

高血圧の 大規模ゲノム解析

竹内史比古

国立国際医療研究センター(NCGM)研究所

<http://www.fumihiko.takeuchi.name>

ゲノムビッグデータ解析の新潮流

2020.01.17 @日本学会議



血圧の

1. ゲノムワイド関連解析 (GWAS)

- 遺伝的成因の全体像
- 発症予測

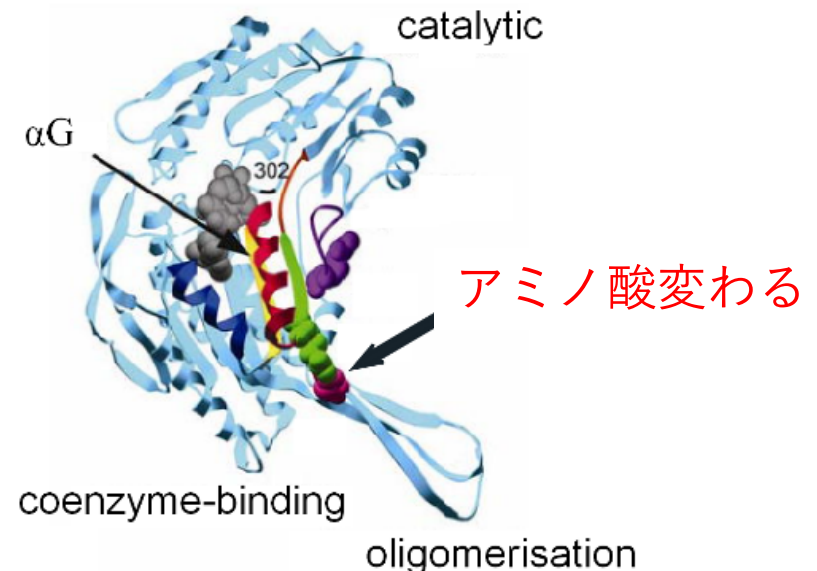
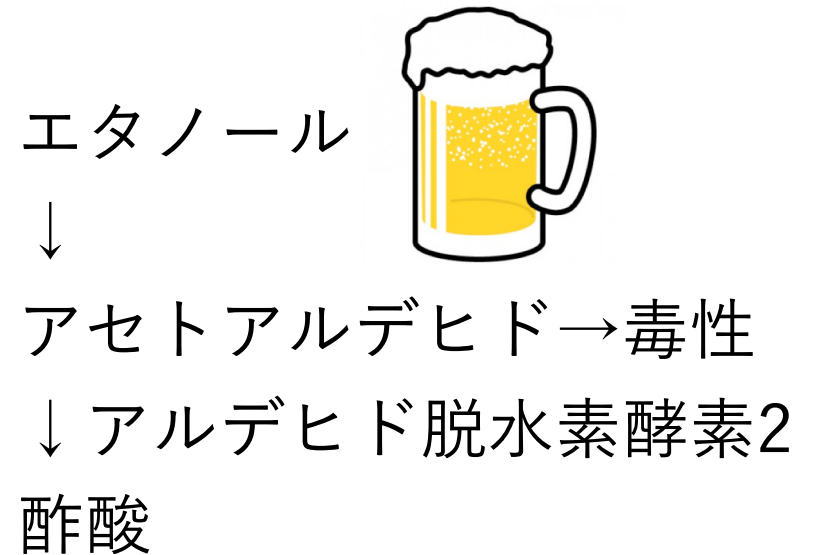
2. 薬剤標的探索

ヒトの疾患感受性遺伝子を見つけて病気を解明する

- 疾患感受性遺伝子とは
 - DNA変異により、その病気の罹り易さ（感受性）が変わる遺伝子
- 見つける意義
 - 病気の仕組みの解明
 - 創薬ターゲットの探索
 - 個人の発症予測・至適治療法の選択（Precision medicine）
- 疾患感受性遺伝子の探索法
 - 多因子の、ありふれた疾患
 - ゲノムワイド関連解析
 - 単一遺伝子による稀な疾患（メンデル型疾患）
 - 連鎖解析

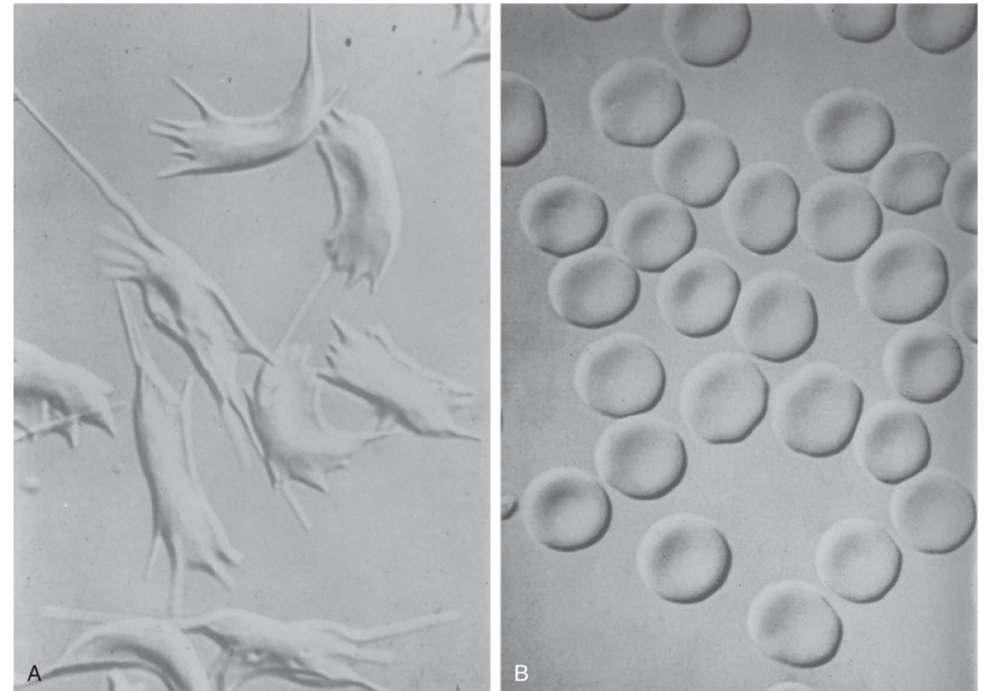
DNA変異と体質

- アルデヒド脱水素酵素2遺伝子
- DNA変異
 - c.1510G>A (p.Glu504Lys) rs671
 - . . . ACT **G**AA GTG . . .
 - . . . ACT **A**AA GTG . . .
- 遺伝型**GG**の人
 - お酒飲める
- 遺伝型**AG**の人
 - 酵素活性が1/16
 - お酒を飲むと赤くなる
- 遺伝型**AA**の人
 - 酵素活性ない
 - お酒が飲めない



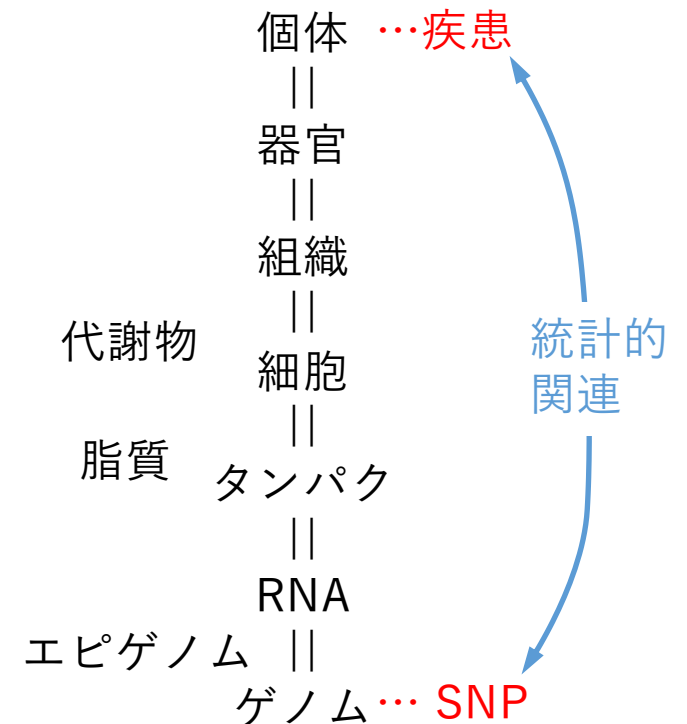
DNA変異と疾患

- ヘモグロビン β 遺伝子
- 変異
 - c.20A>T (p.Glu7Val) rs334
 - β^S アリル
- 変異ホモ接合
 - ヘモグロビンが凝集
 - 赤血球が鎌状になる
 - 赤血球が壊され貧血に
- 変異ヘテロ接合
 - 貧血ない
 - マラリア原虫に感染しにくい



関連解析のコンセプト

- 疾患と一塩基多型（SNP）に着目
 - 生体階層構造の両端に離れている
 - 統計的関連（相関）が、ヒトでの因果関係を示唆する
 - 中間は、ブラックボックスとしてよい
- 関連解析
 - 疾患と関連するSNPを見つける
 - 関連SNPの位置にある遺伝子が、疾患感受性遺伝子のはず
 - ゲノムワイドに関連SNPsを探索するのが、ゲノムワイド関連解析（GWAS）
 - モデル生物実験に展開、逆に検証



SNPと疾患の統計的関連

- i 番目の人の遺伝子型を

$$x_i = 0, 1, 2$$

- 例、SNPがG/Aのとき、
0 (GG), 1 (AG), 2 (AA)

- i 番目の人の血圧値を y_i

- 両者の統計的関連を線形回帰で検定

- 誤差 $\varepsilon_i \sim \text{Normal}(0, \sigma^2)$
- 尤度を最大化する $\hat{\alpha}, \hat{\beta}$ を求める
- 帰無仮説： $\beta = 0$

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

血
圧

SNPと
血圧の
「相関」

遺
伝
子
型

ゲノムワイド関連解析 (GWAS)

- 目標
 - ゲノムワイドに、ありふれた (頻度 $\geq 1\%$) SNPs全てについて疾患との関連を検定する
- 計測
 - 各被験者で、代表的なSNPsをマイクロアレイで測定し、残りのSNPsの情報は推測
- 統計解析
 - SNPsは 6×10^6 個あるが、染色体上で近傍のものは相関している (連鎖不平衡) ので、統計的に独立なものは正味 10^6 個
 - 約 10^6 回の多重検定を行うので、擬陽性を抑えるために、有意水準を $0.05/10^6 = 5 \times 10^{-8}$ と厳しくしないとイケない。ゲノムワイド有意とよぶ
 - 検出力を上げるためには、罹患者・健常者を数千～数万人調べる必要がある

ゲノムワイドに網羅的に調べる
↓
有意水準を厳しくする
↓
多数のサンプルが必要

GWASに必要なサンプルサイズ

- 小さな効果のSNPを検出するには、多数のサンプルが必要
- 有意水準 5×10^{-8} のもとで、検出力を 80% にするには、 $40/R^2$ 人必要
 - R^2 (決定係数) : 疾患・形質 y の分散のうち、SNP遺伝子型 x で説明される割合。これは相関係数の二乗。

SNPのSBPへの効果 [mm Hg]	R^2 [%]	必要なサンプルサイズ [人数]
1	0.13	32,000
0.5	0.06	64,000
0.25	0.03	128,000

- $F_{1,N-2}$ 分布の非心度パラメータが40になればよい
- SNPの効果 b 、アレルの頻度が p のとき、 $R^2 = 2 p (1-p) b^2$
- アレル頻度0.5、SBPの標準偏差を20 mmHgとした

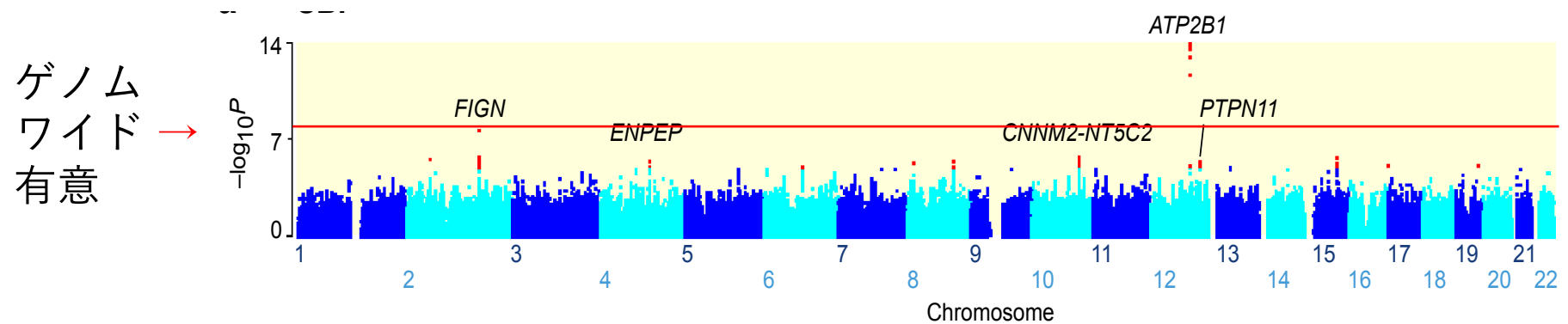
(高)血圧の大規模GWAS

Study	Publication	年	スクリーニング症例数 [万人]				追試症例数 [万人]	ゲノムワイド有意なSNPsの数	新規SNPsの数
			欧米	東アジア	南アジア	アフリカ			
WTCCC	Nature 447:661	2007	0.5					0	0
Global BPgen	Nat Genet 41:666	2009	3				11	8	8
CHARGE	Nat Genet 41:677	2009	3				3	8	8
AGEN-BP	Nat Genet 43:531	2011		2			3	10	5
ICBP	Nature 478:103	2011	7				13	29	16
COGENT	Am J Hum Genet 93:545	2013				3	10	5	3
iGEN-BP	Nat Genet 47:1282	2015	4	3	3		22	35	12
CHD Exome+, ExomeBP, GoT2D	Nat Genet 48:1151	2016	17		3		16	51	30
CHARGE+ Exome	Nat Genet 48:1162	2016	12			2	18	70	31
Cardio-Metabochip	Nat Genet 48:1171	2016	20				14	66	17
GERA	Nat Genet 49:54	2017	9	0.7		0.3		75	39
UK Biobank	Nat Genet 49:403	2017	14				19	107	32
AGEN-BP	Nat Commun 9:5052	2018		13			16	92	19
UK Biobank	Nat Genet 50:1412	2018	76				25	535	535
MVP	Nat Genet 51:51	2019	46				47		261

血圧関連遺伝子座は合計1016箇所

GWASで同定された血圧関連 遺伝子座 (1/4)

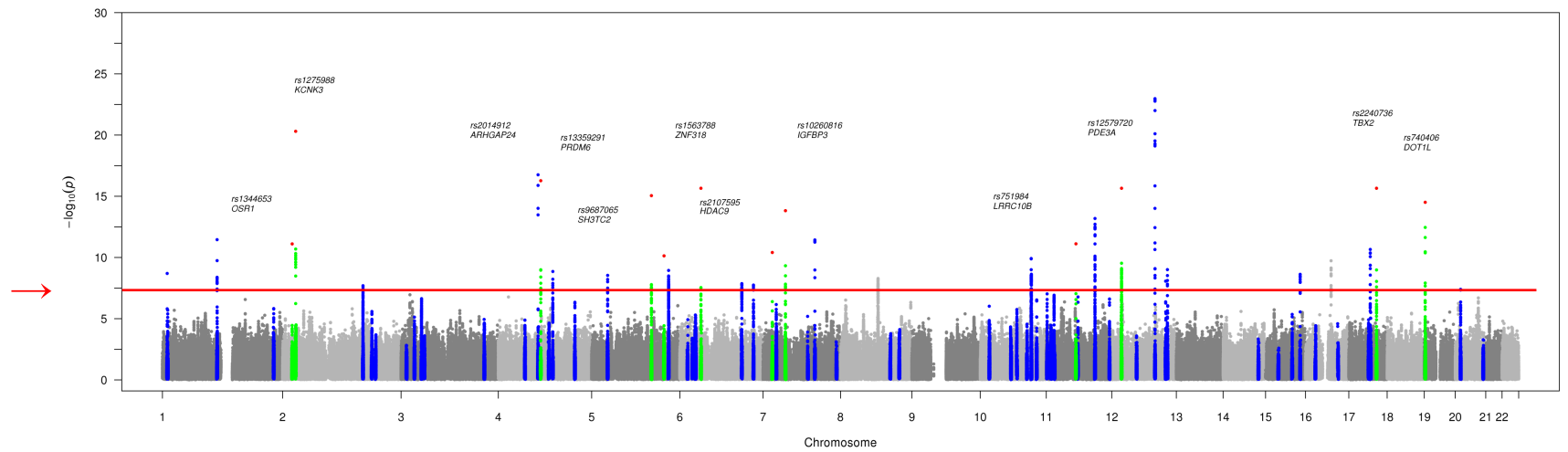
2万人のGWAS



GWASで同定された血圧関連 遺伝子座 (2/4)

10万人のGWAS

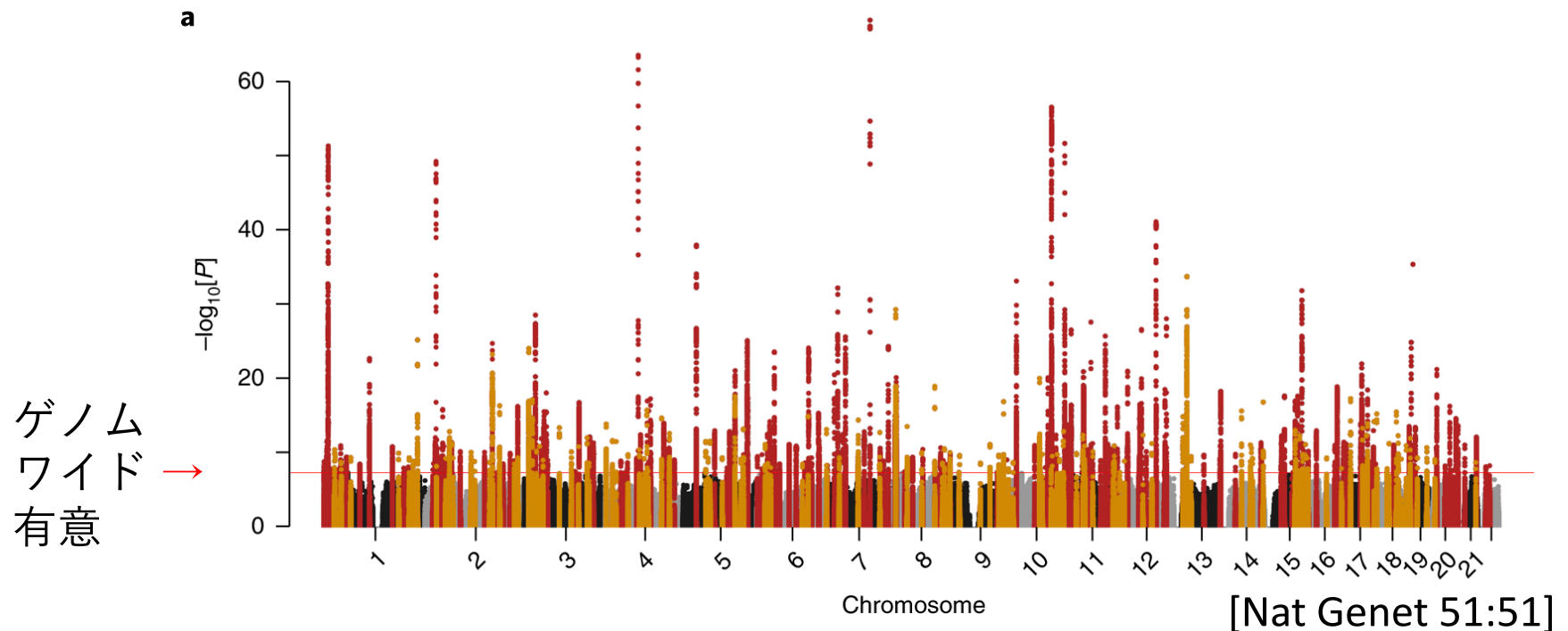
ゲノム
ワイド
有意



GWASで同定された血圧関連 遺伝子座 (3/4)

- >1000遺伝子座が血圧と有意に関連
- ゲノム全体に渡る

46~76万人のGWAS



GWASで同定された血圧関連 遺伝子座 (4/4)

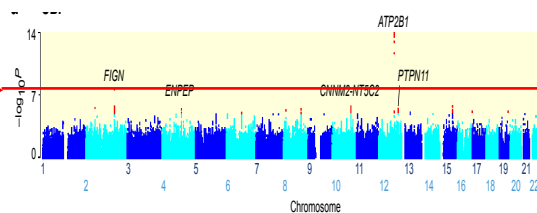
ビッグデータの恩恵！

2万人のGWAS

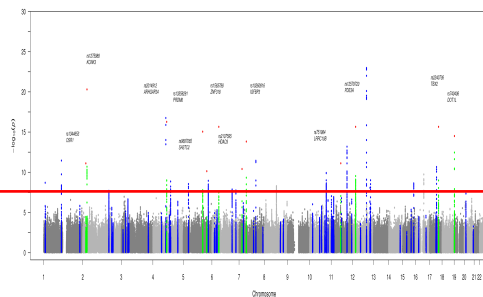
10万人のGWAS

46~76万人のGWAS

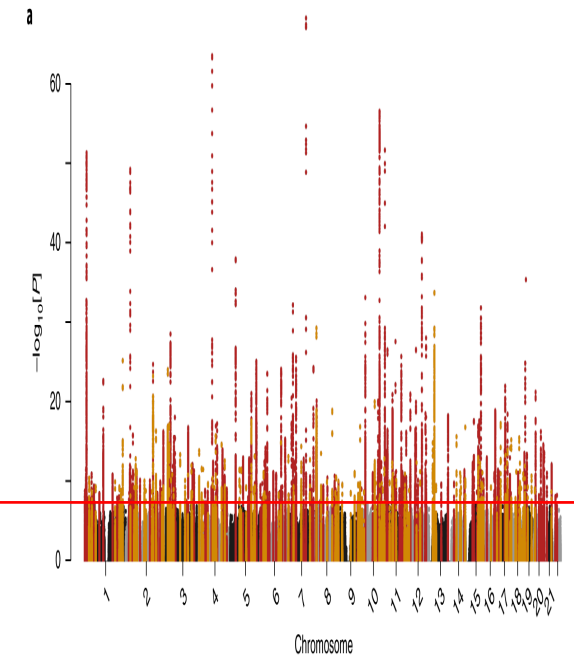
ゲノム
ワイド
有意



[Nat Genet 43:531]



[Nat Genet 47:11]



[Nat Genet 51:51]

GWASメタ解析

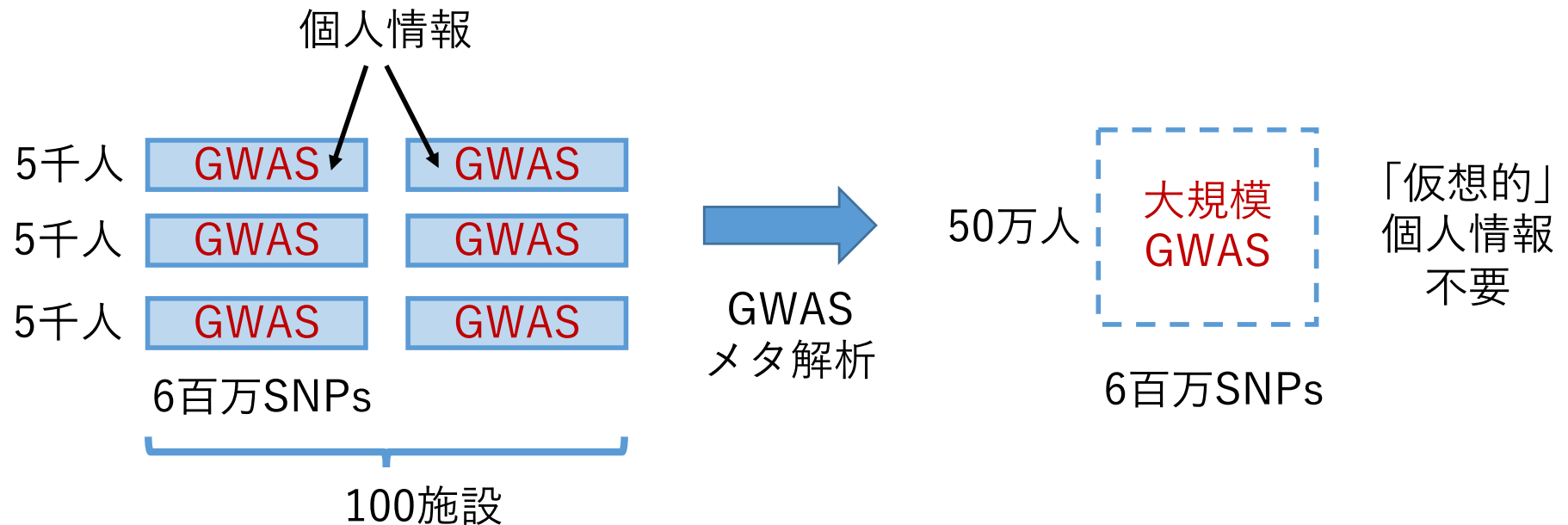
- 1施設のGWASではサンプル数に限りがある
 - ➔ 複数のGWASをメタ解析する
- 個別GWASで、SNPの効果を推測
 - i 番目の人のSNP遺伝子型を x_i
 - i 番目の人の血圧値を y_i
 - 血圧に対するSNPの効果 β を線形回帰で推定
- 複数GWASで推定された効果をメタ解析で統合
 - j 番目の研究における効果の推定値が β_j 、標準誤差が s_j
 - $1/s_j^2$ で重み付けした平均
 - 全体での効果の推定値 β 、標準誤差が s

$$\beta = \frac{\sum_j \frac{\beta_j}{s_j^2}}{\sum_j \frac{1}{s_j^2}}$$

$$s = \sqrt{\frac{1}{\sum_j \frac{1}{s_j^2}}}$$

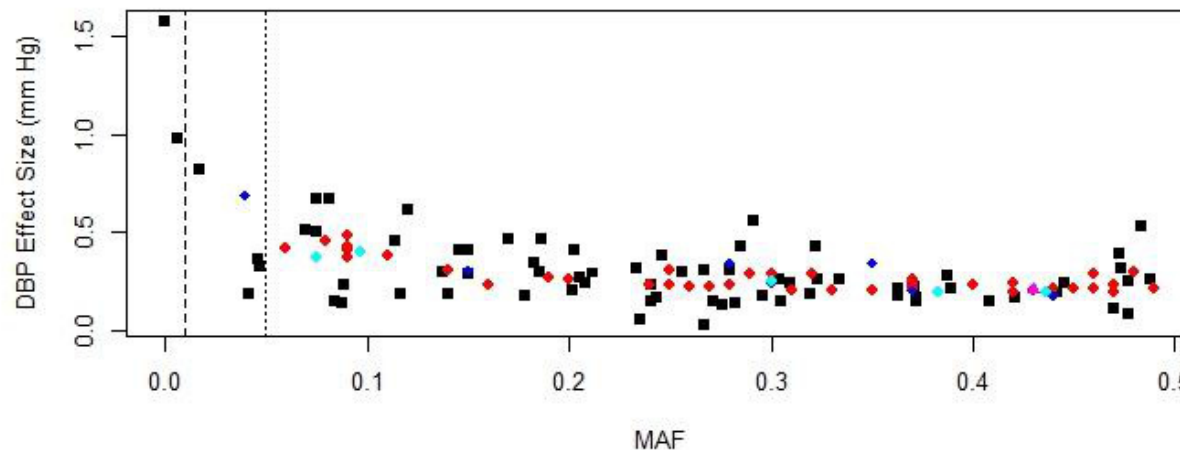
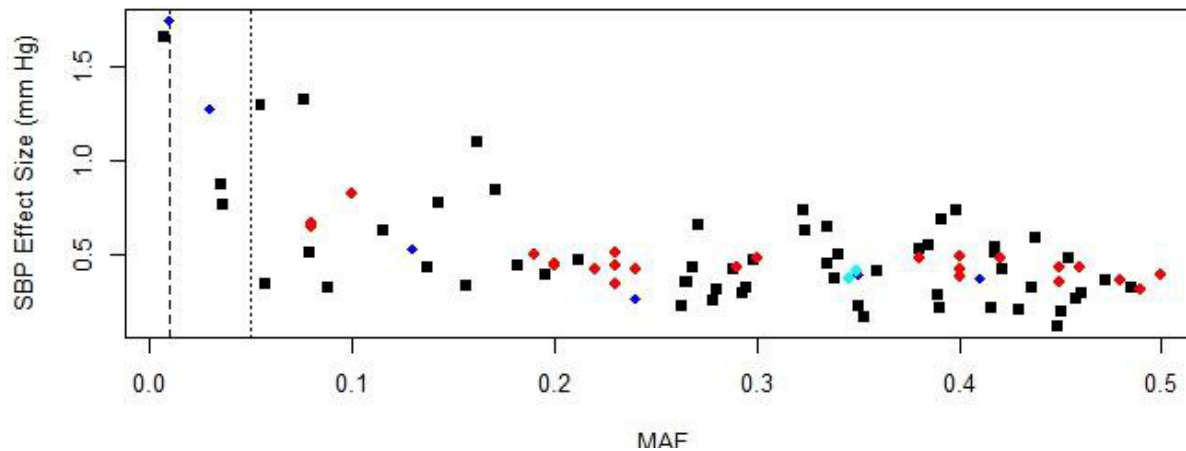
☆ビッグデータ解析としてのGWAS

- GWASメタ解析により、多施設の研究を統合し、ひとつの仮想的な大規模GWASにできる
- メタ解析では、個人情報（遺伝型、疾患）が不要
- 相関（線形な効果）のみを調べるので統合できる



GWAS有意なSNPsの効果は弱い

- SNPの効果は弱く、SBPで0.15~1 mmHg程度



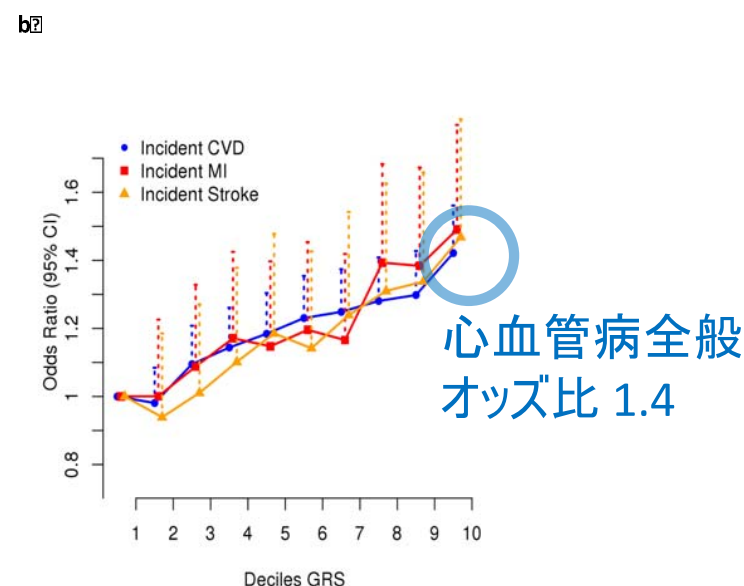
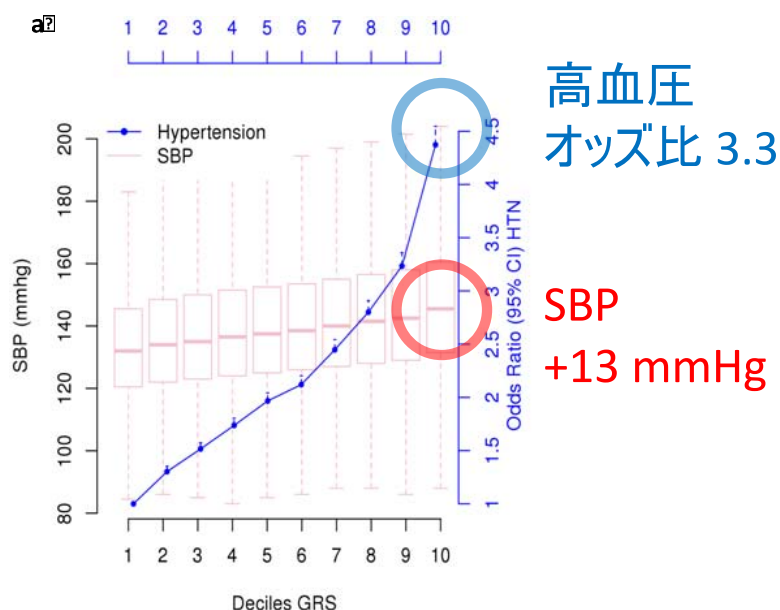
ヒト集団における血圧の遺伝的成因の全体 = (狭義の) 遺伝率、と考えると

- 遺伝率 = 形質分散を遺伝で説明できる割合
- 血縁者で推定した遺伝率
 - SBP 16% [PLoS Genet 2:e132]
 - 高血圧罹患 28% [Nat Genet 48:980]
- GWASでゲノムワイド有意だったSNPsの遺伝率
 - 個別SNPの効果は弱いですが、901関連遺伝子座で合わせてSBP 11%
- ゲノムワイド有意に達しないものも含めた、SNPs全体の遺伝率
 - SBP 16% [Nat Genet 50:746]
 - 少しでもSBPに影響するSNPsは、全ゲノムの5%程度を占める
 - 高血圧罹患 32% [Nat Genet 48:980]
- 本態性高血圧の遺伝的成因は、個々には効果の微弱な遺伝的多型が、相当多く組み合わさっている
- 稀な変異（頻度<1%）による成因は遺伝率としては僅か

遺伝的リスクスコア

- 各人のゲノムから疾患リスク（ここでは血圧予測）を計算
 - 血圧GWASで同定された901関連遺伝子座を使用
 - 個々のSNPsの効果は弱く、SBPで0.15~1 mmHg程度
- 遺伝的リスクスコアにより一般集団を10等分し、**スコア最大と最小のグループを比較**

UK Biobank
29万人
[Nat Genet
50:1412]



- 顕著な効果とは言えないが、将来的に高血圧の発症予測・予防への応用が期待される

遺伝的リスクスコアの臨床応用

- ゲノムワイド多遺伝子リスクスコア (GPS)
 - ゲノムワイドに全多型を用いる
 - GWASで観測された効果と連鎖不平衡で重み付け
- UK Biobank [Nat Genet 50:1219]
 - 29万人で検証
 - GPS検査により、罹患オッズ比>3となる人の割合を計算
- 多くのありふれた疾患について、GPSの臨床応用は技術的には可能になってきた

疾患	オッズ比 >3の人
冠動脈疾患	8.0%
心房細動	6.1%
2型糖尿病	3.5%
炎症性腸疾患	3.2%
乳癌	1.5%
5疾患のいずれか	19.8%

←家族性高コレステロール血症 (0.4%) より多い

editorial

Nat Genet
2018年9月号
論説

GPS for navigating healthcare

Precision genomic medicine is now technically feasible. Just as global positioning systems revolutionized the logistics of travel, so genome-wide polygenic risk scores (GPSs) now have the potential to inform our trajectories of health and to serve in the prevention and mitigation of many common and complex diseases. We welcome research into the implementation of—and equity of access to—genetic predictors and their integration into clinical and evidence-based medical practice.

前半のまとめ

- ゲノムワイド関連解析（GWAS）では、ヒト集団での、疾患等の表現型と一塩基多型（SNPs）の相関をゲノムワイドに検定することにより、表現型に影響する遺伝子座を探索する。
- 血圧関連遺伝子座は累計>1000箇所。
- ヒト高血圧の遺伝的成因は、個々には効果の微弱な遺伝的多型が、相当多く組み合わせられている。
- GWASで見つかったSNPsは個々の効果が弱いですが、それらを組み合わせたりリスクスコア(GPS)により、precision medicineとしての発症予測・予防への応用が期待される。



血圧の

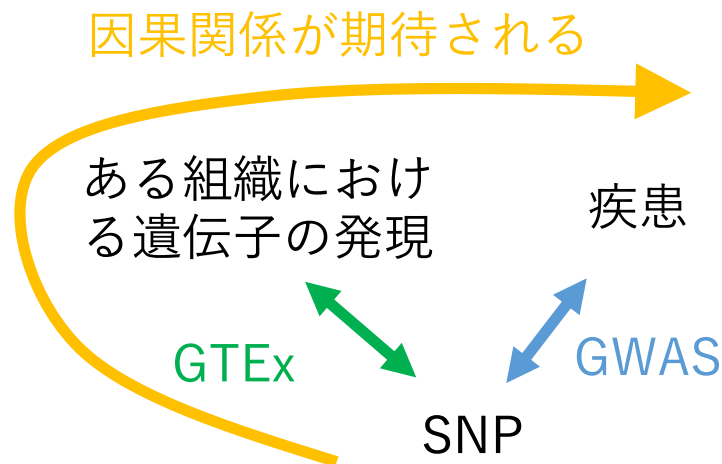
1. ゲノムワイド関連解析 (GWAS)

- 遺伝的成因の全体像
- 発症予測

2. 薬剤標的探索

GWAS SNPから遺伝子へ

- 背景
 - (ある組織で)遺伝子発現が疾患と関連する**遺伝子**を見つけない
 - でも、GWAS被験者では、遺伝子発現は測定していない
 - SNP遺伝子型に基づいて、統計的に推測する
- GTEx (Genotype-Tissue Expression) 参照データ [Nature 550:204]
 - ドナー948名の54組織の全遺伝子発現、およびSNP遺伝子型
 - 腎臓は少なく73名 (しかも組織バルク)
- トランスクリプトームワイド関連解析 (TWAS)
 - GWASデータとGTExデータを統合
 - 関連するか?
 - 疾患 (の遺伝的因子)
 - ある組織における遺伝子の発現 (の遺伝的因子)



UK BiobankのS-PrediXcan解析

https://imlab.shinyapps.io/gene2pheno_ukb_neale

SBPで検索すると

	gene_name	zscore	effect_size	pval	tissue
1	MTHFR	-13.10	-0.20	3.2e-39	TW_Muscle_Skeletal_Elastic_Net_0.5
2	TMEM133	-12.72	-0.05	4.6e-37	TW_Artery_Tibial_Elastic_Net_0.5
3	CLCN6	-12.46	-0.12	1.3e-35	TW_Artery_Tibial_Elastic_Net_0.5
4	NPR3	12.05	0.10	2.0e-33	TW_Lung_Elastic_Net_0.5
5	NPR3	11.93	0.05	8.5e-33	TW_Nerve_Tibial_Elastic_Net_0.5
6	TMEM133	-11.92	-0.05	9.3e-33	TW_Artery_Coronary_Elastic_Net_0.5
7	FES	-11.67	-0.05	1.8e-31	TW_Cells_Transformed_fibroblasts_Elastic_Net_0.5
8	CLCN6	-11.54	-0.04	8.2e-31	DGN_WB_Elastic_Net_0.5
9	CCDC71L	11.47	0.07	1.9e-30	TW_Artery_Aorta_Elastic_Net_0.5
10	FURIN	11.42	0.23	3.5e-30	TW_Artery_Aorta_Elastic_Net_0.5
11	ARHGAP42	-11.41	-0.04	3.7e-30	TW_Artery_Tibial_Elastic_Net_0.5
12	FES	-11.22	-0.11	3.4e-29	TW_Breast_Mammary_Tissue_Elastic_Net_0.5
13	FURIN	11.17	0.08	5.9e-29	TW_Esophagus_Mucosa_Elastic_Net_0.5
14	ATP2B1	-11.17	-0.13	6.0e-29	TW_Artery_Tibial_Elastic_Net_0.5
15	MTHFR	-11.00	-0.07	3.9e-28	TW_Cells_EBV-transformed_lymphocytes_Elastic_Net_0.5
16	FES	-10.78	-0.08	4.1e-27	TW_Artery_Aorta_Elastic_Net_0.5
17	NPR3	10.56	0.06	4.6e-26	TW_Testis_Elastic_Net_0.5
18	FES	-10.48	-0.12	1.0e-25	TW_Nerve_Tibial_Elastic_Net_0.5
19	FES	-10.47	-0.10	1.1e-25	TW_Artery_Tibial_Elastic_Net_0.5
20	FES	-10.47	-0.07	1.1e-25	TW_Adipose_Visceral_Omentum_Elastic_Net_0.5

類似法に

<http://twas-hub.org>

GWASヒット遺伝子のどれが薬剤標的として有望か

- 薬剤標的となる遺伝子
 - 単一遺伝子で血圧(のみ)に顕著な効果が欲しい
- 血圧GWASでゲノムワイド有意だったもの
 - 1000弱の関連**遺伝子**(座)
 - SNPの効果は弱く、SBPで0.15~1 mmHg程度
 - SNPの頻度は >5%
 - **SNP**は進化の過程で残ってきた→効果が弱いのは当然
- 薬剤標的として有望性
 - GWAS SNPの効果 ≠ その遺伝子の効果
 - 他の指標が必要
 - 血圧値が極端な人が、その遺伝子に稀な変異を持つか
 - パスウェイ解析、遺伝子機能解析、モデル動物実験

LDL-CのGWASと薬剤標的遺伝子

Nearby genes	Best SNP	Effect size (SE) mg/dL	P-value	
<i>SORT1</i>	rs629301	-5.65 (0.21)	9.7 E-	171
<i>APOE-C1-C2</i>	rs4420638	7.14 (0.29)	8.7 E-	147
<i>LDLR</i>	rs6511720	-6.99 (0.30)	4.3 E-	117
<i>APOB</i>	rs1367117	4.05 (0.19)	4.5 E-	114
<i>ABCG5/8</i>	rs4299376	2.75 (0.20)	1.7 E-	47
<i>HMGCR</i>	rs12916	2.45 (0.18)	5.1 E-	45
<i>TRIB1</i>	rs2954022	-1.84 (0.17)	2.6 E-	29
<i>PCSK9</i>	rs2479409	2.01 (0.22)	1.9 E-	28
<i>POA1-C3-A4-A5</i>	rs964184	2.85 (0.27)	1.5 E-	26
<i>TIMD4</i>	rs6882076	-1.67 (0.19)	1.9 E-	22
<i>ABO</i>	rs649129	2.05 (0.21)	7.9 E-	22
<i>HPR</i>	rs2000999	2.00 (0.22)	1.8 E-	22
<i>CILP2</i>	rs10401969	-3.11 (0.38)	6.7 E-	22
<i>FADS1-2-3</i>	rs174583	-1.71 (0.19)	1.2 E-	21
<i>TOP1</i>	rs909802	1.41 (0.17)	3.2 E-	19

スタチン

PCSK9阻害薬

Nature
466:707,
Table S2

GWAS SNPのP値が小さい (=ヒト集団での多様性に大きく寄与)

≠ 薬剤標的遺伝子として有効

Omnigenic 仮説

omni- [接頭語] 全ての

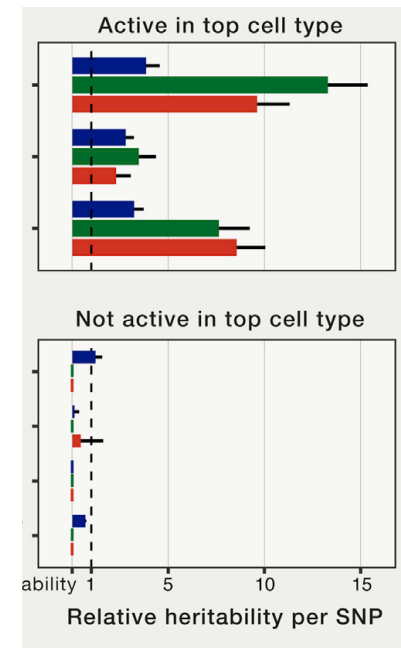
Boyle, Li & Pritchard [Cell (2017) 169:1177]

- なぜこんなに多数の関連遺伝子座があるのだろうか？
- 複雑疾患の遺伝率への寄与は、むしろBが大きいのでは
 - A) その疾患の「コアパスウェイ」の遺伝子
 - B) 疾患に関与する細胞(e.g. 糖尿病のβ細胞)で発現している全遺伝子。Aとネットワークでつながるのでは
- どうやってコア遺伝子を見分けるか
 - 低頻度で効果の強い(e.g. >1 mmHg ?) 変異の探索

クロマチン状態で染色体を領域分け

- 疾患組織のみ活性化
- 中間
- 遍く活性化
- 疾患組織では不活性
- 中間
- 遍く不活性

疾患への寄与の強さ
(遺伝率のenrichment)

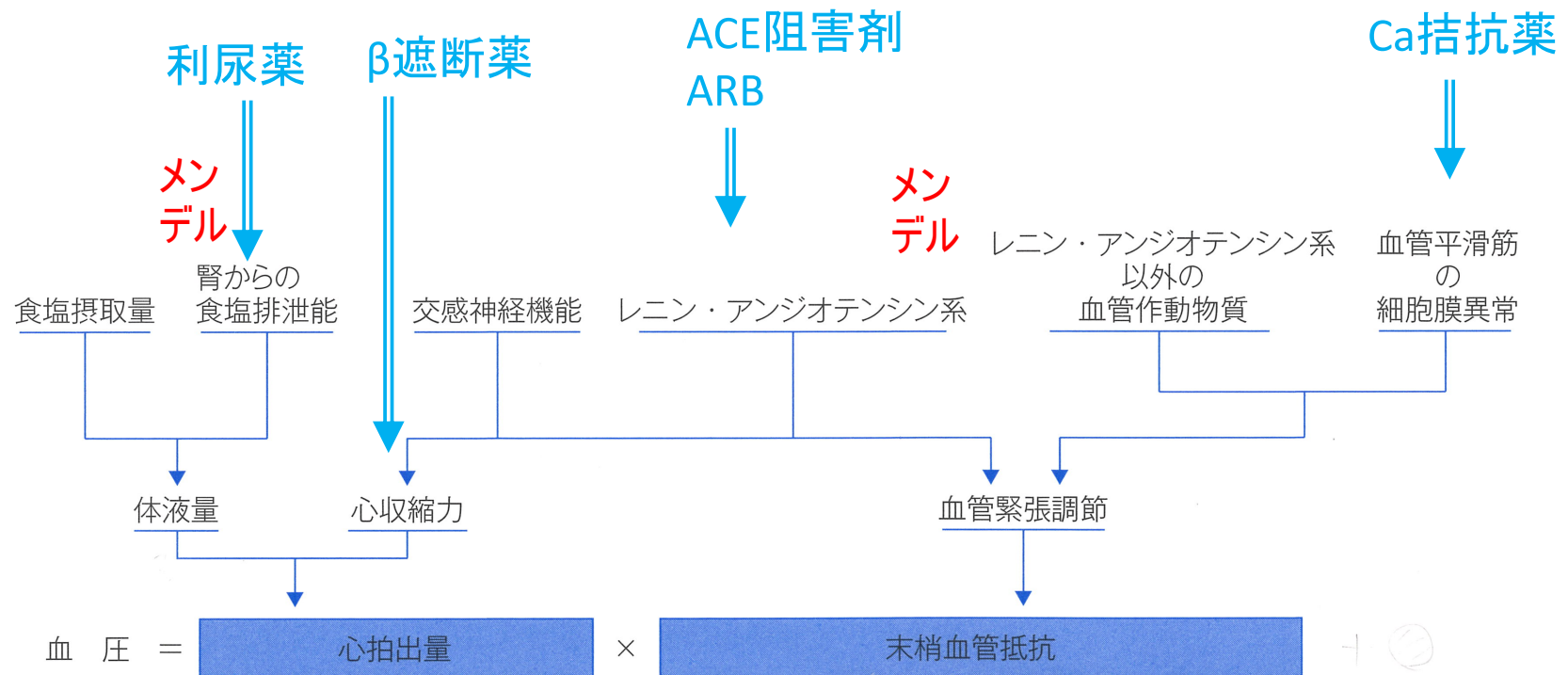


←
←
どちらも強く寄与

- クローン病(免疫)
- リューマチ(免疫)
- 統合失調症(中枢神経)

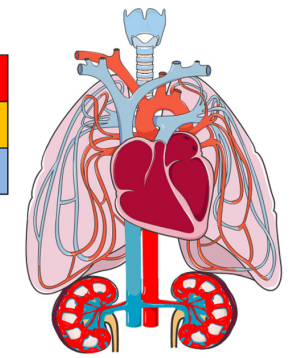
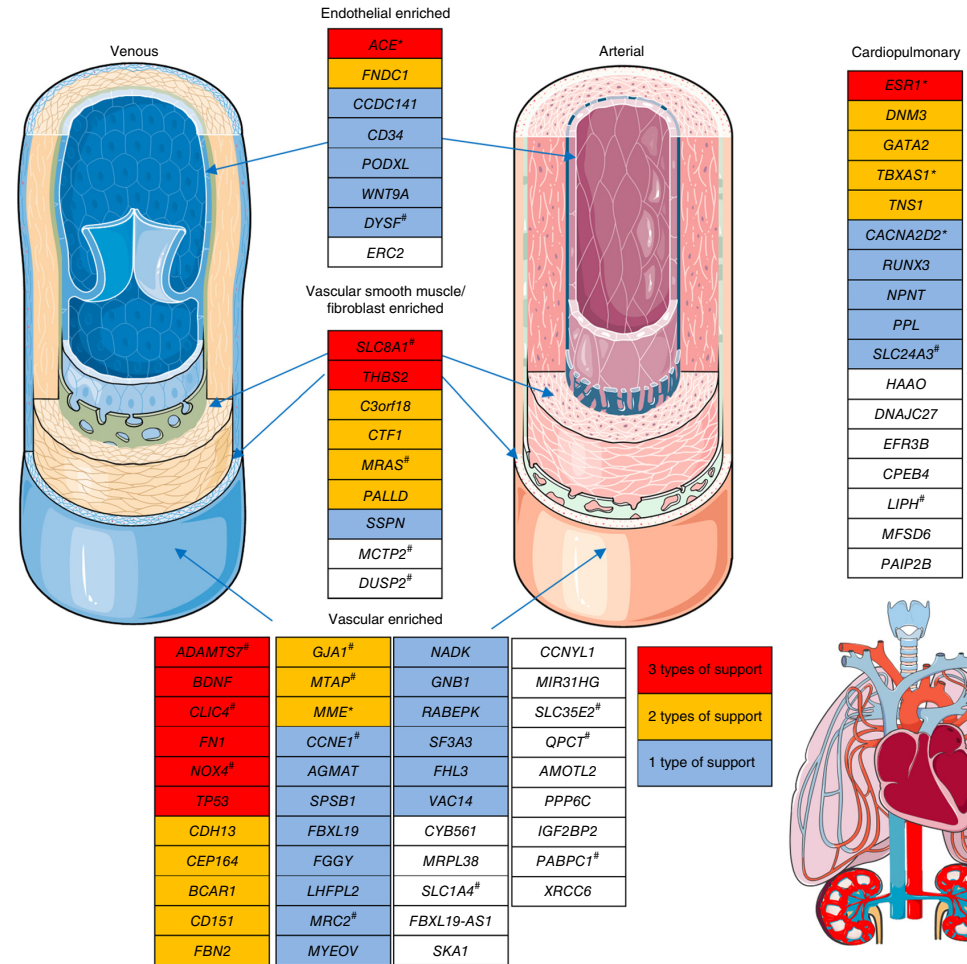
メンデル型高血圧の原因遺伝子

- 腎臓、副腎で機能するものが多い
 - Naチャンネル、Kチャンネル、Clチャンネル
 - アルドステロン合成酵素、受容体
- 降圧剤標的となる、血圧制御因子の一部に限定



血圧GWASで見つかった遺伝子が発現している組織

- 2017年のUK Biobank 14万人スクリーニングで新規に見つかった107遺伝子座 (212遺伝子)
- FANTOM5データで発現している組織
 - 血管 (59遺伝子)
 - 心・肺・腎・大動脈 (17遺伝子)



後半のまとめ

- GWASデータとGTExデータと統合することにより、どの遺伝子・組織が重要か示唆される。
- GWASで見つかった遺伝子の一部は薬剤標的遺伝子になりうるが、どれが有望かはGWASだけからは分からず、他の手法が必要。
- 疾患感受性遺伝子が多数あるのは、コア遺伝子に加え、疾患関連組織で発現している全遺伝子が寄与しているからかもしれない(Omnigenic仮説)