Box 3 | **The main components of eukaryotic genomes**

Protein-coding genes

Although most prokaryotic chromosomes consist almost entirely of protein-coding genes[86], such elements make up a small fraction of most eukaryotic genomes (see figure). As a prime example, the human genome might contain as few as 20,000 genes, comprising less than 1.5% of the total genome sequence[16,82].

Introns

Shortly after their discovery, the non-coding intervening sequences within coding genes (introns) were suggested to account for the pronounced discrepancy between gene number and genome size[7]. It has also recently been suggested that most non-coding DNA in animals (but not plants) is intronic, which would imply that most of the genome is transcribed even though protein-coding regions represent a tiny minority[107,108]. At the very least, introns were found to account for more than a quarter of the draft human sequence[16]. Over a broad taxonomic scale, intron size and genome size are positively correlated[109], although within genera a correlation might (for example, *Drosophila*[110]) or might not (for example, *Gossypium*[111]) be observed.

Pseudogenes

Non-functional copies of coding genes, the original meaning of the term 'junk DNA', were once thought to explain variation in genome size[4]. However, it is now apparent that even in combination, 'classical pseudogenes' (direct DNA to DNA duplicates), 'processed pseudogenes' (copies that are reverse transcribed back into the genome from RNA and therefore lack introns) and 'Numts' (nuclear pseudogenes of mitochondrial origin) comprise a relatively small portion of mammalian genomes. The human genome is estimated to contain about 19,000 pseudogenes[46].

Transposable elements

In eukaryotes, transposable elements are divided into two general classes according to their mode of transposition. Class I elements transpose through an RNA intermediate. This class comprises long interspersed nuclear elements (LINEs), endogenous retroviruses, short interspersed nuclear elements (SINEs) and long terminal repeat (LTR) retrotransposons. Class II elements transpose directly from DNA to DNA, and include DNA transposons and miniature inverted repeat transposable elements (MITEs).

Transposable elements (and especially their extinct remnants) make up a large portion of the human genome, with some elements (for example, the SINE *Alu* element) present in more than a million copies. Transposable-element evolution involves complex interactions with the host genome and other subgenomic elements, ranging from parasitism to mutualism. For a review of transposable-element structure, origins, impacts and evolution see REF. 17.

The figure provides a summary of the different components of the human genome. Less than 1.5% of the genome consists of the suspected 20,000–25,000 protein-coding sequences. By contrast, a large majority is made up of non-coding sequences such as introns (almost 26%) and (mostly defunct) transposable elements (nearly 45%). Data are taken from REF. 16.