**Genomic Constraint**

Contact: siwei@broadinstitute.org

Last updated on Oct 27, 2022

The genomic constraint is quantified at a 1kb scale with a signed Z score by comparing the observed variation to an expectation. In brief, the expected number of variants in each tiling 1kb genomic region is predicted using an improved mutational model taking into account both local sequence context and a variety of genomic features. A higher positive Z score (i.e., observing fewer variants than expected) indicates higher constraint. To learn more details about the score construction and application, please view the [full preprint on bioRxiv](https://www.biorxiv.org/content/10.1101/2022.03.20.485034v2).

**File name: constraint\_z\_genome\_1kb.raw.download.txt.gz**

This file contains constraint Z scores of 2,561,057 tiling 1kb genomic windows (of which 1,984,900 passed all quality control [QC] checks – see col 10) on autosomes. Coordinates are in GRCh38 and are 0-based [start,end). Columns:

1. chrom: Chromosome number
2. start: Starting position of the 1kb window
3. end: Ending position of the 1kb window
4. element\_id: Unique window id (chrom-start-end)
5. possible: Number of possible variants in window
6. expected: Number of expected variants in window
7. observed: Number of observed variants in window
8. oe: Observed over expected ratio for variants in window (‘observed’ divided by ‘expected’)
9. z: Constraint Z score of the 1kb window if the window passed QC, otherwise score is NA
10. fail\_qc: Reason(s) for falling QC. Multiple filters are separated by comma. Four possible reasons:

* low\_variant\_quality = proportion of ‘PASS ‘variants <80%
* coverage\_outlier = mean coverage <25 or >35
* low\_possible\_variant = num of possible variants <1000
* z\_outlier = z score <-10 or >10

1. unfiltered\_z: Constraint Z score of the 1kb window, ignorant of QC

**File name: constraint\_z\_genome\_1kb.qc.download.txt.gz**

This file contains constraint Z scores of 1,984,900 tiling 1kb genomic windows (passing all quality control [QC] checks) on autosomes. Coordinates are in GRCh38 and are 0-based [start,end). Columns:

1. chrom: Chromosome number
2. start: Starting position of the 1kb window
3. end: Ending position of the 1kb window
4. element\_id: Unique window id (chrom-start-end)
5. possible: Number of possible variants in window
6. expected: Number of expected variants in window
7. observed: Number of observed variants in window
8. oe: Observed over expected ratio for variants in window (‘observed’ divided by ‘expected’)
9. z: Constraint Z score of the 1kb window

**File name: constraint\_z\_enh\_gene\_roadmaplinks.all.download.txt.gz**

This file contains constraint Z scores of enhancers that are linked to specific genes, across 21 tissue types (ref <https://ernstlab.biolchem.ucla.edu/roadmaplinking/>). For each enhancer, the Z score was computed by taking the highest Z score of all overlapping 1kb windows (as listed in file constraint\_z\_genome\_1kb\_qc.download.txt.gz). Coordinates are in GRCh38 and are 0-based [start,end). Columns:

1. tissue: Tissue name, or ‘Merged’, which indicates the enhancer regions linked to the same gene were merged across different tissue types
2. chrom: Chromosome number
3. start: Starting position of the enhancer
4. end: Ending position of the enhancer
5. enhancer\_id: Unique id of the enhancer (chrom-start-end)
6. gene: Gene name, to which the enhancer is linked
7. ENSG: Ensembl gene ID (Gencode v32)
8. z: Constraint Z score of the enhancer