

# Genome Wide Association Studies (GWAS) and Post GWAS

*Presenters: Diriba Chere*

*Ruolin Bian*

*Upasana Dhakal*

# Presentation outline

- Introduction
  - History
  - success
  - Challenges
- GWAS Application
- What after GWAS

# Introduction

- An approach that involves rapidly scanning markers across the complete sets of DNA, or genomes, of many individuals to find genetic variations associated with a particular trait
- Detecting variants at genomic loci that are associated with complex traits in the population
- Detecting associations between common single-nucleotide polymorphisms (SNPs) and common diseases

# Introduction



## NIH Public Access Author Manuscript

*Science*. Author manuscript; available in PMC 2006 July 18.

Published in final edited form as:

*Science*. 2005 April 15; 308(5720): 385–389. doi:10.1126/science.1109557.

## Complement Factor H Polymorphism in Age-Related Macular Degeneration

Robert J. Klein<sup>1</sup>, Caroline Zeiss<sup>2,\*</sup>, Emily Y. Chew<sup>3,\*</sup>, Jen-Yue Tsai<sup>4,\*</sup>, Richard S. Sackler<sup>1</sup>, Chad Haynes<sup>1</sup>, Alice K. Henning<sup>5</sup>, John Paul SanGiovanni<sup>3</sup>, Shrikant M. Mane<sup>6</sup>, Susan T. Mayne<sup>7</sup>, Michael B. Bracken<sup>7</sup>, Frederick L. Ferris<sup>3</sup>, Jurg Ott<sup>1</sup>, Colin Barnstable<sup>2</sup>, and Josephine Hoh.<sup>7,†</sup>

**“a common form of blindness is associated with variation in the gene for complement factor H, which produces a protein involved in regulating inflammation”.**  
Found two SNPs with significantly altered allele frequency compared to healthy controls.

# Introduction

## ***HTRA1* Promoter Polymorphism in Wet Age-Related Macular Degeneration**

Andrew DeWan,<sup>1</sup> Mugen Liu,<sup>2\*</sup> Stephen Hartman,<sup>3\*</sup> Samuel Shao-Min Zhang,<sup>2\*</sup> David T. L. Liu,<sup>4</sup> Connie Zhao,<sup>5</sup> Pancy O. S. Tam,<sup>4</sup> Wai Man Chan,<sup>4</sup> Dennis S. C. Lam,<sup>4</sup> Michael Snyder,<sup>3</sup> Colin Barnstable,<sup>2</sup> Chi Pui Pang,<sup>4</sup> Josephine Hoh<sup>1,2†</sup>

Age-related macular degeneration (AMD), the most common cause of irreversible vision loss in individuals aged older than 50 years, is classified as either wet (neovascular) or dry (nonneovascular). Inherited variation in the complement factor H gene is a major risk factor for drusen in dry AMD. Here we report that a single-nucleotide polymorphism in the promoter region of *HTRA1*, a serine protease gene on chromosome 10q26, is a major genetic risk factor for wet AMD. A whole-genome association mapping strategy was applied to a Chinese population, yielding a *P* value of  $<10^{-11}$ . Individuals with the risk-associated genotype were estimated to have a likelihood of developing wet AMD 10 times that of individuals with the wild-type genotype.

## ARTICLES

---

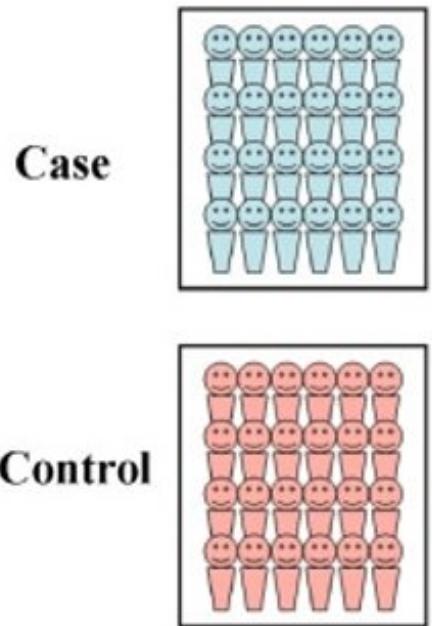
# **Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls**

The Wellcome Trust Case Control Consortium\*

The first large, well-designed GWAS for complex diseases to employ a SNP chip that had good coverage of the genome.

# Steps in GWAS

## Phenotyping



## Genotyping

Commercial array of SNPs

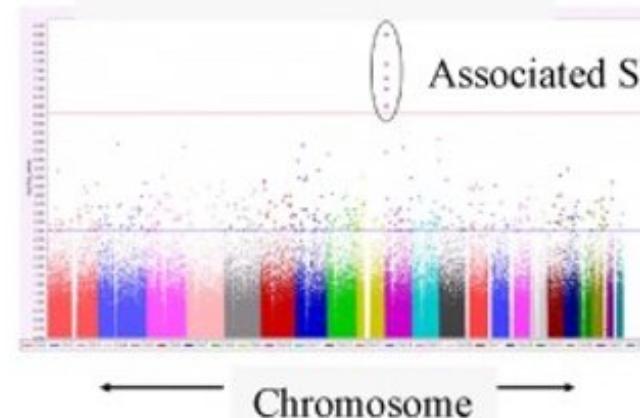


DNA

Information

## Mapping

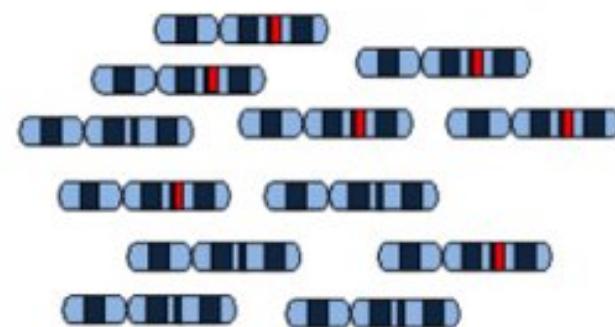
Associated SNP



## Statistics

$$\begin{aligned}
 W &= [1 - \Psi(\mu_2, 0)] \int_{\Phi^{-1}(1-\gamma/2)}^{\infty} \psi(\mu_1, z_1) dz_1 \\
 &+ \int_{\Phi^{-1}(1-\alpha_1/2)}^{\Phi^{-1}(1-\gamma/2)} \psi(\mu_1, z_1) (1 - \Psi(\mu_2, \Phi^{-1}\{1 - \frac{\gamma}{4(1-\Phi(z_1))}\})) dz_1 \\
 &- \Psi(\mu_2, 0) \int_{\Phi^{-1}(1-\gamma/2)}^{\infty} \psi(\mu_1, z_1) dz_1 + \int_{\Phi^{-1}(1-\alpha_1/2)}^{\Phi^{-1}(1-\gamma/2)} \psi(\mu_1, z_1) \Psi(\mu_2, \Phi^{-1}\{\frac{\gamma}{4(1-\Phi(z_1))}\}) dz_1 \\
 &+ [1 - \Psi(\mu_2, 0)] \int_{-\infty}^{\Phi^{-1}(1-\gamma/2)} \psi(\mu_1, z_1) dz_1 + \int_{\Phi^{-1}(1-\gamma/2)}^{\Phi^{-1}(\alpha_1/2)} \psi(\mu_1, z_1) (1 - \Psi(\mu_2, \Phi^{-1}\{1 - \frac{\gamma}{4\Phi(z_1)}\})) dz_1 \\
 &- \Psi(\mu_2, 0) \int_{-\infty}^{\Phi^{-1}(1-\gamma/2)} \psi(\mu_1, z_1) dz_1 + \int_{\Phi^{-1}(1-\gamma/2)}^{\Phi^{-1}(\alpha_1/2)} \psi(\mu_1, z_1) \Psi(\mu_2, \Phi^{-1}\{\frac{\gamma}{4\Phi(z_1)}\}) dz_1,
 \end{aligned}$$

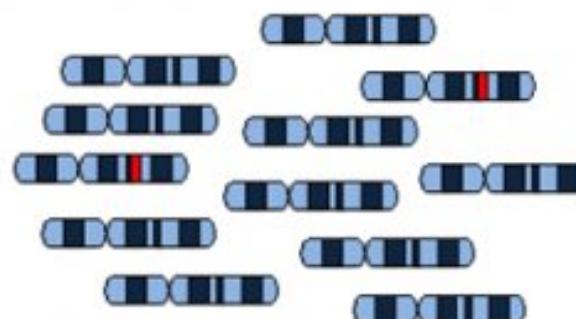
cases



controls



Variant Frequency  
Cases - 58.3%  
Controls - 16.7%





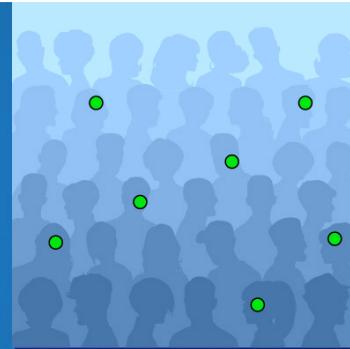
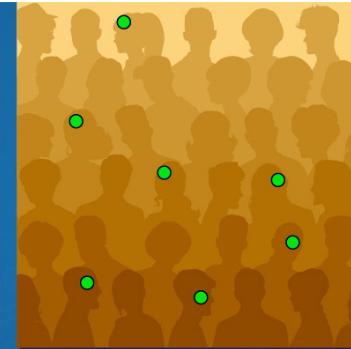
# GWAS – Genome-wide Association Studies

NHGRI FACT SHEETS  
genome.gov

- If one type of the variant is more frequent, the variant is said to be *associated* with the trait.
- The associated SNPs are then considered to mark a region that may influence the trait.

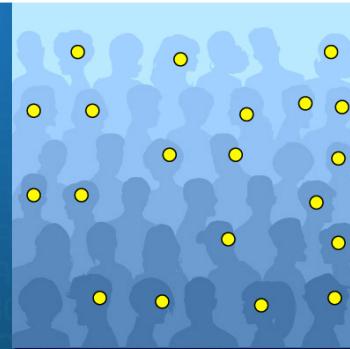
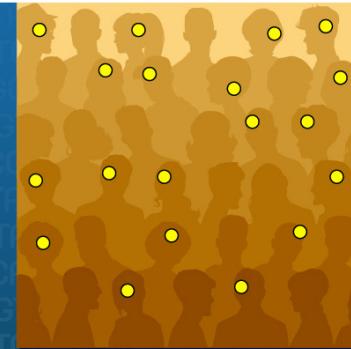


**SNP 1**



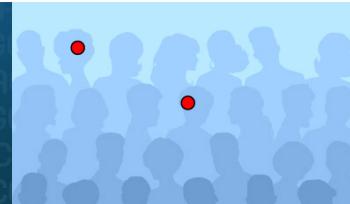
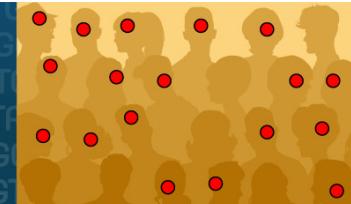
**SNP 1**  
No association to disease

**SNP 2**



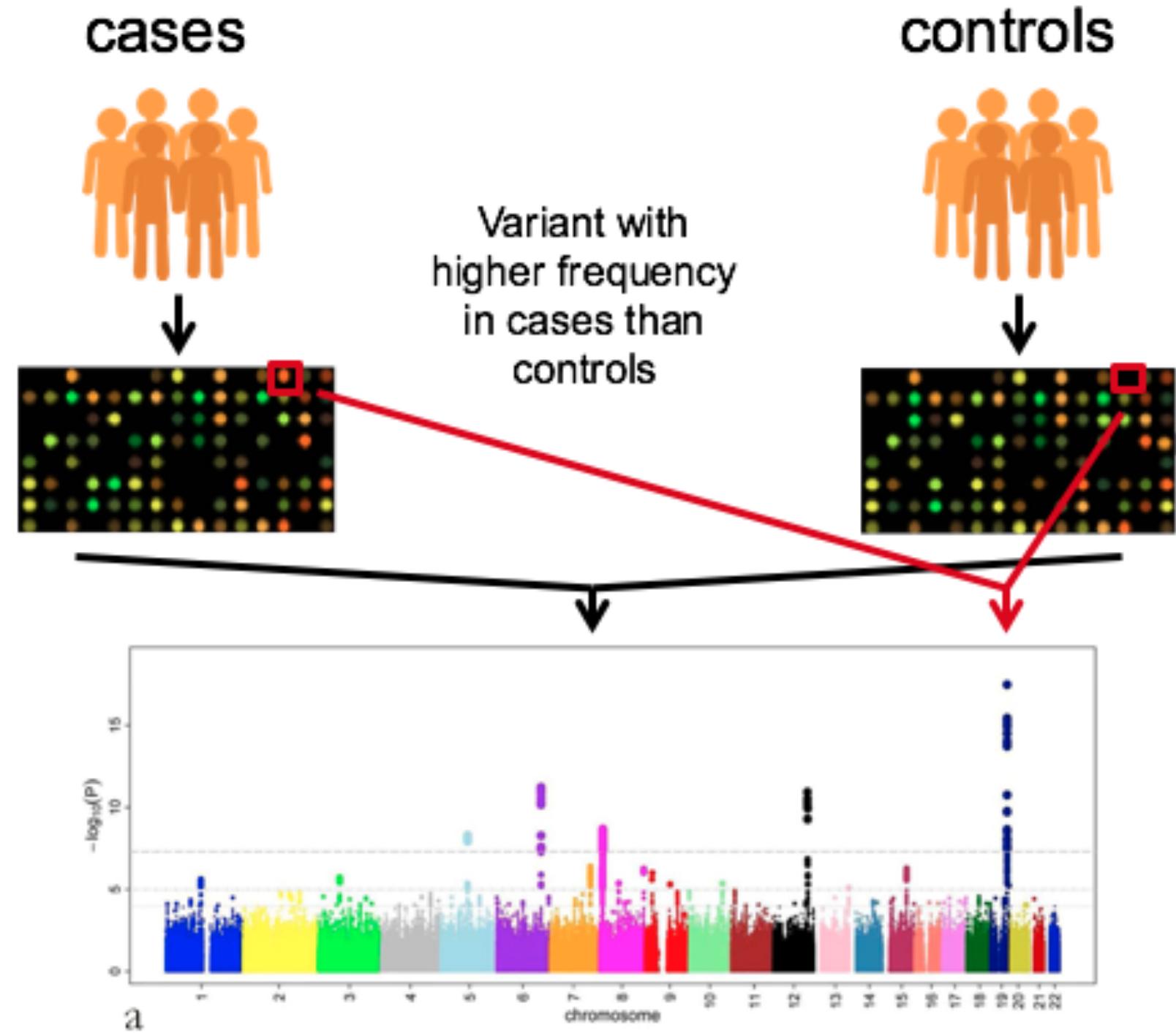
**SNP 2**  
No association to disease

**SNP 3**



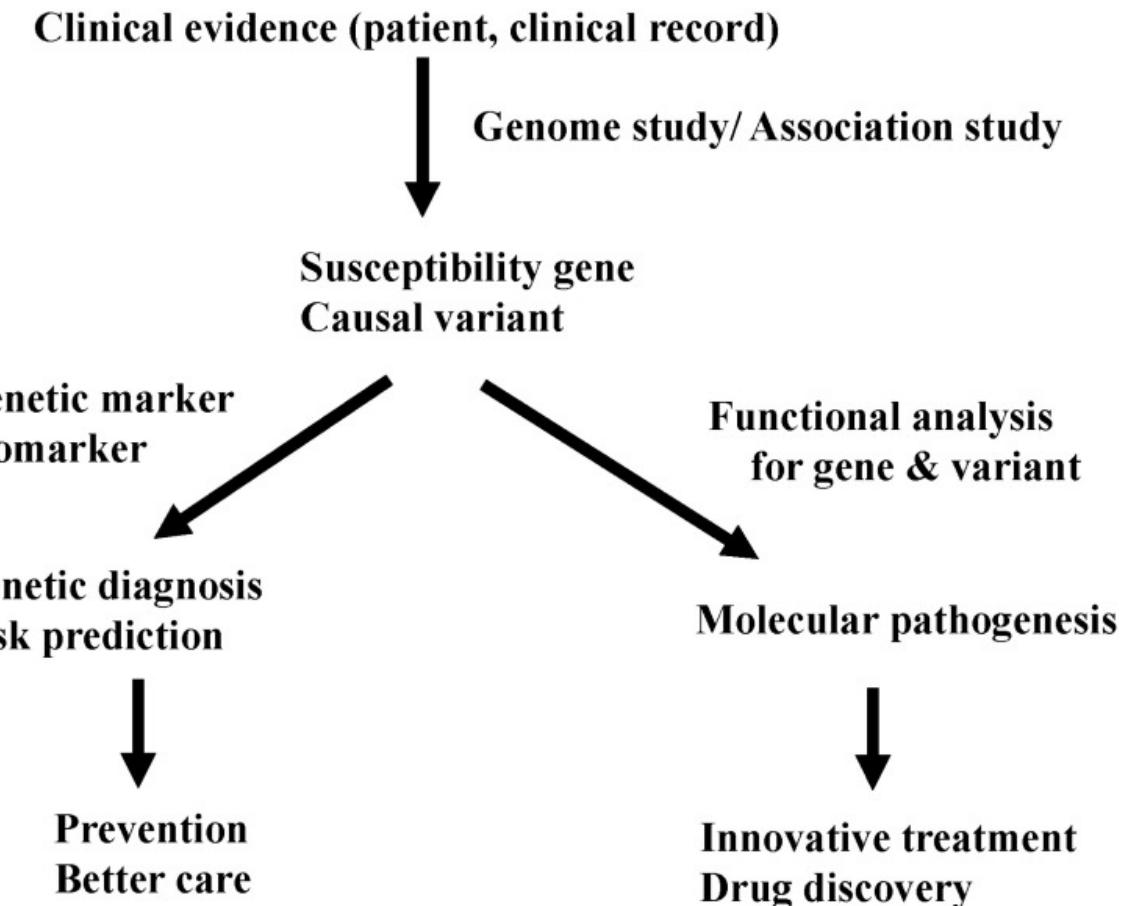
**SNP 3**  
Associated to disease

- Each dot represents a SNP, with the X-axis showing genomic location and Y-axis showing association level.
- A p-value indicates the significance of the difference in frequency of the allele tested between cases and controls
- the probability that the allele is likely to be associated with the trait.



# Importance

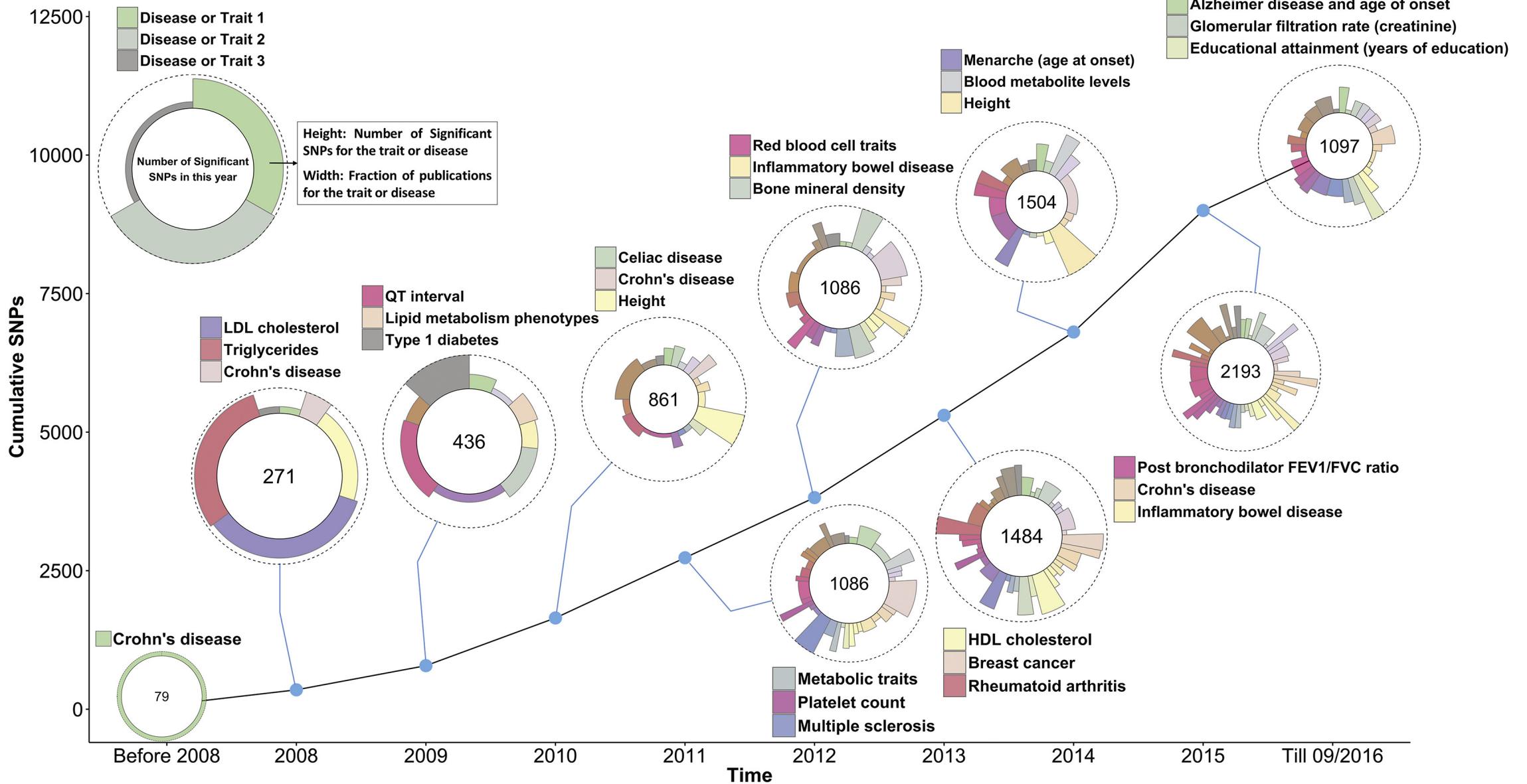
- Identify a marker for genetic diagnosis and risk assessment
- Functional analysis of a gene and its associated variants (Disease mechanism)



# Success

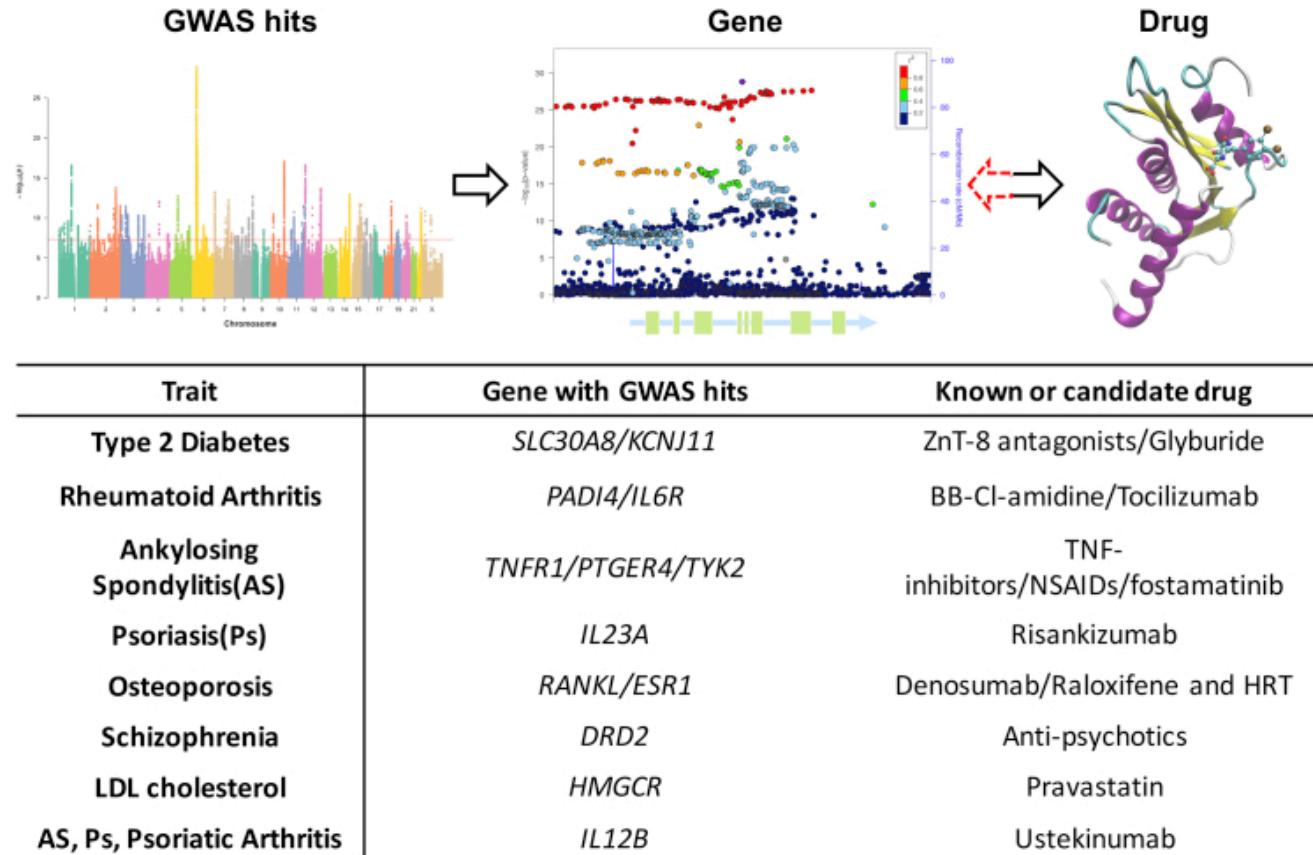
- Helped in detecting associations between common DNA variants and traits.
- Led to a better understanding of the genetic architecture of complex traits and therefore of past natural selection on traits associated with fitness.
- It has led to the discovery of variants, genes, and biological pathways that play a role in specific diseases and disorders.

# GWAS SNP-Trait Discovery Timeline



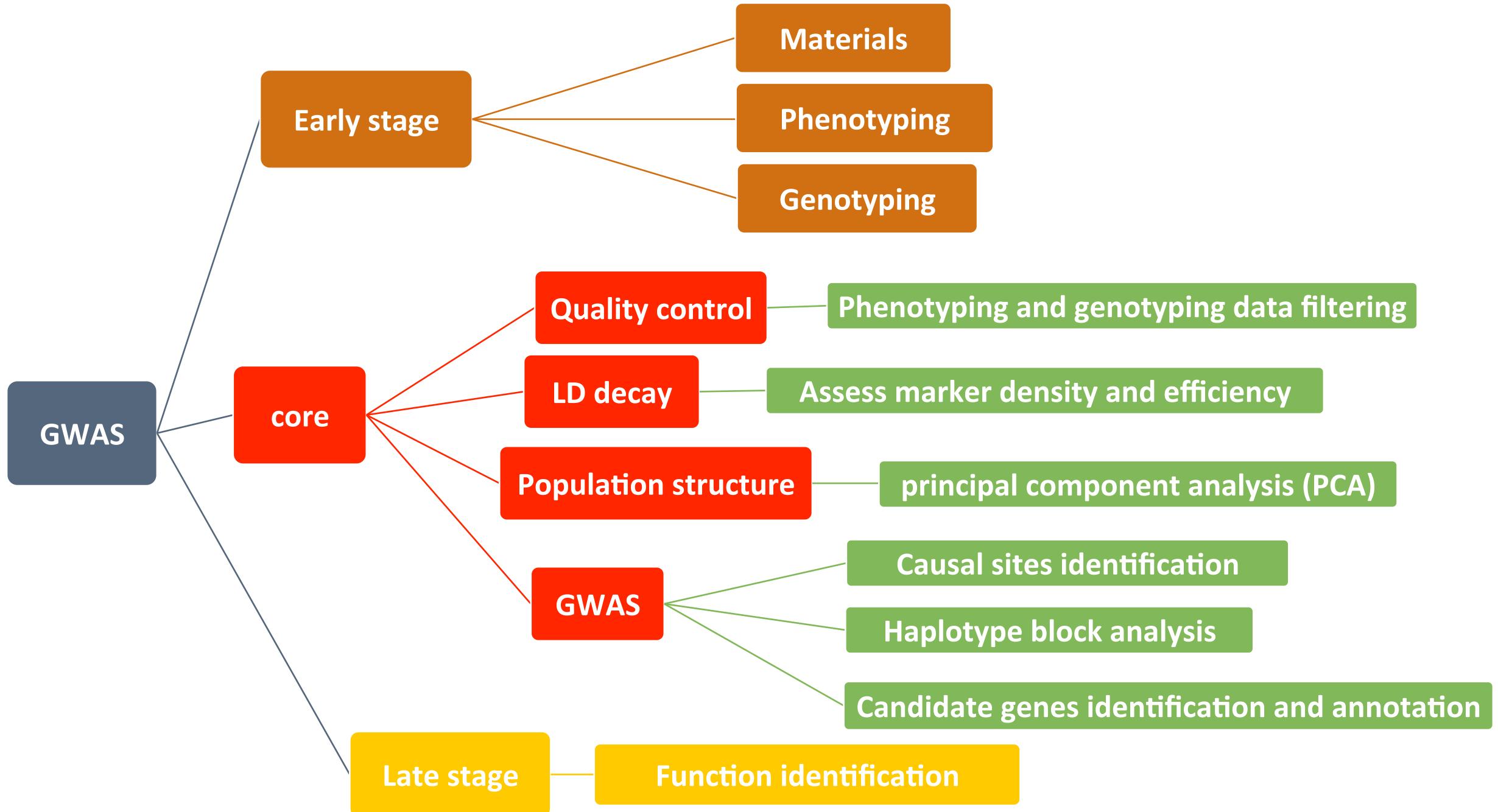
# Success

- GWAS studies have identified over 100 common variant signals for type 2 Diabetes



# Challenges

- “significantly associated” does not mean that the association identified by the study is significant in biology, medicine, or actual life
  - Sanna et al. 2008 reported the identification of a SNP that is associated with height, but its estimated additive effect is only **0.44 cm** on height
- GWAS is not a method to conclude something



# Materials to perform GWAS

- Germplasm resources:

Natural population materials, core germplasm, microcore germplasm, farm species, backbone inbred lines, etc

- F2



Article | Published: 07 September 2016

## Genomic architecture of heterosis for yield traits in rice

Xuehui Huang , Shihua Yang, Junyi Gong, Qiang Zhao, Qi Feng, Qilin Zhan, Yan Zhao, Wenjun Li, Benyi Cheng, Junhui Xia, Neng Chen, Tao Huang, Lei Zhang, Danlin Fan, Jiaying Chen, Congcong Zhou, Yiqi Lu, Qijun Weng & Bin Han

Nature 537, 629–633 (29 September 2016) | Download Citation

### Abstract

Increasing grain yield is a long-term goal in crop breeding to meet the demand for global food security. Heterosis, when a hybrid shows higher performance for a trait than both parents, offers an important strategy for crop breeding. To examine the genetic basis of heterosis for yield in rice, here we generate, sequence and record the phenotypes of 10,074 F<sub>2</sub> lines from 17 representative hybrid rice crosses. We classify modern hybrid rice varieties into three groups, representing different hybrid breeding systems. Although we do not find any heterosis-associated loci



Letter | Published: 09 January 2011

## Genome-wide association study of leaf architecture in the maize nested association mapping population

Feng Tian, Peter J Bradbury, Patrick J Brown, Hsiao-yi Hung, Qi Sun, Sherry Flint-Garcia, Torbert R Rocheford, Michael D McMullen, James B Holland & Edward S Buckler

Nature Genetics 43, 159–162 (2011) | Download Citation

### advantage :

- 1) the number of parents is relatively small ;
- 2) population is not genetically highly structured;
- 3) linkage analysis and association mapping;
- 4) calculate additive effect

### Disadvantage:

Time consuming

- NAM

Islam et al. BMC Genomics (2016) 17:903  
DOI 10.1186/s12864-016-3249-2

- MAGIC

BMC Genomics

Open Access



### RESEARCH ARTICLE

## A MAGIC population-based genome-wide association study reveals functional association of *GhRBB1\_A07* gene with superior fiber quality in cotton

Md Sariful Islam<sup>1</sup>, Gregory N. Thyssen<sup>2</sup>, Johnie N. Jenkins<sup>3</sup>, Linghe Zeng<sup>4</sup>, Christopher D. Delhom<sup>5</sup>, Jack C. McCarty<sup>3</sup>, Dewayne D. Deng<sup>3</sup>, Doug J. Hinchliffe<sup>2</sup>, Don C. Jones<sup>6</sup> and David D. Fang<sup>1\*</sup>

Cotton supplies a great majority of natural fiber for the global textile industry. The negative correlation between yield and fiber quality has hindered breeders' ability to improve these traits simultaneously. A multi-parent advanced generation inter-cross (MAGIC) population developed through random-mating of multiple diverse parents has the ability to break this negative correlation.

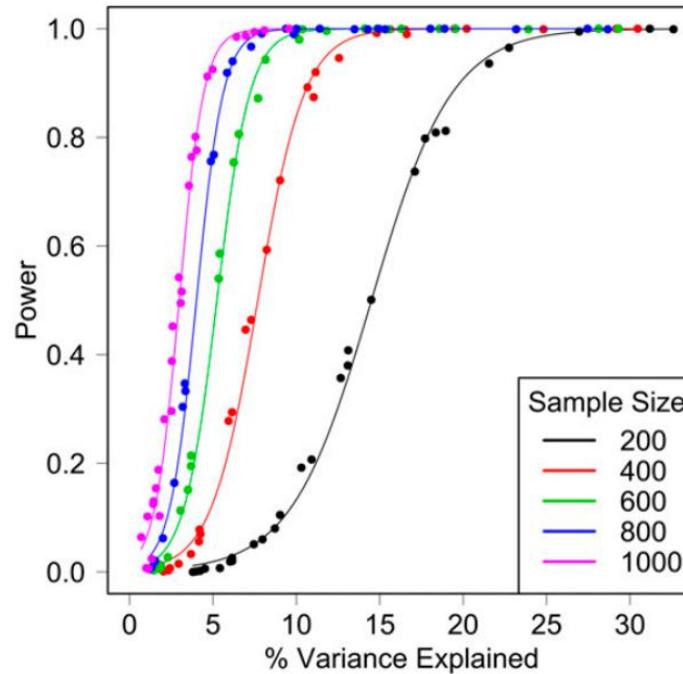
### advantage :

- 1) Keep genetic diversity in progeny
- 2) Obvious heterosis

### Disadvantage:

- 1) Time consuming
- 2) Population structure is hard to control

# Materials to perform GWAS



**Figure 4** Power simulations demonstrate the relationship between power, sample size, and percentage variance explained. The percentage variance explained by the simulated QTL is plotted vs. the power to detect the simulated QTL for five different sample sizes. Points are the mean value from 1000 simulations and curves are logistic regression models fit to the data.

Gatti D M, Svenson K L, Shabalina A, et al. Quantitative trait locus mapping methods for diversity outbred mice[J]. G3: Genes| Genomes| Genetics, 2014, 4(9): 1623-1633.

Species	year	Magazine	Genotyping method	Population size	Trait
Cotton	2017	Nature Genetics	Resequencing	318	Fiber and yield
Cotton	2017	Nature Genetics	Resequencing	352	Fiber quality
Rice	2016	Nature Genetics	Resequencing	176	Agronomic traits
Rice	2016	Nature Genetics	Resequencing	342	Grain shape
Rice	2016	Nature	Resequencing	10,074	Heterotic effects

nature  
genetics

Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice

Kenji Yano<sup>1</sup>, Eiji Yamamoto<sup>2</sup>, Koichiro Aya<sup>1</sup>, Hideyuki Takeuchi<sup>1</sup>, Pei-ching Lo<sup>1</sup>, Li Hu<sup>1</sup>, Masanori Yamasaki<sup>3</sup>, Shinya Yoshida<sup>4</sup>, Hidemi Kitano<sup>1</sup>, Ko Hirano<sup>1</sup> & Makoto Matsuoka<sup>1</sup>

**Consider the sample size:**  
**The lower the effect of a causal site, the larger sample size we need.**

# Genotyping data filtering

## MSP & MAF

- **High MSP (missing rate)** indicates **poor genotype probe performance** and **low genotyping accuracy** (Neale and Purcell, [2008](#); WTCCC, [2007](#))
- SNPs with **low MAF** are more prone to error, as **fewer samples would be within a genotype cluster** (Neale and Purcell, [2008](#); Teo, [2008](#))

176 japonica rice varieties developed in breeding programs conducted in Japan

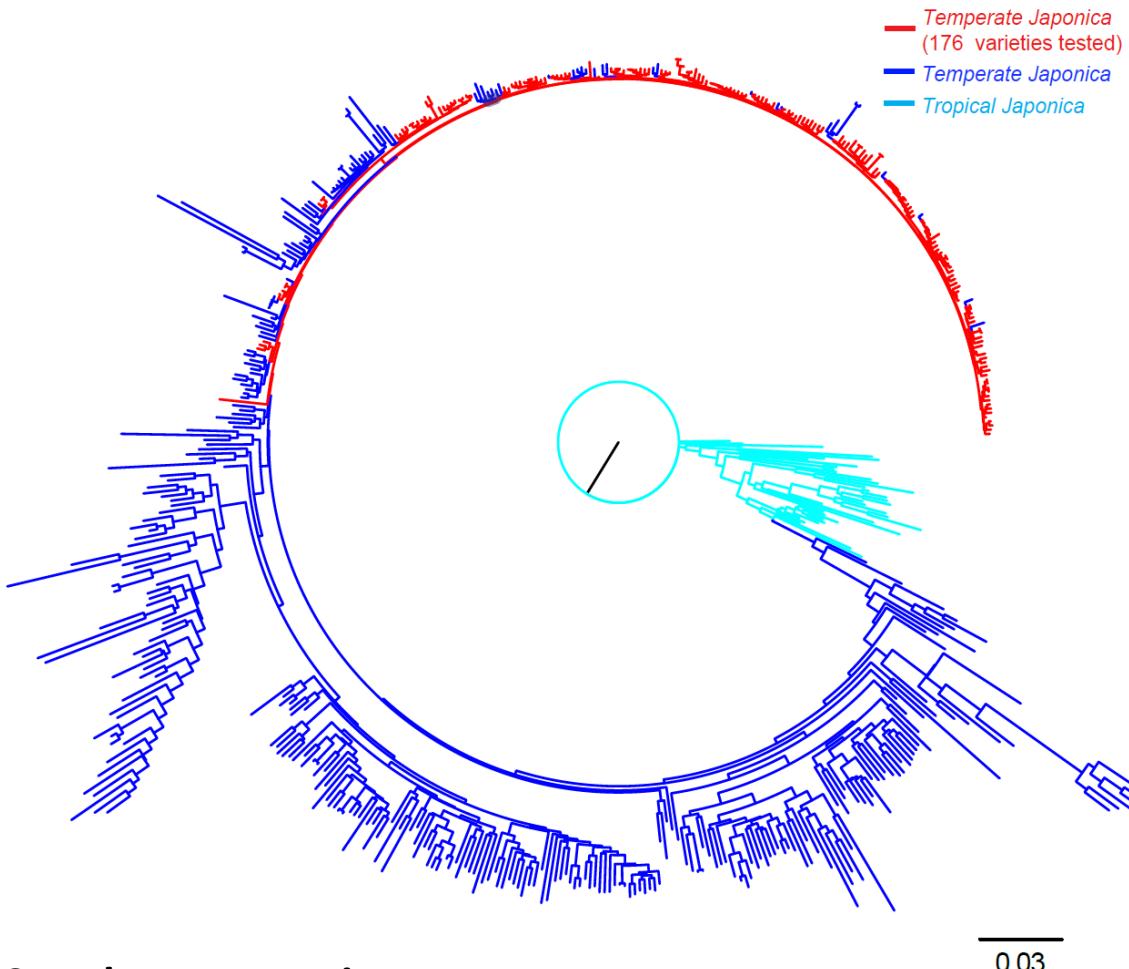
whole-genome sequencing  
GATK: variations discover  
y  
missing rates  $\geq 0.25$   
minor allele frequency  $< 0.05$

**Supplementary Table 4.** Summary of categorized SNPs and INDELs.

Location category	SNPs			INDELs		
promoter_region	110,918			19,318		
five_prime_UTR	4,091			1,325		
CDS	69,804	Synonymous	26,481	3,334	Frameshift	1,678
		Non synonymous	43,323		Non frameshift	1,656
INTRON	65,462			11,683		
three_prime_UTR	7,854			1,947		
Others (Not located in gene region)	168,208			29,937		
Total	426,337			67,544		

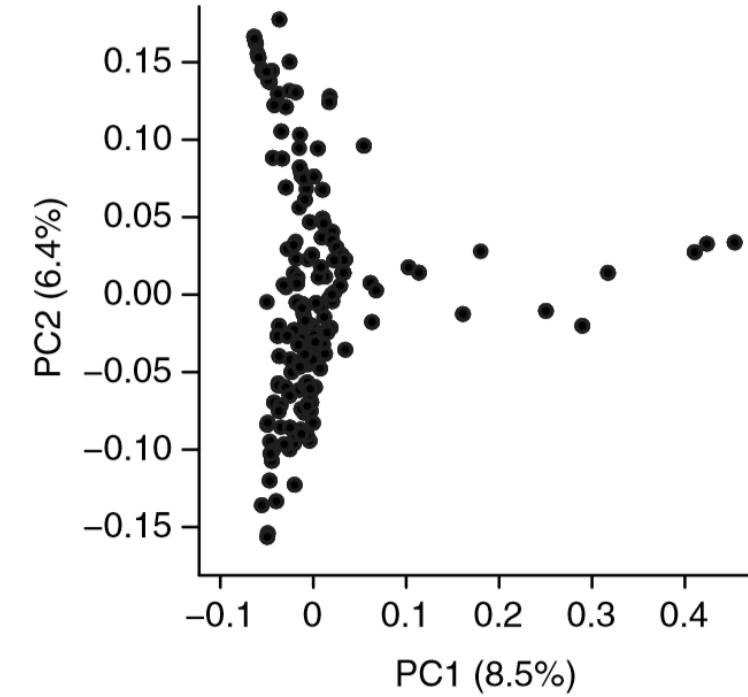
# Quantify the population structure

## (1) Phylogenetic tree construction



**Supplementary Figure 1.**  
Neighbor-joining tree of *Oryza sativa* subspecies *japonica*  
calculated by whole-genome sequence data.

## (2) PCA plot

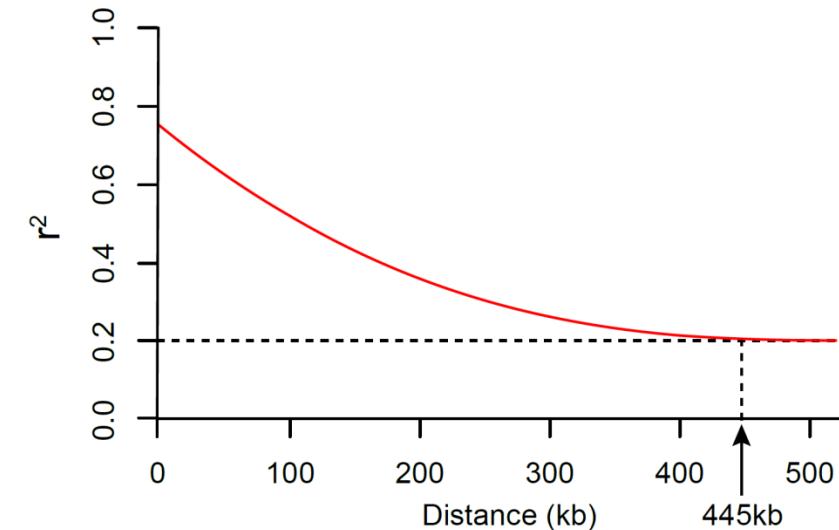
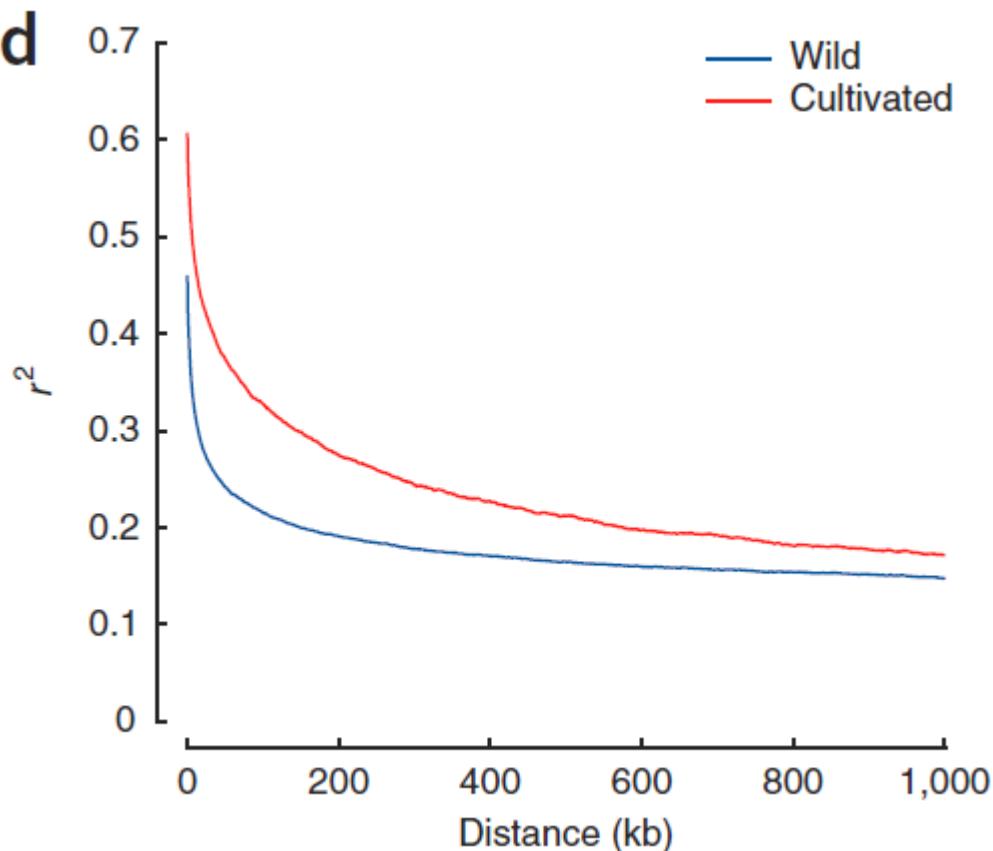


The score plot of principal components showed continuous distribution without any distinct clusters.

# LD decay distance

- LD decay distance determines marker density we needed and the accuracy of GWAS
- GWAS marker density = genome size / LD decay distance

Fig.d Compare LD decay between wild and cultivated soybean



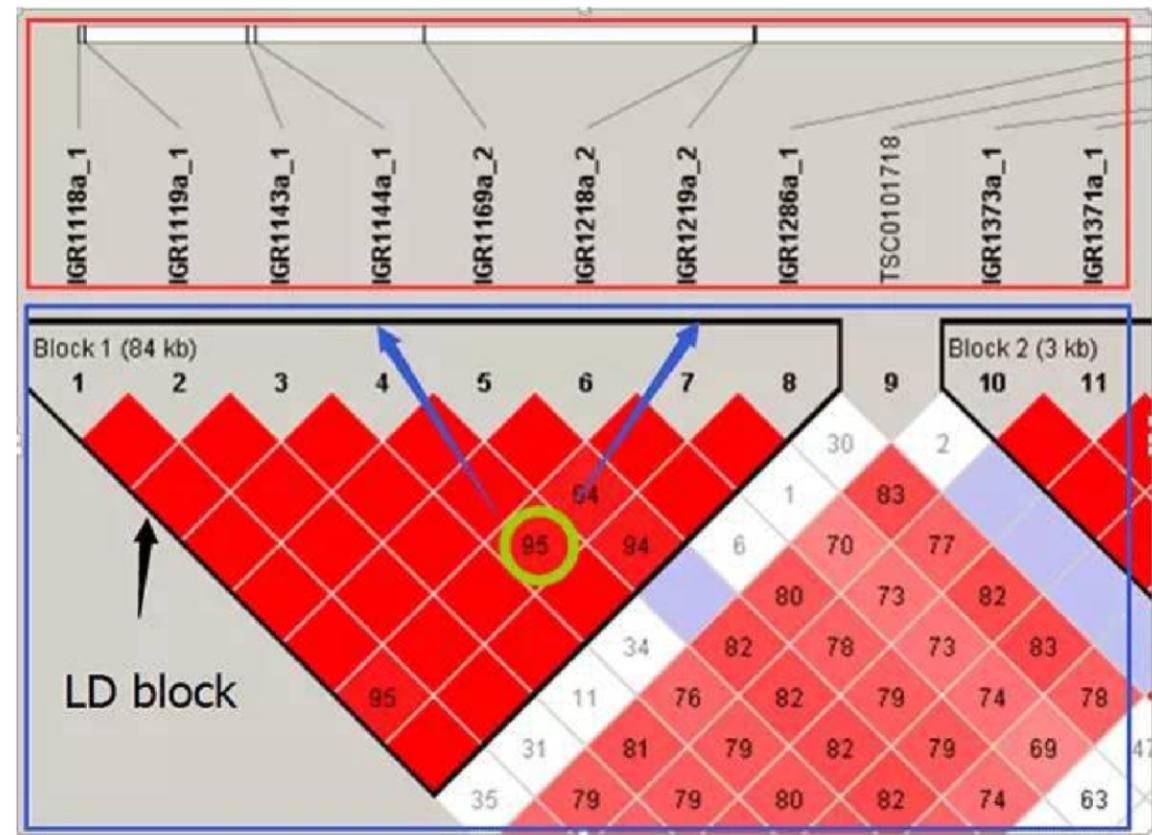
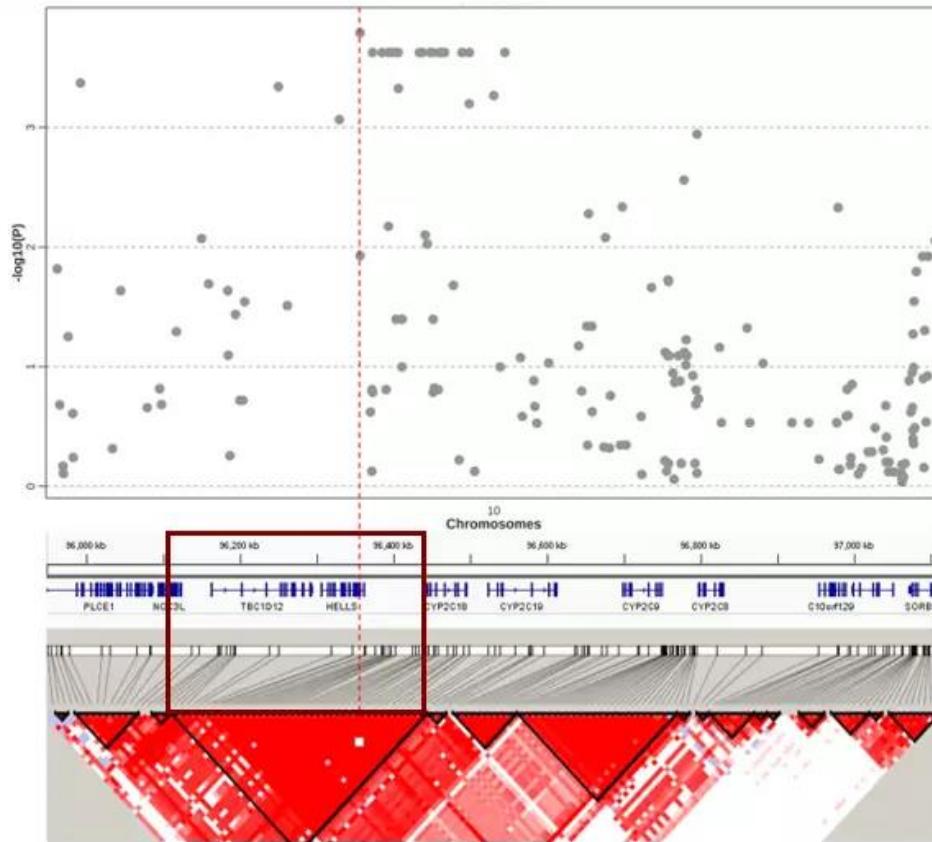
Supplementary Figure 2. Genome-wide average LD decay in the 176 varieties. LD was calculated as the squared Pearson's correlation coefficient ( $r^2$ ).

**The decay of LD with physical distance between SNPs occurred at 445 kb ( $r^2 = 0.2$ )**

- The higher the domestication level, the longer distance LD decay
- LD decay distance decrease with more recombination
- Closer to centromere region, LD decay distance increase

# Validation of rapid gene identification using GWAS

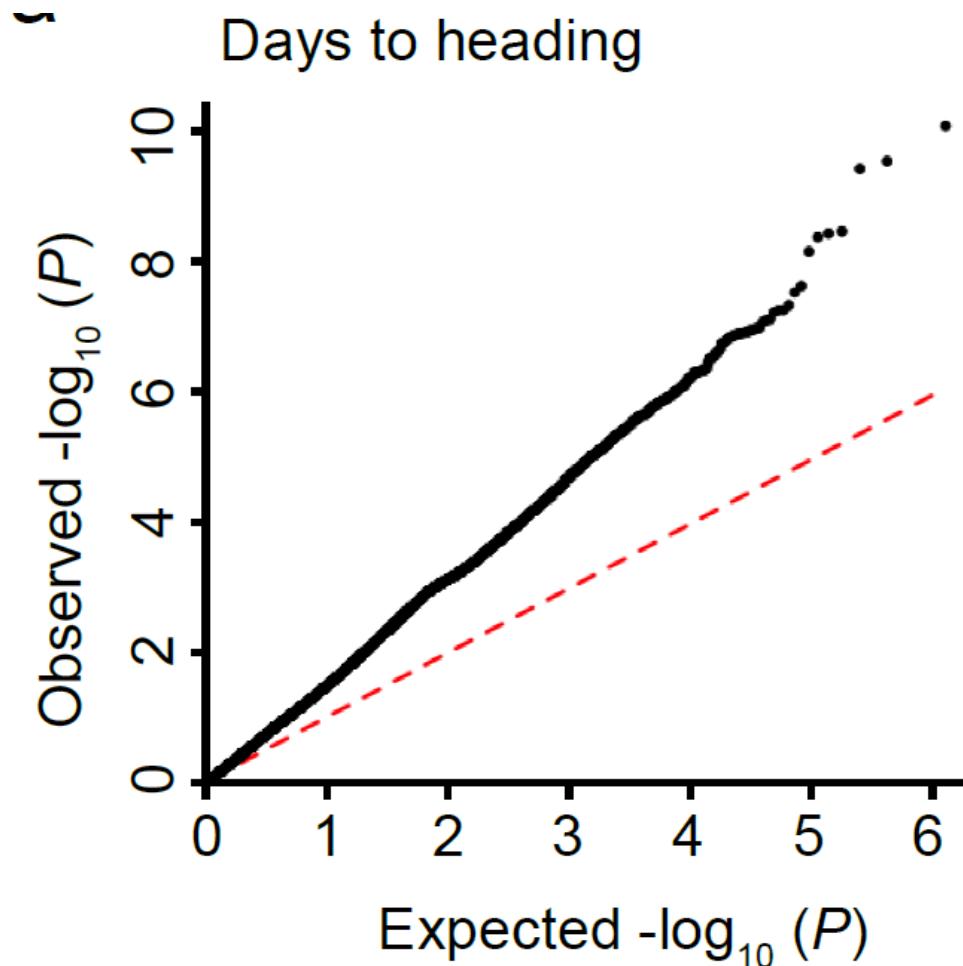
# Haplotype block analysis



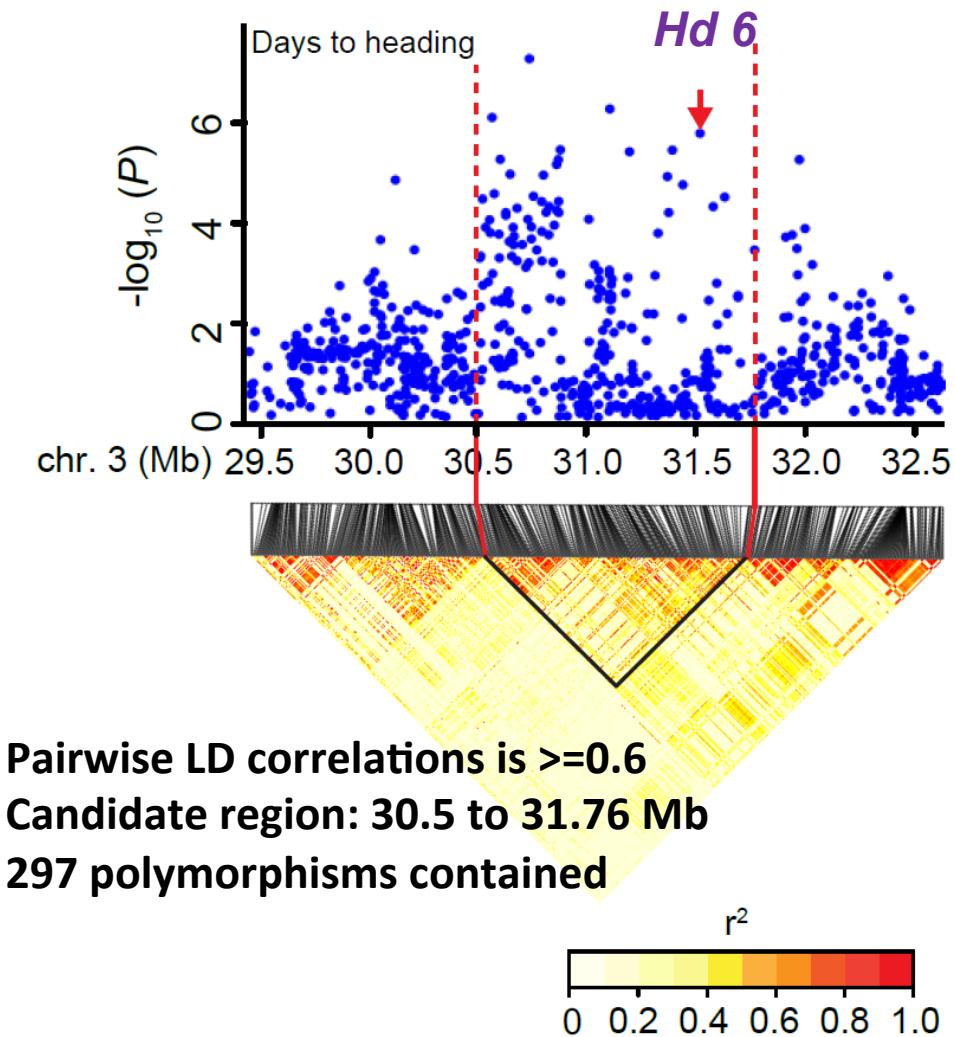
- Haplotypes are inherited together with little chance of contemporary recombination.

# Validation of rapid gene identification using GWAS

Using known heading date (*Hd6* and *Hd2*) genes identify efficiency of GWAS method.



Quantile-quantile plots of observed versus expected  $-\log_{10} P$  values of GWAS results.



Detailed analyses of the peak for days to heading on chromosome 3 (*Hd 6*).

# Validation of rapid gene identification using GWAS

## Group I (1 polymorphism in 1 gene)

Chr.	Position (bp)	P-value	Ref	Alt	Gene ID	Region	Ref.codon	Alt.codon	Ref.aa	Alt.aa	Annotation
3	31,512,460	5.77	A	T	LOC_Os03g55389	CDS	TAG	AAG	*	K	casein kinase II subunit alpha-1, Hd6

## Group II (6 polymorphisms in 6 genes)

Chr.	Position (bp)	P-value	Ref	Alt	Gene ID	Region	Ref.codon	Alt.codon	Ref.aa	Alt.aa	Annotation
3	30,553,943	6.09	-	A	LOC_Os03g53250	promoter_region	-	-	-	-	expressed protein
3	30,637,052	4.94	T	-	LOC_Os03g53400	promoter_region	-	-	-	-	transmembrane BAX inhibitor motif-containing protein, putative, expressed
3	30,790,531	4.92	-	AC	LOC_Os03g53700	promoter_region	-	-	-	-	PHD-finger domain containing protein, putative, expressed
3	30,851,452	5.13	-	C	LOC_Os03g53800	promoter_region	-	-	-	-	periplasmic beta-glucosidase precursor, putative, expressed
3	30,859,447	5.23	A	G	LOC_Os03g53810	promoter_region	-	-	-	-	retrotransposon protein, putative, unclassified, expressed
3	31,361,959	4.89	T	-	LOC_Os03g55120	promoter_region	-	-	-	-	plastocyanin-like domain containing protein, putative, expressed

## Group III (3 polymorphisms in 3 genes)

Chr.	Position (bp)	P-value	Ref	Alt	Gene ID	Region	Ref.codon	Alt.codon	Ref.aa	Alt.aa	Annotation
3	30,868,524	5.43	C	T	LOC_Os03g53820	CDS	GAG	GAA	E	E	retrotransposon protein, putative, unclassified, expressed
3	31,187,258	5.39	G	T	LOC_Os03g54860	INTRON	-	-	-	-	60S ribosomal protein L13a-2, putative, expressed
3	31,384,912	5.42	T	G	LOC_Os03g55140	INTRON	-	-	-	-	transducin family protein, putative, expressed

## Group IV (3 polymorphisms)

Chr.	Position (bp)	P-value	Ref	Alt
3	30,591,452	5.24	A	C
3	30,725,715	7.28	T	A
3	31,096,597	6.26	T	-

Group I: predicted to induce amino acid exchange or to change splicing junctions

Group II: located at the 5' flanking sequences of genes

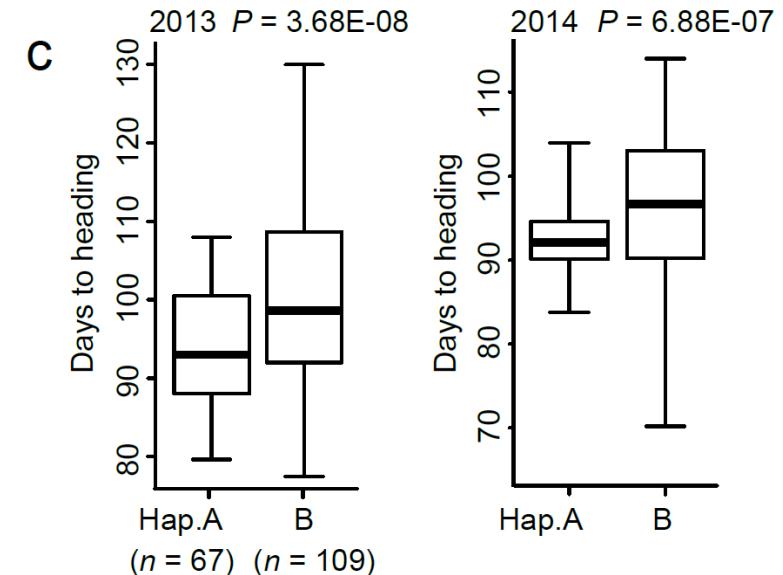
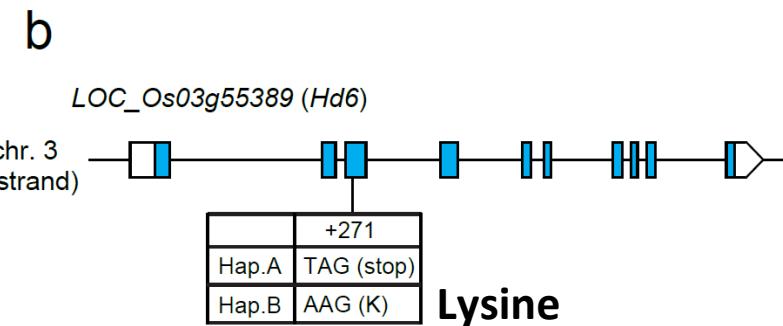
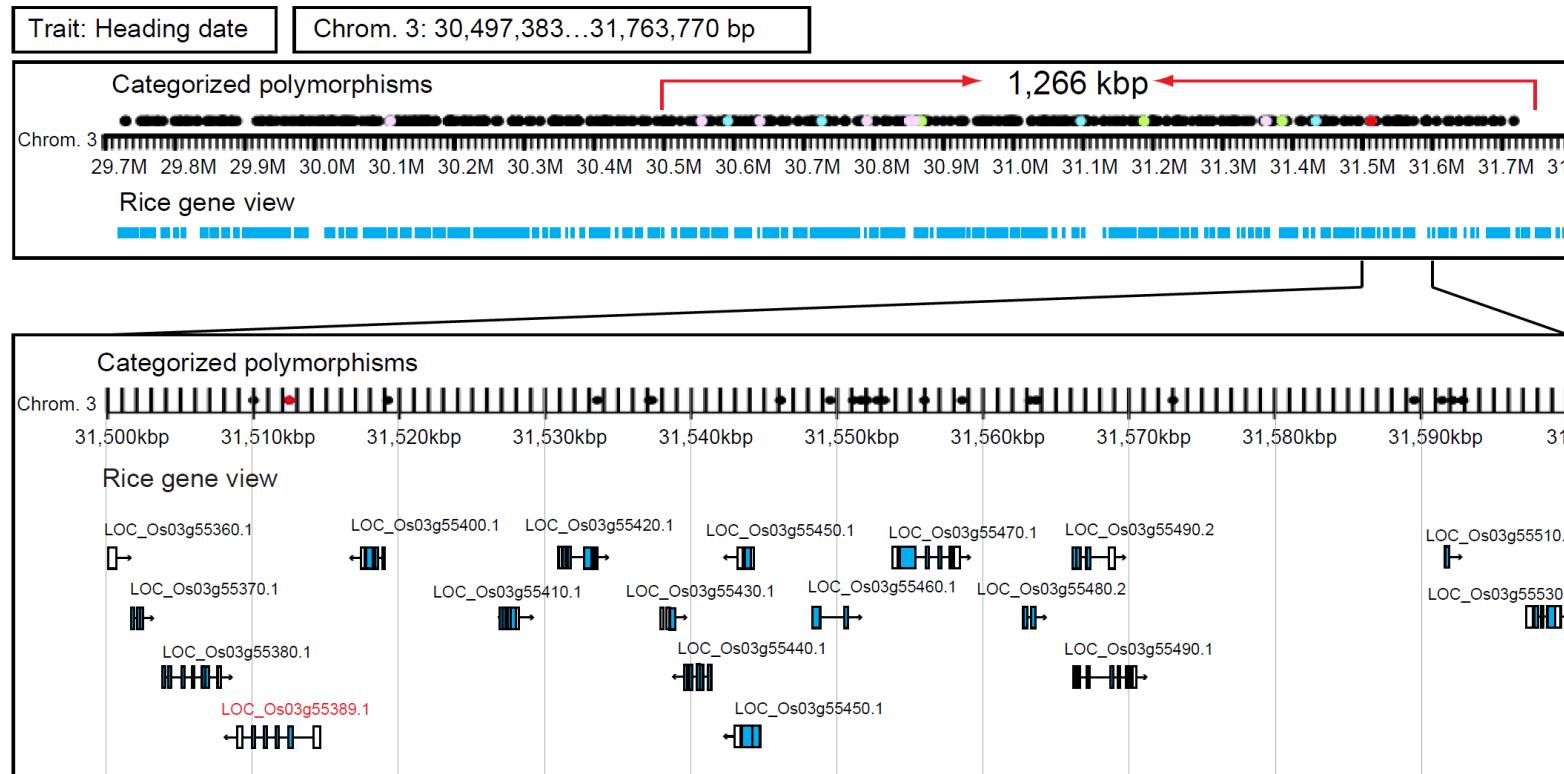
Group III: located in a coding region but not predicted to change an amino acid, an intron or a 3' noncoding sequence

Group IV: located on outside coding regions

Group V: included polymorphisms not significantly associated with trait variation

## Group V (284 polymorphisms)

# Validation of rapid gene identification using GWAS

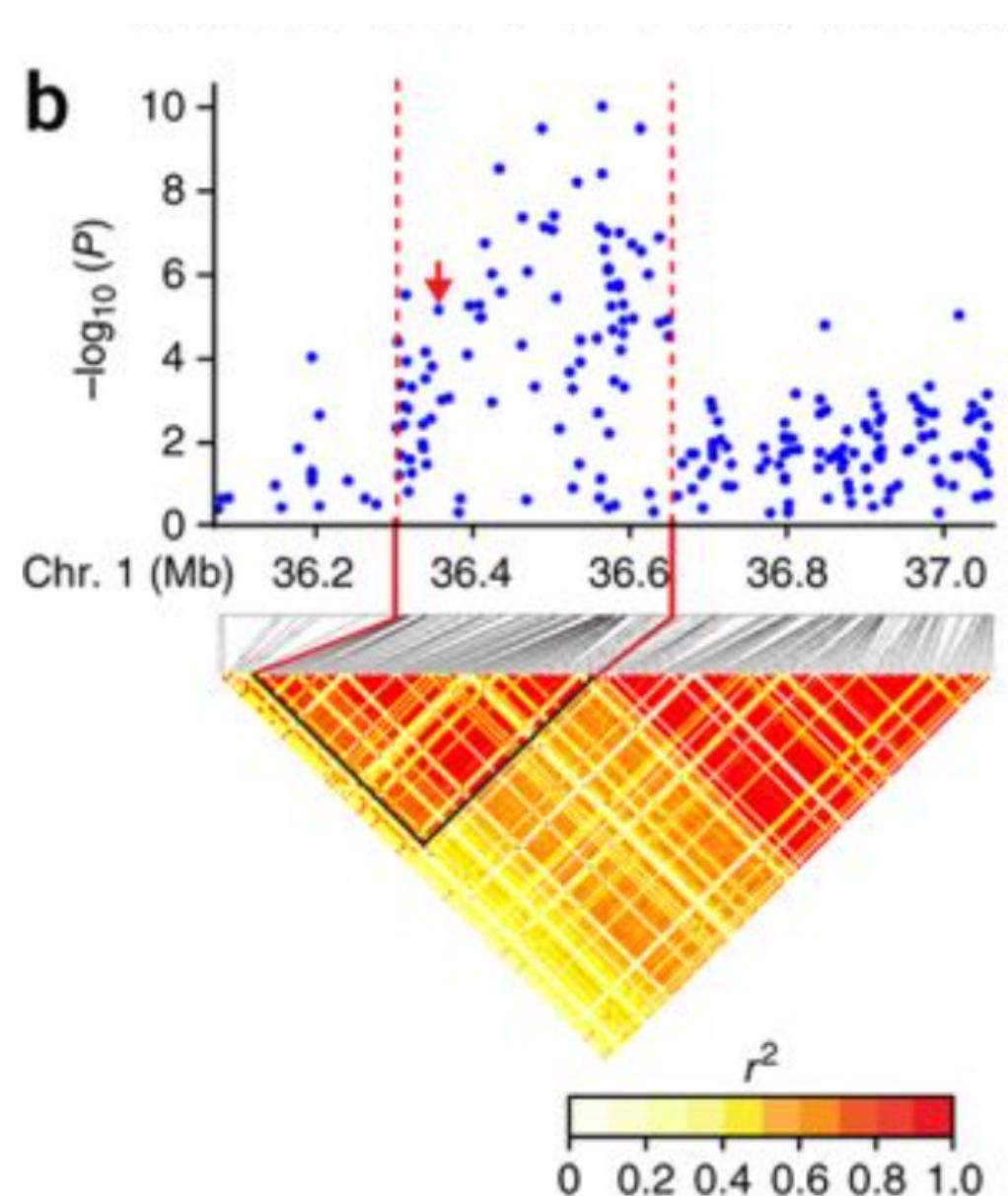
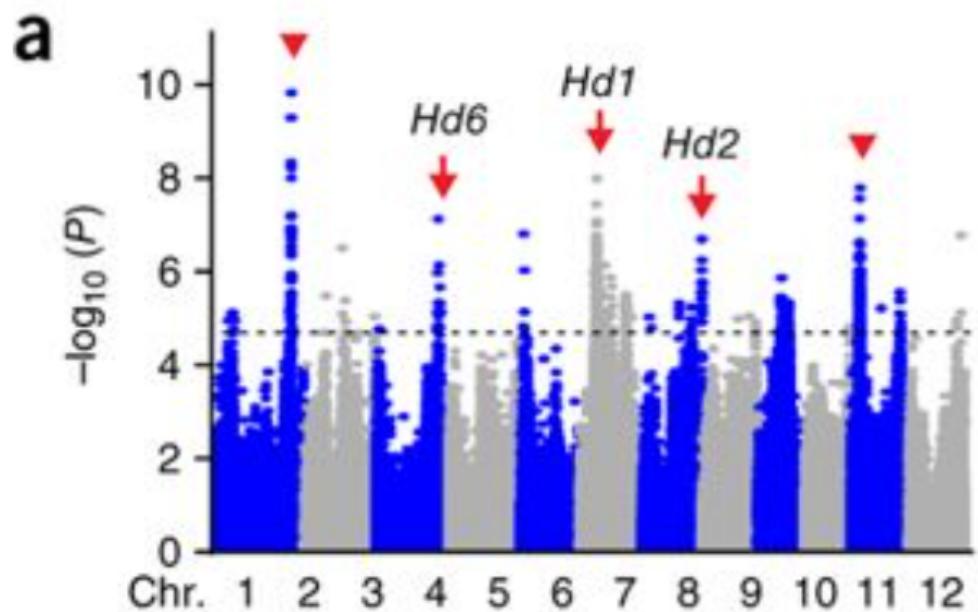


Supplementary Figure 4. Detailed analyses of the peak for days to heading on chromosome 3 (*Hd 6*).

b) Exon-intron structure of *LOC\_Os03g55389* and DNA polymorphisms in this gene.

c) Boxplots for days to heading based on the haplotypes for *LOC\_Os03g55389* in 2013 (left) and 2014 (right).

# Identify new genes for heading date



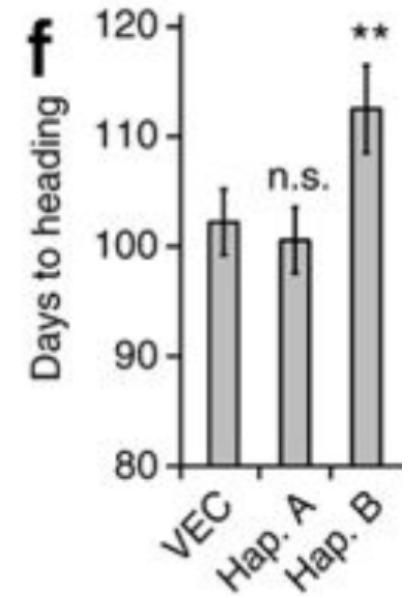
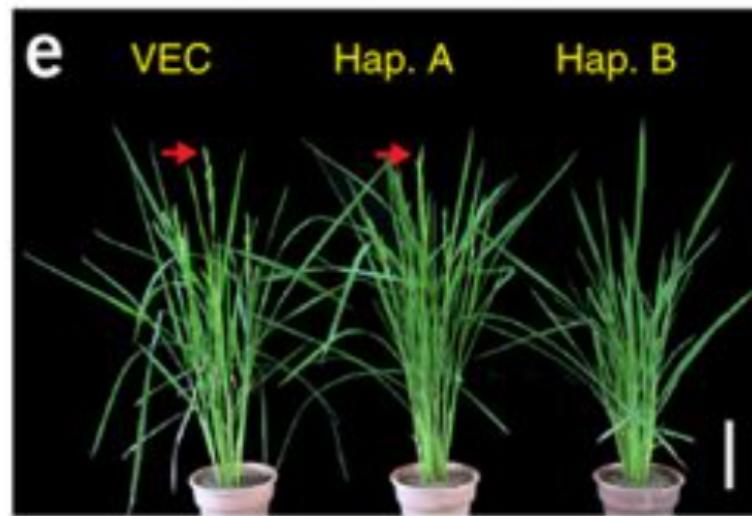
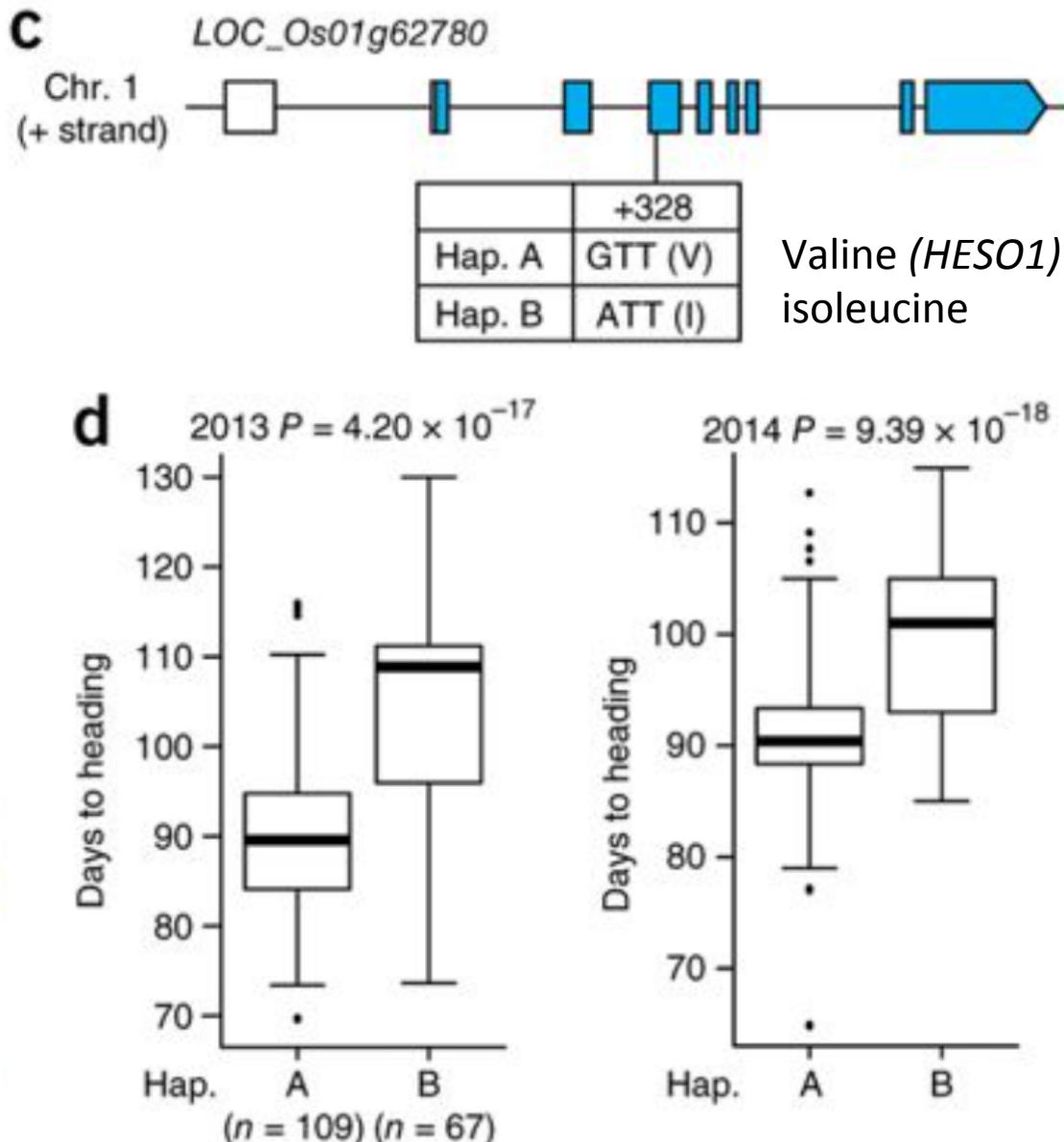
## Polymorphisms on chromosome 1:

- 91 polymorphisms included.
- Eight group I polymorphisms, located within seven genes.

All of these were annotated as transposon-related genes except for LOC\_Os01g62780, a homolog of *Arabidopsis thaliana* HEN1 suppressor 1 (*HESO1*).

*HESO1*: pleiotropic phenotypes including delayed flowering in *Arabidopsis*.

# Identify new genes for new heading date gene



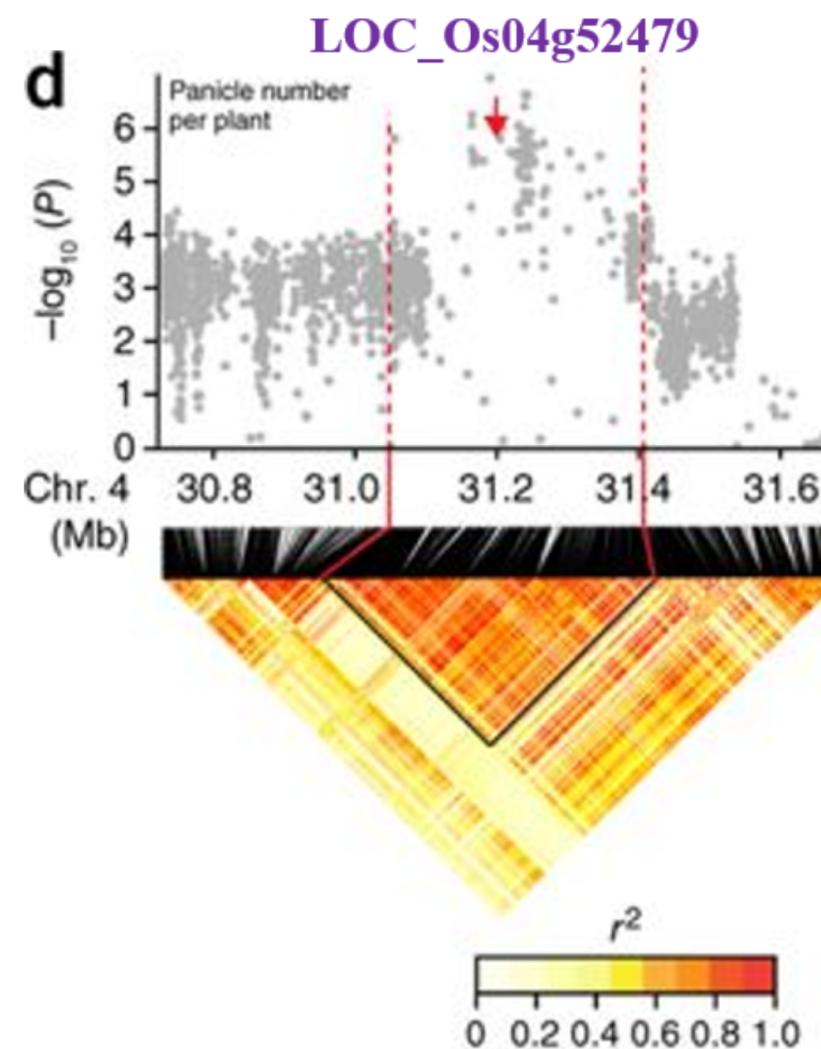
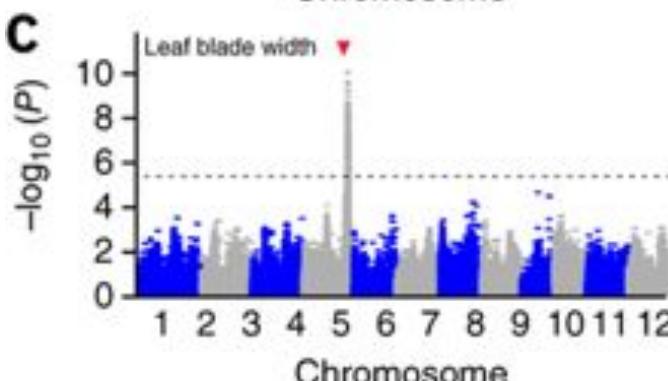
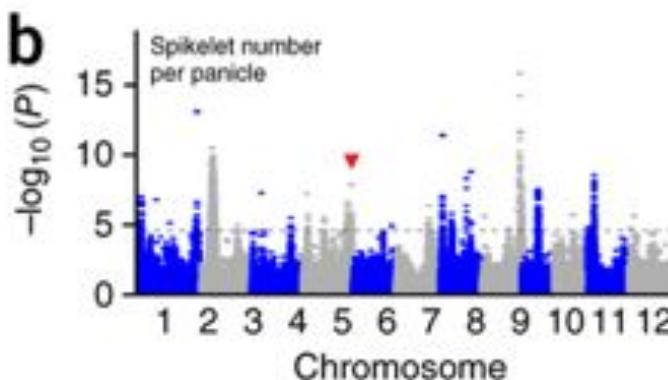
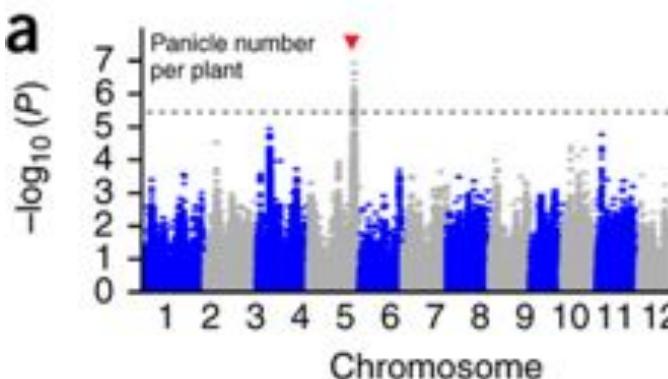
(e) Image of transgenic plants (Nipponbare) transformed with empty vector (VEC), haplotype A (Hap. A) and haplotype B (Hap. B). Red arrows indicate panicle exertion. Scale bar, 15 cm.

(f) Days to heading of the transgenic plants. Error bars, s.d. (n = 20). \*\*P < 0.01; n.s., not significant (Welch's t-test).

*LOC\_Os01g62780* is the causal gene for the peak signal of days to heading on chromosome 1.

# Identify new genes for several traits

GWAS for panicle number per plant, spikelet number per panicle and leaf blade width, and identification of the causal gene for the peak on chromosome 4.



For panic number trait, assigned four polymorphisms to group I, mapped to four genes.

**LOC\_Os04g52479**, encodes **NALLOW LEAF 1 (NAL1)**, has been previously reported to control panicle size and flag leaf width

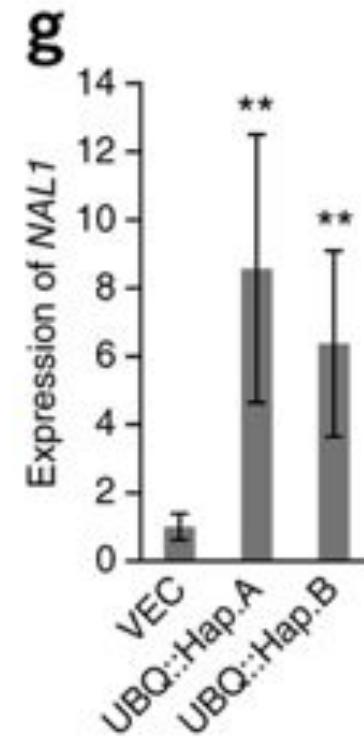
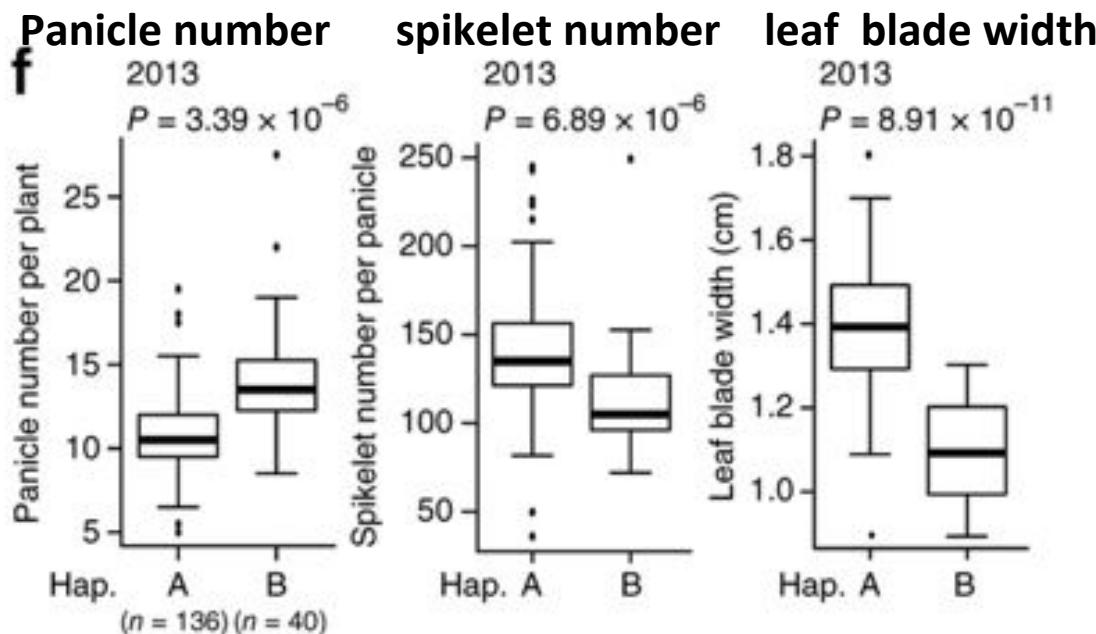
May be pleitropic.

# Identify new genes for several traits

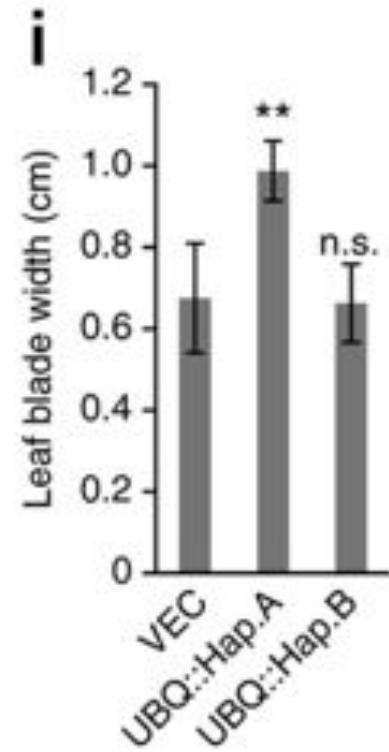
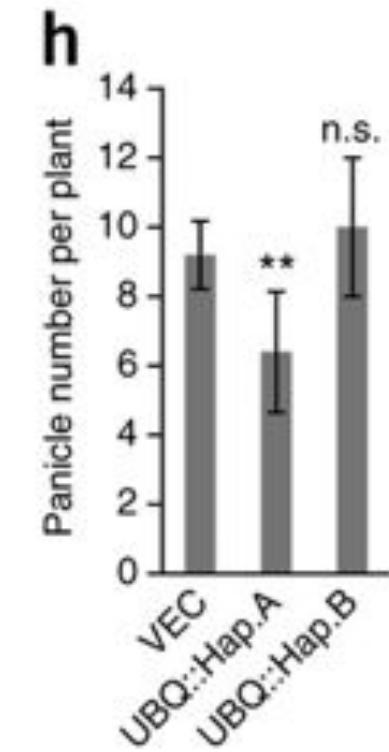
e *LOC\_Os04g52479 (NAL1)*



	+697
Hap. A	CAT (H)
Hap. B	CGT (R)



(h) Panicle number      (i) leaf blade width



(g) Expression of  
*LOC\_Os04g52479 (NAL1)*

Overexpression of haplotype A  
(UBO::Hap.A) and haplotype B  
(UBO::Hap.B).

The plants overexpressing NAL1 from haplotype A showed decreased panicle numbers per plant and increased leaf blade width.

# Summary

- Consider the sample size
- Data filtering
- Characterize population structure
- Perform GWAS and identify causal sites

# What after GWAS?

- Which variants are causal?
- How these variants lead to change in the phenotype?
  - Change in protein
  - Incomplete protein
  - Regulation of gene expression

# LETTER

doi:10.1038/nature17939

## Parkinson-associated risk variant in distal enhancer of $\alpha$ -synuclein modulates target gene expression

Frank Soldner<sup>1</sup>, Yonatan Stelzer<sup>1</sup>, Chikdu S. Shivalila<sup>1,2</sup>, Brian J. Abraham<sup>1</sup>, Jeanne C. Latourelle<sup>3</sup>, M. Inmaculada Barrasa<sup>1</sup>, Johanna Goldmann<sup>1</sup>, Richard H. Myers<sup>3</sup>, Richard A. Young<sup>1,2</sup> & Rudolf Jaenisch<sup>1,2</sup>

Genome-wide association studies (GWAS) have identified numerous genetic variants associated with complex diseases, but mechanistic insights are impeded by a lack of understanding of how specific risk variants functionally contribute to the underlying pathogenesis<sup>1</sup>. It has been proposed that *cis*-acting effects of non-coding risk variants on gene expression are a major factor for phenotypic variation of complex traits and disease susceptibility. Recent genome-scale *enigenetic* studies have highlighted the enrichment of GWAS-

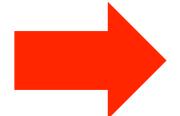
regulatory element through *cis*-regulatory effects on expression is predicted to modulate allele-specific gene expression when measured as the ratio between the modified and the non-targeted allele.

To analyse precisely the expression of two individual alleles in a single multiplex reaction, we adapted TaqMan SNP genotyping assays to quantitative reverse transcription polymerase chain reaction (qRT-PCR; Extended Data Fig. 1a). A common SNP (rs356165 A/G, referred to as SNCA ‘reporter SNP’) was identified in the 3' UTR of

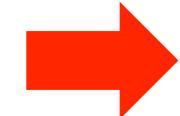
# Background on Parkinson's disease

- Chronic progressive neurodegenerative disease
- Complex trait ; associated with many genetic factors
- GWAS studies have associated many loci to Parkinson's disease
- SNCA is one of the strongest risk loci for Parkinson's disease
- Increased expression of SNCA in brain causes Parkinson's disease
- Studies have identified Parkinson's associates SNPs enriched in distal enhancers

SNPs in SNCA  
locus



Increased  
expression of  
SNCA locus

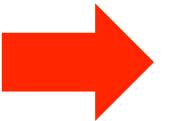


Parkinson's  
disease



Mechanism?

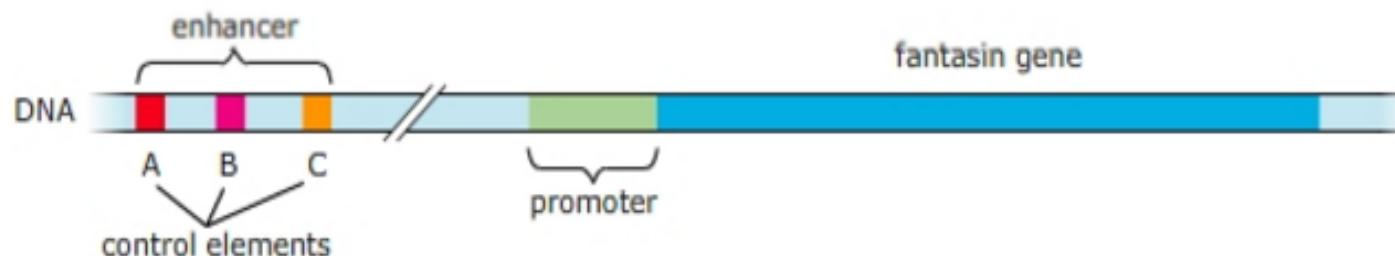
Causal SNPs ?



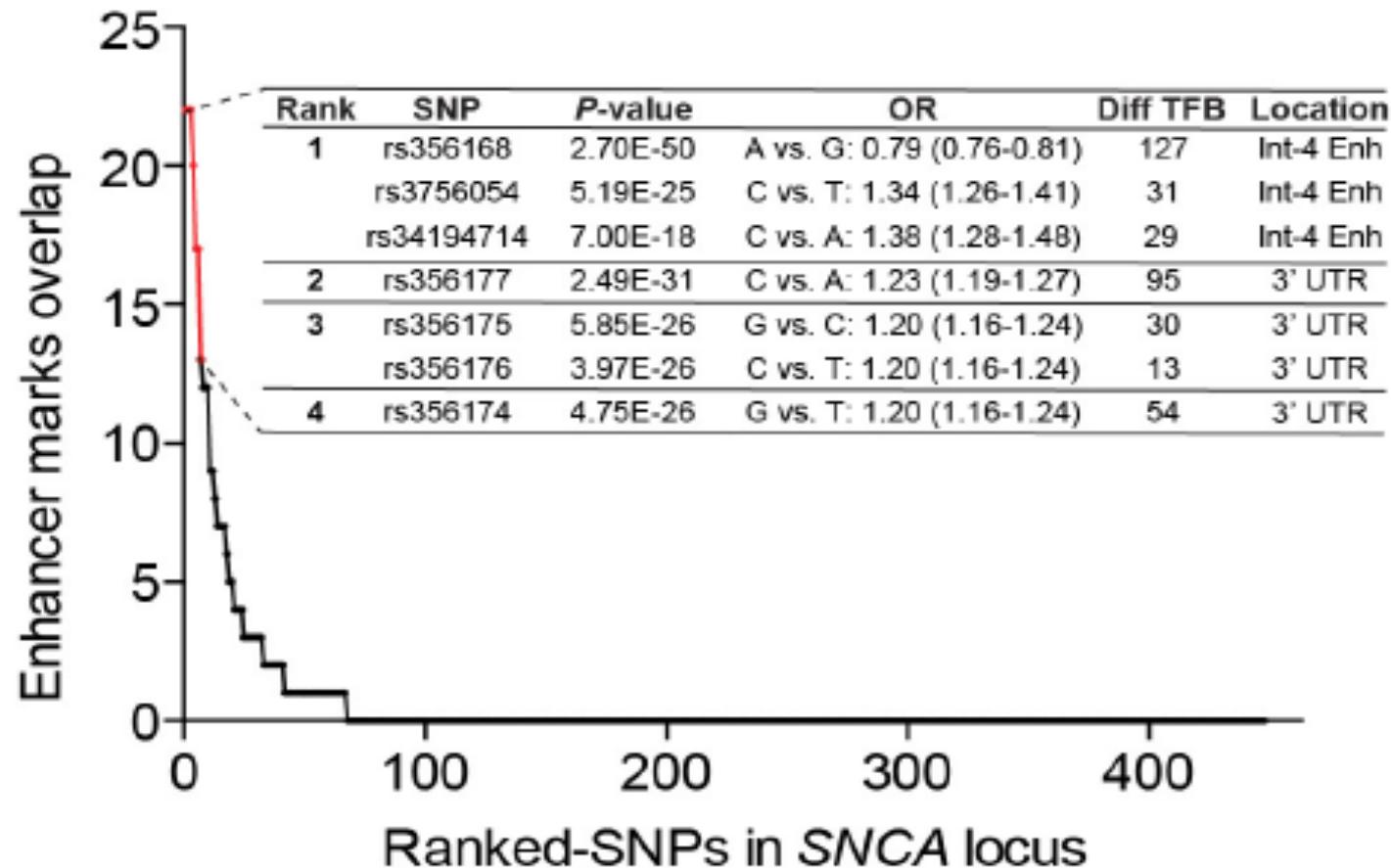
Function ?

# Which variants are causal?

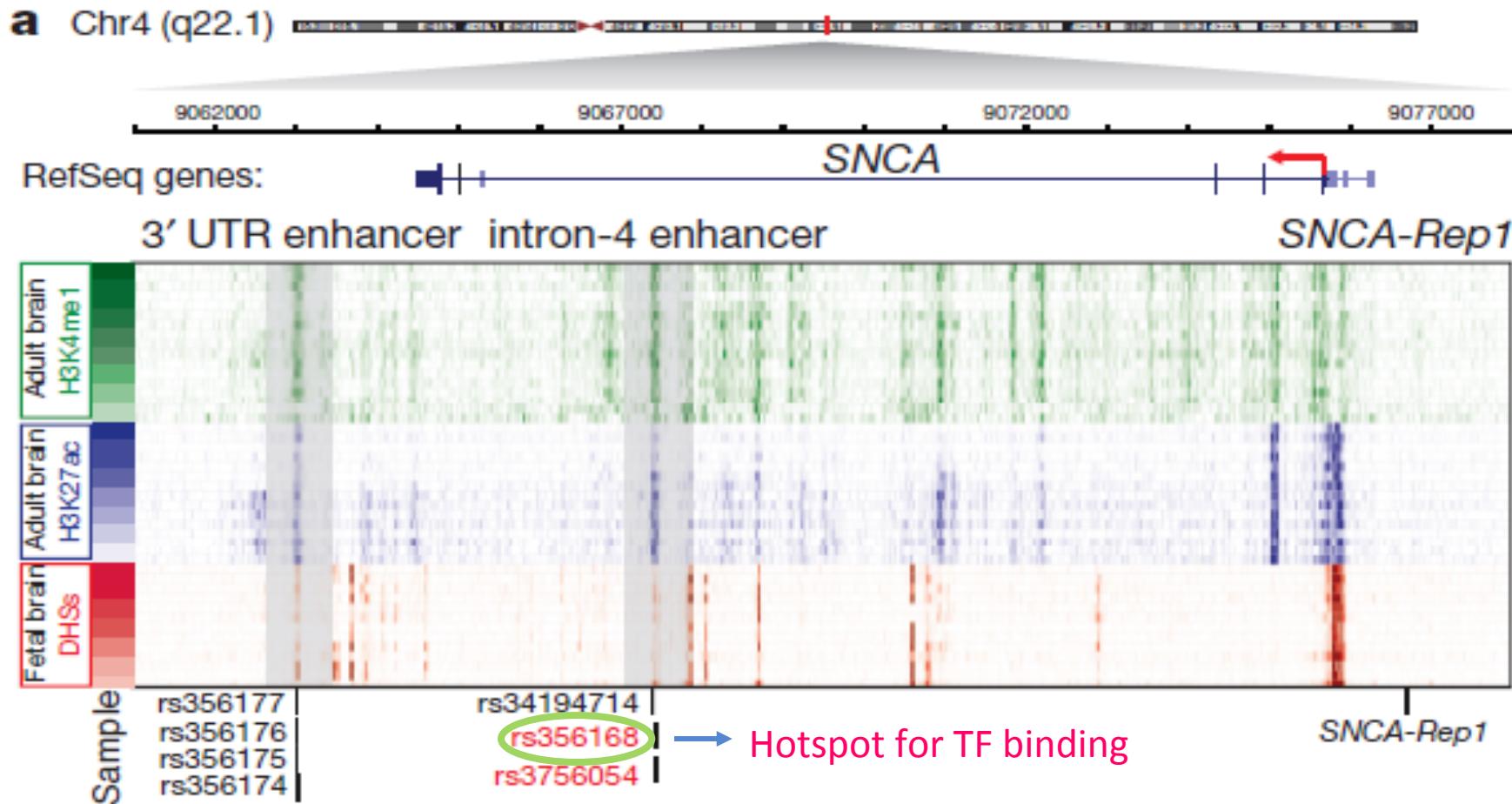
- Approach
  - Intersecting SNPs in SNCA locus with enhancer associated markers
  - Enhancer associated markers: methylation of H3K4, acetylation of H3K27 and DNase I hypersensitive sites



# Which variants are causal?

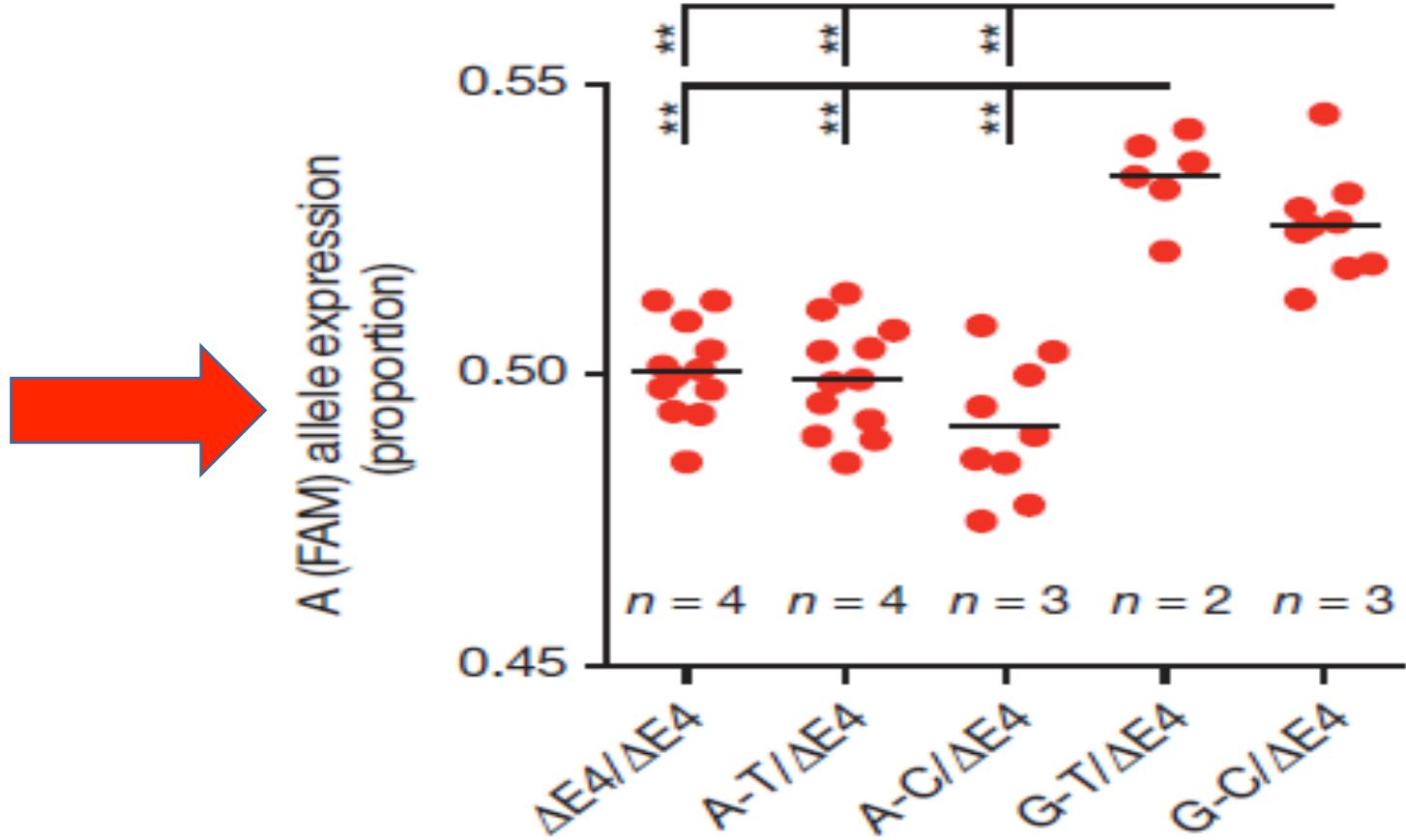
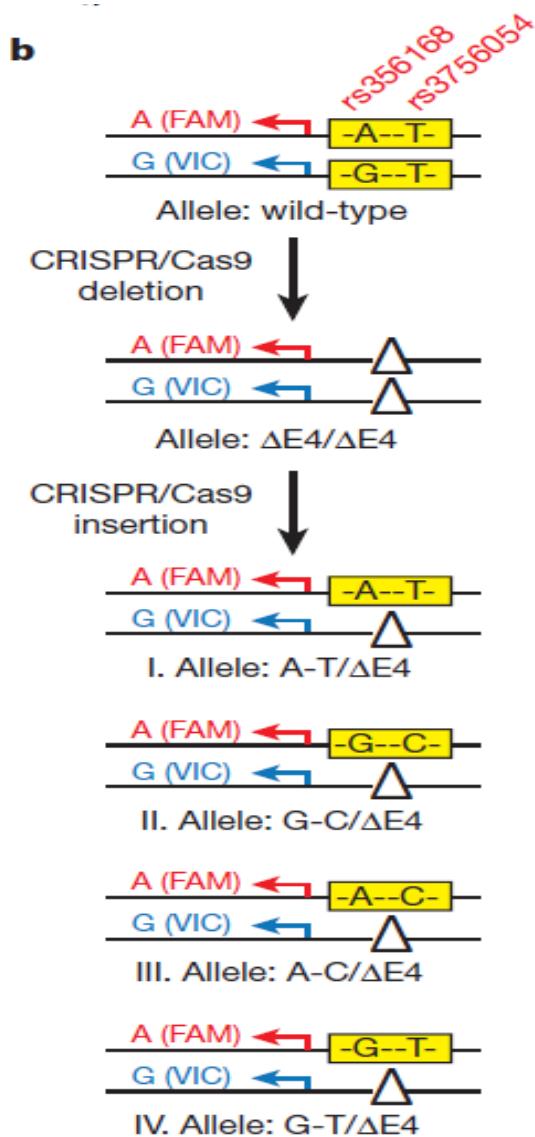


# Which variants are causal?



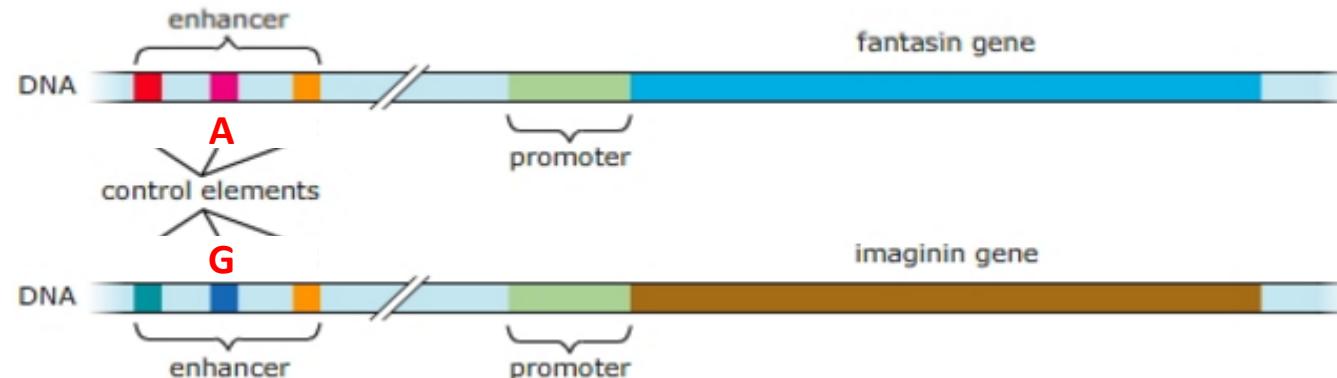
Out of 463 SNPs in SNCA locus, 7 overlapped with enhancer markers

# What is the function of intron 4- enhancer region?



# How does the variant regulate SNCA expression?

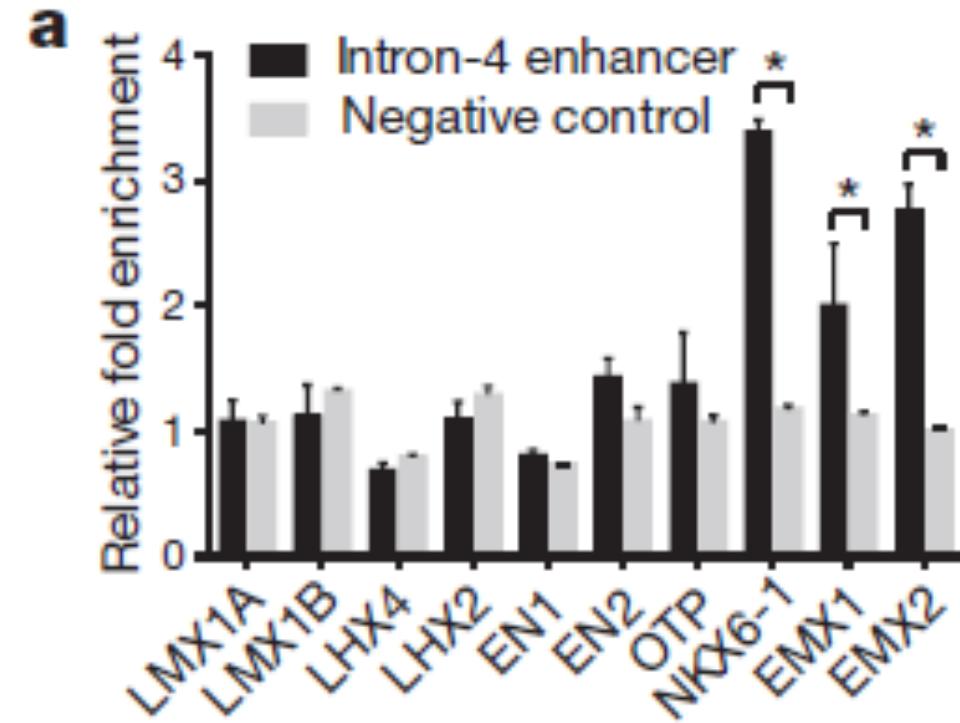
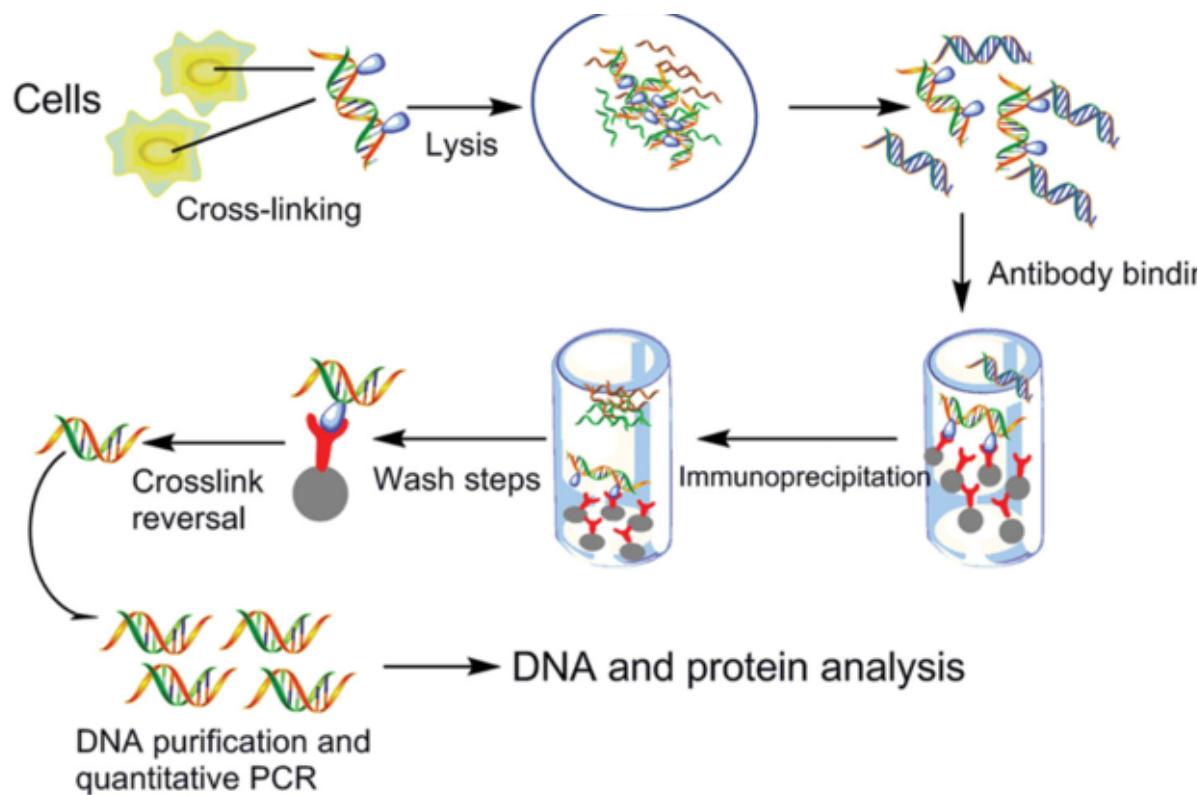
- Hypothesis
  - SNP-specific changes modify enhancer activity by altering transcription factor (TF) binding



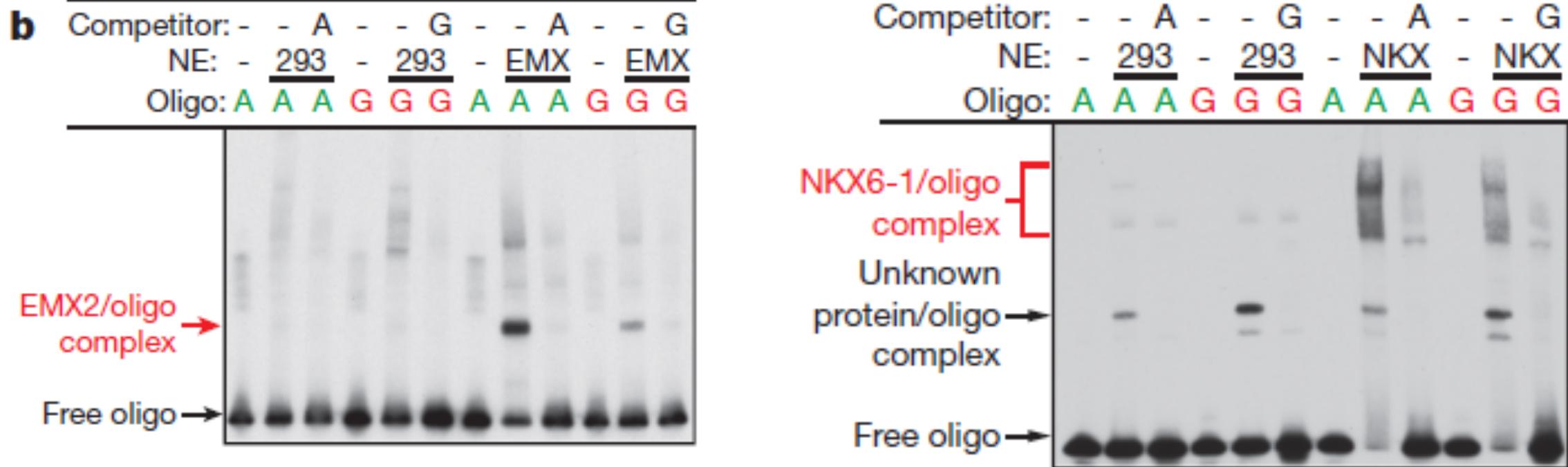
# How does the variant regulate SNCA expression?

- Selection of TF
  - Insilco analysis
  - 10 TF identified
- Specificity of TF
  - Chromatin-immuno precipitation qRT- PCR (Chip- PCR)
  - electrophoretic mobility shift assay analysis (EMSA)

# How does the variant regulate SNCA expression?



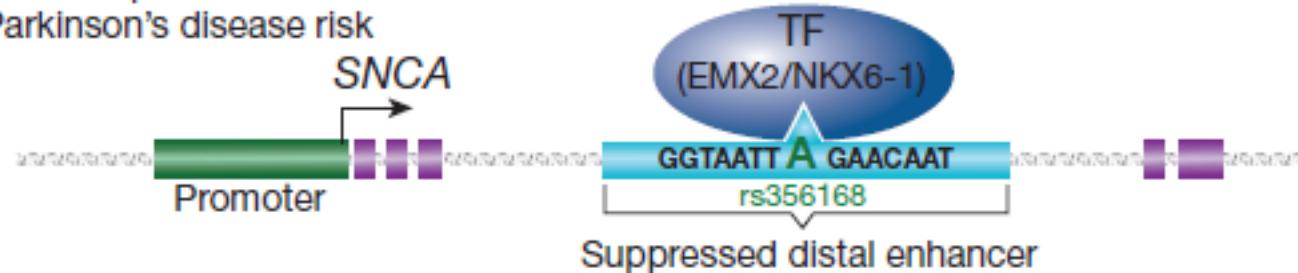
# How does the variant regulate SNCA expression?



# Overall mechanism

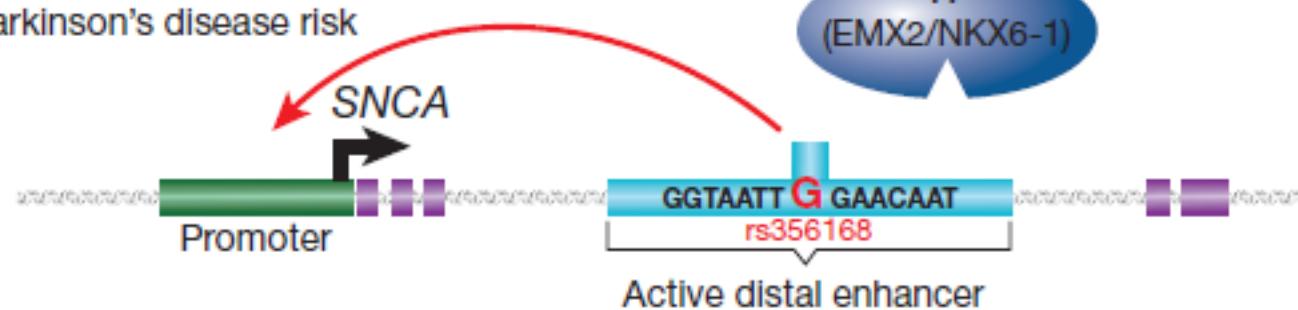
Parkinson's disease protective allele:

- Efficient TF binding
- Decreased SNCA expression
- Decreased Parkinson's disease risk



Parkinson's disease risk allele:

- Reduced TF binding
- Increased SNCA expression
- Increased Parkinson's disease risk



# Reference

- <https://www.ebi.ac.uk/training/online/sites/ebi.ac.uk.training.online/files/resize/GWAS/GWAS3-700x614.png>
- Ikegawa, Shiro. "A short history of the genome-wide association study: where we were and where we are going." *Genomics & informatics* 10.4 (2012): 220
- Sanna S, Jackson AU, Nagaraja R, Willer CJ, Chen WM, Bonnycastle LL, *et al.* Common variants in the *GDF5-UQCC* region are associated with variation in human height. *Nat Genet* 2008;40:198-203.
- Visscher, Peter M., et al. "10 years of GWAS discovery: biology, function, and translation." *The American Journal of Human Genetics* 101.1 (2017): 5-22.
- Soldner, F. *et al.* Parkinson-associated risk variant in distal enhancer of  $\alpha$ -synuclein modulates target gene expression. *Nature* 533, 95–99 (2016).
- Gallagher, M. D. & Chen-Plotkin, A. S. The Post-GWAS Era: From Association to Function. *Am. J. Hum. Genet.* 102, 717–730 (2018).

**Thank you for your time**

