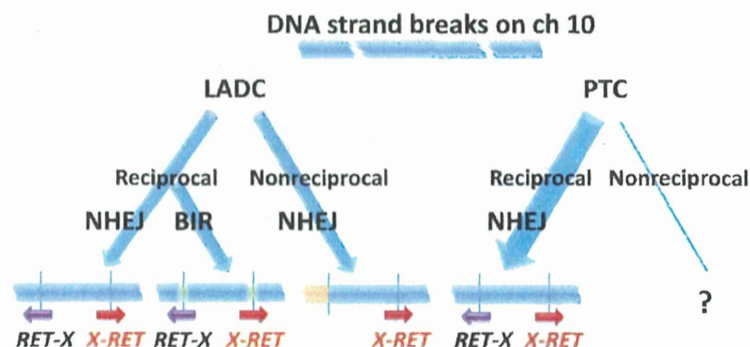


FIGURE 4. Molecular processes underlying *RET* gene fusions in LADC and PTC. Different processes are involved in *RET* fusion in different tumor types. Both reciprocal and nonreciprocal inversions occur in LADC. In LADC, BIR and NHEJ are responsible for DNA end joining in reciprocal inversion, whereas NHEJ is exclusively involved in nonreciprocal inversion. In PTC, reciprocal inversion by NHEJ is dominant. LADC, lung adenocarcinoma; PTC, papillary thyroid carcinoma; NHEJ, nonhomologous end joining; BIR, break-induced replication.



DSBs, which occurs in the context of irradiation, may increase the likelihood of illegitimate repair generating *RET* fusion. On the other hand, in LADC, both DSBs and single-strand breaks formed by multiple causes might trigger rearrangements by multiple DNA repair pathways. The high frequency of nonreciprocal inversion also distinguishes LADC from PTC, for previous study revealed that *RET* fusions result from reciprocal inversion in most cases (43 of 47, 91%).^{14,15} Frequent nonreciprocal inversion is consistent with the observation that *KIF5B-RET* fusion-positive tumors contain deletions of the 5' part of *RET*, as revealed by FISH staining patterns.¹ The present study provides a molecular basis for such a distinct FISH finding and will help to define the criteria used to diagnose *RET*-fusion-positive LADC. Interestingly, FISH analysis also revealed that another driver mutation, *EML4-ALK* fusion, in LADC, caused by a paracentric inversion of chromosome 2, also involves deletion of the 5' region of the *ALK* oncogene locus.^{30,31} Although the structures of breakpoint junctions of *ALK* fusions have not been characterized to the best of our knowledge, these results indicate that a significant fraction of chromosome inversions that cause oncogenic fusions in lung cancer are likely to be nonreciprocal.

Finally, a few issues remain to be elucidated regarding the molecular processes generating oncogenic *RET* fusions. First, although this and previous PTC studies imply that the 2.0-kb region spanning the *RET* exon 11 to intron 11 region is susceptible to DNA strand breaks, the underlying mechanisms remain unknown. For, this region does not exhibit distinctive features known to make DNA susceptible to breaks (Supplementary Fig. 3, Supplementary Digital Content 3, <http://links.lww.com/JTO/A543>; details in Supplementary Notes, Supplementary Digital Content 2, <http://links.lww.com/JTO/A542>). Second, the etiological factors that cause DNA strand breaks, and the factors that determine reciprocal or nonreciprocal inversion and selection of DNA repair pathways, also remain unknown. The mode of joining and breakpoint distribution was irrespective of smoking history, and therefore, DNA damage due to smoking is unlikely to be an important factor. The fact that *RET* fusions are more frequent in LADC of never-smokers than in that of ever-smokers indicates that undefined etiological factors play major roles in the occurrence of *RET* fusions.

ACKNOWLEDGMENTS

We thank Hiromi Nakamura, Isao Kurosaka, Sumiko Ohnami, and Sachiyo Mitani of National Cancer Center

(NCC) Research Institute for data analysis and technical assistance. The NCC Biobank is supported by the NCC Research and Development Fund of Japan. SNP array analysis was performed by the genome core facility of the NCC. This study was supported in part by Grants-in-Aid for Scientific Research on Innovative Areas (22131006), from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, for the Third-term Comprehensive 10-year Strategy for Cancer Control, from the Ministry of Health, Labor, and Welfare, and for the Program for Promotion of Fundamental Studies in Health Sciences, from the National Institute of Biomedical Innovation (NiBio), and by Management Expenses Grants from the Government to the NCC.

REFERENCES

1. Takeuchi K, Soda M, Togashi Y, et al. *RET*, *ROS1* and *ALK* fusions in lung cancer. *Nat Med* 2012;18:378–381.
2. Lipson D, Capelletti M, Yelensky R, et al. Identification of new *ALK* and *RET* gene fusions from colorectal and lung cancer biopsies. *Nat Med* 2012;18:382–384.
3. Kohno T, Ichikawa H, Totoki Y, et al. *KIF5B-RET* fusions in lung adenocarcinoma. *Nat Med* 2012;18:375–377.
4. Ju YS, Lee WC, Shin JY, et al. A transforming *KIF5B* and *RET* gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. *Genome Res* 2012;22:436–445.
5. Gautschi O, Zander T, Keller FA, et al. A patient with lung adenocarcinoma and *RET* fusion treated with vandetanib. *J Thorac Oncol* 2013;8:e43–e44.
6. Drilon A, Wang L, Hasanovic A, et al. Response to Cabozantinib in patients with *RET* fusion-positive lung adenocarcinomas. *Cancer Discov* 2013;3:630–635.
7. Kohno T, Tsuta K, Tsuchihara K, Nakaoku T, Yoh K, Goto K. *RET* fusion gene: translation to personalized lung cancer therapy. *Cancer Sci* 2013;104:1396–1400.
8. Shaw AT, Hsu PP, Awad MM, Engelman JA. Tyrosine kinase gene rearrangements in epithelial malignancies. *Nat Rev Cancer* 2013;13:772–787.
9. Wang R, Hu H, Pan Y, et al. *RET* fusions define a unique molecular and clinicopathologic subtype of non-small-cell lung cancer. *J Clin Oncol* 2012;30:4352–4359.
10. Suehara Y, Arcila M, Wang L, et al. Identification of *KIF5B-RET* and *GOPC-ROS1* fusions in lung adenocarcinomas through a comprehensive mRNA-based screen for tyrosine kinase fusions. *Clin Cancer Res* 2012;18:6599–6608.
11. Yang L, Luquette LJ, Gehlenborg N, et al. Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell* 2013;153:919–929.
12. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. *Pathogenetics* 2008;1:4.
13. Kohno T, Yokota J. Molecular processes of chromosome 9p21 deletions causing inactivation of the p16 tumor suppressor gene in human cancer: deduction from structural analysis of breakpoints for deletions. *DNA Repair (Amst)* 2006;5:1273–1281.

14. Nikiforov YE, Koshoffer A, Nikiforova M, Stringer J, Fagin JA. Chromosomal breakpoint positions suggest a direct role for radiation in inducing illegitimate recombination between the ELET1 and RET genes in radiation-induced thyroid carcinomas. *Oncogene* 1999;18:6330–6334.
15. Bongarzone I, Butti MG, Fugazzola L, et al. Comparison of the breakpoint regions of ELET1 and RET genes involved in the generation of RET/PTC3 oncogene in sporadic and in radiation-associated papillary thyroid carcinomas. *Genomics* 1997;42:252–259.
16. Minoletti F, Butti MG, Coronelli S, et al. The two genes generating RET/PTC3 are localized in chromosomal band 10q11.2. *Genes Chromosomes Cancer* 1994;11:51–57.
17. Klugbauer S, Pfeiffer P, Gassenhuber H, Beimfohr C, Rabes HM. RET rearrangements in radiation-induced papillary thyroid carcinomas: high prevalence of topoisomerase I sites at breakpoints and microhomology-mediated end joining in ELET1 and RET chimeric genes. *Genomics* 2001;73:149–160.
18. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;26:589–595.
19. Cai W, Su C, Li X, et al. KIF5B-RET fusions in Chinese patients with non-small cell lung cancer. *Cancer* 2013;119:1486–1494.
20. Yokota K, Sasaki H, Okuda K, et al. KIF5B/RET fusion gene in surgically-treated adenocarcinoma of the lung. *Oncol Rep* 2012;28:1187–1192.
21. Suzuki M, Makinoshima H, Matsumoto S, et al. Identification of a lung adenocarcinoma cell line with CCDC6-RET fusion gene and the effect of RET inhibitors in vitro and in vivo. *Cancer Sci* 2013;104:896–903.
22. Mahaney BL, Meek K, Lees-Miller SP. Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *Biochem J* 2009;417:639–650.
23. Lieber MR. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* 2010;79:181–211.
24. Bennardo N, Cheng A, Huang N, Stark JM. Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. *PLoS Genet* 2008;4:e1000110.
25. Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 2007;131:1235–1247.
26. Sasaki S, Kitagawa Y, Sekido Y, et al. Molecular processes of chromosome 9p21 deletions in human cancers. *Oncogene* 2003;22:3792–3798.
27. Matsumoto S, Iwakawa R, Takahashi K, et al. Prevalence and specificity of LKB1 genetic alterations in lung cancers. *Oncogene* 2007;26:5911–5918.
28. Lawson AR, Hindley GF, Forshew T, et al. RAF gene fusion breakpoints in pediatric brain tumors are characterized by significant enrichment of sequence microhomology. *Genome Res* 2011;21:505–514.
29. Hamatani K, Eguchi H, Ito R, et al. RET/PTC rearrangements preferentially occurred in papillary thyroid cancer among atomic bomb survivors exposed to high radiation dose. *Cancer Res* 2008;68:7176–7182.
30. Dai Z, Kelly JC, Meloni-Ehrig A, et al. Incidence and patterns of ALK FISH abnormalities seen in a large unselected series of lung carcinomas. *Mol Cytogenet* 2012;5:44.
31. Yoshida A, Tsuta K, Nitta H, et al. Bright-field dual-color chromogenic in situ hybridization for diagnosing echinoderm microtubule-associated protein-like 4-anaplastic lymphoma kinase-positive lung adenocarcinomas. *J Thorac Oncol* 2011;6:1677–1686.

一細胞ゲノム解析

Single-cell genomics of cancer



加藤 護

Mamoru Kato

国立がん研究センター研究所バイオインフォマティクス部門

◎腫瘍内不均質性は、がん細胞がゲノム・エピゲノム変化を経た進化の結果発生し、その程度は抗がん剤耐性や予後のような臨床像とも関係すると考えられている。これら不均質性やがん細胞進化を、その最小単位である単一細胞レベルで調べるため、近年、一細胞シーケンス技術が開発された。この技術は flow cytometry などの方法で細胞を分離し、次世代シーケンサーを使って細胞ひとつひとつのゲノムを配列決定するという技術である。一細胞シーケンスは技術的にまだ課題が多いが、組織をそのまま読む通常のシーケンスではわからないがん細胞の分集団を発見し、これまで不明であったがん細胞進化の詳細を明らかにしつつある。この技術の臨床応用として血中循環腫瘍細胞の変異変化をタイピングし、患者内のがんの状態をモニタリングする研究もはじまっている。



Key Word : 腫瘍内不均質性, がん細胞進化, 一細胞シーケンス

がん組織に腫瘍内不均質性 (intra-tumor heterogeneity) があることは古くから知られていたが、がんゲノム研究がこの不均質性に本格的に取り組みはじめたのはつい最近のことである。次世代シーケンサーと組み合わせられた一細胞シーケンス技術は、この不均質性をもっとも明確に解明できる技術として注目されている。

一細胞シーケンス技術は、多細胞生物において生殖細胞 (精子, 卵細胞)^{1,2)} や神経細胞³⁾ にも適用されているが、本稿ではがん細胞への適用を取り上げる。また、広義には RNA やエピゲノムのシーケンスも含むが、ここではおもにゲノム DNA シーケンスに着目する。

● 腫瘍内不均質性

採取されたがん組織はけっして均質ではない。さまざまな種類の間質細胞を含むこともあるし、がん細胞間でも異なる病理像を呈し、異なるマーカーを発現することもある。病変組織をマウスに異種移植しても、増殖を開始できる (能力をもった) がん細胞もあれば、そうでない細胞もある。こ

のような腫瘍内不均質性は診断や治療にも影響を与えると考えられる。たとえば、検査時において悪性度の高い細胞や化学療法に耐性がありそうな細胞の小集団が他細胞の大集団のなかに埋没している場合、不均質性を考慮しない診断では精度が落ちるであろう。ほかにも腫瘍の不均質性 (多様性) の度合いが高いと腫瘍が悪性化しやすいという報告もある⁴⁾。このように腫瘍内不均質性は生物学的な追求の対象としてだけではなく、臨床医学的にも重要である (本特集・谷内田「腫瘍内多様性とがんゲノム進化」の稿も参照)。

このようながん細胞の不均質性のうち細胞の有糸分裂による継承がかかわる不均質性に関しては、それが生じる理由として、①細胞間の分化階層構造、②がん細胞ゲノムの進化、が考えられている。①はいわゆるがん幹細胞に関連する概念であり、大ざっぱには (ゲノムが同じであったとしても) エピゲノム的な状態の違い、と言い換えることができる。②はがん細胞のクローン性増殖 (clonal expansion)⁵⁾ のことであり、腫瘍進展においてさまざまな体細胞変異を得たがん細胞が生き

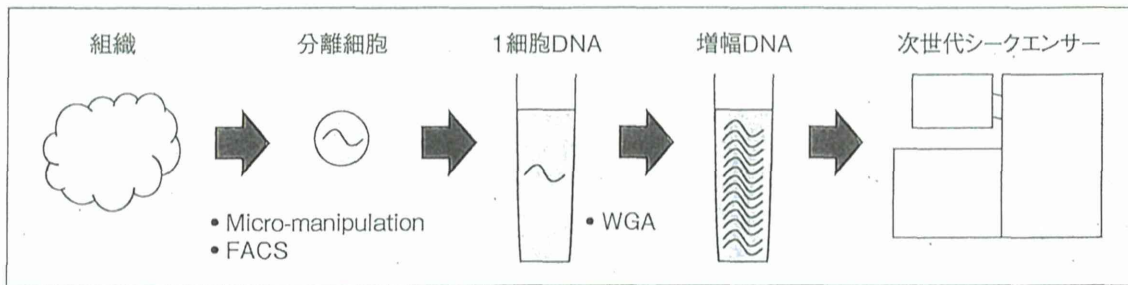


図1 一細胞シーケンスの原理

残り、多様性(不均質性)が生まれるというダーウィン進化にたとえられる過程である。組織像における複雑な不均質性をこれらの描像によって捨象するとすれば、不均質性は「腫瘍進展に伴うゲノム・エピゲノム状態の時間変化の結果現れる、それら状態の細胞間での違い」ということもできる。

不均質性を解明するシーケンズ技術

ゲノム分野で不均質性を研究するおもな手法には2つある。ひとつは組織から細胞塊のままDNAを抽出して次世代シーケンサーでシーケンズし(バルクシーケンズ), リードに含まれる変異文字の割合を数え, それを変異アレル頻度として分析することである。たとえば, 組織にがん細胞の分集団Aが60%, がん細胞の分集団Bが30%存在し, Aに特異的な変異アレルaが座位1に, Bに特異的な変異bが座位2に存在するとしよう。その腫瘍組織をシーケンズすると, 変異aが座位1に30%(通常ヘテロ接合での変異が期待されるため, 60/2), bが座位2に15%検出されることが期待される。逆にもともとどのような分集団が存在するのかわかっていない場合, これらのデータから分集団が2つ存在し, それぞれ組織中60%, 30%を占めていると推測することができる。2つ目の手法は一細胞シーケンズである。ここでは細胞の塊をばらばらにして一細胞にし, 一細胞のDNAを増やして(whole genome amplification: WGA), 次世代シーケンサーにかけられるほど十分な量を得, シーケンズして細胞のゲノムをひとつずつ分析する(図1)。不均質性の最小単位を分析するので, 論理は明快である。

これら2つの手法を比較するとさまざまな点で

一長一短があるが, 一細胞シーケンズでの利点はバルクシーケンズでは座位1と座位2での関連がわからないのに対し, 一細胞シーケンズではわかるという点である(図2)。たとえば, バルクでは座位1で20%, 座位2で20%の頻度があった場合, 通常40%の分集団がひとつ存在すると推測してしまいが, 実はこれら座位間には関連がなく, 2つの分集団がそれぞれ40%存在しているのかもしれない。実験エラーやコピー数変化によって頻度がずれたり, あるいは分集団間の頻度が近接していたりする場合を考えると, 実データでこのような例はけっしてtrivialなケースではないことがわかる。また, この違いは生物学的に異なる解釈を導いてしまう。前者のケースでは単一集団がクロナルに存在しているだけであるが, 後者のケースでは2集団が協調的に存在している可能性が出てくる。

また, 一細胞シーケンズでは低頻度のコピー数変化(copy number alteration: CNA)をより高感度に検出できるという利点もある。CNA検出は次世代シーケンサーを使っても一般にノイズが多く, バルクシーケンズの場合, CNAが細胞集団内で低頻度であるとノイズに埋もれやすい。

一細胞シーケンズにはこのような利点があるが, しかしコストと(後述するように)テクニカルな問題がある。一方, バルクシーケンズはdepthを深く読めばよいだけであり, 条件によってはコストを大幅に下げることができる。分集団の構造を追及せず, 単に点変異の有無を調べるには, バルクのほうが適している可能性もある。

一細胞シーケンズ技術

一細胞シーケンズのプロセスは, 手順①: 組

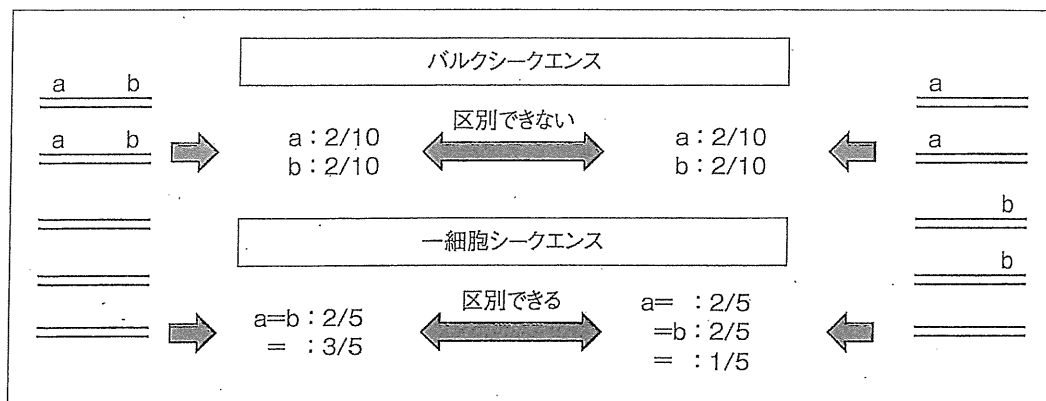


図 2 バルクシーケンスと比較した一細胞シーケンスの利点

バルクシーケンスでの頻度(2/10 など)は DNA コピー数(に比例したリード数)が単位となり、一細胞シーケンスでの頻度は細胞数が単位となる。一細胞シーケンスにおいて“=”は細胞内での連鎖を表す(相, すなわちハプロタイプは決定されない)。

織から細胞を分離し, 手順②: 一細胞の DNA を増やし, 手順③: シーケンスし, 手順④: 情報処理を行うことである。いずれの工程もバルクほどは確立されておらず, 技術的困難が伴う。それは多くの論文⁶⁻⁸⁾が, データクオリティの記述に多くのスペースを費やしていることからうかがえる。

手順①: 組織からの細胞分離

手順①では, 手動による微小操作(連続希釈とマイクロピペッティング)と fluorescence activated cell sorting (FACS) が使われる⁹⁾。固形腫瘍の場合はその前に酵素によって細胞をばらばらにしなければならない。細胞のサンプリングデザインとしては, ランダムサンプリング⁶⁾とマーカー選別によるターゲットサンプリング(*e. g.*, 腫瘍細胞と正常細胞の分離¹⁰⁾)がある⁹⁾。微小操作は比較的簡単に行えるが, 自動化ができず, サンプル汚染にも注意しなければならない^{8,9)}。FACS はある程度自動化できるが, 操作に習熟が必要である。最近では微小流路を利用して細胞分離から核酸増幅まで行う C1 (Fluidigm 社) という一括システムが登場している。これまではトランスクリプトームに対応していたが, 執筆時点(2014年2月)でゲノムにも対応予定とのことである。自動化されており, 閉鎖系のためサンプル汚染の可能性も少ないと考えられる。

手順②: 一細胞の DNA を増やす

手順②の WGA には, PCR ベースの WGA⁶⁾, multiple displacement amplification

(MDA)^{7,10,11)}, そして最近開発された MALBAC⁸⁾ という手法が使われる。WGA で問題になるのはゲノムが全域にわたって(かつ, 均質に)増幅されるか, という問題である。出発の量が一細胞の DNA (倍数体) 1 コピーであることから, その難しさが想像できる。CNA を同定する場合は Navin ら⁶⁾ が示したように, ゲノムの 6% 程度が増幅されていれば解像度 50k bps 程度の CNA が得られるので, (均質に増幅されてさえいけば) 増えない領域があっても大きな問題とはならないかもしれない。しかし, 重要遺伝子を SNV を通して発見・分析しようとする場合は, 増えない領域の SNV を見落とす可能性があるので問題となる。これは allelic dropout (ADO) と呼ばれ, Wang らのグループ^{7,10,11)} や Zong らはこの率を, (おもに正常細胞の) 同じサンプルに対し, バルクでヘテロであるにもかかわらず, シングルでホモとなった座位の数によって近似的に評価している。それによれば, ADO 率は(エキソームシーケンスにおいて) 40~43%^{7,10)} である。Zong ら⁸⁾ によれば, MDA は 65%, 彼らが開発した MALBAC では 1% のことである。

手順③: シーケンス

手順③は通常の方法ととくに変わりはない。

手順④: 情報処理

手順④では, Narvin ら⁶⁾ はマッパーに Bowtie を使い, CNA 検出に独自の Varbin¹²⁾ というアルゴリズムを使っている。CNA 検出の原理は depth 分析である(増幅領域には多くのリードがマップ

表 1 一細胞シーケンスの代表的研究

著者	出版年	腫瘍	細胞数*1	細胞の分離	WGA 法	対象変異	対象範囲	カバレッジ
Navin ら	2011	乳がん	200	FACS	PCR-based	CNA	WG	6% (<1x)
Hou ら	2012	血液腫瘍	90→58	Micro-manipulation	MDA	SNV	WE	≥5x が ≥70%
Hu ら	2012	腎がん	25→22	Hou らに準じる	Hou らに準じる	SNV	WE	≥5x が ≥80%
Li ら	2012	膀胱がん	66→55	Hou らに準じる	Hou らに準じる	SNV	WE	≥5x が ≥60%
Zong ら	2012	(大腸がん細胞株*2)	13→5	Micro-manipulation	MALBAC	SNV, CNA	WG	Mean 25x

*1: シークエンス成功の歩留まりが報告あるいは推察されるものには矢印を付けた。Zong ら以外では最終的にデータとして使われた細胞数、Zong らの場合、平均 depth 数が約 25x 以上のものを選んだ。

*2: 技術開発研究のため、()でくくった。

WG: whole genome, WE: whole exome.

される。欠失領域には少ないリード)。整数コピー数の同定も行っており、ガウシアン・カーネル密度推定法を使って depth を分析している。SNV に関して Wang のグループ^{7,10,11)}では、彼らが開発している SOAP を中心にして分析を行っている。Zong らも含め、通常のコールと違う点は複数細胞(Li ら¹⁰⁾、Zong ら⁸⁾では 3 細胞)で同じ変異が検出されることをコールの条件としていることである。SNV のエラー率(誤検出率)は ADO 率とは逆に、バルクでホモなのにシングルではヘテロ、という原理で近似的に評価している。それによれば、 $10^{-4} \sim 10^{-5}$ 程度^{7,10,11)}、MALBAC では 0(計算母数以下)で⁸⁾、きわめて低い。変異頻度の計算に関し、Hou らと Li らの研究においてはコールされた変異の頻度を細胞数から直接計算するのではなく、細胞ごとのリード数を考慮するベイズ推定によって計算している(手法は文献¹³⁾参照)。これは各細胞において、depth が十分に取れない状況に対処するためである。変異が検出された後の分析においては、標準的な分析法(系統樹分析、主成分分析、変異頻度分析)以外に現在のところとくに深い分析手法はみあたらないが、バルクと違っていくつか注意しなければならない点がある。ひとつは細胞間に系統関係があるため、細胞間で変異が関連していることである。つまり、バルクにおいて 100 症例で頻度 40% であることと、シングルにおいて 100 細胞で頻度 40% であることとは意味が違う。前者は変異イベントが 40 回起こったことを表すが、後者は共通祖先に変異イベントがたった 1 回起こって子孫細胞に受け継がれたことを表す。このため種内変異を扱う集団遺伝学のア

イデアを取り入れた分析手法や解釈を考えに入れなければならない。また、ADO による避けがたい false negative にも注意しなければならない。

一細胞シーケンスの代表的研究

本格的な一細胞シーケンスがはじまったのは、2011 年の Navin らの研究からである⁶⁾(表 1)。彼らは FACS で細胞を分離し、PCR ベースの WGA 法で DNA を増幅し、Illumina 社のシーケンサーで、乳がん臨床サンプル 2 症例から 200 細胞をシーケンスした。と同時に、2 つの細胞株から 14 細胞をシーケンスし、バルクシーケンスと比べてデータ品質も調べている。彼らが注目したのは CNA で、このためとくに深い depth は必要がなかった。シーケンスカバレッジは 6% 程度であるが、全ゲノム領域が対象である。研究デザインでひとつ注目すべきは、細胞がランダムサンプリングされていることである。よって一見腫瘍と無関係かもしれない細胞が、場合によっては 60% もシーケンスされている。しかしこのため、組織学的に推定された腫瘍細胞率との比較が可能になったり、間質細胞に埋もれた pseudo-diploid という新しい細胞分類が発見されたりしている。また彼らは、細胞集団内では均質であるが集団間としては異なるというがんの不均質性を発見した。このことから、がん細胞は漸次的な進化をするのではなく、古生物学でいう断続平衡進化(まったく変化のない期間と突発的に変化する期間を交互に繰り返す進化)をすることを提唱した。

続く 2012 年には Hou らが、1 症例 90 細胞の一

細胞エクソームシーケンシングを行っている⁷⁾。対象は細胞分離が容易な血液腫瘍である。この研究では正常細胞株からの2細胞に対し whole genome の一細胞シーケンセスも行われ、技術的側面とデータ品質についても多く記述されている。Depth の低さ(～WGA 非増幅)と GC-content との間には相関があったが、リピートや染色体位置との相関はなかったようである。90細胞をシーケンセスしたが、データ品質チェックによって最終的に使われた細胞数は58で、腫瘍50細胞中712 SNV が発見された。変異の集団頻度に関し、細胞数から計算された変異頻度とバルクでリード数から計算された変異頻度はよい相関を示した($R^2=0.75$)。がん細胞進化過程での自然選択の痕跡、および分集団はとくに発見されなかった。

Xu ら¹¹⁾は Hou らの方法を踏襲し、腎がん1症例17細胞中260個のSNVを発見したが、分集団は発見されなかった。

Li ら¹⁰⁾は(改良された)Hou らの方法を用いて、膀胱がんの1症例66細胞をシーケンセスし、さらに99症例をバルクでシーケンセスして、それぞれで検出された変異を比較した。データ品質チェックを通った44細胞に3つの分集団と443個の変異を検出し、*NIPBL* や *CFTR* といった新規ドライバー遺伝子とその3集団共通にみられることを発見した。さらに、変異頻度の分析によって non-synonymous 変異が synonymous 変異に比べいくぶん高頻度にずれていることが示された。一細胞分析の場合、進化的に中立であっても遺伝的連鎖(有糸分裂による継承の際の染色体連鎖)があるため non-synonymous のほうが高い理由はなく、これを説明するためがん細胞には漸次的な自然選択がかかっていることを提唱した(一見すると Navin らの説と相反するが、とくに詳しくは述べられていない)。

● おわりに

一細胞シーケンセスはバルクシーケンセスに比べまだ技術的に確立されておらず、コストもかかる。しかし、腫瘍内不均質性の最小単位を分析するため、不均質性とそれに付随するがん細胞進化

にもっとも明確なデータを与える。集団“内”では均質であるが、集団“間”では異質である分集団の発見や、断続平衡進化・漸次進化の提唱にみられるように、腫瘍内不均質性や腫瘍進展の解明を強力に促進する。これまでの研究は一時刻点のスナップショットデータであったが、今後多時刻点でのデータが得られれば、その最小単位の明快なダイナミクスが明らかにされるであろう。また、それに従ってより強力な情報解析手法も開発されねばならない。

著者は、国立がんセンターの柴田龍弘博士、新井康仁博士、筆宝義隆博士らの協力を得、多時刻点の一細胞シーケンセスや、集団遺伝学の考えを取り入れた新しい分析手法の研究を行っている。一細胞シーケンセス技術の臨床応用としては circulated tumor cell (CTC) への応用があり、CTC 検出システムで血中から分離された腫瘍細胞をカウントするだけでなく、次世代シーケンサーを用いて細胞の変異検出を行ってモニタリングや治療効果因子の同定に役立てようとする研究が行われている^{14,15)}。アメリカ NIH は一細胞研究の重要性を認識し、一細胞シーケンセスを含む単一細胞研究に5年間で約90億円(US\$ 90 million)の助成を2012年に公表しており¹⁶⁾、Single Cell Analysis Program として助成が開始されている。

文献

- 1) Hou, Y. et al. : *Cell*, **155** : 1492-1506, 2013.
- 2) Lu, S. et al. : *Science*, **338** : 1627-1630, 2012.
- 3) Evrony, G. D. et al. : *Cell*, **151** : 483-496, 2012.
- 4) Maley, C. C. et al. : *Nat. Genet.*, **38** : 468-473, 2006.
- 5) Nowell, P. C. : *Science*, **194** : 23-28, 1976.
- 6) Navin, N. et al. : *Nature*, **472** : 90-94, 2011.
- 7) Hou, Y. et al. : *Cell*, **148** : 873-885, 2012.
- 8) Zong, C. et al. : *Science*, **338** : 1622-1626, 2012.
- 9) Shapiro, E. et al. : *Nat. Rev. Genet.*, **14** : 618-630, 2013.
- 10) Li, Y. et al. : *Gigascience*, **1** : 12, 2012.
- 11) Xu, X. et al. : *Cell*, **148** : 886-895, 2012.
- 12) Baslan, T. et al. : *Nat. Protocols*, **7** : 1024-1041, 2012.
- 13) Yi, X. et al. : *Science*, **329** : 75-78, 2010.
- 14) Ni, X. et al. : *Proc. Natl. Acad. Sci. USA*, **110** : 21083-21088, 2013.
- 15) Heitzer, E. et al. : *Cancer Res.*, **73** : 2965-2975, 2013.
- 16) Owens, B. : *Nature*, **491** : 27-29, 2012.

6. 発がんドライバー変異の同定

David Tamborero, Abel Gonzalez-Perez, Nuria Lopez-Bigas

シーケンシング技術の爆発的進展によって腫瘍ゲノムの特徴がわかり、がんという疾病は臨床でみられるように分子レベルにおいても均質でないことが明らかになってきた。がんの異なるステージで、腫瘍細胞が獲得する表現型を担うさまざまな変異を同定することは、がん化の過程を理解し、新しい治療介入法を開発するために必要なさらなる解析の基礎となる。本稿では、現在の再シーケンシング計画によって生み出される大規模データの分析や、腫瘍形成を促進する変異・遺伝子・パスウェイの同定に役立つ計算機的手法について論ずる。

はじめに

次世代シーケンシング技術によって、腫瘍細胞ゲノムにおける体細胞変化の完全解明への扉が開かれた¹⁾。がん研究は変容しつつある。たとえ腫瘍の種類が同じでも、がんのサンプルは体細胞異常に関し、きわめて不均質であることもわかってきた²⁾。細胞が経験する複製の回数やDNA維持装置の欠陥、環境からの攻撃といった要因が、体細胞変異の変異率とパターンを規定するが³⁾、がん細胞に見出されるほんの一握りの体細胞変異だけが腫瘍形成の原因となる。残りのものは、がんが引き起こすゲノム不安定性の二次的偶

発事象にすぎない⁴⁾。前者と後者（すなわちドライバー変異とパッセンジャー変異）を区別することは、がん生物学研究および新規治療介入法の開発における最重要課題の1つである。

腫瘍ゲノム再シーケンシングが生み出す大量データの解釈は、近年大きな進歩があったとはいえ、いまだ多くの点で大変な仕事である。本稿では、がんドライバー変異の同定を行う最新の計算機的方法を振り返り、その主要な結果を論ずる。ただし本稿では、体細胞の単一塩基変異および短いフレームシフト変異を対象とする。また、エクソームにおける変異を評価する方法にのみ焦点を当てる。非コード領域（non-coding region）における変異の役割は依然として明らかではないが、近い将来には全ゲノムシーケンスのコストが下がり、その分野で飛躍的進展があらうことは十分に予見できる。

【キーワード&略語】

発がんドライバー、体細胞変異、計算機的方法

ICGC: The International Cancer Genome Consortium (国際がんゲノムコンソーシアム)

TCGA: The Cancer Genome Atlas (がんゲノムアトラス)

1 がんに対する変異の影響を同定する

再シーケンシングによって、腫瘍ゲノムにおける

Identification of oncogenic driver mutations

David Tamborero¹⁾/Abel Gonzalez-Perez¹⁾/Nuria Lopez-Bigas^{1) 2)}: Research Unit on Biomedical Informatics, Department of Experimental and Health Sciences, Universitat Pompeu Fabra¹⁾/Institució Catalana de Recerca i Estudis Avançats (ICREA)²⁾ (ポンペウ・ファブラ大学実験健康科学部生物医学情報研究ユニット¹⁾/カタラーナ高等研究所²⁾)

体細胞変異が検出される。その役割を調べる最初のステップは、ゲノム要素という文脈においてそれらを意味づけることである。すなわち、コーディング配列の場合、突然変異をもつタンパク質コード遺伝子を同定し、その影響を調べる。終止変異、フレームシフト変異といったタンパク質産物を切り詰める変異はおそらくタンパク質の不活化を引き起こすだろうし、一方、同義変異はタンパク質機能に対してはるかに穏やかな効果をおよぼす。これら両極端の間にある非同義変異がタンパク質機能に与える影響を調べることが、計算機的手法の主題である(表)⁵⁾。

当初、計算機ツール群は生殖細胞系列変異のために設計されていた。例えば、SIFT⁶⁾、PolyPhen-2⁷⁾、Mutation Assessor⁸⁾は、アミノ酸進化保存の異なる指標を使ってタンパク質傷害の程度を推測した。Condelはこれらのツールを凌駕すべく、それらツールの出力を総合して、独自の標準化スコアの重み付き平均にもとづき、コンセンサス有害スコアを算出した⁹⁾。

最近開発されたパイオインフォマティクスツールでは、腫瘍の体細胞変異を扱うことができる。fathmmは、がんデータでトレーニングされた隠れマルコフモデルの枠組みで配列保存データを用い、ドライバー変異を識別する¹⁰⁾。CHASMは配列保存に加え、特定のアミノ酸置換やタンパク質構造変化予測に関連した特徴変数に対し、既知のドライバー変異と人工的につくったパッセンジャー変異でトレーニングされたRandom Forestアルゴリズムを使用して、識別を行う¹¹⁾。TransFICはこれらツールのスコアを、遺伝子グループ間での機能変異に対する許容性の違いを考慮に入れて改善し、使用する¹²⁾。なお、スプライス部位変異といった特殊な種類の変異の場合には、他の特別なツールが必要となる⁵⁾。

2 正の選択シグナルを検出する

細胞は増殖するうえでの利点をドライバー変異から受け取り、その結果ドライバー変異は腫瘍のクローン進化過程でより広まっていく。一方、パッセンジャー変異はドライバー変異のそばにいただけで増えていく。それゆえ腫瘍形成に関与する遺伝子は、腫瘍サンプルを見渡したとき、正の自然選択と同じシグナルを出しているはずである。正の選択のシグナルをと

らえるいくつかの方法が、ドライバー遺伝子候補の同定に利用されている(図1, 表)。

1) 変異の頻度による方法

最も直感的な方法は、期待されるバックグラウンド変異率よりも遺伝子が頻繁に変異しているかを評価することである(図1A)。このアイデアは、例えばワシントン大学によって開発されたMuSiC¹³⁾や、Broad研究所によって開発されたMutSig¹⁴⁾といったツールに実装されている。バックグラウンド変異率の推定には、ゲノム上特定の位置に起こる体細胞変異の発生確率に影響を与えるゲノムの特徴が考慮される。例えば、遺伝子の長さ、変異の種類、周りの核酸パターンなどである。MutSig(後にMutSigCVと改名)の最新の実装においては、DNA領域の複製タイミングや遺伝子発現レベルといった因子が、変異率に影響を与える共変量として統計的枠組みに組み込まれ、推定精度を増すために使われている¹⁴⁾。これらの方法によって確かに、がんで高頻度に変異している遺伝子の検出には成功している。しかし変異が低頻度のドライバーは、ほとんど検出できていない。腫瘍形成の全体像を理解するには、低頻度ドライバーも検出しなければならないが、そのためにはより正確なバックグラウンド変異率の推定と、より大きなサンプルサイズが必要であろう。

2) 変異の機能的影響による方法

別の、遺伝子の変異負荷に頼らないアプローチは、サンプル集団において各遺伝子の変異が機能に与える影響を評価する方法である(図1B)。ドライバー変異は、コーディング遺伝子においてタンパク質の機能に影響を与えるはずである。それとは反対に、パッセンジャー変異はランダムに分布するはずである。したがって、機能に大きな影響を与える変異を多くもっていれば、それは正の選択をさし示し、これをもってドライバー遺伝子と判断できる。この方法はバックグラウンド変異率の推定に依存しないため、変異の集団頻度と関係しないドライバー遺伝子の検出に向く。OncodriveFM¹⁵⁾はこの方法を実装している。すなわち、OncodriveFMは遺伝子ごとに各変異の機能的影響を示すいくつかの指標を用い、ある基準値からのずれを計算する。

3) 変異の集中による方法

正の選択シグナルの第三は、タンパク質の一次構造の特定の箇所に変異がまとまっていることである(図1C)。

表 発がんドライバー遺伝子を同定するためのツール一覧

方法	説明
SIFT ⁶⁾	ユーザーの定義するデータベースによって類似タンパク質のマルチプル配列アライメントをつくり、アライメントのすべての位置でのすべての可能な置換に関する標準化確率を計算する。この確率にもとづいて、観察された置換が中立か有害かを分類する。
PolyPhen-2 ⁷⁾	有害・中立アミノ酸変化からなる2つのデータセットによってトレーニングされたナイーブベイズ分類器。主に野生型と変異型アミノ酸の性質の比較に関連した8個の配列ベース、3つの構造ベースの特徴変数が、分類器の作製に使われる。
Mutation Assessor ⁸⁾	アミノ酸残基の進化的保存を調べることで、非同義SNVの機能的影響を予測する。機能全体の保存性をバックグラウンドにして相同配列のマルチプルアライメントをクラスタリングし、タンパク質サブファミリーを決定して、その進化的保存度を利用する。
Condel ⁹⁾	Condel (コンセンサス有害スコア) は、非同義SNVによる機能的影響スコアを組み合わせる方法である。有害および中立非同義SNVのデータセットについて各種ツールが算出するスコアの相補的累積分布から得られた値を、各種ツールの結果を組み合わせるための重みとして使う。
fathmm ¹⁰⁾	がんの“病原性的重み”を変異のモデルにおける許容度で表し、重みをもつ保存タンパク質ドメインと相同配列のアライメントを隠れマルコフモデルで表現する。配列保存性を隠れマルコフモデルと組み合わせ、がん体細胞変異が機能におよぼす効果を予測する。
CHASM ¹¹⁾	専門家がまとめたCOSMIC由来のドライバー変異と、ランダムシミュレートで得られたパッセージ変異にもとづいてトレーニングされたRandom Forest分類器。アミノ酸残基の生理化学的性質や、タンパク質・DNAのマルチプルアライメント由来のスコア、領域ベースのアミノ酸配列組成、UniProtKB 特徴表からのタンパク質局所的構造の性質予測および注釈など、86個の多様な特徴変数 (SNVBox データベースで入手可能) を使う。
TransFIC ¹²⁾	TransFIC (がんの機能的影響変換スコア) は、変異がタンパク質機能に与える影響を評価する他の方法が出す機能的影響スコアを、類似タンパク質における機能的影響変異の許容度を考慮に入れて変換する。変換によって、体細胞変異が細胞に与える影響を補正したと解釈する。
MuSiC ¹³⁾	シークエンスカバレッジや遺伝子の長さ、核酸変化の種類といった各観察変異の基準確率を規定する複数の特徴を考慮に入れて、期待されるより多くの変異をもつ遺伝子を同定する。
MutSigCV ¹⁴⁾	高頻度に変異する遺伝子を同定する。各遺伝子で期待される変異荷重に関連した追加データも取り入れて、バックグラウンドモデルを作製する。その共変量の例としては、がん細胞株で観察されたRNAシークエンسから集めた遺伝子発現データ、HeLa細胞株で測定されたDNA複製のタイミングがある。
OncodriveFM ¹⁵⁾	機能的影響をいくつかの指標値の組で計り、大きな影響をもつ変異に富む遺伝子を同定する。
OncodriveCLUST ¹⁶⁾	サイレント変異を使ってつくられた基準モデルから期待される以上にクラスタ化する変異をもつ遺伝子を同定する。
Active Driver ¹⁷⁾	リン酸化部位に集積する傾向のある変異をもつ遺伝子を同定する。
MEMo ¹⁸⁾	与えられたパスウェイがサンプル間で相互排他的な変異を提示するかを調べ、遺伝子を連結してモジュールを見出す。
HotNet ¹⁹⁾	変異頻度や機能的影響スコアといった指標値が集中する遺伝子モジュールを、与えられたパスウェイ上で連結された遺伝子を通して伝搬する熱拡散モデルによってみつけ出す。

単一の非同義変異が機能におよぼす影響を推定する

腫瘍集団におけるドライバー遺伝子を同定する

腫瘍集団におけるドライバー遺伝子のモジュールを同定する

SNV：一塩基変異