Supporting Information

Won et al. 10.1073/pnas.1112599109

SI Text

Problems with Median Fluorescence Intensity (MFI). If we are given a random sample from a probability distribution that is symmetric about a particular point, then that point is both the mean value and the median of the underlying distribution. Call it μ. One such symmetric distribution is the normal, or Gaussian, distribution. The sample mean value, the arithmetic mean ($\bar{X}$) of the cited random sample, is what is termed a complete, minimal sufficient statistic for μ (1). It follows from theoretical developments that in terms of minimizing variance there is no unbiased function of the random sample that is better than $\bar{X}$ (1). However, when the underlying distribution of the data is Gaussian, then the sample median of the random sample is a reasonably good estimator of μ, with theoretical variance only 1.57 times the theoretical variance of $\bar{X}$ (1). Further, the sample median is robust to outlying observations. Therefore, it may be slightly surprising that in comparing concentrations, as we do in this paper, one can do so much better than conventional sample medians do.

Huber's robust method of estimation, which we have applied to estimating $\{\beta_{ij}\}$ grew from his interest in estimating μ under a particular set of mixture models, which are Gaussian only when one component is absent. We discussed deviation from normality of bead data and observed that the sample mean is very sensitive to outliers or the scale of the distributions of bead measurements. We add that there are problems estimating the variance of a sample median because the variance is inversely proportional to the square of the underlying probability density evaluated at the population median (1). Estimating the value of a density, let alone the reciprocal of its square, is difficult. Indeed, no matter how one estimates the variance of the median, even for large sample sizes, a t-like statistic for comparing sample medians may not lend itself to comparisons, not least here, where sampling distributions are far from normal; and there are additional problems of scaling.

Transform of Fluorescence Measurements and Parameter Estimation. For different conditions, we do not assume equal variance because even after the transform the variance may still vary with i. However, we do assume a common variance across the repeats $j = 1,.....,R_i$ for condition i because they come from the same sample. We transform fluorescence intensities (FI) measurements using the following transformation:

$$T(\cdot) = \log(\cdot - M_{\mathrm{SB}} + s). \qquad \textbf{[S1]}$$

Here, M_{SB} is the pooled 5% trimmed mean blank measurements of the given cytokine, and s is a number that makes the internal term of the log positive for all k. M_{SB} roughly determines the precision of the measurement. Because blank (or buffer alone) measurements are standard in every experiment, it is convenient to use it to adjust the FI.

For parameter estimation, we use all the bead measurements per analyte to construct a distribution of (transformed) fluorescence values per condition. The model parameters μ, $\{\alpha_i\}$, $\{\beta_{ij}\}$, and s in Eq. **1** (main text) are estimated from the data in an iterative fashion. Suppose the values for the difference between the repeats $\{\beta_{ij}\}$ and the transform-related quantity s are fixed. Then Eq. **1** (main text) suggests a standard weighted least squares method for estimating the overall mean μ and the difference between conditions $\{\alpha_i\}$. That is, the error $T(y_{ijk} - \beta_{ij}) - (\mu + \alpha_i)$ between the observation and the model is squared and summed over $i = 0,1,.....,N$. In the sum, each error term is weighted inverse-proportionally to the empirical error standard deviation $\hat{\sigma}_i$, estimated from the residuals $T(y_{ijk} - \beta_{ij}) - (\hat{\mu} + \hat{\alpha}_i)$ of the unweighted least squares fit. Finally, μ and $\{\alpha_i\}$ that minimize the weighted sum is found. Once the values for μ and $\{\alpha_i\}$ are fixed, and the values for $\{\beta_{ij}\}$ are estimated using a nonlinear least squares method. This method requires a good initialization, for which we found satisfactory results with Huber robust regression (2) on repeats for each condition. (Often this initialization is good enough.) We estimate s based on the notion of profile likelihood (3). This process is iterated until convergence.

Selecting the Equivalence Margin Δ. This note deals with the question of how to choose the equivalence margin Δ. As Δ increases, smaller differences are called significant, leading to a loss of specificity and as Δ decreases only large differences are called significant, leading also to a loss of sensitivity. Because it is impossible to make an objective decision when the results are unknown (as is usually the case), it is desirable for us to let the data choose Δ. A seemingly attractive possibility is to use a data-driven machinery such as SAM (4). However, in these assays one does not usually have the luxury of hundreds of p-values such as in microarray analyses (for which SAM was designed). Instead, we choose Δ at a point in which the estimated power is reasonably high. As we increase Δ we note a sharp drop in the estimated power. This drop occurs as noise gives way to signal and is therefore where we wish to set Δ.

The power of the decision rule Eq. **3** (main text) is given as

$$\pi_i(\alpha_i - \alpha_0, \tau_i, \nu; \Delta) = 1 - \Pr\{T_{L,i} \geq t_{\alpha,\nu} \quad \text{and} \quad T_{U,i} \leq t_{1-\alpha,\nu} | \alpha_i - \alpha_0, \tau_i, \nu\}.$$

Under the assumption that $\hat{\alpha}_i - \hat{\alpha}_0$ follows a normal distribution $N(\alpha_i - \alpha_0, \tau_i)$, the vector $(T_{L,i}, T_{U,i})$ has a bivariate noncentral t-distribution with ν degrees of freedom and noncentrality parameters $\delta_{L,i}(\Delta) = (\alpha_i - \alpha_0 + \Delta)/\tau_i$ and $\delta_{L,i}(\Delta) = (\alpha_i - \alpha_0 - \Delta)/\tau_i$ (5).

We estimate the power at the estimated effect size; i.e., evaluate $\pi_i(\hat{\alpha}_i - \hat{\alpha}_0, s_i, \nu; \Delta)$. This estimated power is a nonincreasing function of Δ. For each case-control group and for each analyte, the largest Δ is chosen (Δ^*) below which the average estimated power (over $i = 1,.....,N$) is greater than a threshold. The median of the Δ^*s over the analytes is computed and denoted by $\bar{\Delta}^*$. Then the threshold is varied and the medians Δ^* are plotted. We find the inflection point of the plotted curve and use the value of the threshold that yields the inflection point. This threshold in turn determines Δ^* for each analyte in each case-control group (Fig. S7).

1. Lehmann EL (1983) *The Theory of Point Estimation* (John Wiley and Sons, Hoboken, NJ).
2. Huber PJ (1981) *Robust statistics* (John Wiley and Sons, Hoboken, NJ).
3. Venzon DJ, Moolgavkar SH (1988) A method for computing profile-likelihood-based confidence intervals. *J Roy Stat Soc C* 37:87–94.
4. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Nat'l Acad Sci USA* 98:5116–5121.
5. Phillips KF (1990) Power of the two one-sided tests procedure in bioequivalence. *J Pharmacokinet Phar* 18:137–144.

Fig. S1. Quantile-quantile plots of typical xMap bead fluorescence intensity (FI) data, shown for 21 standard analytes. Straight lines (red) represent the standard normal distribution.

Fig. S2. Typical scale of xMap bead fluorescence intensity data for eotaxin. The standard concentration levels are 1.22, 4.88, 19.53, 78.12, 312.5, 1,250, and 5,000 pg/mL; x4 diluted seven times. The range of data is depicted as the box-and-whisker plot. Also shown are the MFIs in repeat wells (blue dots) and the confidence interval (dashed curves) of the standard curve (solid curve) fitted for the MFIs using a 4PL model (done with *calib* package [http://cran.r-project.org/web/packages/calib/index.html] for R, The R Project for Statistical Computing).

Fig. S3. FPR and TPR of SAxCyB, in comparison with*t*-test (t_fullFI) and *t*-test (t_MFI) at the nominal significance level 0.01 with multiple comparison procedure (MCP) (*A*) and without MCP for 0.05 and 0.01 (inset) (*B*).

Cytokines that were found to be significantly different by MFI t-test but not by SAxCyB

Fig. S4. Cytokines found to be differentially expressed by MFI *t*-tests but not SAxCyB. Experiment and hypotheses testing was done as described in Fig. 3*C*. Hypotheses testing with *t*-tests on MFIs resulted in 12 cases that were significant ($p < 0.05$) by this method but not by SAxCyB. Fluorescence (top) and transformed fluorescence (bottom) measurements of individual beads for all cases. Dots are MFI values. Blue—healthy controls, green—day zero patients, red—month six patients. The data were overlaid for each well separately such that denser colors represent more measurements for that range.

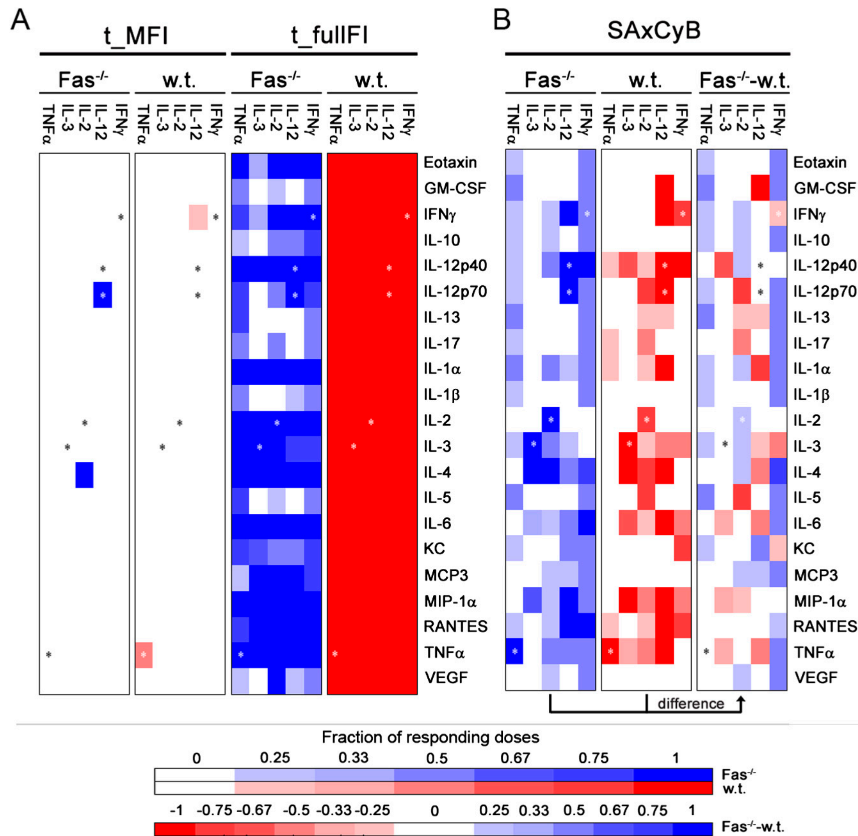

Fig. S5. Use of SAxCyB in analysis of cytokine stimulation assays. T cells from Fas$^{-/-}$ or wild type mice were treated with cytokines (listed above) and synthesis of cytokines (listed to the right) in these cells was measured. Treatments were done in four (IL-2, IL-12, IFNg, and TNFα) or three (IL-3) doses. Fraction of the responding doses is shown. Data were generated from 80 wells for samples and 16 wells for the standard curve. The 80 wells were divided into two sets of 40, where each set consists of one control (untreated) and 19 treatments in 4 or 3 doses, all technically repeated in duplicates. Twenty-one analytes were measured per well. Each well contained 5,791 ± 601 data points, each bead/well combination contained 276 ± 79 events. (*A*) Cytokine expression data was analyzed using median fluorescence intensity (MFI) and using all individual fluorescence data (t_MFI and t_fullFI respectively). (*B*) SAxCyB analysis of the same data and the difference matrix of Fas$^{-/-}$ minus (wild type). Asterisks mark cases where stimulus and response are the same cytokine (and therefore indistinguishable). SAxCyB calls 238 differences (at the 0.05 significance level) compared to 38 and 44 found by commercial software (BeadView and MasterPlex QT respectively). Many of the calls were for data at low MFIs.

Fig. S6. Magnified view of Fig. 2*B* around the zero.

Fig. S7. Illustration of Δ selection using the power estimated from the data. Estimated power is plotted as a function of Δ. Each color represents a difference analyte in the assay. Increasing the threshold from Threshold 3 to Threshold 4 yields the largest change in $\bar{\Delta}^*$ and Threshold 3 is selected.

Table S1. List of analytes used for the sensitivity analysis

Analyte	FPR	TPR	FPR	TPR	FPR	TPR
GM-CSF	0.0000	0.9789	0.0059	0.1354	0.3193	1.0000
ICAM-1	0.0000	0.8597	0.0068	0.1258	0.1433	0.9283
IFNα	0.0000	0.9727	0.0027	0.1288	0.2036	1.0000
IFNβ	0.0000	0.9513	0.0005	0.2031	0.2063	0.9895
IL-17	0.0000	0.9800	0.0000	0.1225	0.3088	1.0000
IL-17F	0.0000	0.9602	0.0000	0.1816	0.1878	0.9927
IL-12p40	0.0000	0.9765	0.0000	0.1841	0.2862	1.0000
IL-12p70	0.0000	0.9912	0.0018	0.1220	0.3147	1.0000
IL-10	0.0000	0.9903	0.0050	0.1548	0.3574	1.0000
IL-8	0.0000	0.9881	0.0000	0.2432	0.2209	1.0000
IL-7	0.0000	0.9708	0.0000	0.1091	0.4395	1.0000
IL-6	0.0000	0.9769	0.0000	0.1036	0.3696	1.0000
IL-5	0.0000	0.9731	0.0000	0.1575	0.3279	1.0000
IL-4	0.0000	0.9822	0.0009	0.1463	0.4068	1.0000
IL-1a	0.0000	0.9752	0.0000	0.1241	0.4526	1.0000
IL-1b	0.0000	0.9816	0.0009	0.2330	0.3610	1.0000
LIF	0.0000	0.9846	0.0087	0.1663	0.2685	1.0000
MCP-1	0.0000	0.9414	0.0009	0.1194	0.2721	1.0000
MIG	0.0000	0.9708	0.0045	0.1573	0.3188	1.0000
MIP-1α	0.0000	0.9394	0.0000	0.1973	0.1937	1.0000
PAI-1	0.0000	0.9782	0.0032	0.1335	0.3728	1.0000
RANTES	0.0000	0.9590	0.0000	0.0924	0.3125	1.0000
SCF	0.0000	0.9721	0.0005	0.1333	0.3143	1.0000
sFas-L	0.0000	0.9782	0.0000	0.0857	0.3755	1.0000
TGFa	0.0000	0.9414	0.0032	0.1636	0.3175	1.0000
TGF-b	0.0000	0.9353	0.0009	0.2168	0.3719	1.0000
TNFα	0.0000	0.9897	0.0000	0.1239	0.3125	1.0000
TNFβ	0.0000	0.9915	0.0027	0.1791	0.2485	1.0000
TRAIL	0.0000	0.9614	0.0005	0.2054	0.3052	0.9943
VCAM-1	0.0000	0.9433	0.0000	0.0641	0.3370	0.9773
VEGF	0.0000	0.9702	0.0000	0.0933	0.2277	1.0000
Eotaxin	0.0002	0.9435	0.0028	0.1176	0.2236	0.9527
HGF	0.0005	0.9630	0.0027	0.0970	0.3370	0.9989
sCD40L	0.0007	0.9384	0.0018	0.1189	0.3342	0.9711
M-CSF	0.0007	0.9699	0.0000	0.1831	0.3565	1.0000
FGF-Basic	0.0011	0.9371	0.0014	0.0995	0.2884	0.9551
MIP-1β	0.0011	0.9569	0.0082	0.2043	0.2812	0.9986
IP-10	0.0014	0.9576	0.0009	0.1530	0.3288	1.0000
GROα	0.0020	0.9034	0.0014	0.1742	0.2358	0.9890
Resistin	0.0023	0.9217	0.0000	0.0683	0.4005	0.9700
ENA-78	0.0029	0.8854	0.0014	0.1113	0.2458	0.9426
IL-13	0.0041	0.9774	0.0018	0.1487	0.2653	1.0000
IL-15	0.0057	0.9368	0.0050	0.1186	0.2073	0.9624
Leptin	0.0066	0.8722	0.0023	0.1192	0.2780	0.9277
G-CSF	0.0084	0.9402	0.0009	0.1637	0.2875	0.9583
IL-1Ra	0.0091	0.8913	0.0005	0.0716	0.3433	0.9393
IFNγ	0.0111	0.9588	0.0000	0.1365	0.2785	0.9927
MCP3	0.0218	0.9468	0.0000	0.0847	0.3220	0.9682
IL-2	0.0240	0.9784	0.0018	0.0654	0.4408	1.0000
NGF	0.0247	0.9162	0.0005	0.1222	0.3324	0.9477
PDGF-BB	0.0683	0.9879	0.0077	0.1223	0.3501	1.0000

Analytes are sorted by the FPR of SAxCyB. All numbers are MCP-adjusted at the nominal significance level 0.05. In the shaded area are the analytes whose SAxCyB FPR is greater than 0.05. Also shown are a two-sample *t*-test ("t_MFI") that employs only MFIs (Median Fluorescence Intensities; therefore two measurements for each instance) and a two-sample *t*-test ("t_fullFI") that employs all bead measurements and pools repeats. The first reference is a common analysis method of xMap data and the second is a naïve use of all the individual bead measurements