

 COMPLEX TRAITS

Repeat, repeat, repeat — gene expression variability explained



PhotoDisc/Getty Images

Short tandem repeats (STRs), which constitute ~1% of the human genome and represent a large source of genetic, and potentially phenotypic, variation, are often considered to be ‘junk’. This view must now be revised in light of the findings by Gymrek *et al.* who have identified STRs on a genome-wide scale that regulate expression of nearby genes — so-called expression STRs (eSTRs).

To overcome the technical difficulties of genotyping STRs (defined as periodic DNA motifs of 2–6 bp spanning a median length of ~25 bp), the team had previously developed the algorithm lobSTR, and used it to catalogue STR variations in whole genomes of individuals in the 1000 Genomes Project. Now, they have linked variations in STR length of 311 of these individuals to the expression levels of nearby genes (obtained by RNA sequencing of the same 311 individuals in the gEUVADIS project), which led to the identification of 2,060 protein-coding genes whose expression was associated with a nearby eSTR variation. The eSTR association signals were robust and reproducible across populations and expression analysis platforms, as 83% of ~800 of the identified eSTRs were confirmed in

an independent cohort using Illumina expression profiles from individuals of African, Asian, European and Mexican ancestry.

“STRs make a significant contribution to gene expression”, says Melissa Gymrek (MIT/Harvard), first author of the study. Indeed, linear mixed model analyses confirmed that many of the identified eSTRs link directly to the gene expression level and could not be explained by tagging other variants in the vicinity of the genes. Moreover, these analyses revealed that 10–15% of the heritable variation in gene expression between individuals is determined by eSTRs.

To further support a regulatory role of eSTRs, Gymrek *et al.* performed integrative genomic analyses. The identified eSTRs reside in conserved regions, co-localize with functional elements including transcription start sites and predicted enhancers, and show strong enrichment in peaks of histone modification.

Although the role of STRs in some Mendelian diseases such as Huntington disease is known, their role in complex quantitative human traits remains elusive. Gymrek *et al.* found significant enrichment of eSTR-associated genes

in genome-wide association studies (GWAS) to predict genes involved in autoimmune diseases, including Crohn’s disease and rheumatoid arthritis. In addition, 12 significant associations between eSTRs and clinical phenotypes were found in ~1,700 individuals from the TwinsUK and UK10K cohorts; 11 of these associations implicated genes not previously detected by GWAS. Incorporating eSTR associations in future GWAS and disease mapping studies will probably lead to the discovery of new genetic variants relevant to human conditions.

The current study was limited by a number of factors, including the facts that STRs have a high genotyping error rate; that 10% of the loci could not be analysed because they are too long to be picked up by current sequencing technologies; that the study assumed a linear relationship between STR length and gene expression; and that only STR length was taken into account without distinguishing between the effect of sequence variations within STRs of identical length. “This is just the tip of the iceberg,” concludes Gymrek, “we believe there are many more eSTRs to be found as we analyse larger and higher powered datasets.”

Liesbet Lieben, Associate Editor,
Nature Reviews Disease Primers

Gymrek, M. *et al.* Abundant contribution of short tandem repeats to gene expression variation in humans. *Nature Genet.* <http://dx.doi.org/10.1038/ng.3461> (2015)