1. **Significance**

This project was designed to address the challenges in Area B (New capabilities advancing precise clinical diagnosis of cancer patients). We will focus on the non-invasive detection of early stage pancreatic cancer, an area where early detection will have clear clinical benefits to the patients. The novel methylation haplotyping approach has many technical advantages (detailed in the proposal below) over existing molecular diagnostic methods in terms of sensitivity and the robustness in the presence of biological variabilities among individual patients. The early proof-of-concept of this strategy has been demonstrated on colon cancer and lung cancer, and promising preliminary data have been generated from pancreatic cancer patients. It has the potential to transform the early detection and intervention of pancreatic cancer.

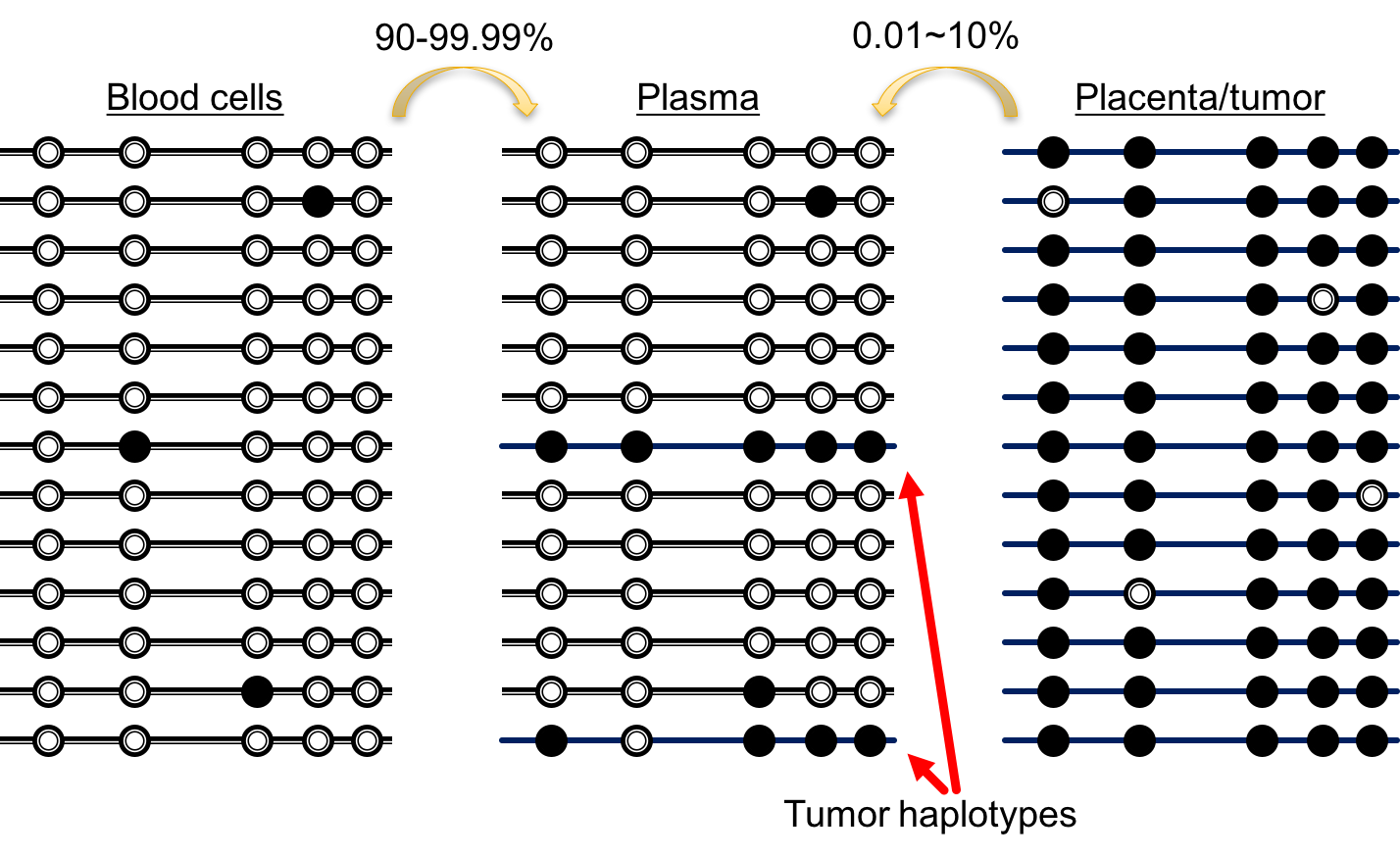
While pancreatic cancer is ranked fourth in cancer mortality as of 2016, it is projected to become the second leading cause of cancer death in the United States by 2030 with a projected 63,000 deaths per year(Rahib, Smith et al. 2014). Advances have been made to reduce mortality of the other cancers but not for pancreatic cancer. Pancreatic cancer tends to initially present to the clinician in an advanced, incurable state where the only treatment options available are systemic chemotherapy or combination therapy(Neoptolemos, Palmer et al. 2017). Due to its aggressiveness and early metastasis, pancreatic cancer is found resistant to most treatment options [5]. Less than 20% of patients who were fortunate to have early resectable tumors could get potentially curative surgeries and much greater chance of 5-year survival up to 40% (stage IA) rather than 5% (overall). Studies attempting to understand the risks of this disease have determined that while majority of cases are sporadic, there are increased risks associated with increased age, smoking, obesity, hepatitis B infection, and diabetes mellitus, and chronic pancreatitis [5]. Regular screening has been recommended for patients with a family history, which accounts for 10% of cases, and patients with chronic pancreatitis, which accounts for up to a third of cases [5]. Whether pancreatic cancer screening in the at risk populations would be beneficial, similar to how screening has been shown life-saving for colorectal and cervical cancers, should be investigated. However, increasing the number of patients with resectable tumors could significantly improve long term survival and treatment options.

To date, there is still no good screening test for pancreatic cancer. Imaging with endoscopic ultrasound (EUS), computed tomography (CT) and/or magnetic resonance (MRI) are commonly used for diagnostic evaluations but while they can detect majority of lesions, the information they provide about each lesion are limited to spatial location and size and not whether the lesion is malignant or not. A blinded multi-center study with 225 asymptomatic high-risk individuals discovered pancreatic lesions in 42% using EUS, 33% using MRI, and 11% using CT, highlighting the prevalence of lesions in high risk individuals(Canto, Hruban et al. 2012). Recent utilization of a positron emission tomography (PET-CT) technique could help to distinguish metastatic disease but it is still too costly and could have high rates of false positives(Jana, Shroff et al. 2015). Since early pancreatic neoplasms may be malignant but not invasive or large enough to cause clinically significant symptoms, an area of immediate help would be to improve the resectability at diagnosis by developing a screening test that have the ability to detect malignant lesions in the localized stage.

There are a limited number of promising molecular biomarkers currently being used or under development for a liquid biopsy test utilizing either blood plasma, urine, pancreatic juice, or bile. A well known biomarker is the blood serum level of CA19-9, a cancer antigen found to have elevated levels in pancreatic cancer patients and is used to monitor disease progress. However, CA19-9 is not a good marker for early detection because it is not specific to pancreatic cancer as other pancreaticobillary conditions may also cause elevated CA19-9, such as in individuals with chronic pancreatitis or cholangitis. In recent years, developments have been made to profile the molecular signatures of pancreatic cancer which led to much better understanding of the potentials for molecular screening. For example, an activating mutation within the *KRAS* gene is found in over 90% of pancreatic tumors. However, a following work have found that only 43% of patients with localized tumors have detectable KRAS mutations in circulating DNA [REF]. Panels for miRNA detection in plasma, bile, or pancreatic juice with 95-100% sensitivity have been identified for discriminating pancreatic cancer from healthy or individuals with chronic pancreatitis, pancreatic cancer from intraductal papillary mucinous neoplasm, or an miRNA signature for better survival in patients with advanced stage [7]. However, these panels have yet to be validated and used in the clinic due to lack of consistency among different studies and lack of a standardized measurement platform for miRNA biomarkers [7]. Proteomics have also been applied to identify a novel panel of 3 proteins (LYVE-1, REG1A, and TFF1) which can discriminate early stage pancreatic cancer patients from healthy individuals with 75.5% sensitivity and 100% specificity(Radon, Massat et al. 2015). Cost, ease of implementation, consistency and accuracy are areas which much be met in developing a screening assay utilizing biomarkers. Therefore, this is an area where a robust, sensitive and specific method for non-invasive diagnosis can lead to major impacts on the clinical intervention and the improvements of clinical outcomes.

[might need a few lines here to close this section in a high note.]

1. **Innovations**



**Figure xx. Methylation haplotypes allow for robust detection of tumor signature in the presence of technical and biological noise.**

This project has several technical innovations on biomarker discovery, methodology for detecting cancer, and the implementation of clinical assay.

First, we will focus on the DNA methylation of circulating DNA molecules. DNA is more stable than metabolites, proteins, or RNA/miRNA. DNA methylation capture the epigenetic state of cells, which is in a layer below DNA mutations in terms of regulatory cascade. While driving mutations can be different from one patient to another or even heterogeneous within the same tumor, they result in consistent methylation signatures related to the hallmarks of cancer. Therefore, a smaller number of methylation-based markers has the potential to cover a spectrum of biological variabilities among cancer patients. In this project, we will start with exhaustive search of informative markers across the entire human genome, using >100 sets of whole genome bisulfite sequencing data. This will ensure the completeness of marker identification. In comparison, some previous efforts typically started with very limited sets of candidate genes or array-based methylation data that cover only 27k or 450k CpG sites among the ~26 million CpG sites in the entire human genome.

Second, we will use a unique approach based on DNA methylation haplotypes, or the co-methylation patterns of multiple adjacent CpG sites along single DNA molecules. The concept of DNA methylation haplotype was initially conceived by us for investigating the pattern of allele-specific DNA methylation(Shoemaker, Deng et al. 2010), and was recently formalized and extended to the study of heterogeneous tissues and non-invasive localization and detection of cancer in plasma(Guo, Diep et al. 2017). Methylation haplotypes are more robust than single CpG sites in the presence of biological and technical noise. They also provide superior sensitivity when analyzing cell-free DNA, in which only very small fractions of DNA molecules came from cancer cells.

Third, we will take advantage of a new observation we made to improve the detection sensitivity and specificity. In matching ctDNA methylation signatures against a database representing 10 human normal tissues and “pan-cancer”, we found that cancer patients’ ctDNA tend to have a higher level of DNA molecules from the normal cells in the body sites where tumor grows. Integrating this information with the methylation signature of cancer-derived DNA molecules in the plasma, we achieved a higher detection sensitivity and specificity than using cancer signature alone. This has been demonstrated on colon cancer and lung cancer. In this project, we will apply this novel concept to the early detection of pancreatic cancer.

Finally, after biomarker discovery, we will develop a clinical assay using a novel targeted bisulfite sequencing strategy. We will benchmark a novel method, called methyl-AmpliSeq, which was developed by Thermo Fisher with our assistance, with another method, BSPP, that we pioneered, optimized and benchmarked in the past few years. After validation with ~300 clinical samples, we will deliver a highly sensitive and robust assay that can be readily deployed in the clinical community.

Methylation haplotype increase the signal-to-noise ratio and the robustness in detecting low level of signals.

Integration of tissue-of-origin and tumor signals improves the detection sensitivity & specificity.

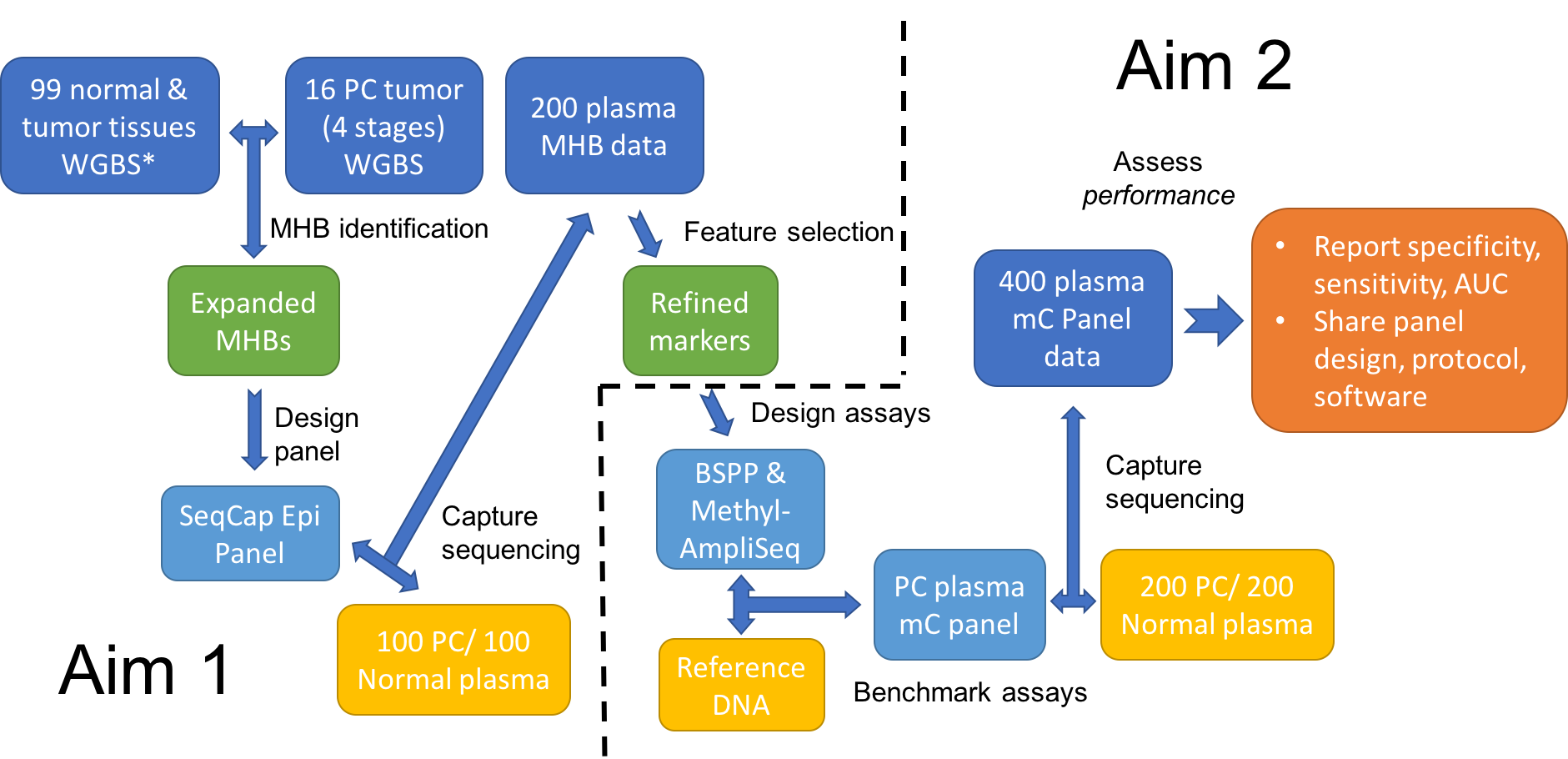
Detection of early stage pancreatic cancer, early detection of metastasized tumors.

Two layers of integration to overcome variability across different patients.

1. **Approach**

**C0. Overall Strategy**

Areas to improve: more tumor tissue samples; breakdown into tumor stages (benign versus stage I versus II or later, metastatized) ; improving library conversion efficiency; more thoughtful selection of targets for the initial screen; a new targeted bisulfite sequencing approach



**Figure 1. Overview of the project design.**

**C1. Preliminary Results** [1-2 page]



**Figure xx. Preliminary results of tissue-of-origin mapping on the plasma DNA from 10 pancreatic cancer patients.**

Under a sub-heading *"Preliminary Data",* address the following aspects:

* Describe the current state of development of the technology/tool/capability proposed. For application to be responsive, this description should offer a clear, rigorous evidence for the feasibility of the proposed approach and attaining the stage suitable for advanced development aimed at implementation.
* Preliminary studies on which the application is predicated should amount to a successfully completed phase I development (e.g., in terms of demonstrating the general utility of the concept, building a prototype, testing technology/tool in "real life samples, and/or other benchmarks as appropriate).
* For projects proposing a novel combination/integration of several approaches (e.g., experimental approaches in conjunction with integrative data analyses), provide appropriate preliminary data for the components as well as the entire combined approach.

**C2. Specific Aims**:

**Aim 1.** Genome-wide screening and characterization of haplotype-based methylation markers for pancreatic cancer in plasma.

Contrast methylation-based detection with mutation-based detection

There have been promising efforts in distinguishing pancreatic cancer from chronic pancreatitis based on methylation-based detection in plasma(Liggett, Melnikov et al. 2010, Henriksen, Madsen et al. 2016), although the specificity and sensitivity is still not sufficient for clinical application. The markers were selected from a limited set of promoters (56 by Liggett et al.; 28 by Henriksen et al). Single CpG methylation measurements were determined by microarray or methylation-sensitive PCR. Our marker selection will start with the data from the entire human genome, we will use multiplexed target bisulfite sequencing and quantify with MHL calculating from multi-CpG haplotpyes.

Clearly explain the rationale for discovering “pan-cancer” MHBs.

Method: use NimbleGen methyl-SeqCap to maximize the recovery of fragments, target selection: intersection of blood UMRs with MHBs (recompiled with larger updated WGBS data sets); plus regions that are unique unmethylated in germ cells. [Need to estimate the target size, based on Version 2 MHBs from Dinh, plus a full list of data sets used]

Description of samples for initial testing and validation. Provide one-paragraph description of MCC biorepository. Also mention AllCells for normal plasma.

The Biorepository and Tissue Technology resource contains plasma, serum, RNA stabilized buffy coat, urine, viable tumor samples, and formalin-fixed paraffin-embedded tumor — all with associated clinical information. (https://healthsciences.ucsd.edu/research/moores/shared-resources/biorepository/Pages/default.aspx)

Table 1. The first batch of pancreatic cancer patient samples retrieved from UCSD MCC.

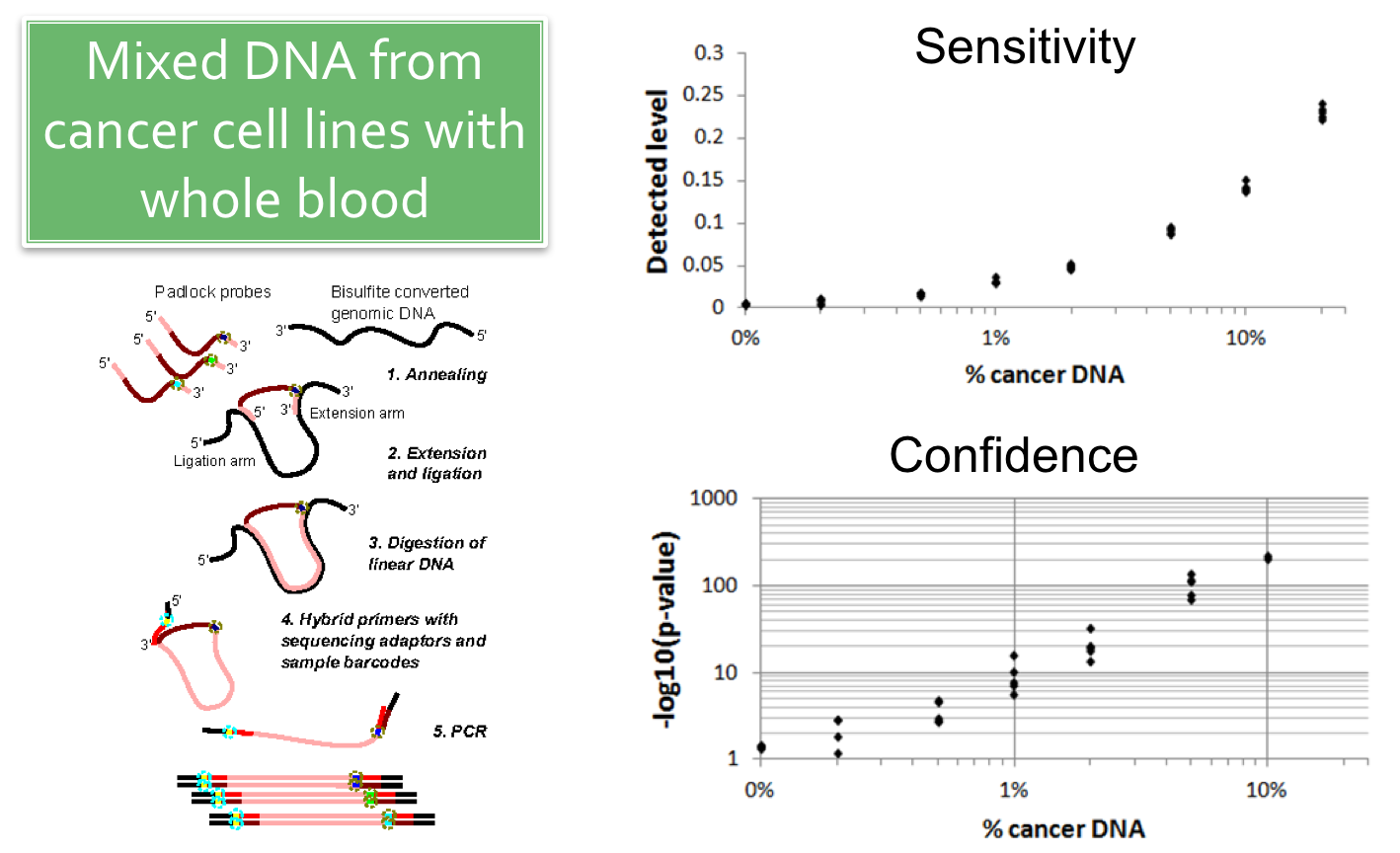
|  |  |  |  |
| --- | --- | --- | --- |
|  | Plasma + Tissue | Plasma only | Total |
| Benign | 0 | 10 | 10 |
| Stage I | 7 | 3 | 10 |
| Stage II | 30 | 31 | 61 |
| Stage III/IV | 2 | 9 | 11 |

Multi-tissue mapping.

Normal plasma: available from AllCells, ~$100 per sample (10mL); Sample preparation/preservation, extraction, QC (Nanodrop, Agilent Bioanalyzer analysis of fragment size distribution)

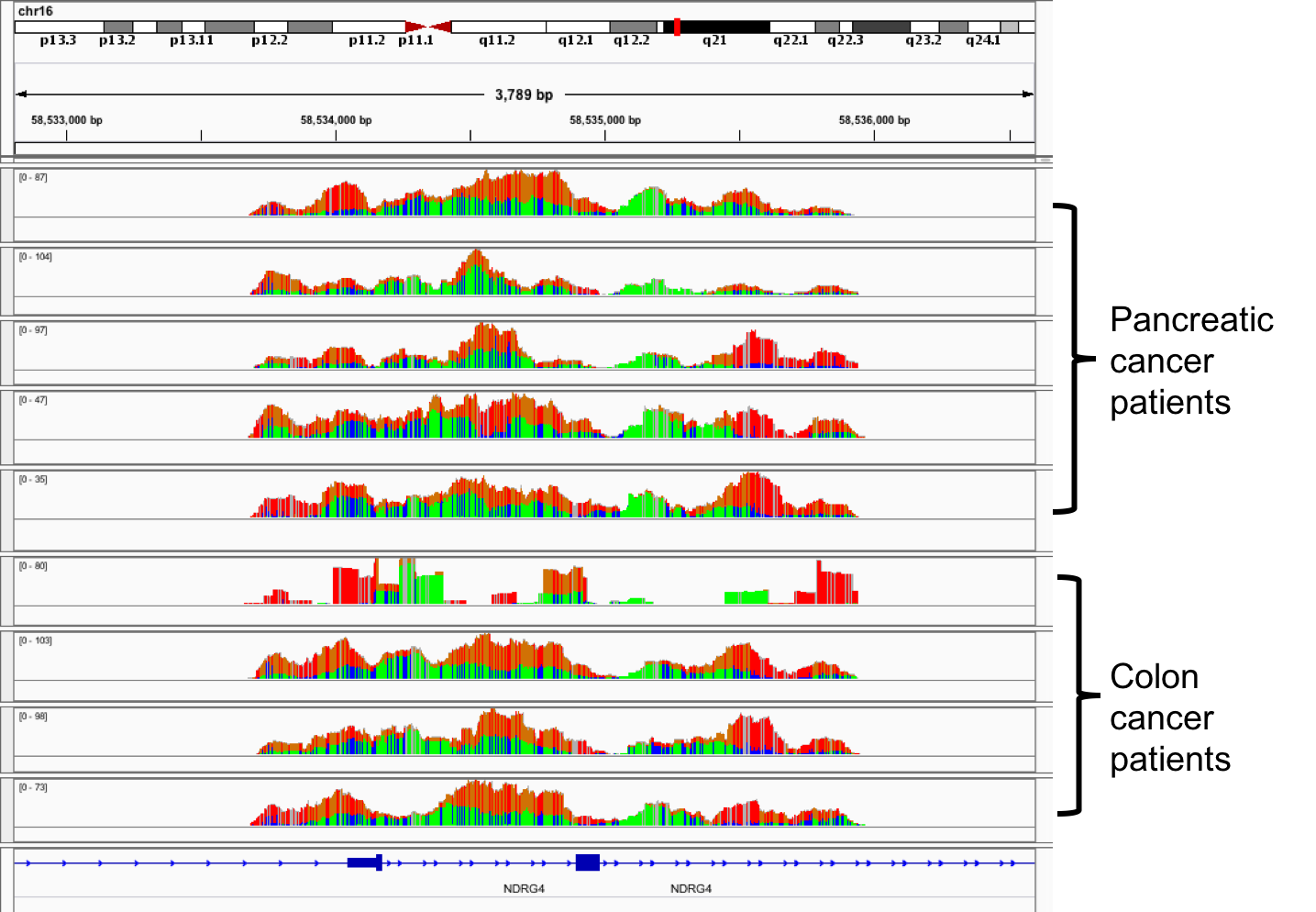
Most previous studies(Imamura, Komatsu et al. 2016) were based on limited sample sizes (<100), we will include samples from 300 patients and 300 normal controls.

Our current MHBs (Version 1.0) were derived from 65 set of WGBS data, including six tumors (one cell line), 10 stem cell samples and 49 human adult tissue samples. To compile an expanded set of MHBs (Version 2.0), we have downloaded and processed 34 sets of recently published WGBS data (12 tumor, 22 normal tissues), for a total of 99 reference data set. Because the 99 samples contain normal pancreas, but have representation of pancreatic cancer, we will generate additional 16 sets of WGBS data from fresh frozen pancreatic cancer tissues (4 samples for each of the four stages I, II, III, IV, all have been identified from UCSD MCC biorepository). This will more than double the number of methylation haplotypes sampled across a wide variety of human normal tissues and cancers, and allow us to define more informative MHBs for biomarker identification.



**Figure xx. Sensitive detection of minor fraction with a combination of BSPP and haplotype-based quantification.** To determine the detection sensitivity, we used a titration series of synthetic DNA mixtures using gDNA from each of the five cancer cell lines (BXPC3, T98G, PANC1, BE2C, U87MG) and human whole blood.

For screening and initial validation of pancreatic cancer specific markers in cell-free DNA, we will use a different experimental strategy, Nimblegen SeqCap Epi, than the scRRBS method used in our recent published study. While scRRBS was initially developed for single-cell methylation sequencing, and has allowed us to generate very informative data, it has a number of limitations. First, ctDNA extracted from human plasma always contain various levels of high molecular weight (HMW) genomic DNA from lysed white blood cells. Such DNA can dilute the signals from cell-free tumor DNA. The scRRBS protocol has equal or even higher efficiency in converting blood cell derived HMW DNA into the resulting sequencing libraries and increases the difficulty in detecting ctDNA. Second, scRRBS targets a specific subset of the human genome that enriched for MspI cutting sites (CCGG), which tend to overlap with CpG islands, it only covers only 44% of the Version 1.0 MHBs, leaving many potentially informative MHBs untapped. For these two considerations, we will switch to a hybridization capture strategy. We will design a customized Nimblegen SeqCap Epi probe set to achieve a full coverage of the Version 2.0 MHBs that we will identified. The hybridization capture protocol also has an additional advantage that the white blood cell derived DNA can be removed by size selection, because they tend to be much longer before and after ligation with sequencing adaptors. This will experimentally improve the signal-to-noise ratio and provide better data for feature selection and classification.



**Figure xx. Targeted bisulfite sequencing of cancer patient plasma DNA with hybridization-based capture.** The genome coverage of a NDRG4 target for 5 pancreatic cancer samples and 4 colon cancer samples is shown with IGV. The first colon cancer library has low complexity evident by the sharp transition of read depth. The other eight libraries have good quality.

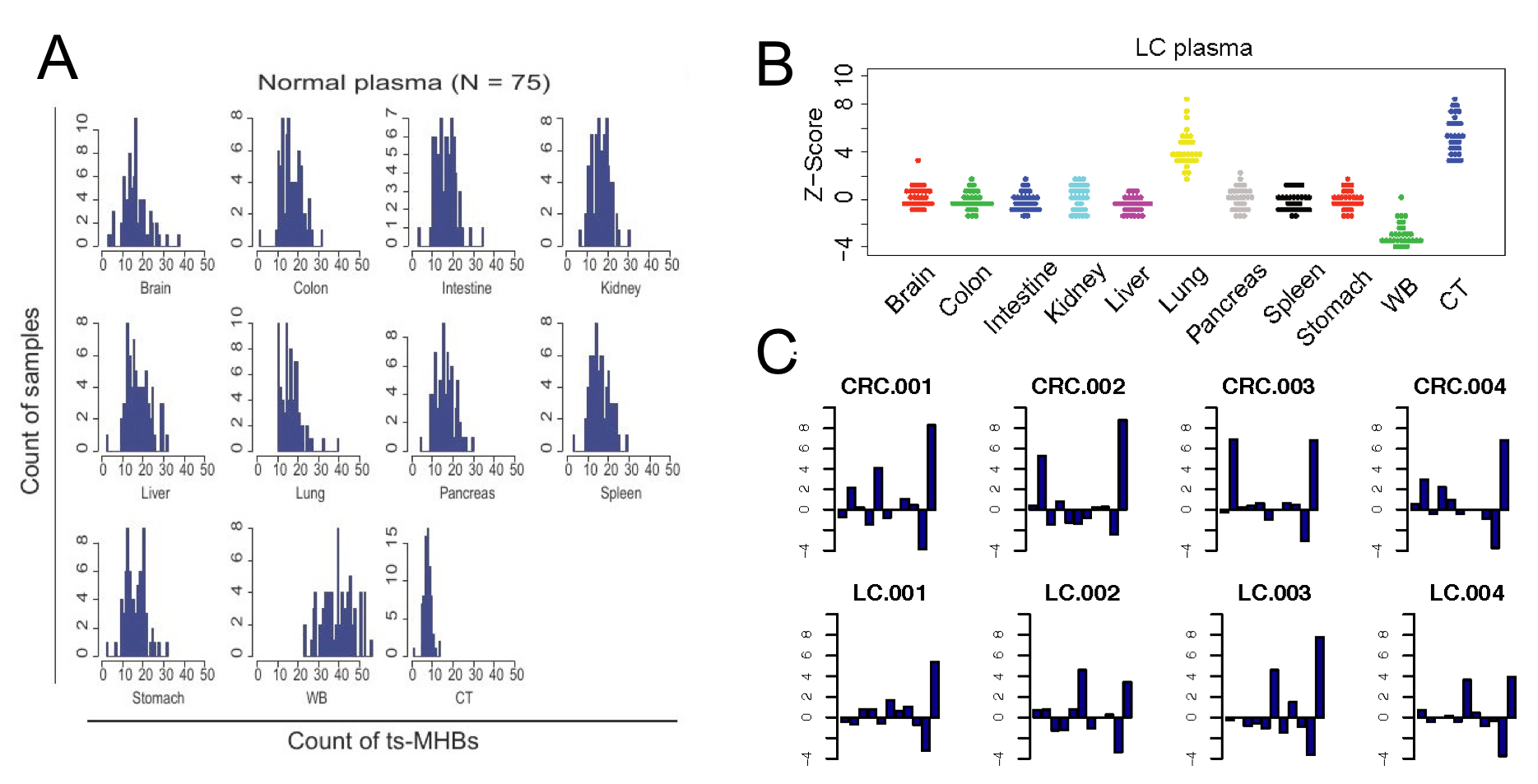
We have had unpublished results with the Nimblegen SeqCap Epi method, using a customized probe set covering 29.4Mb of genomic intervals that have low methylation in human whole blood. We have compared two library construction methods, Kapa Hyper Prep and Swift Accel Prep, and concluded that the Kapa Hyper Prep method leads to higher library complexity and is more robust. We have generated high quality data from 49 human plasma samples, including 8 from pancreatic cancer patients, 8 from colon cancer patients, 8 from lung cancer patients, and 25 from healthy controls. We achieved an average of 62% on-target rate with 10% clonal rate with ~15 million HiSeq PE150bp sequencing reads. The entire experimental workflow and data analysis pipeline for analyzing hybridization capture data has been established. For this project, we will only need to design a different probe set for Version 2.0 MHBs, covering 15-20Mb of genomic targets, for the screening of pancreatic cancer biomarkers in the initial set of 100 patient plasma, 100 control plasma, and 25 frozen tumor tissues from a fraction of the same patients.

How to deal with sparse bisulfite sequencing data and inter-patient variability: identify MHBs that sufficiently cover the diverse cancer-specific signatures of human cancer, for plasma sample of interest, thresholding the MHL for each MHB, then sum up the binarized data for all MHBs of cancer (or a specific tissue type) to derive a combined score.

Plan for marker identification. Existing markers: 9 solid tissue + WBC, additional tissues: Bladder, heart, muscle, vessel (“Aorta”), Thymus, esophagus, Ovary, brain (cerebellum, frontal lobe), placenta; Pan-cancer (CRC+Lung+PC); Stage I PC;

Identification of tissue-specific MHBs, cross validation with DMRs.

Potential problems and alternative solutions: heterogeneity among multiple tumor subtypes; differences among different disease stages are more subtle than expected; common signatures shared between pancreatitis and pancreatic cancer



**Figure xx. Simultaneous mapping of cfDNA to multiple tissue-of-origin and cancer.** (**A**) A total of 2880 ts-MHBs were identified for 10 normal tissues of interest, plus cancer. For a set of 75 normal plasma, we empirically determine the null distribution of the abundance score (defined as the number of ts-MHBs that contain MHL above a threshold) for each tissue. (**B**) For the cancer plasma samples, we calculate the deviation from the null distribution (as Z-scores) for the 10 normal tissues of interest plus cancer. Plasma samples from Lung Cancer (LC) patients clearly showed enrichment in both normal lung and cancer tissue (CT). (**C**) Distribution of Z-scores across the 10 normal tissues (in the same order as in B) and cancer for individual cancer patients. Colon cancer (CRC) patients tend to have high Z-scores for both colon and cancer, whereas Lung Cancer (LC) patients tend to have high Z-scores for both lung and cancer. Whether a patient has cancer, where the cancer is growing, and whether there is metastasis can be determined from the Z-score distribution after setting appropriate cutoff to control for specificity and sensitivity.

**Aim 2.** Development of a targeted methylation panel for non-invasive pancreatic cancer detection, and validation with clinical samples.

Method:

Technical benchmarking of two methods for efficient targeted bisulfite sequencing: Option A: circularization of bisulfite converted DNA, followed by hRCA and padlock capture; Option B: Methyl-AmpliSeq. Target both DNA strands, and tile across the MHBs to improve recovery. Prepare a titration series of HCT116 and NA12878 fragmented gDNA, determine limit of detection, conversion efficiency, library complexity, cost, workflow

Commit on one method for clinical validation, 200 plasma samples from pancreatic cancer patients, 200 from normal donors.

Determine the success rate (sensitivity) of detecting Stage I or II pancreatic cancer at a certain combination of specificity; Determine the success rate of distinguish benign pancreatic lesions from pancreatic cancer (stage I or above).

Potential problems and alternative solutions: If both options A/B do not have sufficient recovery, design a smaller version of NimbleGen methylSeqCap.

We will evaluate two different targeted methylation sequencing methods for implementing a methylation-based clinical assay. Key factors that we will consider include: (i) sensitivity and efficiency in converting low-input cfDNA into sequencing libraries; (ii) accuracy, precision and robustness in the measurement of DNA methylation level and methylation haplotype loads; (iii) assay cost; (iv) simplicity and robustness of the workflow; (v) suitability for broad dissemination to research and clinical labs.

BSPP (bisulfite padlock probes) is the first large-scale targeted methylation sequencing method that we developed and reported on Nature Biotechnology(Deng, Shoemaker et al. 2009). We have applied this methods to multiple biological studies(Ruiz, Diep et al. 2012, Gu, Liu et al. 2014, Plongthongkum, van Eijk et al. 2014), and further optimized the experimental and computational workflows over the years(Diep, Plongthongkum et al. 2012). BSPP was included in an international technical benchmarking study organized by the EU BluePrint consortium. In that study over 20 groups designed their assays for a common sets of genomic targets, and performed the measurements on a common set of reference DNA provided by the BluePrint consortium. The results were submitted to the a data analysis group in Italy led by Christoph Bock for unbiased comparison. The outcomes of this benchmarking study have been published on Nature Biotechnology last year(consortium 2016). Among all the methods included, three methods (BSPP, Raindance, Illumina 450k beadchip) are the three that are highly multiplexed, meaning that these assays can measure multiple CpG sites (from tens to hundreds of thousands) in one single reaction, which is essentially for ctDNA-based detection because the yield of DNA extraction from plasma is very limited. BSPP outperformed the other two assays in terms of accuracy. Internally, we have successfully implemented BSPP on ctDNA, either without any pre-amplification or on WGBS sequencing libraries. Therefore, it is a promising candidate for implementing the clinical assay using the markers that we will identified in Aim 1.

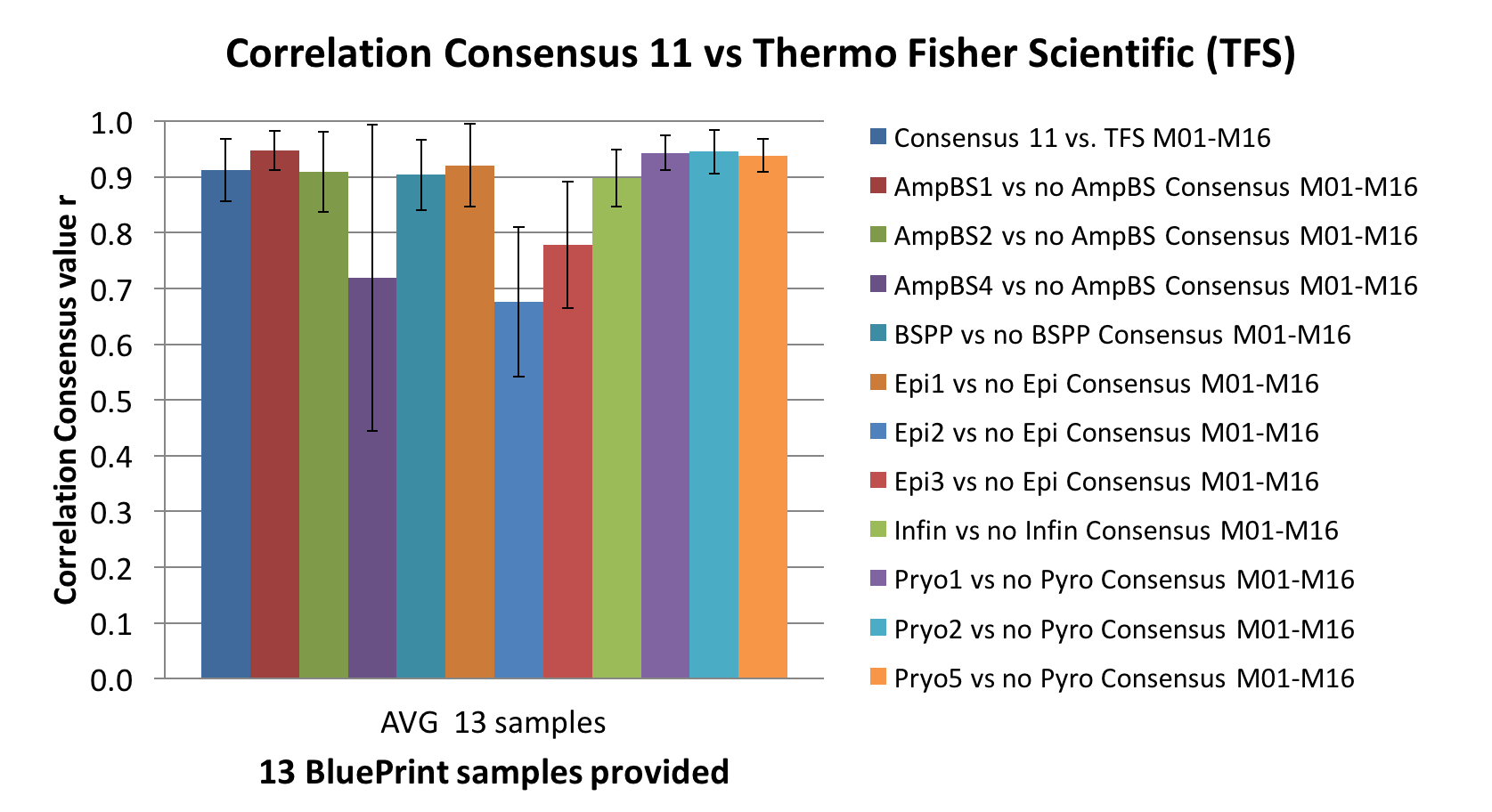


Figure xx. Technical benchmarking of Thermo Fisher Scientific’s Methyl-AmpliSeq against 11 other methods using a common set of reference DNA samples.

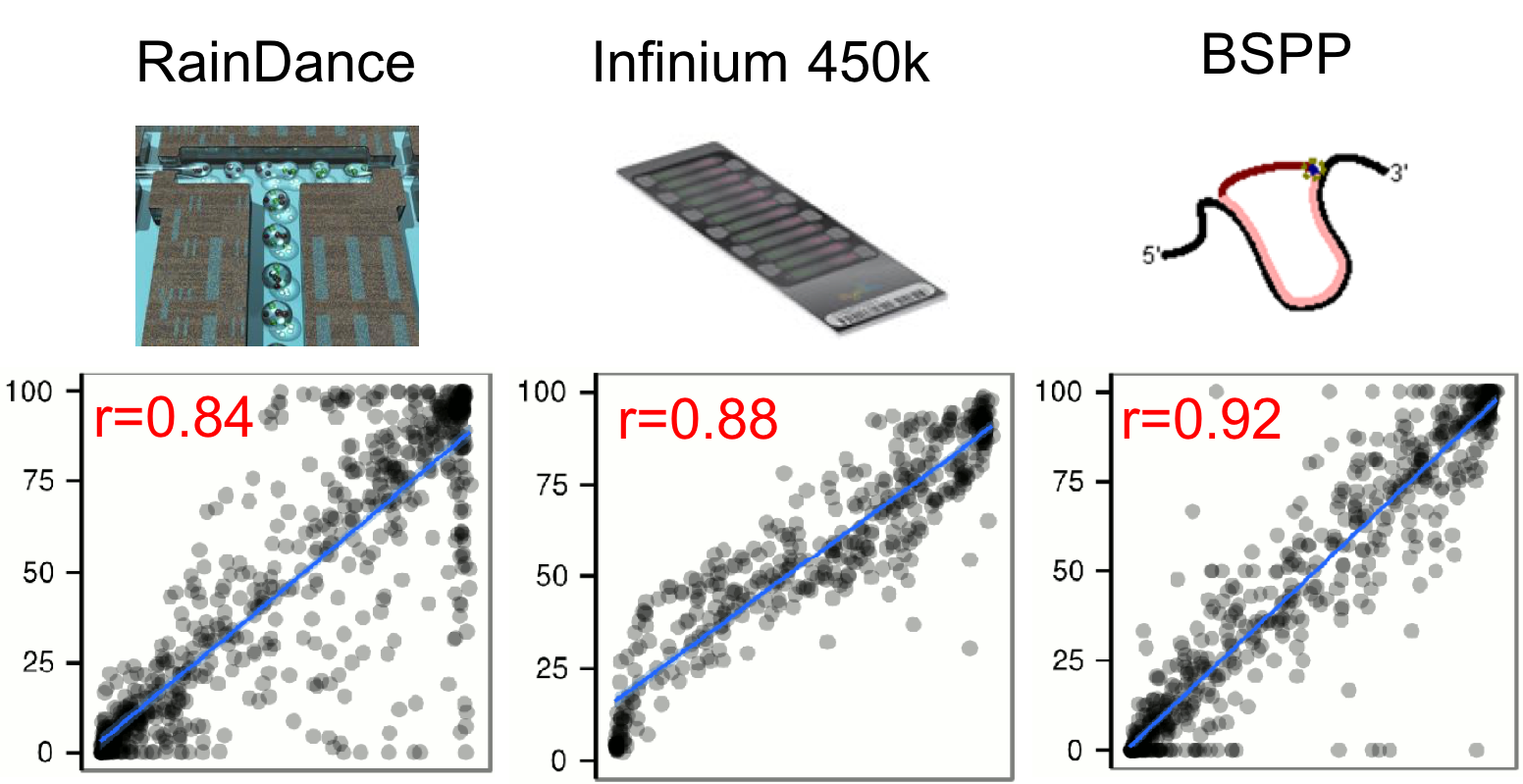


Figure xx. Technical benchmarking of three multiplexed target bisulfite sequencing methods. BSPP has higher concordance with the consensus values than Raindance and Illumina Infinium 450k. (Figure adapted from Bock et al. Nature Biotechnology, 2016).

AmpliSeq is a highly multiplexed PCR technology developed and commercialized by Thermo Fisher. It has been widely applied to targeted DNA sequencing for identifying germline and somatic mutations, and to RNA sequencing for quantifying gene expression and detecting fusion transcripts. This technology is very mature and has been implemented as clinical tests in many CLIA/CAP settings. An R&D group in Thermo Fisher initiated a collaborative project with us in 2016 to develop Methyl-AmpliSeq, by optimizing the primer design, amplification reagents/protocols, sequencing library construction protocol, and bioinformatics pipeline for the compatibility to bisulfite converted DNA. The method developed has been successfully completed. To benchmark this new assay, we used the same set of reference DNA provided by the EU BluePrint project, such that the results are directly comparable with all the published results from 20+ different groups. We showed that Methyl-AmpliSeq has very good performance, similar to BSPP, and superior to Raindance and Illumina 450k beadchip.

Technical benchmarking.

In this project, we will directly benchmark BSPP against Methyl-AmpliSeq using the markers identified in Aim 1, on cell-free DNA from pancreatic cancer patients. Methyl-AmpliSeq is relatively new and has not been tested on any cfDNA sample yet, so a technical benchmarking is necessary prior to the phase II validation on 400 (200 cancer patients + 200 normal controls) plasma samples. For this we will take high-quality genomic DNA from two cell lines (NA12878 and PANC1), generate ~160bp fragments with Covaris, and perform end-polishing to micmic cell-free DNA. We will then create a series of synthetic mixtures at 0.1%, 0.2%, 0.5%, 1%, 2%, 5%, 10%, 20% (PANC1 fraction), and run the two assays each in 5 replicates on these DNA samples. With the resulting 80 sets of data, we will determine and compare multiple metrics, including mapping rate, on-target rate, quantification accuracy, precision, limit of detection and limit of quantification. In the next stage, we will further apply both assay on ~100 cfDNA WGBS libraries constructed in Aim 1, and compare the measurements with the results based on Nimblegen SeqCap Epi.

Technical validation. [separate Aim 1 & Aim 2]

Analytic sensitivity & specificity: mix tumor DNA (Stage I, II, III) with normal plasma DNA at different ratios, perform the assay

Performance measures: LOD, LOQ, specificity, sensitivity, AUC

Diagnostic accuracy

Clinical validation

Specificity, sensitivity, AUC

Essential aspect for this FOA is a rigorous technical validation of the proposed technology/tool/capability. As appropriate to specific projects, such technical validation should be applied to new approaches, systems, materials, devices, mathematical models, etc. The goal should be substantiating the readiness of the proposed solutions for adoption by targeted end-user research communities. Specifically, validation should ensure that the transformative, enabling functionality of the proposed technologies/tools/capabilities for cancer research and/or clinical care is thoroughly verified and objectively confirmed (using appropriate benchmarks and performance measures) by the end of the project period.

We can emphasize the BLUEPRINT benchmarking publication, and that AmpliSeq BS has been validated using the same DNA samples.

**Key terms for the purposes of this FOA:**

**Enabling technology/tool/capability** denotes either a single new technology, method, approach, capability that offers transformative potential or a novel combination of appropriate components that collectively offer enabling characteristics/transformative potential for the targeted area of priority. Methylation haplotyping, a full set of MHBs.

**Advanced technology development** refers to research activities and other actions beyond the initial demonstration of feasibility, proof-of-concept, prototype stage, pilot studies, etc. Targeted sequencing of informative MHBs.

**Technical validation** refers to steps/actions needed to rigorously verify that the proposed new technology/tool/capability meets appropriate performance measures. The performance measures to be assessed should provide sufficient and objective proof that the novel technology/tool/capability is ready for adoption by targeted end-users and will offer genuine potential for transformative impact on the targeted area. Benchmarking with reference samples, from BluePrint or NA12878/HCT116 titration series; followed by validation with clinical samples (N>100).

Under a sub-heading *"Technical Validation",* address the following aspects:

* Explain clearly and precisely how the entire project and the individual elements of the proposed technology/tool/capability will be validated for the attainment of the intended qualities.
* Describe specific, objectively assessable (and quantifiable, as appropriate) *Performance Measures* for technical validation of the new technologies/tools-enabled capability to be achieved. *Performance Measures* should be clearly defined, scientifically justified, and quantitative whenever applicable.
* *Performance Measures* should address, as appropriate for the project, such aspects as:
* Statistical power analysis to support accuracy and precision of performance and also to define the number of specimens needed[How do we calculate the statistical power? Technical validation versus clinical validation ] Limit of detection, limit of quantification, replicates required for accuracy, detection of imprecision, PPV (positive predictive value, beyond the scope of this study), we can follow this guideline: <http://www.sciencedirect.com/science/article/pii/S2212066116300230> ; ‘rule of 3’, see PMID 20664632
* The means by which each of the proposed *Performance Measures* will be assessed.
* Explanation on how the proposed *Performance Measures* will provide the means for objectively assessing progress towards Specific Aims of the project. (Note: for some Specific Aims, it may be sufficient to define a single *Performance Measure* but other Specific Aims may require multiple *Performance Measures*).
* "Strategic" *Performance Measures* that could substantiate the expectations of the potential transformative/enabling impact that the new technology might have on cancer research or clinical oncology.
* While addressing *Potential Pitfalls and Alternative Approaches*, explain intended strategy(-ies) in case of inability to attain some critical performance benchmarks.

Under a sub-heading *"Path to Implementation",* address the following aspects:

* General timeline for development during the project period as well as well as a stage of readiness for the implementation to be attained by the end of the project period. Starting from Year 2: 4 months for assay design & initial testing, 4 months for technical benchmarking, 8 months for clinical validation, 8 months for manuscript writing, polishing protocol and computational pipeline.
* An outline of subsequent timeline to ultimate implementation.
* If applicable to your project, include:
* Plans to address regulatory requirements that might be needed either during the project period (e.g., experimental device status) and/or beyond (e.g., future clinical trials).
* Briefly address the intended steps/approaches during the project period and beyond to facilitate/promote/encourage adoption of the new technology/tool/ capability to be developed.

Prospective study: beyond the scope of this project (significantly more efforts and longer project time).

**Deliverables, Milestones & Timeline:**

**Resource Sharing Plan**: Individuals are required to comply with the instructions for the Resource Sharing Plans as provided in the SF424 (R&R) Application Guide, with the following modification:

* All applications, regardless of the amount of direct costs requested for any one year, should address a Data Sharing Plan.

Deposit all raw data to NCBI SRA, release all methylation haplotpyes, MHBs and markers on Zhang Lab website

* Resource Sharing Plans are expected for this FOA, and should be cohesive with the description under "*Path to Implementation",* including strategies intended beyond the period of the award.

AmpliSeq community panel

Canto, M. I., R. H. Hruban, E. K. Fishman, I. R. Kamel, R. Schulick, Z. Zhang, M. Topazian, N. Takahashi, J. Fletcher, G. Petersen, A. P. Klein, J. Axilbund, C. Griffin, S. Syngal, J. R. Saltzman, K. J. Mortele, J. Lee, E. Tamm, R. Vikram, P. Bhosale, D. Margolis, J. Farrell, M. Goggins and C. American Cancer of the Pancreas Screening (2012). "Frequent detection of pancreatic lesions in asymptomatic high-risk individuals." Gastroenterology **142**(4): 796-804; quiz e714-795.

consortium, B. (2016). "Quantitative comparison of DNA methylation assays for biomarker development and clinical applications." Nat Biotechnol **34**(7): 726-737.

Deng, J., R. Shoemaker, B. Xie, A. Gore, E. M. LeProust, J. Antosiewicz-Bourget, D. Egli, N. Maherali, I. H. Park, J. Yu, G. Q. Daley, K. Eggan, K. Hochedlinger, J. Thomson, W. Wang, Y. Gao and K. Zhang (2009). "Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming." Nat Biotechnol **27**(4): 353-360.

Diep, D., N. Plongthongkum, A. Gore, H. L. Fung, R. Shoemaker and K. Zhang (2012). "Library-free methylation sequencing with bisulfite padlock probes." Nat Methods **9**(3): 270-272.

Gu, Y., G. H. Liu, N. Plongthongkum, C. Benner, F. Yi, J. Qu, K. Suzuki, J. Yang, W. Zhang, M. Li, N. Montserrat, I. Crespo, A. Del Sol, C. R. Esteban, K. Zhang and J. C. Izpisua Belmonte (2014). "Global DNA methylation and transcriptional analyses of human ESC-derived cardiomyocytes." Protein Cell **5**(1): 59-68.

Guo, S., D. Diep, N. Plongthongkum, H. L. Fung, K. Zhang and K. Zhang (2017). "Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA." Nat Genet **49**(4): 635-642.

Henriksen, S. D., P. H. Madsen, A. C. Larsen, M. B. Johansen, A. M. Drewes, I. S. Pedersen, H. Krarup and O. Thorlacius-Ussing (2016). "Cell-free DNA promoter hypermethylation in plasma as a diagnostic marker for pancreatic adenocarcinoma." Clin Epigenetics **8**: 117.

Imamura, T., S. Komatsu, D. Ichikawa, T. Kawaguchi, M. Miyamae, W. Okajima, T. Ohashi, T. Arita, H. Konishi, A. Shiozaki, R. Morimura, H. Ikoma, K. Okamoto and E. Otsuji (2016). "Liquid biopsy in patients with pancreatic cancer: Circulating tumor cells and cell-free nucleic acids." World J Gastroenterol **22**(25): 5627-5641.

Jana, T., J. Shroff and M. S. Bhutani (2015). "Pancreatic cystic neoplasms: Review of current knowledge, diagnostic challenges, and management options." J Carcinog **14**: 3.

Liggett, T., A. Melnikov, Q. L. Yi, C. Replogle, R. Brand, K. Kaul, M. Talamonti, R. A. Abrams and V. Levenson (2010). "Differential methylation of cell-free circulating DNA among patients with pancreatic cancer versus chronic pancreatitis." Cancer **116**(7): 1674-1680.

Neoptolemos, J. P., D. H. Palmer, P. Ghaneh, E. E. Psarelli, J. W. Valle, C. M. Halloran, O. Faluyi, D. A. O'Reilly, D. Cunningham, J. Wadsley, S. Darby, T. Meyer, R. Gillmore, A. Anthoney, P. Lind, B. Glimelius, S. Falk, J. R. Izbicki, G. W. Middleton, S. Cummins, P. J. Ross, H. Wasan, A. McDonald, T. Crosby, Y. T. Ma, K. Patel, D. Sherriff, R. Soomal, D. Borg, S. Sothi, P. Hammel, T. Hackert, R. Jackson, M. W. Buchler and C. European Study Group for Pancreatic (2017). "Comparison of adjuvant gemcitabine and capecitabine with gemcitabine monotherapy in patients with resected pancreatic cancer (ESPAC-4): a multicentre, open-label, randomised, phase 3 trial." Lancet **389**(10073): 1011-1024.

Plongthongkum, N., K. R. van Eijk, S. de Jong, T. Wang, J. H. Sul, M. P. Boks, R. S. Kahn, H. L. Fung, R. A. Ophoff and K. Zhang (2014). "Characterization of genome-methylome interactions in 22 nuclear pedigrees." PLoS One **9**(7): e99313.

Radon, T. P., N. J. Massat, R. Jones, W. Alrawashdeh, L. Dumartin, D. Ennis, S. W. Duffy, H. M. Kocher, S. P. Pereira, L. Guarner posthumous, C. Murta-Nascimento, F. X. Real, N. Malats, J. Neoptolemos, E. Costello, W. Greenhalf, N. R. Lemoine and T. Crnogorac-Jurcevic (2015). "Identification of a Three-Biomarker Panel in Urine for Early Detection of Pancreatic Adenocarcinoma." Clin Cancer Res **21**(15): 3512-3521.

Rahib, L., B. D. Smith, R. Aizenberg, A. B. Rosenzweig, J. M. Fleshman and L. M. Matrisian (2014). "Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States." Cancer Res **74**(11): 2913-2921.

Ruiz, S., D. Diep, A. Gore, A. D. Panopoulos, N. Montserrat, N. Plongthongkum, S. Kumar, H. L. Fung, A. Giorgetti, J. Bilic, E. M. Batchelder, H. Zaehres, N. G. Kan, H. R. Scholer, M. Mercola, K. Zhang and J. C. Izpisua Belmonte (2012). "Identification of a specific reprogramming-associated epigenetic signature in human induced pluripotent stem cells." Proc Natl Acad Sci U S A **109**(40): 16196-16201.

Shoemaker, R., J. Deng, W. Wang and K. Zhang (2010). "Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome." Genome Res **20**(7): 883-889.