Integrated Multiomics Approach Driven by Artificial Intelligence to Realize Whole-Course Management of Cancer

By Mao Mao, Feng Chen, Shiyong Li, and Guolin Zhong, SeekIn

Introduction and Background

We have explored to use one tube of blood to screen all the cancers since 2014, which was the time that a molecular test for colon cancer screening (i.e. Colorguard) was approved in the United States. As having been exploring mutation and methylation based approaches along the way, a large-scale non-invasive prenatal test (NIPT) study for incidental findings of cancer during pregnancy that we initiated in 2016 shed some light on how to solve the problem.[1] The proof-of-concept study started with collecting samples from nearly two million pregnant women of which 639 women were found to be positive for multiple chromosomal aneuploidies on the initial shallow whole genome sequencing (sWGS, 0.1x coverage) data of NIPT. We also applied seven clinically validated plasma tumor markers (PTMs) along with aneuploidy for achieving better sensitivity and specificity for cancer detection.[1] There was a similar study using aneuploidy derived from NIPT data for identifying cancer cases in pregnant women in Belgium.[2] These findings inspired us to use the combination

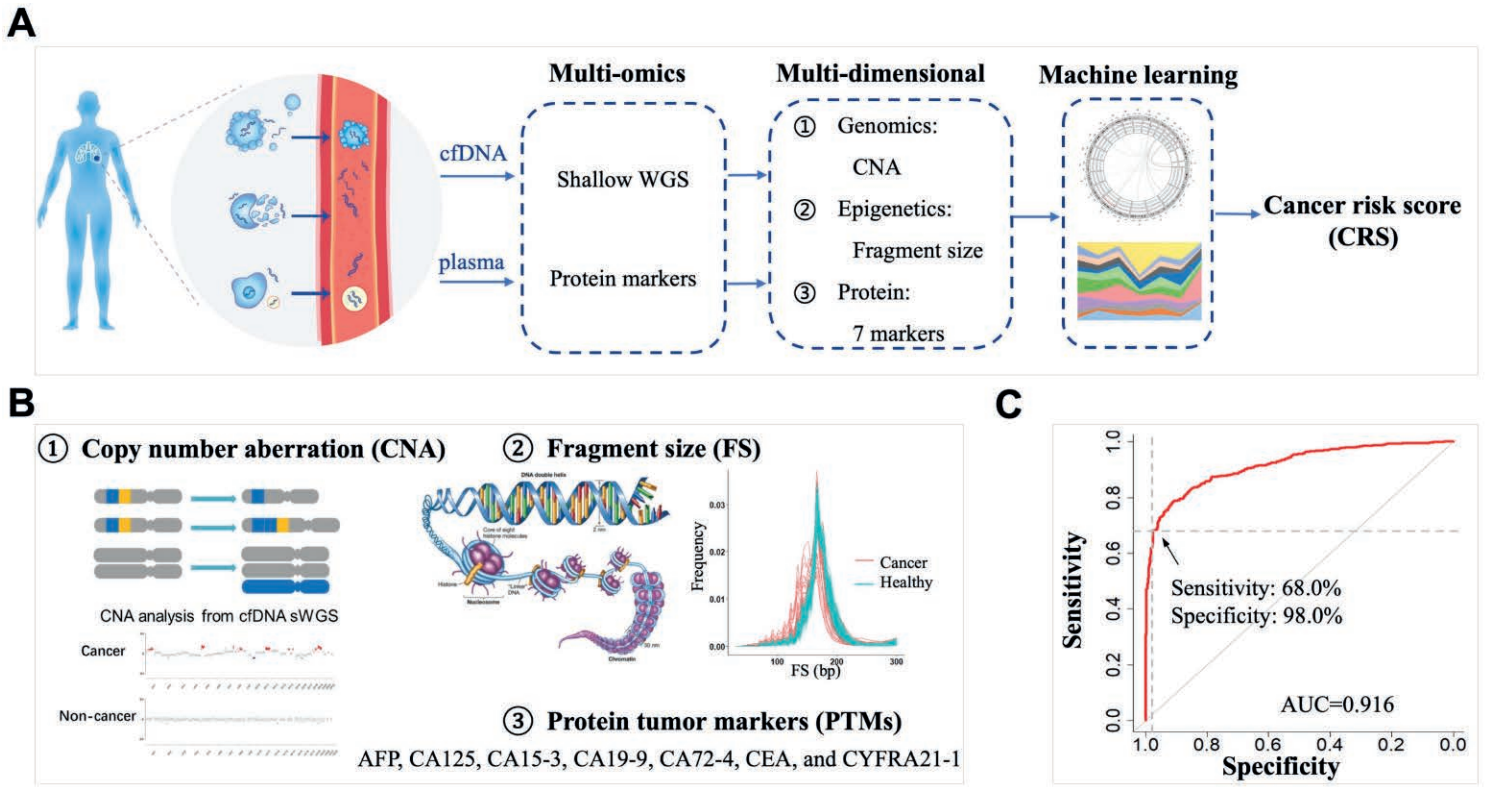


Figure 1: Technical edge of SeekIn tests.

of sWGS and PTMs to search for common cancer features in blood for pan-cancer early detection.

Shown in Figure 1 is the technological base of SeekIn's tests: a novel multi-dimensional cancer risk score (CRS) model that incorporates copy number aberrations (CNAs) and fragment size (FS) via sWGS from cell-free DNA (cfDNA), and seven plasma protein tumor markers (AFP, CA125, CA15-3, CA19-9, CA72-4, CEA, and CYFRA21-1) in a single 8ml blood draw. The copy number aberrations, fragment size pattern, and protein tumor markers (PTMs) represent the common features of cancer at the genomic, the epigenetic and the proteomic levels respectively, and they are integrated by the artificial intelligence (AI). As cancer is caused by different etiologies, multi-omics data (e.g. CNAs, FS pattern, and PTMs) can increase the sensitivity for cancer signal detection from blood and boost the accuracy of cancer early detection.

The core tests for pan-cancer consist of these three assays: SeekInCare, SeekInCure and SeekInClarity. The three tests are using the same core technology but can be implemented in different applications and scenarios. Each test will be explained in the following three sections.

1. SeekInCare

Upon completion of the technical proof of concept and the additional validations, SeekIn launched its first test, SeekInCare. SeekInCare is a first-in-class blood-based pan-cancer early detection test that was launched as a laboratory-developed test (LDT) in China in late 2018. SeekInCare is for individuals with the elevated risks of cancer, e.g., smokers, those suffering from chronic hepatitis, the elderly, etc., and should be used in conjunction with the approved conventional single cancer type screening tests.

As the first-in-class blood-based pan-cancer early detection test, SeekInCare takes a panoramic view of plasma cell-free cancer signatures and a multi-omics approach incorporating genomic and epigenetic alterations in conjunction with clinically validated protein biomarkers. Equipped with proprietary AI- and big data-driven CRS algorithm, SeekInCare exhibits superior effectiveness to detect dozens of cancer types with high specificity.[3]

As shown in **Table 1**, Grail used cfDNA methylation panel for multi-cancer early detection. Specificity for cancer signal detection was 99.5% and overall sensitivity for cancer signal detection was 51.5%. Cancer signals were detected

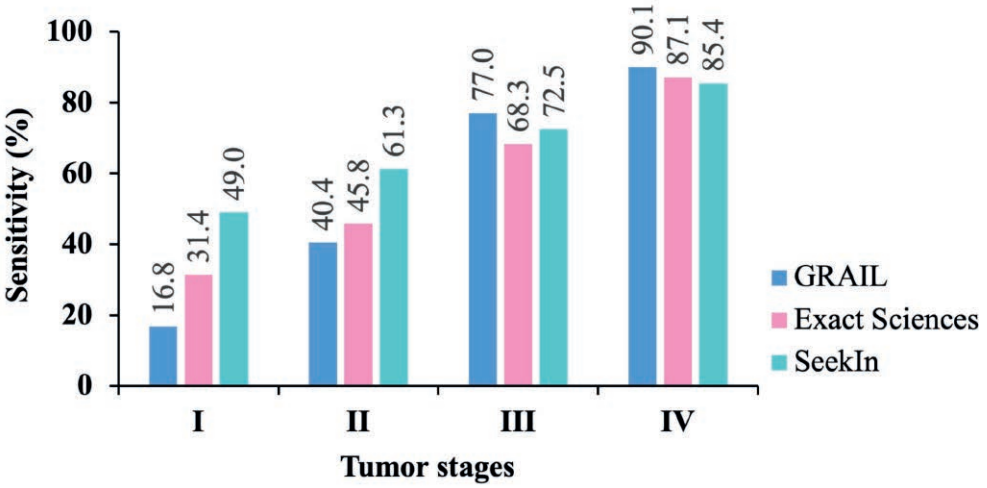


Figure 2: The sensitivity of multi-cancer detection under each tumor stage from case-control validation studies (Table 1) is compared among the three companies.

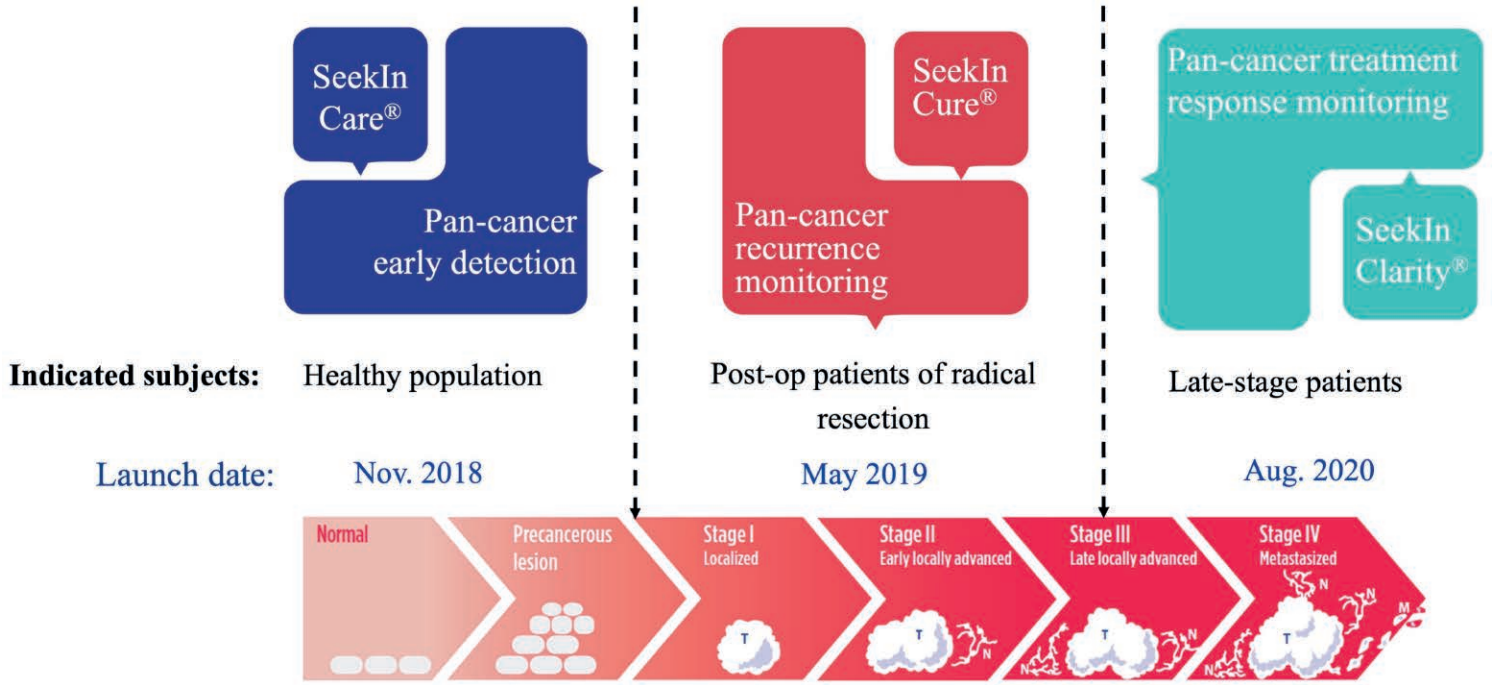


Figure 3: Portfolio of pan-cancer tests.

across >50 cancer types. Overall accuracy of cancer signal origin prediction in true positives was 89%.[4] Exact Science used a combination of four different types of biomarkers – aneuploidy, methylation, protein and mutation. The inclusion of these four biomarker classes in the study resulted in an overall sensitivity of 61.0%, while maintaining the specificity at 98.2%.[5]

The SeekInCare test identified 419 cancer patients with 68.0% sensitivity at 98.0% specificity, resulting in an AUC (area under the curve) of 0.916. And the accuracy of TOO (tissue of origin) prediction is 64% and 78% for TOP1 and TOP2 respectively.[6] Compared with multi-cancer detection in case-control validation cohorts from GRAIL and Exact Sciences, SeekInCare had a similar performance if not better across different tumor stages (**Figure 2**). Furthermore, more data we accumulate to optimize the AI model, more accurate our predictions may become.

As a next-generation sequencing based test that depicts a panoramic view of the whole genome from shallow whole genome sequencing (sWGS, 3x coverage) of cfDNA, this method of molecular testing is more affordable and comprehensive compared to targeted-capturing based tests that rely on deep sequencing at the selected regions in human genome. Moreover, the multi-omics features, CNA, FS, and PTMs, utilized in SeekInCare are not limited to one cancer type and could detect all the cancer types simultaneously.

2. SeekInCure

We provide minimal residual disease (MRD) detection and postoperative recurrence monitoring for cancer patients diagnosed at early stage, SeekInCure (launched in 2019), and treatment response assessment for advanced cancer patients, SeekInClarity (launched in 2020). Our test portfolio for pan cancers is illustrated in **Figure 3**.

As an extension of our cancer detection method, SeekInCure comprehensively utilizes the cancer recurrence index (CRI) to detect minimal residual disease (MRD) and monitor recurrence for the patients with radical surgery, which is used the same multi-omics approach as SeekInCare.

Several studies have evaluated the utility of ctDNA-based MRD detection as a predictive and prognostic test in disease monitoring of several types of cancers, however, the majority of ctDNA-based MRD tests is tumor-informed, which requires tumor tissue analysis to identify individual specific cancer mutations first.[7-9] However, SeekInCure, which incorporates genomic and epigenetic alterations in conjunction with clinically validated protein markers, is a tumor-naive MRD test.

For this test, patients are advised to take a test at 1 month after radical surgery for MRD analysis, and receive this test every 3 months over the following two years, and every 6 months for the next three years, and then take the test every subsequent year thereafter to monitor the potential cancer recurrence (**Figure 4A**).

We conducted a prospective study to assess the potential for broad clinical utility of SeekInCure for the MRD detection and recurrence monitoring in hepatocellular carcinoma (HCC) patients. MRD-positive (MRD+) patients had worse survival than MRD-negative (MRD-) patients. The median overall survival (OS) time of

Table 1: The performance comparison of multi-cancer detection in case-control validation studies among different companies.

Company	Method	Cancer Cases	Normal Cases	Cancer Types	Sensitivity (%)	Specificity (%)	TOO* (%)
GRAIL[4]	cfDNA methylation panel	2823	1254	>50	51.5	99.5	TOP1: 89
Exact Sciences[5]	Mutation panel (plasma + WBC) + Methylation panel + REALSeqS + Proteins	566	566	15	61.0	98.2	No
SeekIn[6]	sWGS + 7 proteins	616	898	27	68.0	98.0	TOP1: 64 TOP2: 78

*TOO: tissue of origin; TOP1: predict top one most probable source of tissue; TOP2: predict top two most probable source of tissue

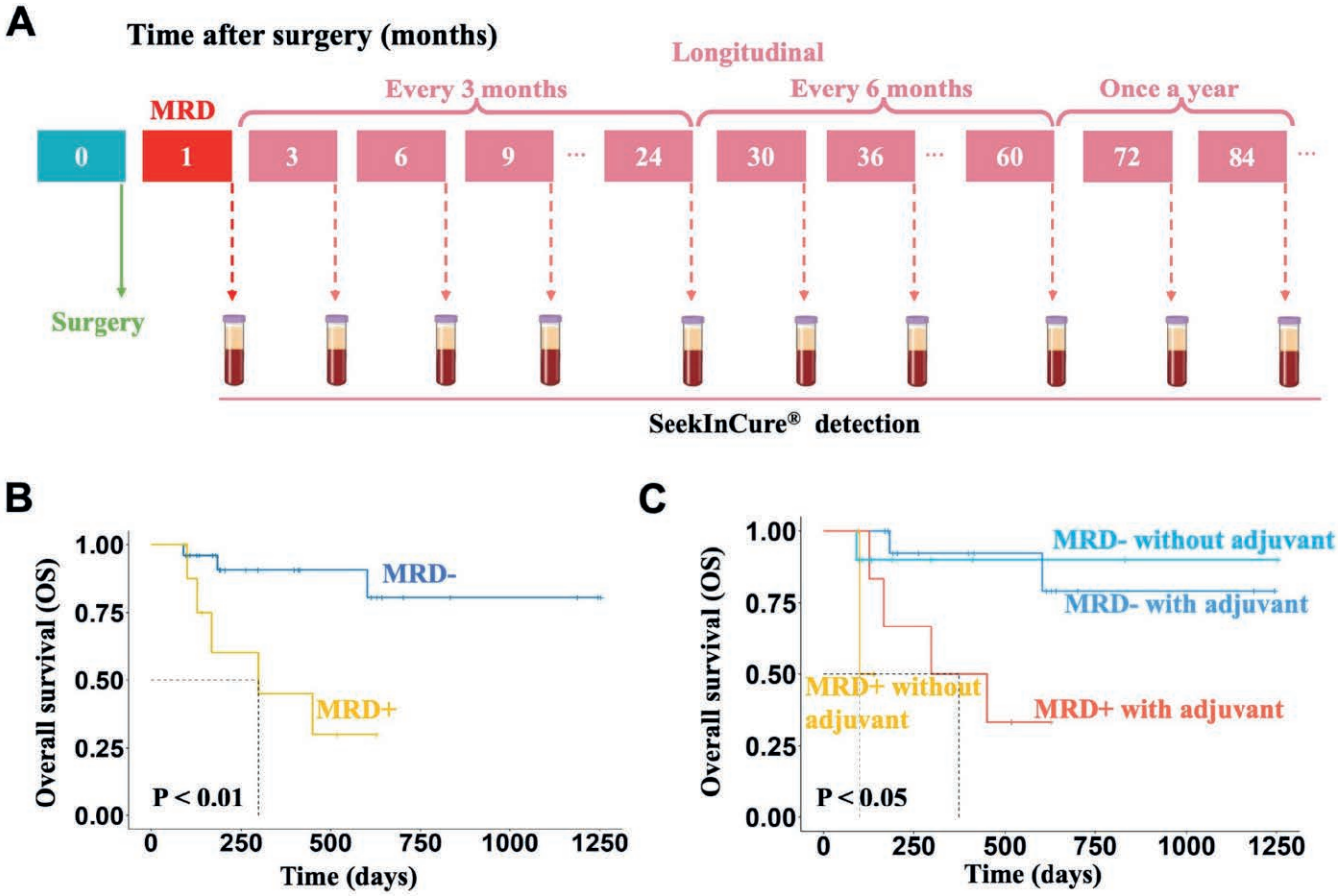


Figure 4: SeekInCure is used to detect minimal residual disease and monitor recurrence in cancer patients with radical surgery.

MRD+ patients was 298 days, while that of MRD- patients was not reached (P<0.01, **Figure 4B**). The MRD- patients with adjuvant therapies had no survival benefit over those without adjuvant therapies, however, MRD+ patients with adjuvant therapies had better OS than those without adjuvant therapies (P<0.05, **Figure 4C**). This indicates that patients with MRD+ are classified as a high-risk population and need further adjuvant treatment to improve their survival, and MRD- patients may not need any additional adjuvant treatment. We also found that patients with pre-/post-operative samples both negative had a very favorable outcome (100% OS), which may represent the potentially cured population with less aggressive cancers, regardless of stage (data not shown).

3. SeekInClarity

Imaging analysis (i.e. Response Evaluation Criteria in Solid Tumors, RECIST) is often used

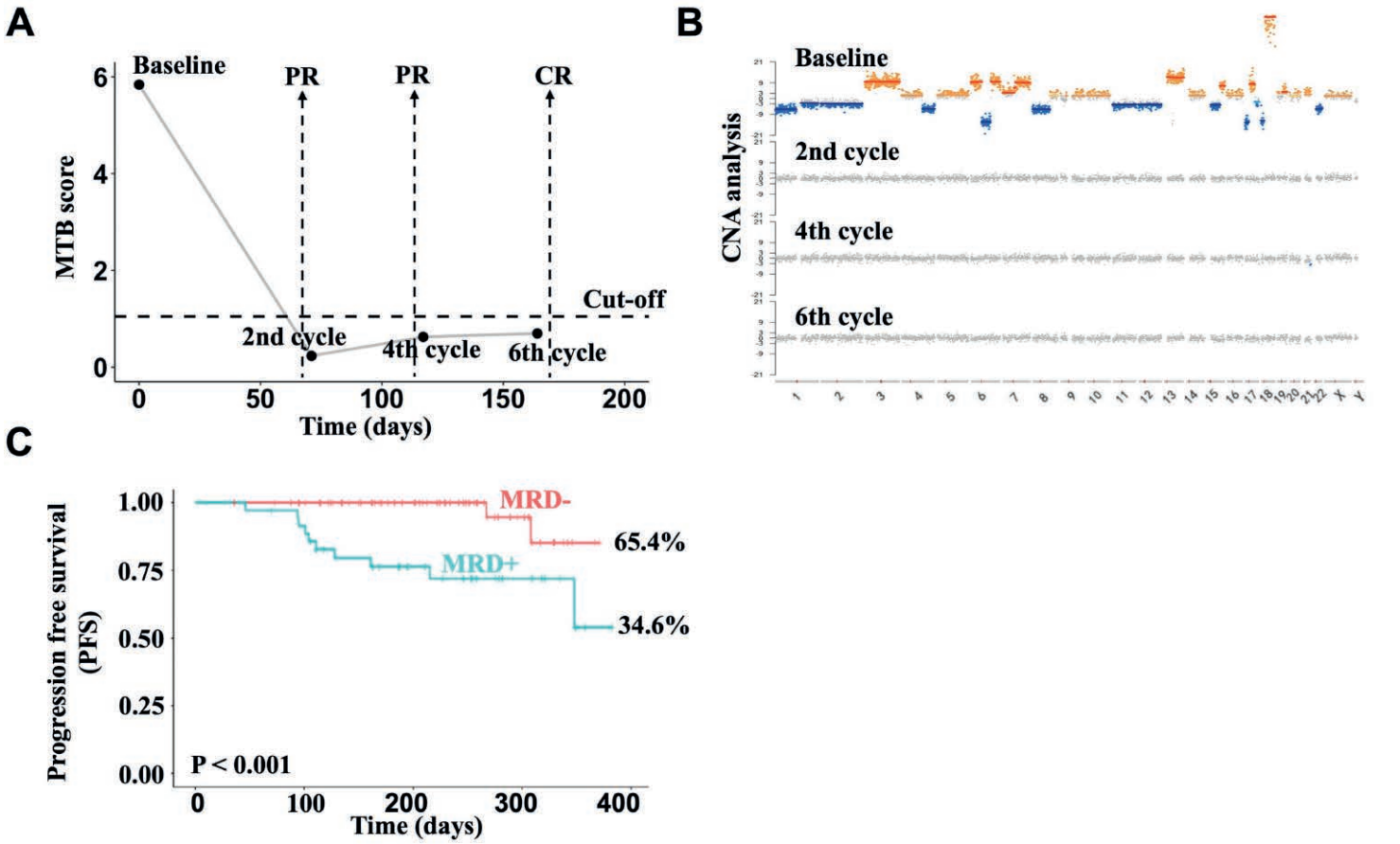


Figure 5: SeekInClarity® for lymphoma patients treatment response assessment

to assess disease burden and to evaluate treatment response for late stage cancer patients. However, it relies on the presence of macroscopic tumor burden. Due to this limitation, patients can be incorrectly labeled as having complete remission even though minimal residual disease (MRD) may still be present. And the subsequent volume change of the tumor can take several cycles of treatments to be seen. In addition, imaging techniques like PET-CT have relatively high cost and expose patients to radiation that can result in downstream negative health consequences. In order to improve upon current clinical ability to monitor cancer treatment response, the development of novel response assessment techniques is needed.

An accurate assessment of treatment response is essential to providing the best care for patients with cancer especially as the response rates of different cancer treatments range from 30% to 70%. Circulating tumor DNA (ctDNA) is a promising non-invasive tumor biomarker that can aid in tumor monitoring throughout cancer care management. SeekInClarity, a blood-based multi-omics assay, integrates the copy number aberrations (CNAs) and the fragment size (FS) profiles of the cell-free DNA and seven protein tumor markers (PTMs) to quantify the molecular tumor burden (MTB) for monitoring cancer treatment response.

An ongoing prospective study that already recruited 114 lymphoma patients across diverse subtypes and treatment regimens was conducted to assess treatment response and predict patient prognosis based on blood via our cost-effective SeekInClarity assay. One patient with baseline MTB+ became MTB- after treatment. Imaging evaluation was PR (partial response), PR and CR (complete response) after 2, 4 and 6 cycles of treatment respectively, MTB score was in concordance with the clinical imaging evaluation (**Figure 5A**). The corresponding changes in CNA profiles are shown in **Figure 5B**. 36 out of 104 patients at landmark test (i.e. after one or two cycles of treatment) were MRD+, who have worse progression free survival (PFS) than the patients with MRD- status (HR: 8.5, 95% CI:2.6-27.4, P-value < 0.001) (**Figure 5C**). Compared with clinical imaging, SeekInClarity can assess the response of treatment in advance, even after only one cycle of treatment. This study presents a strong demonstration that SeekInClarity can be considered as an in vivo drug sensitivity test. Furthermore, high sensitivity and significant lead

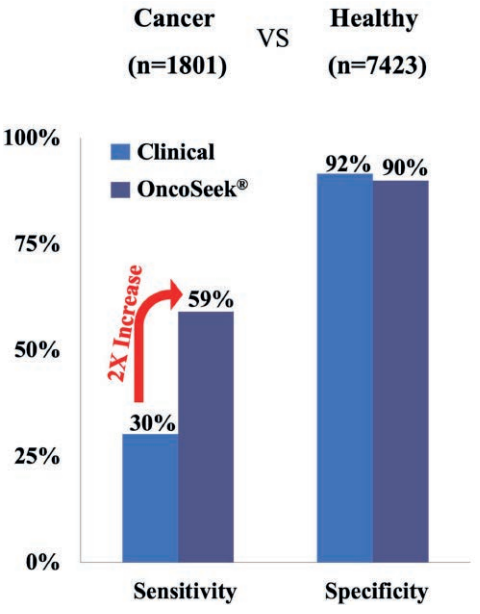


Figure 6: The performance of cancer detection of clinical method and OncoSeek based on the protein tumor markers.

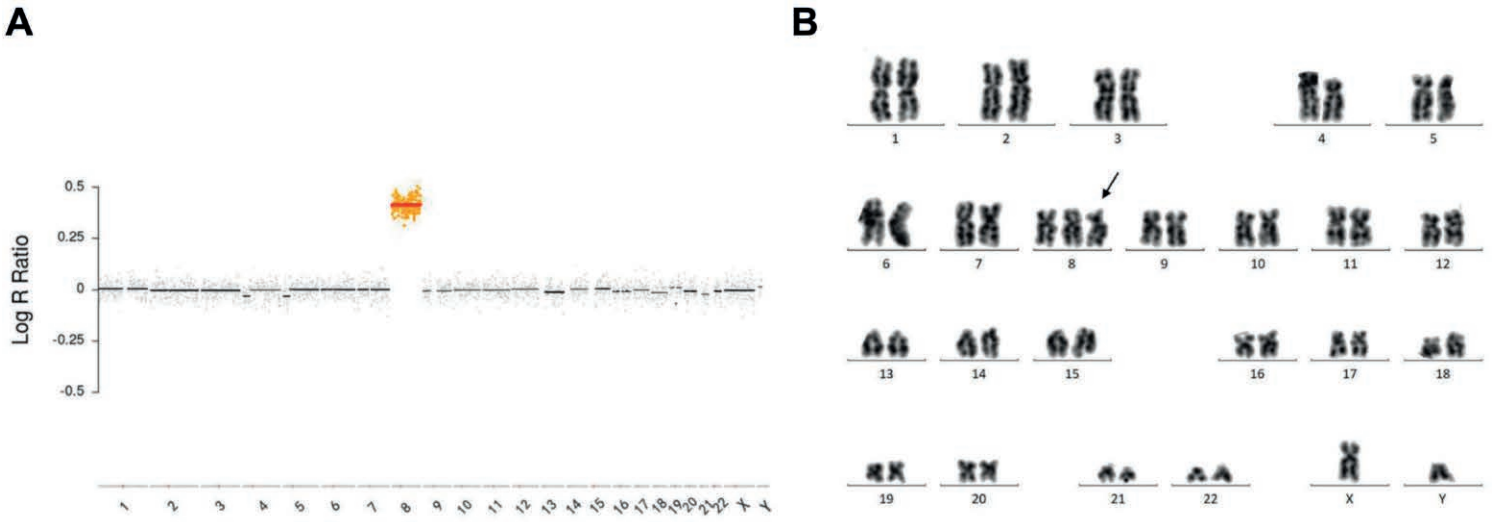


Figure 7: An example of CNA profiling of LeukoPrint reveals abnormal karyotype at molecular level.

time in monitoring treatment response before imaging can elicit a meaningful impact on patient outcomes by informing timely treatment decisions for personalized cancer management.

A path to the broader use of our core technologies

Having developed a set of core technologies, we next wanted to develop tests that would share the same technological principles with our above three pan-cancer tests. We imagined these tests would be simple to perform and could be produced in large volumes based protein tumor markers (PTMs) or shallow whole genome screening (sWGS) for cancer diagnostics. Based on these concepts, the following three tests – OncoSeek, LeukoPrint and PanCanSeek were born.

4. OncoSeek

We designed OncoSeek to be an cost-effective assay that utilized protein tumor markers and clinical information to train and build an artificial intelligence (AI) cancer detection model. The assay quantifies AFP, CA125, CA15-3, CA19-9, CA72-4, CEA and CYFRA21-1 by immunoassay analysis and uses the AI model to distinguish cancer patients from healthy individuals by our probability of cancer (POC) algorithm.

Based on the analysis of nearly 10,000 samples, the sensitivity was determined to be 30% and 59% for the current clinical method and OncoSeek respectively, at around 90% specificity (**Figure 6**). In summary, the sensitivity of OncoSeek for cancer detection is almost two times higher than

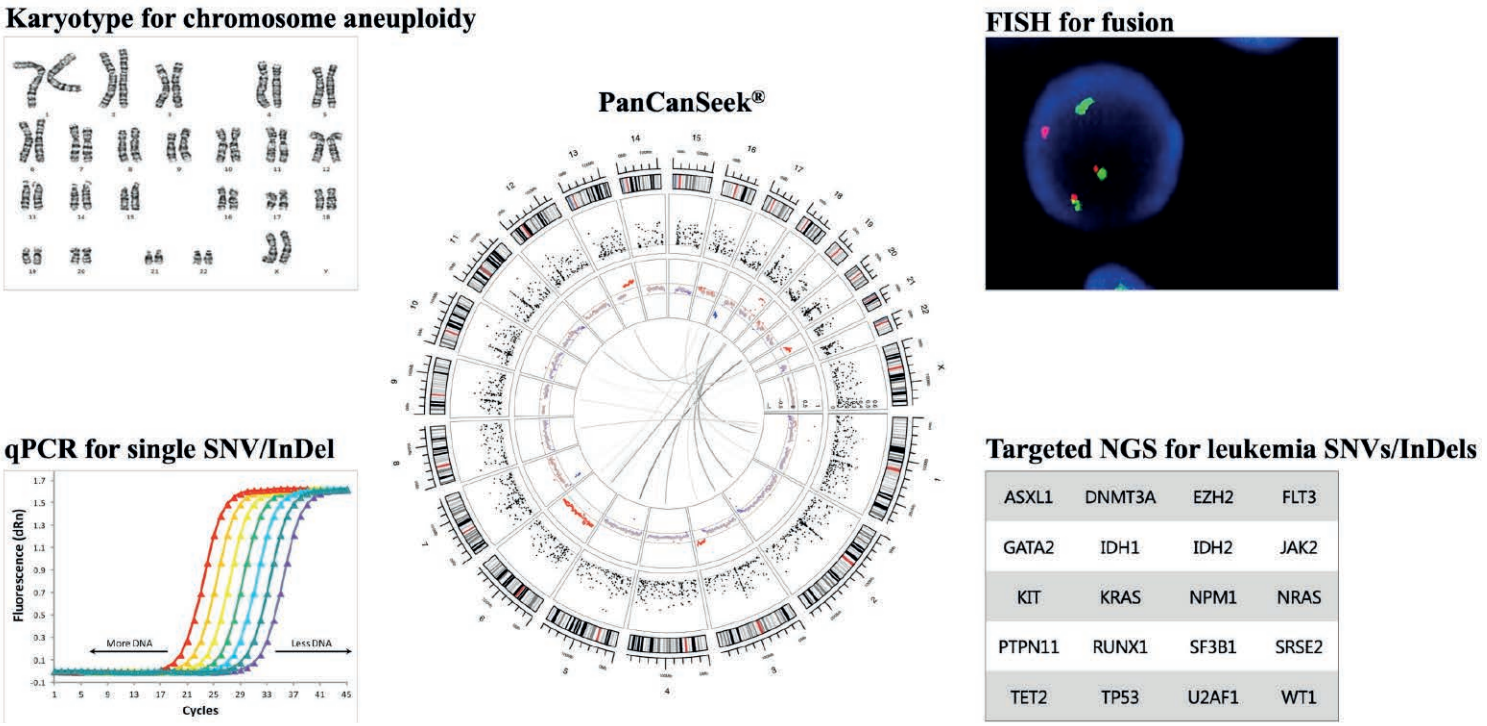


Figure 8: In clinical, karyotyping, FISH, qPCR, and cancer panel were used to detect chromosome aneuploidy, gene fusion, single SNV/InDel and multiple SNVs/InDels respectively.

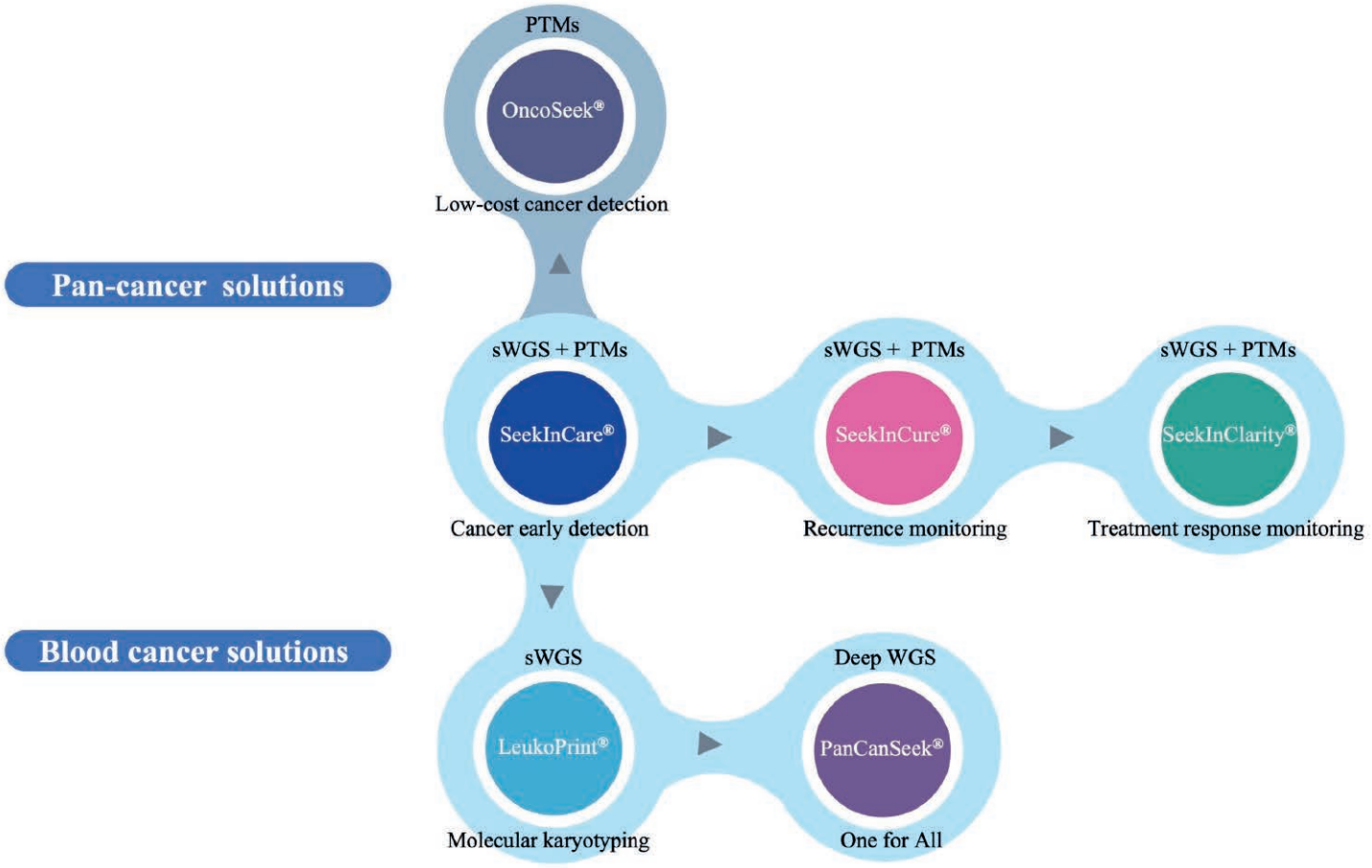


Figure 9: The connectivity of six SeekIn's tests.

the current clinical method. That is to say, twice as many cancer patients can be correctly identified as positive when use OncoSeek comparing to the standard clinical method.

5. LeukoPrint

The LeukoPrint test is designed to depict CNA patterns in patients with hematological malignancies via shallow whole-genome sequencing (sWGS, 1x coverage). LeukoPrint provides a panoramic view of significantly amplified and deleted regions across the entire genome at a resolution of 5Mb via an automated algorithm. **Figure 7A** gives an example of the panoramic landscape of genome of an myelodysplastic syndrome (MDS) patient depicted using LeukoPrint. We can see the clear-cut amplification of the entire chromosome 8 above baseline that was consistent with the result reported by using conventional karyotyping (**Figure 7B**).

LeukoPrint molecular karyotyping test can help hematologists make diagnostic and prognostic stratification according to guidelines laid out by

the European Leukemia Net (ELN), the Revised International Prognostic Scoring System (IPSS-R), and the Revised International Staging System (R-ISS). The significant advantages of LeukoPrint are high accuracy, high sensitivity, and using peripheral blood so that patients can avoid the pain of frequent sampling from bone marrow.

In our previous study,[10] LeukoPrint was utilized to depict genomic CNA profiles from the bone marrows of 137 newly diagnosed acute myelocytic leukemia (AML)/MDS patients. The results are 98.1% concordance of CNA profiles with cytogenetics and/or fluorescence in situ hybridization (FISH). And we further demonstrated that LeukoPrint enabled detection of CNA events occurring in low-proportion populations of leukemic cells from bone marrow, which could be overlooked by cytogenetics.

6. PanCanSeek

PanCanSeek was developed as the world's first integrated solution for leukemia cytogenetics and molecular diagnostics. In **Figure 8**, we show how PanCanSeek covers four different detection techniques – karyotyping for CNV, FISH for gene fusion, qPCR for single mutation and targeted NGS panel for multiple mutations. All the mutations from one leukemia patient were shown in Circos plot from inside to outside circle. PanCanSeek is performed on the next-generation sequencing platform, through high-depth whole-genome sequencing (50x coverage) of leukemia patient bone marrow samples, all the mutation information of the whole genome can be obtained, so that molecular typing, prognosis stratification, treatment guidance, and efficacy prediction can be performed faster, more accurately and more effectively. We positioned this test as a more comprehensive test, in essence, an upgraded version of LeukoPrint in terms of the sequencing depth.

Conclusion

Our story began with developing various novel diagnostic tests on various novel technical platforms, and brought up to the point where SeekIn has developed a cutting-edge novel diagnostics pipeline with our core technologies from which we developed six novel and cost-effective applications for cancer diagnostics (**Figure 9**). We foresee each test has its own clinical utility depending on each patient's need.

We dedicated ourselves to the goal of developing "one size fits all" diagnostics that would be a benefit for all populations in the world. To show our commitment to this goal, we filed for regulatory approval in the European Union and successfully obtained the CE-IVD Mark for all these six tests. In terms of quality, SeekIn tests perform as well as or better than the comparable tests on the market.

SeekIn's mission is to be a leader in cancer early detection, postoperative MRD detection and recurrence monitoring, treatment response evaluation, and novel molecular tests for leukemia patients. Our aspirations are two-fold: first, achieve at least a 15% reduction in cancer mortality rate by offering the means to improve clinical decision-making for early and late-stage cancer patients; and second, advance equitable cancer care through innovation in precision medicine. In reaching these two goals, SeekIn aims to make cutting-edge technologies accessible and affordable worldwide. PMQ

Ms. Guolin Zhong

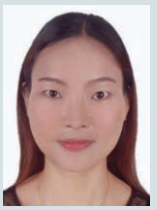
Guolin Zhong has served as Director of Laboratory at SeekIn Since February 2022, previously having the position of Lab Manager since joining SeekIn in 2018, focusing on daily management of the lab, research and development. Ms. Zhong has nearly ten years of work experience in independent third-party clinical testing services and clinical research. She has successfully participated in the growth and development of SeekIn's laboratory from early start-up to full commercial operations. She is also a co-inventor of several SeekIn's technologies and has published several papers. She received a master's degree in Cell Biology from the University of Science and Technology of China.

Mr. Shiyong Li

Shiyong Li has held the position of Senior Director of Bioinformatics at SeekIn since April 2018, as one of 4 initial employees. He has more than 10 years of work experience in bioinformatics analysis. Prior to joining the company, he has served as a senior bioinformatician in the research and development of BGI oncology and has been responsible for many major international cancer projects, including Asian Cancer Research Group (ACRG) projects, Sequencing Quality Control (SEQC) projects. More than 10 scientific research articles have been published, of which 4 are published as the first author or co-author. He also achieved 6 granted inventions patents.

Dr. Mao Mao

Dr. Mao Mao is Founder and CEO of SeekIn Inc. He is also Adjunct Professor of several universities in Asia. He was CSO of BGI Genomics responsible for R&D pipeline and oncology business. He was SVP for in vitro diagnostics (IVD) product development and clinical lab service at WuXi AppTec. He was President of Asian Cancer Research Group (ACRG), a non-profit focusing on the genomic research of the prevalent cancers in Asia. He held several senior positions at Pfizer and Merck in the past. He also practiced as obstetrician/gynecologist and genetic consultant. As one of the founding members of the National Human Genome Center in Shanghai, he established the first high-throughput DNA sequencing facility and pioneered genome research in China. He published 100+ articles in the peer-reviewed journals including Nature, Cell, and Nature Genetics. He is a strong advocate of advancing equitable cancer care through innovation.



Ms. Feng Chen

Feng Chen is the project manager at SeekIn. She used to work at BGI, with more than 7 years of marketing experience and overseas work experience. She has successfully managed hundreds of technology service projects, mainly serving many key universities (Stanford, UC Berkeley, Tulane University, etc.) and research institutions (NIH, CSMC, Scripps Research, etc.) in the United States and Canada and the world's top 50 pharmaceutical companies. She was also the First Person in the laboratory to successfully construct the Aspergillus oryzae transformation system during her graduate studies.

References

1. Xing Ji, Jia Li, Yonghua, et al. Identifying occult maternal malignancies from 1.93 million pregnant women undergoing noninvasive prenatal screening tests. Genetics in Medicine. 2019;21(10):2293-2302.
2. Lenaerts L, Brison N, Maggen C, et al. Comprehensive genome-wide analysis of routine non-invasive test data allows cancer prediction: A single-center retrospective analysis of over 85,000 pregnancies. EClinicalMedicine. 2021;35:100856.
3. Meng Z, Ren Q, Zhong G, et al. Noninvasive Detection of Hepatocellular Carcinoma with Circulating Tumor DNA Features and α-Fetoprotein. The Journal of Molecular Diagnostics. 2021;23(9):1174-1184.
4. E. A. Klein, D. Richards, A. Cohn, et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. Annals of Oncology. 2021;32(9):1167-1177.
5. C. Douville, C. Nobles, H.J. Hwang, et al. Multi-cancer early detection through evaluation of aneuploidy, methylation, mutation, and protein biomarkers in plasma. Annals of Oncology. 2022;33(S7):S575.
6. Mao M., Shiyong Li, Qingqi Ren, et al. Integrating multi-omics features for blood-based pan cancer early detection. Poster at The 2022 Early Detection of Cancer Conference in Portland.
7. Zhang JT, Liu SY, Gao W, et al. Longitudinal Undetectable Molecular Residual Disease Defines Potentially Cured Population in Localized Non-Small Cell Lung Cancer. Cancer Discovery. 2022;12(7):1690-1701.
8. Bruna Pellini, Aadel A. Chaudhuri, et al. Circulating Tumor DNA Minimal Residual Disease Detection of Non–Small-Cell Lung Cancer Treated With Curative Intent. Journal of Clinical Oncology. 2022;40(6):567-575.
9. Aparna R. Parikh, Emily E. Van Seventer, Giulia Siravegna. Minimal Residual Disease Detection using a Plasma-only Circulating Tumor DNA Assay in Patients with Colorectal Cancer. Clinical Cancer Research. 2021;27 (20):5586-5594.
10. Xiaodong Lyu, Tao Li, Dandan Zhu, et al. Whole-genome sequencing as an alternative to analyze copy number abnormalities in acute myeloid leukemia and myelodysplastic syndrome. Leukemia & Lymphoma. 2022; 63(10):2301-2310.