Pseudo-code for Tavtigian BRCA1 RNG domain mutation screen

Brett Milash

02/07/2017

Given the wild-type BRCA1 RNG domain DNA sequence, coordinates of the

region that will be synthesized, and the sub-region that will be mutated:

1. Generate the wild-type protein sequence encoded by the mutated

region.

2. Generate all mis-sense substitutions to the WT protein sequence that

are possible from single nucleotide substitutions to the WT

nucleotide sequence, and the corresponding DNA sequence that

encodes that mis-sense substitution.

3. Generate protein sequences that comprise an alanine scan, ie the

protein sequences that arise from an insertion of a single

alanine amino acid at each position in the WT protein sequence,

and the corresponding DNA sequences that code for each

protein sequence.

4. For each DNA / protein sequence pair generated in steps 2 and 3:

4a. Generate 5 to 8 variations on this sequence by adjusting

the wobble bases in codons to alter the nucleotide

sequences without altering the protein sequence.

4b. Apply the following tests to screen the nucleotide sequences

from 4a, to exclude the following:

- new splice donors / acceptors

- GGGG sequences

- poly-X of length 6 or greater for any nucleotide

- sequences with GC content > 0.75 or below 0.25.

- rare codon usage

- stable secondary structures

- exon exclusion signals

4c. Report the sequences that pass the filters in 4b in a the

proper format for ordering oligonucleotides.