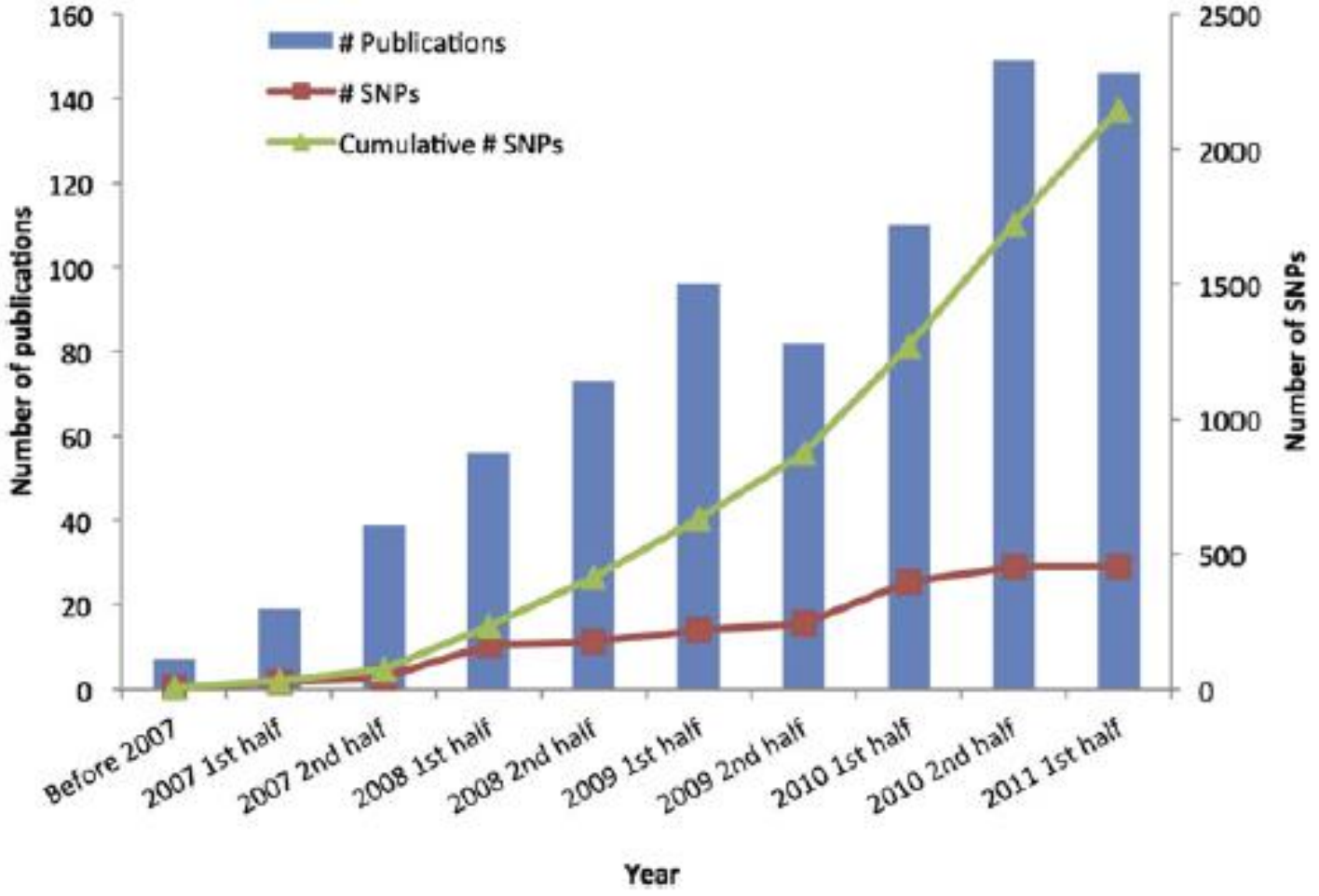
The 15 Korea-Japan-China
Bioinformatics Symposium 2017

Allele-specific functional genomics in post-GWAS era

Hirofumi Nakaoka
Division of Human Genetics,
National Institute of Genetics, Japan

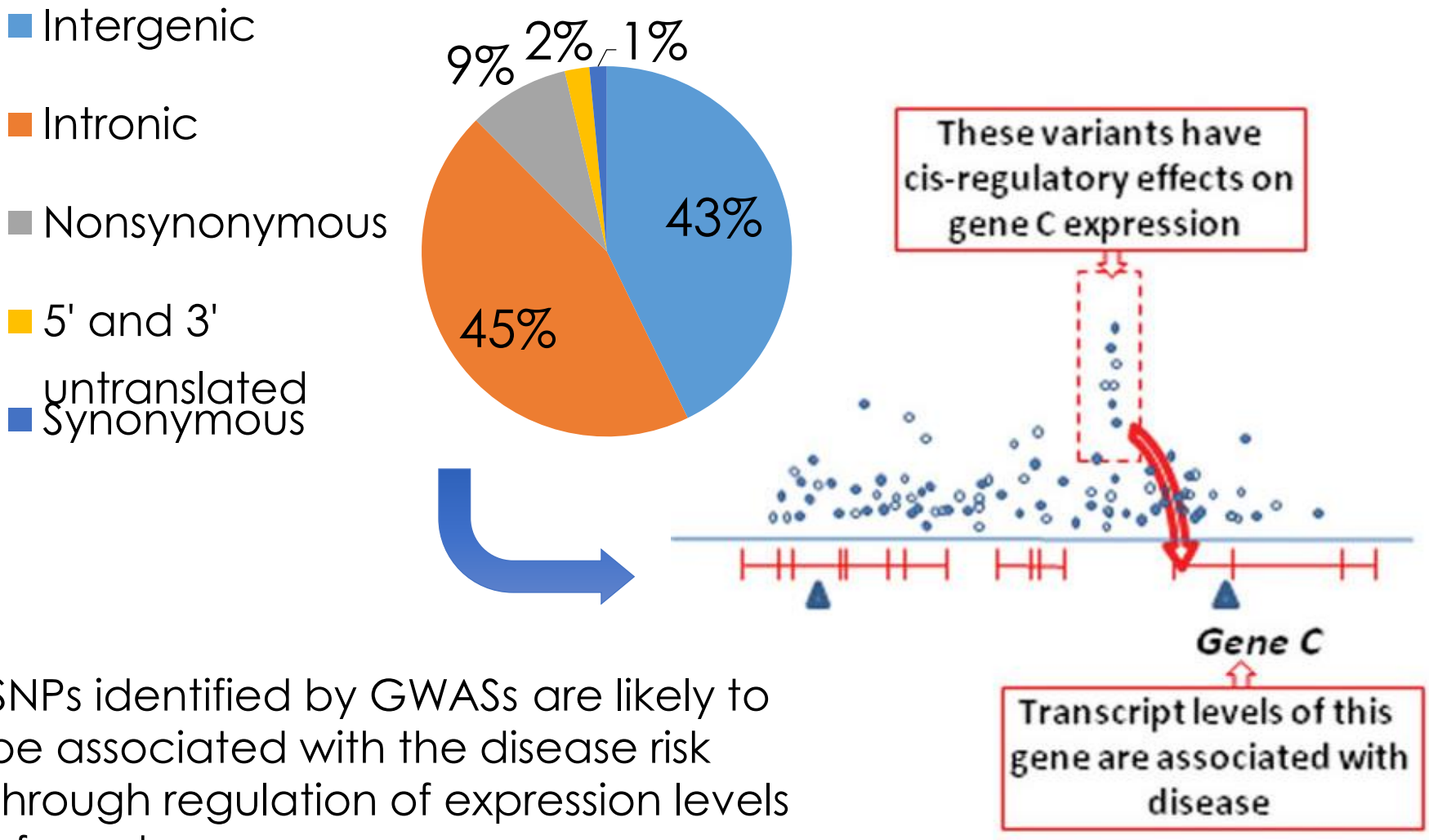
Findings from genome-wide association studies (GWASs)

~7,000 trait/disease-associated SNPs



Visscher et al. 2012 Am J Hum Genet

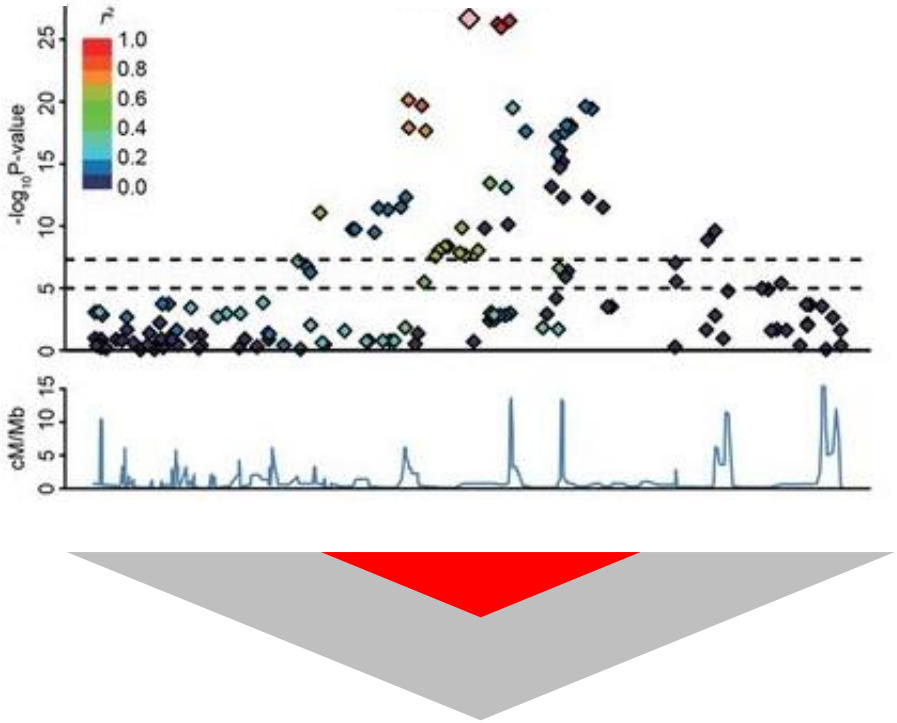
Annotation of SNPs identified by GWASs



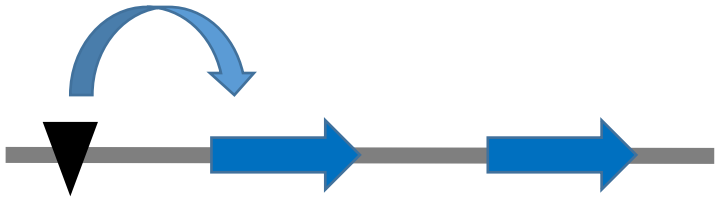
SNPs identified by GWASs are likely to be associated with the disease risk through regulation of expression levels of nearby genes

Hindrof et al. 2009 PNAS

How GWAS SNPs exert the effects on the risk of diseases are largely unknown

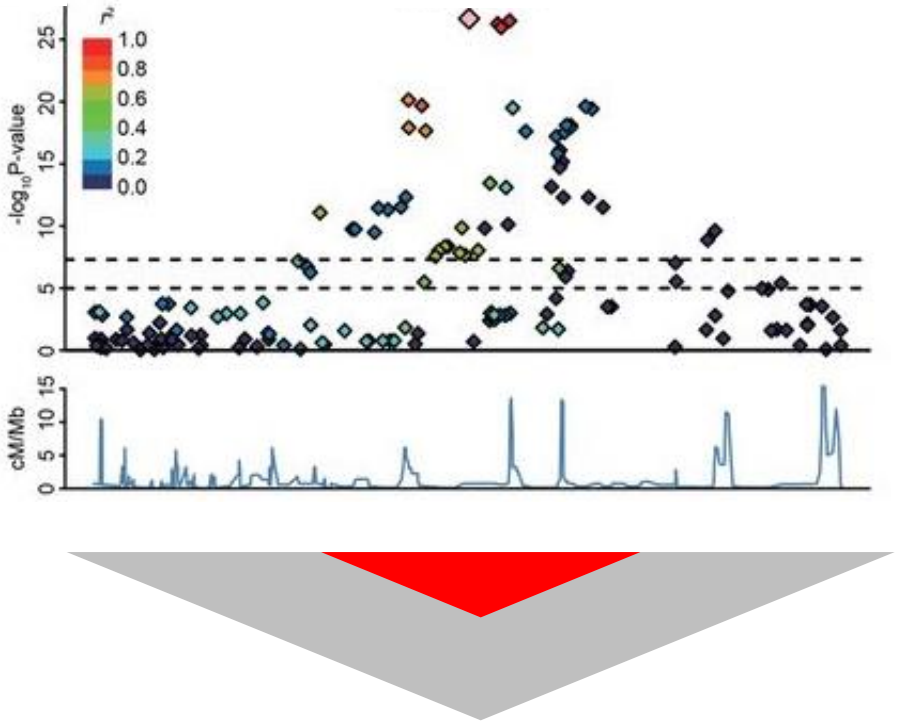


SNPs identified by GWASs are markers of true causal variants (linkage disequilibrium)

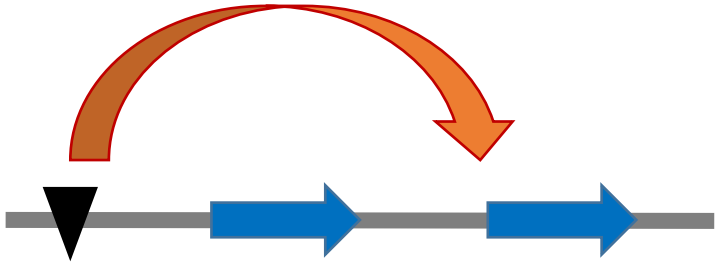


Genes closest to GWAS SNPs have been reported as "susceptibility" genes

How GWAS SNPs exert the effects on the risk of diseases are largely unknown

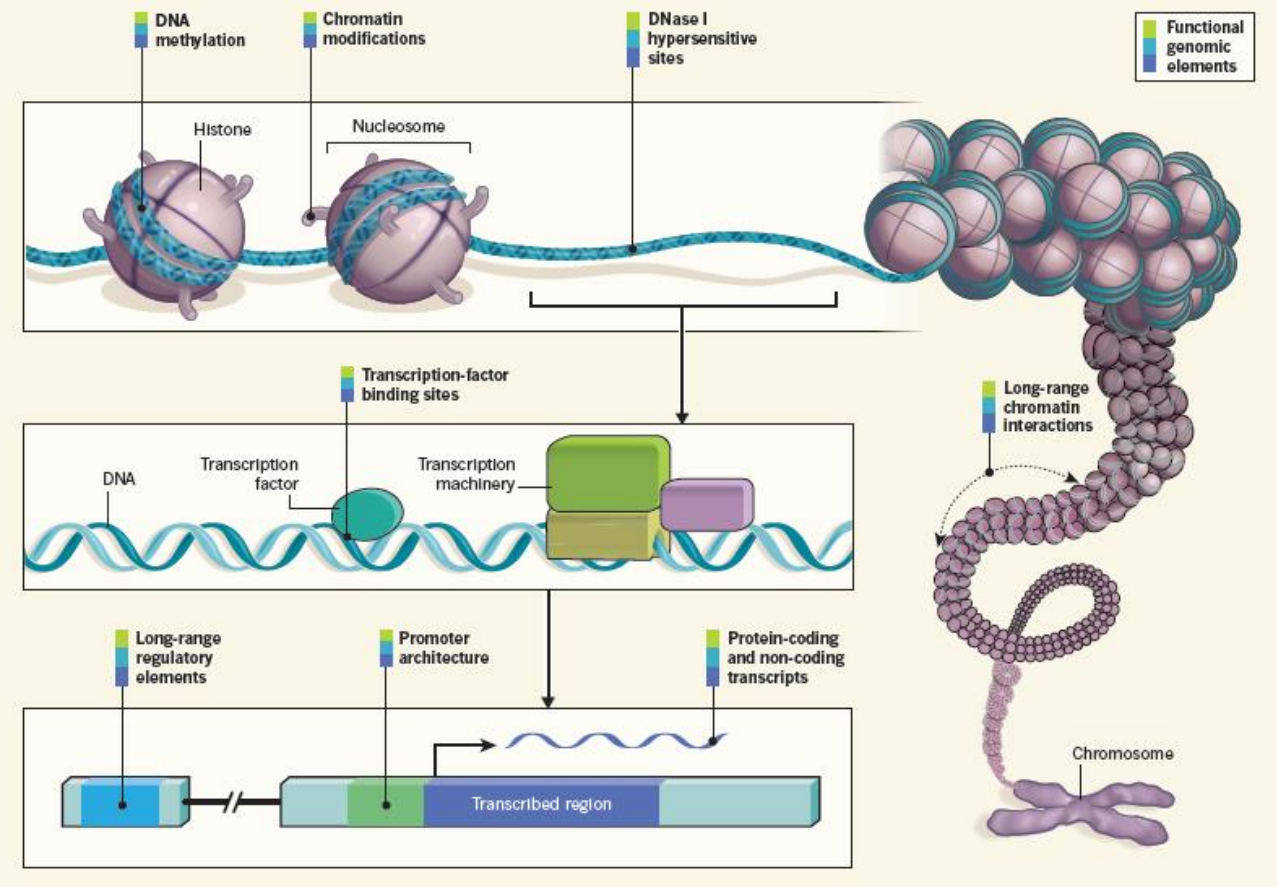


SNPs identified by GWASs are markers of true causal variants (linkage disequilibrium)



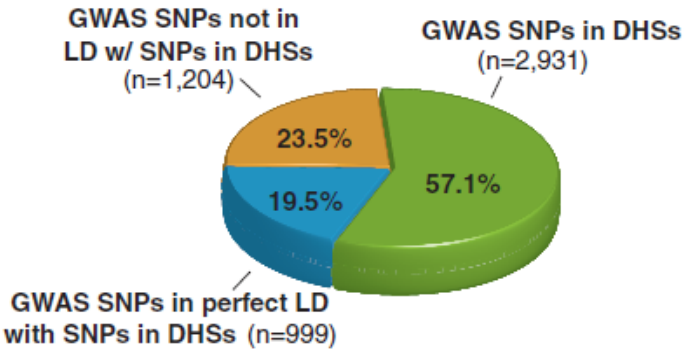
Possibility of regulation of distant genes through chromatin interactions

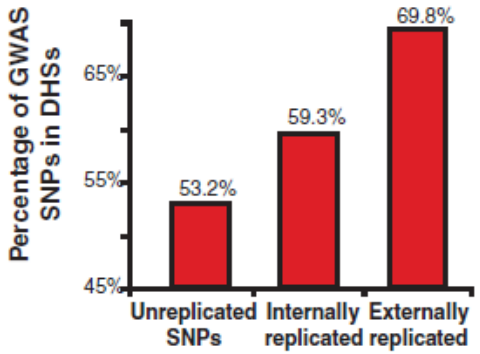
Functional elements across the human genome explored by ENCODE project

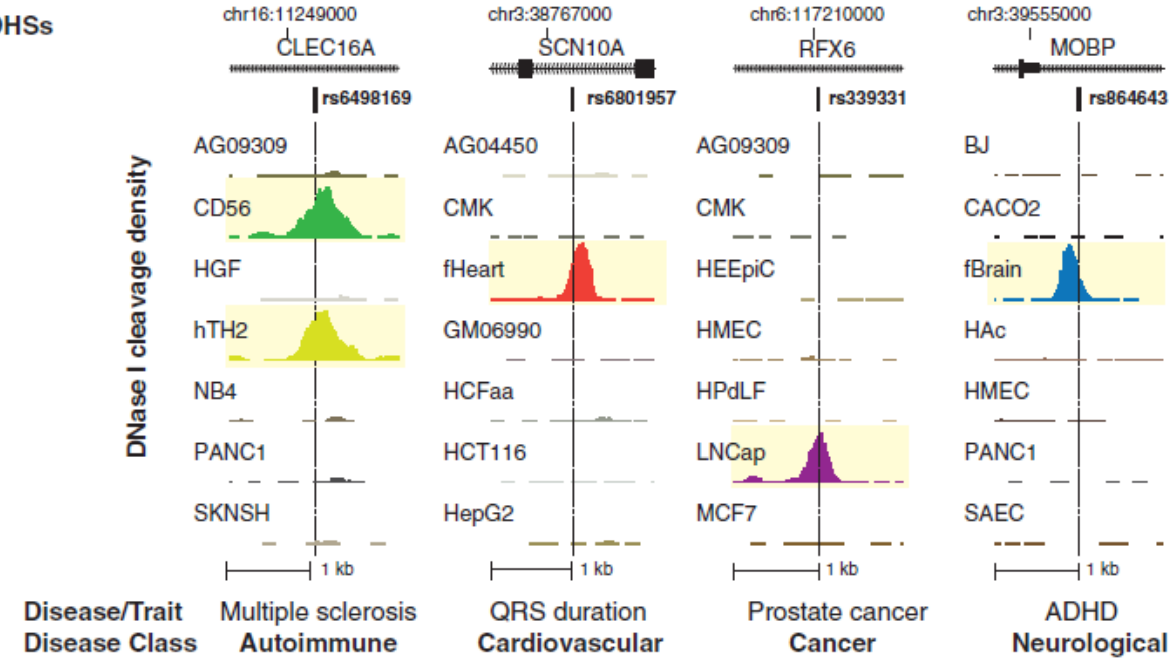


Ecker et al. Nature 2012

Majority of SNPs identified by GWASs are located on DNase I hypersensitive sites (DHSs), signatures of open chromatin status



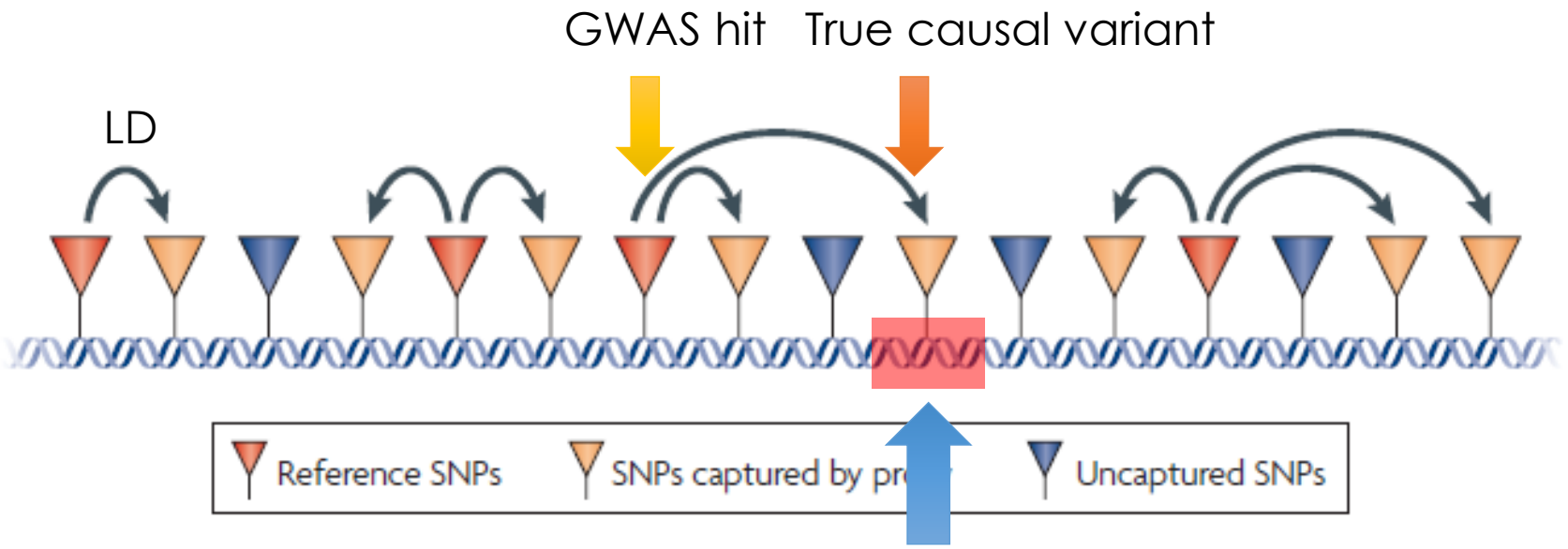




DHSs in cell types relevant to the diseases

Maurano et al. Science 2012

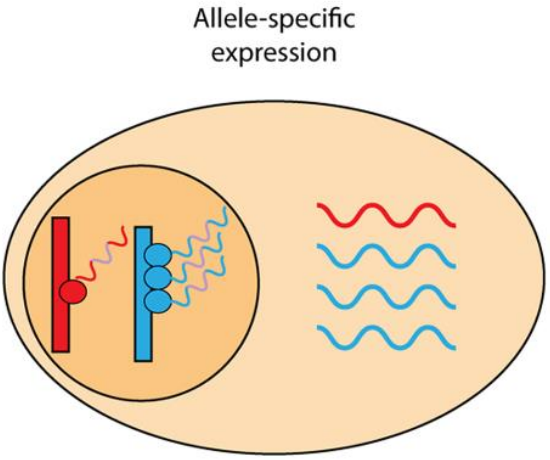
Fine-mapping by combining a comprehensive set of genetic variants and information about regulatory elements



Regulatory element (DHS) in a cell type relevant to the disease

Allele specific functional genomics

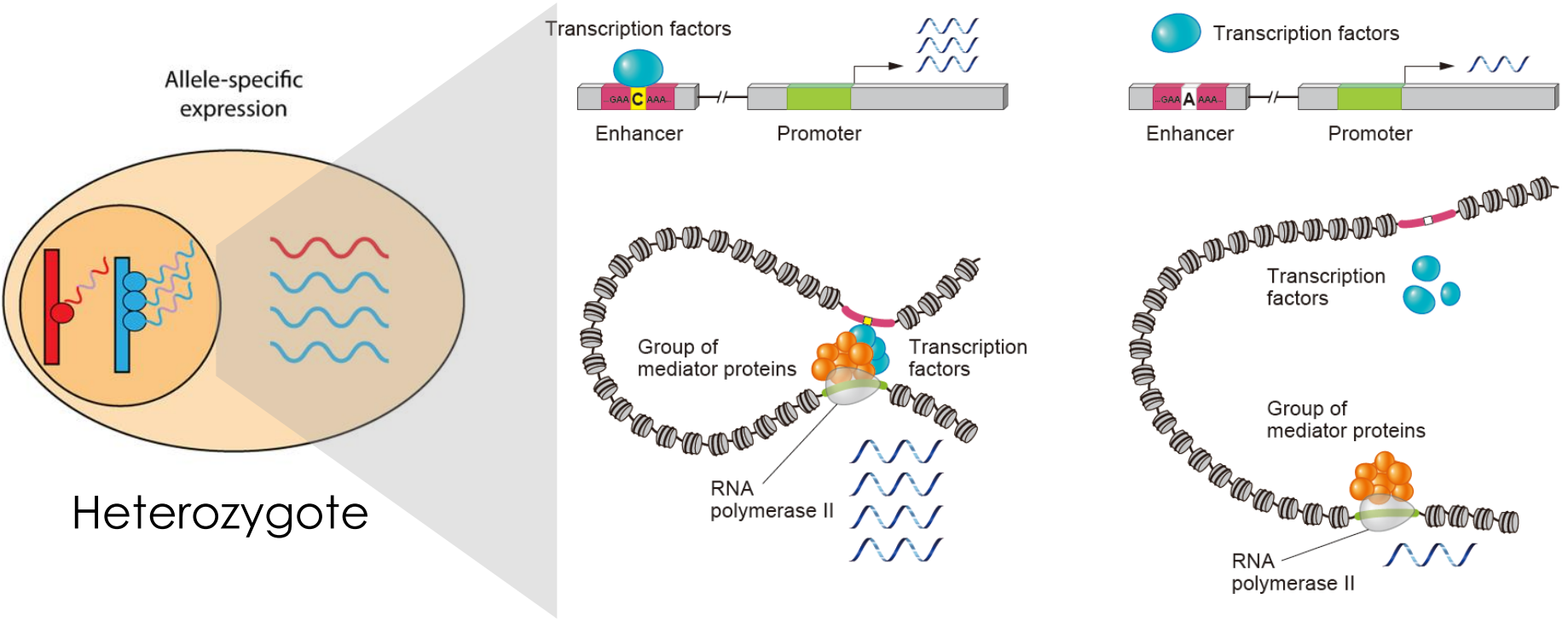
Use of heterozygous samples allow us to directory compare functional activities of two alleles within the same cellular environment



Heterozygote

Allele specific functional genomics

Use of heterozygous samples allow us to directory compare functional activities of two alleles within the same cellular environment



Allelic differences in distinct stages of transcriptional regulation:

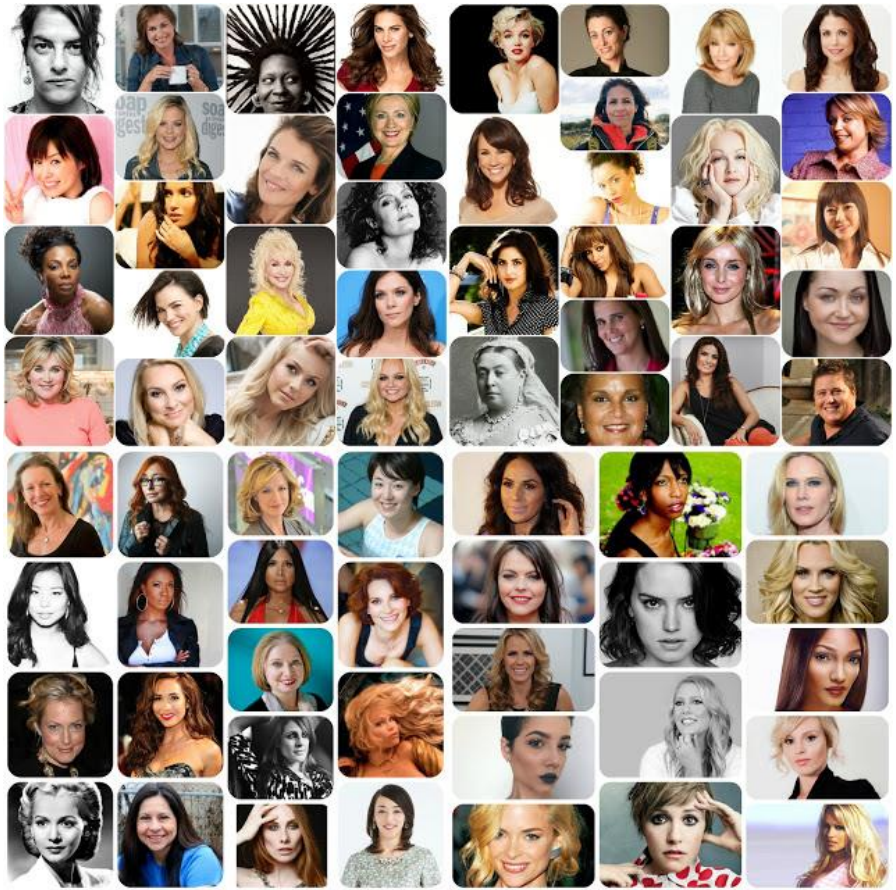
i) Binding affinities of TF

ii) Strength of chromatin interactions

iii) Gene expression levels

Endometriosis
(자궁 내막증,子宫内膜异位症,子宫内膜症)

- Estrogen-dependent inflammatory disease
- 10% of women of reproductive age
- Main symptoms are pelvic pain and heavy or irregular menstrual bleeding, which associated with dysmenorrhea and infertility.

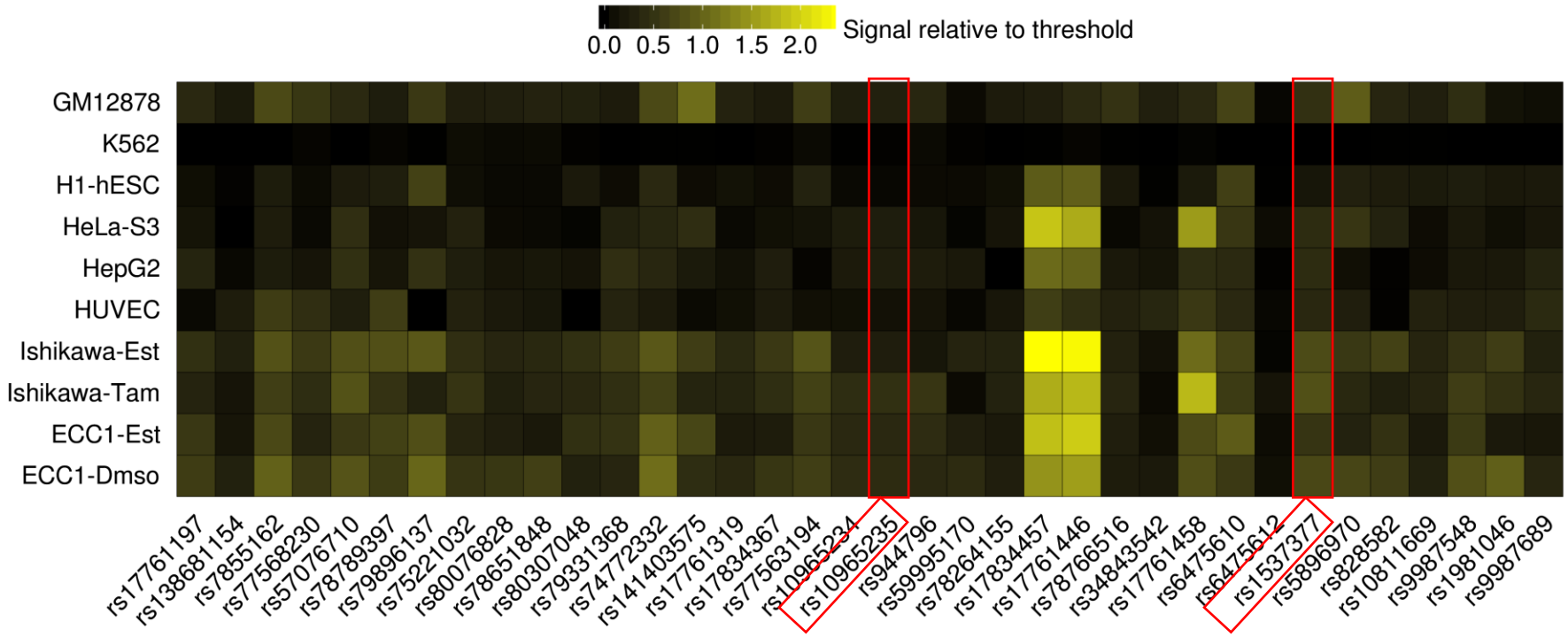


GWASs have identified several risk loci (1p36, 2p14 , 2p25, 2q23, 6p22, 7p15, 9p21, and 12q22)

Aya Matsuura

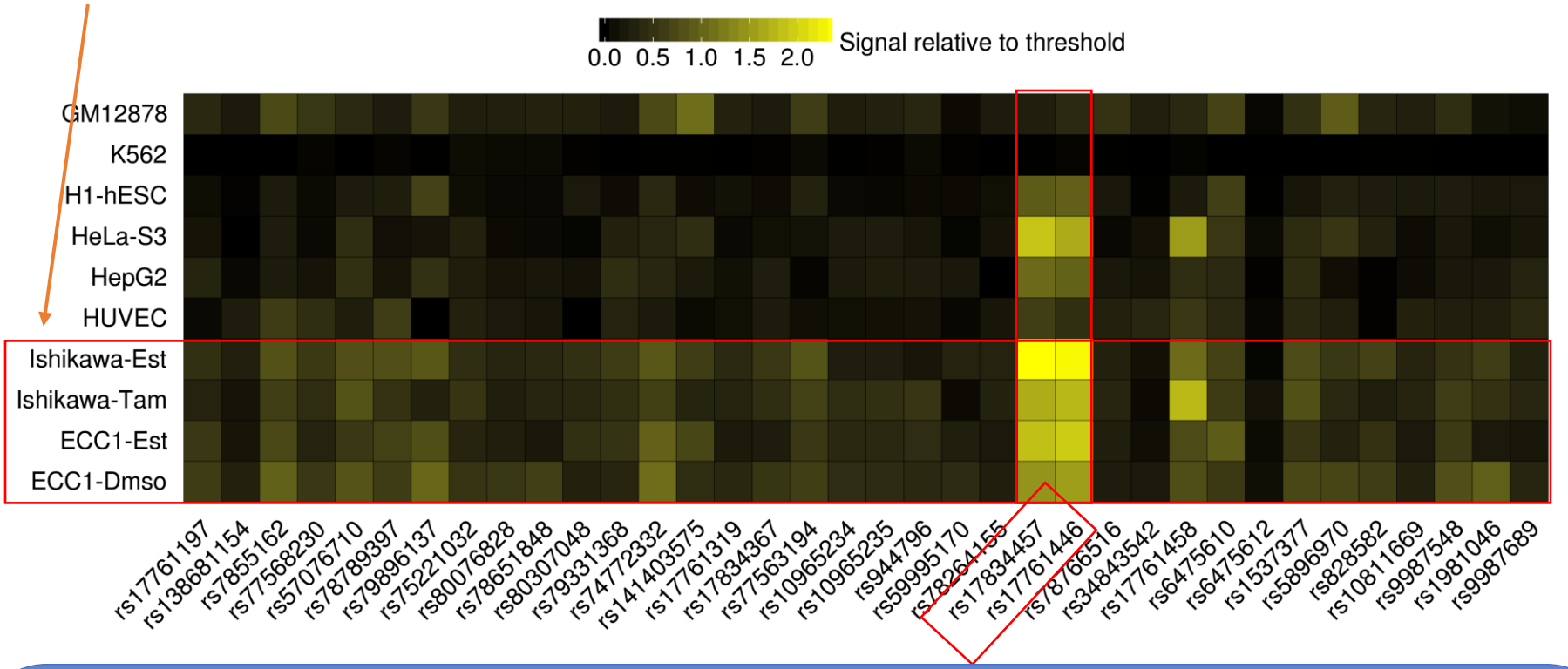
Nanako Fujisaki

DNase-seq signals at SNPs that are in strong LD with the original GWAS SNPs



SNPs identified by GWAS of endometriosis are not located on DHS.
(= less likely to be regulatory SNPs)

Endometrial cell lines



rs17761446 and rs17834457 located in an intron of *ANRIL*, and 123 kb apart from transcription start site of *ANRIL*

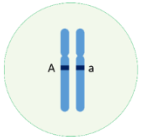
⇒ The DHS is a distal enhancer contacting with promoter of *ANRIL* through chromatin looping interaction?

Enhancer activity differs according to the SNP allele?

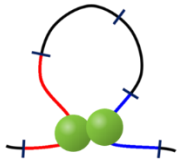
Detection of allele-specific chromatin interactions: AS3C-seq

A. 3C library preparation

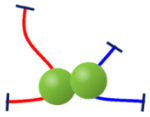
1.Cell lines heterozygous for target SNP



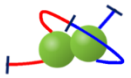
2. Crosslinking and chromatin isolation



3. Digestion by restriction enzyme



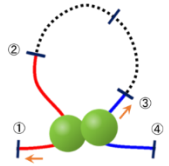
4. Proximity ligation

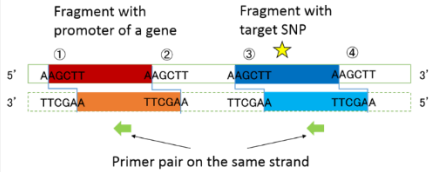


5. Reverse cross-links

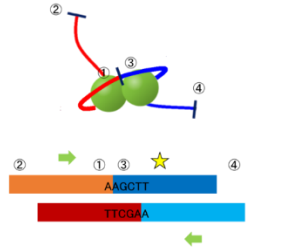
B. PCR amplification followed by sequencing

6. Unidirectional primer design





7. Detection of interaction by unidirectional primer design



Specific PCR amplification of restriction fragments ligated between ① and ③ (exclusion of self-ligation products and uncut linear genome sequences)

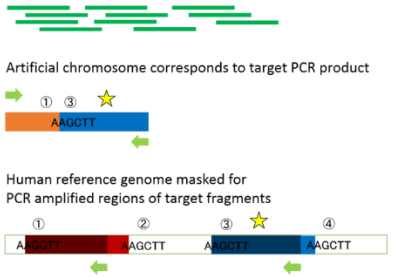
8. PCR product of interacting fragments

9. Fragmentation and adapter ligation by Nextera system

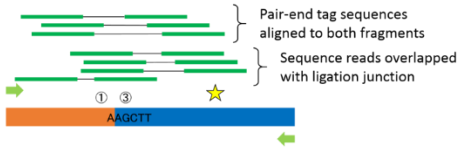
10. Deep sequencing on MiSeq

C. Bioinformatics analysis

11. Align NGS reads to modified genome with "artificial chromosome" and "masked reference genome"



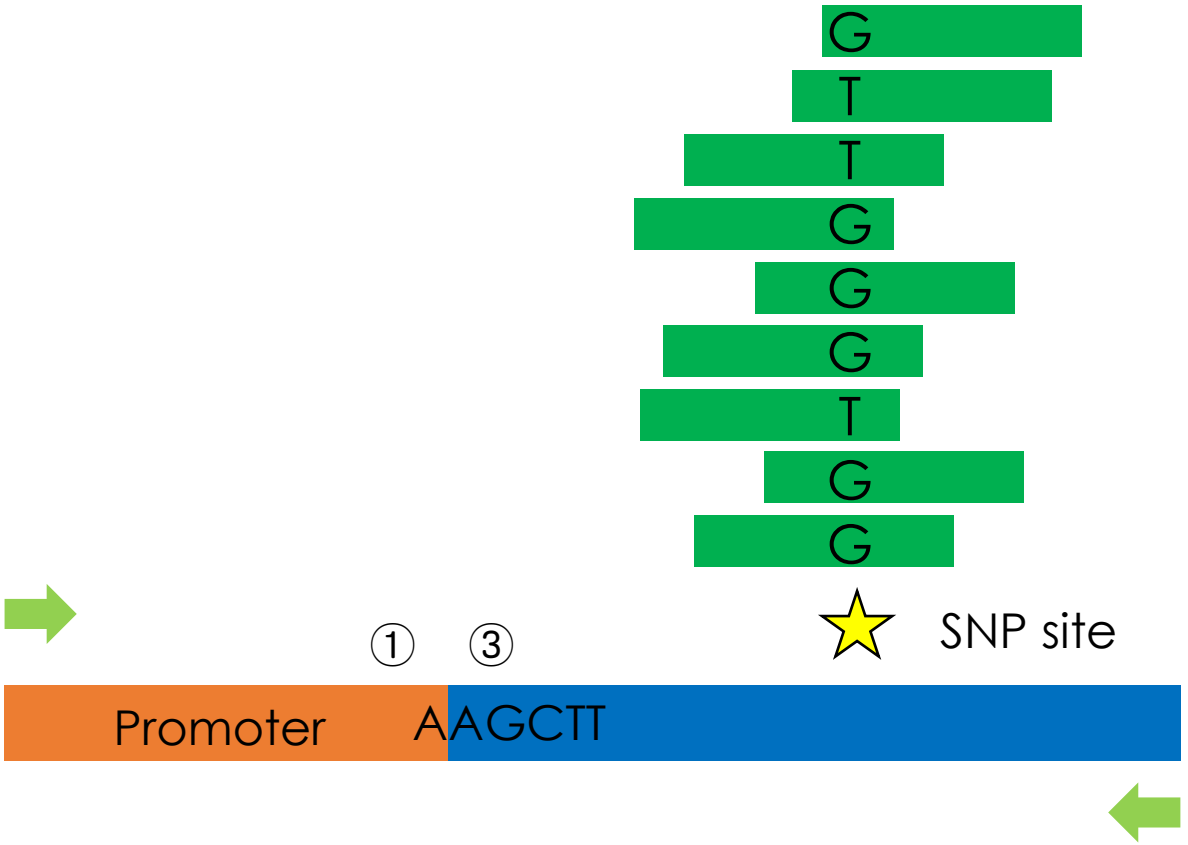
12. Detection of sequence reads supporting the presence of chromatin interaction



Chromosome conformation capture (3C) assay detects two genomic loci interacting within the three-dimensional (3D) nuclear space.

Nakaoka et al. PLOS Genet 2016

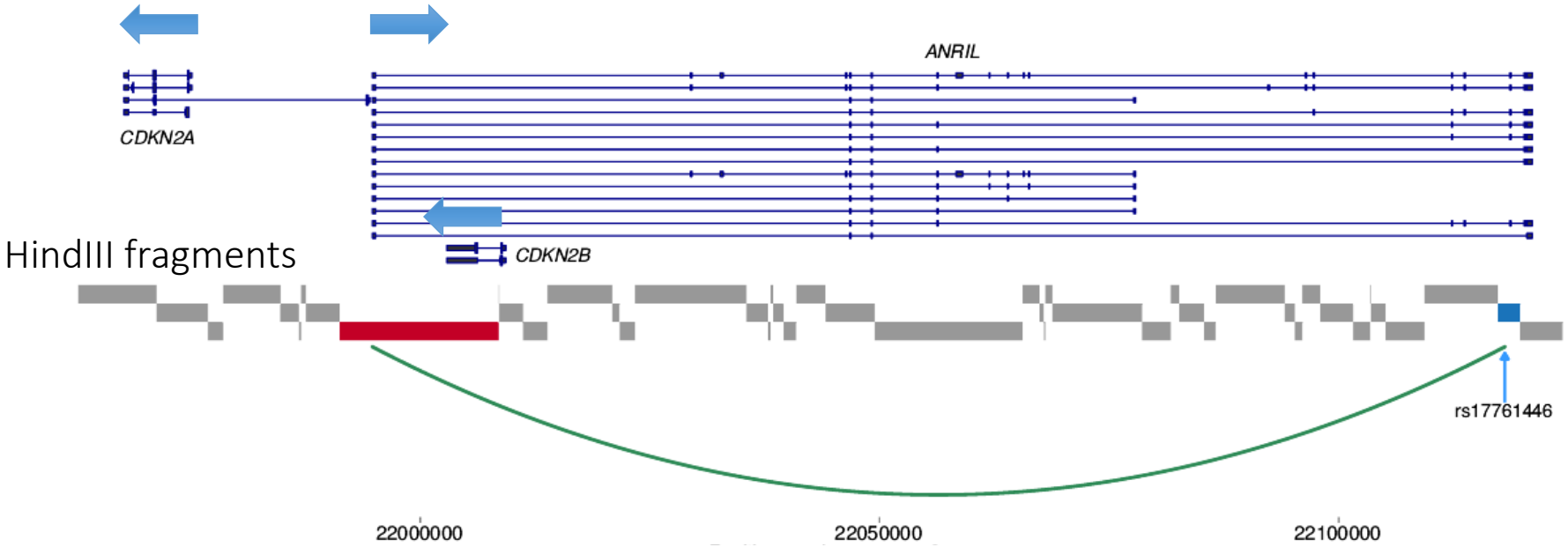
AS3C-seq allows to evaluate relative strength of chromatin interaction between two SNP alleles

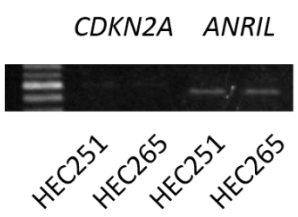


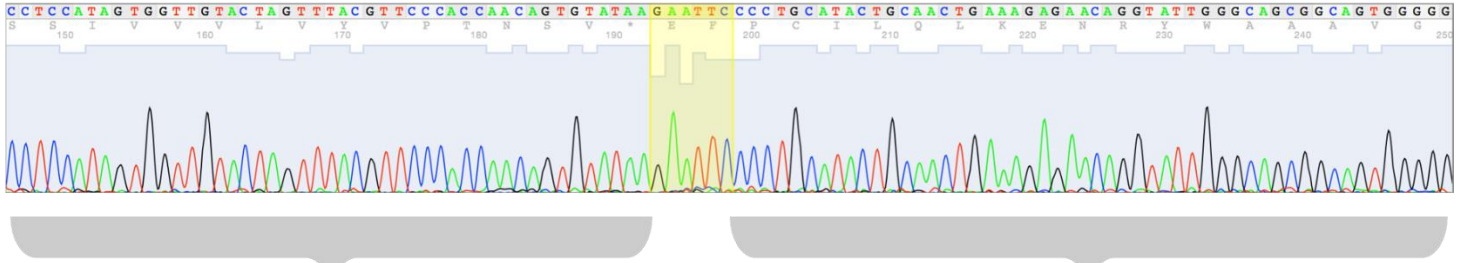
Deviation from 50:50 in the allele-specific read counts of next generation sequence

⇒Difference in the strength of chromatin interaction between the two alleles

Chromatin loops formed between the candidate causal SNP and the fragments around 9p21 genes







Chr9:22121975-22122021
Fragment with rs17761446

Chr9:21998110-21998161
Fragment with ANRIL promoter

Allele specific chromatin interaction between promoter of *ANRIL* and rs17761446

Strength of chromatin interaction with *ANRIL* promoter of "G" allele was two times greater than that of "T" allele.

Ratio between two alleles (G and T) was highly significantly deviated from 50:50 ($P<10^{-16}$; binomial test)

Transcription factor binding motif

Generating sequences for SNV

Reference sequence

TATGCAGAGTGATGCGTGCTGCG

Alternated sequence

TATGCAGAGTGGTGCGTGCTGCG

30bp 30bp

Query to manually curated PWM database, specialized in homo sapience

HOCOMOCO

Kulakovskiy et al. 2012
Nucleic Acids Res

Detection of TF bound to the target sites

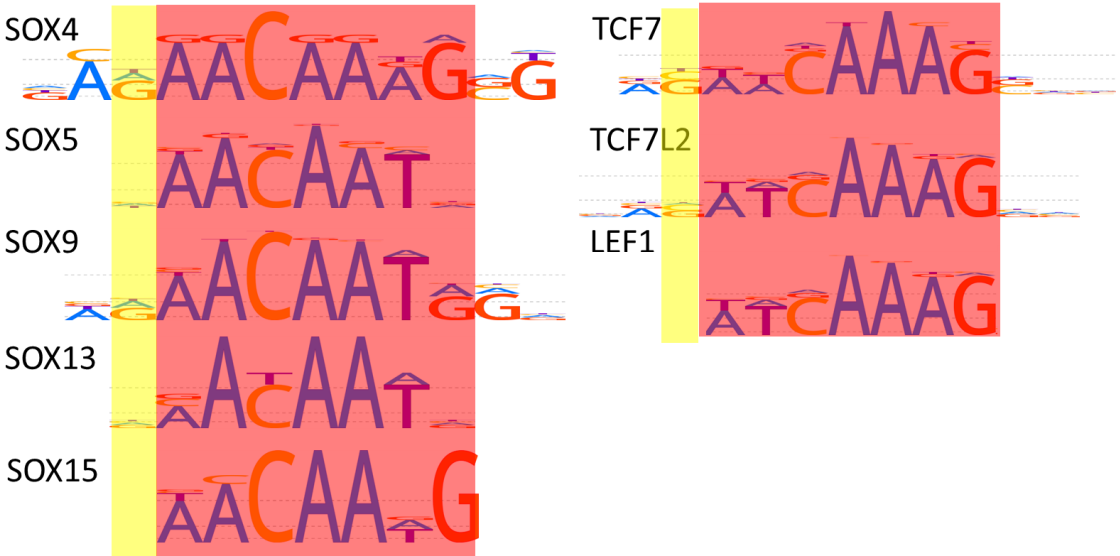
Motifsuite:
Claeys et al. Bioinformatics 2012

SNP site

Weak chromatin interaction

>hg19|rs17761446
ATGATCACATGCCTTCTTGGCAGATGTTTCTAACAAAGGAGAGAGTTGCCCAGGGTGGGGC
>hg19|rs17761446_alt
ATGATCACATGCCTTCTTGGCAGATGTTTCGAACAAAGGAGAGAGTTGCCCAGGGTGGGGC

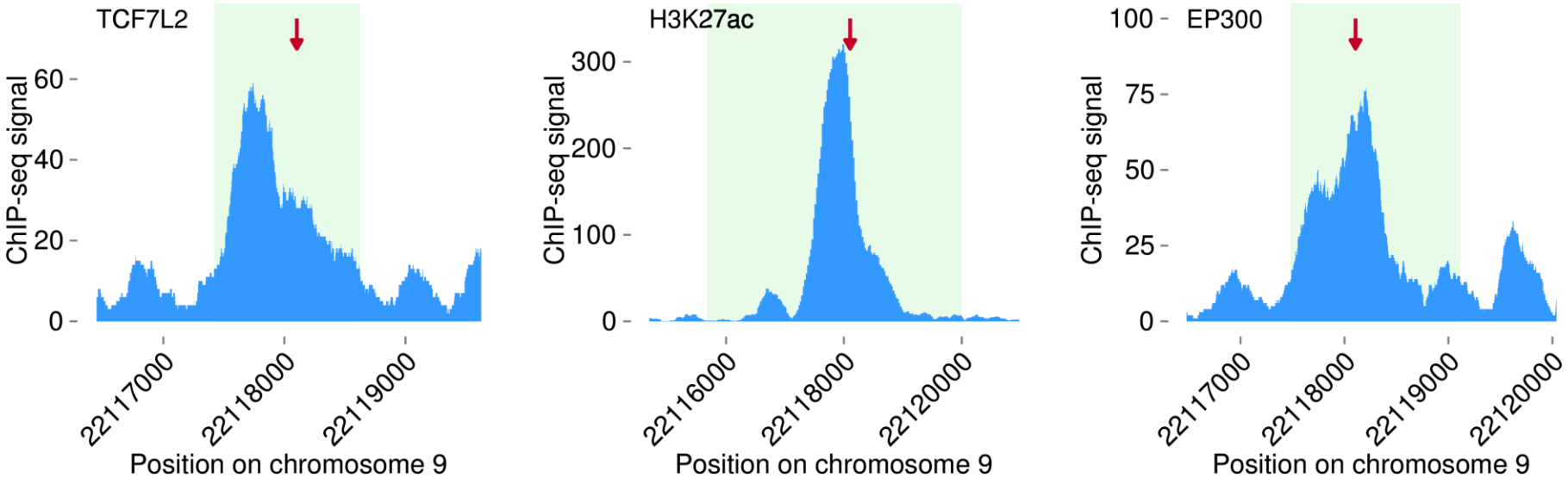
Strong chromatin interaction

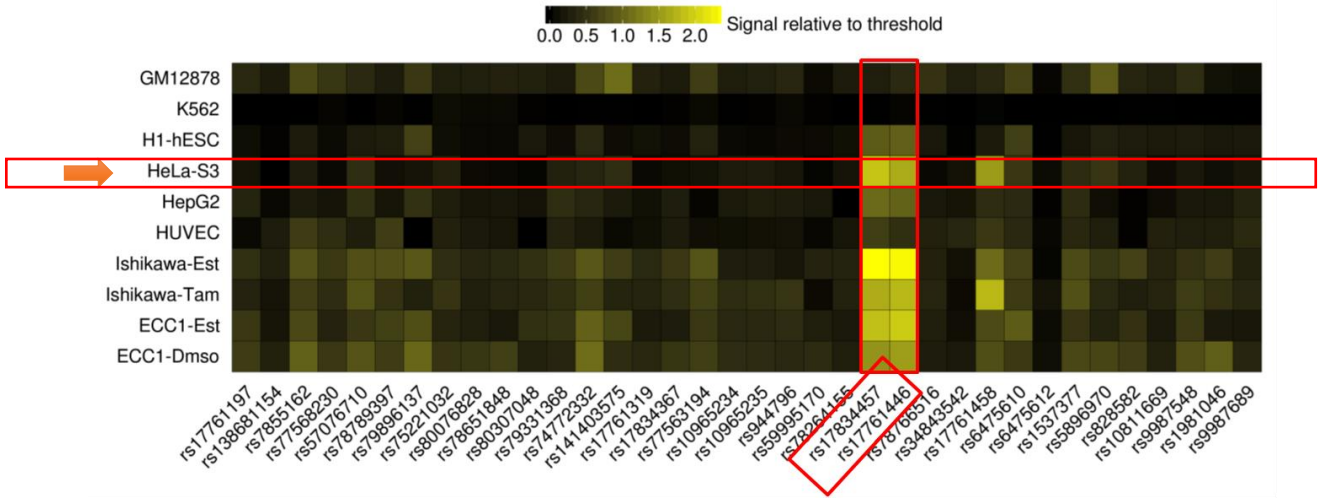


rs17761446 overlap with HMG of transcription factors:
Sox and TCF/LEF families

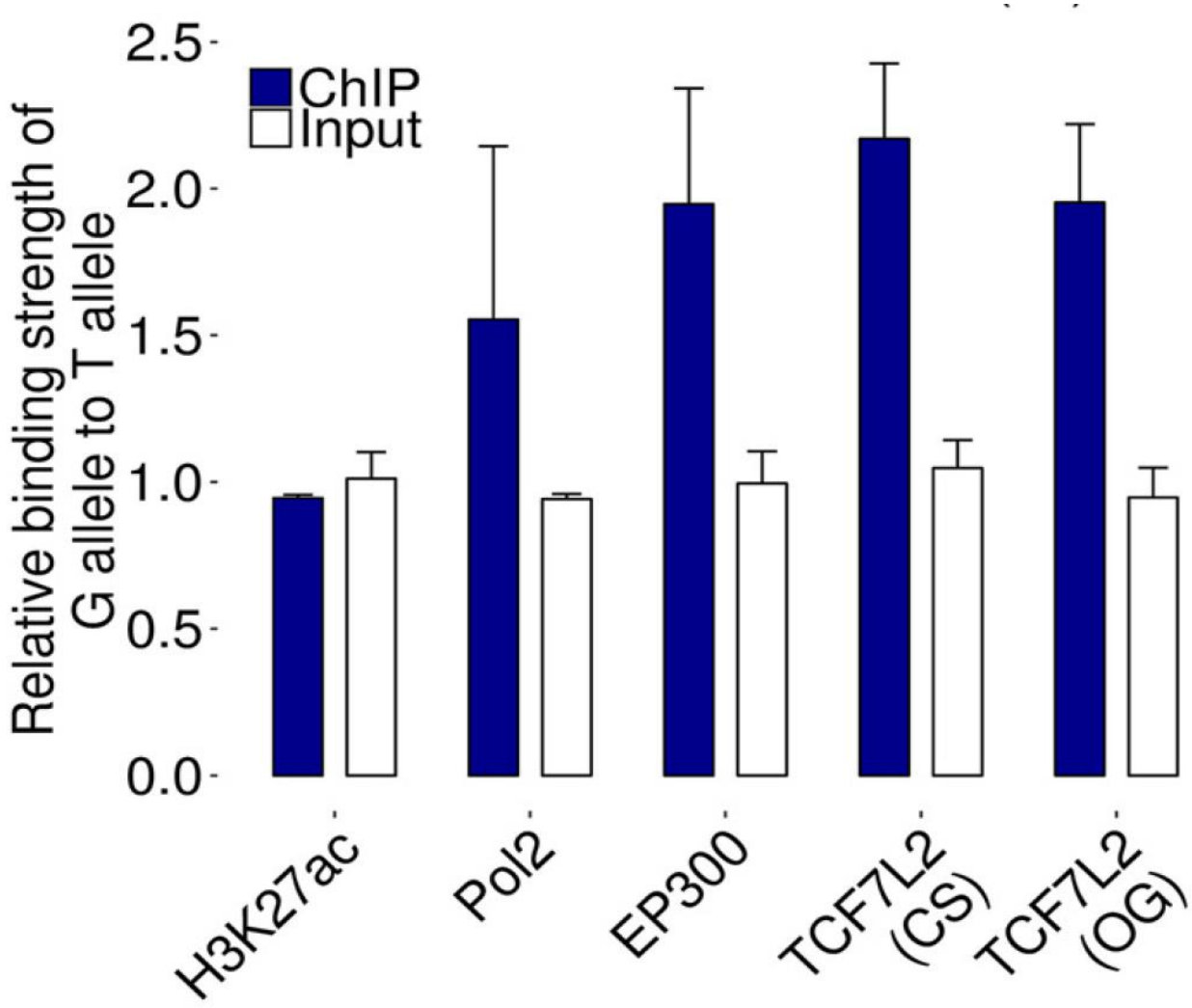
(rs17834457 did not overlap with any TF)

ENCODE ChIP-seq analyses





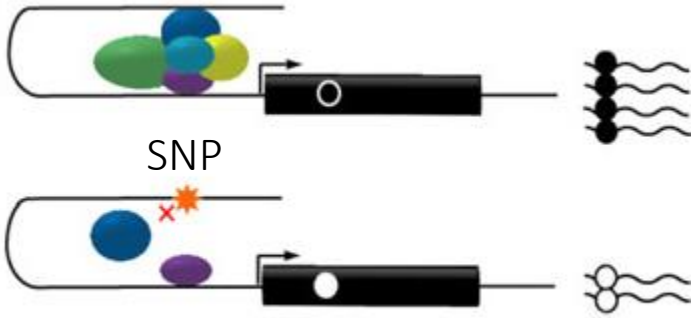
Allele specific binding of TCF7L2 to the candidate SNP site



Endometrial carcinoma cell line (HEC251)

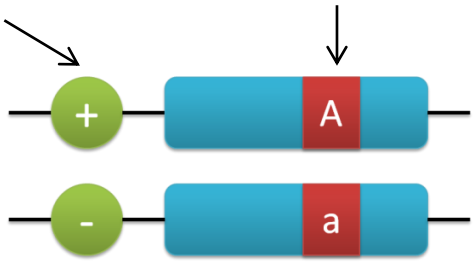
Allele specific expression (ASE) analysis

Cis-acting expression QTL



Candidate SNP SNP within transcribed region

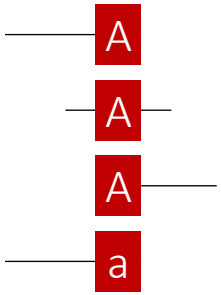
Haplotype phasing



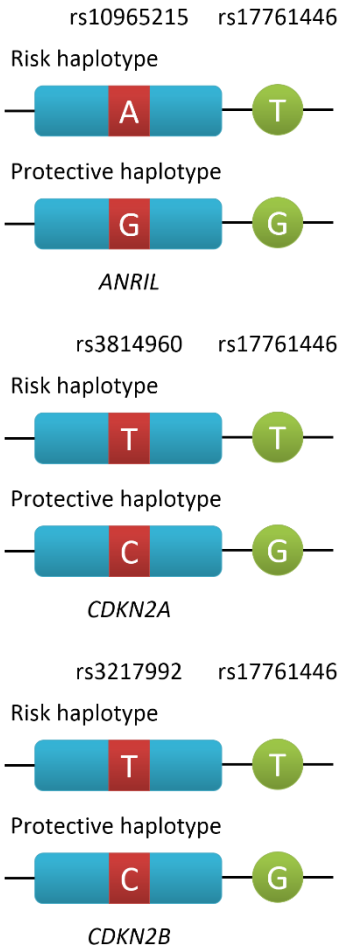
ANRIL

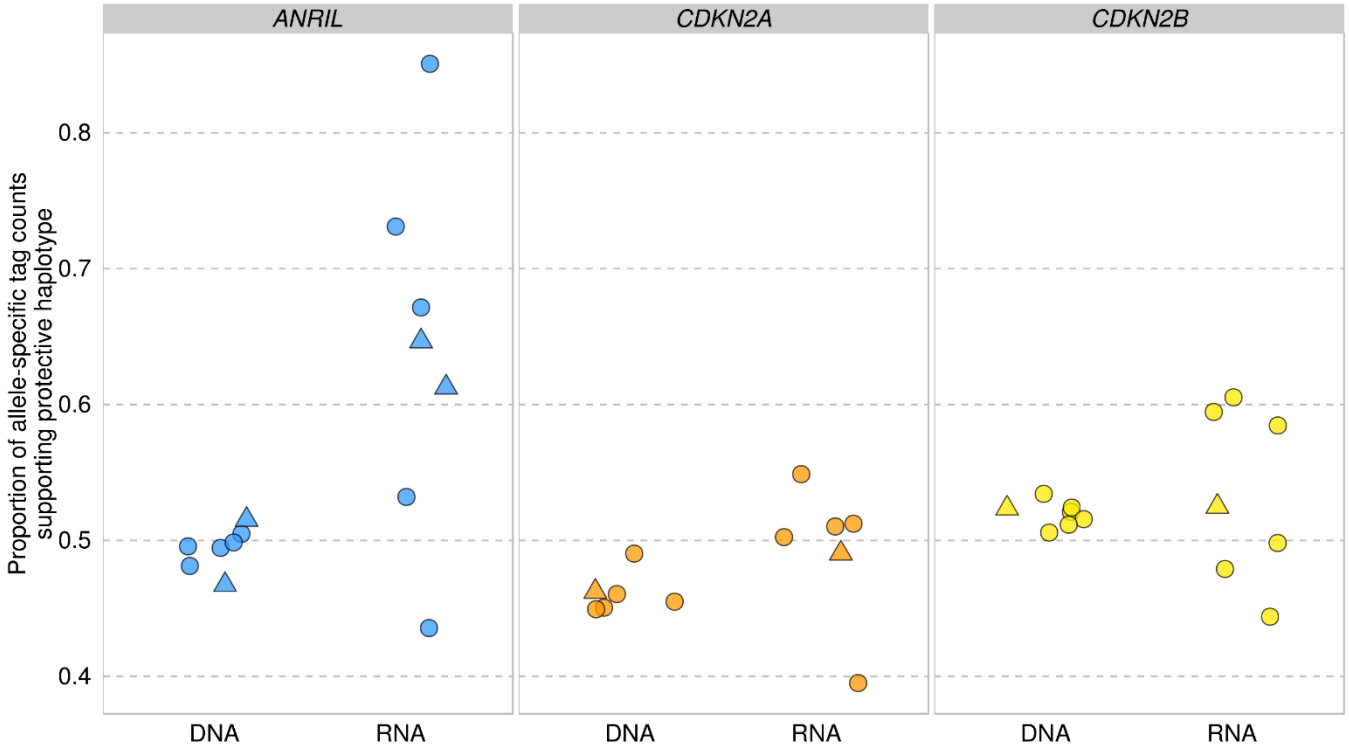
RNA-seq

A : a ≠ 50 : 50



ASE analysis in endometrial cell lines and normal endometrial tissues

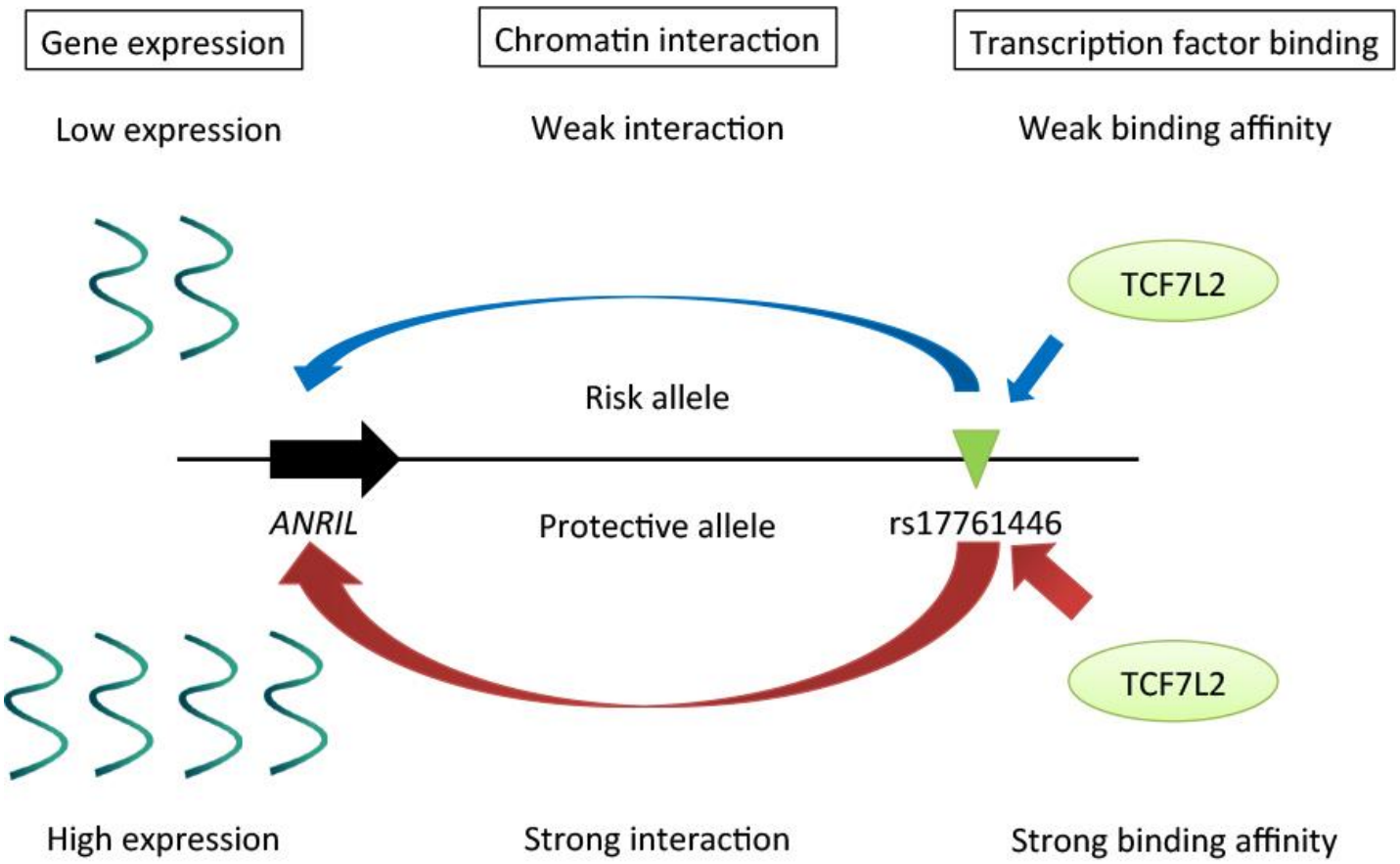




Protective G allele

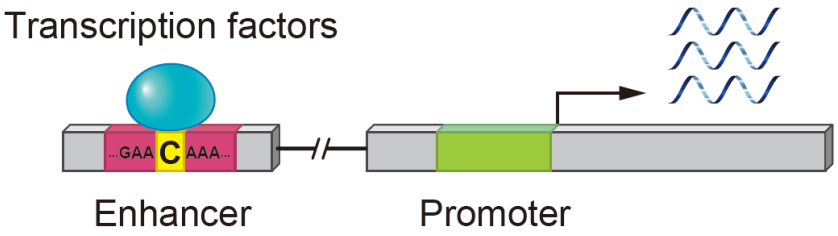
⇒two times greater expression of *ANRIL* (P = 3.9 x 10^{-3})

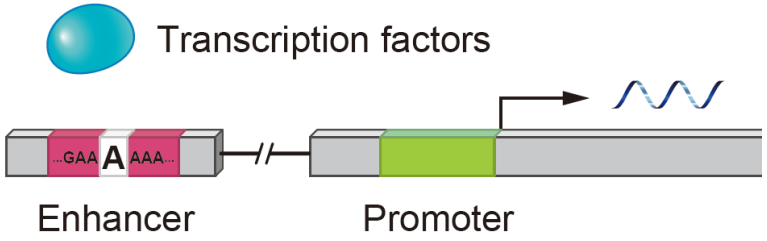
Summary

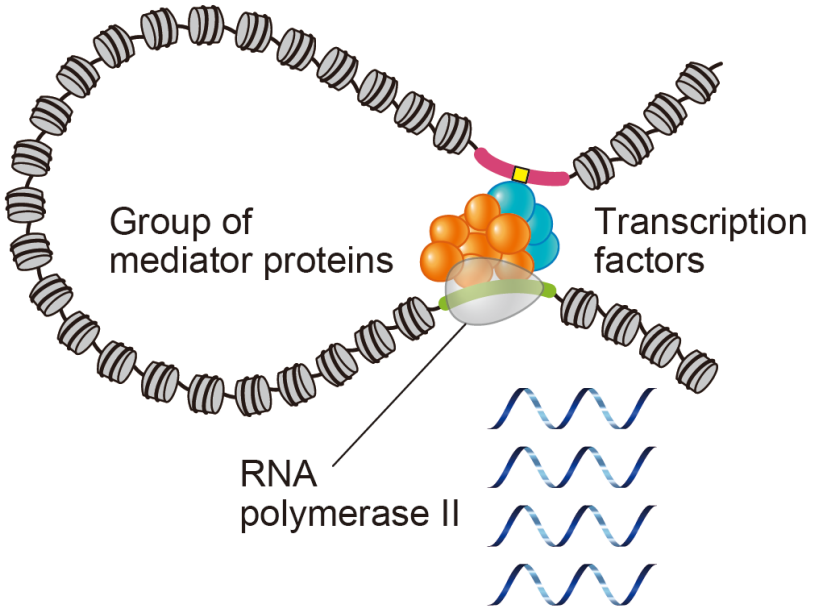


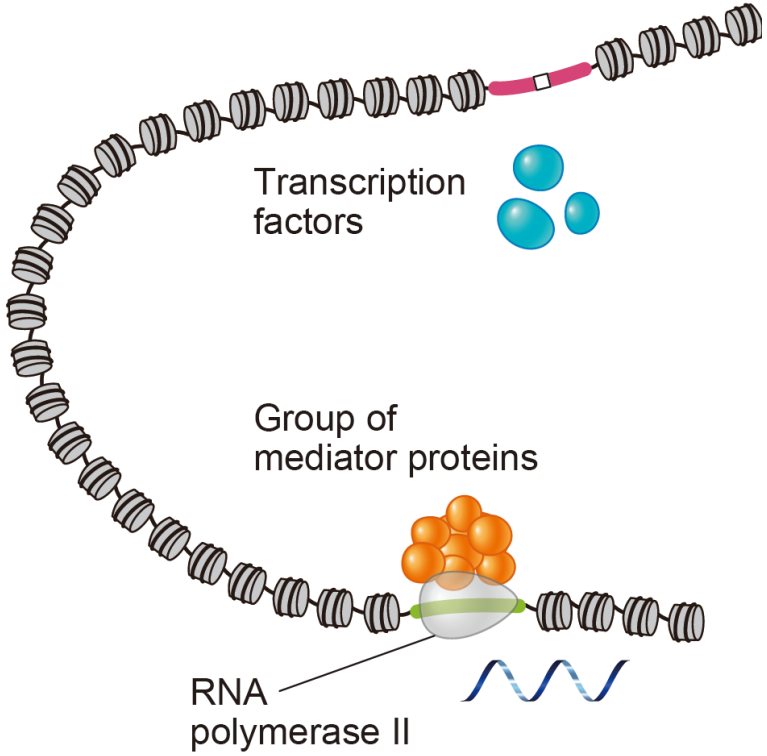
Preferential bindings of TCF7L2 and its coactivator EP300 to the protective G allele of rs17761446 lead to stronger chromatin interaction with the promoter of *ANRIL*, which in turn activate transcription of the non-coding RNA

Plausible transcriptional mechanism underlying disease-associated variants









Acknowledgement

Division of Human Genetics,
National Institute of Genetics
- Ituro Inoue
- Aishwarya Gurumurthy (3C)
- Takahide Hayano (ChIP)
- Somayeh Ahmadloo (ChIP)
- Waleed Omer (Immunofluorescence)
- Chihiro Hata (PCR)
- Masahiko Takahashi (NGS)
- Takayasu Mori (ChIP)
- Junko Kajiwara
- Junko Kitayama
- Yumiko Sato

Niigata Univ.
- Kenichi Tanaka
- Kosuke Yoshihara

Nihon Medical College
- Shigeo Akira

Kanazawa Univ.
- Kazuyoshi Hosomichi (NGS)