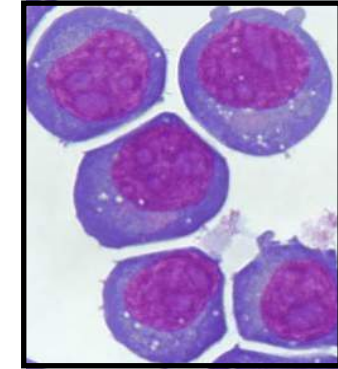
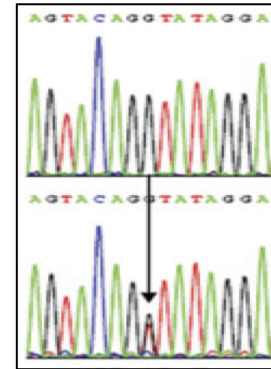
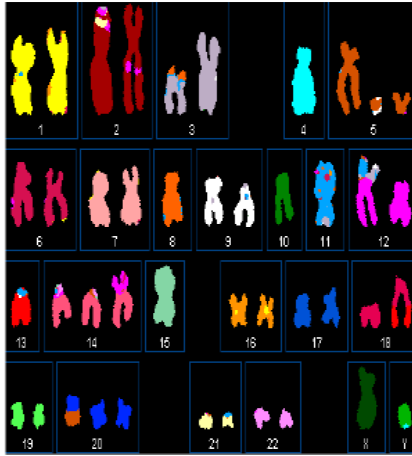




# イルミナHiSeq2000による がんのゲノムシーケンス解析

理化学研究所 ゲノム医科学研究センター(CGMM)

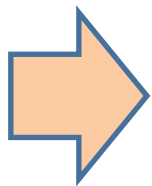
バイオマーカー探索・開発チーム チームリーダー  
中川 英刀 (なかがわ ひでわき)



がんは、遺伝的因子に加えて、様々な環境的因子(発癌物質、感染、炎症など)の作用によって、細胞のゲノムに変異が蓄積し、その結果、正常な分子経路が破綻して無秩序な細胞増殖、さらには浸潤、転移する

## Cancer = “Disease of Genome”

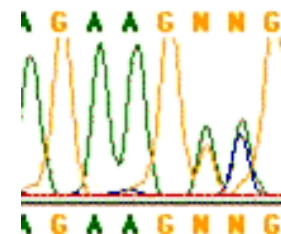
がんの本質であるゲノム異常をすべて解明する



- ✓分子レベルでのがんの病態解明
- ✓新規のがんの治療法、診断法、予防法の開発

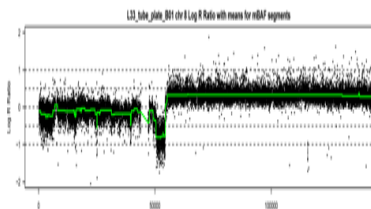
# がんのゲノム異常

✓ **ポイント変異** アミノ酸変化をともなう

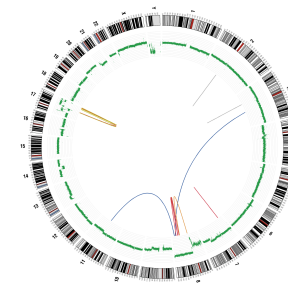
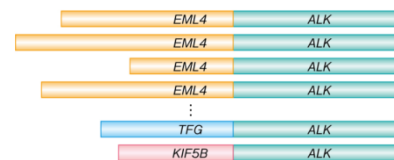


✓ **Short indel (insertion & deletion)**

✓ **コピー数異常** 増幅(Gain) 欠失(Loss)



✓ **構造異常** 染色体異常 転座、逆位など  
>> 融合遺伝子



■ RNA editingやDNA メチル化 (エピゲノム異常)

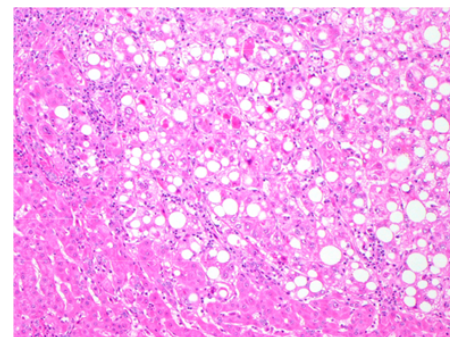
# がんの体細胞変異>> 分子標的・バイオマーカー

変異遺伝子	変異の種類	がんの種類	分子標的阻害剤
<i>BCR-ABL</i> 融合	構造異常(転座)	慢性骨髄性白血病	グリベック (2001)
<i>HER2</i>	遺伝子増幅	乳がん、胃がん	ハーセプチン(1998)
<i>EGFR</i>	遺伝子変異・増幅	肺がん、大腸がん	イレッサ、ターセバ アービタックス
<i>c-KIT</i>	遺伝子変異	消化管間質腫瘍	グリベック
<i>KRAS</i>	遺伝子変異	大腸がん、膵がん	EGFR阻害薬の バイオマーカー
<i>BRAF</i>	遺伝子変異	メラノーマ、大腸がん	2011 診断薬とともに FDA認可
<i>PI3KCA</i>	遺伝子変異	肺がん 大腸がん	治験中
<i>FLT3</i>	遺伝子変異	急性骨髄性白血病	治験中
<i>MET</i>	遺伝子変異・増幅	肺がん、メラノーマ	治験中
<i>BRCA1/2</i>	遺伝子変異	乳がん、卵巣がん	PARP1阻害剤の バイオマーカー
<i>EML4-ALK</i> 融合	構造異常(逆位)	肺がん	2011 診断薬とともに FDA認可





# 臨牀的、病理学的に質の高いサンプルの確保



既往歴  
タバコ、飲酒などの暴露  
家族歴  
腫瘍マーカーなどの検査データ  
治療方法  
転帰

ヒトやがんの“多様性”を考えた場合、  
多数のサンプルを収集し解析する必要性がある

# がんゲノムの多様性

がんのゲノムは、

- ◆ がんの種類によって異なる
- ◆ 発がんの原因によって異なる
- ◆ 人種や民族によって異なる
- ◆ 個々のがん(患者さん)によって異なる

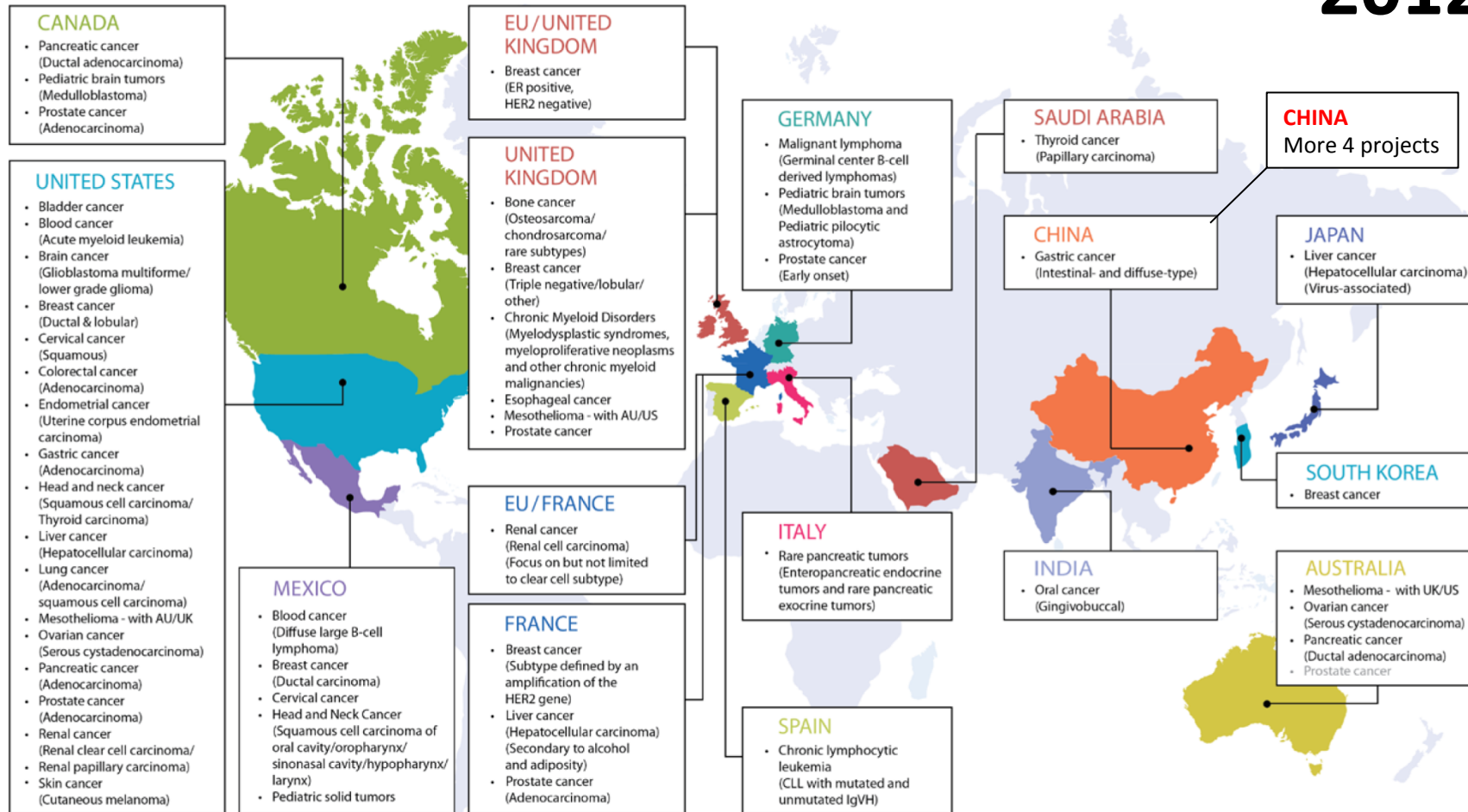


# 国際がんゲノムコンソーシアム

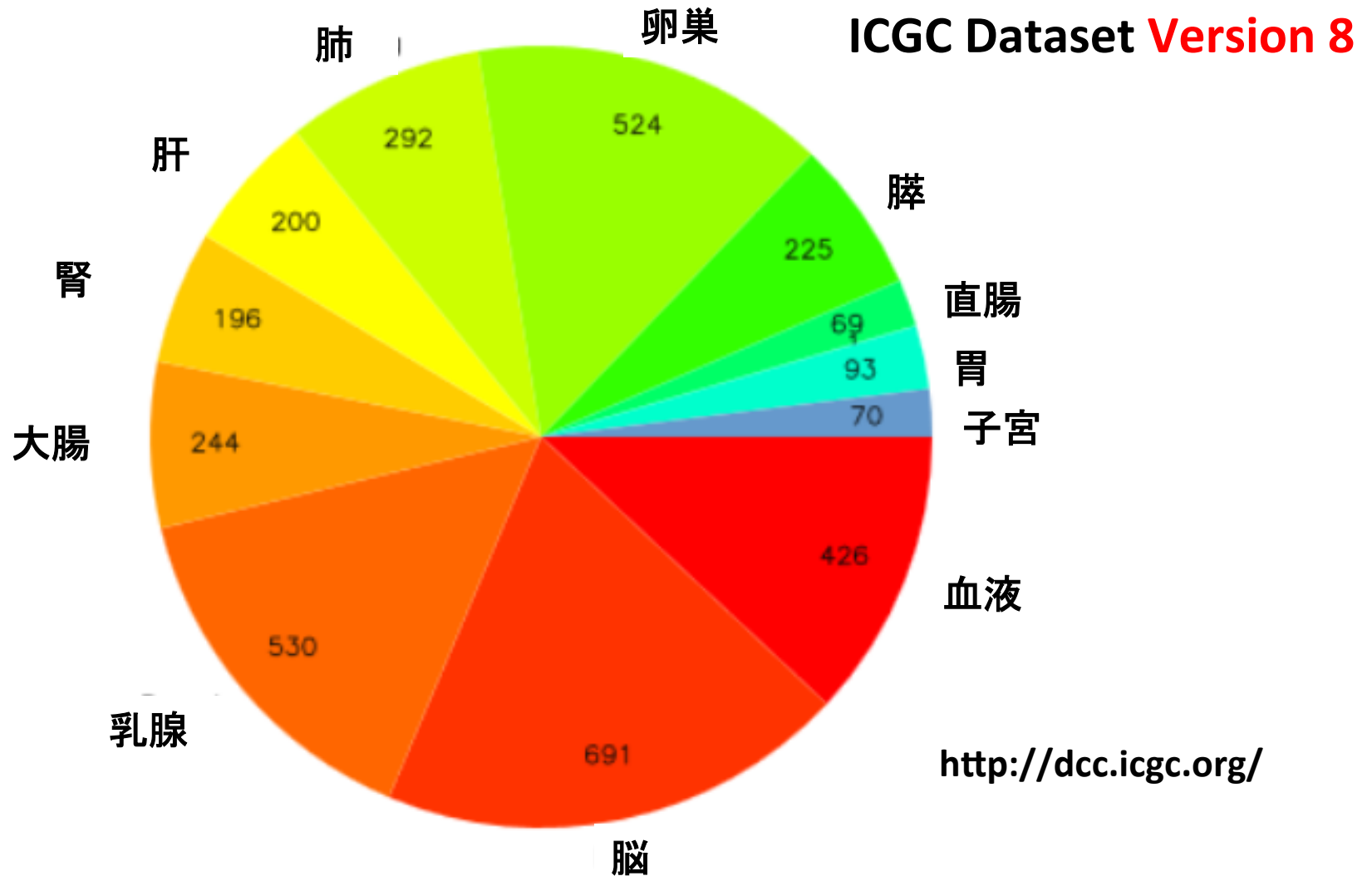
## ICGC

### 47 projects launched

### 2012



2012年3月：合計**3561**例のがんのデータが登録・公開



**500x47 = 23,500 個**のがんゲノムデータベースの構築へ

# 今後のがんゲノム解析研究の戦略について

シーケンスコストが急激に低下し  
WGSが数百ドルの時代へ

- ① 多様性のあるがんゲノムに対して、大量のがんサンプルを解析する
- ② 特殊で稀ながんを少数解析して、がんのある一面を明らかにする



今後、WGSへ倫理的に対応でき、病理学的・臨床的情報を伴った質の高いサンプルを収集して、凍結保管しておく必要がある

# 倫理面でのがんゲノム研究のポイント

- ✓ 究極の個人情報扱うので、**厳重なセキュリティが必要**
- ✓ WGSに対応したinformed consentの取得
- ✓ 胚細胞変異 (germline) もあつかうため、**incidental**な変異への対応を考慮



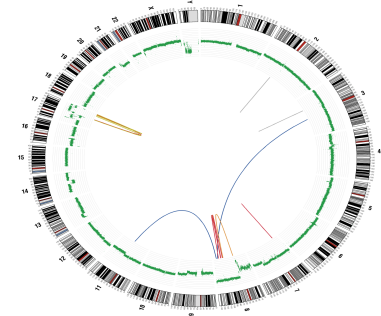
例えば、肝がんの全ゲノムシーケンスによって、BRCA1の germline 変異が同定された

どうすべきか？



# がんゲノムシーケンスの方法戦略

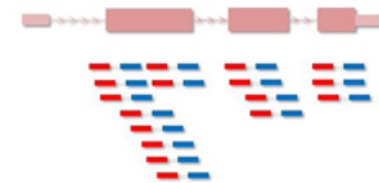
◆ Whole genome sequencing (WGS): 3Gb



◆ Exome sequencing

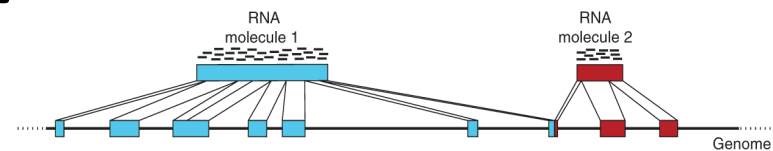
About 1-2% of whole genome: 30-40Mb

Exon capture technology >>NGS



◆ Gene expression (RNA) analysis

RNA-Seq, miRNA-Seq

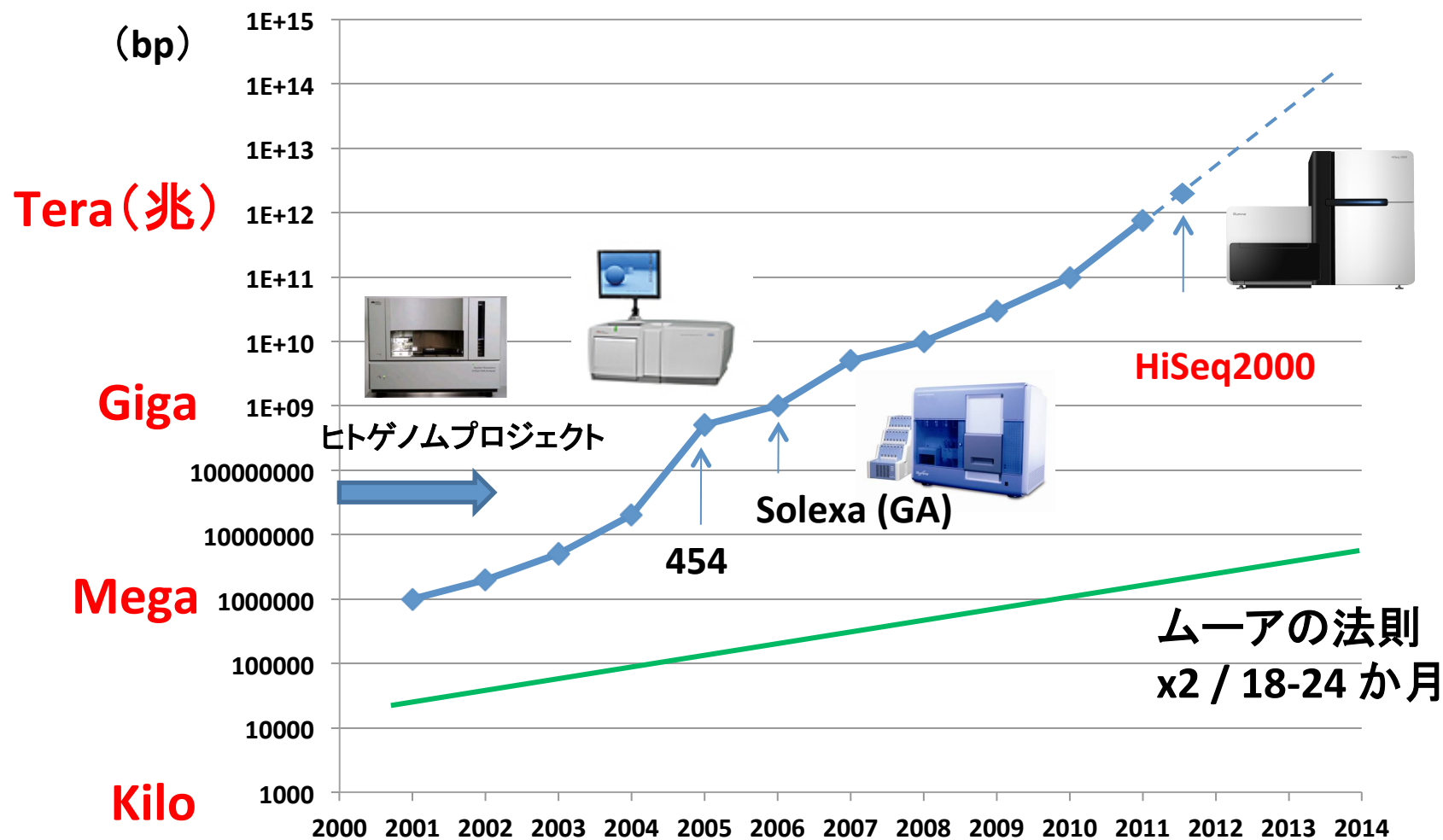


◆ Epigenetic analysis

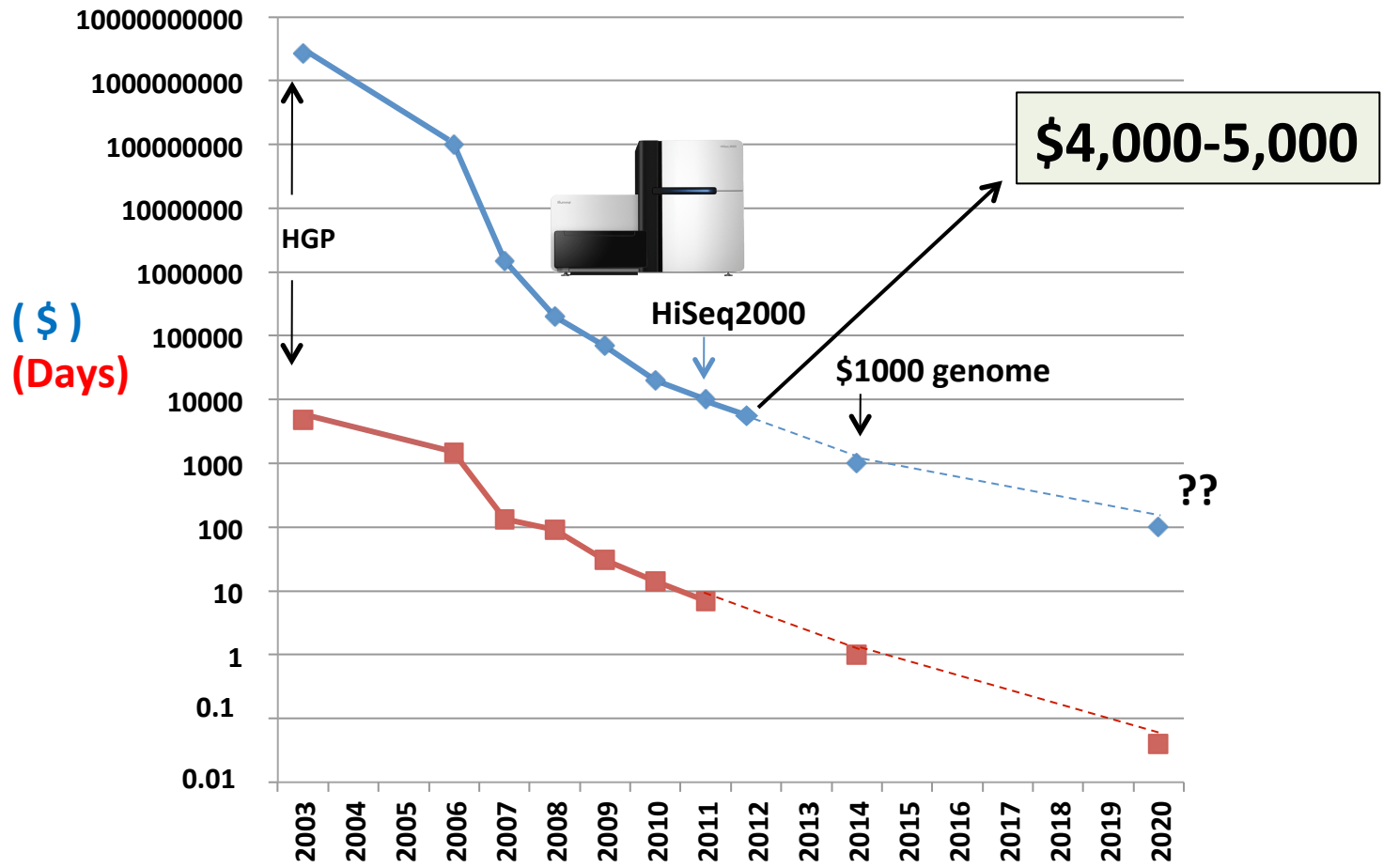
bisulfite treatment, MeDIP>> NGS

# 爆発的なDNAシーケンス技術の進歩

1台のシーケンサーで1か月間に読める塩基数



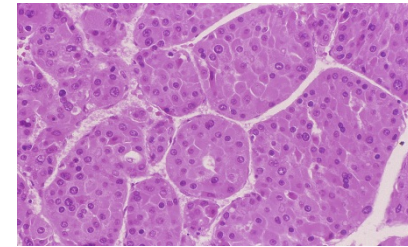
# ヒト全ゲノムシーケンスにかかるコストの推移



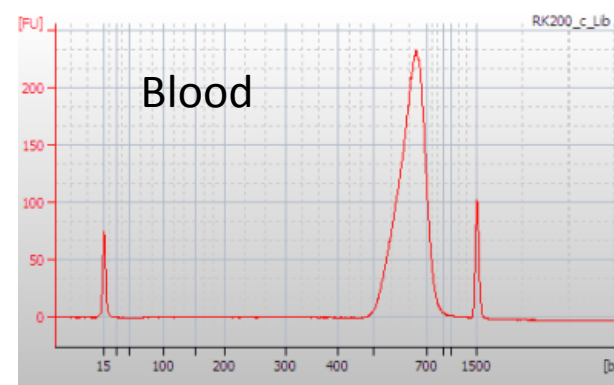
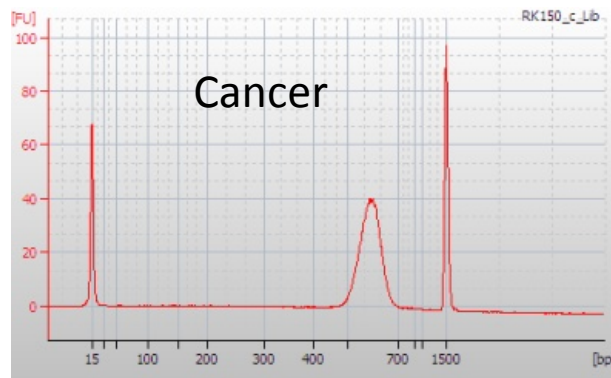
**1,000ドルゲノムの時代へ**

さらに  
100\$へ

- 血液
- がん組織の病理評価  
active cancer cellが80%以上を確認



1. gDNA の抽出
2. PicoGreen法とゲル染色によるgDNAの定量と評価
3. 1.0  $\mu\text{g}$ を用いて、PE ライブラリーの作製  
Insert size 500-600 bp



4. がん:4レーン、血液:3レーン分を、HiSeq2000にて解析

全ゲノムシーケンス

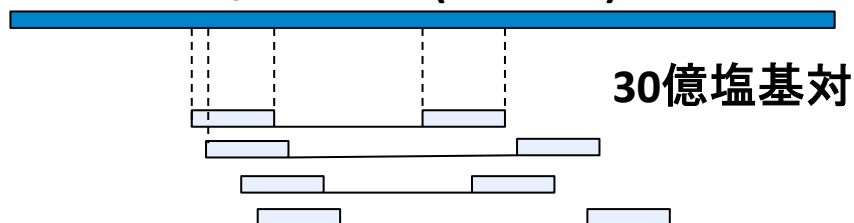
1,200億塩基 (x40, がん)  
900億塩基 (x30, 血液)



プロセッシング、マッピング

Read filtering, PCR duplicationの除去など

ヒトゲノム参照配列 (GRCh37)へマッピング



...GATTATTCCAGGTAT....

正常ゲノム

vs

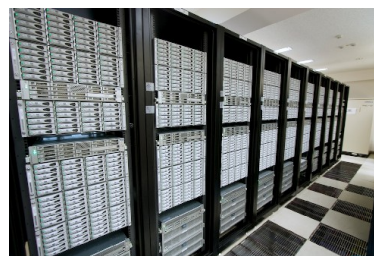
...GATTATTGCAGGTAT....

癌ゲノム

ポイント変異、コピー数異常、  
構造異常を検出

...GATTATT **G** CAGGTAT....

体細胞変異



高度な情報科学、遺伝学、数学、統計学的アプローチが必須

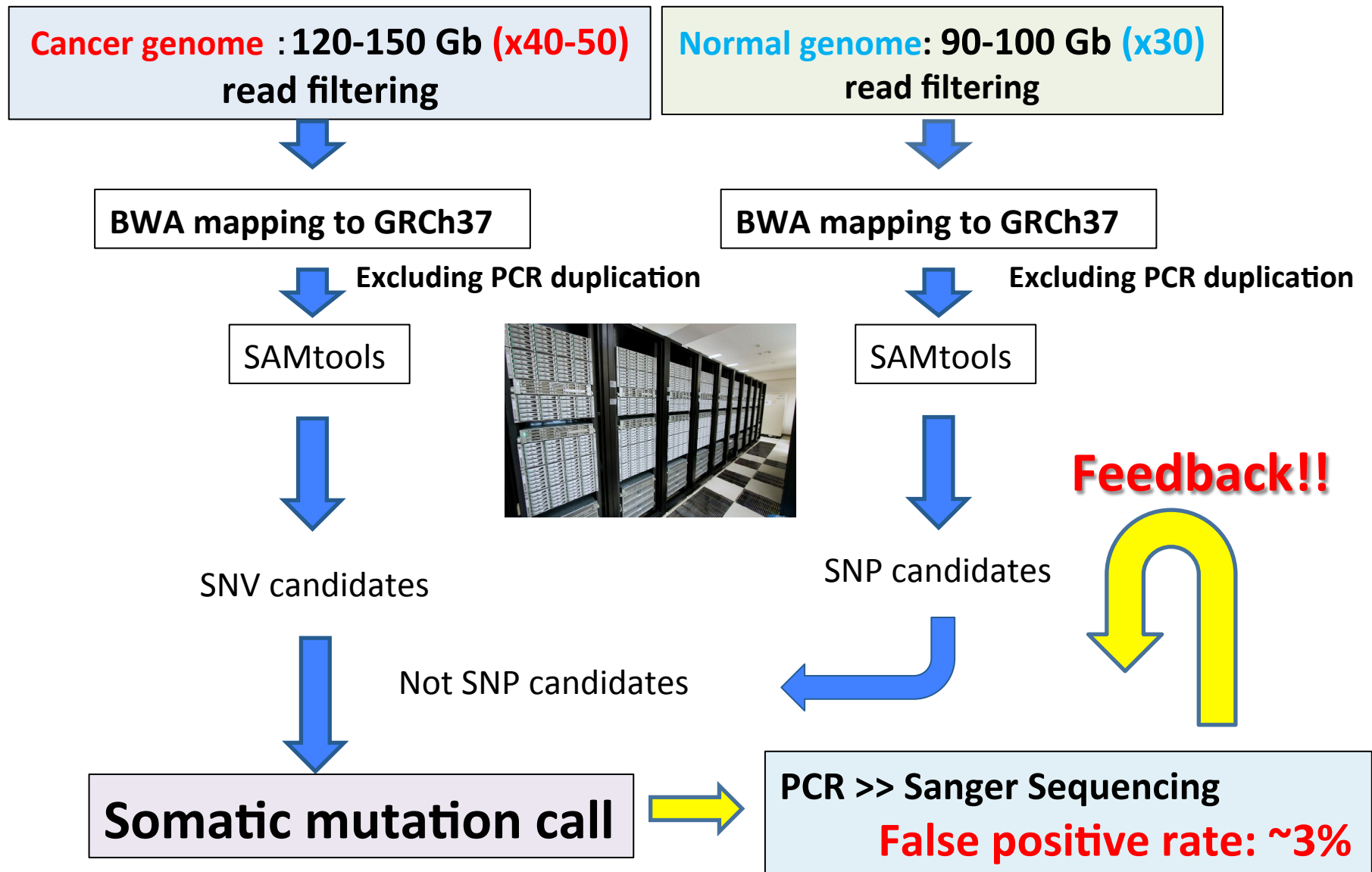
統合解析、生物学的解釈、分子標的や  
driverの同定など

- RNA-Seq
- メチル化DNA解析
- 臨床病理データ

ポイント遺伝子変異

**Somatic substitution and short indel  
in cancer genome**

# Call for somatic point mutations





# Call for somatic point mutations in cancer genome

**Cancer genome** : 120-150 Gb (x40-50)

**Normal genome**: 90-100 Gb (x30)

Mapping to reference genome (GRCh37)

Mapping to reference genome  
(GRCh37)

Read count in every nucleotide  
of human genome

Read count in every nucleotide  
of human genome

Chr2: 85,446,193

T,A,A,A,A,A,A,A,A,A,A,G,G,G,G,G,G,G

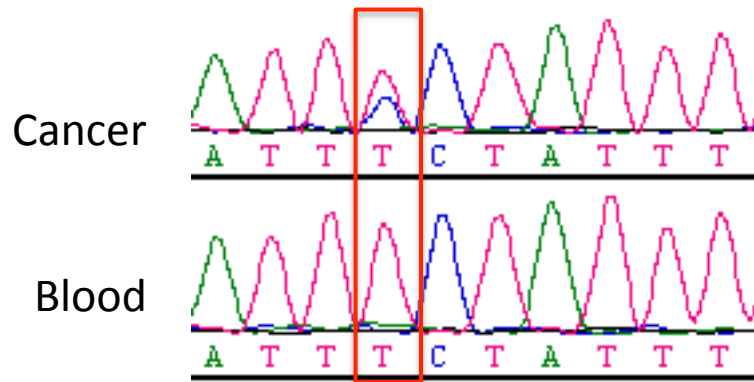
A,A,A,A,A,A,A,A,A,A,A,A,A,A,C

- Sequencing error (C/G > A/T)
- Mapping error
- Low depth region
- Reference genome
- Tumor heterogeneity

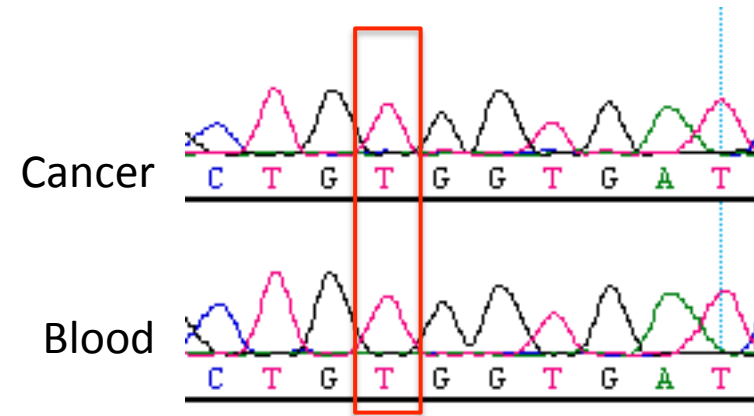
**Error of  
Mutation Call**

# Validations by Sanger sequencing

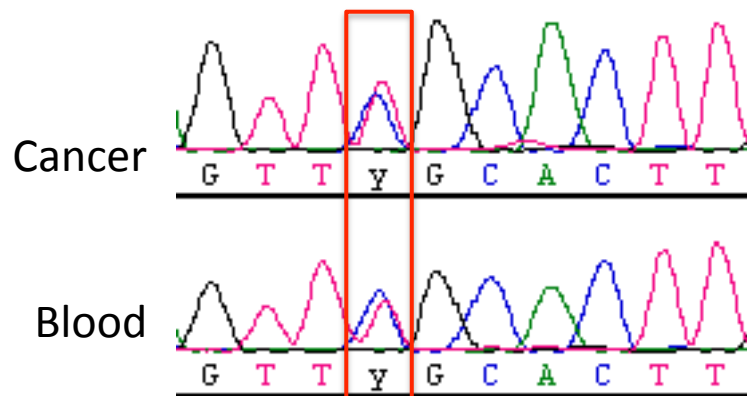
## Somatic substitution



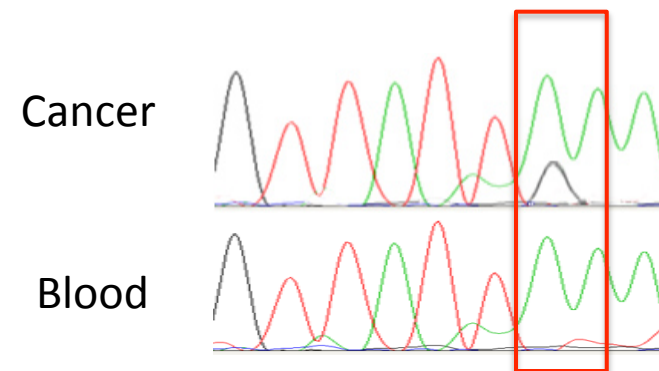
## False positive



## SNPs



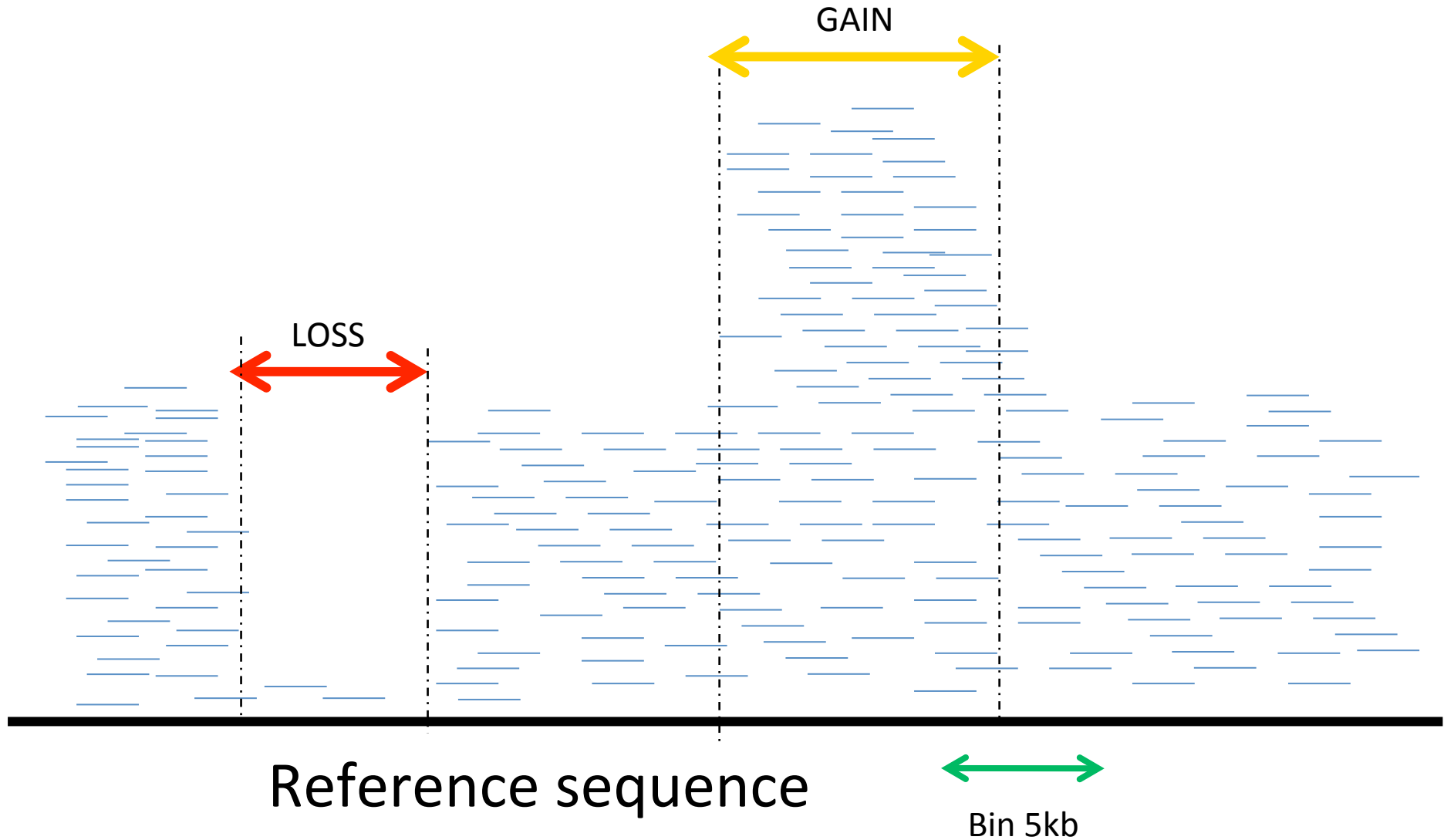
## Tumor heterogeneity, or noise



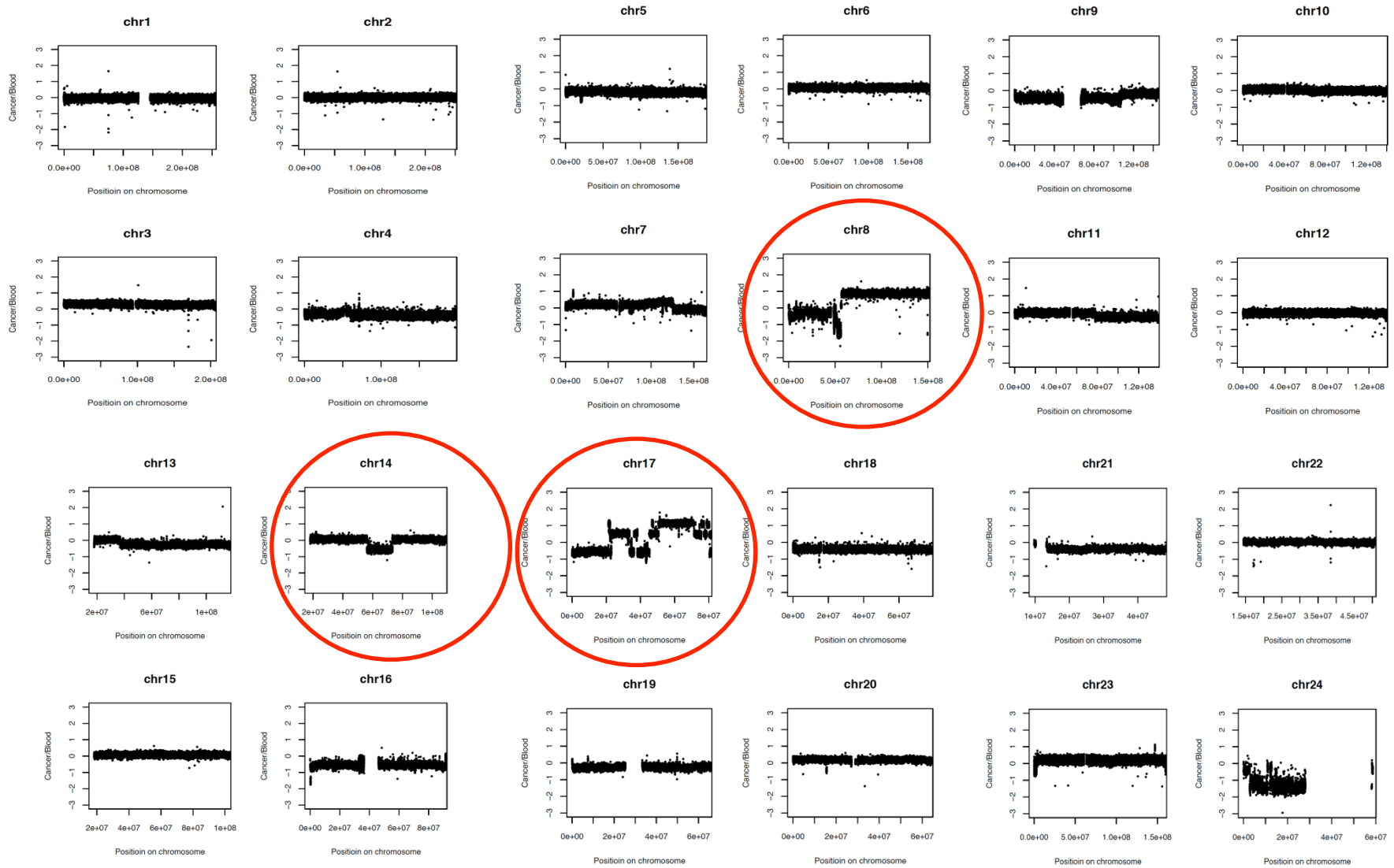
コピー数異常

**Somatic Copy number alterations  
in cancer genome**

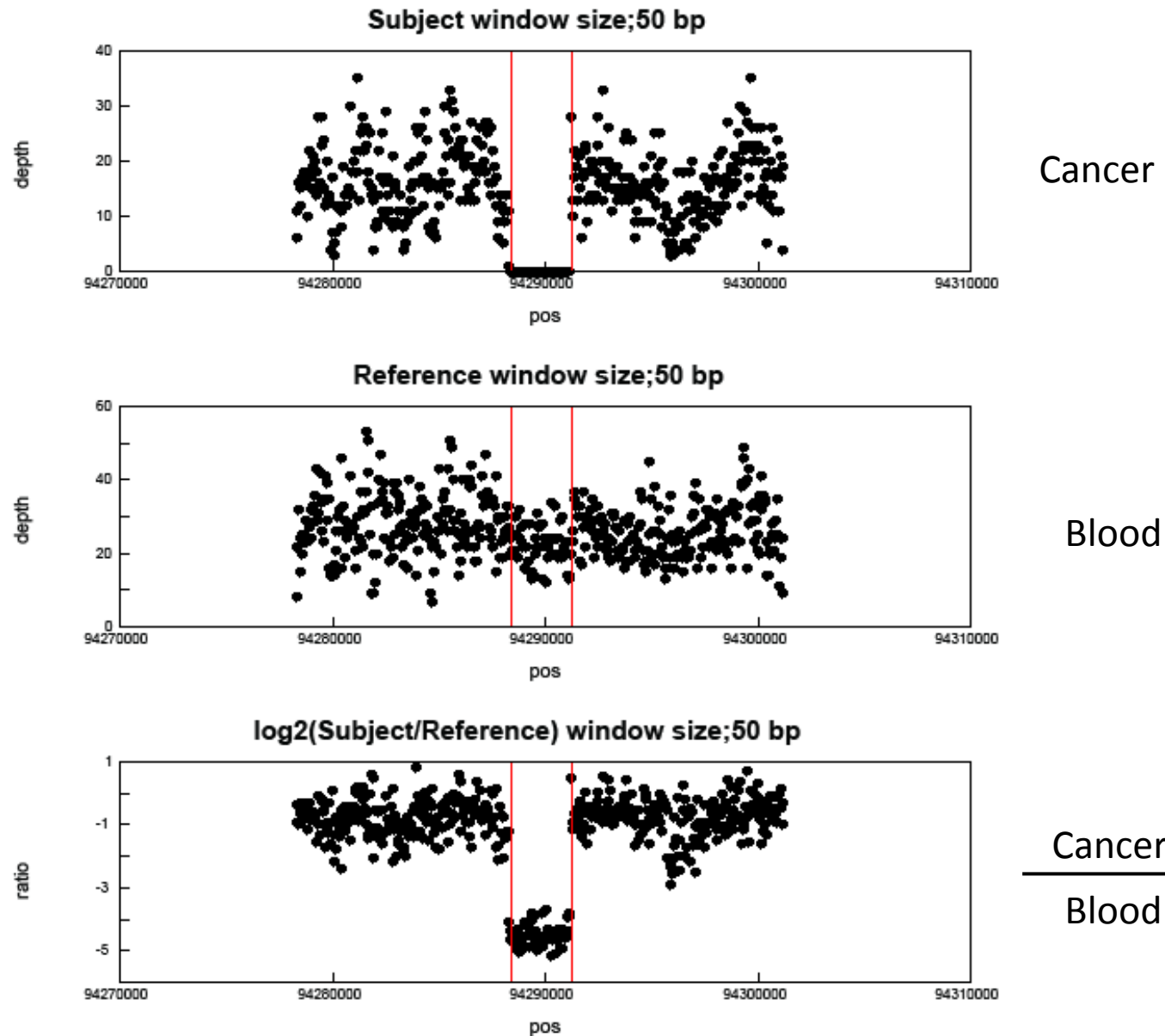
# Copy Number Alteration (CAN) call by NGS



# Copy number alternations in RK003 HCC



# Detection of middle-size deletion (50 bp resolution)



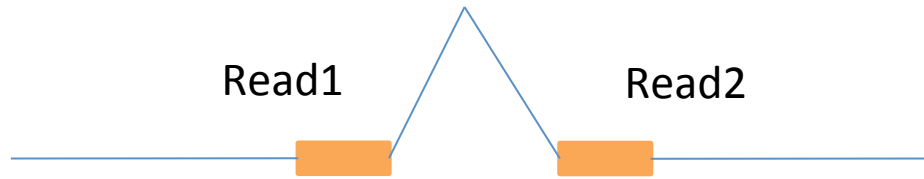
ゲノム構造異常

**Genomic rearrangements  
in cancer genome**

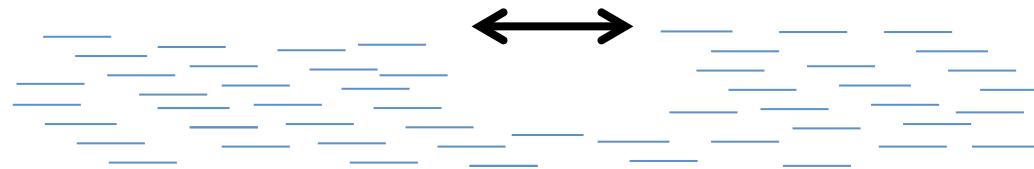


# Call for Structure Variants by NGS

Pair end



Copy number



Split read

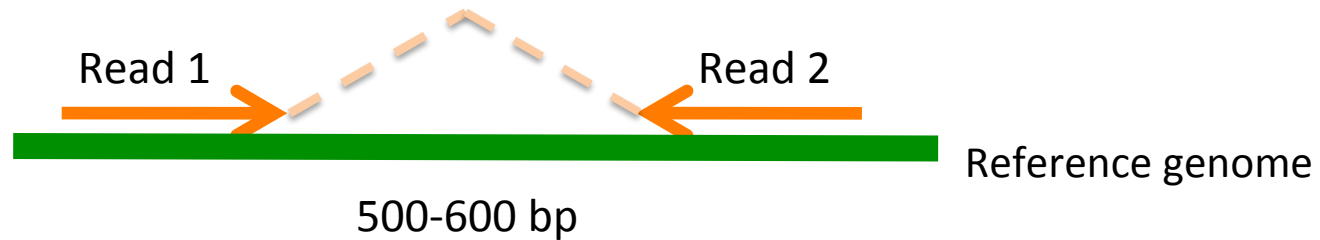


*de novo* assembly

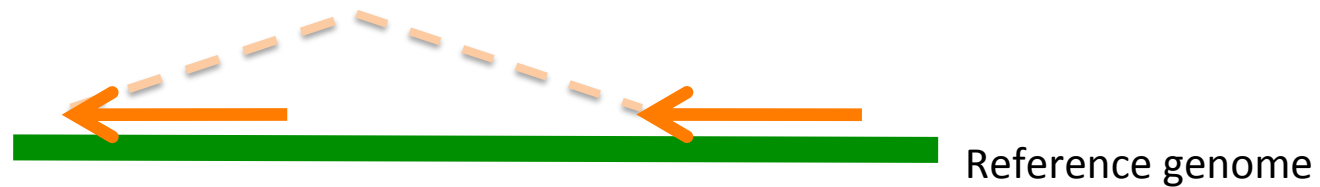


# Paired-end sequencing detects genomic rearrangements

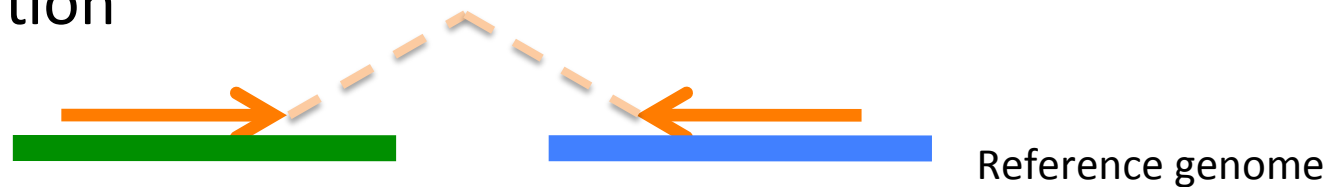
Normal pair



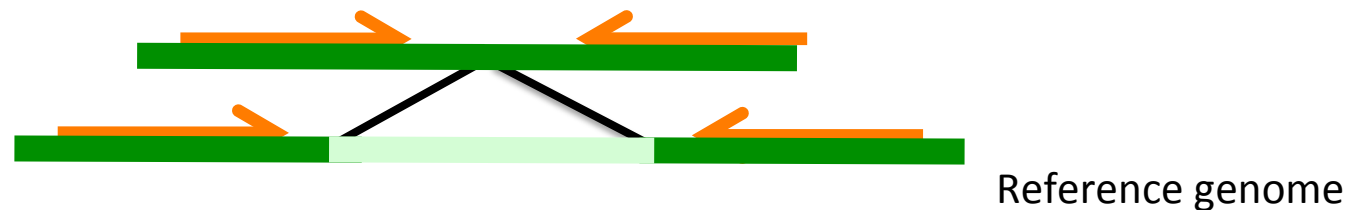
Inversion



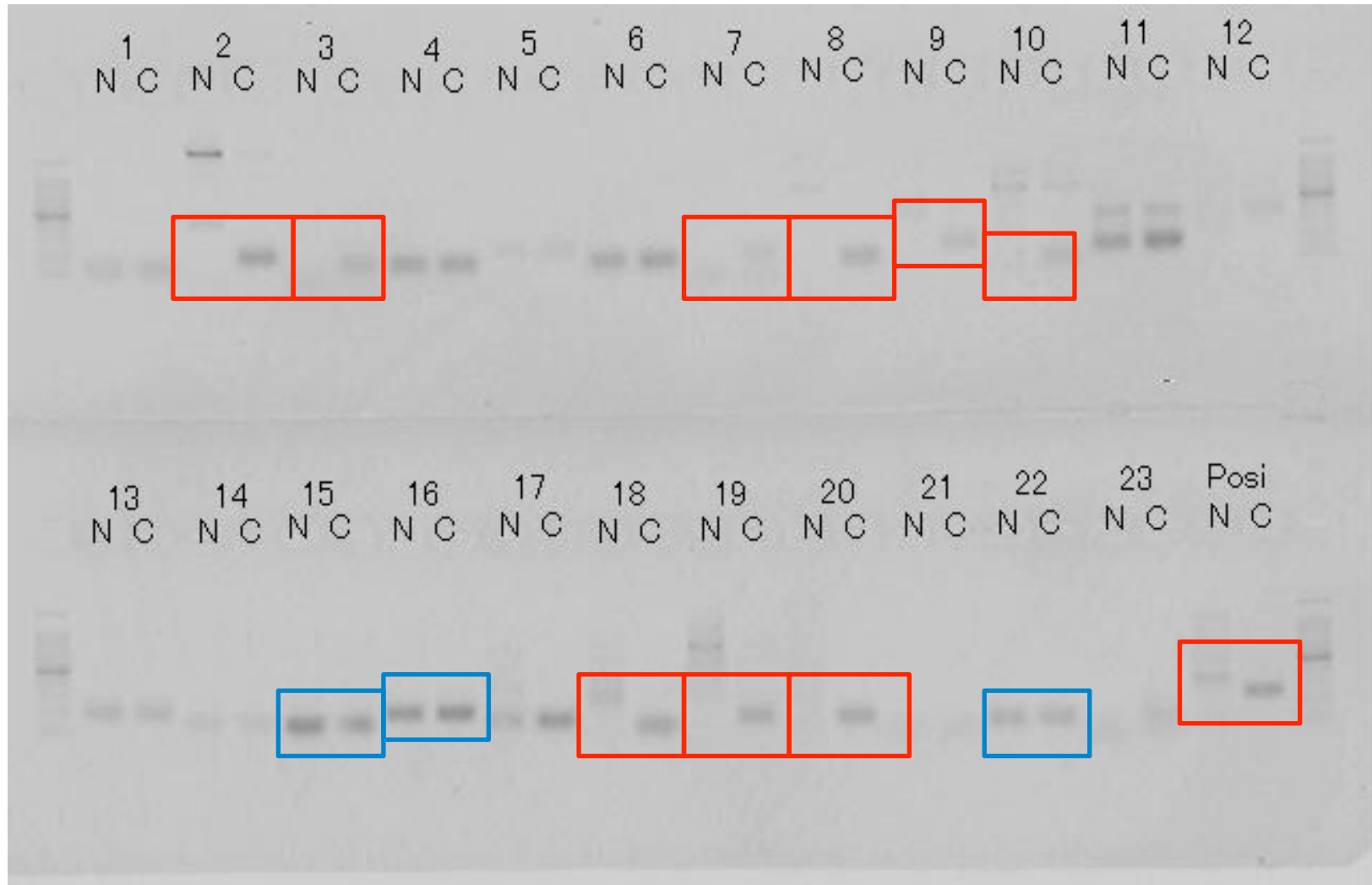
Translocation



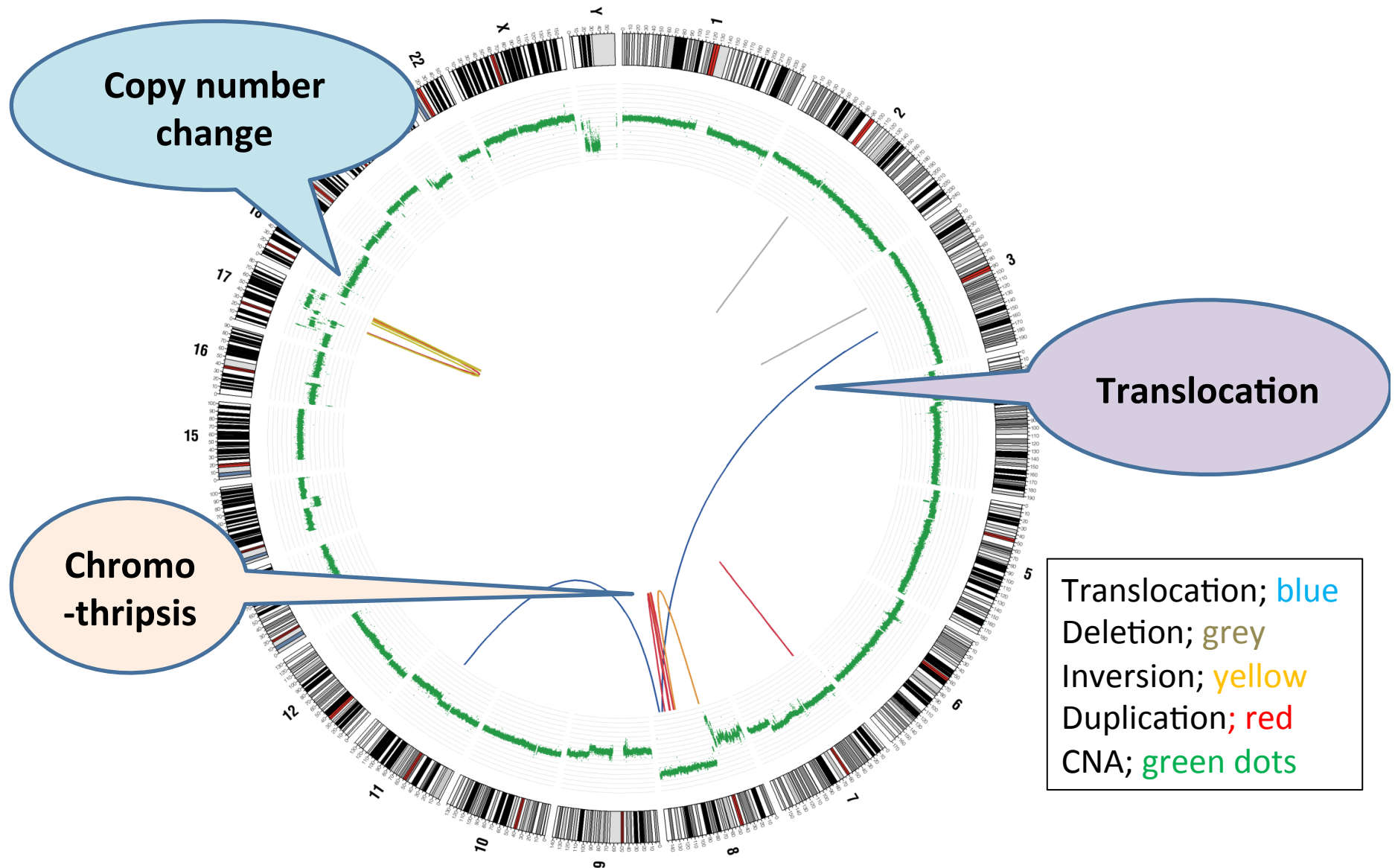
Deletion



# Validation of genomic rearrangements by PCR and Sanger sequencing



# Circos plot for genomic rearrangements and CNV of a HCV-associated HCC (RK003)




# がんSomatic Mutation検出の精度

## がんゲノム vs 正常ゲノム

- |   |             |      |       |
|---|-------------|------|-------|
| □ | SNV         | 偽陽性率 | 3%以下  |
| □ | Short indel | 偽陽性率 | 10%以下 |
| □ | 構造異常        | 偽陽性率 | 20%以下 |
- ✓ 基本的に、WGS解析の精度は、**depth** と**mappingの精度**に依存する
    - >> depthの低いゲノム領域、mapping errorが起きやすい領域 (repeat領域) は、間違いが多い
  - ✓ がんゲノムの場合、intra-tumor heterogeneityの問題もあり、low populationのmutationの有無も精度評価に影響する (偽陽性率、偽陰性率の計算)

# 肝がんゲノムプロジェクト

Liver Cancer

Japan 



International  
Cancer Genome  
Consortium

研究資金提供機関



理化学研究所



医薬基盤研究所

サンプル採取機関

和歌山県立医科大

広島大学

大阪府立成人病センター

東京女子医大

国立がん研究センター中央病院



理研ゲノム医科学研究センター

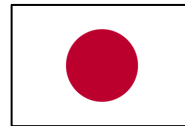
シーケンス配列解析

国立がん研究センター

コンピューターリソース



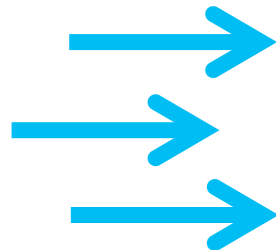
東大医科研 ヒトゲノム解析センター



東大先端科学技術研究センター

500例のゲノム情報、診療情報を登録

他国研究機関



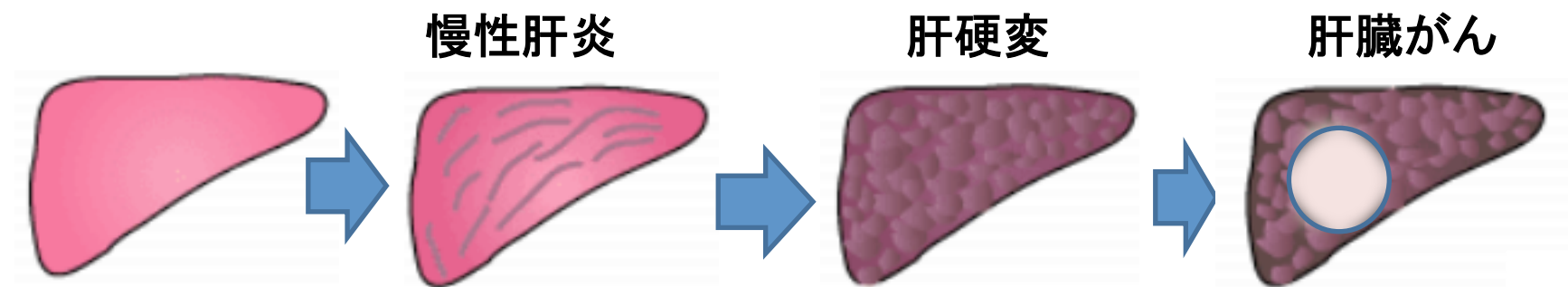
International Cancer Genome Consortium



一般公開



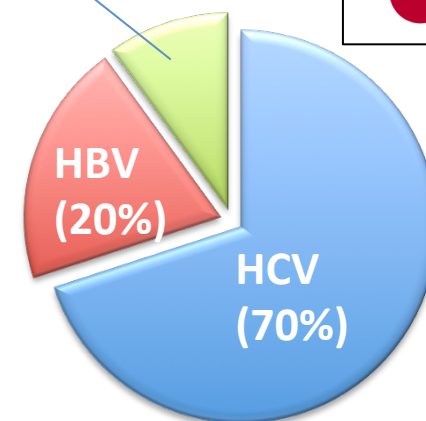
# 原発性肝臓がん



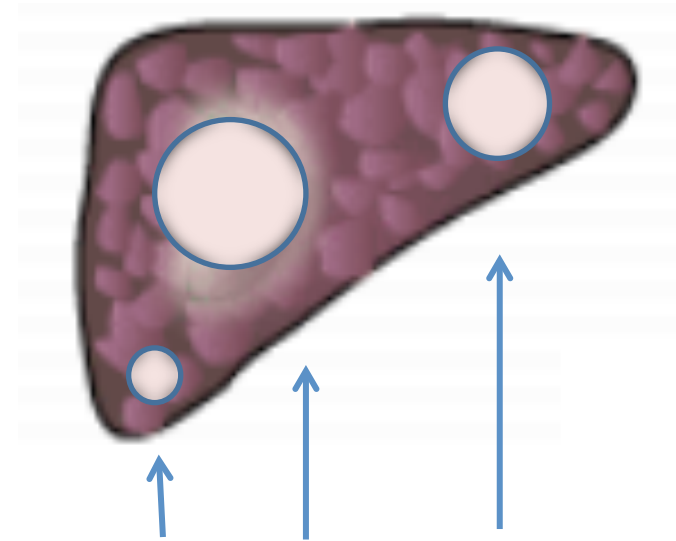
多様な要因

**B型肝炎ウイルス(HBV)**  
**C型肝炎ウイルス(HCV)**  
アルコール  
糖尿病、肥満、脂肪肝  
アフラトキシン  
など、多数

非B非C (10%~)



## 多中心性腫瘍 Multi-centric tumors (MC)



肝硬変といった強い発がんリスクの背景をもつ肝臓のもと、転移ではなく、独立した肝臓がんが同時性または異時性に多発してくる。特にC型肝炎関連。

# 25人からの 27例の肝臓がん のWGS

11人のB型肝炎

12人のC型肝炎  
(2人の多中心性を含む)  
14例の腫瘍

2人の非ウイルス性

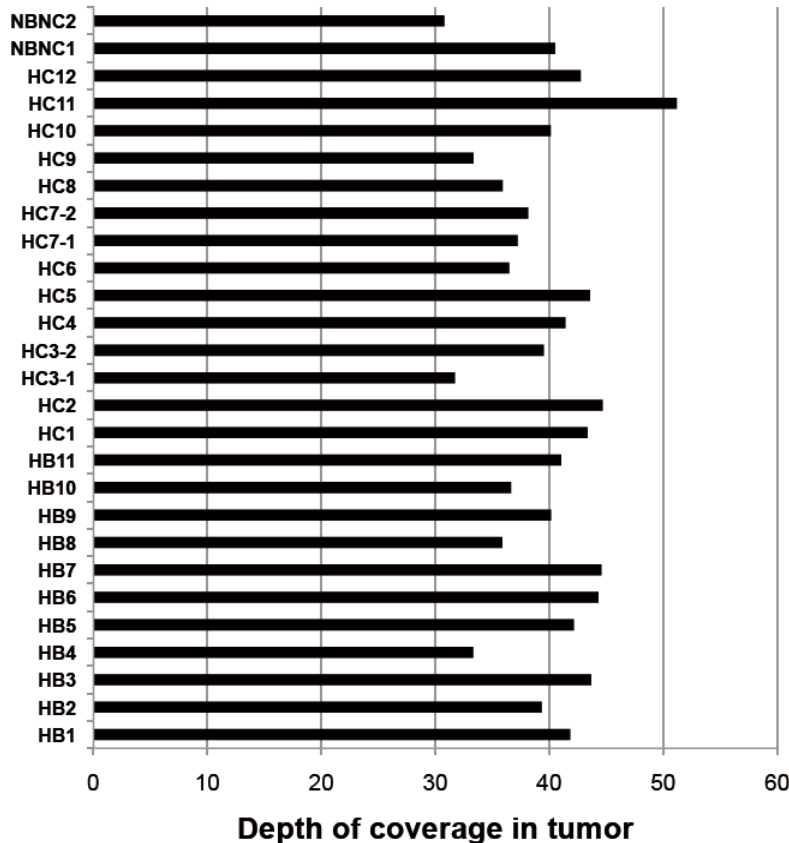
(Fujimoto et al. Nature Genetics 2012)

ID	年齢	性別	肝炎	習慣性飲酒	腫瘍径(mm)
HB1	56	M	HBV	+	33
HB2	46	M	HBV	-	24
HB3	74	M	HBV	+	60
HB4	61	M	HBV	-	30
HB5	38	M	HBV	+	18
HB6	57	M	HBV	+	30
HB7	58	M	HBV	+	22
HB8	68	F	HBV	+	45
HB9	60	M	HBV	-	80
HB10	56	M	HBV	+	63
HB11	41	F	HBV	+	30
HC1	62	M	HCV	-	87
HC2	71	M	HCV	-	45
HC3-1	69	M	HCV	+	20
HC3-2	69	M	HCV	+	20
HC4	61	M	HCV	+	18
HC5	58	M	HCV	+	24
HC6	61	M	HCV	+	35
HC7-1	64	M	HCV	+	27
HC7-2	64	M	HCV	+	25
HC8	73	F	HCV	-	21
HC9	66	M	HCV	-	40
HC10	62	F	HCV	-	20
HC11	71	F	HCV	-	35
HC12	57	M	HCV	+	43
NBNC1	81	F	nBnC	-	70
NBNC2	73	M	nBnC	+	60

# Average depth of coverage with uniquely mapped reads

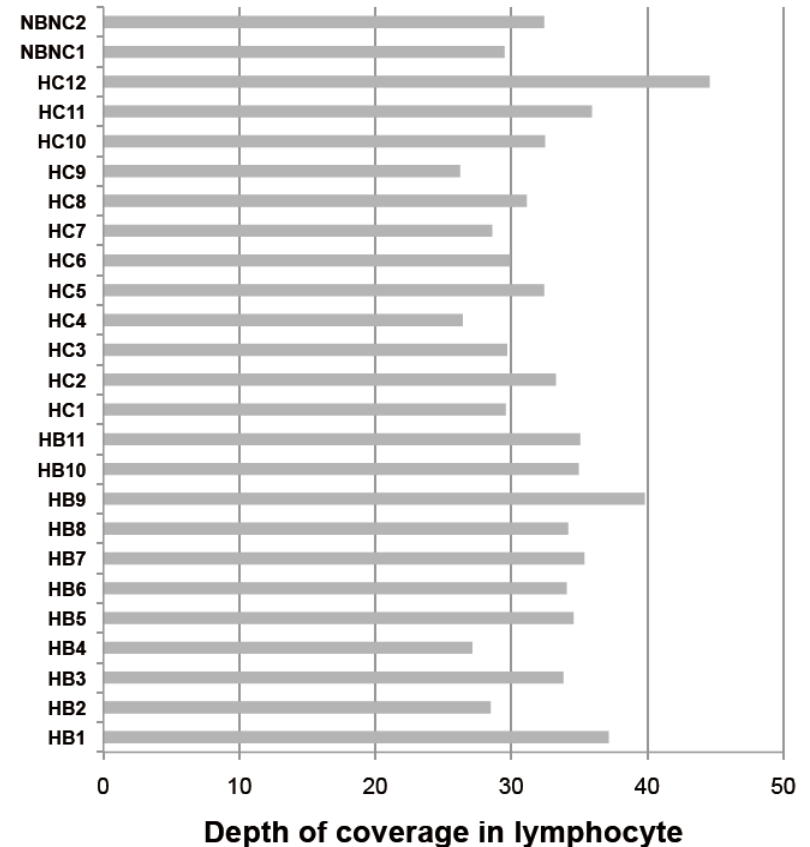
(a) Cancer

x40 depth



(b) Blood

x30 depth



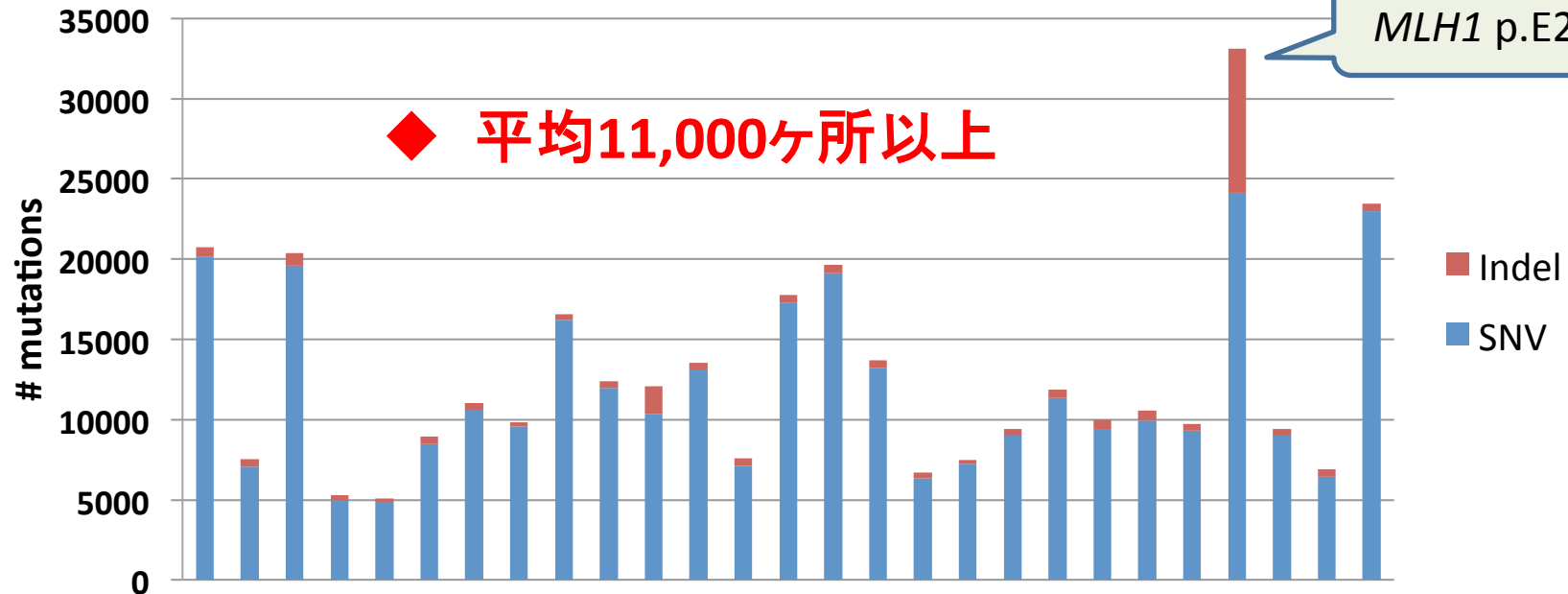
合計で約7テラベース (7兆塩基)

新聞1000年分

(Fujimoto et al. Nature Genetics 2012)

# Somatic mutation numbers of 27 HCCs

(Fujimoto et al. Nature Genetics 2012)



On average,

- ◆ **11,810** point mutations including **70** protein-altering mutations
- ◆ **813** indels including **6** coding indels
- ◆ **21** genomic rearrangements
- ◆ **2.1** HBV genome integrations

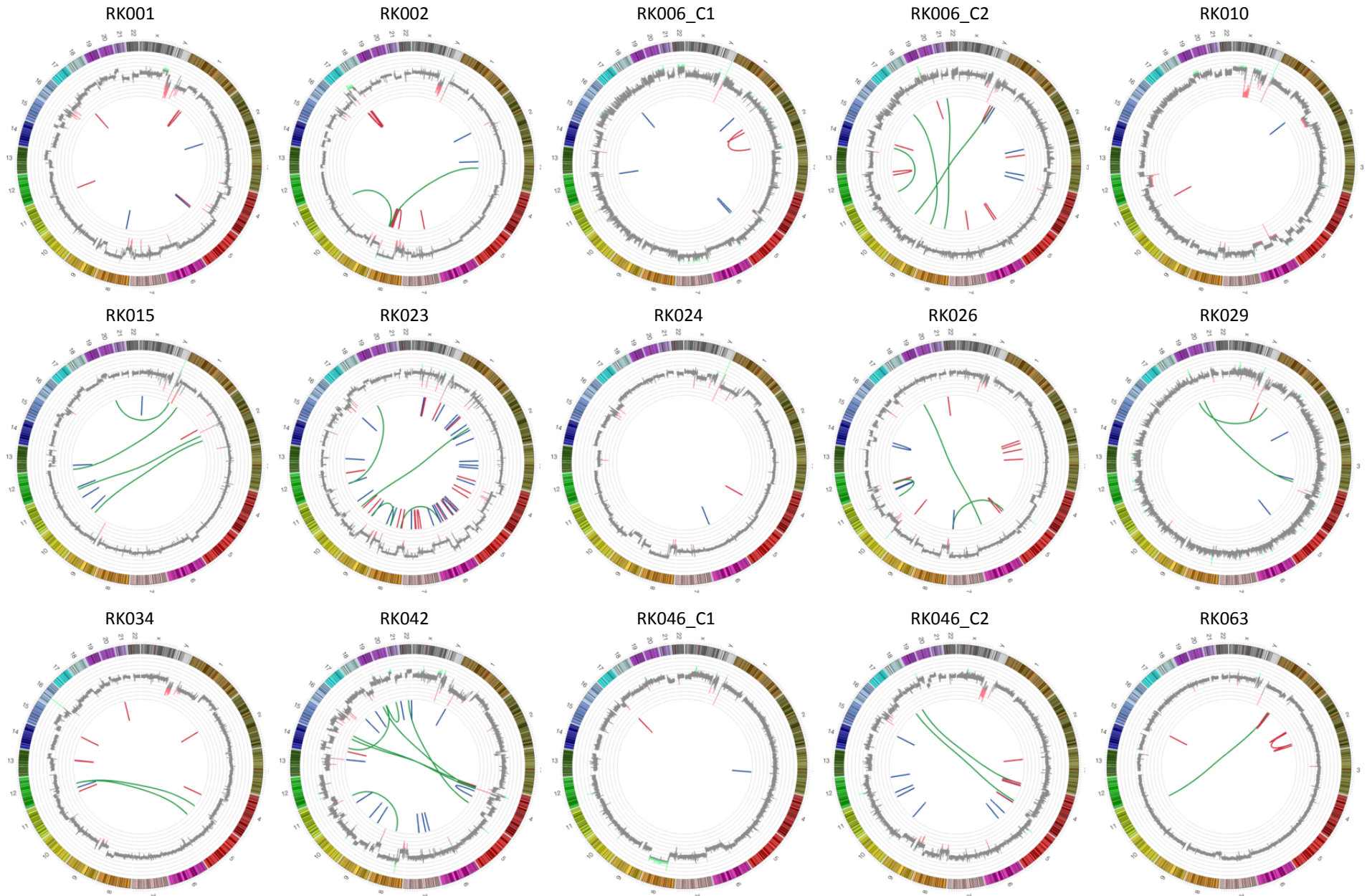


International  
Cancer Genome  
Consortium

**77** HCC genomes

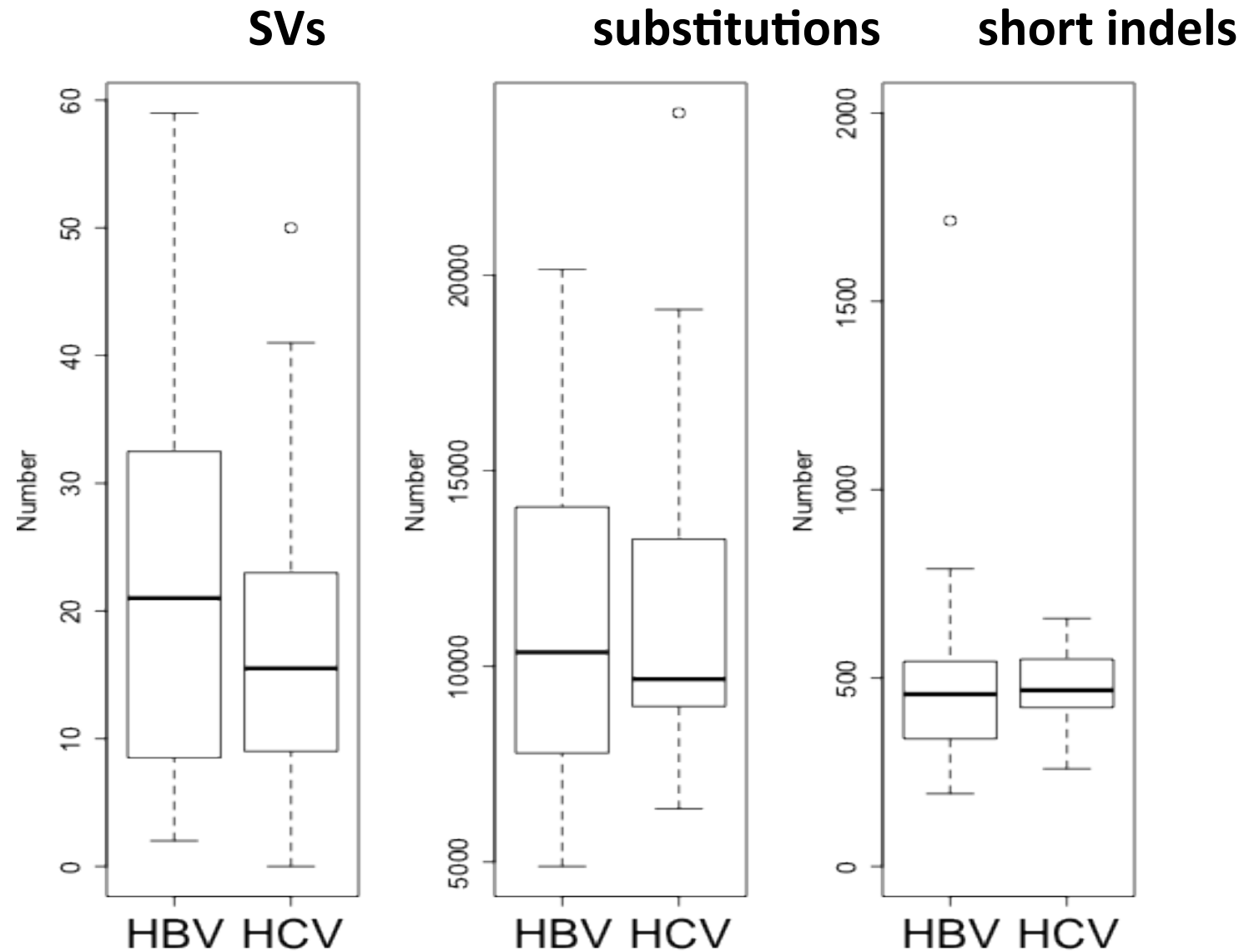
<http://dcc.icgc.org/>

# Circos Plots of Liver Cancer Genomes



Blue; stretched pair, Green; chromosomal translocation, Red; inversion or intra-chromosomal translocation

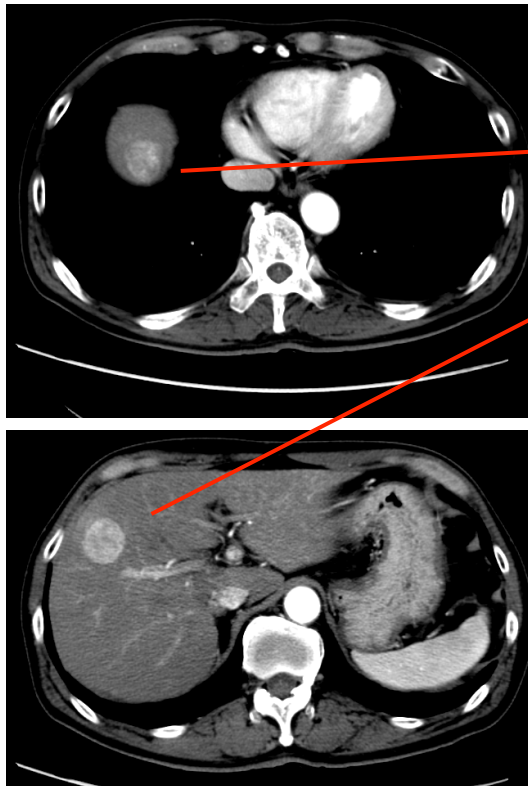
# B型肝炎関連とC型肝炎関連の肝臓がんの違い



(Fujimoto et al. Nature Genetics 2012)



# 多中心性腫瘍MC間の比較



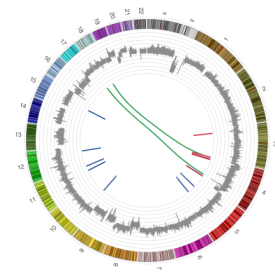
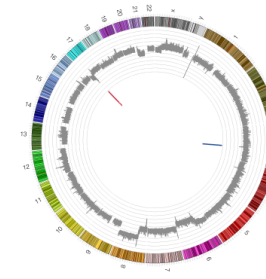
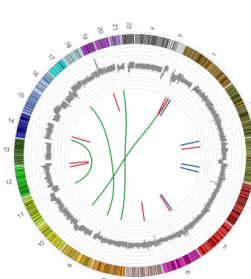
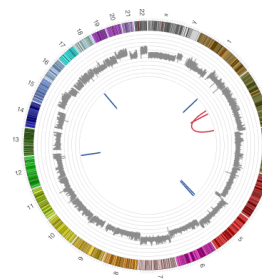
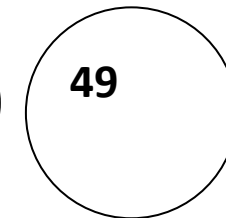
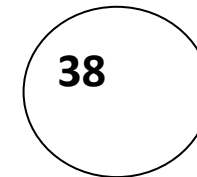
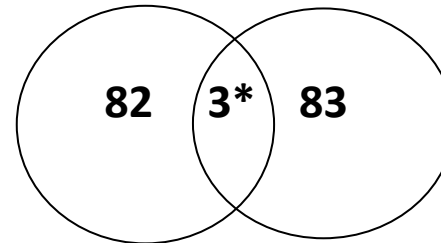
WGS

腫瘍1

腫瘍2

腫瘍1

腫瘍2



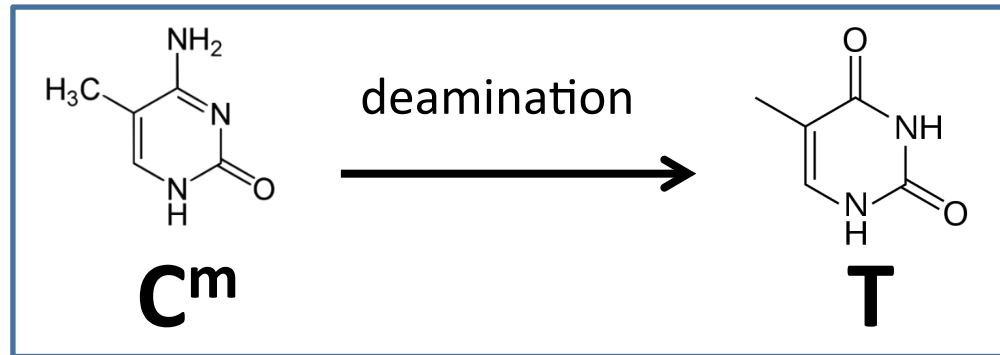
同じ肝臓から発生したMC腫瘍でも、ゲノムの変異場所は全く異なる

- >> MCは異なったゲノム変異を経てがん化
- >> **肝臓がんゲノムの“多様性”**



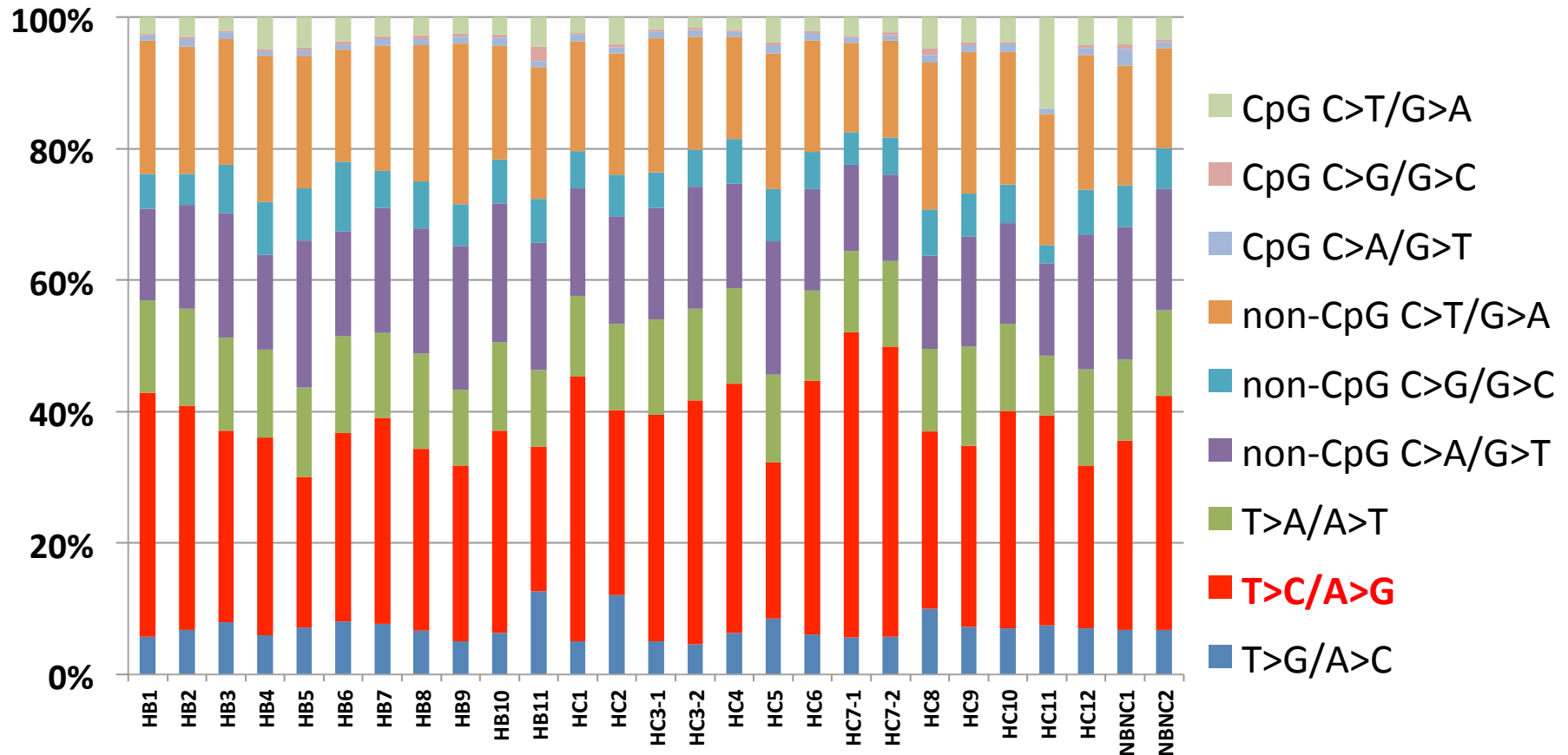
## 塩基配列の置換パターン

- ✓ 塩基置換パターンは、化学反応とDNA修復機構によってきまる



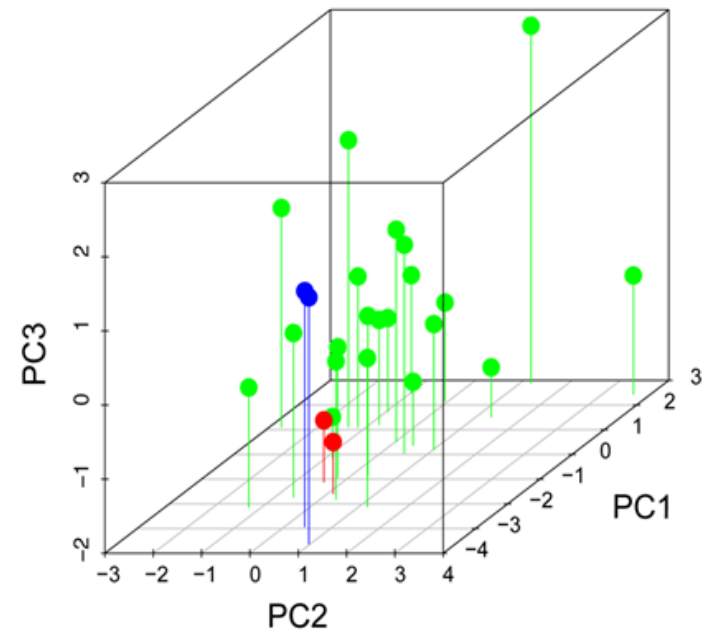
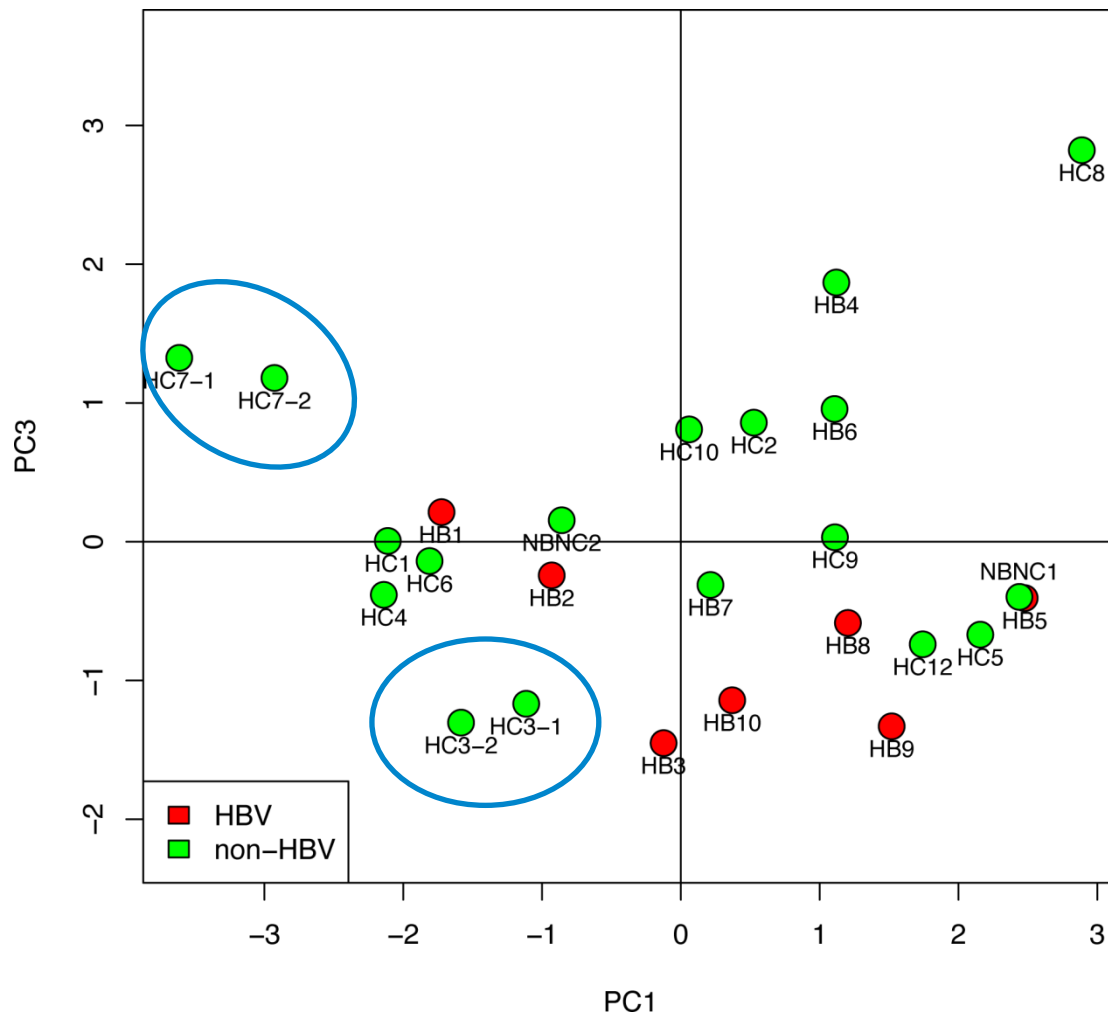
- ✓ その数やパターンは、発がん原因(物質)の暴露の影響を受ける  
タバコ関連の肺がんや紫外線関連のメラノーマは、10万個以上の塩基置換(変異)あり
- ✓ 発生臓器別に特徴的な塩基置換パターンがあり得る  
肝がんや腎がんは、T > C/A > Gが多い

# Genome-wide Somatic Substitution Patterns of 27 HCCs



(Fujimoto et al. Nature Genetics 2012)

# PCAによる肝臓がん変異の塩基配列置換パターン

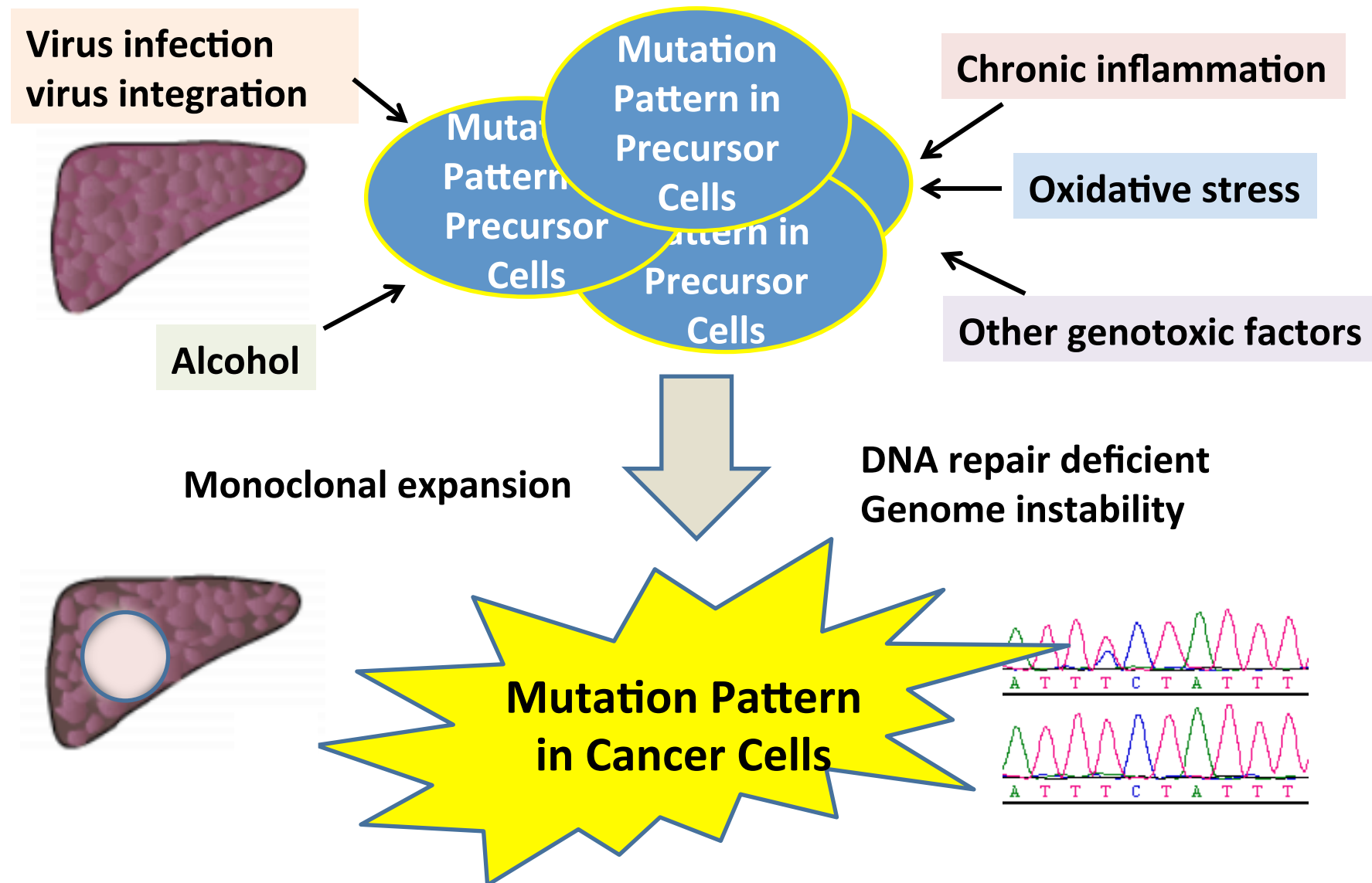


$P$ 値 = 0.00050

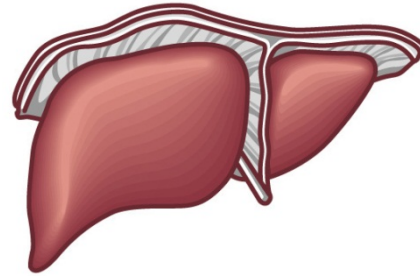
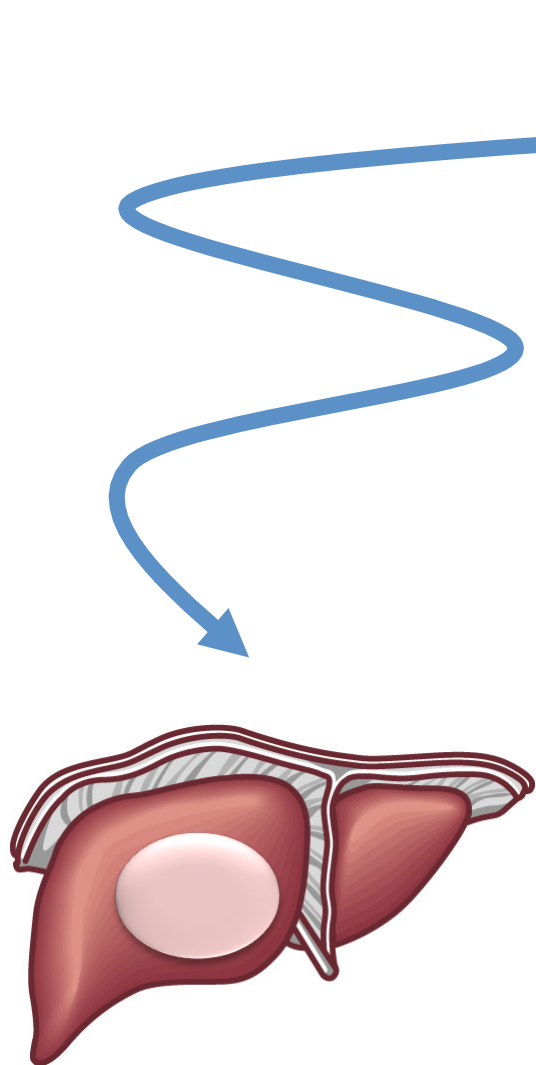
(Fujimoto et al. Nature Genetics 2012)



Somatic substitution patterns of HCCs is determined by their **etioloical backgrounds** where they developed



# Driver or Passenger Mutations





## 200~ recurrently mutated genes in 27 HCC genomes

Adjusting for the gene length and the regional deviation of somatic mutation rate

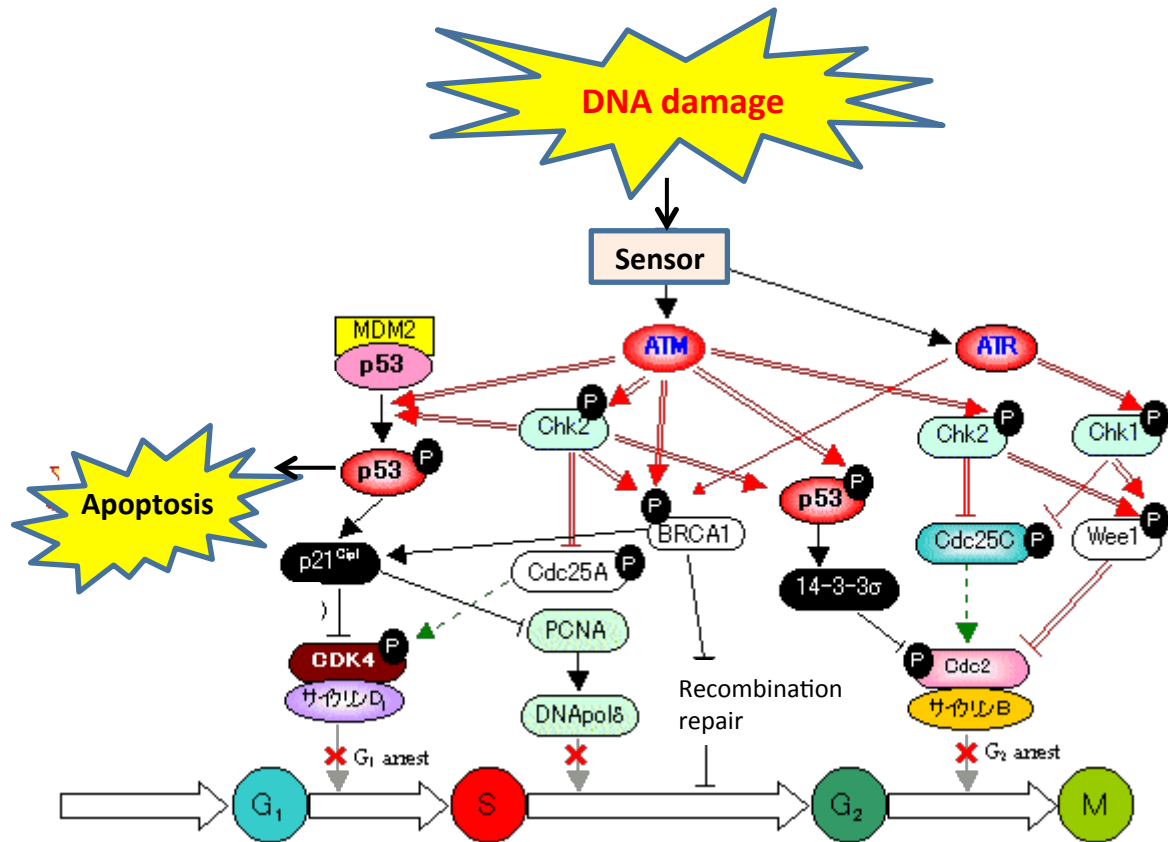
Gene	CDS length	Indel	Missense	Nonsense	Splice site	Total	P-value	q-value	Frequency in validation set
<i>TP53</i>	1,218	0	11	0	3	14	0	0	NA
<i>ERRFI1</i>	1,397	1	0	2	0	3	0.00020	0.0034	3.1% (2/65)
<i>ZIC3</i>	1,412	0	3	0	0	3	0.0005	0.0041	3.3% (4/120)
<i>CTNNB1</i>	2,398	0	3	0	0	3	0.0015	0.0071	NA
<i>GXYLT1</i>	1,351	0	3	0	0	3	0.0013	0.0071	0.8% (1/120)
<i>OTOP1</i>	1,859	1	2	0	0	3	0.0015	0.0071	0.8% (1/120)
<i>ALB</i>	1,882	3	0	0	0	3	0.0022	0.0089	3.3% (4/120)
<i>ATM</i>	9,415	1	4	0	0	5	0.0037	0.013	5.0% (6/120)
<i>ZNF226</i>	2,424	1	1	1	0	3	0.0043	0.014	3.3% (4/120)
<i>USP25</i>	3,260	1	2	0	0	3	0.0051	0.015	0% (0/120)
<i>WWP1</i>	2,857	2	1	0	0	3	0.0060	0.016	7.7% (5/65)
<i>IGSF10</i>	7,892	0	4	0	0	4	0.0091	0.023	3.3% (4/120)
<i>ARID1A</i>	6,934	2	1	0	0	3	0.011	0.026	10% (12/120)
<i>UBR3</i>	5,819	0	3	0	0	3	0.018	0.041	0.8% (1/120)
<i>BAZ2B</i>	6,643	0	3	0	0	3	0.024	0.050	1.6% (2/120)

(Fujimoto et al. Nature Genetics 2012)



# Aberrations of **ATM-p53 pathway** in 27 HCCs

ID	<i>TP53</i>	<i>ATM</i>	<i>ATR</i>
HB1			
HB5			
HB7			
HB8			
HB9			
HB10			
HB11			
HC2			
HC7-1			
HC9			
HC10			
HC11			
HC1			
HC3-1			
HC3-2			
HC5			
HB2			
HC7-2			
HC4			
HB3			
HB4			
HB6			
HC6			
HC8			
HC12			
NBNC1			
NBNC2			



## Gene set enrichment (GSE) analysis for genes with nonsense, coding indel and splice-site mutations (deleterious mutations)

Category	Term	Count	%	<i>P</i> -value	List Total	Pop Hits	Pop Total	Fold Enrichment	<i>q</i> -value
SP_PIR_KEYWORDS	phosphoprotein	162	55.1	6.92E-10	292	7263	19235	1.469292662	0.00000023
INTERPRO	IPR013032:EGF-like region, conserved site	19	6.5	8.02E-07	260	293	16659	4.154909425	0.00049
INTERPRO	IPR000742:EGF-like, type 3	15	5.1	2.14E-06	260	194	16659	4.954103886	0.0006
INTERPRO	IPR006210:EGF-like	15	5.1	3.25E-06	260	201	16659	4.781572905	0.0007
SP_PIR_KEYWORDS	egf-like domain	15	5.1	1.15E-05	292	230	19235	4.296083979	0.0019
UP_SEQ_FEATURE	domain:EGF-like 1	12	4.1	2.25E-06	292	120	19113	6.545547945	0.0039
SMART	SM00181:EGF	15	5.1	3.00E-05	175	201	9079	3.871641791	0.0043
SP_PIR_KEYWORDS	polymorphism	208	70.7	4.75E-05	292	11550	19235	1.18628951	0.0052
SP_PIR_KEYWORDS	calcium	28	9.5	9.42E-05	292	803	19235	2.2969515	0.0078
UP_SEQ_FEATURE	sequence variant	218	74.1	1.06E-05	292	11992	19113	1.189901144	0.0092
SP_PIR_KEYWORDS	<b>bromodomain</b>	6	2.0	2.89E-04	292	39	19235	10.13435195	0.019
SP_PIR_KEYWORDS	<b>chromatin regulator</b>	12	4.1	4.17E-04	292	213	19235	3.711171136	0.023
INTERPRO	IPR006209:EGF	10	3.4	1.66E-04	260	127	16659	5.045124167	0.025
SP_PIR_KEYWORDS	disease mutation	42	14.3	5.45E-04	292	1591	19235	1.738955426	0.026
INTERPRO	<b>bromodomain</b>	6	2.0	3.69E-04	260	40	16659	9.610961538	0.044
SP_PIR_KEYWORDS	tumor suppressor	9	3.1	0.0011571	292	137	19235	4.327442256	0.047

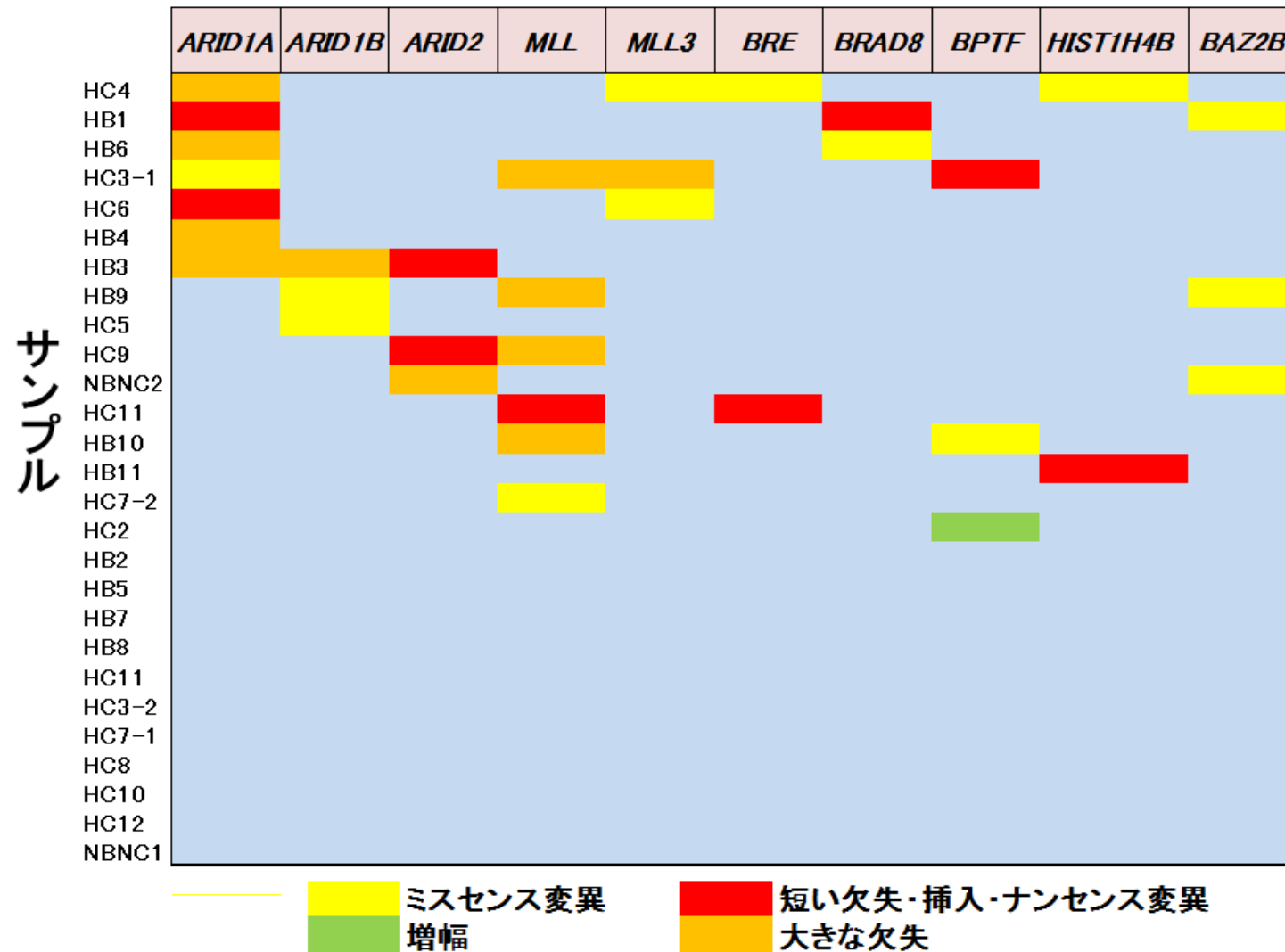
q-value was obtained by Benjamini and Hochberg's FDR method

(Fujimoto et al. Nature Genetics 2012)

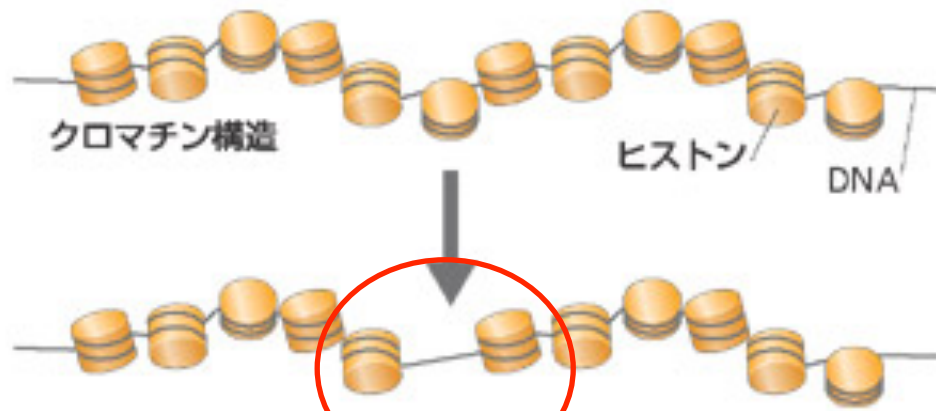
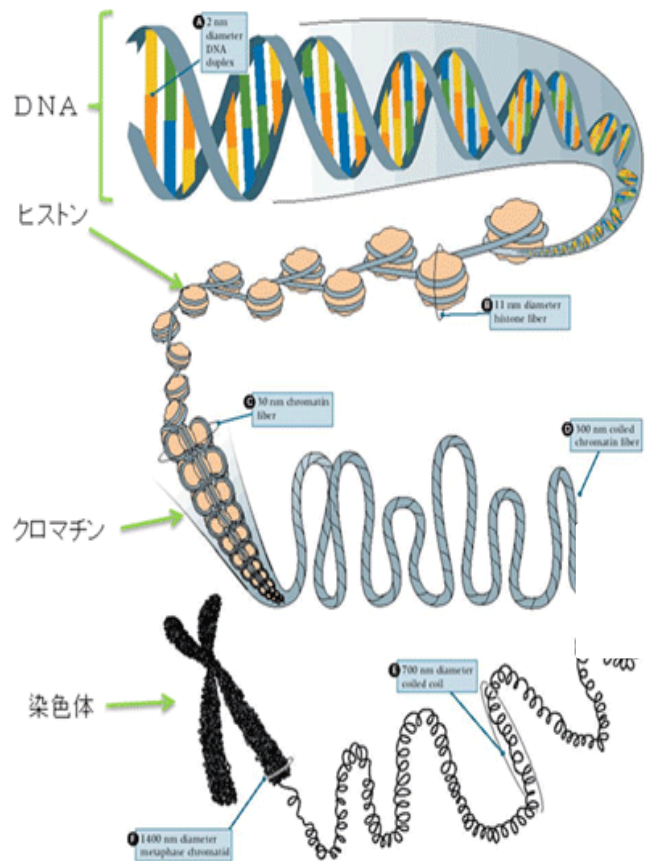
# Mutations of Chromatin regulator genes in HCCs

(Fujimoto et al. Nature Genetics 2012)

## クロマチン制御遺伝子

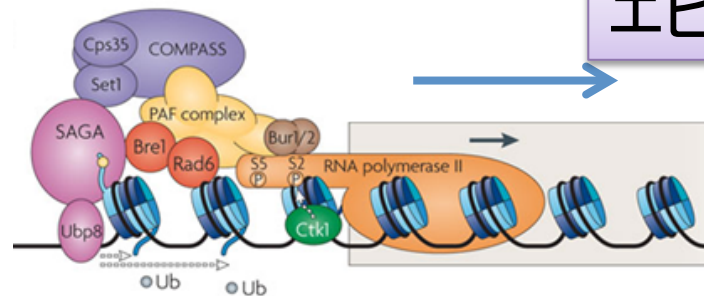


# クロマチン制御とは？



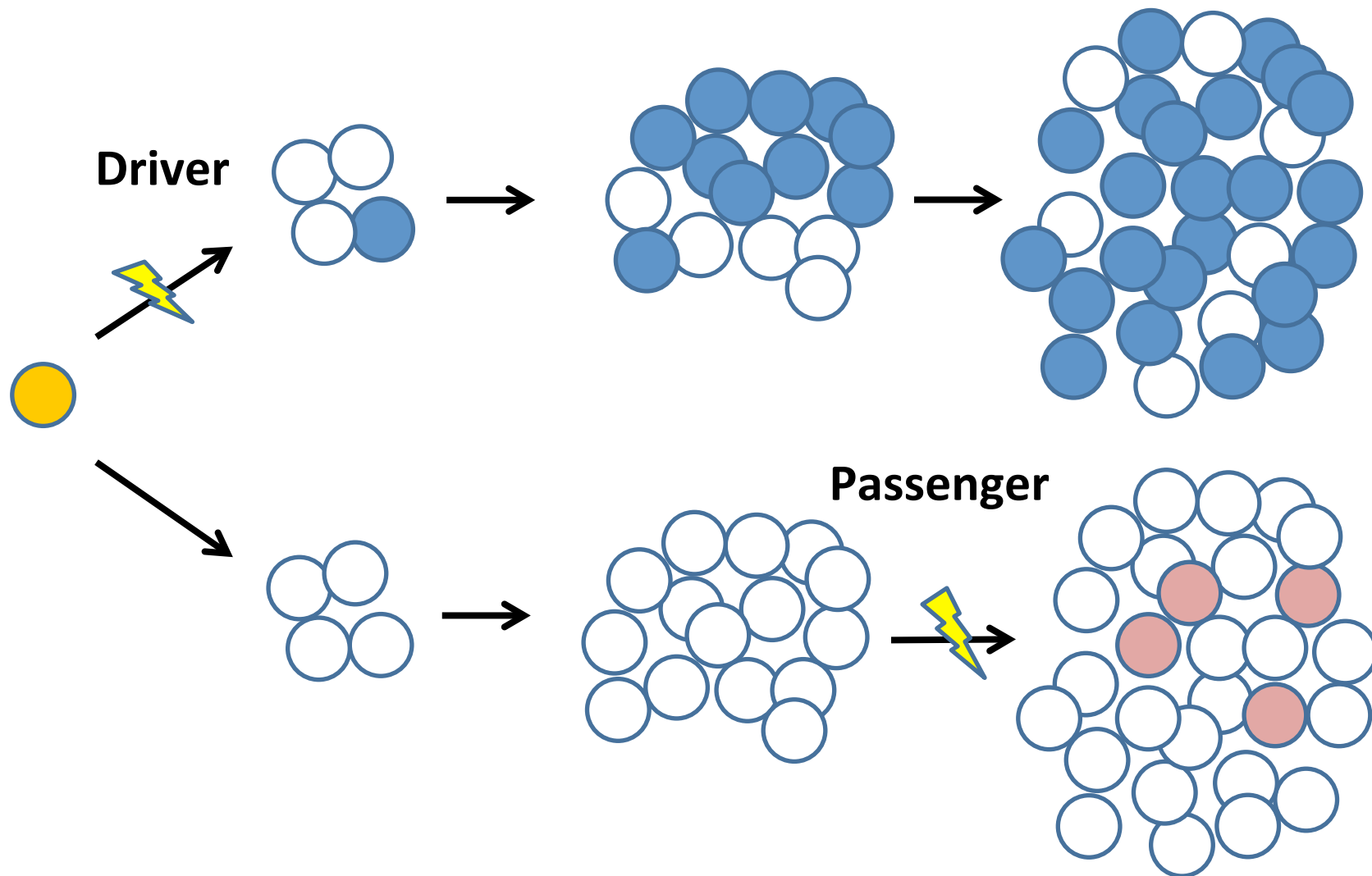
構造がほどける

エピゲノム



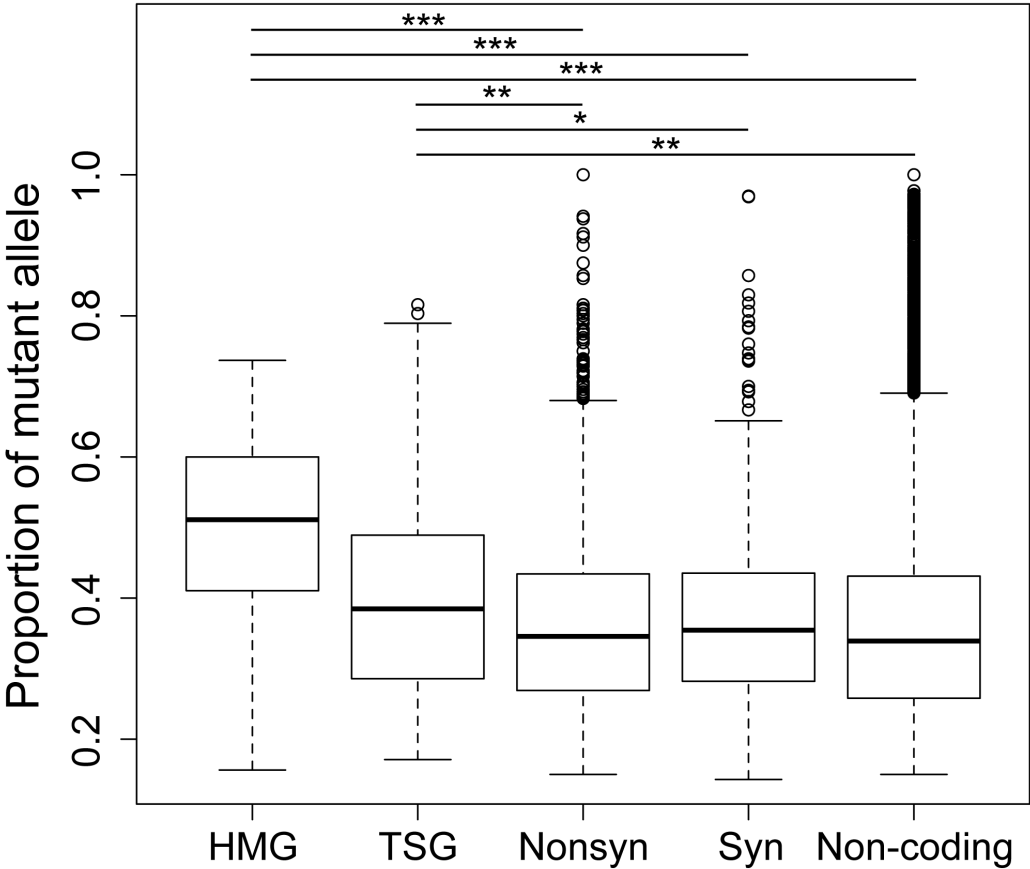
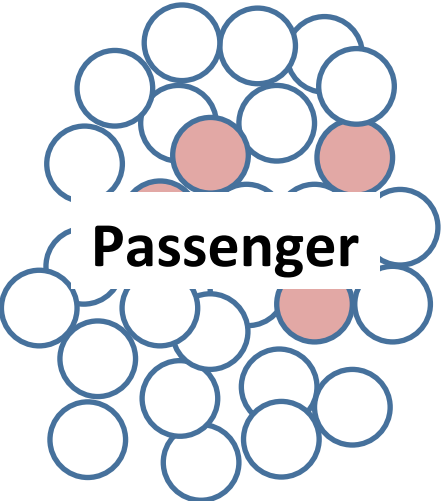
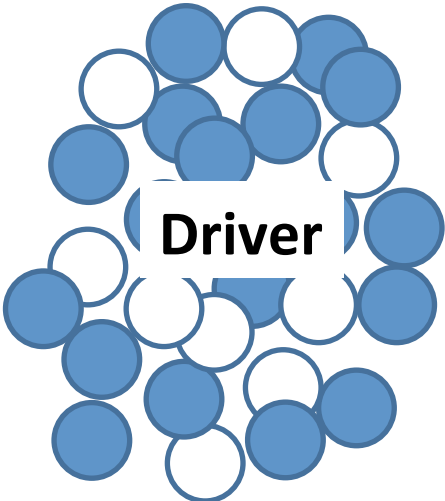
がんゲノム異常が、エピゲノム異常をひきおこす

# Cancer cell evolution



# Mutant allele proportions of point mutations

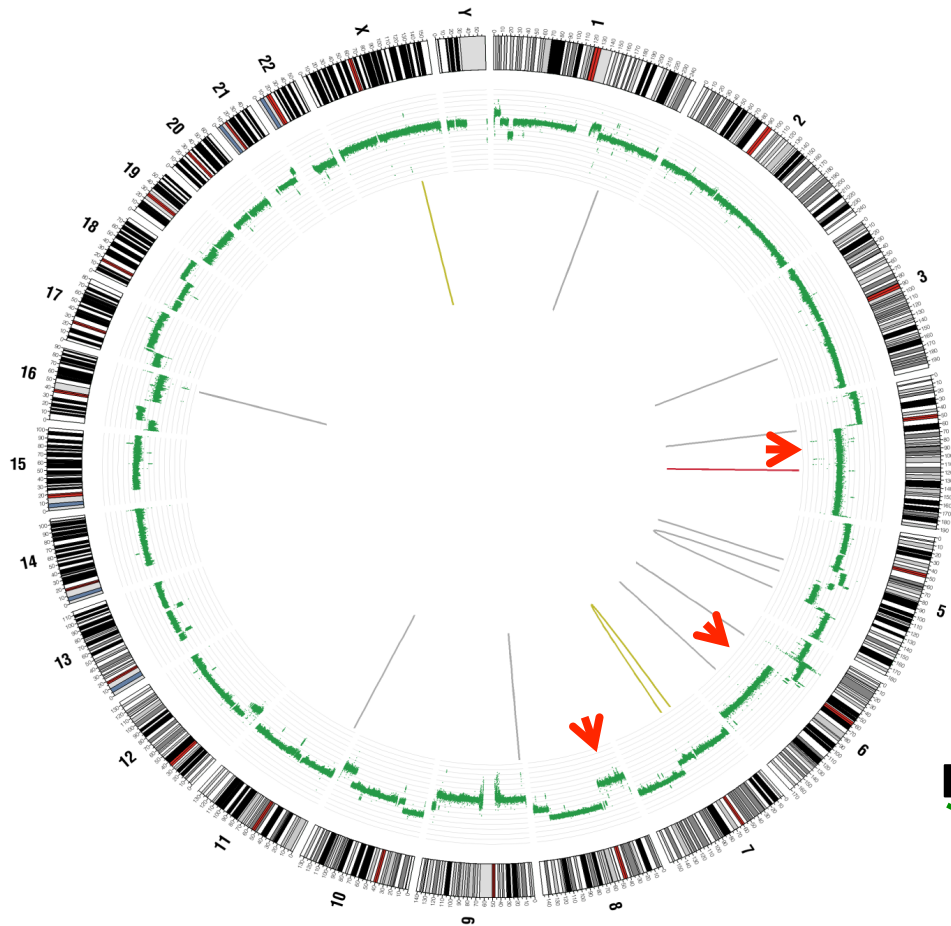
(Fujimoto et al. Nature Genetics 2012)



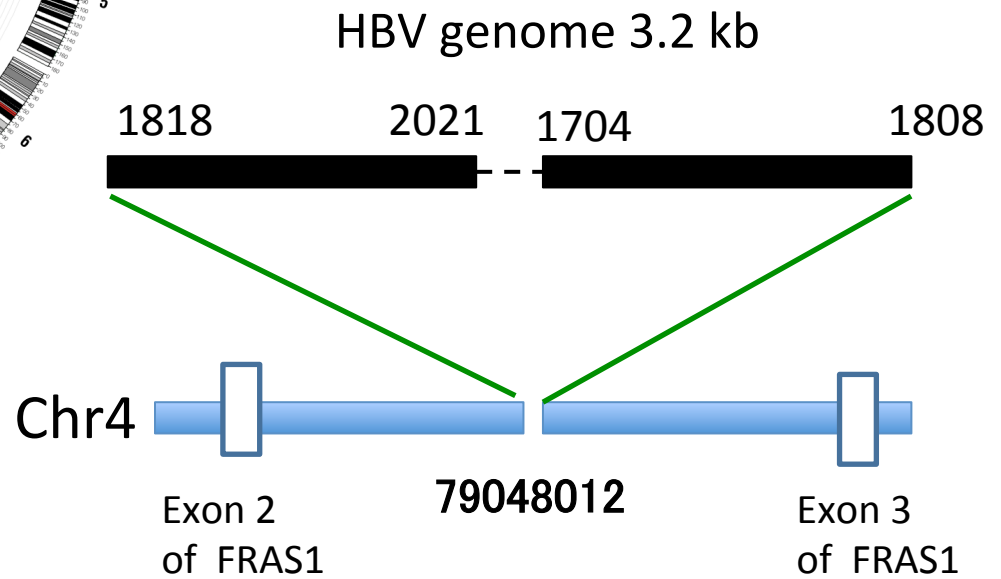
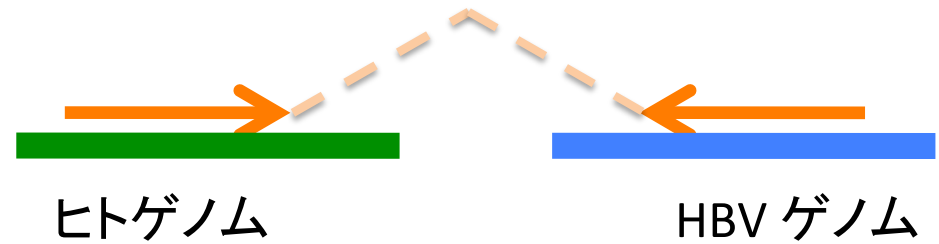
Highly mutated genes

Tumor suppressor genes

# B型肝炎ウイルスのintegration



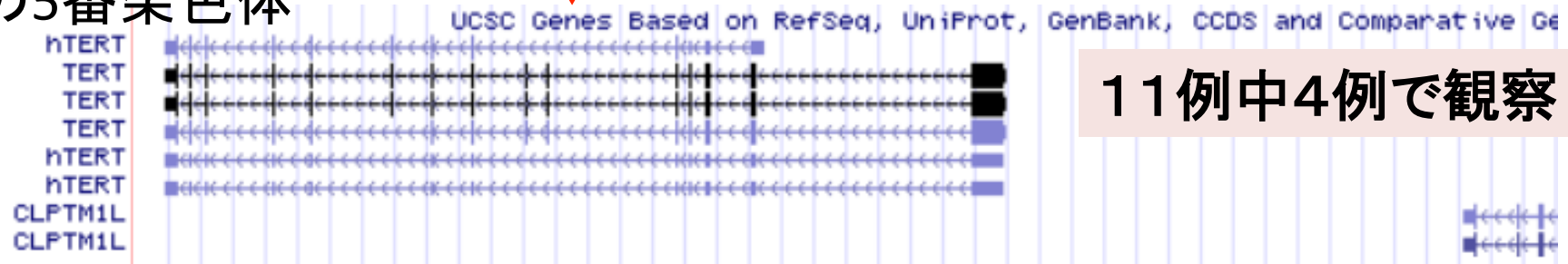
↑ HBV integration



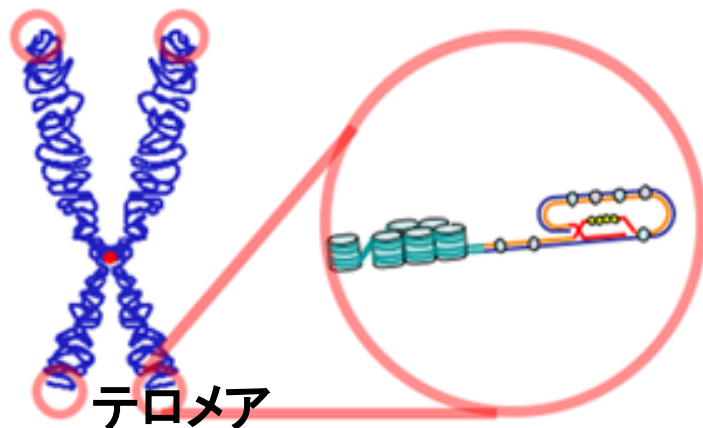
# B型肝炎ウイルスゲノムのヒトゲノムへの挿入

HBVゲノム挿入

ヒトの5番染色体



Sung et al. (Nature Genetics 2012) : 18/81 (22%) で、*TERT*へのintegrationを検出



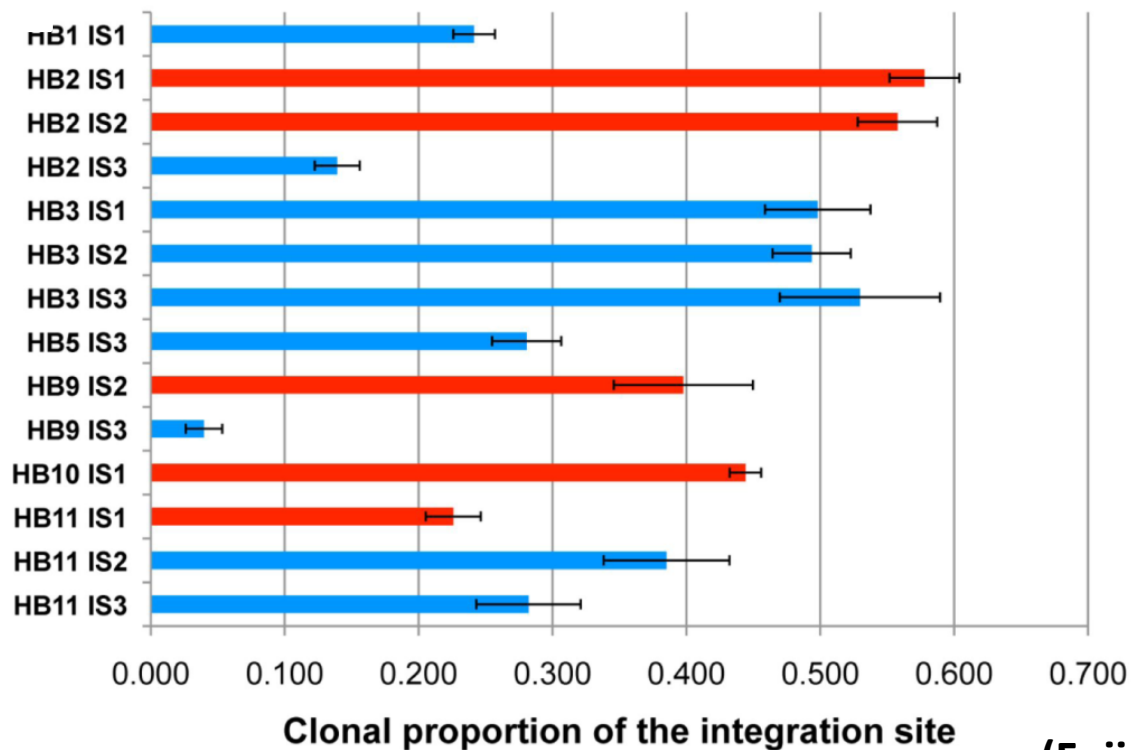
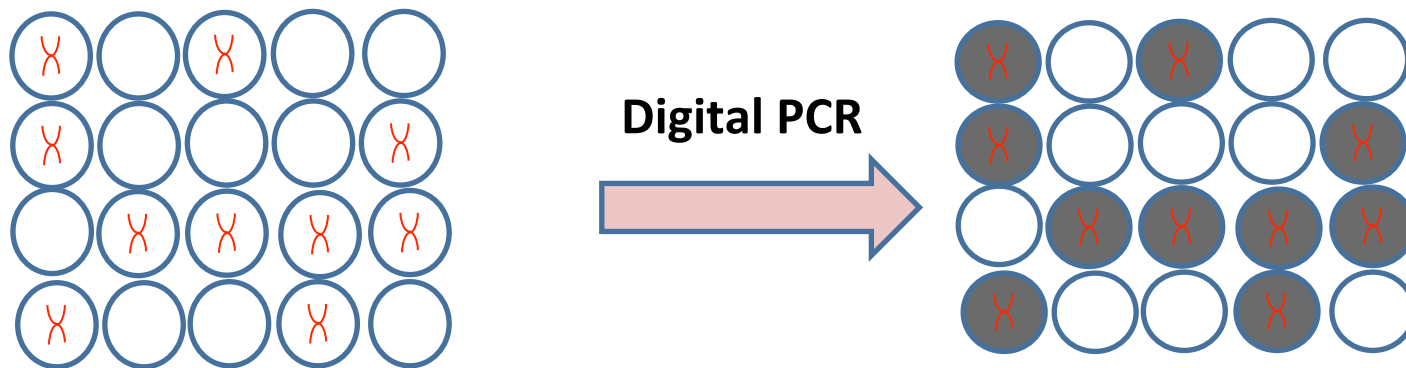
**TERT**  
(テロメア逆転写酵素)

テロメアの長さの維持に必須

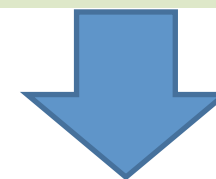
(Fujimoto et al. Nature Genetics 2012)



# Proportion of HBV integration sites (IS) in cancer cell-populations of HBV-integrated HCCs



TERT integrationは、  
tumor populationの  
多くで起こっている



肝癌の進化の過程に  
おいて、初期の段階  
でおきている

(Fujimoto et al. Nature Genetics 2012)

# 現状での(がん)WGSの要点

1. 試薬のコストは現在50万円以上だが、近い将来1000ドル以下へ (>> \$ 100 ?)
2. 現在のプラットフォームの精度は、depth/coverage (どれだけ、沢山読んだか)に依存
3. 多くのerrorは、少ないread数 (low depth)、または mapping エラーによるもの
4. ヒト参照ゲノム (build 37) にない配列が数Mbほどある
5. さらに精度をあげるためには、mappingの精度をあげるための長く正確なreadが必要
6. データ解析には巨大なITインフラと人材が必要
7. 今後、同定された大量の遺伝子変異・多型の生物学的・臨床的”解釈”をどう行うかが最大の問題

# Acknowledgements

(敬称略)

## 理化学研究所CGM

藤本 明洋  
角田 達彦  
中野 かおる  
来田 里奈  
佐々木 亜弥  
Nguyen Hai Ha  
阿倍 哲夫  
Keith Boroevich  
重水 大智  
青木 正幸  
久保 充明



## 国立がん研究センター研究所

柴田 龍弘  
十時 泰  
中釜 齐



## 和歌山県立医科大学第2外科

上野 昌樹  
山上 裕機



## 広島大学医学部

川上 由育  
有廣 光司  
大段 秀樹  
茶山 一彰



## 東京大学医科学研究所 ヒトゲノム解析センター

宮野 悟  
田中 洋子  
白石 友一  
渋谷 哲朗  
中村 祐輔



## 大阪府立成人病センター外科

山田 晃正  
後藤 邦仁  
石川 治



## 東京女子医科大学消化器外科

有泉 俊一  
山本 雅一

