An Analytical Workflow for Investigating Cytokine Profiles


Abstract
Understanding cytokine profiles of disease states has provided researchers with great insight into immunologic signaling associated with disease onset and progression, affording opportunities for advancement in diagnostics and therapeutic intervention. Multiparameter flow cytometric assays support identification of specific cytokine secreting subpopulations. Bead-based assays provide simultaneous measurement for the production of ever-growing numbers of cytokines. These technologies demand appropriate analytical techniques to extract relevant information efficiently. We illustrate the power of an analytical workflow to reveal significant alterations in T-cell cytokine expression patterns in type 1 diabetes (T1D) and breast cancer. This workflow consists of population-level analysis, followed by donor-level analysis, data transformation such as stratification or normalization, and a return to population-level analysis. In the T1D study, T-cell cytokine production was measured with a cytokine bead array. In the breast cancer study, intracellular cytokine staining measured T cell responses to stimulation with a variety of antigens. Summary statistics from each study were loaded into a relational database, together with associated experimental metadata and clinical parameters. Visual and statistical results were generated with custom Java software. In the T1D study, donor-level analysis led to the stratification of donors based on unstimulated cytokine expression. The resulting cohorts showed statistically significant differences in poststimulation production of IL-10, IL-1β, IL-8, and TNFβ. In the breast cancer study, the differing magnitude of cytokine responses required data normalization to support statistical comparisons. Once normalized, data showed a statistically significant decrease in the expression of IFNγ on CD4+ and CD8+ T cells when stimulated with tumor-associated antigens (TAAs) when compared with an infectious disease antigen stimulus, and a statistically significant increase in expression of IL-2 on CD8+ T cells. In conclusion, the analytical workflow described herein yielded statistically supported and biologically relevant findings that were otherwise unapparent.

Key terms
cytokine profile; exploratory data analysis; visualization; cytokine bead array; high-throughput; flow cytometry; diabetes mellitus Type 1; breast cancer

Although cytokines can be used as biomarkers of disease prognosis or progression, a single cytokine measured in isolation is rarely informative. Rather, constellations of cytokines measured from individual cells, or mixtures of cytokines in stimulated supernatants, provide more potent information. For example, recent studies such as those involving IL-17 and experimental autoimmune encephalomyelitis (EAE), a murine model for multiple sclerosis (MS), have illustrated that disease state is more accurately reflected by the entire cytokine milieu rather than the presence or absence of a single cytokine (1–5). IFNγ, traditionally thought of as a proinflammatory Th1 cytokine, leads to some amelioration of EAE (4). IL-6 and TGFβ both traditionally thought of as anti-inflammatory Th2 cytokines, are necessary for the generation of the proinflammatory IL-17 population (1–3,5). These findings highlight...
Investigating Cytokine Profiles

the importance of examining multiple cytokines simultaneously to decipher their roles in the overall immune environment.

Bettes et al. (6) showed that measurement of up to five cytokine-producing functions on CD8+ T cells in patients with HIV increased the ability to differentiate the responses of long-term nonprogressors from those of progressors. Other investigators have begun to use this approach to evaluate T cell responses to agents such as Leishmania major (7), vaccinia (8), and other viral infections (9). Similarly, mixtures of cytokines can be measured in stimulated cell supernatants using multiplex analytical techniques such as cytometric bead arrays. Although cell-type-specific information is lost using this technique, the ability to analyze a large number of cytokines in parallel is facilitated. Examples of studies using this approach to define disease biomarkers include the measurement of Th2 cytokine bias in tears of allergic subjects (10), Th1 cytokine bias in PBMCs of patients with prostate cancer (11), Th1 bias in active visceral leishmaniasis (12), and proinflammatory bias in the synovial fluid of patients with antirefractory lyme arthritis (13).

When a large number of cytokines and/or cell phenotypes is measured in a single study, the complexity of the data generated can make analysis daunting. Generally, researchers use visualization techniques such as scatter plots, bar graphs, and box-whisker plots, coupled with traditional statistical techniques such as Student’s t test or the Wilcoxon rank sum test, to identify differences across cohorts. When such data sets are analyzed only in bulk, donor-specific patterns are overlooked. Thus, the full predictive value of the data is compromised.

Here, we provide examples of how sophisticated clinical or translational studies measuring multiple cytokines can be productively analyzed. We present a workflow of visualization and statistical techniques that emphasizes the importance of inspecting donor-level readouts. The workflow consists of population-level analysis, followed by the examination of donor-level profiles, data transformation steps as suggested by the donor-level profiles, and a return to population-level analysis (Fig. 1). The precise data transformation steps are not known a priori, but rather are determined based on patterns suggested by the donor profiles. Thus, this approach is aligned with the thinking of advocates of cytomics (14) who argue in favor of exhaustive bioinformatics knowledge extraction to avoid the inadvertent loss of information associated with a priori hypotheses.

We demonstrate this workflow on data generated by two different high-throughput flow cytometric methods, each applied to a different disease. The first study used a cytokine bead array to measure the T-cell cytokine production of Type 1 diabetes and control subjects in response to stimulation with agonistic CD3 and CD40 antibodies. Previous work (15) showed that CD4+CD40+ T cells (T_{CD40}) are significantly expanded in the peripheral blood of T1D donors. This increase occurs irrespective of HLA-DR haplotypes, suggesting a role for CD40 on T cells in the human disease process. In a murine in vitro study, CD40 was shown to act as a costimulatory molecule (16). Examining cytokine secretion patterns in primary T cell enriched samples through the use of anti-CD3, anti-CD40, and the combination of anti-CD3 and anti-CD40 stimulation schemes provides one route to test for costimulatory and immunomodulatory function of CD40 in human disease. Data were collected for 10 cytokines (IFNγ, IL-1β, IL-2, IL-4, IL-5, IL-6, IL-8, IL-10, TNFα, and TNFβ) for each of four stimulation conditions in enriched T cell samples. The metrics of interest were MFI and fold change.

The second study (17) used intracellular cytokine staining, leveraging lyophilized stimulation, and staining plates to inspect T cell responses to tumor-associated antigens (TAAs) in breast cancer patients. One goal of the study was to compare a protective immune response, such as to cytomegalovirus, to a presumably nonprotective response, such to a TAA. The TAAs studied were CEA, MAGE-A3, and HER2/neu. The metrics of interest were background-subtracted percent positive values for IFNγ, IL-2, and TNFα as expressed by CD4+ and CD8+ T cells. In both the studies, summary metrics from flow cytometry data were combined with clinical data in a rich analytical environment, as described elsewhere (18).

METHODS AND MATERIALS

Subjects and Blood Samples: Type 1 Diabetes Study

In the T1D study, peripheral blood was collected from healthy controls (n = 7) and donors with type 1 diabetes (T1D) (n = 12), after informed consent. Lymphocytes were separated on Ficoll-Paque Plus (GE Healthcare, Princeton, NJ) per manufacturer’s instructions. Briefly, whole blood was collected from patients, diluted 1:1 with PBS (phosphate buffered saline) and placed on top of an equal volume of Ficoll-Paque Plus. The sample was spun for 30 min at 400g. The lymphocyte layer was then removed and subsequently washed twice in three times the lymphocyte layer volume in PBS centrifuging at 100g for 10 min to remove platelets. T cells were then purified on nylon wool (ZeptoMetrix Corporation, Buffalo, NY)
as previously described (19). T cells were plated in 300 µl of Dulbecco’s Modified Eagle Medium (DMEM) supplemented with 10% fetal calf serum (FCS) in 96-well round-bottomed plates at a concentration of 2 × 10⁶ cells/ml. Cells were then treated with agonistic antibodies at 10 µg/ml. Cells were left untreated or stimulated with either anti-CD3, anti-CD40, or a combination thereof, for 48 h, utilizing biotinylated anti-CD3 and/or anti-CD40 antibodies for 30 min followed by streptavidin (2 µg/ml) to induce signal capping. Mouse anti-human biotinylated CD3 antibody was obtained from AbD Serotec (Raleigh, NC), and anti-human CD40 was obtained from BD Pharmingen (San Diego, CA). Resulting supernatant was removed and combined with bead mixtures from a human Th1/Th2 10 plex cytokine bead kit (Bender MedSystems, Vienna, Austria) per manufacturer’s instructions. All samples were acquired and analyzed on a FACSCalibur flow cytometer with CellQuest Pro software (BD Biosciences, San Jose, CA). Representative gating is shown in Figure S1, which is available online.

Subjects and Blood Samples: Breast Cancer Study

Peripheral blood was collected after obtaining informed consent from 41 healthy volunteers (21 females) and 21 females with primary breast cancer. Patients enrolled in this study were within 5 years of diagnosis, with stable disease, and not receiving chemotherapy at the time of the study. All had received prior surgery and/or chemotherapy to reduce their initial tumor burden. Freshly isolated PBMCs were added to each well of a lyophilized stimulation plate containing peptides and controls and then mixed with a micropipettor to reconstitute the lyophilized pellet. Peptide mixtures, consisting of 15 amino acid residues, overlapping by 11 amino acids each, were designed to span the sequences of antigens such as CMV pp65, CEA, MAGE-3, and the intracellular domain (ICD) of HER2/neu. The stimulation plate was incubated for 2 h at 37°C, and then Brefeldin A was added. Incubation continued for another 4 h. The plate was then treated with EDTA, and cells were fixed, permeabilized, and stained as described elsewhere (17). Staining was carried out using lyophilized stain plates, formulated with two antibody cocktails: anti-IFNγ FITC, CD69 PE, CD4 PerCP-Cy5.5, CD3 APC; and anti-TNFα FITC, anti-IL-2 PE, CD4 PerCP-Cy5.5, CD3 APC.

All samples were acquired and analyzed using a FACSCalibur flow cytometer with high-throughput plate loader and CellQuest Pro software (BD Biosciences, San Jose, CA). A minimum of 10,000 CD3+CD4+ events per sample was collected.

Systems and Software

For each experiment, summary statistics (e.g., percent positive, MFI) exported after analysis of the flow cytometry FCS files were loaded into MySQL, a relational database (MySQL AB, Uppsala, Sweden). Relevant clinical data was also loaded into the database and combined with the flow data as appropriate. All statistical tests were performed in the R software environment (www.r-project.org). A custom Java application, leveraging JFreeChart (www.jfree.org/jfreechart), generated graphical analyses.

The Java application is a highly configurable system that executes a SQL query and graphs the resulting data according to various configuration parameters. For example, the graph can be either a set of scatter plots (e.g., Fig. 2) or a set of bar graphs (e.g., Fig. 3). Legends and axis labels are also controlled by configuration parameters. This architecture allows the same code base to be used to implement the analytical workflow used in this work. The workflow consists of the following steps:

1. Perform population-level exploratory data analysis (EDA) comparing summary statistics across cohorts, readouts (e.g., percent positive, MFI, and fold change), and stimulation environments.
2. Inspect donor-level profiles, showing donor-level readouts across various stimulation environments.
3. Based on donor-level profiles, perform additional data transformation steps, such as normalization or stratification of donors into new cohorts based on phenotype.
4. Perform population-level EDA on transformed data.

The exact data transformation steps are determined based on the findings from the donor-level analysis and are not necessarily driven by a priori hypotheses.

RESULTS

Initial Population-Level Analysis: T1D Study

In the T1D study, we were interested in the effect of stimulation with anti-CD3, anti-CD40, and the combination of anti-CD3 and anti-CD40. A scatter plot array (Fig. 2) of small multiples (20,21) summarizes the effects of stimulation on cytokine expression. For each cohort (control, T1D), there are three different sets of data: unstimulated versus anti-CD3 stimulation, unstimulated versus anti-CD40 stimulation, and unstimulated versus anti-CD3 and anti-CD40 stimulation. MFI readouts for 10 cytokines are shown (IFNγ, IL-1β, IL-2, IL-4, IL-5, IL-6, IL-8, IL-10, TNFα, and TNFβ). P values from Wilcoxon rank sum tests with a 95% confidence interval are shown. At this stage of analysis, we are looking for an overall sense of experimental results. Thus, statistical correction for multiple comparisons is not made. This small multiple approach provides a succinct mental model of population-level experimental results, including visual and statistical elements.

When using a bead array, MFI values can be converted to concentrations using a standard curve based on calibration beads. Although our standard curves were acceptable, with R-squared values ranging from 0.96 to 0.99 with a mean of 0.98, MFI values for IL-10, and IL-1β were frequently higher than the highest calibrated concentration. Thus, reporting MFI values was deemed more reliable in this case.

Figure 2 suggests the statistically significant differences in cytokine expression between unstimulated samples and stimulated samples. However, these population-level analyses mask...
Donor-level cytokine profiles. Additional techniques are required to investigate these profiles.

Donor-Level Cytokine Profiles Suggest Cohort Stratification

Donor-level cytokine profiles can be displayed for all four stimulation environments and all 10 cytokines. MFI is plotted on a common vertical axis for a particular donor (Fig. 3a). Each donor profile is captioned with donor identifier, donor cohort (T1D or control), gender, HLA DR alleles (where available), percent CD40 positive T cells, age at blood draw, and age of disease onset (where available/applicable). This technique facilitates visual inspection of donor-level patterns. One pattern that emerges is that some donors (e.g., 91, 92, and 101) show high levels of IL-10 and IL-1β before any stimulation. Note that since each donor-level profile is plotted into the same vertical space, the data is implicitly normalized.

In this experiment, one of the areas of interest was the change in cytokine expression given stimulation. One mechanism to calculate this effect of stimulation is fold change (22). For each donor, for each stimulation environment, for each cytokine, a fold change metric is calculated, equal to MFI_{stim}/MFI_{unstim}. For example, for a particular donor, the MFI of IL-2 when stimulated with anti-CD3 is divided by the MFI of IL-2 with no stimulation. All resulting values are plotted for each donor, providing a donor-level cytokine profile based on post-stimulation changes in cytokine expression (Fig. 3b). One of the patterns that emerge is that for donors showing a high

**Figure 2.** Stimulation results in statistically significant cytokine expression for both control (n = 7) and T1D (n = 12) cohorts. This scatter plot array shows differences between unstimulated samples and stimulated samples. Each row is specific to either control or T1D cohort. The first two rows compare untreated samples to those treated with anti-CD3; the next two rows, untreated samples to those treated with anti-CD40; and the final two rows, untreated samples to those treated with the combination of anti-CD3 and anti-CD40. Representative flow cytometry output and gating strategy is available on-line.
Figure 3. Donor-level profiles show the existence of an IL-10\textsuperscript{high} phenotype. (a) For six donors, MFI values are shown for 10 cytokines in each of four treatment environments. Each donor profile is captioned with donor identifier, donor cohort (T1D or Control), gender, HLA DR alleles where available, percent CD40 positive T cells, age at blood draw, and age of disease onset. For the first donor (Donor 84), the common abbreviation for each cytokine is shown on the horizontal axis. For subsequent donors, a shorter abbreviation is used. Donors 91, 93, and 101 show relatively high levels of IL-10 before stimulation. (b) For the same six donors, fold change values (MFI stimulated/MFI unstimulated) are shown for each of three stimulations. Donors 91, 93, and 101 show low levels of change in IL-10 in response to stimulation.
level of IL-10 before stimulation (e.g., 91, 93, and 101), fold change values are relatively low. This observation can be explored in greater depth by stratifying the donors based on the "Unstimulated IL-10 High" phenotype.

As shown in Table 1, unstimulated MFI values for IL-10 suggest that there is a reasonable breakpoint at \(1800\). Donors with an IL-10 readout above this point can be classified as IL-10\(_{\text{Hi}}\), while those below this breakpoint are classified as IL-10\(_{\text{Lo}}\).

### Population-Level Analysis After Donor Stratification

Focusing on T1D donors, grouped into IL-10\(_{\text{Hi}}\) \((n = 6)\) and IL-10\(_{\text{Lo}}\) \((n = 6)\) cohorts, Figure 4 presents statistically significant differences in fold changes. Again, a Wilcoxon rank sum test with a confidence interval of 95% is used. As the denominator of the fold change metric, unstimulated levels have a direct mathematical relationship to the computed fold changes. The results presented here suggest a biological relationship as well. Donors that show high levels of expression of IL-10 before stimulation do not show a large increase in expression after stimulation. When stimulated with anti-CD3, the mean fold change of IL-10 for the IL-10\(_{\text{Hi}}\) cohort was 0.73, representing on average a decrease in IL-10 expression given anti-CD3 stimulation. The mean fold change of IL-10 for the IL-10\(_{\text{Lo}}\) cohort was 49.3. The two cohorts show statistically significant differences in responses for IL-10 (all stimulations), IL-1\(_{\beta}\) (all stimulations), IL-8 (anti-CD40 stimulation), and TNF\(_{\beta}\) (anti-CD3 stimulation).

Stratifying these donors into IL-10\(_{\text{Hi}}\) and IL-10\(_{\text{Lo}}\) cohorts yields results that were not visible when all T1D donors were treated as a single cohort. Notice that in Figure 2 the scatter plot arrays comparing unstimulated responses to stimulated responses show no significant differences for IL-10. Often, the natural variation that exists in human population masks phenomena that apply only to a subset of the population. In this study, inspecting the donor-level profiles led to the identification of important phenotypes, as defined by

<table>
<thead>
<tr>
<th>DONOR COHORT</th>
<th>MFI IL-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1D</td>
<td>19</td>
</tr>
<tr>
<td>T1D</td>
<td>31</td>
</tr>
<tr>
<td>T1D</td>
<td>34</td>
</tr>
<tr>
<td>T1D</td>
<td>34</td>
</tr>
<tr>
<td>Control</td>
<td>47</td>
</tr>
<tr>
<td>Control</td>
<td>52</td>
</tr>
<tr>
<td>Control</td>
<td>59</td>
</tr>
<tr>
<td>Control</td>
<td>63</td>
</tr>
<tr>
<td>Control</td>
<td>69</td>
</tr>
<tr>
<td>T1D</td>
<td>112</td>
</tr>
<tr>
<td>T1D</td>
<td>486</td>
</tr>
<tr>
<td>Control</td>
<td>1822</td>
</tr>
<tr>
<td>T1D</td>
<td>2342</td>
</tr>
<tr>
<td>T1D</td>
<td>2390</td>
</tr>
<tr>
<td>T1D</td>
<td>2903</td>
</tr>
<tr>
<td>T1D</td>
<td>2915</td>
</tr>
<tr>
<td>T1D</td>
<td>3634</td>
</tr>
<tr>
<td>T1D</td>
<td>3869</td>
</tr>
<tr>
<td>Control</td>
<td>4325</td>
</tr>
</tbody>
</table>

Figure 4. T1D IL-10\(_{\text{Hi}}\) and IL-10\(_{\text{Lo}}\) donors show significantly different responses to stimulation. This scatter plot array shows differences in fold change between IL-10\(_{\text{Hi}}\) \((n = 6)\) and IL-10\(_{\text{Lo}}\) \((n = 6)\) T1D donors across 10 cytokines and 3 stimulation environments. The IL-10\(_{\text{Hi}}\) and IL-10\(_{\text{Lo}}\) cohorts are defined based on unstimulated levels of IL-10. Because unstimulated levels are the denominator of fold change, the significant differences between the two cohorts for IL-10 fold change for all three environments are not unexpected. However, the fact that all of the IL-10\(_{\text{Hi}}\) donors show very little increase in IL-10 or IL-1\(_{\beta}\) production in response to stimulation is of interest.
baseline cytokine production. These phenotypes are essential for understanding cytokine profiles.

**Initial Population-Level Analysis: Breast Cancer Study**

In the analysis of T cell responses to breast cancer antigens, flow cytometry files were gated as shown in Figure 5a. Data were reported as percent cytokine positive events (of CD3+CD4+ or CD3+CD4– lymphocytes), after subtraction of unstimulated background. CD3+CD4+ cells are hereafter referred to as CD4 cells, while CD3+CD4– cells are referred to as CD8 cells.

Cross-cohort cytokine expression for multiple stimulation environments is captured in a scatter plot array (Fig. 5b). This figure displays background-subtracted percent positive readings for six cell lineage-cytokine pairs (CD4+/IFNγ+, CD4+/IL-2+, CD4+/TNFα+, CD8+/IFNγ+, CD8+/IL-2+, and CD8+/TNFα+), two cohorts (control and breast cancer), and two stimulation environments (CMV and HER2/neu). P values from Wilcoxon rank sum tests with a 95% confidence interval are shown. Again, because we are looking for an overall sense of experimental results, statistical correction for multiple comparisons is not made.

**Donor-Level Cytokine Profiles Suggest Data Normalization**

Donor-level cytokine profiles can be displayed by plotting donor responses for each of the six cell lineage-cytokine pairs (CD4+/IFNγ+, CD4+/IL-2+, CD4+/TNFα+, CD8+/IFNγ+, CD8+/IL-2+, and CD8+/TNFα+) for multiple stimulation environments (Fig. 6a). This technique facilitates the identification of cytokine profile phenotypes. For example, both donors 506 and 515 show a CD8+/TNFα+ “High” feature in response to stimulation with multiple antigens. All donors show a CD8+/IL-2+ “Low” feature in response to stimulation with flu. Yet one limitation of this technique is that readouts across stimulation environments must be of comparable ranges. Alternatively, data of disparate ranges can be normalized onto a common scale.

A donor’s response to viral stimulation can be compared to that of donor’s responses to TAAs such as CEA, HER2/neu, and MAGE-3, by focusing on CMV-seropositive breast cancer patients. However, the magnitude of the TAA-specific T cell responses was generally lower than that of CMV-specific T cell responses. Thus, the normalization is appropriate. In Figure 6b, the six readouts (CD4+/IFNγ+, CD4+/IL-2+, CD4+/TNFα+, CD8+/IFNγ+, CD8+/IL-2+, and CD8+/TNFα+) for a representative donor for CMV and MAGE-3 are plotted.

---

**Figure 5.** Gating and subsequent data processing yield background-subtracted percent positive values for six readouts per donor. (a) A representative sample shows expression of IFNγ, TNFα, and IL-2 on CD3+CD4+ and CD3+CD4– lymphocytes. Results are shown for both HER2/neu stimulation and an unstimulated control. The percent positive values for the unstimulated control are subtracted from the corresponding values for the stimulated sample. (b) A scatter plot array shows the background subtracted percent positive readouts for both breast cancer (n = 21) and control (n = 21) cohorts. There are no significant differences given CMV stimulation, but there are significant differences in IL-2 expression given HER2/neu stimulation.
Figure 6. Data normalization is suggested by donor-level profiles. (a) Readouts for multiple stimulation environments (Flu, HIV, CEA, HER2, and MAGE-3) are shown for four donors. Donors 506 and 515 show a CD8+/TNFα+ high response to stimulation with all three TAAs. However, different dynamic ranges on cytokine responses suggest the need for data normalization. (b) For a representative donor, six readouts (CD4+/IFNγ+, CD4+/IL-2+, CD4+/TNFα+, CD8+/IFNγ+, CD8+/IL-2+, and CD8+/TNFα+) are plotted on a common vertical scale for each of two stimulation environments (CEA and MAGE-3). In comparison to the relatively large CMV response, the pattern of the MAGE-3 response is difficult to discern. (c and d) Responses are plotted onto two different scales, but into the same physical height. Even though the numeric scales are different, the pattern of response can be compared visually. (e) The values for each antigen are normalized by range. This means that for a particular donor for a particular stimulation environment, the six readouts are mapped onto a scale from 0 to 1. (f) For all CMV seropositive breast cancer patients (n = 13), percent positive values are shown for CMV and the three TAAs. In general, CMV responses are higher than the TAA responses. A statistical test would not provide additional insight. (g) A scatterplot array comparing normalized CMV responses to normalized TAA responses for all CMV seropositive breast cancer donors shows statistically significant differences in CD4+/IFNγ+, CD8+/IFNγ+, and CD8+/IL-2+ subsets.
on a common vertical scale. In comparison to the relatively large CMV response, the pattern of the MAGE-3 response is difficult to discern. In Figures 6c and 6d, responses are plotted onto two different scales, but into the same physical height. Even though the numeric scales are different, the pattern of response can be compared visually. In Figure 6e, the values for each antigen are normalized by range. Arithmetically, the cytokine responses have been mapped onto a scale from 0 to 1, where 0 represents the lowest level of cytokine expression and 1 represents the largest. This allows cytokine profiles to be compared visually and statistically, regardless of the original magnitude of response. Other authors who used flow cytometry in clinical and translational studies (6,23,24) have also used normalization techniques.

The specific normalization formula used in this study is

\[
\frac{x_i - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}
\]

where \(x_i\) is the cytokine response for a particular donor for a particular antigen, \(x_{\text{min}}\) is the smallest of these same responses, and \(x_{\text{max}}\) is the largest of these responses.

**Population-Level Analysis After Data Normalization**

A scatter plot comparing readouts of CMV and the three TAAs for the CMV seropositive breast cancer patients are shown in Figure 6f (\(n = 13\); one patient at stage I, six at stage II, one at stage III, three at stage IV, two of unknown stage; age range of 34–79 years, with a mean of 49 years). Again, the magnitude of the CMV responses is so much larger than the TAA responses that a statistical comparison is not informative. However, once normalized, standard statistical techniques can be applied to the data (Fig. 6g). Using a Wilcoxon signed rank test, which is a paired test, with Holm correction for multiple comparison, a significantly lower proportion of CD4+ cells expressing IFN\(\gamma\) was found in all three TAA responses, compared to the CMV response in the same donors (\(P \leq 0.0103\)). Correspondingly, a significantly lower proportion of CD8+ cells expressing IFN\(\gamma\) was found in all three TAA responses (\(P \leq 0.004\)). In contrast, a significantly higher proportion of CD8+ T cells expressing IL-2 was found in all three TAA responses, compared to the CMV response in the same donors (\(P \leq 0.009\)). Together, these data show a consistent difference in the cytokine profile between TAA and CMV responses, with TAA responses consisting of proportionally fewer IFN\(\gamma\)+ T cells and more CD8 IL-2+ cells. This finding is important because it indicates that the pattern of cytokine response to TAA is fundamentally different from that to CMV, suggesting that the immunotherapies directed against TAAs may need to alter these T cell profiles to be effective. This observation could not have been substantiated without data normalization.

**DISCUSSION**

Cytokines are chemical messengers that mediate interactions between various cell subtypes, directing further downstream activity toward proinflammatory, anti-inflammatory, or quiescent states. However, studies that consider small numbers of cytokines will provide an incomplete picture of the immune environment. Flow cytometry supports complex studies of the cytokine milieu, yielding multiple parameters per cell in the case of intracellular cytokine flow cytometry, or multiple parameters per aliquot in the case of cytokine bead arrays. In this work, the bead array measured 10 cytokines per donor per treatment. The breast cancer data were generated with a plate-based assay consisting of 96 wells per donor, with six readouts per well. Both of these experiments represent the types of data sets able to provide insight into cytokine milieus and profiles.

Yet complex data sets present substantial data analysis challenges. To address these challenges, we have presented a workflow that includes population-level analysis, donor-level analysis, and data transformation. This workflow is applicable to other areas of research examining large numbers of parameters per donor, such as phospho-flow. Phospho-flow cytometry measures multiple intracellular protein phosphorylation events in each aliquot. Resulting data support the mapping of signaling networks at a single cell level (22,23).

Currently, various researchers are presenting techniques to summarize the results of highly dimensional clinical and translational studies (6,22–24,26). Much of this analysis is at the population level, comparing one cohort to another using traditional visualization and statistical techniques (11–13) or using data-mining techniques such as clustering (22,23) to automatically group donors into cohorts. Perambakam (11) presents three scatter plots and applies Student’s t test. Ansari (12) presents three bar graphs and applies both Student’s t test and the Wilcoxon signed rank test. Our work suggests that valuable observations may be overlooked when analysis stops at these sorts of cross-cohort comparisons.

In the T1D study, a comprehensive presentation of population-level metrics and relevant statistical tests provided an overview of experimental results. In the next step in the workflow, the donor-level profiles revealed IL-10\(\text{high}\) and IL-10\(\text{low}\) phenotypes. Donors with an IL-10\(\text{high}\) phenotype showed a different response pattern to stimulation than donors with the IL-10\(\text{low}\) phenotype. This differing response was demonstrated both by inspection of donor-level profiles of fold change and by subsequent population-level statistical analysis. These results showed statistically significant differences in IL-10, IL-1\(\beta\), and TNF\(\alpha\). Recall that the scatter plot array of unstimulated responses to stimulated responses showed no significant differences in IL-10 (Fig. 2). Yet the donor-level analysis helped us identify important results associated with baseline IL-10 levels.

In the breast cancer study, stimulation with different peptide pools resulted in cytokine expression of very different magnitudes. Cytokine expression was normalized onto a scale of 0 to 1, for each donor and for each stimulation environment. The Wilcoxon signed rank test, a standard statistical test, was then applied to normalized data. The normalized data demonstrates that, compared to a CMV response, TAA responses show decreased IFN\(\gamma\) on CD4 and CD8 T cells and increased IL-2 on CD8 T cells. These statistically significant results have implications for the design of immunotherapies. Our analytical workflow was an essential component of identifying these results.
This novel approach is both incremental and iterative, allowing us to develop a solid understanding of our experimental results. Initial population-level analysis such as the scatter plot array in Figure 2 provides a succinct visual and statistical summary of an experiment. Such a summary serves as foundation for informed conversations among all members of the research team. All experimental readouts (e.g., 10 cytokines) are analyzed, preempting speculation about which readouts may or may not be relevant. In the next increment of analysis, donor profiles are inspected, leading to data transformation steps. Techniques such as donor stratification and data normalization are relatively easy to understand, particularly when compared with intricate machine learning algorithms. Furthermore, because these data transformations are selected based on patterns observed in donor profiles (e.g., stratification based on IL-10, as suggested by Fig. 3) and are followed by population-level analysis, we fully appreciate why a particular technical approach was selected. Accordingly, we are well prepared to evaluate the biological relevance of the results and to propose additional data transformations. Thus, this workflow offers a methodology for inspecting, interrogating, and appreciating the results of complex multiparameter clinical and translational studies.

In conclusion, we have demonstrated the power of an analytical workflow consisting of population-level analysis followed by donor-level analysis, data transformation, and then a return to population-level analysis. Furthermore, the analytical cycle can be repeated with different transformations. This workflow was not designed to support particular hypotheses, but rather to encourage and support the type of hypothesis-generation that results from EDA on large multiparameter data sets. In both presentations here examined, hypothesis-driven donor profiles led to a data transformation step appropriate for that particular data set. Subsequent population analysis tested the hypotheses generated by examination of the donor profiles. This workflow yields scientific insight not readily available from population-level analysis alone.

ACKNOWLEDGMENTS

The authors thank Dr. Peter Gottlieb of the Barbara Davis Childhood Diabetes Center, University of Colorado at Denver and Health Sciences Center, for providing T1D patient samples. Donor specimens for the breast cancer study were collected, in part, at the General Clinical Research Center Facility at the University of Washington (NIH M01-RR-00037).

LITERATURE CITED


