

## 補足説明

個人情報保護委員会事務局から解析方法・データ保存・個人識別の基準各項目につき質問があったので、これに応じて作成したものが以下の文書である。

## 解析方法

### 1 DNA 塩基配列決定 ( [https://en.wikipedia.org/wiki/DNA\\_sequencing](https://en.wikipedia.org/wiki/DNA_sequencing) )

#### i. 古典的手法

初期の DNA 塩基配列：プライマー延長法を原理とするもの

全ゲノム塩基配列決定：Direct-Blotting-Electrophoresis-System GATC 1500 ( EU のゲノム解読計画で使用 ); AB370I 半自動 DNA シーケンスシステム( Applied Biosystem )  
Genesis 2000 (Dupont) (1980 年代終わり頃); ショットガン法。

#### ii. 次世代

Single-molecule real-time sequencing (Pacific Bioscience)

Ion semiconductor (Ion Torrent sequencing)

Pyrosequencing (454LifeSciences)

Sequencing by synthesis (Illumina)

Sequencing by ligation (SoLiD sequencing)

Chain termination (Sanger sequencing)

#### iii. 開発中

Nanopore DNA sequencing

Tunnelling currents DNA sequencing

Sequencing by hybridization

Sequencing with mass spectrometry

Microfluidic Sanger sequencing

Microscopy-based techniques

RNAP sequencing

*In vitro* virus high-throughput sequencing

### 2 短い縦列反復配列タイピング: PCR して電気泳動により反復回数を測定

### 3 一塩基多型タイピング: BeadChip (Illumina)、Axiom (Affymetrix)、TaqMan (Life Technologies)

注意：以上のリストは、上市或いは上市前のものを含め、全ての解析法を網羅している訳では無い。

## 保存

- 1 . 一般に電子媒体で保存され、そこから情報を引き出し使用されている ( 例 <http://www.ncbi.nlm.nih.gov/gene/> )

現在一般的なゲノム解析法のいずれにおいても、ヒトゲノム約 30 億塩基対のうちの一部がゲノムデータとして取得されている。即ち、

- 短い縦列反復配列 ( Short Tandem Repeat, STR ) タイピング解析では、特定遺伝子座における STR の反復の回数を測定し、その回数をデータとして取得する。回数は、例えば 3 回だったり 4 回だったりする。
- 一塩基多型 ( Single Nucleotide Polymorphism, SNP ) タイピング解析では、特定遺伝子座に位置する SNP のアレルを測定し、その塩基をデータとして取得する。アレルは、例えばアデニン ( A ) だったりシトシン ( C ) だったりする。
- DNA 配列解読解析では、特定遺伝子座における数百塩基に渡る領域について、塩基配列を測定してデータとして取得する。塩基配列は、例えば AAGCCTACAC のように塩基が連なっている。

## ヒトゲノム ( [https://en.wikipedia.org/wiki/Human\\_genome](https://en.wikipedia.org/wiki/Human_genome) )

ゲノム情報の一意性の議論においては、ヒトゲノムがどのような構築になっているかの理解が必要である。

1 . ヒトの遺伝情報は、22 対の常染色体に加えて、女性では 1 対の X 染色体、男性では一本の X 染色体と一本の Y 染色体に存在する。従って、常染色体上の遺伝子は原則 1 対 ( 2 個 ) あることになる。加えて、必ず母親に由来するミトコンドリアの遺伝子がある。人間とチンパンジーの間のゲノムの違いは 1.2%、人間の間での違いは 0.1%と云われている。

2 . ヒトの全ゲノムの内、タンパクをコードしている領域は全体の 1.5%で残りはタンパクをコードしない。遺伝子をコードしない領域は ( 非コード領域 )

Pseudogene

Noncoding RNA

Introns and untranslated regions of mRNA

Regulatory DNA sequence

Repetitive DNA sequences ( 反復配列 ) : minisatellite, microsatellite, satellite; short interspersed nuclear element ( SINE ) ; DNA transposon; LTR retrotransposon; long interspersed nuclear element ( LINE ); rDNA 等

( [http://www.nature.com/nrg/journal/v13/n1/box/nrg3117\\_BX1.html](http://www.nature.com/nrg/journal/v13/n1/box/nrg3117_BX1.html) )

これら反復配列は、全ゲノムのほぼ半分になる。

3 . この内、反復配列は、個人特定の為に法医学で汎用されている技術である

( <http://aboutforensics.co.uk/dna-analysis/> )

4 . ヒトのゲノム解読は緒についたばかりで、特に、非コード領域の機能については不明な点が多い ( 例 : 脳細胞での発現が高い retrotransposon の生理的意味など ) 。解読進展に伴いゲノム情報の個人情報としての各情報の評価も変化する可能性がある。

## 一意性の判断

1. 多くの研究班員の指摘するように、
  - 同一対象者から情報が得られる遺伝子座の数、
  - 当該遺伝子座の対立遺伝子、
  - 遺伝子頻度（調査対象集団で替わり得る）、
  - 調査母集団人口、
  - 調査対象集団の地域特異性等、が一意性に影響する。
2. 個人識別の目安として次の様な提案がなされている。
  - ゲノムデータによる個人識別としては、便宜的ではあるものの、100万人から1人を特定できるなら個人識別できると便宜的に定義できる（しかし、1億人から1人を特定できると書き換えても、便宜的である事には変りない）。
  - 対象集団で多様性が高く、しかも互い連鎖不平衡にない遺伝子座であれば、30遺伝子座を測定すれば、約100万人から1人を特定でき、個人識別できることになる。
  - DNA塩基配列解読においては、極めて低頻度な配列が発見されることがあり、その場合はその配列を持つ人が特定できることになる。
  - ゲノム解析法においては、多くの遺伝子座を測定するほど、より高い確率で個人を識別できるようになる。
3. ゲノムデータの為の生体試料は「ヒトゲノム・遺伝子解析研究に関する倫理指針」、或いは、OECD指針に基づいて作成された「遺伝子関連検査に関するベストプラクティスガイドライン」に従って入手されている筈なので、インフォームドコンセントを得る手続がある筈である。この段階で、ゲノムデータから個人を特定する可能性(一意性)についての説明が求められた場合には前項のような説明が可能である。