<https://hgdownload.cse.ucsc.edu/goldenpath/hg38/bigZips/analysisSet/>

The "analysis set" is a version of the genome prepared for next-gen

sequencing read alignment. It contains no alternate sequences, no

patches, fixes or haplotypes, only the main chromosomes. For more

information, see https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/

The files here come from NCBI, and were converted into UCSC formats.

The files below are the original 2014 release from NCBI, without decoys

or alt-aware BWA files.

For a full description of the "analysis set" concept, see NCBI's README file:

https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA\_000001405.15\_GRCh38/seqs\_for\_alignment\_pipelines.ucsc\_ids/

These files are no longer updated. For the latest analysis

set see the FTP NCBI directory at the address above.

Files included in this directory:

hg38.analysisSet.2bit - analysis set sequence

hg38.analysisSet.fa.gz - analysis set sequence

hg38.analysisSet.chroms.tar.gz - analysis set sequence one file per chromosome

The analysis set sequence is masked as mentioned in ../README.txt,

repeats from RepeatMasker and Tandem Repeats Finder

(with period of 12 or less) are shown in lower case;

non-repeating sequence is shown in upper case.

The sequences in the file are otherwise identical to the NCBI file

GCA\_000001405.15\_GRCh38\_no\_alt\_analysis\_set.fna.gz

hg38.fullAnalysisSet.2bit - all of the sequence from the above set,

plus all of the alt-scaffolds from the GRCh38 ALT\_REF\_LOCI\_\* assembly units.

hg38.fullAnalysisSet.chroms.tar.gz - all of the sequence from the above set,

plus all of the alt-scaffolds from the GRCh38 ALT\_REF\_LOCI\_\* assembly units.

The analysis set sequence is masked as mentioned in ../README.txt,

repeats from RepeatMasker and Tandem Repeats Finder

(with period of 12 or less) are shown in lower case;

non-repeating sequence is shown in upper case.

md5sum.txt - checksums of files in this directory

[Index of /genomes/all/GCA/000/001/405/GCA\_000001405.15\_GRCh38/seqs\_for\_alignment\_pipelines.ucsc\_ids (nih.gov)](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA_000001405.15_GRCh38/seqs_for_alignment_pipelines.ucsc_ids/)

