

ゲノム医療におけるデータサイエンティストの役割と育成

富田秀太,^{*,a,†} 森田瑞樹,^{b,††} 山下範之,^c 平沢 晃,^d 豊岡伸一^e

Training Medical Staff with Basic Skills for Data Science in Genomic Medicine

Shuta Tomida,^{*,a,†} Mizuki Morita,^{b,††} Noriyuki Yamashita,^c Akira Hirasawa,^d and Shinichi Toyooka^e

^aOkayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences; 2-5-1 Shikata-cho, Kita-ku, Okayama 700-8550, Japan; ^bDepartment of Biorepository Research and Networking, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences; 2-5-1 Shikata-cho, Kita-ku, Okayama 700-8550, Japan;

^cCenter for Innovative Clinical Medicine, Okayama University Hospital; 2-5-1 Shikata-cho, Kita-ku, Okayama 700-8550, Japan; ^dDepartment of Clinical Genomic Medicine, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences; 2-5-1 Shikata-cho, Kita-ku, Okayama 700-8550, Japan; and ^eDepartment of General Thoracic Surgery, Breast and Endocrinological Surgery, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences; 2-5-1 Shikata-cho, Kita-ku, Okayama 700-8550, Japan.

(Received August 28, 2019)

The development of specialized training programs for medical personnel, particularly nurses, clinical laboratory technicians, and pharmacists, is considered critical for the promotion of genomic medicine throughout Japan. Specifically, medical personnel skilled at analyzing and understanding high-throughput genomic data are in high demand. In this symposium, we will introduce the basic knowledge and skills necessary for processing genomic data.

Key words—genomic medicine; tumor mutation burden; biomedical data science; bioinformatics

1. はじめに

ゲノム医療の担い手として、看護師、臨床検査技師、薬剤師を主な対象とした「ゲノム医療従事者」の育成による統合的なゲノム医療体制の整備が喫緊の課題となっている。とりわけ次世代シーケンサーを用いた遺伝子パネル検査とそのデータ解析はゲノム医療の根幹であるにもかかわらず、その担い手が極端に不足している。しかしながら、データサ

イエンティストの雇用が困難な状況においては、基本的なデータ解析の流れを理解するとともに、解析結果の解釈ができるメディカルスタッフの育成が、現実的な対応として求められている。本シンポジウムでは、遺伝子パネル検査の流れと結果を紹介するとともに、「遺伝子変異量 (tumor mutation burden; TMB)」の概念についても解説を行う。

2. ヒトゲノムのサイズとエクソン領域

ヒトゲノムのサイズは約 30 億塩基であり、giga base (Gb = 1000000000 base = 10 億塩基) の単位を用いると、3 Gb と表すことができる。また mega base (Mb = 1000000 base = 100 万塩基) の単位を用いると、3000 Mb と表すことができる (Fig. 1)。ここで、タンパク質をコードしているエクソン (exon) 領域は、ヒトゲノムのわずか 1–1.5% 程度であることから、全エクソン領域は 30–45 Mb (3000–4500 万塩基) である。ちなみに、次世代シーケンサー (next generation sequencer; NGS) の出力は一般的に Gb や Mb で評価されることが多い。具体的には、1 リードあたり 300 塩基のリードを全部で 5000 万リード (50 M リード) 解析できる NGS の出力は、300 b / リード × 50 M リード =

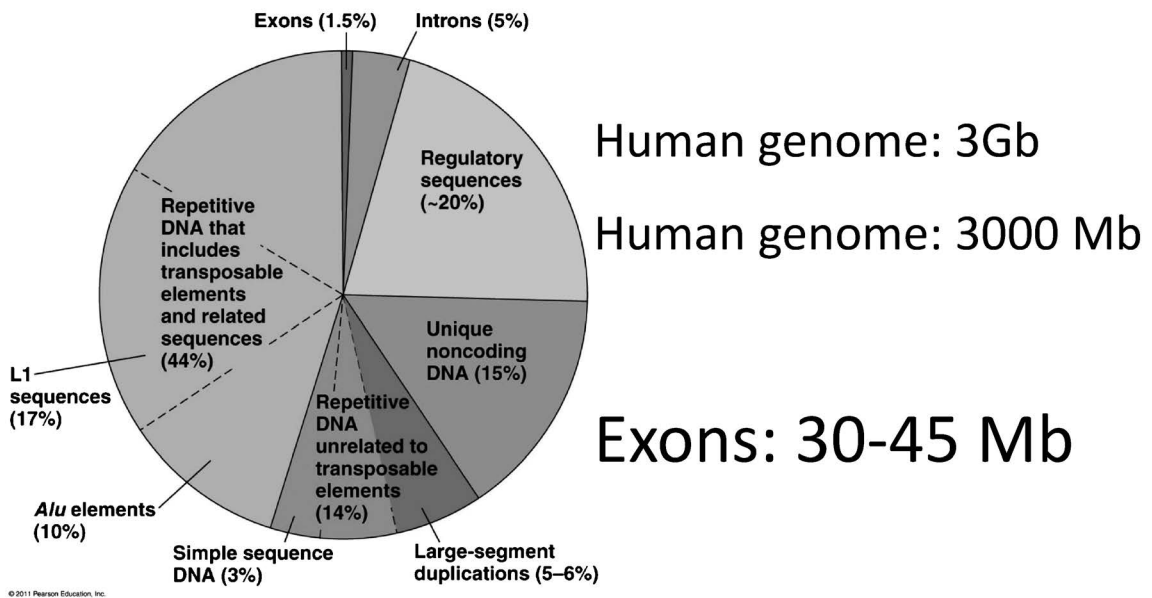
^a岡山大学大学院医歯薬学総合研究科 (〒700-8550 岡山市北区鹿田町 2-5-1), ^b岡山大学大学院医歯薬学総合研究科クリニカルバイオバンクネットワークワーキング事業化研究講座 (〒700-8550 岡山市北区鹿田町 2-5-1), ^c岡山大学病院新医療研究開発センター (〒700-8550 岡山市北区鹿田町 2-5-1), ^d岡山大学大学院医歯薬学総合研究科臨床遺伝子診療科 (〒700-8550 岡山市北区鹿田町 2-5-1), ^e岡山大学大学院医歯薬学総合研究科呼吸器・乳腺内分泌外科 (〒700-8550 岡山市北区鹿田町 2-5-1)

現所属: [†]岡山大学病院ゲノム医療総合推進センター (〒700-8550 岡山市北区鹿田町 2-5-1), ^{††}岡山大学大学院ヘルスシステム統合科学研究科 (〒700-8550 岡山市北区鹿田町 2-5-1)

*e-mail: shuta.tomida@okayama-u.ac.jp

本総説は、日本薬学会第 139 年会シンポジウム S49 で発表した内容を中心に記述したものである。

Exons: 1.0%-1.5% of human genome



https://bio1151b.nicerweb.com/Locked/media/ch21/human_genome.html

Fig. 1. Genomic Size of Human Genome

NGS data analysis flow

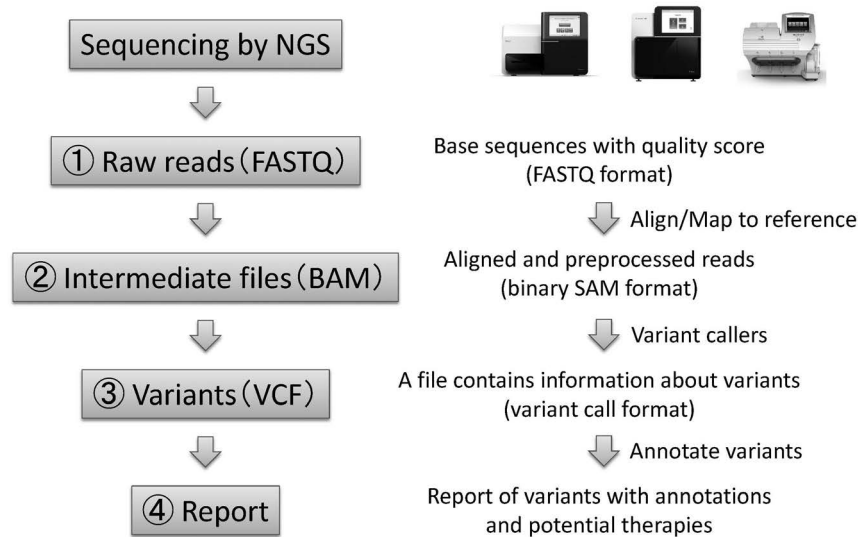


Fig. 2. Experimental Flow of NGS-based Clinical Sequencing

15000 Mb = 15 Gb (150 億塩基) の解析能力と表される。

3. 遺伝子変異の解析の流れ

NGS を用いた遺伝子変異解析の流れと出力ファ

イルについて概説する (Fig. 2)。

①NGS からは、FASTQ 形式の塩基配列データが得られる (元データともいう)。

②FASTQ ファイルに含まれるリードを、リファ

Somatic mutation frequency

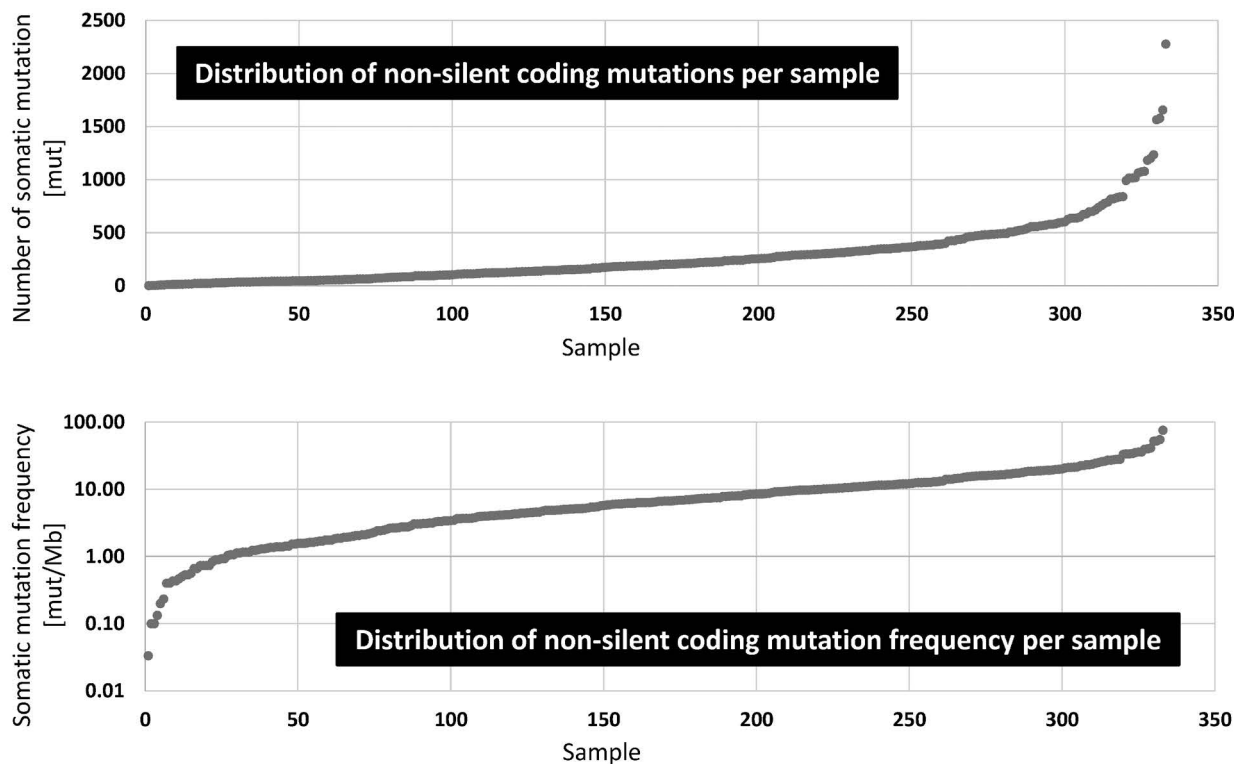


Fig. 3. Distribution of Somatic Mutation Frequency for Lung Adenocarcinomas

Cancer panel test reporting TMB

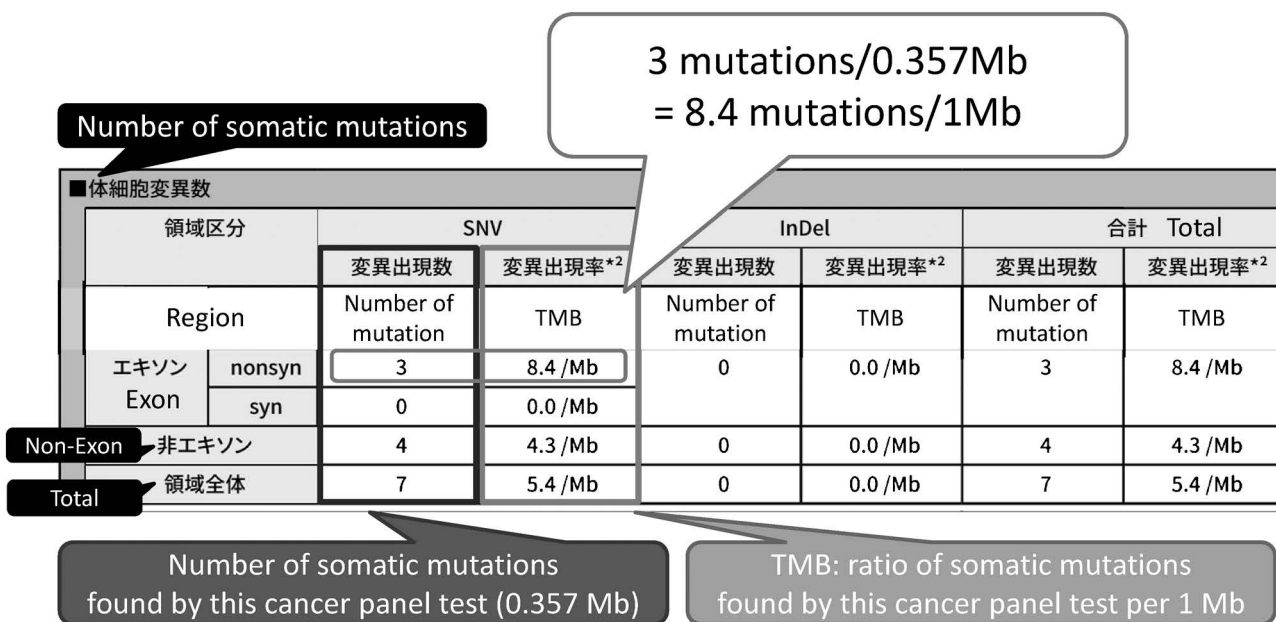


Fig. 4. TMB of Panel-based Test for Mutation Profile

レンスとなるゲノム配列（ヒトゲノム配列など）のどの位置（座位）に対応するのか解析することをマッピングという。その結果のファイル（SAMファイル）を圧縮したものをBAMファイルという。

③マッピング結果を基に、リードの塩基とリファレンスゲノムの塩基を比較することで、一塩基変異（single nucleotide variant; SNV）や挿入欠失変異（small insertions and deletions; indels）の遺伝子変異を抽出する（VCFファイル）。

④VCFファイルの情報を基に、アミノ酸変化やそれに伴う病的意義の有無、データベースの登録情報、論文の有無などの情報を付加（アノテーション）した結果をレポートとして出力。

4. 遺伝子変異量（TMB）

一般的に、TMBは下記の式で表される：

$TMB = \text{遺伝子変異数} / \text{解析対象となった遺伝子領域の塩基数 (Mb)}$

遺伝子変異数：腫瘍細胞から抽出された核酸を用いて、がん関連遺伝子のみを対象としたがん遺伝子パネル解析や全エクソン領域を解析対象とした全エクソン解析を実施することによって明らかとなった遺伝子変異の数

Figure 3はヒト肺腺がん症例335サンプルの全エクソン解析に基づき、1症例あたりに同定された遺伝子変異数と、解析対象となった遺伝子領域の塩基数（30 Mb）当たりの体細胞遺伝子変異数（塩基置換と挿入欠失変異を含む）の分布を示したものである。¹⁾ また近年、RizviらによりTMBの値が高い腫瘍は免疫チェックポイント阻害薬である抗PD-1抗体への感受性が高いことが報告されており、治療の指標としてもTMBの評価が検討されている。²⁾

5. 遺伝子パネル検査におけるTMB

全ゲノム解析や全エクソン解析ではなく、がん関連遺伝子の解析に特化したがん遺伝子パネル検査の結果から、TMBを算出する方法について以下にまとめる。

①がん遺伝子パネル検査を用いて抽出された遺伝子変異数を算出

②がん遺伝子パネル検査の仕様書等を参考に、解析ターゲット領域長を確認

例として、仕様書によると、National Cancer Center (NCC) オンコパネルの解析ターゲット領域長の合計は1.42 Mbとなっている。

③遺伝子パネル検査に基づく $TMB = \text{①} / \text{②}$

Figure 4では、解析ターゲットとするエクソン領域において遺伝子変異が3つ見つかっている。この遺伝子パネル検査における、エクソン領域の解析ターゲット領域長は0.357 Mbとなっており、エクソン領域における遺伝子変異量は8.4個/Mbとなっている。

6. おわりに

本シンポジウムでは、遺伝子パネル検査の流れを概説するとともに、「遺伝子変異量（TMB）」の概念と遺伝子パネル検査におけるTMBの算出方法についても解説を行った。これらの知見をエキスパートパネルでの議論に役立てて頂ければ幸いである。

謝辞 本シンポジウムは、国立研究開発法人日本医療研究開発機構（AMED）のゲノム創薬基盤推進研究事業・ゲノム創薬研究の推進に係る課題解決に関する研究（A課題）のゲノム医療従事者の育成プログラム開発（A-3班）として実施した研究成果を基にしている。

利益相反 開示すべき利益相反はない。

REFERENCES

- Lawrence M. S., Stojanov P., Polak P., Kryukov G. V., Cibulskis K., Sivachenko A., Carter S. L., Stewart C., Mermel C. H., Roberts S. A., Kiezun A., Hammerman P. S., McKenna A., Drier Y., Zou L., Ramos A. H., Pugh T. J., Stransky N., Helman E., Kim J., Sougnez C., Ambrogio L., Nickerson E., Shefler E., Cortés M. L., Auclair D., Saksena G., Voet D., Noble M., DiCara D., Lin P., Lichtenstein L., Heiman D. I., Fennell T., Imielinski M., Hernandez B., Hodis E., Baca S., Dulak A. M., Lohr J., Landau D. A., Wu C. J., Melendez-Zajgla J., Hidalgo-Miranda A., Koren A., McCarroll S. A., Mora J., Crompton B., Onofrio R., Parkin M., Winckler W., Ardlie K., Gabriel S. B., Roberts C. W. M., Biegel J. A., Stegmaier K., Bass A. J., Garraway L. A., Meyerson M., Golub T. R., Gordenin D. A., Sunyaev S., Lander E. S., Getz G., *Nature*, **499**, 214–218 (2013).
- Rizvi N. A., Hellmann M. D., Snyder A.,

Kvistborg P., Makarov V., Havel J. J., Lee W., Yuan J., Wong P., Ho T. S., Miller M. L., Rekhtman N., Moreira A. L., Ibrahim F., Bruggeman C., Gasmi B., Zappasodi R., Mae-

da Y., Sander C., Garon E. B., Merghoub T., Wolchok J. D., Schumacher T. N., Chan T. A., *Science*, **348**, 124–128 (2015).