to find genes that underlie QTLs[7]. Malek *et al.*[1] have determined transcriptional profiles from 18 rat strains, documenting changes in transcript levels for both genders, four tissues (heart, lung, liver and kidney) and two environmental conditions. Environmental stressors like hypoxia, exercise or dietary salt intake are used to unmask deficiencies in compensatory mechanisms maintaining cardiovascular homeostasis that otherwise would not be detected.

Application to myocardial infarction

The authors provide compelling examples of how the expression database can be used in combination with the wealth of physiological data to formulate hypotheses about the genes contributing to cardiac infarct size and to pulmonary vasculature remodeling. First, the examples presented highlight that primary candidate genes can be identified. The high heritability of variation in mammalian gene expression[8,9] has suggested that identification of the genetic determinants of gene expression may give insights into the molecular basis of complex traits. Transcript abundance may act as an intermediate phenotype between genomic DNA sequence and more complex whole-body phenotypes. Second, downstream effects can be identified using the TREX database. Variation in transcriptional regulation may also point to modifier genes that have a crucial role in the disease process. However, one note of caution is required. Correlation of phenotype and expression data does not necessarily imply causality.

It is not obvious whether the CSS offer advantages over other methods for the identification of genes. As a QTL is a locus that contributes to a phenotype that is measured quantitatively, the main drawback of CSS is that it does not allow the fractionation of a large QTL into many loci with smaller effects[10]. An alternative strategy that integrates expression data and QTL mapping data, as recently demonstrated in rat and mouse recombinant inbred strains, may prove advantageous[11–13] and could be applied to CSS strains. This approach enables the mapping of expression QTLs that are primary control points for gene expression across the genome[14,15]. Together with emerging large-scale SNP resources, this may allow a more rapid identification of refined regions of interest for a given QTL and eventual identification of the underlying gene and mechanism.

At this stage, already one can expect that the current resource generated by the authors will significantly enhance our capabilities to perform in depth correlations between transcriptional profiles and a range of cardiovascular physiological phenotypes. But there is even more to expect. Both CSS panels consisting of 22 strains each have been completely established, and phenotype characterization for more than 213 cardiovascular traits is expected to be completed mid-2006 (http://pga.mcw.edu/pga). We can only hope that transcriptional profiling of these additional strains keeps the pace. Such studies should provide the tools required for a more integrative view of biological systems to understand and treat disease better.

1. Malek, R.E. *et al. Nat. Genet.* **38**, 234–239 (2006).
2. Lindsey, J.R. in *The Laboratory Rat* (eds. Baker, H.J., Lindsey, J.R. & Weisbroth, S.H.) 1–36 (Academic, New York, 1979).
3. Gibbs, R.A. *et al. Nature* **428**, 493–521 (2004).
4. Cowley, A.W. *et al. J. Physiol. (Lond.)* **554**, 46–55 (2004).
5. Singer, J.B. *et al. Science* **304**, 445–448 (2004).
6. Darvasi, A. *Nat. Genet.* **18**, 19–24 (1998).
7. Aitman, T.J. *et al. Nat. Genet.* **21**, 76–83 (1999).
8. Cheung, V.G. *et al. Nat. Genet.* **33**, 422–425 (2003).
9. Schadt, E.E. *et al. Nature* **422**, 297–302 (2003).
10. Flint, J. *et al. Nat. Rev. Genet.* **6**, 271–286 (2005).
11. Hubner, N. *et al. Nat. Genet.* **37**, 243–253 (2005).
12. Chesler, E.J. *et al. Nat. Genet.* **37**, 233–242 (2005).
13. Bystrykh, L. *et al. Nat. Genet.* **37**, 225–232 (2005).
14. Damerval, C. *et al. Genetics* **137**, 289–301 (1994).
15. Jansen, R.C. & Nap, J.P. *Trends Genet.* **17**, 388–391 (2001).

The rise and fall of the ape Y chromosome?

Chris Tyler-Smith, Kevin Howe & Fabrício R Santos

The sequence of a second chimpanzee Y chromosome has been determined. It confirms the degradation of four genes on the chimpanzee lineage, reveals the recent gain of one on the human lineage and emphasizes the low Y-chromosomal genetic diversity within western chimpanzees.

When you want a bus, you wait for ages, and then two come along at the same time. Now it seems to be the same with chimpanzee Y chromosomes: after waiting years since the first small-scale studies[1], two independent finished sequences have appeared within the space of a few months: one from David Page's group[2] and the other from Asao Fujiyama's team[3] on p. 158 of this issue of *Nature Genetics*. Why is there so much interest in this chromosome?

Chris Tyler-Smith and Kevin Howe are at the Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK, and Fabricio R. Santos is in the Departamento de Biologia Geral, ICB, Universidade Federal de Minas Gerais, Av. Antonio Carlos 6627, Caixa Postal 486 31.270-910, Belo Horizonte, MG, Brasil. e-mail: cts@sanger.ac.uk

Why sequence the chimpanzee Y—twice?

The chimpanzee (*Pan troglodytes*), along with the bonobo or 'pygmy chimpanzee' (*Pan paniscus*), is our closest living relative—we have been called 'the third chimpanzee'[4]—so the differences between our genomes must underlie the differences between our phenotypes, and help us to understand what makes us human. The Y chromosome determines male sex. Unlike the other chromosomes, it is therefore haploid and male-specific, consequently escaping meiotic recombination over most of its length, and subject to more rapid evolution because of its permanent location in the error-prone male germ line[5]. The chimpanzee Y sequence should therefore increase our understanding of chromosomal and male-lineage biology and has been eagerly awaited. But the chimpanzee genome has already been sequenced, hasn't it? Yes and no. A sequence has been published[6], but that was a draft with 3.6× coverage, full of gaps and particularly incomplete on the X and Y chromosomes, which received only 1.8× coverage each. In addition, the Y chromosome is enriched in repeated sequences, which are difficult to assemble correctly. So there are good reasons for making a special

effort to produce a high-quality finished sequence of the Y separately from the rest of the genome.

One chimpanzee Y sequence should tell us about the fixed differences from humans, so why sequence two, when we have just sequenced one human Y? The real reasons may have more to do with the politics of science than a coordinated program in ape genomics, but we can take advantage of the availability of two such sequences to learn about both the reliability of chromosome-sized sequences and the diversity within chimpanzees: not all differences between the human and chimpanzee reference sequences are fixed in each species.

Chimpanzees are ascribed to four subspecies with different geographical locations in Africa[7]. The genomic sequence[6] and most of one finished Y chromosome[2] were derived from a single captive-born male western chimpanzee (*P. t. verus*) called Clint, whereas the other Y was from Gon, whose origin was unclear but could also be attributed to *P. t. verus* using Y-SNPs[3]. The finished sequences cover only about half of the chromosome, omitting the more difficult palindromic regions, but the current paper extends the previous sequence from ~9.5 Mb to ~12.7 Mb. Both sequences claim high accuracy and completeness, with error rates of 1 per 204 kb (ref. 2) or 1 per ~500 kb (ref. 3), and only small gaps remaining.

Human-chimpanzee differences

The two sequences should be very similar to one another, and it is gratifying to find that they are (**Fig. 1**). Both studies illustrate the rapid evolution of the Y chromosome at all scales. For example, the nucleotide substitution rate was 1.78% compared with the genome-wide average of 1.23%; there are many retroelement differences, including several chimpanzee-specific endogenous retrovirus (CERV) insertions; and there have been large-scale structural rearrangements. These include relocation of the centromere (perhaps through a ~5-Mb insertion into the human Y chromosome followed by loss of the original centromere, or neocentromere formation) and inversions, in which a ~5-Mb inversion on the chimpanzee lineage and a ~1.5-Mb inversion on the human lineage could account for the gross features of the observed structures.

Particular attention has been paid to the gene content of the Y chromosome, and here the conclusions appear to differ somewhat. All genes annotated as pseudogenes in humans are also pseudogenes in chimpanzees, but one gene present in humans, *CD24L4*, represents a recent human-specific gain. *VCY* and *VCY1B* have a complex history. This protein family ends with a repeating motif of ten amino acids.

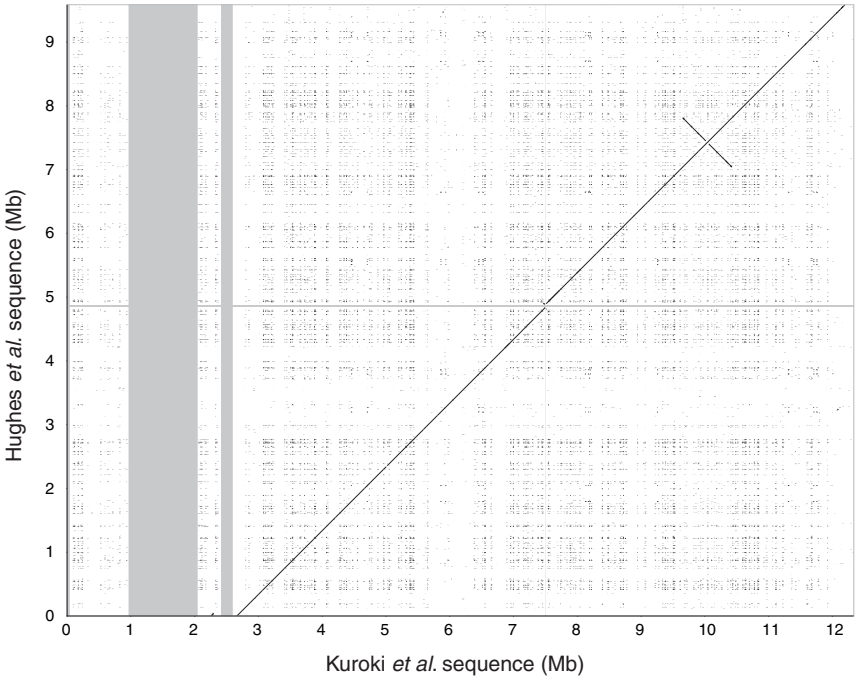

Figure 1 Comparison of the Kuroki *et al.* and Hughes *et al.* chimpanzee Y chromosome assemblies. BLAST hits of >90% between the two sequences are plotted. Shaded areas denote sequence gaps.

The human *VCY*s have one copy of this motif; chimpanzee *VCY*s have eight but are truncated by a frameshift after one and thus end up the same length as the human proteins. Several genes are significantly shorter in the chimpanzee, and are discussed below.

Variation within chimpanzees

Comparison of the ~9.5 Mb aligned sequences shows many small differences: 747 SNPs and 407 indels (**Supplementary Table 1** online), or about one event per 8 kb. This is far higher than the estimated sequence error rates, so most must represent polymorphisms within chimpanzees. The SNP spectrum mirrors the chimpanzee-human SNP differences. Half of the indels involve a single nucleotide, and the medium-sized ones (2–22 bp) are dominated by microsatellite unit differences. Consequently, they show a strong bias toward changes involving even numbers of nucleotides because of the predominance of di- and tetranucleotide microsatellites. The largest indels lie mostly in simple-sequence DNA. The nucleotide diversity between Clint and Gon is low (0.008%, excluding indels) but is consistent with the previous report of zero SNPs in 2,787 bp from 77 *P. t. verus* Y chromosomes[1]. Human Y chromosomes show similar (*e.g.*, 0.010%)[8] or slightly higher (*e.g.*, 0.034%) diversity[9], whereas chimpanzee Y chromosomes as a whole show considerably higher diversity (0.067%)[1].

Kuroki *et al.*[3] identify three genes (*CYorf15B*, *USP9Y*, *TBL1Y*) that are substantially shorter in chimpanzees than humans, whereas Hughes *et al.*[2] identified these and two more: *TMSB4Y* and *CYorf15A*. So, are chimpanzee Y chromosomes polymorphic for their gene content? Closer examination suggests not. *TMSB4Y* carries the same splice donor site mutation in Gon as in Clint, so this apparent difference represents a variation in annotation. The apparent truncation of *CYorf15A* in Clint is also carried by Gon, but it represents a frameshift polymorphism or sequence error in humans rather than a chimpanzee-human fixed difference (**Supplementary Note** online). Thus, chimpanzees have experienced truncation of at least four genes, but none of these are polymorphic between the two Y chromosomes examined. As some healthy men lack *PRKY*, *TBL1Y* and *AMELY*[10], humans are polymorphic for loss or frameshift in three or four genes. These observations call into question the suggested contrast between the gene decay on the chimpanzee Y and conservation on the human Y chromosome[2] and leave open the question of whether there has been a recent selective sweep on the chimpanzee Y.

General lessons

The chimpanzee Y sequences demonstrate both the high reproducibility of current sequencing projects and the difficulties of

reliable annotation. But although sequencing multiple copies of a mammalian chromosome is now a reality, it still requires a substantial investment of money and effort, so thought and coordination in choosing the source material are needed: how different would our view be if the second chimpanzee Y chromosome had been from another subspecies? How representative of all human Y chromosomes is the one sequence available? We clearly need more human Y chromosome sequences, and they should be chosen from diverse haplogroups.

On a more sober note, chimpanzees are now highly endangered[11]. These genomic studies illustrate the scientific importance of the great apes. The organizers of the chimpanzee sequencing projects are to be congratulated for avoiding one failure of the human genome project and ensuring that cell lines from both Clint and Gon were established, and it is to be hoped that Convention on International Trade in Endangered Species (CITES) regulations will not unnecessarily hinder their distribution. We now look forward to genomic analyses of the bonobo, gorilla and orangutan, including, we hope, all species and subspecies.

Note: Supplementary information is available on the Nature Genetics website.

1. Stone, A.C., Griffiths, R.C., Zegura, S.L. & Hammer, M.F. *Proc. Natl. Acad. Sci. USA* **99**, 43–48 (2002).
2. Hughes, J.F. *et al. Nature* **437**, 100–103 (2005).
3. Kuroki, Y. *et al. Nat. Genet.* **38**, 158–167 (2006).
4. Diamond, J. *The Rise and Fall of the Third Chimpanzee* (Vintage, London, 1992).
5. Jobling, M.A. & Tyler-Smith, C. *Nat. Rev. Genet.* **4**, 598–612 (2003).
6. The Chimpanzee Sequencing and Analysis Consortium. *Nature* **437**, 69–87 (2005).
7. Jobling, M.A., Hurles, M.E. & Tyler-Smith, C. *Human Evolutionary Genetics* (Garland Science, New York and Abingdon, 2004).
8. Shen, P. *et al. Proc. Natl. Acad. Sci. USA* **97**, 7354–7359 (2000).
9. Hammer, M.F. *Nature* **378**, 376–378 (1995).
10. Santos, F.R., Pandya, A. & Tyler-Smith, C. *Nat. Genet.* **18**, 103 (1998).
11. Walsh, P.D. *et al. Nature* **422**, 611–614 (2003).