Analysis of the Drosophila melanogaster proteome dynamics during the embryo early development by a combination of label-free proteomics approaches

Bertrand Fabre[1,2,4]*, Dagmara Korona[3,4]*, Arnoud Groen[1,2,4], Jakob Vowinckel[2,4], Laurent Gatto[2,4,5], Mike J Deery[1,2,4], Markus Ralser[2,4,6], Steven Russell[3,4], Kathryn S Lilley[1,2,4]

Affiliations

[1] Cambridge Centre for Proteomics, Department of Biochemistry, University of Cambridge, Cambridge, U.K

[2] Department of Biochemistry, University of Cambridge, University of Cambridge, Cambridge, U.K

[3] Department of Genetics, University of Cambridge, University of Cambridge, Cambridge, U.K

[4] Cambridge Systems Biology Centre, University of Cambridge, Cambridge, U.K

[5] Computational Proteomics Unit, Department of Biochemistry, University of Cambridge, Cambridge, U.K

[6] The Francis Crick Institute, Mill Hill Laboratory, The Ridgeway, London NW7 1AA

*These authors contributed equally to this work.

Correspondence: Pr. Kathryn S. Lilley, Ph.D., Cambridge Centre for Proteomics, Cambridge Systems Biology Centre, Department of Biochemistry, University of Cambridge, Cambridge CB2 1QR, U.K. Telephone: 01223 760255; e-mail: k.s.lilley@bioc.cam.ac.uk

Abbreviations: MS, Mass-Spectrometry; LC, Liquid Chromatography; DDA, Data-Dependent Acquisition; DIA, Data Independent Acquisition; LFQ, label-free Quantification; SWATH, Sequential Window Acquisition of all THeoretical fragment-ion; PRM, Parallel Reaction Monitoring; qRT-PCR, quantitative Real Time Polymerase Chain Reaction.

Keywords: Early embryo development; Label-free quantification; Mass-spectrometry; SWATH

Total number of words: 8051

Abstract

During embryogenesis, organisms undergo considerable cellular remodelling requiring the combined action of thousands of proteins. In the case of the well studied model *Drosophila melanogaster,* transcriptomic studies, most notably from the modENCODE project, have described in detail changes in gene expression at the mRNA level across development. Although such data are clearly very useful

for understand how the genome is regulated during embryogenesis, it is important to understand how changes in gene expression are reflected at the level of the proteome. In this study, we describe a combination of two quantitative label-free approaches, SWATH and Data Dependent Acquisition, to monitor changes in protein expression across a timecourse of *Drosophila melanogaster* embryonic development. We demonstrate that both approaches provide robust and reproducible methods for the analysis of proteome changes. In a preliminary analysis of *Drosophila* embryogenesis, we identified several pathways, including the heat-shock response, nuclear protein import and energy production, that are regulated during embryo development. In some cases changes in protein expression mirrored transcript levels across development, whereas other proteins showed signatures of post-transcriptional regulation. Taken together, our pilot study provides a solid platform for a more detailed exploration of the embryonic proteome.

Significance of the study

Lewis Wolpert is credited with the famous quote that gastrulation is the most important time in your life, highlighting the prominence of just one of the many steps in embryogenesis needed to build an organism from a single cell. Many genetic, biochemical and genomic studies have provided us with a better understanding of the role of many individual proteins play in key aspects of embryogenesis. However, in truth we have very little information on the dynamics of the wider proteome across development and this is much needed if we wish to understand how the complexity of development is controlled. In this study we use a combination of label-free proteomic quantification methods to monitor protein expression changes over a time course of *Drosophila* embryo development. We identify regulation of key pathways, such as nuclear protein import, heat-shock response and energy production systems during embryogenesis, providing new insights on proteome dynamics and a foundation for further detailed functional studies of the developmental proteome.

Introduction

Embryogenesis, the process in which a zygote develops from a fertilised egg cell into a complete organism, encompasses key events such as axis specification, gastrulation, germ layer formation and organogenesis [1]. The Drosophila embryo, due to its short generation time and well established genetic tools, has proved to be a tractable model for characterising many of the molecular aspects underpinning development and a wealth of studies have provided important insights that are often transferable to other systems [2-4]. During embryogenesis, a single diploid cell, the fertilized egg, undergoes a series of rapid nuclear division over a two hour period to form the blastoderm, comprising some 6,000 cells. This is followed by further cell divisions, cell movement and cellular differentiation to generate the germ layers and specific tissues of the larva [5]. Development is a highly regulated and complex process involving hundreds of regulatory factors and characterised by the spatial and temporal expression of the majority of genes in the genome [6]; from approximately 1000 genes at the blastoderm stage to over 10,000 during subsequent stages of development and morphogenesis [6-9]. The dynamic and complex cellular remodelling that characterises embryogenesis is, to a large extent, dependent upon fine tuning of the proteome to regulate cellular processes at each stage of development. As part of the modENCODE project, RNA-Seq and tiling microarrays were used to comprehensively sample the transcriptome across embryonic development: combining these data with previous *in situ* hybridisation studies provides a detailed view of transcriptome dynamics [6-9].

However, changes in transcript levels or isoforms are not always reflected as variation in proteins or proteomes [10] and the relationship between the transcriptome and proteome remains poorly characterised in most species and across development. In order to study relevant changes in protein expression and abundance, we require sensitive technologies capable of interrogating the entire proteome [11].

Mass-spectrometry (MS) has emerged as a powerful tool for proteome analysis. It can be deployed to identify the protein content of a sample, measure the quantity of proteins, provide insights into protein structure and identify molecules that proteins interact with [12]. The combination of high resolution mass-spectrometry, electrospray ionisation sources, robust liquid chromatography (LC) and bioinformatics analysis has enabled the identification of thousands of proteins from several organisms [13]. Quantitative MS is the method of choice for monitoring expression changes in the proteome [12], with Data Dependent Acquisition (DDA) methods based on the fragmentation of the most abundant peptides in a sample used for peptide identification [13]. For protein quantification, methods based on stable isotopic labelling have been widely used and been shown to provide very good accuracy [13]. However, these methods cannot be applied in many experimental workflows and can be prohibitively expensive in studies involving many biological samples [14]. In parallel to stable isotopic labelling methods, quantitative label-free (LFQ) approaches have been developed [12-14]. The reduced cost and ease with which such approaches can be adapted to any experimental workflow in comparison to isotopic labelling methods have contributed to the growing interest in LFQ approaches. In general, LFQ methods compare the signal intensity of a peptide between different LC-MS runs and different strategies have been described based on extracted ion chromatogram (XIC) of the MS1 signal of the peptides (MS1-based XICs) or XIC of the MS2 signals of the fragments (MS2-based XICs) [13]. Spectral counting, which compares the number of MSMS events for all the peptides of a protein between different LC-MS runs, has also been extensively used to monitor protein expression changes [13]. LFQ approaches have been shown to be suitable for comparing protein expression levels between different conditions or to provide good approximations of absolute protein abundance [15], and when used in conjunction with carefully designed experiments and relevant statistical methods, they provide good alternatives to isotopic labelling methods [12-14].

As opposed to DDA approaches, targeted proteomics approaches only quantify a selection of proteins of interest [16]. Methods including SRM (Selected Reaction Monitoring) [17] or PRM (Parallel Reaction Monitoring) [18] are based on the co-selection of peptides of interest and their fragments. In these acquisition modes, a peptide is selected within a defined m/z range and fragmented while all the other ions are discarded. During SRM, the ions resulting from the fragmentation step are subsequently filtered to detect only the expected fragments from the selected precursor [17]. In PRM, all the fragments from the precursor ion are detected in a unique high resolution scan [18]. These methods enable the MS analysis to focus only on the ions, and thus peptides/proteins, of interest which results in improved sensitivity and quantification accuracy. However, only few hundreds proteins are routinely quantifiable per MS run, which does not allow for broad proteome coverage [19, 20]. The main applications of targeted methods are therefore in hypothesis-driven experiments that require high sample numbers, when confirming results from other proteomics experiments or, due to the high sensitivity and accuracy, in biomarker characterisation [16, 21].

Recently, a new approach based on a Data Independent Acquisition mode, Sequential Window Acquisition of all THeoretical mass spectra (SWATH), has been developed [22]. In this approach, after a first MS scan, sequential wide m/z window ranges (swaths) of precursor ions are selected for fragmentation, resulting in mixed MSMS spectra. The quantification is performed through *in silico* SRM-like analysis. Typically, extracted ion chromatograms of the fragments of a peptide are generated

from MSMS spectra and used for comparative analysis between different MS runs. A *priori* information is required for SWATH experiments, typically a spectral library containing both peptide spectra and retention times. This library can be built from separate DDA acquisitions [23] or directly from the SWATH data [24]. Peak areas obtained from extracted ion chromatograms of each fragment from each peptide around the predicted retention time are used to generate the quantification data. The correlation between all the fragments and the precursor, as well as the order of abundance of the fragments compared with the spectra in the library, are used to verify the identity of the precursor. Since all the precursors in the particular swath are fragmented, the SWATH strategy potentially provides information on a large fraction of the proteome and, in theory, combines the best of SRM and DDA approaches: a sensitive and accurate quantification with good proteome depth [12].

In this study, we used a combination of different label-free approaches, global, semi-targeted and targeted, to monitor the dynamics of the *D. melanogaster* embryonic proteome. We first compared the performance of SWATH and MS1-based XICs. We found that, in general, label-free quantification methods provide a robust method for the identification and quantitation of proteins between different conditions. Encouragingly, we found that the SWATH approach showed very good reproducibility. Finally, we used a combination of the different label-free quantification methods to provide, for the first time, a comprehensive analysis of the *D. melanogaster* proteome at different stages of embryonic development. The expression profiles of selected proteins were validated using PRM and their transcripts measured by quantitative real time polymerase chain reaction (qRT-PCR). Importantly, we observed good agreement on trends in protein expressions between all the quantitative methods we employed. Our study identified some clusters of differentially regulated proteins across embryonic development, including importins and heat-shock proteins, as well as remodelling of the global energy production systems of the embryo. This study provides new insights into the proteome dynamics during *D. melanogaster* embryogenesis and serve as a foundation for our ongoing functional characterisation of the developmental proteome.

Materials and Methods

Fly Lines, embryo collection and protein extraction

Flies of the sequenced *D. melanogaster* iso-1 strain from the Bloomington Stock Centre (y^1; Gr22b^{iso-1} Gr22d^{iso-1} cn^1 CG33964^{iso-1} bw^1 sp^1; LysC^{iso-1} MstProx^{iso-1} GstD5$_{iso-1}$ Rh6^1) were kept on standard yeast-cornmeal media in a 12-h light-dark cycle at 25 °C and 75% relative humidity. Embryos were collected from cages on apple juice agar plates over a 16 hours period, or every 4.5 hours after egg laying and aged to produce a crude developmental time course (0-4.5h, 4.5-9h, 9-13.5h, 13.5-18h and 18-22.5h). After collection, embryos were dechorionated with 50% bleach, washed with water, frozen in liquid nitrogen and kept at -80°C. Three independent replicates were collected for each time point. Embryos were lysed in Tris 50mM pH 7.5, 4% SDS, protease inhibitor (Complete, Roche) using a Dounce homogenizer (50 strokes per sample). Each sample was then boiled for 5 minutes at 95°C.

Sample preparation for mass-spectrometry analysis

In gel digestion was used for sample preparation. A detergent compatible (DC) protein assay (Bio-Rad) was used to measure protein concentration. Loading buffer (Tris 40mM pH 7.5, 2% SDS, 10% glycerol, 25 mM DTT final concentration) was added to the samples and they were boiled for 5 minutes at 95°C. Alkylation of cysteines was performed using iodoacetamide at a final concentration of 60mM for 30 min at room temperature in the dark. 100μg of protein per condition were loaded on a SDS-PAGE gel (acrylamide concentration of 4% for the stacking and 12% for the resolving). Proteins were

concentrated in one band in the resolving gel (for the embryo developmental time course and to compare SWATH and DDA label-free quantification methods) or fully separated (for library construction). Gels were stained with colloidal coomassie and were cut and washed with a 50% acetonitrile (ACN), 25mM Ammonium Bicarbonate (AB) solution. Gel bands were dried with ACN and swollen with trypsin (Promega) in 50mM AB. 2µg or 500ng of trypsin were respectively used for single concentrated bands or bands from a full separation. Protein digestion was performed overnight at 37°C. The resulting peptides were extracted from the gel by two incubations in 10% formic acid, acetonitrile (1:1) for 15 min at 37 °C. The collected extractions were pooled with the initial digestion supernatant, dried in a Speed-Vac and resuspended in 3% acetonitrile, 0.1% formic acid for mass-spectrometry analysis. HRM peptides (Biognosys) were added to each sample before injection on the mass spectrometer.

Data Dependent Acquisitions (DDA) Mass Spectrometry

Mass spectrometry analyses were performed on a TripleTOF 6600 mass spectrometer fitted with a Duospray ion source (AB SCIEX) and coupled to an ACQUITY UPLC System (WATERS). 5µg (for the embryo developmental time course) of sample was injected onto a MicroLC column (150 mm long x 0.3 mm inner diameter) with ChromXP C18CL, 300 Å pore size, 3 µm diameter particles (AB SCIEX). Samples were run using a 49 min gradient from 3–40% solvent B (solvent A 0.1% formic acid, 5% DMSO in water; solvent B: 0.1% formic acid, 5% DMSO in acetonitrile) at a flow rate of 5 µl/min. Data were acquired using an ion spray voltage of 5.5 kV, curtain gas at 25 psi and nebulizer gas at 10 psi. A DDA method was set up with the MS survey range set between 350 and 1250 *m/z* (250 ms accumulation time) followed by dependent MS/MS scans with a mass range set between 100 and 2000 *m/z* (100 ms accumulation time) of the 20 most intense ions in the high sensitivity mode with a 2+ to 5+ charge state. Dynamic exclusion was set for a period of 15 sec and a tolerance of 50 ppm. Rolling collision energy was used.

Analysis of Data Dependent Acquisitions (DDA) files

The DDA .wiff files were analysed with MaxQuant [25] version 1.5.2.8 software using default settings. The minimal peptide length was set to 7. Trypsin/P was used as digestion enzyme. Search criteria included carbamidomethylation of cysteine as a fixed modification, oxidation of methionine and acetyl (protein N terminus) as variable modifications. Up to two missed cleavages were allowed. The mass tolerance for the precursor was 0.07 and 0.006 Da for the first and the main searches respectively, and for the fragment ions was 50 ppm. TOF recalibration was allowed. The DDA files were searched against the *D. melanogaster* UniProt fasta database (July 2014, 41,773 sequences) in which the Biognosys iRT peptide sequences (11 entries) were added. The identifications were filtered to obtain a FDR of 1% at the peptide and the protein levels. No filter was applied to the number of peptides per protein.

For the MS1-based XICs quantification of DDA analysis, the re-quantification and match between runs modules of MaxQuant [25] were used and the LFQ normalization method [26] was enabled (DDA LFQ) or not (DDA LF). Only unique peptides were used for quantification. When the LFQ normalization method was not used, the data were normalized by the median intensity in each condition. Fold change thresholds at 2 and 0.5 with an adjusted p-value inferior to 0.05 were used to consider a protein up or down-regulated. The p-values were adjusted using a Benjamini-Hochberg correction.

SWATH data acquisition

The LC gradient and TripleTOF 6600 mass spectrometer were set up as for the DDA acquisition as described above, but operated in SWATH mode using the following parameters: Acquisition of a 100-ms survey scan was followed by acquisition of 40 fragment ion spectra from 40 precursor isolation windows (swaths). The swaths were overlapping by 1 *m/z* and thus cover a range of 400-1250 *m/z*. Swath isolation windows were as follow: 399.5–422.0, 421.0–441.2, 440.2–456.6, 455.6-471.5, 470.5-484.7, 483.7-497.3, 496.3-510.0, 509.0-522.1, 521.1-533.6, 532.6-545.2, 544.2-555.6, 554.6-566.6, 565.6-577.0, 576.0-587.5, 586.5-597.4, 596.4-607.8, 606.8-618.3, 617.3-629.3, 628.3-640.3, 639.3-651.3, 650.3-663.4, 662.4-675.5, 674.5-688.7, 687.7-701.9, 700.9-715.1, 714.1-728.8, 727.8-743.1, 742.1-758.5, 757.5-774.5, 773.5-791.0, 790.0-809.1, 808.1-827.8, 826.8-848.2, 847.2-871.3, 870.3-897.1, 896.1-927.9, 926.9-965.9, 964.9-1015.4, 1014.4-1097.3, 1096.3-1249.7. The SWATH MS2 spectra were recorded with an accumulation time of 40 ms and cover 100-2000 *m/z*. The collision energy for each window was determined according to the calculation for a charge 2+ ion centred upon the window with a spread of 15.

SWATH analysis

For the generation of the SWATH assay library, 10 bands of a 1D gel fractionation of 100µg of a protein sample from embryos collected over a 16 hours period were acquired in Information Dependent Acquisition mode on a TripleTOF 6600 mass spectrometer (AB Sciex). To each peptide sample, HRM peptides (Biognosys) were added before analysis.

All the files were analysed using MaxQuant with similar parameters than for the DDA data analysis. The library was built using Spectronaut 7 from Biognosys [25], using the default settings, from the resulting combined file from the MaxQuant analysis. Spectronaut 7 (Biognosys) [27] was used to analyse the SWATH experiments. Default settings were used except for the retention time prediction type that was set to dynamic iRT with a correction factor for window of 2. Different Q-value were used to filter the reported intensity values (Q-values from 10^{-1} to 10^{-8}). For the embryo development time course experiment and DDA and SWATH quantification methods comparison, a Q-value of 10^{-5} was used. For each proteins, the three peptides with the highest intensities were used for quantitative analysis. The data were normalized by the median protein intensity in each condition. Fold change thresholds at 2 and 0.5 with an adjusted p-value inferior to 0.05 were used to consider a protein up or down-regulated. The p-values were adjusted using a Benjamini-Hochberg correction.

Data Dependent Acquisition (DDA) and Parallel Reaction Monitoring (PRM) on a Q-Exactive mass-spectrometer

A list of peptides from DDA and SWATH analysis of the embryo developmental time course was prepared for PRM validation (at least 3 peptides per protein). An embryo sample was first run in DDA mode and analysed with MaxQuant to obtain the retention times of the peptides which were used to set up a scheduled PRM assay. Samples were trapped on a 100µm × 2cm, C18, 5µm, 100Å trapping column (Acclaim PepMap 100) in µlPickUp Injection mode at 4 µL/min flow rate for 10 minutes. Samples were then loaded on a RSLC, 75µm × 50cm, nanoViper, C18, 3µm, 100Å column (Acclaim, PepMap) retrofitted to an EASY-Spray source with a flow rate of 300nl/min (buffer A: HPLC H2O, 0.1% formic acid, buffer B: 100% ACN, 0.1% formic acid). A 120 minutes gradient was performed as follow; 0-3 min: 2% buffer B, 3-90 min: 2->40% buffer B, 90.3-95 min: 40%->90% buffer B, 95.3-120 min: 2% buffer B). Peptides were transferred to the gaseous phase with positive ion electrospray ionization at 2.1kV. In DDA the top 10 precursors were acquired between 400 and 1600 *m/z* with a 2Th (Thomson) selection window, dynamic exclusion of 30 seconds, normalised collision energy (NCE) of 25 and resolution of 70,000. For PRM, precursors were targeted in a 2Th selection window around the *m/z* of interest. Precursors were fragmented in HCD mode with NCE energy of 25. MS1 was performed at a

70,000 resolution, an AGC target of 3e6 and a maximum C-trap fill time of 200ms; MS/MS was performed at 35,000 resolution, an AGC target of 5e4 and a maximum C-trap fill time of 100ms. Spectra were analysed using Skyline with manual validation. Skyline quantitation data was exported to excel and the quantitation data was normalized against the TIC of the MS runs. The list of the peptides followed by PRM is provided in Supplementary Table 2.

Data availability

All the mass spectrometry data have been deposited to the ProteomeXchange Consortium [28] via the PRIDE partner repository with the dataset identifier PXD003178.

The results from the DDA LFQ and SWATH analysis of the time course experiment, the PRM results and the spectral library used in this study are provided in the supplementary tables.

Total RNA Isolation, cDNA Synthesis and Quantitative Real-time PCR

For the qRT-PCR experiments, 2 biological replicates, independent from the replicates used in the MS experiments, were analysed. Total RNA was isolated using the TRIzol method (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol. The quality of the extracted RNA was verified by gel electrophoresis and optical density measurements. First-strand cDNA was synthesized with the SuperScript® III Reverse Transcriptase Kit (Invitrogen) according to manufacturer instructions, the synthesized cDNAs were stored at −20°C for further qRT-PCR.

qRT-PCR was performed in the iCycler iQ system (Bio-Rad, Hercules, CA) and an IQ SYBR Green Supermix (Bio-Rad) using SensiMix™ SYBR® Hi-ROX Kit (Bioline). Data analysis was performed using comparative C_T method following established protocols [29]. The rp49 (RpL32) mRNA was used as internal control and changes in mRNA levels were calculated as a fold-change relative to the value for the 0-4.5 hours sample. Primers were designed with Primer3 or retrieved from other databases [28]. Sequences of all primers used for qRT-PCR analyses are as following:

Annotation symbol	protein	symbol	F name	Forward primer sequence	R name	Reverse primer sequence
CG4799	P52295	Pen	pen_F	CTGGCACAGATCAACAGACT	pen_R	CCTGGATCTGCTTCTGGTTAC
CG4916	P23128	me31B	me31b_F3	GCCAAAGGACAACCGATTCAA	me31b_R3	TCCCATCCCTTCTCGAATATACC
CG2637	M9NDC3	Fs(2)Ket	FS_F2	TTCGTTCATCAAGCGCATCG	Fs_R2	AAACGAAGTGCACAGATCGC
CG7939	P04359	RpL32	RpL32-F	AGCATACAGGCCCAAGATCG	RpL32-R3	TGTTGTCGATACCCTTGGGC

Results and discussion

Comparison of the performances of SWATH and DDA label-free quantification approaches

SWATH has been recently introduced as a label-free quantitative proteomic approach that combines the advantages of DDA and targeted proteomic methods [22], however, few studies have compared SWATH quantification robustness and accuracy with other quantitative proteomic approaches [30, 31]. We first compared MS1-based XICs and SWATH label-free quantification methods (Figure 1A). Indeed, to our knowledge there have been no published comparison between the MS1-based XICs label-free approach using MaxQuant, which is the most commonly used software when performing this kind of quantification, and the SWATH label-free quantification method. A protein sample prepared from 0-16hr *D. melanogaster* embryos and digested with trypsin was used as a standard to benchmark the different quantification methods. Four injection replicates were used to compare these quantification methods. For the MS1-based XICs, MaxQuant [25] analysis was performed with (DDA LFQ) or without (DDA LF) the LFQ normalization method [26] (Figure 1A). The LFQ normalization method developed by Cox *et al* [26] is a suite of algorithms used to normalize the MS1-based XICs between different LC-

MS/MS runs. For the SWATH analysis, a library containing 1750 proteins was produced from a 0-16 hr embryo extract separated on a 1D SDS-PAGE gel, digested with trypsin and analysed in DDA mode on the same platform using MaxQuant and Spectronaut (Biognosys) [27]. We found that 89% of the proteins present in the spectral library were previously identified in other *Drosophila melanogaster* embryo proteomics studies [32, 33] (Supplementary Figure 2). Spectronaut was used to process the SWATH data and the average of the 3 highest peptide areas based on the sum of the fragment intensities were used to obtain a protein abundance index for the quantification at the protein level (Figure 1A). This approach has been shown to be suitable for label-free quantification approaches and provides good accuracy [34, 35]. Only the peptides of very high quality (Q-value $< 1.10^{-5}$ corresponding to an FDR of 0.001%) were used for the SWATH analysis since these showed low CVs and a good run to run reproducibility at the peptide level when compared to DDA (Supplementary Figure 2). The results showed that SWATH generates more robust quantification data than DDA MS1-based XICs quantification on the same platform when the LFQ normalisation method is not used (DDA LF) (Figure 1B-1C and Figure 2A-B). Indeed, SWATH analysis provides better run to run intensity reproducibility (Figure 1B), generates lower CVs (Figure 1C) and less extreme ratios (Log$_2$ ratios > 1 or < -1) when comparing the technical replicates (Figure 2A). In contrast, more proteins were quantified with the DDA MS1-based XICs quantification but at the cost of poor quality quantification (Figure 2B). However, when the LFQ normalization method was used (DDA LFQ), the quantification quality was greatly improved and more similar to that achieved with the SWATH analysis (Figure 1B-C and Figure 2A-B). The observed median CVs are very close for both methods (3.05 and 5.19 for the DDA LFQ and SWATH, respectively) and both quantification approaches produce few extreme ratios (0.24 and 0.1% of Log$_2$ ratios between the technical replicates > 1 or < -1 for the DDA LFQ and SWATH, respectively) (Figure 1B-C and Figure 2A-B). Thus, our analysis indicates that SWATH and DDA MS1-based XICs quantifications using the LFQ normalisation method in MaxQuant are robust tools for the analysis of changes in protein expression at the proteome level. Moreover, it has been shown that since the quantifications are performed at a different level (MS1 level for DDA MS1-based XICs and MS2 level for SWATH) the methods provide complementary information, with different errors in signal extraction and discrepancies between the two approaches can be easily highlighted [36]. We therefore decided to use a combination of these two quantitative proteomics approaches to study the dynamics of the proteome during *D. melanogaster* embryogenesis.

Monitoring protein expression during *D. melanogaster* embryonic development by Label-Free quantification and SWATH approaches.

Despite its importance for all members of the Animal kingdom, very few quantitative data are available on the modulation of the proteome during embryonic development for any species [10, 33]. To begin to characterise how the proteome is regulated during embryogenesis, we used the *D. melanogaster* embryo as model since it is exceptionally well characterised at the genetic and genomic level [1]. Embryos were collected across development at 5 different time windows (0-4.5h, 4.5-9h, 9-13.5h, 13.5-18h and 18-22.5h) with three independent biological replicates subject to analysis in both DDA and SWATH acquisition modes (Figure 3A). The resulting files were analysed with MaxQuant using the LFQ normalization method or with MaxQuant and Spectronaut for DDA MS1-based XICs (DDA LFQ) and SWATH analysis respectively (Figure 3A). We identified over 400 proteins with each method that were quantified at all timepoints (430 with DDA LFQ and 475 with SWATH). All the timepoints were compared to the 0-4.5h timepoint. Importantly, we found that the coefficients of variation between biological replicates were low (median CV around 10%) for both approaches, demonstrating good data quality (Supplementary Figure 3). First, we observed that the protein ratios are more divergent at the

later stages of embryonic development, indicating a global remodelling of the proteome across development (Supplementary Figure 4). However, the overall profile of the proteins at different stages look similar on a SDS-PAGE, suggesting that the expression of the most abundant proteins remains relatively constant during embryogenesis (Supplementary Figure 5). Encouragingly, both datasets generated by DDA LFQ and SWATH resulted in similar profiles (Figure 3C). Only few discrepancies were observed between the two methods and the problems in signal extraction were identified using Skyline [37] (Supplementary Figure 6A-B).

Proteins regulated across *D. melanogaster* embryogenesis.

To focus on a high confidence set of proteins we selected those that were differentially expressed at least two-fold (corrected p-value <0.05) compared to the first timepoint. We identified 31 proteins (15 up and 16 down) encoded by 31 genes for the DDA analysis and 64 proteins (36 up and 28 down) encoded by 64 genes for the SWATH analysis. The latter were significantly enriched (Benjamini and Hochberg corrected p= 3.7e-3) for the Gene Ontology term small molecule metabolic process, in particular multiple subunits of the ATP Synthase complex (see below). We also noticed that in both lists there was a significant enrichment for genes expressed in the embryonic/larval muscle system (Benjamini and Hochberg corrected p<1.0e-3). Comparing both lists we identified 49 proteins that were common to both and 13 of these are expressed in the embryonic musculature. We elaborate on some of these commonly identified differentially expressed proteins below.

Among the differentially expressed proteins we identified, the alpha and beta subunits of the importin nuclear pore complex were down regulated over the timecourse (Pendulin and Fs(2)Ket respectively; Figure 4A). The importin heterodimer targets hundreds of proteins to the nuclear-pore complex (NPC) and facilitates their translocation across the nuclear envelope. The alpha subunit binds specifically to substrates containing either a simple or bipartite NLS (nuclear localization signal) motif [38] and the beta subunit mediates docking of the substrate to nucleoporins [38]. Both subunits of the heterodimer share very close expression profiles and our data support the view that regulation of the nuclear import system may play an important role in development [38]. An increase in the expression of Pen and Fs(2)Ket was also reported between the 2-4h and 10-12h embryonic time points using SILAC quantification [33], with reported ratios very close to those observed in our study at 0-4.5h and 9-13.5h time points (Supplementary Figure 7).

We also observed decreased expression of two heat-shock proteins, Hsp26 and Hsp27, over the course of embryogenesis (Figure 4B), supporting previous studies that reported a reduction in *Hsp27* expression during early development [39]. In addition, it has been shown that the subcellular localization of Hsp27 and Hsp26 varies between cellular compartments depending on developmental stages [39-41], indicating that the developmental profiles of Hsp26 and Hsp27 may be complex and suggesting that these proteins may have several roles during development. In general, our knowledge about the roles of heat-shock proteins across embryogenesis remains limited and Hsp27 and Hsp26 represent two interesting candidates for further investigations.

The DEAD-box ATP-dependent RNA helicase Me31B, is involved in translation silencing through mRNA decapping [42], was also found to decrease during embryogenesis (Figure 4C). As with the other four proteins discussed above, there is a strong maternal contribution of *me31b* mRNA and this appears to be reflected at the protein level [43]. Although Me31B appears to be required for normal oogenesis, the role of this protein during embryonic development is poorly understood. We suspect that declining protein levels reflect, to a large extent, depletion of maternally contributed protein and mRNA. In support of this notion, we note that modENCODE expression profiles demonstrate that transcripts for

all five of these genes dramatically decline at the onset of zygotic gene expression [6]. Interestingly, Gouw *et al* also observed a decrease of the expression of Me31B at the protein level during the very early steps of the embryo development (180-270min versus 0-90min time points) [32].

In contrast, we observe increasing expression of the DNA-binding protein Modulo during embryogenesis (Figure 4D). Interestingly, RNA-seq data indicates relatively stable transcript levels across embryogenesis, suggesting post-transcriptional regulation. Modulo binds DNA and associates with chromatin [44] but it also contains four RNA Recognition Motifs and has been shown to provides a sequence-specific RNA binding activity as part of a nucleolar ribonucleoprotein complex [45]. It is speculated that Modulo plays a vital role in the regulation of gene expression during Drosophila development. Due to its ability bind to nucleic acids and proteins, it may control the activity of genes critical for the process of morphogenesis of several embryonic territories via chromatin changes. Therefore, Modulo seems to be an interesting candidate to study in the context of development.

As with Modulo, the level of the Failed Axon Connections (Fax) protein was also found to increase over embryonic development (Figure 4D) yet exhibit relatively stable transcript levels. This protein is the founding member of the conserved FAX family that, together with the Abl tyrosine kinase, is involved in axon development [46]. Mutations in *fax* act as dominant genetic enhancers of *Abl* mutant phenotypes and the Fax protein is expressed in a similar pattern to Abl in the embryonic mesoderm and axons of the central nervous system [46]. The increasing expression of Fax in the latter stages of embryonic expression is consistent with the role Fax plays in axonogenesis, since the majority of axon development occurs in the latter half of embryogenesis. In both these cases we have identified proteins encoded by transcripts that are relatively stable over embryonic development yet show significant changes in protein abundance, emphasising the need to gather more quantitative data on the proteome. Hughes *et al* also observed an increase in the expression of Modulo and Fax using SILAC quantification with ratios very close to those measured in our study (Supplementary Figure 7) [33].

Energy production systems remodelling during *D. melanogaster* embryogenesis

We found several enzymes involved in different aspects of energy metabolism steadily increasing during embryogenesis, with peak expression at the 13.5-18h timepoint, before declining at the end of embryonic life (Figure 5). We observed this trend with proteins involved in the TCA cycle (Malic enzyme (Men) and Citrate synthase (Kdn)), amino acid metabolism (Arginine kinase (Argk)), lipid biosynthesis (Fatty acid synthase 1 (FASN1)) and electron transport chain respiration (Glycerol-3-phosphate dehydrogenase (Gpdh) and Cytochrome c proximal (Cyt-c-p)). During early development, the embryo undergoes considerable cellular remodelling which is likely to require high levels of energy and the induction of genes implicated in the main energy production pathways has already been observed at the mRNA level [47]. In contrast, subunits of the ATP synthase complex do not share the same profile (Figure 5A and B). Rather, we observed that 5 subunits (A, B, D, G and O) are all down regulated prior to peaking at 13.5-18h timepoint (Figure 5B). The profiles of some of these subunits are significant, and are similar to that of Prohibitin 2 which has been found to interact with the complex IV of the respiratory chain [48] (Figure 5B). The expression profile of the alpha subunit, representative for these results was confirmed by Parallel Reaction Monitoring (Supplementary Figure 8). These data suggests that the embryo regulates ATP production at key steps of development by modulating the expression level of ATP synthase as well as enzymes involved in the aerobic and anaerobic energy production pathways. It is possible the regulation may be related to changes in O_2 availability, for example, it is known that embryonic tracheal development is regulated by changing O_2 [49]. In general, few data are available on the regulation of metabolic pathways during embryogenesis and our new data hint at a complex interplay between metabolic pathways during development.

Validation of protein expression changes by parallel reaction monitoring and quantitative real time PCR

In order to validate our quantitative proteomics approaches we used parallel reaction monitoring (PRM) to measure expression changes. In the case of the alpha and beta importin subunits, Me31B and Modulo, the PRM data showed very good agreement with the DDA LFQ and SWATH data (Figure 6). We also checked mRNA expression for the importin subunits and Me31B by reverse transcriptase-quantitative PCR (Figure 6) and in all three cases saw that the trends in transcript levels mirrored the protein levels. In addition we note that the trend in transcript levels we measured were in broad agreement with the RNA-seq data from modENCODE. Taken together, these validations highlight the accuracy and reproducibility of the quantitative label-free methods we have used here.

Concluding remarks

In this study we have applied label-free quantification methods to monitor dynamics in the *D. melanogaster* embryonic proteome. We compared the performance of SWATH and DDA MS1-based XICs quantification methods, showing that both approaches provide a robust way to measure variation at the proteome level. For DDA MS1-based XICs quantification we found that LFQ normalization with MaxQuant software gave the most robust performance, whereas SWATH gave comparable results without any sophisticated normalization. However, it would be interesting to apply the MaxQuant LFQ normalization algorithms to SWATH data to see if it can improve SWATH quantification quality and analytical coverage. Using a combination of DDA MS1-based XICs and SWATH quantification approaches, we found that the expression of more than 70 proteins was regulated across embryonic development. Validation of some of our results with orthogonal proteomics experiments or mRNA analysis increases our confidence in the reliability of these approaches when applied to comparative analysis of the developmental proteome. We found that proteins involved in heat-shock response, nuclear protein import or energy production pathways were differentially expressed across development, some showing expected declines due to utilisation of maternal stores, whereas others showed hallmarks of post-transcriptional regulation. We believe the preliminary work reported here is the first proteomics study covering a substantial portion of Drosophila embryogenesis and sets the scene for more detailed analysis. Although more in depth analysis of the proteome will provide additional insights, this study highlights the complexity of proteome dynamics during development and identifies some regulated biochemical pathways likely to be important. Finally, this study serves as a foundation for more detailed, higher resolution analyses to better characterise differential protein expression during embryogenesis in this well-established model system.

References

[1] Fristrom, J. W., The developmental biology of Drosophila. *Annu Rev Genet* 1970, *4*, 325-346.

[2] Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A*., et al.*, The genome sequence of Drosophila melanogaster. *Science* 2000, *287*, 2185-2195.

[3] An, P. N., Yamaguchi, M., Bamba, T., Fukusaki, E., Metabolome analysis of Drosophila melanogaster during embryogenesis. *PloS one* 2014, *9*, e99519.

[4] Fortini, M. E., Skupski, M. P., Boguski, M. S., Hariharan, I. K., A survey of human disease gene counterparts in the Drosophila genome. *J Cell Biol* 2000, *150*, F23-30.

[5] Weigmann, K., Klapper, R., Strasser, T., Rickert, C*., et al.*, FlyMove--a new way to look at development of Drosophila. *Trends Genet* 2003, *19*, 310-311.

[6] Graveley, B. R., Brooks, A. N., Carlson, J. W., Duff, M. O*., et al.*, The developmental transcriptome of Drosophila melanogaster. *Nature* 2011, *471*, 473-479.

[7] Manak, J. R., Dike, S., Sementchenko, V., Kapranov, P*., et al.*, Biological function of unannotated transcription during the early development of Drosophila melanogaster. *Nat Genet* 2006, *38*, 1151-1158.

[8] Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E*., et al.*, Systematic determination of patterns of gene expression during Drosophila embryogenesis. *Genome Biol* 2002, *3*, RESEARCH0088.

[9] Tomancak, P., Berman, B. P., Beaton, A., Weiszmann, R*., et al.*, Global analysis of patterns of gene expression during Drosophila embryogenesis. *Genome Biol* 2007, *8*, R145.

[10] Sun, L., Champion, M. M., Huber, P. W., Dovichi, N. J., Proteomics of Xenopus development. *Molecular human reproduction* 2015.

[11] Artieri, C. G., Fraser, H. B., Transcript length mediates developmental timing of gene expression across Drosophila. *Mol Biol Evol* 2014, *31*, 2879-2889.

[12] Bensimon, A., Heck, A. J., Aebersold, R., Mass spectrometry-based proteomics and network biology. *Annu Rev Biochem* 2012, *81*, 379-405.

[13] Cox, J., Mann, M., Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem* 2011, *80*, 273-299.

[14] Bantscheff, M., Lemeer, S., Savitski, M. M., Kuster, B., Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Analytical and bioanalytical chemistry* 2012, *404*, 939-965.

[15] Fabre, B., Lambour, T., Bouyssié, D., Menneteau, T*., et al.*, Comparison of label-free quantification methods for the determination of protein complexes subunits stoichiometry. *EuPA Open Proteomics* 2014, *4*, 5.

[16] Ebhardt, H. A., Root, A., Sander, C., Aebersold, R., Applications of targeted proteomics in systems biology and translational medicine. *Proteomics* 2015, *15*, 3193-3208.

[17] Lange, V., Picotti, P., Domon, B., Aebersold, R., Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular systems biology* 2008, *4*, 222.

[18] Gallien, S., Duriez, E., Crone, C., Kellmann, M*., et al.*, Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Molecular & cellular proteomics : MCP* 2012, *11*, 1709-1723.

[19] Gallien, S., Kim, S. Y., Domon, B., Large-Scale Targeted Proteomics Using Internal Standard Triggered-Parallel Reaction Monitoring (IS-PRM). *Molecular & cellular proteomics : MCP* 2015, *14*, 1630-1644.

[20] Soste, M., Hrabakova, R., Wanka, S., Melnik, A*., et al.*, A sentinel protein assay for simultaneously quantifying cellular processes. *Nature methods* 2014, *11*, 1045-1048.

[21] Picotti, P., Aebersold, R., Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature methods* 2012, *9*, 555-566.

[22] Gillet, L. C., Navarro, P., Tate, S., Rost, H*., et al.*, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular & cellular proteomics : MCP* 2012, *11*, O111 016717.

[23] Schubert, O. T., Gillet, L. C., Collins, B. C., Navarro, P*., et al.*, Building high-quality assay libraries for targeted analysis of SWATH MS data. *Nature protocols* 2015, *10*, 426-441.

[24] Tsou, C. C., Avtonomov, D., Larsen, B., Tucholska, M*., et al.*, DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nature methods* 2015, *12*, 258-264, 257 p following 264.

[25] Cox, J., Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* 2008, *26*, 1367-1372.

[26] Cox, J., Hein, M. Y., Luber, C. A., Paron, I*., et al.*, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & cellular proteomics : MCP* 2014, *13*, 2513-2526.

[27] Bruderer, R., Bernhardt, O. M., Gandhi, T., Miladinovic, S. M*., et al.*, Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Molecular & cellular proteomics : MCP* 2015, *14*, 1400-1410.

[28] Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A*., et al.*, ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature biotechnology* 2014, *32*, 223-226.

[29] Schmittgen, T. D., Livak, K. J., Analyzing real-time PCR data by the comparative C(T) method. *Nature protocols* 2008, *3*, 1101-1108.

[30] Bourassa, S., Fournier, F., Nehme, B., Kelly, I*., et al.*, Evaluation of iTRAQ and SWATH-MS for the Quantification of Proteins Associated with Insulin Resistance in Human Duodenal Biopsy Samples. *PloS one* 2015, *10*, e0125934.

[31] McQueen, P., Spicer, V., Schellenberg, J., Krokhin, O*., et al.*, Whole cell, label free protein quantitation with data independent acquisition: quantitation at the MS2 level. *Proteomics* 2015, *15*, 16-24.

[32] Gouw, J. W., Pinkse, M. W., Vos, H. R., Moshkin, Y*., et al.*, In vivo stable isotope labeling of fruit flies reveals post-transcriptional regulation in the maternal-to-zygotic transition. *Molecular & cellular proteomics : MCP* 2009, *8*, 1566-1578.

[33] Hughes, C. S., Foehr, S., Garfield, D. A., Furlong, E. E*., et al.*, Ultrasensitive proteome analysis using paramagnetic bead technology. *Molecular systems biology* 2014, *10*, 757.

[34] Fabre, B., Lambour, T., Delobel, J., Amalric, F*., et al.*, Subcellular distribution and dynamics of active proteasome complexes unraveled by a workflow combining in vivo complex cross-linking and quantitative proteomics. *Molecular & cellular proteomics : MCP* 2013, *12*, 687-699.

[35] Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P., Geromanos, S. J., Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Molecular & cellular proteomics : MCP* 2006, *5*, 144-156.

[36] Rardin, M. J., Schilling, B., Cheng, L. Y., MacLean, B. X*., et al.*, MS1 Peptide Ion Intensity Chromatograms in MS2 (SWATH) Data Independent Acquisitions. Improving Post Acquisition Analysis of Proteomic Experiments. *Molecular & cellular proteomics : MCP* 2015, *14*, 2405-2419.

[37] MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M*., et al.*, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010, *26*, 966-968.

[38] Goldfarb, D. S., Corbett, A. H., Mason, D. A., Harreman, M. T., Adam, S. A., Importin alpha: a multipurpose nuclear-transport receptor. *Trends Cell Biol* 2004, *14*, 505-514.

[39] Michaud, S., Marin, R., Tanguay, R. M., Regulation of heat shock gene induction and expression during Drosophila development. *Cell Mol Life Sci* 1997, *53*, 104-113.

[40] Glaser, R. L., Wolfner, M. F., Lis, J. T., Spatial and temporal pattern of hsp26 expression during normal development. *EMBO J* 1986, *5*, 747-754.

[41] Pauli, D., Tonka, C. H., Tissieres, A., Arrigo, A. P., Tissue-specific expression of the heat shock protein HSP27 during Drosophila melanogaster development. *J Cell Biol* 1990, *111*, 817-828.

[42] Nishihara, T., Zekri, L., Braun, J. E., Izaurralde, E., miRISC recruits decapping factors to miRNA targets to enhance their degradation. *Nucleic Acids Res* 2013, *41*, 8692-8705.

[43] de Valoir, T., Tucker, M. A., Belikoff, E. J., Camp, L. A*., et al.*, A second maternally expressed Drosophila gene encodes a putative RNA helicase of the "DEAD box" family. *Proc Natl Acad Sci U S A* 1991, *88*, 2113-2117.

[44] Bantignies, F., Goodman, R. H., Smolik, S. M., The interaction between the coactivator dCBP and Modulo, a chromatin-associated factor, affects segmentation and melanotic tumor formation in Drosophila. *Proc Natl Acad Sci U S A* 2002, *99*, 2895-2900.

[45] Perrin, L., Romby, P., Laurenti, P., Berenger, H*., et al.*, The Drosophila modifier of variegation modulo gene product binds specific RNA sequences at the nucleolus and interacts with DNA and chromatin in a phosphorylation-dependent manner. *J Biol Chem* 1999, *274*, 6315-6323.

[46] Hill, K. K., Bedian, V., Juang, J. L., Hoffmann, F. M., Genetic interactions between the Drosophila Abelson (Abl) tyrosine kinase and failed axon connections (fax), a novel protein in axon bundles. *Genetics* 1995, *141*, 595-606.

[47] Tennessen, J. M., Baker, K. D., Lam, G., Evans, J., Thummel, C. S., The Drosophila estrogen-related receptor directs a metabolic switch that supports developmental growth. *Cell Metab* 2011, *13*, 139-148.

[48] Strub, G. M., Paillard, M., Liang, J., Gomez, L*., et al.*, Sphingosine-1-phosphate produced by sphingosine kinase 2 in mitochondria interacts with prohibitin 2 to regulate complex IV assembly and respiration. *FASEB J* 2011, *25*, 600-612.

[49] Simon, M. C., Keith, B., The role of oxygen availability in embryonic development and stem cell function. *Nat Rev Mol Cell Biol* 2008, *9*, 285-296.

Figures legends

Figure 1: Comparison of label-free quantification methods

A) Overall strategy for comparing SWATH and Data Dependent Acquisition label-free quantification methods. Proteins were extracted from a Drosophila embryo sample, digested with trypsin and retention time normalisation (HRM) peptides spiked into the peptide sample before injection. Four analytical replicates were performed for SWATH and DDA acquisition modes on an ABSciex Triple-TOF 6600. The resulting files were analysed with MaxQuant with (DDA LFQ) or without (DDA LF) the LFQ module enabled or with MaxQuant and Spectronaut for the DDA and SWATH analysis, respectively. B) The Log_{10} intensities of the injection replicates were plotted pairwise to evaluate the run to run reproducibility of each quantification method. C) The Coefficient of Variation (CV) was calculated for each protein quantified and plotted as function of the intensities normalized to the maximal intensity.

Figure 2: Comparison of the number of extreme values generated by label-free quantification methods

A) Ratio were calculated pairwise between the different injection replicates. Intensities, normalized to the maximal intensity, were plotted as function of the ratios for each quantification method. B) Table summarising the comparison of label-free quantification methods.

Figure 3: Monitoring the dynamics of the proteome during embryogenesis by a combination of SWATH and DDA label-free quantification

A) Overall strategy to measure changes in protein expression levels across *D. melanogaster* embryonic development by SWATH and Data Dependent Acquisition label-free quantification methods. Embryos were collected for a developmental time course of five 4.5 hour timepoints. Proteins were extracted, digested with trypsin and HRM peptides spiked into the peptide sample before injection. The samples were analysed using both SWATH and DDA acquisition modes on an ABSciex Triple-TOF 6600. The resulting files were analysed with MaxQuant with the LFQ module (DDA LFQ) enabled or with MaxQuant and Spectronaut for the DDA and SWATH analysis, respectively. B) Venn diagram representing the overlap between the proteins identified in our DDA LFQ and SWATH analysis. C) Correlation between the ratio obtained with the DDA LFQ and SWATH analysis. The ratios measured by DDA LFQ and SWATH (4.5-9h/0-4.5h, 9-13.5h/0-4.5h, 13.5-18h/0-4.5h and 18-22.5h/0-4.5h) are plotted for each protein. The values of the fitted linear regression are reported on the graph.

Figure 4: Proteins regulated across *D. melanogaster* embryonic development.

The ratios for each protein at each timepoint compared with the mean value at the 0-4.5h timepoint were calculated. The fold change obtained from the DDA label-free quantification across the time course were represented using histograms for: A) importin alpha (Pen) and beta (Fs(2)Ket); B) the heat-shock proteins Hsp27 and Hsp26; C) the ATP-dependent RNA helicase Me31b (Me31B); D) the DNA-binding proteins Modulo (Mod) and Failed Axon Connections (Fax). Error bars represent the standard deviations between the biological replicates. Benjamini-Hochberg corrected p-values are shown for the comparison between the 0-4.5h timepoint and the 18-22.5h (A, B and C) or the 13.5-18h (D) timepoints.

Figure 5: Remodelling of the energy production pathways during embryogenesis

The ratios for different proteins involved in different metabolic pathways were calculated for each timepoint as compared to the 0-4.5h timepoint. A) The fold changes obtained from the DDA label-free quantification across the time course were represented using histograms for Arginine kinase (Argk), Citrate synthase (Kdn), Glycerol-3-phosphate dehydrogenase (Gpdh), Malic enzyme (Men), Cytochrome c proximal (Cyt-c-p) and the Fatty acid synthase 1 (FASN1). B) The fold changes obtained from the DDA label-free quantification across the time course were represented using histograms for five ATP synthase (ATPsyn) subunits and Prohibitin 2 (PHB2). Error bars represent the standard deviations between the biological replicates. Benjamini-Hochberg corrected p-values are shown for the comparison between the 0-4.5h timepoint and the 13.5-18h (A) or the 9-13.5h (B) timepoints.

Figure 6: Validation of expression profiles of a subset of proteins by PRM and qRT-PCR

The expression profiles of: A) Pen; B) Fs(2)Ket; C) Me31B and D) Mod, were measured using Parallel Reaction Monitoring (PRM) and quantitative real time PCR (RT-qPCR). The ratios were calculated for timepoint in comparison with the 0-4.5h timepoint. Error bars represent the standard deviations between the biological replicates. Benjamini-Hochberg corrected p-values are shown for the comparison between the 0-4.5h timepoint and the 18-22.5h (A, B and C) or the 13.5-18h (D) timepoints.

Figure 1

Figure 2

A
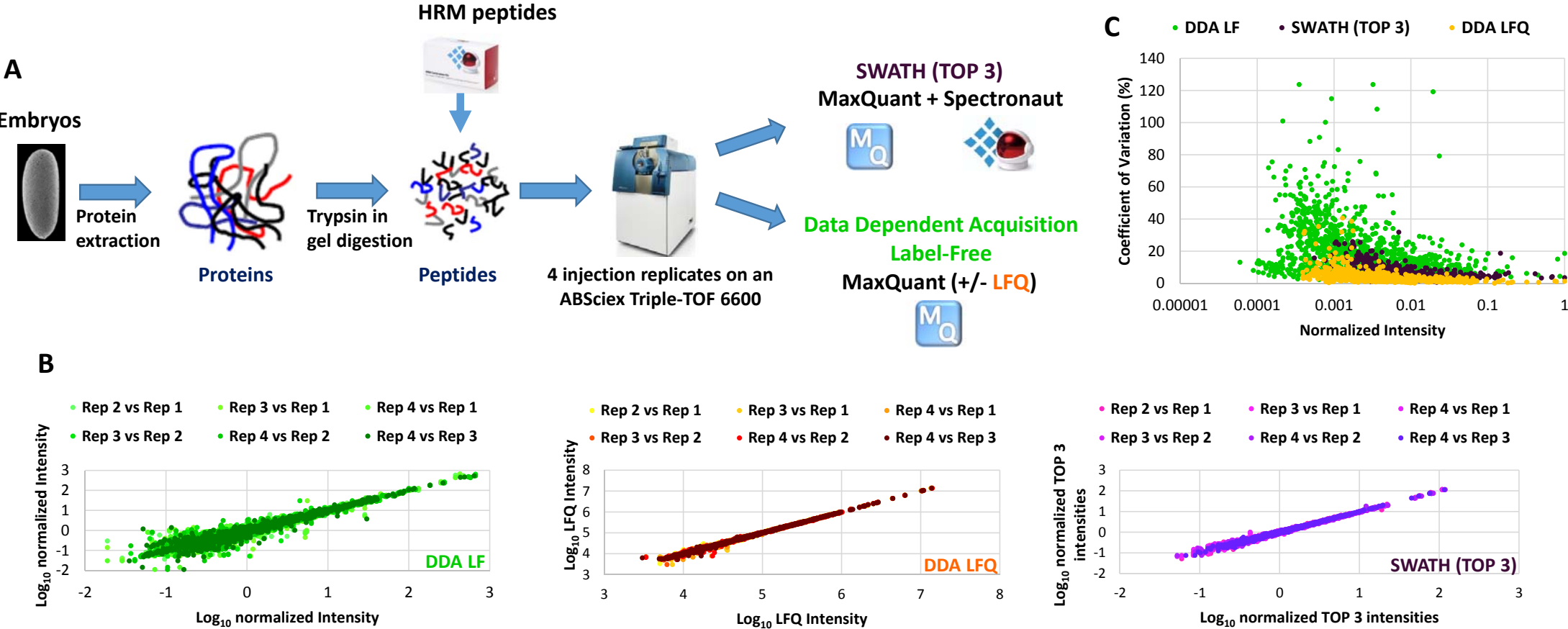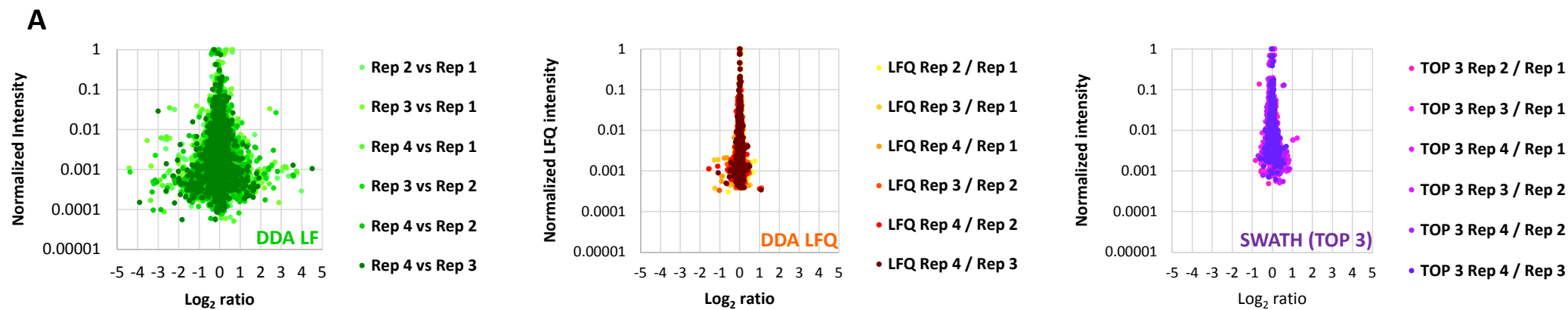
B

	Median CV	Mean CV	Mean R²	Mean slope	Number of proteins quantified	% of Log$_2$ Fold change > 1 or <-1
DDA LF	16.20	21.50	0.938	0.975	1007	8.95
DDA LFQ	3.05	4.13	0.996	1.001	692	0.24
SWATH (TOP 3)	5.19	6.63	0.991	0.968	524	0.10

Figure 3

A

4.5h time points embryos → **Protein extraction** → **Proteins** → **Trypsin digestion** → **Peptides** → **LC-MS/MS analysis on a Triple-TOF 6600**

HRM peptides

SWATH
MaxQuant + Spectronaut

Data Dependent Acquisition Label-Free (LFQ)
MaxQuant

B

SWATH

74 356 98

DDA LFQ
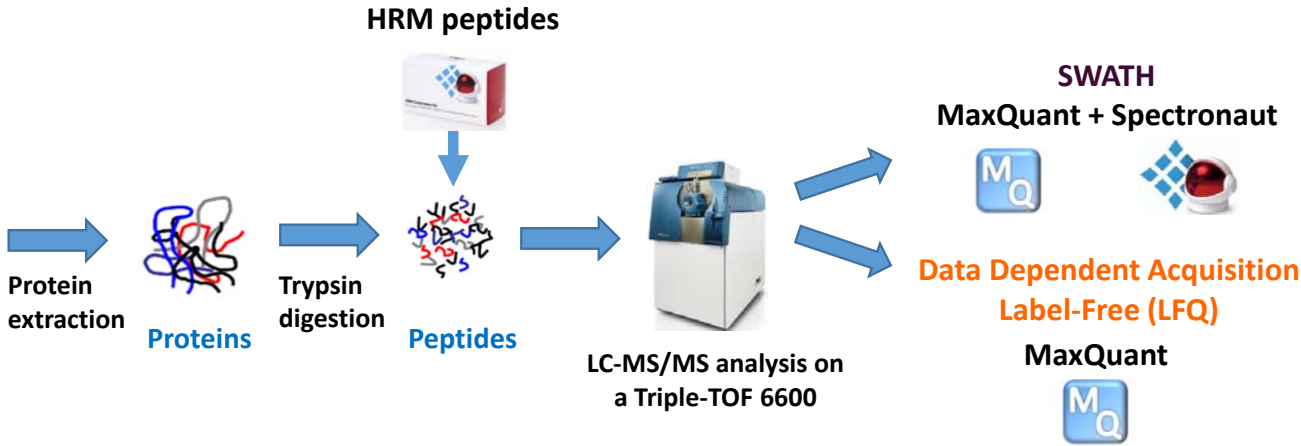
C

• 18-22.5h/0-4.5h   • 13.5-18h/0-4.5h   • 9-13.5h/0-4.5h   • 4.5-9h/0-4.5h

$y = 1.1218x - 0.0284$
$R^2 = 0.7953$

$y = 1.1065x + 0.0105$
$R^2 = 0.748$

$y = 1.049x - 0.0833$
$R^2 = 0.7724$

$y = 0.7466x - 0.0269$
$R^2 = 0.5046$

Fold Change SWATH

Fold Change DDA-LFQ

Figure 4

Figure 5

A

Legend: □ 0-4.5h ■ 4.5-9h ■ 9-13.5h ▥ 13.5-18h ▨ 18-22.5h

B

ATP synthase
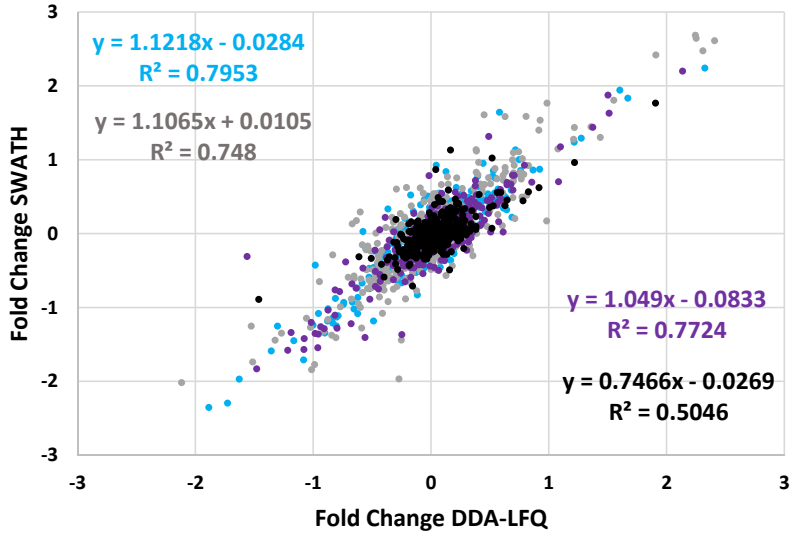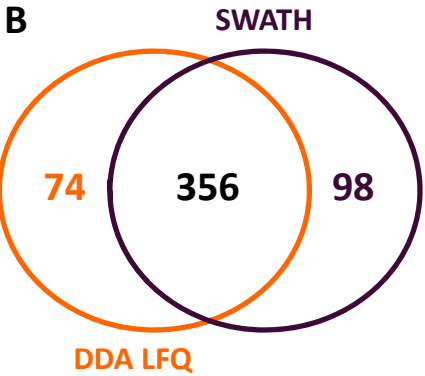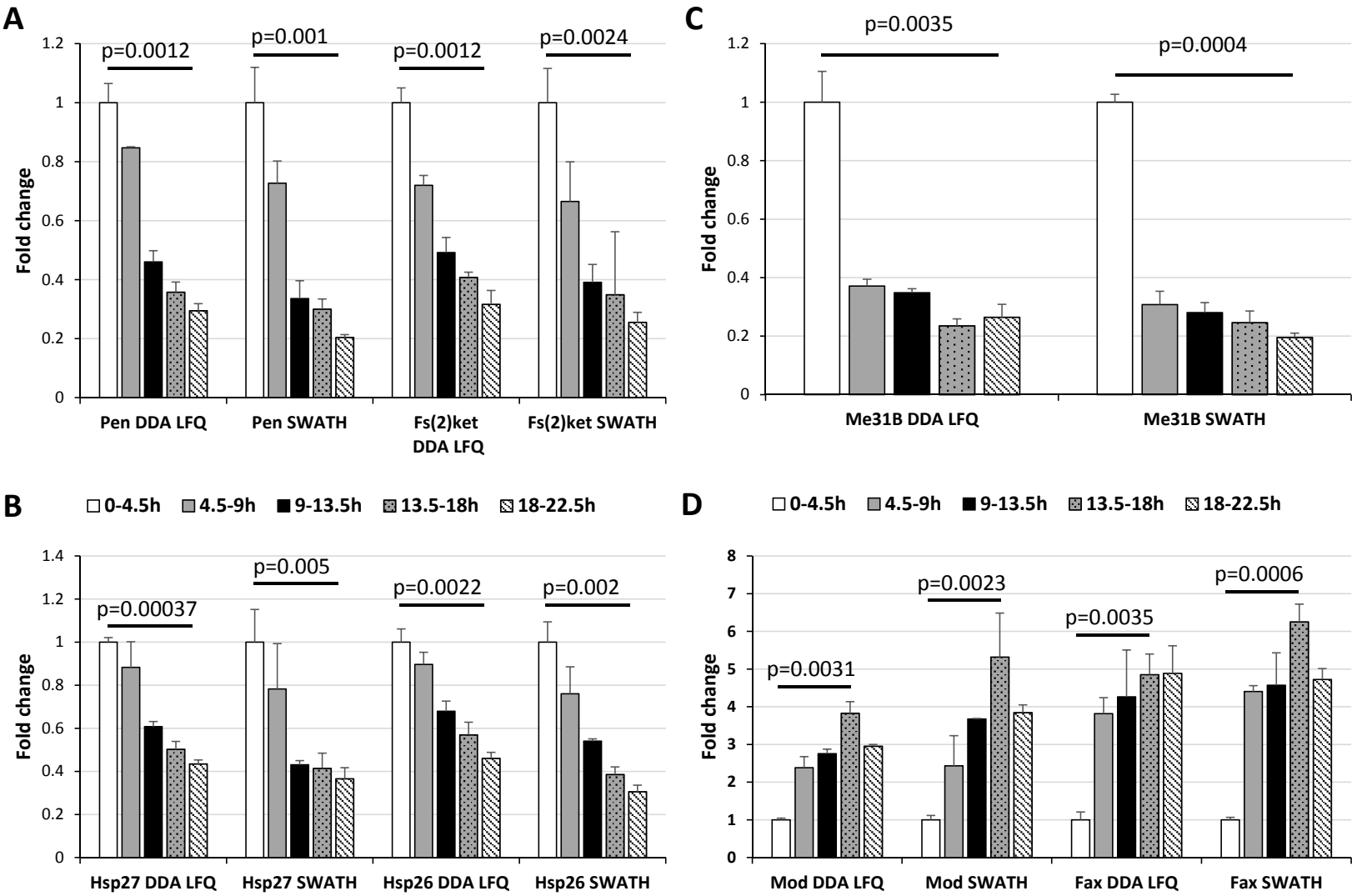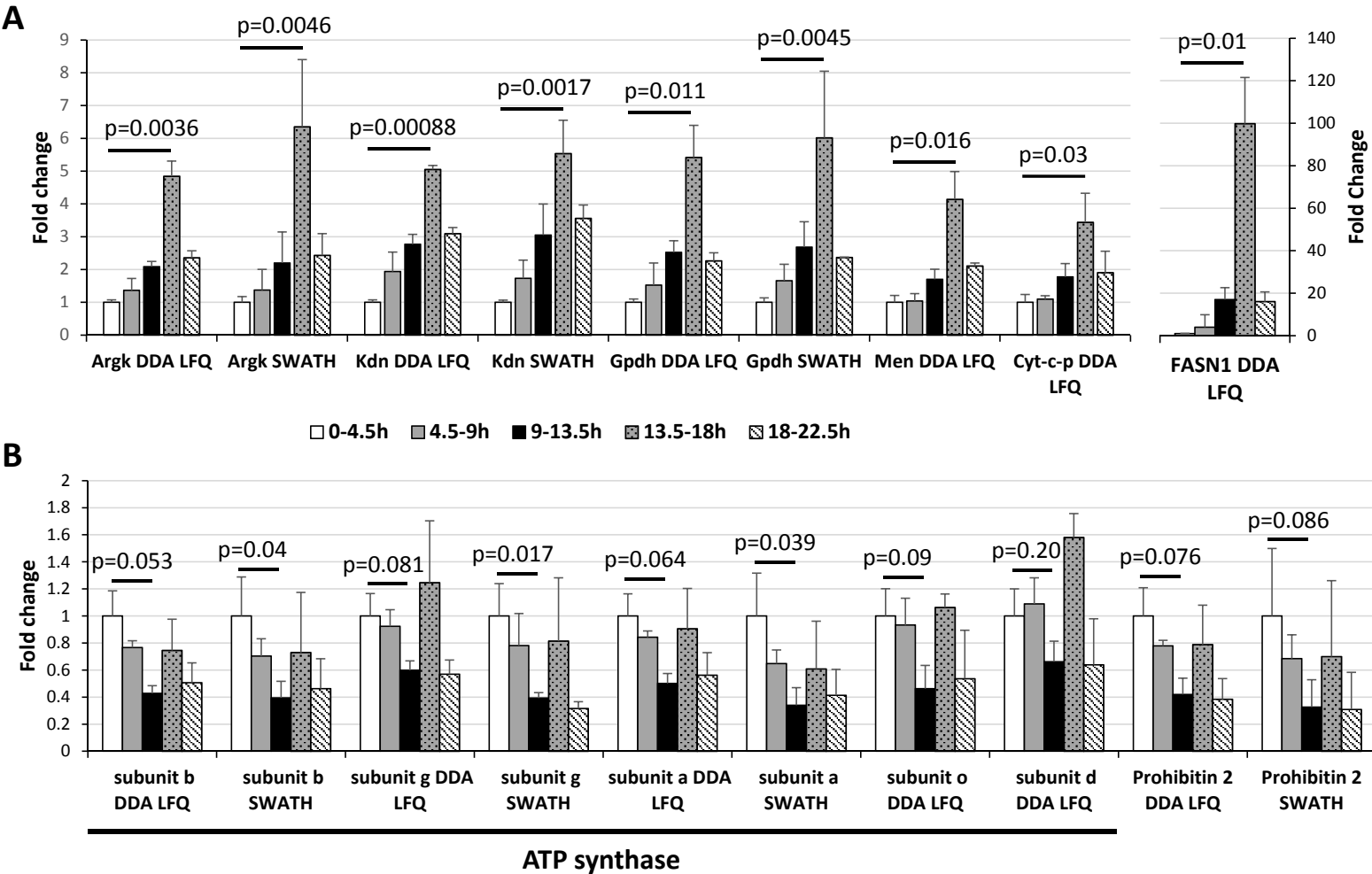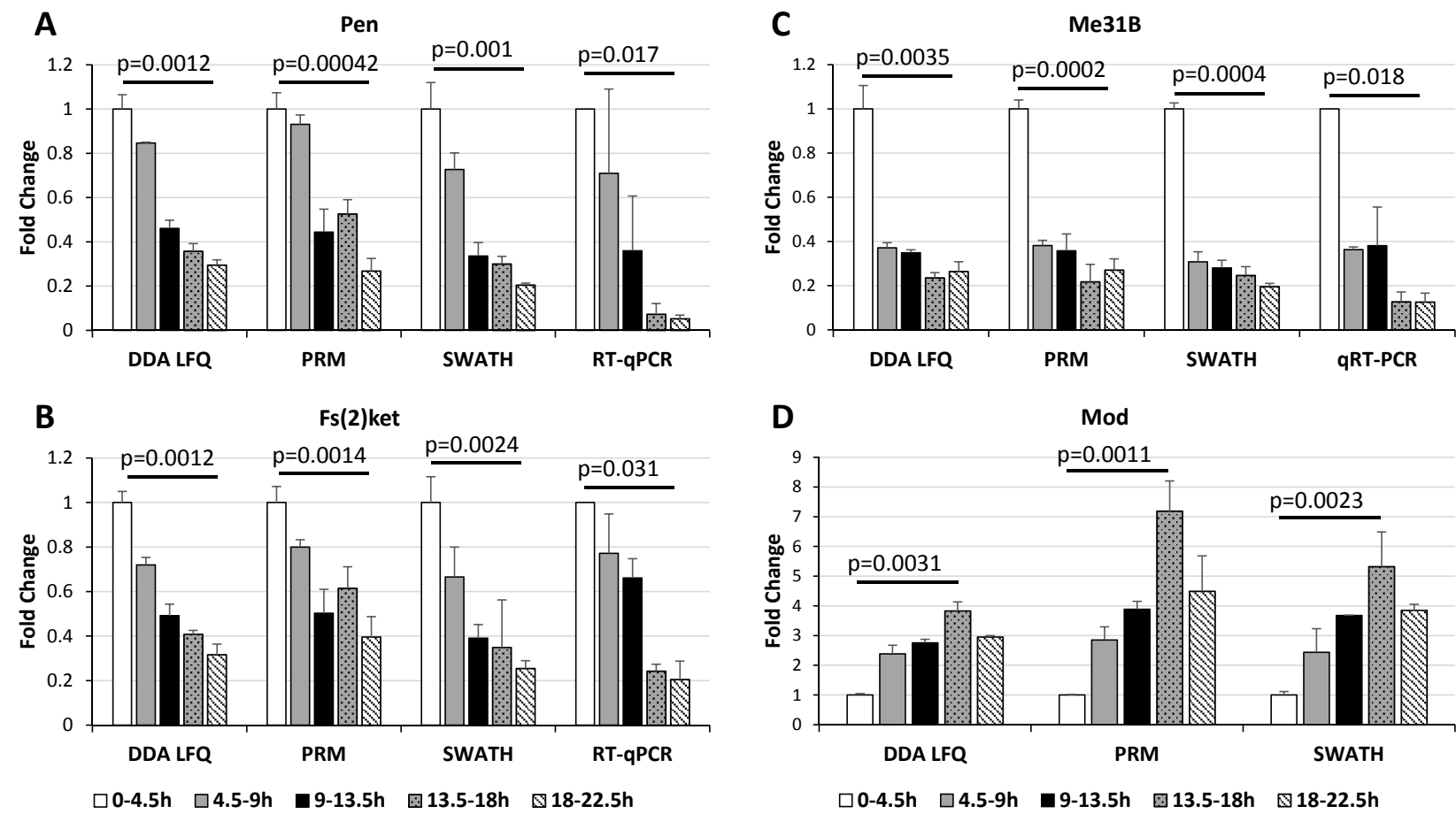
Figure 6

Supplementary Information

Figure legends

Supplementary Figure 1: Comparison of Label-Free quantification methods at the peptide level

A) The Coefficient of variation (CV) was calculated for each peptide quantified using different Q-value filtering for the SWATH analysis. The CVs were plotted as function of the intensities normalized to the maximal intensity. B) Summary of the CVs, run to run reproducibility and identified peptides per run values for the different Q-values filters used for SWATH and DDA analysis. C) The Log_{10} intensities of two representative injection replicates were plotted to evaluate the run to run reproducibility of the SWATH and DDA Label-Free quantification method at the peptide level. D) The Coefficient of variation (CV) was calculated for each peptide quantified using the SWATH and DDA Label-Free quantification method. The CVs were plotted as function of the intensities normalized to the maximal intensity.

Supplementary Figure 2: Comparison of proteins identified in different studies

Venn diagram representing the overlap between the proteins identified in our study with those reported in Gouw *et al*, 2009 and Hughes *et al*, 2014. Only proteins identified with at least two peptides are included in the comparison.

Supplementary Figure 3: Distribution of the coefficient of variations between the biological replicates in the developmental time course

The Coefficient of variations (CV) between the biological replicates were calculated for each protein for every stage and for DDA LFQ and SWATH methods. CVs values were analysed using box plots with R. The red lines indicate CVs of 10 and 20%.

Supplementary Figure 4: Distribution of protein ratios across the developmental time course

The ratios for each replicate at each stage as compared to the mean of the three 0-4.5h stage replicates were plotted for both quantitative proteomics methods. The number of proteins (counts) is plotted as a function of the Log_2 ratio.

Supplementary Figure 5: SDS-PAGE profiles of the different time points

The proteins extracted from embryos at each stage were separated using SDS-PAGE and stained with a colloidal Coomassie solution.

Supplementary Figure 6: Skyline analysis of discrepancies between the DDA LFQ and SWATH ratios

The data were re-analysed with Skyline [37] and two examples for the quantification of the proteins Rps15 (A) and RpL22 (B) were examined to understand why the DDA LFQ and SWATH ratios are different for these proteins. In the case of Rps15 (A), the signal for the MS2 SWATH analysis is close to

the noise which results in poor accuracy for the quantification. The signal for the MS1 XICs based quantification for RpL22 (B) seems contaminated by a wrong extraction of the signal (for the M+1 isotope, purple line). The results of the Skyline analysis show that neither of these two proteins are differentially expressed.

Supplementary Figure 7: Comparisons with other data sets

The ratios for and the 2-4h and 10-12h time points from Hughes *et al*, 2014 and from the 0-4.5h and 9-13.5h time points in our study were compared for the proteins Pen, Fs(2)Ket, Mod and Fax. Standard deviation between the biological replicates are reported.

Supplementary Figure 8: ATP synthase subunit alpha expression profile

Histograms of ratios for the ATP synthase subunit alpha levels compared to the 0-4.5h time point for three different quantitative proteomic approaches (DDA Label-Free quantification, Parallel Reaction Monitoring and SWATH). Error bars represent the standard deviations between the biological replicates. Benjamini-Hochberg corrected p-values are shown for the comparison between the 0-4.5h and the 9-13.5h time points.

Supplementary Figure 1

A

- Q-value < 1.10E-1
- Q-value < 1.10E-2
- Q-value < 1.10E-3
- Q-value < 1.10E-4
- Q-value < 1.10E-5
- Q-value < 1.10E-6
- Q-value < 1.10E-7
- Q-value < 1.10E-8
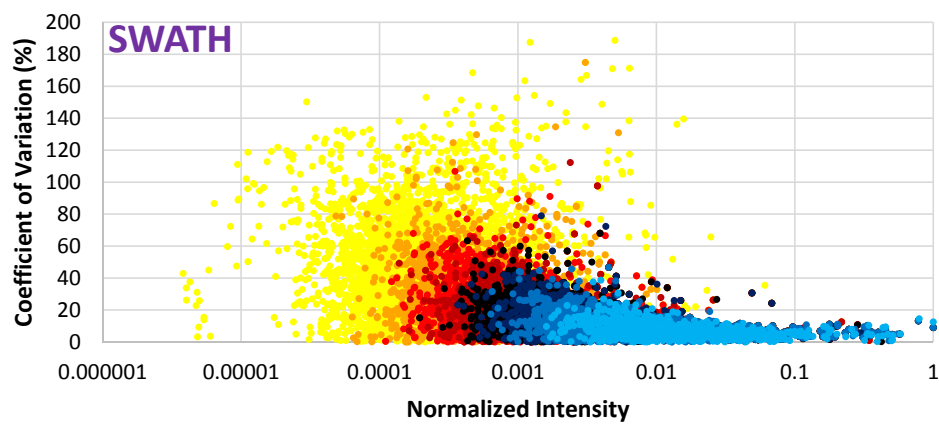
B

Q-value	median CV	mean CV	Mean R²	Mean slope	Number of peptides identified per run
1.00E-01	21.54	28.59	0.880	0.901	12360
1.00E-02	16.22	19.48	0.957	0.940	8888
1.00E-03	13.72	16.03	0.974	0.950	6642
1.00E-04	12.02	13.88	0.981	0.958	4910
1.00E-05	10.43	11.87	0.986	0.965	3464
1.00E-06	8.98	10.40	0.989	0.971	2310
1.00E-07	7.57	8.70	0.992	0.977	1304
1.00E-08	4.64	5.68	0.993	0.970	315
DDA LF	8.74	13.04	0.910	0.962	6165

C

D
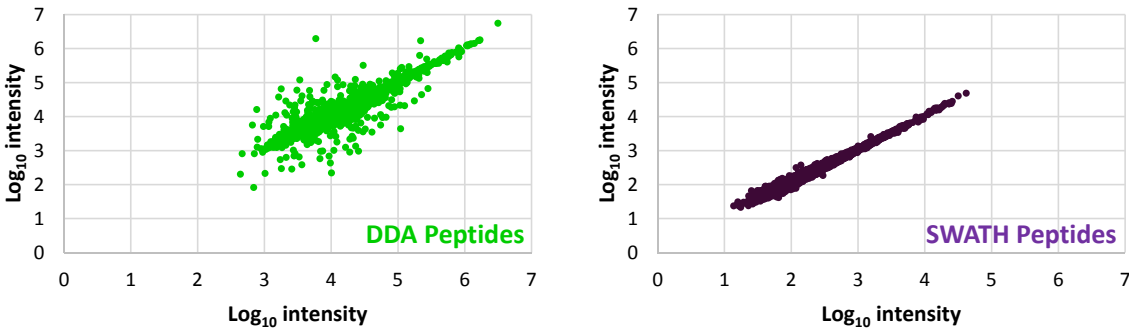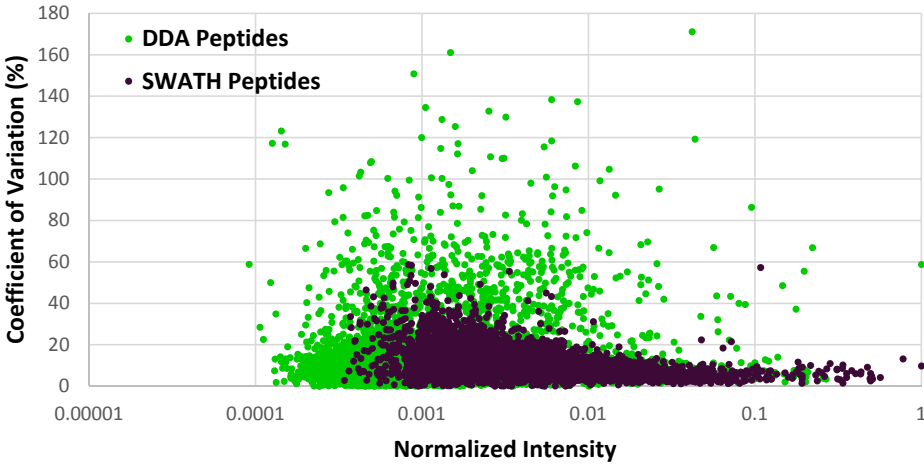
Supplementary Figure 2
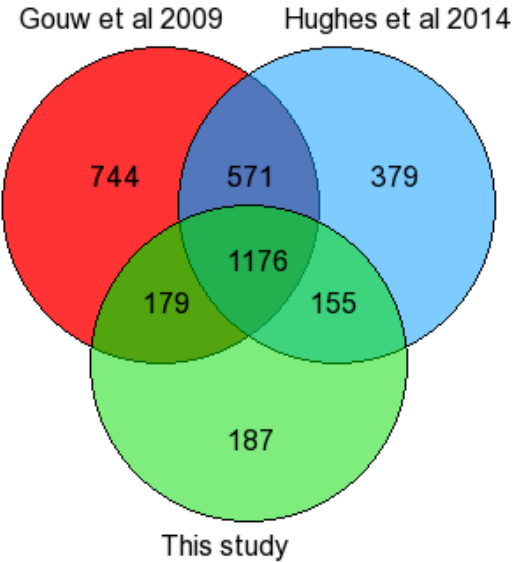
Supplementary Figure 3
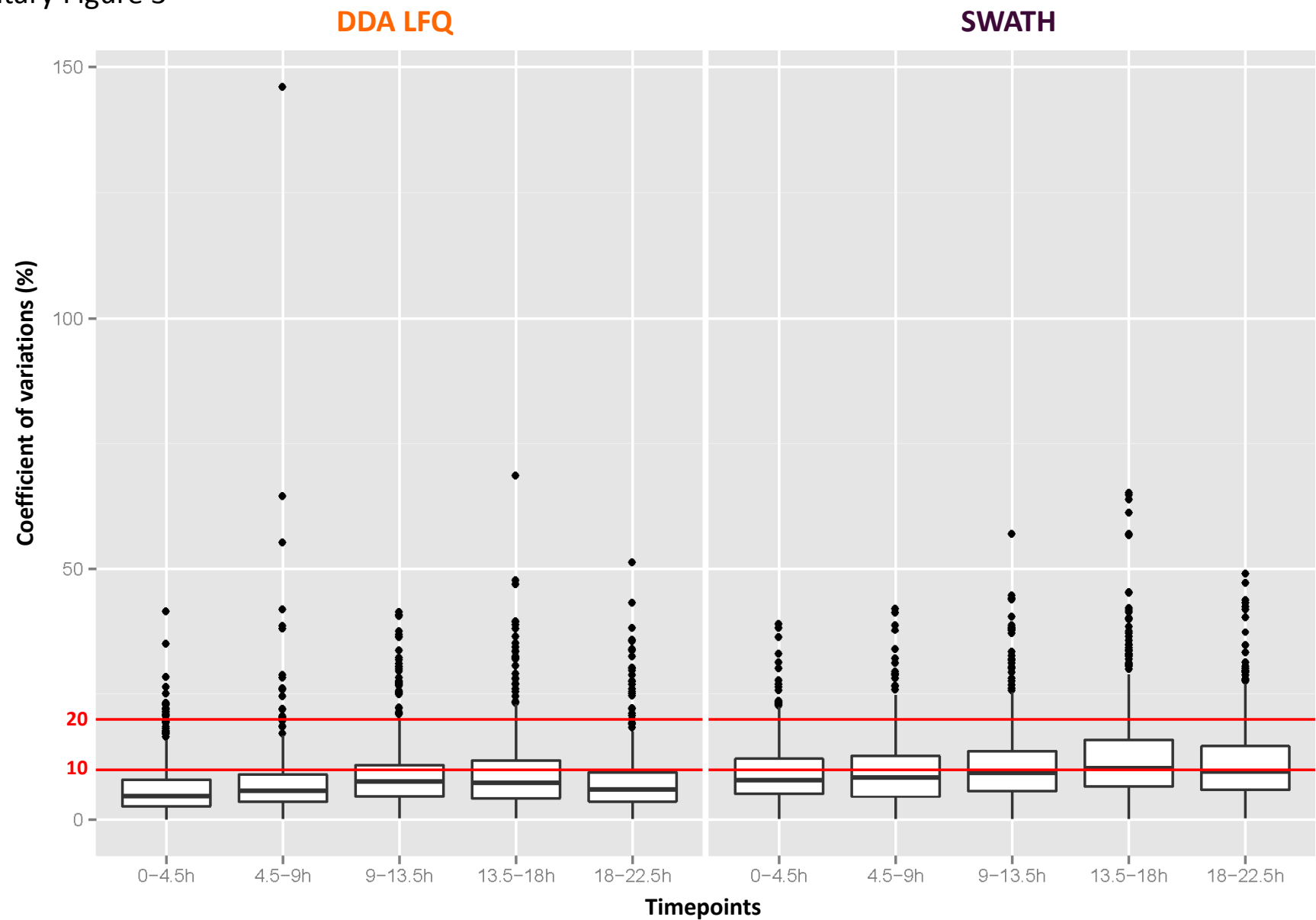
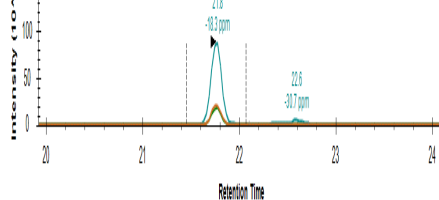Supplementary Figure 4
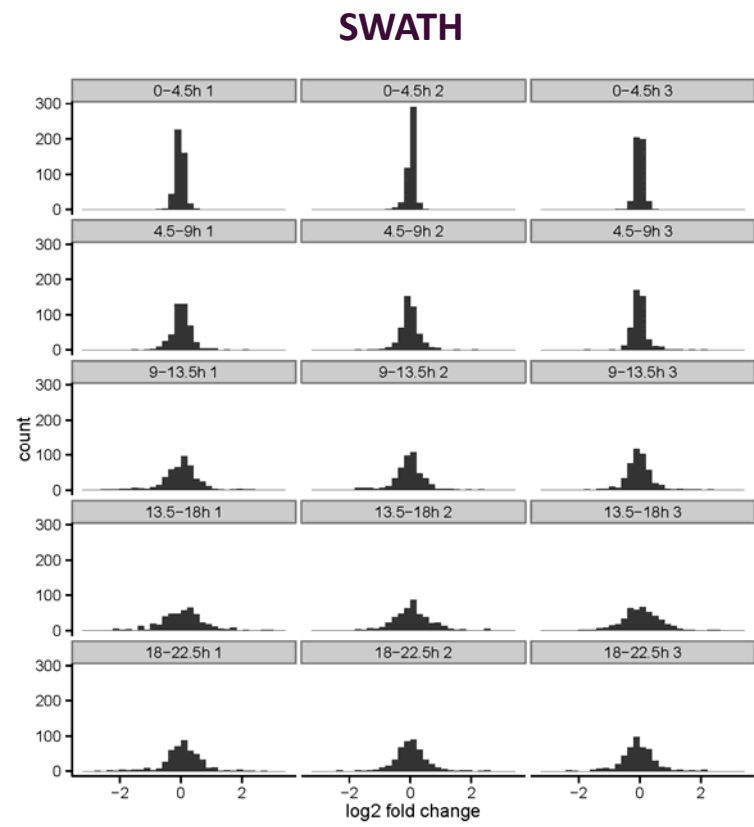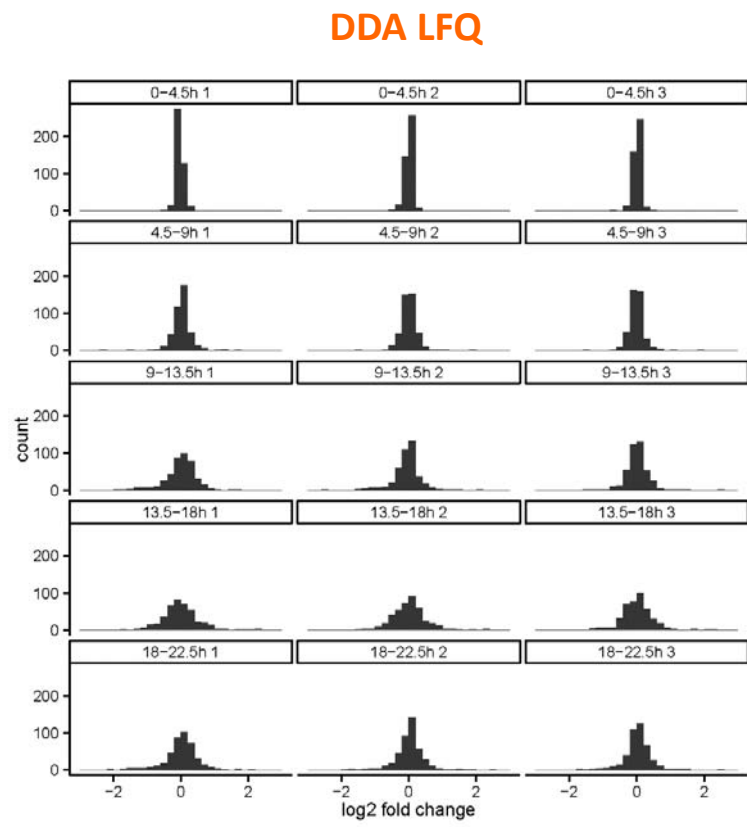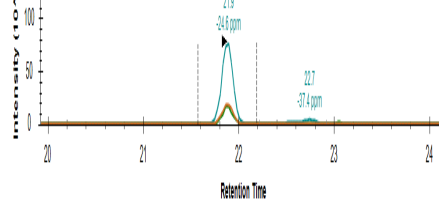
Supplementary Figure 5

Supplementary Figure 6A
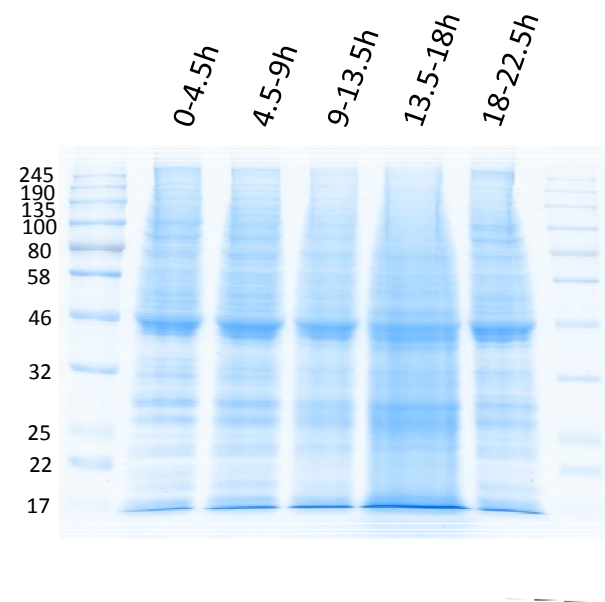
Supplementary Figure 6B

Replicate 1   Replicate 2   Replicate 3

MS1 Filtering
(DDA acquisition)
0-4.5h timepoint
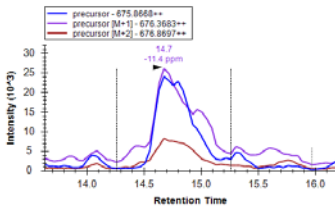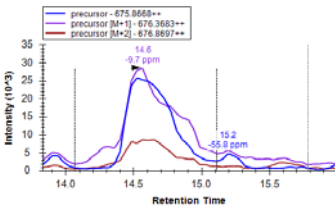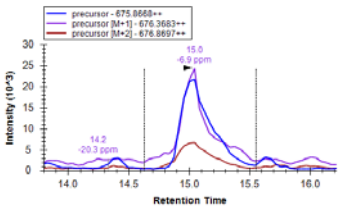
MS1 Filtering
(DDA acquisition)
9-13.5h timepoint
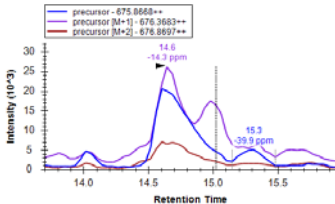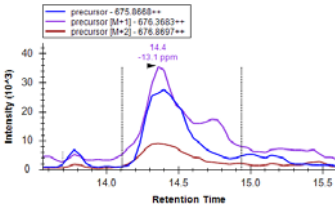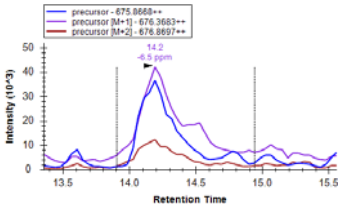
MS2 filtering
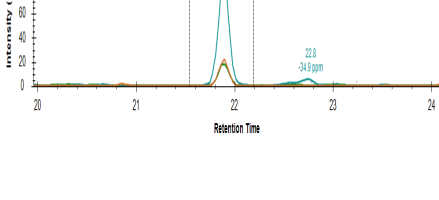(SWATH acquisition)
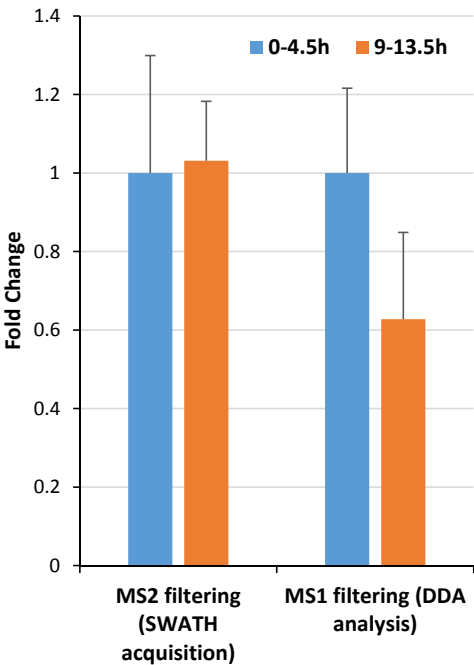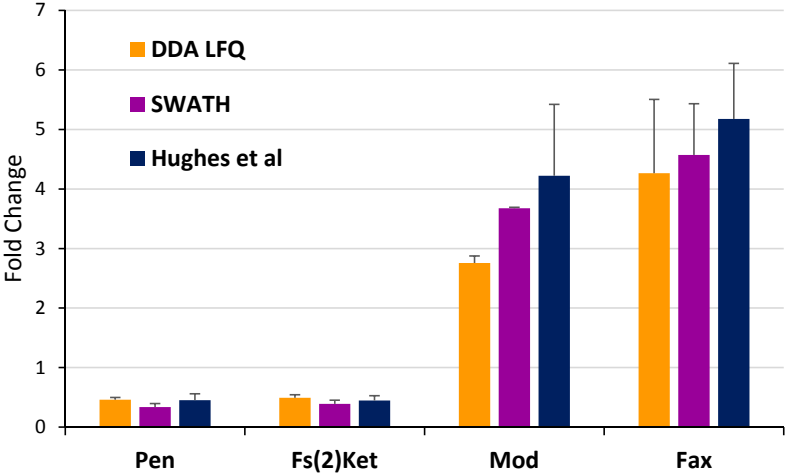0-4.5h timepoint

MS2 filtering
(SWATH acquisition)
9-13.5h timepoint
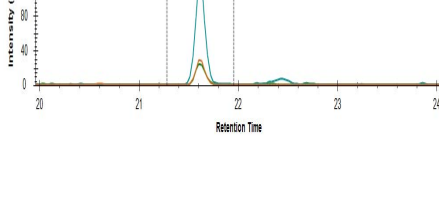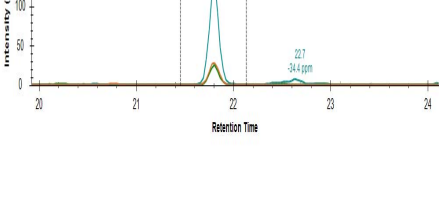
Protein: RpL22 (P50887)
Peptide: VNNLGNNVTFER

Supplementary Figure 7

Supplementary Figure 8

ATP synthase subunit alpha

(Fold Change vs. DDA LFQ, PRM, SWATH; time groups: 0-4.5h, 4.5-9h, 9-13.5h, 13.5-18h, 18-22.5h; p=0.064, p=0.047, p=0.039)