Resource Summary Report

Generated by NIF on May 8, 2025

HS3D - Homo Sapiens Splice Sites Dataset

RRID:SCR_002939
Type: Tool

Proper Citation

HS3D - Homo Sapiens Splice Sites Dataset (RRID:SCR_002939)

Resource Information

URL: http://www.sci.unisannio.it/docenti/rampone/

Proper Citation: HS3D - Homo Sapiens Splice Sites Dataset (RRID:SCR_002939)

Description: Data set of Homo Sapiens Exons, Introns and Splice regions extracted from GenBank Rel.123 with an aim of giving standardized material to train and to assess the prediction accuracy of computational approaches for gene identification and characterization. From the complete GenBank (Primate Sequences Division) Rel.123 (162,557 entries), entries of Human Nuclear DNA including Complete CDS and more than one Exon have been selected, and 4523 exons and 3802 introns have been extracted from these entries. Details about extracted exons and introns are reported (Locus, number, Start and End position in the entry, sequence, length, G+C content, presence of not AGCT data (nucleotide scan check)). Statistics are also reported (overall nucleotides, average G+C content, nucleotide scan check results, number of not GT starting / AG ending introns, minimum / maximum / average length, length standard deviation). 3799+3799 donor and acceptor sites, as windows of 140 nucleotides around each splice site have been extracted. After discarding sequences not including canonical GTAG junctions (65+74), including insufficient data (not enough material for a 140 nucleotide window) (686+589), including not AGCT bases (29+30), and redundant (218+226) there are 2796+ 2880 windows. Finally, there are 271,937 + 332,296 windows of false splice sites, selected by searching canonical GTAG pairs in not splicing positions. The false sites in a range of +/- 60 from a true splice site are marked as proximal.

Abbreviations: HS3D

Synonyms: Homo Sapiens Splice Sites Dataset, HS3D (Homo Sapiens Splice Sites Dataset)

Resource Type: data or information resource, data set

Keywords: human genome, splice, exon, intron, region, gene, dna, nucleotide, splice region

Funding:

Resource Name: HS3D - Homo Sapiens Splice Sites Dataset

Resource ID: SCR_002939

Alternate IDs: nif-0000-02988

Record Creation Time: 20220129T080216+0000

Record Last Update: 20250507T060107+0000

Ratings and Alerts

No rating or validation information has been found for HS3D - Homo Sapiens Splice Sites Dataset.

No alerts have been found for HS3D - Homo Sapiens Splice Sites Dataset.

Data and Source Information

Source: SciCrunch Registry

Usage and Citation Metrics

We found 7 mentions in open access literature.

Listed below are recent publications. The full list is available at NIF.

Zeng Y, et al. (2019) A high-performance approach for predicting donor splice sites based on short window size and imbalanced large samples. Biology direct, 14(1), 6.

Meher PK, et al. (2019) Evaluating the performance of sequence encoding schemes and machine learning methods for splice sites recognition. Gene, 705, 113.

Zhang Y, et al. (2018) Discerning novel splice junctions derived from RNA-seq alignment: a deep learning approach. BMC genomics, 19(1), 971.

Meher PK, et al. (2016) Identification of donor splice sites using support vector machine: a computational approach based on positional, compositional and dependency features. Algorithms for molecular biology : AMB, 11, 16.

Meher PK, et al. (2016) Prediction of donor splice sites using random forest with a new sequence encoding approach. BioData mining, 9, 4.

Chen W, et al. (2014) iSS-PseDNC: identifying splicing sites using pseudo dinucleotide composition. BioMed research international, 2014, 623149.

Galperin MY, et al. (2005) The Molecular Biology Database Collection: 2005 update. Nucleic acids research, 33(Database issue), D5.