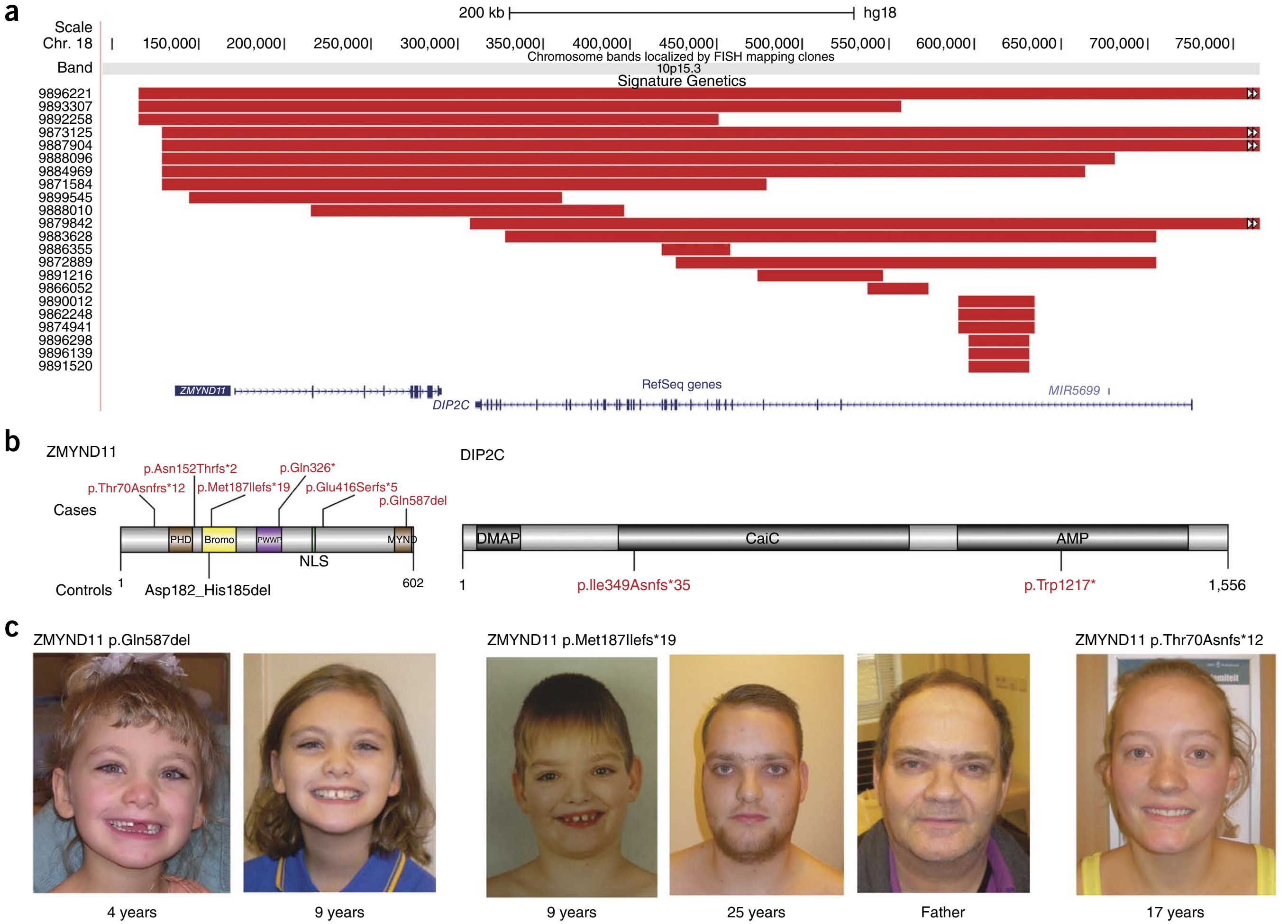


Figure 2 Truncating *ZMYND11* mutations and associated phenotypes. (**a**) CNV data refine a focal CNV deletion region (red bars) containing two genes (*ZMYND11* and *DIP2C*). (**b**) Targeted resequencing identified five truncating variants and one single-amino-acid deletion predicted to behave as loss-of-function variants by removing a critical binding residue in the MYND domain (Gln587). Analysis of control resequencing and exome data identified no additional truncating events in *ZMYND11* but highlighted two truncating mutations in *DIP2C*. PHD, plant homeodomain; Bromo, bromodomain; PWWP, conserved ProTrpTrpPro motif; NLS, nuclear localization sequence; MYND, zinc finger MYND type (myeloid, Nervy and DEAF-1); DMAP, DNA methyltransferase–associated protein; CaiC, crotonobetaine/carnitine-CoA ligase; AMP, AMP-dependent synthetase/ligase. (**c**) Phenotypic assessment (summarized in **Table 5**) showed a consistent phenotype characterized by mild intellectual disability accompanied by speech and motor delays, as well as complex neuropsychiatric behavioral and characteristic facial features. See the **Supplementary Note** for additional photographs of the affected individuals and clinical descriptions. We obtained informed consent to publish the photographs.

Table 5 Brief phenotypic description of cases with *ZMYND11* loss-of-function variants

Case	Age at examination	Sex	Alteration	Inheritance	Cognitive	Speech delay	Social difficulties	Behavioral problems	Facial dysmorphism
Adelaide20124	4 and 9 years	F	p.Gln587del	*De novo*	Global DD	+	+		+
Adelaide3553	22 years	M	p.Asn152Thrfs*26		Global DD	+		+	
DNA-017151	17 years	F	p.Thr70Asnfs*12	*De novo*	Normal IQ	+	+	+	+
DNA04-02424	41 years	M	p.Gln326*		Mild ID	+	+	+	+
DNA05-04370		M	p.Glu416Serfs*5		Severe ID	+	+		+
DNA-013587	25 years	M	p.Met187Ilefs*19	Inherited	Global DD	+	+	+	+
Father of DNA-013587		M	p.Met187Ilefs*19	Carrier	DD			+	

DD, developmental delay; ID, intellectual disability; M, male; F, female.

MICROBIAL EVOLUTION

Global epistasis makes adaptation predictable despite sequence-level stochasticity

Sergey Kryazhimskiy,[1,3]*† Daniel P. Rice,[1,3]* Elizabeth R. Jerison,[2,3] Michael M. Desai[1,2,3]†

Epistatic interactions between mutations can make evolutionary trajectories contingent on the chance occurrence of initial mutations. We used experimental evolution in *Saccharomyces cerevisiae* to quantify this contingency, finding differences in adaptability among 64 closely related genotypes. Despite these differences, sequencing of 104 evolved clones showed that initial genotype did not constrain future mutational trajectories. Instead, reconstructed combinations of mutations revealed a pattern of diminishing-returns epistasis: Beneficial mutations have consistently smaller effects in fitter backgrounds. Taken together, these results show that beneficial mutations affecting a variety of biological processes are globally coupled; they interact strongly, but only through their combined effect on fitness. As a consequence, fitness evolution follows a predictable trajectory even though sequence-level adaptation is stochastic.

SCIENCE sciencemag.org

27 JUNE 2014 • VOL 344 ISSUE 6191

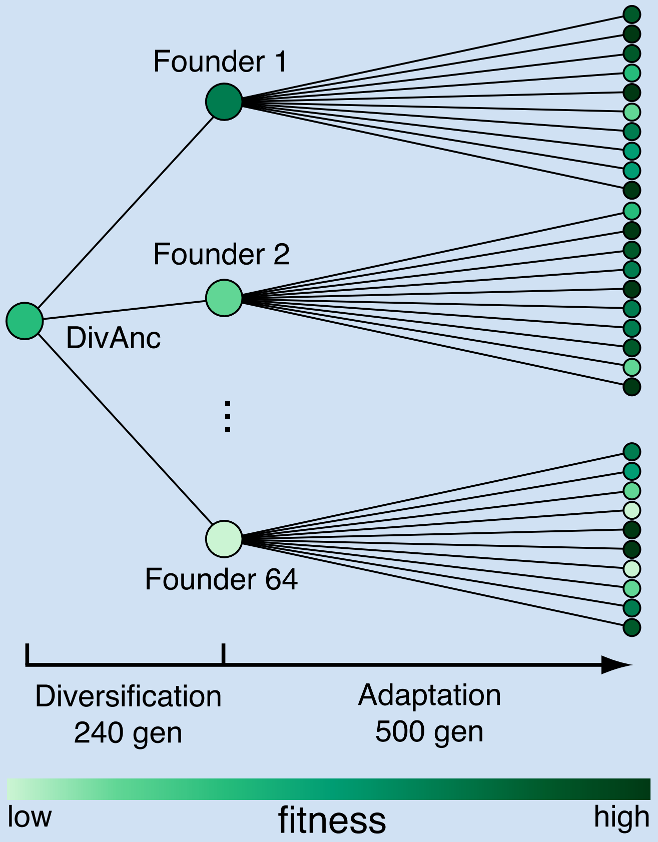


Figure S1. Experimental design. We created many independent lines from a single clone (DivAnc) which came from a previous evolution experiment in the same environment (15) and evolved each of them for 240 generations (Diversification). We then selected a single “Founder” clone from 64 of these lines (chosen to span a range of fitness) and evolved 10 independent replicate populations descended from each Founder for 500 generations (Adaptation).

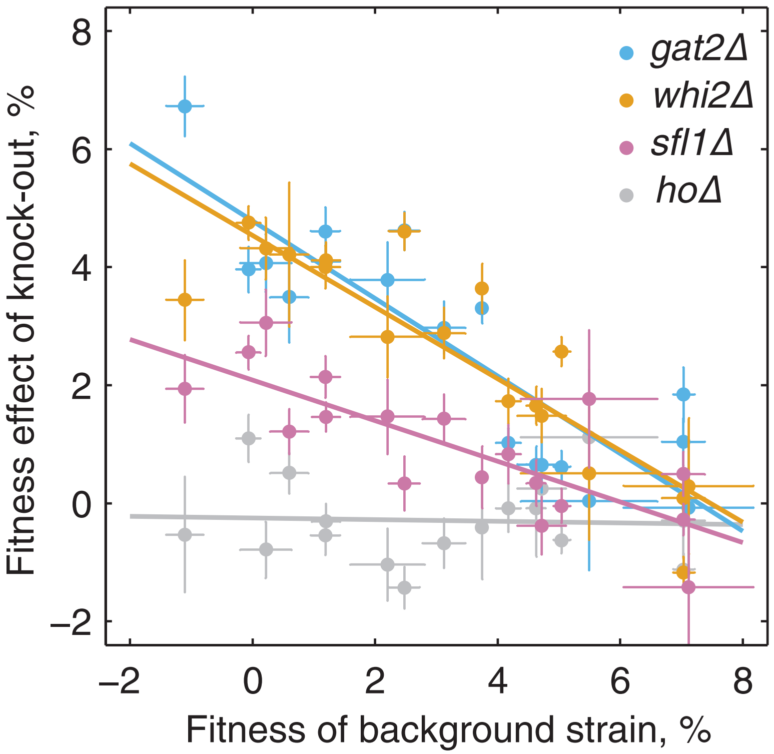


Fig. 3. Diminishing-returns epistasis among specific mutations. The fitness effect of knocking out genes *gat2*, *whi2*, and *sfl1* declines with the fitness of the background strain. The *ho* knockout is a negative control. Error bars are SEM over biological replicates.

- “Yet despite their lack of apparent functional relationship, these mutations are globally coupled by diminishing-returns epistasis; their effects are strongly mediated by background fitness but are otherwise essentially independent of the specific identity of mutations present in the background. The biological basis of this global coupling remains unknown”.