

## ヒトゲノム全体に含まれる遺伝子数への追加資料

### ■追加資料

#### ・ヒトゲノム全体に含まれるタンパク質をコードする遺伝子数について

タンパク質をコードする遺伝子数は、研究成果により常に変動しており、どの研究成果を参照するかによって数字が異なります。例えば、高等学校の教科書「生物基礎」では、ヒトゲノム全体に含まれるタンパク質をコードする遺伝子数として「約2万2000個」、「2万500個」、「約2万個」（教科書会社によって異なる）と記されています。タンパク質をコードする遺伝子数の現時点（2020年6月3日時点）での最新情報を確認したい方は以下資料をご参照ください。

#### 追加資料1：

Ensembleにおける「Human assembly and gene annotation」（下記）にアクセスしていただき、「Gene counts (Primary assembly)」における「Coding genes」をご覧ください。タンパク質をコードする遺伝子数が記されています。2020年6月3日現在で20,438個と記されています。

[http://www.ensembl.org/Homo\\_sapiens/Info/Annotation](http://www.ensembl.org/Homo_sapiens/Info/Annotation)

#### 追加資料2：

GENCODEにおける「Human: Statistics about the current GENCODE Release」（下記）にアクセスしていただき、「Protein-coding genes」をご覧ください。タンパク質をコードする遺伝子数が記されています。2020年6月3日現在で19,959個と記されています。

<https://www.genecodegenes.org/human/stats.html>

#### 追加資料3：

NCBIにおける「NCBI Homo sapiens Updated Annotation Release 109.20200522」（下記）にアクセスしていただき、「protein-coding」行の「GRCh38.p13 Primary Assembly」列をご覧ください。タンパク質をコードする遺伝子数が記されています。2020年6月3日現在で19,691個と記されています。

[https://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/Homo\\_sapiens/109.20200522/](https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Homo_sapiens/109.20200522/)

#### 【追加資料提示の背景】

今般追加資料を提示した理由は、本「一家に1枚ヒトゲノムマップ」の遺伝子数と、高等学校の教科書「生物基礎」などで示されている遺伝子数とが乖離してきたことにあります。

「1家に1枚ヒトゲノムマップ」においては、ヒトゲノム全体に含まれる遺伝子数について「ヒトゲノム全体に含まれる遺伝子の数は、約4万300個—約3万7300個※（注：本訂正については別紙「お詫びと訂正」をご覧ください）であると推定されています。※複数の推定値のうち、総遺伝子数が最も多いデータベースに基づく数値です。」と記載し、またそのリファレンスとして「Homo sapiens Genome:Statistics, Annotation Release 104 statistics (NCBI)」を参照しています。本リファレンスは、下記から御覧いただけます。

<https://www.ncbi.nlm.nih.gov/projects/mapview/stats/BuildStats.cgi?taxid=9606&build=104&ver=0>

本リファレンス結果では、「Genes」（遺伝子）としてタンパク質をコードする遺伝子以外も含まれています。本「一家に1枚ヒトゲノムマップ」においては中央箇所「染色体・DNA・遺伝子・タンパク質」の注釈として『※タンパク質をつくり出さない非遺伝子領域は「ジャンク（がらくた）DNA」などとよばれていましたが、こうした部位にも、重要な機能が隠されていることがわかりはじめてきました。』と記載してきたことなどから、2006年の第1 版出版時に上記NCBI によるデータベースにおける各染色体の“Genes”を遺伝子数として提

示することとしました（第1版では総計約2万6800個と提示）。その後、当方針に則り2013年まで改訂を重ねてきました。しかしながら、研究の進展に伴いタンパク質をコードしない遺伝子数が増加しており、総遺伝子数が、より一般的なタンパク質をコードする遺伝子数と乖離していくこととなりました。すでに述べた通り、高等学校の教科書「生物基礎」では、ヒトゲノム全体に含まれる遺伝子数として「約2万2000個」、「2万500個」、「約2万個」（教科書会社によって異なる）と記されていますが、これらはヒトゲノム全体に含まれるタンパク質をコードする遺伝子数です。

このような背景を踏まえ、タンパク質をコードする遺伝子数を確認する手段として、上記追加資料1～3を提示させていただきました。

以上