

日本における Phenotyping の必要性と可能性

中島直樹^{*1}

*1 九州大学病院メディカル・インフォメーションセンター

Needs and Benefits of Phenotyping in Japan

Naoki Nakashima ^{*1}

*1 Medical Information Center, Kyushu University Hospital

Abstract

Human phenotype should be integrated results of personal factors including environment and daily behaviors from birth to today in addition to genotype. Phenotype also continues to be changed during life. Therefore, we have to chase personal phenotype continuously and correctively to understand genomic, environmental and behavior factors on forming phenotype. However, phenotype is not described appropriately in electronic health record or in any record system at least in Japan. Thus, we should develop “Phenotyping” on EHR or any other records to detect phenotype in a certain accuracy level. This phenotyping technology has been already used in the National Database and the MID-NET etc. in Japan.

In this symposium, after I introduce several examples of phenotyping done in Japan, I present how phenotyping using real world data is important in development of clinical decision support system, medical artificial intelligence, and precision medicine. I also present what are issues in development of phenotyping.

Keywords: Phenotype, Phenotyping, Genotype, Real World Data, Precision Medicine

1. 緒言、Phenotyping とは？

遺伝学の歴史は、主として同じ表現型 (Phenotype) を保有するグループに共通する遺伝型 (Genotype) を検知すること (Genotyping) によって発展してきた。例えば、目が青い、頭髪が黒い、身長が高いなどが Phenotype であり、これらに対応する Genotype を追求する。さらには、癌や糖尿病などの疾患にも疾患感受性遺伝子が見い出され、その保有により疾患の罹患率が上がることが証明されてきた^{1,2)}。

Phenotyping とは、生命体が示す表現型 (Phenotype) を把握する、という意味であり、ヒトでは、健康であること、疾患を発症したこと、発症した時期、病気の程度、どのような薬剤に反応したか、なども Phenotype と言え。Genotype は一生を通して不変 (静的情報) であるのに対して、Phenotype は変化し続ける (動的情報)。例えば、あるヒトは 40 歳台までは健康だったのが、50 歳で糖尿病になり、60 歳で肺癌に罹患する、というように変遷する。つまり情報学的に考えると、Phenotyping の方が Genotyping よりも複雑でかつ長期間 (生命体の寿命期間) のデータを処理することとなる。

また実験動物と異なり、複雑な社会生活を営むヒトは Genotype のみで Phenotype が決まるわけではない。生下時から個体を取り巻く環境 (大気、水、食料、喫煙 (副流煙)、薬物、その他の環境)、およびその個体に特徴的な生活行動 (運動行動、食行動、薬剤アドヒアランス、飲酒、喫煙、睡眠、労働、レジャー、性行動、その他の行動) にも影響されることが知られている。つまり、ヒトを苦しめる病態 (Phenotype) は、Genotype に加えて、その年齢までの個人の環境、生活行動による総合的な結果の現れ、と捉えることが可能である。すでに、気象情報や大気汚染情報などの公的情報あるいは個別のスマホなどを用いた環境因子や生活行動に関するデータが蓄積されやすい状況が揃いつつある現代社会で、これらのデータを用いた解析の中で明らかにしたいものは、やはり病態 (Phenotype) に関連する遺伝、環境、行動因子であろう。

このような、医療分野のみならず多くの社会システムの中で

蓄積された莫大なデータは「real world data」(以下 RWD) と呼ばれ、その活用が始まっている。電子カルテの普及やレセプトオンライン化の推進により日本の医療にも情報化時代が到来した³⁾。しかしながら、これまでの情報化の目的が、1) 診療報酬の請求、2) 医師法等に準拠するための診療録の記載、3) 他の医療者が個別症例を理解するための記録、であり、データの 2 次利用が目的ではなかったため、解析に用いることに適したデータの質や構造となっていない。例えば、「保険傷病名」と呼ばれる傷病名は必ずしも正確に病態を表しておらず、疾患の発症や変化に関しては、多くの場合フリーテキスト中に記載されているに過ぎないか、あるいは記載されていない。つまり RWD である電子カルテやレセプト情報は Phenotype の正確な記載に乏しいのである⁴⁾。

このことは、これからの日本におけるゲノム研究の推進のみならず、データ駆動型医学研究 (以下 DDMS) の実施、診療判断支援システム (以下 CDSS) の発展、そして Precision Medicine の展開などに大きな影響を及ぼす。そこで、医療情報学は、今後2つのことを果たさなければならないであろう。

一つは、既に蓄積している RWD から病態を抽出する Phenotyping の技術開発である。できるだけ正確に抽出しなければならないが限界はある。そこで、感度、特異度、陽性的中率などでその精度を明らかにすることが重要である。これには、RWD を保有する事業者の協力と手続き、および Gold Standard (以下 GS) を得るための専門医の協力等が必要である。

もう一つは、Phenotype を正確に、構造的に記載する社会システム、Electronic Health Record (以下 EHR) の開発と普及である。これには上述の Phenotyping の研究を含めた長い期間が必要である。

本シンポジウムでは、日本において現在どのように Phenotyping が必要とされ、どのような状況下にあるか、について述べたい。

2. 日本のRWDに存在する信頼できるPhenotype

電子カルテやレセプト情報には信頼性の高い Phenotype 情報はないと述べたが、いくつかの例外もある。例えば、Diagnosis Procedure Combination (以下 DPC) のいわゆる「資源病名」は、信頼性の高い情報と言える⁵⁾。これは、そもそも DPC 自体が解析を意識した制度であること、および入院医療費がその資源病名に依存するためである。つまり、アップコーディング(本来の病名よりも診療報酬の高い病名とすること)のリスクはあるが、これは明らかな不法行為であるため、高い精度が保たれる。

また、疾患登録システムは、多くの場合には専門医が判断した真の症例を登録することから、信頼性が高い Phenotype 情報を持っていると考えられる。例えば、がん登録への登録症例や、指定難病に登録された症例などがこれに該当する。さらには、学会などが主体で運用している疾患登録コホート研究においても、登録症例はその疾患に関しては信頼性の高い Phenotype を持つと考えられる。

全国レベルでの悉皆性の高低はそれぞれのシステムによって異なるが、DPC やがん登録などに関しては、少なくとも個々の医療施設での悉皆性はほぼ得られているであろう。

これら Phenotype が高い信頼性で得られている症例は、例えば次項の Phenotyping の精度を調査する際の GS として用いることも可能である。

3. 日本で用いられ始めている Phenotyping 技術

Phenotype は Genotype とセットで遺伝学の中で世に出てきた概念ではあるが、Phenotyping 技術は必ずしも Genotype との突合を目的として発展しているわけではない。以下に述べるように、日本でも、Genotype を含まない様々な研究や事業においても既に用いられ始めている。

3.1 悉皆性が高い DDMS の有用性と欠点

調べたい対象=母集団が(例えば日本人全体、のように)大きすぎるために、まず標本を抽出して統計学的に処理し、間接的に母集団を推定することが従来からの疫学研究の基本的な考え方である。例えば、地域コホート研究では、通常その地域から標本(A)を抽出解析し、地域全住民(B)の動向を推測し、日本人全体(C)をさらに間接的に推測する。つまり、非常に精密にAを調査することが可能であるが、B、Cに関してはそれぞれの選択バイアスに応じて精度が低下する。

その一方で、近年悉皆性の高いデータベースを2次利用したDDMSが発展してきた。これは、母集団自体を統計学的に直接処理することが出来る手法でもある。その代表例であるレセプト情報・特定健診等情報データベース(以下NDB)は、レセプトオンライン化がほぼ100%近い形まで普及したこと、特定健診を対象者(40歳から74歳の国民)の50%超の利用にまで増加していること、などから全国民の悉皆的な解析を可能としたものである。年間20億件のレセプトデータと、年間2600万件の特定健診データを突合する巨大な匿名化データベースである。また、匿名化された個票ベースでの経年解析も可能であり、日本人全体を直接、詳細に解析することができる。地域コホートでは推測しにくい日本の地域別の差異を比較することなども可能である。その他にもコストが安い、短期間で結果を得る、などのメリットは多い。

NDBの規模までは行かなくとも、例えばある病院が完全に電子カルテ化されれば、病院受診者に関しては悉皆データベースとなるわけであり、その病院を研究対象としたこと以外

の選択バイアスを避けることが出来る。2015年から始まった「データヘルス計画」では、保険者レベルで被保険者データベースを用いてこれを実施しているわけである。

しかしながら、問題はデータの質である。NDBは、前述したように保険傷病名などのデータ品質が低い、つまり病名のみでは病態=Phenotype)の把握が不正確であるという課題の他にも、医療扶助などの公費データが含まれない、死亡などの保険資格喪失データが無い、などの多くの課題があり、これらを許容するか、あるいは何らかの処理により向上せねばならない。このため、NDBのデータ品質を用いてできる限り高い精度で病態を把握するためのPhenotypingの技術開発は、NDBを用いる多くの事業に役立つ。

例えば指定難病の対象疾患は、その指定条件に有病者が国民の0.1%を超えないこと(つまり12.6万人以下であること)、と明記されており、指定難病への疾患申請時には全国の有病者数の把握が重要である。つまり、根拠のある有病者数が無い場合、「指定難病」という病態をNDBから出来るだけ正確にPhenotypingしなければならない。これに関して、2017年度厚生労働科学研究事業「1型糖尿病の実態調査、客観的診断基準、日常生活・社会生活に着目した重症度評価の作成に関する研究(代表者・田嶋尚子・慈恵医大名誉教授)」が、1型糖尿病の有病者数調査をNDBを用いて実施しており、2017年度中に推定結果を報告する予定である。

また、末期腎不全症例の透析や腎移植は患者の生活の質低下のみならず、医療費増大に与えるインパクトとしても大きな国家的課題である。しかし、その実数などの全国的な集計は学会主導のアンケート調査でしか算出されていない。これに関しては、2017年度日本医療研究開発機構(以下AMED)「慢性腎臓病(CKD)進行例の実態把握と透析導入回避のための有効な指針の作成に関する研究(代表・山縣邦弘・筑波大学教授)」がNDBでの推定や定点観測を試みており、2017年度中に、一定の結果を報告する予定である。

なお、これら田嶋班と山縣班は、AMED2016年度(2017年度も実施)・臨床研究等ICT基盤構築研究事業、「エビデンスの飛躍的創出を可能とする超高速・超学際次世代NDBデータ研究基盤構築に関する研究(代表・満武日裕・医療経済研究機構)」と連携して実施しており、NDBの保険傷病名、処方材料、加算などを用いたPhenotypingを活用するものである。筆者は、これらに研究分担者として参加しており、成果についてもシンポジウムで報告する。

3.2 薬剤疫学へのPhenotyping技術の応用と実用化

近年、薬剤疫学領域では薬剤の安全性向上のために、選定した医療機関(群)での調査ではあるが、調査対象薬剤の投与全例における副作用サーベイランスの手法を進展させてきた。これは、米国におけるSentinel Initiativeの動きから始まり、薬禍の対策を急ぐ日本においても、厚生労働科研事業や医薬品医療機器総合機構(以下PMDA)のMIHARIプロジェクトなどで進められてきた⁶⁾。

その集大成とも言えるのが、厚生労働省とPMDAが2011年度に始めた薬剤DDMSの基盤的的事业である医療情報データベース(以下MID-NET)である。10の協力医療施設グループ(23医療機関)のレセプト情報とSS-MIX2を用いて、PMDAの複合施設統合データ処理センターを中心にデータ

抽出・解析を行う現在 400 万人規模のデータベース事業である。リアルタイム性が強く、またデータバリデーションを厳格に実施しており、DDMS としてはデータ品質が高いことが特徴である。2018 年度からは、協力医療施設以外の研究者や製薬企業などの第三者活用を含めた本格稼働が開始される予定である⁶⁾。

これら厚生労働科研事業や、MIHARI プロジェクト、MID-NET では、副作用病態を検知するために、Phenotyping 技術を用いている。つまり、保険傷病名や検査結果、処方薬剤、材料、加算などによる Phenotyping の対象薬剤投与前後の差分で薬剤副作用の発現の有無を統計学的に示すものであり、陽性的中率や感度などの Phenotyping の精度が薬剤の評価につながるため慎重に行う必要性が高い。

筆者が代表研究を行っている 2017 年度 AMED 「MID-NET を用いた医薬品等のベネフィット・リスク評価のためのデータ標準化の普及に関する研究」では、過去の厚生労働科研事業、MIHARI、MID-NET、あるいは論文などを精査して Phenotyping ライブラリーを策定している。

また、2017 年度 AMED 「アウトカム定義のバリデーション等に関する研究(代表・宇山佳明・PMDA 医療情報活用推進室室長)」では、MID-NET で用いる Phenotyping の開発・精度向上に向けて機械学習などを使いながら研究を進めている。

これらの薬剤疫学の Phenotyping の可能性と限界についてもシンポジウムで報告する予定である。

3.3 疾患登録コホート研究への Phenotyping の応用

近年、いくつもの臨床学会が、公的競争研究費を利用した疾患登録コホート研究を開始しており、その多くは前述のように専門医が対象疾患と診断した症例を登録するものである。しかしながら、診断症例全てを登録することが種々の理由で困難となることで、登録症例の選択バイアスが問題になる場合もある。一方、日本腎臓学会が厚生労働省の臨床効果データベース事業に採択されて構築を開始した「包括的 CKD データベース構築事業(以下 J-CKD-DB)」では Phenotyping 技術を活用して、症例を抽出し登録している。SS-MIX2 から抽出した尿蛋白陽性や糸球体ろ過率低下が3か月以上に渡って見られる症例を CKD とみなし研究対象とする、という方法である。

このように、診断基準がクリアな場合には、J-CKD-DB のように Phenotyping により症例を抽出し、不同意例を除いて原則全例を登録することが、バイアスの改善、症例数の増大の上で有用である。但し、同意の取得方法などに関して倫理委員会による承認を得ていることが前提である。

4. 増大する Phenotype 情報の重要性

疾患ゲノム解析に必須な情報は個人の genotype と Phenotype であるが、Phenotype は年齢とともに移り変わる動的情報のため継続的な取得が必要である。バイオバンクも近年は、試料提供者の病態を含む臨床情報を継続的に取得することを重要視している⁷⁾が、診療録などから人手でデータを移行入力することによる、手間や誤入力はリスクとなる。そこで EHR などからの直接の電子データ捕捉による効率的かつ正確な病態の継続的取得が望まれる。

2015 年 1 月 20 日のオバマアメリカ合衆国大統領の一般教

書演説において、“Precision Medicine Initiative”が発表され注目されている。これは、特にがん領域で Genotype までを考慮して治療法を選択する個別化医療のことを指している⁸⁾。日本においても、2017 年 5 月に厚生労働省の「がんゲノム医療推進コンソーシアム懇談会」(座長：間野博行・国立がん研究センター研究所長)が、「がんゲノム医療中核拠点病院(仮称)」構想を提案し、2017 年度中に 7 病院程度が指定される予定である⁹⁾。これは、がん患者に医療保険で遺伝子パネル検査を実施し、数百のがん遺伝子などを調査し、治療法選択に用いるもので、まさに Precision Medicine である。

筆者は、今後の Precision Medicine はがん領域から慢性疾患領域へ拡大することのみならず、個人の環境や生活行動までも考慮に入れた真の個別化医療を行う方向に進むことを予想する(図1)。スマートフォンの爆発的な広がりなどによって IoT を活用した個人の環境や生活行動の把握が高度化し、Personal Health Record を中心にデータ管理が進めば、Phenotype に関連する遺伝子のみならず生活習慣、暴露環境を明らかにすることが可能となるであろう。

その中核となる phenotype を動的情報として正確に把握するための Phenotyping 技術開発は必須と言えよう。

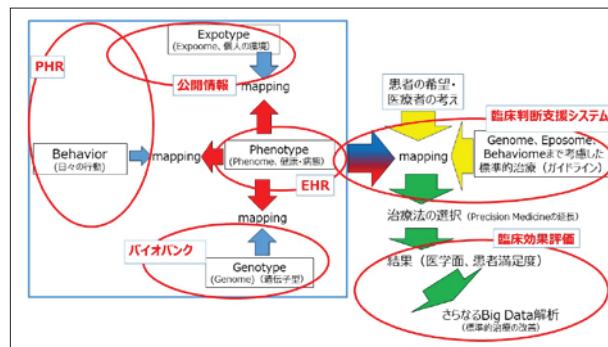


図1 Phenotype を中心とした次世代 Precision Medicine。Phenome は Genome のみならず、Exposome、Behavior にも影響され形成される。標準的治療は個別の Phenome に紐づいた遺伝、環境、行動因子も考慮して定義され、さらにそれを用いた個別の治療効果の解析が治療法を改善する。

述べてきたように日本の電子カルテは患者病態の抽出・解析には適した構造を有していない。つまり、現在は Phenotype を正確に継続的に記述/入力する EHR が日本には存在しない。そこで、電子カルテやレセプトのような診療業務用のデータベースの中の構造化データから一定のアルゴリズムで Phenotype を抽出する手法、Phenotyping の開発が行われている。

構造化データに頼る場合の Phenotyping の限界として、数値などに乏しい診断基準の疾患の Phenotyping 精度が悪いことが挙げられる。例えば、顆粒球減少症などは、顆粒球が 500/uL 未満であることが記載されていれば、陽性的中率は極めて高く、一般検血と白血球分画の実施率が高いため感度も高い。一方で、うつ病などは、患者の自覚症状が必ずしも明確ではないために診断率が低く、精神科以外の診療科における傷病名や処方の頻度も低く、血液検査などの検体検査もないため、Phenotyping における陽性的中率、感度ともに

極めて低くなる。

これに対して、非構造化データである自由文には、Phenotyping に有用な情報が潜在的に存在している可能性が高い。日々の診療記録や、退院記録、手術記録、画像レポート、などである。これらに対して機械学習を用いた自由文解析による Phenotyping の精緻化も期待される。さらに将来は、医用画像の生データなどが Phenotyping の材料になるかもしれない。

しかしながら、いつまでも重要な Phenotype の把握方法を現在のような解析には向かない電子カルテやレセプトデータを用いた Phenotyping に依存し続けることは無いであろう。つまり、次世代の電子カルテや EHR に Phenotype の効率的かつ正確な入力機能を実装すれば良いのである。今日の Phenotyping の技術開発はそのための礎にもなるであろう。

5. 結論

今後の医療情報学の責務の一つとして、AI や CDSS、Precision Medicine など今後の医療を担うシステムの開発に必要な Phenotype 情報を現在の EHR やレセプトデータなどの RWD から抽出するための Phenotyping 技術の開発、およびその知見を利用し、Phenotype を正確に格納する次世代 EHR の開発が挙げられる。

6. 謝辞

本シンポジウム発表は、2017 年度厚生労働科学研究費補助金「1 型糖尿病の実態調査、客観的診断基準、日常生活・社会生活に着目した重症度評価の作成に関する研究（代表者・田嶋尚子・慈恵医大）」、2017 年度 AMED 「慢性腎臓病（CKD）進行例の実態把握と透析導入回避のための有効な指針の作成に関する研究（代表・山縣邦弘・筑波大学）」2016 年度（2017 年度も実施）AMED・臨床研究等 ICT 基盤構築研究事業、「エビデンスの飛躍的創出を可能とする超高速・超学際次世代 NDB データ研究基盤構築に関する研究（代表・満武巨裕・医療経済研究機構）」、2017 年度 AMED 「MID-NET を用いた医薬品等のベネフィット・リスク評価のためのデータ標準化の普及に関する研究（代表・中島直樹・九州大学）」、2017 年度 AMED 「アウトカム定義のパリテーション等に関する研究（代表・宇山佳明・PMDA）」などの成果を含む。関係者に感謝する。

参考文献

- 1) Johannsen W. The genotype conception of heredity. *American Naturalist* 45(531): 129-159, 1911.
- 2) Taylor, Peter; Lewontin, Richard. "The Genotype-Phenotype Distinction". *Stanford Encyclopedia of Philosophy* 2017. [https://plato.stanford.edu/entries/genotype-phenotype/ (accessed 2017-Sep-1)]
- 3) 大江和彦. 電子カルテからの医療ビッグデータベース構築 *実験医学* 34: 792-798, 2016
- 4) 喜連川優. ビッグデータ *実験医学* 34: 761-763, 2016
- 5) 松田晋哉. 医療政策決定へのビッグデータ活用の可能性.

実験医学 34: 830-835, 2016.

6) 石黒智恵子, 宇山佳明. PMDA における医薬品の安全性評価を目的とした医療情報データベースの薬剤疫学的利活用: MIHARI Project と MID-NET 薬理と治療 44(suppl-1): 5012-5016, 2016.

7) 鈴木洋一, 山本雅之. ゲノムコホート研究とバイオバンク - 個別化予防と個別化医療の基盤づくりのために. *医学のあゆみ* 250(5): 321-325, 2014.

8) The White House. The Precision Medicine Initiative. 2015 [https://obamawhitehouse.archives.gov/node/333101] (accessed 2017-Sep-1)]

9) 厚生労働省 がんゲノム医療中核拠点病院(仮称)等の指定要件に関するサブワーキンググループ [http://www.mhlw.go.jp/stf/shingi/other-kenkou.html?tid=474999] (accessed 2017-Sep-1)]