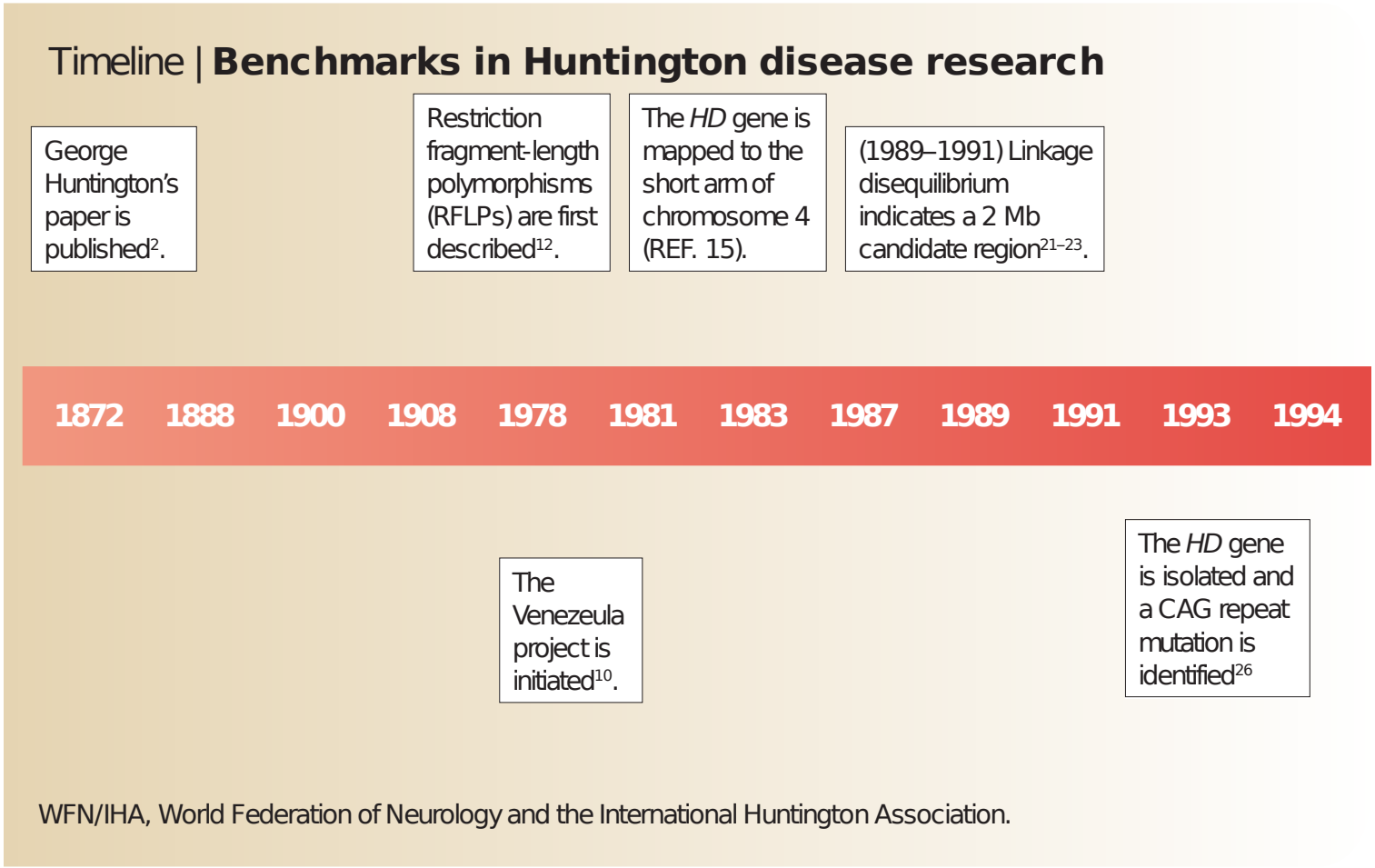
An introduction to discovering genomic variation... Genome Wide Association Studies (GWAS)

Jeff Barrett

wellcome trust sanger institute

October 19, 2010

Early Mendelian gene mapping: *HD*



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*

***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,[2] Carol Lin,[1] Lakshmi Srinidhi,[1] Glenn Barnes,[1] Sherryl A. Taylor,[1] Marianne James,[1] Nicolet Groot,[1] Heather MacFarlane,[1] Barbara Jenkins,[1] Mary Anne Anderson,[1] Nancy S. Wexler,[3] and James F. Gusella[1†]
[1]Molecular Neurogenetics Unit
Massachusetts General Hospital
and Department of Genetics
Harvard Medical School
Boston, Massachusetts 02114
[2]Department of Neurology
Boston University Medical School
Boston, Massachusetts 02118
[3]Hereditary Disease Foundation
1427 7th Street, Suite 2
Santa Monica, California 90401

Group 2:
Gillian P. Bates, Sarah Baxendale, Holger Hummerich, Susan Kirby, Mike North, Sandra Youngman, Richard Mott, Gunther Zehetner, Zdenek Sedlacek, Annemarie Poustka, Anna-Maria Frischauf, 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] Laura Riba-Ramirez,[1] Manish Shah,[1] Vincent P. Stanton,[1] Scott A. Strobel,[2] Karen M. Draths,[2] Jennifer L. Wales,[2] Peter Dervan,[2] and David E. Housman[1]
[1]Center for Cancer Research
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
[2]Division of Chemistry and Chemical Engineering
California Institute of Technology
Pasadena, California 91125

Group 4:
Michael Altherr, Rita Shiang, Leslie Thompson, Thomas Fielder, and John J. Wasmuth
Department of Biological Chemistry
University of California
Irvine, California 92717

Group 5:
Danilo Tagle, John Valdes, Lawrence Elmer, Marc Allard, Lucio Castilla, Manju Swaroop, Kris Blanchard, and Francis S. Collins
Department of Internal Medicine and Human Genetics
and The Howard Hughes Medical Institute
University of Michigan
Ann Arbor, Michigan 48109

Group 6:
Russell Snell, Tracey Holloway, Kathleen Gillespie, Nicole Datson, Duncan Shaw, and Peter S. Harper
Institute of Medical Genetics
University of Wales College of Medicine
Cardiff, CF4 4XN, Wales

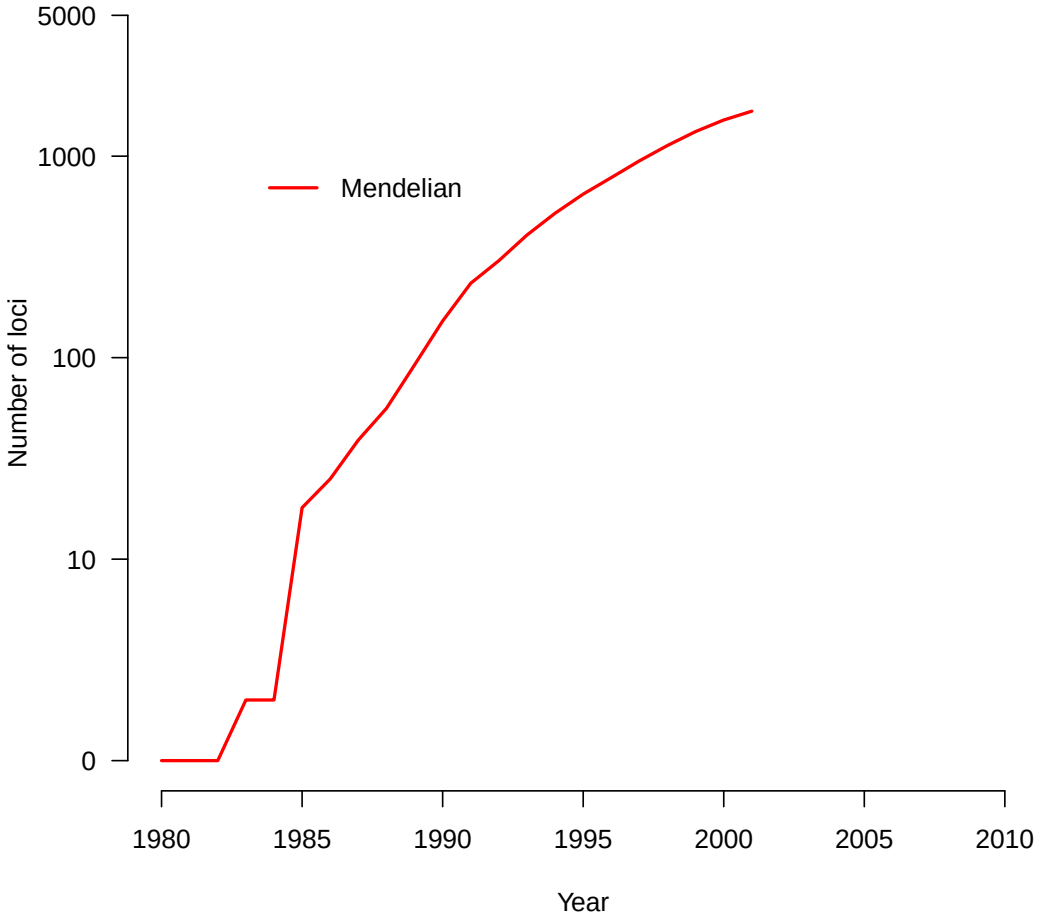
†*Correspondence should be addressed to James F. Gusella.*

Introduction

“Figure 4”

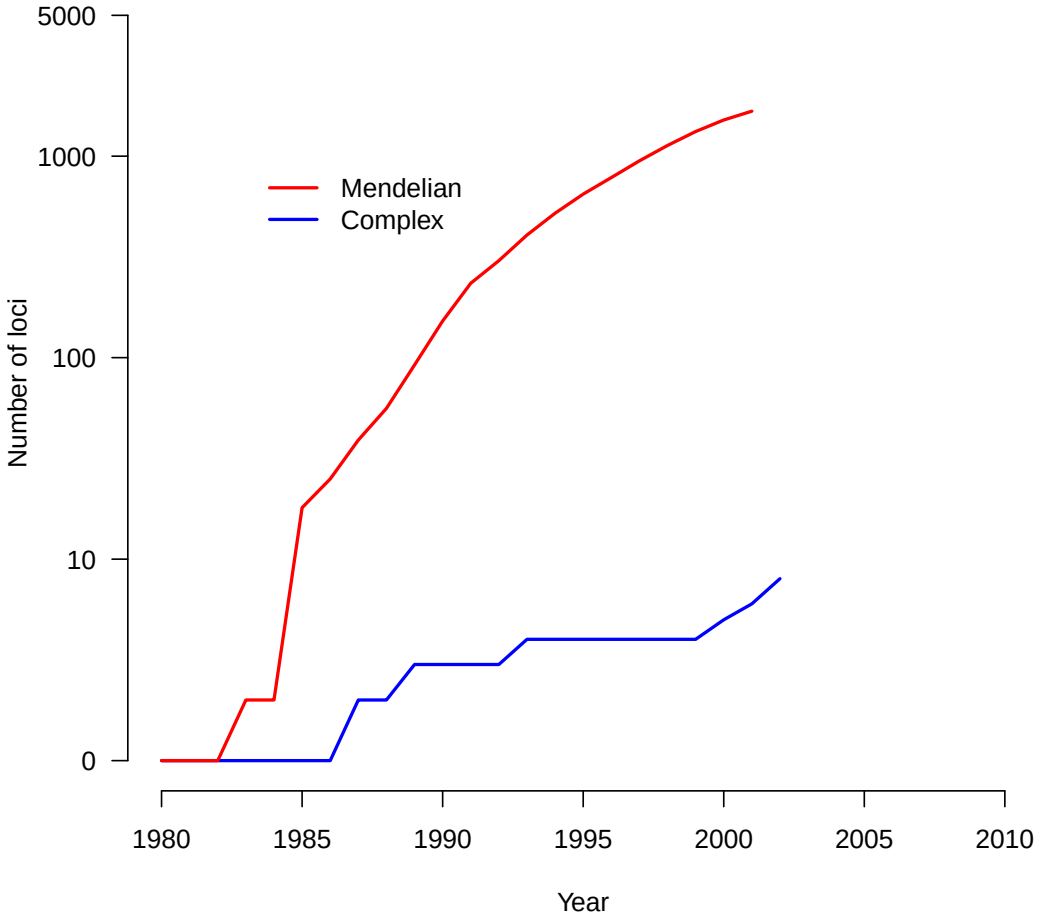
wellcome trust sanger institute

Mendelian gene mapping accelerated rapidly...



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

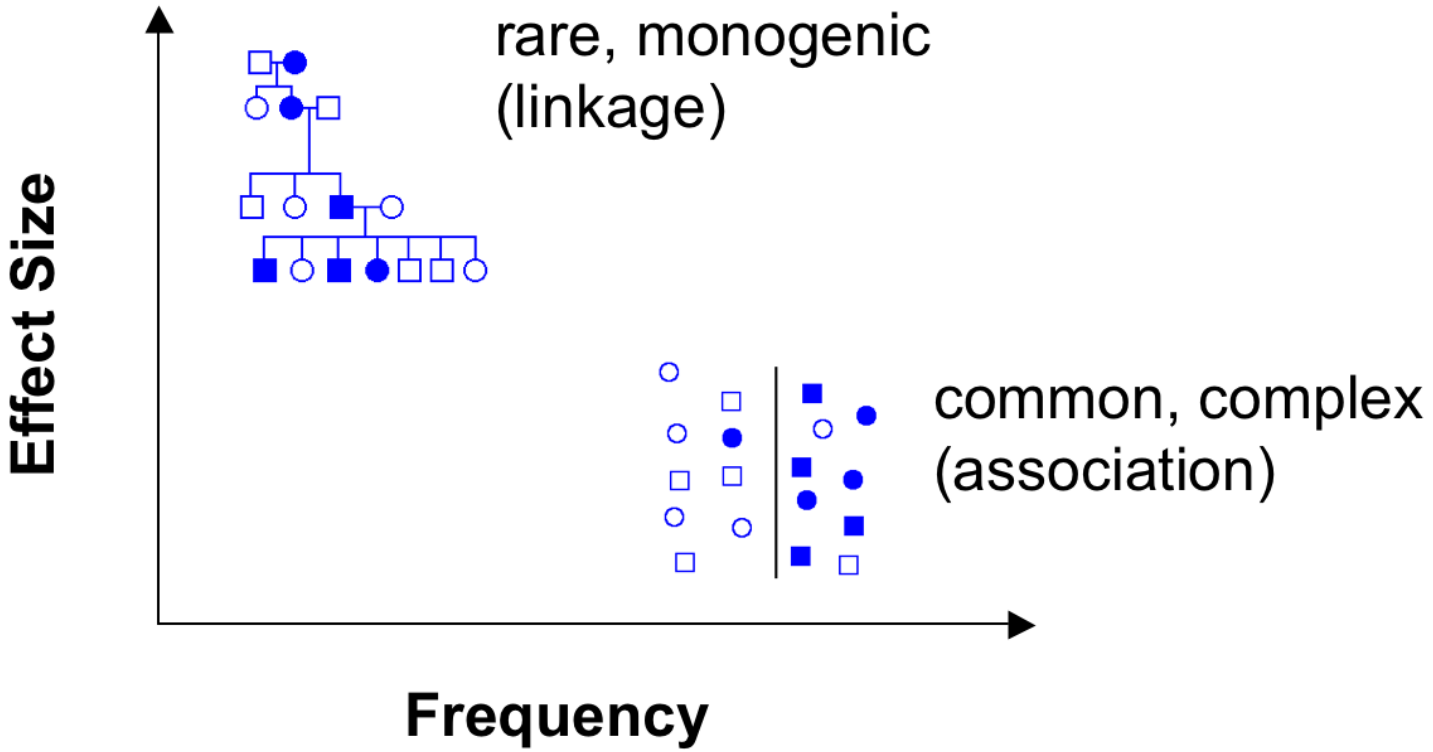
. . . but this success did not translate to complex disease



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

wellcome trust sanger institute

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).

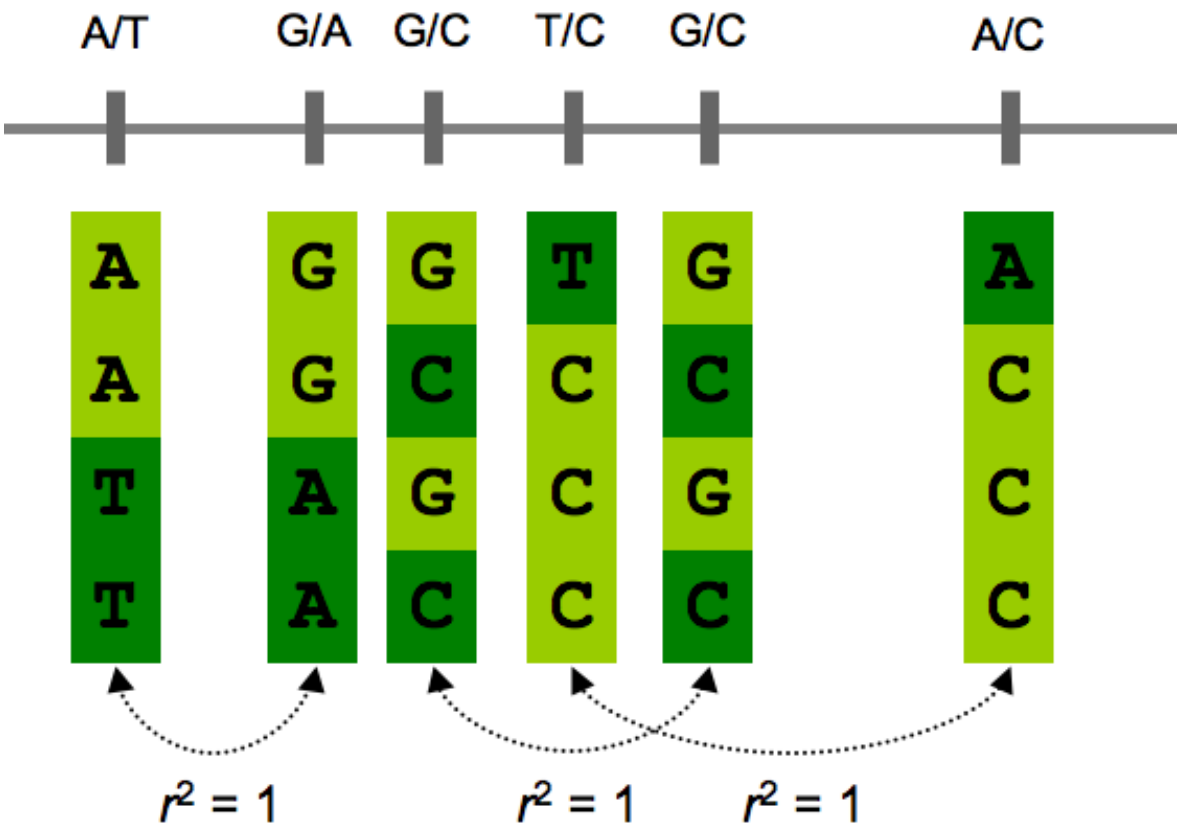
wellcome trust sanger institute

Obstacles to genome-wide association studies

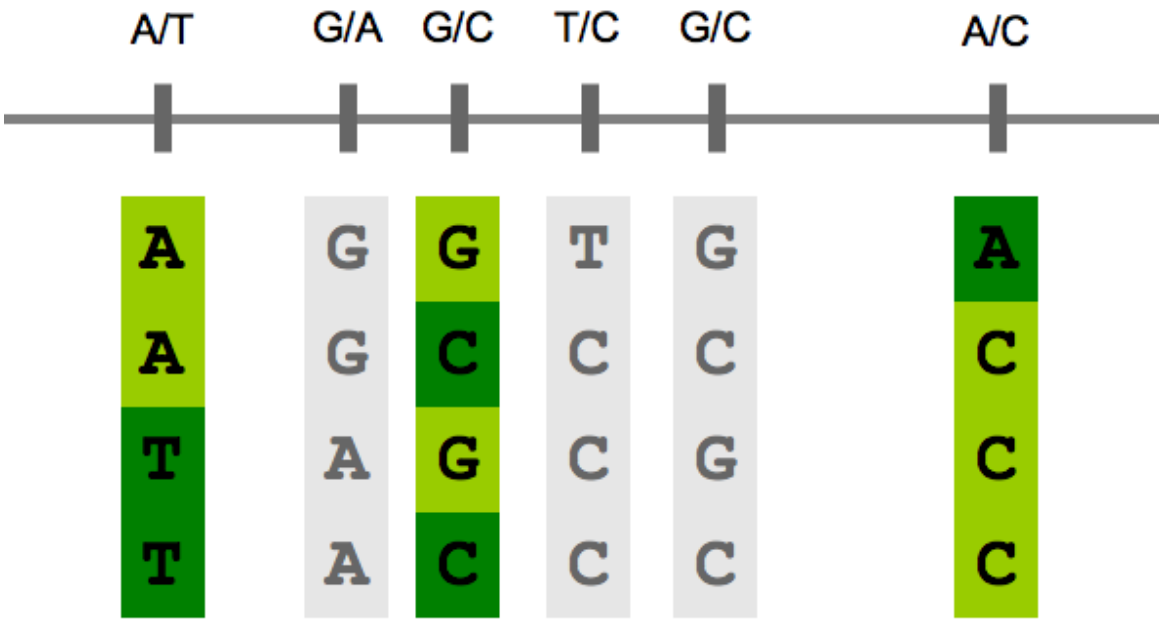
- Complete sequence of the human genome, followed by an extensive catalogue of common variation (especially SNPs) between individuals (no need to guess at functionally relevant genes).
- An understanding of the patterns of correlation between nearby SNPS, called linkage disequilibrium or LD (can capture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

wellcome trust sanger institute

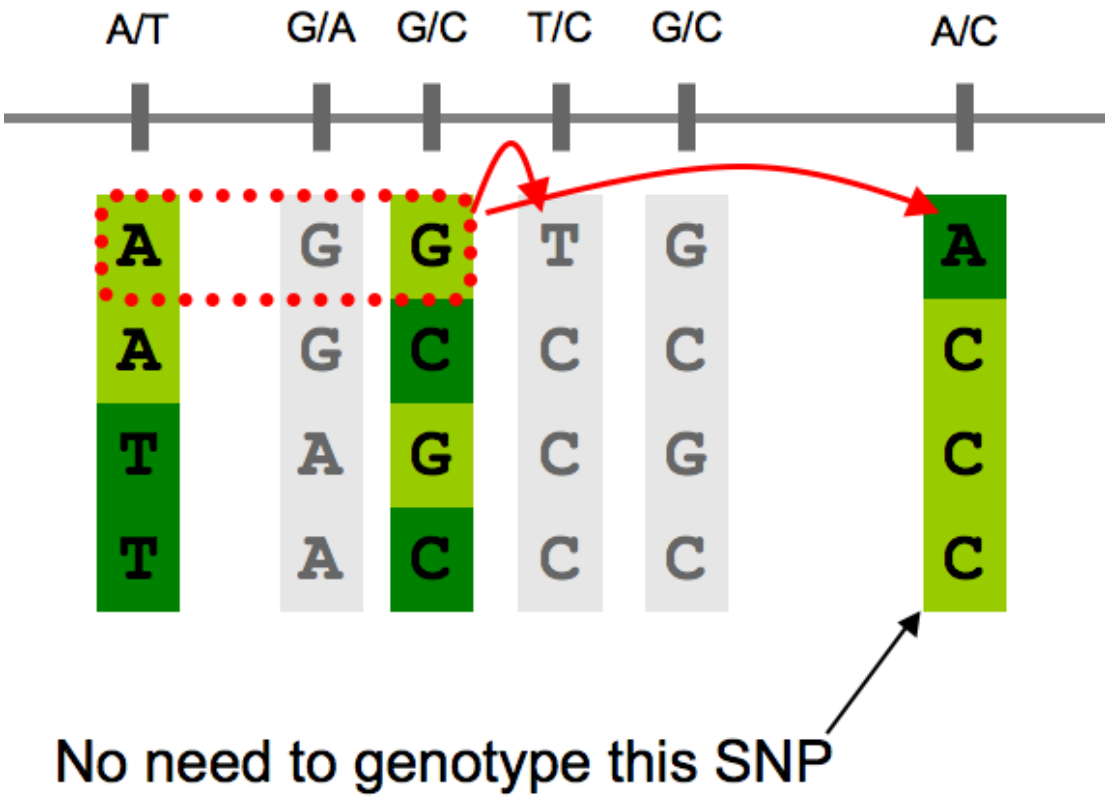
‘Tagging’ gains efficiency via LD



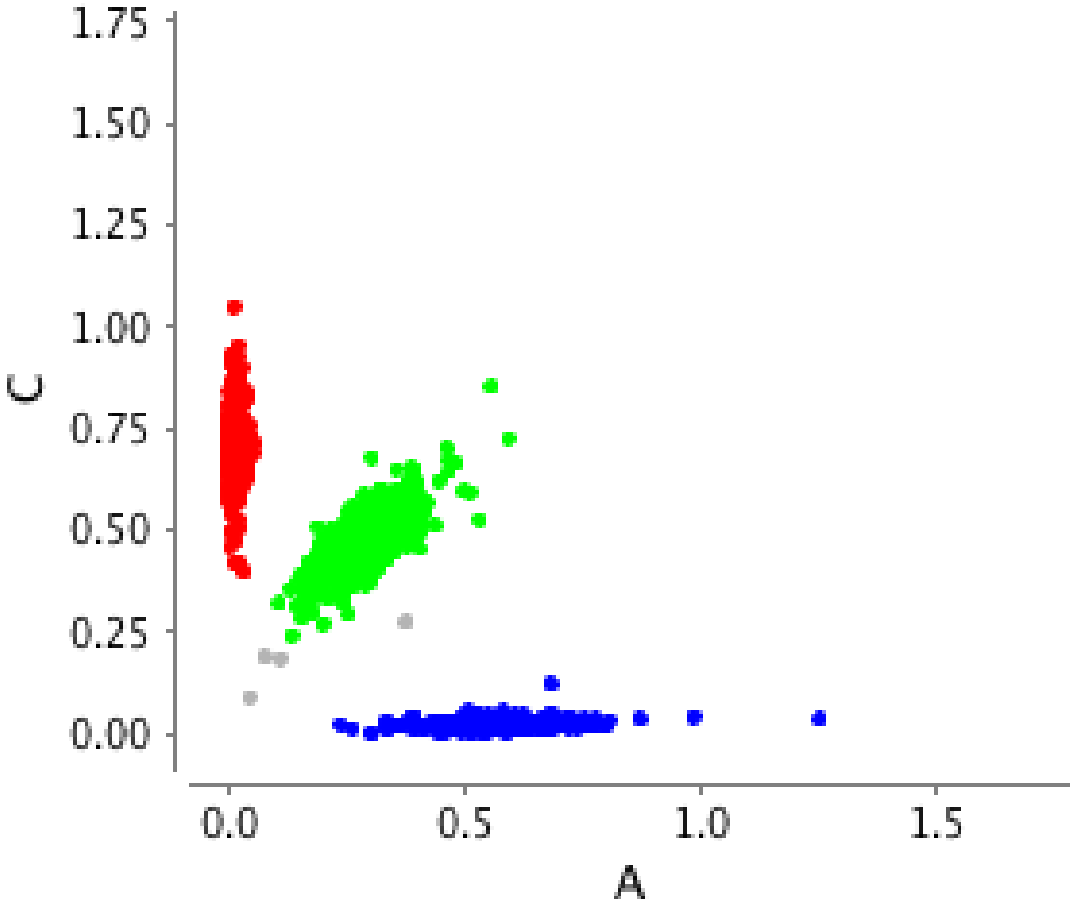
‘Tagging’ gains efficiency via LD



‘Tagging’ gains efficiency via LD

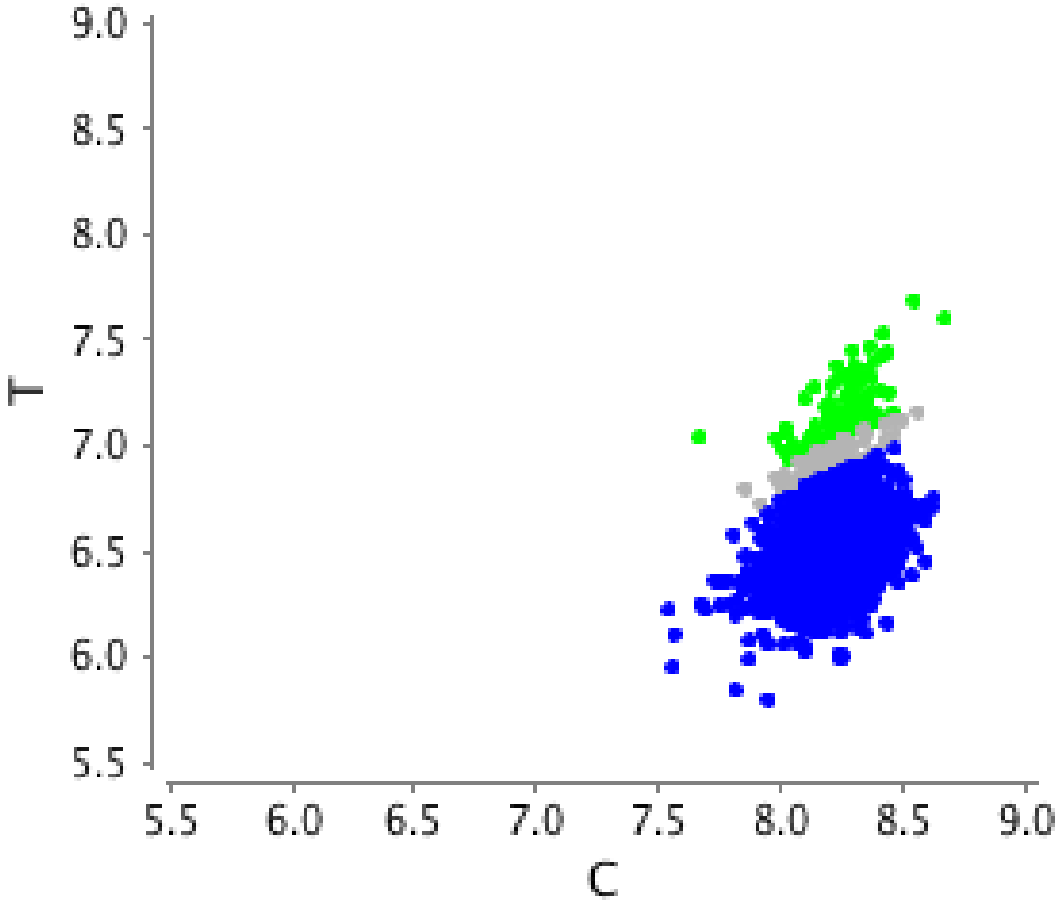


From intensity measurements to genotypes



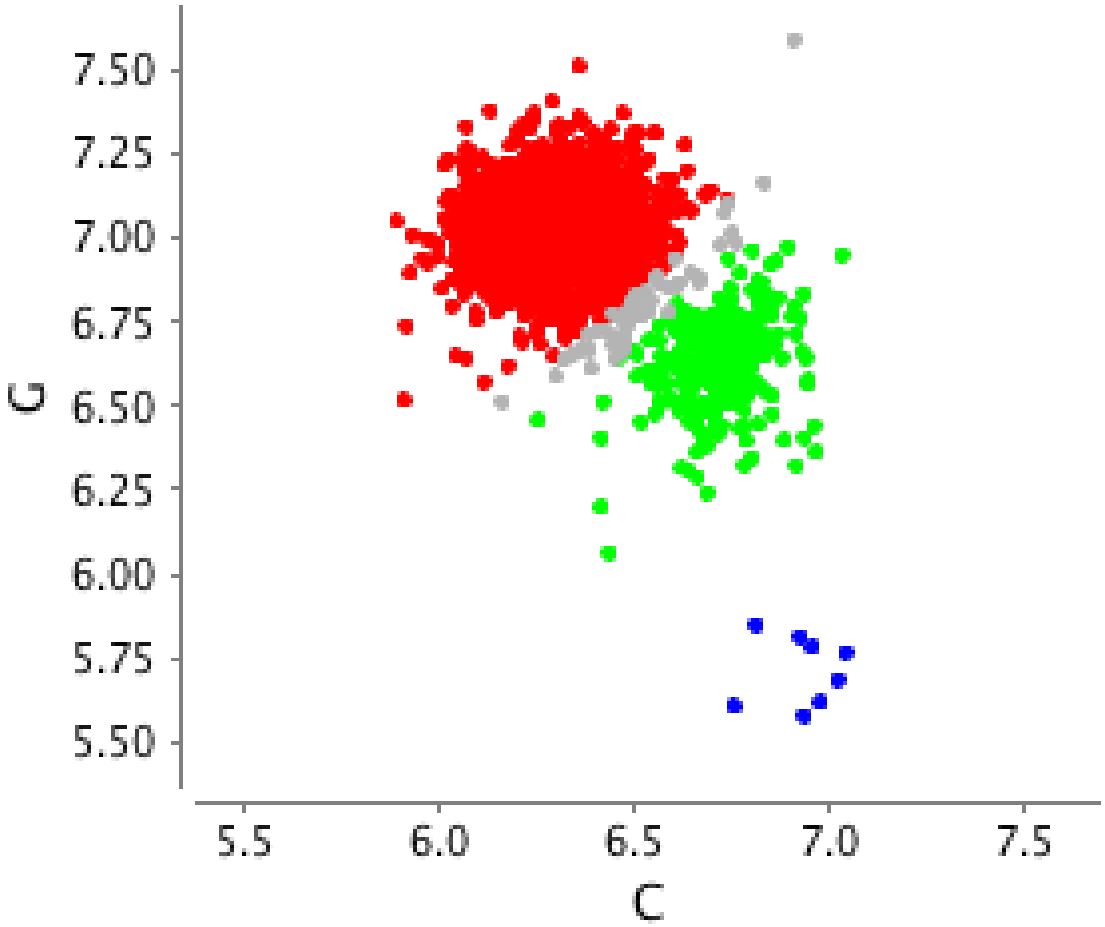
wellcome trust sanger institute

From intensity measurements to genotypes



wellcome trust sanger institute

From intensity measurements to genotypes



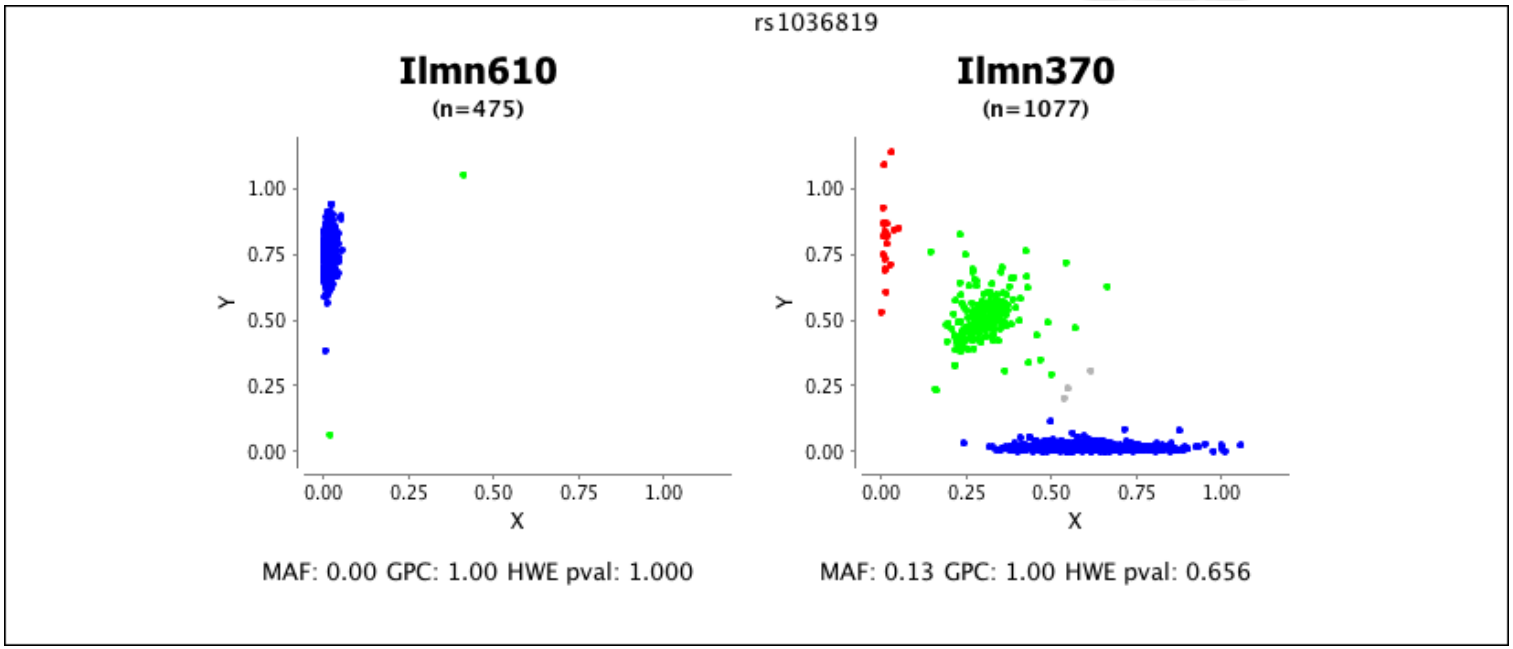
wellcome trust sanger institute

Clean data matters!

Sciencexpress Report

Genetic Signatures of Exceptional Longevity in Humans

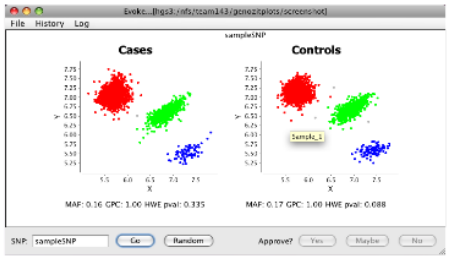
Paola Sebastiani,[1]* Nadia Solovieff,[1] Annibale Puca,[2] Stephen W. Hartley,[1] Efthymia Melista,[3] Stacy Andersen,[4] Daniel A. Dworkis,[3] Jemma B. Wilk,[5] Richard H. Myers,[5] Martin H. Steinberg,[6] Monty Montano,[3] Clinton T. Baldwin,[6,7] Thomas T. Perls[4]*



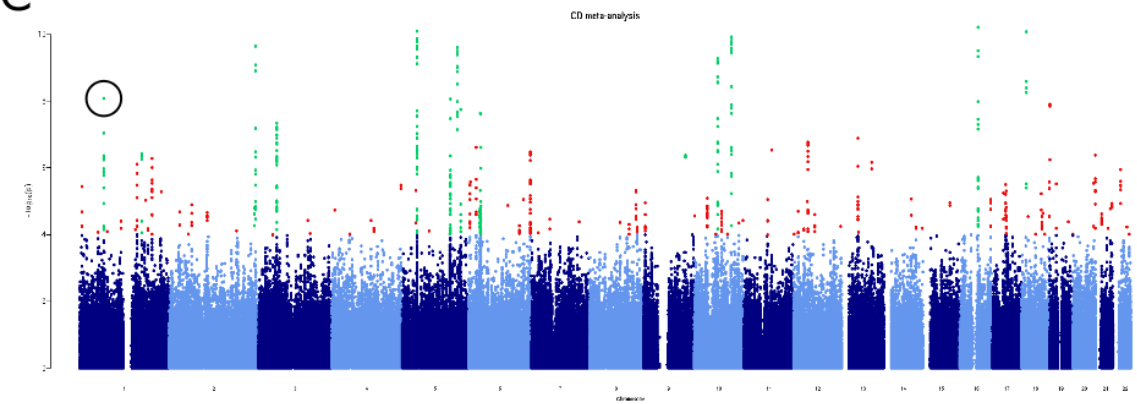
Genome wide association studies

A

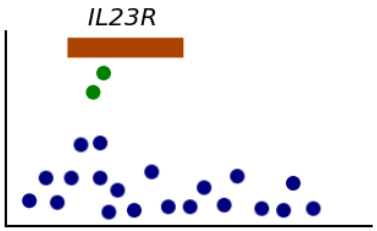
B



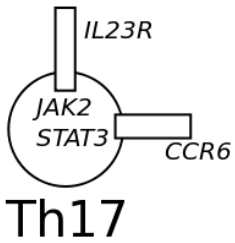
C



D



E



wellcome trust sanger institute

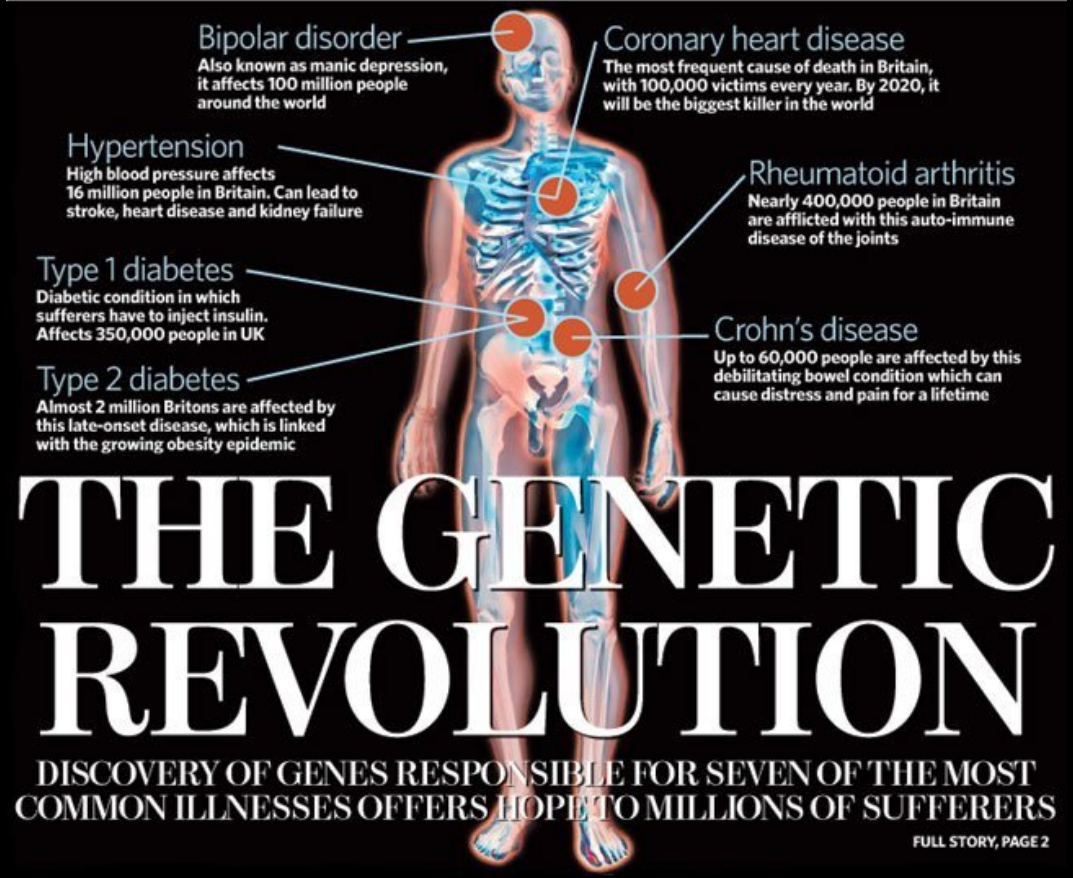
(Ireland, €1) 70p
Thursday 7 June 2007
www.independent.co.uk

THE INDEPENDENT

Tracey Emin

Exclusive: How I created the show of my life

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

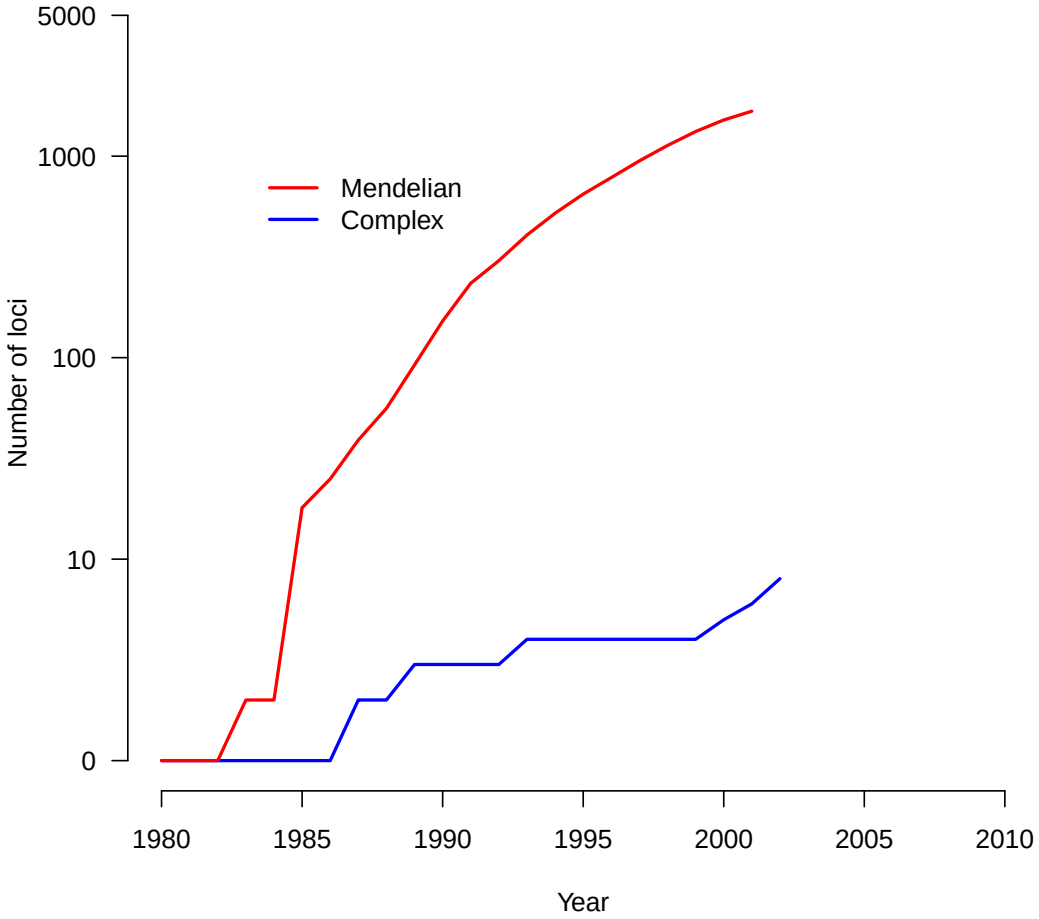


THE GENETIC REVOLUTION

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

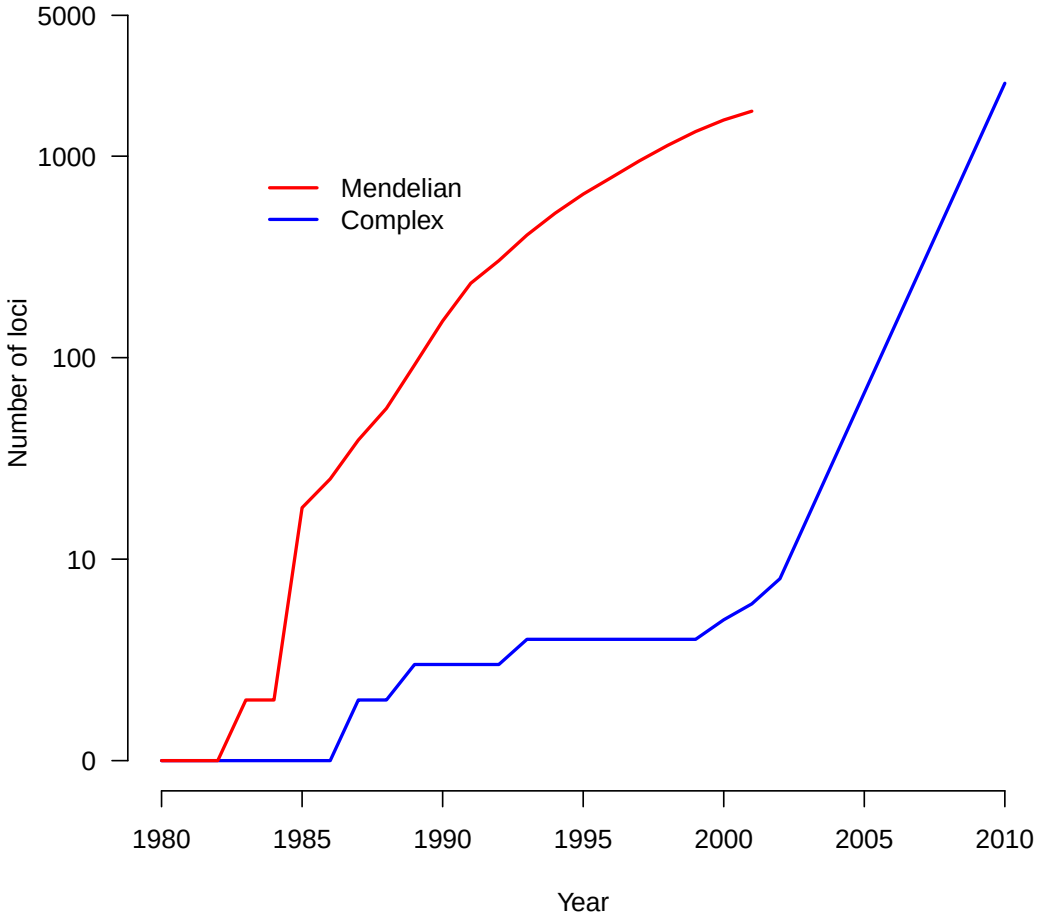
FULL STORY, PAGE 2

GWAS revolutionized complex disease genetics



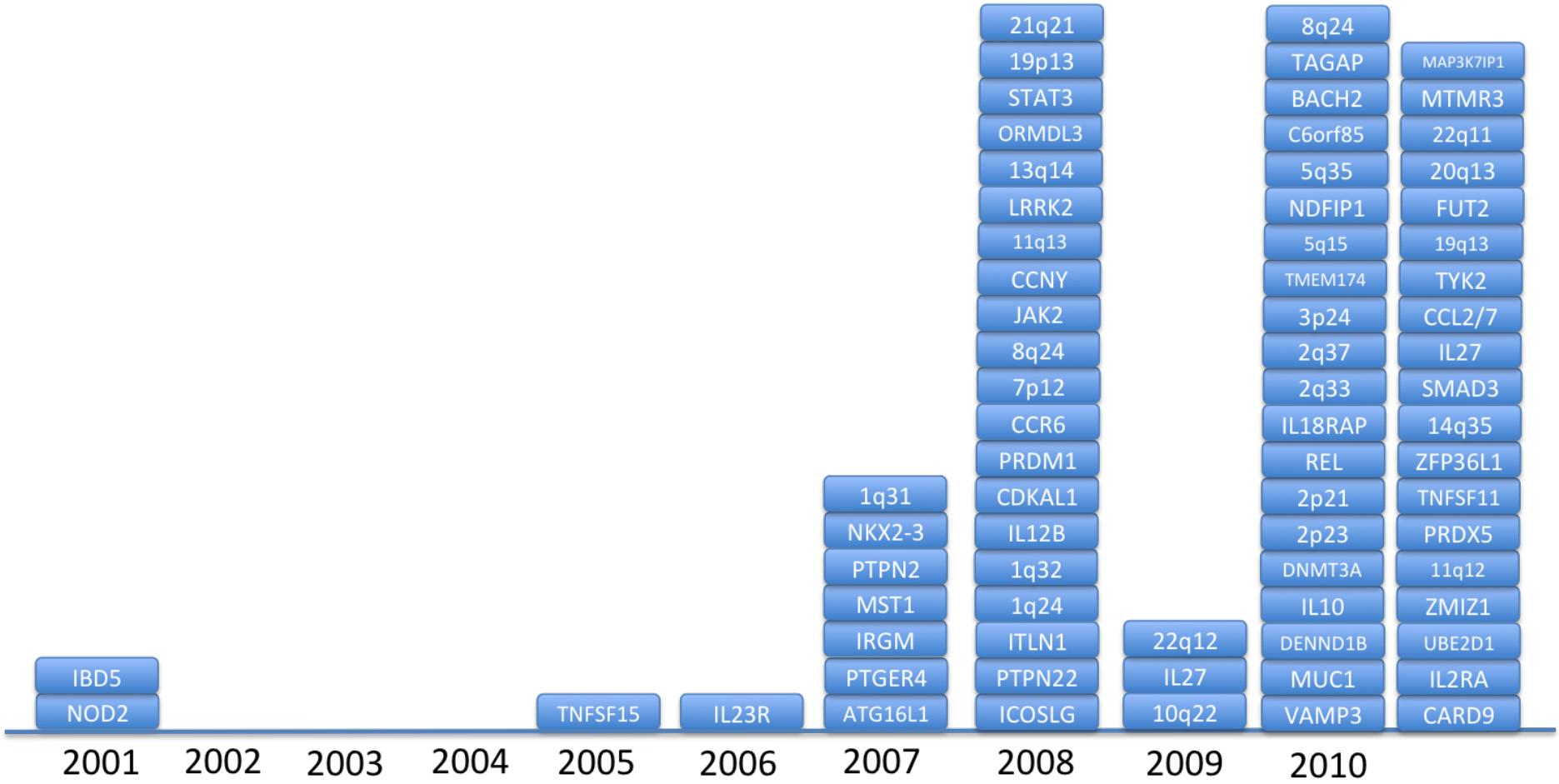
Adapted from Glazier *et al. Science.* 2002, and NHGRI GWAS catalog.

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

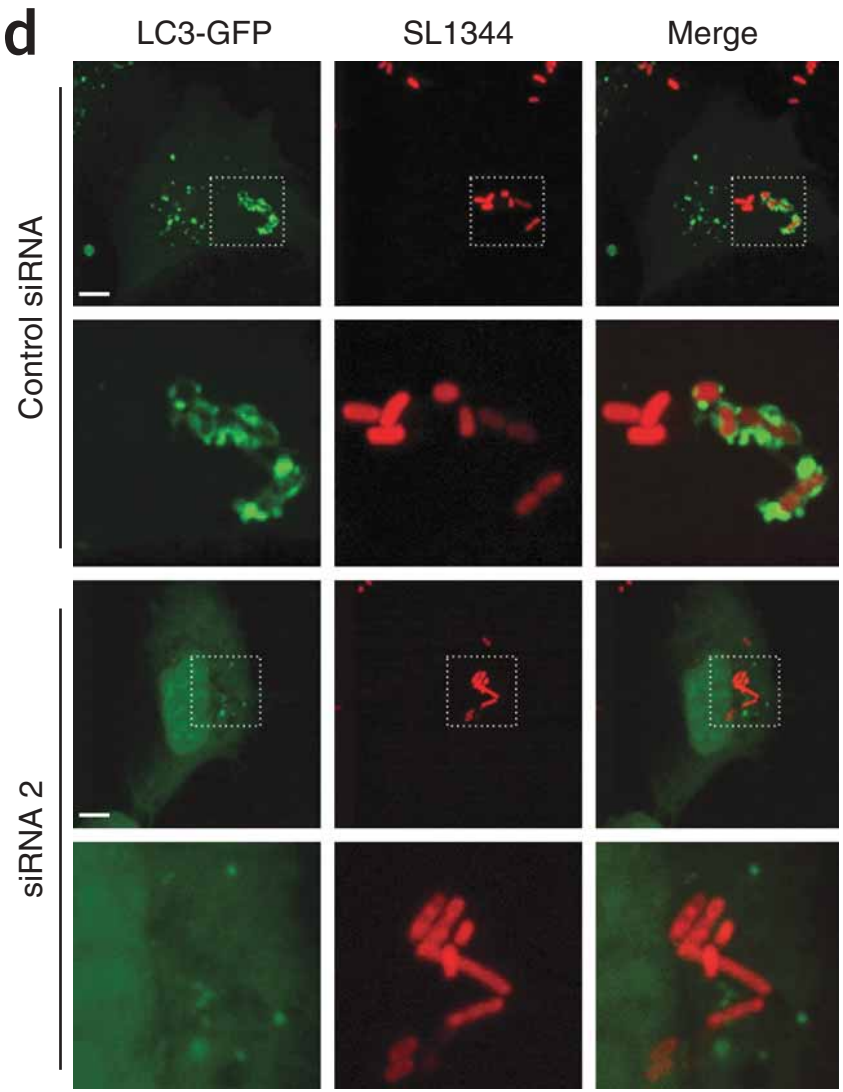


Disease genetics has two parallel goals

1. Understanding disease biology

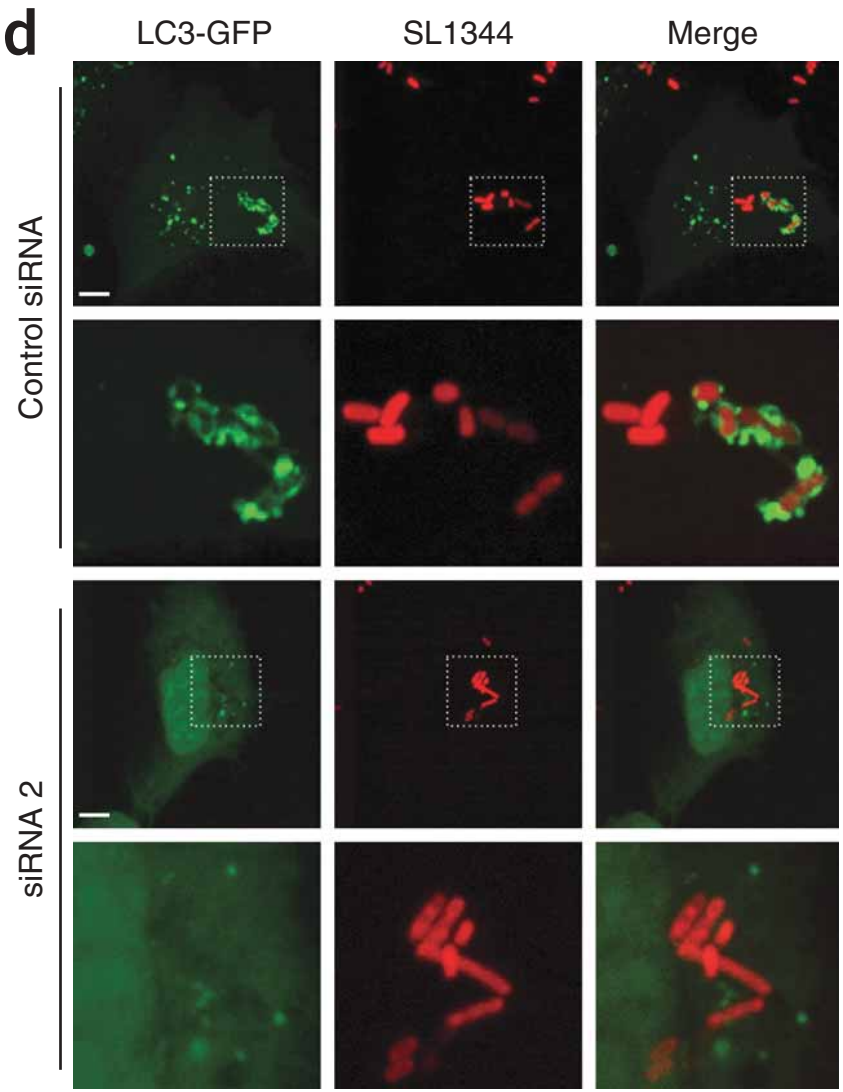
2. Predicting disease risk/outcome

Crohn's disease and autophagy



Rioux *et al. Nat Genet.* 2007.

Crohn's disease and autophagy

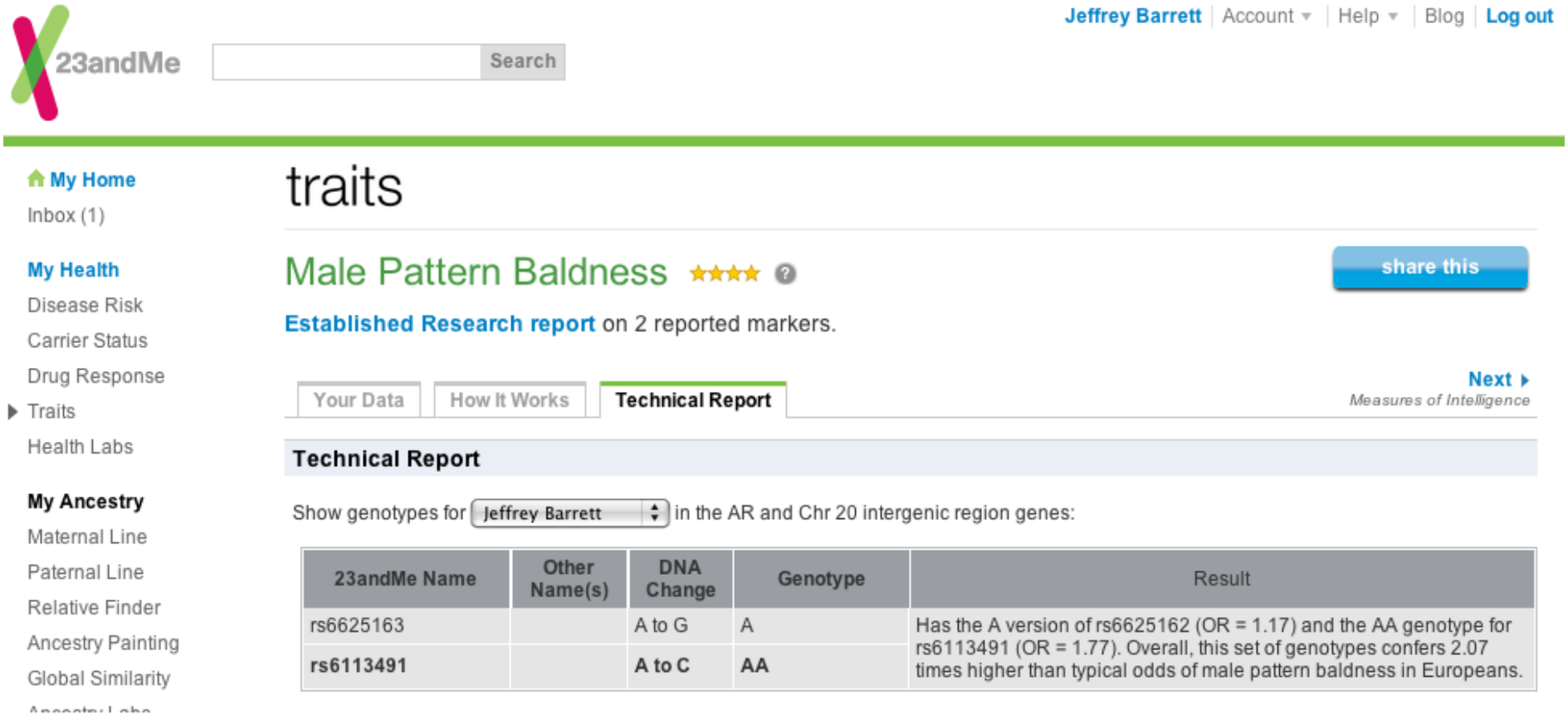


- Publications on autophagy and Crohn's before Hampe *et al.*: 0
- Pubs in subsequent three years: 92

Rioux *et al. Nat Genet.* 2007.

wellcome trust sanger institute

Can genes predict disease?



wellcome trust sanger institute

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.

wellcome trust sanger institute

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!

wellcome trust sanger institute