
VTT: Evaluation of the validity of the data analysis and conclusions of the QBIX study

Study code: QBIX

EVALUATION REPORT

Signatures:

Evaluation report was prepared by:

Jouni Junnila, evaluator
Senior Biostatistician
4Pharma Ltd

14.3.2014

Date

Evaluation report was reviewed/approved by:

Teppo Huttunen
Manager, Statistics
4Pharma Ltd

14.3.2014

Date

Table of Contents

1	Primary objectives	4
2	Design and type of the study	4
3	Evaluation of statistical methods used in the article.....	4
4	Results of the conducted statistical analyses	5
4.1	Lipidomics, Wilcoxon rank sum test.....	5
4.2	Lipidomics, linear mixed effect models	7
4.3	Metabolomics, Wilcoxon rank sum test	9
4.4	Analyses related to seroconversion.....	11
4.5	The representativeness of the case report.....	12
5	Evaluation of the sufficiency of the included dataset.....	30
6	Conclusions	31
7	Execution of statistical analyses	31
8	Hardware and software	32

1 Primary objectives

To investigate if there is any research misconduct in a single study of Prof. Oresic (Dysregulation of lipid and amino acid metabolism precedes islet autoimmunity in children who later progress to type 1 diabetes, published on December 15, 2008, in “The Journal of Experimental Medicine”)ⁱ. In particular the following topics will be evaluated from statistical perspective:

- Evaluation of the correctness of the data analysis presented in the mentioned article.
- Are the conclusions presented in the article correct and in balance with statistical significances?
- How much the case report of the 9-year old child does impact the overall study results? Do the results of this one child direct the whole data analysis to a certain direction?

2 Design and type of the study

This is a longitudinal study of changes in serum metabolome of children who later progressed to type 1 diabetes. A matched control group of children who remained non-diabetic was used for comparisons.

The data related to birth was not given to 4Pharma; therefore the data analysis related to birth-data is not done. However, evaluation of the presentation of the results, also related to birth data can be conducted.

3 Evaluation of statistical methods used in the article

In principal, the statistical methods used in the statistical analyses presented in the article are valid and cannot be considered to introduce significant bias to the results.

The difference between the progressors and non-progressors was analyzed with Wilcoxon rank sum test in both the lipidomics and metabolomics datasets by age cohort. This nonparametric procedure is not sensitive for distribution assumptions but is less powerful than other parametric methods. One particular problem in using this method in this kind of setting is the fact that it ignores the fact that the data has a matched structure. This also reduces the power of the analyses, but shouldn't introduce bias favoring finding significant differences between progressors and non-progressors.

In addition to the Wilcoxon rank sum tests certain sum variables in the lipidomics data were analyzed with linear mixed effects models (using the same derived data as with the Wilcoxon rank sum tests). This kind of model is more powerful and takes into consideration the intra-subject variation. The estimate for the overall difference between the two groups is averaged over the age cohorts (10 cohorts). The analysis does not consider the fact that the difference between the groups could be dependent on age. In addition, a fairly large difference detected in one age cohort can lead to stating that the group difference is significant (even though it would be significant in only one age cohort). The analysis would be less biased if we would have more even amount of observations per age cohort. Therefore in this report also analyses using only the data from first 4 years of the lifespan are provided.

A big amount of statistical comparisons are conducted in the article. The “Statistical methods” section states that these multiple comparisons were accounted for by calculating the false discovery rates (FDR) by age cohort. However, the p-values presented e.g. in the heatmaps (figures 3a, 4 and 5) are actually the raw p-values. Also, in the results text the raw p-values < 0.05 are interpreted as significant differences. From the evaluator's perspective this kind of procedure does not really

account for multiple comparisons. Although the FDRs are presented, they are in principal ignored in the interpretation. More conservative and less biased method would be to first define an acceptable false discovery rate and then adjusting the p-values accordingly. In the article the presented FDRs are partly clearly unacceptable (for the metabolomics data). However this fact is ignored in the text of the article.

4 Results of the conducted statistical analyses

4.1 Lipidomics, Wilcoxon rank sum test

The raw p-values of the conducted Wilcoxon rank sum tests by age cohort are presented in Table 1. It must be noted that there were some small differences in the number of observations presented in the article vs analyses conducted by 4Pharma (maximum of 2 observations by age cohort / group combination). The reason for the differences was left unclear but shouldn't have a significant effect on the results. The new results are in line with the results presented in the article.

	wilcox01	wilcox02	wilcox03	wilcox04	wilcox05	wilcox06	wilcox07	wilcox08	wilcox09	wilcox10
Cer.d18.1.22.0.	0.02	0.10	0.82	0.25	0.78	0.41	0.60	0.87	0.49	0.06
PC.18.0.0.0.	0.01	0.00	0.70	0.14	0.92	0.90	0.20	0.80	0.21	0.37
PC.18.1.0.0.	0.22	0.05	0.87	0.58	0.38	0.57	0.58	0.66	0.23	0.44
PC.18.2.0.0.	0.33	0.14	0.04	0.55	0.07	0.20	0.81	0.22	1.00	0.16
PC.18.3.0.0.	0.33	0.04	0.82	1.00	0.47	0.97	0.89	0.28	0.13	1.00
PC.14.0.18.2.	0.08	0.22	0.91	0.08	0.10	0.33	0.01	0.14	0.05	0.68
PC.18.2.16.1.	0.02	0.02	0.19	0.04	0.10	0.07	0.04	0.02	0.11	0.77
PC.18.0.18.1.	0.30	0.12	0.74	0.50	0.77	0.69	0.78	0.52	0.58	0.77
PC.18.2.18.2.	0.40	0.34	0.70	0.87	0.25	0.86	0.02	0.17	0.17	0.68
PC.16.0.20.5.	0.14	0.01	0.87	0.18	0.80	0.95	0.05	0.07	0.06	0.21
PC.18.0.20.4.	0.42	0.16	0.59	0.58	0.60	0.76	0.92	0.98	0.76	0.21
PC.18.2.20.4.	0.21	0.04	0.59	0.57	0.95	0.66	0.18	0.14	1.00	0.02
PC.38.7.	0.44	0.16	0.52	0.11	0.33	0.15	0.04	0.21	0.02	0.10
PC.18.0.22.6.	0.28	0.02	0.54	0.69	0.94	0.93	0.12	0.09	0.17	0.02
PC.32.1e.	0.78	0.07	0.00	0.13	0.58	0.08	0.36	0.01	0.17	0.95
PC.34.1e.	0.29	0.20	0.01	0.53	0.79	0.02	0.26	0.04	0.76	0.68
PC.16.0e.20.4.	0.24	0.01	0.25	0.54	0.62	0.99	0.55	0.04	0.46	0.10
PC.38.4e.	0.46	0.00	0.12	0.53	0.45	0.73	0.70	0.56	0.90	0.51
PC.18.1e.20.4.	0.17	0.01	0.45	0.92	0.54	0.86	0.36	0.10	0.38	0.21
PE.38.1.	0.98	0.75	0.38	0.61	1.00	0.75	0.53	0.83	0.32	0.95
PE.38.2.	0.53	0.09	0.03	0.44	0.83	0.27	0.29	0.41	0.53	1.00
SM.d18.1.16.0.	0.54	0.30	0.96	0.82	0.68	0.61	0.29	0.02	0.85	0.86
SM.d18.1.16.1.	0.52	0.73	0.08	0.47	0.85	0.41	0.63	0.08	0.80	0.59
SM.d18.1.22.0.	0.07	0.34	0.58	0.99	0.67	0.30	0.17	0.52	0.95	0.25
SM.d18.1.23.1.	0.19	0.12	0.10	0.50	0.89	0.16	0.16	0.94	0.90	0.51

SM.d18.1.24.0.	0.07	0.03	0.90	0.62	0.83	0.82	0.08	0.38	1.00	0.06
SM.d18.1.24.1.	0.35	0.08	0.50	0.39	0.66	0.30	0.73	0.33	0.29	0.51
SM.d18.1.24.2.	0.43	0.38	0.47	0.43	0.26	0.26	0.92	0.07	0.95	0.77
SM.d18.1.25.1.	0.30	0.20	0.29	0.82	0.92	0.25	0.40	0.80	0.95	0.21
TG.42.0.	0.12	0.54	0.04	0.40	0.39	0.40	0.97	0.80	0.15	0.95
TG.16.0.14.0.16.0.	0.14	0.29	0.13	0.29	0.23	0.73	0.94	0.76	0.13	1.00
TG.18.1.16.0.12.0.	0.07	0.32	0.17	0.17	0.26	0.73	0.68	0.36	0.06	1.00
TG.46.2.	0.08	0.21	0.21	0.24	0.21	0.84	0.73	0.91	0.06	1.00
TG.16.0.16.0.16.0.	0.04	0.40	0.14	0.33	0.17	0.90	0.83	0.41	0.35	0.86
TG.18.1.14.0.16.0.	0.02	0.40	0.52	0.20	0.75	0.97	0.33	0.11	0.13	0.59
TG.16.0.18.2.14.1.	0.04	0.07	0.20	0.19	0.34	0.75	0.51	0.98	0.03	0.68
TG.16.0.18.1.15.0.	0.15	0.42	0.11	0.48	0.47	0.44	0.68	0.56	0.11	0.59
TG.16.0.18.0.16.0.	0.21	0.44	0.03	0.09	0.04	0.48	0.81	0.33	0.38	1.00
TG.16.0.16.0.18.1.	0.02	0.50	0.92	0.18	1.00	0.22	0.81	0.06	0.71	0.21
TG.18.1.16.0.16.1.	0.03	0.35	0.61	0.46	0.83	0.31	0.46	0.16	0.90	0.44
TG.16.1.16.1.18.1.	0.03	0.05	0.43	0.31	0.54	0.84	0.18	0.83	0.17	0.13
TG.16.0.18.1.17.0.	0.14	0.32	0.04	0.06	0.24	0.88	0.92	0.44	0.23	0.77
TG.18.1.18.1.15.0.	0.08	0.18	0.12	0.06	0.36	0.90	0.97	0.94	0.11	0.95
TG.16.0.18.1.18.0.	0.09	0.45	0.03	0.03	0.10	0.62	0.36	0.17	0.19	0.37
TG.18.2.16.0.18.1.	0.05	0.07	0.23	0.61	0.89	0.46	0.75	0.91	0.85	0.37
TG.16.0.18.2.18.2.	0.02	0.01	0.12	0.54	0.75	0.99	0.20	0.80	1.00	0.16
TG.16.0.18.3.18.2.	0.12	0.00	0.08	0.54	0.99	0.61	0.92	0.10	0.38	0.44
TG.17.0.18.1.18.2.	0.03	0.10	0.01	0.12	0.27	0.64	0.58	0.94	0.38	0.21
TG.18.0.18.1.18.1.	0.02	0.08	0.00	0.02	0.07	0.34	0.05	0.49	0.32	0.21
TG.18.1.18.1.18.1.	0.02	0.18	0.11	0.38	0.53	0.38	0.23	0.87	0.80	0.86
TG.18.2.18.1.18.1.	0.05	0.04	0.12	0.60	0.87	0.76	0.33	0.98	1.00	1.00
TG.18.1.18.2.18.2.	0.03	0.02	0.09	0.68	0.78	0.62	0.36	0.83	0.71	1.00
TG.56.6.	0.04	0.01	0.08	0.28	0.39	0.80	0.73	0.72	0.07	0.21

The false discovery rates in some age cohorts were large (age cohorts 4-5, 7-8 and 9-10). This enlarges the possibility of stating a difference to be significant even though the difference would be there just by chance, presenting for example the q-values would have been recommendable. However, the data from other age cohorts does not have this problem, as the FDRs are fairly low.

The article abstract states that “Individuals who developed diabetes had reduced serum levels of phosphatidylcholine (PC) at birth”. In other words as written in the abstract (where usually the main results are presented) this sentence seems to state that there is a meaningful and significant difference in some of the PCs between the progressors and non-progressors at birth as tested by the Wilcoxon rank sum test. However, at birth, only one raw p-value for the group difference is below 0.05 in the lipidomics dataset (PC(18:2/18:2)). The FDR for this cohort is estimated at 0.94. This means that there is in fact 94 % probability that even the one detected difference is a false positive and only a 6 % probability that there really would be an actual difference in PC(18:2/18:2). Therefore, the PC levels at birth cannot be said to be reduced in individuals who developed diabetes in general.

4.2 Lipidomics, linear mixed effect models

The linear mixed effects model was fitted for the following variables: Total TG, Total ether PCs, Total PCs, and Total sphingomyelins. Some discrepancies were found from the results. Line plots were presented for Total TG and Total ether PC in the article. In the line plot related to Total TG it seems that for some reason the values for controls clearly go up in the age of 2-3 years, as reproduced in Figure 1 of this document.

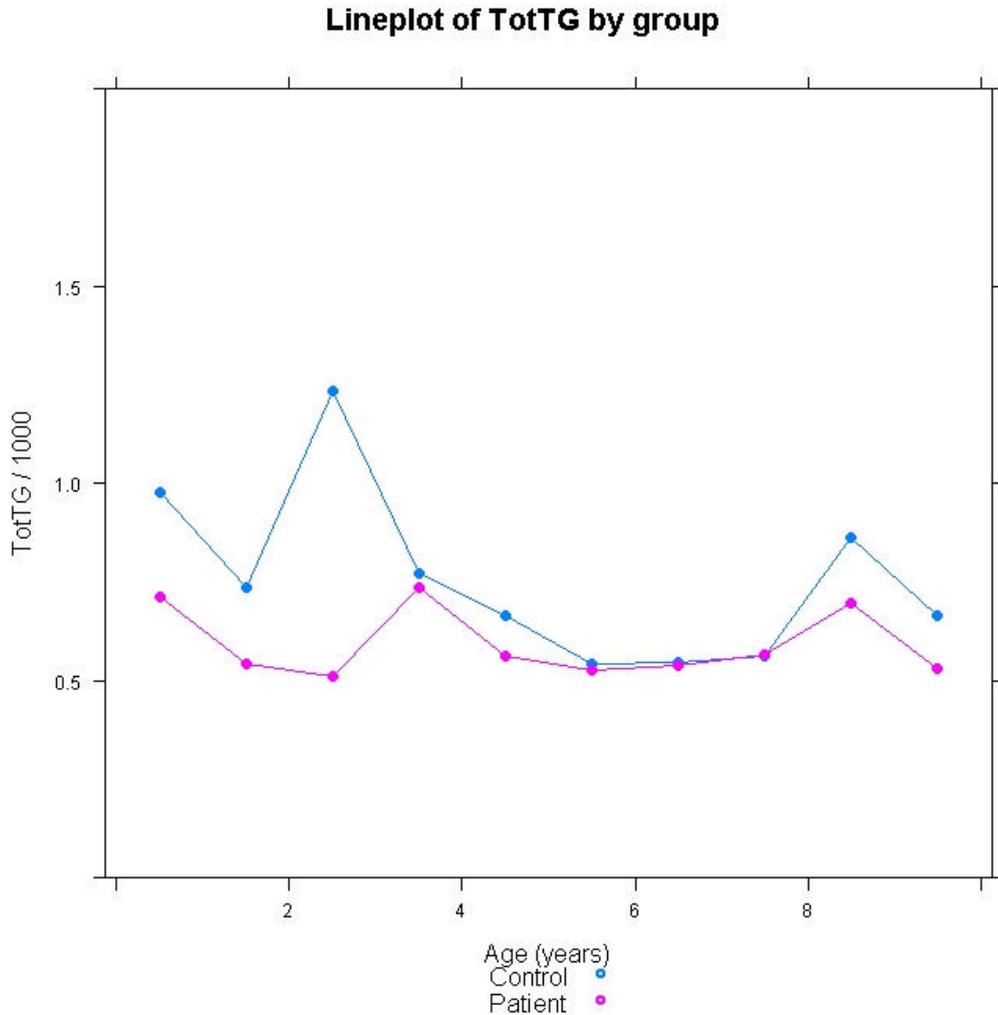


Figure 1. Line plot by group for Total TG.

However this apparent increase in the values of non-progressors is due to only one single extremely high observation (subject ID=-41.5). A more representative figure is achieved by removing this subject from the data. This kind of figure is presented in Figure 2 of this document.

Lineplot of TotTG by group (one extreme outlier removed)

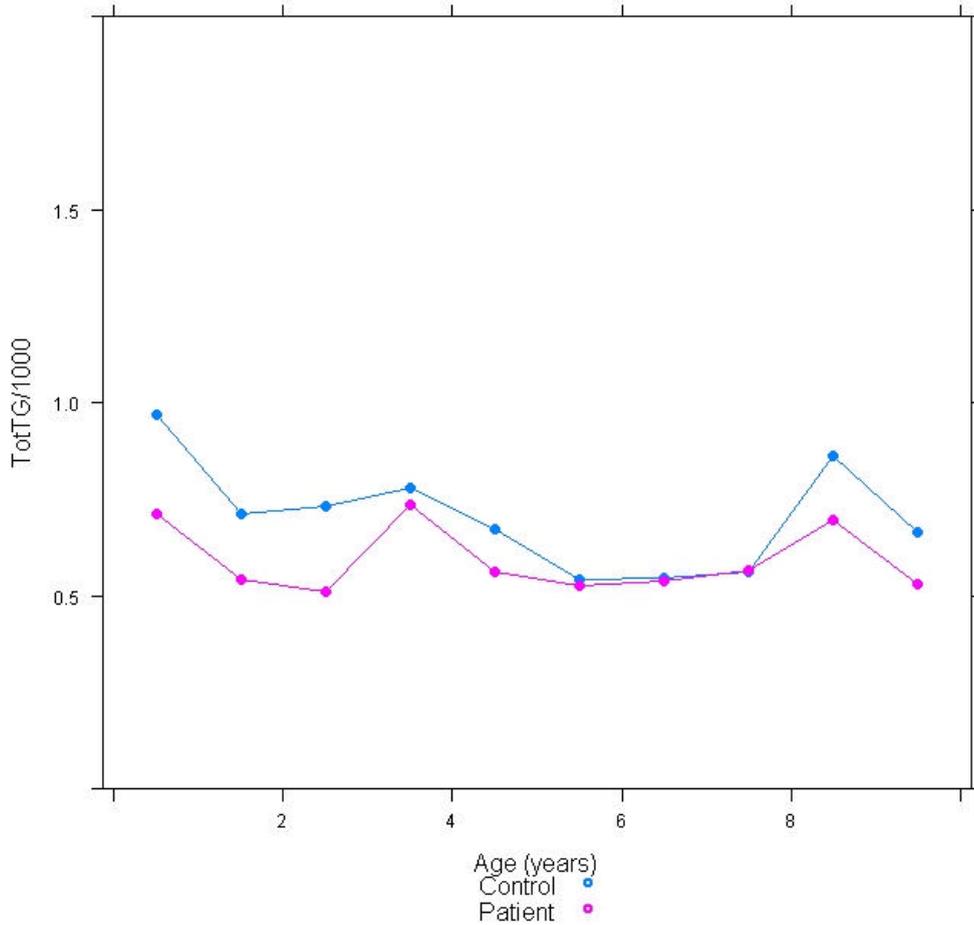


Figure 2. Line plot by group for Total TG (One extreme outlier removed).

From the perspective of statistical significance, intuitively one would think that we get better results with the data from Figure 1 than with the data from Figure 2. However, this is not the case, as due to the extreme outlier, the standard errors are clearly bigger with the full data, increasing the p-values. It seems that this outlier has been noted by the authors and they actually present the results of the model excluding the outlier. So the analysis dataset is not the same than the data presented in the Figure 3 B. This operation can be seen as a way to falsely convince the reader, whereas the presented p-value is actually correct.

Table 2. Type 3 tests of fixed effects (full data)

	numDF	denDF	F-value	p-value
Group	1	126	4.07	0.0457
Year	9	448	1.04	0.4081

Table 3. Type 3 tests of fixed effects (ID -41.5 removed)

	numDF	denDF	F-value	p-value
Group	1	125	7.39	0.0075
Year	9	444	2.79	0.0034

For the models related to Total Ether PC and Total PC the fitted results are comparable with the article (not exactly the same though due to data issues mentioned earlier).

The results related to Total Sm were interpreted as “progressors showed a consistent decrease ($P = 0.09$)”. The p-value is above the cut-off limit of 0.05 so the result cannot be interpreted as consistent decrease. Secondly, the model fitted by 4Pharma gives a clearly higher p-value for the group difference (> 0.2). The reason for this difference between the models fit is left unclear.

The biggest differences were mainly seen with older ages. The older age cohorts were very small, making it more probable that the differences detected there are generated by chance. Therefore, a sensitivity analysis using only the data from the younger age groups is vital and should give more robust results than the full data.

For all of the four variables the group difference gets smaller, when using a subset of the data (< 4 years). However, the difference remains for TotTG, TotPcE and TotPC significant (p-values range from 0.01 to 0.04). TotSM is insignificant as with the full data.

The interpretation of the linear mixed effects models in the article is mainly ok but in some cases a bit biased. It is mentioned in several occasions that the values analyzed with linear mixed models were consistently down-regulated throughout the follow-up. However, with some of these variables the mean curves actually cross at some point during the follow-up. Our view is that this kind of data cannot therefore be interpreted so strongly (even though the p-value for the overall group difference would be significant) and modeling younger age groups separately would have been reasonable. The Wilcoxon rank sum tests already give an impression that there might not be that consistent difference between the groups.

4.3 Metabolomics, Wilcoxon rank sum test

The raw p-values of the conducted Wilcoxon rank sum tests by age cohort are presented in Table 4.

Table 4. Wilcoxon rank sum-tests for differences between progressors and non-progressors (metabolomics)

	wilcox01	wilcox02	wilcox03	wilcox04	wilcox05	wilcox06
Citric.acid	0.59	0.81	0.00	0.13	0.81	0.10
Succinic.acid	0.07	0.80	0.08	0.01	0.92	1.00
Creatinine	0.43	0.89	0.33	0.50	0.10	0.45
Ribitol	0.97	0.46	0.93	0.28	0.06	0.29
Tryptophan	0.48	0.55	0.43	0.35	0.38	0.06
Tyrosine	1.00	0.55	0.34	0.43	0.81	0.61
L.idofuranose	0.65	0.80	0.51	0.19	0.38	0.35
X1H.Indole.3.acetic.acid	0.59	0.10	0.73	0.35	0.05	0.07
Ethylamine	0.71	0.75	0.04	0.35	0.86	0.82
X2.hydroxypyridine	0.26	0.97	0.29	0.63	0.51	0.96
X2.Monopalmitin.ether	0.34	0.92	0.38	0.63	0.65	0.33
Ketoleucine	0.62	0.34	0.12	0.48	0.21	0.29
X2.Propenoic.acid	0.74	0.60	0.86	0.19	0.81	0.10
X4.7.10.13.16.19.Docosaheptaenoic.acid	1.00	0.64	0.79	0.15	1.00	0.09
Phenol	0.48	0.44	0.73	0.13	0.34	0.96
X9.12.Octadecadienoic.acid..Z.Z.	0.65	0.60	0.12	0.39	0.05	0.29
X9.12.Octadecadienoic.acid..Z.Z	0.23	0.70	0.60	0.53	0.02	0.89
X9.Octadecenoic.acid..Z.	0.46	0.45	0.48	0.42	0.97	0.78

Acetic.acid	0.48	0.31	0.26	0.01	0.22	0.21
Aminomalonic	0.11	0.62	0.60	0.53	0.97	0.20
beta.Sitosterol	0.59	0.48	0.19	0.76	0.75	0.68
X2.Hydroxybutanoic.acid	0.54	0.34	0.43	0.48	0.20	0.18
Butylated.Hydroxytoluene	0.43	0.70	0.86	0.19	0.17	0.55
Cholest.5.en.3.ol	0.77	0.97	0.66	0.79	0.51	0.49
Cholesterol	0.87	0.86	0.66	0.04	0.49	0.24
Decanoic.acid	0.84	0.34	0.10	0.60	0.03	0.96
d.Erythrotetrofuranose	0.97	0.12	0.54	0.35	0.70	0.89
D.Fructose	0.48	0.97	0.38	0.74	0.97	0.82
D.Glucuronic.acid	0.27	0.97	0.01	0.68	0.07	0.75
Dodecanoic.acid	0.65	0.34	0.54	0.85	0.89	0.81
d.Xylose	0.34	0.48	0.48	0.48	0.31	0.68
Eicosane	0.90	0.55	1.00	0.79	0.75	0.89
Galactose	0.77	0.92	0.48	0.91	0.34	0.18
Glucopyranose	1.00	0.97	0.03	0.48	0.25	1.00
Glutamic.acid	0.43	0.53	0.43	0.85	0.81	0.68
Glycine	0.71	0.25	1.00	0.06	0.42	1.00
Hexadecanoic.acid	0.77	0.65	0.29	0.06	0.31	0.75
Hexanoic.acid	0.59	0.05	0.57	0.20	0.89	0.29
D.Galactofuranose	0.43	0.21	0.01	0.68	0.09	0.68
Inositol	0.25	0.65	0.46	0.11	0.55	0.18
L.Alanine	0.30	0.70	0.96	0.31	0.97	0.55
L.Aspartic.acid	0.27	0.97	0.79	0.43	0.60	0.61
L.Glutamine	0.80	0.34	0.54	0.74	0.34	0.07
Linolenic.acid	0.77	0.86	0.86	0.43	1.00	1.00
L.Isoleucine	0.09	0.97	0.34	0.79	0.42	0.29
L.Leucine	0.13	0.36	0.25	0.62	0.97	0.85
L.Ornithine	0.65	0.42	0.79	0.48	0.28	0.07
L.Proline	0.15	0.42	0.60	0.25	0.75	0.96
L.Tyrosine	0.87	0.34	0.86	0.19	0.92	1.00
Maltose	0.09	0.23	0.72	0.10	0.49	0.74
Mannose	0.76	1.00	0.79	0.91	0.11	0.61
Monopalmitin	0.39	0.97	0.43	0.46	0.34	0.75
Myo.Inositol	0.84	0.75	0.19	0.31	0.70	0.82
L.Threonine	0.11	1.00	1.00	0.97	0.34	0.61
L.Phenylalanine	0.21	0.65	0.38	0.71	0.28	0.43
n.Butylamine	0.34	0.92	0.08	0.79	0.20	0.33
Nonanoic.acid	0.59	0.42	0.54	0.39	0.81	0.82
Norvaline	0.13	0.81	0.43	0.74	0.31	0.43
n.Pentadecanoic.acid	0.54	0.81	0.63	0.57	0.55	0.55
Octadecanoic.acid	0.84	0.55	0.93	0.48	0.86	0.55
Palmitoleic.acid	0.43	0.97	1.00	0.91	0.25	0.10

Phosphoric.acid	0.77	0.51	1.00	0.01	0.42	0.33
Lactic.acid	0.77	0.51	0.73	0.19	0.81	0.33
Rhamnose	0.03	0.25	0.96	0.16	0.28	0.21
Ribose	0.34	0.28	0.12	0.79	0.97	0.89
Serine	0.48	0.86	0.54	0.15	0.92	0.55
Sucrose	0.27	0.23	0.60	0.68	0.70	0.29
Tetradecanoic.acid	0.48	0.40	0.86	0.97	0.12	0.33
Threonic.acid	0.77	0.75	0.60	0.63	0.81	0.10
Threonine	0.30	0.31	0.22	0.48	0.53	0.55
Tyramine	0.51	0.81	0.22	0.97	0.33	0.10
Urea	0.39	0.86	0.93	0.39	0.38	0.38
Uridine	0.59	0.53	0.93	0.31	0.86	0.61
alpha.Ketoglutaric.acid	0.89	0.51	0.82	0.27	0.85	0.66
gamma.Aminobutyric.acid	1.00	0.75	0.34	0.97	0.92	0.75

In general the new results are in line with the results presented in the article. The false discovery rates in all age cohorts were very large (at least 0.50). This makes the presented results very unstable and in practice almost impossible to interpret. This fact is ignored in the interpretation of the results in the article.

4.4 Analyses related to seroconversion

The analyses related to seroconversion consisted of data divided into four timeframes based on time of seroconversion (18-9 months before, 9-0 months before, 0-9 months after and 9-18 months after seroconversion). Only some lipids and metabolites were selected for this analysis. The selected metabolites and lipids were analyzed as their own groups, so false discovery rates were estimated separately for metabolites and lipids.

4 different antibodies were used (IAA, GADA, ICA and IA-2A). In addition, a combined analysis was made, where the seroconversion was calculated from the first detected antibody.

The results presented in the article could not be fully repeated by 4Pharma. It is possible (or likely), that there are some differences in the definitions of the time frames (e.g. length of month), but it seems that this does not fully explain the differences. The full explanation for the differences is left unsolved. In general, larger p-values are achieved by 4Pharma, compared to the article results. However, in few cases also smaller p-values are achieved.

Assuming that the data has been correctly derived for lipidomics, the interpretations are solid combined with low FDRs and larger amount of subjects as in the metabolomics data.

For the metabolites the results presented in the article are suspicious. The data in some cases is extremely small, making the estimated group differences unstable. For some comparisons there are no significant differences detected (for example GADA – just before seroconversion). However, in this case the FDR is estimated at 0.60. This is technically impossible. As there are no significant differences detected, there can be no false discoveries either. These kind of clear errors raise a concern of the quality of the presented results.

As said earlier the interpretation of the results of metabolites is practically just guessing and the interpretations also with this analysis are biased.

4.5 The representativeness of the case report

One subject (ID=40), progressing to Type 1 diabetes had been selected to be reported as a case report. Selected metabolome and lipidomics data for this patient is presented in Figures 2 A, B and C in the article. According to information received from VTT this case report has been used also in many other occasions as a representative example of metabolomics data from a child developing to Type 1 diabetes. In this evaluation the question of interest was to investigate is this case report really a representative example of metabolomics in children developing to Type 1 diabetes.

The selected subject (ID=40) had two matched controls, who didn't progress to diabetes. Figures 3-12 summarize the same 10 variables as in the articles figure 2 for these three subjects.

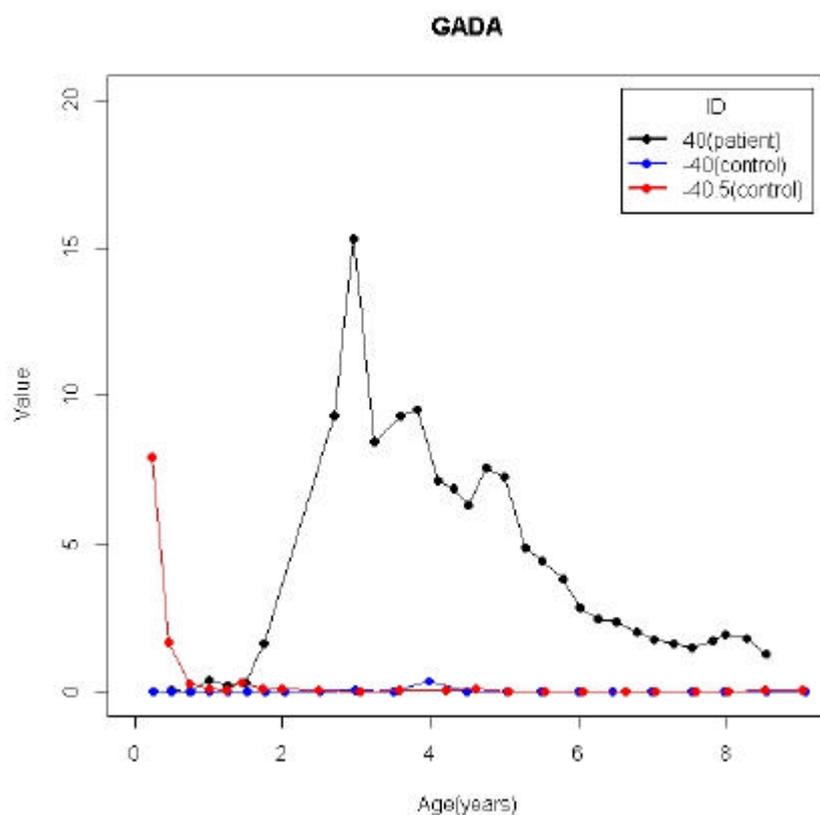


Figure 3. GADA values for the selected three subjects.

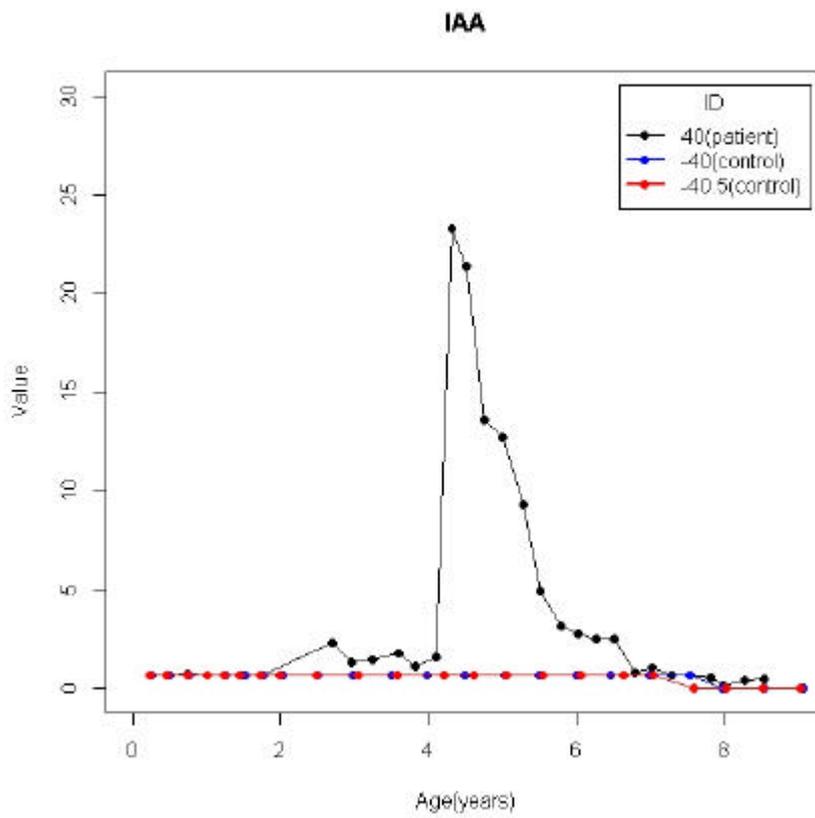


Figure 4. IAA values for the selected three subjects.

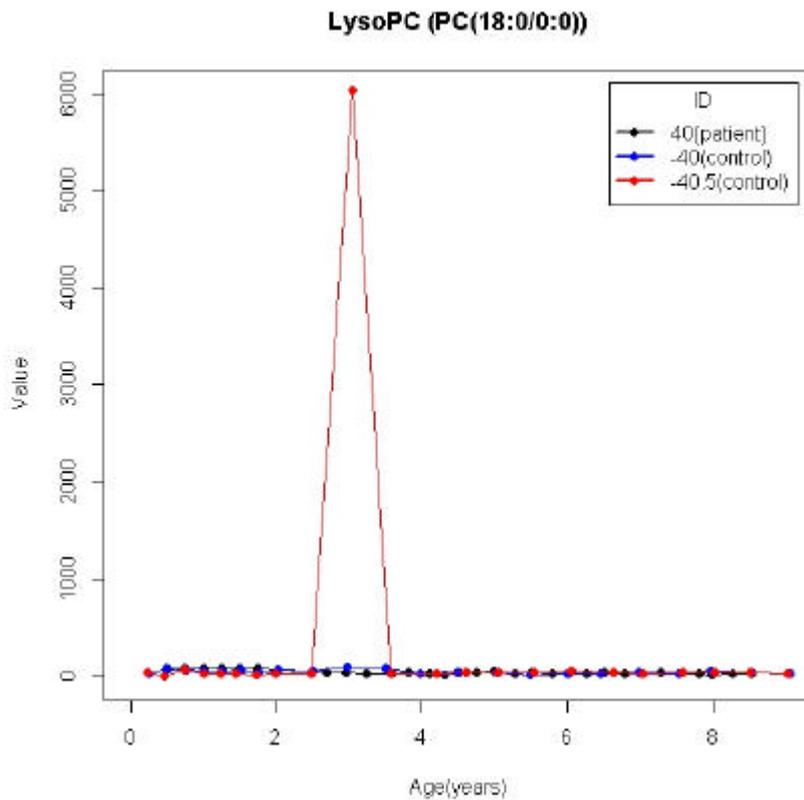


Figure 5a. LysoPC values for the selected three subjects.

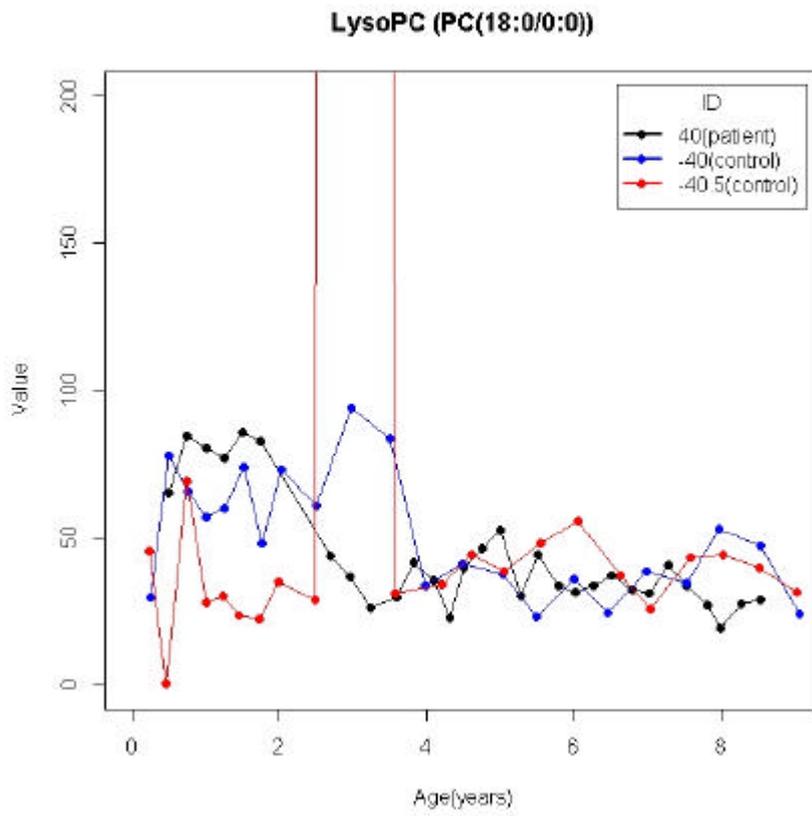


Figure 5b. LysoPC values for the selected three subjects (cutted y-axis).

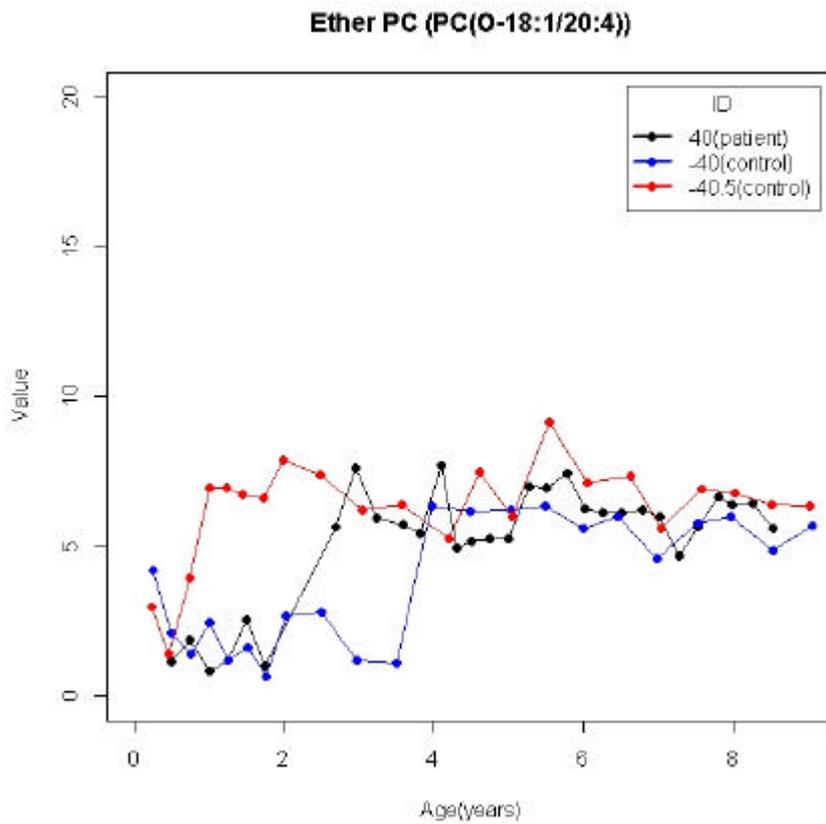


Figure 6. Ether PC values for the selected three subjects.

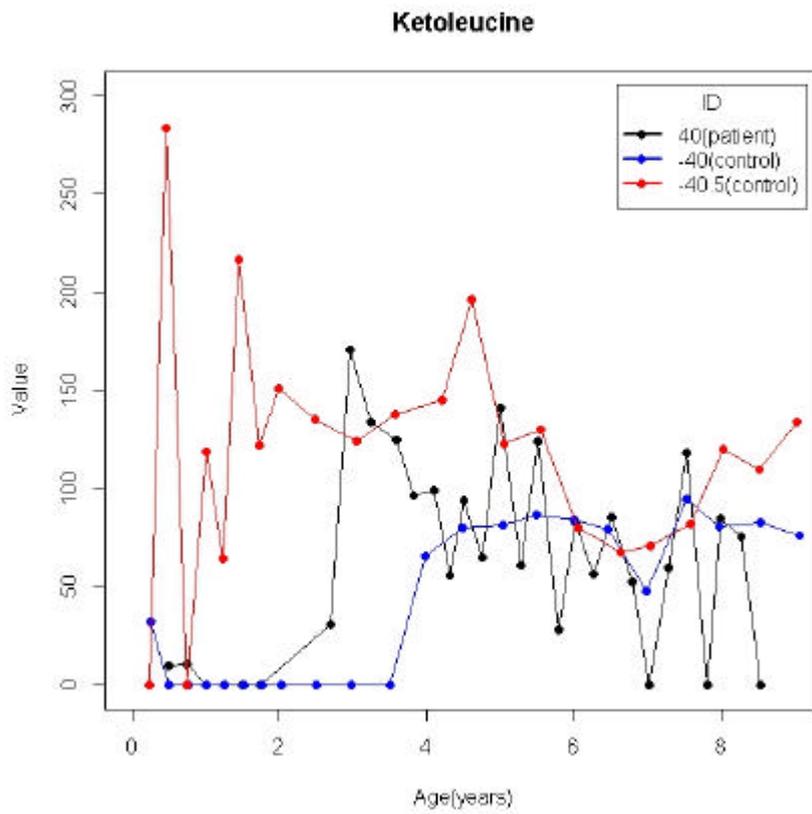


Figure 7. Ketoleucine values for the selected three subjects.

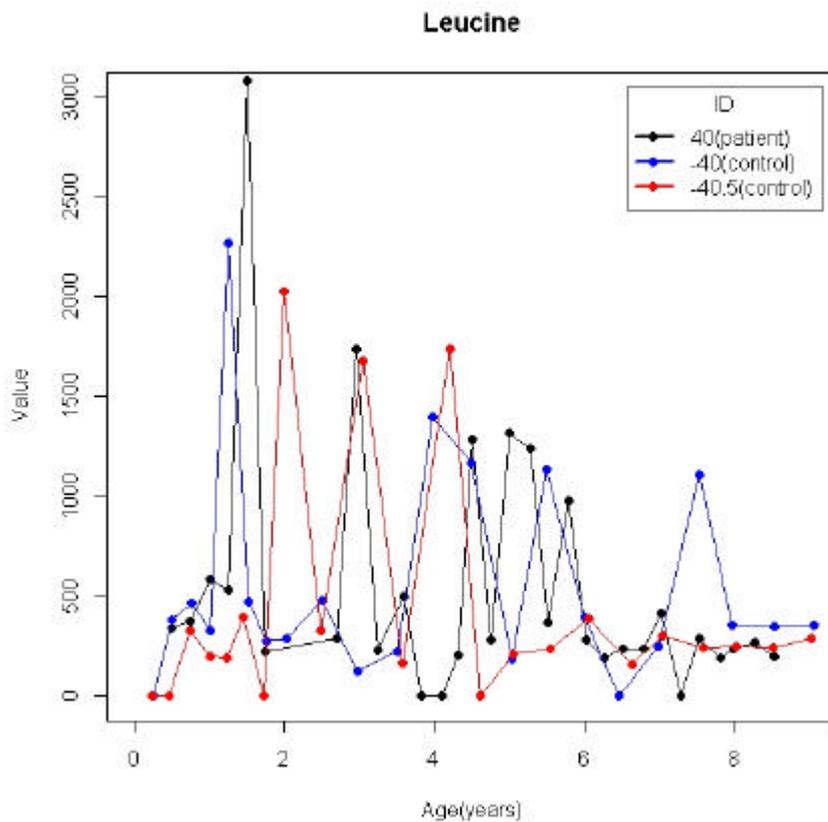


Figure 8. Leucine values for the selected three subjects.

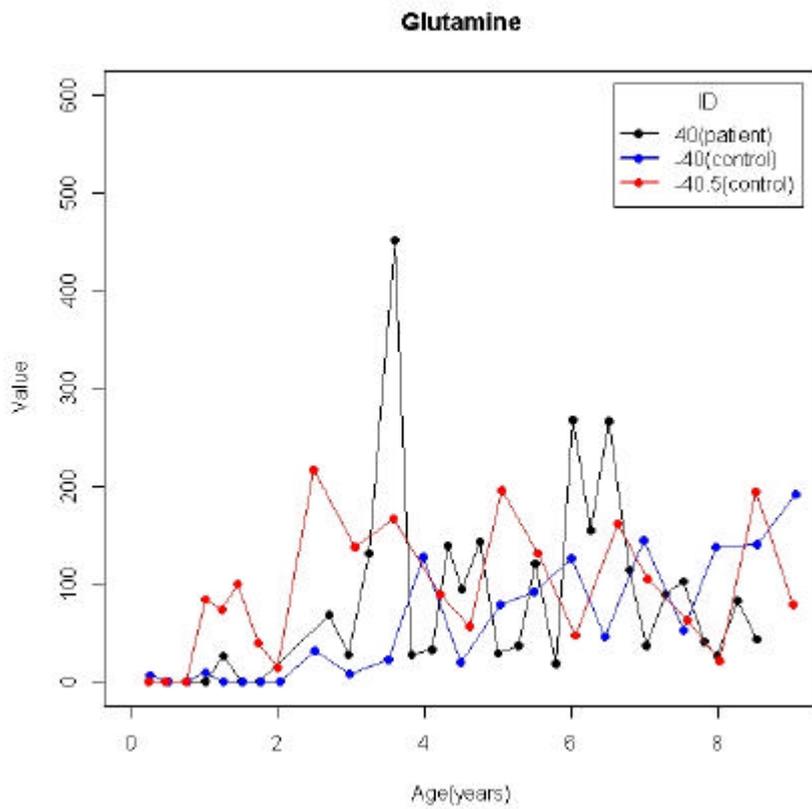


Figure 9. Glutamine values for the selected three subjects.

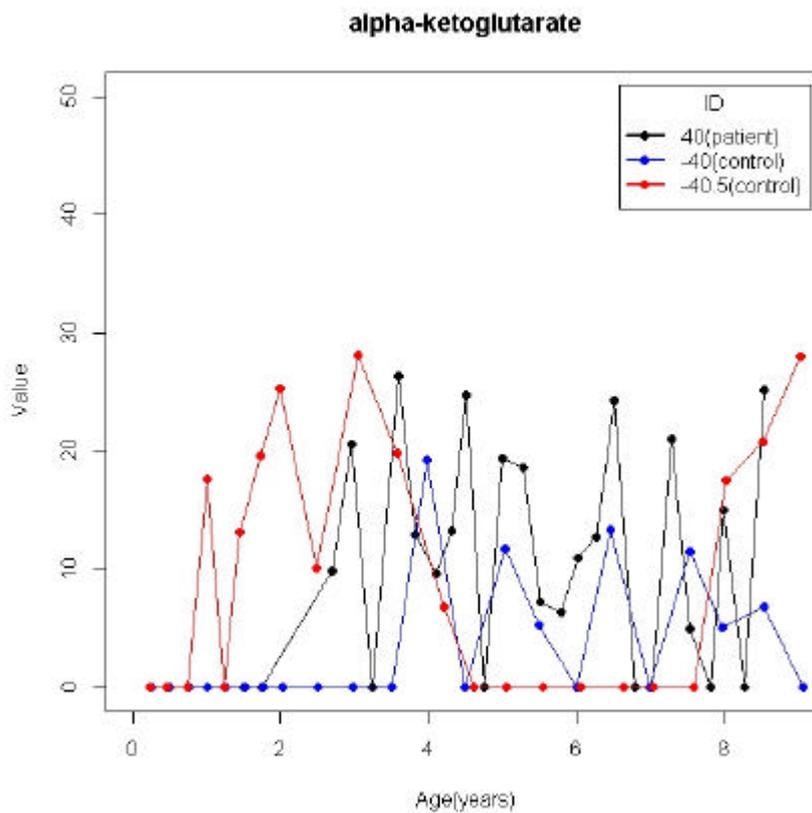


Figure 10. alpha-ketoglutarate values for the selected three subjects.

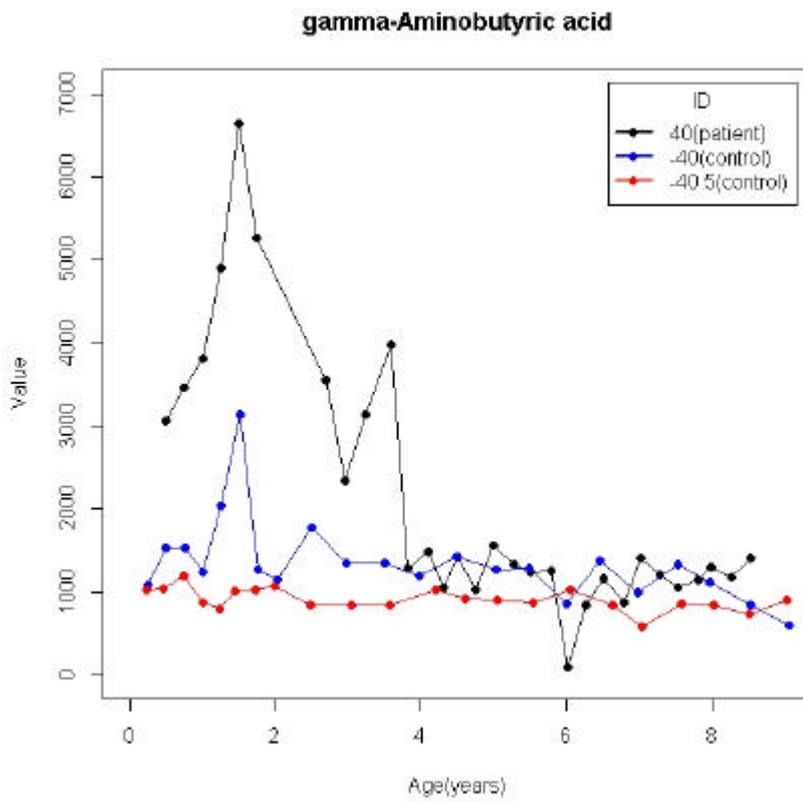


Figure 11. gamma -aminobutyric values for the selected three subjects.

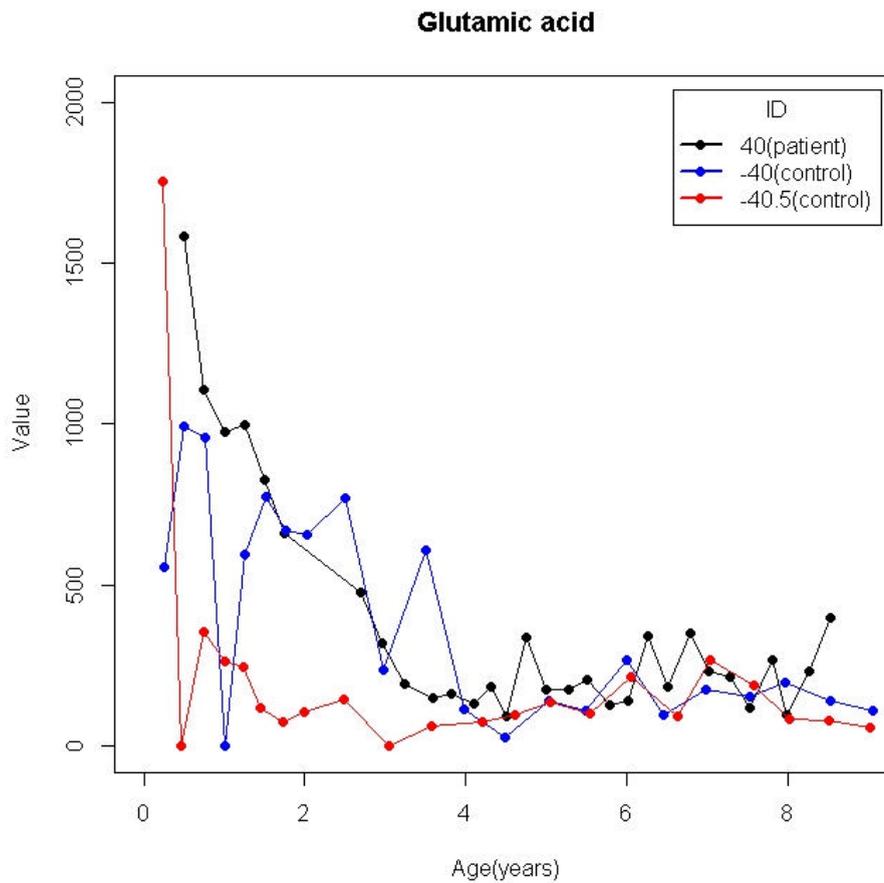


Figure 12. Glutamic acid values for the selected three subjects.

Based on the above figures, it seems that despite the group (Progressor/Control) quite large increases can occur in the selected metabolites. Similar peaks that are presented in the article as a signal of progressing to diabetes can be seen also with controls. With some metabolites even bigger increases could be seen with controls than with the progressor. Only with GABA a bigger increase is seen with the progressor than the non-progressors. Overall, based on the above curves for metabolites (figures 5-12), there is no clear difference between the progressor and its two matched controls.

The autoantibodies (GADA & IAA) are clearly different with the patient compared to the controls. Another way on investigating the representativeness of the case report is to look at the individual curves of the selected metabolites for all subject included in the study. This kind of evaluation is presented in Figures 13-22 of this document.

Lineplots by Subject (GADA)

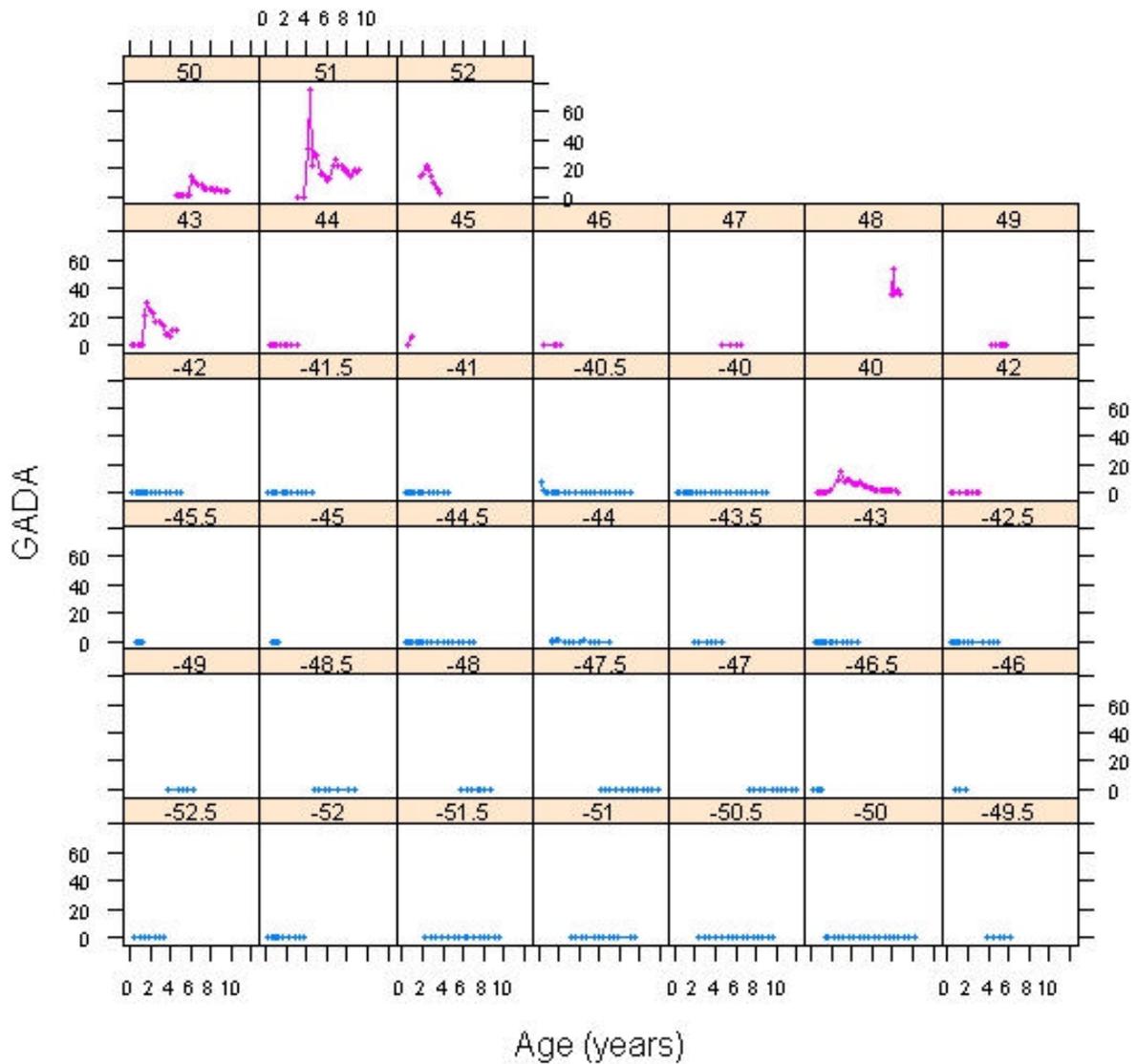


Figure 13. GADA values during the follow-up for all subjects.

Lineplots by Subject (IAA)

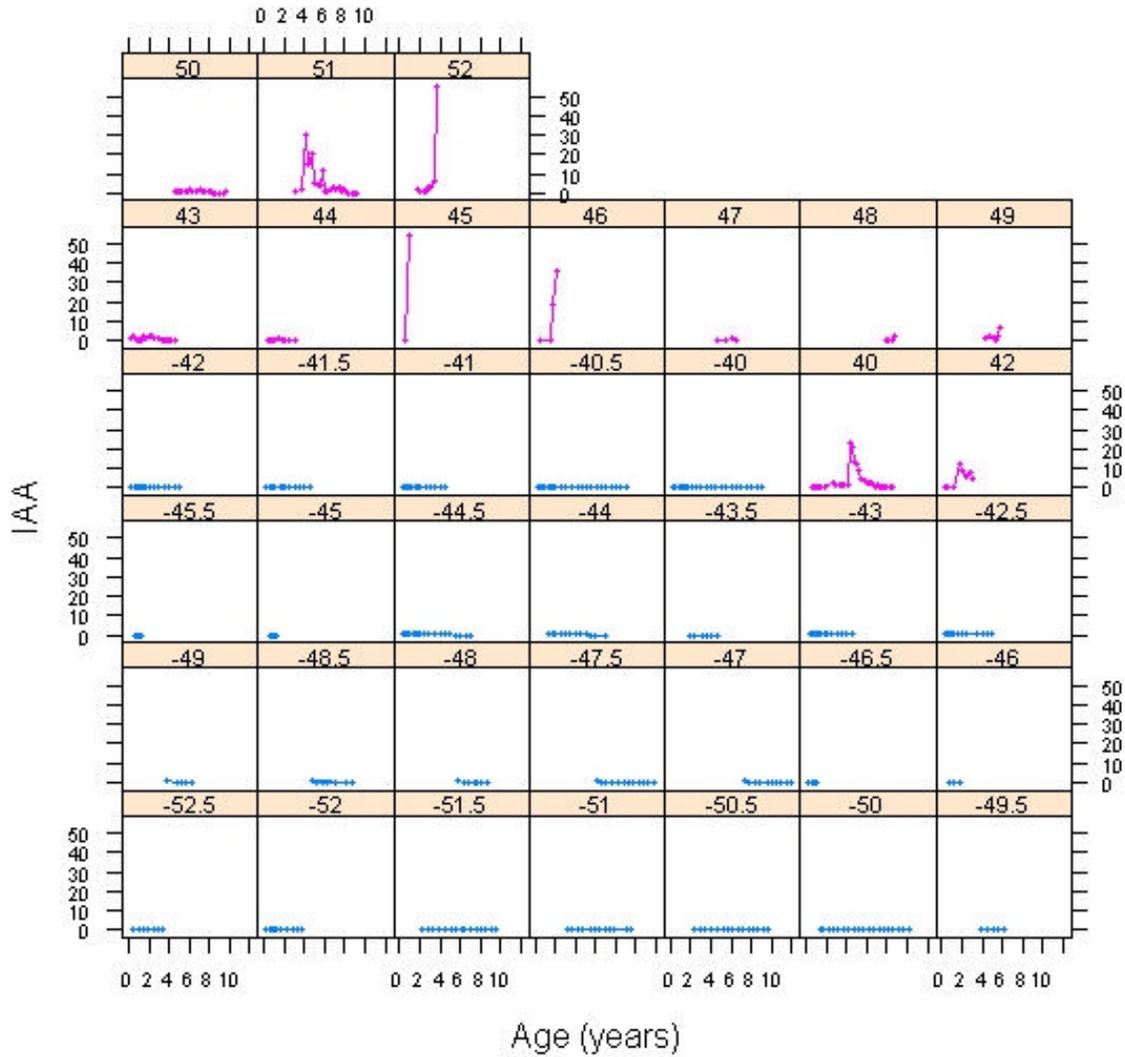


Figure 14. IAA values during the follow-up for all subjects.

Lineplots by Subject (LysoPC) [1]

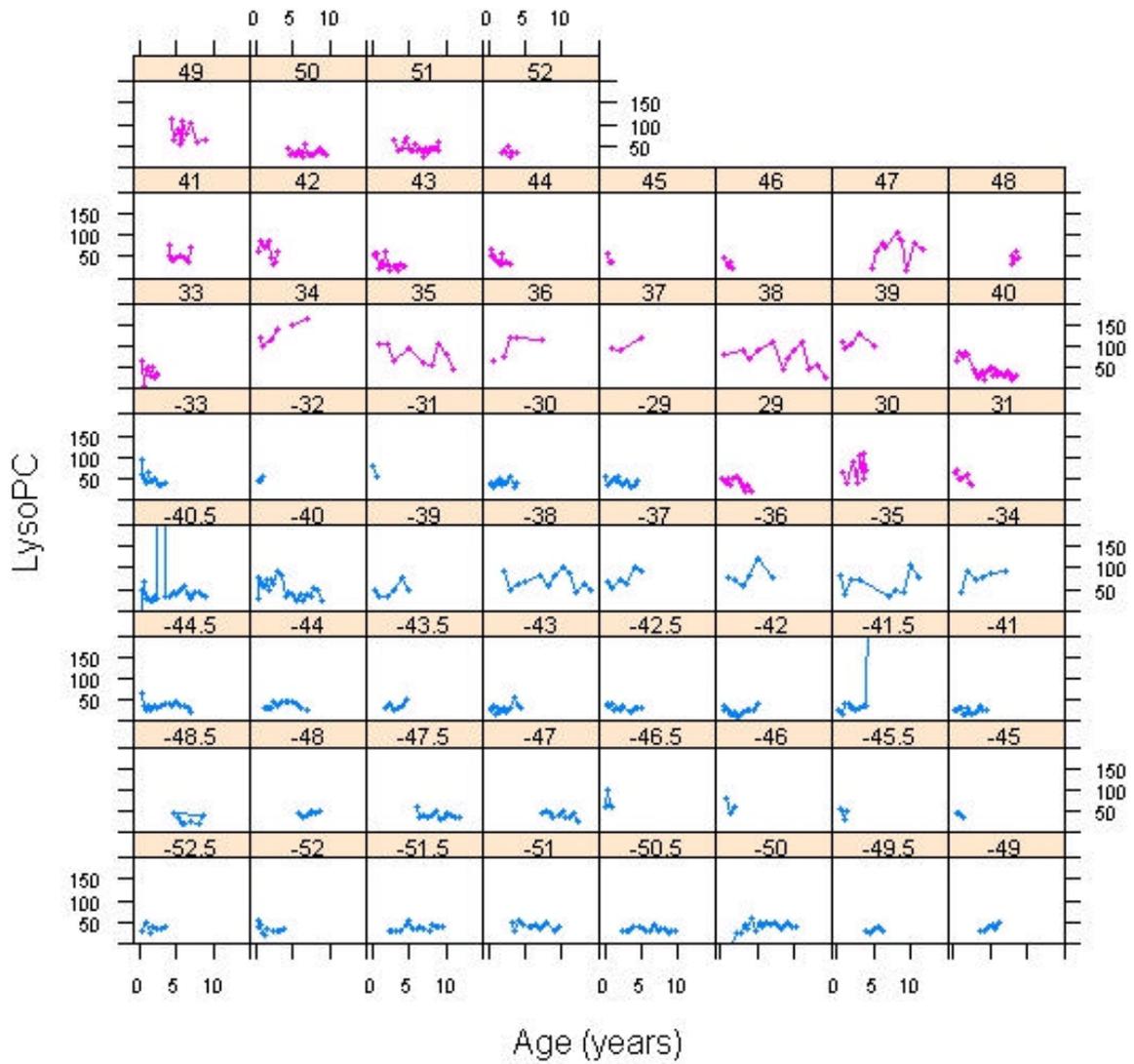


Figure 15a. Lyso PC values during the follow-up for all subjects.

Lineplots by Subject (LysoPC) [2]

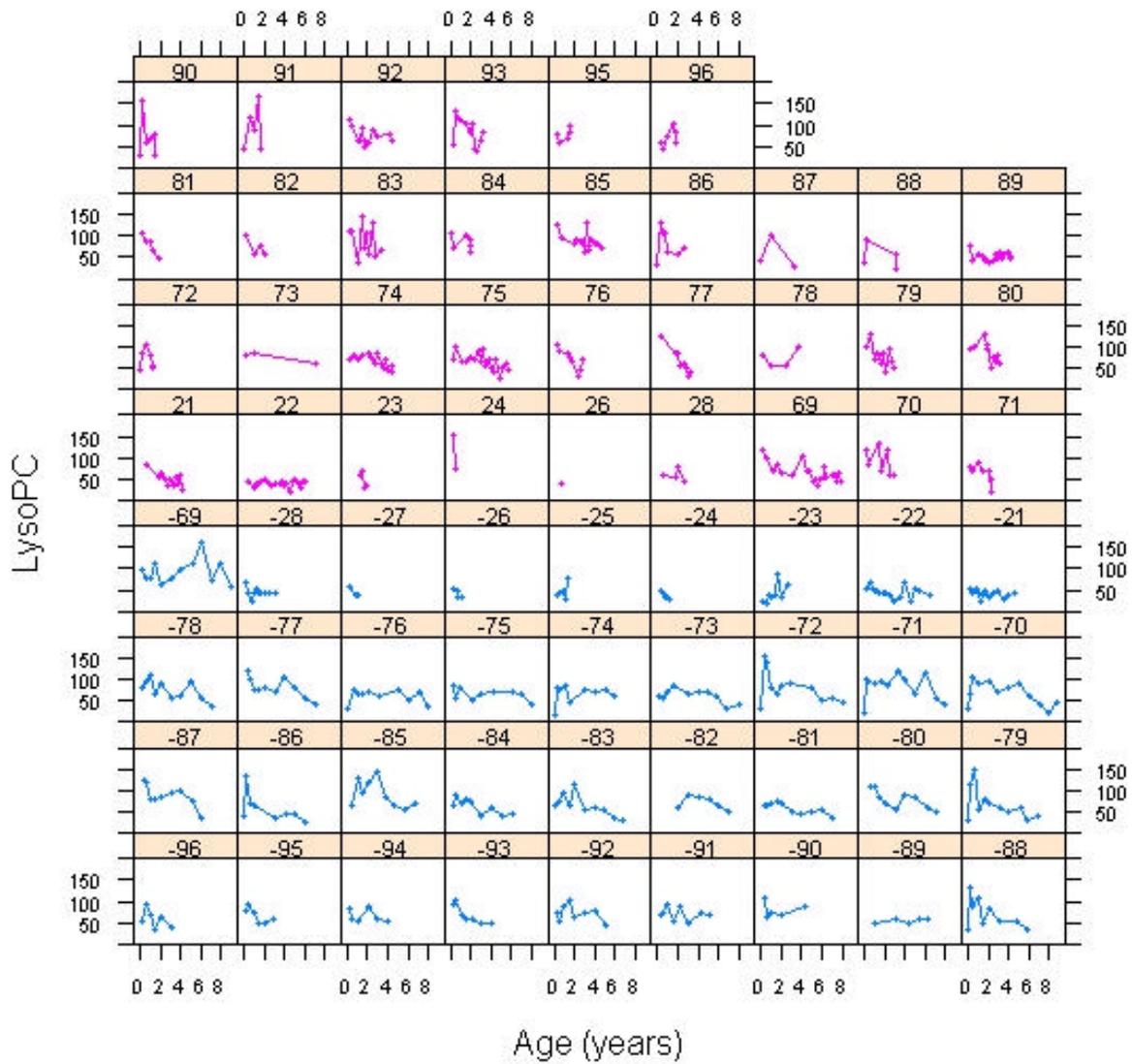


Figure 15b. Lyso PC values during the follow-up for all subjects.

Lineplots by Subject (EtherPC) [1]

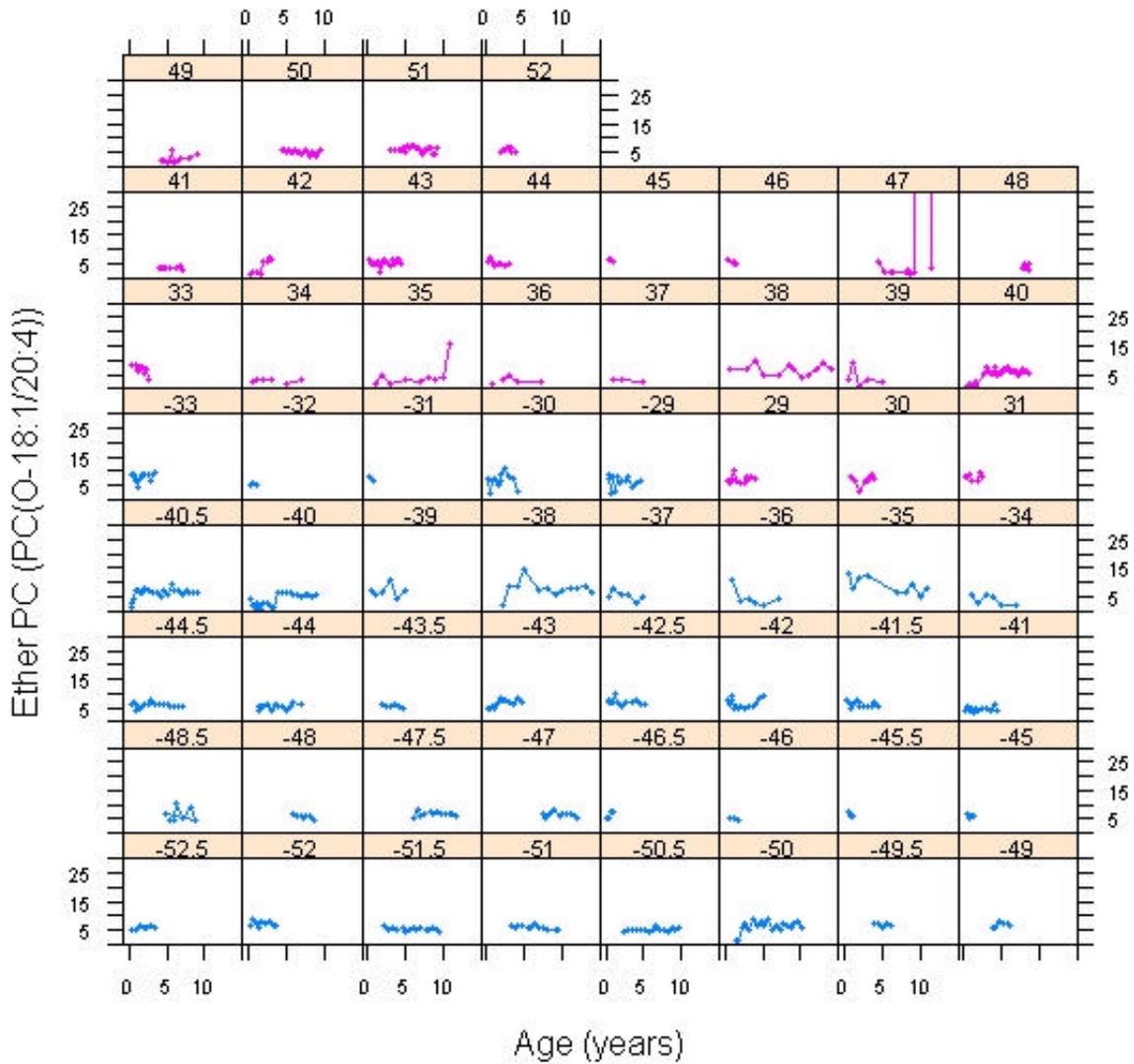


Figure 16a. Ether PC values during the follow-up for all subjects.

Lineplots by Subject (EtherPC) [2]

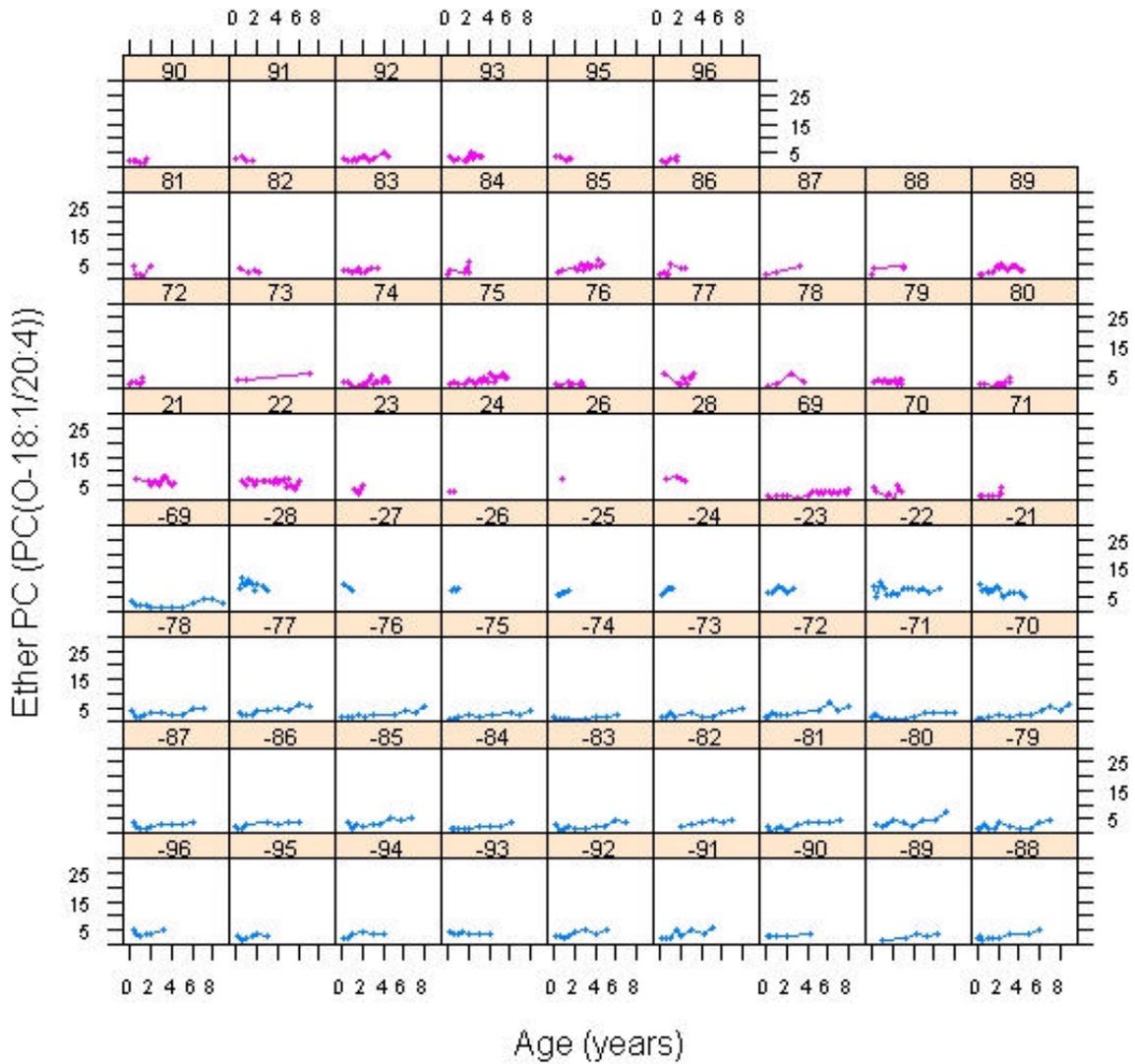


Figure 16b. Ether PC values during the follow-up for all subjects.

Lineplots by Subject (Ketoleucine)

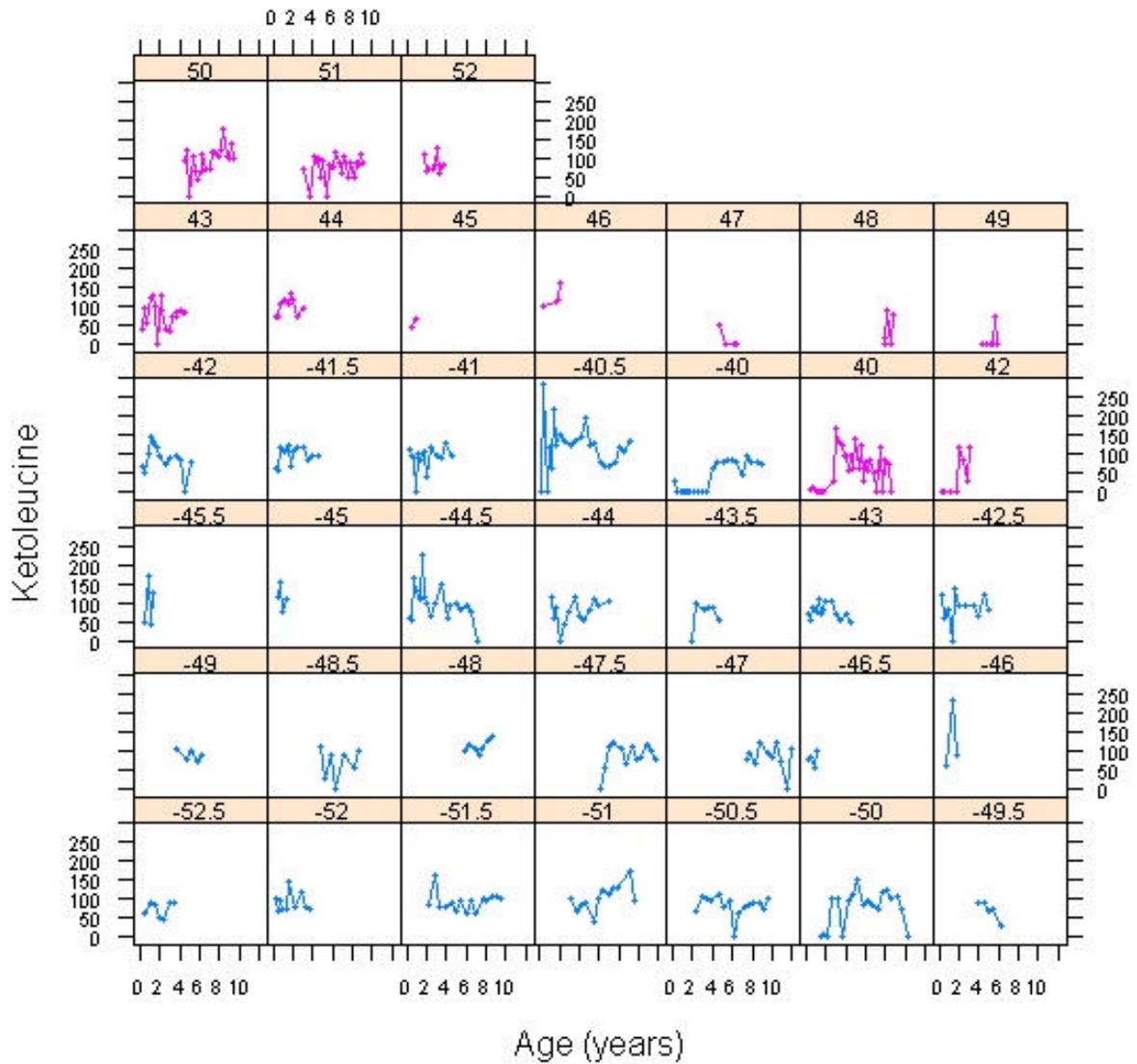


Figure 17. Ketoleucine values during the follow-up for all subjects.

Lineplots by Subject (Leucine)

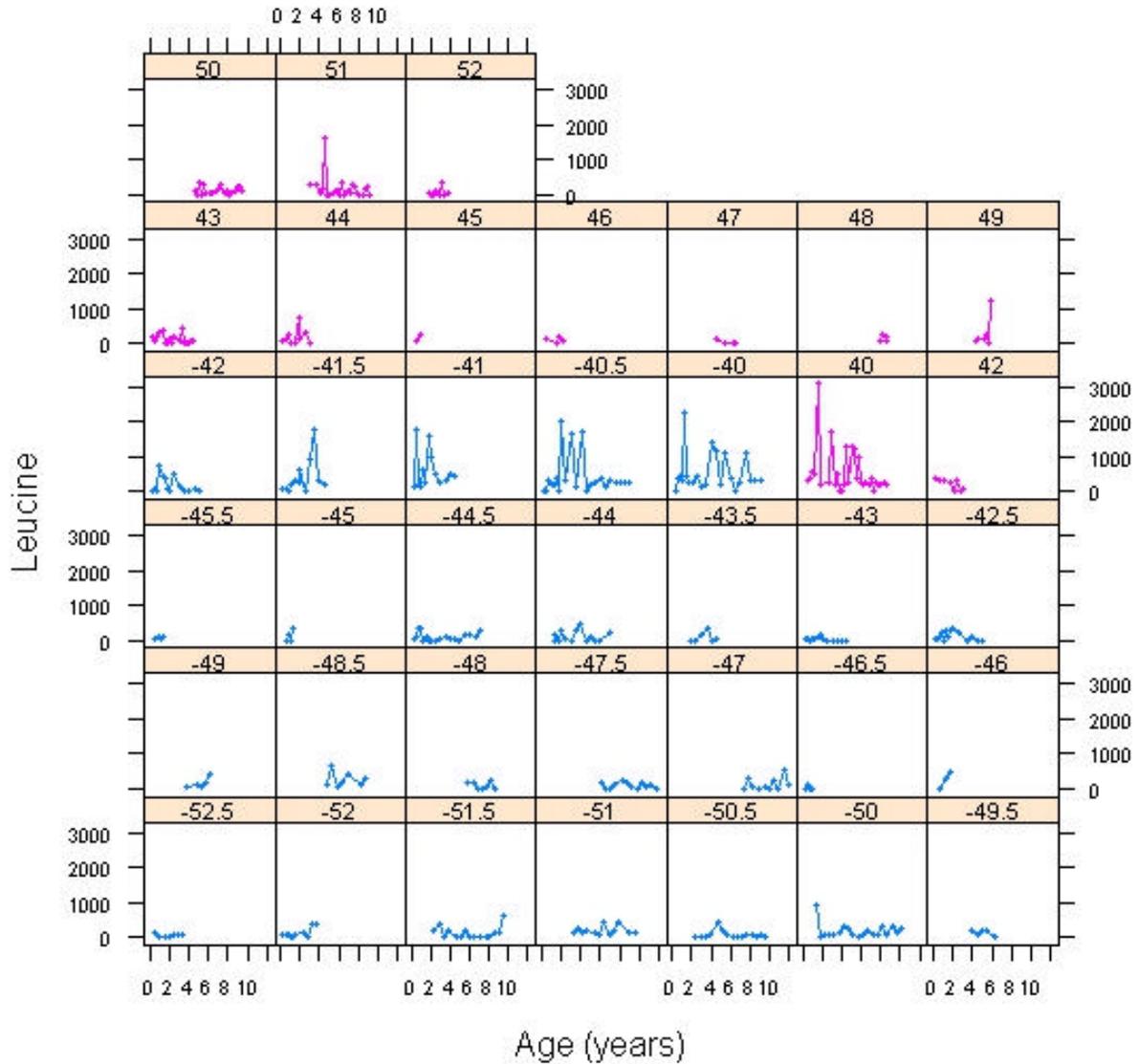


Figure 18. Leucine values during the follow-up for all subjects.

Lineplots by Subject (Glutamine)

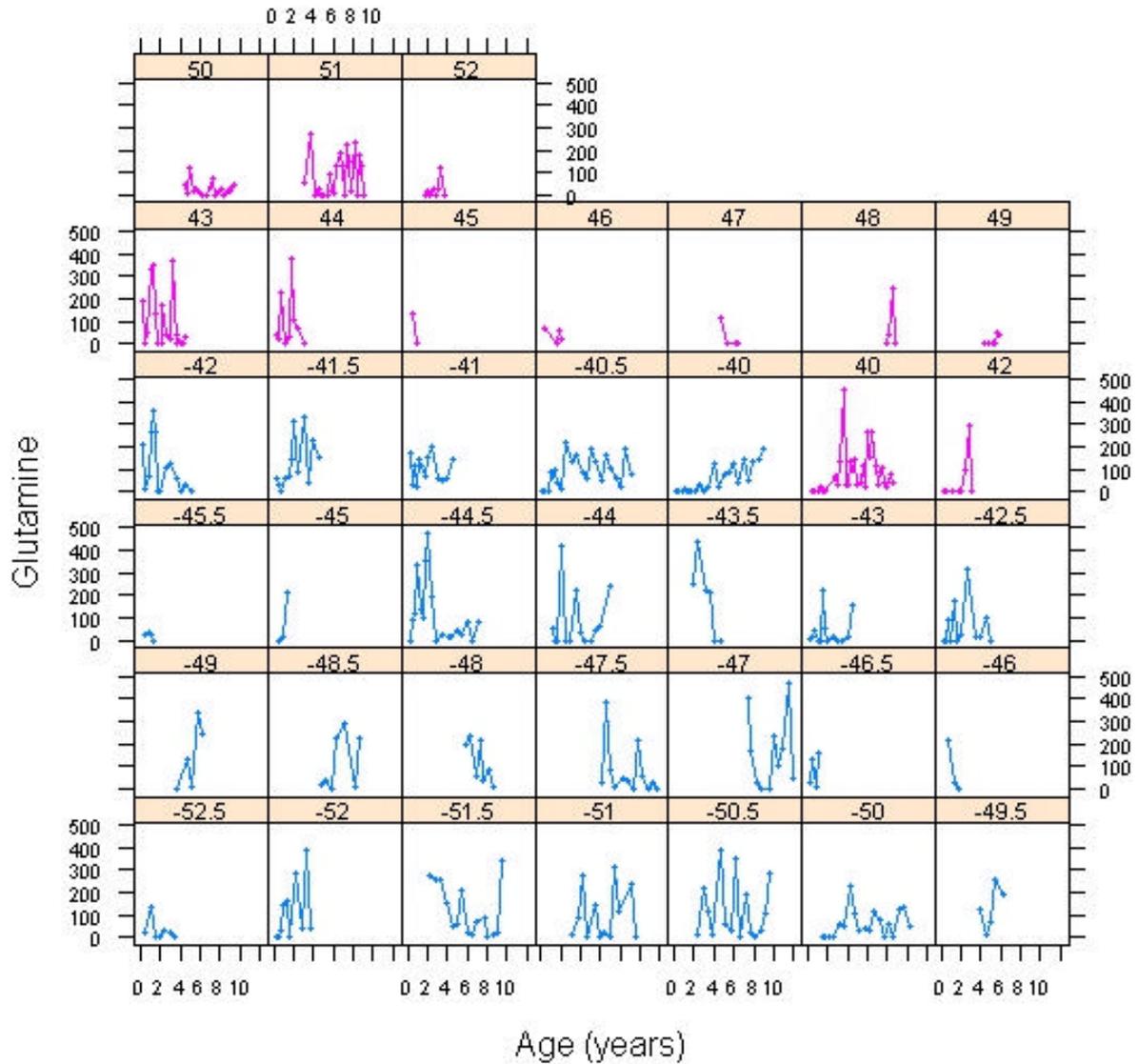


Figure 19. Glutamine values during the follow-up for all subjects.

Lineplots by Subject (alpha-ketoglutarate)

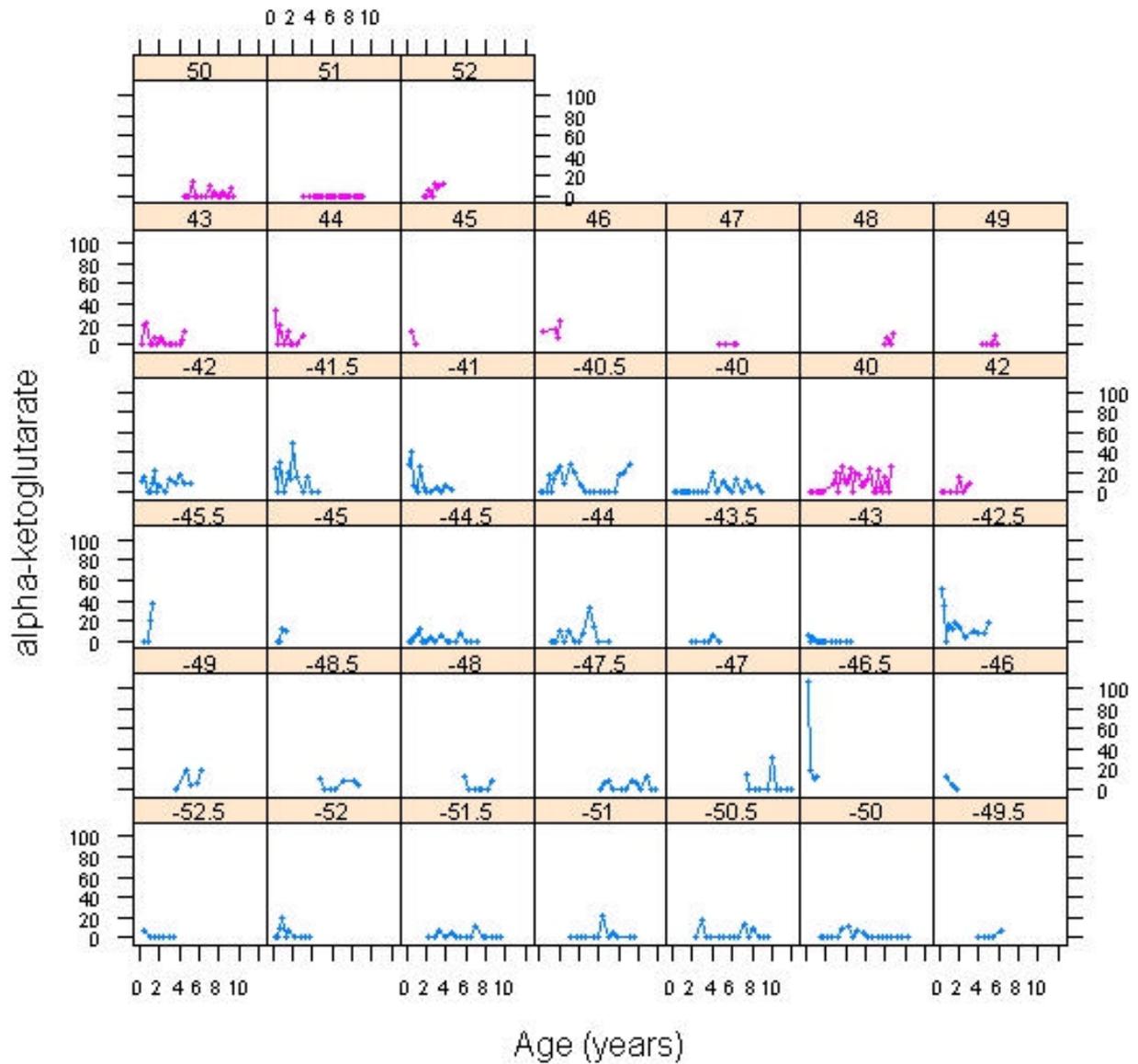


Figure 20. alpha-ketoglutarate values during the follow-up for all subjects.

Lineplots by Subject (gamma-Aminobutyric acid)

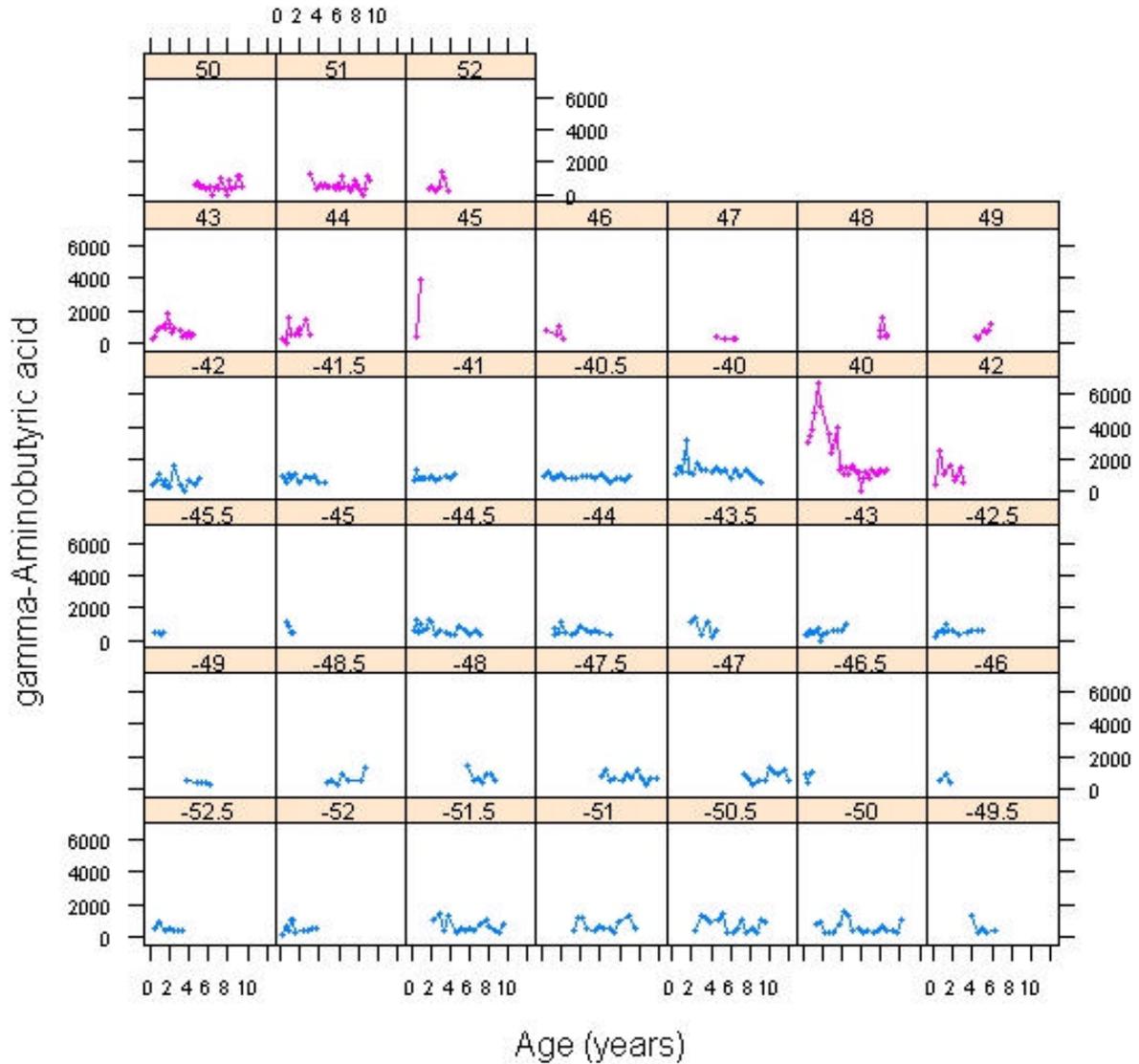


Figure 21. gamma-aminobutyric acid values during the follow-up for all subjects.

Lineplots by Subject (Glutamic acid)

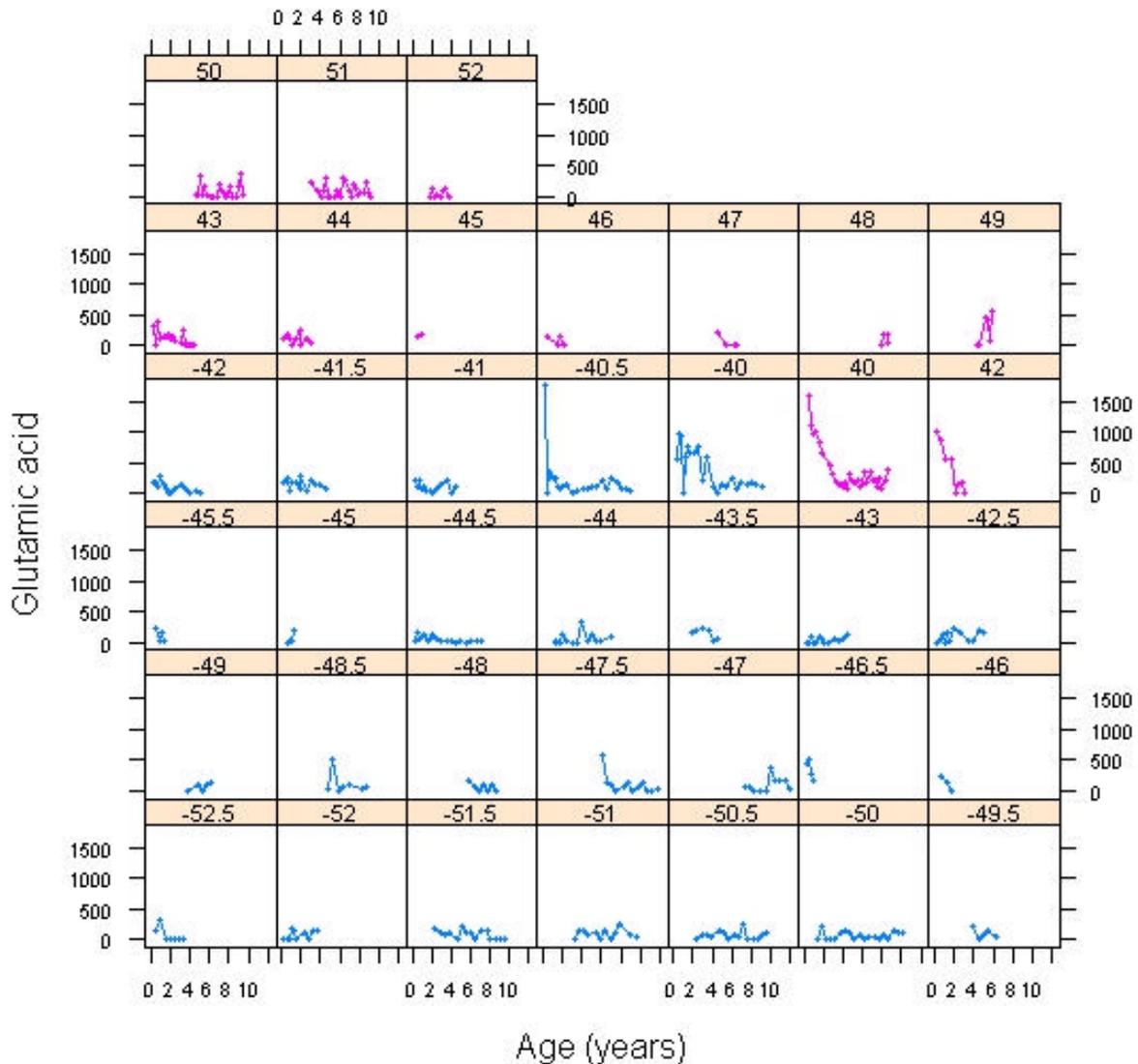


Figure 22. Glutamic acid values during the follow-up for all subjects.

The individual curves show that there is quite significant between-subject variation present in most of the selected metabolites. There seems to be no clear difference between the progressors and non-progressors in any of the selected metabolites. In other words the detected increases in the metabolite seems to be due to some other reason than later developing to type 1 diabetes. Naturally, the autoantibodies (IAA and GADA) are different between progressors and non-progressors.

5 Evaluation of the sufficiency of the included dataset

The abstract of the article states the following: “Serum metabolite profiles were compared between sample series drawn from 56 children who progressed to type 1 diabetes and 73 controls who remained nondiabetic and permanently autoantibody negative.” The included children have very varying length of follow-up. Thus, the number of children in some of the conducted comparisons can be clearly lower than the total number of included children.

With the lipidomics data the number of children decreases quite heavily after the first four years of lifespan, ending up to comparing 5 progressors against 10 non-progressors at the last age cohort (9-10 years). In other words we have data of 9-10 years olds for only about 10 % from all included children. This is clearly a big limitation of the study. Therefore e.g. the linear mixed effects models should have been analyzed also for a smaller age range to achieve more reliable results.

The metabolites data is very small to start with. Only 13 progressors and 26 nonprogressors were included in this dataset. However, the group comparisons include even smaller set of children. For example a maximum of 6 progressors were included in the comparisons by age cohort. The small sample added with the fact that the matching has not been taken into account in the analyses, makes all the statistical analyses related metabolites quite unstable and the interpretations less reliable. The influence of possibly only one abnormal progressor is large, so the results should be interpreted very cautiously. Especially, the results related to the metabolomics data collected after birth are not at all clear enough to be used in predicting the onset of diabetes on a individual level.

6 Conclusions

The article under evaluation had both good and bad qualities. Some of the presented results were repeated in the evaluation and seem to be on a solid basis. These include the analyses related to lipidomics. Both the Wilcoxon tests and the linear mixed effects modelling seem to be correct and the interpretations are in line with the results of statistical analyses.

The biggest concerns raised during the evaluation include

- The omission of the high FDR- values (metabolites (follow-up data) and lipidomics(at birth)). The authors state, that multiplicity has been taken into consideration, but still report and interpret the raw p-values and totally ignore that the FDR-values in some cases are very high, making the interpretations based on raw p-values unreliable.
- The presentation of the subject no 40 as a representative sample of the behaviour of a child who later develop to type 1 diabetes. Even though this has not been that strongly stated in the article, it has notified by VTT that this has been done in several other occasions. This evaluation report shows that in general the subject 40 does not clearly differ from its matched controls, nor there is a trend difference between the progressors and non-progressors overall.
- The analysis results related to seroconversion are suspicious (metabolites). It is noted, that there might be some differences in derivation of the data, but there are still definite errors in the presented results. Even if the results would be assumed to be correct, they are still interpreted falsely. Every raw p-value < 0.05 is interpreted as a significant difference, even though the FDR in this case as well are very high.

The statistical methodology used in the article is a bit low powered in some cases. However, this is not a major issue and if anything should flatten the results.

The main concerns brought up in this evaluation are issues that could have been detected by a sophisticated statistical review (from the Journal reviewers). The only issues that possibly could be a sign of some kind of deception are the analyses related to seroconversion for metabolites. However, this possible data issue could not be investigated further, as the derived data for this analysis is not accessible from the article.

7 Execution of statistical analyses

Statistical analyses and evaluations were performed by 4Pharma Ltd.

8 Hardware and software

Statistical analyses were performed with R for Windows[®] version 3.0.2 (R Foundation for Statistical Computing, Vienna, Austria) and with SAS[®] System for Windows, version 9.3 (SAS Institute Inc., Cary, NC, USA).

.