Development of Quality Control Methods for Reliable NGS Measurement of Methylome in Circulating DNA

Willey, JC[1], Crawford, EL[1], Craig, D.J.[1], Morrison, T[2], He, H.-J.[3], Herman, J. G.[4]

[1]The University of Toledo College of Medicine and Life Sciences Department of Medicine, Toledo, OH, [2]Accugenomics, Inc. Wilmington, NC, [3]National Institutes of Standards and Technology, [4]University of Pittsburgh

Introduction

Detection of aberrant levels and patterns of methylation in circulating free genomic DNA (cfDNA) may be useful in early detection of malignancy or in monitoring known malignancies for recurrence or treatment resistance. For example, altered methylation in lung cancer DNA may be linked to prognosis and may serve as a biomarker or therapeutic target (1,2). However, methylation detection in cfDNA is limited by sample size and quality as well as technical error. The goal of this study was to develop methods to improve accurate and sensitive detection of methylation in cfDNA using synthetic spike-in internal standards (IS) that enable target- and assay-specific limits of detection and control for technical error. For this proof-of-principle study, we targeted a region within the SOX2 gene.

Samples

Genomic DNA specimens with various levels of methylation were provided by the National Institute of Standards and Technology (NIST). These reference materials for measurement of DNA methylation were created by combining native state gDNA from a cell line with varying amounts of 100% *in vitro* methylated gDNA from the same cell line. GM24385

Sample A: Native state methylation only

Sample B: Sample A mixed with 5% methylated

Sample C: Sample A mixed with 20% methylated

Sample D: Sample A mixed with 50% methylated

Sample E: 100% methylated

Because each DNA region in the cell line has a different level of native methylation, expected levels at particular targeted regions were approximate.

References
(1) Hoang, P.H. and Landi, M.T. DNA Methylation in Lung Cancer: Mechanisms and Associations with Histological Subtypes, Molecular Alterations, and Major Epidemiological Factors. Cancers **2022**, 14, 961.
(2) Hulbert, A., Jusue-Torres, I., Stark, A., Chen, C., Rodgers, K., Lee, B., Griffin, C., Yang, A., Huang, P., Wrangle, J., Belinsky, S.A., Wang, T., Yang, S.C., Baylin, S.B., Brock, M.V., and Herman, J. G. Early Detection of Lung Cancer Using DNA Promoter Hypermethylation in Plasma and Sputum Epigenetic Lung Cancer Screening. Clinical Cancer Research **2017** 23(8), 1998.
(3) EpiTect® Bisulfite Handbook, Qiagen, Inc. February 2020.

Materials and Methods

Each NIST sample (A-E) alone or mixed with a commercially available IS mix (Accugenomics, Inc.) was bisulfite treated using the Epitect Bisulfite Kit (Qiagen, Inc.) which converts unmethylated cytosine (C) to uracil (U)(3).

|	Original Sequence	After bisulfite sequence	(3)
Unmethylated DNA	N-C-G-N-C-G-N-C-G-N	N-U-G-N-U-G-N-U-G-N	
Methylated DNA	N-C-G-N-C-G-N-C-G-N	N-C-G-N-C-G-N-C-G-N	

Following conversion, samples were PCR-amplified using primers designed to target specific regions and to anneal regardless of methylation status. Q5U® Hot Start High-Fidelity DNA Polymerase (New England Biolabs, Inc.) was used to reduce PCR errors and read through U. The replication and amplification process replaces U with thymine (T).

An additional PCR was conducted to add sample-specific barcodes and sequencing adaptors.
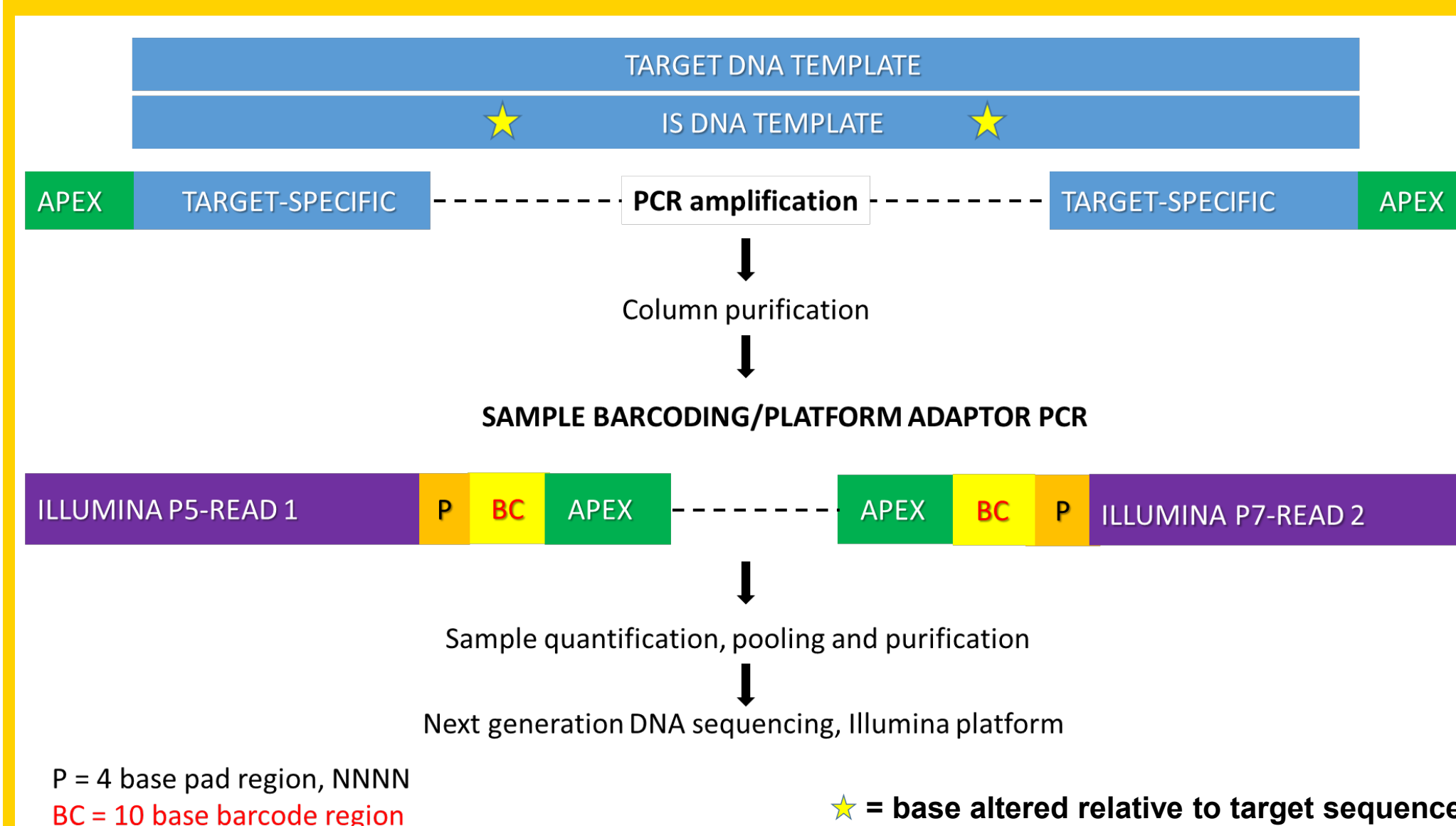
Purified libraries were sequenced at The University of Toledo on an Illumina MiSeq Next Generation Sequencer using Illumina MiSeq V2 Nano reagents.

Bioinformatic Analysis was conducted using Qiagen CLC Genomics Workbench.
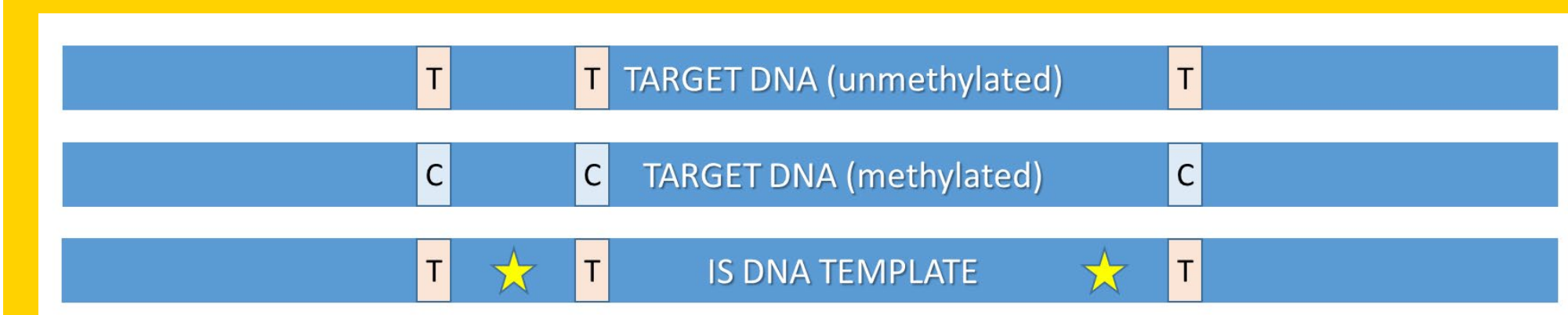
Ratios of sequencing read counts for methylated vs unmethylated and NT vs IS were calculated for each base position in the targeted region. IS values were used to control for position specific technical error at the methylation sites.

Results

• Variation in methylation among targeted methylation sites in the native DNA (Sample A, blue circles)
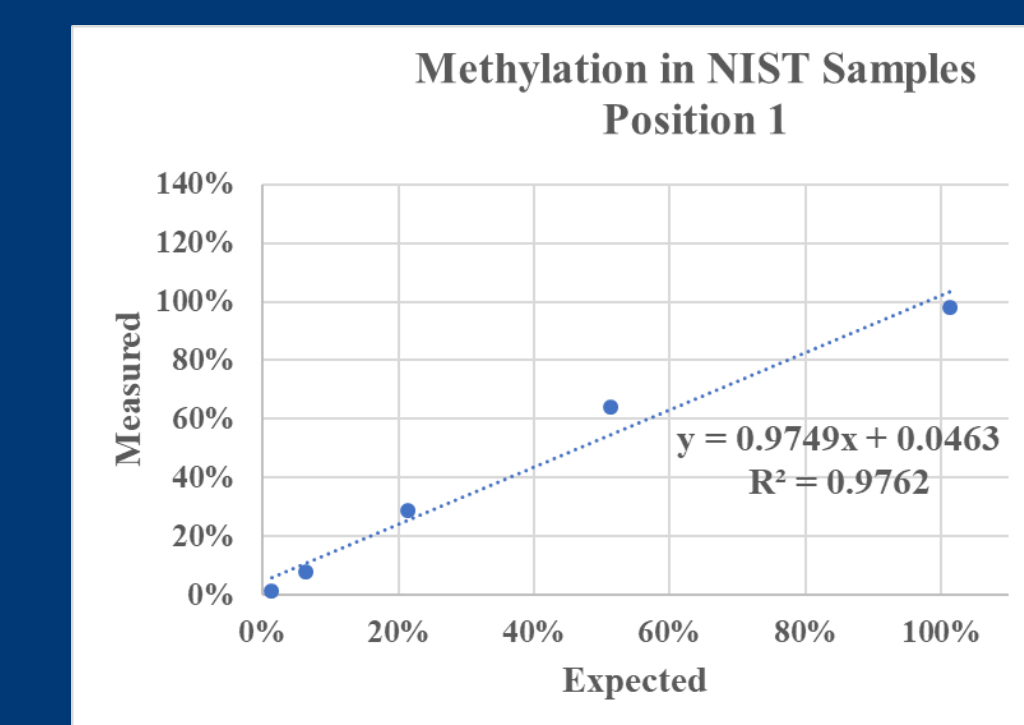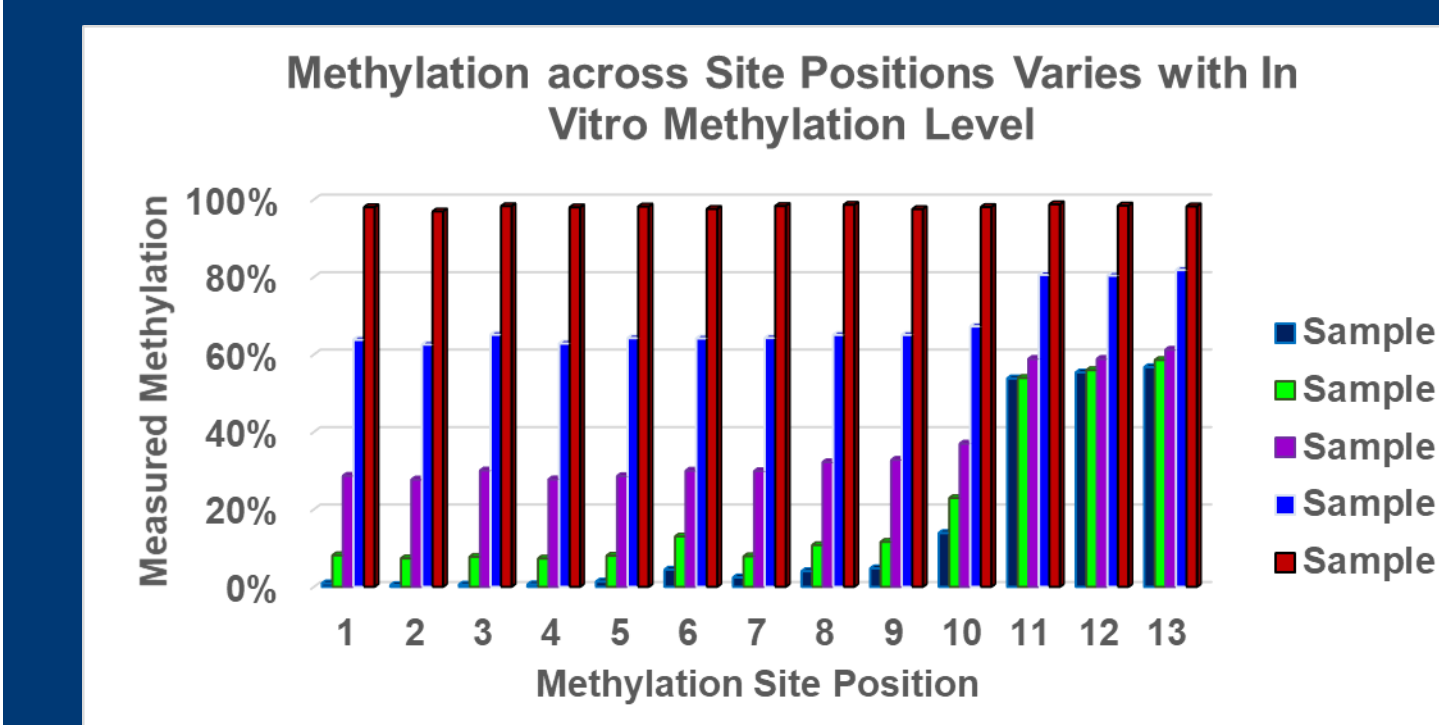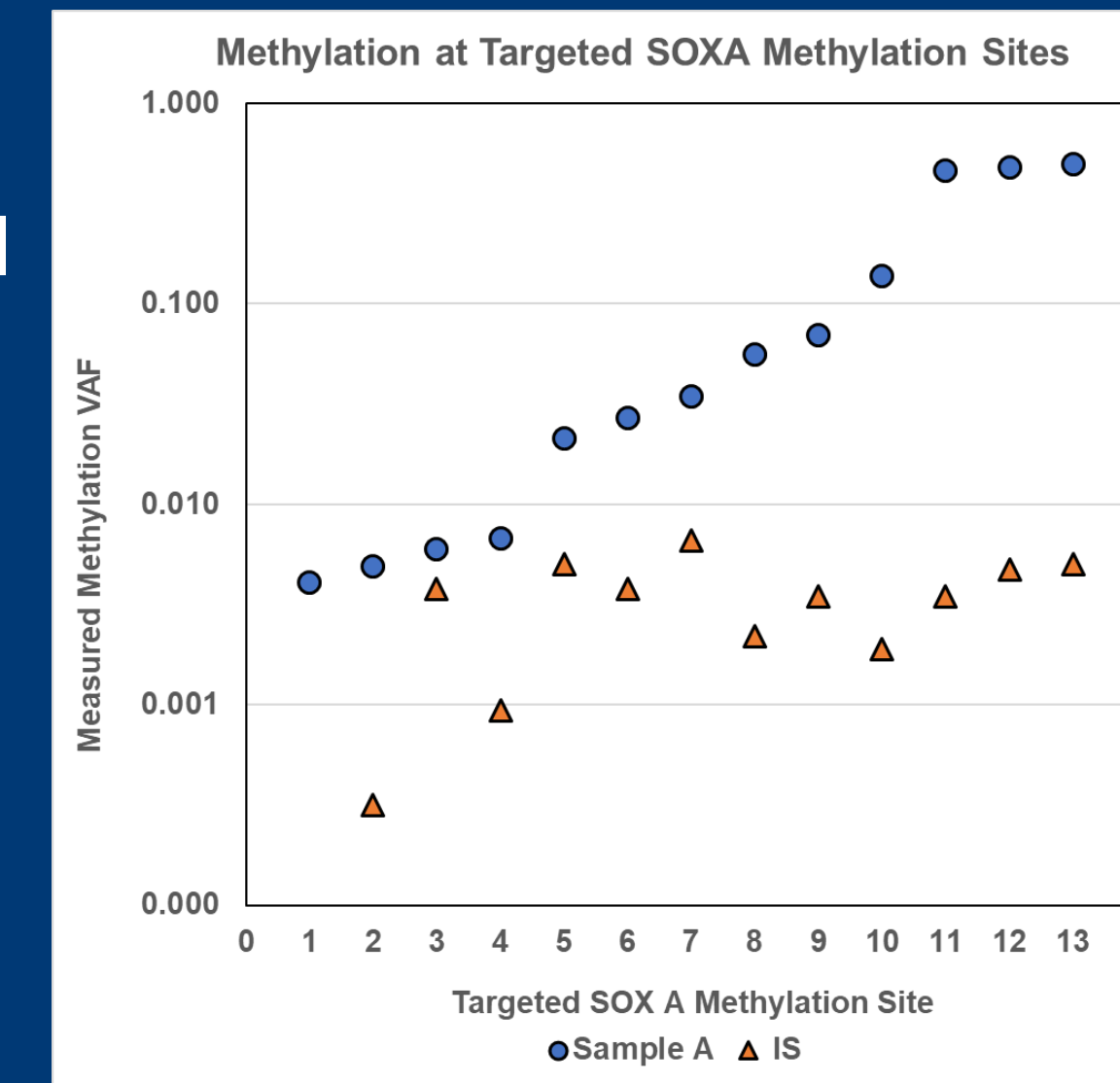 • Mean methylated allele frequency (mMAF): Sample A: 15%
• Low level methylation measured in IS controls (IS, gold triangles)
  • Mean across 13 sites: 0.34%
  • STDEV: 0.18%
  • CV: 0.53

The effect of *in vitro* methylation varies across potential methylation sites with varying levels of native methylation. Methylation increased as expected at each position across samples to a maximum plateau (Ave. 95% in targeted region). Left: measured methylation in all samples across 13 potential methylation sites. Right: measured vs. expected methylation at potential methylation site 1.

Conclusions

• A synthetic internal standard (IS) spike-in enabled measurement of technical error at each measured methylation site in the targeted SOX2 region of the NIST reference material.
• Novel reference materials prepared by NIST and provided to this lab for Beta-testing enabled these preliminary studies to assess the utility of synthetic internal standards in analysis of methylated cfDNA.
• Based on these results, there is a clear need, and a path to optimization of IS design. Re-design will
  • Avoid base alterations relative to the endogenous template that may be affected by methylation status. This change is expected to lead to increased sensitivity and decreased technical error.
  • Optimize settings for bioinformatic analysis with CLC Genomics Workbench to improve separation of target sequence from IS sequence.
• Testing for use in cfDNA analysis will require selection of regions with less native methylation and/or creation of synthetic, completely unmethylated templates.

Ongoing Work

Based on results from this pilot study, we changed the IS design so that only T>A or A>T changes were introduced in the IS relative to endogenous template sequence. This will enable all spanned potential methylation sites to be evaluated.

The optimized design was used to prepare IS that will be used to target four gene regions known to have altered methylation in lung cancer and that promise to serve as a biomarker for early detection.

The newly designed IS sequences were cloned into plasmids, isolated, sequenced and linearized. Three methods of quantification were used to determine concentrations and a mixture of these IS has been created.

Cloned IS sequences were PCR-amplified and quantified. This product will enable comparison of plasmid IS (which may have some low level of methylation due to propagation in E. coli) to completely unmethylated IS.

NIST sample A has been mixed with plasmid IS and with amplicon IS and bisulfite treated. Library generation and sequencing are in progress.

Collection of cfDNA from human subjects with suspected or known lung cancer has begun and is ongoing. Once protocols for methylation detection are optimized, these samples will be evaluated and correlated to diagnosis.