国際塩基配列データベースと遺伝学的検査におけるバリアントの標準的な表記法について

2015年7月8日発行　文責　中山智祥

1. 国際塩基配列データベースについて

塩基配列データベースは、国際塩基配列データベース協力体制（International Nucleotide Sequence Database Collaboration: INSDC）によって収集、アーカイブされている。INSDCは日本DNAデータバンク（DNA Data Bank of Japan: DDBJ）、米国National Center for Biotechnology Information（NCBI）、欧州の European Bioinformatics Institute（EBI） で運営されている European Nucleotide Archive（ENA）の3つの拠点で構成・連携され、インターネットを経由したデータと情報の相互交換がなされている。国際諮問委員会（International Advisory Committee: IAC）と国際実務者会議（International Collaborative Meeting: ICM）の２つの会議が定期的に開催されている（<http://www.ddbj.nig.ac.jp/insdc/insdc-j.html#insd>）。研究者は決定した塩基配列データを上記3つの機関のいずれかにsubmitすれば、そのデータは3つの機関が維持するデータベースであるDDBJ、GenBank、EMBL-Bankのすべてに登録・公開される。

1. 参照配列について

世界中の様々な研究者の努力によってヒトゲノムプロジェクトが2003年に完了し、国際塩基配列データベースを通じてその配列を基本情報（参照配列）として利用できるようになっている。ヒトではGenome Reference Consortiumがsubmitしている参照配列があり、2015年春現在Build38版になっている。細かい修正が加えられると、patch releaseとしてpナンバーが加えられていく。2015年4月3日の最新版はGenome Reference Consortium Human Build 38 patch release 3 (GRCh38.p3)である。参照配列の表記法は、モデル参照配列はXで始め、既知（known）の参照配列はNで始める。そこで接頭辞としてNC\_ (染色体chromosomes)、NM\_ (mRNAs)、NP\_ (proteins)、NT\_ (constructed genomic contigs)を用いる。NG\_は genomic regionsあるいはgene clusters (例immunoglobulin region) を表す。

1. HGVS表記法についての変遷

（<http://www.hgvs.org/>）

さて、ヒトゲノム、RNA、コード領域のDNA、タンパク質配位におけるバリエーションの表記法についてはHuman Genome Variation Society (HGVS)によって確立されたものを標準として現在NCBIが承認している。

表記法の標準化に関する議論は1993年に発表された以下の2つの論文によって始まった。

1. Beaudet AL, Tsui LC: A suggested nomenclature for designating mutations. Hum Mutat. 1993;2(4):245-8.
2. Beutler E: The designation of mutations. Am J Hum Genet. 1993 Sep;53(3):783-5.

その後、議論され、変更の上、広がり、最終的に主に受け入れられた。現在の規定は2000年の以下の論文が基本となっている。

1. den Dunnen JT, Antonarakis SE: Mutation Nomenclature Extensions and Suggestions to Describe Complex Mutations: A Discussion. Hum Mutat. 2000; 15: 7-12.

この規定はその後追加・改訂されているが、それでもカバーされない複雑なケースについては彼ら2名の以下メールアドレス宛にて連絡を取ることを勧めている。

J.T.den\_Dunnen@LUMC.nl

Stylianos.Antonarakis@medecine.unige.ch

1. HGVS表記法についての見解

変異（mutation）と多型（polymorphism）という単語は使用せず、**sequence variant、 alteration、allelic variant**を使用する。すなわち変異には単に変化を示す時や、疾患原因の変化を示す場合があり、一方多型には疾患原因の変化ではないという意味や一般集団中1%以上に頻度を示すものという意味があり混乱を来しているからである。よってSNPまたはSingle Nucleotide Polymorphismという単語も用いない。

同様に病原性の（pathogenic）という単語を使用せず、**affects function**を使用するよう推奨している。これはpathogenicという単語を聞いた非熟練者が短絡的に疾患原因を想起するのに対し、熟練者は特定の状況下で疾患原因になると理解することに依る。

バリアントは以下の5つのカテゴリーに分類される。

1. affects function（機能に影響あり） ：機能に影響している場合
2. probably affects function ：おそらく機能に影響している場合
3. unknown ：機能に影響しているか不明の場合

variants of unknown significance (VUS)ともいう。

1. probably does not affect function ：おそらく機能に影響していない場合
2. does not affect function ：機能に影響していない場合
3. HGVS表記法具体例

HGVS表記法によると参照配列のナンバー、コロン「:」、バリエーション情報（位置および変化）を順に空白なしに記載することになっている。参照配列のナンバーについてはもしバージョンアップ（修正版）されていたら必ずそれも含む。

例を示すとNM\_020631.3:c.1940T>Cが意味するところは、*PLEKHG5*遺伝子mRNAの修正 3版を参照配列としコーディング参照配列のスタートコドンのAを+1とした場合に1940に位置する塩基であるTがCに変化するバリアントということになる。これはHGVSデータによるとNC\_000001.10:g.6529504A>G, NC\_000001.11:g.6469444A>G, NG\_007978.1:g.55566T>C, NG\_029910.1:g.1752T>C, NM\_001042663.1:c.2108T>C, NM\_001042664.1:c.1940T>C, NM\_001042665.1:c.1940T>C, NM\_001265592.1:c.2177T>C, NM\_001265593.1:c.2147T>C, NM\_001265594.1:c.1940T>C, NM\_020631.3:c.1940T>C, NM\_020631.4:c.1940T>C, NM\_198681.3:c.2171T>C, NP\_001036128.1:p.Phe703Ser, NP\_001036129.1:p.Phe647Ser, NP\_001036130.1:p.Phe647Ser, NP\_001252521.1:p.Phe726Ser, NP\_001252522.1:p.Phe716Ser, NP\_001252523.1:p.Phe647Ser, NP\_065682.2:p.Phe647Ser, NP\_941374.2:p.Phe724Ser, XM\_005263485.1:c.2171T>C, XP\_005263542.1:p.Phe724Serと同一のものであり、OMIMデータではallelic variantsの611101.0001として登録されている。dbSNPデータベースではrs63750315 SNPとなっている。

研究者がヒトバリエーションをsubmitする場合、NCBIへではインプット用のexelファイルが用意されており（http://www.ncbi.nlm.nih.gov/projects/SNP/tranSNP/tranSNP.cgi）以下のようなデータを入れるようにしている。申請者のe-mail、HGVS表記法、疾患名（OMIMかMESHのID）、臨床的意義、自施設でのID、OMIM allelic variant ID、独立して観察された回数、自施設でのwebsiteへのリンク、germline（生殖細胞系列）かsomatic（体細胞）か、バリアントの別称、その他の情報。

このことからNCBIやHGVSではHGVS表記法とOMIMのallelic variants IDを重視しており、NCBIのdbSNPのrsナンバー（登録されたSNP）を必須としていないが、日本での認知度ではOMIMよりdbSNPが高いと思われるので、補助データとして報告書などへの記載は有用であると考える。

1. 遺伝学的検査の報告書に記載すべきこと

2007年に出たOECD (organisation for economic co-operation and development) guidelines for quality assurance in molecular genetic testing（以下OECDガイドライン）には遺伝学的検査の報告書には最低限以下の情報を含めるべきであるとある。

1. 被験者識別番号
2. 医療者氏名と連絡先
3. 検査の指示と検査解釈に関係する医学情報
4. 検査方法（解析範囲と限界、感度、特異度を含む）
5. サンプルの種類（解釈のために必要な）
6. サンプル受領日
7. 検査実施施設名と住所（再外部委託の場合の施設を含む）
8. 検査結果
9. 結果の解釈と検査実施施設へ供給されたその他情報
10. 報告書承認者情報
11. 検査実施施設の連絡先
12. 報告書発行日

あと必要に応じて、以下の情報も含むべきである。

1. 有資格の医療者による遺伝カウンセリングの推奨
2. 他の血縁者への影響
3. フォローアップ検査の推奨

外部委託検査所によって報告された検査結果のすべての必要不可欠な関連した要素については、検査をオーダーした医療者への報告書に含めなければならない。

日本ではOECDガイドラインの原則を尊重、遵守しつつ、国内事情も考慮した形で遺伝子関連検査を実施する検査施設の質保証の実務に関する、「遺伝子関連検査に関する日本版ベストプラクティスガイドラインTentative Guideline（暫定文書）」を2010年に特定非営利活動法人日本臨床検査標準協議会（JCCLS）遺伝子関連検査標準化専門委員会が作成・公表した。その中の「3.4　結果の報告の質」には以下のようなことがらを追加している。

1. サンプル採取日を記載する。
2. 検体の状態に関する情報（必要に応じて、添加物、輸送や保存の状態等）を記載する。
3. 検査結果の中間報告を行った場合は、その内容を最終報告に反映させる。
4. 検査結果の解釈に必要な情報（報告書には、臨床的意義を含めた適切な医学的解釈を記載するとともに、依頼元に対して十分な情報提供に努めること。）を記載する。
5. 報告書には、塩基配列の記載に標準的な方法を用いる等、国際的に容認された用語と命名法を使用すべきである。ただし多数の高頻度変異で標準的命名法と一致しない名称がついている場合併用を認める。どの命名法を使用しているかを示すべきである。
6. 検査結果を正確に解釈するためには、被検者・家族の情報が必要であることを、被検者に知らせるべきである。
7. 検査施設は、報告に関連するすべての情報を医療法に従い最低５年間記録・保存しておかなければならない。ただし、検査実施施設の事情、技術要件によって可能な限り長期間とする。
8. 遺伝学的検査の報告書フォーム例

|  |  |  |  |
| --- | --- | --- | --- |
| 報告日 |  |  |  |
| 依頼者情報 | 施設名 |  |  |
|  | 依頼者 |  | ご机下 |
|  |  | この度は私たちが実施している遺伝学的検査をご依頼くださいまして誠にありがとうございました。 |  |
|  | ご住所 |  |  |
|  | お電話 |  |  |
|  | FAX |  |  |
|  | e-mail |  |  |
|  | Webサイト |  |  |
|  |  |  |  |
| 被験者データ | 被験者識別番号（依頼施設） |  |  |
|  | 被験者識別番号（検査実施施設） |  |  |
|  | 被験者の医学的情報 |  |  |
|  |  |  |  |
| 検体データ | 検体の種類 |  |  |
|  | 検体採取日 |  |  |
|  | 抗凝固剤 |  |  |
|  | 検体受領日 |  |  |
|  | 検体輸送状態 |  |  |
|  | 受領後検体保存条件 |  |  |
|  | 核酸抽出法 |  |  |
|  | DNA回収量 |  |  |
|  | DNA保存条件 |  |  |
|  |  |  |  |
| 遺伝学的検査データ | 検査の適用理由 |  |  |
|  | 検査方法 |  |  |
|  | 解析遺伝子 |  |  |
|  | 解析範囲（exon） |  |  |
|  | 解析の限界 |  |  |
|  | 感度、特異度 |  |  |
|  | 検査結果 |  |  |
|  | バリアントHGCS表記 |  |  |
|  | バリアントその他の表記 |  |  |
|  |  | 1) affects function： 機能に影響 |  |
|  |  | 2) probably affects function： おそらく機能に影響 |  |
|  |  | 3) unknown： 機能に影響しているか不明の場合：VUSともいう。 |  |
|  |  | 4) probably does not affect function ： おそらく機能に影響していない |  |
|  |  | 5) does not affect function： 機能に影響していない |  |
|  | 検査結果解説 |  |  |
|  | 既知か未知か |  |  |
|  | 頻度 |  |  |
|  | 結果解釈 |  |  |
|  | 関連文献 |  |  |
|  | 検査結果解説捕捉 |  |  |
|  |  |  |  |
| 遺伝カウンセリング体制 | 遺伝カウンセリングの推奨 |  |  |
|  | 他の血縁者への影響 |  |  |
|  | フォローアップ検査の推奨 |  |  |
|  |  |  |  |
| 検査実施施設情報 | 施設名 |  |  |
|  | 報告書作成者 |  |  |
|  | 報告書責任者 |  | 役職 |
|  | 所属 |  |  |
|  | 住所 |  |  |
|  | 電話 |  |  |
|  | FAX |  |  |
|  | e-mail |  |  |
|  | Webサイト |  |  |