RESEARCH HIGHLIGHTS

HIGHLIGHT ADVISORS

SEAN B. CARROLL

UNIVERSITY OF WISCONSIN,
USA

NANCY J. COX

UNIVERSITY OF CHICAGO, USA

RALPH J. GREENSPAN

THE NEUROSCIENCES
INSTITUTE, CALIFORNIA, USA

YOSHIHIDE HAYASHIZAKI

RIKEN GENOMIC SCIENCES
CENTER, JAPAN

PETER KOOPMAN

UNIVERSITY OF QUEENSLAND,
AUSTRALIA

LEONID KRUGLYAK

FRED HUTCHINSON CANCER
RESEARCH CENTER, USA

STANLEY MALOY

SAN DIEGO STATE UNIVERSITY,
USA

BARBARA MEYER

UNIVERSITY OF CALIFORNIA,
BERKELEY, USA

JOHN QUAKENBUSH

THE INSTITUTE FOR GENOMIC
RESEARCH, USA

NADIA ROSENTHAL

EMBL MONTEROTONDO, ITALY

NORIYUKI SATOH

KYOTO UNIVERSITY, JAPAN

MARC VIDAL

DANA-FARBER CANCER
INSTITUTE, BOSTON, USA

VIRGINIA WALBOT

STANFORD UNIVERSITY, USA

DETLEF WEIGEL

MAX PLANCK INSTITUTE FOR
DEVELOPMENTAL BIOLOGY,
GERMANY

PHIL ZAMORE

UNIVERSITY OF
MASSACHUSETTS, USA

LEONARD I. ZON

CHILDREN'S HOSPITAL,
BOSTON, USA

HUMAN GENOMICS

Enjoy the complete coverage

2001 was a landmark year in genomics — in February, the draft sequence of the human genome was published. Although the word 'draft' was dropped by many, the International Human Genome Sequencing Consortium (IHGSC) did not forget that there was work left to do. In a follow-up paper in *Nature*, they now report the results of the effort to finish what they had left undone in 2001.

The 2001 draft sequence omitted 10% of the euchromatin and was interrupted by ~150,000 gaps; moreover, many local segments remained unordered and the corresponding clones were not oriented. The finishing process involved producing finished maps — "...continuous and accurate paths of overlapping large-insert clones spanning the euchromatic region of each chromosome arm" — and

finished clones — "...continuous and accurate nucleotide sequence across each large-insert clone". The 'finished' product (known as Build 35) now contains 2.85 billion nucleotides interrupted by only 341 gaps. The euchromatic human genome is now ~99% complete, with an error rate of only 1 in 10⁵ bases.

Among the many interesting facts about the genome that the report considers are the gaps that remain in the current sequence. Thirty three of them fall within heterochromatin, which was not targeted by the human genome project owing to its repetitive nature. A further 35 gaps are at the boundaries between euchromatin and heterochromatin, but the remaining 273 gaps lie within euchromatin and are mostly associated with segmental duplications. The IHGSC intend to close all of these euchromatic gaps, but rather than persevere with the existing strategies, which have so far failed in these cases, they propose to develop 'focused

strategies', such as using DNA from a single haploid source and new kinds of large insert libraries.

As the quality of the sequence went up, the estimate of the total number of protein coding genes went down. The 2001 prediction was in the region of 30,000; now, the IHGSC is talking about 20,000–25,000; the figures are derived from a combination of high quality sequence and gene annotation, which has now been manually augmented, taking into consideration gene structure, ESTs and transcript evidence. As the members of the IHGSC point out, this number refers to protein-coding genes only and point out that almost nothing is known about the non-coding transcripts, despite the recent flood of reports about them.

So here we are, the euchromatic sequence of the human genome has essentially been completed. But what of its heterochromatic portion? Will that be sequenced? The IHGSC says that to overcome the cloning and sequencing problems, new technologies would have to be developed. But will knowing the sequence of heterochromatic regions actually be that helpful in understanding this portion of the genome?

Magdalena Skipper

References and links

ORIGINAL RESEARCH PAPER International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 21 October 2004 (doi: 10.1038/nature03001)

FURTHER READING Eichler, E. E. *et al.* An assessment of the sequence gaps: unfinished business in a finished human genome. *Nature Rev. Genet.* 5, 345–354 (2004)