Mixture Models for Single Cell Assays

Greg Finak[1], …Others …[1], and Raphael Gottardo[1]

[1]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center (FHCRC), Seattle, WA

February 2, 2012

Abstract

Flow cytometry, specifically the intracellular cytokine staining (ICS) assay is an important tool to assess immune response to vaccines in clinical trials, providing both phenotypic and functional information about the immune system. The goal of the ICS assay is to determine whether an individual's immune system responds to a vaccine, as well as the nature of the response. ICS assays test multiple antigens and multiple cytokines across hundreds of individuals, with the usual goal of identifying individuals in whom some rare subpopulation of cells is greater in a stimulated conditions than in a control. Maximizing sensitivity and specificity to detect responders to vaccination is one of the most important considerations in designing an ICS assay, particularly due to the rarity of the cell subpopulations being measured. The measured cell subpopulations are usually rare, such as antigen–specific cytokine producing T–cells, and are often measured as a fraction of their parent population. One assay typically measures multiple cytokines simultaneously. Consequently, ICS data is typically analyzed as a series of 2x2 contingency tables using Fisher's exact test. This is problematic for a number of reasons. Fisher's exact test can be overly conservative for small counts. Additionally, the data in the 2x2 table for the ICS assay does not truly have fixed margins since the stimulated and unstimulated cells come from independent experiments. In this paper we present a coherent empirical-Bayes mixture model framework that models cytokine–specific T-cell response across all individuals in a study simultaneously, while modelling the stimulated and unstimulated cell counts independently. We show that our model increases the sensitivity and specificity for positivity calls compared to the classical approach, using simulations and real data from vaccine trials.

1 Introduction

Intracellular cytokine staining assays are used in the context of vaccine trials to provide a sensitive and quantitative assessment of antigen–specific T cell response to vaccine, including phenotyping and measurement of multiple effector functions [1–3]. 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 [1, 4, 5].

2 Materials and Methods

2.1 Vaccine Trial ICS Dataset Description

HVTN054 is a phase 1 (safety and efficacy) trial of an adenoviral vector vaccine in individuals without prior immunity [6]. The vaccine vector expressed Gag, Pol and Env proteins from multiple HIV clades [6]. 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 [1, 6]. The cytokines measured were IFNg (Interferon–γ), IL2 (Interleukin–2), TNFa (Tumor necrosis factor–α) and IL4 (Interleukin 4) [1]. The sample size consisted of 20 vaccine and four placebo recipients. Statistical analysis of the original positivity calls is described in the original publication [6].

2.2 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 [7].

2.3 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.

2.3.1 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 [6]. 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.

2.4 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

IS THIS CORRECT, OR IS IT WITHIN INDIVIDUALS ACROSS STIMULATIONS?

. 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.

2.5 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:

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

with marginal log–likelihood:

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

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:

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

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

2.6 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:

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

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 [8] 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. The observed data can be represented in the following way:

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}

Where the vector of observed counts for individual i in the stimulated or unstimulated sample is denoted: $\bar{n}_{\{s,u\}i} = \{n_{\{s,u\}ij}\}; j \in \{1 \dots 4\}$, 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:

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

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

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

$$\tag{19}$$

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

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:

$$\mathrm{L}(\bar{n}_s, \bar{n}_u, N_s, N_u|\bar{\alpha}_u, \bar{\alpha}_s) = \prod_{i=0}^{P} \frac{\mathrm{B_j}(\bar{\alpha}_u + \bar{n}_{ui})\mathrm{B_j}(\bar{\alpha}_s + \bar{n}_{si})}{\mathrm{B_j}(\bar{\alpha}_s)\mathrm{B_j}(\bar{\alpha}_u)} \cdot \frac{N_{si}!}{\prod_{j=1}^{J} n_{sij}!} \cdot \frac{N_{ui}!}{\prod_{j=1}^{J} n_{uij}!} \tag{20}$$

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 (20).

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.

Results

Simulations

HVTN054 ICS Data

Discussion

Conclusions

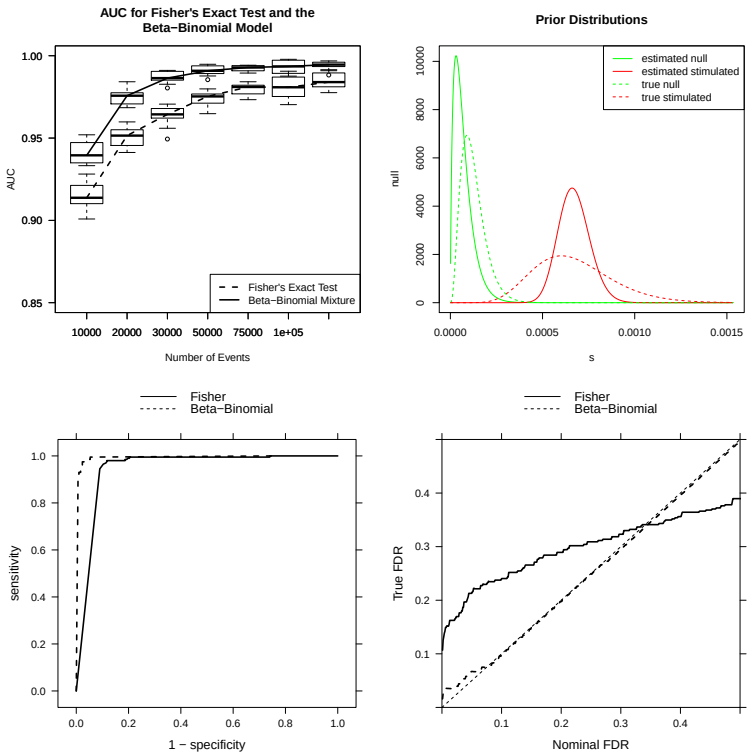

Figure 1: Performance of the Beta-binomial mixture vs Fisher's exact test in simulated data. Data were simulated from a model with hyper–parameters estimated from day 28, Gag1 stimulated, CD4+, IL2 expressing T–cells in the HVTN054 data set. We simulated 500 observations, with a response rate of 40%, and increasing numbers of cells, ten times each. The performance, measured by the AUC, of the beta–binomial mixture compared to Fisher's exact test is shown in the first panel, as a function of increasing number of events. The estimated and true prior distributions 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 Beta–binomial model for the same simulated data set are shown in the third panel. The observed vs expected false discovery rate for Fisher's exact test and the Beta–binomial model are shown for the same data set in the fourth panel.

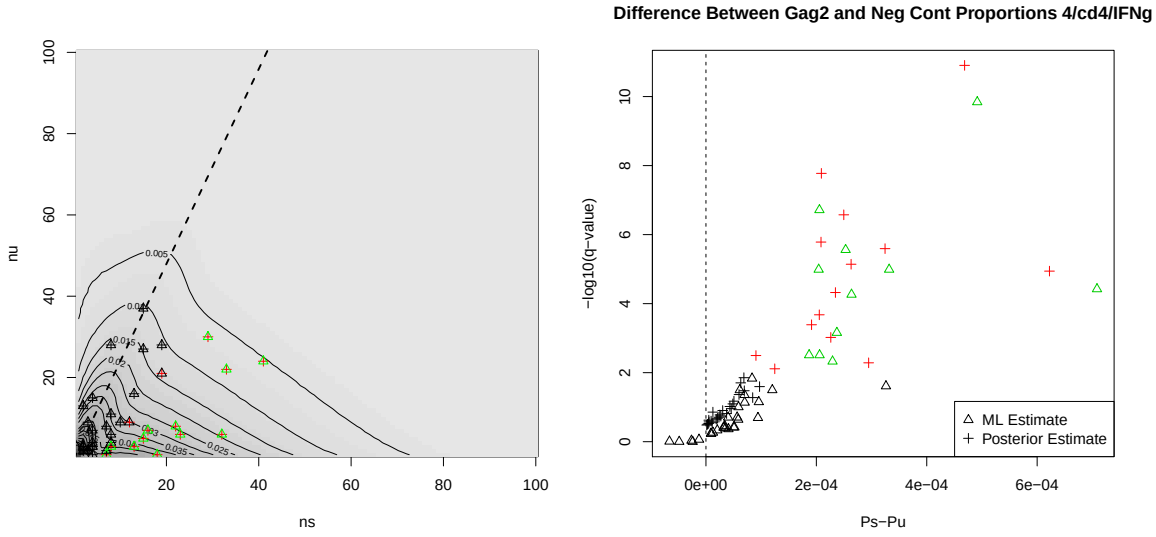

Figure 2: Likelihood surface and volcano plot for IFNg producing, CD4+ T-cells in Gag2 stimulated vs control samples on day 28. A significant difference between control and stimulated samples is called at the 1% FDR threshold (red) for the Beta–binomial model, and at 1% for FDR adjusted p–values from 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)

References

[1] 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.

[2] 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.

[3] 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.

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

[5] 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.

[6] 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.

[7] 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.

[8] 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.