

Introduction to Genetics

遺伝性疾患の全エクソーム解析 (Whole exome sequencing in genetic diseases)

宮武聡子 (Satoko Miyatake)

横浜市立大学附属病院遺伝子診療部 (Clinical Genetics Department, Yokohama City University Hospital)

横浜市立大学大学院医学研究科遺伝学教室 (Department of Human Genetics, Yokohama City University Graduate School of Medicine)

Keywords: Whole exome sequencing, next generation sequencer

ゲノム解析手法の技術革新により、ゲノム医学は長足の進歩を遂げてきた。1977年にサンガー法が考案され、¹ ゲノム上の特定の領域や遺伝子のシーケンス解析が可能となった。1998年に現在使われている96サンプル対応のキャピラリーシーケンサーが登場し、データ出力量が飛躍的に増大した。² 2005年より次世代 (第二世代) シーケンサーが登場し、超並列的なシーケンス手法によって全エクソームや全ゲノムといった網羅的な塩基配列決定が可能となった。またサンプルDNAから検索したい遺伝子領域のみを分画すればターゲットシーケンスとなり、特定領域をディープにシーケンスすれば、低頻度モザイク変異が検出できる。サンプルに相補的DNA (complementary DNA)を使えばRNA-Seqが可能である。

次世代シーケンサーの登場によってヒト疾患責任遺伝子の探索ストラテジーも変化した。かつては、家系の連鎖解析や患者のゲノム構造異常解析から得た疾患関連座位(locus)に存在する遺伝子の中でめぼしい候補遺伝子の変異解析を行っていくのが常套手段であったが、次世代シーケンサーがスタンダードな解析手法となった今は、家系構成員の網羅的なゲノム解析をまず行い、遺伝形式が合致する病的所見を探し出す手法が効率よい。この手法であれば次世代を残さない (家系解析が行えない) 孤発性の疾患といった、従来の遺伝学的手法では解析の手がかりをつかむことさえできなかった遺伝性疾患でも探索可能で、実際に次世代シーケンサーがヒトの遺伝学的研究に使用され始めた2010年頃から、これまで原因の特定が困難であった多くの疾患の責任遺伝子が同定された。³ Online Mendelian Inheritance in Man (OMIM; <https://www.omim.org/>) によると、原因が解明された遺伝性疾患は2007年1月には2048疾患が登録されていたが、2017年5月10日現在、4994疾患となっている。全エクソーム解析の解析コストダウンによって2013年ごろより、診断を目的としたクリニカルエクソームも試みられるようになりその臨床上の有用性も認識されるようになった。全エクソーム解析を臨床応用することで、これまで、診断を元に単一の遺伝子を調べるのが一般的だった遺伝子検査のあり方が、遺伝子のほぼ全部を一度に調べるというものに大きく変革した。臨床応用が進む中で、疾患との関連があるかどうか判断が難しいバリエーション (Variant of unknown significance; VUS) を見つけたときの解釈、膨大なゲノムデータの管理、利用、プライバシー保護、偶発的に見つかる所見への倫理的判断と対応といった問題も表出している。技術的には第二世代の限界を補う第三世代シーケンサーが登場し実用され始めている。本セッションでは、臨床現場でも耳にすることが多くなった“全エクソーム解析”について紹介する。

全エクソーム解析とは

ヒトには20000個程度の遺伝子があるといわれているが、全エクソーム解析は、ヒトゲノム配列のうち、ヒトの全遺伝子のcoding exon (exonの中でアミノ酸への翻訳情報を持っている部分)の配列のみ抽出してシーケンスを決定する方法である。exonはヒトゲノムのうち1-2%程度の領域に過ぎない

が、遺伝子に関係する疾患の85%の原因がexonにあるともいわれており、⁴ 全ゲノムの配列を調べるよりも効率の良い解析方法として普及した。

横浜市大遺伝学教室における全エキソーム解析のワークフロー（図1）

1) エキソームライブラリーの作成（ハイブリダイゼーションによる方法）

ヒトのDNA 2-3 μ gを超音波破碎装置もしくは酵素などによって、ランダムに断片化する。次に切断したDNA断片にアダプター配列を付加させてPCR反応で増幅する。ここで全エキソン領域の配列に相補的なプローブのセットを用意する。このプローブに磁気で吸着されるような標識を施しておき、増幅したDNA断片とハイブリダイズさせる。磁気によってプローブとハイブリダイズした断片を回収、精製する。個々のサンプルが識別できるよう、インデックス配列を含むプライマーを使ってPCR反応で断片を増幅させる。こうして作成した断片の集まりが全エキソームのライブラリである。

2) シーケンシング

シーケンサーによって、シーケンスの原理は異なるが、私たちの教室ではIllumina社のHiSeq200/2500を使用している。Illuminaプラットフォームでは、まずブリッジPCRという方法でライブラリに含まれる各断片を基板上でクラスター状にクローン化させる（クラスター合成）。その後一塩基合成反応(SBS: sequence by synthesis)という手法により、基板上で各断片をテンプレートとして、この断片の相補鎖を一塩基ずつ逐次合成させる。この際合成に使われる4種類の塩基には4種類の別々の蛍光色素を標識したものをあらかじめ用意しておく。一塩基合成したら反応を一度停止させ、基板をカメラで撮影し蛍光シグナルをイメージとして読み取る。これを一サイクルとしてくり返すと、連続する画像情報から各断片の塩基配列を取得することができる。蛍光シグナルの減衰などによって、このようなサイクル反応は劣化していくため、リード長（シングルリード）は100塩基対（bp）程度のショートリードとしている。

3) データ処理

シーケンシングデータは、各断片の配列情報（リード）に関するデジタルデータである。各リードはヒトゲノムのリファレンス配列と比較し、ヒトゲノムの中のどこかの配列に由来するリードなのかをマッピングする。次にマッピングされたリードとリファレンス配列を比較し（ジェノタイピング）リファレンスと異なる塩基配列を検出する。そして検出された変異配列が、どの遺伝子のどこかの塩基の置換にあたるのか、それにより遺伝子がコードするタンパク質の当該アミノ酸がどのように置換されるかなどの情報を付加する。

ヒトゲノム上のある一箇所の配列情報を正確に取得するためには、一箇所を複数のリードでカバーすればよい。何個のリードでカバーできたかをカバレッジと呼ぶが、私たちの教室ではターゲット領域の約90-95%の領域は20カバレッジ以上の厚みで情報を取得するようにしている。ヒトゲノムは約30億bpあり、その1-2%の領域を上流、下流の余剰を含めてライブラリにキャプチャーすると5000万bp程度であるので、1つのエキソームライブラリから得られたシーケンスデータは、約5000万箇所について、その95%以上の領域が20リード以上でカバーされた塩基情報であるから膨大なデータといえる。

4) 目的変異の検出

ヒトゲノムの標準配列と異なる塩基は1人のエキソームデータから約30万個得られる。これを目視で確認していくことは現実的でないので、解析の目的に応じてデータを絞り込む。まず、この中には配列変化（バリエーション）として検出されても、シーケンスデータの品質が悪く偽陽性である可能性が高いものが含まれるのであらかじめデータから除外する。そしてエキソン領域とエキソン-イントロン境界領域を優先して検索するために、それらを外れる領域（たとえばエキソン-イントロン境界領域から遠く離れたイントロン領域とか、エキソンの中で、非翻訳領域 (untranslated region; UTR) と呼ばれるアミノ酸の翻訳に関する情報を持たない部分、など）はデータから除外する。さらに希少遺

伝性疾患の責任遺伝子変異の検索を目的とする場合、一般集団にある程度以上の頻度で認められるバリエーション(common variant)は検索対象外であるのであらかじめデータから除外する。一般集団における各バリエーションの頻度は、dbSNP (<https://www.ncbi.nlm.nih.gov/projects/SNP/>), 1000genome (<http://www.internationalgenome.org/>), ESP (NHLBI Exome Sequencing Project; <http://evs.gs.washington.edu/EVS/>), ExAC (Exome Aggregation Consortium; <http://exac.broadinstitute.org/>), HGVD (Human Genetic Variation Database; <http://www.hgvd.genome.med.kyoto-u.ac.jp/>)などいくつかのデータベースの情報や、教室内で集積している健常日本人575人のデータなどを参考にしている。これらの絞込み作業によって、1人当たり低頻度なバリエーション(rare variant)が300-600個程度含まれるデータへサイズダウンさせる。この作業の後、解析対象としている疾患に合わせて、さらにデータを絞り込む。すでに責任遺伝子が知られている疾患であれば、当該の遺伝子にバリエーションがないか、類似の症状をしめす別疾患の責任遺伝子のバリエーションがないかまず検索する。この作業を正しく行うには、臨床診断が非常に重要であり、網羅的に遺伝情報があるからといって、臨床情報なくしては正しい結果にはたどり着けない。該当のバリエーションがなければ新規の責任遺伝子を探索するために両親の全エクソーム解析を追加し、想定される遺伝形式に合致するrare variantを抽出する。

病的バリエーションの判定

バリエーションには点変異と、塩基の挿入欠失の2つのパターンがある。点変異には、シノニマス(アミノ酸置換にならないもの)、ミスセンス(アミノ酸が置換されるもの)、ナンセンス(ストップコドンに置換されるもので、タンパク質合成がそこで終了するもの)がある。挿入・欠失変異には、フレームシフト(コドンの読み枠が変わるもの、多く場合早期にストップコドンが出現する)、インフレーム(コドンの読み枠が変わらないもの、いくつかのアミノ酸の挿入・欠失となる)がある。またエクソン-イントロン境界に点変異や挿入欠失変異が起こった場合は、スプライスサイトを変化させ、mRNAが生成される際にエクソンの読みとばしやイントロンの挿入などが起こる。その結果、コドンの読み枠がずれればフレームシフトと同じ効果をもつ。機能的には、シノニマス以外はコードされるタンパク質に何らかの影響を及ぼすと考えられる。ナンセンス変異、フレームシフト変異、スプライスサイト変異などは、タンパク質の構造を大きく変化させたり、変異mRNAやタンパク質の不安定性や有害性から細胞内でこれらが早期に分解されてほとんど存在しなくなるため、多くの場合病的効果をもつと考えられる。一方ミスセンス変異については病原性の判断は難しいことがしばしばある。一アミノ酸の置換によってタンパク質の機能が変化していることを、細胞や組織のレベルで証明できること(機能解析)が理想であるが、時間、コストの面から現実的ではない。ミスセンス変異がおこっているアミノ酸の進化的な保存性や、タンパク質の機能的なまとまり(ドメイン)との位置関係、アミノ酸置換によって、本来のアミノ酸とどれくらい性質が異なるものへ置き換わるのか、一般集団における頻度は遺伝形式を考えて妥当なのか、すでに同じ疾患の患者さんで報告されたアミノ酸置換ではないか、などといった観点で判断することが多い。

現状で各研究室が独自の観点で判断していると思われる病的バリエーションの判定については、世界的には標準化した判定基準を作る試みがなされており、2015年、アメリカ臨床遺伝学会(American College of Medical Genetics; ACMG)がバリエーションの病的効果をスコア化して病原性を判定できるガイドラインを発表した。⁵ 当教室でもこれにのっとって判定しているが、必ずしも完全ではない。

全エクソームデータで可能な他の解析

網羅的な遺伝子変異解析のほかに、家系の構成員のデータを統合した連鎖解析や、コピー数異常領域の検出が可能である。また同一疾患の大規模コホートの全エクソームデータがあれば、コントロールコホートのデータと統計学的な比較をすることで、メンデル遺伝には従わないが低頻度で病的な効果を持つ、いわゆるrare variantを検出することも可能である。

全エクソーム解析の短所

現行の全エクソーム解析は第二世代シーケンサーを使用している。このため原理上、リピートや同じ塩基の連なり、10塩基以上の挿入欠失変異といった複雑な配列、4つの塩基の分布が偏っている（グアニン・シトシンが多く存在している領域、など）配列などは解析困難である。またロングリードでないことから、特定配列の伸長をみる解析（リピート病など）やハプロタイプ情報は取得できない。最近実用されつつある、第三世代シーケンサーはこうした短所を補う特性を持っておりこれまでアプローチ困難だった領域の解明が期待されている。

全エクソーム解析の臨床診療への応用

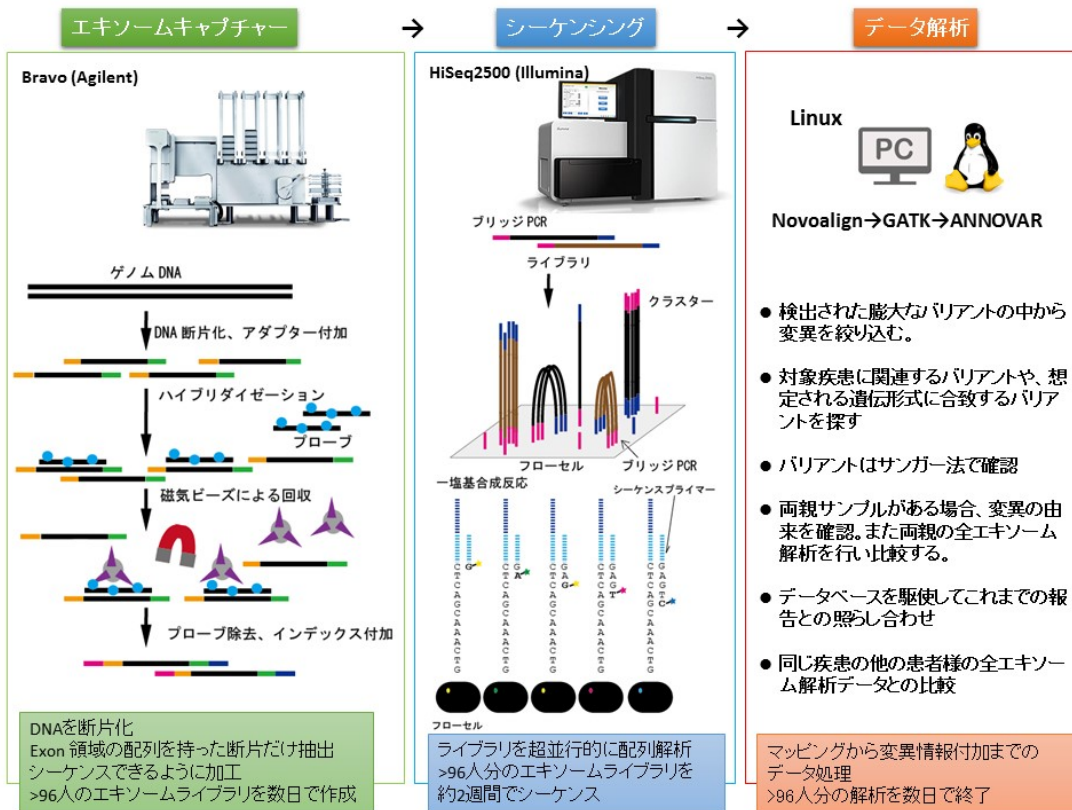
遺伝学的な背景を持つ疾患を対象に、クリニカルエクソームを行うと、対象疾患によって相違があるが、その診断率は概ね25-30%といわれている。⁶⁻⁸ 当教室での2016年度の解析実績は4513検体中原因同定率36.8%であった。網羅的な解析によって、従来知られていなかった遺伝型-表現型連関に気づくことがあり、遺伝子で規定されるある一疾患の表現型の広がり、これまで臨床診断学で考えられ、分類されてきたよりも幅広いことがわかってきた。⁹

全エクソーム解析の問題点

全エクソーム解析の臨床応用における問題点としていくつか挙げると、①病的意義のはっきりしないバリエーションの判断、②解析対象疾患とは関係のない、しかし健康上問題になると予想される遺伝子変異を意図せず見つける可能性（偶発所見; incidental findings）、③膨大なデータをどう管理し、プライバシーに配慮しながら活用していくか、④日本では保険診療になっていない、などがある。②については、ACMGが、既知の遺伝子変異が偶発的に見つかった場合、積極的に被験者に結果を返すべき遺伝子のリストをガイドラインとして発表している。^{10 11} 最新版のガイドラインには59遺伝子が含まれており、家族性腫瘍や不整脈、心筋症や大動脈瘤・解離、悪性高熱、代謝異常症などに関連するもので、早期に知ることによって予後の改善が見込める疾患が対象である (actionable gene)。日本では偶発所見についてのガイドラインはなく、個々の研究室に対応を任されている。④については、日本では研究の中で行わざるを得ず、必ずしも臨床的に適応がある症例が解析を受けられる状況ではなかったが、平成27年より日本医療研究開発機構 (AMED)主導のもと、未診断疾患イニシアチブ Initiative on Rare and Undiagnosed Diseases (IRUD: アイラッド) 事業が行われており、遺伝的要因が疑われる未診断例が広く解析を受けられる。<http://www.amed.go.jp/program/IRUD/> 臨床検査としての解析品質の担保などクリアすべき条件はあるものの、保険診療化に向けた検討がなされていく可能性がある。

図1 次世代シーケンサーのワークフロー

横浜市大遺伝学教室での次世代シーケンサーを使った全エクソーム解析のワークフロー



Reference

1. Sanger, F., Nicklen, S. & Coulson, A.R. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 74, 5463-5467 (1977).
2. Kircher, M. & Kelso, J. High-throughput DNA sequencing--concepts and limitations. Bioessays. 32, 524-536 (2010).
3. Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D., Shannon, P.T., Jabs, E.W., Nickerson, D.A., Shendure, J. & Bamshad, M.J. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet. 42, 30-35 (2010).
4. Majewski, J., Schwartzentruber, J., Lalonde, E., Montpetit, A. & Jabado, N. What can exome sequencing do for you? J Med Genet. 48, 580-589 (2011).
5. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K. & Rehm, H.L. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 17, 405-424 (2015).
6. Yang, Y., Muzny, D.M., Reid, J.G., Bainbridge, M.N., Willis, A., Ward, P.A., Braxton, A., Beuten, J., Xia, F., Niu, Z., Hardison, M., Person, R., Bekheirnia, M.R., Leduc, M.S., Kirby, A., Pham, P., Scull, J., Wang, M., Ding, Y., Plon, S.E., Lupski, J.R.,

- Beaudet, A.L., Gibbs, R.A. & Eng, C.M. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med.* 369, 1502-1511 (2013).
7. Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endeley, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N., Dufke, A., Cremer, K., Hempel, M., Horn, D., Hoyer, J., Joset, P., Ropke, A., Moog, U., Riess, A., Thiel, C.T., Tzschach, A., Wiesener, A., Wohlleber, E., Zweier, C., Ekici, A.B., Zink, A.M., Rump, A., Meisinger, C., Grallert, H., Sticht, H., Schenck, A., Engels, H., Rappold, G., Schrock, E., Wieacker, P., Riess, O., Meitinger, T., Reis, A. & Strom, T.M. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet.* 380, 1674-1682 (2012).
 8. de Ligt, J., Willemsen, M.H., van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., del Rosario, M., Hoischen, A., Scheffer, H., de Vries, B.B., Brunner, H.G., Veltman, J.A. & Vissers, L.E. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med.* 367, 1921-1929 (2012).
 9. Miyatake, S., Osaka, H., Shiina, M., Sasaki, M., Takanashi, J., Haginoya, K., Wada, T., Morimoto, M., Ando, N., Ikuta, Y., Nakashima, M., Tsurusaki, Y., Miyake, N., Ogata, K., Matsumoto, N. & Saitsu, H. Expanding the phenotypic spectrum of TUBB4A-associated hypomyelinating leukoencephalopathies. *Neurology.* 82, 2230-2237 (2014).
 10. Green, R.C., Berg, J.S., Grody, W.W., Kalia, S.S., Korf, B.R., Martin, C.L., McGuire, A.L., Nussbaum, R.L., O'Daniel, J.M., Ormond, K.E., Rehm, H.L., Watson, M.S., Williams, M.S. & Biesecker, L.G. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med.* 15, 565-574 (2013).
 11. Kalia, S.S., Adelman, K., Bale, S.J., Chung, W.K., Eng, C., Evans, J.P., Herman, G.E., Hufnagel, S.B., Klein, T.E., Korf, B.R., McKelvey, K.D., Ormond, K.E., Richards, C.S., Vlangos, C.N., Watson, M., Martin, C.L. & Miller, D.T. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med.* 19, 249-255 (2017).