

バイオフィーマティクス最前線

西郷, 浩人
九州大学大学院システム情報科学研究院情報学部門 : 准教授

<https://hdl.handle.net/2324/1664245>

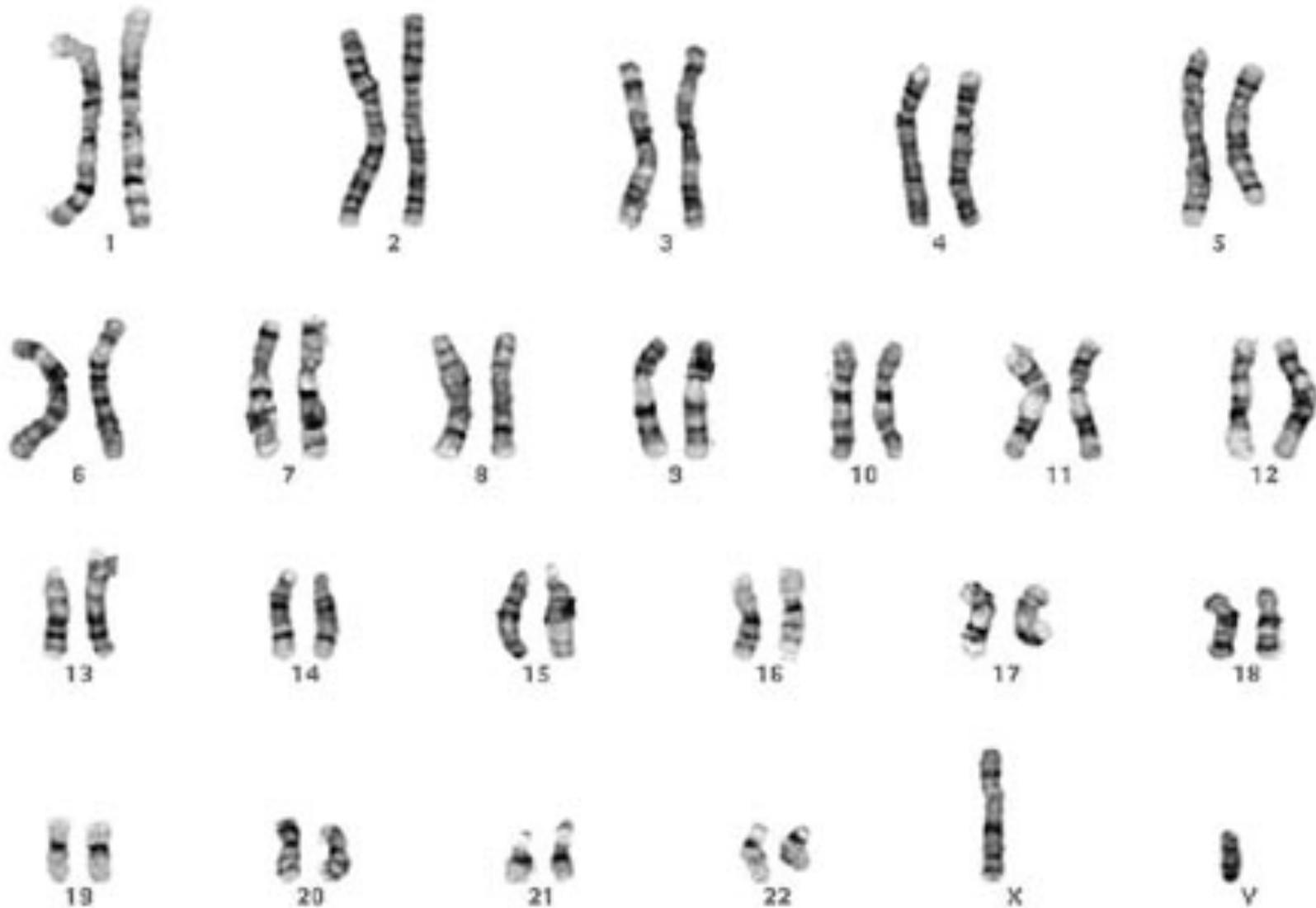
出版情報 : 九州大学大学院システム情報科学府・研究院先端サマーセミナー. 2016, 2016-08-05
バージョン :
権利関係 :

バイオインフォマティクス 最前線

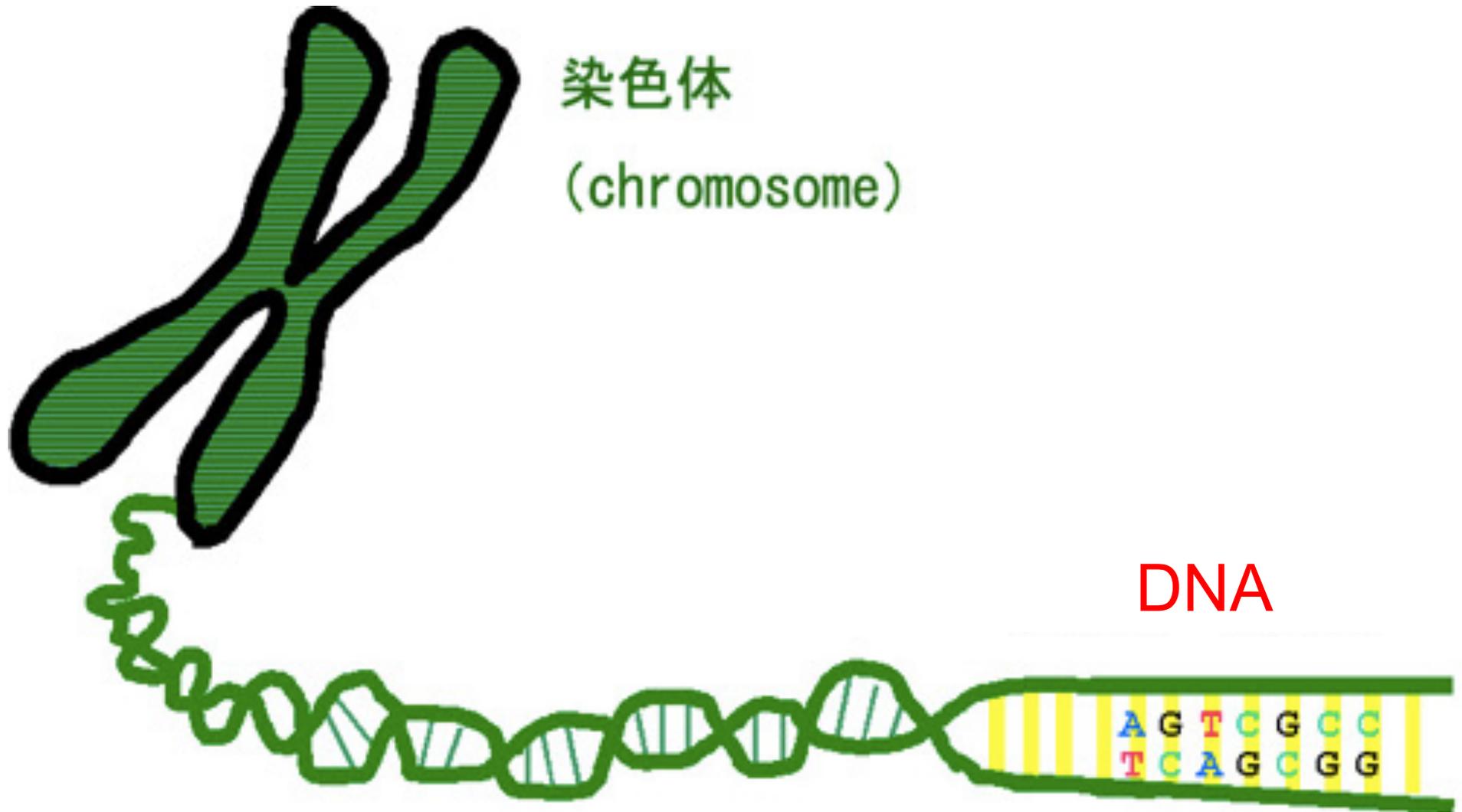
システム情報科学研究所 情報学部門
西郷浩人

遺伝情報から表現型まで

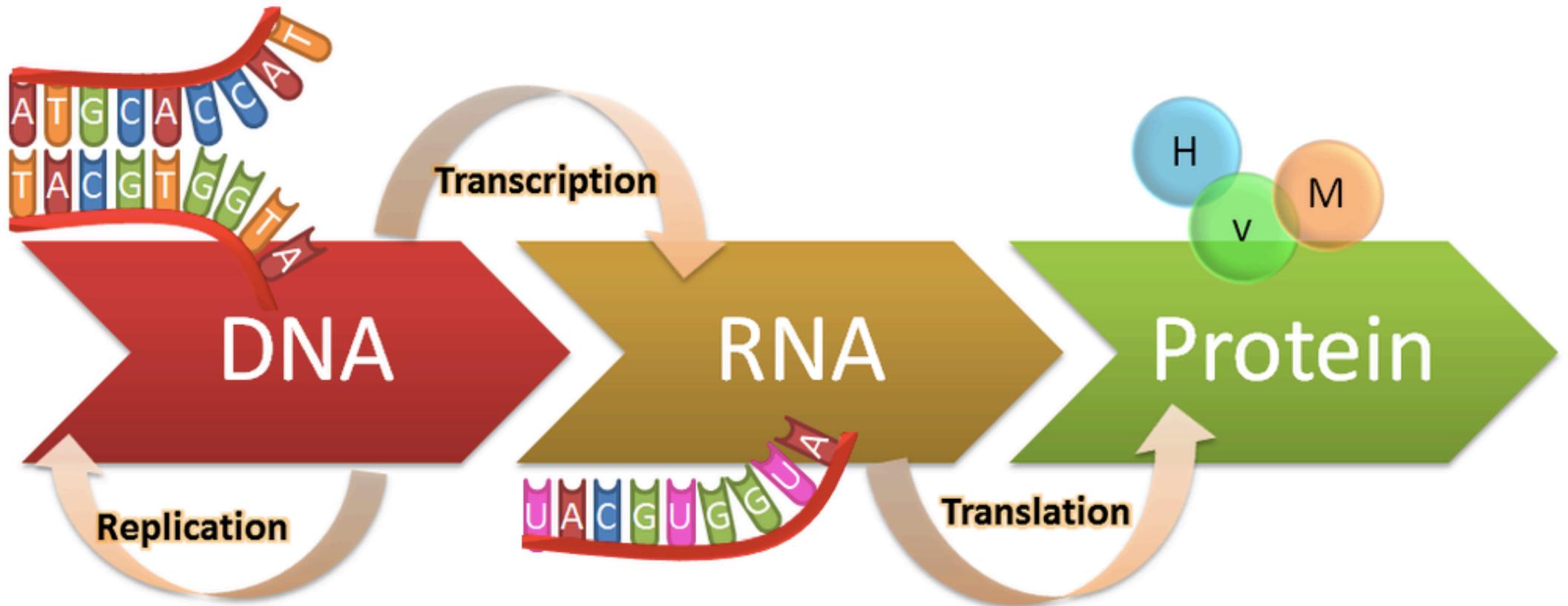
23本の染色体



染色体上のDNA



セントラルドグマ

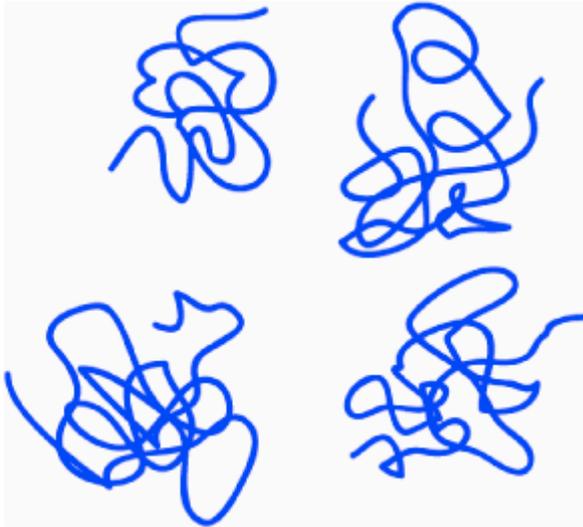


Universal Genetic Code

		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G

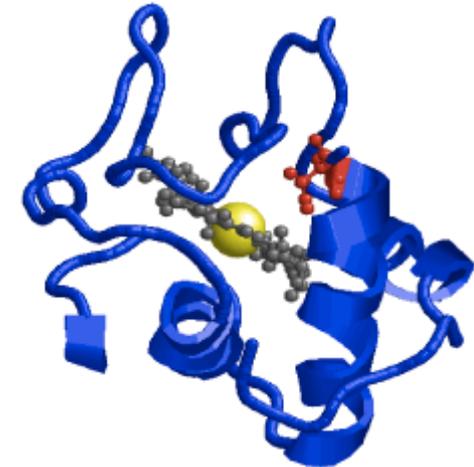
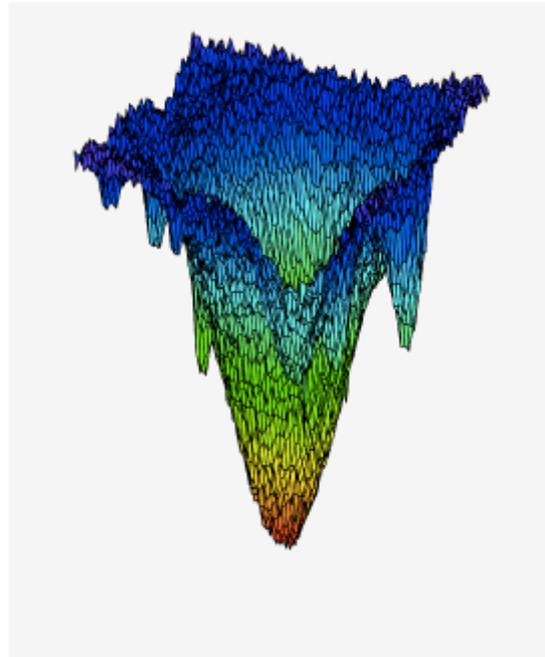
Third letter

タンパク質のおりたたみ



Unfolded states

An astronomical number of conformations. A 100 residue protein, with 2 conformations per residue has 2^{100} or 10^{30} different conformations



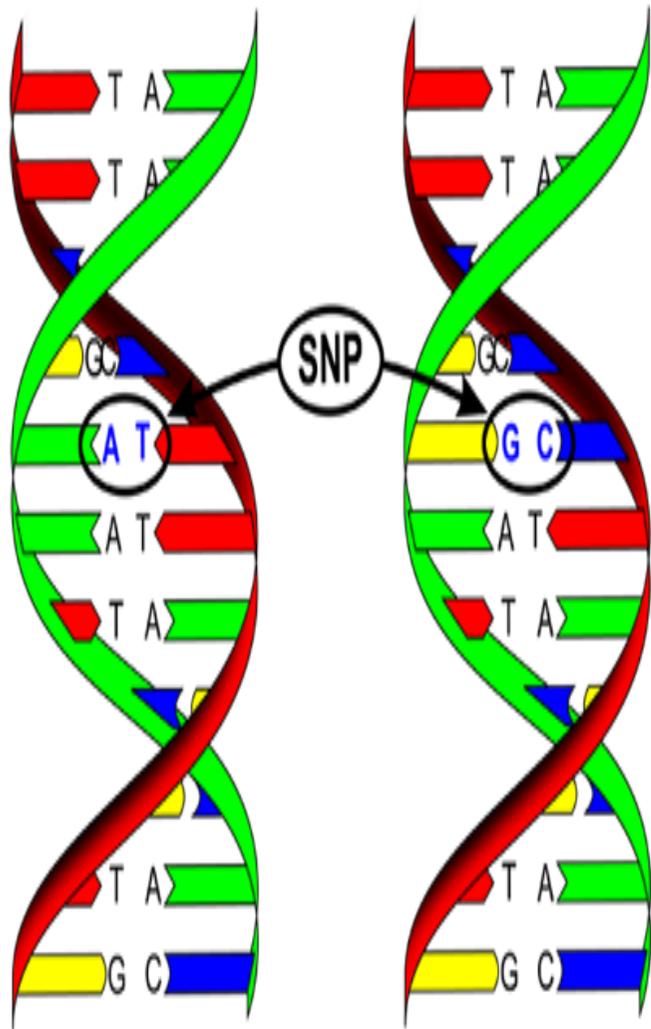
Folded or Native State

A single conformation (or, more correctly, a collection of similar conformational sub-states)

どれくらいの情報量？

- 大腸菌：460万塩基
 - 約4千遺伝子
- 酵母：1500万塩基
 - 約6千遺伝子
- ヒト：30 億塩基
 - 約3万遺伝子

Single Nucleotide Polymorphism (SNP)



DNA ≒ 30億塩基対

↓ 各個人間で**約0.1%**異なる部分

SNP ≒ 300万塩基

SNPは**個人差**を生み出す

肌の色



体質



ライフサイエンス データベース

- Sequence (DNA, Amino Acid)
- Genomics
- Mutation/Polymorphism
- Protein domain/family
- Protein 3D
- Proteomics (Mass Spectrometry)
- Metabolic networks
- Expression
- Disease/Drug

歷史

- 1950s
 - Watson & Click DNA model
- 1960s
 - Sequence Alignment
- 1970s
 - PDB (Protein Data Bank)
 - Sanger sequencing
- 1980s
 - GenBank Database
 - PCR (Polymerase Chain Reaction)
 - NCBI

歷史

- 1990s
 - WWW
 - BLAST
 - Bacterial genome
 - Yeast genome
- 2000s
 - **Draft Human genome**
- 2010s
 - Next-generation sequencer
 - Personal genome

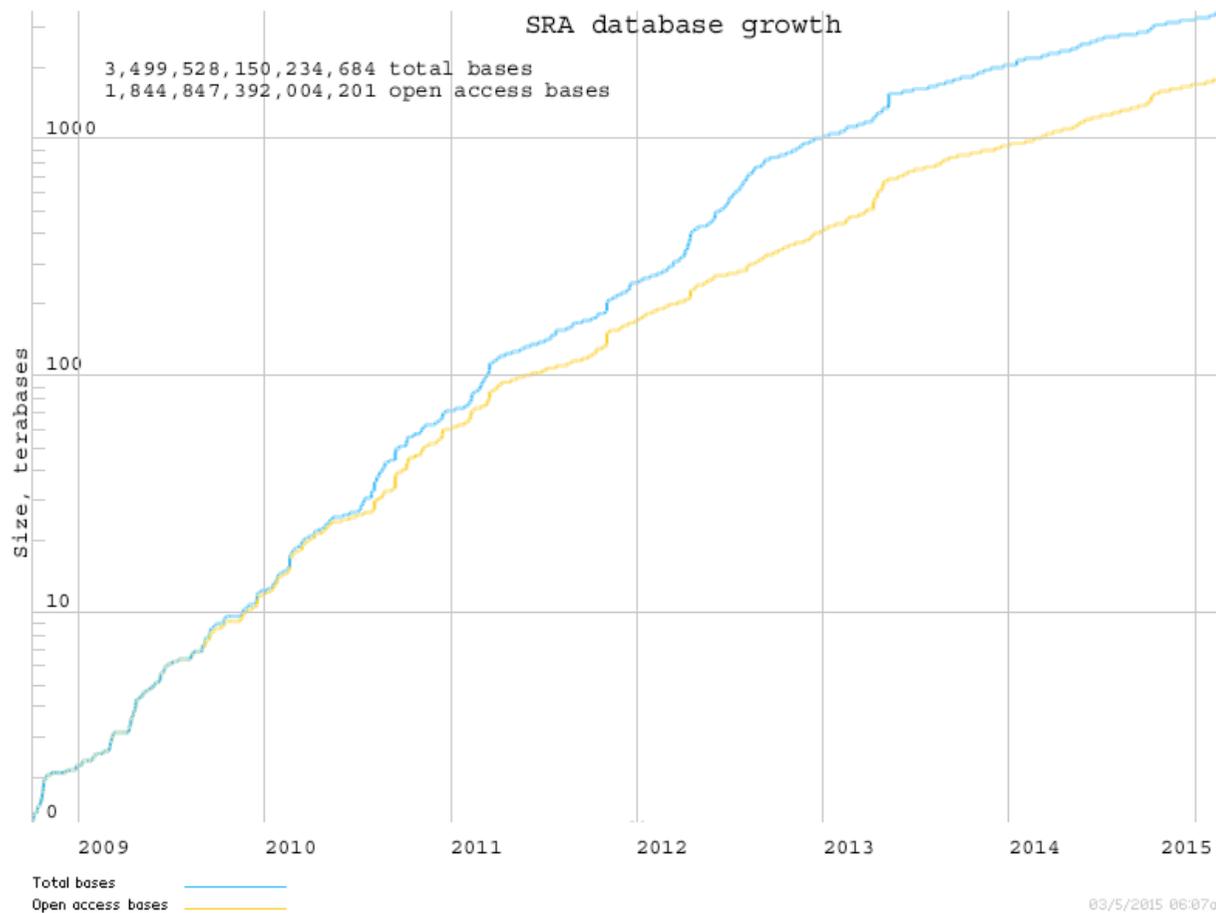
Draft human genome in 2000



Milestones contd.

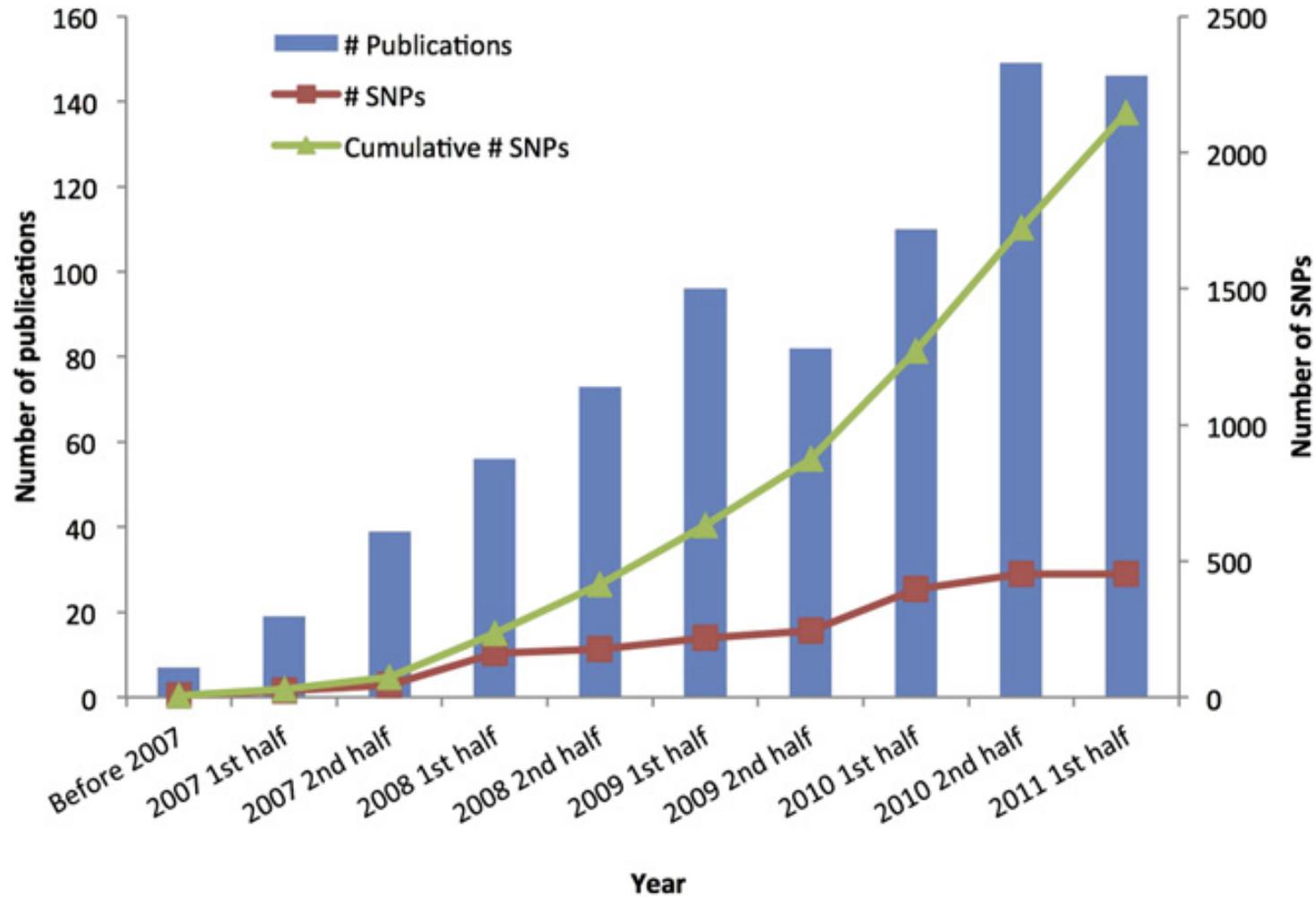
- 1990s
 - WWW
 - BLAST
 - Bacterial genome
 - Yeast genome
- 2000s
 - Draft Human genome
- 2010s
 - **Next-generation sequencer**
 - GWAS (Genome-Wide Association Study)

NGS (次世代シーケンサー) によるデータの増加



NCBI, EBI, DDBJ に登録された配列データの総量の増加

SNPの解読の進展



遺伝子健康診断の時代の到来

- 遺伝子健康診断を開始している企業の例
 - 23andme (米google)
 - yahoo!
 - dena
 - gene quest
- 宮城県、岩手県では震災復興の一環として個人のゲノムを調べて健康管理に活かす事業を国が展開
- イギリス、アイルランド等では数千人単位でゲノムを調べて国民に適した医療の研究

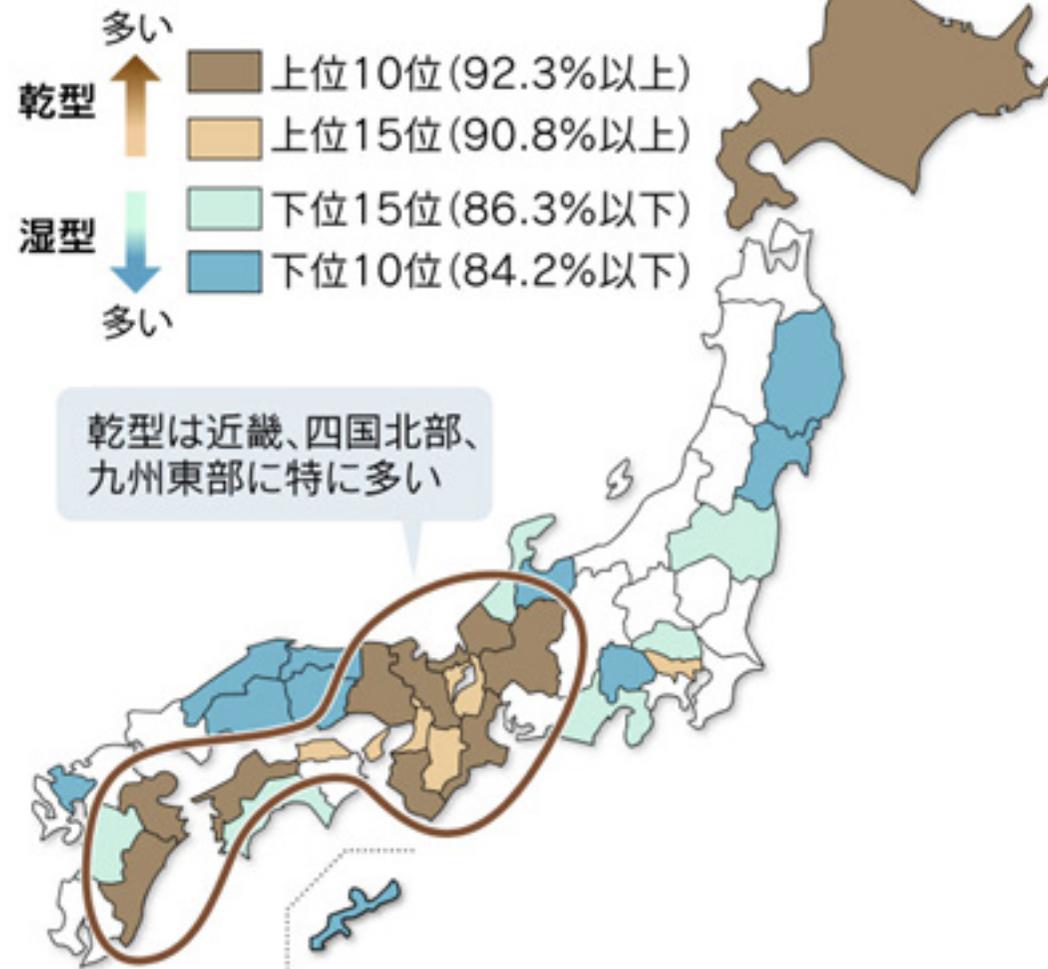
遺伝子健康診断の時代の到来

- 遺伝子診断でわかる表現型の例（まだ少ない）
 - アルコール分解酵素のあるなし
 - 耳垢の型（乾いている、湿っている）
 - 1型糖尿病の発症率
 - 光くしゃみ反応
- 保険が適用される遺伝病
 - 「デュシェンヌ型筋ジストロフィー」
 - 「福山型筋ジストロフィー」
 - 「脊髄性筋萎縮症」など



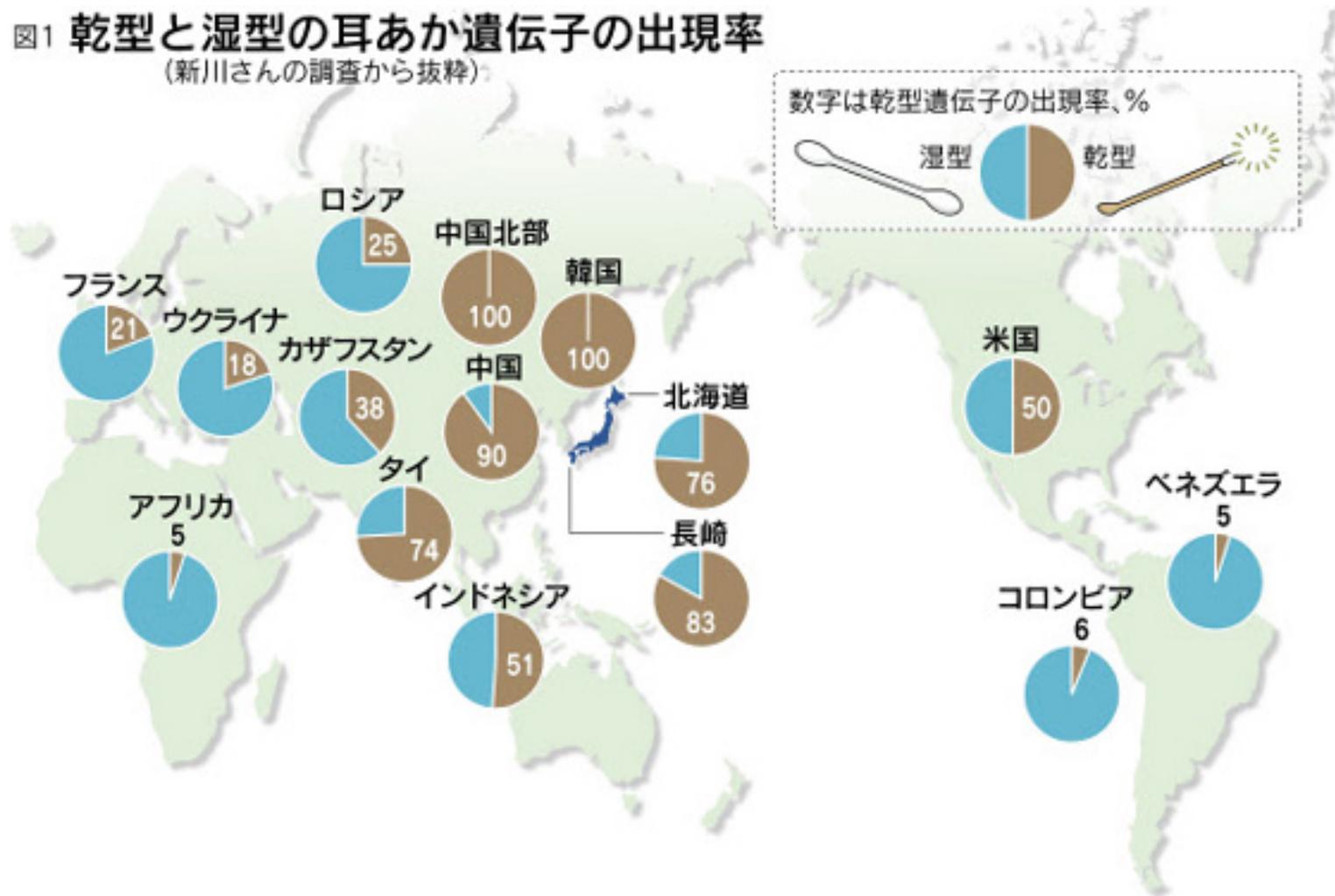
耳垢型を決めるSNP(rs17822931)の分布

図5 乾型遺伝子の出現率



耳垢型を決めるSNP(rs17822931)の分布

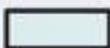
図1 乾型と湿型の耳あか遺伝子の出現率
(新川さんの調査から抜粋)

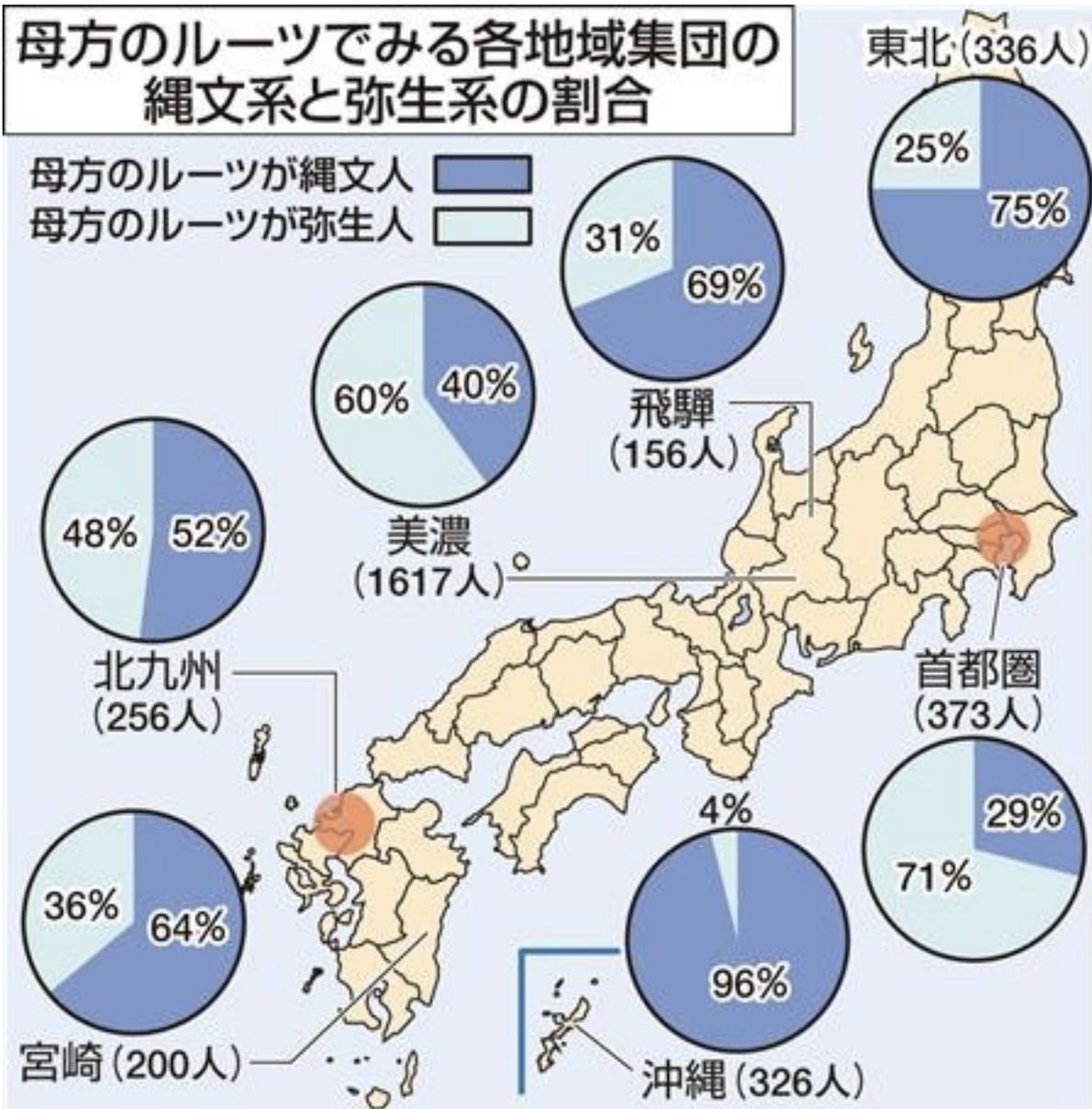


アルコールの分解を決める SNP(rs671)の分布



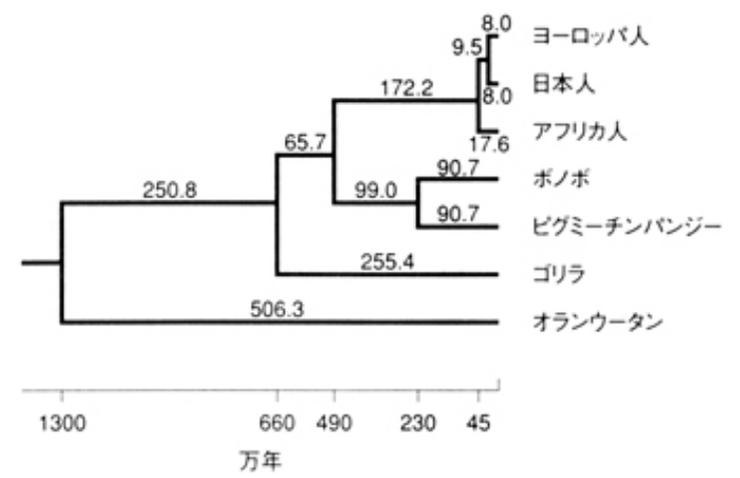
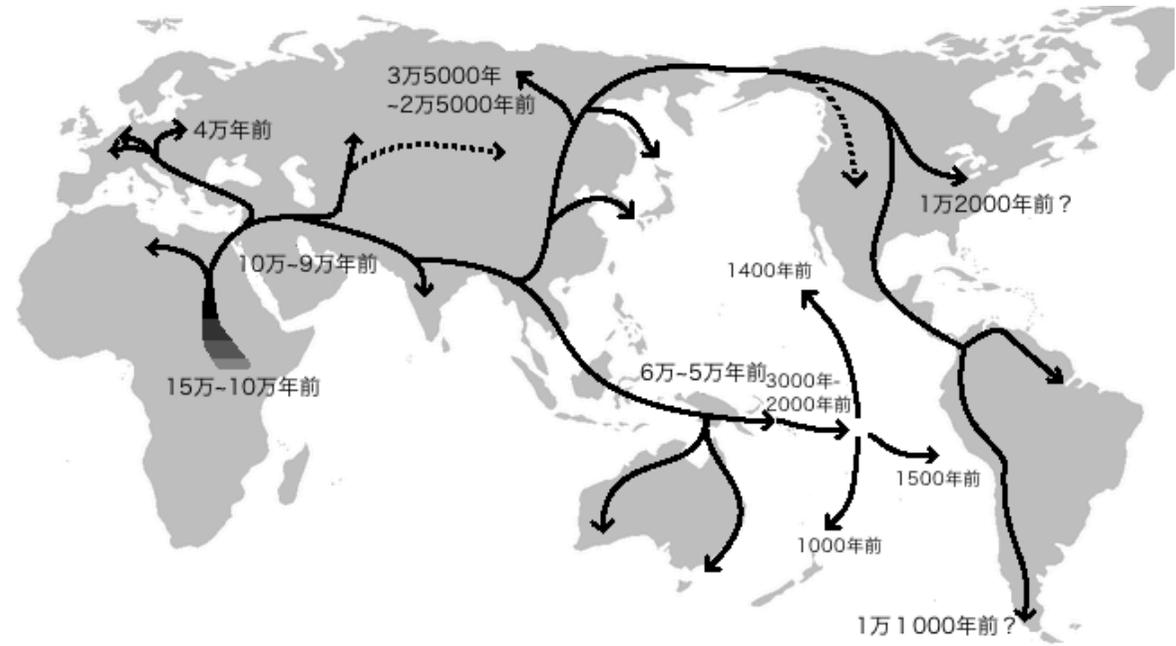
母方のルーツでみる各地域集団の縄文系と弥生系の割合

母方のルーツが縄文人 
 母方のルーツが弥生人 



2008年08月04日 産経新聞ニュース

遺伝子の違いから自分の起源を知る



その他

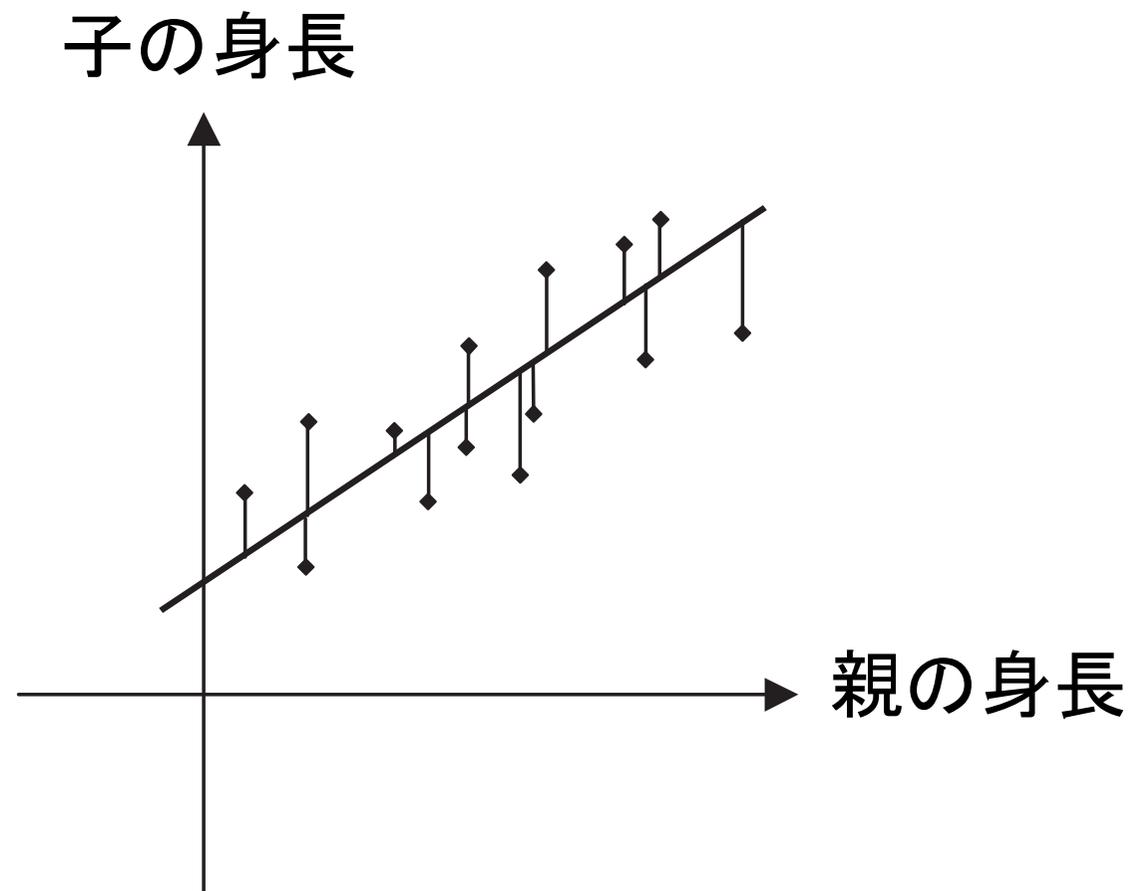
ゲノムシーケンシングのコストは1人10万円

SNPタイピングのコストは1~2万円

説明出来ない遺伝率の謎

表現型の遺伝可能性 (Heritability)

親の身長と子の身長の相関が高ければ、身長 (表現型) は遺伝すると考えられる。



ゲノムワイド相関解析 (GWAS)

目標は、表現型を遺伝子型で説明すること

子の身長(表現型)

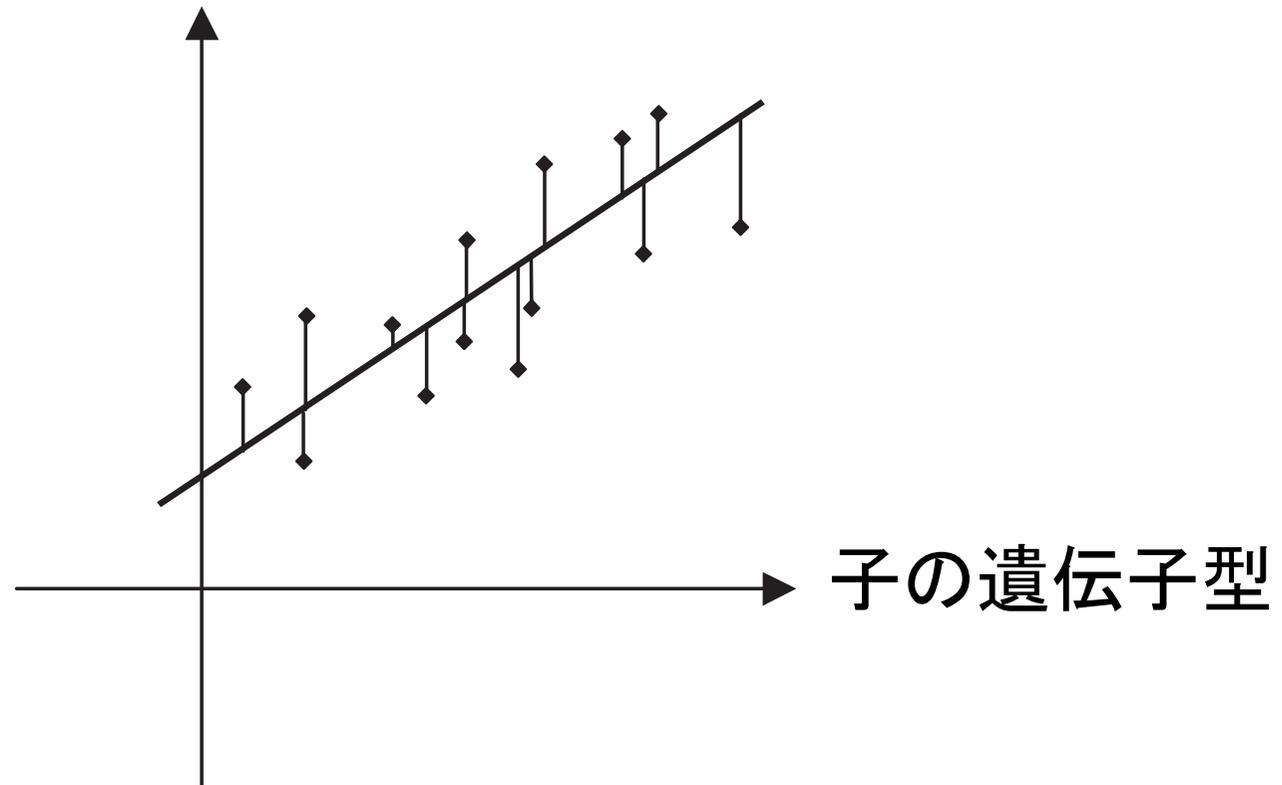


Table 1. Population Variation Explained by GWAS for a Selected Number of Complex Traits

Trait or Disease	h² Pedigree Studies	h² GWAS Hits^a	h² All GWAS SNPs^b
Type 1 diabetes	0.9 ⁹⁸	0.6 ^{99, c}	0.3 ¹²
Type 2 diabetes	0.3–0.6 ¹⁰⁰	0.05–0.10 ³⁴	
Obesity (BMI)	0.4–0.6 ^{101,102}	0.01–0.02 ³⁶	0.2 ¹⁴
Crohn's disease	0.6–0.8 ¹⁰³	0.1 ¹¹	0.4 ¹²
Ulcerative colitis	0.5 ¹⁰³	0.05 ¹²	
Multiple sclerosis	0.3–0.8 ¹⁰⁴	0.1 ⁴⁵	
Ankylosing spondylitis	>0.90 ¹⁰⁵	0.2 ¹⁰⁶	
Rheumatoid arthritis	0.6 ¹⁰⁷		
Schizophrenia	0.7–0.8 ¹⁰⁸	0.01 ⁷⁹	0.3 ¹⁰⁹
Bipolar disorder	0.6–0.7 ¹⁰⁸	0.02 ⁷⁹	0.4 ¹²
Breast cancer	0.3 ¹¹⁰	0.08 ¹¹¹	
Von Willebrand factor	0.66–0.75 ^{112,113}	0.13 ¹¹⁴	0.25 ¹⁴
Height	0.8 ^{115,116}	0.1 ¹³	0.5 ^{13,14}
Bone mineral density	0.6–0.8 ¹¹⁷	0.05 ¹¹⁸	
QT interval	0.37–0.60 ^{119,120}	0.07 ¹²¹	0.2 ¹⁴
HDL cholesterol	0.5 ¹²²	0.1 ⁵⁷	
Platelet count	0.8 ¹²³	0.05–0.1 ⁵⁸	

Table 1. Population Variation Explained by GWAS for a Selected Number of Complex Traits

Trait or Disease	h² Pedigree Studies	h² GWAS Hits^a	h² All GWAS SNPs^b
Type 1 diabetes	0.9 ⁹⁸	0.6 ^{99, c}	0.3 ¹²
Type 2 diabetes	0.3–0.6 ¹⁰⁰	0.05–0.10 ³⁴	
Obesity (BMI)	0.4–0.6 ^{101,102}	0.01–0.02 ³⁶	0.2 ¹⁴
Crohn's disease	0.6–0.8 ¹⁰³	0.1 ¹¹	0.4 ¹²
Ulcerative colitis	0.5 ¹⁰³	0.05 ¹²	
Multiple sclerosis	0.3–0.8 ¹⁰⁴	0.1 ⁴⁵	
Ankylosing spondylitis	>0.90 ¹⁰⁵	0.2 ¹⁰⁶	
Rheumatoid arthritis	0.6 ¹⁰⁷		
Schizophrenia	0.7–0.8 ¹⁰⁸	0.01 ⁷⁹	0.3 ¹⁰⁹
Bipolar disorder	0.6–0.7 ¹⁰⁸	0.02 ⁷⁹	0.4 ¹²
Breast cancer	0.3 ¹¹⁰	0.08 ¹¹¹	
Von Willebrand factor	0.66–0.75 ^{112,113}	0.13 ¹¹⁴	0.25 ¹⁴
Height	0.8 ^{115,116}	0.1 ¹³	0.5 ^{13,14}
Bone mineral density	0.6–0.8 ¹¹⁷	0.05 ¹¹⁸	
QT interval	0.37–0.60 ^{119,120}	0.07 ¹²¹	0.2 ¹⁴
HDL cholesterol	0.5 ¹²²	0.1 ⁵⁷	
Platelet count	0.8 ¹²³	0.05–0.1 ⁵⁸	

Table 1. Population Variation Explained by GWAS for a Selected Number of Complex Traits

Trait or Disease	h² Pedigree Studies	h² GWAS Hits^a	h² All GWAS SNPs^b
Type 1 diabetes	0.9 ⁹⁸	0.6 ^{99, c}	0.3 ¹²
Type 2 diabetes	0.3–0.6 ¹⁰⁰	0.05–0.10 ³⁴	
Obesity (BMI)	0.4–0.6 ^{101,102}	0.01–0.02 ³⁶	0.2 ¹⁴
Crohn's disease	0.6–0.8 ¹⁰³	0.1 ¹¹	0.4 ¹²
Ulcerative colitis	0.5 ¹⁰³	0.05 ¹²	
Multiple sclerosis	0.3–0.8 ¹⁰⁴	0.1 ⁴⁵	
Ankylosing spondylitis	>0.90 ¹⁰⁵	0.2 ¹⁰⁶	
Rheumatoid arthritis	0.6 ¹⁰⁷		
Schizophrenia	0.7–0.8 ¹⁰⁸	0.01 ⁷⁹	0.3 ¹⁰⁹
Bipolar disorder	0.6–0.7 ¹⁰⁸	0.02 ⁷⁹	0.4 ¹²
Breast cancer	0.3 ¹¹⁰	0.08 ¹¹¹	
Von Willebrand factor	0.66–0.75 ^{112,113}	0.13 ¹¹⁴	0.25 ¹⁴
Height	0.8 ^{115,116}	0.1 ¹³	0.5 ^{13,14}
Bone mineral density	0.6–0.8 ¹¹⁷	0.05 ¹¹⁸	
QT interval	0.37–0.60 ^{119,120}	0.07 ¹²¹	0.2 ¹⁴
HDL cholesterol	0.5 ¹²²	0.1 ⁵⁷	
Platelet count	0.8 ¹²³	0.05–0.1 ⁵⁸	

Common SNPs explain a large proportion of the heritability for human height

Jian Yang¹, Beben Benyamin¹, Brian P McEvoy¹, Scott Gordon¹, Anjali K Henders¹, Dale R Nyholt¹, Pamela A Madden², Andrew C Heath², Nicholas G Martin¹, Grant W Montgomery¹, Michael E Goddard³ & Peter M Visscher¹

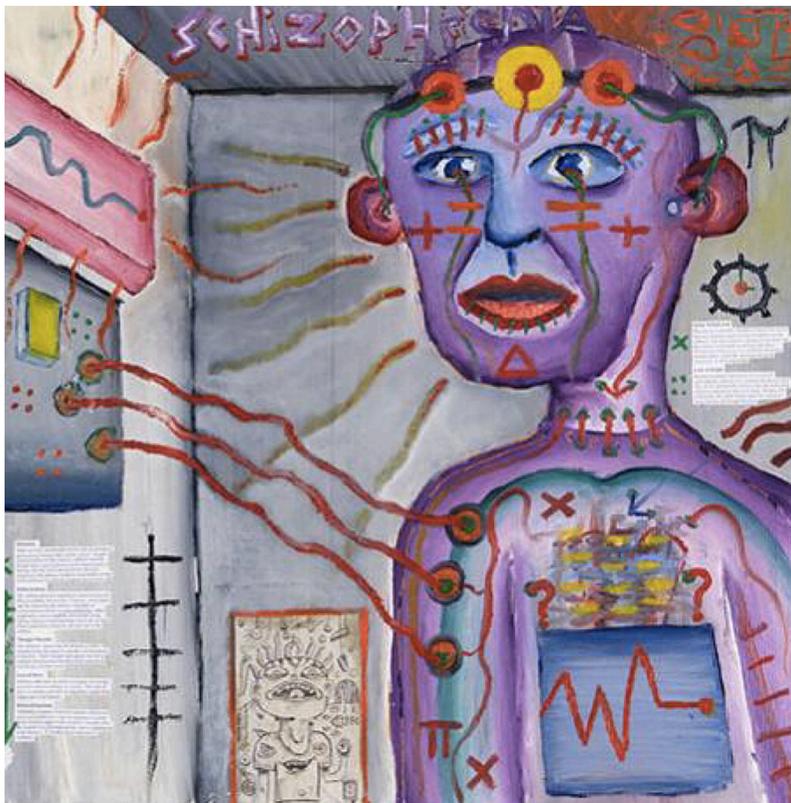
Table 1. Population Variation Explained by GWAS for a Selected Number of Complex Traits

Trait or Disease	h² Pedigree Studies	h² GWAS Hits^a	h² All GWAS SNPs^b
Type 1 diabetes	0.9 ⁹⁸	0.6 ^{99, c}	0.3 ¹²
Type 2 diabetes	0.3–0.6 ¹⁰⁰	0.05–0.10 ³⁴	
Obesity (BMI)	0.4–0.6 ^{101,102}	0.01–0.02 ³⁶	0.2 ¹⁴
Crohn's disease	0.6–0.8 ¹⁰³	0.1 ¹¹	0.4 ¹²
Ulcerative colitis	0.5 ¹⁰³	0.05 ¹²	
Multiple sclerosis	0.3–0.8 ¹⁰⁴	0.1 ⁴⁵	
Ankylosing spondylitis	>0.90 ¹⁰⁵	0.2 ¹⁰⁶	
Rheumatoid arthritis	0.6 ¹⁰⁷		
Schizophrenia	0.7–0.8 ¹⁰⁸	0.01 ⁷⁹	0.3 ¹⁰⁹
Bipolar disorder	0.6–0.7 ¹⁰⁸	0.02 ⁷⁹	0.4 ¹²
Breast cancer	0.3 ¹¹⁰	0.08 ¹¹¹	
Von Willebrand factor	0.66–0.75 ^{112,113}	0.13 ¹¹⁴	0.25 ¹⁴
Height	0.8 ^{115,116}	0.1 ¹³	0.5 ^{13,14}
Bone mineral density	0.6–0.8 ¹¹⁷	0.05 ¹¹⁸	
QT interval	0.37–0.60 ^{119,120}	0.07 ¹²¹	0.2 ¹⁴
HDL cholesterol	0.5 ¹²²	0.1 ⁵⁷	
Platelet count	0.8 ¹²³	0.05–0.1 ⁵⁸	

Table 1. Population Variation Explained by GWAS for a Selected Number of Complex Traits

Trait or Disease	h² Pedigree Studies	h² GWAS Hits^a	h² All GWAS SNPs^b
Type 1 diabetes	0.9 ⁹⁸	0.6 ^{99, c}	0.3 ¹²
Type 2 diabetes	0.3–0.6 ¹⁰⁰	0.05–0.10 ³⁴	
Obesity (BMI)	0.4–0.6 ^{101,102}	0.01–0.02 ³⁶	0.2 ¹⁴
Crohn's disease	0.6–0.8 ¹⁰³	0.1 ¹¹	0.4 ¹²
Ulcerative colitis	0.5 ¹⁰³	0.05 ¹²	
Multiple sclerosis	0.3–0.8 ¹⁰⁴	0.1 ⁴⁵	
Ankylosing spondylitis	>0.90 ¹⁰⁵	0.2 ¹⁰⁶	
Rheumatoid arthritis	0.6 ¹⁰⁷		
Schizophrenia	0.7–0.8 ¹⁰⁸	0.01 ⁷⁹	0.3 ¹⁰⁹
Bipolar disorder	0.6–0.7 ¹⁰⁸	0.02 ⁷⁹	0.4 ¹²
Breast cancer	0.3 ¹¹⁰	0.08 ¹¹¹	
Von Willebrand factor	0.66–0.75 ^{112,113}	0.13 ¹¹⁴	0.25 ¹⁴
Height	0.8 ^{115,116}	0.1 ¹³	0.5 ^{13,14}
Bone mineral density	0.6–0.8 ¹¹⁷	0.05 ¹¹⁸	
QT interval	0.37–0.60 ^{119,120}	0.07 ¹²¹	0.2 ¹⁴
HDL cholesterol	0.5 ¹²²	0.1 ⁵⁷	
Platelet count	0.8 ¹²³	0.05–0.1 ⁵⁸	

Schizophrenia



<http://en.wikipedia.org>

<http://newhealthadvisor.com>

Bipolar disorder



<http://www.uwf.edu>

<http://www.health.com>

説明出来ない遺伝率の謎

Vol 461|8 October 2009|doi:10.1038/nature08494

nature

REVIEWS

Finding the missing heritability of complex diseases

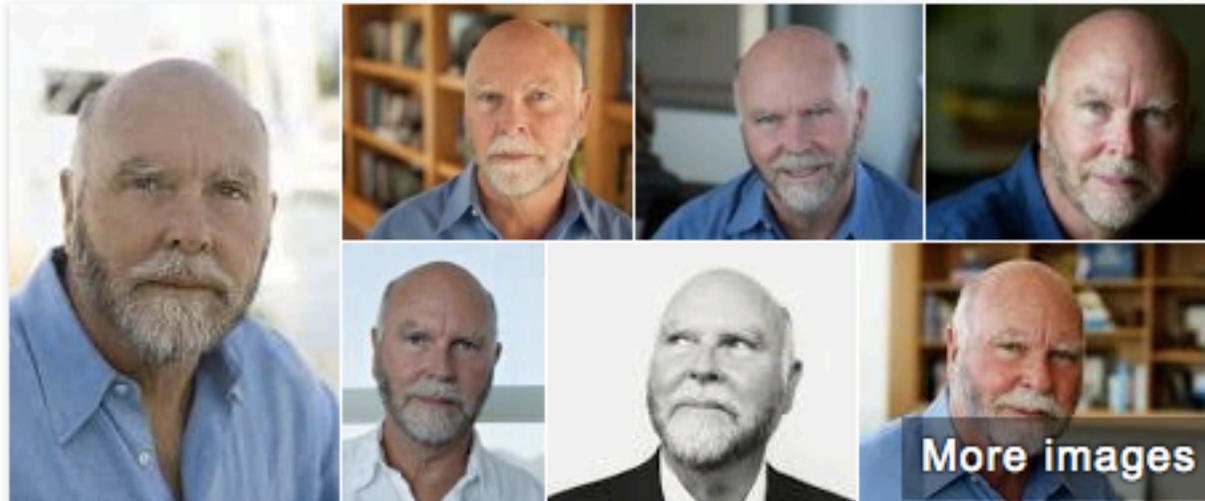
Teri A. Manolio¹, Francis S. Collins², Nancy J. Cox³, David B. Goldstein⁴, Lucia A. Hindorf⁵, David J. Hunter⁶, Mark I. McCarthy⁷, Erin M. Ramos⁵, Lon R. Cardon⁸, Aravinda Chakravarti⁹, Judy H. Cho¹⁰, Alan E. Guttmacher¹, Augustine Kong¹¹, Leonid Kruglyak¹², Elaine Mardis¹³, Charles N. Rotimi¹⁴, Montgomery Slatkin¹⁵, David Valle⁹, Alice S. Whittemore¹⁶, Michael Boehnke¹⁷, Andrew G. Clark¹⁸, Evan E. Eichler¹⁹, Greg Gibson²⁰, Jonathan L. Haines²¹, Trudy F. C. Mackay²², Steven A. McCarroll²³ & Peter M. Visscher²⁴

説明出来ない遺伝率の主な原因

- シーケンシング精度の不足
- モデルの間違い
 - 複雑な表現型は数100以上のSNPの関与
 - 遺伝子間相互作用の考慮
- 病気の原因は患者に共通のSNPではなく、患者毎に異なるSNPである可能性

個人情報としてのゲノム

自分のゲノムを公開する人(その1)



Craig Venter



Geneticist

John Craig Venter is an American biotechnologist, biochemist, geneticist, and entrepreneur. He is known for being one of the first to sequence the human genome and the first to transfect a cell with a synthetic genome.

[Wikipedia](#)

Born: October 14, 1946 (age 69), Salt Lake City, Utah, United States

Organizations founded: J. Craig Venter Institute, [more](#)

Spouse: [Heather Kowalski](#) (m. 2008), [Claire M. Fraser](#) (m. 1981–2005), [Barbara Rae](#) (m. 1968)

Books: [Life at the Speed of Light: From the Double Helix to the Dawn of Digital Life](#), [more](#)

Education: [University of California, San Diego](#) (1975), [more](#)

Movies: [The Future is Now!](#)

自分のゲノムを公開する人(その2)

事業推進担当者: 富田 勝 (とみた まさる)

[事業推進担当者](#) | [特別研究教員](#) | [RA・特別研究員](#)



慶應義塾大学先端生命科学研究所

昭和32年生

慶應義塾大学先端生命科学研究所

〒252-8520 藤沢市遠藤5322

TEL/FAX:0466-47-5099

E-mail:mt@sfc.keio.ac.jp

URL:<http://www.iab.keio.ac.jp/>



Japanese

Google™ カスタム検索

Search

[About DDBJ](#)

[How to Use](#)

[Report/Statistics](#)

[FAQ](#)

[Contact Us](#)

▶ Personal whole genome sequencing data of Prof.Masaru Tomita was released

Last updated : 2012.8.1.

Personal whole genome sequencing data of Prof.Masaru Tomita (Institute for Advanced Biosciences, Keio University) was released as DRA000583. Data is available at [DRASearch](#).

ゲノム情報は誰のものか？

1. 自分もの
2. 直系1親等まで(親、子)
3. 直系2親等まで(祖父母、孫)
4. 直系3親等まで(曾祖父母、曾孫)

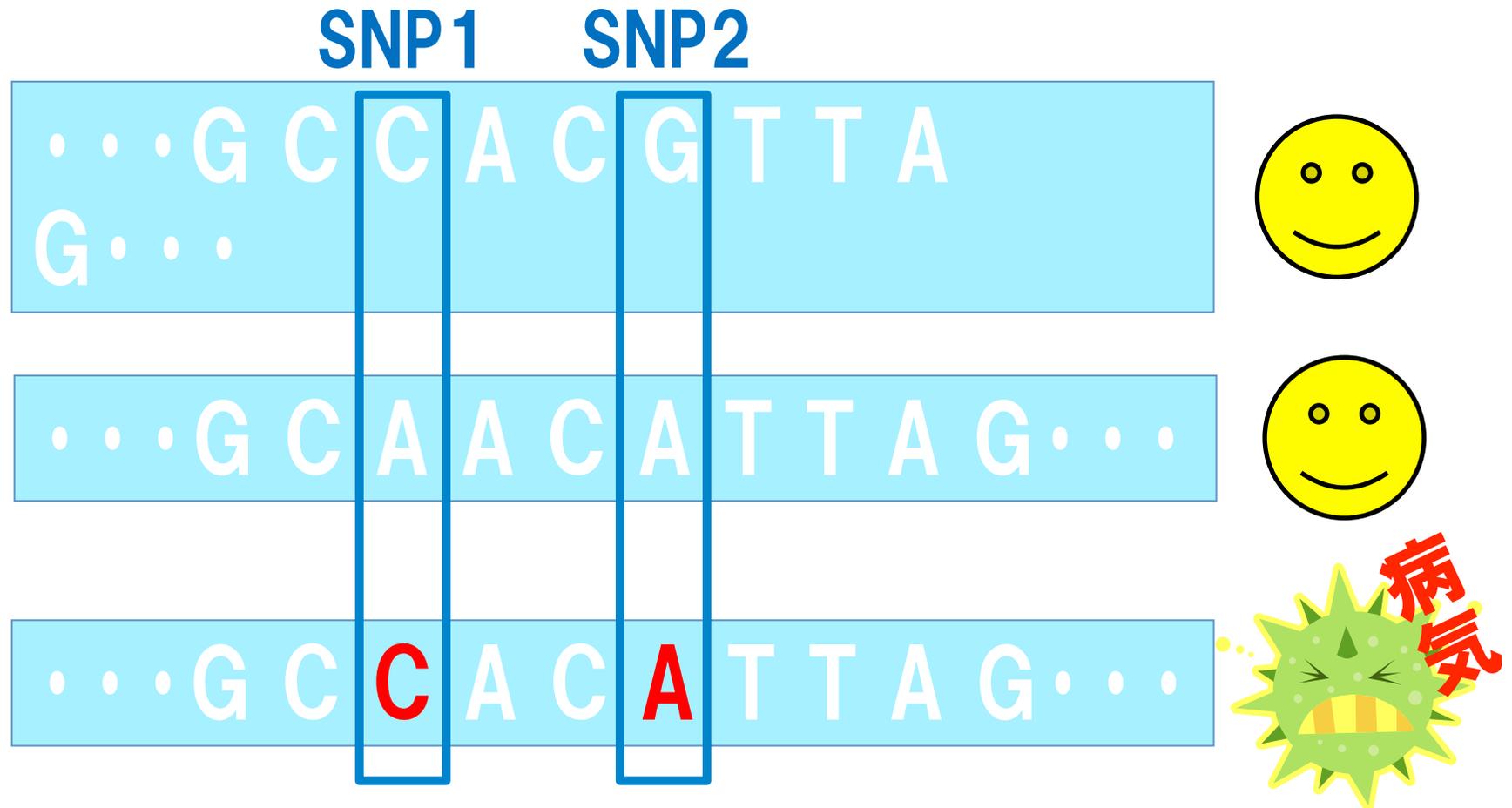
原因が1つのSNPだけとは限らない！

身長や手足の長さといった量的形質、
もしくは癌のような複雑な表現型には複数の遺伝的、
環境的要因があると考えられている。

iPS細胞の活性には3つ以上の遺伝子が必要

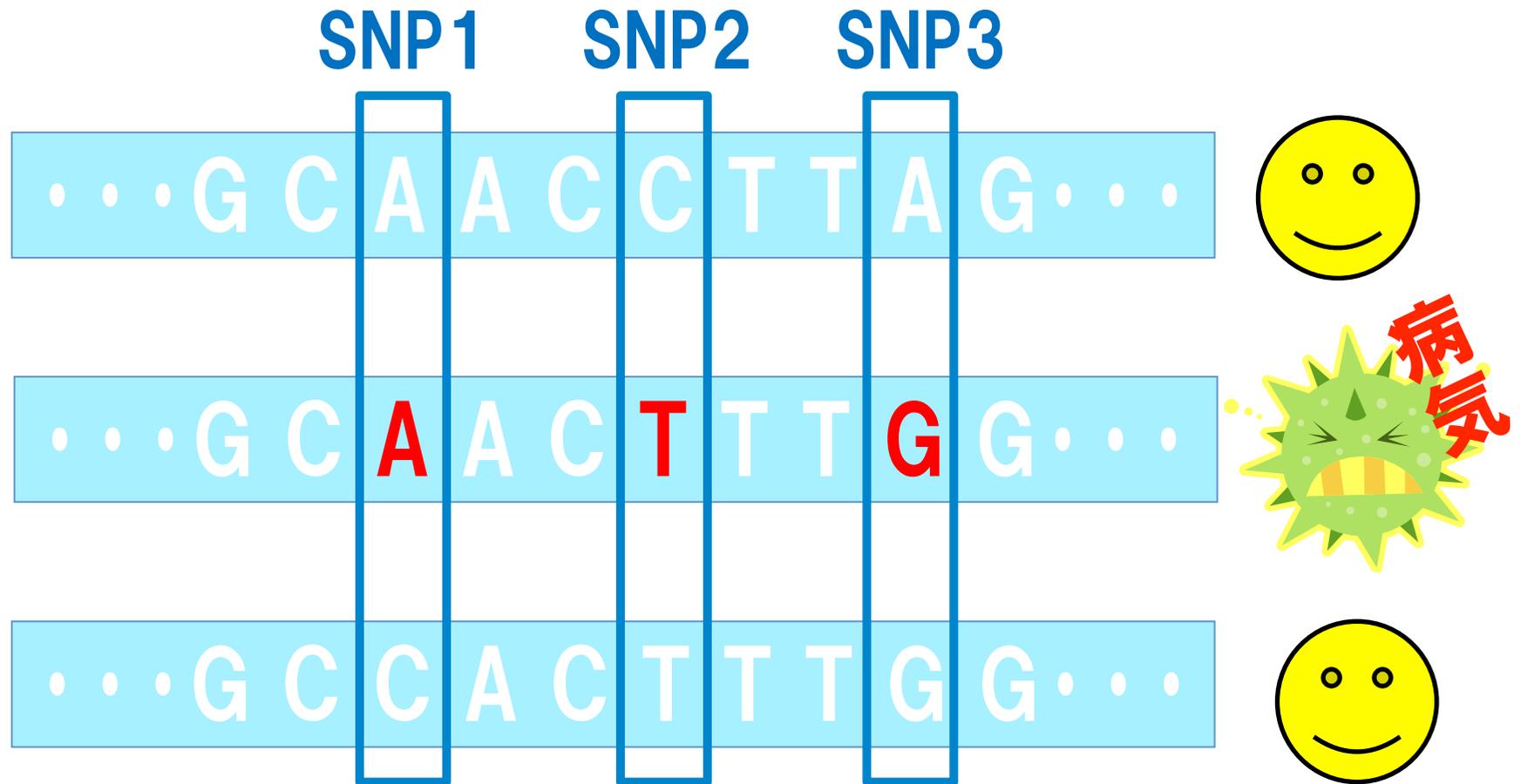
SNP間相互作用

例：2次の相互作用の場合



SNP間相互作用

例：3次の相互作用の場合



SNP組み合わせを全て調べれば、病気の原因となっているSNPが分かる

チャレンジ:膨大な計算量

2次の相互作用を見つけないとき、
(SNP1,SNP2) (SNP1,SNP3)
の1つ1つに相互作用があるかないかを見ていけばよい

2次の相互作用の場合 $300万C_2 \cong 10^{12} = 1兆通り$

3次の相互作用の場合 $300万C_3 \cong 10^{18} = 100京通り$



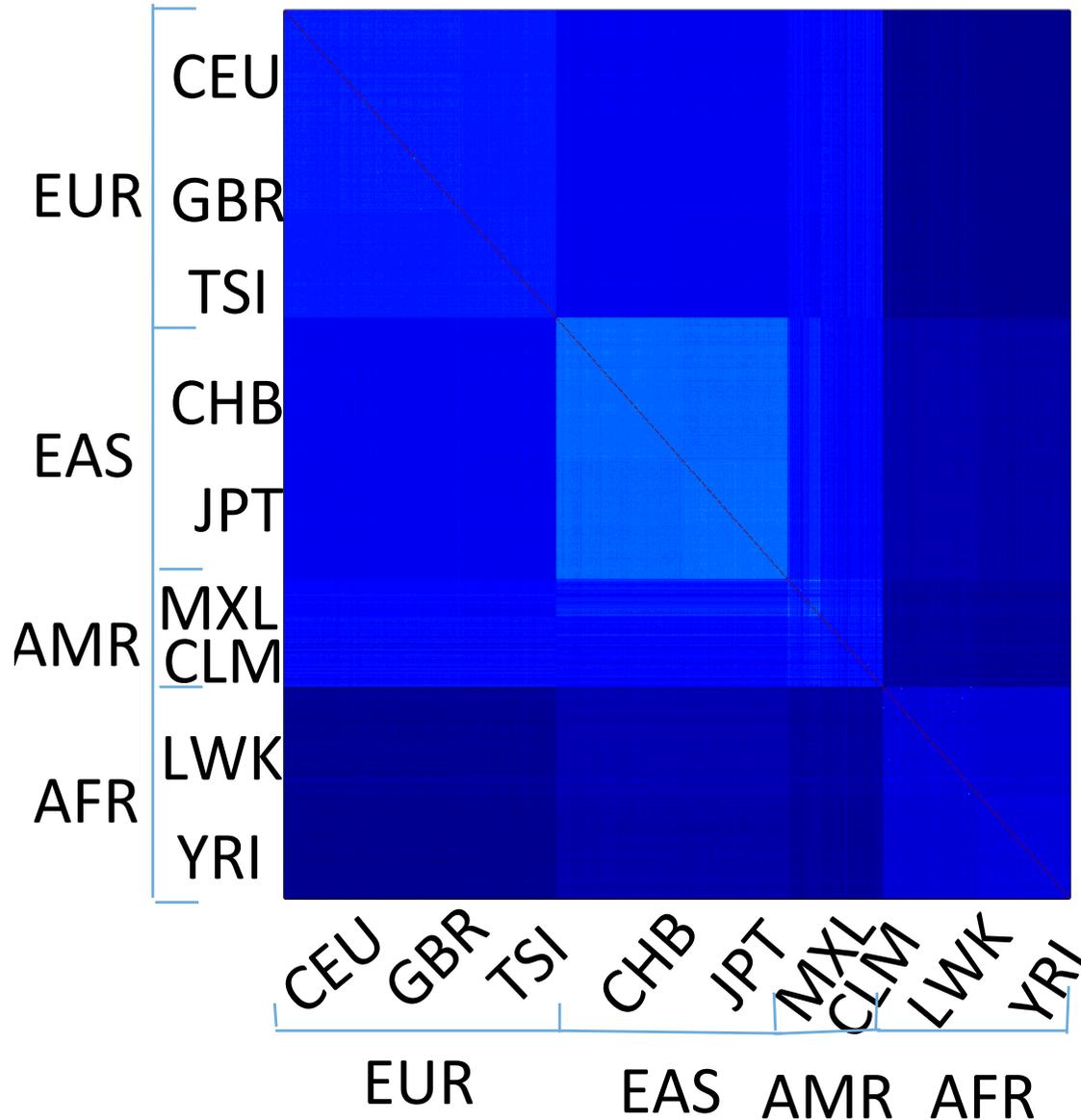
スパコン京を用いても計算困難

まとめ

- 多くの生物種のゲノムが解読されているが、遺伝子型と表現型の対応はまだほとんど分かっていない。
- これからは個人ゲノムの時代。
 - ゲノム創薬によるテーラーメイド医療
 - ゲノム診断による将来の病気の予防

付録

人種間の遺伝子の違い



CEU : 祖先の居住地が北および西ヨーロッパだったユタ州住民

GBR : イギリス

TSI : イタリア

CHB : 中国

JPT : 日本

MXL : ロサンゼルス

CLM : コロンビア

LWK : ケニア

YRI : ナイジェリア

