De Novo multi-omics pathway analysis (DMPA) designed for prior data independent 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]

[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: De novo multi-omics cell signaling pathway analysis

Abbreviations used: ADAM, a disintegrin and metalloprotease; ICD, intracellular domain; RTK, receptor tyrosine kinase; DMPA, de novo multi-omics pathway analysis; WGCNA, weighted correlation network analysis; LOPC, low-order partial correlation, TransNet, Transkingdom Network Analysis;

New tools for cell signaling pathway inference from multi-omics data that are independent of previous knowledge are needed. Here we propose a new de novo method, the de novo multi-omics pathway analysis (DMPA), to model and combine omics data into regulatory complexes and pathways. DMPA 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, metabolic reactions, epigenetic trait associations and signaling pathways. DMPA was benchmarked against existing module and network discovery and multi-omics integration methods and outperformed previous methods in module and signaling pathway discovery especially when applied to datasets with low sample sizes and zero-inflated data. Transcription factor, kinase, subcellular location and function prediction algorithms were devised for transcriptome, phosphoproteome and interactome regulatory complexes and pathways, respectively. To apply DMPA 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. The pathways modeled with DMPA reflected both the predicted function and the direction of the predicted function in validation experiments.

Introduction

Different approaches to model omics data from different levels of cell signaling into cell signaling networks have been developed to understand the interplay, combined effect and causality of cell signaling changes. Previous approaches may be divided into two main categories: methods that are reliant on or independent of previous knowledge. The methods that rely on previous knowledge include the classical gene set enrichment analysis [1], its derivatives, and multiple topology-based methods that use previous data as a backbone to infer pathway networks [2]. Modeling based on previous knowledge poses three major caveats. Firstly, previous knowledge is often incomplete signifying that all present genes, proteins, transcripts and post-translational modification sites are often not covered by the available databases. Secondly, the uneven acquirement of previous knowledge leads to skewed results in endorsing relationships that have already been discovered and extensively researched before. Thirdly, molecular associations are often context-dependent indicating that the associations found in a different context may not universally apply.

Robust methods that are independent on previous knowledge have been generally based on only one measure of molecular association, mainly correlation or partial correlation, between the expression level of different cell signaling molecules [3–6]. Approaches that combine more than one robust molecular association measures such as correlation have not been sufficiently explored. More sophisticated efforts utilizing machine learning methods such as regressions, Bayesian networks and ODE models have been developed to solve single signaling pathway modeling problems such as gene regulatory networks [7,8], but the transferability of these methods to multi-omics data is largely unknown. Most accurate renderings of cell signaling networks have been acquired with methods that are tracking changes in time-course data [7,9]. Unfortunately, the cost and feasibility of acquiring time-course data present a limitation especially in patient samples and in animal models.

The current array of multi-omics integration methods mainly consist of dimensionality reduction and clustering methods that aim to stratify multi-omics data according to disease subtype or patient characteristics and assist in biomarker discovery [10–12]. The module discovery methods such as WGCNA and correlation networks have been expanded and applied to multi-omics data [13,14], but beyond these efforts the multi-omics integration methods that aim to learn cell signaling networks seem to rely on previous knowledge [15–20]. A need for a robust scalable benchmark method that can handle different omics data, is able to track molecular associations from single time point omics data, is independent on previous knowledge and outperforms methods that are based on more than one measure of molecular association, is evident.

68 # Results

69

70 **The underlying concepts of the de novo pathway analysis (DMPA)**

71 We set out to develop a new robust de novo method to infer regulatory modules from different omics
72 data and to combine the modules into multi-omics pathways. Literature search of previous reports of
73 the key features of omics data predicting molecular associations produced the following observations:
74 First, multiple sources indicate the ability of correlation of expression values to predict molecular
75 association to a certain extent [21,22]. Secondly, the findings from an extensive human interactome
76 study indicate that conserved stoichiometry between two proteins inside the same sample can be
77 predictive of protein-protein interaction networks [23]. From the grounds of these two concepts, the
78 basis for the combined score between two proteins, transcripts or phosphosites in an omics dataset
79 was devised to reflect the strength of association. In Figure 1A the mathematical formulation for the
80 correlation and stoichiometry score is presented. The correlation score was based on the
81 nonparametric Spearman correlation which was consequently ranked and adjusted to values between
82 1 and 0 to allow for a linear scale. The stoichiometry score was based on the relation of the third
83 quartile and first quartile value of relative abundances of the two cell signaling molecules or their
84 modifications in an omics dataset and similarly ranked and adjusted to a linear scale from 1 to 0. The
85 combined score was expressed as a simple non-weighted multiplication of the correlation and
86 stoichiometry scores similarly ranked and adjusted to a linear scale between 1 to 0. Nonparametric
87 formulation was chosen to overset technical variation and to desensitize the model to outlier values.

88 To reduce computational demand ($O(n^2)$ for run time), a cut-off for the combined score was
89 devised to consider only strong associations. In the current version of the inference method the cut-
90 off parameter is controlled by two parameters that reflect the number of associations considered for

91 the molecules in the data (parameter C) and the predicted amount of non-associating features in the
92 data (parameter S).

93 The inference of regulatory modules from the combined score after the cutoff was designed
94 to follow the nearest neighbor principle, which assumes that the combination of combined scores
95 between 3 cell signaling molecules is more predictive than a single combined score between two
96 molecules or their states in the dataset. The inference method is outlined in Figure 1B and utilizes the
97 combined scores above the chosen cut-off threshold to find maximal scoring 3 member cliques. These
98 3 member cliques are consequently combined to find larger complexes and trimmed to remove
99 members that are present in a higher scoring complex. The principles of the combination and
100 trimming are outlined in more detail in the Materials and methods section.

101 The combination of the regulatory complexes was designed to follow a similar principle to
102 the initial inference of the regulatory complexes (Figure 1C). A median expression value of all the
103 members of a complex in each sample is used as a representation of the expression values of a
104 complex and the combined score is calculated based on these values. To allow equal combination of
105 down- and up-regulated complexes, an absolute value was taken from the Spearman correlation to
106 equally weigh positive and negative correlations. The complexes were then combined further into
107 pathways based on the combined scores above a certain threshold with the same rules for combining
108 and trimming the maximal scoring 3 complex cliques that were used in the initial regulatory complex
109 inference. The de novo mulri-omics pathway analysis is available both as matlab code and
110 executables in Mendeley Data: https://data.mendeley.com/datasets/m3zggn6xx9/draft?a=71c29dac-
111 714e-497e-8109-5c324ac43ac3.

112

113 **Validation of regulatory complexes modeled with DMPA**

114 The performance of the derived regulatory complex inference method was tested on available
115 interactome [23], phosphoproteome [24], and transcriptome [25] data. While molecular associations
116 are context-dependent, certain associations are more frequently reproduced in different contexts and
117 sometimes even stable. Therefore, a level of conservation is expected from a score that would
118 accurately reflect the probability of molecular association. The conservation of the correlation,
119 stoichiometry and combined scores were examined by modeling a set of proteins, phosphorylation
120 sites or transcripts from one dataset into regulatory complexes and examining the median of the
121 corresponding scores derived from the same complexes in another dataset. A median score above 0.5,
122 the median score expected if the scores would be randomly assigned to molecular associations,
123 indicated conservation.

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

129 To validate that the regulatory complexes reflect discovered molecular associations, a method
130 to count the number of known associations in the modeled regulatory complexes was devised. Known
131 protein-protein interactions, kinase substrate phosphosite relationships, and transcription factor target
132 gene annotations were downloaded from STRING [26], PhosphoSitePlus [27] and ENCODE [28]
133 databases, respectively. The sums of all co-occurrences of proteins known to interact, substrate sites
134 known to be phosphorylated by the same kinase, and transcripts known to be the target gene for the
135 same transcription factor inside the modeled regulatory complexes were used as validation scores.
136 The validation scores were calculated from the regulatory complexes inferred with correlation,

137 stoichiometry and combined scores to compare the effect of the different scores on the accuracy of
138 the model. The validation scores from the real modeled regulatory complexes were compared to the
139 validation scores of randomly formulated regulatory complexes of the same size to calculate the
140 probability that a similar validation score would be acquired by random. The probability density of
141 the validation scores of these randomized regulatory complexes are indicated with a dashed line and
142 the validation scores for the regulatory complexes modeled with different scores with solid lines in
143 the Figure 2B-D. The corresponding P-values drawn from the cumulative probability densities of the
144 randomized regulatory complexes for the validation scores of the regulatory complexes modeled with
145 different scores are visualized in Figure 2E.

146 The regulatory complexes modeled with the combined scores consistently reflected known
147 molecular associations in all interactome, phosphoproteome and transcriptome data. In interactome
148 data, the co-occurrence of two proteins in the same regulatory complex predicted a protein-protein
149 interaction between the two proteins more often than random assignment to regulatory complexes of
150 the same size. Additionally, the median STRING score, that reflects the confidence of the interaction
151 based on the amount and type of evidence on the protein-protein interaction, was consistently higher
152 in the regulatory complexes modeled with the combined score than in the randomized regulatory
153 complexes (Supplementary Figure 1). In the phosphoproteome data, the assignment to regulatory
154 complexes modeled with the combined score indicated that the two phosphorylation sites were
155 controlled by the same kinase more often than by random assignment. In the transcriptome data, in
156 turn, two transcripts in the same regulatory complex modeled with the combined score were more
157 likely to be the target gene for the same transcription factor than randomly chosen transcripts. This
158 effect was similarly observed in transcription factor target gene annotations acquired from REMAP
159 and Literature libraries from the ChEA3 database [29] (Supplementary Figure 2). The regulatory
160 complexes modeled with only the correlation or stoichiometry score varied in their ability to find
161 known molecular associations based on the modeled data. The regulatory complexes modeled with

162 stoichiometry score were sufficient, although not as consistent as the regulatory complexes modeled
163 with the combined score, in reflecting known protein-protein interactions in interactome data.
164 However, the regulatory complexes modeled with stoichiometry score failed to reproduce known
165 molecular associations in phosphoproteome and transcriptome data. The regulatory complexes
166 modeled with the correlation score were more or less able to capture known transcription factor target
167 gene relationships in the transcriptome data but were insufficient in finding known protein-protein
168 interactions and kinase substrate relationships in the interactome and phosphoproteome data,
169 respectively. All in all, the regulatory complexes inferred with the combined score consistently
170 captured known molecular associations in all tested omics data, while the complexes inferred with
171 other scores were found lacking in reproducibility and performance across omics datatypes.

172 To examine the scalability of DMPA to omics data other than the initially tested interactome,
173 phosphoproteome and transcriptome data, the conservation of the combined score was also similarly
174 examined in methylome, metabolomics and protein acetylation data. The combined score was
175 discovered indeed to be significantly conserved also in these omics data (Supplementary Figure 3A).
176 The ability of the regulatory complexes modelled with DMPA to predict known molecular
177 associations from available metabolomics and methylation data was additionally examined. The
178 regulatory complexes modelled with DMPA were able to predict metabolites involved in the same
179 metabolic reaction in metabolomics data and methylated chromatin regions associated with the same
180 trait in methylome data (Supplementary Figure 3B). These data indicate that DMPA is scalable and
181 can additionally be applied to other omics data than the initially validated interactome,
182 phosphproteome and transcriptome data.

183

184 **Validation of regulatory complex combination to pathways by DMPA**

185 Since signaling pathways are a combination of molecular associations, they are similarly expected to
186 be partially conserved. The conservation of the regulatory complex combination to pathways was
187 assessed by modeling a multi-omics dataset into regulatory complexes with either correlation,
188 stoichiometry or combination score, combining the regulatory complexes with either correlation,
189 stoichiometry or combined score and examining the median of the corresponding scores derived from
190 the same modeled pathways in another multi-omics dataset. Multi-omics data from normal and
191 cancer tissue samples from the LinkedOmics [30] database were utilized. Of all the tested score
192 combinations for the analysis, inferring the regulatory complexes with the combined score and
193 combining the inferred regulatory complexes into pathways with a combined score was the only score
194 combination for the analysis that demonstrated statistically significant conservation (Figure 3A).

195 To ensure that DMPA is also able to combine the inferred regulatory complexes into pathways
196 that reproduce known signaling pathways, a validation method was devised. The sum of all co-
197 occurrences of signaling molecules in separate regulatory complexes combined by the DMPA in
198 pathway annotations was used as a validation score for the modeled pathways. The reference pathway
199 annotations were downloaded from the PathwayCommons database [31]. The validation scores were
200 compared to the validation scores calculated from 10,000 randomly combined pathways of the same
201 size to estimate the probability that random assignment of the regulatory complexes would produce
202 pathways that equally reflect known pathway annotations. The probability density of the validation
203 scores of these randomly combined pathways are indicated with a dashed curve and the validation
204 scores for the regulatory complexes modeled with different scores with solid lines in Figure 3B. The
205 corresponding P-values drawn from the cumulative probability densities of the randomized pathways
206 for the validation scores of the regulatory complexes modeled with different scores and combined
207 with different scores are visualized in Figure 3C. The pathways inferred and combined with the
208 combined score consistently reproduced associations represented in known pathway annotations. The

209 second-best performers were the versions of the analysis method where the pathways were inferred
210 with either the stoichiometry score or the combined score and combined with the stoichiometry score.
211 The correlation score-based versions of the analysis method performed worst in finding molecular
212 associations present in known pathway annotations. Taken together, the results from the validations
213 and the level of conservation of the combined score in other datasets strongly imply that the DMPA
214 can find molecular associations and pathways that reflect previously observed cell signaling modules.

215

216 **Validation of the design decisions of DMPA**

217 To explore the validity of the design decision of DMPA, the effect of different design decisions of
218 DMPA on the ability of the inferred molecular associations to reflect known molecular associations
219 was examined as in Figure 2B-E. Since the P-values in many cases approached zero the relative
220 distance of the validation score from the median value of the random distribution was used as a
221 measure instead of the P-value like in Figure 2B-E. First, the effect of the choice to rank and adjust
222 the stoichiometry and correlation score on the amount of known molecular association in the inferred
223 molecular associations was assessed. In both interactome and transcriptome data the choice to rank
224 and adjust the stoichiometry and correlation score (combination, ranked) increased the number of
225 known associations in the inferred molecular associations against the version of the combination
226 score where the unadjusted correlation score and the inverse of the unadjusted stoichiometry score
227 were multiplied (combination, raw) (Supplementary Figure 4A). In phosphoproteome data the choice
228 to rank and adjust the correlation and stoichiometry had no significant effect. Second, the effect of
229 the design choice to find the maximally scoring 3 node cliques for each signaling event in the data
230 instead of a linear approach, where for each signaling event only the maximally scoring signaling
231 event is selected was examined. For both interactome and phosphoproteome data searching the
232 maximally scoring 3 node cliques instead of the single maximally scoring signaling event, increased

233 the amount of previously discovered molecular associations in the inferred molecular associations
234 (Supplementary Figure 4B). For transcriptome data the design choice had no significant effect. Taken
235 together the results imply that the design decisions of DMPA have mostly a positive effect or a neutral
236 effect depending on the examined omics data on discovering true positives in the inferred molecular
237 associations. In combination the effect of adhering to both design decisions increased the number of
238 known associations in the inferred molecular associations for all tested omics data types.

239

240 **Sensitivity analysis of the parameters of DMPA**

241 To assess the sensitivity of DMPA to parameter choice a sensitivity analyses were conducted. First
242 the effect of parameter choice of the parameters that control the size of the regulatory modules was
243 examined with available interactome data. The change in these parameters affected the size and
244 amount of the regulatory modules but had no negative effect on the number of known associations in
245 the inferred molecular associations against randomly modelled molecular associations
246 (Supplementary Figure 5). This suggests that the parameters can be safely adjusted if larger or smaller
247 modules are desired from the inference method.

248 Second, the effect of the parameter choice of the parameters C and S that control the cut-off value of
249 the combined score was assessed in both available omics and simulated data (Figure 1B). To examine
250 the effect of the parameter S choice on the amount of known molecular associations in the inferred
251 molecular associations and on the algorithm performance, published transcriptome data was analyzed
252 with different S parameter values. Different S parameter values had a significant effect on the run
253 time but no adverse effect on the number of known molecular associations in the inferred molecular
254 associations (Supplementary Figure 6). To test the effect of the C and S parameters for the true
255 positive and false positive rate DMPA was applied to normally distributed, negative binomially
256 distributed and beta distributed simulated data (Supplementary Figure 7-9). DMPA was able to

257 accurately discover the true modules among the randomly modelled data. The performance of the
258 inference method was, however, discovered to be sensitive to the parameter S to an extent. When the
259 parameter S was set too low the likelihood of false positives in the regulatory modules was increased.
260 In turn when the parameter S was set too high true positives were lost. It is note of that DMPA never
261 misassigned molecules that belong to a regulatory module into another regulatory module, but rather
262 the false positives always came from the simulated pool that had no true module assignment. The
263 proportion of the molecules or their states that have no association with other molecules or their states
264 in the dataset in real omics datasets is expected to be low. The raw correlation and stoichiometry
265 scores of the molecules that were assigned or not assigned into regulatory modules were examined
266 and a significant drop in the maximum raw correlation and inverted raw stoichiometry score was
267 discovered for the molecules that had no true module assignment (Supplementary Figure 10A). To
268 this end a script to suggest values for the parameter S was devised to help the user to set the parameter
269 S (Supplementary Figure 10B). The effect of the parameter C was also examined and it was
270 discovered that the value for parameter C should be set only based on the size of the dataset (3 for
271 small datasets, 30-300 molecules in the dataset; 1 for larger datasets, 500< molecules in the dataset).

272 Third, the sensitivity of DMPA to the parameter that controls the zero-inflated version of the analysis
273 was assessed. Zero-inflated interactome data and non-zero-inflated transcriptome data was analyzed
274 both with the zero-inflated (weighted stoichiometry score) and non-zero-inflated (non-weighted
275 stoichiometry score) version of DMPA and the amount on known associations in the inferred
276 molecular associations against randomly modelled molecular associations was estimated. In the zero-
277 inflated interactome data (over 30% missing values), the zero-inflated version of the analysis
278 significantly increased the number of known associations in the inferred modules (Supplementary
279 Figure 11). In contrast, in the non-zero-inflated transcriptome data (less than 25% missing values)
280 the zero-inflated version discovered fewer known associations than the non-zero-inflated version.

281 The choice of zero-inflated and non-zero-inflated version of the analysis has a significant effect on
282 the accuracy of the inference and should be set based on the number of missing values in the dataset.

283

284 **DMPA outperforms previous approaches**

285 To assess the performance of DMPA, DMPA was benchmarked against relevant module discovery
286 and multi-omics integration methods. Transcriptome datasets were chosen for the benchmarking of
287 module discovery methods since it allowed the comparison against regulatory network inference
288 methods that are specifically designed to be used with gene expression data. WGCNA [3], LOPC [5]
289 and the DREAM4 challenge winning regulatory network inference method GENIE3 [32] were
290 selected for comparison. Since LOPC returns network edges instead of modules, a separate validation
291 strategy was devised to benchmark against LOPC which is detailed in the Materials and Methods
292 section. For GENIE3 a module was defined as the genes predicted to be controlled by the same
293 upstream gene. Although all WGCNA, LOPC and GENIE3 were able to discover significantly more
294 known molecular associations in the inferred molecular associations than random modelling, DMPA
295 was the most consistent performer across datasets and in most cases discovered most known
296 molecular associations (Figure 4A-B). Strikingly, a clear difference in the performance of DMPA
297 and the other tested methods could be seen with datasets with low sample sizes. To further assess the
298 performance of DMPA and the other module discovery methods with low sample sizes, the module
299 discovery methods were run with datasets with different sample sizes and the amount of known
300 molecular associations in the inferred molecular associations against the randomly assigned
301 molecular associations were examined. DMPA required only 6 samples (3 control and 3 treatment
302 samples) to discover known molecular associations while WGCNA and GENIE3 required over 30
303 (Figure 4C). An attempt to run zero-inflated interactome data with WGCNA and LOPC was made,
304 but the current R implementations of these approaches were unable to handle the zero-inflated

305 interactome data. DMPA in contrast has the advantage of performing without the imputation of the
306 missing values of the zero-inflated data. These results imply that DMPA is more scalable, has less
307 limitations and can discover more known associations than the tested module and network discovery
308 methods.

309 DMPA was additionally benchmarked against another multi-omics integration method that is
310 independent of prior knowledge. Since DMPA already performed better in module discovery than
311 WGCNA and correlation networks, the examination of the performance of DMPA against these
312 methods with multi-omics integration was considered redundant. Instead DMPA was benchmarked
313 against the closest discovered analogue method, Transkingdom Network Analysis (TransNet) [33].
314 The aim of TransNet is to discover causal links between two different omics datasets. TransNet first
315 models the omics datasets into correlation networks, assigns dense subnetworks into modules, and
316 then utilizes correlation to find the connecting nodes between the modules of the omics datasets.
317 Available multi-omics datasets were analyzed both with TransNet and DMPA and the ability of the
318 inferred pathways to represent known signaling pathways was estimated as in Figure 3B-C. The
319 ability to connect inferred modules into signaling pathways was also estimated by a version of the
320 validation method utilized in Figure 3B-C where only the connections but not the modules themselves
321 were randomized. As per its aim, TransNet was able to significantly discover connections between
322 the modules that represented known cell signaling pathways, but the modules themselves were not
323 representative of cell signaling pathways as suggested by the lack of a significant P-value from the
324 validation method that also randomized the modules (Figure 4D). DMPA, however, was able to
325 achieve a significant P-value with both versions of the validation method indicating that the pathways
326 inferred with DMPA are closer to known cell signaling pathways.

327 **Case study: DMPA of cleaved TYRO3 signaling in melanoma cells**

328 To demonstrate the performance of DMPA with a freshly acquired multi-omics dataset, an in-house
329 multi-omics dataset from the cleavage-dependent signaling of the receptor tyrosine kinase TYRO3
330 was acquired. To this end, a stable TYRO3 knock-down WM-266-4 cell line was created and the
331 cells were transfected with wild-type and cleavage-resistant variants of TYRO3 (ΔGS, ΔADAM) or
332 control vector (Supplementary Figure 12A). The ΔGS variant, resistant to proteolytic cleavage by the
333 gamma-secretase complex, has been described before [34]. The ΔADAM variant, resistant to
334 cleavage by a disintegrin and metalloprotease (ADAM), was created by predicting the ADAM17 of
335 ADAM10 cleavage sites of TYRO3 with a 0th order Markov chain and mutating the key residues in
336 the predicted cleavage sites. The ability of the ΔADAM variant to still act as functional receptor was
337 validated by examining its cleavage (Supplementary Figure 12B), autophosphorylation
338 (Supplementary Figure 12C) and ability to activate a known substrate with western blot analyses
339 (Supplementary Figure 12D). The expected effect of the mutations in the ΔADAM variant on the
340 subcellular location of TYRO3 was examined with immunofluorescence (Supplementary Figure
341 12E). The analyses strongly suggested that the ΔADAM variant was indeed a functional receptor that
342 was more resistant to cleavage than the wild-type receptor. Interactome, phosphoproteome, proteome
343 and transcriptome data were acquired from the cells expressing the wild-type or cleavage-resistant
344 variants of TYRO3 or the control vector (Supplementary Figure 12F). A differential expression
345 analysis was conducted to discover the differential interactome, phosphoproteome, transcriptome and
346 proteome associated with the full-length of the cleaved intracellular domain (ICD) of TYRO3
347 (Supplementary Figures 13-18). The details of the methodology to validate the ΔADAM variant and
348 to acquire the multi-omics data is provided in the Supplementary methods.

349 The differentially expressed signaling events associated with the expression of full-length TYRO3 or
350 the cleaved TYRO3 ICD were analyzed with the DMPA first to extract regulatory complexes. The
351 DMPA discovered 24 previously undescribed interactome, 110 phosphoproteome, 46 transcriptome

352 and 130 proteome regulatory complexes that associated with the expression of full-length TYRO3
353 (Supplementary Figure 19, Supplementary Table 2). The expression of the cleaved TYRO3 ICD in
354 turn was associated with 53 new interactome, 84 phosphoproteome, 45 transcriptome and 136
355 proteome regulatory complexes (Supplementary Figure 20, Supplementary Table 2). The inferred
356 regulatory complexes were further combined with the DMPA into pathways, resulting in a total of 51
357 and 46 unique signaling pathways associating with the signaling stimulated by the full-length and the
358 released ICD of TYRO3, respectively (Supplementary Table 1).

359

360 **Enrichment analyses for modules and pathways inferred with the DMPA**

361 To contextualize the function of the modules and pathways inferred with DMPA, an enrichment
362 analysis (please see Materials and methods for details) to predict the upstream transcription factor,
363 upstream kinase, subcellular location and the biological process for each transcriptome,
364 phosphoproteome and interactome module or pathway, respectively, was devised (See supplementary
365 materials and methods for details). The predicted upstream transcription factors, kinases and
366 subcellular locations for the transcriptome, phosphoproteome and interactome modules are provided
367 in Supplementary Tables 3-5. To ensure the validity of the upstream transcription factor and kinase
368 predictions, the predicted transcription factors and kinases were compared to those known to be
369 regulated by TYRO3. Several transcription factors known to be regulated by TYRO3 were predicted
370 to regulate the modeled regulatory transcriptome complexes, such as STAT3 [35], MYC [36,37], and
371 MITF [38] (Supplementary Table 3). Furthermore, the most common transcription factor predicted
372 to regulate the transcriptional complexes associated with TYRO3 signaling was MYC, with the
373 respective gene amplified in the WM-266-4 cell background [39]. Similarly, several kinases reported
374 to be associated with TYRO3, such as AKT [40,41], mTOR [40,42], GSK3 [41,42], MEK1 [42],
375 PKC [42], and CAMK2 [42], were predicted to regulate the phosphoproteome complexes associated

376 with full-length and cleaved ICD of TYRO3 (Supplementary Table 5). The predicted subcellular
377 localizations of the interactome complexes of full-length and the cleaved ICD of TYRO3
378 (Supplementary Table 4), in turn, were compared to the reported localization difference between the
379 wild-type and the ΔGS and ΔADAM variants of TYRO3 (Supplementary Figure 12E) (Merilahti *et*
380 *al*, 2017) to ensure the validity of the predicted subcellular locations. The differences in the predicted
381 localizations of the interactome complexes reflected the known difference in the subcellular
382 localization of full-length and cleaved ICD of TYRO3 (Supplementary Table 4). Thirteen out of the
383 54 (24%) cleaved ICD interactome complexes were predicted to localize into the nucleus, while
384 nuclear localization was predicted for only 2 out of the 21 (8 %) full-length interactome complexes
385 ($P = 0.032$ against frequency of nuclear assignment in the interactome complexes of cleaved TYRO3
386 ICD). Moreover, 8 out of the 21 (33%) full-length interactome complexes, but only 5 out of the 54
387 (9%) ICD interactome complexes, were predicted to localize into the plasma membrane ($P < 0.0001$
388 against frequency of plasma membrane assignment in the interactome complexes of full-length
389 TYRO3). Both the full-length TYRO3 interactome complexes (5 out of 21; 20%) and the ICD
390 interactome complexes (11 out of 54, 20%) were equally predicted to be localized in the cytoskeletal
391 structures ($P = 0.0025$ for full-length TYRO3 and $P < 0.0001$ for TYRO3 ICD against frequency of
392 cytoskeleton assignment in randomly modelled interactome complexes).

393 The summary of the predicted functions for each pathway associated with the signaling of
394 full-length and ICD of TYRO3 is displayed in Figure 5 and Supplementary Figure 21. The functional
395 enrichment analysis of the modelled pathways of full-length and ICD of TYRO3 was able to identify
396 unique pathways associated with various cellular processes including ones involved in the
397 pathogenesis of cancer. A significant difference was noted in the number and direction of pathways
398 related to growth and cell cycle, cell adhesion and motility, cell morphology, cell death, and immune
399 response between the cells expressing the full-length versus the ICD of TYRO3 (Figure 5). These

400 findings indicate that a differential response to these functions could be enacted by the full-length
401 and ICD of TYRO3.

402

403 **Functional validation of the functional enrichment analyses of the DMPA**

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

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

420 Finally, the two-dimensional cellular morphology of the WM-266-4 transfectants were
421 examined from thresholded confocal images with the ImageJ MorphoLibJ plugin (Figure 6C). As
422 indicated by the ICD pathway 46 and the full-length pathway 9 (Figure 5D), the cells expressing
423 wild-type TYRO3 exhibited distinct morphology from the vector control cells or cells expressing

424 either the ΔGS or ΔADAM variant of TYRO3 (Figure 6C). The morphological difference was
425 captured by the convexity measure (the ratio between the cell area and its convex area), the lower
426 value of which suggests a shape with more protrusions. A greater amount of cell protrusions in the
427 cells expressing wild-type TYRO3 was also predicted by the DMPA (ICD pathway 1 in Figure 5D).
428 Furthermore, the WM-266-4 cells expressing wild-type TYRO3 or either of the TYRO3 mutants
429 demonstrated less circular cell shape than the vector control cells, indicating that the full-length
430 TYRO3 additionally regulates cell morphology as suggested by the full-length pathway 9
431 (Supplementary Figure 22). Taken together, these validation experiments suggest that the function
432 enrichment of the pathways modeled with the DMPA can predict both the function and the direction
433 of the function from purely molecular multi-omics data.

434 **Discussion**

435

436 The increased demand for a ground-up analysis method able to model multi-omics data into cell
437 signaling pathways motivated us to create the de novo multi-omics pathway analysis, DMPA. The
438 basis of the DMPA lies in the calculation of two robust metrics of molecular association, the
439 correlation and stoichiometry score, and the consequent modeling of the cell signaling molecules into
440 regulatory complexes based on the nearest neighbor principle. The DMPA was able to consistently
441 find relevant biological associations in both published and freshly acquired multi-omics data. The
442 method showed consistent performance and the combined score-based networks outperformed
443 correlation-based network models in all tested data types. DMPA outperformed previous methods in
444 both module and pathway discovery.

445 The advantages of the DMPA lie in its independence on previous knowledge, robustness,
446 which makes it easily scalable to other omics data types, low computational demand and ability to
447 work with low sample sizes. The proof of the predictive ability to discover relevant molecular
448 association opens up further applications for the combined score in more sophisticated and complex
449 inference methods. Previous multi-omics integration methods have been focused mostly on feature
450 extraction and matrix factorization strategies. While the inferred latent factors have been found
451 accurate in stratifying samples into clusters and can be predictive of clinical outcome, there is little
452 evidence that the discovered factors would represent actual signaling pathways in the cell [10,11].
453 Here we provide evidence that the pathways modelled with the DMPA reflect previously described
454 signaling pathways.

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

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

467 We acquired multi-omics data from the cleavage-resistant and wild-type TYRO3 variants
468 and utilized the DMPA to shed light on the unknown cleavage-associated signaling pathways of the
469 TYRO3 RTK in melanoma cells. In addition to signaling as a full-length RTK, TYRO3 has been
470 proposed to undergo regulated intramembranous proteolysis (RIP) by sequential proteolysis by the
471 gamma-secretase complex followed by cleavage by an ADAM protease, releasing a soluble ICD with
472 potential signaling activities [34] .The DMPA discovered a total of 51 and 46 differential unique
473 signaling pathways for the full-length TYRO3 and the cleaved TYRO3 ICD, respectively. An
474 enrichment analysis to infer the biological process the pathways were involved with was also devised.
475 A difference in the amount and the direction of the pathways associated with growth and cell cycle,
476 cell motility and adhesion, immune response, chromosome organization, cell death, and cell
477 differentiation were observed between the full-length and cleaved TYRO3 ICD. Many of these
478 functions may be critically involved e.g., in the progression or therapeutic responses in melanoma.
479 For example, TYRO3 has been implicated in the proliferation, tumorigenesis, chemoresistance and
480 motility of melanoma cells [38,43–45]. Validation experiments confirmed that the DMPA with the
481 function prediction was able to accurately predict both the function and the direction of the function
482 of the cells expressing either wild-type or cleavage-impaired variants of TYRO3. These findings also

483 represent the first comprehensive characterization of signaling pathways stimulated by the gamma-
484 secretase-released ICD of a RTK as compared to the canonical signaling *via* the full-length receptor.

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

494 # Materials and methods

495

496 A more comprehensive description of the materials and methods can be found in the Supplementary
497 Materials and Methods.

498

499 **The inference of regulatory complexes and their combination in the DMPA**

500 Matlab R2016a (MathWorks) was used as the coding environment. The pair-wise association
501 between two proteins, phosphosites or transcripts was determined by using a combined score. The
502 combined score is derived from the multiplication of the correlation and the stoichiometry score the
503 mathematical formulations of which are presented in Figure 1A. The correlation score for all the
504 possible protein, phosphosite or transcript combinations was calculated with the Spearman
505 correlation, ranked and adjusted to derive scores from 1 to 0 in equal increments. The stoichiometry
506 score for all the possible protein, phosphosite or transcript combinations was calculated by dividing
507 the third quartile value (Q_3) of relative abundances with the first quartile (Q_1) value and similarly
508 ranked and adjusted to derive scores from 1 to 0 in equal increments. Two versions of the
509 stoichiometry score were devised for non-zero-inflated and zero-inflated data. In the non-zero-
510 inflated version of the stoichiometry score, relative abundance data were considered only from
511 samples in which the expression value for both intracellular molecules or their states was non-zero.
512 In the zero-inflated version of the stoichiometry score (weighted stoichiometry score), the
513 stoichiometry score is punished (inflated) if one of the two intracellular molecules or their states in
514 the same sample has a missing value. The stoichiometry score, in turn, is rewarded (deflated) if both
515 intracellular molecules or their states have missing values in the same sample.

516 Cut-off parameters C and S were devised to reduce computational time to only consider
517 combined scores higher than a certain threshold. The cut-off parameter C is determined by the
518 minimum number of interactions considered for intracellular molecules that can be set in the analysis.
519 The cut-off parameter S is determined by the number of intracellular molecules or their states in the
520 dataset that are estimated to not have any association.

521 The regulatory complex inference analysis was designed to utilize the nearest neighbor
522 concept to find the highest scoring (sum of the combined score between all 3 members) 3 member
523 networks for all intracellular molecules in the list and to combine and trim the networks based on
524 common network members and combined scores as described in the following order: 1) Networks
525 with 2 common members are combined until no two complexes with 2 common members can be
526 found. 2) Networks with only one common member are combined if the size of both networks is
527 under or equal to the set parameter value and the network score is higher or equal to 2 indicating a
528 strong association. The parameter for this joining (parameter 6 in regulatory complex inference,
529 maximum size of a complex to be joined with only one common node) can be set if smaller regulatory
530 complexes are preferred. 3) Networks with 2 common members are again combined until no two
531 complexes with 2 common members can be found. 4) Two out of the 3 members of the remaining 3
532 member networks are removed if 2 out of three members are already present in another complex. 5)
533 The remaining 3 member networks are joined if they still have a common member. This step can be
534 omitted if smaller or larger regulatory complexes are preferred (parameter 7 in regulatory complex
535 inference). The analysis allows for an intracellular molecule or their state to be a member for more
536 than one regulatory complex.

537 The same strategy was used for the combination of the regulatory complexes into larger
538 modules as was the derivation of the regulatory complexes in the first place with a few modifications.
539 The median value of all complex members in each sample was used to combine the regulatory
540 complexes and the absolute value of the Spearman correlation was used to allow equal combination

541 of up- and downregulated regulatory complexes. The default settings were used for the analysis of
542 the TYRO3 signaling pathways.

543

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

545 The data on known protein-protein interactions, kinase substrate phosphosite relationships,
546 transcription factor target gene relationships, methylation trait relationships and metabolic reactions
547 were acquired from the STRING, PhosphoSitePlus, ChEA3, and Rhea databases, respectively [26–
548 29,46,47]. The data on annotated signaling pathways was acquired from PathwayCommons [31]. The
549 validation interactome, phosphoproteome, transcription and metabolomics data were acquired from
550 the publications of Bath et al and Karayel et al, and from ArchS4 and MetaboLights database,
551 respectively [24,25,48,49]. The methylation data and multi-omics data for the regulatory complex
552 combination to pathways was acquired from the LinkedOmics database [30], where a combination of
553 either proteomics, transcriptomics, methylomics, phosphoproteomics, protein array or acetylation
554 data from the same samples were analyzed and combined. The sum of all protein-protein interactions,
555 kinase substrate phosphosite relationships or transcription factor target gene relationships in the
556 modeled regulatory complexes was used as the validation score for interactome, phosphoproteome
557 and transcriptome regulatory complex validation, respectively. The sum of co-occurrences of two
558 complex members from two different combined complexes in the same pathway annotation for all
559 the combined complexes was used as a validation score for regulatory complex combination. To
560 validate the regulatory complexes, the validation score from the modelled regulatory complexes was
561 compared to the sum of all protein-protein interactions, kinase substrate phosphosite relationships or
562 transcription factor target gene relationships in randomly modeled regulatory complexes of same size
563 to derive an empirical probability density function from 1000 rounds of simulation. To validate the
564 combination of regulatory complexes into pathways, the validation score from the modelled pathways

565 were compared to the sum of all co-occurrences of two complex members from two different
566 randomly modelled and combined complexes of same size in the same pathway to derive an empirical
567 probability density function from 1000 rounds of simulation. An Epanechnikov kernel was used to
568 fit the simulated data into the empirical probability density function. The corresponding P-values
569 were drawn from the empirical cumulative distribution function. The relative distance of the
570 validation score from the median of the empirical probability distribution was calculated by
571 subtracting the median value from the validation score and by dividing the remainder with the median
572 value. To assess the effect of different scores on the accuracy of the modeled regulatory complexes
573 and pathways, the complexes were modeled with either correlation score, stoichiometry score or the
574 combined score and combined with either correlation score, stoichiometry score or the combined
575 score. The conservation of the different scores was assessed in each data type by deriving the median
576 score of the modeled complexes in different datasets from the one used for the initial modeling.

577

578

579 **Benchmarking**

580 The same datasets that were used for the validation of the regulatory complexes and pathways inferred
581 with DMPA were used for benchmarking. The R implementations of WGCNA [3] , LOPC [50],
582 GENIE3 [32] and TransNet [33] was used according to developers instructions. The power for the
583 WGCNA analysis was set based on the scale free topology model fit and mean connectivity as
584 instructed by the WGCNA documentation. The genes predicted to be controlled by the same upstream
585 gene in GENIE3 were considered as part of the same module. The validation score and statistical
586 significance of the validation score was estimated as for the regulatory complexes for modules
587 derived with WGCNA and GENIE3. Since LOPC only returns network edges, a separate validation
588 strategy to compare the performance of LOPC and DMPA was developed. The amount of known

589 transcription factor target gene relationships in the network edges returned by LOPC as determined
590 by the ENCODE database [28] was calculated and used as a validation score. The validation score
591 was compared to the sum of all transcription factor target gene relationships in randomly modeled
592 network of same size to derive an empirical probability density function from 1000 rounds of
593 simulation. An Epanechnikov kernel was used to fit the simulated data into the empirical probability
594 density function. The corresponding P-values and relative distance from the random median were
595 estimated from the empirical cumulative distribution function. To compare DMPA against LOPC the
596 amount of known transcription factor target gene relationships in the final edges in the regulatory
597 complexes as suggested by the remaining combined scores after cut-off and trimming were
598 considered as network edges. The validation score and the empirical probability distribution were
599 similarly estimated for the networks suggested by DMPA. The validation score for the pathways
600 suggested by the TransNet analysis were estimated as those for the DMPA. The clusters connected
601 by the connected nodes were considered a pathway. A version of the pathway validation method
602 where only the connections between different modules were randomized was developed and applied
603 to both TransNet and DMPA.

604

605 **Immunofluorescence and confocal microscopy**

606 To detect ectopically expressed V5-tagged TYRO3 constructs, WM-266-4 transfectants were
607 cultured on coverslips in serum-free conditions overnight. The cells were fixed with 3%
608 paraformaldehyde, and permeabilized with 0.1% Triton X-100. Cells were stained with anti-V5
609 (13202; Cell Signaling Technologies) and AlexaFluor 555 goat anti-rabbit (Molecular Probes). After
610 labeling the nuclei with 4',6-diamid-ino-2-phenylindole (DAPI; Sigma-Aldrich), the cells were
611 mounted with Mowiol 40-88 (Sigma-Aldrich). Images were acquired with Zeiss LSM 880 confocal

612 microscope (Zeiss). Morphological analysis were carried out from confocal slices in plane with the
613 plasma membrane using MorphoLibJ v 1.4.2.1 [51] plugin in Image J.

614

615 **Live-cell imaging of cell growth**

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

619

620 **Real-time cell adhesion assay**

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

626

627

628 **Information visualization and statistics**

629 For statistical analysis, the GraphPad Prism software (Version 9.02; GraphPad Software) and Matlab
630 2016a (MathWorks) were utilized. The details of the statistical testing are described in figure legends.
631 All datasets were tested for normality and parametric or non-parametric testing was used accordingly.
632 Heatmaps were generated with Morpheus (https://software.broadinstitute.org/morpheus). The

633 visualizations of the inferred regulatory complexes and pathways were created with Cytoscape
634 (Version 3.8.2; Shannon *et al*, 2003) and Affinity Designer (Version 19.0; Serif)

635

636 **Data availability**

637 The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium
638 via the PRIDE partner repository with the dataset identifier PXD028881. Reviewer account details:
639 Username: reviewer_pxd028881@ebi.ac.uk Password: 9e92pZvD. The RNAseq data have been
640 deposited to the Gene Expression Omnibus database with the identifier: GSE190431. The reviewer
641 token (password) to access the dataset in GEO: qxohymmyfvivdyp. The DMPA is available both as
642 matlab code and executables in Mendeley Data:
643 https://data.mendeley.com/datasets/m3zggn6xx9/draft?a=71c29dac-714e-497e-8109-5c324ac43ac3.

644

645 **Source data files**

646 The raw source data for the western blots in Supplementaty Figure 12 are available in
647 Supplemenrary Figure 12 – source data 1.zip file. The zipped folder includes both the raw images
648 and the Supplemenrary Figure 12 Annotated Immunoblots.pdf file where the bands presented in
649 Supplemenrary Figure 12 are annotated and indicated in the full gel images. The raw source data for
650 Figure 6A and B is available in the Figure6_source_data.xlsx file that includes the numerical values
651 of all replicate experiments of Figure 6A and B.

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, the Paavo Nurmi Foundation, the Aarne Koskelo Foundation, the Paulo Foundation, 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 DMPA. TYRO3 knock-down cells were prepared by VKO. KV and KE wrote the manuscript with the help of JAMM.

References

1. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A [Internet]. Proc Natl Acad Sci U S A; 2005 [cited 2021 Mar 1];102:15545–50. Available from: https://pubmed.ncbi.nlm.nih.gov/16199517/

2. Ma J, Shojaie A, Michailidis G. A comparative study of topology-based pathway enrichment analysis methods. BMC Bioinformatics [Internet]. BioMed Central Ltd.; 2019 [cited 2021 Mar 1];20:546. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-3146-1

3. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics [Internet]. BioMed Central; 2008 [cited 2021 Mar 1];9:559. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-559

4. Schäfer J, Strimmer K. Learning Large-Scale Graphical Gaussian Models from Genomic Data. AIP Conf Proc. 2005;776:263–76.

5. Zuo Y, Yu G, Tadesse MG, Ressom HW. Biological network inference using low order partial correlation. Methods [Internet]. NIH Public Access; 2014 [cited 2023 Jan 9];69:266. Available from: /pmc/articles/PMC4194134/

6. De La Fuente A, Bing N, Hoeschele I, Mendes P. Discovery of meaningful associations in genomic data using partial correlation coefficients. BIOINFORMATICS [Internet]. 2004 [cited 2023 Jan 9];20:3565–74. Available from: http://mendes.vbi.vt.edu/tiki-index.php?page=Software

7. Chen S, Mar JC. Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. BMC Bioinformatics [Internet]. BioMed Central Ltd.; 2018 [cited 2021 Mar 1];19:232. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-018-2217-z

8. Pratapa A, Jalihal AP, Law JN, Bharadwaj A, Murali TM. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. Nat Methods [Internet]. Nature Research; 2020 [cited 2021 Mar 1];17:147–54. Available from: https://www.nature.com/articles/s41592-019-0690-6

9. Köksal AS, Beck K, Cronin DR, McKenna A, Camp ND, Srivastava S, et al. Synthesizing Signaling Pathways from Temporal Phosphoproteomic Data. Cell Rep [Internet]. Elsevier B.V.; 2018 [cited 2021 Mar 1];24:3607–18. Available from: https://pubmed.ncbi.nlm.nih.gov/30257219/

10. Vahabi N, Michailidis G. Unsupervised Multi-Omics Data Integration Methods: A Comprehensive Review. Front Genet [Internet]. Frontiers Media S.A.; 2022 [cited 2023 Jan 9];13. Available from: /pmc/articles/PMC8981526/

11. Cantini L, Zakeri P, Hernandez C, Naldi A, Thieffry D, Remy E, et al. Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer. Nat Commun 2021 121 [Internet]. Nature Publishing Group; 2021 [cited 2023 Jan 10];12:1–12. Available from: https://www.nature.com/articles/s41467-020-20430-7

12. Leng D, Zheng L, Wen Y, Zhang Y, Wu L, Wang J, et al. A benchmark study of deep learning-based multi-omics data fusion methods for cancer. Genome Biol [Internet]. BioMed Central Ltd; 2022 [cited 2023 Jan 10];23:1–32. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-022-02739-2

13. Zoppi J, Guillaume JF, Neunlist M, Chaffron S. MiBiOmics: an interactive web application for

multi-omics data exploration and integration. BMC Bioinformatics [Internet]. BioMed Central Ltd; 2021 [cited 2023 Jan 10];22:1–14. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-020-03921-8

14. Eicher T, Kinnebrew G, Patt A, Spencer K, Ying K, Ma Q, et al. Metabolomics and multi-omics integration: A survey of computational methods and resources. Metabolites [Internet]. 2020 [cited 2023 Jan 10];10. Available from: www.mdpi.com/journal/metabolites

15. Bonnet E, Calzone L, Michoel T. Integrative multi-omics module network inference with Lemon-Tree. PLoS Comput Biol [Internet]. PLoS Comput Biol; 2015 [cited 2023 Jan 10];11. Available from: https://pubmed.ncbi.nlm.nih.gov/25679508/

16. Shoaie S, Karlsson F, Mardinoglu A, Nookaew I, Bordel S, Nielsen J. Understanding the interactions between bacteria in the human gut through metabolic modeling. Sci Rep [Internet]. Sci Rep; 2013 [cited 2023 Jan 10];3. Available from: https://pubmed.ncbi.nlm.nih.gov/23982459/

17. Vaske CJ, Benz SC, Sanborn JZ, Earl D, Szeto C, Zhu J, et al. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. Bioinformatics [Internet]. Bioinformatics; 2010 [cited 2023 Jan 10];26. Available from: https://pubmed.ncbi.nlm.nih.gov/20529912/

18. Dugourd A, Kuppe C, Sciacovelli M, Gjerga E, Gabor A, Emdal KB, et al. Causal integration of multi-omics data with prior knowledge to generate mechanistic hypotheses. Mol Syst Biol [Internet]. John Wiley & Sons, Ltd; 2021 [cited 2023 Jan 10];17:e9730. Available from: https://onlinelibrary.wiley.com/doi/full/10.15252/msb.20209730

19. Liu A, Trairatphisan P, Gjerga E, Didangelos A, Barratt J, Saez-Rodriguez J. From expression footprints to causal pathways: contextualizing large signaling networks with CARNIVAL. npj Syst Biol Appl [Internet]. 2019 [cited 2023 Jan 10];5. Available from: https://doi.org/10.1038/s41540-019-0118-z

20. Koh HWL, Fermin D, Vogel C, Choi KP, Ewing RM, Choi H. iOmicsPASS: network-based integration of multiomics data for predictive subnetwork discovery. npj Syst Biol Appl [Internet]. 2019 [cited 2023 Jan 10];5. Available from: https://doi.org/10.1038/s41540-019-0099-y

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

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

23. Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, Toyoda Y, et al. A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. Cell. 2015;163:712–23.

24. Batth TS, Papetti M, Pfeiffer A, Tollenaere MAX, Francavilla C, Olsen J V. Large-Scale Phosphoproteomics Reveals Shp-2 Phosphatase-Dependent Regulators of Pdgf Receptor Signaling. Cell Rep. ElsevierCompany.; 2018;22:2784–96.

25. Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, et al. Massive mining of publicly available RNA-seq data from human and mouse. Nat Commun. Springer US; 2018;9.

26. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: A database of predicted functional associations between proteins [Internet]. Nucleic Acids Res. Nucleic Acids Res; 2003 [cited 2021 Mar 1]. p. 258–61. Available from: https://pubmed.ncbi.nlm.nih.gov/12519996/

27. Hornbeck P V., Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, et al. PhosphoSitePlus: A comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res [Internet]. Nucleic Acids Res; 2012 [cited 2021 Mar 1];40. Available from: https://pubmed.ncbi.nlm.nih.gov/22135298/

28. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The Encyclopedia of DNA elements (ENCODE): Data portal update. Nucleic Acids Res [Internet]. Oxford University Press; 2018 [cited 2021 Mar 1];46:D794–801. Available from: https://pubmed.ncbi.nlm.nih.gov/29126249/

29. Keenan AB, Torre D, Lachmann A, Leong AK, Wojciechowicz ML, Utti V, et al. ChEA3: transcription factor enrichment analysis by orthogonal omics integration. Nucleic Acids Res [Internet]. Oxford University Press; 2019 [cited 2021 Mar 1];47:W212–24. Available from: https://pubmed.ncbi.nlm.nih.gov/31114921/

30. Vasaikar S V., Straub P, Wang J, Zhang B. LinkedOmics: Analyzing multi-omics data within and across 32 cancer types. Nucleic Acids Res. Oxford University Press; 2018;46:D956–63.

31. Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur Ö, Anwar N, et al. Pathway Commons, a web resource for biological pathway data. Nucleic Acids Res. 2011;39:685–90.

32. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. PLoS One [Internet]. Public Library of Science; 2010 [cited 2023 Jan 19];5:e12776. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0012776

33. Rodrigues RR, Shulzhenko N, Morgun A. Transkingdom Networks: A Systems Biology Approach to Identify Causal Members of Host–Microbiota Interactions. Methods Mol Biol [Internet]. Humana Press Inc.; 2018 [cited 2023 Jan 10];1849:227–42. Available from: https://link.springer.com/protocol/10.1007/978-1-4939-8728-3_15

34. Merilahti JAM, Ojala VK, Knittle AM, Pulliainen AT, Elenius K. Genome-wide screen of gamma-secretase–mediated intramembrane cleavage of receptor tyrosine kinases. Heldin C-H, editor. Mol Biol Cell [Internet]. 2017;28:3123–31. Available from: http://www.molbiolcell.org/doi/10.1091/mbc.e17-04-0261

35. Tsai CL, Chang JS, Yu MC, Lee CH, Chen TC, Chuang WY, et al. Functional genomics identifies hepatitis-induced STAT3-TyrO3-STAT3 signaling as a potential therapeutic target of hepatoma. Clin Cancer Res. 2020;26:1185–97.

36. Chen D, Liu Q, Cao G, Zhang W. TYRO3 facilitates cell growth and metastasis via activation of the Wnt/β-catenin signaling pathway in human gastric cancer cells. Aging (Albany NY). 2020;12:2261–74.

37. Dufour F, Silina L, Neyret-Kahn H, Moreno-Vega A, Krucker C, Karboul N, et al. TYRO3 as a molecular target for growth inhibition and apoptosis induction in bladder cancer. Br J Cancer. Springer US; 2019;120:555–64.

38. Zhu S, Wurdak H, Wang Y, Galkin A, Tao H, Li J, et al. A genomic screen identifies TYRO3 as a MITF regulator in melanoma. Proc Natl Acad Sci. 2009;106:17025–30.

39. Kraehn GM, Utikal J, Udart M, Greulich KM, Bezold G, Kaskel P, et al. Extra c-myc oncogene copies in high risk cutaneous malignant melanoma and melanoma metastases. Br J Cancer [Internet].

Br J Cancer; 2001 [cited 2021 Nov 22];84:72–9. Available from: https://pubmed.ncbi.nlm.nih.gov/11139316/

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

41. Guo H, Barrett TM, Zhong Z, Fernndez JA, Griffin JH, Freeman RS, et al. Protein S blocks the extrinsic apoptotic cascade in tissue plasminogen activator/N-methyl D-aspartate-treated neurons via Tyro3-Akt-FKHRL1 signaling pathway. Mol Neurodegener [Internet]. Mol Neurodegener; 2011 [cited 2021 Nov 29];6. Available from: https://pubmed.ncbi.nlm.nih.gov/21291561/

42. Fujita K, Chen X, Homma H, Tagawa K, Amano M, Saito A, et al. Targeting Tyro3 ameliorates a model of PGRN-mutant FTLD-TDP via tau-mediated synaptic pathology. Nat Commun 2018 91 [Internet]. Nature Publishing Group; 2018 [cited 2021 Nov 29];9:1–21. Available from: https://www.nature.com/articles/s41467-018-02821-z

43. Demarest SJ, Gardner J, Vendel MC, Ailor E, Szak S, Huang F, et al. Evaluation of Tyro3 expression, Gas6-mediated akt phosphorylation, and the impact of anti-Tyro3 antibodies in melanoma cell lines. Biochemistry. 2013;52:3102–18.

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

45. Tworkoski KA, Singhal G, Szpakowski S, Zito CI, Bacchiocchi A, Muthusamy V, et al. Phosphoproteomic Screen Identifies Potential Therapeutic Targets in Melanoma. Mol Cancer Res. 2011;9:801–12.

46. Chèneby J, Gheorghe M, Artufel M, Mathelier A, Ballester B. ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. Nucleic Acids Res [Internet]. Oxford Academic; 2018 [cited 2022 Jan 14];46:D267–75. Available from: https://academic.oup.com/nar/article/46/D1/D267/4602873

47. Bansal P, Morgat A, Axelsen KB, Muthukrishnan V, Coudert E, Aimo L, et al. Rhea, the reaction knowledgebase in 2022. Nucleic Acids Res [Internet]. Oxford University Press; 2022 [cited 2023 Jan 19];50:D693. Available from: /pmc/articles/PMC8728268/

48. Karayel Ö, Xu P, Bludau I, Velan Bhoopalan S, Yao Y, Ana Rita FC, et al. Integrative proteomics reveals principles of dynamic phosphosignaling networks in human erythropoiesis. Mol Syst Biol. 2020;16:1–22.

49. Haug K, Cochrane K, Nainala VC, Williams M, Chang J, Jayaseelan KV, et al. MetaboLights: a resource evolving in response to the needs of its scientific community. Nucleic Acids Res [Internet]. Oxford Academic; 2020 [cited 2023 Jan 19];48:D440–4. Available from: https://academic.oup.com/nar/article/48/D1/D440/5613675

50. Guo W. RLowPC: Inference of co-expression gene network using relevance low order partial correlation from large scale expression data. 2016.

51. Legland D, Arganda-Carreras I, Andrey P. MorphoLibJ: integrated library and plugins for mathematical morphology with ImageJ. Bioinformatics [Internet]. Oxford Academic; 2016 [cited 2021 Sep 21];32:3532–4. Available from: https://academic.oup.com/bioinformatics/article/32/22/3532/2525592

52. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 2003;13:2498–504.

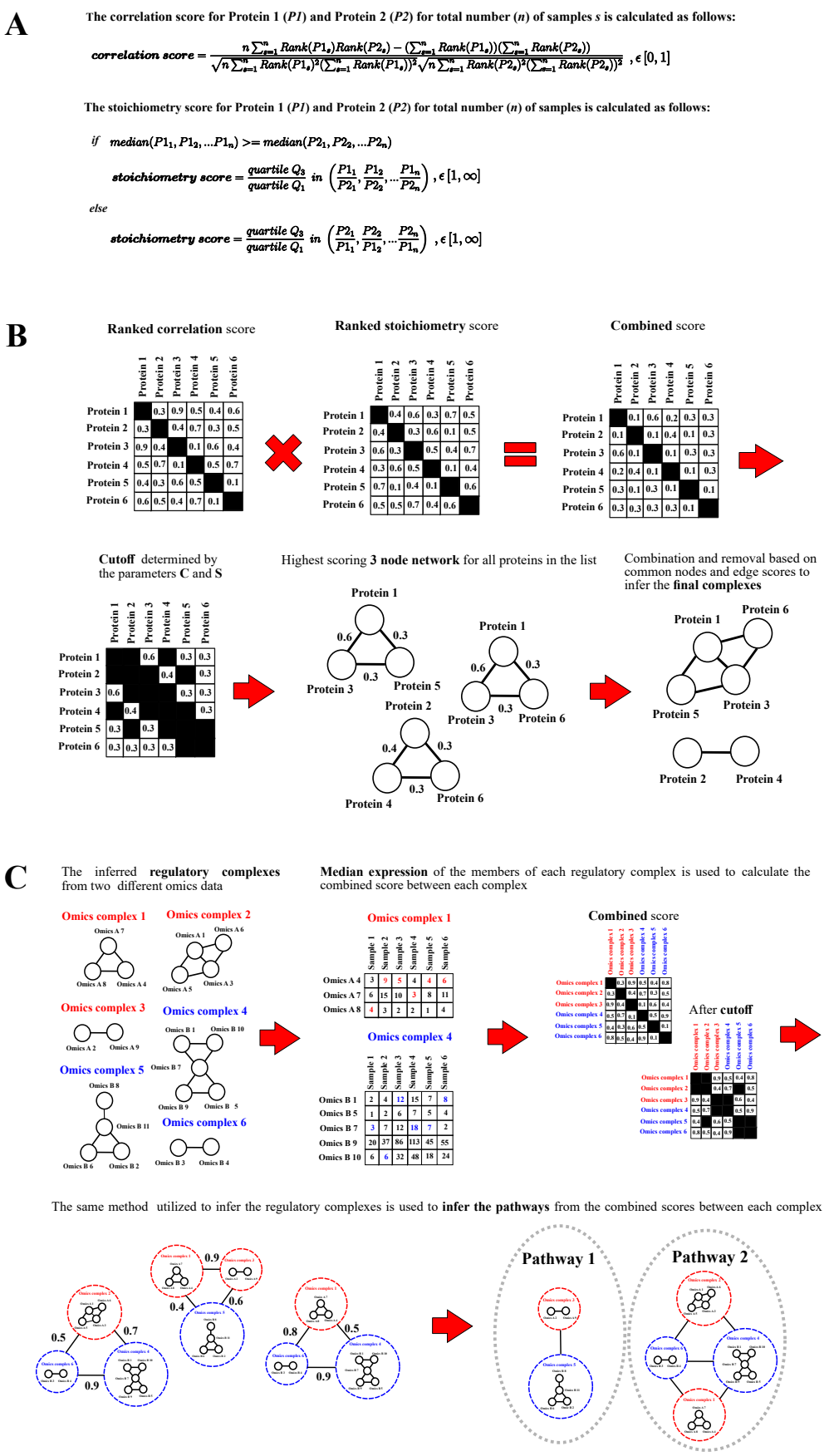


Figure 1. The underlying concepts of the de novo multiomics pathway analysis.
A: Mathematical formulation of the correlation (Spearman correlation) and stoichiometry scores.
B: Schematic representation of the inference of regulatory omics complexes in the de novo multi-omics pathway analysis.
C: Schematic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

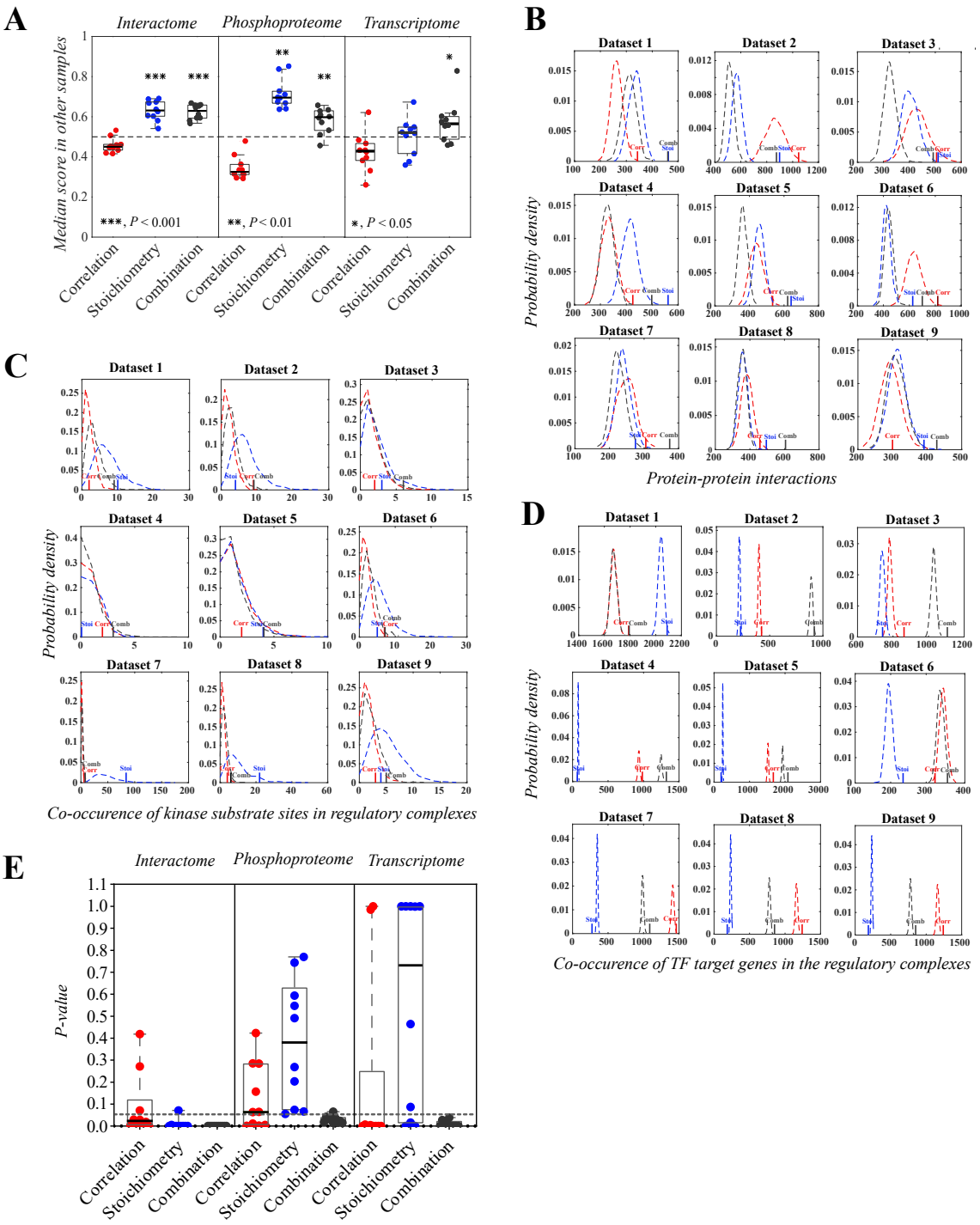


Figure 2. Validation of regulatory complexes modeled with the DMPA.
A: The conservation of correlation, stoichiometry and combined score in interactome, phosphoproteome and transcriptome data. One dot represents the median score between all complex members of all modeled complexes of a dataset that was not utilized for the initial regulatory complex inference. The dashed line represents the median score of randomized complexes. For statistical testing, two-tailed one sample T-test against the theoretical median value of 0.5 was utilized.
B-E: Empirical probability densities simulated by randomizing the members of the modeled regulatory complexes into complexes of the same size and the corresponding P-values for the complexes modeled from either interactome (B, E), phosphoproteome (C, E) or transcriptome data (D, E) by either correlation (Corr), stoichiometry (Stoi) or combination score (Comb). The sum of co-occurrences of either interacting proteins as determined by the STRING database, phosphorylated kinase substrate residues as determined by the PhosphoSitePlus database or transcription factor target genes as determined by the ENCODE database in the modeled complexes were used as a validation score. The empirical probability densities are visualized with dashed lines of the corresponding color and the score for the modeled network in solid line. One dot in the boxplot represents one P-value for one dataset and the horizontal line the median value. The datasets were acquired from published omics data. Red: correlation score; blue: stoichiometry score; grey: combined score.

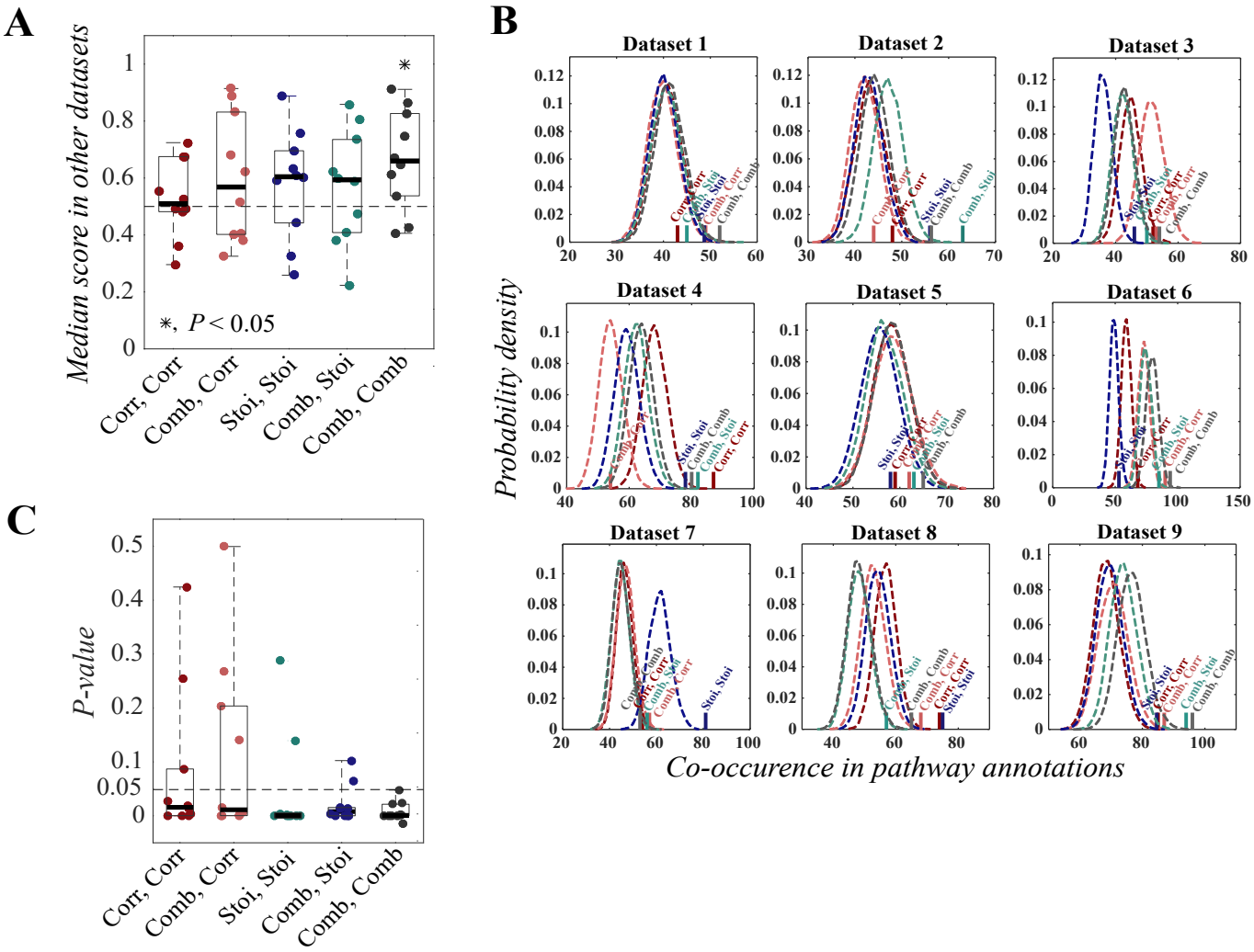


Figure 3. Validation of regulatory complex combination to pathways with DMPA.
A: The conservation of correlation, stoichiometry and combined score in complexes modeled with either correlation (Corr), stoichiometry (Stoi) or combined (Comb) score and combined with either correlation, stoichiometry or 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 line the median value. The datasets were acquired from published multi-omics data.

A

	Dataset 1			Dataset 2			Dataset 3			Dataset 4			Dataset 5		
Dataset size	n: 6 transcripts: 3077			n: 91 transcripts: 1697			n: 104 transcripts: 2003			n: 105 transcripts: 2345			n: 100 transcripts: 3102		
Method	WGCNA	GENIE3	DMPA	WGCNA	GENIE3	DMPA	WGCNA	GENIE3	DMPA	WGCNA	GENIE3	DMPA	WGCNA	GENIE3	DMPA
P-value	0.617	0.980	0.000	0.052	0.000	0.027	0.025	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000
Distance from randomized median	-0.0062	-0.0087	0.0706	0.0426	0.0273	0.0314	0.0647	0.0778	0.0748	0.0566	0.0302	0.0602	0.1015	0.1082	0.0777
Transcripts included in the final modules (%)	100.00	99.64	65.13	100.00	90.81	98.82	100.00	85.62	91.91	100.00	69.13	85.46	100.00	0.00	86.40
Run time (min)	0.63	1.17	12.82	0.29	25.23	26.93	0.23	37.06	28.59	0.52	56.43	60.75	0.54	74.40	154.29

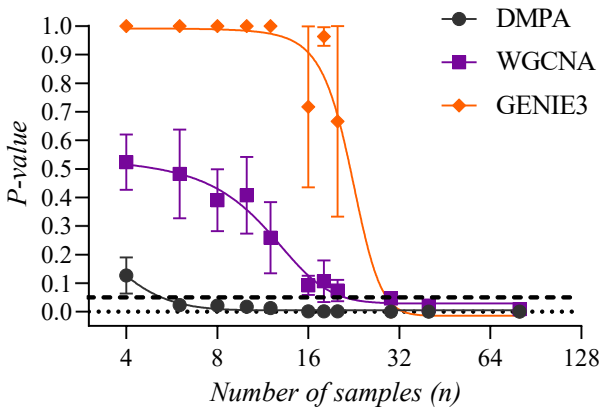
	Dataset 6			Dataset 7			Dataset 8			Dataset 9			Dataset 10		
Dataset size	n: 346 transcripts: 958			n: 188 transcripts: 3119			n: 12 transcripts: 3119			n: 62 transcripts: 3187			n: 175 transcripts: 2601		
Method	WGCNA	GENIE3	DMPA	WGCNA	GENIE3	DMPA	WGCNA	GENIE3	DMPA	WGCNA	GENIE3	DMPA	WGCNA	GENIE3	DMPA
P-value	0.716	1.000	0.039	0.000	0.000	0.000	0.294	1.000	0.022	0.000	0.000	0.000	0.025	0.000	0.000
Distance from randomized median	-0.0387	-0.0983	0.0588	0.0882	0.0827	0.1013	0.0097	-0.2200	0.0310	0.1094	0.0886	0.0937	0.0497	0.0261	0.0945
Transcripts included in the final modules (%)	100.00	99.37	96.35	100.00	90.38	97.02	100.00	70.79	61.33	100.00	80.29	98.09	100.00	95.39	97.35
Run time (min)	0.13	71.83	4.82	0.99	184.45	112.39	1.32	2.76	10.40	1.06	36.34	164.09	0.53	116.49	73.92

B

	Dataset 1		Dataset 2		Dataset 3		Dataset 4		Dataset 5	
	n:6 transcripts: 3077		n:91 transcripts: 1697		n:104 transcripts: 2003		n:105 transcripts: 1086		n:100 transcripts: 3102	
Method	LOPC	DMPA	LOPC	DMPA	LOPC	DMPA	LOPC	DMPA	LOPC	DMPA
P-value	NA	0.000	0.001	0.000	0.000	0.000	0.002	0.000	0.021	0.000
Relative distance from random median	NA	0.085	0.057	0.131	0.058	0.190	0.094	0.278	0.072	0.234
Transcripts included in the network (%)	NA	65.13	27.87	98.82	22.97	91.91	29.47	97.88	4.80	86.40
Run time (min)	NA	12.82	347.70	26.93	497.40	28.59	4.67	6.29	28.78	154.29

	Dataset 6		Dataset 7		Dataset 8		Dataset 9		Dataset 10	
	n:346 transcripts: 958		n:188 transcripts: 3119		n:12 transcripts: 3119		n:62 transcripts: 3187		n:175 transcripts: 2601	
Method	LOPC	DMPA	LOPC	DMPA	LOPC	DMPA	LOPC	DMPA	LOPC	DMPA
P-value	0.260	0.000	0.000	0.000	NA	0.000	0.000	0.000	0.000	0.000
Relative distance from random median	0.250	0.386	0.906	0.385	NA	0.116	0.275	0.725	0.157	0.205
Transcripts included in the network (%)	16.28	96.35	3.49	97.02	NA	61.33	20.93	98.09	44.60	97.35
Run time (min)	30.84	4.82	90.16	112.39	NA	10.40	18.72	164.09	107.14	73.92

C



D

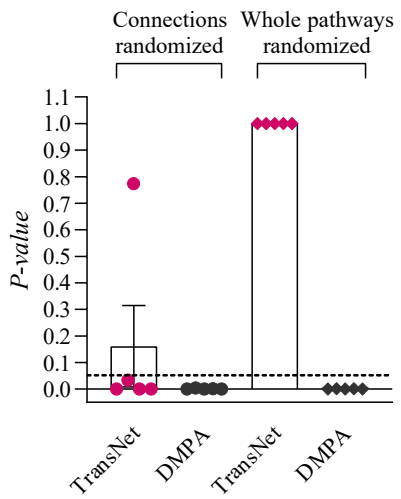


Figure 4. The performance of DMPA against previous methods.
A: The performance of DMPA against WGCNA and GENIE3 in module discovery. The amount of known transcription factor target gene relationships in transcriptome data, runtime and percentage of transcripts included in the inferred modules was assessed.
B: The performance of DMPA against LOPC in network discovery. The amount of known transcription factor target gene relationships in transcriptome data, runtime and percentage of transcripts included in the inferred networks was assessed.
C: The performance of DMPA against WGCNA and GENIE3 in module discovery in transcriptome datasets with different sample sizes. The statistical signficance of the amount of known transcription factor target gene relationships in the inferred modules was assessed. Mean ± SEM.
D: The performance of DMPA against TransNet in pathway discovery. Multi-omics data was analyzed with both methods and the amount of known associations in pathway annotations from the pathway Commons database in the inferred pathways was assessed. The statistical significance of the amount of the known associations in the inferred pathways was determined with two different statistical methods, where either only the connections between the modules or the whole pathways were randomized. Mean ± SEM.

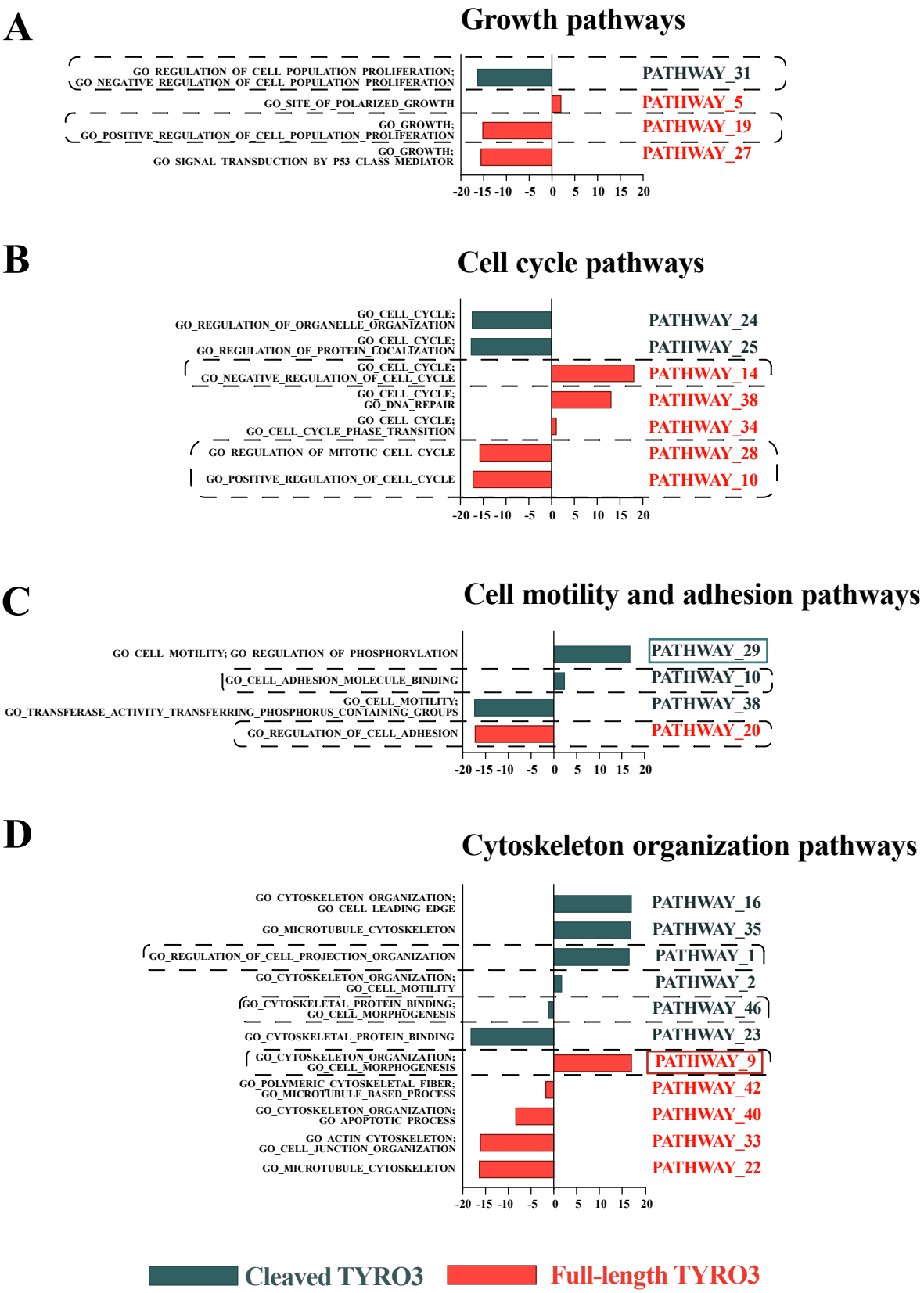


Figure 5. The functional categorization of the selected full-length and cleaved TYRO3 pathways inferred with the DMPA. The predicted function of the multi-omics pathways of the full-length and the cleaved ICD of TYRO3 inferred with the DMPA. The bars represent the median pseudolog2 fold change of all 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 21. For further details on the pathways see supplemental tables 1 and 2. The cell behavior predicted by the pathways surrounded with the dashed line is examined in Figure 6A-C.The highlighted pathways are visualized in Figure 6D-E.

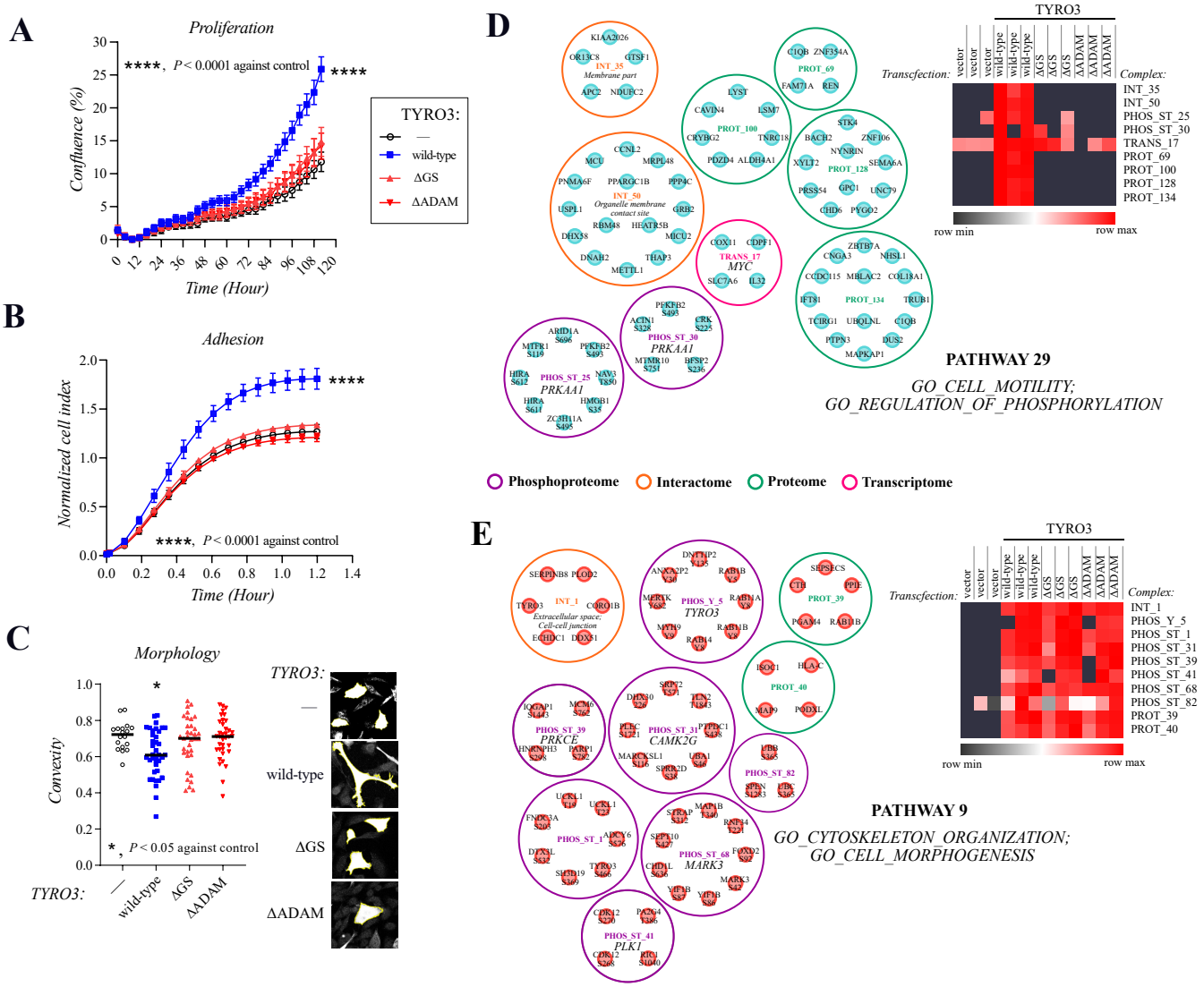


Figure 6. Functional validation and composition of the selected pathways of full-length and cleaved ICD of TYRO3. A: Proliferation of WM-266-4 transfectants was measured with live-cell imaging. For statistical testing, the non-parametric Friedman 2-way ANOVA and the Dunn's multiple comparison test with a multiple test correction for the P-values was utilized. The symbols represent the mean and the whiskers the standard deviation of the values. A representative plot from 1 out of 3 independent replicate experiments is shown (n=6). **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.The symbols represent the mean and the whiskers the standard deviation of the values. A representative plot from 1 out of 4 independent replicate experiments is shown (n=5). **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 cleaved 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.