

## Talk

## Searching for novel genes and pseudogenes in the human Y chromosome based on ancestral coding signals



Alejandro Rubio Valle, Carlos Sánchez Casemiro-Soriguer, Juan Jiménez Martínez and Antonio J. Pérez Pulido(1)

1. Área de Genética, Centro Andaluz de Biología del Desarrollo, Ctra. de Utrera, km. 1, 41013, Sevilla.

**Keywords:** human Y chromosome; coding sequences; pseudogenes

### ABSTRACT

**Motivation:** Human Y chromosome has several features that contribute to an extreme variation due to the lack of a homologous partner for crossing over, high rate of sequence amplification and low evolutionary pressure [1]. For these reasons, we think that the Y chromosome could be a perfect candidate in order to discover new coding and fossil regions such as pseudogenes.

Genome finding is one of the greatest hits in modern biology. However, in silico identification of small and complex coding sequences is still challenging. Jiménez et al [2] developed AnAblast, a computer tool which has been successful in uncovering new genes, as well as fossil-coding sequences. This program generates profiles of accumulated alignments of conserved coding signals using a low-stringency BLAST strategy [2].

**Methods:** We have used AnAblast to localize new coding regions in the chromosome Y. After that, AnAblast-generated profiles were introduced into a genome browser, along with other informative data such as repeats and RNA expression data. The candidate's list obtained was complemented by careful BLAST, InterPro and peaks analysis. Moreover, we performed a search on the tool Genome Data Viewer (GDV) to check each result.

**Results:** We have been able to identify some chromosome Y regions that fulfill different requirements: (1) regions without previous annotations as pseudogenes, genes or non-coding regions (Ensembl track); (2) regions without previous annotations as interspersed repeats and low complexity (RepeatMasker track); and (3) regions with expression profiles (RNA-seq of testis). The best candidate to be a new coding region was localized at Y:9912876-9919657 (-). Blast and InterPro analysis indicated similarity with serine-proteases which are found in rodents and another organism such as *Rousettus aegyptiacus* (Egyptian fruit bat). After the search on GDV, we observed that only the first bat's exon was not found in our candidate. In spite of this, we found a methionine codon in our candidate (more specifically in the first exon). Furthermore, the Y chromosome has a 5'-truncated copy of this region.

**Conclusions:** We have found some chromosome Y regions which could be new coding genes or pseudogenes. Thus, this in silico research provides a powerful protocol to search novel genes and fossil regions in the whole human genome. Although we added several RNA-seq tracks that showed the expression of these regions, clinical trials should be performed to verify our candidates.

### REFERENCES

1. Hughes, J. F. & Page, D. C. (2015) The Biology and Evolution of Mammalian Y Chromosomes. *Annu. Rev. Genet.* 49:507–27
2. Jimenez, J. et al. (2015) AnAblast: a new in silico strategy for the genome-wide search of novel genes and fossil regions. *DNA Res.* 22:439-449.