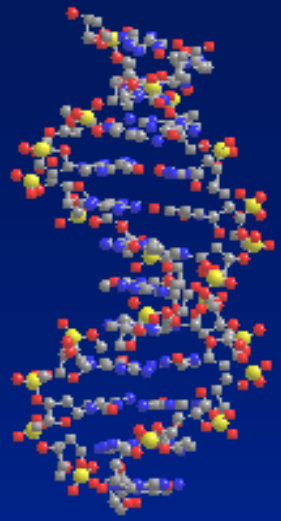
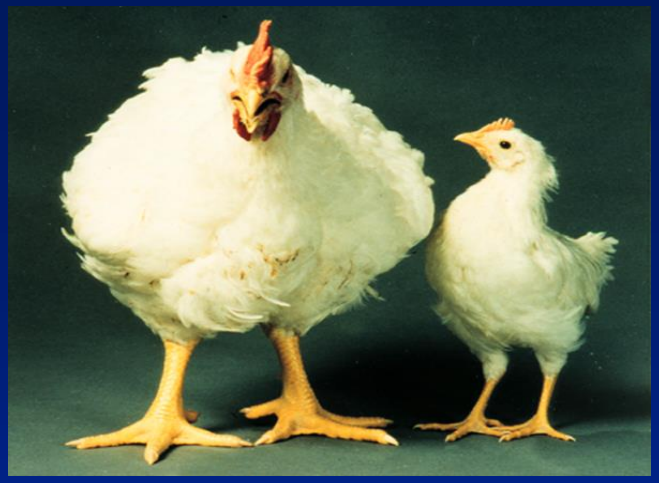
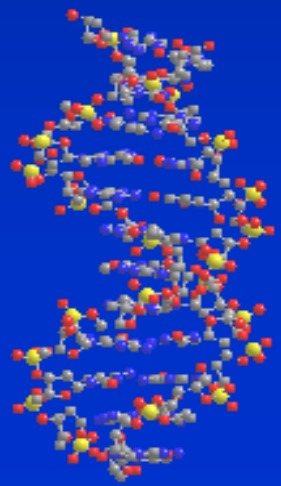
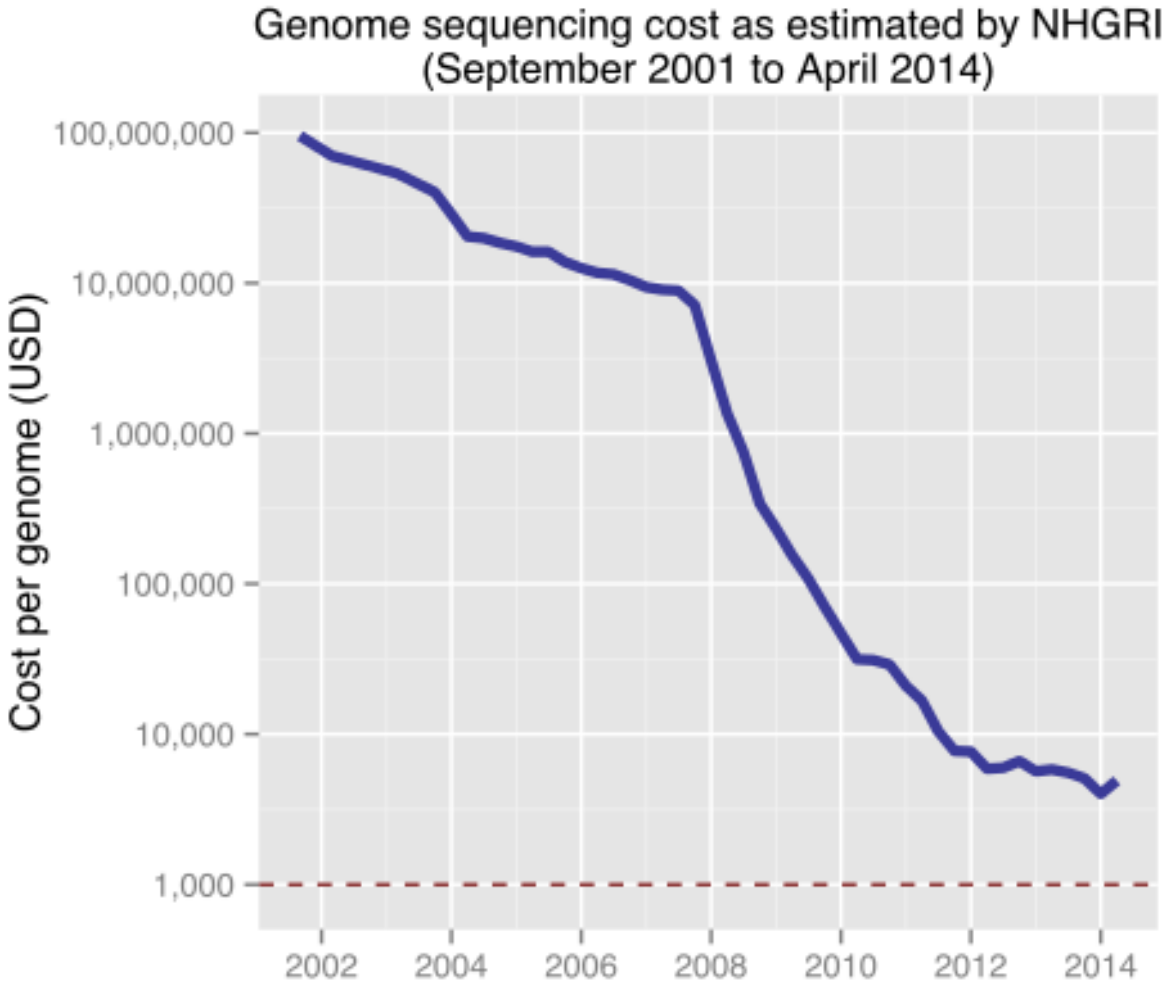


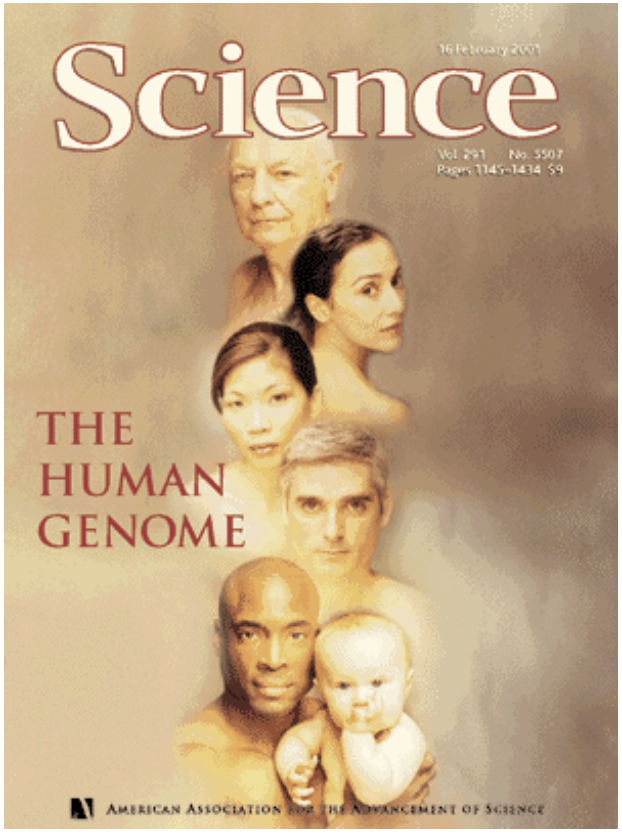


From sequence data to genomic prediction




INFUSE
DEFIER


Genome sequencing cost as estimated by NHGRI
(September 2001 to April 2014)
Cost per genome (USD)
100,000,000
10,000,000
1,000,000
100,000
10,000
1,000
2002
2004
2006
2008
2010
2012
2014


Science
THE HUMAN GENOME
AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

Course overview

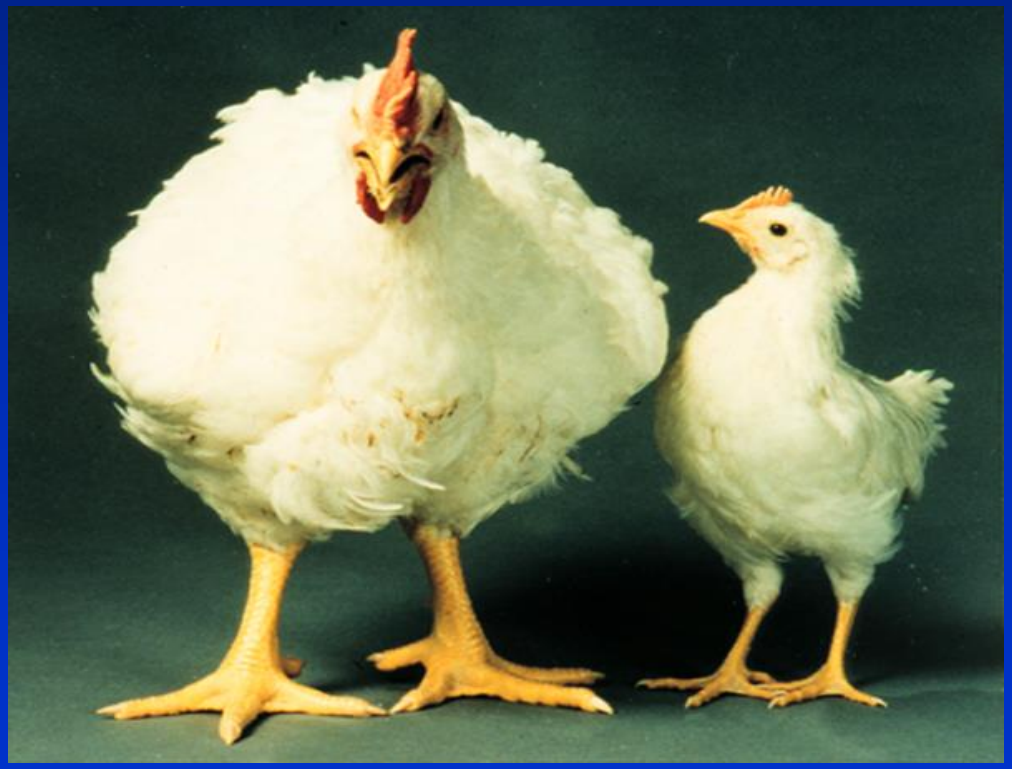
- Day 1
  - Introduction
  - Generation, quality control, alignment of sequence data
  - Detection of variants, quality control and filtering
- Day 2
  - Imputation from SNP array genotypes to sequence data
- Day 3
  - Genome wide association studies with SNP array and sequence variant genotypes
- Day 4 & 5
  - Genomic prediction with SNP array and sequence variant genotypes (BLUP and Bayesian methods)
  - Use of genomic selection in breeding programs

Course overview

- Day 1
  - Introduction
  - Generation, quality control, alignment of sequence data
  - Detection of variants, quality control and filtering
- Day 2
  - Imputation from SNP array genotypes to sequence data
- Day 3
  - Genome wide association studies with SNP array and sequence variant genotypes
- Day 4 & 5
  - Genomic prediction with SNP array and sequence variant genotypes (BLUP and Bayesian methods)
  - Use of genomic selection in breeding programs

Quantitative traits

- Genetic variation observed for many (all?) traits of economic importance in livestock and plant species
- One gene or many?



Yield in Rice

nature genetics

Genome-wide association studies of 14 agronomic traits in rice landraces

Xuehui Huang[1,2,10], Xinghua Wei[3,10], Tao Sang[4,10], Qiang Zhao[1,2,10], Qi Feng[1,10], Yan Zhao[1], Canyang Li[1], Chuanrang Zhu[1], Tingting Lu[1], Zhiwu Zhang[5], Meng Li[5,6], Danlin Fan[1], Yunli Guo[1], Ahong Wang[1], Lu Wang[1], Liuwei Deng[1], Wenjun Li[1], Yiqi Lu[1], Qijun Weng[1], Kunyan Liu[1], Tao Huang[1], Taoying Zhou[1], Yufeng Jing[1], Wei Li[1], Zhang Lin[1], Edward S Buckler[5,7], Qian Qian[3], Qi-Fa Zhang[8], Jiayang Li[9] & Bin Han[1,2]

Uncovering the genetic basis of agronomic traits in crop landraces that have adapted to various agro-climatic conditions is important to world food security. Here we have identified ~3.6 million SNPs by sequencing 517 rice landraces and constructed a high-density haplotype map of the rice genome using a novel data-imputation method. We performed genome-wide association studies (GWAS) for 14 agronomic traits in the population of *Oryza sativa indica* subspecies. The loci identified through GWAS explained ~36% of the phenotypic variance, on average. The peak signals at six loci were tied closely to previously identified genes. This study provides a fundamental resource for rice genetics research and breeding, and demonstrates that an approach integrating second-generation genome sequencing and GWAS can be used as a powerful complementary strategy to classical biparental cross-mapping for dissecting complex traits in rice.

Yield in Rice

nature genetics

"our results suggest that multiple loci with relatively small effects contribute to the phenotypic variance"

Genome-wide association studies of 14 agronomic traits in rice landraces

Xuehui Huang[1,2,10], Xinghua Wei[3,10], Tao Sang[4,10], Qiang Zhao[1,2,10], Qi Feng[1,10], Yan Zhao[1], Canyang Li[1], Chuanrang Zhu[1], Tingting Lu[1], Zhiwu Zhang[5], Meng Li[5,6], Danlin Fan[1], Yunli Guo[1], Ahong Wang[1], Lu Wang[1], Liuwei Deng[1], Wenjun Li[1], Yiqi Lu[1], Qijun Weng[1], Kunyan Liu[1], Tao Huang[1], Taoying Zhou[1], Yufeng Jing[1], Wei Li[1], Zhang Lin[1], Edward S Buckler[5,7], Qian Qian[3], Qi-Fa Zhang[8], Jiayang Li[9] & Bin Han[1,2]

Uncovering the genetic basis of agronomic traits in crop landraces that have adapted to various agro-climatic conditions is important to world food security. Here we have identified ~3.6 million SNPs by sequencing 517 rice landraces and constructed a high-density haplotype map of the rice genome using a novel data-imputation method. We performed genome-wide association studies (GWAS) for 14 agronomic traits in the population of *Oryza sativa indica* subspecies. The loci identified through GWAS explained ~36% of the phenotypic variance, on average. The peak signals at six loci were tied closely to previously identified genes. This study provides a fundamental resource for rice genetics research and breeding, and demonstrates that an approach integrating second-generation genome sequencing and GWAS can be used as a powerful complementary strategy to classical biparental cross-mapping for dissecting complex traits in rice.

Human height

NATURE GENETICS | ARTICLE

日本語要約

Defining the role of common variation in the genomic and biological architecture of adult human height

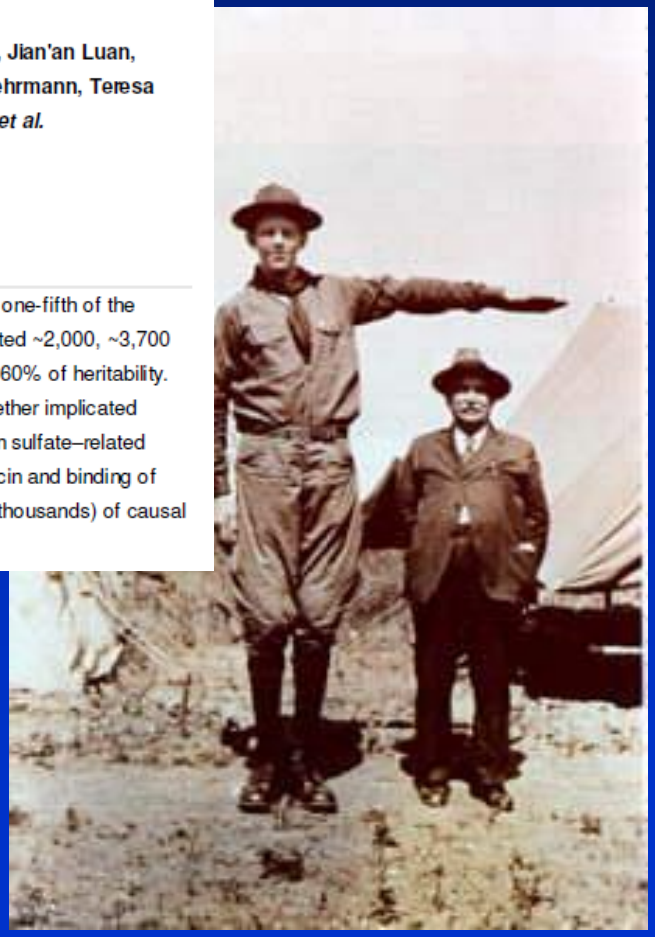
Andrew R Wood, Tonu Esko, Jian Yang, Sailaja Vedantam, Tune H Pers, Stefan Gustafsson, Audrey Y Chu, Karol Estrada, Jian'an Luan, Zoltán Kutalik, Najaf Amin, Martin L Buchkovich, Damien C Croteau-Chonka, Felix R Day, Yanan Duan, Tove Fall, Rudolf Fehrmann, Teresa Ferreira, Anne U Jackson, Juha Karjalainen, Ken Sin Lo, Adam E Locke, Reedik Mägi, Evelin Mihailov, Eleonora Porcu ***et al.***

Nature Genetics **46**, 1173–1186 (2014) doi:10.1038/ng.3097

Received 18 December 2013 Accepted 29 August 2014 Published online 05 October 2014

Abstract

Using genome-wide data from 253,288 individuals, we identified 697 variants at genome-wide significance that together explained one-fifth of the heritability for adult height. By testing different numbers of variants in independent studies, we show that the most strongly associated ~2,000, ~3,700 and ~9,500 SNPs explained ~21%, ~24% and ~29% of phenotypic variance. Furthermore, all common variants together captured 60% of heritability. The 697 variants clustered in 423 loci were enriched for genes, pathways and tissue types known to be involved in growth and together implicated genes and pathways not highlighted in earlier efforts, such as signaling by fibroblast growth factors, WNT/β-catenin and chondroitin sulfate–related genes. We identified several genes and pathways not previously connected with human skeletal growth, including mTOR, osteoglycin and binding of hyaluronic acid. Our results indicate a genetic architecture for human height that is characterized by a very large but finite number (thousands) of causal variants.



The case of the missing heritability

When scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. **Brendan Maher** shines a light on six places where the missing loot could be stashed away.

If you want to predict how tall your children might one day be, a good bet would be to look in the mirror, and at your mate. Studies going back almost a century have

Even though these genome-wide association studies (GWAS) turned up dozens of variants, they did "very little of the prediction that you would do just by asking people how tall their parents are", says Joel Hirschhorn at the Broad Institute in Cambridge, Massachusetts, who led one of the studies[3].

contribute to a variety of traits and common diseases. But even when dozens of genes have been linked to a trait, both the individual and cumulative effects are disappointingly small and nowhere near enough to explain earlier estimates of heritability. "It is the big topic in the genetics of common disease right

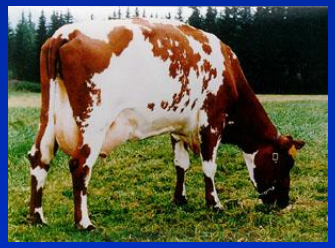
ILLUSTRATIONS BY O. PARKINS

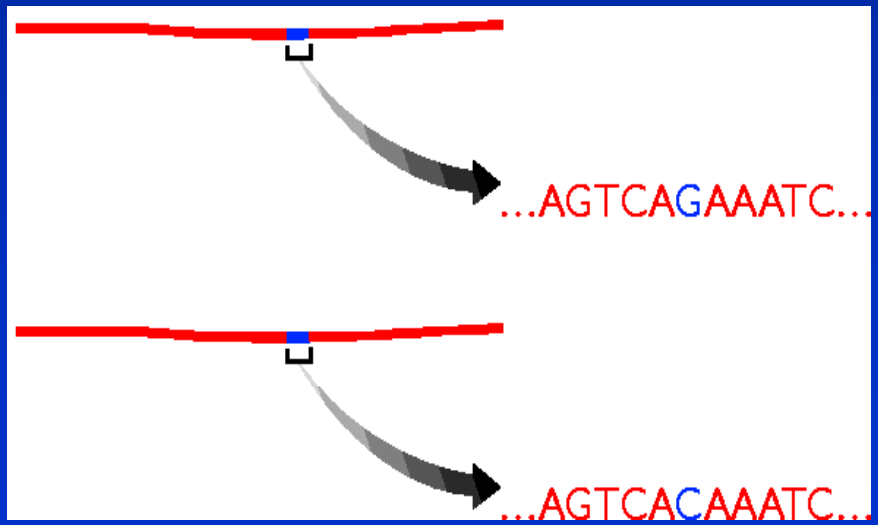
Quantitative traits

- Large number of causative mutations (quantitative trait loci, QTL) for most complex traits
- Variance explained by individual markers will be small
- Genome wide association studies -> powerful experiments!
- Genomic prediction -> Use large numbers of DNA markers to simultaneously track all QTL

The Revolution

- As a result of sequencing animal and plant genomes, have a huge amount of information on variation in the genome
  - at the DNA level
- Most abundant form of variation are Single Nucleotide Polymorphisms (SNPs)



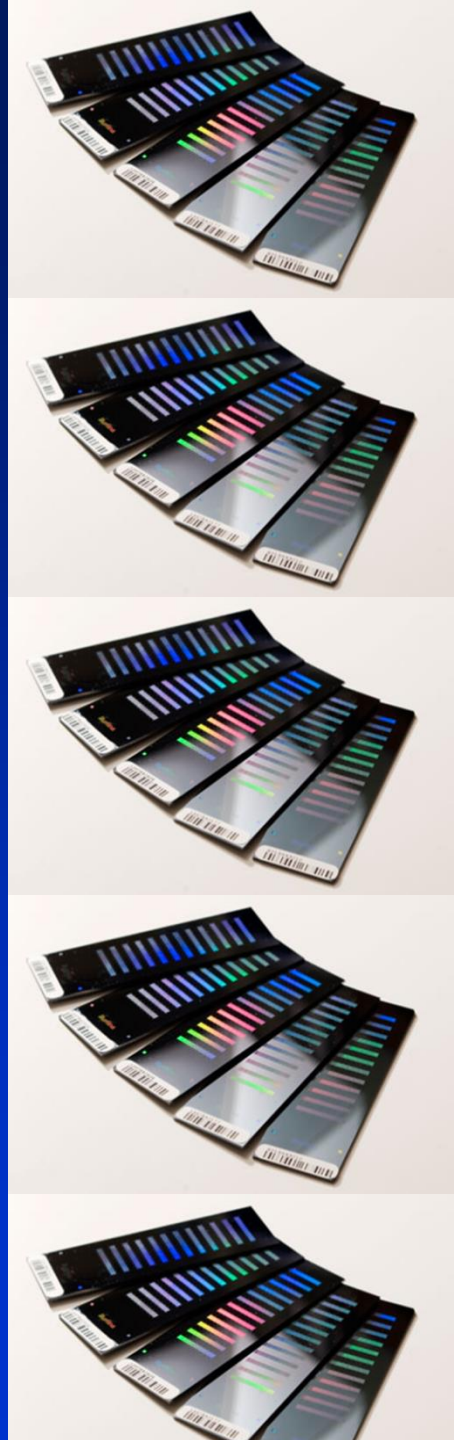


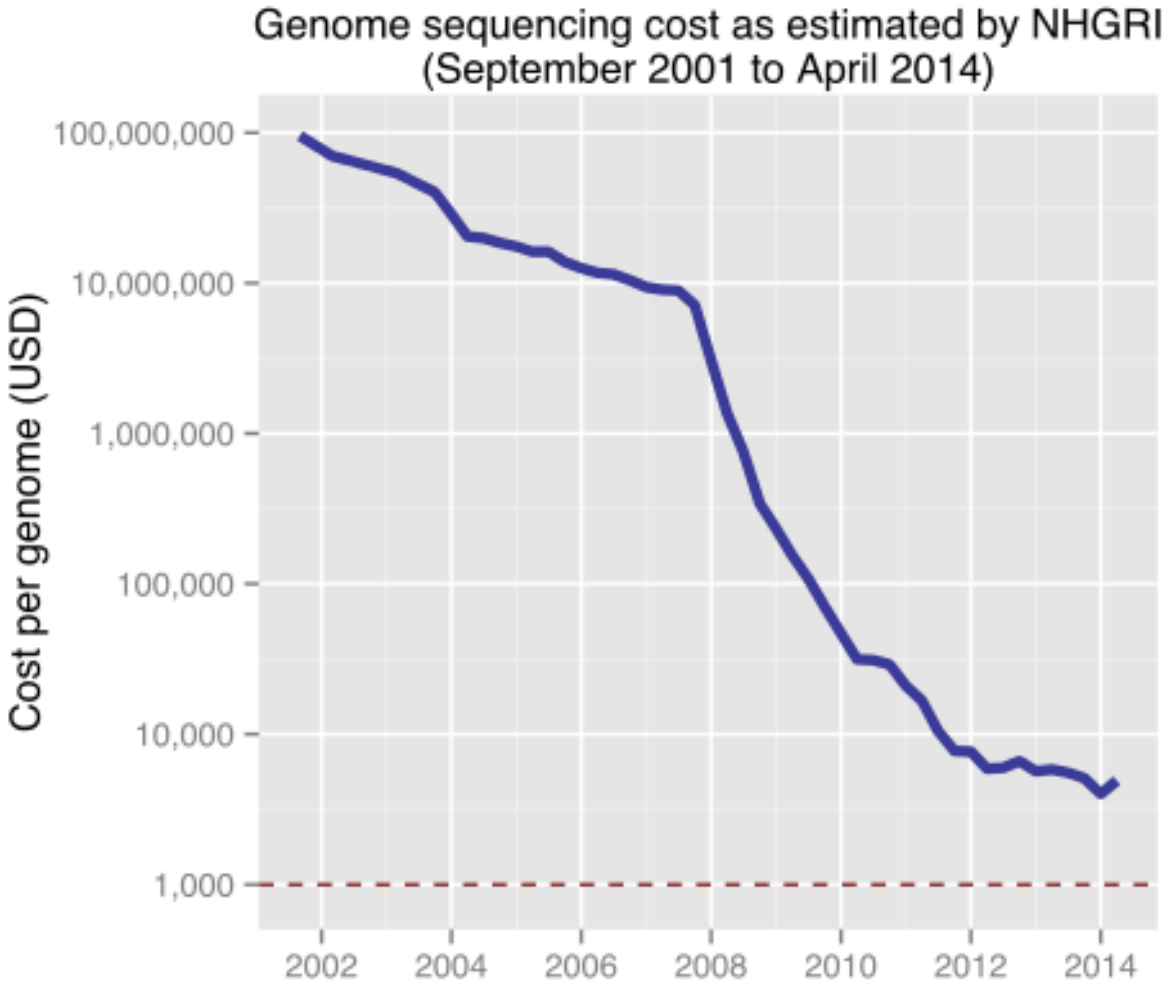


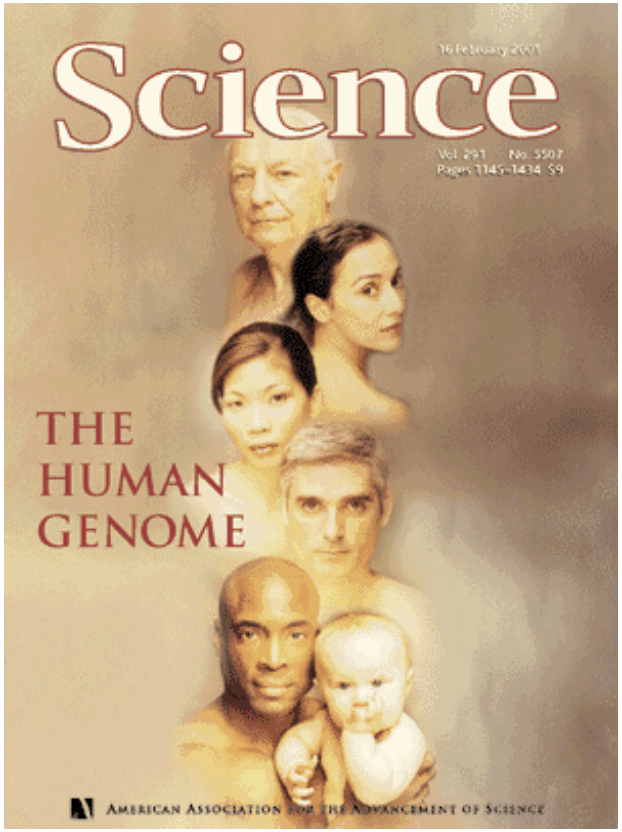
- 1000 Genomes project (Pilot)
- ~15 mill SNPs
- ~7 mill SNPs with minor allele >5%
- ~100,000-300,000 cSNPs
- ~50,000 nonsynonymous cSNPs -> change protein structure
- Every individual carries 250-300 loss of function mutations!

The Revolution

- SNP chips available for
  - Sheep, Cattle (50K, 800K), Pigs,
  - Chickens
  - Salmon
  - Horse, Dog
- Plants
  - Maize, Wheat
  - Cotton, Soybean under development
- Cost?
  - ~ $100-200 USD for 60K SNPs
- Genotyping by re-sequencing?
  - 40 million SNPs in cattle
  - Insertion deletions
  - Copy number variants?




Genome sequencing cost as estimated by NHGRI
(September 2001 to April 2014)
Cost per genome (USD)
100,000,000
10,000,000
1,000,000
100,000
10,000
1,000
2002
2004
2006
2008
2010
2012
2014


Science
THE HUMAN GENOME
AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

Sequence data vs SNP arrays

- Genome wide association study
  - Straight to causative mutation
- Genomic selection (all hypotheses!)
  - No longer have to rely on LD, causative mutation actually in data set
    - Higher accuracy of prediction?
    - Better prediction across breeds/populations?
    - Better persistence of accuracy across generations

- **But** have sequencing errors, genotype errors, expense…….

Aim

- Provide you with genome wide association and genomic prediction methodologies to exploit high density genotypes, up to whole genome sequence data, in livestock and plant improvement