As normal as normal can be?

Nigel P Carter

Two papers report that large-scale copy-number variations, ranging in size from 100 kb to 2 Mb, are distributed widely throughout the human genome, and that a high proportion of them encompass known genes. This unexpected level of genome variation has implications for our view of human genetic diversity and phenotypic variation.

There has been much focus recently on normal genome variation, particularly in the form of single-nucleotide polymorphisms[1], and the extent of segmental duplication in the human genome is now well-documented[2]. But the importance of normal copy-number variation involving large segments of DNA has been largely unappreciated, as only a handful of instances have been reported[3–8]. Now, using DNA microarrays (array comparative genomic hybridization) to screen the human genome for changes in copy number, two studies[9,10] report a substantial degree of large-scale copy-number variation (LCV) in the human population. John Iafrate and colleagues[9], reporting on page 949 of this issue, describe their use of large insert clone arrays for this purpose, and Sebat *et al.*, reporting in *Science*[10], describe similar findings using oligonucleotide arrays. The conclusion is that >200 large segments of the genome vary severalfold in copy number in the human population. This unexpected level of LCV forces us to re-evaluate our view of the structure of the normal human genome.

How many, and how common?
Iafrate *et al.* report 255 variable loci in 55 individuals, whereas Sebat *et al.* report 76 LCVs in 20 individuals, with average differences between any two individuals of 12.4 and 11 LCVs, respectively. The two sets of loci can be viewed at the Genome Variation Database (http://projects.tcag.ca/variation). The two studies report only 11 loci in common (within 1 Mb). It is not clear whether the loci described in each study were mapped onto the same

Nigel P. Carter is in The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK.
e-mail: npc@sanger.ac.uk

build of the human genome, but it seems at first glance that each study underestimates the number of LCVs in the human population.

How can we explain this discrepancy? One factor might be the resolution of the arrays used. Neither array achieves both high resolution and complete coverage of the genome (~150 kb every 1 Mb with a detection size of ~50 kb (ref. 9) versus ~1 probe every 35 kb with a detection size of 105 kb (ref. 10)). Moreover, it is common practice in selecting clones or probes for array comparative genomic hybridization to avoid regions that hybridize to

more than one genomic location or show variation in normal hybridizations. Such arrays are probably biased away from LCVs. Further studies with genome-wide clone tiling path arrays or higher density oligonucleotide arrays will be required to settle this issue, but many more LCVs probably remain to be discovered.

How common are LCVs? In both studies, approximately one-half of the variable loci were polymorphic in only one individual, whereas one-half showed copy-number variation in more than one individual. Only a few of the most common LCVs were detected in both
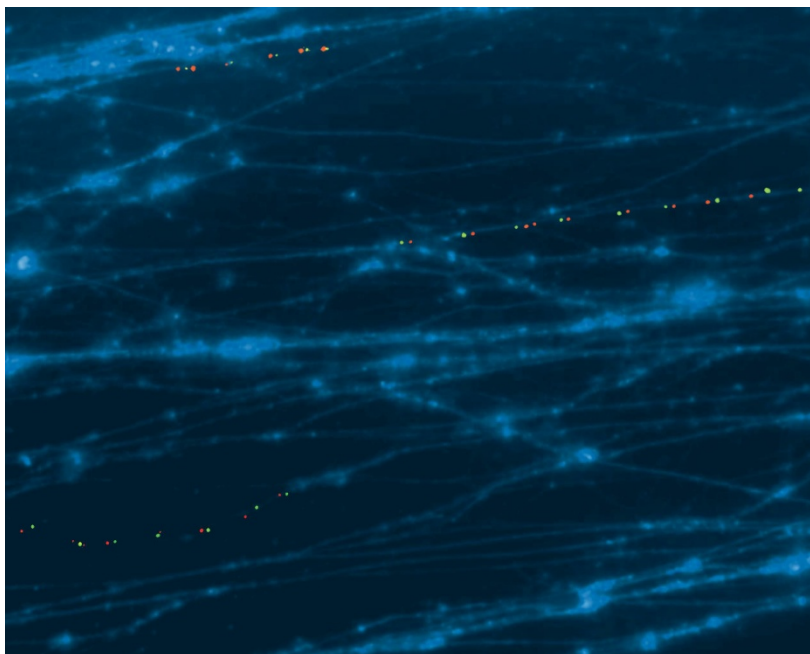

Figure 1 Fluoresence *in situ* hybridization on stretched DNA fibers shows copy-number variation at the *AMY1A–AMY2A* locus on 1p21.1. The image shows hybridization of a 5′ amylase gene probe (red) and a 3′ amylase gene probe (green) to DNA fibers (blue) from three different individuals, each with a different number of tandem copies of the variable segment.

Photo courtesy of Charles Lee

studies (*e.g.*, 7q35, 14q32.33). For example, the most common variant in the Iafrate *et al.* study, located at 1p21.1 (*AMY1A–AMY2A*) and present in >49% of the individuals studied (**Fig. 1**), was not detected by Sebat *et al.* Again, does this discrepancy reflect a technical difference between the two studies or the very different ethnic mix of the two study groups? We will need to study many more individuals from a range of ancestral backgrounds to arrive at an accurate picture of the frequency of the more common LCVs.

LCVs, duplications and disease
Analysis of the reference sequence shows that ~5% of the human genome is duplicated[2]. These segmental duplications, defined as multiple regions sharing at least 1 kb of 90% identical sequence, are thought to have had a key role in human genome evolution[11] and may be responsible, through nonallelic homologous recombination (NAHR), for many chromosome rearrangements leading to disease[12]. Both Iafrate *et al.* and Sebat *et al.* report a higher than expected association of LCVs with known segmental duplications and with regions associated with human genetic disease or cancer. This suggests that LCVs and other genomic rearrangements might have a common mechanistic basis.

It also has been suggested that large segmental duplications could complicate sequence assembly and lead to gaps in the sequence[13].

Iafrate *et al.* note that 12.7% of LCVs are located in the 100 kb of gaps in the current sequence assembly. Similarly, LCVs with high sequence homology might be assembled out of the reference sequence. Furthermore, the reference sequence was produced from clone libraries generated from a small number of individuals, and so most LCVs would not be represented in the libraries. This raises the question, "What is the sequence of the normal human genome?" Much more detailed sequence analysis of LCVs in a large number of individuals will be needed to address this issue.

In both studies, a high proportion of LCVs overlapped with known genes. Further studies of these genes in individuals with different copy numbers will be interesting, as copy-number differences will probably be found to influence gene expression[14]. Alternatively, regulatory mechanisms could compensate for differences in copy number between individuals. While LCVs have been identified in phenotypically normal individuals, we cannot determine the phenotypic consequences of such large polymorphisms. Some of these LCVs may be associated with age-related susceptibilities to disease, and deletion polymorphisms may reveal recessive mutations with phenotypic consequences. A recent study of 50 individuals with learning disability and dysmorphology identified five LCVs that were inherited from normal parents and so did not segregate with the disease phenotype[15]. These

observations underscore the importance of identifying LCVs in the normal population so that we can gauge the importance of copy-number changes in individuals with diseases.

Both Iafrate *et al.* and Sebat *et al.* report that LCVs are frequently located in regions of the genome that are susceptible to rearrangement, particularly by NAHR. The copy-number variation found in LCVs could certainly be generated by NAHR, suggesting that there may be a common mechanism for disease-associated and normal copy-number variation. If this turns out to be true, the LCVs themselves may point to unstable regions of the genome at which new disease-associated rearrangements may be found in the future.

1. Eichler, E.E. & Frazer, K.A. *Genome Biol.* **5**, 318 (2004).
2. Bailey, J.A. *et al. Science* **297**, 1003–1007 (2002).
3. Campbell, R.D. *et al. Exp. Clin. Immunogenet.* **7**, 69–84 (1990).
4. Riley, B. *et al. Genomics* **79**, 197–209 (2002).
5. Blunt, T. *et al. Ann. Hum. Genet.* **58**, 19–24 (1994).
6. Gilles, F. *et al. Genomics* **70**, 364–374 (2000).
7. Hollox, E.J. *et al. Am. J. Hum. Genet.* **73**, 591–600 (2003).
8. Kulski, J.K. *et al. Immunol. Rev.* **190**, 95–122 (2002).
9. Iafrate, A.J. *et al. Nat. Genet.* **36**, 949–951 (2004).
10. Sebat, J. *et al. Science* **305**, 525–528 (2004).
11. Bailey, J.A. *et al. Genome Biol.* **5**, R23 (2004).
12. Shaw, C.J. & Lupski, J.R. *Hum. Mol. Genet.* **13** Spec No 1, R57–R64 (2004).
13. Eichler, E.E. *et al. Nat. Rev. Genet.* **5**, 345–354 (2004).
14. Townson, J.R. *et al. Eur. J. Immunol.* **32**, 3016–3026 (2002).
15. Shaw-Smith, C. *et al. J. Med. Genet.* **41**, 241–248 (2004).

Cancer chromosomes in crisis

Ronald A. DePinho & Kornelia Polyak

The benign-to-malignant transition in human breast cancer is associated with a marked increase in chromosomal aberrations. A new study suggests that telomere dysfunction and its associated bridge-fusion-breakage cycles may drive this episodic instability, thereby providing aspiring cancer cells with the multiple genetic aberrations needed to achieve a fully malignant state.

Ronald A. DePinho is at the Center for Applied Cancer Science and in the Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA, and the Departments of Medicine and Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. Kornelia Polyak is in the Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA, and the Departments of Medicine and Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.
e-mail: ron_depinho@dfci.harvard.edu

Chromosomal aberrations such as amplifications, deletions and complex translocations are pervasive in human epithelial cancers. These are the main cancers affecting the aged, which has motivated efforts to elucidate the mechanisms leading to chromosomal instability[1]. Several mechanisms of genetic instability have been proposed, including mutations in mitotic checkpoint genes that control chromosome segregation, and loss of telomere capping function resulting in dysfunctional telomeres[2]. The latter model is supported by mouse knock-

out studies that have established a link between telomere dysfunction, increased epithelial cancers and radically altered cytogenetic profiles typical of those found in human epithelial cancers[3]. Studies of human primary tumors and epithelial cultures have also supported the idea that telomere dysfunction and its associated bridge-fusion-breakage (BFB) cycles are important in shaping the cancer genome[4–6]. But it is not yet known at which stage of tumorigenesis telomere-induced chromosomal instability unfolds.