Pathway-directed weighted testing procedures for the integrative analysis of gene expression and metabolomic data

Laila M Poisson\textsuperscript{a}, Arun Sreekumar\textsuperscript{b}, Arul M Chinna\textsuperscript{c}, and Debashis Ghosh\textsuperscript{d}

\textsuperscript{a}Department of Public Health Sciences, Henry Ford Hospital, Detroit, MI
\textsuperscript{b}Medical College of Georgia Cancer Center, Medical College of Georgia, Augusta, GA
\textsuperscript{c}Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, MI
\textsuperscript{d}Departments of Statistics and Public Health Sciences, Penn State University, University Park, PA

Abstract

We explore the utility of p-value weighting for enhancing the power to detect differential metabolites in a two-sample setting. Related gene expression information is used to assign an a priori importance level to each metabolite being tested. We map the gene expression to a metabolite through pathways and then gene expression information is summarized per-pathway using gene set enrichment tests. Through simulation we explore four styles of enrichment tests and four weight functions to convert the gene information into a meaningful p-value weight. We implement the p-value weighting on a prostate cancer metabolomics dataset. Gene expression on matched samples is used to construct the weights. Under certain regulatory conditions, the use of weighted p-values does not inflate the type I error above what we see for the un-weighted tests except in high correlation situations. The power to detect differential metabolites is notably increased in situations with disjoint pathways and shows moderate improvement, relative to the proportion of enriched pathways, when pathway membership overlaps.

1. Background

There is currently an explosion of high-throughput technologies for assessing global snapshots of the molecular behavior of cells. Assays exist for measuring DNA sequence, copy number, mRNA transcript levels, protein presence and abundance, as well as metabolite abundance in biological samples of healthy and diseased tissues \cite{1, 2}. In the cancer setting, assessment of gene expression is a prominent tool. Gene expression microarrays have been a mainstay in gene expression for many years and can measure over...
40,000 known and estimated genes and gene elements per array. New sequencing
technologies, which can be used to measure mRNA abundance, are not limited by an array
design and can also measure many thousands of mRNA elements in a single experiment.

A less commonly used platform has been that for measuring metabolites [3]. The
metabolome is the cellular complement of small molecules and includes amino acids, fatty
acids, simple carbohydrates, and exogenous drugs within the cell. Metabolomic profiles are
generated by mass spectrometry (MS) preceded by either gas or liquid chromatography (GC
or LC). Metabolic profiling can currently detect hundreds of metabolites, compared to the
tens of thousands of mRNA species measured.

The motivating example in this paper considers data from a prostate cancer study in which
both gene expression and metabolites were measured on the same set of subjects [4]. In
particular, samples come from two groups: cancer and benign. The intergration of such
molecular data sources for exploration and discovery is of growing interest [5, 6, 7]. A
variety of visualization tools have recently been introduced that can map metabolite and
gene expression information onto metabolic pathways, such as [8, 9, 10]. A few have some
analysis capability such as correlation measures [9] or enrichment analysis [8].

While there have been previous studies in which metabolite and gene expression data have
been simultaneously measured (e.g., see [11, 12, 13, 14, 15]), much of this work has been in
model organisms and the statistical techniques used have been based on clustering and
pathway analysis. By contrast, in this work, our scientific goal centers on identification of
differential molecules, which represents a more supervised analysis. Work in this area is has
been limited, though sparse regression methods appear to have promise [16, 17].

We use pathway based procedures to integrate the information from the two platforms and
we utilize the theory of weighted multiple testing as developed by Genovese et al. (2006)
[18]. We review this work in section 1.1. While the theory proposed there is general, it has
not been applied to the particular problem considered here in which the two element types
do not map in a one-to-one fashion. In this paper, we develop practically useful weighting
methods for the specific problem of integrating gene expression into analysis of metabolite
differential intensity. A key assumption that will be used is that molecules act in pathways.
The Kyoto Encyclopedia for Genes and Genomes [19, 20, 21] will be used to define
pathways to construct weights. In the section 2 we describe the weighting methods that we
explored. Two simulations are employed to study these weights in sections 3.1 and 3.2 and
we apply our findings to the motivating data example in Section 3.3. We conclude in Section
4 with some discussion and recommendations.

1.1. Weighted multiple testing and applications to genomic analysis

Consider a set of \( m \) tests in which a positive constant weight, \( w_i \), is applied to each p-value,
\( P_i \) for test \( i = 1, \ldots, m \), according to its perceived importance. Holm showed that the
sequentially rejective Bonferroni test (a.k.a. Holm’s Test [22]) could be generalized to use
the new p-value, \( P_i^* = P_i/w_i \), to assess significance. Specifically, if the ordered weighted p-
values are written as \( P_{(1)}^* \leq \cdots \leq P_{(m)}^* \) then we reject \( H_{(i)} \) when \( P_{(j)}^* \leq \alpha j \left( \sum_{k=j}^{m} w_k \right) \), \( j = 1, \ldots, i \) where \( H_{(i)} \) and \( w_{(i)} \) are the hypothesis and weight associated with weighted p-value
\( P_{(i)}^* \). This weighted test is designed to control the family-wise error rate (FWER), i.e. the
probability that at least one null hypothesis is falsely rejected, and does not require that the
multiple tests be independent. Holm’s only requirement is that \( w_i \geq 0 \).

weighting with application to genomics studies. Instead of Holm’s test, they use the
Benjamini-Hochberg (BHT) step-down test [23] as the basis of their method. With the BHT, $H_{0(1)}, \ldots, H_{0(i)}$ are rejected for $\max_{i=1, \ldots, m} \{P_i : P_i \leq \alpha i/m\}$ where $H_{0(i)}$ is the hypothesis associated with ordered p-value $P_i$. The BHT controls the false discovery rate (FDR), i.e. the expected rate of incorrectly rejected null hypotheses among all rejected null hypotheses. For testing hundreds, or thousands, of hypotheses controlling the FDR is less conservative than controlling the FWER. Genovese et al. (2006) [18] also define a weight $w_i \geq 0$ for test $i$ resulting in the weighted p-value, $P_i^w = P_i / w_i$. Again there is no requirement on the independence of the tests. However, they allow that the set of weights $W = \{w_1, \ldots, w_m\}$ be random variables and they additionally require that $\sum_{i=1}^{m} w_i = 1$ to maintain control of the FDR. Then the weighted-BHT rejects the null hypotheses, $H_{0(1)}, \ldots, H_{0(i)}$ for $\max_{i=1, \ldots, m} \{P_i^w : P_i^w \leq \alpha i/m\}$, where $H_{0(i)}$ is the null hypothesis associated with ordered weighted p-value $P_i^w$, while controlling the FDR at level $\alpha$.

Roeder et al. (2006) [24] applied the weighted-BHT to large-scale genomic studies where thousands of tests are performed and controlling the error rates leads to a loss of power. In the context of metabolomic profiling we have hundreds of tests, not thousands. In this setting the low power to detect differential metabolites can arise in part from small sample sizes [4], as well as from increased variability due to factors such as diet [25, 26] and diurnal rhythm [27]. Yet we know that metabolites do not act in isolation within the cell. Here we use p-value weights to add information about the behavior of other molecular components such as gene transcripts in an effort to add power to nominate metabolites of interest. Through simulation we find that the power can be improved without raising the Type I error rates above the levels seen in unweighted tests, in the absence of error rate controlling methods such as Holm’s Test or the BHT.

2. Methods

We define genes $\gamma_g (g = 1, \ldots, G)$ which have been tested for differential ability, say between cases ( $\psi = 1$) and controls ( $\psi = 0$), producing the test statistics $T_{g}^\psi (g=1, \ldots, G)$. Additionally, we define pathways $\xi_k (k = 1, \ldots, K)$ and classify each gene by its differential ability ( $|T_{g}^\psi|\text{large}$) and inclusion in the pathway $\xi_k (\gamma_g \in \xi_k)$ according to Table 1. In section 2.1 we describe four pathway enrichment tests and identify the per-pathway test statistics $S_k$ they define. In sections 2.2.1 and 2.3 we describe simulation models that explore the utility of these test statistics, $S_k$, as weights for tests of per-metabolite differential ability.

2.1. Enrichment test methods

As described in Table 2 we consider tests based on either binary or continuous differential expression results. Specifically, binary tests require that the per-gene tests are thresholded to categorize each gene as “differential” or “non-differential”. Continuous tests are based on a continuous measure of the differential expression per gene, such as the t-statistic from a two-sample test. Additionally we consider both competitive and self-contained enrichment tests [28]. A competitive test compares the genes in the set $\xi_k$ to the genes in $\xi_k^\gamma$. The null hypothesis is that $\xi_k$ contains the same proportion of differential genes, say $\pi_k$, as $\xi_k^\gamma$. A self-contained test considers only the genes within the set $\xi_k$ and ignores the genes in $\xi_k^\gamma$. The hypothesis is that there are no more differential genes than expected where the expected value is determined a priori, i.e. 5% based on an $\alpha = 0.05$ error rate, or by sample permutation. Competitive tests allow selection of a “best” set, i.e. one that is enriched above the rest, but they are limited in that a given set $\xi_k$ with $\pi_k$ percent differential genes will receive a different test statistic depending upon the proportion of differential genes, say $\pi_k$. 

Genomics. Author manuscript; available in PMC 2013 May 01.
in the set $\xi_k$. Self-contained tests will always give the same result for the same set of data since the test of $\xi_k$ does not depend on $\xi_k$. However, if differential genes are uniformly distributed across all pathways, that is $\pi_k = \pi$ for all pathways $k = 1, \ldots, K$, then all pathways will be called enriched by a self-contained test if $\pi$ is great enough.

2.1.1. Directional Hypergeometric Test—We assume that Table 1 has a hypergeometric distribution. For $X_k$, the number of differential genes in the pathway of interest $\xi_k$, let $X_k \sim \text{Hypert}(G_{\xi}, G_{\xi_k}, D)$. Conduct a directional test of enrichment, $\Pr(X_k \geq x_k | G_{\xi}, G_{\xi_k}, D)$. This test differs from the Fisher’s Exact test in that it does not consider the depletion of differential genes in a pathway $\xi$ as an interesting case. The hypergeometric distribution assumption provides an exact p-value for this test, without permutation. Because the hypergeometric test uses the genes of $\xi_k$ to define the null proportion of differential genes, this is a competitive test. The test statistic of interest is $S_k = X_k$.

2.1.2. Weighted Kolmogorov-Smirnov Test—The Kolmogorov-Smirnov (K-S) test, also a competitive test, compares the test statistics of genes in pathway $\xi_k$ to those in $\xi_k$ using a single ranked list. The degree of separation of the genes in $\xi_k$ and in $\xi_k$ in the ranking indicates the level of enrichment. Specifically, begin by ranking the vector of t-statistics $T^G$ as $(T^G_{(1)}, T^G_{(2)}, \ldots, T^G_{(g)})$. Construct a corresponding pathway inclusion indicator vector $k(\gamma)$ where $k(\gamma(g)) = 1$ if $\gamma(g) \in \xi_k$ and 0 otherwise. The statistic $S_k = \max_h |P(k, h) - P(k, h)|$ is then the maximum deviation of the empirical distributions, where $P(k, h) = \sum_{g \in h} (\nu_g I(\gamma(g)))/(\sum_{g} \nu_g I(\gamma(g)))$, $P(\xi_k, h) = \sum_{g \in h} (1 - I(\gamma(g)))/(G - G_{\xi_k})$, and $\nu_g \in [0, 1]$ is a weight for $\gamma(g)$. The weighted K-S test as proposed by Subramanian et al. (2005) [29] for pathway enrichment testing uses $\nu_g = |\text{corr}(x_{ig}, \psi_i)|$, where $x_{ig}$ is the gene expression level for gene $g$ and person $i (i = 1, \ldots, N \text{ persons})$ and $\psi_i = 1$ for cases and 0 for controls. This weight emphasizes genes that cluster in the tails of the ranked list by giving them higher weight. In our simulations we define $\nu_k$ based on a function of the per-gene test statistic $\nu_g = |T^G_{(g)}|/(1 + |T^G_{(g)}|)$. We choose this weight for convenience in our simulation but it has a monotonic relationship with the Pearson’s correlation coefficient in the Subramanian $\nu_k$ and thus only the magnitude of the test statistic will be affected. Significance is determined using permutation sampling of the sample labels $\psi_i$ to construct the null distribution or, in the simulation, draws from the null distribution of test statistics.

2.1.3. Binomial Test of Proportions—Also called Tukey’s Higher Criticism [28], the binomial test of proportions is self-contained such that only genes contained in pathway $\xi_k$ are considered for the test of that pathway (i.e. the top row of Table 1). Specifically we assume that $X_k$, the number of differential genes in the pathway of interest $\xi_k$, is distributed $X_k \sim \text{Bin}(G_{\xi_k}, a)$ where $a$ is set a priori. Under $H_0$ we may set $a = 0.05$, the Type I error rate for each per-gene test. In the following, significance, i.e., large $X_k/G_{\xi_k}$, is determined using permutation sampling of the sample labels $\psi_i (i = 1, \ldots, N \text{ persons})$, to construct the null distribution or, in the simulation, draws from the null distribution of test statistics. The test statistic of interest is $S_k = X_k/G_{\xi_k}$ for each pathway $k$.

2.1.4. Sum of Squared Test Statistics—The test statistic from the the sum of squared test statistics method is simply $S_k = \sum_{g \in \xi_k} (T^G_{(g)})^2$, that is the sum of the squared per-gene test statistics in the set $\xi_k$ [30]. Significance is determined using permutation sampling of the
sample labels, $\psi_i (i = 1, \ldots, N \text{ persons})$, to construct the null distribution or, in the simulation, draws from the null distribution of test statistics. This is a self-contained test since $S_k$ and its null distribution consider only the genes in $\xi_k$.

2.2. Simulations

Two simulation models were used to explore the various weight functions. The first simulation uses simple construction to explore specific properties of the data such as correlation, sample size, and effect size. The second simulation is derived from real data and thus has a more complex but realistic construction. This allows us to view the different weight options in a realistic application while retaining knowledge of the true nature of the data. Interested readers can find the simulation R code in the supplemental materials online.

2.2.1. Simulation Model I—Z-scores are simulated from a standard multivariate normal distribution to represent the per-gene test statistics of differential expression (i.e. $T_g^G$, $Z_g^G$) and per-metabolite test statistics of differential intensity ($Z_m^M$). A constant correlation between like elements, i.e. gene-gene ($\rho_{GG}$) and metabolite-metabolite ($\rho_{MM}$), and a constant but lesser correlation between gene and metabolite ($\rho_{GM}$) within a pathway are assumed. For simplicity we assume that pathways are disjoint, that is no element appears in multiple pathways and there is no correlation between elements in different pathways. The case of non-disjoint pathways will be considered in Simulation II; see section 2.3.

We model each pathway to have $N_k^G$ genes and $N_k^M$ metabolites. We draw a vector of z-scores ($Z^G_k$, $Z^M_k$), where $Z^G_k = (z^G_1, \ldots, z^G_{N_k^G})$ and $Z^M_k = (z^M_1, \ldots, z^M_{N_k^M})$, from

$$(Z^G, Z^M)_k \sim MVN((\beta, \phi), \Sigma).$$

The variance covariance matrix is defined per pathway as

$$\Sigma = \begin{bmatrix}
1 & \rho_{GG} & \cdots & \rho_{GM} & \rho_{GM} \\
\rho_{GG} & 1 & \cdots & \rho_{GM} & \rho_{GM} \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\rho_{GM} & \cdots & 1 & \rho_{MM} & \rho_{MM} \\
\rho_{GM} & \rho_{GM} & \cdots & \rho_{MM} & 1
\end{bmatrix}_{(N_k^G+N_k^M) \times (N_k^G+N_k^M)}$$

where $\rho_{GM} < \min(\rho_{GG}, \rho_{MM})$. Under the null model $(\beta, \phi)$ is a vector of zeros.

Under the alternative model, z-scores are simulated from a multivariate normal distribution with the same variance-covariance matrix, $(\Sigma)$, as the null model but with shifted means $(\beta_1 = \cdots = \beta_{N_k^G} = \beta > 0$ and $\phi_1 = \cdots = \phi_{N_k^M} = \phi > 0)$. Genes and metabolites are drawn from this alternative model according to a Bernoulli ($\pi_k$) distribution thereby assigning some elements to be truly differential. The probability of differential elements can differ for genes, $\pi_k^G \in [0, 1]$, and metabolites, $\pi_k^M \in [0, 1]$. We also allow $\pi_k^G$ and $\pi_k^M$ to differ by pathway $(k)$ thereby defining some pathways to be enriched. We retain the simulated state of differential intensity for each metabolite, $m \in (m, \ldots, M)$ in the vector $H$ where $H_m = 0$ for the null case and $H_m = 1$ for the differential case. P-values are calculated from the simulated z-scores using the standard normal distribution, i.e. $p = 2 \Phi(|Z| \geq z_{\alpha/2})$. 

Genomics. Author manuscript; available in PMC 2013 May 01.
We consider a scenario with 50 pathways and allow the following parameters to vary:

- Alternative means ($\phi, \beta$): (1.5, 2), (2,3)
- Pathway size ($N_k^M, N_k^G$): (3, 20), (5, 40)
- Percentage of enriched pathways: 10%, 20%
- Correlation between like elements ($\rho_{MM}, \rho_{GG}$): 0.2, 0.4, 0.6
- Correlation between gene and metabolite ($\rho_{GM}$): 0.0, 0.10, 0.15, 0.25, 0.50 where $\rho_{GM} < \min(\rho_{GG}, \rho_{MM})$

To attain the desired level of enrichment, we set $(\pi_k^M, \pi_k^G) = (0.75, 0.50)$ for enriched pathways and $(\pi_k^M, \pi_k^G) = (0, 0)$ for non-enriched pathways when $(N_k^M, N_k^G) = (3, 20)$. For the larger pathways, $(N_k^M, N_k^G) = (5, 40)$, we set $(\pi_k^M, \pi_k^G) = (0.50, 0.25)$ for enriched pathways and $(\pi_k^M, \pi_k^G) = (0, 0)$ otherwise.

Each of the four enrichment tests described above was applied to the gene expression z-scores for each of the 50 pathways. For tests requiring it, the null distribution was simulated by the generation of 1000 null vectors of z-scores ($\pi_k^G = \pi_k^M = 0.05$ for all $k$). The enrichment test statistic ($S_k$) and p-value ($P_{Sk}$) for each pathway ($k = 1, \ldots, 50$) was retained. Each of four weight functions was then applied to convert $S_k$ or $P_{Sk}$ into a per-pathway weight that increases with increasing significance of the pathway and that is non-negative. The functions are:

(A) $\omega_k = -\log_{10}(P_{Sk})$

(B) $\omega_k = |S_k|$

(C) $\omega_k = \Phi(S_k - \mu - \mu)$

(D) $\omega_k = \exp(S_k \mu)$

The standardized test statistic, $\tilde{S}_k = (S_k - E(S_0))/\sqrt{Var(S_0)}$, uses the 1000 null vectors to determine the null mean, $E(S_0)$, and null variance, $Var(S_0)$. Estimates of the mean and variance were determined from the hypergeometric distribution for the directional hypergeometric test. Here, $\Phi(\cdot)$ is the cumulative distribution function (CDF) for the standard normal distribution and we set $\mu = 2$ unless otherwise noted.

With four enrichment scores and four weight functions we are exploring sixteen separate weights for each pathway, $\omega_{kj}$ where $j = 1, \ldots, 20$. The pathway level weights can be applied to the simulated metabolite p-values within a given pathway such that $\upsilon_{mj} = \omega_{kj}$ for metabolite $m$ in pathway $k$ and weight option $j$. The per-metabolite weights are then $\omega_{mj} = \upsilon_{mj}/\upsilon_{j}$ where $\upsilon_j = \sum_{m=1}^{M} \upsilon_{mj}$ and thus $\tilde{\upsilon}_j = M^{-1} \sum_{m=1}^{M} \upsilon_{mj} = 1$. The per-metabolite p-values are determined from the z-score vector $\vec{z}_m$ by comparing the z-scores to a standard normal distribution, i.e. $p_{mj} = 2Pr(Z_m^M \geq \upsilon_{mj}/\upsilon_j)$. The weighted per-metabolite p-value, $p_{mj}^w$, is calculated by $p_{mj}^w = p_{mj} / w_{mj}$. This results in sixteen weighted p-values for each metabolite, i.e. an $M \times 16$ matrix.

To assess the Type I error rate for each method, with respect to metabolites, we simulated the situation of completely null data by generating Z-scores under a model where $\pi_k^G = \pi_k^M = 0$ for all $k$. We also simulated a second null setting in which we assume that there are differentially expressed elements but that they are not associated with the pathways. Here we set $\pi_k^G > 0$ and $\pi_k^M > 0$ to be constant non-zero rates for all pathways, $k \in (1, \ldots, K)$, to generate differentially expressed elements uniformly across all pathways. The second null model helps us to determine error rates and to assess any power loss from the marginal weighting of the null pathways. The power, or the probability of correctly identifying a differential result, is assessed using the true state of metabolite differential intensity, $H_m$, as simulated by the Bernoulli($\pi_k^M$) draws. We use receiver operating characteristic (ROC) curves, varying the
significance threshold for \( P_m^{M} \), and the associated area under the curve (AUC) to compare the properties of the different methods.

### 2.3. Simulation Model II

The second simulation makes use of the data structure of the KEGG pathways between genes and metabolites to define the pathways [19, 20, 21]. This introduces overlapping pathways and pathways of varying sizes into the simulation. Rather than drawing the data from a multivariate normal distribution we use bootstrap resampling of published gene expression data [31] to populate our vector of per-gene test statistics. We specifically focus on those metabolites measured in Sreekumar et al. (2009) [4] and the genes measured in Varambally et al. (2005) [31]. The gene and metabolite information for each human pathway map in KEGG was acquired from KEGG version 50 (April 2009) using perl scripts and the KEGG API.

These data include 98 KEGG pathways with measurements of both genes and metabolites; 76 of these pathways have between 10 and 100 genes measured. For the simulation we selected eight of the 76 pathways to be enriched. We then selected 15 (10.2% of 147) of the metabolites to be differential. These metabolites were selected in such a way that they are members of between one and five pathways with up to two of the pathways being enriched (see Table 3). By examining metabolites in multiple pathways we can assess the affect of weight summaries across pathways. By examining metabolites that are not in enriched pathways we can examine potential power loss due to down-weighting.

The Varambally gene expression data were analyzed per gene using a two-sample t-test with pooled variance. All 126 permutations of the samples were run and the t-test recalculated to form the permutation null distribution. To prevent overcounting, the t-statistics were averaged across probes by gene symbol prior to gene set enrichment testing. The per-gene test statistic data are simulated by randomly sampling the 2375 mapped genes, with replacement, from the t-statistic matrix (original t-statistic and 126 permutation statistics). To induce enrichment, the genes in the enriched pathways were randomly selected according to a Bernoulli(\( \pi \)) distribution from the subset of genes that were differential (n=177 at \( \alpha = 0.05 \)), where \( \pi \in (0.2, 0.5) \). That is on average 100\% of the genes in an enriched pathway were selected to be differential with an effect size as seen in the Varambally data.

The 147 metabolite p-values were drawn from a uniform distribution on [0,1]. The p-values of those 15 metabolites that were selected to be differential were chosen from a Beta(3, 37) distribution. The shape parameters were chosen for a mean of 0.075 and a relatively narrow variance to provide the marginal p-values of interest. Specifically, the probability of selecting a p-value less than 0.05 is approximately 31% whereas the probability of selecting a p-value greater than 0.2 is less than 1%. All four enrichment tests, as described in section 2.1, were run on the gene expression data and each of the four weight functions, as defined in section 2.2.1, were calculated. To accomodate metabolites, \( \lambda_m \) that belong to multiple pathways we summarized the \( \omega_k \) values across pathways within metabolite, that is \( \nu_{mj} = R(\omega_k \mid \lambda_m \in \xi_k) \). We consider both the median and the 75th percentile as the summary function, \( R(\cdot) \). The per-metabolite weights were then calculated by \( w_{mj} = \nu_{mj} / \bar{w}_j \), so that \( \bar{w}_j = M^{-1} \sum_{m=1}^{M} w_{mj} = 1 \). The gene-set enrichment tests, weight functions, and per-metabolite weights are calculated for each of 1000 generated gene expression datasets and metabolite p-value vectors.
3. Results

3.1. Simulation I: Disjoint Pathways

Z-scores are simulated from a standard multivariate normal distribution to represent the per-gene test statistics of differential expression and per-metabolite test statistics of differential intensity. The percentage of differentially expressed genes (\(\pi_k^G\)) and differential metabolites (\(\pi_k^M\)) is varied per pathway, \(k = 1, \ldots, 50\), to simulate enrichment. A constant correlation between like elements, i.e. gene-gene (\(\rho_{GG}\)) and metabolite-metabolite (\(\rho_{MM}\)), and a constant but lesser correlation between gene and metabolite (\(\rho_{GM}\)) within a pathway are assumed. For simplicity we assume that pathways are disjoint, that is no element appears in multiple pathways and there is no correlation between elements in different pathways. The case of non-disjoint pathways will be considered in Simulation II, see section 3.2.

Each of the four enrichment tests are used to summarize the simulated differential gene expression scores. The four weight functions are then applied to insure that weights are increasing with increasing pathway significance and are non-negative. In Figures 1-3 these weights are labelled A–D and color coded such that the unweighted (raw) p-values are black, and the four weight functions are (A), \(\omega_k = -\log_{10}(p)\), green; (B), \(\omega_k = |\tilde{S}|\), purple; (C), \(\omega_k = \Phi(\tilde{S} - \tilde{\mu})\), orange; and (D), \(\omega_k = \exp(\tilde{S}\tilde{\mu})\), red.

To assess whether Type I error control is preserved, we considered two null models. The first null model selects no genes and no metabolites to be differential (\(\pi_k^G = \pi_k^M = 0\) for all \(k\)) and was simulated for varying correlations. We find that the Type I error is near the nominal level except for the highest correlation levels (i.e. \(\rho_{GM} = 0.5\)). The interquartile range of the error terms includes the nominal 0.05 error rate except for the exponential weight function (D, red) which is conservative; see Supplemental Figure S1.

Under the second null model we assume that there are differential metabolites and genes in the dataset but that they are uniformly distributed across the dataset, that is they are simulated without pathway enrichment. Again the interquartile range of the error rates for each method contains the nominal \(\alpha = 0.05\) level with some inflation in the high correlation cases (e.g. \(\rho_{GG} = \rho_{MM} = 0.6\) and \(\rho_{MG} = 0.5\)). As expected, there is some power lost when there is no enrichment of the pathways, i.e. the gene expression data are not informative by pathway. On average the loss is between 0.5–0.15 points on the AUC scale, except for the exponential weight function where the conservative error rates are reflected in poor power; see Supplemental Figure S2.

When enrichment of differential metabolites and gene expression is simulated in a subset of the pathways three of the four weight functions show robust increases in power over the unweighted case (raw, black); the p-value weight (A, green), the standardized test statistic (B, purple), and the CDF transformation (C, orange). Average ROC curves, across the simulation runs, are used to compare the sensitivity and specificity of correctly identifying metabolites simulated to be differential. Figure 1 shows these average receiver operating characteristic (ROC) curves from 100 simulated datasets where the alternative means are \((\phi, \beta) = (2, 3)\). This representative plot is of data with pathway sizes \((N_k^G, N_k^M) = (3, 20)\), for genes and metabolites respectively, and correlations \(\rho_{GG} = \rho_{MM} = 0.20\) and \(\rho_{GM} = 0.10\).

Ten of 50 pathways (20%) are simulated to be enriched with \((\pi_k^G, \pi_k^M) = (0.75, 0.50)\) for \(k \in (1, \ldots, 10)\) and \((\pi_k^G, \pi_k^M) = (0, 0)\) otherwise. Increasing the correlation to \(\rho_{GG} = \rho_{MM} = 0.6\) and \(\rho_{GM} = 0.15\) provides only a marginal increase in the AUC the exponential (D, red) weight function; see Supplemental Figure S3. The other weight functions appear to have a minimal loss of power, e.g. AUC=0.98 versus AUC=0.96 for the hypergeometric test p-
value weight (A, green) in this higher correlation model. When we reduce the effect size of
the differential elements to have alternative means \((\phi, \beta) = (1.5, 2)\) there is still a substantial
increase in power for the the p-value weight (A, green), the standardized test statistic (B,
purple), and the CDF transformation (C, orange) with AUC values of 90 or greater; see
Supplemental Figures S4 and S5.

3.2. Simulation II: KEGG Based Pathways

The second simulation makes use of the data structure of the KEGG pathways between
genes and metabolites to define the pathways. This introduces overlapping pathways and
pathways of varying sizes into the simulation. Rather than drawing the data from a
multivariate normal distribution we use bootstrap resampling of published gene expression
data \cite{31} to populate our vector of per-gene test statistics. We specifically focus on those
metabolites measured in Sreekumar et al. \cite{2009} and the genes measured in Varambally
et al. \cite{2005}; 98 pathways include measurements for both genes and metabolites.

Per-metabolite p-values are drawn from a uniform(0,1) distribution, excluding 15
metabolites (10.2\% of 147) that are selected to be differential and have p-values drawn from
a beta(3,37) distribution. These metabolites were selected in such a way that they are
members of between one and five pathways with up to two of the pathways being enriched
(see Table 2). By examining metabolites in multiple pathways we can assess the affect of
weight summaries across pathways. By examining metabolites that are not in enriched
pathways we can examine potential power loss due to down-weighting. The gene-set
enrichment tests, weight functions, and per-metabolite weights are calculated for each of
1000 generated gene expression datasets and metabolite p-value vectors.

Figure 2 shows the frequency of significant p-values (at \(\alpha = 0.05\)) for each of the 15
differential metabolites across the 1000 simulated data sets where 20\% \((\pi = 0.2)\) of the
genes in an enriched pathway are drawn as differential; see Supplemental Figure S6 for \(\pi =
0.5\). The color scheme is retained from the first simulation (see Section 3.1). Additionally we
denote the median summary weights by circles and the 75th percentile weights by triangles.
Notice that these weights are similar but not exact for the four metabolites in a single
pathway. Even though their per-metabolite measures \(v_{mj} = \omega_{kj}\) for metabolite \(m\), in
pathway \(k\) with weight option \(j\), will be identical the final weights, \(w_{mj}\), are standardized
against \(\bar{v}_j\) which will vary depending on the other \(v_{m'}\) values \((m' \neq m)\), see 2.2.1. The
black squares represent the unweighted test p-value and, as expected by the Beta(3,37)
distribution used in simulation, they are significant in about 30\% of the datasets.

We see in Figure 2 that, for a given metabolite, the improvement in power is related to the
proportion of enriched pathways associated with the metabolite. Specifically, the greater the
proportion of enriched pathways the more likely the weighting will have a positive effect.
This is due to our choice of the median and 75th percentile as a summary measures for the
per-metabolite weight contribution when there are multiple pathways associated with the
metabolite. Of the two summary measures considered, the 75th percentile (triangles) tends
to perform better than the median (circles). A lower proportion of enriched pathways is
required to achieve up-weighting with the 75th percentile than with the median. However,
there are cases where the 75th percentile is not more powerful, such as for the two
metabolites with only one enriched metabolite (i.e. 1/1). Although the median and 75th
percentiles are the same for this single value, these per-metabolite weights must be
standardized so that \(w = 1\) and thus the resulting weights are dependent on the size of the
other per-metabolite weights in the data.

Finally, it is interesting to note that the amount of down-weighting for the metabolites with a
low proportion of enriched associated pathways is less for the self-contained enrichment
tests (i.e. binomial and SST) than for the competitive tests (i.e. hypergeometric and KS-test). The extent of up-weighting also appears higher for the dichotomized test methods (i.e., binomial and hypergeometric). These patterns persist when the strength of the pathway enrichment is increased (e.g. $\pi = 0.5$; see Figure S6).

Yet, power increases cannot be described without considering the rate of false positive findings. The 132 metabolites that were not chosen to be differential can be used to determine false positive rates by considering the percentage of significant calls per metabolite across the 1000 simulated data sets. In Figure 3 these percentages are plotted as boxplots per weight function. The nominal 5% error rate is noted with a black horizontal line. For most weight methods the majority of the metabolites have error near the nominal line, i.e. the box contains 5%. The unweighted tests (black) are tightly centered at 0.05 as expected by the simulation design. However, the exponential function (D, red) behaves poorly for all four enrichment tests with error rates nearing 50% for some metabolites. The negative log-p-value method (A, green) has higher error for the hypergeometric test than for the other enrichment tests. The CDF function (C, orange) has low error in the self-contained tests (binomial and sum of squared statistic) and the standardized test statistic performs well except for the hypergeometric test. There is no obvious difference between the median summary and the 75th percentile summary with respect to the error rates (see Figure 3). Increasing the pathway enrichment from $\pi = 0.2$ to $\pi = 0.5$ increases the false positive calls in the competitive tests, but the self-contained tests retain appropriate error control except for the exponential function (D, red); see Figure S7.

### 3.3. Application to Motivating Data

We apply the method of p-value weighting to the motivating data example of the Sreekumar et al. (2009) [4] metabolite data and matched gene expression data (GEO:GSE8511). We begin by assessing the pathway enrichment of the gene expression data. Two-sample t-tests were used to assess the difference of gene expression between localized prostate cancer and adjacent benign tissues, per gene. All four enrichment methods were then applied. When required, permutation p-values were calculated from 1000 permutations of the sample labels. Permutation p-values are limited in precision by the number of permutations used. Here this resulted in little discrimination of pathways for the two self-contained tests; binomial and sum of squared statistics. The statistic $S_k$, however, showed potential for discriminating between the $k$ pathways.

Of the two weight functions using $S_k$, Simulation I lead us to favor the CDF weighting function (C, orange) and the standardized test statistic (B, purple). The CDF function was more robust to heavy tails (see Figure S8) but the standardized test statistic (B, purple) outperformed the CDF weighting function (C, orange) in both simulation for all but the K-S test. The two self-contained enrichment statistics appeared to be best at controlling down weighting and false positives.

Before selecting between the two weighting functions we decided to further investigate the parameter $\mu$, the presumed differential effect under the alternative distribution, used in the CDF function. In the simulations we had fixed $\mu = 2$ but in the application dataset the distribution of $S_k$ across $k$ shows that this choice may not be optimal. The median $S_k$ is near 6 for the hypergeometric, binomial and sum of squared test statistic methods and is near to zero only for the weighted K-S approach. By assigning $\mu = 2$ we may be severely tempering the upper range of test statistics and thereby reducing the ability of the weights to differentiate pathway contributions. To assess this we looked back at the data of Simulation II under the sum of squared test ($\pi = 0.2$). The range of the $S_k$ values is less in the simulated data with median $S_k$ between 2.5 and 4.5 for the sum of squares method suggesting that $\mu = 2$ may not be optimal.

*Genomics. Author manuscript; available in PMC 2013 May 01.*
Using the first 500 simulated data sets from Simulation II ($\pi = 0.2$) we assessed how varying $\tilde{\mu}$ affected the frequency of detection for each of the 15 simulated differential metabolites; see Figure 4. In the upper panels we see that the detection rate for each metabolite increases at first and then decreases as $\tilde{\mu}$ becomes large. In contrast the error rates shown in the bottom panels decrease overall (boxes) but a handful of metabolites are falsely discovered at increasing rates. It is interesting to note that when a median summary is used for combining weights for a metabolite in multiple pathways (Figure 4, left) the frequency of detecting truly differential metabolites quickly decreases for a majority of the metabolites. In contrast, if the 75th percentile is used to combine the weights across pathways (Figure 4, right) then the power loss is less and the error rates do not increase as quickly.

From these results we consider assigning $\tilde{\mu}=P_{75}(\hat{S})$, i.e., the 75th percentile of the $S_k$ values accross all $k$. To test this adaptive $\tilde{\mu}=P_{75}(\hat{S})$ we analyzed the remaining 500 Simulation II datasets. We used the sum of squares statistic and the CDF weighting function with either $\tilde{\mu} = 2$ or $\tilde{\mu} = P_{75}(S_k)$ across all $k$. The 75th percentile statistic was used to summarize the pathway weight components ($\omega_k$) when a metabolite was represented in multiple pathways. The fixed estimate of $\tilde{\mu} = 2$ shows more consistent, though marginal, increases in detection rates among the fifteen differential metabolites; see Table 3. The adaptive estimate of $\tilde{\mu}=P_{75}(\hat{S})$ shows stronger gains but they are balanced by slightly stronger losses. Error rates are near to the nominal 0.05 rate, however, the adaptive estimate gives a wider spread of errors with one metabolite being falsely detected in up to 11.8% of the datasets.

The detection rates for the adaptive $\mu$ version of the CDF function are over 70% for several metabolites which is better than we saw for any of the other weight functions; see Figure 2. For the analysis of the Sreekumar et al. (2009) [4] metabolite data we accept the potential losses of the adaptive method in favor of the potential for strong gains. Thus we choose to use the sum of squares enrichment test for gene expression with the CDF weight function, and $\tilde{\mu}=P_{75}(\hat{S})$. Additionally we use the seventy-fifth percentile summary for metabolites that participate in multiple pathways.

Twenty-five metabolites were found to be significant at $p < 0.05$ by both the weighted and unweighted p-values. There was a loss of eight metabolites by the weighted method; homocysteine, asparagine, bradykinin, cysteine, leucine, malate, N-acetylaspartate, and oxalate. However, there was a gain of ten metabolites resulting in a net gain of two metabolites; N-acetylneuraminic acid, adenine, argininosuccinate, aspartate, glycerol, guanosine, hypoxanthine, orotidine-5'-phosphate, spermine and xanthosine. In the original publication [4] leucine was listed as a metabolite up-regulated from benign to metastatic disease so this loss is notable. Additionally there was an enrichment of amino acids detected in the differential metabolites originally and some are lost with the weighted analysis (e.g. leucine, cysteine) but aspartate is gained. Finally, sarcosine, which was of primary interest originally, had a decreased p-value in the weighted analysis (0.029 unweighted; 0.016 weighted) suggesting that this finding is supported by gene-set enrichment results.

Interestingly, we now significantly detect a decrease of spermine in tumor cells, which has been shown elsewhere to be decreased in prostate cancer tissues and spermine dysregulation was shown to be associated with increased probability of biochemical recurrence [32]. Additionally, we have detected the purine and purine sugars adenine, guanosine, xanthosine, and hypoxanthine. Along with adenosine, which was detected originally, these all react with the enzyme purine-nucleoside phosphorylase, an enzyme with oncogenic effects when dysregulated [33]. Further, metabolites of purine nucleoside phosphorylase have been found to
be potential serum biomarkers in pancreatic cancer [34]. Thus the use of gene expression weighting has drawn attention to new metabolite concepts that were missed when the two data types were analyzed separately.

4. Discussion

Here we have explored the utility of gene-based p-value weighting to enhance the power to detect differential metabolites. Genes and metabolites were related via a priori defined pathways. Gene set enrichment scores were used to summarize the gene expression data by these pathways. Four enrichment tests of varying style and four weight functions were considered for constructing the per-metabolite weights.

The weight function converts the gene set enrichment information (p-value or test statistic) into a non-negative weight that increases with increasing importance. We chose to explore the CDF (C, orange) and exponential (D, red) weight functions suggested by Roeder et al. (2006) [24], as they present both bounded (C) and unbounded (D) options for weight construction. As expected, the CDF function (C, orange) is better suited to the integration of gene expression and metabolite data by pathways than the exponential function (E, red) which is better suited for strong and sparse regions of upregulation. This is particularly obvious in the second simulation (see Figure 2). Notice that the exponential function (red) shows the potential for great improvement in detection of metabolites with a high proportion of enriched pathway associations, but those metabolites not associated with a sufficient number of enriched pathways were severely down-weighted and rarely detected. Additionally the strong up-weights of metabolites in enriched pathways falsely identified many metabolites that were associated with differential-gene enriched pathways but not simulated to be differential (see Figure 3). The strength of the exponential weight is also amplified by the choice $\hat{\mu}$ which was set at $\hat{\mu} = 2$. When we reduce the parameter to $\hat{\mu} = 1$ we see that the exponential weight method (D, red) is improved (e.g. AUC = 0.61 to 0.71) but not to the level of the CDF transformation weights (C, orange, e.g. AUC = 0.96); see Figure S9.

In contrast, the CDF function (orange) showed a more moderate improvement in detection without the severe down-weighting (see Figure 2). When we allow the parameter $\tilde{\mu}$ to be selected adaptively we see even stronger gains in detection, albeit with slightly stronger down-weighting (see section 3.3). The standardized enrichment test statistic (C, purple) also performed well in simulation and was among the best performers of the second simulation. However, the CDF function of the standardized test statistic is more robust when the tails of the per-gene test statistics are long (see Figure S8).

The p-value weight (A, green) performed well in Simulation I but had mixed results in the more complex second simulation. The high detection rates with the hypergeometric test (Figure 2, top left) were met with higher error rates (Figure 3, top left). Additionally, like the exponential function, these higher detection rates imposed stronger down-weights. For the other three tests the performance is more moderate. Interestingly, these three tests used a permutation p-value which was bounded at $1/1000$ due to the number of permutations used. In contrast, the hypergeometric test used an exact p-value which has greater range, given the sample size, leading to the larger range of weights.

Considering both simulations we recommend using the CDF function ($\omega_k = \Phi(S - \mu)$, with $\bar{S} = (S - E(S_0)) / \sqrt{Var(S_0)}$), especially when the distribution of the per-gene test statistics has long tails. Alternatively, a thresholded enrichment test can be used which ignores the magnitude of the per-gene test statistics by classifying each test as differential or not. In application we discovered that the use of $\mu = 2$ may not be optimal and an adaptive method for estimating $\mu$ was explored. This adaptive method produced more consistent gains for
some differential metabolites but also resulted in more severe losses. The choice of a fixed or adaptive $\mu$ should be made in consideration of the study goals. When a continuous enrichment test is desired, we prefer the self-contained sum of squared statistic test to the weighted Kolmogorov-Smirnov (K-S) test since the self-contained tests appear to maintain control on the false positives and resist heavy down-weighting.

Another appealing feature of the p-value weighting method is that all metabolites can be considered in the analysis. Currently, as few as one third of the metabolites measured are identified [4]. As mass spectrometry libraries are expanded this number will increase but until then those metabolites that are unknown or are not mapped to gene expression are simply awarded a weight of 1. This does not adjust the p-value nor does it affect the requirement that the weights average to 1. Thus the unidentified metabolites are tested as they would have been had weighting not been considered.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

The authors would like to thank Jeremy M.G. Taylor and Bhramar Mukherjee of the University of Michigan Department of Biostatistics for their helpful discussions. AS, a Georgia Cancer Coalition Distinguished Scholar, was supported in part by 1R01CA133458-01, 1R03CA139489-01, RCA145444A. AMC was supported in part by the NIH Prostate Specialized Program of Research Excellence grant P50CA69568. Early Detection Research Network grant U01 111275 and U.S. Department of Defense PC051081 and a Doris Duke Charitable Foundation Distinguished Clinical Investigator Award. DG was supported in part by R01-GM72007 and by the Huck Institute for Life Sciences.

### References


In this article, we show that p-value weighting can be used for disparate data sources where elements are indirectly mapped, such as through pathways.

The methodology is broadly applicable to data integration across multiple types of technologies.

The use of weighted multiple testing procedures leads to improved power for finding candidate biomarkers.
Figure 1. Average ROC curves from Simulation I

Average receiver operating characteristic (ROC) curves depict the sensitivity and specificity of each test method and weight function to detect metabolites simulated to be differential. Each weight function is uniquely colored such that the unweighted (raw) p-values are black, and the four weight functions are (A), $\omega_k = -\log_{10}(p)$, green; (B), $\omega_k = |S|$, purple; (C), $\omega_k = \Phi(S - \mu)$, orange; and (D), $\omega_k = \exp(S\mu)$, red. The mean area under the curve (AUC) estimate and associated standard error are provided in the table below each plot.
Figure 2. True positive rates for each of the fifteen differential metabolites

The metabolites are noted on the vertical axis by the number of enriched pathways with which each is associated. The weight functions are color coded as before ((A), $\omega_k = -\log_{10}(p)$, green; (B), $\omega_k = |\tilde{S}|$, purple; (C), $\omega_k = \Phi(\tilde{S} - \bar{\mu})$, orange; and (D), $\omega_k = \exp(\tilde{S}\bar{\mu})$, red) with the median summary denoted by circles and the 75th percentile summary noted by triangles. The unweighted result is noted by black squares.
Figure 3. Boxplots of per-metabolite false positive rates
Each method is represented by a boxplot and each data point represents a null metabolite. For instance, the exponential weights (D) tend to have low error rates overall, but a handful of metabolites are called significant over 400 times. The weight functions are color coded as before with the median summary in the left box and the 75th percentile summary in the right box.
Figure 4. Adaptive estimation of $\mu$.
The effect of differing $\mu$ from 1 to 15 is shown on power to detect each of the 15 differential metabolites (one curve per metabolite, top panels) and the Type I error in the remaining 132 metabolites (one point per metabolite, bottom panels). For metabolites involved in multiple pathways, both median (left panels) and seventy-fifth percentile (right panels) summaries are considered.
Table 1
Gene classification underlying enrichment testing

Genes are classified by differential expression as well as by inclusion in the set of interest. Competitive tests consider the entire table whereas self-contained tests focus on the first row.

<table>
<thead>
<tr>
<th></th>
<th>Differential</th>
<th>Not Differential</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>In Pathway $\xi_k$</td>
<td>$X_k$</td>
<td>$G_k - X_k$</td>
<td>$G_{\xi_k}$</td>
</tr>
<tr>
<td>In Pathway Complement $\xi^C_k$</td>
<td>$D - X_k$</td>
<td>$G_{\xi^C_k} - (D - X)$</td>
<td>$G_{\xi^C_k}$</td>
</tr>
<tr>
<td>Total</td>
<td>$D$</td>
<td>$G - D$</td>
<td>$G$</td>
</tr>
</tbody>
</table>
Table 2

Four gene-set enrichment tests considered

The competitive hypergeometric and weighted Kolmogorov-Smirnov tests compare the level of differentiation in the set of interest to all other sets. The self-contained binomial and sum of squared statistics tests are global tests of differentiation within the set. Thresholding of per-gene tests prior to enrichment testing is required for the hypergeometric and binomial tests.

<table>
<thead>
<tr>
<th></th>
<th>Competitive</th>
<th>Self-contained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary</td>
<td>Hypergeometric</td>
<td>Binomial</td>
</tr>
<tr>
<td>Continuous</td>
<td>Weighted...</td>
<td>Sum of Squared Statistics</td>
</tr>
</tbody>
</table>
### Table 3

**Fifteen differential metabolites**

The metabolites selected to be simulated as differential are associated with up to five pathways, of which, up to two pathways are simulated as enriched.

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Pathways</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Number Enriched</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Comparing estimators of $\tilde{\mu}$

Using 500 simulated datasets from SimulationII ($\pi = 0.2$), we calculate the percentage of times the metabolites are detected to be differential. The fifteen differential metabolites are listed by the proportion of enriched pathways in which each is included. An adaptive and a fixed estimation of $\tilde{\mu}$ are compared to the unweighted results. Median and maximum Type I error rates for the 132 non-differential metabolites are given in the bottom rows.

<table>
<thead>
<tr>
<th>Proportion Enriched</th>
<th>$\tilde{\mu} = P_{75}$</th>
<th>$\tilde{\mu} = 2$</th>
<th>Unwt.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/5</td>
<td>0.656</td>
<td>0.484</td>
<td>0.320</td>
</tr>
<tr>
<td>1/5</td>
<td>0.408</td>
<td>0.398</td>
<td>0.350</td>
</tr>
<tr>
<td>0/5</td>
<td>0.276</td>
<td>0.308</td>
<td>0.298</td>
</tr>
<tr>
<td>1/4</td>
<td>0.536</td>
<td>0.480</td>
<td>0.340</td>
</tr>
<tr>
<td>0/4</td>
<td>0.414</td>
<td>0.384</td>
<td>0.308</td>
</tr>
<tr>
<td>2/3</td>
<td>0.700</td>
<td>0.476</td>
<td>0.310</td>
</tr>
<tr>
<td>1/3</td>
<td>0.518</td>
<td>0.366</td>
<td>0.284</td>
</tr>
<tr>
<td>0/3</td>
<td>0.344</td>
<td>0.416</td>
<td>0.320</td>
</tr>
<tr>
<td>2/2</td>
<td>0.706</td>
<td>0.484</td>
<td>0.328</td>
</tr>
<tr>
<td>1/2</td>
<td>0.726</td>
<td>0.484</td>
<td>0.294</td>
</tr>
<tr>
<td>0/2</td>
<td>0.368</td>
<td>0.412</td>
<td>0.320</td>
</tr>
<tr>
<td>1/1</td>
<td>0.512</td>
<td>0.382</td>
<td>0.252</td>
</tr>
<tr>
<td>1/1</td>
<td>0.496</td>
<td>0.404</td>
<td>0.324</td>
</tr>
<tr>
<td>0/1</td>
<td>0.192</td>
<td>0.240</td>
<td>0.308</td>
</tr>
<tr>
<td>0/1</td>
<td>0.160</td>
<td>0.162</td>
<td>0.280</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Median Error</th>
<th>Max Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.044</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>0.118</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>0.048</td>
<td>0.076</td>
</tr>
</tbody>
</table>