جمع‌آوری داده

استفاده از داده‌های COSMIC (https://cancer.sanger.ac.uk/cosmic)

توضیحات فایل:

[column number:label] Heading

[1:A] Gene name - The gene name for which the data has been curated in COSMIC. In most cases this is the accepted HGNC identifier.

- اسم ژن – با آزمایش‌های COSMIC معتبر هستند.

https://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=BRAF

[2:B] Accession Number - The transcript identifier of the gene.

- کد شناسایی ژن

[3:C] Gene CDS length - Length of the gene (base pair) counts.

- تعداد نوکلئوتیدهای ژن که پروتین کد می‌کند.

[4:D] HGNC id - if gene is in HGNC, this id helps linking it to HGNC.

- در صورت داشتن کد شناسایی تایید شده HUGO Gene Nomenclature Committee

[5:E] Sample name, Sample id, Id tumour - A sample is an instance of a portion of a tumour being examined for mutations. The sample name can be derived from a number of sources. In many cases it originates from the cell line name. Other sources include names assigned by the annotators, or an incremented number assigned during an anonymisation process. A number of samples can be taken from a single tumour and a number of tumours can be obtained from one individual. A sample id is used to identify a sample within the COSMIC database. There can be multiple ids, if the same sample has been entered into the database multiple times from different papers.

- نمونه سرطانی که برای جهش‌هایش بررسی می‌شود. اسم نمونه می‌تواند براساس اسم رده سلولی باشد تا اسم فرد به طور مستقیم درج نشود.

[8:H] Primary Site - The primary tissue/cancer from which the sample originated. More details on the tissue classification are avaliable from here. In COSMIC we have standard classification system for tissue types and sub types because they vary a lot between different papers.

[9:I] Site Subtype 1 - Further sub classification (level 1) of the samples tissue of origin.

[10:J] Site Subtype 2 - Further sub classification (level 2) of the samples tissue of origin.

[11:K] Site Subtype 3 - Further sub classification (level 3) of the samples tissue of origin.

[12:L] Primary Histology - The histological classification of the sample.

[13:M] Histology Subtype 1 - Further histological classification (level 1) of the sample.

[14:N] Histology Subtype 2 - Further histological classification (level 2) of the sample.

[15:O] Histology Subtype 3 - Further histological classification (level 3) of the sample.

- دسته‌بندی بافت و زیربافت توسط COSMIC انجام شده است.

[16:P] Genome-wide screen - if the entire genome/exome is sequenced.

- آیا کل ژنوم توالی‌یابی شده است؟

[17:Q] GENOMIC_MUTATION_ID - Genomic mutation identifier (COSV) to indicate the definitive position of the variant on the genome. This identifier is trackable and stable between different versions of the release.

- کد شناسایی مختصات جهش که در بین ژنوم‌های مرجع (GRCh37 و GRCh38) یکسان است.

  اگر مختصات ژنوم جهش مشخص نباشد مقدار null می‌گیرد.

[18:R] LEGACY_MUTATION_ID - Legacy mutation identifier (COSM) that will represent existing COSM mutation identifiers.

- کد شناسایی مختصات جهش که در بین ژنوم‌های مرجع (GRCh37 و GRCh38) یکسان است.

  ممکن است مختصات نامعلوم ژنوم جهش را شناسایی کند.

[19:S] MUTATION_ID - An internal mutation identifier to uniquely represent each mutation on a specific transcript on a given assembly build.

- کد شناسایی جهش در یک رونویسی مشخص از ژنوم مرجع مشخص

[20:T] Mutation CDS - The change that has occurred in the nucleotide sequence. Formatting is identical to the method used for the peptide sequence.

- کد شناسایی منطقه کدکننده رشته DNA که در آن جهش رخ داده است. این جهش می‌تواند باعث توقف تولید پروتئین (Nonsense)، تغییر آمینواسید (Missense)، بدون تغییر امینواسید (Coding silent) شود و یا خارج از محدوده کدکننده (Intronic) باشد.

[21:U] Mutation AA - The change that has occurred in the peptide sequence. Formatting is based on the recommendations made by the Human Genome Variation Society.

- کد شناسایی منطقه کدکننده پروتین در رشته آمینواسیدها که در آن جهش رخ داده است.

[22:V] Mutation Description - Type of mutation at the amino acid level (substitution, deletion, insertion, complex, fusion, unknown etc.)

- نوع جهش (جایگزینی، حذف، اضافه، ترکیبی از حذف و اضافه و جایگزینی، و جابه‌جایی کروموزومی یا حذف کروموزومی یا وارونگی کروموزومی که باعث همجوشی دو ژن مجزا می‌شود)

[23:W] Mutation zygosity - Information on whether the mutation was reported to be homozygous , heterozygous or unknown within the sample.

- یک سلول را homozygous گویند اگر برای یک ژن خاص محتوای ژنتیکی در کروموزوم‌های جفتی یکسان باشد. یک سلول را heterozygous گویند اگر برای یک ژن خاص محتوای ژنتیکی در کروموزوم‌های جفتی متفاوت باشد.

[24:X] LOH - LOH Information on whether the gene was reported to have loss of heterozygosity in the sample: yes, no or unknown.

- پدیده Loss of heterozygosity که در اثر cross در کروموزوم رخ داده است و باعث از دست رفتن کامل ژن می‌شود. اگر عملکرد این ژن در منطقه از دست رفته سرکوب کردن تومور(tumor suppressor gene) باشد، فرد دچار سرطان می‌شود.

[25:Y] GRCh - The coordinate system used -

37 = GRCh37/Hg19

38 = GRCh38/Hg38

- از چه ژنوم مرجعی برای تعیین مختصات جهش استفاده شده است.

[26:Z] Mutation genome position - The genomic coordinates of the mutation.

- مختصات جهش با درج شماره کروموزوم و موقعیت شروع و پایان جهش

[27:AA] Mutation strand - postive or negative.

- رشته مثبت، رشته‌ای که از آن رونویسی صورت می‌گیرد و رشته منفی، رشته مکمل آن است.

[28:AB] SNP - All the known SNPs are flagged as 'y' defined by the 1000 genomes project, dbSNP and a panel of 378 normal (non-cancer) samples from Sanger CGP sequencing.

- آیا جهش در منطقه SNP رخ داده است؟

[29:AC] Resistance Mutation - mutation confers drug resistance

- آیا جهش باعث مقاومت در برابر دارو شده است؟

[30:AD] FATHMM prediction - More information about FATHMM (Functional Analysis through Hidden Markov Models) is available from here. FATHMM descriptors -

Pathogenic = Defined as Cancer or Damaging.

Neutral = Defined as Passenger or Tolerated.

- پیشگویی نتیجه عملکردی جهش براساس اینکه در منطقه کدکننده یا غیر کد کننده باشد.

[31:AE] FATHMM Score - The scores are in the form of pvalues ranging from 0 to 1. Pvalues greater than 0.5 are pathogenic while less than 0.5 are benign. Pvalues close to 0 or 1 are the high confidence results which are more accurate. The results are annotated as 10 feature groups (separately for coding and non coding variants) details of which can be found in the original FATHMM-MKL paper.

- هرچه قدر امتیاز FATHMM به ۰ یا ۱ نزدیک باشد دقیق‌تر است. امتیاز ۰ به معنای بی‌خطر بودن نتیجه جهش و امتیاز ۱ به معنای سرطانی بودن نتیجه جهش است.

[32:AF] Mutation somatic status - Information on whether the sample was reported to be Confirmed Somatic, Previously Reported or Variant of unknown origin -

Confirmed Somatic = if the mutation has been confimed to be somatic in the experiment by sequencing both the tumour and a matched normal from the same patient.

Previously observed = when the mutation has been reported as somatic previously but not in current paper.

Variant of unknown origin = when the mutation is known to be somatic but the tumour was sequenced without a matched normal.

- تایید جسمی بودن جهش براساس اینکه توالی‌یابی بافت سرطانی با توالی‌یابی بافت سالم همان فرد مطابقت داده شده یا خیر.

[33:AG] Pubmed_PMID - The PUBMED ID for the paper that the sample was noted in, linking to pubmed to provide more details of the publication.

- اطلاعات بیشتر در مورد نمونه

[34:AH] Id Study - Lists the unique Ids of studies that have involved this sample.

[35:AI] Sample Type,Tumour origin - Describes where the sample has originated from including the tumour type.

- روش دسترسی به نمونه سرطانی (surgery – NOS, xenograft, surgery-fixed)
- در چه مرحله‌ای از سرطان، نمونه برداشته شده است. (primary, metastasis)

[37:AK] Age - Age of the sample (if this information is provided with the publications).