Topic 12 – DNA technology

Branko Aleksic, MD, PhD
Department of Psychiatry, Nagoya University

Overview of genetic study (in psychiatry)

year 1990 2000 2005 2010

International HapMap Project

1000 Genomes
A Deep Catalog of Human Genetic Variation

hallmarks

Human Genome Project (1999)

HapMap Project(2002)

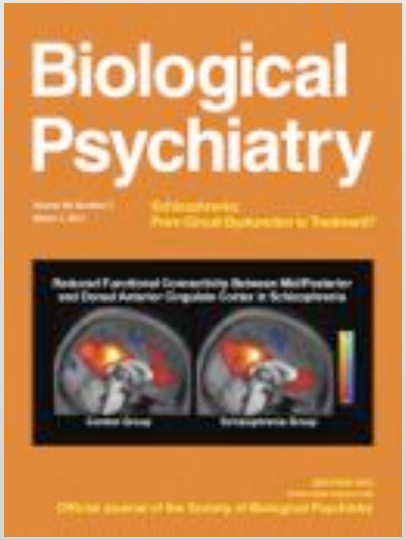
1000genomes Project(2008)

Method	Linkage	Association SNP	Association LD haplotype	GWAS	MEGA RESEQ
Sample size	Small (less than 500)		Medium (more than 1000)	Large (more than 10,000)	
Findings in psychiatry	APOE (1993) for AD	5HTTLPR 5HTTVNTR (1996) for MDD BP	DTNBP1(2002), NRG1(2002), for SCZ	ZNF804A(2008) MHC(2009) MIR137 (2011) for SCZ	

Today's topics:

- Japanese Schizophrenia GWAS
  - SNP-based analysis
  - Polygenic component analysis
  - Follow-up

Schizophrenia GWAS



M. Ikeda, B. Aleksic et al, 2011 Biol Psychiatry

History of Schizophrenia Diagnosis

- Emil Kraepelin (1856-1926): dementia praecox
  - Early onset
  - Progress to dementia
  - Cf. Demetia presenilis (Alzheimer's disease) and manic depressive insanity

- Eugen Bleuler (1857-1939): schizophrenia
  - did not necessarily have an early onset
  - Doesn’t (always) progress to dementia

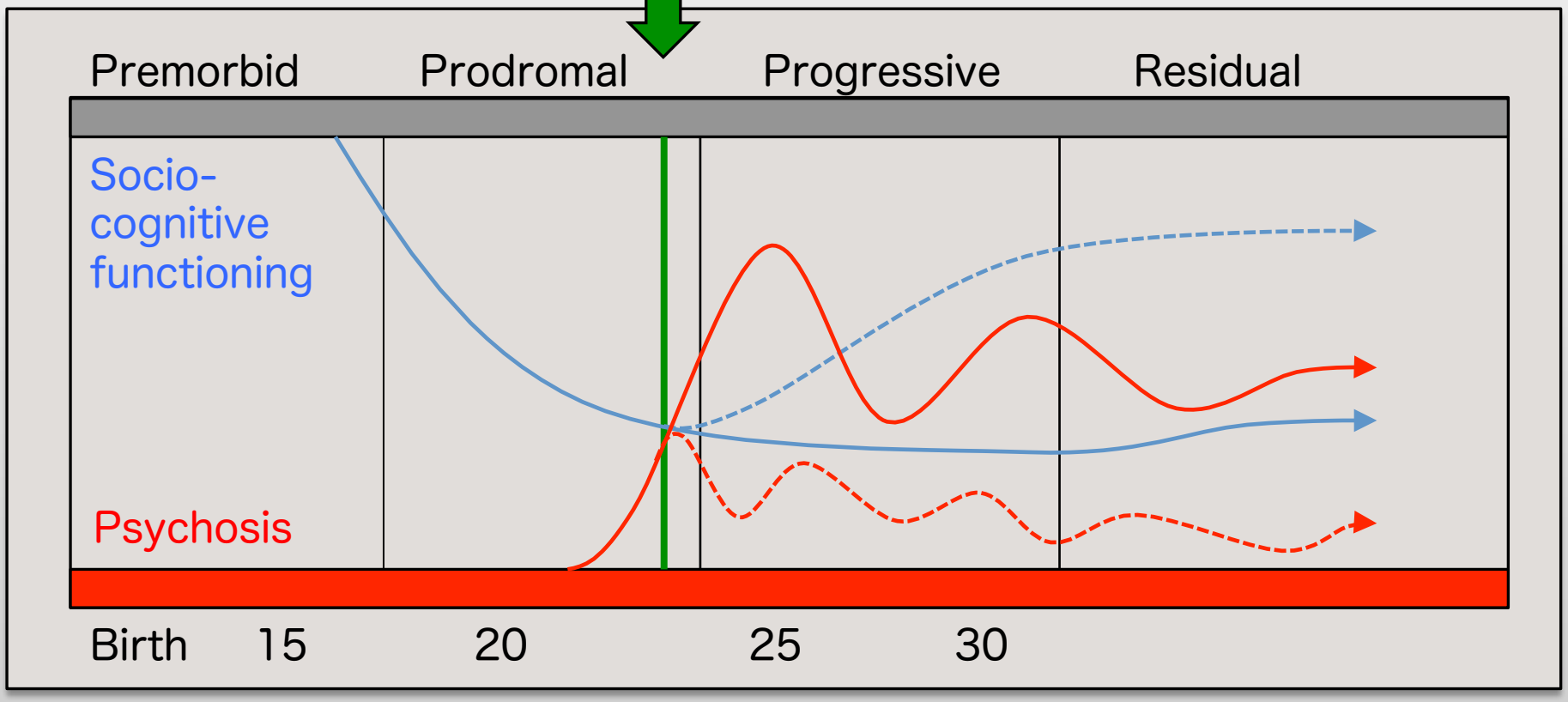
Schizophrenia now

- Common psychiatric disorder
- Onset is usually in early adulthood
- Characterized by:
  - Positive symptoms (hallucinations)
  - Negative symptoms (lack of motivation, social withdrawal)
  - Cognitive symptoms (decreasing of IQ)
- Usually prolonged medication is required

Clinical course of Schizophrenia

With treatment

Without treatment



Approximate risks

	Schizophrenia	Schizoaffective	Bipolar disorder
Risk in general population	0.8-1%	0.3%	0.3-1%
Risk to siblings/ first degree rels	10%	2-3%	5-10%
Monozygotic (MZ) twin concordance	45%	40%	40%
Dizygotic (DZ) twin concordance	5-10%	5%	5%
Risk in adoption studies	8%	-	14% (but small sample)

Heritability: examples

• Schizophrenia	80%+
• Autism	80%+
• Bipolar	60%
• Unipolar depression	40-70%
• Alcohol dependence	50-60%
• Anxiety disorders	20-30%

Our Platform (affy 5.0)

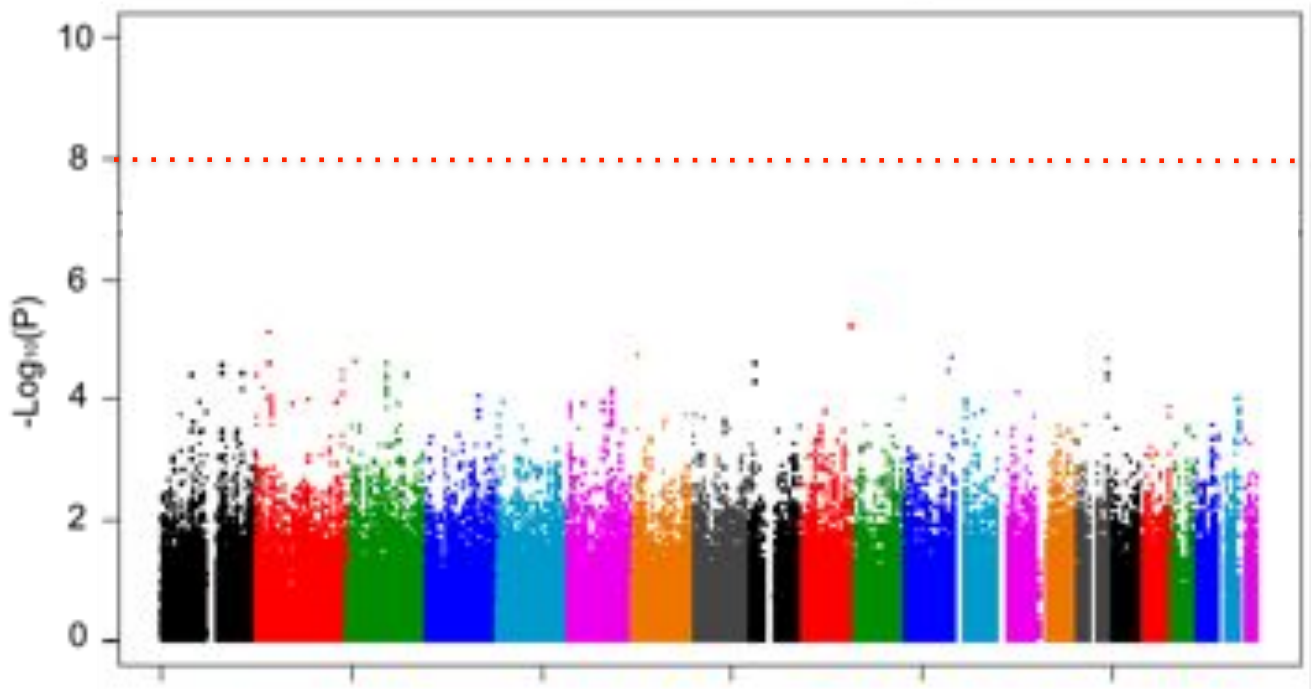
Samples (SNP-based analysis)

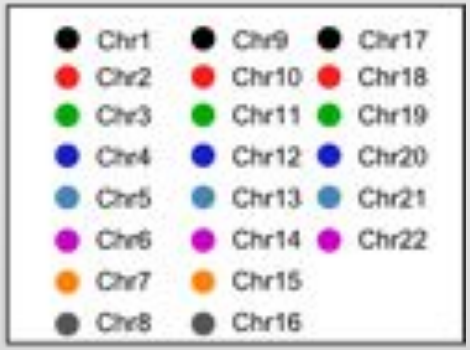
Meta Analysis

- GWAS (JPN)
  - 575 schizophrenia
  - 564 controls
- Rep_JPN
  - Replication1 (main sample for follow-up)
    - 1511 schizophrenia
    - 1517 controls
  - Replication2 (additional con: JPN, public database)
    - 934 controls (Genotyped by Illumina550)
- Replication3 (WTCCC_scz: UK)
  - 479 schizophrenia
  - 2938 controls (Genotyped by Affy 500K)

Manhattan Plot

: GWAS sample (~1,200 sample)





NO SNPs with genome-wide significance ***(P<5x10^{-8})***

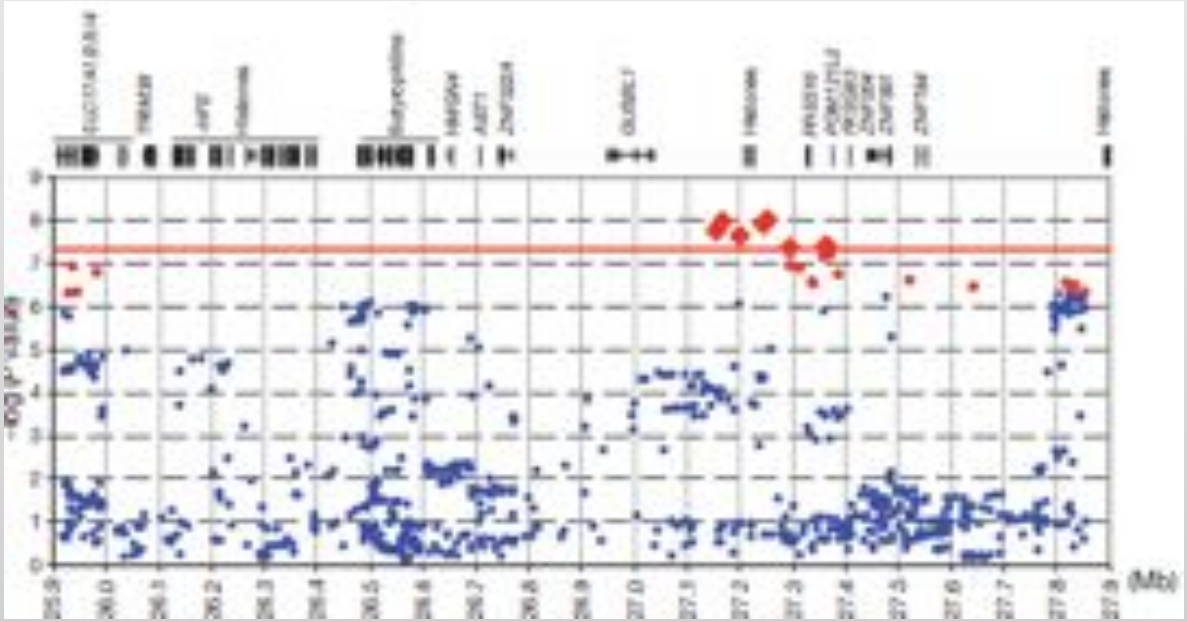
Result: CMH analysis
: meta analysis (~8,000 sample)

			Meta _all		
CHR	SNP	closest gene	MAF	P_{CMH}	OR
2	rs11895771	SULT6B1	0.49	**$3.7X10^{-5}$**	0.84
7	rs1011131	LOC392288	0.070	$1.2X10^{-4}$	1.30
14	rs1176970	LOC644919	0.15	$1.4X10^{-4}$	1.22
1	rs4908274	COL11A1	0.28	$3.1X10^{-4}$	1.20
6	rs2294424	C6orf105	0.41	$5.0X10^{-4}$	1.15
2	rs13010889		0.15	0.0011	0.85
2	rs17026152		0.26	0.0012	0.85
6	rs2787566	GRIK2	0.039	0.0014	1.34
6	*rs2071286*	*NOTCH4*	*0.19*	*0.0014*	*0.87*
8	rs17462248		0.20	0.0017	1.16

Again, NO SNPs with genome-wide significance ($P<5x10^{-8}$)

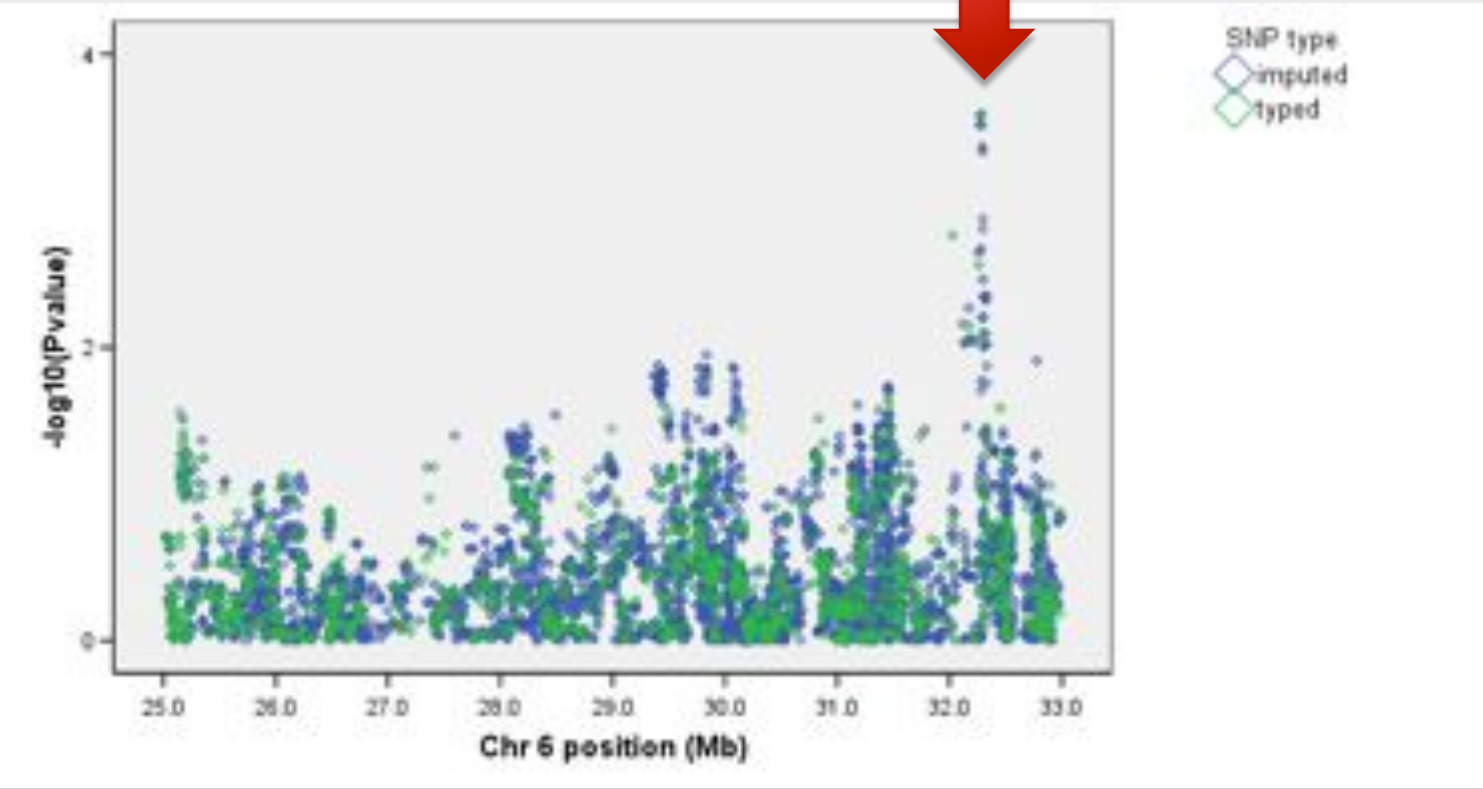
Promising candidate region reported in already published GWASes focused on schizophrenia

- Several candidate genes with genome-wide significance ($5X10^{-8}$)
  - <u>MHC region on Chr6</u> by SGENE, ISC, MGS, PGC

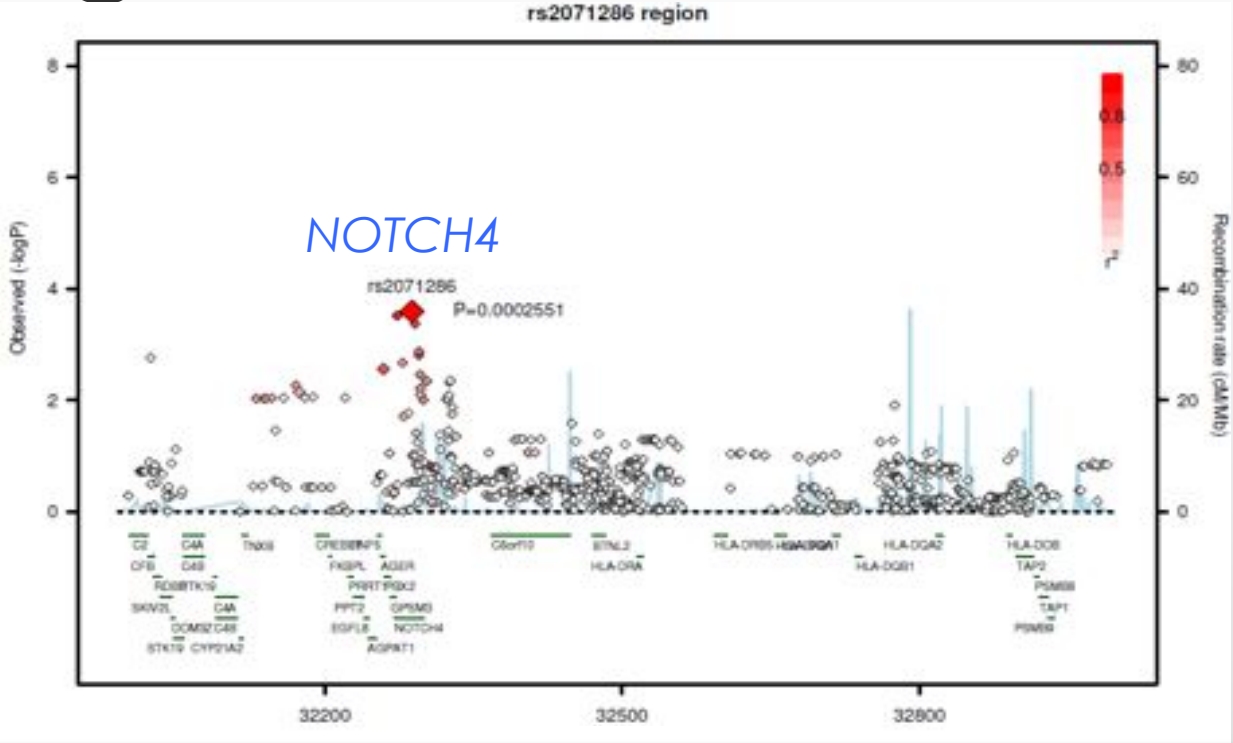


Japanese GWAS data low magnification on MHC

- MHC region on Chr6

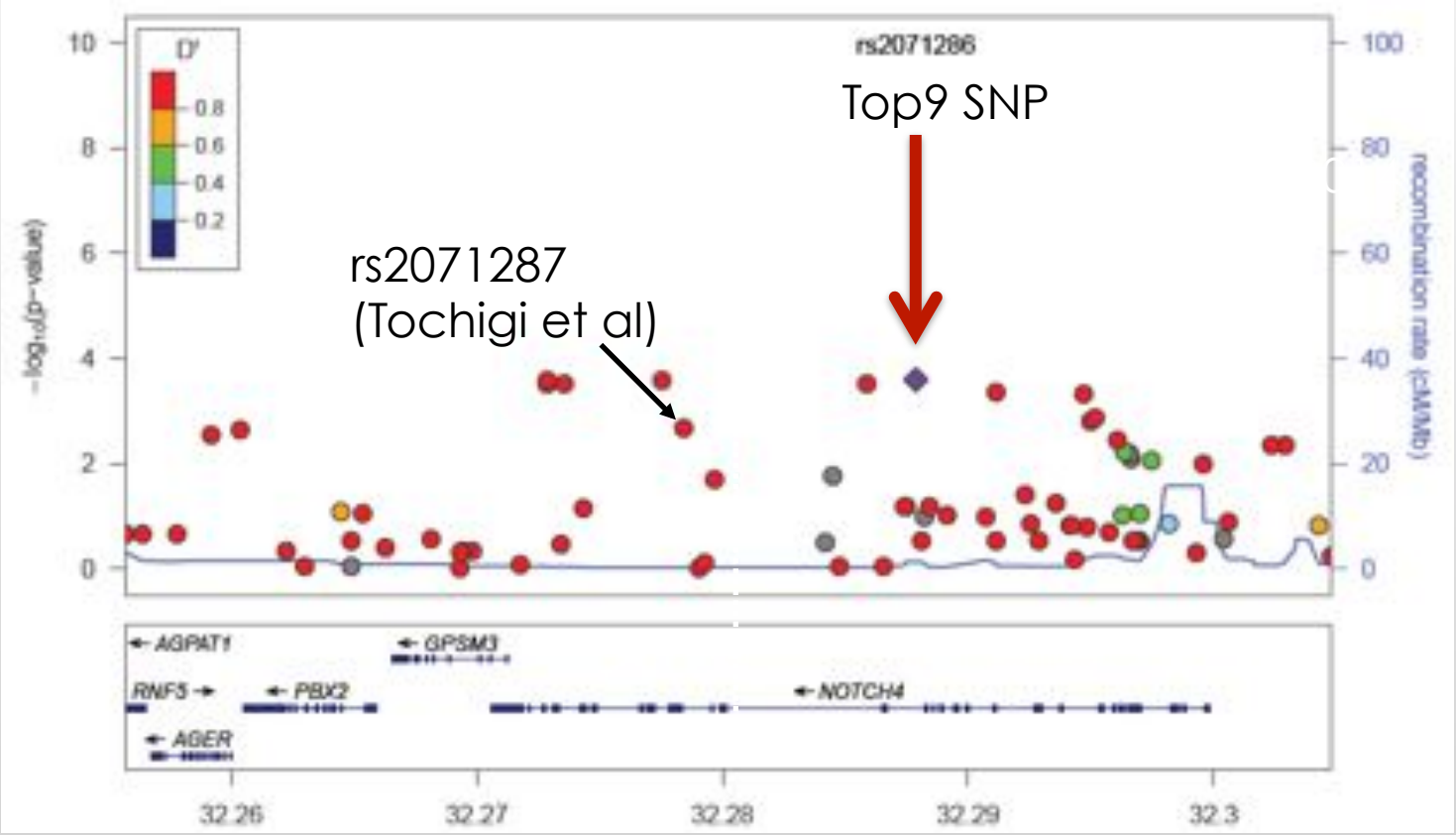


Zooming in!



			Meta _all				
CHR	SNP	closest gene	MAF	P_{CMH}	OR	L95	U95
2	rs11895771	SULT6B1	0.4892	$3.7X10^{-5}$	0.84	0.7708	0.9117
7	rs1011131	LOC392288	0.07039	$1.2X10^{-4}$	1.30	1.136	1.48
14	rs1176970	LOC644919	0.1524	$1.4X10^{-4}$	1.22	1.10	1.354
1	rs4908274	COL11A1	0.28	$3.1X10^{-4}$	1.20	1.087	1.324
6	rs2294424	C6orf105	0.4128	$5.0X10^{-4}$	1.15	1.063	1.244
2	rs13010889		0.1524	0.0011	0.85	0.7716	0.9375
2	rs17026152		0.2618	0.0012	0.85	0.7673	0.9369
6	rs2787566	GRIK2	0.03883	0.0014	1.34	1.12	1.609
6	rs2071286	NOTCH4	0.1929	0.0014	0.87	0.791	0.9459
8	rs17462248		0.1988	0.0017	1.16	1.056	1.268

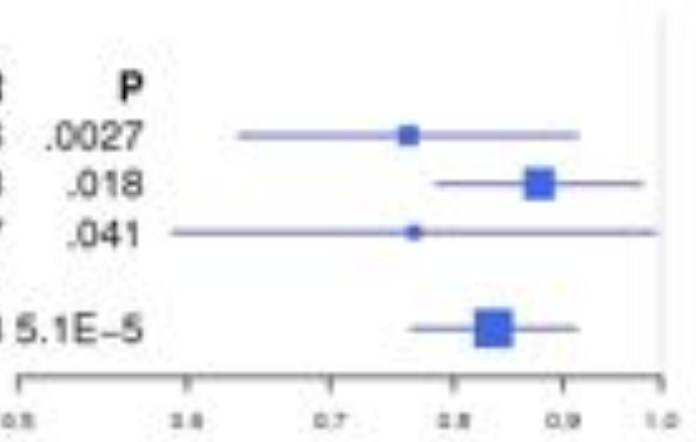
Maximum magnification



Follow up analysis for *NOTCH4* (rs2071287)

Meta_JPN + previously published data (JPN)

Study	Case T allele	Control T allele	OR	P
JPN_GWAS	0.31	0.37	0.76	.0027
Rep–JPN	0.31	0.34	0.88	.018
Tochigi(2007)	0.32	0.38	0.77	.041
Summary			0.83	5.1E–5



~2,500 SCZ
~2,500 CON

Further Replication

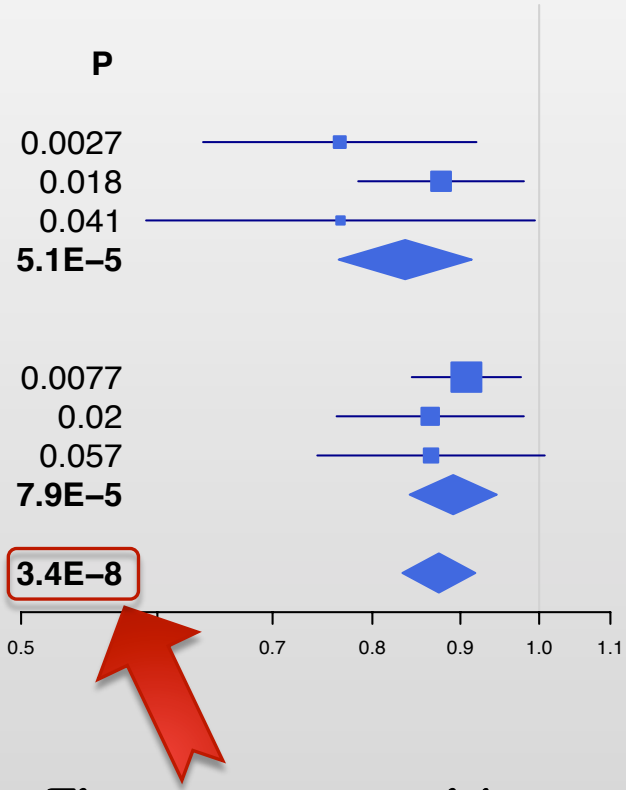
- Samples used in the previous paper (Ikeda et al)
  - JPN_GWAS: 542 SCZ vs 525 CON
  - REP_JPN: 1471 SCZ vs 1493 CON
  - Tochigi et al: 241 SCZ vs 290 CON

- New samples
  - REP1:3150 SCZ vs 3483 CON, Mid-East JPN
  - REP2: 672 SCZ vs 5321 CON, Mid-West JPN
  - REP3: 569 SCZ vs 1622 CON, South Island: Shikoku Island JPN

Extra samples : ~4,400 SCZ
~10,000 CON

Meta-analysis

	N cases	N controls	Case T allele	Control T allele	OR	P
Prior study (Ikeda et al, 2010)						
JPN_GWAS (2010)	542	525	0.31	0.37	0.76	0.0027
Rep_JPN (2010)	1471	1493	0.31	0.34	0.88	0.018
Tochigi et al (2007)	241	290	0.32	0.38	0.77	0.041
–Summary1 (Ikeda et al, 2010)	**2254**	**2308**	**0.31**	**0.35**	**0.83**	**5.1E–5**
Current study						
REP1 (Mid East JPN)	3173	3540	0.33	0.35	0.91	0.0077
REP2 (Mid West JPN)	672	5321	0.31	0.34	0.86	0.02
REP3 (South Island JPN: Shikoku)	569	1622	0.28	0.31	0.87	0.057
–Summary2 (replication:REP1–3)	**4414**	**10483**	**0.32**	**0.34**	**0.89**	**7.9E–5**
ALL combined	**6668**	**12791**	**0.32**	**0.34**	**0.87**	**3.4E–8**



Total

~6,700 SCZ

~13,000 CON

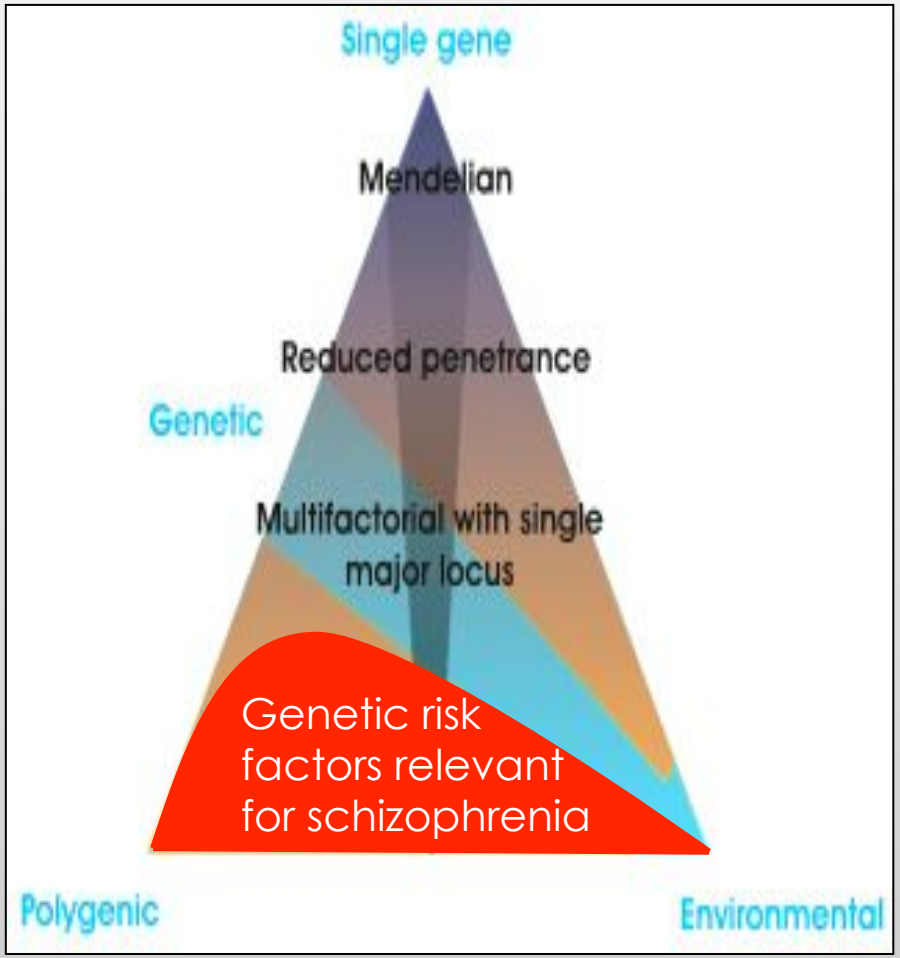
First genome wide significant P-value in Japanese schizophrenia population based sample

Today's topics:

- Japanese Schizophrenia GWAS
  - SNP-based analysis
  - Polygenic component analysis
  - Follow-up

Next we explored the concept of polygenic Component Analysis (PSA) using our dataset…

- We know that schizophrenia is polygenic
- We know that in case of common SNPs risk effect size is small
- Recent studies showed that in case of schizophrenia common variants do have an important role en masse (cumulative risk effect)

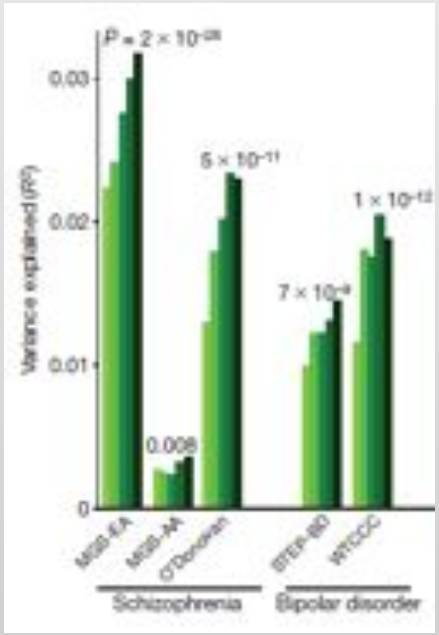


GWAS in schizophrenia (2)

Polygenic Component Analysis

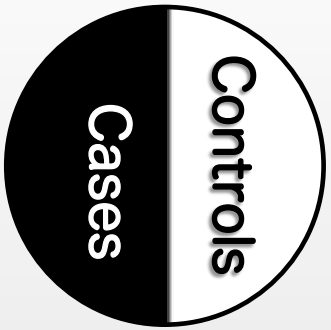
- Method used by International SCZ Consortium (ISC)
- Non stringent definition of risk SNPs *(e.g. P<0.3)*

$$\Sigma xi$$



Two GWAS sample sets are needed for polygenic component analysis !

GWAS data set



Typical genome wide association analysis

Polygenic component analysis

Discovery sample set

Target sample set

Discovery sample set

	SNP1	SNP2	SNP3	SNP4	-----
Risk allele	A	T	G	C	
P	.10	.26	.68	.05	

Calculate P value for SNPs/alleles based on discovery sample

Discovery sample set

$P_T<0.5$

	SNP1	SNP2	SNP3	SNP4	-----
Risk allele	A	T	G	C	
P	.10	.26	.68	.05	

Assign non stringent statistical threshold (P_T : P threshold)and define **risk SNPs/alleles (e.g. $P_T<0.5$)** based on discovery sample

Discovery sample set

$P_T<0.5$

	SNP1	SNP2	SNP3	SNP4	-----
Risk allele	A	T	G	C	
P	.10	.26	.68	.05	

Check the genotypes in TARGET sample

Target sample set

	SNP1	SNP2	SNP3	SNP4
case1	AA	TT		CT
case2	AT	TT		CT
con1	AT	TC		TT
con2	TT	TC		TT

Discovery sample set

$P_T<0.5$

	SNP1	SNP2	SNP3	SNP4	-----
Risk allele	A	T	G	C	
P	.10	.26	.68	.05	

Count number of risk alleles in cases and controls

(TARGET sample)

Target sample set

	SNP1	SNP2	SNP3	SNP4
case1	AA	TT		CT
case2	AT	TT		CT
con1	AT	TC		TT
con2	TT	TC		TT

Polygenic score

SNP1 SNP2 SNP4

2 + 2 + 1 = 5

1 + 2 + 1 = 4

1 + 1 + 0 = 2

0 + 1 + 0 = 1

Discovery sample set

$P_T<0.5$

	SNP1	SNP2	SNP3	SNP4	-----
Risk allele	A	T	G	C	
P	.10	.26	.68	.05	

Calculate mean polygenic score

(TARGET sample)

Target sample	SNP1	SNP2	SNP3	SNP4
case1	AA	TT		CT
case2	AT	TT		CT
con1	AT	TC		TT
con2	TT	TC		TT

Polygenic score

SNP1 SNP2 SNP4	Mean
2 + 2 + 1 = 5	4.5
1 + 2 + 1 = 4	
1 + 1 + 0 = 2	1.5
0 + 1 + 0 = 1	

Discovery sample set

$P_T<0.5$

	SNP1	SNP2	SNP3	SNP4	-----
Risk allele	A	T	G	C	
P	.10	.26	.68	.05	

Compare mean values

TARGET sample

Target sample

	SNP1	SNP2	SNP3	SNP4
case1	AA	TT		CT
case2	AT	TT		CT
con1	AT	TC		TT
con2	TT	TC		TT

Polygenic score

SNP1 SNP2 SNP4	Mean
2 + 2 + 1 = 5	4.5
1 + 2 + 1 = 4	
1 + 1 + 0 = 2	1.5
0 + 1 + 0 = 1	

Samples used for Polygenic Score Analysis (PSA)

- Datasets
  - JPN
  - UK
    - WTCCC Schizophrenia — 479 SCZ vs 2938 CON (O'Donovan et al ,2008)
    - WTCCC Bipolar — 1868 BP vs 2938 CON (WTCCC, 2007)

- Statistical analysis
  - Logistic Regression
  - P value
  - Nagelkerke Pseudo R^2 as measure of explained variability
    - The more variability explained, the better the model

Polygenic component analysis (PCA) -Central hypothesis-

- Can PCA predict
  - status (schizophrenia or healthy) within Japanese GWAS sample
  - status within UK sample based on Japanese SNPs and vice versa

Discovery/Target pair
Within JPN samples: random Division

Case

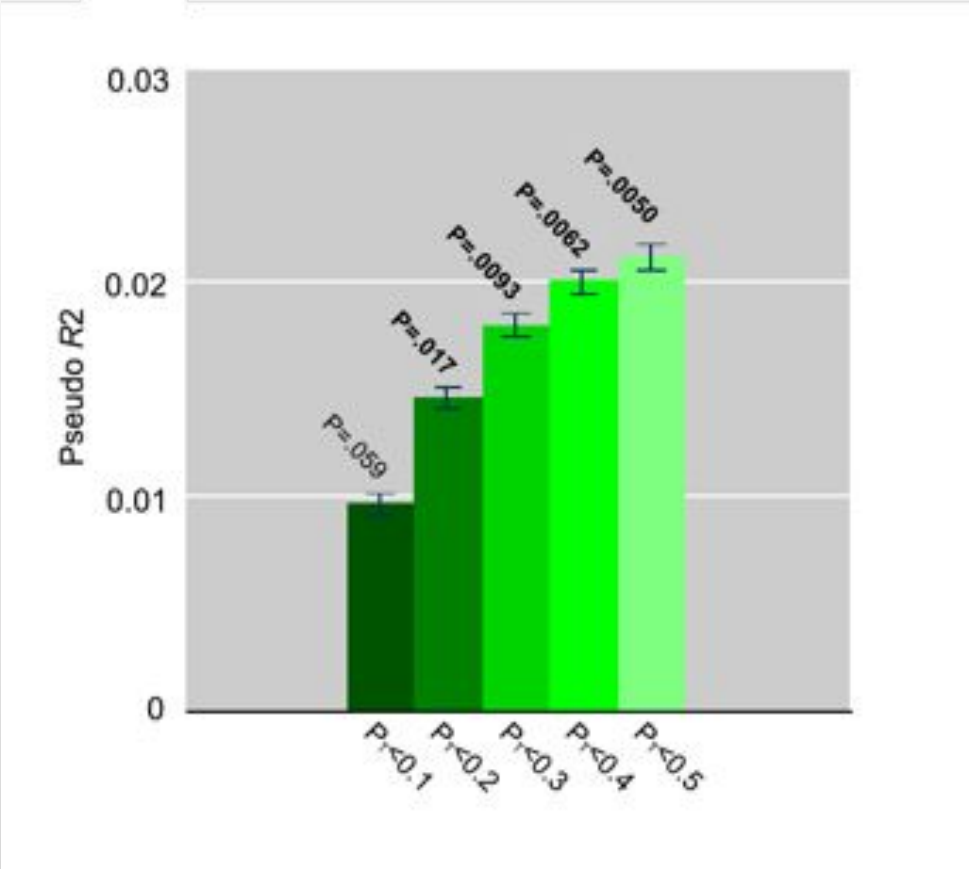
Control

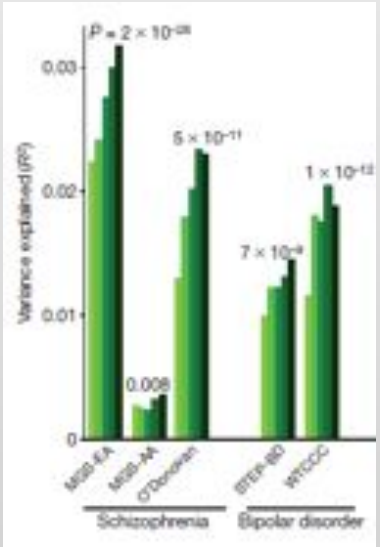
Only one GWAS dataset in JPN sample …

Discovery set
For ‘risk’ alleles

Target set
For polygenic scores

Within JPN samples
1st/2nd: discovery/target pair



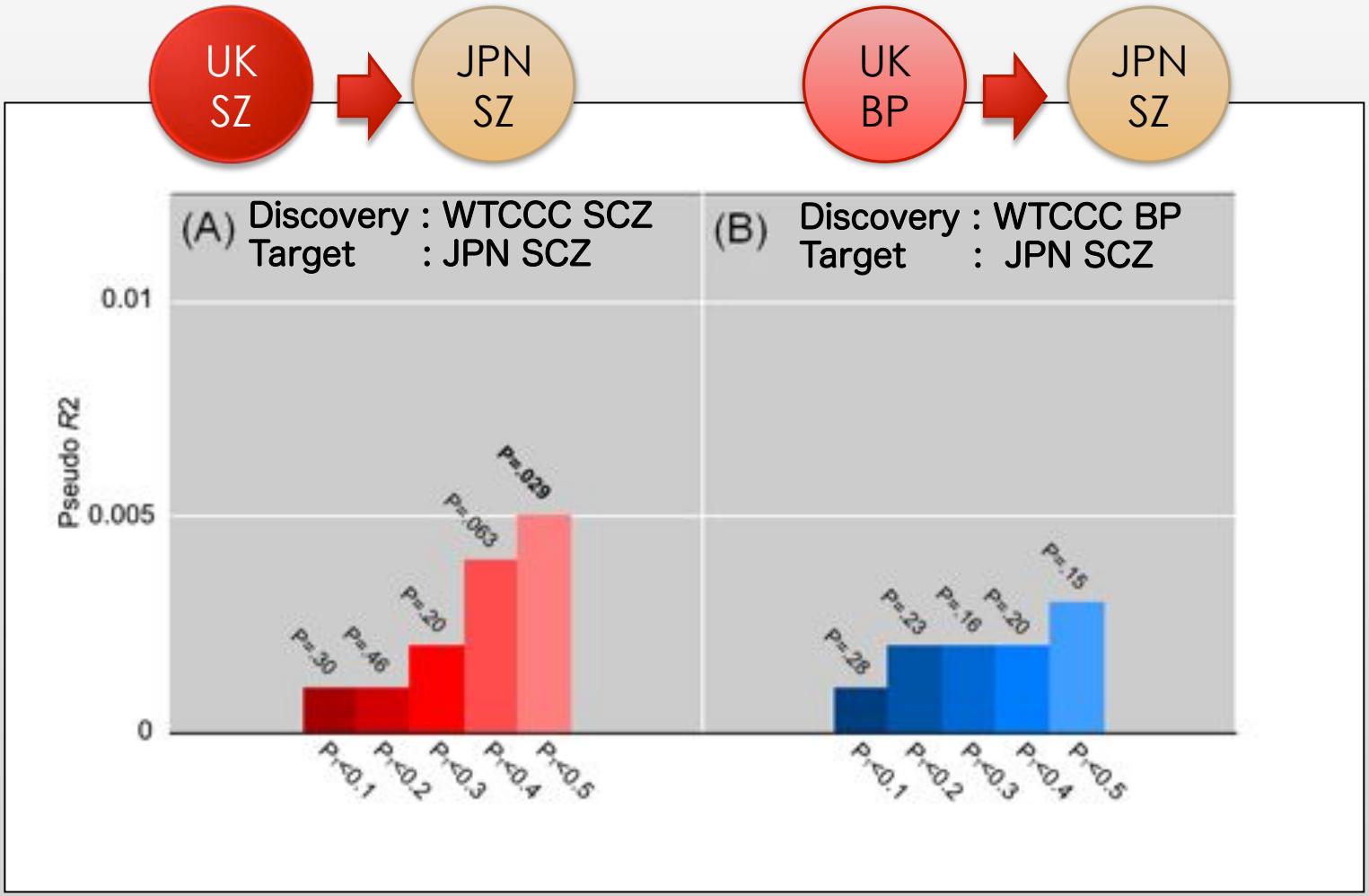


Significant enrichment (risk alleles)
R^2~2% (P_T<0.5) → en masse increase the risk

R^2 in ISC 3%

PCA Results

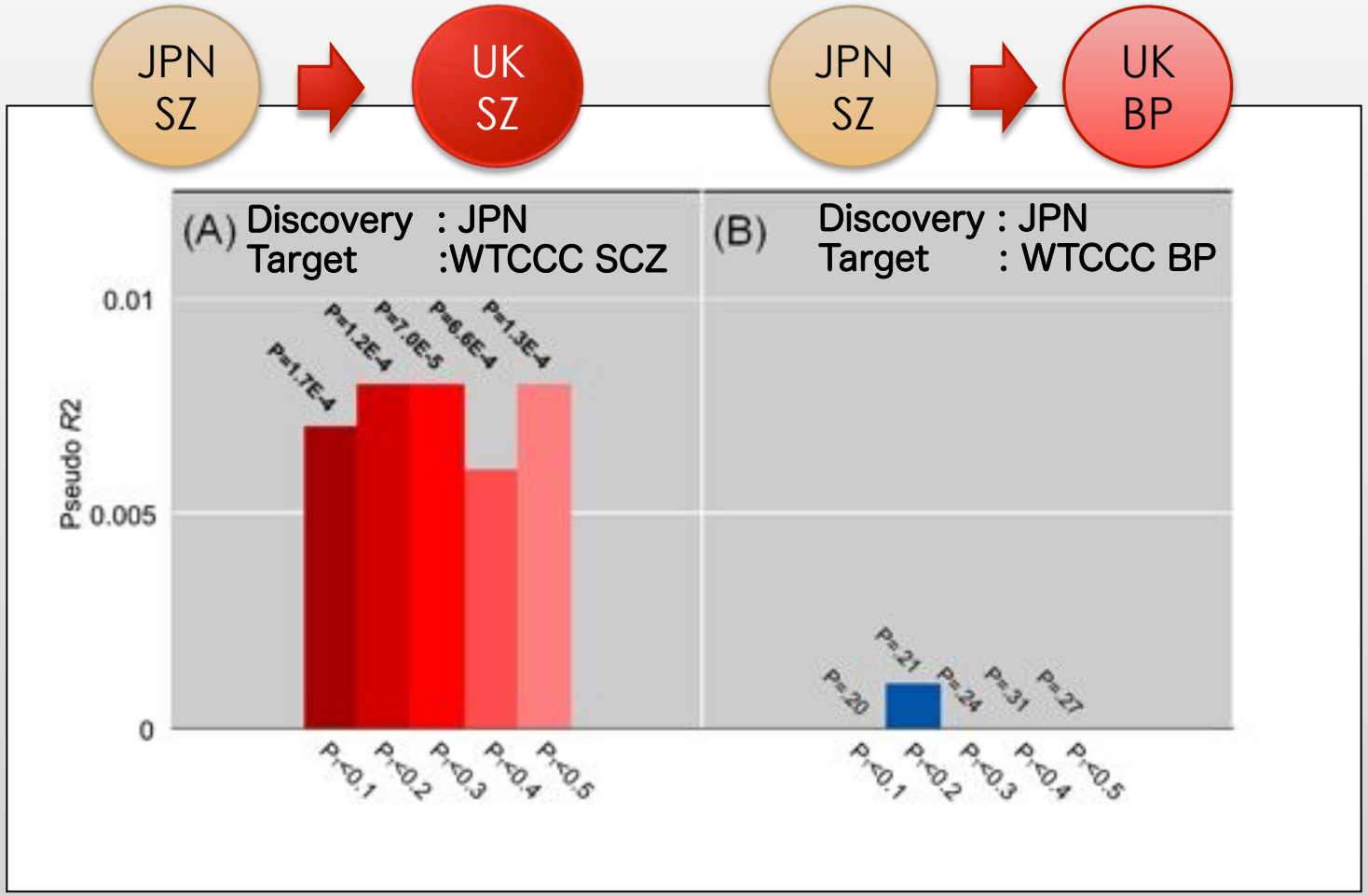
(UK schizophrenia and bipolar SNPs vs. Japanese schizophrenia SNPs)



Low pseudoR^2~0.5% (P_T<0.5)

PCA Results

(Japanese schizophrenia SNPs vs. UK schizophrenia and bipolar SNPs)



Low pseudoR^2~0.7% (P_T<0.5)

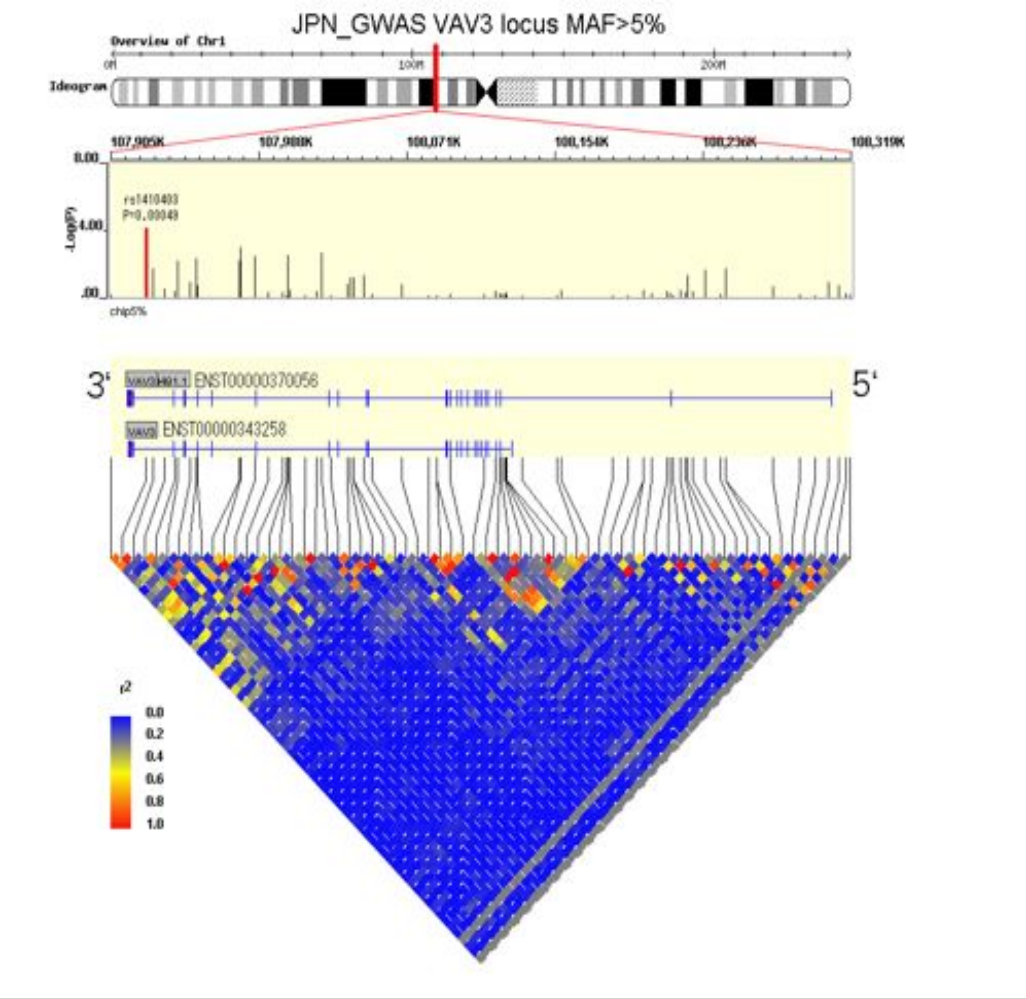
Summary: GWAS

- Effect size of risk SNPs (common) is very small
- It is important to chase sub GWAS P-threshold in underpowered sample sets
- We replicated the observation of a polygenic component to schizophrenia within the Japanese population ($p = .005$)
- Our trans Japan-UK analysis of schizophrenia also revealed a significant correlation (best $p = 7.0 \times 10^{-5}$) in the polygenic component across populations
- These results indicate a shared polygenic risk of schizophrenia between Japanese and Caucasian samples, although we did not detect unequivocal evidence for a novel susceptibility gene for schizophrenia

Today's topics:

- Japanese Schizophrenia GWAS
  - SNP-based analysis
  - Polygenic component analysis
  - Follow-up

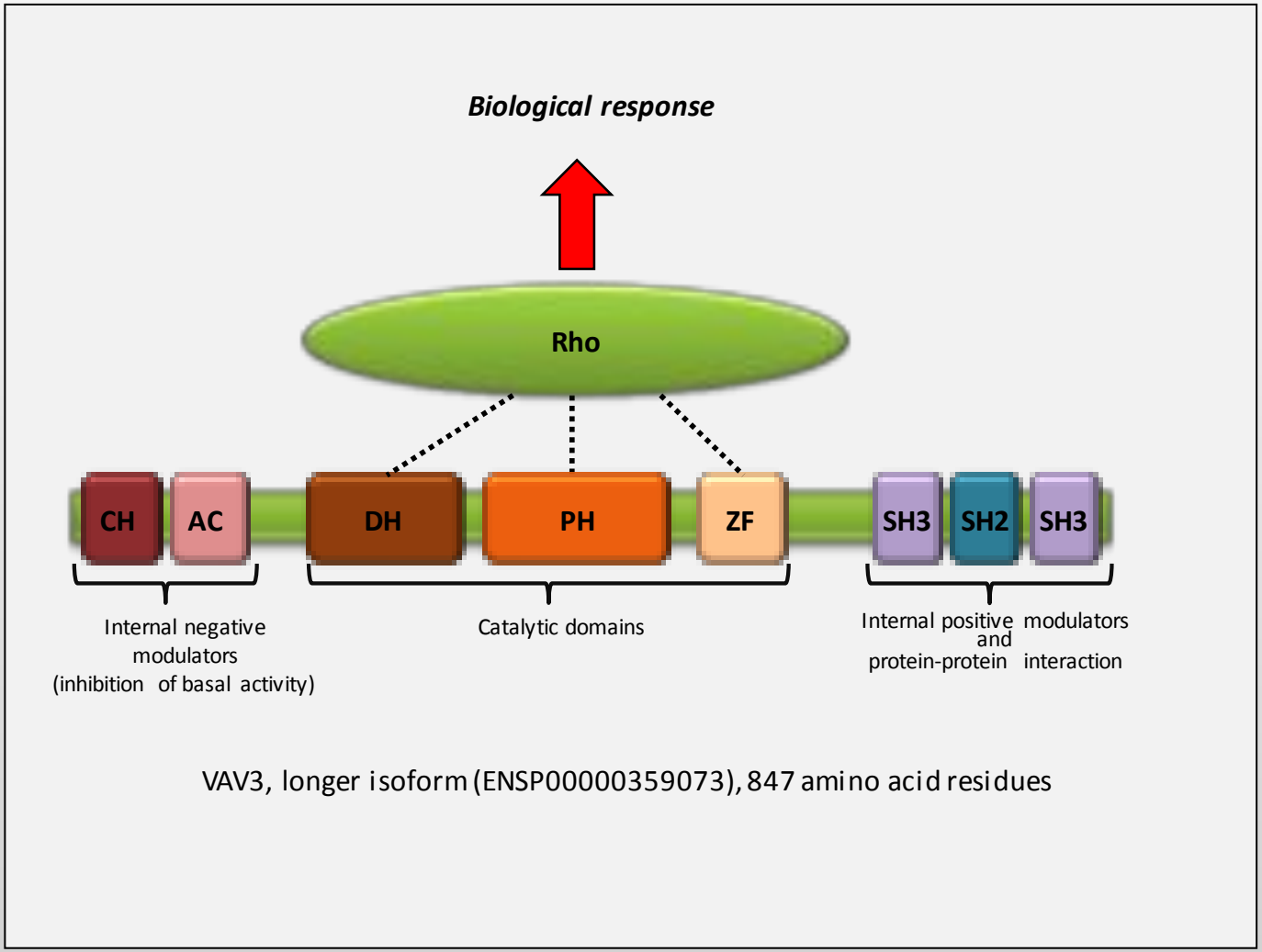
Association signal in JGWAS



Gene centered data:

- Location: 1p13.3
- Size: 400 kbps
- 2 isoforms has been reported

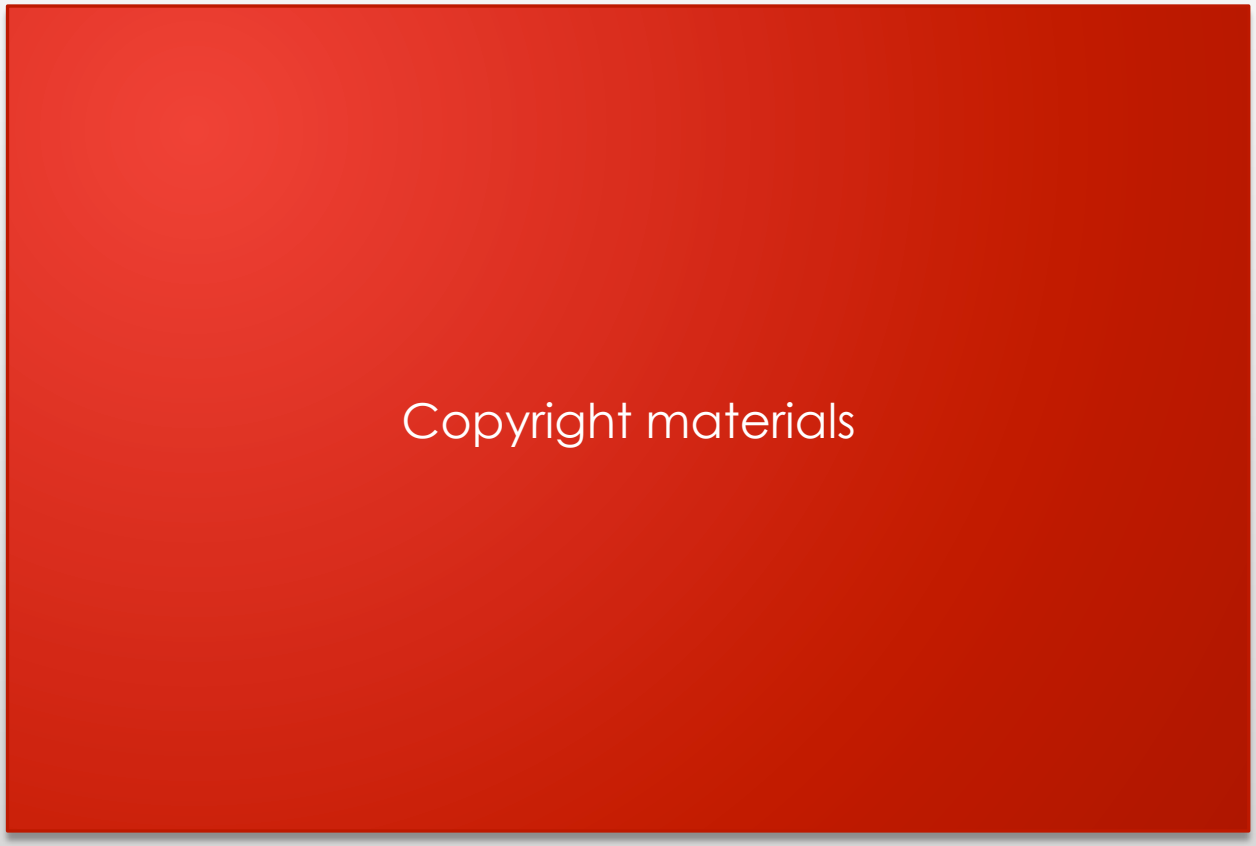
VAV3 structure

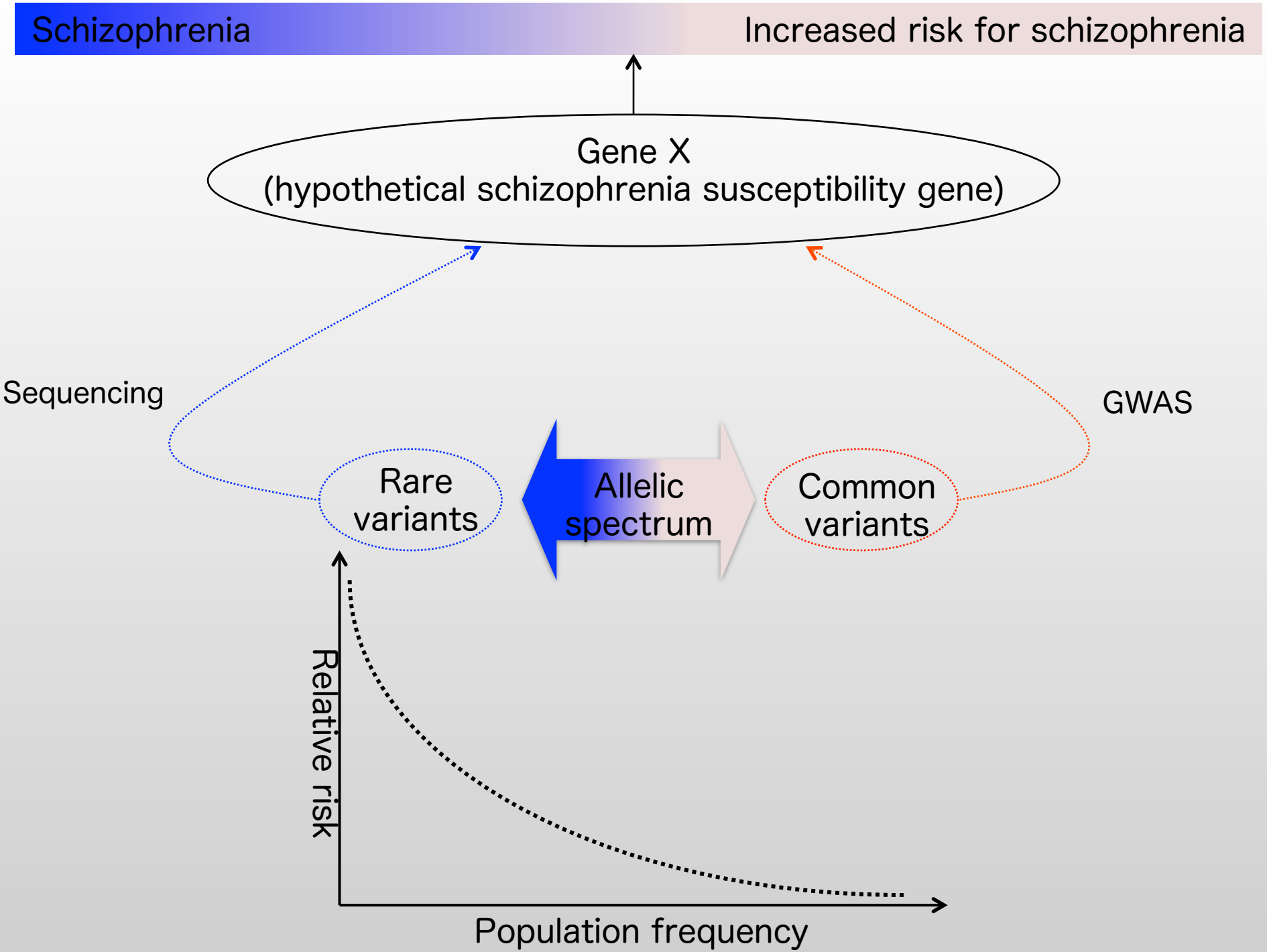


MOLECULAR AND CELLULAR BIOLOGY, Nov. 1999, p. 7870–7885

VAV3 and Schizophrenia

- Related to the axon guidance (process identified as disturbed in schizophrenic patients)
- Identified by linkage study in Japanese population




Schizophrenia
Increased risk for schizophrenia
Gene X
(hypothetical schizophrenia susceptibility gene)
Sequencing
GWAS
Rare variants
Allelic spectrum
Common variants
Relative risk
Population frequency

Our study of *VAV3*

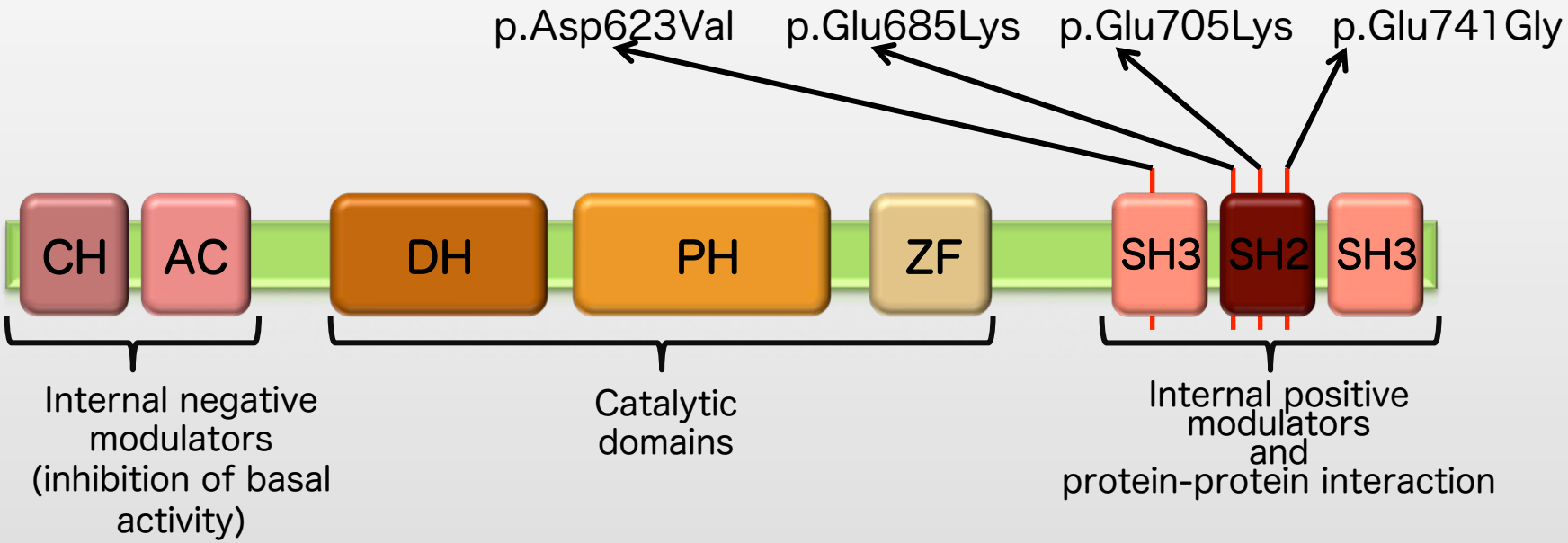
- Mutation screening and association analysis of associated rare variants
- Check the effect of associated common SNP on brain morphology (MRI)

FOCUS ON THE RARE VARIANTS

Mutation screening -strategy-

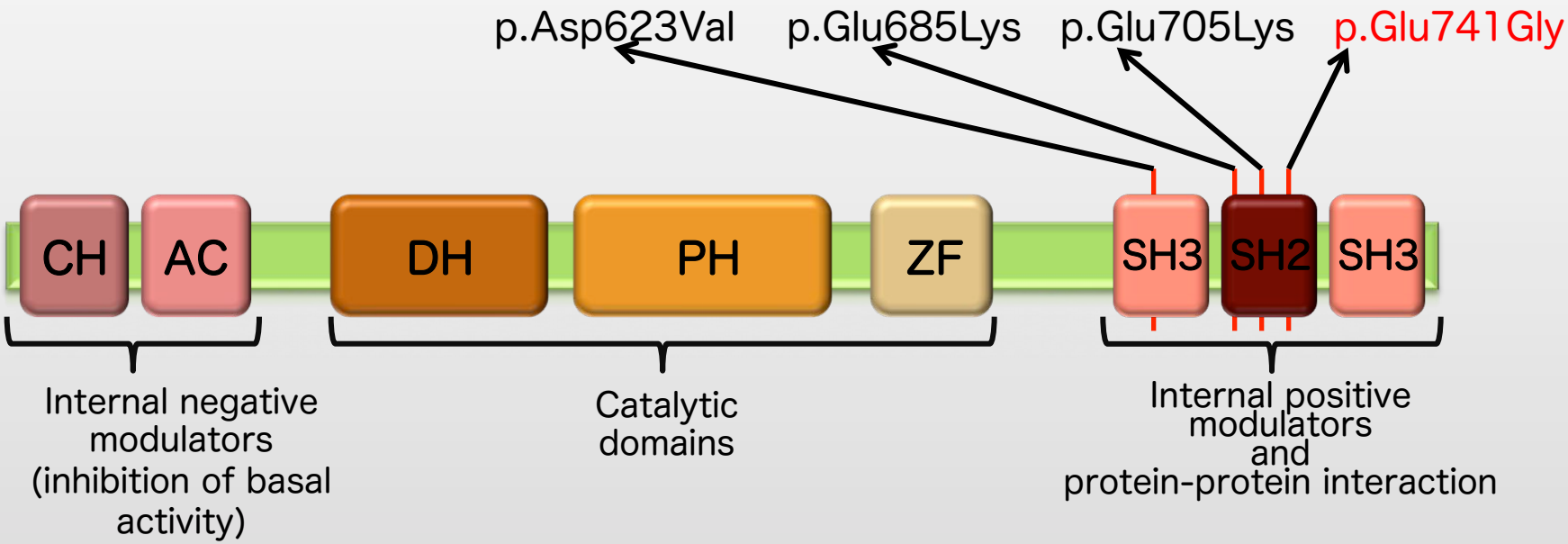
- Custom array resequencing method
- Focusing on exons only
  - Novel rare missense/nonsense variants
- Screening only cases (N=321)
- Follow up candidates in large case-control sample

Discovered variants



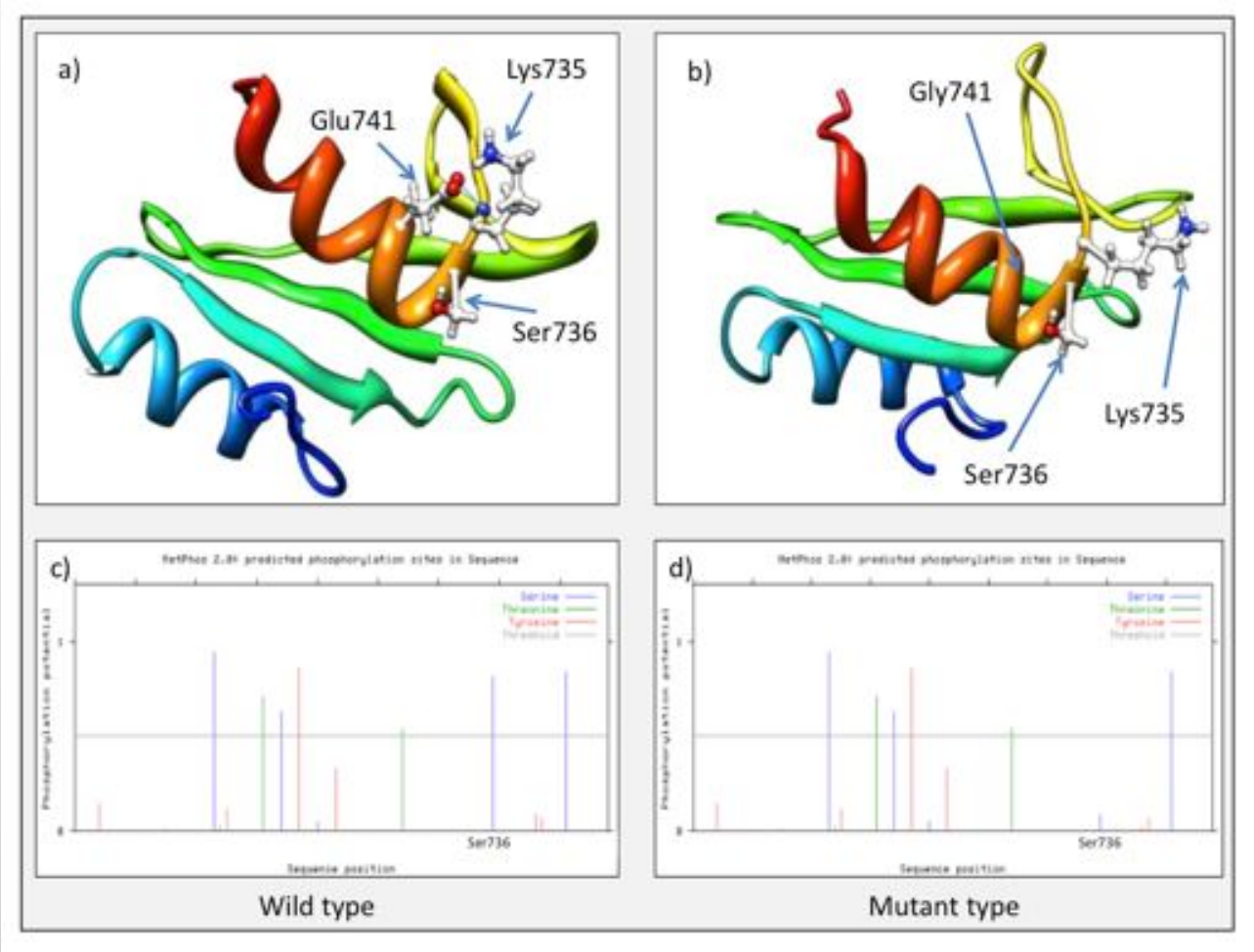
Species	Protein	p.Asp623Val	p.Glu685Lys	p.Glu705Lys	p.Glu741Gly	Pairwise Alignment Scores vs H.sapiens		
						Protein identity %	DNA Identity %	Substitution Rate
Homo sapiens	NP_006104.4	L Q A G D T V	L Q A E T E L	K E S G E Y A	L M E L V E Y	n/a	n/a	n/a
Canis lupus familiaris	XP_537047.2	I Q A G D T V	L Q A E T E L	K E S G E Y A	L M E L V E Y	97.1	92.9	0.075
Bos taurus	XP_615898.4	I Q A G D T V	L Q A E T E L	R E S G E Y A	L M E L V E Y	95.7	92.8	0.076
Mus musculus	NP_065251.2	I Q A G D T V	L Q A E T E L	K E S G E Y A	L M E L V E Y	95.4	90.5	0.101
Rattus norvegicus	XP_227600.4	I Q A G D T V	L Q A E T E L	K E S G E Y A	L M E L V E Y	94.8	89.6	0.112
Gallus gallus	NP_996745.1	I Q I G D T I	L Q A E S E L	K E S G E Y A	L M E L V D Y	86.3	81.3	0.215
Danio rerio	XP_687553.3	A Q I G D V I	H H A E S E L	R E S R E Y A	V L G L V E Y	71.2	68.2	0.414
Caenorhabditis elegans	NP_001041223.1	F A K G D R I	A K A E S T L	K N R K Q T A	T V E L V Q Y	35.3	46.0	0.953

Association analysis



Chr	Variant	Physical position[1]	Protein domain	M[2]	JMut (minor allele count)	m[2]	JPN_GWAS (MAF) N=1100				Rep_JPN (MAF) N=3000				Meta analysis N=4200	
							Cases[3]	Control[3]	P_{allele}	OR[4]	Cases[3]	Control[3]	P_{allele}	OR[4]	P_{CMH}[5]	OR[4]
1	p.Asp623Val	107,986,810	N-SH3	A	2	T	0.0006964	0.0008993	0.8561	0.7742	0.0003344	0	0.3171	NA	0.6649	1.662
1	p.Glu685Lys	107,947,271	SH2	G	1	A	0.0006974	0.001821	0.4151	0.3824	0.0003336	0	0.3168	NA	0.8415	0.8246
1	p.Glu705Lys	107,947,211	SH2	G	3	A	0.0007022	0.0009074	0.8557	0.7737	0.0006658	0.0003311	0.5605	2.011	0.7354	1.355
1	p.Glu741Gly	107,940,485	SH2	A	7	G	0.004972	0.01087	0.09038	0.4547	0.0074480	0.0117400	0.0897	0.6314	0.02065	0.5821

In-silico modeling of SH2 domain (VAV3)

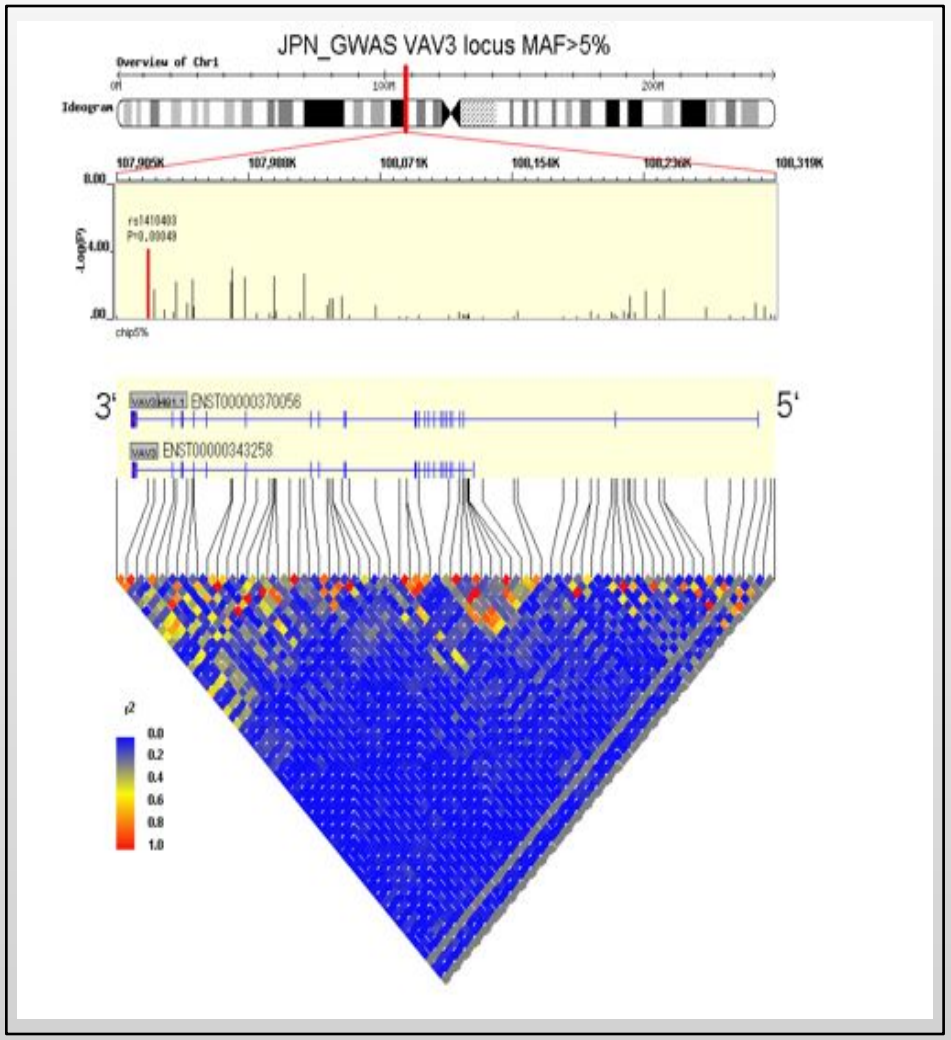


COMMON VARIANTS

Voxel based morphometry

- We followed up rs1410403 (SNP associated with schizophrenia in JGWAS
- $P_{CMH}=9.3\times10^{-4}$, odds ratio=0.86
- case control sample was comprised of 100 patients with schizophrenia (38.3 ± 13.0) and 264 healthy controls (36.7 ± 11.9)
- All magnetic resonance imaging was performed on a 1.5T GE Sigma EXCITE system

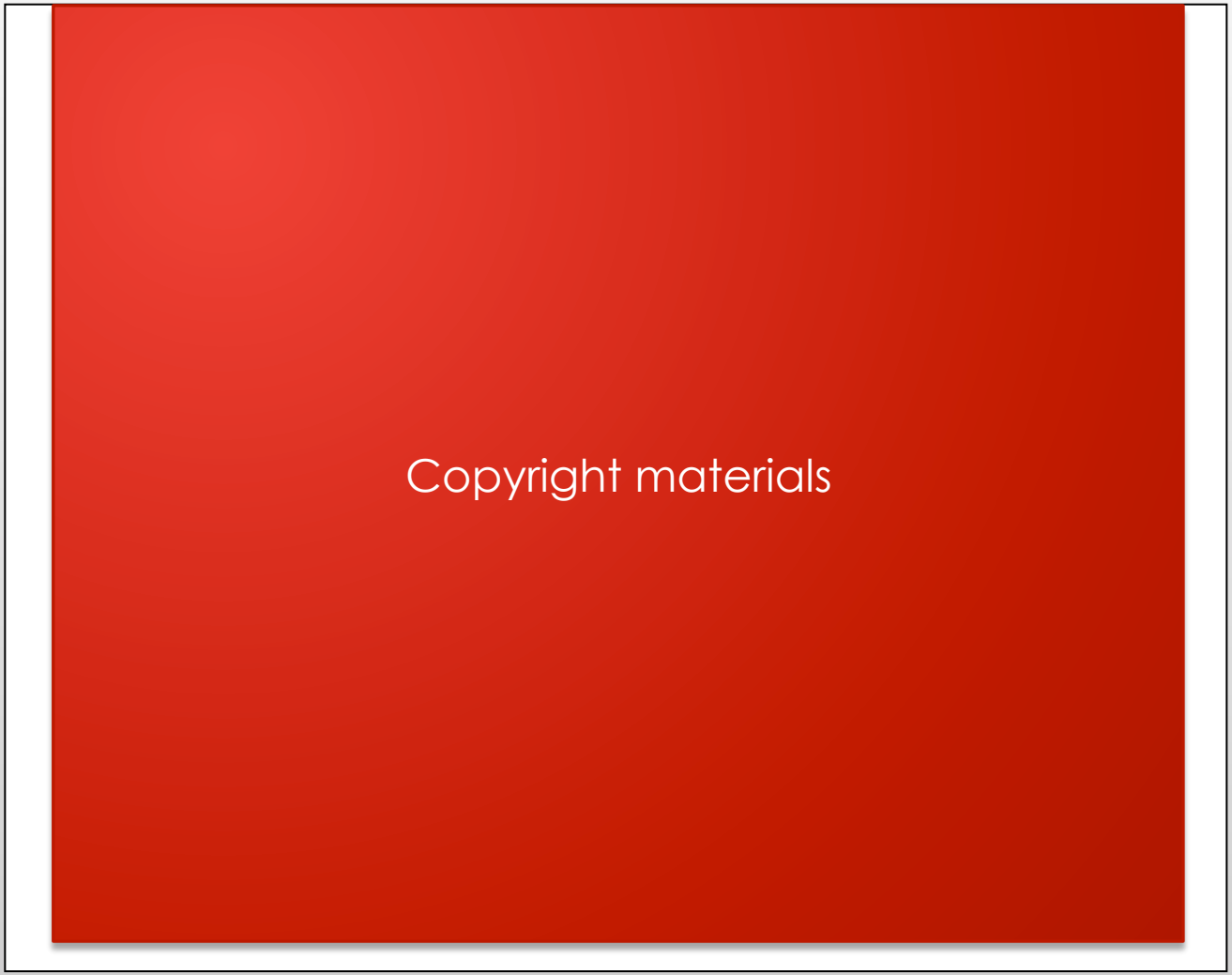
Association signal in JGWAS



Gene centered data:

- Location: 1p13.3
- Size: 400 kbps
- 2 isoforms has been reported

Effect on the brain morphology



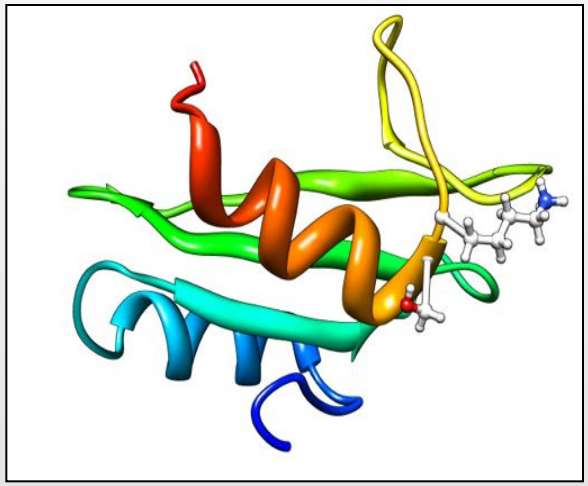
- Rs1410403 (A/G) was followed up
- Minor allele frequency (G) in *CONTROLS* of this SNP was 37%
- Minor allele frequency (G) in *CASES* of this SNP was 32%
- OR=0.86

Summary of findings

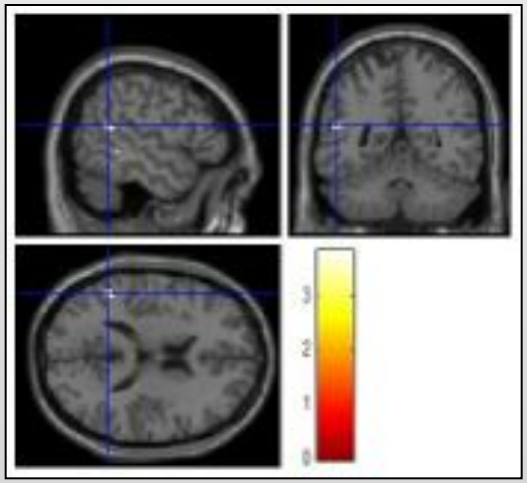
- p.Glu741Gly might be responsible for genetic susceptibility to schizophrenia-biological consequences are unknown (OR=0.52)

- Rs1410403 might influence volume of gray matter in schizophrenic patients (OR=0.86)

VAV3 effects of associated variants



- Change in protein structure
- Biological consequences unknown



- Protein structure unchanged
- Variation in gray matter volume

Odds ratio: Low — High

Frequency: Low — High

Challenges and future directions

- Consideration of the effect of environmental factors such as maternal infection or drug use
- Consideration of epigenetic mechanism
- Use of high-throughput whole genome sequencing
  - Has potential to detect virtually all SNPs/SNVs
  - Will provide comprehensive information of individual at DNA-single base pair level
  - Very costly

Take Home Messages

- Strong genetic basis of SZ proven from age-old family studies to the ultra modern GWAS
- Specific genes and loci are not definitely established (i.e. lack of consistent replication)
- Problem arising from multiple factors
  - Lack of operationalized phenotypes
  - Presence of large number of risk variants with relatively small effect size
  - Cost, manpower and expertise inadequacy

Thank you for your attention!
Any questions or comments?