Combinatorial Mixture Models for Single Cell Assays

Greg Finak[1], SC De Rosa[2], and Raphael Gottardo[1]

[1]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center (FHCRC), Seattle, WA
[2]HIV Vaccine Trials Network, Fred Hutchinson Cancer Research Center (FHCRC), Seattle, Wa

March 22, 2012

Abstract

Immunological endpoints in vaccine trials are measured through a number of different assays that often provide single–cell measurements of multiple, specific intracellular or cell surface proteins, or mRNA expression levels of genes in single cells from specific cell sub–populations. While these measurements are usually continuous, they are often discretized and analyzed as counts of single cells. For example, in flow cytometry, individual cells are often classified as either positive or negative for a marker based on a predetermined threshold. One such assay is the intracellular cytokine staining assay that is used to assess an individual's immune response to a vaccine by measuring the abundance antigen–specific T–cell subpopulations producing different cytokines. Cells are classified as cytokine positive or negative and the resulting cell counts are used to identify experimental conditions leading to an increase in cytokine producing cells. ICS assays are generally used to identify "vaccine responders". These are individuals whose immune system produces significantly more cytokine positive cells in response to antigen stimulation than at baseline. The rarity of these cell populations makes maximizing the sensitivity and specificity of the assay a primary concern. The standard approach to analyze such data, Fisher's exact test, is problematic for two reasons. It can be overly conservative in detecting true differences for small counts, and the data generally do not meet the assumptions of the test. Specifically, the total cell counts across conditions are generally not fixed since these are generated by independent experiments. In this paper we present a coherent framework based on mixtures of Beta–binomial or Dirichlet–Multinomial distributions for analyzing count data derived from single–cell assays. Using the ICS assay as a motivating example, our method models cytokine–specific T-cell response across all individuals simultaneously, sharing information across observations using empirical Bayes methods. The model represents the responder and non-responder observations as separate components in the model, stimulated and unstimulated cell counts are modelled independently. Using simulations and real–world vaccine trial data, we show that our model increases the sensitivity and specificity for positivity calls in ICS assays compared to Fisher's exact test.

1 Introduction

Single–cell assays are an important tool in immunology, providing a functional and phenotypic snapshot of the immune system at a given time. These assays typically measure multiple variables simultaneously on individual cells in a homogeneous mixture such as whole blood. These variables are used to classify individual cells in the mixture into more homogeneous sub–populations based on phenotypic or functional differences. Such single–cell assays are used for immune monitoring of disease, vaccine research, and diagnosis of haematological malignancies [1–3].

A motivating example from vaccine trial research is the flow cytometric intracellular cytokine staining (ICS) assay, which is used to identify individuals whose immune system responds to a vaccine. Upon vaccination, antigen in the vaccine is taken up and presented to CD4 or CD8 T–cels via antigen presenting cells. Not all T–cells can recognize all antigens. Those that recognize antigens in the vaccine become *activated* and produce a variety of cytokines that further promote the immune response. After activation, this antigen–specific subpopulation proliferates and can persist in the immune system for some time providing *memory* that can more rapidly recognize the same antigen again in the future. None the less, these antigen–specific T–cell subpopulations a very small fraction of the total number of CD4 and CD8 T–cells. The ICS assay measures the number of antigen–specific T–cells in whole blood by measuring cytokine production in response to activation following stimulation by an antigen that was present in the original vaccine. Individual cells are labelled using fluorescently conjugated antibodies against phenotypic markers (CD3, CD4, and CD8) and functional markers (cytokines) of the cell subpopulations of interest [2, 4, 5]. A sufficiently large number of cells must be collected to ensure that the rare cell populations can be detected. Subsequently, each individual cell is classified as either positive or negative for each maker based on predetermined thresholds, then the number of cells matching each subpopulation phenotype is counted. These counts are compared between antigen stimulated and unstimulated samples from an individual to identify significant differences. Assessing a broad T cell response to a vaccine is particularly important in HIV vaccine trials, where the search for immune correlates of protection against HIV progression and infection is ongoing [4, 6, 7].

The ICS assay in flow cytometry is just one example of the applications of single–cell technologies to immunology. Single–cell gene expression technologies such as Fluidigm are enabling researchers to measure the expression of 96 genes in 96 single cells from a homogeneous population sorted by flow cytometry[8]. This technology has been used to identify signatures of immune cell sub–populations that correlate with different types of vaccine [9]. To date, published analysis of Fluidigm data ignore cells where transcripts are undetected, focusing solely on continuous gene expression measurements. However, the proportion of single cells expressing individual genes also carries information and should be evaluated.

We develop our framework in the context of ICS data. While there is no standard approach to analyzing ICS assay data, Fisher's exact test, as well as more ad–hoc methods, are commonly applied on a per–cytokine basis to identify observations where the number of cytokine producing cells is greater in the stimulated than the control data [4, 10, 11]. Problems with Fisher's exact when the test assumptions are not met have been reviewed in the literature [12]. In the context

of single–cell assay data the assumption of fixed margins is not satisfied since the observations for treatment and control conditions arise from independent experiments, thus the p–values from Fisher's exact test are incorrect. More importantly, Fisher's exact test does not share information across observations, resulting in low power when counts are small.

The framework developed in this paper addresses these issues explicitly. We present a combinatorial mixture model framework for the analysis of single–cell assay data in which multivariate measurement are made individual cells. The model is used to identify observations where a significant difference exists between paired treatment and control samples with respect to the number of cells expressing different combinations of proteins, genes, or other measured properties of the cells.

Materials and Methods

Vaccine Trial ICS Dataset Description

HVTN054 is a phase 1 (safety and efficacy) trial of an adenoviral vector vaccine in individuals without prior immunity [13]. The vaccine vector expressed Gag, Pol and Env proteins from multiple HIV clades [13]. Vaccine was given at two increasing doses, as well as a placebo. T–cell responses to antigens in the vaccine were measured via the ICS assay [4, 13]. The cytokines measured were IFNg (Interferon–γ), IL2 (Interleukin–2), TNFa (Tumor necrosis factor–α) and IL4 (Interleukin 4) [4]. The sample size consisted of 20 vaccine and four placebo recipients. Statistical analysis of the original positivity calls is described in the original publication [13].

Data Import, Preprocessing, and Gating

The gated ICS assay data was imported into R from the original flowJo workspaces (version 6, TreeStar Inc, Ashland, OR) using the BioConductor tool, *flowWorkspace* (v 1.1.6) and ncdfFlow (v 1.1.4). Data were preprocessed using the flowJo–defined compensation matrices and data transformations extracted from the workspace file, and gated using methods from the flowCore package (v 1.19.2) to extract counts of cytokine positive and negative T–cells for each sample and stimulation [14].

Statistical Analysis of Responder and Non–responder calls

Below, we summarize the methods for statistical analysis of responder and non–responder calls in the published trial, as well as the methods compared in this paper.

Statistical Analysis in the Published Trial

The methodology for statistical analysis and calling responders and non–responders in the original vaccine trial is described in the original publication [13]. In general, a participant is called a "responder" to an antigen stimulation if, for a given cytokine, the number of cytokine–positive T-cells

in the antigen–stimulated sample is significantly greater (for some statistical measure of significance) than the number of cytokine–positive T-cells for the negative control (unstimulated) sample from the same individual. In the original trial, significance was measured via one–sided Fisher's exact test for each participant and cytokine, comparing stimulated against unstimulated samples from that individual. A discrete Bonferroni adjustment for multiple comparisons was applied, and stimulations with an adjusted p–value below $\alpha = 0.00001$ were called positive.

Statistical Analysis of Responder and Non–responder Calls for Direct Comparison Against the Bayesian Mixture Model Approach

Positivity calls for vaccine responders and non–responders depend upon the selection of an appropriate threshold. Therefore, to compare different methods of analysis, comparable thresholds for positivity must be selected for the methods. Our mixture modelling approach is fit within each stimulation, and we make positivity calls based on a false discovery rate calculated across individuals, within each stimulation, whereas the originally published analysis makes multiple testing adjustments within individuals, across cytokines. In order to have comparable response rates, we reanalyzed the ICS data using Fisher's one–sided exact test (as described in the original publication) but made positivity calls based on the false discovery rate computed across individuals within each stimulation.

Two Competing Beta–Binomial Models

Our approach to modelling an individual's response to vaccine using ICS data takes a Bayesian approach. We model all observations (individuals) simultaneously for each combination of cytokine and stimulation (including the unstimulated samples). For a given cytokine, we let n_s be the number of cytokine–positive cells in the stimulated sample, N_s the total number of cells in the stimulated sample, and n_u, N_u, the number of cytokine–positive and total number of cells in the unstimulated sample, respectively. Note that usually, $N_u \neq N_s$. The observed count data $\mathbf{y}$ is a matrix of size $4 \times P$, where P is the number of participants. For the i'th individual $\mathbf{y}_i = \langle N_{si}, n_{si}, N_{ui}, n_{ui} \rangle$, and can be represented as the following contingency table:

Table 1: 2 x 2 contingency table of counts for cytokine positive and cytokine negative events between stimulated and unstimulated conditions

	Cytokine	
	Negative	Positive
Stimulated	$N_{si} - n_{si}$	n_{si}
Unstimulated	$N_{ui} - n_{ui}$	n_{ui}

The positive cell counts for stimulated and unstimulated samples from the same individual are

modelled as:

$$\text{if } p_{si} = p_{ui} \equiv p_{0i}; \qquad n_{si} \sim \text{Bin}(N_{si}, p_u);\ n_{ui} \sim \text{Bin}(N_{ui}, p_{ui}) \tag{1}$$

$$\text{if } p_{si} > p_{ui}; \qquad n_{si} \sim \text{Bin}(N_{si}, p_{si});\ n_{ui} \sim \text{Bin}(N_{ui}, p_{ui}) \tag{2}$$

Where p_{si} and p_{ui} are the unobserved proportions. Equation (1) represents the *null* hypothesis where there is no difference between the stimulation and the control. Equation (2) represents the alternate hypothesis where the cytokine response is stronger in the stimulation than in the control. We place a common Beta prior on the p_{si} and p_{ui} across individuals, as shown:

$$p_{si} \sim \text{Beta}(\alpha_s, \beta_s) \tag{3}$$

$$p_{ui} \sim \text{Beta}(\alpha_u, \beta_u) \tag{4}$$

If $p_{si} = p_{ui}$ we assume that $\alpha_s = \alpha_u$ and $\beta_s = \beta_u$, thus sharing the hyper–parameters between the null and alternative model for the unstimulated samples, such that β_u, α_u hyper–parameters are equal for both the stimulated and unstimulated models. Given this formulation, the posterior probability of the data given that it is generated by model (1), is:

$$\Pr(y_i|\alpha_u, \beta_u) = \binom{N_{si}}{n_{si}} \binom{N_{ui}}{n_{ui}} \frac{\text{B}(n_{si} + n_{ui} + \alpha_u, N_{si} - n_{si} + N_{ui} - n_{ui} + \beta_u)}{\text{B}(\alpha_u, \beta_u)} \tag{5}$$

with marginal log–likelihood:

$$\mathcal{L}(\alpha_u, \beta_u|\mathbf{y}) = \sum_{i=1}^{P} \left[\log \binom{N_{si}}{n_{si}} + \log \binom{N_{ui}}{n_{ui}} + \right. \tag{6}$$

$$\left. \log\left(\text{B}(n_{si} + n_{ui} + \alpha_u, N_{si} - n_{si} + N_{ui} - n_{ui} + \beta_u)\right)\right] - P \log\left(\text{B}(\alpha_u, \beta_u)\right)$$

Thus, in the case of no response to stimulation, the counts for the stimulated and unstimulated samples are modelled as draws from the same "unstimulated" Beta–binomial distribution. Note that the unobserved parameters, p_{si}, p_{ui} have been integrated out to give the marginal log–likelihood.

If the data is generated by model (2), the posterior probability of the data is given by:

$$\Pr(y_i|\alpha_u,\beta_u,\alpha_s,\beta_s) = \binom{N_{ui}}{n_{ui}}\binom{N_{si}}{n_{si}}\frac{\mathrm{B}(n_{ui}+\alpha_u, N_{ui}-n_{ui}+\beta_u)}{\mathrm{B}(\alpha_u,\beta_u)}\frac{\mathrm{B}(n_{si}+\alpha_s, N_{si}-n_{si}+\beta_s)}{\mathrm{B}(\alpha_s,\beta_s)}\cdot$$
$$\frac{\int_{p_{ui}=0}^{1}\left(\frac{1}{\mathrm{B}(n_{ui}+\alpha_u, N_{ui}-n_{ui}+\beta_u)}p_{ui}^{n_{ui}+\alpha_u-1}(1-p_{ui})^{N_{ui}-n_{ui}+\beta_u-1}\right)\left(\mathrm{I}_{1-\mathrm{p_{ui}}}(N_s^i-n_s^i+\beta_s, n_s^i+\alpha_s)\right)dp_{ui}}{\int_{p_{ui}=0}^{1}\left(\frac{1}{\mathrm{B}(\alpha_u,\beta_u)}p_{ui}^{\alpha_u-1}(1-p_{ui})^{\beta_u-1}\right)\left(\mathrm{I}_{1-\mathrm{p_{ui}}}(\beta_s,\alpha_s)\right)dp_{ui}} \tag{7}$$

with marginal log–likelihood:

$$\begin{aligned}\mathcal{L}(\alpha_s,\alpha_u,\beta_s,\beta_u|\mathbf{y}) = &-P\log\left(\mathrm{B}(\alpha_u,\beta_u)\right) - P\log\left(\mathrm{B}(\alpha_s,\beta_s)\right) + \\ &\sum_{i=0}^{P}\Bigg\{\log\binom{N_{ui}}{n_{ui}} + \log\binom{N_{si}}{n_{si}} + \log\left(\mathrm{B}(n_{ui}+\alpha_u, N_{ui}-n_{ui}+\beta_u)\right) + \\ &\log\left(\mathrm{B}(n_{si}+\alpha_s, N_{si}-n_{si}+\beta_s)\right) + \\ &\log\Bigg[\int_{p_{ui}=0}^{1}\left(\frac{1}{\mathrm{B}(n_{ui}+\alpha_u, N_{ui}-n_{ui}+\beta_u)}p_{ui}^{n_{ui}+\alpha_u-1}(1-p_{ui})^{N_{ui}-n_{ui}+\beta_u-1}\right) \\ &\left(\mathrm{I}_{1-\mathrm{p_{ui}}}(N_{si}-n_{si}+\beta_s, n_{si}+\alpha_s)\right)dp_{ui}\Bigg] \\ &-\log\Bigg[\int_{p_{ui}=0}^{1}\left(\frac{1}{\mathrm{B}(\alpha_u,\beta_u)}p_{ui}^{\alpha_u-1}(1-p_{ui})^{\beta_u-1}\right) \\ &\left(\mathrm{I}_{1-\mathrm{p_{ui}}}(\beta_s,\alpha_s)\right)dp_{ui}\Bigg]\Bigg\}\end{aligned} \tag{8}$$

The ratio of integrals in (7) accounts for the different normalizing constants due to the constraints $p_{si} > p_{ui}$ on the prior and the posterior distributions. We call this the *constrained* model.

The term $I_{1-p_{ui}}(\beta_s,\alpha_s) = 1 - I_{p_{ui}}(\alpha_s,\beta_s) = Pr(p_{si} > p_{ui};\alpha_s,\beta_s)$, is just the CDF of Beta distribution with parameters α_s, β_s, leaving a 1–dimensional integration for the ratio of normalizing constants.

Without constraints, ($p_s \neq p_u$), the ratio of integrals for the normalizing constant in (7) is dropped. We call this the *unconstrained* model.

1.1 The Mixture of Beta–Binomials

Although we have specified the two models for the data, we do not know which observation was generated by which model. Clearly, not all individuals are expected to exhibit an immune response to a stimulation. Any individual observation, y_i, could either be generated by model (1) or by model (2). We capture this uncertainty with a mixture framework of the two competing beta–binomial models. The likelihood for the mixture is given by:

$$L(\alpha_s, \beta_s, \alpha_u, \beta_u, \pi_k | \mathbf{y}) = \prod_{i=1}^{P} \left[\pi_1 f_1(y_i|\theta_1) + \pi_2 f_2(y_i|\theta_2)\right], \quad \sum_{k=1}^{2} \pi_k = 1 \tag{9}$$

Where $\theta_1 = \{\alpha_u, \beta_u\}, \theta_2 = \{\alpha_u, \beta_u, \alpha_s, \beta_s\}$, π_1 is the fraction of observations exhibiting no response to stimulation, π_2 the fraction of observations exhibiting a response to stimulation, and $f_1 = Pr(y_i|\alpha_u, \beta_u), f_2 = Pr(y_i|\alpha_u, \beta_u, \alpha_s, \beta_s)$ from (5) and (7), above.

The unobserved component memberships are treated as missing data and modelled as random variables $\mathbf{z}_i = \{z_{i1}, (1 - z_{i1})\}$

$$z_{ik} = \begin{cases} 1 & \text{if observation } i \text{ is from the } k\text{'th model (component)} \\ 0 & \text{otherwise} \end{cases}$$

Each $\mathbf{z}_i$ follows an independent multinomial distribution with one trial and parameters $\boldsymbol{\pi} = \{\pi_1, 1 - \pi_1\}$. Given the z_i's, the complete data log–likelihood is:

$$\mathcal{L}_c(\alpha_s, \beta_s, \alpha_u, \beta_u, \pi_k | \mathbf{y}, \mathbf{z}) = \sum_{i=1}^{P} \sum_{k=1}^{2} z_{ik} \left[\log \pi_k + \log f_k(y_i|\theta_k)\right] \tag{10}$$

In this form, we use the expectation–maximization (EM) algorithm [15] to fit the model.

E–step

Given the model parameters $\boldsymbol{\Psi} = \{\alpha_u, \beta_u, \alpha_s, \beta_s, \pi_k\}$, and the data $\mathbf{y}$, we estimate the unobserved component memberships, $\mathbf{Z_i}$ by computing the conditional expectation of the $\mathbf{Z_i}$'s, $\mathbb{E}_{\boldsymbol{\Psi}}(\mathbf{Z_i}|\mathbf{y}_i)$:

$$\tilde{z}_{ik} = \frac{\pi_k f_k(\mathbf{y}_i|\theta_k)}{\sum_{k=1}^{2} \pi_k f_k(\mathbf{y}_i|\theta_k)} \tag{11}$$

M–step

Finally, given the $\tilde{z}_{ik}$, we update the estimates of the model parameters to maximize the conditional expectation of the complete–data log–likelihood. The mixing proportions are given by:

$$\hat{\pi}_k = \frac{\sum_i^P \tilde{z}_{ik}}{n} \tag{12}$$

There is no closed form for the model hyper–parameters, $\alpha_u, \beta_u, \alpha_s, \beta_s$, and they are estimated via numerical optimization using R's *optim* function. For this purpose they are re–parameterized as $\mu_u = \frac{\alpha_u}{\alpha_u+\beta_u}$ and $S = \alpha_u + \beta_u$ (likewise for the α_s, β_s), corresponding to the mean and sample size of the prior distributions.

Initialization

We initialize the z_{ik}'s using Fisher's exact test to assign each observation to either the $p_{si} = p_{ui}$ or $p_{si} > p_{ui}$ components. We then use the $\hat{z}_i$'s to initialize the hyper–parameters to their method–of–moments estimates:

$$\hat{\alpha} = \hat{\mu}\left(\frac{\hat{\mu}(1-\hat{\mu})}{\hat{\sigma}^2} - 1\right) \tag{13}$$

$$\hat{\beta} = (1-\hat{\mu}\left(\frac{\hat{\mu}(1-\hat{\mu})}{\hat{\sigma}^2} - 1\right) \tag{14}$$

Where $\hat{\mu}$ and $\hat{\sigma}^2$ are the sample mean and sample variance estimates, given the z_{ik}'s.

Generalization to Multiple Cytokines and Polyfunctionality with the Multinomial Dirichlet

The model can be generalized to handle multiple cytokines in a single stimulation, in order to assess polyfunctional cytokine responses of T–cells. We use the Multinomial–Dirichlet family of distributions to model counts of events in two *different* 2x2 contingency tables. Here we consider only the *unconstrained* case ($p_s \neq p_u$). The observed data can be represented in the following way:

Where the vector of observed counts for individual i in the stimulated or unstimulated sample is denoted: $\bar{n}_{\{s,u\}i} = \{n_{\{s,u\}ij}\}; j \in \{1 \ldots 4\}$, and j indexes the cells of the appropriate contingency table shown in Table 2. The counts are modelled as draws from different multinomial distributions:

$$\text{if } \bar{p}_{si} = \bar{p}_{ui}; \qquad \bar{n}_{ui} \sim \mathcal{M}(\bar{p}_{ui}, N_{ui}); \bar{n}_{si} \sim \mathcal{M}(\bar{p}_{ui}, N_{si}) \tag{15}$$

$$\text{if } \bar{p}_{si} \neq \bar{p}_{ui}; \qquad \bar{n}_{ui} \sim \mathcal{M}(\bar{p}_{ui}, N_{ui}); \bar{n}_{si} \sim \mathcal{M}(\bar{p}_{si}, N_{si}) \tag{16}$$

with Dirichlet priors on the proportions:

$$\bar{p}_{si} \sim \text{Dir}(\bar{\alpha}_s); \bar{p}_{ui} \sim \text{Dir}(\bar{\alpha}_u) \tag{17}$$

Table 2: Contingency tables for counts of cells expressing two cytokines between stimulated and unstimulated conditions. $n_{\{s,u\}j}$ denotes observed counts for stimulated or unstimulated table cell j, and individual i

Stimulated			Unstimulated		
	Cytokine A			Cytokine A	
	Negative	Positive		Negative	Positive
Cytokine B			Cytokine B		
Negative	n_{si1}	n_{si2}	Negative	n_{ui1}	n_{ui2}
Positive	n_{si3}	n_{si4}	Positive	n_{ui3}	n_{ui4}

For the null component, where $\bar{p}_s = \bar{p}_u$ the marginal likelihood is given by:

$$\mathrm{L}(\bar{n}_s, \bar{n}_u, N_s, N_u|\bar{\alpha}_u) = \prod_{i=0}^{P} \frac{\mathrm{B_j}(\bar{\alpha}_u + \bar{n}_{ui} + \bar{n}_{si})}{\mathrm{B_j}(\bar{\alpha}_u)} \cdot \frac{N_{si}!}{\prod_{j=1}^{J} n_{sij}!} \cdot \frac{N_{ui}!}{\prod_{j=1}^{J} n_{uij}!} \tag{18}$$

Where $\mathrm{B_j}$ is the j–dimensional Beta function: $\frac{\prod_{j=1}^{J}\Gamma(\alpha_j)}{\Gamma(\sum \alpha_j)}$.

The marginal likelihood for a component where $p_{sj} \neq p_{uj}$ for all j, is given by:

$$\mathrm{L}(\bar{n}_s, \bar{n}_u, N_s, N_u|\bar{\alpha}_u, \bar{\alpha}_s) = \prod_{i=0}^{P} \frac{\mathrm{B_j}(\bar{\alpha}_u + \bar{n}_{ui})\mathrm{B_j}(\bar{\alpha}_s + \bar{n}_{si})}{\mathrm{B_j}(\bar{\alpha}_s)\mathrm{B_j}(\bar{\alpha}_u)} \cdot \frac{N_{si}!}{\prod_{j=1}^{J} n_{sij}!} \cdot \frac{N_{ui}!}{\prod_{j=1}^{J} n_{uij}!} \tag{19}$$

Without loss of generality, if only some p_j are different between stimulated and unstimulated samples, the appropriate components of α_j can be substituted in the calculation of the likelihood eq (19).

The marginal likelihood of the non–responding component, given in eq (18) can be rewritten as:

$$\begin{aligned} l(\bar{\alpha}_u|D) = \sum_i \Bigg[&\log\Gamma(\sum_j \alpha_{uj}) - \log\Gamma(\sum_j \alpha_{uj} + n_{uij} + n_{sij}) + \sum_j (\log\Gamma(\alpha_{uj} + n_{uij} + n_{sij}) \\ &- \log\Gamma(\alpha_{uj})) + \log\Gamma(N_{si} + 1) + \log\Gamma(N_{ui} + 1) - \\ &\left[\sum_j (\log\Gamma(n_{sij} + 1) + \log\Gamma(n_{uij} + 1)) \right] \Bigg] \end{aligned} \tag{20}$$

The gradient is given by:

$$\frac{\partial l}{\partial \alpha_{uj}} = \sum_i \Psi(\sum_j \alpha_{uj}) - \Psi(\sum_j \alpha_{uj} + n_{uij} + n_{sij}) + \Psi(\alpha_{uj} + n_{uij} + n_{sij}) - \Psi(\alpha_{uj}) \tag{21}$$

and the Hessian is:

$$j \neq k; \quad \frac{\partial^2 l}{\partial \alpha_{uj} \alpha_{uk}} = \sum_i \Psi'(\sum_j \alpha_{uj}) - \Psi'(\sum_j \alpha_{uj} + n_{uij} + n_{sij}) \tag{22}$$

$$j = k; \qquad \frac{\partial^2 l}{\partial \alpha_{uj}^2} = \sum_i \Psi'(\sum_j \alpha_{uj}) - \Psi'(\sum_j \alpha_{uj} + n_{uij} + n_{sij}) + \Psi'(\alpha_{uj} + n_{uij} + n_{sij}) - \Psi'(\alpha_{uj}) \tag{23}$$

The marginal likelihood of the responding component in eq 19 can be rewritten as:

$$\begin{aligned} \sum_i \Bigg[\sum_j \big[&\log\Gamma(\alpha_{uj} + n_{uij}) + \log\Gamma(\alpha_{sj} + n_{sij}) - \log\Gamma(\alpha_{uj}) - \log\Gamma(\alpha_{sj}) \\ &- \log\Gamma(n_{sij} + 1) - \log\Gamma(n_{uij} + 1) \big] + \log\Gamma(\sum_j \alpha_{sj}) + \log\Gamma(\sum_j \alpha_{uj}) \\ - \log\Gamma(\sum_j \alpha_{uj} + n_{uij}) &- \log\Gamma(\sum_j \alpha_{sj} + n_{sij}) + \log\Gamma(N_{si} + 1) + \log\Gamma(N_{ui} + 1) \Bigg] \end{aligned} \tag{24}$$

with gradient:

$$\frac{\partial l}{\partial \alpha_{\phi j}} = \sum_i \Psi(\alpha_{\phi j} + n_{\phi ij}) - \Psi(\alpha_{\phi j}) + \Psi(\sum_j \alpha_{\phi j}) - \Psi(\sum_j \alpha_{\phi j} + n_{\phi ij}) : (\phi \in s, u) \tag{25}$$

and Hessian:

$$\frac{\partial^2 l}{\partial \alpha_{\phi j}^2} \overset{k=j}{=} \sum_i \Psi'(\alpha_{\phi j} + n_{\phi ij}) - \Psi'(\alpha_{\phi j}) + \Psi'(\sum_j \alpha_{\phi j}) - \Psi'(\sum_j \alpha_{\phi j} + n_{\phi ij}) \tag{26}$$

$$\frac{\partial^2 l}{\partial \alpha_{\phi j} \alpha_{\phi k}} \overset{k \neq j}{=} \sum_i \Psi'(\sum_j \alpha_{\phi j}) - \Psi'(\sum_j \alpha_{\phi j} + n_{\phi ij}) \tag{27}$$

Mixture Model Complexity

We may wish to detect any of $2^3 = 8$ different possible scenarios where the proportion of events in corresponding cells of the contingency tables are either equal or unequal between stimulated and unstimulated conditions. Such a model would have 8 components and 55 parameters. However, if we recognize that the models can be nested, i.e. that parameters can be shared across components with similar outcomes, then the number of parameters can be reduced to 19, and further to 15 if we only consider components where any one cell of the tables differs between stimulated and unstimulated conditions. This is outlined in Table 3.

Table 3: Nesting of models and parameter counts. Each row is a model component. The three columns correspond to cells two, three, and four of the contingency tables shown in Table 2. An open circle at a position indicates that the component models $p_{sj} = p_{uj}$, and a filled circle indicates that the component models $p_{sj} \neq p_{uj}$. The number of additional parameters that need to be estimated by including each additional component in the mixture model is in the fourth column (number of parameters for proportions + number of parameters for component weights).

Cell of Table			
cell 2	cell 3	cell 4	# of parameters
○	○	○	6+1
○	○	●	2+1
○	●	○	2+1
●	○	○	2+1
○	●	●	0+1
●	●	○	0+1
●	○	●	0+1
●	●	●	0

Model simplifications for K>2 cytokines

In the case of multiple cytokines, there is a combinatorial explosion in the number of parameters of the resulting model. We propose several alternative solutions to reducing the number of parameters in general, when the dimensionality of the Multinomial Dirichlet increases. If we model N cytokines simultaneously, this results in 2^N possible combinations. We let K be the number of possible cytokine combinations being modelled, with hyperparameter vector $\bar{\alpha}_0 = \{\alpha_{00}, \ldots, \alpha_{0K-1}\}$, and $\bar{\alpha}_s = \{\alpha_{s0}, \ldots, \alpha_{sK-1}\}$ for the unstimulated and stimulated distributions, respectively. We formulate several strategies for reducing the number of parameters required to fit MIMOSA to high dimensional data by sharing hyperparameters across some of the K combinations.

Fully shared model hyperparameters

In the most simplistic case, we set up a two mixture component model (responders and non–responders), with a goal of identifying observations with no response (equal to the null), or some response in any of the possible combinations of measured cytokines. For ICS data, the first component of the hyperparameter vector represents cells not expressing any cytokine combination, and is usually much larger than the remaining components. As a first approximation, we can set the $K - 1$ remaining parameters to be all equal $\{\alpha_{01} = \{\alpha_{01}, \ldots, \alpha_{0K-1}\}$ and thus all variation across cells expressing different cytokine combinations in the unstimulated samples are captured by a single parameter. The same simplification is made for modeling the distribution of stimulated samples. As stated

above, such a model can only be fit to two mixture components (the null component and the responder component), but requires estimation of only four parameters. In practice, such a model would likely have a vary high false positive and false negative rate.

Second order hyperparameters as functions of first–order hyperparameters

Moving towards a more realistic case, we can define first–order hyperparameters as those elements of the α vectors corresponding to cells expressing individual cytokines (i.e. cytokine A only, or cytokine B only). Second–order and higher parameters are defined as cells expressing pairs, or triplets of cytokines, and so forth (i.e. A and B, B and C, C and D). Higher order parameters can be expressed as functions of the first–order parameters (i.e. cells expressing A and B can be parameterized as $\alpha_{0ab} = \frac{\alpha_{0a}+\alpha_{0b}}{2}$ or $\alpha_{0ab} = \alpha_{0a}\alpha_{0b}$). This parameterization can be extended to models with more than two mixture components, allowing us to identify different types of responders. For example there may be observations where there are differences in the proportion of cells expressing combination A and B, but also observations where there are differences in the proportion of cells expressing A only, or B only. Distinguishing between these possibilities can increase the power of the model.

Simulation Studies

We examined the performance of the constrained ($p_s > p_u$) and unconstrained ($p_s \neq p_u$) beta–binomial mixture models via simulations. Using hyper parameters estimated from the model fit of the constrained model to data from Gag1 stimulated CD4–positive, IL2 expressing T–cells on day 28 from the HVTN054 data set, we simulated data from the constrained model with 500 observations, a response rate of 40%, an N of 10K, 20K, 30K, 50K, 75K, 100K, and 150K events, and ten independent realizations for each N. The constrained model was fit to this data and the sensitivity and specificity of the model's ability to correctly identify observations from the "responder" and "non–responder" components was evaluated through ROC curve analysis and compared against Fisher's exact test. (Figure 1). The nominal vs observed false discovery rate was also examined for the models and for Fisher's exact test, as well as the accuracy of the estimated prior distributions. In order to assess whether the unconstrained model could be used to effectively fit data from a constrained model and benefit from faster computations using exact closed form expressions rather than Monte-Carlo integration, we fit the same data simulated from the constrained model using the unconstrained model (Figure 1.

To assess the sensitivity of the model to deviations from model assumptions, we repeated the simulations with the cell proportions drawn from truncated normal distributions on $(0, 1)$, rather than beta distributions. The means and variances of the truncated normal distributions were set to the MLE estimates of the beta distributions defined by the α, β hyper parameters estimated from the HVTN054 data set (Figure 2).

Results

Simulations Show MIMOSA Outperforms Fisher's Exact Test and is Robust to Departures from Model Assumptions

We ran simulation studies to assess the performance of the constrained and unconstrained models, as described in the Methods. We found that the both the constrained and unconstrained MIMOSA models out–performed Fisher's exact test with respect to sensitivity and specificity at all values of N, and that the false discovery rate observed for the mixture model more closely reflected the nominal false discovery rate than Fisher's exact test (Figure 1). Furthermore, both models gave reasonable estimates of the true hyper–parameters (Figure 1).

Since the constrained model relies on Monte–Carlo integration, which can be computationally costly, to estimate the normalizing constant in the likelihood calculation, we examined whether the unconstrained model could be used to accurately fit data generated from the constrained model. We found that the unconstrained model performed as well as the constrained model when fitting data generated from the constrained model (Figure 1).

In order to assess the sensitivity of the model to deviations from model assumptions, we fit the unconstrained model to simulated data where the proportions were drawn from truncated normal distributions (Figure 2). Even in these circumstances the unconstrained MIMOSA model outperformed Fisher's exact test and in fact performed about as well as the constrained model fit to constrained data.

MIMOSA Outperforms Fisher's Exact Test in Real–World Vaccine Trial Data from HVTN054

We tested our method on ICS data from HVTN054, a safety and immunogenicity vaccine trial for a replication defective Adenovirus vector HIV vaccine. The data set consisted of 48 individuals who received two doses of vaccine or placebo, and ICS time points for these individuals were available at day 0 and day 28 after vaccination. Responders for each cytokine and stimulation combination were classified using Fisher's exact test and the constrained MIMOSA model at the 1% FDR level, then the response rate within the control and treatment groups were calculated for both methods. Figure 3 shows the response rates for a subset of the ICS data (Gag2 and Pol2 stimulations, for the IL2 and IFNg cytokines, at days 0 and 28, in both CD4 and CD8 T–cell subpopulations). On day 0, we found that the response rates from the MIMOSA model were zero, and equal to response rates computed from Fisher's exact test across all treatment and control groups. The same was true for the response rate in the control groups at day 28. In the treatment groups, at day 28, the response rates for the MIMOSA model were generally higher or equal to those for Fisher's exact test, demonstrating the increased sensitivity and specificity of our method.

The MIMOSA framework generates predictions for individual samples. To better understand the differences between the predictions from the MIMOSA model and Fisher's exact test, we examined the model fit from Gag2 stimulated CD4+ T–cells at day 28 producing IFNg (Figure 4). We

see that two individuals are predicted as responders by the MIMOSA model which are not detected by Fisher's exact test. Both individuals have a larger proportion of IFNg producing antigen specific T–cells in antigen stimulated than unstimulated samples, as measure by the MAP or ML estimates of the proportions (Figure 4).

Testing Multifunctional Response via the Multinomial-Dirichlet model

We applied the Multinomial–Dirichlet extension of MIMOSA to the HVTN054 trial data to identify observations with cells expressing polyfunctional cytokine profiles. As a proof of principle we limited the model to the two–component case (Figure 3), with one component for the null model where we expect no difference between stimulations, and one component for the alternative model where we expect a difference between stimulations. We fit the model to the IL4 and TNFa cytokines for Env3 and Env1 stimulated CD4 T–cells from the day zero and day 28 time points (Figures 5 and 6).

The day 28 stimulations show that the posterior probability of response is generally higher for the multivariate MIMOSA model than for the univariate models, and the day zero stimulation (Figure 5) shows that the multivariate model has fewer false positives than the univariate MIMOSA model fit to distinct cytokine combinations. We compared multivariate MIMOSA against Fisher's exact test. Fisher's exact test was used to test whether the number of cells expressing a specific combination of cytokines (i.e TNFa+/IL4-, TNFa+/IL4+, and so forth) was significantly higher in the stimulated samples than in the unstimulated samples. In this comparison, Fisher's exact test detected only one of the 8 responders detected by the MIMOSA model in the Env3 stimulated, day 28 data, and detected two false positives in the day zero data. Fisher's exact test also failed to identify four of the 12 responders identified by the MIMOSA model in the day 28, Env1 stimulated data.

Discussion

The variety of single–cell assays being adopted by the immunology community is increasing. Flow cytometry, mass cytometry, ELISPOT, Fluidigm, and other single–cell assays can all be analyzed as single–cell count data. Development of effective statistical methods to detect differences in gene or protein expression at the single cell level is becoming increasingly important. Current approaches rely on asymptotic approximations (t–test, or χ^2 test), empirical or ad–hoc methods (2–fold change), or exact tests (Fisher's exact test) where model assumptions are generally not satisfied, all of which can lead to invalid conclusions about the data [4, 10–12, 16]. Most importantly, existing classical methods do not share information across samples, resulting in less power to detect true differences than empirical-Bayes and hierarchical modelling approaches, which are widely applied in the microarray literature [17–19].

The MIMOSA model presented here uses a mixture model framework of Beta–Binomial or Multinomial–Dirichlet distributions to model cell counts in non–responding and responding individuals. Information is shared across non–responders and across responders through an empir-

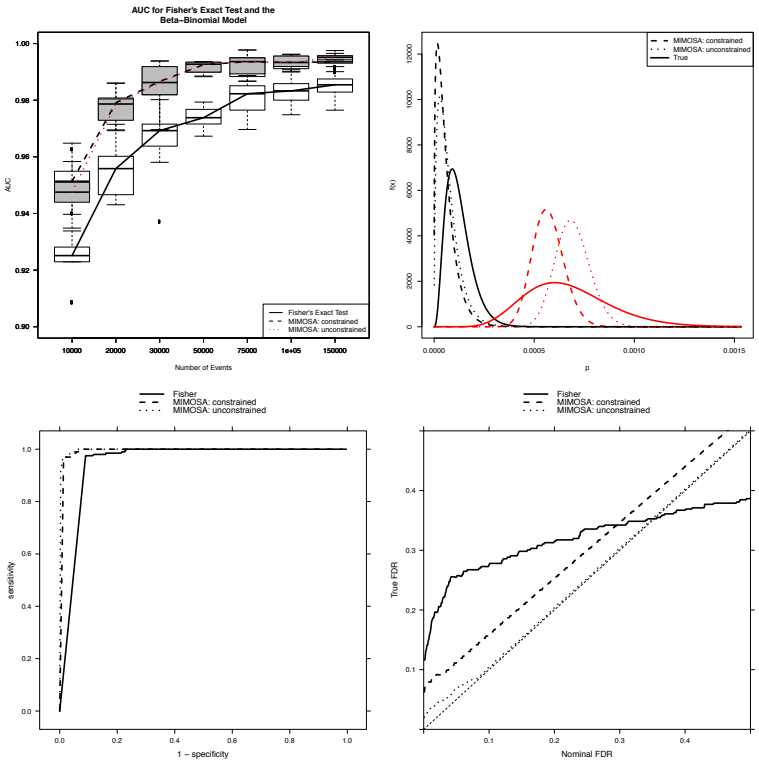

Figure 1: Performance of the constrained and unconstrained Beta–binomial mixture model vs Fisher's exact test on simulated data. Data were simulated from a constrained model with hyper–parameters estimated from a real data set of Gag1 stimulated, CD4+, IL2 expressing T–cells on day 28 from the HVTN054 trial. For total cell counts from 10,000 to 150,000, we simulated ten data sets each of 500 observations with a response rate of 40%. The performance, measured by the AUC (area under the curve), of the constrained and unconstrained Beta–binomial mixture model compared to Fisher's exact test is shown in the first panel, as a function of increasing number of cells. The beta distributions for the estimated and true hyper parameters are shown in the second panel for one simulated data set, with N=150,000 events. The ROC curve for Fisher's exact test and the constrained and unconstrained Beta–binomial model for the same simulation are shown in the third panel. The observed vs expected false discovery rate for Fisher's exact test and the constrained and unconstrained Beta–binomial model are shown in the fourth panel.

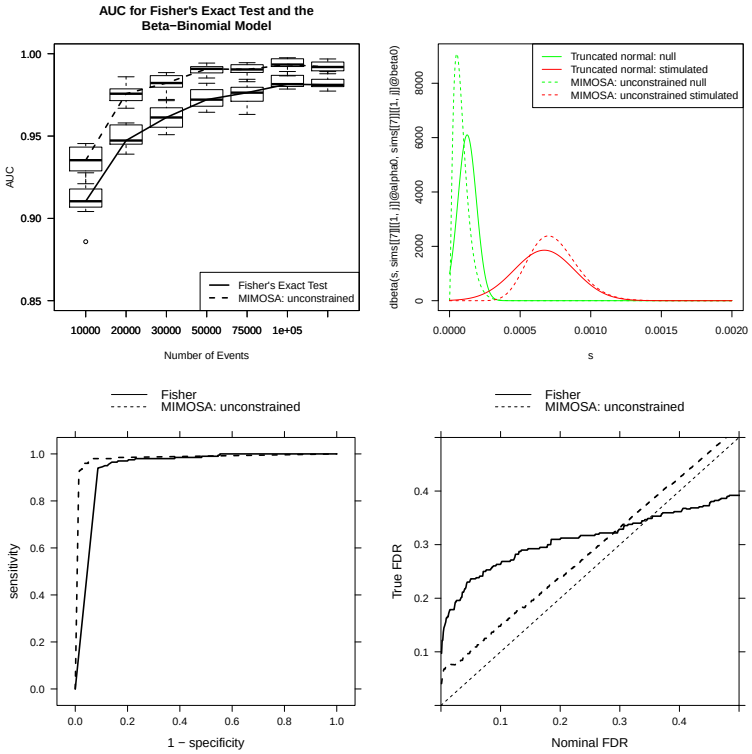

Figure 2: Sensitivity to departures from model assumptions. We generated data from variant of the constrained model ($p_s > p_u$) where the proportions were simulated from truncated normal distributions on $(0, 1)$, rather than from beta distributions. The mean and variance of the normal distributions was given by $\mu = \alpha/(\alpha+beta), \sigma^2 = (\alpha\beta)/((\alpha+beta)^2(\alpha+\beta+1))$, where α, β are the Beta–prior hyper parameters for the MIMOSA model estimated from real data (s, u subscripts omitted for brevity). The AUC for the unconstrained MIMOSA model and Fisher's exact test as a function of event count are shown in the first panel. The beta distributions for the estimated and true hyper parameters are shown in the second panel. The ROC curves for one data set and the observed and true false discovery rates for the model and Fisher's exact test are shown in the third and fourth panels.

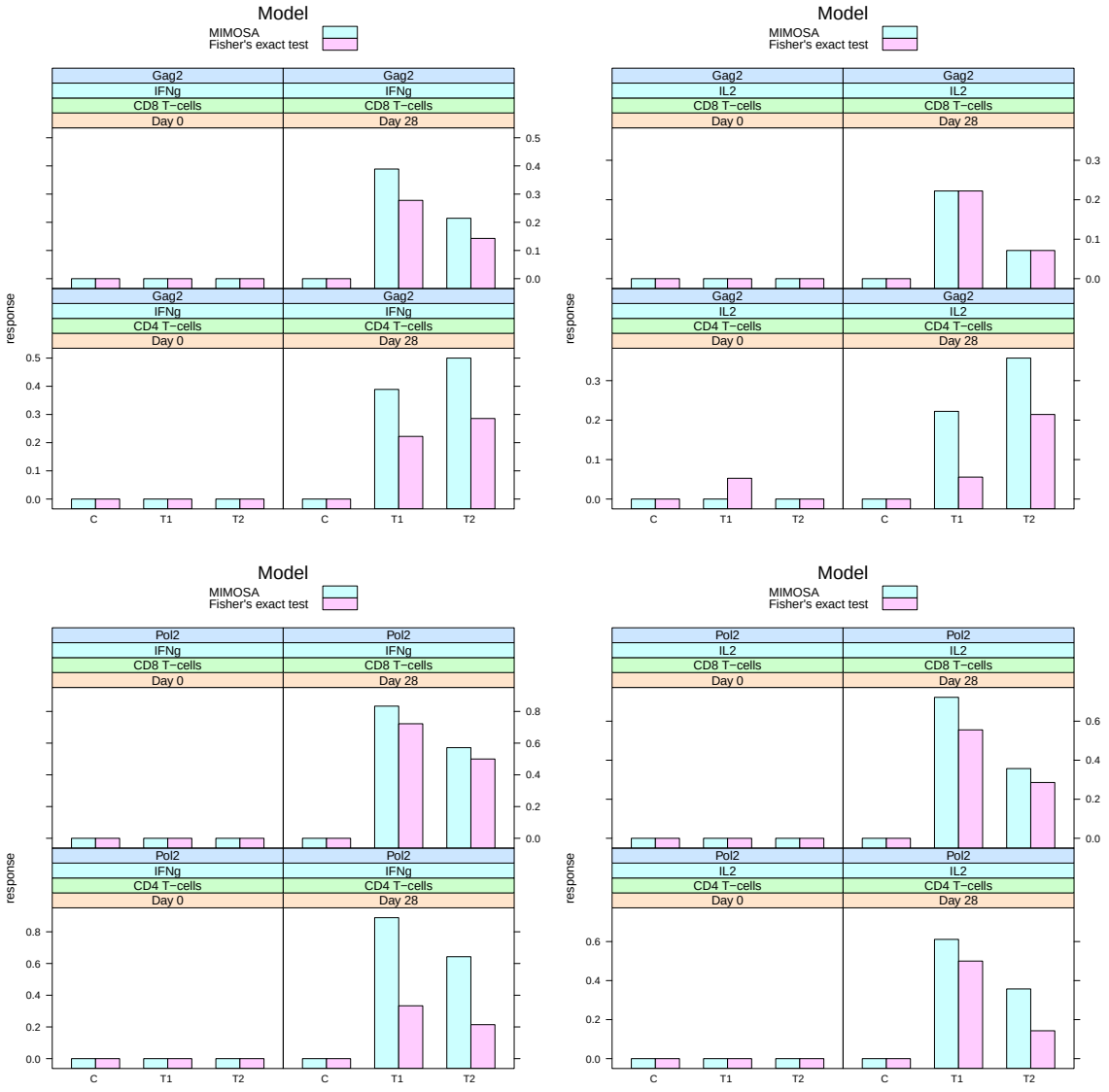

Figure 3: Comparison of MIMOSA and Fisher's exact test for calling responders in ICS data. Response rates for IL2 and IFNg cytokine positivity in Gag2 and Pol2 stimulated samples at days 0 and 28 in the CD4 and CD8 T–cell subpopulations of the HVTN054 ICS data. In the treatment groups, response rates for the MIMOSA model are greater than or equal to those for Fisher's exact test at day 28 but are equal (and zero) at day 0 as expected. In the control groups, response rates are equal (and zero) between Fisher's exact test and the MIMOSA model at day 0 and at day 28, as expected.

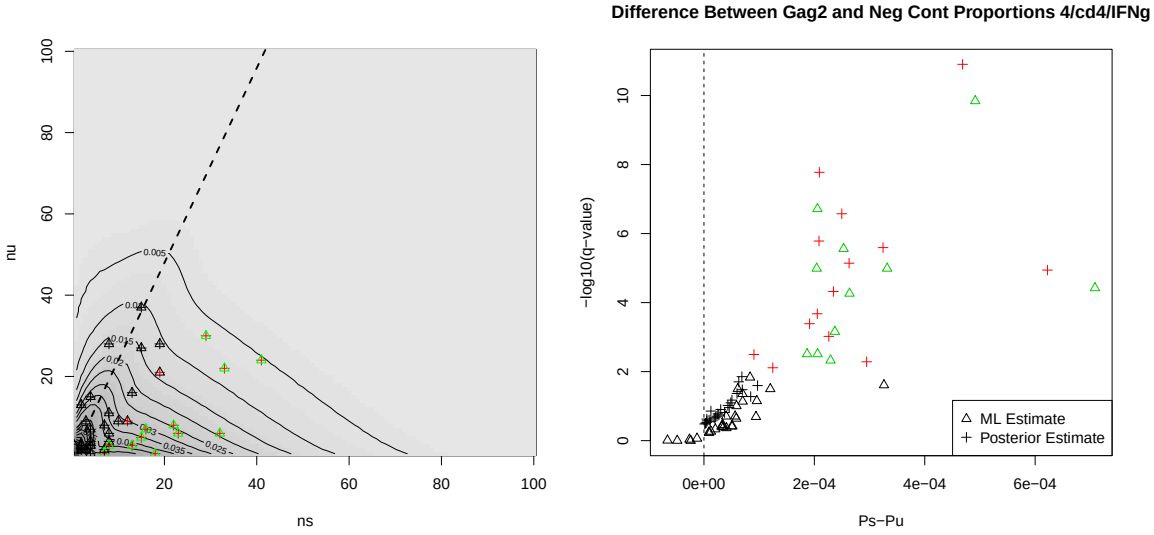

Figure 4: Likelihood surface and volcano plot for IFNg producing, CD4+ T-cells in Gag2 stimulated vs control samples on day 28. A significant difference between control and stimulated samples is called at the 1% FDR threshold (red) for the Beta–binomial model, and at 1% for FDR adjusted p–values from Fisher's exact test (green). The likelihood surface shows the observed counts from stimulated and unstimulated samples. The volcano plot shows the difference between the proportion of cytokine positive cells in the stimulated and unstimulated samples, for the maximum likelihood estimates of the proportions (triangles) and for the MAP estimates (crosses). The effect of shrinking the MAP estimates towards zero can be seen in the volcano plot.

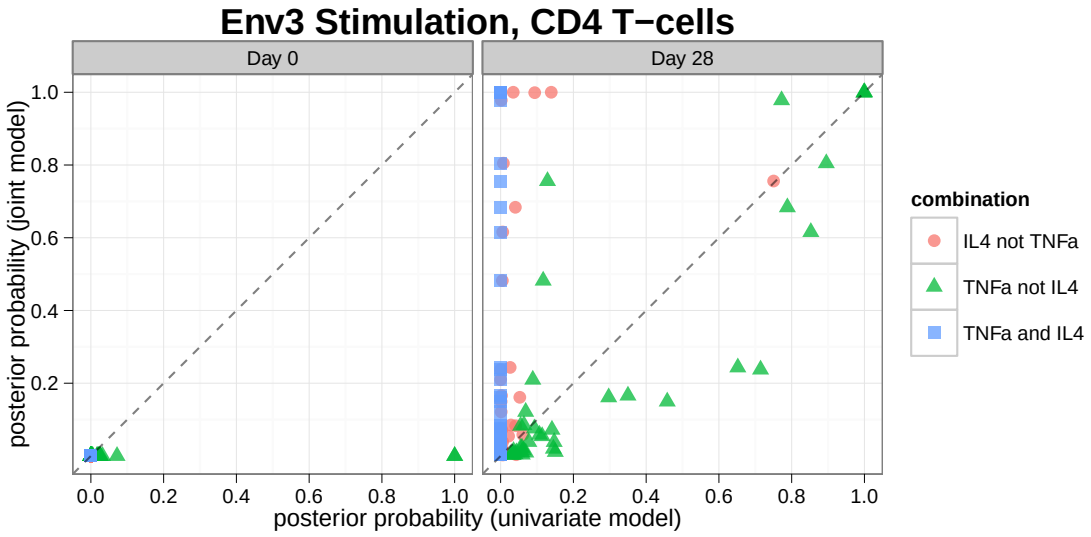

Figure 5: Posterior probabilities from the four–dimensional, Multivariate–Dirichlet, MIMOSA model, compared against posterior probabilities from univariate MIMOSA for each combination of the cytokines IL4 and TNFa, on days zero and 28 for Env–2 peptide stimulated CD4+ T–cells.

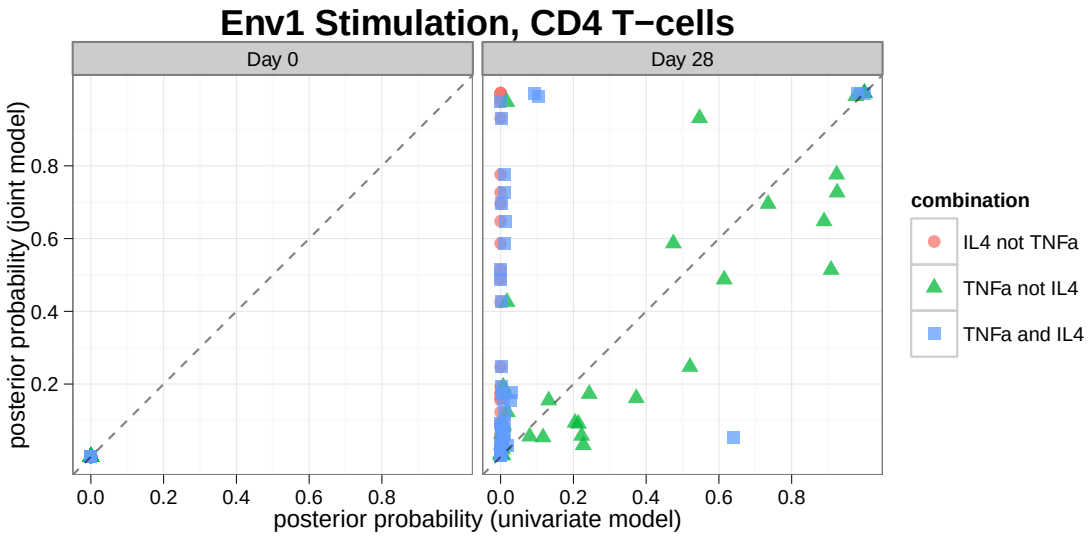

Figure 6: Posterior probabilities from the four–dimensional, Multivariate–Dirichlet, MIMOSA model, compared against posterior probabilities from univariate MIMOSA for each combination of the cytokines IL4 and TNFa, on days zero and 28 for Env–1 peptide stimulated CD4+ T–cells.

ical–Bayes prior, increasing the power to detect true differences between treatment and control conditions compared to Fisher's exact test, even when model assumptions are violated (Figures 1 and 2). The MIMOSA model based on the Beta–Binomial distribution allows us to constrain the alternative hypothesis to the case $p_s > p_u$, where the proportion of cells in the stimulated sample is strictly greater than the proportion of cells in the matched unstimulated sample, our simulations show that we can relax this assumption to the unconstrained case where $p_s \neq p_u$ without compromising the observed false positive or false negative rates (Figures 1 and 2), while saving on computation time.

The analysis of real–world ICS data from a vaccine trial demonstrated that the MIMOSA model outperformed Fisher's exact test in detecting individuals who were responders and non–responders to vaccine across multiple antigen stimulations and multiple cytokines (Figure 3). Importantly, the MIMOSA model was fit within each antigen stimulation $\times$ cytokine $\times$ time point combination but the model was blinded to placebo and vaccine treatment. Despite this, MIMOSA demonstrated a lower false positive rate at day zero (increased specificity), where no response is expected, than Fisher's exact test (Figure 3 Gag2, IL2, CD4+ T–cells, day 0, treatment group T1). Similarly, none of the placebo treated samples were identified as responders by the MIMOSA model (Figure 3 treatment group C). MIMOSA also exhibited an increased sensitivity in real–world data compared to Fisher's exact test, as demonstrated by the higher response rates for 7 of the 8 stimulation/cytokine/cell type combinations at day 28 to which the model was fit, and where a response is expected to be observed. Although it is not possible to validate the responders identified at day 28 for all cases, the conclusion of increased sensitivity is consistent with both the performance of the model in simulations, and with the increased specificity observed on day 0 and in the placebo treated samples.

We examined some of the contradictory observations identified as responders by MIMOSA and non–responders by Fisher's exact test more closely (Figure 4). For Gag2 stimulated CD4+ T–cells expressing IFNg, we see two observations detected as responders by MIMOSA at the 1% FDR level, but not by Fisher's exact test. These two observations had a 2.32 and 3.08 fold increase in the proportion of cytokine–positive cells in the stimulated vs unstimulated samples, respectively, but differences in absolute positive cell counts of 2-3 cells, although the negative cell counts for the simulated and unstimulated samples were substantially different (32K vs 82K cells, and 65K vs 145K cells, respectively), resulting in the significant difference in proportions. This is a clear example where sharing information through the empirical Bayes framework, and eliminating the conditioning on fixed margins inherent in Fisher's exact test, lead to improved sensitivity on real data for the MIMOSA model. This increased power to detect vaccine responders is important since it impacts decision–making about which vaccine candidates move forward to later phase clinical trials.

Polyfunctionality (identifying cells coexpressing multiple proteins, cytokines, or genes, is important in single cells assays since it allows the identification of functionally distinct subsets of cells in heterogeneous cell populations [20]. In vaccine trials, these polyfunctional cell subsets may be correlated with vaccine response or efficacy. Existing methods for identifying polyfunctional cytokine profiles from ICS data either test individual combinations of cytokines separately,

perform global tests at the expense of decreased power when only specific differences are of interest, sometimes empirical in nature, and generally don't share information across observations [4, 10–12, 16, 21]. As others have pointed out, in order to have the most power to detect a true difference, the statistical test selected to detect differences in various combinations of cytokines should take into consideration the cytokine combinations that are of interest [21]. A global test, one that tests for a difference between all proportions of cytokine expressing cells will have lower power than one testing for difference in a specific cytokine combination if only that specific combination is truly different. This approach is inherent in our mixture model framework, where each mixture component representing some pattern of cytokine expression that is of interest.

Multivariate MIMOSA based on the Multinomial-Dirichlet model provides a framework for identifying differences in coexpression in single–cell assays.

We show a proof of principle example with two components to identify responder and non–responder observations for the four–dimensional (two cytokine) case. Sharing of parameters across nested models in the two–cytokine case allows us to reduce the number of parameters. The model is easily extended. First, to multiple components to identify not just responders and non–responders, but responders with respect to each combination of cytokines. Second, extensions to multiple cytokines are straightforward, only requiring increasing the dimensionality of the Multinomial–Dirichlet distribution. With the increase in dimensionality there is a combinatorial explosion in the number of combinations of cytokines, and the resulting number of parameters in the model. We present several possible approaches to reducing the number of parameters in the model by sharing the hyperparamters across different combinations of cytokines. A number of these strategies can be generalized to models with multiple components that allow us to identify observations with differences in specific cytokine combinations between treatment and control samples.

Conclusions

We have developed a combinatorial mixture model framework for identifying differences between treatment conditions in paired observations of cell counts from a variety of single–cell assays. The software is implemented in R and C++, and is freely available from GitHub (`http://www.github.org/finak/MIMOSA`).

References

[1] J D Altman, P A Moss, P J Goulder, D H Barouch, M G McHeyzer-Williams, J I Bell, A J McMichael, and M M Davis. Phenotypic analysis of antigen-specific T lymphocytes. *Science (New York, NY)*, 274(5284):94–96, October 1996.

[2] Michael R Betts, Martha C Nason, Sadie M West, Stephen C De Rosa, Stephen A Migueles, Jonathan Abraham, Michael M Lederman, Jose M Benito, Paul A Goepfert, Mark Connors, Mario Roederer, and Richard A Koup. Hiv nonprogressors preferentially maintain highly functional hiv-specific cd8+ t cells. *Blood*, 107(12):4781–4789, June 2006.

[3] Margaret Inokuma, Corazon dela Rosa, Charles Schmitt, Perry Haaland, Janet Siebert, Douglas Petry, Mengxiang Tang, Maria A Suni, Smita A Ghanekar, Daiva Gladding, John F Dunne, Vernon C Maino, Mary L Disis, and Holden T Maecker. Functional T cell responses to tumor antigens in breast cancer patients have a distinct phenotype and cytokine signature. *J Immunol*, 179(4):2627–2633, August 2007.

[4] H Horton, EP Thomas, JA Stucky, I Frank, Z Moodie, Y Huang, YL Chiu, MJ McElrath, and SC De Rosa. Optimization and validation of an 8-color intracellular cytokine staining (ics) assay to quantify antigen-specific t cells induced by vaccination. *Journal of immunological methods*, 323(1):39–54, 2007.

[5] Stephen C De Rosa, Fabien X Lu, Joanne Yu, Stephen P Perfetto, Judith Falloon, Susan Moser, Thomas G Evans, Richard Koup, Christopher J Miller, and Mario Roederer. Vaccination in humans generates broad t cell cytokine responses. *J Immunol*, 173(9):5372–5380, November 2004.

[6] S Plotkin. Correlates of protection induced by vaccination. *Clinical and Vaccine Immunology*, 2010.

[7] Jerome H Kim, Supachai Rerks-Ngarm, Jean-Louis Excler, and Nelson L Michael. Hiv vaccines: lessons learned and the way forward. *Current opinion in HIV and AIDS*, 5(5):428–434, September 2010.

[8] JS Jang, VA Simon, RM Feddersen, F Rakhshan, DA Schultz, MA Zschunke, WL Lingle, CP Kolbert, and J Jen. Quantitative miRNA Expression Analysis Using Fluidigm Microfluidics Dynamic Arrays. *BMC Genomics*, 12(1):144, 2011.

[9] Lukas Flatz, Rahul Roychoudhuri, Mitsuo Honda, Abdelali Filali-Mouhim, Jean-Philippe Goulet, Nadia Kettaf, Min Lin, Mario Roederer, Elias K Haddad, Rafick P Sékaly, and Gary J Nabel. Single-cell gene-expression profiling reveals qualitatively distinct CD8 T cells elicited by different gene-based vaccines. *Proceedings of the National Academy of Sciences*, 108(14): 5724–5729, April 2011.

[10] Wendy L Trigona, James H Clair, Natasha Persaud, Kara Punt, Margaret Bachinsky, Usha Sadasivan-Nair, Sheri Dubey, Lynda Tussey, Tong-Ming Fu, and John Shiver. Intracellular staining for HIV-specific IFN-gamma production: statistical analyses establish reproducibility and criteria for distinguishing positive responses. *Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research*, 23(7): 369–377, July 2003.

[11] Elizabeth Sinclair, Douglas Black, C Lorrie Epling, Alexander Carvidi, Steven Z Josefowicz, Barry M Bredt, and Mark A Jacobson. CMV antigen-specific CD4+ and CD8+ T cell IFNgamma expression and proliferation responses in healthy CMV-seropositive individuals. *Viral immunology*, 17(3):445–454, 2004.

[12] Michael A Proschan and Martha Nason. Conditioning in 2 x 2 tables. *Biometrics*, 65(1): 316–322, March 2009.

[13] Laurence Peiperl, Cecilia Morgan, Zoe Moodie, Hongli Li, Nina Russell, Barney S Graham, Georgia D Tomaras, Stephen C De Rosa, M Juliana McElrath, and the NIAID HIV Vaccine Trials Network. Safety and immunogenicity of a replication-defective adenovirus type 5 hiv vaccine in ad5-seronegative persons: A randomized clinical trial (hvtn 054). *PLoS ONE*, 5 (10):e13579, October 2010.

[14] F Hahne, N LeMeur, RR Brinkman, B Ellis, P Haaland, D Sarkar, J Spidlen, E Strain, and R Gentleman. flowcore: a bioconductor package for high throughput flow cytometry. *BMC Bioinformatics*, 10(1):106, 2009.

[15] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977.

[16] Marcus Dittrich and Paul V Lehmann. Statistical analysis of ELISPOT assays. *Methods Mol Biol*, 792:173–183, 2012.

[17] CM Kendziorski, MA Newton, and H Lan. On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles - Kendziorski - 2003 - Statistics in Medicine - Wiley Online Library. *Statistics in ...*, 2003.

[18] M A Newton, C M Kendziorski, C S Richmond, F R Blattner, and K W Tsui. On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data. *Journal of Computational Biology*, 8(1):37–52, 2001.

[19] Gordon K Smyth, Joëlle Michaud, and Hamish S Scott. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics (Oxford, England)*, 21(9):2067–2075, May 2005.

[20] Jeffrey M Milush, Brian R Long, Jennifer E Snyder-Cappione, Amedeo J Cappione, Vanessa A York, Lishomwa C Ndhlovu, Lewis L Lanier, Jakob Michaëlsson, and Douglas F Nixon. Functionally distinct subsets of human NK cells and monocyte/DC-like cells identified by coexpression of CD56, CD7, and CD4. *Blood*, 114(23):4823–4831, November 2009.

[21] M Nason. Patterns of Immune Response to a Vaccine or Virus as Measured by Intracellular Cytokine Staining in Flow Cytometry: Hypothesis Generation and Comparison of Groups. *Journal of Biopharmaceutical Statistics*, 16(4):483–498, August 2006.