Exome ID Panel Application Note

CUSTOMER COLLABORATIVE SOLUTION FROM AGENA BIOSCIENCE

Darryl Irwin, Priscilla Hunt, and Marisa Pearce*
*Corresponding author: Marisa.Pearce@AgenaBio.com

INTRODUCTION

A sample can be mislabeled at point of origin or mixed up during transfer from one laboratory to another. Complex laboratory workflows involve multiple stages, introducing numerous steps with potential for sample mishandling: an improperly sterilized pipette, a lab worker momentarily distracted, a misread label, or a typo on a record sheet. It has been estimated that ~2% of samples stored in biobanks and other repositories are incorrectly annotated. Furthermore, cell banks report that ~20% of cell lines sent for inclusion in their repositories for use by researchers are improperly identified. The challenge for the research community is to ensure that the correct samples are used and reported in research studies. These types of sample mix-ups prove difficult to pinpoint and deconvolute, which could undermine the accuracy of results and scientific merit of translational and clinical research studies.

Translational and clinical research studies are frequently performed on archival clinical specimens, including surgical resections, fine needle aspirates, blood, and plasma. Archival samples are commonly collected over many years, with various methods for sample stabilization, including fresh-frozen tissue and formalin-fixed, paraffin-embedded (FFPE) tissue, and varied storage environmental conditions such as temperature and humidity. This results in highly varied levels of degradation and the introduction of artifacts due to the chemical transformation capabilities of stabilizing agents, such as formalin, that affect the overall nucleic acid quality and quantity available for genetic analysis. All these factors highlight the clear need to ensure sample identity and integrity prior to embarking on molecular-based experiments.

OVERVIEW

This application note describes the development of the Exome ID Panel for use on the MassARRAY® System. This panel, analogous to Agena Bioscience's iPLEX® Pro Sample ID Panel, was developed in response to a customer's request for a panel for use in exome sequencing and cytogenetic microarray workflows.

In December 2012 the iPLEX Pro Sample ID Panel was launched, supported by MassARRAY analysis software and database, providing a rapid method for comparing a sample to others and to score for a match or unexpected mismatch. This panel contains 44 SNPs and 3 gender markers for sample identification and 5 DNA copy number controls for sample quantification and quality assessment in a single multiplexed SNP genotyping reaction requiring ≤ 10 nanograms of DNA per sample. The iPLEX Pro Sample ID Panel provides a highly accurate and rapid method for identifying and quantifying genomic DNA samples prior to PCR-based assay methods on the MassARRAY System or alternative genetic analysis technologies.

While genomic DNA samples can enable interpretation of single nucleotide (SNV) and copy number variation (CNV), use of corresponding RNA samples is needed to provide additional information, in terms of drug response, disease progression, or other applications. A panel solution to address this need was developed. The resultant custom Exome ID Panel includes identity markers present in exome sequencing/microarray data to facilitate direct chain-of-custody matching. In addition, due to the requirement for increased intact DNA template quality and quantity for downstream analysis, quality assessment markers span a larger intact template molecular weight size range and

For Research Use Only. Not for use in diagnostic procedures.

generate a degradation curve/template size ladder that matches varied input requirements of subsequent genetic analysis technologies.

ASSAY DESIGN AND DATA ANALYSIS

Sample identification markers (SNPs) are based on the following marker inclusion criteria:

Primary Inclusion Criteria:

- SNPs located in gene exon regions
- SNPs with high minor allele frequency (>0.4 MAF)
- Present on cytogenetic arrays:
    - Illumina® HumanCytoSNP-12 DNA Analysis BeadChip
    - Illumina HumanOmniExpress (Plus)
    - Affymetrix® CytoScan® 750K Array
    - Affymetrix CytoScan HD

Secondary Inclusion Criteria:

- Present on Illumina Infinium® HumanCoreExome BeadChip (26/42 SNPs)

Gender markers (3) utilize chromosome X/Y paralog sequences where the 2 paralogues can be co-amplified by a single PCR primer pair and distinguished with single base extension assays by differential bases or short INDEL.

Sample quality and quantity markers (5 X DNA copy number control assays) utilize competitive PCR with an internal synthetic template for quantification of absolute numbers of starting intact template. Marker target loci are based on the following marker inclusion criteria:

- Gene coding regions commonly used as reference genes in published cancer research studies.
- Exons allowing PCR amplicon sizes ranging from 100-500bp on both cDNA and DNA template (short exons excluded).
- Somatically stable (based on COSMIC, ENSEMBL and CONAN):
    - Low level germline and somatic copy number variation (not observed as double deleted).
    - Low level polymorphisms and somatic mutations (synonymous, missense, nonsense and INDELs).

A total of 42 SNPs (exonic) and 3 gender markers for sample identification, and 5 DNA copy number controls for sample quantification and quality assessment (100-500bp),

SNPS WITHIN THE EXOME ID PANEL		
COPY NUMBER CONTROLS — Albumin_100	rs12102203	rs4664475
PPIA_200	rs12594531	rs4924
Albumin_300	rs132655	rs495680
PDCD2_400	rs1344	rs586421
TBP_500	rs140679	rs6090043
GENDER MARKERS — AMEL_XY	rs1657502	rs6420424
ARSD_XY	rs17548783	rs6587161
TGIF2L_XY	rs2246209	rs6977125
rs1045728	rs2273171	rs707507
rs10495563	rs2301771	rs7653897
rs1065457	rs2736588	rs773901
rs1127379	rs2889732	rs7949414
rs1128536	rs3393	rs843364
rs1131906	rs3737434	rs8473
rs11356	rs3743165	rs9131
rs11998387	rs3884596	rs9897794
rs1200349	rs4478844	

Figure 1:
Exome ID Panel. This figure includes 42 SNPs (designated by reference sequence (rs)), 5 DNA copy number assays (Gene_Amplicon Length), and 3 gender markers *(AMEL, ARSD, and TGIF2L)*.

were multiplexed into a single iPLEX Pro SNP genotyping reaction, using Assay Design Suite software (Figure 1).

The Exome ID Panel is analyzed using iPLEX Pro Sample ID Panel reports within MassARRAY analysis software. Please contact Agena Bioscience Customer Support for detailed information on the Sample ID Panel report capabilities and how it can be used for this specific panel along with additional plot generation.

In brief, MassARRAY analysis software generates a database to readily manage sample data and interpret relationships.

- Samples can be queried against a database to quickly determine expected and unexpected matches using either germline vs. germline OR germline vs. tumor comparisons, taking into account that tumor samples may be subject to loss-of-heterozygosity.
- Automated reports provide the SNP profile, gender ID, and copy number for every sample run.
- A historical algorithm compares sample(s) against the entire database of samples previously run.
- A local algorithm compares a sample against other samples within the same run.

The Exome ID Panel includes quality assessment markers (5) spanning a 100-500bp intact template size range. Sample degradation curve/template size ladders can be generated by plotting the absolute number of intact amplifiable template molecules reported vs. target amplicon length (Figure 2).

DEGRADATION CURVES/TEMPLATE SIZE LADDERS GENERATED FROM THE EXOME ID PANEL
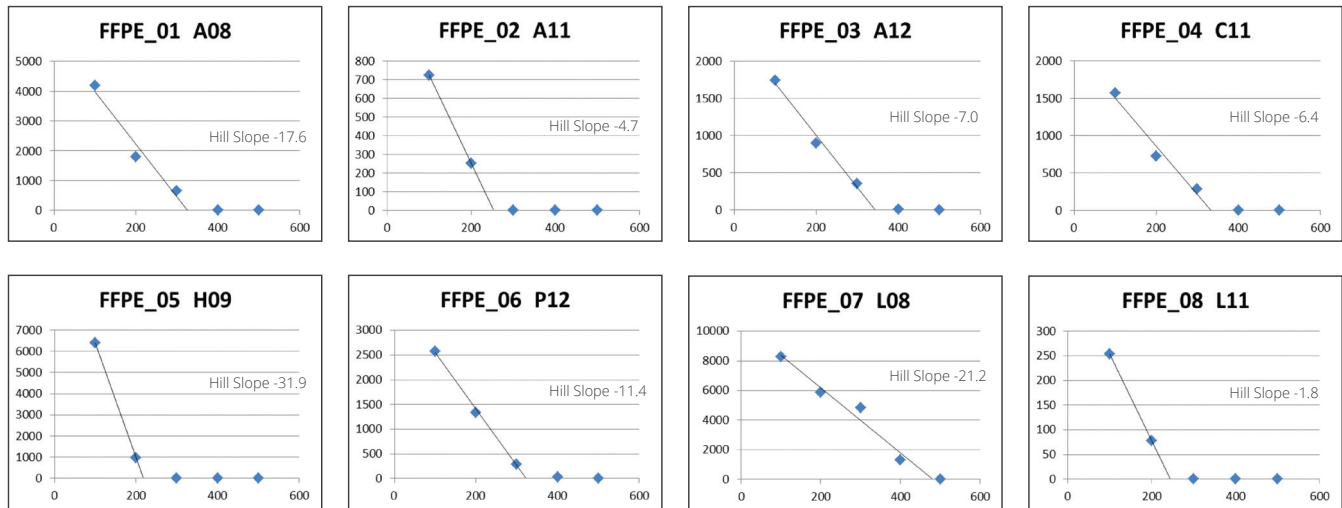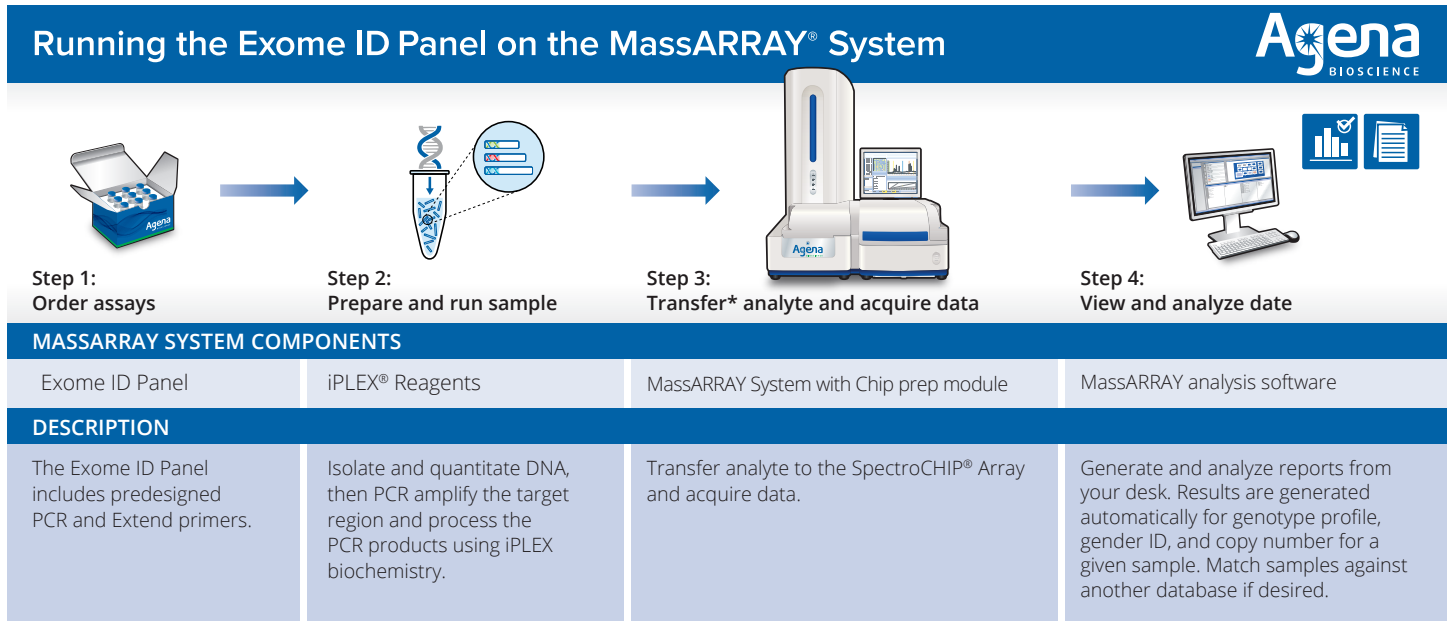
Figure 2:
Degradation curves/template size ladders generated from the Exome ID Panel on FFPE samples. This figure depicts 8 FFPE samples with varying levels of degradation/template fragmentation. The X axis is the target assay amplicon size (bp) and the Y axis is the number of amplifiable intact template molecules for each target amplicon size.

As all extracted DNA samples result in some level of degradation, the Exome ID Panel degradation curves are normalized to the average degradation observed in extracted DNAs from 94 HapMap cell lines (Coriell Cell Repositories, NJ, USA). Reference sample degradation can be modified to meet user requirements – to utilize an alternative reference data set please contact Agena Bioscience Customer Support. Samples showing a positive hill slope have lower levels of degradation than the reference, whereas samples with a negative hill slope are more degraded than the reference. The magnitude of the hill slope indicates the extent of degradation. Slope and range of the curve are used to match sample quality to the input requirements of various downstream genetic analysis technologies.

As expected, extracted DNA from clinical FFPE samples shows high levels of degradation due to DNA cross-linking and cleavage during the formalin fixation and paraffin embedding process as well as sample archival environment conditions. Samples depicted in Figure 2 include surgical and fine needle aspirate biopsies, which were all FFPE samples of lung, colon, or skin (melanoma) cancers.

Degradation hill slopes and amplifiable template copy number ranges vary greatly in clinical FFPE samples, requiring accurate quality assessment, as provided by the Exome ID Panel, to match sample quality with the varied input requirements of downstream genetic analysis technologies. All samples depicted have sufficient short amplicon (100bp) intact template for downstream MassARRAY somatic mutation analysis. Longer amplicon analysis requirements will be compromised in many of these samples.

Running the Exome ID Panel on the MassARRAY® System

Step 1:
Order assays

Step 2:
Prepare and run sample

Step 3:
Transfer* analyte and acquire data

Step 4:
View and analyze date

MASSARRAY SYSTEM COMPONENTS			
Exome ID Panel	iPLEX® Reagents	MassARRAY System with Chip prep module	MassARRAY analysis software
DESCRIPTION			
The Exome ID Panel includes predesigned PCR and Extend primers.	Isolate and quantitate DNA, then PCR amplify the target region and process the PCR products using iPLEX biochemistry.	Transfer analyte to the SpectroCHIP® Array and acquire data.	Generate and analyze reports from your desk. Results are generated automatically for genotype profile, gender ID, and copy number for a given sample. Match samples against another database if desired.

Figure 3: Overview of the workflow for the Exome ID Panel.
*Dispensing options vary by SpectroCHIP Array format.

SUMMARY

The custom Exome ID Panel was developed as a collaborative solution for confirming chain of custody in complex exome sequencing and cytogenetic microarray workflows. This panel takes advantage of the iPLEX Pro Sample ID reports facilitated by MassARRAY analysis software. The Exome ID Panel includes a total of 42 SNPs (high minor allele frequency and located within exon regions), 3 gender markers for sample identification, and 5 DNA copy number controls for sample quantification and quality assessment. Clinical research DNA samples exhibit highly varied quantity and quality, resulting in the risk of false positive and negative results in biomarker research studies. The incorporation of nucleic acid quality markers, spanning a PCR amplicon range of 100-500bp, provides additional quality assessment via a degradation curve/template size ladder, which can be matched to the varied input sample quality requirements and amplicon sizes of downstream genetic analysis technologies. These 50 assays are multiplexed into a single SNP genotyping reaction requiring ≤ 10 nanograms of DNA per sample, providing a rapid, cost effective, and highly accurate method for sample identification and quantification.

WEB: agenabioscience.com
US: +1.877.4.GENOME

EU: +49.40.899676.0
AP: +61.7.3088.1600

CN: +86.21.6427.0566

51-20016R3.0 0317