Cell PRESS

Copy-number variation: the end of the human genome?

Paul H. Dear

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK

Copy-number variation (CNV) – the presence of additional or missing segments of chromosomes in some individuals – has been found to be abundant in humans and adds another dimension of variation to the genome. Copy-number variants have already been associated with some diseases and disease susceptibilities and are likely to prove as significant as sequence polymorphisms in this respect. Changes in copy number of parts of the genome are known to be a feature of many cancers, and their analysis is expected to reveal genes involved in carcinogenesis. This article will present a somewhat biased and occasionally speculative discussion of the current and future significance of CNV with a particular focus on the potential of molecular copy-number counting in the analysis of small, damaged or heterogeneous samples.

Introduction

The human genome ceased to exist in ~2004. As far back as the 1990s, microsatellites were known as stutters in the code that varied in length from person to person, but these were mostly considered convenient and inconsequential genetic tags, useful for forensic scientists and for tracing pedigrees. Point mutations were also known to be scattered around the genome, sometimes with dramatic consequences such as sickle-cell anaemia or cystic fibrosis, but more often benign. Even after the discovery of thousands and then millions of single-nucleotide polymorphisms (SNPs) at the very end of the 20th century (The International HapMap Project, http://snp.cshl.org) [1,2], it was still possible to speak of *the* human genome in the sense of a reference sequence peppered with allelic variants.

However, no sooner than the community had embraced SNPs, another type of variation started to come to the fore. It was, of course, known that duplications or deletions of parts of the genome could occur in rare instances. For example, cancers had long been known to be associated with aneuploidy or structural rearrangements giving rise to extra or missing copies of parts of the genome. Such 'structural variants' had also been found in non-cancer tissues, in both normal and diseased individuals [3–6].

But the increasing use of array-based comparative genomic hybridization (CGH, in which two genomes, labelled with different fluorophores, are hybridized competitively to an array of cloned DNA fragments, synthetic oligonucleotides or other targets) soon led to the realization that the presence of additional or missing copies of some segments of DNA was indeed a widespread feature of the genomes of all healthy humans [7,8]. The quest to catalogue SNPs also helped to drive the discovery of copy-number variations (CNVs), as the dense oligonucleotide arrays that had been developed to identify SNPs proved well-suited to detecting copy-number changes through comparative hybridization. Large-scale sequencing using newly developed technologies also proved effective in revealing CNVs, with some sequences being over- or under-represented in the sequence data according to their copy number in the individual genome being sequenced. Figure 1 outlines some of the most common methods for detecting or quantifying CNVs.

With CNVs added to SNPs, the human genome has ceased to exist as a defined entity and is instead becoming a fluid continuum. Certainly, the 5% of the genome already known to display CNV [9] is considerably larger than the 0.1–1% of sequence difference that was once believed to distinguish one human from another [10,11]. How does this newly appreciated variability affect our view of the genome and the search for the genomic basis of human variation and disease?

A brief history of CNVs

The first CNV in humans – discovered in the early 1900s – was found to be widespread throughout all populations, with a range of striking but largely benign phenotypic consequences. A long period of evolutionary adaptation enables men to survive quite well with only a single X chromosome and a vestigial Y in place of the missing copy. This is the exception, however, as the first human autosomal aneuploidy to be identified, trisomy 21 [12], has a range of generally adverse phenotypic consequences, as do all other human autosomal aneuploidies, most of which are non-viable.

Partial chromosomal duplications were also found to be associated with some inherited disorders, and some other duplications were shown to be present as apparently benign variants in physically normal individuals (reviewed in [13]). However, the first large, systematic structural comparisons of the genomes of healthy humans were only carried out in 2004 and used hybridization to identify large numbers of CNVs that are present at significant frequency [7,8]. Further studies [14–16] have revealed well over a thousand CNVs, which at one time were believed to account for up to 15% of the human genome, a figure inflated somewhat by the limited spatial resolution of the methods used and consequent overestimation of the sizes of the genomic regions involved. A recent large-scale study [9] puts the figure closer to 5%.

Corresponding author: Dear, P.H. (phd@mrc-lmb.cam.ac.uk)

0167-7799/$ – see front matter doi:10.1016/j.tibtech.2009.05.003 Available online 1 July 2009

CNVs and phenotype

Almost by definition, most CNVs have either no phenotypic consequences or only subtle or benign ones. If they had significant effects, the deleterious alleles would have been largely selected out of the population or would be defined as mutations and found in a small number of heterozygotes and in an even smaller number of affected individuals. In fact, given that CNVs seem to have arisen and persisted in multiple independent events in some parts of the genome but are absent from others, it seems likely that selective pressure weeds out most CNVs other than those that are benign or subtle in their effects or that confer a mixture of beneficial and harmful effects depending on their genomic and environmental context (like the sickle-cell mutation in haemoglobin). In this respect, they are comparable to the majority of SNPs.

Nevertheless, it was hoped and expected (and in some cases has been shown) that some CNVs do indeed have phenotypic consequences (for a recent review, see [17]). In some cases, these associations have been discovered by looking for CNVs in genes known to be implicated in disease, such as the α-synuclein gene involved in Parkinson's disease [18] or the *APP* gene, which produces the amyloid precursor protein implicated in familial Alzheimer's disease [19–21]. In other cases, potential phenotypic effects of CNVs have been discovered by association studies; for instance, the *chemokine ligand 3-like 1* (*CCL3L1*) CNV was found to be associated with HIV-1 susceptibility, rheumatoid arthritis and type 1 diabetes [22,23]. The race is now on to find further associations, which might prove to be more frequent and more significant than associations involving SNPs alone.

Copy-number changes and cancer

CNVs have clearly been shown to have the potential to indirectly influence a healthy individual's susceptibility to cancer, for example by varying the gene dosage of tumour suppressors or oncogenes (see, for example, [24]).

A more direct and immediate role of copy-number change is seen in cancerous cells themselves, which frequently display CNVs that are absent in the patient's normal cells in characteristic (although cancer-type-specific) parts of the genome. Arguably, such copy-number changes in cancer cells should not be called 'variants' because they do not fall within the spectrum of normal human variation. Aneuploidy, double-minutes and non-reciprocal translocations have long been recognized in many cancers and are one form of CNV. But cancers have also been shown to gain additional copies of certain smaller genomic regions (for example, [25,26]). Such gains are of particular interest when they are found in many patients with a given cancer type or in the early stages of cancer because they can then be inferred to harbour so-called 'driver genes' that favour the growth of abnormal cells. By contrast, in the later stages of many cancers, 'genomic chaos' often ensues, which makes it difficult to determine which copy-number changes are significant or causative and which are purely incidental or of minor importance.

My own interest in copy-number changes is largely in this context and involves the investigation of genomic gains and losses that arise in early cancerous or pre-cancerous cells in lung and colon. However, studying copy-number changes in early disease is particularly challenging for three reasons. First, biopsies from such early lesions are exceedingly scarce as patients seldom present for examination until the disease is at an advanced stage. Second, biopsies from early lesions are likely to be extremely small and to contain only islands of abnormal cells amongst a sea of normal tissue. Finally, the vast majority of potentially informative biopsies are buried in clinical archives that have accumulated over many years, having been formalin-fixed, wax-embedded and stored in a cupboard; one could hardly imagine a better way of destroying DNA.

MCC and μMCC – a digital genomics approach to CNVs

Conventional methods for analysing genomes, which are based on microarrays, quantitative PCR (qPCR) or sequencing, have great difficulty in measuring CNVs in small or badly degraded specimens, such as fixed biopsies, because the amount of intact DNA available is simply insufficient. Although such limited DNA amounts can be expanded using various whole-genome amplification methods [27,28], this runs the risk of introducing biases and artefacts and, moreover, fails to work entirely for many samples [27,29].

To overcome these limitations, we have recently developed the method of molecular copy-number counting (MCC) [30], which is almost embarrassingly simple in its conception. In brief, the sample DNA is dispensed at limiting dilution into several aliquots, and a multistep, multiplexed single-molecule PCR is used to directly count the number of molecules representing chosen sequences that are scattered amongst the aliquots. If one or more of the sequences are 'references' believed to be present at normal copy, the copy number of other sequences can be simply inferred (Figure 2). Moreover, additional sequences can then be analysed to pinpoint the exact boundaries of the affected region.

MCC was first successfully applied to examine the copy-number changes associated with non-reciprocal translocations in renal carcinoma cell lines, from which abundant DNA was available [30]. The ability to examine copy-number changes in very fine detail revealed, in this case, that the boundaries of the genomic regions involved are complex: rather than a simple genomic 'cut and paste', we found local insertions, deletions and inversions at the translocation points. It will be interesting to see what proportion of copy-number changes have similar small-scale disruptions at their boundaries. However, it seemed likely that the method would lend itself well to analysing more difficult samples, such as those from biopsies, because it requires only minuscule amounts of DNA. We therefore set about analysing biopsies of pre-cancerous lung epithelium obtained from the members of the University College London Hospital Early Lung Cancer Project.

The initial results, however, were distorted owing to the extreme fragmentation of the DNA; as with any PCR-based method, longer target sequences are more susceptible than shorter ones to destruction caused by DNA damage, thereby introducing bias into the results and giving erro-

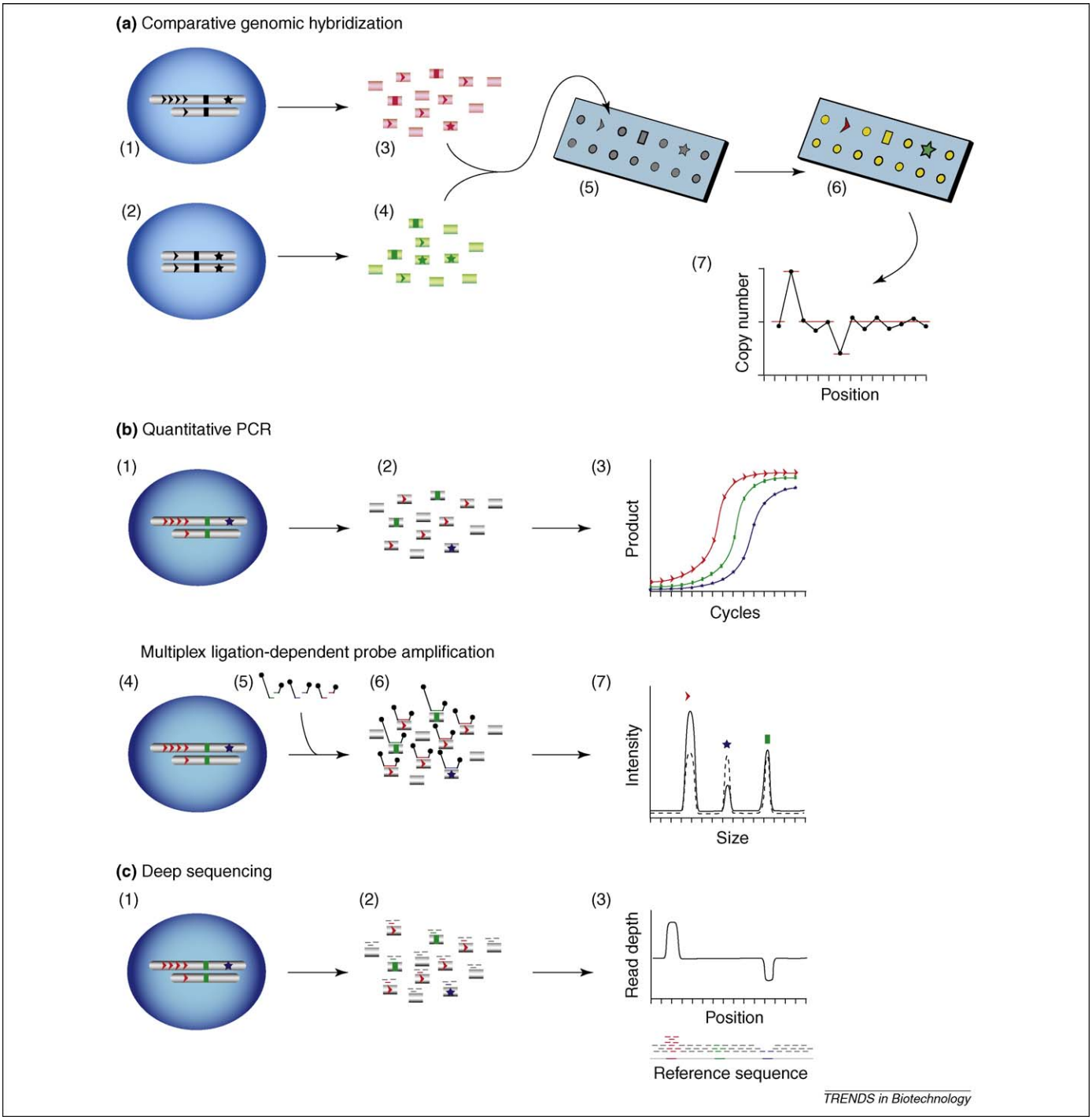

Figure 1. Some of the methods commonly used for analysing CNVs. **(a)** In comparative genomic hybridization (CGH), a test sample (1) is compared to a normal diploid reference sample (2); the grey bars represent a pair of chromosomes. The black chevron, rectangle and star represent three genomic loci that are present at two copies each in the reference sample but are present at five, two and one copy, respectively, in the test sample. DNA is prepared from cells of each type and labelled with either a red fluorophore (3) or a green fluorophore (4). Equal amounts of the two labelled DNAs are then mixed and hybridized to a microarray (5), which contains fixed probes (usually oligonucleotides or cloned fragments) corresponding to the genomic sequences of interest. After hybridization and washing, the slide is imaged and the fluorescence is recorded (6). Sequences that are present at higher copy in the test sample than in the reference (for example, the 'chevron' locus) give a predominantly red fluorescence; those that are present in lower copy in the test than in the reference (for example, the 'star' locus) give a predominantly green fluorescence. Sequences that are present at equal copy in the test and reference samples (e.g. the 'rectangle' locus) give equal amounts of red and green fluorescence and hence appear yellow. The fluorescence ratio (red:green signal) from each spot is quantified and plotted as a graph reflecting relative copy-number as a function of genomic position (7; black points and line); these data are normally smoothed and segmented to reduce noise and give an estimate of the relative copy number along segments of the genome (red bars). Several variations on this procedure are commonly used. Typically, up to tens of thousands of different sequences are analysed in a single experiment but require around a microgram of DNA from the test and reference samples. **(b)** In quantitative PCR (qPCR), DNA from a test sample (1) is prepared; the locus represented by the red chevron is present at five copies per cell, the locus represented by the green rectangle is present at two copies per cell and the locus represented by the blue star is present at one copy per cell. The DNA fragments (2) are amplified by PCR for each of the loci of interest, and the amount of PCR product is monitored at each cycle and plotted as a graph (3). PCR product from the 'red chevron' locus (red curve) appears sooner than that from the less abundant 'green rectangle' and 'blue star' loci, reflecting the relative abundances of these sequences (and hence their relative copy numbers) in the test sample. Many variations on this format are used with a variety of internal references and calibration curves. Typically, such assays are used for only one or a few sequences at a time and are most robust when starting with fairly large amounts (tens of nanograms) of test DNA. Multiplexed ligation-dependent probe amplification (MLPA) (lower schematics) is an enhancement of qPCR, capable of analysing many tens or even hundreds of loci simultaneously [38]. DNA from cells (4) is prepared and mixed with probes (5) for all loci to be analysed. Each probe consists of two DNA molecules whose ends match

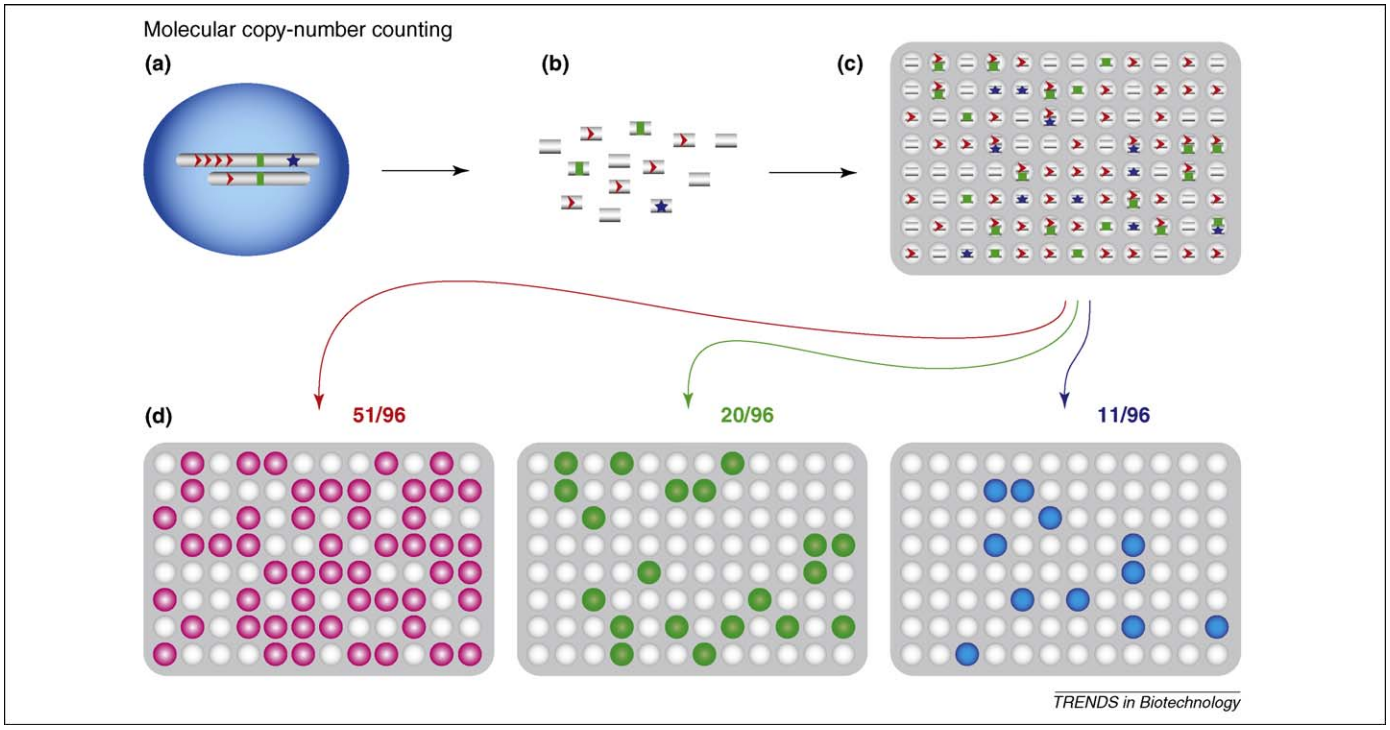

Figure 2. Molecular copy-number counting (MCC). A test sample **(a)** is depicted containing five, two and one copy per cell of the loci represented by the red chevron, green rectangle and blue star, respectively. DNA is prepared from these cells **(b)** and dispensed at limiting dilution into a 96-well microtitre plate **(c)**. At this extreme dilution, individual DNA molecules are peppered across the plate in some, but not all wells. A multiplex PCR procedure is used to detect **(d)** and to count wells that are positive for the 'red chevron', 'green rectangle' and 'blue star' loci. In this example, 51 of the 96 wells test positive for the 'red chevron' locus, 20 test positive for the 'green rectangle' locus and 11 test positive for the 'blue star' locus, reflecting their relative copy numbers (5:2:1) in the test sample. Typically, up to a few hundred loci can be studied in a single experiment, starting from samples containing only a few tens of cells, which corresponds to only tens of picograms of DNA.

neously higher measured copy-numbers for the shorter targets. But by choosing targets of short and approximately uniform length, this problem could be overcome, and this led to the modified method of μMCC. As reported recently [31], μMCC was able to reliably and robustly measure copy-number changes in small numbers of precancerous cells taken from selected regions of fixed, embedded and stained biopsies. Moreover, the ease with which additional targets can be examined (in contrast to, for example, arrays that must be re-made with a new set of probes for each iteration) makes it easy to 'zoom in' on the boundaries of any observed copy-number changes and to identify the precise sequences flanking the genomic amplicon. Thus, we hope to be able to delineate both the stages in carcinogenesis at which regional genomic amplification begins and the precise regions that are amplified across a wide spectrum of patients.

μMCC makes possible the detailed analysis of a wealth of clinical archive specimens, which are difficult to tackle by other methods owing to their low DNA content. In addition, this method might also be particularly beneficial when only small numbers of circulating tumour cells exist or, in the case of lung cancer, when analysing sputum samples. Nevertheless, like all methods, it also has some limitations. In its present form, μMCC is not well suited to the analysis of more than several hundreds of loci at a time. In this regard, microarray-based methods are superior in being able (given sufficient DNA) to scan the entire genome in a single experiment. To address this issue, we are currently attempting to implement μMCC on a microfluidic platform, which will hopefully help to accelerate experiments while reducing the costs, making it practical to analyse more loci.

Location, location, location!

There is one elephant in the room (or perhaps a variable number of elephants) when CNVs are being discussed. Where in the genome do the 'extra' copies lie? CGH data are commonly presented (after suitable smoothing) as a copy-number profile aligned against the reference genome. The peaks and plateaus in such plots give the unintentional and subliminal impression that the additional copies of genomic segments are somehow superimposed on their 'normal' copies at the same point in the genome, like frozen replication bubbles (which, for all we know, some of them might be). Databases such as Ensembl (http://www.ensembl.org) use coloured bars to indicate parts of the genome known to be copy-number variable but do not show where

consecutive parts of the target sequence; they also carry tails (black) which differ in length for each probe but which all carry common sequences at their distal ends (black dots). The probes anneal to the genomic DNA (6), and ligase is used to join the annealed ends. All ligated probes are then amplified simultaneously using PCR primers complementary to the common sequences at the ends of each probe. The products are size-fractionated (for example, by capillary gel electrophoresis) and detected (7), and the intensity of the peak (solid line) corresponding to each probe reflects the relative abundance of the target sequence in the genome. The results are normally compared against a similar profile obtained from a reference genome (dotted line). **(c)** In deep-sequencing approaches, DNA is extracted from cells (1) and sequenced using second-generation technologies (2; short horizontal lines represent individual sequence reads). Sequence reads are then aligned against the reference genomic sequence (3, lower). Regions of increased (red) or decreased (blue) copy number are then identified by their greater or smaller read depth (more or fewer reads aligned against them), respectively (3, upper).

these extra copies lie. In cartoon-style depictions of CNVs, the extra copies are often shown, for convenience, as tandem repeats, with little evidence that this is so (see Figures 1 and 2).

In a few cases, usually those showing large structural abnormalities, the whereabouts of the extra copies are known, but in most cases they are not. Indeed, it is extremely difficult to determine where the further copies of a CNV are located in the genome unless they are large enough to be discerned by fluorescence *in situ* hybridization. Even deep sequencing struggles in this respect: good coverage and a low level of artefacts are needed to reliably identify the new sequences created at the insertion points of a second copy in the genome, and even then it remains difficult or impossible to understand the chromosomal organization of a CNV present on one chromosome when a second, normal chromosome is present. The relatively short read-lengths of most high-throughput sequencing methods also make them vulnerable to being confused by repetitive sequences.

Does this matter? Of course it does, and indeed it matters very much. If CNV has any phenotypic impact, it will arise because the additional copies contain genes that are expressed or mis-expressed, something that we know depends on their genomic environment, or because they create new sequences and disrupt others. The sentence 'The cat sat on cat sat the mat' is the result of a CNV and contains one extra copy of 'cat sat'; but it also contains two new sequences ('on cat' and 'sat the') and has lost one sequence ('on the'). 'The cat sat on the mat cat sat' is identical from a CNV perspective but constitutes a different repertoire of novel sequences. The location of additional copies of copy-number variable sequences is as important as their number, perhaps even more so.

Somatic CNV

Fortunately, in the face of ever-increasing levels of variation being discovered in the human genome, there is at least one touchstone: the genomes of all the cells in a cancer-free individual are the same. However, even in this bedrock, a few cracks have appeared. For example, it is well-known that immune cells undergo rearrangements and deletions at immunoglobulin loci. Cases are also known of genomic mosaicism, even as far as aneuploidy in sub-populations of cells (for example, [32–35]. But, aside from these peculiar exceptions, and barring mutations accumulated as part of the ageing process, the genomes in all of your cells are identical. Or are they?

The only solid arguments *against* widespread genomic variation within a normal individual come from sequencing (but how many individuals have had their genome thoroughly sequenced from two different cell types?) or from whole-animal cloning (but the failure rate of such cloning is enormous, and cloned animals are not usually assessed for donor/clone copy-number anomalies). Copy-number differences between somatic cells could potentially be quite widespread without our being aware of it.

In fact, recently, an intriguing and potentially highly significant paper was published by a team of collaborators in Alabama and Gdansk [36]. They had done what at first glance might seem a rather silly, yet very insightful experiment: a comparative genomic array hybridization between the DNA of pairs of identical (monozygotic) twins. It was a silly experiment because, naturally, monozygotic twins have identical genomes. But, as the authors had no doubt suspected and anticipated, they found copy-number differences at several loci between the members of each pair. In some cases, correlations could be suggested between inter-twin CNVs and phenotypic differences.

If CNVs occur between identical twins, they must also occur between the different tissues of a single individual. After all, the CNVs between twins must have arisen by spontaneous mutations during the early stages of embryogenesis, either just before or just after the embryo split into two individuals. It would be reasonable to assume that CNVs should also arise in cases in which the embryo did

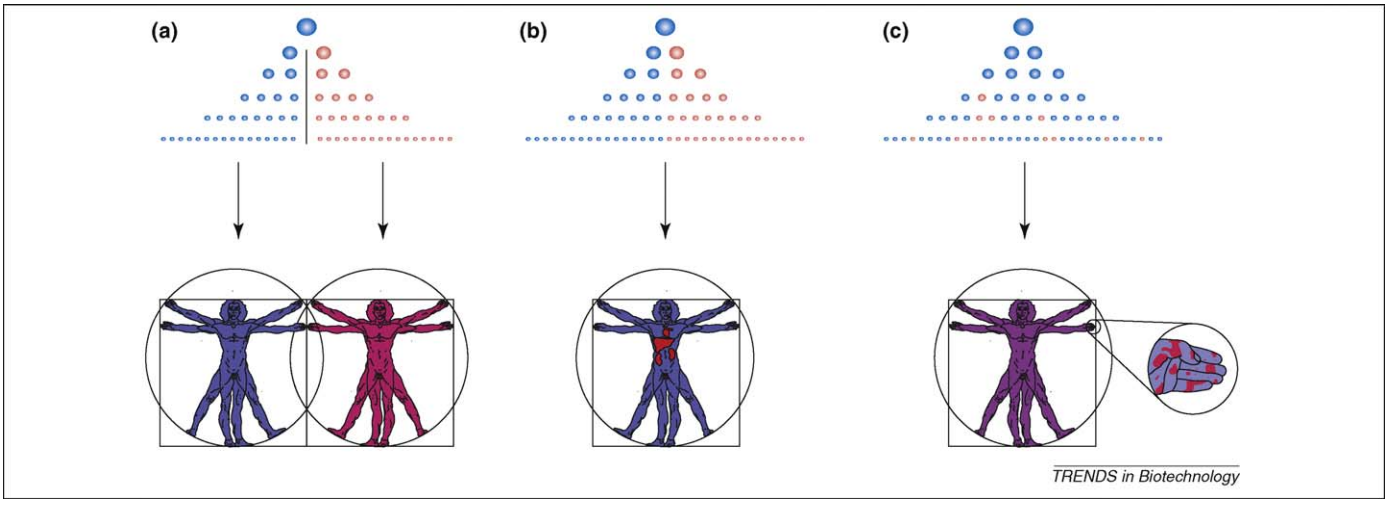

Figure 3. Somatic CNV. The upper parts of each diagram represent the process of cell division from a fertilized zygote; cells shown in red have acquired a spontaneous copy-number change that is then passed on to their descendents. **(a)** Copy-number differences between monozygotic twins, as reported by Bruder *et al.* [36], imply that a spontaneous CNV arose at or about the time at which the developing embryo split (black vertical bar) into two individuals. The result is a pair of monozygotic twins differing at the CNV locus (red and blue human figures below). **(b)** If a spontaneous CNV arose at a similarly early point during the development of a single (non-twinning) embryo, the result would be an individual who is chimeric for copy-number at that locus, with only some tissues (red) showing the new copy-number variant. This phenomenon was reported by Piotrowski *et al.* [37]. **(c)** However, if CNVs arise spontaneously in later stages of embryogenesis or after birth, they will result in an individual with micro-chimerism, in which one or more patches of cells, or individual cells, differ in copy-number from their neighbours.

not split, leading to a copy-number chimerism in the individual (Figure 3).

That this could indeed occur was reported in a follow-up paper [37]. The authors used CGH to reveal copy-number differences between different tissues that were sampled post-mortem from single individuals who had died from apparently non-genomic causes.

Their findings have yet to be widely tested and verified, but we should not be surprised that somatic CNV might indeed be widespread. The genome was not designed in a simple, logical way, and the general rule is instead that 'if it can happen, it will happen'. My suspicion would thus be that somatic CNV will turn out to be quite common, either occurring through genomic 'accidents' similar to those associated with cancer or (possibly) as a regulated process in some cell types, similar to that seen in the immunoglobulin loci of immune cells.

Certainly, if the results of Bruder *et al.* [36] and Piotrowski *et al.* [37] are confirmed and found to be typical, then we should expect to see increasing numbers of somatic CNVs when we look at either small groups of cells or individual cells. As noted by the authors of these two papers, they could only have detected a minority of the potential CNVs within an individual. Variations arising after the first few cell divisions would be confined to small sub-populations of cells and would not have been detected by array hybridization involving several micrograms of DNA extracted from millions of cells. If somatic CNVs arise with an approximately equal probability at each cell division, we should not be surprised to identify many more of them if we compare the genomes of small numbers cells (Figure 3). We hope that μMCC will be applicable in this context.

Might somatic CNVs have a large impact on the study of genomic disease? Surely, they cannot be *very* common, as they would have been noticed before. However, CNV is considerably more difficult to detect than point mutations and, in any case, many diseases that seem to have a genetic cause (but lack, as yet, a genetic explanation) are also quite rare. It is not inconceivable that some of these diseases are attributable to copy-number changes in the affected tissues that are not present in the blood or buccal cells commonly used for genomic analysis.

If (and there is still a big 'if') somatic CNVs turn out to be a real and general phenomenon, this would have profound consequences for all aspects of molecular biology and, particularly, for the growing field of personal genomics. Giving a few drops of blood or a swab of buccal cells for genomic analysis is straightforward: let's hope it does not become necessary to have a whole-body biopsy to obtain a complete picture of one's genome.

Conclusions

Much like SNPs, CNVs in the human genome have long been known to exist but have only recently been shown to be extremely abundant and to contribute much of the genomic diversity between individuals. Just as the explosion of SNP data began with a cataloguing of polymorphisms, so CNV research has, until recently, concentrated on identifying the variants and measuring their prevalence.

CNVs, however, are likely to prove far more complex than SNPs. For any given single-nucleotide substitution, at most four alternatives exist. For any given CNV, the precise boundaries of the affected sequence might differ from person to person. In the case of increases in copy number, the number of extra copies and their precise locations in the genome might also differ. Additional copies of a genomic segment might also, of course, harbour their own SNPs. All of these factors make CNVs far harder to catalogue and interpret than SNPs alone.

The question of the genomic location of additional copies of CNVs is of prime importance and is not fully addressed by current approaches. A second copy of a gene might result in: (i) additional expression if the new copy is functional; (ii) no change in expression if the genomic location of the new copy is unfavourable; (iii) the disruption of another gene at the point where the new copy is inserted; or (iv) the creation of an entirely new chimeric protein. A partial second copy might even abolish expression of the first copy by inserting within it. Hence, efforts to find a simple correlation between copy-number changes and disease might in many cases be frustrated.

Add to this the as-yet-unconfirmed possibility of widespread somatic CNV, and it is clear that the human genome will continue to challenge us for many years to come.

Disclosure statement

A patent for molecular copy-number counting has been applied for by the UK Medical Research Council; P.H.D. is one of the named authors of this patent.

References

1 Cargill, M. *et al.* (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* 22, 231–238

2 Sachidanandam, R. *et al.* International SNP Map Working Group (2001) A map of human genome sequence variation containing 1,42 million single nucleotide polymorphisms. *Nature* 409, 928–933

3 Craig-Holmes, A.P. *et al.* (1973) Polymorphism of human C-band heterochromatin. I. Frequency of variants. *Am. J. Hum. Genet.* 25, 181–192

4 Carnevale, A. *et al.* (1976) The segregation of C-band polymorphisms on chromosomes 1, 9 and 16. *Am. J. Hum. Genet.* 28, 412–416

5 Ottolenghi, S. *et al.* (1974) The severe form of alpha thalassaemia is caused by a haemoglobin gene deletion. *Nature* 251, 389–392

6 Miller, O.J. and Breg, W.R. (1976) Autosomal chromosome disorders and variations. *N. Engl. J. Med.* 294, 596–598

7 Sebat, J. *et al.* (2004) Large-scale copy number polymorphism in the human genome. *Science* 305, 525–528

8 Iafrate, A.J. *et al.* (2004) Detection of large-scale variation in the human genome. *Nat. Genet.* 36, 949–951

9 McCarroll, S.A. *et al.* (2008) Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.* 40, 1166–1174

10 Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921

11 Venter, J.C. *et al.* (2001) The sequence of the human genome. *Science* 291, 1304–1351

12 Lejeune, J. *et al.* (1959) Mongolism; a chromosomal disease (trisomy). *Bull. Acad Natl. Med.* 143, 256–265

13 Buckland, P.R. (2003) Polymorphically duplicated genes: their relevance to phenotypic variation in humans. *Ann. Med.* 35, 308–315

14 Tuzun, E. *et al.* (2005) Fine-scale structural variation of the human genome. *Nat. Genet.* 37, 727–732

15 Redon, R. *et al.* (2006) Global variation in copy number in the human genome. *Nature* 444, 444–454

16 Sharp, A.J. *et al.* (2005) Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* 77, 78–88
17 Estivill, X. and Armengol, L. (2007) Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genet.* 3, e190
18 Singleton, A.B. *et al.* (2003) α-Synuclein locus triplication causes Parkinson's disease. *Science* 302, 841
19 Sleegers, K. *et al.* (2006) APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy. *Brain* 129, 2977–2983
20 Rovelet-Lecrux, A. *et al.* (2006) APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat. Genet.* 38, 24–26
21 Cabrejo, L. *et al.* (2006) Phenotype associated with APP duplication in five families. *Brain* 129, 2966–2976
22 Gonzalez, E. *et al.* (2005) The influence of *CCL3L1* gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 307, 1434–1440
23 McKinney, C. *et al.* (2008) Evidence for an influence of *chemokine ligand 3-like 1* (*CCL3L1*) gene copy number on susceptibility to rheumatoid arthritis. *Ann. Rheum. Dis.* 67, 409–413
24 Shlien, A. *et al.* (2008) Excessive genomic DNA copy number variation in the Li-Fraumeni cancer predisposition syndrome. *Proc. Natl. Acad. Sci. U. S. A.* 105, 11264–11269
25 Tonon, G. *et al.* (2005) Common and contrasting genomic profiles among the major human lung cancer subtypes. *Cold Spring Harb. Symp. Quant. Biol.* 70, 11–24
26 Andre, F. *et al.* (2009) Molecular characterization of breast cancer with high-resolution oligonucleotide comparative genomic hybridization array. *Clin. Cancer Res.* 15, 441–451
27 Little, S.E. *et al.* (2006) Array CGH using whole genome amplification of fresh-frozen and formalin-fixed, paraffin-embedded tumor DNA. *Genomics* 87, 298–306
28 Bredel, M. *et al.* (2005) Amplification of whole tumor genomes and gene-by-gene mapping of genomic aberrations from limited sources of fresh-frozen and paraffin-embedded DNA. *J. Mol. Diagn.* 7, 171–182
29 van Beers, E.H. *et al.* (2006) A multiplex PCR predictor for aCGH success of FFPE samples. *Br. J. Cancer* 94, 333–337
30 Daser, A. *et al.* (2006) Interrogation of genomes by molecular copy-number counting (MCC). *Nat. Methods* 3, 447–453
31 McCaughan, F. *et al.* (2008) Microdissection molecular copy-number counting (microMCC) – unlocking cancer archives with digital PCR. *J. Pathol.* 216, 307–316
32 Ferrier, S. (1964) Mongol child-mosaic parent. Study of 2 families. *J. Genet. Hum.* 13, 315–336
33 Rittinger, O. *et al.* (2008) The changing phenotype in diploid/triploid mosaicism may mimic genetic syndromes with aberrant genomic imprinting: follow up in a 14-year-old girl. *Eur. J. Med. Genet.* 51, 573–579
34 Yang, A.H. *et al.* (2003) Chromosome segregation defects contribute to aneuploidy in normal neural progenitor cells. *J. Neurosci.* 23, 10454–10462
35 Westra, J.W. *et al.* (2008) Aneuploid mosaicism in the developing and adult cerebellar cortex. *J. Comp. Neurol.* 507, 1944–1951
36 Bruder, C.E. *et al.* (2008) Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *Am. J. Hum. Genet.* 82, 763–771
37 Piotrowski, A. *et al.* (2008) Somatic mosaicism for copy number variation in differentiated human tissues. *Hum. Mutat.* 29, 1118–1124
38 Schouten, J.P. *et al.* (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* 30, e57