

平成28年10月11日 (火)  
17:00-18:30

神戸大学  
計算化学教育センター

## 計算生命科学の基礎Ⅲ

ゲノムに記された遺伝ビッグデータを読む  
ーヒトゲノム計画から大規模個人ゲノム解読時代の到来までー

岩手医科大学 いわて東北メディカル・メガバンク機構  
生体情報解析部門 部門長代理 特命教授  
清水 厚志



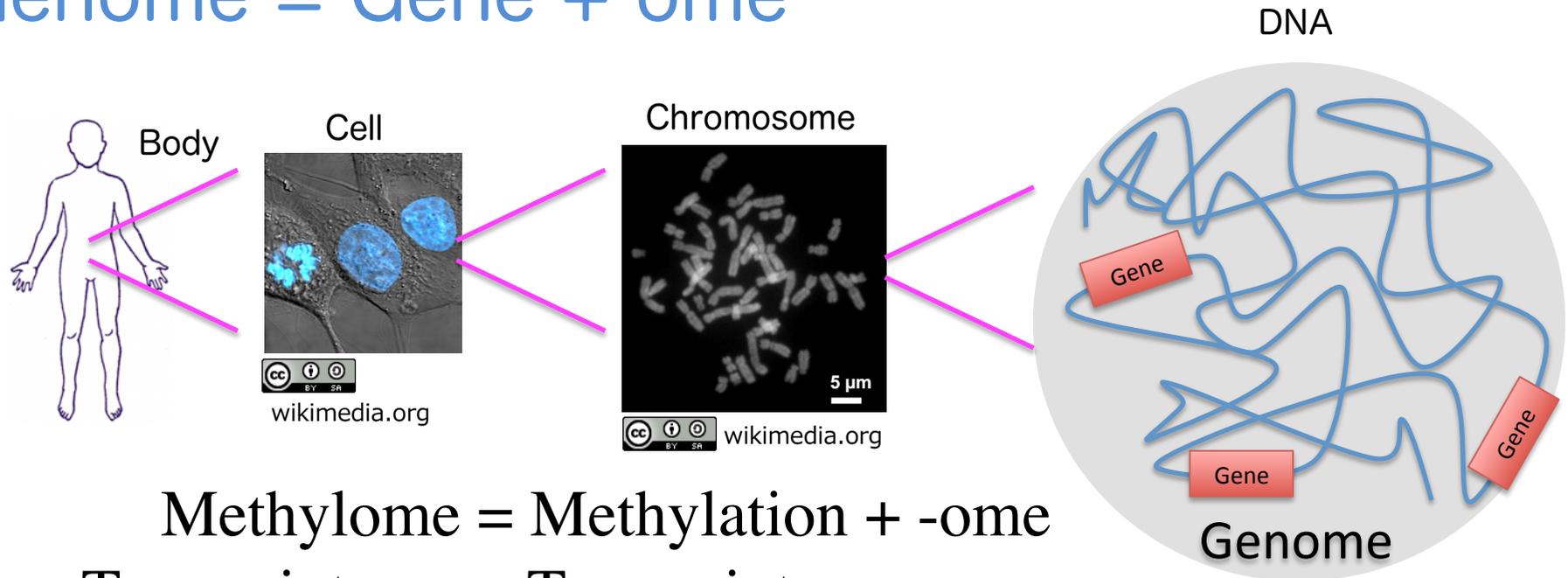
いわて東北メディカル・メガバンク機構  
IWATE TOHOKU MEDICAL MEGABANK ORGANIZATION

- ヒトゲノム計画
- 次世代シーケンサー概要
- 個人ゲノム解読
- ゲノム解読の課題

# ヒトゲノム

---

Genome = Gene + ome



Methylome = Methylation + -ome

Transcriptome = Transcript + -ome

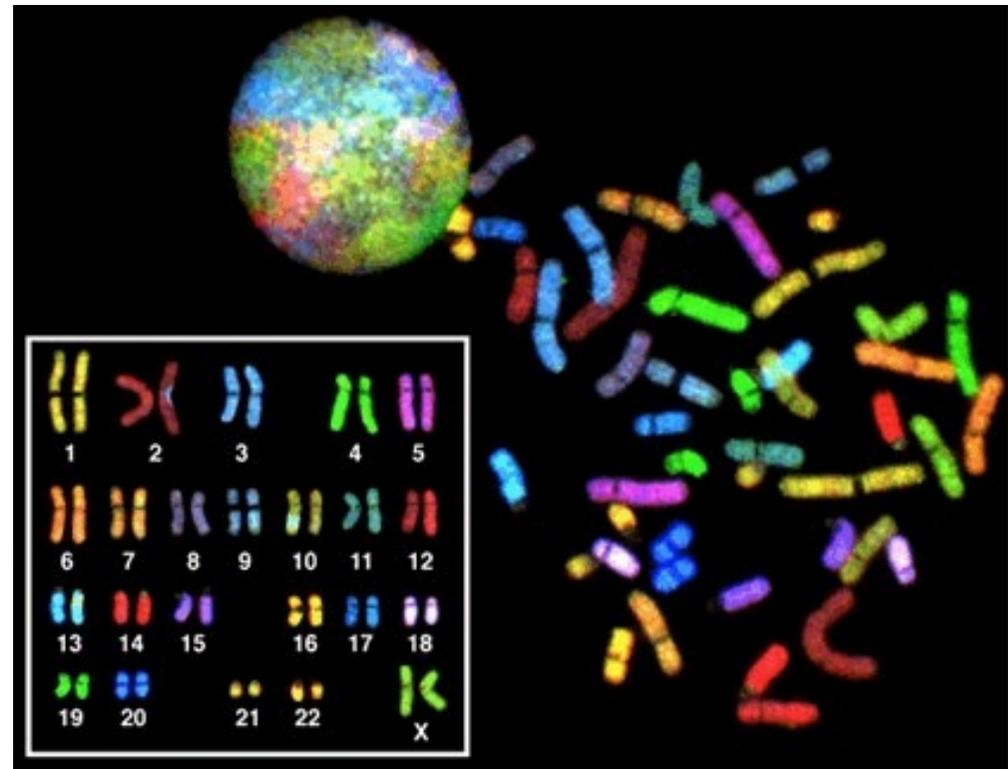
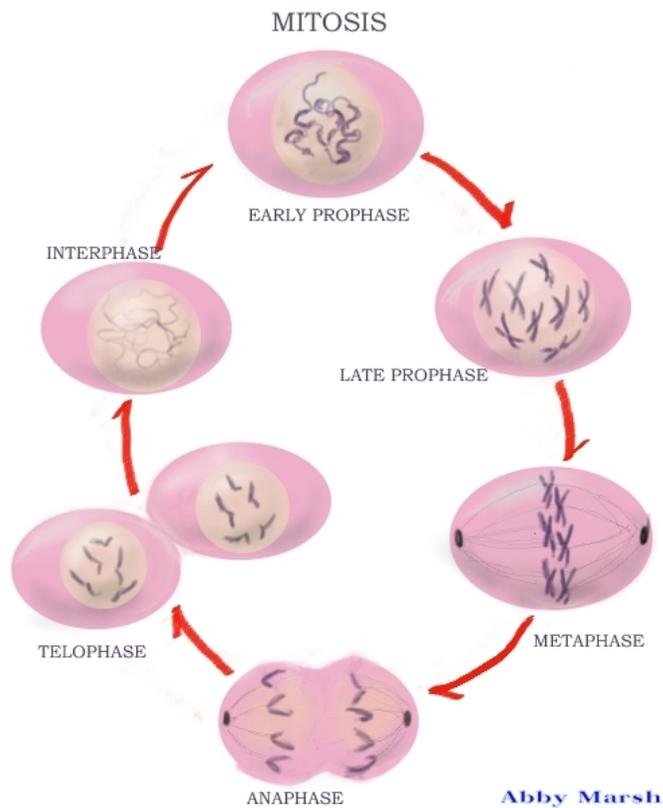
Proteome = Protein + -ome

Metabolome = Metabolism + -ome

and kinome, lipidome, glycome etc ...

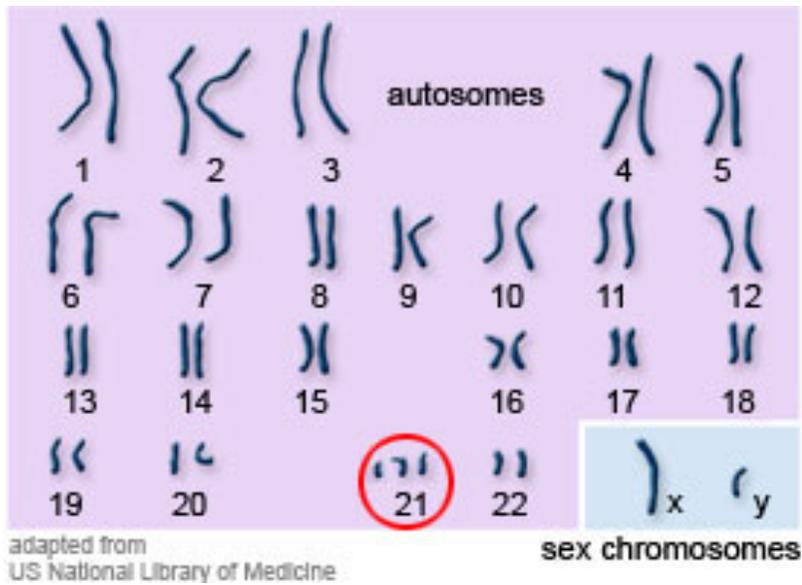
omics

$2n = 46$



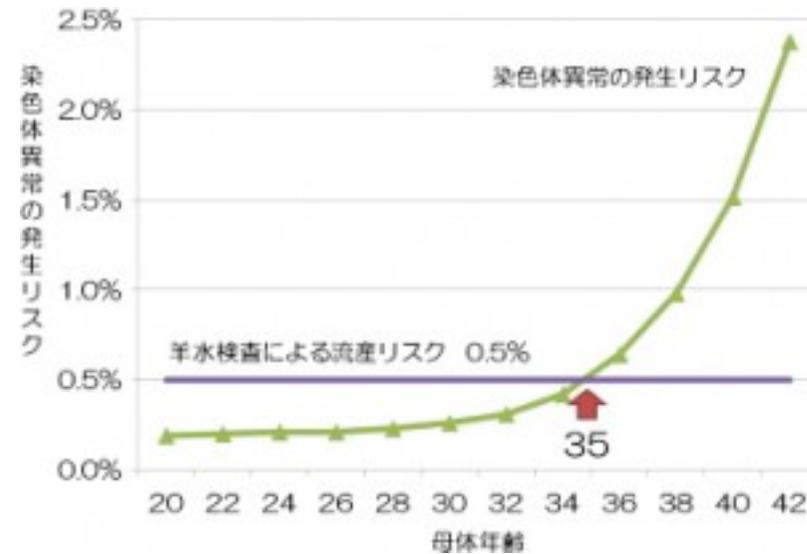
© from National Institutes of Health

## ダウン症候群 (47,XY,+21)



© from National Institutes of Health

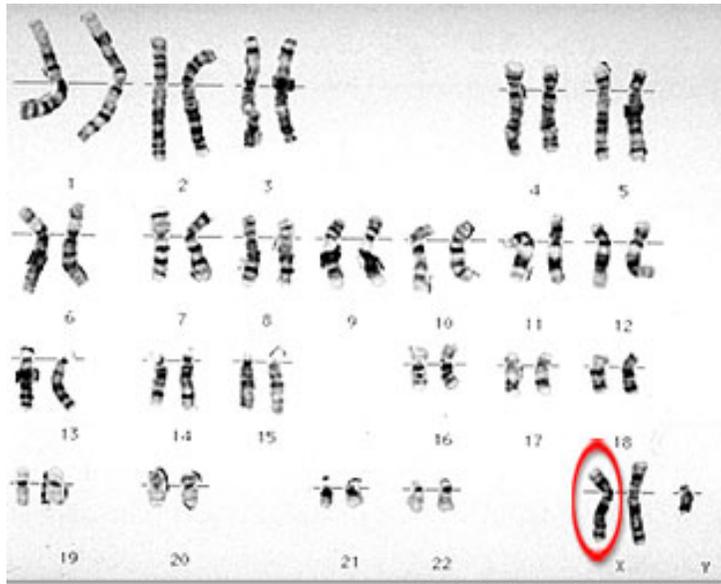
691 出産に 1 件 (米国)



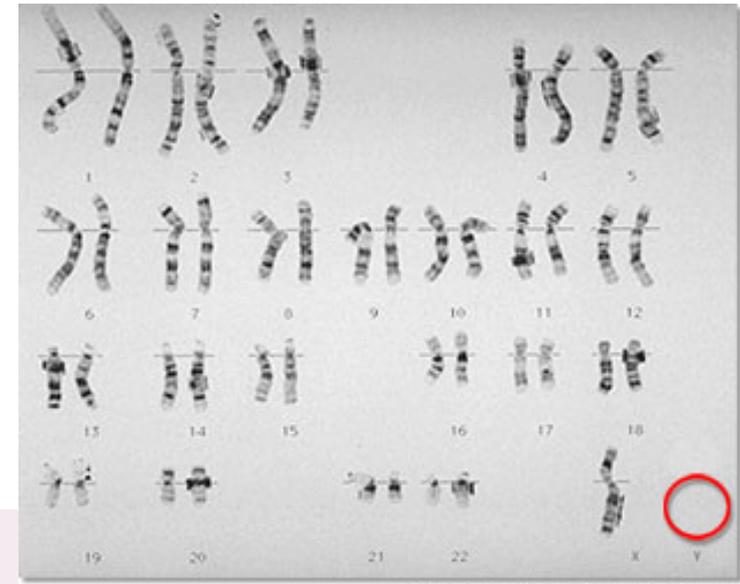
<http://www.prenatal-diagnosis.info/> Oct. 4 2016

## クラインフェルター症候群 47, XXY

## ターナー症候群 45, XO



medgen.genetics.utah.edu



medgen.genetics.utah.edu

1/500 to 1/1000  
live male births

1/2000 to 1/5000  
phenotypic females

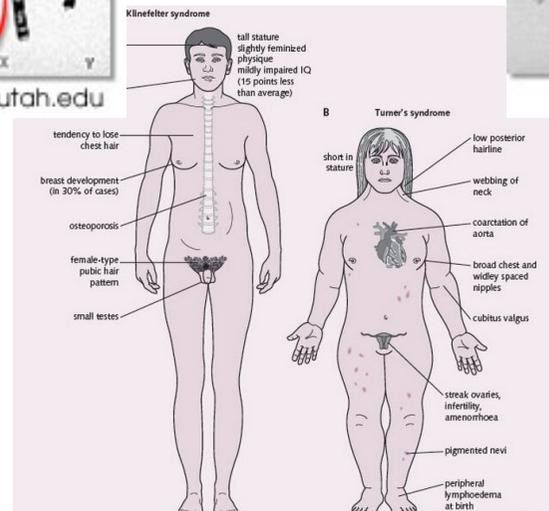
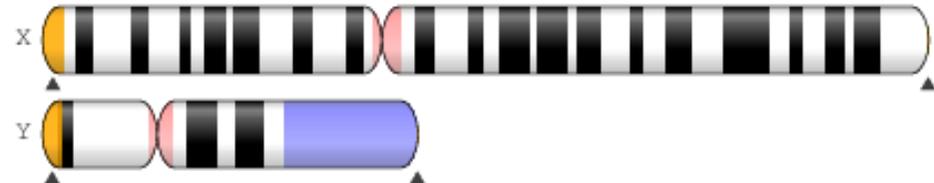
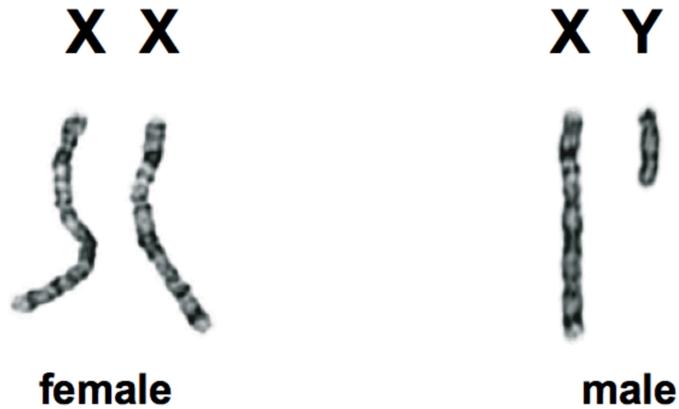


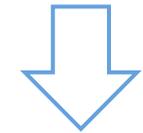
Image source : genmed.yolasite.com



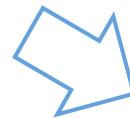
▲ PAR; Pseudoautosomal region

<http://genomeref.blogspot.jp/2013/09/tech-tip-pseudo-autosomal-region-of.html> Oct.4 2016

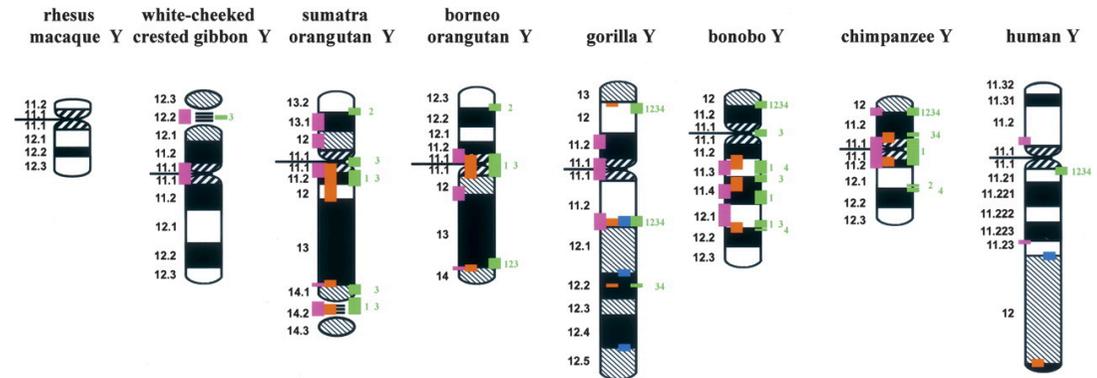
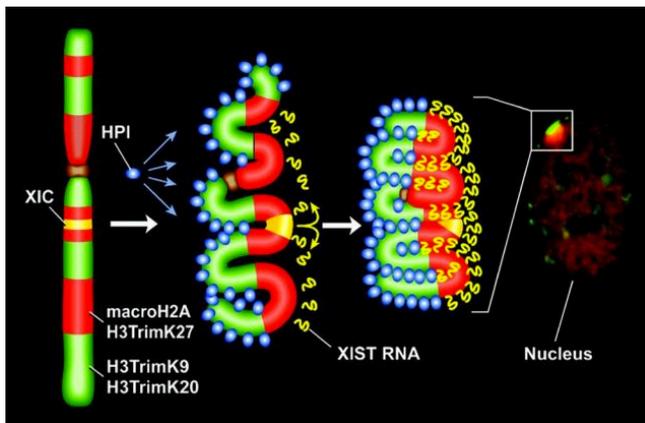
CC BY-NC-SA by Jenny Graves



X染色体不活化により遺伝子機能としては1本として扱える。



ゲノム解析の際には XとYを区別する必要がある。



Genome Res. 2008 Jul; 18(7): 1030–1042.

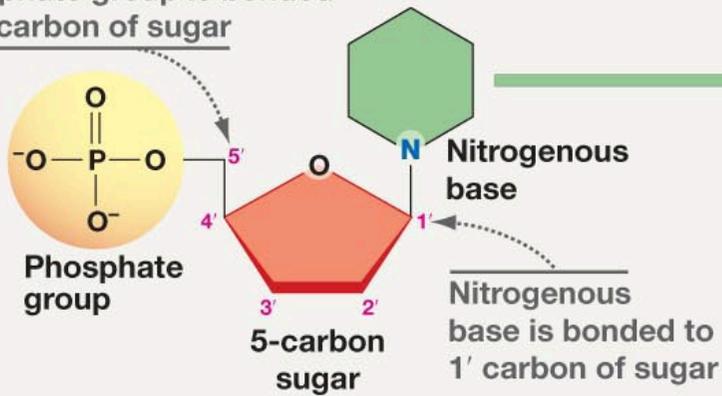
# DNAシーケンシング

---

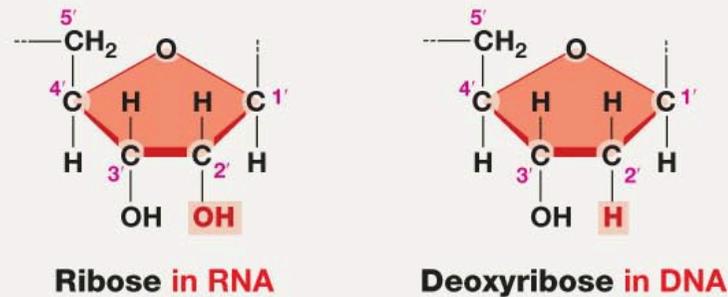
# A C G T / U

## (a) Nucleotide

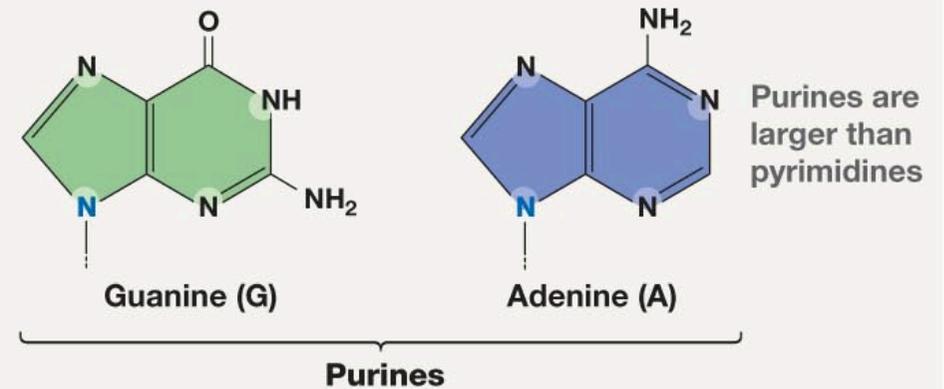
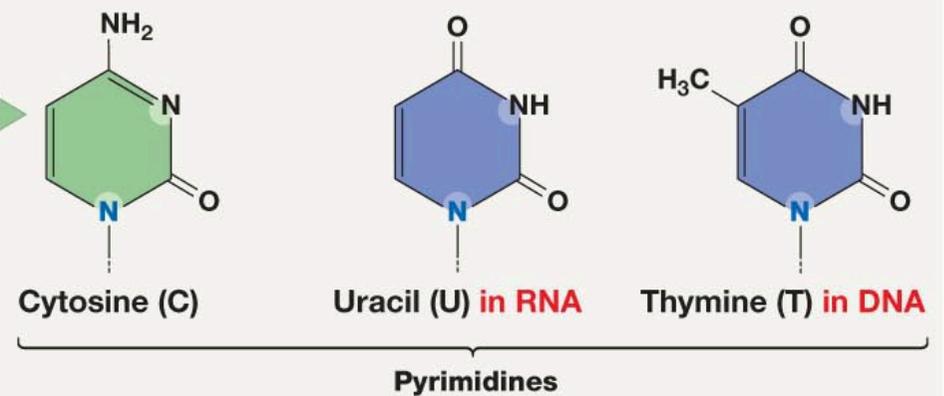
Phosphate group is bonded to 5' carbon of sugar



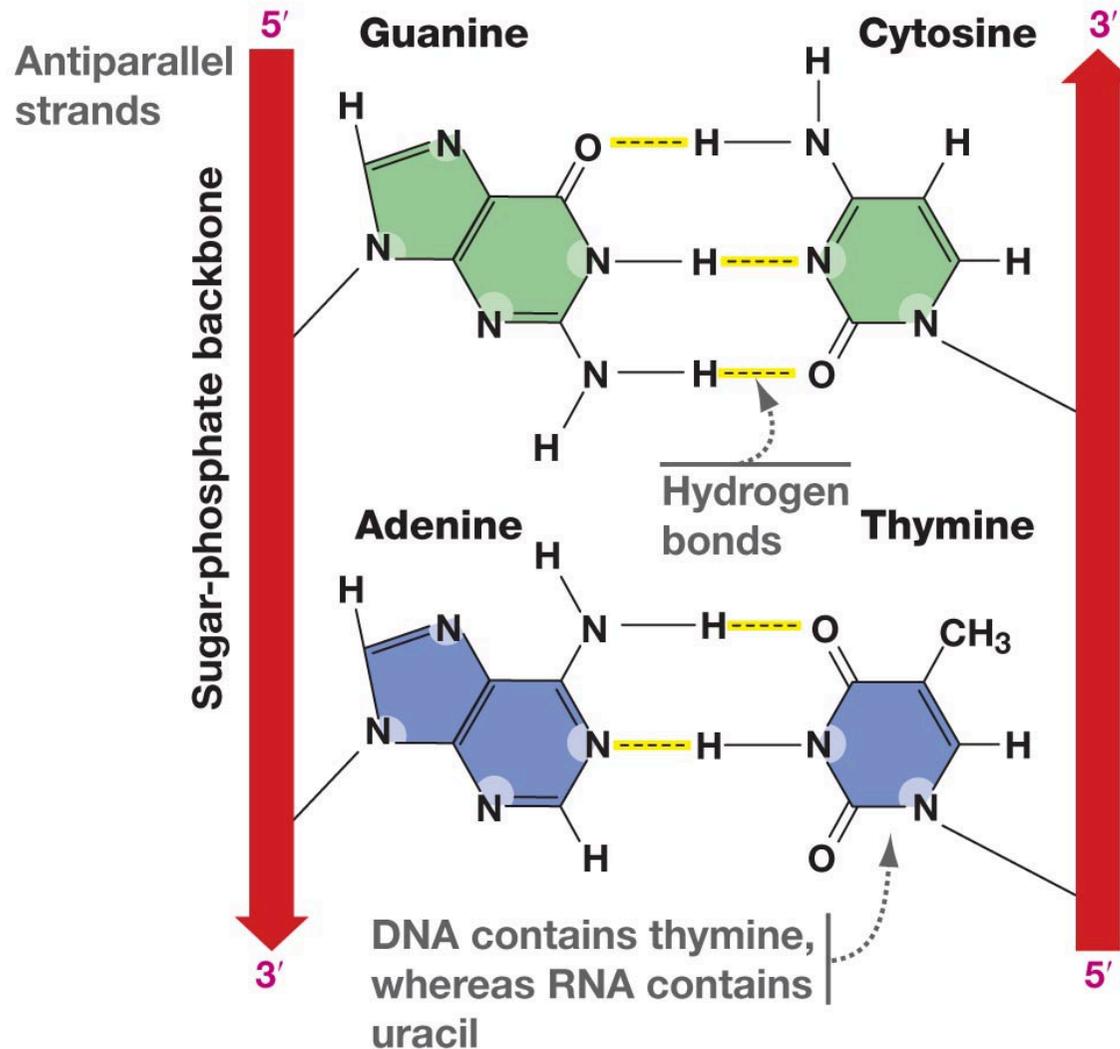
## (b) Sugars



## (c) Nitrogenous bases



## (b) Hydrogen bonds form between G-C pairs and A-T pairs.



水素結合  
A-Tは2本  
G-Cは3本

二重らせんの  
安定性は水素  
結合によって  
維持

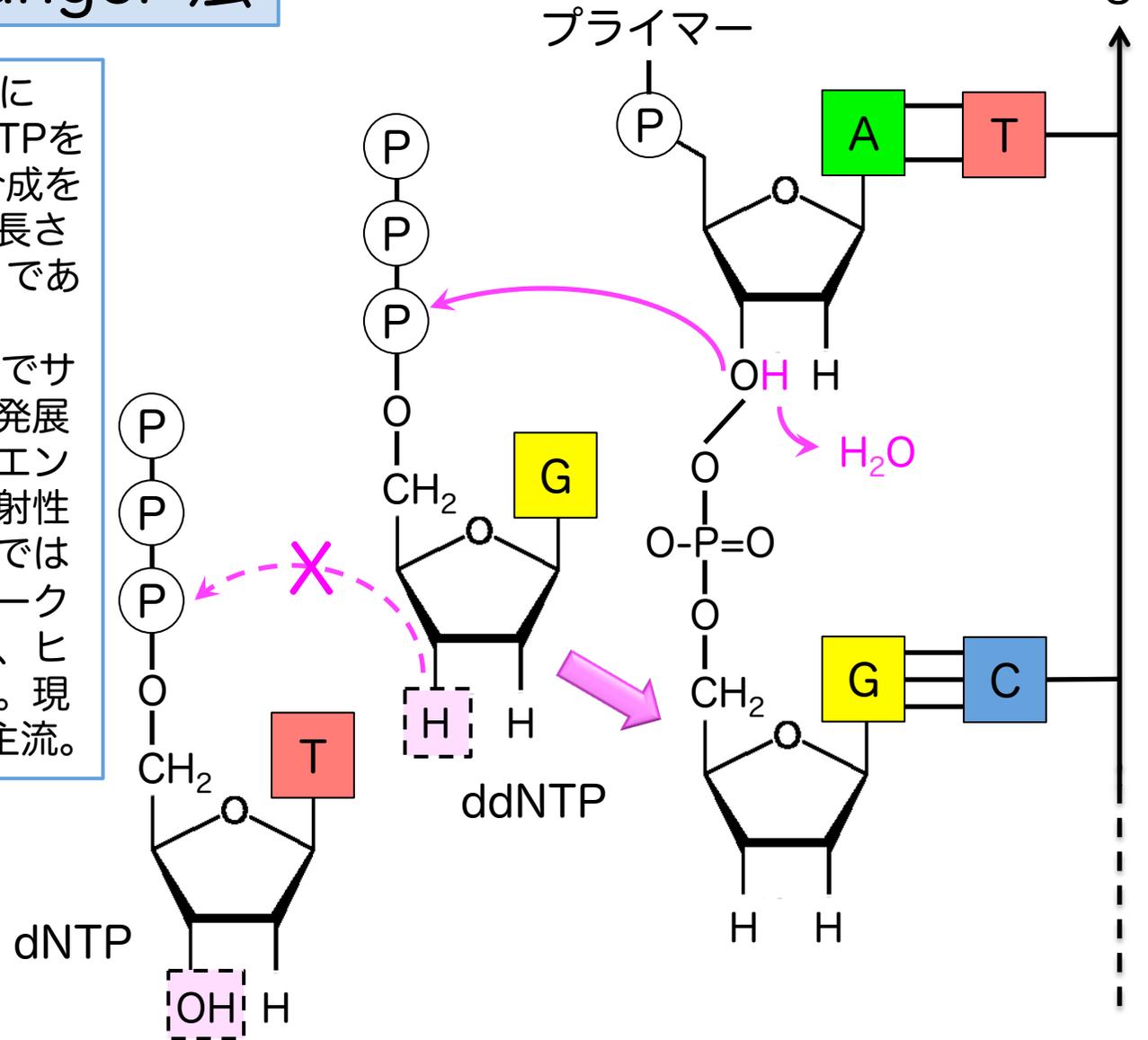
## Dideoxy 法、Sanger 法

鋳型DNA

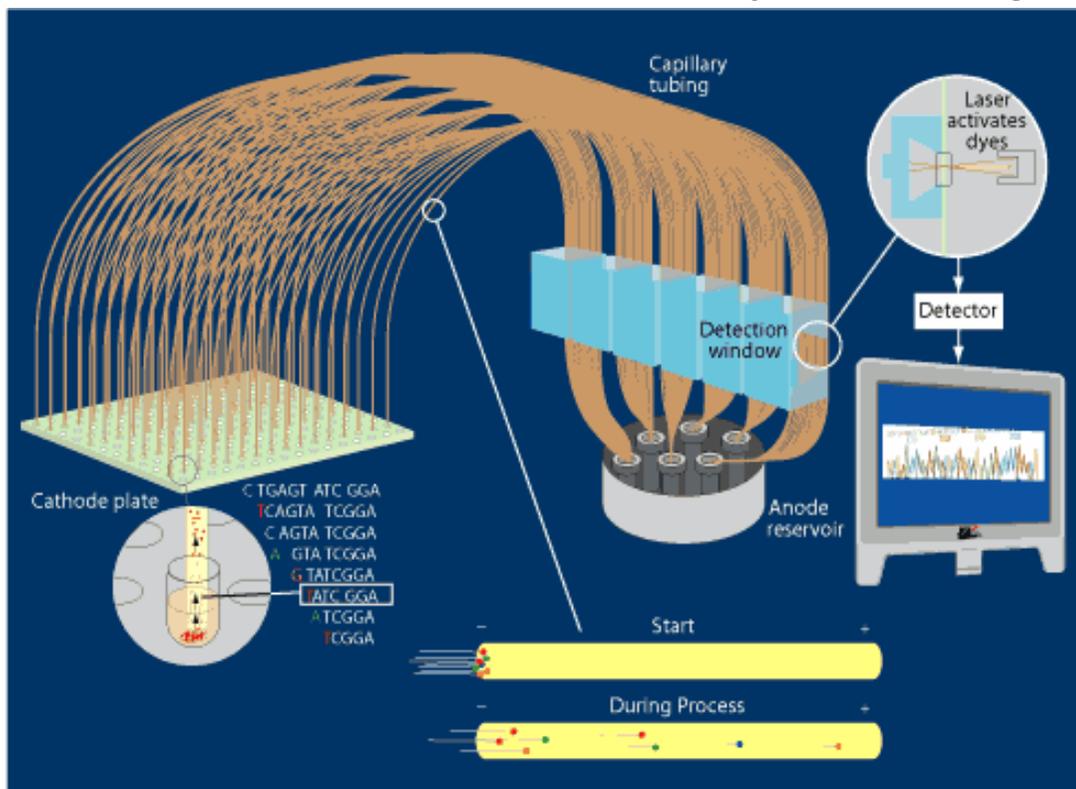
3'

DNA鎖合成に必要なdNTPに3'-OH基を欠いた一定量ddNTPを混ぜることで、「DNA鎖の合成をランダムに終了させ、異なる長さのDNA鎖を合成させる方法」である。

PCRの原理を組み込むことでサイクルシーケンシング法に発展し、微量の試料からのシーケンシングが可能となった他、放射性ラベルから蛍光ラベル、ゲルではなくキャピラリーを用いるシーケンサーが開発されたことで、ヒトゲノム計画の主流となった。現在ではDye terminator 法が主流。

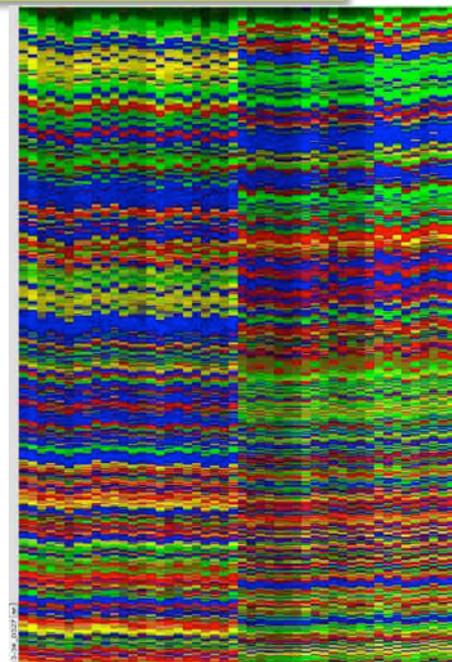


## Chain-termination法 (dideoxy 法、Sanger 法)

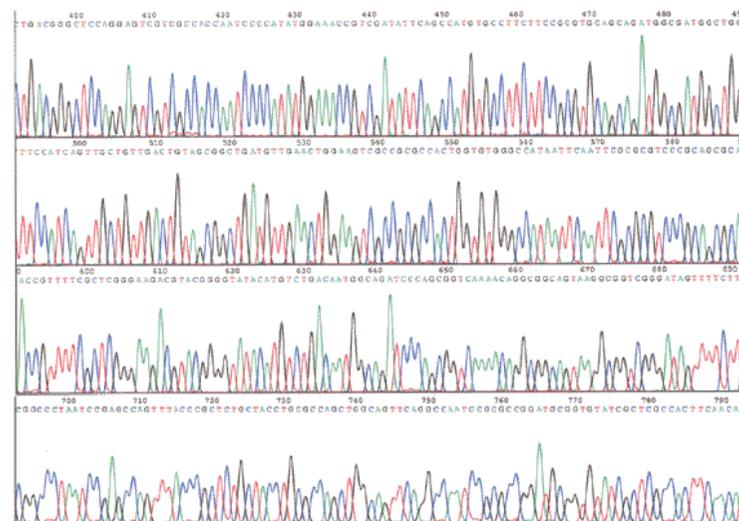


<http://jgi.doe.gov/archived-educator-resources/sanger-sequencing-archive/how-sanger-sequencing-is-done/step-10-capillary-sequencing/> Oct.4 2016

キャピラリーと呼ばれる極細の金属チューブに粘性の低いゲルをつめることで平板のゲルの作成が必要なくなった。自動化が進み、解読速度が大幅に向上した。



[http://www.adfg.alaska.gov/index.cfm?adfg=fishinggeneconservationlab.microsatellite\\_dna\\_analysis](http://www.adfg.alaska.gov/index.cfm?adfg=fishinggeneconservationlab.microsatellite_dna_analysis) Oct.4 2016



by sigmaaldrich

第一世代（キャピラリーシーケンサー）と次世代シーケンサーの違い  
塩基配列決定時に電気泳動を必要としない。

第二世代と第三世代シーケンサーの違い  
鋳型のPCR増幅が必要ない。

第三世代と第四世代シーケンサーの違い  
蛍光色素を使わない。

## 次世代シーケンサー

### 第二世代

Roche



Illumina



Life Tech



Complete Genomics

QIAGEN



Azco Bio



### 第三世代

Helicos



PacBio

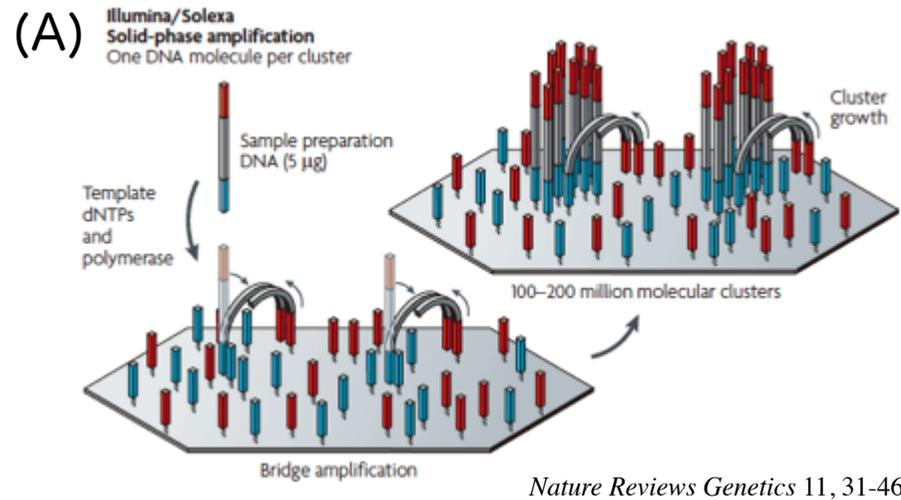


### 第四世代

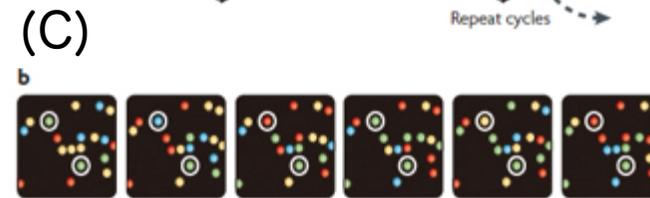
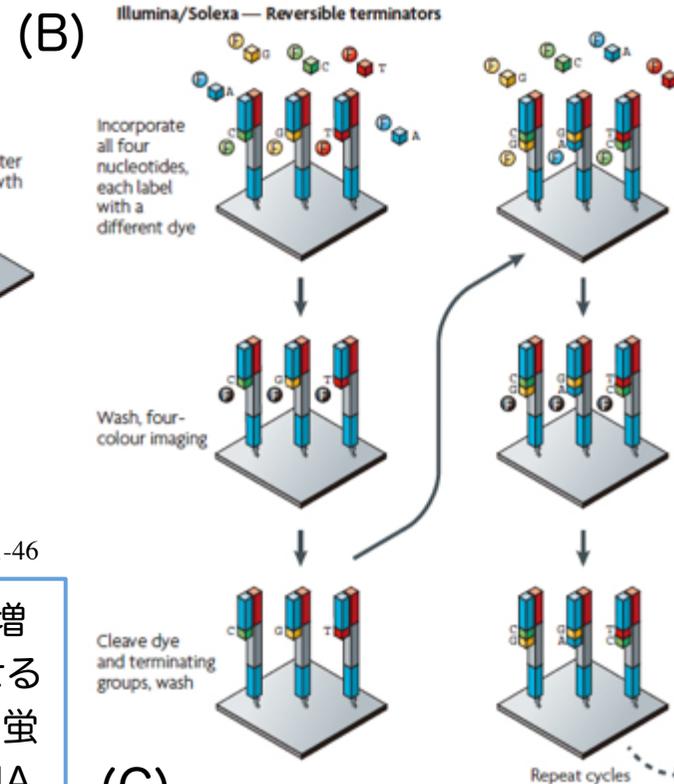
Oxford



## Polony Sequencing 法



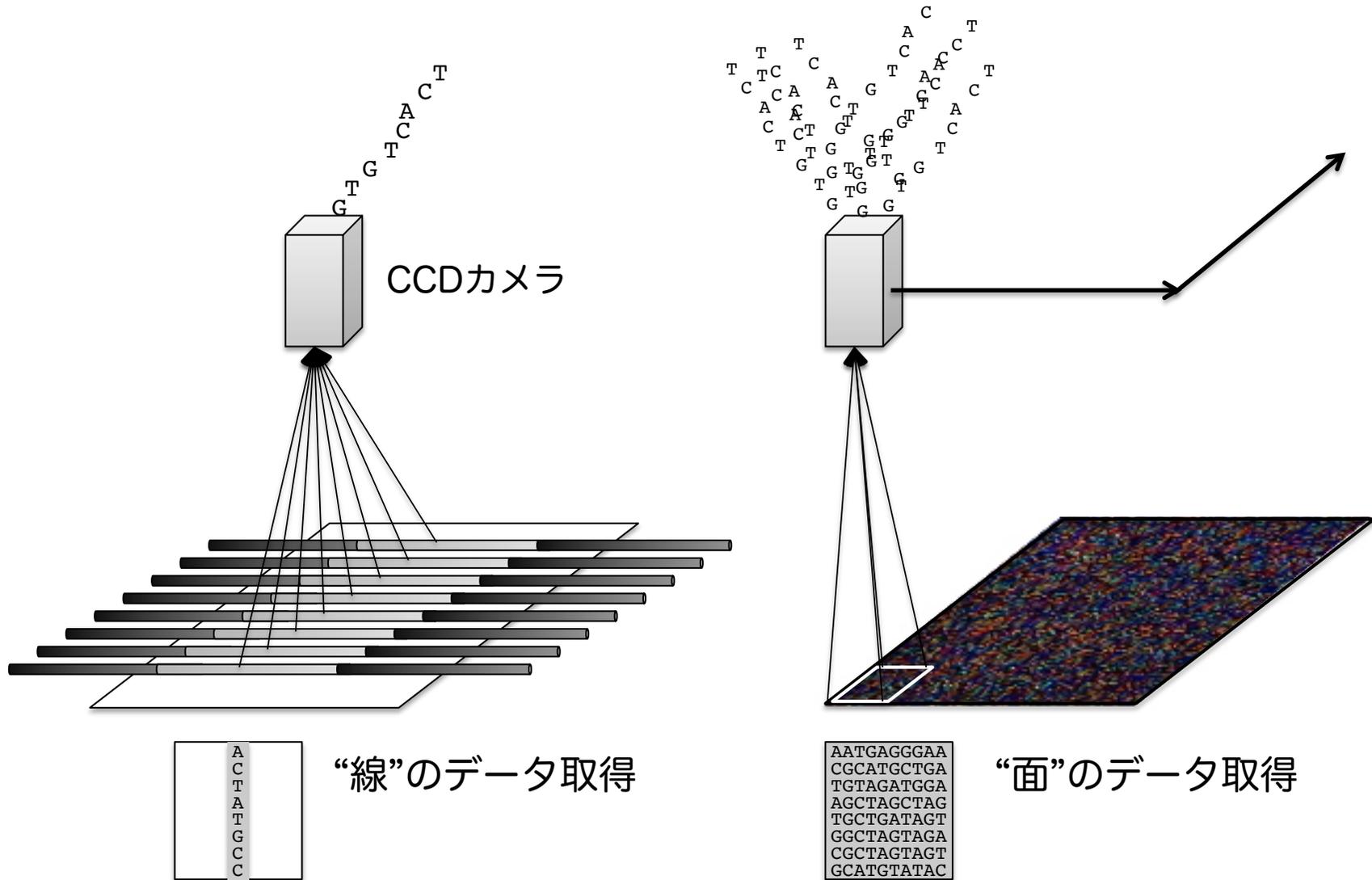
プライマーが結合した基盤上で鋳型をPCRにより増幅し、鋳型DNAのクラスター (Polony) を形成させる (A)。増幅した鋳型にプライマーをアニールさせる。蛍光ラベルした4種類の塩基を順次加え、撮影したDNAポリメラーゼの1反応ごとのイメージを配列に変換する (B, C)。世界で最も使用されている次世代シーケンサーである。

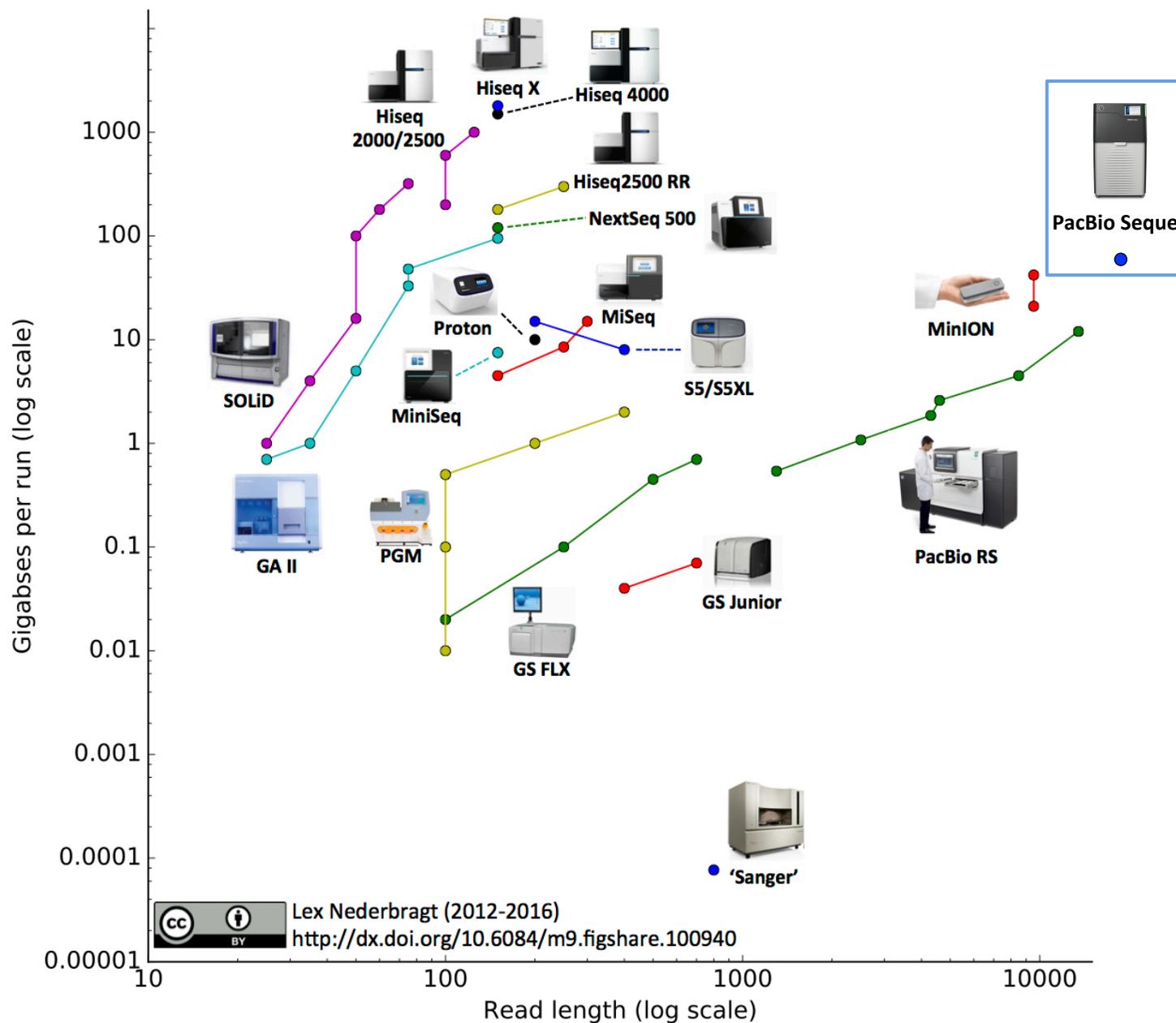


C G A T  
 T G C A  
 Top: CATCGT  
 Bottom: CCCCCC  
**HiSeq2500**



## キャピラリーシーケンサーと次世代シーケンサーの違い

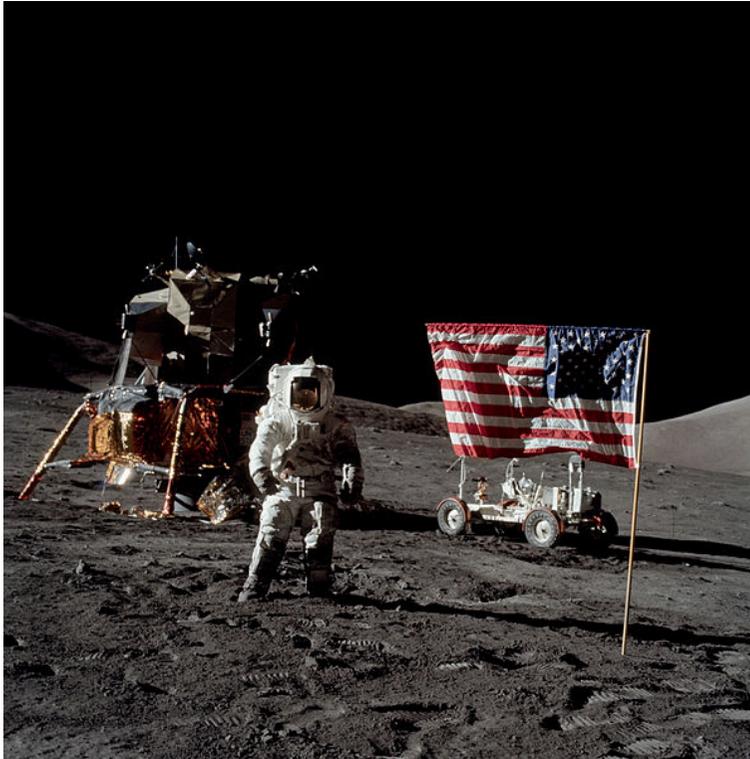




# ヒトゲノム計画

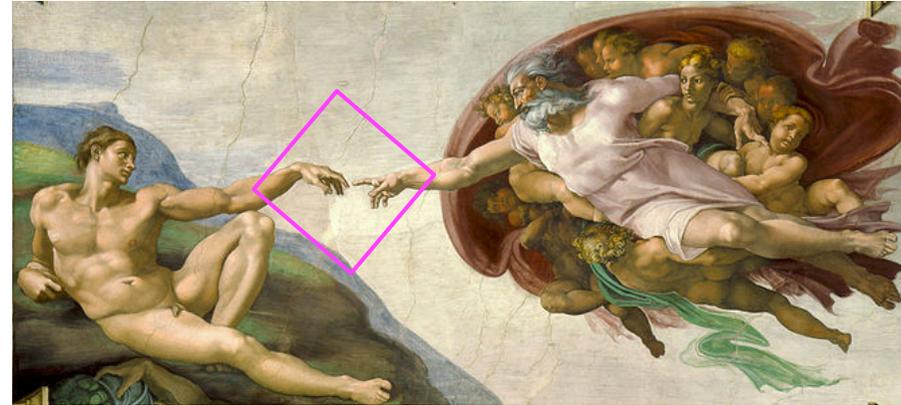
---

## Apollo program (1961-1972)



\$25.4-billion

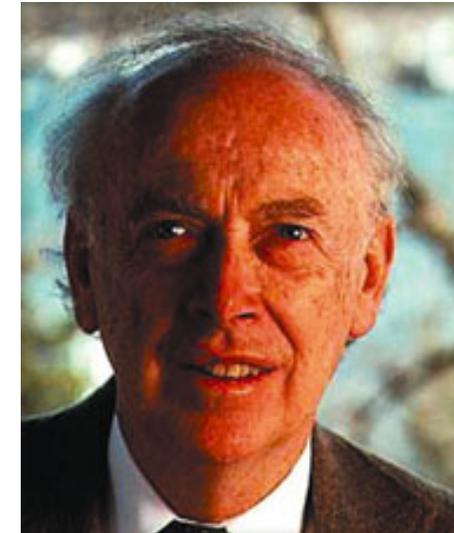
## Human Genome Project (1991-2003)



\$2.7-billion



- 1928年 形質転換因子の発見 (グリフィス)
- 1945年 形質転換因子がDNAであることを証明 (エイブリーら)
- 1952年 DNAが遺伝物質であることの証明 (ハーシーとチェイス)
- 1953年 **DNA 2重らせんの発見 (ワトソンとクリック)**
- 1968年 遺伝暗号表の完成 (コラーナら)
- 1970年代 DNA配列決定法の確立 (サンガー、マキサム、ギルバート)
- 1984年 ヒトゲノム計画発案 (シンスハイマー)
- 1985年 ヒトゲノム計画初の会合 (シンスハイマー)
- 1987年 DOE (エネルギー省) からヒトゲノム解析予算獲得 (デリシ)
- 1988年 ヒトゲノム解析機構 (HUGO) の創設 (ワトソン)  
National Center for Human Genome Research の創設  
NIH (アメリカ国立衛生研究所) からゲノム解析予算獲得 (ワトソン)



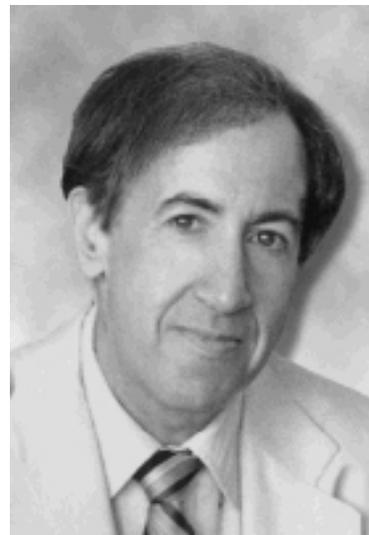
James D. Watson



19

Robert L. Sinsheimer

<http://oralhistories.library.caltech.edu/>



Charles DeLisi

<http://www.osti.gov/>



The Human Genome Organisation

# ヒトゲノム解読の歴史 (2)

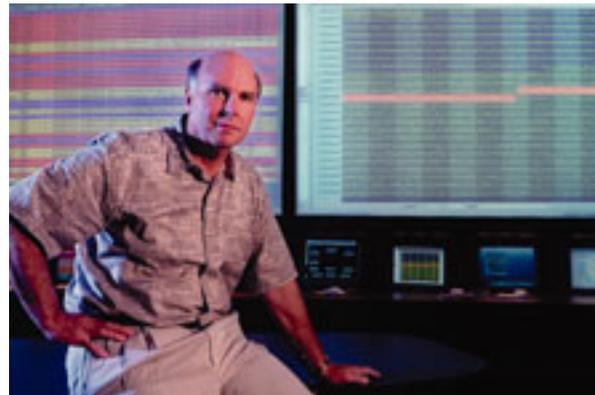
- 1991年 ヒトゲノム計画開始 (ワトソン、15年で完成を目指す)
- 1992年 ヒトゲノム計画のリーダーがコリンズに交代
- 1994年 ヒトゲノム遺伝子地図の完成
- 1996年 バミューダ会議 (第1回国際戦略会議)
- 1998年 Celera社がヒトゲノムを3年で解読すると発表 (ベンター)
- 1999年 22番染色体配列発表 (日・英・米)
- 2000年 21番染色体配列発表 (日・独)
- 2000年 ヒトドラフト (概要) 配列発表
- 2001年 ヒトドラフト配列論文発表



Francis Collins



Bermuda MTG



Craig Venter



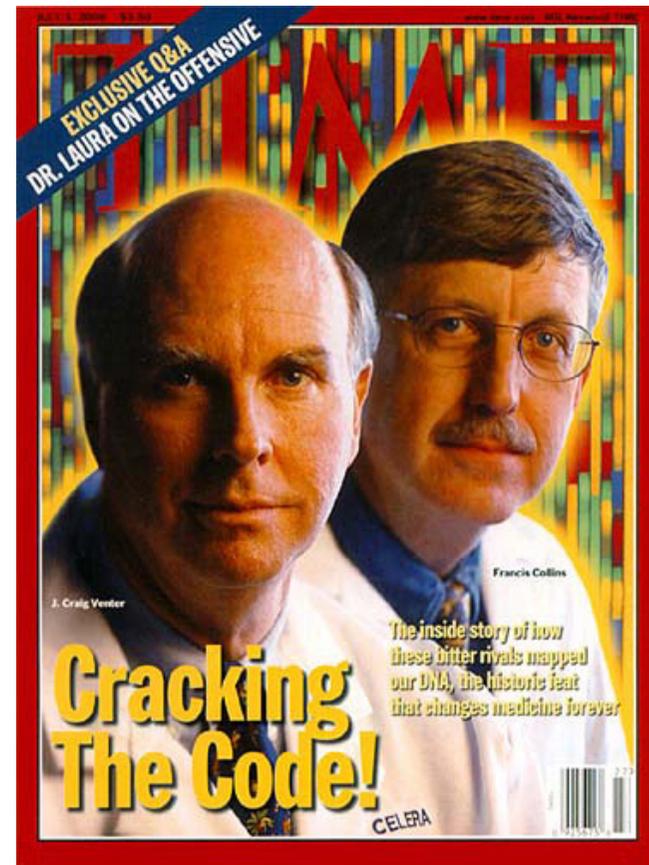
Chr.22 Press release

## Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium\*

\* A partial list of authors appears on the opposite page. Affiliations are listed at the end of the paper.

The human genome holds an extraordinary trove of information about human development, physiology, medicine and evolution. Here we report the results of an international collaboration to produce and make freely available a draft sequence of the human genome. We also present an initial analysis of the data, describing some of the insights that can be gleaned from the sequence.



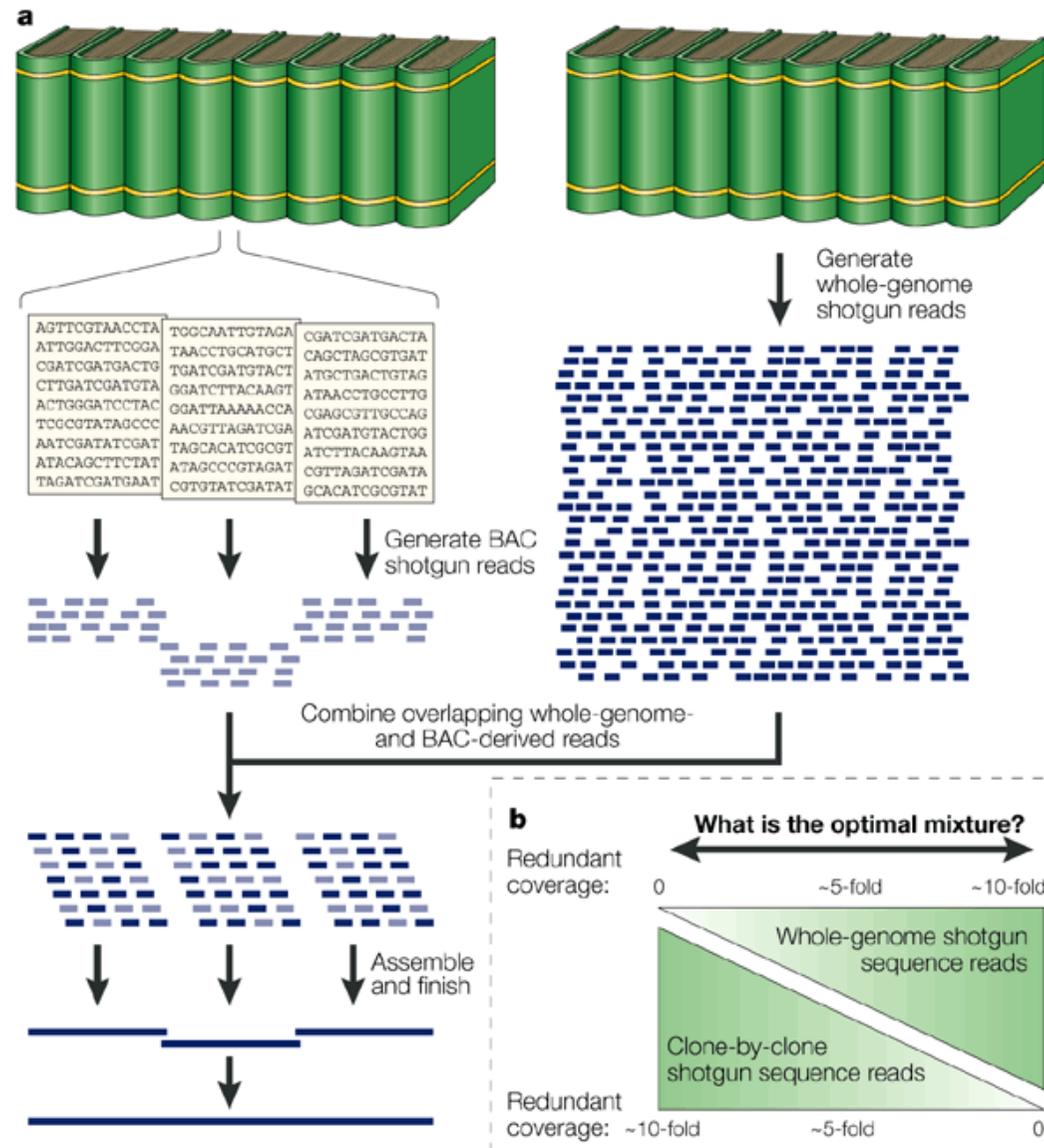
Time cover, July 3, 2000

和田 (藤山) 榎  
理研

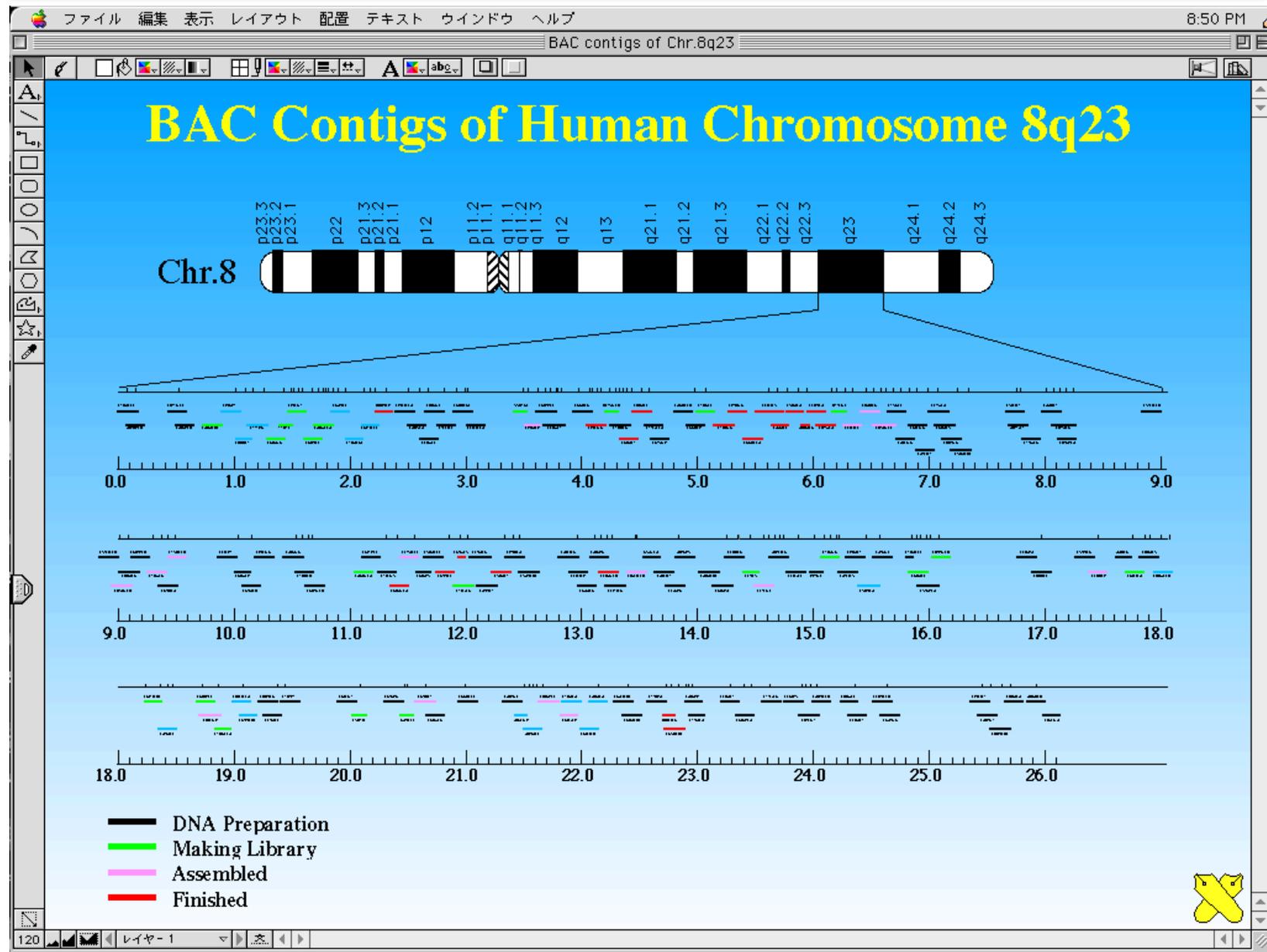
清水 蓑島  
慶應

菅原  
DDBJ

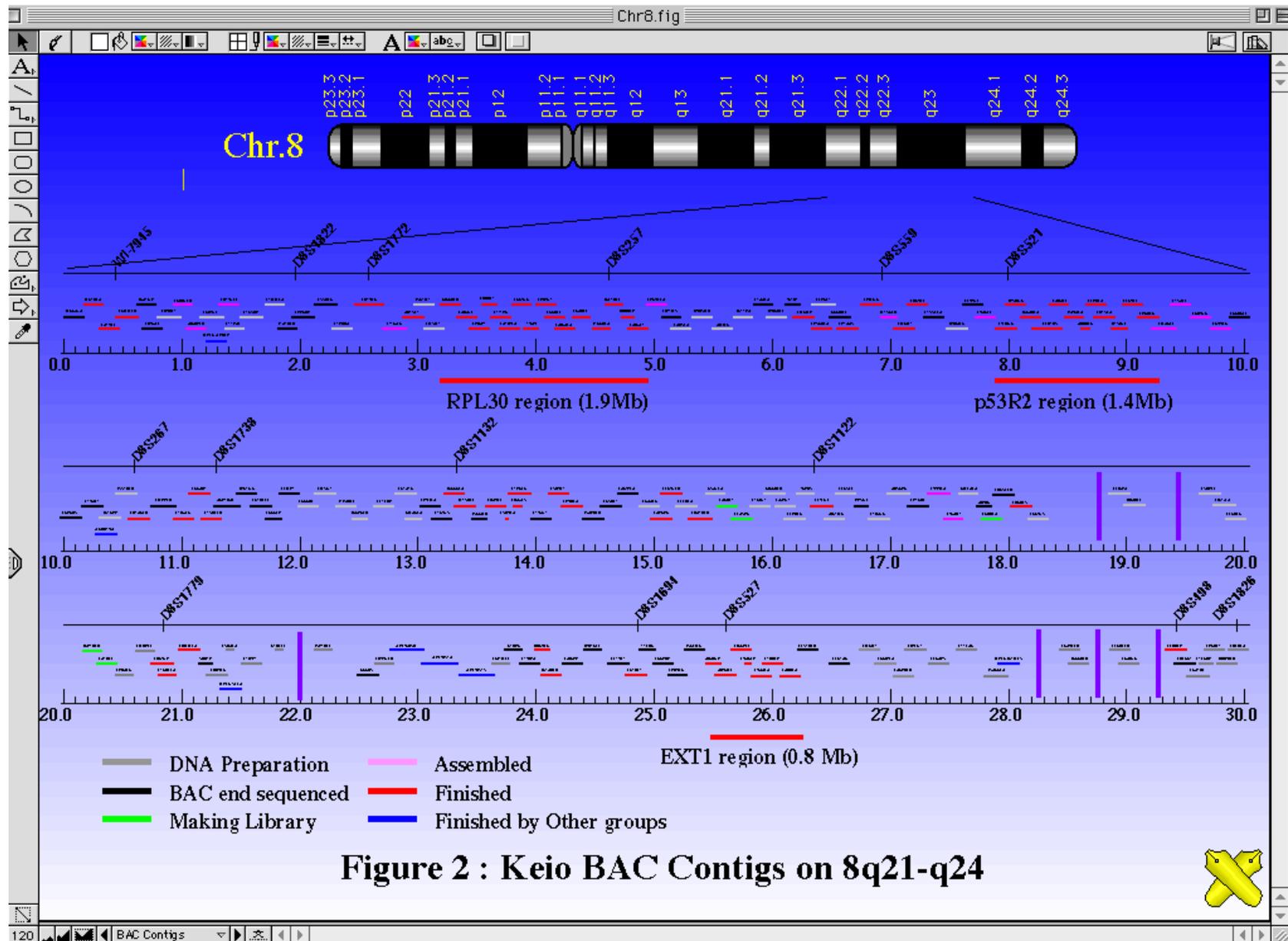
平成12年6月26日



# 2000年頃のヒト8番染色体の解読状況



# 2001年頃のヒト8番染色体の解読状況



2002年 横浜会議 (第12回最終国際戦略会議)

2003年4月 全ヒトゲノム配列発表

2004年10月 ヒトゲノム計画終了宣言

**articles**



Yokohama MTG

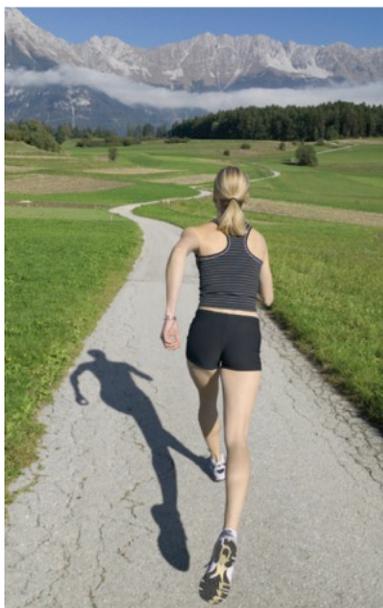
## Finishing the euchromatic sequence of the human genome

International Human Genome Sequencing Consortium\*

\* A list of authors and their affiliations appears in the Supplementary Information

The sequence of the human genome encodes the genetic instructions for human physiology, as well as rich information about human evolution. In 2001, the International Human Genome Sequencing Consortium reported a draft sequence of the euchromatic portion of the human genome. Since then, the international collaboration has worked to convert this draft into a genome sequence with high accuracy and nearly complete coverage. Here, we report the result of this finishing process. The current genome sequence (Build 35) contains 2.85 billion nucleotides interrupted by only 341 gaps. It covers ~99% of the euchromatic genome and is accurate to an error rate of ~1 event per 100,000 bases. Many of the remaining euchromatic gaps are associated with segmental duplications and will require focused work with new methods. The near-complete sequence, the first for a vertebrate, greatly improves the precision of biological analyses of the human genome including studies of gene number, birth and death. Notably, the human genome seems to encode only 20,000–25,000 protein-coding genes. The genome sequence reported here should serve as a firm foundation for biomedical research in the decades ahead.

*Nature* Oct 21; 431 931–45 (2004)



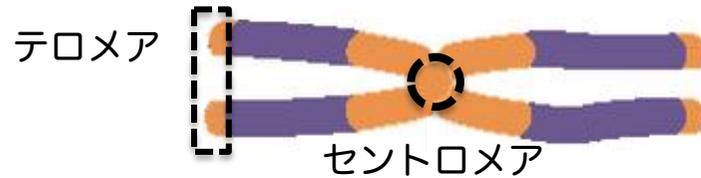
## End of the beginning

Lincoln D. Stein

Just over three years ago, it was announced that a first draft of the human genome sequence had been completed. Gaps and errors remained, but the job of fixing those problems is now largely done.



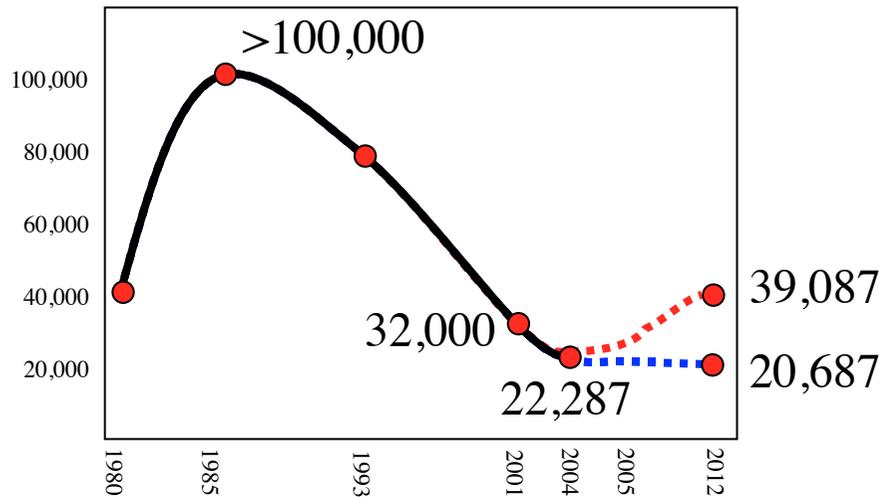
## 真正クロマチン領域の95%を解読



紫が真性クロマチン

オレンジがヘテロクロマチン

Gene Number



Sequence : 2.85 Gb

Gap : 341 (308 in euchromatin)

Protein-coding

Gene Number : 22,287 (20,000~25,000)

Error Rate: ~1 / 100,000 bases

Tiling Path: 26,720 clones

# ヒトゲノムの構成 (2000年頃)

