Why does less than 4% of the human genome encode functional genes?

Introduction

- 3300 million bases, estimates of up to 100,000 genes. Only 4% of genome!
- Why have such a large genome – lots of ATP used in replication for division
- Certain aspects have been identified suggesting some possible functions, most 'junk'
- The nature of genomes can be investigate by Cot analysis – kinetics of reannealing
- Identified 3 types of DNA – highly repetitive (satellites), intermediate repetitive (transposable elements) and non-repetitive (genes, intergenic DNA and introns)
- There is no known function for most of the non coding DNA

Introns

- Every gene has introns – non coding DNA in between exons, variable length.
- Can accumulate mutation without affecting gene. Sequence unimportant *except* splice site consensus sequences – GU – AG – rule: Involvement of snRNP's U1 – U6.
- Introns removed by splicing as part of the processing of mRNA to form mature mRNA
- Thought to be junk DNA with no function but: *Alternative RNA splicing* suggests function
- Splice mRNA at different points and produce different polypeptides from one gene
- Increases coding capacity of genes and offers a way of regulating gene expression

Functional Chromosome

- DNA replication origin is a sequence where replication is initiated upon binding of pre-replication complex – involving licensing factors – ORC, Cdc6 Mcm2-7.
- Efficient cell division requires chromosomes to join spindle via kinetechore. Achieved by having centromere – holds sister chromatids together by having special chromatin organisation. Consists of satellite sequences. Possibly protein binding sequences also
- Telomeres. Lagging strand cannot be synthesised all the way to the end of chromosome, due to 5' – 3' nature of polymerases and the need for primers to initiate. Each division the end of the chromosomes would shorten – eventually impinging on coding DNA. Solution: at ends of chromosomes are extensive repeat sequences of ~6bp = telomeres. In dividing cells, telomere length maintained by reverse transcriptase – telomerase – adding repeats.

Tandemly repeated DNA

- Satellites – large repetitions of short nucleotide sequence (<10% genome) not transcribed. No known function of sequences. Concentrated in centromere region – possible function?
- Micro- and Minisatellites – repeats of shorter sequences throughout genome. Polymorphism in repeat number (DNA fingerprinting); high mutation rate - . Can cause replication slippage and unequal crossing over. Pathogenic – CGG repeats and Fragile X.

Transposable elements

- Longer repetitive sequences dispersed single copies not in tandem arrays. 'Selfish DNA'
- Are mobile – or more usually were mobile in the past – transposition. Can cause change in gene expression by their movement in genome. There are two types:
- a) DNA intermediate – part of transposon codes for an enzyme that excises and then inserts itself at a different locus. Conserves number of elements
- b) RNA intermediate – via reverse transcriptase (retroelements) – associated with retroviruses. Increases number of elements. Examples include L1 elements (<15% genome) and Alu elements. Such elements make up a large portion of human genome.