

遺伝統計解析を通じた 疾患病態解明・ゲノム創薬への展望

岡田 随象

東京医科歯科大学 疾患多様性遺伝学分野

テニュアトラック講師



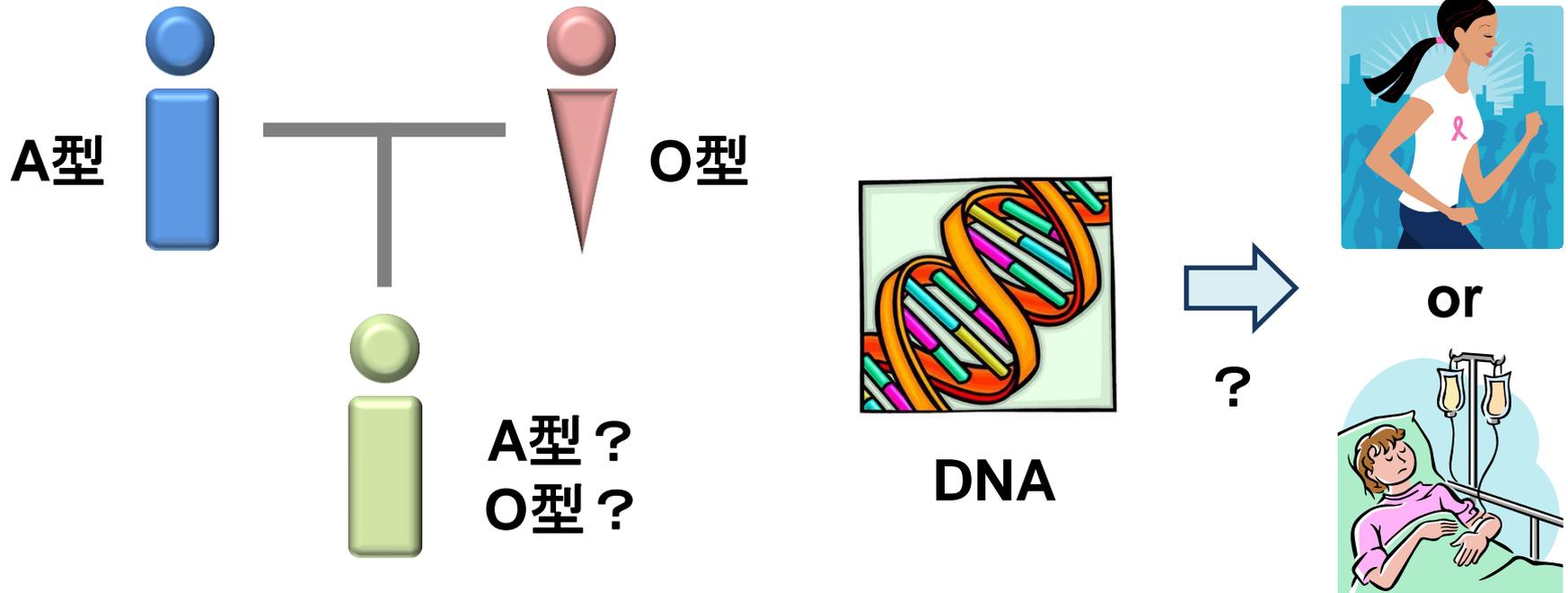
【講演内容】

- ゲノム解析(遺伝統計解析)とは？
- ゲノムワイド関連解析とその成果
- 疾患病態解明・ゲノム創薬への展望
- これからのゲノム解析



【ゲノム解析(遺伝統計解析)とは?】

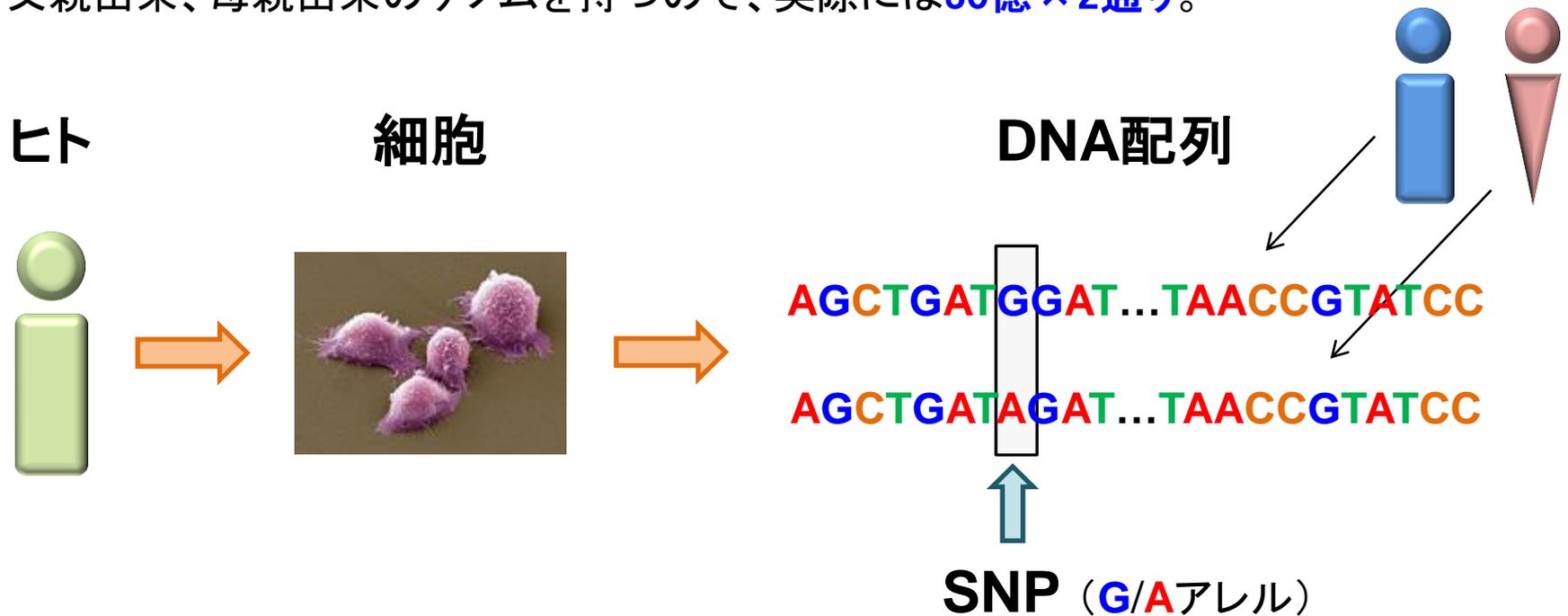
- ・「**遺伝情報**」と「**形質情報**」との結びつきを、統計解析を通じて評価すること。



- ・**遺伝情報**・・・生物の細胞の中に、DNAとして保存されている。
ヒト個人間ではほとんど同じだが(99.9%)、少しずつ違う。
- ・**形質情報**・・・ヒトから得られる情報。個人間で違うものを対象とすることが多い。
(血液型、性別、身長・体重、病気 etc.)

【遺伝情報とは？】

- ・ 生き物の体は、**細胞**が集まって構成されている。
- ・ 各細胞の中に、遺伝情報が**ゲノム(DNA)**として保存されている。
- ・ ゲノムは、**A、T、G、C**、の4種類の塩基の配列で構成されている。
- ・ 塩基配列の長さは生物種によって異なり、ヒトゲノムの場合、**約30億の塩基配列**。
- ・ 父親由来、母親由来のゲノムを持つので、実際には**30億×2通り**。



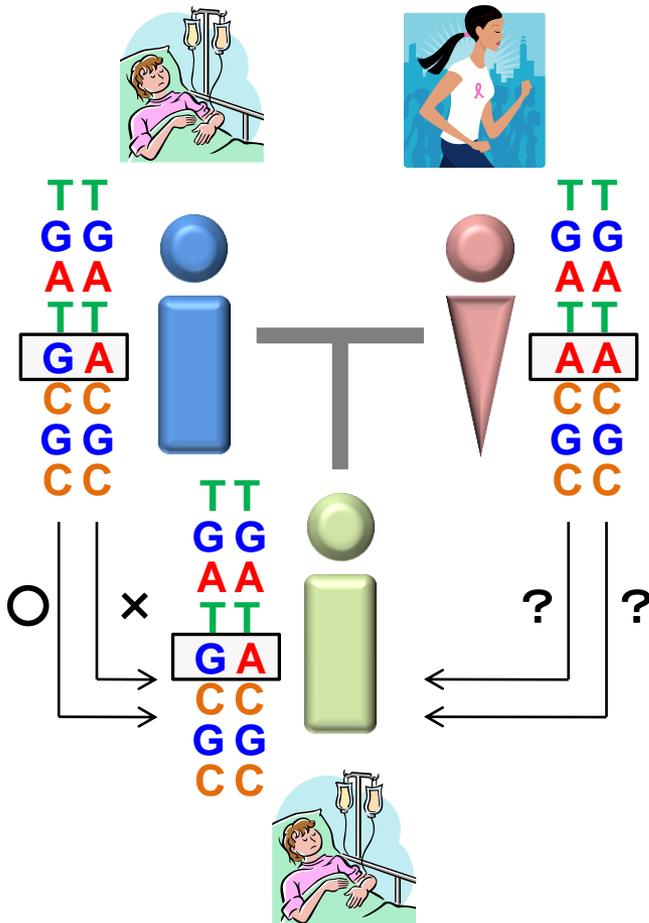
- ・ ヒトゲノム上の塩基配列は、ヒト個人間/個人内で少しずつ異なる(**多型**)。
- ・ 最も代表的なのは、一塩基の違い(一塩基多型、SNP: single nucleotide polymorphism)。

【遺伝統計解析の色々】

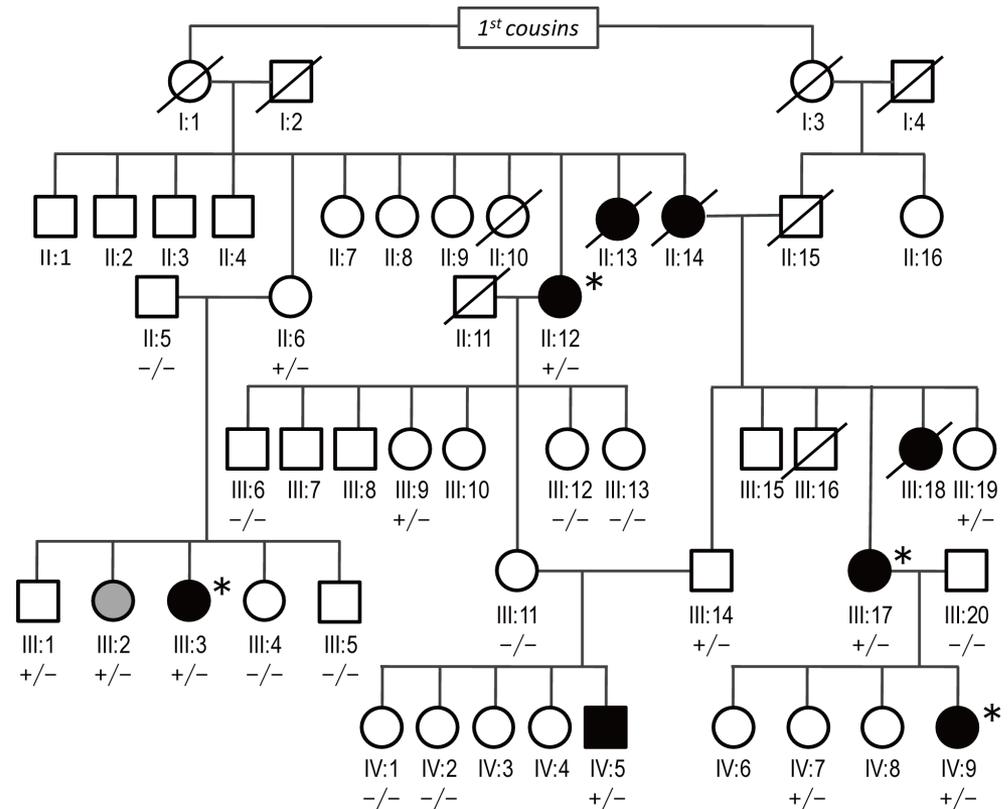


連鎖解析 (Linkage analysis)

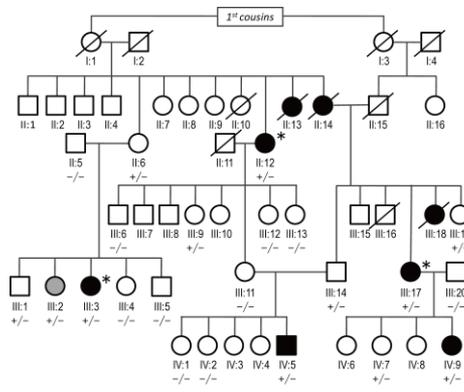
- ・ 家系内での遺伝情報の伝達と、対象形質の受け継がれ方を比較することにより、形質の原因となっている多型を同定することができる。
- ・ 古くはメンデルの法則までさかのぼることのできる、伝統のある遺伝統計解析手法。



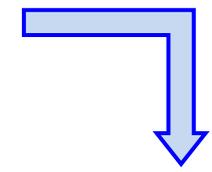
【家族性関節リウマチにおける実施例】



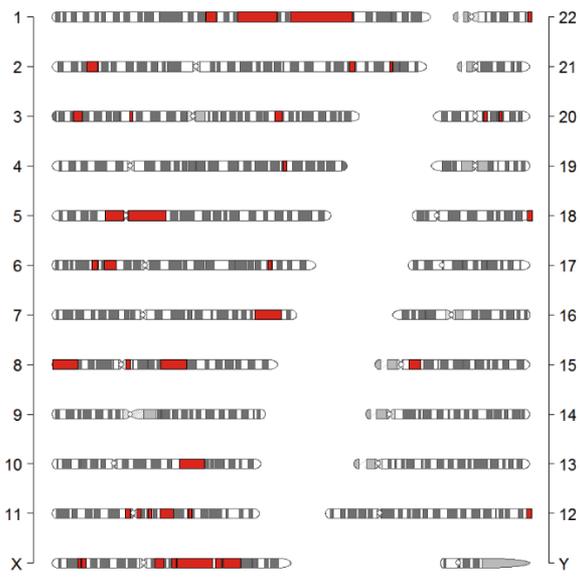
(*PLoS One*. Okada Y. et al. 2014.)



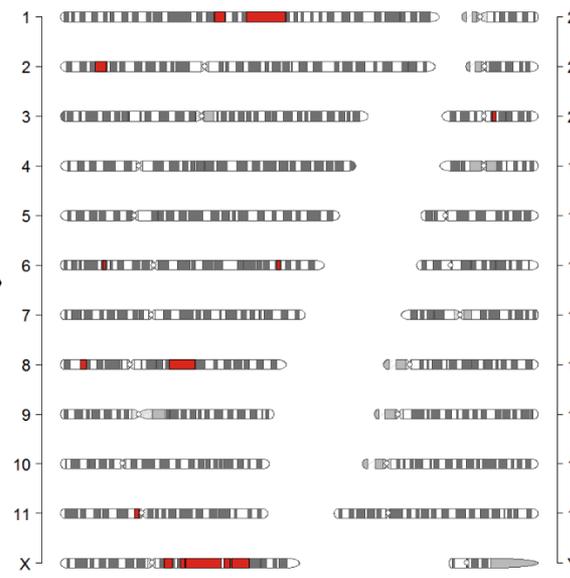
遺伝統計解析



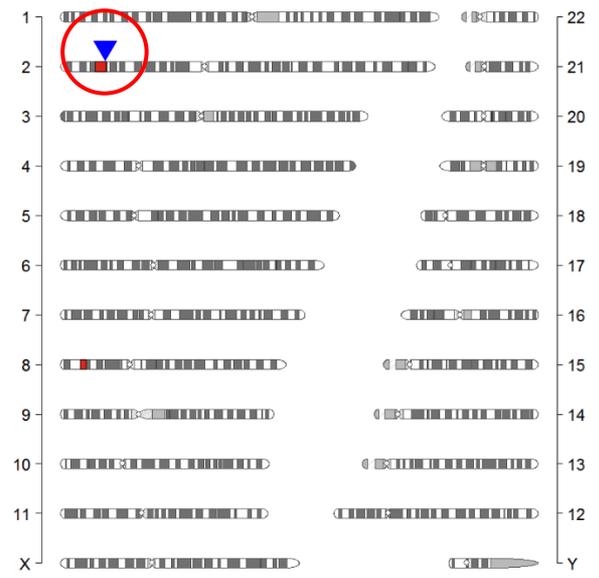
IBD stretches among 4 RA cases



IBD stretches among 5 RA cases and 1 ACPA-positive control



IBD stretches among 5 RA cases, 1 ACPA-positive control, and ≤ 9 ACPA-negative controls



■ ... IBD stretch

▼ ... Candidate causal SNV

・ 遺伝統計解析の第一目的は、疾患原因遺伝子の“住所”をつきとめること

患者対照群解析 (Case control study)

- ・ アレルの組み合わせ (ジェノタイプ) の頻度差を、患者群と対照群とで比較する。
- ・ 患者群で有意に頻度の高いアレルは、疾患罹患リスクを有すると考えられる。

QTL解析 (Quantitative trait locus analysis)

- ・ 連続量を持つ形質 (ex. 身長、臨床検査値) の、ジェノタイプ別平均値の差を評価する。

～患者対象解析の例～

	SNP ジェノタイプ		
	GG	GA	AA
患者群 (N=200)	32	96	72
対照群 (N=200)	50	100	50

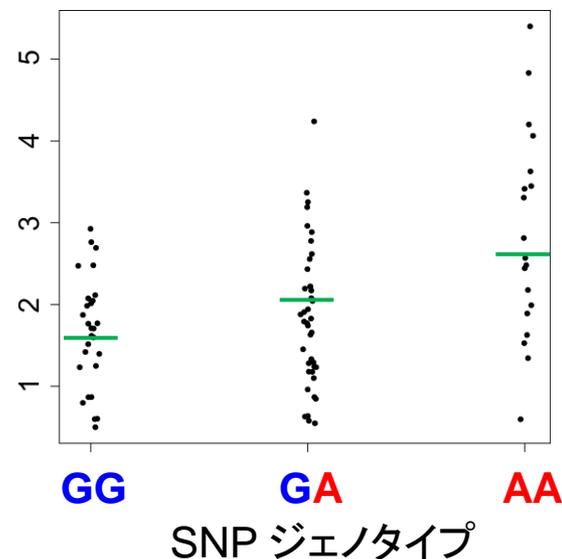
患者群のアレルA頻度 = 0.6

対照群のアレルA頻度 = 0.5

アレルAのオッズ比 = 1.50 95%CI: 1.13 – 1.98

Cochran-Armitage's trend test P-value = 0.0047

～ QTL解析の例～



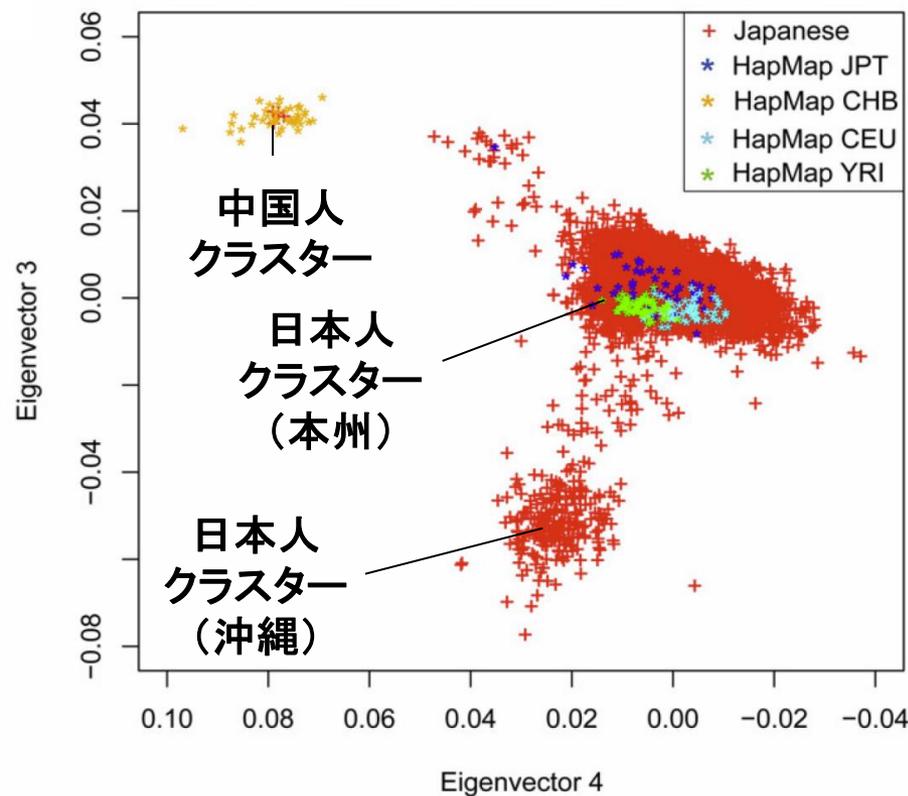
各ジェノタイプ毎の平均値差 = 0.5

線形回帰 P-value = 0.00023

【主成分分析による日本人集団の分類】

	SNP1	SNP2	SNP3	SNP4	SNP5
Sample1	1	0	1	1	1
Sample2	1	2	1	2	1
Sample3	0	1	1	1	1
Sample4	0	2	0	2	1
Sample5	1	1	1	0	1
Sample6	1	1	1	1	1
Sample7	2	0	0	2	0
Sample8	1	1	0	1	1
Sample9	1	1	2	2	2
Sample10	0	2	1	1	2

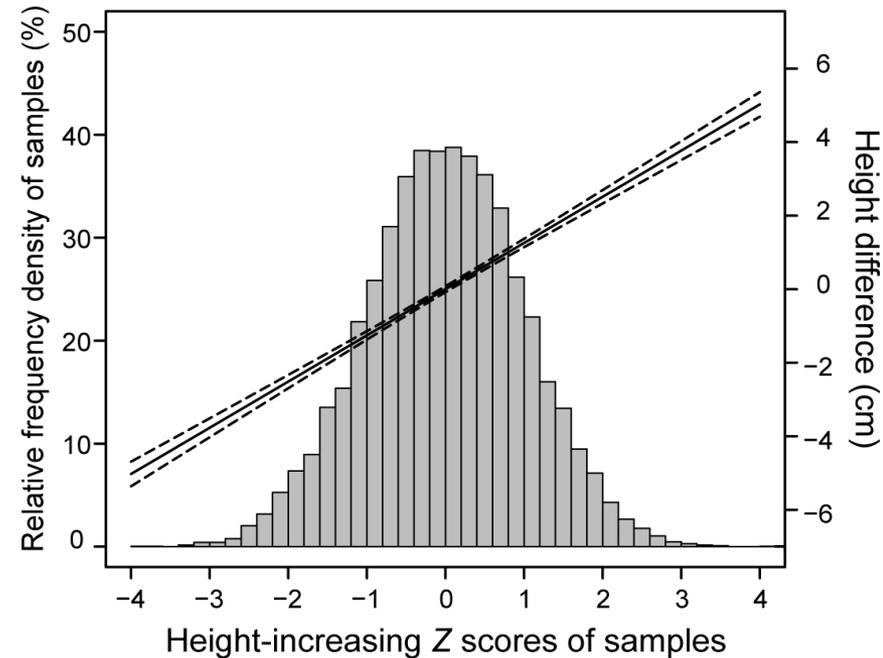
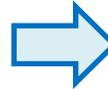
PCA
→



- ・ 日本人集団7000人のジェノタイプデータを対象に主成分分析 (PCA) を行った。
- ・ 遺伝的背景が、**本州と沖縄とで(ほんの少しだけ)違う**ことがわかった。

【成人身長予測モデルの構築】

Effect size	SNP1	SNP2	SNP3	...	SNP51	Normalized Score
	0.1	0.2	0.05	...	0.08	
Sample1	0	0	1		0	0.14
Sample2	2	0	0		2	-0.91
Sample3	1	2	0		1	-0.62
Sample4	0	1	2		2	1.28
Sample5	0	0	1		0	0.72
Sample6	2	1	0	...	1	-0.59
Sample7	2	1	1		2	1.12
Sample8	1	0	1		1	-0.33
Sample9	1	1	2		1	0.39
Sample10	1	1	0		2	-0.30



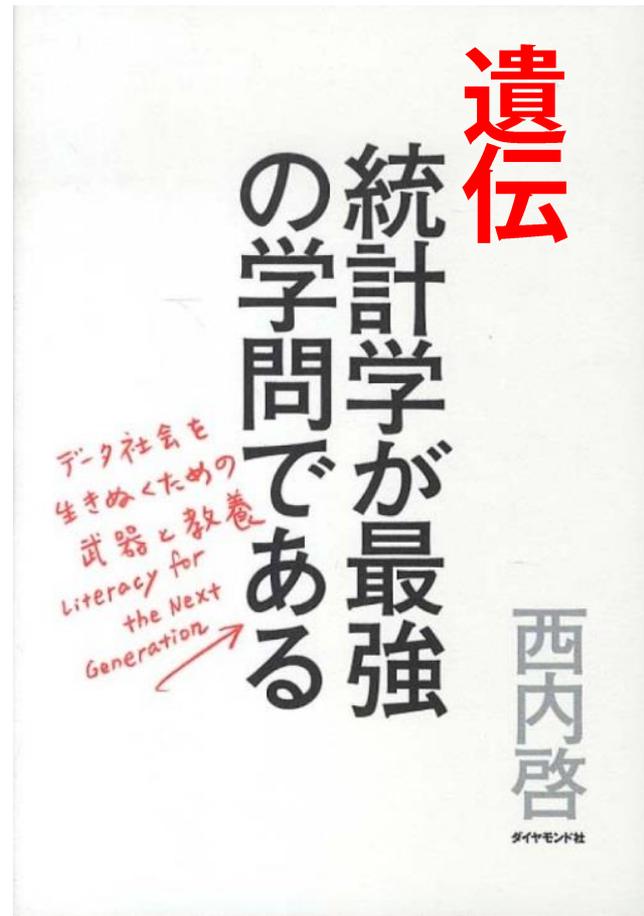
$$0 \times 0.1 + 0 \times 0.2 + 1 \times 0.05 + \dots + 0 \times 0.08 = 0.14$$

$$2 \times 0.1 + 0 \times 0.2 + 0 \times 0.05 + \dots + 2 \times 0.08 = -0.91$$

$$1 \times 0.1 + 2 \times 0.2 + 0 \times 0.05 + \dots + 1 \times 0.08 = -0.62$$

- 成人身長との関連が報告された計51領域の組み合わせが、身長分布に与える影響を評価したところ、身長の分散の約5%を説明することが判明した。

【遺伝統計学と統計学の違い】



～ 真に「最強」なのは、**遺伝統計学**である ～

【遺伝統計学は何故「最強」なのか？】

遺伝
統計学



DNA

RNA

Protein

Cell

Tissue

Organ



病気との
関連の強さ

中

大

相関関係

小

大

因果関係

大

小

- ・ 統計学で最も難しいこと(の一つ)は、**因果関係の証明**。
- ・ 遺伝統計学においては、**遺伝情報** → **形質情報の因果関係**が担保されている。
- ・ 因果関係に基づく故に結果・解釈が**頑強**であり、**応用性・再現性**が高い。
- ・ 統計学者・情報学者の遺伝統計学への参入が急増中(**就職市場でも「最強」**)。

【次世代シーケンサーを用いた解析】

ARTHRITIS & RHEUMATISM

Vol. 65, No. 8, August 2013, pp 1975–1979

DOI 10.1002/art.38011

© 2013, American College of Rheumatology

EDITORIAL

Entering the Age of Whole-Exome Sequencing in Rheumatic Diseases: Novel Insights Into Disease Pathogenicity

Yukinori Okada and Robert M. Plenge

If you sequenced the genome of a patient with a rare disease to search for a pathogenic mutation, what would you expect to find? In this issue of *Arthritis & Rheumatism*, Belot et al (1) and Özçakar et al (2) consider the question in greater detail. In this editorial, we provide a practical perspective on interpreting whole-exome sequencing, with a focus on its application to disease pathogenicity and the development of new drugs/treatment strategies.

mutations from whole-exome sequencing is not as straightforward as it might at first seem. Here, we describe how to interpret findings from a typical whole-exome sequencing study (Figure 1). We focus on data interpretation rather than technical features of data generation.

First, it is important to consider the genetic mode of inheritance by analyzing disease segregation in the family, because this will set expectations for the type of

・ゲノム解析における**レアバリエント解析**や**エクソーム解析**についてのreviewです。

(*Arthritis Rheumatism*. Okada Y. et al. 2013)

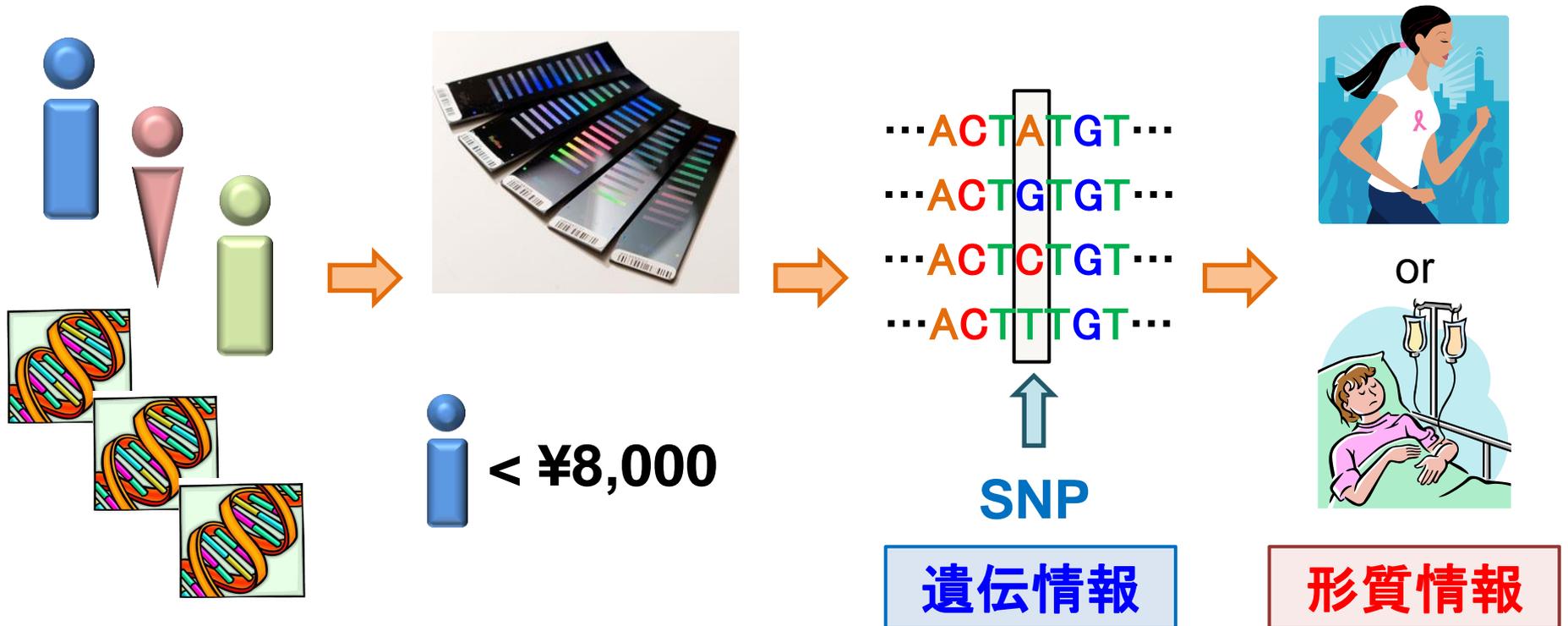
【講演内容】

- ゲノム解析(遺伝統計解析)とは？
- **ゲノムワイド関連解析とその成果**
- 疾患病態解明・ゲノム創薬への展望
- これからのゲノム解析



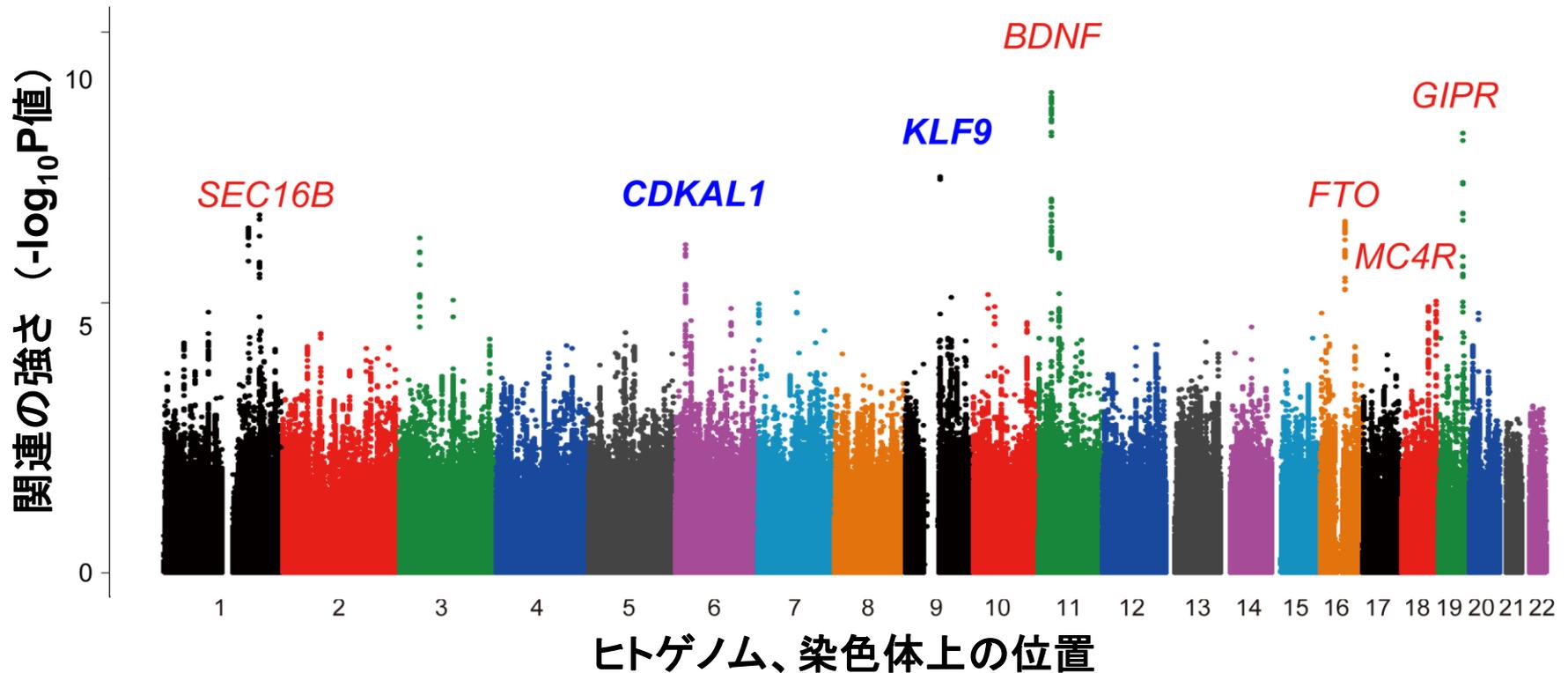
【ゲノムワイド関連解析 (Genome-wide association study; GWAS)】

- ・ **遺伝情報**と**形質情報**との結びつきを評価する、**遺伝統計学**の一手法。
- ・ **数百人～数万人**を対象に、ヒトゲノム全体を網羅する**数百万箇所**の**一塩基多型 (single-nucleotide polymorphism; SNP)**のタイピングを実施し、対象形質との関連を評価する手法。



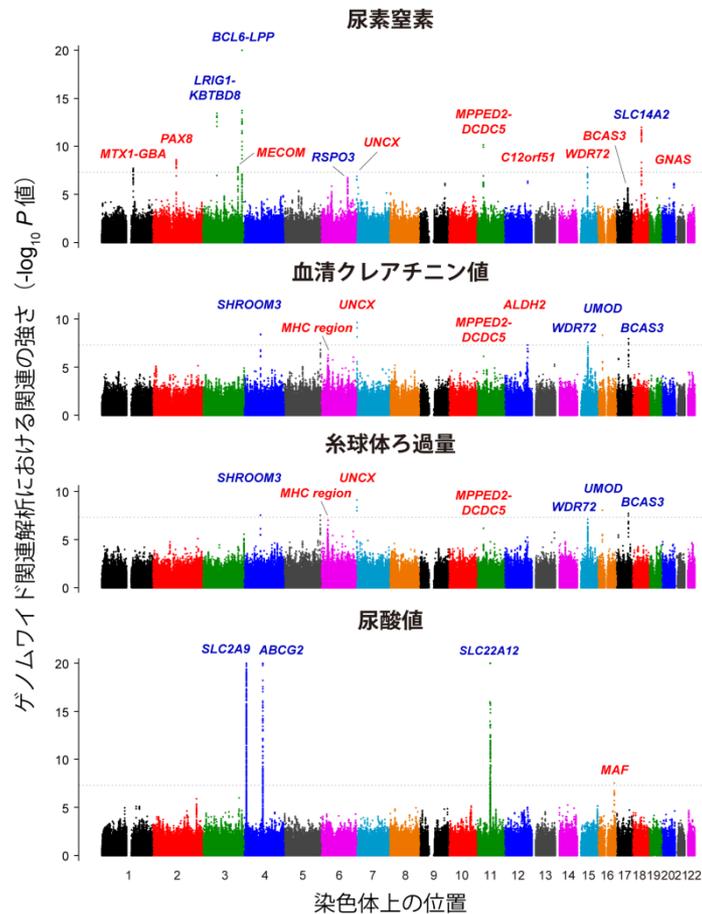
【ゲノムワイド関連解析の一例 ～肥満～】

Manhattan plot (マンハッタン街の高層ビル群に似てるので)



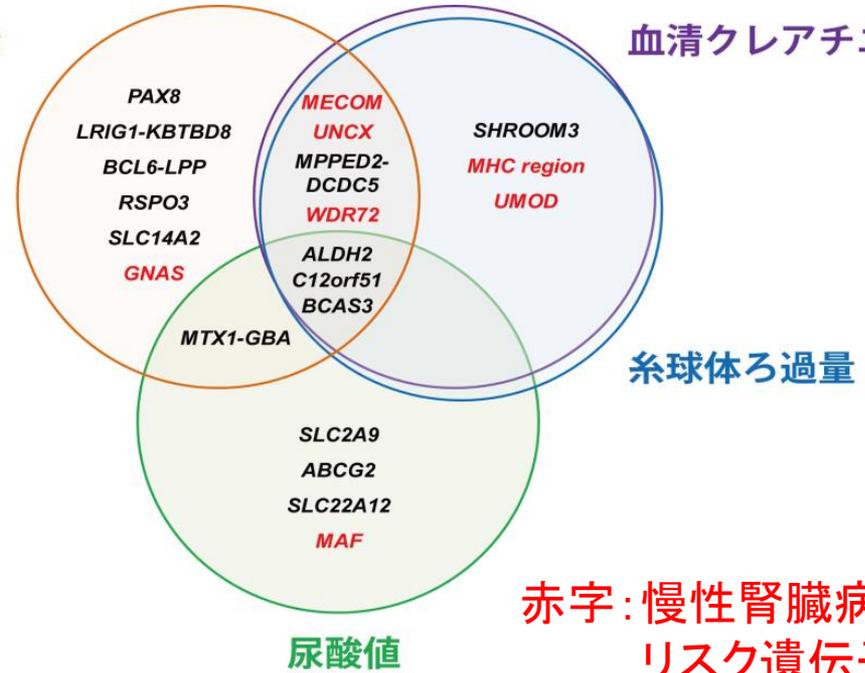
- AGEN consortiumとの共同研究を通じて、**日本人集団26,620名**に対して、肥満の程度 (Body Mass Index: BMI) と250万SNPとの関連を評価。
- **7個の関連遺伝子**を同定し、そのうち**2個が新規の発見**であった。

【ゲノムワイド関連解析の一例 ～腎機能指標～】



尿素窒素

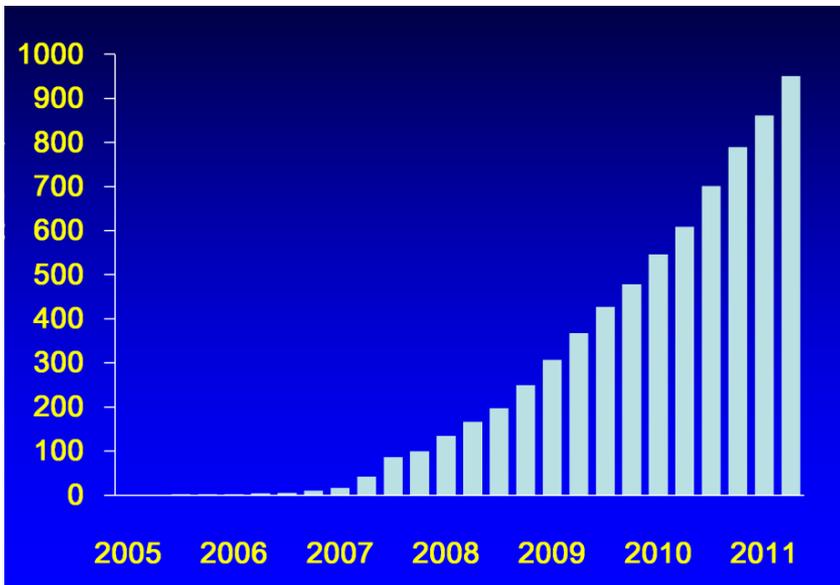
血清クレアチニン値



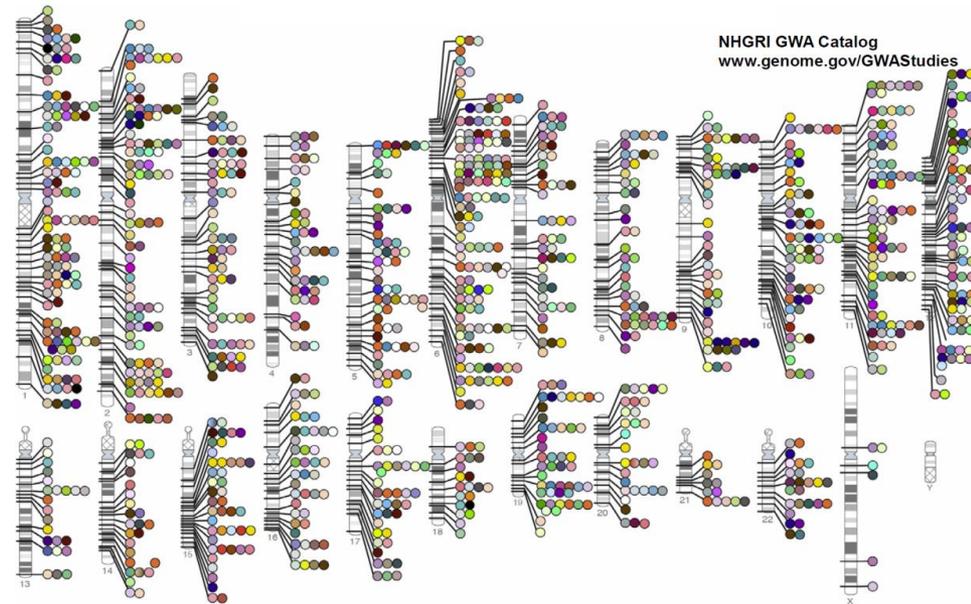
- 東アジア人集団71,149名に対して、腎機能指標(尿素窒素、血清クレアチニン値、糸球体ろ過量、尿酸値)と250万SNPとの関連を評価。
- 21個の関連遺伝子を同定し、そのうち12個が新規の発見であった。

- SNPタイピング技術の発達に伴い、2000年代後半以降、世界中の研究施設でゲノムワイド関連解析が実施されている。
- 2012年までに、300以上の形質に対し1,000以上のゲノムワイド関連解析が報告されている。

GWASの報告数(2005-2011)



GWASで同定された関連遺伝子のマップ



<http://www.genome.gov/GWASStudies/>

Science誌が発表した”Scientific breakthrough of the year 2007” は・・・
⇒ **1位: GWAS**、**2位: iPS細胞**、であった。

Functional SNPs in the lymphotoxin- α gene that are associated with susceptibility to myocardial infarction

Kouichi Ozaki¹, Yozo Ohnishi¹, Aritoshi Iida², Akihiko Sekine², Ryo Yamada³, Tatsuhiko Tsunoda⁴, Hiroshi Sato⁵, Hideyuki Sato⁵, Masatsugu Hori⁵, Yusuke Nakamura^{2,6} & Toshihiro Tanaka¹

Genes screened	13,738
Exon	11,694
Synonymous	1,493
Non-synonymous	1,491
Non-coding region	8,710
Intron	43,910
5' flanking region	2,038
3' flanking region	1,428
Other regions*	6,601
Total	65,671

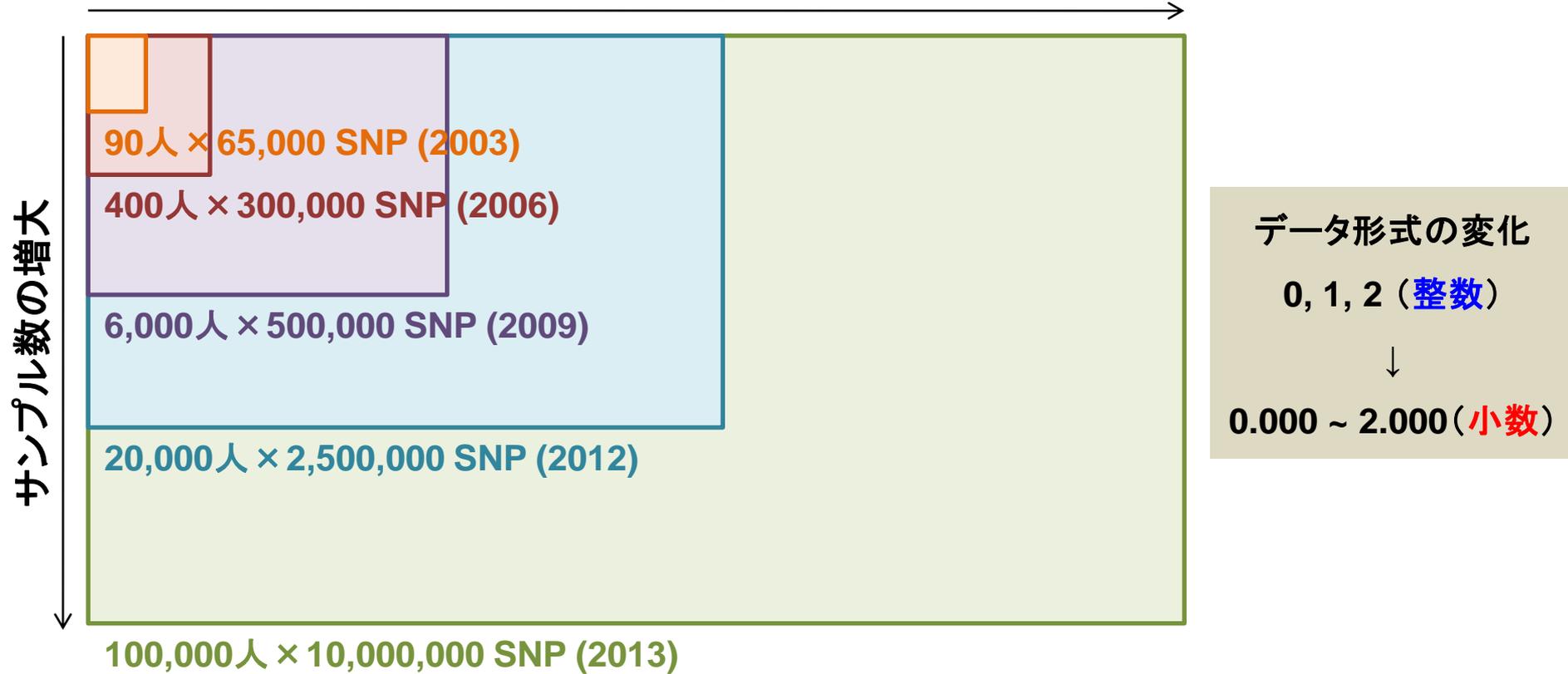
*These SNPs were not mapped within gene regions.

ゲノムワイドな
~65,000 SNPs

- ゲノムワイド関連解析は、**2002年**に日本の**理化学研究所**で**世界に先駆けて**実施された。

【ゲノムワイド関連解析 = ビッグデータ解析】

SNP数の増大



2003年 ⇒ 2013年でデータ容量が $\times 10^5$ 倍 (+ α) に増大！！

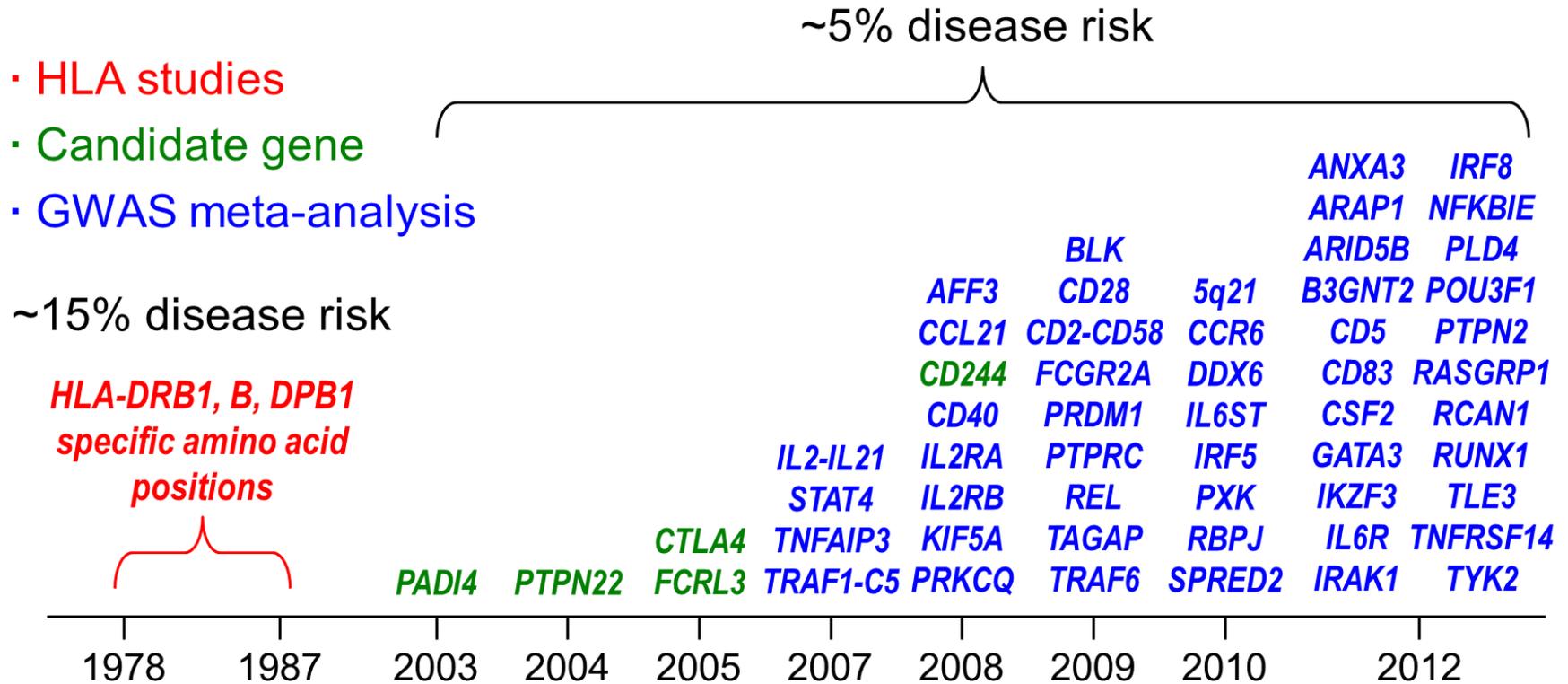
- ・ タイピング技術の商用化やコスト低下に伴い、解析対象となるジェノタイプデータ容量が急激に増加した。
- ・ **ビッグデータ**に対応した解析技術の開発、解析環境の整備が求められている。

【関節リウマチ（Rheumatoid arthritis : RA）】



- ・ 関節リウマチ(RA) : 関節破壊を生じる自己免疫疾患
- ・ 罹患率は0.5%~1%

【関節リウマチ (Rheumatoid arthritis: RA) とGWAS】

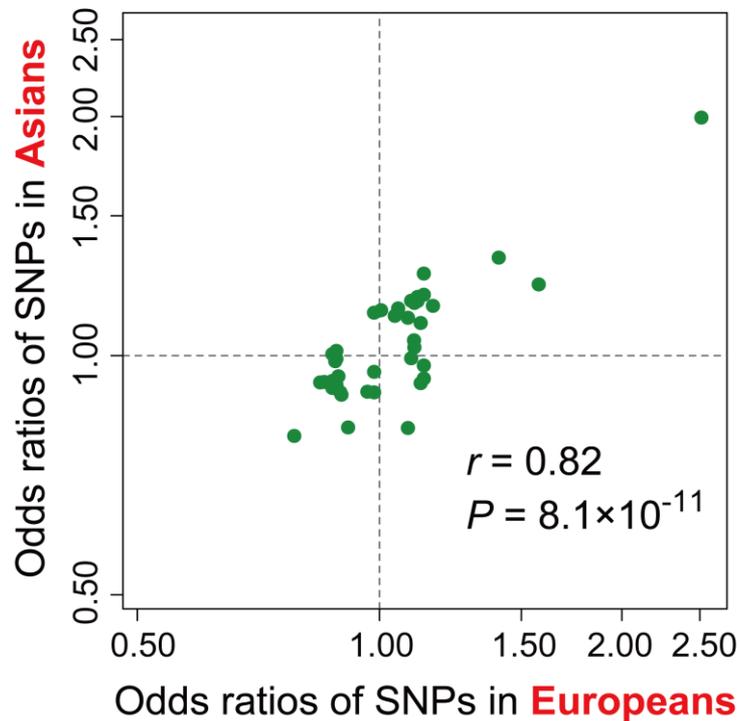


- RAのHeritability (遺伝因子が罹患リスクに占める割合): ~50%
- 2012年までに、60個のRA感受性遺伝子領域が同定されている。

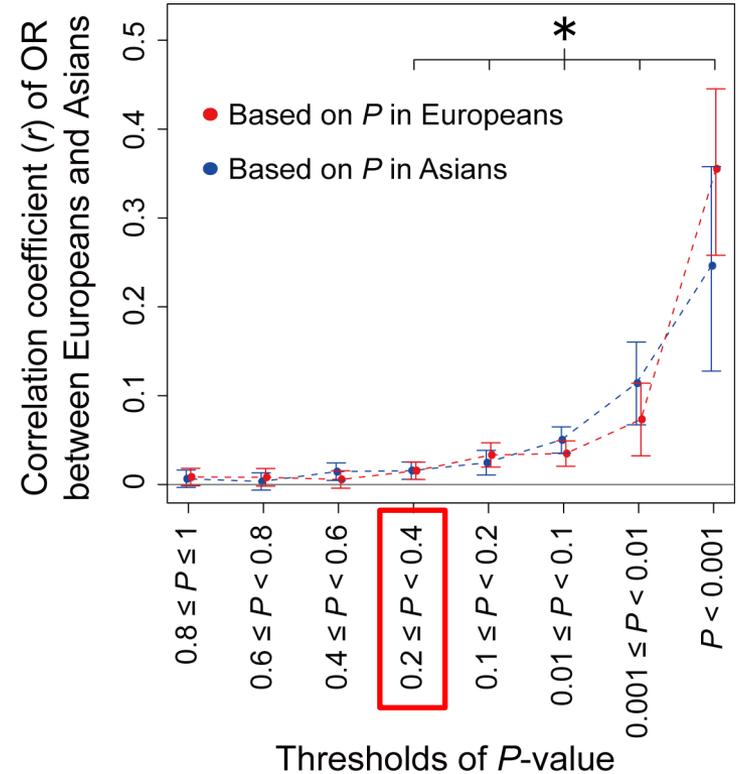
(*Nature Genetics*. Stahl EA et al. 2010, Okada Y et al. 2012, Eyre S et al. 2012)

【RA遺伝リスクの人種間における共有】

既知のRA感受性領域



未同定のRA感受性領域



- ・ RA遺伝リスクの**人種間における共有**が、**既知および未同定のRA感受性領域**において観察された。
- ・ **複数人種を対象としたGWAS**の実施より、さらなる感受性領域の発見が期待された。

欧米人と日本人をあわせたGWASを実施すれば、
より多くのRA感受性領域が見つかるのでは？



Mega-GWAS projectが始動

【国際共同研究チームによる、世界規模のRA GWASの実施】



- ・ **国際共同研究チーム**を通じて、**複数人種、10万人以上**を対象としたRA GWASを実施。
- ・ **25以上の研究グループ**が参画。

【国際共同研究チームによる、世界規模のRA GWASの実施】

Trans-ethnic GWAS meta-analysis

19,234 RA cases and 61,565 controls



In silico replication study

3,708 RA cases and 5,535 controls

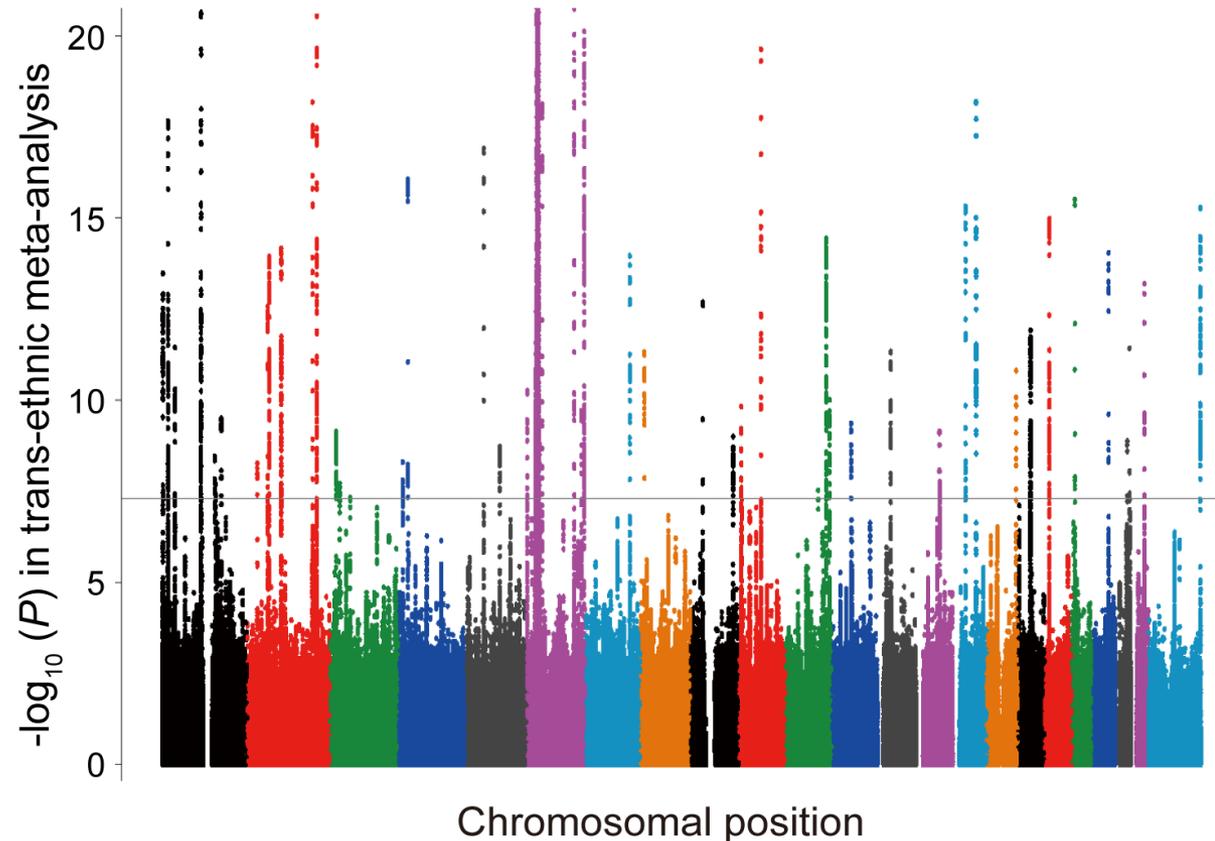


De novo replication study

6,938 RA cases and 6,658 controls



42 novel loci with $P < 5 \times 10^{-8}$



- 1000 Genome imputationにより、X染色体を含む**1000万SNP**を解析。
- **42の関節リウマチ感受性領域を新規に同定**（**計101領域**）。

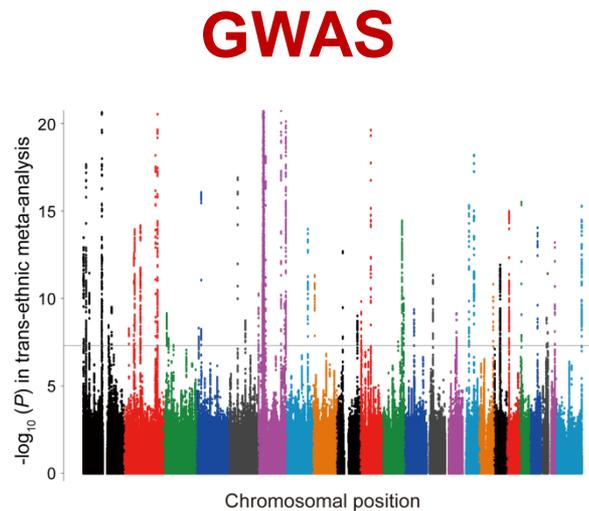
(*Nature*. Okada Y et al. 2014)

【講演内容】

- ゲノム解析(遺伝統計解析)とは？
- ゲノムワイド関連解析とその成果
- 疾患病態解明・ゲノム創薬への展望
- これからのゲノム解析

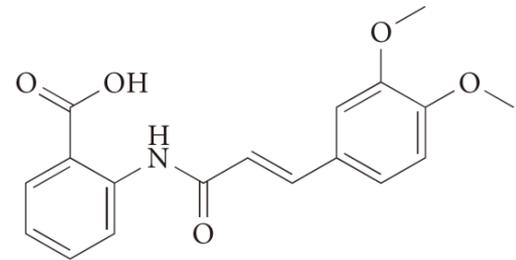


GWASの結果をどのように活用すれば、 疾患病態解明や新規創薬に貢献できるのか？



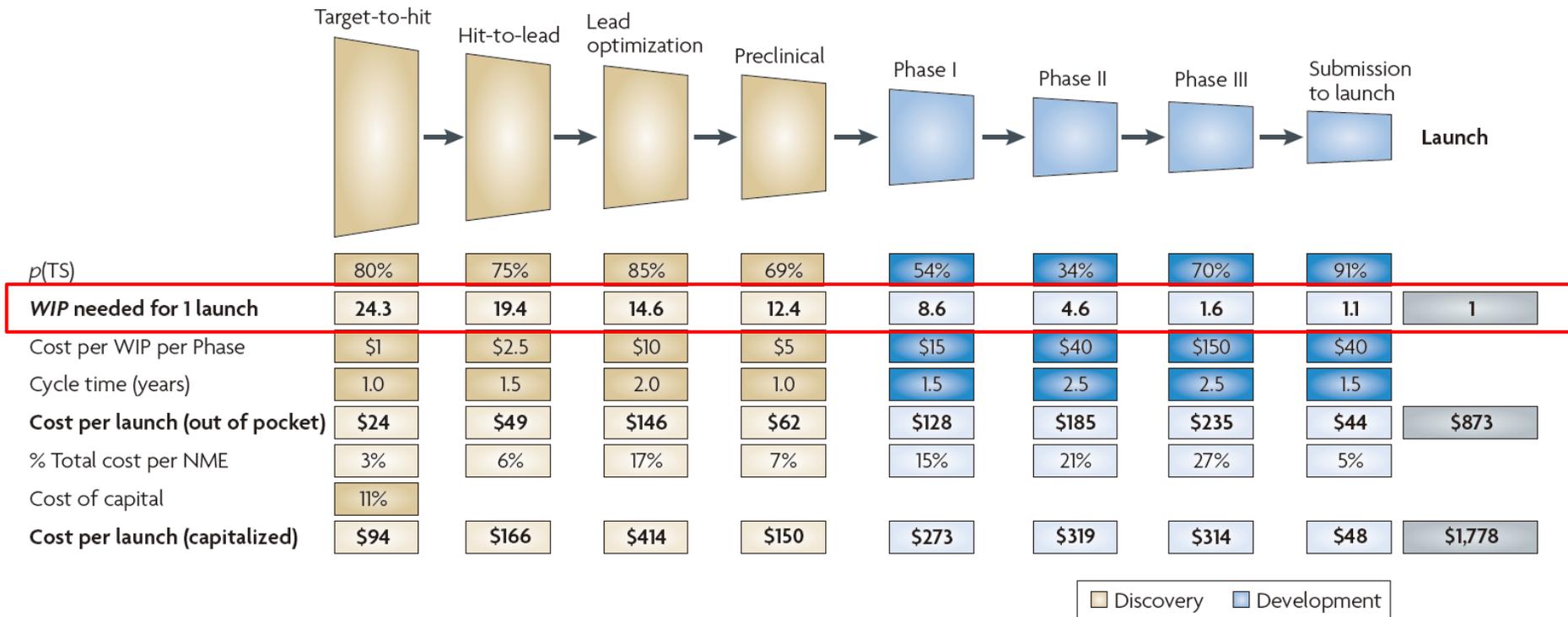
Biology

Drug



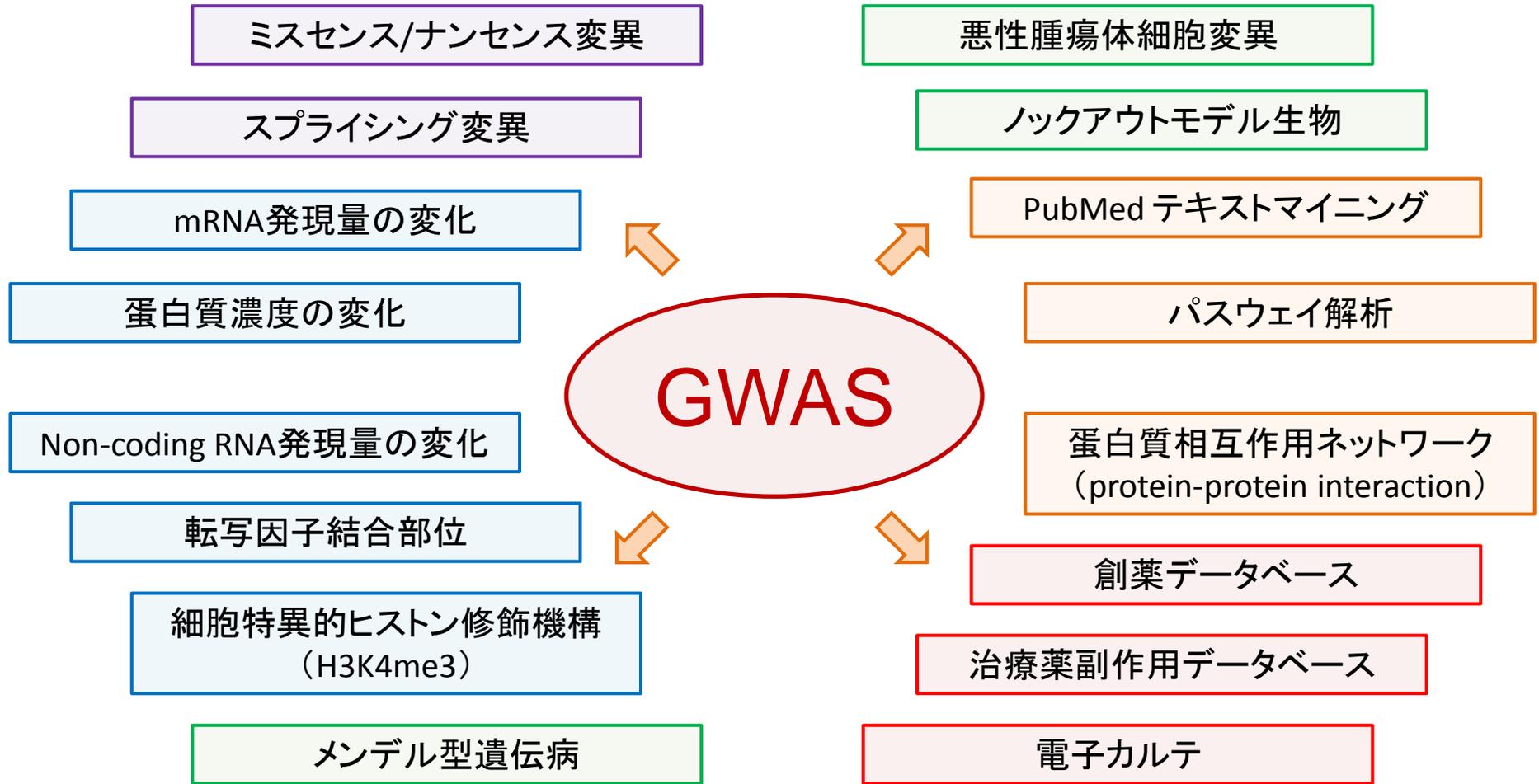
(*Nature*. Okada Y et al. 2014)

【ゲノム創薬を取り巻く現状】



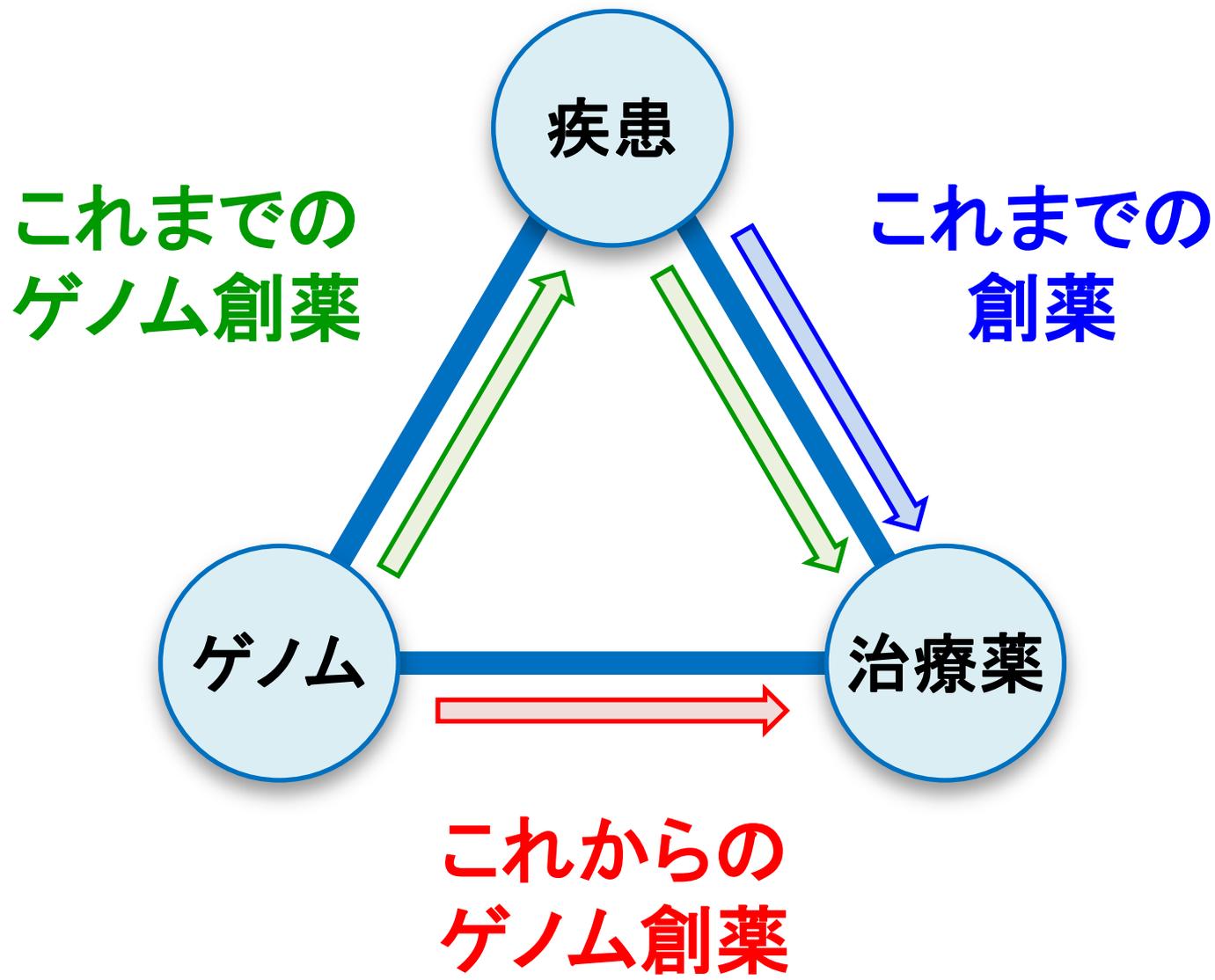
- Clinical trial drugの9割は市場に出ず、新規創薬コストの改善が急務。
- ゲノム創薬 (ex. pharmacogenetics) に関する論文のうち、基準を満たしたArticleはわずか (~2%) で、大部分はReviewだった (2009年時点)。
- 「ゲノム創薬が大事なこと」は皆が知っているが、
「ゲノム創薬が何なのか」は皆がわかっていない。

【GWASデータを用いた疾患病態・新規創薬へのアプローチ】



- ・ **GWASデータ**を、多様な**生物学的データベース**や**創薬データベース**と統合することにより、**疾患病態の解明**や**新規創薬**が可能となる。

【ゲノム創薬の新しいモデル】

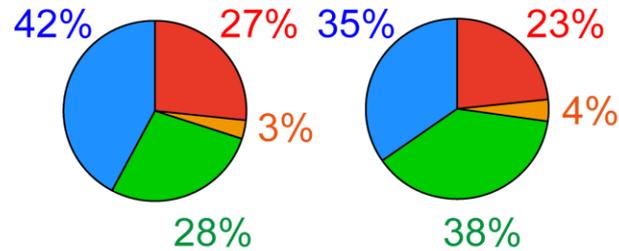


【RAリスクSNPの機能分類】

Explained heritability

European

Asian



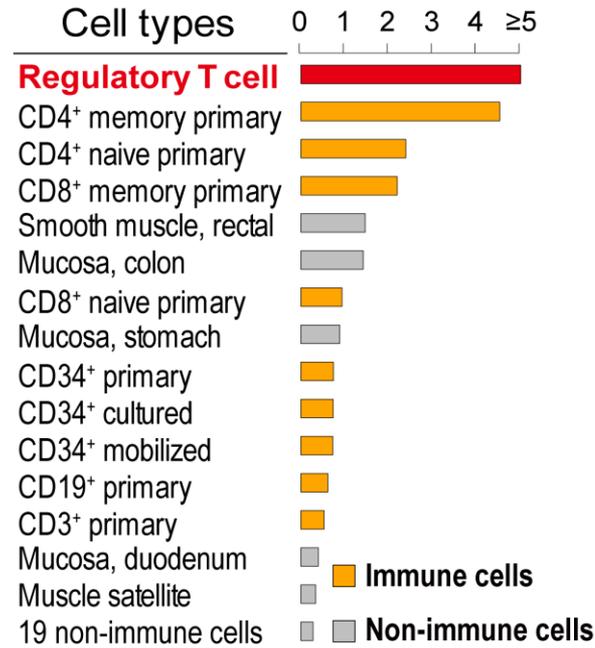
Missense SNPs

Synonymous SNPs

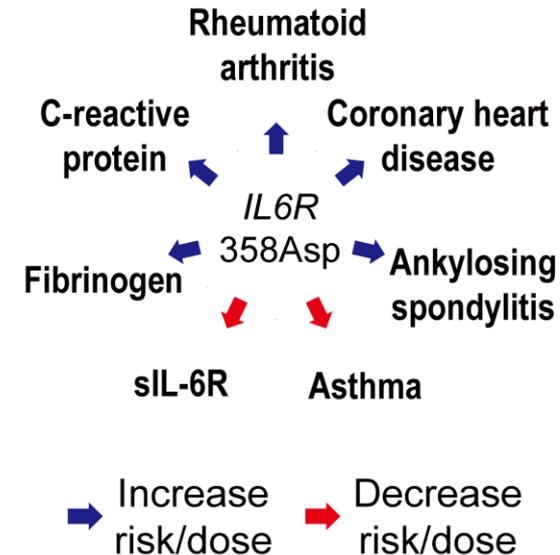
Non-coding SNPs (cis-eQTL)

Non-coding SNPs (others)

$-\log_{10}(P)$ for H3K4me3 enrichment



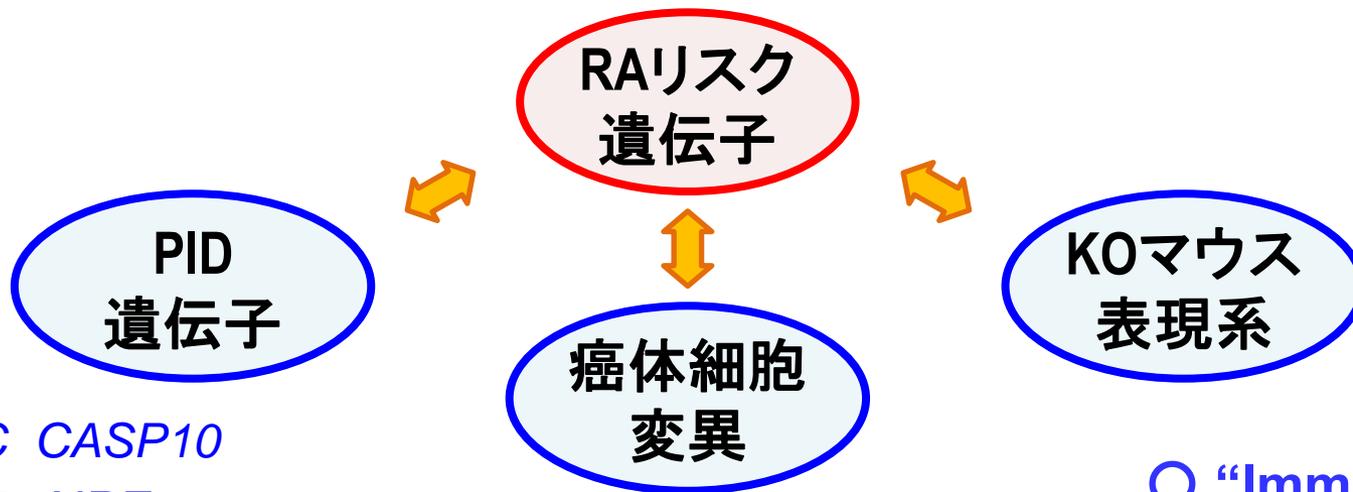
Pleiotropic effect of *IL6R* 358Asp



・ RAリスクSNPの疾患病態への寄与は...

- (1) **ミスセンス変異** および **細胞特異的な cis-eQTL**。
- (2) **制御性T細胞**における**ヒストン修飾機構 (H3K4me3)**。
- (3) 他のヒト疾患と共通した**多面的関連 (pleiotropic effects)**。

【RAリスク遺伝子と機能性遺伝子群との重複】



PTPRC CASP10
RAG1/2 AIRE
CD40 IL2RA
ATM IFNGR2
TYK2 IRF8
UNG MVK
CASP8 C5

○ Hematological cancers

・ Lymphoma

・ Lymphocytic leukemia

× Solid cancers

○ “Immune”

○ “Hematopoietic”

× “Sensory”

× “Neurological”

× “Body size”

・ RA感受性領域内の遺伝子との重複が、下記の遺伝子群で認められた。

(1) 原発性免疫不全症候群(primary immunodeficiency:PID) 遺伝子。

(2) 血液細胞由来悪性腫瘍の体細胞変異遺伝子。

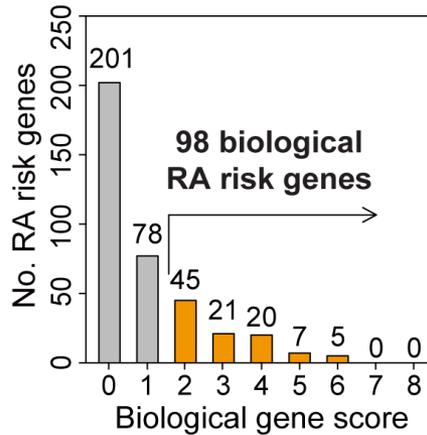
(3) ノックアウトマウス表現系(Immune/hematopoietic traits)。

【生物学的データベースを用いたRAリスク遺伝子の選択】

377 genes from
100 non-MHC
RA risk loci

Prioritization criteria

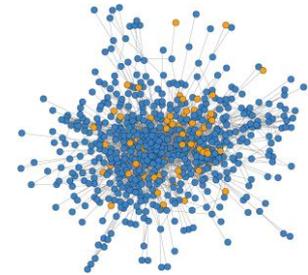
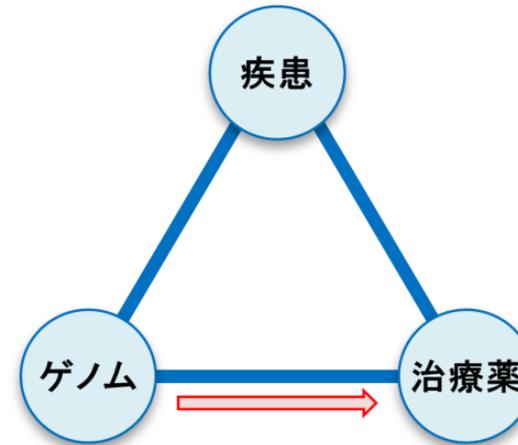
1. Missense SNPs
2. cis-eQTL
3. PubMed text mining
4. PPI
5. PID genes
6. Hematological cancer
7. KO mouse phenotype
8. Molecular pathway



蛋白質間相互作用 ネットワーク

98 biological
RA risk genes

871 drug
target genes



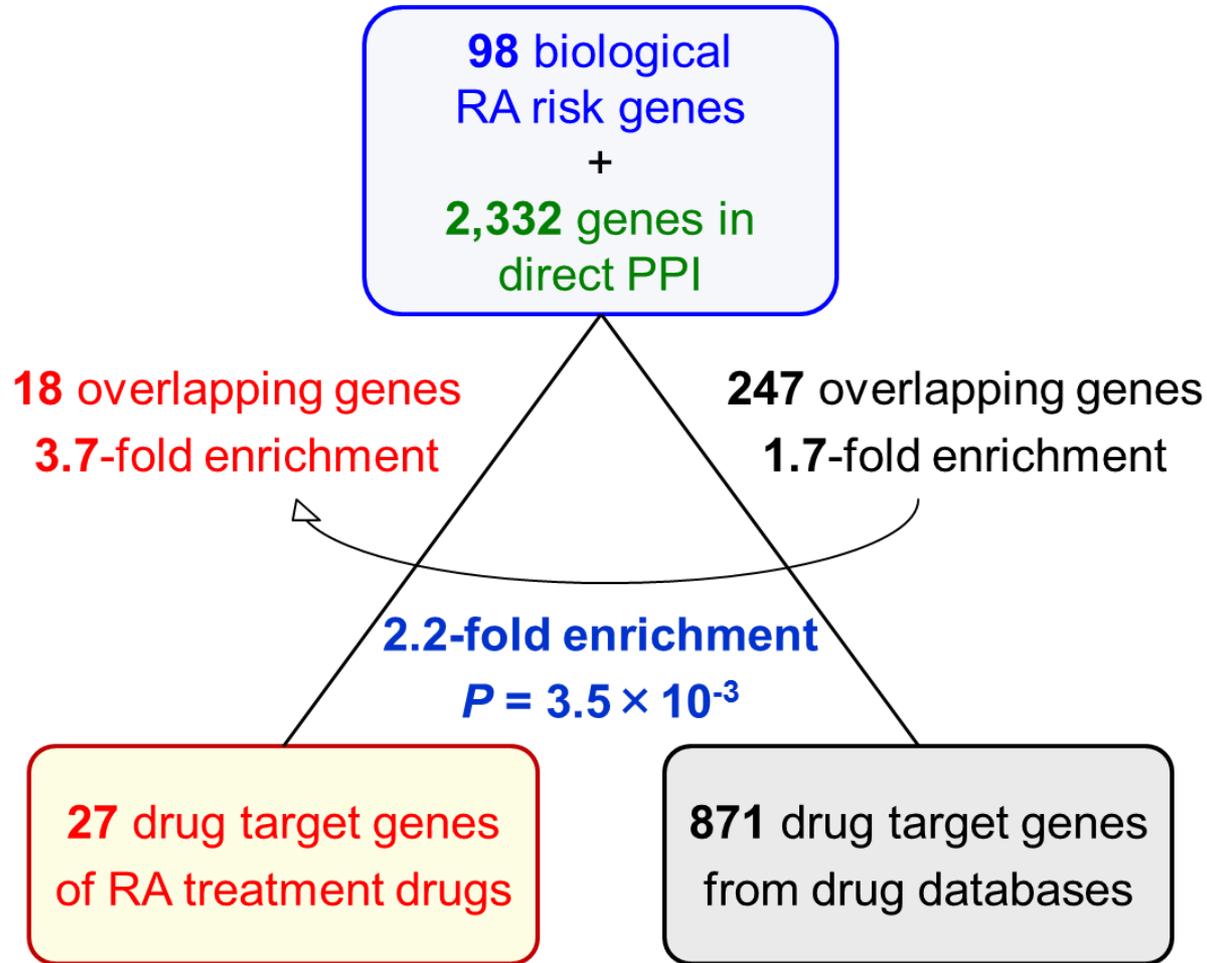
- ・ RA感受性領域に含まれる377遺伝子と生物学的データベースを統合。
- ・ 二つ以上の基準を満たした98遺伝子をRA原因遺伝子候補として選択。
- ・ 創薬データベース上の治療薬ターゲット遺伝子とのつながりを検討

【 生物学的データベースを用いたRAリスク遺伝子の選択 】

関節リウマチ関連一塩基多型	関節リウマチ感受性遺伝子	関節リウマチの病態への関与							細胞特異的な発現制御(ヒストン修飾:H3K4me3)							創薬ターゲット遺伝子			
		遺伝子蛋白質配列の変異	遺伝子発現量の変化	論文データベース上の情報	蛋白質間相互作用ネットワーク	原発性免疫不全症候群原因遺伝子	白血病原因遺伝子	ノックアウトマウスにおける形質	サイトカインシグナル・パスウェイ	制御性T細胞	CD4陽性メモリー細胞	CD4陽性ナイーブ細胞	CD8陽性メモリー細胞	CD8陽性ナイーブ細胞	CD34陽性プライマリー細胞		CD34陽性培養細胞	CD34陽性動員細胞	CD19陽性プライマリー細胞
rs2301888	<i>PADI4</i>																		
rs2476601	<i>PTPN22</i>																		
rs2228145	<i>IL6R</i>																		
chr1:161644258	<i>FCGR2B</i>																		
rs11889341	<i>STAT4</i>																		
rs1980422	<i>CD28</i>																		
rs3087243	<i>CTLA4</i>																		
rs45475795	<i>IL2</i>																		
rs7752903	<i>TNFAIP3</i>																		
rs1571878	<i>CCR6</i>																		
rs4272	<i>CDK6</i>																		
chr7:128580042	<i>IRF5</i>																		
rs10985070	<i>C5</i>																		
rs706778	<i>IL2RA</i>																		
rs773125	<i>CDK2</i>																		
rs1633360	<i>CDK4</i>																		
rs10774624	<i>SH2B3</i>																		
rs8083786	<i>PTPN2</i>																		
rs4239702	<i>CD40</i>																		
rs73194058	<i>IFNGR2</i>																		
rs3218251	<i>IL2RB</i>																		

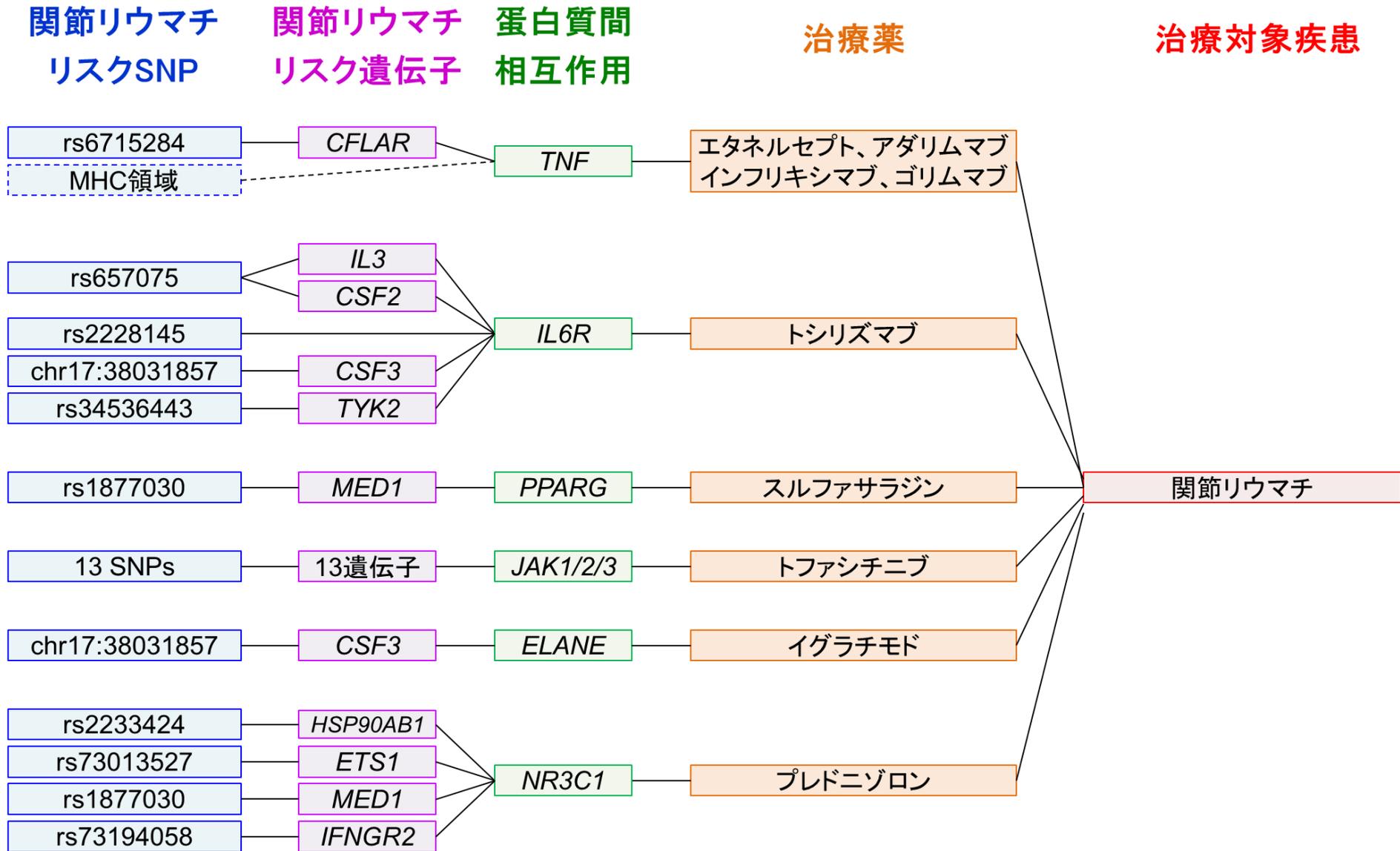
- RA感受性領域内の全遺伝子について、生物学的機能分類を実施。

【RAリスク遺伝子とRA治療薬のつながり】



- RAリスク遺伝子は、蛋白質相互作用ネットワークを介して、RA治療薬のターゲット遺伝子とつながっていることが明らかになった。

【RAリスク遺伝子とRA治療薬のつながり】



【RAリスク遺伝子を利用したDrug repurposing】

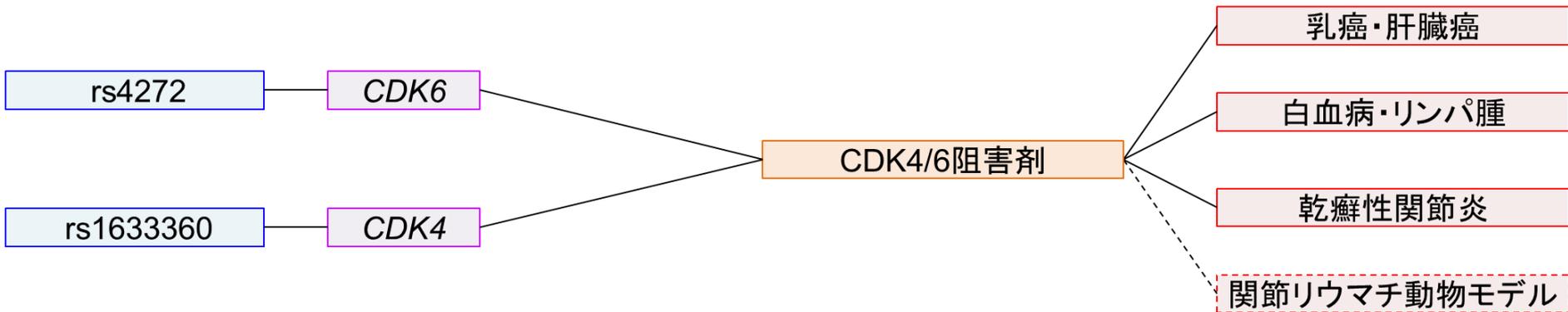
関節リウマチ
リスクSNP

関節リウマチ
リスク遺伝子

蛋白質間
相互作用

治療薬

治療対象疾患



- ・ 既存の治療薬を他の疾患の治療への適用拡大 (**Drug repurposing**) が、新規創薬のコスト改善に貢献すると注目されている。
- ・ **GWAS**リスク遺伝子を直接ターゲットとした**治療薬**に注目することで、Drug repurposing候補を見出すことができる。
- ・ CDK4/6遺伝子をターゲットとした**CDK4/6阻害剤**を有力候補として同定。

【講演内容】

- ゲノム解析(遺伝統計解析)とは？
- ゲノムワイド関連解析とその成果
- 疾患病態解明・ゲノム創薬への展望
- **これからのゲノム解析**



~ Genetics/Gene to ??? (in next 5-10 years) ~

▪ G to A (allele)

▪ G to Q (RNA/protein QTL)

▪ G to E (epigenetics)

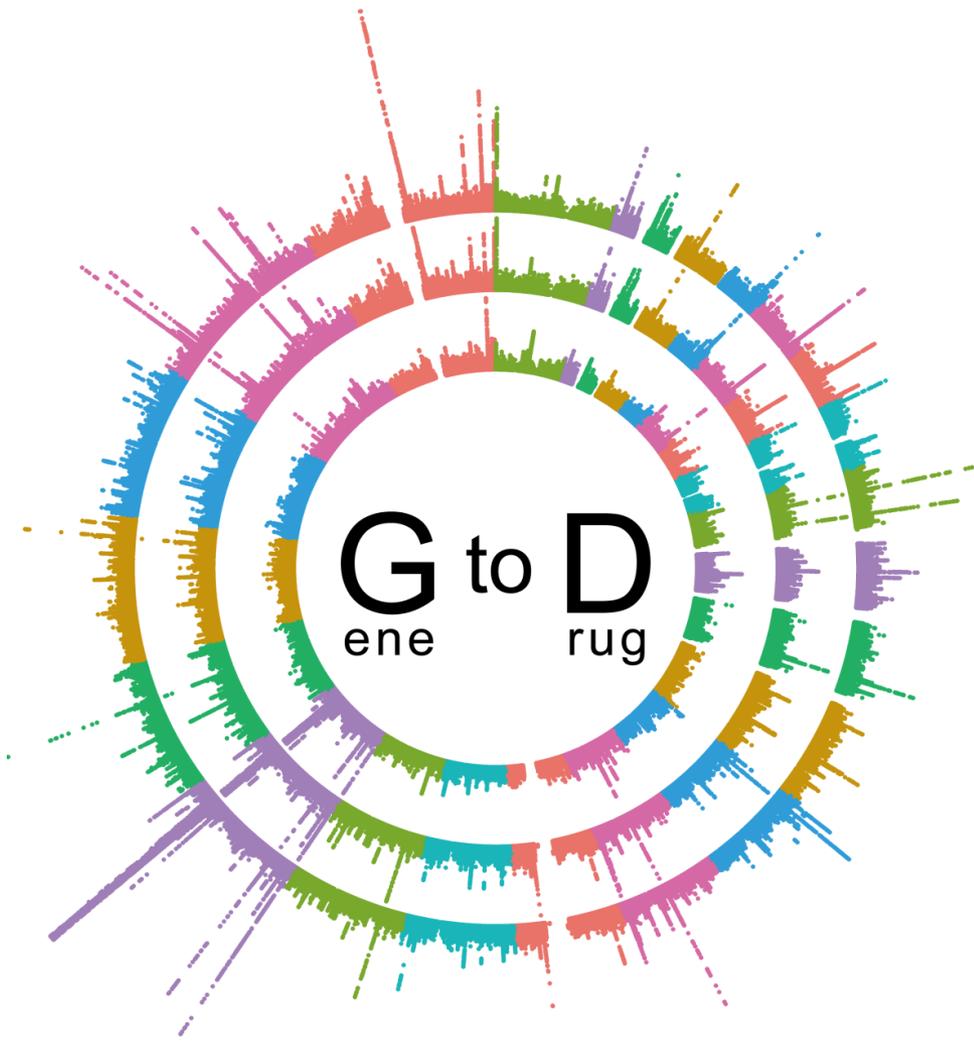
▪ G to M (molecular biology)

▪ G to S (system biology)

▪ G to P (phenome)

▪ G to C (clinical application)

▪ G to D (drug discovery)

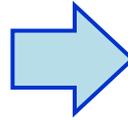


Much progress !!

Now trying !

Not yet ...

【大きなデータ から 質の高いデータ へ】



- ・「データの大きさ」を求める研究が**国際共同研究**に移行していく一方で、「データの質の高さ」を各**研究施設**が追求していく傾向にあります。
- ・データの質 = **ゲノム情報**だけでなく、**組織別**(血清および病理検体)の**エピゲノム情報**、および**臨床情報**(電子カルテ由来の疾患名、検査値、投薬状況)にアクセスが可能かどうか。

【結語】

- 大規模な遺伝情報が得られる時代となり、**遺伝統計学**の分野は大きな進歩を見せている。
- 複数人種を対象とした、**10万人×1000万SNP**の**ビッグデータ解析**を通じて、**42**のRA感受性遺伝子領域を新規に同定し、**総計101領域**とした。
- 生物学的データベースに対する網羅的解析を通じて、RAの疾患病態を新たに同定した(**原発性免疫不全症候群**および**血液細胞由来悪性腫瘍**)。
- 創薬データベースとの照合を通じて、**RAリスク遺伝子**と**RA治療薬ターゲット遺伝子**のつながりが明らかとなった。
- GWASリスク遺伝子を直接ターゲットとした治療薬に注目することで、**Drug repurposing候補**を見出すことができる。
- **ゲノム解析を通じて、疾患の病態解明やゲノム創薬に貢献できる可能性**が示された。

～ RA GWAS結果、遺伝子リスト、創薬ターゲット、ソースコードは下記URLにて公開中 ～

<http://plaza.umin.ac.jp/~yokada/datasource/software.htm>

謝辞(敬称略)

東京大学大学院医学系研究科 アレルギー・リウマチ学教室

— 山本一彦

理化学研究所 統合生命医科学研究センター

— 久保充明 高橋篤 角田達彦 鈴木亜香里 高地雄太 鎌谷洋一郎

東京大学医科学研究所 ヒトゲノム解析センター

— 中村祐輔 松田浩一

京都大学 ゲノム医学センター/医学研究科臨床免疫学

— 松田文彦 三森経世 山田亮 大村浩一郎 寺尾知可史

東京女子医科大学附属膠原病リウマチ痛風センター

— 山中寿 桃原茂樹 猪狩勝則

情報解析研究所

— 鎌谷直之

Brigham and Women's Hospital, Harvard Medical School Broad Institute

— Robert M Plenge, Soumya Raychaudhuri, Paul IW de Bakker

ご質問、ご相談はお気軽に yokada.brc@tmd.ac.jp まで

大学院入学も随時募集中 (医歯学総合研究科 疾患多様性遺伝学分野)