

2004年10月21日

国際ヒトゲノムシーケンシングコンソーシアム

独立行政法人 理化学研究所

慶應義塾大学

国立遺伝学研究所

高精度ヒトゲノムに関する学術論文の発表

日、米、英、仏、独、中の6か国18研究センター等から構成される、国際ヒトゲノムシーケンシングコンソーシアム(IHGSC: International Human Genome Sequencing Consortium)は、ヒトゲノムの高精度配列の最終的な検証と解析を行い、その成果を学術論文としてとりまとめました。論文のとりまとめにあたり、我が国においては、独立行政法人理化学研究所(野依良治理事長)ゲノム科学総合研究センター(GSC)の榊佳之センター長(国際ヒトゲノム機構前会長)らが中心となり、配列の完成及び解析など今回の成果に大きく貢献しました。

ヒトの遺伝設計図であるヒトゲノムの全解読は、人類科学の最も輝かしい成果のひとつとして歴史に残るものです。このヒトゲノムの全解読を目指す「ヒトゲノム計画」は国際協調のもと、1991年より進められ、2003年4月には、本コンソーシアム及び日、米、英、仏、独、中の6か国首脳共同によるヒトゲノム解読完了が宣言されています。その後コンソーシアムは1年以上をかけて得られた高精度配列に対して最終的な検証と解析を進めました。今回の論文はその結果についての報告であり、より正確な構造及び遺伝子組成などのヒトゲノムの特徴解明が行われました。

今回、本ゲノムの99%以上をカバーする約29億塩基対のシーケンズデータを解析した結果、ヒトのタンパク質をコードする遺伝子数は約2.2万個であることなどゲノムの新しい特徴が明らかになりました。なお、ギャップと呼ばれる未決定領域は341か所のみとなりました。

今回の成果によって、ヒトを理解する上で最も基盤となる情報が確定したことになり、今後は医学・医療をはじめとする生命科学の発展が大きく加速化されると期待されています。

本研究成果の詳細は、英国の科学雑誌『Nature』(10月21日号)に掲載されます。

1. 背景

2001年2月、国際ヒトゲノム^{*1}シーケンシングコンソーシアム(IHGSC: International Human Genome Sequencing Consortium)^{*2}と米国の私企業セレーラ・ジェノミクス社は、それぞれヒトゲノムシーケンスのドラフト完成を発表しました。そこではヒトの遺伝子総数が3万~4万、ヒトゲノムの半分は、繰り返しの配列であること、ゲノムには多数の個人差(SNP)が存在すること等、ヒトゲノムの全体像が浮かび上がりました。しかしドラフト配列はあくまでも予備的データであったため、ヒトゲノム全体の約10%の解読は未着手で、15万か所のギャップ(未決定領域)、目標とされる精度99.99%より一桁低い99.9%精度のデータ、さらに決定領域にも遺伝子の位置の間違いや配列の決失など解決すべき重大な課題も多く

残されていました。ゲノム中のわずか1塩基の違いを解析するような病気の診断などの私たちの遺伝子の研究には、基本となるヒト配列データの精度はできるかぎり高くなければなりません。もしそこに多くの間違いがあれば、得られる解析結果の信頼性は低くなる上に、その検証のために多くの時間と予算も無駄にしています。

その後、ヒトゲノムの完全解読を目指して更なる努力が続けられ、既に1999年12月に22番、2000年5月に21番、2001年12月に20番、2003年2月に14番染色体の完全解読については論文発表されてきたが、その他の染色体も99.99%以上の高精度に読み終え、2003年4月解読完了が宣言されています。その後コンソーシアムは得られた高精度配列に対して最終的な検証と解析をさらに1年以上をかけて進めました。今回の論文はその結果についての報告であり、より正確な遺伝子発見、構造及び遺伝子組成などのヒトゲノムの特徴解明がなされました。なお国際コンソーシアムは約2800名の研究者で構成されています。

2. 研究成果

ドラフト配列には未決定領域や低精度箇所以外に、配列の間違ったつなぎ方、方向、染色体上の配置、配列の欠失、これらに伴った間違っただ遺伝子の配置と方向など目に見えない精査すべき領域が多くありました。今回の高精度配列の決定と解析では、全体精度の評価とともにこれら間違っただ領域の検出とその修正、訂正等の作業が行われました。その結果、シーケンス対象である真正クロマチン領域（遺伝子が存在する領域）の99%以上に相当する約2,850Mbの高精度（99.999%）ゲノム配列が決定され、そこに存在するタンパク質をコードした遺伝子及び構造上の特徴が精度高く解析されました。真正クロマチン領域でのクローン不在領域（未決定領域）は250か所、サイズにして約25Mbとなりました。セントロメア領域等シーケンス対象にならなかった箇所を含めると未決定領域は341ヶ所になりました。確定されたタンパク質をコードする遺伝子の数は約2.2万個となり、その中にはマウスやラットにはない遺伝子（免疫、臭覚、生殖関連）が含まれます。一方ヒトにおいて機能を失った遺伝子も見つかりました。しかし、いまだに埋められないギャップが残っており、これについては各国が今後も引き続き努力をしていきます。今回のデータは国立遺伝学研究所日本DNAデータバンク（DDBJ）などの公的DNAデータバンクより公開されています。既に高精度ヒトゲノム情報は、ヒトの病気克服をめざしているHapMapプロジェクト^{*3}や、遺伝子発現やタンパク質ネットワーク研究などのより高次な生命現象の解明をめざした国家プロジェクトであるゲノムネットワークプロジェクト（日本）及びENCODE計画（米国）^{*4}の基盤として活用されています。

3. 今後の展開

今回の成果によって、ヒトを理解する上で最も基盤となる情報が確定したことになります。今後は医学・医療をはじめとする生命科学の発展が大きく加速化されると期待されるだけでなく、食料・人口問題、産業・経済活動、さらには法律、教育など私たちを支える様々な学問や社会活動に今後大きな変革をもたらすであろうと考えられます。

また、ゲノムは生物が生きる仕組みを書き込んだ指示書であるとともに、進化を経て今日に至った進化適応のプロセスが書かれた歴史書でもあります。生物が進化の過程をどのように生き抜いて来たのか、そして現在どのような仕組みで繁栄を保っているのか、その「生命戦略」とも呼ぶべき生命の智慧を解き明かし、それを人類の繁栄に役立てることがポストシーケンス時代のゲノム研究の目標です。

(問い合わせ先)

独立行政法人理化学研究所 横浜研究所
ゲノム科学総合研究センター

センター長 榊 佳之

Tel : 045-503-9171 / Fax : 045-503-9170

(報道担当)

独立行政法人理化学研究所 広報室

Tel : 048-467-9272 / Fax : 048-462-4715

Mail : koho@riken.jp

<補足説明>

※1 ヒトゲノム

ヒトの持つ遺伝情報の一セットのこと。DNAの塩基配列として約31億塩基となる。このうち真正クロマチンと呼ばれ遺伝子等の情報を含む約29億塩基が今回の解析対象となっている。単純配列の繰り返しからなるヘテロクロマチンと呼ばれる約2億塩基の部分は今回の解析対象となっていない。

※2 国際ヒトゲノムシーケンシングコンソーシアム(IHGSC: International Human Genome Sequencing Consortium)

米国、英国、フランス、ドイツ、日本そして中国の公的研究機関の連合

※3 HapMap プロジェクト

ヒトゲノム上の多型情報を医学応用していくために不可欠である、疾患感受性、薬剤に対する効果や副作用に関与する遺伝的要因を発見する上で、役立つ研究ツールとなる地図（ハプロタイプ地図）の作成に関する国際的な枠組み。

※4 ENCODE 計画

米国が、ヒトゲノム解読完了後の新規ゲノム関連大規模プロジェクトとしてスタートさせた、ヒトゲノム上のすべての機能の解明を目指す計画。

<コンソーシアム参加機関>

1. The Wellcome Trust Sanger Institute, United Kingdom
2. Washington University Genome Sequencing Center, USA
3. Whitehead Institute for Biomedical Research, Center for Genome Research, Nine Cambridge Center, USA.
4. US Department of Energy Joint Genome Institute, USA
5. Baylor College Of Medicine Human Genome Sequencing Center, USA
6. RIKEN Genomic Sciences Center, Japan
7. University of Washington Genome Center, USA
8. Genoscope and CNRS UMR-8030, France
9. Department of Molecular Biology, Keio University School of Medicine, Japan
10. Genome Therapeutics Corporation, USA. Current address: Agencourt Bioscience Corp 100 Cummings Center, USA
11. Institute for Systems Biology, USA
12. Department of Genome Analysis, Institute of Molecular Biotechnology, Germany
13. Beijing Genomics Institute, China
14. Max Planck Institute for Molecular Genetics, Germany
15. Department of Genome Analysis, GBF – German Research Centre for Biotechnology., Germany
16. Advanced Center for Genetic Technology, Applied Biosystems Division of Perkin-Elmer Corp., USA
17. Analysis group. Individuals listed below in alphabetical order performed analyses discussed in the manuscript, and/or were responsible for assembling and maintaining the finished sequence in the public database. Includes individuals listed under other headings. (Department of Computer Science, University of California at Santa Cruz, USA ; National Center for Biotechnology Information National Institutes of Health USA; MRC Functional Genetics Unit, University of Oxford, Department of Human Anatomy and Genetics, UK; Department of Genetics, Case Western Reserve University, USA; Wistar Institute, USA)

18. Scientific management National Human Genome Research Institute, National Institutes of Health, USA

Latest summary of HGS Build 35 (July 2004)

Chr	Total finished sequence (kb)	Euchromatic gaps		Heterochromatic gaps		Estimate of total gap size (kb)	Unfinished clones	
		Number	Est. size (kb)	Number	Est. size (kb)		Number	Est. size (kb)
1	222,828	32	1605	2	19510	21115	17	850
2	237,503	20	2512	1	2900	5412	0	0
3	194,636	5	1935	1	1500	3435	0	0
4	187,161	14	1250	1	3000	4250	0	0
5	177,703	5	92	1	340	432	0	0
6	167,318	10	658	1	2300	2958	0	0
7	154,759	11	869	1	4630	5499	0	0
8	142,613	9	662	1	2190	2852	0	0
9	117,781	40	1955	2	18000	19955	12	600
10	131,614	12	1020	1	2515	3535	8	400
11	131,131	7	322	1	4760	5082	0	0
12	130,259	8	795	1	4300	5095	0	0
13	95,560	6	715	2	17200	17915	0	0
14	88,291	1	8	2	17220	17228	0	0
15	81,342	10	737	2	18260	18997	0	0
16	78,885	4	143	2	10000	10143	0	0
17	77,800	9	875	1	7500	8375	0	0
18	74,656	3	97	1	1368	1465	0	0
19	55,786	5	5015	1	340	5355	0	0
20	59,505	4	1157	1	1766	2923	0	0
21	34,170	3	53	2	11620	11673	0	0
22	34,765	11	460	2	14330	14790	0	0
X	150,394	12	750	1	3000	3750	14	700
Y	24,872	9	1480	2	31618	33098	7	350
TOTAL	2,851,331	250	25165	33	200167	225332	58	2900

