Submitted: March 2010, Master's Thesis

Department of Computational Biology, Graduate School of Frontier Science, The University of Tokyo

Immigrants to the Nucleus; Analysis of Mitochondrially Derived Nuclear Genomic Regions (NUMT)

Name: Junko Tsuji
Student ID: 47-086914
Advisor: Paul Horton, associated professor

Mitochondria are cellular organelles which were originally diverged from α-proteobacteria. They contain their own genomes aside from the nuclear genome. In mammals, each double-stranded circular mtDNA consists of approximately 16,000 base pairs. The two strands of mitochondrial DNA (mtDNA) are discriminated by their nucleotide content with the guanine rich strand called the heavy chain, and the cytosine rich strand called the light chain.

In the endosymbiotic evolutionary process, many mtDNA fragments were transferred to the nucleus. This early phase of mtDNA transfer was thought to provide massive relocation of mitochondrial genes to nuclear chromosomes. This conception comes from that the extant mitochondrial proteome is now overwhelmingly encoded by the nuclear genome. However, in the present day, the transfer of genetic element is extremely rare or has ceased in most eukaryotes. Despite this fact, mtDNA are still continuously transferred to the nucleus, producing mtDNA-like nuclear domains called "NUMTs"; NUclear MiTochondrial-like DNAs. NUMTs have been identified over 70 eukaryotes, and their numbers widely differ in each species; some species retain several hundred NUMTs, but others have no detectable NUMTs at all. It was hypothesized that this variation of NUMTs population arises from the difference of mtDNA copy number and its length in each species. However, this hypothesis does not explain some cases, so the reasons behind NUMTs population difference still remain unclear.

Recent studies indicate that the creation of NUMTs is mediated by non-homologous end joining repair; mtDNA fragments are inserted and joined with nuclear break ends when nuclear double strand breaks (DSB) occur (Figure 1). In previous NUMTs studies, the patterns mtDNA contributing to NUMTs and features of NUMTs insertion site were investigated. Most studies indicated transferred mtDNA and nuclear NUMTs insertion site were randomly chosen. Moreover, in the functional NUMTs survey,

it was suggested that most NUMTs exists in introns or intergenic regions. Because of the difference of genetic codes between mitochondria and the nucleus, NUMTs tend to be "dead-on-arrival" pseudogenes. However, several studies provided the evidence of functional NUMTs in particular species (e.g. plants, yeasts, and flies). In humans, only one functional human NUMT has been proposed in the literature: Christos *et al.* discovered a potential functional NUMT in the 3'UTR of the nuclear receptor coactivator2 mRNA (NCOA2). Additionally, the role of somatic cell NUMTs in disease has been demonstrated.

In this study, we mainly focused on human NUMTs. Additionally, also rhesus, mouse and rat NUMTs were investigated for testing and confirming our observations. From analyzing our original datasets, we discovered a specific pattern of mtDNA migration and characteristics of nuclear NUMT integration sites. Our result suggests that the mitochondrial promoter region and its peripheral domains (548bp-1142bp counted by D-loop as the origin) were seldom transferred. In NUMT preferentially integrated sites of human genome, AT-rich oligomers appeared in all NUMTs flank and 90% of NUMTs contained retrotransposons in their flanks (P-value = 0.001). The retrotransposon-encoded endonuclease recognizes AT-rich oligomers (5'-TTTTAA-3'). This suggests retrotransposon-encoded endonuclease may be involved in NUMT insertion. An understanding of mtDNA-transferred and nuclear NUMT-insertion features would help to understand genomic evolution and diversification among living organisms, and also contribute to identify the positions prone to produce DSBs (perhaps also in somatic cells) which are deleterious for organisms. Furthermore, by crosschecking our NUMT dataset against annotation databases, we observed new functional NUMTs candidates in human genome; three non-coding transcripts, and one secreted protein, R-spondin homolog 1 (RSPO1). Applying phylogenetic analysis, we found that these functional NUMTs were inserted; after humans and gorillas diverged and three of them after humans and chimpanzees diverged. Interestingly, the human specific functional NUMTs were non-coding RNAs which are expressed during fetal brain development. Hence, it is conceivable that this element might contribute to the difference between human and chimpanzee brain structure.

Keywords: Nuclear genome, mitochondrial DNA, NUMT, mitochondrial pseudogene, retrotransposon, exogenous DNA element