# EXPANSION: A Webserver to Explore the Functional Consequences of Protein-Coding Alternative Splice Variants in Cancer Genomics

Chakit Arora<sup>\*†1</sup>,Natalia De Oliveira Rosa<sup>\*†1</sup>, Marin Matic<sup>1</sup>, Mariastella Cascone<sup>1</sup>, Pasquale Miglionico<sup>1</sup>, Francesco Raimondi<sup>†1</sup>

Laboratorio di Biologia Bio@SNS, Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126, Pisa, Italy

\*Equally contributing

<sup>†</sup>Corresponding author, Email: <u>francesco.raimondi@sns.it</u>, <u>chakit.arora@sns.it</u>, <u>natalia.deoliveirarosa@sns.it</u>

Lead Contact: francesco.raimondi@sns.it.

#### Abstract

EXPANSION (https://expansion.bioinfolab.sns.it/) is an integrated web-server to explore the functional consequences of protein-coding alternative splice (AS) variants. We combined information from Differentially Expressed (DE) protein-coding transcripts from cancer genomics, together with domain architecture, protein interaction network, and gene enrichment analysis to provide an easy-to-interpret view of the effects of protein-coding splice variants. We retrieved all the protein-coding Ensembl transcripts and mapped Interpro domains and Post-translational modifications (PTMs) on canonical sequences to identify functionally relevant splicing events. We also retrieved isoform-specific protein-protein interactions (PPIs) and binding regions from IntAct to uncover isoform-specific functions via

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/40/) which permits unrestricted reuse distribution and reproduct

(http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

<sup>©</sup> The Author(s) 2023. Published by Oxford University Press.

gene-sets over-representation analysis. Through EXPANSION, users can analyse precalculated or user-inputted DE transcript datasets, to easily gain functional insights on any protein spliceform of interest.

### **Key Words**

Protein domains, protein interaction networks, alternative splicing, functional genomics

#### Introduction

Alternative splicing is a widespread and complex process in eukaryotic cells that plays a critical role in regulating expression gene and generating protein diversity(Nilsen and Graveley, 2010). The process is regulated by various factors, including cis-acting elements, trans-acting factors, and epigenetic modifications, which work together to determine which exons are included or excluded mature mRNA(Black, from 2003). Dysregulation of alternative splicing

implicated the has been in pathogenesis of various diseases. including cancer, as it can lead to aberrant protein isoforms that may have oncogenic properties or result in the loss of tumor suppressor activity(Kahles et al., 2018), for e.g. roles of alternative splicing of the CD44 and BCL2L1 genes in various types of cancer(Jiang et al., 2014; Brown et al., 2011). RNA sequencing (RNA-seq) technology has greatly enhanced ability to study our

1

59 60 alternative splicing and its functional consequences(Merkin et al., 2012; Wang et al., 2015). Transcriptomics data can be used identifv to differentially expressed transcripts, including alternative splicing isoforms, and investigate the functional implications of alternative splice variants, particularly in the context of cancer.Integration of transcriptomics data with protein functional annotations can help investigate the impact of alternative splicing.

Several methods and associated webservers have integrated the information of protein structures, **PTMs** domains. and interaction networks to explore the consequences of genetic variants (e.g. Mechismo(Betts et al., 2015; Raimondi et al., 2019), Mechnetor(González-Sánchez et al., 2021), DysSysNet(Mosca et al., 2015)) Similar efforts leveraged the information of protein domains and interaction network to predict the functional consequences of aberrant splicing events in cancer, suggesting a potential role as oncodrivers for some of them(Climente-González et al., 2017; Kahraman et al., 2020). A few webservers have also been recently proposed. DIGGER(Louadi, Yuan, et al., 2021) integrates protein-protein interactions (PPI), domainand residue-level interactions information to interpret exon expression to a CanlsoNet(Lay network level.

Downloaded from https://academic.oup.com/bioinformaticsadvances/advance-article/doi/10.1093/bioadv/vbad135/7282943 by Scuola Normale Superiore user on 28 September 2023

60

1

Karakulak et al., 2023) allows to analyse isoform switching events in multiple diseases, by integrating isoform expression data with interaction networks from STRING and ClinVar annotations. Similarly, NEASE detects protein domains affected by alternative splicing and then uses PPIs domain-domain and domain-motif interactions identify interaction to partners likely affected(Louadi, Elkjaer, 2021). Additional et al., online resources include ASCancer Atlas(Wu et al.. 2023), а comprehensive knowledgebase of aberrant splicing in human cancers with highly curated cancer-associated splicing events. which experimentally have been shown to promote tumorigenesis, and their regulatory networks.

webserver to explore the functional implications of alternative splice variants, which integrates alternative transcript differential expression, domain architecture and Posttranslational **Modifications** (PTMs) annotations, isoform-specific protein interaction networks and gene pipeline enrichment The analysis. provides with researchers а comprehensive both toolkit. webbased and API based, to explore the functional effects of alternative splicing events in the context of cancer, as well as in any other condition of interest by allowing the upload and analysis of user-specified DE input dataset.

Here we present EXPANSION, a new

# Methods

Retrieval of protein-coding transcripts and clustering:

A list comprising 19,232 protein-coding genes was acquired from HGNC. This list was employed to execute queries

3
4
5
6
7
8
0
9
10
11
12
13
1/
14
15
16
17
18
19
20
∠∪ ว1
21
22
23
24
25
25
20
27
28
29
30
21
27
32
33
34
35
36
27
27
38
39
40
41
42
43
11
44
45
46
47
48
49
50
50
21
52
53
54
55
56
50
5/
58
59
60

on the EnsembIDB database, utilizing the 'EnsDb.Hsapiens.v86' package within the R environment to retrieve transcript related annotations(Rainer et al., 2019). CD-HIT(Li et al., 2001) was utilized to execute sequence clustering based on similarity. Employing a threshold of 0.6, we successfully grouped closely related protein sequences that likely fulfill analogous functional roles. The clustering was crucial to avoid the alignment of sequences of substantially different proteins originated from the same genetic locus. e.g., Guanine nucleotide-binding protein G(s) subunit alpha. Neuroendocrine secretory protein 55 and ALEX, all coded in the GNAS locus. Since performing MSA on such sequences would lead to variation linked to different protein identity, and not to splicing, we had first to group sequences via clustering, so as to produce MSAs instrumental in detecting alternative splicing variation. We named each cluster based on gene symbol and the Uniprot Accession of the longest, canonical sequence in that cluster.

Multiple sequence alignments (MSAs) and detection of splicing events:

For the sequences in each cluster, wWe generated multiple sequence alignments (MSAs) by using ClustalOmega(Sievers et al., 2011) with default parameters. The MSAs were exploited to detect either inserted, deleted or divergent regions in the alternative isoforms with respect to the canonical Uniprot sequence. То identify any splicing events that could affect structural and functional sites, mapped Interpro domain we definitions(Blum et al., 2021; Hornbeck et al., 2015) and PTMs obtained from PhosphoSitePlus(Hornbeck et al., 2015) onto Uniprot canonical sequences.

Downloaded from https://academic.oup.com/bioinformaticsadvances/advance-article/doi/10.1093/bioadv/vbad135/7282943 by Scuola Normale Superiore user on 28 September 2023

1

Protein isoform-specific interactions and overrepresentation analysis (ORA):

We retrieved protein isoform specific interactions from IntAct (version 1.38.0, R4.2)(del Toro *et al.*, 2022) including. "direct". "physical" and "association" interactions. For each interaction. we retrieved the information of binding regions, wherever available, and mapped it to splicing-variable regions obtained from MSAs, to identify interaction sites affected by alternative splicing.

We integrated gene set overrepresentation analysis (ORA) to identify specific biological processes affected by alternative splicing. The g:profiler-official (v1.0.0)(Raudvere et al., 2019) python package was used to perform the functional ORA of categories significantly enriched in either the whole, or isoform-specific PPI network.

# Integration with transcriptomics DE data:

We integrated such protein-centric analysis modality with transcriptomics data. bv usina RSEM transcript abundances from the TCGA and GTEX databases (16 tissue-types, 19109 samples) which were processed through the TOIL pipeline(Vivian et al., 2017), and which we retrieved from the UCSC Xena Browser(Goldman et al., 2020). We employed EBSeq, а Bayesian approach that provides more estimates of differential accurate expression compared to traditional methods(Leng et al., 2013), to identify differentially expressed transcript isoforms. Instances with a posterior probability differentially of being expressed (PPDE) greater than 0.95 were considered significant.

#### API

Through an application programming interface (API), we enable direct open

access to data. The Python programming language and the Fast API library (https://fastapi.tiangolo.com/lo/, v0.95.1) were used to construct the API. The API documentation website (https://expansion.bioinfolab.sns.it/api/) offers information about how to use the API and the supported methods. The EXPANSION API is intended for multigene queries of the information integrated in the EXPANSION webserver and allows for retrieval of either individual, or combined, analyses generated for the

#### Implementation

The web-server was developed using
Apache2 (https://https.apache.org/)
and Flask (v2.0.2)
(http://flask.pocoo.org/) for the web
frontend and for the internal pipeline to
handle back-end processes. We
additionally used the following Python
and JavaScript libraries at both back-
and front-ends: D3 (v4), jQuery
(v3.2.1), DataTables (v2.3.2),
biopython(v1.79), gprofiler-
official(v1.0.0), Mysql-connector-
python (v8.0.31). To construct the API,
Fast API library (v0.95.1) was used.

Downloaded from https://academic.oup.com/bioinformaticsadvances/advance-article/doi/10.1093/bioadv/vbad135/7282943 by Scuola Normale Superiore user on 28 September 2023

### Web-server usage

queried transcript sets.

The web server provides a graphical framework to analyze protein alternative splice-forms and their functional consequences in a genecentric manners, as outlined in **Figure 1A**. The 'Help' and 'About' pages contain detailed instructions, and users can query pre-calculated datasets (TCGA vs GTEX from Xena/TOIL) or upload their own differential expression (DE) transcript dataset. The format requirements for the

60

1

uploaded DE transcript dataset are also provided in the 'Help' page. The search bar on the landing page allows users to access all available isoforms for a given gene symbol or ENSEMBL transcript id, grouped into clusters identified by the gene symbol and, in parenthesis, the uniprot accession of the canonical sequence representing that cluster (**Figure 1B**)

After loading or querying the dataset, users can examine the differential expression of transcripts in each cancer tissue using the displayed bubble plot (Figure 1B). Each circle represents a transcript colored based on log-fold change and size proportional to DE significance (e.g. PPDE). The cartoon panel on the right displays splicing variation details. including insertions, deletions. variations. divergent and posttranslational modifications (PTMs), using colored boxes and lollipops placed along the protein sequence

architecture diagrams. The transcript structure can be explored interactively by double-clicking on the sequence diagram, revealing a pop-up window showcasing aligned transcript sequences from different isoforms, with a specific focus on the selected position. Furthermore. PTMs are highlighted using asterisks within the MSA viewer's pop-up window, offering an additional layer of insight to the visualization. Users can export both the bubble plot and the domain architecture diagram in SVG or PNG format.

The central panel displays interaction networks mediated by the protein isoforms based on interaction data from IntAct. Users can adjust the number of interactors using toggles for MI score, interaction type, and maximum number of isoform-specific interactors (**Figure 1C**). Each node represents a protein, and its domain architecture can be visualized by

60

clicking on the node to expand it. The node colors are representative of isoform specificity. Region-specific interaction edges are drawn if binding region information is provided, with corresponding edges highlighted in red if a splicing event affects a binding region.

Finally, users visualize overcan (ORA) representation analysis of functional categories for the genes in the network. User can specify whether to calculate ORA for either the entire, or isoform-specific networks, allowing of distinct biological comparison mediated by isoformprocesses specific interactors (Figure 1C). The results of the enrichment analysis are presented in a table that includes the genes linked to each significant term. The analysis excludes Uniprot accessions associated with alternative isoforms lacking distinct interactors.

**PRKACB** isoforms as a case study

PRKACB, a protein kinase A catalytic subunit, is characterized by alternative splicing isoforms, which are expressed in a tissue specific fashion and are associated to a number of disorders, including cancer (Ramms *et al.*, 2021; Taylor *et al.*, 2021).

We analyzed them via 'EXPANSION', including a set of sixteen reported PRKACB isoforms, nine of which show significant differential expression in cancer tissues. These isoforms, namely 'ENST00000370680', 'ENST00000417530',

'ENST00000436133',

'ENST00000614872',

'ENST0000370688' and 'ENST00000610457', are exclusively upregulated different in tissues; transcripts 'ENST00000450730' and 'ENST00000370682' are both up and down-regulated based on tissue types: while 'ENST0000394838' is exclusively down-regulated in Stomach cancer (Figure 2A). Notably, 5 of these

1

isoforms, i.e., 'ENST00000370680', 'ENST00000417530'. 'ENST00000450730', 'ENST00000436133', and 'ENST00000370688', exhibit Cterminus deletions affecting the 'Protein kinase-like domain superfamily' (Figure 2A). These deletions may lead to functional impairments, potentially disrupting protein activity and interactions.

The network of PRKACB-isoform specific interactions divulges which interactions (or interactors) are most likely to be affected in terms of splicevariants (Figure 2B). Interestingly, among the 9 different DE isoforms, three ('ENST0000417530', 'ENST00000450730', and 'ENST00000436133') display altered binding regions for the interactor 'PITX2 (Q99697)'. PITX2 is a wellknown transcription factor that plays a crucial role in embryonic development and tissue differentiation, particularly in

the heart, eye, and pituitary gland (Chinchilla et al., 2011). Aberrant expression and functional alterations of PITX2 have been implicated in various types of cancer. In certain contexts, PITX2 acts as a tumor suppressor by regulating cell cycle progression, apoptosis, DNA and repair mechanisms. Conversely, in other instances, PITX2 can promote tumor growth and metastasis by enhancing cell survival, angiogenesis, and invasion (Fung et al., 2012). Additionally, PITX2 has been associated with cancer stemness and drug resistance, further emphasizing its impact on tumor behavior (Zhang et al., 2012). The modified PRKACB-PITX2 interaction observed in these isoforms suggests a potential impact on downstream signaling pathways, gene expression, and cellular processes regulated by both proteins (Cox et al., 2002).

Furthermore, we discovered that the splicing-affected binding regions (210-218) within these three isoforms also exhibit otherwise missina an ubiguitination at position '214' (Figure 2A). This aberrant ubiquitination event is unique to these isoforms and may contribute to the altered dynamics of protein-protein interactions. The alteration of ubiquitination at this specific site could have implications for protein stability, degradation, or other regulatory mechanisms. Incidentally, these three isoforms lack interactor information in Intact, which results in no ORA enrichment results. However, for those isoforms for which this information is available for, e.g., Discussion

P22694-1 P22694-2, distinct and enrichment terms are found which might hint towards specific biological roles (Figure 2C). In particular, PRKACB canonical isoform (P22694-1) is associated to cellular components such as "cAMP-dependent protein kinase complex" (FDR=2.7E-5) and pathways such as "CREB1 phosphorylation through the activation of Adenylate Cyclase" (FDR=3E-4), while the alternative isoform P22694-2 cellular compartments such as "Cyliary base" (FDR=1E<sup>-10</sup>) and processes such as "Lipid metabolism pathway" (FDR=1E<sup>-10</sup>).

Downloaded from https://academic.oup.com/bioinformaticsadvances/advance-article/doi/10.1093/bioadv/vbad135/7282943 by Scuola Normale Superiore user on 28 September 2023

Overall, EXPANSION web server takes gene-centric queries, and provides a multimodal annotation of alternative splicing isoforms to help predicting the functional impact of such variation. This analysis encompasses the identification of variants with respect to aligned sequences, domain architecture, PTMs, and interaction networks mediated by the protein isoform considered. Furthermore, overrepresentation analysis of functional categories is performed for each isoform-

60

1

distinct specific group to compare biological processes mediated by isoformspecific interactors. These findings are then mapped onto the DE transcripts across cancers to obtain an integrated view of the functional consequences of differentially expressed spliceforms, whilst giving the possibility to process with the same pipeline additional DE transcript datasets We also provide access to our webserver through an API interface, which allows for multigene queries.

EXPANSION currently includes only human isoforms and a possible area of improvement will be considering protein isoforms from additional model organisms. It would also be interesting to integrate experimental, isoformspecific interactions from IntAct with computationally predicted ones. This will be particularly useful for protein

## Data Availability

isoforms that are currently not reported in IntAct. It would also be interesting to integrate in the webserver methods to predict the effect of splicing to PPIs (e.g. based on protein language models(Brandes *et al.*, 2022) or graph neural networks).

While some of functionalities presented are in part included in other resources (see the Introduction for an overview), others, such as IntAct isoform specific networks are not. We therefore believe that our web-toolkit will offer researchers a comprehensive exploration protein alternative of splice-forms and their potential functional consequences which is not fully available elsewhere, providing valuable insights for the study of these complex phenomena.

#### https://mc.manuscriptcentral.com/bioadv

EXPANSION is freely available at <a href="http://expansion.bioinfolab.sns.it/">http://expansion.bioinfolab.sns.it/</a>. The code of the scripts used for

available

is

https://github.com/raimondilab/expansi

Datasets associated to this resource are available at the following URL:

https://doi.org/10.5281/zenodo.822912

0.

at:

**EXPASION** 

on

## **Figures**

Figure 1:A) Workflow and web-server functionality. The figure showcases the step-by-step process and various tools available to analyze protein alternative splice-forms and their functional consequences, enabling users to query pre-calculated datasets, upload their own data, visualize transcript expression, explore splicing variations, examine protein interactions, and perform over-representation analysis of functional categories; B) (snapshot of the first half of the main landing page when a user submits or uploads a query. Main features include (i) Search bar for accessing isoform clusters (ii) A color-coded summary of statistics which for various assessments (iii) Bubble plot visualizing differential expression of transcripts in each cancer tissue. (iv) Splicing variations and post-translational modifications displayed in cartoon panels, (v) a click to open pop-up window to visualize MSA and (vi) Download options; C) snapshot of the second half of the main landing page when a user submits or uploads a query displaying (i) Interactions mediated by protein isoform shown based on IntAct data. Users can adjust interactors using toggles such as MI score and type of interaction. Region-specific interaction edges highlighted if the splicing event affects the binding region and ii) Over-representation analysis (ORA) of functional categories for genes in the network. Comparison of biological processes mediated by isoform-specific interactors. Each of these is complemented with export options.

**Figure 2**: A) Differential Expression and Protein Domain Changes in PRKACB Splice Variants. An analysis of differential expression for 16 PRKACB isoforms highlights significant changes in 9 isoforms within cancer samples, as denoted by larger bubbles in the bubble plot on the left panel. Five isoforms (e.g., ENST00000370680, ENST00000417530, ENST00000450730, ENST00000436133, and ENST00000370688) display C-terminus deletions affecting the 'Protein kinase-like domain superfamily', potentially influencing protein function and interactions, as shown in the domain architecture diagram on the right panel. Three of these isoforms (ENST00000417530, ENST00000450730, and ENST00000436133) exhibit modified binding regions with the interactor PITX2 and are shown in highlighted boxes. B) Modified PRKACB Isoform Interactions with PITX2. The network of PRKACB isoform-specific interactions reveals alterations in interaction partners due to splicing variations. On the left, the full interaction network of PRKACB from Intact is shown using default settings. On the right the altered interactions due to binding regions are revealed when 'binding region' only is selected. Altered PRKACB-PITX2 interactions suggest potential downstream effects on signaling pathways and cellular processes governed by both proteins. C) Overall Enrichment Analysis (ORA) of PRKACB Isoform-Specific Interactions. Enrichment analysis uncovers distinct terms for isoforms with available interaction information, such as P22694-1 and P22694-2, offering insights into potential biological roles linked to specific PRKACB isoforms.

## Author contributions

C.A. performed the analysis, contributed the bioinformatic pipeline and wrote the manuscript; N.D.O.R contributed the bioinformatic pipeline and developed the webserver; M.M. contributed the

# Funding

1 2 3

4

5

6

7 8

9

10

11

12

13 14

15

16

17

18 19 20

21 22 23

24

25

26 27

28

29 30

31 32 33

34

35 36

37 38

39

45 46 47

48

49 50

51

52 53 54

55 56 57

58

59

60

This work was supported by the Italian Association for Cancer Research (AIRC) under My First AIRC Grant (MFAG) 2020 [grant number 24317] to F.R. We received additional support by the bioinformatic pipeline; M.C. performed the analysis; P.M. contributed to the bioinformatic pipeline; F.R. conceived the project, contributed the bioinformatic pipeline and wrote the manuscript.

the Italian Ministry of University and Research through the Department of excellence "Faculty of Sciences" of Scuola Normale Superiore (to F.R.).

# Acknowledgements

We gratefully acknowledge the CINECA award, in collaboration with AIRC, for the availability of high performance computing resources and support. We gratefully

# Data Availability

The code of the scripts used for EXPASION is available at: https://github.com/raimondilab/expansion acknowledge computational resources of the Center for High Performance Computing (CHPC) at Scuola Normale Superiore.

Datasets associated to this resource are available at the following URL: https://doi.org/10.5281/zenodo.8229120.

5

6 7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23 24

25

26

27

28 29

30

31

32

33

34

35 36

37

38

39

40 41

42

43

44

45

46 47

48

49

50

51 52

53

54

55

56

57 58

59 60

# References

- Betts,M.J. *et al.* (2015) Mechismo : predicting mechanistic impact of mutations and modifications on molecular interactions. *Nucleic Acids Res*.
- Black,D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*, **72**, 291–336.
- Blum,M. *et al.* (2021) The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res*, **49**, D344–D354.
- Brandes,N. *et al.* (2022) Genome-wide prediction of disease variants with a deep protein language model. *bioRxiv*, 2022.08.25.505311.
- Brown,R.L. *et al.* (2011) CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. *J Clin Invest*, **121**, 1064–1074.
- Climente-González,H. *et al.* (2017) The Functional Impact of Alternative Splicing in Cancer. *Cell Rep*, **20**, 2215–2226.
- Goldman,M.J. *et al.* (2020) Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol*, **38**, 675–678.
- González-Sánchez,J.C. *et al.* (2021) Mechnetor: a web server for exploring protein mechanism and the functional context of genetic variants. *Nucleic Acids Res*, **49**, W366–W374.
- Hornbeck,P. V. *et al.* (2015) PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. *Nucleic Acids Res*, **43**, D512–D520.
- Jiang,H. *et al.* (2014) Prognostic value of CD44 and CD44v6 expression in patients with non-small cell lung

cancer: meta-analysis. *Tumour Biol*, **35**, 7383–7389.

- Kahles,A. *et al.* (2018) Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. *Cancer Cell*, **34**, 211-224.e6.
- Kahraman,A. *et al.* (2020) Pathogenic impact of transcript isoform switching in 1,209 cancer samples covering 27 cancer types using an isoformspecific interaction network. *Scientific Reports 2020 10:1*, **10**, 1–15.
- Lay Karakulak,T. *et al.* (2023) CanlsoNet: a database to study the functional impact of isoform switching events in diseases. *Bioinformatics Advances*, **3**.
- Leng,N. *et al.* (2013) EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics*, **29**, 1035–1043.
- Li,W. *et al.* (2001) Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics*, **17**, 282–283.
- Louadi,Z., Yuan,K., *et al.* (2021) DIGGER: exploring the functional role of alternative splicing in protein interactions. *Nucleic Acids Res*, **49**, D309–D318.
- Louadi,Z., Elkjaer,M.L., *et al.* (2021) Functional enrichment of alternative splicing events with NEASE reveals insights into tissue identity and diseases. *Genome Biology 2021* 22:1, **22**, 1–22.
- Merkin,J. *et al.* (2012) Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science*, **338**, 1593–1599.
- Mosca,R. *et al.* (2015) dSysMap: exploring the edgetic role of disease mutations. *Nat Methods*, **12**, 167–8.

3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
20
21 22
~~ 23
∠_) 24
∠4 25
25
20
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Nilsen,T.W. and Graveley,B.R. (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature*, **463**, 457–463.

- Raimondi,F. *et al.* (2019) Rare, functional, somatic variants in gene families linked to cancer genes: GPCR signaling as a paradigm. *Oncogene*, **38**, 6491–6506.
- Rainer,J. *et al.* (2019) ensembldb: an R package to create and use Ensemblbased annotation resources. *Bioinformatics*, **35**, 3151–3153.
- Ramms,D.J. *et al.* (2021) Gαs–Protein Kinase A (PKA) Pathway Signalopathies: The Emerging Genetic Landscape and Therapeutic Potential of Human Diseases Driven by Aberrant Gαs-PKA Signaling. *Pharmacol Rev*, **73**, 155–197.
- Raudvere,U. *et al.* (2019) g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res*, **47**, W191–W198.
- Sievers,F. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, **7**, 539.
- Taylor,S.S. *et al.* (2021) PKA Cβ: A forgotten catalytic subunit of cAMPdependent protein kinase opens new windows for PKA signaling and disease pathologies. *Biochemical Journal*, **478**, 2101– 2119.
- del Toro,N. *et al.* (2022) The IntAct database: efficient access to finegrained molecular interaction data. *Nucleic Acids Res*, **50**, D648–D653.
- Vivian,J. *et al.* (2017) Toil enables reproducible, open source, big biomedical data analyses. *Nat Biotechnol*, **35**, 314–316.
- Wang,E.T. *et al.* (2015) Antagonistic regulation of mRNA expression and splicing by CELF and MBNL proteins. *Genome Res*, **25**, 858–871.

Wu,S. *et al.* (2023) ASCancer Atlas: a comprehensive knowledgebase of alternative splicing in human cancers. *Nucleic Acids Res*, **51**, D1196– D1204.



a) LFC	Transcript significantly expressed Binding regions affected 6 4 2 072-4-6-8-10													<ul> <li>Deletions</li> <li>Phosphorylation</li> <li>Insertions</li> <li>Ubiquitylation</li> <li>Divergent positions</li> <li>Other</li> </ul>			
1086420	72-4	-6 -8	8-10				Doi PTI	maiı Ms a	ns a affec	ffec cted	ted						Protein kinase-like domain superfamily (IPR01100)
P22694-1 ENST00000370689 *			$\bigcirc$	$\bigcirc$	$\bigcirc$		$\bigcirc$	$\bigcirc$	$\bigcirc$	$\bigcirc$		$\bigcirc$	$\bigcirc$		$\bigcirc$		
B1APG1 ENST00000417530	_		$\bigcirc$		) 🔵						)						
P22694-3 ENST0000610703			$\bigcirc$		$\bigcirc$	$\bigcirc$		$\bigcirc$		$\bigcirc$			$\bigcirc$		$\bigcirc$		1 88 176 264 351
A0A087WVC4[ENST00000610457					$\bigcirc$	$\bigcirc$	$\bigcirc$			$\bigcirc$		$\bigcirc$	$\bigcirc$		$\bigcirc$		1 88 176 264 351 
B1APG0 ENST00000450730			$\bigcirc$														1 88 176 264 351
B1APG2 ENST00000413538										$\bigcirc$							1 88 176 264 351 ••••••••••••••••••••••••••••••••••••
P22694-7 ENST00000370682			$\bigcirc$											$\bigcirc$			1 88 176 264 351
P22694-10 ENST00000394839															$\bigcirc$		1 88 176 264 351
B1APG3 ENST00000370680			$\bigcirc$					$\bigcirc$		$\bigcirc$				$\bigcirc$	$\bigcirc$	$\bigcirc$	1 88 176 264 351 ••••••••••••••••••••••••••••••••••••
P22694-8 ENST00000370688			$\bigcirc$				$\bigcirc$			$\bigcirc$		0				$\bigcirc$	1 88 176 264 351
B1APF9 ENST00000370684		)									)						1 88 176 264 351 ••••••••••••••••••••••••••••••••••••
P22694-6 ENST00000614872			$\bigcirc$			$\bigcirc$	$\bigcirc$						$\bigcirc$			$\bigcirc$	1 88 176 264 351 ፵፼፼፼ ፼፼፼፼ ፼ 
P22694-2 ENST00000370685		)														$\bigcirc$	1 88 176 264 351 
P22694-9 ENST00000446538			$\bigcirc$		$\bigcirc$					$\bigcirc$	$\bigcirc$	$\bigcirc$	$\bigcirc$		$\bigcirc$		1 88 176 264 351
B1APF7 ENST00000436133		$\bigcirc$	$\bigcirc$		$\bigcirc$					$\bigcirc$		)			$\bigcirc$		1 88 176 264 351 ••••••••••••••••••••••••••••••••••••
B1APF8 ENST00000432111			$\bigcirc$		$\bigcirc$		$\bigcirc$			$\bigcirc$							1 88 176 264 351 ••••••••••••••••••••••••••••••••••••
A0A0A0MS54 ENST00000394838			$\bigcirc$												$\bigcirc$		1 88 176 264 351
																	1 88 176 <u>264</u> 351
	UTERUS	- KIDNEY	BRAIN	- TESTIS	LIVER	- PROSTATE	LUNG	BLADDER	- BREAST	- OVARY	SKIN	- PANCREAS	- THYROID	COLON	- ESOPHAGUS	- STOMACH	
																	Binding Regions Only

Page 18 of 18

