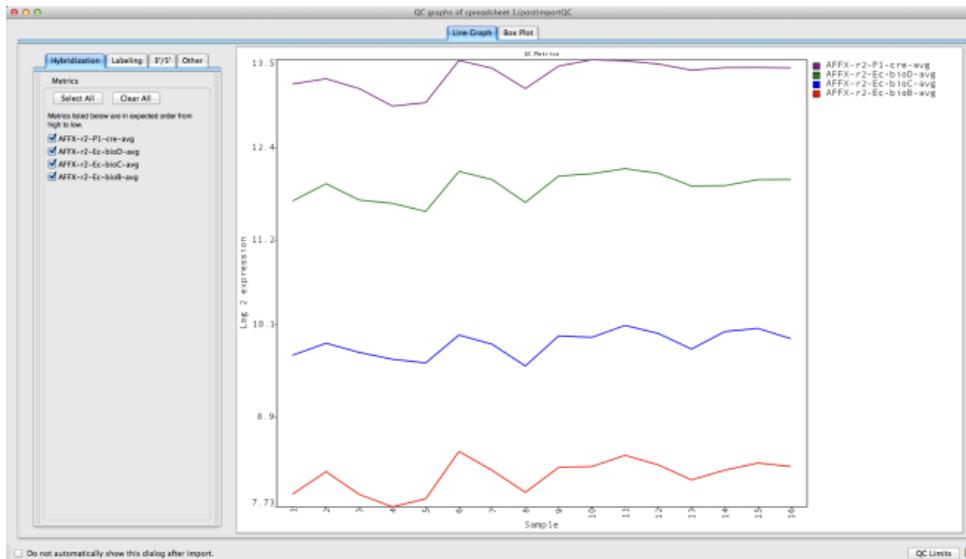


## Appendix 1: QC for microarray

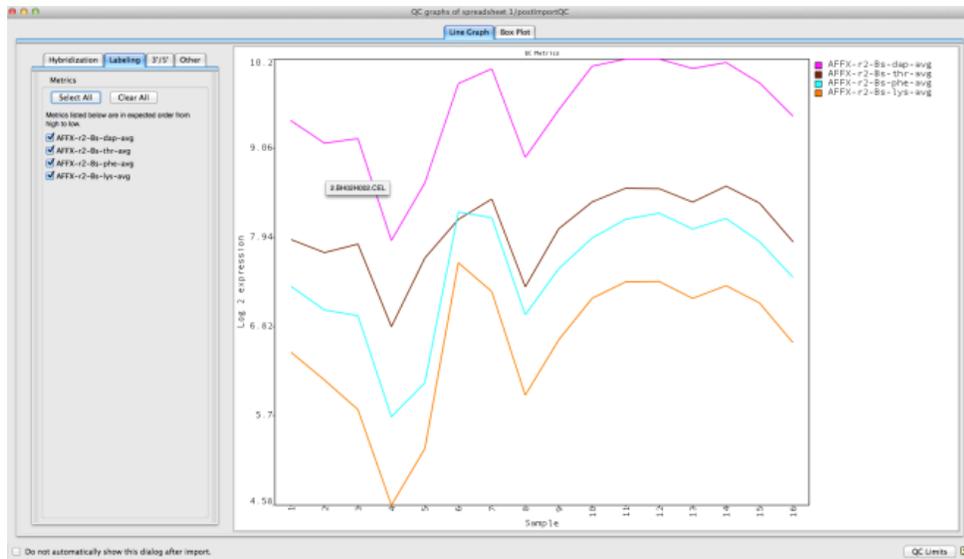
QC Metrics generated using Partek® Genomics Suite software version 6.6.

Quality control steps were performed to ensure quality of normalized and summarized data to enable comparison of HME50 cell line samples hybridized to multiple microarrays. The gene expression profiles for HME50, HME50-5E, HME50-hTERT and HMET cell lines were analyzed using Partek® software, version 6.6 Copyright © 2012 Partek Inc., St. Louis, MO, USA. To assess gene expression, raw Affymetrix CEL files for each sample (BH02H001.CEL through BH02H016.CEL) were imported using RMA import settings and all files were saved in specified output folder. After CEL files are imported, quality control (QC) metrics are generated and quality assessment results displayed in postImportQC spreadsheet.

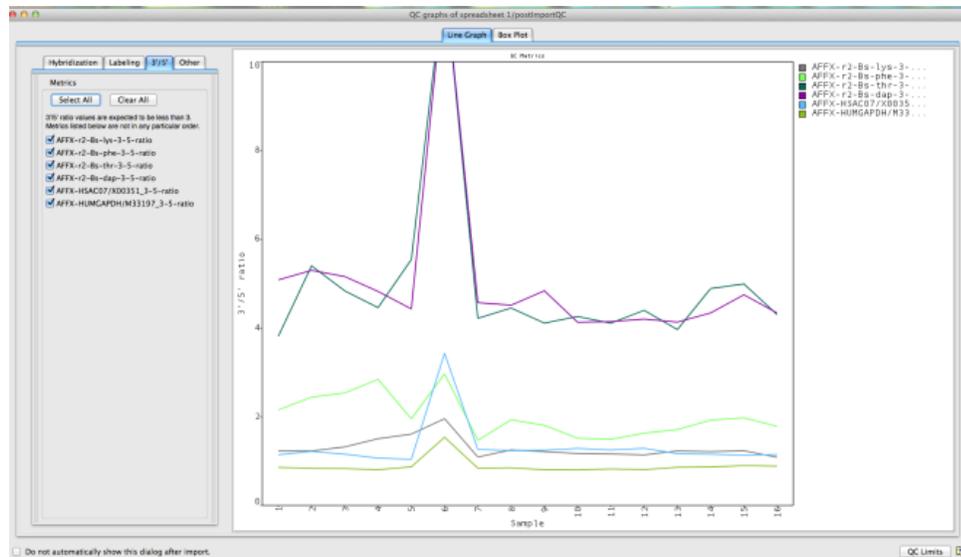
The QC metrics categories include graphs for hybridization metrics depicted in **Figure S.1.** (log expression signal box plot generated from normalized and summarized probe set signal values), labeling depicted in **Figure S.2.**, and 3'/5' values (3'/5' ratios of labeling controls and GAPDH and beta-actin) depicted in **Figure S.3.** that provide QC information from control and experimental probes. The hybridization and labeling control metrics are listed and observed in expected order from high to low.



**Figure S.1. PGS 6.6 Hybridization Control.** The hybridization controls are in the expected order from high to low: AFX-rs-P1-cre-avg > AFX-r2-Ec-bioD-avg > AFX-r2-EcbioC-avg > AFX-r2-Ec-bioB-avg. All 16 samples are on X axis in same order as import and the y-axis represents log<sub>2</sub> values of the normalized spike in control probe-sets.



**Figure S.2. PGS 6.6 Labeling Control.** The Labeling controls line graph shows that the four unlabeled polyA controls spiked in increasing concentrations of Lys<Phe<Thr<Dap into the sample prior to labeling are detected in expected order with Lys as lowest to Dap as highest concentration.



**Figure S.3. 3'/5' ratio plot.** PGS 6.6 calculates and plots the 3'/5' ratio of GAPDH, beta-actin and all labeling controls used and displays under the 3'/5' tab after import of the arrays. The 3'/5' ratio compares the abundance of the signal at the 3' end over the abundance at the 5' end and a ratio of 3 or less is acceptable.

Sample information was added by creating a new column with categorical attribute for cell lines (factor) and quadruplicates were grouped as HME50 (BH02H001.CEL to BH02H004.CEL), HME50-5E (BH02H005.CEL to BH02H008.CEL), HME50-hTERT (BH02H009.CEL to BH02H012.CEL), HMET (BH02H013.CEL to BH02H016.CEL). Principal component analysis and hierarchical clustering was then performed following the QC metrics analysis and preliminary exploratory analysis.