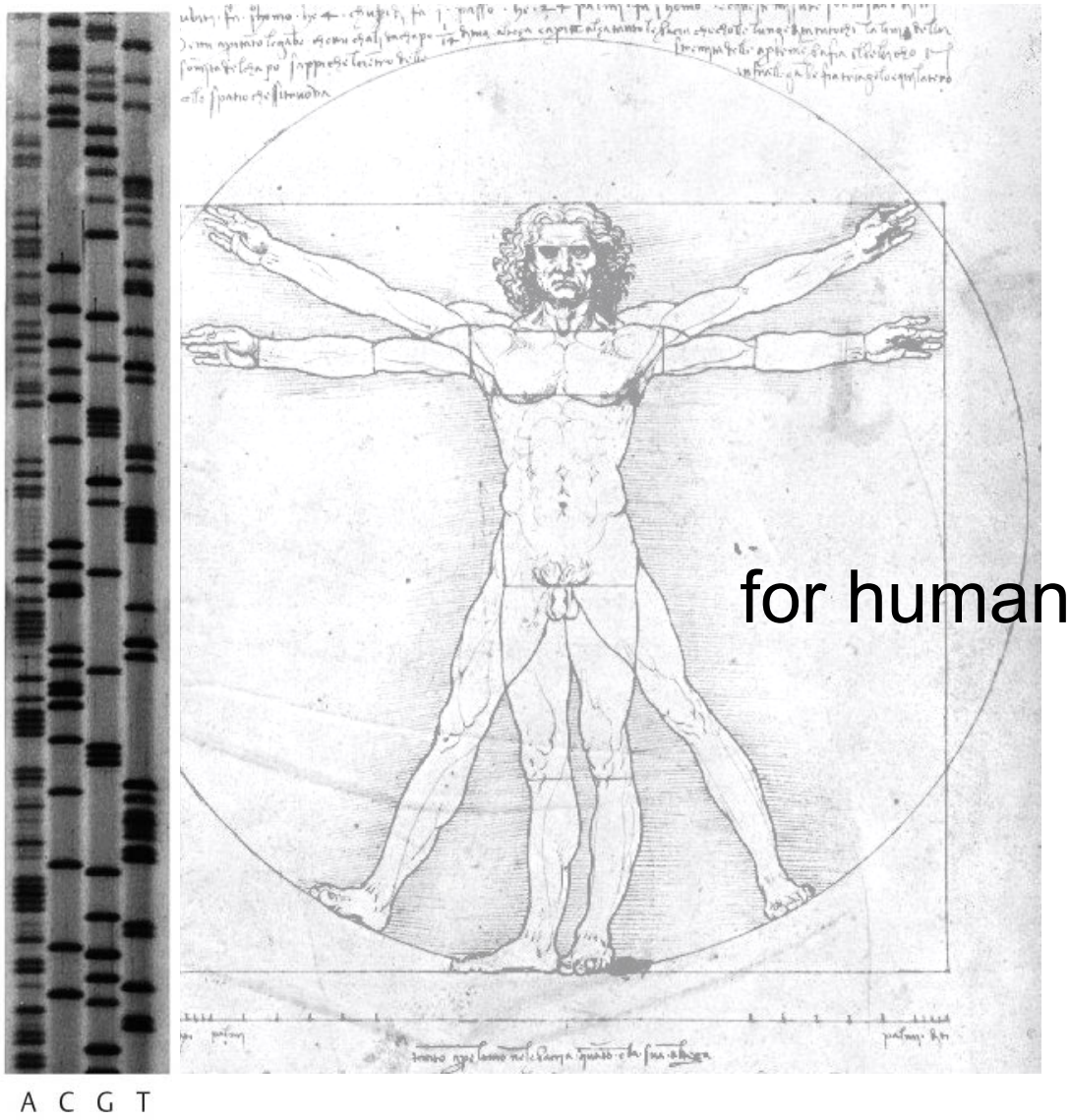


GWAS

for human aging-related loci

Matthias Platzer

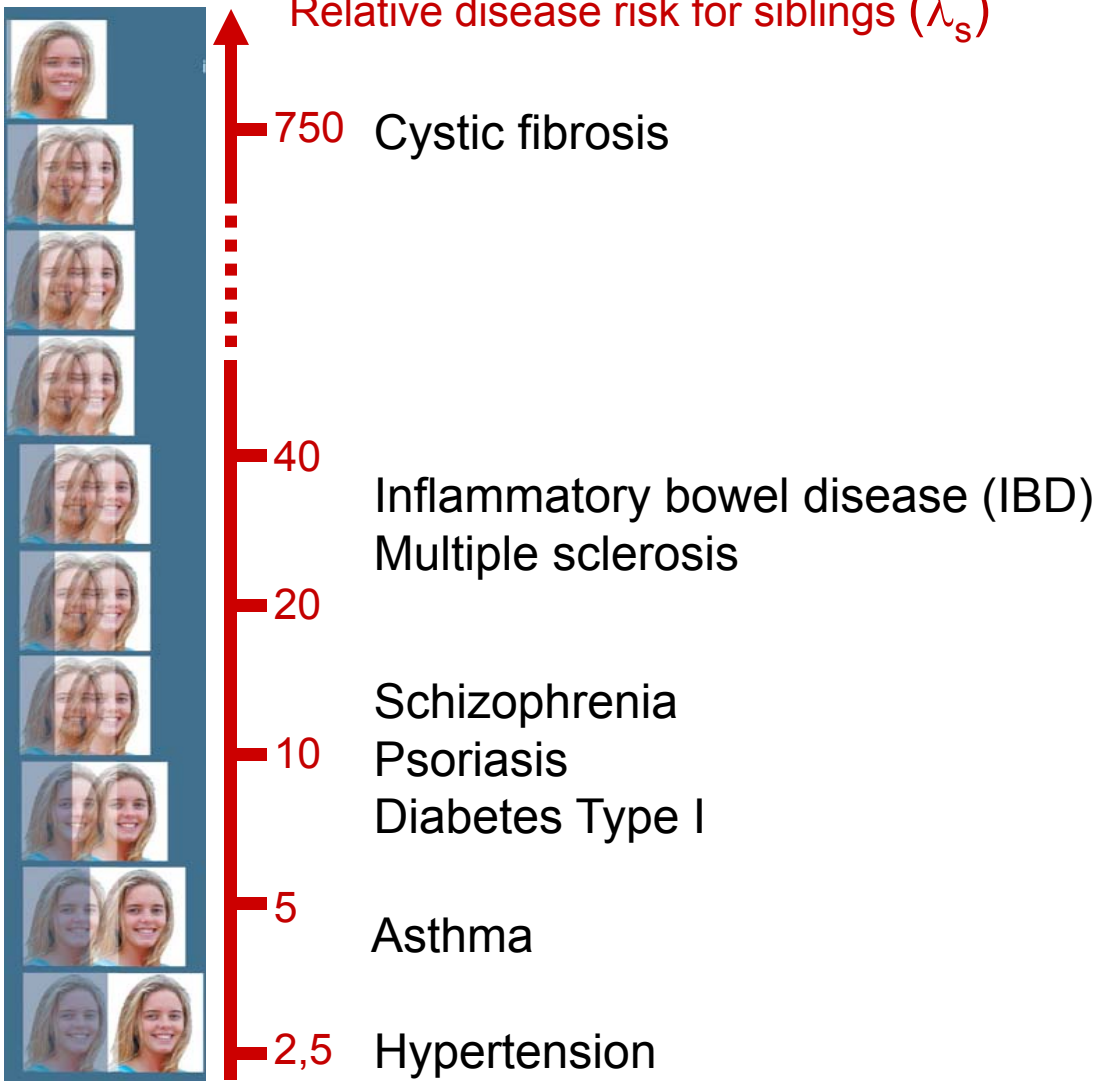
Genome Analysis
Leibniz Institute for Age Research
- Fritz Lipmann Institute (FLI)

Genetic risk

Siblings

We used to think our fate
was in our stars.
Now we know,
in large measure,
our fate is in our genes

James Watson, 1989



Human genetics

Limitations & advantages

non-human models

+ experimental intervention
+ genetic manipulation
- limited genetic & phenotypic variation

humans

- no experimental intervention
- no genetic manipulation
+ wealth of phenotypic information reflecting huge genetic variability

Human aging

Genetic predisposition

- heritability of human longevity: ~15 to 30%
  - twin-studies, large population-based samples
- greater genetic influences on longevity once an individual achieves age 60
- much larger genetic contribution to other aspects of aging
  - healthy physical aging (wellness)
  - physical performance
  - cognitive function
  - bone aging
- both exceptional longevity and a healthy aging phenotype linked to the same region / common genetic pathways?

GWAS

Genome-Wide Association Studies

Definition *NIH*

Study of genetic variation across the entire human genome designed to identify genetic associations with observable traits (blood pressure, weight), or the presence or absence of a disease or condition.

Aim

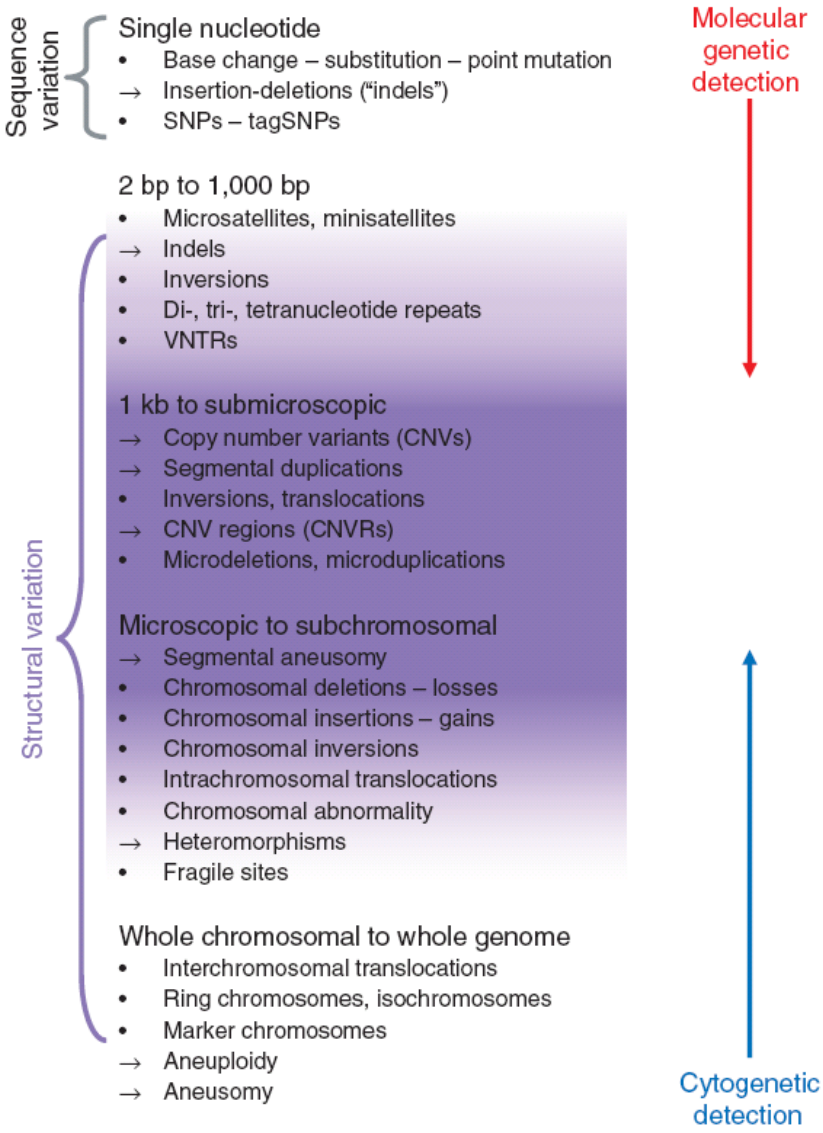
- increased understanding of basic biological processes affecting human health,
- improvement in the prediction of disease and patient care,
- ultimately the realization of the promise of personalized medicine.

Research tools

High-throughput, cost-effective methods for genotyping

Genetic variation

Lexicon



Scherer et al., Nat Genet Suppl 39:s7 2007

Genetic variation

Terminology

Mutation

= event causing genetic variation

substitution, insertion, deletion, inversion

Polymorphism

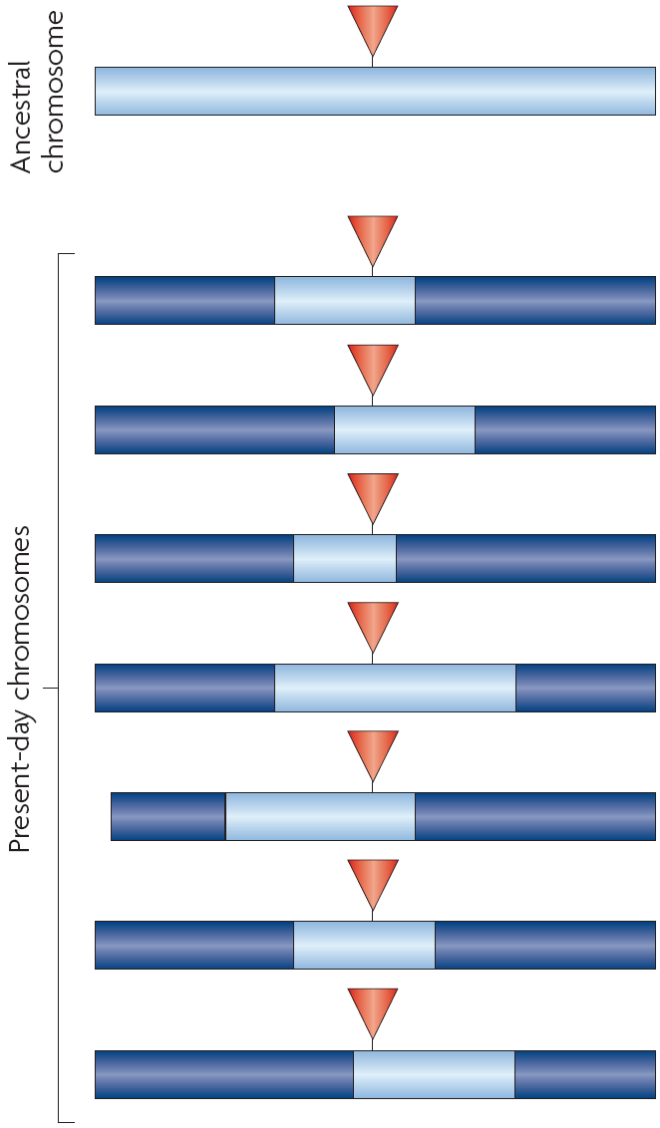
= condition of a variation, when it is established with frequency ≥1% in a population

Mutation *in medical genetics*

= rare variation with a population frequency <1%

Genetic variation

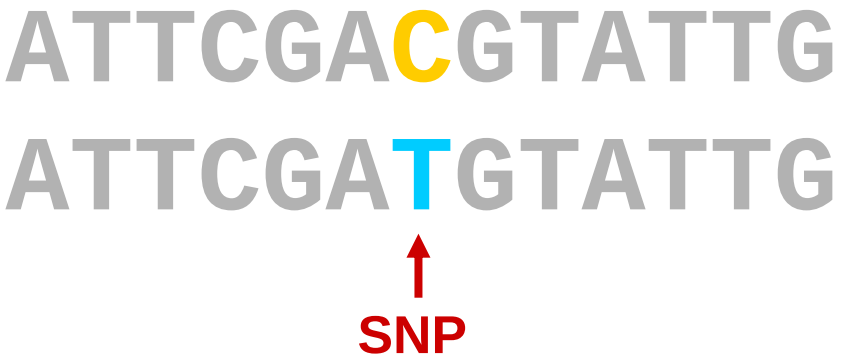
Linkage disequilibrium (LD)



Kruglyak. Nat Rev Genet 9:314 2008

Genetic variation

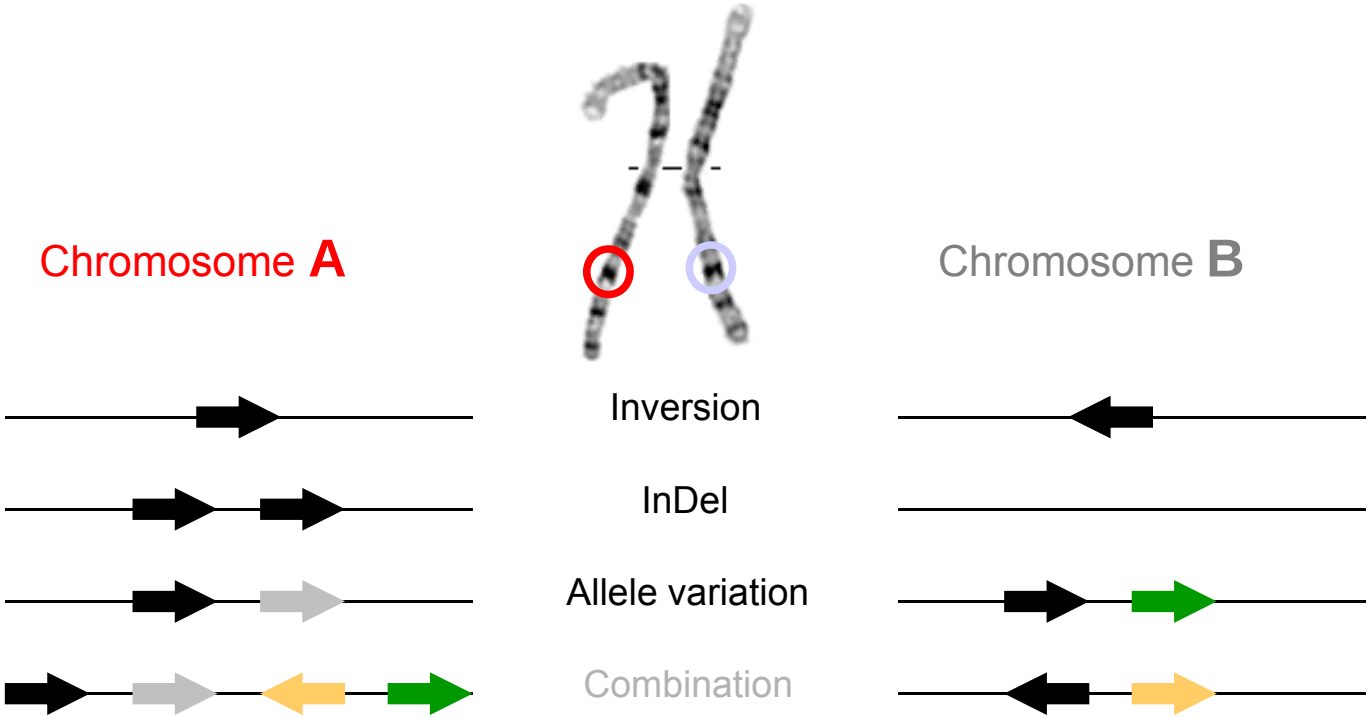
Single Nucleotide Polymorphism (SNP)



- as a rule bi-allelic
- 12 Mio SNPs genom-wide 1/250 bp
- 2 individuals differ in ~300.000 SNPs 1/10.000 bp
- ~5% of SNPs, e.g. 600.000 SNPs with phenotyp (?)
  50-100.000 SNPs with clinical relevance (?)

Genetic variation

Structural variations



Genetic variation

Segmental duplications

genomic regions >**1kb**

with nt identity >**90%**

Human genome

5.3% segmentally duplicated

87% of all segmental duplications >**50 kb**

Genetic variation

Summary

Genomes of any two individuals in the human population differ more at the structural level than at the nucleotide sequence level.

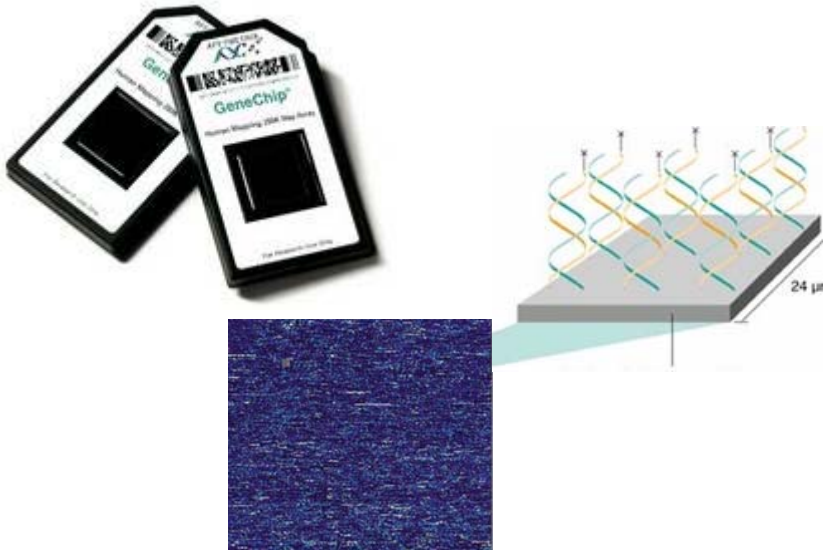
Differences between individuals

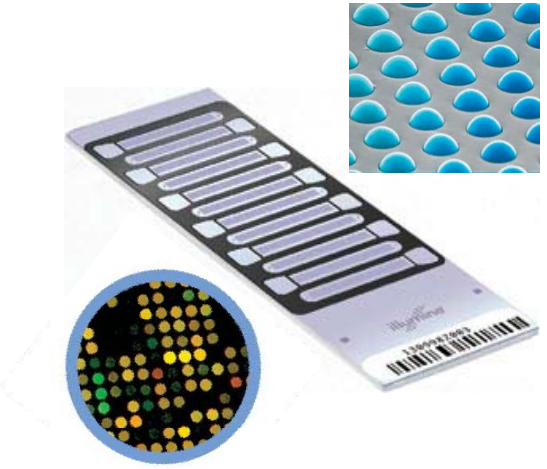
- CNV: >4 Mb >1/800 bp > 0.12 %
- SNP: 2.5 Mb 1/1,200 bp 0.08 %

Sebat, Nat Gen Suppl 39:s3 2007

GWAS prerequisites

High-throughput array-based genotyping





Affymetrix

Human SNP Array 6.0

>1.8 million markers

906,600 SNPs

946,000 for CNVs

Illumina

Human 660W-Quad BeadChip

2.6 million markers / four samples

550,000 tag SNPs

100,000 for CNVs

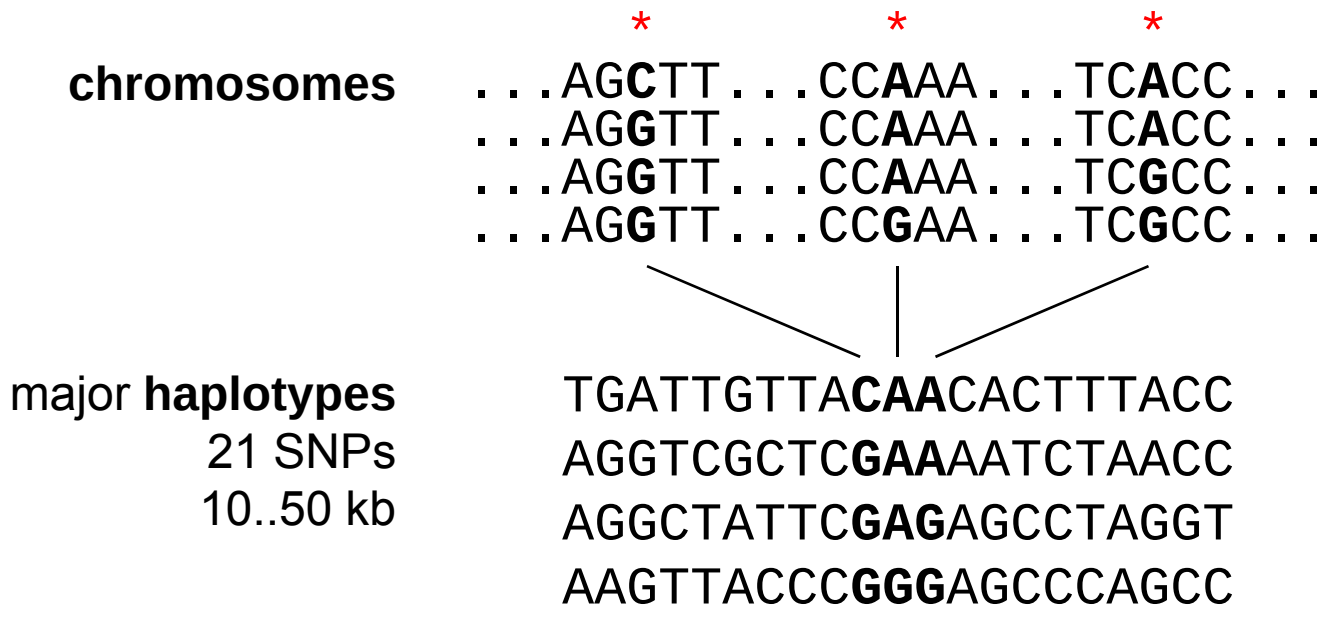
5,000 common CNVs

Genetic variation

Terminology

Haplotype

= set/region physically linked polymorphism



GWAS prerequisites

Haplotype maps

Vol 449|18 October 2007|doi:10.1038/nature06258 nature

ARTICLES

A second generation human haplotype map of over 3.1 million SNPs

The International HapMap Consortium*

We describe the Phase II HapMap, which characterizes over 3.1 million human single nucleotide polymorphisms (SNPs) genotyped in 270 individuals from four geographically diverse populations and includes 25–35% of common SNP variation in the populations surveyed. The map is estimated to capture untyped common variation with an average maximum r^2 of between 0.9 and 0.96 depending on population. We demonstrate that the current generation of commercial genome-wide genotyping products captures common Phase II SNPs with an average maximum r^2 of up to 0.8 in African and up to 0.95 in non-African populations, and that potential gains in power in association studies can be obtained through imputation. These data also reveal novel aspects of the structure of linkage disequilibrium. We show that 10–30% of pairs of individuals within a population share at least one region of extended genetic identity arising from recent ancestry and that up to 1% of all common variants are untaggable, primarily because they lie within recombination hotspots. We show that recombination rates vary systematically around genes and between genes of different function. Finally, we demonstrate increased differentiation at non-synonymous, compared to synonymous, SNPs, resulting from systematic differences in the strength or efficacy of natural selection between populations.

TIHC. Nature 449:851 2007

GWAS prerequisites

HapMap Project

270 individuals from

4 geographically diverse populations:

YRI Africans: 30 trios (Yoruba in Ibadan, Nigeria)
CEU European: 30 trios of northern/western ancestry (Utah, US; CEPH collection)
CHB Chinese: 45 unrelated Han individuals (Beijing, China)
JPT Japanese: 45 unrelated individuals (Tokyo, Japan)

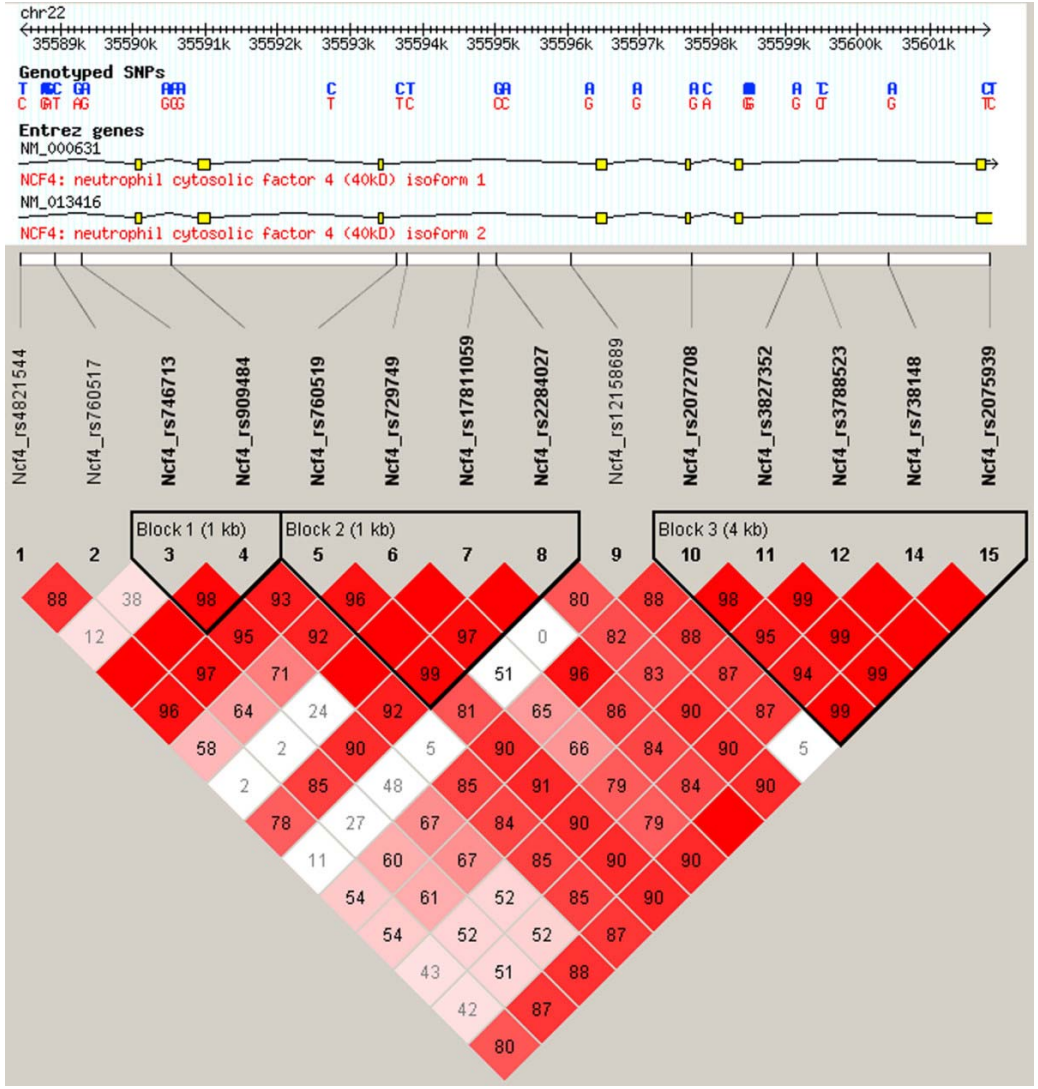
3.1 million human SNPs genotyped

~25–35% of common SNP variation in the populations surveyed

TIHC. Nature 449:851 2007

GWAS prerequisites

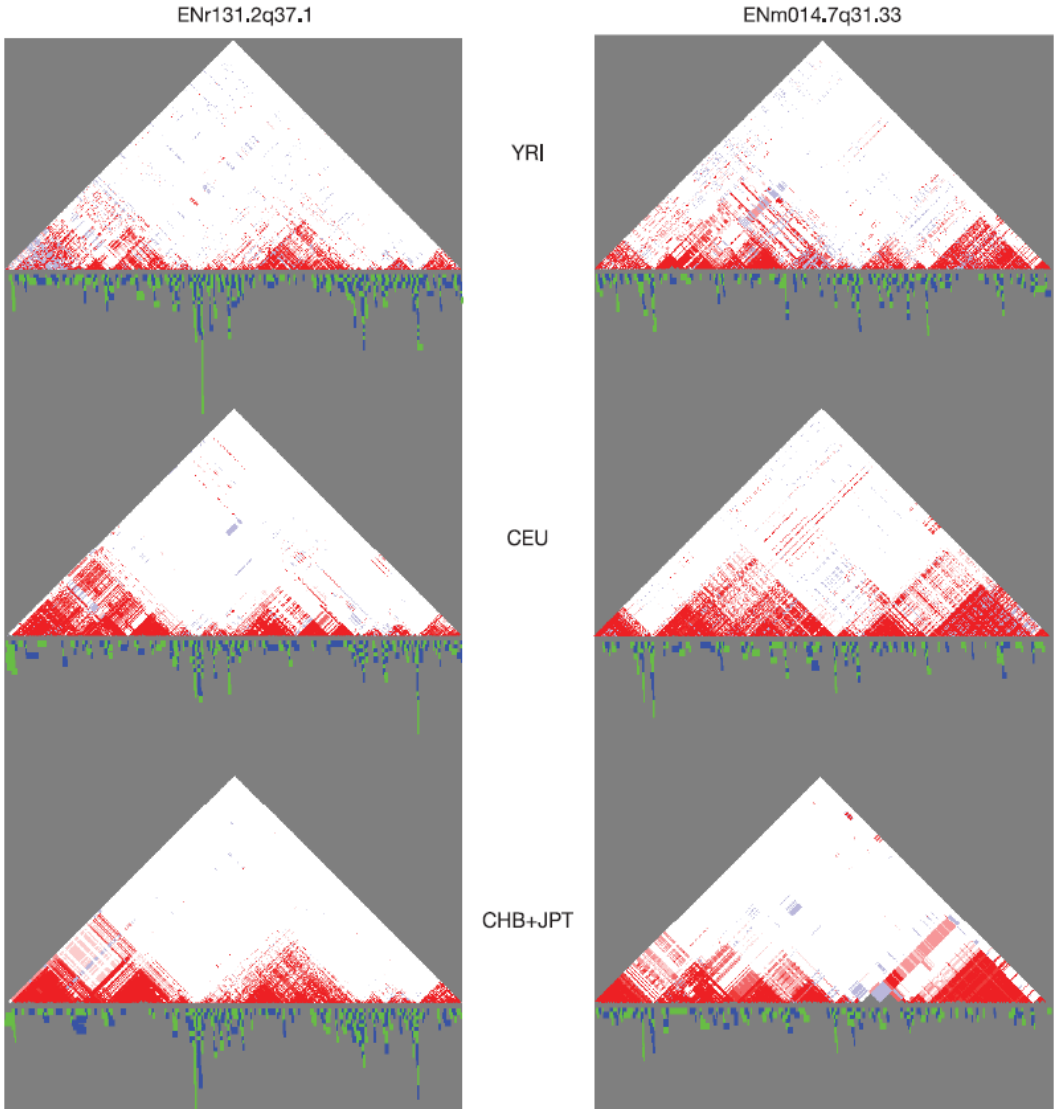
Haplotype blocks



Olsson et al. Arthritis Res & Ther 9:r98 2007

GWAS prerequisites

Haplotype blocks



Africans

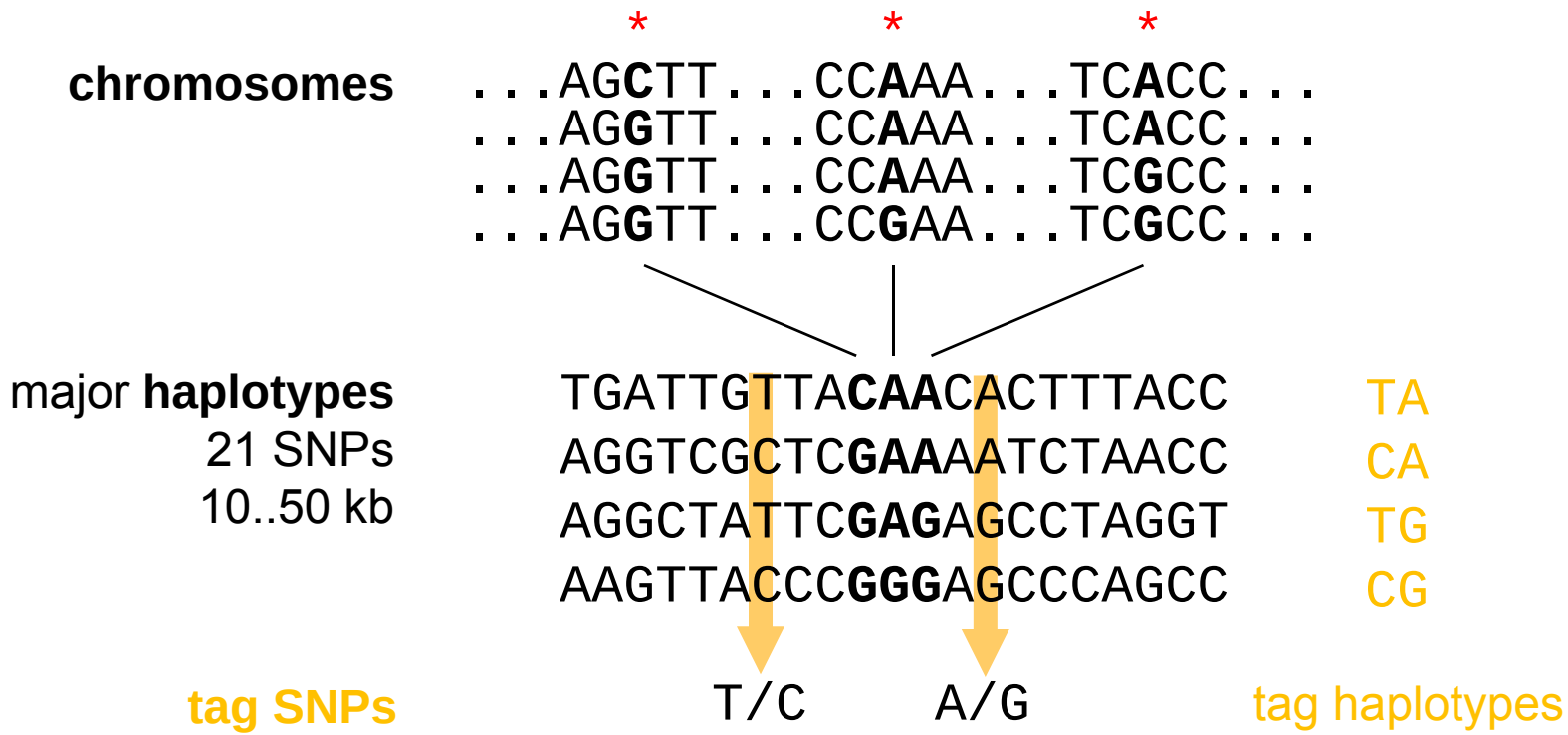
Europeans

Asians

TIHC. Nature 437:1299 2005

Haplotype

= set/region physically linked polymorphism



GWAS prerequisites

HapMap Project

Conclusions

- HapMap is estimated to capture the 65-75% untyped common SNPs with a likelihood of 90-96% depending on population (average maximum r^2).
- Current generation of commercial genome-wide genotyping products captures 3.1 million HapMap SNPs with 80% in African and up to 95% in non-African populations.

TIHC. Nature 449:851 2007

GWAS prerequisites

Large, well-phenotyped study groups

WTCCC Wellcome Trust Case Control Consortium (GB)

17,000 samples

- 2,000 from each of seven diseases
  - *type 1 diabetes, type 2 diabetes, coronary heart disease, hypertension, bipolar disorder, rheumatoid arthritis, Crohn's disease*
- 3,000 controls also from England, Scotland and Wales

KORA Kooperative Gesundheitsforschung in der Region Augsburg

20,000 samples since 1985 *coronary heart disease*

POPGEN Schleswig- holstein Biobank für eine Medizin der Zukunft

since 2003 aiming at 30,000 controls + study groups for:

- aging *3,000 centenarians*
- diseases *inflammation, heart, cancer, nervous system*

GWAS prerequisites

Large, well-phenotyped study groups

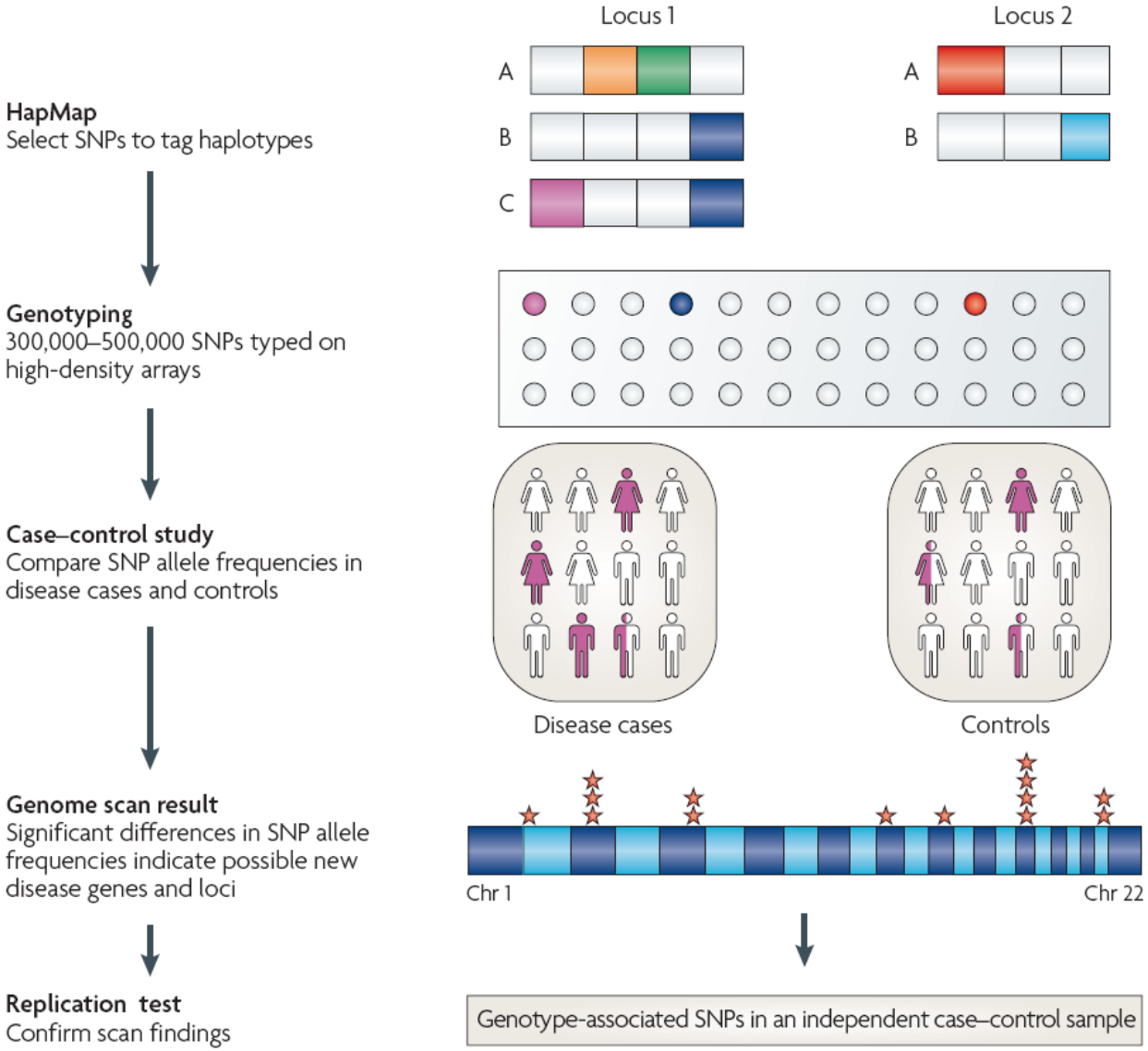
GEHA Genetics of Healthy Aging (EU 2004-9)

2,650 90^{+} sib-pairs 5,300 samples

2,650 young ethnically matched controls

GWAS

Scheme



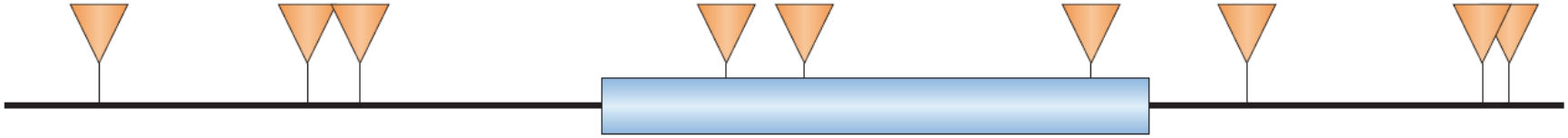
Mathew. Nat Rev Genet 9:9 2008

GWAS

Alternative designs

Direct:
catalogue and test all functional variants for association

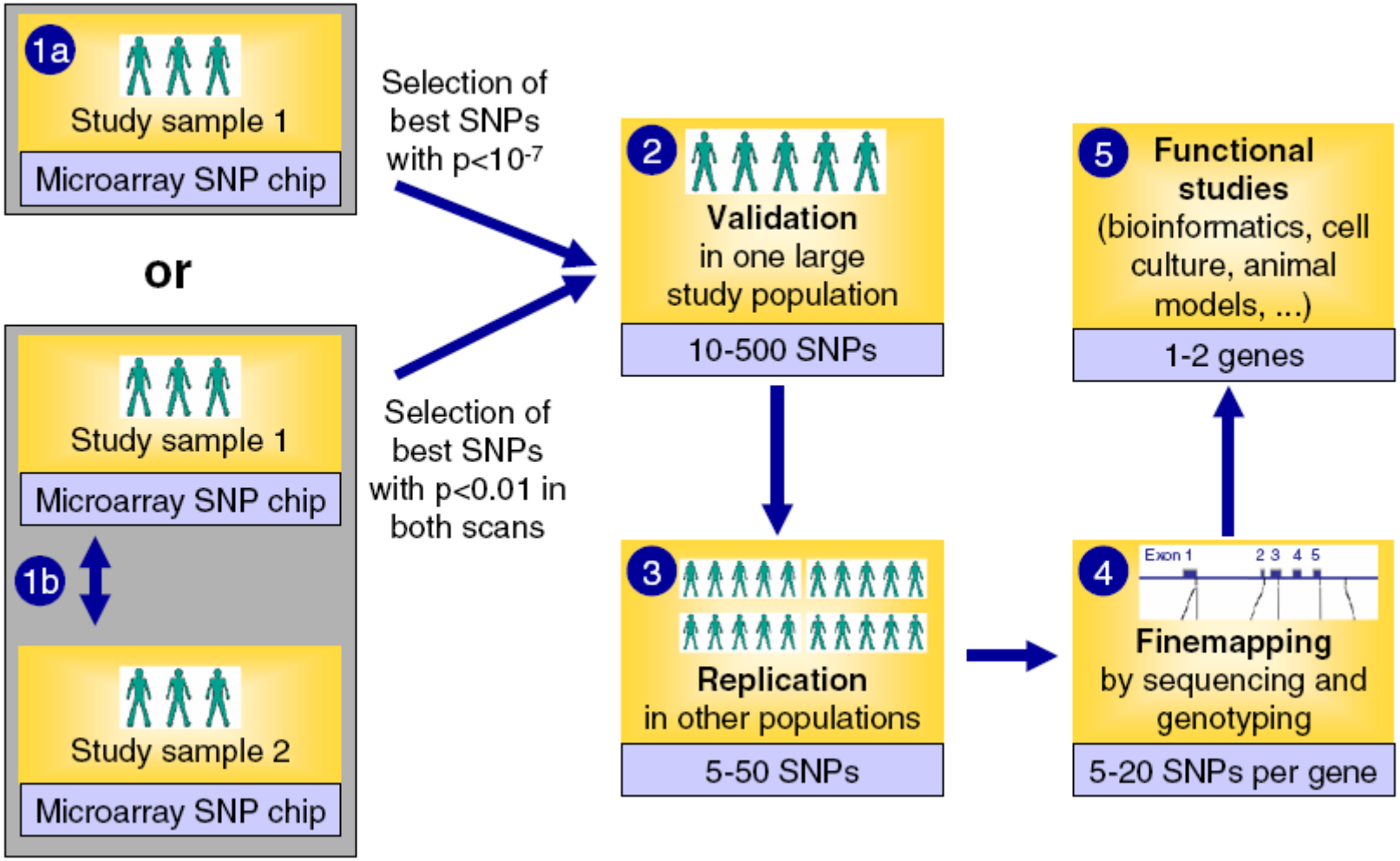
Indirect:
use a dense SNP map and test for linkage disequilibrium



Kruglyak. Nat Rev Genet 9:314 2008

GWAS

Statistical design & follow-up



Kronenberg. Exp Geront 43:39 2008

GWAS

Current stage

carried-out GWAS: 180
within last 12 month: 100

Identified many novel genes involved in complex diseases.

Some genes are associated with several phenotypes.

A genome-wide scan for linkage to human exceptional longevity identifies a locus on chromosome 4

Annibale A. Puca*†, Mark J. Daly‡, Stephanie J. Brewster§, Tara C. Matise¶, Jeffrey Barrett‡, Maureen Shea-Drinkwater‖, Sammy Kang¶, Erin Joyce§, Julie Nicoli*, Erica Benson§, Louis M. Kunkel*, and Thomas Perls‖

Contributed by Louis M. Kunkel, July 2, 2001

Substantial evidence supports the familial aggregation of exceptional longevity. The existence of rare families demonstrating clustering for this phenotype suggests that a genetic etiology may be an important component. Previous attempts at localizing loci predisposing for exceptional longevity have been limited to association studies of candidate gene polymorphisms. In this study, a genome-wide scan for such predisposing loci was conducted by using 308 individuals belonging to 137 sibships demonstrating exceptional longevity. By using nonparametric analysis, significant evidence for linkage was noted for chromosome 4 at D4S1564 with a MLS of 3.65 ($P = 0.044$). The analysis was corroborated by a parametric analysis ($P = 0.052$). These linkage results indicate the likelihood that there exists a gene, or genes, that exerts a substantial influence on the ability to achieve exceptional old age. Identification of the genes in humans that allow certain individuals to live to extreme old age should lead to insights on cellular pathways that are important to the aging process.

1st genome-wide linkage scan

Samples & Methods

Samples:

308 individuals from 137 sibships with exceptional longevity

- 98 years for at least one member of the sibship (the proband)
- siblings male >91 years; >95 years (sample)
- represents 5% oldest individuals in the birth cohort based on U.S. and Canadian life tables
- predominantly of European descent

Linkage Analysis:

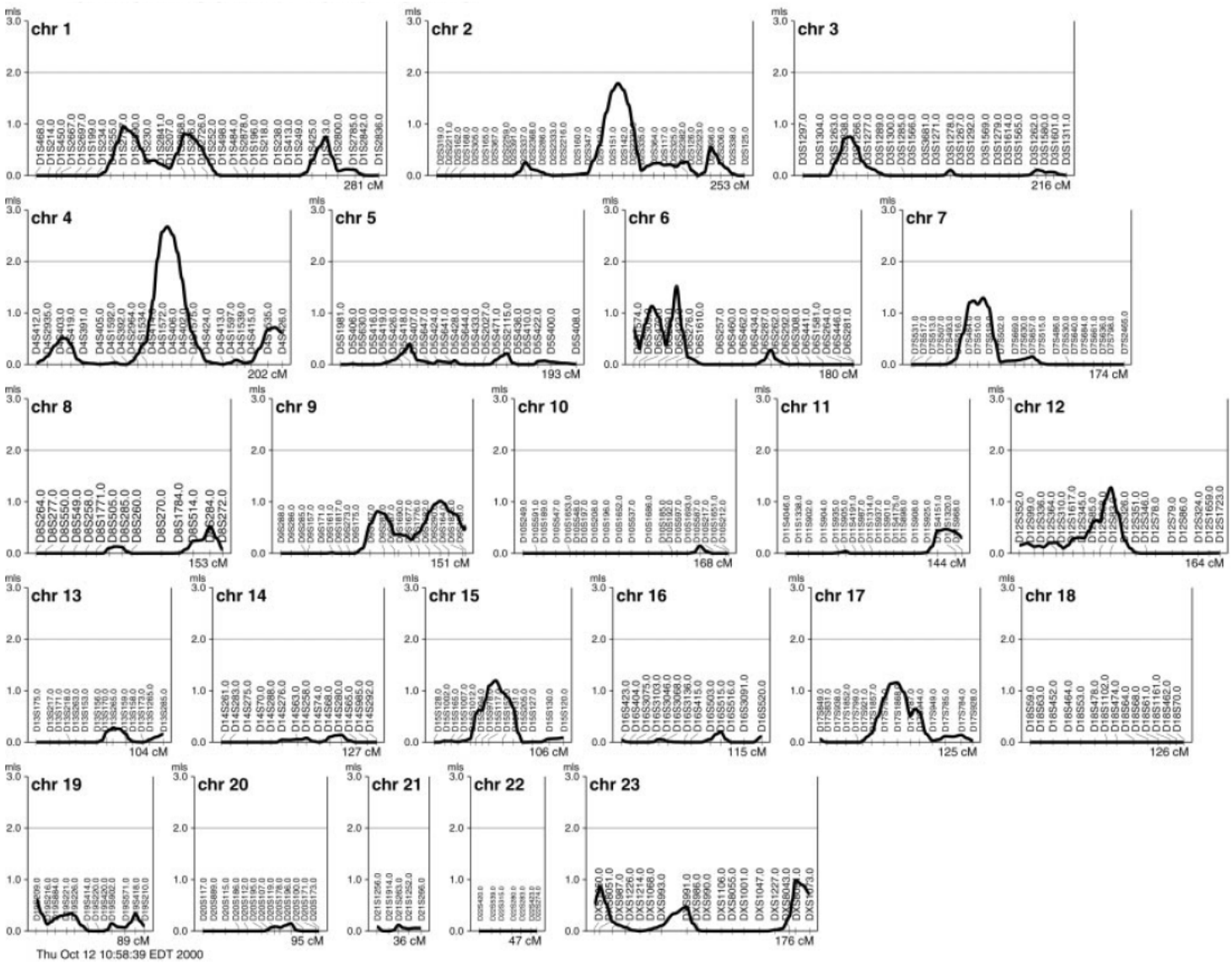
ABI Prism Linkage Mapping Set, Version 2

400 microsatellite markers

Puca et al. PNAS 98:10505 2001

1st genome-wide linkage scan

Results



Puca et al. PNAS 98:10505 2001

1st genome-wide linkage scan

Results

Of the 4-fold risk to siblings of centenarians (λs) to achieve at least their early nineties, the degree of excess allele sharing indicates that a locus in the D4S1564 region could explain 1.65-fold of that risk,

Puca et al. PNAS 98:10505 2001

BMC Medical Genetics

BioMed Central

Research

Open Access

Genetic correlates of longevity and selected age-related phenotypes: a genome-wide association study in the Framingham Study

Kathryn L Lunetta[1,3], Ralph B D'Agostino Sr[1,5], David Karasik[4], Emelia J Benjamin[1,2], Chao-Yu Guo[1,2], Raju Govindaraju[1,2], Douglas P Kiel[4], Margaret Kelly-Hayes[1,2], Joseph M Massaro[1,5], Michael J Pencina[1,5], Sudha Seshadri[1,2] and Joanne M Murabito*[1,2]

Address: [1]The National Heart Lung and Blood Institute's Framingham Heart Study, Framingham, MA, USA, [2]Section of General Internal Medicine and the Departments of Neurology, Cardiology, and Preventive Medicine, Boston University School of Medicine, Boston, MA, USA, [3]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA, [4]Hebrew Senior Life Institute for Aging Research and Harvard Medical School, Boston, MA, USA and [5]Statistics and Consulting Unit, Department of Mathematics, Boston University, Boston, MA, USA

Email: Kathryn L Lunetta - klunetta@bu.edu; Ralph B D'Agostino - ralph@bu.edu; David Karasik - karasik@hrca.harvard.edu; Emelia J Benjamin - emelia@bu.edu; Chao-Yu Guo - chaoyu@bu.edu; Raju Govindaraju - drgraju@bu.edu; Douglas P Kiel - kiel@hrca.harvard.edu; Margaret Kelly-Hayes - mkhayes@bu.edu; Joseph M Massaro - jmm@bu.edu; Michael J Pencina - mpencina@bu.edu; Sudha Seshadri - suseshad@bu.edu; Joanne M Murabito* - murabito@bu.edu

* Corresponding author

Published: 19 September 2007

BMC Medical Genetics 2007, **8**(Suppl 1):S13 doi:10.1186/1471-2350-8-S1-S13

Framingham GWAS

Study sample

Framingham Heart Study (FHS)

- longitudinal family-based community sample
- participants have been well-characterized throughout adulthood
  - prospectively ascertained risk factors
  - diseases
- continuously followed until death

Lunetta et al. BMC Med Genet 8:s13 2007

Framingham GWAS

Study sample

- community-based sample: 1,345 Framingham Study participants
  - 330 largest families
  - original cohort: 258
  - offspring: 1,087
  - 149 deaths
  - 713 participants age ≥65
- 5 longevity and aging traits
  - age at death /mean 83 [46..99]
  - morbidity-free survival at age 65 years /CVD, dementia, and cancer
  - age at natural menopause /mean 50.2 [38..57]
  - walking speed
  - biologic age by osseographic scoring system
- Affymetrix 100K SNP GeneChip
  - 70,987 autosomal SNPs (genotypic call rate ≥80%, minor allele frequency ≥10%, Hardy-Weinberg test p ≥ 0.001)

Lunetta et al. BMC Med Genet 8:s13 2007

Framingham GWAS

Result

None of the associations achieved genome-wide significance

These data generate hypotheses and serve as a resource for replication as more genes and biologic pathways are proposed as contributing to longevity and healthy aging

Lunetta et al. BMC Med Genet 8:s13 2007

Framingham GWAS

Add-ons

- simple low p-value SNP ranking strategy
- SNP selection due to associations with more than one related phenotype
  - age at death & morbidity-free survival at age 65
  - biologic age and walking speed
- SNP associations within 79 candidate genes and regions
  - NCBI search term "longevity"
  - Science of Aging Knowledge http://sageke.sciencemag.org/cgi/genesdb

Lunetta et al. BMC Med Genet 8:s13 2007

Framingham GWAS

Discussion

FOXO - forkhead box group O transcription factors

- targets of insulin-like signaling
- involved in DNA repair and resistance to oxidative stress
- *FOXO1A* - increased mortality attributable to diabetes related deaths in participants aged ≥85
- *FOXO3A* - age at natural menopause; implicated in oocyte death, depletion of functioning ovarian follicles, and infertility in mice

SOX5 - potentially related to musculoskeletal function

WRN - Werner Syndrome

- longitudinal study of ageing Danish twins: possible association between a successful aging trait and 3 SNPs in *WRN*

KL (Klotho)

- in mouse lead to a syndrome resembling human aging
- functional variant linked to human longevity

Lunetta et al. BMC Med Genet 8:s13 2007

Framingham GWAS

Limitations

- systematic DNA collection began 1995 and hence the GWAS participants are likely healthier than the full FHS sample
- priori candidate genes without any SNP within 60 kb on the chip: *ACE*, *LAMINA*, *SIRT2 and SIRT3*
- epistasis or gene-environment interactions not examined

Lunetta et al. BMC Med Genet 8:s13 2007

Genome-wide Association Analysis Reveals Putative Alzheimer's Disease Susceptibility Loci in Addition to *APOE*

Lars Bertram,[1,6] Christoph Lange,[2,6] Kristina Mullin,[1] Michele Parkinson,[1] Monica Hsiao,[1] Meghan F. Hogan,[1] Brit M.M. Schjeide,[1] Basavaraj Hooli,[1] Jason DiVito,[1] Iuliana Ionita,[2] Hongyu Jiang,[2] Nan Laird,[2] Thomas Moscarillo,[4] Kari L. Ohlsen,[5] Kathryn Elliott,[5] Xin Wang,[5] Diane Hu-Lince,[5] Marie Ryder,[5] Amy Murphy,[2] Steven L. Wagner,[5] Deborah Blacker,[3,4] K. David Becker,[5] and Rudolph E. Tanzi[1,*]

Alzheimer's disease (AD) is a genetically complex and heterogeneous disorder. To date four genes have been established to either cause early-onset autosomal-dominant AD (*APP*, *PSEN1*, and *PSEN2*[1–4]) or to increase susceptibility for late-onset AD (*APOE*[5]). However, the heritability of late-onset AD is as high as 80%,[6] and much of the phenotypic variance remains unexplained to date. We performed a genome-wide association (GWA) analysis using 484,522 single-nucleotide polymorphisms (SNPs) on a large (1,376 samples from 410 families) sample of AD families of self-reported European descent. We identified five SNPs showing either significant or marginally significant genome-wide association with a multivariate phenotype combining affection status and onset age. One of these signals ($p = 5.7 \times 10^{-14}$) was elicited by SNP rs4420638 and probably reflects *APOE*-ε4, which maps 11 kb proximal ($r^2 = 0.78$). The other four signals were tested in three additional independent AD family samples composed of nearly 2700 individuals from almost 900 families. Two of these SNPs showed significant association in the replication samples (combined p values 0.007 and 0.00002). The SNP (rs11159647, on chromosome 14q31) with the strongest association signal also showed evidence of association with the same allele in GWA data generated in an independent sample of ~1,400 AD cases and controls ($p = 0.04$). Although the precise identity of the underlying locus(i) remains elusive, our study provides compelling evidence for the existence of at least one previously undescribed AD gene that, like *APOE*-ε4, primarily acts as a modifier of onset age.

[1]Genetics and Aging Research Unit, Mass General Institute for Neurodegenerative Disease (MIND), Department of Neurology, Massachusetts General Hospital, Charlestown, MA 02129, USA; [2]Department of Biostatistics, [3]Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA; [4]Gerontology Research Unit, Department of Psychiatry, Massachusetts General Hospital, Charlestown, MA 02129, USA; [5]TorreyPines Therapeutics, La Jolla, CA 92037, USA
[6]These authors contributed equally to this work
*Correspondence: tanzi@helix.mgh.harvard.edu
DOI 10.1016/j.ajhg.2008.10.008.

GWAS of Alzheimer's disease

Samples & Results

Samples:

1,376 samples from 410 AD families

self reported European descent

GWAS:

GeneChip Human Mapping 500K Array Set (Affymetrix)

500,668 SNPs

Results:

5 SNPs significantly associated

rs4420638 11 kb proximal *APOE-ε4*

Replication:

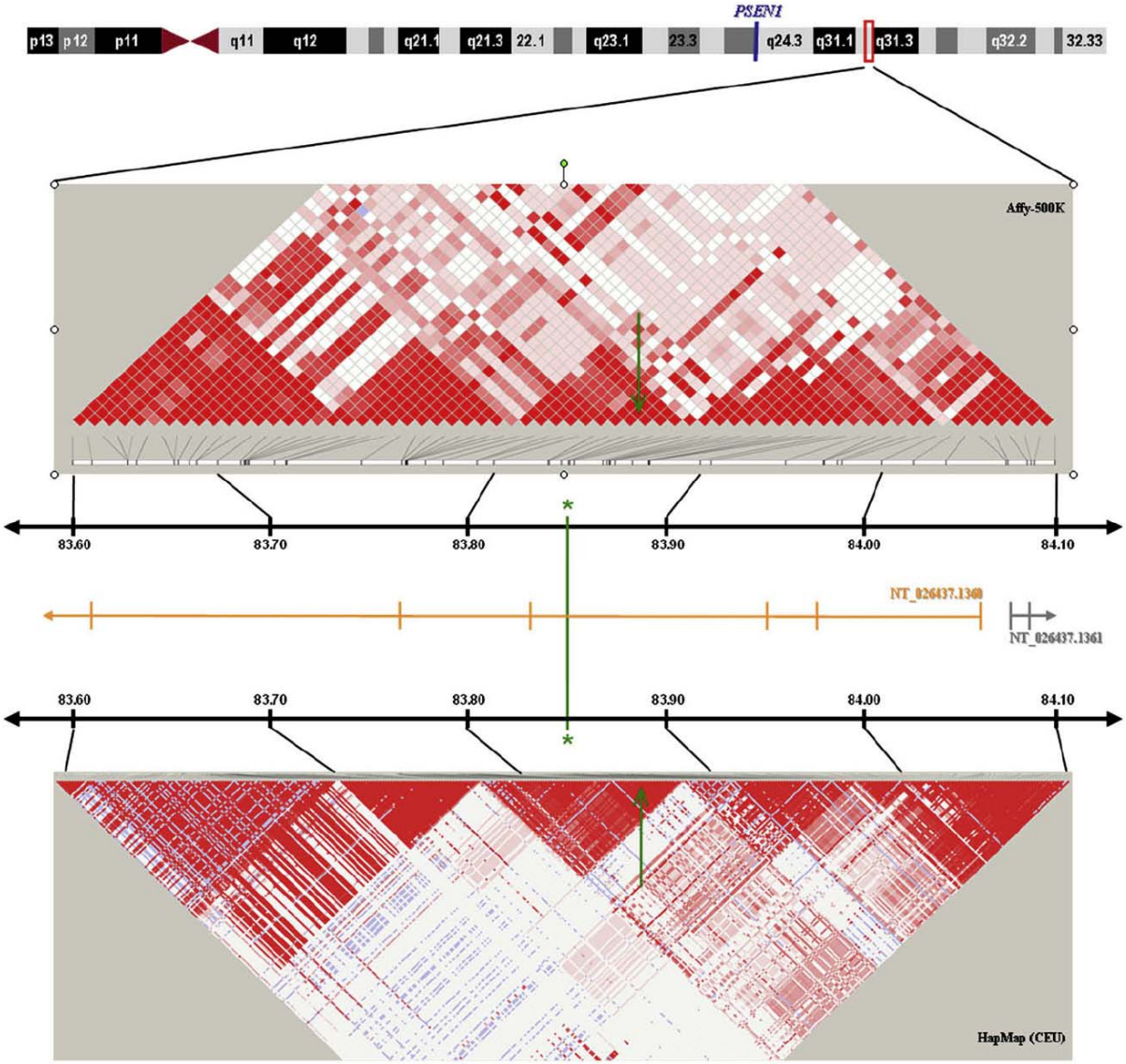
2 SNPs significant 2,700 individuals from almost 900 families

1 SNP significant 1,400 cases & controls

Bertram et al. Am J Hum Genet Epub Oct 29 2008

GWAS of Alzheimer's disease

14q31 association signal



Bertram et al. Am J Hum Genet Epub Oct 29 2008

GWAS of Alzheimer's disease

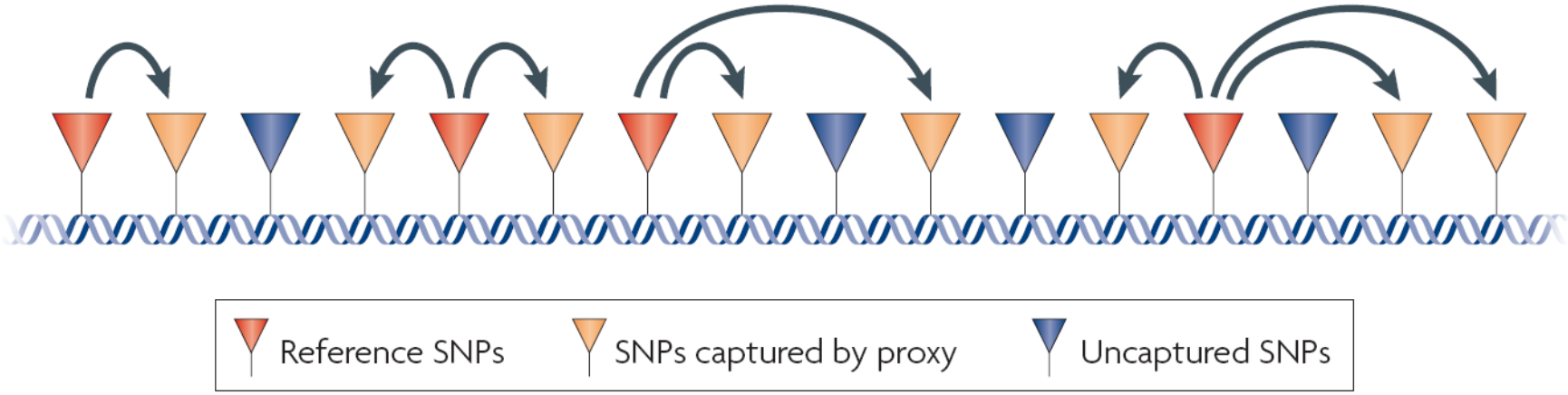
Conclusion

Existence of at least one previously undescribed AD gene that, like *APOE-ε4*, primarily acts as a modifier of onset age

Bertram et al. Am J Hum Genet Epub Oct 29 2008

GWAS

Array-based open questions



Kruglyak. Nat Rev Genet 9:314 2008

GWAS

Outlook

Problem

Case & control sampling !

Question

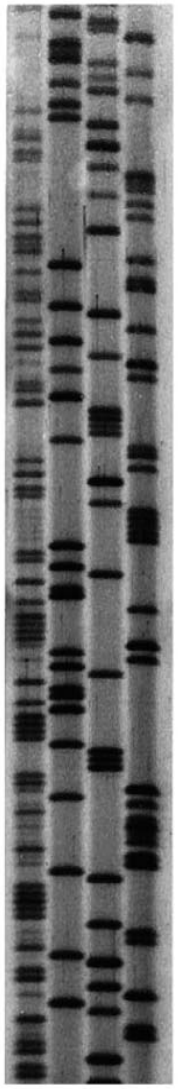
Are there high-frequency, small-effect polymorphisms affecting healthy aging & longevity ?

NIA sponsored Longevity Consortium

http://www.longevityconsortium.org

opportunity of collaboration with other investigators to replicate important findings in additional cohorts

Lunetta et al. BMC Med Genet 8:s13 2007



genome.fli-leibniz.de
Teaching