CALIFORNIA STATE UNIVERSITY SAN MARCOS

PROJECT SIGNATURE PAGE

PROJECT SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE

MASTER OF SCIENCE

IN

BIOTECHNOLOGY

PROJECT TITLE: Development of a universal contamination testing protocol

AUTHOR: Rita Pitts

DATE OF SUCCESSFUL DEFENSE: April 23, 2013

THE PROJECT HAS BEEN ACCEPTED BY THE PROJECT COMMITTEE IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

Betsy Read
PROJECT COMMITTEE CHAIR

DATE

Darren Davis
PROJECT COMMITTEE MEMBER

DATE

Julie Jameson
PROJECT COMMITTEE MEMBER

DATE

Jill Litschewski
PROJECT COMMITTEE MEMBER

DATE
EXECUTIVE SUMMARY

Development of a universal contamination testing protocol for index adapters used for multiple Illumina, Inc. platforms

Sponsor by Illumina, Inc.
Rita R. Pitts
April 23, 2013

Professional Science Master’s Degree Program
California State University, San Marcos

Next generation sequencing, NGS, is revolutionizing the world of science. Illumina, Inc.’s NGS chemistry begins with sample preparation which involves fragmenting genomic DNA (gDNA) and attaching oligonucleotide adapters. DNA fragments are then amplified, in a process referred to as cluster generation and sequenced in parallel. The adapters added during the sample preparation process are required for Illumina’s NGS chemistry. Most of the adapters have a unique index sequence (index adapter) incorporated in them, which allows sample tracking and multiplexing. Illumina, Inc. has a diverse NGS product portfolio that currently uses 220 index adapters from seven different sample preparation product lines. Each manufactured index adapter requires quality control (QC) contamination testing. The purpose of this testing is to determine if the indexed adapters under evaluation are cross-contaminated with any other indexed adapter. This practice is particularly important because the end-user utilizes the Illumina, Inc. supplied adapter sequences for sample identification and cross-contamination can result in sample tracking errors. It is imperative that contamination testing be conducted on all manufactured adapters; however the majority of the QC testing of these product lines involves labor-intensive sample preparation protocols and instrument time intensive sequencing runs. The project aimed to move the contamination testing of index adapters with library preparation protocols to a shorter PCR-based protocol, introduce a faster cluster generation protocol, and combine multiple QC contamination processes into a single protocol. The single protocol allows all product lines to be tested on a single sequencing run, with data analyzed using a common algorithm. This will result in approximately a 40% reduction in sequencing runs for contamination testing, leading to subsequent reductions in costs as a single index contamination sequencing run has an internal price tag of about $3000. Ultimately, 220 different Illumina, Inc. reagents would be positively affected by the implementation of a universal contamination protocol. This change will greatly benefit the turnaround or cycling time (i.e. time it takes to go through quality control) for these particular reagents, providing QC the ability to focus more time, attention, and money on other important areas.

A new protocol was created to integrate all existing index adapter contamination protocols into a single workflow. Optimal PCR conditions and optimal adapter concentrations were identified. The fast clustering protocol for cluster generation used with the new protocol/workflow produced more accurate estimates of contamination levels compared to the current sample preparation protocol/workflow used with the standard cluster recipe. The new fast cluster recipe reduces the cluster generation step by two hours. The overarching goal of this investigation was to develop a method that would be more time and cost effective for index adapter contamination testing. The protocol is still in the validation stage as more sequencing runs are needed. Once these data points have been evaluated and the responsible parties for transfer to production have been satisfied, the new protocol will be implemented, resulting in a more efficient process estimated to save the company $132,000 per year.
Development of a universal contamination testing protocol for index adapters used for multiple Illumina, Inc. platforms

May 14, 2012
Rita Pitts
Darren Davis
Presentation Outline

- Next Generation Sequencing (NGS)
  - Sample Preparation
  - Cluster Generation
  - Genomic sequencing using Illumina, Inc. technology

- Problem/Solution: Creating a Universal Index Adapter Contamination Protocol
  - Importance/Impact
  - Methods/Materials
  - Major Findings
  - Discussion

- Conclusions/Future Directions
What is Next Generation Sequencing? (NGS)
Capillary Electrophoresis-based Sanger Sequencing

http://spectrum.ieee.org/biomedical/diagnostics/gene-sequencings-industrial-revolution
Illumina, Inc.’s Sequencing by Synthesis (SBS) Technology
Illumina Inc. NGS Workflow

Start with Sample Preparation

http://www.dkfz.de/gpcf/illumina_hiseq_technology.html
Illumina Inc. NGS Workflow cont.
Cluster Generation (Clonal Amplification-Clustering)

http://www.dkfz.de/gpcf/illumina_hiseq_technology.html

www.illumina.com
Illumina Inc. NGS Workflow cont.

Illumina Inc.’s Sequencing By Synthesis (SBS)

Add 4 Fl-NTP's + Polymerase  
Add 4 Fl-NTP's + Polymerase  
Add 4 Fl-NTP's + Polymerase
Index Adapters

Index Adapters are an essential element in Library Preparation

The Problem: Seven unique protocols for current contamination testing for index adapters
Current Contamination Testing

- **What:**
  - The contamination test is a short GA sequencing run performed on all indexed adapters to determine the index sequence.

- **Why:**
  - Aim is to confirm that indexed adapters are not cross-contaminated with other indexed adapters.
  - The index sequence is used by customers to identify samples. Cross-contamination of adapters could result in sample tracking errors.

- **How:**
  - Developed on product-specific basis
    - TruSeq Adapter Plate bulk – PCR → sequencing
    - TruSeq – PCR → sequencing
    - TruSeq smRNA – PCR → sequencing
    - TruSeq Targeted RNA – PCR → sequencing
    - TruSeq Custom Amplicon – library prep → sequencing
    - Nextera – library prep → sequencing
Current practice cost to the company

- Complex training required
- Waste of resources
- Time and cost inefficiency
- Single sequencing run costs $3000 (internal cost)
- Long turnaround time for protocols with sample prep
- Lack of scalability
Cost Analysis

- Proposed protocol results in a decrease in annual sequencing runs from 157 to 113.
- This is a cost saving of $132,000 per year
The Solution: Create a universal PCR-based contamination test
Proposed Workflow

- PCR adapter using product-specific primers
- Use universal PCR protocol
- Quantitate library on Bioanalyzer
- Fast clustering on cBot (<2 hours current time 4.5 hours)
- Sequence on GA
- Run different platforms on the same flowcell
- Analyze data with universal script
PCR Optimization

- Each PCR reaction contained
  - 12.5 μL PML
  - 1 μL 10 μM PCR primer
  - Ultrapure water + adapter to a final volume of 25 μL

- PCR conditions were modeled after the TruSeq SmRNA™
  - 98°C, 30 seconds
  - 10 cycles of:
    - 98°C, 10 seconds
    - 60°C, 30 seconds
    - 72°C, 30 seconds
  - 72°C, 5 minutes x 1 cycle
Titration results for determining optimal PCR conditions
Summary & Conclusions

- PCR product was seen in all adapters for all concentrations tested.
- 2 μL of undiluted adapter was chosen as the best quantity to use for both TruSeq Custom Amplicon™ and Nextera™ DNA because:
  - PCR product was generated for all adapters at this concentration with high yields (~275-350 nM)
  - It was away from a performance edge (the lowest concentrations gave less product)
  - There is little yield benefit beyond this input quantity for Nextera™ DNA Sample prep
  - It gave the cleanest product for N702
  - It will be simplest logistically – the same volume as the current TruSeq SmRNA™ and TruSeq Custom Amplicon™ protocols, but with no dilution step.
PCR library optimization for fast clustering

Fast vs Standard cBot Recipe
Cluster Density

Cluster density (K/mm²)

N501
N702
AR001
R709
A501
A701
RPI14

Standard
Fast
PCR library optimization for fast clustering cont.

![Fast cBot Recipe
Cluster Density vs Library Conc](image)

- Cluster Density (K/mm²)
- Library conc (pM)
- A701-Std
- A702-Std
- A701-Fast
- A702-Fast
PCR-based protocol validation

- TruSeq Custom Amplicon™ and Nextera™ adapters currently have a library prep procedure for contamination testing.
- The PCR-based protocol was validated for these products by preparing adapters with known amounts of contamination.
- The current library prep protocols and the new PCR-based protocol were compared for each contaminated adapter sample.
New protocol validation; TruSeq Custom Amplicon™

**New Workflow Validation - TruSeq Custom Amplicon**

- Expected Contamination Level
- Actual Contamination Level - Standard Clustering
- Actual Contamination Level - Fast Clustering
- Actual Contamination Level - Current Sample Prep Protocol

**TruSeq Custom Amplicon™ Adapter**

A708  A706  A710  A504  A506  A508
New protocol validation; Nextera™DNA
Conclusions

- PCR was successful for all adapters with the new protocol
- Fast clustering protocol gives comparable cluster density
- New PCR-based contamination test protocol accurately detects contamination
- The new contamination test protocol offers significant turnaround time improvements
- The new contamination test protocol is predicted to save the company $132,000 annually
Future Directions

- Complete PCR-based protocol validation for Nextera™
- Reanalyze existing data to validate script for TruSeq SmRNA and TrueSeq Amplicon platforms
- Determine variability in current and new protocols
  - Re-cluster and re-sequence contaminated samples
  - Analyze historical data for adapter contamination
  - Generate data from multiple users