

# 入門編

# がんの全ゲノム解析

令和5年度

# 目次

#	内容	執筆者
<a href="#">第I章</a>	<a href="#">がんゲノム医療概論</a>	池上恒雄/山口貴世志/古川洋一
<a href="#">第II章</a>	<a href="#">サンプル調製</a>	永澤慧/鈴木穰
<a href="#">第III章</a>	<a href="#">遺伝学的検査とNGSのメリット</a>	南谷泰仁
<a href="#">第IV章</a>	<a href="#">がんゲノムデータの1次解析</a>	中川英刀
<a href="#">第V章</a>	<a href="#">1次解析から得られた結果の臨床的な意義づけ</a>	井本逸勢
<a href="#">第VI章</a>	<a href="#">臨床におけるがんゲノム医療（抗がん剤選択、治験情報活用）</a>	藤澤孝夫
<a href="#">第VII章</a>	<a href="#">がんゲノム医療データベースの活用とデータ共有</a>	鎌田真由美
<a href="#">(コラム)</a>	<a href="#">がんゲノム医療データの統計的活用</a>	加藤護
<a href="#">(コラム)</a>	<a href="#">蓄積情報によるデジタル創薬やドラッグ・リポジショニング</a>	岡田随象/小沼貴裕
<a href="#">(コラム)</a>	<a href="#">AIによるリアルワールドデータやPHRとの統合解析</a>	小貫真希
<a href="#">(コラム)</a>	<a href="#">がんゲノム医療におけるELSI（倫理的・法的・社会的課題）</a>	武藤香織

# 用語の前提

本テキストでは下記に基づいて説明する。

- **変異** : mutationに対応した和訳として使用する。本来、mutationは、例えばリファレンスゲノムと比較した際に見つかるゲノムの違いやがんゲノムで生じている後天的な変化の両方の意味を持つ言葉である。また、pathogenic（病原性）、またはnon-pathogenic（非病原性）のような特定の意味はない。特に、先天的な変異は生殖細胞系列変異、後天的な変異は体細胞変異と呼ばれる。最近ではvariantをmutationと同義な言葉として使用することが増えてきている。SNVは一塩基変異（Single Nucleotide Variant）を示す。
- **多型** : polymorphismに対応した和訳であり、一般集団において1%以上のアレル頻度で見られる変異を表すことが多い。SNPは一塩基多型（Single Nucleotide Polymorphism）を示す。
- **CNVとSV** : Copy Number Variant/Copy Number VariationとStructural Variant/Variationの略語とする。Copy Number Variation/Structural Variationの場合は生殖細胞系列変異の場合を、Copy Number Variant/Structural Variantの場合は生殖細胞系列変異と体細胞変異（後天的な変異）も含む用語を示す。日本語では、CNVはコピー数変異またはコピー数異常、SVは構造変異または構造異常と呼ぶことが多い。「異常」と呼ばれる際には正常ゲノムと比較して見出された体細胞変異や病原性に関連する生殖細胞系列変異のことを示している。

# 第 I 章

# がんゲノム医療概論

東京大学医科学研究所 臨床ゲノム腫瘍学分野 講師 **山口貴世志**  
東京大学医科学研究所 臨床ゲノム腫瘍学分野 教授 **古川洋一**

- ✓ がんゲノムの理解に必要な基礎知識（池上）
- ✓ がんの全ゲノム解析に用いられるシーケンス技術（山口）
- ✓ がんの全ゲノム解析結果の解釈（古川）

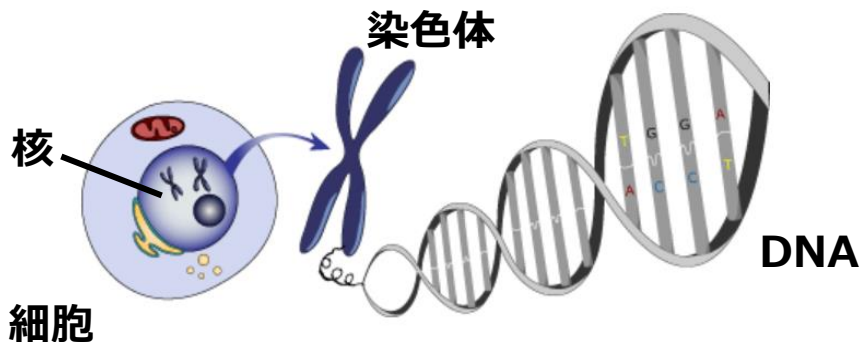
まず、がんゲノムの理解に必要な基礎知識を身につける

● ゲノム

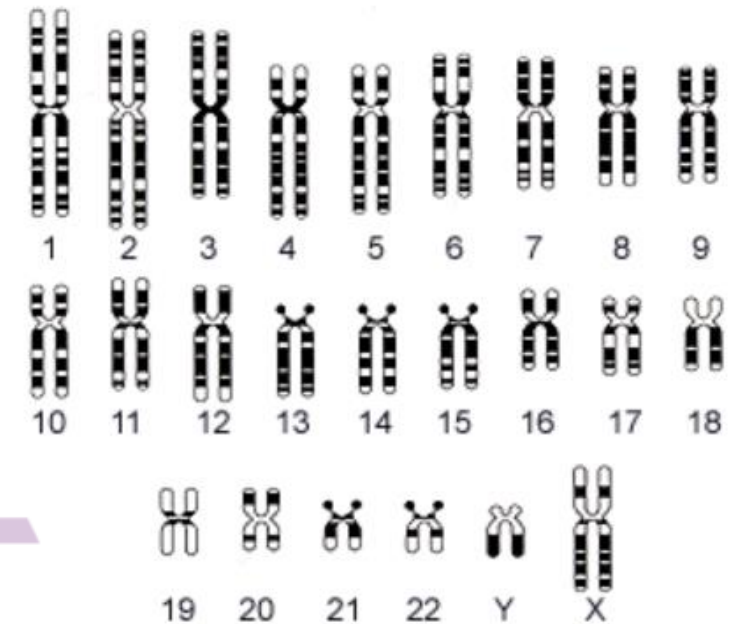
- ✓ **ゲノム**とは、生物体を構成する細胞に含まれる遺伝情報の総体のことを指す
- ✓ ヒトの場合23本の染色体（常染色体22本+性染色体1本）およびミトコンドリアDNAに含まれる約30億塩基対のDNA情報
- ✓ 細胞内で様々な機能を担うタンパク質をコードする“**遺伝子**”が2万数千種類存在
- ✓ 遺伝子が占める領域はゲノム全体の2%に過ぎず、その他の領域は遺伝子の発現制御に関わる領域や機能未知の領域

● DNA

- ✓ **DNA**とは、遺伝子を構成している物質のことを指す
- ✓ アデニン（A）、チミン（T）、グアニン（G）、シトシン（C）という4種類の塩基が共有結合により鎖のように一列に繋がり、さらにアデニンとチミン、グアニンとシトシンが水素結合により対（塩基対）になることにより2本の鎖がらせん状の構造（2重らせん構造）を形成



ヒトの染色体は23本で1セット  
細胞内には2セットが存在



(図はすべて放射線医学総合研究所ウェブサイトから。一部改変  
[https://www.nirs.qst.go.jp/research/review/radgenomics/index\\_re\\_genome.php](https://www.nirs.qst.go.jp/research/review/radgenomics/index_re_genome.php))

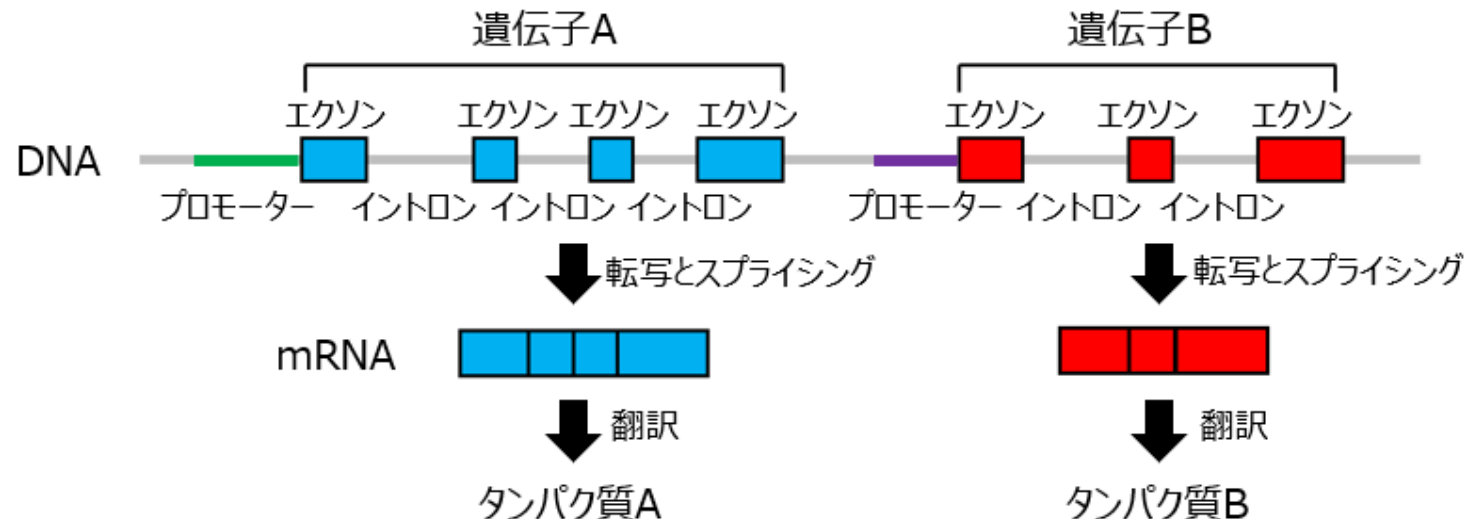
遺伝子とは、細胞内で何らかの機能を担うタンパク質のアミノ酸配列情報をコードしているゲノム内の領域のことを指す

- ヒトの体は約37兆個の細胞からできており、それぞれの細胞には2万数千個の遺伝子がある
- 個々の遺伝子は臓器や組織ごとに発現時期や発現量が異なり、それらが適切にコントロールされることによってヒトの体の恒常性が保たれる
- それぞれの遺伝子にはその発現を制御するプロモーター、遺伝情報がコードされている部分（**エクソン**）、エクソン間に介在し遺伝情報がコードされていない部分（**イントロン**）があり、遺伝子のDNAからmRNAが合成される過程（**転写**）でイントロンが除去され、エクソンが連結される（**スプライシング**）
- 転写で合成されたmRNAの情報に基づいて、細胞質のリボソームでタンパク質を合成する過程を**翻訳**という

遺伝子の構造及び発現の仕組み（右図）

（図）セントラルドグマ

- 遺伝子A、Bはそれぞれのプロモーターにより発現制御され、転写とスプライシングによりmRNAが合成される
- 転写で合成されたmRNAは細胞質のリボソームにおいて翻訳され、タンパク質A、Bとなる
- このDNAからタンパク質までの流れのことを**セントラルドグマ**（central dogma）という



# 点変異・欠失・挿入

点変異・欠失・挿入とは、DNA複製時のエラーか損傷後のDNA修復のミスにより発生するものを指す

- **点変異**：DNA配列上の1塩基が別の塩基に置き換わる変異
- **欠失**：DNA配列上の塩基が失われる変異
- **挿入**：DNA配列上に新たな塩基が付加される変異
- **非同義変異 (Nonsynonymous mutation)**：アミノ酸が置換する変異
  - ✓ **ミスセンス変異**：アミノ酸が別のアミノ酸に置換する変異
  - ✓ **ナンセンス変異**：アミノ酸が終止コドンに置換する変異
- **同義変異 (Synonymous mutation)**：アミノ酸が置換しない変異（サイレント変異と同義）
- **フレームシフト変異**：塩基の欠失や挿入により読み枠がずれる変異。アミノ酸の変化に加え、終止コドンのずれを生じる
- **点変異、欠失、挿入の例**（右図）
  - ✓ 点変異の例では赤字で示した1塩基が参照配列でのTからGに置き換わっており、その結果アミノ酸がセリン（Ser）からアラニン（Ala）に置換されている（ミスセンス変異）
  - ✓ 欠失の例では1塩基（T）が失われ、挿入の例では赤字で示したGが新たに付加されることにより、その部位以降のアミノ酸配列がずれている（フレームシフト変異）

参照配列

```
His Ser Pro Val Pro...
CATTCACCTGTACCA
GTAAGTGGACATGGT
```

点変異

```
His Ala Pro Val Pro...
CATGCACCTGTACCA
GTACGTGGACATGGT
```

欠失

```
His His Leu Tyr...
CATCACCTGTACCA
GTA GTGGACATGGT
```

挿入

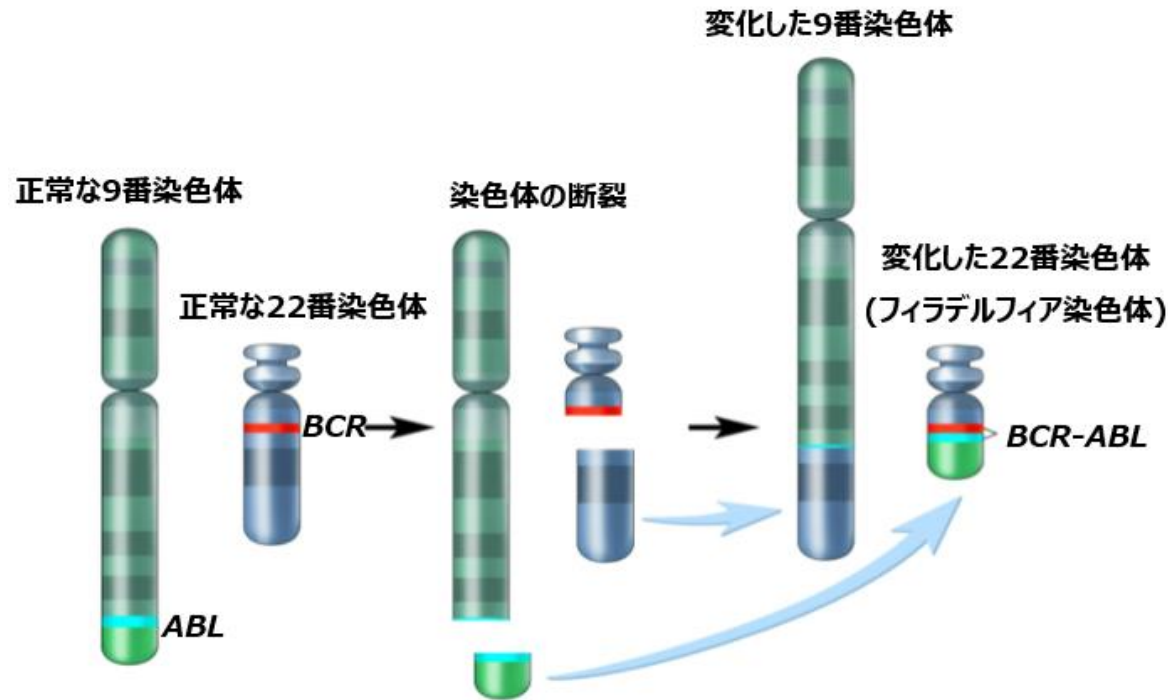
```
His Val Thr Cys Thr...
CATGTCACCTGTACCA
GTACAGTGGACATGGT
```



# コピー数異常・構造異常

コピー数異常や構造異常は細胞分裂の時に発生しやすい。これらが起きると遺伝子の発現が失われたり、過剰に発現したり、新たな機能を持つ融合遺伝子が生じたりするという問題が生じる

- **コピー数異常 (CNV)** : 細胞内に染色体のある領域が通常は2コピー存在するが、0または1コピーに減少していたり、3コピー以上に増加していたりすること
- **構造異常 (SV)** : 染色体の一部または全体が欠けたり、他の染色体と不適切に結合すること (欠失、重複、逆位、転座など)。転座により生じた融合遺伝子や転座した遺伝子のエンハンサー (発現制御領域の一種) により切断点近傍の遺伝子が高発現することがある。例: *BCR-ABL*、*EML4-ALK*



## BCR-ABL融合遺伝子 (左図)

慢性骨髄性白血病の90%以上に見られる融合遺伝子で、9番染色体長腕 (9q34) に座位するABL遺伝子と22番染色体長腕 (22q11) に座位するBCR遺伝子との相互転座により生じる。その結果チロシンキナーゼ活性が亢進した融合タンパク質が産生され、細胞増殖が促進される

(NIHのウェブサイトの図を改変 <https://nci-media.cancer.gov/pdq/media/images/533336.jpg>)

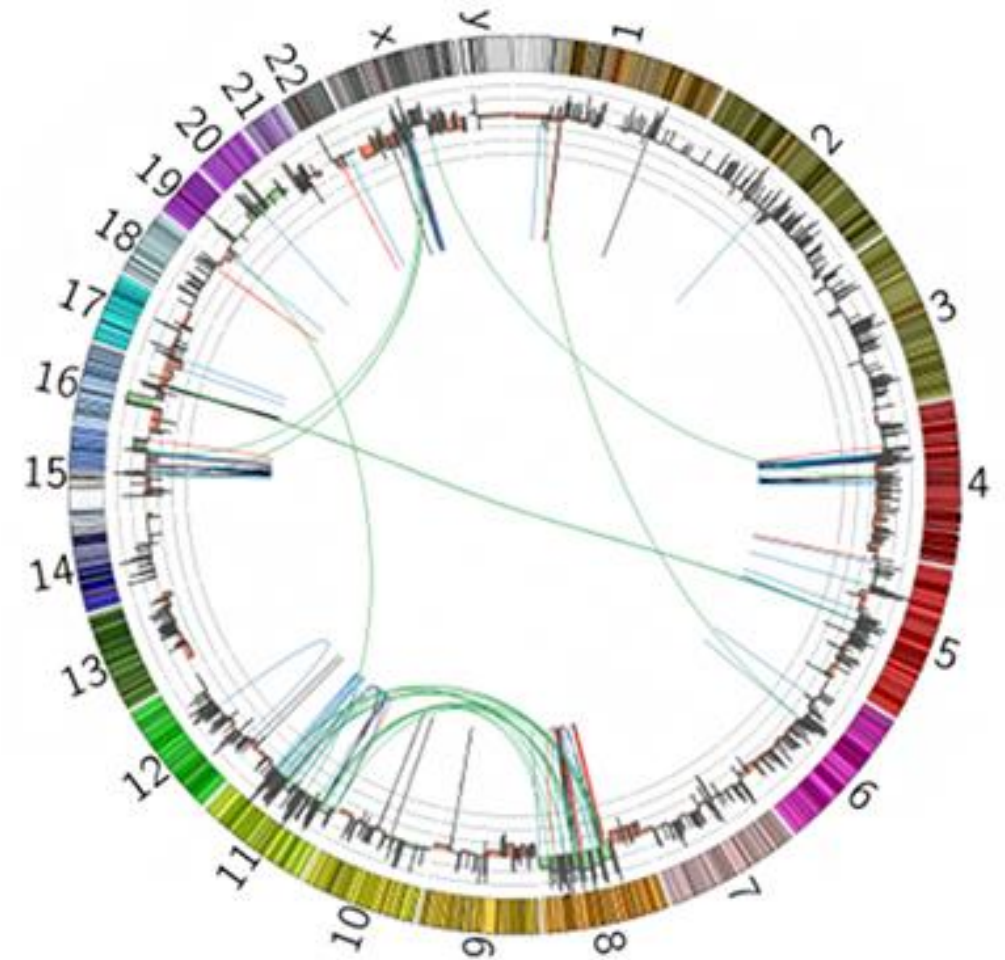
# Circos plot

Circos plotとは、ゲノム構造の視覚化のために利用されるツールである

- **Circos plot**では、1番から22番染色体、X染色体Y染色体を環状に並べて、がんの全ゲノムレベルでの染色体異常を示す
- 染色体上の2座標間の構造異常による結合を時計回りに環状に配置された染色体間の線で表現する
- 膨大な量のゲノム情報を非常に効率的に視覚化できるということが最大の利点

## Circos plotの例（右図）

肝細胞がんの1症例の全ゲノムレベルでの染色体異常を表したもの。もっとも外側の環は染色体を、内側の環は染色体の一部分のコピー数の変化（増幅や欠失）を、さらに内側のループは染色体の一部分同士の再編成（転座：構造異常の一種）を表している。この症例では8番染色体と11番染色体の間で、多数の転座が見られる



(理化学研究所ウェブサイトより [https://www.riken.jp/press/2020/20200206\\_1/index.html](https://www.riken.jp/press/2020/20200206_1/index.html))

がん遺伝子とは、がんを促進する潜在的機能を有するタンパク質をコードする遺伝子のことを指す

- **がん遺伝子**には、変異により本来の制御から外れて、細胞増殖シグナル促進能、細胞周期促進能、細胞死抑制能などを有する遺伝子が多い
- 点変異、増幅、染色体転座による融合遺伝子の形成や、染色体転座によるエンハンサー活性が高い領域への移動によりがん遺伝子が活性化される
- がん遺伝子の異常（変異、増幅、遺伝子融合など）が見られた場合、その遺伝子産物（タンパク質）に対する薬剤（分子標的薬）が奏効する可能性がある

例： EGFR変異に対するゲフィチニブ（イレッサ®）  
HER2増幅に対するトラスツズマブ（ハーセプチン®）  
BCR-ABL融合遺伝子に対するイマチニブ（グリベック®）

代表的ながん遺伝子

遺伝子	主な機能	変異の多い癌
KRAS	低分子GTP結合タンパク質	膵癌、大腸癌
BRAF	セリンスレオニンキナーゼ	メラノーマ、大腸癌
PIK3CA	脂質キナーゼ	乳癌、子宮体癌
EGFR	受容体型チロシンキナーゼ	非小細胞肺癌
HER2	受容体型チロシンキナーゼ	乳癌、胃癌
MET	受容体型チロシンキナーゼ	腎癌、非小細胞肺癌
KIT	受容体型チロシンキナーゼ	消化管間質腫瘍
RET	受容体型チロシンキナーゼ	甲状腺髄様癌
MYC	転写因子	悪性リンパ腫、乳癌
BCR-ABL	融合遺伝子（チロシンキナーゼ）	慢性骨髄性白血病
EML4-ALK	融合遺伝子（チロシンキナーゼ）	非小細胞肺癌

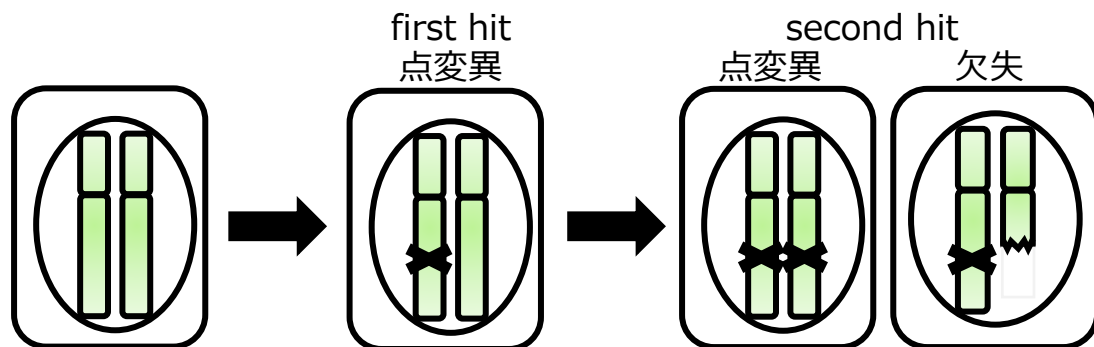
がん抑制遺伝子とは、がんを抑制する機能を有するタンパク質をコードする遺伝子のことを指す

- **がん抑制遺伝子**は、DNA修復機能、アポトーシス誘導能、細胞増殖シグナル抑制能、細胞周期抑制能などを有する遺伝子が多い
- 欠失、点変異、プロモーターのメチル化などによりがん抑制遺伝子が不活化される
- **2ヒット仮説**（2コピーの遺伝子がともに失われることによりがんを生じるという説）、**ハプロ不全**（1コピーの遺伝子が失われるだけでもがん化に寄与すること）、**ドミナントネガティブ効果**（変異したアレルから生じるタンパク質が野生型のアレルから生じるタンパク質の機能を阻害する現象）などの機序でがん化に寄与する

アレル (対立遺伝子)

- ✓ 相同な遺伝子座を占める遺伝子に複数の種類がある場合にその個々の遺伝子を意味する
- ✓ ヒトをはじめ2倍体の生物はそれぞれ父母に由来する二つのアレルを有する
- ✓ 野生型とは集団の中で最も多数を占めている型を指し、野生型アレル由来のタンパク質は正常な機能を有するとみなされる

がん抑制遺伝子の2ヒット仮説

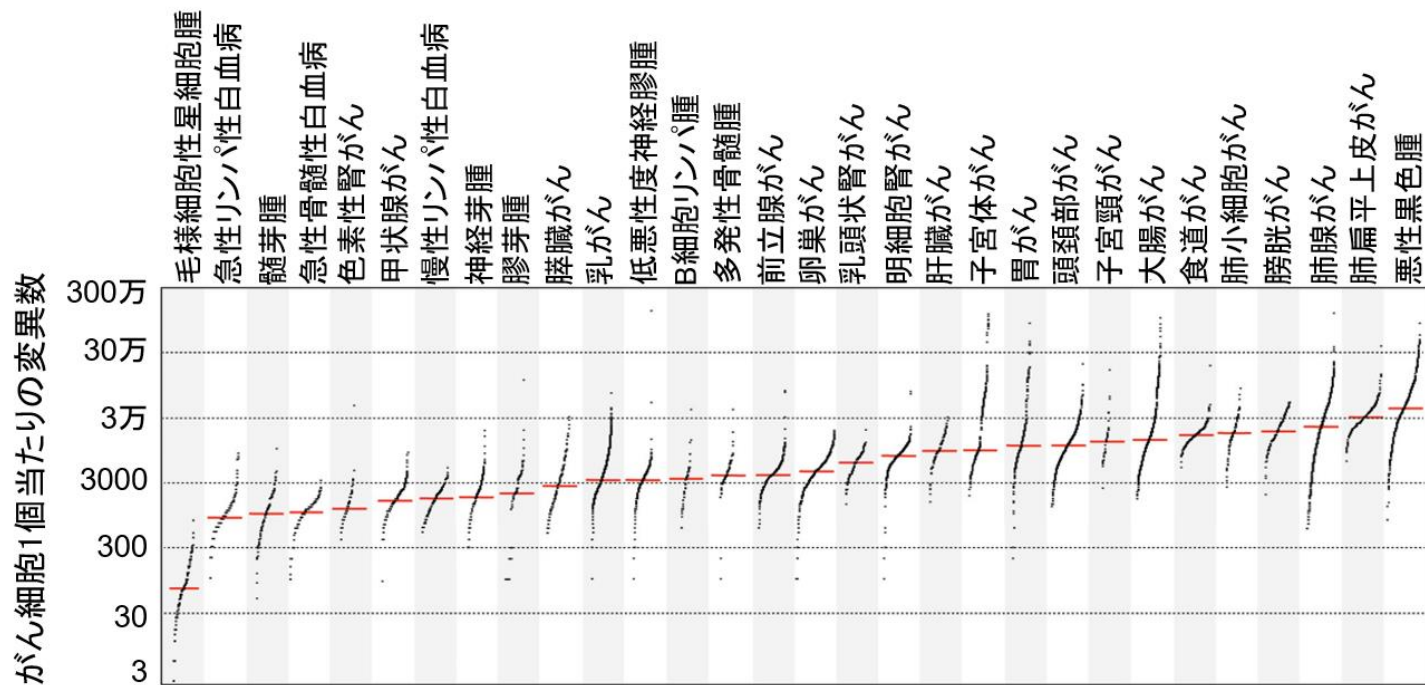


代表的ながん抑制遺伝子

遺伝子	主な機能	変異の多い癌
TP53	転写因子	様々な臓器由来の癌
RB1	細胞周期制御	網膜芽細胞腫、肺癌
PTEN	脂質・タンパク質ホスファターゼ	子宮体癌、膠芽腫
BRCA1	相同組み換え修復	乳癌、卵巣癌
BRCA2	相同組み換え修復	乳癌、卵巣癌
MLH1	ミスマッチ修復	大腸癌、子宮体癌
MSH2	ミスマッチ修復	大腸癌、子宮体癌
APC	β-catenin制御	大腸癌
SMAD4	転写因子	膵癌、大腸癌
VHL	ユビキチンリガーゼ	腎癌
STK11	セリンスレオニンキナーゼ	肺癌、子宮頸癌

## がんは遺伝子変異が蓄積して生じる

- ほとんどのがんは喫煙、生活習慣、加齢など後天的要因により体細胞に遺伝子異常が蓄積することにより生じるが、一部のがん（遺伝性のがん）は親から原因となる遺伝子変化を受け継ぎ、そこに後天的要因が加わることで生じる
- 1個のがん細胞には数百から数百万個の変異が存在することが分かっている
- がんの治療法を開発するには、ゲノムを解析し変異を同定し、発がんメカニズムの詳細を知ることが不可欠



(Nature 500:415-21,2013. doi: 10.1038/nature12477を改変)

ゲノム不安定性とは、ほとんどのがん細胞にみられ、細胞ががん化に必要な数の突然変異を蓄積するのに必須な特質と考えられる

- がん細胞では**DNA損傷修復機構**（細胞内で生じる活性酸素やX線、抗がん剤など外来刺激により生じたDNAの損傷を修復する仕組みで、2本鎖の切断の修復にはBRCA1、BRCA2などのタンパク質が関与する）や**ミスマッチ修復機構**（DNAの複製ミスによる塩基対のミスマッチや短い欠失や挿入を修復し、完全な塩基対に戻す仕組みでMLH1、MSH2などのミスマッチ修復タンパク質が関与する）の異常により染色体数や塩基配列の異常が生じやすい
- ゲノム不安定性には下記種類が存在する（ただし、両者を併せ持つことは非常にまれ）：
  - ✓ **染色体不安定性（Chromosomal instability/CIN）**：DNA損傷修復機構の異常により染色体の欠失や重複を高頻度に生じる
  - ✓ **マイクロサテライト不安定性（Microsatellite instability/MSI）**：DNAミスマッチ修復機構の異常によりゲノム上の繰り返し配列（マイクロサテライト）に高頻度に変異を生じる

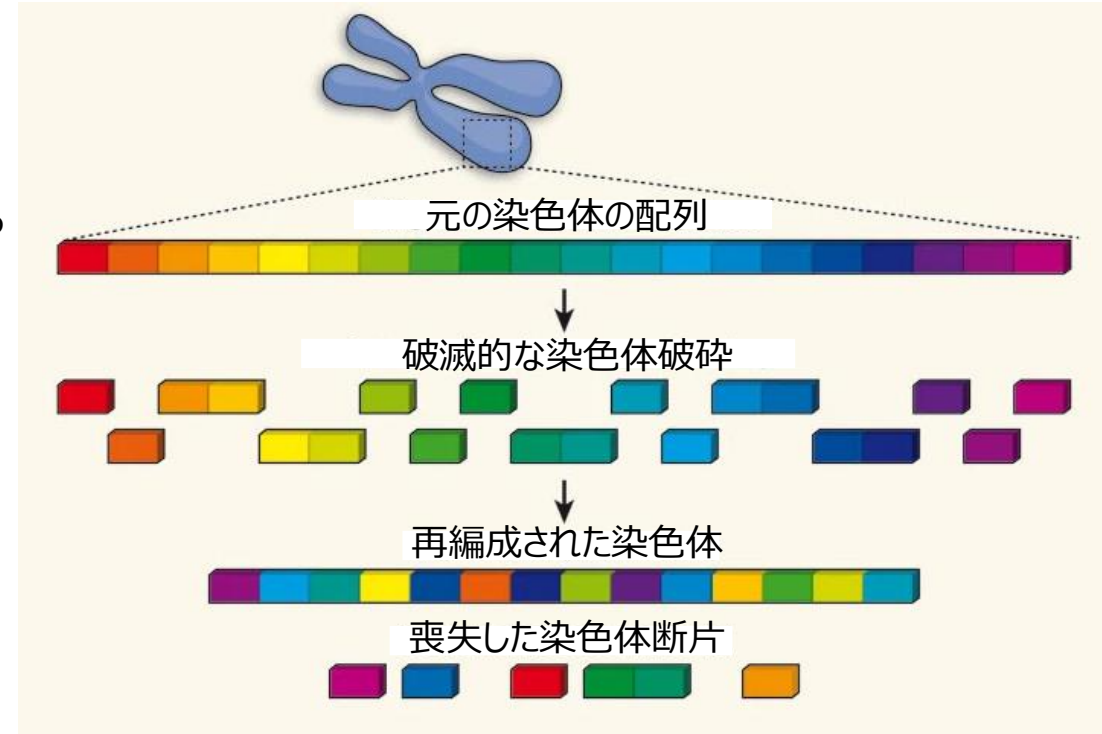
散発性大腸癌における染色体不安定性とマイクロサテライト不安定性

	CIN	MSI
特徴	異数性（Aneuploidy） 増幅（Amplification） 欠失（LOH）	MLH1プロモーターメチル化 腫瘍遺伝子変異量高値 （TMB* high）
頻度	70-80%	20-30%
部位	左側結腸、直腸	右側結腸
関連遺伝子	APC、KRAS	BRAF、TGFB2

\*TMB：Tumor mutation burden（詳細説明は後述）

クロモスリプシスとは、一度の破滅的なイベントにより1つ、もしくはごく少数の染色体において数十～数千箇所にも及ぶ再編成がおこる現象のこと

- **Chromothripsis** : chromosome (染色体) の略語である*chromo*と、ギリシア語で「粉々に砕く」という意味を表す*thripsis*に由来する造語
- 単一のイベントによって染色体が多数の断片に破碎された後、ゲノム修復機構により染色体の再構築が行われるが、損傷の程度が大きいため、何百もの異なる突然変異（重複、欠失、逆位などあらゆる現象）が“修復された”染色体に一度に組み込まれる
- クロモスリプシスを起こした細胞の多くは**アポトーシス**（プログラムされた細胞死であり、細胞が自分の役目を終えたり、不要になるとみずから死ぬ（自殺）現象）により排除されるが、一部の細胞で融合遺伝子の形成、がん抑制遺伝子の破壊や欠失、がん遺伝子の増幅などによりがん化が促進される
- 長期間にわたり遺伝子異常が蓄積して発がんするという既存の概念に相反する現象
- 従来のゲノム解析方法では発見されていなかったが、次世代シーケンサーを用いた全ゲノム解析で、がんゲノムの塩基配列をくまなく解析をすることが可能になったことにより発見された



(Nature 470:476-7,2011. doi:10.1038/470476aを改変)

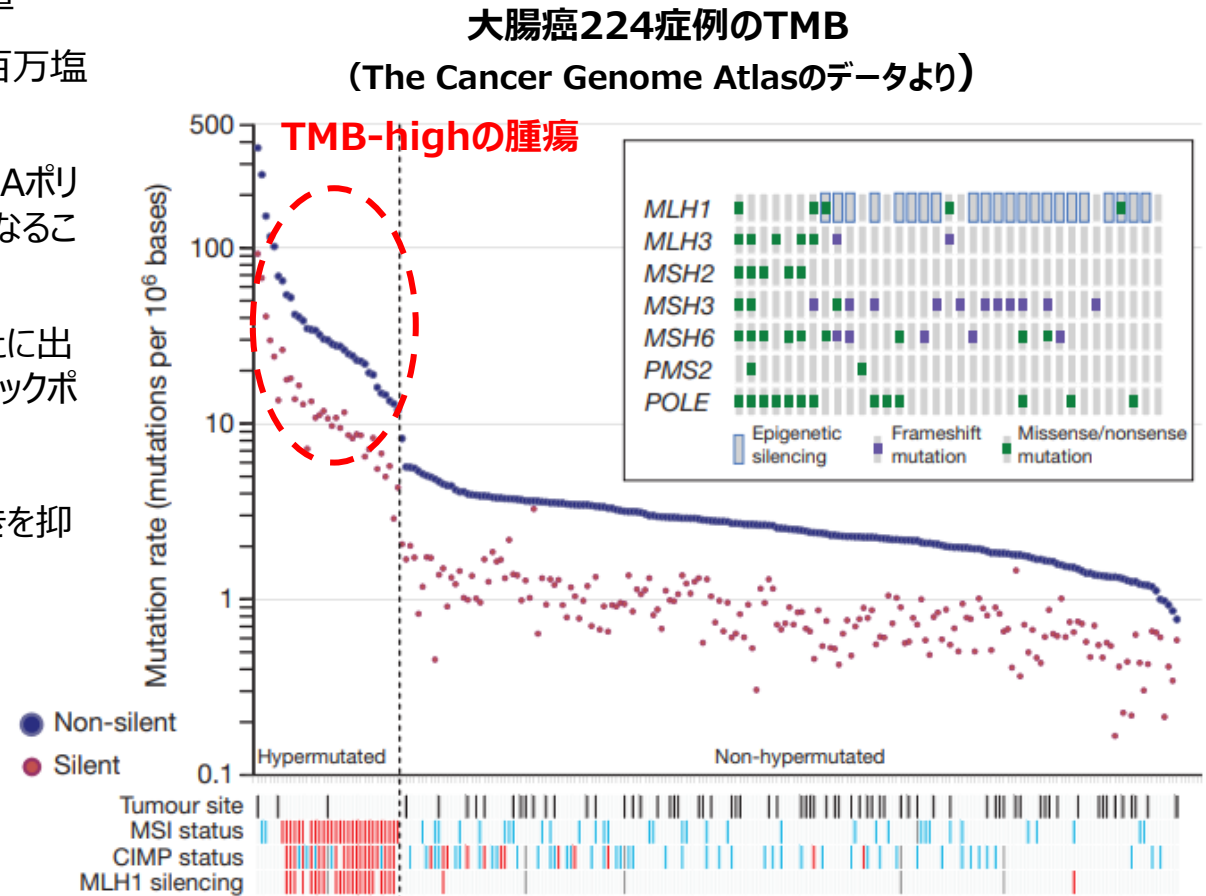
# Tumor mutation burden (TMB)

Tumor mutation burden (TMB) とは、腫瘍遺伝子変異量のことを指す

- がん細胞中の後天的に獲得された（蓄積された）体細胞変異の総量を示す値
- 次世代シーケンシングによりがん細胞と正常細胞の遺伝子配列を比較して、百万塩基（Mb）あたり何塩基の変異が入っているかで表される
- マイクロサテライト不安定性高度陽性（MSI-H）やPOLEやPOLD1などのDNAポリメラーゼ遺伝子に変異を有する腫瘍では**TMB-high**（10変異以上/Mb）となることが多い
- TMB-highの腫瘍は**ネオアンチゲン**（がん細胞特有の遺伝子変異によって新たに出現したがん抗原）がより多く誘導され、PD-1抗体やPD-L1抗体などの免疫チェックポイント\*阻害剤が効果を示しやすい

\***免疫チェックポイント**：免疫が自己の細胞を攻撃しないよう、免疫細胞の働きを抑制する仕組み。がん細胞はこの仕組みを利用して免疫の攻撃を回避する

- 右の図は大腸がん224例のTMBを示したもの



(Nature 487:330-7,2012. doi:10.1038/nature11252を改変)

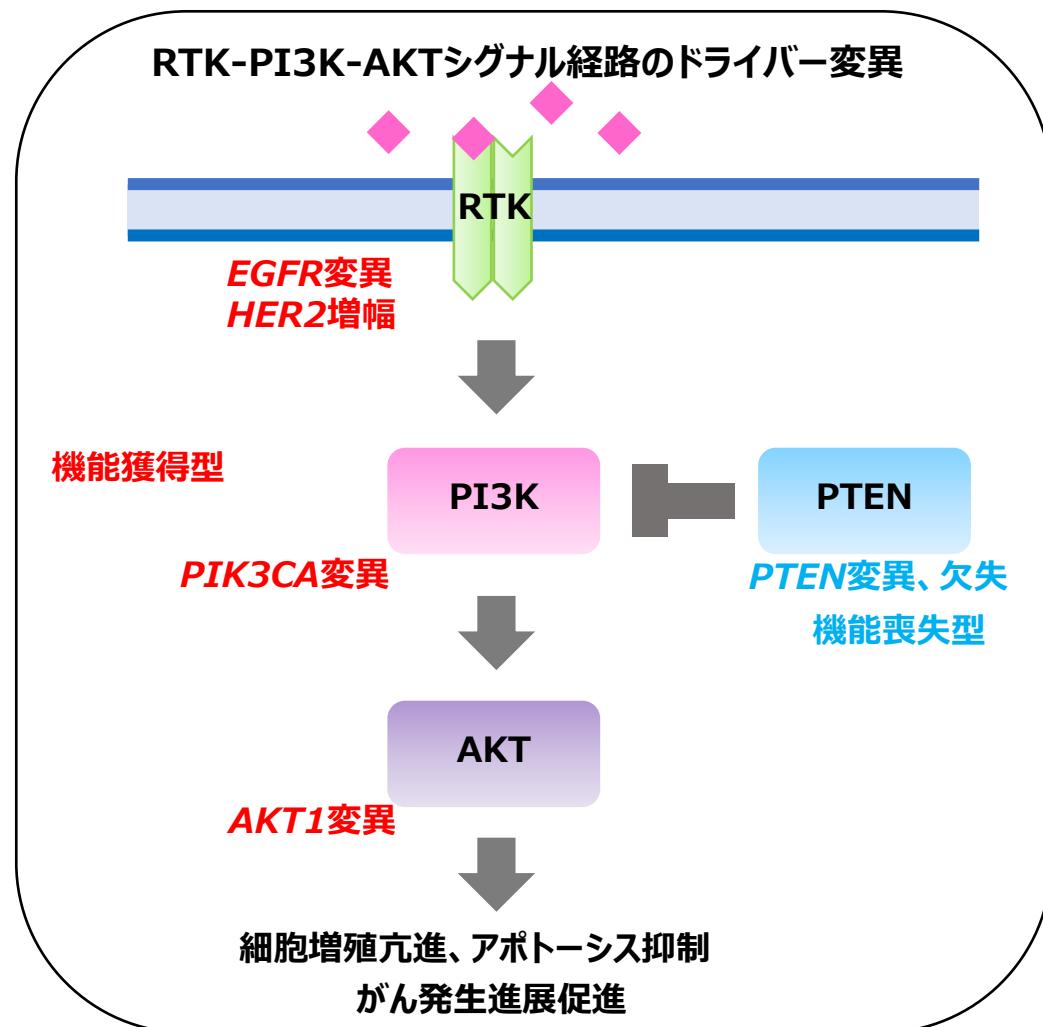


# ドライバー変異とパッセンジャー変異

一方、すべての変異ががん化の原因ではないということに注意しなければならない

- **ドライバー遺伝子 (Driver genes)** : 細胞のがん化 (発がんやがんの悪性化) に直接関わりをもつ遺伝子 (がん遺伝子、がん抑制遺伝子を含む)
- **ドライバー変異** : がんの発生や進展の直接的な原因となる遺伝子変異
- **機能獲得型変異** : コードするタンパク質が新たに機能を獲得する変異のことであり、がん遺伝子の活性化型ミスセンス変異が多い  
例 : KRAS変異、BRAF変異、PIK3CA変異
- **機能喪失型変異** : コードするタンパク質が機能を失う変異。がん抑制遺伝子のミスセンス変異やコードするタンパク質のtruncation (終止コドンが生じて翻訳が途切れ、短いタンパク質が生じること) を生じるナンセンス変異やフレームシフト変異が多い  
例 : PTEN変異、APC変異、BRCA1/2変異
- **パッセンジャー変異** : がんのゲノム不安定性によってランダムに生じるがん化に無関係な遺伝子変異

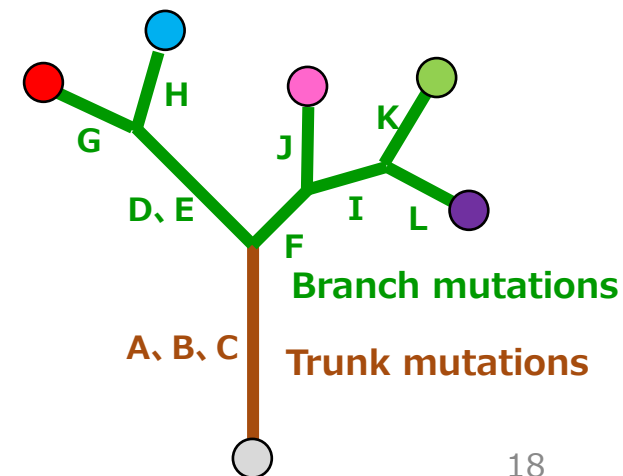
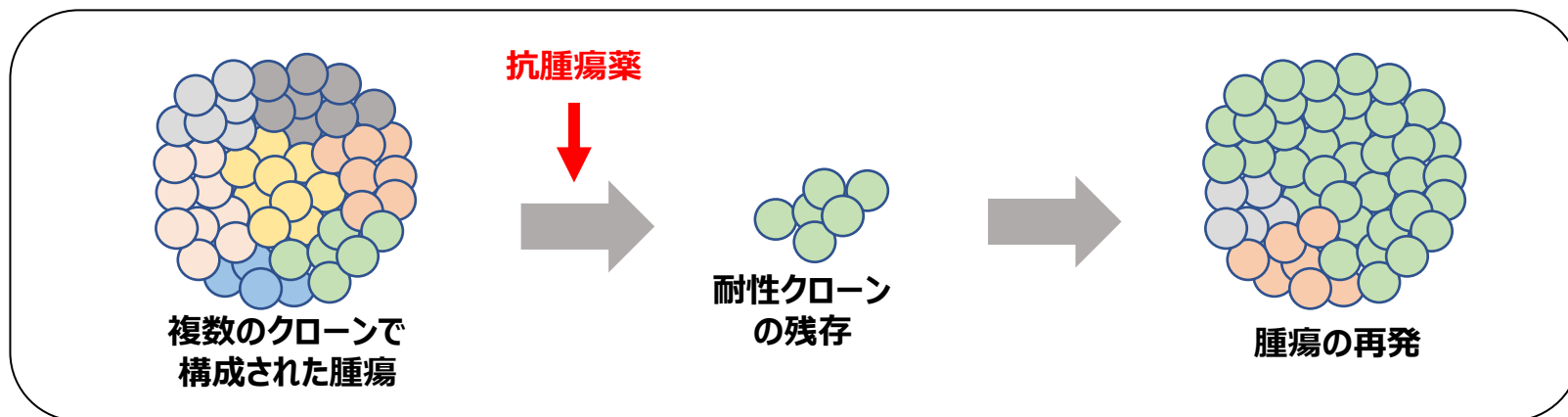
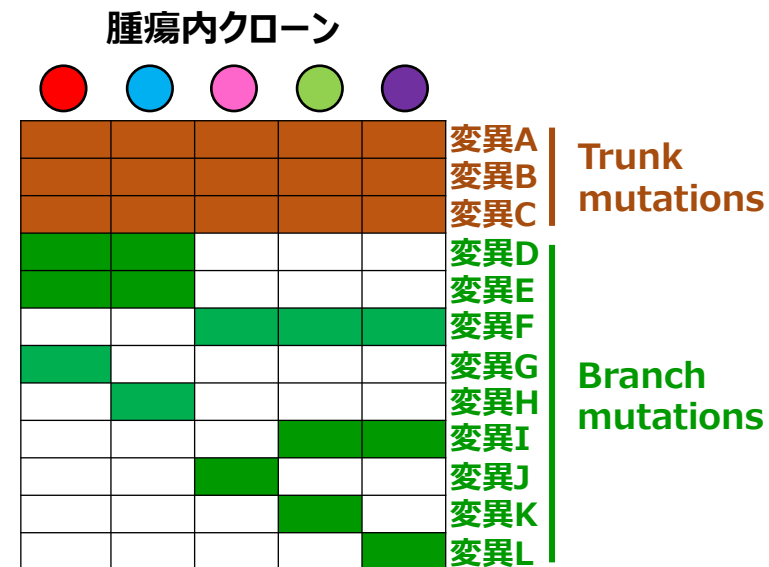
右の図は受容体型チロシンキナーゼ (EGFR、HER2など) が活性化するPI3キナーゼ (PI3K)、AKTを介するシグナル伝達経路の関連遺伝子にヒトのがんで好発するドライバー変異を示している



# 腫瘍内不均一性 (Tumor heterogeneity)

腫瘍は、遺伝子変異の不均一な細胞集団（クローン）の集合体のことを指す

- 右の図は腫瘍内クローンを表す図であり、変異A、B、Cはすべてのクローンで共通であるのに対し、変異D～Lは一部のクローンのみに存在する
- **Trunk mutation**：腫瘍内のがん細胞のすべてのクローンに共通に保存されている、がんの進化の過程で最も重要な変異
- **Branch mutation**：腫瘍内のがん細胞の一部のクローンにのみ存在する変異。腫瘍内不均一性の原因となる
- **治療抵抗性との関わり**：治療により腫瘍が縮小しても治療抵抗性のクローンが残存し、やがてそのようなクローンが増大してくることにより再発に至る
- 下記の図は複数のクローンで構成された腫瘍に対して抗腫瘍薬を投与した後、耐性クローンが残ると、腫瘍が再発する可能性があることを示している



日本でもいよいよゲノム情報を医療に活用する時代が到来した

- **精密医療 (Precision medicine) ・個別化医療 (Personalized medicine)** : 患者のゲノムを解析し、その結果を治療方針の決定や疾病予防など健康管理に活用する医療。2015年米国オバマ大統領が宣言した“Precision Medicine Initiative”により加速
- がんゲノム医療 : がんの領域には精密医療が最も早期に導入された
- 米国でのがんゲノム医療
  - ✓ 2017年11月にがんパネル検査FoundationOne® CDxとMSK-IMPACTをFDAが承認
- 日本でのがんゲノム医療
  - ✓ 2019年6月に保険収載
  - ✓ がんパネル検査
    - NCCオンコパネルシステム 124遺伝子
    - FoundationOne® CDx 324遺伝子
    - FoundationOne® Liquid CDx 324遺伝子
  - ✓ がんゲノム医療を受けられる病院
    - ✓ がんゲノム医療中核拠点病院 12施設
    - ✓ がんゲノム医療拠点病院 33施設
    - ✓ がんゲノム医療連携病院 188施設

## Precision Medicine Initiative (2015)

100万人以上の米国市民からゲノム情報、環境、ライフスタイルなどのデータを集め、精密医療につなげることを目指す計画



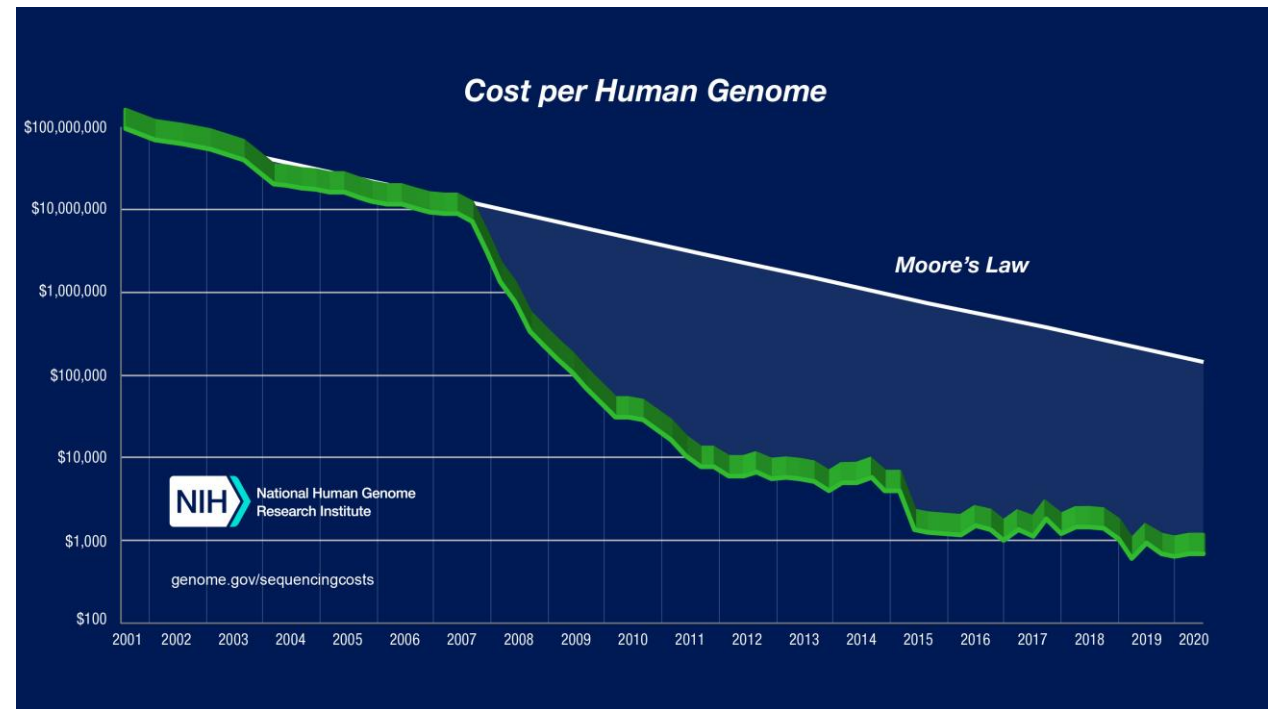
( <https://obamawhitehouse.archives.gov/precision-medicine>より)

# シーケンス技術の進歩と“全ゲノム”の時代へ

シーケンスの精度やスループットは向上し、コストは低下し続けている

- ヒト全ゲノムの解読コストは2015年には1,500ドルを下回り、その後1,000ドルを切ったからは、価格はほぼ横ばいで推移している
- シーケンスコストの低下や計算速度の向上に伴い、“全エクソン”からヒトゲノム全体を解析する“全ゲノム”時代に突入した
- 2005年頃より登場した次世代シーケンス技術は既に成熟しており、現在の全ゲノム解析の中心的役割を果たす
- 短いDNA鎖（ショートリード）と高い精度で配列を解析する次世代シーケンス技術から、“第3世代”と呼ばれるロングリードシーケンス解析も進んでいる
- Ultima Genomicsによる100ドルゲノム時代の到来？

下記図は、ヒトゲノム配列の解析コストを示したグラフとなっている  
NGSの出力データ量はムーアの法則を超えるスピードで上昇している



(<https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>)



(Science誌 2022年6月)

# シーケンサーの種類

シーケンス技術は世代交代を繰り返し、進化してきた

## ● 第一世代（キャピラリーシーケンサー）

- ✓ キャピラリーと呼ばれるゲルを充填したガラスチューブの中で電気泳動を行う
- ✓ 自動化により解読速度が向上したが、大規模な解析には不向き
- ✓ ヒトゲノム計画で使用された（PRISM3700）



サーモフィッシャーサイエンティフィック社  
Applied Biosystems SeqStudio Genetic Analyzer

## ● 第二世代（次世代シーケンサー、NGS）

- ✓ 次世代シーケンスは、**Next-Generation Sequencing**の頭文字をとって**NGS**と呼ばれている
- ✓ ゲノムを短く切断し、PCRで増幅する
- ✓ 多くの反応を並列的に処理し、大規模に解析できるようになった



イルミナ社 NovaSeq 6000  
©Illumina, Inc. All rights reserved.



サーモフィッシャーサイエンティフィック社  
Ion S5

## ● （第三・四世代）

- ✓ NGSとは異なる基盤技術により配列が解析が行われる
- ✓ 長鎖DNAの解析が可能になった



パシフィック バイオサイエンス社 Sequel II



オックスフォードナノポア社  
PromethION

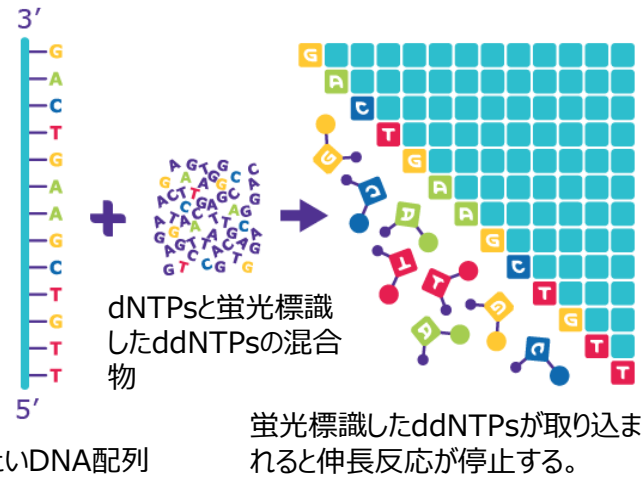
# サンガーシーケンスの原理

サンガーシーケンス（**ダイターミネーター法**）とは、キャピラリー電気泳動で1塩基違いのDNA断片を分離し、流れてきた順に3'末端の蛍光を検出することで、塩基配列を決定する技術のことである

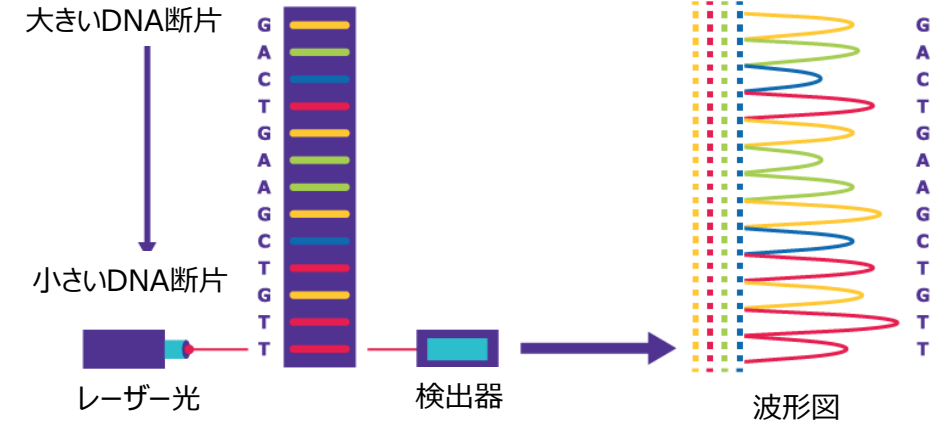
## キャピラリー電気泳動によるシーケンス （ダイターミネーター法）

1. DNAは5'から3'に向かって合成され、ランダムにdNTPsとddNTPsが取り込まれる。最終的に末端が各塩基の蛍光で標識された1塩基違いのDNA断片が生成される
2. ゲルの中を移動する速度は、短い断片は早く流れ、長い断片は時間がかかる
3. レーザーを当て、蛍光標識されたDNA断片が通過すると励起し、光センサーで検出する。その後、シグナルから塩基情報に変換する

### 1 蛍光色素で標識した ddNTPsを取り込ませる



### 2 電気泳動で大きさの違うDNAを分ける



(MERCK社ホームページ(一部改変) : <https://www.sigmaaldrich.com/JP/ja/technical-documents/protocol/genomics/sequencing/sanger-sequencing>)

dNTPs (dATP、dCTP、dGTP、dTTP、デオキシヌクレオチド三リン酸) : DNAの原料

ddNTPs (ddATP、ddCTP、ddGTP、ddTTP、ジデオキシヌクレオチド三リン酸) : DNAの原料となるが伸長は停止する

# 次世代シーケンスとは？

次世代シーケンスは、大規模なデータ量を生み出すことのできる最初のシーケンス技術である

## ● 次世代シーケンサーの特徴：

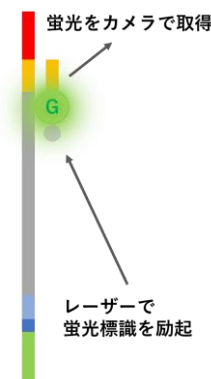
- ✓ ヒトゲノム計画で使用され、当時、広く普及していたサンガー法（第一世代）に対して、全く異なる技術を基盤とするシーケンサーが出現し“次世代”と呼ばれるようになった
- ✓ サンガー法と比べシーケンスリードが短いため、**ショートリードシーケンス**とも呼ばれている
- ✓ 超並列的に反応させることで大規模なデータ量を生み出すことが可能
- ✓ 様々な原理が採用されているが、基本的には1塩基ごとの伸長反応を利用し塩基配列を決定している

現行の主力機種とデータ量

	プラットフォーム	リード長	データ量
イルミナ	NovaSeq 6000 System	~2×250 bp	~6.0 Tb
	HiSeq 4000 System	~2×150 bp	~1.5 Tb
サーモフィッシャーサイエンティフィック	Ion GeneStudio S5 System	1×200 bp	~25 Gb
MGI	DNBSEQ-T7	~2×150 bp	~6.0 Tb
	DNBSEQ-G400	~2×200 bp (~1×400 bp)	~1.4 Tb

DNAポリメラーゼやリガーゼを用いて逐次DNA合成反応を行い、**蛍光・発光・電位変化**などを検出することで塩基配列を決定する

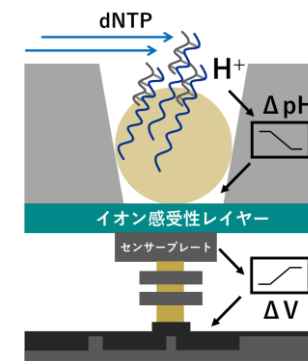
### SBS法 (Sequencing By Synthesis)



### 蛍光色を読み取る方法

異なる蛍光でラベルしたdNTPの取り込みを、蛍光の色を読み取ることによって、どの塩基が取り込まれたかわかる（全ゲノム解析で用いられる）

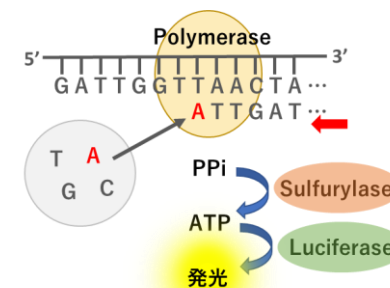
### イオン半導体 シーケンシング法



### 半導体チップを用いる方法

核酸の取り込み時に放出されるH+を半導体チップで検出する

### パイロシーケンス法



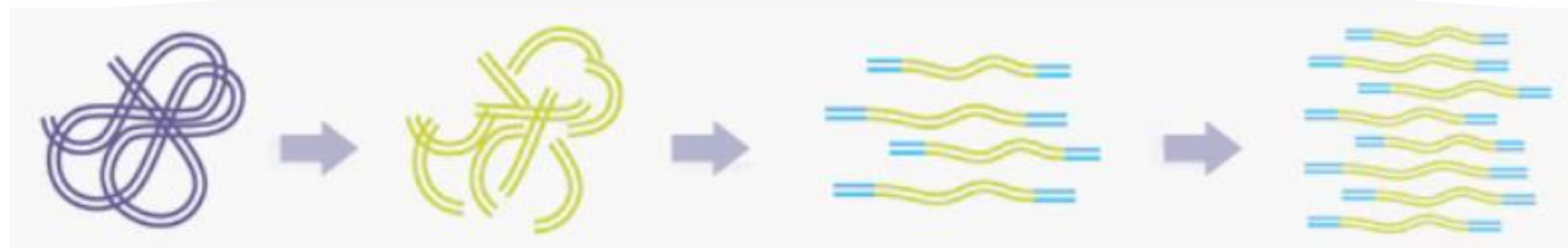
### 発光を利用する方法

核酸の取り込み時に放出されるピロリン酸を発光反応に用いることで検出する

（イルミナおよびサーモフィッシャーサイエンティフィック社の資料を参考に作成）

サンプルをシーケンサーにかける前には下準備が必要となる（DNAをそのまま機器に投入しても配列情報は得られない）  
→ DNAライブラリを調製する必要あり

下記では一般的なライブラリ調製方法を説明をする



ゲノム DNA  
(~1 µg)



超音波破碎装置や  
酵素を使って、数百  
bpの長さになるよう  
に断片化する



アダプターと呼ばれる  
短いDNAを両末端に  
付加させる  
→DNAライブラリ



DNAライブラリを  
PCR\*で増やす  
(オプション)

\*PCR: Polymerase Chain Reaction  
(ポリメラーゼ連鎖反応)

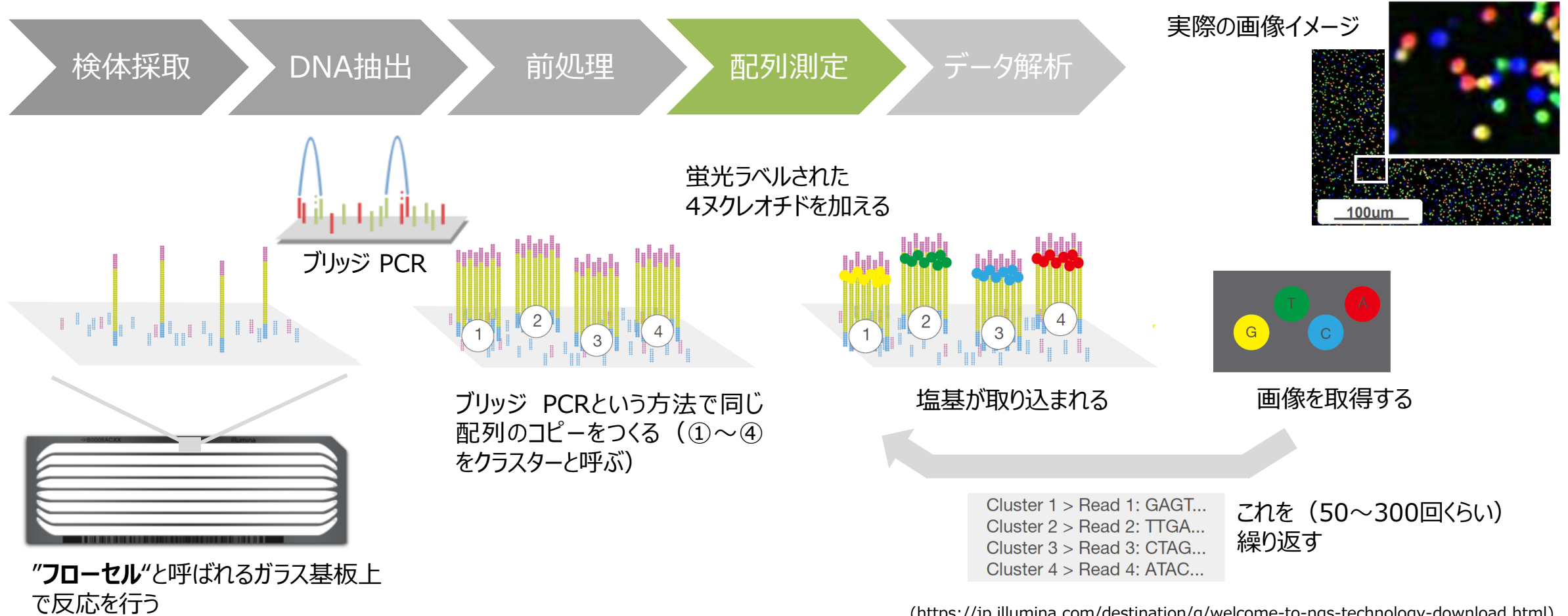
(サーモフィッシャーサイエンティフィック社ホームページ  
<https://www.thermofisher.com/jp/ja/home/brands/thermo-scientific/molecular-biology/thermo-scientific-specialized-molecular-biology-applications/sequencing-thermo-scientific/next-generation-sequencing-library-preparation-thermo-scientific.html>)

令和5年度がんの全ゲノム解析に関する人材育成推進事業



# NGSの原理 (SBS法を例に)

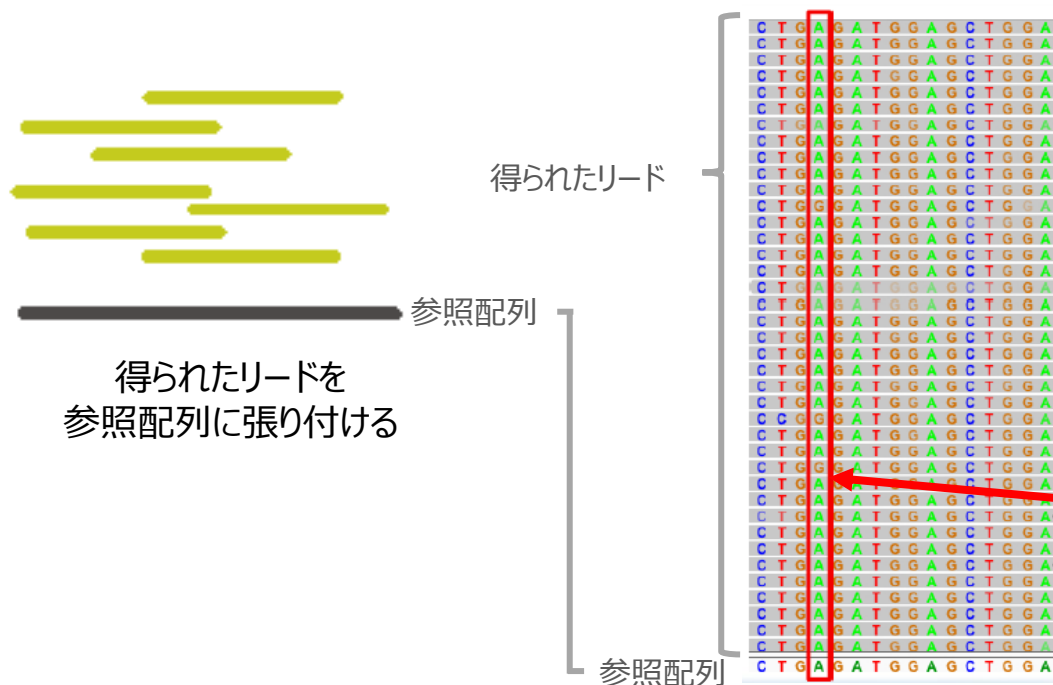
SBS (Sequencing By Synthesis) 法では、並列で1つずつ塩基を伸長し配列を決定する



(<https://jp.illumina.com/destination/g/welcome-to-ngs-technology-download.html>)

# NGS解析のアウトプット

NGSでは膨大な量の塩基配列データが一度に得られる。読み取ったリード配列を参照配列（リファレンス）にマッピングを行い、リファレンスと比較して（わずかな）配列の違いを明らかにする

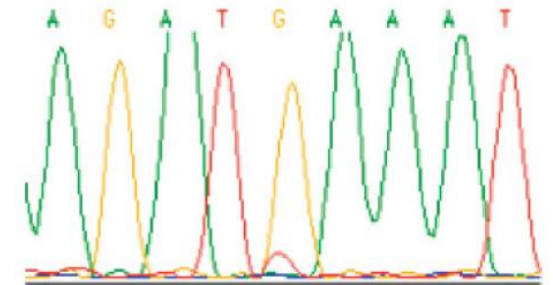


- 配列のエラー率が高いので、ゲノム上の同じ位置を繰り返して読む必要がある
- 統計的手法によりエラーと変異を区別する
- 繰り返し数をdepthという
- 低頻度の変異も検出できる

A>G (8%)

サンガーシーケンスでは

得られた波形から配列を決定する  
一般的に20%以下の低頻度変異は、波形が小さくなってしまい、検出することが難しい



G: 87.8%  
T: 12.2%

(J Hum Genet. 2015 May;60(5):227-231)

# NGSの原理2 (CPAS法)

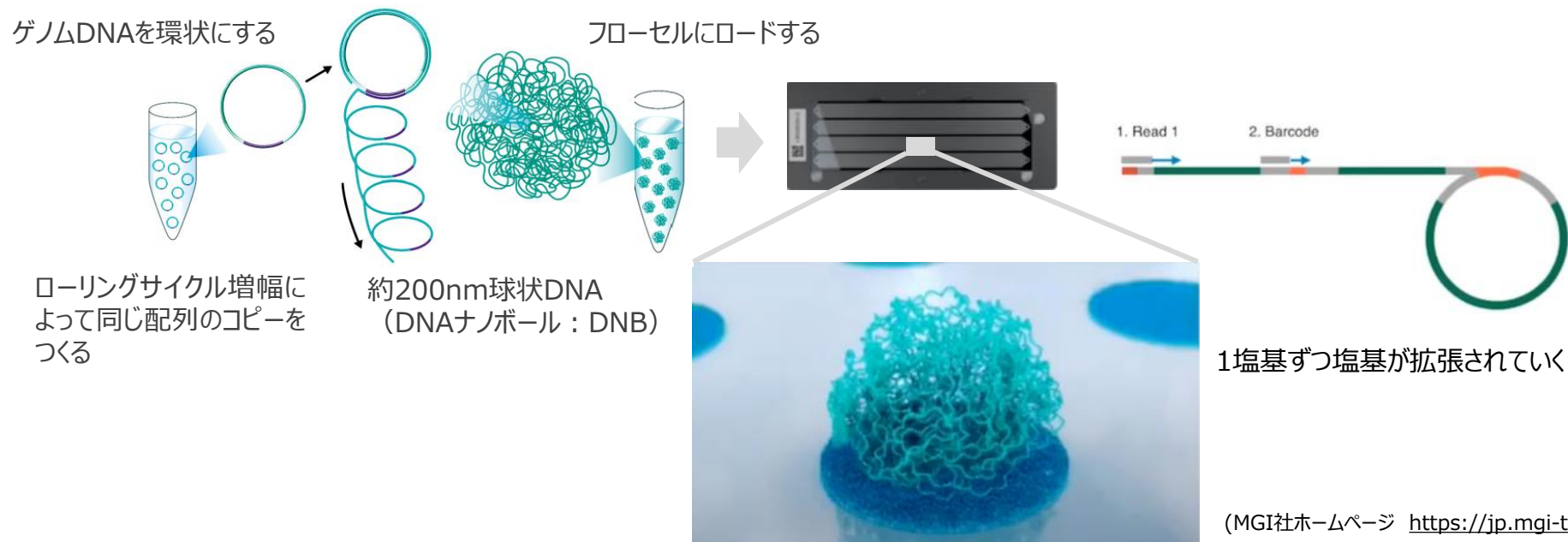
全ゲノム解析に適したもう一つのNGS技術に**CPAS法** (Combinational Probe Anchor Synthesis) がある

- CPAS法でのシーケンスは、「DNAナノボール (DNB)」上の蛍光標識されたプローブの画像をデジタル信号に変換することによって決定される。ローリングサイクル増幅によってDNBを作成することには、いくつかの利点がある
- 利点：
  - ✓ ローリングサイクル増幅は線形増幅のため、PCRのバイアスを低減させることができる
  - ✓ インデックスホッピング\*が抑制できる
  - ✓ フローセル上でPCR (ブリッジPCR) を行わないため、リアルPCRフリーのシーケンスが可能である



MGI Tech社  
DNBSEQ-T7

\***インデックスホッピング**：複数ライブラリをプーリングし、シーケンスをした際にライブラリ間でインデックスの組み換えが起こり、別のライブラリ由来のリードが含まれてしまう現象

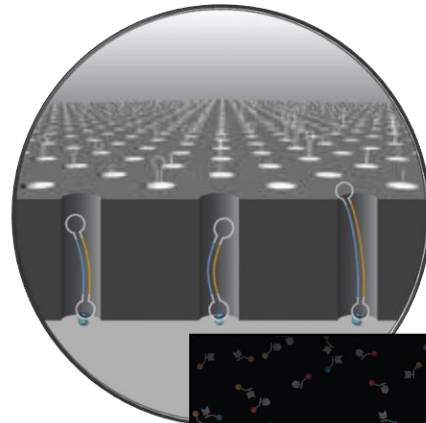


(MGI社ホームページ <https://jp.mgi-tech.com/products/>)

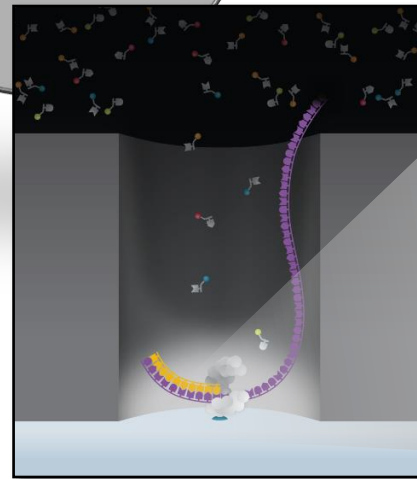
# NGSの次の新しい技術：SMRTテクノロジー

パシフィック バイオサイエンス社のSingle Molecule, Real-Time (SMRT) シークエンシングテクノロジーを紹介する

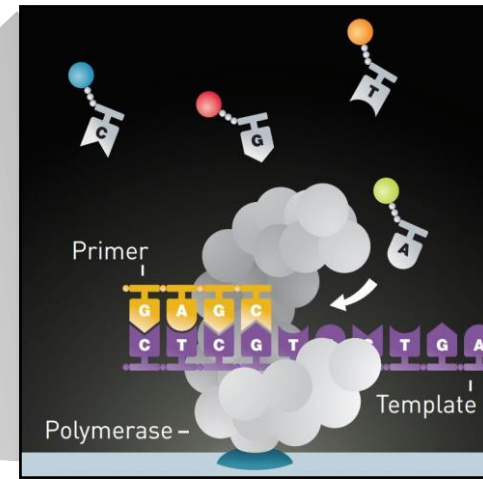
- **原理：**  
長く断片化したnative DNAをテンプレートにして、DNA合成を百万単位の小孔内で同時に行う
- **利点：**
  - ✓ 数十kbpの高精度なロングリードシーケンスが可能である
  - ✓ エピジェネティクス情報を同時に解析できる
  - ✓ GC含量に影響を受けにくく、均一なデータカバレッジが得られる



異なる色の蛍光分子で標識された各ヌクレオチド (A、T、G、C) が、ウェルに運ばれる



DNAポリメラーゼとテンプレートDNAの複合体は、微小ウェルの底部に固定される



固定されたポリメラーゼがシークエンシングエンジンとして働く

- ✓ 取り込まれたヌクレオチドが数十ミリ秒保持され、蛍光を発する
- ✓ リン酸基が切断され、付加された蛍光分子が離れる
- ✓ 次の相補する塩基が取り込まれ、同じ過程が繰り返される



Sequel II

(パシフィック バイオサイエンス社資料「SMRT テクノロジーとSequel IIe システム概要」から抜粋)

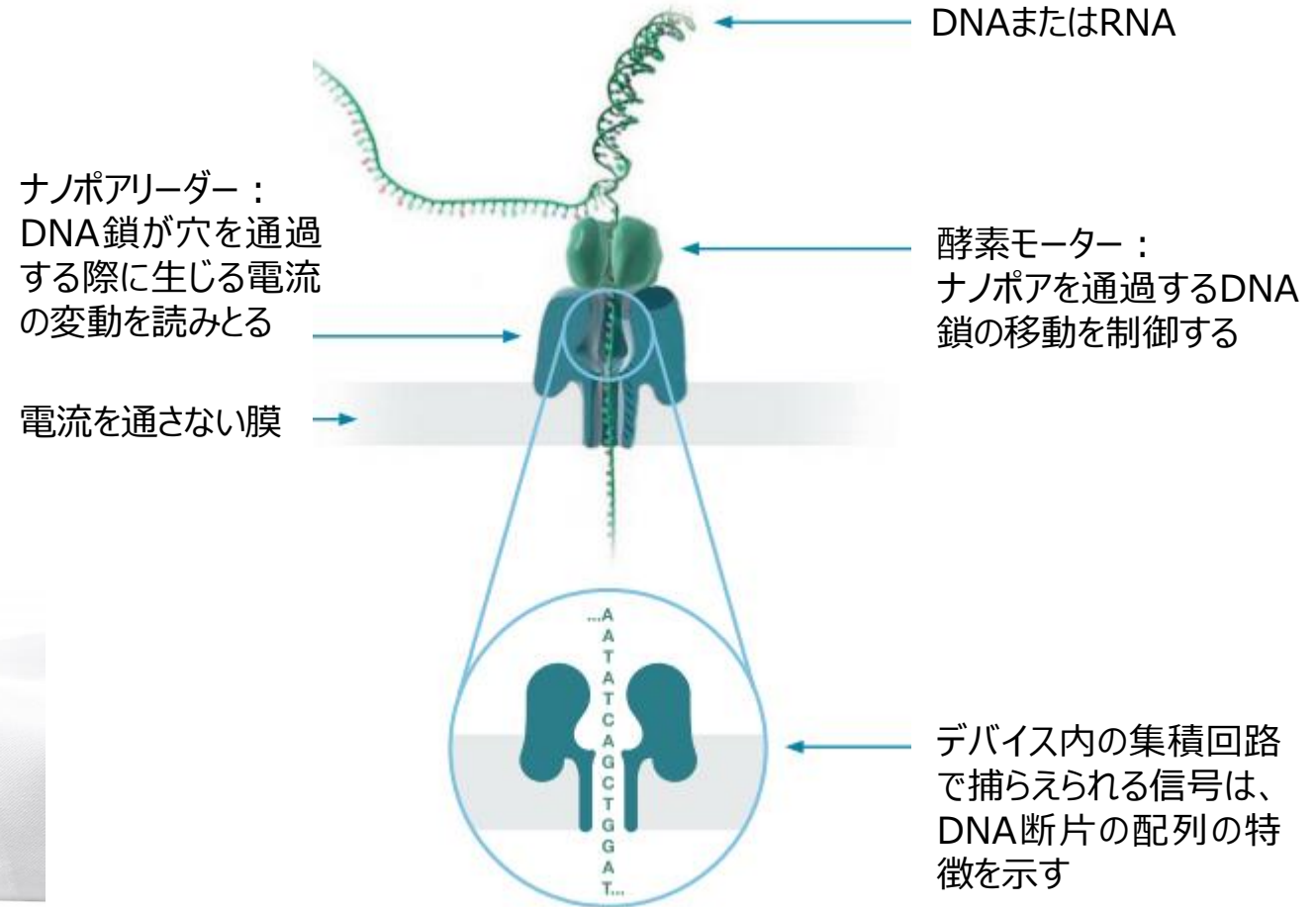
# NGSの次の新しい技術：ナノポア

Oxford Nanopore Technologies社のナノポアシーケンス技術を紹介する

- **原理：**  
DNAが膜タンパク質の微細孔（ナノポア）を通り抜けるときに生じるイオン電流の変化を読み取り塩基配列を決定する（参考となる動画サイト：  
<https://nanoporetech.com/jp/how-it-works>）
- **利点：**
  - ✓ 理論上、読める長さに制限はなく、400万塩基以上の長さのDNAも解析可能である
  - ✓ 塩基の配列だけでなく、メチル化などの化学修飾も検出できる
  - ✓ リアルタイムに解析できる
  - ✓ とても小さくどこでも解析できる（MinION）
- **課題：**エラー率が高い



MinIONポータブルシーケンサー



(Oxford Nanopore Technologies社 製品カタログから抜粋)

# 全ゲノム・エキソーム・パネル解析概要

全ゲノム解析 (WGS) は、遺伝子を含むゲノムの全領域を網羅的に解析する

- **WGS** : ゲノム全領域

- ✓ 約3Gb
- ✓ 平均カバレッジ (depth) は、
- ✓ コントロール : 30X~50X、腫瘍 : 80X~
- ✓ 決定塩基数 (throughput) は、300Gb~

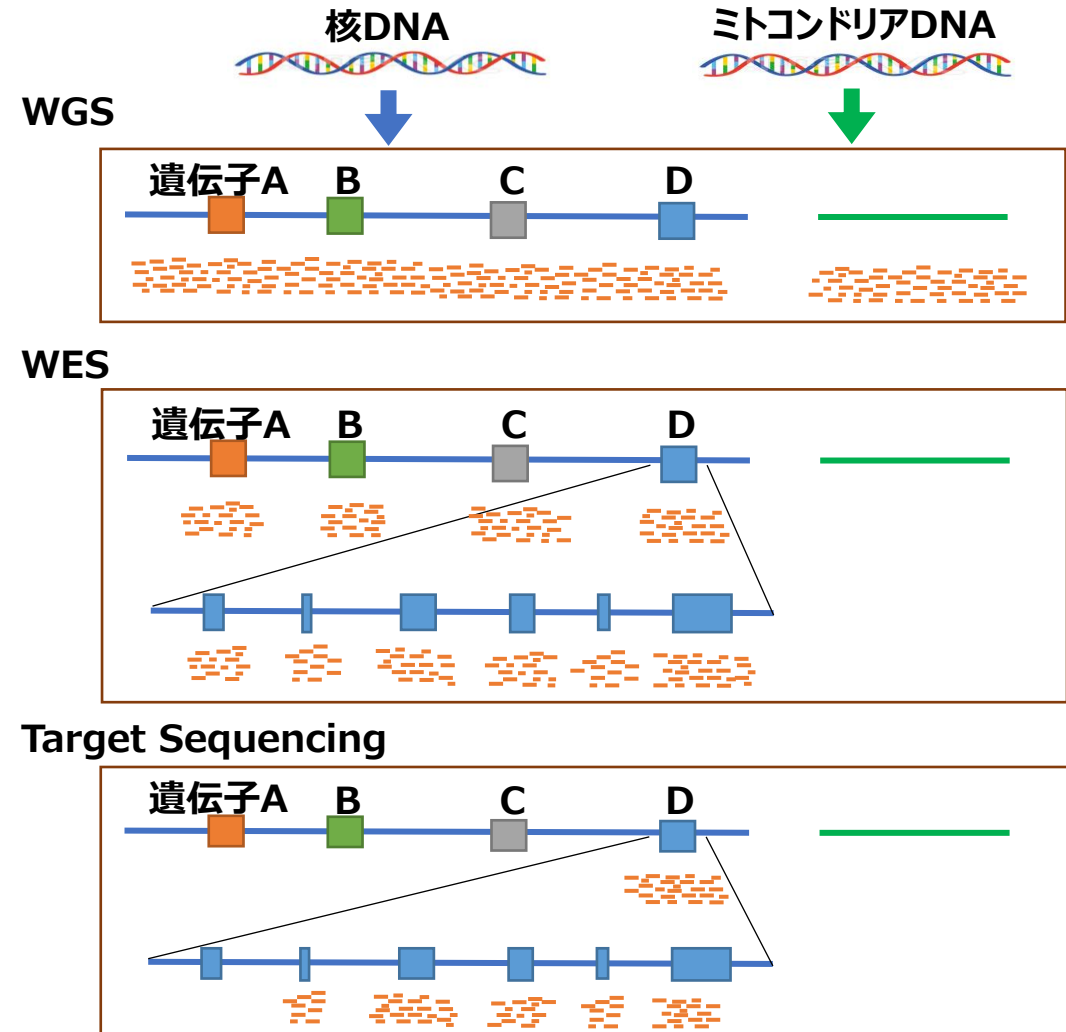
- **WES** : 全遺伝子の全エクソン領域

- ✓ 約50Mb
- ✓ 平均カバレッジ (depth) は、100X~
- ✓ 決定塩基数 (throughput) は、5~10 Gb

- **がんパネル解析 (Target Sequencing)** : 特定遺伝子の特定領域\*

- ✓ 数10kb~数 100 kb (数 10~数100 遺伝子)
- ✓ 平均カバレッジ (depth) は、400X~1,000X
- ✓ 決定塩基数 (throughput) は、0.1~1 Gb

\*プロモーター領域、ミトコンドリア遺伝子などを含むものもある。



## WGSとWESの比較をする

### ● WGSの長所 (WESと比較して)

- ✓ 遺伝子の非コーディング領域 (プロモーターや調節領域を含む) のバリエーション・変異も検出可能
- ✓ ゲノム構造のバリエーション・異常 (転座・逆位・遺伝子融合を含む) の検出が可能
- ✓ ウイルスゲノムなど、外来病原体のゲノムの存在や挿入部位も検出可能
- ✓ ミトコンドリアゲノムの変化も検出可能
- ✓ より正確な変異シグニチャーが得られる
- ✓ 多数の遺伝子多型の情報が得られる
- ✓ 将来の医療応用・開発に役立つ情報も得られる

### ● WGSの短所 (WESと比較して)

- ✓ データ量が大きく、アッセイコストが高い
- ✓ データ解析に必要とされる、高速なデータの転送、解析システムを要する
- ✓ 大容量データの保存場所が必要である
- ✓ 多くのバリエーション・変異の評価が必要

変異の位置・検出	WGS	WES	Target Sequencing
コード領域の変異	○	○	○ 限られた遺伝子のみ
スプライトサイト変異	○	○	○ 限られた遺伝子のみ
プロモーター領域変異	○	△	△ 限られた領域のみ
調節領域変異	○	△	△ 限られた領域のみ
コピー数変異	○	○	○ 限られた遺伝子のみ
構造変異	○	△	
ウイルスゲノム挿入	○	△	
ミトコンドリア変異	○	△	
変異シグニチャー	○	△	



- ✓ アッセイコストは減少傾向にある
- ✓ 様々なコンピューターシステムの開発が進んでいる
- ✓ クラウドなど、ストレージ方法の進歩が広がっている
- ✓ データベースの充実、人工知能の開発など研究の進歩が見込まれる

全ゲノム解析+RNAシーケンス解析で治療薬が発見できる可能性が高まる

100人の末期がん患者に対して：

● **全ゲノム解析 + RNAシーケンス解析**を実施

- ✓ 100人中78人が解析可能
- ✓ 解析可能な78人中55人にactionable mutationあり
- ✓ 55人中34人が解析結果により治療を受けた  
(25人 FDA approved; 8人Off-label; 1人Clinical trial)
- ✓ 治療を受けた患者の中で14人に何らかの効果が認められた

● **がんプロフィール解析 (AmpliSeq Cancer Panel)** を実施

- ✓ 100人中81人が解析可能
- ✓ 解析可能な81人中59人にmutationを同定 (22人に変異なし)
- ✓ Mutationの55%はTP53遺伝子変異 (治療薬なし)

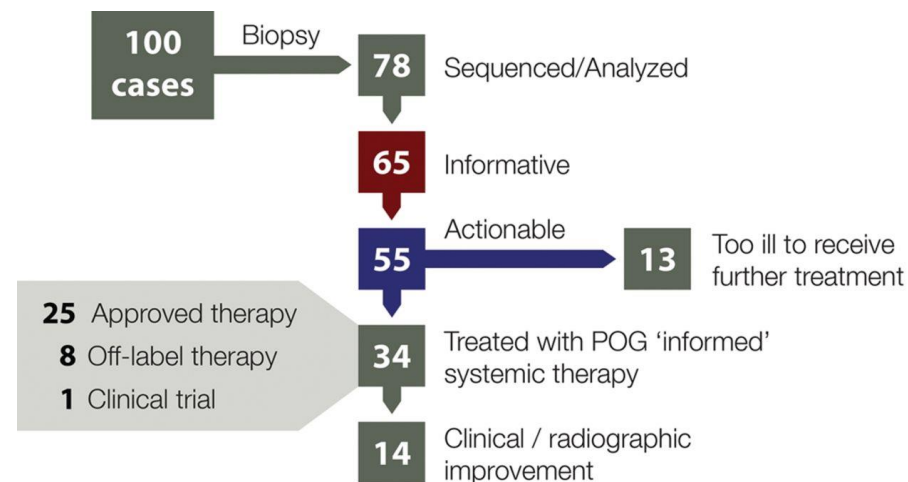
全ゲノム解析+RNAシーケンス解析 >>> プロファイル解析



PRECISION MEDICINE IN PRACTICE

**Lessons learned from the application of whole-genome analysis to the treatment of patients with advanced cancers**

Janessa Laskin,<sup>1</sup> Steven Jones,<sup>2</sup> Samuel Aparicio,<sup>3</sup> Stephen Chia,<sup>1</sup> Carolyn Ch'ng,<sup>2</sup> Rebecca Deyell,<sup>4</sup> Peter Eirew,<sup>3</sup> Alexandra Fok,<sup>2</sup> Karen Gelmon,<sup>1</sup> Cheryl Ho,<sup>1</sup> David Huntsman,<sup>3,5</sup> Martin Jones,<sup>2</sup> Katayoon Kasaiian,<sup>2</sup> Aly Karsan,<sup>1,2</sup> Sreeja Leelakumari,<sup>2</sup> Yvonne Li,<sup>2</sup> Howard Lim,<sup>1</sup> Yussanne Ma,<sup>2</sup> Colin Mar,<sup>6</sup> Monty Martin,<sup>6</sup> Richard Moore,<sup>2</sup> Andrew Mungall,<sup>2</sup> Karen Mungall,<sup>2</sup> Erin Pleasance,<sup>2</sup> S. Rod Rassekh,<sup>4</sup> Daniel Renouf,<sup>1</sup> Yaoqing Shen,<sup>2</sup> Jacqueline Schein,<sup>2</sup> Kasmitan Schrader,<sup>7</sup> Sophie Sun,<sup>1</sup> Anna Tinker,<sup>1</sup> Eric Zhao,<sup>2</sup> Stephen Yip,<sup>1</sup> and Marco A. Marra<sup>2,7</sup>



(Laskin J et al. Cold Spring Harb Mol Case Stud 2015;1:a000570)



# 全ゲノム解析による治療薬選択

全ゲノム解析により、ドライバー変異を網羅的に発見できる

- がん転移巣における体細胞変異の頻度 (2,520ペア検体/2,399患者)
  - ✓ Total somatic variants ~70,000,000 (平均 ~28,000/tumor)
  - ✓ SNV ~59,500,000 (平均 ~24,000)
  - ✓ MNV ~840,000 (平均 ~300)
  - ✓ Ins/Del (Indel) ~9,600,000 (平均 ~3,800)
  - ✓ SV ~600,000
- ドライバー変異とその分類
  - ✓ **Candidate driver mutations: 13,384 (平均~5.6/patient)**  
(Coding mutations: ~7,400; Non-coding SNV: 615 ;  
ホモ欠失: ~2,700 ; 遺伝子融合: 276 ; 遺伝子増幅: ~2,400)
  - ✓ 遺伝性腫瘍の原因遺伝子の生殖細胞系列変異 : 189 (~7.9%)
- Actionable mutation & their types
  - ✓ **Actionable event (Level A & B) が1,480 例 (62%) に見つかった**  
(Level A-on label: 18%; Level A-off label: 13%; Level B: 31%)
  - ✓ SNV (62%) , MNV (1.6%) , Indel (6.5%) , Amplification (25.5%) , Fusion (2.4%) , MSI (2.0%)

ドライバー変異のうち、遺伝子融合やInversion、発現調節領域の変化などはWESでは見つからないかもしれない

## Article

### Pan-cancer whole-genome analyses of metastatic solid tumours

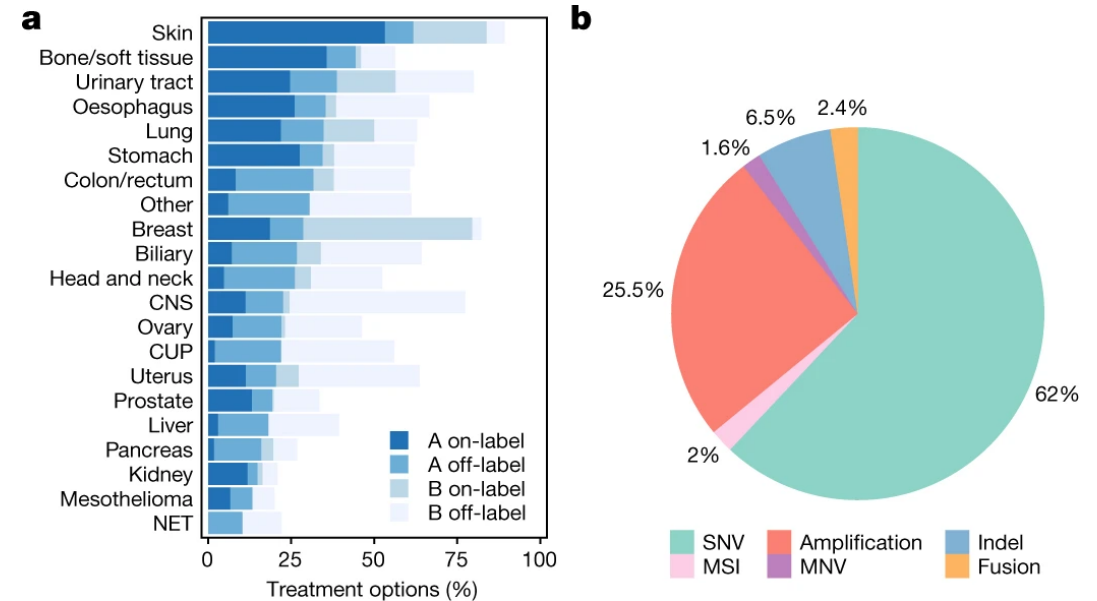
<https://doi.org/10.1038/s41586-019-1689-y>

Received: 9 September 2018

Accepted: 20 September 2019

Published online: 23 October 2019

Peter Priestley<sup>1,2,12</sup>, Jonathan Baber<sup>1,2,12</sup>, Martijn P. Lolkema<sup>3,4</sup>, Neeltje Steeghs<sup>3,5</sup>, Ewart de Bruijn<sup>1</sup>, Charles Shale<sup>2</sup>, Korneel Duyvesteyn<sup>1</sup>, Susan Haidari<sup>1,3</sup>, Arne van Hoeck<sup>6</sup>, Wendy Onstenk<sup>1,3,4</sup>, Paul Roepman<sup>1</sup>, Mircea Voda<sup>1</sup>, Haiko J. Bloemendal<sup>7,8</sup>, Vivianne C. G. Tjan-Heijnen<sup>9</sup>, Carla M. L. van Herpen<sup>9</sup>, Mariette Labots<sup>10</sup>, Petronella O. Witteveen<sup>11</sup>, Egbert F. Smit<sup>3,5</sup>, Stefan Sleijfer<sup>3,4</sup>, Emile E. Voest<sup>3,5</sup> & Edwin Cuppen<sup>1,3,6\*</sup>



(Level A: presence of biomarkers with either an approved therapy or guidelines; level B: biomarkers with strong biological evidence or clinical trials that indicate that they are actionable.)

がんゲノム解析時に得られる生殖細胞系列の情報は、患者およびその近親者の治療・予防に有用である

- 生殖細胞系列の病的バリエーションで治療法・治療薬を選択（2021年9月現在）
  - ✓ *BRCA1/BRCA2* 遺伝子の生殖細胞系列に病的バリエーションがある乳がん・卵巣がん・前立腺がん・膵がん患者に対して、PARP阻害剤（オラパリブ）が適応になる。卵巣癌ではニラパリブも適応
  - ✓ *RET* 遺伝子の生殖細胞系列に病的バリエーションがある甲状腺髄様がんの場合、甲状腺全摘+リンパ節郭清が推奨される
  - ✓ *APC* に病的バリエーションがある大腸がんでは大腸（結腸+直腸）全摘が基本であるが、ポリープの数や年齢によって切除範囲や再建方法を検討する
  - ✓ *MLH1, MSH2, MSH6, PMS2, EPCAM* 遺伝子に病的バリエーションがある腫瘍は、マイクロサテライト不安定性のことが多い。MSI検査が未実施の場合は実施を推奨する。MSI検査で陽性であれば、免疫チェックポイント阻害剤（ペムブロリズマブ）が適応
- 遺伝性腫瘍の診断により、サーベイランスやフォローアップが変わる
  - ✓ 遺伝性腫瘍の場合、原発巣に対するフォローアップだけでなく、それぞれの疾患に応じたサーベイランスを生涯行う必要がある（サーベイランスの詳細は各疾患のガイドラインを参照）
- リスク低減手術、化学予防などの予防法の選択肢がある
  - ✓ *BRCA1/BRCA2* 遺伝子の生殖細胞系列に病的バリエーションがある乳がん患者（HBOC患者）に対しては、リスク低減乳房切除術や、リスク低減卵管卵巣摘出術を検討する（HBOC診療の手引きを参照）
  - ✓ HBOC患者に対する卵巣がん化学予防のために、低用量経口避妊薬または低用量エストロゲン・プロゲステロン配合薬を投与することがある（HBOC診療の手引きを参照）
- （遺伝学的な）近親者のリスク評価、発症前診断に利用可能
  - ✓ 家族歴等を含めた総合的なリスク評価が必要であり、遺伝性腫瘍の専門医あるいは臨床遺伝専門医、認定遺伝カウンセラーなどとの連携で対応することが必要である

# 生殖細胞系列バリエーションの同定

遺伝性疾患、先天性疾患などの原因となる生殖細胞系列のバリエーションの同定では、WGSはWESよりも検出感度の点で優れている。その為、遺伝学的検査においてWESで原因を同定できない場合には、追加でWGSやRNA-seqなどを行う必要がある

## ● エクソン領域のバリエーション

- ✓ GCリッチな配列部分は、WESのCoverageが低い  
(GCリッチ領域はCaptureによるDNAの濃縮が困難)

## ● 非コード領域のバリエーション

- ✓ WESでは、Exonを中心にCaptureする為、非コード領域のバリエーションをほとんど同定できない (GWASなどで見つかった体質に関わる多くのSNP、薬剤代謝遺伝子の発現に関わるプロモーターのSNP情報は解析できない)

## ● CNV解析・SV解析

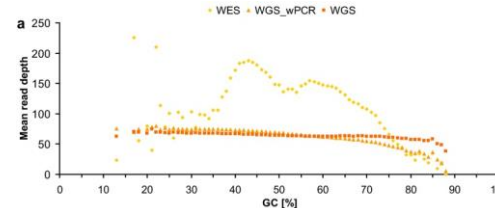
- ✓ WESでは、ExonをCapture用キットのバージョン等により、濃縮される領域にばらつきを生じる可能性あり。同一キットを用いたデータとの比較が必要な為、WGSよりも、CNVの感度・精度が劣る
- ✓ WESでは、構造変化のBreakpointを決定すること困難  
(欠失・重複・転座や逆位の確認方法に限界がある)

Hum Genet (2016) 135:359–362  
DOI 10.1007/s00439-015-1631-9

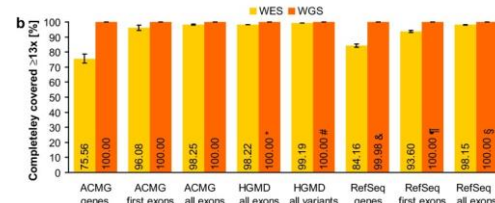
### SHORT REPORT

#### Clinical sequencing: is WGS the better WES?

Janine Meienberg<sup>1</sup> · Rémy Bruggmann<sup>2</sup> · Konrad Oexle<sup>1</sup> · Gabor Matyas<sup>1,3</sup>



図a : PCR-freeのWGSでは、WESに比べてGC含有量に関わらずcoverageが均等である



図b : 遺伝子のコーディング領域の解析でも、WGSはWESに比べ領域の網羅性がすぐれている

#### Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants

Aziz Belkadi<sup>a,b,1</sup>, Alexandre Bolze<sup>c,1,2</sup>, Yuval Itan<sup>c</sup>, Aurélie Cobat<sup>a,b</sup>, Quentin B. Vincent<sup>a,b</sup>, Alexander Antipenko<sup>c</sup>, Lei Shang<sup>c</sup>, Bertrand Boisson<sup>c</sup>, Jean-Laurent Casanova<sup>a,b,c,d,e,3,4</sup>, and Laurent Abel<sup>a,b,c,3,4</sup>

<sup>a</sup>Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM U1163, 75015 Paris, France; <sup>b</sup>Paris Descartes University, Imagine Institute, 75015 Paris, France; <sup>c</sup>St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY 10065; <sup>d</sup>Howard Hughes Medical Institute, New York, NY 10065; and <sup>e</sup>Pediatric Hematology-Immunology Unit, Necker Hospital for Sick Children, 75015 Paris, France

### METHODS

#### Comparison of Exome and Genome Sequencing Technologies for the Complete Capture of Protein-Coding Regions

Stefan H. Lelieveld,<sup>1</sup> Malte Spielmann,<sup>2,3</sup> Stefan Mundlos,<sup>2,3</sup> Joris A. Veltman,<sup>1,4</sup> and Christian Gilissen<sup>1\*</sup>

Human Mutation

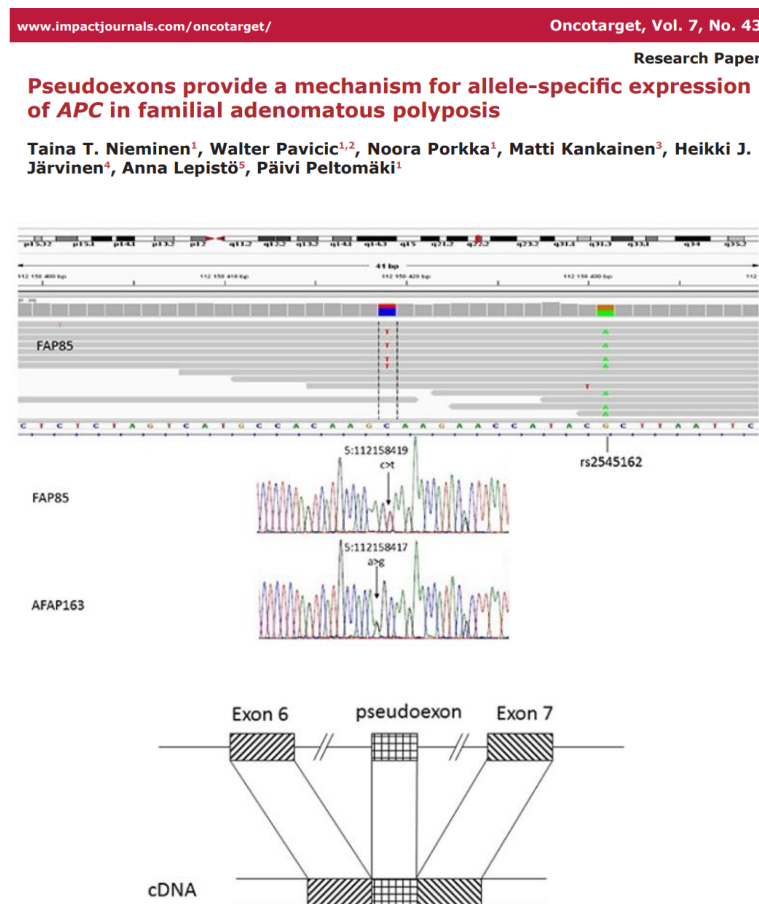


# 全ゲノム解析による病的バリエーションの同定

原因遺伝子がわかっているにもかかわらず、全ゲノム解析でしか見つからない病的バリエーションもある

- APC遺伝子のdeep intronic regionのバリエーションが新たなスプライスサイトを生み、イントロン内の一部が pseudoexonとして転写され、APCタンパク質に truncationを生じる
- APC遺伝子のプロモーター領域を含む欠失が、同アレル\*から転写される産物の減少を引き起こして病気の原因となる  
\*ヒトでは（常染色体上の）遺伝子は1対（2つ）存在する。アレルとは、その個々の遺伝子（対立遺伝子）を意味する
- APC遺伝子プロモーターの転写因子結合領域の病的バリエーションが、Gastric Adenocarcinoma and Proximal Polyposis of the Stomach (GAPPS) の原因となる

WGSは十分なカバレッジがあれば、WESよりも高い精度で CNVを検出できる。したがって、マイクロアレイやMultiplex Ligation-Dependent Probe Amplificationの代替検査として、エキソン単位の欠失・増幅などのバリエーション同定にも有用である



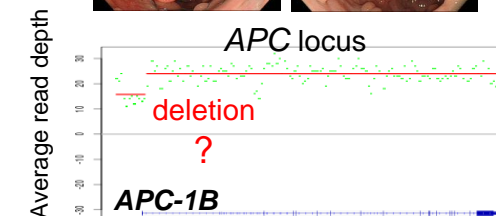
Deep intronic variants may cause pathogenic splicing

## SCIENTIFIC REPORTS

OPEN Reduced expression of APC-1B but not APC-1A by the deletion of promoter 1B is responsible for familial adenomatous polyposis

Received: 10 February 2016  
Accepted: 22 April 2016  
Published: 24 May 2016

Kiyoshi Yamaguchi<sup>1,2</sup>, Satoshi Nagayama<sup>2,3</sup>, Eligo Shimizu<sup>1,2</sup>, Mitsuhiro Komura<sup>2</sup>, Rui Yamaguchi<sup>1</sup>, Tetsuo Shibuya<sup>1</sup>, Masami Arai<sup>2</sup>, Seira Hatakeyama<sup>1</sup>, Tsuneo Ikenoue<sup>1</sup>, Masashi Ueno<sup>2</sup>, Satoru Miyano<sup>1,4,5</sup>, Seiya Imoto<sup>2</sup> & Yoichi Furukawa<sup>2</sup>



A deletion of the promoter and non-coding exon in APC is responsible for polyposis

バイオマーカーとは、「正常な生物学的過程、発病の過程、もしくは治療介入による薬理学反応を反映する、測定及び評価可能な特性」と定義されている。多くは遺伝子・遺伝子産物（DNA/RNA）やタンパク質、代謝産物などの生体内物質である。がん診療におけるバイオマーカーには、がんの診断・分類や治療薬の効果・副作用の予測、あるいは治療抵抗性の予測に用いられるもの、予後の予測に用いられるものなど、多様なバイオマーカーが存在する。

## ● バイオマーカーの分類

- ✓ 診断マーカー（早期診断・腫瘍分類）：AFP, CEA, 血中カルシトニン, *IDH*遺伝子変異, Philadelphia染色体など
  - ✓ 薬効予測マーカー（効果予測・抵抗性予測）：HER2発現, *EGFR*変異, *ALK*融合遺伝子, *KRAS*変異, *ABL*変異(T315I), *EGFR*(T790M)変異など
  - ✓ 毒性マーカー：KL-6, SP-Dなど（間質性肺炎）, ヒト脳性ナトリウム利尿ペプチド（BNP）, 同前駆体N端フラグメント（NT-proBNP）など（心毒性）
  - ✓ 代謝マーカー（用量決定）：*UGT1A1*遺伝子多型（イリノテカン代謝）, *TPMP*遺伝子多型（アザチオプリン代謝）, *DPD*（カペシダビン代謝）
  - ✓ 予後マーカー（反応、進展、再発）：TP53, オンコタイプDx（乳癌）, マンマプリント（乳癌）など
  - ✓ 感受性・リスクマーカー（発症予測）：HTLV-1抗体（ATL）, HBs抗原（肝がん）など
  - ✓ モニタリングマーカー：PSA, AFP, CEA, 血中カルシトニン, CTC（末梢循環癌細胞）, ctDNA(血中循環腫瘍DNA)など
  - ✓ 代替マーカー：Progression-free survival (PFS), Disease-free survival (DFS)など、臨床試験のエンドポイントの代わりに用いられるなど
- 特定の薬剤の効果・抵抗性を予測するバイオマーカーが、レスポンドを予測するためのコンパニオン診断薬として開発され、薬剤投与の決定の際に用いられる。
  - 適切な患者に、適切な薬剤を、適切な量投与するためにバイオマーカーが必要である。また、バイオマーカーは薬剤開発のfailureを減らし、費用対効果を上げるためにも有用である。多くのバイオマーカーが提唱されてきたが、それらのエビデンスレベルを検討して、バイオマーカーとしてのクオリティーを正しく評価する必要がある。

分子標的薬とは、疾患に関わる遺伝子または遺伝子産物を標的として特異的に作用する薬剤。2000年頃から分子標的薬の開発が盛んになった。抗がん剤の場合、従来の細胞障害性薬剤に比べ、腫瘍細胞特異性が高く、著効するケースが多い。しかし、薬価が高い。また、標的・特異的な副作用が認められる症例もある

## ● 分子標的薬の分類

- ✓ 低分子化合物
- ✓ 抗体薬
- ✓ ペプチド（ワクチンなど）
- ✓ 核酸医薬
- ✓ その他

## ● 標的分子の多様性

- ✓ 細胞表面分子
- ✓ 増殖因子・受容体
- ✓ キナーゼ
- ✓ プロテアソーム
- ✓ 血管新生関連分子
- ✓ 細胞周期関連分子
- ✓ アポトーシス関連分子
- ✓ DNA修復関連分子（PARPなど）
- ✓ 免疫チェックポイント分子
- ✓ その他

## 従来の（細胞障害性）抗がん剤と分子標的治療薬の比較

	細胞障害性抗がん剤	分子標的治療薬
種類	代謝拮抗薬、アルキル化剤など	低分子阻害剤、モノクローナル抗体など
作用機序	DNAやRNAの合成や修復に関与する酵素や、細胞分裂に関与する分子に直接作用し、殺細胞的に働く	特定の分子を標的として、癌細胞の増殖を特異的に抑制したり、進展過程を阻害したりする
標的	細胞内のDNAやRNA	癌細胞だけでなく、間質細胞や免疫細胞で発現する分子も標的となる
腫瘍選択性	低い（経験的）	高い
バイオマーカー	バイオマーカーがないものが多い	効果と関連するバイオマーカーがあるものが多い
副作用	重篤な副作用が多い	少ない（ただし、標的分子に特異的な副作用あり）
薬価	安価	高価

# 分子標的薬開発とコンパニオン診断

コンパニオン診断（CDx）検査とは、薬の安全性と効果の情報を提供するin vitro検査。厚生労働省は、2013年7月に「コンパニオン診断薬等及び関連する医薬品の承認申請に係る留意事項について」を通知し、新規治療薬の開発にコンパニオン診断薬の開発が必要であることを示した

## 分子標的薬の適応とCDx診断の例

- HER2を標的分子とする抗体薬トラスツズマブ（ハーセプチン）は、HER2を発現する乳がんに加え、HER2を発現する胃癌にも適応拡大された。
  - ✓ 免疫染色あるいはFISHでHER2の発現を調べて陽性の患者
  - ✓ FoundationOne CDx がんゲノムプロファイル検査でコピー数増幅のある患者
- イマチニブ（グリベック）は慢性骨髄性白血病の治療薬として開発された分子標的薬。Philadelphia染色体を持つ急性リンパ性白血病や、KIT（CD117）陽性消化管間質腫瘍、*FIP1L1-PDGFRα*陽性の好酸球増多症候群又は慢性好酸球性白血病にも有効
  - ✓ 免疫染色でKIT（CD117）が陽性の消化管間質腫瘍（GIST）患者
  - ✓ 染色体検査で、Philadelphia染色体陽性の急性リンパ性白血病患者
  - ✓ 遺伝子/染色体検査で*FIP1L1-PDGFRα*陽性の好酸球増多症候群又は慢性好酸球性白血病患者
- ペムブロリズマブ（抗PD-1抗体）は、腫瘍の発生臓器に関係なくマイクロサテライト不安定性（MSI）検査陽性、またはMMRタンパク免疫染色陰性の腫瘍に効果が高い（MSI検査も要）
- FoundationOne CDxがんゲノムプロファイル検査や、オンコマインDx Target TestマルチCDx検査などの、がん遺伝子プロファイル検査がCDxとして承認されつつある。

## 日本で承認されているCDx検査の例

分子標的薬	腫瘍	標的分子	コンパニオン診断
トラスツズマブ	乳癌, 胃癌	HER2	HER2発現、コピー数
イマチニブ	CML, ALL, GIST, HES/CEL	BCR-ABL, KIT, PDGFRα	Philadelphia染色体, KIT (CD117) 免疫染色, PDGFRα融合遺伝子検査
ゲフィチニブ	非小細胞肺癌	EGFR	EGFR遺伝子検査
セツキシマブ	大腸・直腸癌	EGFR	KRAS, NRAS遺伝子検査
オラパリブ	乳癌, 卵巣癌, 前立腺癌, 膵癌	PARP	BRCA1/2遺伝子検査, 相同組換え修復欠損
ベムラフェニブ	悪性黒色腫	BRAF	BRAF遺伝子検査
ダブラフェニブ, トラメチニブ	悪性黒色腫, 非小細胞肺癌	BRAF, MEK	BRAF遺伝子検査
クリゾチニブ	非小細胞肺癌	ALK・ROS1	ALK, ROS1遺伝子検査
エヌトレクチニブ	固形癌, 非小細胞肺癌	NTRK1/2/3	NTRK1/2/3融合遺伝子, ROS1 遺伝子検査
ペムブロリズマブ	固形癌, 非小細胞肺癌, 乳癌, 食道癌	PD-1	MSI検査, MMRタンパク免疫染色, PD-L1免疫染色
ニラパリブ	卵巣癌	PARP	相同組換え修復欠損

PMDA コンパニオン診断薬等の情報(令和3年9月28日版より抜粋)

# 第Ⅱ章

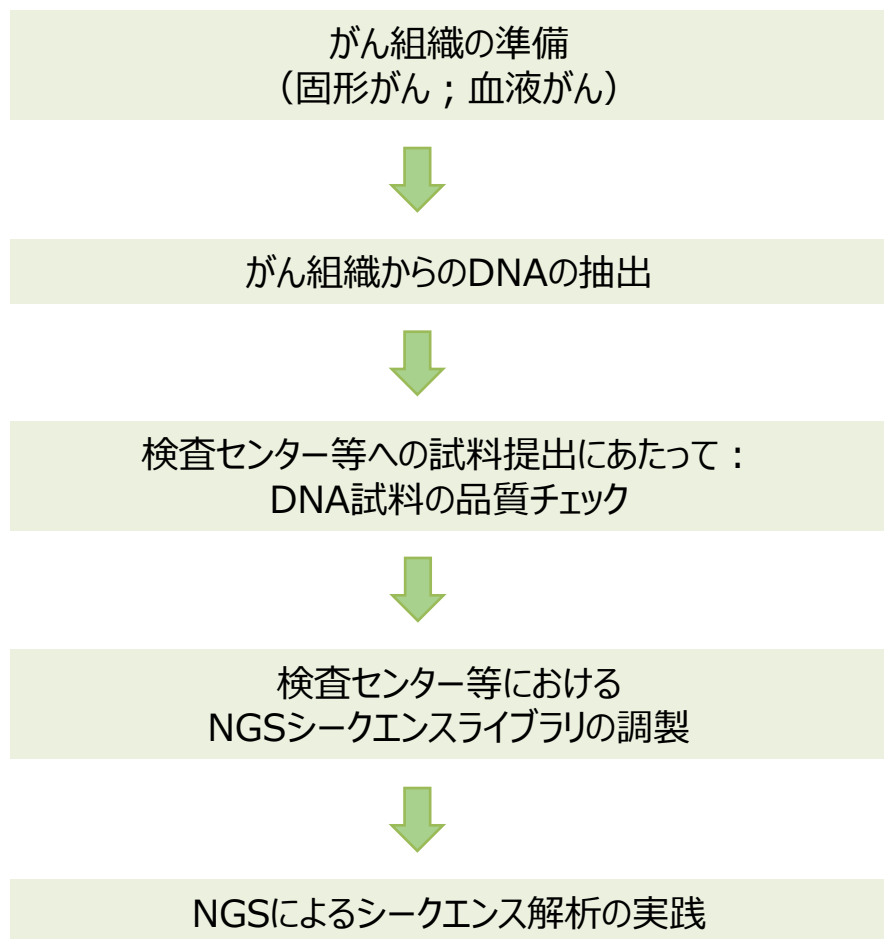
# サンプル調製

東京大学大学院新領域創成科学研究科  
メディカル情報生命専攻/生命データサイエンスセンター 教授  
鈴木穰

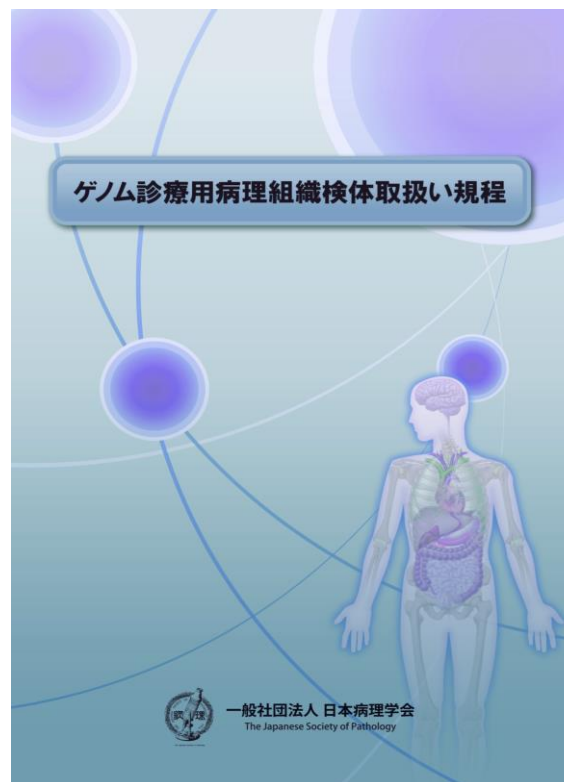


1. 概要：NGSによる鋳型調製～ゲノム解析の流れ
2. 生体サンプルの取得からNGSにかける前までの処理について
  - ✓ 固形がんにおける試料調製
    - 新鮮凍結検体とFFPE検体
    - 組織からのDNAの抽出
  - ✓ 血液がんにおける試料調製
3. シークエンス鋳型（シークエンスライブラリ）の調製
4. シークエンス解析と結果例

がんのNGS解析は以下の工程でがん組織を採取、ゲノムDNAを抽出して行う



### ゲノム診療用病理組織検体取扱い規程

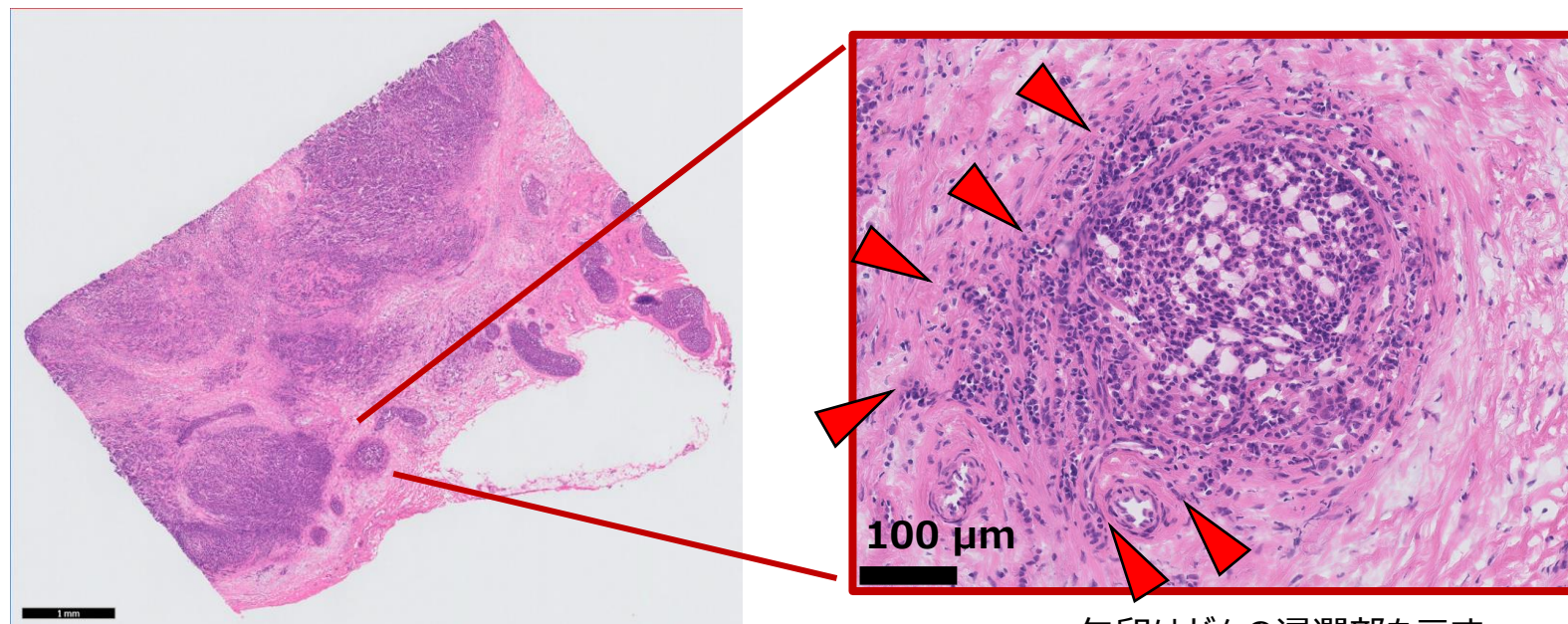


各関連学会が、NGS解析に用いる DNA抽出用に詳細にマニュアルをまとめている

(参考文献：ゲノム診療用病理組織検体取扱い規程 - 日本病理学会  
[https://pathology.or.jp/genome\\_med/pdf/textbook.pdf](https://pathology.or.jp/genome_med/pdf/textbook.pdf)  
参考文献：日本病理学会「ゲノム研究用病理組織検体取扱い規程」  
<https://genome.pathology.or.jp/houhou/index.html>  
参考文献：日本血液学会「造血器腫瘍ゲノム検査ガイドライン」  
<http://www.jshem.or.jp/genomgl/>)

NGS解析に用いるゲノムDNAの採取に際しては、病理診断に用いる生検試料、手術検体試料の一部を用いることが多い

乳がん (DCIS) 浸潤部の拡大図



右にがん組織切片のイメージ図を示す。  
図は乳がん (Ductal Carcinoma In Site;  
DCIS) の新鮮凍結検体である

矢印はがんの浸潤部を示す

(Nagasawa et al Com Biol 2021)

# 新鮮凍結（Fresh Frozen/FF）検体

病理診断用標本にはFFPE検体(ホルマリン固定)が用いられることが多い。そのため多くの場合、新鮮凍結検体の作製にあたっては、独立に試料を作成する必要がある

- 新鮮凍結検体とは以下のように作成された検体を指す
- 凍結に際しては、一般には、断熱容器等にいった液体窒素にチューブを浸して急速凍結する
- 組織の劣化が懸念されるために、凍結後、試料は液体窒素（-196℃）、少なくとも-80℃の超低温フリーザで保存する必要がある。以降、温度管理には細心の注意を払う
- 検体採取後に化学的反応を施していないために、それ以降のゲノムDNAの抽出の過程は比較的容易である
- 新鮮凍結検体から病理切片を作成して、がん細胞部分の分取等を行う必要がある場合、FFPE検体とは異なる注意が必要である。例えば、一般に、新鮮凍結検体では、病理切片の解像度、および組織像の保存度はFFPE検体に劣る
- ゲノムDNAに限らず、DNAよりもさらに劣化しやすい生体高分子（RNA、タンパク質）等の解析にも適している
- 細菌、ウィルス等が死滅してない可能性があるために全ての新鮮凍結検体はバイオハザード対象検体として扱う



新鮮凍結した検体  
(液体窒素タンクに保存)

近年のゲノム解析技術の進展により、FFPE検体からも精度の高いNGS解析を行うことが可能となった。ただし、固定・包埋の手順によっては後のDNA解析に悪影響を及ぼす可能性があるために以下の点には十分に注意する

### 影響を及ぼす可能性のある要因

- 試料作成までの時間：組織の劣化にともないDNAの品質も低下する可能性がある。ただしその影響は新鮮凍結検体に比べれば限定的である。
  - ✓ 血流停止から摘出までの時間
  - ✓ 摘出から固定までの時間
- 固定条件：適切な固定条件の設定が以下のNGS解析に最も大きな影響を及ぼす
  - ✓ ホルマリン固定液の組成、濃度、pH等
  - ✓ 固定の時間と温度
- 固定後：固定によりゲノムDNAは基本的に安定な状態となる。ただし適切な包埋条件の設定、その後の標本の保存条件は依然として重要である
  - ✓ 脱水・透徹条件の設定
  - ✓ 適切なパラフィンの選択と包埋条件の設定
  - ✓ 適切な標本の保存。フリーザーでの保存は必須ではないが、湿度、温度のコントロールされた環境に保つことが望ましい



ホルマリンで固定した検体  
(正常部とがん部)

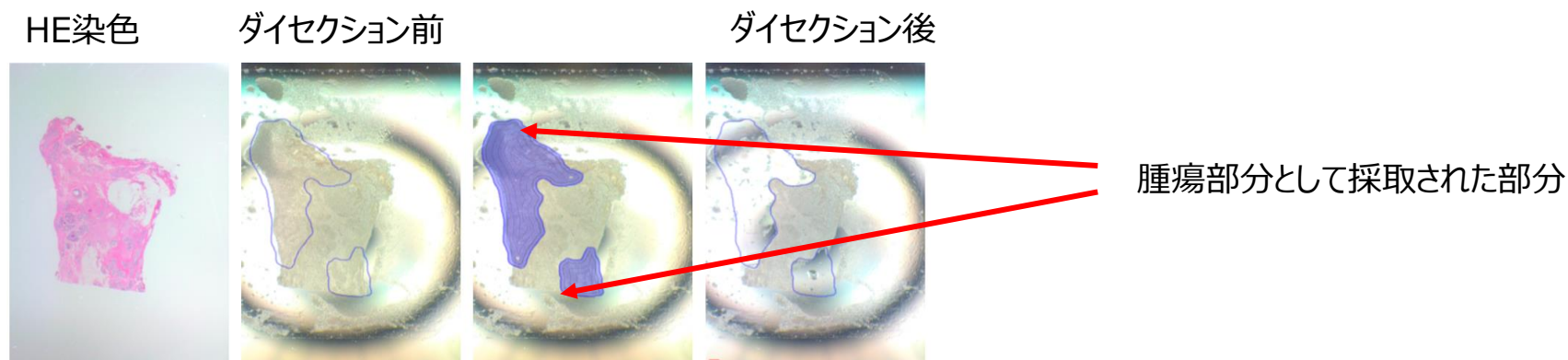
# 新鮮凍結検体とFFPE検体に共通した注意点

一般的な注意点は次ページに記す。また、どちらの場合でも特に検体中のがん細胞含有率には十分に注意する

- **がん細胞含有率**：提出する試料におけるがん細胞の含有率には細心の注意を払うべきである  
正常細胞からもゲノムDNAが抽出されるために、多くの正常細胞が混入していた場合、NGSデータにおけるがん細胞由来のゲノムDNAの割合が相対的に薄まることで、解析精度の低下を招く可能性がある

※米国がんゲノム解析プロジェクト等ではがん細胞含有率が30%以上であることが推奨されている

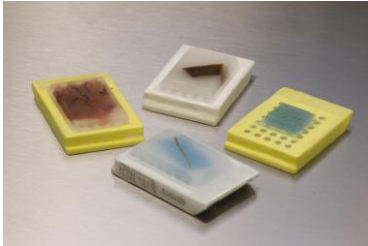
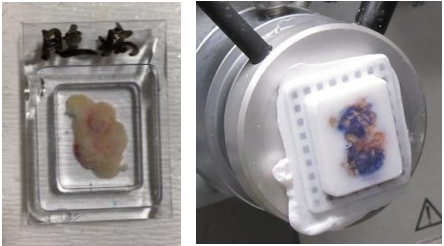
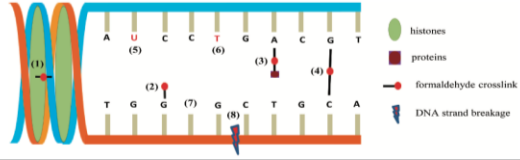
- がん細胞に対して間質細胞の割合が高いがん種/症例については、**レーザーマイクロダイセクション法（LMD）**等を用いた細胞の分取が必要となる場合もある（すい臓がん、血液循環がん細胞等では特にそのがん細胞由来のゲノムDNAの割合が低い）
- がん細胞の間でも高度にゲノム変異が進行している可能性がある症例、あるいは発がん起源が複数あると思われる症例についても同様の注意が必要である
- 近年になって、正常細胞（少なくとも非がん細胞）であっても、体細胞性突然変異が生じている例が報告されている（血液中のクローン性造血あるいはがん細胞周辺の間質細胞等）



FFPE検体からは比較的容易に腫瘍部分を分取して、がん細胞含有率を上げることが可能である

# がん組織検体の準備のまとめ

以下にFFPE検体、新鮮凍結検体を用いる場合のそれぞれの注意点および共通した注意点をまとめる

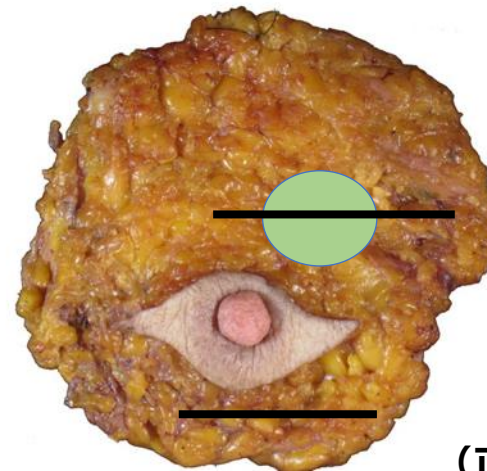
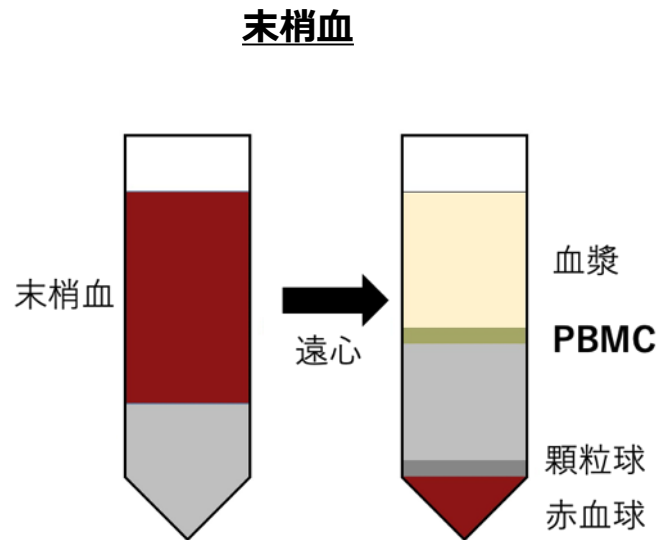
	FFPE検体	新鮮凍結検体
外観		
共通する注意点	<ul style="list-style-type: none"> <li>✓ 阻血時間の短縮・記録</li> <li>✓ 固定までに時間がかかる場合は、低温保存</li> <li>✓ 全ての検体はバイオハザード対象検体として扱う</li> <li>✓ がん細胞含有率に注意</li> </ul>	
各固定方法の注意点	<p><b>DNAの劣化がおこることを知っておくことは大切である</b></p> <ul style="list-style-type: none"> <li>✓ DNAと高分子間の共有結合を介した架橋形成</li> <li>✓ 架橋形成、高温での保存、紫外線によるゲノムの断片化</li> </ul> <p><b>核酸塩基の化学修飾がおこることも知っておく</b></p> <ul style="list-style-type: none"> <li>✓ 加水分解に伴う脱アミノ化でシトシンがウラシルに置換される</li> <li>✓ その後の PCR 増幅反応によってチミンができて最終的にシトシン&gt;チミン置換になる</li> </ul> <p>→ シークエンスデータの解釈に影響することがある（変異の偽陽性検出等）</p> 	<p><b>乾燥を防止する</b></p> <ul style="list-style-type: none"> <li>✓ 保存は密閉容器（ジップロック、蓋付き瓶など）が望ましい</li> </ul> <p><b>温度管理は徹底する</b></p> <ul style="list-style-type: none"> <li>✓ ディープフリーザーの扉開閉時間、故障に備える</li> </ul> <p>→ これらの保存環境の不備はDNAの劣化に直結する</p>

# 正常対照検体の準備

がん組織と同様の手順で、正常細胞ゲノムDNAを採取する組織を準備する

- 固形腫瘍における正常対照のゲノムDNAの採取は、おもに末梢血から密度勾配遠心法を用いて分離された単球やリンパ球を含む単核球（PBMC：Peripheral Blood Mononuclear Cells）より採取される（下図左）
- あるいは、手術検体内に腫瘍部とは別部位に非腫瘍部が存在する場合、正常対照のゲノムDNAとして非腫瘍部から採取も可能である（下図右）
- 組織切断面（割の方向）・採取の手法は施設内でよく相談の上実施することが望ましい

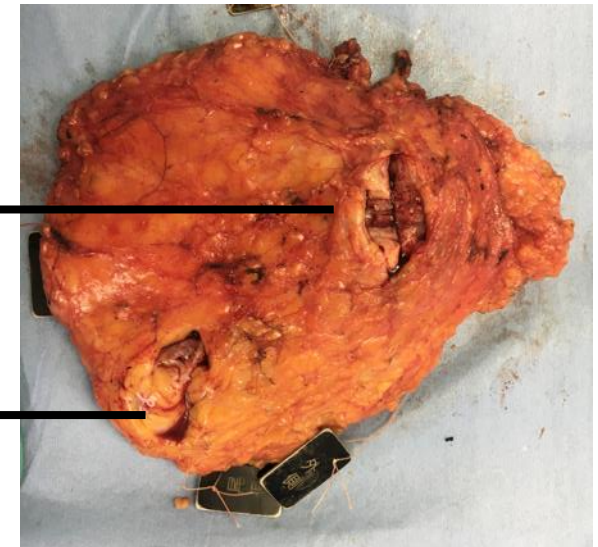
## 正常細胞の準備



## 近接非がん部位

腫瘍部

非腫瘍部  
(正常乳腺部)





# がん組織からのDNAの抽出（1）

## がん組織からDNAを抽出する

- リキッドバイオプシー検体をはじめ、がん検体をそのまま検査センターに提出できる場合もあるが、自施設でDNAを抽出してから送付する場合も多い。DNAの抽出法およびその後の品質確認の方法はFFPE検体と新鮮凍結検体で異なる
- 新鮮凍結検体からのDNA抽出は比較的容易である。多くのキットが市販されているが、そのいずれを用いても高品質のDNAが安定的に抽出可能である
- FFPE検体からゲノムDNAを抽出するには、FFPE専用のDNA抽出キットを用いる。いくつかのキットが市販されている（下図）。これらのキットでは、DNAからホルマリンの架橋を外す脱クロスリンク過程、パラフィンを除去する過程が含まれている。これらが不完全だった場合、NGS解析に大きな影響を及ぼすことがある

FFPE検体からのゲノム抽出に用いる代表的なキットを図示する

ホルマリン固定に際して導入された塩基置換を修復する酵素反応を追加するものもある



### タカラバイオ ; NucleoSpin® DNA FFPE XS

([https://catalog.takara-bio.co.jp/product/basic\\_info.php?unitid=U100006792](https://catalog.takara-bio.co.jp/product/basic_info.php?unitid=U100006792))

### QIAamp DNA FFPE Tissue Kit

(<https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/genomic-dna/qiaamp-dna-ffpe-tissue-kit/>)

※※※新鮮凍結用に最も頻繁に用いられるキット“Dneasy”のFFPE版である

# がん組織からのDNAの抽出 (2)

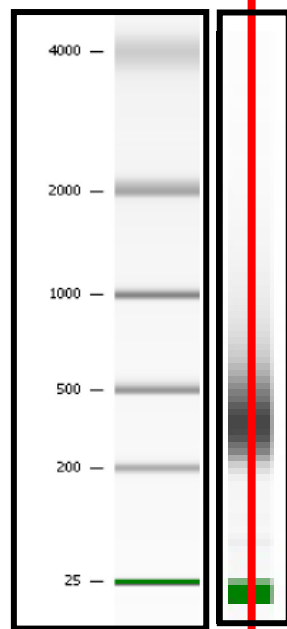
FFPE検体、新鮮凍結検体のどちらからの抽出であってもDNAの長さや量の定量が必要となる場合がある

- DNAの長さや量の定量には、一般的には、従来の吸光度分析よりも、アジレント社のBioanalyzerやTape Station等の電気泳動装置が用いられる
- 検査センターに提出する際に、その計測結果ファイルの添付を要求されることも多い

## アジレントBioanalyzer (オンチップゲル電気泳動装置)



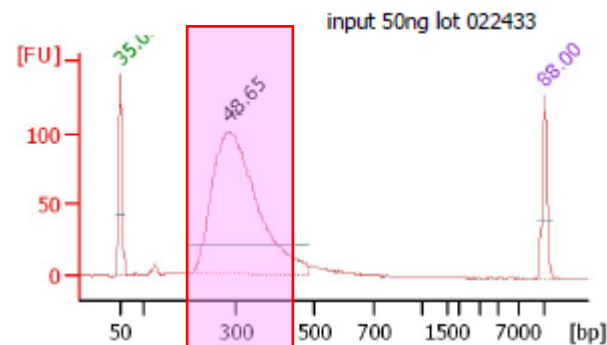
(詳細は下記のHPを参照  
<https://www.chem-agilent.com/contents.php?id=51>)



従来のゲル電気泳動を模したイメージ

定量的：“バンドの濃さ”を蛍光で定量  
高感度：1ng程度のDNAも測定可能

範囲を指定してその長さのDNA量を算出



Overall Results for sample 2 : input 50ng lot 022433

Number of peaks found: 1

Peak table for sample 2 : input 50ng lot 022433

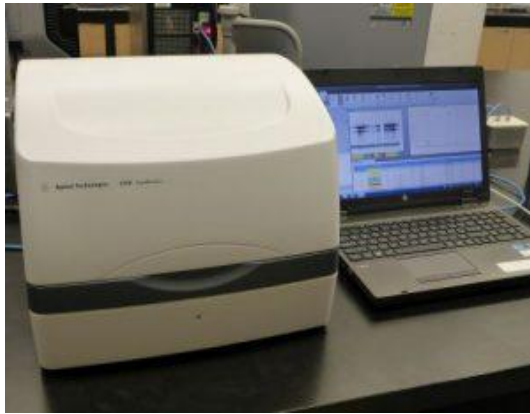
Peak	Size [bp]	Conc. [ng/μl]	Molarity [nmol/l]	Observations
2	285	48.18	256.0	

断面のイメージ

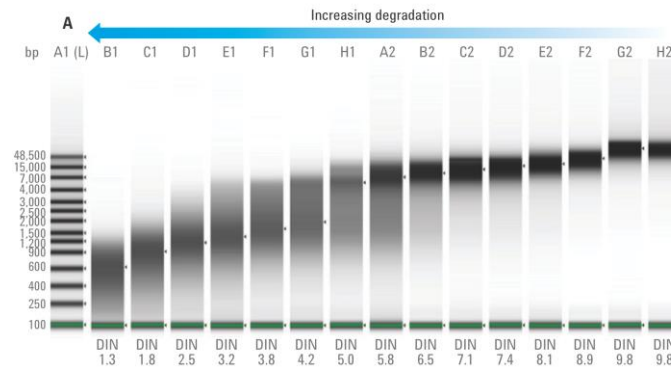
# がん組織からのDNAの抽出（3）

抽出されたDNAの品質管理は細心の注意を払って行う。DNAの品質はNGSデータの品質に大きく影響する

- アジレント社のBioanalyzer/Tape Station等の電気泳動結果の例を以下に示す



アジレント社Tape Station  
(並列型オンチップゲル電気泳動装置)



- Tape Stationでは、チップ内に内蔵されたマイクロ流路で従来のゲル電気泳動が行われる。流路の各位置に泳動されたDNAが蛍光色素で検出される。微小流路と好感度蛍光検出を用いることで微量検体の定量的解析が可能である
- 特にFFPE検体からのDNAの品質確認には、異なる長さの2か所の領域を増幅し（例：50～100 bp程度の増幅領域と100～300bp程度の増幅領域等）、その差をリアルタイムPCRによりCt値の差（ $\Delta$ Ct値）として検出することも多い。ホルマリンによる断片化、あるいはDNAの架橋により、PCR反応（NGS解析のモデルとなる）が阻害される程度を検証するものである

# ゲノム断片化～FFPE工程によるDIN値の低下

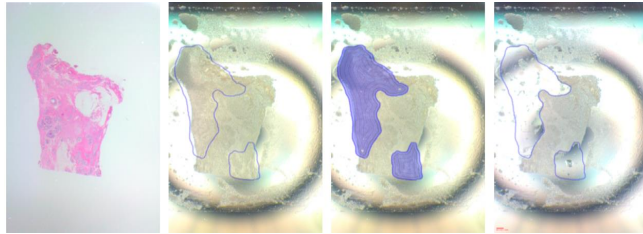
FFPEの工程によりDNAの断片化と質的劣化が進むことに注意する

- **DIN (DNA Integrity Number) 値**はアジレント社のAgilent 2200 TapeStationシステムを用いたGenomic DNA ScreenTape assayで測定したデータをgDNAの分解度に応じて1～10（値が高いほど高品質）にスコア化した値であり、FFPE検体のDNA品質評価が可能である。一般的な次世代シーケンズ解析にはDNAの断片化を行うので、直接DIN値が問題になることはない。ただしこの値が著しく低くなっている場合、組織の採取・保存状態がよくないことを示す、として注意が必要である

Qubit = 吸光度（実際には蛍光）で濃度測定する機器

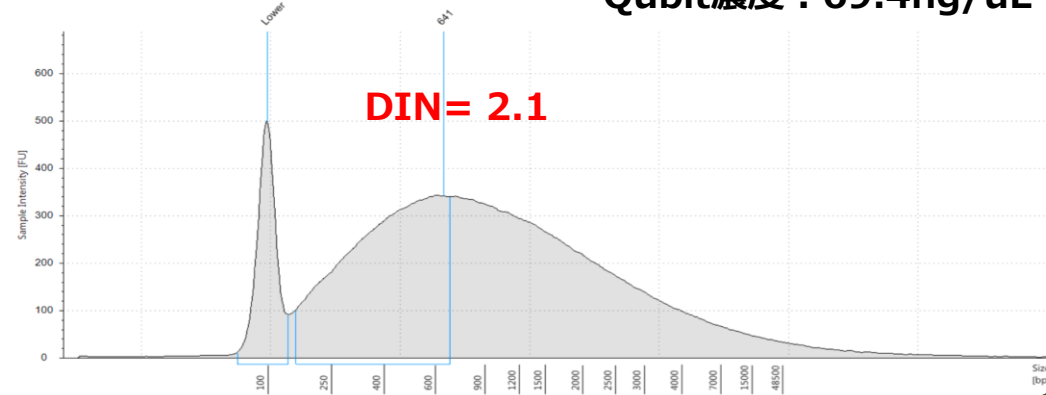
QubitではDNAの長さにはかれないことに注意

## FFPE検体（2011年アーカイブ検体）

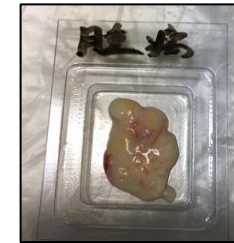


G1: D9

Qubit濃度 : 69.4ng/uL



## 新鮮凍結検体（手術からゲノム抽出まで約30分）



E1: Case10DCIS

Qubit濃度 : 9.62ng/uL



最終的にNGS解析に必要なDNA量は目的によって異なる

- 1~10  $\mu$ g

**全ゲノム解析（PCRフリー）：**

PCRでの増幅を行わないことで、最も定量性等の精度に優れた解析が可能である。がん種によるが100万細胞程度（概ね0.5- 1cm<sup>3</sup>の立方程度；ただし検体の種類と場所による）の検体が必要となる

- 100 ng~1  $\mu$ g

**全ゲノム解析（PCR法）、エキソーム解析、パネル検査：**

一般的に行われる手法。PCRでの増幅を行うために定量性、データの均一性に劣るが、比較的小さながん組織からも解析が可能である。組織片の大きさによるが、生検検体の解析にはこの手法が用いられることが多い

- 1 ng~100 ng

**微小がんゲノム解析、がん局所ゲノム解析、血液循環がん細胞/がんゲノム解析（リキッドバイオプシー）：**

初期がん等の微小病変あるいは局所に限定してがんゲノムを解析する場合。微量のゲノムDNAに対して、集中的なPCRをかけて増幅する。血液中に循環がん細胞/がんゲノム解析についても用いられる

### 血液がんについては固形がんと異なる注意点がいくつかある

- 日本血液学会では造血器腫瘍ゲノム検査ガイドラインを作成している。ここでは疾患によりゲノム検査の重要性とその意義がグレードに分けて明示されている
- 血液がんでは、固形がんとはゲノムの検査/解析を行う目的が異なる
- 固形がんにおいては遺伝子変異の検出は主として分子標的薬の選択に活用されるのに対して、血液がんではこれに加えて、病型の診断自体に有用と期待されている
- 血液がんでは、遺伝子変異の解析結果が、造血幹細胞移植術の是非またそのタイミングを決定するという意味での予後予測にも活用できることが期待されている
- 血液がんでは高頻度にみられる遺伝子変異の種類が固形がんとは異なることから、着目すべき遺伝子の選択も固形がんとは異なる (※)

(※) 実際、遺伝子パネル検査に適切な遺伝子の選別とその活用方法について、現在、急ピッチで議論されている。詳細は第III章を参照

日本血液学会 造血器腫瘍ゲノム検査ガイドライン 2021年版

ホーム このガイドラインについて 疾患・病期別パネル検査推奨度 遺伝子・疾患別エビデンスレベル

検索

フリーワード検索 (全項目)  検索する

▼ 項目別検索 (自動の項目で絞り込みが可能です)

遺伝子 A-Z Index

1007 件のデータが検索されました。

1 / 21 ページ中 ※項目名をクリックすると階層または詳細に並び替わります。

遺伝子	別名	該当する造血器疾患	臨床的有用性	遺伝子変異の機能的作用とその種類	関連する染色体構造変化
ABL1		MPAL with t(9;22)(q34.1;q11.2); BCR-ABL1	1	機能獲得 (融合: BCR/ABL1)	t(9;22)(q34;q11)
ABL1		MPAL with t(9;22)(q34.1;q11.2); BCR-ABL1	1	機能獲得 (変異: T315I など)	
ABL1		CML, BCR-ABL1+	1	機能獲得 (融合: BCR/ABL1)	t(9;22)(q34;q11)
ABL1		CML, BCR-ABL1+	2	機能獲得 (融合、再構成: BCR/ABL1以外)	
ABL1		CML, BCR-ABL1+	1	機能獲得 (変異: T315I など)	
ABL1		B-ALL/LBL with t(9;22)(q34.1;q11.2); BCR-ABL1	1	機能獲得 (融合: BCR/ABL1)	t(9;22)(q34;q11)
ABL1				機能獲得 (融合: NUP214/ABL1)	t(9;9)(q34;q34)

(<http://www.jshem.or.jp/modules/genomgl/>)

# 血液がんにおける試料調製について（1）

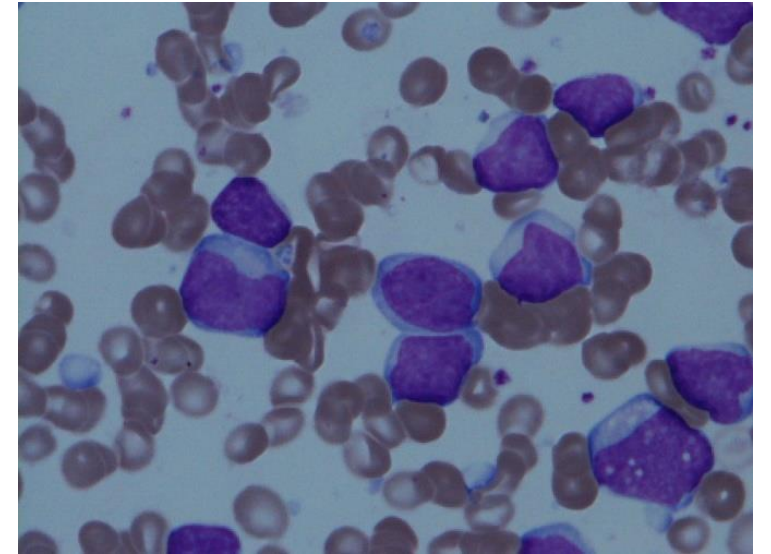
血液がんにおける試料調製にはいくつかの細かい注意点が存在する。

- 血液がんは一般的に進行がはやく、検体の提出から結果返却までの時間（**Turn around time/TAT**）の影響を受け易い為、よりスピーディな検体の提出が必要である

※結果返却の時間短縮を目指すには：

- ✓ 造血器腫瘍シーケンス検査の実施体制：エキスパートパネルの在り方
  - ✓ 臨床的意義に関する情報集約：エビデンスレベル・治験情報
- について、関係者一体となった連携体制の構築が必須であることも議論されている

- 固形がんの場合と同様にがん細胞含有率が一定以上であることが望ましい
- がん細胞含有率を調べる方法としては、白血病では骨髓塗抹標本を用いて測定されるが、悪性リンパ腫では固形がんと同様に病理標本を用いた評価が行われる
- 固形がんと異なりフローサイトメトリー（細胞ソート法）によるがん細胞数の計測も補助的に用いられる
- 固形がんの場合のレーザーマイクロダイセクション（LMD）と同様に、血液がんにおいてもFACSにより選別された細胞のみからゲノムDNAを抽出してNGS解析に用いることで、マイナーな細胞集団の解析も可能となる



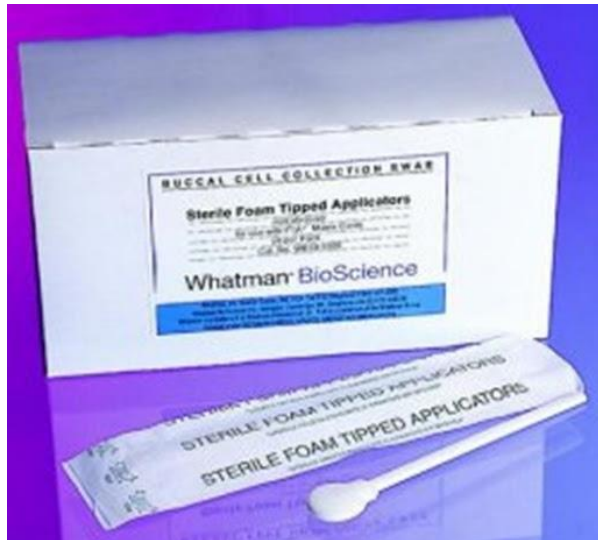
急性白血病の塗抹標本

# 血液がんにおける試料調製について (2)

特に検体採取にあたっての注意点を記す。ゲノムDNAの正常対照には口腔粘膜細胞から採取したDNAが用いられる

- 唾液による消化を防ぐために、採取後、検体は速やかに乾燥することが望ましい
- 血液細胞の混入を最小限に防ぐために、採取前には患者にうがいをしてもらうことが望ましい

## QIAGEN WB100032



(<https://www.midlandsci.com/50555.html>)



比較的幅の広いものを利用すると細胞/DNAの回収量が良い  
左右10回程度づつ、よく口腔内をこするように指導する

(<https://www.fishersci.se/shop/products/whatman-fta-card-applicators/11380844>)



# ゲノムDNAからのNGS鑄型の調製

ゲノムDNAを用いてシーケンス解析用の試料調製（鑄型調製）が行われる

- 多くの局面で、ゲノムDNAに対してのNGS解析用の鑄型調製は検査センターで行われる
- ただし、状況によって（精製されたDNAの質・量が検査センターの要求基準に満たない場合等）、自施設で鑄型調製が必要な場合もある
- DNAの破碎：ゲノムDNAをNGS解析に適した300-700 bpの断片へと破碎する。破碎方法は酵素による切断と超音波による物理的な剪断に大別される

- ✓ **酵素法**：切断活性を有したトランスポゾンのゲノムへの挿入により行われる。ゲノムへの挿入はランダムな箇所について起こるが、ある程度のバイアス（詳細は論文によっても異なる）があるとする報告もある。切断断端は一般的に平滑端となるために、それに続く反応を効率よく行うことが可能である。微量のDNAからの鑄型調製に向く
- ✓ **超音波破碎法**：DNAを超音波による物理的なエネルギーをかけて破碎する。剪断はゲノムに対して完全にランダムな場所にかかる。切断断端は一般的に2本鎖の間でそろわないために、末端の平滑化反応が必要となる。一定量のDNAロスが発生するため、極微量のDNAからの鑄型調製には向かない



※酵素法にはイルミナ社のキット“NEXTERA”が用いられることが多い

	酵素法	超音波法
破碎方法の原理	✓ トランスポゾンのゲノムへの挿入	✓ 物理的エネルギーの付加
代表的キット/機器	Nextera kit（イルミナ）	COVARIS（エムエス機器）
利点	<ul style="list-style-type: none"> <li>✓ 96穴プレート等での反応が可能であり、自動化（大量検体処理）に向く</li> <li>✓ 微量検体を扱うことができる</li> </ul>	✓ 場所のバイアスなくDNAを破碎することが可能である
欠点	✓ ある程度、切断点に偏りがあるとする論文がある	✓ 末端が平滑でないので、破碎後の反応の前にその修復反応が必要になる



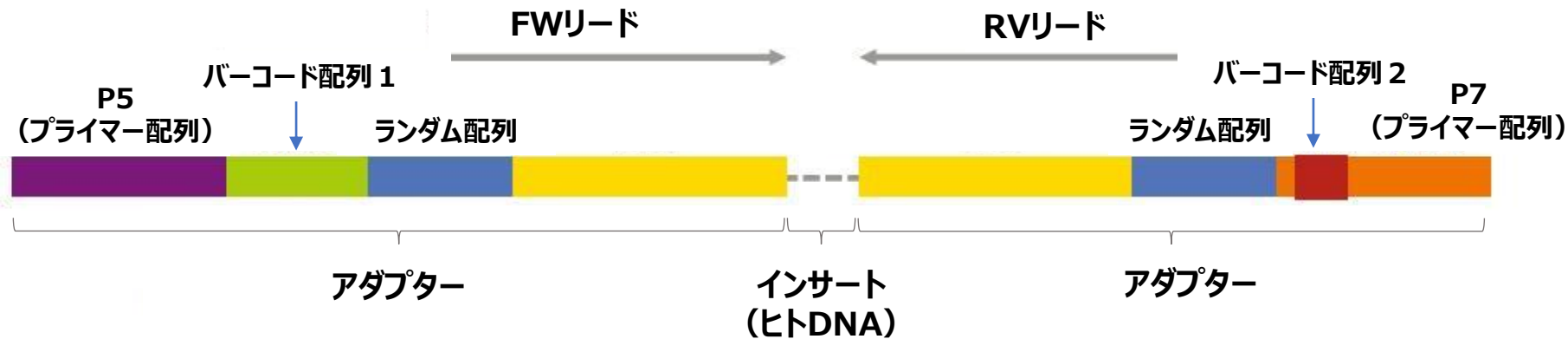
※超音波破碎にはエムエス機器社のCOVARISが用いられることが多い

# シーケンスライブラリの調製

DNA末端へ付加されたアダプター配列にはサンプルIDを区別するバーコードが含まれている

- 代表的な鋳型調製キットを示す（以下）。ライブラリ調整反応ではDNAの末端にアダプターが付加される。アダプターにはNGSのシーケンスの読み取りに必要な配列（P5およびP7プライマー配列）およびバーコード配列（後述）が含まれている。この工程を終了した産物を“シーケンス鋳型”（シーケンスライブラリ）と呼ぶ

## シーケンス鋳型の分子構造



TruSeq DNA PCR-Free HT Library Prep Kit

- DNAの断片化にトランスポゾンを用いた反応の場合、トランスポゾンによる切断と同時にアダプターを結合させるキットもある
- 必要に応じてP5/P7の配列を用いてPCRによる増幅を行う
- アダプターの部分には、サンプルバーコード（インデックス：インサートの前後に8 + 6塩基）と呼ばれる配列が存在する。検体によって異なるアダプター配列を用いて鋳型調製を行うことで、検体ごとに異なる配列を末端に導入することが可能である。これにより、多検体をまとめてシーケンス反応を行っても、産出される配列データの末端配列を手掛かりに、もともとの検体へとデータを振り分けることが可能である（標準的には、イルミナバーコードキットが用いられる。これには最大384検体を区別することができるアダプターが含まれている）

# 鋳型調製 (シーケンスライブラリ) の実例

鋳型調製の実例を示す。各ステップ終了時の典型的な“Tape Station” (p.50ご参照) によるDNAの長さとの量の定量的結果である

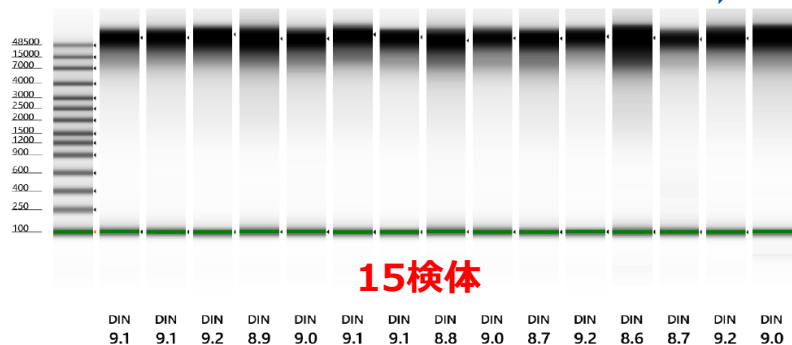
抽出DNA



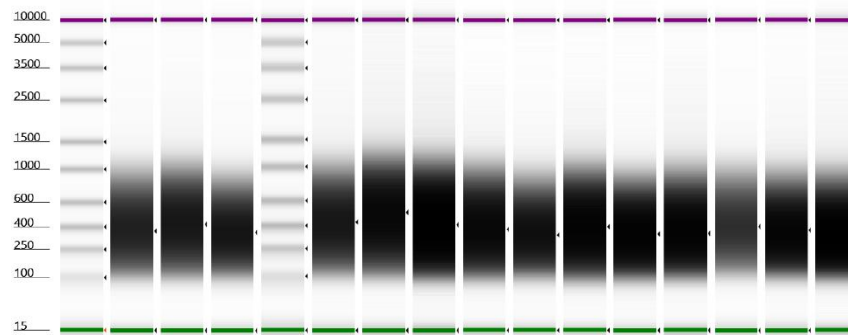
DNA破碎



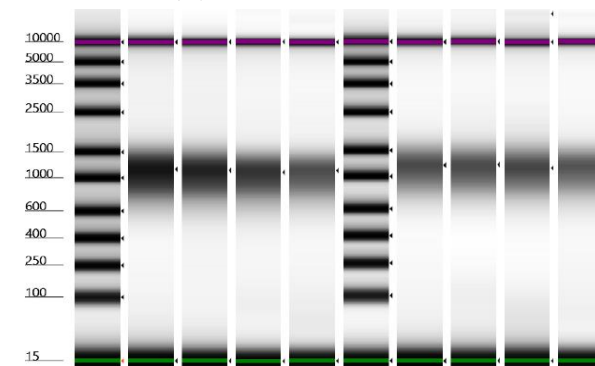
シーケンスライブラリ



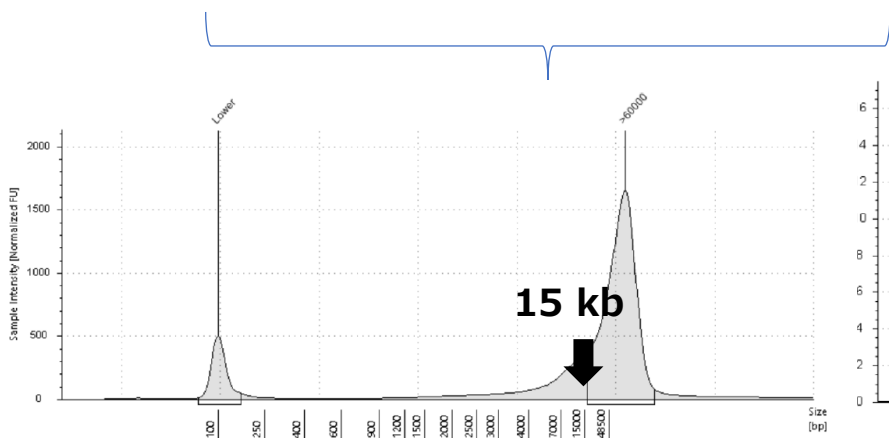
15検体



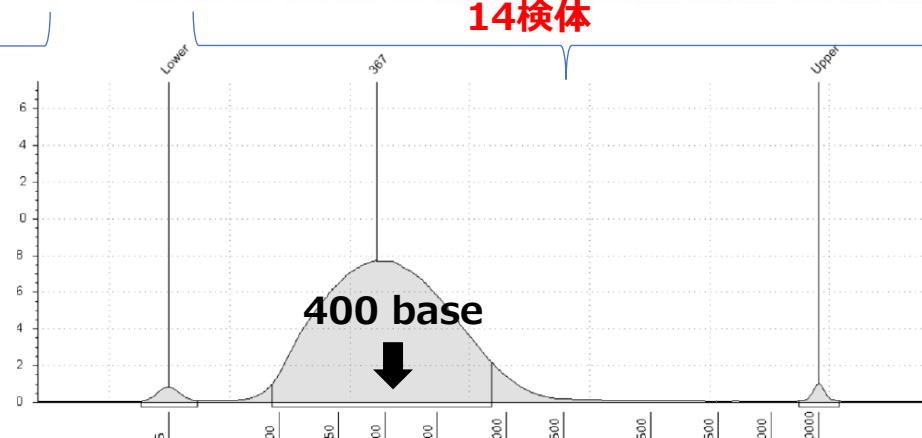
14検体



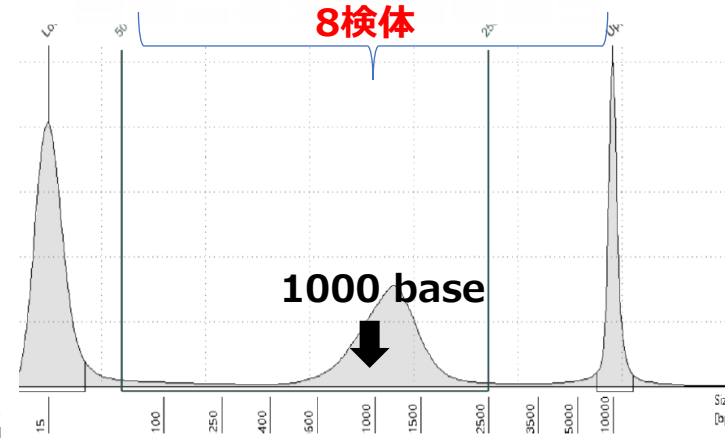
8検体



代表例



代表例



代表例

- ライブラリサイズは、インサートサイズピークが通常400-600 bp程度にくるように調製する。ただし、目的によってインサートサイズをより長く/短く調製することも可能である。また、DNA濃度は、3 nM以上あると後のシーケンス反応の際、濃縮が不要となる

# シーケンスの読み取り

実際のシーケンス（DNA塩基配列の決定）は以下の通り行われる

鋳型調製を終えたシーケンスライブラリをDNAシーケンサーにかける



30X（1塩基が平均30本の配列でカバーされる）のシーケンス深度でヒト全ゲノムを解読とした場合（1検体あたり90Gbあたりの塩基配列データが必要とした場合）、1解析（3Tb）で30人程度のヒト全ゲノム解析を行うことができる（1回の“ラン”では2枚分なので60人分の解析が可能である）



そのため、30人分のシーケンス鋳型を混ぜてシーケンス解読を行い、インデクス配列を指標に各検体へと割り振る、という手法が一般的である

- ✓ ヒトの全ゲノム解析においては、一般的に、最も読み取り性能の高いイルミナ社のシーケンサーである**NovaSeq6000**が用いられることが多い
- ✓ その中でも出力が最大であり検体あたりのコストが最小となる**フローセルS4**が頻繁に用いられる（S4においては解析単位（1“フローセル”と呼ぶ）あたり3Tb程度の塩基配列を解読することができる）



Nova Seq  
6000

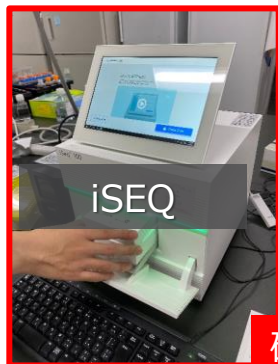
# DNAシーケンサー

イルミナ社が提供するヒトゲノム解析に用いられるDNAシーケンサーの一覧を示す

(詳細は以下のHPを参照 <https://www.illumina.com/systems/sequencing-platforms.html>)

ヒトゲノムの解析には鋳型確認用にiSEQ、データ産生にはNovaSeqが用いられることが多い。これらのシーケンサーを赤枠で示す

超小型



確認用

ハイスループット



ロースループット ©Illumina, Inc. All rights reserved.

集団規模の  
ヒト全ゲノム解析



最大出力6Tb程度/run  
(3Tb程度の出力のS4x2枚  
の実装が可能)

解析用

プロダクション規模

# フローセルの性能

Nova Seqに用いられるそれぞれのフローセルの性能を示す

## フローセルあたりのシーケンスアウトプットとラン時間

NA: Not applicable (用いられない)

フローセルタイプ	SP	S1	S2	S4
1 × 35 bp				280-350 Gb (14時間)
2 × 50 bp	65-80 Gb (13時間)	134-167 Gb (13時間)	333-417 Gb (16時間)	N/A
2 × 100 bp	134-167 Gb (19時間)	266-333 Gb (19時間)	667-833 Gb (25時間)	1600-2000 Gb (36時間)
2 × 150 bp	200-250 Gb (25時間)	400-500 Gb (25時間)	1000-1250 Gb (36時間)	2400-3000 Gb (44時間)
2 × 250 bp	325-400 Gb (38時間)	N/A	N/A	N/A

2 x: forward+ reverse リードの両側読みを指す



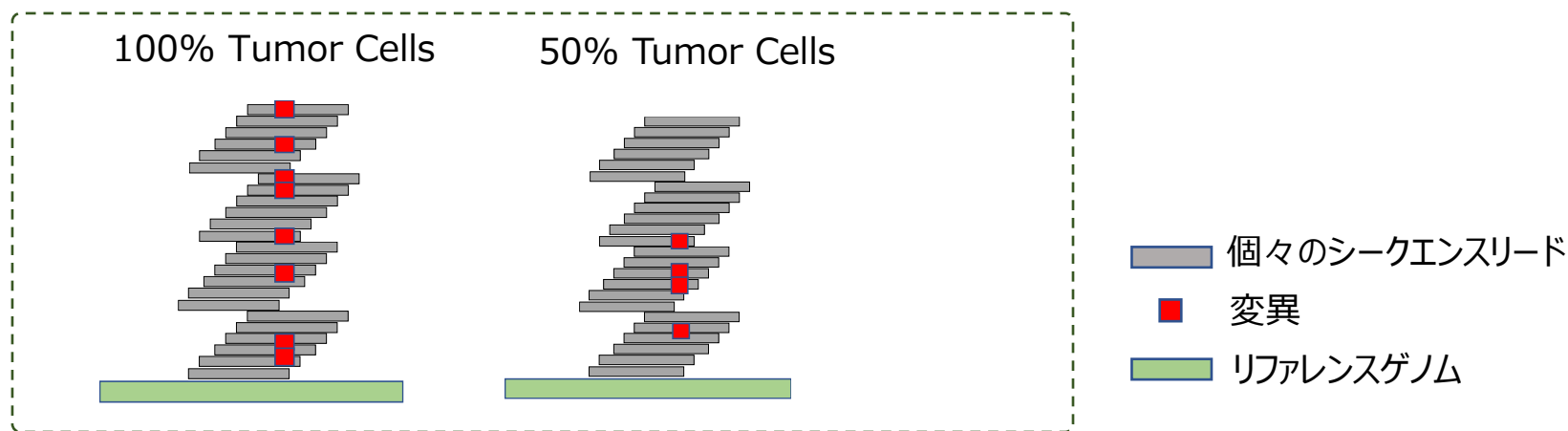
**S4フローセル** : ヒトがん全ゲノム解析に主として用いられる

## 主要アプリケーションのサンプルスループット\*

フローセルタイプ	S1	S2	S4
ランあたりのヒト全ゲノム	~8	~16	~48
ランあたりのヒト全エクソーム	~80	~200	
ランあたりのヒトパネルシーケンス	>100	>200	

### サンプル調製に関連したシーケンスデータのパターンを示す

- がんゲノムの場合、がん細胞含有率が必ずしも100%ではない。そのため30Xよりも高い深度（60X程度の深度）で解析を行うことも多い

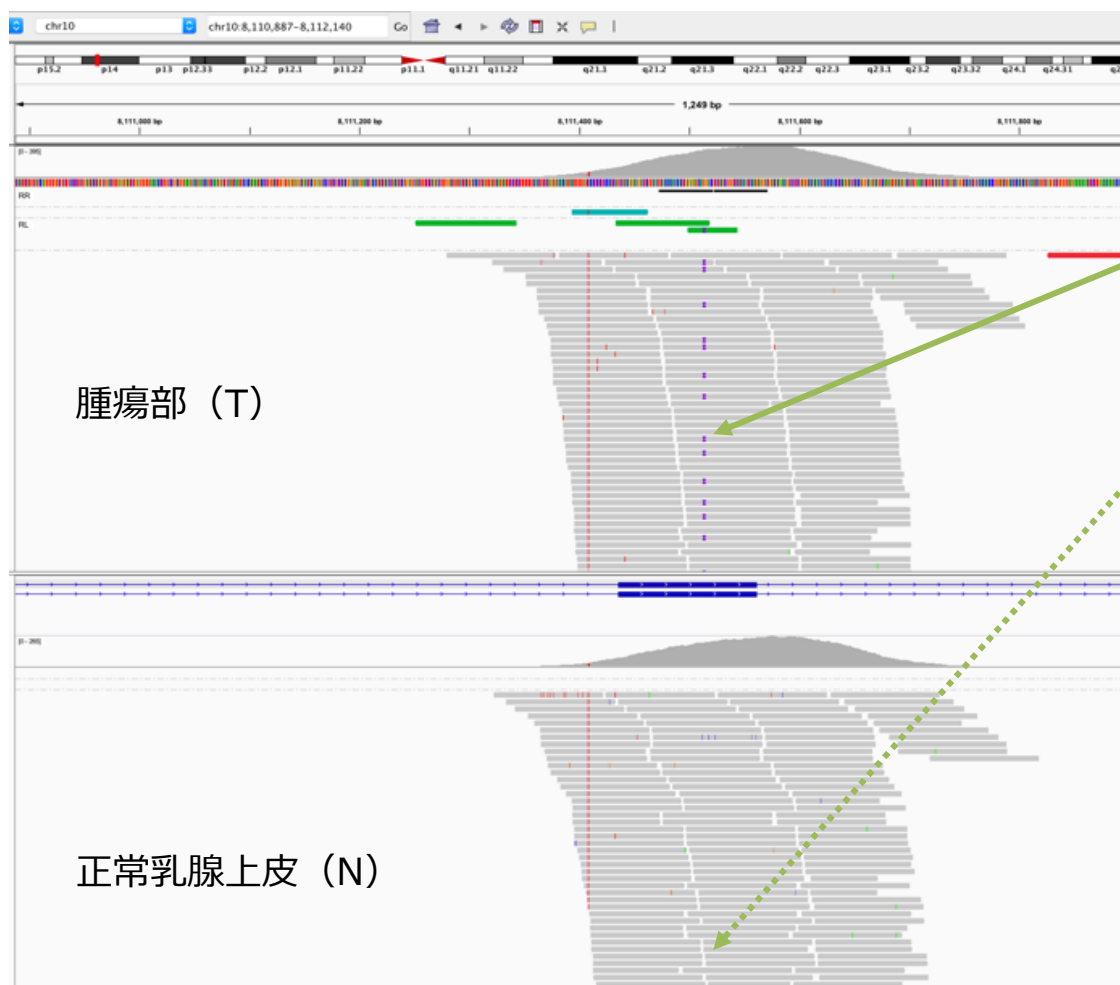


変異は片方の染色体にのみ入ることも多いので、50%のリードは正常塩基を持つ。具体的な例は次ページに示す

- がんゲノム解析において検出された塩基変異が、体細胞変異であることを確認するために同一個人の非がん部（多くの場合は血液が用いられる）についても同様のシーケンス解析を行う
- がんゲノムを“T”（Tumor）、正常部位ゲノムを“N”（Normal）と呼ぶ。体細胞突然変異“T”のみに存在して“N”には存在しない塩基変異として検出されるこれを“T/Nペアでの解析”と称する
- 正常ゲノムのデータベースの充実により“T”のみでの解析を行って、差分をとるべき正常部位には正常DNAデータベースを参照する、という手法も最近、取られ始めている。ただし、欧米に比して日本人正常ゲノムのデータ蓄積量はまだ十分とは言えない

# T/Nデータの実例

がんゲノム解析のデータの可視化例を示す



腫瘍部でのみ色の変化  
(塩基変異がある ;  
exon6\_c.1312dupA\_p.S437fs )  
→体細胞変異

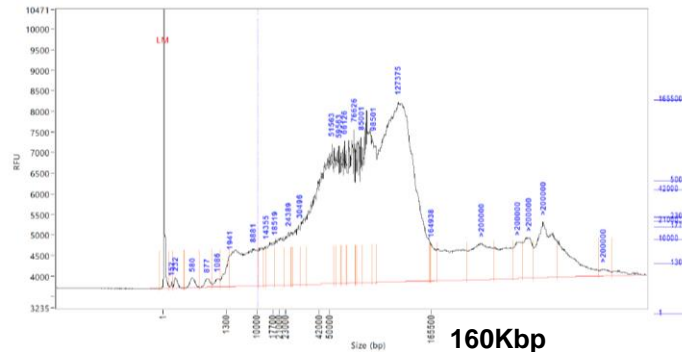
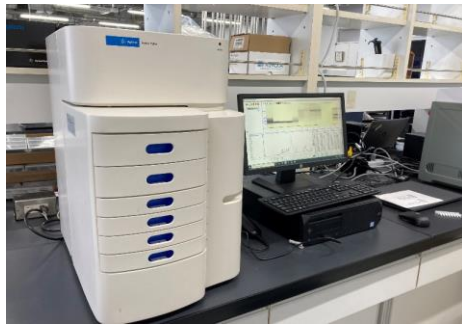
配列データの表示 (可視化) にはIntegrated  
Genome Viewer (IGV) を用いた



# 長鎖DNA解析用の鋳型調整

長鎖DNA解析に向けては、鋳型調製が大きく異なる。

- ✓ FFPE検体では、DNAの分解が進んでいることが多い。長鎖DNA解析には新鮮凍結検体を用いる。
- ✓ 前述のTape Stationでは数十kb以上のゲノムDNAを正確に計測することは不可能である。その目的にはAgilent社が販売するFemtoPluseを用いる。ただしあまり国内で導入例の多くない機器であるために、ゲノムDNAが概ねTape Stationの測定限界以上の長さであることが確認できればよい、とする場合も多い。

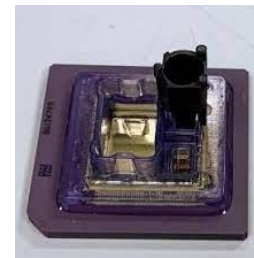


図：左図：FEMTOPLUSEの外観：右図：FEMUTOPULSEを用いた電気泳動パターン：150kb程度のサイズのDNAまで計測できている。

- ✓ シークエンサーには第一章で紹介したナノポア社製のPromethIONとPacBio社製のSequel IIがある。このいずれについても1回の解析あたり1-10 $\mu$ g程度の比較的大量のDNAが必要になる。
- ✓ 抽出されたDNAを用いて鋳型調製を行う。調製には下記に示す各シーケンサーに専用のキットを用いる。



SEQUEL II



図：鋳型調整キットとフローセルの外観  
(左図：ナノポアPromethION; 右図：PacBio Sequel II)

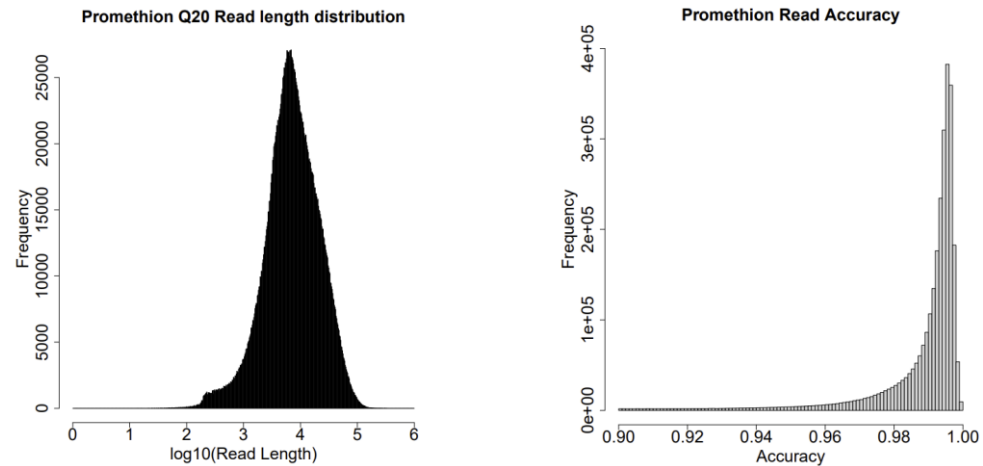
<https://www.pacb.com/wp-content/uploads/Sequel-II-System-v7.0-and-SMRT-Link-v7.0-Overview-Customer-Training.pdf>

- ✓ 近年、いくつかの受託解析会社もできているので、十分な量と品質のゲノムDNAが抽出できた時点で、そちらに送付することが可能な場合もある。

# 長鎖DNA解析データ例：PromethIONの場合

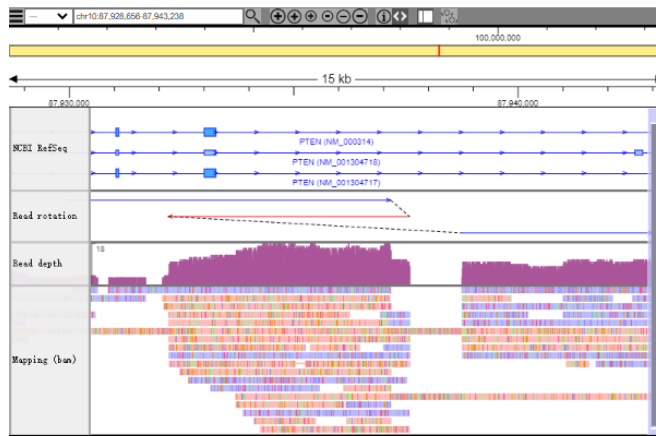
長鎖DNA解析では短鎖DNA解析に比べて、読み取り塩基精度が落ちる。

- ✓ 通常の1本のリードでの塩基精度は90%であった。しかし近年の機器開発で大きな精度向上がみられている（最新のQ20だと99%程度の精度が得られる）



図：がん細胞株を対象に行ったPromethIONを用いた長鎖ゲノム解析  
左図：読み取り長の分布、右図：個々のシーケンスリードの読み取り精度

- ✓ 長鎖DNA解析では一般に一塩基変異の解析精度に劣るが、逆位・染色転座といったより大域的な構造異常の検出に優れる。



図：肺腺癌細胞株中に見出された逆位 + 重複の例：赤い部分で塩基の並びの方向性が逆転している。ただし1塩基置換は発生していないことに注意（Sakamoto et al Genome Res 2021）。

# 第Ⅲ章






# 遺伝学的検査とNGSのメリット

東京大学医科学研究所附属病院 血液腫瘍内科 教授

南谷泰仁

# 既存遺伝学的検査とNGS検査

既存の遺伝学的検査とNGS検査で検出可能な変異を以下にまとめる

検査手法	 SNVs indels	 CNVs	 リピート 伸長異常	 構造異常	 ミトコンドリア 異常
サンガー法	●				●
PCR法	●	●	●	変異による	
FISH		●		●	
Karyotype		大きな異常は可能		●	
染色体マイクロアレイ		●			
NGSパネル検査	●	限定的		限定的	パネルに含む場合
NGSエクソーム解析	●	限定的		限定的	●
NGS全ゲノム解析	● <sup>1</sup>	● <sup>1</sup>	● <sup>2</sup>	● <sup>3</sup>	● <sup>1</sup>

**CMA**=chromosomal microarray; **CNV**=Pg3ご参照; **FISH**=fluorescence in situ hybridization; **Indel**=small insertion/deletion; **NGS**=next-generation sequencing; **PCR**=polymerase chain reaction; **SNV**=single nucleotide variant; **WES**=whole-exome sequencing; **WGS**=whole-genome sequencing

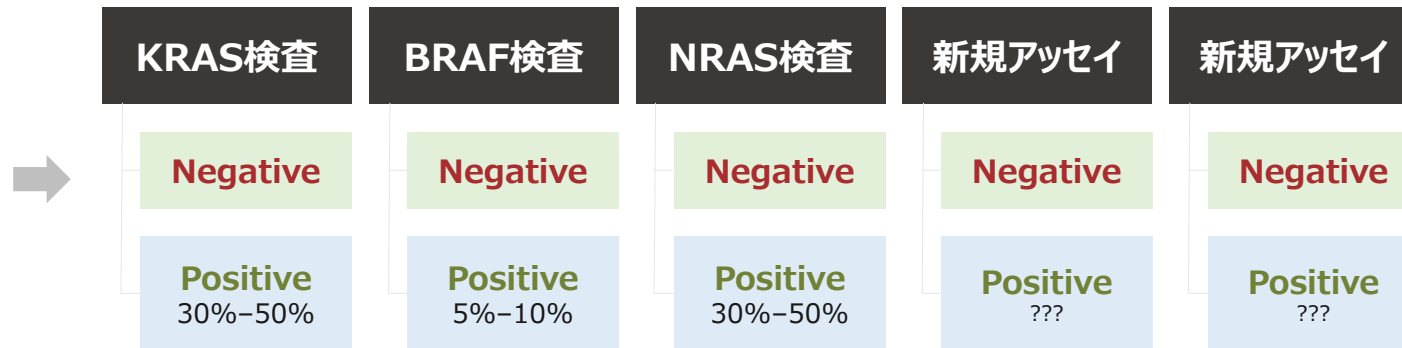
(**References:** **1.** Lionel AC, Costain G, Monfared N, et al. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genet Med.* 2018 Apr;20(4):435-443. doi: 10.1038/gim.2017.119. Epub 2017 Aug 3. **2.** Dolzhenko E, van Vugt JJFA, Shaw RJ, et al. Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Res.* 2017;27(11):1895-1903. doi: 10.1101/gr.225672.117. **3.** Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics.* 2016;32(8):1220-1222. [http://doi.org/10.1093/bioinformatics/btv710.](http://doi.org/10.1093/bioinformatics/btv710))

# 既存遺伝学的検査とNGS検査

既存の遺伝学的検査とNGS検査の特徴について説明する



腫瘍組織検体

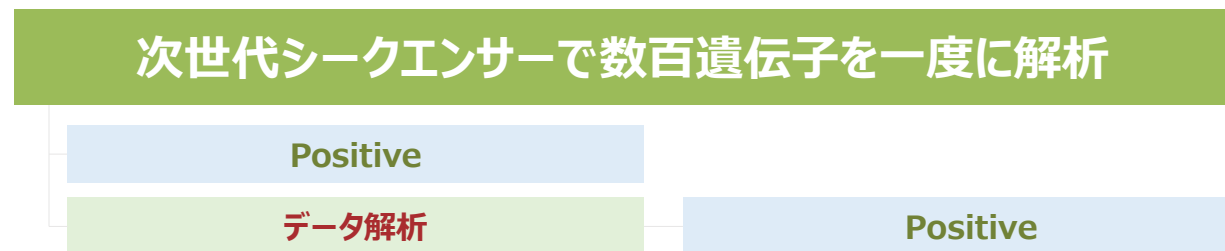


## 既存の遺伝学的検査

シングルアッセイの繰り返し  
組織検体、時間、がかさむ  
1つ1つの検査は安価



腫瘍組織検体



## NGS検査

1度の網羅的解析  
組織検体、時間を節約  
検査は高価（合計コストは安価）

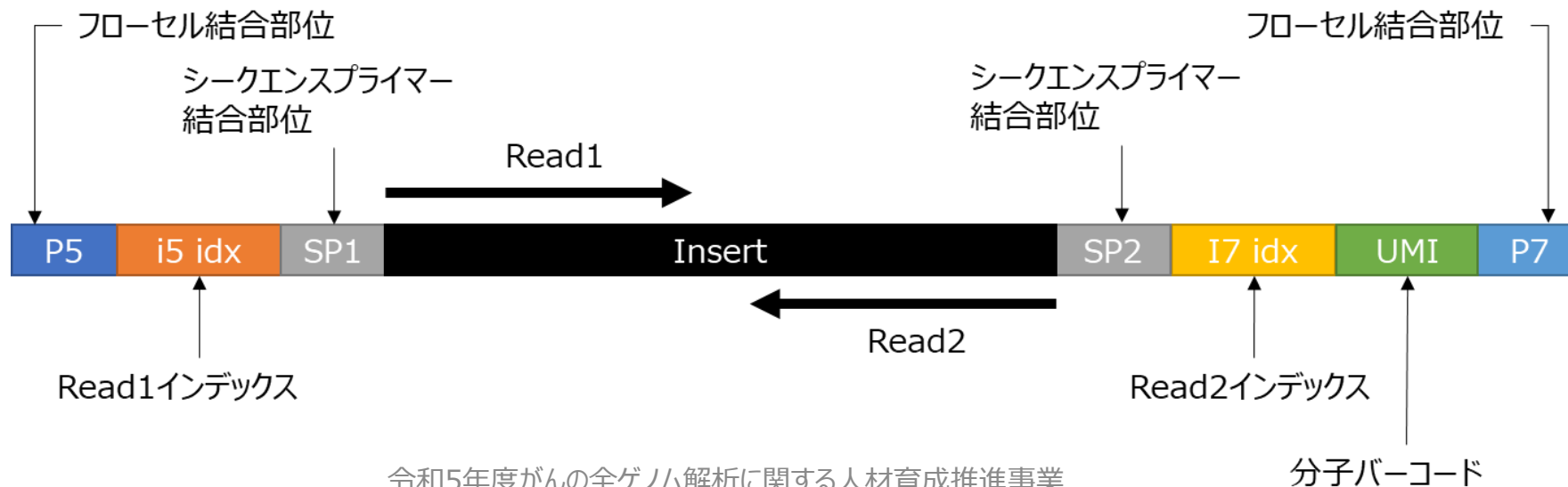
(<sup>§</sup>Roth AD<sup>1</sup>, Tejpar S, Delorenzi M, et al. Prognostic role of KRAS and BRAF in stage II and III resected colon cancer: results of the translational study on the PETACC-3, EORTC 40993, SAKK 60-00 trial. J Clin Oncol. 2010 Jan;28(3): 466-474.)

# 短鎖シーケンスのreadの構造

短鎖シーケンスの原理を理解するためにはリードの構造を理解することが必要である

- ペアエンドは、Insertの両側からreadを読む方法である
  - ✓ 短鎖シーケンスはreadが長くなるとシーケンスのクオリティが低下する。そのため、read長を抑え、代わりに両側から読む方がクオリティの低下が抑えられる
  - ✓ ペアエンドリードは構造異常の検出に有利である（後述）
- サンプル特有のインデックス配列の付与によって多検体の同時処理が可能になる（次スライド）
- 分子バーコードを含めることができる
  - ✓ アダプター部位のそばに数塩基のランダムな配列を含めることができ、これを**分子バーコード**とよぶ
  - ✓ 分子バーコードはエラーコールの除去に有効だが、代わりに取得するべきデータ量が増加する

短鎖シーケンスのリードの構造（例）



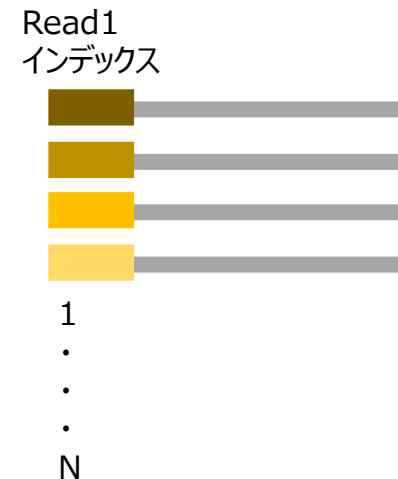
# Indexによる複数sampleの測定

短鎖シーケンスの原理を理解するためにはリードの構造を理解することが必要である

- サンプル特有のインデックス配列の付与によって多検体の同時処理が可能になる
  - ✓ **インデックス**は、多くの検体をまとめて処理をおこない、解析する時点で検体毎のデータに分けるために用いられる
  - ✓ インデックスとreadの情報が合わないことを**ホッピング**とよび、エラーコールの原因となる
  - ✓ インデックスをRead1側とRead2側の両方に設定することができる (**デュアルインデックス**)
  - ✓ インデックスは、Read1側とRead2側に、対応のある重複のないものを設定することで、精度を高めることができる (**ユニークデュアルインデックス**)

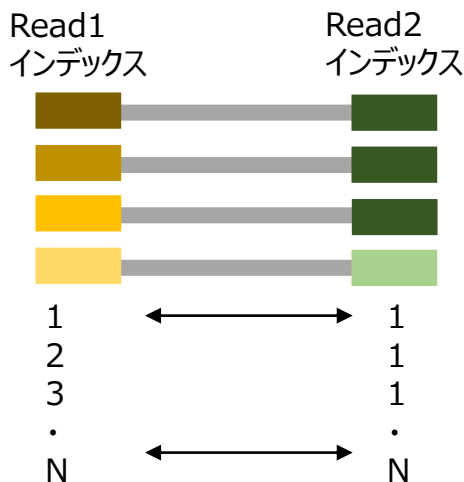
## Single index

N種類のサンプルを区別できる  
Hoppingが生じた場合には、  
対応できない



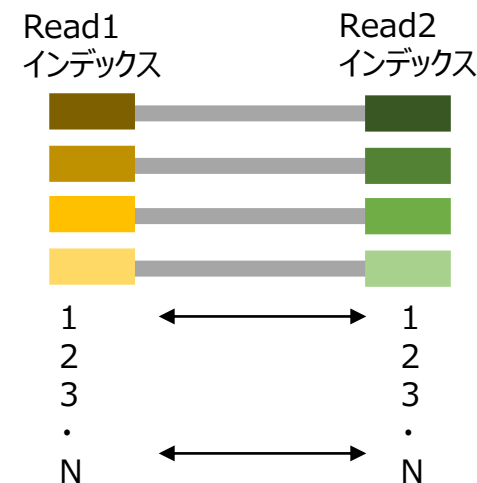
## Non-unique Dual index

Read1とRead2のインデックスの対応がない場合：Read1インデックスの数 x Read2インデックスの数通りの組み合わせを区別可能だが、(この場合  $N \times N = N^2$  とおり) ホッピングの影響を受けやすい



## Unique Dual index

Read1とRead2のインデックスの対応がある場合：区別可能な検体数はNとおりとなるが、両方のインデックスが合っているreadのみを割り振るため、ホッピングの影響を最小限に抑えることができる



# 短鎖シーケンサーの原理

## 主な短鎖シーケンサーの原理を理解する

### ● Illumina社のシーケンサーの原理（右図）

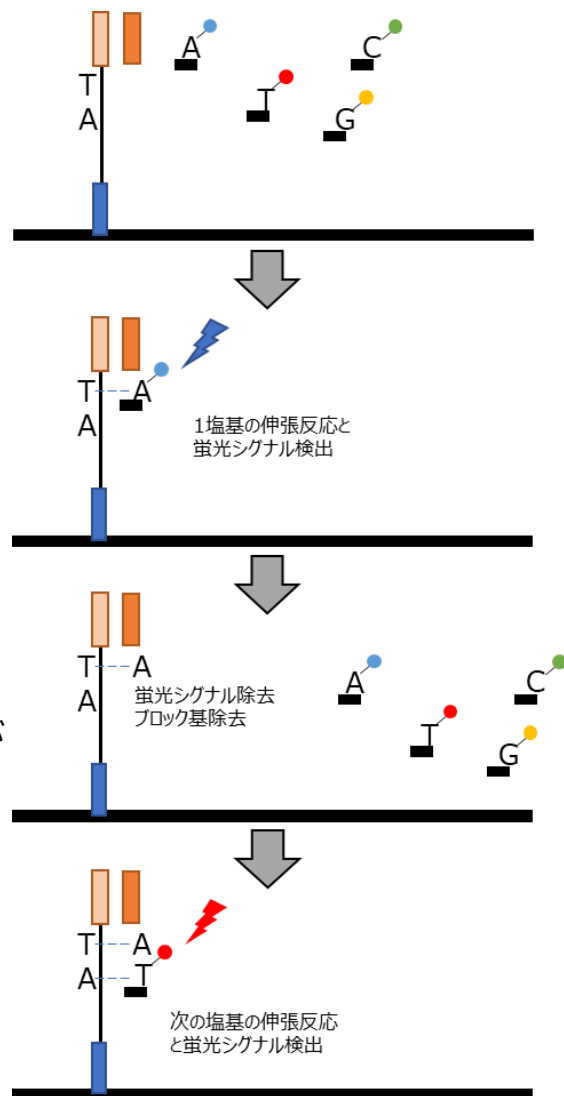
- ✓ ライブラリをテンプレートとして、1塩基毎の伸長反応をおこない、取り込まれた塩基に付加された蛍光シグナルを読み込む
- ✓ これを繰り返すことで、連続した塩基配列の情報を取得する
- ✓ アダプターを逆向きに接地することで、逆のアダプターから配列を読み、ペアエンドを実現している

### ● ThermoFisher社のIon Torrentシステムの原理

- ✓ 1塩基毎の伸長反応をおこなう点はIllumina社のシステムと同様であるが光学系を利用しない
- ✓ DNA伸長時に放出される水素イオン濃度を測定する
- ✓ 4つの塩基を順に送り込み、伸長反応が生じた際のみ水素イオンが生じるため、どの塩基が伸長反応に使用されたかが検出できる

### ● MGI社のDNBSEQシステムの原理(第1章CPAS法参照)

- ✓ Illumina社システムのクラスター生成の代わりにSingle strand DNAからDNA nanoballを作ること、蛍光強度を保つ。Linearな増幅系を採用することで、エラーが増幅しない



## Illumina社のシーケンサーの原理

伸長反応に用いるdNTPには、以下の特徴をもつ

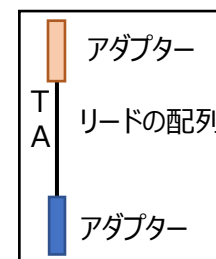
- ① A、T、C、Gの塩基を区別する異なった蛍光シグナルが付加されている
- ② 3'位がブロック基が付いており、伸長反応が阻害される

1塩基ずつ伸長反応を行う。相補的な塩基が取り込まれる。他のdNTPを除去し、蛍光を読み取る

ブロック基を除去し、さらに伸長反応を1塩基分おこなう。新たに相補的なdNTPが取り込まれる

これを繰り返す

実際には検出可能な蛍光シグナルの強度を確保するために、同じリードを局所で増幅している（これを**クラスター生成**という）





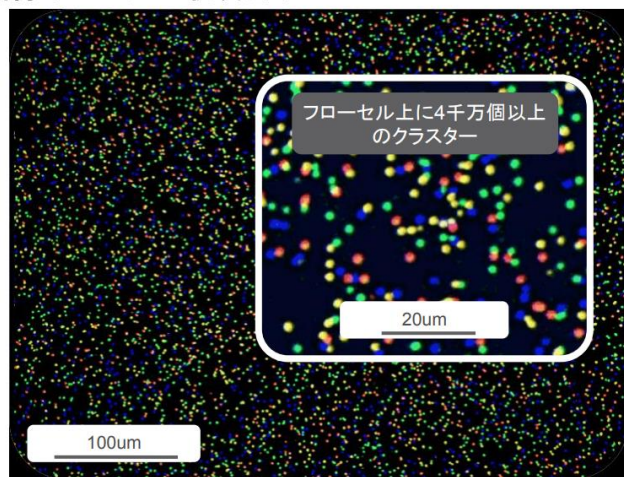
# FASTQファイルの生成とその内容

FASTQファイルはシーケンサーが出力した塩基配列情報とシーケンスクオリティ情報を含む

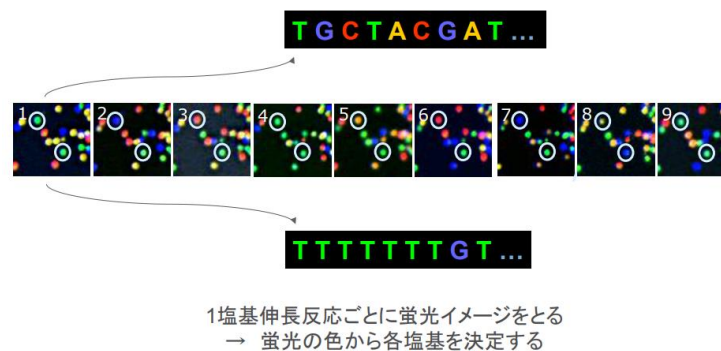
- **FASTQファイル**には、シーケンサーから出力されたリードの情報が含まれる
  - ✓ 同じreadから得られる配列情報をつなげて作成される
  - ✓ テキストベースであり、ヒトが判読できる
  - ✓ 4行で1つのリードを記述する
  - ✓ 配列情報と、コールのクオリティスコアが含まれ、マッピング情報は含まれない
- クオリティスコアも1塩基あたり1文字で表現される
  - ✓ **クオリティスコア**は、1文字毎に1塩基のスコアを示す
  - ✓ ASCII文字コードに変換し、33を引いた数字で示される
  - ✓ 一部のシーケンサーではクオリティの評価を簡略化し、4段階で評価している

## illumina社のシーケンサーの塩基情報の取得

### 画像イメージの取り込み



### 画像蛍光シグナルから塩基への変換



## FASTQファイルの情報

- ✓ 1行目：readの名前
- ✓ 2行目：塩基配列
- ✓ 3行目：区切り (+の1文字)
- ✓ 4行目：クオリティスコア

```
例
@read1_locusAAA
ACACGTACGTACGTACGT...
+
@B(92==9>2@=@##!...
```

(画像はイルミナ社資料より抜粋)

# 長鎖シーケンサーとその原理

## 長鎖シーケンサーの原理を理解する

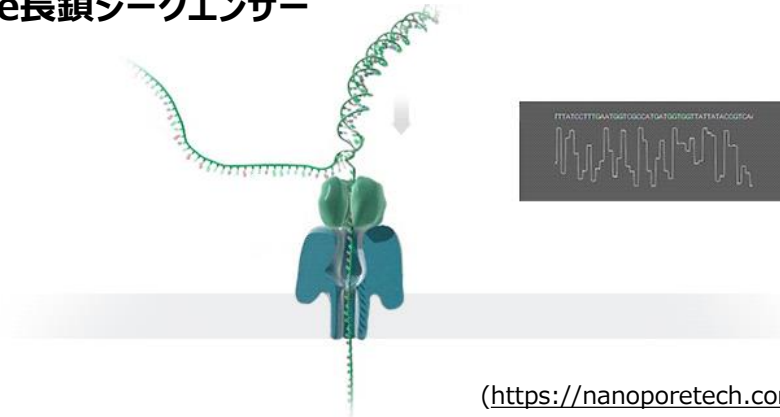
### ● Nanopore方式の原理

- ✓ Oxford nanopore technologies (ONT) が採用するシーケンサーでは、膜たんぱく質中に埋め込まれたナノサイズの穴（ナノポア）にDNA1本鎖を通過させ、各塩基が通過する際のイオン電流の変化で配列情報を読み出す
- ✓ 一分子のDNAを対象としてシーケンスをおこなうため、DNAの増幅が不要である
- ✓ DNAの修飾として、メチル化シトシンの検出も可能である
- ✓ 数百Kbの長さまで読むことができる

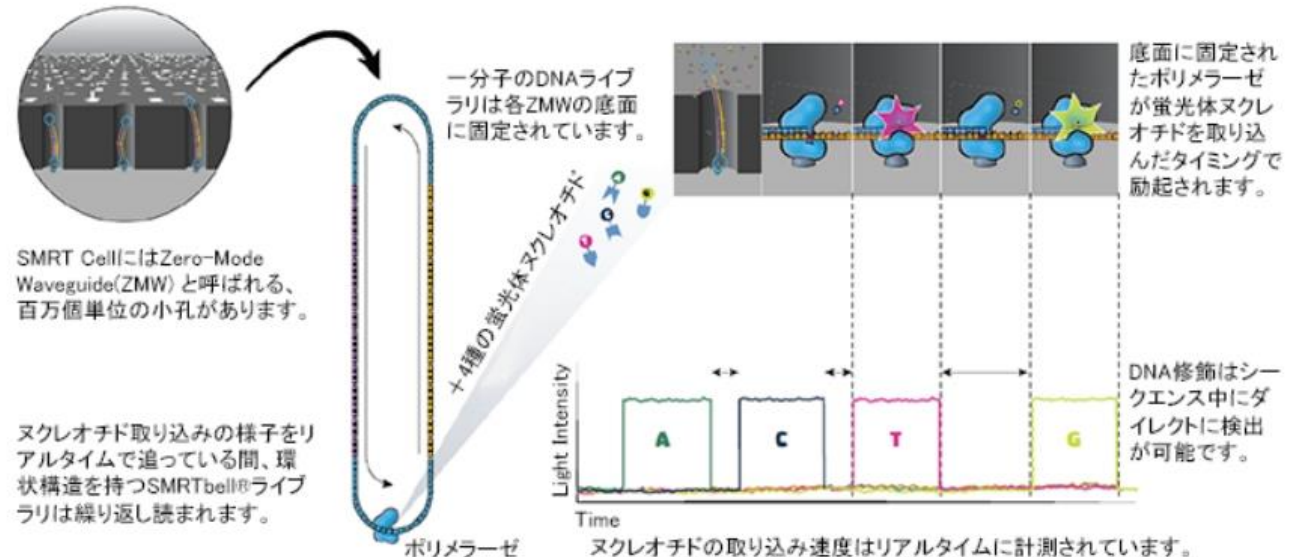
### ● PacBio社のSMRT方式の原理(第1章SMRTテクノロジー参照)

- ✓ PacBio社は、二重鎖DNAの両端にアダプターを付加して環状の1本鎖DNAをテンプレートとする
- ✓ 1本のテンプレートDNAを1つの小孔に入れ、その中でテンプレートDNAの相補鎖を合成する。この反応を1分子リアルタイムリード (SMRT) とよぶ。1塩基の伸長反応毎に、取り込まれた塩基に付着した蛍光標識のシグナルが読まれる
- ✓ HiFiモードは、環状のテンプレートを繰り返し読むことで高い精度を実現している (99.999%)

## Nanopore長鎖シーケンサー



## PacBio SMRT



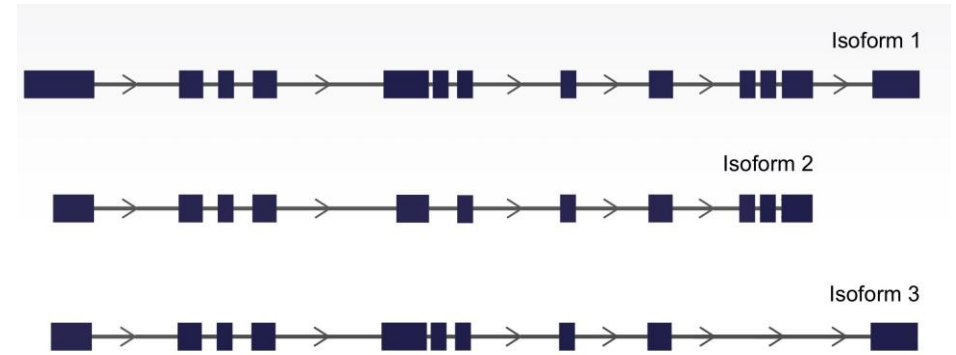
(https://www.digital-biology.co.jp/allianced/products/pacbio/)

長鎖シーケンシングの主な利用用途は下記の通り

- 繰り返し配列領域のゲノム配列の決定
  - ✓ 繰り返し配列部位の、繰り返し数を決定することが可能となる
  - ✓ タンデムに連続する繰り返し配列は、ある種の疾患の発症と関与しており、これまで*FMR1*、*ATXN10*、*HTT*などの遺伝子の繰り返し配列の決定に長鎖シーケンスが寄与した
- ハプロタイプの決定
  - ✓ 2つの異なるvariantが同じアレルに存在するか（cis）、異なるアレルに存在するか（trans）を判定することを**ハプロタイプの決定（phasing）**という
  - ✓ SNPsのphasingをおこなうことで、変異アレルが父由来か母由来かを判定することが可能となる
  - ✓ ハプロタイプの決定を拡張することで、ヒト染色体の全長の配列が決定された
- 偽遺伝子（pseudogene）の同定
  - ✓ 偽遺伝子は機能を有する遺伝子に重複が生じ、さらに変異によって機能を喪失したものと定義される
  - ✓ オリジナルの機能を有する遺伝子の変異を検出する際に、短鎖sequenceでは偽遺伝子との混同が生じるため、偽遺伝子に特異的な変異とphaseの決定をおこなうことで、変異が機能遺伝子に生じたものであることを確認できる（ADPKDの*PKD1*遺伝子）

- 構造異常の決定
  - ✓ 大きな欠失などの構造異常の同定に役立つ
- RNAにおけるsplicing isoformの同定
  - ✓ 全量のmRNAの配列を決定することで、局所的なsplicing variantでは決定できないisoformを決定できる

## mRNA全長のsplicing isoformの同定



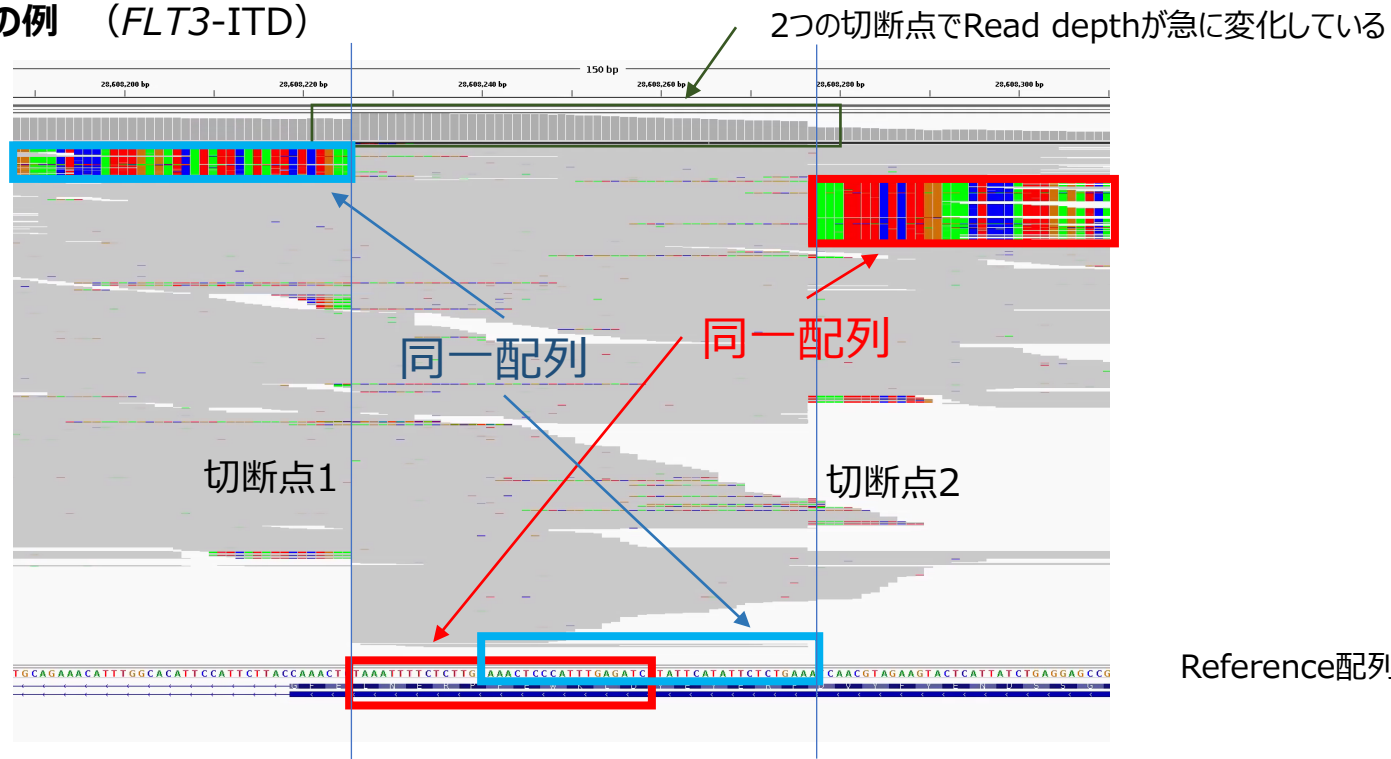
(Front Genet. 2019 May 7;10:426)

短鎖シーケンスでは各isoformの全体像を知ることは困難である

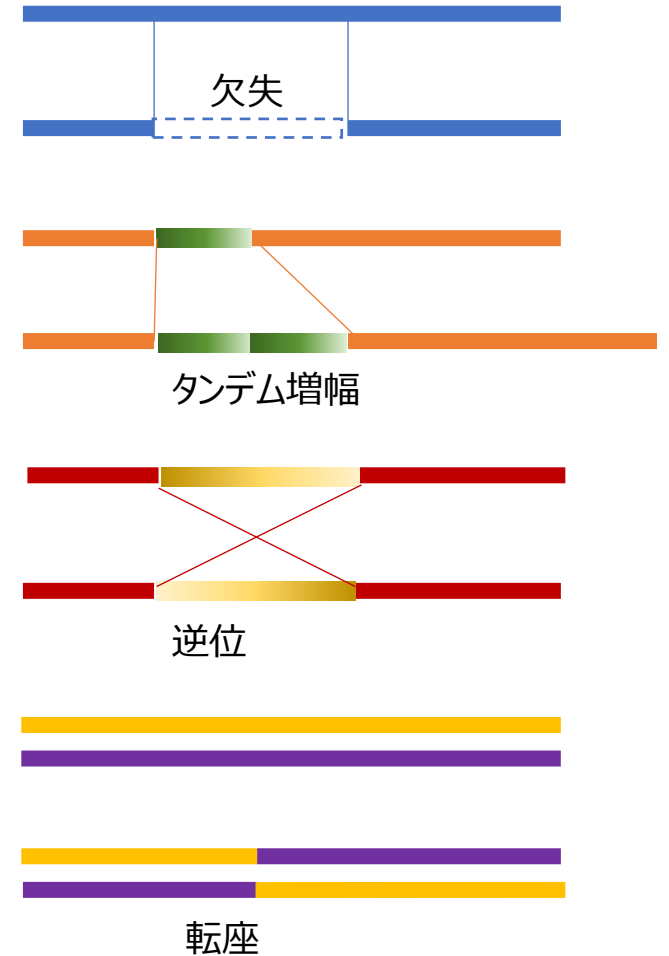
融合遺伝子を形成するタイプの構造異常の同定に、全ゲノム解析は有用である

- 構造異常は、「欠失」、「挿入（タンDEM増幅）」、「逆位」、「転座」に分類される
- 切断点が2カ所存在する
  - ✓ 切断点の周囲には、組み替え後の配列がみられる

## タンDEM増幅の例 (FLT3-ITD)



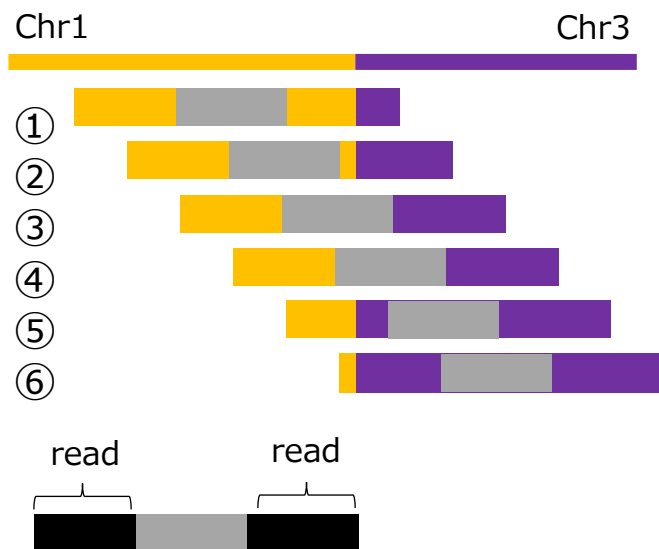
## 構造異常の分類



全ゲノムシーケンスにより既存の技術では検出されない構造異常の検出が見込める

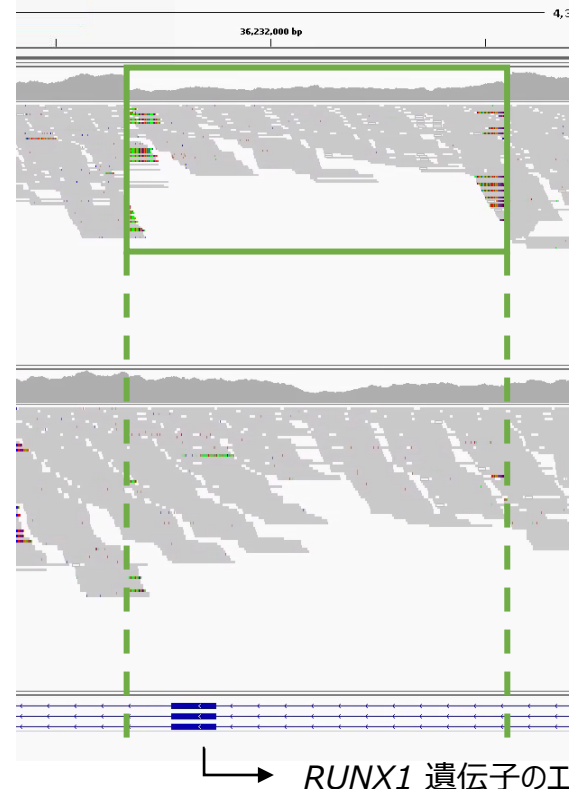
- 切断点がイントロンに生じた場合には、短鎖シーケンスでは、ペアエンドを読むことで構造異常の検出率が向上する
- 全ゲノムsequenceによる構造異常の検出は、標的シーケンスや全エキソンシーケンスでは検出されない構造異常を検出することができることがある

## ペアエンドリードによる構造異常の検出



Chr1とChr3の転座をペアエンドの短鎖シーケンスで読む場合、リード内部で切断点を持つ①②⑤⑥に加えて、両リードが異なる部位にマッピングされる③④のリードも転座をサポートするため、構造異常の検出率が向上する

## WGSにおける構造異常の検出



骨髄異形成患者  
骨髄検体

同一患者の正常対照  
検体（口腔スワブ）

→ RUNX1 遺伝子のエキソン3

この例では、RUNX1遺伝子のexon3を挟むintron領域に切断点をもつ欠失が生じている。標的シーケンスや全エキソンシーケンスでは、このような構造異常は検出できない

# 遺伝子転写の変化をきたす異常の検出-1

全ゲノムシーケンスによって非コード領域の変異を検出する意義を理解する

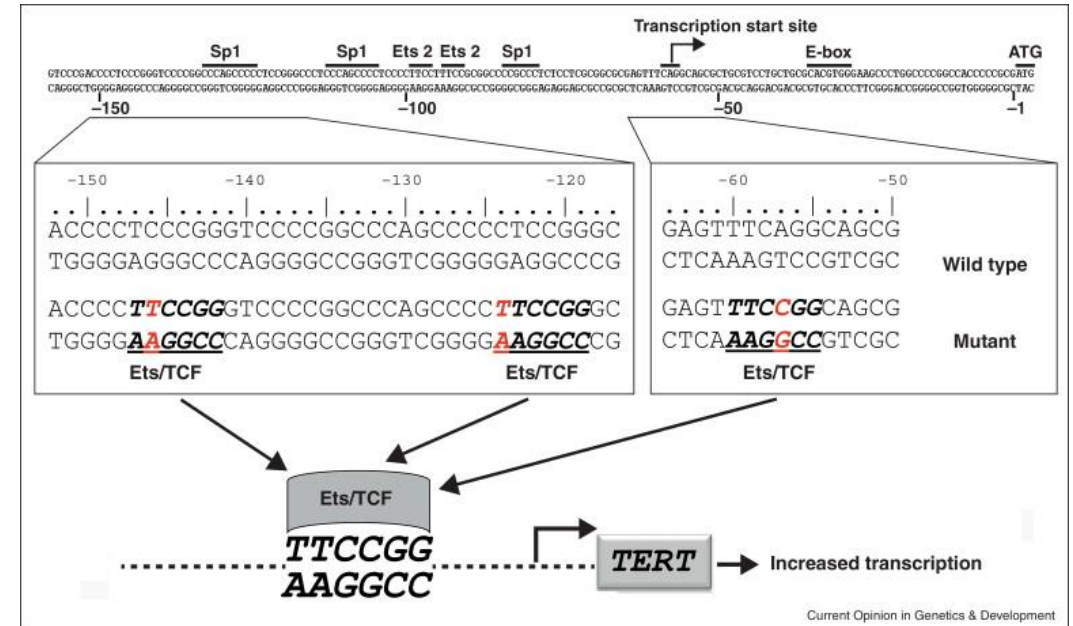
## ● 全ゲノムシーケンスによる非コード領域の変異

- ✓ 非コード領域の異常は、エンハンサ、プロモータ、3'-UTR、5'-UTR、転写因子の結合部位、lncRNA、smallRNAなどの、機能ドメイン単位で考慮するとよい
- ✓ 非コード領域の異常を解釈するには、splicingや転写調節に関与する幅広い生物学の知識が必要である
- ✓ がんにおける変異の集積、特定のがん種への偏在がみられることがあり、意義のある変異であることを示唆する
- ✓ できる限りRNA-seqを併用することで、ゲノムの変化が遺伝子発現量、isoformの変化などを生じているかを確認することが望ましい
- ✓ さらにメチル化アレイ、ChIP-seq、ATAC-seq等を併用することで、ゲノムの異常が疾患を引き起こすメカニズムを解明することができる
- ✓ 非コード領域の変化が遺伝子の構造や発現を変化させるメカニズムを知ることが重要である

## ● プロモータ領域の変異

- ✓ *TERT* 遺伝子のプロモータ領域の変異は多くのがん種で認められる
- ✓ 変異の結果、ETSという転写因子が結合できるようになりプロモータが活性化される

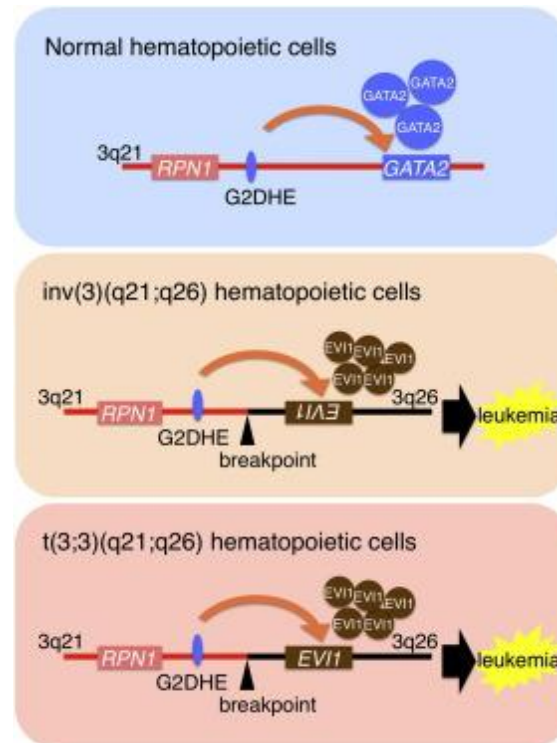
### *TERT* 遺伝子のプロモータ領域の変異



(Current Opinion in Genetics & Development 24, 2014, 30-37)

## ● エンハンサー置換を生じる構造異常

- ✓ t(3;3)(q21;q26)もしくはinv(3;3)(q21;q26)によるGATA2エンハンサーの異常
- ✓ 通常GATA2エンハンサ(G2DHE)はGATA2遺伝子の転写を活性化している
- ✓ 上記の構造異常によってG2DHEがEVI1遺伝子の活性を亢進すると同時に、GATA2の活性がなくなり、ヘテロ欠失をきたす
- ✓ その結果、白血病を発症する

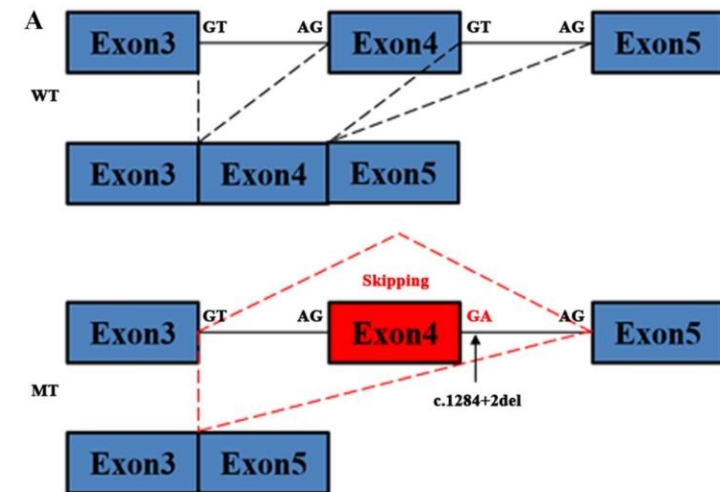


Cancer Cell,25(4), 2014, 415-427

## ● Splicing siteの変異

- ✓ **Splicing**はGT-AG配列の間で生じる
- ✓ この配列に変異があると、正しいsplicingが生じないため、splicing後の転写産物に異常が生じる
- ✓ RNA-seqで実際の転写産物の構造を確認することが必要である

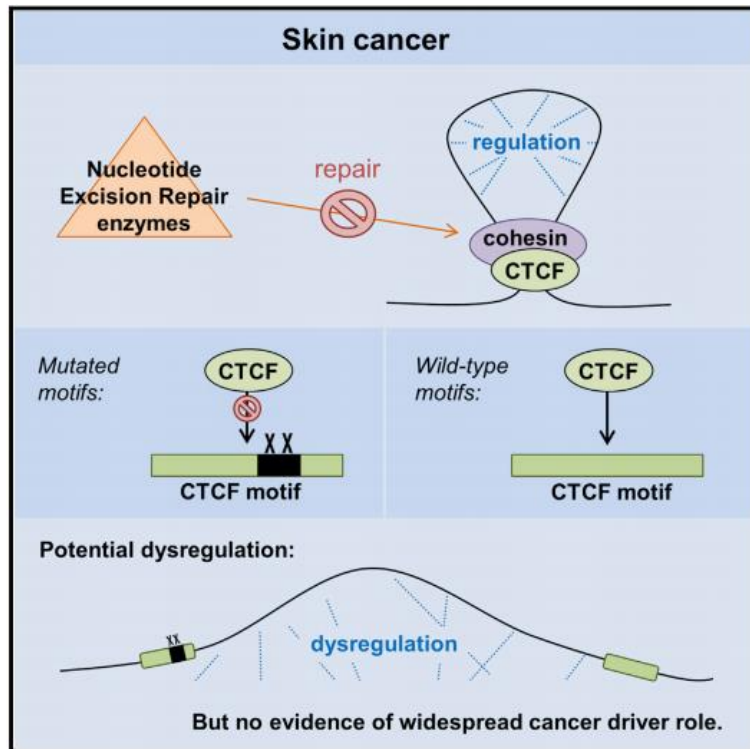
### Splicing部位の変異とsplicing異常



(International J of Oncol, 54(3), 859-68, 2019)

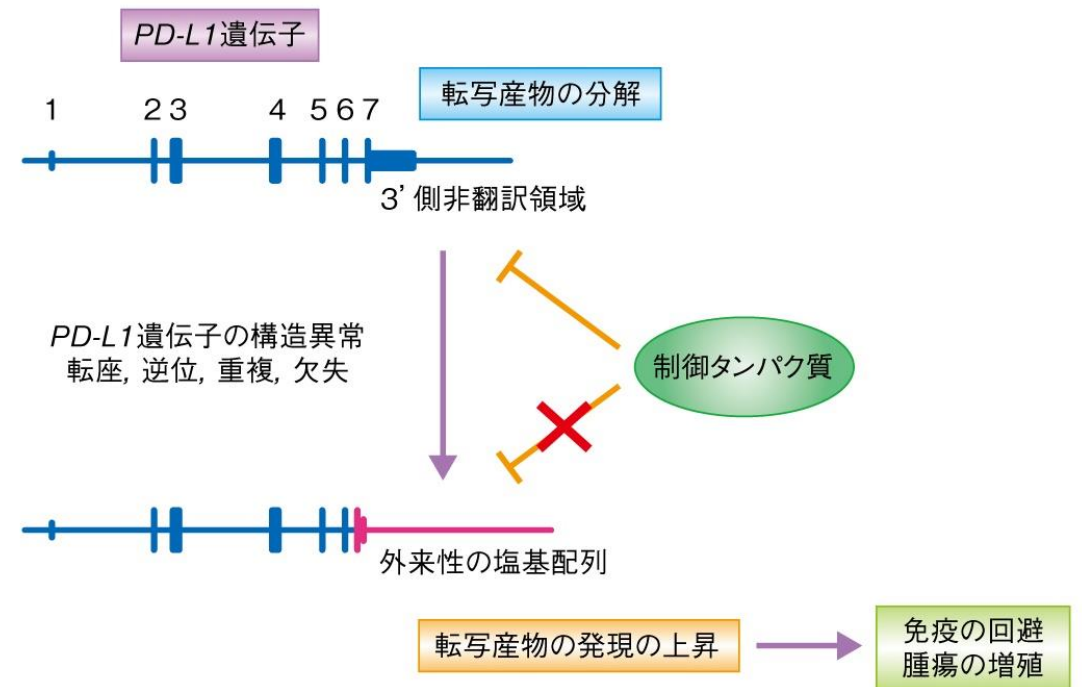
- ゲノム3次元構造を制御する部位の変異
  - ✓ CTCFは、ゲノムの三次元構造 (topologically associating domain/TAD) を規定する因子である
  - ✓ CTCF結合サイトの変異によって3次元構造が変化し、遺伝子発現の変化が生じる

## CTCF結合部位の変異によるゲノム3次元構造の変化



(Cell Reports 17, 2865-2872, 2016)

- RNA安定性
  - ✓ mRNAの3'-UTR は、しばしば転写後に遺伝子発現に影響を与える調節配列を含んでいる (調節タンパク質やmiRNAによる制御)
  - ✓ *PD-L1* 3' 非翻訳領域の構造異常による欠損は、RNAの制御蛋白の脱制御によって転写産物の異常な増加をきたす



©2016 片岡圭亮・小川誠司 Licensed under a Creative Commons 表示 2.1 日本 License



# コピー数変化の検出

## NGSによるコピー数解析の原理を理解する

### ● NGSを利用したコピー数変化の検出

- ✓ NGSでは、シーケンス深度を利用してコピー数の変化を検出する
- ✓ コピー数の変化部位には構造異常を伴う事が多い

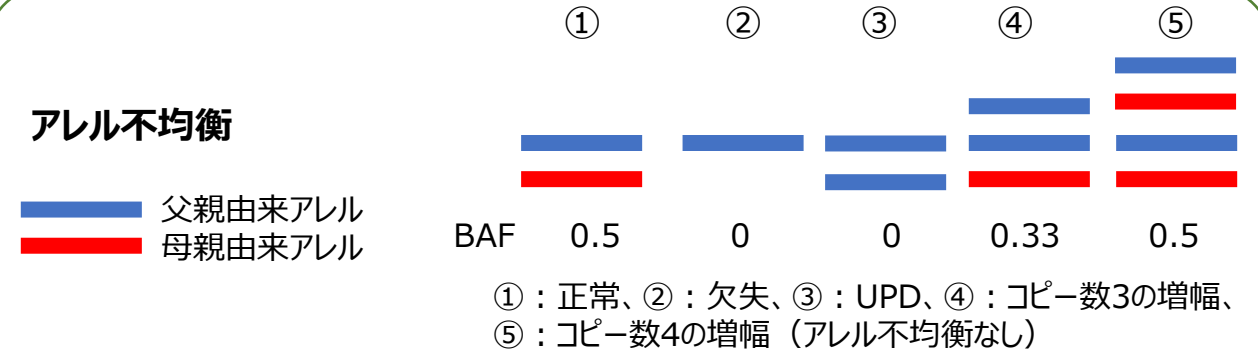
### ● アレル不均衡とその検出

- ✓ 常染色体と女性のX染色体は父親由来のアレルと母親由来のアレルが1本ずつ存在するため、正常な状態ではその構成比は1 : 1である
- ✓ コピー数の変化に伴い、アレルの構成比が1 : 1からずれることを**アレル不均衡**という
- ✓ 特に片方のアレルが欠失することをヘテロ接合性の喪失 (loss of heterozygosity: **LOH**) とよび、特に片親由来のアレルが2本に増幅し他方のアレルが欠失することを片親性ダイソミー (uniparental disomy: **UPD**) とよぶ。UPDは必然的にLOHを伴う
- ✓ LOHは、癌抑制遺伝子の正常アレルの欠失を通じてがん化過程によくみられる現象である
- ✓ NGSにおけるアレル不均衡の検出には、SNPsを利用することが一般的である

### ● 全ゲノム解析におけるコピー数変化の検出特有の問題点

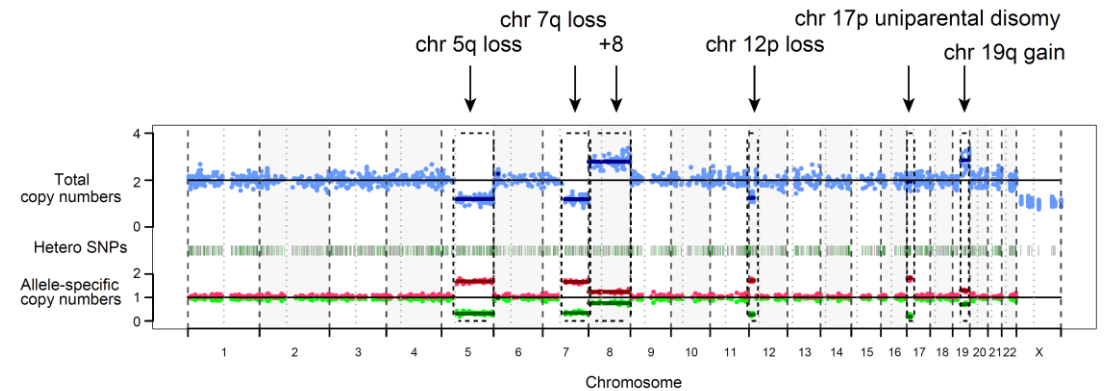
- ✓ 特に非コード領域や繰り返し配列には、個人差が大きいことに留意し、個人差が大きな場所を解析から除外するか、同一患者の正常対照検体と比較することが望ましい

### アレル不均衡



アレル不均衡はB-allele-frequency (BAF) という指標で示される。少ない方のアレル比で、0~0.5の値を取る

### NGSによるコピー数解析の例



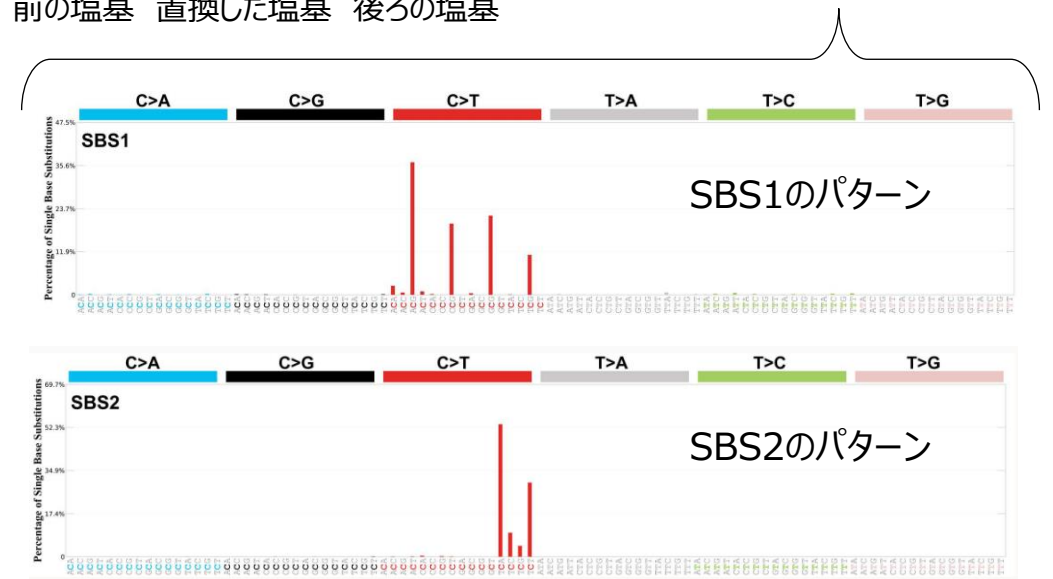
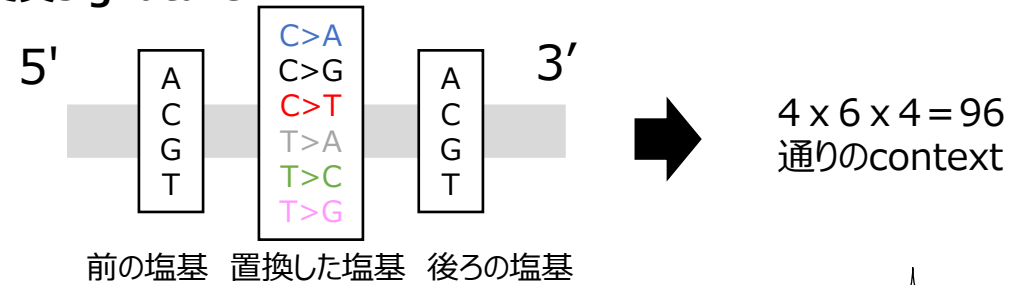
上段の青い線が、コピー数の変化を示す。下段の赤と緑の線が、アレル不均衡を示す (基線からの逸脱が不均衡の大きさを示す)。欠失はアレル不均衡を伴う。Chr17はコピー数が変化していないが、アレル不均衡を示し、UPDを生じていると考えられる。

# 変異signature-1

変異signatureは、単塩基置換型の体細胞変異のパターンを指す。COSMICによって94種類の変異パターンが提唱されている

- **1塩基置換 (Single base substitutions/SBS)** は、代表的な変異signatureであり、1塩基置換とその前後1塩基のパターンによって定義される
  - ✓ SBSとして、C>A、C>G、C>T、T>A、T>C、T>Gの6パターンがある。相補的な置換も含める。例えばG>TはC>Aに含まれる
  - ✓ 従って、前後の塩基を含めたSBSのcontextは、 $4 \times 6 \times 4 = 96$  通りある
- COSMIC データベースには、SBS signatureのパターンとしてSBS1~SBS94が登録されている (2021年8月時点)
  - ✓ 各サンプルのSBSのパターンを、これらの既知のCOSMICのSBS signatureが組み合わさったものとして考え、その構成比に分解することができる
  - ✓ ただし、WGSの変異データから、COSMICが提唱するSBS signatureパターンとは別の新たなsignatureパターンを抽出することもできる。その目的に、NMF (non-negative matrix factorization) など、クラスタリング手法が用いられる

## 変異signature



(<https://cancer.sanger.ac.uk/signatures/>)

変異signatureは、変異を生じる原因によって特徴的なsignatureを示すことから、変異signatureの解析は各腫瘍の変異を生じる原因を推定するのに用いられる

- COSMICのSBSパターンの中には、変異の原因が想定されているものがある
  - ✓ それぞれのSBSパターンにうち、変異の原因が特定されているものがあり、主なものを右の図にまとめる
  - ✓ 特に、SBS1はCpGジヌクレオチドのシトシンが脱アミノ化によってチミンに変化するパターンを示し、経時的に増加するため患者の年齢とも増加し、分子clockといわれる。正常細胞を含むすべての細胞に生じる変化であるがclone性の増殖を示す腫瘍ではNGSで検出される
  - ✓ COSMICのSBS signatureパターンの構成はがん腫によって特徴があり、そのがん腫の変異原を推定するのに使われる

変異signatureとその原因

SBSパターン	想定される原因
SBS1	Spontaneous deamination of 5-mC
SBS2	APOBEC deaminase
SBS4	tobacco
SBS6	Defective DNA mismatch repair
SBS7	UV light
SBS9	Polymerase $\eta$
SBS11	Temozolomide treatment
SBS13	APOBEC deaminase
SBS15	Defective DNA mismatch repair
SBS18	Reactive oxygen species

# 変異signature-3

変異signatureには、strand-biasを示すものがある。また、単塩基置換以外の二塩基置換やindelのパターンを分類したものが提唱されている

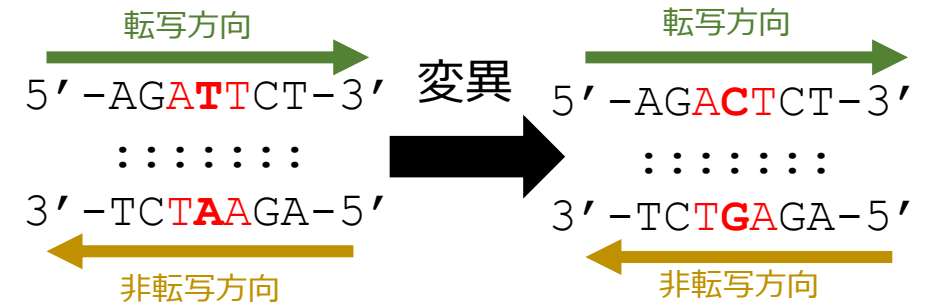
- Strand biasをもつsignatureがある

- ✓ 転写を受ける箇所 (intronを含む) のSBSについては、転写の方向が定まるため、相補的な変異パターンを区別した場合、転写方向によって変異パターンに偏りがみられることがあり、これを**strand bias**とよぶ
- ✓ Strand biasを伴う変異は、特定のSBSパターンに限定される (例 SBS16, )
- ✓ Strand biasを生じる原因として、transcription-coupled nucleotide excision repair や非転写鎖の過剰なDNA damage が想定されている

- 近年、変異signatureの新しい構成要素として、二塩基置換 (**Doublet base substitutions/DBS**) や短鎖挿入欠失 (**Small insertions and deletions/ID**) も提唱されている

- ✓ DBSは連続する2塩基が、異なる2塩基に置換するものを指す。例えばCCからTTへの置換を主な要素とするDBSは、DBS1とよばれ、紫外線の影響で生じる。また、CCから、主にAAに置換され、少数のAG、ATへの置換も含むものをDBS2とよび、喫煙の影響で生じる
- ✓ IDは1~50b.p. の短い塩基の欠失、挿入のパターンである。挿入、欠失の長さ、それらの変異が生じる場所 (繰り返し配列の長さなど) を要素とするcontextによって分類される

## Strand bias

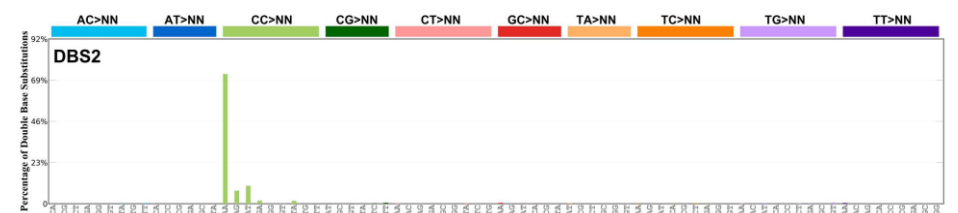


この変異は転写方向のA(T>C)Tと表現される。



この変異が転写方向と逆方向で生じる頻度を区別した場合に、転写方向と頻度の間に関連があることを、strand biasとよぶ

## DBS変異signatureの例



(<https://cancer.sanger.ac.uk/signatures/>)

# Catastrophic events

ゲノムに無数の変異が集中して生じる現象を**catastrophic event**とよぶ。全ゲノムシーケンスはこれらの検出に強力な役割を果たす

## ● Chromothripsis

- ✓ 1～数本の染色体が多数の断片に粉碎され、不完全な修復を受ける結果、欠失、重複などのコピー数異常と、あらゆる構造異常、無数の突然変異が組み込まれる現象をChromothripsisとよぶ

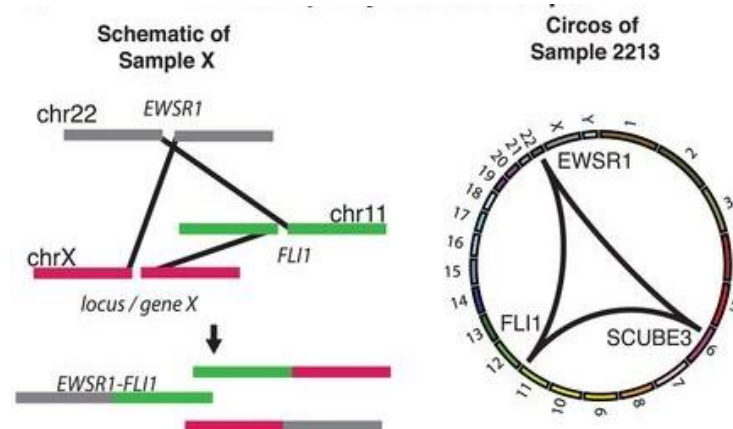
## ● Chromoplexy

- ✓ 複数の染色体の間でDNAの分裂と再構成が生じる現象をChromoplexyとよぶ
- ✓ 最初の記述は前立腺がんでおこなわれたが、ユーイング肉腫などにもみられる
- ✓ 機序は不明だが、核内の転写活性ハブに転写の盛んな遺伝子をコードするDNAが集まり、その場でDNAの断裂と再結合が生じると想定される

## ● Kataegis

- ✓ 遺伝子変異がゲノム上で局所的に高頻度に発生する現象をkataegisとよぶ
- ✓ APOBECなどの活性化誘導シチジンデアミナーゼの活性の変化が関与していると考えられており、乳がんを高頻度にみられる

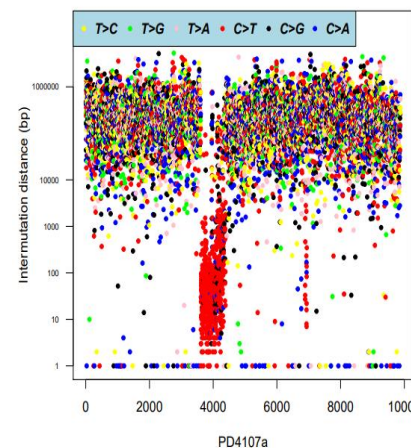
## Chromoplexyの例



ユーイング肉腫に生じた Chromoplexyの2例。左の例では第11, 第22, X染色体の間に、右の例で第6, 第11, X染色体の間に Chromoplexyが生じ、EWSR1-FLI1融合遺伝子が生じている。

(Science 31 Aug 2018:Vol. 361, Issue 6405)

## Kataegisの例



Rainfall plotは、kataegisを示すのに有効である。横軸に変異の場所を、縦軸に、隣の変異箇所との間隔 (b.p.) を示す。変異が集中している箇所は、縦軸の値が小さく表示される。色は塩基の変化を示す。この図にみられるように、乳がんが生じる kataegisは、C>Tの変異が多く、赤色で示される。

(Cell. 149 (5): 979-93)

非コード領域のシーケンスによって、テロメア長の測定や非ヒトゲノムの検出が可能である

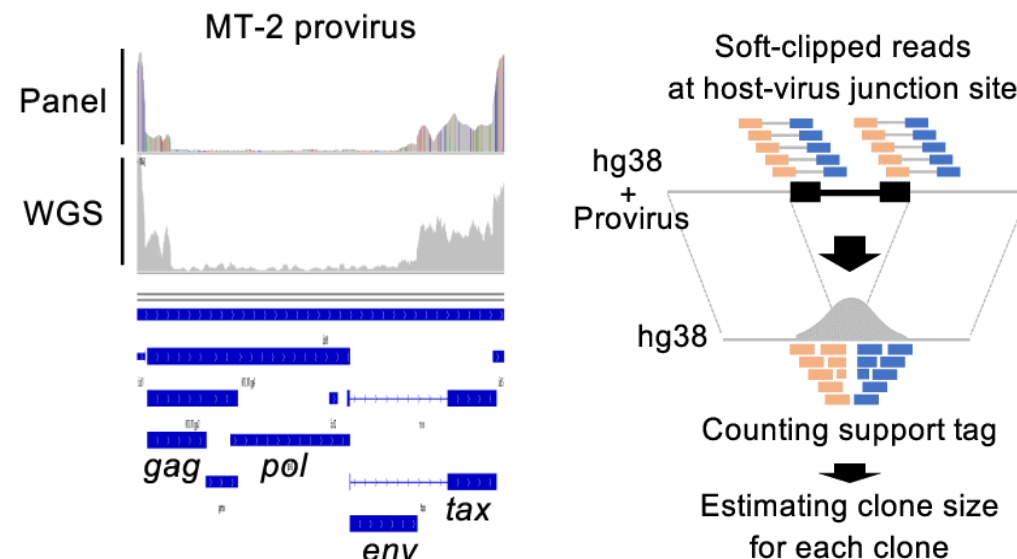
## ● テロメア長の測定

- ✓ テロメアは染色体末端を保護する構造で、哺乳類では (TTAGGG) の繰り返し配列と、テロメア構成タンパク質から構成される
- ✓ テロメアの繰り返し配列の長さは、細胞分裂に従って短縮するが、一部のがんではテロメアの伸長を司る酵素 (テロメラーゼ) の活性が亢進している
- ✓ 全ゲノムシーケンスのデータを用いて上記の繰り返し配列の長さを測定することで、テロメアの長さを測定することができる

## ● 非ヒトゲノムの混入と検出

- ✓ 全ゲノムシーケンスではシーケンスをおこなうゲノムの選択をおこなわないために、非ヒトゲノムの混入を認めることがある
- ✓ ヒトゲノムにマッピングされないリードを、非マップリードとして、抽出することができる
- ✓ 検体の処理の間に増殖した微生物ゲノムの混入を認めることがある。例えば、乾燥が不十分な口腔粘膜スワブに真菌が繁殖し、検出されることがあるため、スワブの十分な乾燥は重要である
- ✓ 病原体微生物の検出をおこなう目的で、目的とする微生物のゲノムシーケンスにマッピングをさせて、検出・定量をすることがある
- ✓ ウイルスゲノムがヒトゲノムに取り込まれることをgenomic integrationとよび、HBV,HPVやHTLV-1でみられる。クローン性の増殖を来す悪性腫瘍の場合、integration siteが同じクローンが増殖をしている

## HTLV-1ウイルスのマッピングとintegration siteの同定



HTLV-1ウイルスの塩基配列に対してマッピングをおこなうことで、HTLV-1ウイルスの定量と、配列の変化を観察することができる。ATL患者ではHost DNAの配列を調べることで、ホストゲノムのintegration部位が判明する。それを多くのリードに対しておこなうことで、integration部位の種類と割合を計算することができる。

(Nature communications volume 12, Article number: 4821 (2021), in press)

様々な遺伝子検査とその特徴を理解し、適した検査を選択できるようにすることが必要である

## ● CGH/CGHアレイ

- ✓ 腫瘍組織と対照組織から抽出したゲノムDNAを異なる蛍光色素で標識をしたのちにゲノム試料にハイブリダイゼーションさせて、蛍光色素強度を比較することで、コピー数変化を解析する方法である。現在はほぼ下記のSNPsアレイに置き換わっている
- ✓ SNPsアレイは全ゲノム関連解析を目的として、ヒトゲノムの数十～百万カ所のSNPsをタイピングするplatformとして開発されたが、シグナルを定量的に扱うことで、コピー数変化とアレル不均衡の検出ができるため、コピー数解析を目的として用いられることがある

## ● MLPA法

- ✓ 鋳型DNAにハイブリダイズしたプローブをPCRにより増幅することにより、対象ゲノム領域のコピー数変化を定量的に解析する方法である

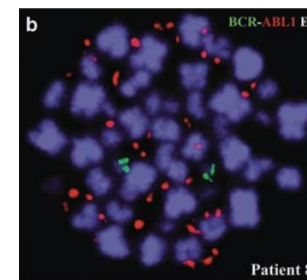
## ● 核型分析

- ✓ G分染法は、網羅的な核型異常を検出できるが、分解能、感度が低い
- ✓ FISH法はG分染法と比べ分解能と感度が高いが、目的とした異常の検出しかできない。SKY-FISH法を用いると、染色体を染め分けることができ、複雑な異常も検出できるようになる
- ✓ FISH法では染色体外の増幅を検出することができる（二重微小染色体）
- ✓ 染色体数の判定ができることは大きな利点である。例えば2倍体と4倍体の区別はSNPsアレイやNGSでは原理的に不可能であり、多倍体を呈するがんの解析では染色体数の確認が必要となることもある

## ゲノムのコピー数変化を目的としたゲノム検査

手法	特徴
SNPsアレイ	比較的安価、分解能は高い。アレル不均衡の検出が可能。
MLPA	特殊な装置が不要。感度は低い。特定の領域のコピー数変化の検出に向いている。
G分染	網羅的であるが、手間・時間がかかる。感度、分解能が低い。生細胞が必要で、分裂が盛んな細胞の異常が出やすい。
FISH	標識プローブの設計によって、転座や欠失も検出可能である。目的とした異常しか検出できない。
WGS	網羅的でアレル不均衡の検出も可能。感度はシーケンス深度に依存。領域によっては精度が得られない。高価。

## FISH法による染色体外増幅の同定



*NUP214-ABL1* 融合遺伝子の染色体外増幅

*BCR-ABL1* のプローブを用いて、*ABL1* が染色体外で増幅していることを示した

(Leukemia (2009) 23, 125–133)

## ● PCR法

- ✓ PCRはコントロールされた系を用いると感度が高く安価であるが、目的とした遺伝子異常しか検出できない
- ✓ 融合遺伝子を用いた微小残存病変（MRD）の評価には、逆転写-定量PCR（RT-qPCR）が高感度で、定量的な評価が可能な点から、臨床では頻用される
- ✓ Droplet digital PCR（ddPCR）は希釈した検体を多くのwellに分けてPCRをかけ、PCR産物が得られたwellの数をカウントすることで、定量性を持たせた方法である。PCR標的による増幅特性の違いを吸収することができる。一般にNGSより感度が高い

## ● Sanger シークエンス

- ✓ 安価であるが、限られた変異の解析しかできない点と、感度が低い点から、目的を限定する必要がある
- ✓ 欠失、挿入に弱く、これらの変異があるとそれ以降の波形が乱れる

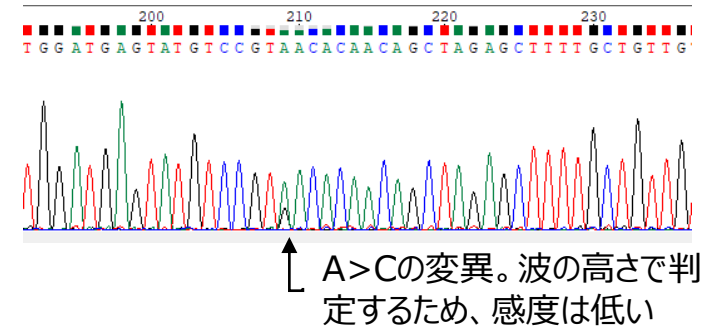
## ● 全エクソンシーケンス

- ✓ 全coding領域を対象とする。UTRやmiRNAを含めることもある

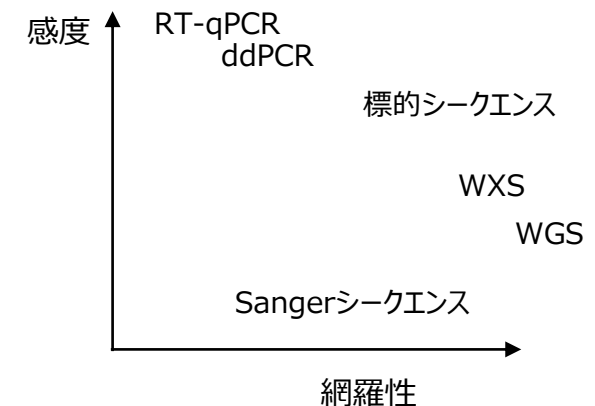
## ● 標的シーケンス

- ✓ 検査領域を絞ることで、コストを下げることができ、データの扱いも容易となる。シーケンス深度を確保しやすく、検出感度が高くなる。設計次第でnon-coding領域を含めることができる

## Sangerシーケンスの例



## 変異検出に対する各検査の位置づけ





# 第Ⅳ章

# がんゲノムデータの1次解析

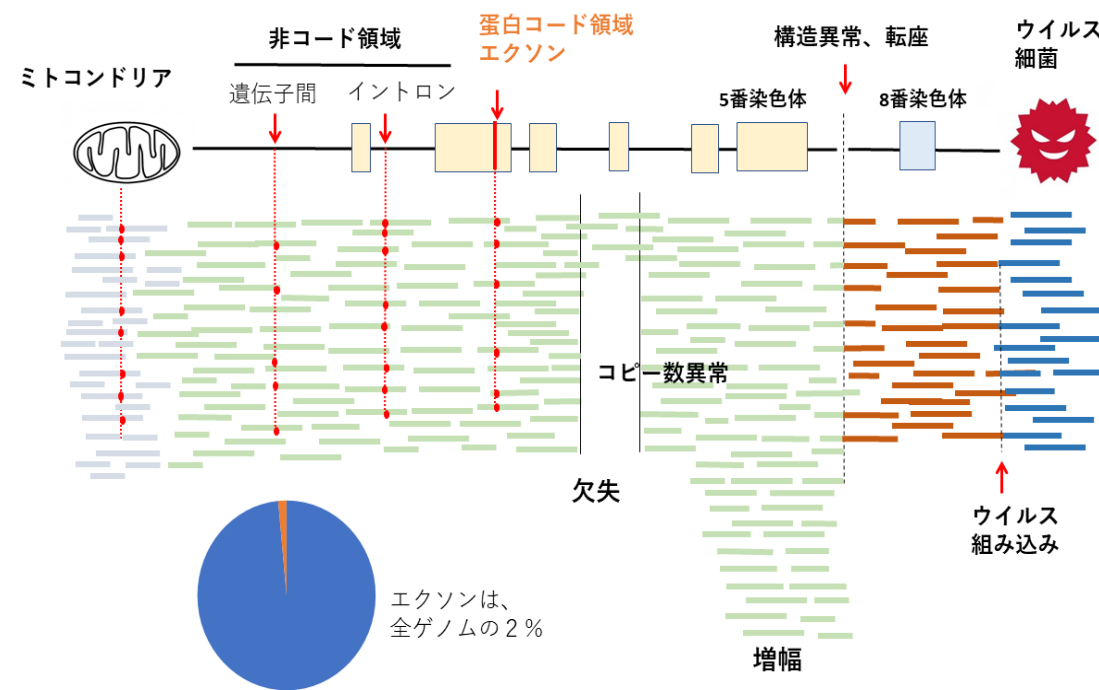
理化学研究所 生命医科学研究センター チームリーダー

中川英刀

# がん全ゲノムシーケンスでわかること

## がんは、ゲノム不安定の疾患である

- がんは、**ゲノム不安定性**を基本に、さまざまな変異がゲノムに蓄積して発生し、さらなるゲノム変異の蓄積によって進化するものである。NGSというゲノム解析技術の高度化と大規模化に伴い、がんゲノム医療にて数百の遺伝子のタンパクコード領域での変異同定が診断として行われるようになった。また、難病領域においては、**全ゲノム配列**（約30億塩基対）のうち、約2%を占めるエクソンの配列の決定も、診断としておこなわれるようになってきている。がんゲノムの場合、がん腫や個々のがんによって変異部位は異なるが、全ゲノム上で数千から数十万か所の様々なパターンの**体細胞変異**が蓄積しており、それらをすべて把握することによって、個々のがんの生物学的、進化論的、そして臨床的な理解が深まり、治療法選択の幅が広がる
- ゲノム医療で使われているターゲットシーケンスとの違い  
**ターゲットシーケンス**では、PCRやhybridizationによって、100-300遺伝子のエクソン領域に標的をしぼり濃縮して（～1Mb）、deepに（500x以上）シーケンスを行い、エクソン内のSNV/indelの体細胞変異を検出する。一方、**全ゲノムシーケンス（WGS）**では、がんおよび正常臓器（多くは、血液）から特定の標的にしぼって濃縮せずに、すべてシーケンスを行うことになる。その結果、がん遺伝子のエクソン以外に、以下の領域の変異を解析することになる（右図）
  - ✓ 全遺伝子のエクソンの変異（タンパクコード領域、5'、3'非コード領域）
  - ✓ イントロン、プロモーターなどの発現制御領域の変異
  - ✓ コピー数異常（CNV）
  - ✓ 染色体間転座を含む構造異常（SV）
  - ✓ ミトコンドリアゲノムの変異
  - ✓ 外来性、内在性ウイルスの組み込み、細菌のゲノム解析



- WGSでは、がんDNAは50～150x、正常DNAは30x（全ゲノム3Gbに対して）のdepth/coverageでシーケンスがおこなわれるが、解析が難しい領域を含む極めて広いゲノム領域を浅いdepthにてシーケンス解析を行うため、ターゲットシーケンスと比べて（500x～）、変異検出の精度が低く、エラーが多く含まれる可能性がある

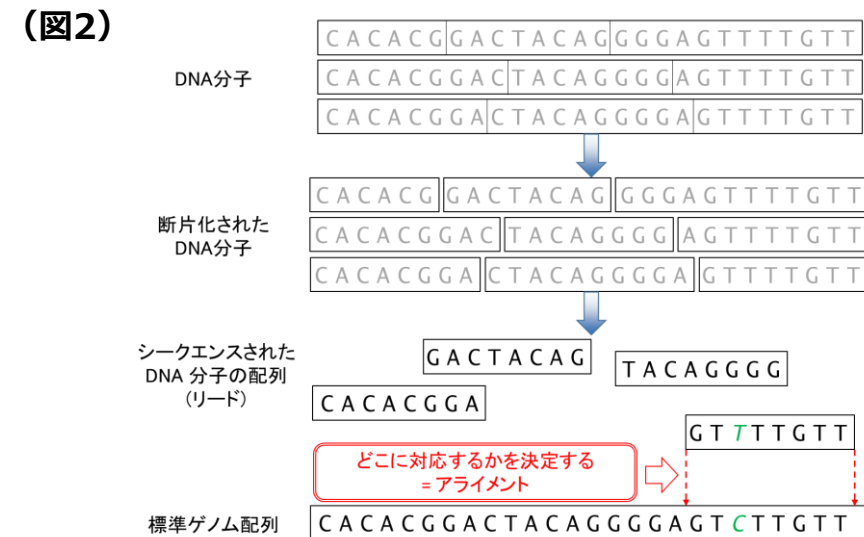
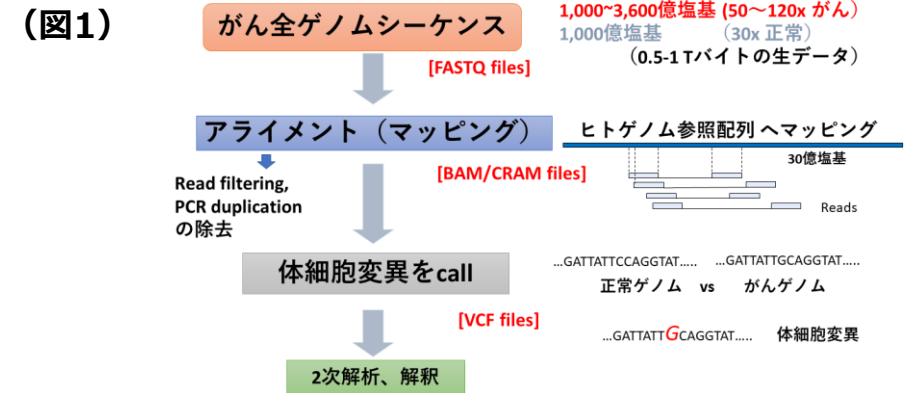
- 下記は、変異等の検出（一次解析）までと、その結果を利用した発展的解析へのワークフローを示す



# がんWGS解析のワークフロー 1

## がん全ゲノムシーケンス（WGS）解析の基本ワークフロー 1：アライメントまたはマッピング

- がん組織と正常組織（多くは、血液）のDNAを、ランダムに分断して、NGSライブラリーを作成し、シーケンスを行う。産出されたシーケンスデータは、100-150bpの長さのATCGの配列データであり（**short read**）、それぞれの塩基決定のクオリティスコアも付加されたファイルである（**Fastqファイル**）。WGSにおけるがんDNAの情報は、通常50x~150xの情報量であり、正常DNAは30xの情報量である（数百ギガバイト）。変異コールの精度は、腫瘍の**purity**（腫瘍細胞の割合やがん細胞の均一度）と、シーケンスの**depth/coverage**に依存するところが多い
- 解析の最初のステップにて、まず、これら短い配列のreadを、**ヒトゲノム参照配列**に対して**アライメント**（または**マッピング**）を行う（図1）。ヒト参照配列については、これまではGRCh37（hg19）を用いてきたが、最新のバージョンであるGRCh38に対してアライメントを行うことが多くなってきている。アライメントのアルゴリズム/ソフトは、高速かつ正確なものが求められ、現在は**bwa-mem**を用いることが多い
- **アライメント**（または**マッピング**）は、約30億塩基からなるヒトゲノム参照配列と、個々の100-150bpのリード配列のATGC文字の並びを比較して、最も一致する位置を決める作業である（図2）。数億個以上のreadのアライメント作業のため、コンピューターの負荷が最もかかる。また、100-150bpで、相同性の高い配列（似た配列）は、ヒトゲノム上には多数あるため、アライメント（マッピング）エラーも起きやすく、これが後述の変異コールのエラー誘発につながる
- アライメント作業の後に、PCR duplication（全く同じread配列）の除去や質の悪いreadの除去などのフィルタリングも行う。たとえば通常のWGSのデータのPCR duplicationは、~10%であり、これが高いとデータやDNAの質の不良が疑われる。

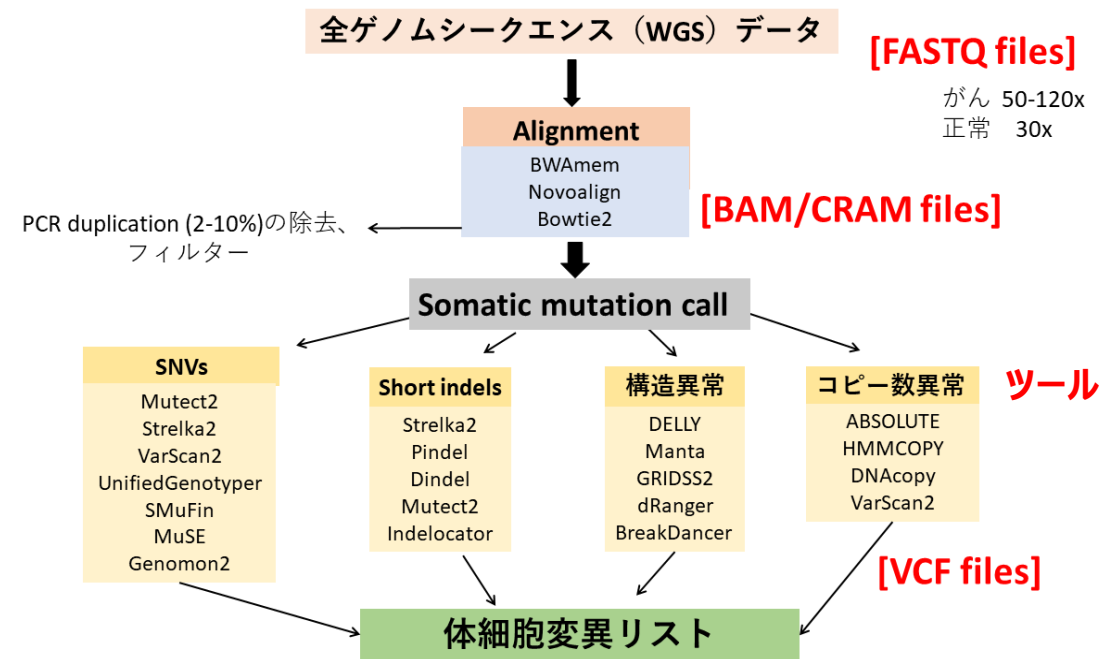


(加藤護、「最新がん個別化医療」、癌と化学療法、2016より)

# がんWGS解析のワークフロー2

## がん全ゲノムシーケンス解析のワークフロー2：体細胞変異コール

- 参照配列にアライメントされたreadの情報は、**BAMファイル (Binary Alignment Map)** として作成される。CRAM files (compressed columnar) は、BAM ファイルを圧縮したファイルである
- 次にこのBAMファイルから、体細胞変異の検出を行うが（一次解析）、様々な検出ソフトウェア（コンピュータ・ツール、caller）が変異のタイプに応じて開発されてきている
  - ✓ **SNV (single nucleotide variant)**：1塩基の塩基置換
  - ✓ **Indel (short insertion/deletion)**：1～数十塩基の挿入/欠失
  - ✓ **CNV (copy number variants)**：本来2コピーあるゲノムにおいて、数百kb～染色体レベルの領域のコピー数の変化
  - ✓ **SV (structural variants)**：構造異常
- 変異検出までのソフトウェアやコマンドの組合わせを**パイプライン**という
- 体細胞SNV callは、がん組織と正常組織とのread数の比較によって行われる。それぞれのパイプラインによって癖があり、SNV callの感度、特異度を特徴づけている。統計的手法やフィルターの違いが大きい。詳細については、次に述べる
- 同定された個々の変異は、多くの場合**VCF ファイル**として生成されて記載されて、次の2次解析や解釈を行う。その記載情報を右に記す

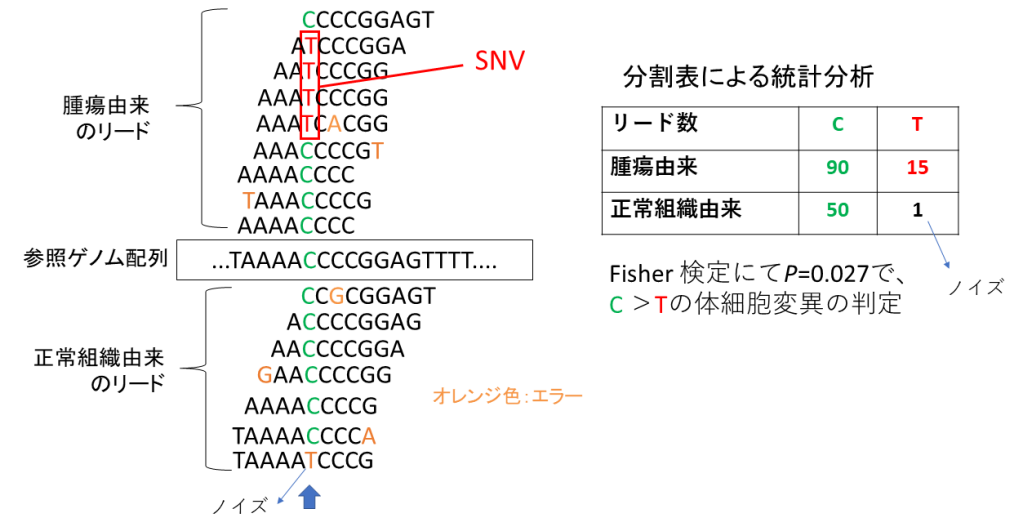


1. CHROM：染色体番号
2. POS：塩基の位置
3. ID：塩基（変異）のID。該当変異が dbSNP に登録されていれば、rs から始まるアクセッション番号を記載
4. REF：ゲノム参照配列上の塩基
5. ALT：シーケンシングデータ上の塩基
6. QUAL：シーケンシングデータのクオリティスコア
7. FILTERの有無など

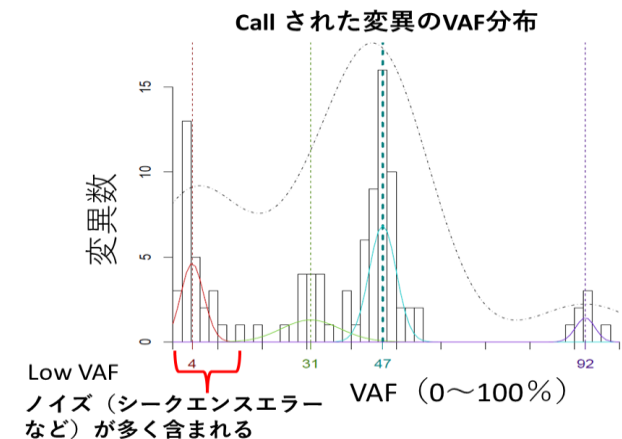
# 体細胞変異SNV Caller

NGSの特性を理解したうえで、ノイズの除去と体細胞変異SNV callを行う

- 体細胞SNV callは、基本的にがん組織と正常組織とのread数の比較によって行われる。それぞれのパイプラインによって、統計処理やフィルターの手法が異なり、SNV callの感度、特異度を特徴づけている  
以下のツールがよく使われている：Mutect2、Strelka2、VarScan2、Genomon2等
- 概念図を右に示す。参照ゲノム配列にアライメントされたリードにおいて、緑Cの位置のC（正常）とT（変異またはエラー）のリード数をカウントして、正常と腫瘍での数を統計的に比較し、有意に変異リードが多いか統計値を計算する（図では、分割表でのFisher検定）。この統計処理の手法はツールによって異なる
- 統計処理の前後にフィルターをかけて、ノイズやエラーを除去するが、このフィルターのかけ方は、パイプラインによって様々である。フィルターの例を以下に示す
  - ✓ VAF値（アライメントされたreadの内の変異readの割合）が低い変異はエラーが多い、という経験則があり、右図のように、低VAFの変異を統計的手法やリードの質の情報を用いて、noise（エラー）と変異とを分離する
  - ✓ 全ゲノムを30x~150xの平均depthでシーケンスしても、低いdepthの領域があり（特にGCが多い領域）、実際depthが低い領域は変異コールのエラーが多い。Read数が閾値よりも低いSNVは除くことがある
  - ✓ 近くに繰り返し配列やindelがある領域は、マッピングエラーが多いので、要注意
  - ✓ 正常組織由来の複数のサンプルにおいて、低VAF値でコールされる変異は、シーケンスエラーやマッピングエラーの可能性が高いので、除去する



(加藤護、「最新がん個別化医療」、癌と化学療法、2016より改変)

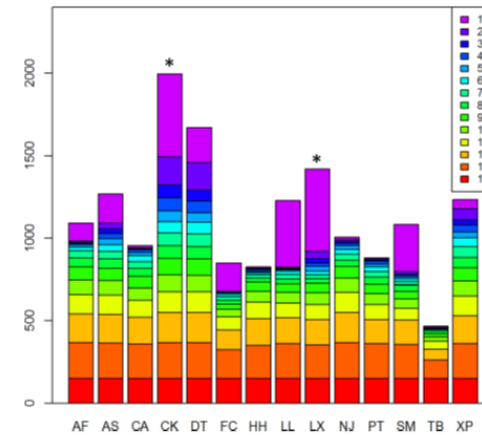


# 体細胞変異SNV callの精度

体細胞変異の検出は、パイプラインごとによって異なるので、精度評価は重要

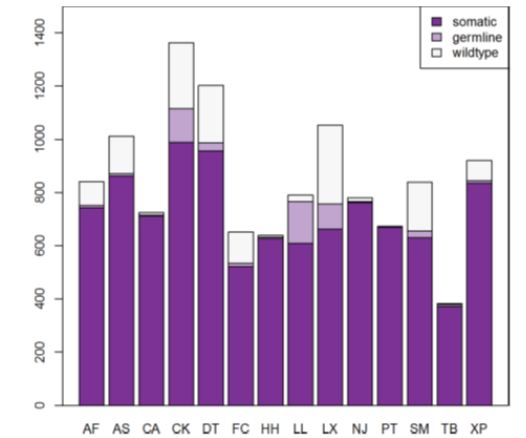
- WGSは、臨床で使われるターゲットシーケンスと比べて（500x~）、広い領域を低いdepth/coverageで解析するため、変異コールの精度が低い
- エクソン外のゲノム領域（全ゲノムの98%）は、繰り返し配列が多く、アライメントがうまくいかず、間違っただ変異コールをしたり、変異コールができないことも多い
- したがって、解析パイプラインを構築した場合、他のアルゴリズムとの比較や正解データセットとの比較によって、その精度を把握しておく必要があり、それによって、構築したパイプラインの特性を熟知しておく必要がある
- 世界の14か所で開発された変異コールについて、同じがん全ゲノムシーケンスのデータを用いて解析し、全ゲノム上での変異を同定した。それぞれのパイプラインで同定された変異数は、大きく異なることが分かる。赤の変異は、共通して同定された変異の数（図左）
- 右図の濃い棒は、全ゲノムシーケンスした元のDNAから、左図にて同定されたほとんどすべての変異について、別のNGS解析にて再解析を行い、実際に確認された変異の数を示す。薄い棒は、体細胞変異ではなく、生殖細胞系のvariantsであったものを、白は、確認できなかった変異の数を示す（図右）
- パイプライン共通でなかったり、間違いであった変異の多くは、depth/coverageが低い領域、リピート配列かその周辺の領域（アライメントエラーが起きやすい）にあるものであった

14個の解析パイプラインでcallしたSNVの数



世界中のゲノム解析センターのpipeline

確認できたSNVの数（正解数）



共通してcallされた変異

(Tyler S, et al. A comprehensive assessment of somatic mutation detection in cancer using whole genome sequencing. *Nat Commun* 6: 10001 (2015))

# Short Indel解析

## NGSの特性上、Indel やリピート配列の解析はエラーが多い

- PCRやNGS解析はDNAポリメラーゼの反応であり、そのエラーはDNAポリメラーゼの特性と一致する。すなわち、繰り返し配列には、エラーが入りやすく、NGSの解析においても、シーケンスエラーが生じており、繰り返し配列のindel（例えば、AAAAA > AAAAA）の検出は難しい
- 数塩基以上のindel があるread は、参照配列へのアライメントに間違いがはいることがあり、indel 周辺にエラーが入りやすい。右図のように、AAATの挿入変異がある場合、参照配列に対してアライメントをすると、アライメントの間違いが起きやすくなり、挿入変異の検出ができなかったり、周辺に別の変異があるとコールされることがある（図1）
- Indel変異に特化したソフトウェアは、PINDEL, Strelka2, MUTECT2などが使われる。Indel 周辺のアライメントエラーを減らすために、local realignmentやassemblyを行って、前述のアライメントエラーを解消する
- 右に記している左下の図は、がん全ゲノムシーケンスの同じデータに対し、14個の異なる解析パイプラインで同定した全ゲノム上での変異。それぞれのパイプラインで同定された変異数は、大きく異なることが分かる。赤の変異は、共通して同定された変異の数（図2左）
- 右下の図の濃い棒は、全ゲノムシーケンスした元のDNAから、左図にて同定されたほとんどすべての変異について、別のNGS解析にて再解析を行い、実際に確認された変異の数を示す。薄い棒は、体細胞変異ではなく、生殖細胞系のvariants であったものを、白は、確認できなかった変異の数を示す（図2右）

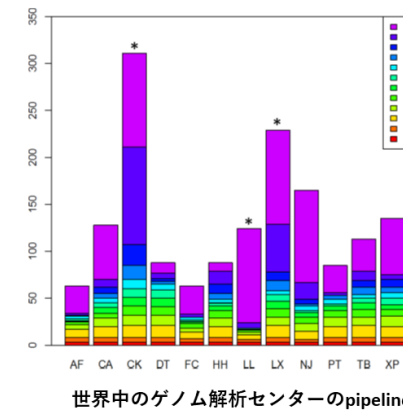
(図1)

参照配列	TTTTAGGTCATAAAC ---- AATTAAGTCAACA	
がんゲノム	TTTTAGGTCATAAACAAATAATTAAGTCAACA	(AAAT の挿入)
リード 1	TTTTAGGTCATAAAC ---- AAAT	} アライメント間違い
リード 2	TTTTAGGTCATAAAC ---- AAATAATT	
リード 3	TTTTAGGTCATAAACAAATAATTAAGTCAACA	アライメント正しい
リード 4	CAAAAC ---- AATTAAGTCAACA	} アライメント間違い
リード 5	AT ---- AATTAAGTCAACA	
リード 6	T ---- AATTAAGTCAACA	

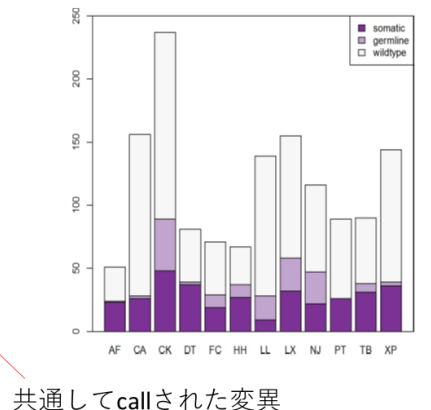
オレンジは、ミスマッチ（アライメントではある程度のミスマッチは許容）

(図2)

14個の解析パイプラインで callしたindelの数



確認できたindelの数（正解数）



(Tyler S, et al. A comprehensive assessment of somatic mutation detection in cancer using whole genome sequencing. *Nat Commun* 6: 10001 (2015))

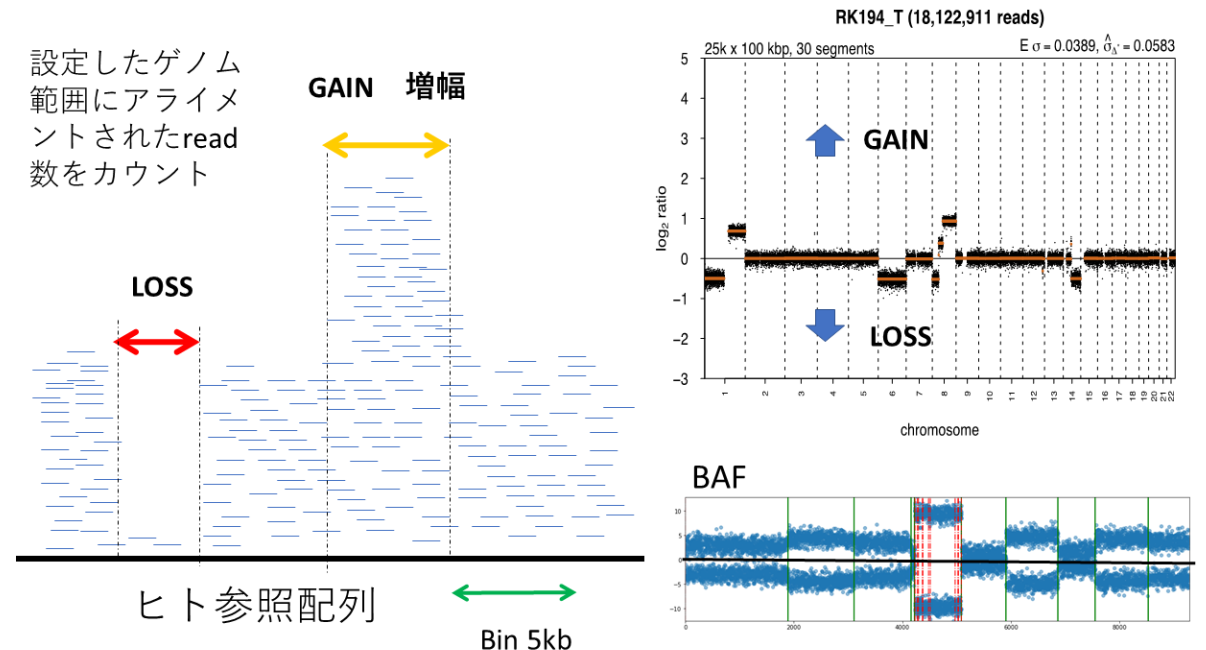


# コピー数異常 (CNV) 解析

CNVは、高い頻度でがんゲノム解析で検出され、機能的にも重要である。

- がんゲノムで、SNV変異に次いで頻度が高く、かつ重要な変異は、**コピー数異常 (CNV/Copy Number Variation)** である。がんゲノム研究の早期において、コピー数異常が集積する領域から多くのがん関連遺伝子が同定されてきた
- ヒトゲノムの常染色体は、2コピー (父方と母方由来) 存在する。がんゲノムは、様々な領域、または染色体の全体にわたって、コピー数が1コピー (**Loss**または**LOH**)、0コピー (**homozygous deletion**) に減ることがある。その領域には、腫瘍抑制遺伝子を含むことが多い
- 3コピー、4コピー～数十コピーにまで増える (**Gain**) 場合もある。比較的小さい領域 (100bp～数MBbp) が数十コピーにまで増える現象を局所増幅 (**focal amplification**) といい、その領域にはHer2、MYC、EGFRなど、がん遺伝子が含まれることが多い
- 参照配列にアライメントされたread数を、数kb～100kbの幅にてカウントし、他の領域や正常組織における当該領域のカウントと比較して、多ければGain、少なければLossと判定する。
- SNPチップでの解析と同様に、**BAF (B-allele frequency)** の情報を用いた**LOH (loss of heterozygous)** 解析もCNVの検出感度が高い。生殖細胞系SNPでヘテロ (A/B) の部位のread数の不均衡を反映するものであり、正常では同じ比率で0.5であるが、どちらかのアレルが欠損したり、増幅するとその比が0.5からずれる (図右下)

(図) CNV検出手法 (左) と例 (右)



- **腫瘍purity** : 臨床サンプルには、多くの非腫瘍細胞 (間質、免疫細胞など) が含まれており、膀胱がんなどでは、腫瘍細胞の割合は5%を切るものが多く、purityの低いサンプルの解析は一般的には難しい。CNVは、ほぼすべてのがん細胞で起きている現象なので、このCNVの情報を利用して、DNAレベルでのサンプルの腫瘍purityを推定することができる

# 構造異常 (SV) 解析

## WGSで構造異常を効率的に検出できる

- 構造異常 (大きな欠失、挿入、逆位、転座) は、WGSによって、効率よく検出できるようになった。しかし、現在のshort-read sequencing 技術では、検出精度は低いことがあり、long-read sequencing 技術の発展が待たれる
- 慢性骨髄性白血病の原因であるフィラデルフィア染色体 (9番と22番染色体間の転座によって、BCR-ABL融合遺伝子が発現する (図1) )。ALK遺伝子についても、2番染色体短腕の逆位によって、EML4-ALK融合遺伝子が発現する

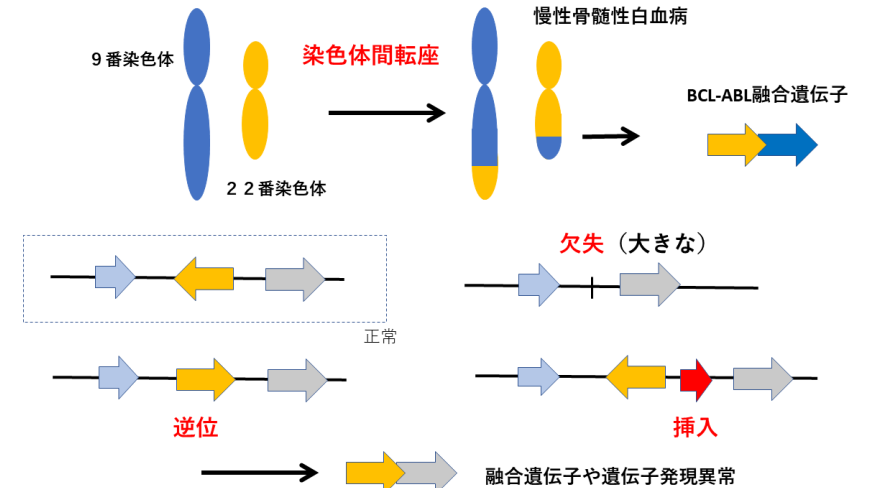
### ● 構造異常の種類 (図1)

- ✓ 大きな欠失 (deletion)
- ✓ 逆位 (inversion)
- ✓ 挿入 (insertion)
- ✓ 転座 (translocation) 染色体間、染色体内
- ✓ 複雑な構造異常 chromothripsisなど

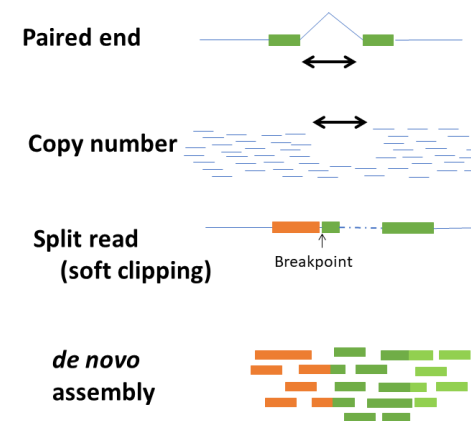
### ● Short-read sequencing による構造異常の検出手法 (図2)

- ✓ **Pair-end**の情報を利用して、その間の距離から、欠失や挿入部位や大きさをみる。Read の向きを考慮することによって、逆位や重複 (同じ領域がすぐ横に挿入)、転座の状態を決めることもできる (図3)
- ✓ コピー数異常の情報
- ✓ **Soft clipping** (150bpのreadの一部が正確にアライメント) によって見つかったsplit read。構造異常のbreakpointを直接検出する
- ✓ ヒト参照配列にアライメントせずに、SVの候補read をアセンブルして、構造異常のbreakpoint 付近の配列を構築する

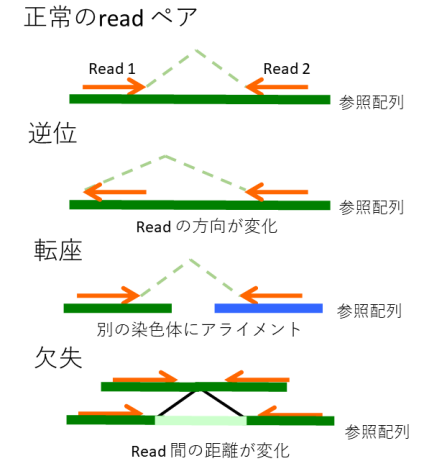
(図1) 構造異常の種類を示す図



(図2) 構造異常の検出方法の例



(図3) Paired end解析の例

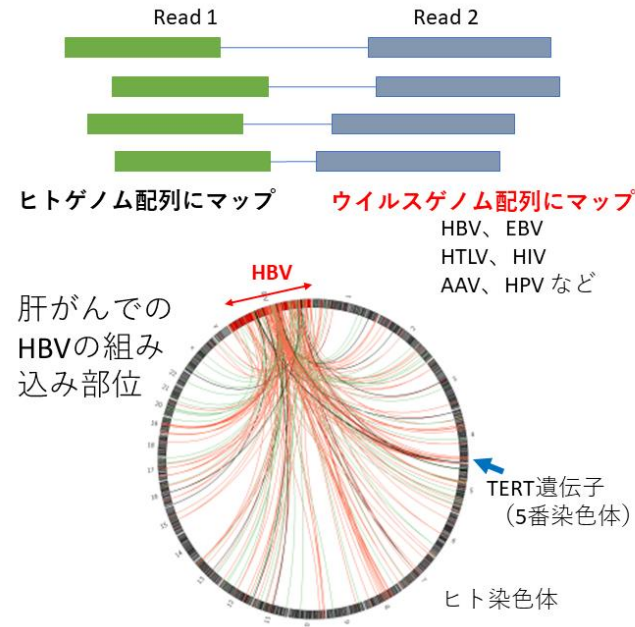


# ウイルスゲノム+レトロトランスポゾン

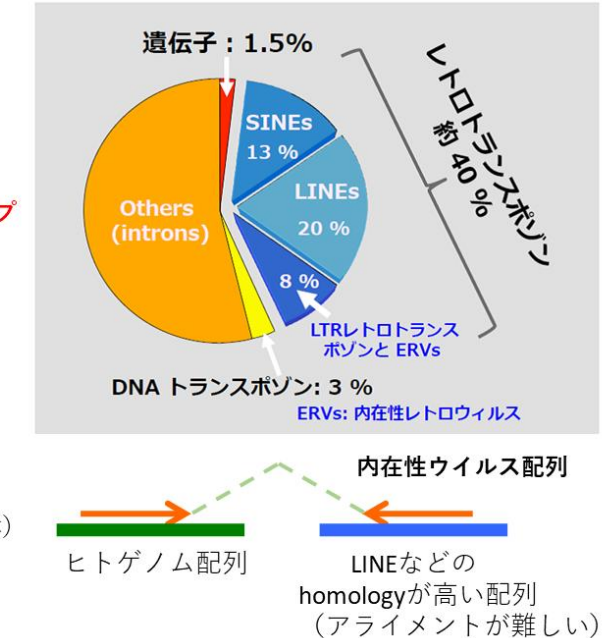
WGS解析により、ヒトゲノム以外の配列（ウイルスや細菌）も検出可能

- ヒト肝炎ウイルス（HBVとHCV）、ヒトパピローマウイルス（HPV）、EBウイルス（EBV）、HTLV-1、HIVなど、さまざまウイルス感染からの発がんが確認されており、一部のウイルスゲノムは、ヒトゲノムへの組み込みが報告されている
- HBV, AAV, レトロウイルス（HIV, HTLV-1）は、ヒトゲノムに高い頻度で組み込みを起こし、ゲノム不安定性やがん関連遺伝子の発現異常につながることもある
- バイアスのない形で、ゲノムシーケンスを行うWGSによって、さまざまなヒトゲノム以外の微生物由来のread（ウイルス、細菌、寄生虫）が検出される
- 構造異常の解析と同様に、pair-endの情報やsplit readの情報にて（ヒトゲノムとウイルスゲノムのキメラ）、ヒトゲノムへの組み込む部位をピンポイントで同定することができる（図1上）
- HBV感染によって発症する肝がんにおいては、多数のHBV組み込み部位が同定されており、5番染色体のTERT遺伝子領域にHBVの組み込みが多く観察される（図1下）
- ヒトゲノム配列（約30億塩基対）の約40%は、**内因性レトロウイルス**やそれ由来の配列（LINE、SINEなどの繰り返し配列）である（図2上）
- **内因性レトロトランスポゾン**は、自身のDNA配列から転写されたRNA配列を、逆転写反応でDNAにコピーし、このコピーをゲノムの別の場所に挿入することで転移する。がんにおいても、このレトロトランスポゾン（内因性レトロウイルス）によって、構造異常などの変異が起こり、がん関連遺伝子の発現にも関与することがある

（図1）外来性ウイルスの組み込み



（図2）内在性ウイルスの組み込み



- Pair-endの情報やsplit readの情報にて、レトロトランスポゾン（内因性レトロウイルスの組み込み）を、専用に構築されたアルゴリズムによって検出することができるが、ホモロジーが高い（似た配列が多い）ので、アライメントの作業を緻密に行うことが重要である（図2下）

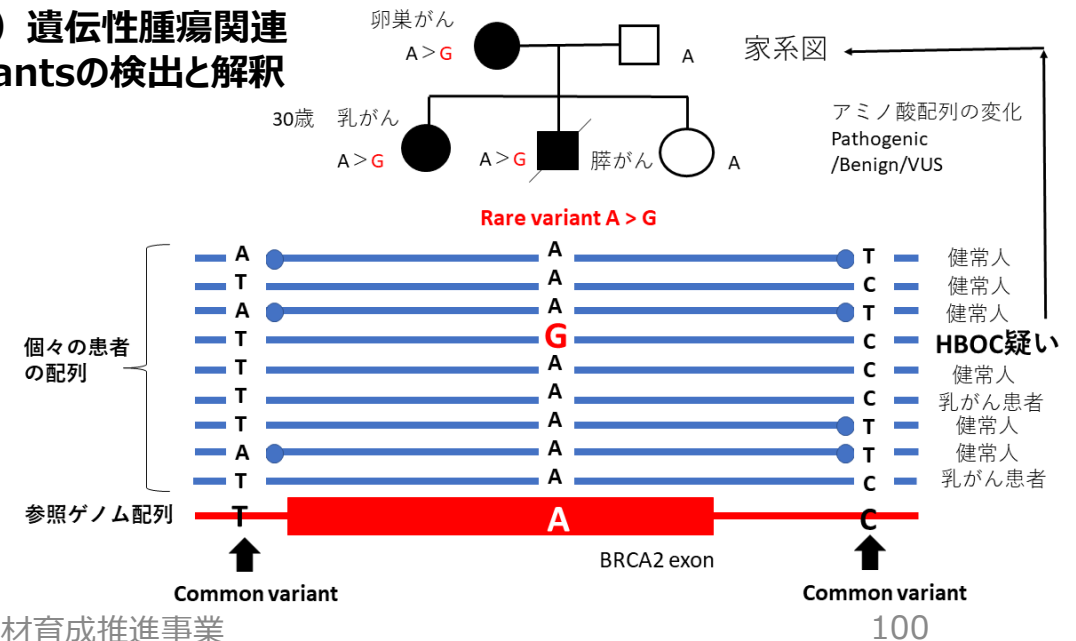
# 生殖細胞系列バリエーション

## 生殖細胞系列 (germline) 変異の解析は、遺伝性腫瘍の診断と治療に直結する

- 一般的ながんにおいても、生殖細胞系変異に起因するがん（**遺伝性腫瘍**）が1-5%ほど含まれており、その検出と解釈は臨床上重要である。*BRCA1/2*の生殖細胞系変異のあるがん（**HBOC**(遺伝性乳がん卵巣がん症候群)）に対してPARP1阻害剤、リンチ症候群/MSI陽性症例に対して免疫チェックポイント阻害剤での治療ができるようになり、遺伝性腫瘍の解析は、がんゲノム医療の主目的の1つである
- 臨床的に頻度が高い**ACMGのガイドライン**にある下記の遺伝子などの生殖細胞系変異が見つかる可能性が高く、同定作業と解釈は必須である
  - ✓ **BRCA1 BRCA2 PALB2** (主に卵巣、乳腺、前立腺、膵胆道がん)  
相同組み換え修復関連分子 (HBOC, HRD関連)
  - ✓ **MSH2 MLH1 MSH6 PMS2** (主に大腸、胃、子宮体がん、リンチ症候群関連) *PMS2* は、偽遺伝子が複数あり変異解析には注意を要する。
  - ✓ **TP53** (主に血液腫瘍、肉腫、小児腫瘍など、リフラウメニー症候群関連)
- 生殖細胞系列バリエーションの同定は、正常DNAのWGSのデータから、参照ヒトゲノムへアライメントを行った後に、標準的なツールである**GATK HaplotypeCaller**を用いて行う。これは、参照ゲノム配列に対してアライメントを行い、QCの後、参照ゲノム配列に対しての差異を検出する（体細胞変異は、個人の正常ゲノムに対しての差異を検出）。正常DNAのNGSデータのVAFは理論的には50%であり、low VAFのvariantsは含まない。しかし正常血液DNAの解析で*TP53*のlow VAFのvariantsが見つかった場合、CHP(骨髄の体細胞変異によるクローン性造血) との鑑別が必要となるので注意。

- 多くの**common variants** (いわゆる**SNP**) が含まれているため、dbSNPや人種特異的なSNPデータベースを用いてフィルターを行う (図)。**Rare variants**(希なバリエーション)の解釈は、アミノ酸配列の変化、一般集団内の出現頻度や家系内発生などの情報を考慮し、**ACMGのガイドライン**に基づいて、**pathogenic/benign/VUS(判定不能)**の判定が行われる
- GWASによって、多数のがん発生に関わる**common variants**が同定されており、それらを組み合わせて**PRS (polygenic risk score)**にてがんの遺伝性を診断する試みがある

(図) 遺伝性腫瘍関連 variantsの検出と解釈

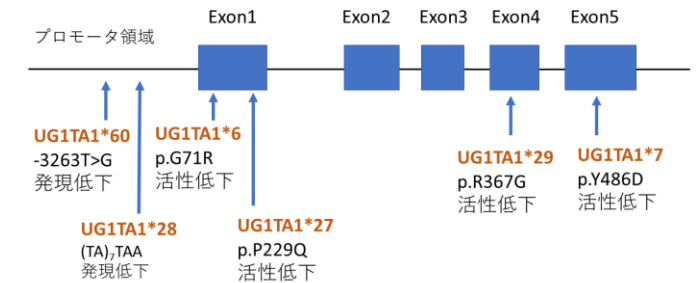


# Pharmacogenomics (PGx)

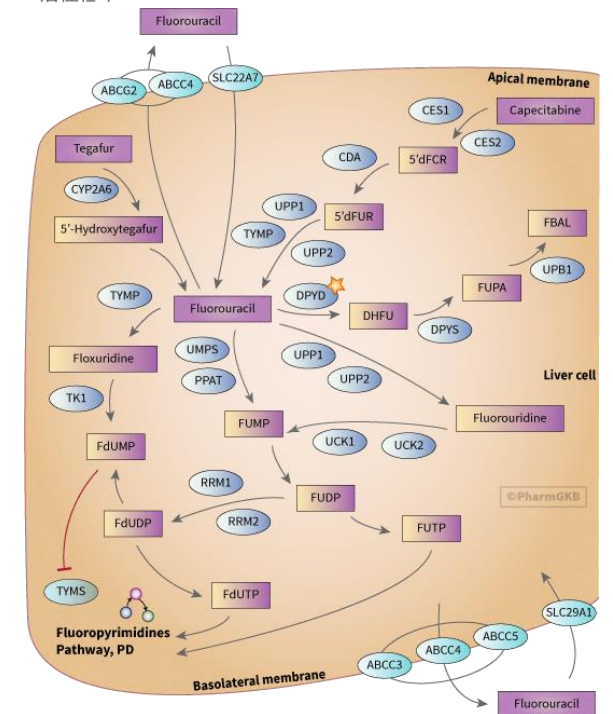
主に生殖細胞系のゲノム情報を用いて、がんの薬物療法の効果や副作用を予測し、薬剤選択や投与量調節をおこなう

- PGx検査として、大腸がんなどの治療で用いられる抗がん剤イリノテカンによる副作用（重篤な下痢など）の可能性を調べるヒト遺伝子診断薬〔UDP-グルクロン酸転移酵素をコードする遺伝子（UGT1A1）多型検査〕が2008年より保険適用となっている。UGT1A1には、プロモータ領域やエクソンに、様々なその酵素活性や発現に影響を与えるバリエーション(variants)が報告されている（上図）。
- チオプリン製剤（イムラン、アザニン）は、自己免疫疾患や白血病の治療で用いられるが、その副作用として重度の白血球減少や脱毛がおこることがある。NUDT15は、これら薬剤の代謝に関わる遺伝子であり、東アジア人に多いcodon139における遺伝子多型（p.R139C、p.R139H）などにより、その酵素活性が著しく低下する。NUDT15遺伝子型検査も、保険適応になっている。
- CYP2D6などのチトクロームP450ファミリーに属する薬物代謝酵素は、タモキシフェンや抗がん剤の代謝に関与し、その多型は酵素活性に影響をあたえ、薬剤の効果や副作用の出現と関連する。タモキシフェンはCYP2D6、5-FUは、DPYD、DPYS、CYP2A6などのvariantsと関連する（下図）。プラチナ製剤については、GSTP1代謝酵素に加えて、ERCC1などのDNA修復関連の遺伝子の多型と関連するという報告がある。これら代謝関連の遺伝子の多くは、コピー数変化や偽遺伝子の存在があり（類似配列が多い）、バリエーションコールは慎重に行わなければならない。
- （広義のPGxにおいては）様々な融合遺伝子や体細胞変異の有無が、それらの分子を標的とする薬剤の効果予測、使用条件となっている。ALK融合遺伝子、EGFR遺伝子変異など多数。
- 免疫チェックポイント阻害剤は、（体細胞変異の）MSI検査を行い、MSI陽性のすべての癌腫について効果が期待できるものとして、使用適応されている。

UGT1A1遺伝子の多型



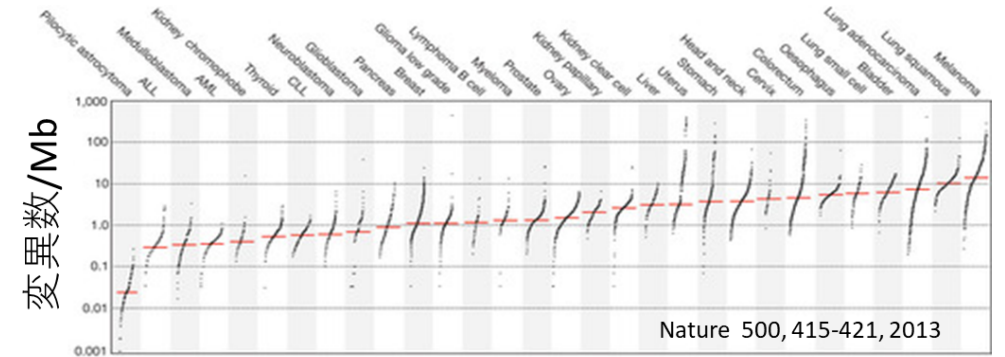
5-FUの代謝経路 (pharmgkb.org) より



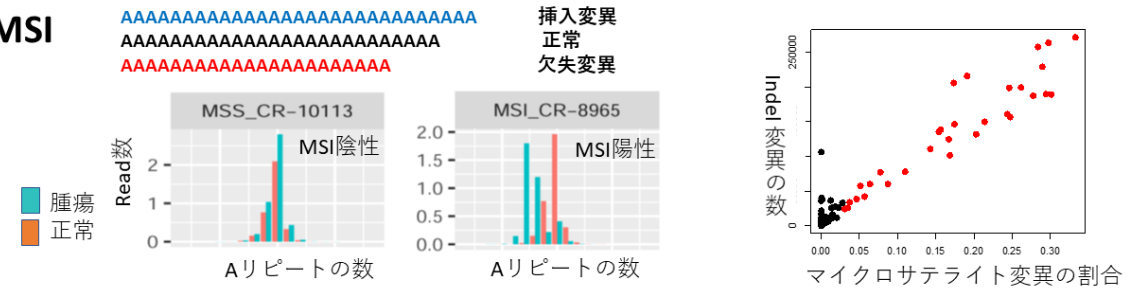
### マクロゲノム解析 SNV/indel の総数 (TMB、MSI)

- 全ゲノムシーケンスのデータから、個々の遺伝子の変異 (木をみる) だけでなく、ゲノム全体のパターン解析を行う (森をみる) 必要あり
- ゲノム全体での変異の数**TMB: (tumor mutation burden)** は、最も簡便なマクロゲノム解析であり、これまではターゲットシーケンスやエクソームのデータから推測計算が行われていた。WGSによって、実際のTMBを計測することができる。TMBは癌腫によって大きく異なり、紫外線曝露によって発生するメラノーマ、たばこ発がんの肺癌は一般的に多く(数万~数十万)、環境発がんの関与がすくない小児がんや造血腫瘍はTMBは小さい(図上)
- TMBは、免疫チェックポイント阻害剤の効果と関連するという報告があるが、TMBの分布は、腫瘍の種類によって異なるので、TMBの閾値は慎重に設定すべきである
- **MSI (microsatellite instability)** の診断は、現在は5個ほどのマクロサテライトマーカー (BAT25、BAT26など) の塩基配列の長さの変化をもって診断されている。MSIは、遺伝性腫瘍であるリンチ症候群の診断、および免疫チェックポイント阻害剤の適応に使われている
- MSIは、DNA修復遺伝子 (MLH1, MSH2など) の変異やその機能低下が原因である。MSI陽性腫瘍では、WGSでの変異数 (SNV, indel) が、MSI陰性腫瘍の10倍近く多くなる (TMBが極めて高くなる)。しかし、TMBが高いからといって、MSI陽性とは言えない
- WGSの場合、マイクロサテライトを含むリピート配列の解析は難しく、マイクロサテライトや繰り返し配列に特化したアルゴリズムによって解析する必要がある

### TMB



### MSI



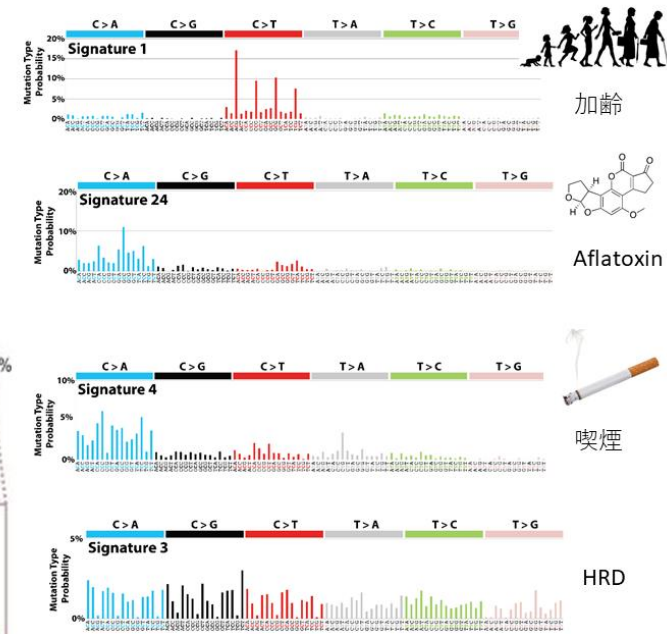
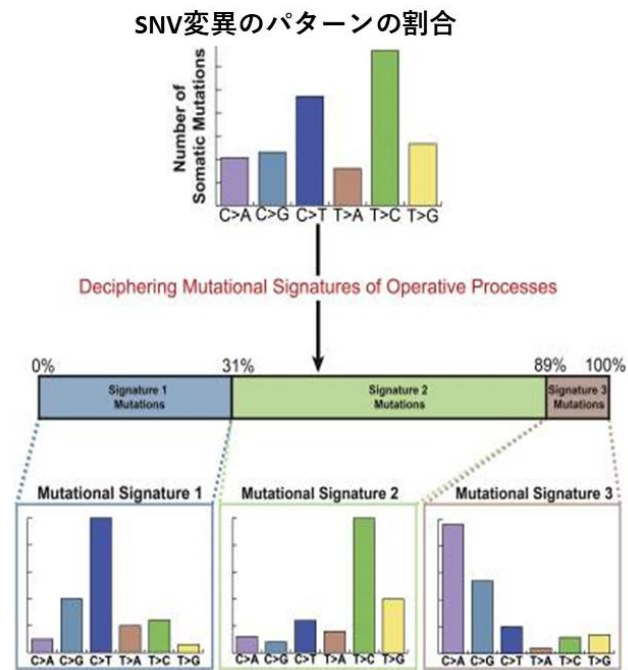
- 繰り返し配列は、正常DNAにおいても、DNAポリメラーゼのエラーのため長さが変わるので (図下左)、背景の変異確率を計算した上で、正常とがんでの繰り返し配列の長さを比較し、変異コールを行う (Fujimoto et al. Genome Res 2020)
- Indel 変異の数は、MSIと強く相関するので (図下右)、全ゲノム上でのindel 変異のTMBをもって、MSIの有無を判定することも可能である

# 変異Signature解析

## マクロゲノム解析 全SNVのパターンを見るSignature解析にて、発がん素因を推測

- 全ゲノムシーケンスのデータから、個々の遺伝子の変異（木をみる）だけでなく、ゲノム全体のパターン解析を行う（森をみる）必要あり
- 変異は、DNAの化学反応である。DNAはさまざまな化学物質や物理的ストレスによって修飾されたり、ダメージを受ける。修飾されたDNAは、細胞内のDNA修復機構によって修復されるが、修復が不十分だと、変異形成につながる。変異Signatureは、外的発がん要因（環境要因）への暴露と内的ゲノム修復要因の低下を反映している
- 100～数十万個同定されるがんのSNV変異は、基本的には図の6つの塩基置換パターンに分類される（C>A, C>G, C>T, T>A, T>C, T>G）
- がんの変異はそれぞれの素因に対応した特有の塩基置換パターンを持っている。この変異パターン（Signature）を、変異箇所とその前後を含め3つの塩基置換のパターンととらえ、それらを数学的に分解し、これまで90個ほどのSignatureパターンが同定された（図左）。このSignatureパターンが組み合わさって、それぞれのがんの変異Signature が形成されている
- **COSMIC (Catalogue of Somatic Mutation of Cancer)** では、これら92個の変異Signature について、環境暴露因子や内因性のゲノム不安定との関連が意義づけられている。しかし、多くの変異Signature の意義は依然不明である

Alexandrov et al. Cell Rep. 2013;3(1):246-59.



変異Signature の例（図右、COSMIC Mutational Signature v2より）

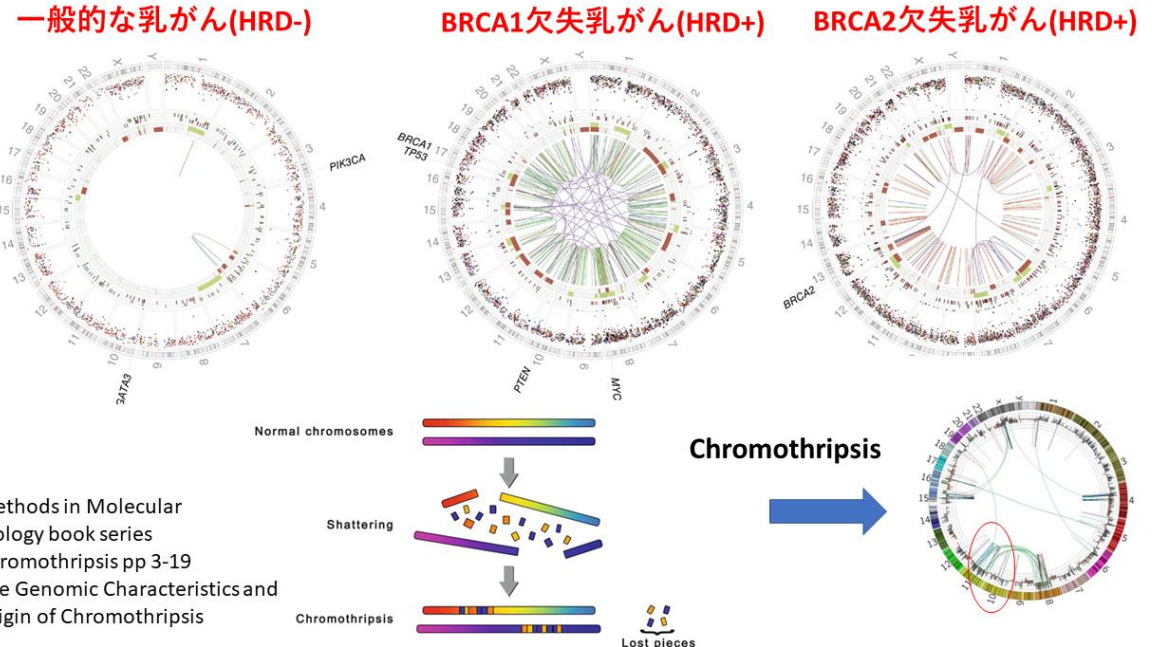
- ✓ 加齢に伴うSignature (C>T)
- ✓ 発癌物質Aflatoxin のSignature
- ✓ 喫煙やアルコール曝露によるSignature
- ✓ 内因性ゲノム不安定であるHRDと関連するSignature

## マクロゲノム解析 SVのパターン解析 (HRD、Chromothripsis)

- 全ゲノムシーケンスのデータから、個々の遺伝子の変異 (木をみる) だけでなく、ゲノム全体のパターン解析を行う (森をみる) 必要あり
- BRCA1/2遺伝子変異などにより相同組換え修復 (DNAを修復する仕組みの一つ) に欠損する状態となり、これを**相同組換え修復欠損 (HRD)**という。PARP阻害剤の適応判定のためにコピー数異常のパターン診断が行われているが、がんのWGSのデータを用いることによってHRDの状態をより精度高く判定することができる
- HRD は、卵巣がんの40-50%、乳がん20%、前立腺がん10%、膵臓がんの10%に観察される (Nguyen L, et al. Nat Commun. 11:5584, 2020.)
- HRDの判定は、以下の情報を組み合わせて判定される (Davis H, et al. Nature Medicine 23, 517-525(2017))
  - ✓ SVの数
  - ✓ Inversion の数 (BRCA1変異と関連)
  - ✓ SNV Signature (SBS3やSBS5)
  - ✓ SVのbreakpoint周辺のindel数
  - ✓ CNVの大きさや数

- **Chromothripsis** (下図) とは、一回の染色体の破壊的な粉砕化に起因し、1-2個の染色体に限局性に数十から数百もの構造異常が生じる現象である。通常はコピー数異常は伴わない (欠失や増幅の前段階では? という最近の報告あり)。ほぼ全てのがん種において2~3%の頻度で生じ、高悪性度のがんの発症やTP53変異、放射線などのゲノム不安定との関連が提唱されている

Davis H, et al. Nature Medicine 23, 517-525(2017)





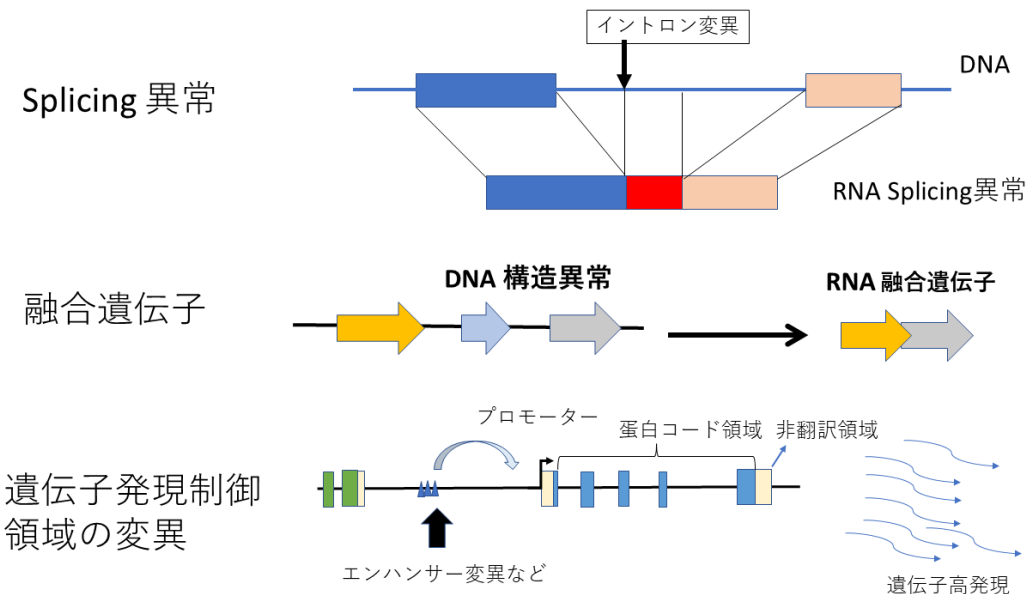
# RNAシーケンスとの統合解析

RNAseq との統合解析は、変異の意義や解釈には必須

- イントロンの変異や構造異常が検出できたとしても、それらの変異の効果や意義については、多くが不明である
- 全ゲノムシーケンスに加えて、同じがん組織由来のRNAシーケンスを行い、全ゲノム情報と統合することで、より変異の意義づけができることになる
- イントロンやexon-intron 付近のアミノ酸置換を伴わない変異は、RNAseqをおこなうと、**splicing 異常**を生じさせていることがある
  - ✓ 構造異常によって、**融合遺伝子**が発現することがある
  - ✓ **遺伝子制御領域**（エンハンサーなど）の構造異常やプロモータのSNVによって、下流遺伝子の高発現を誘導することがある。最も頻度の高いプロモーター変異は、*TERT* 遺伝子上流の変異（SNV, SV, ウイルス組み込み）である

● RNAシーケンス解析の一般的な手法

- ① Total RNAよりpolyA選択にてmRNAを精製
- ② RNAを化学的に断片化
- ③ cDNAの合成とともに、NGSのライブラリー作成
- ④ NGSによるシーケンス（通常は20M read以上のデータをとる）
- ⑤ シーケンスリード（FASTQ）のQC
- ⑥ STARやTopHat2/HISAT2にて参照RNA配列（RefSeqなど）に対してアライメントを行う

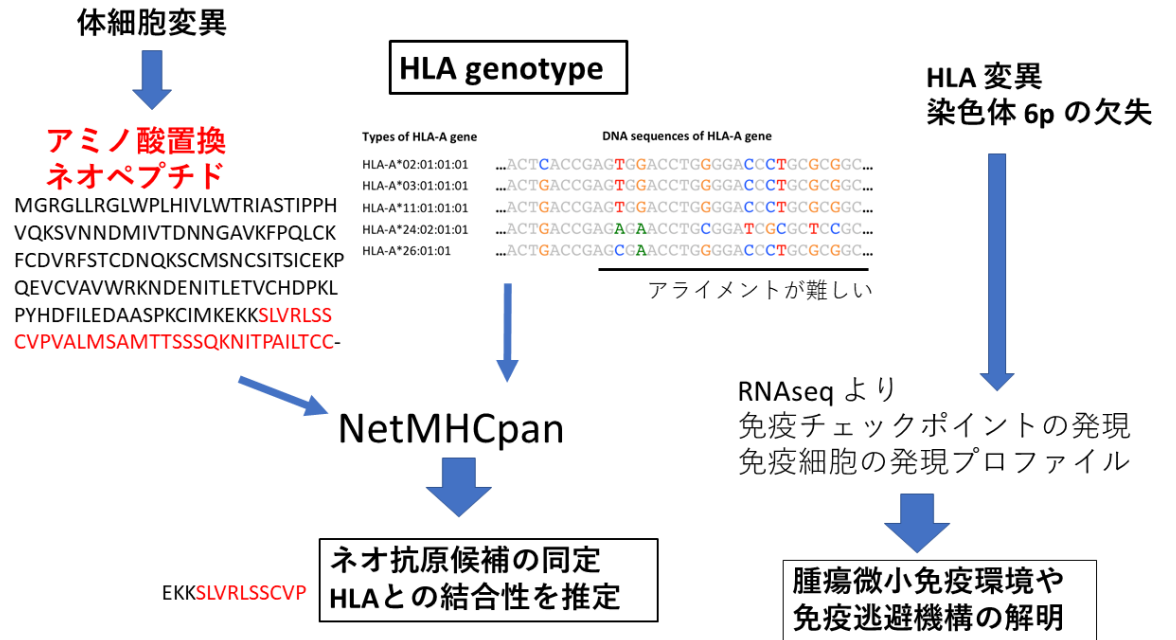


- ⑦ どの遺伝子領域に何本リードがマッピングされたかを、遺伝子の長さで補正した上で計算し、個々の遺伝子の発現量データとする（発現量の単位の例 **FKPM** : fragment per kilobase of exon per million reads mapped)
- ⑧ 融合遺伝子の検出：SV検出と同様の手法
- ⑨ Splicing異常の検出：これも慎重にアライメントを行う必要がある
- ⑩ 新規RNAの同定

免疫の疾患でもあるがんにおいては、HLAなどのゲノム免疫解析は治療選択において重要

- がんはゲノム変異が蓄積して発生し進展するが、初期の段階では、変異からプロセスされるネオ抗原が認識され免疫細胞によって排除される。しかし、ある段階で、**免疫逃避機構**を獲得して、増殖し、肉眼的ながんが成立するものと考えられている (**immune editing仮説**)
- がんゲノムの情報から、このimmune editingについて解析するのががん免疫ゲノム解析であり、以下の解析項目が含まれる
  - ✓ 免疫認識の中心であるHLAの型や変異の同定
  - ✓ HLAの型と体細胞変異情報から**ネオ抗原**を予測する
  - ✓ RNAシーケンスのデータから、免疫チェックポイント分子の発現、免疫細胞の発現を推定する
  - ✓ TCR/BCRの配列情報より、T細胞やB細胞の多様性を解析する
- **HLA型決定**：HLAには極めて多くのvariantsがあり、その組み合わせは、膨大な数（約 $1.4 \times 10^{21}$ ）に上る。NGSを用いてこの極めて多様な配列を決定する方法が開発されてきている。極めて多様なHLA領域にreadをアライメントすることが要であり、WGSからのreadを、HLA型のデータベース（IPD-IMGT/HLA database）にあるすべてのシーケンス配列に対してアライメントを行い、HLAの型を決めるアルゴリズムが主体である
- HLAminer、ATHLATES、Optitype、ALPHLARDなど、様々なHLA callerが開発されてきているが、WGSのデータからアライメント作業を綿密にしなければならない

(図) がん免疫ゲノム解析の例

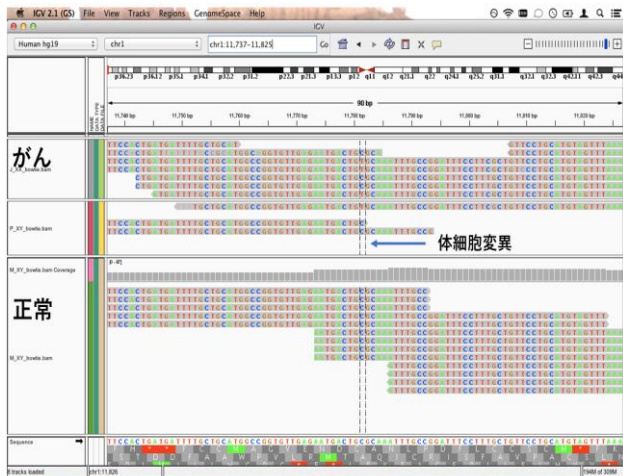


- HLA領域（染色体6p）には、コピー数異常などの体細胞変異が入っており、これによって免疫逃避を獲得している場合もあるので、HLA領域の変異解析も重要である
- HLA型と体細胞変異 > アミノ酸の変化の情報を入力することで、NetMHCpanなどのアルゴリズムでHLAと結合するネオ抗原を予測することができる

# がんゲノムの可視化 (Visualization)

同定される膨大な変異データを可視化することで、がんゲノムの理解および解釈を深める

- **IGV (Integrative Genomics Viewer)** は、マッピングデータを視覚的に確認することができる大変便利なツールである。リファレンスとなるゲノムファイル（及び、インデックス付与したfaiファイル）と、マッピング結果のソート済みBAMファイルを入力することで変異部分を視覚化することができる
  - ✓ 変異リファレンスに対して異なる塩基配列をもつリードの該当塩基は、色分けされてハイライト表示される。ズームアウトした場合は、縞模様のように表示される
  - ✓ Pair-end間の距離情報にしたがって色分けされて表示される。deletion（赤）、insertion（青）、ペアとなるリードが他の染色体にマッピングされた場合、そのほかの色で表示される。

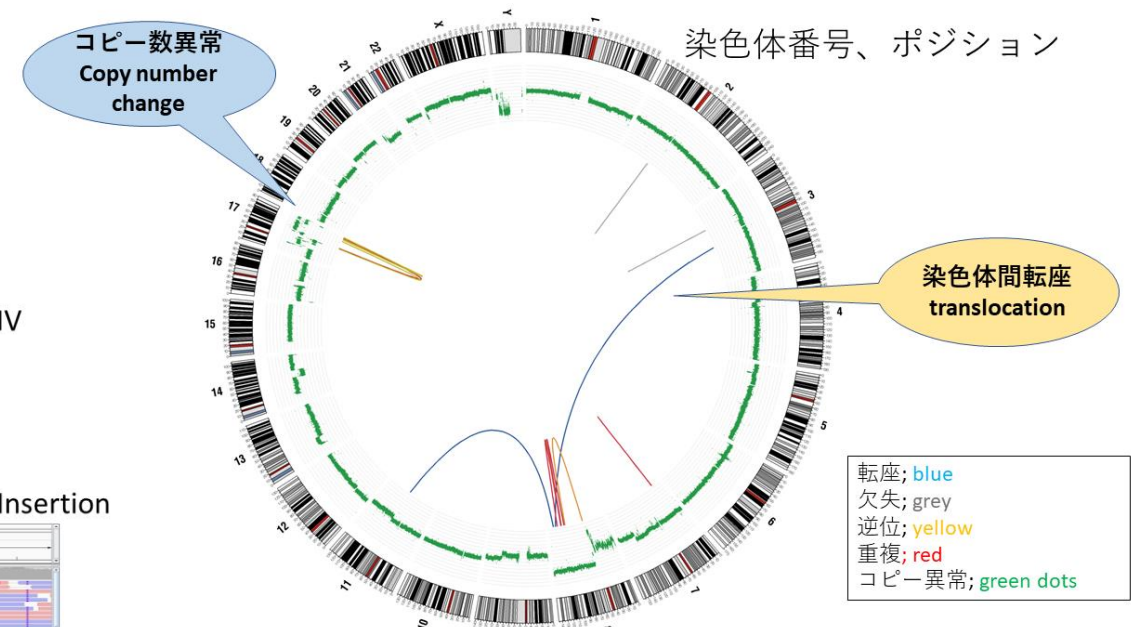


SNV



Insertion

- **Circos plot** : ゲノム構造の視覚化の際によく利用されるプロット。染色体1-22、X、Yの各ポジションでのコピー数異常（緑）や構造異常を、1つの図でわかりやすく視覚化できる



# がんゲノムデータベース

大規模ながんゲノムプロジェクトにて、データベースが構築され公開されており、変異頻度の取得や変異の解釈に使われる

- **TCGA** (The Cancer Genome Atlas) : 2007年に始まった米国の大規模がんゲノムプロジェクト。30種類の癌腫についてそれぞれ500以上のサンプルについて、exome、WGS、RNAseq、SNP chip (CNV)、DNA methylationチップ、そしてRPPAでのプロテオームのデータも揃えており、データは制限公開である
- **ICGC** (International Cancer Genome Consortium) : TCGAを含む16か国のがんゲノム解析データを集約したものであり、exome/WGSについては、25,000例のデータを公開した。WGSのデータについては、PCAWGとして、2700例のpan-cancerのデータを統一したパイプラインにて解析して、その結果を公開している (Nature 2020)
- **TARGET** (Therapeutically Applicable Research to Generate Effective Treatments) : 米国での小児がんを中心とするゲノム解析のデータベース
- **COSMIC** : UK Sangerセンターにて、論文発表などで公開された変異データをキュレーション(変異の病的意義などの解釈)した上で集積し公開している。世界最大規模
- **GEL** (Genomic England) : UKのゲノム医療プロジェクトで、がんと難病のWGSを臨床の場で10万人以上行い、その結果を患者に還元している。がんゲノムのデータは、多くのFFPEからのDNAのWGSが含まれている
- **HMF** (Hartwig Medical Foundation) : オランダにて、再発性がんの生検サンプルからの臨床WGSを行い、臨床試験に繋げている
- **GENIE** : 米国AACRが主導で、臨床でのがんゲノムシーケンス (ターゲットシーケンス) のデータと臨床データを収集したもの
- **cBioPortal** : GENIEやTCGAのデータを含めて公開しているポータルサイトで、様々な解析ツールを実装している
- **CIVIC** : がんの変異の解釈データも公開している
- **ClinVar** : 生殖細胞系列variantsを中心に解釈もつけて公開
- **MGeND** : 日本でのがんゲノムを含む様々な疾患ゲノムデータを収集し、その臨床的意義についても付加している日本人のClinVar

データベース		アクセス	症例数	データの種類
TCGA	米国のがんゲノムプロジェクト	制限公開	85,000	WGS, WES, RNA, DNAメチル化、CNV
ICGC	国際がんゲノムプロジェクト	制限公開	25,000	WGS, WES, RNA
TARGET	小児がん	制限公開		WGS, WES, RNA
COSMIC	論文発表された変異データ	公開	149,000	論文発表された変異データ
GEL	UK 臨床ゲノム診断のデータ	制限公開	26,000	WGS
HMF	オランダ臨床ゲノム診断のデータ	制限公開	5,200	WGS
GENIE	米国臨床ゲノム診断のデータ	制限公開	111,000	臨床ゲノム診断での変異データが主
cBioPortal	TCGA/TARGETのデータを含む	公開		臨床ゲノム診断での変異データが主
CIVIC	がんの変異の解釈	公開		がんの変異の解釈
ClinVar	variantsデータと解釈で最大級	公開		生殖細胞系列variantsを中心に中心
MGeND	日本人のvariantsの解釈	公開		遺伝性腫瘍、難病、がんゲノムを含む

# 第 V 章

## 1次解析から得られた結果の 臨床的な意義づけ

愛知県がんセンター研究所 所長

**井本逸勢**

1. がんの全ゲノム解析から得られたバリエントの臨床的意義づけの背景として、
  - 必要性を理解する
  - 対象を理解する
2. がんの全ゲノム解析から得られたバリエントの注釈づけについて、
  - 過程を理解する
  - データベース（DB）・バイオインフォマティクスツールを理解する
3. がんの全ゲノム解析から得られたバリエントの臨床的意義づけについて、
  - 過程を理解する

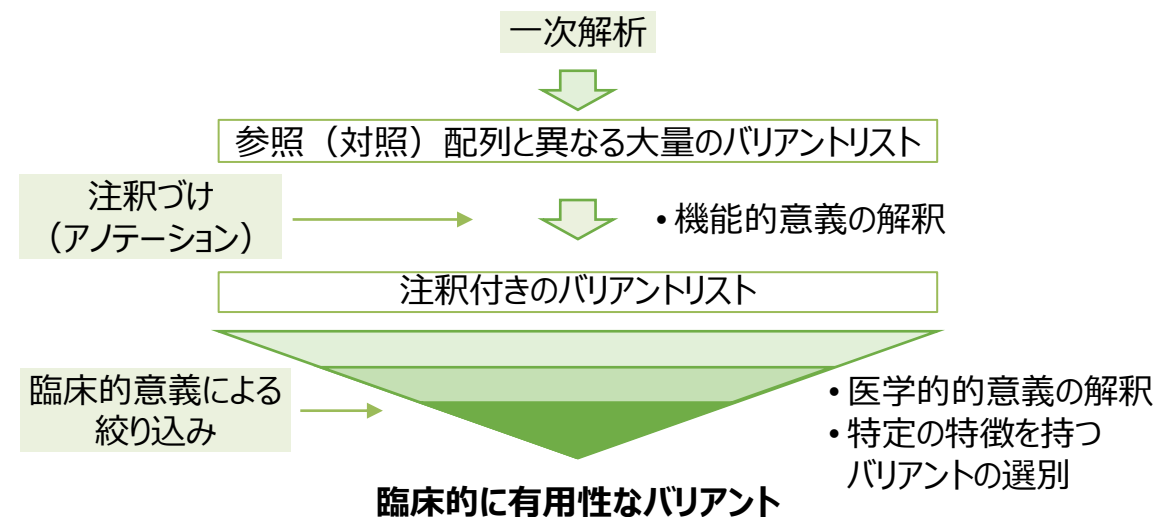
参照（対照）配列と異なる無数のバリエントから臨床的に有用なものを絞り込むには、臨床的意義の解釈が必要である

## ● バリエントの注釈づけ・臨床的意義づけの概要

- ✓ 個々人のがん、末梢血（対照）の全ゲノム解析を行うと、ヒトゲノムの参照配列（リファレンスゲノム）と異なる無数の一次構造変化（バリエント）が検出される
- ✓ バリエントのうち、臨床的に意義のあるバリエントのみが、診断マーカーや治療標的として臨床的有用性がある
- ✓ そのため、一次解析から得られた各バリエントに、遺伝子機能への影響の有無などの解釈に役立つ各種情報を付加する『注釈づけ（アノテーション）\*』を行った後、その注釈情報に基づき臨床的に意義のあるバリエントを絞り込む『バリエントの臨床的意義づけ』が行われる
- ✓ 臨床的意義づけの結果に基づき、臨床的有用性（Actionable：治療薬の選択、診断、予後予測、予防など医学管理に用いることができる）が検討される（がんの治療薬選択 → 入門編第VI章参照）

\* **アノテーション（annotation）**：塩基配列データに遺伝子構造や遺伝子機能の情報、文献情報などを注釈づけする事を指す。大規模の処理が必要な為、作業は通常先にコンピュータを用いて自動的に行われる（自動アノテーション）。注釈結果に対する人手による検証と修正処理は、アノテーションとは別に**キュレーション（curation）**と呼ばれる

## 全ゲノム解析の1次解析から得られた結果の臨床的意義づけの過程



## ● がんの場合の臨床的有用性の例

- ✓ 病態の解釈（腫瘍原性）
  - ✓ 疾患素因関連
  - ✓ 治療効果予測マーカー → ✓ **治療薬選択**
  - ✓ 診断マーカー
  - ✓ 予後予測
- （入門編第VI章参照）

# 対象となるバリエント（1）

• がんの解析における注釈づけ・臨床的意義づけの対象となるバリエントには、体細胞と生殖細胞系列の2種類がある

● がんゲノム解析で腫瘍/非腫瘍（T/N）ペア検体が解析対象となる場合

✓ 腫瘍細胞特異的体細胞バリエント（体細胞変異）

対照はペアとなる非腫瘍細胞ゲノム。パッセンジャー遺伝子と区別して、腫瘍の病態に寄与する**ドライバー遺伝子バリエント\***の検出が目的となるが、腫瘍組織に含まれる腫瘍細胞の割合や腫瘍細胞の多様性（組織内に含まれる腫瘍細胞の不均一性）が体細胞バリエント同定の感度・特異度に影響し得ることに注意する必要がある

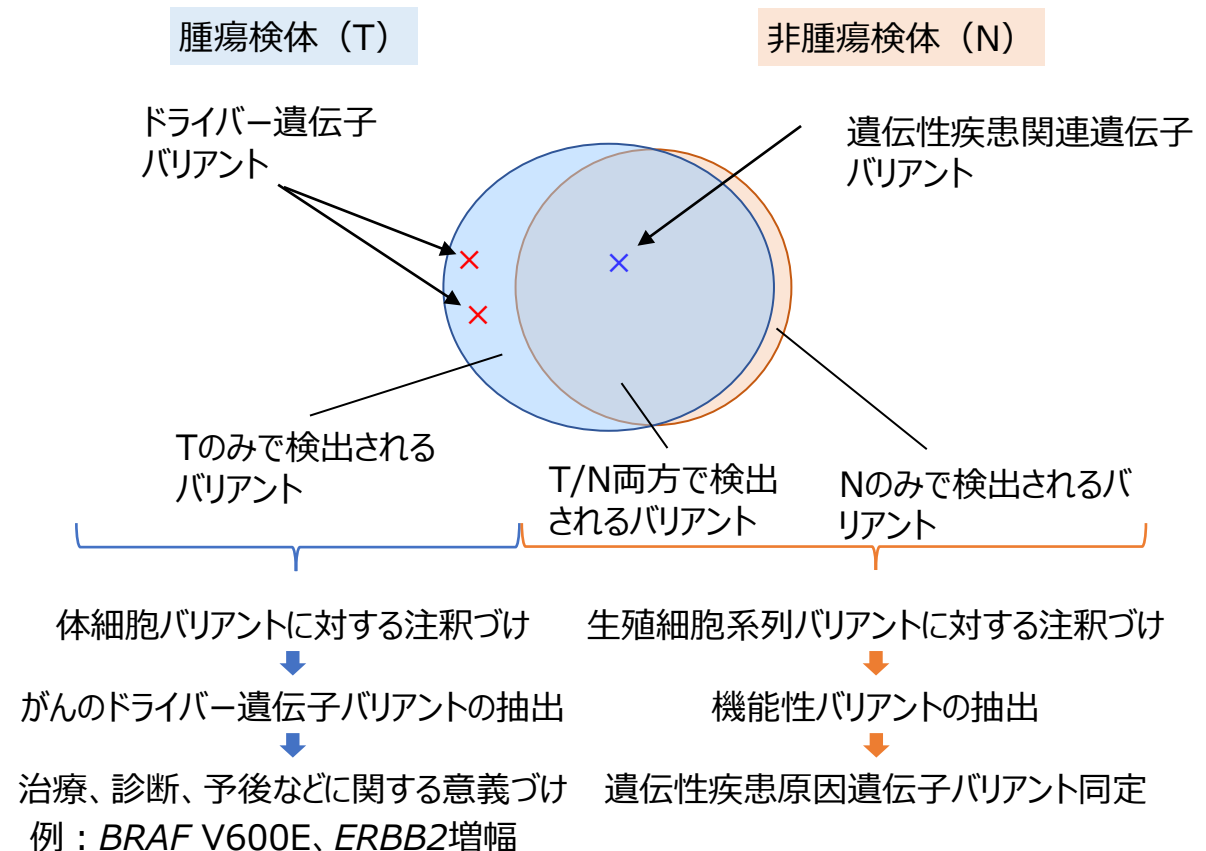
✓ 非腫瘍細胞で検出される生殖細胞系列バリエント

対照はヒトリファレンス（参照）ゲノム。集団内で頻度の高い多型など疾患への寄与度が低いバリエントを除外することにより、遺伝性疾患の原因となる病的バリエントの検出が行える可能性がある（網羅的ながんゲノム検査におけるGermline findings）

\***ドライバー遺伝子バリエント**：ヒトの細胞では様々な要因によりゲノムDNAに損傷が生じる為、体細胞では細胞分裂のたびにバリエントが蓄積していく。大多数は細胞増殖などの表現型に影響を与えない中立的变化であるが、細胞に対してがん化に有利な性質を与える「ドライバー（遺伝子）バリエント」と呼ばれる変化が生じることがあり、がんでは～5個程度が認められる（1）。がんでは認められるそれ以外のほとんどの中立的变化は**パッセンジャーバリエント**とされるが、近年、その中に弱いドライバーとしての機能を持つものが含まれることが報告されている（2）

（1）Campbell PJ, et al. Nature 2020;578:82-93、（2）Kumar S, et al. 2020;180:915-927）

## がんの腫瘍/非腫瘍（T/N）検体でのゲノム解析で検出されるバリエント

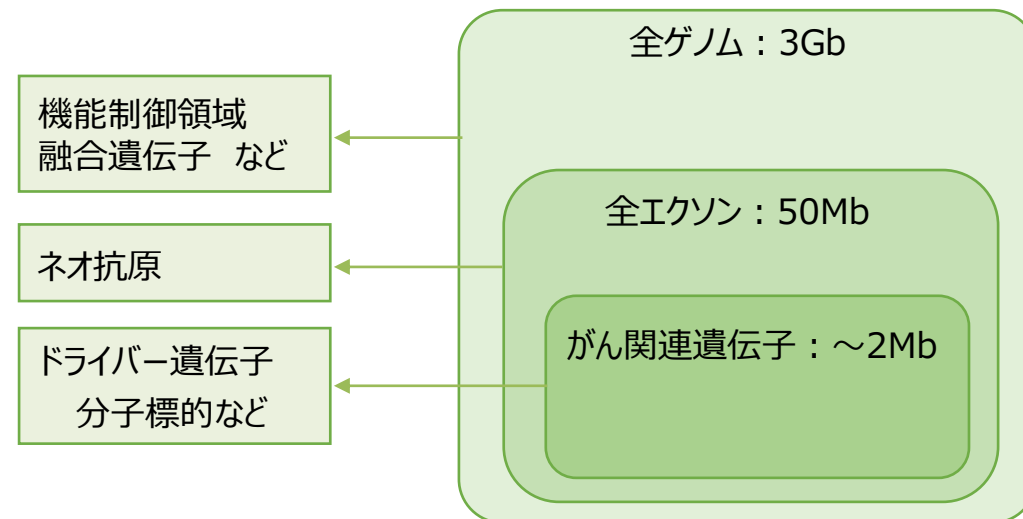




# 対象となるバリエント (2)

- 全ゲノム解析を行う場合に注釈づけ・臨床的意義づけの対象となるバリエント (入門編第IV章「がん全ゲノムシーケンスでわかること」参照)
- 全ゲノム解析で検出可能なバリエントの種類
  - ✓ 一塩基置換 (SNV)
    - ミスセンス (アミノ酸置換)、ナンセンス (終止コドン)、など
  - ✓ 小さな欠失/重複 (INDEL)
  - ✓ 大きな構造異常 (SV) (注)
    - 欠失、重複・増幅などのコピー数変異 (CNV)
    - 挿入
    - 逆位
    - 縦列反復
    - 転座
  - ✓ 繰り返し配列
  - ✓ その他 (ミトコンドリア配列、ウイルス・細菌ゲノム、Mutation signatureなど)
- 全ゲノム解析の対象となるゲノム領域
  - ✓ 遺伝子のエクソン中タンパクコード・非コード領域、イントロン領域 (スプライシング領域を含む)
  - ✓ 遺伝子制御領域を含む遺伝子間の非コード領域 (注)
  - ✓ ミトコンドリア
  - ✓ ウイルス・細菌

## がんのゲノム解析対象領域と期待される臨床的に有用なバリエント情報



(注) 入門編第IV章「RNAシーケンスとの統合解析WGTS」参照。RNAシーケンスで転写産物の配列や量を検出することで、ゲノムの一次構造変化による融合遺伝子、スプライシング異常、転写制御への影響が解釈される場合がある

- がんの全ゲノム解析の1次解析で得られた各バリエーションの意義を解釈するには、まず様々な情報を付与する必要がある

## ● 入力

- ✓ バリエーションとして認識されたリファレンス（参照）ゲノムと異なる一次構造変化がその位置や変化とともにテキスト形式で記録されたVCFファイル\*が、リストとして入力される

## ● 情報の付与（注釈づけ）

- ✓ 個々のバリエーションに、遺伝子産物に与える影響（アミノ酸配列変化など）、各種データベース（DB）情報（アレル頻度など）、機能変化予測などのバイオインフォマティクスツール実行結果を追記する
- ✓ 大規模なDNA配列データに対応したアノテーションプログラムがある
  - ANNOVAR
  - SnpEff
  - Variant Effect Predictor (VEP, Ensembl)
  - Variant Annotation Integrator (VAI, UCSC Genome Browser)

## ● 出力

- ✓ アノテーション結果が追記されたバリエーションのリストが、VCFファイル\*（あるいはHTML形式）として出力される

バリエーションアノテーションプログラムの比較

プログラム名	VCF入力	VCF出力	HTML出力	ライセンス
SnpEff	○	○	○	フリー
ANNOVAR	○	○	対応なし	アカデミックフリー
VEP	○	○	○	フリー
VAI	○	○	○	フリー

\***VCFファイル (variant call format)** : BAMファイルからバリエーションを検出するソフトウェアを用いて参照ゲノムと比較してゲノムのどの位置にどのような変化があるかを検出した結果をタブ区切りのテキスト形式で記録したもの（拡張子は「.vcf」）。最初の8列が固定され、9列目には解析に使用したソフトに応じたデータのフォーマットが書かれ、10列目以降に各サンプルの実際のデータが書かれている（サンプルが複数ある場合は11列目、12列目と列を増やす）

- バリアントの注釈づけは複数の項目に関して多面的に行われる（ANNOVERを例に注釈付けされる項目の例を示す）
- 遺伝子ベースの注釈づけ（どの遺伝子にどのように影響するか）
  - ✓ バリアントと既知遺伝子との位置関係（エクソン内、イントロン内、遺伝子間、など）に着目し、どの遺伝子の機能に影響を及ぼしそうかを推定する
  - ✓ バリアントがタンパクのコードの変化を引き起こすかどうかや影響を受けるアミノ酸を特定して、タンパク機能への影響の推定に用いる
  - ✓ RefSeq、UCSC、ENSEMBL、GENCODE、AceView、など多数の遺伝子定義が利用可能である
- 領域ベースの注釈づけ（ゲノム上の特定の機能領域に影響するか）
  - ✓ さまざまな特定のゲノム領域（転写調節領域など）とバリアントとの関係を同定して、遺伝子発現に影響する可能性、などを推定する
  - ✓ 種間の高度保存領域、予測される転写因子結合部位、マイクロRNAや他のnon-coding RNA、セグメント重複（segmental duplication）領域、GWASヒット領域、ゲノムバリアントのDB、DNase I hypersensitivity sites、ENCODE H3K4Me1/H3K4Me3/H3K27Ac/CTCF結合領域、ChIP-Seqのピーク、RNA-Seqのピーク、など様々な領域に対応する

## バリアントと既知遺伝子との位置関係の評価の例

評価	説明
exonic	タンパクコード領域内のバリアント
splicing	splicing junctionから2 bp以内のイントロン内バリアント
ncRNA	遺伝子の定義の中でタンパクをコードする注釈づけがされていない転写産物（non-coding RNA）中のバリアント
UTR5	5' untranslated region (UTR) 内のバリアント
UTR3	3' untranslated region (UTR) 内のバリアント
intronic	イントロン内のバリアント
upstream	転写開始点から1 kb上流以内にあるバリアント
downstream	転写終止点から1 kb下流以内にあるバリアント
intergenic	遺伝子間領域内のバリアント

- フィルターベースの注釈づけ（DBや各種プログラムを利用：後述）
  - ✓ 特定のDBでの登録状況や機能推定プログラムから機能を推定する
  - ✓ dbSNPでの登録の有無や、gnomAD、1000genome、NHLBI-ESP 6500、などの集団DBでの出現頻度から、疾患と関係しないバリアント（一般集団で見られる多型）の可能性を推定する
  - ✓ 各種の予測プログラムを用いて、バリアントの機能有害性予測スコアを算出する（例：SIFT/PolyPhen/LRT/MutationTaster/MutationAssessor/FATHMM/MetaSVM/MetaLR/GERPスコアの計算や、複数のスコアを統合したCADD、など）
  - ✓ COSMIC、ICGC、NCI60など体細胞バリアントDBへの登録も考慮する

- 配列バリエーションの種類別の遺伝子機能への影響を説明する
- 遺伝子の翻訳領域内の塩基配列のバリエーション
  - ✓ 置換・欠失・重複・挿入・逆位・変換などで、タンパクレベルでの停止コドンへの置換（ナンセンス変異）、読み枠のずれ（フレームシフト変異）が起こると機能を失うことが多い
  - ✓ 置換・逆位・変換・挿入・欠失などで、アミノ酸での置換（ミスセンス変異）が起こると、機能に影響する場合としない場合がある
- 遺伝子の翻訳領域外の塩基配列バリエーション
  - ✓ スプライシング部位の塩基の変化やプロモーター領域など機能部位の塩基変化は、スプライシング異常によるアミノ酸配列の変化や遺伝子発現量の変化を介して機能に影響する場合がある
- 大きな領域の構造異常（SV）
  - ✓ ゲノムコピー数変化（CNV）や融合遺伝子形成などが生じると、遺伝子発現量や遺伝子機能の変化を介して機能に影響する場合がある
- 機能獲得型と機能喪失型のバリエーション（臨床的意義づけの過程（3）参照）
  - ✓ 機能獲得型のバリエーションは、特定の機能部位（Hot spot）に生じるミスセンス変異など、機能活性化につながる特定の変化であることが多い
  - ✓ 機能喪失型のバリエーションは、機能を失わせる変化であればよいため、部位も変化の種類も様々である

DNA塩基配列のバリエーションの種類例

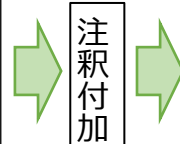
種類	参照配列と比較した場合の塩基配列の変化	
Substitution (>)	置換 1つの塩基が1つの塩基と置き換えられた変化	AGTCTGAGCTA ↓ AGTCTGA <u>A</u> CTA
Deletion (del)	欠失 1つ以上の塩基（多くは～数塩基）が削除された変化	AGTCTGAGCTA ↓ AGTCTGACTA
Duplication (dup)	重複 1つ以上の塩基（多くは～数塩基）のコピーが元の3'側に挿入された変化	AGTCTGAGCTA ↓ AGTCTGAG <u>G</u> CTA
Insertion (ins)	挿入 1つ以上の塩基（多くは～数塩基）が挿入された変化で、重複以外の場合	AGTCTGAGCTA ↓ AGTCTGAG <u>TT</u> CTA
Inversion (inv)	逆位 1つ以上の塩基（多くは～数塩基）が元の配列と逆方向かつ相補的に挿入された変化	AGTCTGAGCTA ↓ AG <u>TT</u> CAGGCTA
Conversion (con)	変換 1つ以上の塩基（多くは～数塩基）が別の場所の塩基配列と取り替えられた変化	AGTCTGAGCTA ↓ ACAGT AGTACAGT <u>G</u> CTA
Deletion-insertion (delins)	欠失挿入 1つ以上の塩基（多くは～数塩基）が1つ以上の塩基（多くは～数塩基）と取り替えられた変化で、置換、逆位、変換以外の場合	AGTCTGAGCTA ↓ AGTCT <u>I</u> CTAGCTA

# 注釈つきVCFファイルの出力

- 個々のバリエントにあらかじめ選択した様々な項目に関して注釈が付加され、ひとつのリストとしてVCFファイルに出力される
- 1行に1バリエントの情報がまとめて表示され、各項目（列）で特定の条件を満たすバリエントを絞り込める（フィルターをかけられる）
  - ✓ 出力される注釈付加項目やその記載方法などは、使用するプログラムやあらかじめ選択した項目・DBの種類に依存する

## 一次解析後の入力VCFファイル

		固定項目						オプション項目 + サンプル毎データ	
染色体	塩基位置	DBのID	参照配列	バリエント	QUAL	FILTER	INFO	FORMAT	SAMPLE
CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	SAMPLE
① chr1	12837308	rs201834787	T	A	1180.77	SnpcCluster	AC=1; AF=0.500; AN=2; BaseQRankSum=-1.953; DB; DP=680; Dels=0.00; FS=10.651; HaplotypeScore=10.0798; MLEAC=1; MLEAF=0.500; MQ=57.09; MQ0=3; MQRankSum=-12.885; QD=1.74; ReadPosRankSum=7.394	GT:AD:DP:GQ:PL	0/1:567,113:680:99:1209,0,17508
② chr1	12837311	rs199576535	G	A	1138.77	SnpcCluster	AC=1; AF=0.500; AN=2; BaseQRankSum=-10.236; DB; DP=676; Dels=0.00; FS=12.535; HaplotypeScore=11.7626; MLEAC=1; MLEAF=0.500; MQ=57.07; MQ0=3; MQRankSum=-13.823; QD=1.68; ReadPosRankSum=7.079	GT:AD:DP:GQ:PL	0/1:563,111:676:99:1167,0,17774
....									



## 情報が付加された部分のVCFファイル（ANNOVARを利用した例）

遺伝子との関係	遺伝子名	エクソン内の意味	アミノ酸の変化	高度保存領域	セグメント重複領域	Esp6500での頻度	1000genomeでの頻度	dbSNP番号	Dbexomeでの検出	機能有害性の評価						
Func.refGene	Gene.refGene	GeneDetail.refGene	ExonicFunc.refGene	AAChange.refGene	phastConsElements46way	GenomicSuperDups	Esp6500si_all	1000g2012apr_all	snp138	Dbexome20131010	LJB23_SIFT_score	LJB23_SIFT_score_Converted	LJB23_SIFT_pred	LJB23_Polyphen2_HDIV	LJB23_Polyphen2_pred	LJB23_Polyphen2_HVAR
① exonic	PRAMEF12		nonsynonymous SNV	PRAMEF12: NM_001080830: exon3:c.T1018A:p.S340T				rs201834787	0	0.41	0.59	T	0	B	0	....
② exonic	PRAMEF12		nonsynonymous SNV	PRAMEF12: NM_001080830: exon3:c.G1021A:p.V341I				rs199576535	0	0.54	0.46	T	0.015	B	0.015	
....																

- バリアントの集団中での頻度情報から病的表現型との関連の有無を推定する目的で参照する海外のDBの例を挙げる

- **用途**
  - （一般）集団内の頻度が高く病的表現型と関連しない可能性が高いバリアントを除外する
- **特徴**
  - 規模の大きいDBが整備されている
  - エクソン領域だけでなく、全ゲノム解析でのデータも充実している
- **注意点**
  - ✓ 日本を含め東アジア人のデータが少ない為、これらのみを使用してアノテーションを行った場合には偽陽性の判定となる可能性がある
  - ✓ 重篤な小児疾患などは除かれたデータセットになっているが、様々なプロジェクト群が含まれているため注意が必要である

## 海外のバリアントの集団頻度情報のDB（2022年6月8日現在）

Webサイト	URL	特徴
gnomAD	<a href="https://gnomad.broadinstitute.org/">https://gnomad.broadinstitute.org/</a>	V2.1.1（エクソーム約12.6万検体+全ゲノム約1.5万、参照ゲノムはGRCh37/hg19）、v3.1.2（全ゲノム約7.6万検体、参照ゲノムはGRCh38/hg38）から成るヒト集団ゲノムバリアント情報DB。50bp以上の構造バリアントはSVs v2.1、ミトコンドリアゲノム情報もv3.1（5.6万検体）で公開。ホモ接合体の有無の情報も得られる
1000genome	<a href="https://www.internationalgenome.org/home">https://www.internationalgenome.org/home</a>	日本を含め世界の26の集団からの3,202人の全ゲノム解析データ情報（参照ゲノムはGRCh38/hg38）
Exome Variant Server	<a href="https://evs.gs.washington.edu/EVS/">https://evs.gs.washington.edu/EVS/</a>	ESP6500SI-V2（6,503検体）のエクソームのバリアント情報
dbSNP	<a href="https://www.ncbi.nlm.nih.gov/snp">https://www.ncbi.nlm.nih.gov/snp</a>	50bp以下のバリアント（RefSNP）の収集、公開。Build 155（2021.06.16）で約10億が登録される
dbVar	<a href="https://www.ncbi.nlm.nih.gov/dbvar">https://www.ncbi.nlm.nih.gov/dbvar</a>	50bpを越える構造バリアントを収集、公開。約745万が登録される

# バリアントの集団内頻度情報（国内）

- バリアントの集団中での頻度情報から病的表現型との関連の有無を推定する目的で参照する国内のDBの例を挙げる

- **用途**

- ✓ （一般）集団内の頻度が高く病的表現型と関連しない可能性が高いバリアントを除外する

- **特徴**

- ✓ 日本人特有のバリアントを含むため、人種差での偽陽性・偽陰性を減らせる
- ✓ 希少疾患などは除かれたデータセットになっている
- ✓ 全ゲノム解析でのデータ、構造バリアント情報も年々増加している

- **注意点**

- ✓ 日本国内でも規模の大きなデータが整備されつつあり信頼性はあがっているが、海外に比べてまだ小規模で単独で参照するには不十分。この為、現時点でも、集団中の頻度評価は、重複のない複数の海外のDBと組み合わせて行うことが現実的である

## 国内のバリアントの集団頻度情報のDB（2022年6月8日現在）

Webサイト	URL	特徴
<b>Human Genetic Variation Database (HGVD)</b>	<a href="https://www.hgvd.genome.med.kyoto-u.ac.jp/">https://www.hgvd.genome.med.kyoto-u.ac.jp/</a>	京都大学が運営し、1,208検体のエクソームと3,248検体の一般的なバリアントのジェノタイピングから得られたバリアント頻度情報を公開するDB
<b>Japanese Multi Omics Reference Panel (jMorp)</b>	<a href="https://jmorp.megabank.tohoku.ac.jp/202112/">https://jmorp.megabank.tohoku.ac.jp/202112/</a>	東北メディカル・メガバンク機構が運営し、14,000検体（14KJPN）の全ゲノム解析から得られた日本人バリアント頻度情報の公開DB。0.01%以上の頻度のバリアントをほぼ網羅。8,300検体の構造バリアント（8.3KJPN-SV）も公開されている。大部分が健常人と考えられ、日本で公開される最大のサンプルサイズを持つ健常人コントロールDB
<b>TogoVar</b>	<a href="https://togovar.biosciencedbc.jp/?">https://togovar.biosciencedbc.jp/?</a>	バイオサイエンスデータベースセンター（NBDC）とライフサイエンス統合データベースセンター（DBCLS）が共同開発し、研究プロジェクト横断的に集約した日本人バリアントの頻度情報を提供するDB。JGA-NGS（125検体のエクソーム）とJGA-SNP（約18万検体のSNP）に加え、GEM-J WGA（jMorpを含む7,69検体の全ゲノム）や他のDBの情報も表示可能

# バリエーションの有害性予測ツール（1）

• バリエーションが分子機能に与える効果（主に機能喪失）を予測するバイオインフォマティクスツール・DBの例を挙げる

● 用途

- ✓ 様々な根拠からバリエーションの機能への効果を予測し病原性を推定する

● 特徴

- ✓ 複数の予測ツールでの結果を統合してスコア化するツール（CADD、など）やDB（dbNSFP、など）がある

● 注意点

- ✓ 複数のツールの結果が一致していても、類似の評価基準（タンパク構造、進化的保存性、など）を用いた結果は同傾向になりやすい。このため、予測ツールの結果を単独で疾患リスクとの関連性の根拠（病原性の判定根拠）として使用すべきではない点に留意する

## バリエーションが分子機能に与える効果を予測するバイオインフォマティクスツール・DBの例

ツール名	URL	特徴
dbNSFP	<a href="https://sites.google.com/site/jpopgen/dbNSFP">https://sites.google.com/site/jpopgen/dbNSFP</a>	非同義SNV（nsSNV）やスプライスサイトSNV（ssSDNV）のバリエーションに関して、38種の予測ツール（アルゴリズム）による機能予測と様々なDBによるアノテーション情報を統合し病原性を評価したDB
CADD	<a href="https://cadd.gs.washington.edu">https://cadd.gs.washington.edu</a>	SNVとIndelに対し様々なアノテーション情報を統合して客観的な重み付けを行い病原性をスコア化する
MutationTaster	<a href="http://www.mutationtaster.org/">http://www.mutationtaster.org/</a>	非同義置換のSNVだけでなく、イントロン、イントロン・エクソン境界部（MaxEntScan）、Indel、同義置換などのバリエーションも評価対象
PolyPhen-2	<a href="http://genetics.bwh.harvard.edu/pph2">http://genetics.bwh.harvard.edu/pph2</a>	非同義置換のSNVについて、タンパク構造と進化的保存性をもとに予測
PROVEAN	<a href="http://provean.jcvi.org/index.php">http://provean.jcvi.org/index.php</a>	非同義置換のSNVやIndelについて、タンパク構造と進化的保存性をもとに予測
SIFT	<a href="https://sift.bii.a-star.edu.sg/">https://sift.bii.a-star.edu.sg/</a>	非同義置換のSNVについて、進化的保存性をもとに予測



# バリエーションの有害性予測ツール（2）

- バリエーションのスプライスサイトへの影響を予測するバイオインフォマティクスツールの例を挙げる

- **用途**
  - ✓ 様々な手法でバリエーションのスプライスサイトへの影響を予測し病原性を推定する
- **特徴**
  - ✓ スプライスサイトだけでなく、SRタンパクが結合するExonic splicing enhancer（ESE）モチーフを予測するツールなども利用可能である
- **注意点**
  - ✓ 複数のツールの結果が一致していても、類似の評価基準（タンパク構造、進化的保存性、など）を用いた結果は同傾向になりやすい。このため、予測ツールの結果を単独で疾患リスクとの関連性の根拠（病原性の判定根拠）として使用すべきではない点に留意する

## バリエーションのスプライスサイトへの影響を予測するバイオインフォマティクスツールの例（2022年6月8日現在）

ツール名	URL	特徴
GeneSplicer	<a href="http://www.cbcb.umd.edu/software/GeneSplicer/gene_spl.shtml">http://www.cbcb.umd.edu/software/GeneSplicer/gene_spl.shtml</a>	スプライスサイトの予測ツール
NetGene2	<a href="http://www.cbs.dtu.dk/services/NetGene2/">http://www.cbs.dtu.dk/services/NetGene2/</a>	ニューラルネットワークを用いたスプライスサイト予測ツール
NNSplice	<a href="https://www.fruitfly.org/seq_tools/splice.html">https://www.fruitfly.org/seq_tools/splice.html</a>	ニューラルネットワークを用いたスプライスサイト予測ツール
MaxEntScan	<a href="http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_score_seq.html">http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_score_seq.html</a> または <a href="http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html">http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html</a>	5' splice sites (3-bp exon + 6-bp intron) と 3' splice sites (20-bp intron + 3-bp exon) の配列の保存性をもとに影響を評価するツール
ESEfinder3.0	<a href="http://krainer01.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi">http://krainer01.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi</a>	Exonic splicing enhancer（ESE）モチーフやスプライスサイト・ブランチサイトの予測ツール
SPIDEX	—	ANNOVARに搭載される深層学習を用いたスプライスバリエーションの予測ツール
SpliceAI	<a href="https://github.com/Illumina/SpliceAI">https://github.com/Illumina/SpliceAI</a>	ニューラルネットワークを用いた遺伝子内のすべての置換、1塩基挿入、1-4塩基欠失のスプライシングへの影響をアノテーションするツール

# バリエーションと疾患との関連づけ情報（1）

- 主に生殖細胞系列のバリエーションの病原性の判定に使える、既知病原性バリエーションDBや推定Webツールの例を挙げる
- 用途
  - ✓ 検出されたバリエーションが、遺伝性疾患の原因遺伝子に見られる既知の病原性または非病原性バリエーションと一致しているかを検索する
- 特徴
  - ✓ 主に遺伝性疾患の原因となる既知のバリエーションのDBが複数公開されている
- 注意点
  - ✓ 古い文献中のバリエーションには、その後病原性が否定されているものが含まれている（除かれていない）可能性があり、正解が自明であるとは言えない点に留意する

## 主に単一遺伝子疾患とバリエーションとの関連づけ情報のDBの例（2022年6月8日現在）

Webサイト	URL	特徴
ClinVar	<a href="https://www.ncbi.nlm.nih.gov/clinvar">https://www.ncbi.nlm.nih.gov/clinvar</a>	遺伝性疾患など臨床的に重要な疾患の表現型とバリエーションとの関係を収載した公開アーカイブ。臨床的解釈付与はデータの登録者に任される（正解を保証していない）。2,228の登録者から1,531,964種の臨床的解釈付きのユニークなバリエーションが登録されている。同一のバリエーションに対し複数の異なる解釈が登録される場合があり、医学的診断や推奨される診療方針を提供する設計にはなっていない
Human Gene Mutation Database (HGMD®)	<a href="http://www.hgmd.org">http://www.hgmd.org</a>	ヒト遺伝性疾患に関して、論文発表された生殖細胞系列のバリエーションを包括的に集めたDB。有償のProfessional版では年に4回アップデートされるが、無償公開版では有料版に比べて情報が3年以上古い
Locus Specific Mutation Databases	<a href="https://www.hgvs.org/locus-specific-mutation-databases">https://www.hgvs.org/locus-specific-mutation-databases</a>	Human Genome Variation Society (HGVS) のホームページ上で提供される、遺伝子ごとのバリエーションDBのリスト（カタログ）
MGeND	<a href="https://mgend.med.kyoto-u.ac.jp/">https://mgend.med.kyoto-u.ac.jp/</a>	ゲノムバリエーション情報と臨床的特性を疾患横断的に統合する日本のDB。臨床的解釈付与はデータの登録者に任されるため、同一のバリエーションに対し複数の異なる解釈が登録される場合がある
VarSome	<a href="https://varsome.com/">https://varsome.com/</a>	バリエーションのACMG AMP分類での病原性解釈を、133のDBや解析ツールを用いて自動で行うツール

# バリエーションと疾患との関連づけ情報 (2) 第 V 章 1次解析から得られた結果の臨床的な意義づけ

- 主にがん特有の体細胞バリエーションの病原性予測や治療法選択に有用なバリエーションや薬効のエビデンスに関するDBの例を挙げる
- **用途**
  - ✓ 体細胞バリエーションのがんの病態への関与（ドライバー遺伝子変異かどうか、治療抵抗性に関わるか、など）や薬剤の効果判定マーカー（Actionableかどうか）としての意義を調べる
- **特徴**
  - ✓ DBによりカバーされる遺伝子やバリエーション、データの信頼性、エビデンスの区分などが異なるために、複数のDBを参照し統合的に判断する必要がある
- **注意点**
  - ✓ 海外のDBは日本の薬事承認やガイドラインなどの状況を反映していない
  - ✓ がんで報告がある = ドライバー遺伝子変異とは言えないことに留意する
  - ✓ DBで報告数の少ない体細胞バリエーションの病原性判定は難しい場合がある
  - ✓ DBによりカバーされる遺伝子やバリエーション、データの信頼性、エビデンスの区分などが異なるために、複数のDBを参照し統合的に判断する必要がある

## 主にがんの体細胞バリエーションの病原性や治療法決定に有用なDBの例 (2022年6月8日現在)

Webサイト	URL	特徴
<b>Catalogue Of Somatic Mutations In Cancer (COSMIC)</b>	<a href="https://cancer.sanger.ac.uk/cosmic">https://cancer.sanger.ac.uk/cosmic</a>	英国Sanger研究所が運営する、様々なプロジェクトで収集されたがんの体細胞バリエーションとその関連情報を集積した大規模知識DB。現在v96が最新。既知のバリエーション以外に、遺伝子の構造や機能情報なども統合的に得られる
<b>Clinical Interpretations of Variants in Cancer (CIViC)</b>	<a href="https://civicdb.org/home">https://civicdb.org/home</a>	ワシントン大学で開発された、バリエーションの臨床的解釈用の知識DB。研究者参加型でキュレーションを行い、3,171バリエーション、471遺伝子、490薬剤、8,859エビデンス情報、などが提供される
<b>OncoKB</b>	<a href="https://www.oncokb.org/">https://www.oncokb.org/</a>	メモリアル・スローンケタリングがんセンター（MSKCC）で開発された、バリエーションのエビデンスレベルの知識DB。688遺伝子、5,730バリエーション、133薬剤の組み合わせのエビデンスレベルを提供
<b>Clinical Knowledgebase (CKB)</b>	<a href="https://www.jax.org/clinical-genomics/ckb">https://www.jax.org/clinical-genomics/ckb</a>	ジャクソン研究所で開発された、バリエーションの機能への影響、治療、治験情報を提供する知識DB
<b>Cancer Genome Interpreter (CGI)</b>	<a href="https://www.cancergenomeinterpreter.org/home">https://www.cancergenomeinterpreter.org/home</a>	がんの体細胞バリエーションの機能評価（ドライバーかどうか）、対応する薬剤、治療エビデンスレベルの解釈を、他の知識DBのデータなども参照して自動化する多目的プラットフォーム

# 臨床的意義づけの過程 (1)

遺伝性疾患の原因となる生殖細胞系列バリエントの臨床的意義づけの場合のエビデンス基準を説明する

● **ACMG\*-AMPガイドライン(2015)による病原性分類のエビデンス**

- ✓ (一般) 集団中の頻度データ
- ✓ 機能有害性の予測データ
- ✓ 機能情報 (実験的解析、hot spotや機能ドメイン内など)
- ✓ Segregation (分離 = 家系内でのバリエントと表現型の一致) の有無や程度
- ✓ 顕性遺伝疾患の場合、新生変異 (*de novo*) か
- ✓ 潜性遺伝疾患の場合、既知の病的バリエントの対側アレルにあるか
- ✓ 表現型を説明できる遺伝子に同定されたバリエントか

などの項目ごとにVery strongからSupportingまで異なるレベルのエビデンスを集積して分類に用いる

(Richards S, et al. Genet Med. 2015;17:405-424.)

\***ACMG** : American College of Medical Genetics (米国臨床遺伝・ゲノム学会)

**ACMG-AMPガイドラインでの病原性に関するエビデンス (病的基準のみ抜粋)**

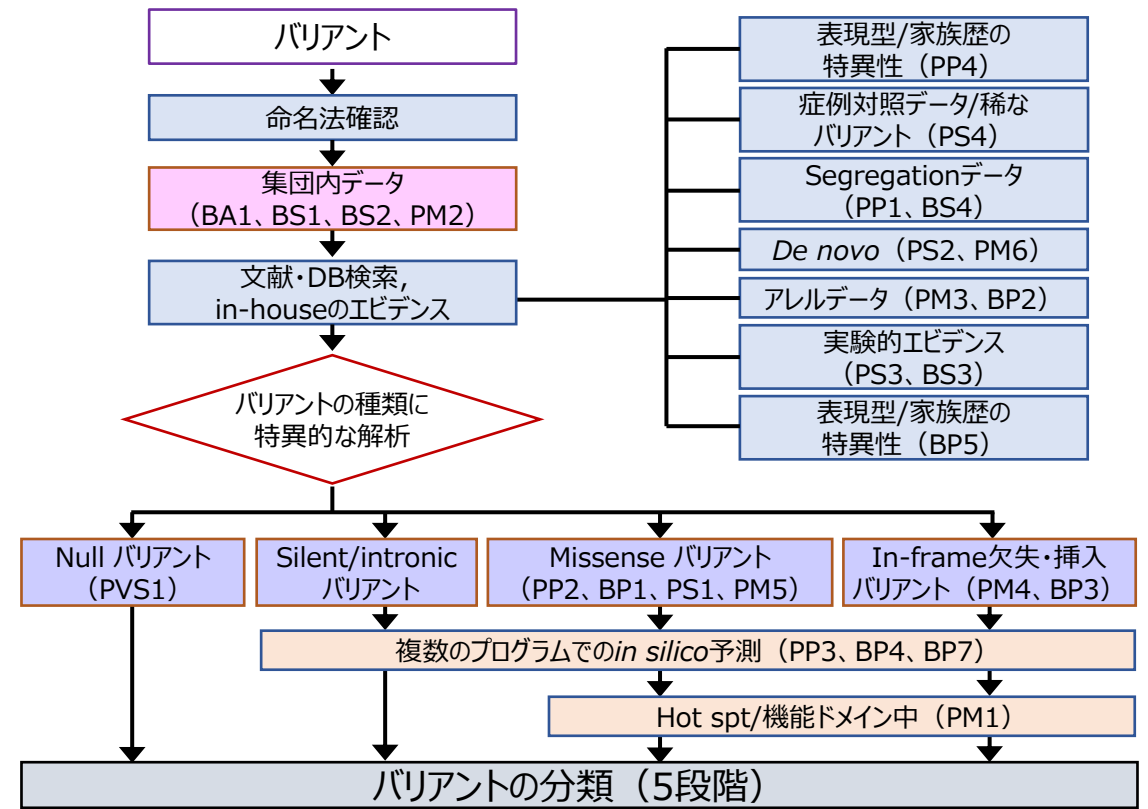
根拠の強さ	小基準	基準
Very strong	PVS1	疾患の発症に機能喪失が知られている遺伝子における予測されたNullバリエント (ナンセンス、フレームシフト、エクソンから±2bpのスプライス部位、開始コドンなどのバリエントや、単一もしくは複数エクソンの欠失)
Strong	PS1	過去に病的バリエントとして確立しているミスセンスバリエントと同一のアミノ酸置換を起こすバリエント
	PS2	罹患のない両親に認められないことが確認された <i>de novo</i> のバリエント
	PS3	よく確立された <i>in vivo</i> もしくは <i>in vitro</i> の機能解析で機能に有害な影響を与えることが示されたバリエント
Mode-rate	PS4	罹患患者集団におけるバリエント頻度が健常対照集団における頻度よりも有意に上昇しているバリエント
	PM1	変異のホットスポットやよく確立された機能ドメインに位置する良性バリエントではないバリエント
		PM2
	PM3	劣性遺伝疾患において、他の既知の病的バリエントと <i>trans</i> (反対側の染色体) に認められたバリエント
		PM4
	PM5	別のアミノ酸置換を伴うミスセンスバリエントが過去に病的バリエントとして報告されているミスセンスバリエント
PM6		両親に認められないことの確認が取れていない <i>de novo</i> バリエント
Supporting	PP1	複数の家系内罹患患者で認められ、非罹患患者で認められないバリエント
	PP2	ミスセンスバリエントが疾患の主原因で、良性のミスセンスバリエントの頻度が低い遺伝子におけるミスセンスバリエント
	PP3	複数の機能予測プログラムで遺伝子や遺伝子産物に対し病原性が予測されるバリエント
	PP4	患者の表現型や家族歴から原因が予測される単一の遺伝子に認められるバリエント
	PP5	信頼できる情報源が最近病的であると報告しているバリエント

# 臨床的意義づけの過程 (2)

遺伝性疾患の原因となる生殖細胞系列バリエントのエビデンスに基づく病原性分類の方法（病原性の付与）について説明する

- 遺伝性疾患の原因となるバリエントの絞り込みの特徴
  - ✓ 主に機能喪失型のバリエントを抽出していく
  - ✓ 特定の遺伝性疾患特有の表現型、遺伝形式、家族の罹患状況とバリエントの有無の一致性 (segregation) などの情報も有用になる
- ACMG-AMPガイドラインによる病原性分類
  - ✓ 何項目エビデンスが当てはまるかで5段階に分類する
    - Pathogenic/likely pathogenic : 病的
    - Uncertain(Unknown) significance (VUS) : 分類不能
    - Likely benign /benign : 非病的

生殖細胞系列のバリエントの病原性を体系的に解釈するためのワークフロー



ACMG-AMPガイドラインでのバリエントの病原性分類基準 (病的基準のみ抜粋)

Pathogenic	PVS	PS					
	PVS	PM	PM				
	PVS	PM	PP				
	PVS	PP	PP				
	PS	PS					
	PS	PM	PM	PM			
	PS	PM	PM	PP	PP		
Likely pathogenic	PS	PM	PP	PP			
	PS	PM	PM				
	PS	PP	PP				
	PM	PM	PM				
	PM	PM	PP	PP	PP		PP
	PM	PP	PP	PP	PP		PP
	PM	PP	PP	PP	PP		PP

がん特異的な体細胞バリエントの病的意義の分類を記す

## ● ドライバー遺伝子バリエントのパスセンジャーからの区別

- ✓ **がん抑制遺伝子**の場合は、機能喪失型（不活性化）バリエント

例：停止コドンへの変化・フレームシフト変異・スプライシング部位の変異・SVなどによる遺伝子産物の短縮化（Truncation）、大きな欠失

- ✓ **がん遺伝子**の場合は、機能獲得型（活性化）バリエント

例：ミスセンス変異、遺伝子産物の短縮化による分子内機能阻害の解除、遺伝子増幅、SVによる融合遺伝子

- ✓ 予測モデルや既知バリエントとの比較など様々なバイオインフォマティクスツールによりドライバー遺伝子バリエントの抽出の自動化が可能である

(Rogers MF, et al. Brief Bioinform. 2021;22:bbaa250 .

Nakagawa H, Fujita M. Cancer Sci. 2018;109:513-522.)

- ✓ その他、Tumor mutation burden (**TMB**)、マイクロサテライト不安定性 (Microsatellite instability、**MSI**)、Homologous recombination defect (**HRD**)、**Mutational signature**など、特定の遺伝子の変化によらないバリエントの出現パターンを指標とした治療選択マーカーがある（右記、[入門編第VI章参照](#)）

## ● **MSI-H（高頻度マイクロサテライト不安定性/Microsatellite Instability）**

- ✓ ミスマッチ修復遺伝子の機能不全により、後天的にマイクロサテライトリピート数の変化が高頻度に認められる状態
- ✓ 免疫チェックポイント阻害薬の治療効果予測マーカー
- ✓ 全ゲノム解析用のMSI検出プログラムとして、MIMcallなどがある (Fujimoto A, et al. Genome Res. 2020;30:334-346.)

## ● **TMB（腫瘍遺伝子変異量/Tumor Mutation Burden）**

- ✓ 腫瘍ゲノムで後天的に獲得された体細胞変異の総量を意味し、解析領域1Mbあたりの体細胞変異数で示す
- ✓ ネオアンチゲンの代替バイオマーカー（surrogate biomarker）として、免疫チェックポイント阻害薬への反応性を予測できる可能性

## ● **HRD（相同組換え修復欠損/Homologous Recombination Defect）**

- ✓ 相同組み換え遺伝子の機能不全によりコピー数やSVが増加する
- ✓ PARP阻害剤の治療効果予測マーカーとして適応判定に用いる
- ✓ 全ゲノム解析用のHRD推定プログラムとしてCHORDなどがある (Nguyen L, et al. Nat Commun 2020;11:5584. )

## ● **Mutational signature（変異スペクトラム）**

- ✓ 体細胞変異の組み合わせを指し、多くは既知の発がん要因と関連している
- ✓ 治療や予後推定にも有用な可能性が報告されている
- ✓ プロファイリングプログラムとして、SigProfiler toolsなどがある COSMIC (<https://cancer.sanger.ac.uk/signatures/tools/>)

# 臨床的意義づけの過程 (4)

がん特異的な体細胞バリエントの臨床的意義づけ（エビデンスに基づきActionableかどうかを判断する）について説明する

● がん特異的な体細胞バリエントの臨床的意義に基づく分類

- ✓ 機能への影響や介入可能性（治療薬）の有無に基づき分類する
- ✓ **ASCO**\*1、**CAP**\*2、**AMP**\*3による4段階の分類システム（右図）など、様々な臨床的意義づけの評価法がある
- ✓ 薬剤の開発状況や臨床試験の結果により介入可能性などのエビデンスは常に変化するため、治療、診断、予後におけるバリエントの臨床的意義は継続的に再評価が必要である



治療選択に関する詳細は入門編第VI章参照

- ✓ 臨床的意義づけまでのレポーティングを行う有償のInterpretationツールとして、QCI（Qiagen社）、SOPHiA DDM（SOPHiA GENETICS社）、MH Guide（Molecular Health社）なども利用可能である

\*1 **ASCO** : American Society of Clinical Oncology（米国臨床腫瘍学会）

\*2 **CAP** : College of American Pathologist（米国病理医協会）

\*3 **AMP** : Association for Molecular Pathology（米国分子病理学会）

\*4 **FDA** : Food and Drug Administration（米国食品医薬品局）

**ASCO、CAP、AMPが作成した臨床的意義に基づく体細胞バリエント分類**

Tier 1	Tier 2	Tier 3	Tier 4
強い臨床的意義をもつバリエント	臨床的意義の可能性のあるバリエント	臨床的意義の不明なバリエント	良性か良性の可能性が高いバリエント
<p><u>Level A エビデンス</u></p> <p>専門家のガイドラインに含まれる<b>FDA</b>*4承認薬がある</p>	<p><u>Level C エビデンス</u></p> <p>異なるがん種でFDA承認薬や治験薬がある</p> <p>一定のコンセンサスが得られた複数の小規模な研究が報告されている</p>	<p>一般集団や特定の亜集団のDB、あるいは全がんや特定のがん特異的なバリエントDBで、一定頻度検出されない</p> <p>がんとの関連を明示するエビデンスが報告されていない</p>	<p>一般集団や特定の亜集団のDBで、一定頻度検出される</p> <p>がんとの関連を示すエビデンスが報告されていない</p>
<p><u>Level B エビデンス</u></p> <p>当該分野の専門家のコンセンサスを得た十分な症例数のある臨床試験がある</p>	<p><u>Level D エビデンス</u></p> <p>前臨床試験で治療的意義あり、またはコンセンサスのない少数の症例報告がある</p>		

(Li MM, et al. J Mol Diagn. 2017;19:4-23.)

# 臨床的意義づけの過程（5）

Tier 1やTier 2に含まれてくるがん特異的な体細胞バリエントの特徴を説明する

臨床的意義を持つかその可能性のあるがん特異的な体細胞バリエントが示す特徴の例

特徴	強い臨床的意義をもつバリエント	臨床的意義の可能性のあるバリエント
FDA承認治療薬、専門家のガイドライン、研究的治療	治療：FDA承認治療薬や強い根拠のある研究的治療がある 診断：専門家のガイドラインにあるかコンセンサスのある根拠の報告がある 予後：専門家のガイドラインにあるかコンセンサスのある根拠の報告がある	治療：異なる臓器でFDA承認治療薬がある、またはある程度根拠のある研究的治療がある 診断：専門家のガイドラインにはないが信頼できる報告がある 予後：専門家のガイドラインにはないが信頼できる報告がある
バリエントの種類	活性型、不活性型（ミスセンス、ナンセンス、欠失挿入、スプライシング）、コピー数異常、融合遺伝子	活性型、不活性型（ミスセンス、ナンセンス、欠失挿入、スプライシング）、コピー数異常、融合遺伝子
バリエントのアレル頻度	ほとんどがモザイク（がん組織中のがん細胞だけで見られる）	ほとんどがモザイク（がん組織中のがん細胞だけで見られる）
集団内検出頻度（集団DB）	ないか極めて稀	ないか極めて稀
機能予測プログラムのスコア	ほとんどが機能破壊性だが参考程度	ほとんどが機能破壊性だが参考程度
生殖細胞系列での病的バリエントDB	存在することがある	存在することがある
体細胞（がん）での病的バリエントDB	ほとんどが存在する	存在することが多い
パスウェイへの関与	疾患関連パスウェイ上にある	疾患関連パスウェイ上にある
文献（機能解析、集団解析、など）	治療：コンセンサスのある根拠の報告がある 診断：コンセンサスのある根拠の報告がある 予後：コンセンサスのある根拠の報告がある	治療：異なる臓器でFDA承認治療薬を使った報告（目的外使用での効果検証）があるかいくつかの臨床試験の報告がある 診断：コンセンサスのある根拠の報告がある 予後：コンセンサスのある根拠の報告がある



# 第 VI 章

## 臨床におけるがんゲノム医療 (抗がん剤選択、治験情報活用)

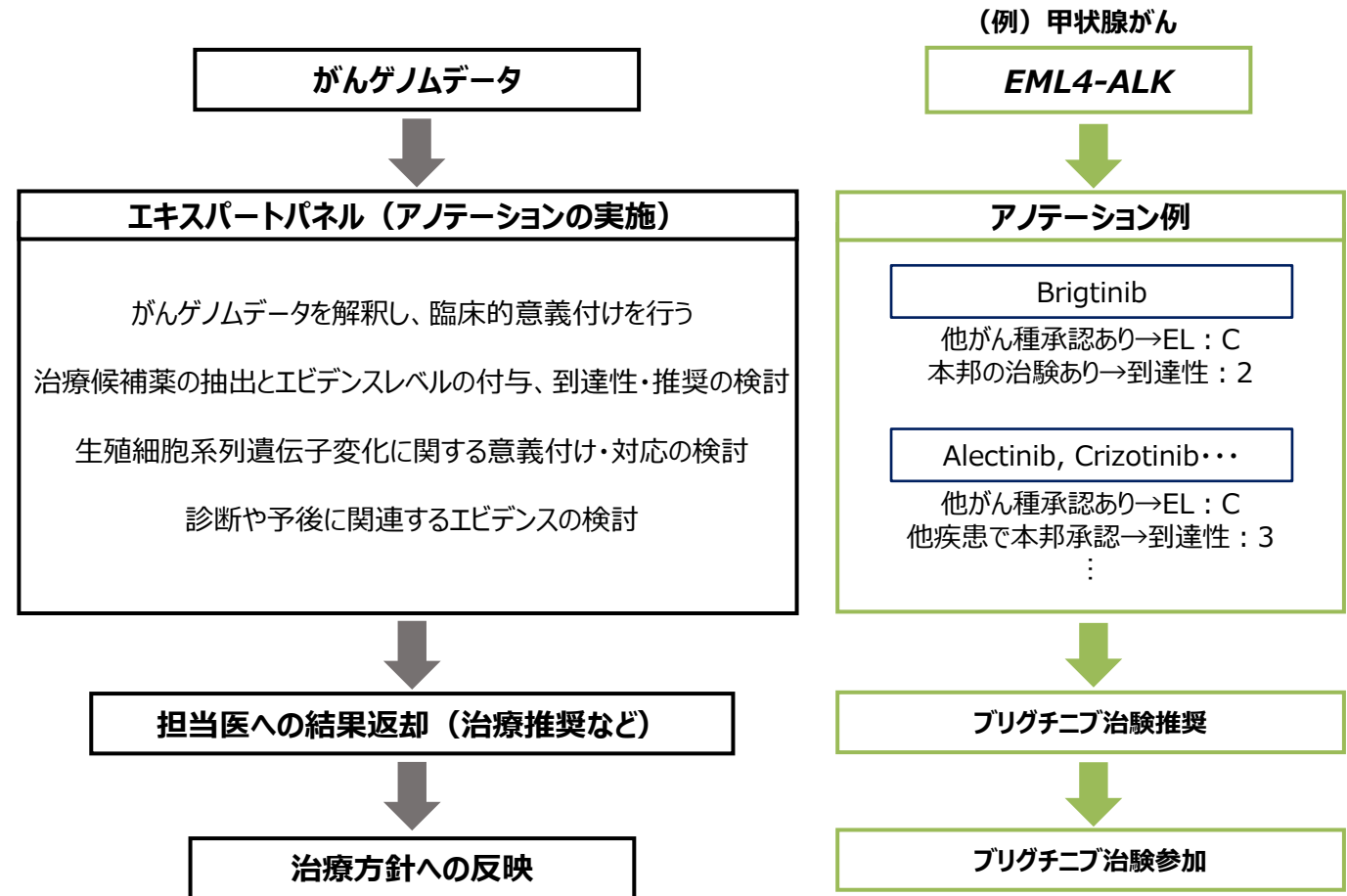
国立がん研究センター東病院トランスレーショナルリサーチ支援室

藤澤孝夫

# がんゲノム医療とエキスパートパネル

下記にて、がんゲノム医療とエキスパートパネルについて説明する

- 遺伝子パネル検査などのがん遺伝子解析の結果を元に治療方針を検討することを**がんゲノム医療**という
- がんゲノム医療の流れは、①遺伝子解析の実施、②遺伝子解析結果に基づく候補治療薬剤の検討、③候補治療薬剤による治療の3段階からなる
- 遺伝子パネル検査で得られる結果は遺伝子バリエーションやバイオマーカーそのものであり、実際の治療方針に結びつけるためには臨床的意義付け（**アノテーション**）が必要となる
- 遺伝子解析結果とそれに対するアノテーションについて総合的に検討する会議体を「**エキスパートパネル**」と呼ぶ
- 遺伝子解析などの分子プロファイリング結果に基づき治療方針を検討する会議体のことを“Molecular Tumor Board”と呼び、エキスパートパネルもMolecular Tumor Boardの一型である



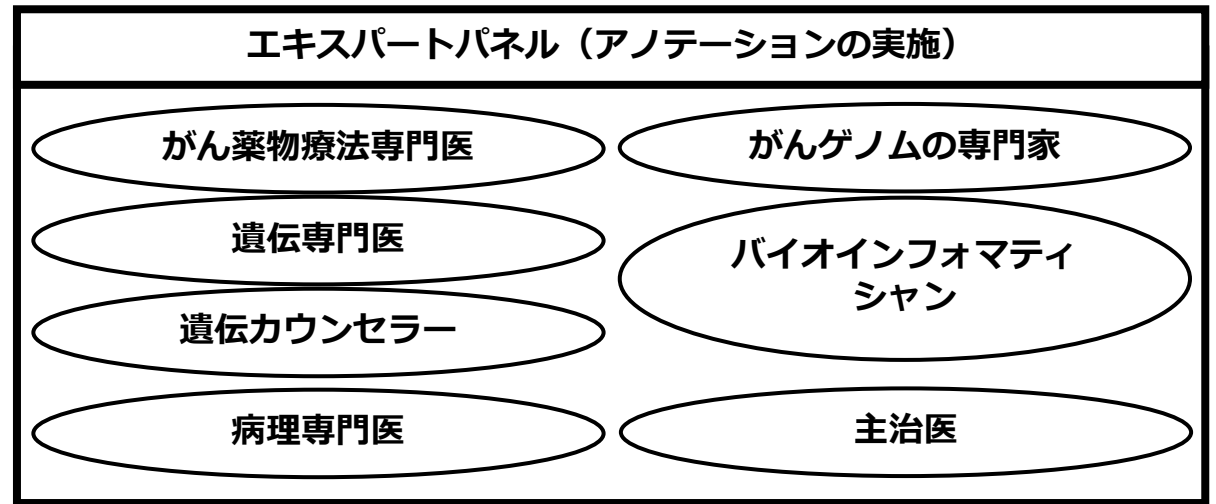
# エキスパートパネル（1）構成要件

まず、エキスパートパネルの構成要件について説明する

- 厚生労働省が指定する「がんゲノム医療中核拠点病院」の指定要件によると、**エキスパートパネルの構成員**には下記を含む必要がある
  1. がん薬物療法に関する専門的な知識および技能を有する診療領域の異なる常勤の医師
  2. 遺伝医学に関する専門的な知識及び技能を有する医師
  3. 遺伝医学に関する専門的な遺伝カウンセリング技術を有する者
  4. 病理学に関する専門的な知識及び技能を有する常勤の医師
  5. 分子遺伝学やがんゲノム医療に関する十分な知識を有する専門家
  6. 次世代型シーケンサーを用いた遺伝子解析等に必要バイオインフォマティクスに関する十分な知識を有する専門家
  7. 検討を行う対象患者の主治医または当該主治医に代わる医師
- 下記の条件を満たす場合には、リアルタイムで協議可能な方法でのエキスパートパネルの開催は必要とせず、ファイル共有サービスや電子メールを介した持ち回り協議が可能
  1. 遺伝子異常が検出されなかった場合
  2. 検出されたすべての遺伝子異常が以下の(1)～(4)のいずれかに該当する場合

(\*エビデンスレベルは「次世代型シーケンサー等を用いた遺伝子パネル検査に基づくがん診療ガイドライン 第2.1版」での定義による)

  - (1) エビデンスレベルA\*の遺伝子異常
  - (2) エビデンスレベルR\*の遺伝子異常
  - (3) MSI-High
  - (4) TMB-High



(がんゲノム医療推進コンソーシアム運営会議、平成31年3月8日)

# エキスパートパネル（2） 検討内容

エキスパートパネルでの検討内容には、検査全体の質や遺伝子異常ごとのアノテーション内容が含まれる

- エキスパートパネルで最低限議論すべき内容は下記とされる

- ✓ 検査全体に関して  
検体およびデータの品質について
- ✓ 各遺伝子異常に関して
  - ① 遺伝子異常に対する生物学的意義付け（がん化能など特定の形質獲得に寄与するかどうか等）を行う
  - ② 遺伝子異常に対応する治療薬を確認する
  - ③ 患者基本情報（年齢・性別・がん種など）を考慮した上で遺伝子異常に対応する具体的な候補薬とエビデンスレベル、及び承認状況や治験状況を踏まえたavailabilityランクを付ける
  - ④ 必要に応じて③で挙げられた候補薬剤から、患者状態やavailabilityを考慮して、優先的に推奨されるものがあるかなどについて検討する
  - ⑤ 診断や予後に関するエビデンスの解釈
  - ⑥ 生殖細胞系列遺伝子異常を認める（または疑われる）場合は、関連するガイドライン・ガイダンス・提言に従い、その意義付け及び対応について検討する

## エキスパートパネルで議論すべき内容

- 検査全体に関して  
検体およびデータの品質
- 各遺伝子異常に関して
  - ① 遺伝子異常に対する生物学的意義付け
  - ② 遺伝子異常に対応する治療薬
  - ③ 遺伝子異常に対応する具体的な候補薬とエビデンスレベル、到達性
  - ④③で挙げられた候補薬剤について推奨されるものがあるかの検討
  - ⑤ 診断や予後に関するエビデンスの解釈
  - ⑥ 生殖細胞系列遺伝子異常に関する意義付け・対応

（がんゲノム医療推進コンソーシアム運営会議、平成31年3月8日）

- ✓ 利用する知識データベースはキュレーションがどのように実施されるかによって情報の質が大きく左右される。キュレーションの方法（オート、マニュアル等）や、キュレーターの質といった実施体制が公開された知識ベースであることが望ましい。また、古い情報を元に誤った意義付けを行うリスクを避ける為、定期的に知識ベースの情報が更新されており、併せて古い情報の整理が実施されている知識ベースを選ぶことが望ましい
- ✓ 知識ベースはupdateされるため、使用したversionの明記が望ましい

# エキスパートパネル（3） 報告内容

エキスパートパネルでの議論内容のうち、報告すべき内容には治療候補薬や二次的所見に関連する情報などが含まる

- エキスパートパネルでの議論内容のうち、報告書または診療録に記載すべき内容は下記とされる
  - ✓ 検体およびデータの品質
  - ✓ 検出された遺伝子異常の生物学的意義付け
  - ✓ 遺伝子異常に対応する具体的な候補薬とエビデンスレベル
  - ✓ 候補薬の適応状況や治験情報を踏まえたavailabilityランク
  - ✓ 二次的所見の有無とその意義付け
  - ✓ レポートの記載範囲
  - ✓ 意義付けに用いた知識データベースの内訳とそのアクセス日
  - ✓ 臨床的意義づけは、網羅性を持ったものではなく、今後変わりうるものである、という留意事項

注：造血器腫瘍においては、「診断」や「予後」における遺伝子異常のエビデンスをもとに、治療薬の選択や幹細胞移植治療の適応が決まることがある為、日本血液学会「造血器腫瘍ゲノム検査ガイドライン」に記載されている「診断」や「予後」に関連した遺伝子異常のエビデンスもあわせて記載すべきである

## エキスパートパネルの報告書に記載すべき内容

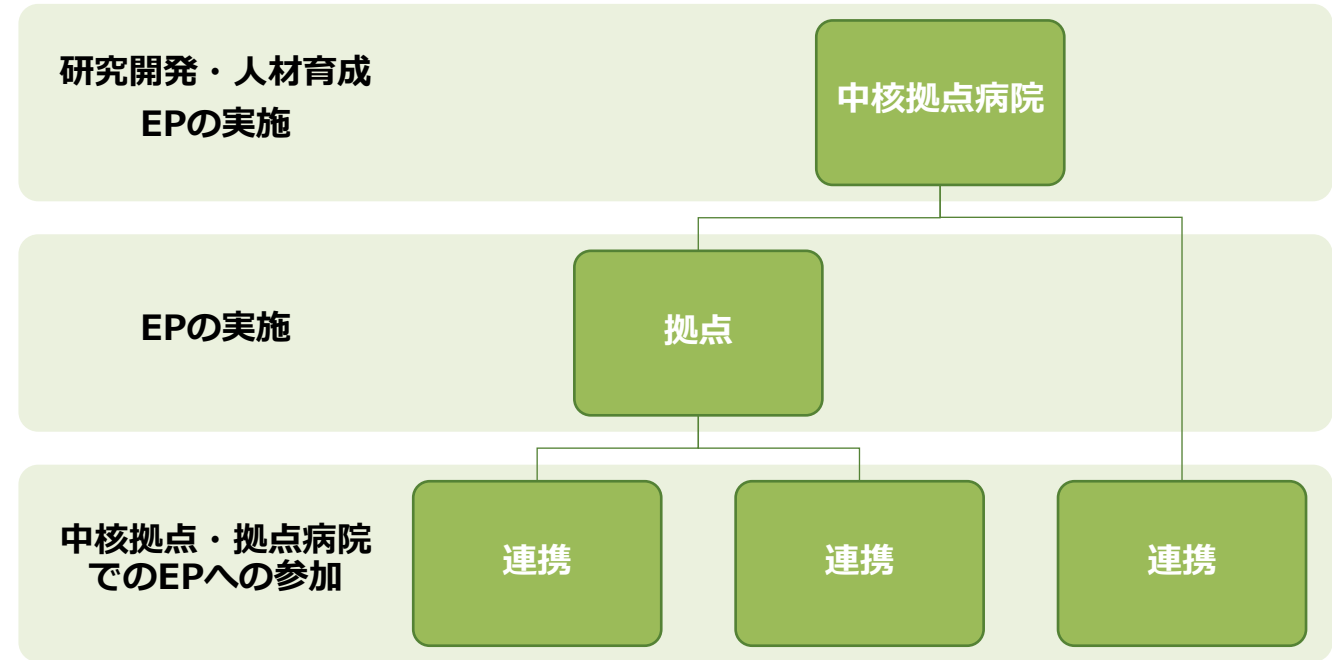
- 検体およびデータの品質
- 検出された遺伝子異常の生物学的意義付け
- 遺伝子異常に対応する具体的な候補薬とエビデンスレベル
- 候補薬の適応状況や治験情報を踏まえたavailabilityランク
- 二次的所見の有無とその意義付け
- レポートの記載範囲
- 意義付けに用いた知識データベースの内訳とそのアクセス日
- 臨床的意義づけは、網羅性を持ったものではなく、今後変わりうるものである、という留意事項

(がんゲノム医療推進コンソーシアム運営会議、平成31年3月8日)

# エキスパートパネル（４） 実施施設の現状

がんゲノム医療の実施施設は、がんゲノム医療中核拠点病院・拠点病院・連携病院の3種類からなる

- がんゲノム医療の実施施設は、がんゲノム医療中核拠点病院・拠点病院・連携病院の3種類からなる
- これら施設の業務は検査の患者説明・検体準備・シーケンス実施（外注可）、エキスパートパネル開催、レポート作成、結果の患者説明、治療などのほか、研究開発や人材育成が含まれる
- 中核拠点病院は上記の全てを実施可能である
- 拠点病院は研究開発・人材育成については中核拠点病院と連携するが、その他の全てが実施可能である
- 連携病院では研究開発・人材育成については中核拠点病院と連携するほか、エキスパートパネルは中核拠点病院または拠点病院で実施されるものに参加する形となる
- これら施設については厚生労働大臣への申請が必要である
- 2022年5月1日時点で、12の中核拠点病院・33の拠点病院・188の連携病院が存在する
- がんゲノム医療の拡大に伴い、エキスパートパネル実施施設への負担が懸念される

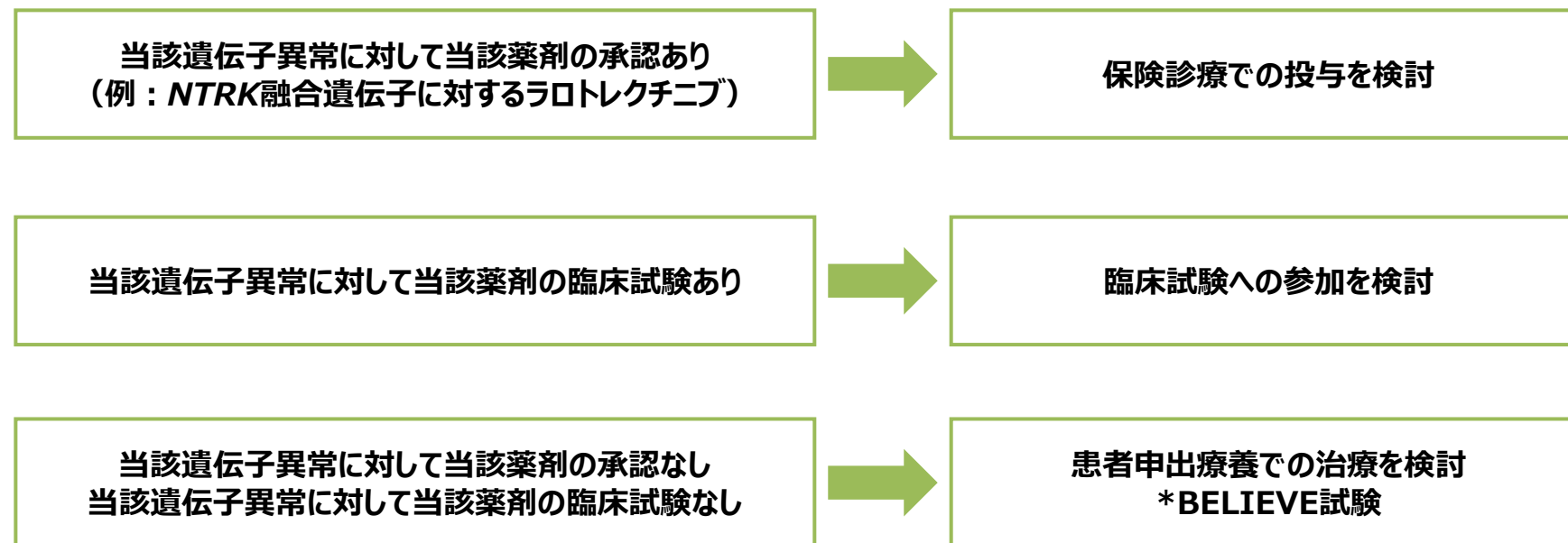


(がんゲノム医療推進コンソーシアム運営会議,令和元年12月5日)

# がんゲノム医療での治療様式

本邦では、がんゲノム医療により行われる治療様式として保険診療、臨床試験、患者申出療養などがある

- 検出された遺伝子バリエントと候補薬剤について承認がある場合は保険診療での実施が可能となる  
（\*遺伝子パネル検査がコンパニオン診断薬でない場合も、エキスパートパネルが推奨した場合はあらかじめのコンパニオン検査は実施不要\*）
- 検出された遺伝子バリエントと候補薬剤に対応する臨床試験が存在する場合は、臨床試験での実施が検討される
- 検出された遺伝子バリエントに対する候補薬剤の投与が保険診療・臨床試験下での実施が困難な場合には、患者申出療養制度を用いれば薬剤費用以外を保険診療として実施することが可能であるが、（1）患者申出療養制度には申請の困難さ、（2）高額な薬剤費用による経済的負担などの問題がある。上記2点の解消のため、企業による薬剤無償提供を背景とした患者申出療養に対する受け皿試験（「遺伝子パネル検査による遺伝子プロファイリングに基づく複数の分子標的治療に関する患者申出療養（NCCH1901：BELIEVE試験）」）が実施されている
- がんゲノム医療における治療到達性は概ね10-20%にとどまり、今後の検討課題である



# エビデンスタイプの種類と到達度

遺伝子バリエーションの解釈のエビデンスには腫瘍原性、疾患素因関連、治療効果予測、診断、予後予測の5つのタイプがある

- 遺伝子バリエーションの解釈には主に以下の5つのエビデンスタイプがある（右表）
  - (1) 腫瘍原性エビデンス (Oncogenic)
  - (2) 疾患素因関連エビデンス (Predisposing)
  - (3) 治療効果予測エビデンス (Predictive)
  - (4) 診断エビデンス (Diagnostic)
  - (5) 予後予測エビデンス (Prognostic)
- このうち治療効果予測エビデンスは、治療候補薬の検討に直結するため重要である
- また、疾患素因関連エビデンスに関しても生殖細胞系列遺伝子変化に関連するためエキスパートパネルでの重要な検討事項となる

エビデンスタイプ	内容
腫瘍原性 (Oncogenic)	細胞のがん化に寄与する遺伝子変化のことを意味する。同定された遺伝子変化が腫瘍の発生原因となるドライバー変異であるのか、あるいは偶発的に生じたパッセンジャー変異なのかどうか重要な観点になる
疾患素因関連 (Predisposing)	がん化関連の生殖細胞系列遺伝子変化がこのカテゴリーに入る
治療効果予測 (Predictive)	薬剤や放射線治療などの感受性や耐性に関連するマーカー。バリエーションが薬剤感受性の場合、承認薬標的マーカーであるかや臨床試験の受け入れ基準になっているかといった治療法へのアクセスが可能かどうかの情報は治療選択上の重要な情報となる
診断 (Diagnostic)	患者の診断に関連するマーカー
予後予測 (Prognostic)	がんの進行、重症度、生命予後などに関連するマーカー

(次世代型シーケンサー等を用いた遺伝子パネル検査に基づくがん診療ガイドライン 第2.1版)



# 治療効果予測エビデンスレベル

治療効果予測エビデンスは、その重要性に鑑みレベル付け（エビデンスレベル）がなされ、治療方針の検討に重要である

- このうち治療効果予測エビデンスは、遺伝子バリエーションの有無による薬剤感受性を示しており、治療候補薬の検討に直結する為、重要である
- 一方、治療候補薬の検討においてはエビデンスの内容による重み付けが必要であり、この指標としてエビデンスレベルが設定されている
- エビデンスレベルの定義は日米欧により異なる。米国では米国臨床腫瘍学会（**ASCO**）、米国病理医協会（College of American Pathologists；**CAP**）、米国分子病理学会（Association for Molecular Pathology；**AMP**）の3学会が合同で作成したRecommendationが存在し、欧州では欧州臨床腫瘍学会（**ESMO**）によるthe ESMO Scale for Clinical Actionability of molecular Target（**ESCAT**）が発表されている
- 本邦では日本における薬事承認や治療ガイドラインでの推奨を組み込んだExpert Panel Working Group（**EPWG**）分類が汎用されている（右表）

基準	EL*1	EL*1に基づく対応例
当該がん種、国内承認薬がある	A	国内承認薬がある場合はコンパニオン診断薬の結果も踏まえて治療薬の投与を推奨するが、各種ガイドラインに従う。それ以外では、治験・先進医療・薬価基準収載医薬品の適応外使用等の評価療養や患者申出療養等の保険外併用療養費制度の利用を推奨するが、各種ガイドラインに従う
当該がん種、FDA承認薬がある	A	
当該がん種、ガイドラインに記載されている	A	
当該がん種、統計的信憑性の高い臨床試験・メタ解析と専門家間のコンセンサスがある	B	科学的根拠があり、治験・先進医療・薬価基準収載医薬品の適応外使用等の評価療養や患者申出療養等の保険外併用療養費制度の利用を考慮する
他がん種、国内またはFDA承認薬がある	C	科学的根拠があり、治験・先進医療・薬価基準収載医薬品の適応外使用等の評価療養や患者申出療養等の保険外併用療養費制度の利用を考慮する
他がん種、統計的信憑性の高い臨床試験・メタ解析と専門家間のコンセンサスがある	C	
がん種に関わらず、規模の小さい臨床試験で有用性が示されている	C	
がん種に関わらず、症例報告で有用性が示されている	D	科学的根拠が十分ではないが、治験・先進医療・薬価基準収載医薬品の適応外使用等の評価療養や患者申出療養等の保険外併用療養費制度の利用等を、EPのコンセンサスの基で考慮する
前臨床試験（in vitroやin vivo）で有用性が報告されている	E	一定の科学的根拠があるが、ヒトへの投与がないことから、当該薬剤の使用は積極的に推奨されないが、治験の対象マーカーとなっている場合はEP*2のコンセンサスを基に、治験への登録を考慮する。近い将来ELが上がるが見込まれるため、CCAT*3への情報登録を行う
がん化に関与することが知られている	F	現時点で治療選択に関する科学的根拠はないが、治験の対象マーカーとなっている場合はEPのコンセンサスを基に、治験への登録を考慮する。近い将来エビデンスレベルが上がるが見込まれるため、CCAT*3への情報登録を行う
薬剤耐性に関与することが知られている	R	EPのコンセンサスを基に、治療選択を決定する

\*1EL：エビデンスレベル、\*2EP：エキスパートパネル、\*3CCAT：がんゲノム情報管理センター

（次世代型シーケンサー等を用いた遺伝子パネル検査に基づくがん診療ガイドライン 第2.1版）

到達性とは、当該薬剤への到達のしやすさを到達手段（保険診療、臨床試験…）を加味しレベル付けしたものである

- 治療効果予測、特に薬剤感受性エビデンスのある場合においても、薬剤の承認状況や臨床試験の適格基準などは様々である。この為、薬剤への到達性という指標でマーカーと薬剤の関連性を評価することが望ましい
- このためアノテーションにおいてはエビデンスレベルのみならず薬剤到達性についても付与することが必要である
- 下表にEPWG分類による到達性の分類を示す
- 特に到達度1は保険診療での治療実施に直結するほか、到達度2は国内臨床試験への参加可能性を示唆するため非常に重要である

指標概要	番号	指標詳細
当該がん種、国内承認薬がある	1	当該がん種において、当該バイオマーカーを適応とした国内承認薬が存在する
当該がん種、国内臨床試験がある	2	当該がん種において、当該バイオマーカーを受け入れ基準とした国内臨床試験が存在する
他がん種、国内承認薬がある（適応外）	3	他がん種において、当該バイオマーカーを適応とした国内承認薬が存在する
当該がん種、海外臨床試験がある	4	当該がん種において、当該バイオマーカーを受け入れ基準とした海外臨床試験が存在する
がん種に関わらず、FDA承認薬がある	5	がん種に関わらず、当該バイオマーカーを適応としたFDA承認薬が存在する
上記以外	6	上記のどれにもあてはまらない

（次世代型シーケンサー等を用いた遺伝子パネル検査に基づくがん診療ガイドライン 第2.1版）

アノテーションにおいては、公開データベース（Database：DB）の活用が重要である

- アノテーションの実施には、（1）検出された遺伝子バリエーション・バイオマーカーの生物学的意義、（2）検出された遺伝子バリエーション・バイオマーカーと関連する治療候補薬剤とそのエビデンスレベルの検索、（3）検出された遺伝子バリエーション・バイオマーカーに対応する治療候補薬剤のAvailabilityの3点に関する情報を収集・吟味する必要があり、その情報量は膨大となりうる
- 上記に関連する知識を集約した公開DBとして下表のものが存在し、これらの活用が実際のアノテーションには重要である
- 本章では、治療候補薬剤の検索に関連する、治療効果予測エビデンスに関するDBと臨床試験に関するDBについて概説する

種類	具体例
バリエーションの生物学的意義に関するDB	ClinVar、COSMIC
治療効果予測エビデンスに関するDB	OncoKB、CIViC、Cancer Genome Interpreter
臨床試験に関するDB	ClinicalTrial.gov、UMIN-CTR、JapicCTI、jRCT

OncoKB、CIViC、Cancer Genome Interpreterの治療効果予測エビデンスについて説明する

- 遺伝子バリエーション・バイオマーカーと対応する治療候補薬剤およびそのエビデンスレベルの評価には、OncoKB、CIViC、CGIが有用である
- これらのDBをもとに治療候補薬剤のエビデンスを検索し、エビデンスレベルを評価することでアノテーションを行う

Webサイト	URL	特徴
<b>OncoKB</b> (Precision Oncology Knowledge Base)	<a href="https://www.oncokb.org/">https://www.oncokb.org/</a>	<ul style="list-style-type: none"><li>✓ 米国Memorial Sloan Kettering Cancer Center (MSKCC) によるがんゲノム医療の知識ベース</li><li>✓ バリエーションの治療効果予測・診断・予後予測エビデンスについて網羅している</li><li>✓ 特に治療効果予測エビデンスについてはFDAでの承認状況を含め網羅しており、候補薬剤の検討について有用である</li></ul>
<b>CIViC</b> (Clinical Interpretation of Variants in Cancer)	<a href="https://civicdb.org/home">https://civicdb.org/home</a>	<ul style="list-style-type: none"><li>✓ 米国ワシントン大学による、がんにおけるバリエーションの臨床的解釈のための知識ベース</li><li>✓ 研究者（キュレーター）によるエビデンスの追加が可能である</li><li>✓ 腫瘍原性・疾患素因関連・治療効果予測エビデンス・診断・予後予測エビデンスについて網羅しており、遺伝子内の個々のバリエーションに関するエビデンスについても充実している</li></ul>
<b>CGI</b> (Cancer Genome Interpreter)	<a href="https://www.cancergenomeinterpreter.org/home">https://www.cancergenomeinterpreter.org/home</a>	<ul style="list-style-type: none"><li>✓ スペインBarcelona Biomedical Genomics Labによる知識ベース</li><li>✓ バリエーションのドライバーとしての意義の評価や候補薬剤・治療エビデンスレベルについて検討が可能である</li></ul>

## ClinicalTrial.gov、UMIN、JapicCTIの臨床試験情報について説明する

- 遺伝子バリエーションと対応する治療候補薬剤の治療到達性の評価には、薬剤の承認状況の確認に加え臨床試験情報の収集が必要となる
- 臨床試験情報の収集には、ClinicalTrial.gov、UMIN-CTR、JapicCTI、jRCTなどのDBが有用である
- ClinicalTrial.govは海外臨床試験まで含めて網羅されている
- UMIN、JapicCTI、jRCTは国内臨床試験のみが対象であるが、国内臨床試験に関する網羅性は高い

Webサイト	URL	特徴
<b>ClinicalTrial.gov</b>	<a href="https://clinicaltrials.gov">https://clinicaltrials.gov</a>	<ul style="list-style-type: none"><li>✓ 米国NIHとFDAが共同で運用する臨床試験登録データベース</li><li>✓ 200以上の国からの臨床試験情報を提供する世界最大の臨床試験登録サイトである</li></ul>
<b>UMIN-CTR</b> (UMIN臨床試験登録システム)	<a href="https://upload.umin.ac.jp/cgi-open-bin/ctr/index.cgi?function=02">https://upload.umin.ac.jp/cgi-open-bin/ctr/index.cgi?function=02</a>	<ul style="list-style-type: none"><li>✓ 大学病院医療情報ネットワーク (UMIN) による臨床試験データベース</li><li>✓ 治験に加え、アカデミアによる臨床試験などの情報が充実している</li></ul>
<b>JapicCTI</b> (Japic Clinical Trials Information)	<a href="https://www.clinicaltrials.jp/cti-user/trial/Search.jsp">https://www.clinicaltrials.jp/cti-user/trial/Search.jsp</a>	<ul style="list-style-type: none"><li>✓ 日本製薬工業会加盟の製薬企業を中心とした団体である財団法人Japicによる臨床試験データベース</li><li>✓ 企業治験の情報が充実している</li></ul>
<b>jRCT</b> (臨床研究実施計画・研究概要公開システム)	<a href="https://jrct.niph.go.jp/search">https://jrct.niph.go.jp/search</a>	<ul style="list-style-type: none"><li>✓ 厚生労働大臣に対して実施計画の提出などの届出手続を行うためのシステム</li><li>✓ 特定臨床研究及び一部の治験等が登録されている</li></ul>

# データベース活用実践例：OncoKB

例として、甲状腺がんでのALK融合遺伝子に関する臨床レベルでの治療効果予測エビデンスを検索する  
OncoKBはFDA承認状況を含めた治療効果予測エビデンスがシンプルにまとめられており有用である

The screenshot shows the OncoKB website interface. At the top, there is a navigation bar with the OncoKB logo and links for Levels of Evidence, Actionable Genes, Cancer Genes, API / License, About, News, and FAQ. The main heading reads "Welcome to OncoKB™" and "MSK's Precision Oncology Knowledge Base". Below this, statistics are displayed: 688 Genes, 5730 Alterations, 133 Cancer Types, and 112 Drugs. A search bar is present with the placeholder text "Search Gene / Alteration / Drug". Below the search bar, there are four tabs: "Therapeutic Levels" (highlighted in blue), "Diagnostic Levels", "Prognostic Levels", and "FDA Levels". Under the "Therapeutic Levels" tab, five evidence levels are listed: Level 1 (FDA-approved drugs, 43 Genes), Level 2 (Standard care, 24 Genes), Level 3 (Clinical evidence, 31 Genes), Level 4 (Biological evidence, 25 Genes), and Level R1/R2 (Resistance, 11 Genes). A red box highlights Levels 1, 2, and 3. A blue arrow points from a text box on the right to the "Therapeutic Levels" tab. Another red arrow points from a text box on the right to the Level 3 category.

“Therapeutic Levels”が  
治療効果予測エビデンスに該当

Level 3までがclinical evidence  
(EPWG A-D)に相当→選択

<https://www.oncokb.org>

# データベース活用実践例：OncoKB

例として、甲状腺がんでのALK融合遺伝子に関する臨床レベルでの治療効果予測エビデンスを検索する  
OncoKBはFDA承認状況を含めた治療効果予測エビデンスがシンプルにまとめられており有用である

The screenshot shows the OncoKB interface. At the top, there's a navigation bar with 'OncoKB' logo and links for 'Levels of Evidence', 'Actionable Genes', 'Cancer Genes', 'API / License', 'About', 'News', and 'FAQ'. On the right, there's a search bar, 'Account', and the Memorial Sloan Kettering Cancer Center logo.

Below the navigation bar, there are filter sections. The 'Therapeutic' section is expanded, showing six levels of evidence: Level 1 (FDA-approved drugs, 43 Genes), Level 2 (Standard care, 24 Genes), Level 3 (Clinical evidence, 31 Genes), Level 4 (Biological evidence, 25 Genes), Level R1 (Standard care, 8 Genes), and Level R2 (Clinical evidence, 6 Genes). A blue arrow points to these levels with the text: "Therapeutic Levels"の Level 1, 2, 3を選択した状態.

Below the therapeutic levels, there are sections for 'Diagnostic (for hematologic malignancies only)', 'Prognostic (for hematologic malignancies only)', and 'FDA-Recognized Content'. A red box highlights the '56 actionable genes' dropdown menu, with a red arrow pointing to it from the text: 検索する遺伝子名を選択 (今回はALK).

Below the filters, there's a summary: 'Showing 284 clinical implications (56 genes, 55 cancer types, 3 levels of evidence)'. There are buttons for 'Associations' and 'Reset filters'.

The main table has columns for Level, Gene, Alterations, Cancer Types, and Drugs. The first few rows are:

Level	Gene	Alterations	Cancer Types	Drugs
1	ABL1	BCR-ABL1 Fusion	B-Lymphoblastic Leukemia/Lymphoma	Dasatinib
1	ABL1	BCR-ABL1 Fusion	B-Lymphoblastic Leukemia/Lymphoma	Imatinib
1	ABL1	BCR-ABL1 Fusion	B-Lymphoblastic Leukemia/Lymphoma	Ponatinib
1	ABL1	BCR-ABL1 Fusion	Chronic Myelogenous Leukemia	Asciminib
1	ABL1	BCR-ABL1 Fusion	Chronic Myelogenous Leukemia	Bosutinib
1	ABL1	BCR-ABL1 Fusion	Chronic Myelogenous Leukemia	Dasatinib
1	ABL1	BCR-ABL1 Fusion	Chronic Myelogenous Leukemia	Imatinib

<https://www.oncokb.org>

# データベース活用実践例：OncoKB

例として、甲状腺がんでのALK融合遺伝子に関する臨床レベルでの治療効果予測エビデンスを検索する  
OncoKBはFDA承認状況を含めた治療効果予測エビデンスがシンプルにまとめられており有用である

OncoKB™ Levels of Evidence Actionable Genes Cancer Genes API / License About News FAQ

Memorial Sloan Kettering Cancer Center

Therapeutic

- Level 1: FDA-approved drugs (1 Gene)
- Level 2: Standard care (1 Gene)
- Level 3: Clinical evidence (0 Genes)
- Level 4: Biological evidence (0 Genes)
- Level R1: Standard care (0 Genes)
- Level R2: Clinical evidence (0 Genes)

Diagnostic (for hematologic malignancies only)

Prognostic (for hematologic malignancies only)

FDA-Recognized Content

ALK

Select a cancer type

5 drugs

Showing 10 clinical implications (1 gene, 3 cancer types, 2 levels of evidence) Associations Reset filters

Level	Gene	Alterations	Cancer Types	Drugs
1	ALK	Fusions	Anaplastic Large-Cell Lymphoma ALK Positive	Crizotinib
1	ALK	Fusions	Non-Small Cell Lung Cancer	Alectinib
1	ALK	Fusions	Non-Small Cell Lung Cancer	Brigatinib
1	ALK	Fusions	Non-Small Cell Lung Cancer	Ceritinib
1	ALK	Fusions	Non-Small Cell Lung Cancer	Crizotinib
1	ALK	Oncogenic Mutations	Non-Small Cell Lung Cancer	Lorlatinib
2	ALK	Fusions	Inflammatory Myofibroblastic Tumor	Brigatinib

ALKを選択した状態

ALK遺伝子異常に関するエビデンスの一覧が表示  
→スクロール

<https://www.oncokb.org>



# データベース活用実践例：OncoKB

例として、甲状腺がんでのALK融合遺伝子に関する臨床レベルでの治療効果予測エビデンスを検索する  
OncoKBはFDA承認状況を含めた治療効果予測エビデンスがシンプルにまとめられており有用である

OncoKB™ Levels of Evidence Actionable Genes Cancer Genes API / License About News FAQ

Memorial Sloan Kettering Cancer Center

+ Prognostic (for hematologic malignancies only)

+ FDA-Recognized Content

ALK x Select a cancer type 5 drugs

Showing 10 clinical implications (1 gene, 3 cancer types, 2 levels of evidence) Associations Reset filters

Level	Gene	Alterations	Cancer Types	Drugs
1	ALK	Fusions	Anaplastic Large-Cell Lymphoma ALK Positive	Crizotinib
1	ALK	Fusions	Non-Small Cell Lung Cancer	Alectinib
1	ALK	Fusions	Non-Small Cell Lung Cancer	Brigatinib
1	ALK	Fusions	Non-Small Cell Lung Cancer	Ceritinib
1	ALK	Fusions	Non-Small Cell Lung Cancer	Crizotinib
1	ALK	Oncogenic Mutations	Non-Small Cell Lung Cancer	Lorlatinib
2	ALK	Fusions	Inflammatory Myofibroblastic Tumor	Brigatinib
2	ALK	Fusions	Inflammatory Myofibroblastic Tumor	Ceritinib
2	ALK	Fusions	Inflammatory Myofibroblastic Tumor	Crizotinib
2	ALK	Fusions	Inflammatory Myofibroblastic Tumor	Lorlatinib

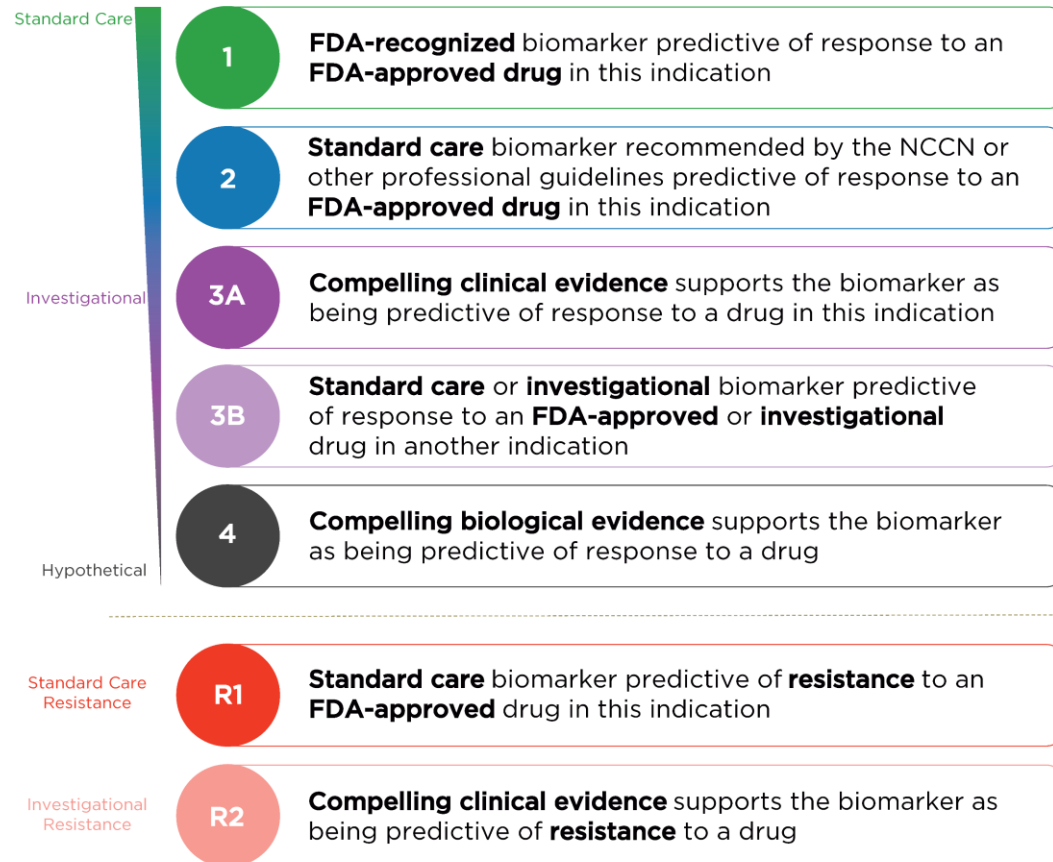
- 1：FDA承認あり→EPWG分類Aに相当
- 2：ガイドライン推奨あり→EPWG分類Bに相当
- 3：臨床的エビデンスあり→EPWG分類C/Dに相当（3Aが当該がん種、3Bが他がん種）

AlterationがFusionsでのエビデンス：  
Crizotinib, Alectinib, Brigatinib, Ceritinib, Lorlatinibについて他がん種でLevel 1/2のエビデンス  
→EPWG分類では他がん種Level A/B  
→EPWG分類でLevel Cに相当

<https://www.oncokb.org>

# データベース活用実践例：OncoKB

例として、甲状腺がんでのALK融合遺伝子に関する臨床レベルでの治療効果予測エビデンスを検索する  
OncoKBはFDA承認状況を含めた治療効果予測エビデンスがシンプルにまとめられており有用である



1：FDA承認あり→EPWG分類Aに相当  
2：ガイドライン推奨あり→EPWG分類Bに相当  
3：臨床的エビデンスあり→EPWG分類C/Dに相当  
(3Aが当該がん種、3Bが他がん種)

<https://www.oncokb.org/levels>

# データベース活用実践例：CIViC

例として、甲状腺がんでのALK融合遺伝子に関する臨床レベルでの治療効果予測エビデンスを検索する  
CIViCはエビデンス、遺伝子、バリエーションに関する情報が端的にまとめられており有用である

The screenshot shows the CIViC website interface. On the left, a dark navigation menu is highlighted with a red box, containing the following items: KNOWLEDGEBASE, Assertions, Evidence, Genes, Variants, Variant Groups, Clinical Trials, Diseases, Drugs, Phenotypes, Sources, and Variant Types. The main content area features a 'Become an Editor!' banner, a 'Knowledgebase Statistics' table, and sections for 'News & Events' and 'Live Curation Activity'.

Knowledgebase Statistics				
	Total	Weekly	Monthly	Yearly
Total Assertions	44			
Total Evidence	8,960			
Total Genes	472			
Total Variants	3,195			
Total Contributors	324			
Total Diseases	336			
Total Drugs	490			
Total Sources	3,171			
Total Revisions	33,138			
Total Comments	59,444			

Evidence (エビデンス)、  
Genes (遺伝子)  
Variants (バリエーション)  
など、さまざまな種類の情報が閲覧可

<https://civicdb.org/welcome>

# データベース活用実践例：CIViC

例として、甲状腺がんでのALK融合遺伝子に関する臨床レベルでの治療効果予測エビデンスを検索する  
CIViCはエビデンス、遺伝子、バリエントに関する情報が端的にまとめられており有用である

① 左の“KNOWLEDGEBASE”から“Genes”を選択

② “Names”に遺伝子名を入力

③ 遺伝子名をクリック

Name	Aliases	Diseases	Drugs	Count	Count	Count
ALK	ALK, ALK1 +2	Lung Non-small Cell... +25	Crizotinib +18	55	213	2

<https://civicdb.org/welcome>

# データベース活用実践例：CIViC

例として、甲状腺がんでのALK融合遺伝子に関する臨床レベルでの治療効果予測エビデンスを検索する  
CIViCはエビデンス、遺伝子、バリエントに関する情報が端的にまとめられており有用である

④ "Description"や"MyGeneInfo"に遺伝子の詳細の記載あり  
⑤ "ALK Variants"で"EML4-ALK"を選択

<https://civicdb.org/welcome>

# データベース活用実践例：CIViC

例として、甲状腺がんでのALK融合遺伝子に関する臨床レベルでの治療効果予測エビデンスを検索する  
CIViCはエビデンス、遺伝子、バリエントに関する情報が端的にまとめられており有用である

⑥

**Description**

The EML4-ALK fusion variant 1 consisting of ALK kinase domain (exons 20-29) fused to EML4 exons 1-13 is the most common EML4-ALK variant, and was discovered in non-small cell lung cancer. Multiple EML4 breakpoint shavbe described with differential sensitivity to inhibitors with variant 1 showing greater sensitivity than 3a in cell lines. EML4-ALK is crizotinib sensitive; however, several mutations that confer resistance mutations have been described in case studies. In the only clinical trial for crizotinib that included determination of EML4-ALK variant type in a subset of its participants, a very high response rate was observed, although the numbers were insufficient to validate correlation of variant type to outcome. Preclinical studies with this variant have indicated sensitivity to Hsp90 inhibitors.

EID	Disease	Drugs	DIT	DESC	EL	ET	ED	CS	VO	R
EID262	Lung Non-small Cell...	Crizotinib	N/A	≡	C	👁	👍	👍	...	4★
EID1207	Lung Non-small Cell...	Crizotinib	N/A	≡	C	👁	👍	👍	...	4★
EID7482	Colorectal Cancer	Nivolumab	N/A	≡	C	👁	👎	👎	...	3★
EID1121	Lung Acinar Adenocarcinoma	Erlotinib	N/A	≡	C	👁	👍	👎	...	2★

⑥“EML4-ALK”に関する詳細を“Description”で閲覧可能

<https://civicdb.org/welcome>

# データベース活用実践例：CIViC

例として、甲状腺がんでのALK融合遺伝子に関する臨床レベルでの治療効果予測エビデンスを検索する  
CIViCはエビデンス、遺伝子、バリエントに関する情報が端的にまとめられており有用である

①左の“KNOWLEDGEBASE”から“Evidence”を選択  
②“Names”に遺伝子名を入力  
③“ET”（Evidence Type）を選択  
④“Predictive”が効果予測エビデンスに相当

EID	Gene	Variant	Disease	Drugs	DESC	EL	ET	ED	CS	VO	R
EID2344	ALK	F1174L	Neuroblastoma	Crizotinib		B	Predictive				5★
EID2345	ALK	F1174L	Neuroblastoma	Crizotinib		B	Predictive				5★
EID4858	ALK	Fusion	Lung Non-small Cell...	Crizotinib		B	Predictive				5★
EID7571	ALK	R1275Q	Neuroblastoma	N/A		B	Predictive				5★
EID1175	ALK	Fusion	Lung Non-small Cell...	Ceritinib		B	Predictive				4★
EID1200	ALK	Fusion	Lung Non-small Cell...	Crizotinib		B	Predictive				4★

<https://civicdb.org/welcome>

# データベース活用実践例：CIViC

例として、甲状腺がんでのALK融合遺伝子に関する臨床レベルでの治療効果予測エビデンスを検索する  
CIViCはエビデンス、遺伝子、バリエントに関する情報が端的にまとめられており有用である

Quicksearch Home About CIViC Help Sign In / Sign Up

Evidence / Home

Evidence Items

The clinical evidence statement is a piece of information that has been manually curated from trustworthy medical literature about a variant or genomic 'event' that has implications in cancer predispositi... Expand

Browse All Evidence (8,960 total, showing 30 of 205 filtered)

EID	Gene	Variant	Disease	Drugs	DESC	EL	ET	ED	CS	VO	R
EID1279	ALK	Fusion	Lung Non-small Cell...	Alectinib		B	Predictiv	All	All	All	4★
EID4835	ALK	Fusion	Lung Non-small Cell...	Brigatinib		B					4★
EID7286	ALK	Fusion	Lung Non-small Cell...	Lorlatinib		B					4★
EID7533	ALK	Fusion	Lung Non-small Cell...	Alectinib		B					4★
EID7535	ALK	Fusion	Lung Non-small Cell...	Ceritinib		B					4★
EID7590	ALK	Fusion	Lung Non-small Cell...	Lorlatinib		B					4★

⑤ "CS" (Clinical Significance) を選択  
⑥ "Sensitivity/Response" が感受性データ

<https://civicdb.org/welcome>



# データベース活用実践例：CIViC

例として、甲状腺がんでのALK融合遺伝子に関する臨床レベルでの治療効果予測エビデンスを検索する  
CIViCはエビデンス、遺伝子、バリエントに関する情報が端的にまとめられており有用である

EID	Gene	Variant	Disease	Drugs	DIT	DESC	EL	ET	ED	CS	VO	R
EID1187	ALK	Fusion	Lung Non-small Cell...	Crizotinib	N/A	A	A	o	o	o	o	5★
EID1198	ALK	Fusion	Lung Non-small Cell...	Crizotinib	N/A	A	A	o	o	o	o	5★
EID1199	ALK	Fusion	Lung Non-small Cell...	Crizotinib	N/A	A	A	o	o	o	o	5★
EID1282	ALK	Fusion	Lung Non-small Cell...	Alectinib	N/A	A	A	o	o	o	o	5★
EID7284	ALK	Fusion	Lung Non-small Cell...	Alectinib	N/A	A	A	o	o	o	o	5★
EID7285	ALK	Fusion	Lung Non-small Cell...	Ceritinib	N/A	A	A	o	o	o	o	5★
EID7872	ALK	Fusion	Lung Non-small Cell...	Alectinib	N/A	A	A	o	o	o	o	5★
EID8657	ALK	Fusion	Lung Non-small Cell...	Alectinib	N/A	A	A	o	o	o	o	5★
EID2341	ALK	R1279G	Neuroblastoma	Crizotinib	N/A	B	B	o	o	o	o	5★
EID2342	ALK	R1279G	Neuroblastoma	Crizotinib	N/A	B	B	o	o	o	o	5★
EID4858	ALK	Fusion	Lung Non-small Cell...	Crizotinib	N/A	B	B	o	o	o	o	5★
EID1175	ALK	Fusion	Lung Non-small Cell...	Ceritinib	N/A	B	B	o	o	o	o	4★
EID1200	ALK	Fusion	Lung Non-small Cell...	Crizotinib	N/A	B	B	o	o	o	o	4★
EID1201	ALK	Fusion	Lung Non-small Cell...	Crizotinib	N/A	B	B	o	o	o	o	4★
EID1237	ALK	Mutation	Lung Non-small Cell...	Ceritinib	N/A	B	B	o	o	o	o	4★
EID1241	ALK	NPM-ALK	Anaplastic Large Cell...	Crizotinib	N/A	B	B	o	o	o	o	4★
EID1279	ALK	Fusion	Lung Non-small Cell...	Alectinib	N/A	B	B	o	o	o	o	4★
EID4835	ALK	Fusion	Lung Non-small Cell...	Brigatinib	N/A	B	B	o	o	o	o	4★
EID7286	ALK	Fusion	Lung Non-small Cell...	Lorlatinib	N/A	B	B	o	o	o	o	4★
EID7533	ALK	Fusion	Lung Non-small Cell...	Alectinib	N/A	B	B	o	o	o	o	4★
EID7535	ALK	Fusion	Lung Non-small Cell...	Ceritinib	N/A	B	B	o	o	o	o	4★
EID7990	ALK	Fusion	Lung Non-small Cell...	Lorlatinib	N/A	B	B	o	o	o	o	4★
EID7865	ALK	EML4-ALK L1196M	Lung Non-small Cell...	Ensartinib	N/A	B	B	o	o	o	o	4★
EID945	ALK	Fusion	Lung Non-small Cell...	Pemestresad	N/A	B	B	o	o	o	o	3★
EID1343	ALK	Fusion	Inflammatory...	Crizotinib	N/A	B	B	o	o	o	o	3★
EID1272	ALK	Fusion	Lung Non-small Cell...	Alectinib	N/A	B	B	o	o	o	o	3★

AlterationがFusionsでのエビデンス：  
Crizotinib, Alectinib, Brigatinib, Ceritinib, Lorlatinibについて  
他がん種でLevel A/B/Cのエビデンス  
→EPWG分類では他がん種Level A-Dに相当  
Crizotinibについて当該がん種でLevel Cのエビデンス  
→EPWG分類では当該がん種Level Cに相当  
→総合し、EPWG分類でLevel Cに相当

Evidence Items  
The clinical evidence statement is a piece of information that has been manually curated from trustworthy medical literature about a variant or genomic 'event' that has implications in cancer predispositi... Expand

View Evidence Model Docs  
View Evidence Curation Docs

Browse All Evidence (8,960 total, showing 1 of 1 filtered)

EID	Gene	Variant	Disease	Drugs	DIT	DESC	EL	ET	ED	CS	VO	R
EID5954	ALK	Fusion	Thyroid Gland Anaplastic...	Crizotinib	N/A	C	A	o	o	o	o	3★

https://civicdb.org/welcome

# データベース活用実践例：CIViC

例として、甲状腺がんでのALK融合遺伝子に関する臨床レベルでの治療効果予測エビデンスを検索する  
CIViCはエビデンス、遺伝子、バリエーションに関する情報が端的にまとめられており有用である

Level	Name	Definition	Example and further comments
A	Validated association	Proven/consensus association in human medicine.	"AML with mutated NPM1" is a provisional entity in WHO classification of acute myeloid leukemia (AML). This mutation should be tested for in clinical trials and is recommended for testing in patients with cytogenetically normal AML. Validated associations are often in routine clinical practice already or are the subject of major clinical trial efforts.
B	Clinical evidence	Clinical trial or other primary patient data supports association.	BRAF V600E is correlated with poor prognosis in papillary thyroid cancer in a study of 187 patients with PTC and other thyroid diseases. The evidence should be supported by observations in multiple patients. Additional support from functional data is desirable but not required.
C	Case study	Individual case reports from clinical journals.	A single patient with FLT3 over-expression responded to the FLT3 inhibitor sunitinib. The study may have involved a large number of patients, but the statement was supported by only a single patient. In some cases, observations from just a handful of patients (e.g. 2-3) or a single family may also be considered a case study/report.
D	Preclinical evidence	In vivo or in vitro models support association.	Experiments showed that AG1296 is effective in triggering apoptosis in cells with the FLT3 internal tandem duplication. The study may have involved some patient data, but support for this statement was limited to in vivo or in vitro models (e.g. mouse studies, cell lines, molecular assays, etc.).
E	Inferential association	Indirect evidence.	CD33 and CD123 expression were significantly increased in patients with NPM1 mutation with FLT3-ITD, indicating these patients may respond to combined anti-CD33 and anti-CD123 therapy. The assertion is at least one step removed from a direct association between a variant and clinical relevance.

A：証明されている/コンセンサスあり→EPWG A/B相当  
B：臨床試験または他がん種で支持するデータあり→EPWG C相当  
C：症例報告レベル→EPWG D相当  
D：前臨床レベル→EPWG E相当

<https://civic.readthedocs.io/en/latest/model/evidence/level.html>

# 第Ⅵ章

# がんゲノム医療データベースの 活用とデータ共有

京都大学大学院 医学研究科附属 医療DX教育研究センター  
医療データ活用部門 准教授

**鎌田真由美**

## 情報共有のためのバリエーションデータベースと臨床的判断のための知識データベース

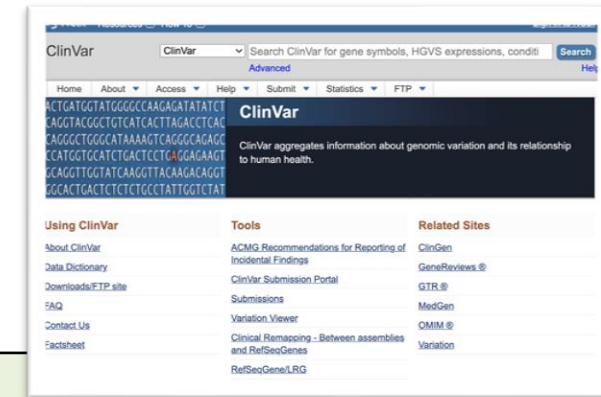
- がんゲノム医療では、パネル検査などで検出された遺伝子バリエーションに基づき、治療を提供する。その際、バリエーションの解釈とエビデンスレベルを与える「臨床的意義付け」を行う。臨床的意義づけを効率的かつ正確に行うために、これまでに共有されている根拠となる情報および専門家による議論を共有（データシェアリング）することが重要
- これまでに文献や研究で明らかになっている遺伝子バリエーションの病原性を集約したデータベースと、治療効果や予後・診断のような臨床的な意義を、複数の情報源から取得し整理した知識データベースがある
- この章では、がんゲノム医療でよく使われるデータベースと知識ベースについて紹介するとともに、利用の際に注意すべき・留意すべき点についてみていく
- 本章で紹介する内容：
  - 代表的なデータベース：ClinVar, COSMIC
  - がんゲノムでよく用いられる便利な知識ベース：VarSome, CIViC, OncoKB, JAX CKB, CGI,
  - ゲノムデータの多様性と日本での取り組み
    - データベース：MGeND, jMorp, TogoVar
  - データシェアリングについて

ClinVarは、ゲノム医療で世界的に広く利用されている疾患バリエーションデータベースである

- **ClinVar** は、臨床的に重要なゲノムバリエーションデータベースとして、世界的に広く利用されているデータベース。米国国立衛生研究所 (NIH) の元、国立生物工学情報センター (NCBI) と国立医学図書館 (NLM) により運営されている。
- 運用開始当初はメンデル遺伝疾患が主体だったが、現在は多因子疾患や体細胞が関わる非メンデル遺伝疾患も対象に、研究機関や検査会社等から広くデータを受け付けている
- がんゲノム医療では、家族性腫瘍における生殖細胞系列変異の解釈に活用されることが多い
- 収載されているバリエーションは、情報提供元において、Pathogenic (病原性あり) や Benign (良性) などの「臨床的意義」が付与されている。メンデル遺伝病においては、American College of Medical Genetics and Genomics (ACMG) / Association of Molecular Pathology (AMP) によるガイドラインでの用語を標準としているが、非メンデル遺伝疾患に関しても、ユーザーニーズに合わせた表現の導入を進めている (表)

### ClinVar

(<https://www.ncbi.nlm.nih.gov/clinvar/>)



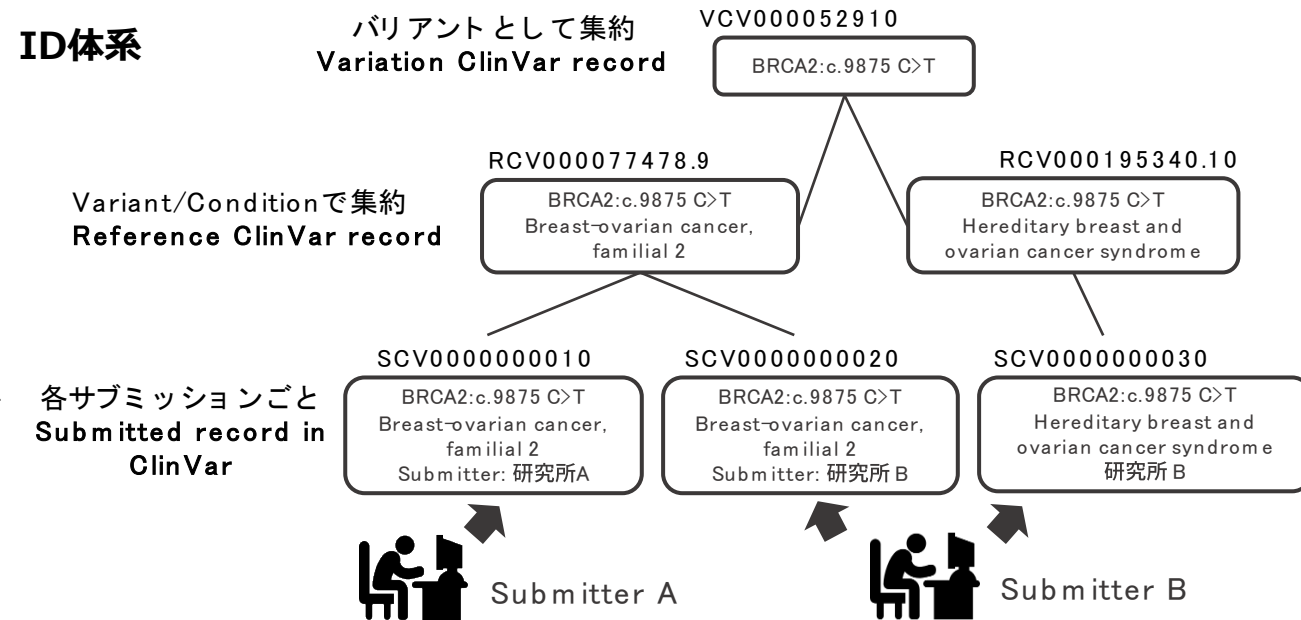
Clinical Significance	用語で表される臨床的解釈
Benign	メンデル遺伝病に関わるバリエーションに対してACMG/AMPで推奨される用語
Likely benign	Benign … 病原性なし
Likely pathogenic	Likely benign … 病原性なしの可能性が高い
Pathogenic	Likely pathogenic … 病原性ありの可能性が高い
Uncertain significance	Pathogenic … 病原性あり Uncertain significance … 解釈不明 (上記のいずれにも分類できなかったもの)
drug response	疾患には影響を及ぼさないが、薬剤応答に影響するバリエーション
association	GWAS解析などで同定され、臨床的な解釈がなされたバリエーション
risk factor	疾患の要因ではないが、その発症リスクを高めるバリエーション
protective	疾患 (感染症を含む) のリスクを軽減するようなバリエーション
Affects	乳糖不耐症のような疾患以外の表現型にかかわるバリエーション
other	上記タームに適するものがないバリエーション

(<https://www.ncbi.nlm.nih.gov/clinvar/docs/clinsig/> より抜粋)

## ClinVarの登録データIDとその構造についての説明は下記の通り

- 提供者から登録されたバリエーションデータには、「Submitted record in ClinVar」という SCV accession number が割り当てられている。同じバリエーションと病状/表現型の組み合わせは、ひとつの「Reference ClinVar record」として集約され、RCV accession number が割り当てられる。また、RCVはバージョン概念を持っている。集約されるバリエーションの情報が更新されると、サブバージョン（"."ドット以降の数字）が変更される
- Webページの検索結果として表示されるバリエーション情報は、同一のバリエーションで集約されたもの。これは「Variation ClinVar record」として表現され、VCV accession number が割り当てられる。ここで、ユニークな整数識別子として各バリエーションのセットに対して割り当てられるVariation IDがある。VCV accession numberは、このVariation IDと全部で9桁になるよう挿入される0によって構成されている
- Variation IDは、接合ヘテロ接合体やハプロタイプなど、複数のバリエーションを包括した形で解釈付がされている場合に、それらのバリエーションセットに対して付与される。一方、個々のバリエーションに対しては、Allele IDというものが付与される
- ClinVarでは各データ提供者独自の臨床的解釈付けを許容しており、キュレーション方法や用いるエビデンスの違い等によって、同一のバリエーションに対して複数の異なる解釈が登録されることがある。もし相反する解釈が登録されたとしても、ClinVarで一方のみを採用することはしていない。同一バリエーションに対する解釈は、次で述べるReview status や登録された解釈の組み合わせに応じて集約してされている

### ID体系



### Variation IDとAllele ID

バリエーションセットとしてのエントリ

NM\_000140.3(FECH):c.[1224T>A;1225C>T;1231T>G]  
Variation ID: 561 (=VCV000000561)  
Description: **Haplotype**

NM\_000140.3(FECH):c.1224T>A  
Variation ID: 242755  
Description: SNV  
Allele ID: 38381

NM\_000140.3(FECH):c.1225C>T  
Variation ID: 242756  
Description: SNV  
Allele ID: 38382

NM\_000140.3(FECH):c.1231T>G  
Variation ID: 242821  
Description: SNV  
Allele ID: 15600

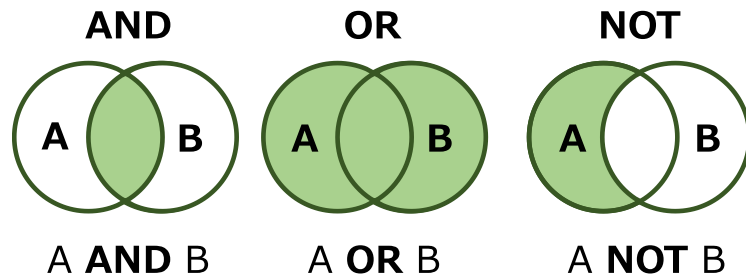
## ClinVar Review Status : 情報の信頼性

- ClinVarでは各バリエントに対し、星★の数による5段階レビューを付与している。ユーザーはこのレビューによって、情報の信頼性を評価することができる
- バリエントの登録において、臨床的な重要性を判定するための基準及びそれを示す文書（Assertion criteria）が付与されている場合には星が1つ付くが、そうでない場合、星は付かない。複数の提供者から同じ基準（解釈付け）でバリエントが登録された場合には、星が2つ付く。一方、複数の登録があったとしても解釈が異なっている場合は、星は1つになる
- NIHには、複数の疾患領域の専門家によるキュレーションの方法の議論や方針決定等を行う **ClinGen** (<https://clinicalgenome.org/>) という研究プロジェクト組織がある。ClinVarとClinGenとの間には、協力体制が構築されており、ClinVarエントリに対して、ClinGenに認められた各疾患領域の専門家によりキュレーションが実施されると3つ星が付与される。また、ACMGのようなガイドラインに合致し、ClinGenのSteering Committeeによりレビュー・承認されたバリエントには、星4つが付与される

星の数	意味	
★★★★	Practice guideline	ガイドラインで決定された場合
★★★★☆	Reviewed by expert panel	各疾患領域の専門家グループによって解釈が決定された場合
★★☆☆	Criteria provided, multiple submitters, no conflicts	複数の提供者から同一の解釈がエビデンスと合わせて登録されている場合
★☆☆☆	Criteria provided, conflicting interpretations OR single submitter	エビデンスが提供されているが単一の提供者のみによる解釈もしくは解釈が提供者間で相反している場合
☆☆☆☆	No assertion (criteria) provided	解釈が提供されていない、もしくは、基準やエビデンスが提供されていない場合

## ClinVarでの検索

- ClinVarでは、トップページから、遺伝子名、HGVSによるバリエーション表現、アミノ酸と残基番号で表記されるタンパク質の変化、染色体位置、疾患名/表現型、登録者名、SVC/RCV/VCVアクセッション番号 をキーワードとした検索が可能
- 特に指定がない場合、ClinVarに収録されているすべてのデータ項目を対象にキーワードの検索を行う。また、検索結果（エントリ）は、遺伝子位置でソートされて表示されるため、場合によっては目的と異なるエントリが検索結果上位に並ぶことがある
- 検索キーワードはスペースで自動的に区切られてしまうため、文字列を一つの検索フレーズとして指定したい場合は、ダブルクォーテーション (") で囲んで検索を行う
- 疾患名/表現型を指定した検索を行うには、"breast cancer"[dis] のように、[dis] を検索キーワードに併記をすることで指定ができる（表）
- 複数のキーワードを組み合わせて検索を行う場合、論理演算子「AND」「OR」「NOT」を用いることができる。



キーワード	略号	検索例
疾患名/表現型	[dis]	"breast cancer" [dis]
遺伝子名	[gene]	BRAF[gene]
染色体位置	[chr] AND [chrpos][chrpos37] ※ [chrpos37]: GRCh37 ※ [chrpos]: GRCh38	7[chr] AND 140753336 [chrpos] 17[chr] AND 43000000:44000000[chrpos37] ※ 範囲指定は「:」で行う

(<https://www.ncbi.nlm.nih.gov/clinvar/docs/help/> より一部抜粋)



## ダウンロードデータの活用についての注意点を挙げる

- ClinVarのデータは、TSV・XML・VCF形式でダウンロードサイトから取得可能。XMLおよびタブ区切りデータでは、すべてのバリエーションセットを対象としている。一方、現在VCF形式には、バリエーションの開始・終了のゲノム位置が明確なエントリのみを対象としている
  - ✓ **XMLデータ**：VCFエントリごとにまとめられているVariationセット (ClinVarVariationRelease\_\*.xml.gz) とRCVごとに詳細がまとめられているフルセット (ClinVarFullRelease\_\*.xml.gz) を利用することができる
  - ✓ **タブ区切りデータ**：様々な観点でまとめられたデータを利用することが可能。variant\_summary.txt はバリエーションを軸に集約されたデータ。gene\_specific\_summaryでは、遺伝子ごとのデータ登録数と Pathogenic もしくは Likely pathogenic と報告された Alleleの数などがまとめられている
  - ✓ **VCFデータ**：後続の解析にも使いやすい形式である。特に、よく用いられる SnpEff や VEP などのアノテーションツールでも利用することができる
- ClinVarのダウンロードサイトには、上述のバリエーションデータ以外にも、疾患名に関する定義ファイルやデータ登録テンプレートなど、様々なファイルが共有されている。ClinGenのディレクトリには、3つ星レビューを得る際のキュレーションに関するドキュメント類が置かれている
- バリエーションデータは毎週・毎月更新される。XMLの毎週のリリースはアーカイブされない。ここで注意が必要なのは、臨床的意義が更新される場合もあること。どのデータベースを使用する場合も同様だが、使用するデータのリリース日を明記することは重要

## ダウンロードサイト

(<https://ftp.ncbi.nlm.nih.gov/pub/clinvar/>)

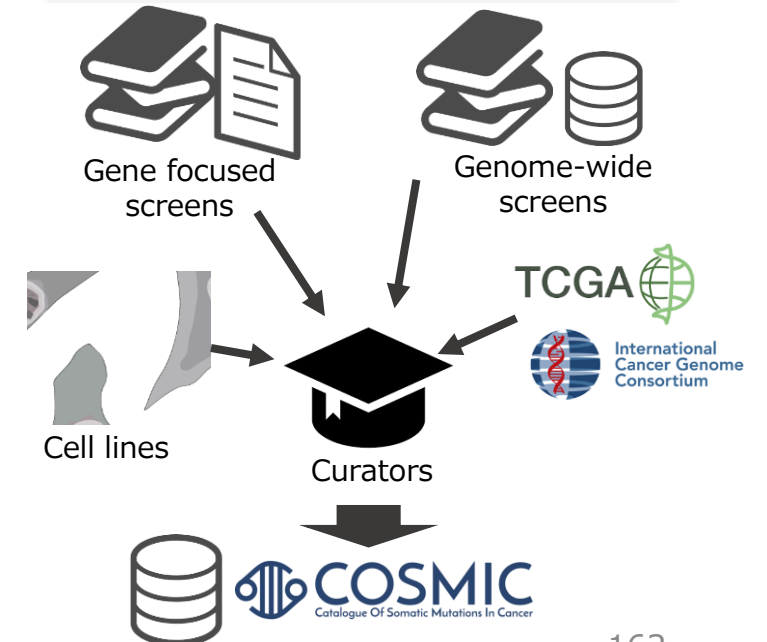
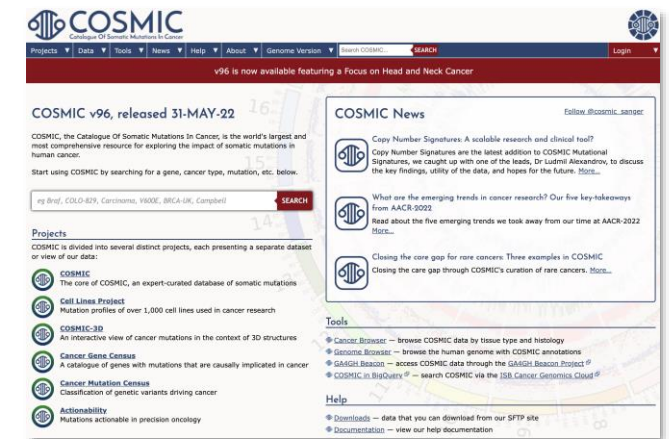
## Index of /pub/clinvar

Name	Last modified	Size
<a href="#">Parent Directory</a>		-
<a href="#">ClinGen/</a>	2018-12-14 09:17	-
<a href="#">document_archives/</a>	2014-04-24 08:19	-
<a href="#">presentations/</a>	2021-06-23 17:39	-
<a href="#">release_notes/</a>	2022-06-20 08:44	-
<a href="#">submission_examples/</a>	2020-08-03 13:46	-
<a href="#">submission_templates/</a>	2022-03-15 14:32	-
<a href="#">tab_delimited/</a>	2022-06-26 09:05	-
<a href="#">temp/</a>	2021-11-03 15:14	-
<a href="#">vcf_GRCh37/</a>	2022-06-26 01:00	-
<a href="#">vcf_GRCh38/</a>	2022-06-26 01:00	-
<a href="#">xml/</a>	2022-06-26 01:06	-
<a href="#">xsd_public/</a>	2022-06-20 08:44	-
<a href="#">xsd_submission/</a>	2020-05-22 13:21	-
<a href="#">ConceptID_history.txt</a>	2022-06-28 10:14	1.1M
<a href="#">README.txt</a>	2022-03-15 14:28	43K
<a href="#">README_VCF.txt</a>	2017-11-14 07:47	7.0K
<a href="#">clinvar_public.xsd</a>	2022-04-25 15:39	77K
<a href="#">clinvar_submission.xsd</a>	2020-05-22 13:21	124K
<a href="#">disease_names</a>	2022-06-28 10:15	4.2M
<a href="#">gene_condition_source_id</a>	2022-06-28 10:25	1.1M

COSMICは、大規模ながん体細胞変異データベースである

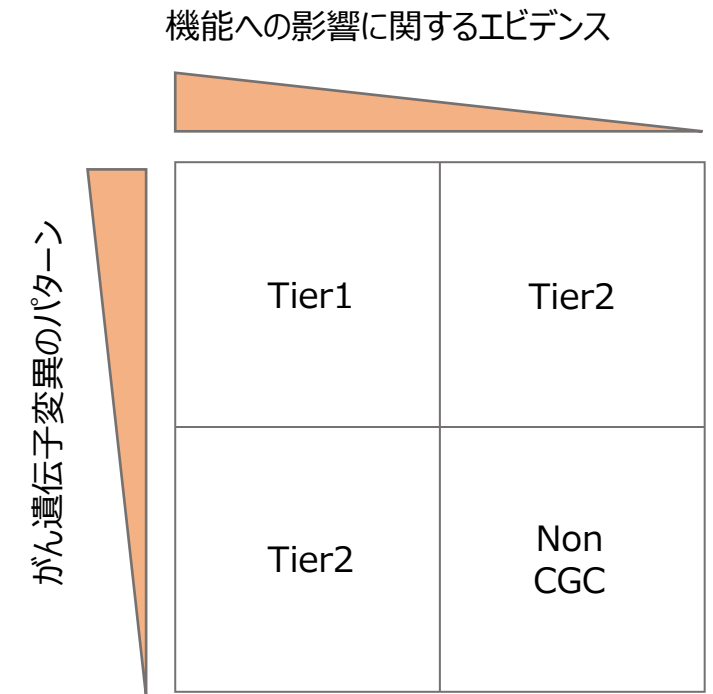
- **COSMIC** (Catalogue Of Somatic Mutations In Cancer) は、英国 Wellcome Trust Sanger研究所が運営する、がんに関する体細胞変異情報を集約した、大規模データベースである。 ClinVarと同様に、世界的に広く参照されている。今回紹介するCOSMICデータベースと、細胞株と薬剤感受性の網羅的データを集約したCell Lines project が、COSMICの情報リソース源となっている
- がんに関連する遺伝子および体細胞変異について、各腫瘍タイプ・臓器でどのくらい報告があるのか（稀な変異かどうか）を確認できるとともに、がんの特徴的なSignatureもカタログ化されている
- COSMICデータベースは、主に以下の2つのデータから構成されている。
  - ✓ **High Precision Data** : 専門家が査読つき論文から手動でキュレーション&抽出したバリエーションセット。 論文は、後述する、がん関連遺伝子カタログCancer Gene Census を対象に、PubMedから検索し、本文もしくは Supplementary material に記載されるサンプルとバリエーションのデータを抽出している。Negative サンプル（変異がないサンプル）情報も収載しており、正確な頻度情報を確認することができる
  - ✓ **Genome-wide Screen Data** : 大規模なゲノムスクリーニングに関する報告論文からアップロードされたものや、TCGAやICGCなどの他のデータベースから取得したものも含むデータセット。 Cancer Gene Censusに限らず、ゲノム全体を網羅していることから、疾患の偏りのない分子プロファイルが可能。また、バリエーションを持たない遺伝子を検出・検証することが可能であり、更に、新規なドライバー遺伝子の発見も期待ができる

COSMIC (<https://cancer.sanger.ac.uk/cosmic>)



## COSMIC Cancer Gene Census

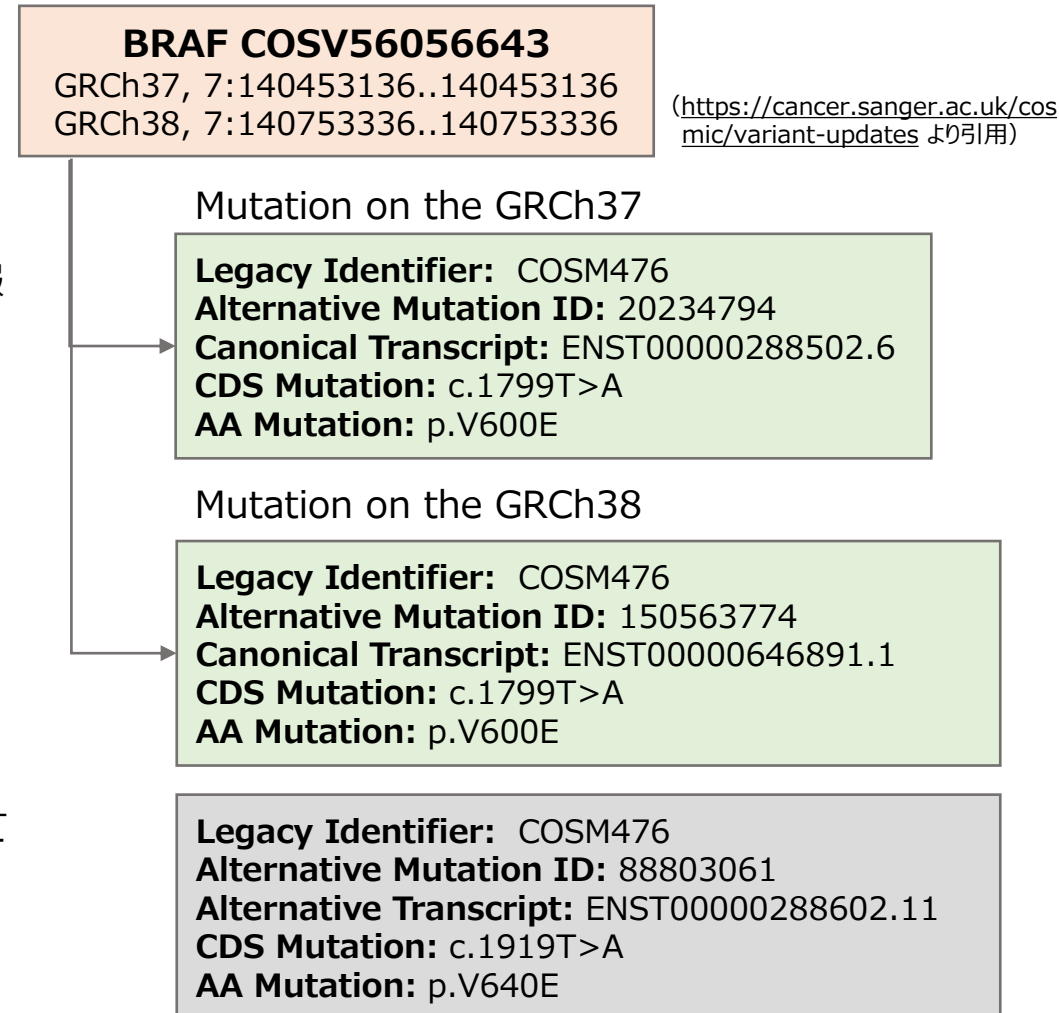
- **Cancer Gene Census (CGC)** は、COSMICで定義される、がんとの因果関係が指摘されている変異を持つ遺伝子のカタログのこと。ここでの定義は、新しいエビデンスが出てきたときに更新されるものであり、固定ではないことに注意が必要。リストアップされる遺伝子は、2つのグループ (Tier) に分けられる。
- 「発がんに機能的に関与していることを示すエビデンスがあるか」と「文献キュレーションで明らかになっている体細胞変異のパターンが存在するか」の2軸を基準に、グループは決定される。機能解析と変異解析どちらからも強力なエビデンスが得られている場合のみ「Tier1」として認定される。一方、がんのドライバーとして典型的な変異パターンを持つが、機能的に特徴づけられていない遺伝子や、がんへの関与について機能的な報告がされているが、体細胞における変異が証明されていない遺伝子は、「Tier2」に分類される (図)
- CGCでは、変異が認められている細胞種 (somatic/germline) などの情報に加え、がんにおける役割 (Oncogene/TSG (Tumor Suppressor Gene) /Fusion) の情報も付与されている



(Sondka, Z. et al., *Nat Rev Cancer* **18**, 696–705 (2018) Figure 2より引用)

## COSMICで使われるIDについて説明する

- **Mutation somatic status** : 変異が確認された細胞に関する情報。「Confirmed Somatic」は、同一個人に対し、生殖細胞アレルとは異なる腫瘍サンプルアレルにあることが確認できたバリエーションのこと。「Previously Reported」は、生殖細胞系列アレルの情報は提供されていないが、同じバリエーションが別患者の正常-腫瘍サンプルペアで「Confirmed Somatic」になっている場合に付与される。「Variant of unknown origin」はデータソースに生殖細胞アレルに関する記載がなく、判別することができない場合に付与される情報
- **Mutation ID** : 2019年のv.90リリース以降、COSVから始まるIDが Genomic Mutation ID として使用されている。COSVは、ゲノム上の変異体の確定的な位置を示すゲノム変異識別子で、リリースの異なるバージョン間でも追跡可能なID。また、このIDは、異なるアセンブリ (GRCh37とGRCh38) 間でも同じ。ダウンロードデータには、これまで使用されていたCOSM/COSNから始まるID (Legacy Mutation Identifier) も記載されている。COSMは Coding regionの変異に対するID、COSNは Non-coding 変異に対するID (図)
- **Sample ID** : 変異が見つかったサンプルに対するユニークなIDのこと。論文をもとに、変異が発見されたサンプルには Sample name が付与される。この名前は、論文などの様々な情報源から取得されるが、多くの場合は、細胞株名。異なる論文で共通するサンプル名がつく場合、本当に同じかどうか不明な場合は、同じ Sample name に対して複数のSample IDが付与される
- **Study ID** : データ元となる研究・論文に対し、Study IDが付与されている。サンガー研究所により実施された研究やICGC/TCGAからインポートされた場合にはCOSUから始まるIDが、論文 (Publication) に対してはCOSPから始まるIDが付与されている



## COSMIC Webブラウザでは各遺伝子に対する様々なアノテーション情報が確認できる

- COSMICでは、トップページから、遺伝子名、組織、がんタイプ、HGVSによる変異記載、Sample IDなどでの検索が可能

### 遺伝子ページ (e.g. BRAF)

- ✓ BRAF遺伝子全長に対し、どの領域にどのくらい変異の報告があるか確認できる
- ✓ インタラクティブに領域を絞るなどの操作が可能

- ✓ CGCが、キュレーションされているか等、遺伝子に関するアノテーション
- ✓ 融合遺伝子での報告があれば相手の遺伝子についての記載
- ✓ 耐性変異に関連する薬剤および同薬剤に紐づく他の遺伝子情報等

✓ 各変異がどのくらい報告されているか確認できる (画像は "Count" でソート)

### バリエーション詳細ページ (e.g. BRAF p.V600E)

- ✓ 組織別の変異の分布や、変異を有するサンプルの情報が確認できる

### COSMICデータのダウンロードに際しての注意点を記す

- COSMICデータは、アカデミック研究者はユーザー登録後、無料で利用することができる。その他に対しては、QIAGEN商用ライセンス（有償）で利用可能
- ダウンロードデータは主に、タブ区切りのデータ（TSV）として提供されている
  - ✓ **High Precision** : CosmicCompleteTargetedScreensMutantExport.tsv.gz
  - ✓ **Genome-wide screen** : CosmicGenomeScreensMutantExport.tsv.gz
  - ✓ **全ての変異** : CosmicMutantExport.tsv.gz
- 各ファイルの全ての情報をダウンロードすることも可能だが、COSMICに集約されているデータは膨大（v.96 データで、ユニークなバリエーション（COSV ID）数は、23,399,170）。その為、特定の遺伝子・サンプル・がん種に絞ってダウンロードすることも可能。また、コマンドラインを用いて取得することも可能
- COSMICでは、変異データの他、SV、CNVのデータ、TCGA由来のGene expression とメチル化データも提供されている
- VCF形式でも coding、non-coding 別にデータが提供されている。ここで、VCF形式では、タブ区切りデータと同様の記載（3'shifted）されたVCFと、VCF標準（5'shifted）に変更した正規化バージョンのVCFの、2つの種類がある。他のデータベースから取得したデータと統合して解析する場合には、バージョンの確認が必要
- 更新頻度は年に4回

### COSMIC ダウンロードページ（一部）

(<https://cancer.sanger.ac.uk/cosmic/download>)

The screenshot displays three download options for COSMIC data:

- COSMIC Complete Mutation Data (Targeted Screens)**: A tab separated table of the complete curated COSMIC dataset (targeted screens) from the current release. It includes all coding point mutations, and the negative data set. Buttons: Download Whole File (Login Required), Download Filtered File (Login Required), Scripted download (Login Required). File: CosmicCompleteTargetedScreensMutantExport.tsv.gz
- COSMIC Mutation Data (Genome Screens)**: A tab separated table of coding point mutations from genome wide screens (including whole exome sequencing). Buttons: Download Whole File (Login Required), Download Filtered File (Login Required), Scripted download (Login Required). File: CosmicGenomeScreensMutantExport.tsv.gz
- COSMIC Mutation Data**: A tab separated table of all COSMIC coding point mutations from targeted and genome wide screens from the current release. Buttons: Download Whole File (Login Required), Download Filtered File (Login Required), Scripted download (Login Required). File: CosmicMutantExport.tsv.gz

VarSomeは、キュレーション過程が可視化された、使いやすい知識データベースである

- ここからは、代表的な知識データベースを見ていく
- まずは、がんに特化したデータベースではないが、2022年6月時点で139のデータソースから情報を集約している、**VarSome** を紹介する
- VarSome.comは、バリエーションに関する専門知識をグローバルに共有することを目的とした、**コミュニティ主導型のプロジェクト**。現在、30万人以上のヘルスケアおよびライフサイエンスのユーザーが参加しており、バリエーションの分類、研究成果や関連する文献へのリンク、エビデンスを共有している
- このプロジェクトにより構築されているデータベースが、**VarSome**。後述するGA4GHのメンバーでもあるスイスのSAPHETOR社により運用されている。掲載されているバリエーションの数がとても多く、解釈しやすいインターフェースが特徴。VarSomeは、ライセンスが必要な一部データソースを除き、無料で参照可能だが、フルバージョンは有償。データへのアクセスはAPIを介して可能であるが、フリープランの場合は、17番染色体のみアクセス可能
- VarSomeでは、各バリエーション詳細ページで、自動化されたACMG・AMPガイドラインに基づくバリエーション分類を提供している。各エビデンスカテゴリに該当する情報を、自動的にデータベースなどから抽出しているが、追加のエビデンスが得られた場合には、マニュアルでアノテーションすることも可能。臨床的なエビデンスは、ClinVarやUniprot、MitoMap、そしてVarSomeユーザーにより登録された文献などから取得されてる

**VarSome**  
(<https://varsome.com/>)

chr7:140453136-A-C (BRAF:p.V600G)

ACMG Classification - Educational use only Version: 10.1.5

Verdict: **Pathogenic**  
Gene BRAF is associated with cancer

NM\_004333.6, canonical, protein length 767, gene BRAF, missense variant

Automated criteria  Show summary view

Rule	Explanation	Show failed criteria
PP5 Very Strong	ClinVar classifies this variant as Pathogenic, rated 3 stars, reviewed by expert panel. VarSome users have linked 1 article stating the variant is Pathogenic (2310010). Using strength Very Strong because of the evidence presented by ClinVar and VarSome users.	<input type="checkbox"/>
PM1 Strong	Using strength Strong because Hot-spot of length 17 amino-acids has 48 non-VUS missense/in-frame variants (48 pathogenic and 0 benign), pathogenicity = 100.0%, qualifies as a dense hot-spot.	<input type="checkbox"/>
PM2 Moderate	Variant not found in gnomAD exomes (good gnomAD exomes coverage = 82.5). Variant not found in gnomAD genomes (good gnomAD genomes coverage = 30.6).	<input type="checkbox"/>

Alternative variant chr7:140453135 CA=AT (Val600Asp) is classified Pathogenic, 0 stars, by ClinVar (and confirmed using ACMG).  
Alternative variant chr7:140453136 AC=CG (Val600Arg) is classified Pathogenic by a VarSome user in article 25763513 (and confirmed using ACMG).  
Alternative variant chr7:140453136 AC=CT (Val600Arg) is classified Pathogenic by a VarSome user (and confirmed using ACMG).  
Alternative variant chr7:140453136 AC=TT (Val600Lys) is classified Pathogenic by a VarSome user in article 29483135 (and confirmed using ACMG).

VarSomeでは解釈のための情報を網羅的に確認することができる

- バリエント詳細ページでは、AMP/ACMGガイドラインによる分類とその根拠となる情報を確認できるほか、ClinVarやCIViCなど、がんゲノムに関わる重要なデータベースの情報を一覧で確認することができる。

バリエント詳細ページ (e.g. KRAS G12D)

General Information SNV KRAS(NM_004985.5):c.35G>A (p.Gly12Asp)	AMP Classification Tier I	ICGC Samples No data available	ACMGによる分類 Pathogenic	Pathogenicity Scores 17
Genes KRAS	JAX CKB No data available	Cancer Hotspots Samples: 758 Type: Pancreatic Cancer, Color...	ClinVar Pathogenic Submissions: 15	Expression Data Top: Cells - EBV-transformed I... Tissues: 54
Community Contributions	CIVIC Lung Non-small Cell Carcinoma	GDC Samples: 200 Type: Colorectal Cancer, Endo...	Frequencies Gnomad Genomes $f = 0.0000131$ Mean Coverage: 31.1	GWAS No data available
Publications Variant: 1562 Gene: 337	PMKB Tier I Evidence: 13 Adenocarcinoma	cBioPortal Samples: 4482 Type: Pancreatic Cancer, Color...	MitoMap No data available	Structural Browser
Transcripts NM_004985.5 - missense RefSeq: 4 Ensembl: 12	Cancer Samples Summary Samples: 5440 Sources: 3 Tier I	PharmGKB No data available	Conservation Scores phyloP100: 7.892	Beacon Network
Region Browser	Cosmic No data available	DoCM Pathogenic Diseases: 9 Drugs: 0	Uniprot Variants Pathogenic	Protein Viewer

※ スクリーンショットは2022年6月時点であることを注意

e.g. ACMGガイドラインに基づく分類

分類結果

Verdict Pathogenic

Automated criteria Show summary view Enable clinical evidence

Pathogenic  
PS4 Strong PM3 Moderate PP4 Supporting PM1 Strong PM2 Strong PM5 Strong PP5 Strong PS3 Strong  
PP2 Supporting PP3 Supporting PM4 Moderate PM6 Moderate PP1 Supporting PS2 Strong PVS1 Vary Strong

Benign  
BP2 Supporting BP5 Supporting BA1 Stand Alone BP1 Supporting BP3 Supporting BP4 Supporting BP6 Supporting BP7 Supporting  
BS1 Strong BS2 Strong BS3 Strong BS4 Strong

Rule Explanation Show failed criteria

PM1 Strong Using strength Strong because hot-spot of length 17 amino-acids has 39 missense/in-frame variants (29 pathogenic variants and 10 uncertain variants), which qualifies as a dense hot-spot.

PM2 Strong Using strength Strong because the position is strongly conserved (phyloP100way = 7.89 is greater than 7.2). GnomAD exomes allele count = 1 is less than 5 for AD unable to check gnomAD exomes coverage. GnomAD genomes allele count = 2 is less than 5 for AD with good gnomAD genomes coverage = 31.1.

PM5 Strong Using strength Strong because 7 pathogenic alternative variants identified. Alternative variant chr12:25245350 C>A (Gly12Trp) is classified Pathogenic by the VarSome community in article 33075161 (and confirmed using ACMG). Alternative variant chr12:25245350 CC>GA (Gly12Ser) is classified Pathogenic by the VarSome community in article 32308773 (and confirmed using ACMG). Alternative variant chr12:25245350 C>G (Gly12Val) is classified Pathogenic by UniProt Variants (and confirmed using ACMG). Alternative variant chr12:25245350 C>A (Gly12Ala) is classified Pathogenic by the VarSome community (and confirmed using ACMG). Alternative variant chr12:25245351 C>A (Gly12Cys) is classified Pathogenic by the VarSome community (and confirmed using ACMG). Alternative variant chr12:25245351 C>G (Gly12Arg) is classified Pathogenic by the VarSome community (and confirmed using ACMG). Alternative variant chr12:25245351 C>T (Gly12Ser) is classified Pathogenic by UniProt Variants (and confirmed using ACMG).

PP5 Strong Using strength Strong because the VarSome community has classified this variant as Pathogenic, citing 19 articles (35075146, 33917572, 33538228, 32810914, 31409810 and 14 more).

PS3 Strong UniProt Variants classifies this variant as Pathogenic, backed by functional studies (requires user validation) mentioned in 40 articles (29298116, 23607916, 23475782, 23103869, 22662154 and 35 more), and also citing 1071 articles (30891959, 24997135, 24936796, 24913807, 24866168 and 1066 more), associated with Acute myeloid leukemia, Carcinoma of pancreas, Cerebral arteriovenous malformation and 16 more.

PP2 Supporting ClinVar classifies this variant as Pathogenic, 2 stars (multiple consistent, 32 submissions), backed by functional studies (requires user validation) mentioned in 2 articles (29298116 and 21079152), and also citing 45 articles (31891627, 31836588, 30936194, 30443000, 25695684 and 40 more), associated with Acute Myeloid Leukemia, Acute Myeloid Leukemia by Fab Classification, Acute Myeloid Leukemia, Adult and 39 more.

PP3 Supporting 253 out of 258 non-VUS missense variants in gene KRAS are pathogenic = 98.1% which is more than threshold of 40.2%, and 270 out of 479 clinically reported variants in gene KRAS are pathogenic = 56.4% which is more than threshold of 35.2%.

Pathogenic computational verdict based on 11 pathogenic predictions from BayesDel\_addAF, DANN, DEOGEN2, EIGEN, FATHMM-MKL, M-CAP, MVP, MutationAssessor, MutationTaster, PrimateAI and SIFT vs 1 benign prediction from LIST-52.

該当するエビデンスが色付きで示されている



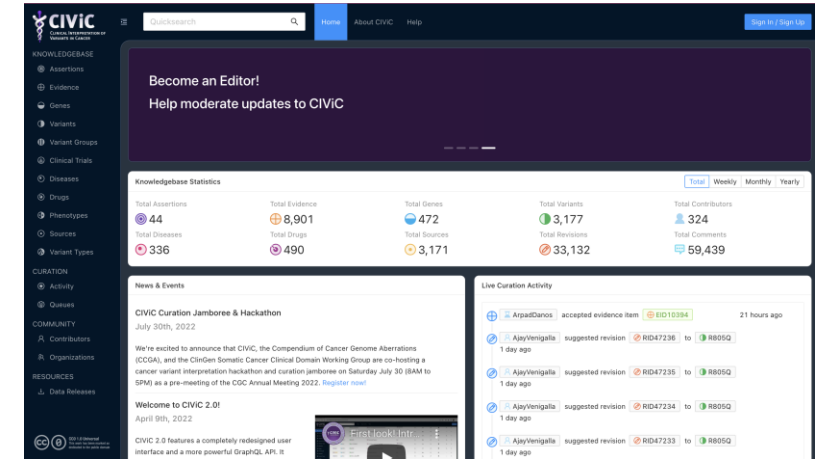
各エビデンスの詳細を確認できる



CIViCは、オープンアクセス・オープンソースな、がん知識データベースである

- **CIViC** (Clinical Interpretation of Variants in Cancer) は、Washington University School of Medicineにより運営される、がん知識データベースのこと。VarSomeと同様、コミュニティ主導型のプロジェクトにより構築されている。プレジジョンメディシンの実現において、必要な臨床的解釈は自由に入手でき、多様なコミュニティでオープンに議論されるべき、との理念に基づき、オープンアクセス・オープンソース・オープンライセンスなウェブデータベース
- 登録対象となるバリエントは、Evidence statement (臨床的な関連性についての記述) があるものに限られる。キュレーションは、当該領域の専門家(研究者・臨床医)によるマニュアルキュレーションで実施がされるが、アカウントを作成し申請することで、誰でも参加が可能。コミュニティにキュレーターとして参加することで、バリエントに対するエビデンスの追加・修正を投稿することが可能。キュレーターにより投稿された内容は、エディターによりレビューされ、承認・棄却がなされる。情報の公開には、最低でも2名のレビューが必要
- 臨床的解釈を支える各エビデンスは、Evidence items (EID) として管理され、それらによりサポートされるバリエントの臨床的根拠をまとめたものはCIViC Assertionとして管理されている。CIViCでは、エビデンスの質に対するキュレーターの信頼度を1~5の星の数で示す「Evidence Rating」という項目がある。信頼度は、研究規模、研究デザイン、検証方法、再現性などの複数要素によって決定される。また、各エビデンスには、A/B/C/D/Eからなる「Evidence Level」が付与されている(表)
- データは誰でも登録なしに使うことができ、APIを用いても自由にアクセス可能。データセットは、全てのレコードを毎晩・毎月取りまとめて更新し、Data Releasesページで公開されている。Gene/Variants/Evidenceなどの情報ごとに取りまとめられたTSV形式と、それら情報を取りまとめたVCF形式でも取得可能。TSVデータは、承認されたエビデンスレコードを持つバリエントのみが含まれている。VCF形式では、完全なゲノム座標を持つバリエントのみが含まれている為、ゲノム座標がキュレーションされていないバリエントも含むTSV形式とは、バリエント数が異なる場合がある

CIViC (https://civicdb.org/welcome)

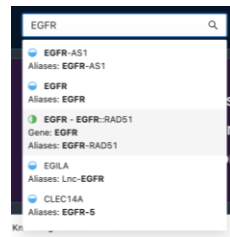


### Evidence Level

Level	Name	定義
A	Validated association	バリエントとの関連性が医学的に証明されている、もしくは、臨床的なコンセンサスが得られている
B	Clinical evidence	臨床的関連性を裏付ける大規模な臨床試験や5人以上の主張を裏付ける患者データに由来
C	Case study	臨床雑誌に掲載された個々の症例報告
D	Preclinical evidence	臨床上の主張を裏付けるin vivoまたはin vitroの実験から得られたもの
E	Inferential association	間接的に関連付けるもの

CIViCでは臨床的解釈のための様々なエビデンスを確認することができる

- トップページ Quick Search から、遺伝子名・変異名をキーワードに検索が可能。



遺伝子詳細ページ  
(e.g. EGFR)

登録されている変異リスト

フィルタで目的の変異を検索可能

EGFR Variants 3 Total Filter: L858Q

Buttons: Ex19 del L858R, L858Q, L858R

変異  
詳細ページ  
(e.g. L858R)



変異 x 疾患  
に対する  
薬剤の情報  
確認できる

※ スクリーンショットは2022年6月時点であることに注意

Evidence Level

Evidence Type

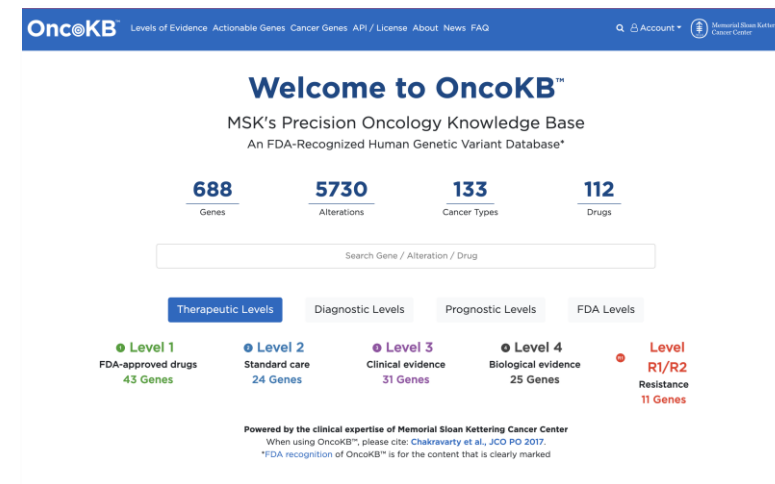
EID	Disease	Drugs	EL	ED	CS	VO	R
EID2994	Lung Non-small Cell...	Erlotinib	A	👁	👁	👁	5★
EID2997	Lung Non-small Cell...	Afatinib	A	👁	👁	👁	5★
EID4860	Lung Non-small Cell...	Dacomitinib	B	👁	👁	👁	5★
EID275	Lung Non-small Cell...	Gefitinib, Erlotinib	B	👁	👁	👁	4★
EID879	Lung Adenocarcinoma	Afatinib	B	👁	👁	👁	4★
EID982	Lung Adenocarcinoma	Afatinib	B	👁	👁	👁	4★
EID2621	Lung Non-small Cell...	Gefitinib	B	👁	👁	👁	4★

Type	Definition
Predictive	治療効果に対する変異の影響に関するエビデンス
Prognostic	重症度や患者生存に対する変異の影響に関するエビデンス
Diagnostic	診断に対する変異の影響に関するエビデンス
Predisposing	生殖細胞変異の病原性に関するエビデンス
Oncogenic	体細胞変異の腫瘍の病因であるかに関するエビデンス
Functional	変異による遺伝子産物の機能変化有無に関するエビデンス

OncoKBは、実用的なラベル情報が利用可能な、体細胞がんの知識データベースである

- **OncoKB** (Precision Oncology Knowledge Base) は、米国Memorial Sloan Kettering Cancer Center (MSK) により運用される知識データベース。体細胞バリエーションによる影響とその臨床的な意義について集約している。初めてFDA承認されたがん臨床シークエンス検査である、MSK-IMPACTにより得られた研究成果も、OncoKBでは公開されている。2022年現在は、体細胞バリエーションを対象としているが、今後、生殖細胞バリエーションにも拡張を計画している
- 収載されているバリエーションのキュレーションは、MSKの各疾患領域を専門とする臨床医・臨床研究者の監督の下、キュレーターによってマニュアルで実施されている。各バリエーションは治療効果や予後情報、生物学的な影響などと合わせて検証され、治療上の意義を4つのレベルで分類している。
  - Level1: FDA承認薬に対するバイオマーカーとして認められているバリエーション
  - Level2: 標準治療においてバイオマーカーとして用いられるバリエーション
  - Level3: 臨床的エビデンスによりバイオマーカーとして見込まれるバリエーション
  - Level4: 生物学的実験により薬剤応答を示すバイオマーカーとして見込まれるバリエーション
- さらに、キュレーションされたアノテーション情報として、生物学的な機能への影響 (Gain-of-function/Loss-of-function etc…) や、発がんへの影響 (Oncogenic/Likely Oncogenic etc…) などのラベル情報も提供している。これらの判断基準は、詳細なSOP (<https://sop.oncokb.org/>) としてまとめて公開されている
- 検索以外のデータアクセスは、APIを用いる。これには事前登録と、それにより取得可能なトークンが必要。また、Pythonで提供されるOncoKB-Annotator (<https://github.com/oncokb/oncokb-annotator>) を使うことで、手元のMAFファイルへのアノテーションも可能

OncoKB (<https://www.oncokb.org/>)



バリエーションに付与されているラベル情報 (一部)

項目	ラベル
治療上の意義	Level1 - 4
生物学的効果	Gain-of-function, Likely Gain-of-function, Loss-of-function, Likely Loss-of-function, Switch-of-function, Likely Switch-of-function, Neutral, Likely Neutral, Inconclusive, Unknown
発がんを誘発するか否か	Oncogenic, Likely Oncogenic, Likely Neutral, Inconclusive, Unknown, Resistance

## 遺伝子ごとの検索、アノテーションレベルでの検索も可能

✓ トップページから「治療」「診断」「予後」「FDA」各々のアノテーションによる遺伝子の検索

Therapeutic Levels: Level 1 (43 Genes), Level 2 (24 Genes), Level 3 (31 Genes), Level 4 (25 Genes), Level R1/R2 (11 Genes)

Diagnostic Levels: Level Dx1 (22 Genes), Level Dx2 (53 Genes), Level Dx3 (57 Genes)

✓ 「Actionable genes」タブから、アノテーションを組み合わせた検索も可能

Showing 285 clinical implications (97 genes, 61 cancer types, 3 levels of evidence)

Level	Gene	Alterations	Cancer Types	Drugs
1	ABL1	BCR-ABL1 Fusion	B-Lymphoblastic Leukemia/Lymphoma	Dasatinib
1	ABL1	BCR-ABL1 Fusion	B-Lymphoblastic Leukemia/Lymphoma	Imatinib
1	ABL1	BCR-ABL1 Fusion	B-Lymphoblastic Leukemia/Lymphoma	Ponatinib
1	ABL1	BCR-ABL1 Fusion	Chronic Myelogenous Leukemia	Asciminib
1	ABL1	BCR-ABL1 Fusion	Chronic Myelogenous Leukemia	Bosutinib

条件を満たす  
臨床的意義

## 遺伝子ページ (e.g. BRAF)

**BRAF**  
Oncogene  
Highest level of evidence: Level 1 • Level Dx2 • FDA Level 2

Also known as BRAF1  
BRAF, an intracellular kinase, is frequently mutated in melanoma.

✓ 遺伝子全長・腫瘍タイプでの変異分布

Annotated Alterations: Therapeutic, Diagnostic, FDA-Recognized Content

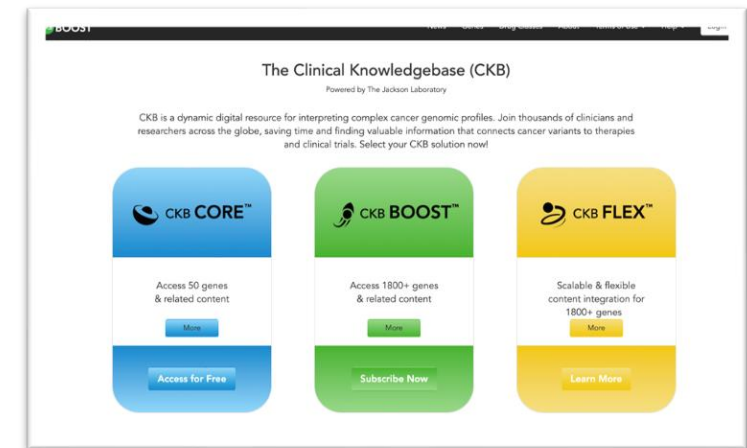
Alteration	Oncogenic	Mutation Effect	Citations
AGAP3-BRAF Fusion	Likely Oncogenic	Likely Gain-of-function	4
AGK-BRAF Fusion	Oncogenic	Gain-of-function	5
AKAP9-BRAF Fusion	Oncogenic	Gain-of-function	1
Amplification	Oncogenic	Gain-of-function	5
ARMCI0-BRAF Fusion	Likely Oncogenic	Likely Gain-of-function	1

✓ 遺伝子が有する変異の各種アノテーション情報  
✓ Oncogenicity, Mutation Effect (GoF/LoF)  
✓ 薬剤とエビデンスレベル etc...

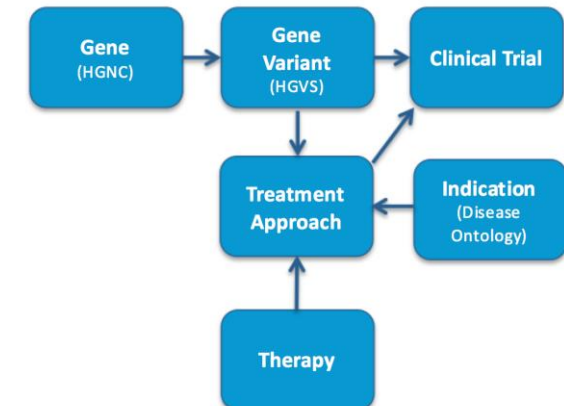
JAX CKBは、文献キュレーションにより治療に関するアノテーションを提供する知識データベースである

- **JAX CKB** (JAX Clinical Knowledgebase) は、様々な実験マウスの供給元として有名な米国 The Jackson Laboratory により開発・運用されている、体細胞がんを対象とした知識データベース。がんゲノムパネル検査JAX-CTPに対する知識ベースとして開発されてきたもので、現在は一部無料公開されている。有料プランでは、全登録情報を閲覧することができる
- キュレーションは、機械学習により抽出された科学論文に対するマニュアルキュレーションによって実施されている。対象とする文献は全て査読つき論文。直接的な情報がない場合も、類似する変異に対する情報がある場合には「predicted」のラベルが付与され、アノテーションされる。このキュレーションにより、各遺伝子・変異に対する、治療知識、診断/予後情報、臨床試験などの情報が付与されている
- 無料で使用できる一般公開版では、よく知られている82のドライバー遺伝子に対し、遺伝子、バリエーション、薬剤、薬剤クラス、適応症、および臨床試験の軸から検索することができる

JAX CKB (<https://ckbhome.jax.org/>)



CKB Data Model Overview:



# Cancer Genome Interpreter

Cancer Genome Interpreterは、in silico 解析を駆使したがんアノテーションシステムである

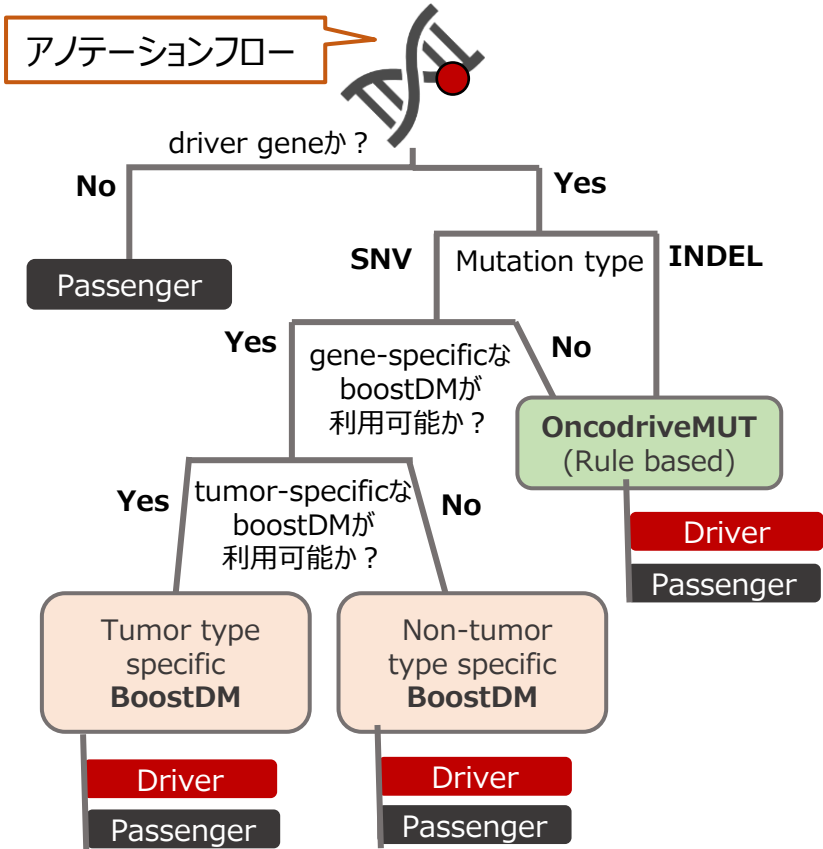
- **Cancer Genome Interpreter (CGI)** は、疾患の原因となるバリエーションや、治療の対象となるバリエーションを特定することを目的に、Barcelona Biomedical Genomics Labにより運営されているアノテーションシステムである。同Labが開発してきたバイオインフォマティクスツールおよび公開データベースに集積された情報を活用し、アノテーションを行っている
- TCGAなどから取得した66のがん種・28,000の腫瘍サンプルを対象に、同ラボが開発した解析パイプライン IntOGen (<https://www.intogen.org>) で得られたバリエーションをデータリソースとしている。CGIでは、これらのバリエーションに対し、DriverかPassengerかのラベルを付与している。このラベル判定は、同ラボが開発したバイオインフォマティクスツールを活用している。(右図)
  - SNV : 各がん遺伝子に対する機械学習によるboostDMもしくはOncodriveMutにより予測された結果
  - Indel : OncodriveMutによって予測した結果
  - CNVや転座などの構造変異 : COSMICのCancer Gene Censusに掲載されている情報などに基づくアノテーション
- CGIのアノテーションでキーとなるboostDMは、既存のがんコホート研究データから各がん遺伝子および各がん種ごとに特徴を抽出して学習モデルを構築し、高精度な予測を実現したもの (Muiños, F. et al., *Nature* 596, 428–432 (2021))。また、OncodriveMutは、既存のがんコホート研究データと健常人データに対する解析によって各遺伝子の特徴量を抽出し、ルールベースな分類でバリエーションをDriver/Passengerのラベルを推定する (Tamborero, D. et al., *Genome Med* 10, 25 (2018))。各手法ともに、ハイインパクトな国際誌に掲載されており、精度検証も示されているが、計算機による予測結果であることには留意が必要
- CGIでは、手持ちのバリエーションリストをSubmitすると、Driver/Passengerなどの情報をアノテーションしてくれる。結果はTSVでダウンロードが可能で、その後の解析に活用することができる。また、CGIをローカルに動かしたい場合については、申請し承認されることで利用が可能



CGI

(<https://www.cancergenomeinterpreter.org>)

アノテーションフロー



(<https://www.cancergenomeinterpreter.org/faq#q05>より引用)

ゲノムデータには人種の偏りが指摘されており、データの多様性が求められている

- これまでに紹介したデータベースはいずれもキュレーションにおける重要な知識基盤であり、非常に有益。一方で、主に欧米人集団でのバリエーション情報が中心となっている
- 遺伝性腫瘍のパネル検査結果を人種ごとに比較した結果、アジア・ヒスパニックではVUSの割合が多いことが報告されている (Caswell-Jin, J. L. et al, *Genetics in Medicine*, 20 (2): 234-39,2018)
- 非小細胞性肺癌においても、ドライバー遺伝子の体細胞変異の頻度が、東アジア人集団と欧米人集団で異なっていることが報告されている (Kohno, T. et. al., *Transl. Lung Cancer Res*, 4(2):156-164, 2015)
- この他にも、GWAS CatalogやdbGaPに収載されているデータの解析によって、アフリカ、ラテンアメリカ、アジアを祖先とする集団を対象とした研究が、ヨーロッパの集団を対象とした研究に比べ、有意に少なかったことも指摘されている (Landry, L. G. et al., *Health Affairs*, 37 (5): 780-85, 2018)
- このように、現在のゲノムデータに対し、データの偏りが指摘されており、さらに、その偏りがバリエーション解釈に影響することが懸念されている。その為、各民族集団でのデータ蓄積により、**ゲノムデータ多様性**を高めることが、世界的にも求められている

# 日本人ゲノムデータの共有

## 日本人集団でのゲノムデータの集積・共有が必要

- 日本においても、これまでにゲノムデータの蓄積・共有するためのプロジェクトおよびデータベース構築が進められてきた
- 現在、生命科学研究では、その成果を論文で発表するだけでなく、元データや解析プロトコルを公開・共有することが求められている。ヒト関連研究の成果公開の受け皿となるべく、国立研究開発法人科学技術振興機構（JST）バイオサイエンスデータベースセンター（NBDC）が、2013年に**NBDCヒトデータベース**を設立。データ提供および利用の際の審査を行う審査委員会の機能がついたデータベースで、現在、国立遺伝学研究所生命情報・DDBJセンターと共同で運用されている
- また、頻度情報を共有するデータベースとして、京都大学が公開している1208人の日本人ゲノムアレル頻度データベース**HGVD**（Human Genetic Variation Database）や、東北メディカル・メガバンクが収集したサンプルに対する解析で得られたバリエーション情報を公開している**jMorp**などがある。また、NBDCヒトデータベースに収載されているゲノムデータの再解析で得られたアレル頻度情報を公開している**TogoVar**もある。TogoVarおよびjMorpの詳細については後述
- このように研究により得られたデータの蓄積および健常人を対象とした頻度情報の集約が進められている。一方、先に紹介したClinVarのような、疾患に関わるゲノムバリエーションを集約し公開するデータベースはこれまでなかった。そこで、国立研究開発法人日本医療研究開発機構（AMED）では、2016年より、**臨床ゲノム情報統合データベース整備事業**を開始。次スライド以降で、この事業の取り組みを紹介する

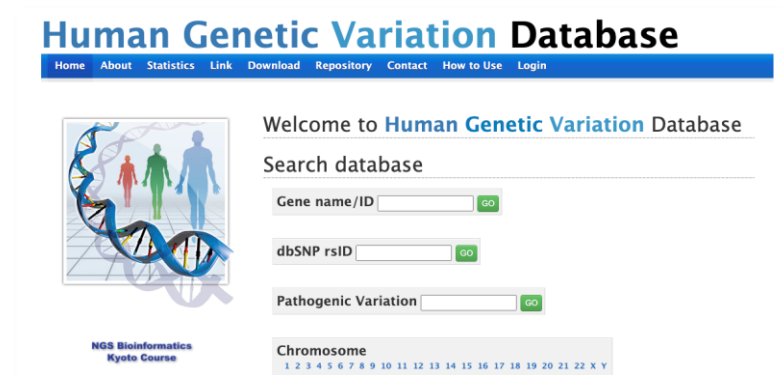
## NBDCヒトデータベース

(<https://humandbs.biosciencedbc.jp/>)



## HGVD

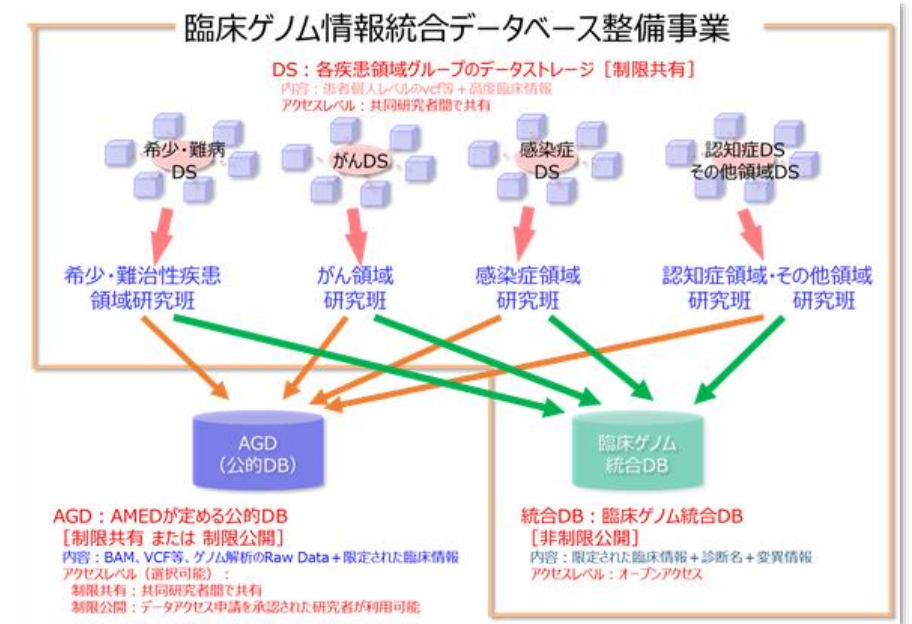
(<https://www.hgvd.genome.med.kyoto-u.ac.jp/>)





### 日本人疾患ゲノム情報の集積と共有に向けた取り組みについて説明する

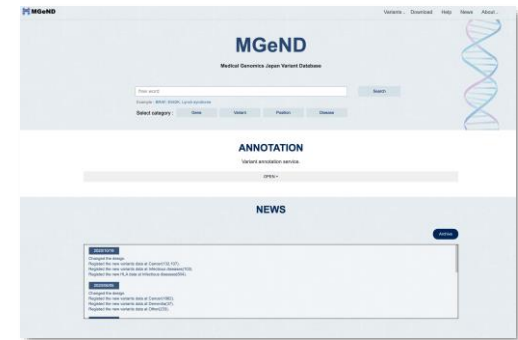
- 2016年に開始された「臨床ゲノム情報統合データベース整備事業」では、「がん」「稀少・難治性疾患」「感染症」「認知症」「難聴」を対象に、患者リクルートからゲノム解析およびその解釈を行う1次研究班と、公開可能な基本情報とゲノムデータを収載する臨床ゲノム統合データベースを構築する2次研究班の、2段階で進められた
- 1次研究班では、各疾患領域におけるゲノム研究実績を持つ11の研究機関からなり、ゲノムデータと、その解釈に必要な患者個人レベルの詳細な臨床情報を集約する、データストレージを構築してきた
- 国立国際医療研究センター・慶應義塾大学・京都大学からなる2次研究班では、1次研究班のデータストレージに集積された臨床ゲノムデータのうち、公開可能な情報を疾患横断的に集約・統合するデータベースとして、**MGeND** (Medical Genomics Japan Variant Database/エムジェンド) を開発してきた。MGeNDは、2018年3月に公開されている
- MGeNDは、誰でも利用可能な「非制限公開データベース」で、稀少疾患に代表される単一遺伝子疾患のみならず、感染症などの多因子疾患も対象に、広くバリエーションデータを収載している。バリエーション単位で統合されたデータに対し、疾患横断的な検索が可能
- 疾患領域によって、主要なゲノム解析方法は異なる。多因子疾患を対象とした研究においては、ゲノムワイド関連解析 (GWAS ; Genome Wide Association Study) が広く用いられている。また、感染症など、免疫にかかる疾患においては、免疫応答の中心的役割を担うタンパク質HLA (Human Leukocyte Antigen ; ヒト白血球抗原) のアリル型に着目した研究が多く行われている。そこで、MGeNDでは、ClinVar等で扱われる「SNV、short indel、SV、CNV」データに加え、GWAS解析で得られるバリエーションセット、そしてHLAアリル型頻度データも対象としている



(<https://www.amed.go.jp/program/list/14/01/006.html>)

### MGeND

(<https://mgend.med.kyoto-u.ac.jp/>)



## MGeNDは、日本人疾患関連バリエーションデータベースである

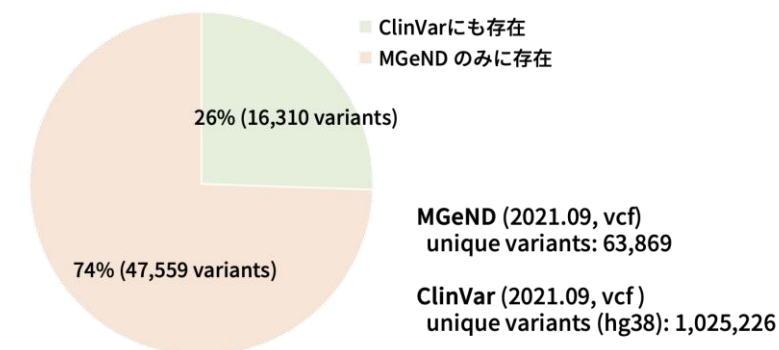
- MGeNDでは提供されたバリエーションデータのみならず、ClinVarをはじめとする、バリエーション解釈でよく用いられる公開データベース情報を統合した、知識データベース。検索結果一覧や各バリエーションページで、データベース横断的に情報を確認することが可能
- MGeNDに登録されたバリエーションデータを表に示す。各疾患領域で多くのバリエーションデータが提供されている。先述のClinVarに掲載されているデータとの比較（図）から、その約8割はMGeNDのみに登録されているバリエーションデータであることがわかっている。このことから、各集団でのデータ集積の取り組みの重要性を認識することができる
- MGeNDでは、染色体位置とバリエーション変化（Reference、Alteration）をバリエーションのIDとして管理している。異なるアセンブリ間のバリエーションは、データベース内部で相互に紐付けがされており、各バリエーションページで互いに遷移可能。
- 登録者からのデータ提供に対しては、MGSから始まるSubmission IDが付与されている。各研究ごとに付与されており、同一研究におけるデータ更新の場合は、同一のSubmission IDが付与されるが、履歴はサブバージョンにてデータベース内部で管理されている。TMGSから始まるIDは、データ受付時に発行される仮管理IDである
- MGeNDに掲載されているデータは、VCF・TSV形式で取得が可能。また、トップページからアクセス可能なアノテーションシステムでは、手持ちのバリエーションデータ（VCF・TSV形式）に、MGeND登録情報やClinVarやToMMoなどの頻度情報を付与することができる。アノテーションデータはそのままCSVやVCFでダウンロードでき、後続の解析に活用可能

## MGeNDに登録されたデータ

（2021年9月時点）

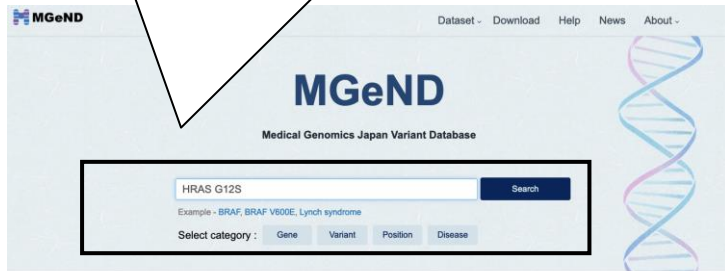
Data type	Variants	GWAS	HLA allele data
がん	185,828 (147,086)	---	---
希少疾患 / 難病	15,399 (2,989)	---	---
感染症	1,276 (1,276)	155,100 (155,100)	1,979 (1,803)
認知症 / 難聴	11,408 (7,668) APOE: 17,876 (5,451)	1,171 (410)	---
その他	1,100 (770)	14,321,737 (584)	---

## ClinVarに登録されているバリエーションとの比較



## MGeNDでは、疾患横断的な検索とアノテーションが可能

- ✓ 遺伝子名・変異表記・疾患名・rs番号・Submission IDなどのキーワードで検索が可能
- ✓ フリーワード検索と、カテゴリ（Gene・Variant・Position・Disease）を絞った検索が可能



### NEWS

#### VCF Annotation Result

Genome : hg19  
Search target variant : MGeND data only  
Showing 1 to 10 of 11 entries

Variant name	AA change	CDS	Japanese frequency	TogoVar	Entry
chr1:g.21584775C>T (rs138879998)	USH2A R4493H	ENST00000307340.7:c.13478G>A, ENST00000307340.8:c.13478G>A, ENST00000674083.1:c.13478G>A, more	●●●○	lgv4774526	1
chr1:g.241667357T>A		ENST00000366560.3:c.1093A>T, NM_001143.3:c.1093A>T	○○○○		0
chr3:g.178936076C>G	PIK3CA L540V	ENST00000263967.3:c.1618C>G, ENST00000263967.4:c.1618C>G,	○○○○		1

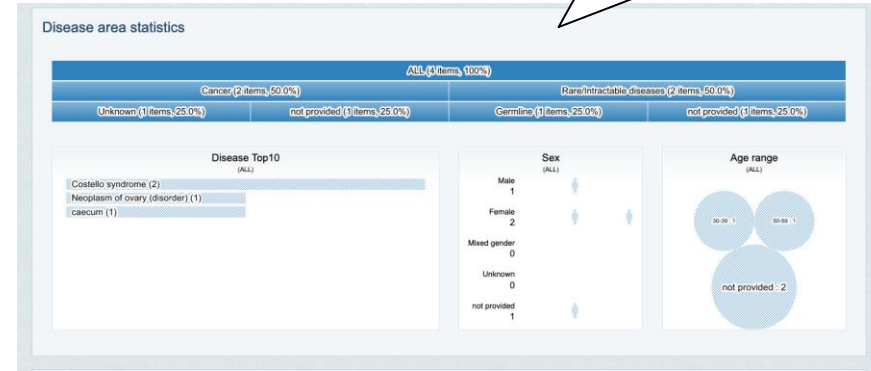
- ✓ 手持ちの変異リスト(TSV/VCF形式)にアノテーションが可能（現在は100変異まで）
- ✓ 結果はCSV/VCFでダウンロード可能

### 検索結果一覧

Variant name	AA change	Entry	Origin	Type	Annotation	ClinVar
NC_000011.9:g.534289G>T	HRAS G12S	1	germline	Variant	Pathogenic	
NC_000011.9:g.534289C>T (rs104894229)	HRAS G12S	4	germline, not provided, unknown	Variant	Pathogenic	Pathogenic
NC_000011.9:g.534289C>A (rs104894229)	HRAS G12C	2	unknown	Variant	not provided, Pathogenic	Pathogenic
NC_000012.11:g.25398285C>T (rs121913530)	KRAS G12S	94	not provided, somatic, unknown	Variant	not provided, Pathogenic	Pathogenic
NC_000012.12:g.25245351C>T (rs121913530)	KRAS G12S	3	somatic	Variant	Pathogenic	Pathogenic
NC_000012.11:g.25398285C>A (rs121913530)	KRAS G12C	305	not provided, somatic, unknown	Variant	not provided, Pathogenic	Pathogenic, Likely pathogenic

- ✓ 検索結果一覧では、登録情報とClinVarの情報と併せて確認できる
- ✓ Filteringで詳細な絞り込みが可能

### バリエーション詳細ページ



- ✓ 変異が検出された、疾患・性別・年齢層の分布が確認できる

Clinical significance	Last evaluated	Condition	Origin	Submission ID	Submitter	Institute	Citation
Pathogenic	2020/04/20	caecum	not provided	MGS000042 (TMGS000093)	Hitoshi Nakagama	National Cancer Center Japan	
Pathogenic		Costello syndrome	germline	MGS000009 (TMGS000039)	Shoji Tsuji	Tokyo University	
Pathogenic		Neoplasm of ovary (disorder)	unknown	MGS000021 (TMGS000080)	Manabu Muto	Kyoto University	
Pathogenic		Costello syndrome		MGS000001 (TMGS000001)	Kenjiro Kosaki	Keio University	

- ✓ 変異詳細と提供者情報が確認できる

# jMorp (ToMMo)

jMorpは、日本人アレル頻度データベースである

- がんゲノム解析では、体細胞バリエーションのフィルタリングや生殖細胞系バリエーションの解釈に、SNP情報を活用する。そのため、バリエーションの日本人集団におけるアレル頻度情報は重要な情報
- **jMorp** (Japanese Multi-Omics Reference Panel) は、東北大学東北メディカル・メガバンク機構 (ToMMo) により運用されるデータベース。東北メディカル・メガバンク計画で得られた日本人を対象としたゲノム・オミクスデータの解析結果を集約・公開している
- jMorpでは、ゲノムバリエーションに関するデータだけでなく、血漿メタボローム/プロテオーム解析や遺伝子発現解析の結果データも収載・公開されている。また、日本人から得られたゲノム配列情報により構築された参照配列 (2022年6月時点での最新版は JG2.1.0) も公開している
- ゲノムバリエーションに関しては、集団内で約0.01%以上の頻度をもつバリエーションを対象に、その頻度情報を「ToMMo \*.\*KJPN」として公開されている。\*.\*Kは、解析の際に使用されたサンプルの数を示す。2015年に初めて公開されたのは1KJPNで、1000人の方を対象に得られた頻度データである。2022年6月30日時点での最新バリエーションデータは、ToMMo 38KJPNで、3.8万人を対象としたゲノム解析により得られたバリエーションのアレル頻度情報である
- 公開されているToMMoの頻度データは、GRCh37/hg19でのポジションであることに注意が必要。頻度データは、VCF形式でダウンロードが可能

jMorp (<https://jmorp.megabank.tohoku.ac.jp/>)

The screenshot shows the jMorp website interface. At the top, there is a navigation bar with links for Repository, GWAS, Downloads, Help, and Login. Below this is a header with two hands and the text "Welcome to Japanese Multi Omics Reference Panel." The main content area is divided into two columns. The left column, titled "Datasets available on jMorp", lists various omics data types with corresponding icons and buttons: Phenome (Metagenome, PGx), Metabolome (Metabolome), Proteome (Proteome), Transcriptome (ToMMo ISO-Seq, IMM Transcriptome), Methylome (IMM Methylome), Genome Variation (38KJPN (short-read based SNV/INDEL), 38KJPN-HLA (short-read based HLA), JSV1 (long-read based SV), 8.3KJPN-SV (short-read based SV)), and Genome Sequence (JG2.1.0). The right column, titled "jMorp release 202206", contains a news item dated June 30th, 2022, about the release of the 38KJPN frequency panel. It includes a "More" button. Below this is a "jMorp Publication" section with a citation for Tadaka et al. (2020) and Takayama J. et al. (2021).

# TogoVar

TogoVarは、日本人のバリエーションアレル頻度を集約した知識データベースである

- **TogoVar (日本人ゲノム多様性統合データベース)** は、NBDCとライフサイエンス統合データベースセンター (DBCLS) が共同で開発したデータベースで、日本人のバリエーションアレル頻度と関連する公的データベース情報をワンストップで取得できる知識データベースのこと
- TogoVarでは、先述のClinVarやHGVD, jMorp (ToMMo 8.3KJPN) のデータを収載している他に、NBDCヒトデータベースに寄託されたゲノムデータに対する統一パイプライン再解析で得られたバリエーションアレル頻度情報を収載している
  - ✓ JGA-NGS/SNP: NBDCヒトデータベースのJGA (Japanese Genotype-phenotype Archive) に登録された個人レベルのゲノム配列データ (各々125/183,884サンプル) を再解析して検出したバリエーションアレル頻度データ
  - ✓ GEM-J WGA: AMEDが推進する**GEM Japan** (Genome Medical Alliance Japan) プロジェクトの一環として実施された研究によるデータで、7609サンプルを対象に得られたバリエーションアレル頻度データ
- 各バリエーションアレルは、tgVから始まるTogoVar IDにより管理されている。バリエーションアレルは、GRCh37 (hg19) を対象に表記されているので、検索ではその点に留意が必要。VCF形式で取得可能なダウンロードデータでは、GRCh38にマッピングされたデータも利用可能

TogoVar (<https://togovar.biosciencedbc.jp/>)

variant ID	rsID	position	ref/alt	type	gene	consequence
tgV8222268	rs144727480	1:10117	C>T	Deletion		Intergenic variant
tgV8222268	rs144727480	1:10145	A>G	Deletion		Intergenic variant
tgV8222268	rs127528992	1:10147	C>T	Deletion		Intergenic variant
tgV8222268	rs128668604	1:10150	G>T	Deletion		Intergenic variant
tgV8222268		1:10151	C>T	Deletion		Intergenic variant
tgV8222268		1:10152	A>G	Deletion		Intergenic variant
tgV8222268	rs118401806	1:10168	C>T	Deletion		Intergenic variant
tgV8222268	rs126627190	1:10173	C>T	Deletion		Intergenic variant
tgV8222268	rs148947538	1:10174	C>T	Deletion		Intergenic variant
tgV8222268		1:10179	C>T	Deletion		Intergenic variant
tgV8222270		1:10225	C>T	Deletion		Intergenic variant
tgV8222270	rs1287031179	1:10248	A>G	Deletion		Intergenic variant
tgV8222270	rs118401806	1:10248	A>G	Deletion		Intergenic variant
tgV8222272	rs786958100	1:10247	T>C	SNV		Intergenic variant
tgV8222276	rs140194108	1:10255	A>G	Deletion		Intergenic variant
tgV8222272		1:10255	A>G	Insertion		Intergenic variant
tgV8222281	rs1307882531	1:10263	A>G	Deletion		Intergenic variant
tgV8222281		1:10291	C>T	Insertion		Intergenic variant
tgV8222281	rs145427725	1:10291	C>T	SNV		Intergenic variant
tgV8222281	rs100211756	1:10321	C>T	SNV		Intergenic variant
tgV8222288		1:10327	T>A	Deletion		Intergenic variant
tgV8222288	rs150966722	1:10330	C>T	Deletion		Intergenic variant

## 日本人/アジア人集団でのバリエーションアレル頻度情報

Dataset	Population	Allele count	Total	Frequency	Genotype count	Filter status	Quality score
GEM-J WGA	Japanese	2,279	14,990	0.152		PASS	688936
JGA-NGS	Japanese	7	250	0.028		LowCoverage	162.84
JGA-SNP	Japanese	108	836	0.129	6 / 96 / 316	PASS	
ToMMo 8.3KJPN	Japanese	2,574	16,758	0.154		PASS	
gnomAD Genomes	Total	4,701	31,318	0.150		PASS	2639158.3

## 日本でのデータシェアリングポリシーについて説明する

- ここまで様々なデータベース・知識データベースを紹介したが、ゲノムデータは機微な情報であることから、データの共有・公開の範囲については段階的に分けて考える必要がある
- AMEDでは、2016年にデータ共有に関する指針「データシェアリングポリシー」を発表した。このポリシーでは、以下の3つのカテゴリを規定している
  - ✓ 制限共有データ：データ登録者が規定するデータマネジメントプランに記載された研究者、及びデータアクセス申請を承認された研究者間で共有するデータ
  - ✓ 制限公開データ：当該データベースの規約等に従って、利用目的、利用方法等を明らかにした上で、データアクセス申請を承認された研究者が利用することが可能なデータ
  - ✓ 非制限公開データ：アクセスに制限なく誰でも利用することが可能なデータ
- 先に紹介した臨床ゲノム情報統合データベース整備事業において、1次研究班が共有する患者個人レベルの詳細な臨床情報とゲノムデータは、「制限共有データ」として、MGeNDに集約されるデータは「非制限公開データ」として定義されている。NBDCヒトゲノムデータベースのJGAは「制限公開」、AGD（AMEDゲノム制限共有データベース）は、「制限共有」（NBDCでは「グループ共有」と呼んでいる）
- データ共有に関する国際的な取り組みとして、**GA4GH**（Global Alliance for Genomics and Health）を紹介する。GA4GHは、各国間でのゲノムデータの共有によりゲノム医療の発展を目指す国際協力組織のこと。その実現に向け、各国での取り組みを共有し、またそれらに基づき国際標準を構築している。2022年時点で、90カ国、600超の機関（各国政府の研究費配分機関、研究機関、企業等）が加盟している。日本からも、15を超える機関が参画しており、AMEDからも各事業の成果を統合するGEM Japan（GEnome Medical Alliance Japan）プロジェクトが参画している

### 制限共有 (グループ共有)



承認された研究者で共有が可能

### 制限公開



承認された研究者が利用可能

### 非制限公開



誰でも利用可能



**Global Alliance**  
for Genomics & Health  
Collaborate. Innovate. Accelerate.

(<https://www.ga4gh.org/>)

## データ共有のためには、どのような情報が必要か

- 本章では、がんゲノム医療に関わる代表的なデータベースと、データ共有の重要性について紹介してきた。今後、自身の研究で得られたバリエントデータを登録することがあるかもしれない。登録する内容は、各データベースのアクセスレベル・共有範囲により異なるが、ここでは、MGeNDのバリエントデータ登録に必要な項目を通して、どのようなデータを準備する必要があるのか、確認する。MGeNDバリエントデータ登録フォーマットは、ClinVarに準拠しているため、必須項目はおおよそ共通していると捉えることができる
- 以下は、バリエントに関する必須項目（この他に登録者情報は必要）
  - ✓ **疾患名**：OMIM、MeSH、MedGen、Orphanet、HPO、ICD10 などのコードで指定できる。  
該当するコードがない場合は、自由記載で記載可能だが、表記揺らぎの観点からコードでの入力を推奨している
  - ✓ **遺伝子名**：HGNCに基づくGene Symbol（遺伝子名）を指定
  - ✓ **遺伝型情報**：ゲノム位置（染色体番号・位置・Reference・Alteration）もしくはHGVS表記（e.g. NM\_001127500.1:p.Ile284Met）で記載。  
パネル検査では、結果として標的分子のアミノ酸位置のみが返却される場合がある。その場合は、どの配列（先のHGVSの例だと、NM\_001127500.1）を基準としているかも合わせて提出。基準となる配列が不明な場合、ゲノム位置を特定できないためMGeNDでは登録ができない。どのトランスクリプト配列に基づく表記かは、とても重要
  - ✓ **臨床的意義**：ClinVarで使用されているタームに準拠（スライド5を**ご参照**）
  - ✓ **基本情報**（MGeNDのみ）：年齢層（unknown、not provided含む）・性別（「不明」「混合」等を含む）
- 必須ではないが、以下の情報も、バリエントの検出された背景を知る上で重要なため、可能な限り提出
  - ✓ 解析プラットフォーム・解析パネル・解析ソフトウェアおよびバージョン、等

# コラム がんゲノム医療データの統計的活用

国立がん研究センター 研究所 生物情報学分野 分野長

加藤護



# がんゲノムデータの統計的利用 (1)

(コラム) がんゲノム医療データの統計的活用

がんゲノム学\*データの統計的利用により、主要ながん種における遺伝子異常の集団頻度は、ほぼ解明された

- がんゲノムデータを生成する基本的技術には主に下記があげられる
  - ✓ 次世代シーケンサー
  - ✓ 次世代シーケンサー・データから遺伝子異常を検出する情報技術
  - ✓ 検出された遺伝子異常を注釈付けする情報技術
  - ✓ がん関連遺伝子の知識をまとめたデータベース
- がんゲノムデータは、大まかに3つに分類できる
  - ✓ \***がんゲノム学**のデータ：がんゲノムの性質・特徴を明らかにする学問
  - ✓ **がんゲノム臨床研究**のデータ：がんゲノムの性質・特徴を利用した技術による、患者への臨床的な有用性を主題とする研究
  - ✓ **がんゲノム医療**のデータ：がんゲノムの性質・特徴を利用した技術による医療
- がんゲノム学データについて
  - ✓ 検出された遺伝子異常を統計的に分析することにより、主要ながん種におけるがん関連遺伝子とそのDNA異常、および集団頻度は、ほぼ解明された
  - ✓ 研究主体の例
    - 国際がんゲノムコンソーシアム
    - がんゲノムアトラス
    - 個々の研究グループ
  - ✓ 体細胞変異を主対象とし、多くは全エキソームシーケンスである。がん組織のトランスクリプトームを含む場合もある
  - ✓ 集積されたデータは多数の研究で再利用され、データを産出した元プロジェクトでは尽くせなかった発見もなされてきた

## がんゲノム学データの統計的利用

- がん関連遺伝子とその異常の解明
- 第一期 (2008年頃～)



## がんゲノム臨床研究データの統計的利用

- 薬剤適合率の見積もり
- 第二期 (2013年頃～)



## がんゲノム医療データの統計的利用

- 薬剤奏効の傾向、希少症例群の解明
- 最近 (2019年頃～)

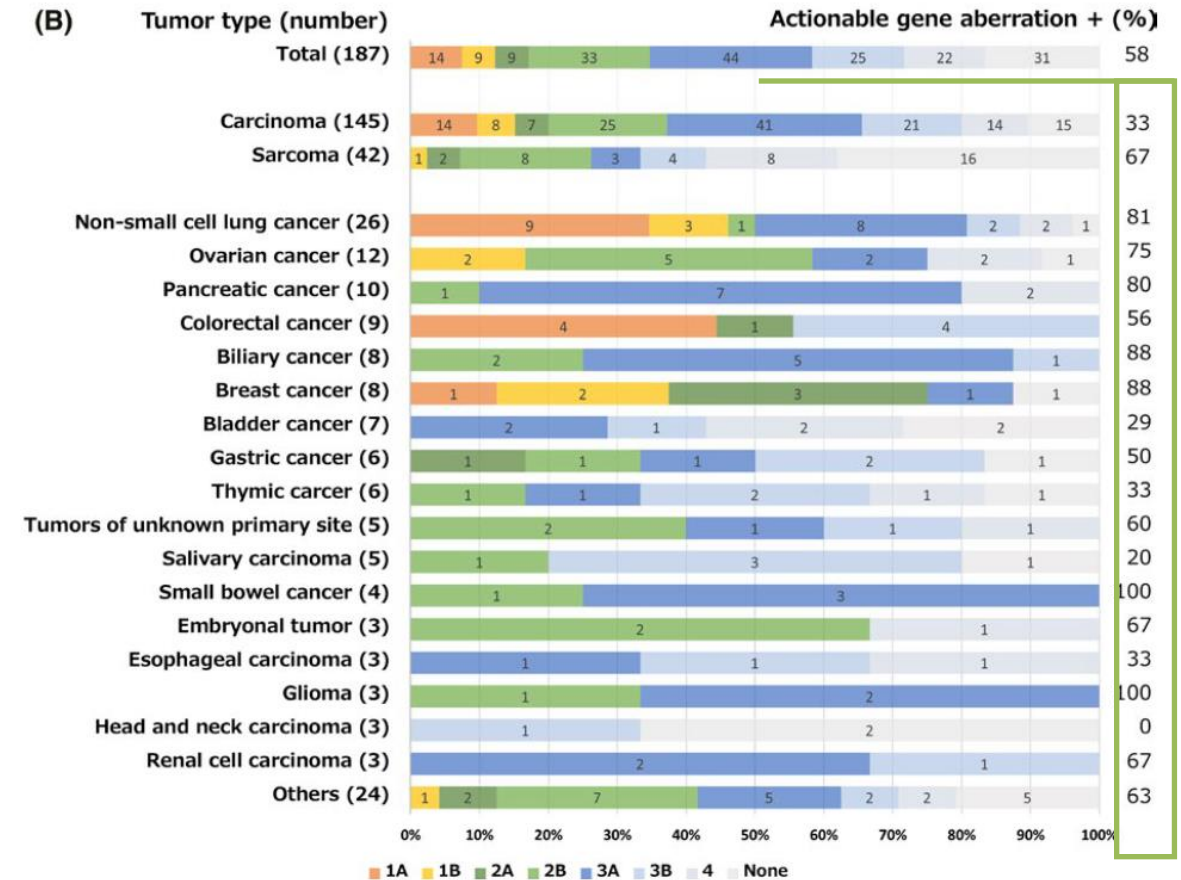
# がんゲノムデータの統計的利用 (2)

(コラム) がんゲノム医療データの統計的活用

がんゲノム臨床研究データの統計的利用により、がんゲノム医療の薬剤適合率は50%前後であることが分かった

がんゲノム臨床研究データについて

- がんゲノム臨床研究：がんゲノム学に比し、より臨床的性格の強い研究
- 研究の例
  - ✓ MSK-IMPACT (Zehir et al, Nat Med, 2017)
  - ✓ TOP-GEAR (Sunami et al, Cancer Sci, 2019)
  - ✓ SCRUM-Japan GOZILA (Nakamura et al, Nat Med, 2020)
- より医療現場に即した設定
  - ✓ FFPE 試料や、血中遊離DNA試料。対してがんゲノム学では、コストがかかる凍結試料が主体
  - ✓ 包括的がんゲノムプロファイリング検査で使われるようなターゲット遺伝子・シーケンスであり、対象遺伝子数は約100-500遺伝子。対してがんゲノム学では、全エキソームシーケンスであり、約2万遺伝子
- 検出された遺伝子異常と分子標的薬のデータベースとを結び付けて統計分析にすることにより、多くの研究を総合すると、**検出された遺伝子異常のうち約半数が分子標的薬の対象となることが判明した**。例えば右の研究では、対象となった全がん種の58%の症例で、(当該研究が定義するエビデンス・レベル3A以上の)分子標的薬がマッチした
- 諸事情のため**実際の投与率は約10-20%程度であることが判明した**



(Sunami et al, Cancer Sci, 2019)

# がんゲノム・データの公開データベース

(コラム) がんゲノム医療データの統計的活用

- がんゲノム・データを整理した公開データベース（ポータルサイト）が存在する。しかし現時点、臨床データは不足している
- 再利用可能なように、個々のプロジェクトでコールされた遺伝子異常をまとめた公開データベース（ポータルサイト）が存在する。以下は、代表的なデータベースである
- これらのデータベースに蓄積されたデータを統計分析し、ゲノムの性質を解明していくには問題ないが、精密医療に直接適用されるような発見をするには、現時点、臨床データが不足している（たとえば追加の効果予測マーカーを見出すには、使用した薬剤や薬剤応答のデータが必要）

## ✓ NCI Genomic Data Commons

(<https://gdc.cancer.gov/>)

- 米国国立がん研究所が管理。米国を主体に、多くのプロジェクトのデータが整理・集積されている

## ✓ cBioPortal (<https://www.cbioportal.org/>)

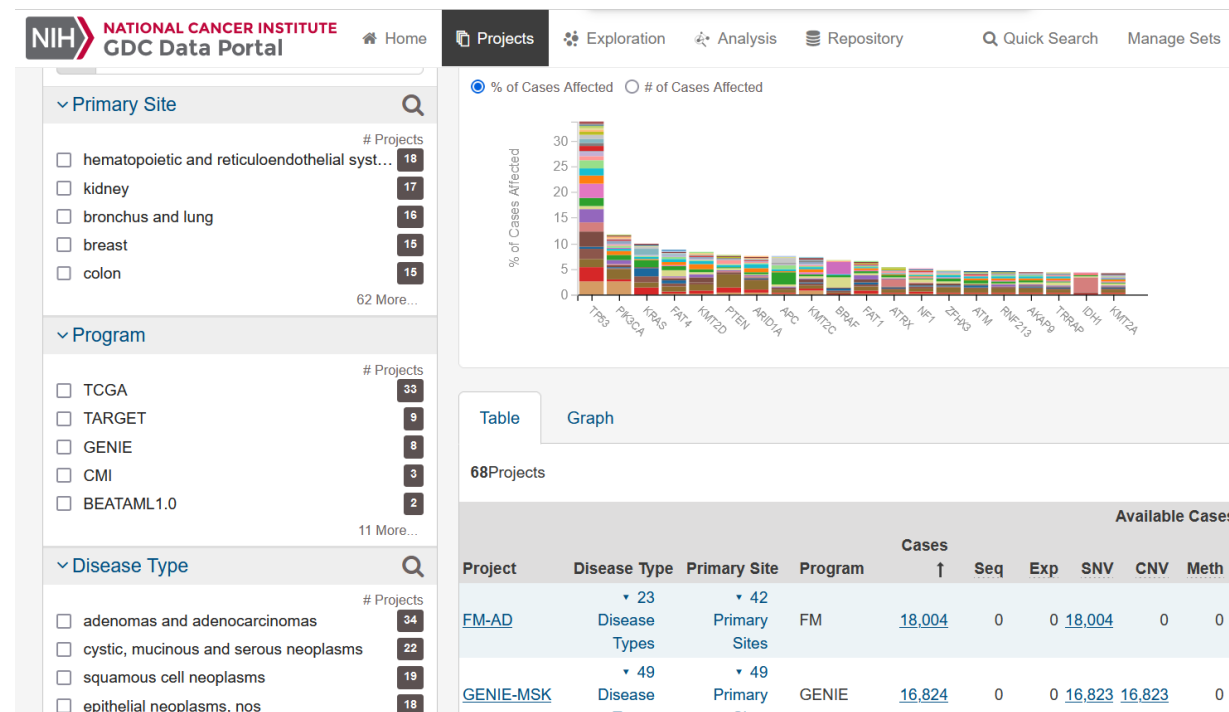
- メモリアル・スローン・ケタリングがんセンターが管理。個々の論文ごとにもまとまっている

## ✓ ICGC-ARGO (<https://www.icgc-argo.org/>)

- ICGC（国際がんゲノムコンソーシアム）が管理。正確にはデータベース・サイトではないが、ICGC 25k（ICGC 第一期のプロジェクト群）やPCAWG のデータ、さらに現在実施されているICGC-ARGO のデータが管理されている

## ✓ AMED AGD (<https://gr-sharingdbs.biosciencedbc.jp/agd>)

- JST NBDC が管理。本邦のプロジェクトにおける、コールデータよりもFASTQのような生データの収集に注力



(NCI Genomic Data Commons)

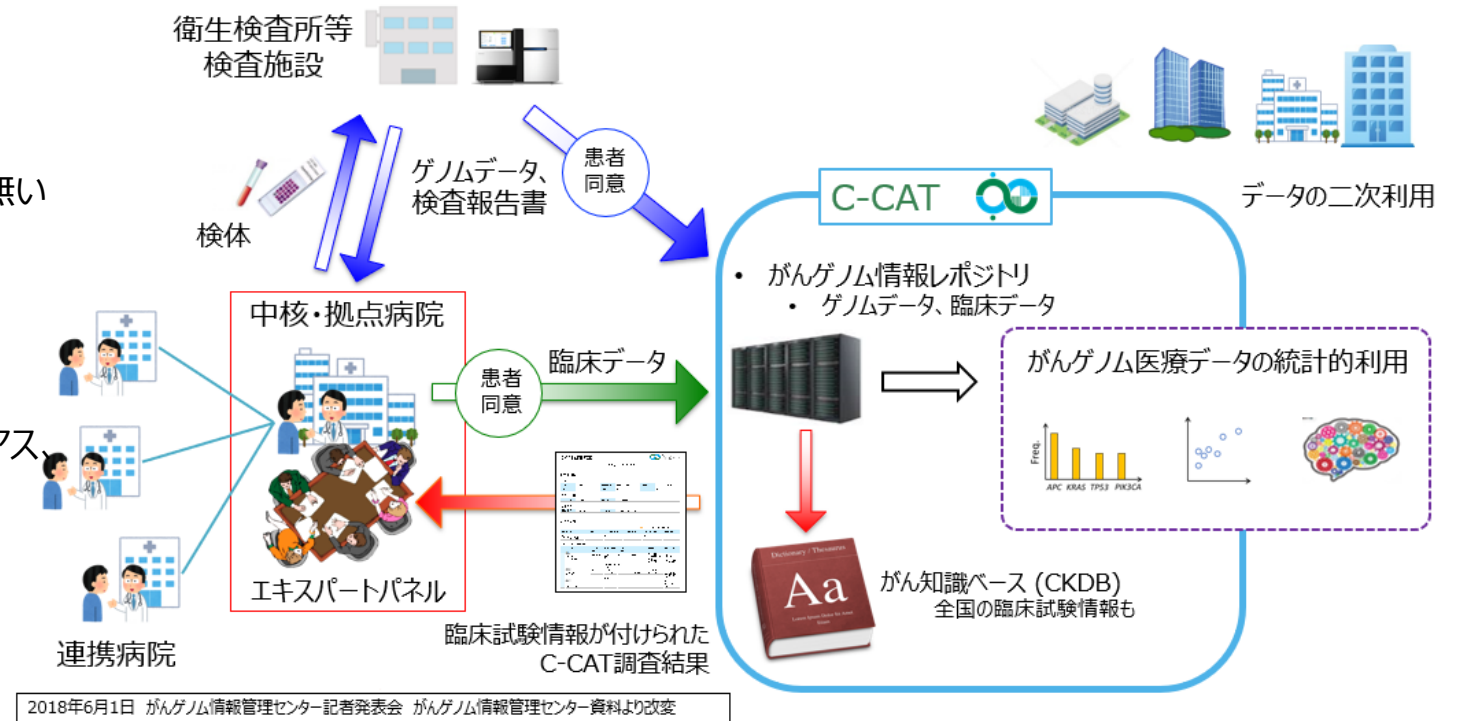
# がんゲノム医療データ

(コラム) がんゲノム医療データの統計的活用

がんゲノム医療が実施されたデータを蓄積して、“実世界（リアルワールド）”データとして活用する動きが始まっている

- がんゲノム学やがんゲノム臨床研究のような“研究”のデータではなく、**包括的がんゲノムプロファイリング検査**として医療的に実施されたがんゲノム“医療”のデータ。日本では、国立がん研究センターの**がんゲノム情報管理センター（C-CAT）**が、保険診療で実施されている包括的がんゲノムプロファイリング検査のデータを**集積している**
- 奏効率や、奏効の詳細な分析
- 希少患者群におけるゲノム異常の分析
- メリット
  - ✓ データを収集する研究プロジェクトを立ち上げる必要が無い
  - ✓ サンプルサイズが大きい
  - ✓ 電子カルテにある臨床データと結合しやすい
- デメリット
  - ✓ 臨床試験データに比べると、データの質が落ちる（バイアス、層別の問題）
  - ✓ ターゲット遺伝子・シーケンス
- 必然的に、探索研究となる
- 臨床データと結びついた分析は、**clinico-genomic analysis（臨床ゲノム解析）**とも呼ばれる

## 国立がん研究センター がんゲノム情報管理センター（C-CAT）



2018年6月1日 がんゲノム情報管理センター記者発表会 がんゲノム情報管理センター資料より改変

(がんゲノム情報管理センター資料より改変)

# がんゲノム医療データの統計的活用例1

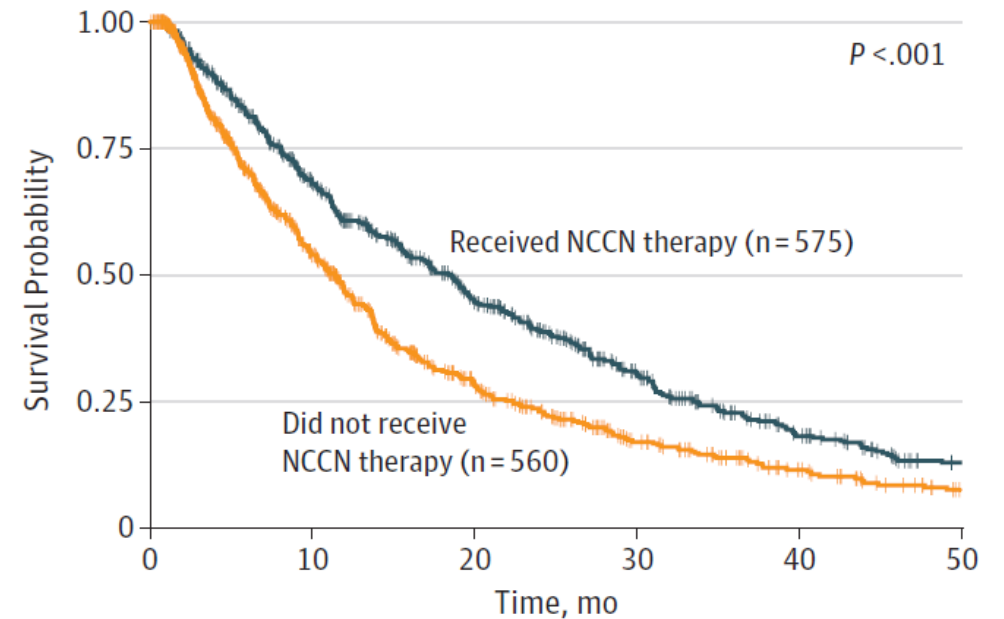
(コラム) がんゲノム医療データの統計的活用

電子カルテのデータと統合されたがんゲノム医療データの統計分析は、がんゲノム研究における従来の発見を再確認した

● Singal et al, 2019, JAMA

- ✓ 2011-2018 年の間に包括的がんゲノムプロファイリング検査を実施し、蓄積された38がん種28,998症例のデータベースの中から抽出された、非小細胞性肺がんの4,064 症例
- ✓ 臨床データは主として電子カルテから抽出され、校閲
- ✓ エンドポイント：全生存期間、maximal therapeutic response、など
- ✓ NCCN (National Comprehensive Cancer Network) ガイドラインにあるドライバー遺伝子の変異を持つ症例のうち、NCCN ガイドラインに従って分子標的薬治療を受けた症例群では、そうでなかった症例群に比べ、より長い全生存期間が観察された (右図)
- ✓ 抗PD-1/PD-L1 治療を受けた症例のうち、腫瘍変異負荷が高い (TMB 20 以上) 症例群では、低い症例群に比べ、より長い全生存期間が観察された
- ✓ 臨床ゲノム (“clinico-genomic”) データベースにおける、電子カルテ・データと統合された包括的がんゲノムプロファイリング検査データの解析は、これまでがんゲノム学やがんゲノム臨床研究が発見してきた相関関係を再確認した

B NCCN-directed therapy



No. at risk

Received NCCN therapy	2	168	142	91	59	32
Did not receive NCCN therapy	55	196	93	54	28	13

(Singal et al, 2019, JAMA)

# がんゲノム医療データの統計的活用例2

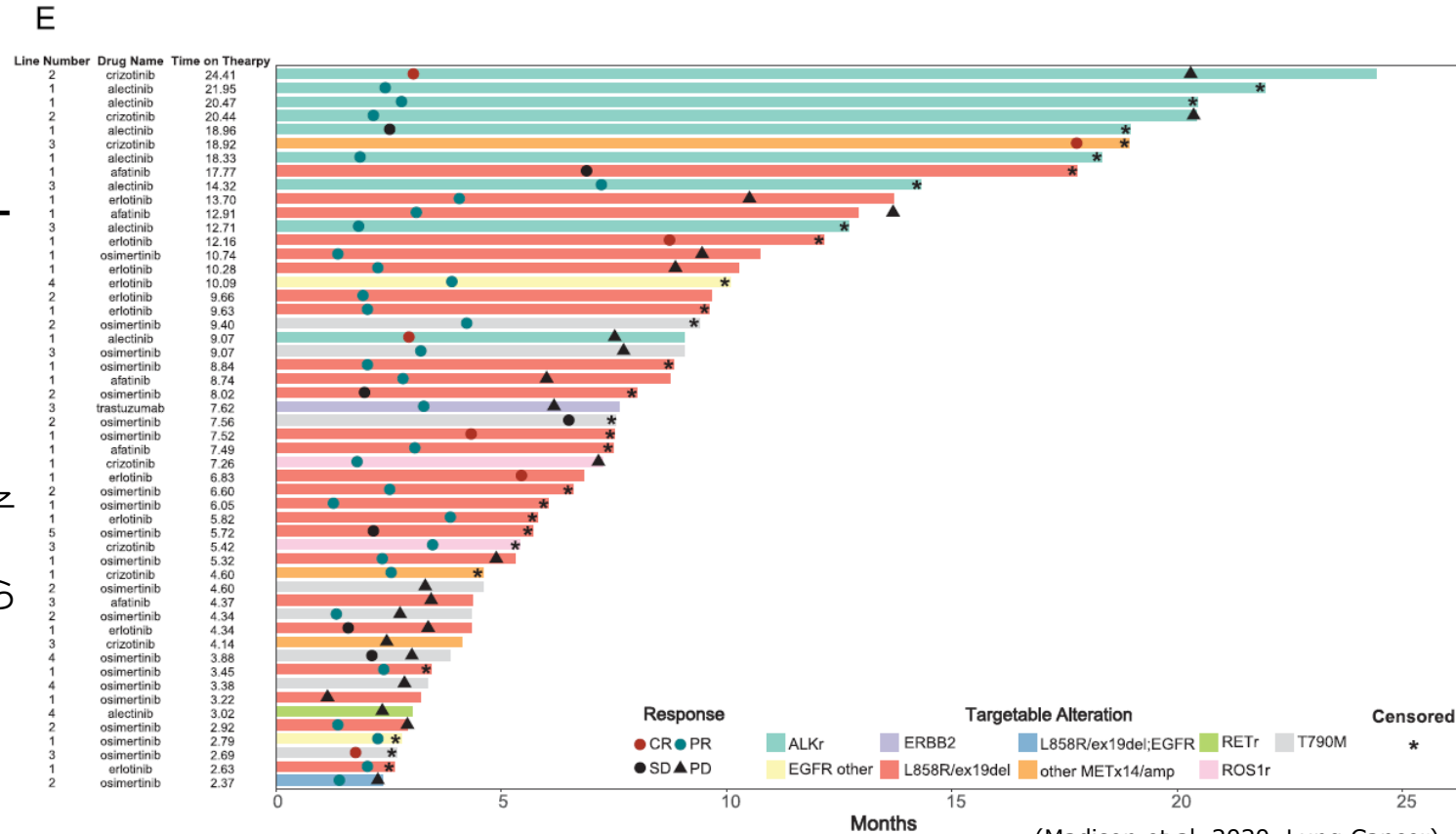
(コラム) がんゲノム医療データの統計的活用

がんゲノム医療データの統計分析によれば、cfDNA 検査の臨床的有用性は、組織検査に匹敵する

## ● Madison et al, 2020, Lung Cancer

- ✓ リキッドバイオプシー (cfDNA) による包括的がんゲノムプロファイリング検査
- ✓ 組織による包括的がんゲノムプロファイリング検査との有用性を比較
- ✓ 非小細胞性肺がんの6,491 症例 (リキッドバイオプシーによる2016-2019 年での934 症例 and/or 組織検査による2012-2019 年での5,570 症例。13症例は重なっているが、独立して集計)
- ✓ Singalら (2019, JAMA) と同じような仕組みで構築された臨床ゲノムデータベースを利用
- ✓ 右図で示されるように、リキッドバイオプシーで適合した分子標的薬が処方された症例に対し、処方薬・治療継続期間・最良効果測定時点・PD までの期間が電子カルテからほぼ自動的に抽出された
- ✓ 主に電子カルテから抽出された分子標的薬治療の実世界 ("real-world") 無増悪期間と全奏効率は、リキッドバイオプシーと組織検査とで、ほぼ同じであった

最良効果測定の時点とPD (進行) までの期間、治療継続期間を示すスマープロット



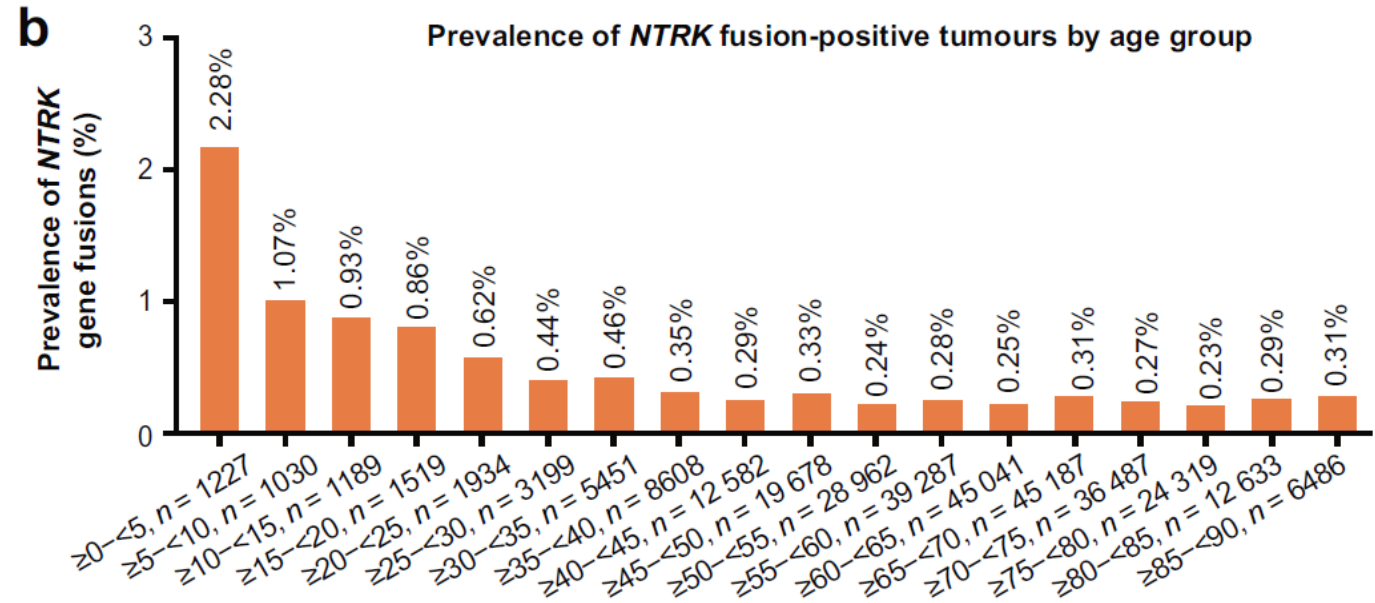
(Madison et al, 2020, Lung Cancer)

# がんゲノム医療データの統計的活用例3

(コラム) がんゲノム医療データの統計的活用

がんゲノム医療データの統計分析によれば、*NTRK* 融合遺伝子を持つ希少がん症例は5歳以下で比較的頻度が高い

- Westphalen et al, 2021, npj Precision Oncology
  - ✓ *NTRK* 融合遺伝子を持つがん症例は希少症例であり、罹患率は約0.3%と言われる
  - ✓ 包括的がんゲノムプロファイリング検査を2013-2019年に実施した、75 がん種にわたる295,676 症例のゲノムデータベースを利用
  - ✓ うち、45 がん種の889 症例が*NTRK* 融合遺伝子を持ち、罹患率は実際0.30%であった
  - ✓ 新規の融合遺伝子パートナーが、58 個見つかった
  - ✓ 最も罹患率が高かった年齢層は、5歳以下 (2.28%)
  - ✓ 最も罹患率が高かったがん種は、唾液腺がん (2.62%)
  - ✓ 大腸がん群において、*NTRK* 融合遺伝子陽性とマイクロサテライト不安定性とで、相関があった
  - ✓ また、このマイクロサテライト不安定性大腸がん群においては、*NTRK* 融合遺伝子陽性と*BRAF* 変異陽性とは、相互排他的であった



(Westphalen et al, 2021, npj Precision Oncology)

# コラム 蓄積情報によるデジタル創薬や ドラッグ・リポジショニング

大阪大学 大学院医学系研究科 遺伝統計学

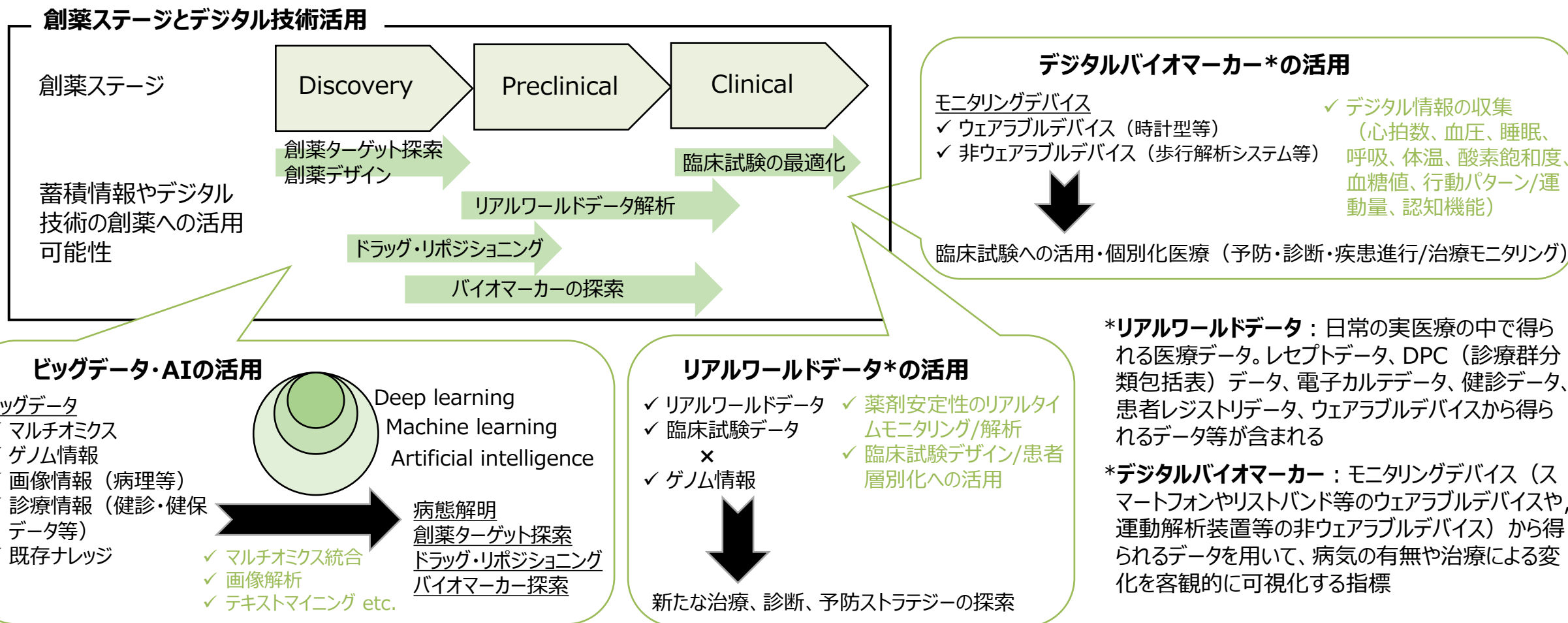
岡田随象、小沼貴裕



# 創薬におけるデジタル技術活用

(コラム) 蓄積情報によるデジタル創薬や  
ドラッグ・リポジショニング

- 医薬品の研究開発の生産性は近年下降傾向にあり (Nature Reviews Drug Discovery 2012 11:191-200)、従来の研究開発ではできなかったことを可能にするデジタル技術を活用した創薬 (デジタル創薬) への期待が高まってきている



\*リアルワールドデータ：日常の実医療の中で得られる医療データ。レセプトデータ、DPC (診療群分類包括表) データ、電子カルテデータ、健診データ、患者レジストリデータ、ウェアラブルデバイスから得られるデータ等が含まれる

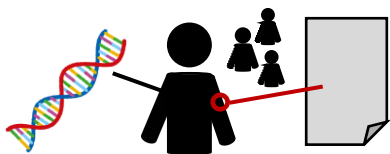
\*デジタルバイオマーカー：モニタリングデバイス (スマートフォンやリストバンド等のウェアラブルデバイスや、運動解析装置等の非ウェアラブルデバイス) から得られるデータを用いて、病気の有無や治療による変化を客観的に可視化する指標

# 大規模ゲノム情報の創薬への活用

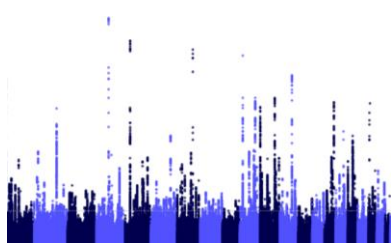
(コラム) 蓄積情報によるデジタル創薬や  
ドラッグ・リポジショニング

- 近年、UK-Biobank（イギリス）、FinnGen（フィンランド）、ToMMo（東北大学東北メディカル・メガバンク機構）等の前向きコホート研究のバイオバンクが、大規模なゲノム情報・表現型情報を公開するなど、疾患ゲノム情報が大規模に入手できるようになってきた。それらの情報を用いた、創薬に関連する様々な段階への活用（ゲノム創薬）が盛んに試みられている

大規模ゲノム情報・  
電子カルテ情報

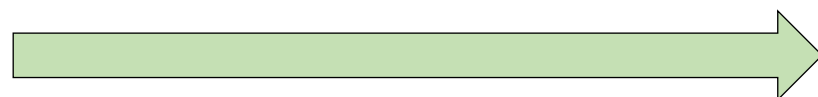
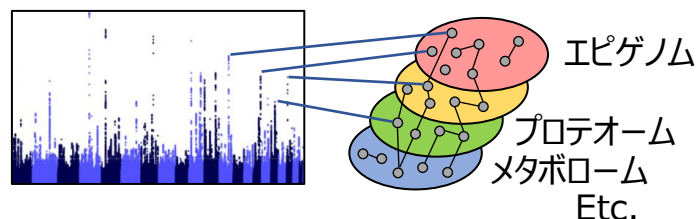


大規模ゲノム・  
表現型情報



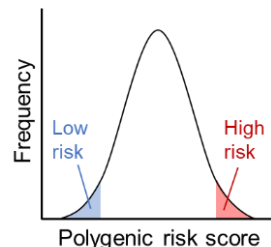
大規模ゲノム・表現型情報を用いた  
**Genome-Wide Association Study (GWAS)**

横断的オミクス解析  
-大規模ゲノム情報とマルチオミクスデータとの統合-



Risk variants

$$\sum_i \beta_i \times \text{variant}_i$$



**Polygenic risk score\*解析**  
-個人ごとの遺伝的変異の組み合わせによる  
疾患リスクの定量-

**ゲノム創薬による新たな創薬アプローチ**

大規模ゲノム情報・表現型情報に基づくデジタル創薬

- ✓ 機械学習/AI活用によるネットワーク解析
- ✓ 薬剤データベース情報との統合、等
- 解析薬剤ターゲット探索、ドラッグ・リポジショニング\*、バイオマーカー探索

**個別化医療への応用**

- 個人別の疾患リスク予測による
- ✓ 臨床試験における患者層別化
- ✓ 個別化医療/予防医療への貢献

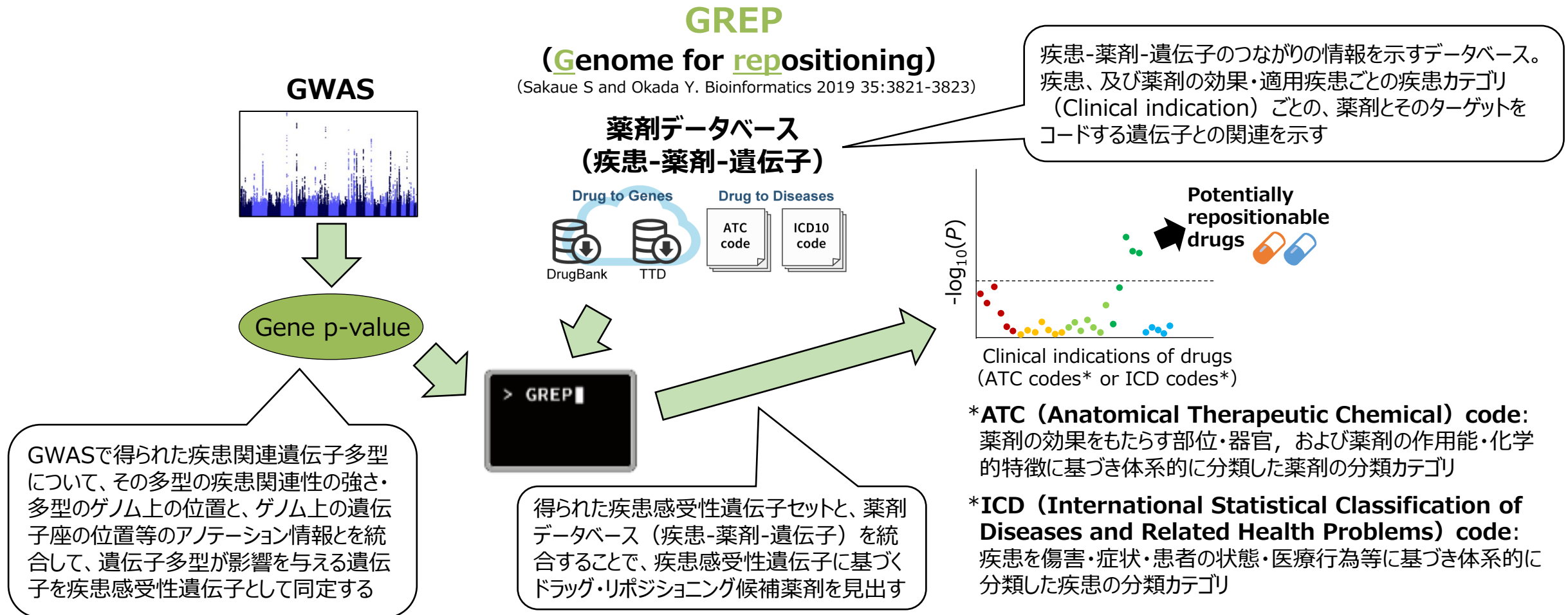
\***ドラッグ・リポジショニング**：ある疾患に既に有効であることがわかっている治療薬から、別の疾患にも有効である治療薬を見出すこと。特に、既に臨床使用が承認されている治療薬の場合は、開発期間の短縮や開発コストの低減が期待できる

\***Polygenic risk score**：これまでのゲノム解析で推定された、一つずつの遺伝子変異がヒトの形質に与える効果量（ $\beta$ ）を用いて、個人ごとの遺伝子変異の組み合わせとそれらの効果量を掛け合わせて和をとったスコアであり、その個人の疾患リスクの推定値を示す

# 大規模ゲノム情報を用いた薬剤探索

(コラム) 蓄積情報によるデジタル創薬や  
ドラッグ・リポジショニング

- 大規模ゲノム解析 (GWAS) で同定された疾患感受性に関与する遺伝子変異の情報と、オミクス情報、及び薬剤データベースを統合的に解析することで、これまで見出されていなかった、疾患に有効と期待される薬剤を見出すことが可能となる



# 大規模ゲノム情報を用いた薬剤探索

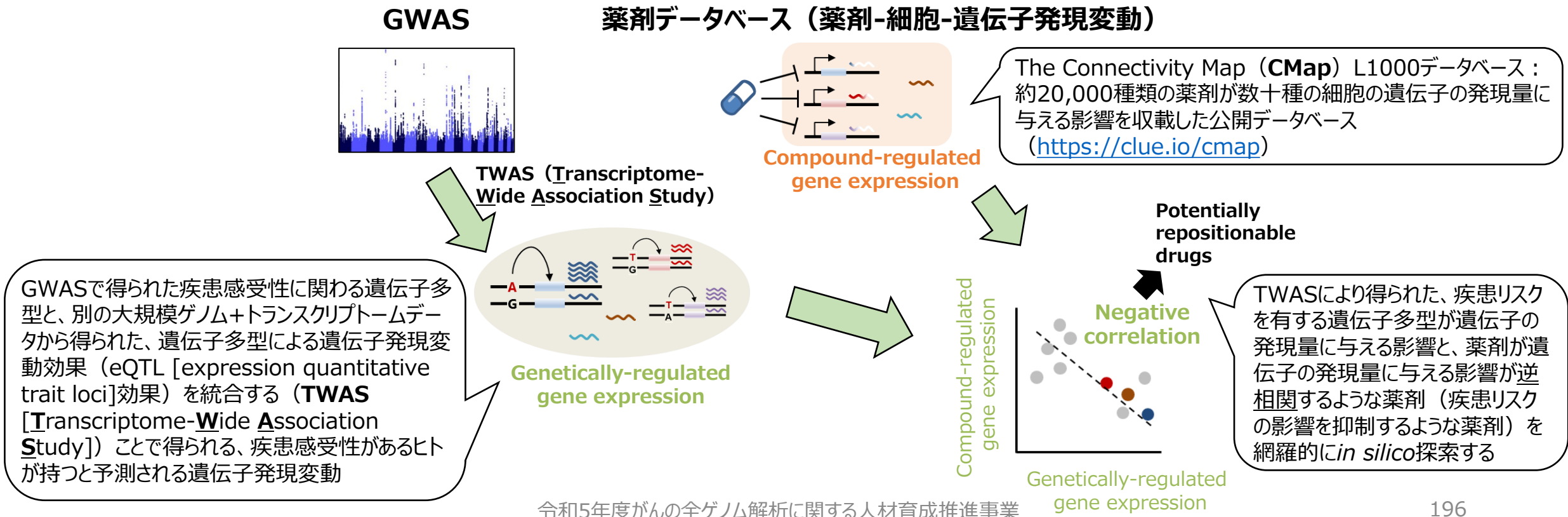
(コラム) 蓄積情報によるデジタル創薬や  
ドラッグ・リポジショニング

- 大規模ゲノム解析 (GWAS) で同定された疾患感受性に関与する遺伝子変異の情報と、オミクス情報、および薬剤データベースを統合的に解析することで、これまで見出されていなかった、疾患に有効と期待される薬剤を見出すことが可能となる

## Trans-Phar

(Integration of **Trans**criptome-wide association study and **Phar**macological database)

(Konuma T et al. Hum Mol Genet 2021 30:294-304)



# コラム

## AIによるリアルワールドデータや PHRとの統合解析

株式会社ヒューマノーム研究所  
小貫真希

# リアルワールドデータ解析とは

(コラム) AIによるリアルワールドデータやPHRとの統合解析

リアルワールドデータ解析に至るまでの手順概要は下記の通り

## リアルワールドデータの解析手順

全ゲノム解析が進み、がんの原因変異や分子標的薬による適切な投薬が進むだろう。しかし、患者が直ぐに寛解することはなく、病院への通院や退院後も体調の急変への対応が必要である。ゲノム解析が進めば進むほど、病院の外での体調管理、それに伴ったデータ計測が重要性を増してくる。

計測されたデータは、適切な手法を用いて解析することで、体調の異変などを自動で検知することが可能となる。その解析手法として、AIは非常に注目されている。医療分野においても、AIは研究開発で用いられている。その一例として、医用画像を用いた診断が挙げられる。これには、任意の画像とそれに対応する病名などのラベルをひとまとめにしたデータセットと深層学習を用いて、画像から病名を推定するモデルが数多く提案されている。この生成されたモデルを用いると、熟練の医師の診断と同程度の推定精度になるという報告もある。

AIを用いたデータ解析は、人類の繁栄に大きく貢献する可能性がある。ただし、AIの実世界への応用には、大きな障壁がある。多くのAIの研究では、**既に整理されたデータ**（例として、MNISTの手書きデータベースが挙げられる）を用いて成果を挙げているからだ。そのような研究では、AIの理論の発展に特化しており、計測で用いたセンサの特徴や、データに存在するノイズ、データ計測の困難さなどが考慮されないことが多い。実世界への応用に必要なのは、**データ計測から解析まで一気通貫で行うリアルワールドデータの解析**の知見である。本コラムでは、筆者が主導で実施した取り組みを基に、リアルワールドデータ解析の難しさを説明する。

筆者は、最新のウェアラブル機器等を用いたデータ計測と、そのデータ計測に参加していた方々のアンケートデータを用いた、「主観データと客観データの統合データ解析」を実施した。その解析の時系列は図1（右上）に示した通りである。実際にデータ解析を実施するには、データ計測するための実験参加者の募集や、参加者への研究説明、計測機器の設定、そしてデータ計測をしなければならない。データ解析を開始するまでに、4ヶ月もの期間を必要とした。このような下準備があって、はじめてリアルワールドデータの解析が実施可能となる。**長期間のデータ計測の実施の必要性**が、リアルワールドデータを扱う上での難しさと言える。

## 図1 実施したリアルワールドデータの解析手順

(図の引用：<https://humanome.jp/female-specific-cancer-report/> , 2021年9月18日時点)



データ解析で最も重要なのは、**良質なデータ**を取得することである。図1の工程3の計測期間におけるデータ取得方法が、良質なデータを取得するために重要である。次ページでは、このデータ取得の際の重要な点について、筆者が実施した取り組みを例に説明する。また、良質なデータを取得しても、どのような解析を実施すればよいかは、データの傾向を把握しないと判断できない。リアルワールドデータでは、**我々が予期していないデータが取得されている可能性がある**ため、事前知識のみでデータ解析するのは危険である。よって、次ページ以降では、データ解析結果だけでなくデータ自体の傾向についても記述する。

# リアルワールドデータ取得

(コラム) AIによるリアルワールドデータやPHRとの統合解析

## リアルワールドデータ取得の事例：女性がんサバイバーのQOL向上応援プロジェクト

### 女性がんサバイバーの方々のデータ取得事例

筆者が中心となり、女性特有のがんに直面した方々にご協力いただき、デジタルデバイスを活用した日常生活を検証するプロジェクトを実施した。本プロジェクトの実施概要は表1のとおりである。実際にごんに罹患された方々を対象に、3ヶ月もの期間データ計測を実施した。本ページでは、そのデータ計測での工夫点について中心に説明する。

表1 女性がんサバイバーの方々のデータ計測の実施概要

調査対象者	女性特有のがんに直面した女性100名	
	上記参加者の現在の治療内容 経過観察中：36%、ホルモン治療中：37% 抗がん剤等の治療中：27%	
対象者属性	がん腫	乳がん：68%、卵巣がん：14% 子宮体・頸がん：13%、その他：5%
	年齢	40代：50%、50代：34% 30代：11%、60代：5%
調査項目	Fitbit Inspire HRによる活動量・睡眠調査 毎日のアンケート、SNSへの投稿状況調査	
調査実施期間	2020年8月10日～2020年11月9日	

データ解析を実施して意味のある結果を導出する為には、多くの良質なデータを取得しなければならない。本プロジェクトでは、良質なデータを取得する為に、あらゆる工夫をした。

表1の調査実施期間から分かるように、本プロジェクトはコロナ禍で実施された。本来であれば、プロジェクト参加者と対面でプロジェクトの説明会を実施するが、今回は全てオンラインで完結した。参加者の中には、電子機器に慣れていない方もいる為、プロジェクト説明の難易度は普段に比べて非常に高い状態であった。その為、説明会前に参加者へのオンライン会議の仕様の周知や、説明会当日の電話対応を徹底した。そのおかげもあり、大きな問題もなくデータ計測を進めることができた。

肝心のデータ取得の部分であるが、まず、活動量と睡眠に関しては、Fitbit社から提供されるAPIを利用したシステムを構築し、遠隔でデータ取得をした。この方法では、プロジェクト参加者のデータ取得状況を逐一確認でき、データ計測できていない参加者を特定することができる。そして、機器自体を取り付けていなかったり、機器トラブル等の理由でデータ計測ができていない人に対しては、個別に対応することで、データ計測を継続することができた。

また、今回オンラインによる「毎日のアンケート」を実施した。本アンケートは、「主観的睡眠」、「運動内訳」、「体調」、「食事評価」を調査実施期間中に、スマートフォンやPCで毎朝記入していただいた。本アンケートの記入に関する所要時間は1分ほどであるが、毎日の記入は失念されることが多い。そこで、毎朝のアンケート記入願いメールの送付と、朝にアンケートの記入が終了していない人へ確認メールを昼過ぎに送信することで、プロジェクト終了まで回答率を維持することに成功した。

コロナ禍におけるデータ計測において、**100名中97名が本プロジェクトを完遂**し、さらに、上述の**アンケート回収率は平均93%を達成**した。これも、上述のような工夫と、本プロジェクトに協力していただいた参加者の御理解がなければ達成できない数字である。

# データの傾向確認

## リアルワールドデータを解析前の取得データの大まかな特徴を把握

### がんサバイバーの方々のデータ特徴の把握

データ計測の工夫により、がんサバイバーの方々の良質なデータを取得することに成功した。そのデータを用いてデータ解析をするために、取得されたデータがどのような傾向かを把握する必要がある。本ページでは、そのデータ傾向の一例を示す。

本プロジェクトの参加者は、非常に健康への意識が高い人が多いというのが分かった。その理由として、図2、3を見て分かるように、歩数や運動量が一般女性の平均を大きく上回っていたことが挙げられる。やはり、がんサバイバーの方々を対象としたこともあり、良好な健康を維持する意識の高さが伺える。

図2（右上）を見ると分かるように、本プロジェクト参加者の1日の平均歩数は9000歩を超えていた。これは、国民健康・栄養調査で示された女性の1日の平均歩数と比べると、非常に大きい数値であることが分かる。日頃より意識して歩かなければ、この数値は達成できない。

図3（右下）では、1週間の運動時間について示した。図3（a）から分かるように、がんサバイバーガイドラインで推奨されている運動時間150分/週に対して、本プロジェクト参加者の平均運動時間236分/週であった。ガイドラインの推奨時間と比べて、80分以上も本プロジェクト参加者は運動時間が多いという結果であった。そして、図3（b）の運動時間に関する本プロジェクトの参加者の割合を見ると、60%以上の方々が、がんサバイバーガイドラインの推奨時間を上回る運動時間であった。

さらに、プロジェクト参加者の中で、1日の平均運動時間を治療中と経過観察中の2群に分けて比較した。その結果、両者に差はほとんど見られず、経過観察中でも継続的に運動を実施していた。

上述から、本プロジェクトの参加者は非常に健康への意識が高く、活動的な方々が多かった。このように意識の高い方々によって、本プロジェクトの完遂率が非常に高くなったと言ってもよい。

図2 女性がんサバイバーの方々のデータ計測参加者の平均歩数

(図の引用：<https://humanome.jp/female-specific-cancer-report/>，2021年9月18日時点)

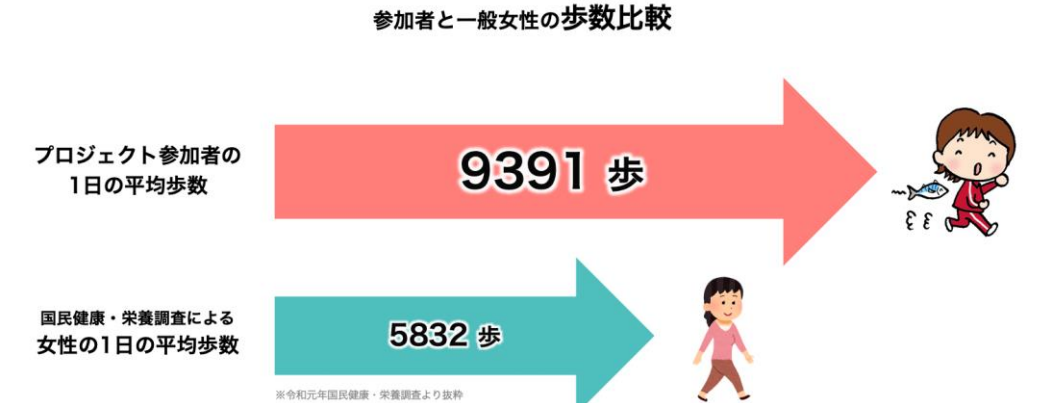
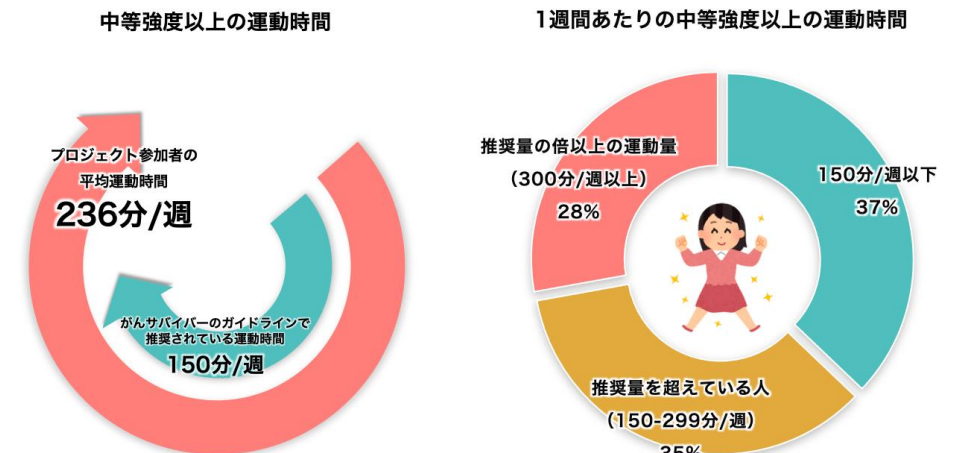


図3 女性がんサバイバーの方々のデータ計測参加者の平均運動量

(図の引用：<https://humanome.jp/female-specific-cancer-report/>，2021年9月18日時点)





# データの統合解析

(コラム) AIによるリアルワールドデータやPHRとの統合解析

## 主観データと客観データの統合解析結果

### 毎日のアンケートデータと活動量計のデータとの関係性について

前述のように、筆者が実施したプロジェクトでは、プロジェクト参加者にFitbitを装着していただいた他に、毎日のアンケートを記載していただいた。それらのデータを統合解析した結果を本ページでは示す。

今回の解析結果であるが、本記事のタイトルにもあるように（簡易ではあるが）AIを用いた。AIは、機械学習や数理統計手法を用いて、データから規則性や判断基準を学習し、それに基づいて目的の出力を導出する手法である。このとき、計算機が適切に規則や判断基準を学習できるようなデータの作成は、AIを構築する者の腕の見せ所である。以下では、入力データ（= Fitbitから得られた客観データ）と出力データ（= プロジェクト参加者の主観評価）を適切に前処理を実施し、それらが相関関係にあるかを確認する。そして、線形回帰を用いて入力データから、主観評価の推定の可能性について示す。

毎日のアンケートでは、主観的睡眠を毎日評価していただいた。この睡眠の評価と、Fitbitで取得した運動時間データとの関連性を調べた結果が図4（右上）である。人によって、運動を多くした日には熟睡度が良く/悪くなる人がいることが分かった。この結果から、どのくらいの運動量にすれば、良い睡眠状態を維持することが可能なのかを導ける可能性がある。

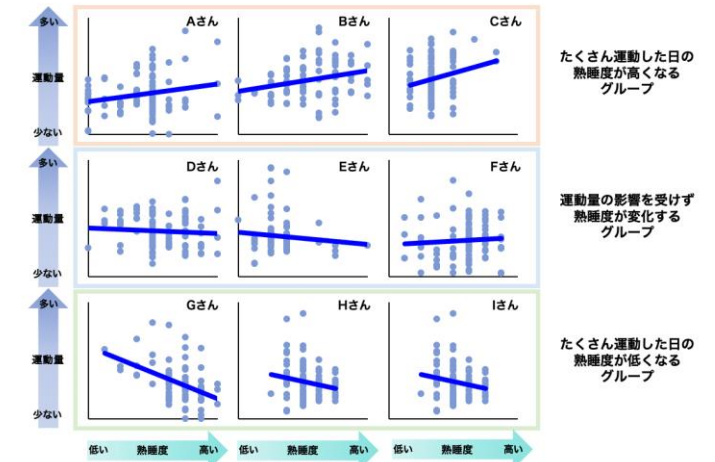
図5（右下）には、毎日のアンケートの体調の評価と、Fitbitで取得した心拍データの関連性を示した。図中のグラフは、横軸が10分以上安静時の心拍データで、右にいくほど心拍数が大きいことを示している。縦軸は、体調の評価となっており、上に行くほど体調が良いことを示している。図から、**体調が良くなると心拍数が小さくなる参加者群（グループ1）**と、**体調が良くなると心拍数が大きくなる参加者群（グループ2）**という、2つのグループに分割されることが分かった。しかし、なぜこのようなグループに分けられるのかの確証は得られていない。さらなる解析により、原因を解明していきたい。

上述のように、主観データと客観データをAIによる統合解析することにより、個人差が把握しやすくなった。この解析結果を基に、個人差に着目した健康増進の取り組みに繋がられる可能性は大いにある。前述のように**良質なデータを取得しなければ**、このような結果を得られないことができないが、**人間が健康を維持するには必要な労力**だと思われる。筆者は、積極的に今回紹介したものと同様のプロジェクトを実施し、将来の人類の健康増進に寄与するために挑戦していきたい。

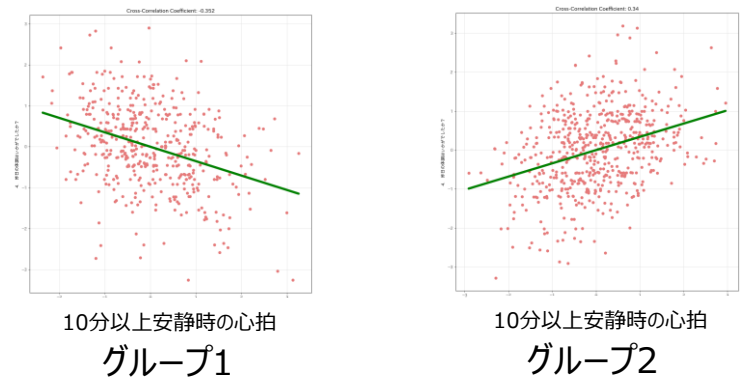
## 図4 運動量と熟睡度の関連性

（図の引用：<https://humanome.jp/female-specific-cancer-report/> , 2021年9月18日時点）

運動量と熟睡度の関連性



## 図5 体調の評価と心拍データの関係性



# コラム

# AIによるリアルワールドデータや PHRとの統合解析

東大医科研ヒトゲノム解析センター

**武藤香織**

# ELSI (エルシー) とは？

**ELSI**とは、新規科学技術を研究開発し、社会実装する際に生じる技術的課題以外のあらゆる課題のことを指す (大阪大学ELSIセンター)

- **Ethical, Legal and Social Implications/Issues**の頭文字をとったもの
- 医学・医療分野の文脈では、萌芽的な技術の研究開発段階から臨床試験、医療としての普及の過程で生じる、倫理、法、そして幅広く社会全体（心理、経済、産業、報道・ソーシャルメディア、コミュニケーション等）の諸問題を指している
- **由来は、米国で1990年に始まった「ヒトゲノム解析計画」**
  - ✓ DNAの二重らせん構造の発見者の一人であり、同計画の責任者でもあったJ. ワトソンが、研究の進展による新たな「倫理的・社会的課題 (ethical and societal implications)」への懸念を表明
  - ✓ 「全研究予算の3%程度をこれらの領域に投じるべきである。このことを通じて、20世紀前半に優生学の名の下に、不完全な遺伝学的知識の恐るべき誤用があったことを常に意識すべき」と明言
  - ✓ 1989年には国立衛生研究所 (National Institute of Health, NIH) とエネルギー省が、ELSIワーキンググループを共同で設置した。その委員長には、かつて優生学が淘汰の対象としていた遺伝性神経難病であるハンチントン病の当事者でもあるナンシー・ウェクスラー氏を迎え、今日も続く研究事業として“ELSI program”が誕生
- **日本では、ELSIは多義的に使用され、意味が混乱している**
  - ✓ ①一つの研究分野の総称、② ELSIを研究するプロジェクトの固有名詞、③生命科学の研究プロジェクトの外部助言委員会の固有名詞、④既に実装済みの技術の課題群、⑤関連法令・倫理指針、倫理審査の手続きなどの規制枠組み、などの用法がみられる
- **国の科学技術政策の根幹に位置付けられるようになった**
  - ✓ 第5期科学技術基本計画 (2016～20年度) では「倫理的・法制度的・社会的課題」として、第6期科学技術・イノベーション基本計画 (2021～25年度) では、「ELSI」として取り上げられている
  - ✓ 2015年 健康・医療戦略推進本部ゲノム医療実現推進協議会より、ゲノム医療分野のELSIについて本格的な検討が開始された
- **欧州では、ELSA (Aspects) として始まり、現在は“Responsible Research and Innovation (責任ある研究&イノベーション)” という概念に発展**

- がんの全ゲノム解析は、研究と臨床を架橋する観点から、倫理的配慮が不可欠

- 倫理的な原則・指針の歴史

- ✓ 1997年 ユネスコ「ヒトゲノムと人権に関する宣言」  
ヒトゲノムは人類の遺産
- ✓ 2000年 ユネスコ「ヒト遺伝情報に関する国際宣言」  
研究結果について知られるか否かを定める権利
- ✓ 2000年 科学技術庁「ヒトゲノム研究に関する基本原則」  
文部科学省・厚生労働省・経済産業省「ヒトゲノム・遺伝子解析研究に関する倫理指針」（2001-2021年）の骨格

- 現在、遵守すべき指針「人を対象とする生命科学・医学系研究に関する倫理指針」（2021年）のポイント

- ✓ インフォームド・コンセント（電磁的方法も可）で研究結果の説明方針を決定、理解を得る
- ✓ インフォームド・コンセントを取得する機関共同研究に係る研究計画は、原則として、一つの倫理審査委員会による一括した審査
- ✓ 研究実施の最終責任者は機関の長から研究責任者に変更。研究開始前に、各機関の長から研究の実施許可を得る必要あり（フライング実施は倫理指針違反!）
- ✓ 研究結果の公表時には、対象集団への配慮を

- 研究結果の説明方針について

- ✓ がんの全ゲノム解析において、主たる結果や所見のみならず、当該研究実施に伴って二次的に得られた結果や所見として、遺伝性腫瘍やその他の遺伝性疾患のリスクが判明する可能性がある
- ✓ 倫理指針の規定では、以下の通り
  - 健康状態等を評価するための情報としての精度や確実性、研究対象者の健康等にとって重要な事実の有無などを考慮した方針決定が必要
  - 研究対象者が16歳以上の未成年者の場合で代諾者に説明する際は、研究対象者の意向確認と尊重が必要
- ✓ AMED小杉班より、詳細な提言が公開されている  
<https://www.amed.go.jp/news/seika/kenkyu/20200121.html>
  - 「ゲノム医療における情報伝達プロセスに関する提言 -その1：がん遺伝子パネル検査を中心に（改定第2版）」
  - 「がん遺伝子パネル検査二次的所見患者開示 推奨度別リスト（Ver2.0\_20191210）」
  - 「ゲノム医療における情報伝達プロセスに関する提言—その2：次世代シーケンサーを用いた生殖細胞系網羅的遺伝学的検査における具体的方針（改定版）」

- 個人情報保護、遺伝的特徴に基づく差別、守秘義務などが課題

- 用語の整理

(厚労省「ゲノム医療等の実現・発展のための具体的方策について（意見とりまとめ）」（2016）、個人情報保護法）

- ✓ 「ゲノムデータ」は、塩基配列を文字列で表記したもの  
→ 法律上、「個人識別符号」に該当。個人情報としての保護が必要であり（利用目的を特定した同意取得、変更前の利用目的との関連性を有すると合理的に認められる範囲を超えた変更はできない）
- ✓ 「ゲノム情報」は、塩基配列に解釈を加え意味を有するもの
- ✓ 「遺伝情報」は、ゲノム情報の中で子孫へ受け継がれるもの  
→ 法律上、「要配慮個人情報」に該当。一段高い保護の対象（本人の同意がない取得や第三者への提供を原則禁止）

- 日本では、遺伝的特徴・遺伝情報に基づく差別禁止を法定化していない

- ✓ 保険：2018年より生命保険協会がガイドラインを検討中とされる
- ✓ 雇用：「雇用管理分野における個人情報のうち健康情報を取り扱うに当たっての留意事項（2017年）」にて、遺伝性疾患に関する情報については、職業上の特別な必要性がある場合を除き、事業者は、労働者から取得すべきではない、とされている

- 患者本人の意に反する血縁者に対するリスクの説明：守秘義務と説明義務

- ✓ 診療の場合、患者の検査結果について、本人の意思に反して、医療従事者が家族に伝えることは、守秘義務違反にあたる可能性がある
  - がん遺伝子パネル検査では、病状の悪化その他の事態に備え、あらかじめ守秘義務を解除してもらうため、検査の同意取得時に、検査結果の共有を希望する相手と連絡先を尋ねる方針（空欄でも受検は可能）
- ✓ HIVや精神科疾患での他者危害リスク等は例外として定着したが、ゲノム医療での議論は未確定
  - 英国では、ハンチントン病患者である父親の意思に従い、医師が血縁者へ遺伝的リスクを伝達されずに妊娠・出産した女性による医師の説明義務違反を問う裁判があった。2020年に原告敗訴が確定したが、「他者に深刻な害を及ぼす場合に、他者への説明を考慮する義務」が指摘された（ABC v St Georges NHS Trust）
- ✓ 研究の場合、研究者が法的説明義務を負っていないこと、臨床検査より不確実な結果となる可能性等から、より慎重に考慮する必要あり
  - 倫理指針では、研究対象者が当該研究により得られた結果等の説明を希望しない場合には、その意思を尊重。但し、研究対象者やその血縁者等の生命に重大な影響を与えることが判明し、且つ、有効な対処方法があるときは、研究責任者に報告が必要、とされる

## ● 患者・市民とともに、がんゲノム医療を発展させる社会へ

### ● 患者・市民参画の推進

- ✓ 「医学研究・臨床試験における患者・市民参画」とは、医学研究・臨床試験プロセスの一環として、研究者が患者・市民の知見を参考にすること (AMED)
- ✓ 健康・医療戦略推進本部 ゲノム医療実現推進協議会「中間とりまとめ」(2015年7月)
  - 研究計画に対して、研究者は研究対象者から評価を求められる機会が少なく、研究対象者は受動的な関与に留まっている。また、研究対象者が長期にわたって研究の趣旨に賛同し、研究協力の意思を継続できているかどうか把握できていない  
→ **電磁的インフォームド・コンセントで研究者や企業と長期的な信頼関係を維持**
- ✓ がん対策推進基本計画 (第3期) (2017年)
  - AMEDは、海外の研究体制と同様、我が国でも患者やがん経験者が研究のデザインや評価に参画できる体制を構築するため、平成30 (2018) 年度より、患者及びがん経験者の参画によって、がん研究を推進するための取組を開始する。また、国は、研究の計画立案と評価に参画可能な患者を教育するためのプログラムの策定を開始する  
→ **研究者のパートナーとなる患者・市民人材の育成**

### ✓ 「がんゲノム医療推進コンソーシアム懇談会」報告書 (2017年)

- ここに構築するがんゲノム医療の基盤は国民共有の財産である。広く国民の理解及び賛同なくしては成り立たず、「がんの撲滅」も現実のものとはならない。本懇談会では、がんゲノム医療の基盤整備は、国民が主体的に参加し、その恩恵も国民が享受すべきものであることを銘記して、検討を進める  
→ **「全ゲノム解析等実行計画」における患者・市民パネルの創設**

### ● がん全ゲノム解析をもっと社会に開かれた存在に (2021年3月調査より)

- ✓ 「全ゲノムを解析することで何が分かり、どのように治療が変わっていかうとしているのかを知りたいです」
- ✓ 「ゲノム解析が進むことで人類はどのような医療を受けられるようになるのか、具体的なイメージがほしいです」
- ✓ 「例えばこの機械に入れただけでゲノムを解読できるものだと一般の方は思っているし、それをいつも診てくれる医師が簡単に結果を出してくれると言ったイメージを持っていることが多いです。現実にはゲノム解析になくはない人がいます。その人たちの存在が余りにも世に出ていない為その人材がものすごく不足していると思います。このゲノム解析に関わる多くの人バイオインフォマティクスのこととかバイオインフォマティシヤンのことをもう少し知ってもらう機会があるといいのではないかと思います。意外に医療者がこだわらないところに一般の人はこだわります。私などもこの人たちの存在を知ったとき、こんな難しいことを解析する仕事をする人がいるんだととても感動しました」

# INDEX

単語	参照頁
1000genome	<a href="#">118</a>
2ヒット仮説	<a href="#">12</a>
ACMG (American College of Medical Genetics/ 米国臨床遺伝・ゲノム学会)	<a href="#">100</a> 、 <a href="#">124</a> 、 <a href="#">125</a>
Actionable Mutation	<a href="#">32</a> 、 <a href="#">33</a>
AMED AGD	<a href="#">187</a>
AMP (Association for Molecular Pathology/ 米国分子病理学会)	<a href="#">127</a> 、 <a href="#">137</a>
ASCO (American Society of Clinical Oncology/ 米国臨床腫瘍学会)	<a href="#">127</a> 、 <a href="#">137</a>
ATC (Anatomical Therapeutic Chemical) code	<a href="#">195</a>
BAF (B-allele frequency)	<a href="#">81</a> 、 <a href="#">97</a>
BAMファイル (Binary Alignment Map)	<a href="#">91</a> 、 <a href="#">93</a>
Branch mutation	<a href="#">18</a>
CADD	<a href="#">115</a> 、 <a href="#">120</a>
CAP (College of American Pathologist/ 米国病理医協会)	<a href="#">127</a> 、 <a href="#">137</a>

単語	参照頁
Catastrophic event	<a href="#">85</a>
cBioPortal	<a href="#">108</a> 、 <a href="#">187</a>
CCAT (がんゲノム情報管理センター)	<a href="#">137</a>
CDx (コンパニオン診断)	<a href="#">39</a>
Central dogma (セントラルドグマ)	<a href="#">7</a>
CGH/CGHアレイ	<a href="#">87</a>
CGI (Cancer Genome Interpreter)	<a href="#">123</a> 、 <a href="#">140</a> 、 <a href="#">156</a> 、 <a href="#">174</a>
Chromoplexy	<a href="#">85</a>
Chromothripsis (クロモスリプシス)	<a href="#">15</a> 、 <a href="#">85</a> 、 <a href="#">91</a> 、 <a href="#">104</a>
CIN (Chromosomal instability/ 染色体不安定性)	<a href="#">14</a>
Circos plot	<a href="#">10</a> 、 <a href="#">107</a>
CIViC (Clinical Interpretations of Variants in Cancer)	<a href="#">108</a> 、 <a href="#">123</a> 、 <a href="#">147</a> 、 <a href="#">169</a>
CKB (Clinical Knowledgebase)	<a href="#">123</a> 、 <a href="#">156</a>

# INDEX

単語	参照頁	単語	参照頁
ClinicalTrial.gov	<a href="#">141</a>	Deletion (欠失)	<a href="#">93</a> 、 <a href="#">98</a> 、 <a href="#">116</a>
ClinVar	<a href="#">108</a> 、 <a href="#">122</a> 、 <a href="#">139</a> <a href="#">157</a>	depth	<a href="#">26</a> 、 <a href="#">30</a>
CMap (Connectivity Map)	<a href="#">196</a>	DIN値 (DNA Integrity Number)	<a href="#">52</a>
CNV (Copy Number Variant/Variation)	<a href="#">3</a> 、 <a href="#">9</a> 、 <a href="#">35</a> 、 <a href="#">93</a> 、	DNA	<a href="#">6</a>
COSMIC (Catalogue of Somatic Mutation of Cancer)	<a href="#">162</a>	DNA損傷修復機構	<a href="#">14</a>
CPAS法 (Combinational Probe Anchor Synthesis)	<a href="#">27</a>	DNB (DNAボール)	<a href="#">27</a>
DB (データベース)	<a href="#">114</a> 、 <a href="#">118</a> 、 <a href="#">120</a> 、 <a href="#">122</a>	Driver genes (ドライバー遺伝子)	<a href="#">17</a>
dbNSFP	<a href="#">120</a>	ELSA (Ethical、 Legal、 and Social Aspects)	<a href="#">203</a>
DBS (Doublet base substitutions/ 二塩基置き換え)	<a href="#">84</a>	ELSI (Ethical、 Legal、 and Social Implications/Issues)	<a href="#">203</a>
dbSNP	<a href="#">118</a>	EPWG (Expert Panel Working Group)	<a href="#">137</a>
dbVar	<a href="#">118</a>	ESCAT (ESMO Scale for Clinical Actionability of molecular Target)	<a href="#">137</a>
ddPCR (Droplet digital PCR)	<a href="#">88</a>	ESEfinder3.0	<a href="#">121</a>
Deep intronic region	<a href="#">36</a>	ESMO (欧州臨床腫瘍学会)	<a href="#">137</a>



# INDEX

単語	参照頁
Exome Variant Server	<a href="#">118</a>
FASTQ	<a href="#">73</a> 、 <a href="#">92</a>
FDA (Food and Drug Administration/ 米国食品医薬品局)	<a href="#">127</a> 、 <a href="#">142</a>
FFPE (Formalin Fixed Paraffin Embedded/ ホルマリン固定パラフィン包埋) 検体	<a href="#">44</a> 、 <a href="#">47</a> 、 <a href="#">50</a>
FISH	<a href="#">68</a> 、 <a href="#">87</a>
FKPM (Fragment per Kilobase of Exon per million reads mapped)	<a href="#">105</a>
Focal amplification (局所増幅)	<a href="#">97</a>
FoundationOne CDx	<a href="#">39</a>
GA4GH (Global Alliance for Genomics and Health)	<a href="#">182</a>
GEL (Genomics England)	<a href="#">108</a>
GEM Japan (GENome Medical Alliance Japan)	<a href="#">181</a> 、 <a href="#">182</a>
GeneSplicer	<a href="#">121</a>
GENIE	<a href="#">108</a>

単語	参照頁
gnomAD	<a href="#">118</a>
GREP (Genome for repositioning)	<a href="#">195</a>
GWAS (Genome Wide Association Study)	<a href="#">177</a> 、 <a href="#">194</a> 、 <a href="#">195</a> 、 <a href="#">196</a>
G分染	<a href="#">87</a>
HGMD (Human Gene Mutation Database)	<a href="#">122</a>
HGVD (Human Genetic Variation Database)	<a href="#">119</a> 、 <a href="#">176</a>
HLA型決定	<a href="#">106</a>
HMF (Hartwig Medical Foundation)	<a href="#">108</a>
Homozygous deletion	<a href="#">97</a>
HRD (Homologous Recombination Defect/ 相同組み換え修復欠損)	<a href="#">104</a> 、 <a href="#">126</a>
ICD (International Statistical Classification of Diseases and Related Health Problems) code	<a href="#">195</a>
ICGC (International Cancer Genome Consortium)	<a href="#">108</a> 、 <a href="#">187</a>
ICGC-ARGO	<a href="#">187</a>

# INDEX

単語	参照頁
ID (Small insertions and deletions/短鎖挿入欠失)	<a href="#">84</a>
IGV (Integrative Genomics Viewer)	<a href="#">107</a>
Indel	<a href="#">33</a> 、 <a href="#">68</a> 、 <a href="#">93</a> 、 <a href="#">96</a>
Index (インデックス)	<a href="#">70</a>
Insert	<a href="#">68</a>
Insertion (挿入)	<a href="#">98</a> 、 <a href="#">116</a>
JapicCTI (Japic Clinical Trials Information)	<a href="#">141</a>
JAX CKB	<a href="#">173</a>
JGA (Japanese Genotype-phenotype Archive)	<a href="#">181</a>
jMorp (Japanese Multi-Omics Reference Panel)	<a href="#">180</a>
jRCT (臨床研究実施計画・研究概要公開システム)	<a href="#">141</a>
Kataegis	<a href="#">85</a>
Locus Specific Mutation Databases	<a href="#">122</a>

単語	参照頁
LOH (Loss of heterozygous/ヘテロ接合性の喪失)	<a href="#">81</a> 、 <a href="#">97</a>
Mb	<a href="#">16</a>
MGeND (Medical Genomics Japan Variant Database/エムジェンド)	<a href="#">177</a>
MLPA法	<a href="#">87</a>
MSI ( Microsatellite instability/ マイクロサテライト不安定性)	<a href="#">14</a> 、 <a href="#">126</a>
MutationTaster	<a href="#">120</a>
NBDCヒトデータベース	<a href="#">176</a>
NCI Genomic Data Commons	<a href="#">187</a>
NetGene2	<a href="#">121</a>
NGS (次世代シーケンス)	<a href="#">20</a> 、 <a href="#">21</a> 、 <a href="#">23</a>
NNSplice	<a href="#">121</a>
Nonsynonymous Mutation (非同義変異)	<a href="#">8</a>
OncoKB	<a href="#">123</a> 、 <a href="#">139</a> 、 <a href="#">171</a>

# INDEX

単語	参照頁
PCR (Polymerase Chain Reaction)	<a href="#">24</a> 、 <a href="#">53</a> 、 <a href="#">68</a> 、 <a href="#">88</a>
Personalized medicine (個別化医療)	<a href="#">19</a>
Phasing (ハプロタイプの決定)	<a href="#">75</a>
PolyPhen-2	<a href="#">120</a>
Precision Medicine (精密医療)	<a href="#">19</a>
PROVEAN	<a href="#">120</a>
PRS (Polygenic risk score)	<a href="#">100</a> 、 <a href="#">194</a>
Pseudo-exon	<a href="#">36</a>
Pseudogene (偽遺伝子)	<a href="#">75</a>
Purity	<a href="#">92</a> 、 <a href="#">97</a>
Read	<a href="#">70</a> 、 <a href="#">77</a> 、 <a href="#">92</a>
SBS (Single base substitutions/ 1塩基置き換え)	<a href="#">82</a>
SBS法 (Sequencing by Synthesis)	<a href="#">23</a> 、 <a href="#">25</a>

単語	参照頁
Short read (ショートリード)	<a href="#">92</a>
SIFT	<a href="#">120</a>
SMRTテクノロジー (Single Molecule、 Real-Timeシーケンシングテクノロジー)	<a href="#">28</a>
SNV	<a href="#">3</a> 、 <a href="#">68</a> 、 <a href="#">93</a>
Soft clipping	<a href="#">98</a>
SPIDEX	<a href="#">121</a>
Strand bias	<a href="#">84</a>
SV (Structural Variant・Variation/構造異常)	<a href="#">3</a> 、 <a href="#">9</a> 、 <a href="#">68</a> 、 <a href="#">75</a>
Synonymous Mutation (同義変異)	<a href="#">8</a>
TARGET (Therapeutically Applicable Research to Generate Effective Treatments)	<a href="#">108</a>
TAT (Turn around time)	<a href="#">55</a>
TCGA (The Cancer Genome Atlas)	<a href="#">108</a>
Throughput (決定塩基数)	<a href="#">30</a>

# INDEX

単語	参照頁
TogoVar	<a href="#">119</a> 、 <a href="#">181</a>
ToMMo (東北大学東北メディカル・メガバンク機構)	<a href="#">180</a> 、 <a href="#">194</a>
Translocation (転座)	<a href="#">98</a>
Trans-Phar (Integration of Transcriptome-wide association study and Pharmacological database)	<a href="#">196</a>
Trunk mutation	<a href="#">18</a>
Tumor heterogeneity (腫瘍内不均一性)	<a href="#">18</a>
Tumor mutation burden (TMB)	<a href="#">14</a> 、 <a href="#">16</a> 、 <a href="#">102</a> 、 <a href="#">126</a>
TWAS (Transcriptome-Wide Association Study)	<a href="#">196</a>
UMIN-CTR (UMIN臨床試験登録システム)	<a href="#">141</a>
VarSome	<a href="#">122</a> 、 <a href="#">156</a> 、 <a href="#">167</a>
VCFファイル (Variant Call Format)	<a href="#">114</a> 、 <a href="#">117</a>
WES (Whole Exome Sequencing/ 全エクソン・エキソームシーケンス)	<a href="#">30</a> 、 <a href="#">31</a> 、 <a href="#">68</a>
WGS (Whole Genome Sequencing/全ゲノム解析)	<a href="#">30</a> 、 <a href="#">68</a> 、 <a href="#">88</a> 、 <a href="#">90</a> 、 <a href="#">92</a>

単語	参照頁
アデニン (A)	<a href="#">6</a>
アノテーション	<a href="#">111</a> 、 <a href="#">114</a> 、 <a href="#">130</a> 、 <a href="#">131</a> 、 <a href="#">139</a>
アポトーシス	<a href="#">15</a>
アライメント/マッピング	<a href="#">92</a>
アレル	<a href="#">12</a>
アレル不均衡	<a href="#">81</a>
イオン半導体シーケンシング法	<a href="#">23</a>
遺伝子	<a href="#">6</a>
インデックスホッピング	<a href="#">27</a>
イントロン	<a href="#">7</a>
エキスパートパネル	<a href="#">130</a> 、 <a href="#">137</a>
エクソン/エキソン	<a href="#">7</a>
核型分析	<a href="#">87</a>

# INDEX

単語	参照頁
片親性ダイソミー	<a href="#">81</a>
がん遺伝子	<a href="#">11</a>
がんゲノム医療	<a href="#">19</a> 、 <a href="#">130</a>
がんゲノム学	<a href="#">185</a>
がん細胞含有率	<a href="#">46</a>
がんプロファイル解析	<a href="#">32</a>
がん抑制遺伝子	<a href="#">12</a>
機能獲得型変異	<a href="#">17</a>
機能損失型変異	<a href="#">17</a>
キュレーション	<a href="#">108</a> 、 <a href="#">111</a>
グアニン (G)	<a href="#">6</a>
クオリティスコア	<a href="#">73</a>
クラスター生成	<a href="#">72</a>

単語	参照頁
ゲノム	<a href="#">6</a>
サンガーシーケンス (ダイターミネーター法)	<a href="#">22</a> 、 <a href="#">26</a>
シーケンスライブラリ	<a href="#">58</a>
シトシン (C)	<a href="#">6</a>
腫瘍Purity	<a href="#">97</a>
腫瘍検体 (T)	<a href="#">112</a>
新鮮凍結 (Fresh Frozen/FF) 検体	<a href="#">44</a> 、 <a href="#">46</a> 、 <a href="#">47</a> 、 <a href="#">52</a>
スプライシング	<a href="#">7</a>
ターゲットシーケンス	<a href="#">90</a>
多型	<a href="#">3</a>
チミン (T)	<a href="#">6</a>
デジタルバイオマーカー	<a href="#">193</a>
デュアルインデックス	<a href="#">71</a>

# INDEX

単語	参照頁
テロメラーゼ	<a href="#">86</a>
転写	<a href="#">7</a>
点変異	<a href="#">8</a>
ドミナントネガティブ効果	<a href="#">12</a>
ドライバー遺伝子バリエーション	<a href="#">112</a>
ドライバー変異	<a href="#">17</a> 、 <a href="#">33</a>
ドラッグ・リポジショニング	<a href="#">194</a>
ナノポア (Nanopore) シーケンス	<a href="#">29</a> 、 <a href="#">74</a>
ナンセンス変異	<a href="#">8</a> 、 <a href="#">17</a> 、 <a href="#">116</a>
ヌクレオチド	<a href="#">28</a>
ネオアンチゲン	<a href="#">16</a>
バーコード	<a href="#">58</a>
バイオマーカー	<a href="#">37</a>

単語	参照頁
パイプライン	<a href="#">93</a>
パイロシーケンス法	<a href="#">23</a>
パッセンジャーバリエーション	<a href="#">112</a>
パッセンジャー変異	<a href="#">17</a>
ハプロ不全	<a href="#">12</a>
非腫瘍検体 (N)	<a href="#">112</a>
フレームシフト変異	<a href="#">8</a> 、 <a href="#">116</a>
フローセル	<a href="#">25</a> 、 <a href="#">60</a> 、 <a href="#">62</a>
分子バーコード	<a href="#">70</a>
分子標的薬	<a href="#">11</a> 、 <a href="#">38</a>
ペアエンド	<a href="#">70</a>
変異	<a href="#">3</a>
変異Signature	<a href="#">82</a>

# INDEX

単語	参照頁
ホッピング	<a href="#">71</a>
ミスセンス変異	<a href="#">8</a>
ミスマッチ修復機構	<a href="#">14</a>
免疫チェックポイント	<a href="#">16</a>
ユニークデュアルインデックス	<a href="#">71</a>
リアルワールドデータ	<a href="#">193</a>
リキッドバイオプシー	<a href="#">190</a>
リファレンス	<a href="#">26</a>
レーザーマイクロダイセクション法 (LMD)	<a href="#">46</a>

\*単語が記載されている全頁を記しているわけではありません。

# 監修者

本テキストの監修者は下記の通りとなります。

監修者	所属	監修章
井元清哉	東京大学医科学研究所 ヒトゲノム解析センター 健康医療インテリジェンス分野 教授	第5章、コラム②
加藤護	国立がん研究センター 基盤的臨床開発研究コアセンター オミックスコア バイオインフォマティクス部門 部門長	第4章、コラム③
山口類	愛知県がんセンター研究所 システム解析学分野 分野長	第7章、コラム③
南谷泰仁	東京大学医科学研究所 附属病院 血液腫瘍内科 教授	第6章、コラム①、コラム④
松田浩一	東京大学大学院 新領域創成科学研究科 メディカル情報生命専攻 メディカルサイエンス講座 教授	第1章、コラム④
鈴木健介	イルミナ株式会社	第2章、第3章、コラム①

本テキストに掲載する著作物の複製権、上映権、譲渡権、公衆送信権（送信可能化権を含む）は厚生労働省が保有します。本テキストを無断で複製する行為（コピー、スキャン、印刷など）は、著作権法上で限られた例外（「私的使用のための複製」など）を除き禁じられています。