

An introduction to discovering genomic variation... Genome Wide Association Studies (GWAS)

Jeff Barrett



October 19, 2010

Early Mendelian gene mapping: *HD*

Timeline | **Benchmarks in Huntington disease research**

George Huntington's paper is published².

Restriction fragment-length polymorphisms (RFLPs) are first described¹².

The *HD* gene is mapped to the short arm of chromosome 4 (REF. 15).

(1989–1991) Linkage disequilibrium indicates a 2 Mb candidate region^{21–23}.

1872 1888 1900 1908 1978 1981 1983 1987 1989 1991 1993 1994

The Venezeula project is initiated¹⁰.

The *HD* gene is isolated and a CAG repeat mutation is identified²⁶

WFN/IHA, World Federation of Neurology and the International Huntington Association.

Mapping *HD* needed collaboration

Cell, Vol. 72, 971–983, March 26, 1993, Copyright © 1993 by Cell Press

A Novel Gene Containing a Trinucleotide Repeat That Is Expanded and Unstable on Huntington's Disease Chromosomes

The Huntington's Disease Collaborative Research Group*

Introduction

*The Huntington's Disease Collaborative Research Group comprises:

Group 1:

Marcy E. MacDonald,¹ Christine M. Ambrose,¹ Mabel P. Duyao,¹ Richard H. Myers,¹ Carol Lin,¹ Lakshmi Srinidhi,¹ Glenn Barnes,¹ Sherry A. Taylor,¹ Marianne James,¹ Nicole Groot,¹ Heather MacFarlane,¹ Barbara Jenkins,¹ Mary Anne Anderson,¹ Nancy S. Wexler,¹ and James F. Gusella^{1†}

¹Molecular Neurogenetics Unit
Massachusetts General Hospital
and Department of Genetics
Harvard Medical School
Boston, Massachusetts 02114

²Department of Neurology
Boston University Medical School
Boston, Massachusetts 02118
³Hereditary Disease Foundation
1427 7th Street, Suite 2
Santa Monica, California 90401

Group 2:

Gillian P. Bates,¹ Sarah Bayliss,¹ Hölger Hummerich,¹ Susan Kirby,¹ Mike North,¹ Sandra Youngman,¹ Richard Mott,¹ Gunther Zehetner,¹ Zdenek Sedlacek,¹ Annemarie Poustka,¹ Anna-Maria Pirschaut,¹ and Hans Lehrach

Genome Analysis Laboratory
Imperial Cancer Research Fund
Lincoln's Inn Fields
London, WC2A 3PX, England

Group 3:

Alan J. Buckler,¹ Deanna Church,¹
Lynn Doucette-Stamm,¹ Michael C. O'Donovan,¹

Laure Ribes-Ramirez,¹ Manish Shah,¹ Vincent P. Stanton,¹ Scott A. Strobel,² Karen M. Draths,² Jennifer L. Walea,² Peter Dervan,² and David E. Housman¹

¹Center for Cancer Research
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
²Division of Chemistry and Chemical Engineering
California Institute of Technology
Pasadena, California 91125

Group 4:

Michael Atharr,¹ Rita Shiang,¹ Leslie Thompson,¹ Thomas Fielder,¹ and John J. Wasmuth
Department of Biological Chemistry
University of California
Irvine, California 92717

Group 5:

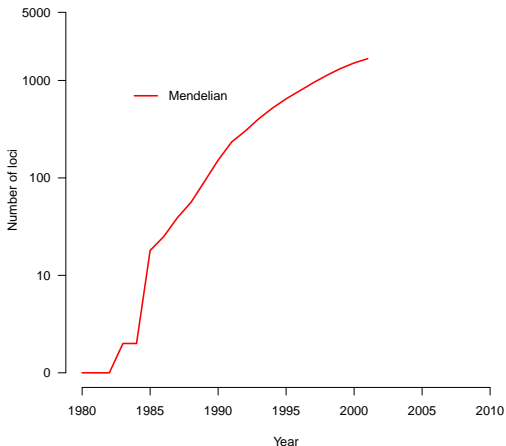
Danilo Tagle,¹ John Valdes,¹ Lawrence Elmer,¹ Marc Allard,¹ Lucio Castilla,¹ Manju Swaroop,¹ Kris Blanchard,¹ and Francis S. Collins
Department of Internal Medicine and Genetics
and The Howard Hughes Medical Institute
University of Michigan
Ann Arbor, Michigan 48109

Group 6:

Russell Snell,¹ Tracey Holloway,¹ Kathleen Gillespie,¹ Nicole Datson,¹ Duncan Shaw,¹ and Peter S. Harper
Institute of Medical Genetics
University of Wales College of Medicine
Cardiff, CF4 4XN, Wales

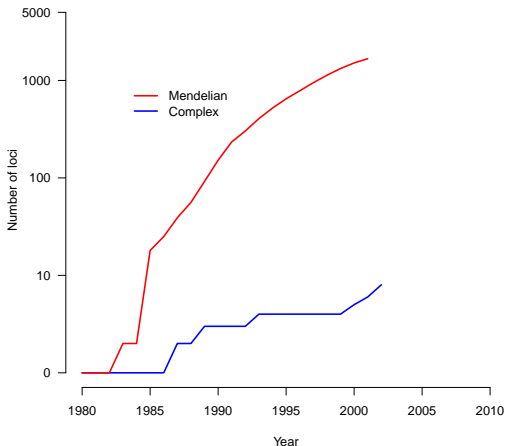
[†]Correspondence should be addressed to James F. Gusella.

Mendelian gene mapping accelerated rapidly...



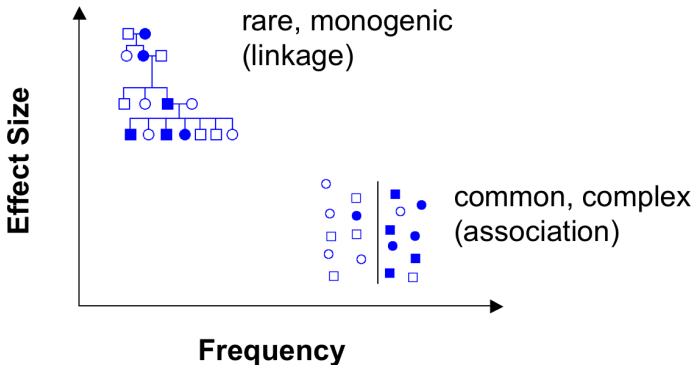
Adapted from Glazier *et al. Science*. 2002.

... but this success did not translate to complex disease



Adapted from Glazier *et al. Science*. 2002.

Different diseases need different methods of gene hunting



Obstacles to genome-wide association studies

Obstacles to genome-wide association studies

- ▶ Complete sequence of the human genome, followed by an extensive catalogue of common variation (especially SNPs) between individuals (no need to guess at functionally relevant genes).

Obstacles to genome-wide association studies

- ▶ Complete sequence of the human genome, followed by an extensive catalogue of common variation (especially SNPs) between individuals (no need to guess at functionally relevant genes).
- ▶ An understanding of the patterns of correlation between nearby SNPs, called linkage disequilibrium or LD (can capture all the genetic information with a small number of markers).

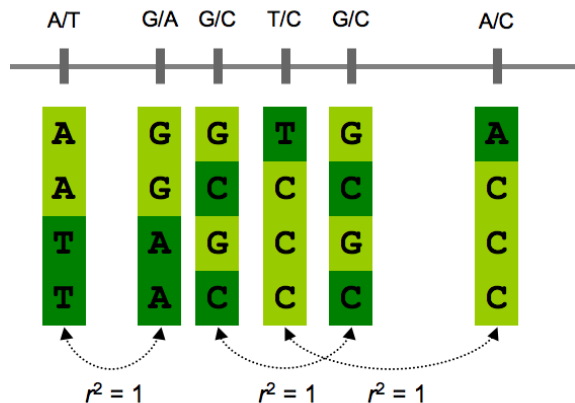
Obstacles to genome-wide association studies

- ▶ Complete sequence of the human genome, followed by an extensive catalogue of common variation (especially SNPs) between individuals (no need to guess at functionally relevant genes).
- ▶ An understanding of the patterns of correlation between nearby SNPs, called linkage disequilibrium or LD (can capture all the genetic information with a small number of markers).
- ▶ An inexpensive and accurate means of genotyping hundreds of thousands of SNPs (practical implementation of possibilities above).

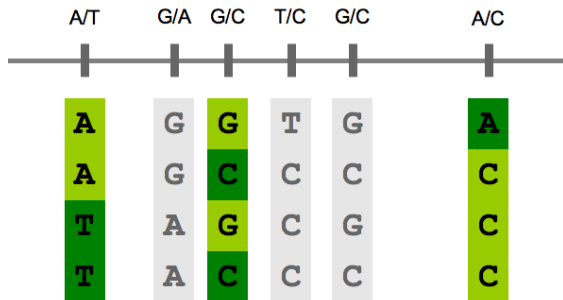
Obstacles to genome-wide association studies

- ▶ Complete sequence of the human genome, followed by an extensive catalogue of common variation (especially SNPs) between individuals (no need to guess at functionally relevant genes).
- ▶ An understanding of the patterns of correlation between nearby SNPs, called linkage disequilibrium or LD (can capture all the genetic information with a small number of markers).
- ▶ An inexpensive and accurate means of genotyping hundreds of thousands of SNPs (practical implementation of possibilities above).
- ▶ Large collections of thousands of disease cases and healthy controls (to ensure power to find very weak effects).

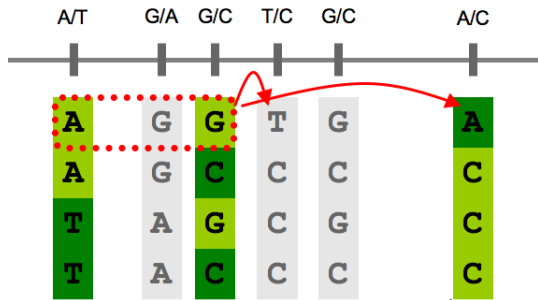
'Tagging' gains efficiency via LD



'Tagging' gains efficiency via LD

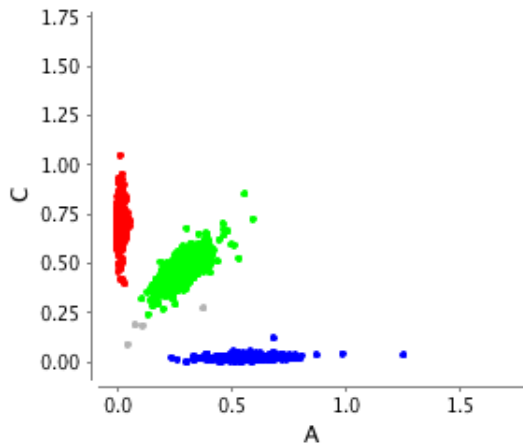


'Tagging' gains efficiency via LD

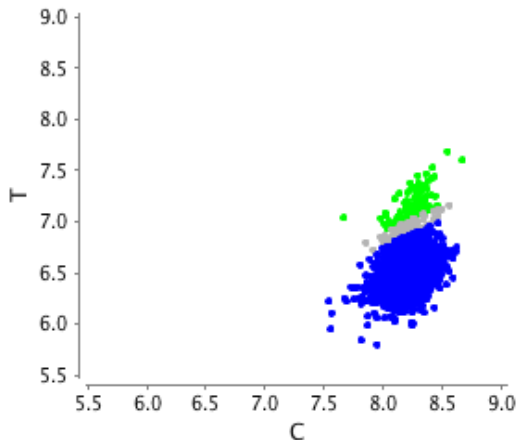


No need to genotype this SNP

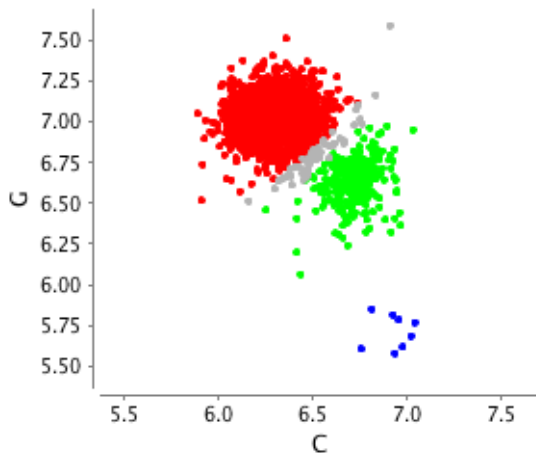
From intensity measurements to genotypes



From intensity measurements to genotypes



From intensity measurements to genotypes



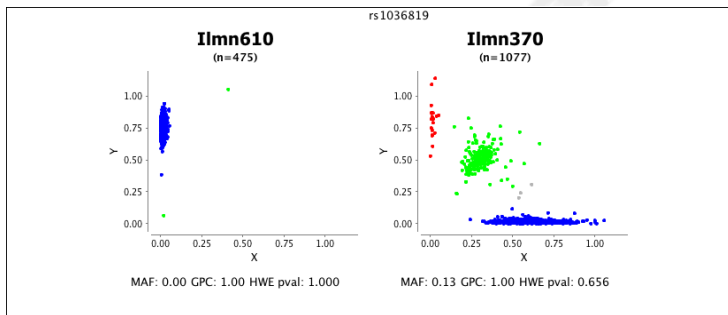
Clean data matters!

Scienceexpress

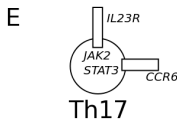
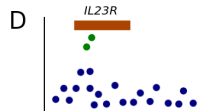
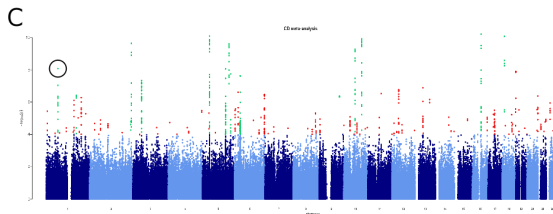
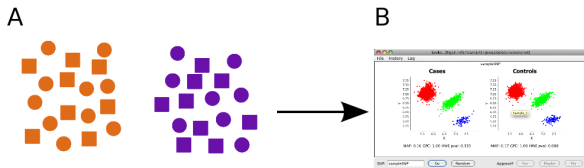
Report

Genetic Signatures of Exceptional Longevity in Humans

Paola Sebastiani,^{1*} Nadia Solovieff,¹ Annibale Puca,² Stephen W. Hartley,¹ Efthymia Melista,³ Stacy Andersen,⁴ Daniel A. Dworkis,³ Jemma B. Wilk,⁵ Richard H. Myers,⁵ Martin H. Steinberg,⁶ Monty Montano,³ Clinton T. Baldwin,^{6,7} Thomas T. Perls^{4*}



Genome wide association studies



(Ireland, €1) 70p
Thursday 7 June 2007
www.independent.co.uk
• 16,000 1.44

THE INDEPENDENT

Tracey Emin

Exclusive: How I created the show of my life

PLUS YOUR CHANCE TO OWN A LIMITED-EDITION ARTWORK IN EXTRA



Bipolar disorder
Also known as manic depression, it affects 100 million people around the world

Coronary heart disease
The most frequent cause of death in Britain, with 100,000 victims every year. By 2020, it will be the biggest killer in the world

Hypertension
High blood pressure affects 76 million people in Britain. Can lead to stroke, heart disease and kidney failure

Rheumatoid arthritis
Nearly 400,000 people in Britain are afflicted with this auto-immune disease of the joints

Type 1 diabetes
Diabetic condition in which sufferers have to inject insulin. Affects 350,000 people in UK

Crohn's disease
Up to 60,000 people are affected by this debilitating bowel condition which can cause distress and pain for a lifetime

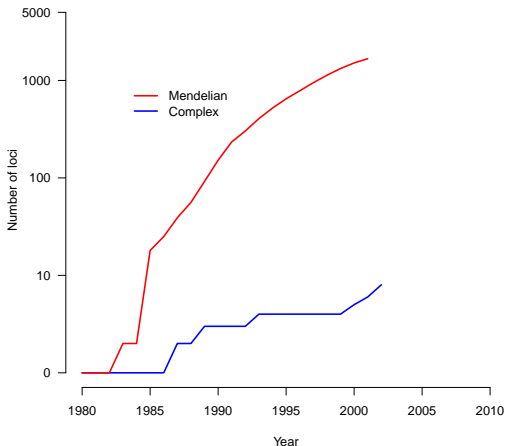
Type 2 diabetes
Almost 2 million Britons are affected by this late-onset disease, which is linked with the growing obesity epidemic

THE GENETIC REVOLUTION

DISCOVERY OF GENES RESPONSIBLE FOR SEVEN OF THE MOST COMMON ILLNESSES OFFERS HOPE TO MILLIONS OF SUFFERERS

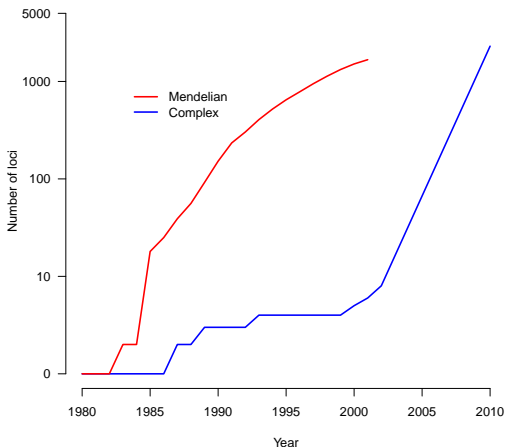
FULL STORY, PAGE 2

GWAS revolutionized complex disease genetics



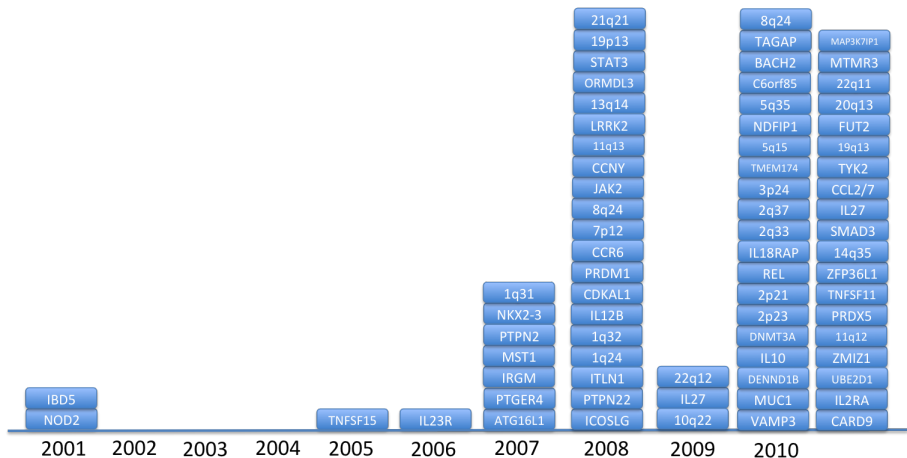
Adapted from Glazier *et al. Science*. 2002, and NHGRI GWAS catalog.

GWAS revolutionized complex disease genetics



Adapted from Glazier *et al. Science*. 2002, and NHGRI GWAS catalog.

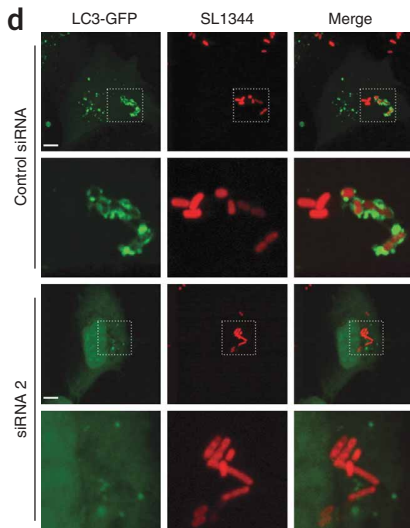
GWAS of Crohn's disease have been very successful



Disease genetics has two parallel goals

1. Understanding disease biology
2. Predicting disease risk/outcome

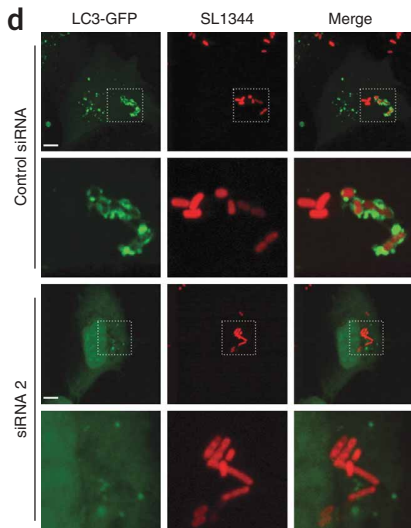
Crohn's disease and autophagy



Rioux *et al.* *Nat Genet.* 2007.

Intro to GWAS

Crohn's disease and autophagy



- ▶ Publications on autophagy and Crohn's before Hampe *et al.*: 0
- ▶ Pubs in subsequent three years: 92

Rioux *et al.* *Nat Genet.* 2007.

Can genes predict disease?


 Search

[Jeffrey Barrett](#) | [Account](#) ▾ | [Help](#) ▾ | [Blog](#) | [Log out](#)

My Home

Inbox (1)

My Health

Disease Risk

Carrier Status

Drug Response

Traits

Health Labs

My Ancestry

Maternal Line

Paternal Line

Relative Finder

Ancestry Painting

Global Similarity

Health Labs

traits

Male Pattern Baldness ★★★★★ ?

Established Research report on 2 reported markers.

[share this](#)
[Your Data](#)
[How It Works](#)
[Technical Report](#)
[Next](#) ▸

Measures of Intelligence

Technical Report

Show genotypes for in the AR and Chr 20 intergenic region genes:

23andMe Name	Other Name(s)	DNA Change	Genotype	Result
rs6625163		A to G	A	Has the A version of rs6625162 (OR = 1.17) and the AA genotype for rs6113491 (OR = 1.77). Overall, this set of genotypes confers 2.07 times higher than typical odds of male pattern baldness in Europeans.
rs6113491		A to C	AA	

UK10K: GWAS 2.0



UK
10K

RARE GENETIC VARIANTS IN HEALTH AND DISEASE

- ▶ 4,000 deeply phenotyped controls whole-genome sequenced.
- ▶ 6,000 cases (autism, schizophrenia, obesity, 8 rare) exome sequenced.
- ▶ Association analysis of low frequency and rare variation.
- ▶ *Might* be more useful in prediction.

Conclusions

- ▶ Advances in disease genetics happen when technologists, clinicians, analysts come together in the right mix.
- ▶ GWAS have revolutionized our ability to link genomic function with disease.
- ▶ Progress has been slower on prediction, but we haven't given up hope yet!