Mapping SardiNIA homoplasy mutations to predicted g-quadruplex-forming sequences

André Rush, Jr.
Mentored by Robert Brosh and Sanjay Bharti

Introduction

G-quadruplexes (G4) are inhibitory, guanine-rich DNA sequences that could lead to many medical complications. These mitochondrial DNA G-quadruplexes are likely to pose replication barriers that represent hot-spots for mutation. Complex traits are influenced by both environmental and genetic factors. The SardiNIA Study was initiated by NIA and carried out to identify the genetic basis of aging-related phenotypes in a founder population. In previous work, mtDNA variation and copy number in lymphocytes of ~2,000 Sardinians were assessed using a novel algorithm that identified mtDNA homoplasy and heteroplasy variants (Ding et al., 2015). Sardinians tend to live longer than average, so they were useful in studying phenotypes as people age. These studies demonstrated an increase in heteroplasy with age, a relatively high heritability of mitochondrial copy number, and its association with metabolic traits. The goal of the current study was to determine the extent with which mutations map to predicted mtDNA G4-forming sequences (Bharti et al., 2013). Not all mutations are negative, but those in g-quadruplex regions were hypothesized to have deleterious effects.

Methods and Materials

The Quadruplex forming G-Rich Sequences (QGRS) predictor algorithm was used to assess whether G4-predicted (G4P) sequences reside in close proximity to known mitochondrial DNA deletion breakpoints associated with genetic disorders, cancers, and aging. In excel, the base pairs, were first manually organized into mutated and non-mutated DNA. To make this more efficient, the R programming language was used to automatically organize the base pairs. After that, the mutations were organized into transitions and transversions. Furthermore, the mutations were separated into G4 and non-G4 sequences. This was to determine whether or not these are truly deleterious mutations or if they are common mutations. So whether or not a mutation was more frequent, it could determine where they are enriched. The reverse complement of all of the sequences were used in the pattern Finder program. The G4 sequences were put together in substrates that were then characterized by their mutation frequencies. The mutations that had frequencies such as 2000 were largely disregarded because they were indicated to be founder mutations. If the whole sample size had the mutations, then this meant that this was one of the genes that led to the population in the first place. Finally, these mutations were categorized into those that do and do not form in G4 regions.

Results

Two-hundred-ninety-seven (35%) base pairs were mapped to replication barriers that represent hot-spots for mutation. Complex traits are influenced by both environmental and genetic factors. The SardiNIA Study was initiated by NIA and carried out to identify the genetic basis of aging-related phenotypes in a founder population. In previous work, mtDNA variation and copy number in lymphocytes of ~2,000 Sardinians were assessed using a novel algorithm that identified mtDNA homoplasy and heteroplasy variants (Ding et al., 2015). Sardinians tend to live longer than average, so they were useful in studying phenotypes as people age. These studies demonstrated an increase in heteroplasy with age, a relatively high heritability of mitochondrial copy number, and its association with metabolic traits. The goal of the current study was to determine the extent with which mutations map to predicted mtDNA G4-forming sequences (Bharti et al., 2013). Not all mutations are negative, but those in g-quadruplex regions were hypothesized to have deleterious effects.

Conclusion

Two-hundred-ninety-seven (35%) base pairs were mapped to predicted G4-forming sequences; 212 (71%) of the 297 were located in genes encoding proteins of the oxidative phosphorylation enzyme complexes, and 63 (21%) were in predicted G4 of the D-loop where mutations are known to frequently occur. The remaining 532 nucleotide positions scoring positively for mutation (65%), including those with peak mutation frequency (≥ 444 mutations), mapped to mtDNA sequences not predicted to form G4. Of the 80 predicted G4 motifs, 42 mtDNA G4 motifs displayed homoplasmy mutations participants. Two one-tailed binomial tests were used where α = 0.05 and it seemed that the non-G4-forming sequences caused more mutations overall, yet those mutations were much less unique than those in the G4-forming sequences. This was supported by a p-value of 0.0062 < 0.05. Also, the p-value for the mutations enriched in non-G4 regions was 0.012 < 0.05. Most mutations are enriched in non-G4 regions but the unique mutations are enriched in G4 regions. Unique mutations were defined as mutations conserved in less than 20% of the total mutations (811 nucleotides).

References