Introduction
Agilent recently had the opportunity to speak with Dr. Atsushi Shimizu, a professor at the Iwate Medical and Dental Research Institute and Deputy Director of the Iwate Medical Megabank Organization, about our customizable SureSelect target enrichment kit. In our conversation, we asked about his work, his thoughts on our custom NGS panels and Targeted DNA Analysis kit for epigenomic analysis, and how these technologies have benefited his research.

Q. Please tell us about the Iwate Medical Megabank (IMM) and the Tohoku University Tohoku Medical Megabank (ToMMo) Organizations.

These organizations have jointly implemented the Tohoku Medical Megabank (TMM) project, the goal of which is to support recovery from the 2011 Great East Japan Earthquake, which marks its 10th anniversary at the end of FY2020. In addition to the questionnaires for cohort studies, information such as health checkups and biological samples are provided by the researchers. Due to storing this information, these organizations are able to function as a biobank, enabling the transfer of biological samples (and corresponding biological information) to outside research institutions.

Our division of biomedical information analysis consists of a wet lab and a dry lab, where DNA and RNA are extracted and used to obtain genomic data from donated samples. In addition, the department is developing technologies to more efficiently obtain genomics data from these samples, and is also researching methods to predict the individual’s risk of disease onset (and personalized prevention strategies) using the information obtained.
Q. Please give us an overview of your research using SureSelect custom design.

We have been using next-generation sequencing (NGS) to examine DNA methylation information of TMM project participants to examine CpG methylation rates and analyze their association with disease.

As a prelude to conducting epigenomic analysis with capture sequencing, we examined whether the already designed capture sequencing system would be effective in assessing the risk of developing the multifactorial diseases we were targeting.

As you know, whole blood, which is collected and used without further processing, contains a variety of blood cell components. We wanted to reduce the number of cell types to make it easier to find markers, so we focused on monocytes and CD4-positive T lymphocytes by using a cell sorter to extract these cells from the blood of TMM project participants. Whole-genome bisulfite sequencing of monocytes and CD4-positive T lymphocytes from roughly 100 people revealed the enormous diversity of methylation states at 30 million CpG site. Additionally, we were able to examine known epigenomic markers associated with diseases to further analyze associations.

When performing genomic analyses of hereditary diseases, focused exome analysis is used to collect rare polymorphisms in order to compare patients with diseases to healthy patients. On the other hand, in the case of multifactorial diseases, it is rare to find loss-of-function mutations in exons, so we collect and analyze single-nucleotide variants (SNVs) with a large inter-individual diversity.

We believed that this would also hold true for epigenomic analysis, so we hypothesized that regulatory regions are important for diseases such as cancer (in which cell characteristics are altered) and multifactorial diseases, (in which there is a large inter-individual diversity).

We then conducted whole-genome bisulfite sequencing analysis on roughly 100 people to investigate the relationship between regions with different DNA methylation rates and known disease-associated epigenomic markers.

We found that areas of high epigenomic diversity were important in the analysis of multifactorial diseases, so we decided that capture sequencing should be used to collect and analyze high-diversity areas. So, to target peripheral blood mononuclear cells (PBMCs) for analysis, we created a custom panel of genes targeted towards monocytes and CD4-positive T lymphocytes and with target enrichment of areas exhibiting high methylation diversity.

However, since large cohorts and historical biobanks are essentially only preserved with whole blood samples, this raised several experimental considerations. Therefore, we next performed whole-genome bisulfite sequencing of neutrophils, which make up the majority of whole blood, to analyze regions with greater methylation diversity.

Using the results, we then designed a second custom panel to collect regions of high methylation diversity in neutrophils along with monocytes and CD4-positive T lymphocytes for epigenomic association analysis of whole blood (Figure 1).

We are now in the process of designing and analyzing a third panel that not only encompasses as many epigenomic markers of disease as possible, but also removes genes that did not work in the panels we designed for whole blood.

What makes our design unique is that it focuses specifically on areas of relevance for multifactorial diseases, rather than the standard regulatory areas that are found in off-the-shelf catalog products.

![Whole-Genome Methylation Area Map](image)

**Figure 1.** Overview of the target enrichment NGS method used in PBMCs, CD4-positive T lymphocytes, and neutrophils.
Q. While a comprehensive view of CpGs is ideal, I believe you also considered the fact that adding more CpGs to your design increases the costs. How did you select which to examine when creating your design?

The advantage of target enrichment sequencing is that, while it is expensive to try to get a comprehensive analysis, costs can be significantly reduced by excluding less important areas. For example, when analyzing SNVs in the genome, one can analyze a very small amount of data because the data is digital: either a 1 (present) or a 0 (absent).

However, because methylation analysis is analog data with a diversity of methylation rates, a more advantageous approach is to gain depth with capture sequencing of your areas of interest than to perform bisulfite sequencing of the entire genome. Therefore, there is an advantage to capture sequencing in epigenomic analyses.

The basic principle of our custom design was to load the epigenomic markers that had been discovered at that time and exclude the areas where we had designed probes but could not get reads. In the case of our exome analyses, we increased the number of probes in the regions where we could not obtain reads. However, regions that require a large amount of data to obtain a read are unsuitable for capture-based sequencing. Therefore, we decided to exclude these regions from the analysis because of the cost they would add to methylation analyses.

Q. Why did you choose Agilent's custom design?

Compared to other products, Agilent makes it easier to design panels and has experience with exomes and with arrays. We chose Agilent because they are responsive to users when they consult with us. As a user, I feel that the strength of Agilent strength is their excellent customer support so I can talk to them without worry.

Q. Besides custom design, what other Agilent products have you found helpful?

For reagents, we also use SureSelect XT Human Methyl-seq, an Agilent catalog product. We have not yet designed a set that covers both cancer and multifactorial diseases, so we use SureSelect XT Human Methyl-seq for analysis of cancer and other serious diseases, and custom designs for multifactorial and lifestyle diseases.

In terms of instrumentation, we use the TapeStation system (a fully automated electrophoresis system). Additionally, as we may process samples from up to 1000 patients, our library preparation is automated using a Bravo Automated Liquid Handling system (NGS library preparation automation system). I am especially grateful for the support from Agilent, which has developed a new program for us to use.

Q. How do you feel about your satisfaction with our company?

The Agilent employees who serve as our contacts in Japan show their commitment to supporting the promotion of research by researchers, which makes me feel comfortable asking for help. In addition, when I had a problem, they provided good support and so I did not have to worry about it. I often worry about whether or not a company will be able to handle a new project, but I don't have to worry about that with Agilent.
Q. Can you tell us about your research plans for the future?

Our research group is currently conducting a prospective cohort study. This is not a study to investigate the causes of disease, but rather to find changes prior to the onset of disease and prevent them. The epigenome of multifactorial diseases, which we have been focusing on, is gradually changing and is affected by changes in the epigenome. As such, we aim to find changes in disease markers at an early stage and use them to help patients improve their lifestyle. Additionally, TMM is conducting a three-generation cohort study and we are thinking about doing an epigenomic analysis of the cord blood of newborn babies when they are born. There is a hypothesis that lifestyle in childhood and the maternal environment during pregnancy affect the future development of disease, which is called DOHaD, and research is being conducted in humans as well as in mice. From this point of view, I think it would be meaningful to proceed with the research in a prospective cohort if we could establish a system for identifying future susceptibility to diseases by analyzing cord blood methylation and improving lifestyle habits at the earliest possible stage.

Q. Once you’ve explored the methylation markers associated with multifactorial diseases, are you considering what tools you can use to put them into practice?

If you are targeting a small number of CpGs, I think one would choose the pyrosequencing method for methylation analysis. Alternatively, Dr. Kanai (Department of Pathology, Keio University School of Medicine) has created a system for calculating DNA methylation using the high-performance liquid chromatography method, which may be used when the number of target CpGs is smaller. However, when analyzing the risk of developing a disease with genomic information, a small number of polymorphisms alone do not explain the disease. It is a large number of weakly affected polymorphisms that form a disease. If the analysis continues at this rate, perhaps tens or hundreds of thousands of epigenomic locations will need to be analyzed to properly assess the risk of disease development. Then it may be necessary to make the capture sequencing method cheaper or to create a microarray that can analyze the target region.

Q. Are there differences between races in areas where there is significant individual variation in methylation status?

At this point, we have not directly compared the data from large-scale methylation analyses, but analysis of the polygenic model of genomic information has shown that race has a large effect. Therefore, the markers for the epigenome are likely to also differ between races, and I suspect that future analysis of a large number of CpGs will show a completely different pattern.

Q. What are your requirements for Agilent?

Agilent has been developing microarrays, NGS, and other instruments needed to explore new medical and life sciences as soon as they are invented. This has allowed not only the most advanced researchers, but a wide range of researchers to take advantage of the new technology, broadening the scope of research as a whole. As new devices are invented in the future, we will continue to develop new products in the form of kits.