Detection call algorithms for high-throughput gene expression microarray data

Kellie J. Archer and Sarah E. Reese

Submitted: 4th September 2009; Received (in revised form): 13th October 2009

Abstract
Extensive methodological research has been conducted to improve gene expression summary methods. However, in addition to quantitative gene expression summaries, most platforms, including all those examined in the MicroArray Quality Control project, provide a qualitative detection call result for each gene on the platform. These detection call algorithms are intended to render an assessment of whether or not each transcript is reliably measured. In this paper, we review uses of these qualitative detection call results in the analysis of microarray data. We also review the detection call algorithms for two widely used gene expression microarray platforms, Affymetrix GeneChips and Illumina BeadArrays, and more clearly formalize the mathematical notation for the Illumina BeadArray detection call algorithm. Both algorithms result in a $P$-value which is then used for determining the qualitative detection calls. We examined the performance of these detection call algorithms and default parameters by applying the methods to two spike-in datasets. We show that the default parameters for qualitative detection calls yield few absent calls for high spike-in concentrations. When genes of interest are expected to be present at very low concentrations, spike-in datasets can be useful for appropriately adjusting the tuning parameters for qualitative detection calls.

Keywords: gene expression; microarray; detection calls; present

INTRODUCTION
After hybridization, washing, and scanning a microarray, a gene expression summary method is applied to the normalized intensities. These gene expression summaries are subsequently analyzed using statistical methods for performing gene-level class comparisons, deriving a phenotypic classifier, and in clustering applications. Commercially produced arrays provide software for calculating gene expression summaries, such as the MAS5 method for Affymetrix GeneChips [1]. Most research seeking to improve output from gene expression platforms has therefore focused on developing improved gene expression summary methods, some examples include the robust multiarray average (RMA) [2], GC-RMA [3] or the Model Based Expression Index (MBEI) [4] methods. However, a question frequently asked of a microarray experiment is, ‘Is the gene present in the given sample?’ Most platforms, including all those examined in the MicroArray Quality Control project, provide a qualitative detection call result for each gene on the platform [5]. These detection call algorithms are intended to render an assessment of whether or not each transcript is reliably measured. Since inclusion of transcripts not reliably measured in statistical analyses yields results not easily interpreted, a frequent use of the results from detection call algorithms is for dimension reduction or filtering [5–11]. Interestingly, there is no consistent filtering method being applied. Some filter genes called Absent on all arrays [6, 9]; others have recommended filtering by fraction...
present, which can be filtering out probe sets not called Present among at least 50% of the samples in one treatment group for small sample sizes, or for larger experiments, filtering out probe sets not called Present among at least 25% of the samples in one treatment group [8]. Regardless, filtering on detection call has been demonstrated to increase both the number of differentially expressed genes detected [6] and reduce the number of false positive findings [7, 8, 10]. Moreover, even when performing cross-platform comparisons, cross-platform correlation increased as the percent present call filter was increased [12].

Aside from filtering, detection call results have been used in a variety of other ways. Detection calls have been used for estimating the percent of genes present in a given hybridization, which is commonly used as a quality control measure [9–11, 13–18]. When evaluating replicate arrays, call concordance has been used as a measure of reproducibility, defined for two replicate hybridizations as the sum of probe sets called Absent, Marginally Present, or Present on both arrays [13]. In a study comparing two different microarray platforms, to obtain one detection call result per gene for each platform, Fisher’s method for combining \( P \)-values was applied to the detection call \( P \)-values from five samples hybridized to both platforms. Thereafter, the number of concordantly present and absent genes on both platforms served as a measure of cross-platform reproducibility [19]. In another study that compared the equivalence between use of standard poly(RNA) and total RNA as the starting template for subsequent cDNA reactions, detection call concordance among the hybridized microarrays was examined as the primary outcome [20].

Detection call \( P \)-values have also been used to weight gene expression values when clustering samples [21]. Specifically, for five different expression summary algorithms, the investigators examined whether using weighted or unweighted Pearson’s correlation in agglomerative hierarchical clustering, with the weights defined as 1-Detection Call \( P \)-value, improved the performance. They found that inclusion of detection call \( P \)-value weights improved the performance of the clustering result for MBEI, Probe Profiler, and RMA expression summaries.

Others have used detection call results in class comparisons and class predictions. For example, investigators identified estrogen-regulated genes by calculating a confidence score for each gene, of which one of the four components was a score based on the percent of samples having a present call for the given gene [22]. A similar method was used for identifying genes regulated by selective estrogen receptor modulators [23]. Detection call results have also been directly used in deriving a phenotypic classifier, which had better accuracy than a classifier derived using the gene expression summaries and the Prediction Analysis for Microarray algorithm [24]. These authors concluded that detection calls are especially useful for deriving a classifier when the study includes arrays of multiple chip formats, because the detection calls do not require any specific normalization.

In this paper, we review the detection call algorithms for two widely used gene expression microarray platforms, Affymetrix GeneChips and Illumina BeadArrays. Both algorithms result in a \( P \)-value which is then used for determining the detection call rate. Because two possible errors can be made when declaring a gene to be present or absent using these detection call algorithms, namely, one can declare a gene to be present when in fact it is absent (Type I Error) or declare a gene to be absent when in fact it is present (Type II Error), we examined the performance of these detection call algorithms using two low level spike-in datasets.

**DETECTION CALL ALGORITHMS**

**Affymetrix detection call algorithm**

An Affymetrix GeneChip is characterized as being composed of several perfect match (PM) and their corresponding mismatch (MM) probes that interrogate for a single gene. The PM is the exact complementary sequence of the target genetic sequence, composed of 25 base pairs. Each PM probe has a corresponding MM probe, which has the same sequence with exception that the middle base (13th) position is complementary to that in the PM [25]. The underlying idea behind the MM probes is to account for background. The set of PM/MM pairs that interrogate for a specific gene is called a probe set. As part of the output for an Affymetrix GeneChip, the GeneChip Operating Software includes a column labeled ‘Detection call’ that indicates whether the probe set is present (P), marginally present (M) or absent (A) in the given sample. The original detection call algorithm was based on a decision tree, applied to two
GeneChip outputs: the number of times the PM intensity exceeded the MM intensity for a given probe set and the average log(PM/MM) signal [25]. Some investigators noted deficiencies with the first rendition of the detection call algorithm and proposed an alternative detection call algorithm based on a permutation test [26], which was implemented in the dChip user-developed software [27]. With the release of Affymetrix’s Microarray Suite Software (MAS) version 5 came the current implementation of Affymetrix’s detection call algorithm [28]. It is based on a hypothesis test using the PM and MM intensities for the probe set and presumes the MM intensities are accurate estimates of gene-specific background. The Affymetrix detection call algorithm declares a probe set to be present, marginally present, or absent by using the Wilcoxon signed-rank test in the following manner [28]. First, probe pairs that have saturated MM intensities (intensity ≥46 000) are discarded. Additionally, probe pairs where PM and MM are within τ of each other are discarded. If all probe pairs for a probe set are saturated, the probe set is automatically labeled ‘Present’ and the Detection call P-value set to 0. Thereafter, for probe set i, using each of its remaining probe pairs j, a discrimination score \( R_{ij} = (PM_{ij} - MM_{ij}) / (PM_{ij} + MM_{ij}) \) is calculated and the Wilcoxon Signed Rank test is conducted for the probe set with the null and alternative hypotheses specified as

- \( H_0: \text{median}(R_{ij}) = {\tau} \)
- \( H_A: \text{median}(R_{ij}) > {\tau} \) (corresponds to testing the gene is present)

where the default value for τ is 0.015, which was empirically selected from a Latin Square experiment but is a user-adjustable parameter [28].

The decision rule used to declare genes as present, absent, or marginally present are based upon the resulting P-values using the following thresholds: if \( P < \alpha_1 \) then the probe set is declared present; if \( \alpha_1 < P < \alpha_2 \) the probe set if declared marginally present; if \( P > \alpha_2 \) the probe set is declared absent. The original defaults set \( \alpha_1 = 0.04 \) and \( \alpha_2 = 0.06 \); current defaults in the GeneChip Operating Software are \( \alpha_1 = 0.05 \) and \( \alpha_2 = 0.065 \).

**Illumina detection call algorithm**
The Illumina gene expression platform is a bead-based microarray technology whereby for each gene interrogated, a 50 nt length oligo interrogating a specific gene is attached to a bead, and multiple beads (~30) of the same type are randomly included on the array as a means to more precisely measure that gene’s expression [29]. For each hybridized BeadArray, the output from the Illumina GenomeStudio software includes for each beadtype a column labeled ‘Detection.Pval’ [30]. In calculating the Illumina detection P-value, let \( i \) index the hybridized BeadArrays (samples), \( g \) index beadtypes or genes, and \( b_{ig} \) represent expression for beadtype \( g \) for the \( i \)th BeadArray. Illumina includes control features on each BeadArray, such as negative control beadtypes which are randomly permuted sequences known not to exist in the given organism’s genome. Let the \( N \) negative control beadtypes be represented by \( b_{ig}^{neg} \), such that the mean of the negative control beadtypes for BeadArray \( i \) can be given by

\[
\mu_{ig}^{neg} = \frac{1}{N} \sum_{j=1}^{N} b_{ig}^{neg}
\]

Similarly, the standard deviation (\( \sigma_{ig}^{neg} \)) of the negative control beadtypes for BeadArray \( i \) is calculated. Thereafter, a z-score is calculated for each gene on BeadArray \( i \) as

\[
Z_{ig} = \frac{b_{ig} - \mu_{ig}^{neg}}{\sigma_{ig}^{neg}}
\]

These \( g \) z-scores are then ranked using a specialized function

\[
R_{ig} = \begin{cases} 
0 & \text{if } Z_{ig} < \min_{g \in \text{neg}} \{Z_{ig}\} \\
\text{rank}(Z_{ig}) & \text{if } \min_{g \in \text{neg}} \{Z_{ig}\} \leq Z_{ig} \leq \max_{g \in \text{neg}} \{Z_{ig}\} \\
N & \text{if } Z_{ig} > \max_{g \in \text{neg}} \{Z_{ig}\}
\end{cases}
\]

Thereafter, the detection call P-value for gene \( g \) on BeadArray \( i \) is given by

\[
DPV_{ig} = 1 - \frac{R_{ig}}{N}.
\]

Although the Illumina BeadStudio and GenomeStudio software does not provide a qualitative detection call, it has been recommended that an alpha level such as 0.05 or 0.01 can be used for making Present/Absent calls [31].

**METHODS**

**Affymetrix low level spike-in dataset**
Previously, the detection call algorithm and its user-specified parameters were applied to a Latin Square experiment where genes were spiked in at concentrations of 0, 0.25, 0.5, 1, 2, 4, 8, 16, 32, 64,
the hypothesis that the Kolmogorov-Smirnov test was performed to test
dition) should follow a uniform distribution, a
for the 0 spike-in concentration (the null con-
marginal concentration was calculated.
Moreover, because the distribution of 
list in Supplementary Table S1 were extracted.
The present/marginally present/absent call frequen-
caries for each spike-in concentration was uniform.

<table>
<thead>
<tr>
<th>GeneChip</th>
<th>Spike group</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>3823p_P1a.cel</td>
<td></td>
<td>0</td>
<td>0.75</td>
<td>1.5</td>
<td>3</td>
</tr>
<tr>
<td>3823p_P1b.cel</td>
<td></td>
<td>0</td>
<td>0.75</td>
<td>1.5</td>
<td>3</td>
</tr>
<tr>
<td>3823p_P1c.cel</td>
<td></td>
<td>0</td>
<td>0.75</td>
<td>1.5</td>
<td>3</td>
</tr>
<tr>
<td>3823p_P2a.cel</td>
<td></td>
<td>3</td>
<td>0</td>
<td>0.75</td>
<td>1.5</td>
</tr>
<tr>
<td>3823p_P2b.cel</td>
<td></td>
<td>3</td>
<td>0</td>
<td>0.75</td>
<td>1.5</td>
</tr>
<tr>
<td>3823p_P2c.cel</td>
<td></td>
<td>3</td>
<td>0</td>
<td>0.75</td>
<td>1.5</td>
</tr>
<tr>
<td>3823p_P3a.cel</td>
<td></td>
<td>1.5</td>
<td>3</td>
<td>0</td>
<td>0.75</td>
</tr>
<tr>
<td>3823p_P3b.cel</td>
<td></td>
<td>1.5</td>
<td>3</td>
<td>0</td>
<td>0.75</td>
</tr>
<tr>
<td>3823p_P3c.cel</td>
<td></td>
<td>1.5</td>
<td>3</td>
<td>0</td>
<td>0.75</td>
</tr>
<tr>
<td>3823p_P4a.cel</td>
<td></td>
<td>0.75</td>
<td>1.5</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>3823p_P4b.cel</td>
<td></td>
<td>0.75</td>
<td>1.5</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>3823p_P4c.cel</td>
<td></td>
<td>0.75</td>
<td>1.5</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

Illumina spike-in dataset
A spike-in dataset from a previously published study was used to investigate methods for background esti-
mation, bead-level summarization, and differential
expression analysis for Illumina BeadArray data [33].
These data were subsequently used for identifying an appropriate variance-stabilizing transformation
[34]. However, the investigators did not examine the sensitivity and specificity of the detection call
P-values, so we have re-analyzed the Dunning
spike-in dataset here. Briefly, transcripts correspond-
ing to 33 beadtypes were spiked into the back-
ground sample and hybridized to Mouse-6 version
BeadChips.

The Affymetrix detection call algorithm was applied using the GCOS defaults of
α_1 = 0.05 and
α_2 = 0.065 and the detection calls for the probe sets
listed in Supplementary Table S1 were extracted. The present/marginally present/absent call frequen-
cies for each spike-in concentration was calculated.
Moreover, because the distribution of P-values
for the 0 spike-in concentration (the null con-
tion) should follow a uniform distribution,
a Kolmogorov-Smirnov test was performed to test
the hypothesis that the P-value distribution for the
0 spike-in concentration was uniform.

As a comparison to the algorithmic-based Illumina detection call method, we additionally
applied a two-sample t-test comparing mean bead-
level expression for each beadtype to the mean
expression of the negative control bead-level data.
Among the negative control beads outlier beads
were defined as those having an intensity exceeding
the median intensity + 3 × median absolute deviation
and were removed from the detection call
analysis. Again, letting i index the hybridized
BeadArrays, g index beadtypes, k index the bead-level
data within each beadtype such that
b_{igk} represents
the bead-level expression for bead k within beadtype
for the i-th BeadArray. For negative control bead-
types the bead-level expression is represented by
b_{igneg}.

For each beadtype g on BeadArrays i, a two-sample

The t-test was applied to statistically compare its mean bead-level expression $\left(1/K \sum_{k=1}^{K} b_{ik}\right)$ to the mean bead-level expression among the negative control bead types $\left(1/\left(\sum_{g=1}^{G} \sum_{k=1}^{K} b_{igk}\right)\right)$ resulting in one $P$-value per bead type ($p_{ig}$). Bead-level data can be analyzed using the beadarray package [35] in the R programming environment [36].

**RESULTS**

**Affymetrix low level spike-in dataset**

The frequencies and percent of detection calls for each spike-in concentration (Table 2) reveals that the Affymetrix algorithm had a 2.6–5.1% error rate for the 0, 1.5 and 3 pM concentrations and a 15.4% error rate for the 0.75 pM concentration, for an overall error rate of 6.73%. A classification tree predicting 0 versus non-zero spike concentration using the detection call $P$-values selected 0.08429 as the optimal $P$-value cutpoint for declaring a probe set present/absent. Application of this cutpoint for declaring the spike-in probe sets either present or absent resulted in a 4.1% error rate (Table 3). Boxplots of the detection call $P$-values by spike-in concentration revealed that the distribution of $P$-values for the null condition seem to depart from a uniform distribution (Figure 1). This was confirmed by application of the Kolmogorov-Smirnov test which resulted in a $P = 0.0498$, indicating the $P$-values were not uniformly distributed.

The sensitivity of the detection call result was investigated in a number of ways. First, we note that none of the probe-level intensities for any of the spiked-in probe sets were saturated and only 6 of the 3432 spiked-in probe pairs among the 12 GeneChips were discarded for the PM and MM being within $\tau$ of one another. Including these probe pairs and re-performing the Wilcoxon signed rank test resulted in slightly different $P$-values for these six probe sets but no change in the detection call, therefore, the filtering procedure did not affect the detection call results. Probes within a probe set are ordered by their probe interrogation position with the _at1 probe being most 5’ and the _at11 probe being the most 3’. Therefore, if the RNA was degraded, probe intensities are expected to vary systematically by probe interrogation position [37]. To investigate whether there was a dependence of $R_{ij}$ on probe position $j$, for each spiked-in probe set we plotted the $R_{ij}$ values against probe position $j$ for all 12 GeneChips (data not shown, figures freely available upon request). We did not observe any dependence of $R_{ij}$ upon $j$. Researchers recently identified that intensities of probes containing multiple runs of guanines were discordant with respect to intensities of neighboring probes within the same probe set [38], a finding that was subsequently empirically confirmed by

<table>
<thead>
<tr>
<th>Spike concentration</th>
<th>Absent</th>
<th>Marginally present</th>
<th>Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 pM</td>
<td>74 (94.9%)</td>
<td>0</td>
<td>4 (5.1%)</td>
</tr>
<tr>
<td>0.75 pM</td>
<td>10 (12.8%)</td>
<td>2 (2.6%)</td>
<td>66 (84.6%)</td>
</tr>
<tr>
<td>1.5 pM</td>
<td>1 (1.3%)</td>
<td>1 (1.3%)</td>
<td>76 (97.4%)</td>
</tr>
<tr>
<td>3 pM</td>
<td>2 (2.6%)</td>
<td>1 (1.3%)</td>
<td>75 (96.2%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Spike-in concentration</th>
<th>0 pM</th>
<th>Non-zero concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absent</td>
<td>71</td>
<td>6</td>
</tr>
<tr>
<td>Present</td>
<td>7</td>
<td>228</td>
</tr>
</tbody>
</table>

**Figure 1:** Boxplots of detection call $P$-values by spike-in concentration for Affymetrix low level spike-in dataset. Dashed line represents $\alpha_1 = 0.05$ while the dotted line represents $\alpha_2 = 0.065$.  

---

**Table 2:** Frequency and percent of detection calls by spike-in concentration for the Affymetrix low level spike-in dataset

<table>
<thead>
<tr>
<th>Spike concentration</th>
<th>Absent</th>
<th>Marginally present</th>
<th>Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 pM</td>
<td>74 (94.9%)</td>
<td>0</td>
<td>4 (5.1%)</td>
</tr>
<tr>
<td>0.75 pM</td>
<td>10 (12.8%)</td>
<td>2 (2.6%)</td>
<td>66 (84.6%)</td>
</tr>
<tr>
<td>1.5 pM</td>
<td>1 (1.3%)</td>
<td>1 (1.3%)</td>
<td>76 (97.4%)</td>
</tr>
<tr>
<td>3 pM</td>
<td>2 (2.6%)</td>
<td>1 (1.3%)</td>
<td>75 (96.2%)</td>
</tr>
</tbody>
</table>

**Table 3:** Results from decision rule applied to the detection call $P$-values

<table>
<thead>
<tr>
<th>Spike-in concentration</th>
<th>Absent</th>
<th>Non-zero concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absent</td>
<td>71</td>
<td>6</td>
</tr>
<tr>
<td>Present</td>
<td>7</td>
<td>228</td>
</tr>
</tbody>
</table>

---

**Figure 1:** Boxplots of detection call $P$-values by spike-in concentration for Affymetrix low level spike-in dataset. Dashed line represents $\alpha_1 = 0.05$ while the dotted line represents $\alpha_2 = 0.065$.  

---

**Table 2:** Frequency and percent of detection calls by spike-in concentration for the Affymetrix low level spike-in dataset

<table>
<thead>
<tr>
<th>Spike concentration</th>
<th>Absent</th>
<th>Marginally present</th>
<th>Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 pM</td>
<td>74 (94.9%)</td>
<td>0</td>
<td>4 (5.1%)</td>
</tr>
<tr>
<td>0.75 pM</td>
<td>10 (12.8%)</td>
<td>2 (2.6%)</td>
<td>66 (84.6%)</td>
</tr>
<tr>
<td>1.5 pM</td>
<td>1 (1.3%)</td>
<td>1 (1.3%)</td>
<td>76 (97.4%)</td>
</tr>
<tr>
<td>3 pM</td>
<td>2 (2.6%)</td>
<td>1 (1.3%)</td>
<td>75 (96.2%)</td>
</tr>
</tbody>
</table>

**Table 3:** Results from decision rule applied to the detection call $P$-values

<table>
<thead>
<tr>
<th>Spike-in concentration</th>
<th>Absent</th>
<th>Non-zero concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absent</td>
<td>71</td>
<td>6</td>
</tr>
<tr>
<td>Present</td>
<td>7</td>
<td>228</td>
</tr>
</tbody>
</table>
an analysis of 6685 Affymetrix GeneChip CEL files [39]. Therefore, to investigate whether runs of guanine were associated with the $R_{ij}$ for each probe sequence the maximum guanine run length was obtained. Among the probe sequences associated with the 26 spiked-in probe sets, there were 91, 127, 49, 12, 6 and 1 probe sequences having maximum guanine runs of length 1, 2, 3, 4, 5 and 6. No probe sequence had a guanine run greater than length 6. Thereafter, for each hybridized GeneChip the $R_{ij}$’s for probes associated with the spiked-in probe sets were plotted against the maximum guanine run length (data not shown, figures freely available upon request). No association between $R_{ij}$ and maximum guanine run length was observed for these probe sets. Finally, we note that replicate hybridizations yielded consistent results (Supplementary Material, Figures 1–4).

**Illumina spike-in dataset**

The frequencies of detection calls for each spike-in concentration reveals that the Illumina algorithm had a 4.5% error rate for the 0 pM concentration when using $\alpha = 0.01$ and a 7.6% error rate with using $\alpha = 0.05$ (Table 4). However, 93.9, 92.4, 79.5 and 28.8% of transcripts spiked at 0.01, 0.03, 0.1 and 0.3 pM concentrations were declared Absent at the $\alpha = 0.01$ level. These percentsl slightly decline when the $\alpha$-level was adjusted to 0.05. Boxplots of the Illumina detection call P-values by spike-in concentration appear in Figure 2 (left panel). As seen in Table 4 and the left panel of Figure 2, none of the transcripts spiked in at 3 pM or greater were called absent and 2.3% of transcripts spiked in at 1 pM were called absent when using $\alpha = 0.05$. Therefore, the Illumina detection call algorithm accompanied by the recommended qualitative thresholds performs well when the transcript is either truly absent or present at a concentration at 1–3 pM or greater. However, the assay had low sensitivity at the spike-in concentrations of 0.01, 0.03, 0.1 and 0.3 pM.

When using the two-sample t-test and comparing mean bead-level expression for each beadtype to the mean bead-level expression among all negative control beadtypes, there were more transcripts called present for the 0 pM spiked transcripts compared to the Illumina Detection method, but fewer absent calls for non-zero spiked transcripts (Table 5). In fact, for the 0.1 pM concentration, half as many beadtypes were declared absent using the two sample t-test compared to using the default Illumina detection call methodology. Boxplots of the t-test P-values by spike-in concentration appear in Figure 2 (right panel). These detection call P-values were consistent among replicate hybridizations as well (Supplementary Material Figures 5–10). Again, as seen in Table 5 and the right panel of Figure 2 none of the transcripts spiked in at 3 pM or greater were called absent and 0.8% of transcripts spiked in at 1 pM were called absent when using $\alpha = 0.05$. Therefore, both methods are able to identify transcripts spiked-in at high concentrations as present.

**CONCLUSION**

In this article, we reviewed various uses of detection call results in microarray experiments. We also reviewed the detection call algorithms for two widely used commercial platforms, the Affymetrix GeneChip and Illumina BeadArray. We then applied the two detection call algorithms to two spike-in datasets for evaluating the performance of the default/recommended parameters. We conclude that the default settings are likely to work well in applications where genes of interest are expected to be present at 1 pM concentration or higher. Genes present in low abundance, such as some transcription factors, may be routinely labeled as absent when using the Illumina detection call algorithm since...
the assay had low sensitivity for the 0.01, 0.03, 0.1 and 0.3 pM spike-in concentrations. Since both platforms include user-adjustable parameters, particularly the α-level used in making a qualitative statement of whether a given transcript is present or absent, we recommend that the tuning parameters be adjusted depending upon whether the investigator wishes to filter out all transcripts truly absent, thereby potentially removing some low level transcripts that are truly present, versus retaining as many transcripts that may be present. Consideration of how to appropriately adjust the tuning parameters can be derived from application of supervised learning methods, such as classification trees, to detection call $P$-values in a low-level spike-in experiment.

### Key Points
- The default parameters for qualitative detection calls yield few absent calls for high spike-in concentrations.
- When genes of interest are expected to be present at very low concentrations, spike-in datasets can be useful for appropriately adjusting the tuning parameters for qualitative detection calls.
- Alternative, statistically-based detection call methods such as a two-sample $t$-test comparing gene intensities to negative control expression may result in improved performance compared to ad hoc detection call algorithms.
Table 5: When comparing bead-level intensities for each beadtype to the bead-level intensities of negative control beadtypes using a two-sample t-test, frequency and percent of detection calls by spike-in concentration for the Illumina spike-in dataset using Illumina recommended thresholds of $x = 0.01$ and $x = 0.05$

<table>
<thead>
<tr>
<th>Spike concentration</th>
<th>$x = 0.01$</th>
<th>$x = 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Absent</td>
<td>Present</td>
</tr>
<tr>
<td>0 pM</td>
<td>103 (78.0%)</td>
<td>29 (22.0%)</td>
</tr>
<tr>
<td>0.01 pM</td>
<td>97 (73.5%)</td>
<td>35 (26.5%)</td>
</tr>
<tr>
<td>0.03 pM</td>
<td>88 (66.7%)</td>
<td>44 (33.3%)</td>
</tr>
<tr>
<td>0.1 pM</td>
<td>52 (39.4%)</td>
<td>80 (60.6%)</td>
</tr>
<tr>
<td>0.3 pM</td>
<td>14 (10.6%)</td>
<td>118 (89.4%)</td>
</tr>
<tr>
<td>1 pM</td>
<td>3 (2.3%)</td>
<td>129 (97.7%)</td>
</tr>
<tr>
<td>3 pM</td>
<td>0</td>
<td>132 (100%)</td>
</tr>
<tr>
<td>10 pM</td>
<td>0</td>
<td>132 (100%)</td>
</tr>
<tr>
<td>30 pM</td>
<td>0</td>
<td>132 (100%)</td>
</tr>
<tr>
<td>100 pM</td>
<td>0</td>
<td>132 (100%)</td>
</tr>
<tr>
<td>300 pM</td>
<td>0</td>
<td>132 (100%)</td>
</tr>
<tr>
<td>1000 pM</td>
<td>0</td>
<td>132 (100%)</td>
</tr>
</tbody>
</table>

SUPPLEMENTARY DATA
Supplementary data are available online at http://bib.oxfordjournals.org/.

Acknowledgements
The authors would like to thank Suzanne Dee from Affymetrix for making the Affymetrix low level spike-in dataset available.

References
24. Reme T, Hose D, De Vos J, et al. A new method for class prediction based on signed-rank algorithms applied to...


