Introduction
Metabolomics is a rapidly growing field of post-genomic biology, aiming to comprehensively characterize the small molecules in biological systems. Nonbiological systematic biases from instrument calibration or the order of sample injection account for the most significant errors in LC-TOF-MS data [1]. Here we present a workflow using a UHPLC/SIEVE quadrupole Orbitrap platform and automated data analysis software for untargeted metabolic profiling of plasma samples for biomarker discovery from the Zucker diabetes fatty (ZDF) rat model. The optimal conditions for sample preparation, liquid chromatography (LC), column, mass spectrometry (MS), and data processing parameters are explored.

Methods
Sample Preparation
Plasma samples were deproteinized with organic solvent. Four extraction solvent systems including methanol (MeOH), acetonitrile (ACN), acetone, and 1:1:1 of the above were tested in this work. Endogenous metabolites were reconstituted in methanol/water (19) containing isotopically labeled internal standard (IS), d5-hippuric acid for LC-MS analysis. Solvent blank, pooled QC, and biological samples were analyzed in a randomized injection order.

Liquid Chromatography (or more genetically Separations)
UHPLC separation was performed on a Thermo Scientific™ Dionex™ UltiMate™ 3000 HPG (high-pressure gradient) pump using Thermo Scientific™ HyperSil GOLD™ RP C18 column at 450 μL/min, column temperature at 55 °C. LC-solvents were 0.1% FA (A) and 0.1% FA in MeOH (B). Apply linear gradient from 0.5–50% B for 5.5 min, followed by increasing to 98% at 6 min, hold 98% B for 6 min, then decrease to 0.5% at 13 min, then equilibrate for another 2 min.

Mass Spectrometry
The Thermo Scientific™ Q Exactive™ mass spectrometer was operated under electrospray ionization (ESI) positive, and polarity switching modes. Full scan (m/z 67–1050) used resolution 70,000 with automatic gain control (AGC) target of 1×10^5 ions and a maximum ion injection time (IT) of 35 ms. Data-dependent MS/MS were acquired on a “Top5” data-dependent mode using the following parameters: resolution 17,500, AGC 1×10^5 ions; maximum IT 80 ms; 2.0 amu isolation window; normalized collision energy 35% ± 50% with a 1.0% Apex trigger 2–4 s, and dynamic exclusion 6 s. Source ionization parameters were: spray voltage, 3.9 kV; capillary temperature, 300 °C; and S-Lens level, 45.

Data Analysis
Differential analyses of the obese versus lean plasma were performed using Thermo Scientific™ SIEVE™ 2.1 software which executes background subtraction, component detection, peak alignment, differential analysis (Figure 1). Statistical results, putative IDs, and pathways were obtained through searching ChemSpider and KEGG™. Metabolites of interest were identified via MS/MS mass spectral database matching. The raw files were converted to mzCloud format using ProteoWizard and also analyzed by XCMS Online [2] to compare the results.

FIGURE 1. Untargeted metabolomics workflow

Results

Challenges in Untargeted Metabolomics Study
- Complexity of biological samples
- Diversity of small molecule metabolites: polar and non-polar analytes
- Ionization requires both positive and negative ion mode
- Wide range of concentrations
- No universal method for chromatographic separation
- Multiple sources of variability
- Structure elucidation of unknowns is expensive: lack of synthetic standards

Preparing for the UHPLC-MS Data Acquisition
Prior to the real samples analysis, a solvent blank with internal standard (IS) is injected at the beginning to check the solvent and the LC-MS status. The injections of the real samples should be randomized in order to eliminate systematic bias. Triplicate injections of the pooled plasma are intermittently repeated throughout the whole batch to validate consistent performance of the overall system. The experimental design and run sequence are shown in Figure 2.

FIGURE 2. UHPLC/MS Experimental Design and Run Sequence. Left, schematic showing the vials and sample names. Right, detailed content and overall time for each step.

UHPLC provides fast chromatography for high throughput analysis, the typical peak width is 4–6 seconds. Our method can baseline resolve Isolucine and Leucine, generating peak width 1.2 s at FWHM. Refer to Figure 4. For such narrow peaks, Q Exactive mass spectrometer operating at 70,000 resolution acquires >15 point across the peak without losing sensitivity (A). The QC of each run was performed by monitoring the intensity of IS, d5-hippuric acid (B), and the overall base peak (C). When all samples were finished, the selected ion chromatograms can be quickly viewed with RawMeat 2.1 (a free app with SIEVE 2.1 software). By inspecting the R1 and intensity of the IS, runs with large retention time drift and bad injections can be excluded from the following data analysis.

FIGURE 4. Method validation, data quality control in metabolomics application

Comparing Differential Analysis Results from SIEVE 2.1 software and XCMS Online
The results from SIEVE 2.1 software are compared to XCMS Online (Table 1) [2]. As shown in Figure 5, the significant components (p-value < 0.05, ratio > 2.0) identified by SIEVE 2.1 software and XCMS Online are similar, while the SIEVE software started with a much smaller number of components because of its capability to automatically remove solvent background.

TABLE 2. Representative metabolites that are significantly changed

Conclusion
An efficient and robust workflow for untargeted metabolomics is presented here. The reliable high-resolution, accurate-mass (HR/AM) performance of the Q Exactive LC-MS system eliminates the need for technical replicates on biological samples. The superior S/N in Orbitrap data allows efficient data reduction in SIEVE 2.1 software, resulting in much reduced data analysis effort. KEGG pathway visualization allows quick access to biological pathway mapping. The MS^3 spectral library mzCloud facilitates accurate compound identification.

References (if necessary)