An unbiased pathway analysis (UPA) designed for multi-omics inference of cell signaling pathways

Katri Vaparanta¹, ², ³, *, Johannes A. M. Merilahti¹, ², ³, *, Veera K. Ojala¹, ², ³, and Klaus Elenius¹, ², ³, ⁴

*These authors contributed equally to this work

¹Turku Bioscience Centre, University of Turku and Åbo Akademi University, 20520 Turku, Finland
²Mediticy Research Laboratories, University of Turku, 20520 Turku, Finland
³Institute of Biomedicine, University of Turku, 20520 Turku, Finland
⁴Department of Oncology, Turku University Hospital, Turku, 20521, Finland

Address correspondence to: Klaus Elenius (klaus.elenius@utu.fi)

Running Head: Unbiased multiomics analysis of cell signaling

Abbreviations used: ADAM, a disintegrin and metalloprotease; CTF, carboxy-terminal fragment; GSI IX, gamma-secretase inhibitor IX; ICD, intracellular domain; mCTF, membrane-anchored C-terminal fragment; RIP, regulated intramembrane proteolysis; RTK, receptor tyrosine kinase; TAM, TYRO3, AXL, and MER.
New tools for unbiased cell signaling pathway inference from multi-omics data are needed. Here we propose a new unbiased method, the unbiased pathway analysis (UPA), to model and combine omics data into regulatory complexes and pathways. UPA was validated with publicly available omics data and was found accurate in discovering protein-protein interactions, kinase substrate phosphosite relationships, transcription factor target gene relationships, and signaling pathways. Transcription factor, kinase, subcellular location and function prediction algorithms were devised for transcriptome, phosphoproteome and interactome regulatory complexes and pathways, respectively. To apply UPA in a biologically relevant context, interactome, phosphoproteome, transcriptome and proteome data were collected from analyses carried out using melanoma cells to address gamma-secretase cleavage-dependent signaling characteristics of the receptor tyrosine kinase TYRO3. Significant differences were discovered between full-length and gamma-secretase cleaved TYRO3 when signaling modules associated with cell cycle and growth, cell adhesion and motility, cell morphology, or immune responses were compared. The pathways modeled with UPA reflected both the predicted function and direction of the predicted function in validation experiments.
**Introduction**

The vast amount of high throughput data from the different levels of cell signaling has created a need to derive accurate models on the complex cell signaling networks. Previous efforts of cell signaling landscape construction may be divided into two main categories: methods that are reliant or independent on previous knowledge. The methods that rely on previous knowledge include the classical gene set enrichment analysis (Subramanian et al, 2005), its derivatives, and multiple topology-based methods that use previous data as a backbone to infer pathway networks (Ma et al, 2019). Modeling on the basis of previous knowledge poses three major caveats. Firstly, previous knowledge is often incomplete signifying that all present genes, proteins, transcripts and post-translational modification sites are often not covered by the available databases. Secondly, the uneven acquisition of previous knowledge leads to skewed results in endorsing relationships that have already been discovered and extensively researched before. Thirdly, molecular associations are often context-dependent indicating that the associations found in a different context may not universally apply.

Robust methods that are independent on previous knowledge have been generally based on only one measure of molecular association, mainly correlation, between different genes, proteins, transcripts and post-translational modification sites (Care et al, 2019; Langfelder & Horvath, 2008). Integrative approaches combining more than one robust molecular association measure, however, have not been explored. More sophisticated efforts utilizing machine learning methods such as regressions, Bayesian networks and ODE models have been developed, but only to solve single signaling pathway modeling problems such as gene regulatory networks (Chen & Mar, 2018; Pratapa et al, 2020). Most accurate renderings on the interrelationships of the molecular landscape of the cell have been acquired with methods that are tracking changes in time-course data (Chen & Mar, 2018; Köksal et al, 2018), the availability of which is often a limitation. A need for a robust scalable
benchmark method that is both able to track molecular associations from single time point omics data independent on previous knowledge and is based on more than one measure of molecular association is evident.

The canonical receptor tyrosine kinase (RTK) signaling pathway is initiated by the ligand-activated full-length RTK dimer phosphorylating downstream substrates at the plasma membrane. Recently, it has been proposed that RTKs may also signal by a process called gamma-secretase-mediated regulated intramembrane proteolysis (RIP) (Merilahti & Elenius, 2019). In this process the extracellular ectodomain of an RTK is first proteolytically cleaved generating a soluble RTK ectodomain in the extracellular space as well as a membrane-anchored C-terminal fragment (mCTF). This primary cleavage event (ectodomain shedding) is followed by a secondary cleavage step within the transmembrane domain leading to release of a soluble intracellular domain (ICD) of the RTK into the cytosol (Merilahti & Elenius, 2019). Typically, the primary cleavage is performed by a member of the disintegrin and metalloprotease (ADAM) family of proteases, and the secondary cleavage by the gamma-secretase complex (Güner & Lichtenthaler, 2020). The soluble RTK ICD released in the process includes a functionally competent kinase domain and may promote specific cellular responses upon translocation to subcellular structures, such as the nucleus or the mitochondria (Carpenter & Liao, 2013). For example, the ICDs have been reported to function as transcriptional co-regulators in the nucleus or control the function of Bcl-like proteins in the mitochondria (Vidal et al., 2005; Naresh et al., 2006; Merilahti & Elenius, 2019). While a relatively common phenomenon – approximately half of all human RTKs have been shown to be cleaved by gamma-secretase activity (Merilahti et al., 2017) – relatively little is known about the molecular pathways involved in or the functional significance of RTK signaling via RIP.

Here we set out to create a novel unbiased method to infer signaling pathways from one-time-point multi-omics data. To apply the new unbiased pathway analysis (UPA) to a relevant research question, the canonical versus RIP-regulated signaling of an RTK, TYRO3, was addressed in
melanoma cells. TYRO3 is a member of the TAM RTK family consisting of TYRO3, AXL, MER, that has been reported to be overactive and associated with poor prognosis in melanoma (Zhu et al, 2009b; Graham et al, 2014; Hsu et al, 2019). Additionally, TYRO3 was recently identified as a substrate for gamma-secretase (Merilahti et al, 2017). The results presented here validate the novel UPA in discovering relevant cell signaling modules in previously acquired as well as in freshly acquired multi-omics data. Evidence is provided that the modeled signaling pathways can predict both the cellular function and the direction of the cellular function. Furthermore, we provide the first assessment of cleavage-dependent TYRO3 signaling pathways. The results indicate that the RIP-released ICD of the TYRO3 RTK is capable of activating a unique set of signaling pathways that is different from the canonical TYRO3 signaling activated by the full-length receptor.
Results

The underlying concepts of the unbiased pathway analysis (UPA)

We set out to develop a new robust unbiased method to infer regulatory complexes from different omics data and to combine the complexes into multi-omics pathways. Literature search of previous reports of the key features of omics data predicting molecular associations produced the following observations: First, multiple sources indicate the ability of correlation of expression values to predict molecular association to some extent (Eisen et al., 1998; Imamura et al., 2010). Secondly, the findings from an extensive human interactome study indicate that conserved stoichiometry between two proteins inside the same sample can be predictive of protein-protein interaction networks (Hein et al., 2015). From the grounds of these two concepts, the basis for the combined score between two proteins, transcripts or phosphosites in an omics dataset was devised to reflect the strength of association. In Figure 1A the mathematical formulation for the correlation and stoichiometry score is presented. The correlation score was based on the nonparametric Spearman correlation which was consequently ranked to values between 1 and 0 to allow for a linear scale. The stoichiometry score was based on the relation of the third quartile and first quartile value of relative abundances of the two proteins, transcripts or phosphosites in an omics dataset and similarly ranked to a linear scale from 1 to 0. The combined score was expressed as a simple non-weighted multiplication of the correlation and stoichiometry scores similarly ranked to a linear scale between 1 to 0. Nonparametric formulation was chosen to overset technical variation and to desensitize the model to outlier values.

To reduce computational demand (O(n^2) for run time), a cut-off for the combined score was devised to consider only strong associations. The following inference of regulatory complexes is sensitive to the cut-off parameter C in a fashion that setting the parameter too high will lead to insufficient models. This can be easily overcome by lowering the parameter C until a lower parameter...
value will not lead to changes in the model. In the current version of the inference method the parameter C is controlled by the number of associations considered for each molecule in the data.

The inference of regulatory complexes from the combined score follows the nearest neighbor principle, assuming that while a single high combined score can be random chance, the combination of combined scores between 3 cell signaling molecules would be predictive. The inference method is outlined in Figure 1B and utilizes the combined scores above a chosen cut-off threshold to find maximal scoring 3 member cliques. These 3 member cliques are consequently combined to find larger complexes and trimmed to remove members that are present in a higher scoring complex. The principles of the combination and trimming are outlined in more detail in the Materials and methods section.

The combination of the regulatory complexes follows a similar principle to the initial inference of the regulatory complexes (Figure 1C). A median expression value in each sample is used as a representation of the expression values of a complex and the combined score is calculated based on these values. To allow equal combination of down- and up-regulated complexes, an absolute value was taken from the Spearman correlation to equally weigh positive and negative correlations. The complexes were then combined further into pathways based on the combined scores above a certain threshold with the same rules for combining and trimming the maximal scoring 3 complex cliques that were used in the initial regulatory complex inference. The unbiased pathway analysis is available both as matlab code and executables in Mendeley Data.

**Validation of regulatory complexes modeled with UPA**

The performance of the derived regulatory complex inference method was tested on available interactome (Hein *et al*, 2015), phosphoproteome (Batth *et al*, 2018), and transcriptome (Lachmann
et al, 2018) data. While molecular associations are context-dependent, certain associations are more frequently reproduced in different contexts and sometimes even stable. Therefore, a level of conservation is expected from a score that would accurately reflect the probability of molecular association. The conservation of the correlation, stoichiometry and combined scores were examined by modeling a set of proteins, phosphosites or transcripts from one dataset into regulatory complexes and examining the median of the corresponding scores derived from the same complexes in another dataset. A median score above 0.5, the median score expected if the scores would be randomly assigned to molecular associations, indicated conservation.

In all examined omics data, the combined score was statistically significantly conserved, while the stoichiometry score was significantly conserved only in the mass spectrometry-derived interactome and phosphoproteome data (Figure 2A). The correlation score in turn was below the 0.5 line in all tested omics data, indicating that the correlation score alone is least likely to predict molecular associations (Figure 2A).

To validate that the regulatory complexes reflect discovered molecular associations, a method to count the number of known associations in the modeled regulatory complexes was devised. Known protein-protein interactions, kinase substrate phosphosite relationships, and transcription factor target gene annotations were downloaded from STRING (von Mering et al, 2003), PhosphoSitePlus (Hornbeck et al, 2012) and ENCODE (Davis et al, 2018) databases, respectively. The sums of all co-occurrences of proteins known to interact, substrate sites known to be phosphorylated by the same kinase, and transcripts known to be the target gene for the same transcription factor inside the modeled regulatory complexes were used as validation scores. The validation scores were calculated from the regulatory complexes inferred with correlation, stoichiometry and combined scores to compare the effect of the different scores on the accuracy of the model. The validation scores from the real modeled regulatory complexes were compared to the validation scores of randomly formulated regulatory complexes of the same size to calculate the probability that a similar validation score would
be acquired by random. The probability density of the validation scores of these randomized regulatory complexes are indicated with a dashed line and the validation scores for the regulatory complexes modeled with different scores with solid lines in the Figure 2B-D. The corresponding P-values drawn from the cumulative probability densities of the randomized regulatory complexes for the validation scores of the regulatory complexes modeled with different scores are visualized in Figure 2E.

The regulatory complexes modeled with the combined scores consistently reflected known molecular associations in all interactome, phosphoproteome and transcriptome data. In interactome data, the co-occurrence of two proteins in the same regulatory complex predicted a protein-protein interaction between the two proteins more often than random assignment to regulatory complexes of the same size. Additionally, the median STRING score, that reflects the confidence of the interaction based on the amount and type of evidence on the protein-protein interaction, was consistently higher in the regulatory complexes modeled with the combined score than in the randomized regulatory complexes (Supplementary Figure 1). In the phosphoproteome data, the assignment to regulatory complexes modeled with the combined score indicated that the two phosphorylation sites were controlled by the same kinase more often than by random assignment. In the transcriptome data, in turn, two transcripts in the same regulatory complex modeled with the combined score were more likely to be the target gene for the same transcription factor than randomly chosen transcripts. This effect was similarly observed in transcription factor target gene annotations acquired from REMAP and Literature libraries from the ChEA3 database (Keenan et al, 2019) (Supplementary Figure 2).

The regulatory complexes modeled with only the correlation or stoichiometry score varied in their ability to find known molecular associations based on the modeled data. The regulatory complexes modeled with stoichiometry score were sufficient, although not as consistent as the regulatory complexes modeled with the combined score, in reflecting known protein-protein interactions in interactome data. However, the regulatory complexes modeled with stoichiometry score failed to
reproduce known molecular associations in phosphoproteome and transcriptome data. The regulatory complexes modeled with the correlation score were more or less able to capture known transcription factor target gene relationships in the transcriptome data but were insufficient in finding known protein-protein interactions and kinase substrate relationships in the interactome and phosphoproteome data, respectively. All in all, the regulatory complexes inferred with the combined score consistently captured known molecular associations in all tested omics data, while the complexes inferred with other scores were found lacking in reproducibility and performance across omics datatypes.

Validation of regulatory complex combination to pathways by UPA

Since signaling pathways are a combination of molecular associations, they are similarly expected to be partially conserved. The conservation of the regulatory complex combination to pathways was assessed by modeling a multi-omics dataset into regulatory complexes with either correlation, stoichiometry or combination score, combining the regulatory complexes with either correlation, stoichiometry or combined score and examining the median of the corresponding scores derived from the same modeled pathways in another multi-omics dataset. Multi-omics data analyzed from the same normal and cancer tissue samples from the LinkedOmics (Vasaikar et al, 2018) database were utilized. Of all the tested score combinations for the analysis, inferring the regulatory complexes with the combined score and combining the inferred regulatory complexes into pathways with a combined score was the only score combination for the analysis that demonstrated statistically significant conservation (Figure 3A).

To ensure that the created UPA is also able to combine the inferred regulatory complexes into pathways that reproduce known signaling pathways, a validation method was devised. The sum of all co-occurrences of signaling molecules in separate regulatory complexes combined by the UPA in
pathway annotations was used as a validation score for the modeled pathways. The reference pathway annotations were downloaded from the PathwayCommons database (Cerami et al., 2011). The validation scores were compared to the validation scores calculated from 10,000 randomly combined pathways of the same size to estimate the probability that random assignment of the regulatory complexes would produce pathways that equally reflect known pathway annotations. The probability density of the validation scores of these randomly combined pathways are indicated with a dashed curve and the validation scores for the regulatory complexes modeled with different scores with solid lines in Figure 3B. The corresponding P-values drawn from the cumulative probability densities of the randomized pathways for the validation scores of the regulatory complexes modeled with different scores and combined with different scores are visualized in Figure 3C. The pathways inferred and combined with the combined score consistently reproduced associations represented in known pathway annotations. The second-best performers were the versions of the analysis method where the pathways were inferred with either the stoichiometry score or the combined score and combined with the stoichiometry score. The correlation score-based versions of the analysis method performed worst in finding molecular associations present in known pathway annotations. Taken together, the results from the validations and the level of conservation of the combined score in other datasets strongly imply that the UPA is able to find molecular associations and pathways that reflect previously observed cell signaling modules.

Melanoma cell model to study TYRO3 signaling

Since resources for multi-omics data acquired from the same conditions are limited, a single-time-point multi-omics dataset was collected from samples processed in parallel. To apply the analysis to a biologically relevant context, differences in signaling promoted by different TYRO3 receptor constructs expressed in melanoma cells were addressed. The model of analyzing the signaling of
TYRO3 constructs sensitive or resistant to proteolysis-dependent signaling by RIP was chosen to learn how the signaling via a soluble ICD of a RTK differs from the canonical signaling of a full-length RTK. For prediction of how the TYRO3 constructs mutated at the ADAM cleavage sites (TYRO3 ΔADAM) or at the gamma-secretase cleavage site (TYRO3 ΔGS) produce different signaling fragments, please see Figure 4A.

To study RIP-dependent signaling of TYRO3, the endogenously expressed TYRO3 was knocked down in WM-266-4 melanoma cells by lentiviral transduction of TYRO3-specific shRNAs targeting a sequence in the 3’-UTR of TYRO3 mRNA. Cells with knocked-down endogenous TYRO3 expression were subsequently rescued for TYRO3 expression by transfection with constructs encoding C-terminally Myc- and His-tagged or V5-tagged wild-type human TYRO3 (TYRO3 wild-type) or similarly tagged TYRO3 ΔADAM and TYRO3 ΔGS variants. As a control, the cells were transfected with a construct encoding GFP (vector control). The TYRO3 ΔGS construct has been described before (Merilahti et al., 2017). The TYRO3 ΔADAM construct was designed by identifying the cleavage sites for ADAM10 and ADAM17 in the extracellular TYRO3 amino acid sequence with a prediction method utilizing published cleavage sites in other substrates (Figure 4B; please see Materials and methods).

TYRO3 cleavage site mutants are functional receptors

The functionality of the cloned RIP-resistant TYRO3 mutants was validated by examining their susceptibility to cleavage (Figure 4B), subcellular localization (Figure 4C), phosphorylation (Figure 4D), and ability to phosphorylate known substrates (Figure 4E).

TYRO3 ΔGS mutant has been shown to be resistant to gamma-secretase activity in HEK293 cells (Merilahti et al., 2017). As expected, also in the WM-266-4 melanoma cell model, expression of
TYRO3 ΔGS promoted accumulation of the membrane-anchored C-terminal fragment (mCTF) to a similar extent as treatment of cells expressing wild-type TYRO3 with a chemical gamma-secretase inhibitor GSI IX (Figure 4B). The expression of the ΔADAM TYRO3 construct in turn did not associate with accumulation of fragments of the size of either the soluble intracellular domain (ICD) or mCTF, consistent with the accumulation of these fragments being secondary to ADAM cleavage (Figure 4A and B). Further in accordance with the model in which proteolytic cleavage of the full-length TYRO3 is a prerequisite for nuclear translocation of soluble ICD fragment (Figure 4A), nuclear accumulation of C-terminal epitopes was reduced when the ΔADAM and ΔGS TYRO3 constructs were compared to wild-type TYRO3 in confocal immunofluorescence analyses of WM-266-4 transfectants (Figure 4C).

To test whether the different TYRO3 constructs retained autophosphorylation activity, WM-266-4 transfectants were analyzed by TYRO3 immunoprecipitation followed by anti-phosphotyrosine Western analysis. All three variants demonstrated tyrosine phosphorylation of TYRO3 (Figure 4D). Similarly, a Western analysis of the phosphorylation status of STAT3, a known substrate for TYRO3 (Tsai et al., 2020), indicated that all three variants stimulated phosphorylation of an activating residue in STAT3 (Figure 4E). These findings indicate that the mutant TYRO3 receptors are functional kinases in our cell model and can activate a known downstream component of TYRO3 signaling.

**Multi-omics data of cleavage-dependent TYRO3 signaling**

To acquire multi-omics data on cleavage-dependent TYRO3 signaling (Figure 4F), the WM-266-4 vector control cells or cells expressing the TYRO3 variants were lysed and subjected to either RNA purification or protein extraction for the analysis of transcriptomes and proteomes, respectively. The extracted proteins were further subjected to His-tag pull-down for the analysis of TYRO3 variant
interactomes. For phosphoproteome analysis, the extracted proteins were digested to peptides and subjected to TiO\textsubscript{2} enrichment. The extracts for all proteome, interactome and phosphoproteome analyses were subsequently analyzed by LC-MS/MS with data-dependent acquisition. The purified total RNA was transformed into an mRNA sequencing library and sequenced with Illumina HiSeq3000. Statistical testing was utilized to discover the differentially expressed proteins, interacting proteins, transcripts and phosphorylation sites associating with the expression of each TYRO3 variant, as described in Materials and methods.

The coprecipitating proteins, proteins, phosphorylation sites and transcripts significantly different (P ≤ 0.05, fold change ≥ 1.5) from the vector control cells but not between cells expressing the different TYRO3 variants were considered to represent signaling mediated by the full-length TYRO3 (Figure 4A). In contrast, the coprecipitating proteins, proteins, phosphorylation sites and transcripts that were unique for the cells expressing wild-type TYRO3 – i.e. significantly different between cells expressing wild-type TYRO3 and vector control cells or cells expressing TYRO3ΔGS or TYRO3ΔADAM – were considered to represent signaling mediated by the released ICD of TYRO3. As a result, a total of 117 differentially coprecipitating proteins (Supplementary Figure 7, Supplemental Table 6), 636 differentially expressed phosphorylation sites (Supplementary Figure 8, Supplementary Table 7), 276 differentially expressed transcripts (Supplementary Figure 10, Supplementary Table 8), and 744 differentially expressed proteins (Supplementary Figure 11, Supplementary Table 9) were discovered to associate with the full-length TYRO3. In turn, a total of 343 differentially coprecipitating proteins (Supplementary Figure 7, Supplementary Table 6), 479 differentially expressed phosphorylation sites (Supplementary Figure 9, Supplementary Table 7), 242 differentially expressed transcripts (Supplementary Figure 10, Supplementary Table 8) and 791 differentially expressed proteins (Supplementary Figure 12, Supplementary Table 9) were discovered to associate with the ICD of TYRO3.
UPA of TYRO3 signaling

The differentially expressed signaling events associated with the expression of full-length TYRO3 or TYRO3 ICD were analyzed with the UPA first to extract regulatory complexes. The UPA discovered 24 previously undescribed interactome, 110 phosphoproteome, 46 transcriptome and 130 proteome regulatory complexes that associated with the expression of full-length TYRO3 (Supplementary Figure 4, Supplementary Table 2). The expression of TYRO3 ICD in turn was associated with 53 new interactome, 84 phosphoproteome, 45 transcriptome and 136 proteome regulatory complexes (Supplementary Figure 5, Supplementary Table 2). The inferred regulatory complexes were further combined with the UPA into pathways, resulting in a total of 51 and 46 unique signaling pathways associating with the signaling stimulated by the full-length and the released ICD of TYRO3, respectively (Supplementary Table 1).

Functional enrichment analysis for pathways inferred with the UPA

To contextualize the function of the pathways inferred with UPA, an enrichment analysis (please see Materials and methods for details) to predict the biological process each pathway is involved with was devised. The summary of the predicted functions for each pathway associated with the signaling of full-length and ICD of TYRO3 is displayed in Figure 5 and Supplementary Figure 3. The functional enrichment analysis of UPA was able to identify unique pathways associated with various cellular processes including ones involved in the pathogenesis of cancer. A significant difference was noted in the number and direction of pathways related to growth and cell cycle, cell adhesion and motility, cell morphology, cell death, and immune response between the cells expressing the full-length versus the ICD of TYRO3 (Figure 5). These findings indicate that a differential response to these functions could be enacted by the full-length and ICD of TYRO3.
Functional validation of the functional enrichment analyses of the UPA

To validate the ability of function enrichment of the UPA to predict differential cellular functions based on molecular data, we randomly chose 3 cellular functions predicted to go to opposite directions by the signaling pathways activated with the ICD or the full-length TYRO3. First, to assess proliferation induced by the ICD or the full-length TYRO3, the growth of WM-266-4 transfectants was analyzed using live-cell imaging (Figure 6A). The function prediction of the ICD pathway 31 and the full-length pathways 19 and 10 (Figure 5A-B) indicated that the cleavage and release of the ICD of TYRO3 would promote cellular growth, while the expression of the full-length TYRO3 would down-regulate the positive regulation of proliferation. Indeed, the WM-266-4 cells expressing wild-type TYRO3 proliferated significantly faster than the vector control cells or cells expressing either the ΔGS or the ΔADAM variants of TYRO3 (Figure 6A).

As another read-out to validate the functional enrichment analyses of the pathways inferred with UPA, the adhesion rate of the WM-266-4 transfectants to fibronectin coated wells was investigated by real-time cellular impedance measurement (Figure 6B). Again, as predicted by the ICD pathway 10 and the full-length pathway 20 (Figure 5C), the cells expressing wild-type TYRO3 adhered to fibronectin with a greater affinity than vector control cells or cells expressing either the ΔGS or ΔADAM variant of TYRO3 (Figure 6B).

Finally, the two-dimensional cellular morphology of the WM-266-4 transfectants were examined from thresholded confocal images with the ImageJ MorphoLibJ plugin (Figure 6C). As indicated by the ICD pathway 46 and the full-length pathway 9 (Figure 5D), the cells expressing wild-type TYRO3 exhibited distinct morphology from the vector control cells or cells expressing either the ΔGS or ΔADAM variant of TYRO3 (Figure 6C). The morphological difference was captured by the convexity measure (the ratio between the cell area and its convex area), the lower
value of which suggests a shape with more protrusions. A greater amount of cell protrusions in the cells expressing wild-type TYRO3 was also predicted by the UPA (ICD pathway 1 in Figure 5D).

Furthermore, the WM-266-4 cells expressing wild-type TYRO3 or either of the TYRO3 mutants demonstrated less circular cell shape than the vector control cells, indicating that the full-length TYRO3 additionally regulates cell morphology as suggested by the full-length pathway 9 (Supplementary Figure 4). Taken together, these validation experiments suggest that the function enrichment of the pathways modeled with the UPA are able to predict both the function and the direction of the function from purely molecular multi-omics data.

Transcription factor enrichment analysis function of UPA

Since the UPA was able to predict transcription factor target gene relationships in available transcriptome data (Figure 2D-E and Supplementary Figure 2), a transcription factor enrichment analysis was created (see Materials and methods for details) to predict a transcription factor for each transcriptome regulatory complex. To ensure the validity of the transcription factor predictions, the predicted transcription factors were compared to transcription factors known to be regulated by TYRO3. Indeed, several transcription factors known to be regulated by TYRO3 were predicted to regulate the modeled regulatory transcriptome complexes, such as STAT3 (Tsai et al, 2020), MYC (Chen et al, 2020; Dufour et al, 2019), and MITF (Zhu et al, 2009a) (Supplementary Table 3).

Furthermore, the most common transcription factor predicted to regulate the transcriptional complexes associated with TYRO3 signaling was MYC, with the respective gene amplified in the WM-266-4 cell background (Kraehn et al, 2001). These observations indicate that the transcription factor prediction function was able to produce both molecule and cellular context specific suggestions of upstream transcription factors for the regulatory transcriptome complexes.
Subcellular location enrichment analysis function of UPA

To predict a subcellular compartment for each interactome complex inferred with UPA, a subcellular location enrichment analysis for interactome complexes was devised (see Materials and methods). The basis of the subcellular location enrichment analysis relies on two observations. First, the ability of UPA to discover previously reported protein-protein interactions in available interactome data (Figure 2B,E). Secondly, the need for shared physical location in order for protein-protein interactions to occur. The predicted subcellular localizations of the interactome complexes of full-length and the ICD of TYRO3 (Supplementary Table 4) were compared to the reported localization difference between the wild-type and the ΔGS and ΔADAM variants of TYRO3 (Figure 4C) (Merilahti et al, 2017) to ensure the validity of the predicted subcellular locations. The differences in the predicted localizations of the interactome complexes reflected the known difference in the subcellular localization of full-length and ICD of TYRO3 (Supplementary Table 4). Thirteen out of the 54 (24%) ICD interactome complexes were predicted to localize into the nucleus, while nuclear localization was predicted for only 2 out of the 21 (8 %) full-length interactome complexes ($P = 0.032$ against frequency of nuclear assignment in the interactome complexes of TYRO3 ICD). Moreover, 8 out of the 21 (33%) full-length interactome complexes, but only 5 out of the 54 (9%) ICD interactome complexes, were predicted to localize into the plasma membrane ($P < 0.0001$ against frequency of plasma membrane assignment in the interactome complexes of full-length TYRO3). Both the full-length TYRO3 interactome complexes (5 out of 21; 20%) and the ICD interactome complexes (11 out of 54, 20%) were equally predicted to be localized in the cytoskeletal structures ($P = 0.0025$ for full-length TYRO3 and $P < 0.0001$ for TYRO3 ICD against frequency of cytoskeleton assignment in randomly modelled interactome complexes). These findings indicate that the subcellular location prediction method was able to recapitulate previously reported subcellular location differences.
Kinase enrichment analysis function of UPA

To predict upstream kinases for phosphoproteome complexes inferred with UPA, a kinase enrichment analysis was created. Similarly to UPAs ability to discover previously reported kinase substrate phosphosite relationships in available phosphoproteome data (Figure 2C,E), a kinase enrichment analysis based on reported kinase substrate phosphosite annotations was devised. To augment the significant lack of available kinase substrate phosphosite annotations, an additional method to predict the upstream kinase for phosphoproteome complexes based on reported kinase substrate protein-protein interactions following the concepts of the current kinase enrichment benchmark method KEA3 (Kuleshov et al, 2021) was developed (see Materials and methods). The predicted upstream kinases were referenced to kinases known to be regulated by TYRO3. Indeed several kinases reported to be associated with TYRO3, such as AKT (Brown et al, 2012; Guo et al, 2011), mTOR (Brown et al, 2012; Fujita et al, 2018), GSK3 (Guo et al, 2011; Fujita et al, 2018), MEK1 (Fujita et al, 2018), PKC (Fujita et al, 2018), and CAMK2 (Fujita et al, 2018), were predicted to regulate the phosphoproteome complexes associated with full-length and ICD of TYRO3 (Supplementary Table 5). These results indicate that the kinase enrichment function of UPA is able to produce molecule specific suggestions of upstream kinases.

UPA enables the discovery of previously reported and novel molecular associations

Since the modeled pathways are not limited by the existing pathway annotations, the UPA allows for the discovery of novel molecular associations involved in a biological process. To examine the novel molecular associations indicated to be involved in the same biological process by the UPA, selected signaling pathways of full-length and ICD of TYRO3 were visualized (Figure 6D-E). The cell motility pathway 29 of TYRO3 ICD included proteins, transcripts and phosphorylated residues of proteins known to associate with motility or migration in almost all included regulatory complexes.
(APC2 in INT_35; DNAH2, GRB2, MCU in INT_50; HMGB1, NAV3 in PHOS_ST_35; CRK in PHOS_ST_50; SLC7A6 in TRANS_17, LYST in PROT_100; GPC1, SEMA6A, STK4 in PROT_128 and COL18A1 in PROT_134 in Figure 6D). Similarly, proteins and phosphorylated residues of proteins known to associate with cytoskeleton organization were found in almost all included regulatory complexes in the cytoskeleton organization pathway 9 of full-length TYRO3 (CORO1B in INT_1; MYH9 and RAB11A in PHOS_Y_5; SH3D19 in PHOS_ST_1; MARCKSL1 and PLEC in PHOS_ST_31; IQGAP1 in PHOS_ST_39; MAP1B, MARK3 and SEPT10 in PHOS_ST_68 in Figure 6E). The interplay of most of these molecules in regulating cell motility or cytoskeleton organization has not been described, indicating that the UPA provides indications of novel signaling modules. The cytoskeleton organization pathway 9 of full-length TYRO3 (Figure 6E) paired the interactome complex INT_1 containing TYRO3 to the phosphoproteome complex PHOS_Y_5 containing MERTK and Rab proteins. Since TAM RTKs have been indicated to heterodimerize and cross-phosphorylate (Seitz et al, 2007; Brown et al, 2012) and Y682 is identified as an autophosphorylation site for MERTK, the discovery of this signaling by the UPA enforces the credibility of the method in additionally discovering signaling modules that reflect known cell signaling events.
Discussion

The increased demand for a ground-up analysis method able to model multi-omics data into cell signaling pathways motivated us to create an unbiased pathway analysis, UPA. The basis of the UPA lies in the calculation of two robust metrics of molecular association, the correlation and stoichiometry score, and the consequent modeling of the cell signaling molecules into regulatory complexes based on the nearest neighbor principle. The UPA was able to consistently find relevant biological associations in both freely available and freshly acquired multi-omics data. The method showed consistent performance and the combined score-based networks outperformed correlation-based network models in all tested data types.

The advantages of the UPA lie in its independence on previous knowledge, robustness, which makes it easily scalable to other omics data types, and low computational demand. The proof of the predictive ability to discover relevant molecular association opens up further applications for the combined score in more sophisticated and complex inference methods. Previous unbiased multi-omics pathway analysis methods have been focused mostly on feature extraction and matrix factorization strategies (Argelaguet et al., 2018; Mo et al., 2013). While the inferred latent factors have been found accurate in stratifying samples into clusters and can be predictive of clinical outcome, there is little evidence that the discovered factors would represent actual signaling pathways in the cell. Here we provide evidence that the pathways modelled with the UPA reflect previously described signaling pathways.

Since the UPA is based on the inherent variation between samples in the omics data, the quality of the omics data poses a limitation to the analysis. Data wrought with technical variation, or inflated with missing values or lack of repetitions, will lead to loss of model accuracy. The
nonparametric formulation and a zero inflated version of the stoichiometry score was devised to battle two of these limitations. Since the correlation, stoichiometry and combined scores are ranked and thus dependent on the choice of the signaling molecules included in the dataset, a choice of the cut-offs for the data to be modeled will also significantly affect the final modeled pathways. Involving or omitting proteins, post-translational modifications or transcripts from the omics data used for the analysis will influence the inferred final pathways. For accurate inference of the multi-omics pathways, the multi-omics data should be acquired from the same samples. The UPA does not support combination of multi-omics data that have been acquired from different samples even though they would represent the same tissue type or cell line.

We acquired multi-omics data from the RIP-resistant and wild-type TYRO3 variants and utilized the UPA to shed light on the unknown RIP-associated signaling pathways of the TYRO3 RTK in melanoma cells. The UPA discovered a total of 51 and 46 differential unique signaling pathways for the full-length TYRO3 and the RIP-released TYRO3 ICD, respectively. An enrichment analysis to infer the biological process the pathways were involved with was also devised. A difference in the amount and the direction of the pathways associated with growth and cell cycle, cell motility and adhesion, immune response, chromosome organization, cell death, and cell differentiation were observed between the full-length and RIP-released TYRO3 ICD. Many of these functions may be critically involved e.g. in the progression or therapeutic responses in melanoma. For example, TYRO3 has been implicated in the proliferation, tumorigenesis, chemoresistance and motility of melanoma cells (Demarest et al, 2013; Shao et al, 2018; Zhu et al, 2009a; Tworkoski et al, 2011). Validation experiments confirmed that the UPA with the function prediction was able to accurately predict both the function and the direction of the function of the cells expressing either wild-type or cleavage-impaired variants of TYRO3. These findings also represent the first comprehensive characterization of signaling pathways stimulated by the gamma-secretase released ICD of a RTK as compared to the canonical signaling via the full-length receptor.
The UPA segregates all signaling events into pathways independent on the amount of previous knowledge on the modeled molecules. This feature is crucial in discovery of novel signaling modules. Since the UPA is able to predict molecular associations, it can predict upstream transcription factors of transcripts, kinases of phosphorylation sites, and subcellular locations of proteins in the regulatory complexes, for which these features are unknown. Similarly, the UPA can predict connections to biological functions for proteins, transcripts or post-translationally modified proteins in the pathways for which no previous connection to the biological function has been discovered. The pathways inferred with UPA can be re-entered into UPA to acquire super pathways and thus broader connections between the modelled pathways can additionally be discovered with the UPA.
Materials and methods

Cell culture and transfections

WM-266-4 human melanoma cancer cells and HEK293T cells were cultured in DMEM medium (Life Technologies) supplemented with 10% (weight/vol) fetal calf serum (FCS) (Promocell). For the transfection of WM-266-4 cells, jetPRIME (Polyplus-transfection) transfection reagent was used according to manufacturer’s protocol. For the production of lentiviral particles, the HEK293T cells were transfected with the lentiviral packaging and shRNA plasmids with FuGENE6 (Roche) according to the manufacturer’s instructions.

ADAM10 and ADAM17 cleavage site prediction

The cleavage site prediction method was modified from a previously published transcription factor binding prediction method (Lähdesmäki et al., 2008). The positional relative frequencies of the amino acid sequences in putative ADAM10 and ADAM17 cleavage sites were calculated from the results of two published peptide screens (Caescu et al., 2009; Tucher et al., 2014). The cleavage sites for ADAM10 and ADAM17 were predicted with a sliding window analysis calculating the probability of a sequence window of 10 residues having an ADAM cleavage motif based on the positional relative frequencies of amino acid residues with a 0th order Markov chain. The probability for lacking an ADAM cleavage motif was calculated from the same sequence windows with 0th order Markov chain using the relative frequencies of amino acid residues in proteins as determined by Uniprot (Bateman et al., 2021) as positional frequencies. The sequence windows with at least 2 times higher probability for hosting an ADAM cleavage motif than lacking an ADAM cleavage motif were considered probable cleavage sites. Additional constraint for the cleavage motifs to reside within 35 amino acids

ADAM10 and ADAM17 cleavage site prediction

The cleavage site prediction method was modified from a previously published transcription factor binding prediction method (Lähdesmäki et al., 2008). The positional relative frequencies of the amino acid sequences in putative ADAM10 and ADAM17 cleavage sites were calculated from the results of two published peptide screens (Caescu et al., 2009; Tucher et al., 2014). The cleavage sites for ADAM10 and ADAM17 were predicted with a sliding window analysis calculating the probability of a sequence window of 10 residues having an ADAM cleavage motif based on the positional relative frequencies of amino acid residues with a 0th order Markov chain. The probability for lacking an ADAM cleavage motif was calculated from the same sequence windows with 0th order Markov chain using the relative frequencies of amino acid residues in proteins as determined by Uniprot (Bateman et al., 2021) as positional frequencies. The sequence windows with at least 2 times higher probability for hosting an ADAM cleavage motif than lacking an ADAM cleavage motif were considered probable cleavage sites. Additional constraint for the cleavage motifs to reside within 35 amino acids
from the transmembrane domain was added, due to known ectodomain size constraints of gamma-secretase substrates (Funamoto et al, 2013). The method was validated by ensuring its capability in finding published cleavage sites in the peptides described in Goth et al, 2015. P408/L409 and A418/G419 of TYRO3 were predicted to serve as cleavage sites with highest predicted relative cleavage site probabilities for both ADAM10 and ADAM17.

Plasmids and cloning

The pDONR223-TYRO3ΔGS vector including a TYRO3 insert with a mutated gamma-secretase cleavage site (I449A; TYRO3 ΔGS) has been previously described (Merilahti et al, 2017). Constructs encoding TYRO3 with mutated ADAM cleavage sites (P408A/L409P/G419P triple mutation; TYRO3 ΔADAM) were generated using synthetic DNA fragments (Integrated DNA Technologies) and assembled into pDONR223-based TYRO3 Gateway entry plasmid (Addgene kit #1000000014) (Johannessen et al, 2010; Yang et al, 2011) using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs). Wild-type TYRO3, TYRO3 ΔGS and TYRO3 ΔADAM constructs were cloned from pDONR223 plasmids into pEZY-myc-his (Addgene #18791) (Guo et al, 2008) and pMAX-DEST (Addgene #37631) (Klezovitch et al, 2008) expression plasmids using Gateway Cloning Technology with LR Clonase II Enzyme Mix (Life Technologies) to allow expression of C-terminally Myc-his-tagged and V5-tagged proteins, respectively. Mutations were verified by sequencing.

For the production of lentiviral particles carrying shRNAs, the following plasmids were used: pLKO.1-puromycin plasmid (containing the shRNA) (Sigma-Aldrich), pRSV-Rev, pMDLg/pRRE and pMD2.G (a gift from Didier Trono; Addgene plasmids #12253, #12251 and #12259). Restriction digestion with EcoRI was used to verify the integrity of the commercial plasmids.
**Antibodies**

Anti-TYRO3 (5585), anti-V5 (13202), anti-phospho-STAT3 (8862), and anti-STAT3 (9972) antibodies were purchased from Cell Signaling Technology. Anti-actin (MA1-744) antibodies were purchased from Thermo Fisher Scientific. Anti-β-tubulin (T7816) and anti-phosphotyrosine (4G10; 5-321) antibodies were purchased from Sigma-Aldrich.

**Immunoprecipitation and Western analysis**

For immunoprecipitation experiments, WM-266-4 vector control cells or cells expressing Myc-tagged TYRO3 constructs were incubated in serum-free conditions overnight before lysis into lysis buffer (0.1% Triton X-100, 10 mM Tris-Cl pH 7.4, 150 mM NaCl including Pierce protease and phosphatase inhibitor mini tablet (Thermo Fisher Scientific)). The cell lysates were pre-cleared with 50 µl Pierce protein G magnetic beads (Thermo Fisher Scientific) at 4°C for 1 hour and subjected to affinity enrichment with Pierce anti-c-Myc magnetic beads (Thermo Fisher Scientific) at 4°C overnight. The beads were washed five times with 5×TBS-T buffer (125 mM Tris, 750 mM NaCl, 0.25% Tween-20) and once with ultrapure water. The proteins were denatured and eluted from the beads by incubation at 95°C for 10 minutes in SDS-PAGE loading buffer.

For Western analyses, cell lysates were prepared as above and run on SDS–PAGE gels. The separated proteins were transferred to nitrocellulose membranes, which were incubated with indicated primary antibodies and IRDye-conjugated secondary antibodies (LI-COR) to label the proteins of interest. The IR-signals were detected with the Odyssey CLx imaging system (LI-COR).
Generation of stable TYRO3 knock-down cell lines

For stable downregulation of TYRO3, TYRO3-targeted lentiviral shRNA plasmid TRCN0000231528 (MISSION, Sigma-Aldrich) was used. The control lentiviral shRNA plasmid (Addgene plasmid #1864) was a kind gift from David Sabatini (Sarbassov et al., 2005). Third generation lentiviral packaging system (Addgene) was used in HEK293T cells to produce shRNA-carrying lentiviruses. Growth medium was changed every 24 hours post-transfection. Virus-containing medium was collected after 48 and 72 hours, filtered and titered. The virus-containing medium was used to infect WM-266-4 cells at a multiplicity of infection of 2 in the presence of 8 μg/ml Polybrene (Sigma-Aldrich). Cells were maintained in the presence of 1 μg/ml puromycin (Sigma-Aldrich) to select the cells that stably express the lentiviral shRNA plasmids.

Immunofluorescence and confocal microscopy

To detect ectopically expressed V5-tagged TYRO3 constructs, WM-266-4 transfectants were cultured on coverslips in serum-free conditions overnight. The cells were fixed with 3% paraformaldehyde, and permeabilized with 0.1% Triton X-100. Cells were stained with anti-V5 (13202; Cell Signaling Technologies) and AlexaFluor 555 goat anti-rabbit (Molecular Probes). After labeling the nuclei with 4',6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich), the cells were mounted with Mowiol 40-88 (Sigma-Aldrich). Images were acquired with Zeiss LSM 880 confocal microscope (Zeiss). EzColocalization plugin (Stauffer et al., 2018) in ImageJ (version 1.53c) was used to analyze colocalization from confocal slices taken from the middle of the nucleus. The colocalization of V5-signals with DAPI was calculated using Pearson correlation coefficient (Manders et al., 1992) to measure the nuclear localization of TYRO3. Morphological analysis were
carried out from confocal slices in plane with the plasma membrane using MorphoLibJ v 1.4.2.1 (Legland et al, 2016) plugin in Image J.

**Live-cell imaging of cell growth**

The WM-266-4 cell transfectants were plated in serum-free medium at 30,000 cells/24-well. The wells were imaged every 2 hours for 120 hours and cell confluence was measured with Incucyte ZOOM (Sartorius).

**Real-time cell adhesion assay**

The WM-266-4 cell transfectants were detached using 5 mM EDTA 6 to 12 hours after transfection. The cells were plated at 15,000 cells/well onto fibronectin-coated (5 μg/cm²) xCELLigence E-plate (Agilent) wells in serum free medium. Real time cell impedance was measured with the xCELLigence RTCA analyzer (Agilent) for 24 hours. The resulting cell index values were normalized to the number of cells remaining in the wells after 24 hours.

**RNA-sequencing**

The WM-266-4 cell transfectants were starved in serum-free medium overnight and lysed. Total RNA was extracted using NucleoSpin RNA Plus kit (Macherey-Nagel).

The quality of the RNA samples was ensured using Advanced Analytical Fragment Analyzer (Agilent). The sequencing library was created with 300 ng of sample with TruSeq Stranded mRNA HT Kit (Illumina) and indexed with IDT for Illumina TruSeq RNA UD Indices according to manufacturer’s protocol. Genome-wide strand-specific RNA-sequencing (RNA-seq) was performed
at Turku Bioscience Centre sequencing core with Illumina HiSeq3000 using 75 bp paired-end reading.

Mass spectrometry sample preparation

The WM-266-4 cell transfectants were lysed in mass spectrometry lysis buffer (6 M guanidine hydrochloride, 100 mM Tris-HCl pH 8.5, 5 mM Tris(2-carboxyethyl)phosphine (TCEP), 10 mM chloroacetamide) or affinity enrichment lysis buffer (70 mM octyl-β-D-glucopyranoside, 25 mM Tris-HCl pH 7.5, 150 mM NaCl, Pierce protease and phosphatase inhibitor mini tablet). The lysates were centrifuged at 20000 x g for 6 minutes and the supernatants were collected. The protein concentration of the lysates was measured with Bradford protein assay (Bio-Rad) before proceeding to affinity enrichment or to protein digestion to peptides.

Affinity enrichment of mass spectrometry samples

The WM-266-4 cell transfectants were subjected to protein crosslinking with 2 mM DTBP for 10 minutes. The crosslinking reaction was quenched by incubation with 50 mM Tris-HCl pH 7.5 for 15 minutes before cell lysis. Equal amounts of WM-266-4 cell lysates were pre-cleared with Pierce protein G magnetic beads (Thermo Fisher Scientific) at 4 °C for 1 hour and subjected to affinity enrichment with HisPur Ni-NTA magnetic beads (Thermo Fisher Scientific) at 4 °C overnight. The beads were washed five times with 5×TBS-T buffer (125 mM Tris, 750 mM NaCl, 0.25% Tween-20) and once with ultrapure water. The proteins were denatured, alkylated and eluted from the beads at 95 °C for 10 minutes in elution buffer (6 M guanidine hydrochloride, 100 mM Tris-HCl pH 8.5, 10 mM Tris(2-carboxyethyl)phosphine (TCEP), 10 mM chloroacetamide).
Protein digestion to peptides

Proteins enriched in affinity enrichment or purified cell lysates were digested with lys-C (enzyme/protein ratio 1:100) for 1 hour at 37 °C. The samples were diluted 1:10 with 50 mM NH₄HCO₃ and digested with trypsin (enzyme/protein ratio 1:100) at 37 °C overnight.

Sample desalting

Sep-Pak tC18 96-well plate (Waters) was activated with 100% methanol and conditioned with 0.1% TFA in 80% acetonitrile (Thermo Fisher Scientific). Digested peptides were acidified to a pH 3 with trifluoroacetic acid (TFA) and desalted with the activated and conditioned Sep-Pak tC18 96-well plate. The peptides were washed with 0.1% TFA and eluted with 0.1% formic acid in 50% acetonitrile. Samples were dried in Hetovac vacuum centrifuge (Heto Lab Equipment) and stored dry in -20 °C until analysis with mass spectrometer.

Phosphopeptide enrichment

Phosphopeptides were enriched from desalted and dried peptides using Pierce high-select TiO₂ phosphopeptide enrichment kit (Thermo Fisher Scientific) according the manufacturer’s instructions. Following elution, phosphopeptides were dried with Hetovac vacuum centrifuge and stored as dry in -20 °C until analysis with mass spectrometer.
Mass spectrometry

Dried peptide samples were resuspended in 0.1% formic acid and sample concentrations were measured with Nanodrop 1000 (Thermo Fisher Scientific). Equal amounts of samples were analyzed on an Easy-nLC 1000 coupled to a Orbitrap Fusion Lumos instrument (Thermo Fisher Scientific) at Turku Bioscience Centre Proteomics Core Services. Peptides were loaded on in-house packed 100 µm × 2 cm precolumn packed with ReproSil-Pur 5 µm 200 Å C18-AQ beads (Dr. Maisch) using 0.1% formic acid in water (buffer A) and separated by reverse phase chromatography on a 75 µm × 15 cm analytical column packed with ReproSil-Pur 5 µm 200 Å C18-AQ beads (Dr. Maisch). All separations were performed using a 60 minute gradient ranging from 8% buffer B (80% acetonitrile in 0.1% formic acid) to 21% in buffer B in 28 minutes and to 36% buffer B in 22 minutes and ramped to 100% buffer B in 5 minutes at flow rate of 300 nl/minute. The washout followed at 100% buffer B for 5 minutes.

All MS spectra were acquired on the orbitrap mass analyzer and stored in centroid mode. For data-dependent acquisition experiments, full MS scans were acquired from 300 to 1600 m/z at 120,000 resolution with fill target of 7E5 ions and maximum injection time of 50 ms. The most abundant ions on the full MS scan were selected for fragmentation using 1.6 m/z precursor isolation window and beam-type collisional-activation dissociation (HCD) with 30% normalized collision energy for a cycle time of 3 seconds. MS/MS spectra were collected at 15,000 resolution with fill target of 5E4 ions and maximum injection time of 100 ms. Fragmented precursors were dynamically excluded from selection for 35 seconds.
Protein identification and quantification

MS/MS spectra were searched with Metamorpheus (version 0.0.304) (Solntsev et al., 2018) against human proteome containing known post-translational modifications (downloaded from Uniprot on 19.2.2019). The mass spectrometry files were calibrated, possible post-translational modifications were searched and peptides and proteins were identified and quantified using FlashLFQ algorithm (Millikin et al., 2018). The following parameters were used for the post-translational modifications search. Cysteine carbamidomethylation and methionine oxidation were set as constant and variable modifications, respectively and G-DPM search in Metamorpheus was additionally used to discover other modifications. Search results were filtered to a 1% FDR at PSM level. Peptides were accepted with search engine score above 5.

RNAseq data analysis

The RNAseq reads were quality checked with FastQC (Babraham Bioinformatics), quality and adapter trimmed with PRINSEQ (Schmieder & Edwards, 2011) and Trimmomatic (Bolger et al., 2014), and pseudoaligned with kallisto v 0.46.0 (Bray et al., 2016) to human transcriptome Ensembl v96 (Yates et al., 2020) to retrieve TPM (transcripts per million) values. The batch effect between experiment 1 and experiments 2 and 3 was corrected with batchelor (Haghverdi et al., 2018) and the TPM values were normalized by library size. Differential expression between the vector control sample and samples representing the different variants of TYRO3 was analyzed with DeSeq2 (Love et al., 2014). Transcripts with fold change over 1.5 and FDR-adjusted P-value lower or equal than 0.05 were chosen for further analysis. The transcripts significantly different between the vector control sample and all of the samples representing the TYRO3 variants (wild-type, ∆ADAM or ∆GS) were considered to reflect the signaling mediated by the full-length TYRO3 receptor. The transcripts significantly different between the vector control sample and the sample representing wild-type
TYRO3, but not between the vector control sample and the samples representing ∆ADAM or ∆GS variant of TYRO3, were considered to reflect the signaling mediated by the soluble ICD of TYRO3.

Mass spectrometry data analysis

The interactome, proteome and phosphoproteome data were normalized to the sum of intensities of all detected proteins in the sample. A probability density function was fitted with Epanechnikov kernel to median normalized intensities of different treatments to estimate the P-value for differential expression from the cumulative density function. Proteins with fold change over 1.5 and FDR adjusted P-value lower or equal to 0.05 were chosen for further analysis. The coprecipitating proteins, proteins and phosphorylated residues significantly different between the vector control sample and all of the samples representing the TYRO3 variants (wild-type, ∆ADAM or ∆GS) were considered to reflect the signaling mediated by the full-length TYRO3 receptor. The proteins significantly different between the sample representing wild-type TYRO3 and the vector control sample as well as between the sample representing wild-type TYRO3 and samples representing the ∆ADAM or ∆GS variant of TYRO3, were considered to reflect the signaling mediated by the soluble ICD of TYRO3.

The inference of regulatory complexes and their combination in the UPA

Matlab R2016a (MathWorks) was used as the coding environment. The pair-wise association between two proteins, phosphosites or transcripts was determined by using a combined score. The combined score is derived from the multiplication of the correlation and the stoichiometry score the mathematical formulations of which are presented in Figure 1A. The correlation score for all the possible protein, phosphosite or transcript combinations was calculated with the Spearman correlation and ranked to derive scores from 1 to 0 in equal increments. The stoichiometry score for
all the possible protein, phosphosite or transcript combinations was calculated by dividing the third quartile value (Q₃) of relative abundances with the first quartile (Q₁) value and similarly ranked to derive scores from 1 to 0 in equal increments. Two versions of the stoichiometry score were devised for non-zero-inflated and zero-inflated data. In the non-zero-inflated version of the stoichiometry score, relative abundance data were considered only from samples in which both intracellular molecules were present. In the zero-inflated version of the stoichiometry score (weighted stoichiometry score), missing values in the same sample of only one of the intracellular molecules was punished by inflating the stoichiometry score and missing values in the same sample of both intracellular molecules was rewarded by deflating the stoichiometry score.

A cut-off parameter C was devised to reduce computational time to only consider combined scores higher than a certain threshold. The cut-off parameter was determined by the minimum number of interactions considered for intracellular molecules that can be set in the analysis (the parameter C_num). To find the set threshold the parameter C was decreased by an increment than can be set in the analysis by changing the step increment parameter.

The regulatory complex inference analysis was designed to utilize the nearest neighbor concept to find the highest scoring (sum of the combined score between all 3 members) 3 member networks for all intracellular molecules in the list and to combine and trim the networks based on common network members and combined scores as described in the following order: 1) Networks with 2 common members are combined until no two complexes with 2 common members can be found. 2) Networks with only one common member are combined if the size of both networks is under or equal to 4 and the network score is higher or equal to 2 indicating a strong association. The parameters for this joining (maximum size of a complex to be joined with only one common node and minimum complex score for the complexes to be joined by only one common node) can be set if smaller regulatory complexes or regulatory complexes with a higher association scores are preferred. 3) Networks with 2 common members are again combined until no two complexes with 2 common
members can be found. 4) Two out of the 3 members of the remaining 3 member networks are removed if 2 out of three members are already present in another complex. 5) The remaining 3 member networks are joined if they still have a common member. This step can be omitted if smaller regulatory complexes are preferred. The analysis allows for a protein, phosphosite or a transcript to be a member for more than one regulatory complex.

The same strategy was used for the combination of the regulatory complexes into larger modules as was the derivation of the regulatory complexes in the first place with a few modifications. The median value of all complex members in each sample was used to combine the regulatory complexes and the absolute value of the Spearman correlation was used to allow equal combination of up- and downregulated regulatory complexes. The default settings were used for the analysis of the TYRO3 signaling pathways.

Validation of regulatory complexes and their combination

The data on known protein-protein interactions, kinase substrate phosphosite relationships and transcription factor target gene relationships were acquired from the STRING, PhosphoSitePlus and ChEA3 databases, respectively (von Mering et al., 2003; Hornbeck et al., 2012; Keenan et al., 2019; Davis et al., 2018; Chèneby et al., 2018). The data on annotated signaling pathways was acquired from PathwayCommons (Cerami et al., 2011). The validation interactome, phosphoproteome and transcription data were acquired from the publications of Batth et al and Karayel et al, and from ArchS4 database, respectively (Lachmann et al., 2018; Batth et al., 2018; Karayel et al, 2020). The validation data for the regulatory complex combination to pathways was acquired from the LinkedOmics database (Vasaikar et al., 2018), where a combination of either proteomics, transcriptomics, methylomics, phosphoproteomics, protein array or acetylation data from the same samples were modeled and combined. The sum of all protein-protein interactions, kinase substrate
phosphosite relationships or transcription factor target gene relationships in the modeled regulatory complexes was used as the validation score for interactome, phosphoproteome and transcriptome regulatory complex validation, respectively. The sum of co-occurrences of two complex members from two different combined complexes in the same pathway annotation for all the combined complexes was used as a validation score for regulatory complex combination. The validation score was compared to the sum of all protein-protein interactions, kinase substrate phosphosite relationships, transcription factor target gene relationships or co-occurrences of two complex members from two different randomly combined complexes in the same pathway annotation in randomly modeled regulatory complexes or combined pathways of same size as in the modeled data to derive an empirical probability density function from 1000 rounds of simulation. An Epanechnikov kernel was used to fit the simulated data into the empirical probability density function. The corresponding P-values were drawn from the empirical cumulative distribution function. To assess the effect of different scores on the accuracy of the modeled regulatory complexes and pathways, the complexes were modeled with either correlation score, stoichiometry score or the combined score and combined with either correlation score, stoichiometry score or the combined score. The conservation of the different scores was assessed in each data type by deriving the median score of the modeled complexes in different datasets from the one used for the initial modeling.

Transcription factor and kinase prediction

An enrichment analysis to predict a transcription factor for each transcriptome and a kinase for each phosphoproteome regulatory complex was devised. Transcription factor target gene annotations from the ChEA3 data resource (Keenan et al., 2019) were used and only the target gene annotations of the transcription factors present in the dataset with at least 5 transcripts were considered. The annotations of the upstream kinase for each phosphosite was acquired and combined from the results of a kinase
knock-down screen (Sugiyama et al., 2019), RegPhos 2.0 (Huang et al., 2014), PhosphoSitePlus (Hornbeck et al., 2012), and KinaseNET (Kinexus BioInformatics Corporation) databases. The annotations of upstream kinases based on protein-protein interactions were accumulated from BioGrid (Oughtred et al., 2021), Cheng et al. (Cheng et al., 2014), Harmonizome (Rouillard et al., 2016), HIPPIE (Schaefer et al., 2012), Mentha (Calderone et al., 2013), Mint (Zanzoni et al., 2002), PIPs (McDowall et al., 2009), PSOPIA (Murakami & Mizuguchi, 2014), Reactome (Jassal et al., 2020), and STRING (von Mering et al., 2003) databases. The kinase annotations were filtered according to the LFQ intensity of the kinases in the proteome and phosphoproteome data. Only substrate phosphosite or protein-protein interaction annotations of the kinases that were present in the corresponding samples with at least the LFQ intensity of 10000 were considered.

The enrichment analyses were created to follow the 3 steps as described: Step 1) The annotation most enriched in all the transcripts or phosphosites in all the complexes is given to the complexes for which the annotation gives the highest statistically significant enrichment score (the number of transcripts or phosphorylated residues found in the annotation divided by the number of transcripts or phosphorylated residues in the complex; P-value equal or less than the set threshold). The most enriched annotation and the complexes for which the annotation gives the highest enrichment score are removed. For the remaining dataset the same process is repeated until no complex remains. If more than one annotation provides the same overall and complex specific enrichment, all annotations are given as equally likely to the complexes. To derive a P-value for the enrichment of each annotation in each regulatory complex 10,000 random sets of equal size from all transcripts or phosphorylated residues identified in the transcriptome or phosphoproteome data are drawn. The respective enrichment of each of these randomized sets is derived and a probability density function is fitted with an Epanechnikov kernel to the enrichment scores. The P-values for the enrichment are drawn from the corresponding empirical cumulative distribution function. Step 2) Secondary transcription factor and kinase predictions are performed by assessing whether any of the
primary transcription factors or kinases would be statistically significantly enriched also in the regulatory complexes were another transcription factor or kinase is assigned as the primary. P-values are similarly acquired for the secondary transcription factors and kinases. If more than one annotation provides the same complex specific enrichment with a P-value lower than 0.05, all annotations are given as equally likely to the complex. Step 3) If any unannotated complexes remain, all remaining annotations are searched to find the first statistically significant annotation with the highest enrichment score. If more than one annotation provides the same enrichment score with a P-value lower than the set threshold, all annotations are given as equally likely to the complex.

**Subcellular location prediction**

An enrichment analysis to predict the subcellular location of each regulatory interactome complex was devised. The annotations of subcellular locations was acquired from the knowledge, experiments and text mining channels of COMPARTMENTS database (Binder et al, 2014). The enrichment score (i.e., relative frequency) of each annotation for each interactome complex was calculated. To derive a P-value for the enrichment of each annotation in each regulatory complex, 1,000 random sets of equal size from all possible proteins identified in the interactome data were drawn. The respective enrichment of each of these randomized sets was derived and a probability density function was fitted with an Epanechnikov kernel to the enrichment scores. The P-values for the enrichment were drawn from the corresponding empirical cumulative distribution function. Annotations with a P-value lower or equal to 0.05 were considered significant. The subcellular location annotation with a highest enrichment score and lowest P-value were selected.
Function prediction

An enrichment analysis to predict the biological process involved with the inferred pathways was devised. The Gene Ontology resource (Ashburner et al., 2000; Carbon et al., 2021) annotations for biological processes were acquired from the MSigDB v7.3 (Liberzon et al., 2011). The enrichment score (i.e., relative frequency) of each annotation for each pathway was calculated. To derive a P-value for the enrichment of each annotation in each pathway 10,000 random sets of equal size from all possible proteins, transcripts and phosphorylated proteins identified as significantly altered in the condition were drawn. The respective enrichment of each of these randomized sets was derived and a probability density function was fitted with an Epanechnikov kernel to the enrichment scores. The P-values for the enrichment were drawn from the corresponding empirical cumulative distribution function. Annotations with a P-value lower or equal to 0.05 were considered significant. The function annotation with a highest enrichment score and lowest P-value were selected. Contextually irrelevant annotations, such as specific functions of certain non-skin tissues or non-melanoma cell types, were ignored.

Information visualization and statistics

For statistical analysis, the GraphPad Prism software (Version 9.02; GraphPad Software) and Matlab 2016a were utilized. The details of the statistical testing are described in figure legends. All datasets were tested for normality and parametric or non-parametric testing was used accordingly. Heatmaps were generated with Morpheus (https://software.broadinstitute.org/morpheus). The visualizations of the inferred regulatory complexes and pathways were created with Cytoscape (Version 3.8.2; Shannon et al., 2003) and Affinity Designer (Version 19.0; Serif)
Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD028881. The RNAseq data have been deposited to the Gene Expression Omnibus database with the identifier: GSE190431. The UPA matlab code and executables are available through Mendeley data.
Acknowledgements

We thank Maria Tuominen for skillful technical assistance. This work was supported by the Academy of Finland, Cancer Foundation Finland, the Sigrid Juselius Foundation, the Turku University Central Hospital, Orion Research Foundation, K. Albin Johanssons Foundation, the Cancer Society of Southwestern Finland, the Jenny and Antti Wihuri Foundation, the Maud Kuistila memorial Foundation, Emil Aaltonen Foundation, the Finnish Foundation for Cardiovascular Research, Finnish Cultural Foundation and the Varsinais-Suomi Regional Fund of the Finnish Cultural Foundation.

Competing interests

The authors declare that they have no conflict of interests.

Author contributions

JAMM, KV, KE designed the study, JAMM, KV, and VKO performed the experimentation. KV created the unbiased pathway analysis. TYRO3 knock-down cells were prepared by VKO. JAMM, KV and KE wrote the manuscript.

References


Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, Toyoda Y, Gak IA, Weisswange I, Mansfeld J,

M (2012) PhosphoSitePlus: A comprehensive resource for investigating the structure and
function of experimentally determined post-translational modifications in man and mouse.
Nucleic Acids Res 40

(Maywood) 244: 83–99

phosphorylation networks in mammals. Database 2014: 1–12

of signal transduction networks using phosphorylation dynamics data. BMC Bioinformatics 11


Johannessen CM, Boehm JS, Kim SSY, Thomas SR, Wardwell L, Johnson LA, Emery CM, Stransky
kinase pathway reactivation. Nature 468: 968–72

Karayel Ö, Xu P, Bludau I, Velan Bhoopalan S, Yao Y, Ana Rita FC, Santos A, Schulman BA, Alpi

Keenan AB, Torre D, Lachmann A, Leong AK, Wojciechowicz ML, Utti V, Jagodnik KM,

Kinexus BioInformatics Corporation KinaseNET.

A causal role for ERG in neoplastic transformation of prostate epithelium. Proc Natl Acad Sci
USA 105: 2105–10

Köksal AS, Beck K, Cronin DR, McKenna A, Camp ND, Srivastava S, MacGilvray ME, Bodik R,

c-myc oncogene copies in high risk cutaneous malignant melanoma and melanoma metastases.
Br J Cancer 84: 72–79

Ma’Ayan A (2021) KEA3: improved kinase enrichment analysis via data integration. Nucleic
Acids Res 49: W304–W316

BioRxiv preprint doi: https://doi.org/10.1101/2022.02.05.479228; this version posted February 6, 2022. The copyright holder for this
preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.


1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113
The correlation score for Protein 1 ($P_1$) and Protein 2 ($P_2$) for total number ($n$) of samples $s$ is calculated as follows:

$$\text{correlation score} = \frac{\sum_{i=1}^{n} \text{Rank}(P_1)_{i} \cdot \text{Rank}(P_2)_{i} - \left(\sum_{i=1}^{n} \text{Rank}(P_1)_{i}\right) \left(\sum_{i=1}^{n} \text{Rank}(P_2)_{i}\right)}{\sqrt{\left(\sum_{i=1}^{n} \text{Rank}(P_1)_{i}\right)^2 - \left(\sum_{i=1}^{n} \text{Rank}(P_1)_{i}\right)^2} \cdot \sqrt{\left(\sum_{i=1}^{n} \text{Rank}(P_2)_{i}\right)^2 - \left(\sum_{i=1}^{n} \text{Rank}(P_2)_{i}\right)^2}} \cdot (s-1)$$

The stoichiometry score for Protein 1 ($P_1$) and Protein 2 ($P_2$) for total number ($n$) of samples is calculated as follows:

$$\text{stoichiometry score} = \frac{\sum_{i=1}^{n} \text{Rank}(P_1)_{i} \cdot \text{Rank}(P_2)_{i} - \left(\sum_{i=1}^{n} \text{Rank}(P_1)_{i}\right) \left(\sum_{i=1}^{n} \text{Rank}(P_2)_{i}\right)}{\sqrt{\left(\sum_{i=1}^{n} \text{Rank}(P_1)_{i}\right)^2 - \left(\sum_{i=1}^{n} \text{Rank}(P_1)_{i}\right)^2} \cdot \sqrt{\left(\sum_{i=1}^{n} \text{Rank}(P_2)_{i}\right)^2 - \left(\sum_{i=1}^{n} \text{Rank}(P_2)_{i}\right)^2}} \cdot (s-1)$$

### Figure 1. The underlying concepts of the unbiased pathway analysis.

**A**: Mathematical formulation of the correlation (Spearman correlation) and stoichiometry scores. **B**: Schematic representation of the inference of regulatory omics complexes for the unbiased multi-omics pathway inference analysis. **C**: Schematic representation of the principles of combining the regulatory omics complexes into pathways.
Figure 2. Validation of regulatory complexes modeled with the UPA. A: The conservation of correlation, stoichiometry and combined score in interactome, phosphoproteome and transcriptome data. One dot represents the median score between all complex members of all modeled complexes of a dataset that was not utilized for the initial regulatory complex inference. The dashed line represents the median score of randomized complexes. For statistical testing, two-tailed one sample T-test against the theoretical median value of 0.5 was utilized. B-E: Empirical probability densities simulated by randomizing the members of the modeled regulatory complexes into complexes of the same size and the corresponding P-values for the complexes modeled from either interactome (B, E), phosphoproteome (C, E) or transcriptome data (D, E) by either correlation (Corr), stoichiometry (Stoi) or combination score (Comb). The sum of co-occurrences of either interacting proteins as determined by the STRING database, phosphorylated kinase substrate residues as determined by the PhosphoSitePlus database or transcription factor target genes as determined by the ENCODE database in the modeled complexes were used as a validation score. The empirical probability densities are visualized with dashed lines of the corresponding color and the score for the modeled network in solid line. One dot in the boxplot represents one P-value for one dataset and the horizontal line the median value. Red: correlation score; blue: stoichiometry score; grey: combined score.
Figure 3. Validation of regulatory complex combination to pathways with the UPA. A: The conservation of correlation, stoichiometry and combined score in complexes modeled with either correlation (Corr), stoichiometry (Stoi) or combined (Comb) score and combined with either correlation, stoichiometry or combined score. Corr, Corr: modeled with correlation score, combined with correlation score. Comb, Corr: modeled with correlation score, combined with combined score. Stoi, Stoi: modeled with stoichiometry score, combined with stoichiometry score. Comb, Stoi: modeled with stoichiometry score, combined with combined score. Comb, Comb: modeled with combined score, combined with combined score. One dot in the boxplot represents the median score between all complexes of all inferred pathways of a dataset that was not initially utilized for the pathway inference. The horizontal line represents the median value and the dashed line the median score for randomized pathways. For statistical testing, two-tailed one sample T-test against the theoretical median value 0.5 was utilized. B-C: Empirical probability densities simulated by randomizing the modeled regulatory complexes into pathways of the same size and the corresponding P-values. Acetylation, methylation, proteomics, phosphoproteomics, transcriptomics and protein array data from LinkedOmics was modeled and combined. The sum of co-occurrences of genes, transcripts and proteins in a pathway annotation in separate combined regulatory complexes as determined by the PathwayCommons database were used as a validation score. One dot in the boxplot represents a P-value for one dataset and the horizontal line the median value.
Figure 4. Functional validation of RIP-resistant TYRO3 receptor variants. A: A Schematic depicting the signaling mediated by wild-type and regulated intramembrane proteolysis (RIP)-resistant variants (ΔGS, ΔADAM) of TYRO3. Since the cleavage events in RIP are sequential, blocking only the gamma-secretase cleavage will allow the receptor still to signal through the membrane-anchored C-terminal fragment (mCTF) in addition to the full-length receptor. Both cleavage events are blocked when the primary shedding performed by ADAM proteases is inhibited leading the receptor to signal only through the canonical full-length receptor. The wild-type receptor in turn is able to signal through the canonical full-length receptor, the mCTF and the released soluble intracellular domain (ICD).

B: Western analysis of WM-266-4 cells expressing the indicated TYRO3 variants and treated with or without the gamma-secretase inhibitor GSI IX. The full-length receptor and the different cleavage products of TYRO3 are indicated.

C: Confocal microscopy analysis of TYRO3 in WM-266-4 cells expressing the indicated V5-tagged variants of TYRO3. V5 signal is shown in red and DAPI-stained nuclei in blue. Nuclear localization is presented as Pearson correlation coefficient of TYRO3-V5 co-localizing with DAPI within the cells. For statistical testing, the non-parametric Kruskal-Wallis ANOVA was utilized. The post hoc analyses were conducted with the Mann-Whitney U test and the resulting P-values were corrected with the method of Benjamini, Krieger and Yekutieli. One dot represents one cell and the horizontal line the median value.

D: Western analysis of the autophosphorylation of TYRO3 in WM-266-4 cells expressing the indicated TYRO3 variants.

E: Western analysis of the phosphorylation status of TYRO3 downstream effector STAT3 in WM-266-4 transfectants.

F: The workflow of interactome, phosphoproteome, proteome and transcriptome data acquisition from WM-266-4 transfectants.
Figure 5. The functional categorization of the selected full-length and TYRO3 ICD pathways inferred with the UPA. The predicted function of the multi-omics pathways of the full-length and ICD of TYRO3 inferred with the UPA. The bars represent the median pseudolog2 fold change of all the proteins, transcripts and phosphorylated residues in the pathway against the control condition. For a full list of predicted functions for the pathways of full-length and ICD of TYRO3 see Supplementary Figure 3. For further details on the pathways see supplemental tables 1 and 2. The highlighted pathways are visualized in Figure 6D-E.
Figure 6. Functional validation and composition of the selected pathways of full-length and ICD of TYRO3. A: Proliferation of WM-266-4 transfectants was measured with live-cell imaging. For statistical testing, the non-parametric Friedman 2-way ANOVA and the Dunn’s multiple comparison test with a multiple test correction for the P-values was utilized. The symbols represent the mean and the whiskers the standard deviations of the values. A representative plot from 1 out of 3 independent replicate experiments is shown. B: Real-time adhesion of WM-266-4 transfectants was explored with the xCELLigence cell impedance measurement system. For statistical testing, the parametric 2-way ANOVA and the Dunn’s multiple comparison test with a multiple test correction for the P-values was utilized. A representative plot from 1 out of 4 independent replicate experiments is shown. C: The morphology of WM-266-4 transfectants was analyzed from thresholded confocal images taken in plane with the plasma membrane with MorphoLibJ plugin of ImageJ. For statistical testing, the non-parametric Kruskal-Wallis ANOVA was utilized. The post hoc analyses were conducted with the Mann-Whitney U test and the resulting P-values were corrected with the method of Benjamini, Krieger and Yekutieli. Convexity: the ratio between the convex perimeter and the real perimeter. One dot represents one cell and the horizontal line the median value. Combined results from 4 independent experiments is shown. D-E: Visualization of the cell motility pathway 29 of TYRO3 ICD (D) and the cytoskeleton organization pathway 9 of full-length TYRO3 (E). The median expression of all the transcripts, proteins or phosphorylated residues in the indicated samples and regulatory complexes is presented in the heatmaps. Individual complexes are separated by circles. The inferred regulatory complexes are identified by a letter combination (INT for interactome, PHOS_ST for serine/threonine phosphoproteome, PHOS_Y for tyrosine phosphoproteome, TRANS for transcriptome, PROT for proteome) and a number. Predicted transcription factors, kinases and subcellular locations are indicated with italic letters under the complex identifier.