



## OPEN ***rx*COV is a quantitative metric for assessing immunoassay analyte fidelity**

Rhonda M. Brand<sup>1,2</sup>, Danielle Pitlor<sup>3</sup>, E. Jeffrey Metter<sup>4</sup>, Beth Dudley<sup>1</sup>, Eve Karloski<sup>1</sup>, Ashley Zyhowski<sup>2</sup>, Randall E. Brand<sup>1,5</sup> & Shikhar Uttam<sup>5,6</sup>✉

Immunoassay based bioanalytical measurements are widely used in a variety of biomedical research and clinical settings. In these settings they are assumed to faithfully represent the experimental conditions being tested and the sample groups being compared. Although significant technical advances have been made in improving sensitivity and quality of the measurements, currently no metrics exist that objectively quantify the fidelity of the measured analytes with respect to noise associated with the specific assay. Here we introduce *ratio of cross-coefficient-of-variation (rxCOV)*, a fidelity metric for objectively assessing immunoassay analyte measurement quality when comparing its differential expression between different sample groups or experimental conditions. We derive the metric from first principles and establish its feasibility and applicability using simulated and experimental data. We show that *rxCOV* assesses fidelity independent of statistical significance, and importantly, identifies when latter is meaningful. We also discuss its importance in the context of averaging experimental replicates for increasing signal to noise ratio. Finally, we demonstrate its application in a Lynch Syndrome case study. We conclude by discussing its applicability to multiplexed immunoassays, other biosensing assays, and to paired and unpaired data. We anticipate *rxCOV* to be adopted as a simple and easy-to-use fidelity metric for performing robust and reproducible biomedical research.

Biosensing assays such as antibody based enzyme-linked immunoabsorbent assay (ELISA) and its multiplexed counterpart (mELISA), are ubiquitously used methods that utilize antigen-antibody binding to determine with high sensitivity, concentration of analytes in a wide variety of bioanalytical settings<sup>1</sup>. Their ease-of-use coupled with improvements in measurement technology and quality control<sup>2,3</sup> has made these immunoassays a mainstay in diverse biomedical research and its clinical applications<sup>4,5</sup>. More recently, nucleic acid-based probes, such as aptamers have also been developed<sup>6,7</sup>. Aptamers are short single-stranded RNA or DNA sequences that can replace antibodies in ELISA to bind targets of interest with improved specificity and affinity<sup>6-8</sup>. The resulting enzyme-linked aptamer assay (ELAA) has the potential to provide fast, low-cost, and improved target detection alternatives to ELISA<sup>8</sup>. However, measurement of target expression in these biosensing assays is fundamentally confounded by intrinsic biological heterogeneity, variations in preanalytical variables between different sample groups, variability is user-dependent assay implementation, and imperfection of the assay itself due to varying degrees of target-dependent cross-reactivity and affinity<sup>9-12</sup>. These confounding factors together combine to generate stochastic noise that we refer to as assay-associated noise. Despite improvements both in the quality and biochemistry of biosensing technologies, assay-associated noise continues to confound analyte measurements. This noise is especially important to account for when comparing sample groups or experimental conditions with analyte expression in low signal to noise regime, or when differences in expression between the comparison groups are small. In such scenarios comparison of analyte expression between the two groups in the background of assay-associated noise can falsely suggest significance, or conversely the lack of it. Therefore, it is essential to establish the validity of differential analyte expression with respect to assay-associated noise, which we refer to as analyte fidelity. Establishing analyte fidelity is necessary for subsequent analyses to be meaningful.

<sup>1</sup>Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA. <sup>2</sup>Magee Women's Research Institute, University of Pittsburgh, Pittsburgh, PA, USA. <sup>3</sup>Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA, USA. <sup>4</sup>Department of Neurology, University of Tennessee, Memphis, TN, USA. <sup>5</sup>UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA, USA. <sup>6</sup>Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA, USA. ✉email: shf28@pitt.edu

Currently, protocol-level experimental techniques are used to ameliorate these assay-associated noise effects by carefully implementing best practices in assay design<sup>3</sup>. Additionally, repeat measurements can be performed to average out noise effects and improve the analyte signal to noise ratio (SNR). However, currently no quantitative metric exists to ascertain if the SNR has sufficiently improved, or how many experimental replicates are sufficient for each individual experiment to ensure good SNR. As a result, researchers use an ad-hoc number, usually justified based on logistics, cost or accepted practice. This can lead to inconsistent findings and reduced translational efficacy, particularly in research where reproducibility is a key criterion such as in biomarker discovery and drug development. Furthermore, in mechanism-focused basic research, lack of analyte fidelity can validate incorrect hypotheses resulting in spurious conclusions.

To overcome these challenges, we have developed a simple and easy-to-use quantitative metric that objectively characterizes analyte fidelity when comparing different sample groups or treatment conditions in individual experiments. As a result, it is also capable of identifying spurious statistical significance. We have developed the metric from first principles and show that it naturally emerges as a ratio of two modified forms of coefficient-of-variation (COV) statistic (see Eq. 2c). We, therefore, refer to it as *ratio of cross-Coefficient-of-Variation* (*rxCOV*). COV itself has been used as a unit-free interpretable measure of precision in assessing reliability and repeatability of collected data in a range of disciplines including biology and medicine<sup>13–15</sup>. The two modified COV terms together quantify the relative effect size of the differential analyte expression with respect to the assay-associated noise. We show that this relative effect size objectively characterizes analyte fidelity. Furthermore, we demonstrate that *rxCOV* can determine whether the number of experimental replicates is sufficient to ensure good SNR and prevent spurious findings. Finally, as the fidelity metric is computed separately for each analyte and for any two sets of sample groups or conditions, it can be easily computed in parallel for multiple individual analytes in a multiplexed setting. It may also be extended to measure analyte fidelity between multiple different sample groups or experimental conditions.

## Results

**Fidelity metric: *rxCOV*.** We derive *rxCOV* via a two step process. We first define a function that quantifies relative change between the differential analyte expression and the measurement noise associated with the assay. We next use this function to derive the fidelity metric.

*Function of relative change.* Let random variables  $X$  and  $Y$  respectively represent the expressions of an analyte of interest in samples belonging to two different patient groups or experimental conditions being compared. And let  $X'$  and  $Y'$  represent their repeat measurements on aliquots drawn from the same samples. The difference between the two sets of measurements on the same sample, defined as,  $N_X = |X - X'|$  and  $N_Y = |Y - Y'|$  represents the assay-associated variations in analyte expression. To ensure robustness, we combine  $N_X$  and  $N_Y$  into a single worst-case-scenario random variable  $N = \max(N_X, N_Y)$ . The differential analyte expression between the two samples is given by  $Z = |X - Y|$ . By construction,  $N$  and  $Z$  are ratio-scale variables.

For differential expression  $Z$  to represent a valid difference in the analyte expression between the two groups—whether statistically significant or not—its effect size should be greater than that of  $N$ . That is, the effect, or magnitude of variation in the differential analyte expression between the groups should be greater than the magnitude of variation in assay-associated noise. We note that this formulation is distinct from common metrics of effect-size such as Cohen's  $d$ <sup>16</sup>. They are typically used to quantify the effect-size of differential analyte expression between the sample groups, particularly in the context of power analysis. Our formulation, on the other hand, aims to assess the *validity* of the differential analyte expression computed from measurements made in presence of assay noise. It, therefore, is fundamental to all downstream statistical analyses including power analysis.

Mathematically we can express the above relation as  $\mathcal{M}(Z) \geq \mathcal{M}(N)$ , where  $\mathcal{M}$  denotes a measure of the effect-size. Re-expressing it as  $\log_{10} \frac{\mathcal{M}(Z)}{\mathcal{M}(N)} \geq 0$ , we obtain the fidelity condition in terms of the log-based function of relative change:  $f(\mathcal{M}(Z), \mathcal{M}(N)) = \log_{10} \frac{\mathcal{M}(Z)}{\mathcal{M}(N)}$ , which satisfies the following important properties<sup>17</sup>,

$$f(\mathcal{M}(Z), \mathcal{M}(N)) = 0 \Leftrightarrow \mathcal{M}(Z) = \mathcal{M}(N), \quad (1a)$$

$$f(\mathcal{M}(Z), \mathcal{M}(N)) > 0 \Rightarrow \mathcal{M}(Z) > \mathcal{M}(N), \quad (1b)$$

$$f(\mathcal{M}(Z), \mathcal{M}(N)) < 0 \Rightarrow \mathcal{M}(Z) < \mathcal{M}(N), \quad (1c)$$

$$f(\alpha \mathcal{M}(Z), \alpha \mathcal{M}(N)) = \alpha f(\mathcal{M}(Z), \mathcal{M}(N)), \forall \alpha > 0, \text{ and} \quad (1d)$$

$$f(\mathcal{M}(Z), \mathcal{M}(N)) \text{ is a continuous and monotonically increasing function of } \mathcal{M}(Z) \text{ when } \mathcal{M}(N) \text{ is fixed,} \quad (1e)$$

As a result,  $f$  defines a well-behaved general measure of fidelity that is independent of measurement units and increases or decreases based on whether the effect size of  $Z$  is greater or less than  $N$ . Since the log function satisfies the properties  $\log(1) = 0$  and  $\log(pq) = \log(p) + \log(q), \forall p, q > 0$ <sup>18</sup>, we can naturally extend  $f$  to combine different measures of relative change into a single log-based sum of relative changes. This extension will also satisfy Eq. (1) properties. We next use this extension to define *rxCOV* fidelity metric as a function of first- and second-order measures of effect-size.

**rxCOV.** We consider mean ( $\mu$ ) and standard deviation ( $\sigma$ ) as two measures  $\mathcal{M}_1(\cdot)$  and  $\mathcal{M}_2(\cdot)$  respectively of the effect size of  $Z$  and  $N$ . The resulting functions of relative change  $f(\mu_Z, \mu_N)$  and  $f(\sigma_Z, \sigma_N)$  respectively quantify the *average* and *dispersion* fidelity of the differential analyte expression between the two sample groups, with respect to the assay-associated noise. As an example,  $f(\mu_Z, \mu_N) = 0$  and  $f(\sigma_Z, \sigma_N) = 0$  imply that effect-sizes related to mean (first-order) and dispersion (second-order) of differential analyte expression are the same as that of assay-associated noise. As a result, differences in analyte expression between the two groups or conditions cannot be separated from differences arising due to assay-associated variability. Consequently, the analyte has low-fidelity for this specific assay.

Utilizing the additive property of  $f$ , we combine the first- and second-order fidelity measures of effect size and define our joint fidelity metric *rxCOV* as

$$rxCOV(Z, N) = \log_{10} \frac{\mathcal{M}_1(Z)}{\mathcal{M}_1(N)} + \log_{10} \frac{\mathcal{M}_2(Z)}{\mathcal{M}_2(N)}, \quad (2a)$$

$$rxCOV(Z, N) = \log_{10} \frac{\mu_Z}{\mu_N} + \log_{10} \frac{\sigma_Z}{\sigma_N}, \quad (2b)$$

$$rxCOV(Z, N) = \log_{10} \frac{(\sigma_Z/\mu_N)}{(\sigma_N/\mu_Z)}. \quad (2c)$$

Equation (2c) has an intuitive interpretation: It is a *standardized* measure of dispersion of differential analyte expression between two sample groups, relative to dispersion in assay associated noise. The standardization of  $\sigma_Z$  by  $\mu_N$ , and  $\sigma_N$  by  $\mu_Z$ , ensures that the comparison of relative dispersion is adjusted to the correct scale. It also motivates the name of the fidelity metric *ratio of cross-Coefficient-of-Variations (rxCOV)*. Here, ‘cross’ refers to standardization of  $\sigma_Z$  by  $\mu_N$ , and of  $\sigma_N$  by  $\mu_Z$ . If  $rxCOV(Z, N) > 0$ , then the range of variation in the differential analyte expression after having been adjusted by average assay-associated noise dominates the variation in assay-associated noise adjusted by the mean differential analyte expression. Consequently, the analyte has high fidelity. If, on the other hand,  $rxCOV(Z, N) \leq 0$ , then the analyte has low fidelity.

We emphasize that Eq. (2c) is easy to implement. As the equation shows, computation of *rxCOV* is straightforward. It requires computing four simple statistics: mean and standard deviation of differential analyte expression ( $Z$ ) and assay-associated noise ( $N$ ) measurements. This can be done using any rudimentary statistical software, thereby making *rxCOV* an easy-to-use metric for researchers from diverse backgrounds.

**rxCOV is a smooth function of its variables.** The *rxCOV* metric, given by Eq. (2c), is a function of four variables  $\mu_Z, \mu_N, \sigma_Z, \sigma_N$ , defined on the open region  $\mathcal{D} = (0, \infty)^4 \subset \mathcal{R}^4$ . Its partial derivative with respect to each of the four variables is a hyperbola, a continuous function over the domain  $(0, \infty)^4$ . As a result, *rxCOV* is differentiable over  $\mathcal{D}$ , which in turn ensures that it is also a continuous function over  $\mathcal{D}$ . Furthermore, with hyperbola as the partial derivative of first-order for each of the four variables, the existence of partial derivatives of *rxCOV* of any order is guaranteed. Therefore, *rxCOV* is not only a continuous function over  $\mathcal{D}$ , but it is also a smooth function over it. This property ensures that the small changes in  $\mu_Z, \mu_N, \sigma_Z, \sigma_N$  result in small changes in *rxCOV*. Thus, *rxCOV* can smoothly track analyte fidelity across the whole range of differential analyte expression.

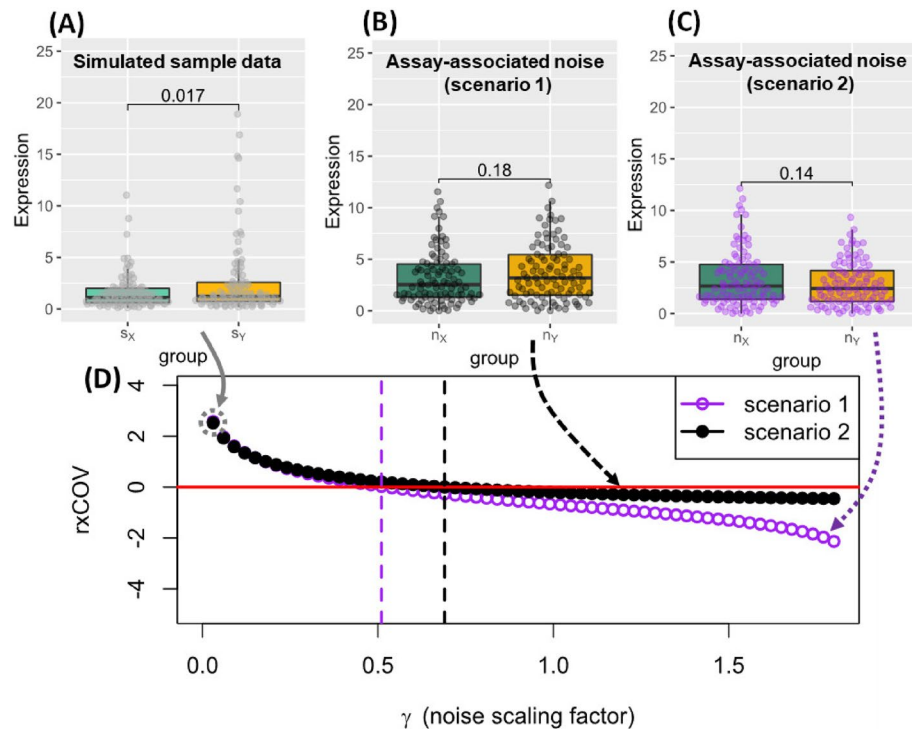
**rxCOV is an objective fidelity metric.** We constructed a dataset comprising simulated expression of a cytokine in two samples groups, denoted by  $s_X$  and  $s_Y$  (Fig. 1A). The expressions are chosen such that the differential expression between the two sets is significant at the 95% confidence level, with a p-value of 0.017. Two scenarios based on two different assay-dependent noise profiles ( $n_X$  and  $n_Y$ ) are considered (Fig. 1B,C). The base noise profiles, however, are not statistically significant at the 95% confidence level. The final dataset for the two scenarios is constructed as follows,

$$d_X = s_X + \gamma n_X, \quad (3a)$$

$$d_Y = s_Y + \gamma n_Y. \quad (3b)$$

Here  $\gamma$  scales the base noise profile to recapitulate a range of assay-dependent noise scenarios, without affecting the relative significance of the noise profiles themselves within each of the two scenarios. The resulting plots of *rxCOV* as a function of  $\gamma$  for the two scenarios are shown in Fig. 1D. The color coded dashed vertical lines indicate where *rxCOV* curve reaches zero. At this zero threshold, the effect-sizes of differential cytokine expression between the two sample groups and that of the assay associated noise with respect to mean and dispersion are the same and cannot be distinguished from each other. Above this threshold the effect size of the differential expression dominates the noise and the analyte has high fidelity, while below this threshold the converse is true and the noise dominates. Therefore, *rxCOV* is an experiment-specific metric that objectively assesses analyte fidelity via the zero-level threshold.

**rxCOV identifies spurious statistical significance.** In Fig. 2A, along with the plot of *rxCOV* as a function of  $\gamma$  for the two scenarios shown in Fig. 1D, we also plot the p-value corresponding to statistical significance of the differential expression between  $d_X$  and  $d_Y$  as a function of  $\gamma$ . In this plot the solid red line indicates statistical significance at the 95% confidence level.

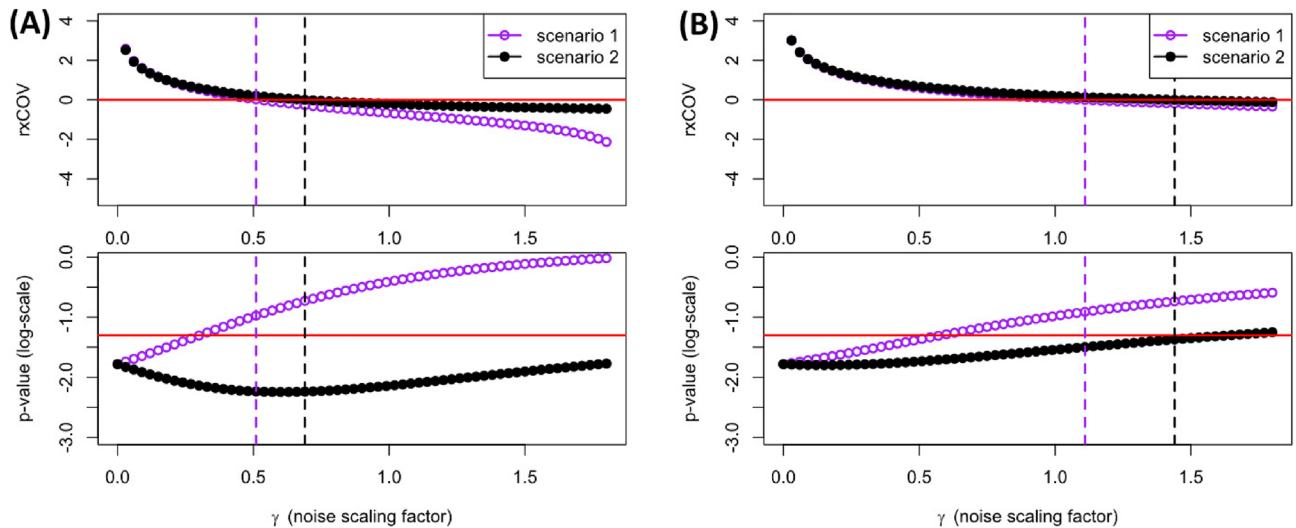


**Figure 1.** Simulated example of  $rxCOV$  as an objective fidelity metric. **(A)** Analyte data simulating hypothetical cytokine expression in two sample groups  $s_X$  and  $s_Y$ . The difference in cytokine expression is significant at the 95% confidence level. **(B)** Baseline scenario 1 representing measurement noise associated with assaying each sample of each group. **(C)** Baseline scenario 2 representing a slightly different assay-associated noise distribution than scenario 1. Data  $d_X$  and  $d_Y$  are generated by scaling these baseline scenarios by the scaling factor  $\gamma$  and adding them to  $s_X$  and  $s_Y$  as formulated in Eq. (3). Assay-associated noise in both scenarios is not significant at the 95% confidence level and remains unaffected by the scaling factor  $\gamma$ . **(D)**  $rxCOV$  plot as a function of  $\gamma$  for Scenarios 1 and 2. When  $\gamma = 0$ ,  $d_X = s_X$  and  $d_Y = s_Y$ , no assay noise is assumed, and the two scenarios coincide. This is indicated by the dashed gray circle. The solid red line corresponds to  $rxCOV = 0$ , the objective fidelity threshold. The purple and black dashed vertical lines correspond to the noise levels at which the  $rxCOV$  plots intersect the solid red line, and indicate the low-fidelity analyte threshold for the two respective scenarios.

Scenario 1 reflects a situation where increasing noise level gradually washes out the significance of the data. Here, the differential expression becomes insignificant before  $rxCOV$  reaches 0, as indicated by the purple dashed vertical line. This result indicates that although, the first- and second-order effect sizes of the differential expression between  $d_X$  and  $d_Y$  are greater than that of assay-associated noise, they are not large enough to ensure statistical significance at the 95% confidence level. Once  $rxCOV$  reaches 0, the assay-associated noise dominates and statistical significance becomes spurious.  $rxCOV$ , therefore, establishes the sufficient condition for the validity of the p-value in assessing statistical significance of the differential analyte expression. Specifically, if  $rxCOV \leq 0$  then any p-value, whether  $\leq 0.05$  or  $\leq 0.001$  is a noise related artifact and is meaningless.

Scenario 2 illustrates a situation where the differential expression between  $d_X$  and  $d_Y$  remains significant beyond the black dashed vertical line, where  $rxCOV$  reaches zero, and the first- and second-order effect size of assay-associated noise begin overwhelming the differential analyte signal.  $rxCOV$  identifies this significance to be spurious. Importantly, it exemplifies how a small stochastic change in the noise distribution from that of scenario 1 can result in wide divergence in p-value behavior of the differential analyte expression, thereby emphasizing the critical importance of a fidelity metric.

**$rxCOV$  and experimental replication.** A common approach for improving the fidelity of analyte expression is to average multiple independent replicates. Figure 2B demonstrates that  $rxCOV$  fidelity metric is able to capture the resulting increase in noise ratio via the rightward shift of the purple and black dashed vertical lines corresponding to  $rxCOV = 0$ . In it, the same two scenarios presented in Fig. 2A are considered, but instead of a single measurement, three independent replicates of  $d_X$  and  $d_Y$  are generated and averaged. As expected, the  $rxCOV$  and p-value plots corresponding to the two scenarios as a function of  $\gamma$  converge towards each other. Furthermore, the values of  $\gamma$  at which  $rxCOV$  reaches zero increases, marked by the rightward shift of the dashed vertical lines, and indicates the averaging-dependent increase in noise range over which the signal maintains its fidelity with respect to assay-associated noise. Interestingly, despite the benefit of data averaging, the risk of spurious significance is not completely removed. This is evident in scenario 2, where p-value remains below 0.05 immediately after  $rxCOV$  reaches 0. Therefore, even though averaging ameliorates noise effect, it



**Figure 2.** Simulated example of  $rxCOV$  identifying spurious statistical significance and capturing effect of experimental replication. (A) Fig. 1D is extended to include the relation between  $rxCOV$  and p-value as a function of  $\gamma$  for the two noise scenarios presented in Fig. 1B,C. The solid red line in the p-value plot (bottom panel) indicates p-value of 0.05. The dashed vertical lines corresponding to the two scenarios continue over from the top panel and correspond to the value of  $\gamma$  where  $rxCOV$  reaches zero. They establish the noise level beyond which p-value based computational of statistical significance becomes spurious. (B) Three replicates of  $s_X$  and  $s_Y$  are generated from the same distribution used in Fig. 1A. Corresponding values of  $d_X$  and  $d_Y$  are computed and averaged to increase analyte SNR. The rightward shift of the purple and black dashed vertical lines coupled with the convergence of the  $rxCOV$  plots for the two scenarios demonstrates the ability of  $rxCOV$  to capture the increased SNR due to data averaging.

remains difficult to ascertain if the number of replicates are enough to overcome the noise level of the specific assay. In such a situation,  $rxCOV$  metric can provide the answer: if the  $rxCOV$  value resulting from using averaged replicates is  $> 0$ , the number of replicates is sufficient. If, on the other hand, it is  $\leq 0$ , then more replicates are required. Importantly,  $rxCOV$  metric can also be used to decide for which analytes data averaging is not needed, particularly in multiplexed scenarios where replication can be limited by cost, sample, and logistical constraints.

**$rxCOV$  is a fidelity metric for paired and unpaired data.** Rewriting the definition of  $Z = |X - Y|$  in terms of the lattice identity  $Z = |X - Y| = \max(X, Y) - \min(X, Y)$ <sup>21</sup> shows that this definition depends on pairwise max and min operations and assumes paired data. For unpaired data, we redefine  $Z$  as  $Z^{up} = X - Y$ , with its mean and standard deviation respectively given by  $\mu_Z^{up} = \mu_X - \mu_Y$  and  $\sigma_Z^{up} = \sqrt{\sigma_X^2 + \sigma_Y^2}$ . The corresponding measures of effect size, then are  $\mathcal{M}_1^{up}(Z) = |\mu_Z^{up}|$  and  $\mathcal{M}_2^{up}(Z) = \sigma_Z^{up}$ , and we can use Eq. (2c) to compute analyte fidelity. We note that the first- and second-order measures of effect-size of assay-dependent noise are always paired because, by experimental design, assay-associated expression variability is quantified via repeated measurement of analyte expression of the same sample. The paired or unpaired status of the data itself is determined by the study and experimental design and is not a matter of preference. Nevertheless, in order to develop a comprehensive understanding of the fidelity metric, we computed the deviation in  $rxCOV$  value for the same data under paired and unpaired assumptions. We first defined difference in the mean and standard deviation via the following relations,

$$\mu_Z^{up} = \mu_Z + \Delta\mu_Z, \Delta\mu_Z \leq 0, \quad (4a)$$

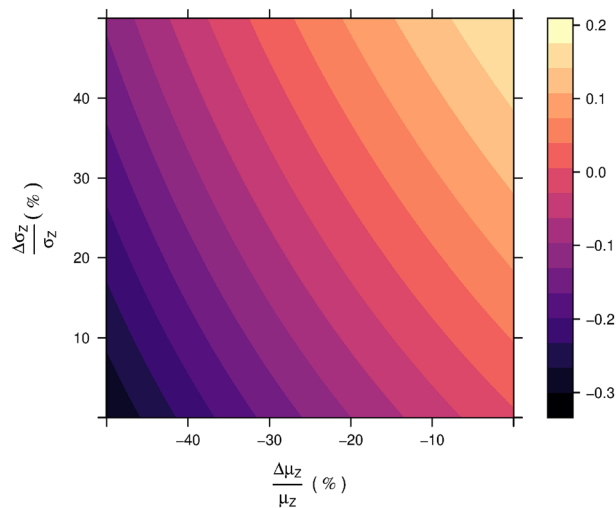
$$\sigma_Z^{up} = \sigma_Z + \Delta\sigma_Z, \Delta\sigma_Z \geq 0, \quad (4b)$$

with the inequalities  $\Delta\mu_Z \leq 0$  and  $\Delta\sigma_Z \geq 0$  following directly from the lattice identity representation of  $Z$ . The relative difference in the  $rxCOV$  fidelity metric between the paired and unpaired assumption can then be expressed as,

$$\Delta rxCOV = \log_{10} \left( 1 + \frac{\Delta\mu_Z}{\mu_Z} \right) + \log_{10} \left( 1 + \frac{\Delta\sigma_Z}{\sigma_Z} \right), \quad (5)$$

with  $\Delta\mu_Z \leq 0, \Delta\sigma_Z \geq 0$ .

Here,  $\frac{\Delta\mu_Z}{\mu_Z}$  and  $\frac{\Delta\sigma_Z}{\sigma_Z}$  respectively capture the relative difference between the paired and unpaired assumption for  $\mu$  and  $\sigma$  with respect to the paired assumption. Figure 3 shows that for a broad range of deviation, from 0% through



**Figure 3.** Relative change in  $rxCOV$  under the paired vs. unpaired assumptions. Plot of  $\Delta rxCOV$  as a function of relative deviation in mean and standard deviation resulting from differences due to the paired and unpaired assumption. The deviation in relative mean value ranges from 0% through 50%, while that in standard deviation ranges from  $-50\%$  through 0%. The colorbar indicates the corresponding deviation in  $\Delta rxCOV$ .

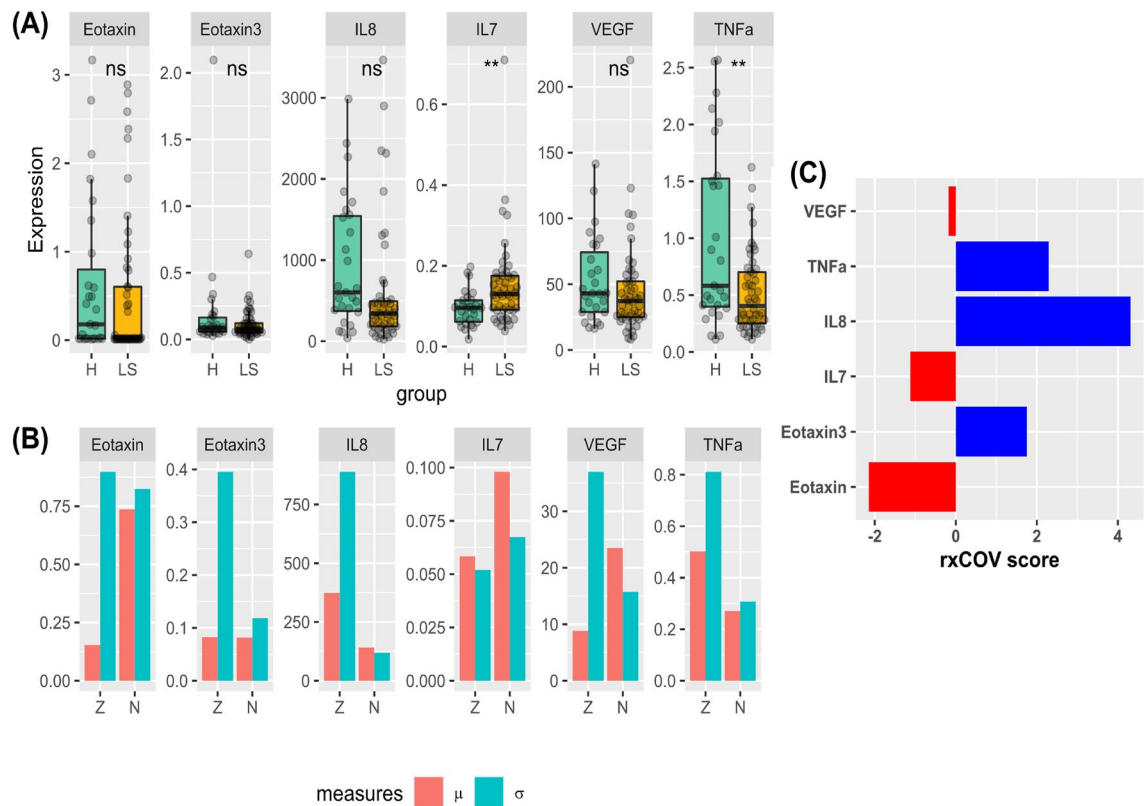
50%,  $\Delta rxCOV$  is bounded within the range  $(-0.301, 0.2)$ , implying that although  $rxCOV$  is sensitive to the paired and unpaired assumption, it remains a relatively stable and bounded metric, for a large range of variation.

**Case study: establishing the fidelity of immune signaling profile in patients with Lynch syndrome.** Figure 4A shows a subset of multiplexed ELISA expression data collected from healthy volunteers (H) and patients with Lynch syndrome (LS) for six chemotactic, proinflammatory and immune suppressive cytokines (see “Materials and methods” section). The differential expression of the six cytokines between healthy volunteers and LS patients captures differences in the immune signaling microenvironment of these groups of patients. The figure also indicates the significance of this difference at the 95% confidence level. Figure 4B plots the mean and standard deviation of differential expression ( $Z$ ) and assay-associated noise ( $N$ ) for the same set of cytokines. The figure makes it evident that despite paying careful attention to experimental protocols that help ameliorate the effects of assay-associated noise, for some cytokines assay-associated noise still overwhelms the differential expression. We note that the interplay between signal and noise is sample and assay dependent. The importance of  $rxCOV$  is in its ability to objectively quantitate this sample and assay dependent interplay and establish the fidelity of each individual cytokine. Figure 4C shows that  $rxCOV$  fidelity metric is able to identify high and low fidelity cytokines. Interestingly, although we see that differential expression of both IL7 and TNF $\alpha$  is statistically significant, the  $rxCOV$  fidelity metric shows that for IL7 this significance is spurious ( $rxCOV < 0$ ) and is an assay-associated artifact. On the other hand, differential expression of TNF $\alpha$  is significant ( $rxCOV > 0$ ). Importantly,  $rxCOV$  also indicates where lack of significance might be an artifact. For example, although both Eotaxin and Eotaxin3 are not statistically significant, this lack of significance for Eotaxin could potentially be an experimental artifact ( $rxCOV < 0$ ), while for Eotaxin3 it is not ( $rxCOV > 0$ ). Thus, these results demonstrate the practical way in which  $rxCOV$  metric can guide experimental decisions in real-time.

## Discussion

We have introduced  $rxCOV$ , a simple and easy-to-use quantitative metric for assessing the fidelity of immuno-assay based analyte measurements comparing different treatment conditions or sample groups. As shown in Eq. (2c),  $rxCOV$  only requires computation of mean and standard deviation of the differential expression ( $Z$ ) and assay-associated noise ( $N$ ).  $Z$  is obtained from assay measurements of the samples under study.  $N$  is also simple to compute and only requires an additional measurement of an aliquot of the same set of samples, as exemplified by the case study. The computations can be easily performed using a low-end computer equipped with a statistical software capable of performing subtraction, and computation of mean and standard deviation of expression values from a sample group. Therefore,  $rxCOV$  is accessible to researchers in diverse research, translational and clinical settings with varying computational expertise. Importantly, unlike other common and important metrics of effect size such as Cohen’s  $d$  and its variations, that are used to quantify the size of the effect—that is, the magnitude of the differential analyte expression—between sample groups or treatment conditions,  $rxCOV$  assesses the validity of the differential analyte expression in presence of assay noise. Additionally, it establishes a sufficient condition for the validity of p-value and statistical significance associated with the differential analyte expression. Thus, it is fundamental to all downstream analyses and interpretation.

The strength of  $rxCOV$  is its objective threshold of zero that demarcates the boundary between experiment-specific analyte fidelity and the lack of it. From the construction of the metric, it is evident that zero is the correct fidelity threshold because it identifies the point of balance between the differential analyte expression and the noise embedded in the measurement of analyte expression for a specific experiment. For  $rxCOV$  values  $> 0$ ,



**Figure 4.** Case study. (A) Boxplot representation of multiplexed ELISA expression of six cytokines measured in Lynch Syndrome (LS) patients with healthy (H) patients acting as controls. (B) Barplots of the mean and standard deviation of the cytokine differential expressions (Z), and their corresponding assay-associated noise (N). (C)  $rxCOV$  metric computed by substituting the mean and standard deviation of the differential expression, shown in (B), for each of the six cytokines. Those with  $rxCOV > 0$  are high-fidelity analytes and are shown in blue. Those with  $rxCOV \leq 0$  are low-fidelity analytes and are shown in red.

the differential signal is stronger than the noise, while it is the opposite for  $rxCOV < 0$ . Using the shift in this threshold, we demonstrated the sensitivity of  $rxCOV$  to averaging experimental replicates, an important validation of the metric being able to measure analyte fidelity. Importantly, it identifies another use of  $rxCOV$  as an objective criterion in determining the optimal number of replicates needed to ensure signal fidelity of an analyte.

$rxCOV$  uses mean and standard deviation summary statistics to quantify the effect-size of properly normalized differential analyte expression and its noise (Eq. 2c). However,  $rxCOV$  is not limited to the use of these statistics in two ways. First, if the mean and standard deviation are skewed by the nature of the study data, they can respectively be replaced by median and interquartile range (IQR) that could be more robust to outliers<sup>22</sup>. Second,  $rxCOV$  can be naturally extended to incorporate additional summary statistics, such as entropy, based on the additive property of the function of relative change as shown in Eq. (2a). However, use of additional statistics should be supported by a rationale relevant for the study and the nature of its data. Typically, our use of first- and second-order summary statistics should be sufficient and accurate in most cases.

For computing summary statistics of assay noise within a sample group or treatment condition associated with measurement of analyte expression,  $rxCOV$  requires a single repeat measurement per sample. This modest requirement follows from the observation that variations in repeat measurements of the same sample fundamentally capture stochastic and assay-associated noise, and are independent of the sample itself. Thus, single repeat measurement per sample, for all samples in the study, provides enough stochastic diversity to reliably compute noise summary statistics. The same summary statistics can also be computed using repeat measurements ( $> 1$ ) of aliquots drawn from a subset of samples. The optimal choice between these range of options is dependent upon sample constraint specific to the study. Computation of  $rxCOV$  can, therefore, be adapted to the constraints and logistics of individual research studies, thereby suggesting its wide applicability.

As  $rxCOV$  is computed per analyte, it can be easily computed for multiple analytes measured in parallel, for example, in the case of multiplexed immunoassays. We demonstrated this generalizability to a multiplexed setting in the case study presented above. There  $rxCOV$  was used to simultaneously assess the fidelity of all six analytes. Additionally, its pairwise computation for any two—paired or unpaired—treatment conditions or sample groups, can also be easily extended to multiple experimental or sample group comparisons, each implemented pairwise. Importantly though, given its simple formulation, increase in number of analytes and sample group comparisons will not unduly increase the computational complexity of  $rxCOV$  based fidelity analysis as  $rxCOV$  only requires computation of four values—two mean and two standard deviation values—per comparison.

$rxCOV$  can also be used with other biosensing assays beyond antibody-based immunoassays. As noted in the Introduction, aptamer-based ELAA assays replace antibody based detection with aptamer probes. Nevertheless, like ELISAs, ELAA measurements are usually based on calorimetric, fluorescence, or chemiluminescence principles, ensuring direct applicability of  $rxCOV$ <sup>8</sup>. Beyond aptamer-based assays,  $rxCOV$  can also be applied to other types of biosensing assays where the measurements are ratio scale variables, that is, analyte expression value of zero is well-defined and biochemically meaningful.

## Conclusion

We have introduced  $rxCOV$ , a simple, quantitative metric to determine whether the differential expression of analytes between two sample or treatment groups is valid with respect to the assay-associated noise in individual experiments.  $rxCOV$  is applicable when measurements are made using antibody-based immunoassays, aptamer-based biosensing assays, and for assays where the measured quantity is a ratio scale variable with zero-value analyte expression being biochemically meaningful. It is easy to use and requires computation of only mean and standard deviation. Its key strength is the objective threshold of zero that establishes experiment-specific analyte fidelity that also provides a sufficient condition for the validity of the p-value based claim of statistical significance. Since  $rxCOV$  is computed per analyte at a very low computational cost, it can be easily computed for multiple analytes measured in parallel as is the case for multiplexed assays. Additionally, it can also be extended to multiple sample group comparisons, each implemented pairwise. Finally,  $rxCOV$  can help determine the optimal number of experiment-specific replicates required to ensure analyte fidelity. We anticipate that these strengths will make  $rxCOV$  an attractive metric in aiding implementation of robust and reproducible biomedical research studies.

## Materials and methods

**Tissue collection.** The data for the case study demonstrating the practical efficacy of  $rxCOV$ , was collected as part of a separate study of patients with Lynch Syndrome (LS), the most common cause of hereditary colorectal cancer. The study (STUDY20010017) was approved by the Institutional Review Board at the University of Pittsburgh, Pittsburgh, PA and adhered to the University guidelines for research involving human subjects. All participants signed informed consent documents. Biopsies of normal-appearing colorectal mucosa were obtained from 16 healthy volunteers and 28 LS patients. All biopsies were obtained during routine colonoscopies performed at University of Pittsburgh Medical Center, Shadyside Hospital, Pittsburgh, PA, USA.

**Materials: chemicals and reagents.** All transport and tissue culture media ingredients were purchased from Thermo Fisher Scientific, (Waltham, Massachusetts), including RPMI 1640 (Cat #22400105), heat inactivated fetal bovine serum (HI-FBS, Cat #10082147) and Antibiotic-Antimycotics (Cat #15240062). Biomarker assays were performed using the Meso Scale Discovery (MSD, Rockville, MD) V-PLEX Human Cytokine 30-Plex Kit (Cat #K15054D).

**Multiplexed ELISA on explant cultures.** Intact biopsies collected from healthy and LS participants were immediately placed into tubes containing tissue culture media comprised of RPMI 1640, 7.5% HI-FBS and 1% antibiotic-antimycotic. Samples were maintained on ice and transported to the laboratory in a biohazard container. Tissues were immediately weighed and placed into prefilled individual wells of a 24-well tissue culture plate containing 1 mL of complete RPMI (cRPMI;RPMI 1640, 10% HI-FBS, 1% antibiotic/antimycotic) culture medium in a biosafety cabinet and incubated for 24 h at 37 °C with 5% CO<sub>2</sub>, as described previously<sup>23,24</sup>. Soluble biomarkers released into the supernatant through 24 h of culture were aliquoted and measured using MSD mELISA. Measured biomarkers included a range of chemotactic, immunosuppressive and proinflammatory cytokines used to profile the immune signaling microenvironment of the tissue samples. All assays were performed according to the manufacturer's instructions. To remove variations due to size of tissue biopsy, we normalized the assayed expression by tissue weight. The mELISA assay was also repeated on a second aliquot from each sample. Differences between these two sets of measurements were used as a quantification of assay-associated experimental variability.

**Software.** R programming language implemented within the RStudio integrated development environment was used to perform all analysis. We note that no specialized packages are required to implement the  $rxCOV$  metric itself.

## Data availability

The simulated and experimental data supporting the development and findings of this work are available from the corresponding author upon reasonable request.

Received: 29 September 2022; Accepted: 29 December 2022

Published online: 03 January 2023

## References

- Vashist, S. K. & Luong, J. H. Immunoassays: An overview. In *Handbook of Immunoassay Technologies* (eds Vashist, S. K. & Luong, J. H.) 1–18 (Academic Press, 2018).
- Bastarache, J., Koyama, T., Wickersham, N. & Ware, L. Validation of a multiplex electrochemiluminescent immunoassay platform in human and mouse samples. *J. Immunol. Methods* **408**, 13–23. <https://doi.org/10.1016/j.jim.2014.04.006> (2014).



3. Cox, K. L. Immunoassay methods. In *Assay Guidance Manual* (eds Markossian, S., Grossman, A., Brimacombe, K. *et al.*) (Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2019).
4. Sakamoto, S. *et al.* Enzyme-linked immunosorbent assay for the quantitative/qualitative analysis of plant secondary metabolites. *J. Nat. Med.* **72**, 32–42. <https://doi.org/10.1007/s11418-017-1144-z> (2018).
5. Lim, M. J. *et al.* An ELISA-based high throughput protein truncation test for inherited breast cancer. *Breast Cancer Res.* **12**, R78. <https://doi.org/10.1186/bcr2722> (2010).
6. Gopinath, S. C., Lakshmi Priya, T. & Awazu, K. Colorimetric detection of controlled assembly and disassembly of aptamers on unmodified gold nanoparticles. *Biosens. Bioelectron.* **15**, 115–123. <https://doi.org/10.1016/j.bios.2013.07.037> (2014).
7. Stoltenburg, R., Reinemann, C. & Strehlitz, B. SELEX-a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol. Eng.* **24**, 381–403. <https://doi.org/10.1016/j.bioeng.2007.06.001> (2007).
8. Toh, S. Y., Gopinath, S. C. & Tang, T. H. Aptamers as a replacement for antibodies in enzyme-linked immunosorbent assay. *Biosens. Bioelectron.* **64**, 392–403. <https://doi.org/10.1016/j.bios.2014.09.026> (2015).
9. Caruso, B., Bovo, C. & Guidi, G. C. Causes of preanalytical interferences on laboratory immunoassays—A critical review. *EJIFCC* **31**, 70–84 (2020).
10. Dhar, P., Samarasinghe, R. M. & Shigdar, S. Antibodies, nanobodies, or aptamers—which is best for deciphering the proteomes of non-model species. *Int. J. Mol. Sci.* **21**, 2485. <https://doi.org/10.3390/ijms21072485> (2020).
11. Lakhin, A. V., Tarantul, V. Z. & Gening, L. V. Aptamers: Problems, solutions and prospects. *Acta Nat.* **5**, 34–43 (2013).
12. Sturgeon, C. M. & Viljoen, A. Analytical error and interference in immunoassay: Minimizing risk. *Ann. Clin. Biochem.* **48**, 418–432. <https://doi.org/10.1258/acb.2011.011073> (2011).
13. Feltz, C. & Miller, G. An asymptotic test for the equality of coefficients of variation from  $k$  populations. *Stat. Med.* **15**, 647–658. [https://doi.org/10.1002/\(SICI\)1097-0258\(19960330\)15:6<647::AID-SIM184>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1097-0258(19960330)15:6<647::AID-SIM184>3.0.CO;2-P) (1996).
14. Krishnamoorthy, K. & Lee, M. Improved tests for the equality of normal coefficients of variation. *Comput. Stat.* **29**, 215–232. <https://doi.org/10.1007/s00180-013-0445-2> (2014).
15. Bedeian, A. G. & Mossholder, K. W. On the use of the coefficient of variation as a measure of diversity. *Organ. Res. Methods* **3**, 285–297. <https://doi.org/10.1177/109442810033005> (2000).
16. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* (L. Erlbaum Associates, 1988).
17. Vartia, Y. O. *Relative Changes and Index Numbers, No. 4 in ETLA A* (The Research Institute of the Finnish Economy, 1976).
18. Lang, S. *Basic Mathematics* (Springer, 1988).
19. Stewart, J. *Calculus: Early Transcendentals* (Brooks/Cole, Cengage Learning, 2012).
20. Thomas, G., Weir, M., Hass, J. & Giordano, F. *Calculus: Early Transcendentals* (Pearson College Div, 2007).
21. Hong, L. The absolute difference law for expectations. *Am. Stat.* **69**, 8–10 (2015).
22. Huber, P. *Robust Statistics* 2nd edn. (Wiley, 2009).
23. Brand, R. M. *et al.* Immunological responsiveness of intestinal tissue explants and mucosal mononuclear cells to ex vivo stimulation. *J. Immunol. Methods* **463**, 39–46. <https://doi.org/10.1016/j.jim.2018.08.009> (2018).
24. Brand, R. M. *et al.* Tofacitinib inhibits inflammatory cytokines from ulcerative colitis and healthy mucosal explants and is associated with pstat1/3 reduction in t-cells. *Am. J. Physiol. Gastrointest. Liver Physiol.* **320**, G396–G410. <https://doi.org/10.1152/ajpgi.00383.2020> (2021).

## Author contributions

R.M.B., E.J.M., and S.U. conceived the concept; D.P. and S.U. developed the principle and the method; B.D., E.K. and R.E.B. collected the patient samples; R.M.B. and A.Z. performed the multiplexed immunoassays; R.M.B., E.J.M., D.P. and S.U. performed the analyses; R.M.B. and S.U. wrote the manuscript. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to S.U.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023