Genome **Medicine**

MUSINGS

DNA structure matters

Albino Bacolla[1]*, David N Cooper[2] and Karen M Vasquez[1]

The human genome and structural variation

One of the most intriguing findings in the wake of the release of the reference genome sequence from the Human Genome Project has been the realization of the extent to which each individual genome differs, not only in terms of single nucleotide polymorphisms, but also in terms of large deletions, duplications and other re-arrangements, a phenomenon now referred to as copy number variation. Some 3,000 protein-coding genes (around 10% of the human gene complement) are known to be associated with copy number variants (CNVs), and two unrelated human genomes may therefore differ quite dramatically in terms of their gene content. Indeed, it is becoming increasingly evident that CNVs are a major source of genetic variation, contributing not only to phenotypic traits but also to inherited disease. A growing number of reports support the role of CNVs in the etiology of complex genomic disorders, such as the Smith-Magenis and Potocki-Lupski syndromes, Charcot-Marie-Tooth disease 1A (CMT1A) and hereditary neuro-pathy with liability to pressure palsies (HNPP), Sotos syndrome, Williams-Beuren syndrome, Pelizaeus-Merz-bacher disease and autism, among others [1].

In light of these findings, it is clear that the nature of the mechanisms underlying CNV formation is of central importance, from both a theoretical and a clinical standpoint. Analyses of CNVs in humans and across lines of *Drosophila melanogaster* have revealed that the sites of chromosomal rearrangements are characterized by either stretches of homology, or little to no homology at all, suggesting that both non-allelic homologous recombina-tion and homology-independent repair are likely to lead to CNV formation. A study [2] also showed that DNA sequences flanking CNV breakpoints often contain repetitive sequence motifs known to form alternative DNA structures, or non-B DNA (various non-canonical types of DNA, including left-handed Z-DNA, triplexes, G-quadruplexes, cruciform and slipped structures). This

is an important conclusion since it implies that DNA structure, rather than the sequence *per se*, may pre-dispose to chromosomal breakage and subsequent repair, thereby promoting CNV formation. These results [2] expand observations made earlier by a number of laboratories, including our own, using different analyses and model systems [3,4]. Recent molecular analyses of novel CNVs, such as the *NRXN1* region associated with autism spectrum and other neurodevelopmental dis-orders [5], and non-recurrent microdeletions of the *FOXL2* gene associated with blepharophimosis-ptosis-epicanthus-inversus syndrome, also support the above conclusions.

What are non-B DNA sequences?

Soon after Watson and Crick's description of the canonical right-handed double-helical B-form of DNA in 1953, it was discovered that the DNA helix can assemble into other structures, and a wealth of information from biophysical studies has served to characterize these non-canonical or non-B structures. The most common include left-handed Z-DNA formed by alternating pyrimidine-purine bases, quadruplex DNA formed by four arrays of two to four guanines each and exemplified by the human telomeric $(TTAGGG)_4$ motif, triplex or H-DNA formed by purine-rich motifs containing mirror repeat symmetry, and cruciform and slipped-out struc-tures formed by inverted and direct repeats, respectively [4]. Basic research over the past few years has been instrumental in demonstrating that non-B-DNA-forming motifs are abundant in mammalian genomes and that specific antibodies or small molecules can be used to detect the resulting non-B structures in living cells. Under certain circumstances, such structures elicit speci-fic cellular responses that may be monitored experi-mentally. For example, Schwab *et al.* [6] found that the absence of the helicase gene *FANCJ* in cultured chicken DT40 cells led to a decrease in replication fork velocity and the accumulation of single-stranded gaps, especially in cells treated with telomestatin, a small molecule that binds and stabilizes quadruplex DNA. The authors postulated that FANCJ prevents the DNA replication machinery from being arrested by physical obstacles such as non-B DNA structures, resolving these via its helicase activity. In the absence of FANCJ, the lagging strand

*Correspondence: albino.bacolla@austin.utexas.edu
[1]Division of Pharmacology and Toxicology, College of Pharmacy, Dell Pediatric Research Institute, The University of Texas at Austin, Austin, TX 78723, USA
Full list of author information is available at the end of the article

polymerase delta is forced to bypass the obstacle-containing Okazaki fragments, leaving behind single-stranded regions and inducing local reorganization of the chromatin. These results are particularly interesting since mutations in *FANCJ* cause the cancer-predisposing disorder Fanconi anemia, characterized by a failure to repair complex DNA lesions, and raise the possibility that rapidly proliferating cancer cells may represent a target for chemotherapeutics that can synergistically stabilize non-B DNA structures and inhibit their clearance [6].

The nuclear genome is not unique in harboring mutations mediated by non-B DNA. The occurrence of intrinsically bent DNA (caused by runs of adenine base pairs known as A-tracts), triplex-forming and quadruplex-forming sequences has been noted in the vicinity of high-frequency mitochondrial genome deletions. Recently, Damas *et al.* [7] reported a detailed analysis of the potential for sections of the mitochondrial genome to adopt stable fold-back (hairpin and cloverleaf-like) structures. This study provides evidence for the role of complex DNA secondary structures in mediating mitochondrial genome deletions, which are associated with various pathologies.

How does non-B DNA form and trigger genomic instability?

Although the full range of generative mechanisms remains to be elucidated, both transcription and DNA replication have been shown to facilitate non-B DNA formation, not only on the separated single DNA strands but also as a consequence of the negative torsional stress they leave behind during translocation. Hence, non-B DNA is likely to form more readily during the S-phase of the cell cycle in rapidly dividing cells than in quiescent cells. Once non-B structures have been formed, at least two mechanisms have been proposed to account for chromosomal breakage: the first is an increase in oxidative damage that has been noted at selected bases within or adjacent to non-B structures [8]; the second is the recognition of these structures by DNA repair or other structure-specific enzymes which, in some cases, induce a DNA damage response [4]. Inagaki *et al.* [9] have elegantly shown that cruciform structures formed on human chromosomes 22 and 11 promote recurrent t(11;22)(q23;q11.2) constitutional translocations (balanced karyotype that would have disease consequences in the offspring) in cell culture, and are recognized by the Holliday junction resolvase GEN1. Following chromosomal breakage, chromosomal fusion proceeds through end-processing via Artemis, a nuclease that promotes homologous recombination and V(D)J recombination, and non-homologous end-joining (NHEJ) proteins. Thus, there appears to be coordination between proteins from distinct repair pathways (homologous recombination and NHEJ) in processing these non-B DNA structures. One intriguing aspect pointed out by Inagaki *et al.* [9] is that, *in vivo*, t(11;22)(q23;q11.2) is observed only in sperm cells but not in somatic cells. Likewise, in cell culture, the translocation is detected only when the appropriate sequences are provided *in trans* on a plasmid, not on the endogenous chromosomes. The authors suggest that, because formation of the large cruciforms leading to t(11;22)(q23;q11.2) and other translocations requires very high levels of torsional stress, this can only be achieved in sperm cells, a life stage in which nucleosome-based chromatin is transiently replaced by a protamine-based configuration. This 'protein swap' would generate widespread torsional stress, which would then promote the formation of non-B DNA. Hence, further work will be required to visualize non-B DNA structures *in vivo* and fully elucidate the factors that facilitate their formation.

In light of the findings outlined above, and the likelihood that not all sequences with the capacity to adopt non-B DNA structures may serve a biological function (for example, many are located outside of genes), their abundance in eukaryotic genomes remains puzzling. One thought-provoking possibility, proposed by Begum and Honjo [10] in the context of the similarities between the genetic mechanisms underlying genome diversity and immune system antibody diversity (such as V(D)J recombination, somatic hypermutation and class-switch recombination), is that the formation of non-B DNA is a prerequisite step in both processes. It may therefore be that the conservation of non-B DNA in extant genomes, and hence the associated disease risk arising from CNVs and other mutations, is the price that we pay for this non-canonical form of DNA having played a key role in our evolutionary history, including development of the recognition of 'self' and defense against external pathological agents.

Author details
[1]Division of Pharmacology and Toxicology, College of Pharmacy, Dell Pediatric Research Institute, The University of Texas at Austin, Austin, TX 78723, USA. [2]Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff CF14 4XN, UK.

References
1. Lupski JR, Stankiewicz PT: *Genomic Disorders: The Genomic Basis of Disease.* Totowa, New Jersey: Humana Press; 2006:426.

2. Cardoso-Moreira M, Arguello JR, Clark AG: **Mutation spectrum of Drosophila CNVs revealed by breakpoint sequencing.** *Genome Biol* 2012, **13**:R119.
3. Cooper DN, Bacolla A, Ferec C, Vasquez KM, Kehrer-Sawatzki H, Chen JM: **On the sequence-directed nature of human gene mutation: the role of genomic architecture and the local DNA sequence environment in mediating gene mutations underlying human inherited disease.** *Hum Mutat* 2011, **32**:1075-1099.
4. Zhao J, Bacolla A, Wang G, Vasquez KM: **Non-B DNA structure-induced genetic instability and evolution.** *Cell Mol Life Sci* 2010, **67**:43-62.
5. Chen X, Shen Y, Zhang F, Chiang C, Pillalamarri V, Blumenthal I, Talkowski M, Wu BL, Gusella JF: **Molecular analysis of a deletion hotspot in the *NRXN1* region reveals the involvement of short inverted repeats in deletion CNVs.** *Am J Hum Genet* 2013, **92**:375-386.
6. Schwab RA, Nieminuszczy J, Shin-Ya K, Niedzwiedz W: **FANCJ couples replication past natural fork barriers with maintenance of chromatin structure.** *J Cell Biol* 2013, **201**:33-48.
7. Damas J, Carneiro J, Goncalves J, Stewart JB, Samuels DC, Amorim A, Pereira F: **Mitochondrial DNA deletions are associated with non-B DNA conformations.** *Nucleic Acids Res* 2012, **40**:7606-7621.
8. Clark DW, Phang T, Edwards MG, Geraci MW, Gillespie MN: **Promoter G-quadruplex sequences are targets for base oxidation and strand cleavage during hypoxia-induced transcription.** *Free Radic Biol Med* 2012, **53**:51-59.
9. Inagaki H, Ohye T, Kogo H, Tsutsumi M, Kato T, Tong M, Emanuel BS, Kurahashi H: **Two sequential cleavage reactions on cruciform DNA structures cause palindrome-mediated chromosomal translocations.** *Nat Commun* 2013, **4**:1592.
10. Begum NA, Honjo T: **Evolutionary comparison of the mechanism of DNA cleavage with respect to immune diversity and genomic instability.** *Biochemistry* 2012, **51**:5243-5256.