

Functional enrichment of alternative splicing events with NEASE reveals insights into tissue identity and diseases

Zakaria Louadi^{1,2}, Maria L. Elkjaer^{3,4,5}, Melissa Klug^{1,6,7}, Chit T. Lio^{1,2}, Amit Fenn^{1,2}, Zsolt Illes^{3,4,5}, Dario Bongiovanni^{6,7,8}, Jan Baumbach^{2,9}, Tim Kacprowski^{10,11}, Markus List^{1,*}, Olga Tsoy^{2,*}

¹ Chair of Experimental Bioinformatics, Technical University of Munich, 85354 Freising, Germany

² Chair of Computational Systems Biology, University of Hamburg, Notkestrasse 9, 22607 Hamburg, Germany

³ Department of Neurology, Odense University Hospital, Odense, Denmark

⁴ Institute of Clinical Research, University of Southern Denmark, Odense, Denmark

⁵ Institute of Molecular Medicine, University of Southern Denmark, Odense, Denmark

⁶ Department of Internal Medicine I, School of Medicine, University hospital rechts der Isar, Technical University of Munich, Munich, Germany

⁷ German Center for Cardiovascular Research (DZHK), Partner Site Munich Heart Alliance, Munich, Germany

⁸ Department of Cardiovascular Medicine, Humanitas Clinical and Research Center IRCCS and Humanitas University, Rozzano, Milan, Italy

⁹ Institute of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5000 Odense, Denmark

¹⁰ Division Data Science in Biomedicine, Peter L. Reichertz Institute for Medical Informatics of Technische Universität Braunschweig and Hannover Medical School, Braunschweig, Germany

¹¹ Braunschweig Integrated Centre of Systems Biology (BRICS), TU Braunschweig, Braunschweig, Germany.

* joint last author

Abstract

Alternative splicing (AS) is an important aspect of gene regulation. Nevertheless, its role in molecular processes and pathobiology is far from understood. A roadblock is that tools for the functional analysis of AS-set events are lacking. To mitigate this, we developed NEASE, a tool integrating pathways with protein-protein and domain-domain interactions to functionally characterize AS events. We show in four application cases how NEASE can identify pathways contributing to tissue identity and cell type development, and how it highlights splicing-related biomarkers. With a unique view on AS, NEASE generates unique and meaningful biological insights complementary to classical pathways analysis.

Keywords: alternative splicing, differential splicing, functional enrichment, systems biology, protein-protein interactions, pathways, disease pathways

Background

Alternative splicing (AS) boosts transcript diversity in human cells [1] and thus contributes to tissue identity [2], cell development [3], and pathology in, e.g., cardiomyopathy [4], muscular dystrophy [5] or autoimmune diseases [6]. It is estimated that up to 30% of disease-associated genetic variants affect splicing [7]. RNA sequencing technologies (RNA-seq) allow the quantification of different types of AS events and detect splicing

abnormalities in disorders. However, RNA-seq utility is currently limited by our incomplete understanding of the functional role of specific exons or the transcripts they contribute to.

A major challenge in AS analysis is the functional interpretation of a set of events, including isoform switching events and differentially spliced exons. The usual approach is to perform gene set enrichment or overrepresentation analysis [8–10]. This approach treats all genes affected by AS equally, neglecting that some AS events may not be functionally relevant at the protein-level [11] or result from noise in the splicing machinery [12]. Furthermore, functional differences between protein isoforms remain uncertain in many cases. A promising strategy to identify relevant AS events is to focus on those that lead to meaningful changes in the protein structure. Recent studies have shown that AS has the potential to rewire protein-protein interactions by affecting the inclusion of domain families [13] and linear motifs [14] or by activating nonsense-mediated decay [15].

This motivated the creation of databases and tools that predict the consequences of individual AS events or isoform switches. IsoformSwitchAnalyzeR [16], tappAS [17], DoChaP [18] and Spada [19] support transcript-level (as opposed to exon-level) analysis to identify isoform switches and their impact on the translation and the resulting isoforms features, such as domains, motifs, and non-coding sites. Exon Ontology [20] and DIGGER [21] support exon-level analysis to identify exon skipping events and their possible impact on the protein structure and function. Spada and DIGGER further consider the impact of AS on protein-protein interactions.

Most existing tools allow investigating AS-driven changes in an explorative fashion but tools for systematic analysis of functional effects of AS are lacking. Exon Ontology performs statistical tests to identify enriched features within a set of skipped exons. One example are domain families affected by AS across proteins more frequently than expected. However, none of the existing tools offer a systems biology view to specifically highlight functional consequences of AS events.

To tackle these limitations, we developed the first tool for functional enrichment of AS events. NEASE (Network-based Enrichment method for AS Events) first detects protein domains affected by AS and then uses an integrated protein-protein interaction (PPI) and domain-domain interaction (DDI) network [21] to identify protein interaction partners likely affected by AS. Next, it employs an edge-level hypergeometric test for gene set overrepresentation analysis. This approach is new in the way genes are selected for the enrichment test. Rather than considering only differentially spliced or expressed genes, which is currently the most common strategy, NEASE uses network information to select genes that are likely affected in the interactome. This is also superior to a simple network enrichment analysis, as we consider only those edges for which an AS contribution seems relevant and for which false positive results are less likely. We evaluated NEASE using multiple datasets from both healthy and disease cohorts. We show that the NEASE approach complements gene-level enrichment, and even outperforms it in scenarios where gene-level enrichment fails to find relevant pathways. Moreover, NEASE generates unique and meaningful biological insights on the exact impact of AS. Furthermore, since the statistical approach is network-based, NEASE can prioritize (differentially) spliced genes and finds new disease biomarkers candidates in case of aberrant splicing. The NEASE Python package, freely available at <https://github.com/louadi/NEASE>, provides multiple functions for

a deeper analysis and visualization of affected protein domains, edges, and pathways (individually or as a set).

Results

Overview of NEASE

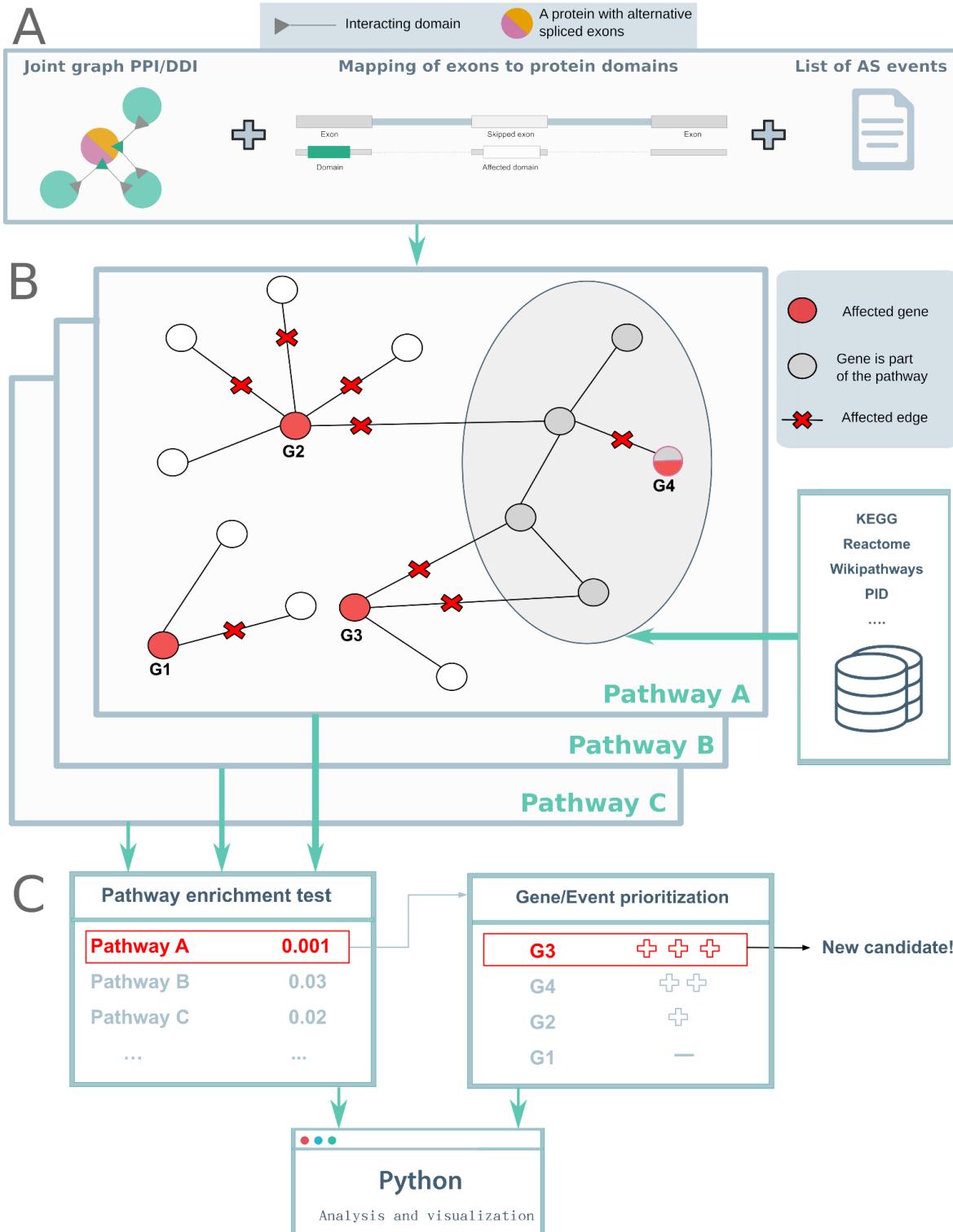


Figure 1. Overview of NEASE. (A) Annotated exons are mapped to Pfam domains. The joint graph of PPIs and DDIs is used to identify the interactions mediated by these domains. (B) For a list of exons/events, NEASE identifies interactions mediated by the spliced domains and pathways that are significantly affected by those

interactions. (C) NEASE provides a corrected p-value, in addition to an enrichment score (NEASE Score) for every pathway (see Methods). The user can further focus on an individual pathway, where NEASE can prioritize genes and find new biomarkers. In this example, the gene G3 was not part of the enriched pathway A but it has the largest number of affected interactions with genes from the pathway.

NEASE uses a hybrid approach that combines biological pathways with PPIs and DDIs to perform functional enrichment of AS. First, we use the structural annotation of known isoforms by mapping protein domains from the Pfam database [22] to the corresponding exons (Figure 1A). Second, we construct a structural joint graph as previously reported [21] by enriching the BioGrid PPI [23] with DDIs from DOMINE [24] and 3did [25] (see Methods). In the joint graph, protein domains are mapped to their mediated interactions. Thus, NEASE addresses the limited exon-level annotation and provides an exon-centric view of the interactome, where exons are represented by their encoded domains and edges represent the binding between the domains. In this way, the impact of AS can be seen as an edgetic change in the network. Analyzed AS events are viewed as a set of affected edges that represent gained or lost PPIs.

We then perform statistical tests to find enriched pathways and most likely responsible genes (Figure 1B). Following, (differential) splicing analysis, a one-sided hypergeometric test is used to test for enrichment of a given pathway or gene set by considering all edges affected by AS in an experiment. A similar test is applied for each spliced gene to prioritize the most relevant events/genes that are affecting a pathway. We further introduce a weighted score (NEASE score) that penalizes hub nodes that are more likely to be connected to the pathway of interest by chance. Notably, this approach also considers genes that are not part of the existing pathway definition but show a significant number of interactions with the pathway, highlighting new putative biomarkers (see Methods for details).

The Python package provides an interactive analysis. Using a list of exons or events, users can run a general enrichment on 12 different pathway databases (collected from the ConsensusPathDB resource [26]), followed up by a specific analysis and visualization for a single affected pathway or module of interest (Figure 1C). To provide analysis for individual isoforms and events, we linked NEASE to our previously developed database DIGGER, which provides an isoform- and exon-centric view of the interactome [21].

NEASE gives insights into the role of the muscle- and neural-specific exons

Recent studies suggest that the regulation of AS occurs in a tissue-specific manner and leads to remodeling of protein-protein interactions [27]. Understanding the functional impact of co-regulated exons is critical in understanding gene regulation. We applied NEASE to tissue-specific exons reported in VastDB, a resource that provides information on multiple types of AS events detected by RNA-seq from different tissue types and developmental stages [28]. We extracted 2,831 exon skipping events and Percent Spliced In values (PSI) from 12 different human tissue types (see Methods). We then performed hierarchical clustering on the z-score standardized PSI values (Figure 2A). The heatmap shows two distinct clusters, where neural-specific and muscle-specific (merged with heart-specific) exons are dominant.

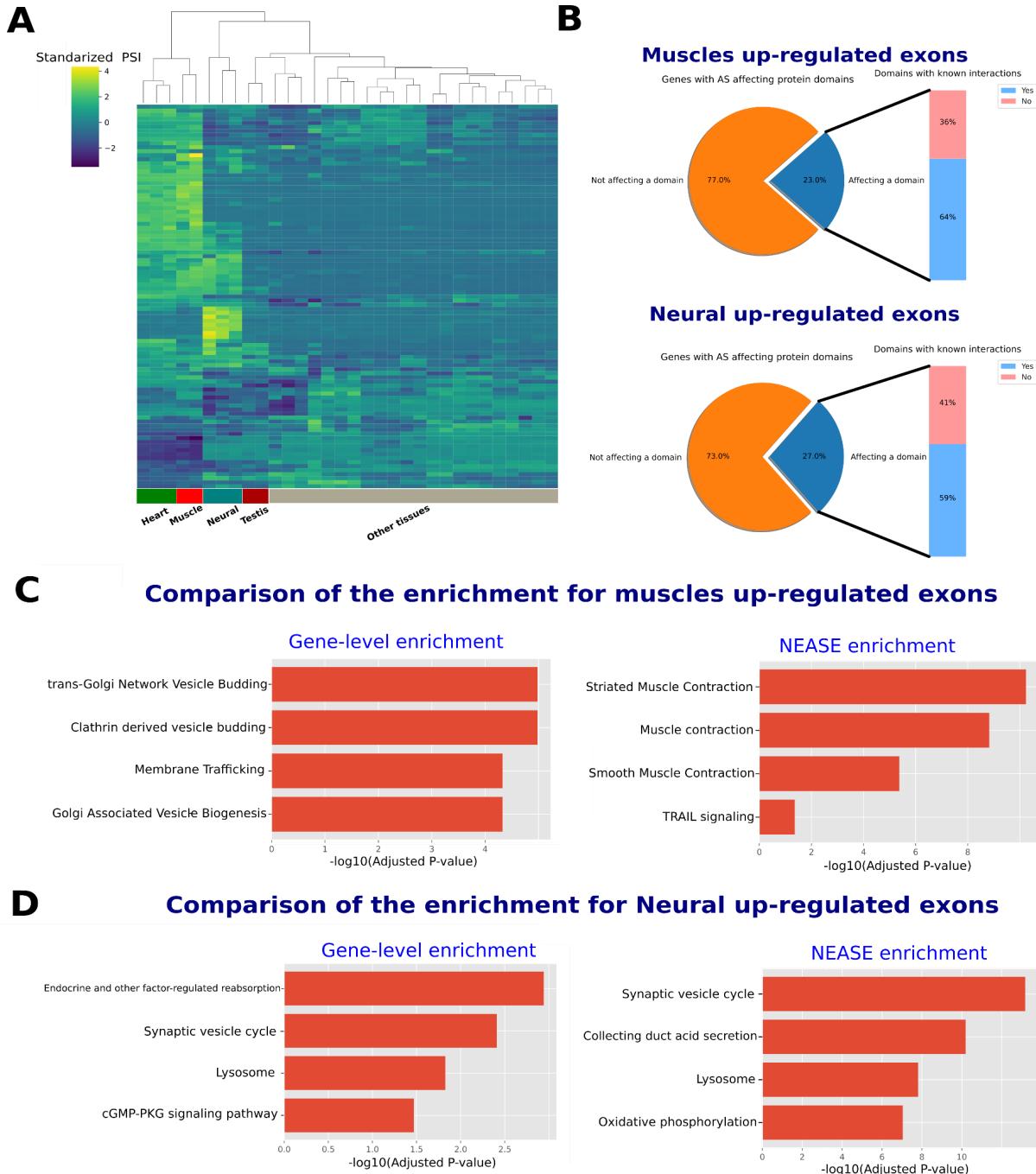


Figure 2. Analysis of tissue-specific exons. (A) Heatmap and hierarchical clustering of standardized PSI values obtained from VastDB. The heatmap only shows events with standard deviation of PSI values ≥ 20 . The heatmap shows that clusters of exons up-regulated in neural tissues and muscle/heart tissues are dominant. (B) NEASE analysis shows that 23% and 27% for both neural and muscle up-regulated exons, respectively, are encoding protein domains. Around 60% of the spliced domains are known to have interaction partners from the PPI and DDI joint graph. (C and D) Comparison between gene-level enrichment and NEASE enrichment for the two sets of exons.

Next, we extracted 66 skipped exons with a high PSI in the muscle tissues and 55 skipped exons with a high PSI in the neural tissues (z-score $\geq +2$, see Methods). We checked how many of these events are overlapping with domains. As shown in Figure 2B, 23% of the upregulated exons in muscle tissues (13) and 27% of the upregulated exons in the neural tissues (17) overlapped with domains. NEASE also provides statistics of how many of these

domains have known binding partners in the joint graph. In the two sets, around 60% of the affected domains have known interactions in our joint graph: 9 binding domains in the muscle tissues and 10 binding domains in the neural tissues (Additional files 2: Tables S4, S5 and Additional file 3: Tables S8, S9). For these groups of events, the exact protein complexes involved can be identified, and NEASE statistical analysis can be performed to determine affected pathways.

First, we ran a gene set overrepresentation analysis (one-sided hypergeometric test), which we refer to as gene-level enrichment, to detect enriched pathways (see Methods). Next, we applied NEASE to the same genes to detect pathways affected by AS. Unlike the gene-level enrichment, the results obtained from NEASE in both sets better explain the functional role of the regulated exons (Figure 2 C and D).

The upregulated exons in heart and muscle tissues were enriched in “Muscle Contraction” pathways (Figure 2C and Additional file 2: Table S7), while, in the gene-level enrichment, the pathways were related to very common subcellular functions such as the Golgi apparatus, which also is an organelle for collecting, modifying or destroying protein products (Figure 2 C and Additional file 2: Table S6). NEASE provides detailed information about the affected domains and their interaction partners (Additional file 1: Table S1). The domain Tropomyosin (Pfam id: PF00261), which is part of the gene TPM1, e.g., is involved in the regulation of muscle contraction via actin and myosin. GAS2 (Pfam id: PF02187) is a domain of DST, a dystonin encoding gene, which plays a role in maintaining the integrity of the cytoskeleton. AS affects its binding with the gene CALM1 that encodes a calcium-binding protein involved in various calcium-dependent pathways like muscle contraction [29].

The exons upregulated in neural tissues showed enrichment in the synaptic vesicle cycle pathway responsible for the communication between neurons (Figure 2D). Gene-level enrichment performed on par with NEASE, resulting in the same pathway but with a lower rank and significance (adjusted p-values: 6.51e-14 using nease and 0.0039 using gene-level, Additional file 3: Tables S10 and S11). Notably, NEASE also detected an enrichment in “oxidative phosphorylation”, which is the initiator for powering all major mechanisms mediating brain information and processing [30]. The neuron's energy demands are remarkable both in their intensity and in their dynamic range and quick changes [31–34]. Therefore, AS could modify oxidative phosphorylation to serve the tissue-specific needs. Experimental studies have also found that several key enzymes in “oxidative phosphorylation” are spliced, e.g. pyruvate kinase (PKM) that shifts from the PKM2 to the PKM1 isoform [35,36]. NEASE also provides a detailed view on the affected mechanisms, such as an exon skipping event in the gene ATP6V0A1 overlapping with the V_ATPase_I domain (PFAM id: PF01496) and affecting the binding with seven other proteins from the complex vacuolar ATPase (V-ATPase) (p-value: 7.14e-17, Figure 3, Additional files 1: Table S2). V-ATPase is required for synaptic vesicle exocytosis [37] The a1-subunit of the V0 domain in ATP6V0A1 was recently shown to be highly expressed in neurons and to be essential for human brain development [38,39]. In another example, NEASE identified two co-regulated events of the genes CLTA and CLTB (Figure 3). CLTA and CLTB genes are involved in Clathrin-dependent endocytosis which forms clathrin-coated vesicles. Both genes play a major role in forming the protein complex of the coated vesicle. Both events affect the same domain Clathrin light chain (Pfam id: PF01086). The Clathrin light chain domain binds to CLTC and CLTCL1 which are the Clathrin heavy chain genes (p-value: 7.356981e-05).

These results suggest that the formation of this complex is co-regulated by AS. A similar finding about the role of the Clathrin light chain in neurons was also described in [40]. NEASE highlights these co-regulated events at the network level (Figure 3). The analysis generated from VastDB using NEASE agrees with the latest studies at transcriptomics and proteomics levels that emphasize the crucial role of AS in the function and development of brain and heart tissues [41–43].

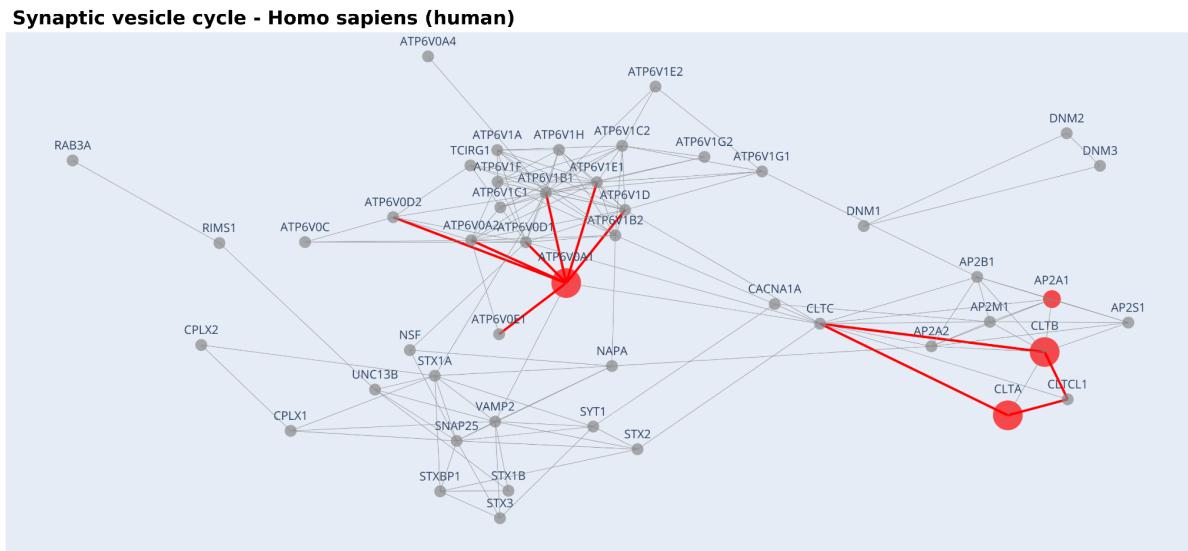


Figure 3: NEASE visually highlights the impact of the AS regulation at the interactome level. The gray nodes represent proteins from the pathway and the red nodes represent genes with AS events. Red edges represent the affected interactions for the nodes with known DDIs. The visualization of the pathway “Synaptic vesicle cycle” from the KEGG database for the exons up-regulated in the neural tissues shows that the splicing in the genes CLTA and CLTB is co-regulated and affects the interactions of the same complex. Similarly, NEASE highlights the importance of the domain ATP6V0A1 which is up-regulated in neural tissues and binds seven proteins from the “Synaptic vesicle cycle” pathway.

NEASE reveals splicing-related differences of reticulated and mature platelets

AS does not only drive tissue-specific regulation but also plays a major role in cell differentiation and maturation. To illustrate an example of the utility of NEASE in such studies, we used the RNA-seq data set from [44] which compares the transcriptome profiles of reticulated platelets and mature platelets from healthy donors. Reticulated platelets are younger [45], larger in size, and contain more RNA [46]. Moreover, they have a prothrombotic potential and are known to be more abundant in patients with diabetes, acute or chronic coronary syndrome, and in smokers [46–48]. Additionally, elevated levels of reticulated platelets in peripheral blood are predictors of insufficient response to antiplatelet therapies (e.g. Aspirin and P2Y12 inhibitors) and are promising novel biomarkers for the prediction of adverse cardiovascular events in different pathological settings [47,49]. A strong enrichment of pro-thrombotic signaling in reticulated platelets was observed in healthy donors [44]. Comparative transcriptomic analysis revealed a differential expression of several pathways in addition to an enrichment of prothrombotic pathways and transcripts of transmembrane proteins as the collagen receptor GPVI, the thromboxane receptor A2 and the thrombin receptors PAR1 and PAR4. Gene set enrichment analysis indicated an upregulation of entire prothrombotic activation pathways as the thrombin PAR1 and integrin GPIIb/IIIa signalling pathway in reticulated platelets.

Since AS has been described to occur in platelets [50], we wanted to investigate the splicing patterns between the previously defined reticulated and mature platelet subgroups. Using MAJIC [51] (see Methods), we found 169 differentially spliced genes. From 25 affected protein domains, 17 have known interactions (68% of affected domains, Figure 4 A, Additional file 4: Tables S12 and S13).

We observed that the enrichment at the gene-level using the Reactome [52] database ranks general cellular pathways higher, including “Membrane Trafficking” and “Vesicle-mediated transport”, and “Golgi-to-ER retrograde transport”. An exception is the “Circadian Clock” pathway, which is hypothesized to be related to platelet activation [53] (Figure 4 B). The pathway “Platelet activation, signaling and aggregation” was less significant in gene-level enrichment (adjusted p-value: 0.061, Additional file 4: Table S14) compared to NEASE enrichment (adjusted p-value: 0.004, Additional file 4: Table S15). Using NEASE, we obtained more meaningful results and unique pathways. As shown in Figure 4C, the most significant pathways in reticulated platelets are G Protein-Coupled Receptor-related. G proteins are essential in the second phase of platelet-dependent thrombus formation [54]. Furthermore, GPCR isoforms are known to have distinct signaling properties [55]. Other relevant pathways associated with platelet activation are “Hemostasis”, “Thromboxane signaling through tp receptor”, and “Platelet homeostasis”. The full tables for enrichment at the gene level and using NEASE are available in the Additional file 4: Tables S14 and S15. The upregulation of these pathways in reticulated platelets emphasizes their previously described prothrombotic phenotype and their involvement in several downstream signaling processes.

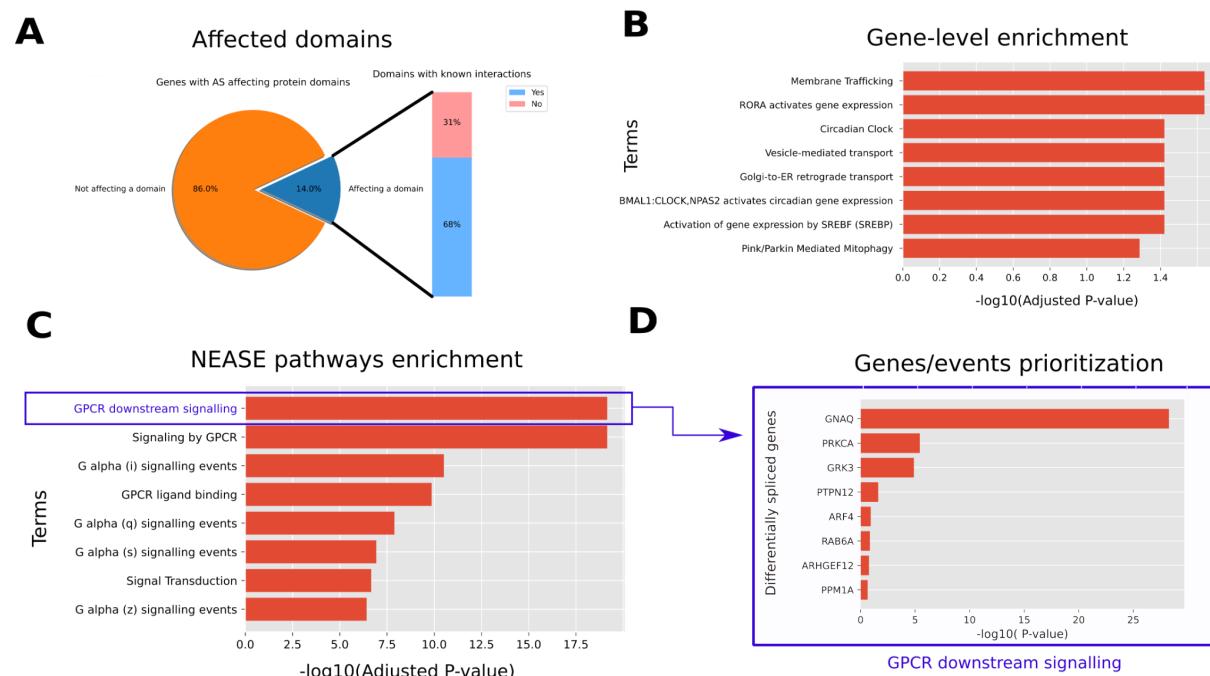


Figure 4. (A) 14 % of differentially spliced exons, between reticulated and mature platelets, are known to encode protein domains. (B) Gene level enrichment of all differentially spliced exons in the Reactome database fails to capture most relevant pathways. (C) In contrast, NEASE shows an enrichment of the GPCR downstream signaling and other related pathways that are well known to be important in platelet activation. (D) A further look at the genes driving the enrichment of the GPCR pathway shows the most relevant genes affected by AS.

We also looked at the individual AS events driving this enrichment. For each affected domain, NEASE tests if it significantly interacts with the GPCR downstream signaling pathway (Additional file 4: Table S16, see Methods). Figure 4C illustrates affected genes and their p-value ranking. The top gene is GNAQ (G-protein subunit alpha q), which is known to be involved in signal transduction in platelets leading to platelet activation [56]. The regulation of the G-protein alpha subunit can be an indication that compared to mature platelets, reticulated platelets are more involved in various signal transduction pathways related to, e.g., pro-thrombotic processes [46]. PRKCA, which also showed different splicing patterns between the two platelet subgroups, plays a major role in the platelet formation process by modulating platelet function [57], megakaryocyte function, and development [58] and negatively regulates pro-platelet formation [59]. Moreover, the regulation of PRKCA binding in reticulated platelets might refer to the young nature of reticulated platelets, which have undergone the pro-platelet formation process more recently than mature platelets [45,60].

NEASE characterizes complex disorders such as Multiple Sclerosis

Multiple sclerosis (MS) is a chronic inflammatory demyelinating disease of the central nervous system. Early in the disease course, MS is characterized by focal lesions in the brain induced by influx of systemic inflammatory cells. These active lesions infiltrated by immune cells and activated microglia are characterized by inflammatory demyelination and axonal loss [61]. The surrounding white matter tissue is termed normal-appearing white matter due to diffuse pathology without focal lesion activity and dense immune activity [62]. The etiology of MS remains unknown. Recently, a systematic literature review found 27 genes that were alternatively spliced in MS patients [63].

We used RNA-Seq of macrodissected areas from postmortem white matter tissue of patients with progressive MS [64]. We compared normal-appearing white matter and active lesions regions from postmortem white matter brains of MS patients. We found 109 differentially spliced genes and 19 affected domains with known interactions. In total, NEASE identified 150 affected interactions (Additional file 5: Tables S17 and S18).

Gene-level enrichment ranks high pathways likely irrelevant that are involved in muscle contraction, cardiac conduction, and membrane Trafficking, with the exception of Ca2+ ion flow across membranes.(Additional file 5: Table S19). Ca2+ is an essential signal molecule for all cell activity. Although deregulation of calcium signaling is related to the pathogenesis of multiple diseases [65], including neurological disorders [66], It is not specific to neuronal tissues. In line with the neurodegenerative and immune-mediated features of MS, NEASE found unique enriched pathways related to brain network signaling and neuronal pathways “Neurotransmitter receptors and postsynaptic signal transmission”, “Transmission across Chemical Synapses”, “Activation of NMDA receptor and postsynaptic events”, “MAPK family signaling cascades”, “Neuronal System”), as well as pathways related to immune responses (“interleukin-17 signaling, “Toll Like Receptor 10 (TLF10) Cascade”) (Table 1 and Additional file 5: Table S20). Two other pathways were related to the uptake of anthrax or bacterial toxins. This could be a result of clean-up from toxic inflammatory processes or increased presence of invaders due to the leaky brain-blood-barrier in MS [67–69]. Additionally, it also supports the theory of infections as the trigger of lesion damage in MS [70].

As shown in Table 1, the pathway “Uptake and function of anthrax toxins” has the best overall adjusted p-value, calculated only based on the total number of edges affecting the pathway. When we also included the number of significant genes and calculated NEASE scores (see Methods), NEASE ranks the pathway “Neurotransmitter receptors and postsynaptic signal transmission” first, and moves pathways such as “Transmission across Chemical Synapses” and “Neuronal System” higher in the rank. These observations illustrate the usefulness of the NEASE score as a complement to the global edge-based enrichment.

Two of the most significant genes in the “Neurotransmitter receptors” pathway were GRIN1 and GRIA1 (Additional file 5: Table S21). GRIN1 encodes GluN1, which is one of the two obligatory subunits for the NMDAR1 receptor, where GRIA1 encodes the AMPAR1 subunit. Their ligand is glutamate, and they are both ionotropic receptors and have been associated with MS disease severity [71–73]. Interestingly, AS of MAP2K4 appeared in both brain-related and immune-related pathways, significantly enriched in active lesions vs normal-appearing white matter (Table 1). MAP2K4 is a mitogen-activated protein kinase (MAPK) orchestrating multiple biological functions [74,75]. AS of MAP2K4 has been found in rheumatoid arthritis [76], as well as in pathways of patients with other autoimmune diseases [77]. MS also precedes autoimmune attack, and therefore AS of MAP2K4 in active lesions detected with NEASE may represent dysregulated immune responses originating from the infiltrating immune cells or inflammatory-activated brain cells. This is supported by previous studies that found (*i*) overactivity of MAPK pathways in microglia (the resident immune cell of the brain) during neurodegeneration [78,79], and (*ii*) increased phosphorylation of MAPK kinases in the systemic immune cells of MS patients [80,81]. A recent study also characterized activated MS-specific pathways in immune cells from blood using phosphoproteomics. Here, MAP2K4 and its interaction partners (e.g. TAK1) were present in MS-specific signaling activity [82]. Future functional studies on the AS of MAP2K4 may help explain if AS could be the reason for increased phosphorylation and overactivity detected in MS. AS of MAP2K4 could result in switching protein conformation, increasing susceptibility to phosphorylation or changing the downstream protein cascade.

With NEASE, we were able to specifically detect AS of genes and related pathways already known to be dysregulated within MS from excitotoxicity to inflammation. The detected AS genes in active lesions vs normal-appearing white matter demonstrate how major components in signaling activities may be fine-tuned/changed from regulation of a homeostatic state to an inflammatory state. Combining NEASE with functional experiments to understand the biological impact of AS could fuel new therapeutic opportunities for complex neurological diseases as MS. Novel developments in genome-editing tools and gene-specific strategies have made it possible to use antisense oligonucleotides or small modulators for splice modification. This is already used in the rare neuromuscular disease, spinal muscular atrophy, where an antisense oligonucleotide binds to a site near splicing to ensure the inclusion of an exon during the splicing event [78].

Table 1 NEASE enrichment obtained from AS comparison between normal-appearing white matter (NAWM) and acute lesions (AL), from multiple sclerosis patients. The highly enriched pathways belong to Neurotransmitter receptors, MAPK, and bacterial infection. Most of these pathways are hallmarks of MS. The NEASE score is obtained after combining the p-value with the number of significant genes. The latter is obtained after individual tests for each gene in the column “Spliced genes” (see Methods).

Pathway name	Spliced genes (number of interactions affecting the pathway)	p-value	adj p-value	NEASE score
Neurotransmitter receptors and postsynaptic signal transmission	GRIA1 (7), ATP2B1 (1), BRAF (4), MAP2K4 (1), GRIN1 (4)	2.128010e-08	0.000009	15.34
Uptake and function of anthrax toxins	ATP2B1 (1), BRAF (5), MAP2K4 (3)	2.468257e-09	0.000004	14.90
Transmission across Chemical Synapses	GRIA1 (7), ATP2B1 (1), BRAF (4), MAP2K4 (1), GRIN1 (4)	2.207774e-07	0.000037	13.31
Uptake and actions of bacterial toxins	ATP2B1 (1), BRAF (5), MAP2K4 (3)	2.560786e-08	0.000009	13.14
Activation of NMDA receptor and postsynaptic events	GRIA1 (2), ATP2B1 (1), BRAF (4), MAP2K4 (1), GRIN1 (3)	2.447069e-06	0.000194	11.22
MAPK family signaling cascades	MYH10 (2), ATP2B1 (1), BRAF (17), MAP2K4 (5), GRIN1 (3)	1.140595e-06	0.000148	10.29
Neuronal System	GRIA1 (7), ATP2B1 (1), BRAF (4), MAP2K4 (1), GRIN1 (4)	2.657946e-06	0.000194	9.65
FCER1 mediated MAPK activation	MYH10 (1), BRAF (7), MAP2K4 (8)	5.455227e-07	0.000082	8.85
RAF/MAP kinase cascade	MYH10 (1), ATP2B1 (1), BRAF (16), MAP2K4 (4), GRIN1 (3)	7.076613e-07	0.000099	8.69
MAPK1/MAPK3 signaling	MYH10 (1), ATP2B1 (1), BRAF (16), MAP2K4 (4), GRIN1 (3)	1.736848e-06	0.000194	8.14
Interleukin-17 signaling	BRAF (6), MAP2K4 (12)	2.203770e-06	0.000194	7.99
MAP kinase activation	BRAF (6), MAP2K4 (12)	2.203770e-06	0.000194	7.99
Toll Like Receptor 10 (TLR10) Cascade	BRAF (6), MAP2K4 (13)	2.618844e-06	0.000194	7.89

NEASE finds new biomarker candidates for Dilated Cardiomyopathy

AS might play a role in driving Dilated Cardiomyopathy (DCM) [83]. DCM is a common heart muscle disease that is often diagnosed with structural abnormalities resulting in impaired contraction. Previous studies have shown a large number of differentially used exons in DCM patients [4,10]. In this analysis, we used a list of 1,212 differentially used exons between DCM patients and controls as reported by Heinig *et al.* [10]. 28% of these exons overlapped with domains (Additional file 6: Tables S22 and S23). In this exon set, both the gene level enrichment and NEASE show very similar results (Additional file 6: Tables S24 and S25). In both methods, we found that the list of exons was enriched in the Dilated cardiomyopathy (DCM) pathway from KEGG, as well as, “Adrenergic signaling in cardiomyocytes”, and “Regulation of actin cytoskeleton”.

In contrast to gene-level enrichment analysis, NEASE is able to score the contribution of alternatively spliced genes that are interacting with but are not part of the DCM pathway, allowing us to highlight putative biomarkers (Table 2, Additional file 6: Table S26, Additional files 1: Figure S1). The Myosin head domain from the gene MYO19 interacts with 6 other genes associated with DCM: (1) MYL2, which triggers contraction after Ca⁺ activation [84]; (2-5) TPM1/TPM2/TPM3/TPM4, which encode the tpm protein - the main regulator of muscle contraction [85]; and (6) ACTG, which encodes actin. Interestingly, MYO19 has not been investigated for its role in DCM, while its interacting genes are associated with DCM [86–89]. Additionally, the gene OBSCN has one affected interaction with the TTN gene [90]. The TTN gene itself is also differentially spliced associated with DCM [90]. OBSCN was recently reported as a new DCM candidate [91,92]. Another interesting example is CACNA1C (Calcium Voltage-Gated Channel Subunit Alpha1 C), an already known DCM candidate [93]. The differentially spliced exon overlaps with the domain Ion_trans (Pfam id: PF00520) that is essential for myocyte contraction [94]. The affected interaction identified is with the Ryanodine receptor 2 (RYR2). In striated muscles, excitation-contraction coupling is mediated by this complex [95]. Both CACNA1C and RYR2 are part of the KEGG DCM pathway [96]. Alterations in Ryanodine Receptors were repeatedly reported to be related to heart failure [97–99].

Table 2: Enrichment of the pathway “Dilated cardiomyopathy (DCM)” from KEGG for the exons differentially used in DCM patients. The table shows the most significant genes (p_value<0.05) (See Methods).

Differentially Spliced Genes	DCM associated	Percentage of affected edges associated with DCM	P-value	Affected binding (edges) associated with DCM
MYO19	No	6/51	0.000002	MYL2,TPM4,TPM3,TPM2,TPM1,ACTG
OBSCN	No	1/2	0.014717	TTN
USP49	No	1/4	0.029217	PRKACA
CACNA1C	Yes	1/4	0.029217	RYR2

Discussion

In spite of its importance for biomarker and therapeutic target discovery, differential AS is still not a routine part of transcriptome analysis. A key reason for this could be the lack of suitable methods and software tools for AS-specific functional analysis. Our method NEASE closes this gap and provides a unique view on the impact of AS complementary to functional insights gained from traditional gene-level enrichment analysis. We applied NEASE to four diverse data sets and show that its results generate novel disease-relevant insights and provide valuable context to prior findings on altered RNA- and protein-expression levels consistent with recent literature.

In many cases, NEASE improves over gene-level enrichment analysis focusing on differentially spliced genes. One potential reason for this could be that not all AS events are necessarily functional [11,12]. NEASE mitigates this by focusing on AS events that affect protein domains. However, it is important to keep in mind that this is not the only way to define functional AS events. AS also affects interacting disordered regions [14] or facilitates nonsense-mediated decay [100].

AS events could also lead to completely different functions or interactions [101], e.g., two isoforms can have different interaction partners depending on the inclusion or loss of a single domain [13]. Such changes in the interactome can not be captured with gene-level enrichment which has a strict focus on nodes rather than edges. With NEASE, we could show that integrating structural information at the exon level and PPI networks helps to identify the functional impact of differentially spliced and co-regulated exons. In practice, we consider both approaches as complementary and recommend running gene-level and edge-level enrichments together (both supported by the NEASE package).

NEASE relies on structurally annotated interactions and existing pathway annotations from databases such as KEGG [96] and Reactome [52]. Leveraging reliable structural information and established pathways likely removes many false positive PPI from considerations. While DDI are generally of high quality, it should be noted that not all AS events are necessarily of functional consequence since other processes such as nonsense-mediated decay need to be considered as well. With our current approach, a large fraction of the PPI network remains unexplored, suggesting that adapting *de novo* network enrichment methods such as KeyPathwayMiner [102] towards AS could be a promising research direction to uncover previously unknown disease mechanisms. NEASE currently considers the immediate neighborhood of a pathway in the PPI network. When carefully considering the expected increase in false positives, one could also increase the size of the pathway neighborhood using, e.g., a fixed radius for shortest paths. While these are attractive approaches, the biases of the PPI towards hubs, as well as the high number of false (or missing) edges of PPI, in its current form, make such approaches hard to control and statistically challenging. Even though NEASE is relatively conservative, we demonstrated that it is simple, robust, and generates meaningful and interpretable results. Thus, it provides an unprecedented opportunity to understand the functional impact of tissue-, developmental- and disease-specific AS in a system biology manner.

While a plethora of gene set enrichment methods have been proposed in recent years, AS is typically not addressed specifically. Thus, NEASE closes an important gap in functional enrichment analysis of transcriptomics data. The analyses described here, confirm the

widespread impact of AS in multiple biological processes and disorders. In the future, we plan to extend NEASE with further model organisms and to add structural annotations covering more types of AS events. Finally, we plan to integrate NEASE with the DIGGER web tool [21] for a seamless downstream analysis of AS in the web browser with the vision of establishing functional AS event analysis as a routine step in transcriptomic analysis.

Methods

NEASE Data sources

We construct a human structurally annotated PPI as described previously [21]. Briefly, we integrate DDI and PPI information into a joint network where DDIs were obtained from 3did (v2019_01 [25]) and DOMINE (v2.0 [24] including high- and mid-confidence interactions) and PPIs were obtained from BioGRID 3.5 [23]. In summary, out of 410,961 interactions from the human interactome 52,467 have at least one domain interaction. The mapping of exons to domains was performed using DIGGER's mapping table that in turn uses BioMart and Pfam annotations [22,103]. We obtain the biological pathways with their gene list from KEGG [52] and Reactome [96] integrated into the ConsensusPathDB database [26].

Statistical tests and pathway scores

Gene-level enrichment is performed using a hypergeometric test from the package GSEAPY (a Python wrapper for Enrichr [104]) by considering all genes with (differential) AS events.

For NEASE enrichment, we filtered the PPI graph $G=(V,E)$, where V is the set of genes and E is the set of edges, to a subgraph $G'=(V', E')$ containing only structurally annotated interactions E' and their nodes V' .

For a submitted query list of exons, NEASE first identifies affected domains that overlap with the exons and their interactions. Let N be two times the number of edges in G' (the degree of the network) and n be the number of affected edges from the query. These edges are then considered using a test modified from [105]: For every pathway P with degree K , let k be the number of affected edges that are connected to P . We model X whose outcome is k as a random variable following a hypergeometric distribution:

$$X \sim \text{Hypergeometric} (n=\text{number of affected edges}, K=\text{degree of } P, N=\text{degree of } G')$$

where k is considered as the number of observed successes out of n draws, from a population of size N containing K success. Subsequently, NEASE tests if the number k is significant using a one-sided hypergeometric test (over-representation). After testing for multiple pathways, the obtained p-values for the edge-level enrichment are corrected, using the Benjamini-Hochberg method [106].

For a pathway of interest, a similar test can be applied to determine if a splicing event significantly affects interactions of a specific gene with this pathway. Here, n is the number of all affected interactions (edges) of a spliced gene and k is the number of affected interactions (edges) across genes that are linked to the pathway of interest. As a result, for every pathway, NEASE provides an overall p-value, as well as the most significant genes. Since the p-value only depends on the overall number of affected edges but not on the number of genes, the p-value can be heavily influenced by hub genes. To reduce this

influence, an optional score (NEASE Score) can be computed by NEASE to scale the natural logarithm of the p-value with the total number of significant genes using a cutoff from the user (for instance p-value<0.05):

$$NEASE\ Score = -\sqrt{G} \times \log_{10}(p - value)$$

where G is the total number of significantly connected genes obtained after testing individual spliced genes. Thus, the NEASE Score prioritizes pathways that are affected by a larger number of spliced genes rather than pathways that have a larger number of affected interactions (edges). The user can choose to rank enrichment based on the adjusted p-value or by the NEASE score.

VastDB events processing

PSI values of the exon skipping events from VastDB were quantified by the developers using vast tools [28,107]. In our analysis, we extracted the PSI values for 32 experiments belonging to 12 main tissues: muscle/heart, neural (whole brain, cortex and peripheral retina), placental, epithelial, digestive (colon and stomach), liver, kidney, adipose, testis, immune-hematopoietic and ovary. We then filtered out the events with low read coverage (VLOW) and performed hierarchical clustering of standardized values (z-scores). For every exon, we calculated the mean of PSI values from the samples of the same tissues. To extract muscle/heart and neural-specific exons, we applied two filters, namely that the z-score of the exon PSI value in the relevant tissue is higher than 2 and that the mean PSI value of the exon across all tissues is between 0.15 and 0.90. The latter ensures that we consider only exons that are part of AS events.

RNA-Seq analysis

Raw RNA-Seq reads for two types of platelets and multiple sclerosis patients were downloaded from the GEO repository (access numbers: GSE126448 and GSE138614). The number of samples and sequencing depth are reported in Additional file 1: Table S3. RNA-Seq reads were aligned to the reference human genome (hg38) using STAR 2.7 [108] in a 2-pass mode and filtered for uniquely mapped reads. Differential AS analysis was performed by MAJIQ [51] with default parameters, and with a threshold of $P(dPSI > 20\%) > 0.95$.

NEASE: The Python package

NEASE's Python package relies on NumPy [109], pandas [110], NetworkX [111], SciPy [112], and Statsmodels [113]. The gene-level enrichment is also supported in the NEASE package using the Python implementation of Enrichr [104]. To speed up the edge hypergeometric test, the total degree of every pathway in the structural PPI, as well as the overall degree of the network were pre-computed. For visualization, we use the complete PPI (not the structural PPI) and extract connected subnetworks from each pathway as well as spliced genes and their interactions with the extracted modules. The position of nodes is computed using the Fruchterman-Reingold force-directed algorithm implemented in NetworkX [114]. The interactive visualization for individual genes and events is implemented with information from the DIGGER database and the Plotly package.

The standard input of the package is a DataFrame object with the exon coordinates and Ensembl IDs of the genes. The package also supports the output of AS differential detection

tools that are event-based such as MAJIQ [51] where NEASE only considers annotated exons. NEASE is released as open-source under the GPLv3 license and available at (<https://github.com/louadi/NEASE>). The code used to produce the results in this manuscript was deposited at Zenodo (<https://doi.org/10.5281/zenodo.4985321>).

Declarations

Availability of data and materials: RNA sequencing data for reticulated platelets was provided by the authors [44] and it is freely available at GEO (access number: GSE126448). Multiple sclerosis Raw sequence was provided by the authors [64] and freely available at GEO (access number: GSE138614). Dilated Cardiomyopathy raw data is available in the European Genome-phenome Archive (Dataset ID: EGAS00001002454), in our analysis we used pre-processed data from the manuscript [10]. Vastdb dataset for humans (hg19) was downloaded from <https://vastdb.crg.eu/wiki/Downloads>. The generated joint graph and the exon mapping database are available on the DIGGER database website <https://exbio.wzw.tum.de/digger/download>. All processed datasets, as well as step-by-step tutorials for using NEASE to reproduce the results presented in this paper, are available at <https://github.com/louadi/NEASE-tutorials> and deposited to Zenodo (<https://doi.org/10.5281/zenodo.4985321>).

Competing interests: The authors declare that they have no competing interests

Funding: This project received financial support from BMBF grant Sys_CARE (no. 01ZX1908A) of the Federal German Ministry of Research and Education. MLE is grateful for financial support from Lundbeckfonden (no. R347-2020-2454).

Authors' contributions: ZL, TK, OT, JB and ML conceived the project. ZL performed the initial experiments, developed the method, and implemented the software. OT and ML supervised the project, provided critical feedback, and helped shape the research. OT prepared the RNA-Seq datasets for the differential splicing analysis. AF, CTL interpreted and discussed the biological insights of the results. MK and DB interpreted and discussed the platelet analysis. MLE and ZI interpreted and discussed the Multiple Sclerosis analysis. ZL wrote the initial draft of the manuscript. All authors contributed to writing the final manuscript and approved the final version.

Acknowledgements: Not applicable.

References:

1. Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, et al. Function of alternative splicing. *Gene*. 2005;344:1–20.
2. Yeo G, Holste D, Kreiman G, Burge CB. Variation in alternative splicing across human tissues. *Genome Biol*. 2004;5:R74.
3. Baralle F, Giudice J. Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol*. Springer Nature; 2017;18. Available from: <http://dx.doi.org/10.1038/nrm.2017.27>
4. Beqqali A. Alternative splicing in cardiomyopathy. *Biophys Rev*. 2018;10:1061–71.

5. Douglas AGL, Wood MJA. Splicing therapy for neuromuscular disease. *Mol Cell Neurosci*. 2013;56:169–85.
6. Esvyukova I, Somarelli JA, Gregory SG, Garcia-Blanco MA. Alternative splicing in multiple sclerosis and other autoimmune diseases. *RNA Biology*. Taylor and Francis Inc.; 2010. p. 462–73. Available from: <https://pubmed.ncbi.nlm.nih.gov/20639696/>
7. López-Bigas N, Audit B, Ouzounis C, Parra G, Guigó R. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett*. 2005;579:1900–3.
8. Karlebach G, Veiga DFT, Mays AD, Chatzipantsiou C, Barja PP, Chatzou M, et al. The impact of biological sex on alternative splicing. *bioRxiv*. 2020 [cited 2021 Jul 12]. p. 490904. Available from: <https://www.biorxiv.org/content/10.1101/490904v2.abstract>
9. Tollervey JR, Wang Z, Hortobágyi T, Witten JT, Zarnack K, Kayikci M, et al. Analysis of alternative splicing associated with aging and neurodegeneration in the human brain. *Genome Res*. 2011;21:1572–82.
10. Heinig M, Adriaens ME, Schafer S, van Deutekom HWM, Lodder EM, Ware JS, et al. Natural genetic variation of the cardiac transcriptome in non-diseased donors and patients with dilated cardiomyopathy. *Genome Biol*. 2017;18:170.
11. Tress ML, Abascal F, Valencia A. Alternative Splicing May Not Be the Key to Proteome Complexity. *Trends in Biochemical Sciences*. Elsevier Ltd; 2017. p. 98–110. Available from: <https://pubmed.ncbi.nlm.nih.gov/27712956/>
12. Melamud E, Moult J. Stochastic noise in splicing machinery. *Nucleic Acids Res*. 2009;37:4873–86.
13. Yang X, Coulombe-Huntington J, Kang S, Sheynkman GM, Hao T, Richardson A, et al. Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell*. 2016;164:805–17.
14. Buljan M, Chalancón G, Eustermann S, Wagner GP, Fuxreiter M, Bateman A, et al. Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol Cell*. 2012;46:871–83.
15. da Costa PJ, Menezes J, Romão L. The role of alternative splicing coupled to nonsense-mediated mRNA decay in human disease. *Int J Biochem Cell Biol*. 2017;91:168–75.
16. Kristoffer V-S, Sandelin A. The Landscape of Isoform Switches in Human Cancers. *Genomics*. 2017; Available from: <http://dx.doi.org/10.1158/1541-7786.MCR-16-0459>
17. delaFuente L, Arzalluz-Luque Á, Tardáguila M, DelRisco H, Martí C, Tarazona S, et al. tappAS: a comprehensive computational framework for the analysis of the functional impact of differential splicing. *Genome Biol. NLM (Medline)*; 2020;21. Available from: <http://dx.doi.org/10.1186/s13059-020-02028-w>
18. Gal-Oz ST, Haiat N, Eliyahu D, Shani G, Shay T. DoChaP: the domain change presenter. *Nucleic Acids Res*. 2021;49:W162–8.
19. Ctor Climente-Gonzá Lez H, Porta-Pardo E, Godzik A, Correspondence EE, Eyras E. The Functional Impact of Alternative Splicing in Cancer. *Cell Reports*. 2017;20:2215–26.
20. Tranchevent L-C, Aubé F, Dulaquier L, Benoit-Pilven C, Rey A, Poret A, et al. Identification of protein features encoded by alternative exons using Exon Ontology.

Genome Res. 2017;27:1087–97.

21. Louadi Z, Yuan K, Gress A, Tsoy O, Kalinina OV, Baumbach J, et al. DIGGER: exploring the functional role of alternative splicing in protein interactions. *Nucleic Acids Res.* Oxford University Press (OUP); 2020; Available from: <https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkaa768/5911747>
22. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* 2019;47:D427–32.
23. Oughtred R, Stark C, Breitkreutz B-J, Rust J, Boucher L, Chang C, et al. The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* 2019;47:D529–41.
24. Yellaboina S, Tasneem A, Zaykin DV, Raghavachari B, Jothi R. DOMINE: a comprehensive collection of known and predicted domain-domain interactions. *Nucleic Acids Res.* 2011;39:D730–5.
25. Mosca R, Céol A, Stein A, Olivella R, Aloy P. 3did: a catalog of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res.* 2014;42:D374–9.
26. Kamburov A, Stelzl U, Lehrach H, Herwig R. The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res.* 2013;41:D793–800.
27. Ellis JD, Barrios-Rodiles M, Çolak R, Irimia M, Kim T, Calarco JA, et al. Tissue-Specific Alternative Splicing Remodels Protein-Protein Interaction Networks. *Mol Cell.* 2012;46:884–92.
28. Tapial J, Ha KCH, Sterne-Weiler T, Gohr A, Braunschweig U, Hermoso-Pulido A, et al. An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Res.* 2017;27:1759–68.
29. Tansey MG, Luby-Phelps K, Kamm KE, Stull JT. Ca(2+)-dependent phosphorylation of myosin light chain kinase decreases the Ca2+ sensitivity of light chain phosphorylation within smooth muscle cells. *J Biol Chem.* 1994;269:9912–20.
30. Hall CN, Klein-Flügge MC, Howarth C, Attwell D. Oxidative phosphorylation, not glycolysis, powers presynaptic and postsynaptic mechanisms underlying brain information processing. *J Neurosci.* 2012;32:8940–51.
31. Sanganahalli BG, Herman P, Blumenfeld H, Hyder F. Oxidative neuroenergetics in event-related paradigms. *J Neurosci.* 2009;29:1707–18.
32. Vergara RC, Jaramillo-Riveri S, Luarte A, Moënne-Locoz C, Fuentes R, Couve A, et al. The Energy Homeostasis Principle: Neuronal Energy Regulation Drives Local Network Dynamics Generating Behavior. *Front Comput Neurosci.* 2019;13:49.
33. Du F, Zhu X-H, Zhang Y, Friedman M, Zhang N, Ugurbil K, et al. Tightly coupled brain activity and cerebral ATP metabolic rate. *Proc Natl Acad Sci U S A.* 2008;105:6409–14.
34. Howarth C, Gleeson P, Attwell D. Updated energy budgets for neural computation in the neocortex and cerebellum. *J Cereb Blood Flow Metab.* 2012;32:1222–32.
35. Magistretti PJ, Allaman I. A cellular perspective on brain energy metabolism and functional imaging. *Neuron.* 2015;86:883–901.
36. Zheng X, Boyer L, Jin M, Mertens J, Kim Y, Ma L, et al. Metabolic reprogramming during

neuronal differentiation from aerobic glycolysis to neuronal oxidative phosphorylation. *Elife*. 2016;5. Available from: <http://dx.doi.org/10.7554/eLife.13374>

37. Hiesinger PR, Fayyazuddin A, Mehta SQ, Rosenmund T, Schulze KL, Zhai RG, et al. The v-ATPase V0 subunit a1 is required for a late step in synaptic vesicle exocytosis in *Drosophila*. *Cell*. 2005;121:607–20.
38. Aoto K, Kato M, Akita T, Nakashima M, Mutoh H, Akasaka N, et al. ATP6V0A1 encoding the a1-subunit of the V0 domain of vacuolar H⁺-ATPases is essential for brain development in humans and mice. *Nat Commun*. 2021;12:2107.
39. Poëa-Guyon S, Amar M, Fossier P, Morel N. Alternative splicing controls neuronal expression of v-ATPase subunit a1 and sorting to nerve terminals. *J Biol Chem*. 2006;281:17164–72.
40. Redlingshöfer L, McLeod F, Chen Y, Camus MD, Burden JJ, Palomer E, et al. Clathrin light chain diversity regulates membrane deformation in vitro and synaptic vesicle formation in vivo. *Proc Natl Acad Sci U S A*. 2020;117:23527–38.
41. Rodriguez JM, Pozo F, di Domenico T, Vazquez J, Tress ML. An analysis of tissue-specific alternative splicing at the protein level. Orengo CA, editor. *PLoS Comput Biol*. Public Library of Science; 2020;16:e1008287.
42. Raj B, Blencowe BJ. Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. *Neuron*. 2015;87:14–27.
43. Su C-H, D D, Tarn W-Y. Alternative Splicing in Neurogenesis and Brain Development. *Front Mol Biosci*. 2018;5:12.
44. Bongiovanni D, Santamaria G, Klug M, Santovito D, Felicetta A, Hristov M, et al. Transcriptome Analysis of Reticulated Platelets Reveals a Prothrombotic Profile. *Thromb Haemost*. 2019;119:1795–806.
45. Ault KA, Knowles C. In vivo biotinylation demonstrates that reticulated platelets are the youngest platelets in circulation. *Exp Hematol*. 1995;23:996–1001.
46. Karpatkin S. Heterogeneity of human platelets. II. Functional evidence suggestive of young and old platelets. *J Clin Invest*. 1969;48:1083–7.
47. Cesari F, Marcucci R, Gori AM, Caporale R, Fanelli A, Casola G, et al. Reticulated platelets predict cardiovascular death in acute coronary syndrome patients. *Thromb Haemost*. Schattauer GmbH; 2013;109:846–53.
48. Guthikonda S, Alviar CL, Vaduganathan M, Arikan M, Tellez A, DeLao T, et al. Role of reticulated platelets and platelet size heterogeneity on platelet activity after dual antiplatelet therapy with aspirin and clopidogrel in patients with stable coronary artery disease. *J Am Coll Cardiol*. Elsevier BV; 2008;52:743–9.
49. Muronoi T, Koyama K, Nunomiya S, Lefor AK, Wada M, Koinuma T, et al. Immature platelet fraction predicts coagulopathy-related platelet consumption and mortality in patients with sepsis. *Thromb Res*. 2016;144:169–75.
50. Nassa G, Giurato G, Cimmino G, Rizzo F, Ravo M, Salvati A, et al. Splicing of platelet resident pre-mRNAs upon activation by physiological stimuli results in functionally relevant proteome modifications. *Sci Rep*. 2018;8:498.
51. Vaquero-Garcia J, Barrera A, Gazzara MR, González-Vallinas J, Lahens NF, Hogenesch

JB, et al. A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife*. 2016;5:e11752.

52. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* 2018;46:D649–55.

53. Scheer FAJL, Michelson AD, Frelinger AL 3rd, Evoniuk H, Kelly EE, McCarthy M, et al. The human endogenous circadian system causes greatest platelet activation during the biological morning independent of behaviors. *PLoS One*. 2011;6:e24549.

54. Offermanns S. Activation of Platelet Function Through G Protein–Coupled Receptors. *Circ Res*. American Heart Association; 2006;99:1293–304.

55. Marti-Solano M, Crilly SE, Malinverni D, Munk C, Harris M, Pearce A, et al. Combinatorial expression of GPCR isoforms affects signalling and drug responses. *Nature*. 2020;587:650–6.

56. Jalagadugula G, Dhanasekaran DN, Kim S, Kunapuli SP, Rao AK. Early growth response transcription factor EGR-1 regulates Galphaq gene in megakaryocytic cells. *J Thromb Haemost*. 2006;4:2678–86.

57. Moore SF, van den Bosch MTJ, Hunter RW, Sakamoto K, Poole AW, Hers I. Dual regulation of glycogen synthase kinase 3 (GSK3) α/β by protein kinase C (PKC) α and Akt promotes thrombin-mediated integrin $\alpha IIb\beta 3$ activation and granule secretion in platelets. *J Biol Chem*. 2013;288:3918–28.

58. Harper MT, Poole AW. Diverse functions of protein kinase C isoforms in platelet activation and thrombus formation. *J Thromb Haemost*. 2010;8:454–62.

59. Williams CM, Harper MT, Poole AW. PKC α negatively regulates in vitro proplatelet formation and in vivo platelet production in mice. *Platelets*. 2014;25:62–8.

60. Ault KA, Rinder HM, Mitchell J, Carmody MB, Vary CP, Hillman RS. The significance of platelets with increased RNA content (reticulated platelets). A measure of the rate of thrombopoiesis. *Am J Clin Pathol*. 1992;98:637–46.

61. Bö L, Dawson TM, Wesselingh S, Mörk S, Choi S, Kong PA, et al. Induction of nitric oxide synthase in demyelinating regions of multiple sclerosis brains. *Ann Neurol*. 1994;36:778–86.

62. Ludwin SK. The pathogenesis of multiple sclerosis: relating human pathology to experimental studies. *J Neuropathol Exp Neurol*. 2006;65:305–18.

63. Hecker M, Rüge A, Putscher E, Boxberger N, Rommer PS, Fitzner B, et al. Aberrant expression of alternative splicing variants in multiple sclerosis - A systematic review. *Autoimmun Rev*. 2019;18:721–32.

64. Elkjaer ML, Frisch T, Reynolds R, Kacprowski T, Burton M, Kruse TA, et al. Molecular signature of different lesion types in the brain white matter of patients with progressive multiple sclerosis. *Acta Neuropathol Commun*. 2019;7:205.

65. Gissel H. Ca $^{2+}$ accumulation and cell damage in skeletal muscle during low frequency stimulation. *Eur J Appl Physiol*. 2000;83:175–80.

66. Maléth J, Hegyi P. Ca $^{2+}$ toxicity and mitochondrial damage in acute pancreatitis: translational overview. *Philos Trans R Soc Lond B Biol Sci*. 2016;371. Available from: <http://dx.doi.org/10.1098/rstb.2015.0425>

67. Minagar A, Alexander JS. Blood-brain barrier disruption in multiple sclerosis. *Mult Scler*. 2003;9:540–9.
68. Claudio L, Raine CS, Brosnan CF. Evidence of persistent blood-brain barrier abnormalities in chronic-progressive multiple sclerosis. *Acta Neuropathol*. 1995;90:228–38.
69. Ortiz GG, Pacheco-Moisés FP, Macías-Islas MÁ, Flores-Alvarado LJ, Mireles-Ramírez MA, González-Renovato ED, et al. Role of the blood-brain barrier in multiple sclerosis. *Arch Med Res*. 2014;45:687–97.
70. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part I: the role of infection. *Ann Neurol*. Wiley; 2007;61:288–99.
71. Baranzini SE, Srinivasan R, Khankhanian P, Okuda DT, Nelson SJ, Matthews PM, et al. Genetic variation influences glutamate concentrations in brains of patients with multiple sclerosis. *Brain*. 2010;133:2603–11.
72. Strijbis EMM, Inkster B, Vounou M, Naegelin Y, Kappos L, Radue E-W, et al. Glutamate gene polymorphisms predict brain volumes in multiple sclerosis. *Mult Scler*. 2013;19:281–8.
73. Wang JH, Pappas D, De Jager PL, Pelletier D, de Bakker PI, Kappos L, et al. Modeling the cumulative genetic risk for multiple sclerosis from genome-wide association data. *Genome Med*. 2011;3:3.
74. Keshet Y, Seger R. The MAP kinase signaling cascades: a system of hundreds of components regulates a diverse array of physiological functions. *Methods Mol Biol*. 2010;661:3–38.
75. Chang L, Karin M. Mammalian MAP kinase signalling cascades. *Nature*. 2001;410:37–40.
76. Shchetytsky K, Protsyuk D, Ronninger M, Diaz-Gallo L-M, Klareskog L, Padyukov L. Gene-gene interaction and RNA splicing profiles of MAP2K4 gene in rheumatoid arthritis. *Clin Immunol*. 2015;158:19–28.
77. Tuller T, Atar S, Ruppin E, Gurevich M, Achiron A. Common and specific signatures of gene expression and protein-protein interactions in autoimmune diseases. *Genes Immun*. 2013;14:67–82.
78. ten Bosch GJA, Bolk J, 't Hart BA, Laman JD. Multiple sclerosis is linked to MAPK/ERK overactivity in microglia. *J Mol Med*. 2021; Available from: <https://doi.org/10.1007/s00109-021-02080-4>
79. Mass E, Jacome-Galarza CE, Blank T, Lazarov T, Durham BH, Ozkaya N, et al. A somatic mutation in erythro-myeloid progenitors causes neurodegenerative disease. *Nature*. 2017;549:389–93.
80. Kotelnikova E, Kiani NA, Messinis D, Pertsovskaya I, Pliaka V, Bernardo-Faura M, et al. MAPK pathway and B cells overactivation in multiple sclerosis revealed by phosphoproteomics and genomic analysis. *Proc Natl Acad Sci U S A*. 2019;116:9671–6.
81. Krentenov DN, Thornton TM, Teuscher C, Rincon M. The emerging role of p38 mitogen-activated protein kinase in multiple sclerosis and its models. *Mol Cell Biol*. 2013;33:3728–34.
82. Bernardo-Faura M, Rinas M, Wirbel J, Pertsovskaya I, Pliaka V, Messinis DE, et al. Prediction of combination therapies based on topological modeling of the immune signaling

network in Multiple Sclerosis. *bioRxiv*. 2019 [cited 2021 Jun 14]. p. 541458. Available from: <https://www.biorxiv.org/content/10.1101/541458v2>

83. Maatz H, Jens M, Liss M, Schafer S, Heinig M, Kirchner M, et al. RNA-binding protein RBM20 represses splicing to orchestrate cardiac pre-mRNA processing. *J Clin Invest*. 2014;124:3419–30.
84. Sheikh F, Lyon RC, Chen J. Functions of myosin light chain-2 (MYL2) in cardiac muscle and disease. *Gene*. 2015;569:14–20.
85. Matyushenko AM, Levitsky DI. Molecular Mechanisms of Pathologies of Skeletal and Cardiac Muscles Caused by Point Mutations in the Tropomyosin Genes. *Biochemistry*. 2020;85:S20–33.
86. Caleshu C, Sakhija R, Nussbaum RL, Schiller NB, Ursell PC, Eng C, et al. Furthering the link between the sarcomere and primary cardiomyopathies: restrictive cardiomyopathy associated with multiple mutations in genes previously associated with hypertrophic or dilated cardiomyopathy. *Am J Med Genet A*. 2011;155A:2229–35.
87. Brody MJ, Hacker TA, Patel JR, Feng L, Sadoshima J, Tevosian SG, et al. Ablation of the cardiac-specific gene leucine-rich repeat containing 10 (Lrrc10) results in dilated cardiomyopathy. *PLoS One*. 2012;7:e51621.
88. Gupte TM, Haque F, Gangadharan B, Sunitha MS, Mukherjee S, Anandhan S, et al. Mechanistic Heterogeneity in Contractile Properties of α -Tropomyosin (TPM1) Mutants Associated with Inherited Cardiomyopathies*. *J Biol Chem*. 2015;290:7003–15.
89. Huang W, Liang J, Yuan C-C, Kazmierczak K, Zhou Z, Morales A, et al. Novel familial dilated cardiomyopathy mutation in MYL2 affects the structure and function of myosin regulatory light chain. *FEBS J*. 2015;282:2379–93.
90. Herman DS, Lam L, Taylor MRG, Wang L, Teekakirikul P, Christodoulou D, et al. Truncations of titin causing dilated cardiomyopathy. *N Engl J Med*. 2012;366:619–28.
91. Marston S, Montgiraud C, Munster AB, Copeland O 'neal, Choi O, Dos Remedios C, et al. OBSCN Mutations Associated with Dilated Cardiomyopathy and Haploinsufficiency. *PLoS One*. 2015;10:e0138568.
92. Marston S. Obscurin variants and inherited cardiomyopathies. *Biophys Rev*. 2017;9:239–43.
93. McNally EM, Mestroni L. Dilated Cardiomyopathy: Genetic Determinants and Mechanisms. *Circ Res*. 2017;121:731–48.
94. Boczek NJ, Ye D, Jin F, Tester DJ, Huseby A, Bos JM, et al. Identification and Functional Characterization of a Novel CACNA1C-Mediated Cardiac Disorder Characterized by Prolonged QT Intervals With Hypertrophic Cardiomyopathy, Congenital Heart Defects, and Sudden Cardiac Death. *Circ Arrhythm Electrophysiol*. 2015;8:1122–32.
95. Mouton J, Ronjat M, Jona I, Villaz M, Feltz A, Maulet Y. Skeletal and cardiac ryanodine receptors bind to the $\text{Ca}^{(2+)}$ -sensor region of dihydropyridine receptor alpha(1C) subunit. *FEBS Lett*. 2001;505:441–4.
96. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28:27–30.
97. Ather S, Respress JL, Li N, Wehrens XHT. Alterations in ryanodine receptors and related

proteins in heart failure. *Biochim Biophys Acta*. 2013;1832:2425–31.

98. Yano M, Yamamoto T, Kobayashi S, Matsuzaki M. Role of ryanodine receptor as a Ca²⁺ regulatory center in normal and failing hearts. *J Cardiol*. 2009;53:1–7.

99. Moccia F, Lodola F, Stadiotti I, Pilato CA, Bellin M, Carugo S, et al. Calcium as a Key Player in Arrhythmogenic Cardiomyopathy: Adhesion Disorder or Intracellular Alteration? *Int J Mol Sci*. 2019;20. Available from: <http://dx.doi.org/10.3390/ijms20163986>

100. Jaffrey SR, Wilkinson MF. Nonsense-mediated RNA decay in the brain: emerging modulator of neural development and disease. *Nat Rev Neurosci*. 2018;19:715–28.

101. Schwerk C, Schulze-Osthoff K. Regulation of apoptosis by alternative pre-mRNA splicing. *Mol Cell*. 2005;19:1–13.

102. List M, Alcaraz N, Dissing-Hansen M, Ditzel HJ, Mollenhauer J, Baumbach J. KeyPathwayMinerWeb: online multi-omics network enrichment. *Nucleic Acids Res*. 2016;44:W98–104.

103. Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database*. 2011;2011:bar030.

104. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44:W90–7.

105. Signorelli M, Vinciotti V, Wit EC. NEAT: an efficient network enrichment analysis test. *BMC Bioinformatics*. 2016;17:352.

106. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1995. p. 289–300. Available from: <http://dx.doi.org/10.1111/j.2517-6161.1995.tb02031.x>

107. Irimia M, Weatheritt RJ, Ellis JD, Parikshak NN, Gonatopoulos-Pournatzis T, Babor M, et al. A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell*. 2014;159:1511–23.

108. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.

109. Van Der Walt S, Chris Colbert S, Varoquaux G. The NumPy array: a structure for efficient numerical computation. *arXiv [cs.MS]*. 2011. Available from: <http://arxiv.org/abs/1102.1523>

110. McKinney W, Others. Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*. Austin, TX; 2010. p. 51–6.

111. Hagberg A, Swart P, S Chult D. Exploring network structure, dynamics, and function using networkx. Los Alamos National Lab. (LANL), Los Alamos, NM (United States); 2008 Jan. Report No.: LA-UR-08-05495; LA-UR-08-5495. Available from: <https://www.osti.gov/biblio/960616>

112. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020;17:261–72.

113. Seabold S, Perktold J. Statsmodels: Econometric and statistical modeling with python. Proceedings of the 9th Python in Science Conference. Austin, TX; 2010. p. 61.
114. Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. Software: Practice and Experience. 1991. p. 1129–64. Available from: <http://dx.doi.org/10.1002/spe.4380211102>