

個人識別性 について：法科学からの視点

分担研究者 大澤 資樹 東海大学医学部基盤診療学系法医学 教授

ゲノム情報から個人を特定する時には、一人分の配列情報があったとしても、個人を特定することは困難である。他の情報と照らし合わせて始めて、個人が特定できる。対照する方法としては、配列の一致を調べるマッチング（異同識別）が単純で一般的な方法となる。それに加えて、ゲノム情報の特徴として血縁関係を調べることによっても個人が特定できる。このアプローチはファミリアルサーチ（血縁関係推定）と呼ぶ。

マッチングに関しては、非血縁者を仮定する限り、連鎖不平衡や連鎖は必ずしも関係なく、一致確率は集団内の対立遺伝子頻度のみ依存する。どの程度の一致確率が求められるかといえば、数千万分の1でも誤認逮捕された英国の事件が有名である。例えば、ある人物のABO式血液型がA型で、別のゲノム情報からA型遺伝子が得られ、一致したとする。しかし、実際には国内には5千万人もA型の人があり、A型が一致したからといって、個人が特定できたかということ、とても言えない。対象とする集団の大きさが重要で、ある程度の数のSNPや繰り返し配列が一致して始めて、特定の個人に結びつく可能性がでてくることになる。

ファミリアルサーチから個人を特定した例としては、厚生労働省が関係する中国残留孤児の身元調査や戦没者慰霊事業が挙げられる。DNAを抽出できるような本人の遺品は残っていないので、もっぱら親族との血縁関係を推定することで個人を特定してきた。逆の言い方をすれば、ゲノム情報から血縁者かどうか分かるリスクがあるということになる。

法科学領域では、この個人識別をSTR（short tandem repeat）と呼ばれる繰り返し配列の解析を中心に行ってきた経緯がある。特に、米国連邦捜査局（Federal Bureau of Investigation）は、CODIS（combined DNA index system）と呼ばれる4塩基の繰り返し単位からなる13 STR座位を犯罪捜査等の個人識別目的で検査する座位と定め、現在では各国で採用され指紋のようにデータベースとして集積されている。市販キットは、CODIS座位と性別判定用座位を含めて、一度に16～23座位を1本のチューブ内で増幅し解析するシステムが採用されており、multiplex系と呼ばれる。

【個人識別性を有する総一致確率について】

ある集団内で個人識別性を有する可能性のある遺伝型の一致確率は以下の式で表すことができる。

$$(1 - p_x)^N \geq 1 -$$

p_x ：ある座位Xでのプロファイルが偶然に一致する確率（random match probability）

$1 -$ ：信頼水準（confidence level）

N：母集団数（sample size）

（参考文献：Budowle, B., Chakraborty, R., Carmody, G., and Monson, K.L., (2000) Source Attribution of a Forensic DNA Profile. Forensic Science Communications, 2 (3). Available online at:

<https://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/july2000/index.htm/source.htm>)

集団サイズ (N) を 1 億人とした時に、この式から要求される一致確率は、次のように計算できる。

	1 -	: 信頼水準 (confidence level)
集団サイズ (N)	0.95	0.99
1 億人 (10 ⁹)	5.13 x 10 ⁻¹⁰	1.01 x 10 ⁻¹⁰

(参考文献: “Forensic DNA typing” 2nd Ed., John M. Butler, Elsevier, pp 513-515, 2005)

【実際に要求される解析座位数について】

仮に、集団を 1 億人と仮定した時に、信頼水準 0.99 で個人を識別するためには、1.0 x 10⁻¹⁰ 程度の総合一致確率 (P) が求められる。そして、互いに独立した何個の座位における遺伝型の出現頻度 (p_x 値) を掛け合わせると、このレベルに達するかを考えてゆくことになる。式で表すと以下のようなものとなり、n がどの程度の数値となるか SNP と STR で検討してみた。

$$P = p_1 \times p_2 \times p_3 \times \dots \times p_n \leq 1.0 \times 10^{-10}$$

任意の二検体間における座位 X における遺伝型が一致 (IBS2) する確率は、
 $p_x = a_i^4 + 4a_i^2a_j^2$ (i ≠ j, a: 対立遺伝子 (アリル) 頻度)
 で計算される。

個人識別に利用される 4 塩基繰り返し構造の STR (マイクロサテライト) は、p_x 値は 0.07 ~ 0.10 をとる。仮に、一致する確率が 0.08 ばかりの場合を考えた時に、座位数が増えた時の複合一致率は、以下ようになる。

座位数 (n)	複合一致率 (P)
1	8 x 10 ⁻²
2	6.4 x 10 ⁻³
...	...
8	1.7 x 10 ⁻⁹
9	1.3 x 10 ⁻¹⁰
10	1.1 x 10 ⁻¹¹
11	8.6 x 10 ⁻¹³

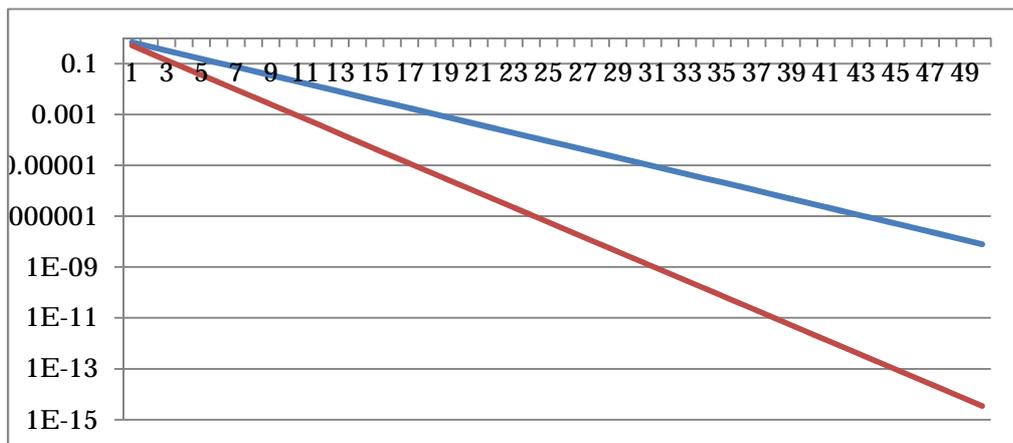
1.0 x 10⁻¹⁰ 以下になるのは、9 ないし 10 座位を検出した時と判断した。

次に、SNP においては、マイナーアリル頻度を以下のように仮定した場合に、二検体間で座位 X において遺伝型が一致する確率は次の表となる。

minor allele	単独一致率 (p _x)
0.05	0.8235375
0.1	0.6886
0.2	0.5136
0.3	0.4246

0.4	0.3856
0.5	0.375

SNPのアリル頻度の分布データを持ち合わせていないので、マイナーアリル頻度が0.1(青線)と0.2(茶線)のみの場合の複合一致確率を示したのが、下のグラフになる。



(本来ならば、もっと正確な試算をしなければならないが、今回の期間内ではこの程度までである。)

そこで、個人を1億人の中から特定するために求められる解析座位数を概算すると以下の表のようになる。

	p_x 値	必要な個数
SNP	0.38 ~ 0.85	40 ~ 50
STR	0.07 ~ 0.10	9 ~ 10

日本人集団を想定した時に、個人識別性を有するゲノム情報とは、SNPではおおむね40座位以上のデータを、STRでは9~10座位以上のデータに含むものに相当すると考えた。ただし、SNPに関しては、座位ごとのアリル頻度により、求められる数はかなり変わってくることはやむをえない。ただ、一般的に、STRの一致率は、SNPの4ないし5座位分の一致率になるとされており、この観点からも大きな間違いはないと考えている。