

がんゲノム研究からみたポスト「京」への期待

宮野 悟

miyano@ims.u-tokyo.ac.jp

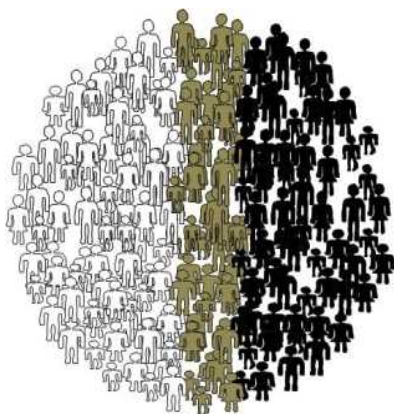
東京大学 医科学研究所 宮野 悟

HPCI計画推進委員会

ポスト「京」の利活用促進・成果創出加速に関するワーキンググループ

2019年2月26日(火) TKP虎ノ門駅前カンファレンスセンター

1/2 →がんを発症
1/3 →がんで亡くなる



「がんでは死なない」
という日本

なぜがんになるの？

DNAはたばこの煙や放射線、ウイルス感染などの環境因子や加齢により変わる

AAGCTCGAATTCGGAGAGATCTCTAGGAGAGAGAAAA

AAGCTCGCATCCGGAGAGA-----TAGGAGAGAGAAAA

TがCに変わった

TCTCがなくなった

警察「免疫細胞」に捕まらなくなった



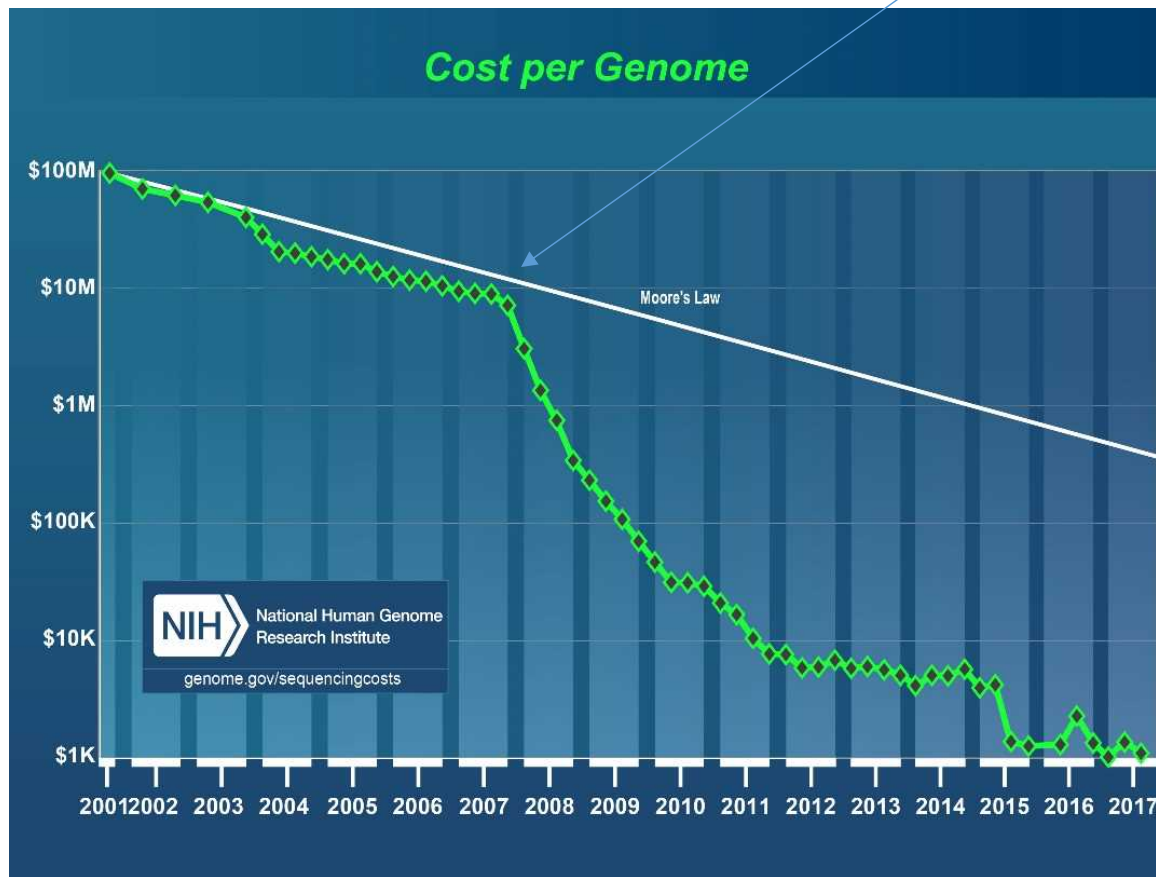
細胞増殖にブレーキが効かなくなった

ほとんどが「修理」されたり、「逮捕」されますが、システムが異常になるとそれができなくなります。

だから、がんのゲノム異常を調べるのです

ヒトゲノムは今600ドル、そして100ドル時代へ 次世代シーケンサーという装置の登場

私が昨年外注した最安価格は8万7千円
(カバレッジ30: 0.9テラ文字)



Illumina社製 NovaSeq 6000
600ドルゲノムを達成

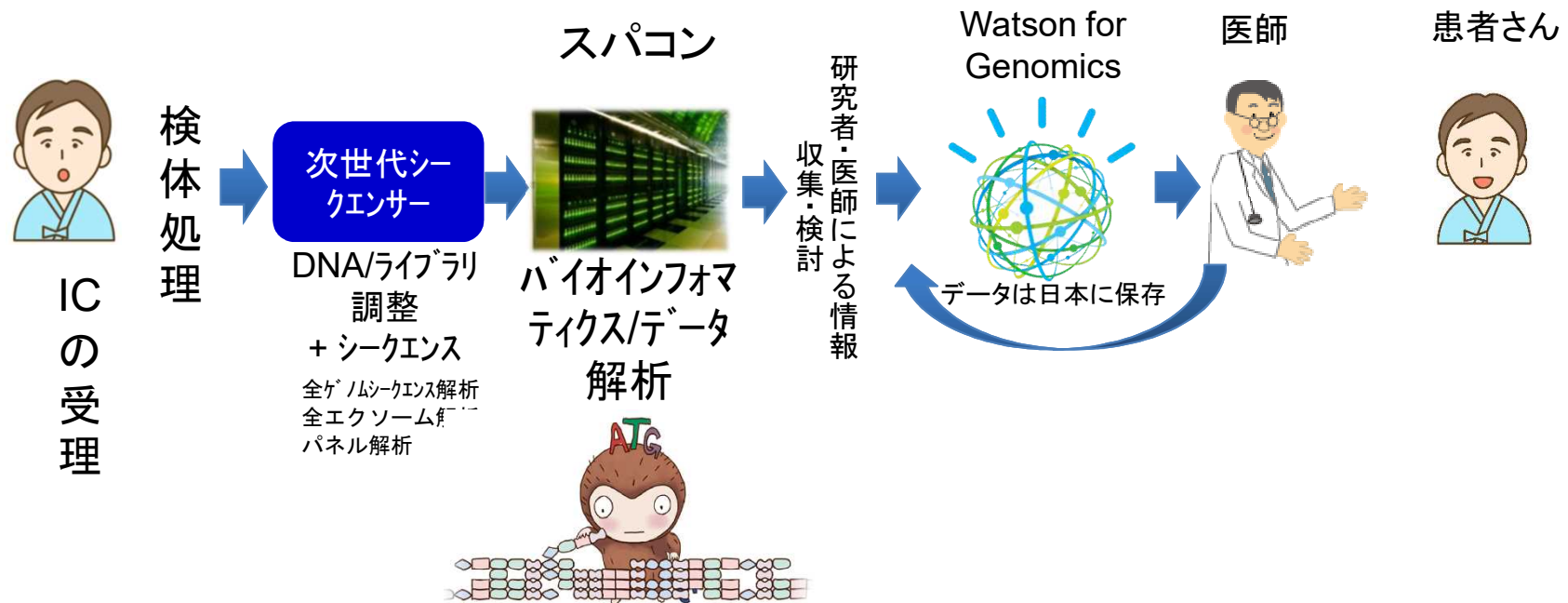
<https://jp.illumina.com/systems/sequencing-platforms/novaseq.html>

https://www.genome.gov/images/content/costpergenome_2017.jpg

次世代シーケンサーによるヒト ゲノム解析のコスト

- 私が今年（2019年1月）国内会社に入札外注したがん1検体あたりの全ゲノムシーケンスコスト（カバレッジ40: 6テラ文字）は15万円。
- 全エクソン（全遺伝子）解析（カバレッジ200）は15万円。
- ただし、情報解析・解釈の費用は含まず。

東大医科研の がんの臨床シーケンスの流れ



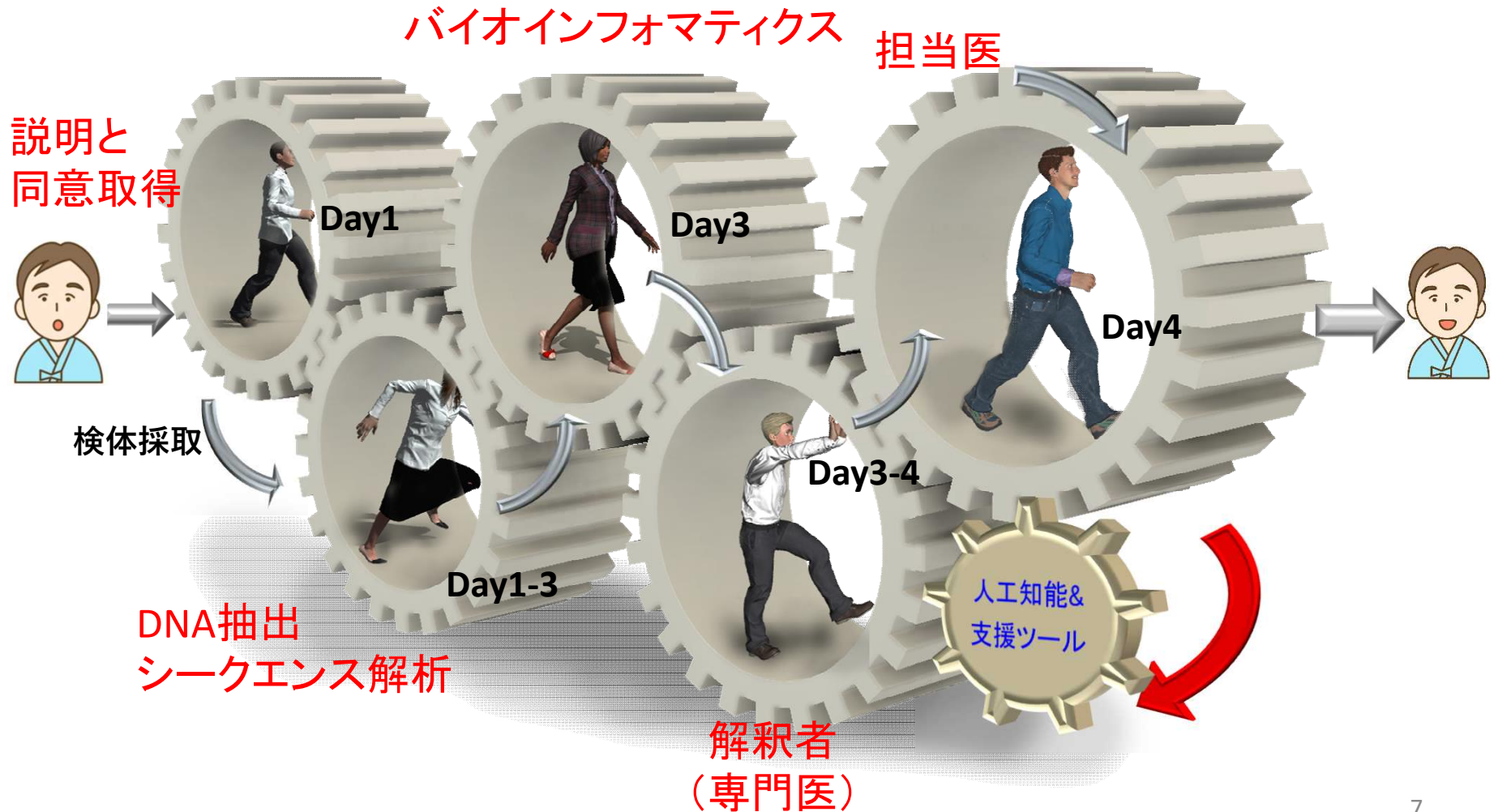
Genomon —The Zen of Cancer Genome

Genomon <https://github.com/Genomon-Project>

ゲノムデータを解析し、がんの変異を正確に暴き出すスパコンソフトウェア

患者さんは待てない

2016年、5日で全遺伝子解析・解釈結果を患者さんに返していた
2017年、全ゲノムシーケンスに基づき→3日と8時間



エピソード1

2018年、急性リンパ性白血病と診断され、 抗がん剤治療により寛解状態*の19歳男性

主治医から患者さんご家族に

「移植せず再発した場合、その後の治療で長期生存できる可能性は
1割以下です。**

移植した場合、2割以上の方が治療による合併症で亡くなります。

あなたの白血病には移植が必要かどうか私には分かりません。

移植...どうしますか？」

と言われる。

寛解状態*： 光学顕微鏡による検査で、白血病細胞が検出限界以下になり一定の治療効果が得られたと考えられる状態。しかし、寛解状態であっても、患者さんの体内には白血病細胞（微小残存病変）が残っていることがあります。

移植**： 造血幹細胞移植

主治医の言葉の背景

診療ガイドラインでは

“抗がん剤治療を行い

- ・ 微小残存病変陰性 ⇨ 移植は不要
- ・ 微小残存病変陽性 ⇨ 移植が必要
- ・ 微小残存病変が測定できない場合、移植を考慮しま
しょう”

とされている。

微小残存病変 (MRD: minimal residual disease)とは、抗がん剤の投与などにより、一定の効果が確認された後でも、患者さんの体内にまだ残っているだろうと想定されるがん病変（細胞）のこと。

参考

- ◆ 急性リンパ性白血病は抗癌剤による治療反応性はよいが、再発が多く、より治療強度の高い「**同種造血幹細胞移植**」を必要とすることが多い。
- ◆ 一般的には抗癌剤治療で寛解状態となった後に移植を行う。
- ◆ 移植を行わず抗癌剤治療のみで治療終了した場合、仮に**再発すると治療を行っても長期生存できるのは1割以下である**。
- ◆ 移植は**2割以上の治療関連死亡を伴う**非常に侵襲性の高い治療法である。幸運にも白血病がしなかったとしても、死亡以外に多くの重篤な合併症を来し、**不便なく日常生活を送れるようになるには数年を要する**。長期的にも**不妊や二次発がん**といった問題がある。
- ◆ 移植を必要としない患者集団もいるが、特に一般臨床検査では、**その集団の同定は容易ではない**。

ご家族のアクション

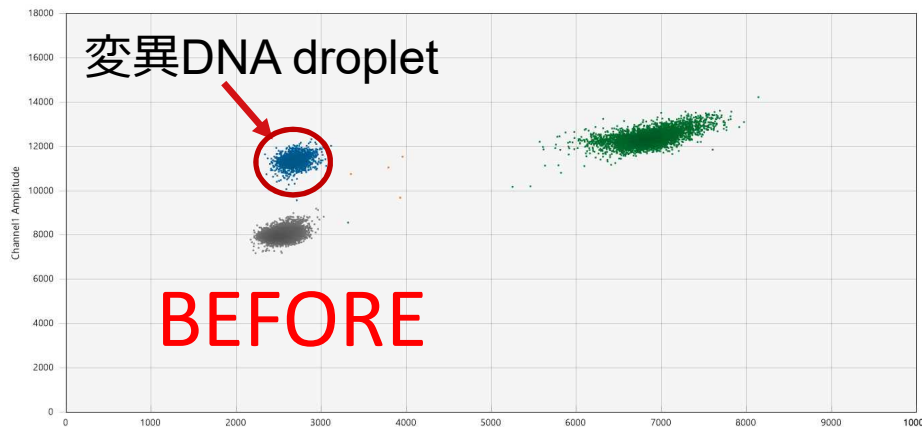
- 東大医科学研究所附属病院ゲノム診療部の古川洋一教授の「医科研でのがんのゲノム医療」の講演を聴き、古川教授に相談にこられる。
- 研究にご参加いただけるとのことと、ゲノムシーケンス解析と微小残存病変測定を希望される。血液腫瘍内科（東條有伸教授）で行うことにした。
- 治療前の試料の全エクソンシーケンス解析では微小残存病変があるとは確定できなかった。
- そこで、全ゲノムシーケンス解析をすると、12番染色体と17番染色体の構造異常による *TAF15-ZNF384* 融合遺伝子が見つかる。→ *TAF15-ZNF384* 融合遺伝子による白血病が報告されている。
- これが微小残存病変マーカーに使える！

全エクソンシーケンス：全遺伝子のタンパク質をコードしているところのみの解析で全ゲノムの2%の解析

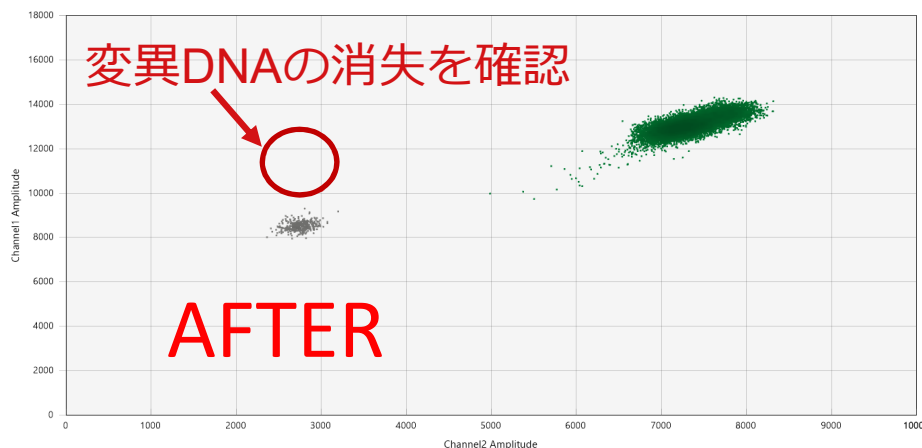
全ゲノムシーケンスでリスクの高い移植を回避

TAF15-ZNF384 融合遺伝子

骨髓液 2018/6/29 (治療開始前)



骨髓液 2018/11/14 (治療中)



TAF15-ZNF384 融合遺伝子による白血病の既報告を調べると世界で19人目の患者さんだった。

治療開始前に、全ゲノム解析により見つかった融合遺伝子を、MRD (微小残存病変) マーカーとして設定。

治療奏功、再発のマーカーに活用。

こんなことが実現できたのは
患者さんに合わせてスパコン・
シークエンサー・人員を運用す
るシステムができていたから

- しかも、ひとりの患者さんに集中して

そもそも
なぜがん研究や医療にス
パコンが必要なのか

今までの解析規模

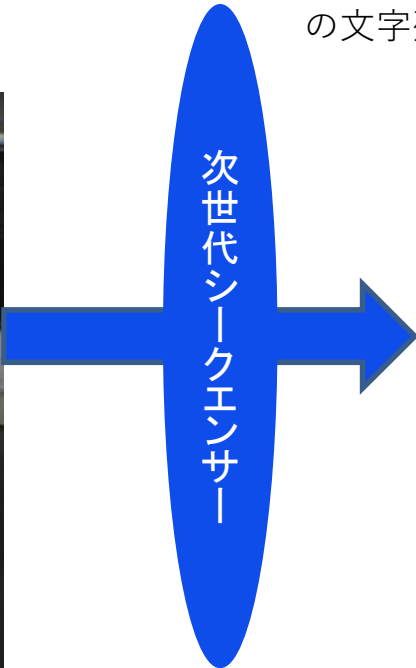
ヒトゲノムシーケンスって、どんなデータ？
生のサンプル スパコンで解析するデータ

DNAとして抽出

- 正常ゲノム：30コピー
- がんゲノム：40コピー

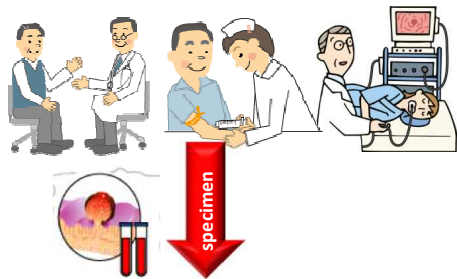
2,100億の
ATCGの文字データ

100文字ぐらいの断片になった21億ピース
の文字列断片がコンピュータに吐き出される



こんな量
じゃない
ですけど

スパコンで**9億 + 12億ピース**の2箱のジグソーパズルを解き、がんのシステム異常の原因を暴き出す！



次世代シーケンサー

Human Genome Center
Institute of Medical Science, University of Tokyo

Supercomputer

- Germline Genome
- Driver genes in genome
- Cancer genome

正常ゲノム
900億文字/100=9億ピース
がんゲノム
1200億文字/100=12億ピース

2箱の「ジグソーパズル」を解く

過去のストーリー

ヒトゲノム解析センタースパコン SHIROKANE “がん研究に必須のインフラ”

日本のトップがんゲノム研究はほぼ全てこのスパコンから出てきた

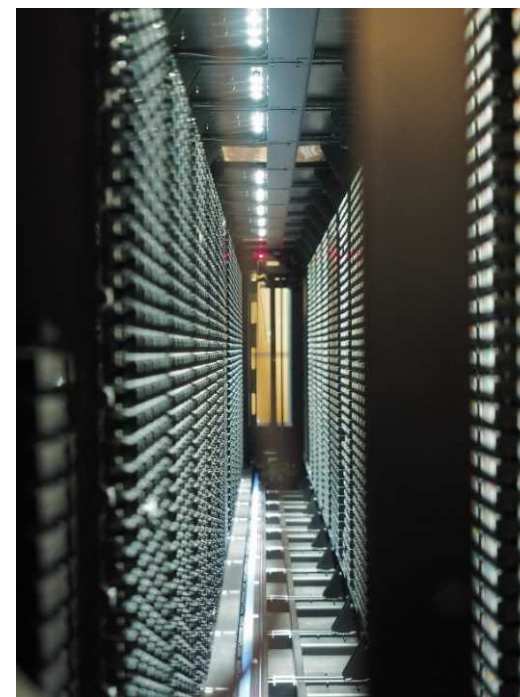
計算能力
(Thin:5GB/Core Fat: n 2TB/node)

高速ストレージ
Lustre File System

通常ストレージ
100PB

2019～1.5PFLOPS

30PB高速ディスクアレイ



IBM社製テープアーカイブ+
1PBニアラインディスク

Genomon GO! Get Mutations!

<https://github.com/Genomon-Project>

Since 2010

A suite of bioinformatics tools for analyzing
Cancer Genomes and RNA-seq data



Yuichi Shiraishi, PhD



Keisuke Kataoka, MD, PhD



Satoru Miyano, PhD



Seishi Ogawa, MD, PhD

Genomon —The Zen of Cancer Genome

Sequence Analysis

「京」コンピュータを用いた大規模データ解析

次世代シーケンスとスパコンの融合により
大規模全ゲノムシーケンスデータの高速解析を実現

(独)理化学研究所提供



HPCI戦略プログラム 分野1

「予測する生命科学・医療および創薬基盤」

大規模生命データ解析

代表:宮野悟₂₀

Genomon Discovery

がんはどのようにT細胞(キラー)から逃れているのか



nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For

Research > Letters > Article

ARTICLE PREVIEW

[view full access](#)
[options](#)

NATURE | LETTER

Aberrant *PD-L1* expression through 3'-UTR disruption in multiple cancers

Keisuke Kataoka, Yuichi Shiraishi, Yohei Takeda, Seiji Sakata, Misako Matsumoto, Se Nagano, Takuya Maeda, Yasunobu Nagata, Akira Kitanaka, Seiya Mizuno, Hiroko Tan Kenichi Chiba, Satoshi Ito, Yosaku Watatani, Nobuyuki Kakiuchi, Hiromichi Suzuki, Tetsuichi Yoshizato, Kenichi Yoshida, Masashi Sanada, Hidehiro Itonaga, Yoshitaka Imaizumi, Yasushi Totoki, Wataru Munakata, Hiromi Nakamura, Natsuko Hama, Kotaru Shide, Yoko Kubuki, Tomonori Hidaka, Takuro Kameda, Kyoko Masuda, Nagahiro Mi Koichi Kashiwase, Koji Izutsu, Akifumi Takaori-Kondo, Yasushi Miyazaki, Satoru Taka Tatsuhiko Shibata, Hiroshi Kawamoto, Yoshiki Akatsuka, Kazuya Shimoda, Kengo Takeuchi, Tsukasa Seya, Satoru Miyano & Seishi Ogawa [Show fewer authors](#)

[Affiliations](#) | [Contributions](#) | [Corresponding author](#)

Nature (2016) | doi:10.1038/nature18294
Received 01 March 2015 | Accepted 29 April 2016 | Published online 23 May 2016

- 遺伝子 *PD-L1* のタンパク質をコードしていない領域の構造異常(本の乱丁)が原因で、*PD-L1* タンパク質(賄賂)が異常に作られ、T細胞(警察)の賄賂受取ポケットに渡される。
- その結果、T細胞はがん細胞を殺さなくなった。
- 複数種類のがんで観察された。

Genomon on スパコンでいっぱい研究成果がでてますが・・・

- **Nature** 2019 加齢に伴う正常組織の遺伝子異常とがん化のメカニズムの解明
- **Nature Microbiology** 2019 慢性活動性 EB ウイルス感染症の原因解明
- **Nature** 2016 免疫系を回避するがん細胞のゲノム異常の解明
- **Nature Genetics** 2015 ヒト白血病ウイルスとT細胞の戦場跡の全貌
- **New England J Medicine** 2015 再生不良性貧血のクローン進化
- **Nature** 2011 骨髄異形成症候群の解明

Journal	#
<i>New Eng J Med</i>	1
<i>Nature</i>	3
<i>Nature Genetics</i>	10
<i>Science</i>	1

(2011 ~ Today) + 60~ publications

こんなことができたのは、スパコンパワーだけでなく、大規模データ解析を考慮したスパコンの運用方法による

ほとんどが、タッチの差で世界一に
スパコンの運用による

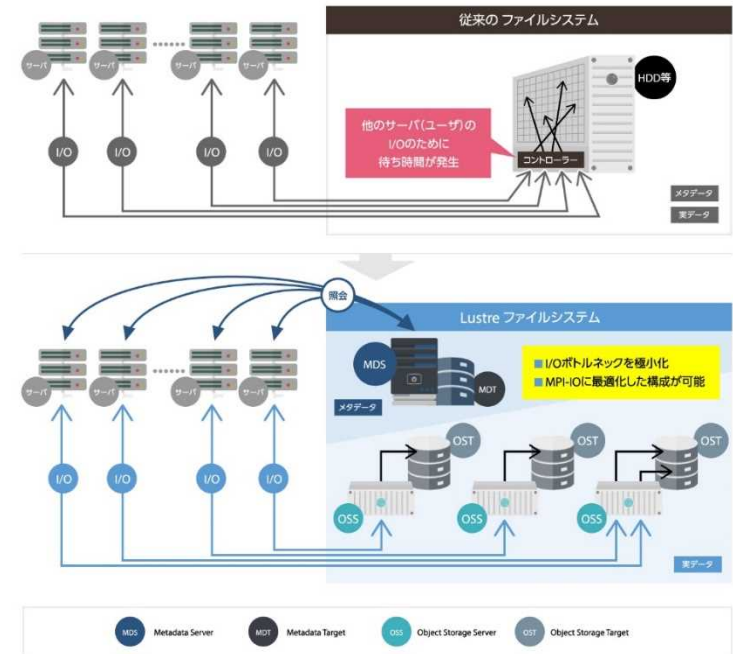
「京」大規模データ解析実行時に判明したことの一部

• ステージング関連

- 必要ノード以上のノードが必要。
 - Quota: 14~29GB/node
- アップロードに長大な時間が必要。
 - ジョブスタートまでにデータのアップロードが完了せず、計算が開始されない。

• 分散ファイルシステム(Lustre)

- MDSへの過負荷でシステムがダウンする。
 - 計算開始時に全ワーカがログファイル生成。
 - splitや並列処理のためにファイルを分割/結合する際に生じる、一時的なファイル増加
 - 2~3,000ファイル/ディレクトリが限界



「京」を使うことは「刑」か？

未来に向けて

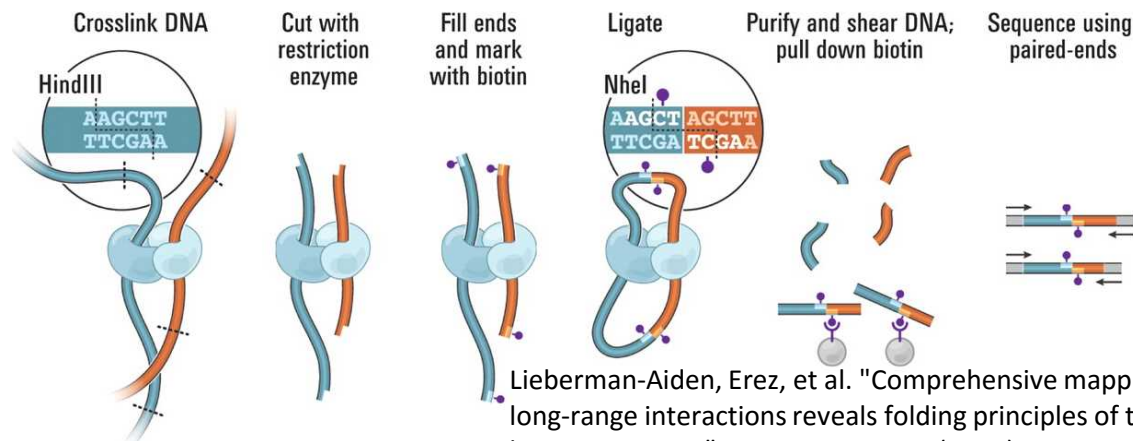
これからの世界トップがんゲノム研究
では100～300コピーを全ゲノムシー
クエンスしないと脱落する

- 研究としては数千のサンプル解析が当たり前
- 予防・医療としては時系列・多領域での全ゲノムシークエンス

Hi-C解析

ゲノム(染色体)の3次元空間内の立体構造を明らかにする

- Hi-C解析とはシーケンシングによって、ゲノム配列そのものからはうかがい知れない染色体の立体構造を、全領域 vs. 全領域の近接性をすべてシーケンシングで決定してしまう。
- 膨大な組み合わせ探索が発生。




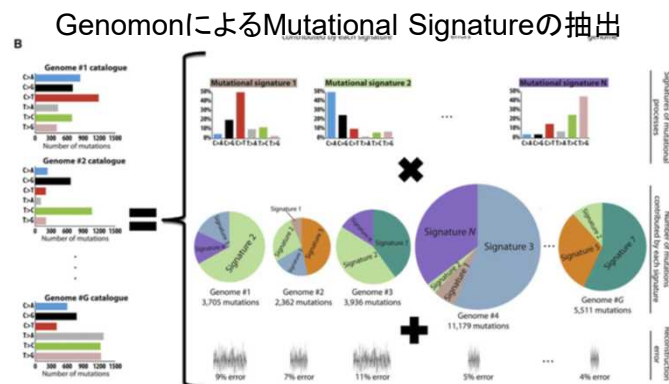
Lieberman-Aiden, Erez, et al. "Comprehensive mapping of long-range interactions reveals folding principles of the human genome." *Science* 326.5950 (2009): 289-293.

- ヒトゲノム解析センターのスパコンではもう無理
- ポスト「京」では大丈夫！
- ただし運用が大問題
- AIの活用が不可欠

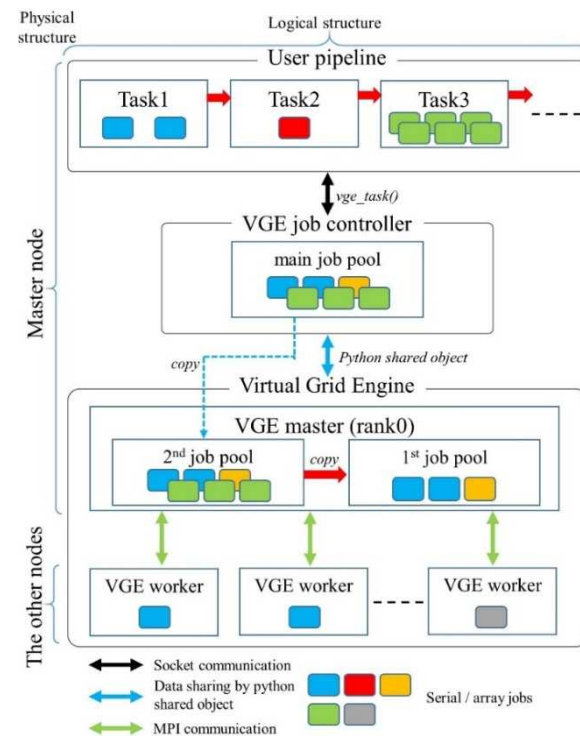
Genomon on ポスト「京」は1000WGS/Dayを達成予定

- ポスト「京」で稼働するVirtual Grid Engineを同時に開発。
- ジョブを投げ込むだけ。

がんの変異を正確に暴き出すソフトウェア
Genomon <https://github.com/Genomon-Project/>
 ゲノモンGO Mutationをゲットしようぜ！

Virtual Grid Engineのワークフロー



VGEは主として3種のプロセスで構成される。VGEは起動時に指定されたプロセス数のMPIプログラムとして動作を開始する(VGE開始)。初期設定を読み込んだのち、マスターランクとそれ以外で動作が区分される。マスターランクの場合、pythonのmultiprocessingモジュールを利用して二つの子プロセスが起動する。

がんの早期発見・個別化医療には、
多数の患者さんのデータに対応で
きるトータルな運用システムが必要。

がんの理解と医療は、もはや生物学・医学の領域から昇華し、新しい次元へのオデッセイが始まったとあってよい。

ポスト「京」はその最前線を走ってほしい。

「がんゲノム研究」で常勝

參考資料

エピソード2

急性骨髄性白血病(AML)

人間ドック白血球、好酸球増多を指摘された。

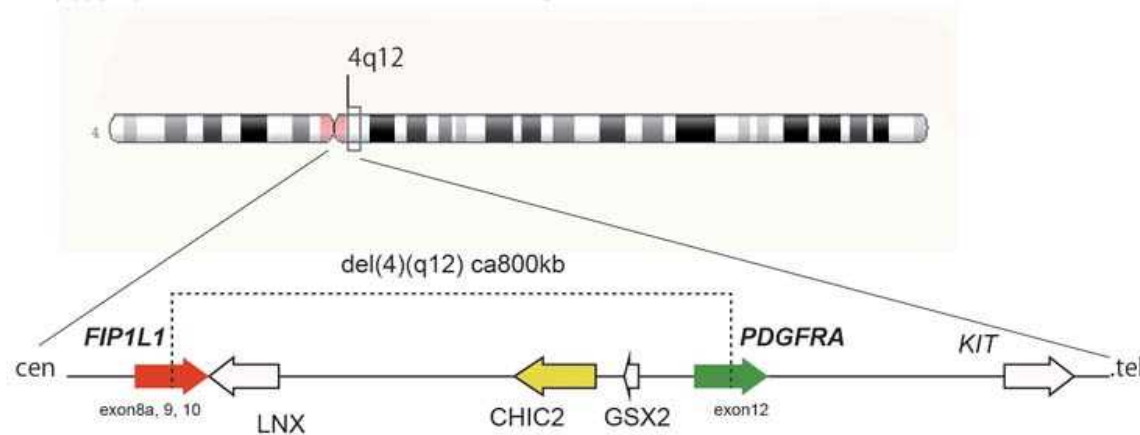
近医受診しAMLが疑われ、セカンドオピニオンを希望し、
精査加療目的に入院。

検査所見

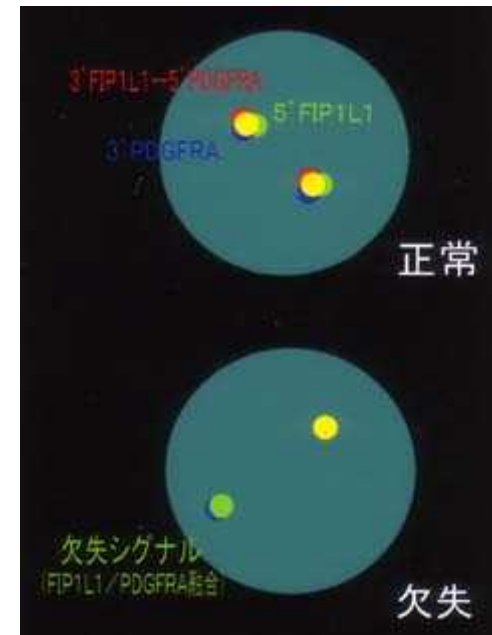
2017/10/05 骨髄

FISH法でFIP1L1-PDGFR α 融合遺伝子が見つかる

del(4)(q12)– FIP1L1-PDGFR α fusion transcript



<http://www.ft-patho.net/index.php?plugin=attach&refer=eosinophilia%A4%AC%C3%F8%CC%C0%A4%CA%B9%FC%BF%F1%C1%FD%BF%A3%C0%AD%BC%F0%E1%E7&openfile=FIP1L1-PDGFR-CEL.jpg>



<http://www.ft-patho.net/index.php?plugin=attach&refer=eosinophilia%A4%AC%C3%F8%CC%C0%A4%CA%B9%FC%BF%F1%C1%FD%BF%A3%C0%AD%BC%F0%E1%E7&openfile=SRL-del4q12.jpg>

ゲノム異常はこれだけか？
他に「悪たれ」はないのか？

⇒ 全ゲノムシーケンスでみるしかない！

全ゲノムシーケンスをすると

- AML症例の全ゲノムシーケンスを行い**7488**個の体細胞変異と**108**個の構造異常を検出。



Genomon —The Zen of Cancer Genome

Sequence Analysis

Watson for Genomicsが全部見ると！

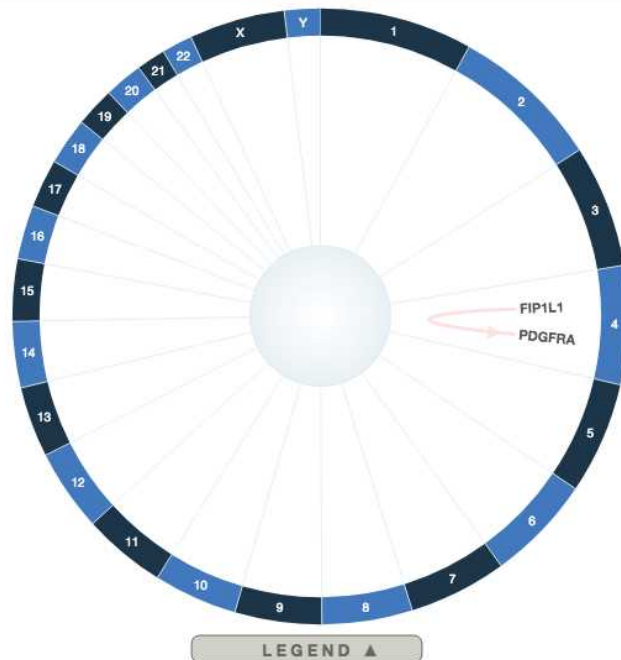
Diagnosis: Acute Myeloid Leukemia
Age: Unspecified Gender: Unspecified

Show: All Level of Evidence

PDF Report

Summary Drugs Molecular Profile

Include all in PDF Report



PDGFRA FIP1L1-PDGFRA



PDGFRA is a tyrosine-protein kinase that plays an essential role in the regulation of embryonic development, cell proliferation, survival and chemotaxis. It is important for the differentiation of bone marrow-derived mesenchymal stem cells and it required for the development of the skeleton, the mucosa lining in the GI tract and of intestinal villi. Depending on the context, PDGFRA signaling either promotes or inhibits cell proliferation and cell migration. In addition, the receptor is also involved in wound healing through cell migration and it plays a role in platelet activation. Somatic mutations in the PDGFRA gene are present in approximately 35% of GIST (PubMed: [12522257](#)). The protein is amplified and mutated in 15% glioblastoma (PubMed: [24120142](#)) and high level of expression defines the proneural subtype of this tumor (PubMed: [20129251](#)).

FIP1L1-PDGFRA [See all notes](#)

The PDGFRA fusions can transform hematopoietic cells and activate downstream signaling activity (PubMed: [12660384](#)). Cells expressing FIP1L1-PDGFRA induced leukemia in vivo (PubMed: [21224473](#)). FIP1L1-PDGFRA causes a truncation of the juxtamembrane domain, which is thought to remove the autoinhibitory activity of the kinase (PubMed: [16690743](#)).

やはり、それしかなかった！
(このWatson解析は10分)

まとめ

- このうち4番染色体長腕の欠失による*FIP1L1-PDGFR α* 融合遺伝子を白血化に寄与するdriverの構造異常として検出。
(4q12領域に約48.76MBの欠失が観察された)
- *FIP1L1*はintron13にbreak pointがあり*PDGFR α* はexon12にbreakpointがある(*Leukemia* (2009) **23**, 271–27).
- 上記融合遺伝子に対してはimatinibなどのTKIが有効と考えられる。(注)ImatinibはAMLの適応を有していない。

パネル解析と全ゲノムシーケンス解析 あなたならどちらを選びますか？ (2015年の知見、もう大昔の話ですが)

Laskin et al. (2015) Lessons learned from the application of whole-genome analysis to the treatment of patients with advanced cancers. Cold Spring Harb Mol Case Stud 1:a000570. doi: 10.1101/mcs.a000570

Canada British Columbiaの
グループの2012～2014の研究結果

<http://molecularcasestudies.cshlp.org/content/1/1/a000570.full>

【要旨】

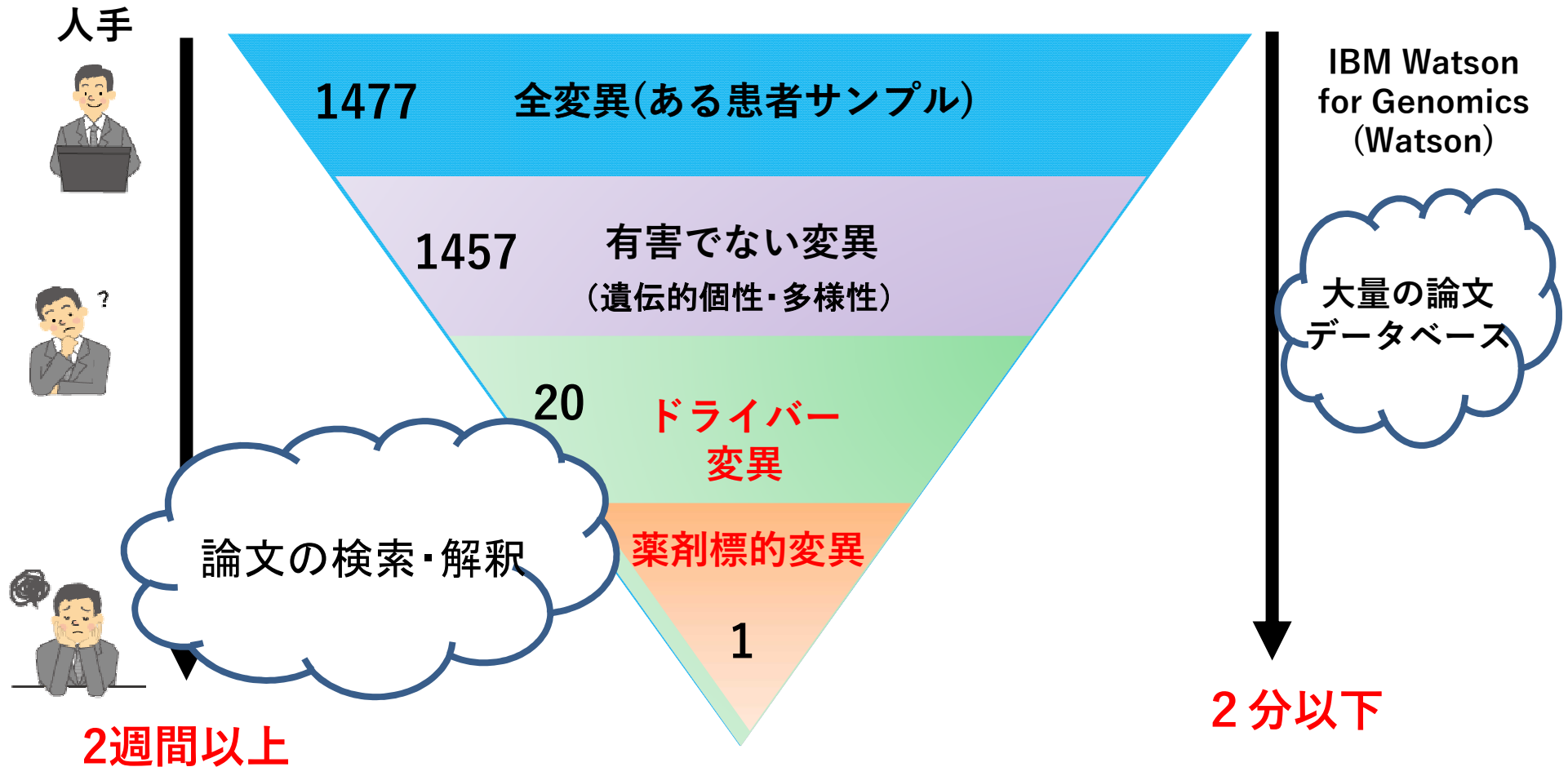
- 末期のがん患者群に対して、全ゲノムシーケンスを用いた場合、78人中、55人においてactionableな変異が見つかり、そのうち23人が実際にそれに基づく治療を受けた。
- 一方、同じ患者群に対してパネルをつかった場合、81人中、73%にinformativeだけれどもactionable（治療には使えない）ではない変異が見つかり、それらの変異のmajority（55%）はTP53のものだった。また23%の患者群は、パネルではなにも見つからなかった。パネルでは、結局actionにつながるものが見つからなかった。

ここでのactionable および informativeの定義は以下のようになっています。

“informative” : “identification of an interesting feature that may not have prognostic or therapeutic relevance at this time”

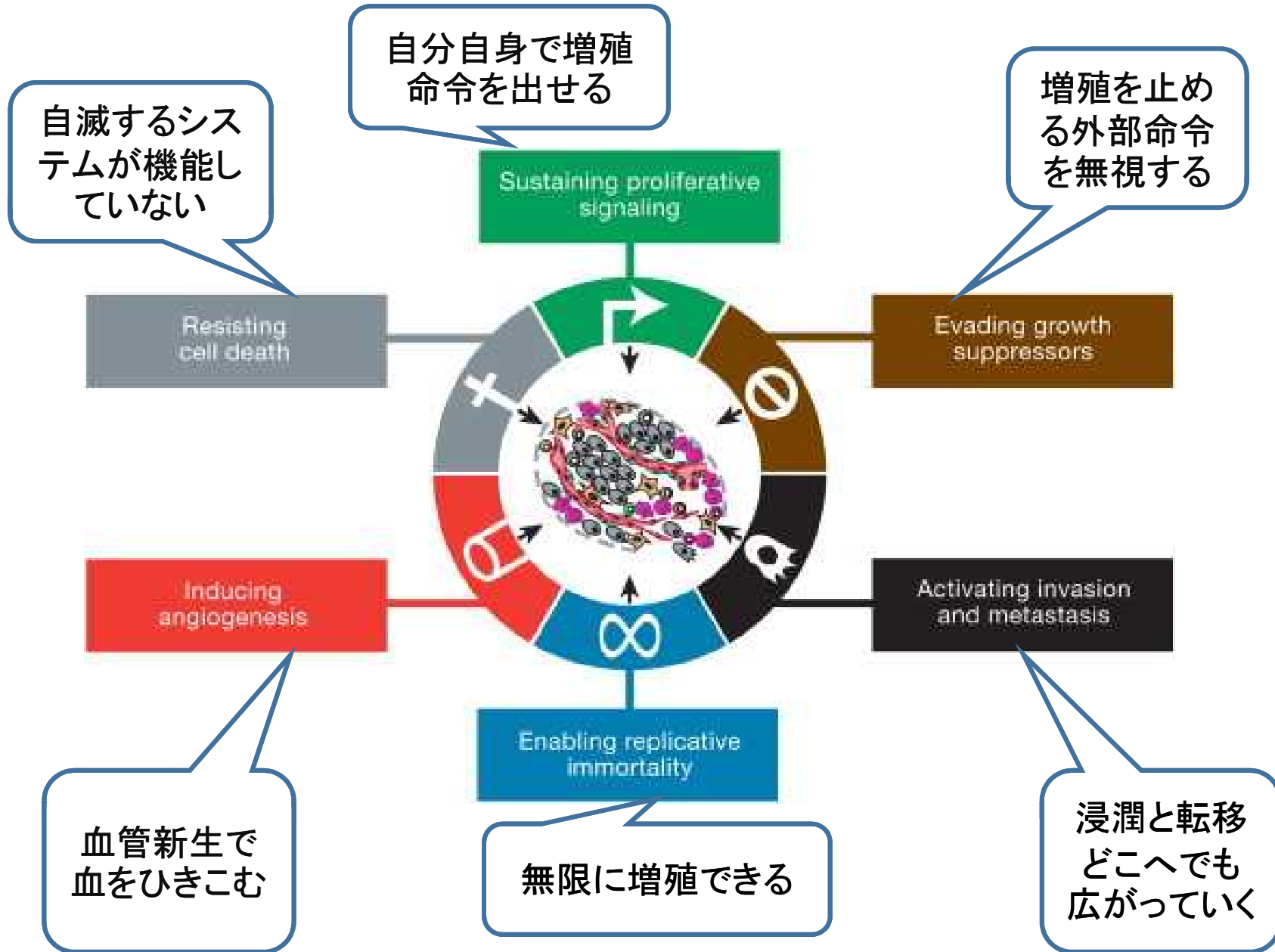
“actionable” : “identification of a potential target or risk factor that affects the treatment plan.”

人の目で変異の整理・解釈（キュレーション）は大変



医科研血液腫瘍内科の
専門医による自験例

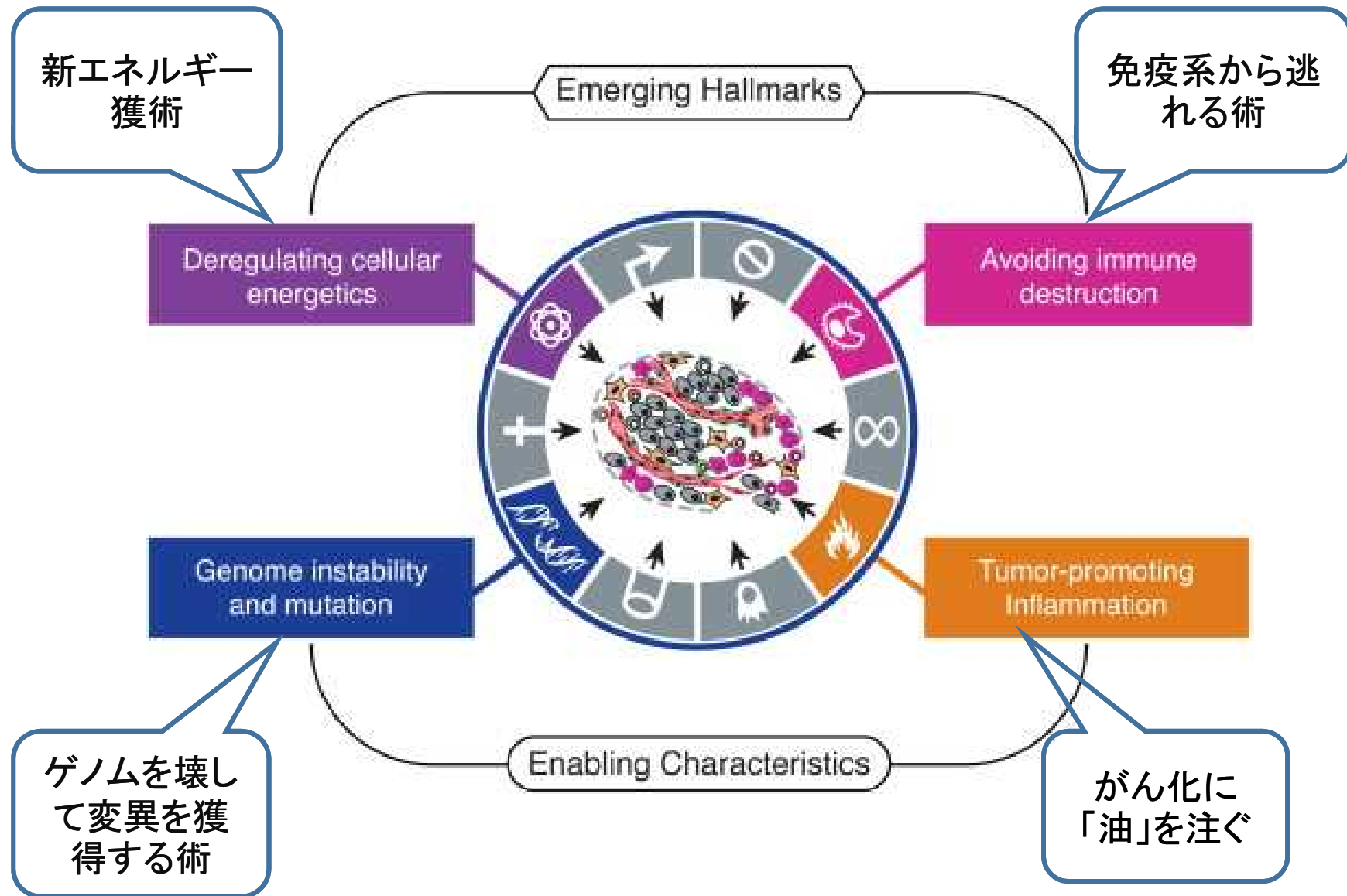
「がん」ってなに？



[Hallmarks of cancer: the next generation.](#)

Hanahan D, Weinberg RA.

Cell. 2011 Mar 4;144(5):646-74. doi: 10.1016/j.cell.2011.02.013. Review.



[Hallmarks of cancer: the next generation.](#)

Hanahan D, Weinberg RA.

Cell. 2011 Mar 4;144(5):646-74. doi: 10.1016/j.cell.2011.02.013. Review.