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

Katri Vaparanta[1, 2, 3, *], Johannes A. M. Merilahti[1, 2, 3, *], Veera K. Ojala [1, 2, 3], and Klaus Elenius[1, 2, 3,4, a]

*These authors contributed equally to this work

[1] Turku Bioscience Centre, University of Turku and Åbo Akademi University, 20520 Turku, Finland

[2] Medicity Research Laboratories, University of Turku, 20520 Turku, Finland

[3] Institute of Biomedicine, University of Turku, 20520 Turku, Finland

[4]Department of Oncology, Turku University Hospital, Turku, 20521, Finland

[a]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.**

33 # Introduction

34 The vast amount of high throughput data from the different levels of cell signaling has created a need
35 to derive accurate models on the complex cell signaling networks. Previous efforts of cell signaling
36 landscape construction may be divided into two main categories: methods that are reliant or
37 independent on previous knowledge. The methods that rely on previous knowledge include the
38 classical gene set enrichment analysis (Subramanian *et al*, 2005), its derivatives, and multiple
39 topology-based methods that use previous data as a backbone to infer pathway networks (Ma *et al*,
40 2019). Modeling on the basis of previous knowledge poses three major caveats. Firstly, previous
41 knowledge is often incomplete signifying that all present genes, proteins, transcripts and post-
42 translational modification sites are often not covered by the available databases. Secondly, the uneven
43 acquirement of previous knowledge leads to skewed results in endorsing relationships that have
44 already been discovered and extensively researched before. Thirdly, molecular associations are often
45 context-dependent indicating that the associations found in a different context may not universally
46 apply.

47 Robust methods that are independent on previous knowledge have been generally based on
48 only one measure of molecular association, mainly correlation, between different genes, proteins,
49 transcripts and post-translational modification sites (Care *et al*, 2019; Langfelder & Horvath, 2008).
50 Integrative approaches combining more than one robust molecular association measure, however,
51 have not been explored. More sophisticated efforts utilizing machine learning methods such as
52 regressions, Bayesian networks and ODE models have been developed, but only to solve single
53 signaling pathway modeling problems such as gene regulatory networks (Chen & Mar, 2018; Pratapa
54 *et al*, 2020). Most accurate renderings on the interrelationships of the molecular landscape of the cell
55 have been acquired with methods that are tracking changes in time-course data (Chen & Mar, 2018;
56 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

82 melanoma cells. TYRO3 is a member of the TAM RTK family consisting of TYRO3, AXL, MER,
83 that has been reported to be overactive and associated with poor prognosis in melanoma (Zhu *et al*,
84 2009b; Graham *et al*, 2014; Hsu *et al*, 2019). Additionally, TYRO3 was recently identified as a
85 substrate for gamma-secretase (Merilahti *et al*, 2017). The results presented here validate the novel
86 UPA in discovering relevant cell signaling modules in previously acquired as well as in freshly
87 acquired multi-omics data. Evidence is provided that the modeled signaling pathways can predict
88 both the cellular function and the direction of the cellular function. Furthermore, we provide the first
89 assessment of cleavage-dependent TYRO3 signaling pathways. The results indicate that the RIP-
90 released ICD of the TYRO3 RTK is capable of activating a unique set of signaling pathways that is
91 different from the canonical TYRO3 signaling activated by the full-length receptor.

92 **Results**

93

94 **The underlying concepts of the unbiased pathway analysis (UPA)**

95 We set out to develop a new robust unbiased method to infer regulatory complexes from different
96 omics data and to combine the complexes into multi-omics pathways. Literature search of previous
97 reports of the key features of omics data predicting molecular associations produced the following
98 observations: First, multiple sources indicate the ability of correlation of expression values to predict
99 molecular association to some extent (Eisen *et al*, 1998; Imamura *et al*, 2010). Secondly, the findings
100 from an extensive human interactome study indicate that conserved stoichiometry between two
101 proteins inside the same sample can be predictive of protein-protein interaction networks (Hein *et al*,
102 2015). From the grounds of these two concepts, the basis for the combined score between two
103 proteins, transcripts or phosphosites in an omics dataset was devised to reflect the strength of
104 association. In Figure 1A the mathematical formulation for the correlation and stoichiometry score is
105 presented. The correlation score was based on the nonparametric Spearman correlation which was
106 consequently ranked to values between 1 and 0 to allow for a linear scale. The stoichiometry score
107 was based on the relation of the third quartile and first quartile value of relative abundances of the
108 two proteins, transcripts or phosphosites in an omics dataset and similarly ranked to a linear scale
109 from 1 to 0. The combined score was expressed as a simple non-weighted multiplication of the
110 correlation and stoichiometry scores similarly ranked to a linear scale between 1 to 0. Nonparametric
111 formulation was chosen to overset technical variation and to desensitize the model to outlier values.

112 To reduce computational demand ($O(n^2)$ for run time), a cut-off for the combined score was
113 devised to consider only strong associations. The following inference of regulatory complexes is
114 sensitive to the cut-off parameter C in a fashion that setting the parameter too high will lead to
115 insufficient models. This can be easily overcome by lowering the parameter C until a lower parameter

116 value will not lead to changes in the model. In the current version of the inference method the
117 parameter C is controlled by the number of associations considered for each molecule in the data.

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

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

135

136 **Validation of regulatory complexes modeled with UPA**

137 The performance of the derived regulatory complex inference method was tested on available
138 interactome (Hein *et al*, 2015), phosphoproteome (Batth *et al*, 2018), and transcriptome (Lachmann

139 *et al*, 2018) data. While molecular associations are context-dependent, certain associations are more
140 frequently reproduced in different contexts and sometimes even stable. Therefore, a level of
141 conservation is expected from a score that would accurately reflect the probability of molecular
142 association. The conservation of the correlation, stoichiometry and combined scores were examined
143 by modeling a set of proteins, phosphosites or transcripts from one dataset into regulatory complexes
144 and examining the median of the corresponding scores derived from the same complexes in another
145 dataset. A median score above 0.5, the median score expected if the scores would be randomly
146 assigned to molecular associations, indicated conservation.

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

152 To validate that the regulatory complexes reflect discovered molecular associations, a method
153 to count the number of known associations in the modeled regulatory complexes was devised. Known
154 protein-protein interactions, kinase substrate phosphosite relationships, and transcription factor target
155 gene annotations were downloaded from STRING (von Mering *et al*, 2003), PhosphoSitePlus
156 (Hornbeck *et al*, 2012) and ENCODE (Davis *et al*, 2018) databases, respectively. The sums of all co-
157 occurrences of proteins known to interact, substrate sites known to be phosphorylated by the same
158 kinase, and transcripts known to be the target gene for the same transcription factor inside the modeled
159 regulatory complexes were used as validation scores. The validation scores were calculated from the
160 regulatory complexes inferred with correlation, stoichiometry and combined scores to compare the
161 effect of the different scores on the accuracy of the model. The validation scores from the real
162 modeled regulatory complexes were compared to the validation scores of randomly formulated
163 regulatory complexes of the same size to calculate the probability that a similar validation score would

164 be acquired by random. The probability density of the validation scores of these randomized
165 regulatory complexes are indicated with a dashed line and the validation scores for the regulatory
166 complexes modeled with different scores with solid lines in the Figure 2B-D. The corresponding P-
167 values drawn from the cumulative probability densities of the randomized regulatory complexes for
168 the validation scores of the regulatory complexes modeled with different scores are visualized in
169 Figure 2E.

170 The regulatory complexes modeled with the combined scores consistently reflected known
171 molecular associations in all interactome, phosphoproteome and transcriptome data. In interactome
172 data, the co-occurrence of two proteins in the same regulatory complex predicted a protein-protein
173 interaction between the two proteins more often than random assignment to regulatory complexes of
174 the same size. Additionally, the median STRING score, that reflects the confidence of the interaction
175 based on the amount and type of evidence on the protein-protein interaction, was consistently higher
176 in the regulatory complexes modeled with the combined score than in the randomized regulatory
177 complexes (Supplementary Figure 1). In the phosphoproteome data, the assignment to regulatory
178 complexes modeled with the combined score indicated that the two phosphorylation sites were
179 controlled by the same kinase more often than by random assignment. In the transcriptome data, in
180 turn, two transcripts in the same regulatory complex modeled with the combined score were more
181 likely to be the target gene for the same transcription factor than randomly chosen transcripts. This
182 effect was similarly observed in transcription factor target gene annotations acquired from REMAP
183 and Literature libraries from the ChEA3 database (Keenan *et al*, 2019) (Supplementary Figure 2).
184 The regulatory complexes modeled with only the correlation or stoichiometry score varied in their
185 ability to find known molecular associations based on the modeled data. The regulatory complexes
186 modeled with stoichiometry score were sufficient, although not as consistent as the regulatory
187 complexes modeled with the combined score, in reflecting known protein-protein interactions in
188 interactome data. However, the regulatory complexes modeled with stoichiometry score failed to

189 reproduce known molecular associations in phosphoproteome and transcriptome data. The regulatory
190 complexes modeled with the correlation score were more or less able to capture known transcription
191 factor target gene relationships in the transcriptome data but were insufficient in finding known
192 protein-protein interactions and kinase substrate relationships in the interactome and
193 phosphoproteome data, respectively. All in all, the regulatory complexes inferred with the combined
194 score consistently captured known molecular associations in all tested omics data, while the
195 complexes inferred with other scores were found lacking in reproducibility and performance across
196 omics datatypes.

197

198 **Validation of regulatory complex combination to pathways by UPA**

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

210 To ensure that the created UPA is also able to combine the inferred regulatory complexes into
211 pathways that reproduce known signaling pathways, a validation method was devised. The sum of all
212 co-occurrences of signaling molecules in separate regulatory complexes combined by the UPA in

213 pathway annotations was used as a validation score for the modeled pathways. The reference pathway
214 annotations were downloaded from the PathwayCommons database (Cerami *et al*, 2011). The
215 validation scores were compared to the validation scores calculated from 10,000 randomly combined
216 pathways of the same size to estimate the probability that random assignment of the regulatory
217 complexes would produce pathways that equally reflect known pathway annotations. The probability
218 density of the validation scores of these randomly combined pathways are indicated with a dashed
219 curve and the validation scores for the regulatory complexes modeled with different scores with solid
220 lines in Figure 3B. The corresponding P-values drawn from the cumulative probability densities of
221 the randomized pathways for the validation scores of the regulatory complexes modeled with
222 different scores and combined with different scores are visualized in Figure 3C. The pathways
223 inferred and combined with the combined score consistently reproduced associations represented in
224 known pathway annotations. The second-best performers were the versions of the analysis method
225 where the pathways were inferred with either the stoichiometry score or the combined score and
226 combined with the stoichiometry score. The correlation score-based versions of the analysis method
227 performed worst in finding molecular associations present in known pathway annotations. Taken
228 together, the results from the validations and the level of conservation of the combined score in other
229 datasets strongly imply that the UPA is able to find molecular associations and pathways that reflect
230 previously observed cell signaling modules.

231

232 **Melanoma cell model to study TYRO3 signaling**

233 Since resources for multi-omics data acquired from the same conditions are limited, a single-time-
234 point multi-omics dataset was collected from samples processed in parallel. To apply the analysis to
235 a biologically relevant context, differences in signaling promoted by different TYRO3 receptor
236 constructs expressed in melanoma cells were addressed. The model of analyzing the signaling of

237 TYRO3 constructs sensitive or resistant to proteolysis-dependent signaling by RIP was chosen to
238 learn how the signaling via a soluble ICD of a RTK differs from the canonical signaling of a full-
239 length RTK. For prediction of how the TYRO3 constructs mutated at the ADAM cleavage sites
240 (TYRO3 ΔADAM) or at the gamma-secretase cleavage site (TYRO3 ΔGS) produce different
241 signaling fragments, please see Figure 4A.

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

253

254 **TYRO3 cleavage site mutants are functional receptors**

255 The functionality of the cloned RIP-resistant TYRO3 mutants was validated by examining their
256 susceptibility to cleavage (Figure 4B), subcellular localization (Figure 4C), phosphorylation (Figure
257 4D), and ability to phosphorylate known substrates (Figure 4E).

258 TYRO3 ΔGS mutant has been shown to be resistant to gamma-secretase activity in HEK293
259 cells (Merilahti *et al*, 2017). As expected, also in the WM-266-4 melanoma cell model, expression of

260 TYRO3 ΔGS promoted accumulation of the membrane-anchored C-terminal fragment (mCTF) to a
261 similar extent as treatment of cells expressing wild-type TYRO3 with a chemical gamma-secretase
262 inhibitor GSI IX (Figure 4B). The expression of the ΔADAM TYRO3 construct in turn did not
263 associate with accumulation of fragments of the size of either the soluble intracellular domain (ICD)
264 or mCTF, consistent with the accumulation of these fragments being secondary to ADAM cleavage
265 (Figure 4A and B). Further in accordance with the model in which proteolytic cleavage of the full-
266 length TYRO3 is a prerequisite for nuclear translocation of soluble ICD fragment (Figure 4A),
267 nuclear accumulation of C-terminal epitopes was reduced when the ΔADAM and ΔGS TYRO3
268 constructs were compared to wild-type TYRO3 in confocal immunofluorescence analyses of WM-
269 266-4 transfectants (Figure 4C).

270 To test whether the different TYRO3 constructs retained autophosphorylation activity, WM-
271 266-4 transfectants were analyzed by TYRO3 immunoprecipitation followed by anti-
272 phosphotyrosine Western analysis. All three variants demonstrated tyrosine phosphorylation of
273 TYRO3 (Figure 4D). Similarly, a Western analysis of the phosphorylation status of STAT3, a known
274 substrate for TYRO3 (Tsai *et al*, 2020), indicated that all three variants stimulated phosphorylation
275 of an activating residue in STAT3 (Figure 4E). These findings indicate that the mutant TYRO3
276 receptors are functional kinases in our cell model and can activate a known downstream component
277 of TYRO3 signaling.

278

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

280 To acquire multi-omics data on cleavage-dependent TYRO3 signaling (Figure 4F), the WM-266-4
281 vector control cells or cells expressing the TYRO3 variants were lysed and subjected to either RNA
282 purification or protein extraction for the analysis of transcriptomes and proteomes, respectively. The
283 extracted proteins were further subjected to His-tag pull-down for the analysis of TYRO3 variant

284 interactomes. For phosphoproteome analysis, the extracted proteins were digested to peptides and
285 subjected to $TiO_2$ enrichment. The extracts for all proteome, interactome and phosphoproteome
286 analyses were subsequently analyzed by LC-MS/MS with data-dependent acquisition. The purified
287 total RNA was transformed into an mRNA sequencing library and sequenced with Illumina
288 HiSeq3000. Statistical testing was utilized to discover the differentially expressed proteins,
289 interacting proteins, transcripts and phosphorylation sites associating with the expression of each
290 TYRO3 variant, as described in Materials and methods.

291 The coprecipitating proteins, proteins, phosphorylation sites and transcripts significantly
292 different ($P \leq 0.05$, fold change $\geq 1.5$ ) from the vector control cells but not between cells expressing
293 the different TYRO3 variants were considered to represent signaling mediated by the full-length
294 TYRO3 (Figure 4A). In contrast, the coprecipitating proteins, proteins, phosphorylation sites and
295 transcripts that were unique for the cells expressing wild-type TYRO3 – i.e. significantly different
296 between cells expressing wild-type TYRO3 and vector control cells or cells expressing TYRO3 ΔGS
297 or TYRO3 ΔADAM – were considered to represent signaling mediated by the released ICD of
298 TYRO3. As a result, a total of 117 differentially coprecipitating proteins (Supplementary Figure 7,
299 Supplemental Table 6), 636 differentially expressed phosphorylation sites (Supplementary Figure 8,
300 Supplementary Table 7), 276 differentially expressed transcripts (Supplementary Figure 10,
301 Supplementary Table 8), and 744 differentially expressed proteins (Supplementary Figure 11,
302 Supplementary Table 9) were discovered to associate with the full-length TYRO3. In turn, a total of
303 343 differentially coprecipitating proteins (Supplementary Figure 7, Supplementary Table 6), 479
304 differentially expressed phosphorylation sites (Supplementary Figure 9, Supplementary Table 7), 242
305 differentially expressed transcripts (Supplementary Figure 10, Supplementary Table 8) and 791
306 differentially expressed proteins (Supplementary Figure 12, Supplementary Table 9) were discovered
307 to associate with the ICD of TYRO3.

308

309 **UPA of TYRO3 signaling**

310 The differentially expressed signaling events associated with the expression of full-length TYRO3 or
311 TYRO3 ICD were analyzed with the UPA first to extract regulatory complexes. The UPA discovered
312 24 previously undescribed interactome, 110 phosphoproteome, 46 transcriptome and 130 proteome
313 regulatory complexes that associated with the expression of full-length TYRO3 (Supplementary
314 Figure 4, Supplementary Table 2). The expression of TYRO3 ICD in turn was associated with 53
315 new interactome, 84 phosphoproteome, 45 transcriptome and 136 proteome regulatory complexes
316 (Supplementary Figure 5, Supplementary Table 2). The inferred regulatory complexes were further
317 combined with the UPA into pathways, resulting in a total of 51 and 46 unique signaling pathways
318 associating with the signaling stimulated by the full-length and the released ICD of TYRO3,
319 respectively (Supplementary Table 1).

320

321 **Functional enrichment analysis for pathways inferred with the UPA**

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

332

333 **Functional validation of the functional enrichment analyses of the UPA**

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

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

350 Finally, the two-dimensional cellular morphology of the WM-266-4 transfectants were
351 examined from thresholded confocal images with the ImageJ MorphoLibJ plugin (Figure 6C). As
352 indicated by the ICD pathway 46 and the full-length pathway 9 (Figure 5D), the cells expressing
353 wild-type TYRO3 exhibited distinct morphology from the vector control cells or cells expressing
354 either the ΔGS or ΔADAM variant of TYRO3 (Figure 6C). The morphological difference was
355 captured by the convexity measure (the ratio between the cell area and its convex area), the lower

356 value of which suggests a shape with more protrusions. A greater amount of cell protrusions in the
357 cells expressing wild-type TYRO3 was also predicted by the UPA (ICD pathway 1 in Figure 5D).
358 Furthermore, the WM-266-4 cells expressing wild-type TYRO3 or either of the TYRO3 mutants
359 demonstrated less circular cell shape than the vector control cells, indicating that the full-length
360 TYRO3 additionally regulates cell morphology as suggested by the full-length pathway 9
361 (Supplementary Figure 4). Taken together, these validation experiments suggest that the function
362 enrichment of the pathways modeled with the UPA are able to predict both the function and the
363 direction of the function from purely molecular multi-omics data.

364

365 **Transcription factor enrichment analysis function of UPA**

366 Since the UPA was able to predict transcription factor target gene relationships in available
367 transcriptome data (Figure 2D-E and Supplementary Figure 2), a transcription factor enrichment
368 analysis was created (see Materials and methods for details) to predict a transcription factor for each
369 transcriptome regulatory complex. To ensure the validity of the transcription factor predictions, the
370 predicted transcription factors were compared to transcription factors known to be regulated by
371 TYRO3. Indeed, several transcription factors known to be regulated by TYRO3 were predicted to
372 regulate the modeled regulatory transcriptome complexes, such as STAT3 (Tsai *et al*, 2020), MYC
373 (Chen *et al*, 2020; Dufour *et al*, 2019), and MITF (Zhu *et al*, 2009a) (Supplementary Table 3).
374 Furthermore, the most common transcription factor predicted to regulate the transcriptional
375 complexes associated with TYRO3 signaling was MYC, with the respective gene amplified in the
376 WM-266-4 cell background (Kraehn *et al*, 2001). These observations indicate that the transcription
377 factor prediction function was able to produce both molecule and cellular context specific suggestions
378 of upstream transcription factors for the regulatory transcriptome complexes.

379

380 **Subcellular location enrichment analysis function of UPA**

381 To predict a subcellular compartment for each interactome complex inferred with UPA, a subcellular
382 location enrichment analysis for interactome complexes was devised (see Materials and methods).
383 The basis of the subcellular location enrichment analysis relies on two observations. First, the ability
384 of UPA to discover previously reported protein-protein interactions in available interactome data
385 (Figure 2B,E). Secondly, the need for shared physical location in order for protein-protein interactions
386 to occur. The predicted subcellular localizations of the interactome complexes of full-length and the
387 ICD of TYRO3 (Supplementary Table 4) were compared to the reported localization difference
388 between the wild-type and the ΔGS and ΔADAM variants of TYRO3 (Figure 4C) (Merilahti *et al*,
389 2017) to ensure the validity of the predicted subcellular locations. The differences in the predicted
390 localizations of the interactome complexes reflected the known difference in the subcellular
391 localization of full-length and ICD of TYRO3 (Supplementary Table 4). Thirteen out of the 54 (24%)
392 ICD interactome complexes were predicted to localize into the nucleus, while nuclear localization
393 was predicted for only 2 out of the 21 (8 %) full-length interactome complexes ($P = 0.032$ against
394 frequency of nuclear assignment in the interactome complexes of TYRO3 ICD). Moreover, 8 out of
395 the 21 (33%) full-length interactome complexes, but only 5 out of the 54 (9%) ICD interactome
396 complexes, were predicted to localize into the plasma membrane ($P < 0.0001$ against frequency of
397 plasma membrane assignment in the interactome complexes of full-length TYRO3). Both the full-
398 length TYRO3 interactome complexes (5 out of 21; 20%) and the ICD interactome complexes (11
399 out of 54, 20%) were equally predicted to be localized in the cytoskeletal structures ($P = 0.0025$ for
400 full-length TYRO3 and $P < 0.0001$ for TYRO3 ICD against frequency of cytoskeleton assignment in
401 randomly modelled interactome complexes). These findings indicate that the subcellular location
402 prediction method was able to recapitulate previously reported subcellular location differences.

403

404 **Kinase enrichment analysis function of UPA**

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

420

421 **UPA enables the discovery of previously reported and novel molecular associations**

422 Since the modeled pathways are not limited by the existing pathway annotations, the UPA allows for
423 the discovery of novel molecular associations involved in a biological process. To examine the novel
424 molecular associations indicated to be involved in the same biological process by the UPA, selected
425 signaling pathways of full-length and ICD of TYRO3 were visualized (Figure 6D-E). The cell
426 motility pathway 29 of TYRO3 ICD included proteins, transcripts and phosphorylated residues of
427 proteins known to associate with motility or migration in almost all included regulatory complexes

428 (APC2 in INT_35; DNAH2, GRB2, MCU in INT_50; HMGB1, NAV3 in PHOS_ST_35; CRK in
429 PHOS_ST_50; SLC7A6 in TRANS_17, LYST in PROT_100; GPC1, SEMA6A, STK4 in
430 PROT_128 and COL18A1 in PROT_134 in Figure 6D). Similarly, proteins and phosphorylated
431 residues of proteins known to associate with cytoskeleton organization were found in almost all
432 included regulatory complexes in the cytoskeleton organization pathway 9 of full-length TYRO3
433 (CORO1B in INT_1; MYH9 and RAB11A in PHOS_Y_5; SH3D19 in PHOS_ST_1; MARCKSL1
434 and PLEC in PHOS_ST_31; IQGAP1 in PHOS_ST_39; MAP1B, MARK3 and SEPT10 in
435 PHOS_ST_68 in Figure 6E). The interplay of most of these molecules in regulating cell motility or
436 cytoskeleton organization has not been described, indicating that the UPA provides indications of
437 novel signaling modules. The cytoskeleton organization pathway 9 of full-length TYRO3 (Figure 6E)
438 paired the interactome complex INT_1 containing TYRO3 to the phosphoproteome complex
439 PHOS_Y_5 containing MERTK and Rab proteins. Since TAM RTKs have been indicated to
440 heterodimerize and cross-phosphorylate (Seitz *et al*, 2007; Brown *et al*, 2012) and Y682 is identified
441 as an autophosphorylation site for MERTK , the discovery of this signaling by the UPA enforces the
442 credibility of the method in additionally discovering signaling modules that reflect known cell
443 signaling events.

444

445 **Discussion**

446

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

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

465 Since the UPA is based on the inherent variation between samples in the omics data, the
466 quality of the omics data poses a limitation to the analysis. Data wrought with technical variation, or
467 inflated with missing values or lack of repetitions, will lead to loss of model accuracy. The

468 nonparametric formulation and a zero inflated version of the stoichiometry score was devised to battle
469 two of these limitations. Since the correlation, stoichiometry and combined scores are ranked and
470 thus dependent on the choice of the signaling molecules included in the dataset, a choice of the cut-
471 offs for the data to be modeled will also significantly affect the final modeled pathways. Involving or
472 omitting proteins, post-translational modifications or transcripts from the omics data used for the
473 analysis will influence the inferred final pathways. For accurate inference of the multi-omics
474 pathways, the multi-omics data should be acquired from the same samples. The UPA does not support
475 combination of multi-omics data that have been acquired from different samples even though they
476 would represent the same tissue type or cell line.

477 We acquired multi-omics data from the RIP-resistant and wild-type TYRO3 variants and
478 utilized the UPA to shed light on the unknown RIP-associated signaling pathways of the TYRO3
479 RTK in melanoma cells. The UPA discovered a total of 51 and 46 differential unique signaling
480 pathways for the full-length TYRO3 and the RIP-released TYRO3 ICD, respectively. An enrichment
481 analysis to infer the biological process the pathways were involved with was also devised. A
482 difference in the amount and the direction of the pathways associated with growth and cell cycle, cell
483 motility and adhesion, immune response, chromosome organization, cell death, and cell
484 differentiation were observed between the full-length and RIP-released TYRO3 ICD. Many of these
485 functions may be critically involved e.g. in the progression or therapeutic responses in melanoma.
486 For example, TYRO3 has been implicated in the proliferation, tumorigenesis, chemoresistance and
487 motility of melanoma cells (Demarest *et al*, 2013; Shao *et al*, 2018; Zhu *et al*, 2009a; Tworkoski *et*
488 *al*, 2011). Validation experiments confirmed that the UPA with the function prediction was able to
489 accurately predict both the function and the direction of the function of the cells expressing either
490 wild-type or cleavage-impaired variants of TYRO3. These findings also represent the first
491 comprehensive characterization of signaling pathways stimulated by the gamma-secretase released
492 ICD of a RTK as compared to the canonical signaling *via* the full-length receptor.

493 The UPA segregates all signaling events into pathways independent on the amount of previous
494 knowledge on the modeled molecules. This feature is crucial in discovery of novel signaling modules.
495 Since the UPA is able to predict molecular associations, it can predict upstream transcription factors
496 of transcripts, kinases of phosphorylation sites, and subcellular locations of proteins in the regulatory
497 complexes, for which these features are unknown. Similarly, the UPA can predict connections to
498 biological functions for proteins, transcripts or post-translationally modified proteins in the pathways
499 for which no previous connection to the biological function has been discovered. The pathways
500 inferred with UPA can be re-entered into UPA to acquire super pathways and thus broader
501 connections between the modelled pathways can additionally be discovered with the UPA.

502

503

504 # Materials and methods

505

506 **Cell culture and transfections**

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

513

514 **ADAM10 and ADAM17 cleavage site prediction**

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

527 from the transmembrane domain was added, due to known ectodomain size constraints of gamma-
528 secretase substrates (Funamoto *et al*, 2013). The method was validated by ensuring its capability in
529 finding published cleavage sites in the peptides described in Goth *et al*, 2015. P408/L409 and
530 A418/G419 of TYRO3 were predicted to serve as cleavage sites with highest predicted relative
531 cleavage site probabilities for both ADAM10 and ADAM17.

532

533 **Plasmids and cloning**

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

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

549

550 **Antibodies**

551 Anti-TYRO3 (5585), anti-V5 (13202), anti-phospho-STAT3 (8862), and anti-STAT3 (9972)
552 antibodies were purchased from Cell Signaling Technology. Anti-actin (MA1-744) antibodies were
553 purchased from Thermo Fisher Scientific. Anti-β-tubulin (T7816) and anti-phosphotyrosine (4G10;
554 5-321) antibodies were purchased from Sigma-Aldrich.

555

556 **Immunoprecipitation and Western analysis**

557 For immunoprecipitation experiments, WM-266-4 vector control cells or cells expressing Myc-
558 tagged TYRO3 constructs were incubated in serum-free conditions overnight before lysis into lysis
559 buffer (0.1% Triton X-100, 10 mM Tris-Cl pH 7.4, 150 mM NaCl including Pierce protease and
560 phosphatase inhibitor mini tablet (Thermo Fisher Scientific)). The cell lysates were pre-cleared with
561 50 μl Pierce protein G magnetic beads (Thermo Fisher Scientific) at 4°C for 1 hour and subjected to
562 affinity enrichment with Pierce anti-c-Myc magnetic beads (Thermo Fisher Scientific) at 4°C
563 overnight. The beads were washed five times with 5×TBS-T buffer (125 mM Tris, 750 mM NaCl,
564 0.25% Tween-20) and once with ultrapure water. The proteins were denatured and eluted from the
565 beads by incubation at 95°C for 10 minutes in SDS-PAGE loading buffer.

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

570

571 **Generation of stable TYRO3 knock-down cell lines**

572 For stable downregulation of TYRO3, TYRO3-targeted lentiviral shRNA plasmid
573 TRCN0000231528
574 (CCGGTTGGTATCTCAGGTCTGAATCCTCGAGGATTCAGACCTGAGATACCAATTTTTG)
575 (MISSION, Sigma-Aldrich) was used. The control lentiviral shRNA plasmid (Addgene plasmid
576 #1864) was a kind gift from David Sabatini (Sarbassov *et al*, 2005). Third generation lentiviral
577 packaging system (Addgene) was used in HEK293T cells to produce shRNA-carrying lentiviruses.
578 Growth medium was changed every 24 hours post-transfection. Virus-containing medium was
579 collected after 48 and 72 hours, filtered and titered. The virus-containing medium was used to infect
580 WM-266-4 cells at a multiplicity of infection of 2 in the presence of 8 μg/ml Polybrene (Sigma-
581 Aldrich). Cells were maintained in the presence of 1 μg/ml puromycin (Sigma-Aldrich) to select the
582 cells that stably express the lentiviral shRNA plasmids.

583

584 **Immunofluorescence and confocal microscopy**

585 To detect ectopically expressed V5-tagged TYRO3 constructs, WM-266-4 transfectants were
586 cultured on coverslips in serum-free conditions overnight. The cells were fixed with 3%
587 paraformaldehyde, and permeabilized with 0.1% Triton X-100. Cells were stained with anti-V5
588 (13202; Cell Signaling Technologies) and AlexaFluor 555 goat anti-rabbit (Molecular Probes). After
589 labeling the nuclei with 4',6-diamid-ino-2-phenylindole (DAPI; Sigma-Aldrich), the cells were
590 mounted with Mowiol 40-88 (Sigma-Aldrich). Images were acquired with Zeiss LSM 880 confocal
591 microscope (Zeiss). EzColocalization plugin (Stauffer *et al*, 2018) in ImageJ (version 1.53c) was used
592 to analyze colocalization from confocal slices taken from the middle of the nucleus. The
593 colocalization of V5-signals with DAPI was calculated using Pearson correlation coefficient
594 (Manders *et al*, 1992) to measure the nuclear localization of TYRO3. Morphological analysis were

595 carried out from confocal slices in plane with the plasma membrane using MorphoLibJ v 1.4.2.1
596 (Legland *et al*, 2016) plugin in Image J.

597

598 **Live-cell imaging of cell growth**

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

602

603 **Real-time cell adhesion assay**

604 The WM-266-4 cell transfectants were detached using 5 mM EDTA 6 to 12 hours after transfection.
605 The cells were plated at 15,000 cells/well onto fibronectin-coated (5 μg/cm$^2$) xCELLigence E-plate
606 (Agilent) wells in serum free medium. Real time cell impedance was measured with the
607 xCELLigence RTCA analyzer (Agilent) for 24 hours. The resulting cell index values were normalized
608 to the number of cells remaining in the wells after 24 hours.

609

610 **RNA-sequencing**

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

613 The quality of the RNA samples was ensured using Advanced Analytical Fragment Analyzer
614 (Agilent). The sequencing library was created with 300 ng of sample with TruSeq Stranded mRNA
615 HT Kit (Illumina) and indexed with IDT for Illumina TruSeq RNA UD Indices according to
616 manufacturer's protocol. Genome-wide strand-specific RNA-sequencing (RNA-seq) was performed

617 at Turku Bioscience Centre sequencing core with Illumina HiSeq3000 using 75 bp paired-end
618 reading.

619

620 **Mass spectrometry sample preparation**

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

628

629 **Affinity enrichment of mass spectrometry samples**

630 The WM-266-4 cell transfectants were subjected to protein crosslinking with 2 mM DTBP for 10
631 minutes. The crosslinking reaction was quenched by incubation with 50 mM Tris-HCl pH 7.5 for 15
632 minutes before cell lysis. Equal amounts of WM-266-4 cell lysates were pre-cleared with Pierce
633 protein G magnetic beads (Thermo Fisher Scientific) at 4 °C for 1 hour and subjected to affinity
634 enrichment with HisPur Ni-NTA magnetic beads (Thermo Fisher Scientific) at 4 °C overnight. The
635 beads were washed five times with 5×TBS-T buffer (125 mM Tris, 750 mM NaCl, 0.25% Tween-
636 20) and once with ultrapure water. The proteins were denatured, alkylated and eluted from the beads
637 at 95 °C for 10 minutes in elution buffer (6 M guanidine hydrochloride, 100 mM Tris-HCl pH 8.5,
638 10 mM Tris(2-carboxyethyl)phosphine (TCEP), 10 mM chloroacetamide).

639

640 **Protein digestion to peptides**

641 Proteins enriched in affinity enrichment or purified cell lysates were digested with lys-C
642 (enzyme/protein ratio 1:100) for 1 hour at 37 °C. The samples were diluted 1:10 with 50 mM
643 $NH_4HCO_3$ and digested with trypsin (enzyme/protein ratio 1:100) at 37 °C overnight.

644

645 **Sample desalting**

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

652

653 **Phosphopeptide enrichment**

654 Phosphopeptides were enriched from desalted and dried peptides using Pierce high-select $TiO_2$
655 phosphopeptide enrichment kit (Thermo Fisher Scientific) according the manufacturer's instructions.
656 Following elution, phosphopeptides were dried with Hetovac vacuum centrifuge and stored as dry in
657 -20 °C until analysis with mass spectrometer.

658

659 **Mass spectrometry**

660 Dried peptide samples were resuspended in 0.1% formic acid and sample concentrations were
661 measured with Nanodrop 1000 (Thermo Fisher Scientific). Equal amounts of samples were analyzed
662 on an Easy-nLC 1000 coupled to a Orbitrap Fusion Lumos instrument (Thermo Fisher Scientific) at
663 Turku Bioscience Centre Proteomics Core Services. Peptides were loaded on in-house packed 100
664 μm × 2 cm precolumn packed with ReproSil-Pur 5 μm 200 Å C18-AQ beads (Dr. Maisch) using 0.1
665 % formic acid in water (buffer A) and separated by reverse phase chromatography on a 75 μm × 15
666 cm analytical column packed with ReproSil-Pur 5 μm 200 Å C18-AQ beads (Dr. Maisch). All
667 separations were performed using a 60 minute gradient ranging from 8% buffer B (80% acetonitrile
668 in 0.1% formic acid) to 21% in buffer B in 28 minutes and to 36% buffer B in 22 minutes and ramped
669 to 100% buffer B in 5 minutes at flow rate of 300 nl/minute. The washout followed at 100% buffer
670 B for 5 minutes.

671 All MS spectra were acquired on the orbitrap mass analyzer and stored in centroid mode. For
672 data-dependent acquisition experiments, full MS scans were acquired from 300 to 1600 m/z at
673 120,000 resolution with fill target of 7E5 ions and maximum injection time of 50 ms. The most
674 abundant ions on the full MS scan were selected for fragmentation using 1.6 m/z precursor isolation
675 window and beam-type collisional-activation dissociation (HCD) with 30% normalized collision
676 energy for a cycle time of 3 seconds. MS/MS spectra were collected at 15,000 resolution with fill
677 target of 5E4 ions and maximum injection time of 100 ms. Fragmented precursors were dynamically
678 excluded from selection for 35 seconds.

679

680 **Protein identification and quantification**

681 MS/MS spectra were searched with Metamorpheus (version 0.0.304) (Solntsev *et al*, 2018) against
682 human proteome containing known post-translational modifications (downloaded from Uniprot on
683 19.2.2019). The mass spectrometry files were calibrated, possible post-translational modifications
684 were searched and peptides and proteins were identified and quantified using FlashLFQ algorithm
685 (Millikin *et al*, 2018). The following parameters were used for the post-translational modifications
686 search. Cysteine carbamidomethylation and methionine oxidation were set as constant and variable
687 modifications, respectively and G-DPM search in Metamorpheus was additionally used to discover
688 other modifications. Search results were filtered to a 1% FDR at PSM level. Peptides were accepted
689 with search engine score above 5.

690

691 **RNAseq data analysis**

692 The RNAseq reads were quality checked with FastQC (Babraham Bioinformatics), quality and
693 adapter trimmed with PRINSEQ (Schmieder & Edwards, 2011) and Trimmomatic (Bolger *et al*,
694 2014), and pseudoaligned with kallisto v 0.46.0 (Bray *et al*, 2016) to human transcriptome Ensembl
695 v96 (Yates *et al*, 2020) to retrieve TPM (transcripts per million) values. The batch effect between
696 experiment 1 and experiments 2 and 3 was corrected with batchelor (Haghverdi *et al*, 2018) and the
697 TPM values were normalized by library size. Differential expression between the vector control
698 sample and samples representing the different variants of TYRO3 was analyzed with DeSeq2 (Love
699 *et al*, 2014). Transcripts with fold change over 1.5 and FDR-adjusted P-value lower or equal than
700 0.05 were chosen for further analysis. The transcripts significantly different between the vector
701 control sample and all of the samples representing the TYRO3 variants (wild-type, ΔADAM or ΔGS)
702 were considered to reflect the signaling mediated by the full-length TYRO3 receptor. The transcripts
703 significantly different between the vector control sample and the sample representing wild-type

704 TYRO3, but not between the vector control sample and the samples representing ΔADAM or ΔGS
705 variant of TYRO3, were considered to reflect the signaling mediated by the soluble ICD of TYRO3.

706

707 **Mass spectrometry data analysis**

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

719

720 **The inference of regulatory complexes and their combination in the UPA**

721 Matlab R2016a (MathWorks) was used as the coding environment. The pair-wise association
722 between two proteins, phosphosites or transcripts was determined by using a combined score. The
723 combined score is derived from the multiplication of the correlation and the stoichiometry score the
724 mathematical formulations of which are presented in Figure 1A. The correlation score for all the
725 possible protein, phosphosite or transcript combinations was calculated with the Spearman
726 correlation and ranked to derive scores from 1 to 0 in equal increments. The stoichiometry score for

727 all the possible protein, phosphosite or transcript combinations was calculated by dividing the third
728 quartile value ($Q_3$) of relative abundances with the first quartile ($Q_1$) value and similarly ranked to
729 derive scores from 1 to 0 in equal increments. Two versions of the stoichiometry score were devised
730 for non-zero-inflated and zero-inflated data. In the non-zero-inflated version of the stoichiometry
731 score, relative abundance data were considered only from samples in which both intracellular
732 molecules were present. In the zero-inflated version of the stoichiometry score (weighted
733 stoichiometry score), missing values in the same sample of only one of the intracellular molecules
734 was punished by inflating the stoichiometry score and missing values in the same sample of both
735 intracellular molecules was rewarded by deflating the stoichiometry score.

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

741 The regulatory complex inference analysis was designed to utilize the nearest neighbor
742 concept to find the highest scoring (sum of the combined score between all 3 members) 3 member
743 networks for all intracellular molecules in the list and to combine and trim the networks based on
744 common network members and combined scores as described in the following order: 1) Networks
745 with 2 common members are combined until no two complexes with 2 common members can be
746 found. 2) Networks with only one common member are combined if the size of both networks is
747 under or equal to 4 and the network score is higher or equal to 2 indicating a strong association. The
748 parameters for this joining (maximum size of a complex to be joined with only one common node
749 and minimum complex score for the complexes to be joined by only one common node) can be set if
750 smaller regulatory complexes or regulatory complexes with a higher association scores are preferred.
751 3) Networks with 2 common members are again combined until no two complexes with 2 common

752 members can be found. 4) Two out of the 3 members of the remaining 3 member networks are
753 removed if 2 out of three members are already present in another complex. 5) The remaining 3
754 member networks are joined if they still have a common member. This step can be omitted if smaller
755 regulatory complexes are preferred. The analysis allows for a protein, phosphosite or a transcript to
756 be a member for more than one regulatory complex.

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

763

764 **Validation of regulatory complexes and their combination**

765 The data on known protein-protein interactions, kinase substrate phosphosite relationships and
766 transcription factor target gene relationships were acquired from the STRING, PhosphoSitePlus and
767 ChEA3 databases, respectively (von Mering *et al*, 2003; Hornbeck *et al*, 2012; Keenan *et al*, 2019;
768 Davis *et al*, 2018; Chèneby *et al*, 2018). The data on annotated signaling pathways was acquired from
769 PathwayCommons (Cerami *et al*, 2011). The validation interactome, phosphoproteome and
770 transcription data were acquired from the publications of Batth et al and Karayel et al, and from
771 ArchS4 database, respectively (Lachmann *et al*, 2018; Batth *et al*, 2018; Karayel *et al*, 2020). The
772 validation data for the regulatory complex combination to pathways was acquired from the
773 LinkedOmics database (Vasaikar *et al*, 2018), where a combination of either proteomics,
774 transcriptomics, methylomics, phosphoproteomics, protein array or acetylation data from the same
775 samples were modeled and combined. The sum of all protein-protein interactions, kinase substrate

776 phosphosite relationships or transcription factor target gene relationships in the modeled regulatory
777 complexes was used as the validation score for interactome, phosphoproteome and transcriptome
778 regulatory complex validation, respectively. The sum of co-occurrences of two complex members
779 from two different combined complexes in the same pathway annotation for all the combined
780 complexes was used as a validation score for regulatory complex combination. The validation score
781 was compared to the sum of all protein-protein interactions, kinase substrate phosphosite
782 relationships, transcription factor target gene relationships or co-occurrences of two complex
783 members from two different randomly combined complexes in the same pathway annotation in
784 randomly modeled regulatory complexes or combined pathways of same size as in the modeled data
785 to derive an empirical probability density function from 1000 rounds of simulation. An Epanechnikov
786 kernel was used to fit the simulated data into the empirical probability density function. The
787 corresponding P-values were drawn from the empirical cumulative distribution function. To assess
788 the effect of different scores on the accuracy of the modeled regulatory complexes and pathways, the
789 complexes were modeled with either correlation score, stoichiometry score or the combined score
790 and combined with either correlation score, stoichiometry score or the combined score. The
791 conservation of the different scores was assessed in each data type by deriving the median score of
792 the modeled complexes in different datasets from the one used for the initial modeling.

793

794 **Transcription factor and kinase prediction**

795 An enrichment analysis to predict a transcription factor for each transcriptome and a kinase for each
796 phosproteome regulatory complex was devised. Transcription factor target gene annotations from the
797 ChEA3 data resource (Keenan *et al*, 2019) were used and only the target gene annotations of the
798 transcription factors present in the dataset with at least 5 transcripts were considered. The annotations
799 of the upstream kinase for each phosphosite was acquired and combined from the results of a kinase

800 knock-down screen (Sugiyama *et al*, 2019), RegPhos 2.0 (Huang *et al*, 2014), PhosphoSitePlus
801 (Hornbeck *et al*, 2012), and KinaseNET (Kinexus BioInformatics Corporation) databases. The
802 annotations of upstream kinases based on protein-protein interactions were accumulated from
803 BioGrid (Oughtred *et al*, 2021), Cheng et al. (Cheng *et al*, 2014), Harmonizome (Rouillard *et al*,
804 2016), HIPPIE (Schaefer *et al*, 2012), Mentha (Calderone *et al*, 2013), Mint (Zanzoni *et al*, 2002),
805 PIPs (McDowall *et al*, 2009), PSOPIA (Murakami & Mizuguchi, 2014), Reactome (Jassal *et al*,
806 2020), and STRING (von Mering *et al*, 2003) databases. The kinase annotations were filtered
807 according to the LFQ intensity of the kinases in the proteome and phosphoproteome data. Only
808 substrate phosphosite or protein-protein interaction annotations of the kinases that were present in the
809 corresponding samples with at least the LFQ intensity of 10000 were considered.

810 The enrichment analyses were created to follow the 3 steps as described: Step 1) The
811 annotation most enriched in all the transcripts or phosphosites in all the complexes is given to the
812 complexes for which the annotation gives the highest statistically significant enrichment score (the
813 number of transcripts or phosphorylated residues found in the annotation divided by the number of
814 transcripts or phosphorylated residues in the complex; P-value equal or less than the set threshold).
815 The most enriched annotation and the complexes for which the annotation gives the highest
816 enrichment score are removed. For the remaining dataset the same process is repeated until no
817 complex remains. If more than one annotation provides the same overall and complex specific
818 enrichment, all annotations are given as equally likely to the complexes. To derive a P-value for the
819 enrichment of each annotation in each regulatory complex 10,000 random sets of equal size from all
820 transcripts or phosphorylated residues identified in the transcriptome or phosphoproteome data are
821 drawn. The respective enrichment of each of these randomized sets is derived and a probability
822 density function is fitted with an Epanechnikov kernel to the enrichment scores. The P-values for the
823 enrichment are drawn from the corresponding empirical cumulative distribution function. Step 2)
824 Secondary transcription factor and kinase predictions are performed by assessing whether any of the

825 primary transcription factors or kinases would be statistically significantly enriched also in the
826 regulatory complexes were another transcription factor or kinase is assigned as the primary. P-values
827 are similarly acquired for the secondary transcription factors and kinases. If more than one annotation
828 provides the same complex specific enrichment with a P-value lower than 0.05, all annotations are
829 given as equally likely to the complex. Step 3) If any unannotated complexes remain, all remaining
830 annotations are searched to find the first statistically significant annotation with the highest
831 enrichment score. If more than one annotation provides the same enrichment score with a P-value
832 lower than the set threshold, all annotations are given as equally likely to the complex.

833

834 **Subcellular location prediction**

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

846

847 **Function prediction**

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

861

862 **Information visualization and statistics**

863 For statistical analysis, the GraphPad Prism software (Version 9.02; GraphPad Software) and Matlab
864 2016a were utilized. The details of the statistical testing are described in figure legends. All datasets
865 were tested for normality and parametric or non-parametric testing was used accordingly. Heatmaps
866 were generated with Morpheus (https://software.broadinstitute.org/morpheus). The visualizations of
867 the inferred regulatory complexes and pathways were created with Cytoscape (Version 3.8.2;
868 Shannon *et al*, 2003) and Affinity Designer (Version 19.0; Serif)

869

870 **Data availability**

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

875

## 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

895 Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, Buettner F, Huber W & Stegle O
896 (2018) Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics
897 data sets. *Mol Syst Biol* 14

898 Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight
899 SS, Eppig JT, *et al* (2000) Gene ontology: Tool for the unification of biology. *Nat Genet* 25: 25–
900 29 doi:10.1038/75556 [PREPRINT]

901 Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, Alpi E, Bowler-Barnett EH,
902 Britto R, Bursteinas B, *et al* (2021) UniProt: The universal protein knowledgebase in 2021.
903 *Nucleic Acids Res* 49: D480–D489

904 Batth TS, Papetti M, Pfeiffer A, Tollenaere MAX, Francavilla C & Olsen J V. (2018) Large-Scale
905 Phosphoproteomics Reveals Shp-2 Phosphatase-Dependent Regulators of Pdgf Receptor
906 Signaling. *Cell Rep* 22: 2784–2796

907 Binder JX, Pletscher-Frankild S, Tsafou K, Stolte C, O'Donoghue SI, Schneider R & Jensen LJ
908 (2014) COMPARTMENTS: unification and visualization of protein subcellular localization
909 evidence. *Database* 2014: bau012–bau012

910 Bolger AM, Lohse M & Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence
911 data. *Bioinformatics* 30: 2114–2120

912 Bray NL, Pimentel H, Melsted P & Pachter L (2016) Near-optimal probabilistic RNA-seq
913 quantification. *Nat Biotechnol* 34: 525–527

914 Brown JE, Krodel M, Pazos M, Lai C & Prieto AL (2012) Cross-phosphorylation, signaling and
915 proliferative functions of the Tyro3 and Axl receptors in Rat2 cells. *PLoS One* 7: 1–11

916 Caescu CI, Jeschke GR & Turk BE (2009) Active-site determinants of substrate recognition by the
917 metalloproteinases TACE and ADAM10. *Biochem J* 424: 79–88

918 Calderone A, Castagnoli L & Cesareni G (2013) mentha: a resource for browsing integrated protein-
919 interaction networks. *Nat Methods* 10: 690–691

920 Carbon S, Douglass E, Good BM, Unni DR, Harris NL, Mungall CJ, Basu S, Chisholm RL, Dodson
921 RJ, Hartline E, *et al* (2021) The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids*
922 *Res* 49: D325–D334

923 Care MA, Westhead DR & Tooze RM (2019) Parsimonious Gene Correlation Network Analysis
924 (PGCNA): a tool to define modular gene co-expression for refined molecular stratification in
925 cancer. *npj Syst Biol Appl* 5: 1–17

926 Carpenter G & Liao HJ (2013) Receptor tyrosine kinases in the nucleus. *Cold Spring*
927 *HarbPerspectBiol* 5: a008979

928 Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur Ö, Anwar N, Schultz N, Bader GD & Sander
929 C (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res* 39:
930 685–690

931 Chen D, Liu Q, Cao G & Zhang W (2020) TYRO3 facilitates cell growth and metastasis via activation
932 of the Wnt/β-catenin signaling pathway in human gastric cancer cells. *Aging (Albany NY)* 12:
933 2261–2274

934 Chen S & Mar JC (2018) Evaluating methods of inferring gene regulatory networks highlights their
935 lack of performance for single cell gene expression data. *BMC Bioinformatics* 19: 232

936 Chèneby J, Gheorghe M, Artufel M, Mathelier A & Ballester B (2018) ReMap 2018: an updated atlas
937 of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments.
938 *Nucleic Acids Res* 46: D267–D275

939 Cheng F, Jia P, Wang Q & Zhao Z (2014) Quantitative network mapping of the human kinome
940 interactome reveals new clues for rational kinase inhibitor discovery and individualized cancer
941 therapy. *Oncotarget* 5: 3697

942 Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov
943 UK, Narayanan AK, *et al* (2018) The Encyclopedia of DNA elements (ENCODE): Data portal
944 update. *Nucleic Acids Res* 46: D794–D801

945 Demarest SJ, Gardner J, Vendel MC, Ailor E, Szak S, Huang F, Doern A, Tan X, Yang W,
946 Grueneberg DA, *et al* (2013) Evaluation of Tyro3 expression, Gas6-mediated akt
947 phosphorylation, and the impact of anti-Tyro3 antibodies in melanoma cell lines. *Biochemistry*
948 52: 3102–3118

949 Dufour F, Silina L, Neyret-Kahn H, Moreno-Vega A, Krucker C, Karboul N, Dorland-Galliot M,
950 Maillé P, Chapeaublanc E, Allory Y, *et al* (2019) TYRO3 as a molecular target for growth
951 inhibition and apoptosis induction in bladder cancer. *Br J Cancer* 120: 555–564

952 Eisen MB, Spellman PT, Brown PO & Botstein D (1998) Cluster analysis and display of genome-
953 wide expression patterns. *Proc Natl Acad Sci U S A* 95: 14863–8

954 Fujita K, Chen X, Homma H, Tagawa K, Amano M, Saito A, Imoto S, Akatsu H, Hashizume Y,
955 Kaibuchi K, *et al* (2018) Targeting Tyro3 ameliorates a model of PGRN-mutant FTLD-TDP via
956 tau-mediated synaptic pathology. *Nat Commun 2018 91* 9: 1–21

957 Funamoto S, Sasaki T, Ishihara S, Nobuhara M, Nakano M, Watanabe-Takahashi M, Saito T, Kakuda
958 N, Miyasaka T, Nishikawa K, *et al* (2013) Substrate ectodomain is critical for substrate
959 preference and inhibition of γ-secretase. *Nat Commun* 4: 2529

960 Goth CK, Halim A, Khetarpal SA, Rader DJ, Clausen H & Schjoldager KT-BGT-BG (2015) A
961 systematic study of modulation of ADAM-mediated ectodomain shedding by site-specific O-
962 glycosylation. *Proc Natl Acad Sci U S A* 112: 14623–8

963 Graham DK, DeRyckere D, Davies KD & Earp HS (2014) The TAM family: phosphatidylserine-
964 sensing receptor tyrosine kinases gone awry in cancer. *Nat Rev Cancer* 14: 769–785

965 Güner G & Lichtenthaler SF (2020) The substrate repertoire of γ-secretase/presenilin. *Semin Cell Dev*
966 *Biol* 105: 27–42

967 Guo F, Chiang M-Y, Wang Y & Zhang Y-Z (2008) An in vitro recombination method to convert
968 restriction- and ligation-independent expression vectors. *Biotechnol J* 3: 370–377

969 Guo H, Barrett TM, Zhong Z, Fernndez JA, Griffin JH, Freeman RS & Zlokovic B V. (2011) Protein
970 S blocks the extrinsic apoptotic cascade in tissue plasminogen activator/N-methyl D-aspartate-
971 treated neurons via Tyro3-Akt-FKHRL1 signaling pathway. *Mol Neurodegener* 6

972 Haghverdi L, Lun ATL, Morgan MD & Marioni JC (2018) Batch effects in single-cell RNA-
973 sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol* 36: 421–

427

Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, Toyoda Y, Gak IA, Weisswange I, Mansfeld J, Buchholz F, *et al* (2015) A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell* 163: 712–723

Hornbeck P V., Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, Latham V & Sullivan 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

Hsu PL, Jou J & Tsai SJ (2019) TYRO3: A potential therapeutic target in cancer. *Exp Biol Med (Maywood)* 244: 83–99

Huang K-Y, Wu H-Y, Chen Y-J, Lu C-T, Su M-G, Hsieh Y-C, Tsai C-M, Lin K-I, Huang H-D, Lee T-Y, *et al* (2014) RegPhos 2.0: an updated resource to explore protein kinase–substrate phosphorylation networks in mammals. *Database* 2014: 1–12

Imamura H, Yachie N, Saito R, Ishihama Y & Tomita M (2010) Towards the systematic discovery of signal transduction networks using phosphorylation dynamics data. *BMC Bioinformatics* 11

Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, Sidiropoulos K, Cook J, Gillespie M, Haw R, *et al* (2020) The reactome pathway knowledgebase. *Nucleic Acids Res* 48: D498–D503

Johannessen CM, Boehm JS, Kim SSY, Thomas SR, Wardwell L, Johnson LA, Emery CM, Stransky N, Cogdill AP, Barretina J, *et al* (2010) COT drives resistance to RAF inhibition through MAP 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 AF, Weiss MJ, *et al* (2020) Integrative proteomics reveals principles of dynamic phosphosignaling networks in human erythropoiesis. *Mol Syst Biol* 16: 1–22

Keenan AB, Torre D, Lachmann A, Leong AK, Wojciechowicz ML, Utti V, Jagodnik KM, Kropiwnicki E, Wang Z & Ma'ayan A (2019) ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Res* 47: W212–W224

Kinexus BioInformatics Corporation KinaseNET.

Klezovitch O, Risk M, Coleman I, Lucas JM, Null M, True LD, Nelson PS & Vasioukhin V (2008) A causal role for ERG in neoplastic transformation of prostate epithelium. *Proc Natl Acad Sci U S A* 105: 2105–10

Köksal AS, Beck K, Cronin DR, McKenna A, Camp ND, Srivastava S, MacGilvray ME, Bodík R, Wolf-Yadlin A, Fraenkel E, *et al* (2018) Synthesizing Signaling Pathways from Temporal Phosphoproteomic Data. *Cell Rep* 24: 3607–3618

Kraehn GM, Utikal J, Udart M, Greulich KM, Bezold G, Kaskel P, Leiter U & Peter RU (2001) Extra c-myc oncogene copies in high risk cutaneous malignant melanoma and melanoma metastases. *Br J Cancer* 84: 72–79

Kuleshov M V., Xie Z, London ABK, Yang J, Evangelista JE, Lachmann A, Shu I, Torre D & Ma'Ayan A (2021) KEA3: improved kinase enrichment analysis via data integration. *Nucleic Acids Res* 49: W304–W316

1014 Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, Silverstein MC & Ma'ayan A
1015 (2018) Massive mining of publicly available RNA-seq data from human and mouse. *Nat*
1016 *Commun* 9

1017 Lähdesmäki H, Rust AG & Shmulevich I (2008) Probabilistic inference of transcription factor
1018 binding from multiple data sources. *PLoS One* 3

1019 Langfelder P & Horvath S (2008) WGCNA: An R package for weighted correlation network analysis.
1020 *BMC Bioinformatics* 9: 559

1021 Legland D, Arganda-Carreras I & Andrey P (2016) MorphoLibJ: integrated library and plugins for
1022 mathematical morphology with ImageJ. *Bioinformatics* 32: 3532–3534

1023 Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P & Mesirov JP (2011)
1024 Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27: 1739–1740

1025 Love MI, Huber W & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-
1026 seq data with DESeq2. *Genome Biol* 15: 550

1027 Ma J, Shojaie A & Michailidis G (2019) A comparative study of topology-based pathway enrichment
1028 analysis methods. *BMC Bioinformatics* 20: 546

1029 Manders EMM, Stap J, Brakenhoff GJ, Van Driel R & Aten JA (1992) Dynamics of three-
1030 dimensional replication patterns during the S-phase, analysed by double labelling of DNA and
1031 confocal microscopy. *J Cell Sci* 103: 857–862

1032 McDowall MD, Scott MS & Barton GJ (2009) PIPs: human protein–protein interaction prediction
1033 database. *Nucleic Acids Res* 37: D651–D656

1034 Merilahti JAM & Elenius K (2019) Gamma-secretase-dependent signaling of receptor tyrosine
1035 kinases. *Oncogene* 38: 151–163

1036 Merilahti JAM, Ojala VK, Knittle AM, Pulliainen AT & Elenius K (2017) Genome-wide screen of
1037 gamma-secretase–mediated intramembrane cleavage of receptor tyrosine kinases. *Mol Biol Cell*
1038 28: 3123–3131

1039 von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P & Snel B (2003) STRING: A database of
1040 predicted functional associations between proteins. *Nucleic Acids Res* 31: 258–261
1041 doi:10.1093/nar/gkg034 [PREPRINT]

1042 Millikin RJ, Solntsev SK, Shortreed MR & Smith LM (2018) Ultrafast Peptide Label-Free
1043 Quantification with FlashLFQ. *J Proteome Res* 17: 386–391

1044 Mo Q, Wang S, Seshan VE, Olshen AB, Schultz N, Sander C, Powers RS, Ladanyi M & Shen R
1045 (2013) Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proc*
1046 *Natl Acad Sci U S A* 110: 4245–4250

1047 Murakami Y & Mizuguchi K (2014) Homology-based prediction of interactions between proteins
1048 using Averaged One-Dependence Estimators. *BMC Bioinformatics* 15: 1–11

1049 Naresh A, Long W, Vidal GA, Wimley WC, Marrero L, Sartor CI, Tovey S, Cooke TG, Bartlett
1050 JMSS & Jones FE (2006) The ERBB4/HER4 intracellular domain 4ICD is a BH3-only protein
1051 promoting apoptosis of breast cancer cells. *Cancer Res* 66: 6412–6420

1052 Oughtred R, Rust J, Chang C, Breitkreutz BJ, Stark C, Willems A, Boucher L, Leung G, Kolas N,

1053 Zhang F, *et al* (2021) The BioGRID database: A comprehensive biomedical resource of curated
1054 protein, genetic, and chemical interactions. *Protein Sci* 30: 187–200

1055 Pratapa A, Jalihal AP, Law JN, Bharadwaj A & Murali TM (2020) Benchmarking algorithms for
1056 gene regulatory network inference from single-cell transcriptomic data. *Nat Methods* 17: 147–
1057 154

1058 Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG & Ma'ayan
1059 A (2016) The harmonizome: a collection of processed datasets gathered to serve and mine
1060 knowledge about genes and proteins. *Database* 2016

1061 Sarbassov DD, Guertin DA, Ali SM & Sabatini DM (2005) Phosphorylation and regulation of
1062 Akt/PKB by the rictor-mTOR complex. *Science (80- )* 307: 1098–1101

1063 Schaefer MH, Fontaine JF, Vinayagam A, Porras P, Wanker EE & Andrade-Navarro MA (2012)
1064 HIPPIE: Integrating Protein Interaction Networks with Experiment Based Quality Scores. *PLoS*
1065 *One* 7: e31826

1066 Schmieder R & Edwards R (2011) Quality control and preprocessing of metagenomic datasets.
1067 *Bioinformatics* 27: 863–864

1068 Seitz HM, Camenisch TD, Lemke G, Earp HS & Matsushima GK (2007) Macrophages and Dendritic
1069 Cells Use Different Axl/Mertk/Tyro3 Receptors in Clearance of Apoptotic Cells. *J Immunol*
1070 178: 5635–5642

1071 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B & Ideker
1072 T (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular
1073 Interaction Networks. *Genome Res* 13: 2498–2504

1074 Shao H, Wang A, Lauffenburger D & Wells A (2018) Tyro3-mediated phosphorylation of ACTN4
1075 at tyrosines is FAK-dependent and decreases susceptibility to cleavage by m-Calpain. *Int J*
1076 *Biochem Cell Biol* 95: 73–84

1077 Solntsev SK, Shortreed MR, Frey BL & Smith LM (2018) Enhanced Global Post-translational
1078 Modification Discovery with MetaMorpheus. *J Proteome Res* 17: 1844–1851

1079 Stauffer W, Sheng H & Lim HN (2018) EzColocalization: An ImageJ plugin for visualizing and
1080 measuring colocalization in cells and organisms. *Sci Rep* 8: 1–13

1081 Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy
1082 SL, Golub TR, Lander ES, *et al* (2005) Gene set enrichment analysis: A knowledge-based
1083 approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102:
1084 15545–15550

1085 Sugiyama N, Imamura H & Ishihama Y (2019) Large-scale Discovery of Substrates of the Human
1086 Kinome. *Sci Rep* 9: 1–12

1087 Tsai CL, Chang JS, Yu MC, Lee CH, Chen TC, Chuang WY, Kuo WL, Lin CC, Lin SM & Hsieh SY
1088 (2020) Functional genomics identifies hepatitis-induced STAT3-TyrO3-STAT3 signaling as a
1089 potential therapeutic target of hepatoma. *Clin Cancer Res* 26: 1185–1197

1090 Tucher J, Linke D, Koudelka T, Cassidy L, Tredup C, Wichert R, Pietrzik C, Becker-Pauly C &
1091 Tholey A (2014) LC-MS based cleavage site profiling of the proteases ADAM10 and ADAM17
1092 using proteome-derived peptide libraries. *J Proteome Res* 13: 2205–2214

1093 Tworkoski KA, Singhal G, Szpakowski S, Zito CI, Bacchiocchi A, Muthusamy V, Bosenberg M,
1094 Krauthammer M, Halaban R & Stern DF (2011) Phosphoproteomic Screen Identifies Potential
1095 Therapeutic Targets in Melanoma. *Mol Cancer Res* 9: 801–812

1096 Vasaikar S V., Straub P, Wang J & Zhang B (2018) LinkedOmics: Analyzing multi-omics data within
1097 and across 32 cancer types. *Nucleic Acids Res* 46: D956–D963

1098 Vidal GA, Naresh A, Marrero L & Jones FE (2005) Presenilin-dependent gamma-secretase
1099 processing regulates multiple ERBB4/HER4 activities. *J Biol Chem* 280: 19777–83

1100 Yang XXX, Boehm JS, Yang XXX, Salehi-Ashtiani K, Hao T, Shen Y, Lubonja R, Thomas SR,
1101 Alkan O, Bhimdi T, *et al* (2011) A public genome-scale lentiviral expression library of human
1102 ORFs. *Nat Methods* 8: 659–661

1103 Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov
1104 AG, Bennett R, *et al* (2020) Ensembl 2020. *Nucleic Acids Res* 48: D682–D688

1105 Zanzoni A, Montecchi-Palazzi L, Quondam M, Ausiello G, Helmer-Citterich M & Cesareni G (2002)
1106 MINT: a Molecular INTeraction database. *FEBS Lett* 513: 135–140

1107 Zhu S, Wurdak H, Wang Y, Galkin A, Tao H, Li J, Lyssiotis CA, Yan F, Tu BP, Miraglia L, *et al*
1108 (2009a) A genomic screen identifies TYRO3 as a MITF regulator in melanoma. *Proc Natl Acad*
1109 *Sci* 106: 17025–17030

1110 Zhu S, Wurdak H, Wang Y, Galkin A, Tao H, Li J, Lyssiotis CA, Yan F, Tu BP, Miraglia L, *et al*
1111 (2009b) A genomic screen identifies TYRO3 as a MITF regulator in melanoma. *Proc Natl Acad*
1112 *Sci* 106: 17025–17030

1113

**A**

The correlation score for Protein 1 (*P1*) and Protein 2 (*P2*) for total number (*n*) of samples *s* is calculated as follows:

$$correlation\ score = \frac{n\sum_{s=1}^{n} Rank(P1_s)Rank(P2_s) - (\sum_{s=1}^{n} Rank(P1_s))(\sum_{s=1}^{n} Rank(P2_s))}{\sqrt{n\sum_{s=1}^{n} Rank(P1_s)^2(\sum_{s=1}^{n} Rank(P1_s))^2}\sqrt{n\sum_{s=1}^{n} Rank(P2_s)^2(\sum_{s=1}^{n} Rank(P2_s))^2}}\ ,\epsilon\,[0,1]$$

The stoichiometry score for Protein 1 (*P1*) and Protein 2 (*P2*) for total number (*n*) of samples is calculated as follows:

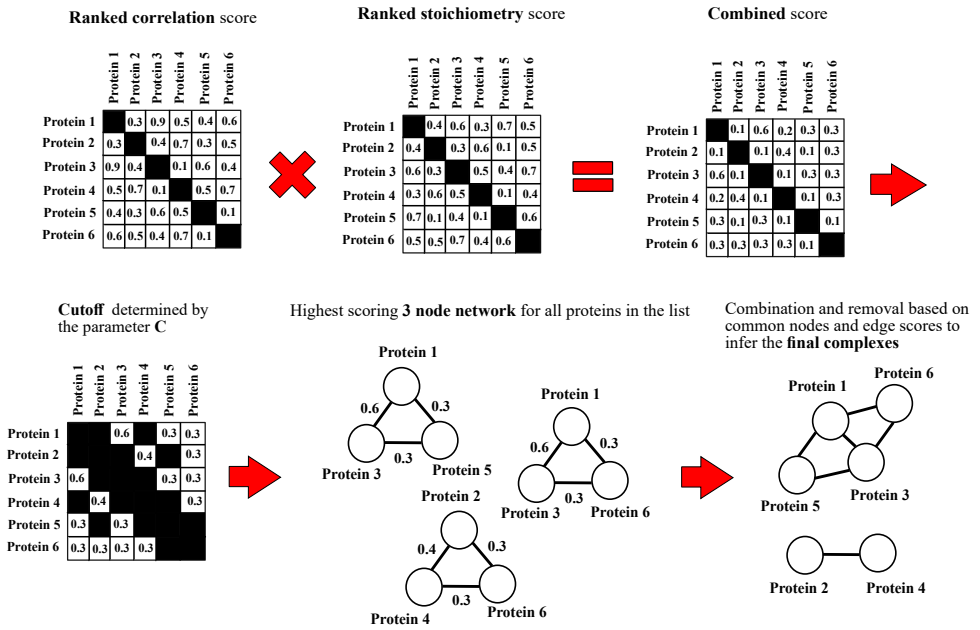
*if* $median(P1_1, P1_2, ...P1_n) >= median(P2_1, P2_2, ...P2_n)$

$$stoichiometry\ score = \frac{quartile\ Q_3}{quartile\ Q_1}\ in\ \left(\frac{P1_1}{P2_1}, \frac{P1_2}{P2_2}, ...\frac{P1_n}{P2_n}\right)\ ,\epsilon\,[1,\infty]$$

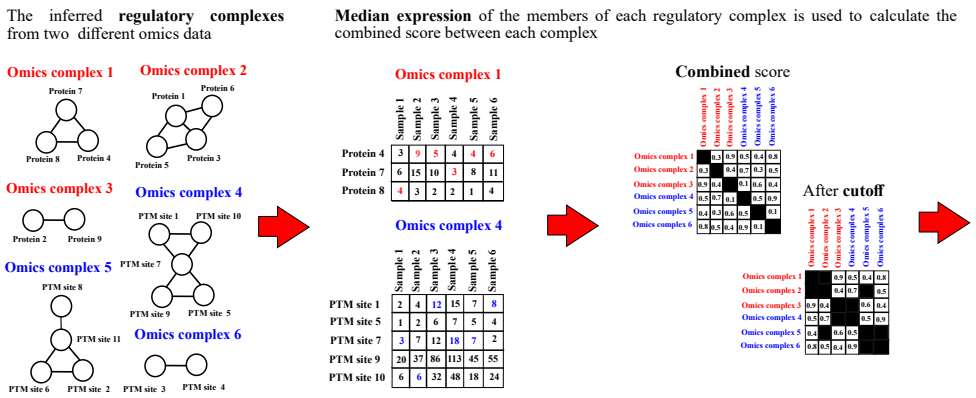
*else*

$$stoichiometry\ score = \frac{quartile\ Q_3}{quartile\ Q_1}\ in\ \left(\frac{P2_1}{P1_1}, \frac{P2_2}{P1_2}, ...\frac{P2_n}{P1_n}\right)\ ,\epsilon\,[1,\infty]$$

**B**




**C**




The same method utilized to infer the regulatory complexes is used to **infer the pathways** from the combined scores between each complex

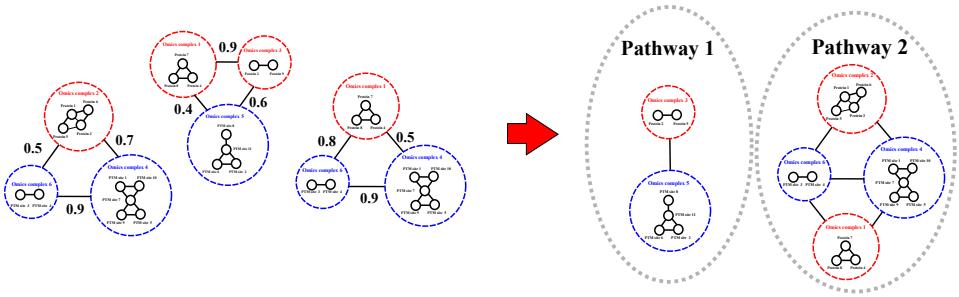



**Figure 1. The underlying concepts of the unbiased pathway analysis. A:** Mathematical formulation of the correlation (Spearman correlation) and stoichiometry scores. **B:** Schematic representation of the inference of regulatory omics complexes for the unbiased multi-omics pathway inference analysis. **C:** Schematic representation of the principles of combining the regulatory omics complexes into pathways.

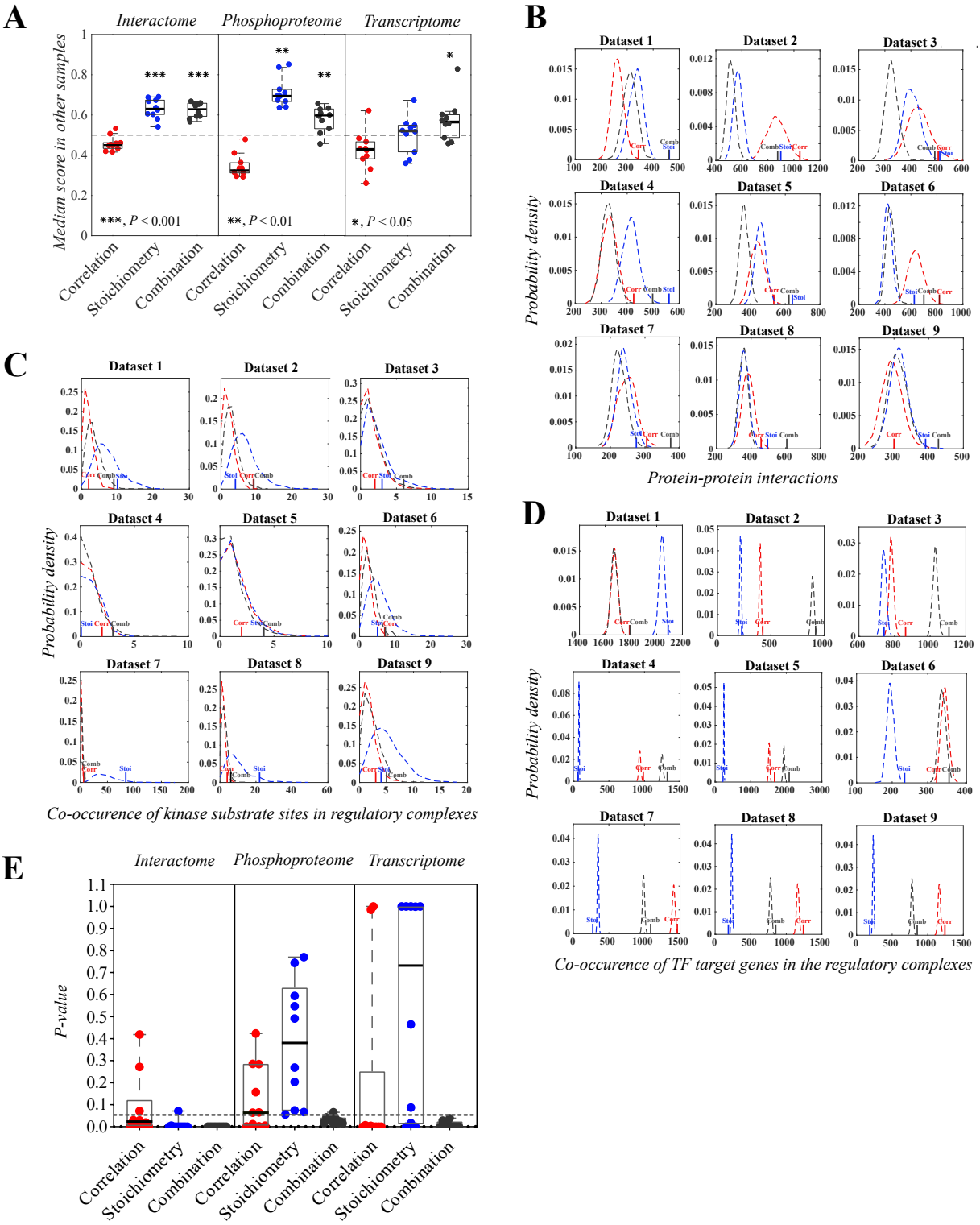



**Figure 2. Validation of regulatory complexes modeled with the UPA. A:** The conservation of correlation, stoichiometry and combined score in interactome, phosphoproteome and transcriptome data. One dot represents the median score between all complex members of all modeled complexes of a dataset that 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 horizintal line the median value. Red: correlation score; blue: stoichiometry score; grey: combined score.

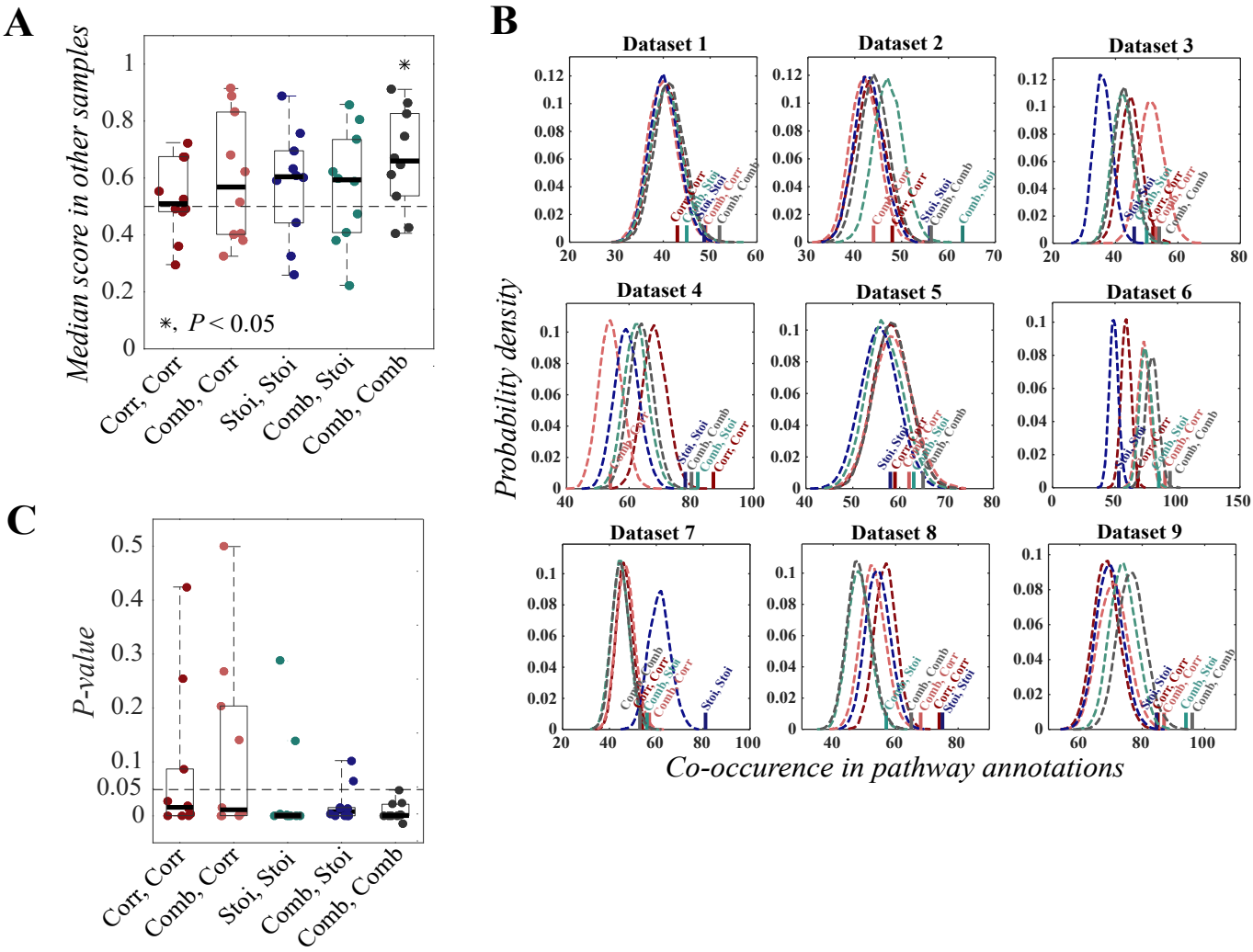



**Figure 3. Validation of regulatory complex combination to pathways with the UPA. A:** The conservation of correlation, stoichiometry and combined score in complexes modeled with either correlation (Corr), stoichiometry (Stoi) or combined (Comb) score and combined with either correlation, stoichiometry 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 horizintal line the median value.

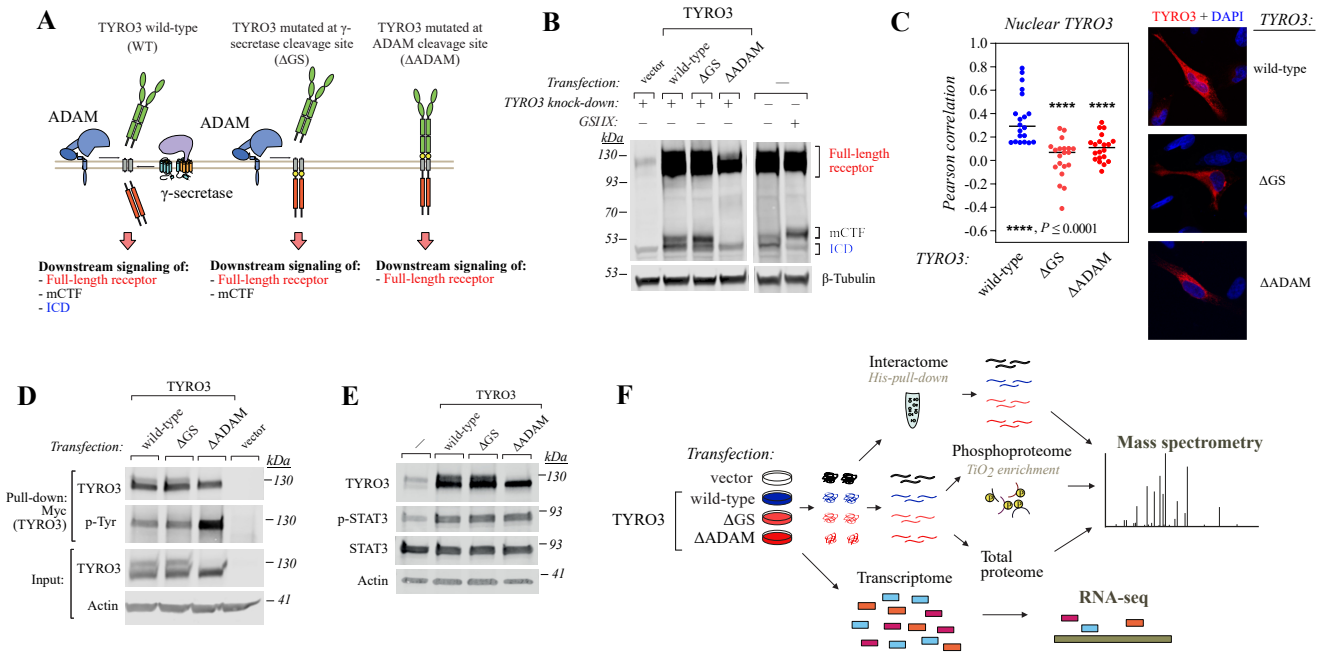



**Figure 4. Functional validation of RIP-resistant TYRO3 receptor variants. A:** A Schematic depicting the signaling mediated by wild-type and regulated intramembrane proteolysis (RIP)-resistant variants (ΔGS, ΔADAM) of TYRO3. Since the cleavage events in RIP are sequential, blocking only the gamma-secretase 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.

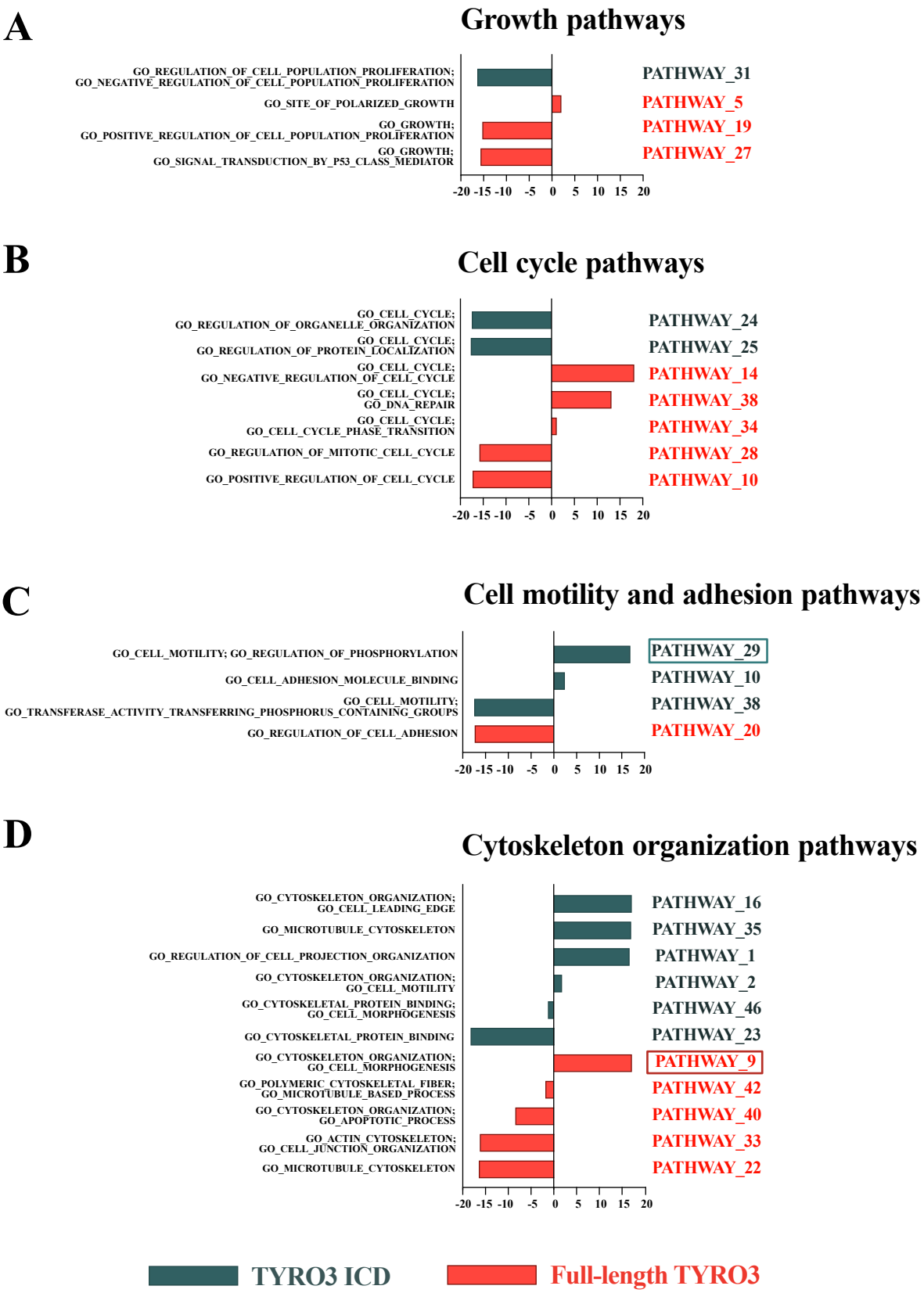



**Figure 5. The functional categorization of the selected full-length and TYRO3 ICD pathways inferred with the UPA.** The predicted function of the multi-omics pathways of the full-length and ICD of TYRO3 inferred with the UPA. The bars represent the median pseudolog2 fold change of all the proteins, 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.

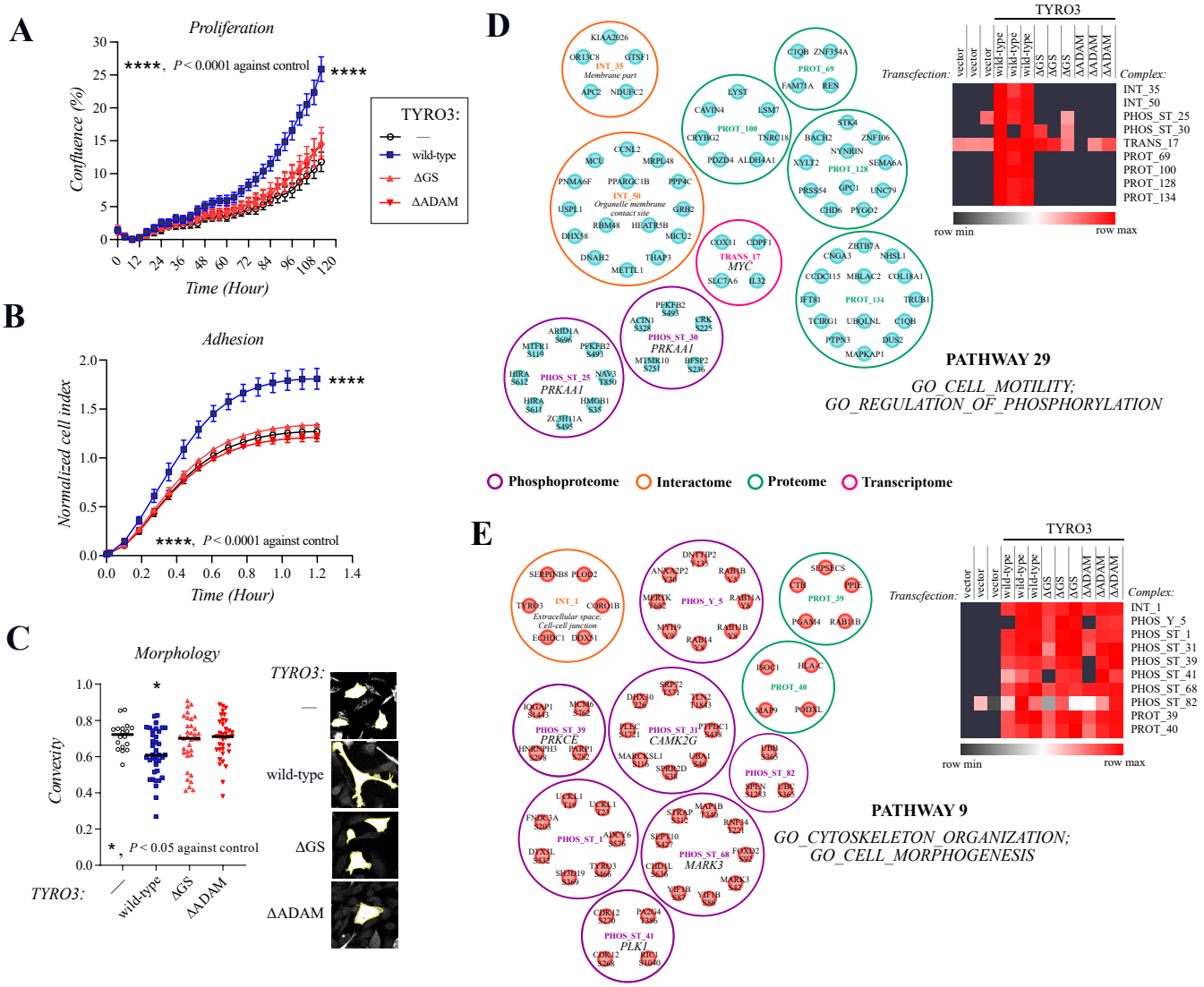



**Figure 6. Functional validation and composition of the selected pathways of full-length and ICD of TYRO3. A:** Proliferation of WM-266-4 transfectants was measured with live-cell imaging. For statistical testing, the non-parametric Friedman 2-way ANOVA and the Dunn's multiple comparison test with a 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 indenpendent 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