# CNV Whole Exon Sequencing Simulator Pipeline

1. Generating copy number variation

This copy number variation simulator, written in python, is to introduce sequences of nucleotides ranging from couple of hundreds to couple of millions in base pairs into the reference genome.

Based on the ideal target regions (e.g., BED), the tool randomly selects a number of regions (exons) and generating copies of non-overlapping sequences.

The tool then randomized the ratio of amplifications and deletions based on the user's input. Within the amplifications, the tool creates a random number of copies for each amplified region and within the deletion, it simply deletes one or both copy of sequences from the reference genome.

The tool then outputs a list of modified regions and a pair of modified genomes containing CNVs

2. Processing CNVs:

The pair of modified genomes is then processed through batch script in Linux to generate whole exon sequencing reads and to secondary analysis.

3. Validation: