

日常の中の遺伝統計学

上辻茂男

Shigeo Kamitsuji, PhD

株式会社スタージェン

開催日時：2015年4月17日（金）17:00-18:00

開催場所：上智大学市谷キャンパス本館201 室



日常のなかの統計学：私の日常のなかの統計学



初めて統計学を学ぶ

- 薦尾 泰俊 名誉教授
慶應義塾大学理工学部管理工学科
- 学部2年時の統計学の授業
 - 数理科学科の学生に甘かった

数理統計学を学ぶ

- 渋谷 政昭 名誉教授
慶應義塾大学理工学部数理科学科
- 授業がおもしろかった（ブラックジョークばかり）
 - テストに出る問題を教えてくれた

データ解析の基礎を学ぶ

- 樫 広計 先生
独立行政法人統計センター理事長
- 「仏の樫」として有名だった
 - いろんなデータを取得し解析した

データサイエンスを学ぶ

- 柴田 里程 名誉教授
慶應義塾大学理工学部数理科学科
- 学部4年生のときにはサバティカルでいなかった
 - 精神面を鍛えてもらった

- 鎌谷 直之 先生
痛風財団理事長、弊社社長、元東京女子医科大学附属膠原病リウマチ痛風センター所長、元理化学研究所センター長
- 遺伝統計学のパイオニア
 - 統計学がやりたくて定年間際で医者をやめた

遺伝統計学を学ぶ

遺伝子ブーム到来：遺伝子検査サービス



NIPT (Noninvasive prenatal genetic testing) 無侵襲的出生前遺伝学的検査

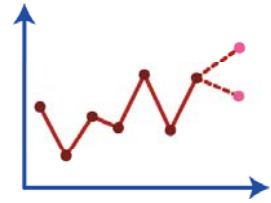
妊婦の母体血から胎児の染色体数を調べる

羊水検査

慶應義塾大学病院医療健康情報サイトKOMPAS
http://kompas.hosp.keio.ac.jp/contents/000379.html

統計学は多様な現象を説明するための道具

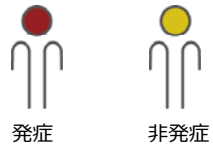
次の日の株価は上がる？



Learning machine

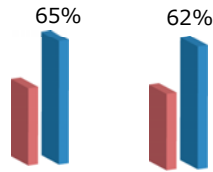
↑ : 53%
→ : 7%
↓ : 40%

病気の原因は飲酒量？



発症

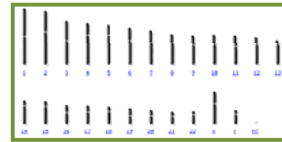
非発症



青 : 1合以上/日
赤 : 1合未満/日

有意差なし

病気の原因はどの遺伝子？



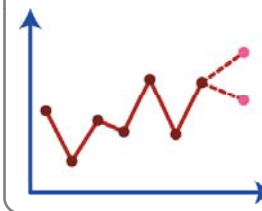
	+	-
発症	64	36
非発症	48	52

$P < 0.05$, $OR = 1.9$

正確な予測に必要なこと

- 因果関係がはっきりしていること
- 原因と結果の間のメカニズムが分かること
- データが正確に観測できること

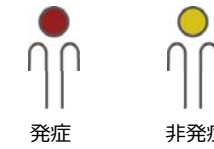
次の日の株価は上がる？



原因を特定できない
メカニズムが分からない

私の博士論文

病気の原因は飲酒量？



発症

非発症

過去の記憶は曖昧
データの改竄

病気の原因はどの遺伝子？



原因はゲノム配列の多様性
メカニズムは遺伝継承法則
安定した観測データ

統計学者の多くが遺伝学者

生物計測学派

大喧嘩

メンデル学派

計測されたデータから
遺伝的要因と形質の関係を探る

遺伝的要因と形質の間では
メンデルの法則が成立している



C リンネ, 18世紀
生物分類学



CR ダーウィン, 1859
ダーウィンの進化論



W ウェルドン, 19世紀
動物学者



G メンデル, 19世紀
メンデルの法則



W ベートソン, 20世紀
Geneticsを提唱



F ゴールトン, 20世紀
帰属



K ピアソン, 20世紀
相関係数, カイ2乗検定



W ヨハンセン, 20世紀
Geneを命名

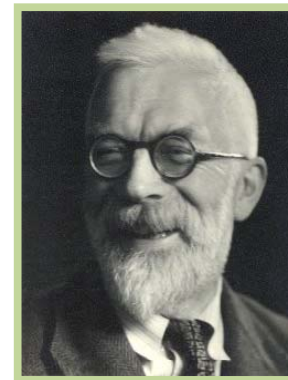


RA フィッシャー, 20世紀
仮説検定、尤度

データありき

モデルありき

Modern Synthesisと遺伝継承法則



Sir Ronald Aylmer Fisher (1890 – 1962)

現代の数値統計学の基礎を築いた1人

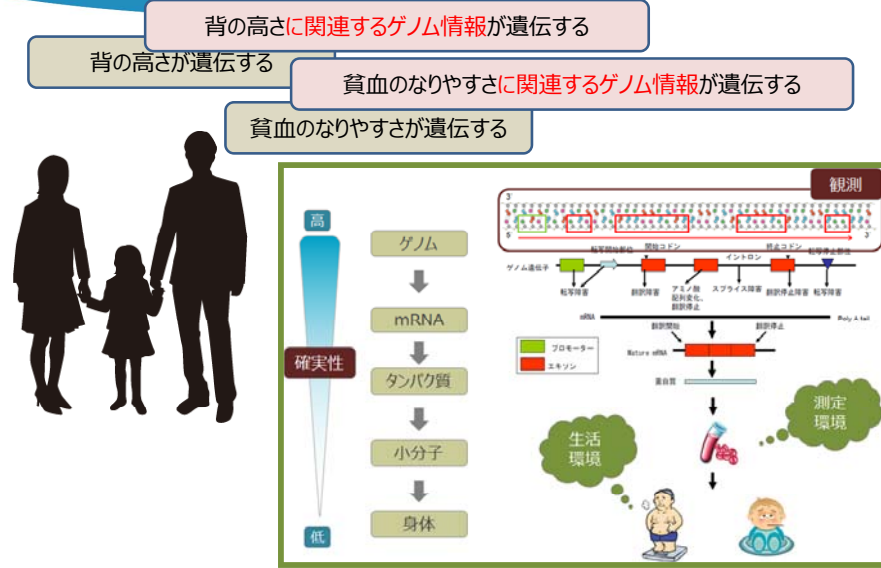
- 分散の概念、分散分析
- 中心極限定理
- 仮説検定、フィッシャーの正確検定
- ランダム化の重要性
- 連鎖解析
- 相加的ポリジーンモデル

Modern synthesis

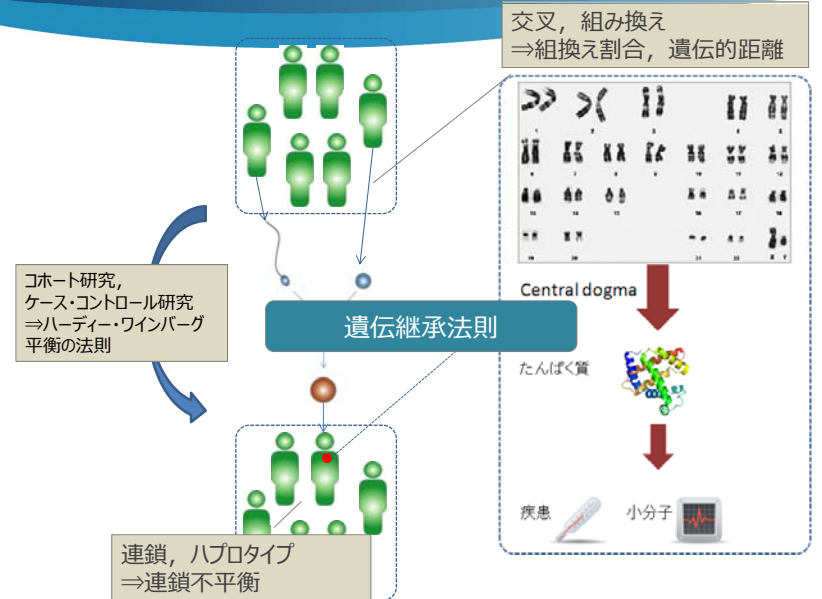
heredityとvariationを数理的に統合

- ダーウィンの進化説
- 突然変異 (mutation)
- メンデルの遺伝継承法則
- 個体間の多様性 (variation) : 生物計測学の概念

ゲノム情報は安定して伝わる



遺伝継承法則は真実のモデル



ゲノム研究環境が成熟してきた 遺伝的多様性の指標「SNP」とHapMapプロジェクト

SNP (一塩基置換, スニップ)

父由来
母由来

International HapMap Project
Home | About the Project | Data | Publications | Tutorial

全ゲノムをほぼカバーするSNP座位のリストが作成された

国際貢献度

9.5%	US
10.1%	Japan
23.7%	UK
24.3%	Canada
32.4%	China

Genomic sequence example:
A A T C G T A G
A A T C G T A G
A A T T G T A G
A A T T G T A G

ゲノム情報取得技術の飛躍的な向上

アレルの観測

技術革新により安価にかつ大量に観測が可能となった

候補遺伝子タイピング
DNAチップ
Massively parallel sequencer

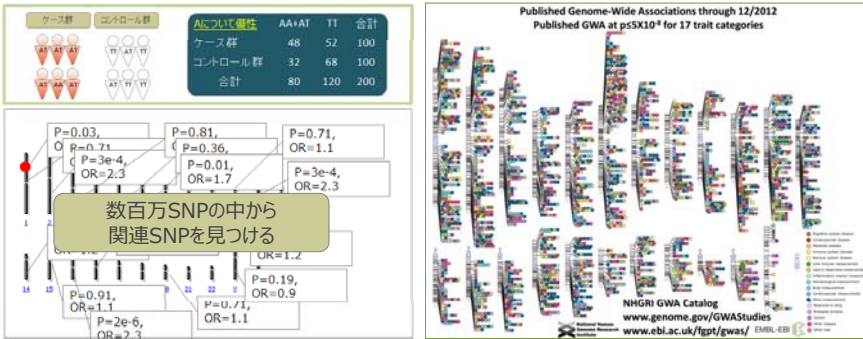
Invader法
TaqMan PCR法
GoldenGate法
...

~1,000,000座位
~3,000,000,000座位

DTC遺伝子検査
新型出生前診断

ゲノムワイド関連研究 (Genome-Wide Association Study, GWAS)

疾患の有無や薬効などに関連する遺伝的要因を網羅的に探索する



世界で行われたゲノムワイド関連研究から得られた因果関係のデータベース

NATURE | VOL. 428 | 1 APRIL 2004 | www.nature.com/nature

Medical genetics

A marker for Stevens-Johnson syndrome

Wen-Hung Chung*, Shuen-Iu Hung†, Hong-Shang Hong*, Mo-Song Hsih‡, Li-Cheng Yang*, Hsin-Chun Ho*, Jer-Yuarn Wu†§, Yuan-Tsong Chen†¶

We studied 44 patients with carbamazepine-induced Stevens-Johnson syndrome (CBZ-SJS), including five with overlapping toxic epidermal necrolysis, in whom the clinical-filled Roujeau's diagnostic

When the CBZ-tolerant group is used as the control, the presence of *B*1502* has a 93.6% positive-prediction value for CBZ-SJS, whereas its absence has a negative-prediction value of 100%. In a test for CBZ-SJS, the *HLA-B*1502* allele should therefore have 100% sensitivity and 97% specificity.



99%の確率で胎児は染色体数異常です。

検査の結果、陽性だったのですが。

本日のお話

専門家以外の方も触れる機会が増えたゲノムビジネスについて紹介します

- 遺伝統計学の概要
 - 家系データと集団データ
 - ゲノムデータの御作法
- 日常の中の遺伝統計学
 - 遺伝子検査
 - 医師を介さずに直接個人に遺伝子検査結果が通知
 - 新しい産業として新規参入企業が目立つ
 - 検査をする側も検査される側も統計的側面を理解していない場合が多い
 - 新型出生前診断の統計的問題にも触れる

ゲノム研究は言葉が分かりづらいので遠慮なく質問してください

遺伝統計学の概要

遺伝統計学に少しふれてみる

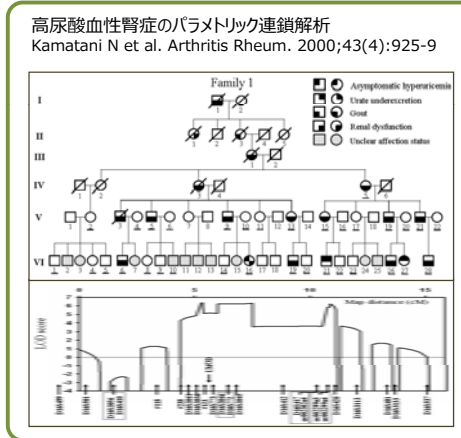


パラメトリック連鎖解析を利用した疾患関連座位の特定

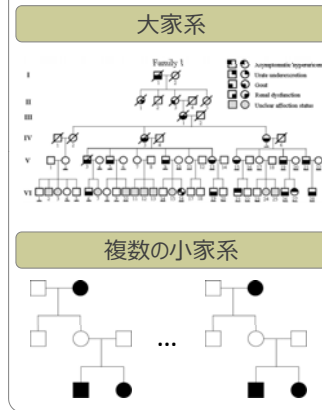
数多くのメンデル型遺伝性疾患の原因座位の特定に貢献した

- 1983. Gusella, J.F. et al.
 - ハンチントン（舞踏）病
- 1988. Monaco, A.P. et al.
 - デュシャンヌ型筋ジストロフィー
- 1989. Kerem, B. et al.
 - Cystic fibrosis (嚢胞性線維症, のうほうせいせんいしやう)
- 1990. Hall, J.M. et al.
 - Breast cancer (乳癌)
- 1993. Saunders, A.M. et al.
 - アルツハイマー
- 2000. Kamatani, N. et al.
 - 高尿酸血症性腎症

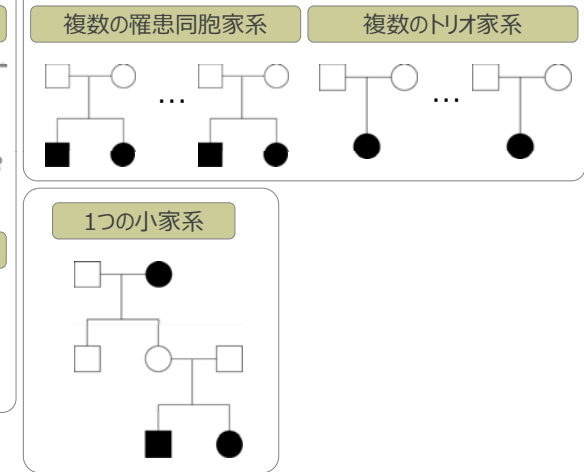
など



パラメトリック連鎖解析



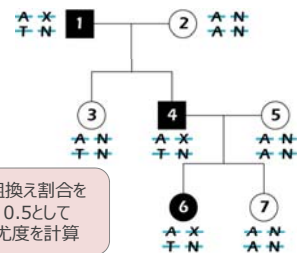
罹患同胞対解析・TDT解析（ノンパラメトリック連鎖解析）



パラメトリック連鎖解析
仮説の下での尤度と検定

帰無仮説の下での尤度

原因座位は近くはない：組換え割合0.5



検定

2つの仮説のうちどちらが尤もらしいか

LODスコア

尤度比の常用対数が3以上か

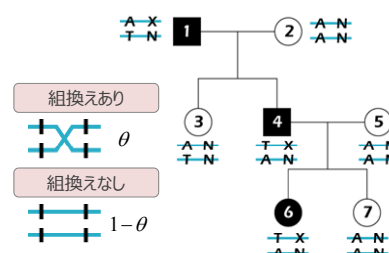
\log_{10} 対立仮説の下での最大尤度

帰無仮説の下での尤度

検定の代わり

対立仮説の下での尤度

原因座位が近くにある：組換え割合は0.5より小さい



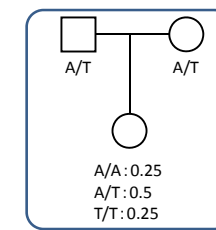
組換え割合θで表わされた尤度

最尤法

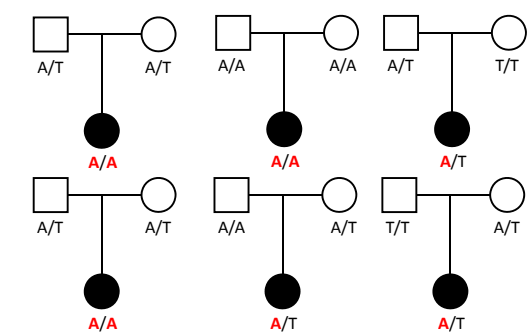
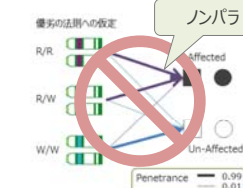
観測が尤も得られる組換え割合を推定

ノンパラメトリック連鎖解析

TDT解析（Transmission Disequilibrium Test）



分離の法則

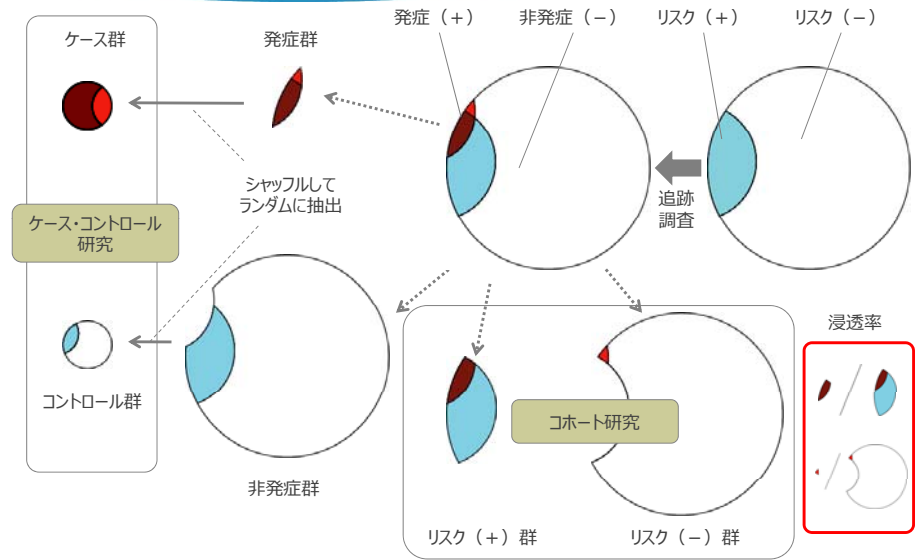
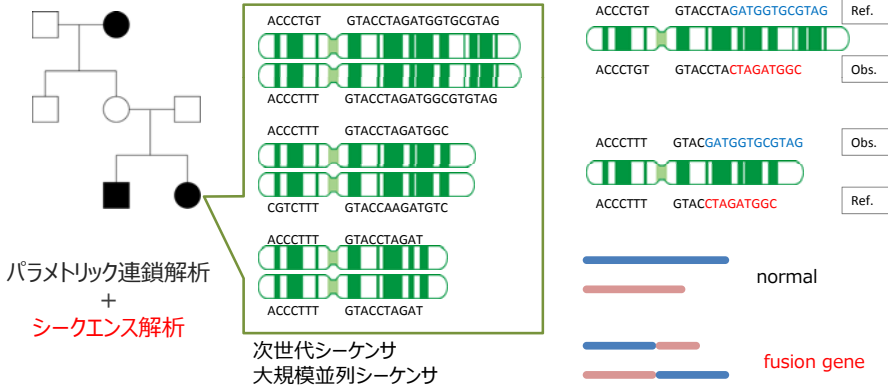


病気の子供と健康な両親のデータを集める

病気の原因SNPであれば、分離の法則に反してリスクアレルが優先的に伝わっているように見える

データの少ないメンデル型疾患（希少疾患）

遺伝的には強い関連性をもつが、ある特別の家系にしかおきない

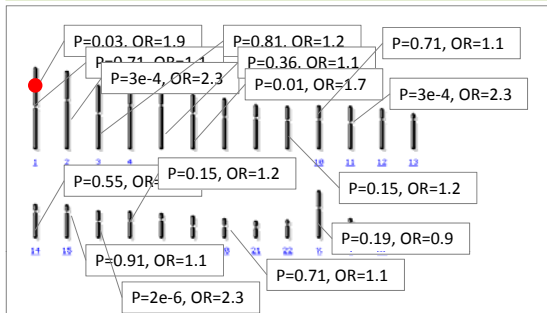


Genome-Wide Association Study, GWAS

ゲノムワイド関連研究

疾患の有無や薬効などに関連する遺伝的要因を網羅的に探索する

ケース群	コントロール群	A1について優性		
AA+AT	TT	AA+AT	TT	合計
48	52	48	52	100
32	68	32	68	100
合計	80	80	120	200

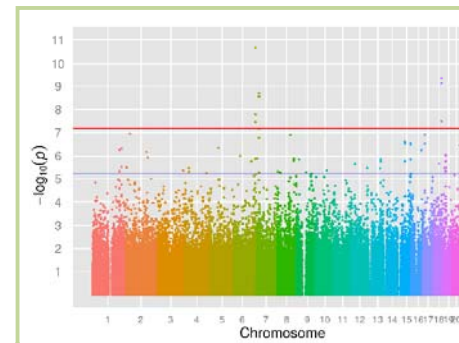


- 2,000,000 SNP程度について解析することが多い
- 大量の検定結果を比較するためタイプIIの誤りの多量性や検出力が問題

ゲノムワイド関連研究結果の視覚化

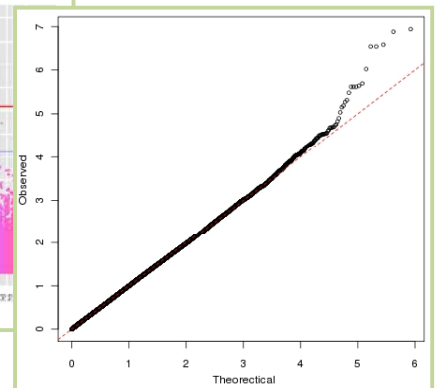
マンハッタンプロット

染色体、SNPの物理的位置順に（横軸）
P値の常用対数の負値（縦軸）をプロット



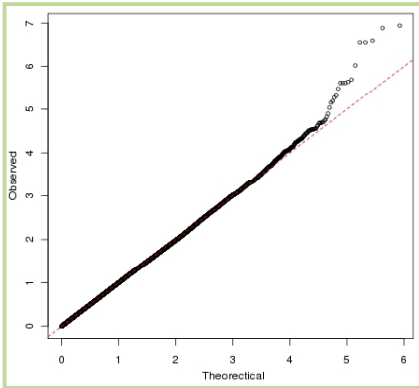
QQプロット

横軸に一様分布の確率点
縦軸に観測P値の確率点をプロット



QQプロット

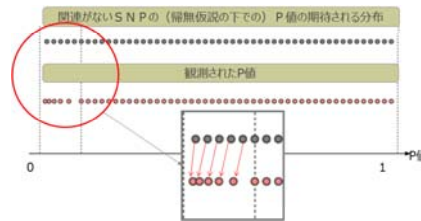
横軸に一樣分布の確率点
縦軸に観測P値の確率点をプロット



GWASでみつかるSNPは数十個

ほとんどのSNPは関連性がない

P値は一樣分布に従う



多重比較法 (multiple comparison method)

検定の規模や検定の結果を考慮して有意水準を設定する方法

検定の規模に基づいて設定する

検定の結果を考慮して設定する

ボンフェローニの不等式に基づいた補正法
(ボンフェローニの補正法、Bonferroni's correction)

FDRに基づく有意水準の設定法
(BH法、FDR=False Discovery Rate)

全体の誤りを5%におさえたいとき

QQプロット
(Quantile-Quantile プロット)

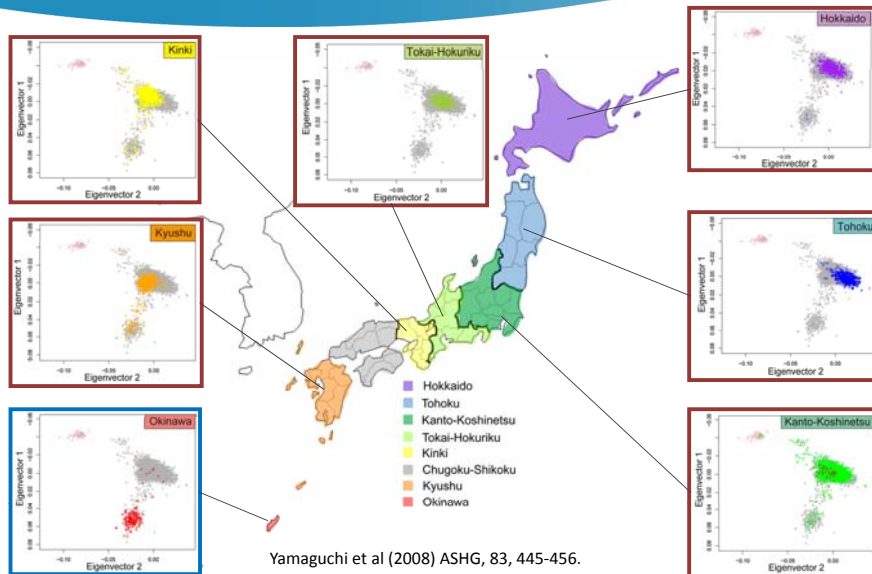
$$1\text{回の検定の有意水準} = \frac{0.05}{\text{SNP座位数}}$$

Permutation test

100万座位の場合、 $0.05/1,000,000=5E-8$

実験前 (研究デザイン) : ボンフェローニの補正法
実験後 : BH法、QQプロット

GWASデータと主成分分析
基本的な統計解析で人種や出身地を特定できる

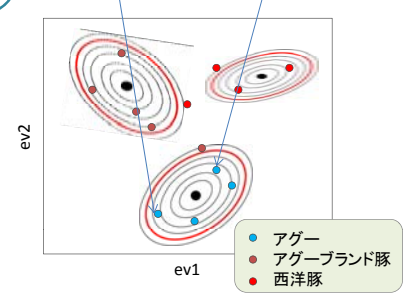
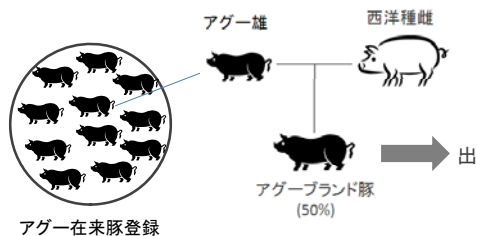
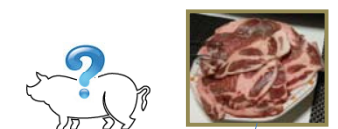


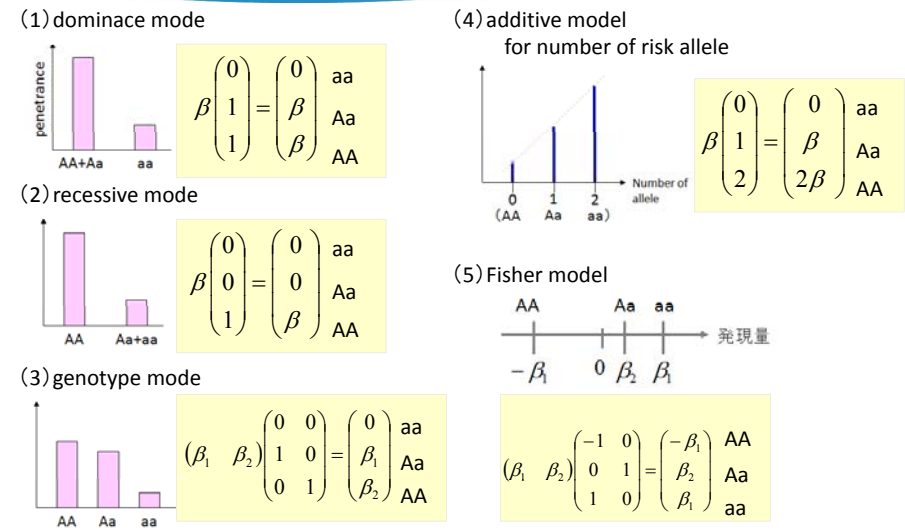
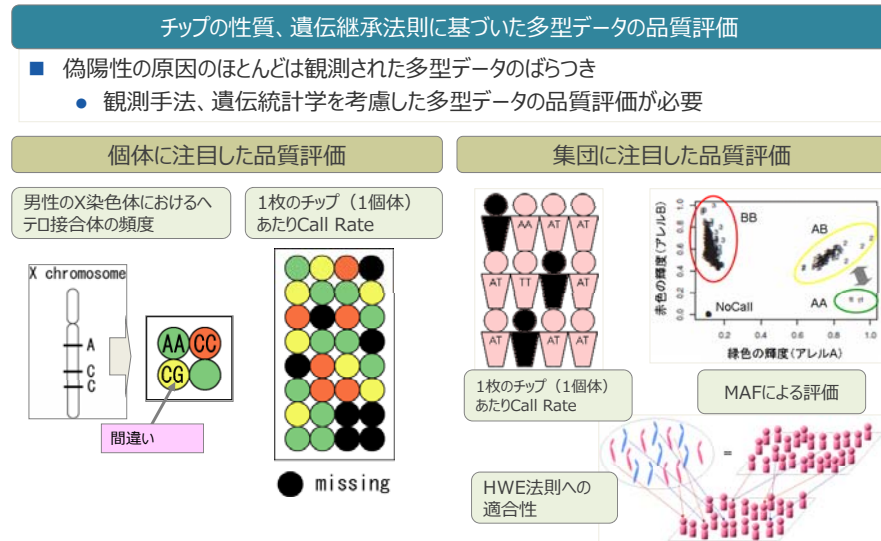
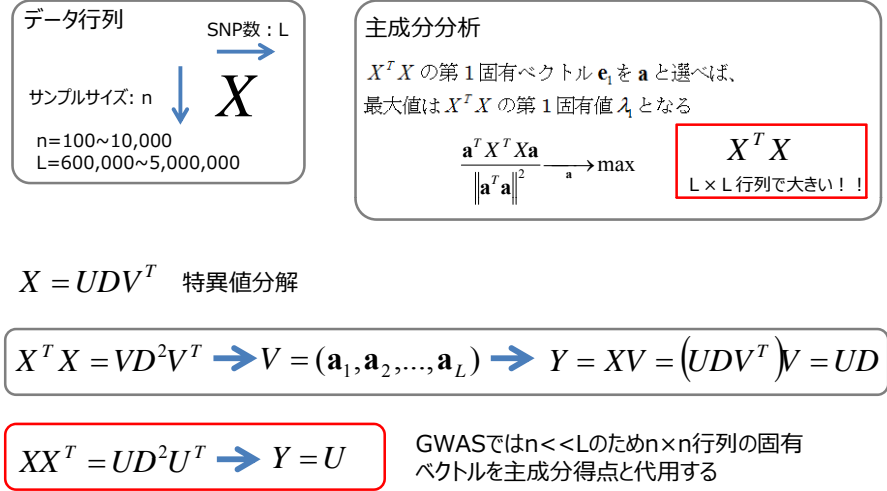
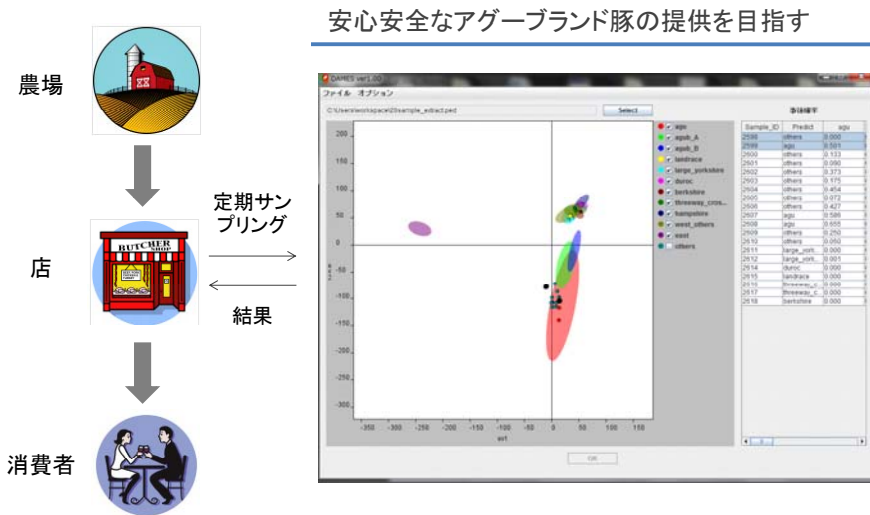
GWASデータと主成分分析
アグーブランド豚の識別方法

沖縄ブランド保護: GWASチップを用いたアグーブランド豚識別



アグー豚(左)と西洋種(右)
稲嶺ら(2009)沖縄県畜産研究センター研究報告, 47.





日常のなかの遺伝統計学

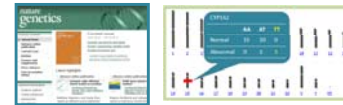
- 1. DTC遺伝子検査サービス
- 2. 無侵襲出生前遺伝学的検査



ヘルスケアへのゲノム研究活用

ヘルスケア

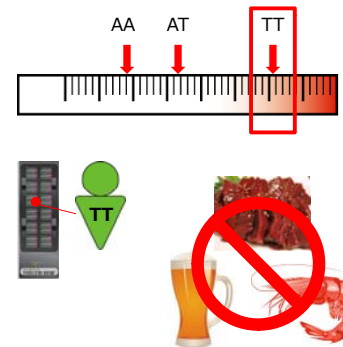
先行研究に基づいて病気のなりやすさを予測して健康維持に利用する



23andMe (<https://www.23andme.com/>)

Google venture

組織を送るとGWASチップでgenotypingする。その結果に基づいてヘルスケア情報を取得できる。



DTC遺伝子検査サービス

DTC遺伝子検査サービス

遺伝子検査結果が直接消費者に提供される (Direct To Consumer) サービス

- ①キットを購入
- ②唾液を採取する
- ③郵送する
- ④自分の結果が反映されたサイトを見ることができる



医師を介さずに遺伝子検査の結果をうけると

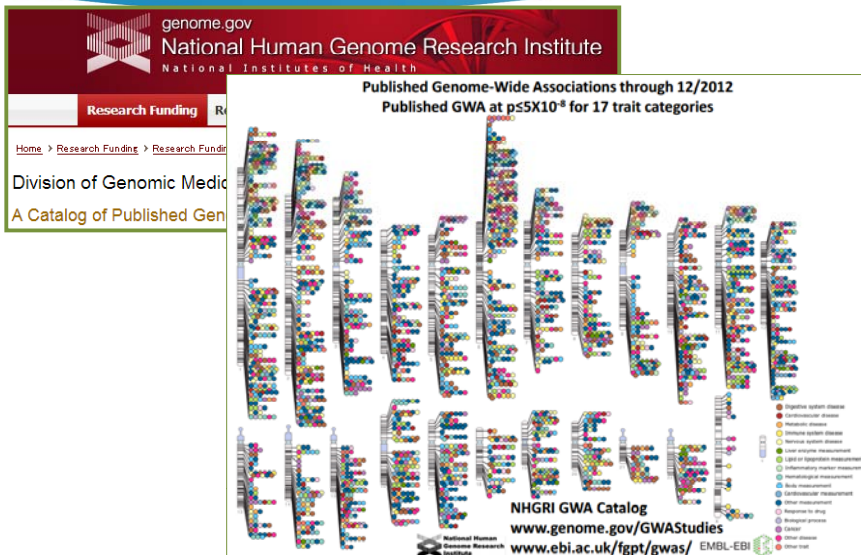
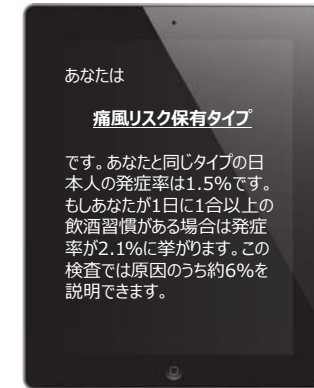
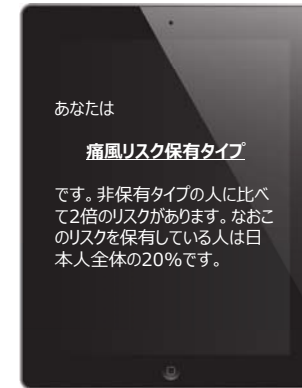
DTC遺伝子検査サービス





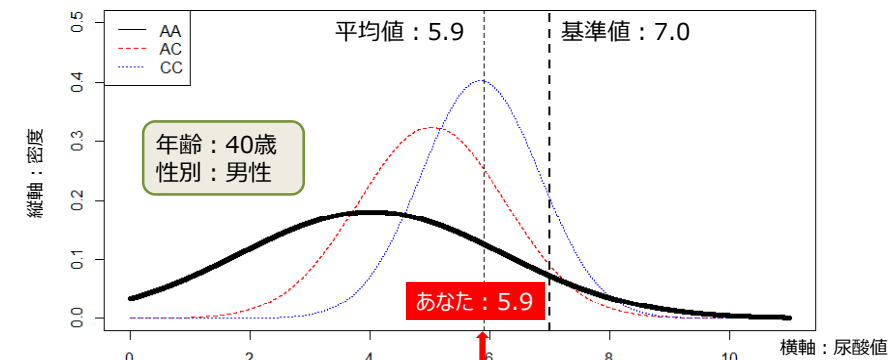
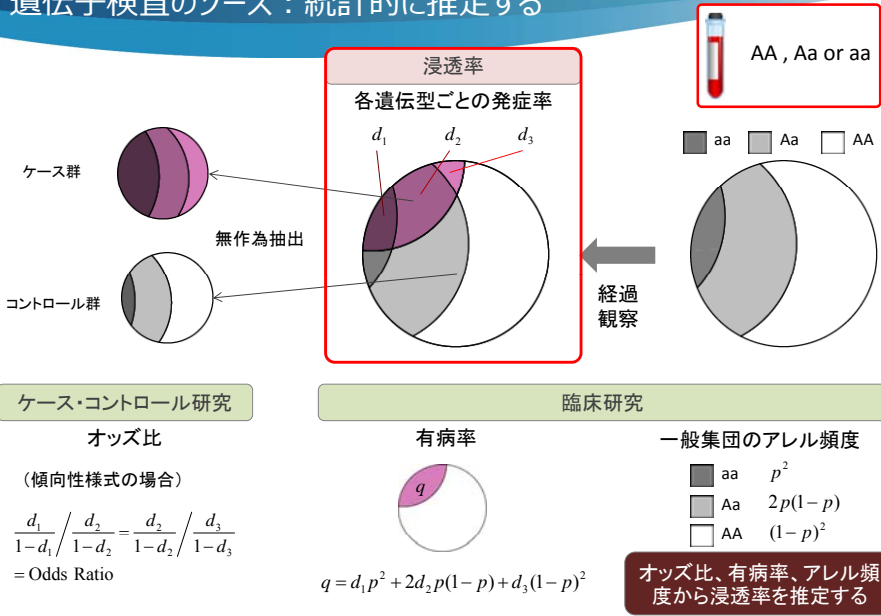
薬局で買える遺伝子検査キットで
抗リウマチ薬メトトレキサートの有効投与量を判断で
きる

検査結果を持って医師のところへ説明を求めにくる

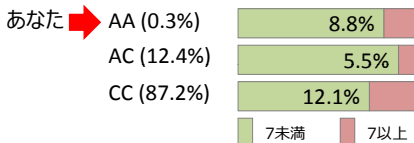


Initial Sample Description	Replication Sample Description	Region	Reported Gene(s)	Mapped Gene(s)	Strongest SNP-Risk Allele	Context	Risk Allele Frequency in Controls	P-value	OR or beta-coefficient and [95% CI]	Platform [SNPs passing QC]
2,115 European ancestry cases, 67,259 European ancestry controls	1,036 European ancestry cases	4q22.1	ABC2	ABC2	rs1481012-A	intron	0.8903	2×10^{-32}	.5484 [0.46-0.64] unit decrease	Affymetrix, Illumina [2,538,056] (imputed)
		4p16.1	SLC2A9	SLC2A9	rs4475146-A	intron	0.2237	4×10^{-26}	.4682 [0.38-0.55] unit decrease	
	18q21.32	NR	CTBP2P3 - HMG1P31	rs2168303-A		0.9953	2×10^{-7}	1.9403 [1.21-2.67] unit decrease		
	4q35.1	NR	CLDN24 - CDKN2AIP	rs17074636-T		0.9862	7×10^{-7}	.7091 [0.43-0.99] unit decrease		

- Risk Allele Frequency
 - 一般集団でリスクをどのくらい持っている人がいるのか
- P-value
 - この結果はどの程度有意なものなのか
- OR
 - このリスクと病気との関連性の強さはどのくらいか



あなたのタイプと基準値7以上の割合



あなたと同じタイプで尿酸値があなた以上の人は19.7%しかいません。適切な生活習慣の改善を心がけましょう。

NATURE | VOL 428 | 1 APRIL 2004 | www.nature.com/nature

Medical genetics

A marker for Stevens-Johnson syndrome

Wen-Hung Chung*, Shuen-Iu Hung†, Hong-Shang Hong*, Mo-Song Hsieh‡, Li-Cheng Yang*, Hsin-Chun Ho*, Jer-Yuarn Wu†§, Yuan-Tsong Chen†¶

We studied 44 patients with carbamazepine-induced Stevens-Johnson syndrome (CBZ-SJS), including five with overlapping toxic epidermal necrolysis, in whom the clinical morphology fulfilled Roujeau's diagnostic criteria¹ (see supplementary information). Patients suffered from widespread skin rashes with blisters, skin detachment and mucosa involvement. Controls included 101 patients who had been on carbamazepine for at least three months without adverse reaction ('CBZ-tolerant') and 93 normal individuals.

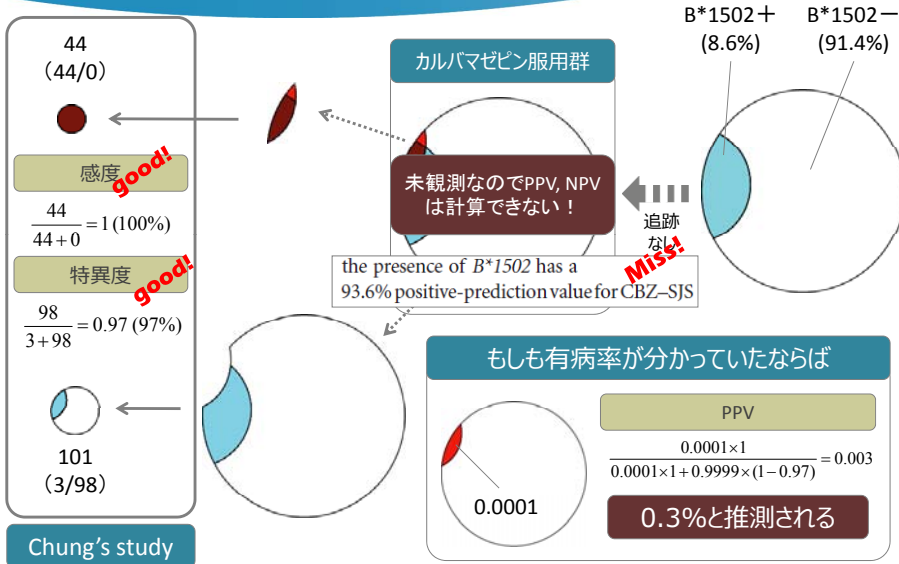
Table 1 Frequency of HLA alleles in patients with Stevens-Johnson syndrome

HLA allele	CBZ-SJS	CBZ-tolerant	Normal
B*1502	44 (100%)	3 (3%)*	8 (8.6%)†
Cw*0801	41 (93.2%)	17 (16.8%)	13 (14%)
A*1101	36 (81.8%)	51 (50.5%)	53 (57%)
DRB1*1202	33 (75%)	12 (11.8%)	18 (19.4%)
B*1502, Cw*0801	41 (93.2%)	3 (3%)	7 (7.5%)
B*1502, A*1101	36 (81.8%)	2 (2%)	6 (6.5%)
B*1502, DRB1*1202	33 (75%)	1 (1%)	5 (5.4%)
B*1502, Cw*0801, A*1101, DRB1*1202	29 (66%)	0 (0%)	3 (3.2%)

*Frequencies (by number and percentage) of individual or combined loci of the B*1502 ancestral haplotype are shown in patients with carbamazepine-induced Stevens-Johnson syndrome (CBZ-SJS; n=44) and in carbamazepine-tolerant (n=101) and normal subjects (n=93). For methods, see supplementary information. †Odds ratio (CBZ-SJS/normal): 2.504 (95% CI, 1.26–49.522); corrected P value P₁=3.13 × 10⁻¹⁰. ‡Odds ratio (CBZ-SJS/normal): 895 (95% CI, 50–15,869); P₂=1.38 × 10⁻¹⁹.

When the CBZ-tolerant group is used as the control, the presence of B*1502 has a 93.6% positive-prediction value for CBZ-SJS, whereas its absence has a negative-prediction value of 100%. In a test for CBZ-SJS, the HLA-B*1502 allele should therefore have 100% sensitivity and 97% specificity.





日常のなかの遺伝統計学

1. DTC遺伝子検査サービス
2. 無侵襲出生前遺伝学的検査



Whole Genome Sequencing

The \$2 Million Genome

James Watson, codiscoverer of the structure of DNA, now has a copy of his very own genome. Will you be next?

By Emily Singer FRIDAY, JUNE 01, 2007

On Thursday, James Watson was handed a DVD containing his entire genome, sequenced in the past few months by 454, a company based in Branford, CT, that's developing next-generation technologies for efficiently reading the genome. At a cost of \$2 million, 454 sequenced Watson's genome for roughly an order of magnitude less than it would have cost using traditional machines. While this is still too expensive for the average Joe, experts say that the advance marks a major milestone toward personal-genome sequencing--and more-personalized medicine--for all.

"We've heard people talking about personalized medicine for the last year or two, but this is the first concrete incarnation of that whole era," says George Weinstock, codirector of the Human Genome Sequencing Center at Baylor College of Medicine, in Houston. Scientists at Baylor collaborated on the genome project.

The \$2 million and two months that it took to sequence Watson's genome is a far cry from the more than ten years and \$3 billion required for the Human Genome Project's reference genome, released in 2003. Scientists ultimately hope to bring the cost down to less than \$10,000, a target price that many believe will be the turning point in genomic medicine. At that price, many people could afford to have their genomes sequenced, and doctors could then use that data to give their patients more-personalized medical advice.

At a press conference at Baylor on Thursday announcing the completion of the genome, 454's founder, Jonathan Rothberg, compared the company's sequencing

DNA's daddy: James Watson, pictured above, predicted the

Massively parallel sequencer

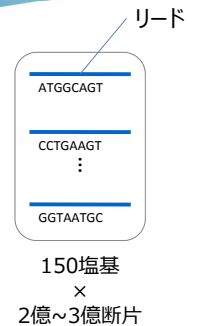
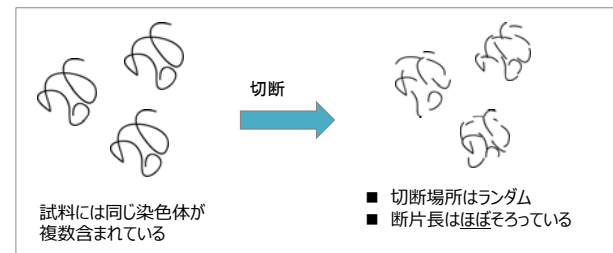


~3,000,000,000座位

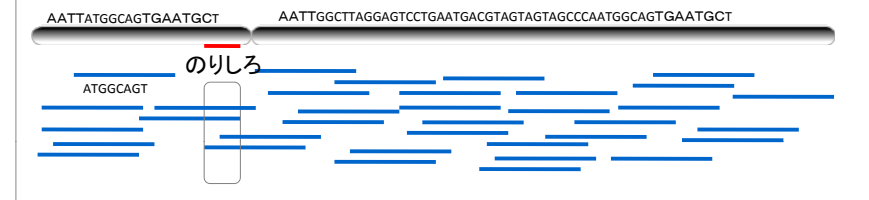
現在30万円程度

数年後には10万円?

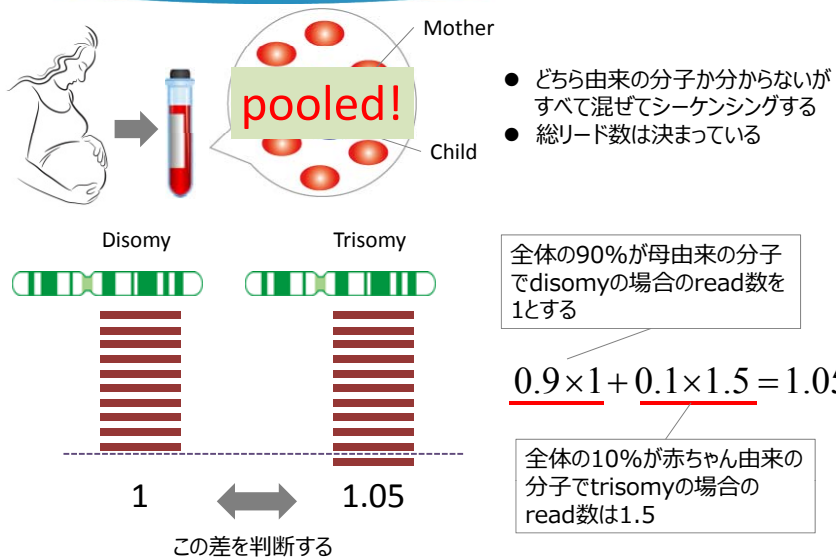
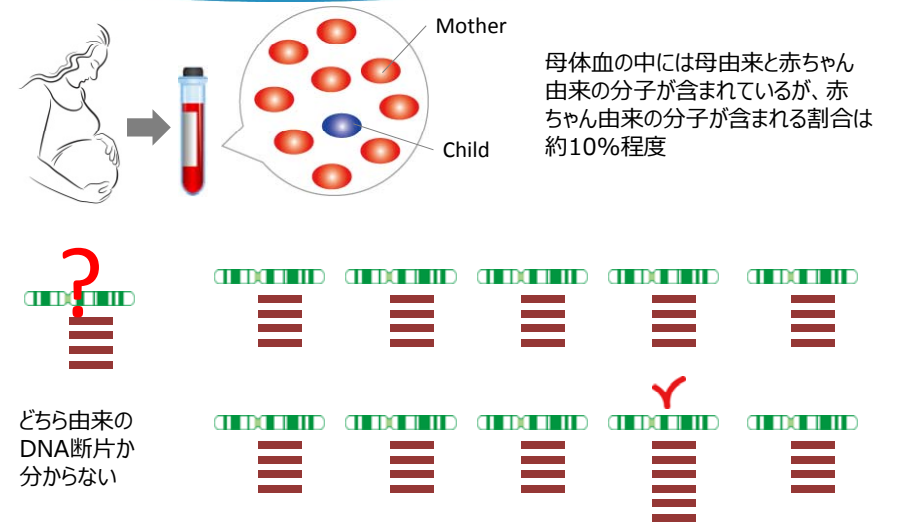
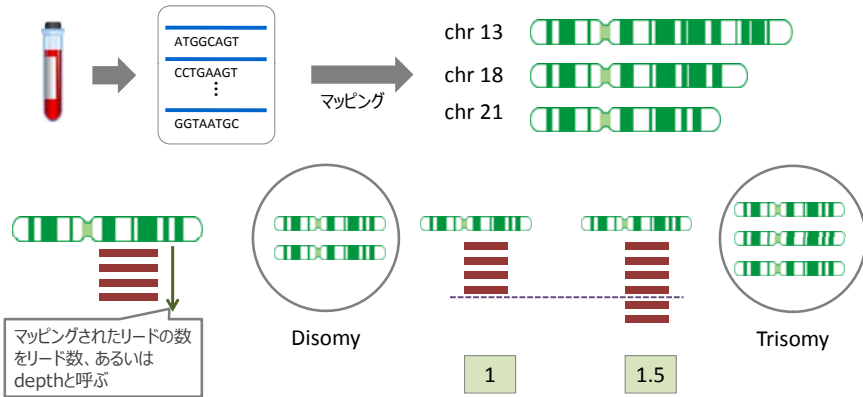
大規模並列シーケンサ



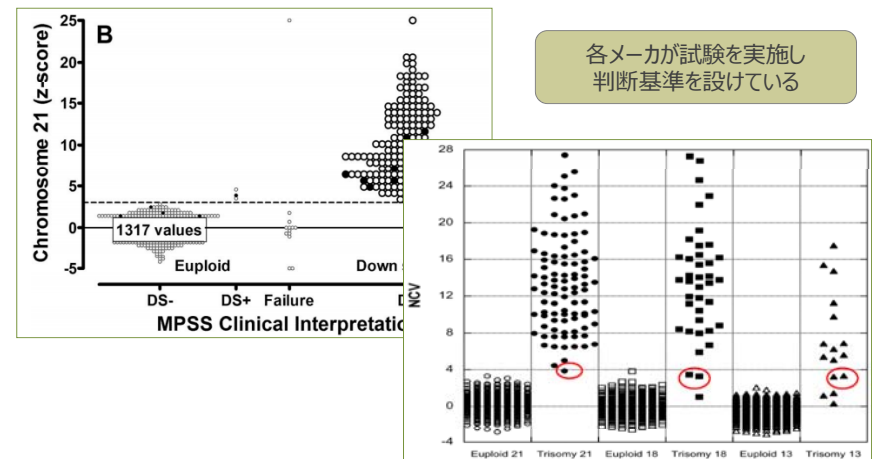
染色体地図との配列比較



- 母体血を利用した赤ちゃんの遺伝学的検査
 - 染色体数の異常を検査する
 - 13、18、21番染色体を対象とし、トリソミーか否かを判定する



Palomaki et al. Genet Med. 2011 Nov;13(11):913-20.



Bianchi et al. Obstet Gynecol. 2012 May;119(5):890-901.



新型出生前診断

検査精度 [編集]

シーケノム社のMPS法については、米国をはじめとする世界の27施設で計1988例を対象としたコホート研究が行われた。21トリソミーの出生前診断では感度sensitivityが99.1%、特異度specificityが99.9%という高い精度が報告されている。ただし、陽性的中率、陰性的中率は対象群の罹患率によって変化することに注意が必要である。また、妊婦の年齢が低いほど、結果が正しい確率は下がる[2]。

感度

赤ちゃんがTrisomyである場合に母体血から陽性と判断できる割合

特異度

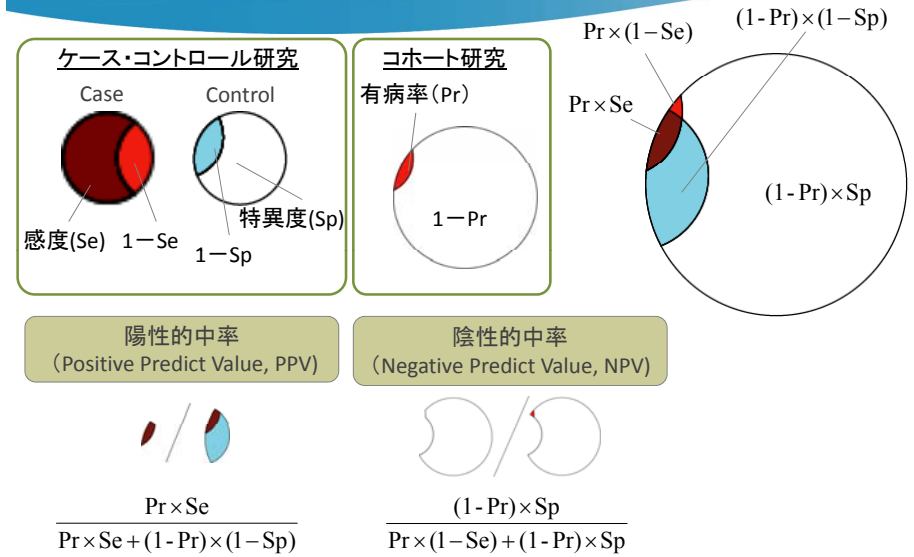
赤ちゃんがTrisomyでない場合に母体血から陰性と判断できる割合

陽性的中率

母体血検査から陽性と判断された赤ちゃんがTrisomyである割合

陰性的中率

母体血検査から陰性と判断された赤ちゃんがTrisomyでない割合



99%の確率で胎児は染色体数異常です。

検査の結果、陽性だったのですが。

有病率と陽性的中率の関係

年齢	有病率	PPV
若年	1/1000	49.8%
35歳以上	1/350	74.0%
40歳以上	1/100	90.9%

有病率を1/1000としてシミュレーション

お母さんは今35歳ですね。今回利用したNIPTにおける**陽性的中率は74.0%**です。つまり陽性だった人の3/4の方に染色体数異常のお子さんが生まれる確率です。逆を言えば、1/4の方には染色体数正常のお子さんが生まれることとなります。確定診断を得るために羊水検査を受けてください。

科学界はより一層の誠実さと正確さで社会の信用を取り戻す一年となるでしょう。ゲノム分野では正確なデータが取得できるGWASに再び注目が集まると考えています。GWASの商用サービスの1つであるDTC遺伝子検査は、23andMeの英国での業務再開も追い風となり市場を拡大できるでしょうか。DTC遺伝子検査は事業者だけでなく消費者にも統計リテラシーが必要です。スタージェンとは遺伝統計学のエキスパートとして社会の統計リテラシー向上に少しでも貢献できるように努力して参ります。本年もどうぞよろしくお願いたします。

株式会社ステージンは
医学・医療・医薬品に関する統計解析技術を用いた発明の技術移転や
事業化, 更には教育やデータ解析を通じて社会に貢献します。



Statistics

遺伝統計学知識のみならず実データ解析の経験も豊富です

Genetics

遺伝継承法則の視点からデータを眺めることができます

Informatics

多量のデータをストレスなく扱うことができます

Medical science

医学的視点で現実的な解釈ができます



沿革

平成11年10月	有限会社ギアマップ設立
平成15年 7月	有限会社テラーエイドに名称変更
平成16年 5月	株式会社ステージンに組織変更
平成18年 1月	東京都台東区蔵前に移転

事業部・研究所

知的財産事業部	特許権の取得・保有・運用業務
遺伝統計解析事業部	遺伝情報と形質との関連研究の支援
情報解析研究所	医療における情報解析、統計解析

従業員数 (2015年4月現在)

従業員 : 11名

鎌谷 直之 会長

医学博士 : 2名 (医師1名)
理学博士 : 2名
農学博士 : 1名



- 遺伝統計学の教科書
 - Balding DJ, Bishop M. and Cannings C. (eds) (2008) Handbook of Statistical Genetics, Third Edition, John Wiley & Sons, Ltd, Chichester, UK.
 - Ott J (1999) Analysis of Human Genetic Linkage, 3rd edition, John Hopkins University Press.
 - 鎌谷直之 (2007) 遺伝統計学入門, 岩波書店.
 - 上辻茂男 (2008) 遺伝統計学へようこそ, BTJジャーナル (計15回連載) . 15回分の連載が総集編として無料ダウンロード可. <http://biotech.nikkeibp.co.jp/btjin/#btji0905>
興味のある方メールください
- 遺伝学の読み物
 - 鎌谷直之 (2009) オンリーワン・ゲノム—今こそ『遺伝と多様性』を知ろう, 星の環会.
- GWAS関連
 - Balding, D.J. (2006) A tutorial on statistical methods for population association studies. Nature Reviews Genetics, 7, 10, 781-791.
 - plink, <http://pngu.mgh.harvard.edu/~purcell/plink/>
- ゲノム研究で重要なサイト
 - NCBI, <http://www.ncbi.nlm.nih.gov/>
 - HapMap project, <http://hapmap.ncbi.nlm.nih.gov/>
 - Genetic Analysis Software, <http://linkage.rockefeller.edu/soft/>

株式会社ステージン 遺伝統計解析事業部
〒111-0051 東京都台東区蔵前4-11-6 KUGAビル8F
tel: (03) 5835-2137 / 2138, fax: (03) 5835-2139
URL: <http://www.stagen.co.jp/>
e-mail: kamitsuji@stagen.co.jp

遺伝統計教育/統計解析/コンサルテーション/
システム開発/技術移転