

疾患遺伝子探索のための遺伝統計解析: ゲノムワイド関連解析とその後

竹内史比古(たけうち ふみひこ)
国立感染症研究所

2011年7月1日

@東大・柏・GCOE談話会

疾患感受性遺伝子とは？

- * 遺伝子の型により、病気の罹り易さ(感受性)が変わる遺伝子
 - * dystrophin (*DMD*; Xp21.1)
 - * 変異による短いタンパク質 → Duchenne 型筋ジストロフィー
 - * hemoglobin, beta (*HBB*; 11p15.5)
 - * Glu6Val → 鎌状赤血球貧血症、でもマラリアには罹らない
 - * ABO (9q34.2)
 - * O型 → 静脈血栓塞栓症に罹りにくい (OR=0.33)
- * 疾患感受性遺伝子を見つける意義
 - * 病気の仕組みの解明
 - * 創薬のターゲットになる
 - * 個人の発症予測・至適治療法の選択(個別化医療)
- * 疾患感受性遺伝子がそもそも存在するか？
 - * 疾患感受性の素因は、遺伝と環境
 - * 家族集積性から遺伝が占める割合(遺伝率)が分かる
 - * 糖尿病 0.5
 - * 身長 0.8

遺伝統計学を用いた疾患感受性 遺伝子探索

- * 3万個の遺伝子を網羅的に調べ、疾患感受性遺伝子を統計的に絞り込む
 - * 全ての遺伝子型について、疾患との相関を調べる
 - * 罹患者はA型が多い、健常者はB型が多い
 - * 統計的有意に相関しているなら、感受性遺伝子のはず
 - * 遺伝子型と疾患の間は、ブラックボックスとしてよい
- * 全ゲノム配列決定は、多数の人ではまだできない
 - ゲノムワイド関連解析が有効



ゲノムワイド関連解析 (GWAS)

- * 目標: ありふれた (頻度 $\geq 5\%$) の一塩基多型 (SNP) の全てについて、疾患との関連を検定する
- * 染色体上で近傍のSNPsは相関しており (連鎖不平衡)、冗長なもの ($r^2 > 0.8$) を省いて、 10^6 SNPs をタイピングできるマイクロアレイが市販されている (数万円)
 - * ヒトゲノム配列決定、dbSNP、HapMap、アレイ技術により実現
- * 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 $0.05/10^6 = 5 \times 10^{-8}$ と厳しくしないといけない
- * 検出力を上げるためには、罹患者・健常者を数千人タイピングする必要がある

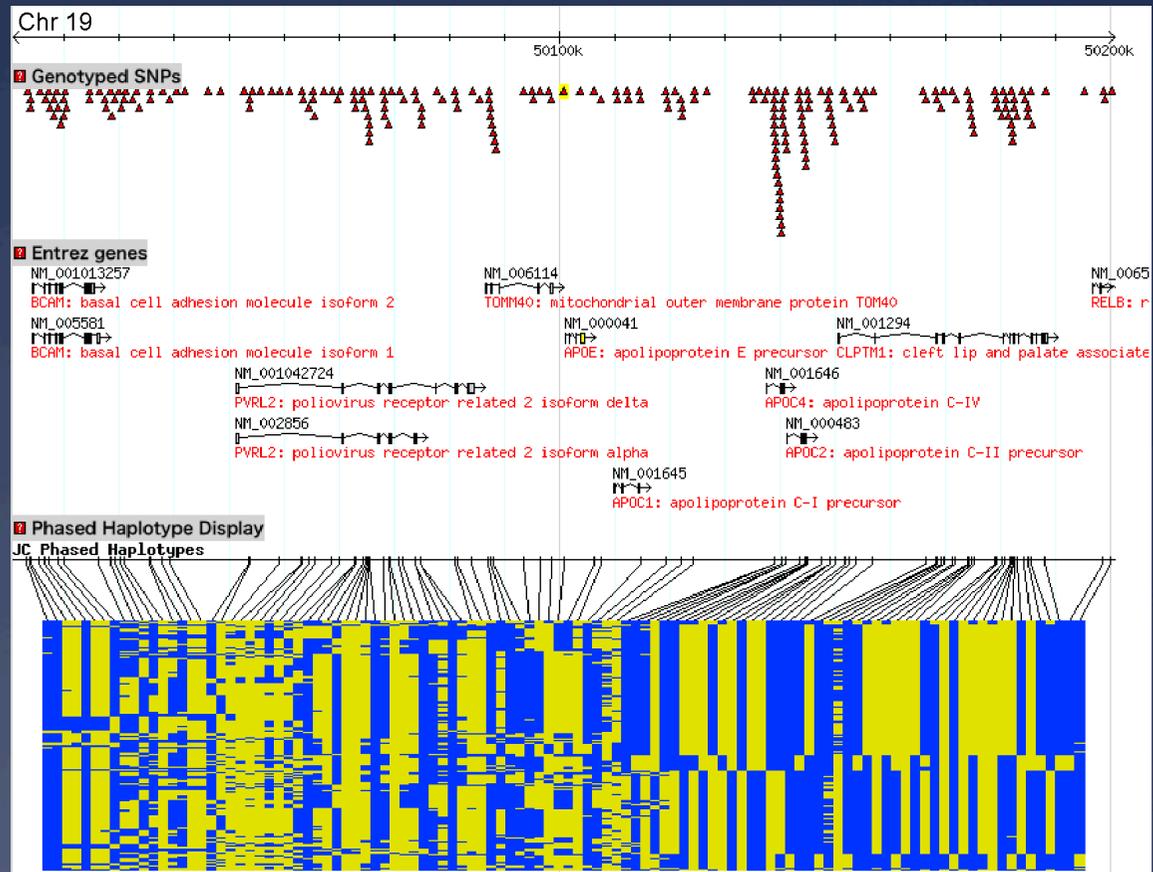
全ゲノムは不要だが、 10^6 SNPs は調べる

↓
有意水準を厳しくする

↓
多数のサンプルが必要

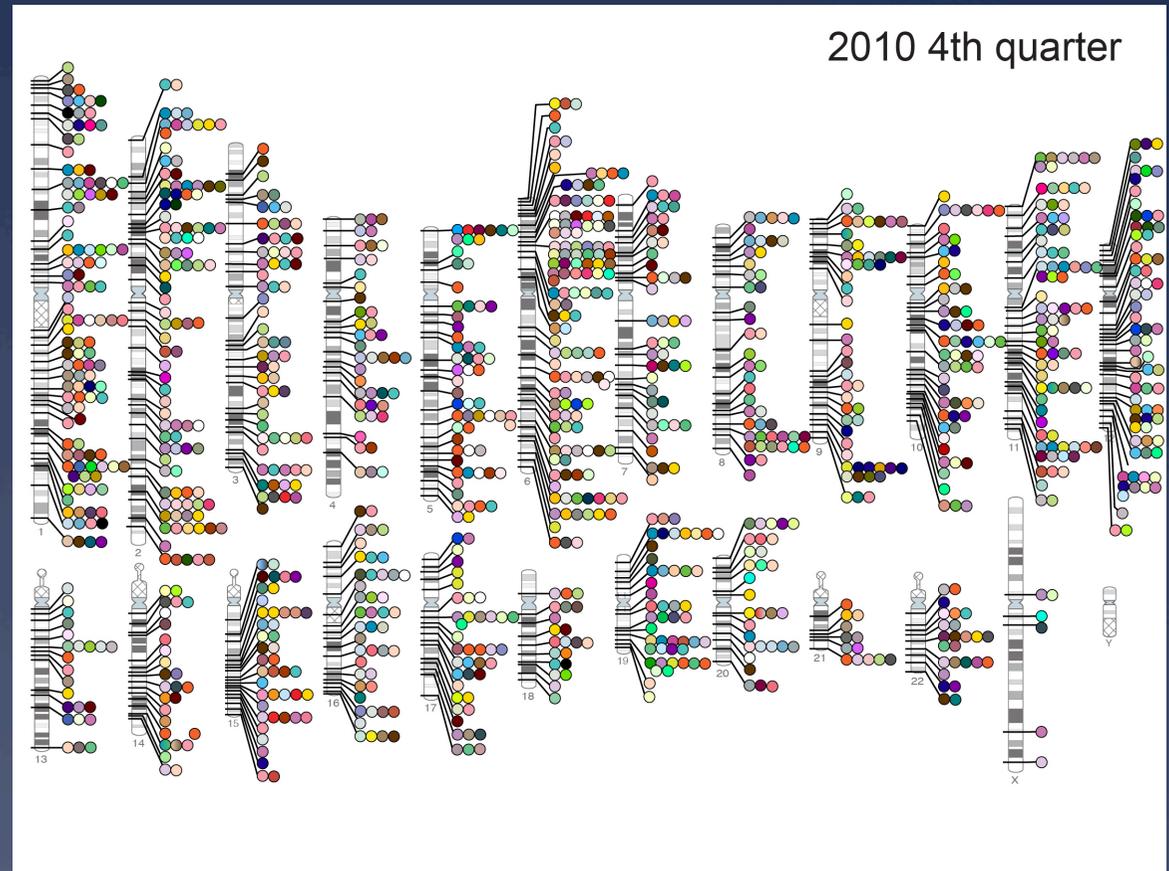
SNPsの相関(連鎖不平衡)

- * 染色体19番の200kbの領域中の108 SNPs
- * 日本人45人(染色体90本)における遺伝子型



GWASの成功

- * 2007年頃から普及
- * 数百の形質と関連する1913 SNPsが報告されている
- * <http://www.genome.gov/gwastudies>



1. ゲノムワイド関連解析 (GWAS)
 - * 糖尿病; 人種差
 - * 検出力の限界
2. GWASのメタ解析
 - * 解像度の限界
3. GWASで見つかった染色体領域の精密マッピング

いずれも私が行ってきた研究を中心に...

糖尿病のGWAS(方法)

目的1. 日本人における感受性遺伝子の網羅的探索

方法1. Hap550 (illumina) を用いたゲノムワイド関連解析

1次: 482,625 SNPs、罹患519人 vs 健常503人+一般964人(GeMDBJ)

↓ $P < 0.0025$

2次: 1,456 SNPs、罹患1,110人 vs 健常1,014人

↓ $P < 7 \times 10^{-5}$, 1+2次パネル

3次: 17 SNPs、罹患4,000人(BioBank) vs 健常4,889人

↓ $P < 0.003$

4 SNPs

目的2. 欧米人と日本人の遺伝素因の比較

方法2. 既報17 SNPsを、上記パネル(罹患 1,629~5,629人 vs 健常 1,517~6,406人)で追試。
さらに、日本人の既存研究とメタ解析を行った。

Mori et al. (2001) *Diabetes* 50:891

Hayashi et al. (2007) *Diabetologia* 50:980; Horikoshi et al. (2007) *Diabetologia* 50:747; Miyake et al. (2008) *J Hum Genet* 53:174

Omori et al. (2008) *Diabetes* 57:791; Horikoshi et al. (2007) *Diabetologia* 50:2461; Horikawa et al. (2008) *J Clin Endocrinol Metab* 93:3136

Yasuda et al. (2008) *Nat Genet* 40:1092; Unoki et al. (2008) *Nat Genet* 40:1098

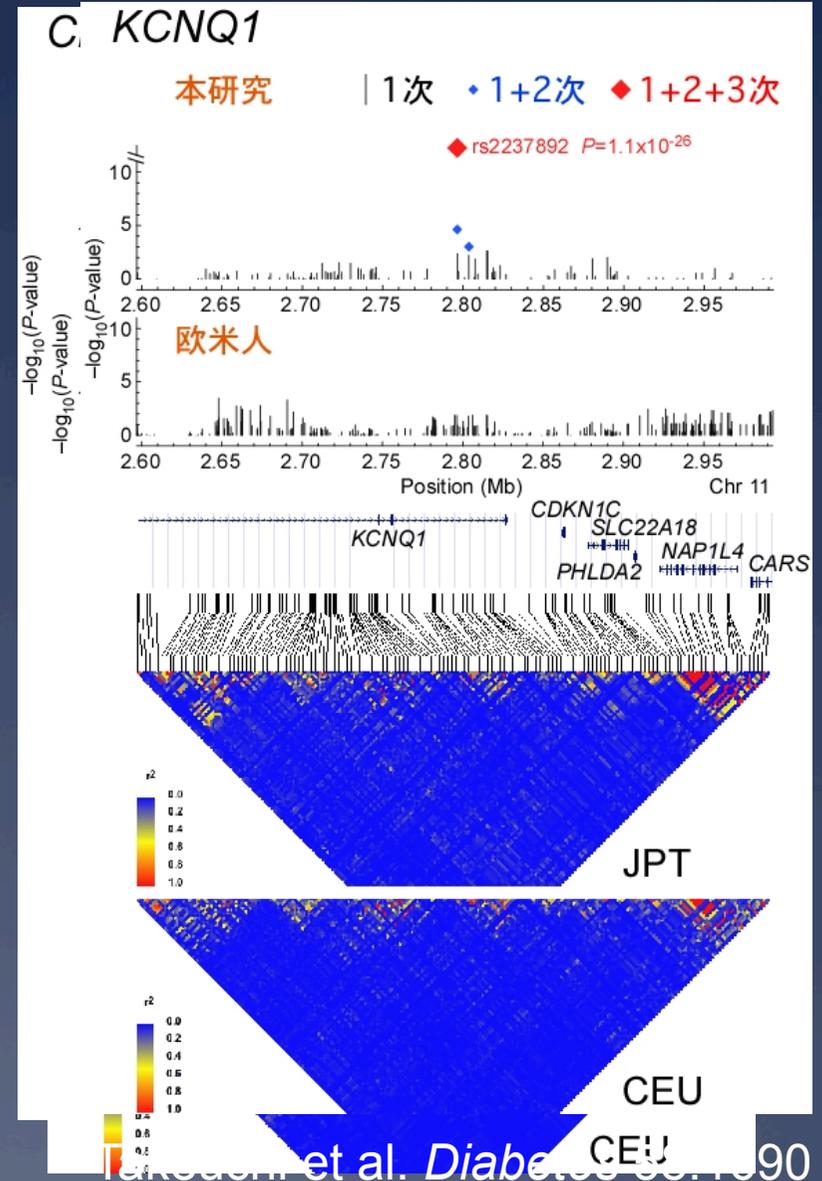
Takeuchi et al. *Diabetes* 58:1690

結果1. ゲノムワイド関連解析

* 4 SNPs がスクリーニングで見出された

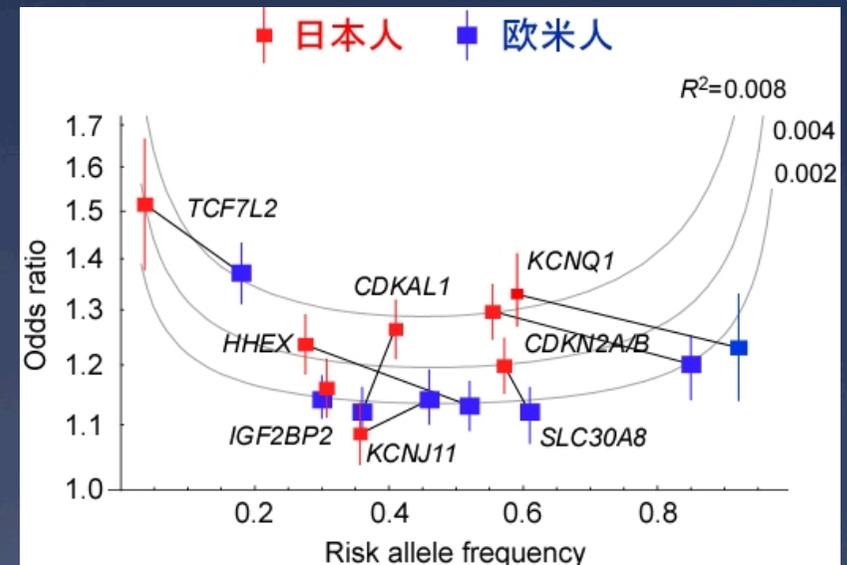
SNP	遺伝子領域	リスクアレル頻度	OR	P値
rs4712523	<i>CDKAL1</i>	0.41	1.27	7.2E-20
rs2383208	<i>CDKN2A/B</i>	0.55	1.34	2.1E-29
rs2237892	<i>KCNQ1</i>	0.59	1.33	1.1E-26
rs10425678	<i>PEPD</i>	0.26	1.14	1.4E-05

- * *CDKAL1*, *CDKN2A/B* は欧米人で既報
- * *KCNQ1* は日本人で既報
- * *PEPD* (peptidase D) のみ新規。ただし、ゲノムワイドな有意水準に達せず、さらなる検証が必要。



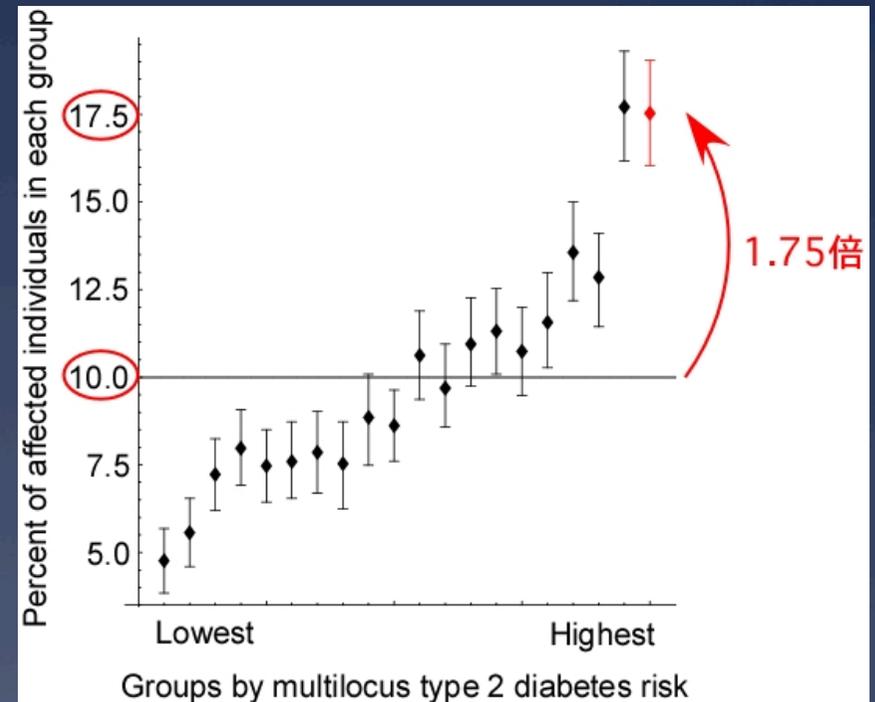
結果2. 関連の強さの人種間比較

- * 日本人でメタ解析を行い、ORとアリル頻度を欧米人と比較した
- * ORが大きく、アリル頻度が0.5に近いSNPは関連が強い
 - * 日本人では *CDKAL1*, *CDKN2A/B*, *KCNQ1*
 - * 欧米人では *TCF7L2*
- * 日本人の方がORが大きい傾向がある:
CDKAL1, *CDKN2A/B*, *HHEX*, *SLC30A8* で有意に異なる
- * 遺伝的浮動により、人種間でアリル頻度は変化している
 - 0 or 1 に近いと、関連が弱くなる
 - * 欧米人での *KCNQ1*
 - * 日本人での *TCF7L2*
- * 両人種で再現性の高い7 SNPs,
KCNQ1



結果3. 複数SNPsの組合せ的効果

- * 個々の OR 1.15~1.3 のSNPsでも、組み合わせれば、糖尿病リスクがより正確に評価できる
- * 全サンプルを用いて、有病率10%の一般集団をシミュレーションで作成
- * 両人種で再現性の高い7 SNPsについて
- * 「リスクアレルを幾つ持つか」で各人を並べ、20階層に分類した。
- * リスクが最大の群(5%)では、有病率が平均の1.75倍
- * しかし7 SNPsで説明できる割合はまだ小さい
 - * 分散の3%
 - * ROC曲線のAUCが0.60
- * *CDKAL1, CDKN2A/B, IGF2BP2, SLC30A8, HHEX, TCF7L2, KCNJ11*



SNPの関連の検定

- * i 番目の人のSNP 遺伝子型を $x_i = 0, 1, 2$
 - * 例、アレルがA/Cのとき、0 (CC), 1 (AC), 2 (AA)

- * 連続形質との関連の検定

- * i 番目の人の形質の値を y_i (例、血糖値)
- * 線形回帰
 - * 誤差 $\varepsilon_i \sim \text{Normal}(0, \sigma^2)$
 - * 帰無仮説: $\beta = 0$

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

- * 疾患との関連の検定

- * i 番目の人の表現型を $y_i = 1$ (罹患), 0 (健常)
- * ロジスティック回帰
 - * $y_i \sim \text{Binomial}(p_i)$
 - * 帰無仮説: $\beta = 0$

$$\log \frac{p_i}{1 - p_i} = \alpha + \beta x_i$$

- * 尤度を最大化する $\hat{\alpha}, \hat{\beta}$ を求める

SNPの関連の検出力

- * y の分散は、 x で説明される部分 (S_R) と残差平方和 (S_E) に分解できる

$$\begin{aligned}\sum_{i=1}^N (y_i - \bar{y})^2 &= \sum_{i=1}^N (\hat{\alpha} + \hat{\beta}x_i - \bar{y})^2 + \sum_{i=1}^N (y_i - \hat{\alpha} - \hat{\beta}x_i)^2 \\ &= S_R + S_E\end{aligned}$$

- * 検定に用いる統計量 $S_R / \{S_E / (N-2)\}$ は
 - * 関連が無いとき (帰無仮説) は $F_{1, N-2}$ 分布に従う
 - * 関連が有るとき (対立仮説) は **非心度パラメータ $N R^2 / \{1 - R^2\}$** の $F_{1, N-2}$ 分布に従う
 - * 連続形質 y の分散のうち、SNP 遺伝子型 x で説明される割合を R^2 とする (決定係数)。
 - * N はサンプルの人数
 - * **有意水準 5×10^{-8} のもとで、検出力が 80% となるのは、非心度パラメータが約 40 のとき**
 - * $R^2 = 0.1$ なら $N = 360$
 - * $R^2 = 0.01$ なら $N = 4000$ (例、日本人での糖尿病に対する *KCNQ1*)
 - * $R^2 = 0.005$ なら $N = 8000$ (例、同じく *CDKAL1*)
 - * $R^2 = 0.001$ なら $N = 40000$
 - * ざっくり $N \doteq 40 / R^2$
- 弱い関連を検出するには多数のサンプルが必要

第1部のまとめ： GWASの成果、残る課題

* まとめ

- * 日本人での2型糖尿病のゲノムワイド関連解析により、感受性遺伝子領域 *CDKAL1*, *CDKN2A/B*, *KCNQ1*, *PEPD* を同定した
- * 欧米人で既報の関連の多くが、日本人でも再現、あるいは似た傾向を示した
- * 両人種に共通した疾患感受性SNPsでは、日本人のほうがORが強い傾向がある

* 関連解析として、残る課題

- * まだ検出できていない関連がある
 - * 数千人の規模のGWASが多数行われたが、同定できたSNPsは遺伝的因子の数パーセントしか説明できていない。
- * 未同定の多型は関連が弱いはず
→ゲノムワイドメタ解析でサンプル数を増やす

1. ゲノムワイド関連解析 (GWAS)。私の行ってきたもの
 - * 検出力の限界
2. GWASのメタ解析
 - * 解像度の限界
3. GWASで見つかった染色体領域の精密マッピング

メタ解析による複数研究の統合

* 線形回帰

- * i 番目の人のSNP 遺伝子型を $x_i = 0, 1, 2$
- * i 番目の人の連続形質の値を y_i (例、血糖値)
- * 誤差 $\varepsilon_i \sim \text{Normal}(0, \sigma^2)$
- * 連続形質に対するSNPの効果は β

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

* 複数の研究で推定された効果を統合する

- * j 番目の研究での効果の推定値が β_j 、標準誤差が s_j
- * $1/s_j^2$ で重み付け
- * 全体での効果の推定値 β 、標準誤差が s
- * 利点: 個人の遺伝子情報は外に出さずに済む

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

東アジア人大規模GWASメタ解析

* 目的

- * 東アジア人で影響の強い高血圧関連遺伝子座の探索

* 方法

* 1次スクリーニング

- * 東アジア人(日本・韓国・中国・台湾・シンガポール)ゲノム疫学コンソーシアム(AGEN)の約2万人を対象としたGWAS結果をメタ解析する。
- * ゲノムワイドに ~200万 SNPs を検定

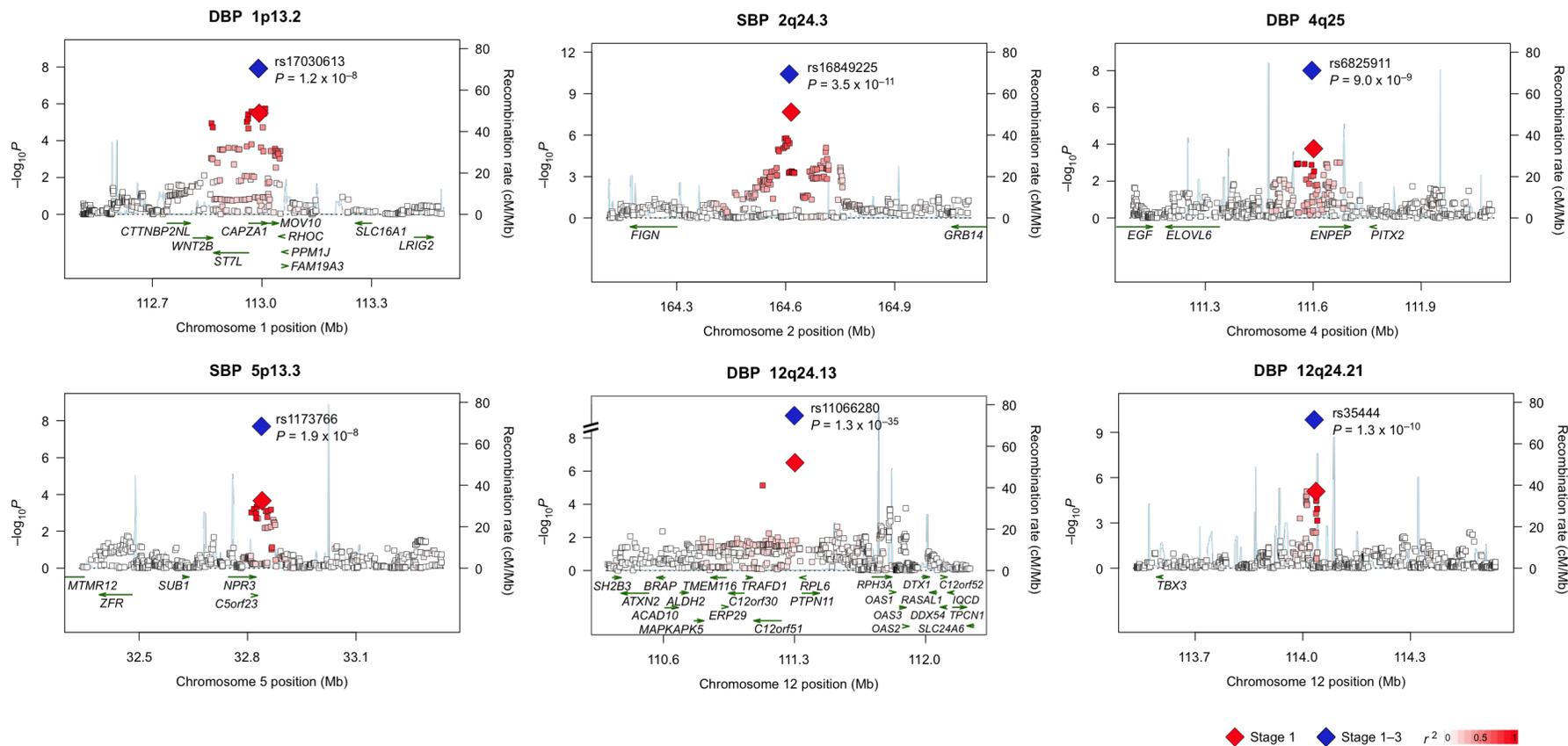
* 2次スクリーニング

- * 日本人1万人をタイピング

* 追試

- * 日本人2万人をタイピング

血圧感受性領域の新規同定



GWASメタ解析で見つかった感受性多型の強さ

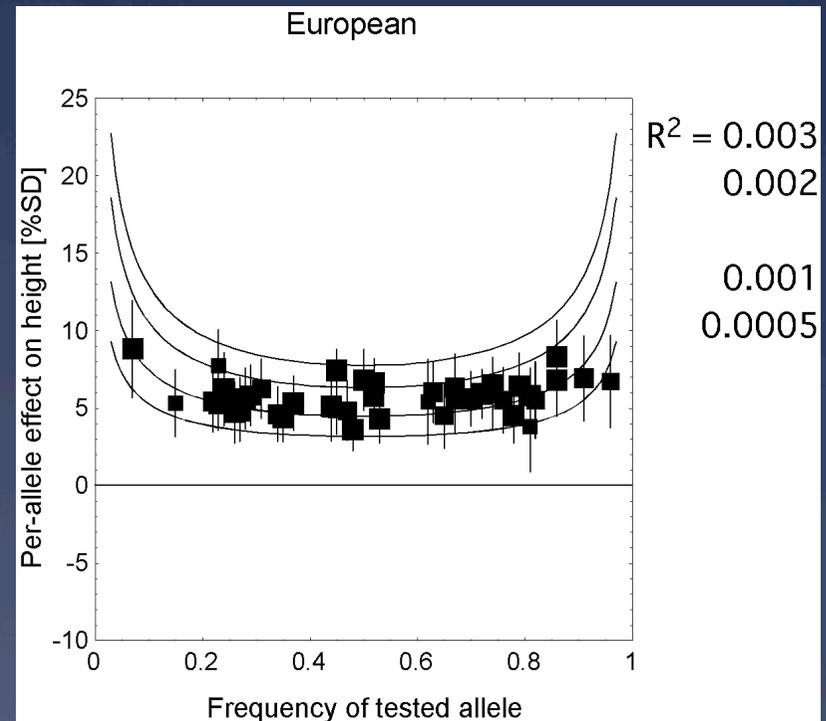
* $R^2=0.0005\sim0.0031$

* 検出力80%となるのは、 $N=13,000\sim80,000$

Chr	SNP ID (pos Build 36.3)	Coded/Other allele	Nearby Gene(s)	N	Coded allele freq.	SBP			DBP		
						Beta (SE), mm Hg	P	R ²	Beta (SE), mm Hg	P	R ²
1	rs17030613 (112,971,190)	C/A	ST7L CAPZA1	49,952	0.49	0.49 (0.11)	8.4E-06	0.0003	0.38 (0.07)	1.2E-08	0.0006
2	rs16849225 (164,615,066)	C/T	FIGN GRB14	49,511	0.61	0.75 (0.11)	3.5E-11	0.0007	0.29 (0.07)	2.7E-05	0.0003
4	rs6825911 (111,601,087)	C/T	ENPEP	49,515	0.51	0.60 (0.11)	7.3E-08	0.0005	0.39 (0.07)	9.0E-09	0.0006
5	rs1173766 (32,840,285)	C/T	NPR3	49,970	0.60	0.63 (0.11)	1.9E-08	0.0005	0.36 (0.07)	1.2E-07	0.0005
12	rs11066280 (111,302,166)	T/A	PTPN11 ALDH2	46,957	0.75	1.56 (0.13)	7.9E-31	0.0024	1.01 (0.08)	1.3E-35	0.0031
12	rs35444 (114,036,820)	A/G	TBX3	49,984	0.75	0.63 (0.13)	7.5E-07	0.0004	0.50 (0.08)	1.3E-10	0.0008

身長と関連する遺伝子多型

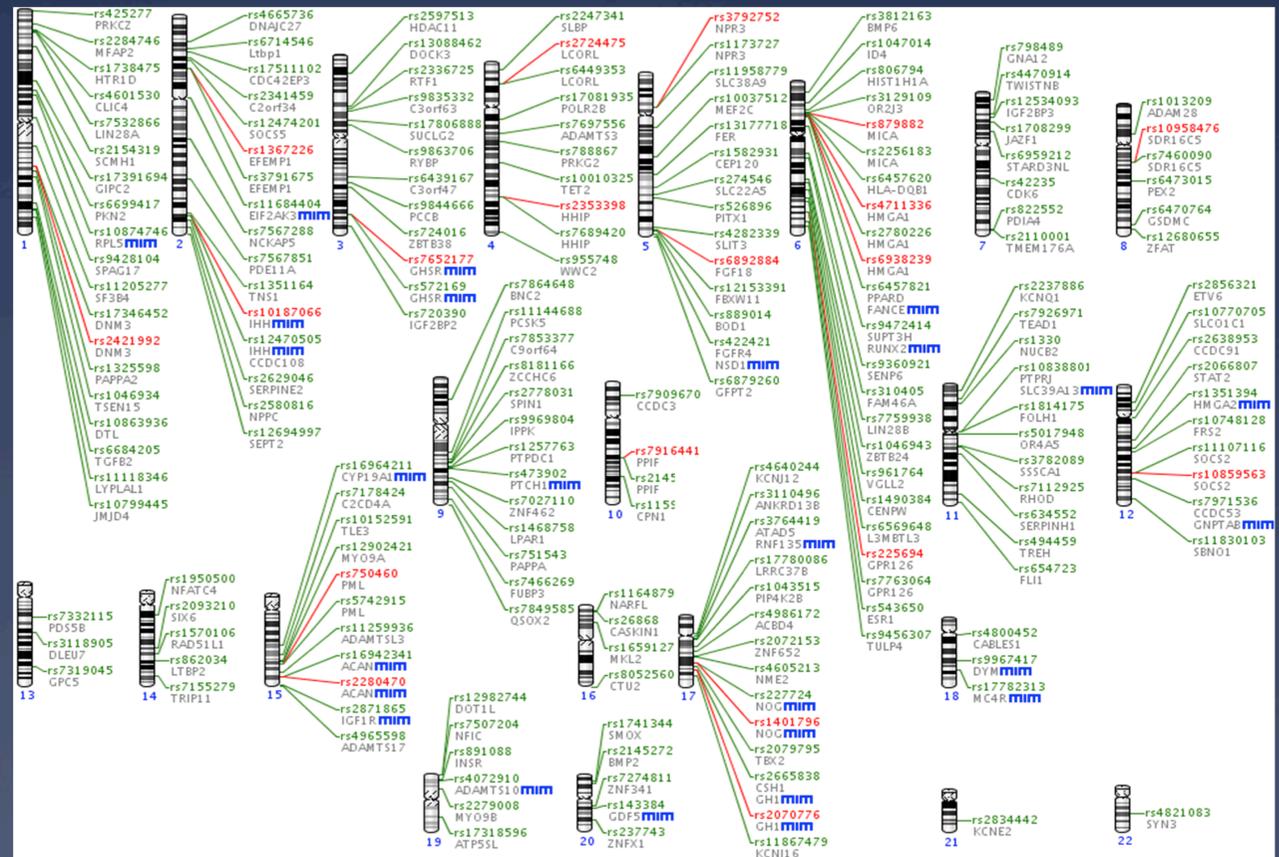
- * 身長の遺伝率は ~ 0.8
- * 2009年までに、欧米人で50の遺伝子領域が同定された
- * 全てが $R^2 < 0.003$



身長と関連する遺伝子多型

* 2010年には、
欧米人183,727
人のGWASメタ
解析で、180の
遺伝子領域が
同定された

Supplementary Figure 2. 199 loci associated with adult height variation. Karyogram displaying the genome location of the 180 height SNPs identified from the primary meta-analysis (green) and the 19 secondary signals (red) discovered in the conditional analysis to be associated with height. The closest genes to the SNPs (gray) are followed by a MIM (blue) label if the gene underlies a skeletal growth-related Mendelian disorder described in OMIM. The plot was created using Affyrmation (<http://genepipe.ngc.sinica.edu.tw/affyrmation/>).



第2部のまとめ： GWASメタ解析の成果と課題

- * GWASメタ解析でサンプル数を積みましたために、関連が弱いものも含め、数多くの関連SNPが同定できた
- * 関連SNPは分かったが、原因変異・原因遺伝子は未同定
 - * 関連SNPの周辺の遺伝子のどれか絞り込めない
 - * noncoding RNAなど、未知の遺伝子かもしれない
 - * なるべく絞り込んで(精密マッピング)、次段階の機能解析を効率的に行いたい

1. ゲノムワイド関連解析 (GWAS)。私の行ってきたもの
 - * 検出力の限界
2. GWASのメタ解析
 - * 解像度の限界
3. GWASで見つかった染色体領域の精密マッピング

GWASで見つかった染色体領域の精密マッピング

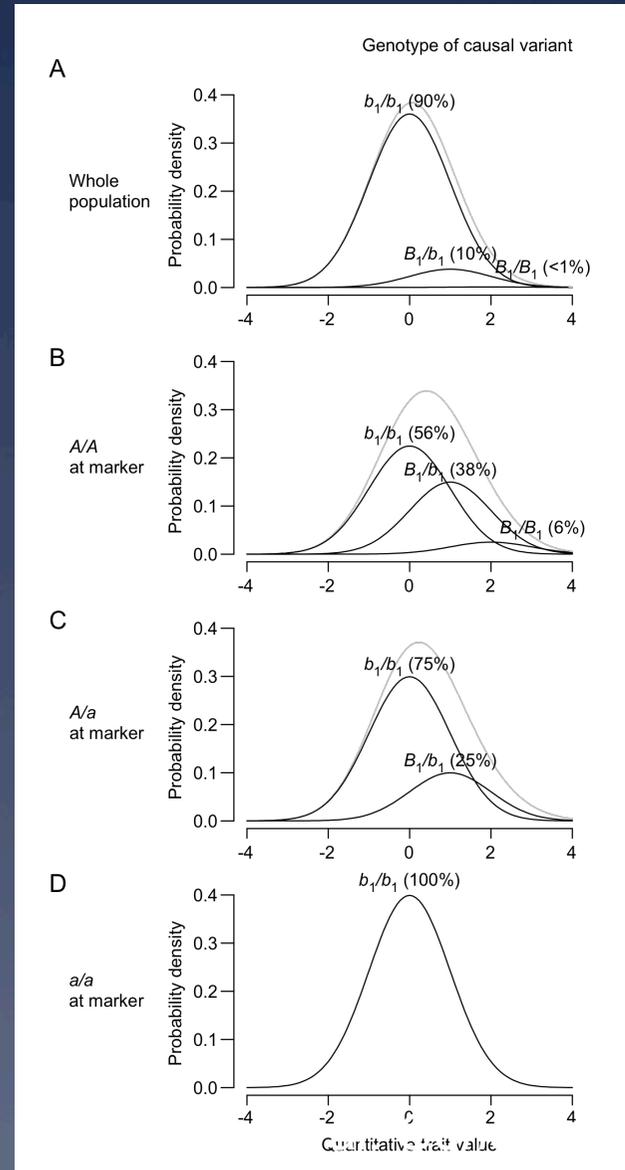
- * ゲノムワイド関連解析により、形質と関連する common (頻度 $\geq 5\%$) な一塩基多型 (SNP) が多数見つかったが、これらは **マーカー** であり、関連の元となる **原因変異** は同定できていない
- * マーカー-SNPと原因変異の関係は？
 - * Indirect association: マーカー-SNPと頻度が同等で、強い相関を示す (連鎖不平衡係数 $r^2 \approx 1$) 原因変異が1つ有る
 - * Synthetic association: マーカー-SNPよりも頻度が低く、相関が強くない原因変異が (複数) 有る [Dickson 他 PLoS Biol 8:e1000294]
- * 原因変異の同定には、大規模な塩基配列再解読・タイピングが必要で、手間が大変！
- * 遺伝統計から、少し見当をつけられないか...

提案1: 遺伝子型間での異分散性から synthetic associationを検出する

- * マーカーSNPと量的形質(QT)の関連がsyntheticかを統計的に検出する
 - * マーカーSNPの3つの遺伝子型間で、QTの分散が不均一ならば、syntheticと判定する
 - * 異分散性は Bartlett's test [Bartlett 1937] で検定する
 - * QT全体の分布は予め rank-based inverse normal transformation [Blom 1958] で正規化しておく
 - * 原因変異が未知でも検出できる！

低頻度原因変異に因る異分散性： モデルケース

- * SNPs and haplotypes
 - * marker SNP with alleles A and a
 - * causal variant with alleles B_1 and b_1
 - * allele B_1 (5% in frequency) is always linked to allele A (20% in frequency); thus existing haplotype classes are AB_1 , Ab_1 and ab_1
- * QT distribution
 - * Normally distributed with the unit variance and the mean equal to 2, 1 and 0 within a subgroup of individuals having genotype B_1/B_1 , B_1/b_1 and b_1/b_1 , respectively
 - * **A:** In the whole population, a mixture of the normal distributions combined according to the frequency of genotypes B_1/B_1 , B_1/b_1 and b_1/b_1
 - * **B:** Individuals with A/A genotype at the marker SNP are enriched with the genotypes of B_1/B_1 and B_1/b_1 at the causal variant, which are minor in the whole population. **分散が大きい**
 - * **C:** Individuals with A/a genotype at the marker SNP
 - * **D:** All individuals with a/a genotype have b_1/b_1 genotype at the causal variant. **分散 = 1**



低頻度原因変異に因る異分散性： APOE遺伝子とLDLコレステロール

- * APOE 遺伝子は LDL コレステロール (LDL-C) 量と関連する
 - * 3つの isoform が2つの原因変異 (MAF <10%) でコードされている
 - * E3 (一般的) に対して、E2 (rs7412, Arg158Cys) では LDL-C 減少、E4 (rs429358, Cys112Arg) では LDL-C 増加
- * マーカー SNP rs405509 (MAF >30%) では異分散性が有意 (4990人で検定)

		Marker SNP	Causal variant	Causal variant
		rs405509	rs7412	rs429358
Isoform	Freq			
E2	5%	C	T	T
E3	26%	C	C	T
E3	60%	A	C	T
E4	10%	A	C	C

Testing heteroscedasticity of SNPs in the APOE locus associated with LDL-C.								
SNP	Genotype	Number of individuals	Distribution of LDL-C level		Association with LDL-C level			Heteroscedasticity
			Mean	Variance	Beta	p-value	R ²	p-value
rs405509 (GWAS SNP)	C/C	462	-0.153	1.182	-0.117	1.0E-07	0.006	0.019
	C/A	2035	-0.050	0.976				
	A/A	2343	0.073	0.978				
rs377702 (GWAS SNP)	T/T	32	-0.487	1.231	-0.191	5.1E-07	0.005	0.583
	T/C	677	-0.149	1.025				
	C/C	4131	0.028	0.991				
rs7412 (causal variant)	T/T	12	-1.302	1.079	-0.651	2.0E-44	0.040	0.92
	T/C	452	-0.584	0.981				
	C/C	4376	0.064	0.960				
rs429358 (causal variant)	T/T	3954	-0.042	0.987	-0.212	1.4E-09	0.008	0.73
	T/C	850	0.185	1.023				
	C/C	36	0.214	1.104				

We first adjusted LDL-C level for body mass index and categories by sex and age (≤ 40 , 41–50, 51–60, ≥ 61 years), and then applied rank-based inverse normal transformation. Individuals under lipid treatment were excluded. Data is shown for 4840 individuals with complete observation from the Amegasaki study in (Takeuchi et al. 2010).

シミュレーションによる検出力の評価

- * ゲノムワイド関連解析で同定されるマーカー SNPを仮定する(寄与率 $R^2=0.00592$; 有意水準 5×10^{-8} のもと、5000人で関連を検定したときの検出力が0.5)
- * Synthetic association について4つのモデルを仮定(次ページ)
- * シミュレーション(1000回)
 - * マーカーSNPと原因変異の遺伝子型とQTをランダムに生成(5000人)
 - * Synthetic associationを検定して、検出力を評価

Synthetic association のモデル

- * SNPs and alleles
 - * Marker SNP has alleles A and a
 - * l causal variants each have alleles B_1 and b_1 , B_2 and b_2 , up to B_l and b_l , where the *causal (low-frequency) allele B_j is linked to marker allele A*
 - * m other causal variants each have alleles C_1 and c_1 , C_2 and c_2 , up to C_m and c_m , where the *causal (low-frequency) allele c_j is linked to marker allele a*
- * Model 1 (マーカーの片方のアリルに、全ての低頻度変異アリルが連鎖している)
 - * All *causal* alleles linked to *marker* allele A have identical effect-size
 - * No *causal* allele linked to allele a
- * Model 2 (マーカーの両アリルに、低頻度変異アリルが均等に連鎖している)
 - * Effect-size is uniform
 - * Cumulative frequencies equal between *causal* alleles B_j and *causal* alleles c_j
- * Model 3
 - * Effect-size of *causal* alleles is uniform
 - * Cumulative frequency of the *causal* alleles B_j is twice the cumulative frequency of *causal* alleles c_j
- * Model 4
 - * Cumulative frequencies are equal between *causal* alleles linked to the two *marker* alleles
 - * Effect-size of *causal* alleles B_j is twice the effect-size of *causal* alleles c_j

結果1: 遺伝子型間での異分散性から synthetic associationを検出する

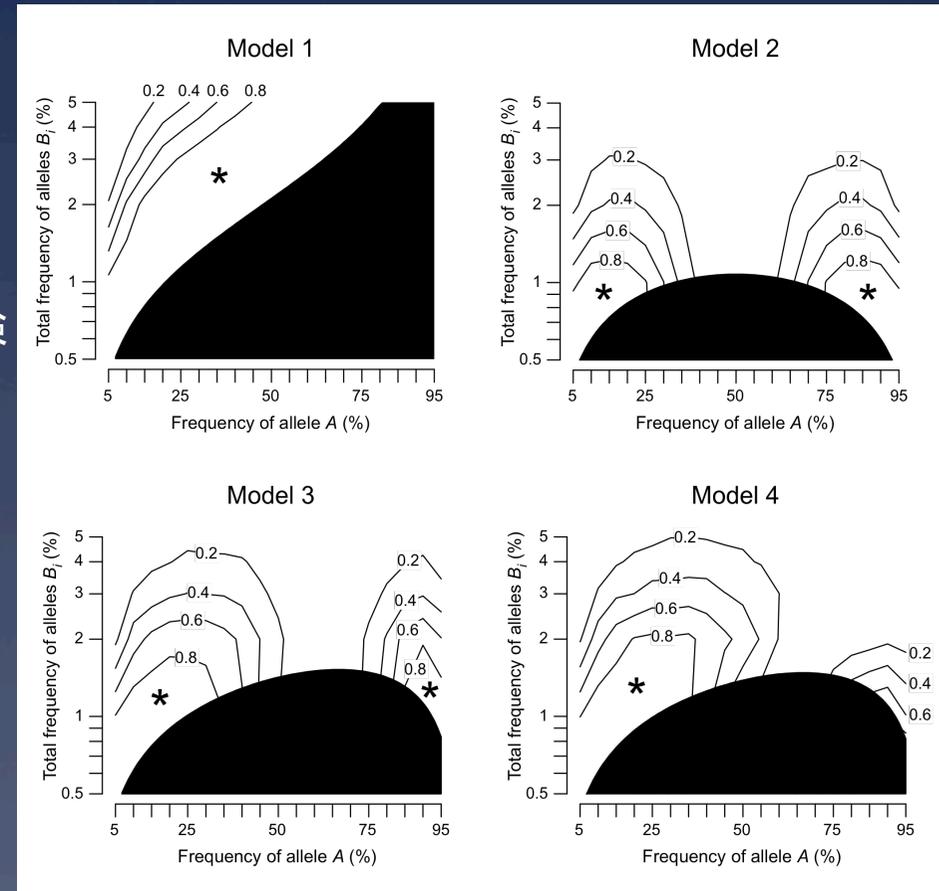
検出力をシミュレーションで評価した

* Model 1

- * マーカーの片方のアレルに、全ての低頻度変異アレルが連鎖している場合
- * (マーカーアレルの頻度) $\geq 45\%$ なら検出可能
- * (マーカーアレルの頻度) = 25% で、(低頻度変異アレルの合計頻度) $< 3\%$ なら検出可能

* Model 2

- * マーカーの両アレルに、低頻度変異アレルが均等に連鎖している場合
- * マーカーアレルの頻度 $\sim 50\%$ では検出できない!



•星(*)の領域で synthetic association が検出可能 (検出力 > 0.8)

•黒塗りの領域は原因変異が存在し得ないので無視

Takeuchi et al. *Genome Res* Epub

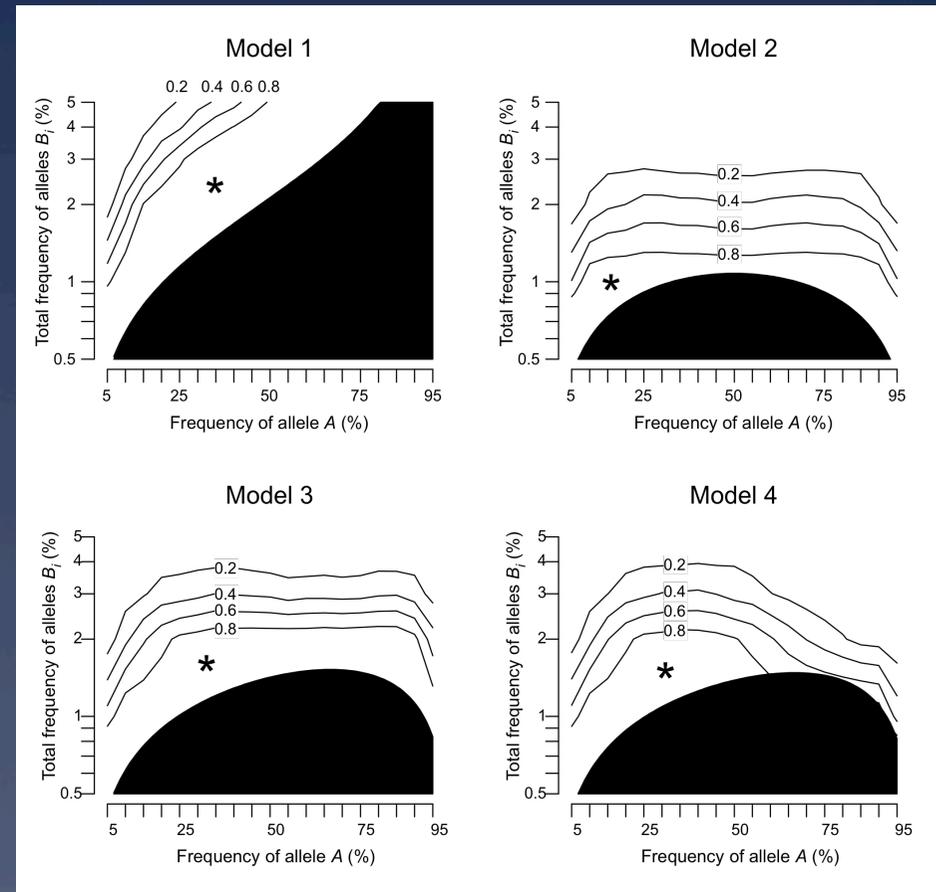
提案2: 異分散性と歪度から synthetic associationを検出する

- * Model 2(マーカーの両アリルに、低頻度変異アリルが均等に連鎖している)については、
 - * 異分散性が生じないが
 - * 各遺伝子型でのQT分布は歪んでいる
- * 各遺伝子型でのQT分布の歪度もsynthetic associationの検出指標になる
- * 提案2: 異分散性と歪度を(Fisherの方法で)組合せてsynthetic associationを検定する

結果2:異分散性と歪度から synthetic associationを検出する

検出力をシミュレーションで評価した

- * Models 1, 3, 4
 - * (マーカーアレルの一方に連鎖する変異アレルの合計頻度) $\leq 2\%$ なら検出可能
- * Model 2
 - * マーカーの両アレルに、低頻度変異アレルが均等に連鎖している場合
 - * (マーカーアレルの一方に連鎖する変異アレルの合計頻度) $\leq 1\%$ なら検出可能
- * いずれの場合も
 - * (変異アレルの合計頻度) $\leq 3\%$ なら検出可能



- 星(*)の領域で synthetic association が検出可能 (検出力 > 0.8)
- 黒塗りの領域は原因変異が存在し得ないので無視

第3部のまとめ

- * ゲノムワイド関連解析で見つかったSNPの関連が低頻度変異に由来する(synthetic)かは、QT分布の遺伝子型間の異分散性と歪度により統計的に検出できる
- * 低頻度原因変異の合計頻度 < 3% の場合、検出力 >80% (有意水準5%、5000人で検定)
- * synthetic な関連が検出されたときは、低頻度多型の探索により、原因変異の同定が期待できる
→ 精密マッピングの方針決定に役立つ

今日のまとめ

1. GWASにより、多数の疾患感受性遺伝子が同定できた。これは今後の疾患解明の手がかりとなる
2. GWASにおいては、 10^6 SNPs を多重検定することから、真の関連を見つけ出すには多数のサンプルが必要になる
 - * サンプルサイズを大きくするためには、多数のGWASのメタ解析が有効である
3. GWASあるいはそのメタ解析で同定される関連多型はあくまでもマーカーであり、原因多型を同定するための精密マッピングが必要
 - * そのための新たな統計手法が必要。その一つとして、遺伝子型ごとの表現型の分散を調べる手法が有効である。

展望

- * 今後数年
 - * 数十人の全ゲノム(あるいは全exon)配列解読はお手頃
 - * メンデル型遺伝病の原因遺伝子同定が進みつつある
 - * 稀だけど重篤
 - * 原因遺伝子が数個
 - * 特定領域なら数千人の配列解読も可能
 - * GWAS領域の精密マッピングが進む
- * 10年後
 - * 全ゲノム配列解読が普通になり、GWASは不要になる
- * 入手できるデータに対応した統計・情報解析の手法が必要

私の今後の研究

- * 3月から国立感染症研究所・病原体ゲノム解析研究センターに着任
- * 次世代シーケンサを用いた感染症研究なら何でも
 - * 病原体(ウイルス・細菌・真菌・寄生虫)のゲノム
 - * 不明病原体の解読
 - * 感染症らしいが、従来法では病原体が分からない
 - * ヒトの感染症感受性・ワクチンの副作用
 - * 病原体とヒトの遺伝子型の相互作用、共進化