

Scalable High-Throughput Array Enables Ethnic Chinese Genome Database Development

Fast Infinium™ array analysis and broad genome coverage support personal genetic testing services and database development for research studies and partnerships.

Introduction

No two humans are genetically identical, whether they are of the same region, race, or family. Sequencing the human genome facilitated our understanding of the role genetics plays in human disease, and the effects that environment, diet, and behavior have on the genetics of human health. It also shed light into the unique genetic differences between ethnicities. Based in Shenzhen, China, WeGene is a personal genetic testing and population genomics company that is capitalizing on the value of understanding these differences among East Asian populations, particularly Chinese ethnicities. In addition to providing genetic testing services to consumers, it is creating an East Asian genome database to support human genome research studies and the development of personalized medicine approaches at research centers worldwide.

Since its founding in 2015, WeGene has grown its personal genetic testing service rapidly, with the resulting data forming the foundation of its East Asian database. The database contains > 200,000 pieces of consumer data, with Chinese samples contributing most of the data. To provide its Chinese customers with a better understanding of their unique genetic profiles and a complete ancestry analysis, WeGene is expanding the database to include more Chinese ethnic groups. The database is expected to grow at > 15,000 data pieces per month, thanks to the company's transition to the Infinium Global Screening Array and iScan™ System solution to achieve higher throughput and broader genome coverage. To scale its laboratory and increase workflow efficiencies, WeGene worked with Illumina Array Lab Services.

iCommunity spoke with Qiang Zheng, Chief Executive Officer, and Gang Chen, PhD, Chief Technology Officer about the rapid growth of WeGene's personal genetic testing services, the increasing value of its Chinese ethnic database, and the role next-generation sequencing (NGS) will play in the development of its future genetic testing products and services.

Q: What is the focus and mission of WeGene?

Qiang Zheng (QZ): As one of the most successful personal genetic testing companies in East Asia, WeGene provides direct-to-consumer (DTC) testing services to customers through online channel marketing and special events. Our mission is to use the genetic data that we gather through our testing service to shine new light on human health, disease research, and drug development, particularly for Chinese ethnic populations.

Q: What is the status of the Chinese DTC genetic testing market?

QZ: The DTC market in China is flourishing. While the proportion of the population accessing genetic testing services is still very small, consumer interest is growing rapidly. There are many companies entering the market because of its potential.

Q: What types of testing services does WeGene provide?

QZ: We provide our customers with data-driven, science-based genetic assessments of their ancestry, as well as health screening for allergies, etc. We integrate DNA and phenotype data, probing and studying the data in depth. Each customer receives a report that outlines the scientific basis, research results, and corresponding research methodology. Together, we review the results to determine how helpful this information is to them.

The lifestyle, phenotypic, and genotypic data that our customers provide form the foundation of our growing genomic database covering all 56 Chinese ethnic groups and several other East Asian populations. Our goal is to build the largest genome database for these Asian populations. We partner with scientific research organizations, enabling our Chinese customers to participate in various genome-related research studies.

Q: Why is the creation of a Chinese ethnic genome database so important?

Gang Chen (GC): Genome technology has had a significant influence on many areas of health and medical research. Yet, most of the genome data available today was not generated from East Asian populations, such as Chinese ethnic groups. There are substantial variations between the genomes of different human population groups, impacting our understanding of disease presentation, course, and prognosis within these groups. Accumulating genomic data for Chinese and East Asian populations and creating reference genomes will be essential for the development of successful personalized therapeutic approaches and diagnostic assays.



Qiang Zheng is Chief Executive Officer and Gang Chen, PhD, is Chief Technology Officer at WeGene in Shenzhen, China.

Q: How are you creating this database?

GC: We feel it is important to interact directly with our customers as we gather genetic data. The value of a database with only genome data is limited. Communicating directly with our Chinese customers enables us to gather genetic, phenotypic, and behavior data for a more comprehensive database. We believe that this type of database will be the most valuable in driving genome research on Chinese population groups. We hope to expand the scale of our genome database quickly, establishing a foundation to support genetic research and application development for Chinese and East Asian population groups.

"The goal of our WeGene Chinese ethnic and East Asian genome database development is to drive the application and study of gene testing technology for the benefit of individual consumers, as well as global research organizations."

Q: What tools and methods are you using to obtain genetic data for these databases?

GC: WeGene uses FDA-approved saliva samplers to ensure the quality of their collection, transport, and storage. A high-throughput microarray platform is used for genetic testing and we perform data quality control and imputation analyses to produce high-quality genetic data to inform our East Asian and Chinese ethnic reference genomes.

Q: What were the limitations of the previous methods and systems that you used to create these databases?

QZ: In the early days after the company was founded, our testing volume and growth rate was small. We considered various solutions, including the Illumina Infinium BeadChip and Affymetrix microarrays. We ultimately chose Affymetrix.

Q: What motivated you switch to a different array solution?

QZ: After two years of development, our sample volume increased significantly. We were having production issues processing such a large number of samples quickly. We needed a more efficient production platform that offered high speed and data quality.

Q: Why did you choose an Illumina array solution?

QZ: We tested and evaluated several genetic testing solutions. Based on our sample volume projections, growth estimates for the China DTC testing market, and data requirements, our best option was a higher performance array-based platform. The only choices were the Illumina array solution, consisting of the Infinium Global Screening Array (GSA) and iScan System, and an array solution from another vendor.

In our final evaluation, we obtained several overseas references, including one from 23andMe, and researched accessory-related issues. We concluded that the Illumina array solution was extremely stable, had a significant worldwide user base, and offered great technical support. It was capable of resolving the production backlog, quality, and efficiency issues exacerbated by our rapidly increasing sample testing volume. Based on efficiency, quality, and service considerations, we chose the Illumina array solution.

Q: Did the sample processing capacity increase after switching to the Illumina workflow?

QZ: With the Illumina array solution, our sample turnaround time decreased from seven to three working days. This substantial efficiency improvement increased our sample processing speed from < 2000 to 15,000 samples per week. That's a significant increase in processing capacity.

Q: Was the ability to automate the Illumina array workflow with a Tecan liquid handling solution an advantage?

QZ: The ability to automate the array workflow is one of the advantages of the Illumina array solution that I value the most. Automation is a key issue for the future in terms of processing large amounts of samples.

Q: What level of data quality have you achieved with the Illumina array solution?

GC: We worked with the Illumina research, development, and production teams on the addition of certain markers to create various custom GSAs to meet our needs. Even if we request the inclusion of many additional custom markers, Illumina synthesis of these arrays has been excellent. These custom GSAs deliver good stability, correlation, and high data quality.

"Based on efficiency, quality, and service considerations, we chose the Illumina array solution."

Q: What are the advantages of using the GSA for genetic testing?

GC: WeGene was part the early GSA design discussions. Since the Broad Institute came out with the initial design solution, we have seen the GSA take a leading position in terms of chip design concepts and design foundation.

The GSA is used in many commercial and scientific research organizations in China and throughout the world, and a large amount of data has been accumulated. The scale and speed of this data accumulation will continue to increase for the foreseeable future. This provides WeGene with a significant advantage as we create our database, providing data compatible with future global research data. Using GSA-based chips in our consumer and scientific research enables WeGene to achieve better results and increases the opportunity to identify disease- and health-associated markers that are ethnicity specific.

The high-quality data generated by the Illumina array solution also provides value to our customers. Other Infinium personal genome products, such as some of the 23andMe DTC tests, have received regulatory approval in the United States. Results that we obtain from performing personal genome testing with the Infinium GSA and iScan System will support our own efforts to gain regulatory approval in China and other regions. The Illumina array solution provides our customers with trustworthy and reliable DTC genetic testing.

Q: Have you started using the Illumina Asian Screening Array (ASA)?

GC: We haven't started using the ASA yet. We will be working with Illumina to create custom ASA chips for Chinese ethnic groups that will support the development of future WeGene genetic testing products.

" With the Illumina array solution, our sample turnaround time decreased from seven to three working days."

Q: What benefits did the Illumina ArrayLab Consulting Service provide in establishing the new laboratory workflow?

QZ: The Illumina ArrayLab consulting team has extensive experience in planning and designing large-scale production laboratory operations. They assisted us in establishing the facility design and determining the number of systems needed for an efficient workflow. From a consumables standpoint, we worked with them to create a production array shipment schedule to support our increasing sample volume. The ArrayLab team also helped us determine staffing and external resource forecasting, and automation and process control requirements necessary for a scalable, high-performance laboratory. The team's experience and guidance enabled us to avoid potential problems we might have faced throughout the construction and development process had we coordinated everything ourselves. The ArrayLab Consulting Services team enabled us to start running our lab on a mature, established model in a faster and more efficient way.

Q: What software pipeline are you using for data analysis?

QZ: Illumina assisted us in designing a high-throughput analysis software pipeline. We have optimized the standard data analysis pipeline provided by Illumina for our large-scale genomics data factory. A significant number of computing nodes are necessary to efficiently process the data we produce. We also developed several cloud components to enable the analysis and interpretation of the data.

Q: Are you assessing NGS for the development of genetic tests to expand your DTC product offering?

GC: WeGene is focused on offering the latest genetic testing technologies and best methods to our customers. To obtain comprehensive personal genome data, we need to sequence the entire genome. Using NGS, particularly whole-genome sequencing (WGS), to develop new genetic testing products will

enable us to provide customers with a better, more informative personal genome testing and analysis service. Because WGS analyzes the whole genome instead of individual genes, it provides a deeper assessment of genetic variation.

We're using the latest sequencing systems in our WGS pilot studies, including the NovaSeq™ 6000 System. Through collaborations with academic research teams at various global organizations, we're performing data analysis and management of thousands of customer whole-genome sequences.

The pilot study sequencing results have been impressive. WGS delivers significantly more information than arrays. We're providing customers participating in the studies with their WGS VCF files so that they can submit them for analysis to third-party data analysis companies. For example, there are services that provide Y chromosome-based data analysis results. Along with the raw sequence data, we also include telomere length data for use in determining cellular age, as well as other relevant analysis results.

WGS data from our pilot studies will be used to further develop our custom personal genome chips for genetic testing in Asian populations. Combined with array imputation and quality control data, it will enable us to improve our custom arrays further.

Q: How is the NovaSeq 6000 System performing in these studies?

GC: The NovaSeq 6000 System is one of the world's foremost sequencing platforms and provides extremely impressive performance, data quality, and stability.

"The ArrayLab Consulting Services team enabled us to start running our lab on a mature, established model in a faster, and more efficient way."

Q: How could NGS positively impact your DTC business?

GC: We believe that as technology progresses and the cost of sequencing decreases, NGS will eventually become the mainstream technology for personal genome testing. That's why we are conducting the necessary research and development to enhance our NGS technical capabilities today. The NGS data from our studies will benefit our Chinese ethnic and East Asian database development, improve our GSA-based offering, and act as the foundation for the next generation of WeGene personal genome testing services. We'll continue to conduct in-depth studies with scientists in various research sectors to tap into the benefits of NGS full genome sequencing data.

Q: How could NGS data benefit the further development of your database?

GC: The goal of our WeGene Chinese ethnic and East Asian genome database development is to drive the application and study of gene testing technology for the benefit of individual consumers, as well as global research organizations. NGS

technology facilitates testing of the entire genome, providing us with detailed genetic information, such as the presence of structural variations, copy number variations, and rare mutations. This data has become valuable for academic researchers studying cancer, autism, and rare disease.

We are convinced that if we make comprehensive, large-scale use of NGS technologies in personal genome products, we can bring the research results from these academic fields to benefit individual consumers quickly. The data on individual consumers will in turn stimulate genome database research.

"We believe that as technology progresses and the cost of sequencing decreases, NGS will eventually become the mainstream testing technology for personal genome testing."

Q: What are WeGene's next steps in expanding its DTC business?

QZ: We want to grow our user base rapidly. Based on our forecasts, we expect to reach one million users in 2019 and over 20 million users in five years. We will continue to promote and educate potential customers about WeGene services in China, East Asia, and globally. We'll continue to pursue research partnerships to enhance the value and utility of our consumer genetic testing services, and Chinese ethnic and East Asian genome databases.

Q: How will Illumina products, systems, and services support these goals?

QZ: Illumina products, systems, and services will be instrumental in lowering costs, improving testing effectiveness, and achieving greater production efficiency with intelligent automation systems to scale our production and analysis services.

Learn more about the products and systems mentioned in this article:

Infinium Global Screening Array (GSA), www.illumina.com/products/by-type/microarray-kits/infinium-global-screening.html

Asian Screening Array (ASA), www.illumina.com/products/by-type/microarray-kits/infinium-asian-screening.html

iScan System, www.illumina.com/systems/array-scanners/iscan.html

NovaSeq 6000 System, www.illumina.com/systems/sequencing-platforms/novaseq.html

Illumina, Inc. • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

© 2018 Illumina, Inc. All rights reserved. All trademarks are the property of Illumina, Inc. or their respective owners. For specific trademark information, see www.illumina.com/company/legal.html. Pub. No. 370-2018-003-A QB 6738

