



TOHOKU
UNIVERSITY

東北メディカル・メガバンクにおける プレシジョン医療の実現のための ゲノム情報と環境要因情報の統合

東北大学

東北メディカル・メガバンク機構

バイオクリニカル分野

田中 博、荻島創一、永家聖

東北メディカル・メガバンク機構
TOHOKU MEDICAL MEGABANK ORGANIZATION



本日の話題

- ゲノム情報と環境情報(生活習慣)を統合して分析することの必要性、およびそのための統合データベース構築の不可欠性
- 東北メディカル・メガバンク計画における「統合データベースdbTMM」
- 「ゲノムと環境情報の相互作用(GxE)」を考慮して疾患の発症相対リスクを計算する、我々が開発した「新しい方法」

1. なぜ

ゲノム情報と環境情報は統合して
解析しなければならないのか
—「統合データベース」の必要性—

疾患の個別化発症予測・個別化予防

コホートデータに基づいて疾患の発症を予測する

東北メディカル・メガバンク計画の主要な目標の一つ

特に重要視される目標は

多因子疾患（「複雑な疾患」）の
（「ありふれた病気」；高血圧，糖尿病）
発症予測に基づいた個別化予防

現在の「ゲノム医学」のアプローチ

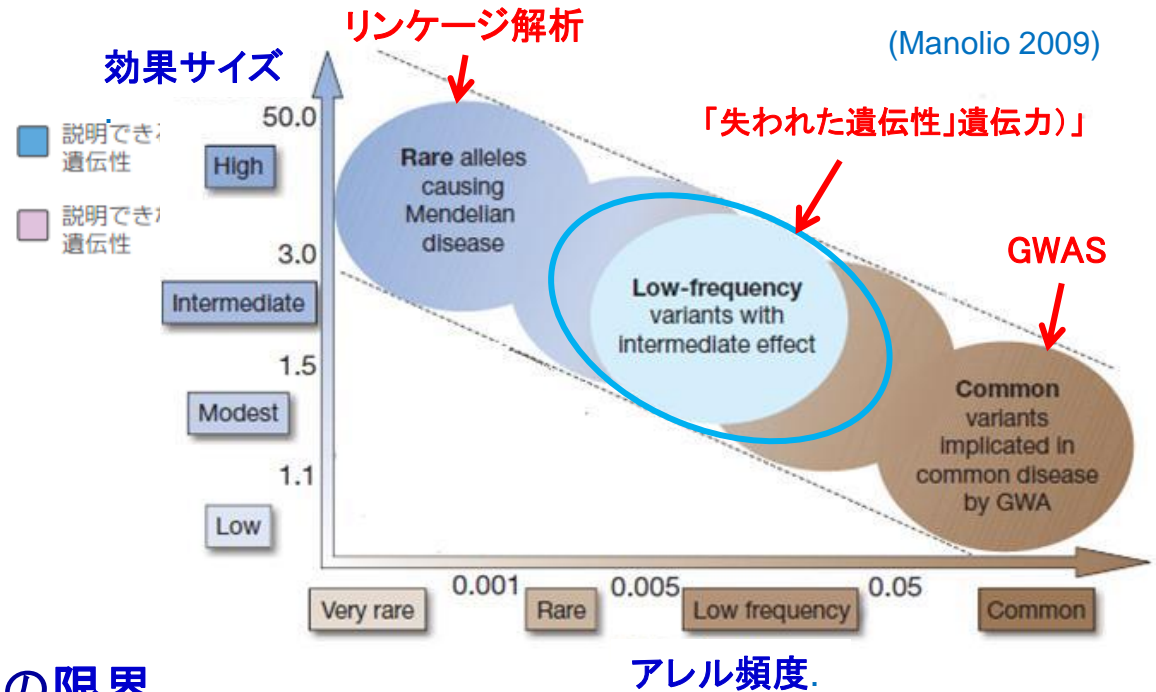
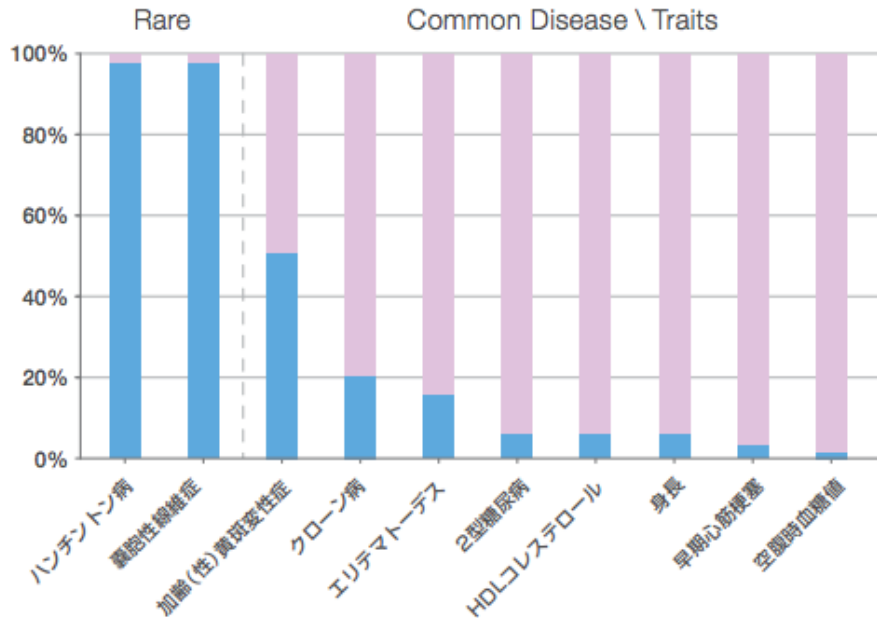
成功した臨床実装

1. 希少未診断疾患の原因遺伝子を病院の現場でシーケンサにより同定する
2. がんのドライバー遺伝子変異を同定し、適切な分子標的薬を処方する
3. 患者の薬剤の代謝酵素の多型性を先制的に同定し、副作用を防ぐ

しかし

多因子疾患の機序同定・発症予測などには全く無力である

現行の「ゲノム医学」的方法の「無効性」



- 「単一遺伝的原因」帰着アプローチの限界**
 - 連鎖解析にしろGWASにしろ、疾病という表現型を**単一の原因遺伝子や単一の遺伝子多型性に帰着する統計的方法論**で解析されている。他の遺伝的要因との相相互作用は考慮されていない。
 - これは**ゲノム情報の変数の数が観測する個体例(多くても数万)に比べ、圧倒的に多い(例、何千万SNV)ため、通常の統計学(変数の数は観測個体数より圧倒的に少ない少ない)が利用できない**からである。したがって**複数のSNVの相相互作用を検出できる多変量解析が利用できない(p>>n 問題)**
- 「失われた遺伝力」の主要な原因**
 - 複数の疾患関連遺伝子間の相相互作用 (epistasis) が取り入れられていない : $G \times G$
 - 環境と遺伝子の相相互作用が取り入れられていない : $G \times E$

(ここでのXは乗算ではなく相相互作用を

遺伝子素因と環境要因の相互作用

稀少な単一遺伝疾患を除き、**大半の疾患は、個人の遺伝的体質と環境(暴露)要因の複雑な相互作用の**

疾患発症の相対リスク=相互作用(ゲノム、環境要因)

- 両者は加算的($G \oplus E$)でもなく乗算的($G \otimes E$)でもない
- **<(G,E) 組合せ特異的な効果>である**

GWASでSNPの相対リスクが低い(1.1~1.3)理由: **GxE組合せ特異的な効果を環境要因の全てに亘って平均しているからである**



<GxE組合せ特異的> 相対リスクの例

- 遺伝素因と環境要因の相互作用の典型例
 - GxEの効果, 加算的でも乗算的でもない
- 大腸がん発症の相対リスク
 - ハワイでの調査 (Le Marchand 2001)
 - 環境要因: 喫煙、焦げた赤肉を嗜好
 - 遺伝素因: CYP1A2, NAT2のタイプ

		CYP1A2 Phenotype \leq Median		CYP1A2 Phenotype $>$ Median	
		Likes rare/medium meat	Likes well-done meat	Likes rare/medium meat	Likes well done meat
Non-Smoker	NAT2 Slow	1	1.9	0.9	1.2
	NAT2 Rapid	0.9	0.8	0.8	1.3
Ever-Smoker	NAT2 Slow	1	0.9	1.3	0.6
	NAT2 Rapid	1.2	1.3	0.9	8.8

疾患発症の相対リスクが

GxE組合せ特異的である

L. Le Marchand, JH. Hankin, LR. Wilkens, et al Combined Effects of Well-done Red Meat, Smoking, and Rapid N-Acetyltransferase 2 and CYP1A2 Phenotypes in Increasing Colorectal Cancer Risk, Cancer Epidemiol. Biomarkers Prev 2001;10:1259-1266

統合データベースの不可欠性

以上から、多因子疾患の発症リスクの評価においては
遺伝素因(SNVなど)と環境要因(生活習慣)の全ての組合せについて
網羅的に計算する必要がある

東北メディカル・メガバンク(TMM) は、ゲノム情報や環境暴露因子に関して多様な項目を調べている。発症予測と個別化予防の実現だけでなく、様々な種類の統合分析のインフラとして、遺伝情報と環境情報を統合的に扱う統合データベースが必要とされる

統合データベース

- ① 遺伝情報と環境情報を同一のデータ空間に格納している
- ② 両者に跨る検索条件を利用でき「横断的に」検索できる

個別化発症予測・個別化予防の基礎となる
時間的に追跡し、生涯型ヘルスケアの基礎となる

3. GxE相互作用を取り入れた 「新しい疾患発症リスク解析」

GxE相互作用を取り入れた 「新しい疾患発症リスク解析」

疾患の発症リスクはGxE組合せ特異的である

我々は網羅的にすべての遺伝素因と環境要因の組合せに関して、疾患発症への効果を評価する必要がある

最初の解析段階として(GE-WAS)

遺伝素因と環境要因を1対1の組合せ効果を網羅的に調べる

<1つの遺伝的素因>

G_i

x

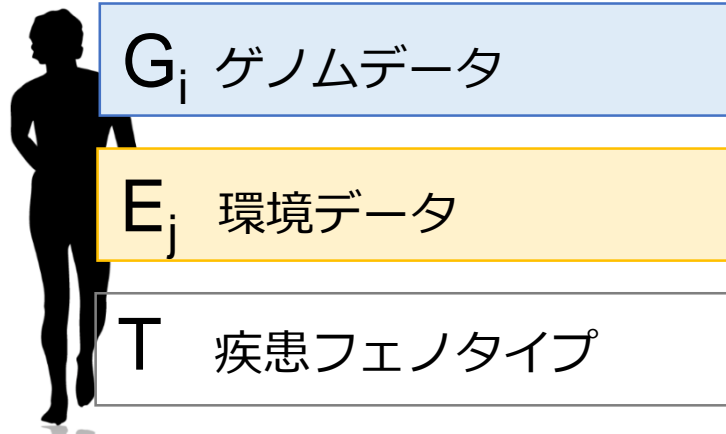
<1つの環境要因 >

E_j

GxE1対1逐次効果をすべてのG、Eの組合せについて計算する

1対1の組合せ以上に高次な組合せについては、次の段階で計算する

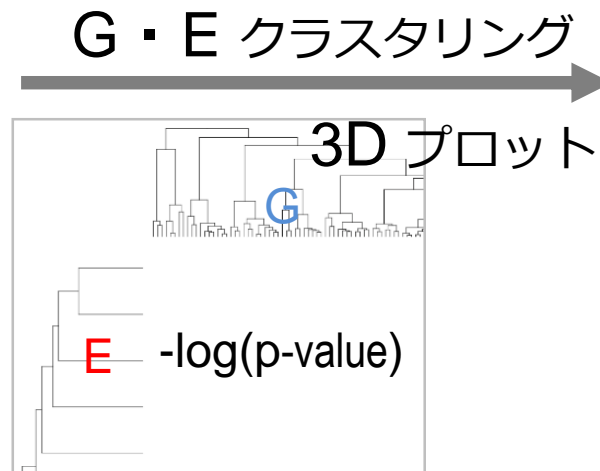
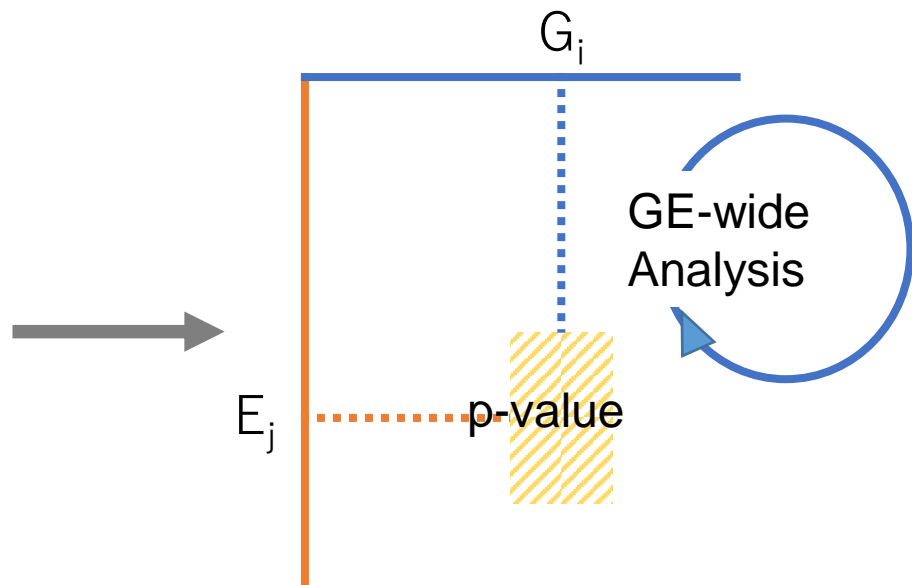
GE-WAS 解析フロー



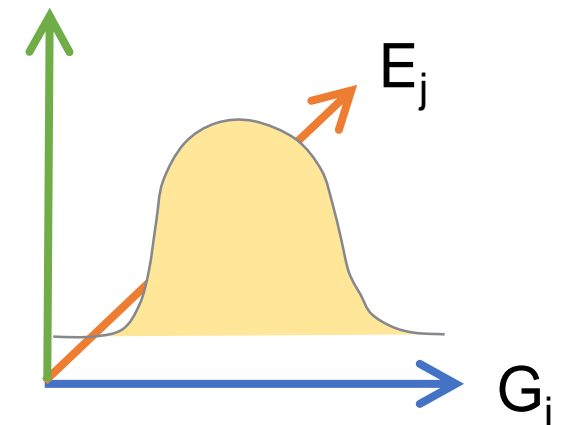
計算

T 対象		G_i			A
		G_i			
T 症例	E_j	aa	aA	AA	5_{ij}
		0	N^1_{ij}	N^3_{ij}	
1	N^2_{ij}	N^4_{ij}	N^6_{ij}		

コクラン-マンテル-ヘンツェル検定



-log(p-value)



GE-WAS 解析フロー

G_i 9番染色体
499,738 SNVs

E_j 検体検査項目41 因子
(正常/異常)

T アレルギー性鼻炎: 167 件
コントロール: 901 件

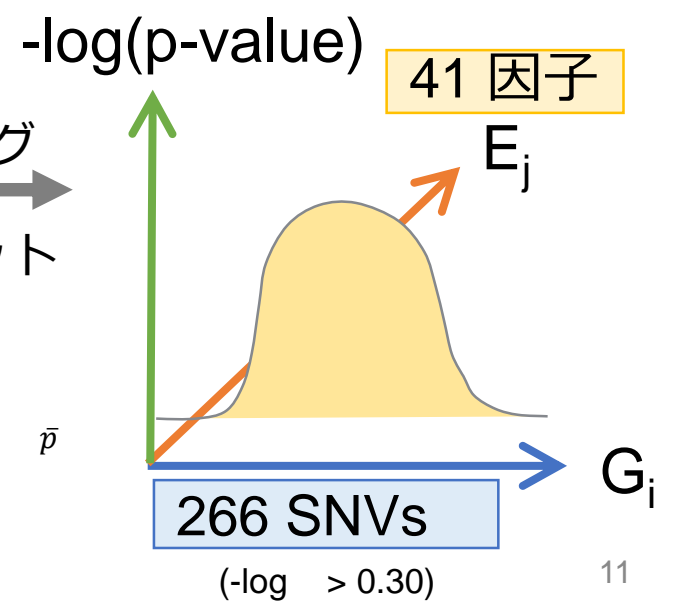
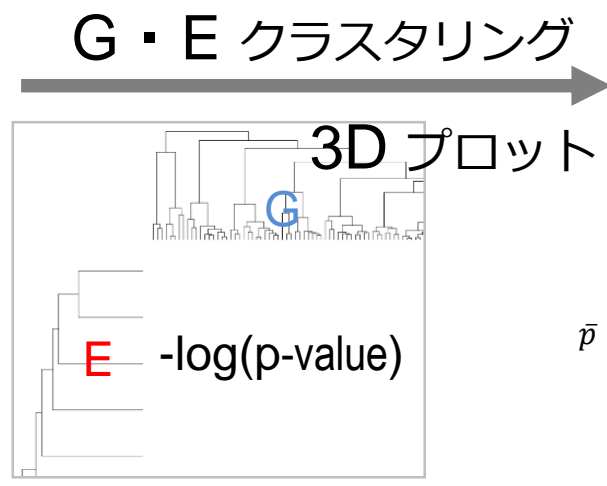
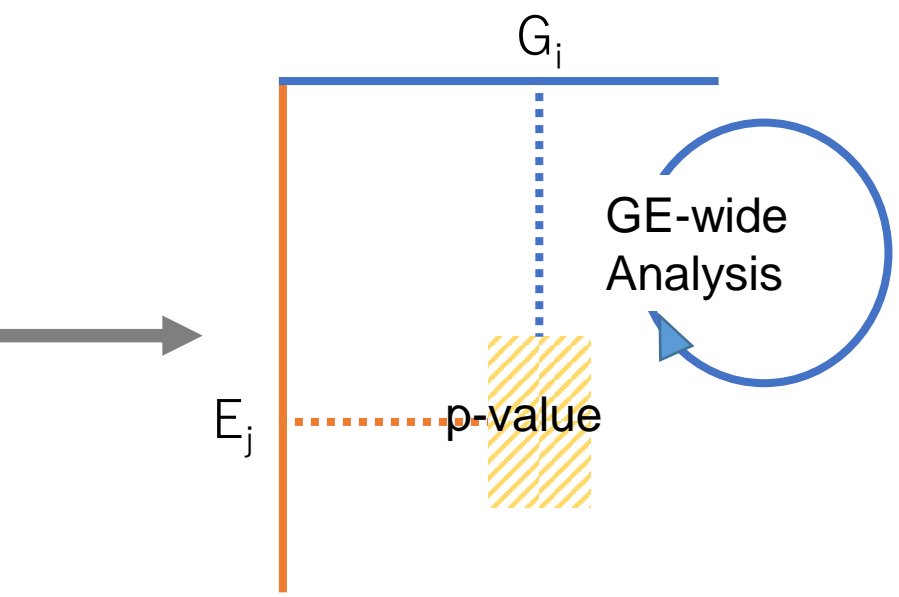
ToMMo 1KJPN

計算

T 対象		G _i		
		aa	aA	AA
T 症例	0	N ¹ _{ij}	N ³ _{ij}	N ⁵ _{ij}
	1	N ² _{ij}	N ⁴ _{ij}	N ⁶ _{ij}

41 因子 1,237 SNVs

コクラン-マンテル-ヘンツェル検定

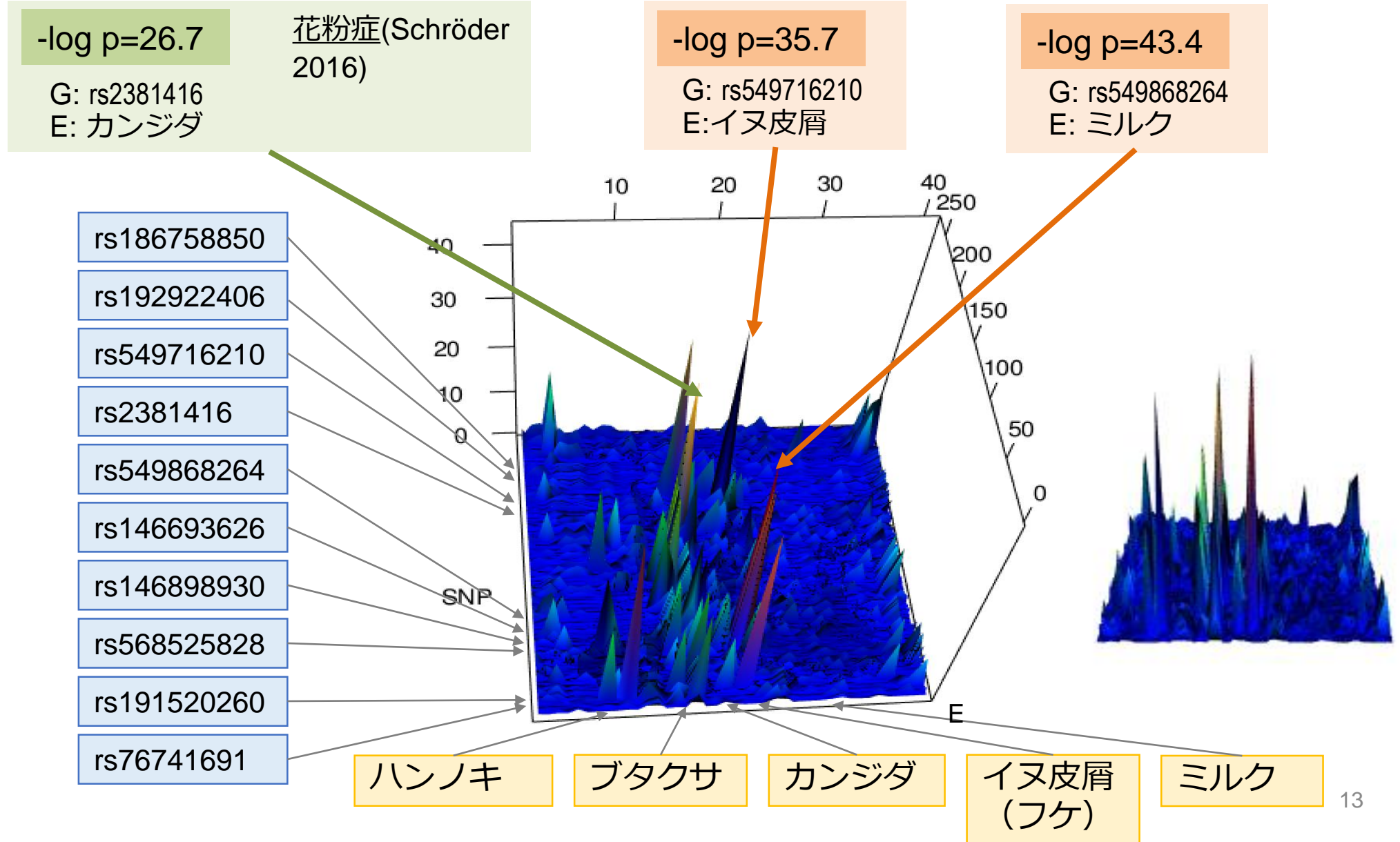


*Sakashita, Clin Exp Allergy, 2008

検体検査項目

No.	項目名	No.	項目名	No.	項目名	No.	項目名	No.	項目名
1	NT-proBNP	11	シラカンバ	21	カニ	31	白血球数	41	I g E (非特異的 I g E)
2	アルブミン/ Cre補正	12	コナヒョウヒダニ	22	ミルク	32	赤血球数		
3	アルブミン定量値	13	ハウスダスト I	23	牛肉	33	血色素量		
4	クレアチニン (血中)	14	ペニシリウム	24	卵白	34	ヘマトクリット値		
5	オオアワガエリ	15	カンジダ	25	ピーナッツ	35	平均赤血球容積		
6	ハルガヤ	16	ネコ皮膚	26	抗体濃度	36	平均赤血球 色素濃度		
7	カモガヤ	17	イヌ皮膚	27	尿素窒素	37	血小板数		
8	ブタクサ混合物	18	小麦	28	尿酸	38	リンパ球		
9	ヨモギ	19	米	29	グルコース	39	好酸球		
10	ハンノキ	20	エビ	30	グリコアルブミン	40	好中球		

アレルギー性鼻炎のGxE ランドスケーププロット



本方法のまとめ

- $\langle 1$ つの遺伝素因 $G_i \rangle \times \langle 1$ つの環境要因 $E_j \rangle$ の1対1の組合せすべての相互作用の偶現表についての有意水準レベル (p_{ij}) を計算する方法を開発した
- アレルギー性鼻炎に関して関係するSNVと環境要因について $-\log p_{ij}$ を計算しGxE2次元平面にてプロットした
- 環境要因とSNVは組合せ特異的な有意水準を示した
- これまでの「**SNV単独の発症相対リスク**」の概念は誤った理解を導く誤謬概念であることが判明した。

Populationタイプのバイオバンクの 長期的目標

- “病気の医学”
 - 19世紀の後半から、細菌病理学(コッホ・パスツール)、細胞形態学(ウィルヒュウ)の確立とともに、急速に発展した近代医学
 - 現在は、ゲノム医学や免疫学、再生医学が領導している
 - “疾患型バイオバンク”の基軸概念
- “人の医学”
 - 個人の疾患発症可能性・疾患感受性を、全生涯的な視座で、個人が生活する全エコシステムから、その「全体性」において理解する
 - 近代の「病気の医学」も次第にこの観点を取り込みつつある
 - 1. 全生涯的な観点から、疾患発症可能性の理解
 - “胎内から墓場まで”; DOHaD 学説, life course ヘルスケアの概念
 - 2. 個人が生きるエコシステムの全体から、発症可能性の理解
 - 環境と遺伝素因の相互作用の場としての腸内細菌叢
 - ” population 型バイオバンク” が追跡する医学の概念

ご清聴ありがとうございました