Population genomics scales up

Next-generation sequencing is the gold standard for studying population genomics, but the cost of library preparation is a limiting factor. Faster library preparation methods are helping to reduce the burden.

Produced by nature custom media

TWIST BIOSCIENCE

Large-scale population studies that combine genomic information with many other sources of data have been made possible by falling sequencing costs and advances in technologies such as next-generation sequencing. *Credit: Tanya Joy / Getty Images*

Philip Awadalla has been studying population genetics for decades, and is still amazed at what scientists have learned from comparing genomes at scale. "There have been massive advancements in our understanding of how genomes evolve," says Awadalla, professor of molecular genetics at the University of Oxford and University of Toronto. "It touches on almost all aspects of biology, from disease applications and healthy ageing to trying to understand where we came from."

Awadalla is also the scientific director of CanPath, a population health study similar to the UK Biobank and the National Institute of Health's All of Us programme. CanPath collects genomic as well as other health and lifestyle data from thousands of volunteers in Canada. The large scale and long duration of these studies also reveals new insights.

"One of my favourite projects in CanPath is that we follow individuals over time," says Awadalla. "They tell us where they live, what they eat, where they work. We integrate all that information with health data and cellular, genomic, epigenomic and proteomic information to see who is developing chronic diseases and who is not."

Studies like this, where genomic information forms a piece of a much larger puzzle, are only possible because the cost of genomic sequencing has been rapidly decreasing, and because new technological advances make it ever easier to collect genetic information at scale.

The library prep problem

As the cost of next-generation sequencing (NGS) has come down, it has become the dominant method of collecting population genomics, supplanting microarray screens.

“When you’re using arrays, you’re largely mapping variation and haplotypes that have already been assayed in subsets of individuals,” says Awadalla. “With sequencing, you get to discover new genetic variation, from different tissues and different populations.”

However, the most popular sequencing methods require the preparation of sequencing libraries in which each genomic DNA sample is fragmented into smaller segments, and each segment is tagged with a short barcode sequence. While sequencing costs have been steadily declining, the expense of preparing sequencing libraries has not. As a result, this process is increasingly driving the cost for studies with hundreds or even thousands of samples — more samples to sequence means more libraries to prepare, after all. But faster and more efficient library prep methods are reducing the burden.

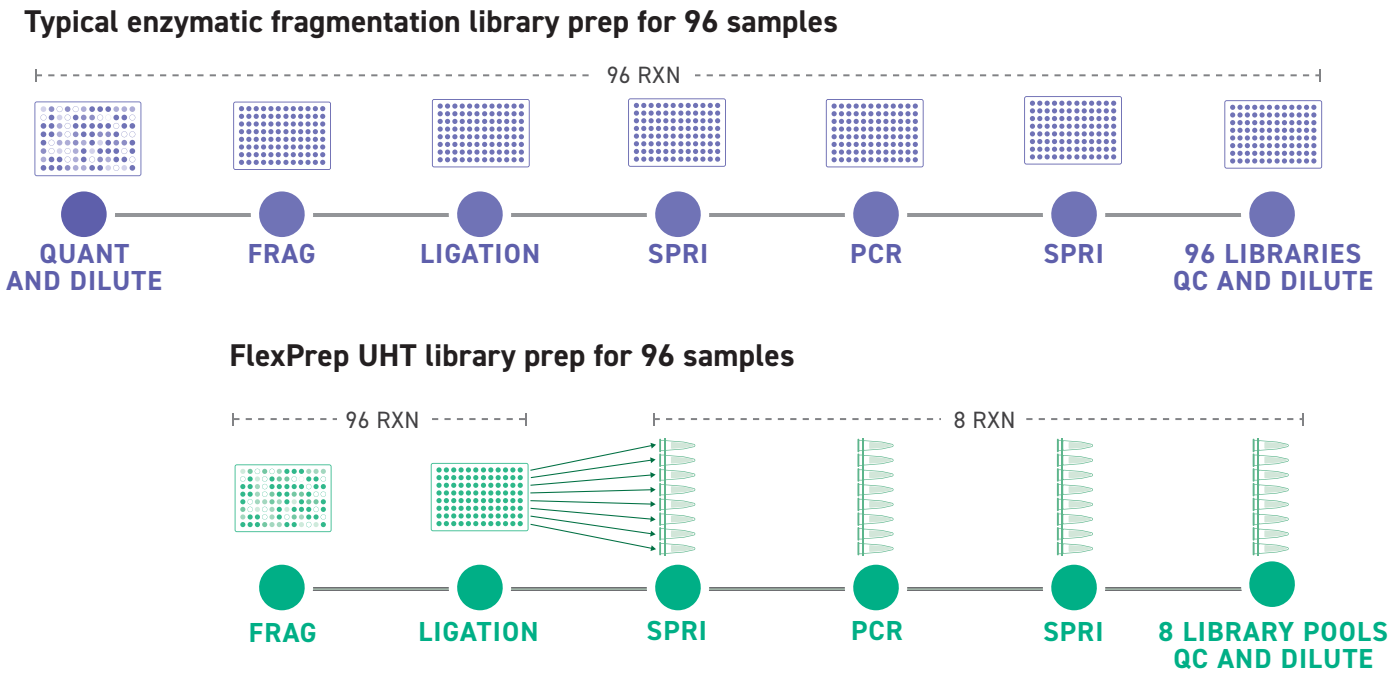
Angela Jones is the next-generation sequencing core manager at Vanderbilt University Medical Center. The core facility she manages is used by researchers at Vanderbilt and by external investigators for a wide range of genomics research. One such project is studying clonal haematopoiesis, a condition where an acquired mutation in a haematopoietic stem cell causes that cell to produce more cells than it should. Vanderbilt researchers developed an assay to detect these types of mutations, called the CHIP Assay[1].

“Last year we did about 150,000 samples through the clonal haematopoiesis assay,” says Jones. The assay uses a targeted NGS panel to discover new mutations in genetic regions known to be affected by this condition. Running so many samples would usually involve a lot of library preparation time and many hours of run time on the facility's devices, but this changed when the core switched to a new library preparation method. “We can do more assays, and we can do them for less.”

A new library preparation method

“Sequencing capacity is growing every year,” says Emily Leproust, CEO of Twist Bioscience, developer of the Twist FlexPrep UHT Library Preparation Kit, which Jones’s facility has been using. “We wanted to help researchers to take better advantage of this capacity, letting them make more libraries at a time in a high-throughput fashion.”

Twist researchers addressed the normalization step in library preparation. This ensures the libraries for each sample all have the same read coverage and depth of sequencing when they are pooled together for sequencing. Without normalization, DNA from some samples would be sequenced more than others, which skews comparisons between samples.



The FlexPrep UHT library preparation kit uses Twist's Normalization by Ligation approach to eliminate the need for upfront and intermediate sample quantitation, streamlining the sequencing workflow. *Credit: Twist Bioscience*

"Manually diluting and portioning out each of those samples is tedious and time- consuming," says Leproust, "so we developed a technology called Normalization by Ligation." This method, which is used in FlexPrep kits, takes advantage of the ligation step of sample preparation, during which adapters are added to each DNA fragment for later identification and analysis. FlexPrep kits use adapters that not only act as barcodes, but also take care of sample normalization. The adapters are added in a limited concentration and interact with fragments in a 1:1 ratio, which can normalize samples in a range of inputs from 30ng to 300ng.

Because this system removes the need for an initial quantification and dilution step, it also reduces the time needed on sample-handling robots. "Tech time is a significant cost for every assay," says Jones, whose facility now uses FlexPrep for several large-scale assays.

Besides the reduced preparation time, the system also speeds up the assay run time thanks to 12 variations of adapters that allow samples to be pooled together. This way, a single 96-well plate can now include 1,152 samples, which also makes it more cost-effective for studies with large sample sizes, like those found in population genomics research. "It allows NGS analysis with a similar cost and timeframe as microarray studies, while revealing more about the genetics," says Leproust.

Awadalla and Jones have both seen how NGS studies have eclipsed microarray analyses in popularity, bringing new opportunities and insights. "What's really exciting about next-generation sequencing is that many of us are now exploring single-cell genomics or single-cell transcriptomics," says Awadalla. This makes it possible to study variation among cells within an individual. "High-throughput, high-coverage can allow you to identify those rare variants, not just among different individuals, but among cells in the sample."

But even when comparing genomes within a population, NGS offers so much more for researchers to explore. "You're not just looking for known variants anymore," adds Jones. "You have the whole genome to look at now. It's a different mindset."

Find out how Twist Bioscience's FlexPrep UHT Library Preparation Kit can support NGS for population genomics and other ultra-high throughput applications.

Reference

[1] Mack, T. et al. J. Mol. Diagn. **26**, 563–573 (2024).