A metabolomics and lipidomics study of mouse models of type 1 diabetes highlights divergent metabolism in purine and tryptophan metabolism prior to disease on-set

Steven A. Murfitt[1], Paola Zaccone[2], Xinzhu Wang[1], Animesh Acharjee[3], Yvonne Sawyer[2], Albert Koulman[3], Lee D. Roberts[3], Anne Cooke[2]., & Julian Leether Griffin[1,3*].

1. The Sanger Building, 80 Tennis Court Road, Department of Biochemistry, Cambridge, CB2 1GA, UK.

2. Department of Pathology, Tennis Court Rd, Cambridge CB21QP, UK.
3. Medical Research Council Human Nutrition Research, The Elsie Widdowson Laboratory, 120 Fulbourn Road, Cambridge, UK.

*corresponding author. Email: jlg40@cam.ac.uk. Tel. +44-1223 764922.

Abstract:

With the increase in incidence of type 1 diabetes (T1DM) there is an urgent need to understand the early molecular and metabolic alterations that accompany the autoimmune disease. This is not least because in murine models early intervention can prevent the development of disease. We have applied a liquid chromatography (LC-) and gas chromatography (GC-) mass spectrometry (MS) metabolomics and lipidomics analysis of blood plasma and pancreas tissue to follow the progression of disease in three models related to autoimmune diabetes: the non-obese diabetic (NOD) mouse, susceptible to the development of autoimmune diabetes, and the NOD-E (transgenic NOD mice that express the I-E heterodimer of the major histocompatibility complex II) and NOD- severe combined immunodeficiency (SCID) mouse strains, two models protected from the development of diabetes. All three analyses highlighted the metabolic differences between the NOD-SCID mouse and the other two strains, regardless of diabetic status indicating that NOD-SCID mice are poor controls for metabolic changes in NOD mice. Comparing NOD and NOD-E mice we show the development of T1DM in NOD mice is associated with changes in lipid, purine and tryptophan metabolism, including an increase in kynurenic acid and a decrease in lysophospholipids, metabolites previously associated with inflammation.

Introduction:

Type 1 diabetes (T1DM) is increasing in prevalence across the globe, particularly in westernised countries with incidences of 8-17 per 100,000 people in Northern Europe and the US, and rising to 35 per 100,000 in Scandinavia [1,2]. The disease arises following the destruction of beta cells in the pancreas, resulting from an autoimmune response, with sufferers having autoantibodies to the 65-kDa isoform of glutamic acid decarboxylase, zinc transporter proteins, phosphatase- related IA-2 molecule, insulin and islet cells [3]. While not everyone with autoantibodies will progress to T1DM, the likelihood increases with the number of autoantibodies with 3-4 antibody types increasing the risk to ~80% [3].

There is an important need for early stage biomarkers of the disease progression in T1DM in order to better understand the early aetiology of the disease and potentially target immunosuppression therapy to prevent the further destruction of pancreatic beta cells. Furthermore, there is significant evidence that immunomodulation by infection can be used to prevent T1DM if targeted to the right clinical window of autoimmunity [4,5]. While autoantibodies do discriminate between type 1 diabetics and healthy individuals, autoantibodies are relatively late stage changes and do not always lead to beta cell destruction [3], confounding their use as reliable biomarkers

Metabolomics, the study of the metabolite complement of a biofluid, cell, tissue or whole organism [6] has been previously used to follow both early and late stage changes in T1DM. Drug induced T1DM using streptozotocin induces a variety of metabolic changes in rodents including metabolites in the citric acid cycle, glucose metabolism, choline turnover and amino acid metabolism [7]. These pathways have also been confirmed to be altered in humans with T1DM, and in particular insulin treatment has also been associated with protein turnover/amino acid metabolism and ketosis in healthy individuals in human intervention studies [8,9]. In addition metabolomics has been used to discriminate type 1 diabetics with early stage kidney disease from those without[10-12]. However, these are late changes in the pathogenesis of T1DM and are unlikely to provide biomarkers of the pre-diabetic state.

A sub-set of the field of metabolomics is lipidomics, which studies the lipid component of biofluids, cells or tissues. Lipidomics has been transformed by recent developments in liquid chromatography mass spectrometry (LC-MS) which allows the rapid analysis of intact lipids from a range of chemical classes including triglycerols, diglycerols, phospholipids, free fatty acids, cholesterol and cholesterol esters and sphingolipids [13]. Metabolomics and lipidomics have been used to also follow early stage disease progression in both mouse models and

human sufferers with T1DM [14-18]. Examining serum from children who subsequently went on to develop T1DM, Oresic and colleagues [14] identified a number of early stage metabolic changes, including reduced concentrations of succinate, phosphatidylcholines and triglycerides at birth. In addition, prior to serum conversion in terms of autoantibodies, individuals had increased pro-inflammatory lysophosphocholines, and reduced concentrations of triglycerides and anti-oxidant ether lipids. Metabolic changes have even been detected in the cord blood of babies that subsequently developed T1DM before the age of 8 [17].

In the present study we have profiled metabolic and lipidomic changes in a mouse model of autoimmune diabetes. The non-obese diabetic (NOD) inbred mouse strain recapitulates the autoimmune nature of T1DM in humans and has allowed researchers to follow the autoimmune destruction of beta cells in the pancreas as well as develop treatments that halt this process [19]. The mouse develops islet infiltration (insulitis) at around 3-4 weeks comprising initially of innate immune cells such as macrophages and dendritic cells followed by T cells and other cells of the immune system[20-23 22-24]. The mice develop diabetes from around 14 weeks of age. This mouse model has been used previously to monitor metabolic changes that accompany the progression of T1DM [16], but in this previous study only NOD mice were considered, and thus, it is difficult to determine what changes were associated with disease progression and what was associated with age-related changes in the pancreas. In the present study we have compared metabolic changes in NOD mice with their transgenic control strain NOD-E mice. The NOD-E mouse is a transgenic mouse generated directly on the NOD background which, unlike the NOD mouse, expresses the major histocompatibility complex (MHC) class II I-E heterodimer [25]. These mice are protected from the development of T1DM and develop negligible insulitis. In the present study we show by metabolomics that the NOD-E mouse is a better control for the metabolic changes associated with the NOD mouse compared with the NOD severe combined immunodeficiency (SCID) mouse strain [26]. [26]. As the NOD-SCID mouse was generated through breeding the SCID mutation onto the NOD mouse background using donor SCID mice from a different genetic background there may be contributions from non-NOD genes in this strain. In addition NOD and NOD-E mice have distinct metabolic trajectories during the ageing process, while aged NOD mice that do not develop T1DM develop a metabolic phenotype more similar to NOD-E mice than NOD mice.

Experimental Procedures:

Animal Experiments:

4

All mice were obtained from stable colonies housed within the Department of Pathology, University of Cambridge. The incidence of diabetes in NOD female mice in this colony is around 80%. Further details as the progression of the disease process are found in the supplementary section. Mice were housed in a temperature- and humidity-controlled facility, with a 12-h:12-h light-dark cycle and access to water and regular chow diet (Caloric content: 11.5% fat, 26.9% protein, 61.6% carbohydrate; RM1; Special Diet Services, UK) *ad libitum*. Blood plasma was collected by retro-orbital bleed into heparinised vials prior to centrifugation. The pancreas was rapidly dissected and a portion of each tissue was snap frozen with liquid nitrogen and stored at -80 $^{\circ}$C until further analysis. All animal work was carried out in accordance with UK Home Office project and personal licences (80/2442 and 70/8442) and was approved by the Ethical Review Committee of the University of Cambridge.

Two studies were performed. The first examined blood plasma and pancreatic tissue taken post mortem from stable colonies of female NOD, NOD-E and NOD-SCID mice (**Supplementary Table 1**). In the second study blood plasma was sampled at six time points (6, 9, 11, 14, 19, and 22 weeks) for female NOD (n=9), NOD-E (n=7) and NOD-SCID mice (n=10) (**Supplementary Table 2**). In this study pancreas tissue was taken at the end of the study.

Analysis of the metabolite composition of faeces
To assess changes in the gut metabolome mouse faeces were collected from individual NOD (n=5 for 6 weeks age and n=6 each for 12 and 18 weeks age) and NOD-E mice (n=5 for 6 and 12 weeks age). 20 mg of faeces were extracted using the chloroform/methanol extraction procedure described below.

Metabolite extraction from the pancreas, blood plasma and faeces
Metabolites (50 µl plasma or 50 mg tissue wet weight) were extracted using a modified Bligh and Dyer method [27]. 600 µl of methanol:chloroform (2:1 v/v) was added to each sample, with tissue samples then being pulverized/homogenised using a TissueLyser (Qiagen, Hilden, Germany) set at 20 Hz for 20 minutes and sonicated for 15 minutes. 200 µl of chloroform and 200 µl of H_2O (Chromasolv® Plus for HPLC, Sigma-Aldrich, Gillingham, UK) were added, vortex mixed, and centrifuged at 17,000 rcf for 15 minutes. The aqueous and organic

layers were then separated. The procedure was repeated to form a double extract to maximise the recovery of metabolites from the protein pellet. The aqueous layer was dried overnight in an evacuated centrifuge (Eppendorf, Hamburg, Germany), while the lipid fraction was dried under a stream of nitrogen gas. Dried samples were stored at -80 °C until analysis.

Lipidomics analysis of blood plasma and pancreas tissue

All solvents used were of LC-MS grade or better and ordered from Sigma Aldrich (Gillingham, UK). All internal standards were obtained from Avanti Polar Lipids (Alabaster, AL, USA) with the exception of undecanoic acid and trilaurin (Sigma Aldrich).

The samples, QC samples (pools of either the plasma and pancreas extracts) and blanks were placed in a pre-defined random order across 96-well plates (Plate+, Esslab, Hadleigh, UK). The extracts were reconstituted in 100 µl of 7.5 mM NH_4Ac IPA:MeOH (2:1), spiked with six internal standards (1,2-di-o-octadecyl-sn-glycero-3-phosphocoline, 1,2-di-O-phytanyl-sn-glycero-3-phosphoethanolamine, C8-ceramide, N-heptadecanoyl-D-erythro-sphingosylphosporylcholine, undecanoic acid, and trilaurin) which were used to correct for variations in m/z across the dataset acquired.

Lipidomics was performed on the extract using chip-based nanospray with an Advion TriVersa Nanomate (Advion, Ithaca, USA) interfaced to the Thermo Exactive Orbitrap (Thermo Scientific, Hemel Hampstead, UK), using a mass acquisition window from 200 to 2000 m/z and acquisition in positive mode. A voltage of 1.2 kV was used in positive mode with an acquisition time of 72 s.

Acquired spectral raw data were processed using an in-house bioinformatics platform based on XCMS [28]. This performed sample-specific mass re-calibration using predefined sets of internal standards and the removal of commonly present contaminant ions (often associated with plasticizers) with help of predefined rejection lists and mass defect filters. The raw data were converted to .mzXML (using MSconvert [29] with peakpick level 1), parsed with R and 50 (scan from 20 to 70) spectra were averaged per sample using XCMS [28], with a signal cutoff at 2000. The files were aligned using the XCMS [28,30] grouping function using "mzClust" with a m/z-window of 22 ppm and a minimum coverage of 60 %. Isotopes were annotated using the CAMERA package in R with the following parameters (maxcharge=1, ppm=5, mzabs=0.001, intval = c("into"), minfrac=0.25) [30].

Automated compound annotation was carried out using both an exact mass search in compound libraries as well as applying the referenced Kendrick mass defect approach. Signal normalisation was performed by summing the intensities of all detected metabolites to a fixed value to produce a correction factor for the efficiency of ionisation. Features of interest were subsequently confirmed using fragmentation experiments on a Thermo Velos Orbitrap mass analyser using an Advion Nanomate to directly infuse the lipid extract or a Surveyor HPLC system for chromatographic separation. The selected masses were isolated with a 1.5 m/z width in the linear ion trap and then fragmented using either linear ion trap with 35% relative collision energy or in the HCD (higher-energy collision-induced dissociation) collision cell, with a range of collision energies from 5% to 75% relative collision energy. All spectra were recorded in the Orbitrap set at 100,000 resolution. As fragmentation data confirms the fatty acids that are present but does not definitively define where the double bonds are placed these annotations are at level 2 of the Metabolomics Standards Initiative [31].

Targeted analysis of aqueous metabolites in blood plasma, pancreas and faeces

All analyses were carried out using a QTRAP 4000 quadrupole mass spectrometer (AB Sciex, Warrington, UK) coupled to an Acquity ultra performance liquid chromatography (UPLC) system from Waters Ltd. (Atlas Park, Manchester, UK) according to published methods [32,33]. In brief, formic acid, ammonium acetate, and valine-d8 were purchased from Sigma-Aldrich. Plasma samples were prepared for LC-MS analyses via protein precipitation with the addition of nine volumes of 74.9:24.9:0.2 vol/vol/vol acetonitrile/methanol/formic acid containing stable isotope-labeled valine-d8 internal standard. The valine-d8 standard was used to calculate concentrations for the endogenous metabolites measured by this method. The samples were centrifuged (10 min, 15,000 *g*, 4 °C), and the supernatants were injected directly. Two columns were used for chromatography: i. an Atlantis HILIC silica 3 μm 2.1 × 150 mm column (Waters Corporation) at 30 °C using 0.1 % formic acid and 10 mM ammonium formate as solvent A and 0.1 % formic acid in acetonitrile as solvent B, at 250 uL/min across a 32 min chromatography run time and ii. a Synergi 4u Polar-RP 80A 4μm 4.6 × 50 mm column (Phenomenex) at 30 °C using 95% water, 5% acetonitrile and 5 mM ammonium acetate as solvent A and 5% water, 95% acetonitrile and 5 mM ammonium acetate as solvent B, at 250 uL/min across a 15 min chromatography run time. This method detected the following metabolites: glycine, alanine, serine, threonine, methionine, aspartate, glutamate, asparagine, glutamine, histidine, arginine, lysine, valine, leucine, pheylalanine, tyrosine, tryptophan, proline, cis/trans hydroxyproline, ornithine, citrulline, taurine, 5-HT, 5-HIAA, serotonin, cysteamine, aminoisobutyric acid, dimethylglycine, homocysteine,

7

argininosuccinate, ADMA/SDMA, allantoin, GABA, anthranilic acid, carnitine, kynurenic acid, 1-methylhistamine, 5'-adenosylhomocysteine, carnosine, histamine, 3-hydroxyanthranilic acid, N-carbomoyl-beta-alanine, cobalamin, thiamine, niacinamide, betaine, choline, phosphocholine, phosphoethanolamine, alpha-glycerophosphocholine, acetylcholine, spermidine, creatine, creatinine, thyroxine, trimethylamine-N-oxide, glucose, adenosine, cytosine, thymidine, xanthosine, 2'-deoxycytidine, isoleucine, xantherenate, kynurenine, beta-hydroxyisovaleric acid, adenine, AMP, cGMP, CMP, guanine, guanosine, methyl-cytosine, PEP, SAH, uracil, cystathionine, histidinol, N-acetyl-glutamate, N-acetyl-glutamine, glucosamine-6-phosphate, IDP and indole. As metabolite identification was confirmed with respect to genuine standards this is a level 1 identification according to the Metabolomics Society Metabolomics Standards Initiative [31].

Total plasma fatty acid analysis using fatty acid methyl esters by Gas Chromatography – Mass Spectrometry of blood plasma, pancreas and faeces

Total fatty acid analysis was performed as described in [34]. The organic fraction was dissolved in methanol:chloroform (750 µl, ratio 1:1) and D_{25}-tridecanoic acid in chloroform (internal standard, at 200 µM, 50 µl; Sigma Aldrich) was added. Acid-catalyzed esterification was used to derivatize the organic phase samples. BF_3-methanol (10%, 0.125 ml; Sigma-Aldrich) was added to the organic phase and incubated at 90 °C for 90 minutes. Water (0.15 ml) and hexane (0.3 ml) were added and the samples vortex mixed for 1 minute and left to form a bilayer. The aqueous phase was discarded and the organic layer, containing fatty acid methyl esters (FAMEs), evaporated to dryness under a stream of nitrogen prior to reconstitution in analytical grade hexane (100 µl; Sigma-Aldrich) before gas chromatography mass spectrometry (GC-MS) analysis.

GC-MS was performed using a Trace GC Ultra coupled to a Trace DSQ II mass spectrometer (Thermo Scientific, Hemel Hempstead, UK). The derivatized organic samples were injected with a split ratio of 20 onto a 30 m × 0.25 mm 70% cyanopropyl polysilphenylene-siloxane 0.25 µm TR-FAME stationary phase column (Thermo Scientific). GC-MS chromatograms were processed using Xcaliber (version 2.0; Thermo Scientific). Each individual peak was integrated and then normalized to total intensity of the FAMEs detected. Overlapping peaks were separated using traces of single ions. Peak assignment was based on mass fragmentation patterns matched to the National Institute of Standards and Technology (USA) library and to previously reported literature. Identification of metabolites from organic phase GC-MS analysis was supported by comparison with a FAME

8

standard mix (Supelco 37 Component FAME Mix; Sigma-Aldrich) and retention time matching. As metabolite identification was confirmed with respect to genuine standards this is a level 1 identification according to the Metabolomics Society Initiative [31].

Univariate statistical analysis

Univariate statistical tests including Student's t-test and ANOVA were performed with Matlab (Mathworks) using the Statistical Toolbox, with the exception of Pearson correlation plots which were performed within the statistics package R (www.r-project.com) as described in [35]. For univariate statistics a Bonferonni correction was applied to correct for multiple testing.

Multivariate statistical analysis

The metabolic profiles obtained, containing integral information, were analysed using the multivariate statistical analysis tools found in SIMCA (version 13; Umetrics AB, Umeå, Sweden). Models were processed using principal components analysis (PCA; an unsupervised technique), partial least squares discriminate analysis (PLS-DA; a supervised classification), orthogonal partial least squares discriminate analysis (OPLS-DA; a supervised classification) and orthogonal partial least squares (O-PLS; a multivariate regression tool used with age set as the y-variable) where appropriate. Lipidomics data were Pareto scaled, in which each variable was centred and multiplied by $1/(S_k)^{1/2}$ where S_k is the standard deviation of the variable. GC-MS total fatty acid data and targeted analysis of aqueous metabolites were unit variance scaled. Model statistics reported include the fraction of the sum of squares for the selected component (R^2) which equates to the percentage of the model variance explained, and the predictive ability (Q^2). Cross-validation was performed to test whether models were over fitted. For PLS-DA models random permutation was used whereby the class membership of individual samples are permutated randomly. In addition ANOVA of the cross-validated residuals (CV-ANOVA) test was performed within Simca to further validate the models [36]. Where relevant OPLS-DA models were further validated by selecting 2/3 of the data randomly and then predicting the class membership of the other 1/3.

Results:

Lipidomics of blood plasma:

Blood plasma from both the cross sectional study and time course study were analysed using direct infusion (DI-)MS in positive ion mode (**Figure 1a**). 250 variables were detected

across the dataset and these were imported into the multivariate models. Initially, the samples from the NOD, NOD-E and NOD-SCID mice were compared. OPLS-DA readily separated the three mouse genotypes according to their lipidomic profiles, regardless of age (**Figure 1b**; for summary statistics see figure legend). This model was improved by a two group comparison where the NOD-SCID mice were compared with a combined group of NOD and NOD-E mice (**Figure 1c**). This was in part driven by the greater long chain containing triglycerides in the two NOD strains, particularly TAG(58:9) (exact mass(m/z): 946.789), TAG(62:14) (m/z = 992.772), TAG(58:10) (m/z = 944.773), TAG(60:11) (m/z = 970.790), TAG(58:11) (m/z = 942.757), TAG(58:8) (m/z = 948.805), TAG(60:12) (m/z = 968.773), TAG(60:10) (m/z = 972.805), TAG(56.9) (m/z = 918.756) and TAG(60:9) (m/z = 974.821). To test the validity of the model further, an OPLS-DA model was built using 2/3 of the data and then predicting the class membership of the remaining third. This correctly picked the classification of 87% of the data.

Removing the NOD-SCID group from the dataset, and comparing the NOD and NOD-E mice for those aged 4-23 weeks, OPLS-DA discriminated the two mouse strains (**Figure 1d**). NOD-E were characterised by increased concentrations of cholesterol esters and lysophosphatidylcholines (**Table 1** for list of metabolite changes) while NOD mice had higher concentrations of a range of triglycerides and diglycerides (**Table 1**). To test the validity of the model further, an OPLS-DA model was built using 2/3 of the data and then predicting the class membership of the remaining third. This correctly picked the classification of 71% of the data.

Considering the time course of the development of type 1 diabetes in the NOD mice, the lipidomics data for the NOD and NOD-E mice was next examined looking at mice less than 14 weeks and mice aged 14 weeks and above. For mice less than 14 weeks, OPLS-DA separated the two genotypes, although the model was relatively weak (**Figure 1e**). This classification was caused by increases in ceramides, cholesterol esters and lyso-phophatidylcholines and three unknown lipids (m/z = 642.619), (m/z = 590.323) and (m/z = 909.548) in NOD-E mice, and triglycerides and an unknown lipid (m/z = 577.520) in NOD mice (**Table 1**). A more robust OPLS-DA model was built for mice aged 14 weeks and older (**Figure 1f**). However, the distinct pattern between cholesterol esters and lyso-phospholipids versus triglycerides found in the overall NOD compared with NOD-E mice comparison, and the younger mice was no longer apparent at this later time point (**Table 1**).

Next we compared lipidomic profiles of NOD mice aged 25-30 weeks who had not developed T1DM (NOD* group) with the NOD and NOD-E lipidomic profiles used in the previous comparisons. No model could be built that discriminated that NOD* group from the

10

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
...

NOD-E mice, indicating they shared a common metabolic profile. However, the NOD and NOD* group was separated by OPLS-DA (**Figure 1g**). To further investigate these profiles, the NOD* group was mapped onto the NOD and NOD-E OPLS-DA model in **Figure 1d**. Using this model to classify the NOD* group, 10/10 samples were classified with the NOD-E group, with the average predicted Y = 0.84 ± 0.14, where Y=1 is a perfect classification as a NOD-E mouse and Y= 0 is a perfect classification as a NOD mouse.

Table 1: Summary of metabolite changes for NOD and NOD-E comparisons described in Figure 1. Key: CE cholesterol ester, LPC lysophosphatidylcholine, TAG triacylglycerols, DI-MS direct infusion mass spectrometry, QqQ triple quadrupole mass spectrometry. MSI Metabolomics Standardisation Initiative for metabolite identification.

Comparison	Increased in NOD	Increased in NOD-E	Method	Adduct	Mass or mass transition	MSI level assignment
All NOD v NOD-E		CE(16:1)	DI-MS	$[M+H]^+$	649.603	2
		LPC(20:4)	DI-MS	$[M+H]^+$	544.340	2
		LPC(22:6)	DI-MS	$[M+H]^+$	568.340	2
		CE(18:1)	DI-MS	$[M+H]^+$	668.634	2
		LPC(22:6)	DI-MS	$[M+H]^+$	590.323	2
		LPC(18:2)	DI-MS	$[M+H]^+$	520.340	2
		LPC(18:1)	DI-MS	$[M+H]^+$	522.356	2
		CE(16:0)	DI-MS	$[M+H]^+$	642.619	2
		PC(34:3)	DI-MS	$[M+H]^+$	756.555	2
		CE(18:4)	DI-MS	$[M+H]^+$	645.559	2
	TAG(56:8)		DI-MS	$[M+NH_4]^+$	920.772	2
	TAG(52:6)		DI-MS	$[M+NH_4]^+$	868.740	2
	TAG(54:4)		DI-MS	$[M+NH_4]^+$	900.804	2
	TAG(52:3)		DI-MS	$[M+NH_4]^+$	874.788	2
	TAG(53:4)		DI-MS	$[M+NH_4]^+$	886.788	2
	TAG(58:10)		DI-MS	$[M+NH_4]^+$	944.773	2
	TAG(54:7)		DI-MS	$[M+NH_4]^+$	894.756	2
	TAG(54:5)		DI-MS	$[M+NH_4]^+$	898.788	2
	TAG(52:5)		DI-MS	$[M+NH_4]^+$	870.756	2
	DAG(41:4)		DI-MS	$[M+NH_4]^+$	704.636	2
	TAG(54:6)		DI-MS	$[M+NH_4]^+$	896.772	2
	TAG(52:4)		DI-MS	$[M+NH_4]^+$	872.772	2
	Argininosuccinate		QqQ		291→70	1
	cobalamin		QqQ		678→147	1
	creatinine		QqQ		114→44	1
	AMP		QqQ		134→107	1
	kynurenic acid		QqQ		190→144	1
		Serotonin	QqQ		177→160	1
		5-hydroxytryptophan	QqQ		221→204	1
		phosphoethanolamine	QqQ		142→44	1
		phosphoenolpyruvate	QqQ		168→	1
		Ornithine	QqQ		133→70	1
Young NOD v NOD-E (less than 14 weeks)		(Cer(t40:0)	DI-MS	$[M+H]^+$	640.604	2
		Cer(t42:0)	DI-MS	$[M+H]^+$	668.633	2
		CE(18:4)	DI-MS	$[M+H]^+$	645.559	2
		CE(20:4)	DI-MS	$[M+H]^+$	673.590	2
		(LPC(20:4)	DI-MS	$[M+H]^+$	544.340	2
		LPC(22:6)	DI-MS	$[M+H]^+$	568.340	2
	TAG(52:3)		DI-MS	$[M+NH_4]^+$	874.788	2
	TAG(51:3)		DI-MS	$[M+NH_4]^+$	860.772	2
	TAG(52:4)		DI-MS	$[M+NH_4]^+$	872.772	2
	TAG(52:5)		DI-MS	$[M+NH_4]^+$	870.756	2
	TAG(52:6)		DI-MS	$[M+NH_4]^+$	868.740	2

11

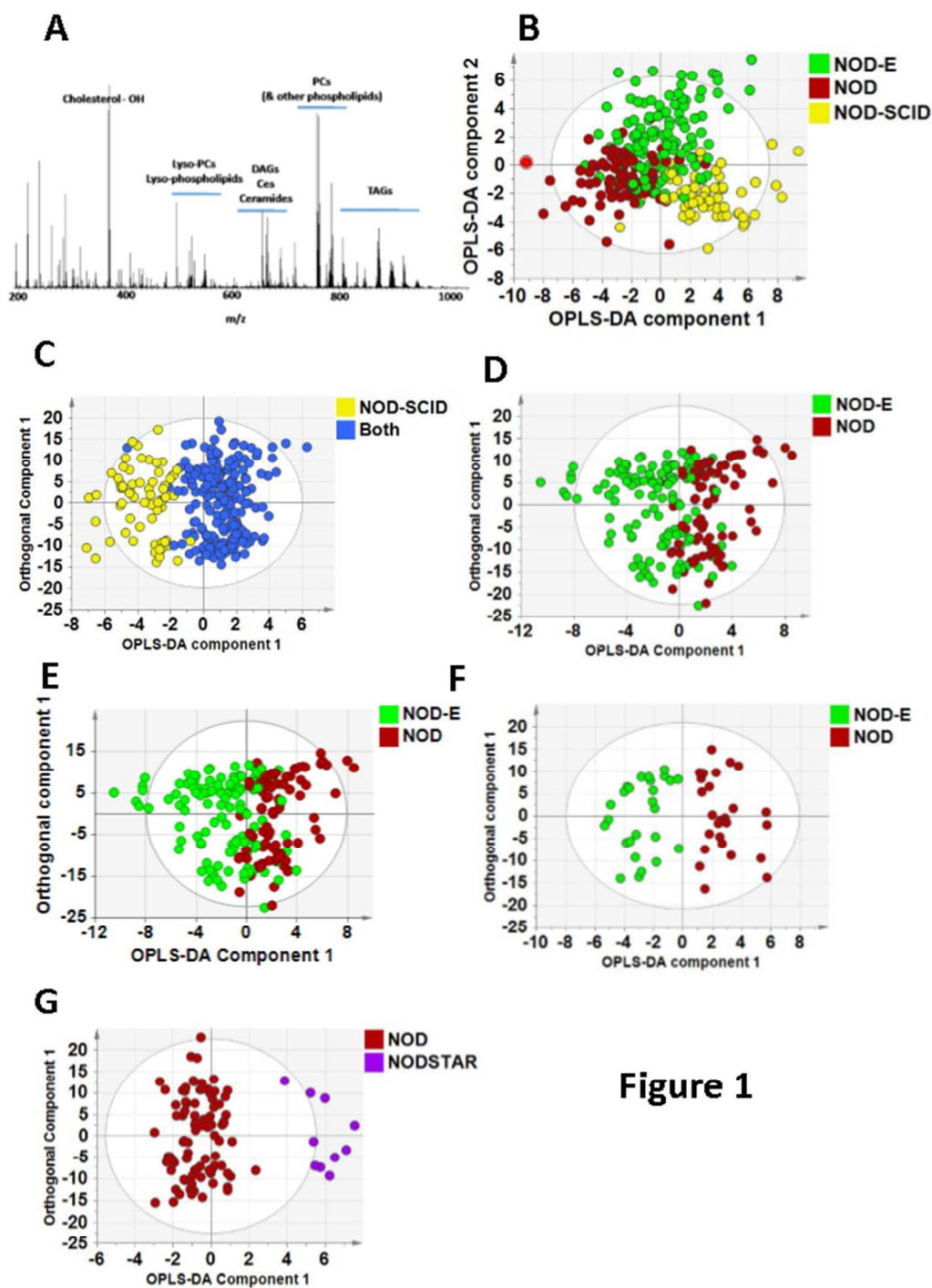| Group	Compound	Compound	Method	Adduct	m/z	n
	TAG(54:5)		DI-MS	$[M+NH_4]^+$	898.788	2
	TAG(54:7)		DI-MS	$[M+NH_4]^+$	894.756	1
	TAG(50:4)		DI-MS	$[M+NH_4]^+$	844.741	1
	TAG(54:6)		DI-MS	$[M+NH_4]^+$	896.772	1
	Arginosuccinate		QqQ		291→70	1
	Cobalamin		QqQ		678→147	1
	creatine		QqQ		132→90	1
		5-hydroxytryptophan	QqQ		221→204	1
		Serotonin	QqQ		177→160	1
		phosphoenolpyruvate	QqQ		168→	1
		Phosphoethanolamine	QqQ		142→44	1
		ornithine	QqQ		133→70	1
Old NOD v NOD-E (14 weeks and older)		TAG(56:6)	DI-MS	$[M+NH_4]^+$	924.807	2
		PC(32:0)	DI-MS	$[M+H]^+$	734.570	2
		TAG(50:2)	DI-MS	$[M+NH_4]^+$	848.772	2
		TAG(56:7)	DI-MS	$[M+NH_4]^+$	922.789	2
		TAG(52:3)	DI-MS	$[M+NH_4]^+$	874.788	2
		TAG(50:1)	DI-MS	$[M+NH_4]^+$	850.789	2
		DG (O-34:3)	DI-MS	$[M+NH_4]^+$	577.520	2
		TAG(50:3)	DI-MS	$[M+NH_4]^+$	846.756	2
	CE(20:4)	LPC(22:6)	DI-MS	$[M+H]^+$	568.340	2
	LPC(20:3)		DI-MS	$[M+H]^+$	690.620	2
	Cer(t43:1)		DI-MS	$[M+H]^+$	546.354	2
	LPC (18:0)		DI-MS	$[M+H]^+$	680.635	2
	CE(20.4)		DI-MS	$[M+H]+$	524.372	2
	AMP		DI-MS	$[M+Na]+$	695.575	2
	Cobalamin		QqQ		134→107	1
	Arginosuccinate		QqQ		678→147	1
			QqQ		291→70	1
		Serotonin	QqQ		177→160	1
		5-hydroxytryptophan	QqQ		221→204	1
		phosphoenolpyruvate	QqQ		168→	1
		Phosphoethanolamine	QqQ		142→44	1

Figure 1: **A:** High resolution mass spectrum of a typical lipid extract from blood plasma derived from the NOD mouse. **B.** OPLS-DA plot of lipidomics of blood plasma from NOD, NOD-E and NOD-SCID mice considered as separate groups. Data is Pareto scaled. $R^2(X)$ = 75%, $R^2(Y)$ = 48%; Q^2 = 34%; p = 3.5*10^{-22} by CV-ANOVA for cross validation. 2

13

components and 6 orthogonal components. **C.** OPLS-DA plot of lipidomics of blood plasma from NOD-SCID mice compared with a combined group of NOD and NOD-E mice (labelled Both). Data is Pareto scaled. $R^2(X) = 69\%$; $R^2(Y) = 68\%$; $Q^2 = 51\%$; $p = 5.8*10^{-34}$. 1 component and 6 orthogonal components. **D.** OPLS-DA plot of lipidomics of blood plasma from NOD and NOD-E mice. Data is Pareto scaled. $R^2(X) = 57\%$, $R^2(Y) = 41\%$, $Q^2 = 23\%$; p = $5.4*10^{-9}$, 1 component and 3 orthogonal components. **E.** OPLS-DA plot of lipidomics of blood plasma from NOD and NOD-E mice under 14 weeks of age. Data is Pareto scaled. $R^2(X) = 43\%$, $R^2(Y) = 32\%$, $Q^2 = 23\%$, p = $1.2*10^{-7}$, 1 component and 1 orthogonal component. **F.** OPLS-DA plot of lipidomics of blood plasma from NOD and NOD-E mice aged 14 weeks and over. Data is Pareto scaled. $R^2(X) = 56\%$, $R^2(Y) = 80\%$, $Q^2 = 54\%$; p=$2.2*10^{-5}$, 1 component and 4 orthogonal components. **G**. OPLS-DA plot of lipidomics of blood plasma from NOD (diabetic) and NOD (non-diabetic referred to as NODSTAR) mice aged 25-30 weeks and over. $R^2(X) = 68\%$; $R^2(Y) = 79\%$; $Q^2 = 56\%$; p = $2.5*10^{-11}$, 1 component and 4 orthogonal components.

To explore the lipidomic changes further, total fatty acid analysis was performed using GC-MS. Examining the different mouse strains, a model could be built that discriminated the three mouse strains (**Figure 2a**), regardless of age of animal, with again the NOD-SCID mice being the most different from the other two mouse strains. To explore this difference further, the NOD-SCID mice were compared with the other two strains treated as a single group. This produced a more robust model (**Figure 2b)**, with NOD-SCID mice having higher concentrations of C16:0, C21:0, C20:5, and C18:3 and decreased concentrations of C16:2, C24:1, C17:0 and C14:0. Similar fatty acid changes were also observed in the Pareto scaled model as evidenced in the associated S-plot (**Figure 2c**). To test the validity of the model further, an OPLS-DA model was built using 2/3 of the data and then predicting the class membership of the remaining third. This correctly picked the classification of 88% of the data. However, NOD and NOD-E mice regardless of age could not be separated according to their total fatty acid profile (data not shown). Similarly no model could be built that discriminated NOD from NOD-E mice when considering animals 14 weeks and over of age, or when considering the animals under 14 weeks of age.

Finally we examined whether we could model time changes in either the NOD or NOD-E mice using PLS to regress age of animal against metabolic profiles. A robust model was built for regressing the blood plasma profile of NOD mice against age (**Figure 2d** shows the actual verses predicted age of mice according to the total fatty acid profile of the blood plasma). This was caused by increases in C18:3, C18:2, and C20:4, and decreases in C16:1, cholesterol, C22:0, and C20:3 as the animals age. However, no model could be produced for the NOD-E mice that passed cross validation.

14

Figure 2: A. OPLS-DA plot of total fatty acid analysis of blood plasma from NOD, NOD-E and NOD-SCID mice considered as separate groups. Data is Univariate scaled. $R^2(X) = 25\%$; $R^2(Y) = 47\%$; $Q^2 = 33\%$, 2 components and 1 orthogonal component. **B.** OPLS-DA plot of total fatty acid analysis of blood plasma from NOD-SCID mice compared with a combined group of NOD and NOD-E mice (labelled Both). Data is univariate scaled. $R^2(X) = 40\%$; $R^2(Y) = 80\%$; $Q^2 = 72\%$; $p = 6.3*10^{-37}$, 1 component and 3 orthogonal components. **C.** S-plot of Pareto scaled variables from the dataset analysed for the plot in **B**. **D.** PLS plot of age against metabolic profile of NOD mice. $Q^2 = 51\%$, $R^2(X) = 27\%$, $R^2Y = 77\%$; $p=4*10^{-6}$, 1 component and 2 orthogonal.
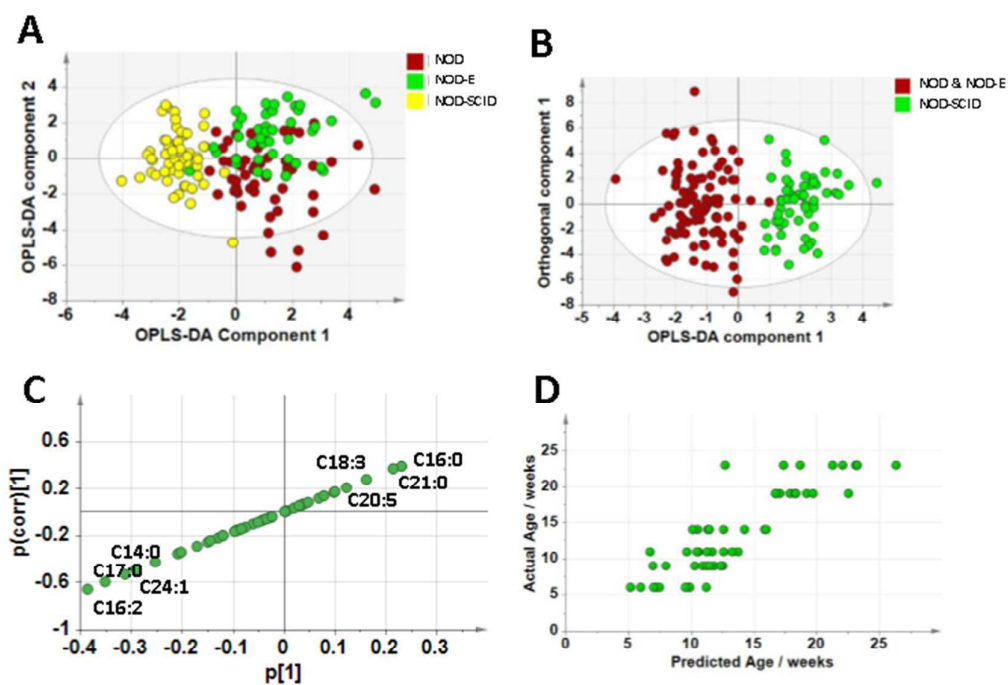
Alterations in amino acid metabolism in blood plasma

LC-MS/MS was used to profile amino acids and nucleotides using a targeted method of 87 metabolites in blood plasma and then multivariate statistics was used to identify clusters within the dataset in terms of comparing different mouse genotypes, as well as exploring time trends by regressing age of animal against metabolic profile. Initially, the three mouse strains were compared regardless of age. OPLS-DA readily discriminated all three groups

according to genotype, with NOD-SCID mice most separated along the first component compared with the other two animal models (**Figure 3a**). Similarly to the lipidomic data, a more robust model could be built separating the NOD-SCID mice from the other two strains (**Figure 3b**). To test the validity of the model further, an OPLS-DA model was built using 2/3 of the data and then predicting the class membership of the remaining third. This correctly picked the classification of 92% of the data. This separation was driven by increases in trimethylamine-N-oxide, kynurenine, 1-methyl-histidine, xanthosine and lysine and decreases in cobalamin, betaine, β-hydroxyisovaleric acid, acetylcholine and inosine diphosphate in the NOD-SCID mice (**Figure 3c**). Taking the metabolites with a Variable Importance Parameter >1 from this model, metabolites involved in nucleotide metabolism, B-vitamins, one-carbon metabolism/choline breakdown and amino acids were responsible for this discrimination (**Supplementary Table 3**). Using the metabolite enrichment software on MetaboAnalyst [37] the most perturbed pathways were glycine, serine and threonine metabolism, biotin/thiamine metabolism, beta-alanine metabolism, tryptophan metabolism and protein biosynthesis.

Underlying the large differences between NOD-SCID mice and the other two strains, a weaker model was produced examining NOD and NOD-E mice, regardless of age. OPLS-DA produced a one component model with three orthogonal components with a $Q^2 = 50\%$ indicating the separation was less marked than models involving the NOD-SCID strain (**Figure 3d**). To test the validity of the model further, an OPLS-DA model was built using 2/3 of the data randomly selected and then predicting the class membership of the remaining third. This correctly picked the classification of 83% of the data. This was associated with an increase in the concentrations of argininosuccinate, cobalamin, creatinine, AMP and kynurenic acid in NOD mice and a relative increase in concentration of serotonin, 5-hydroxytryptophan (5-HT), phosphoethanolamine, phosphoenolpyruvate and ornithine in NOD-E mice (**Figure 3e**). While better models could be built considering mice aged below 14 weeks (**Figure 3f**) and over 14 weeks (**Figure 3g**), similar metabolites drove all three models (**Figure 3e & 3h**).
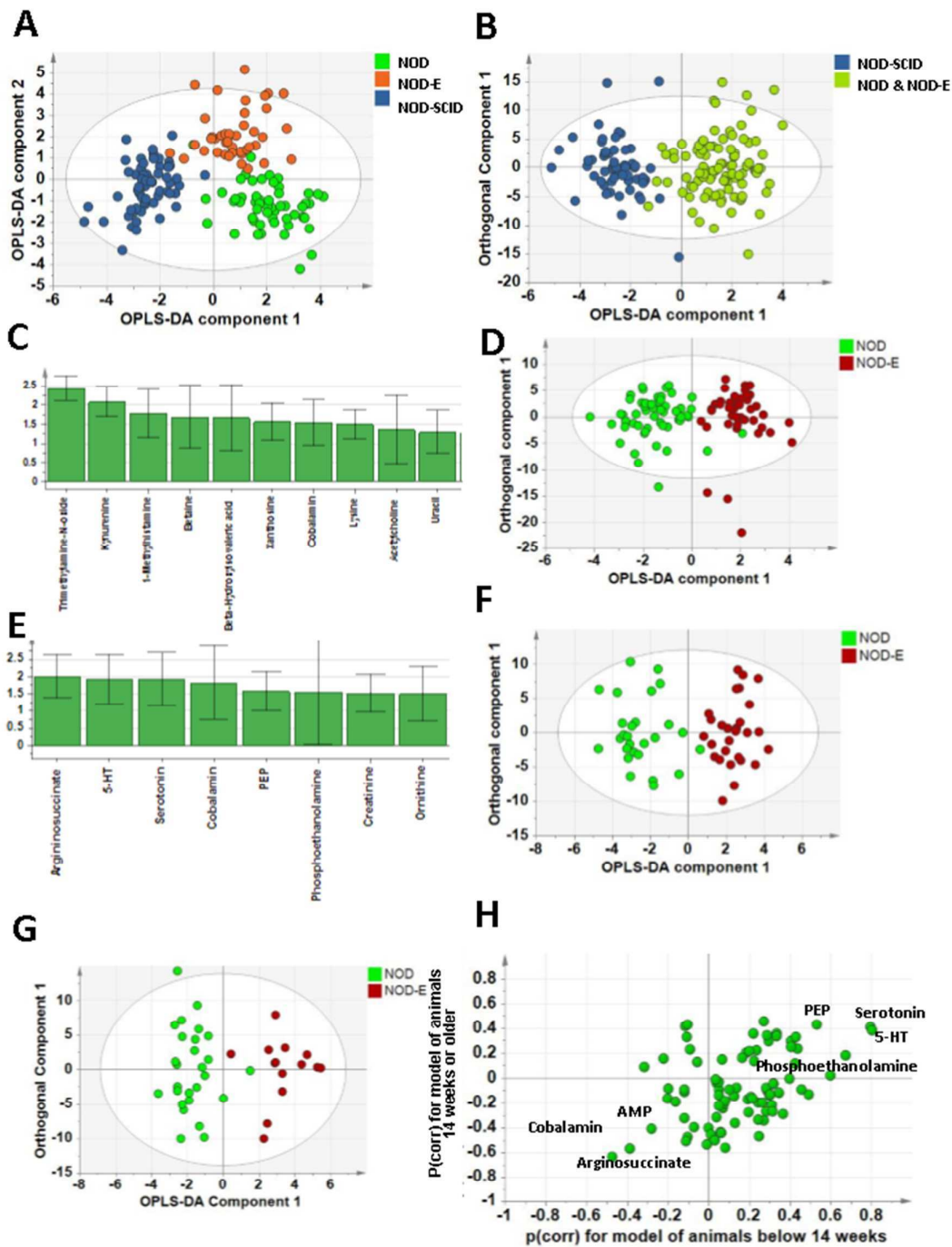
Figure 3: A. OPLS-DA plot of targeted analysis of central metabolism (amino acids, TCA cycle intermediates, nucleotides, and other small aqueous metabolites) of blood plasma from NOD, NOD-E and NOD-SCID mice considered as separate groups. Data is Univariate scaled. $R^2(X) = 56\%$, $R^2(Y) = 72\%$, $Q^2 = 52\%$, 2 components and 4 orthogonal components; CV ANOVA validation $p = 1.5*10^{-22}$. **B.** OPLS-DA plot of targeted analysis of central metabolism of blood plasma from NOD-SCID mice compared with the NOD and NOD-E

mice as a single group. $R^2(X) = 42\%$, $R^2(Y) = 79\%$, $Q^2 = 79\%$, 1 component and 2 orthogonal components; CV-ANOVA validation $p = 2.7*10^{-38}$. **C.** Variable Importance Parameter (VIP) plot to show key metabolites that are most important for the discrimination for the plot in **B**. **D.** OPLS-DA plot of targeted analysis of central metabolism of blood plasma from NOD and NOD-E mice. Data is Univariate scaled. $R^2(X) = 51\%$; $R^2(Y) = 76\%$; CV-ANOVA validation $p = 2*10^{-10}$. **E.** VIP plot to show key metabolites that are most important for the discrimination for the plot in **D**. **F.** OPLS-DA plot of targeted analysis of central metabolism of blood plasma from NOD and NOD-E mice below 14 weeks of age. Data is Univariate scaled. $R^2(X) = 44\%$, $R^2(Y) = 84\%$, $Q^2 = 74\%$, 1 component and 2 orthogonal components; CV ANOVA validation $p = 2.7*10^{-38}$. **G.** OPLS-DA plot of targeted analysis of central metabolism of blood plasma from NOD and NOD-E mice 14 weeks and older of age. Data is Univariate scaled. $R^2(X) = 64\%$, $R^2(Y) = 82\%$, $Q^2 = 51\%$, 1 component and 4 orthogonal components; CV-ANOVA validation $p = 2.7*10^{-38}$. **H.** SUS-plot of discriminatory metabolites for models described in **F** and **G**.

For both non-diabetic and diabetic NOD mice, and NOD-E mice robust PLS models could be built between amino acid profiles and age of animal. For the non-diabetic NOD mice, age was associated with an increase in 5-hydroxytryptophan ($p = 0.017$; One-way ANOVA), serotonin ($p = 0.013$), creatinine ($p < 0.0001$), histamine ($p = 0.0002$) and cytosine ($p < 0.0001$) and decreases in dimethylglycine ($p = 0.0072$) and hydroxyproline ($p=0.005$) producing a robust PLS model (**Figure 4a**). Repeating this analysis with the diabetic NOD mice another highly predictive model was again built between age and amino acid profiles ($R^2X = 45\%$, $R^2Y = 97\%$; $Q^2 = 78\%$; data not shown). As with the non-diabetic animals, ageing was associated with a decrease in hydroxyproline ($p = 0.0002$ using one-way ANOVA). In addition, ageing of the non-diabetic NOD animals was associated with a decrease in arginine ($p = 0.004$), glycerophosphocholine ($p < 0.0001$) and lysine ($p = 0.006$) and increased in trimethylamine-N-oxide ($p<0.0001$), kynurenic acid ($p < 0.0001$), phenylalanine ($p = 0.0005$), glutamine ($p = 0.0003$) and alanine ($p < 0.0001$) (data not shown).

Similarly a robust PLS model was also built for NOD-E mice ($Q^2 = 72\%$, $R^2X = 46\%$, $R^2Y = 92\%$). As with the NOD mice, ageing was associated with a decrease in hydroxyproline ($p < 0.0001$ one-way ANOVA) and an increase in creatinine ($p < 0.0001$). In addition across the ageing time course there was a decrease in xanthosine ($p < 0.0001$) and increases in the concentration of tryptophan ($p = 0.0001$), thiamine ($p < 0.0001$), creatine ($p = 0.0002$), and

indole (p = 0.0003). In all three models the use of SUS plots demonstrated that the metabolic effects of ageing were similar across all three mouse groups (**Figure 4b**).
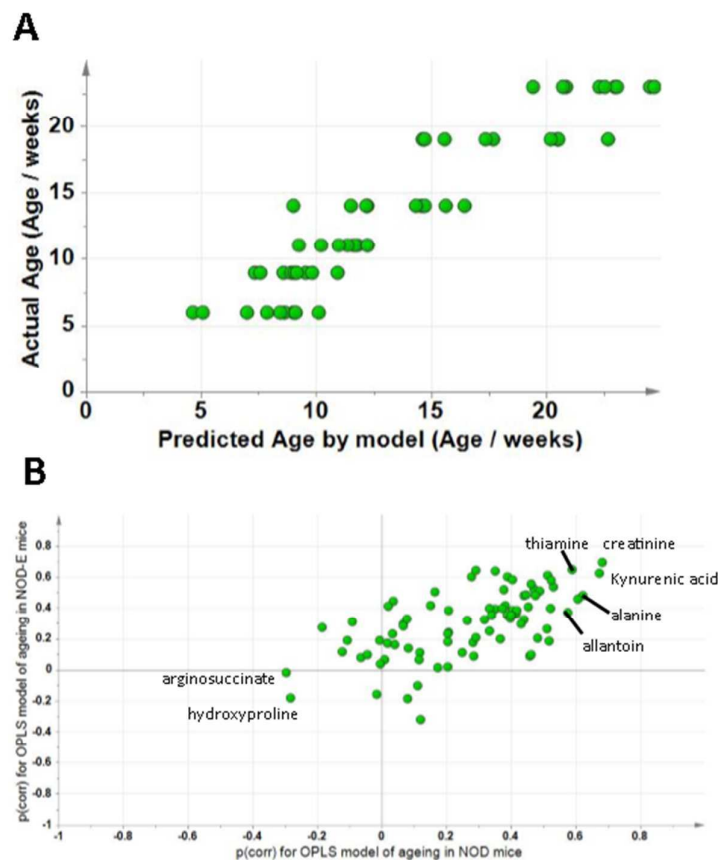
A

B

Figure 4. A: PLS model of targeted analysis of central metabolism of blood plasma from NOD mice against age (R^2X = 53%, R^2Y = 79%, Q^2 = 65%). **B:** SUS-plot of PLS models for NOD mice and NOD-E mice against age. In this plot the loadings for the two PLS models are plotted against each other so that if the two models had the same loadings then these would line up at 45° to the x and y-axes. While some metabolites are in common there is also a degree of scatter from the 45° line showing that other metabolites are responsible for the ageing trends in the NOD and NOD-E mice.

One of the major pathways to be perturbed between the three strains of mice and across the ageing studies was the tryptophan metabolism pathway (KEGG map 00380). To investigate this further we constructed correlation plots between tryptophan and its products detected in the targeted assay for NOD and NOD-E mice (**Figure 5a**). For both strains tryptophan was

highly correlated with the majority of metabolites, as might be expected by the major precursor to many of these metabolites (**Figure 5b**). However, for both mouse strains it was not correlated to serotonin, 5-HT and 5-hydroxyindole acetic acid concentrations, unlike the metabolites in the NAD biosynthesis and xanthurenic pathways, suggesting the latter two pathways are in part regulated by the total concentration of tryptophan. For the NOD-E mice the pathway from kynurenine to xanthurenic acid was more pronounced than that found in NOD mice, while correlations were more evenly spread across metabolites including anthranilic acid and quinolinic acid for NOD mice.
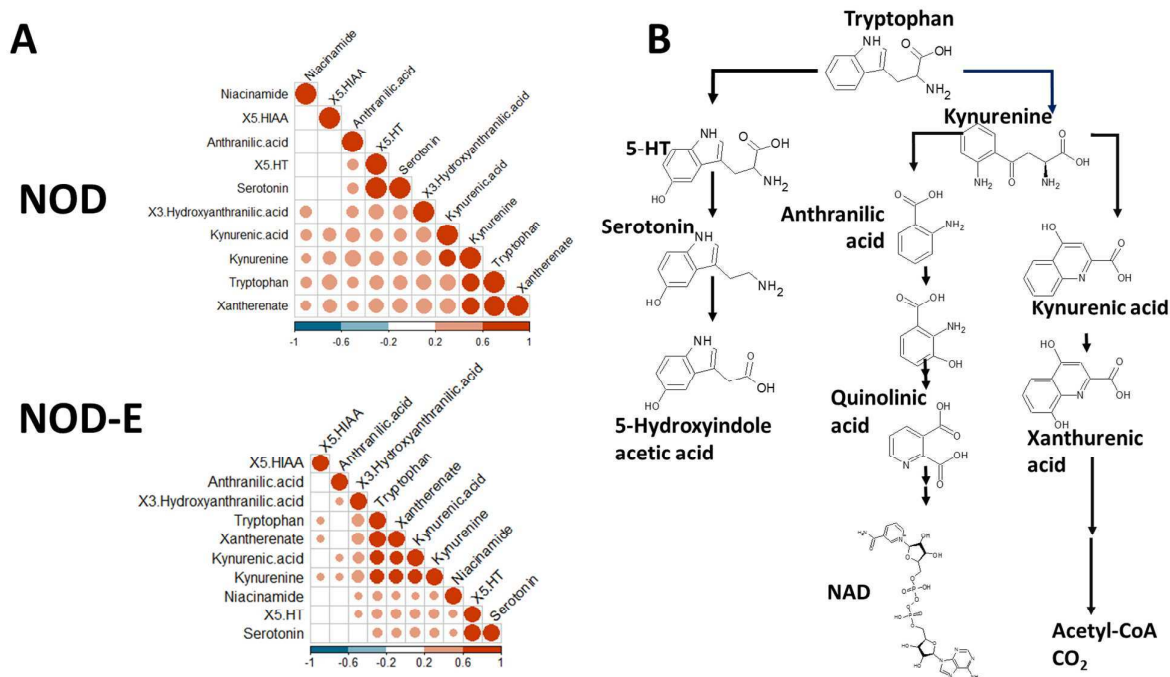

Figure 5. A: Pearson correlation plots for the tryptophan pathway for NOD and NOD-E mice. Correlations were calculated for each pair of metabolites with the circle representing the magnitude of the correlation coefficient. **B:** The tryptophan pathway mapped in the correlation analysis in **A**. Key: 5-HT 5-hydroxytryptophan, 5-HIAA 5-Hydroxyindole acetic acid.

Alterations in metabolism in the pancreas using lipidomics:

Our lipidomic analysis was applied to the pancreas samples to further investigate whether blood plasma based biomarkers reflected changes in the pancreas across the mouse models examined. For the positive ionisation mode analysis of the pancreas extracts (**Figure**

20

6a), 470 variables were detected across the dataset and these were imported into the multivariate models. Examining samples from mice regardless of age, again, the samples from the NOD, NOD-E and NOD-SCID mice were compared and OPLS-DA readily separated the three mouse genotypes according to their lipidomic profiles (**Figure 6b**) demonstrating the profound difference between the three mouse models. This model was improved by a two group comparison where the NOD-SCID mice were compared with a combined group of NOD and NOD-E mice (**Figure 6c**), although one must be aware of the relatively small number of NOD-SCID animals in this comparison. This was in part driven by increased concentrations of sphingomyelins SM(d34:2) (m/z = 701.558, $[M+H]^+$) and SM(d38:2) (m/z = 757.621, $[M+H]^+$], SM(d40:2) (m/z = 785.652, $[M+H]^+$) and the sphingolipid Cer (t34:0) (m/z = 594.581, $[M+K]^+$]) in the pancreas tissue of NOD-SCID mice, and increases in the concentrations of TG (46:1) (m/z = 815.636, $[M+K]^+$), PC (33:4) (m/z = 762.506, $[M+Na]^+$), PC (34:2) (m/z = 758.567, $[M+H]^+$), MG (16:0) (m/z = 313.272, $[M+H-H_2O]^+$ and unknown m/z = 760.135 in the NOD and NOD-E mice.

Excluding samples from the NOD-SCID mice from the analysis, pancreas tissue from NOD and NOD-E mice could readily be distinguished (**Figure 6d**). This was driven by increases in the concentration of lipids PE (O-36:5), (m/z = 724.526, $[M+H]^+$), PE (O-32:0), (m/z = 700.523, $[M+Na]^+$), PE (P-38:6; (m/z = 748.526, $[M+H]^+$), unknown m/z = 753.558 and SM(39:1) (m/z = 773.652, $[M+H]^+$) in the NOD mice and increases in Cer (d34:1) (m/z = 538.518, $[M+H]^+$), DG (40:3) (m/z = 699.592, $[M+H-H_2O]^+$), TG (48:1) (m/z = 843.681, $[M+K]^+$) and DG (40:5) (m/z = 671.565, $[M+H]^+$) lipids in NOD-E mice.

To further examine the metabolic changes in the pancreas we examined NOD and NOD-E mice aged 16 weeks of age and older, comparing NOD mice that had either developed or not developed diabetes. However, no model could be built that discriminated between NOD and NOD-E mice, and a combined NOD-E, NOD and NOD non diabetic model, suggesting that metabolic differences were most apparent at the earlier time points. To confirm this, a robust model was built comparing the lipid profiles of pancreas from mice aged 12 weeks and under **(Figure 6e)**. This was caused by increases in TAG (34:1), (m/z = 647.4563, [M+K]+); PE-Ceramide (d27:2, putative) (m/z = 583.383, [M+H]+); TAG (46:8), (m/z = 801.548, [M+H]+) and PE (O-36:5) or PE(P-36:4); (m/z = 724.526, [M+Na]+) and decreases in PS(O-40:5) (m/z = 824.588, [M+H]+; PE (30:2),(m/z = 660.421, [M+H]+]), PE (32:2), (m/z = 688.451, [M+H]+) and PC (O-38:4), (m/z = 818.601, [M+Na]+) in the pancreas of NOD mice.
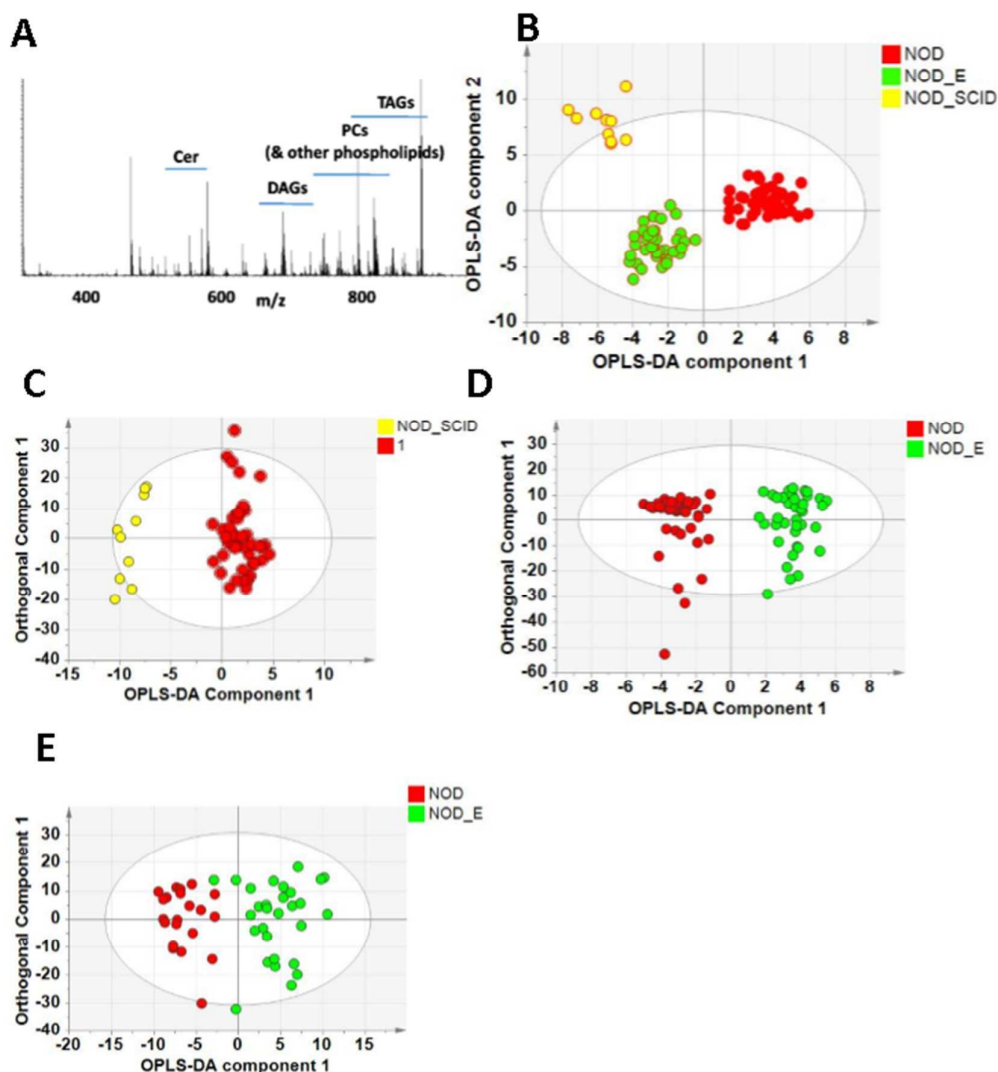
Figure 6. **A:** High resolution mass spectrum of a typical pancreas extract from the NOD mouse. **B:** OPLS-DA plot of lipidomics of the pancreas comparing NOD (red circles), NOD-E (green circles) and NOD-SCID (yellow circles) mice. Data is univariate scaled. $R^2(X) = 79\%$, $R^2(Y) = 89\%$; $Q^2 = 61\%$; $p = 4.5*10^{-12}$ by CV-ANOVA for cross validation. 2 components and 7 orthogonal components. **C:** OPLS-DA plot of lipidomics of the pancreas comparing NOD-SCID mice (yellow circles) compared with the other NOD and NOD-E mice (red circles) treated as a single group. $R^2(X) = 74\%$; $R^2(Y) = 92\%$; $Q^2 = 79\%$; $p = 1.3*10^{-13}$, 1 component and 4 orthogonal components. **D**. OPLS-DA plot of lipidomics of the pancreas comparing NOD (red circles) and NOD-E (green circles) mice across all ages. $R^2(X) = 76\%$; $R^2(Y) = 92\%$; $Q^2 = 60\%$; $p = 7.6*10^{-9}$, 1 component and 6 orthogonal components. **E**. OPLS-DA plot of lipidomics of the pancreas comparing NOD and NOD-E mice under 14 weeks of age.

$R^2(X) = 49\%$; $R^2(Y) = 81\%$; $Q^2 = 63\%$; $p = 5.2 \times 10^{-8}$, 1 component and 2 orthogonal components.

Alterations in amino acid metabolism in the pancreas:

Repeating the amino acid based assay analysis for the pancreas, again the pancreas of the NOD-SCID mice were significantly different from the other two mouse models either treating all three models as distinct groups (data not shown) or comparing the NOD-SCID mice with a combined group of NOD and NOD-E mice (**Figure 7a**). This was driven by increases in riboflavin, thiamine, anthranilic acid, methionine and dimethylglycine in the NOD-SCID mice and increases in phosphocholine, 3-phosphoserine, creatine, cytosine and cytidine in the other two strains. Similar results were produced when the data was Pareto scaled (**Figure 7b**).

Repeating the analysis but comparing the NOD and NOD-E mice another robust OPLS-DA was built ($R^2(X) = 42\%$; $R^2(Y) = 99\%$; $Q^2 = 70\%$; CV-ANOVA $p = 0.04$; **Figure 7c**). The pancreas from NOD mice had increased concentrations of alanine ($p = 0.003$ for a Student's T-test univariate analysis), asparagine ($p = 0.03$), xanthine, NADH, arginosuccinate and glucose while the pancreas from NOD-E mice were most characterised by relative increases in a range of aromatic amino acids including indole ($p = 0.004$), xanthurenate ($p = 0.002$), tryptophan ($p = 0.001$), S-adenosyl homocysteine ($p = 0.0006$), 5-HT ($p = 0.0004$) and serotonin ($p = 0.0002$) as well as methionine ($p = 000003$). S-adenosyl homocysteine, 5-HT, serotonin and methionine also passed the threshold for Bonferroni correction ($p = 0.00047$). Similar results were produced when the data was Pareto scaled (**Figure 7d**).
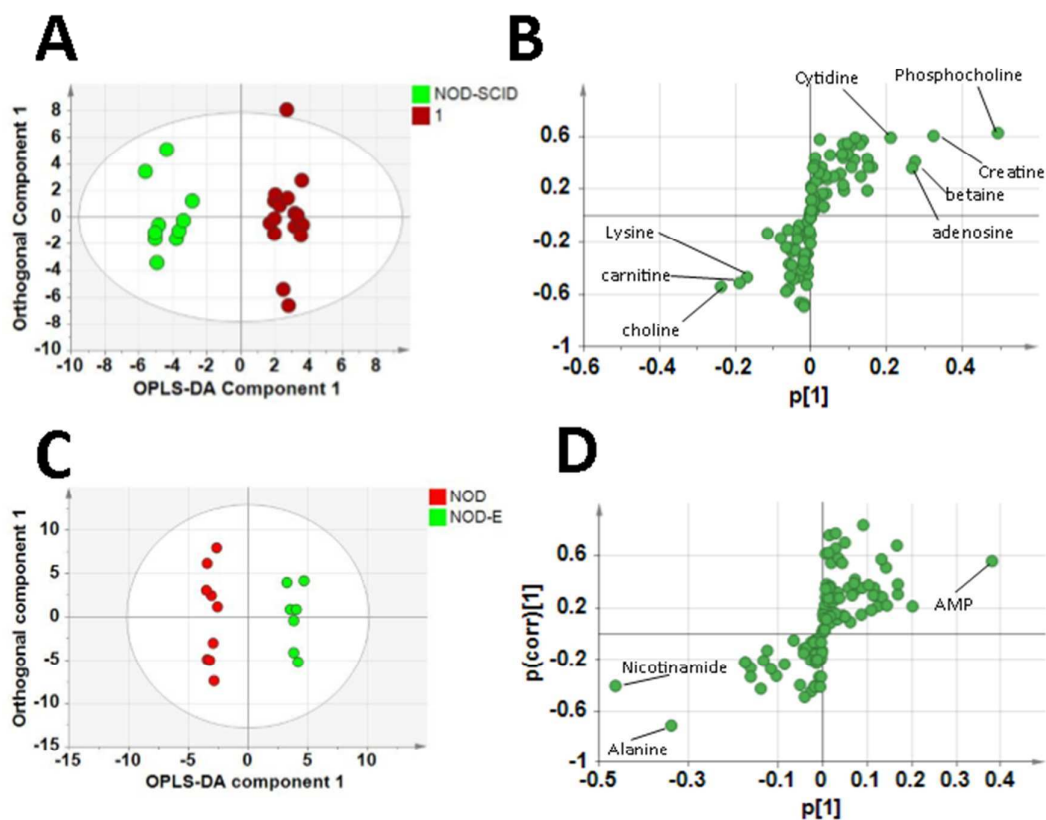
Figure 7: A: OPLS-DA plot of targeted core metabolism of extracts of pancreas tissue comparing the NOD-SCID mouse (green circles) to a combined group of NOD and NOD-E mice (red circles), regardless of age. Data is univariate scaled. $R^2(X) = 37\%$; $R^2(Y) = 96\%$; $Q^2 = 82\%$; CV-ANOVA p = $2.8*10^{-6}$, 1 component and 2 orthogonal components. **B:** S-plot of most discriminatory metabolites in a Pareto scaled OPLS-DA model using the data analysed in **A**. **C:** OPLS-DA plot of targeted core metabolism of extracts of pancreas tissue comparing the NOD mouse (red circles) to NOD-E mice (green circles), regardless of age. Data is univariate scaled. $R^2(X) = 42\%$; $R^2(Y) = 99\%$; $Q^2 = 70\%$; CV-ANOVA p = 0.04, 1component and 2 orthogonal components. **D:** S-plot of most discriminatory metabolites in a Pareto scaled OPLS-DA model using the data analysed in **C**.

Modelling correlated metabolism between blood plasma and the pancreas

To examine whether metabolic changes in the pancreas manifested themselves in blood plasma, OPLS was used to regress changes in aqueous metabolism in the pancreas against the same metabolites detected in blood plasma across 23 week old NOD, NOD-E and NOD-SCID mice. Applying this approach to all the mouse models a weak OPLS model was built (2

components and 2 orthogonal X components; $R^2(X) = 53\%$; $R^2(Y) = 54\%$; $Q^2 = 8\%$;
Supplementary Figure 1a). The most correlated metabolite between the two tissues was
betaine ($p = 0.005$) and β-hydroxyisovalerate ($p = 0.007$). Next we examined correlations
within individual strains. While no model could be built for the NOD-SCID mice and only a
poor model for the NOD mice was produced (2 components and no orthogonal components;
$R^2(X) = 55\%$; $R^2(Y) = 60\%$; $Q^2 = 16\%$), a robust model was built for NOD-E mice (3
components, 3 orthogonal X components and 3 orthogonal Y components; $R^2(X) = 99\%$;
$R^2(Y) = 99\%$; $Q^2 = 100\%$). This was driven by changes in the correlation of serine ($p =
0.046$), citrulline ($p = 0.030$), taurine ($p = 0.033$), aminoisobutyric acid ($p = 0.042$),
homocysteine ($p = 0.039$), GABA ($p = 0.046$), histamine ($p = 0.024$), 3-hydroxyanthranilic
acid ($p = 0.03$), betaine ($p = 0.032$), α-glycerophosphocholine ($p = 0.048$), creatine ($p =
0.036$), cytosine ($p = 0.02$), xanthosine ($p = 0.034$), beta-hydroxyisovaleric acid ($p = 0.03$),
adenine ($p = 0.04$), cytidine ($p = 0.044$), SAH ($p = 0.41$), N-acetylglutamate ($p = 0.036$, N-
acetylglutamine ($p = 0.038$) (**Supplementary Figure 1b**).

Examining changes in the gut metabolome

To investigate whether changes in the mouse metabolome could be related to alterations in
the gut metabolome we examined the metabolite composition of faeces from NOD and
NOD-E mice using GC-MS to examine the total fatty acid compliment and LC-MS/MS to
examine the aqueous metabolome. While no PLS-DA or OPLS-DA models could be built
separating the spectral profiles at 6 weeks or 12 weeks of age (n=6 for both groups), a time
trend was detected using PLS considering either all mice (for total fatty acids: 2 components;
$R^2(X) = 0.42$, $R^2(Y) = 0.56$, $Q^2 = 0.25$); for aqueous metabolites: 4 components, $R^2(X) = 0.71$,
$R^2(Y) = 0.88$, $Q^2 = 0.37$) (data not shown) or NOD mice (for total fatty acids: 2 components
$R^2(X) = 0.45$, $R^2(Y) = 0.88$, $Q^2 = 0.62$; for aqueous metabolites: 2 components, $R^2(X)=0.55$,
$R^2(Y) = 0.74$, $Q^2 = 0.38$) (**Supplementary figure 2**). For the NOD mice this was driven by an
increase in C14:1, C18:3n-6 and C18:3n-3 fatty acids, serotonin, guanosine, creatine and
trimethylamine, and a decrease in C16:0, C18:0 and C20:0 fatty acids, ADMA, taurine,
citrate and α-ketoisocaproic acid.

Discussion:

The NOD mouse and its related strains have been major research tools for
understanding how T1DM arises across the natural life course of the animal [38]. The mouse
develops a number of autoantibodies as it ages including those recognising autoantigens
insulin, IA-2 and glutamic acid decarboxylase that, as in humans, provide markers of

ongoing beta cell destruction. The NOD mouse has been the subject of a number of previous metabolomic studies which have used the reproducible time course and incidence of T1DM in this model to follow metabolic changes that accompany the development of T1DM [16,18,39-41]. However, these studies have focussed solely on NOD mice, and thus it is difficult to disentangle any metabolic changes that are associated with ageing in the animal from the subsequent age related changes associated with the development of T1DM. To address this Fahrmann and colleagues [18] compared NOD mice that developed T1DM with those that did not across the time course. In the present study we took an alternative strategy and compared the metabolic profiles of blood plasma and pancreatic tissue from NOD, NOD-E and NOD-SCID mice, the latter two strains being protected from the development of T1DM compared with NOD mice.

We compared both lipid and aqueous metabolism changes in both blood plasma and pancreas tissue across the three mouse models, and a striking result was the discrimination of NOD-SCID mice from the NOD and NOD-E strains. In terms of both intact lipids and the core metabolic pathways measured in our aqueous assays, NOD-SCID mice were readily discriminated from the other two strains and in cross validation of the associated pattern recognition models, NOD-SCID mice could be classified with 100% success according to their metabolic profiles regardless of age. The profound metabolic differences, which discriminate the NOD-SCID mice from the other two strains indicate that this model is a poor control strain to compare with NOD mice when considering the metabolic changes that accompany T1DM. As a result we have focussed on the discrimination between NOD and NOD-E mice in terms of the metabolic changes that accompany the development of T1DM. We focussed on female mice given the sexual dimorphism of the risk of developing T1DM. Females mice are more likely to develop T1DM and this is thought to be due to both a sex-specific, epistatic interaction with Idd4.1 in females [42] and differential interactions with the gut microflora in male and female NOD mice [43].

Across both lipid and aqueous metabolites datasets in both blood plasma and the pancreas we could discriminate NOD and NOD-E mice readily, and indeed in NOD mice younger than 14 weeks, and hence prior to the onset of T1DM, it was possible to discriminate mice according to their lipid and aqueous profiles from the NOD-E strain. Examining the RIP CD154 x RAG-/- mouse, a different model for T1DM produced by expressing CD154 under the rat insulin promoter crossed into the immuno-deficient recombination-activating gene (RAG) knockout mouse, Overgaard and colleagues have also shown that metabolic changes pre-date the development of overt T1DM. Indeed some of the metabolic changes were similar between the CD154 x RAG-/- mouse and the NOD mouse with decreases in lysophosphatidylcholines and decreased methionine (detected in the

pancreas of NOD mice), as well as increases in ceramides in the CD154 x RAG-/- mouse [15]. Indeed, La Torre and colleagues have even detected decreases in phosphatidylcholines and phosphatidylethanolamines in cord blood of children at birth who subsequently go on to develop T1DM compared with children matched for HLA risk, sex, date of birth and gestational age but who do not develop the disease [17,44]. However, it was not possible to discriminate between NOD mice that did not develop autoimmune diabetes from NOD-E mice at 25-30 weeks, suggesting that metabolism normalises for those NOD mice that avoid autoimmune diabetes in the normal disease aetiology.

In terms of different classes of metabolites, one of the biggest changes was associated with amino acid metabolism, both in terms of discrimination of the three strains of mice and the progression of the disease in NOD mice. Alterations in amino acid metabolism have been reported in a number of previous studies. Lu and co-workers have previously identified alterations in 5-oxoproline, glutamate and homocysteine as being common to T1DM, type 2 diabetes (T2DM) and fulminant type 1 diabetes in patients compared with a control group [45]. Deja and co-workers correlated increased concentrations of alanine and valine with increased concentrations of glycated haemoglobin (HbA1c) using ^{1}H NMR spectroscopy of urine [46]. Indeed, alterations in amino acid metabolism appears to be a central feature of both T2DM and other rarer forms of diabetes. Perhaps the most robust metabolic markers for predicting future T2DM incidence are branch chain amino acids, and Patel and colleagues, investigating A-β+ ketosis-prone diabetes using stable isotope techniques to follow metabolic flux noted that although sufferers had normal oxidation rates of fatty acids and acetyl-CoA, there was increased leucine oxidation, conversion of glutamine to glutamate but impaired conversion of glutamate to 2-oxoglutarate and reduced flux into the citrulline cycle [47]. In the present study blood plasma from NOD mice were characterised by increases in arginosuccinate and a reduction in ornithine, similarly indicating alterations in the citrulline cycle consistent with a reduction in urea production. The turnover of the citrulline cycle, and in particular metabolites around arginine, suggests that there may be increased NO production and could be associated with activated macrophages and subsequent pathology. Xanthine, a purine breakdown product and a precursor for uric acid, was also found to be increased in the pancreas of NOD mice. Increased purine catabolism and the induction of xanthine oxidase has previously been linked to T1DM associated diabetic nephropathy in streptozotocin-dosed rats, presumably through the generation of ROS species in the kidneys [48]. Further evidence of dysfunction in the kidneys is suggested by the increased creatinine concentrations in the blood plasma of NOD mice in the present study.

It has previously been observed that nicotinamide has the ability to both prevent T1DM in prediabetic NOD mice and reverse the pathology in mice that exhibit hyperglycemia [38], and one of the major changes detected in both the pancreas and blood plasma in the current study involved changes in the precursors of nicotinamide, NAD^+ and $NADP^+$. This included an increase in the niacin precursor kynurenic and decreases in serotonin and 5-HT in blood plasma of NOD mice and increases in xanthinine and NADH, and decreases in xanthurenate in the pancreas of NOD mice. The tryptophan pathway produces either kynurenic acid via indoleamine 2,3-dioxygenase (IDO) which ultimately can be used to synthesize NAD^+ for energy generation, or 5-HT and serotonin via tryptophan hydroxylase, which can be used to synthesize melatonin (**Figure 5b**). Considering the three pathways that branch off tryptophan, while NOD-E mice demonstrated high correlations between kynurenine, kynurenic acid and xanthurenic acid, the correlation profiles were more equally spread across metabolites derived from tryptophan, including the NAD-synthesis pathway. This suggests that different regulatory mechanisms are in operation between the two mouse strains. While no mechanism has been identified that links increased nicotinamide metabolism to the development of T1DM in the NOD mouse, the generation of reactive oxygen species in mitochondria, mitochondrial DNA damage and genetic interactions with variants in nicotinamide adenine dinucleotide hydrogen dehydrogenase subunit 2 have all been linked to increased relative risk of T1DM in the NOD mouse [38]. Furthermore, inflammation is known to stimulate the expression of IDO in dendritic cells to increase kynurenic acid concentration [49], and alterations in this pathway have previously been detected in patients with T1DM [50]. In addition, the tryptophan pathway has been highlighted in a number of chronic diseases as well as ageing [51].

Fahrmann and colleagues [18] previously described how changes in ROS species and their associated products accompany the development of T1DM in the NOD mouse. While we did not find changes in the production of ROS products, such as oxidized amino acids and nucleotides, in the present study a reduction in methionine was detected in the pancreas of NOD mice suggesting they may have increased oxidative stress compared with NOD-E mice.

A wide range of lipid species were altered across the three strains and across the time courses, and a number of previous studies have highlighted altered lipid metabolism both pre and post overt T1DM [8,17,52,53]. Alterations in lipid metabolism in female NOD mice has previously been reported by Markle and colleagues [23] who reported alterations in sphingolipid metabolism in female NOD mice who have a higher incidence of T1DM compared with male mice. These changes could be corrected if the female mice were given a faecal transplant from male mice, indicating these changes were in part driven by the gut

microflora. However, sphingolipids are not found in bacteria, suggesting the lipid changes were a downstream effect rather than something that directly acts on sphingolipid metabolism. Furthermore, Makinen and colleagues have also suggested altered sphingolipid metabolism in patients with type 1 diabetes-associated kidney disease [54]. While we did not detect changes in sphingolipid metabolism in blood plasma, there were a number of sphingolipid species altered in the pancreas – being one of the classes of lipids that most readily discriminated NOD-SCID from NOD and NOD-E mice. In addition SM(39:1) was increased in the pancreas of NOD mice and Cer (d34:1) in the pancreas of NOD-E mice.

In blood plasma, the NOD mice were clearly characterised by increases in triglycerides containing longer, polyunsaturated fatty acids, and relative decreases in lysophospholipids containing longer polyunsaturated acids as well as cholesterol esters. Alterations in phospholipids, especially lysophospholipids, and cholesterol esters have previously been described in T1DM with well controlled diabetes compared with healthy controls [17,52,55,56], with Fievet and colleagues explaining this as arising from lipoprotein remodelling and an increased pro-inflammatory state for those with T1DM. Indeed, in children autoantibodies to a number of lysophospholipids have also been described [57], demonstrating the interplay between this class of lipid and autoimmune disease.

Conclusions:

In summary, we have used a combination of lipidomics and metabolomics to profile the development of T1DM in the NOD mouse and compare these changes with NOD-E and NOD-SCID mice, two strains that are protected from T1DM. This approach has highlighted that the NOD-SCID mouse provides a poor control for the NOD mouse in terms of metabolic changes, and indeed the most profound metabolic differences between the mouse strains were between NOD-SCID and the other two strains, regardless of diabetic status, underlining how different the NOD-SCID mice are. When comparing the NOD and NOD-E mice the most consistent metabolic changes are in the tryptophan pathway and in particular in the synthesis of kynurenic acid derivatives, and a reduction in lysophospholipids. Both pathways have previously been implicated in modulating immune responses, underlying the interplay between metabolism and inflammation. As well as highlighting metabolic mechanisms that could be targeted to treat T1DM, such metabolic markers of disease could be used in the future to target immune-suppression therapies to reduce the likelihood of developing T1DM in at risk populations.

Supporting Information:

Author contributions: JG and AC wrote the manuscript. PZ, XW and YS performed animal experiments. SAM and JLG performed sample preparation, GC-MS and NMR spectroscopy. AA and JLG performed data processing. AK, JLG and LDR performed LC-MS. All authors read and approved the final manuscript.

References:

(1)   Kasper, D. L., Braunwald, E., Fauci, A.S., Hauser, S.L., Jameson, J.L.: *Harrison's Principles of Internal Medicine*; 16th ed.; McGraw-Hill: New York, 2005.

(2)   Soltesz, G., Patterson, C.C., Dahlquist, G. Worldwide childhood type 1 diabetes incidence—what can we learn from epidemiology? *Pediatric diabetes* **2007**, *8*, 6-14.

(3)   Knip, M., Veijola, R., Virtanen, S.M., Hyöty, H., Vaarala, O., Akerblom, H.K. Environmental Triggers and Determinants of Type 1 Diabetes. *Diabetes* **2005**, *54*, S125-136.

(4)   Raine, T., Zaccone, P., Mastroeni, P., Cooke, A. Salmonella typhimurium infection in nonobese diabetic mice generates immunomodulatory dendritic cells able to prevent type 1 diabetes. *J Immunol* **2006**, *177*, 2224-2233.

(5)   Cooke, A. Review series on helminths, immune modulation and the hygiene hypothesis: how might infection modulate the onset of type 1 diabetes? . *Immunology* **2009**, *126*, 12-17.

(6)   Griffin, J. L. Metabolic profiles to define the genome: can we hear the phenotypes? *Philos Trans R Soc Lond B Biol Sci* **2004**, *359*, 857-871.

(7)   Zhang, S., Nagana Gowda, G.A., Asiago, V., Shanaiah, N., Barbas, C., Raftery, D. Correlative and quantitative 1H NMR-based metabolomics reveals specific metabolic pathway disturbances in diabetic rats. *Anal Biochem* **2008**, *383*, 76-84.

(8)   Lanza, I. R., Zhang, S., Ward, L.E., Karakelides, H., Raftery, D., Nair, K.S. Quantitative metabolomics by H-NMR and LC-MS/MS confirms altered metabolic pathways in diabetes. *PLoS One* **2010**, *5*, e10538.

(9)   Dutta, T., Chai, H.S., Ward, L.E., Ghosh, A., Persson, X.M., Ford, G.C., Kudva, Y.C., Sun, Z., Asmann, Y.W., Kocher, J.P., Nair, K.S. Concordance of changes in metabolic pathways based on plasma metabolomics and skeletal muscle transcriptomics in type 1 diabetes. *Diabetes* **2012**, *61*, 1004-1016.

(10)   Mäkinen, V. P., Soininen, P., Forsblom, C., Parkkonen, M., Ingman, P., Kaski, K., Groop, P.H.; FinnDiane Study Group. . 1H NMR metabonomics approach to the disease continuum of diabetic complications and premature death. *Mol Syst Biol* **2008**, *4*, 167.

(11)   Mäkinen, V. P., Forsblom, C., Thorn, L.M., Wadén, J., Gordin, D., Heikkilä, O., Hietala, K., Kyllönen, L., Kytö, J., Rosengård-Bärlund, M., Saraheimo, M., Tolonen, N., Parkkonen, M., Kaski, K., Ala-Korpela, M., Groop, P.H.; FinnDiane Study Group. Metabolic phenotypes, vascular complications, and premature deaths in a population of 4,197 patients with type 1 diabetes. *Diabetes* **2008**, *57*, 2480-2487.

(12)   van der Kloet, F. M., Tempels, F.W., Ismail, N., van der Heijden, R., Kasper, P.T., Rojas-Cherto, M., van Doorn, R., Spijksma, G., Koek, M., van der Greef, J., Mäkinen, V.P., Forsblom, C., Holthöfer, H., Groop, P.H., Reijmers, T.H., Hankemeier, T. Discovery of early-stage biomarkers for diabetic kidney disease using ms-based metabolomics (FinnDiane study). . *Metabolomics* **2012**, *8*, 109-119.

(13)   Roberts, L. D., McCombie, G., Titman, C.M., Griffin, J.L. A matter of fat: an introduction to lipidomic profiling methods. *J Chromatogr B Analyt Technol Biomed Life Sci.* **2008**, *871*, 174-181.

(14)   Oresic, M., Simell, S., Sysi-Aho, M., Näntö-Salonen, K., Seppänen-Laakso, T., Parikka, V., Katajamaa, M., Hekkala, A., Mattila, I., Keskinen, P., Yetukuri, L., Reinikainen, A., Lähde, J., Suortti, T., Hakalax, J., Simell, T., Hyöty, H., Veijola, R., Ilonen, J., Lahesmaa, R., Knip, M., Simell, O. Dysregulation of lipid and amino acid metabolism precedes islet autoimmunity in children who later progress to type 1 diabetes. *J Exp Med* **2008**, *205*, 2975-2984.

(15)   Overgaard, A. J., Weir, J.M., De Souza, D.P., Tull, D., Haase, C., Meikle, P.J., Pociot, F. Lipidomic and metabolomic characterization of a genetically modified mouse model of the early stages of human type 1 diabetes pathogenesis. *Metabolomics* **2016**, *12*, 13.

(16)   Sysi-Aho, M., Ermolov, A., Gopalacharyulu, P.V., Tripathi, A., Seppänen-Laakso, T., Maukonen, J., Mattila, I., Ruohonen, S.T., Vähätalo, L., Yetukuri, L., Härkönen, T., Lindfors, E., Nikkilä, J., Ilonen, J., Simell, O., Saarela, M., Knip, M., Kaski, S., Savontaus, E., Orešič, M. Metabolic regulation in progression to autoimmune diabetes. *PLoS Comput Biol* **2011**, *7*, e1002257.

(17)   La Torre, D., Seppänen-Laakso, T., Larsson, H.E., Hyötyläinen, T., Ivarsson, S.A., Lernmark, A., Oresic, M.; the DiPiS study group. Decreased cord-blood phospholipids in young age at onset type 1 diabetes. *Diabetes* **2013**, *62*, 3951-3956.

(18)   Fahrmann, J., Grapov, D., Yang, J., Hammock, B., Fiehn, O., Bell, G.I., Hara, M. Systemic alterations in the metabolome of diabetic NOD mice delineate increased oxidative stress accompanied by reduced inflammation and hypertriglyceremia. *Am J Physiol Endocrinol Metab.* **2015**, *308*, E978-989.

(19)   Ramirez, L., Hamad, A.R. Status of autoimmune diabetes 20-year after generation of BDC2.5-TCR transgenic non-obese diabetic mouse. *World J Diabetes* **2013**, *4*, 88-91.

(20)   Kolb H, B. V., Appels B, Hanenberg H, Kant-werk-Funke G, Kiesel U, Funda J, Schraermeyer U, Kolb-Bachofen V. . Essential contribution of macrophages to islet cell destruction in vivo and in vitro. *J Autoimmun* **1990**, *3*, 117-120.

(21)   Jansen A, H.-D. F., Hooijkaas H, Leenen PJ, Dardenne M, Drexhage HA. Immunohistochemical characterization of monocytes-macrophages and dendritic cells involved in the initiation of the insulitis and beta-cell destruction in NOD mice. *Diabetes* **1994**, *43*, 667-675.

(22)   Fox CJ, D. J. Independent genetic regulation of T-cell and antigen-presenting cell participation in autoimmune islet inflammation. *Diabetes* **1998**, *47*, 331-338.

(23)   Markle JGM, F. D., Adeli K, von Bergen M, Danska J.S Microbiome manipulation modifies sex-specific risk for autoimmunity. *Gut* **2014**, *5*, 485-493.

(24)   Lehuen, A., Diana, J., Zaccone, P., Cooke, A. Immune cell crosstalk in type 1 diabetes. *Nat Rev Immunol.* **2010**, *10*, 501-513.

(25)   Parish, N. M., Acha-Orbea, H., Simpson, E., Qin, S.X., Lund, T., Cooke, A. A comparative study of T-cell receptor V beta usage in non-obese diabetic (NOD) and I-E transgenic NOD mice. *Immunology* **1993**, *78*, 606-610.

(26)   Prochazka, M., Gaskins, H.R., Shultz, L.D., Leiter, E.H. . The nonobese diabetic scid mouse: model for spontaneous thymomagenesis associated with immunodeficiency. *Proc Natl Acad Sci USA* **1992**, *89*, 3290-3294.

(27)   Bligh, E. G., Dyer, W.J. . A rapid method of total lipid extraction and purification. *Can J Biochem Physiol* **1959**, *37*, 911-917.

(28)   Smith, C. A., Want, E.J., O'Maille, G., Abagyan, R., Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **2006**, *78*, 779-787.

(29)   Race, A. M., Styles, I.B., Bunch, J. Inclusive sharing of mass spectrometry imaging data requires a converter for all. *J Proteomics* **2012**, *75*, 5111-5112.

(30)   Tautenhahn, R., Böttcher, C., Neumann, S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* **2008**, *9*, 504.

(31)   Sumner, L. W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A., Fan, T.W., Fiehn, O., Goodacre, R., Griffin, J.L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A.N., Lindon, J.C., Marriott, P., Nicholls, A.W., Reily, M.D., Thaden, J.J., Viant, M.R. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **2007**, *3*, 211-221.

(32)   Roberts, L. D., Souza, A.L., Gerszten, R.E., Clish, C.B. Targeted metabolomics. *Curr Protoc Mol Biol* **2012**, *32*, 11-24.

(33)   Roberts, L. D., Bostrom, P., O'Sullivan, J.F., Schinzel, R.T., Lewis, G.D., Dejam, A., et al. Beta-Aminoisobutyric Acid Induces Browning of White Fat and Hepatic beta-Oxidation and Is Inversely Correlated with Cardiometabolic Risk Factors. *Cell Metab.* **2014**, *19*, 96-108.

(34)   Roberts, L. D., West, J.A., Vidal-Puig, A., Griffin, J.L. Methods for performing lipidomics in white adipose tissue. *Methods Enzymol.* **2014**, *538*, 211-231.

(35)   Castro, C., Kumsiek, J., Lehrbach, N.J., Murfitt, S.A., Miska, E.A, Griffin, J.L. A study of Caenorhabditis elegans DAF-2 mutants by metabolomics and differential correlation networks *Molecular BioSystems* **2013**, *9*, 1632-1642.

(36)   Eriksson, L., Trygg, J., Wold, S. CV-ANOVA for significance testing of PLS and OPLS models. *Journal of Chemometrics* **2008**, *22*, 594-600.

(37)   Xia, J.; Psychogios, N.; Young, N.; Wishart, D. S. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acid Res* **2009**, *37*, W652-660.

(38)   Driver, J. P., Chen, Y-G, Matthews, C.E. Comparative genetics: synergizing human and NOD mouse studies for identifying genetic causation of type 1 diabetes. *Rev Diabet Stud.* **2012**, *9*, 169-187.

(39)   Sysi-Aho, M., Ermolov, A., Gopalacharyulu, P.V., Tripathi, A., Seppänen-Laakso, T., Maukonen, J., Mattila, I., Ruohonen, S.T., Vähätalo, L., Yetukuri, L., Härkönen, T., Lindfors, E., Nikkilä, J., Ilonen, J., Simell, O., Saarela, M., Knip, M., Kaski, S., Savontaus, E., Orešič, M. . Metabolic regulation in progression to autoimmune diabetes. *PLoS Comput Biol.* **2011**, *7*, e1002257.

(40)   Greiner, T. U., Hyötyläinen, T., Knip, M., Bäckhed, F., Orešič, M. The gut microbiota modulates glycaemic control and serum metabolite profiles in non-obese diabetic mice. *PLoS One* **2014**, *9*, e110359.

(41)   Grapov, D., Fahrmann, J., Hwang, J., Poudel, A., Jo, J., Periwal, V., Fiehn, O., Hara, M. Diabetes Associated Metabolomic Perturbations in NOD Mice. *Metabolomics* **2015**, *11*, 425-437.

(42)   Ivakine, E. A., Mortin-Toth, S.M., Gulban, O.M., Valova, A., Canty, A., Scott, C., Danska, J.S. The idd4 locus displays sex-specific epistatic effects on type 1 diabetes susceptibility in nonobese diabetic mice. *Diabetes* **2006**, *55*, 3611-3619.

(43)   Markle, J. G., Frank, D.N., Mortin-Toth, S., Robertson, C.E., Feazel, L.M., Rolle-Kampczyk, U., von Bergen, M., McCoy, K.D., Macpherson, A.J., Danska, J.S. Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. *Science* **2013**, *339*, 1084-1088.

(44)   La Torre, D., Seppänen-Laakso, T., Larsson, H.E., Hyötyläinen, T., Ivarsson, S.A., Lernmark, A., Oresic, M.; DiPiS Study Group. Decreased cord-blood phospholipids in young age-at-onset type 1 diabetes. *Diabetes* **2013**, *62*, 3951-3956.

(45)   Lu J, Z. J., Bao Y, Chen T, Zhang Y, Zhao A, Qiu Y, Xie G, Wang C, Jia W, Jia W. Serum metabolic signatures of fulminant type 1 diabetes. *J Proteome Res* **2012**, *11*, 4705-4711.

(46)   Deja, S., Barg, E., Młynarz, P., Basiak, A., Willak-Janc, E. 1H NMR-based metabolomics studies of urine reveal differences between type 1 diabetic patients with high and low HbAc1 values. *J Pharm Biomed Anal* **2013**, *83*, 43-48.

(47)   Patel, S. G., Hsu, J.W., Jahoor, F., Coraza, I., Bain, J.R., Stevens, R.D., Iyer, D., Nalini, R., Ozer, K., Hampe, C.S., Newgard, C.B., Balasubramanyam, A. Pathogenesis of A⁻β⁺ ketosis-prone diabetes. *Diabetes* **2013**, *62*, 912-922.

(48)   Liu, J., Wang, C., Liu, F., Lu, Y., Cheng, J. Metabonomics revealed xanthine oxidase-induced oxidative stress and inflammation in the pathogenesis of diabetic nephropathy. *Anal Bioanal Chem* **2015**, *407*, 2569-2579.

(49)   van der Goot, A., Nollen EAA. Tryptophan metabolism: entering the field of aging and age-related pathologies. , Trends in Molecular Medicine, 2013, 19(6), 336-344. *Trends in Molecular Medicine* **2013**, *19*, 336-344.

(50)   Oxenkrug, G., van der Hart, M., Summergrad, P. . Elevated anthranilic acid plasma concentrations in type 1 but not type 2 diabetes mellitus. *Integr Mol Med* **2015**, *2*, 365-368.

(51)   van der Goot, A. T., Nollen, E.A. Tryptophan metabolism: entering the field of aging and age-related pathologies. *Trends in Molecular Medicine* **2013**, *19*, 336-344.

(52)   Balderas, C., Rupérez, F.J., Ibañez, E., Señorans, J., Guerrero-Fernández, J., Casado, I.G., Gracia-Bouthelier, R., García, A., Barbas, C. Plasma and urine metabolic fingerprinting of type 1 diabetic children. *Electrophoresis* **2013**, *34*, 2882-2890.

(53)   Godzien, J., Ciborowski, M., Whiley, L., Legido-Quigley, C., Ruperez, F.J., Barbas C. In-vial dual extraction liquid chromatography coupled to mass spectrometry applied to streptozotocin-treated diabetic rats. Tips and pitfalls of the method. *J Chromatogr A* **2013**, *1304*, 52-60.

(54)   Mäkinen VP, T. T., Soininen P, Forsblom C, Peltola T, Kangas AJ, Groop PH, Ala-Korpela M. Sphingomyelin is associated with kidney disease in type 1 diabetes (The FinnDiane Study). *Metabolomics* **2012**, *8*, 369-375.

(55)   Fievet, C., Ziegler, O., Parra, H.J., Mejean, L., Fruchart, J.C., Drouin, P. Depletion in choline containing phospholipids of LpB particles in adequately controlled type I insulin-dependent diabetes mellitus. *Diabete Metab* **1990**, *16*, 64-69.

(56)   Rabini, R. A., Galassi, R., Fumelli, P., Dousset, N., Solera, M.L., Valdiguie, P., Curatola, G., Ferretti, G., Taus, M., Mazzanti, L. Reduced Na(+)-K(+)-ATPase activity and plasma lysophosphatidylcholine concentrations in diabetic patients. *Diabetes* **1994**, *43*, 915-919.

(57)   Bleich, D., Polak, M., Chen, S., Swiderek, K.M., Lévy-Marchal, C. Sera from children with type 1 diabetes mellitus react against a new group of antigens composed of lysophospholipids. *Horm Res.* **1999**, *52*, 86-94.

Figure legends

Figure 1: **A:** High resolution mass spectrum of a typical lipid extract from blood plasma derived from the NOD mouse. **B.** OPLS-DA plot of lipidomics of blood plasma from NOD, NOD-E and NOD-SCID mice considered as separate groups. Data is Pareto scaled. $R^2(X) = 75\%$, $R^2(Y) = 48\%$; $Q^2 = 34\%$; $p = 3.5*10^{-22}$ by CV-ANOVA for cross validation. 2 components and 6 orthogonal components. **C.** OPLS-DA plot of lipidomics of blood plasma from NOD-SCID mice compared with a combined group of NOD and NOD-E mice (labelled Both). Data is Pareto scaled. $R^2(X) = 69\%$; $R^2(Y) = 68\%$; $Q^2 = 51\%$; $p = 5.8*10^{-34}$. 1 component and 6 orthogonal components. **D.** OPLS-DA plot of lipidomics of blood plasma from NOD and NOD-E mice. Data is Pareto scaled. $R^2(X) = 57\%$, $R^2(Y) = 41\%$, $Q^2 = 23\%$; $p = 5.4*10^{-9}$, 1 component and 3 orthogonal components. **E.** OPLS-DA plot of lipidomics of blood plasma from NOD and NOD-E mice under 14 weeks of age. Data is Pareto scaled. $R^2(X) = 43\%$, $R^2(Y) = 32\%$, $Q^2 = 23\%$, $p = 1.2*10^{-7}$, 1 component and 1 orthogonal component. **F.** OPLS-DA plot of lipidomics of blood plasma from NOD and NOD-E mice aged 14 weeks and over. Data is Pareto scaled. $R^2(X) = 56\%$, $R^2(Y) = 80\%$, $Q^2 = 54\%$; $p=2.2*10^{-5}$, 1 component and 4 orthogonal components. **G**. OPLS-DA plot of lipidomics of blood plasma from NOD (diabetic) and NOD (non-diabetic referred to as NODSTAR) mice aged 25-30 weeks and over. $R^2(X) = 68\%$; $R^2(Y) = 79\%$; $Q^2 = 56\%$; $p = 2.5*10^{-11}$, 1 component and 4 orthogonal components.

Figure 2: A. OPLS-DA plot of total fatty acid analysis of blood plasma from NOD, NOD-E and NOD-SCID mice considered as separate groups. Data is Univariate scaled. $R^2(X) = 25\%$; $R^2(Y) = 47\%$; $Q^2 = 33\%$. **B.** OPLS-DA plot of total fatty acid analysis of blood plasma from NOD-SCID mice compared with a combined group of NOD and NOD-E mice (labelled Both). Data is univariate scaled. $R^2(X) = 40\%$; $R^2(Y) = 80\%$; $Q^2 = 72\%$; $p = 6.3*10^{-37}$. **C.** S-plot of Pareto scaled variables from the dataset analysed for the plot in **B**. **D.** PLS plot of age against metabolic profile of NOD mice. $Q^2 = 51\%$, $R^2(X) = 27\%$, $R^2Y = 77\%$; $p=4*10^{-6}$.

Figure 3: A. OPLS-DA plot of targeted analysis of central metabolism (amino acids, TCA cycle intermediates, nucleotides, and other small aqueous metabolites) of blood plasma from NOD, NOD-E and NOD-SCID mice considered as separate groups. Data is Univariate scaled. $R^2(X) = 56\%$, $R^2(Y) = 72\%$, $Q^2 = 52\%$, 2 components and 4 orthogonal components; CV ANOVA validation $p = 1.5*10^{-22}$. **B.** OPLS-DA plot of targeted analysis of central

metabolism of blood plasma from NOD-SCID mice compared with the NOD and NOD-E mice as a single group. R^2(X) = 42%, R^2(Y) = 79%, Q^2 = 79%, 1 component and 2 orthogonal components; CV-ANOVA validation p = 2.7*10^{-38}. **C.** Variable Importance Parameter (VIP) plot to show key metabolites that are most important for the discrimination for the plot in **B**. **D.** OPLS-DA plot of targeted analysis of central metabolism of blood plasma from NOD and NOD-E mice. Data is Univariate scaled. R^2(X) = 51%; R^2(Y) = 76%; CV-ANOVA validation p = 2*10^{-10}. **E.** VIP plot to show key metabolites that are most important for the discrimination for the plot in **D**. **F.** OPLS-DA plot of targeted analysis of central metabolism of blood plasma from NOD and NOD-E mice below 14 weeks of age. Data is Univariate scaled. R^2(X) = 44%, R^2(Y) = 84%, Q^2 = 74%, 1 component and 2 orthogonal components; CV ANOVA validation p = 2.7*10^{-38}. **G.** OPLS-DA plot of targeted analysis of central metabolism of blood plasma from NOD and NOD-E mice 14 weeks and older of age. Data is Univariate scaled. R^2(X) = 64%, R^2(Y) = 82%, Q^2 = 51%, 1 component and 4 orthogonal components; CV-ANOVA validation p = 2.7*10^{-38}. **H.** SUS-plot of discriminatory metabolites for models described in **F** and **G**.

Figure 4. A: PLS model of targeted analysis of central metabolism of blood plasma from NOD mice against age (R^2X = 53%, R^2Y = 79%, Q^2 = 65%). **B:** SUS-plot of PLS models for NOD mice and NOD-E mice against age. In this plot the loadings for the two PLS models are plotted against each other so that if the two models had the same loadings then these would line up at 45° to the x and y-axes. While some metabolites are in common there is also a degree of scatter from the 45° line showing that other metabolites are responsible for the ageing trends in the NOD and NOD-E mice.

Figure 5. A: Pearson correlation plots for the tryptophan pathway for NOD and NOD-E mice. Correlations were calculated for each pair of metabolites with the circle representing the magnitude of the correlation coefficient. **B:** The tryptophan pathway mapped in the correlation analysis in **A**. Key: 5-HT 5-hydroxytryptophan, 5-HIAA 5-Hydroxyindole acetic acid.

Figure 6. A: High resolution mass spectrum of a typical pancreas extract from the NOD mouse. **B:** OPLS-DA plot of lipidomics of the pancreas comparing NOD (red circles), NOD-E (green circles) and NOD-SCID (yellow circles) mice. Data is univariate scaled. R^2(X) = 79%,

$R^2(Y) = 89\%$; $Q^2 = 61\%$; p = $4.5*10^{-12}$ by CV-ANOVA for cross validation. **C:** OPLS-DA plot of lipidomics of the pancreas comparing NOD-SCID mice (yellow circles) compared with the other NOD and NOD-E mice (red circles) treated as a single group. $R^2(X) = 74\%$; $R^2(Y) = 92\%$; $Q^2 = 79\%$; p = $1.3*10^{-13}$. **D**. OPLS-DA plot of lipidomics of the pancreas comparing NOD (red circles) and NOD-E (green circles) mice across all ages. $R^2(X) = 76\%$; $R^2(Y) = 92\%$; $Q^2 = 60\%$; p = $7.6*10^{-9}$. **E.** OPLS-DA plot of lipidomics of the pancreas comparing NOD and NOD-E mice under 14 weeks of age. $R^2(X) = 49\%$; $R^2(Y) = 81\%$; $Q^2 = 63\%$; p = $5.2*10^{-8}$.

Figure 7: A: OPLS-DA plot of targeted core metabolism of extracts of pancreas tissue comparing the NOD-SCID mouse (green circles) to a combined group of NOD and NOD-E mice (red circles), regardless of age. Data is univariate scaled. $R^2(X) = 37\%$; $R^2(Y) = 96\%$; $Q^2 = 82\%$; CV-ANOVA p = $2.8*10^{-6}$. **B:** S-plot of most discriminatory metabolites in a Pareto scaled OPLS-DA model using the data analysed in **A**. **C:** OPLS-DA plot of targeted core metabolism of extracts of pancreas tissue comparing the NOD mouse (red circles) to NOD-E mice (green circles), regardless of age. Data is univariate scaled. $R^2(X) = 42\%$; $R^2(Y) = 99\%$; $Q^2 = 70\%$; CV-ANOVA p = 0.04. **D:** S-plot of most discriminatory metabolites in a Pareto scaled OPLS-DA model using the data analysed in **C**.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For TOC only