

CNV evolution revisited

The size and sheer number of copy number variants (CNVs) in the human genome puts them on a par with SNPs for their potential importance to health and evolution. Previous reports had reasoned that CNVs were frequent because they were beneficial. A more comprehensive analysis now concludes instead that CNVs might persist simply because the genome has difficulty getting rid of them. Knowing how CNVs arise and how selection acts on them should help us to better understand their contribution to genomic variation and disease.

The theory that CNVs have been positively selected was based on the high gene density within CNV-containing regions (CNVRs), as well as their elevated rate of protein evolution. The availability of additional genome-wide CNV data

sets prompted the authors to revisit this hypothesis by asking whether patterns seen in previous studies could be explained by alternative, non-adaptive mechanisms. Four CNV data sets, together consisting of over 6,000 genomic regions, were subjected to two types of analysis to consider which mutational processes give rise to CNVs and the impact of selection on them.

One previously undetected feature of CNVRs is that they are strikingly GC rich — the longer the CNVR, and the more frequent it is, the higher its GC content. Importantly, CNVRs that overlap with variable segmental duplications are more GC rich than those that do not — an observation that the authors attribute to the propensity of GC-rich regions to generate segmental duplications, and to lead to CNVs.

Previously, rapid evolution of CNVRs had been thought to mean that such regions are frequently under positive selection. However, an alternative explanation is given here: recombination in the vicinity of segmental duplications is reduced, thereby leading to the accumulation of mutations that are deleterious but hard to weed out. The fact that the rate of protein evolution is higher in CNVRs that are associated with segmental duplications than in those that are not supports this idea.

CNVRs are enriched for environmental response genes, which

bolstered the idea that CNVRs were positively selected. But what if this bias was caused simply by selection being weaker on non-essential environmental genes? A survey of the genomic distribution of essential genes found that these are indeed under-represented in CNVRs, which could mean that copy number changes in these genes are deleterious. If CNVRs are dangerous places for genes to be, why are they so gene rich? This might simply reflect the fact that genes more frequently reside in GC regions, which are themselves more prone to creating CNVRs.

The genetic features of CNVRs can therefore — based on a data set 10 times larger than before — be explained by non-adaptive processes related to GC content and the low rate of recombination in duplicated regions. This knowledge might aid our ability to understand the phenotypic effects of CNVs: if most copy number changes are deleterious, then disease-causing mutations are most likely to fall outside segmental duplications.

Tanita Casci

ORIGINAL RESEARCH PAPER Nguyen, D.-Q. *et al.* Reduced purifying selection prevails over positive selection in human copy number variant evolution. *Genome Res.* 25 Sep 2008 (doi: 10.1101/gr.077289.108)

FURTHER READING Feuk, L. *et al.* Structural variation in the human genome. *Nature Rev. Genet.* 7, 85–97 (2006)

