

1 **Multi-omics analysis reveals drivers of loss of β -cell function after**
2 **newly diagnosed autoimmune type 1 diabetes:**
3 **An INNODIA[‡] multicenter study**

4 Jose Juan Almagro Armenteros, PhD^{1†}, Caroline Brorsson, PhD^{1†}, Christian Holm Johansen,
5 MSc^{1†}, Karina Banasik, PhD¹, Gianluca Mazzoni, PhD¹, Robert Moulder, PhD^{2,3}, Karoliina
6 Hirvonen, MSc^{2,3}, Tomi Suomi, PhD^{2,3}, Omid Rasool, PhD^{2,3}, Sylvaine FA Bruggraber,
7 PhD⁴, M Loredana Marcovecchio, MD⁴, Emile Hendricks, MD⁴, Naba Al-Sari, PhD⁵, Ismo
8 Mattila, MSc⁵, Cristina Legido-Quigley, PhD⁵, Tommi Suvitaival⁵, PhD, Piotr J Chmura,
9 PhD¹, Mikael Knip, MD^{6,7}, Anke M Schulte, MD⁸, Jeong Heon Lee, PhD⁹, Guido Sebastiani,
10 PhD^{10,11}, Giuseppina Emanuela Grieco, PhD^{10,11}, Laura L Elo, PhD^{2,3,12}, Simranjeet Kaur,
11 PhD¹³, Flemming Pociot, MD¹³, Francesco Dotta, MD^{10,14}, Tim Tree, PhD¹⁵, Riitta
12 Lahesmaa, MD^{2,3}, Lut Overbergh, PhD¹⁶, Chantal Mathieu, MD¹⁶, Mark Peakman, MD^{9,15},
13 Søren Brunak, PhD^{1*} on behalf of the INNODIA[‡] investigators

14
15 ¹Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical
16 Sciences, University of Copenhagen, Copenhagen, Denmark

17 ²Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland

18 ³InFLAMES Research Flagship Center, University of Turku, Turku Finland

19 ⁴Department of Paediatrics, University of Cambridge, Cambridge, England

20 ⁵Steno Diabetes Center Copenhagen, Systems Medicine, Herlev, Denmark

21 ⁶Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University
22 of Helsinki, Helsinki, Finland

23 ⁷Pediatric Research Center, New Children's Hospital, Helsinki University Hospital

24 ⁸Bayer Pharmaceuticals, Berlin, Germany

25 ⁹Immunology & Inflammation Research Therapeutic Area, Sanofi, MA, USA

26 ¹⁰Department of Medicine, Surgery and Neuroscience, Università degli Studi di Siena, Siena,
27 Italy

28 ¹¹ Fondazione Umberto di Mario, ONLUS – Toscana Life Sciences, Siena, Italy

29 ¹²Institute of Biomedicine, University of Turku, Turku, Finland

30 ¹³Steno Diabetes Center Copenhagen, Herlev University Hospital, Region Hovedstaden,
31 Denmark

32 ¹⁴Tuscany Centre for Precision Medicine (CReMeP), Siena, Italy

33 ¹⁵Department of Immunobiology, King's College, London, UK

34 ¹⁶Department of Chronic Diseases and Metabolism, Endocrinology, Katholieke Universiteit

35 **NOTE:** This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Leuven, Leuven, Belgium

36

37 † Shared first author

38 * Corresponding author:

39 Søren Brunak, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and
40 Medical Sciences, University of Copenhagen, Blegdamsvej 3A, DK-2200 Copenhagen,
41 Denmark. Tel: +45-35325000, Email: soren.brunak@cpr.ku.dk.

42 ‡INNODIA: “Innovative approaches to understanding and arresting type 1 diabetes”- List
43 of contributors included with submission.

44

45 **Abstract**

46 **Background**

47 Heterogeneity in the rate of β -cell loss in newly diagnosed type 1 diabetes patients is poorly
48 understood and creates a barrier to designing and interpreting disease-modifying clinical
49 trials. Integrative analyses of complementary multi-omics data obtained after the diagnosis of
50 T1D may provide mechanistic insight into the diverse rates of disease progression.

51 **Methods**

52 We collected samples in a pan-European consortium that enabled the concerted analysis of
53 five different omics modalities in data from 97 newly diagnosed patients. In this study we
54 used Multi-Omics Factor Analysis to identify molecular signatures correlating with post-
55 diagnosis decline in β -cell mass measured as fasting C-peptide.

56 **Results**

57 Two molecular signatures were significantly correlated with fasting C-peptide levels. One
58 signature showed a correlation to neutrophil degranulation, cytokine signaling, lymphoid and
59 non-lymphoid cell interactions and G-protein coupled receptor signaling events that were
60 inversely associated with rapid decline in β -cell function. The second signature was related to
61 translation and viral infection were inversely associated with change in β -cell function. In
62 addition, the immunomics data revealed a Natural Killer cell signature associated with rapid
63 β -cell decline.

64 **Conclusion**

65 Features that differ between individuals with slow and rapid decline in β -cell mass could be
66 valuable in staging and prediction of the rate of disease progression and thus enable smarter
67 (shorter and smaller) trial designs for disease modifying therapies, as well as offering
68 biomarkers of therapeutic effect.

69

70 **Funding**

71 This work is funded by the Innovative Medicine Initiative 2 Joint Undertaking (IMI2 JU)
72 under grant agreement N° 115797 (INNODIA) and N° 945268 (INNODIA HARVEST). This
73 Joint Undertaking receives support from the Union's Horizon 2020 research and innovation
74 program and 'EFPIA', 'JDRF' and 'The Leona M. and Harry B. Helmsley Charitable Trust'.
75

76 **Introduction**

77 Type 1 diabetes is an autoimmune disease involving environmental and genetic factors that
78 trigger immune-mediated pancreatic β -cell dysfunction and destruction that results in insulin
79 loss and symptomatic hyperglycemia requiring lifelong insulin therapy¹. Globally, around 1.2
80 million people below the age of 20 years have type 1 diabetes, with an annually increasing
81 incidence of 3% influenced strongly by geography². Insulin replacement therapy is unable to
82 fully mimic physiological control of blood glucose and therefore, many people living with type
83 1 diabetes develop severe disease complications that are directly attributable to prolonged
84 glycemic exposure³, with markedly reduced life expectancy⁴. In the face of the disease burden
85 and unmet need, international consortia are mobilizing to develop disease-modifying therapies.
86 For example, therapies that maintain even minimal residual C-peptide secretion capacity have
87 been found to have demonstrable clinical benefit⁵. An emerging barrier to this effort is disease
88 heterogeneity. In particular, the rate of decline of β -cell capacity is highly variable, for reasons
89 that remain unclear. As a result, clinical trial designs for disease-modifying therapies are
90 necessarily cumbersome, requiring large sample sizes and prolonged observation. In addition,
91 opportunities for tailored disease-modifying therapies are limited by an unclear understanding
92 of the factors that drive diabetes progression after diagnosis. Gaining this knowledge could
93 provide the attainment for improved participant inclusion in focused designed trials to foster
94 their success and participant benefit⁶.

95
96 Studies with this goal conducted to date have typically been constrained by limitations to cohort
97 size and the number of different data dimensions available for analysis^{7,8}. Thus, it has not been
98 possible to conduct studies with an emphasis on hypothesis-generating, unbiased approaches,
99 and integration of data across pathophysiological systems. These require large-scale, inter-
100 disciplinary research efforts, in which carefully curated longitudinal clinical cohorts are aligned
101 with multi-parametric technology platforms. Such a systems-based approach can address key
102 questions with less bias and generate novel hypotheses on disease drivers. We used this strategy
103 in the setting of a pan-European research consortium in which people with newly diagnosed
104 type 1 diabetes as well as people at risk of developing type 1 diabetes (antibody positive) were
105 enrolled into a master protocol⁹ to conduct a prospective study in search of factors that correlate
106 with the rate of decline in β -cell mass/function. This endeavor was supported by the Innovative
107 Medicines Initiative-2 Joint Undertaking, where INNODIA was created, being a private-public
108 partnership of 40 partners in 16 countries. In the natural history study, people with newly
109 diagnosed type 1 diabetes and unaffected family members are in follow up, allowing deep

110 clinical characterization, as well as multi-omics analysis of samples (blood, urine, stool)
111 collected and analyzed using standardized operating procedures (www.innodia.eu). Here we
112 report the multi-omics analysis of the ‘first 100’ people with newly diagnosed disease. We
113 report the existence of latent factors integrated from transcriptomic, small RNA, genomic,
114 targeted proteomic, lipidomic, metabolomic, and immunomic-level data that show a
115 relationship to subsequent rates of disease progression and have potential value as stratification
116 and therapeutic target identification tools.

117 **Results**

118 **C-peptide decline over time**

119 The individual rates of decline in C-peptide levels over time were calculated, defined as the
120 slope of C-peptide change over 12 months as described in Methods (Supplementary Fig. 1).
121 Two individuals only had one available C-peptide measurement and were not assigned a group
122 as C-peptide decline could not be determined. An overall trend of C-peptide decline over time
123 was observed for the entire cohort (p-value 0.0001) (Fig. 2A). By calculating the C-peptide
124 slopes using a linear mixed-effect model, the participants were divided into terciles (equal-
125 sized), classified as rapid, slow, and increasing progression groups, yielding three groups with
126 distinct progression patterns (Fig. 2B). All three progression groups had a similar estimated
127 baseline C-peptide value (p-value 0.296), but significant C-peptide slopes (p-value < 0.0001).
128 The clinical features for each of the progression groups are shown in Table 1.

129
130 We further analyzed the relationship between age and C-peptide change over time (Fig. 2C) as
131 this was the only significant difference between the progression groups (Table 1). Participants
132 less than ten years old had a significantly lower baseline C-peptide level compared to those
133 older than ten years (p-value < 0.0001), whereas baseline values did not differ between the 10-
134 18 years and >18 years age groups (p-value 0.85). However, the C-peptide change over time
135 was not significantly different between the three age intervals (p-value 0.46). This indicated
136 that age is associated with baseline C-peptide values but the decline in C-peptide over time is
137 similar for all age groups. In addition, when investigating the distributions of C-peptide decline
138 rates, children < 10 years old had a significantly different distribution of decline rates than
139 those aged 10-18 years (p-value 0.0072). However, neither 10-18 year and >18 year groups nor
140 the >18 year and < 10 year groups had different distributions (p-values 0.075 and 0.7,
141 respectively) (Supplementary Fig. 2). These findings indicate a degree of association between
142 the rate of decline and age among children, and as a result, age was included after log
143 transformation as a covariate in our models to correct for potential effects on the relationship
144 between C-peptide slopes and ‘omics.

145
146 Evaluating the association of sex with the C-peptide change over time (Fig. 2D) we found no
147 significant association with baseline C-peptide (p-value 0.64) or slope (p-value 0.16).

148
149
150

151 **Multi-omics integration analysis**

152 The multi-omics data set overview is shown in Fig. 3A. Missing values, seen predominantly in
153 the immunomics data set, are disregarded by MOFA and do not affect the decomposition of
154 the data into latent factors. After training MOFA on the multi-omics data set (Fig. 3B) the latent
155 factors that capture most of the variance across participants were represented by the mRNA
156 and miRNA data. MOFA captures latent factors with common variance across the different
157 data sets, even though certain data types appear to be responsible of most of the captured
158 variance (Factors 1 to 7). This indicates that certain degree of heterogeneity exists across data
159 sets, making the integration more challenging.

160

161 Importantly, however, latent factors 15 and 18 were significantly associated (p-values < 0.1
162 adjusted by Benjamini-Hochberg) with C-peptide slopes (Fig. 3D) but not with age or baseline
163 C-peptide. The amount of variance captured by latent factors 15 and 18 is 2.62% and 1.84%,
164 respectively, indicating that the decline of C-peptide over time is not among the major sources
165 of variance across the participants, but is sufficiently strong to be captured by this method. The
166 differences in the strength of associations for latent factors 15 and 18 with the progression
167 groups and C-peptide slopes indicated that latent factor 15 captures a non-linear association
168 with the progression groups (Fig. 3C), and for latent factor 18 a linear association with the rate
169 of C-peptide decline (Fig. 3E). As the two factors correlate with C-peptide decline, they may
170 contain molecular signatures useful for explaining the differences in disease progression
171 between patients. Therefore, we continued a thorough scrutiny of these factors.

172

173 **Differential gene expression analysis**

174 Differentially expressed genes (DEGs) were identified between the different progression
175 groups, with batch and age groups used as covariates (Fig. 4). P-values were adjusted for
176 multiple testing using the Benjamini-Hochberg procedure and genes with an adjusted p-value
177 < 0.1 were reported as differentially expressed genes (DEG). Fig. 4 shows the volcano plots of
178 the different comparisons together with the genes belonging to latent factors 15 and 18. A total
179 of 339 DEGs were observed comparing the rapid decline group and the increasing group (Fig.
180 4A, Supplementary Table 7), 33 DEGs between the rapid and slow decline group (Fig. 4B,
181 Supplementary Table 8), and 1,205 DEGs between the slow decline group and the increasing
182 group (Fig. 4C, Supplementary Table 9). Additionally, differential gene expression was
183 performed for the C-peptide slopes, hence studying the linear change in gene expression with

184 respect to the rate of decline of C-peptide over time. Here we found 484 DEGs (Fig. 4D,
185 Supplementary Table 10).

186

187 More DEGs were observed when comparing the slow and increasing than between rapid and
188 increasing progression groups, with little overlap between the significant top-ranking genes.
189 Additionally, very few genes were differentially expressed between rapid and slow progression
190 groups. Altogether, these results might indicate that even though the rapid and slow progression
191 groups are more alike, the set of DEGs between these two groups and the increasing group are
192 not the same. To further confirm this, two additional differential expression analyses were
193 performed. One between the rapid-slow groups combined versus the increasing group and
194 another between the slow-increasing groups combined versus the rapid. The analyses showed
195 that the rapid-slow vs increasing comparison produced 1,804 DEGs (Supplementary Table 11),
196 while the slow-increasing vs rapid comparison produced 18 DEGs (Supplementary Table 12).
197 This indicates that the increasing progression group is much more dissimilar in its blood sample
198 expression profile towards the other two groups at the early stage of type 1 diabetes
199 manifestation. Therefore, the underlying biological processes involved in the developing
200 disease progression do not vary much at this early time of the disease manifestation among
201 people experiencing in the future different degrees of loss of β -cell function. At baseline they
202 vary significantly more when comparing people experiencing in the future loss of β -cell
203 function and those experiencing an improvement in β -cell function.

204

205 The similarity between DEGs found by the continuous change in C-peptide levels (n=484) and
206 DEGs between the rapid and increasing group (n=339) showed an overlap of 209 genes. The
207 rapid-slow vs increasing DEGs (n=1,804) had a bigger overlap with the continuous C-peptide
208 levels where 313 of the same DEGs were found. In all cases, the DEGs had the same sign of
209 their log₂ fold changes for both analyses. Therefore, most DEGs were observed for the
210 continuous change were also found when investigating progression groups. Nonetheless, we
211 believe that the linear association of blood gene expression at baseline with the C-peptide
212 slopes is more informative regarding the disease progression. We observe that change in β -cell
213 function follows a gradient, so by categorizing participants into groups, we lose the resolution
214 that the C-peptide slopes are providing.

215

216

217

218 **Annotation of latent factors**

219 To examine the biological pathways in the two most relevant latent factors, we used gene set
220 enrichment analysis. This analysis was divided into genes positively regulated in the rapid
221 decline group and genes negatively regulated in the rapid decline group. Fig. 5 displays the top
222 15 significant pathways in each of the two associated latent factors (p-values < 0.1 adjusted by
223 Benjamini-Hochberg). Negatively regulated genes in latent factor 15 are enriched in the
224 immune system and signaling by G protein-coupled receptors (GPCR) pathways. Of specific
225 interest are pathways associated with innate immunity, such as neutrophil degranulation, with
226 high expression of granule proteins (e.g. CTSS, MPO) pointing to the presence of activated or
227 degranulated neutrophils; and platelet activation, signaling and aggregation (the latter not
228 shown). Also, several pathways pointing to cytokine signaling and interleukin (e.g. IL-1 β)
229 signaling emerge. Regarding the role of GPCRs, several pathways are associated with latent
230 factor 15, such as signaling by GPCRs, downstream signaling and GPCR ligand binding.
231 Positively regulated genes in latent factor 15 are also enriched in immune system pathways,
232 with again an important contribution of the innate immune system, although here it seems that
233 the increase is mainly attributed to resting neutrophils, with for instance high expression of
234 LY96. Collectively these data show that GSEA pathways in innate immunity are mainly
235 associated with the activation status of the neutrophils, with shift in the balance of resting
236 neutrophils versus activated or degranulated neutrophils.

237

238 Negatively regulated genes in latent factor 18 are enriched in influenza infection pathways and
239 mRNA translation pathways, suggesting that viral mRNA replication and translation by host
240 cell machinery are major pathological features of this association. Positively regulated genes
241 in latent factor 18 did not show a particular pattern of enrichment.

242

243 Furthermore, the relation of the latent factors' genes to previous type 1 diabetes publications
244 was studied. Using the Open Targets database¹⁰, a total of 174 genes out of the 668 top genes
245 (three-fold higher than expected by chance) in the latent factor 15 have been previously
246 associated with type 1 diabetes (p-value 0.03). On the other hand, the overlap between the top
247 genes in the latent factor 18 and associated disease genes was not significant. These results
248 indicate that latent factor 15 is capturing a set of genes composed of known disease targets (e.g.
249 INSR, NDUFA4, CTSH) as well as potential new candidate genes already detectable in blood
250 in the early phase of type 1 diabetes.

251

252 **Interpretation of biological networks**

253 Biological networks were constructed for latent factors 15 and 18 separately, based on protein-
254 protein or protein-protein-miRNA interactions (Supplementary Fig. 3 and 4). As the two
255 network types yielded similar results, we selected the protein-protein-miRNA networks for the
256 focus of our interpretations. The latent factor 15 network revealed a diverse set of biological
257 functions (Supplementary Fig. 5), some of them overlapping with the pathways shown in Fig.
258 5. Immune system responses, signaling by different receptors and lipid metabolism are (widely)
259 represented in these clusters. Notably there is an enrichment of lipid metabolism pathways as
260 the lipidomic data also influenced latent factor 15. Some of the genes significantly associated
261 with the C-peptide slopes also appear in several of the clusters, which further validates the
262 biological processes captured by the latent factor. The latent factor 18 network was smaller and
263 had more loosely defined clusters (Supplementary Fig. 6). Nonetheless, it captured a similar
264 set of biological functions compared to the latent factor 15 network. Eukaryotic/viral mRNA
265 translation is the main difference between the two networks, which appear as the biggest and
266 more interconnected cluster of the latent factor 18 network. In this case, only one of the genes
267 in this network was significantly associated with the C-peptide slopes.

268

269 **Immunomics signature**

270 Analysis of immune cell populations identified based on standard markers and their association
271 with the C-peptide slopes revealed that Natural Killer (NK) cells were significantly associated
272 after p-value adjustment, (Fig. 6A), with higher levels of NK cells observed in people with
273 slow disease progression (Fig. 6B). Examination of the relationship between C-peptide slope
274 and NK cell frequency in individual progression groups indicated the strongest correlation was
275 observed among rapid progressors (Fig. 6C). Unsupervised analysis of the immunomics data
276 revealed distinct clusters among the progression groups. Fig. 7A shows a FlowSOM color-
277 density map of CD16 expression levels with node sizes representing the frequency of cells in
278 each cluster. Meta-cluster 19 (MC-19) was assigned as the primary NK (CD56loCD16+) subset
279 based on lineage marker expression and was significantly more abundant in the increasing
280 versus rapid progression groups (8.8 vs 6.3% p-value=0.013). Examination of markers of NK
281 cell activation and differentiation (KLRG1, TIGIT, CD38, and CD57) in the different
282 progression groups showed higher expression of KLRG1, in the increasing group but lower
283 levels of CD38.

284

285 Comparison of the genes associated with latent factor 15 and the leukocyte gene signature
286 matrix LM22¹¹ revealed an overlap of 52 genes and enrichment in immune cell-specific genes
287 (p-value 0.0001). The genes were representative of the following five main groups, T-cell
288 specific, macrophage M1 specific, monocyte specific, neutrophil specific, and eosinophil
289 specific (Supplementary Fig. 7). In this way, in addition to capturing immune-related pathways,
290 the latent factor 15 broadly represented immune cell-specific genes.

291

292 **Other ‘omics associations with disease progression**

293 The association of factors derived from running MOFA without the transcriptomics data, did
294 not yield any significant associations with the progression groups nor the C-peptide slopes. We
295 can draw two conclusions based on these results. Firstly, the variance of the miRNA and
296 lipidomics captured by latent factor 15 might indicate that these two omics data sets are only
297 informative of the disease progression in combination with transcriptomics. Interestingly, an
298 enrichment on lipid metabolism pathways was apparent among the genes associated with latent
299 factor 15, indicating that these pathways might have an effect of disease progression. This
300 shows how integration of both lipidomics and transcriptomics data aid in discovering
301 enrichment in specific pathways. Similarly, the incorporation of transcriptomics and miRNA
302 data allowed the discovery of enrichment in viral infection pathways.

303

304 Secondly, the main difference between transcriptomics and the other omics data was that it was
305 determined from whole blood. The remaining omics are collected either from serum, plasma,
306 or PBMCs. Even though some analytes may give similar results, samples obtained from the
307 same medium are more easily comparable. Therefore, the disparity between omics that we
308 observe in the latent factors and in the linear association of each omics data set might be caused
309 by the source medium. This could be considered both a drawback and an advantage. On one
310 hand, it is undesirable that this disparity exists because correlated analytes across omics cannot
311 be studied. This makes the data integration challenging as we cannot observe the joint effect
312 of multiple omics nor validate whether the analytes associated with the disease progression in
313 one data type can also be observed in a different one. On the other hand, the heterogeneity
314 across omics can be seen as complementary information. Each omics data set is capturing a
315 different source of variation, thus, providing additional information not captured by the other
316 omics types. Based on the current data, we cannot conclude whether the plasma and serum
317 omics were not associated with the disease progression due to the medium or the analytes
318 themselves.

319

320 Discussion

321 In this study, we identified two latent factors associated with β -cell decline. These factors were
322 predominantly influenced by transcriptomics, with secondary contributions of miRNA and
323 lipidomics, respectively. Latent factor 15 revealed an enrichment of immune system pathways,
324 the most significant being associated with neutrophil degranulation, cytokine signaling, and
325 immunoregulatory interactions between lymphoid and non-lymphoid cells. Moreover, there
326 were multiple pathways associated with GPCR signaling events. More detailed GSEA revealed
327 that disease progression (C-peptide slopes) was associated with an altered balance between
328 resting and activated/degranulated neutrophils. This is in keeping with previous studies that
329 showed a temporary decline in the number of circulating neutrophils in people with newly
330 diagnosed type 1 diabetes, compared to healthy controls, as well as high circulating levels of
331 neutrophil extracellular traps (NETs)¹²⁻¹⁴. The previous detection of neutrophils and NETs in
332 the pancreas of deceased subjects affected by type 1 diabetes, and a correlation between
333 circulating neutrophil numbers with C-peptide levels in pre-symptomatic subjects (non-
334 diabetic, at-risk) has implicated that activated neutrophils play a pathogenic role in type 1
335 diabetes¹⁵. Our data add significantly to this hypothesis since we show in longitudinal follow-
336 up that the neutrophil profile at diagnosis associates with rate of disease progression.

337
338 Of further interest, GSEA shows platelet activation to be a feature of latent factor 15, linking
339 our findings to the recent demonstration of a role for activated platelets in the formation of
340 platelet-neutrophil aggregates (PNAs), which are increased in the circulation of subjects during
341 the development of type 1 diabetes¹⁶. It is tempting to speculate that activation of GPCR
342 pathways (also associated with latent factor 15) may play a role in these events since it is a
343 response to a variety of stimuli (chemokines, cytokines, complement fragments) and can trigger
344 neutrophil degranulation¹⁷.

345
346 In contrast, latent factor 18, which was also associated with β -cell decline, is characterized by
347 features of enrichment of viral mRNA translation and subsequent translation by the host cell
348 machinery. There is a considerable body of literature that associates viral infections with early
349 events in type 1 diabetes development as well as peri-diagnosis events¹⁸. As a result, virus
350 infection has often been cited as an autoimmunity-triggering event as well as a disease-
351 precipitating event. Our findings in the context of the present study design are entirely
352 consistent with the latter hypothesis, which could be addressed in follow-up viromic studies

353 targeted to samples in which both the relevant viral mRNA translation signals and negative
354 slope of C-peptide decline are prominent.

355

356 Molecular and cellular signatures of adaptive immune responses were by and large not
357 observed to be associated with β -cell decline in our study, which might at first sight be
358 considered a surprise, given the strong credentials, at genomic, pathological and mechanistic
359 levels, for type 1 diabetes being the archetype of an organ-specific autoimmune disease.
360 However, it is entirely plausible that the detection of such associations is challenging in whole
361 blood or whole mononuclear cell analyses, both because the disease-relevant, β -cell antigen-
362 specific lymphocytes are rare, and even more importantly, because they may be sequestered at
363 inflammatory sites. Certainly, smaller scale studies focused on using appropriately sensitive
364 technologies have identified that the activation and differentiation state of circulating antigen-
365 specific cytotoxic T lymphocytes, for example, correlate with changes in β -cell function after
366 the diagnosis of type 1 diabetes⁸. Amongst the lymphocyte studies presented here, our
367 observation of a prominent NK cell signal related to rapid β -cell decline is of considerable
368 interest. NK cells have features of both innate and adaptive immune cells and play a key role
369 in anti-viral responses. Both pro-inflammatory and regulatory functions have been ascribed to
370 these cells, and functional subtypes can be partially differentiated by surface markers. Previous
371 studies examining the frequency of NK cells in individuals affected by type 1 diabetes have
372 consistently reported lower circulating levels of both proinflammatory and regulatory NK cells
373 when compared to aged-matched non-diabetic subjects^{19,20,10}, potentially reflecting homing to
374 inflammatory sites in the pancreas. Consistent with this, we observed lower circulating levels
375 of NK cells (both effector and regulatory subtypes according to surface markers) in the rapid
376 decline group. Of interest, reduced circulating NK cell levels are also associated with viral
377 infection, linking this observation to the viral signatures already described. Future functional
378 studies will be required to explore the pathological implications of these findings, since the
379 immune phenotyping performed here was limited to expression of CD38 (NK cell activation)
380 and KLRG1 (an inhibitory receptor associated with an exhausted phenotype).

381

382 Key strengths of the study include (i) the setting of a large, longitudinal natural history study
383 conducted across multiple European sites according to standardized clinical and laboratory
384 protocols; (ii) access through the INNODIA network to highly specialized, systems-based
385 technology platforms for parallel multi-parametric analysis; (iii) leverage of new tools in
386 integrated multi-omics factor analysis to discover signatures that are significantly associated

387 with rate of disease progression for the year following diagnosis. This period of the disease is
388 important since it represents the phase during which disease-modifying immunotherapies are
389 typically trialed for their effect on arresting β -cell decline. Factors identifiable at baseline that
390 associate with faster β -cell decline could be used to conduct shorter and smaller trials (e.g. by
391 enrollment of a rapid-decline group), an important step towards de-risking the investment
392 needed to bring disease modifying strategies into clinical use.

393

394 Finally, we want to highlight specificities of the present study that can be interpreted as
395 limitations or strengths. Our INNODIA natural history collection involves individuals from the
396 ages of one up to 45 years, and thus includes the whole lifespan of people living with type 1
397 diabetes. Although this may render interpretation of findings in this clearly heterogenic disease
398 more complex, it may add to the identification of common factors that drive disease,
399 irrespective of age. We used fasting C-peptide as read-out for β -cell function rather than
400 stimulated C-peptide and could demonstrate similar trends in decline of function using this
401 simple parameter. Although collecting throughout Europe, the population studied is almost
402 completely white Caucasian, thus limiting the generalizability of the findings to a global
403 population, where type 1 diabetes is becoming more prevalent in non-Caucasian people.
404 Confirmation of our observations will therefore be needed in more diverse cohorts. Our work
405 reports on a small cohort of just under 100 people with newly diagnosed disease. However, we
406 demonstrate that even in such small numbers, using deeply phenotyped individuals and
407 standardized operating procedures, application of systems biology techniques can lead to
408 significant associations. Here we believe that the collaboration between academics,
409 foundations, industry and people affected by type 1 diabetes and their families within
410 INNODIA was a unique driver. We created strict protocols for follow up where people could
411 be convinced to participate with support of materials created by the PAC (People with diabetes
412 Advisory Committee), we set up a highly standardized sample collection (e.g. even
413 standardizing the pipet tips for miRNA sample collection) and applied homogeneous
414 laboratory procedures. Importantly, we brought all data into a GDPR-conform centralized
415 database, allowing clean data collection and high quality data for input into the analysis.

416

417 In summary, the presented study addressed the drivers of disease heterogeneity in type 1
418 diabetes by leveraging opportunities presented by a highly integrated clinical network featuring
419 embedded research platforms with the capability to generate large systems-level datasets. One
420 of the two factors identified showed correlation to neutrophil degranulation, cytokine signaling,

421 lymphoid and non-lymphoid cell interactions and G-protein coupled receptor signaling events,
422 while the second signature, pathways related to translation and viral infection were inversely
423 associated with change in β -cell function. The derived latent factors were used to identify
424 specific signatures, which were further investigated for biomarker opportunities and
425 mechanistic pathways that correlate with β -cell decline. This shows how multi-omics analysis
426 can be used as an important foundation for the development and testing of disease-modifying
427 therapies in the future.

428 **Methods**

429 **Subjects with type 1 diabetes**

430 For this in depth analysis, we included the first 100 subjects with newly diagnosed (<6 weeks)
431 type 1 diabetes enrolled in the INNODIA natural history study. Using a consecutive
432 recruitment approach, subjects were included based on baseline omics data availability, an even
433 gender distribution and biosample availability, positivity for at least one diabetes-related
434 autoantibody (GADA, IA-2A, IAA, ZnT8A), and age between one to 45 years. Two subjects
435 were excluded due to incomplete ‘omics datasets and one following the detection of a MODY
436 gene mutation. The final analysis cohort comprised 49 male and 48 female study participants
437 (Table 1), the average age at diabetes diagnosis of 13.2 years (SD 8.5; two-ten years n=38; ten-
438 18 years n=41, 18-39 years n=18), mean disease duration of 3.9 weeks (SD 1.5) and at baseline
439 an average total daily insulin dose of 0.51 IU/kg (SD 0.27), HbA1c of 75 mmol/mol (SD 21.3),
440 fasting C-peptide level of 269 pmol/l (IQR 25.7), fasting glucose level of 7.78 mmol/l (IQR
441 2.8) and BMI SDS of 0.327 units (SD 1.1). Fasted C-peptide measurements were made at four
442 visits (Fig. 1A). To define the rate of C-peptide decline over time, we utilized linear mixed-
443 effect models to fit the log-transformed fasted C-peptide from day of diagnosis to 12 months
444 (Fig. 1B). The model was fitted using subject-level random effects and the rate of C-peptide
445 change over time. Mixed-effect models were fit using the lme4 R package²¹ with an
446 unstructured random effects variance-covariance matrix.

447

448 Some individuals did not complete all visits. A total of 69 individuals completed the four visits
449 (baseline, three, six, and 12 months), 21 individuals completed three visits, five individuals
450 completed two visits, and two individuals completed only the baseline visit (not necessarily
451 consecutive visits). At each visit HbA1c was measured and stimulated C-peptide response was
452 determined by mixed meal tolerance test (MMTT) from individuals of at least five years of
453 age. The islet autoantibodies GADA, IA-2A, IAA, and ZnT8A were quantified with the use of
454 specific radiobinding assays as described earlier²².

455

456 **Targeted serum proteomics by liquid chromatography-mass spectrometry (LC-MS/MS)**

457 Targeted proteomics of baseline serum samples from 91 individuals was performed using
458 liquid chromatography with selected reaction monitoring (SRM) mass spectrometry (LC-
459 MS/MS) to measure 195 peptides, representing 98 target proteins and retention time markers.
460 The list of peptides and proteins measured is provided in Supplementary Table 1.

461 The samples were reduced, digested, and alkylated, then spiked with isotopically labelled
462 synthetic analogues of the targets, as previously described²³. The batchwise analyses of the
463 samples were made with the periodic inclusion of three internal QC reference controls. Skyline
464 software²⁴ was used to both develop the acquisition method and perform primary processing
465 of the data, including peak integration and quality assessment. The un-normalized data was
466 subsequently exported and prepared for batch correction and normalization²⁵.

467

468 **Whole blood transcriptomics**

469 We performed transcriptome analysis on whole blood samples collected from 92 individuals
470 as well as four reference RNAs (two anonymous donors) used for normalization and assessing
471 batch effects across sample pools. Prior to RNA extraction, frozen whole blood PAXgene
472 samples were thawed at room temperature for 2 hours and subjected to RNA extraction using
473 PAXgene Blood miRNA Kit (PreAnalytix/QIAGEN, Cat# 763134). Total RNA, including
474 RNA longer than approximately 18 nucleotides, was purified according to the manufacturer's
475 protocol. Sample concentration was measured with a Nanodrop 2000 spectrophotometer and
476 Qubit Fluorometric Quantitation (Thermo Fisher Scientific).

477 The quality of the samples was ensured with the Experion Automated Electrophoresis System
478 (Bio Rad) and Agilent 2100 Bioanalyzer RNA Pico chip. Library preparation and sequencing
479 were carried out at the Finnish Functional Genomics Centre (FFGC). Before starting library
480 preparation, ERCC Spike-in Mix 1 (Invitrogen P/N 4456739) was added to 100 ng RNA
481 according to the kit's protocol. RNAseq libraries were prepared using the TruSeq stranded
482 mRNA HT kit and protocol # 15031047 (Illumina). The quality and quantity of the amplified
483 libraries were measured using Advanced Analytical Fragment Analyzer (Agilent) and Qubit
484 Fluorometric Quantitation, respectively. Pooled libraries were sequenced on an Illumina
485 NovaSeq 6000 instrument, using 2x50 bp paired-end sequencing. bcl2fastq2 Conversion
486 Software v2.20.0.422 was used to convert base call files into FASTQ files. The quality of the
487 raw sequencing reads was checked using the FastQC tool (v. 0.11.14). The sequencing data
488 preprocessing was carried out using R (v. 3.6.1) and the related Bioconductor module (v. 3.9).
489 The reads were aligned to UCSC hg38 human reference genome (downloaded from Illumina
490 iGenomes site https://support.illumina.com/sequencing/sequencing_software/igenome.html)
491 using Rsubread package (v. 1.34.7) and the same package was used to produce the read counts
492 for the RefSeq annotated genes. The gene-wise counts per million (CPM) were generated using
493 the edgeR package (v. 3.26.8).

494

495 **Circulating Small RNAs sequencing**

496 Expression levels of circulating Small RNAs were analyzed from baseline plasma samples of
497 n=91 subjects on the same sample set adopting two different Next Generation Sequencing
498 (NGS) approaches: (i) probes-based sequencing focused on miRNAs (targeted) through HTG
499 EdgeSeq miRNA whole transcriptome assay (Supplementary Table 2); (ii) Small RNAs
500 sequencing using the QiaSeq miRNA/small RNA library preparation kit (untargeted).
501 Plasma samples were shipped to HTG Molecular Inc. in order to be analyzed using rigorous
502 standard procedures. The HTG EdgeSeq miRNA whole transcriptome assay method is an RNA
503 extraction free approach that exploits quantitative nuclease protection assay chemistry using
504 sequence-specific nuclease-protection probes (NPPs). This was followed by an NGS step, in
505 order to allow semi-quantitative analysis of a panel of n=2102 targeted miRNAs (including
506 n=13 housekeeping, n=5 negative process controls, n=1 positive process control and n=2083
507 targeted miRNAs) from 15 μ L of plasma. HTG EdgeSeq Plasma Lysis buffer was added to 15
508 μ L of each plasma sample. Lysed samples were then transferred to a standard 96-well plate.
509 The NPPs were added to the lysed samples followed by the addition of S1 nuclease to digest
510 non-hybridized RNA. The nuclease digestion reaction was then stopped and each processed
511 sample was used as a template for PCR-based library preparation using specifically designed
512 primers (tags), which share common sequences complementary to 5'-end and 3'-end "wing"
513 sequences of the probes and common adapters required for cluster generation on Illumina NGS
514 platform. Libraries were prepared in accordance with HTG standard procedures, HTG EdgeSeq
515 PCR processing, and HTG EdgeSeq AMPure cleanup of Illumina Sequencing Libraries.
516 Libraries concentration was evaluated by the HTG EdgeSeq KAPA Library quantification kit,
517 and each library was normalized and pooled using the HTG EdgeSeq RUO library calculator.
518 Then, pooled libraries were denatured in 2 N NaOH and sequenced (final concentration 4 pM)
519 onto Illumina NextSeq550 platform (High Output kit v.2 cat. FC-404-2005). Data were
520 returned from the sequencer as demultiplexed FASTQ files. The resulting reads (on average
521 79% passing filter, corresponding to a total of 8.33×10^8 passing filter reads) were aligned
522 referring to miRbase v.20 using HTG Parser software. Raw reads were standardized into CPM
523 and filtered through the "limma Edge R" R Bioconductor package. For the QiaSeq Small RNA
524 sequencing, total RNA extraction was performed from 200 μ L of plasma through
525 Serum/Plasma Norgen kit (cat. 55000). Small RNA Libraries were prepared using the QiaSeq
526 miRNA library kit (cat. 331505) following the manufacturer's instructions. QiaSeq strategy
527 assigns Unique Molecular Index (UMI) during reverse transcription step to every mature
528 miRNA molecule, in order to enable unbiased and accurate Small RNAome-wide

529 quantification of mature miRNAs and additional small RNAs by NGS. Libraries quality control
530 (QC) was performed by quantifying their concentration through QUBIT 3.0 spectrofluorometer
531 (Qubit™ dsDNA HS Assay Kit, cat. Q32854) and assessing their quality using capillary
532 electrophoresis in Bioanalyzer 2100 (Agilent High Sensitivity DNA kit cat. 5067-4626). The
533 quality of libraries was evaluated considering electropherograms showing a peak comprised
534 between 175 and 185 bp as high quality. Following QC, all libraries were normalized until 2
535 nM and pooled, denatured in 0.2 N NaOH, and further sequenced (final concentration 175 pM)
536 using the Illumina NovaSeq 6000 platform (NovaSeq 6000 SP Reagent Kit (100 cycles)
537 cat. 20027464, NovaSeq XP 2-Lane Kit cat. 20021664, using the XP protocol applying 75x1
538 single reads). Data were returned from the sequencer as demultiplexed FASTQ files. Resulting
539 reads (on average 87% passing filter, corresponding to a total of 1.11×10^9 passing filter reads)
540 were mapped using QIAGEN Gene Globe data analysis center software, which adopts a
541 sequential alignment strategy to map to different databases (perfect match to miRbase v.21
542 mature, miRBase hairpin, noncoding RNA, mRNA and other RNA, and ultimately a second
543 mapping to miRBase mature, where up to two mismatches are tolerated) using bowtie (bowtie-
544 bio.sourceforge.net/index.shtml). At each step, only mapped sequences were passed to the next
545 step.

546

547 **Plasma Metabolomics (GCxGC-MS)**

548 83 metabolites, covering various amino acids, fatty acids and sugars etc. (Supplementary Table
549 3), was generated from baseline plasma samples as described before²⁶. In brief, 30 μ L blood
550 samples from 93 participants were spiked with 10 μ L of the internal standard mixture (d4-
551 succinic acid, d5-glutamic acid, d8-valine, and d33-heptadecanoic acid.; Sigma Aldrich).
552 Samples were vortex-mixed and incubated on ice for 30 min and centrifuged (10,000 rpm, 3
553 min, 4°C). Finally, 180 μ L of the filtered extracts were transferred to glass vials and evaporated
554 dry before derivatization.

555 The samples were derivatized using a previously described procedure²⁷, where reactive groups
556 are converted into trimethylsilyl derivates, which increases the volatility of the biomolecules.
557 The polar metabolites were then analyzed using a Pegasus 4D (LECO; Saint Joseph; USA)
558 system, which combines two-dimensional chromatographic separation with time-of-flight
559 (TOF) mass spectrometric detection. Identifications were assigned using the National Institute
560 of Standards (NIST) database and Steno Diabetes Center Copenhagen in-house libraries. After
561 data acquisition, the raw data were pre-processed into a peak table with ChromaTOF (LECO;

562 Saint Joseph; USA). Finally, data were post-processed in R by batch correction, truncation of
563 outliers, and imputation of missing values.

564

565 **Plasma Lipidomics (LC-MS)**

566 The lipidome (Supplementary Table 4) was measured from baseline plasma samples of 94
567 individuals and two unrelated control samples in four replicates following lipid extraction from
568 a 10 μ L plasma sample using a chloroform:methanol (2:2 V/V) lipid extraction method²⁸. Nine
569 stable isotope labelled and non-physiological lipid species were spiked as internal standards.
570 (Supplementary Table 5).

571 Samples were analysed in positive and negative ion modes of ultra-high-performance liquid-
572 chromatography mass-spectrometry (UHPLC-MS; Agilent Technologies; Santa Clara, CA,
573 USA) at Steno Diabetes Center Copenhagen as described previously²⁹⁻³⁰. After data
574 acquisition, the raw data were pre-processed into a peak table with MZmine 2. Finally, data
575 were post-processed in R by denoising as normalization to internal standards, batch correction,
576 truncation of outliers, and imputation of missing values.

577

578 Nine stable isotope labelled and non-physiological lipid species were spiked as internal
579 standards (Supplementary data table 2).

580

581

582 **PBMC (cryopreserved) multi-dimensional flow cytometry (Multi-FACS) immunomics**

583 The immunome was examined using cryopreserved peripheral blood mononuclear cells
584 (PBMCs; that were freshly isolated and cryopreserved in Cryostor) from 76 baseline blood
585 samples (Na-heparin) shipped in liquid nitrogen and processed at a single site by Multi-FACS.
586 36-marker Multi-FACS Cytex Aurora panel (plus one viability dye) was set up and a specific
587 flow gating strategy that allowed enumeration of 150 cell populations was applied
588 (Supplementary Table 6).

589 All flow data were also analyzed by OMIQ platform (www.omiq.ai), including unbiased
590 clustering (FlowSOM) and dimensionality reduction (UMAP). In FlowSOM analysis, the
591 median intensity of 36 markers were used to generate 100 nodes, which were further clusters
592 as meta-cluster (k=30, grey area on FlowSOM map under randomized speed and Euclidean
593 distancing metric). FlowSOM map provides a concise representation of the number of cell
594 types and visualization of the differential marker profiles by color-density.

595

596 Samples were processed in five batches (between 12 to 16 samples per batch, consisting of a
597 mixture of samples from each of the five INNODIA immune laboratories) together with two
598 unrelated control samples in each batch using one 2.5×10^6 PBMCs per sample except for two
599 samples where two aliquots were used due to poor percentage of recovery (range from 11.2 to
600 228.8%, average of 68.1%). Viability assessed by trypan blue ranged between 70 and 100%
601 (average 94%). An average of 1.51×10^6 (range 0.54 - 2.8×10^6) cells were used for staining. List
602 of Multi-FACS flow panel antibodies and reagents used is detailed in Supplementary data table
603 3. PBMCs were first stained using Live/dead blue for 15 min at room temperature, washed with
604 FACS buffer (PBS with 0.2% BSA and 2mM EDTA), and incubated with Fc receptor blocker
605 (TruStain FcX Fc; BioLegend) for 10 min at room temperature. Without wash, samples were
606 stained in a 37°C waterbath for 15 min using mastermix 1 (containing antibodies against
607 CXCR3, CD117, CD294/CRTH2, and CD161). Samples were further stained in waterbath for
608 15 min using mastermix 2 (containing antibodies against CXCR5, ICOS, CCR7, and CCR6),
609 followed by 30 min at room temperature using mastermix 3 (Supplementary data table 3).
610 Finally, samples were washed using FACS buffer, then fixed and resuspended in PBS
611 containing 1% paraformaldehyde (Alfa Aesar). Single colour controls were made using PBMC
612 for all colours except for CD294, CD117, CD161, and TCRgd where BD mouse or rat comp
613 beads were used instead due to low cell expression. Single colour controls were subjected to
614 the same buffer and fixed as the multi-colour stained samples. SpectroFlo QC beads were run
615 daily and single colour controls were acquired in the reference library, which was subsequently
616 used for live unmixing during sample acquisition on a Cytex Aurora cytometer. Flow data were
617 analysed using FlowJo software (an example of the gating strategy is shown in Supplementary
618 Fig. 8 and checked by an independent reviewer).

619

620 **Multi-omics data pre-processing, integration, and analysis**

621 For transcriptomics and miRNA data, the DESeq 2 package³¹ was used to normalize the counts
622 using the variance stabilizing transformation (VST). The transcriptomics and miRNA data sets
623 were filtered for low counts (features with less than 10 counts in total or features with zero
624 counts in more than 90% of the samples). The proteomics, metabolomics, lipidomics and
625 immunomics data sets were log₂ transformed. The transcriptomics, proteomics, and
626 immunomics data were corrected for batch effects associated with dataset-specific factors
627 (sequencing at different days or different handling of the samples) using the limma package in
628 R. Finally, all the data sets were corrected for age. The age was log-transformed to account for
629 the growth effect in children (one year difference in adults is not equivalent to one year

630 difference in children). This was necessary due to the high degree of age heterogeneity present
631 in the cohort and the association of the fasted C-Peptide slopes with age.

632

633 The omics data sets were integrated using the general framework in the Multi-Omics Factor
634 Analysis (MOFA) package from 2018³². MOFA performs a dimensionality reduction of the
635 omics data into a lower-dimensional latent space (Fig. 1C). The latent factors generated by
636 MOFA capture sources of global variability across the different omics data sets. Each factor
637 has an underlying weight for every feature, which can be used to annotate the factors in terms
638 of omics features, yielding a specific molecular signature for each factor. MOFA was run with
639 default parameters and 20 latent factors. The model was initialized with different random seeds
640 yielding similar results, generally only altering the number assigned to each factor associated
641 with the C-peptide slopes.

642

643 Latent factors were associated with the C-peptide slopes using the Spearman correlation. Other
644 covariates were also analyzed such as age at baseline and C-peptide at baseline. The association
645 of the latent factors with the progression groups was calculated using the Kruskal-Wallis test
646 and each group was compared using a Mann-Whitney U test.

647

648 Gene Set Enrichment Analysis (GSEA)³³ was performed in order to better characterize the
649 genes with the largest weight in the latent factors. This analysis was performed using the
650 MOFA GSEA function that utilizes a modified version of the principal component gene set
651 enrichment scheme (PCGSE)³⁴. The Reactome database was the gene set annotation used for
652 this analysis. The GSEA was performed separately for genes with a positive and negative
653 weight in each latent factor. This was done to avoid combining genes that are upregulated
654 (positive weight) and downregulated (negative weight) in the latent factor, as these two groups
655 of genes might be involved in different biological pathways. The top 15 significant pathways
656 for each latent factor (positive and negative weights) were selected and grouped by biological
657 pathway.

658

659 Differential gene expression of individual genes was determined with the DESeq 2 package³¹.
660 Models included covariates for the batch variable (different runs) and the age of the people.
661 Age was encoded as a categorical variable defined in three groups, less than ten years, between
662 ten and 18 years, and more than 18 years. Differential gene expression was assessed between
663 rapid and slow groups, rapid and increasing and slow and increasing, respectively.

664

665 For biological network analysis, two types of interaction networks were constructed. One was
666 compiled from the STRING³⁵ database to study protein-protein interactions or associations
667 only. Another network was compiled from the mirTarBase³⁶ (miRNA-gene), which was
668 combined with the STRING network to study protein-protein-miRNA interactions or
669 associations. The STRING database was filtered for high-confidence interactions (combined
670 score above 0.7) and miRNA-gene interactions had to be reported by at least two publications
671 and two non-high-throughput methods. We decided to focus on genes and miRNAs only
672 because those were the omics types with higher weights in the most relevant latent factors.

673

674 The association between the biological networks and the latent factors was performed using
675 the PCSF graph optimization approach. This method allows us to interpret the biological
676 landscape of the interaction network based on the importance/weight of each gene/miRNA in
677 the latent factor. The output is a subnetwork that captures interactions between the
678 genes/miRNAs with a higher importance in the latent factor. In order to select a subset of genes
679 and miRNAs to construct the network, only genes with normalized absolute weights three-fold
680 higher than expected by chance and miRNAs with normalized absolute weights two-fold higher
681 than expected by chance were selected. Grid-search was performed to select the best
682 parameters based on the network that had a high number of genes/miRNA from the latent factor
683 while keeping number of genes/miRNA not observed in the latent factor low
684 ($\mu=0.005$, $\omega=1$, $\beta=5000$). The final network was constructed using 20 runs with noise to edge
685 costs ($r=0.1$) that were combined and clustered using the edge-betweenness algorithm. Gene
686 set enrichment analysis was performed on the clusters obtained from the network.

687 **Figures**

688

689 Fig. 1. Cohort data and analysis overview. A) The cohort consists of 97 people with newly
690 diagnosed type 1 diabetes. Multi-omics data were collected at baseline (within six weeks after
691 diagnosis of type 1 diabetes) and clinical data were collected at baseline and at three, six, and
692 12 months. B) Participants were divided into three groups based on their change of insulin
693 secretion levels (fasted C-peptide measurements) from baseline to 12 months. C) Multi-
694 Omics Factor Analysis was performed to obtain an integrated signature across omics data
695 types followed by differential expression analysis for each omics data type independently.
696

697 Fig. 2: Fasted C-peptide trends over time. Time is represented as days from type 1 diabetes
698 diagnosis. A) C-peptide values over time for all participants. B) C-peptide values over time
699 divided into progression groups (progression terciles). C) C-peptide values over time for
700 different age groups. D) C-peptide values over time for males and females, respectively.
701

702 Fig. 3. A) Overview of the multi-omics data sets, describing the number of features per data
703 set and the level of missing data (white). B) Latent factors obtained from MOFA, the color
704 scale represents the variance captured by each of the latent factors indicating the level of
705 integration of the data types for each factor. C) Association of latent factors 15 and 18 values
706 with the different progression groups. D) Spearman correlation of latent factors with the
707 fasted C-peptide slopes, baseline fasted C-Peptide, Age (log scale), and BMI-SDS. P-values
708 were adjusted by Benjamini-Hochberg. E) Spearman correlation of fasted C-peptide slopes
709 against latent factor values (15 and 18).
710

711 Fig. 4. Volcano plot of differential gene expression between progression groups. The color
712 indicates gene membership to either or both associated latent factors. A) DGE between rapid
713 and increasing progression groups (339 genes are differentially expressed). B) DGE between
714 rapid and slow progression groups (33 genes are differentially expressed). C) DGE between
715 slow and increasing progression groups (1,206 genes are differentially expressed). D) DGE
716 for C-peptide slopes (484 genes are differentially expressed).
717

718 Fig. 5. GSEA was performed separately for genes with positive weights in the latent factor
719 (upregulated in rapid decline) and genes with negative weights in the latent factor

720 (downregulated in rapid decline). Pathways are coloured depending on the main pathway
721 they belong to, according to Reactome. P-values were adjusted using Benjamini-Hochberg.
722
723 Fig. 6. Immunomics association with C-peptide slopes. A) Benjamini-Hochberg adjusted p-
724 value for the linear association of immune cells abundances with the C-peptide slopes
725 (correcting for batch effects and age groups). B) NK cell levels for the different progression
726 groups (as a frequency of total live mononuclear cells). C) Spearman correlation of fasted C-
727 peptide slopes versus NK cell frequency for the different progression groups. D) FlowSOM
728 unbiased cluster analysis on live CD45⁺ PBMCs of Donor D07 (Increasing, left) and K40
729 (Rapid, right) highlighting Metacluster-19 (red-circle) as a primary NK (CD56^{lo}CD16⁺)
730 subset. Colour-density and circle size were overlaid as CD16 intensity and number of events,
731 respectively. E) Bar charts representing NK marker expression on the selected clusters in
732 each progression group.
733

734 **Tables**

735 Table 1. Clinical and demographic data for the type 1 diabetes cohort across progression
736 groups. P-values were calculated using the Kruskal-Wallis test.

737

	Rapid (N=33)	Slow (N=32)	Increasing (N=32)	P-value	Overall (N=97)
Sex					
Female	20 (60.6%)	15 (46.9%)	13 (40.6%)	0.256	48 (49.5%)
Male	13 (39.4%)	17 (53.1%)	19 (59.4%)		49 (50.5%)
Age (years)					
Mean (SD)	11.2 (9.17)	12.9 (7.94)	15.6 (7.94)	0.0125	13.2 (8.49)
Median (IQR) [Min, Max]	9.60 (8.97) [2.01, 38.1]	10.9 (8.68) [2.08, 36.4]	13.4 (5.23) [7.57, 38.8]		11.8 (8.08) [2.01, 38.8]
Age Intervals					
<10	17 (51.5%)	14 (43.8%)	7 (21.9%)	0.16	38 (39.2%)
>10-18	11 (33.3%)	12 (37.5%)	18 (56.3%)		41 (42.3%)
>18	5 (15.2%)	6 (18.8%)	7 (21.9%)		18 (18.6%)
Disease duration (weeks)					
Mean (SD)	3.88 (1.52)	3.80 (1.66)	4.10 (1.43)	0.843	3.93 (1.53)
Median (IQR) [Min, Max]	4.40 (2.50) [0.900, 6.40]	4.35 (2.35) [0.900, 6.30]	4.40 (1.70) [0.700, 6.10]		4.40 (2.20) [0.700, 6.40]
BMI SDS					
Mean (SD)	0.0719 (1.05)	0.397 (1.02)	0.511 (1.20)	0.153	0.327 (1.10)
Median (IQR) [Min, Max]	0.0800 (1.19) [- 2.32, 2.47]	0.350 (1.60) [- 1.56, 2.72]	0.610 (1.35) [- 2.00, 2.28]		0.260 (1.57) [- 2.32, 2.72]
Missing	1 (3.0%)	0 (0%)	0 (0%)		1 (1.0%)
Glucose (mmol/l)					
Mean (SD)	8.43 (5.57)	8.42 (5.33)	6.45 (2.20)	0.22	7.78 (4.69)
Median (IQR) [Min, Max]	7.10 (2.70) [3.70, 31.5]	6.70 (2.93) [3.60, 26.9]	6.10 (2.25) [3.70, 13.8]		6.40 (2.80) [3.60, 31.5]
Insulin dose (IU/kg)					
Mean (SD)	0.544 (0.279)	0.498 (0.247)	0.498 (0.281)	0.819	0.514 (0.268)
Median (IQR) [Min, Max]	0.545 (0.291) [0.136, 1.46]	0.468 (0.243) [0.108, 1.20]	0.500 (0.429) [0.0359, 1.01]		0.500 (0.354) [0.0359, 1.46]
Missing	1 (3.0%)	1 (3.1%)	0 (0%)		2 (2.1%)

HbA1c (mmol/mol)					
Mean (SD)	72.2 (21.4)	74.9 (24.2)	78.0 (18.1)	0.699	75.0 (21.3)
Median (IQR) [Min, Max]	75.0 (27.3) [8.70, 119]	73.0 (29.5) [13.4, 130]	80.3 (22.4) [50.0, 130]		77.5 (25.7) [8.70, 130]
Missing	1 (3.0%)	1 (3.1%)	1 (3.1%)		3 (3.1%)
Baseline C-peptide (pmol/l)					
Mean (SD)	235 (183)	248 (185)	324 (263)	0.296	269 (215)
Median (IQR) [Min, Max]	165 (228) [26.1, 809]	240 (204) [15.0, 986]	264 (268) [25.8, 1290]		224 (239) [15.0, 1290]
12 Month C-peptide (pmol/l)					
Mean (SD)	64.7 (59.8)	194 (137)	432 (206)	<0.001	234 (214)
Median (IQR) [Min, Max]	41.6 (83.5) [11.7, 222]	161 (136) [45.8, 635]	433 (246) [96.7, 896]		163 (283) [11.7, 896]
Missing	7 (21.2%)	9 (28.1%)	5 (15.6%)		21 (21.6%)
GADA					
Negative	5 (15.2%)	11 (34.4%)	8 (25.0%)	0.199	24 (24.7%)
Positive	28 (84.8%)	21 (65.6%)	24 (75.0%)		73 (75.3%)
IA-2A					
Negative	9 (27.3%)	12 (37.5%)	9 (28.1%)	0.615	30 (30.9%)
Positive	24 (72.7%)	20 (62.5%)	23 (71.9%)		67 (69.1%)
IAA					
Negative	4 (12.1%)	10 (31.3%)	10 (31.3%)	0.118	24 (24.7%)
Positive	29 (87.9%)	22 (68.8%)	22 (68.8%)		73 (75.3%)
ZnT8A					
Negative	12 (36.4%)	10 (31.3%)	14 (43.8%)	0.582	36 (37.1%)
Positive	21 (63.6%)	22 (68.8%)	18 (56.3%)		61 (62.9%)
Detectable autoantibodies					
Mean (SD)	3.09 (0.947)	2.66 (0.902)	2.72 (0.851)	0.0961	2.82 (0.913)
Median (IQR) [Min, Max]	3.00 (2.00) [1.00, 4.00]	3.00 (1.00) [1.00, 4.00]	3.00 (1.00) [1.00, 4.00]		3.00 (2.00) [1.00, 4.00]

738
739
740

741 **Other**

742

743 *Data Availability*

744 The data generated and analysed is person-sensitive as it can be used to identify people based
745 on their sequence variation and can be accessed in secure environments only. Access to data
746 can be provided by application to the INNODIA Data Access Committee by emailing
747 Professor Lut Overbergh (lutgart.overbergh@kuleuven.be). Processed results of GSEA
748 analysis is available as supplementary material.

749

750 *Code Availability*

751 The primary software used for this work was Multi-Omics Factor Analysis (MOFA). The
752 code for this software is available at <https://github.com/bioFAM/MOFA2>.

753

754 *Ethics Declaration*

755 The INNODIA study protocol was approved by the London – City & East Research Ethics
756 Committee on 28 October 2016 (REC 16/LO/1750) IRAS Project ID 210497. Subsequently,
757 after translation of the participants’ documentation, approval was obtained from local Ethic
758 authorities throughout the entire INNODIA clinical network.

759

760 *Acknowledgments*

761 We gratefully acknowledge all participants of the INNODIA natural history study which built
762 the basis for the work presented here. Without the outstanding engagement of the people with
763 type diabetes and their relatives and friends participating in the INNODIA clinical efforts,
764 this data and new knowledge gain would not have been possible. The authors also want to
765 thank all the clinical personnel for their dedication in the participant recruitment,
766 characterization, sample collection and preparation that was at the basis of this analysis. We

767 dedicate this work to the late Professor David Dunger, who inspired INNODIA and lay at the
768 basis of this analysis.

769

770 *Author contributions*

771 JJAA, CB, SFAB, DBD, PJC, MK, AMS, GS, LLE, FD, TT, RL, LO, CM, MP, SB designed
772 the study. JJAA, CB, KB, RM, NA-S, IM, CL-Q, TS, JHL, GEG, MKH, TS, OR, PJC, JHL,
773 GS, TT collected data and performed analyses. JJAA, CHJ, TT, LO, CM, MP, SB wrote the
774 manuscript. All authors revised the manuscript for crucial content and approved the final
775 version. All authors had full access to all the data and had final responsibility for the decision
776 to submit for publication. A complete list of the partners of the INNODIA consortium is
777 included separately.

778

779 *Declaration of interests*

780 CM serves or has served on the advisory panel for Novo Nordisk, Sanofi, Merck Sharp and
781 Dohme Ltd., Eli Lilly and Company, Novartis, AstraZeneca, Boehringer Ingelheim, Roche,
782 Medtronic, ActoBio Therapeutics, Pfizer, Imcyse, Insulet, Zealand Pharma, Avotres,
783 Mannkind, Sandoz and Vertex. Financial compensation for these activities has been received
784 by KU Leuven; KU Leuven has received research support for CM from Medtronic, Imcyse,
785 Novo Nordisk, Sanofi and ActoBio Therapeutics; CM serves or has served on the speakers
786 bureau for Novo Nordisk, Sanofi, Eli Lilly and Company, Boehringer Ingelheim, Astra
787 Zeneca and Novartis. Financial compensation for these activities has been received by KU
788 Leuven.

789 S.Br. reports having received funding from INNODIA (grant agreement No 115797), having
790 ownerships in Intomics A/S, Hoba Therapeutics Aps, Novo Nordisk A/S, Lundbeck A/S,
791 ALK A/S and managing board memberships in Proscion A/S and Intomics A/S.

792 MK reports ownership and managing board membership in Vactech Oy.

793 CLQ reports serving on advisory boards of Fondation Alzheimer's and Institute Pasteur, and
794 an affiliation with Novo Nordisk.

795 LO reports having received funding from INNODIA and INNODIA Harvest.

796

797 *Supplementary information*

798 Supplementary information is available for this paper as an appendix.

799

800 References

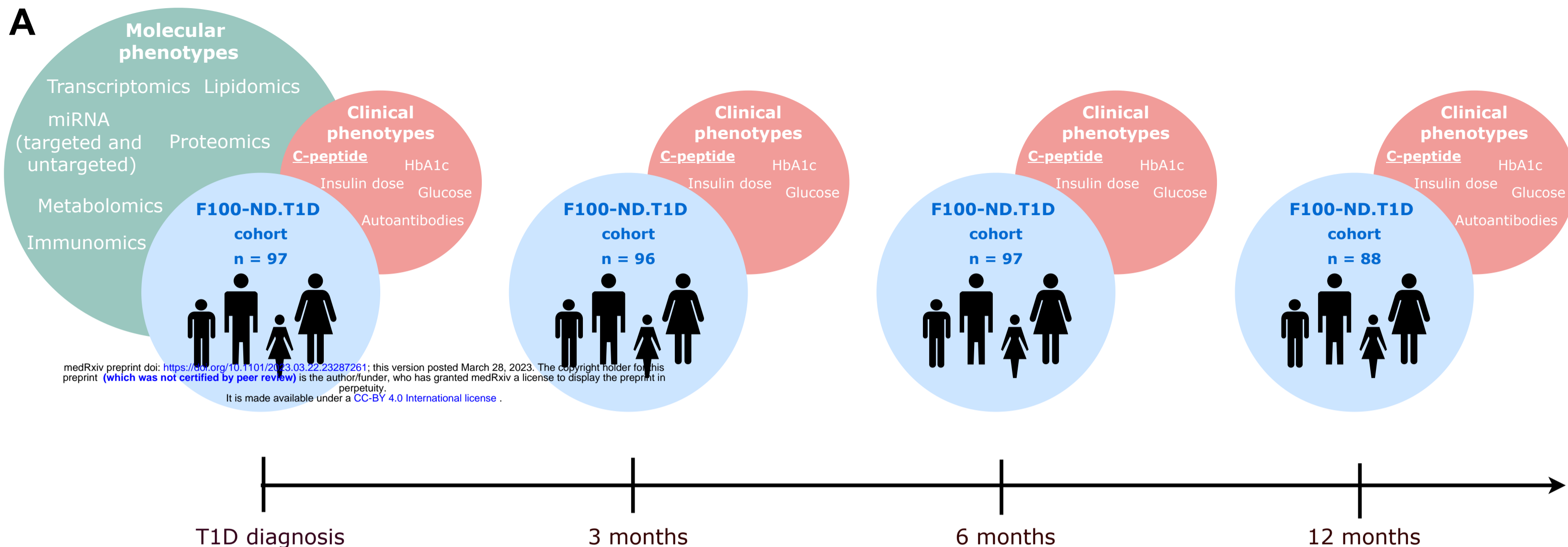
- 801 1. Warshauer, J.T., Bluestone, J.A., Anderson, M.S., 2020. New frontiers in the
802 treatment of Type 1 diabetes. *Cell Metab.* 31, 46–61.
803 <https://doi.org/10.1016/j.cmet.2019.11.017>
- 804 2. IDF, D.A., 2021. International Diabetes Federation. (No. 10th edn). Brussels,
805 Belgium.
- 806 3. Miller, R.G., Orchard, T.J., 2020. Understanding Metabolic Memory: A Tale of Two
807 Studies. *Diabetes* 69, 291–299. <https://doi.org/10.2337/db19-0514>
- 808 4. Rawshani, Araz, Sattar, N., Franzén, S., Rawshani, Aidin, Hattersley, A.T., Svensson,
809 A.-M., Eliasson, B., Gudbjörnsdottir, S., 2018. Excess mortality and cardiovascular
810 disease in young adults with type 1 diabetes in relation to age at onset: a nationwide,
811 register-based cohort study. *Lancet Lond. Engl.* 392, 477–486.
812 [https://doi.org/10.1016/S0140-6736\(18\)31506-X](https://doi.org/10.1016/S0140-6736(18)31506-X)
- 813 5. Tatovic D, Dayan CM. Replacing insulin with immunotherapy: Time for a paradigm
814 change in Type 1 diabetes. *Diabet Med.* 2021 Dec;38(12):e14696. doi:
815 10.1111/dme.14696.
- 816 6. Battaglia, M., Ahmed, S., Anderson, M.S., Atkinson, M.A., Becker, D., Bingley, P.J.,
817 Bosi, E., Brusko, T.M., DiMeglio, L.A., Evans-Molina, C., Gitelman, S.E.,
818 Greenbaum, C.J., Gottlieb, P.A., Herold, K.C., Hessner, M.J., Knip, M., Jacobsen, L.,
819 Krischer, J.P., Long, S.A., Lundgren, M., McKinney, E.F., Morgan, N.G., Oram,
820 R.A., Pastinen, T., Peters, M.C., Petrelli, A., Qian, X., Redondo, M.J., Roep, B.O.,
821 Schatz, D., Skibinski, D., Peakman, M., 2020. Introducing the Endotype Concept to
822 Address the Challenge of Disease Heterogeneity in Type 1 Diabetes. *Diabetes Care*
823 43, 5–12. <https://doi.org/10.2337/dc19-0880>
- 824 7. Speake, C., Skinner, S.O., Berel, D., Whalen, E., Dufort, M.J., Young, W.C.,
825 Odegard, J.M., Pesenacker, A.M., Gorus, F.K., James, E.A., Levings, M.K., Linsley,
826 P.S., Akirav, E.M., Pugliese, A., Hessner, M.J., Nepom, G.T., Gottardo, R., Long,
827 S.A., 2019. A composite immune signature parallels disease progression across T1D
828 subjects. *JCI Insight* 4. <https://doi.org/10.1172/jci.insight.126917>
- 829 8. Yeo, L., Woodwyk, A., Sood, S., Lorenc, A., Eichmann, M., Pujol-Autonell, I.,
830 Melchiotti, R., Skowera, A., Fidanis, E., Dolton, G.M., Tungatt, K., Sewell, A.K.,
831 Heck, S., Saxena, A., Beam, C.A., Peakman, M., 2018. Autoreactive T effector
832 memory differentiation mirrors β cell function in type 1 diabetes. *J. Clin. Invest.* 128,
833 3460–3474. <https://doi.org/10.1172/JCI120555>
- 834 9. Dunger DB, Bruggraber SFA, Mander AP, Marcovecchio ML, Tree T, Chmura PJ,
835 Knip M, Schulte AM, Mathieu C; INNODIA consortium. INNODIA Master Protocol
836 for the evaluation of investigational medicinal products in children, adolescents and
837 adults with newly diagnosed type 1 diabetes. *Trials.* 2022 May 18;23(1):414. doi:
838 10.1186/s13063-022-06259-z.
- 839 10. Oras A, Peet A, Giese T, Tillmann V, Uibo R. A study of 51 subtypes of peripheral
840 blood immune cells in newly diagnosed young type 1 diabetes patients. *Clin Exp*
841 *Immunol.* 2019 Oct;198(1):57-70. doi: 10.1111/cei.13332.
- 842 11. Chen, B., Khodadoust, M.S., Liu, C.L., Newman, A.M., Alizadeh, A.A., 2018.
843 Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol. Biol.*
844 Clifton NJ 1711, 243–259. https://doi.org/10.1007/978-1-4939-7493-1_12
- 845 12. Harsunen MH, Puff R, D'Orlando O, Giannopoulou E, Lachmann L, Beyerlein A, von
846 Meyer A, Ziegler AG. Reduced blood leukocyte and neutrophil numbers in the
847 pathogenesis of type 1 diabetes. *Horm Metab Res.* 2013 Jun;45(6):467-70. doi:
848 10.1055/s-0032-1331226.

- 849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
13. Valle A, Giamporcaro GM, Scavini M, Stabilini A, Grogan P, Bianconi E, Sebastiani G, Masini M, Maugeri N, Porretti L, Bonfanti R, Meschi F, De Pellegrin M, Lesma A, Rossini S, Piemonti L, Marchetti P, Dotta F, Bosi E, Battaglia M. Reduction of circulating neutrophils precedes and accompanies type 1 diabetes. *Diabetes*. 2013 Jun;62(6):2072-7. doi: 10.2337/db12-1345.
 14. Wang Y, Xiao Y, Zhong L, Ye D, Zhang J, Tu Y, Bornstein SR, Zhou Z, Lam KS, Xu A. Increased neutrophil elastase and proteinase 3 and augmented NETosis are closely associated with β -cell autoimmunity in patients with type 1 diabetes. *Diabetes*. 2014 Dec;63(12):4239-48. doi: 10.2337/db14-0480.
 15. Vecchio F, Lo Buono N, Stabilini A, Nigi L, Dufort MJ, Geyer S, Rancoita PM, Cugnata F, Mandelli A, Valle A, Leete P, Mancarella F, Linsley PS, Krogvold L, Herold KC, Elding Larsson H, Richardson SJ, Morgan NG, Dahl-Jørgensen K, Sebastiani G, Dotta F, Bosi E; DRI_Biorepository Group; Type 1 Diabetes TrialNet Study Group, Battaglia M. Abnormal neutrophil signature in the blood and pancreas of presymptomatic and symptomatic type 1 diabetes. *JCI Insight*. 2018 Sep 20;3(18):e122146. doi: 10.1172/jci.insight.122146.
 16. Popp SK, Vecchio F, Brown DJ, Fukuda R, Suzuki Y, Takeda Y, Wakamatsu R, Sarma MA, Garrett J, Giovenzana A, Bosi E, Lafferty AR, Brown KJ, Gardiner EE, Coupland LA, Thomas HE, Chong BH, Parish CR, Battaglia M, Petrelli A, Simeonovic CJ. Circulating platelet-neutrophil aggregates characterize the development of type 1 diabetes in humans and NOD mice. *JCI Insight*. 2022 Jan 25;7(2):e153993. doi: 10.1172/jci.insight.153993.
 17. Othman A, Sekheri M, Filep JG. Roles of neutrophil granule proteins in orchestrating inflammation and immunity. *FEBS J*. 2021 Mar 8. doi: 10.1111/febs.15803.
 18. Vehik K, Lynch KF, Wong MC, Tian X, Ross MC, Gibbs RA, Ajami NJ, Petrosino JF, Rewers M, Toppari J, Ziegler AG, She JX, Lernmark A, Akolkar B, Hagopian WA, Schatz DA, Krischer JP, Hyöty H, Lloyd RE; TEDDY Study Group. Prospective virome analyses in young children at increased genetic risk for type 1 diabetes. *Nat Med*. 2019 Dec;25(12):1865-1872. doi: 10.1038/s41591-019-0667-0.
 19. Wilson RG, Anderson J, Shenton BK, White MD, Taylor RM, Proud G. Natural killer cells in insulin dependent diabetes mellitus. *Br Med J (Clin Res Ed)*. 1986 Jul 26;293(6541):244. doi: 10.1136/bmj.293.6541.244.
 20. Fitas AL, Martins C, Borrego LM, Lopes L, Jörns A, Lenzen S, Limbert C. Immune cell and cytokine patterns in children with type 1 diabetes mellitus undergoing a remission phase: A longitudinal study. *Pediatr Diabetes*. 2018 Aug;19(5):963-971. doi: 10.1111/vedi.12671.
 21. Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
 22. Knip, M., Virtanen, S.M., Seppä, K., Ilonen, J., Savilahti, E., Vaarala, O., Reunanen, A., Teramo, K., Hämäläinen, A.-M., Paronen, J., Dosch, H.-M., Hakulinen, T., Akerblom, H.K., Finnish TRIGR Study Group, 2010. Dietary intervention in infancy and later signs of beta-cell autoimmunity. *N. Engl. J. Med.* 363, 1900–1908. <https://doi.org/10.1056/NEJMoa1004809>
 23. Bhosale, S.D., Moulder, R., Kouvonon, P., Lahesmaa, R., Goodlett, D.R., 2017. Mass Spectrometry-Based Serum Proteomics for Biomarker Discovery and Validation. *Methods Mol. Biol.* Clifton NJ 1619, 451–466. https://doi.org/10.1007/978-1-4939-7057-5_31
 24. MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., MacCoss, M.J., 2010. Skyline: an open source

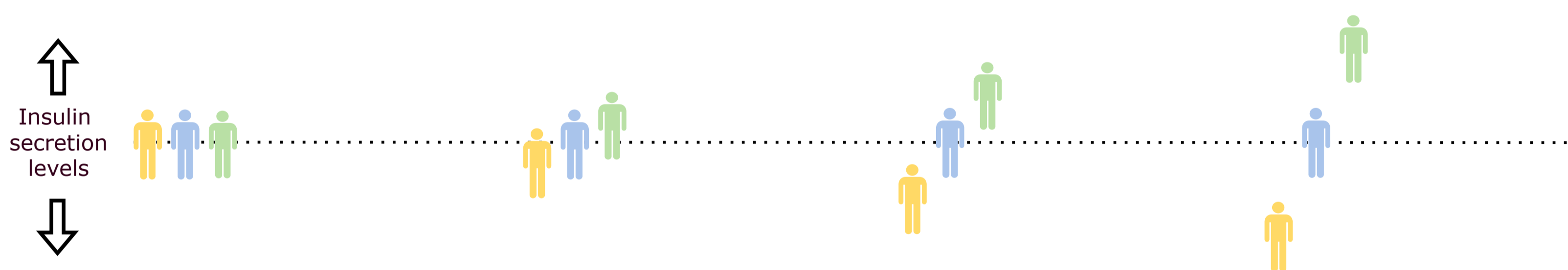
- 898 document editor for creating and analyzing targeted proteomics experiments.
899 Bioinforma. Oxf. Engl. 26, 966–968. <https://doi.org/10.1093/bioinformatics/btq054>
- 900 25. Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects
901 Models Using lme4. *J. Stat. Softw.* 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- 902 26. Orešič M, Anderson G, Mattila I, Manoucheri M, Soininen H, Hyötyläinen T,
903 Basignani C. Targeted Serum Metabolite Profiling Identifies Metabolic Signatures in
904 Patients with Alzheimer's Disease, Normal Pressure Hydrocephalus and Brain Tumor.
905 *Front Neurosci.* 2018 Jan 9;11:747. doi: 10.3389/fnins.2017.00747. PMID: 29375291;
906 PMID: PMC5767271.
- 907 27. Gika, H.G., Theodoridis, G.A., Plumb, R.S., Wilson, I.D., 2014. Current practice of
908 liquid chromatography–mass spectrometry in metabolomics and metabonomics. *J.*
909 *Pharm. Biomed. Anal., Review Papers on Pharmaceutical and Biomedical Analysis*
910 2013 87, 12–25. <https://doi.org/10.1016/j.jpba.2013.06.032>
- 911 28. Folch, J., Lees, M., Sloane Stanley, G.H., 1957. A simple method for the isolation and
912 purification of total lipides from animal tissues. *J. Biol. Chem.* 226, 497–509.
- 913 29. O’Gorman, A., Suvitaival, T., Ahonen, L., Cannon, M., Zammit, S., Lewis, G.,
914 Roche, H.M., Mattila, I., Hyotylainen, T., Oresic, M., Brennan, L., Cotter, D.R., 2017.
915 Identification of a plasma signature of psychotic disorder in children and adolescents
916 from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort. *Transl.*
917 *Psychiatry* 7, e1240. <https://doi.org/10.1038/tp.2017.211>
- 918 30. Tofte, N., Suvitaival, T., Ahonen, L., Winther, S.A., Theilade, S., Frimodt-Møller, M.,
919 Ahluwalia, T.S., Rossing, P., 2019. Lipidomic analysis reveals sphingomyelin and
920 phosphatidylcholine species associated with renal impairment and all-cause mortality
921 in type 1 diabetes. *Sci. Rep.* 9, 16398. <https://doi.org/10.1038/s41598-019-52916-w>
- 922 31. Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and
923 dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
924 <https://doi.org/10.1186/s13059-014-0550-8>
- 925 32. Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J.C., Buettner,
926 F., Huber, W., Stegle, O., 2018. Multi-Omics Factor Analysis—a framework for
927 unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* 14:e8124.
928 <https://doi.org/10.15252/msb.20178124>
- 929 33. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette,
930 M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005.
931 Gene set enrichment analysis: A knowledge-based approach for interpreting genome-
932 wide expression profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550.
933 <https://doi.org/10.1073/pnas.0506580102>
- 934 34. Frost, H.R., Li, Z., Moore, J.H., 2015. Principal component gene set enrichment
935 (PCGSE). *BioData Min.* 8, 25. <https://doi.org/10.1186/s13040-015-0059-z>
- 936 35. Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S.,
937 Doncheva, N.T., Legeay, M., Fang, T., Bork, P., Jensen, L.J., von Mering, C., 2021.
938 The STRING database in 2021: customizable protein-protein networks, and functional
939 characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49,
940 D605–D612. <https://doi.org/10.1093/nar/gkaa1074>
- 941 36. Huang, J., Xiao, Y., Xu, A., Zhou, Z., 2016. Neutrophils in type 1 diabetes. *J.*
942 *Diabetes Investig.* 7, 652–663. <https://doi.org/10.1111/jdi.12469>
- 943
944

Cohort data collection

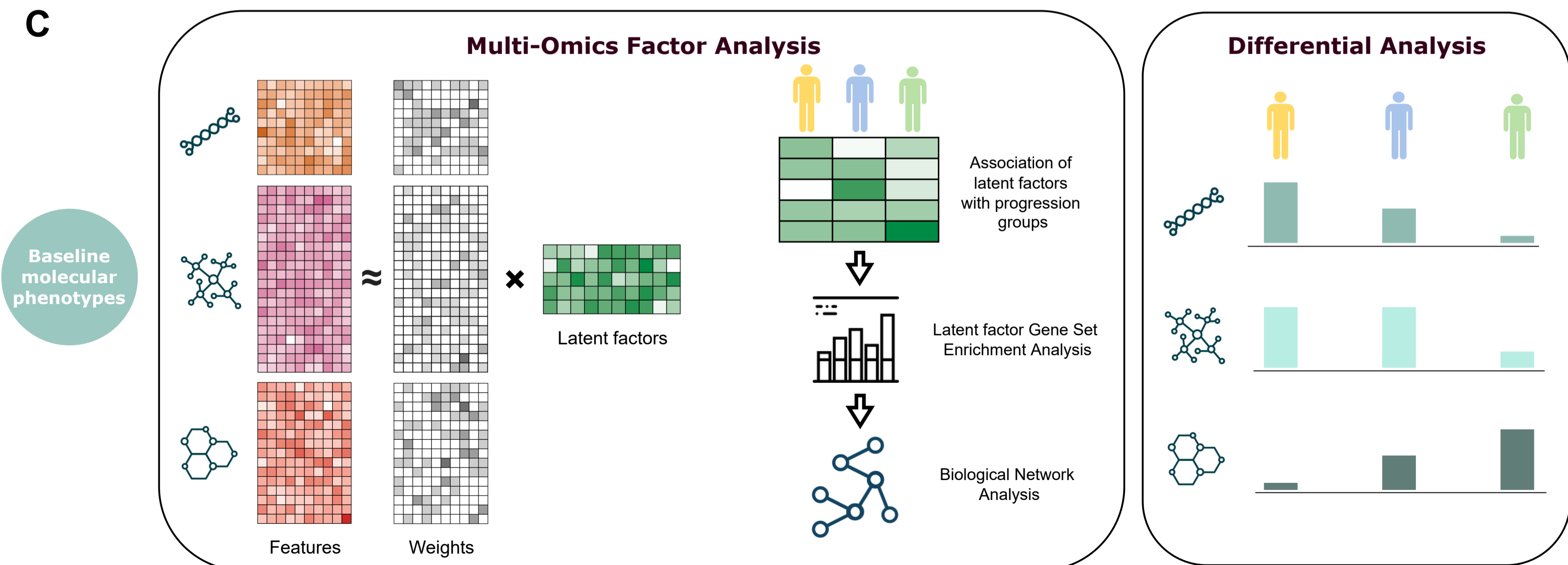
A



B



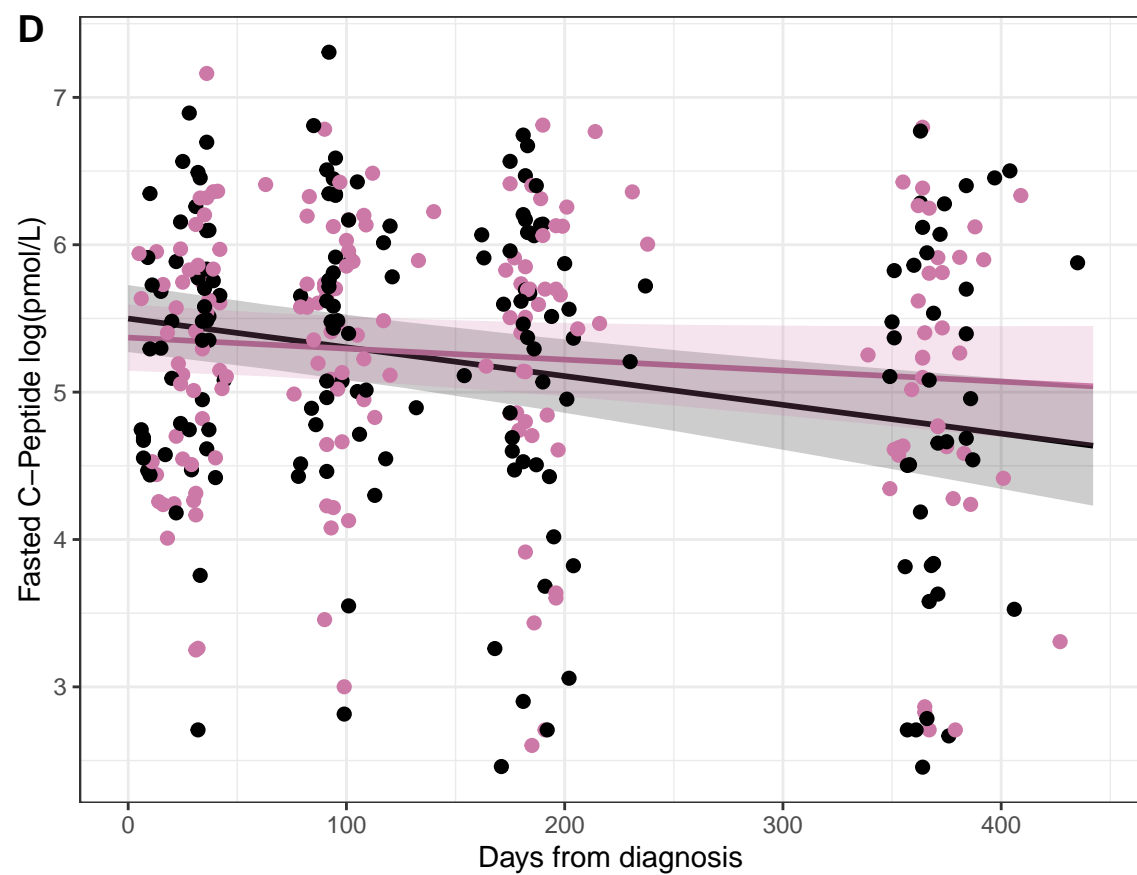
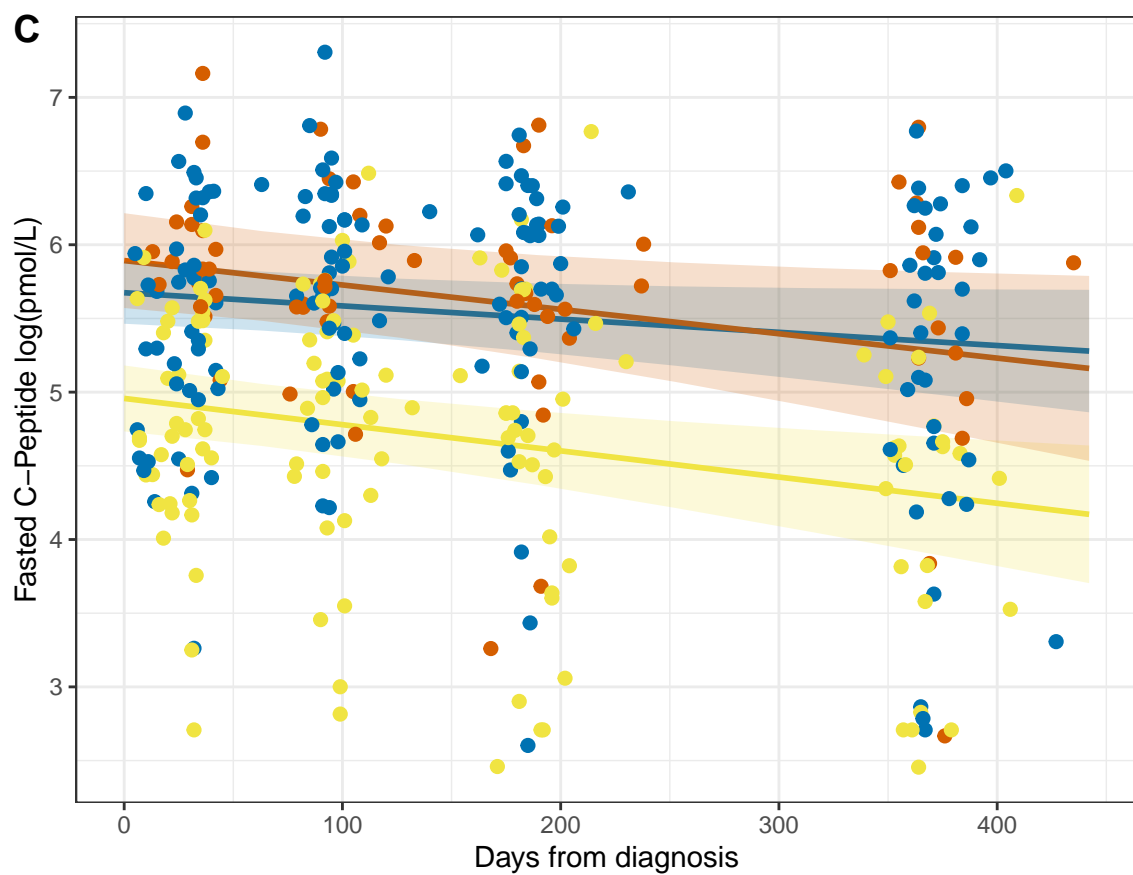
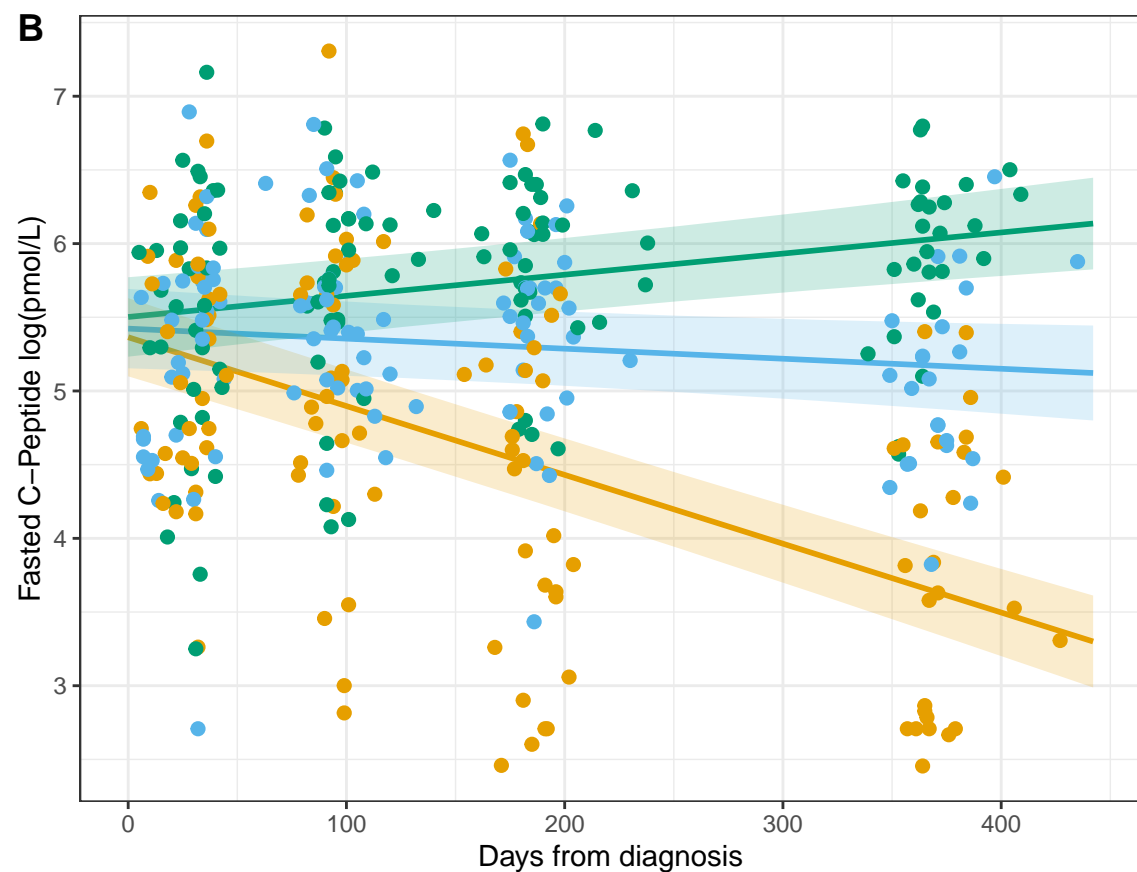
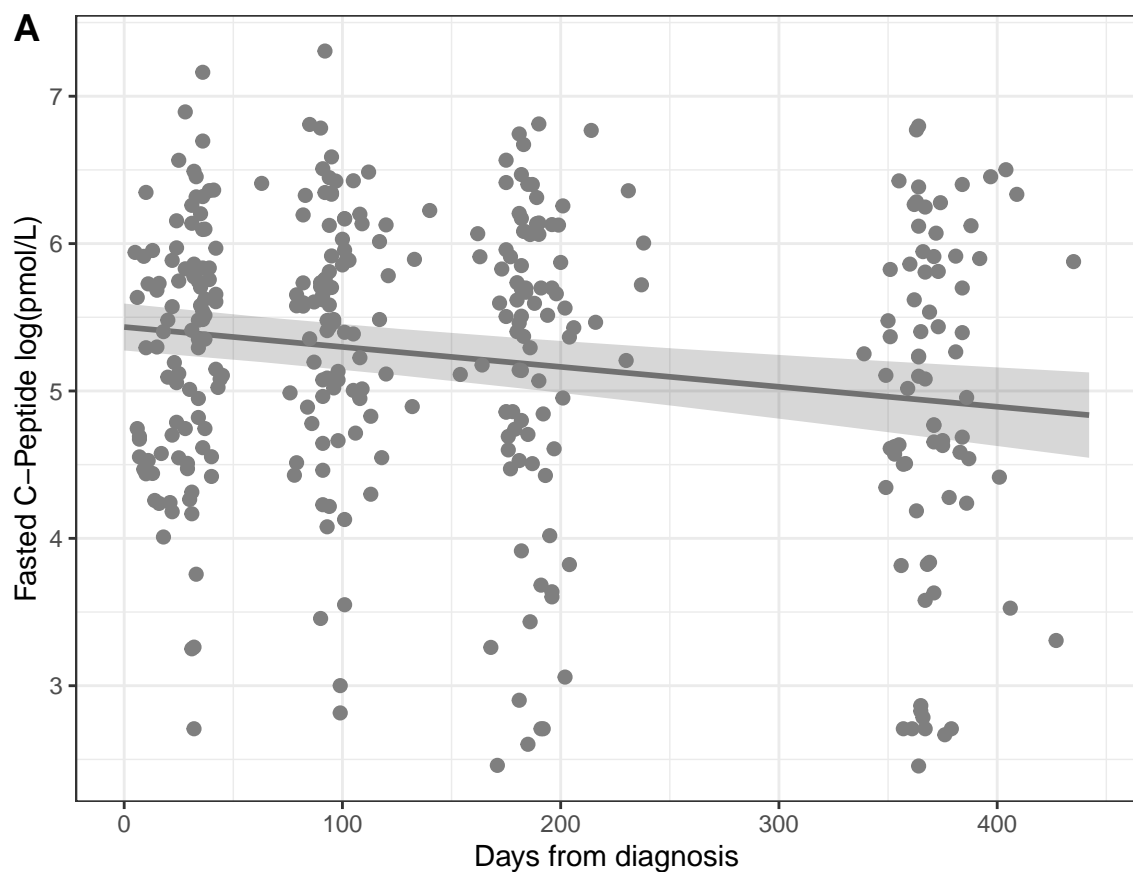
C

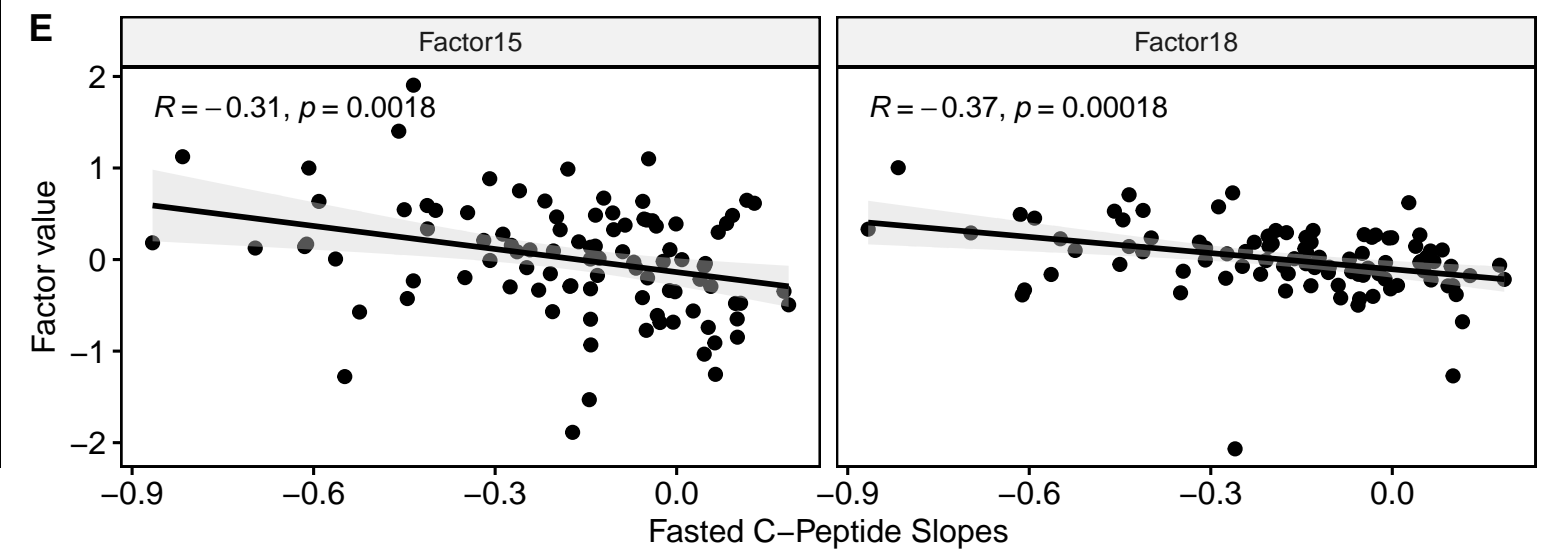
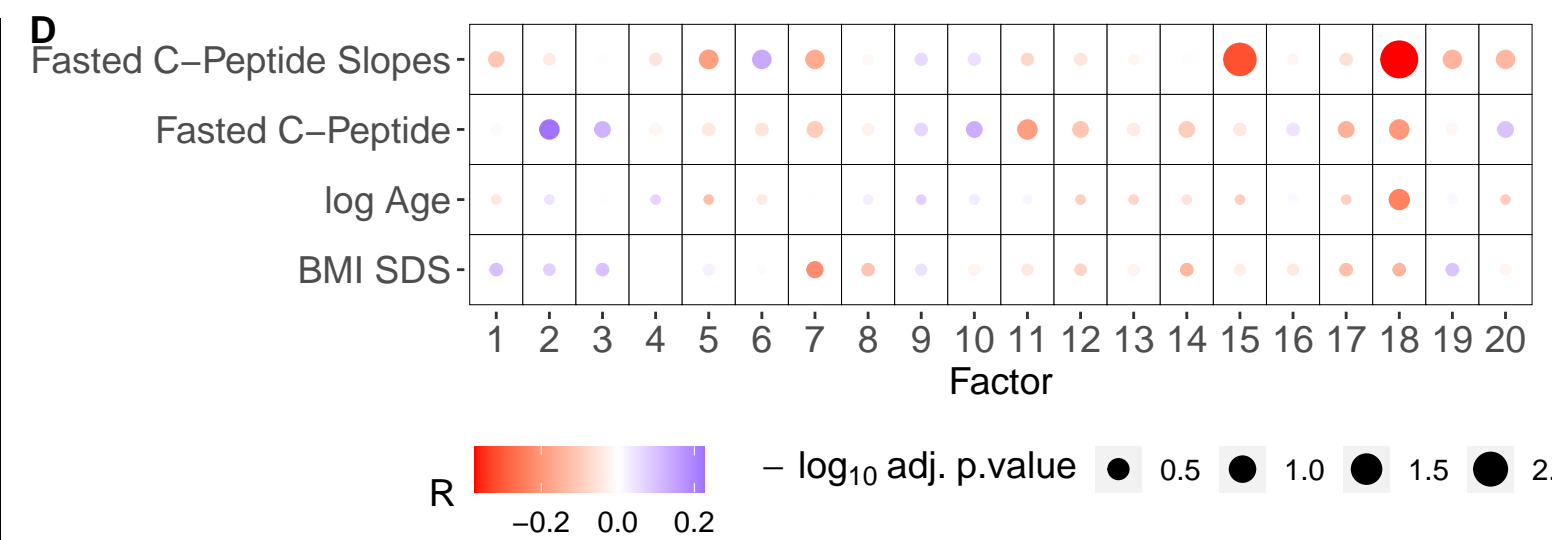
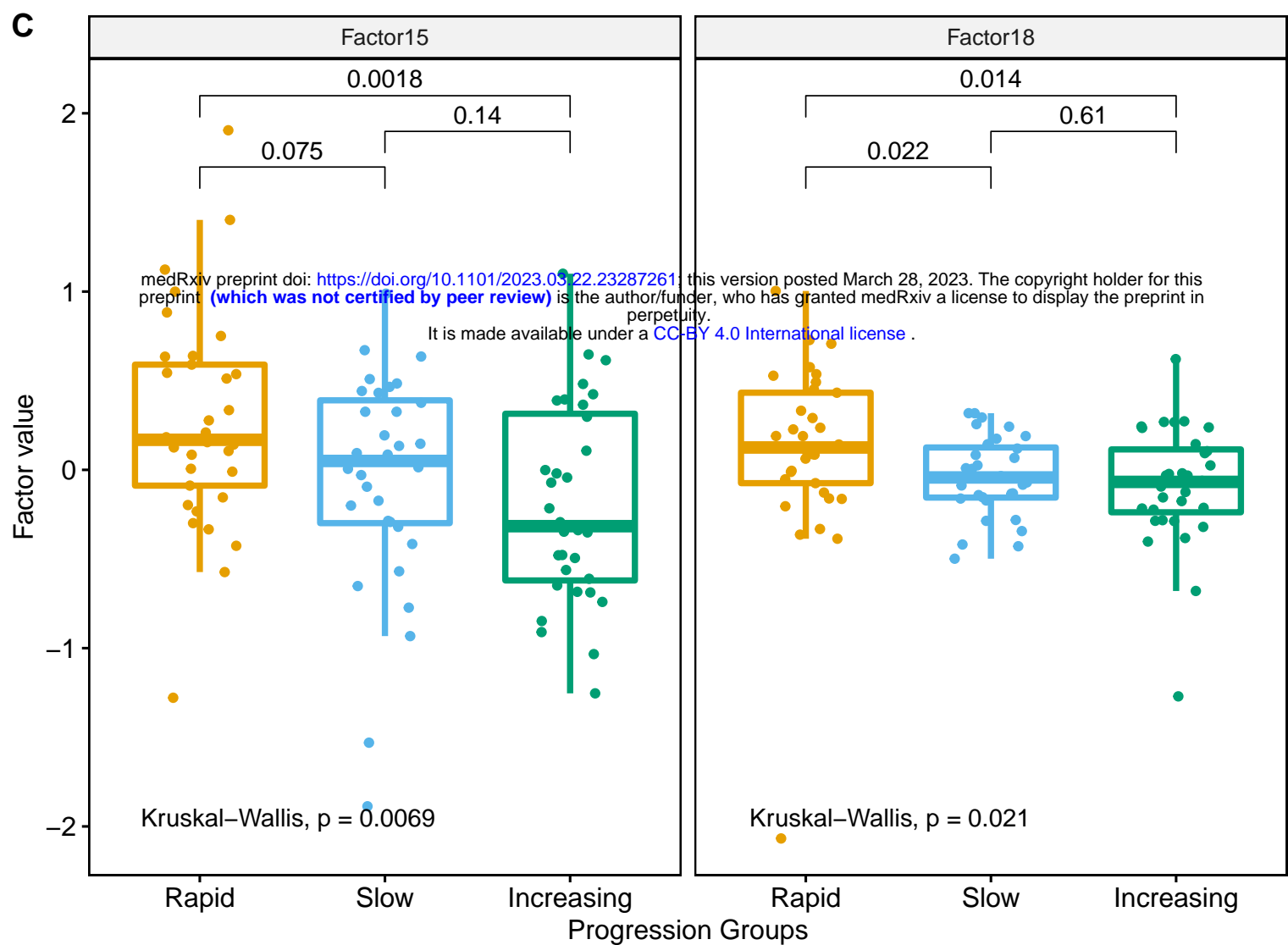
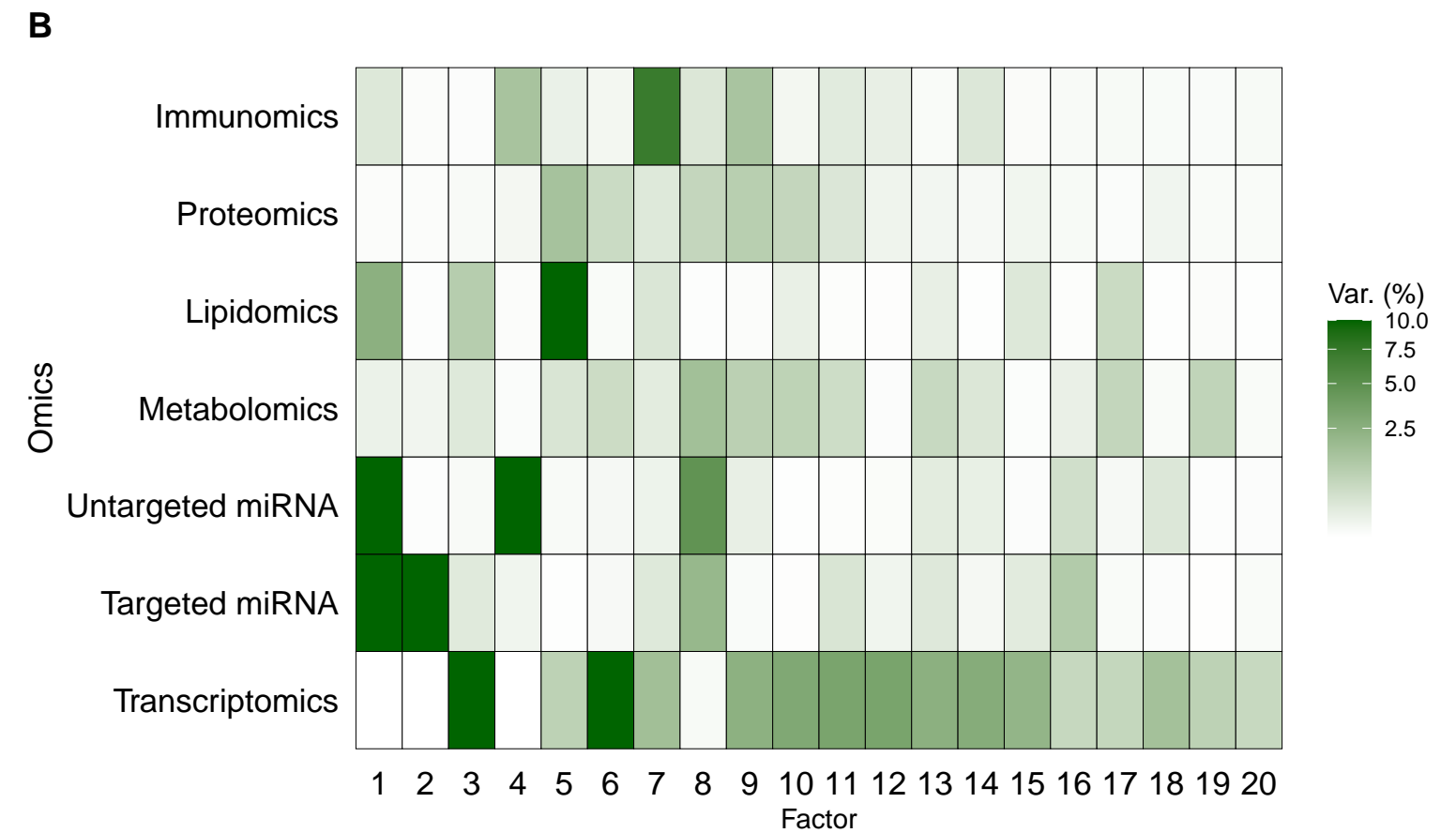
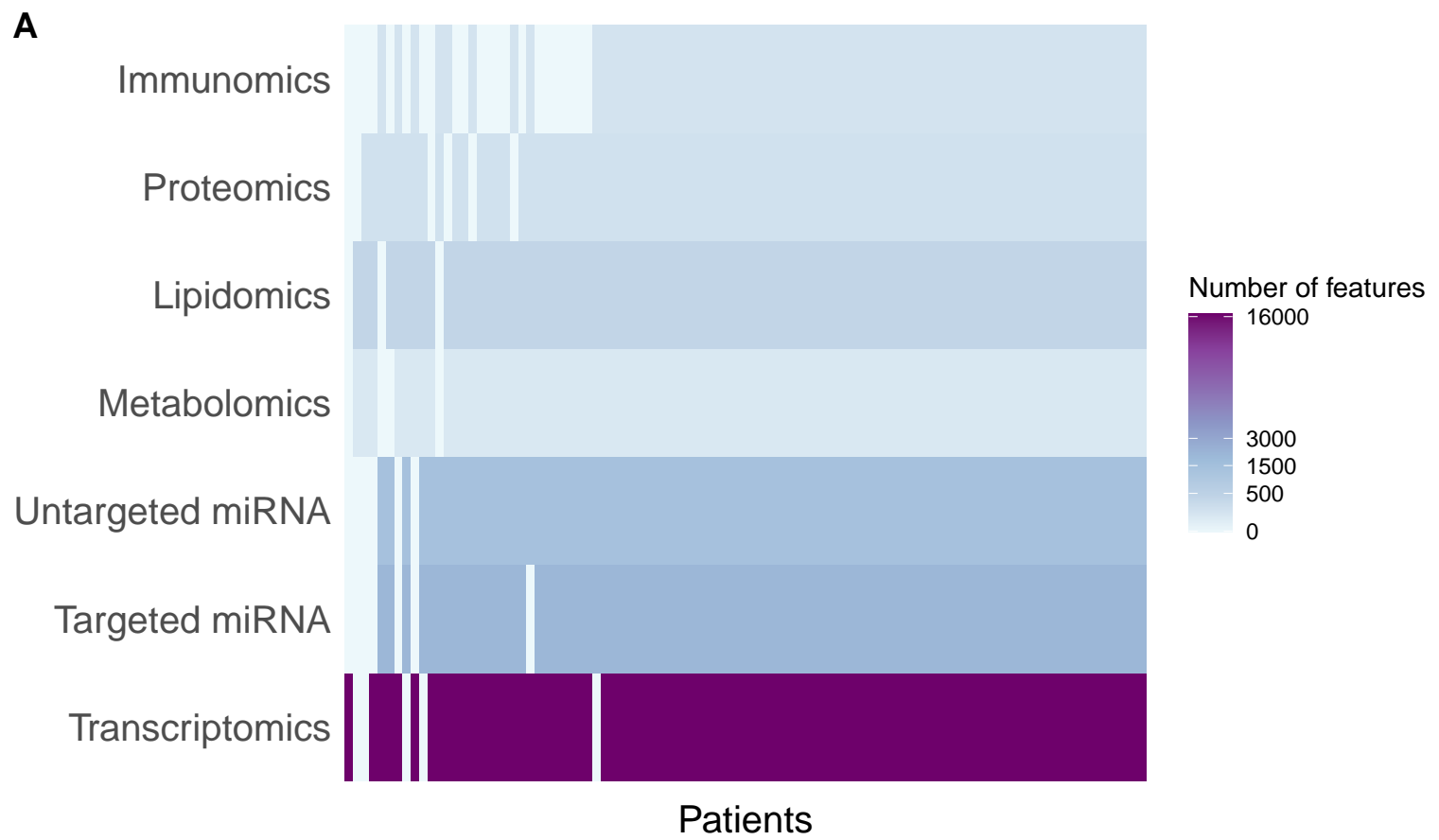


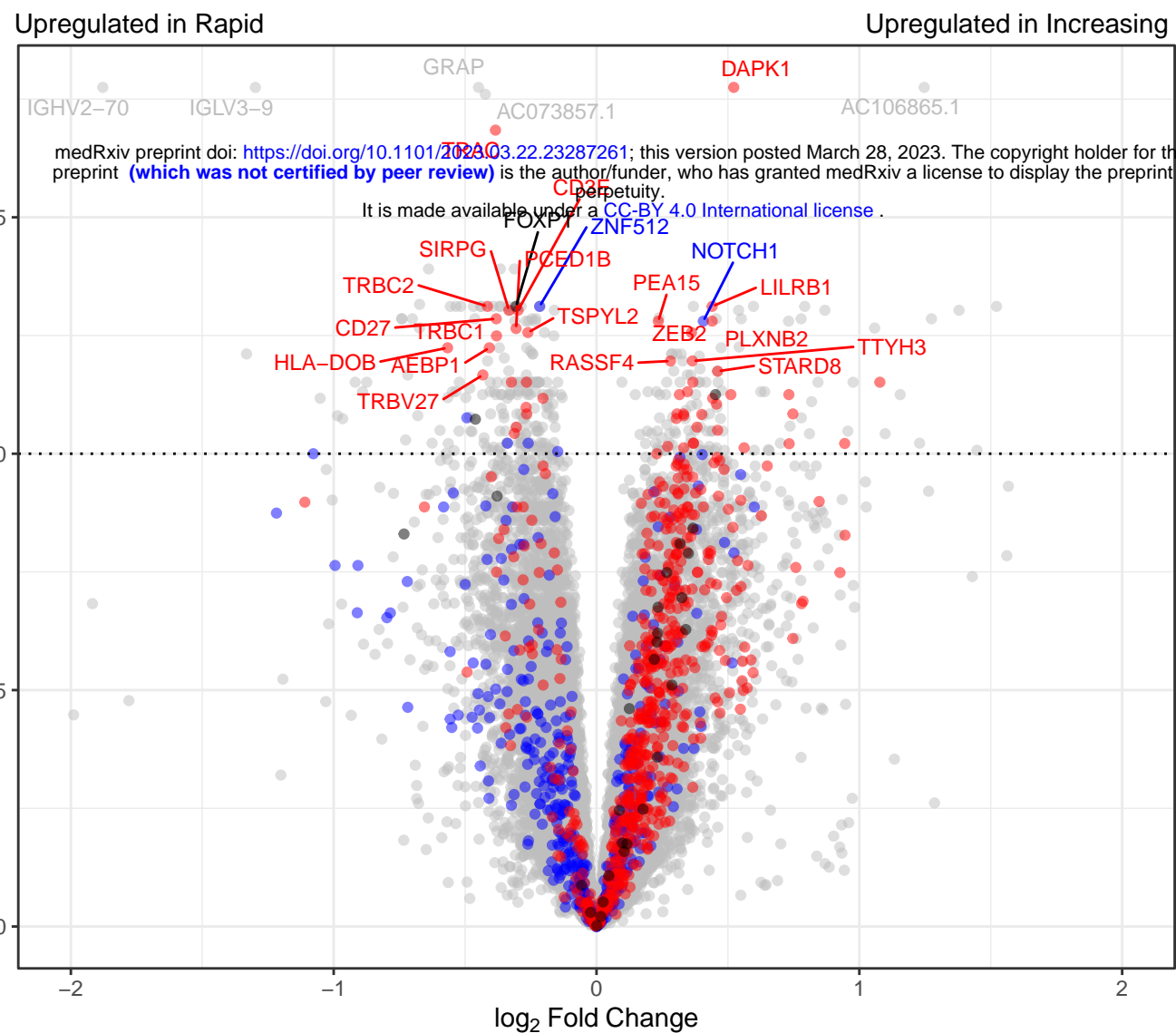
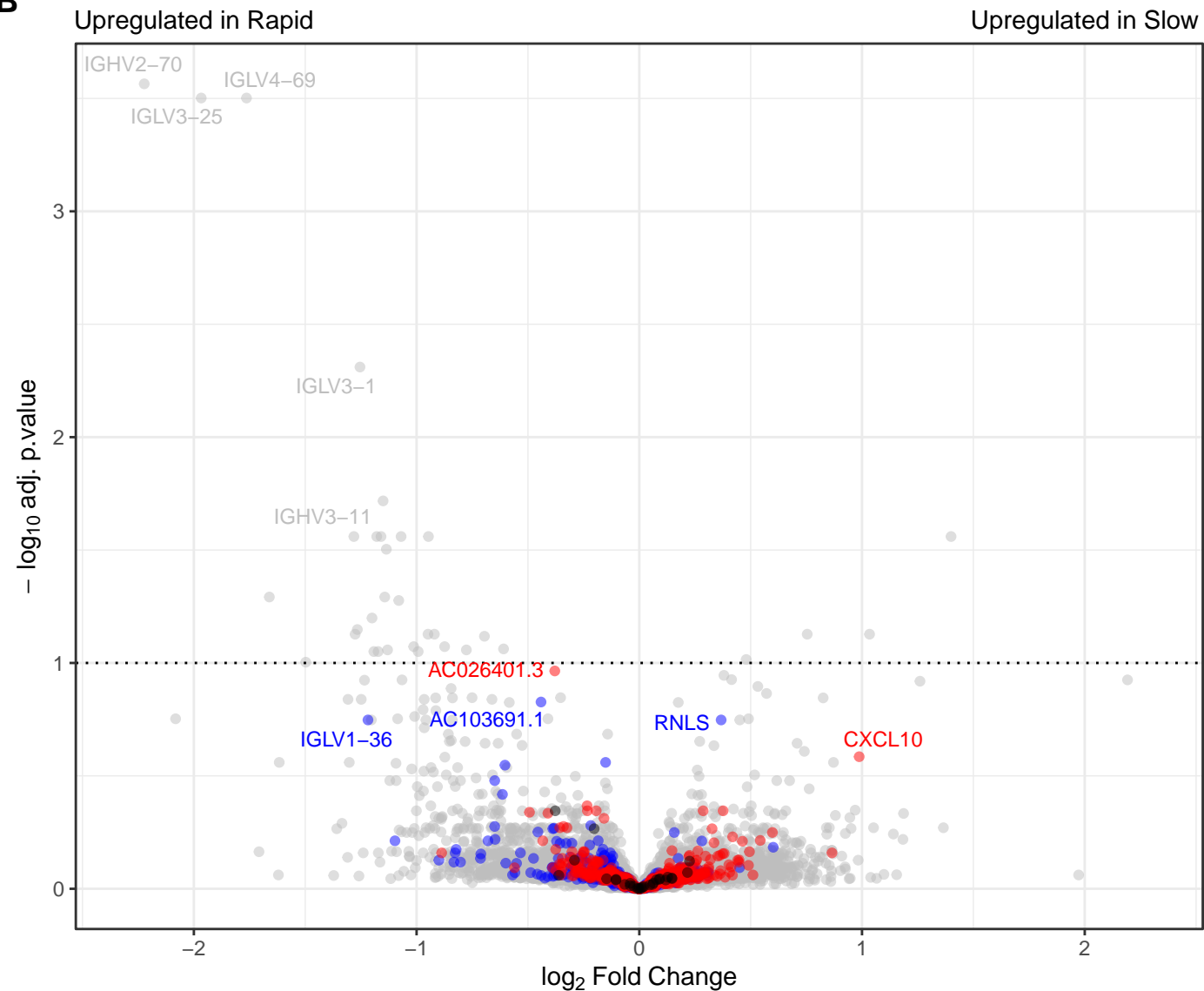
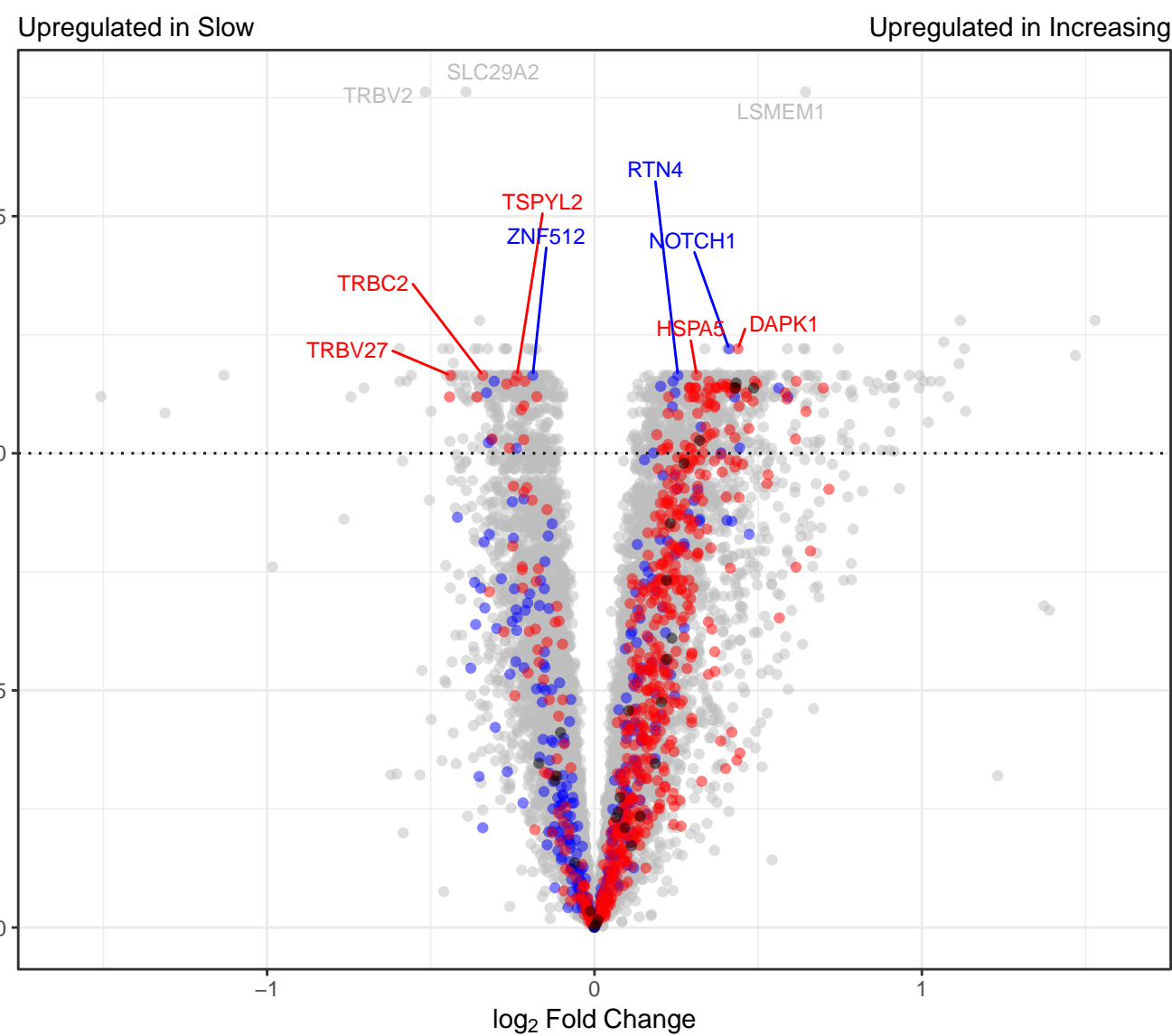
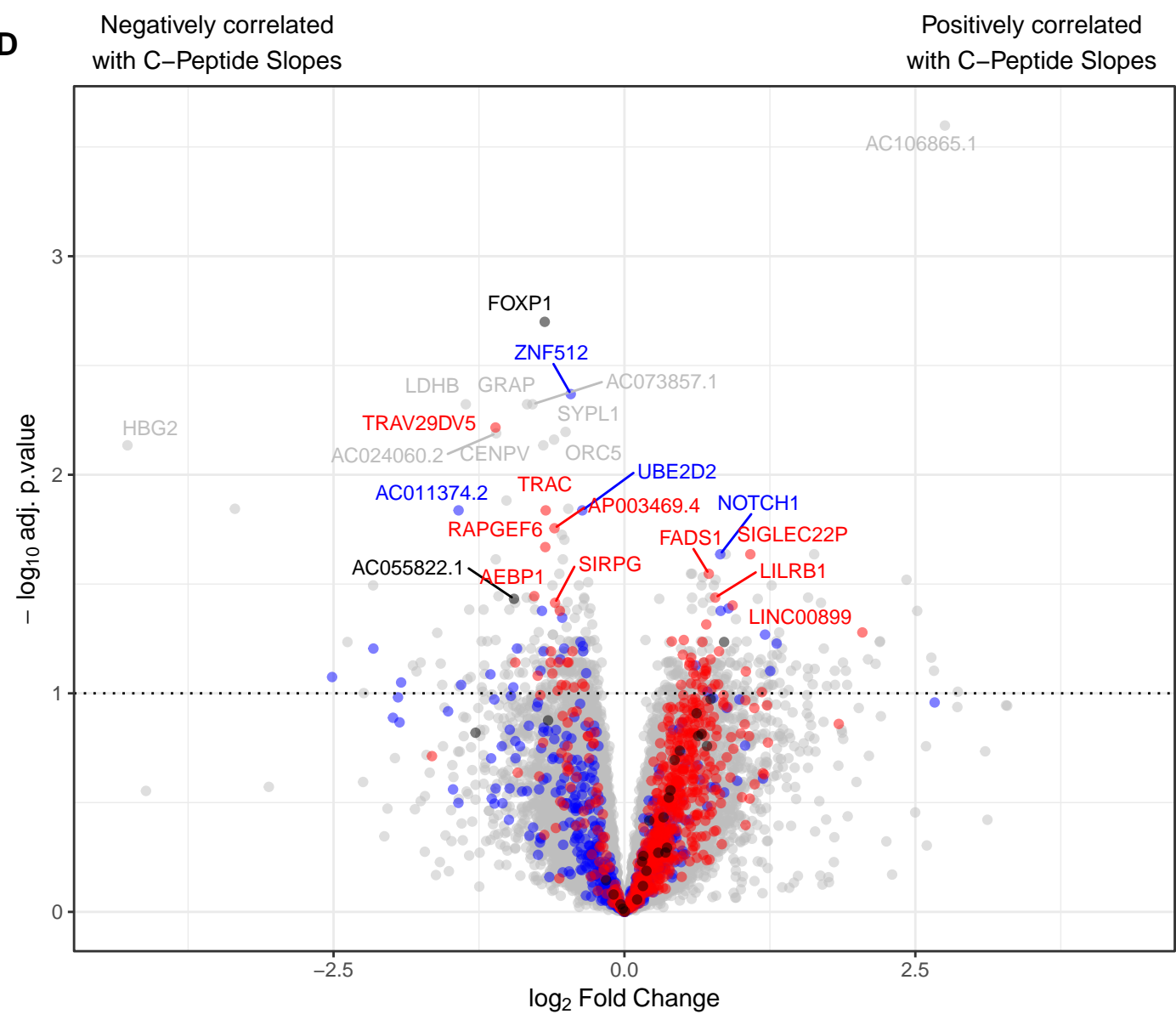
Progression groups — Rapid — Slow — Increasing

Age Intervals — <10 — >10-18 — >18

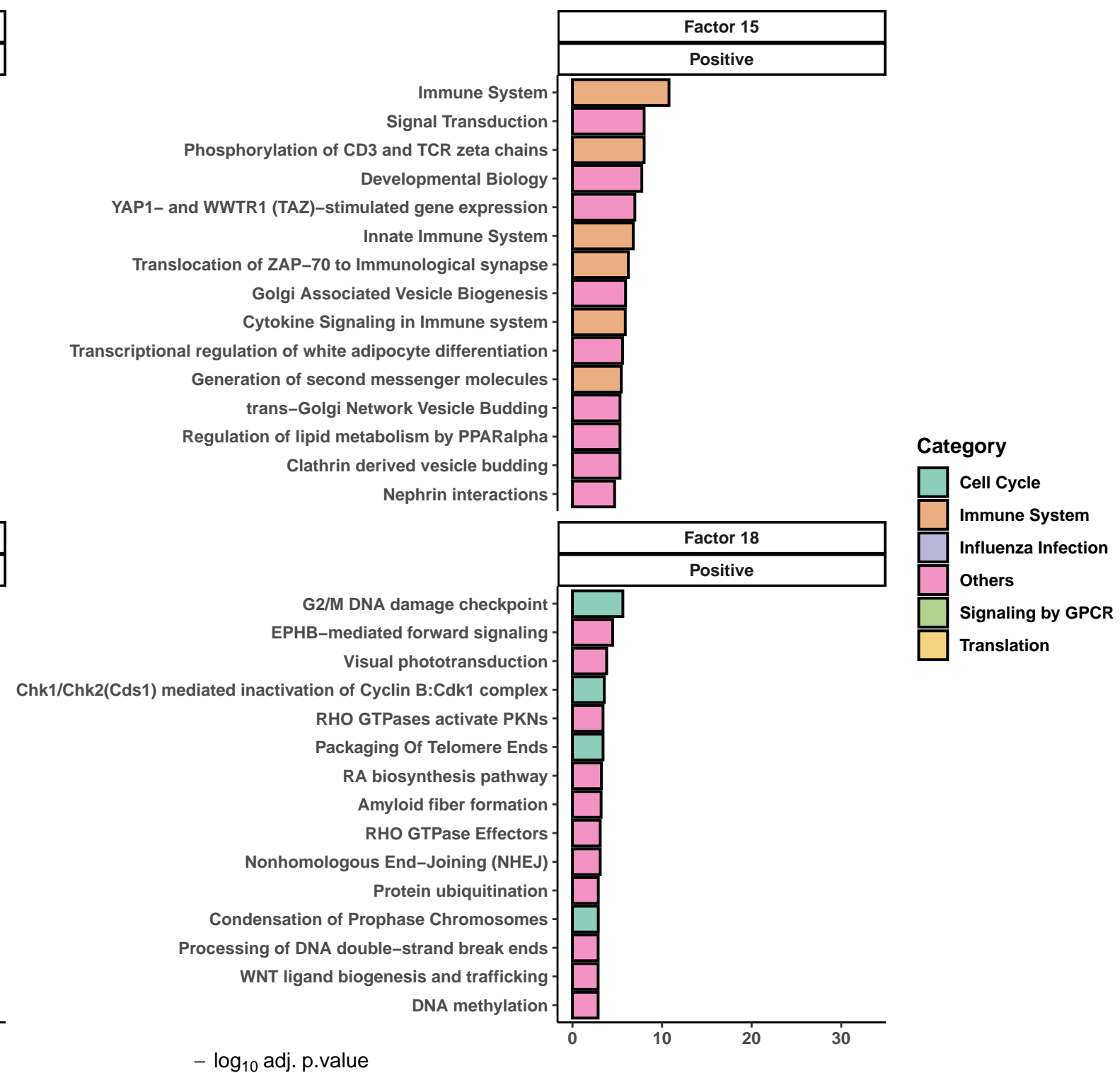
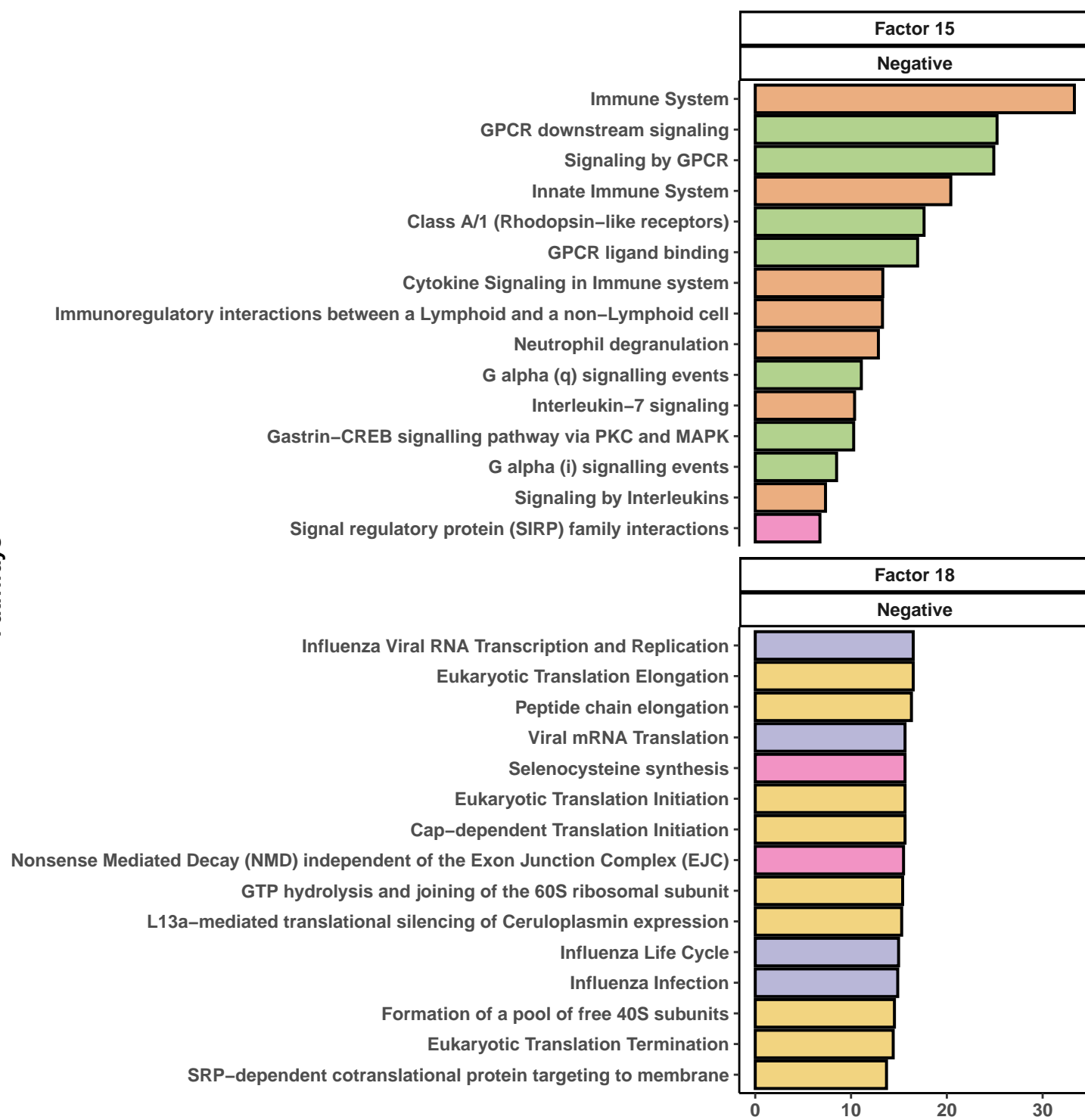
Sex — Female — Male





A**B****C****D**

Pathways



Category

- Cell Cycle
- Immune System
- Influenza Infection
- Others
- Signaling by GPCR
- Translation

$-\log_{10}$ adj. p.value

