# Supplemental material

In the main body of this article a strong emphasis was put in applying the p-value distribution analysis to real data from actual screening campaigns. While a comparison of PVDA predictions with experimental results showed good agreement and thus proved the usefulness of the method, the accessible parameter range, for instance, of the day-to-day variation or of the different magnitude of the standard deviation of negative and positive controls, is limited. The simulations presented here were performed to better explore the validity range of PVDA and to find its limits.

## Supplemental methods

### Monte-Carlo simulation

All simulations and computations were performed in **R** using a conventional desktop computer. Data were generated in a format that resembles real HTS data: one line corresponds to the measurement in one well, the columns indicate the following parameters: date, plate identifier, well identifier, compound type (negative control, intermediate control, positive control, compound), plate-wise normalized signal (expressed as percentage of the difference between the medians of the positive and negative control on each plate), plate-wise variances of the 3 controls.

To simulate a small screen, data were grouped to 384 wells per plate, of which 24 were chosen to be controls and 360 to be compounds. The runs consisted of 100 plates per day, 1000 plates in total, i.e., 360'000 data points. The signal of the controls was drawn from normal distributions with mean equal to 0, 50, and 100 for negative, intermediate, and positive control, respectively. The mean standard deviation was chosen as indicated in Supplemental Table 1. To account for plate-to-plate and day-to-day variations, the standard deviation was adjusted on each plate by two correction factors, each drawn from a uniform distribution and constant for all wells:
$c = \sigma_{set} \cdot sdcor_{d2d}^{x} \cdot sdcor_{p2p}^{y}$ $x, y \sim U(-1,1)$. The day-to-day variability factor is indicated in Suppl. Tab. 1, the plate-to-plate variability factor was set to 1.2, except for the data in Suppl. Fig. 6, where it was increased to 2.

The expected compound signal was chosen to be either 0 for the inactive fraction or drawn from a normal distribution with mean 0 and standard deviation 50 at a ratio according to the set hit rate (Suppl. Tab. 1). In an idealistic way, this ad hoc selection resembled real screen distributions in the absence of gross outliers. Assay variability was added to the compound signal by drawing from a normal distribution with mean equal to the just mentioned expected compound signal and a standard deviation $\sigma(signal)$ that included the same correction factors above and in addition was allowed to depend on the magnitude of the signal. For simplicity, the compound standard deviation at a signal of 0, 50, and 100 was assumed to be equal to the standard deviation of the negative, intermediate, and positive control, and a linear dependence was chosen in between. The resulting data set was then processed by the same PVDA algorithm as the real data. From the positive arm of $T = \frac{signal - cutoff}{\sigma_{neg}}$, the p-value was calculated according to a Z-test, i.e., directly from the normal distribution of $T$.

## Supplemental results and discussion

### Monte-Carlo simulation

The range of validity of PVDA was explored for the following conditions: small (0.01), intermediate (0.1) and large (up to 0.4) hit rate; small (1.2), intermediate (2.0) and large (3.0) day-to-day change of the assay variability; low (2), medium (4) and high (8) standard deviation of the negative control; a signal-independent and a signal-dependent (factor of 4 per 100 percent) standard deviation of the compounds.

Predictions of PVDA are compared with controlled conditions, i.e., the set hit rate with the predicted hit rate (1-FDR) given $H_0 \mid signal = 0$; and with calculated prior information, i.e., the expected number of true positives among all positives, $\frac{\#\{signal>cutoff\}\{signal}>cutoff\}}{\#\{signal>cutoff\}}$, with the predicted confirmation rate (1-FDR) given $H_0 \mid signal = cutoff$.

Results of the simulations using the parameters in Suppl. Tab. 1 are shown in Suppl. Figs. 1-6. Except for Suppl. Fig. 6, which will be discussed separately, all PVDs are regular. Although there is no formal proof, it is tempting to claim that this finding retrospectively justifies the use of PVDA and shows that the underlying assumptions are not dramatically violated. The height of the peak at low p-values reflects the observed hit rate, the widths of the peak is determined by the assay variability.

The plate-wise Z' pattern that was calculated for each of the different parameter sets ranges from constantly high (Suppl. Fig. 1a) to highly variable and often very low (Suppl. Fig. 4d). In fact, if it would be real screening data, those in Suppl. Figs. 2-4 c & d, as well as Suppl. Fig. 6 would not pass our QC criteria. The signal dependent, daily changing assay variability that is illustrated in the corresponding graphs on the right hand side in each figure nicely resembles the profile and range of real data.

The robust estimate of the compound standard deviation (black line) in Suppl. Fig. 4 is increasing from a-d with increasing hit rate, starting to follow closely the negative control (a) and ending to be rather close to the intermediate control (d). The increased proportion of actives among the compounds (0.1-0.4) is causing the distribution to broaden even close to the maximum at 0 signal and is gradually increasing the estimated standard deviation even though the robust MAD is used. This nicely shows that particular care needs to be taken when the compounds' distribution is used to estimate parameters of the inactives, under the assumption that most of the compounds are inactive.

The data in Suppl. Fig. 6 exhibit an irregular PVD with a valley around 0.1-0.3 and a hill for large p-values. This is a definite sign that here the underlying assumptions are violated to a large extent and any prediction based on PVDA is expected to have a large error. The difference between this data and those in Suppl. Fig. 4a is just the fact that the additional plate-to-plate variability is */÷ 2 instead of */÷ 1.2. This increased variability difference between neighboring plates is sufficient to violate the assumption that neighboring plates have a similar assay variability and therefore can be averaged over when estimating the local standard deviation. Lowering the amount of averaging is reducing the problem, but at the cost of a larger estimation error. Without averaging, the PVD is completely regular but the hit rate is overestimated by a factor of 2 (data not shown). Large day-to-day variations of the assay variability do not seem to violate the assumptions of PVDA to the same extent because they only affect the estimate of the standard deviation on plates close to the border between two days, something that can even be avoided entirely. The data were excluded from further analysis.

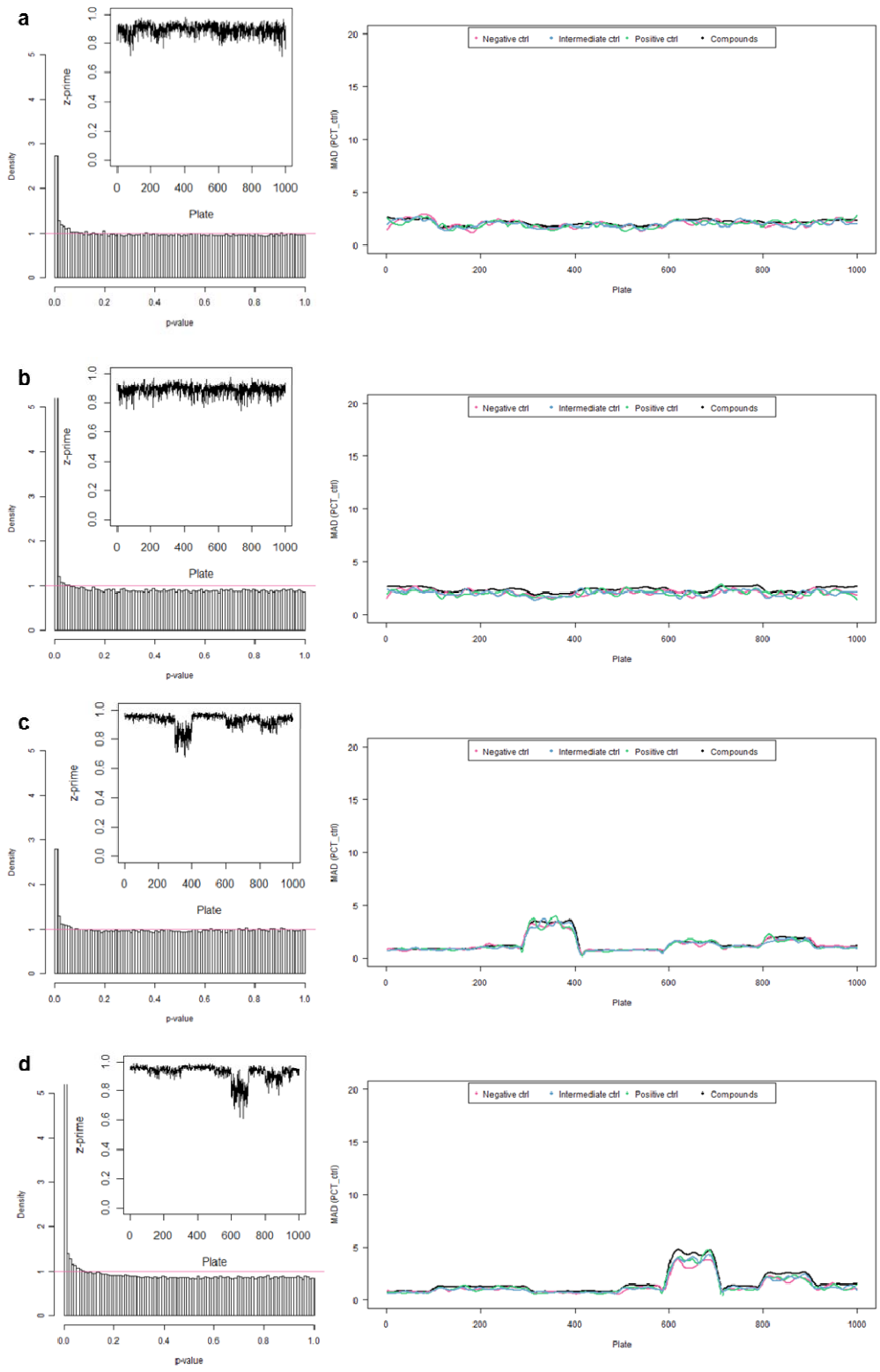
A summary of PVDA predictions for the simulated data is shown in Suppl. Fig. 7. The diagonal lines indicate the line of identity together with a */÷ 1.2 and +/- 0.1 confidence band, respectively. Given the described conditions, larger hit rates are predicted well within the given confidence limit (Suppl. Fig. 7a). With decreasing hit rate the data are reaching a sensitivity limit at about 5%. The confirmation rate relative to a given cutoff of 25 (circles) is predicted much better than the one with a cutoff of 40 (triangles). This is related to the fact that the cutoff was increased only because the assay variability did not allow a reliable hit selection at the lower value (as judged from an irregular PVD). The color code reflects this dependence of the quality of the prediction on the total assay variability, quantified by Z'. For the cases with Z' > 0.75, which we would consider a good quality run, the predicted confirmation rate was only few percent points away from the true value and altogether very high. With decreasing Z' the predicted confirmation rate decreased as well as its reliability (Suppl. Fig. 7b).

The performance of PVDA in the current realistic simulation in the explored parameter space underlines its broad applicability in relevant real screening situations. In particular, for screens with Z' > 0.7, PVDA predicts hit rates and confirmation rates sufficiently well to serve as a planning tool for follow-up experiments. Using PVDA is only valid if the assumptions are fulfilled. A major benefit of PVDA is the fact that the shape of the PVD allows to identify situations when this is not the case.

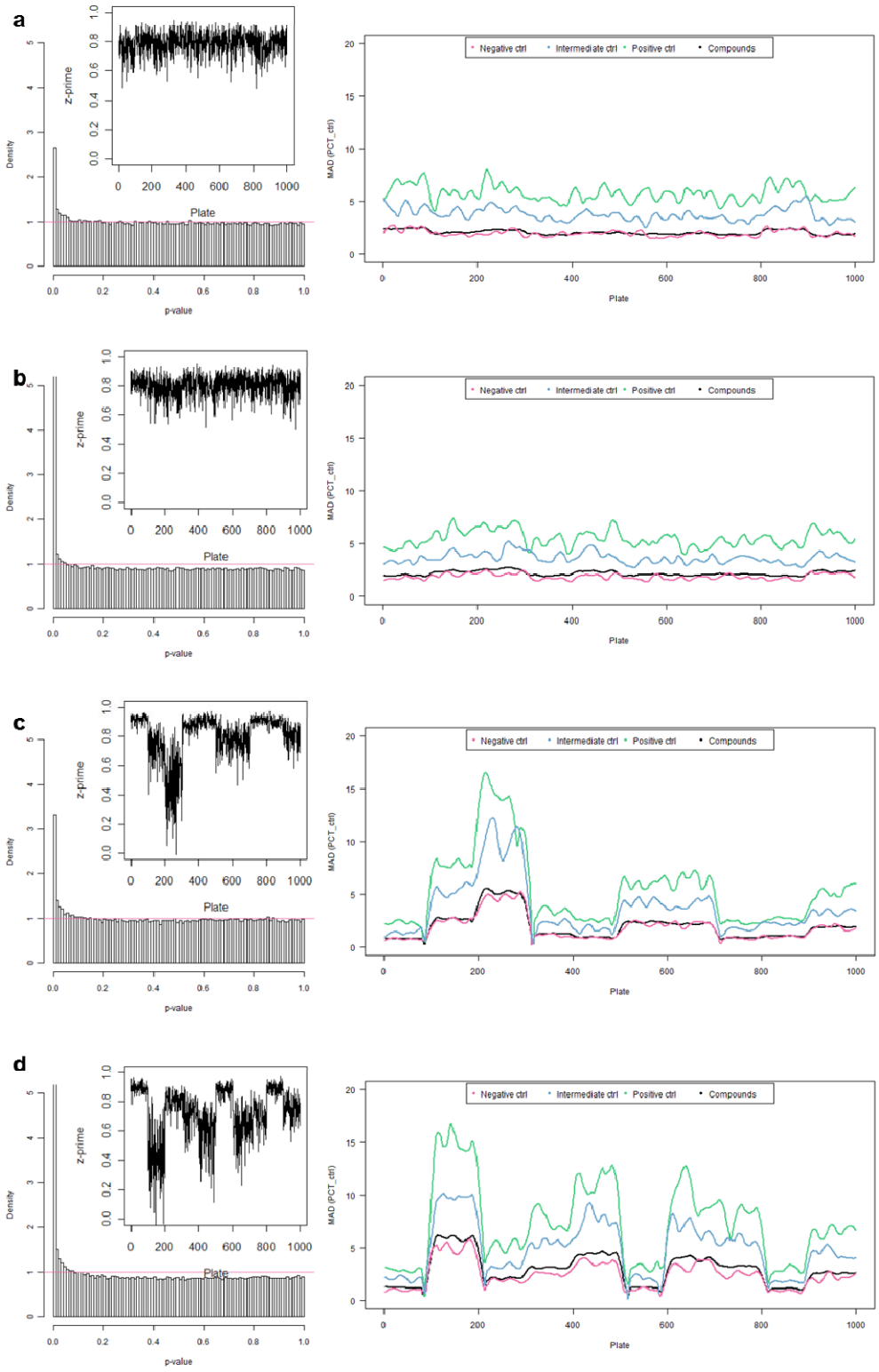
**Supplemental Table 1:** Parameters used in the simulations, the corresponding figure number, and the obtained results (grey).

| Fig | set hit rate | day-to-day sd variability ($sdcor_{d2d}$) | sd neg ctr | sd interm ctr | sd pos ctr | median Z' | predicted hit rate | Confirm cutoff | expected confirm rate | predicted confirm rate |
|---|---|---|---|---|---|---|---|---|---|---|
| 1a | 0.01 | 1.2 | 2 | 2 | 2 | 0.89 | 0.03 | 25 | 0.98 | 0.97 |
| 1b | 0.1 | 1.2 | 2 | 2 | 2 | 0.89 | 0.12 | 25 | 0.98 | 0.95 |
| 1c | 0.01 | 3 | 2 | 2 | 2 | 0.95 | 0.01 | 25 | 0.99 | 0.98 |
| 1d | 0.1 | 3 | 2 | 2 | 2 | 0.94 | 0.15 | 25 | 0.99 | 0.96 |
| 2a | 0.01 | 1.2 | 2 | 4 | 6 | 0.80 | 0.04 | 25 | 0.97 | 0.98 |
| 2b | 0.1 | 1.2 | 2 | 4 | 6 | 0.81 | 0.12 | 25 | 0.97 | 0.95 |
| 2c | 0.01 | 3 | 2 | 4 | 6 | 0.86 | 0.04 | 25 | 0.99 | 1.00 |
| 2d | 0.1 | 3 | 2 | 4 | 6 | 0.76 | 0.12 | 25 | 0.97 | 0.92 |
| 3a | 0.01 | 1.2 | 4 | 6 | 8 | 0.71 | 0.03 | 25 | 0.96 | 0.92 |
| 3b | 0.1 | 1.2 | 4 | 6 | 8 | 0.69 | 0.12 | 25 | 0.95 | 0.90 |
| 3c | 0.01 | 3 | 4 | 6 | 8 | 0.67 | 0.03 | 25 | 0.86 | 0.64 |
| 3d | 0.1 | 3 | 4 | 6 | 8 | 0.76 | 0.12 | 25 | 0.92 | 0.79 |
| 4a | 0.01 | 1.2 | 8 | 8 | 8 | 0.60 | 0.07 | 40 | 0.92 | 0.64 |
| 4b | 0.1 | 1.2 | 8 | 8 | 8 | 0.60 | 0.05 | 40 | 0.90 | 0.73 |
| 4c | 0.01 | 3 | 8 | 8 | 8 | 0.45 | 0.04 | 40 | 0.33 | 0.17 |
| 4d | 0.1 | 3 | 8 | 8 | 8 | 0.58 | 0.14 | 40 | 0.80 | 0.52 |
| 5a | 0.1 | 2 | 2 | 4 | 6 | 0.83 | 0.14 | 25 | 0.98 | 0.95 |
| 5b | 0.2 | 2 | 2 | 4 | 6 | 0.81 | 0.23 | 25 | 0.97 | 0.95 |
| 5c | 0.3 | 2 | 2 | 4 | 6 | 0.81 | 0.32 | 25 | 0.97 | 0.95 |
| 5d | 0.4 | 2 | 2 | 4 | 6 | 0.78 | 0.40 | 25 | 0.97 | 0.94 |
| 6 | 0.1 | 2 | 2 | 4 | 6 | 0.73 | NA | 25 | NA | NA |

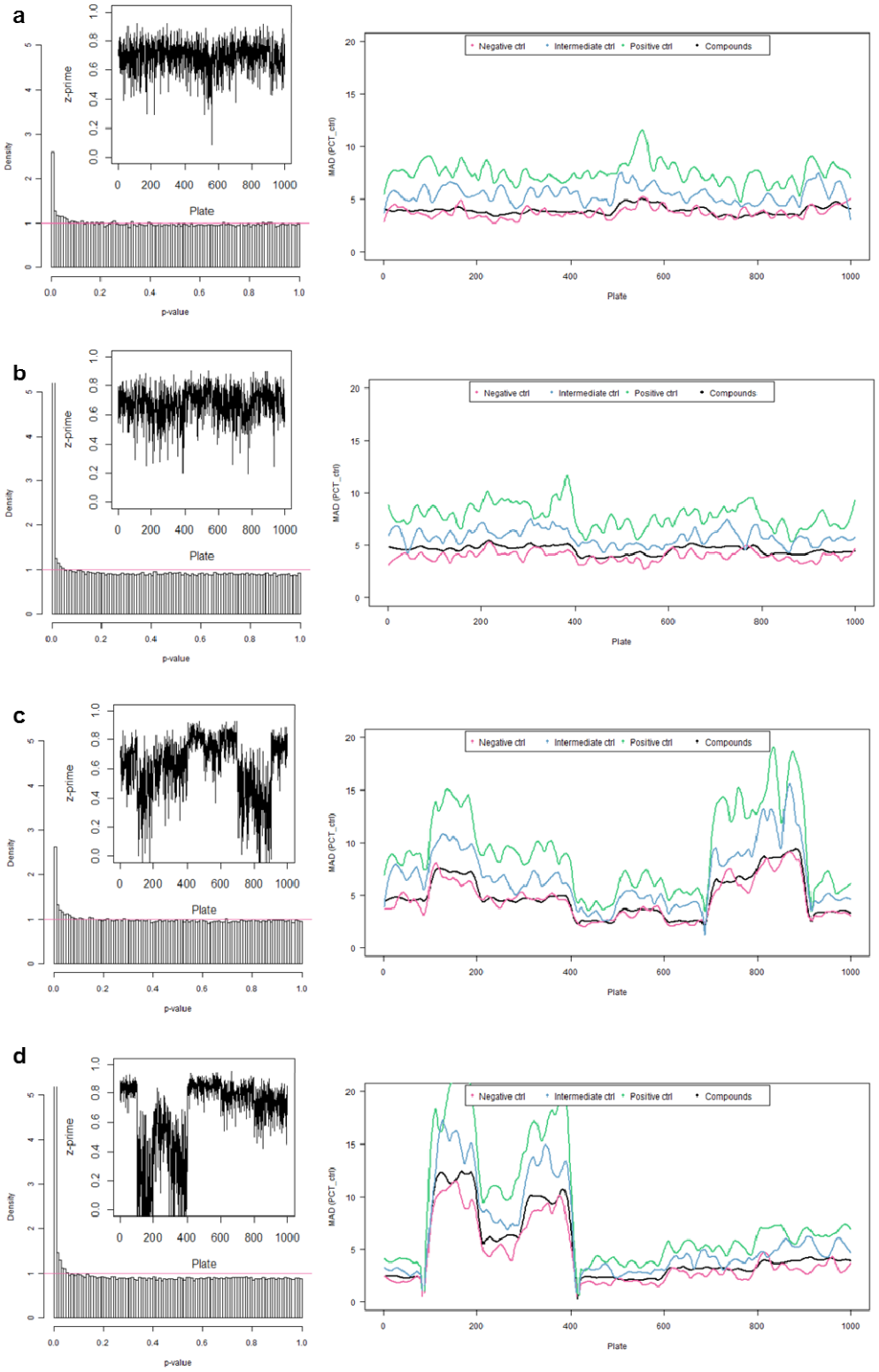
**Supplemental Figure 1:** Simulation results for small, equal variability (sd.n = sd.i = sd.p = 2). (a,b) low, (c,d) high day-to-day variability. (a,c) low, (b,d) high hit rate. P-value distribution (left), plate-wise Z' (inset), smoothed plate-wise robust estimate of the standard deviation (right).



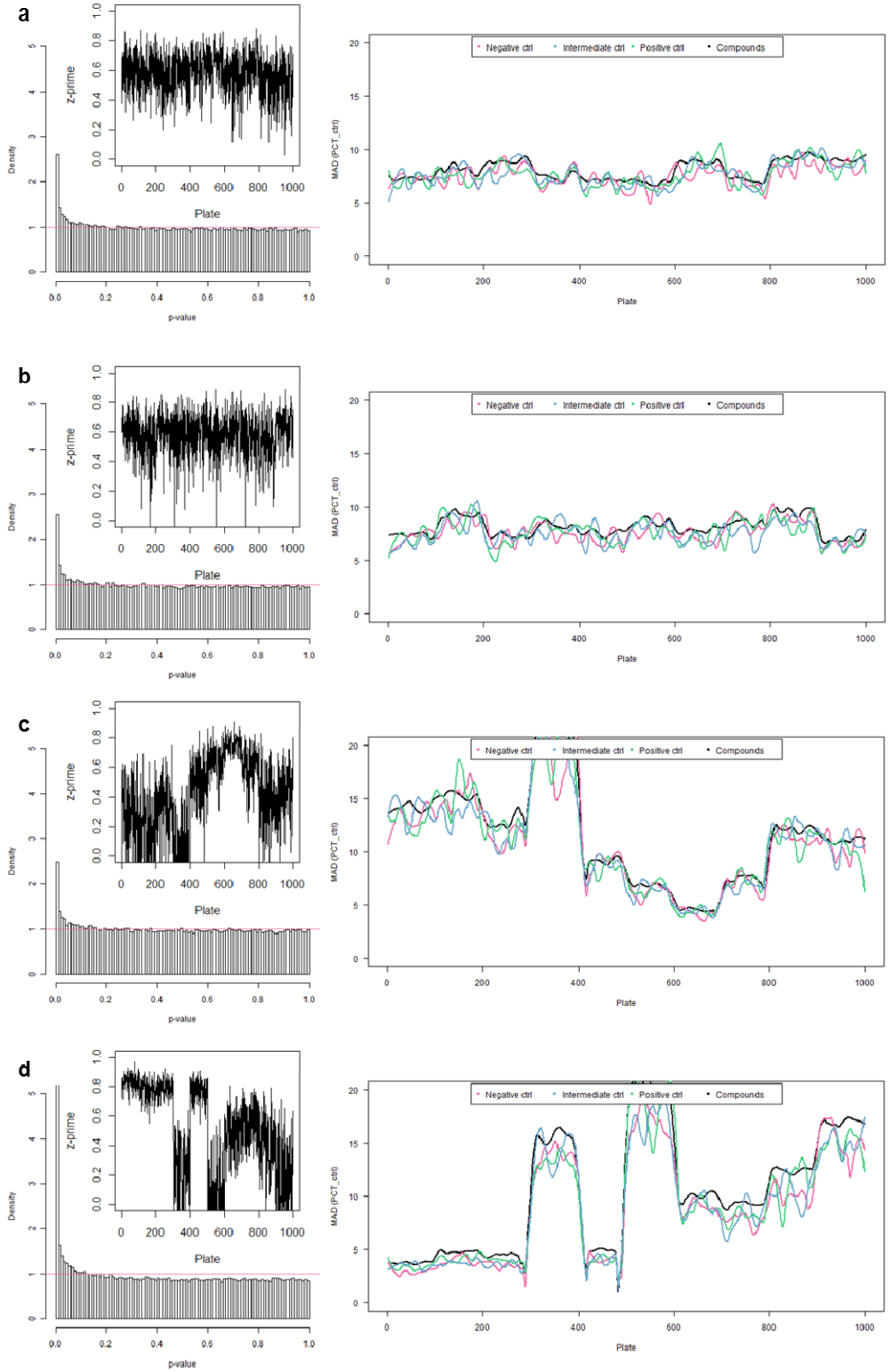
**Supplemental Figure 2:** Simulation results for small, changing variability (sd.n = 2, sd.i = 4, sd.p = 6). (a,b) low, (c,d) high day-to-day variability. (a,c) low, (b,d) high hit rate. P-value distribution (left), plate-wise Z' (inset), smoothed plate-wise robust estimate of the standard deviation (right).



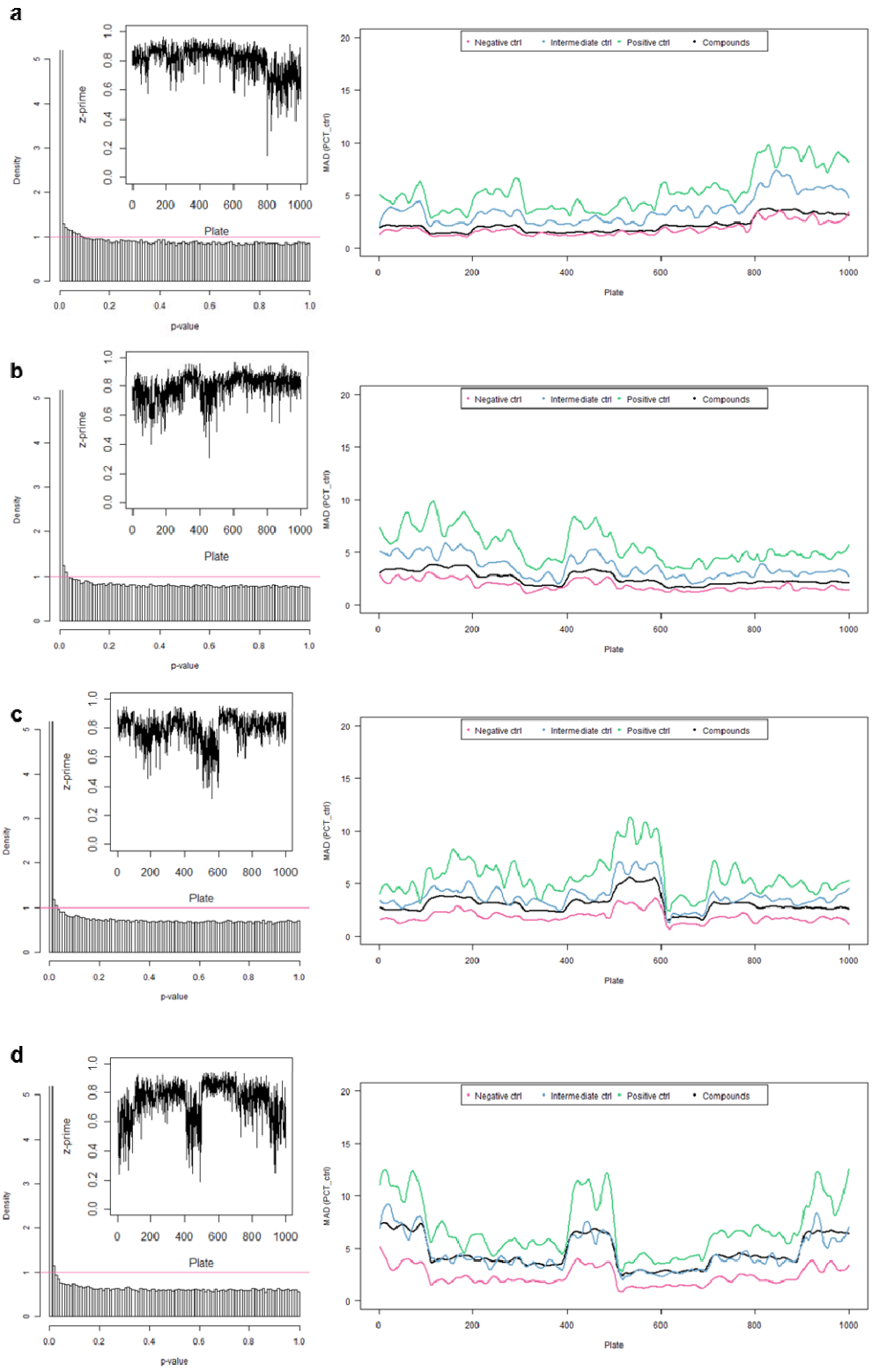
**Supplemental Figure 3:** Simulation results for large, changing variability (sd.n = 4, sd.i = 6, sd.p = 8). (a,b) low, (c,d) high day-to-day variability. (a,c) low, (b,d) high hit rate. P-value distribution (left), plate-wise Z' (inset), smoothed plate-wise robust estimate of the standard deviation (right).



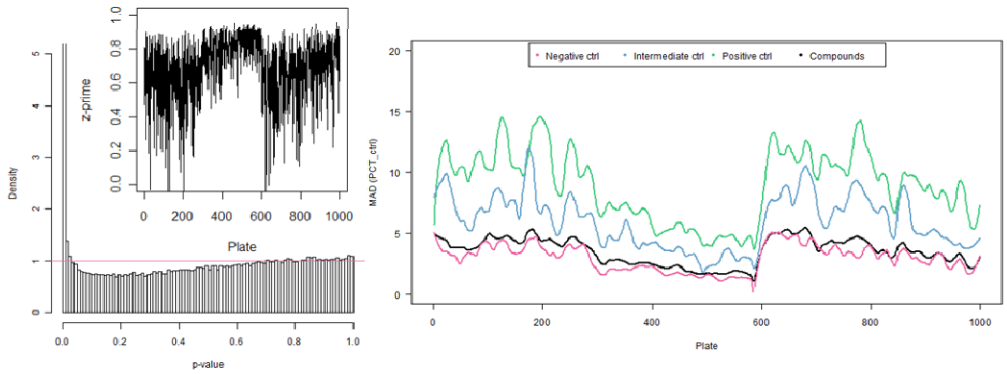
**Supplemental Figure 4:** Simulation results for large, equal variability (sd.n = sd.i = sd.p = 8). (a,b) low, (c,d) high day-to-day variability. (a,c) low, (b,d) high hit rate. P-value distribution (left), plate-wise Z' (inset), smoothed plate-wise robust estimate of the standard deviation (right).



**Supplemental Figure 5:** Simulation results for small, changing variability (sd.n = 2, sd.i = 4, sd.p = 6), medium day-to-day variability, and increasing hitrate: 0.1 (a), 0.2 (b), 0.3 (c), 0.4 (d). PVD (left), plate-wise Z' (inset), smoothed plate-wise robust estimate of the standard deviation (right).



**Supplemental Figure 6**: Simulation results for similar conditions as in Suppl. Fig. 4a but with an additional plate-to-plate variability of */÷ 2 instead of 1.2. P-value distribution (left), plate-wise Z' (inset), smoothed plate-wise robust estimate of the standard deviation (right).




**Supplemental Figure 7**: Comparison of set parameters of the simulation with predicted parameters from the PVDA. (a) Predicted hit rate versus set hit rate, (b) predicted confirmation rate versus observed confirmation rate. Triangles mark the simulations with a constant standard deviation of 8 (Fig. 3), where the confirmation cutoff had to be set to 40 instead of 25 (circles). The color code allows to distinguish simulations resulting in a median Z' below 0.5 (black), above 0.75 (green) or in between (red).