

'Fools' rush in where genome giants fear to tread

Wrangling the last bits of sequence from the human genome could provide a few surprises—enough to drive research into some tricky territory beyond the original goals of the Human Genome Project. For people who work on duplicated and unstable regions of the genome, for instance, work has just begun.

"A few fools go into these regions, and we are among them," says Evan Eichler, assistant professor of genetics at Case Western Reserve University. Eichler's group painstakingly catalogs duplicated sequences in the genome and places them in the correct positions. Because duplicated regions are particularly susceptible to breaks and rearrangements during recombination, they underlie many human genetic diseases and may even drive evolution, Eichler says. They are also the source of many of the missed and misplaced genes in both the Celera and publicly-funded sequences.

The Human Genome Project aimed to sequence at least 95% of each chromosome—excluding centromeric heterochro-

matin—with all gaps sized, oriented and annotated. The sequenced regions were to be 99.99% accurate at the nucleotide level. On 14 April, leaders of the project are expected to announce that they have achieved that aim.

The stated goals have most likely been met, says Huntington Willard, past president of the American Society of Human Genetics. But "it would be a shame if they were to say they are done, full stop," he says. Willard, who plans to sequence heterochromatic regions, says it is critical that researchers continue trying to accurately place duplicated regions on the genome.

"This is heavy lifting but the payout will be fascinating and likely important," says Willard. "It's worth it from any number of perspectives—philosophically, evolutionarily and medically."

Duplicated regions, which can exceed 400 kb, account for half of the known gaps in the sequence, says Eichler, and either get pinned to the wrong place or dropped off the map entirely. That's particularly problematic when a duplicated region is larger

than a bacterial artificial chromosome, the sequencing unit for genome projects.

It is difficult to determine precisely how many duplicated regions are missing, says Julie Korenberg, professor of genetics at the University of California in Los Angeles. At the American Society for Human Genetics annual meeting last fall, Korenberg estimated that 47% of 580 bacterial artificial chromosomes with known duplications contained sequences that were either missing or misassembled in the publicly funded genome sequence. That number has dropped to below 20% so the data are improving, she now says. Similar problems afflict the Celera sequence, she adds.

Duplicated regions are enriched for genes with roles in immunity and development. They may also underlie some of the variability in the human population's susceptibility to disease. "If you are the sort who says you look for treasure under the light," says Willard, "then this is where you should be shining the light."

Charlotte Schubert, New York