医学研究における
次世代シーケンサ技術の活用

大阪府立成人病センター　研究所

久木田洋児

医学研究における次世代シーケンサーの用途

- 個人ゲノム解読 (Personal Genomics) のための技術
  ” human genome re-sequencer”

- 研究
  - 一般的な疾患の遺伝素因探索
    - 全ゲノム相関解析の検出限界以下（付近）の稀な疾患変異探索
  - 単遺伝子疾患原因遺伝子探索（>3000については未だ原因遺伝子が不明）
    - 連鎖解析では解析不能な稀少疾患の小家系の解析

- 診断
  - 既知疾患/遺伝病発症予測
    - 出生前後診断 （血中遊離DNAを用いたダウン症の出生前診断）
  - がんの分子診断
    - 癌組織や血中遊離DNA中の体細胞変異（癌組織の突然変異）の検出・解析

塩基配列変異（突然変異と多型）

- 突然変異（mutation）はゲノム中の塩基配列の違いで、稀であり、ある個人に特異なものであったりする。一般的に表現型に病的影響を与えるものを指すことが多い。

- 多型（polymorphism）は一般集団に見られる塩基配列の違いである。疾患を引き起こす変異ではないが、身長や髪色などの身体的特徴や疾患への感受性、薬剤への応答などに影響している事がわかってきている。多型も始まりは、ある個体に生じた突然変異で、個体の生存に深刻な悪影響を及ぼさなかった場合、集団中に広まり定着したものである。

ヒトゲノム上の塩基配列変異

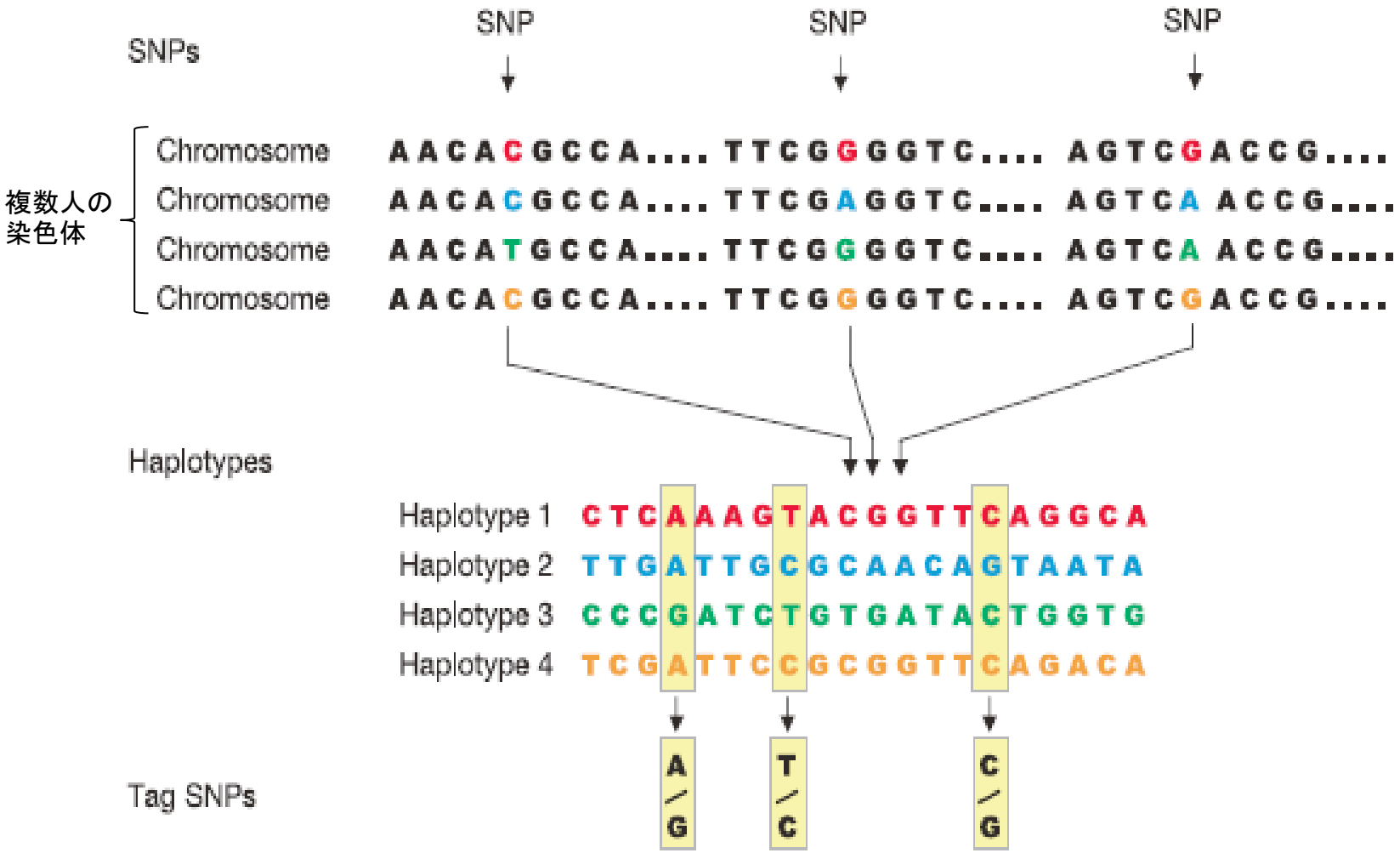
- 1塩基が他の塩基に置き換わっているもの
  - 1塩基置換
  （1塩基の多型のことをSNP，single nucleotide polymorphismという）

- 1から数十塩基以上の欠失/挿入
  - 2塩基から数十塩基を1単位とする配列が並んで繰り返す
    マイクロサテライト，VNTR（variable number of tandem repeat）
  - 数kb以上の配列（領域）の数が個人間で異なる
    コピー数多型（CNV, CNP）

- 配列の向きがレファレンスゲノムと逆になっているもの
  - 逆位 （e.g., ヨーロッパ人集団に見られる17q21.31の 900 kb領域）

- 染色体が途中から切れて、別の染色体につながる
  - 転座

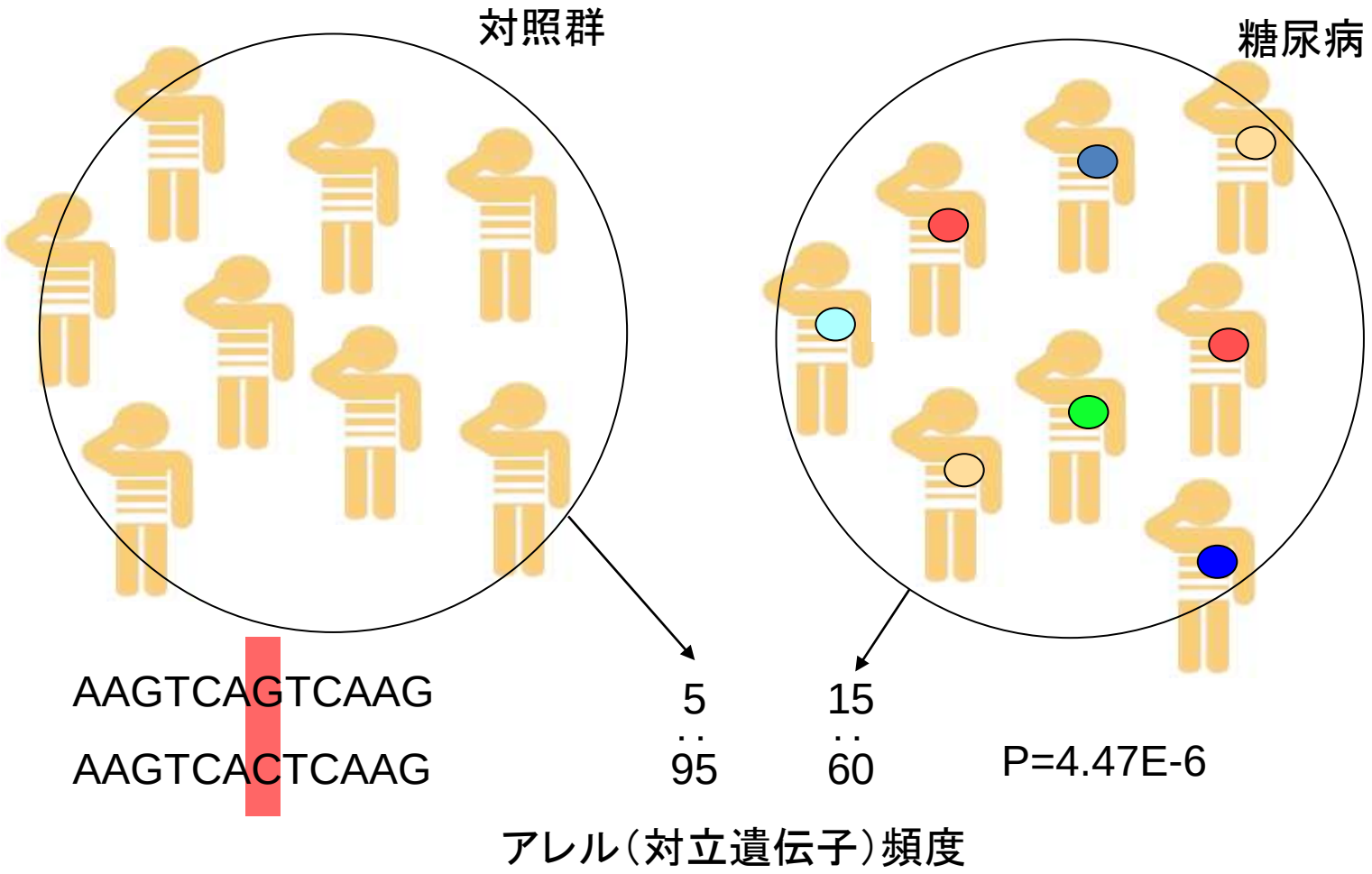
一塩基多型(SNP)とハプロタイプ



The International HapMap Consortium, Nature 2003

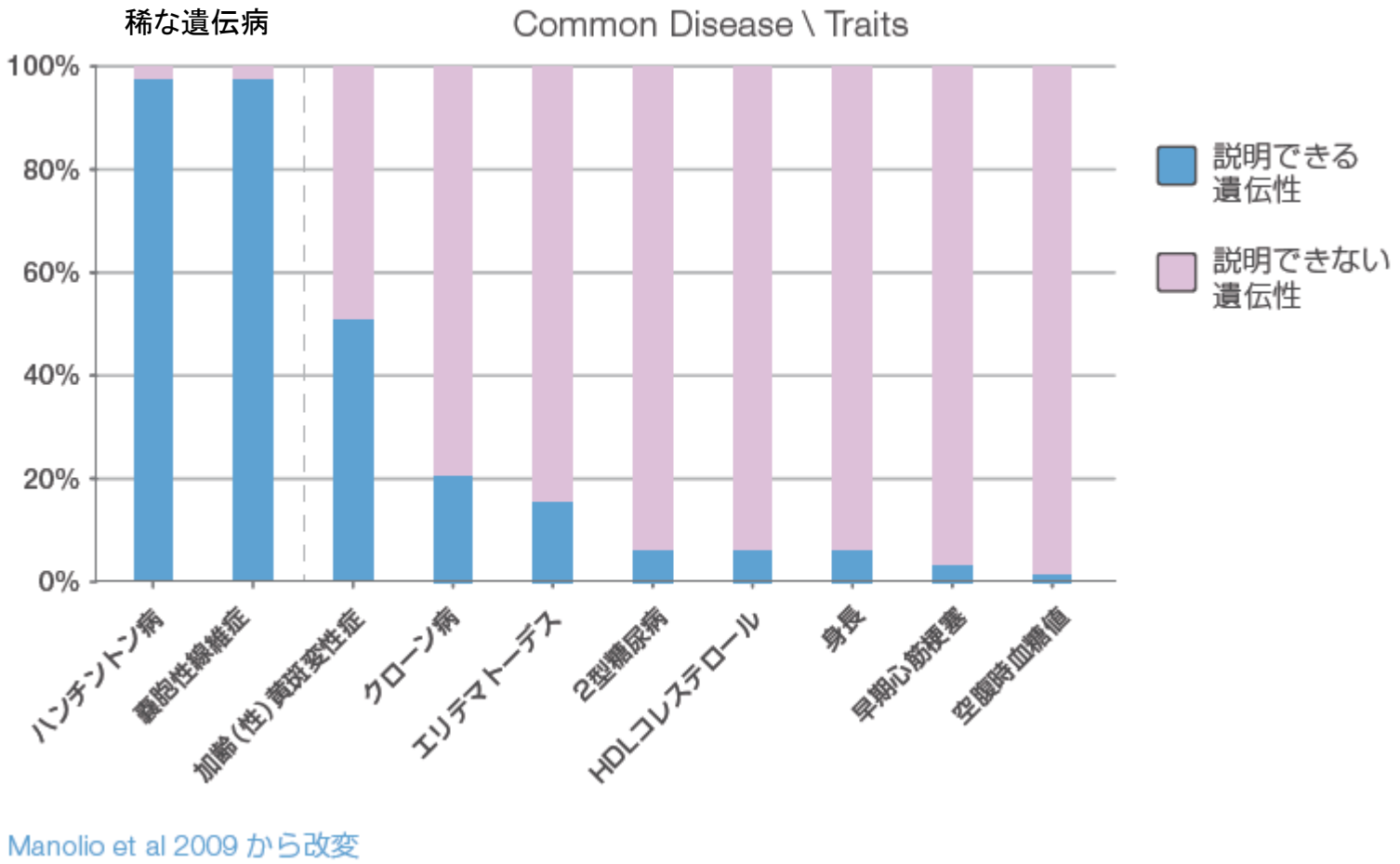
全ゲノム相関解析
genome-wide association study (GWAS)

糖尿病や高血圧などのcommonな疾患は、その疾患集団に高頻度で存在するcommonな遺伝多型で説明できる、という仮定が基になっている。



GWASでは遺伝素因の殆どは説明できない

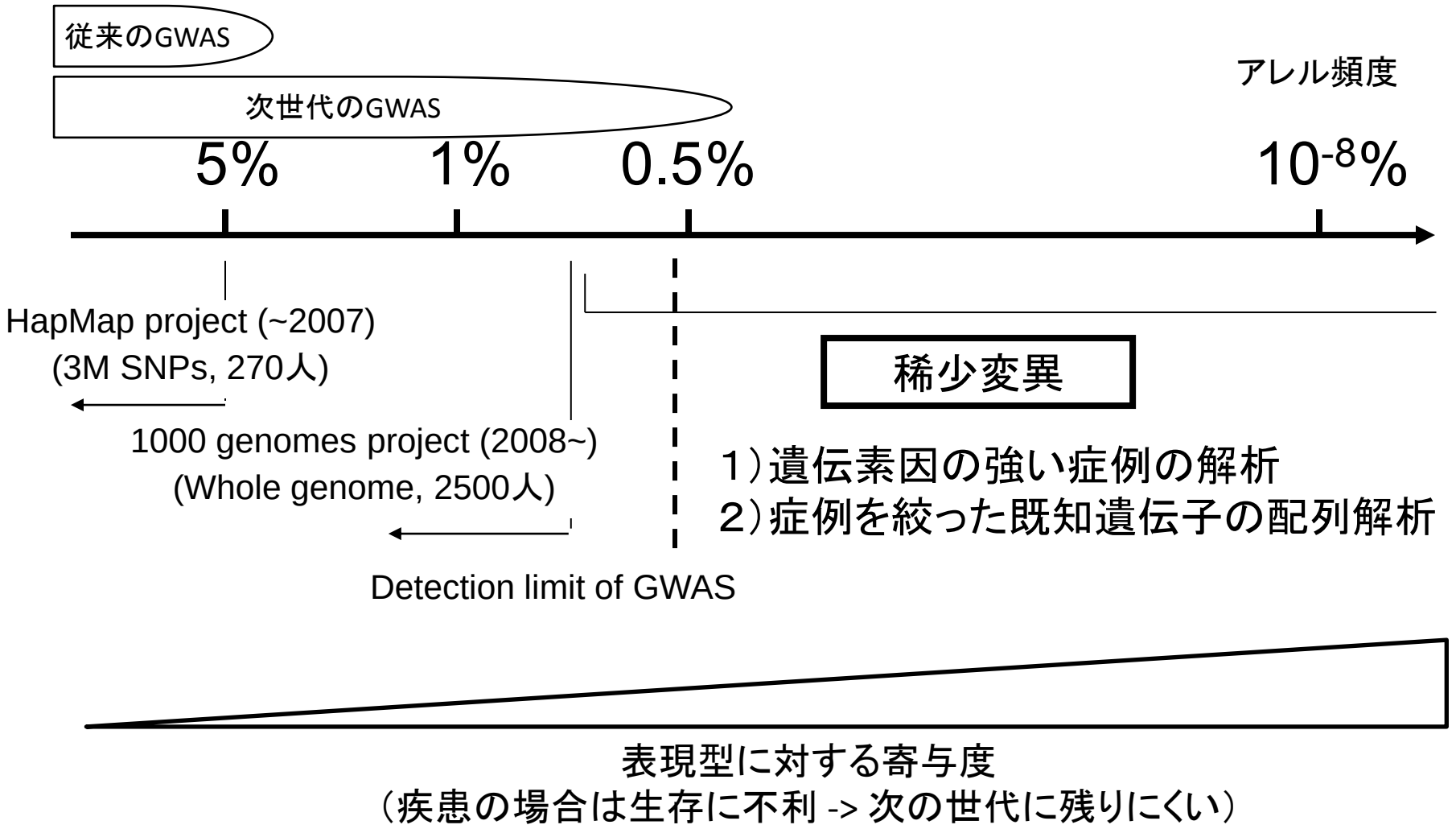
検出されたSNPの多くは、疾患に対する寄与率が低い



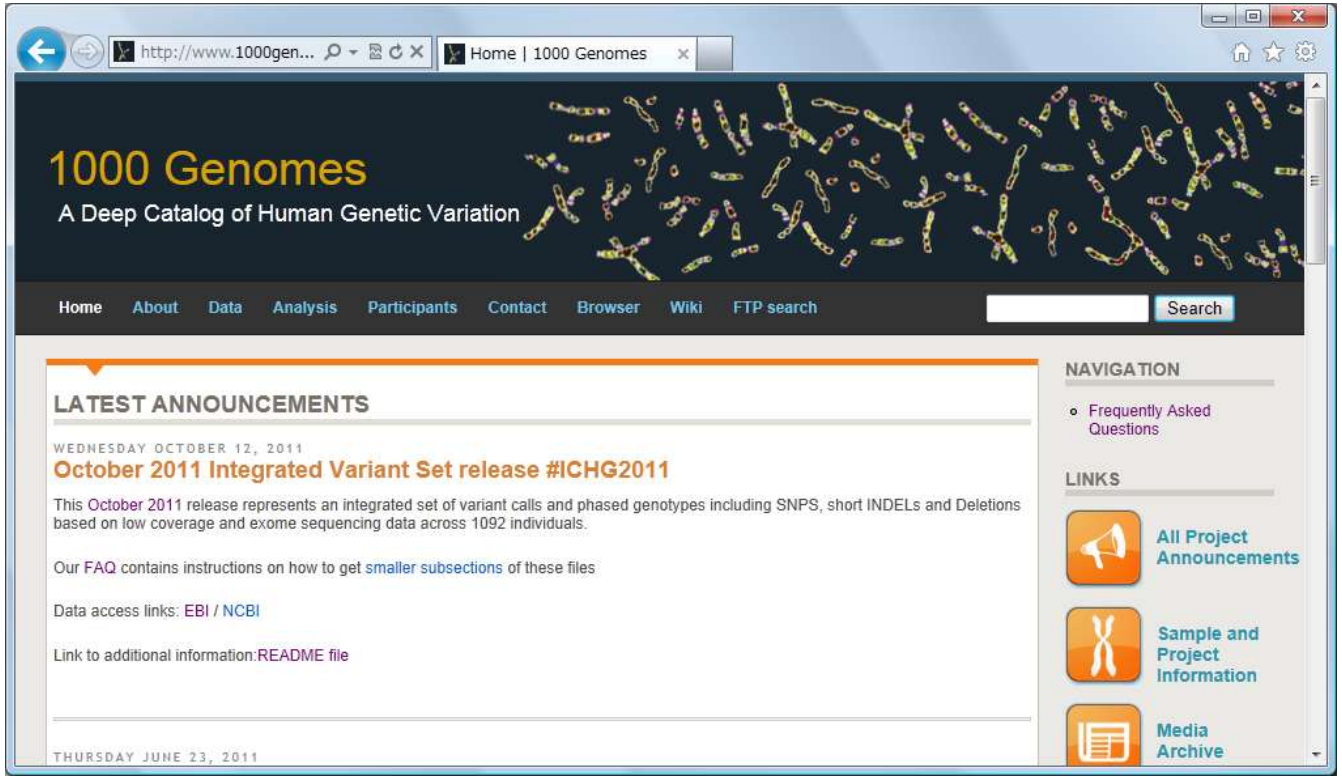
Manolio et al 2009 から改変

Illumina社資料　次世代GWAS研究

疾患関連稀少アレル（変異）の検出



1000 Genomes Project



- 5大陸29集団（ヨーロッパ人（5），東アジア人（6），西アフリカ人（6），アメリカ人（7），南アジア人（5））、計　2500人の全ゲノム配列を決定する。
- 各集団において95%以上の多型変異（頻度0.5-1%以上）を検出する。
- Pilot研究（Nature, 2010），Phase1データが公開され始めている。

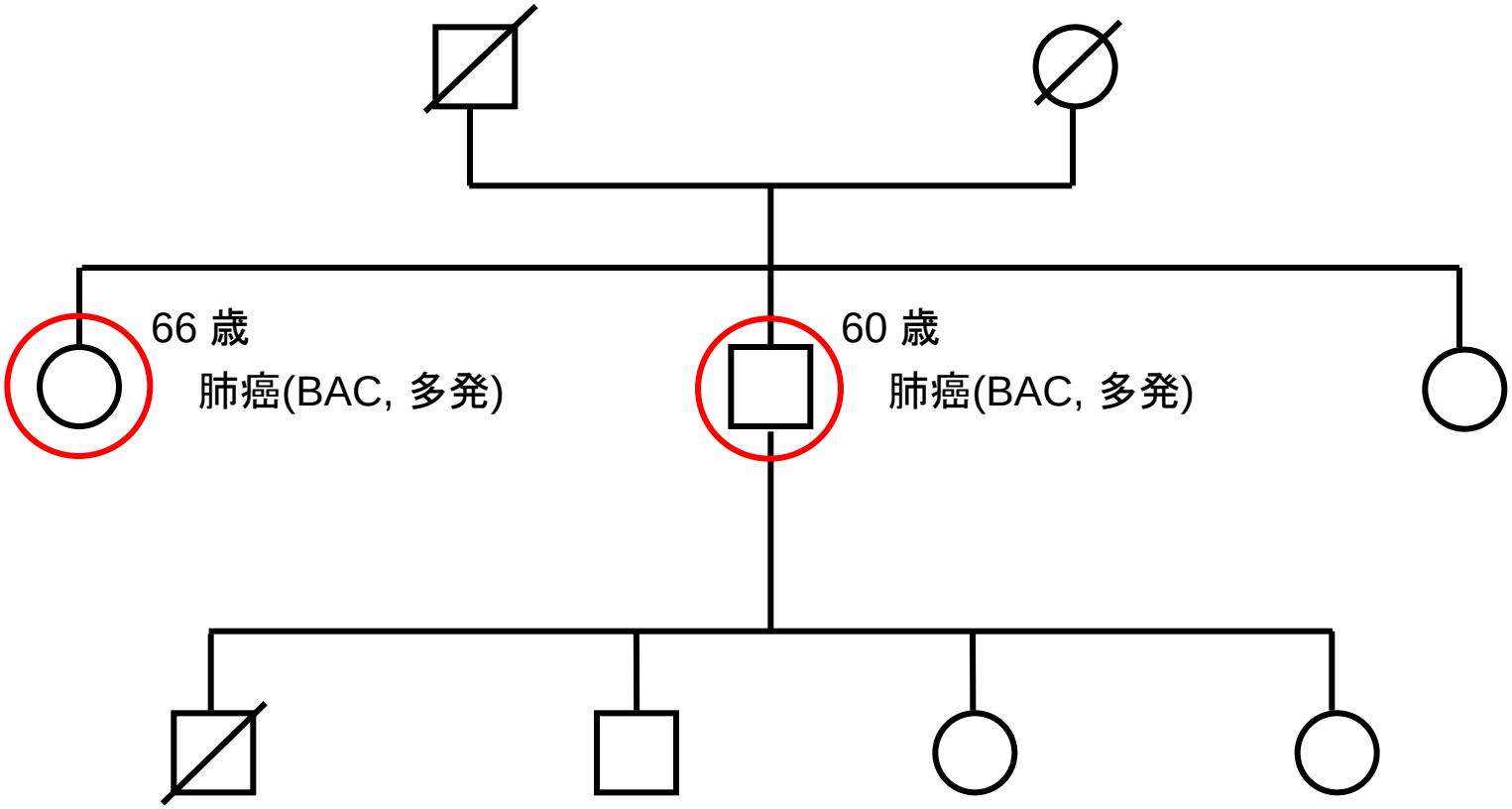
医学研究における次世代シーケンサーの用途

- 個人ゲノム解読 (Personal Genomics) のための技術
  ” human genome re-sequencer”

- 研究
  - 一般的な疾患の遺伝素因探索
    - 全ゲノム相関解析の検出限界以下（付近）の稀な疾患変異探索
  - **単遺伝子疾患原因遺伝子探索（>3000については未だ原因遺伝子が不明）**
    - **連鎖解析では解析不能な稀少疾患の小家系の解析**

- 診断
  - 既知疾患/遺伝病発症予測
    - 出生前後診断 （血中遊離DNAを用いたダウン症の出生前診断）
  - がんの分子診断
    - 癌組織や血中遊離DNA中の体細胞変異（癌組織の突然変異）の検出・解析

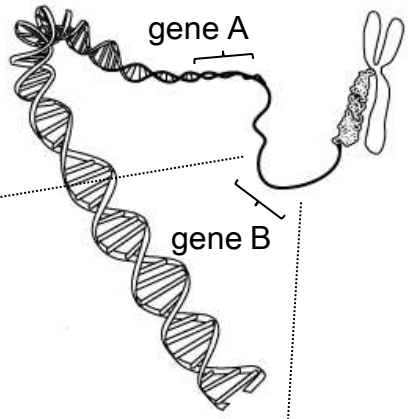
解析対象



BAC：**細気管支肺胞上皮癌**

Exome Sequencing

全遺伝子配列シーケンス



ヒトゲノムレファレンス配列　エキソン　イントロン　エキソン

bwa (*Li and Durbin, 2009, 2010*)でマッピング

冗長度
(50~ x)

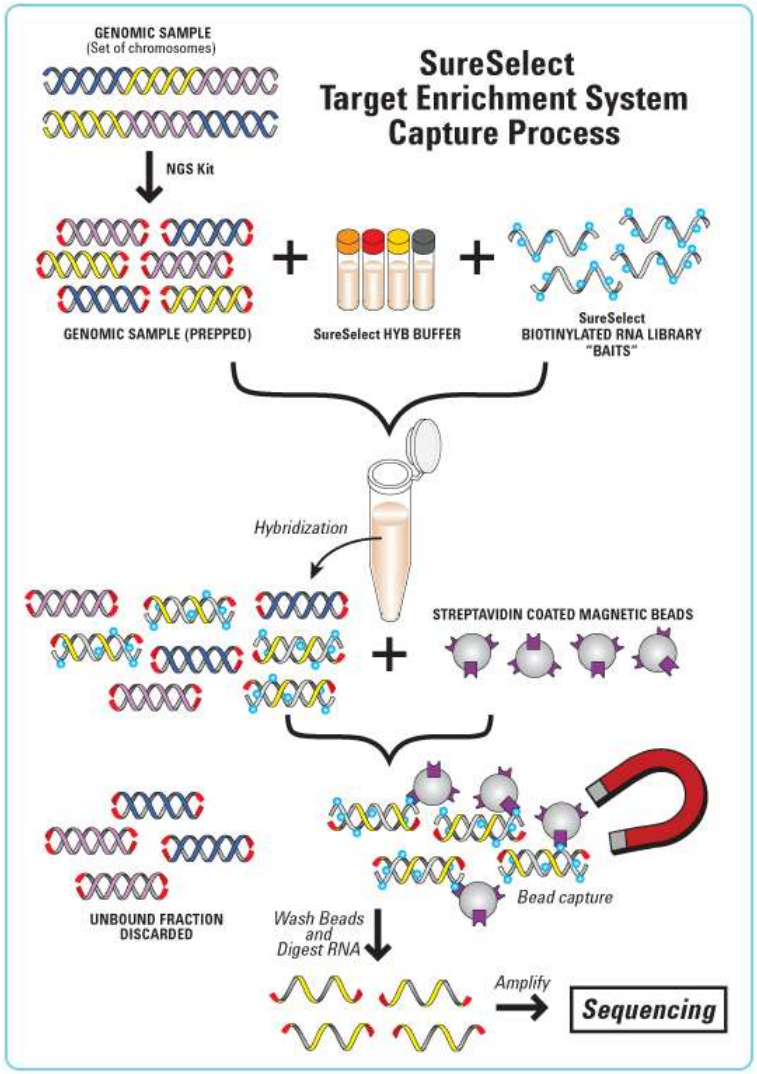
リード配列
(76 bases x 2)

samtools (*Li et al., 2009*)による変異(塩基置換, 挿入, 欠失)の検出

公共データベース(dbSNP b130, 1000 genome project Pilot1)を参照した既知多型の除去

アミノ酸変化を伴う新規変異の抽出

全エキソン配列の抽出



Sure Select Human All Exon Kit

(Agilent Technologies)

- Target regions:
  - 38 Mb (1% of human genome)
  - ~180,000 exons (coding regions)
  - >700 miRNAs
  - >300 non-coding RNA
- input: ~3 ug of genomic DNA

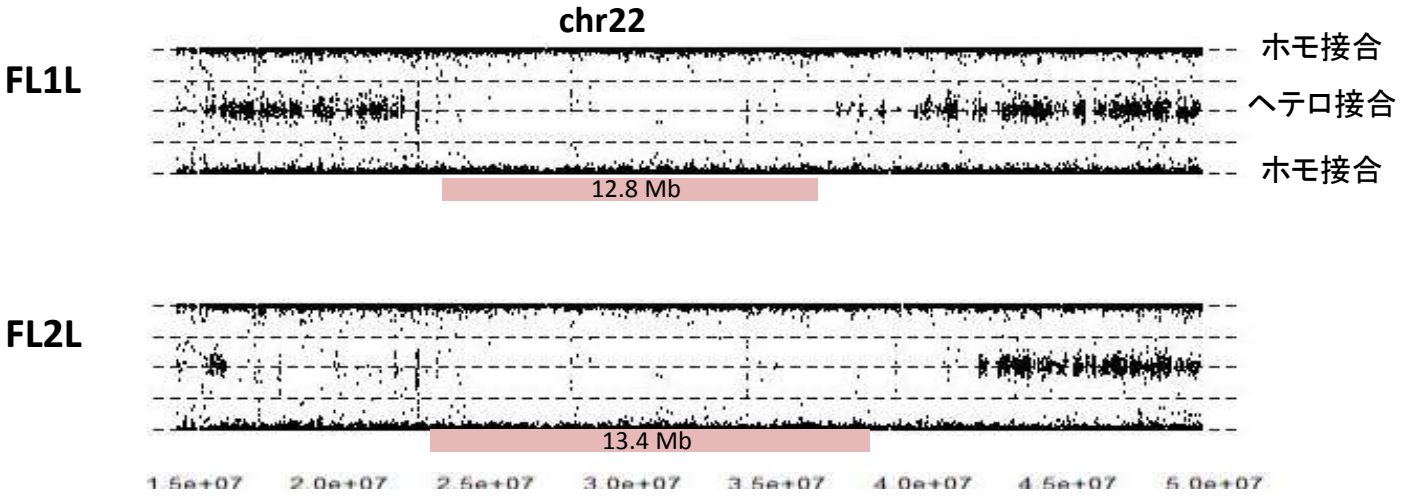
http://www.opengenomics.com/Products/SureSelect_Target_Enrichment_System

Variants Identified by Exome Sequencing

	FL1 (♂)		FL2 (♀)	
SNVs		(novel)		(novel)
synonymous	8,687	(247)	8,674	(236)
non-synonymous	7,350	(388)	7,493	(445)
nonsense	45	(9)	45	(9)
splice site	57	(2)	55	(5)
small indels				
coding indel	210	(74)	207	(66)
splice site	33	(7)	31	(5)

- アミノ酸変化を伴う新規変異を持つ遺伝子数は、FL1（♂）が427遺伝子、FL2（♀）が486遺伝子。その内、233遺伝子には両者に共通の変異があった。

患者には広いホモ接合領域が検出された



Assayed using Omni1 BeadChip (Illumina).
BAF: B allele frequency = intensity of B allele / (intensity of A allele + intensity of B allele)

Total length of homozygous regions (detected by plink software, Purcell *et al.*, 2007)

- FL1L (male): 278 Mb (chr1-22)
- FL2L (female): 210 Mb (chr1-22, X)

親以上の世代で近親婚が行われている（いとこ婚？）

- shared homozygous regions: 72 Mb (chr1-22, X. 612 genes) -> 原因変異の存在が疑われる領域

- 4 homozygous variants/ 4 genes in shared homozygous regions between patients

タンパク質変異機能予測プログラム

遺伝子産物であるタンパク質のアミノ酸配列や構造の保存度をもとに、塩基変異によるアミノ酸変化が与える機能への影響を予測する（影響が無いのか、不活化するのか）。

Method	Algorithm
SIFT (http://sift.jcvi.org)	SIFT uses sequence homology; scores are calculated using position-specific scoring matrices with Dirichlet priors
Polyphen[12] (http://genetics.bwh.harvard.edu/pph/)	Polyphen uses sequence conservation, structure and SWISS-PROT annotation
PMUT[13] (http://mmb2.pcb.ub.es:8080/PMut/)	PMUT provides prediction by neural networks, which use internal databases, secondary structure prediction and sequence conservation
SNPs3D[18] (http://www.snps3d.org/)	SNPs3D is based on support vector machine that uses structural or sequence conservation features
PantherPSEC[19] (http://www.pantherdb.org/tools/csnpScoreForm.jsp)	Panther PSEC uses sequence homology; scores are calculated using PANTHER Hidden Markov model families
MAPP[14] (http://mendel.stanford.edu/SidowLab/downloads/MAPP/index.html)	MAPP considers the physicochemical variation present in a column of a protein sequence alignment to predict the effect of all possible amino acid substitutions on protein function
Align-GVGD[15] (http://agvgd.iarc.fr/agvgd_input.php)	Align-GVGD combines the biophysical characteristics of amino acids and protein multiple sequence alignments

Kumar P et al., Nature Protocols, 4:1073-1082, 2009

マッピング/アラインメント時の障害

- 短いリード配列（今回は76塩基）
  - 低い特異性

    -> 間違った領域にマップされる

- 不完全なレファレンスゲノム配列
  - ギャップ　357か所（GRC37）
  - 個人間の違い（多様性、未知配列の存在）
  - 塩基置換、欠失、挿入、逆位などのゲノム構造変化

    -> 解決策としてde novoアライメントがあるが、ゲノム中の相同配列の存在（ゲノムの約半分を占めるトランスポゾンなどのリピート配列、遺伝子ホモログ、重複領域）が障害になっている

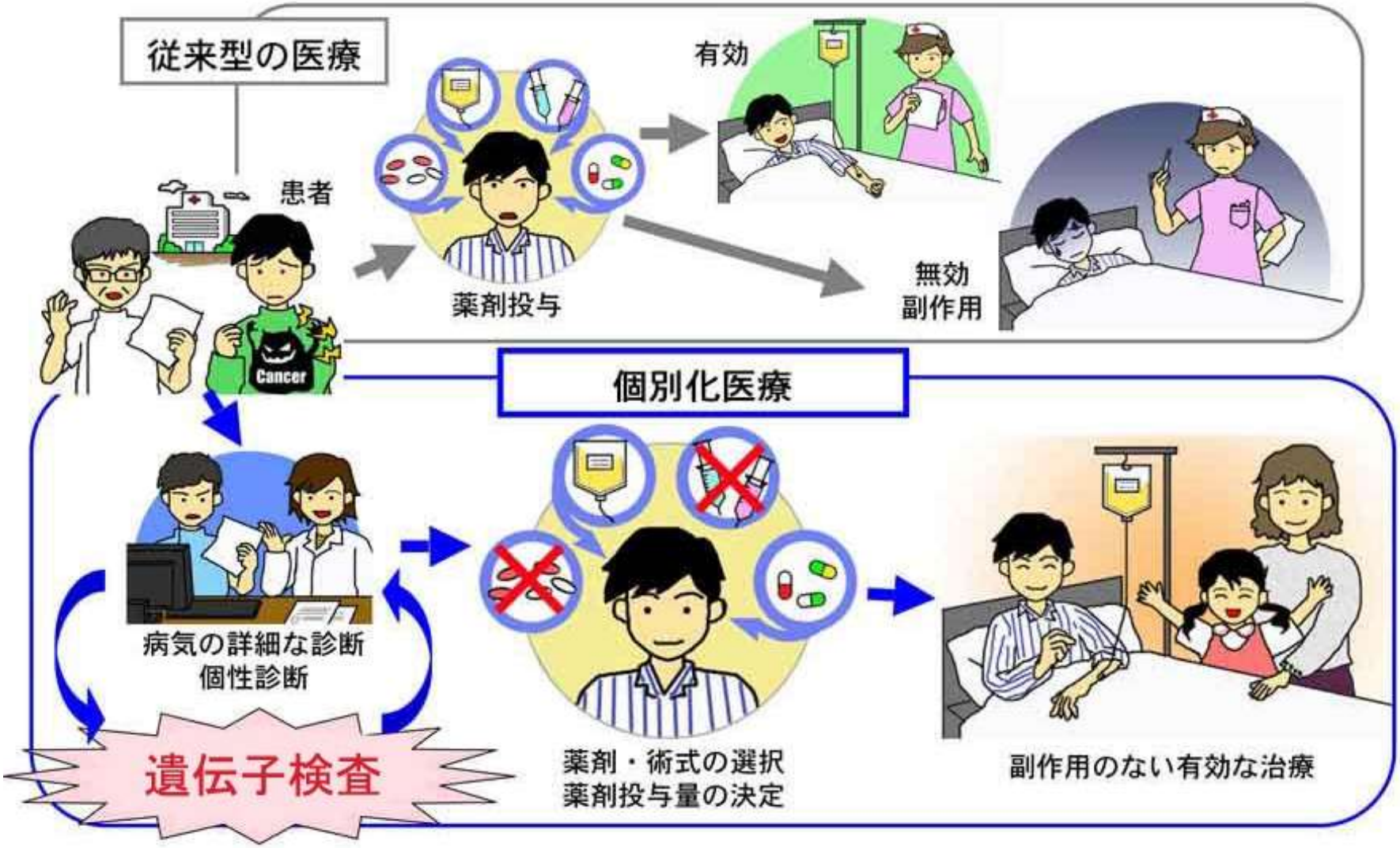
医学研究における次世代シーケンサーの用途

- 個人ゲノム解読 (Personal Genomics) のための技術
  ” human genome re-sequencer”

- 研究
  - 一般的な疾患の遺伝素因探索
    - 全ゲノム相関解析の検出限界以下（付近）の稀な疾患変異探索
  - 単遺伝子疾患原因遺伝子探索（>3000については未だ原因遺伝子が不明）
    - 連鎖解析では解析不能な稀少疾患の小家系の解析

- **診断**
  - **既知疾患/遺伝病発症予測**
    - **出生前後診断 （血中遊離DNAを用いたダウン症の出生前診断）**
  - **がんの分子診断**
    - **癌組織や血中遊離DNA中の体細胞変異（癌組織の突然変異）の検出・解析**

個別化医療


従来型の医療
患者
Cancer
薬剤投与
有効
無効
副作用
個別化医療
病気の詳細な診断
個性診断
遺伝子検査
薬剤・術式の選択
薬剤投与量の決定
副作用のない有効な治療

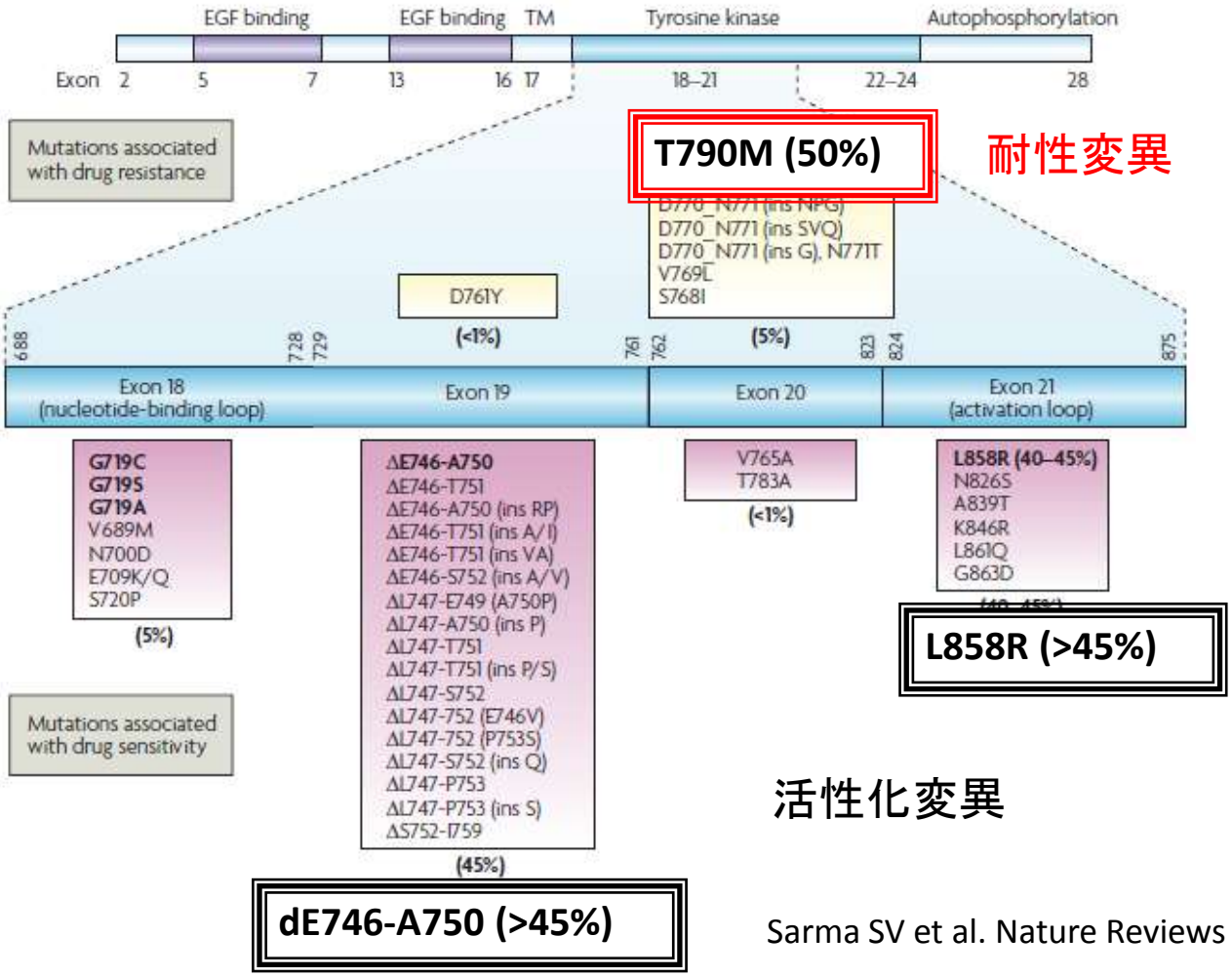
代表的な分子標的薬

一般名	商品名	開発企業	標的分子	国内承認	対象疾患	備考
イマチニブ	グリベック	ノバルティス	bcr-abl	2001.11	慢性骨髄性白血病	
トラスツズマブ	ハーセプチン	ロッシェ	Her2	2001.4	乳癌	Her2強発現のみ
ゲフィチニブ	イレッサ	アストラゼネカ	EGFR	2001.7	肺癌	EGFR mutation(+)のみ
エルロチニブ	タルセバ	ロッシェ	EGFR	2007.1	肺癌、膵癌	EGFR mutation(+)のみ
リツキシマブ	リツキサン	ロッシェ	CD20	2003.9	B細胞性非ホジキンリンパ腫	
イブリツモマブチウキセタン	ゼヴァリン	バイエル	CD20	2008.1	低悪性度B細胞性非ホジキンリンパ腫	放射性免疫療法薬(99Y or 111In)
ゲムツズマブオゾガマイシン	マイロターグ	ワイス	CD33	2005.7	急性骨髄性白血病	Mab+カリケアマイシン
スニチニブ	スーテント	ファイザー	VEGF, PDGF	2008.2	GIST, 腎臓癌	
セツキシマブ	アービタックス	メルク	EGFR	2008.7	大腸癌	K-ras mutation (-)のみ
ベバシズマブ	アバスチン	ロッシェ	VEGF	2007.4	大腸癌	
ソラフェニブ	ネクサバール	バイエル	PDGFR,VEGFR,KIT	2008.11	腎臓癌	
パニツムマブ	ベクチビックス	アムジェン	EGFR	2010.4	大腸癌	K-ras mutation (-)のみ
ボルテゾミブ	ベルケード	ヤンセンファーマ	プロテアソーム	2006.1	多発性骨髄腫	
クリゾチニブcrizotinib		ファイザー	ALK		肺癌	EML4-ALK

□ 個別化医療の対象となる薬剤

肺癌は日本人の癌種別死亡率の1位。患者の約1/4には、癌組織のEGFR遺伝子に活性化型突然変異が見つかる。

活性化及び耐性EGFR変異



Sarma SV et al. Nature Reviews Cancer 7: 169, 2007

- 活性化型突然変異の約90%はエキソン19の欠失変異とエキソン21の塩基置換。
- 耐性患者の半数にT790Mが見つかる。

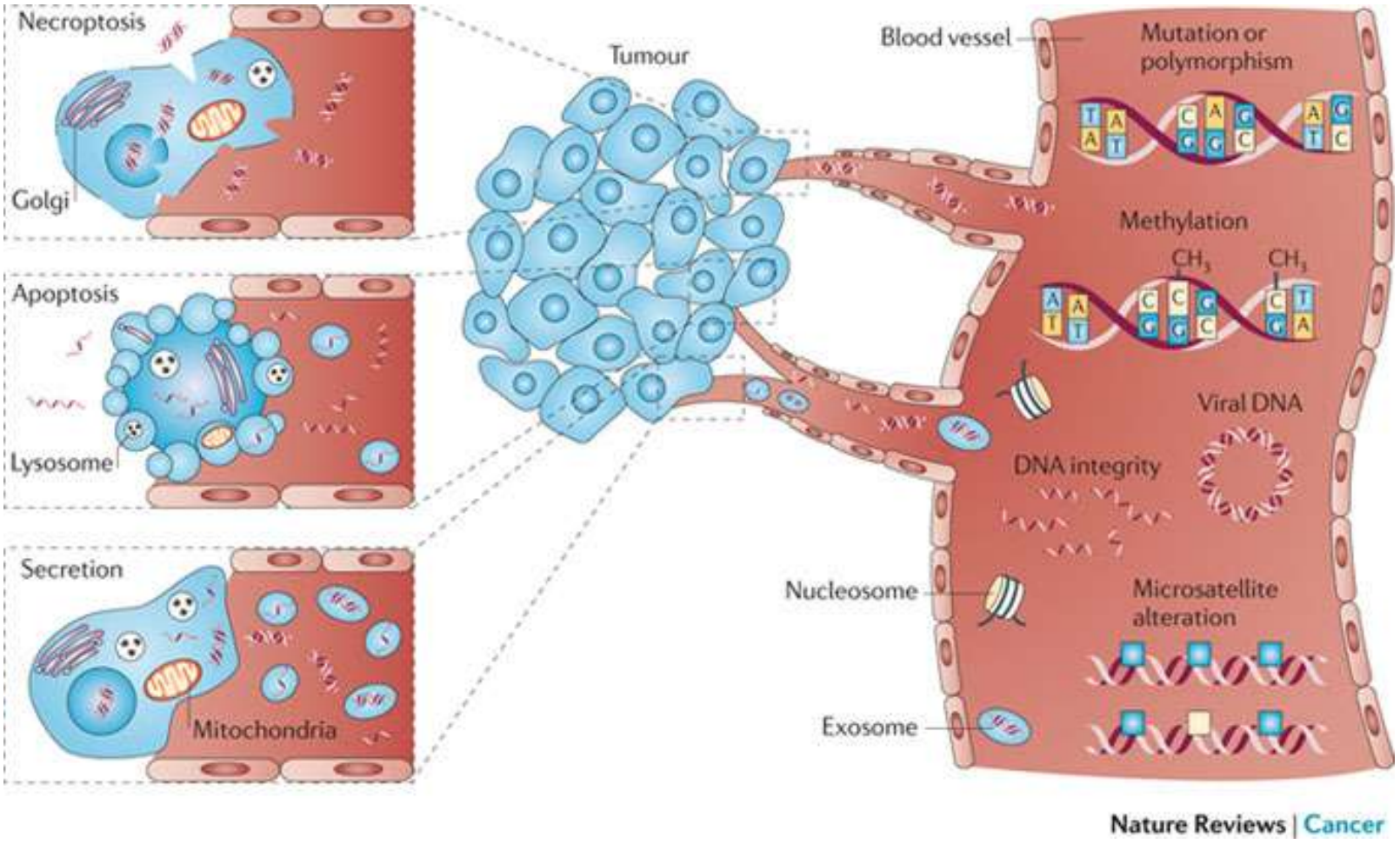
3か所の調査で大部分の患者を判別可

肺癌における非侵襲性遺伝子検査の重要性

1. 他臓器の癌に比べて腫瘍組織採取が難しい
   -> 気管支鏡検査、CTガイド下肺生検（気胸などの合併症）

2. チロシンキナーゼ阻害剤(イレッサやタルセバ)
   適用症例選択のため遺伝子検査は必須

血液中遊離DNA



Schwarzenbach, H. Nature Review Cancer, 11, 426, 2011.

半導体シーケンサー
Life Technology社 Ion Torrent PGM

ARTICLE

doi:10.1038/nature10242

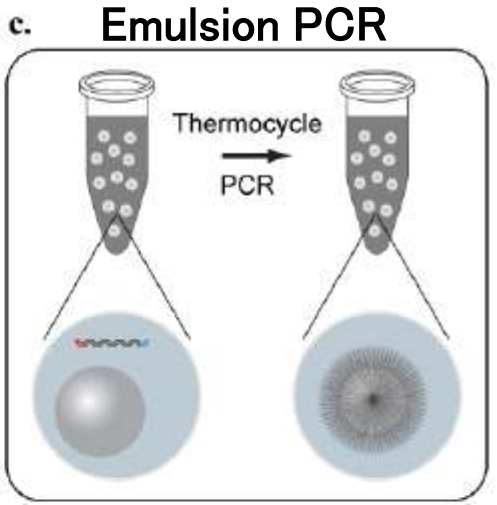
An integrated semiconductor device enabling non-optical genome sequencing

Jonathan M. Rothberg[1], Wolfgang Hinz[1], Todd M. Rearick[1], Jonathan Schultz[1], William Mileski[1], Mel Davey[1], John H. Leamon[1], Kim Johnson[1], Mark J. Milgrew[1], Matthew Edwards[1], Jeremy Hoon[1], Jan F. Simons[1], David Marran[1], Jason W. Myers[1], John F. Davidson[1], Annika Branting[1], John R. Nobile[1], Bernard P. Puc[1], David Light[1], Travis A. Clark[1], Martin Huber[1], Jeffrey T. Branciforte[1], Isaac B. Stoner[1], Simon E. Cawley[1], Michael Lyons[1], Yutao Fu[1], Nils Homer[1], Marina Sedova[1], Xin Miao[1], Brian Reed[1], Jeffrey Sabina[1], Erika Feierstein[1], Michelle Schorn[1], Mohammad Alanjary[1], Eileen Dimalanta[1], Devin Dressman[1], Rachel Kasinskas[1], Tanya Sokolsky[1], Jacqueline A. Fidanza[1], Eugeni Namsaraev[1], Kevin J. McKernan[1], Alan Williams[1], G. Thomas Roth[1] & James Bustillo[1]

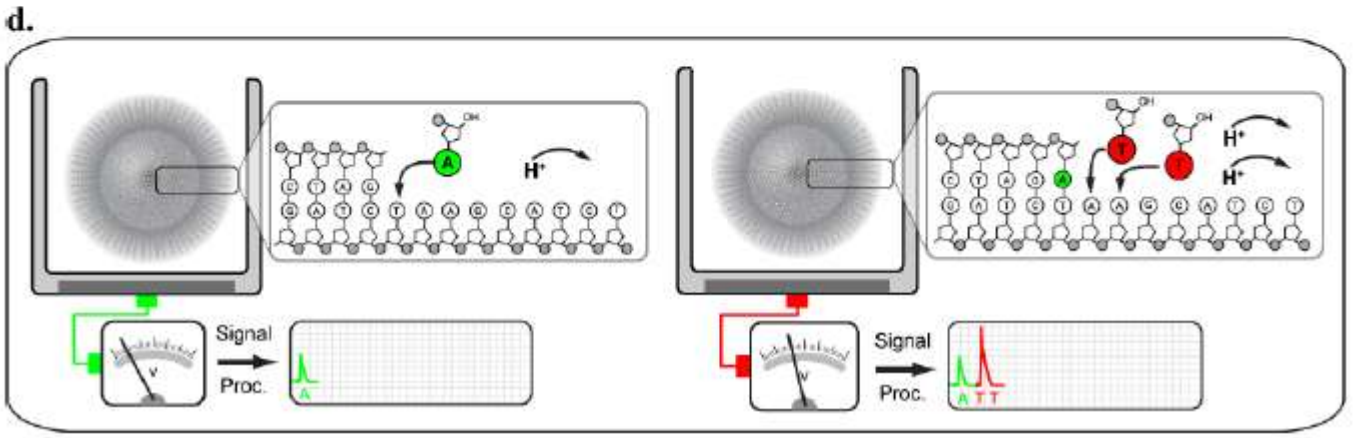
The seminal importance of DNA sequencing to the life sciences, biotechnology and medicine has driven the search for more scalable and lower-cost solutions. Here we describe a DNA sequencing technology in which scalable, low-cost semiconductor manufacturing techniques are used to make an integrated circuit able to directly p DNA sequencing of genomes. Sequence data are obtained by directly sensing the ions produced by DNA polymerase synthesis using all-natural nucleotides on this massively parallel semiconductor-s chip. The ion chip contains ion-sensitive, field-effect transistor-based sensors in perfect register wi which provide confinement and allow parallel, simultaneous detection of independent sequencing r most widely used technology for constructing integrated circuits, the complementary metal-ox (CMOS) process, allows for low-cost, large-scale production and scaling of the device to higher array sizes. We show the performance of the system by sequencing three bacterial genomes, its robus by producing ion chips with up to 10 times as many sensors and sequencing a human genome.

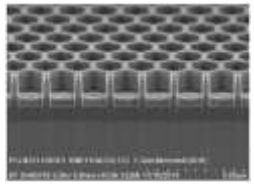
Ion Torrent PGMでのシーケンス

血漿由来DNAを鋳型にした
EGFRエキソンのPCR



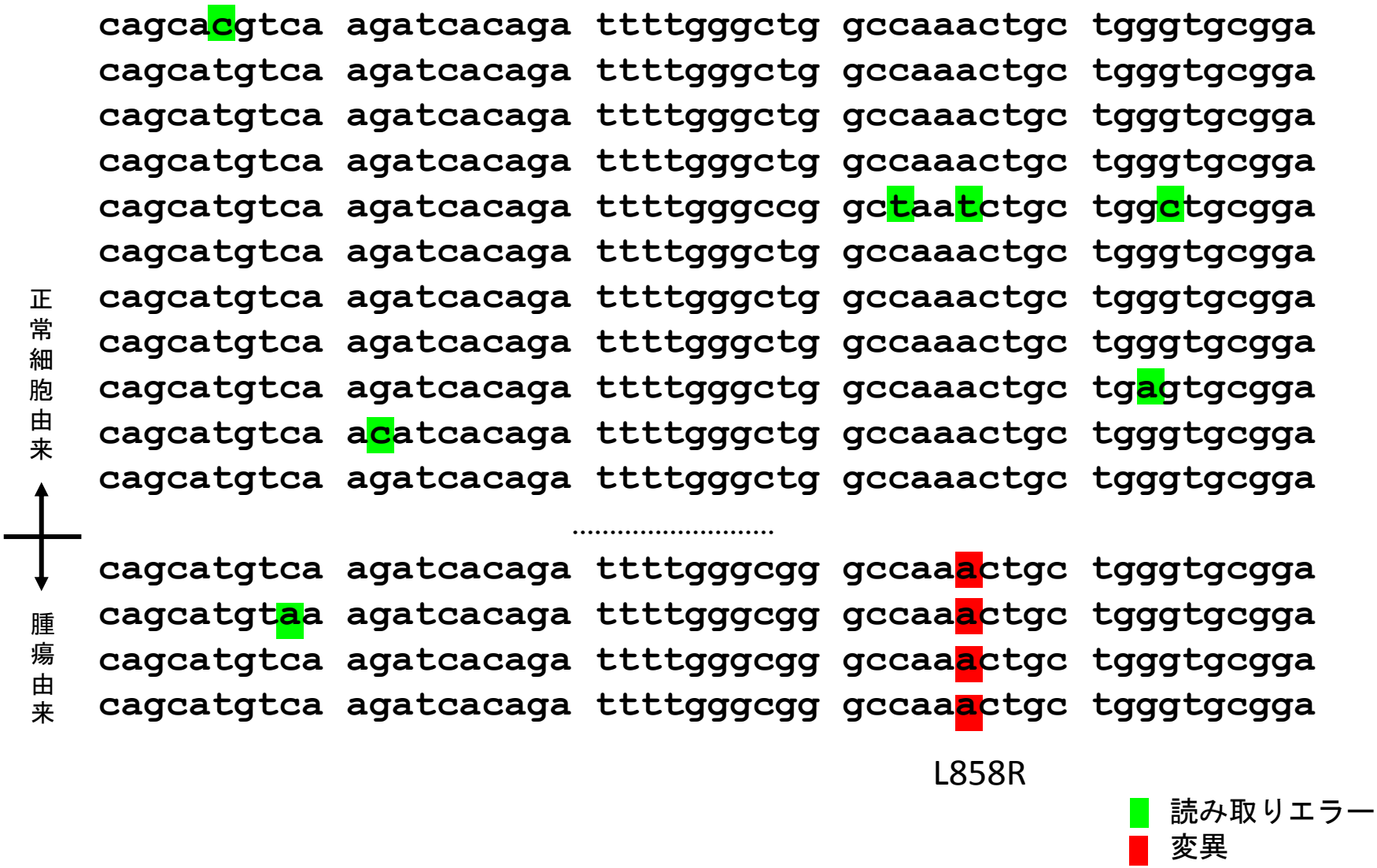
水−油エマルジョン中の1つの水滴中にプライマーの付いたビーズ1個、鋳型DNA1個が入るように調整して行う



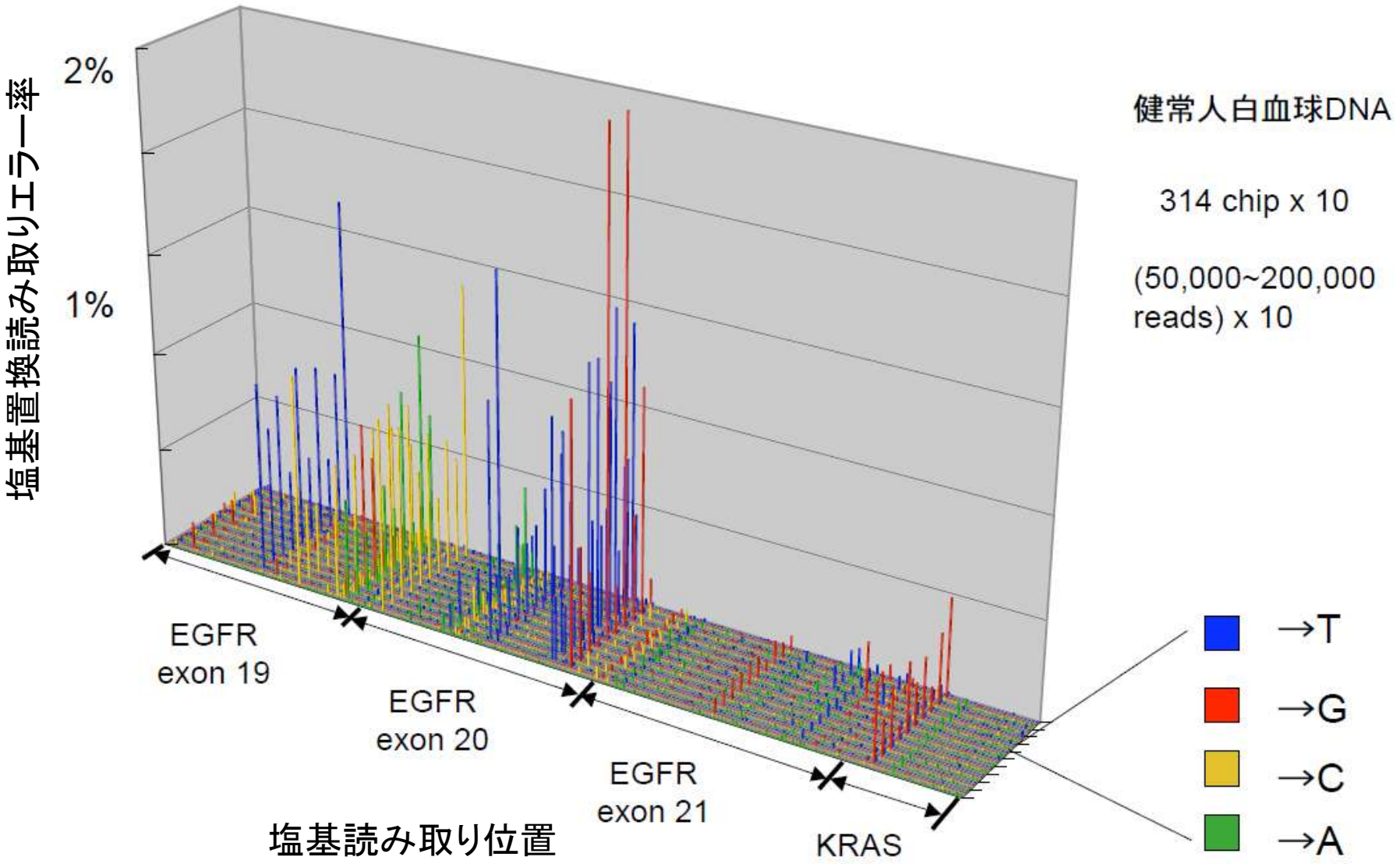


Rothberg JM et al., Nature, 475:348-352, 2011

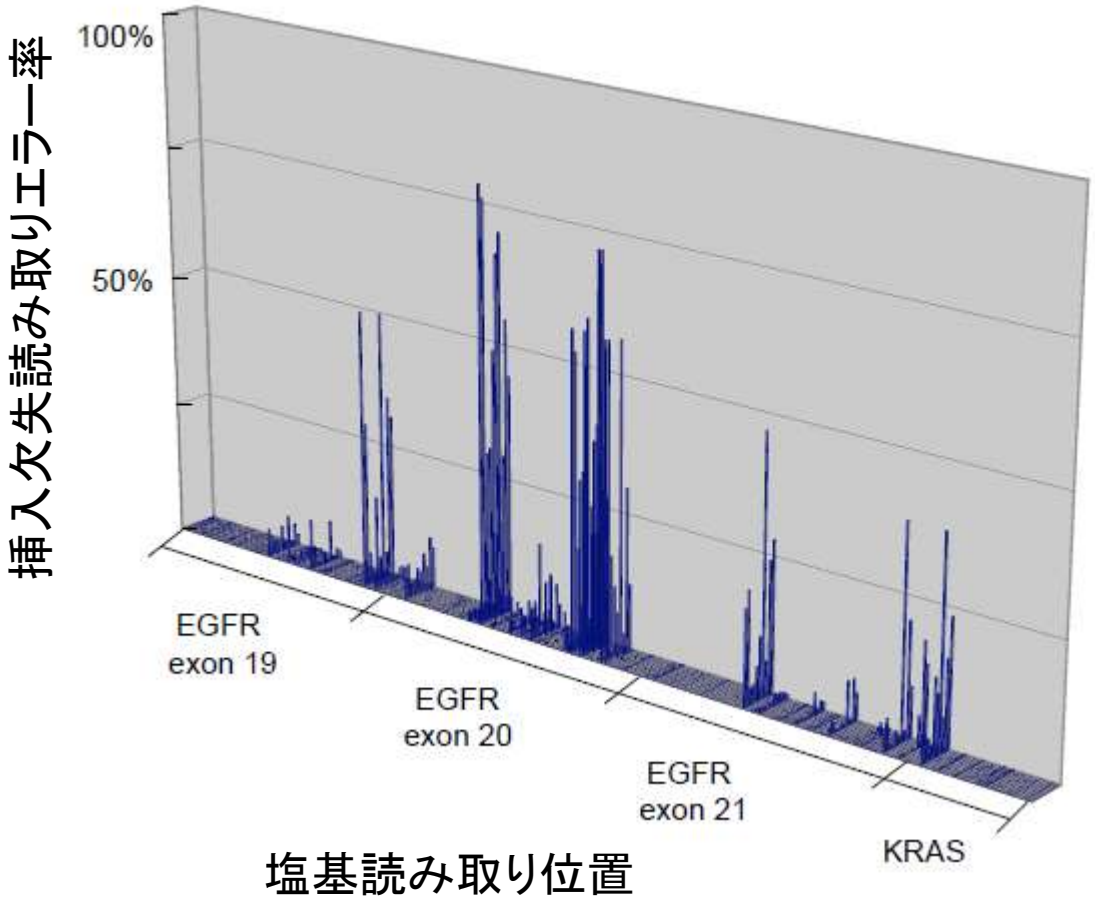
Massive Amplicon Sequencing (MAS)



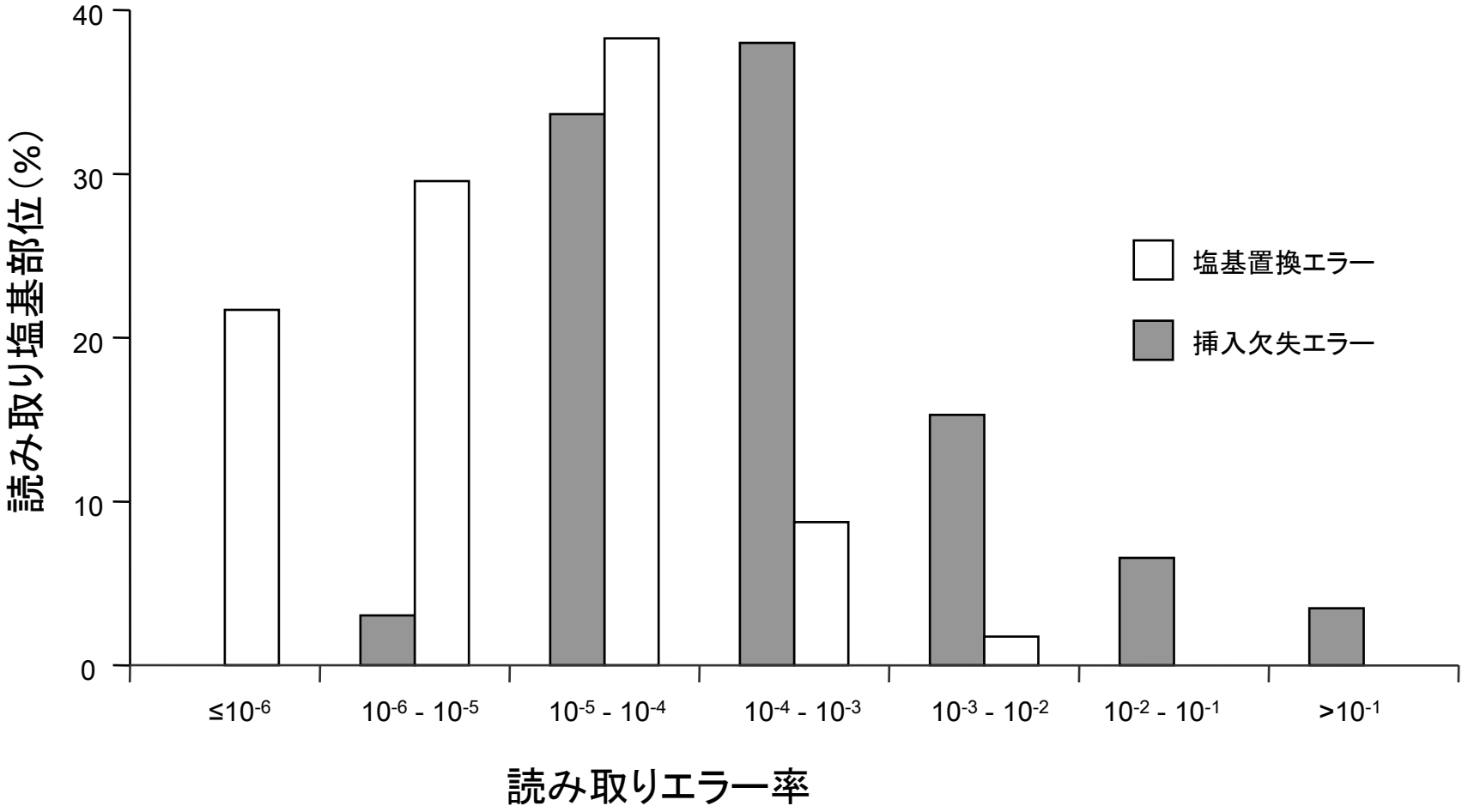
塩基置換読み取りエラー（PGM）



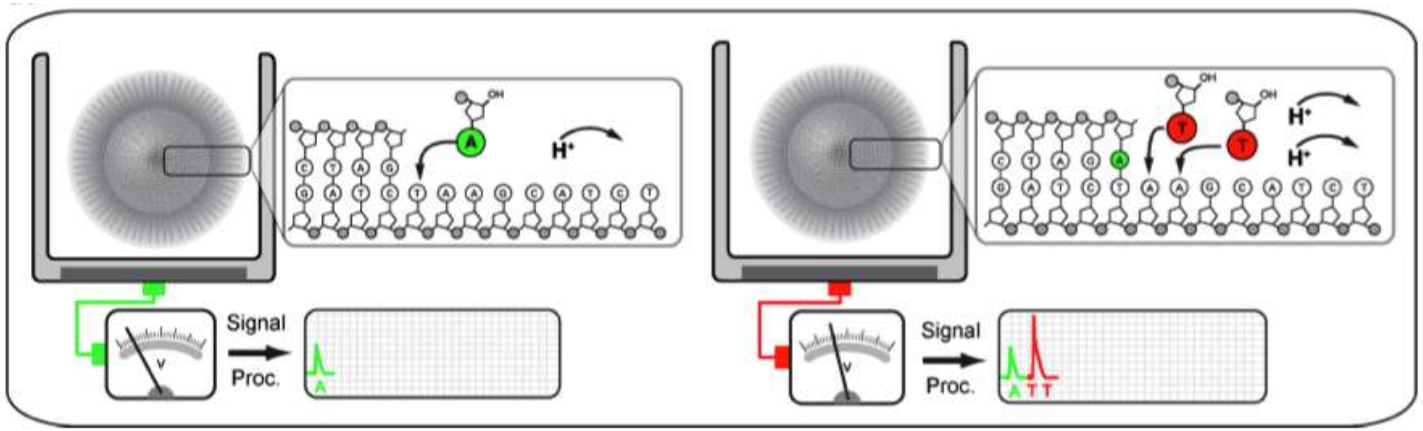
挿入欠失読み取りエラー（PGM）

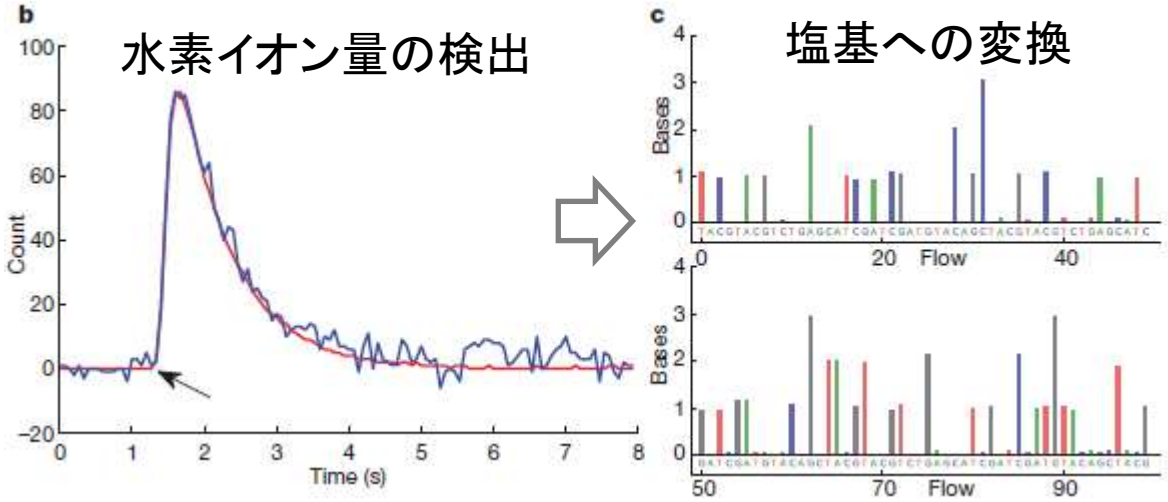


読み取りエラー（塩基置換と挿入欠失）



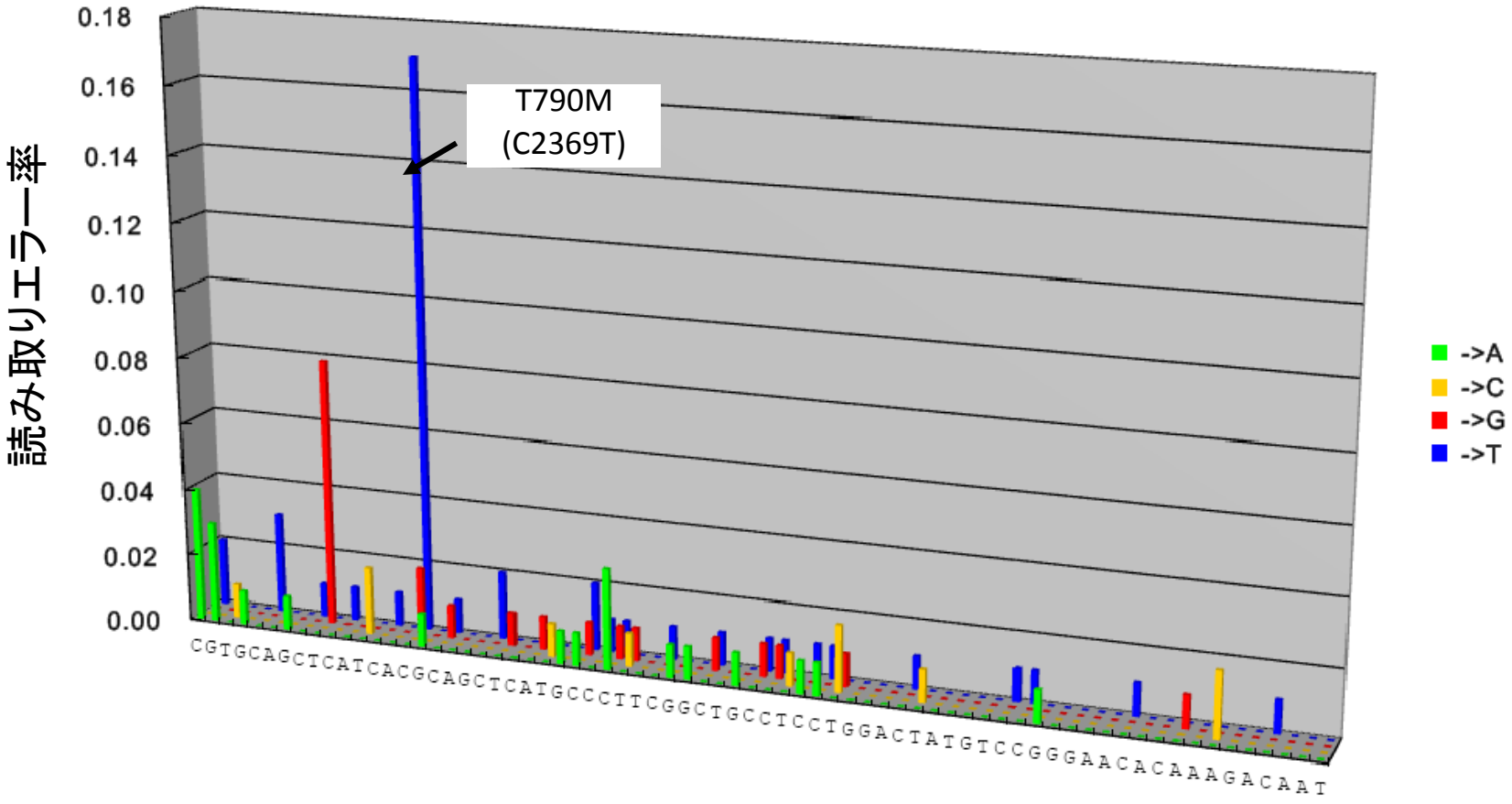
Ion Torrent PGMでのシーケンス





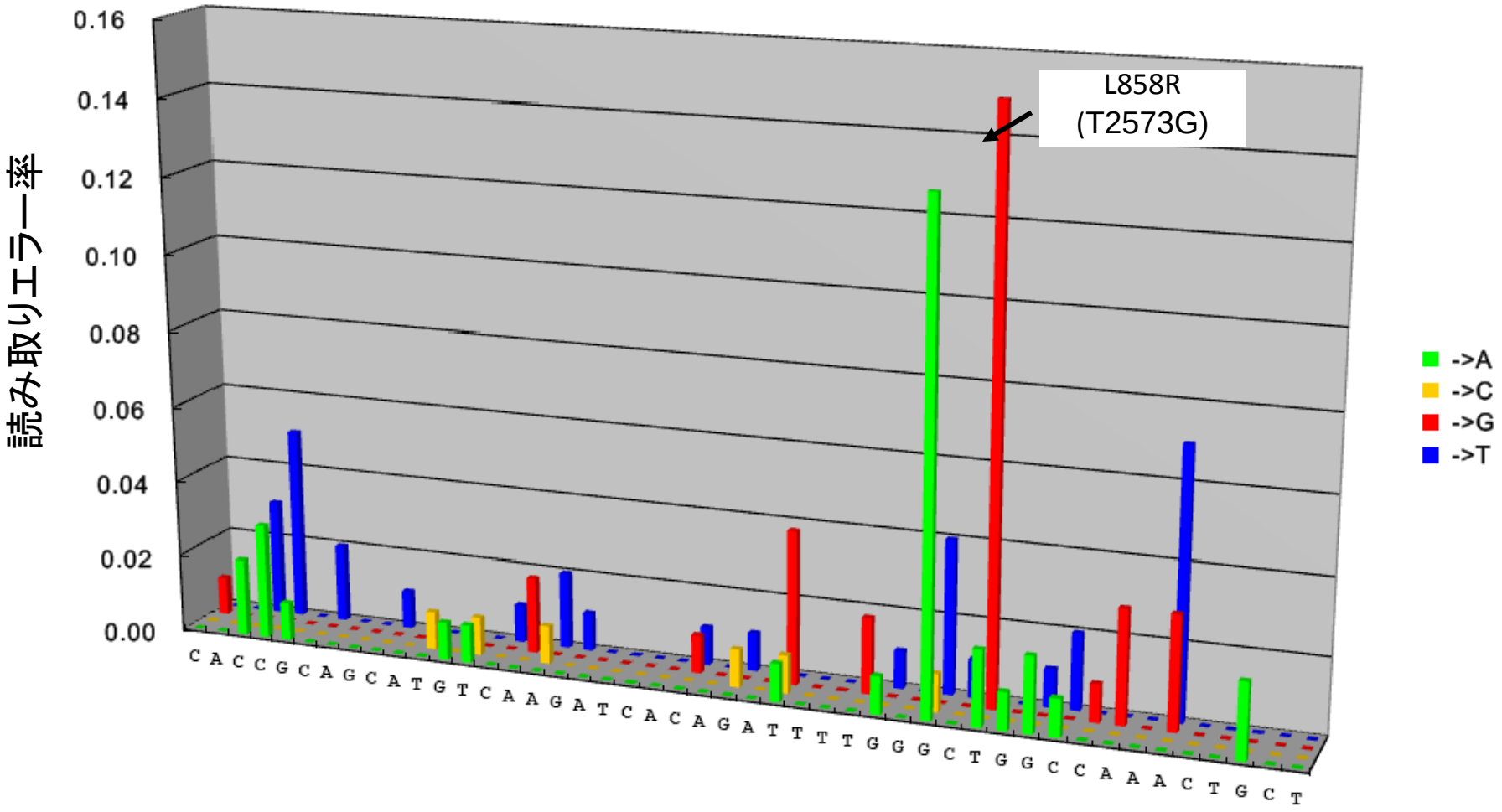
Rothberg, J. et al. Nature 2011, 475, 348.

Exon 20 (T790M) patient no.4



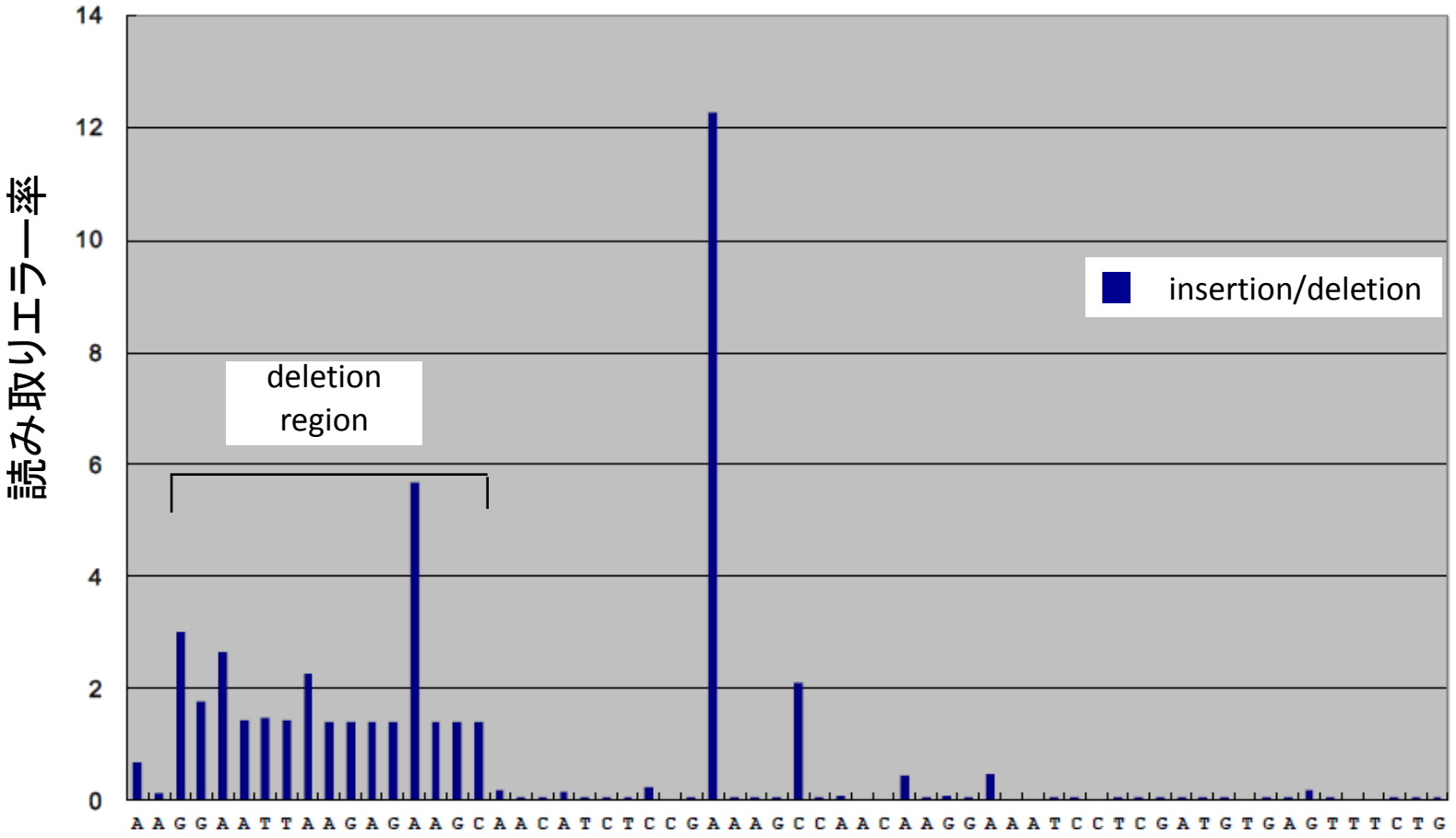
Total read, 98,816; BEAMing（別の検出方法）, 0.08%; Sequencing, 0.17%

Exon 21 (L858R) patient no.4



Total read, 157,201; BEAMing, 0.13%; Sequencing, 0.15%

Exon 19 (deletion) patient no.6



Total read, ~262,000; BEAMing, 1.03% ; Sequencing, ~1.4%

まとめ

- 次世代シーケンサーは集団及び個人レベルでの疾患原因変異の検出（疾患の診断）に有効である。

- 次世代シーケンサーは機種によって塩基配列決定法が異なる。そのためデータの特徴・品質について注意が必要である。情報系解析手法が必要な部分は、

  • 検出システムに依存するものでは、
  - 検出器で検出されたシグナルを塩基配列データに変換部分

  •塩基配列データ解析に関しては、
  - レファレンスゲノム配列の整備に依存する面もあるが、エラーを含んだ配列を正確にマッピング/アライメントする部分
  - 変異部位の影響を知るための変異タンパク質機能予測する部分
  - エラーを含む配列集団中に存在する低頻度だが真の変異を検出する部分

共同研究者

大阪府立成人病センター
研究所

加藤菊也，谷口一也

呼吸器外科

兒玉　憲，岡見次郎，
東山聖彦

呼吸器内科

今村文生，内田純二，
西野和美，熊谷　融，
奥山貴子

東京大学大学院

新領域創成科学研究科
菅野純夫，鈴木穣

奈良先端科学技術大学院大学
バイオサイエンス研究科

加藤順也，加藤規子，
中前伊公子

大阪大学蛋白質研究所

川端猛