Expert Insights

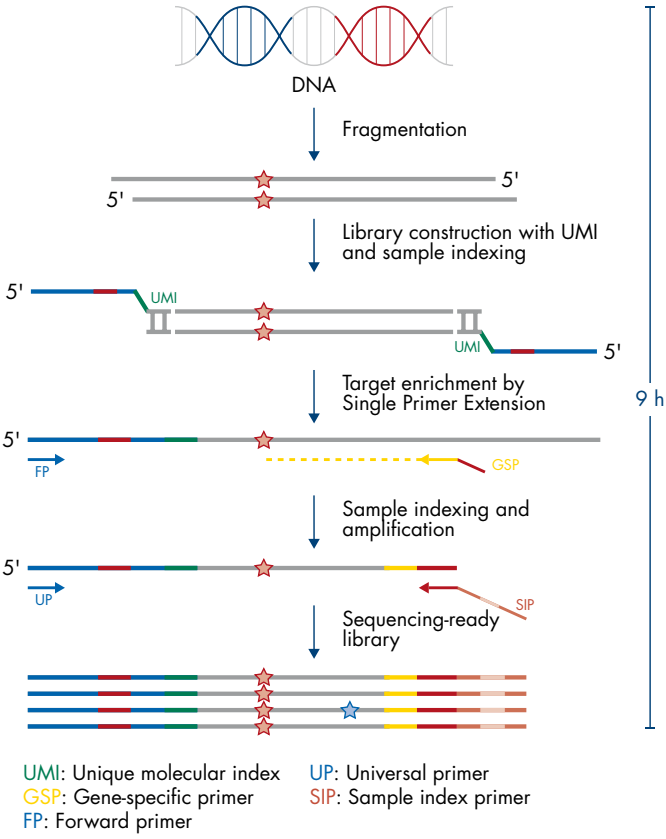
Longitudinal case-control studies using NGS

Pancreatic cancer has one of the worst survival outcomes for any type of cancer, with an 8.5% five-year survival rate according to the National Cancer Institute (1). When pancreatic cancer is detected early – while nodes are pre-metastatic – the five-year survival rate goes up to 25%. Given the high fatality rates for pancreatic cancer, testing for inherited susceptibility may help identify candidates who can participate in screening and prevention programs and potentially detect disease earlier to improve outcomes.

The challenge is knowing where to look.

A team lead by Dr. Fergus Couch at the Mayo Clinic examined the coding regions and consensus splice sites for 21 cancer predisposition genes to determine which ones were associated with the increased risk for pancreatic cancer. This 3030 case and 123,136 control patient study discovered six genes that were independently associated with disease in 5.5% of all pancreatic patients, and 7.9% of patients with a family history of pancreatic cancer. The six genes yielded odds ratios (ORs) between 2.58 to 12.33 (2). Mutations in *CDKN2A* were associated with the highest risk of pancreatic cancer, but were only observed in 0.3% of cases and 0.02% of controls. Additionally, *TP53* (0.2% cases, 0.02% controls), *MLH1* (0.13% cases, 0.02% controls), *BRCA2* (1.9% cases, 0.3% controls), *ATM* (2.3% cases, 0.37% controls) and *BRCA1* (0.6% cases, 0.2% controls) were found to increase patient risk.

Patients were recruited into the Mayo Clinical BioSpecimen resource for Pancreas Research from Mayo Clinic sites in Minnesota, Arizona and Florida. Rather than conducting studies focused on people with a certain type of cancer, and then finding matched controls, Couch's team started with disease-agnostic longitudinal cohort studies. Within those groups, Mayo scientists pulled out huge numbers of participants who eventually developed relevant cancers, and used other participants as the controls. The study's recruitment period was from October 12, 2000 to March 31, 2016.



Dr. Couch and his team used a custom 21-gene QIAseq® Targeted DNA Panel developed and optimized by QIAGEN. Libraries were derived from each DNA sample and then barcoded by dual indices. Sequencing of pools of 768 libraries was performed on a HiSeq® 4000, with 150 bp end reads with a median sequencing read depth of 200x.

The pancreatic cohort is one of many cancer population studies conducted by Dr. Couch's lab that utilize NGS to identify genes associated with different types of cancer. At the peak of his studies, his lab runs 1600 samples per week with only two dedicated technicians. High-quality sequencing data were obtained from 3030 of the 3046 patients.

Figure 1. QIAseq Targeted DNA Panel: Workflow. Isolated DNA, as low as 20 ng, is enzymatically fragmented to generate small pieces of dsDNA. This is followed by the library construction step, in which IL-N7 adapters, unique molecular indices and sample indices are incorporated into DNA fragments generated in the previous step. Library fragments now serve as templates for target enrichment using single primer extension. In this step, targets are enriched using a single gene-specific primer and a universal forward primer. The final step is library amplification and sample indexing (for dual indexing) using the IL-S5 sample index primer and a universal primer. After sequencing, unique molecular indices enable the differentiation of true variant (red asterisk) from false positive (blue asterisk) for sensitive variant detection.

We caught up with Dr. Couch and interviewed him about his study's success and where he wanted to go next.

QIAGEN	*Why did you choose QIAseq as your NGS solution?*
Dr. Couch	Most options were not cost-effective enough to run the volume of samples and fit into our grant's budget. I was referred by a colleague to QIAGEN and performed a quick trial using a catalog QIAseq panel. QIAseq panels are able to get into difficult regions of the genome because they utilize a single primer extension design strategy and both the 3' and 5' fragments have sample indices so up to 1500 samples can be sequenced simultaneously (Figure 2).

While other projects using targeted NGS are considered successful if they achieve 85% coverage, with the custom panel we're using now, we have achieved 99.7% coverage for our target regions. In addition to excellent coverage, the team has been pleased with the uniformity of results. Other options might generate 10,000-fold coverage at one site and just 10-fold coverage at another. That's a problem when you're trying to make sense of so many samples. We knew that we couldn't have these massive outliers in coverage because they would take over the sequencing reaction. The QIAseq panels, on the other hand, deliver a very tight range. The quality of sequence coming out of it is just fantastic. It's been a tremendous success.

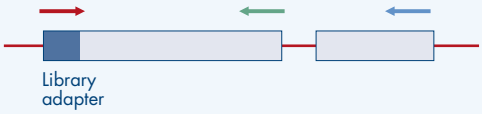


Figure 2. Primer design using single primer extension (SPE). In the single primer extension strategy, only one region-specific primer is used so you are able to design to more regions within the genome (3).

QIAGEN *Did QIAGEN customize any of their chemistry?*
What type of support did QIAGEN provide as you were testing your panel?

Dr. Couch We worked with QIAGEN to design an initial panel. We had a gene list of between 20–26 genes and wanted QIAGEN to create maximal coverage of the exonic regions of the genes. After running the panel at QIAGEN's Frederick site, a few of the primers were changed. The final design had 21 genes with coverage of 99.7%.

QIAGEN spent many hours optimizing the kit configuration so it could seamlessly work with our automation platform and laboratory procedures.

Once we had the data, we needed to analyze it. While we have our own internal analysis pipeline, we worked with the QIAGEN® Bioinformatics analysis team to decode reads with unique molecular indices (UMIs), and to understand library structure to ensure proper analysis of any UMI-containing fragments.

Finally, we needed to have a smooth logistical operation in order to process as many as 6000 samples per month. QIAGEN worked closely with us to develop a manufacturing and shipping schedule that has contributed significantly to the success of our studies.

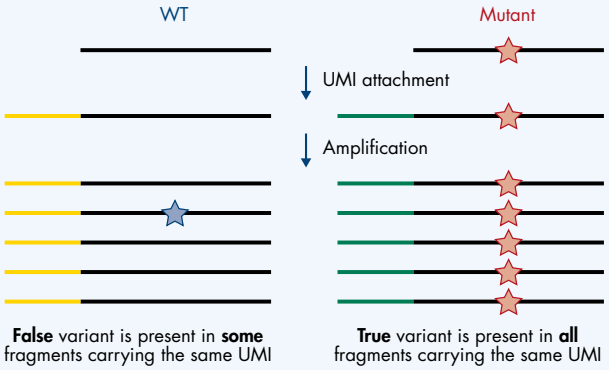


Figure 3. Accuracy in quantification using unique molecular indices (UMIs). A variant identified in a sample represents one of two events: a true or false variant. False variants can be introduced at any step during the workflow, including PCR and sequencing reactions. This results in the inability to accurately and confidently call rare variants (those present at 1% of the sample). Due to PCR duplicates generated in amplification steps, all DNA fragments look exactly the same, and there is no way to tell whether a specific DNA fragment is a unique DNA molecule or a duplicate of a DNA molecule. With UMIs, since each unique DNA molecule is barcoded before any amplification takes place, unique DNA molecules are identified by their unique molecular index, and PCR duplicates carrying the same index are removed, thereby increasing the sensitivity of the panel.

QIAGEN	*What criteria did you take into account when designing your study?*
Dr. Couch	Our objectives were very concrete: We had a defined set of genes and we wanted maximal coverage of the exonic regions within these genes, coupled with high uniformity of enrichment and sequencing. The primary outcome for this study was to determine associations between the candidate genes and pancreatic cancer risk. The secondary outcome was for overall survival. The fact that over 99% of patients had great sequencing data means that the estimates derived from this study are likely a good reflection of risks of this disease for those in the general population with inherited mutations in the predisposition genes.
QIAGEN	*How do you think this research will influence disease surveillance for pancreatic cancer?*
Dr. Couch	It has been known for some time that pancreatic cancer has a germline component. However, this was always associated with family history of pancreatic cancer. My study has shown that even if you don't have a known family history, we can now estimate your risk of developing pancreatic cancer. In addition, we showed that the great majority (83%) of patients with predisposing mutations do not have a family history. We believe this study provides convincing evidence in support of genetic testing of all pancreatic cancer patients, as has recently been proposed by expert panels responsible for formulating medical management guidelines.

We would like to thank Dr. Couch for his insights into his research.

Fergus J. Couch, Ph.D. (left), studies how genetic alterations influence the development of both breast and pancreatic cancer. The long-term goals of his research program are to develop methods that predict an individual's risk of developing breast cancer and facilitate cancer prevention efforts, as well as develop tests that improve selection of treatment for individuals with breast and pancreatic cancer.

Dr. Couch is affiliated with the Mayo Clinic Cancer Center and the Center for Individualized Medicine. His research is supported by the Breast Cancer Research Foundation, the Minnesota Partnership for Biotechnology and Medical Genomics, the National Institutes of Health (NIH) and an NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer.

Raed Samara (right) is a Senior Global Product Manager for NGS solutions and Enterprise Genomic Services. Prior to joining QIAGEN, he was a post-doctoral fellow at the National Cancer Institute conducting research in the field of immuno-oncology, with emphasis on identifying strategies to boost the efficacy of cancer vaccines. He received his Ph.D. degree from Georgetown University in tumor biology. Dr. Samara joined QIAGEN in 2010 as an R&D Scientist and moved to a Global Product Manager role in 2013.

Learn more about QIAseq DNA Panels at **www.qiagen.com/QIAseqpanels**

References

1. SEER Cancer Statistics Review. https://seer.cancer.gov/statfacts/html/pancreas.html Website Accessed August 13, 2018 2. Hu, C. et al. (2018) Association between inherited germline mutations in cancer predisposition genes and risk of pancreatic cancer. JAMA. **319**(23). 2401–2409. 3. QIAseq SPE technology for Illumina: Redefining amplicon sequencing". QIAGEN 2018.

Ordering **www.qiagen.com/shop** | Technical Support **support.qiagen.com** | Website **www.qiagen.com**