OXFORD

Systems biology

Systematic analysis of alternative splicing in time course data using Spycone

Chit Tong Lio [1,2], Gordon Grabert[3,4], Zakaria Louadi[1,2], Amit Fenn[1,2], Jan Baumbach [1,5], Tim Kacprowski [3,4], Markus List [2,†] and Olga Tsoy[1,*,†]

[1]Institute for Computational Systems Biology, University of Hamburg, Notkestrasse 9, Hamburg 22607, Germany, [2]Chair of Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich, Freising 85354, Germany, [3]Division Data Science in Biomedicine, Peter L. Reichertz Institute for Medical Informatics of Technische Universität Braunschweig and Hannover Medical School, Braunschweig 38106, Germany, [4]Braunschweig Integrated Centre of Systems Biology (BRICS), TU Braunschweig, Braunschweig 38106, Germany and [5]Institute of Mathematics and Computer Science, University of Southern Denmark, Odense 5000, Denmark

*To whom correspondence should be addressed.

[†]The authors wish it to be known that, in their opinion, the last two authors should be regarded as Joint Last Authors.

Associate Editor: Valentina Boeva

Abstract

Motivation: During disease progression or organism development, alternative splicing may lead to isoform switches that demonstrate similar temporal patterns and reflect the alternative splicing co-regulation of such genes. Tools for dynamic process analysis usually neglect alternative splicing.

Results: Here, we propose Spycone, a splicing-aware framework for time course data analysis. Spycone exploits a novel IS detection algorithm and offers downstream analysis such as network and gene set enrichment. We demonstrate the performance of Spycone using simulated and real-world data of SARS-CoV-2 infection.

Availability and implementation: The Spycone package is available as a PyPI package. The source code of Spycone is available under the GPLv3 license at https://github.com/yollct/spycone and the documentation at https://spycone.readthedocs.io/en/latest/.

Contact: olga.tsoy@uni-hamburg.de

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Changes in alternative splicing (AS) lead to a differential abundance of gene isoforms between experimental conditions or time points. If the relative abundance of two isoforms of a gene changes between two conditions or time points, this behavior is called isoform switching (IS). While differential isoform expression focus on the change in the expression value of one isoform, IS detects switches of predominantly expressed isoforms between conditions. A change of the predominant isoform appears as an intersection in time course data. However, existing methods for time course change points detection are applied to detect abrupt change between states while IS events are usually slow and gradual changes of isoform expression (Aminikhanghahi and Cook, 2017). IS has a functional impact on the gene when the two switching isoforms perform different functions or when they have different interaction partners. Vitting-Seerup and Sandelin (2017) showed that IS changes the functions of 19% ($N = 2352$) of genes with multiple isoforms in cancer, most of them leading to a protein domain loss. In cardiovascular disease, the IS of Titin causes clinical symptoms of dilated cardiomyopathy (Makarenko et al., 2004). Therefore, the detection and functional interpretation of IS events is a promising strategy to reveal the mechanism of disease development.

However, the above examples refer to molecular snapshots of dynamic processes. In order to study such dynamic processes, like disease progression, we need time course data. By identifying groups of genes with similar temporal expression or AS/IS patterns, we can dissect the disease progression into mechanistic details. A study of mouse retinal development has shown that genes with similar temporal exon usage patterns shared similar biological functions and cell type specificity (Wan et al., 2011). However, existing tools for AS analysis mostly focus on a single condition or two conditions from snapshot experiments. Tools developed for time course data analysis, e.g. TiCoNE (Wiwie et al., 2019), moanin (Varoquaux and Purdom, 2020), TimesVector-web (Jang et al., 2021) focus on gene expression level and neglect splicing. Thus, systematic time course AS analysis is usually done manually. Common approaches are

semi-automatic clustering of temporal patterns of percent-spliced-in (PSI) value (Trincado *et al.* 2017; Xing *et al.* 2020) or differential splicing analysis between pairs of time points (Hooper *et al.* 2020). PSI values indicate the fraction of transcripts carrying an AS event and thus do not directly reflect isoform switches which are crucial for interpreting functional consequences of AS. Iso-MaSigPro uses a generalized linear model to detect differential expression changes along time courses between two experimental groups (Nueda *et al.*, 2018). However, Iso-MaSigPro is limited in time series data with two conditions and it does not provide information like switching points. TSIS, the only available tool to perform AS time course analysis in one condition, detects IS events whose effect lasts across several time points (Guo *et al.*, 2017). However, TSIS treats all IS events similarly, independent of their expression level. As a result, TSIS emphasizes isoforms with low expression while isoforms with comparably high expression levels are expected to be more involved in biological processes.

We introduce Spycone, a splicing-aware framework for systematic time course transcriptomics data analysis. It employs a novel IS detection method that prioritizes isoform switches between highly expressed isoforms over those with minor expression levels, thus focusing on biologically relevant changes rather than transcriptional noise. Spycone operates on both gene and isoform levels. For isoform-level data, the total isoform usage is quantified across time points. We have incorporated clustering methods for grouping genes and isoforms with similar time course patterns, as well as network and gene set enrichment methods for the functional interpretation of clusters. The IS genes within the same clusters are expected to interact cooperatively with other functionally related genes. Thus, we hypothesize that disease mechanisms or developmental changes can be identified with network and functional enrichment methods. We compare the performance of Spycone and TSIS on a simulated and real-world dataset. On the latter, we demonstrate how Spycone identifies network modules that are potentially affected by alternatively spliced genes during SARS-CoV-2 infection.

2 Materials and methods

2.1 Data preprocessing

We demonstrated the performance of Spycone on RNA-seq data from SARS-CoV-2 infected human lung cells (Calu-3) with eight time points and four replicates for each time point (de la Fuente *et al.*, 2020).

For the SARS-CoV-2 dataset, we used Trimmomatic v0.39 (Bolger *et al.*, 2014) to remove Illumina adapter sequences and low-quality bases (Phred score < 30) followed by Salmon v1.5.1 (Patro *et al.*, 2017) for isoform quantification with a mapping-based model, the human genome version 38 and an Ensembl genome annotation version 104.

2.2 Protein–protein interaction network and domain–domain interaction

A protein–protein interaction (PPI) network is obtained from BioGRID (v.4.4.208) (Oughtred *et al.*, 2021) and a domain–domain interaction network from 3did (v2019_01) (Mosca *et al.*, 2014). The edges of the PPI network are weighted according to the number of interactions found between the domains of the protein (nodes of PPI), given by the domain–domain interaction. Weighting PPIs with domain-based information can result in a functionally more interpretable network in diseases and pathways (Shim and Lee, 2016).

2.3 Simulation

We used the SARS-CoV-2 dataset described above as a reference for setting the parameters of a negative binomial distribution of gene expression counts, as well as the parameters of the Poisson distribution of the number of isoforms for each gene.

2.3.1 First-order Markov chain

A first-order Markov chain is used for the simulation of the gene states at each time point. In the simplest form, we defined two gene states: switched or unswitched. Change of the states along the time course depends on the transition probabilities.

We used a Dirichlet distribution to simulate relative abundance for each isoform of a gene. The relative abundance of an isoform is the ratio of the isoform expression to the total gene expression. The outcome of the Dirichlet distribution is *k*-dimensional vectors *x* with real numbers between 0 and 1 such that the sum of the elements in *x* is 1. This is suitable to simulate probability distribution of *k* categories. The Dirichlet distribution is defined as:

$$f(M_t|\alpha_0, \alpha_1, \ldots, \alpha_k) = \frac{1}{\text{beta}(\alpha)} \prod_{i=1}^{k} M_i^{\alpha_i - 1}, \tag{1}$$

$$\text{beta}(\alpha) = \frac{\prod_{i=1}^{k} \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{k} \alpha_i)}, \tag{2}$$

where the parameter is a *k*-dimensional vector governing the distribution of the probabilities. In our case, *k* equals the number of isoforms in a gene, where each isoform will be assigned an *i* value. The higher *i*, the higher the probability of the isoform *i*.

In Model 1, where we assumed that switching isoforms are highly expressed, the α for switching isoforms are $\alpha = \{1, 2, \ldots, s\}*10$, *s* is the number of switching isoforms, while α for the remaining isoforms are 1. To introduce switching events, the isoform probabilities of two highly expressed isoforms are swapped. For instance, if the isoform probabilities of the unswitched state for gene *g* with five isoforms are {0.03, 0.07, 0.1, 0.3, 0.5}. Then, the isoform probabilities for the switched state is {0.03, 0.07, 0.1, 0.5, 0.3}.

In Model 2, where we assumed that isoforms with abundance higher than 0.3 have equal chances to switch, the vector is $\alpha = \{1, 2, \ldots, k\}*10$, *k* is the number of all isoforms. To introduce switching events, the probabilities of two random isoforms will be swapped.

After we simulated abundances for each isoform, we multiplied it to a gene expression mean selected based on real-life dataset to obtain the transcript expression mean (μ_i). The gene expression means are randomly picked among the genes with the same number of isoforms from the real-world dataset.

We simulated time course data with 10 time points and 3 replicates using 10 000 genes. The transcript expression of replicates is sampled from normal distribution with a given transcript mean (μ_i), and the variance is sampled from a gamma distribution as the following:

$$\theta_i \sim \text{gamma}\left(a = \frac{\mu_i + \text{noise}}{\mu_i * \text{noise}}\right) \tag{3}$$

$$\text{Count}t_i \sim \text{normal}(\mu_i, \theta_i). \tag{4}$$

In order to simulate the differences generated for each individual experiment in real life, we tested on noise levels 1, 5 and 10. This setting can ensure isoforms with higher abundance will have a higher variance compared to those with lower abundance. The simulated dataset can be downloaded from doi: 10.5281/zenodo.7228475. The code for generating the benchmarking figures is stored in https://github.com/yollct/spycone_benchmark.

The data were analyzed with Spycone's detect_isoform_switch function and TSIS's iso.switch function. TSIS was further tested in two modes—TSIS and major_TSIS—major_TSIS uses the max.ratio = TRUE parameter.

2.4 Detection of isoform switch events

2.4.1 Spycone

The first step of IS detection is to filter out transcripts that have an average transcripts per million (TPM) <1 over all time points.

Spycone then detects IS events based on the relative abundance of the isoforms. The IS events are defined with the following metrics:

2.4.1.1 Switching points. Switch points refer to the points where two time courses intersect in at least 60% of the replicates. For every pair of isoforms in a gene, Spycone detects all possible switch points for further analysis. For a dataset that has only one replicate, Spycone checks the intersection between isoform pairs in one replicate.

2.4.1.2 Switching probability. As TSIS, Spycone calculates a switching probability for each IS event. A switching probability is the average of the ratio of samples where the relative abundance I of isoform i is higher than isoform j before switch (T_1), and vice versa, the ratio of samples where the relative abundance I isoform i is lower than of isoform j after switch (T_2). If two isoforms switched between time interval T_1 and T_2 the switching probability between isoform i and isoform j is:

$$P(\text{switch}) = [P_{T_1}(\mathrm{I}_{i,t} > \mathrm{I}_{j,t}) + P_{T_2}(\mathrm{I}_{i,t} < \mathrm{I}_{j,t})]/2, \qquad (5)$$

where P denotes the frequency of the respective condition between relative abundance (I) of isoform i and j at each time point t within the time interval T.

2.4.1.3 Significance of switch points. If replicates are available, Spycone calculates the significance of a switch point by performing a two-sided Mann–Whitney U-test between relative abundance before and after the switch point similar to TSIS. For a dataset that has only one replicate, a permutation test is performed, where the time points within a time course are permuted. An empirical P-value is calculated to indicate the probability for the two switching isoforms to have a higher dissimilarity coefficient and higher difference of relative abundance before and after switch. Since the goal here is to select genes that have significant IS, Spycone takes the best switch point for further analysis with the smallest P-value. Other significant switch points will be reported as part of the result for users to investigate.

2.4.1.4 Difference of relative abundance. To quantify the magnitude of changes during IS, Spycone calculates the average difference of relative abundance before and after a switch point. If replicates are available, Spycone calculates the average change of relative abundance. We selected a cutoff of 0.1, where the changes in the relative abundance accounts for at least 10% of the total gene expression. Difference of relative abundance I between switching isoforms i and j is defined as:

$$\text{Diff}_{i,j,s} = \left[\sum_{r=1}^{R} (I^r_{i,s+1} - I^r_{i,s})/R + \sum_{r=1}^{R} (I^r_{j,s+1} - I^r_{j,s})/R \right]/2, \quad (6)$$

where s is a switch point of Isoform i and j; R is the number of replicates.

2.4.1.5 Event importance. Event importance is a novel metric that accounts for the expression level of switching isoforms. We defined event importance of a switch occurs between time point t and $t+1$ as:

$$\text{Event importance} = \sum_{r=1}^{R} \left[\left(\frac{I^r_{aGt}}{\max(I^r_{Gt})} + \frac{I^r_{aGt+1}}{\max(I^r_{Gt+1})} + \frac{I^r_{bGt}}{\max(I^r_{Gt})} + \frac{I^r_{bGt+1}}{\max(I^r_{Gt+1})} \right)/4 \right]/R, \tag{7}$$

where I^r_{aGt} is the relative abundance of isoform a of a gene G at time point t; and R is the total number of replicates. Each I is normalized to the highest relative abundance $\max(I^r_{Gt})$ at the corresponding time point. The metric takes the average of the relative abundance of isoforms i and j before and after switch.

For the analysis, we used IS events with differences of relative abundance higher than 0.2 and event importance higher than 0.3.

2.1.4.6 Dissimilarity coefficient. Dissimilarity coefficients $d_{i,j}$ assess the dissimilarity of the time course between isoforms. It is calculated based on the Pearson correlation $r_{i,j}$ between time course I and J:

$$r_{i,j} = \frac{\text{cov}(I,J)}{\sigma_i, \sigma_j} \tag{8}$$

$$d = \frac{1-r}{2}. \tag{9}$$

The higher coefficient, the less similar are the time courses.

2.4.1.7 Domain inclusion or exclusion. We used the Pfam database v.35.0 (Mistry *et al.*, 2021) to map domains to isoforms. Spycone compares isoforms in the IS event with each other to define if there is a loss/gain of domain.

2.4.1.8 Multiple testing correction. Finally, we implemented multiple testing corrections for IS detection. Available corrections are Bonferroni, Holm–Bonferroni and Benjamini–Hochberg false discovery rate. We use the Benjamini–Hochberg method as default.

2.4.2 TSIS

To detect IS in TSIS, we used the following parameters: (i) the switching probability > 0.5; (ii) difference before and after switch > 10; (iii) interval lasting before and after at minimum one time point; (iv) P-value < 0.05 and (v) Pearson correlation < 0. More detailed descriptions of parameters are found in Guo *et al.* (2017). The above parameters are set with defaults suggested by TSIS, except parameter (iii), since we have a larger interval between time points (12 h at maximum).

2.5 Change of total isoform usage

Isoform usage measures the relative abundance of an isoform. Isoform usage of all isoforms from one gene are summed up to obtain the total isoform usage. We defined the change of total isoform usage as between two consecutive time points:

$$\Delta \text{total isoform usage} = \sum_{A=0}^{n} \left| \left(\frac{I_{AGt1}}{\sum_{A=0}^{n}(I_{AGt1})} - \frac{I_{AGt0}}{\sum_{A=0}^{n}(I_{AGt0})} \right) \right|, \tag{10}$$

where I is the expression of isoform A of gene G at time points t_1 and t_0; and n is the total number of all isoforms for gene G.

2.6 Clustering analysis

The clustering algorithms are implemented using the scikit-learn machine learning package in python (v0.23.2) (Pedregosa *et al.*, 2011) and tslearn (v0.5.1.0) time course machine learning package in python (Tavenard *et al.*, 2020). The available algorithms are K-means, K-medoids, agglomerative clustering, DBSCAN and OPTICS.

The number of clusters is chosen manually by visually checking the Ward distance dendrogram (Supplementary Fig. S6).

2.7 Gene set enrichment and network analysis

For enrichment analysis, Spycone uses g:Profiler and NEASE. g:Profiler is a functional enrichment toolkit for GO terms and pathways (Raudvere *et al.*, 2019). Gene set enrichment method is performed using Fisher's exact test. NEASE (Louadi *et al.*, 2021) is an enrichment method for co-regulated alternative exons. We used NEASE with KEGG and Reactome pathways. For a seamless analysis, the newest version of the NEASE's Python package (v1.1.9) is integrated with Spycone.

Spycone employs DOMINO (0.1.0) (Levi *et al.*, 2021) for active module identification in PPI networks using default parameters.

2.8 Splicing factor co-expression and motif enrichment analysis

List of splicing factors and their position-specific scoring matrices (PSSMs) are obtained from the mCross database (downloaded in 2022), currently only available for *Homo sapiens* (Feng *et al.*, 2019). First, we filtered splicing factors with TPM > 1 in all time points. Next, we calculated the correlation between the relative abundance of each isoform and the expression of splicing factors. We filtered the pairs with correlation >0.7 or <−0.7 and adjusted *P*-value <0.05.

Finally, we performed motif enrichment analysis using the motifs module from the Biopython library (Cock *et al.*, 2009). The motifs module computes the log-odd probability of a specific region in the genome to match the binding motif using the PSSM (Henikoff and Henikoff, 1996). Hence, the higher the log-odd score, the more likely the binding. We compared these scores obtained from the lost, gained and unregulated exons from the same clusters. A Mann–Whitney *U*-test is performed on the sets of scores. Each motif threshold is selected using the distribution of the PSSM score over the frequency of nucleotides (background). The threshold is set at a false positive rate <0.01, meaning the probability of finding the motif in the background is <0.01.

3 Results

3.1 Spycone overview

Spycone is available as a python package that provides systematic analysis of time course transcriptomics data. Figure 1 shows the workflow of Spycone. It uses gene or isoform expression and a biological network as an input. It employs the sum of changes of all isoforms relative abundance (total isoform usage) (de la Fuente *et al.*, 2020) (see Section 2), i.e. the sums of pairwise changes in relative isoform abundance, across time points to detect IS events. It further

provides downstream analysis such as clustering by total isoform usage, gene set enrichment analysis, network enrichment and splicing factors analysis. Visualization functions are provided for IS events, cluster prototypes, network modules and gene set enrichment results.

IS detection. We propose novel metrics for the detection and selection of significant IS across time. IS events are described as a change of the isoform distribution between two conditions (time points). To detect an IS, our algorithm first searches for switch points, i.e. a specific time point where two isoform expression time courses intersect.

The main challenges to detect time course IS are: (i) most genes have multiple isoforms, the changes of the relative abundance can be due to factors other than AS, e.g. RNA degradation. (ii) Most IS have multiple switch points, with different magnitudes of change in abundance; we need to consider how prominent the changes in abundance are to be recognized as an IS event. (iii) Most genes have multiple lowly expressed isoforms that constitute noise and might not be biologically relevant. An ideal IS detection tool, therefore, should prioritize IS events according to their expression level (Supplementary Fig. S1).

Spycone overcomes these challenges by using a novel approach to detect IS events. Spycone uses two metrics: a *P*-value and event importance. The *P*-value is calculated by performing a two-sided Mann–Whitney *U*-test between relative abundance before and after the switch point among the replicates. Event importance is the average of the ratio of the relative abundance of the two switching isoforms to the relative abundance of the isoform with the highest expression between the switching time points (see Section 2). Examples of high and low event importance are illustrated in Figure 2. The event importance will be highest when an IS includes the highest expressed isoform. Similarly, event importance will be low if an IS occurs between two lowly expressed isoforms. We also provide different metrics to comprehensively assess features of the IS events including switching probability, difference of abundance before and after switching and a dissimilarity coefficient (see Section 2).

Clustering analysis for identifying co-spliced genes. Similar to how transcription factors co-regulate sets of genes, in the context of AS, the splicing events of a subset of genes are co-regulated by splicing factors (Barberan-Soler *et al.*, 2011). For genes with important IS events (identified as described above), we want to quantify the impact of splicing regulation between two time points. To this end, Spycone clusters genes by changes in total isoform usage over time to identify co-spliced genes. A previous study showed that clustering performance is highly dependent on the dataset and the clustering method (Javed *et al.*, 2020). Therefore, Spycone offers various clustering techniques, including agglomerative clustering (hierarchical clustering) (Johnson, 1967), K-Means clustering (Hartigan and Wong, 1979), K-Medoids clustering (Park and Jun, 2009), DBSCAN (Ester *et al.*, 1996), OPTICS (Ankerst *et al.*, 1999) and various distance metrics such as euclidean distance, Pearson distance, as well as tslearn (Tavenard *et al.*, 2020) for calculating the dynamic time warping distance measure.

With temporal patterns of the clusters, Spycone dissects context-specific processes in terms of AS. In order to gain functional knowledge of the clusters, Spycone offers g:Profiler (Raudvere *et al.*,

Fig. 1. Overview of the Spycone architecture. Spycone takes count matrices and biological networks as input. We provide isoform-level functions such as isoform switch detection and total isoform usage calculation. Users could also cluster the gene count matrix directly. For downstream analysis, we integrated multiple clustering algorithms and an active modules identification algorithm (DOMINO). We also implemented splicing factors analysis for isoform-level data. Finally, visualizations are provided to better evaluate and interpret the results

Fig. 2. Plots showing the examples of three levels of event importance. Each plot contains all isoforms of a gene. The circle indicates the IS events with the corresponding level of event importance

2019) and NEASE (Louadi *et al.*, 2021) for gene set enrichment analysis. The former conducts classical enrichment analysis for multiple ontologies and pathway databases. The latter combines information from PPI and domain–domain interaction networks and allows to predict functional consequences of AS events caused by a set of IS genes.

Active modules identification. Genes with consistent temporal patterns are thought to be functionally related in terms of co-regulation, molecular interactions or participation in the same cellular processes. To uncover the underlying mechanism that is represented by a temporal pattern, Spycone projects the results of the clustering analysis on a molecular interaction network for active modules identification, i.e. for detection of subnetworks enriched in genes affected by IS. We incorporated DOMINO (Levi *et al.*, 2021) as it has been previously demonstrated the best performance for this task (Lazareva *et al.*, 2021). To elucidate the functional impact of IS events, we further leveraged domain–domain interaction information from the 3did database (Mosca *et al.*, 2014). Spycone identifies domains lost/gained during IS, which might indicate a functional switch, and affected edges in the PPI network. This provides additional insights about the functional consequences of time course IS.

Splicing factor analysis. Spycone also provides splicing factor analysis using co-expression and RNA-binding protein motif search. Splicing factors are a group of RNA-binding proteins that regulate the splicing of genes. We assume that the expression of splicing factors that are responsible for an IS event correlates with the relative abundance of participating isoforms. Spycone calculates the correlation between the expression value of a list of RNA-binding proteins derived from ENCODE eCLIP data (Feng *et al.*, 2019) and the relative abundance of isoforms involved in IS. We implemented PSSM of RNA-binding protein motifs to calculate and detect the potential binding sites along the sequence of the targeted isoforms (see Section 2).

3.2 Evaluation using simulated data

To evaluate the performance of Spycone, we compared its performance (precision and recall) to TSIS using simulated data. TSIS provides an option to filter for IS events that involve only the highest abundance isoform—we refer to the result after filtering as major_TSIS. We aimed to investigate whether the performance of TSIS improves when applying this option.

We use a hidden Markov model to simulate the switching state of the genes at each time point (see Section 2). We simulated two models (Supplementary Fig. S2): Model 1 allows only major isoforms, i.e. those with the highest abundance per gene, to be involved in IS events across time points; Model 2 allows IS to occur between isoforms with relative abundance higher than 0.3. We used Model 2 to show that neither tool is biased towards events that involve only major isoforms.

For both tools, we varied their parameters (difference of relative abundance), to investigate how this affects their precision and recall. We also considered different levels of variance of gene expression, namely 1, 5 and 10, across replicates to mimic the noise (Fig. 3).

In Model 1, Spycone achieved high precision and recall. The precision of TSIS dropped drastically with increasing recall. After filtering major events, TSIS's recall reached 0.5. Spycone performs better in the setting with the highest noise level as it maintains high precision (0.95) and acceptable recall (0.75). In Model 2, Spycone achieved higher precision and recall than TSIS; however, they dropped as the model allows more IS events. We applied spline regression to detect switch points and calculated precision and recall as above (Supplementary Fig. S3). Results showed that spline regression does not improve precision and recall in both tools. Moreover, TSIS has a higher algorithmic complexity of $O(n*\log(n))$ than Spycone with a complexity of $O(n)$, leading to a drastically lower runtime for Spycone in the range of a few minutes rather than hours (Supplementary Fig. S4). In summary, Spycone outperforms TSIS in detecting IS events.
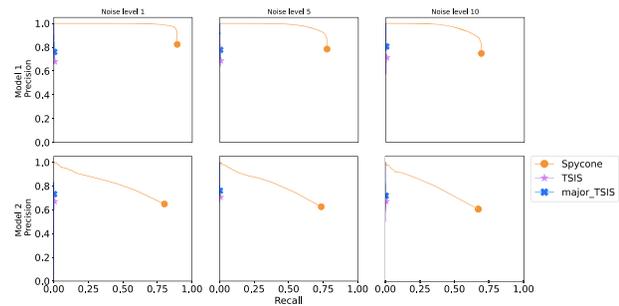
Fig. 3. Precision and recall curves for Spycone, TSIS and major-isoforms filtered TSIS from simulated data of two models (rows) and three noise levels (columns)

3.3 Application to SARS-Cov2 infection data

We applied Spycone to an RNA-seq time course dataset of SARS-CoV-2-infected human lung cells (Kim *et al.*, 2021). The dataset contains eight time points: 0, 1, 2, 4, 12, 16, 24 and 36 h post-infection. We kept isoforms with TPM > 1 across all time points resulting in 36 062 isoforms for IS event detection with Spycone and TSIS. To call an IS significant, we used the following criteria: for Spycone, (i) switching probability > 0.5; (ii) difference of relative abundance > 0.2 before and after the switch; (iii) dissimilarity coefficient > 0.5; and (iv) adjusted *P*-value < 0.05. For TSIS, we used (i) switching probability > 0.5; (ii) difference of expression before and after switch > 10; (iii) correlation coefficient < 0; and (iv) adjusted *P*-value < 0.05. The dissimilarity coefficient from Spycone and the correlation coefficient from TSIS are used to filter for IS events with negatively correlated isoforms. The values are chosen according to the performance on Model 2 simulated data with noise level 10 that showed the best precision. Spycone reported 915 IS events, of which 418 affected at least 1 protein domain. TSIS reported 985 events, of which 417 affected at least one protein domain. On gene level, Spycone reported 745 genes with IS events, TSIS reported 858 genes where 225 genes were found by both Spycone and TSIS (Fig. 4A).

We then used the event importance metric to assess the ability of each method to detect IS events from higher abundance isoforms. We calculated event importance for IS events identified by Spycone, TSIS and major_TSIS (Fig. 4B). Spycone results include mostly events with high importance, while in TSIS events with low importance prevail. Supplementary Table S1 shows the result for the SARS-CoV-2 dataset from the Spycone IS detection. Event importance has no clear prevalence towards overall gene expression and adjusted *P*-value (Supplementary Figs S5 and S6).

To exclude IS events with lowly expressed isoforms, we applied a filter of event importance higher than 0.3 to both Spycone and TSIS results. We calculated the change of total isoform usage of the IS genes across time points and employed Ward linkage hierarchical clustering. This led to four clusters with similar temporal patterns of changes in total isoform usage for Spycone (Fig. 4C, Supplementary Fig. S7, Supplementary Table S2) and four clusters for TSIS (Supplementary Fig. S10A). Each cluster is represented by a cluster prototype, which is the median change of total isoform usage per pair of time points.

IS events that lead to domain gain or loss might break the interactions, hence rewiring the PPI network. Moreover, if the IS events belong to the same cluster, it indicates the synchronized gain or loss of interactions with particular pathways. Our goal is therefore to assess if IS events within clusters rewire interactions with particular pathways during SARS-CoV-2 infection. We performed AS-aware pathway enrichment analysis using NEASE with KEGG (Kanehisa *et al.*, 2016) and Reactome (Jassal *et al.*, 2020) pathway databases for results from Spycone (Supplementary Fig. S8, Supplementary Table S3) and TSIS (Supplementary Table S4, Supplementary Fig S10B). In addition, we performed classical gene set enrichment analysis using g:Profiler. The results are not informative since only five terms are found in Cluster 3 and zero in others.

Overall, clusters with similar prototypes from both tools are enriched in distinct pathway terms. For example, TSIS's Cluster 1

A

Genes with isoform switch events

B

C
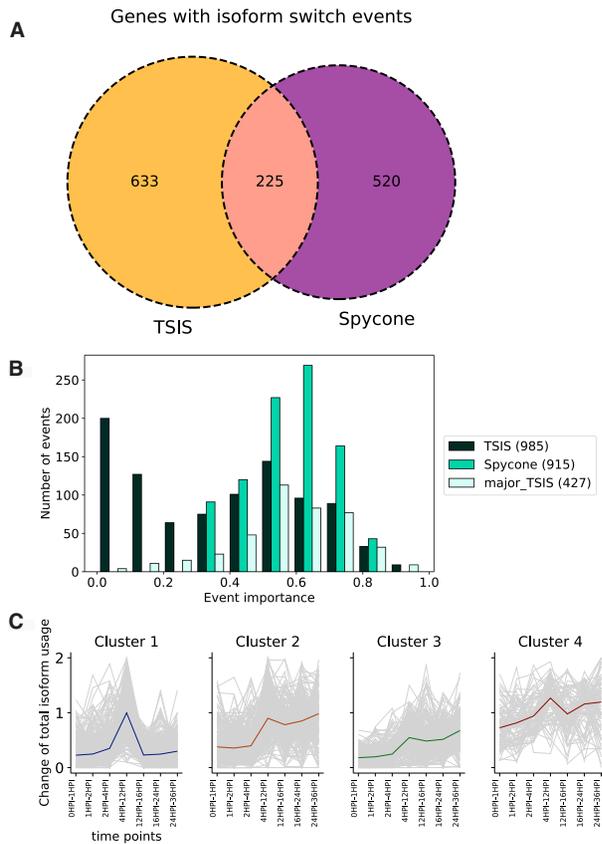
Fig. 4. Comparing IS detection results from Spycone and TSIS (**A**) Venn diagram showing the number of genes detected with isoform switch events by Spycone and TSIS (all). (**B**) The distribution of the detected events in Spycone, TSIS and major_TSIS based on the event importance metric of Spycone. The number of events for each tool is indicated in brackets in the legend. (**C**) Cluster prototypes and all objects show the pattern of the change of total isoform usage across time points

and Spycone's Cluster 1 have a strong peak between 4 and 12 h post-infection. Only transforming growth factor (TGF)-beta signaling is commonly found in both tools. MAPK pathway and DNA damage checkpoint are enriched uniquely in Spycone. TSIS's Cluster 2 and Spycone's Cluster 3 have lower changes of total isoform usage overall. Spycone's clusters showed more unique and relevant terms: 70 enriched Reactome terms in Spycone's clusters and only 7 terms in TSIS's clusters. TSIS's Cluster 3 and Spycone Cluster 2 show an increase of change of total isoform usage after 12 h post-infection. Spycone's cluster is enriched uniquely in protein folding chaperonin complex TriC/CCT and NOTCH signaling pathway. Finally, TSIS's Cluster 4 and Spycone's Cluster 4 have increasing changes of total isoform usage overall. TSIS's cluster is enriched in mitosis-related pathways, cell cycle and tubulin folding. Whereas in Spycone's Cluster 4 is found with signaling by PTK6, interferon, metabolism of proteins, pentose phosphate pathway, etc.

Next, we detected active modules that show over-representation of IS genes from the same cluster based on DOMINO using a PPI network from BioGRID (Oughtred *et al.*, 2021) (see Section 2). Detected active modules suggest the impact of splicing on regulatory cascades and cellular trafficking (Table 1, Fig. 5, Supplementary Fig. S7).

3.3.1 Splicing factor Anaysis
Assuming that multiple IS events occurring between the same time points are co-regulated by the same splicing factor, we perform co-expression and motif analysis. The co-expression analysis yields thirteen significant RNA-binding proteins that are positively or negatively correlated with at least two isoforms of the same gene: in cluster 1 - FUBP3, HLTF, IGF2BP3, ILF3, RBFOX2, RBM22, SF3B1 and TAF15; in cluster 3 - IGF2BP3, RBM22, RPS6, SRSF7 and SUGP2; ($|r| > 0.6$ and adjusted p-value < 0.05) (Table S5). To investigate whether the regulated exons, i.e. the lost or gained exons after IS events, show higher PSSM scores to a certain RNA-binding protein motif than the unregulated exons in a cluster, we applied motif enrichment analysis. We calculated PSSM scores along the flanking regions of the exons 5' and 3' boundaries and excluded the first and last exons in an isoform since these are often regulated by 5'-cap binding proteins and polyadenylation regulating proteins (Zheng, 2004). All exons in the switched isoforms within a cluster are categorized to 1) lost exons, 2) gained exons, and 3) unregulated exons for the analysis (Fig.6, Table S6). RNA-binding proteins with multiple motifs are numbered with an underscore. Each motif is selected with a threshold where the false-positive rate is below 0.01. Position-specific log-odd scores higher than the corresponding threshold are obtained after calculating the PSSM scores of each motif for all exons (see Methods section). The ILF3_9 and ILF3_14 motifs show higher log-odd scores at the 5' end of the lost/gain exons than of the unregulated exons in cluster 1 (one-sided Mann-Whitney U test p-value < 0.05) (Fig. 6A). HLTF_7 and SRSF7_1 motifs show higher log-odd scores at the 3' end (Fig. 6B).

4 Discussion
AS regulates dynamic processes such as development and disease progression. However, AS analysis tools typically compare only two conditions and neglect how AS changes dynamically over time. Currently, the only existing tool for time course data analysis that accounts for splicing is TSIS. TSIS detects temporal IS events but is biased towards IS events between lowly expressed isoforms and does not offer features for downstream analysis which is important for interpreting the functional consequences of IS events.

Spycone, a framework for analysis of time course transcriptomics data, features a new approach for detecting temporal IS events and a new event importance metric to filter out lowly expressed isoforms. We demonstrate that Spycone's IS detection method outperforms TSIS in terms of precision and recall based on simulated data. A key advantage of Spycone is that it explicitly considers how well IS events agree across replicates while TSIS considers averaged expression values among replicates and/or by natural spline-curves fitting. More specifically, Spycone uses a non-parametric Mann–Whitney U-test to test for significant IS and performs multiple testing correction to reduce type I error.

We have demonstrated the usability of Spycone by analyzing time course transcriptomics data for SARS-CoV-2 infection where we found affected signaling cascades. We performed NEASE enrichment on the clusters and compared the results from Spycone and TSIS. Spycone results are enriched in relevant terms such as mitogen-activated protein kinase (MAPK) pathway (Cluster 1), NOTCH signaling (Cluster 2), fibroblast growth factor receptor (FGFRs) and toll-like receptor (TLR) pathways (Cluster 3) and pentose phosphate pathway (Cluster 4). NOTCH signaling pathways are found upregulated in the lungs of infected macaques (Rosa *et al.*, 2021).

The MAPK pathway has a pro-inflammatory effect by interacting with SARS-CoV-2 downstream pathogenesis, especially in patients suffering from cardiovascular disease (Weckbach *et al.*, 2022). TLR 7/8 cascades are related to ssRNA, and there is a study supporting the association of TLR 7/8 with SARS-CoV-2 infection (Salvi *et al.*, 2021). The pentose phosphate pathway is an alternative pathway of glycolysis that produces more reduced NADP (NADPH) oxidase. It is activated during SARS-CoV-2 infection in response to oxidative stress and the activation of the immune response (Yang *et al.*, 2021). Spycone also detected the enrichment of pathways which association with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection has not been characterized yet: kinesins, signaling by NTRKs, degradation of AXIN, signaling by Hedgehog and 5-phosphoribose 1-diphosphate biosynthesis.

The active modules extracted from the clusters highlight mechanisms involved in the host cell response to infection. In Cluster 2

Table 1. Related biological processes and pathways of the respective modules found in clusters

Clusters	Module	Related biological processes
1 (Fig. 5, Supplementary Fig. S9A)	1	RNA splicing and mRNA processing
	2	Cellular protein modification genes, signal transduction in the VEGF signaling pathway
	3	Positive regulation of protein ubiquitination
	4	Protein and intracellular trafficking genes
	5	Protein and intracellular trafficking genes
2 (Supplementary Fig. S9B)	1	Transcription and mRNA splicing
	2	Ras and Rho protein signaling transduction
	3	Ubiquitination
	4	Protein import into the nucleus
	5	Transition of cell cycle to G2/M phase
3 (Supplementary Fig. S9C)	1	Regulation of transcription, cell cycle arrest and protein catabolic process
	2	Transmembrane receptor protein tyrosine kinase signaling pathway, in particular MAPK cascade and ERK cascade
	3	Protein ubiquitination
	4	Histone acetylation
	5	Organelle membrane fusion
4 (Supplementary Fig. S9D)	1	Transcription and apoptotic processes
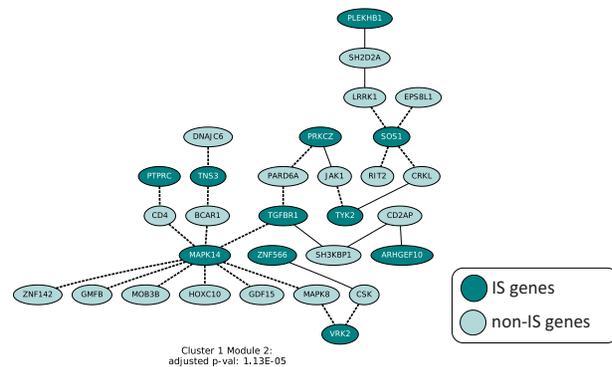
Fig. 5. Spycone results in modules of the PPI network and their corresponding gene set enrichment results. Active network modules are identified using DOMINO. Each node represents a domain of a gene. Darker nodes are the isoform switched genes and lighter nodes are non-IS genes from the PPI. Dashed edges are the affected interactions between the genes due to the loss/gain of domains during the IS events

Fig. 6. (**A**) Boxplots showing the PSSM score difference between lost/gained exons and unregulated exons at the exon 5′ boundaries in logarithmic scale (one-sided Mann–Whitney *U*-test *P*-value < 0.05). (**B**) Boxplots showing the PSSM scores difference between lost/gained exons and unregulated exons at the exon 3′ boundaries in logarithmic scale (one-sided Mann–Whitney *U*-test *P*-value < 0.05). LE, lost exons; GE, gained exons; UE, unregulated exons

Module 1 (Fig. 5B) revealed that interactions between three kinases (MAPK39, AURKC and DCLK2) and a protein chaperone, HSP90AA1 is affected by IS. HSP90 is expressed under the endoplasmic reticulum (ER) stress caused by SARS-CoV-2 and its inhibitor is identified as a therapeutic inhibition target (Wyler *et al.*, 2021). A previous study found that knock down of MAP3K9 reduced SARS-CoV-2 virus replication (Higgins *et al.*, 2021). DCLK2 is differentially expressed in SARS-CoV-2 patients (Alqutami *et al.*, 2021). AURKC would be an interesting candidate to investigate for its role in SARS-CoV-2 infection.

Besides these three kinases, network enrichment analysis highlighted the general importance of kinases in infection development, e.g. JAK1, LYN, TYK2 and PRKCZ (Fig. 5). JAK1 is responsible for interferon signaling (Yan *et al.*, 2021). Inhibition of LYN reduces the efficiency of SARS-CoV-2 virus replication (Meyer *et al.*, 2021). TYK2, which is a key player for IFN signaling, has been associated with cytokine storms in SARS-CoV-2 patients (Solimani *et al.*, 2021). IS events of kinases might cause major rewiring of the transduction cascade, which could lead to altered immune response, cell cycle control and promote viral replication.

Our analysis also suggests an important role of growth factor receptors (FGFR, epidermal growth factor receptor (EGFR) and vascular endothelial growth factor (VEGF)) and their downstream kinases. They are essential for viral infection since they modulate cellular processes like migration, adhesion, differentiation and survival. One example is that activation of EGFR in SARS-CoV-2

can suppress the IFN response and aid viral replication (Klann *et al.*, 2020).

Another key finding is that E3 ubiquitin ligases are affected by IS. They are known to mediate host immune response by removing virus particles. Various virus species hijack the host E3 ubiquitin ligases in favor of viral protein production (Dubey *et al.*, 2021). They are also involved in maintaining TMPRSS2 stabilization during virus entry to the host cells (Chen *et al.*, 2021).

In splicing factor analysis, ILF3 and SRSF7 are identified as a splicing factor affecting the splicing of exons. ILF3 plays a role in antiviral response by inducing the expression of interferon-stimulated genes (Watson *et al.*, 2020). In another computational analysis, SRSF7 is also predicted to have binding potential with SARS-CoV-2 RNA (Horlacher *et al.*, 2021).

Lastly, in order to get confident time course analysis results, one will need high-resolution data in terms of number of time points and sample replicates. Consequently, at least three time points and three

replicates are recommended in Spycone analysis. However, this criterion is rather met due to technical and economical restraints. Thus, Spycone also provides an option for a permutation test with only one replicate for the dataset under investigation. We demonstrated this usage in a tumor development dataset with one replicate (see Supplementary information).

Limitations. Spycone achieves high precision and considerably higher recall than the only competing tool TSIS. Nevertheless, the moderate recall we observe in particular in the presence of noise shows that there is further room for method improvement. In our simulation Model 2, where we allowed for isoform switches between minor isoforms, we observed a reduction in both precision and recall. Spycone identifies only two isoforms that switch per event, but in reality, an event could involve more than two isoforms. In the future, we should consider multiple-isoforms switches to handle more complex scenarios. In addition, the usage of weighted PPI network might introduce selection bias. However, the higher weight gives higher confidence to an interaction, meaning more domains between the proteins are interacting. Therefore, using weighted PPI helps prioritizing interactions with higher confidence. We believe this advantage outweighs the potential bias. Nevertheless, the usage of weighted PPI is optional.

Spycone uniquely offers features for detailed downstream analysis and allows for detecting the rewiring of network modules in a time course as a result of coordinated domain gain/loss. This type of analysis is limited by the availability of the structural annotation. However, the current developments in computational structural biology that could expand the information about domains and domain–domain interactions e.g. AlphaFold2 (Jumper *et al.*, 2021), will greatly strengthen our tool. Lastly, our PSSM-based approach for splicing factor analysis does not allow us to investigate splicing factors that bind indirectly through other adaptor proteins, requiring further experiments that establish binding sites for such proteins. In our future work, we plan to optimize the algorithm and include introns in the analysis.

Spycone was thus far applied exclusively to bulk RNA-seq data. When considering tissue samples, IS switches between time points could also be attributed to changes in cellular composition. An attractive future prospect is thus to apply Spycone for studying IS in single-cell RNA-seq data where dynamic IS events could be traced across cellular differentiation using the concept of pseudotime. However, the current single-cell RNA-seq technologies are limited in their ability to discern isoforms (Arzalluz-Luque and Conesa, 2018).

5 Conclusion

With declining costs in next-generation sequencing, time course RNA-seq experiments are growing in popularity. Although AS is an important and dynamic mechanism it is currently rarely studied in a time course manner due to the lack of suitable tools. Spycone closes this gap by offering robust and comprehensive analysis of time course IS. Going beyond individual IS events, Spycone clusters genes with similar IS behavior in time course data and offers insights into the functional interpretation as well as putative mechanisms and coregulation. The latter is achieved by RNA-binding protein motif analysis and highlights splice factors that could serve as potential drug targets for diseases. Using simulated and real data, we showed that Spycone has better precision and recall than its only competitor, TSIS and that Spycone is able to identify disease-related pathways in the real-world data, as we demonstrated for SARS-CoV-2 infection. In summary, Spycone brings mechanistic insights about the role of temporal changes in AS and thus perfectly complements RNA-seq time course analysis.

Funding

Conflict of Interest: none declared.

Data availability

The SARS-CoV-2 infection RNA-sequencing data are obtained from the GEO database (accession ID GSE157490). The Spycone package is available as a PyPI package. The source code of Spycone is available under the GPLv3 license at https://github.com/yollct/spycone. The code used to produce the result shown in this manuscript is compiled into the Google colab notebook (https://colab.research.google.com/drive/13CjfzZizPlmxzsT-zm6zEgfdFce1fzSC?usp=sharing). This workflow is documented in the AIMe registry: https://aime.report/DXKacH (Matschinske *et al.*, 2021).

References

Aminikhanghahi,S. and Cook,D.J. (2017) A survey of methods for time series change point detection. *Knowl. Inf. Syst.*, 51, 339–367.

Alqutami,F. *et al.* (2021) COVID-19 transcriptomic atlas: a comprehensive analysis of COVID-19 related transcriptomics datasets in different tissues and clinical settings. *Front. Genet.*, 12, 755222.

Ankerst,M. *et al.* (1999) OPTICS: ordering points to identify the clustering structure. *SIGMOD Rec.*, 28, 49–60.

Arzalluz-Luque,Á. and Conesa,A. (2018) Single-cell RNAseq for the study of isoforms—how is that possible? *Genome Biol.*, 19, 1–19.

Barberan-Soler,S. *et al.* (2011) Co-regulation of alternative splicing by diverse splicing factors in *Caenorhabditis elegans*. *Nucleic Acids Res.*, 39, 666–674.

Bolger,A.M. *et al.* (2014) Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, 30, 2114–2120.

Chen,Y. *et al.* (2021) A high-throughput screen for TMPRSS2 expression identifies FDA-approved compounds that can limit SARS-CoV-2 entry. *Nat. Commun.*, 12, 3907.

Cock,P.J.A. *et al.* (2009) Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25, 1422–1423.

de la Fuente,L. *et al.* (2020) tappAS: a comprehensive computational framework for the analysis of the functional impact of differential splicing. *Genome Biol.*, 21, 119.

Dubey,A.R. *et al.* (2021) Biochemical strategies of E3 ubiquitin ligases target viruses in critical diseases. *J. Cell. Biochem.*, 123, 161–182.

Ester,M. *et al.* (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pp. 226–231.

Feng,H. *et al.* (2019) Modeling RNA-Binding protein specificity in vivo by precisely registering protein-RNA crosslink sites. *Mol. Cell*, 74, 1189–1204.e6.

Guo,W. *et al.* (2017) TSIS: an R package to infer alternative splicing isoform switches for time-series data. *Bioinformatics*, 33, 3308–3310.

Hartigan,J.A. and Wong,M.A. (1979) Algorithm as 136: a K-Means clustering algorithm. *J. R. Stat. Soc. Ser. C Appl. Stat.*, 28, 100–108.

Henikoff,J.G. and Henikoff,S. (1996) Using substitution probabilities to improve position-specific scoring matrices. *Comput. Appl. Biosci.*, 12, 135–143.

Higgins,C.A. *et al.* (2021) SARS-CoV-2 hijacks p38ß/MAPK11 to promote viral protein translation. bioRxiv, 2021.08.20.457146.

Hooper,J.E. *et al.* (2020) An alternative splicing program for mouse craniofacial development. *Front. Physiol.*, 11, 1099.

Horlacher,M. *et al.* (2021) Computational mapping of the human-SARS-CoV-2 protein-RNA interactome. bioRxiv, 2021.12.22.472458.

Jang,J. *et al.* (2021) TimesVector-Web: a web service for analysing time course transcriptome data with multiple conditions. *Genes*, 13, 73.

Jassal,B. *et al.* (2020) The reactome pathway knowledgebase. *Nucleic Acids Res.*, 48, D498–D503.

Javed,A. *et al.* (2020) A benchmark study on time series clustering. *Mach. Learn. Appl.*, 1, 100001.

Johnson,S.C. (1967) Hierarchical clustering schemes. *Psychometrika*, 32, 241–254.

Jumper,J. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.

Kanehisa,M. *et al.* (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.*, **44**, D457–D462.

Kim,D. *et al.* (2021) A high-resolution temporal atlas of the SARS-CoV-2 translatome and transcriptome. *Nat. Commun.*, **12**, 5120.

Klann,K. *et al.* (2020) Growth factor receptor signaling inhibition prevents SARS-CoV-2 replication. *Mol. Cell.*, **80**, 164–174.e4.

Lazareva,O. *et al.* (2021) On the limits of active module identification. *Brief. Bioinform.*, **22**, bbab066.

Levi,H. *et al.* (2021) DOMINO: a network-based active module identification algorithm with reduced rate of false calls. *Mol. Syst. Biol.*, **17**, e9593.

Louadi,Z. *et al.* (2021) Functional enrichment of alternative splicing events with NEASE reveals insights into tissue identity and diseases. *Genome Biol.*, **22**, 327.

Makarenko,I. *et al.* (2004) Passive stiffness changes caused by upregulation of compliant titin isoforms in human dilated cardiomyopathy hearts. *Circ. Res.*, **95**, 708–716.

Matschinske,J. *et al.* (2021) The AIMe registry for artificial intelligence in biomedical research. *Nat. Methods*, **18**, 1128–1131.

Meyer,B. *et al.* (2021) Characterising proteolysis during SARS-CoV-2 infection identifies viral cleavage sites and cellular targets with therapeutic potential. *Nat. Commun.*, **12**, 5553.

Mistry,J. *et al.* (2021) Pfam: the protein families database in 2021. *Nucleic Acids Res.*, **49**, D412–D419.

Mosca,R. *et al.* (2014) 3did: a catalog of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res.*, **42**, D374–D379.

Nueda,M.J. *et al.* (2018) Identification and visualization of differential isoform expression in RNA-seq time series. *Bioinformatics*, **34**, 524–526.

Oughtred,R. *et al.* (2021) The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.*, **30**, 187–200.

Park,H.-S. and Jun,C.-H. (2009) A simple and fast algorithm for K-medoids clustering. *Expert Syst. Appl.*, **36**, 3336–3341.

Patro,R. *et al.* (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*, **14**, 417–419.

Pedregosa,F. *et al.* (2011) Scikit-learn: machine learning in python. *J. Mach. Learn. Res.*, **12**, 2825–2830.

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.

Rosa,B.A. *et al.* (2021) IFN signaling and neutrophil degranulation transcriptional signatures are induced during SARS-CoV-2 infection. *Commun. Biol.*, **4**, 290.

Salvi,V. *et al.* (2021) SARS-CoV-2-associated ssRNAs activate inflammation and immunity via TLR7/8. *JCI Insight*, **6**(18), e150542.

Shim,J.E. and Lee,I. (2016) Weighted mutual information analysis substantially improves domain-based functional network models. *Bioinformatics*, **32**, 2824–2830.

Solimani,F. *et al.* (2021) Janus kinase signaling as risk factor and therapeutic target for severe SARS-CoV-2 infection. *Eur. J. Immunol.*, **51**, 1071–1075.

Tavenard,R. *et al.* (2020) Tslearn, a machine learning toolkit for time series data. *J. Mach. Learn. Res*, **21**, 1–6.

Trincado,J.L. *et al.* (2017) SUPPA2 provides fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol.*, **19**, 40.

Varoquaux,N. and Purdom,E. (2020) A pipeline to analyse time-course gene expression data. *F1000Res.*, **9**, 1447.

Vitting-Seerup,K. and Sandelin,A. (2017) The landscape of isoform switches in human cancers. *Mol. Cancer Res.*, **15**, 1206–1220.

Wan,J. *et al.* (2011) Dynamic usage of alternative splicing exons during mouse retina development. *Nucleic Acids Res.*, **39**, 7920–7930.

Watson,S.F. *et al.* (2020) ILF3 contributes to the establishment of the antiviral type I interferon program. *Nucleic Acids Res*, **48**, 116–129.

Weckbach,L.T. *et al.*; EMB Study Group. (2022) Association of complement and MAPK activation with SARS-CoV-2-Associated myocardial inflammation. *JAMA Cardiol.*, **7**, 286–297.

Wiwie,C. *et al.* (2019) Time-resolved systems medicine reveals viral infection-modulating host targets. *Syst. Med. (New Rochelle)*, **2**, 1–9.

Wyler,E. *et al.* (2021) Transcriptomic profiling of SARS-CoV-2 infected human cell lines identifies HSP90 as target for COVID-19 therapy. *iScience*, **24**, 102151.

Xing,Y. *et al.* (2020) Dynamic alternative splicing during mouse preimplantation embryo development. *Front. Bioeng. Biotechnol.*, **8**, 35.

Yan,B. *et al.* (2021) SARS-CoV-2 drives JAK1/2-dependent local complement hyperactivation. *Sci. Immunol.*, **6**, eabg0833.

Yang,H.-C. *et al.* (2021) G6PD deficiency, redox homeostasis, and viral infections: implications for SARS-CoV-2 (COVID-19). *Free Radic. Res.*, **55**, 364–374.

Zheng,Z.-M. (2004) Regulation of alternative RNA splicing by exon definition and exon sequences in viral and mammalian gene expression. *J. Biomed. Sci.*, **11**, 278–294.