SHIMADZU
Excellence in Science

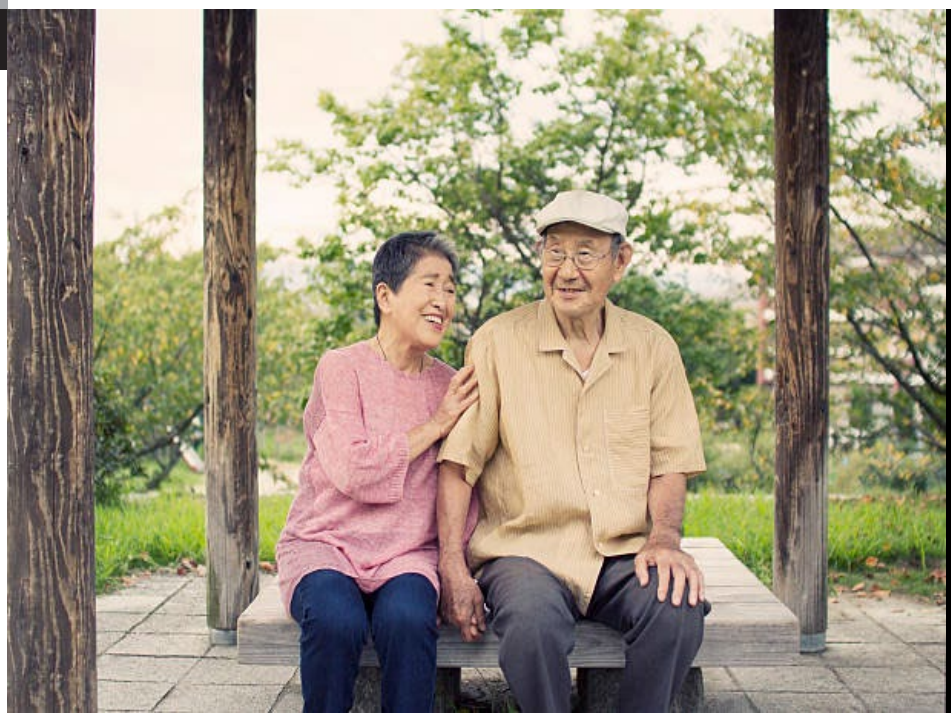
Application Note

No. 98

Life Science

Multi-omics Analysis Using Next-Generation Sequencer and Mass Spectrometer in Longevity Research

Yuki Nakagawa[1], Tsubasa Ibushi[2], Kosuke Kasadera[3], Soshiro Kashio[4]



Life Science

■ Abstract

Using a next-generation sequencer (GridION, Oxford Nanopore Technologies) and a liquid chromatograph-mass spectrometer system (LCMS-9050 with Nexera Mikros™), we compared long-lived Drosophila melanogaster with its wild type. In DNA analysis, 830,000 SNVs/Indels (approximately 40,000 structural variants) were detected by whole genome sequencing, and in RNA analysis, 185 loci with a p-value of 0.05 or less were detected by comprehensive expression analysis. About 1,000 proteins were identified by comprehensive protein analysis, and about 300 metabolites were detected by wide-target analysis. These results were integrated into Multi-omics Analysis Package to study longevity mechanisms.

1. Introduction

In the modern world, demographic change is an important issue. While the world's median population is 30.0 years old, developed countries face challenges such as a declining labor force and rising social security costs as the population ages. In Japan, the median age is 48.4, one of the highest in the world, and together with Italy and Germany, it is attracting attention as a model case of longevity society (Fig. 1).

A detailed analysis of Japanese demographics shows that senility is the third leading cause of death, and the rate has risen sharply in recent years (Fig. 2). Health problems such as nursing care, frailty, and dementia are on the rise as the number of elderly people increases. To address these issues, research is urgently needed to extend healthy life expectancy.

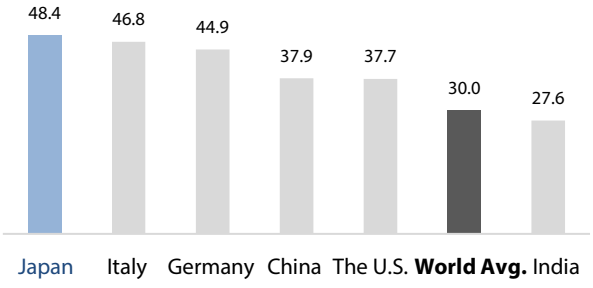


Fig. 1 Median age worldwide (2021 years) [1)]

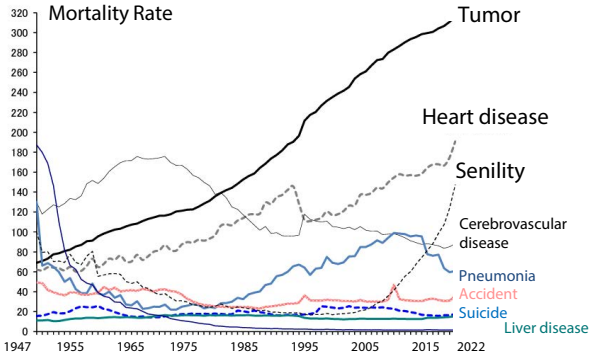


Fig. 2 Annual trends in mortality rates by major cause of death (Population: 100,000) (Japan) [2)]

1 Shimadzu Corporation
2 Shimadzu Techno-Research Inc.
3 Infocom Corporation
4 The University of Tokyo

The secret to improving your health span is tricky and complicated. Therefore, the highest human age, 122 years (Guinness record in 1997), has not been broken for more than 20 years, and only life expectancy has increased [3]. To address these challenges, basic research typically uses model organisms such as flies, mice, and Arabidopsis rather than humans. Drosophila melanogaster is one of the ideal model organisms for studying longevity, with a short generation cycle and various genetic manipulation techniques.

In this application, we present an example of a comprehensive analysis of multiple biochemical processes and molecular mechanisms involved in longevity using multi-omics analysis (DNA, RNA expression, protein, and metabolite analyses), which projects the measurement results of next-generation sequencers and mass spectrometers onto a metabolic pathway diagram (Fig. 3).

2. Experimental

To evaluate Drosophila melanogaster (Wild-type n=2, long-lived n=2, each sample 5-10 Drosophila melanogaster ground together) from eight backcross generations, we measured DNA, RNA, protein, and metabolites using the following three instruments.

Next Generation Sequencer GridION

For DNA measurement, the samples were extracted using the QIAamp DNA Mini Kit, and 200 µL of the extract (O.D. 260/O.D. 280 = less than 2.0) containing approximately 20 µg/mL of DNA was used. GridION used Ligation Sequencing Kit V14 to base call in 5 kHz (400 bps) run mode with R10.4.1 flow cell and SUP accuracy (Guppy Ver.7). 830000 SNV/Indel mutations and 39,000 structural mutations were detected.

For RNA measurement, Total RNA was extracted using the RNeasy® Micro Kit, and 30 µL of the extract (O.D. 260/O.D. 280 = less than 2.0) containing approximately 20 µg/mL of RNA was used for measurement. GridION used the PCR-cDNA Barcoding Kit to perform base calls in 450 bps run mode with R9.4.1 flow cell and HAC accuracy (Guppy Ver.7). 5920 loci were detected, including 185 statistically significant loci (p-value 0.05 or less).

Liquid Chromatograph Mass Spectrometer Nexera Mikros-LCMS9050

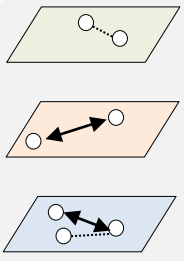
Each sample was bead broken in phosphate-buffered saline (PBS) and then solubilized by adding an equal volume of 10 % SDS in 100 mM TEAB pH8.5. The S-Trap protocol was followed for Trypsin digestion and alkylation with Iodoacetamide [4]. The buffer was dried in a Speed vac, redissolved in 0.1 % formic acid, and assayed by non-target analysis in a Mikros-9050. The Nexera Mikros is a microchannel liquid chromatograph mass spectrometer system, and the LCMS-9050 is a quadrupole time-of-flight (Q-TOF) mass spectrometer. The mobile phase was 0.1 % formic acid water and acetonitrile, and the trap column was YMC-Triart C18, 12 nm, 5 µm, 0.3 mm I.D. 50 mm, and the analytical column was YMC-Triart C18, 12 nm, 1.9 µm, 0.3 mm I.D. 150 mm, and 944 proteins were identified.

Gas Chromatograph Mass Spectrometer GCMS-TQ8040 NX

Using GCMS-TQ8040 NX and Smart Metabolites Database™ Ver. 2, 488 primary metabolites, including organic acids, sugars, nucleic acids, fatty acids, and amino acids, were measured in 23 minutes. Area values were corrected with 2-isopropylmalic acid, an internal standard. Less than 300 metabolites were detected in each sample.

Fig. 4 Equipment images
Multi-omics Analysis Package (top), Nexera Mikros™ and TOF System (bottom left), GCMS-TQ™8040 NX (bottom right)

Single-omics analysis concept

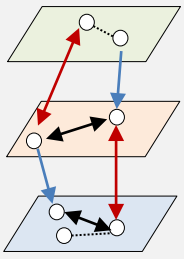


Genomics
Species identification and relative abundance quantification

Metabolomics
Organic acids, sugars, nucleic acids, fatty acids, and amino acids

Phenotype (medical record data)
Age, BMI, race

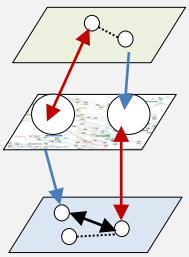
Conventional multi-omics analysis concepts



A series of T-tests based on published papers.

Example) Infection with strain A increases the amino acid tryptophan, resulting in phenotype X

Shimadzu Multi-omics Analysis Package



A series of T-tests based on metabolic pathways and published papers.

Example) Infection with species A leads to the stalling of the tryptophan formation pathway and the accumulation of precursors. Accumulated precursors cause phenotype X.

- PCA on metabolic pathways
- Volcano plots on metabolic pathways

Fig.3 Concepts for single-omics and multi-omics analyses

3. Results

PEAKS Studio XPro software (Infocom Inc.) was used for protein data analysis (Fig. 5) obtained with Mikros-9050 (DDA*1). First, the peptide's amino acid sequence (including post-translational modifications) was calculated from the MS/MS mass peak list by a function called de novo sequencing. The number of peptides detected in the wild type alone was 963, the number detected in the mutant alone was 1,148, and the number detected in common was 3,666 (Fig. 6).

Using the *de novo* sequencing results (list of peptides), Uniprot's Gene Count 13824 database identified protein names at an FDR of 1 % (p-value for multiple testing: 1 % for both peptides and proteins) [5].

*1 DDA stands for Data Dependent Acquisition. It is a method to detect and identify proteins with high to low abundance in descending order. As a comparison, the Data Independent Acquisition method acquires all MS/MS spectra regardless of abundance., resulting in massive data and complex data analysis.

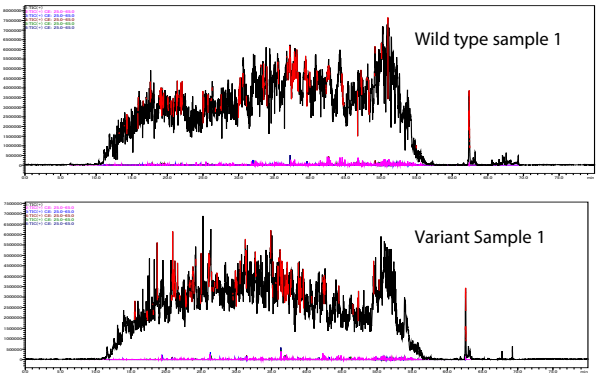


Fig. 5 Comprehensive proteomics measurement chromatogram (LabSolutions)

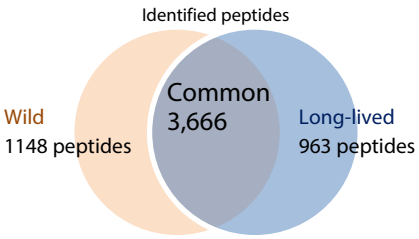


Fig. 6 Peptide identification with PEAKS Studio

A database search was performed on Fixed PTM: Carbamidomethylation and "Variable PTM: Oxidation, Acetylation (N-term)." Fig. 7 shows the integrated peptide scores of the peptides that were hit as peptides of the corresponding proteins. There was no significant difference in the score distribution between the wild and long-lived types, indicating that both the measurement and the analysis were good.

Using these 5,777 peptides, we identified 944 proteins. Of the 944, 162 proteins were detected only in the wild type, 126 proteins were detected only in the long-lived type, and 656 proteins were detected in both groups (Fig. 8). These identified proteins were subjected to ANOVA analysis (Significance > 10, Fold change > 1.3), and 28 statistically significant proteins were visualized in a Volcano plot and heat map (Fig. 9 and Fig. 10). The results showed that 12 proteins, including CG9075 (ATP-dependent RNA helicase eIF4A), a protein involved in germ cell formation, were characteristically high in the wild type. In addition, 16 proteins, including CG2171 (triosephosphate isomerase), essential for energy production (glycolysis pathway), were enriched explicitly in long-lived Drosophila.

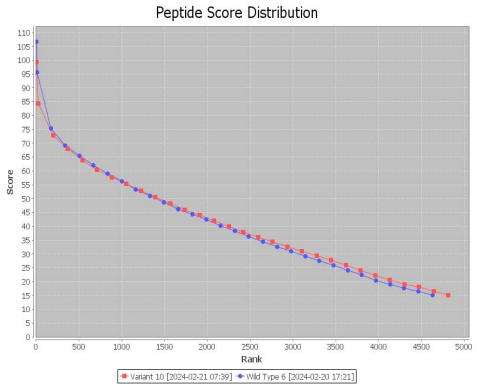


Fig. 7 Peptide identification score distribution

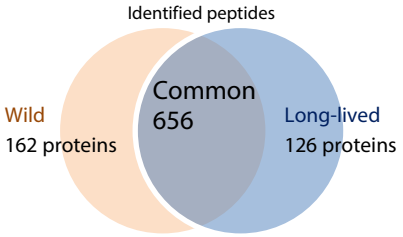


Fig. 8 Protein identification with PEAKS Studio

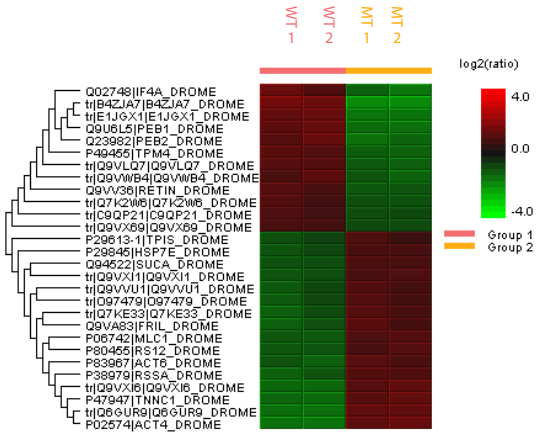


Fig. 9 Heat map analysis in PEAKS Studio
Proteins detected by ANOVA Significance > 10 and Fold change > 1.3

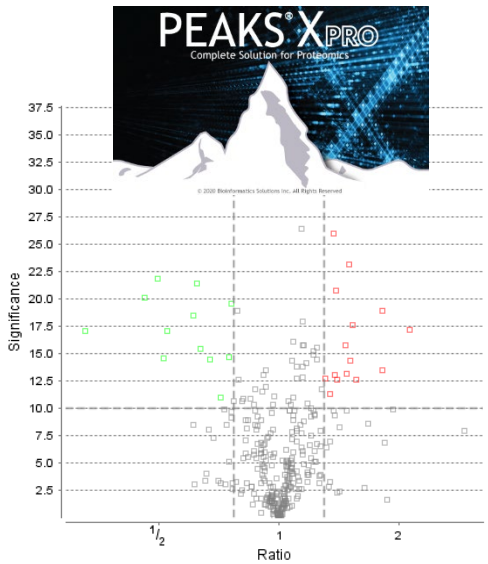


Fig. 10 Volcano plot analysis in PEAKS Studio

Fig. 11 Multi-omics analysis workflow

CG14792 (40S ribosomal protein SA) is one of the proteins detected explicitly as high in the long-lived form in Fig. 9. This protein, which is required for oogenesis and adult development, was detected explicitly at high levels in long-lived Drosophila, suggesting that the process of adult development was delayed in long-lived Drosophila compared with wild-type. CG14792 is required for the formation of the 40 S ribosomal subunit, and the mechanism of the longevity mutation may involve these ribosomes [6].

We can infer differences between sample groups upstream (DNA, RNA) or downstream (Metabolites, phenotypes) by performing statistical analyses such as Volcano plots in proteomics. However, when we measure DNA, RNA, metabolites, etc., we sometimes get results that are not expected from proteomics considerations. Therefore, it is vital to use a Multi-omics Analysis Package to simultaneously analyze not only the metabolites of interest but also their metabolic pathways and to promote multi-omics analysis by using a method that combines not only one statistical analysis method but also multiple statistical analysis methods (e.g., projecting the results of principal component analysis onto a metabolic pathway diagram) (Fig. 11).

Rather than analyzing the results of these next-generation sequencers on their own, the combined analysis of proteins and metabolites with fewer variables (e.g., organic acids, sugars, nucleic acids, fatty acids, amino acids, etc.) is expected to significantly reduce the analysis time of the next-generation sequencers (Fig. 12).

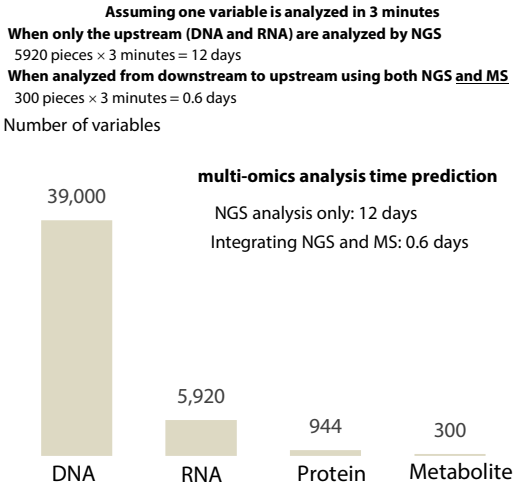


Fig. 12 Number of variables found in each hierarchy

Unexpected Results and Reasons for Connecting Upstream Omics to Metabolomics, and Shimadzu Solution

Not detected

- Concentration is too low
- Method is not optimized

[Shimadzu Solutions]

- Measure by MRM mode using Smart Metabolites Database Ver.2 (more sensitive than Scan)
- Predict the behavior of an undetected compound from those in downstream and upstream of that compound in the metabolic pathway by Multi-omics Analysis Package

Not annotated (or wrongly annotated)

- Buried under matrix peaks
- Retention index is not used in annotation
- Standard solution is not used in annotation
- CID fragmentation is not used in annotation

[Shimadzu Solutions]

- Measure by MRM mode using Smart Metabolites Database Ver.2 (more selective than Scan)

No significant difference between groups Or difference within the same group

- Method is not optimized
- The number of samples is too small

[Shimadzu Solutions]

- Use "Pretreatment Procedure Handbook for Metabolites Analysis"
- Analyze ca. 500 metabolites in 23 mins by Smart Metabolites Database Ver.2, and increase the throughput and the number of samples

Metabolite that should decrease increases in stead or vice versa

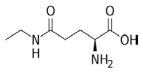
- Method is not optimized
- The number of samples is too small

[Shimadzu Solutions]

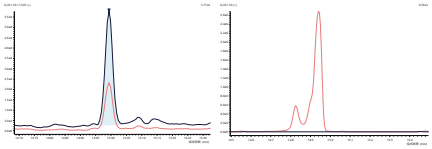
- Use "Pretreatment Procedure Handbook for Metabolites Analysis"
- Analyze ca. 500 metabolites in 23 mins by Smart Metabolites Database Ver.2, and increase the throughput and the number of samples

An example of compounds not detected by Scan

Theanine-2TMS



MRM Scan



Smart Metabolites Database Ver.2

Fig. 13 List of Shimadzu solutions for multi-omics analysis with metabolites

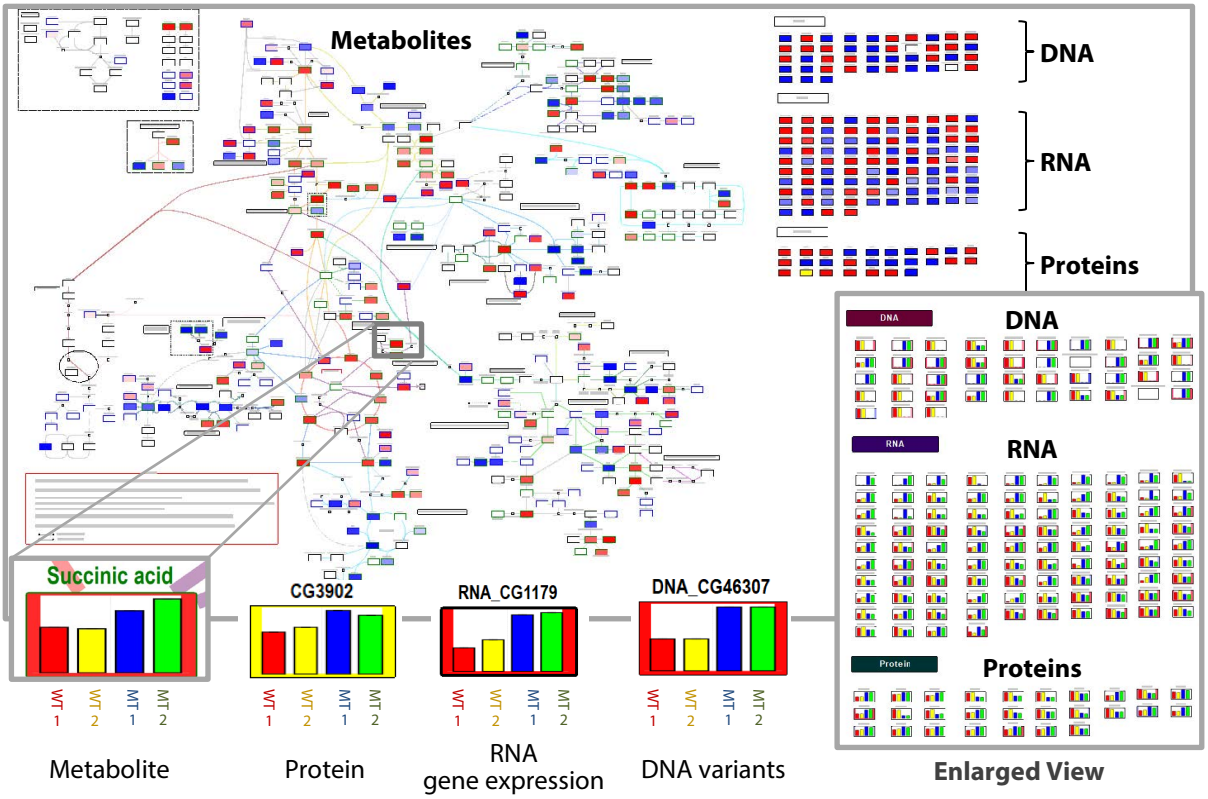


Fig. 14 Multi-omics Analysis Package (Added nodes for DNA, RNA, and protein to standard Metabolite Pathway Map)
*As shown in Fig. 9, only filtered DNA, RNA, and protein are projected. For DNA, 0 for 0/0, 1 for 1/0 or 0/1, and 2 for 1/1 were entered.

However, when linking upstream (DNA and RNA) and downstream (metabolites) omics, confusion can occur when the metabolite of interest may be undetected or mis-annotated. Therefore, it is crucial for multi-omics analysis to selectively detect peaks derived from metabolites of interest in complex biological samples using optimized MRM methods in the Smart Metabolites Database Ver. 2, and to reliably annotate (identify) peaks using methods identified by standard reagents (Fig. 13). For example, the metabolite theanine (Theanine-2TMS) in the Drosophila sample in this report was not detected by Scan and was detected only by MRM.

Using the method shown in Fig. 11 in the previous section, the number of DNA, RNA, and protein variables was narrowed down by filtering and statistical analysis at each level, resulting in 53 DNA, 94 RNA expression mutations, and 316 proteins. They were projected onto a metabolic pathway map, and a correlation analysis was performed using the protein CG3902 as an example on the pathway (Fig. 14). The software automatically identified DNA, RNA, and metabolites positively correlated with the protein CG3902 in red and negatively correlated in blue.

A combined analysis of 654 variables was performed using a Volcano plot to detect variables that were specifically detected in the wild type (top left: 7 in the green box) and variables that were specifically highly detected in the long-lived type (top right: 15 in the red circle) (Fig. 15). Volcano plots identified variables of interest and then projected onto metabolic pathways. The results showed that homocysteine, the final product of alanine metabolism, accumulated explicitly in the long-lived group, while 2-aminoadipic acid, an intermediate of lysine degradation metabolism, decreased in the long-lived group, promoting the metabolic pathway (Fig. 16).

High levels of homocysteine are known to damage the brain and blood vessels, leading to Alzheimer's dementia. The high homocysteine level in longevity-mutated Drosophila suggests that although longevity increases, health span may not increase [7].

Proteins of unknown function, such as CG15616, are also detected at high levels, specifically in long-lived flies, indicating that investigating their association with metabolites and RNAs detected in the same pattern of variation will be an issue for the following study.

In the long-lived form, 2-aminoadipic acid is a metabolite that is present in aging human skin and is a marker of carbonyl oxidation of skin proteins.8). However, its morphology in Drosophila has not been reported, and it is necessary to keep an eye on RNAs and proteins that are also shown in green in metabolic pathways.

In this way, by color-coding the Volcano plot and the metabolic pathway diagram to determine which DNA, RNA, or protein is specifically detected in one group, we can study the relationship between genetic loci, mutations, or proteins of unknown function and metabolic pathways.

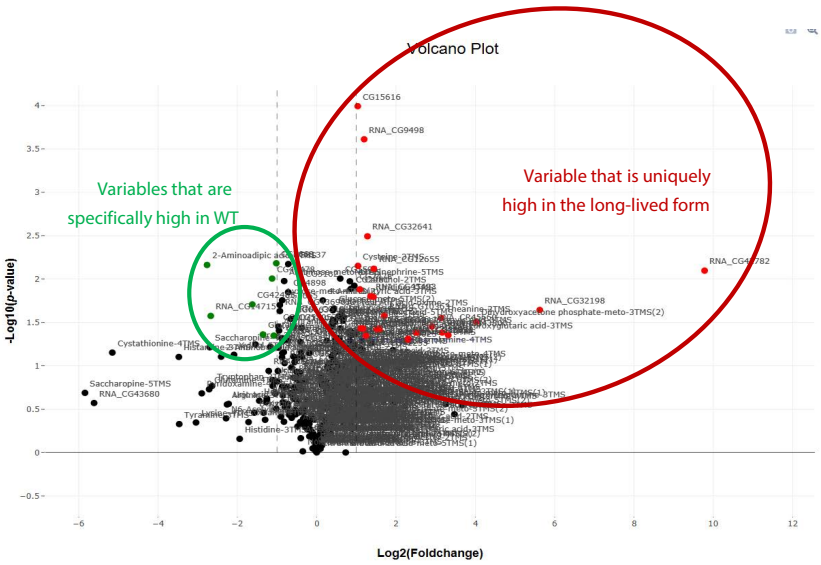
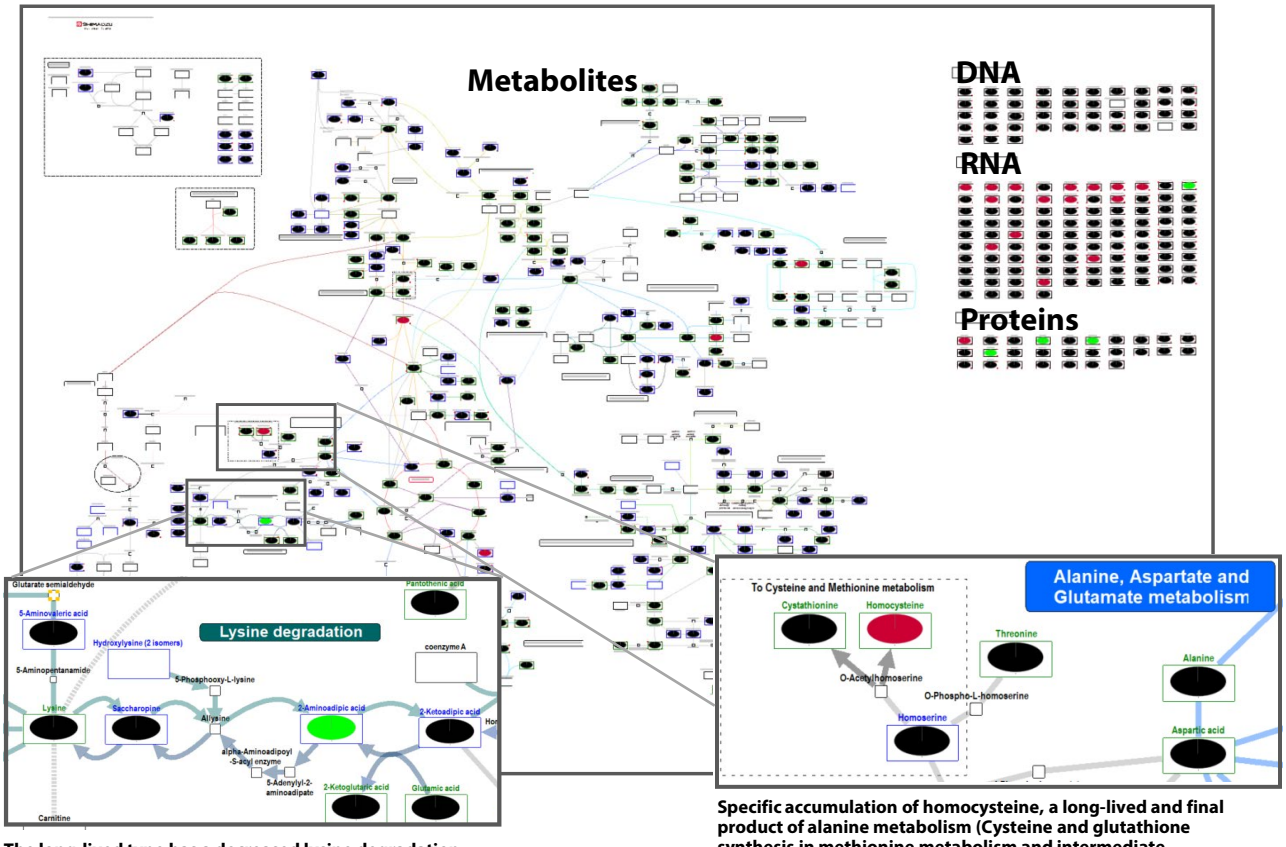


Fig. 15 Volcano plot analysis of wild type (n=2) and long-lived (n=2) (654 variables for DNA, RNA, protein, and metabolite data)



The long-lived type has a decreased lysine degradation metabolic intermediate(i.e., promoted metabolism)

Specific accumulation of homocysteine, a long-lived and final product of alanine metabolism (Cysteine and glutathione synthesis in methionine metabolism and intermediate metabolites of the methionine cycle)

Fig. 16 Analysis of volcano plot results projected onto a metabolic pathway map

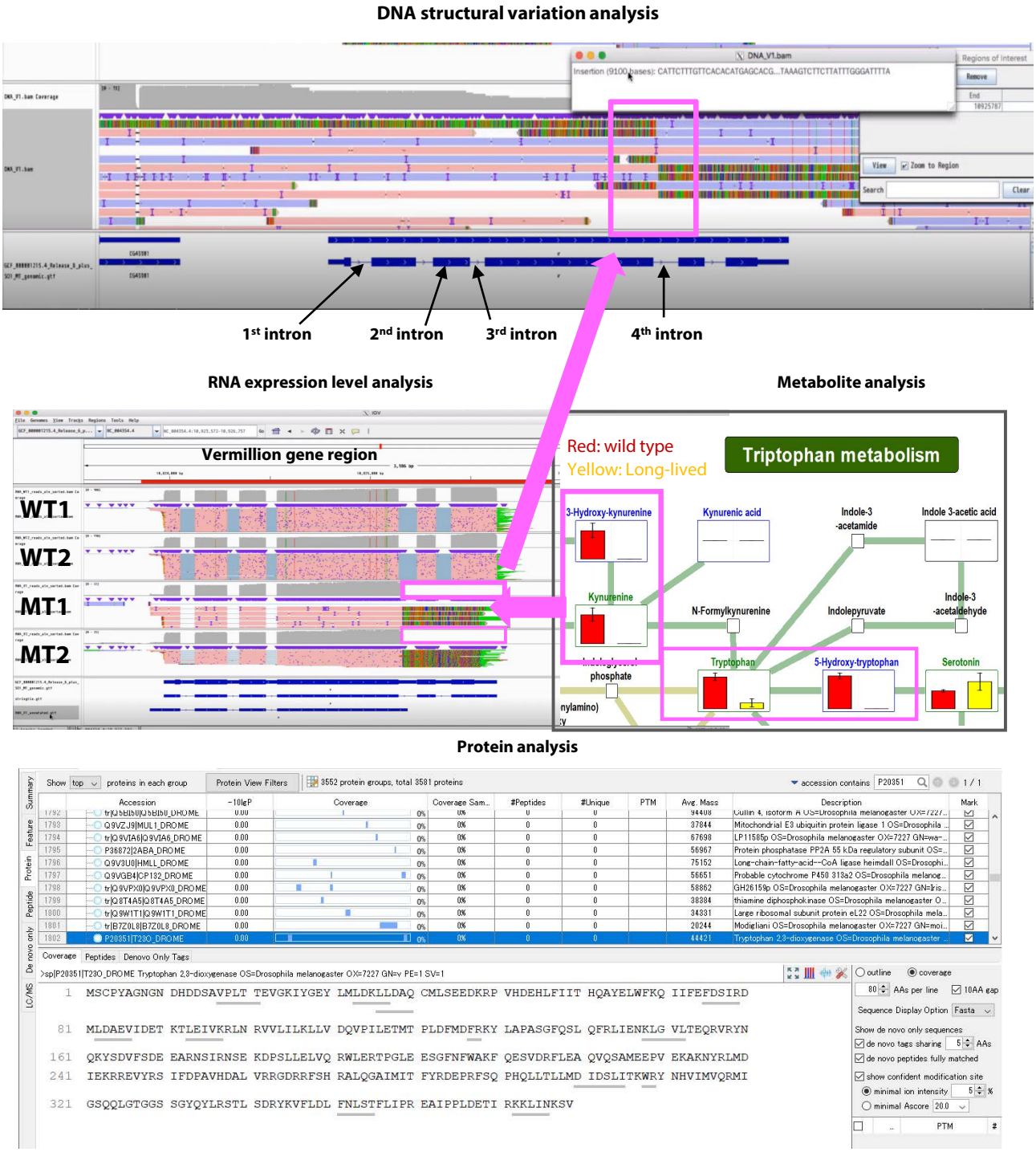


Fig. 17 Suppression of the kynurenine metabolic pathway in long-lived Drosophila and confirmation of TDO2 RNA expression level variation and insertion into intron 4 of the Vermillion locus(Screen capture of IGV for DNA and RNA, PEAKS Studio for proteins, and Multi-omics Analysis Package for metabolites)

In addition, since the kynurenine metabolic pathway was suppressed in the long-lived pathway, the concentration and expression levels of tryptophan 2,3-dioxygenase (TDO2), a rate-limiting enzyme, were confirmed (Fig. 17). Although the proteomic analysis did not provide reliable results due to the low concentration of the corresponding protein, RNA expression analysis showed structural defects in the long-lived Vermillion gene region. Therefore, DNA analysis of the Vermillion locus revealed a clipping near the fourth intron and an insertion about 9,100 bases long.

We improved the efficiency of multi-omics analysis by utilizing metabolic pathway analysis rather than a single analysis of 830,000 SNVs/Indels and 39,000 structural variations, which are genetic differences between wild and long-lived types.

It is possible to identify substrates and products by calculating the ratios of each metabolite, placing them on a grid, and performing correlation analysis between the ratios and the RNA or protein of interest (Fig. 18 on the next page).

Positive correlations of CG31508, a protein that affects stress sensitivity, were automatically colored in red, and negative correlations in blue. The results showed a positive correlation with the ratio of 2-deoxy-D-ribose as a product and 6-phosphate fructose as a substrate. On the contrary, it was negatively correlated with the ratio of Allose as a product and 2-deoxy-D-ribose as a substrate. This suggests that CG31508 accumulates 2-deoxy-D-ribose.

4. Conclusion

We used a next-generation sequencer and a liquid chromatography-mass spectrometer system to compare Drosophila melanogaster and its wild type with long-lived gene mutations. In DNA analysis, 830,000 SNVs/Indels (approximately 40,000 structural variants) were detected by genome-wide analysis. In RNA analysis, 185 loci with a p-value of 0.05 or less were detected by comprehensive expression level analysis. About 1,000 proteins were identified by thorough analysis, about 500 metabolites were determined by wide-target analysis, and about 250 metabolites were detected. These results were integrated into a Multi-omics Analysis Package. By analyzing the results of next-generation sequencers (DNA, RNA) in combination with proteins and metabolites with fewer variables (Organic acids, sugars, nucleic acids, fatty acids, amino acids, etc.), the analysis time of the results of next-generation sequencers is shortened, and multi-omics analysis can be promoted. Using the metabolic pathway analysis in the Multi-omics Analysis Package, we observed suppression of the tryptophan pathway in the long-lived form, and we were able to detect insertion mutations in the Vermillion locus in DNA and RNA. Possible future projects include enrichment analysis as well as DIA measurement in proteomics.

<References>
1) World Population Prospects 2022, United Nations, accessed on Feb. 27th, 2024)
2) 令和4年（2022年）人口動態統計月報年計（概数）の概況 [Summary of the Annual Vital Statistics Report 2022 (2022)], 厚生労働省, accessed on Feb. 27th, 2024
3) 120歳は長寿の上限？スーパーセンテナリアンの世界 [Is 120 the upper limit of longevity? supercentenarian world], 日経新聞, accessed on Feb. 27th, 2024
4) Category: Protocols, PROTIFI, accessed on Feb. 27th, 2024
5) Proteomes Drosophila melanogaster (Fruit fly), UniProt, accessed on Feb. 27th, 2024
6) P38979 RSSA_DROME, UniProt, accessed on Feb. 27th, 2024
7) アルツハイマー病や大病を引き起こす高ホモシステインとは？ [What is high homocysteine that causes Alzheimer's and major diseases?], DIAMOND online, accessed on Feb. 27th, 2024
8) 2-aminoadipic acid is a marker of protein carbonyl oxidation in the aging human skin: effects of diabetes, renal failure and sepsis, David R Sell, accessed on Feb. 27th, 2024

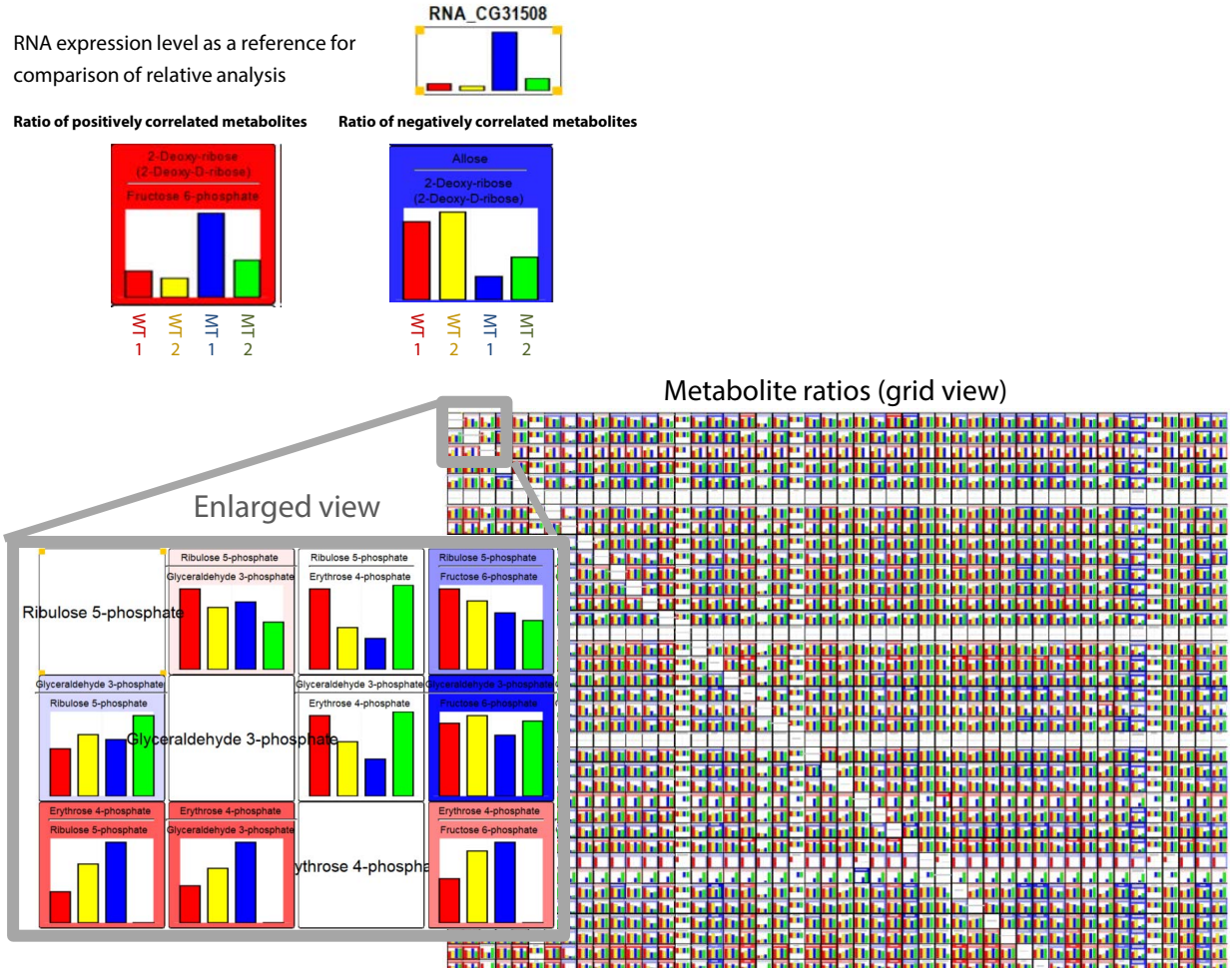


Fig. 18 Search for substrate/product pairs with the same pattern of intergroup differences as a particular RNA expression detected by NGS (e.g., CG31508 was used in this example) (Positive correlation with RNA expression in red, negative correlation in blue)

SHIMADZU
Shimadzu Corporation
www.shimadzu.com/an/

First Edition: Apr. 2024