

Assembly V4.0 :

This is our most recent assembly of 14 million high quality sequences. Approximately 89.9% of these were assembled into contigs with a total length of 542 Mb. These contigs were organized into 921,990 isogroups, representing a total of 1,057,173 isotigs. We have complete (~7K) or incomplete (~10K) protein-coding sequence models for ~17K human refseq proteins. We have 3K additional significant blast hits to non-human protein coding models. With respect to the tissues we have sampled (brain, limb, blood, etc) , we believe we have significant hits to >95% of the transcriptome.

Assembly V3.0:

Approximately 74% of the 3.3 million high quality sequences were assembled into contigs with a total length of 71 Mb. The final unique set of contigs and singletons from the assembly was about 915,442 sequences. We have complete (~3K) or incomplete (~12K) protein-coding sequence models for ~15K human refseq proteins.