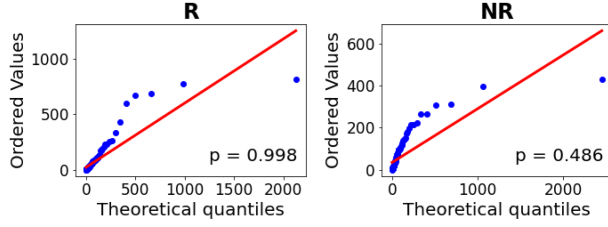


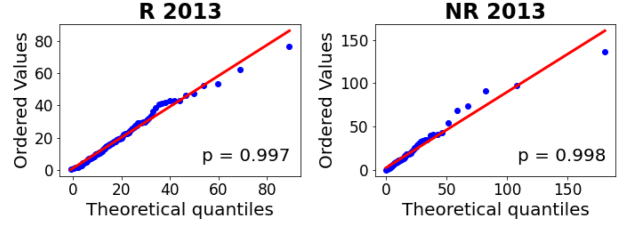
# Supporting Information: Bayesian hypothesis testing reveals that reproducible models in systems biology get more citations

Sebastian Höpfl, Jürgen Pleiss, Nicole E. Radde

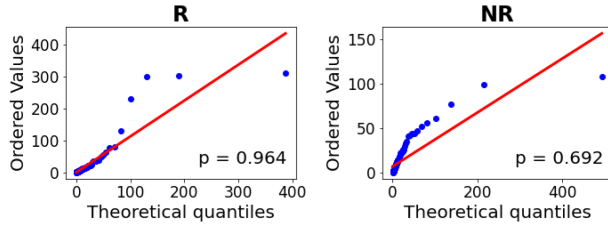
### A. 1985 - 2020



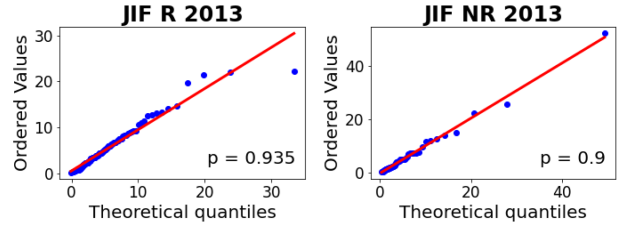
### B. 2013 - 2020



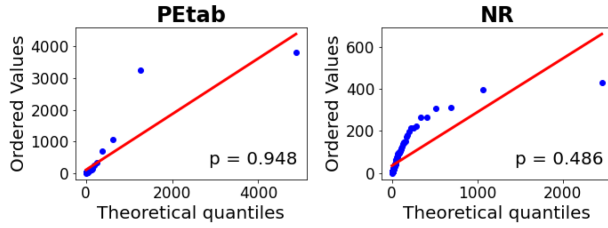
### C. JIF normalized 1985 - 2020



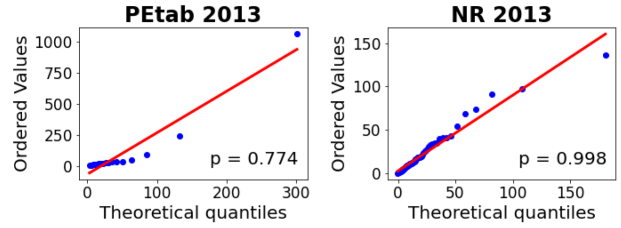
### D. JIF normalized 2013 - 2020



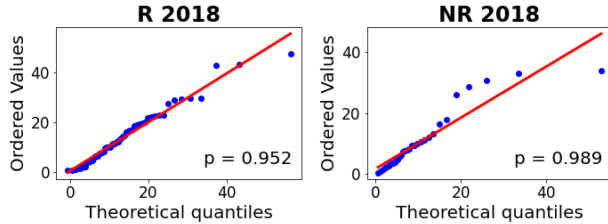
### E. PETab 1985 - 2020



### F. PETab 2013 - 2020

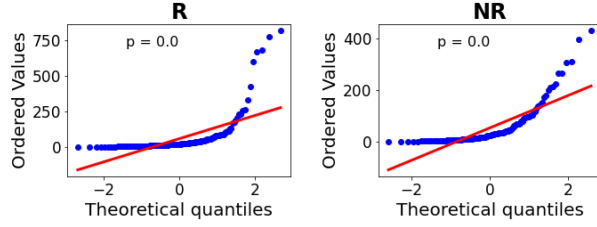


### G. 2018 - 2020

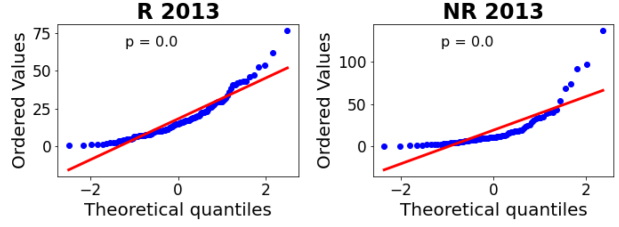


**Supplementary Figure 1: All investigated datasets can be described by a non-central t distribution (NCT).** QQ-plots of the quantiles of the data against the quantiles of the NCT distribution are shown for each comparison with the p-values of the Kolmogorov-Smirnov test for the goodness of the fit to the NCT distribution. A. Citations of papers with reproducible vs. non-reproducible models without year restriction. B. Citations of papers with reproducible vs. non-reproducible models published between 2013 and 2020. C. JIF normalized citations of papers with reproducible vs. non-reproducible models without year restriction. D. Journal Impact Factor (JIF) normalized citations of papers with reproducible vs. non-reproducible models published between 2013 and 2020. E. Citations of papers of the PETab benchmark database vs. papers with non-reproducible models without year restriction. F. Citations of papers of the PETab benchmark database vs. papers with non-reproducible models published between 2013 and 2020. G. Citations of papers with reproducible vs. non-reproducible models published between 2018 and 2020.

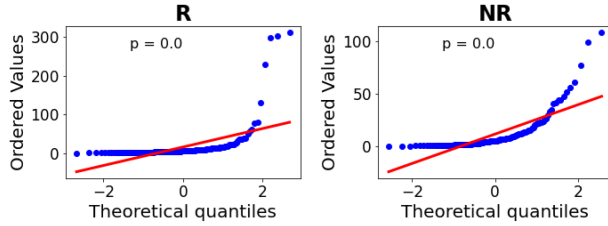
### A. 1985 - 2020



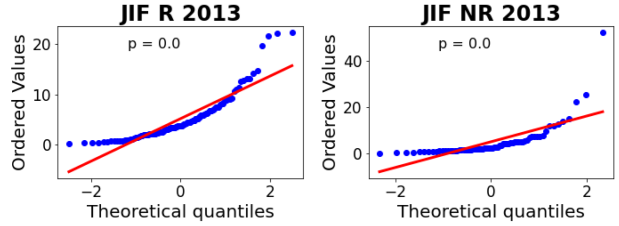
### B. 2013 - 2020



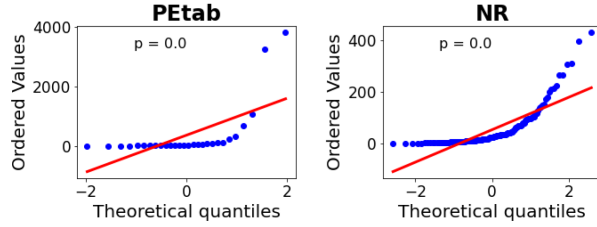
### C. JIF normalized 1985 - 2020



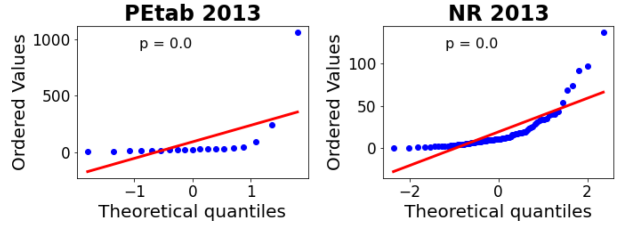
### D. JIF normalized 2013 - 2020



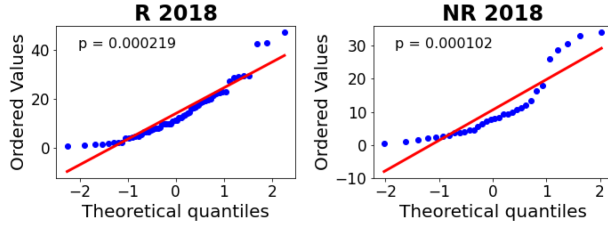
### E. PTab 1985 - 2020



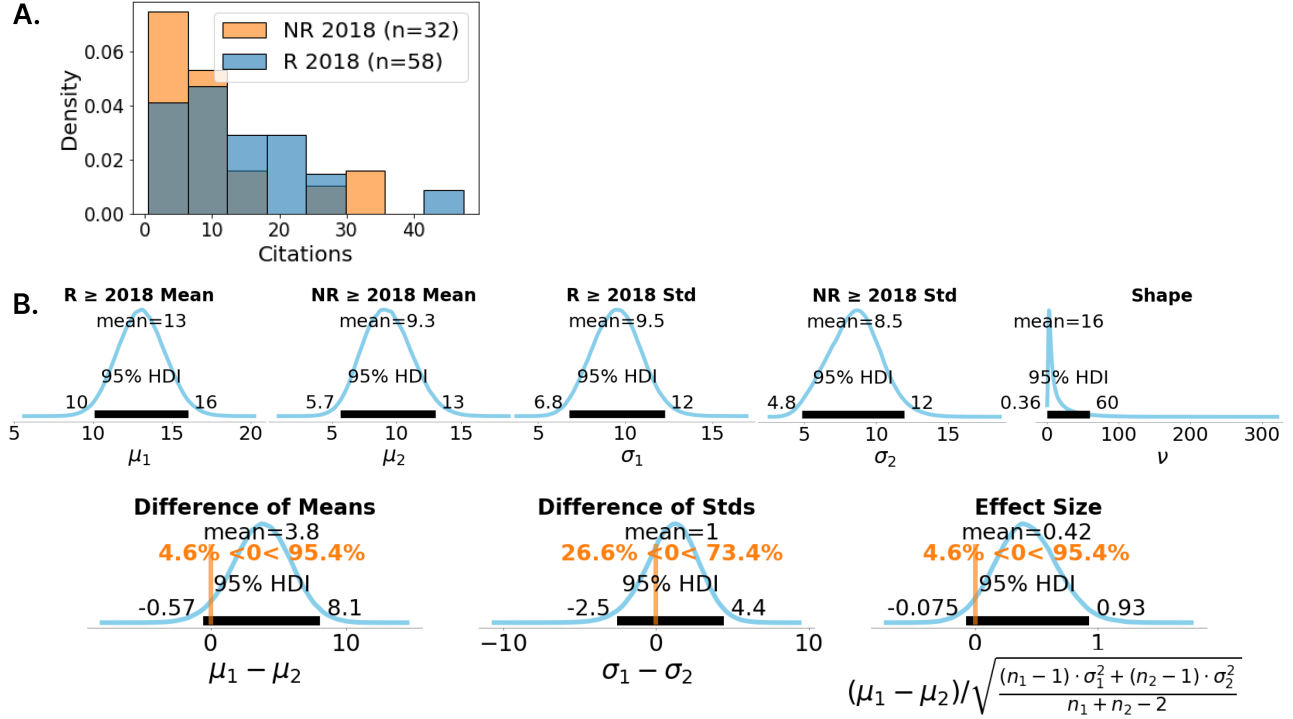
### F. PTab 2013 - 2020



### G. 2018 - 2020



**Supplementary Figure 2: The investigated datasets are not normally distributed according to a Shapiro-Wilk test and QQ-plot investigation.** QQ-plots of the quantiles of the data against the quantiles of the normal distribution are shown for each comparison with the p-values of the Kolmogorov-Smirnov test for the goodness of the fit to the normal distribution. p-values were rounded to six digits, this led to a value of 0 except for G. A. Citations of papers with reproducible vs. non-reproducible models without year restriction. B. Citations of papers with reproducible vs. non-reproducible models published between 2013 and 2020. C. JIF normalized citations of papers with reproducible vs. non-reproducible models without year restriction. D. Journal Impact Factor (JIF) normalized citations of papers with reproducible vs. non-reproducible models published between 2013 and 2020. E. Citations of papers of the PTab benchmark database vs. papers with non-reproducible models without year restriction. F. Citations of papers of the PTab benchmark database vs. papers with non-reproducible models published between 2013 and 2020. G. Citations of papers with reproducible vs. non-reproducible models published between 2018 and 2020.



**Supplementary Figure 3: Reproducible papers are significantly more cited than non-reproducible for the time span 2018 - 2020.** A. Histogram of the number of citations of 58 reproducible (R) and 32 non-reproducible (NR) papers published since 2018. B. Means  $\mu_1$  and  $\mu_2$  and standard deviations  $\sigma_1$  and  $\sigma_2$ , and shape parameter  $\nu$  posterior distributions. C. Posterior distributions of the differences in means (left), standard deviations (center) and the effect size (right). 97.3% of the credibility mass indicate a difference above zero, with a mean difference of 3.4 citations. The mean effect size between reproducible and non-reproducible papers, published since 2018, is 0.4.