

Supplementary Materials for

A practical solution to pseudoreplication bias in single-cell studies

Kip D. Zimmerman, Mark A. Espeland, Carl D. Langefeld

Correspondence to: kdzimmer@wakehealth.edu or clangefe@wakehealth.edu

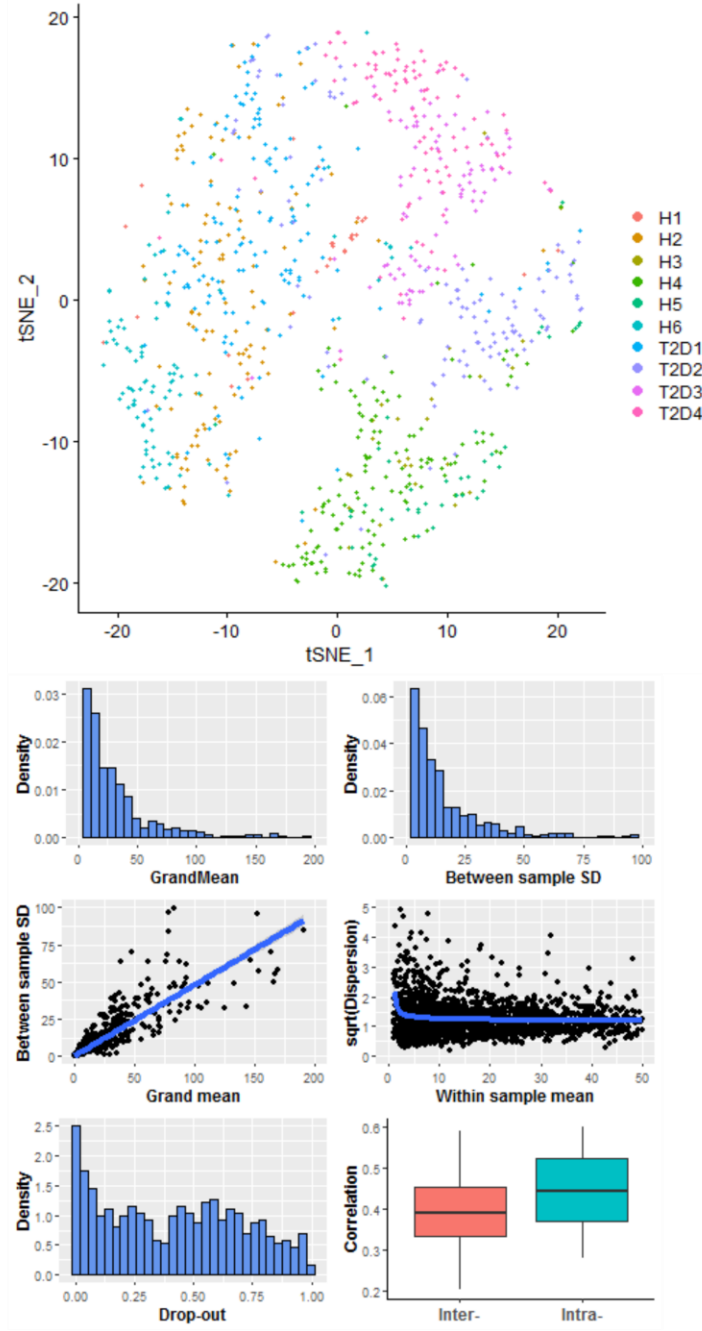
This PDF file includes:

Supplementary Figures 1 to 11

Supplementary Tables 1 to 6

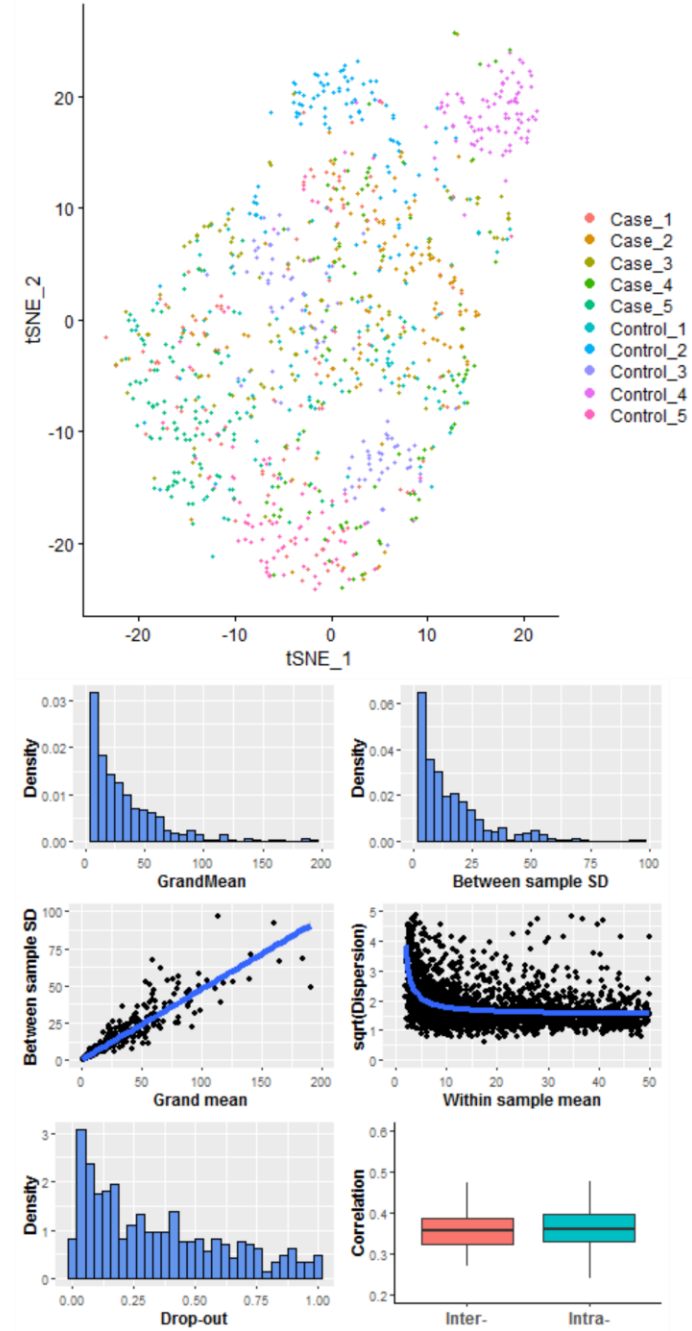
Real data - Pancreatic alpha cells

(10 samples, ~100 cells per sample)



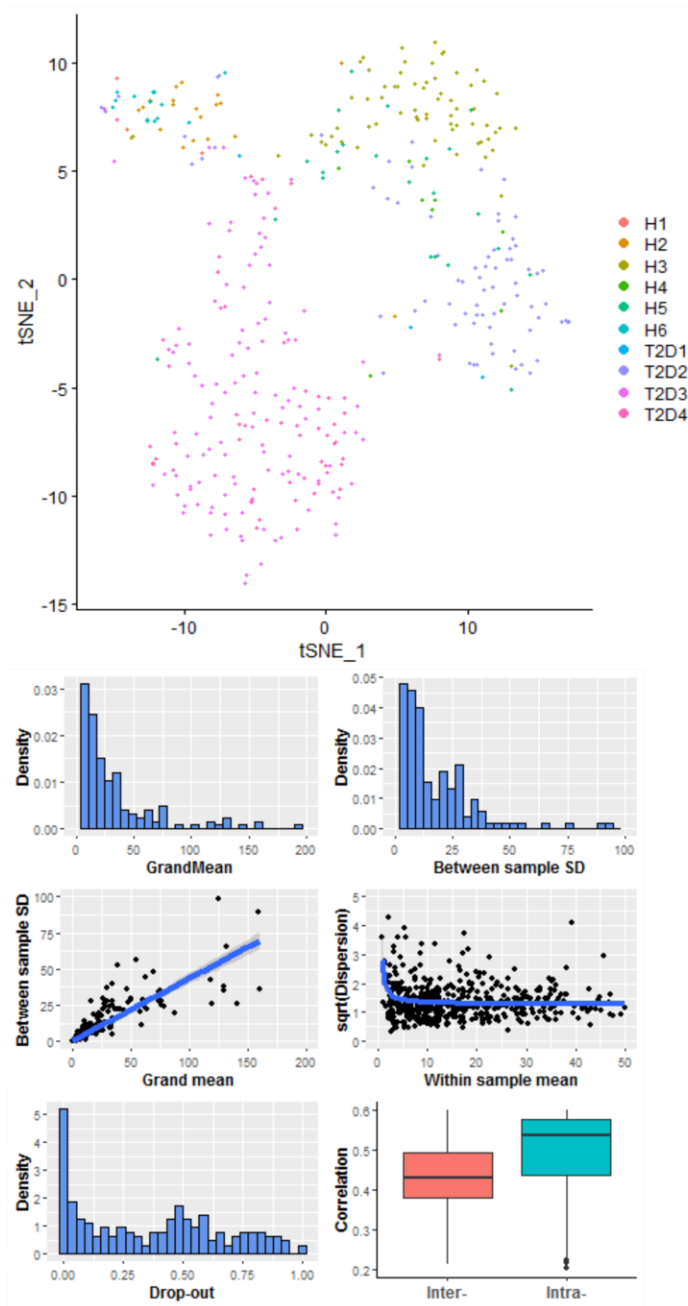
Simulated data

(10 samples, ~100 cells per sample)

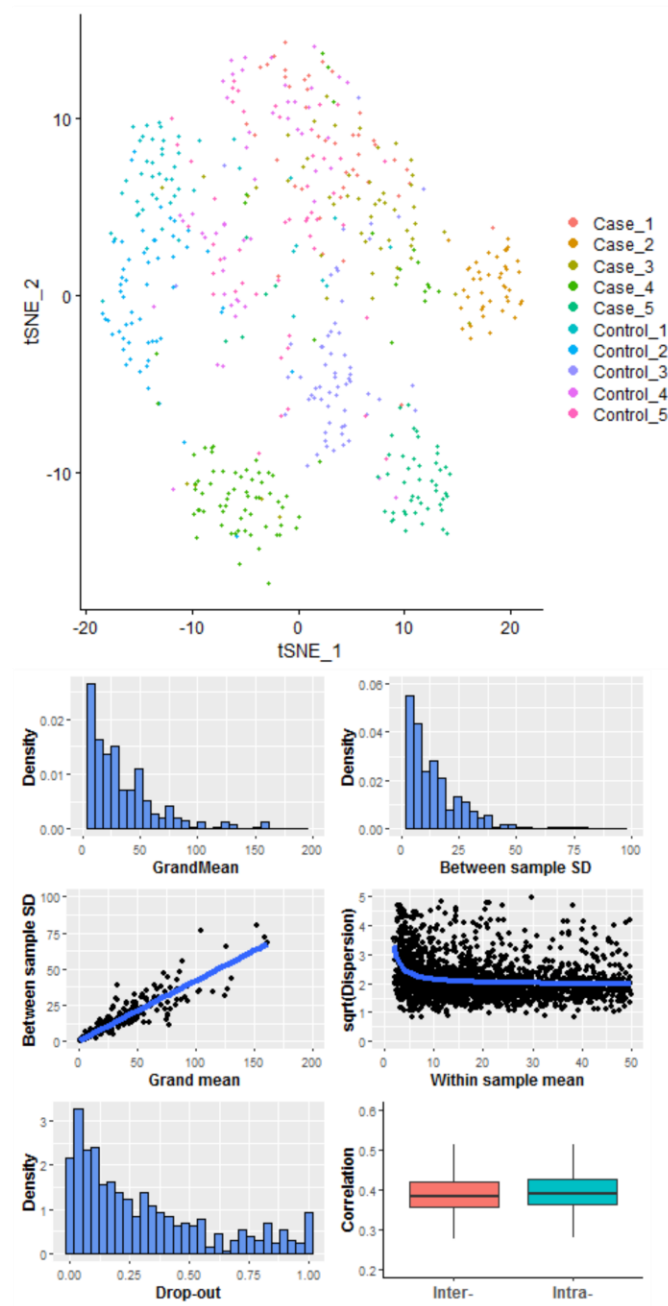


Supplementary Figure 1: Similarities between simulated data and real data. tSNE plots and plots of various parameters demonstrate similarities between simulated and real data. All axes, except tSNE plot axes, are held constant across the two sets of panels. Parameters were estimated on pancreatic alpha cells with 500 genes correlated with a Spearman's coefficient < 0.25 . The simulated data consist of 500 independently simulated genes. Top panels: tSNE plots of the data. The next panel below: histograms of the grand mean and between-sample standard deviation. Middle set of panels: relationships between the grand mean and the between-sample standard deviation and relationships between intra-individual means and dispersion. Bottom set of panels: dropout results and the boxplot demonstrate distributions of intra-individual (blue) and inter-individual (red) correlations. Differences between inter- and intra-individual correlations are less exaggerated in our simulated data. The center line represents the median. The lower and upper box limits represent the 25% and 75% quantiles, respectively. The whiskers extend to the largest observation within the box limit plus or minus one and a half multiplied by the interquartile range.

Real data - Pancreatic ductal cells (10 samples, ~50 cells per sample)

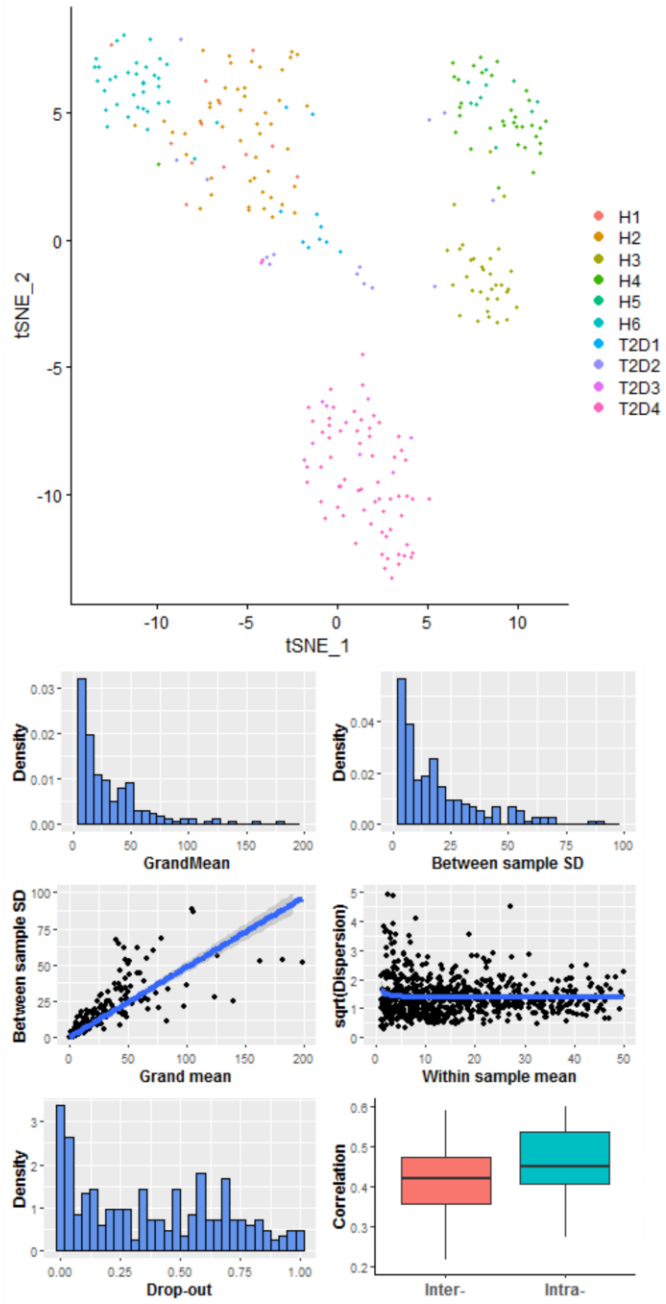


Simulated data (10 samples, ~50 cells per sample)

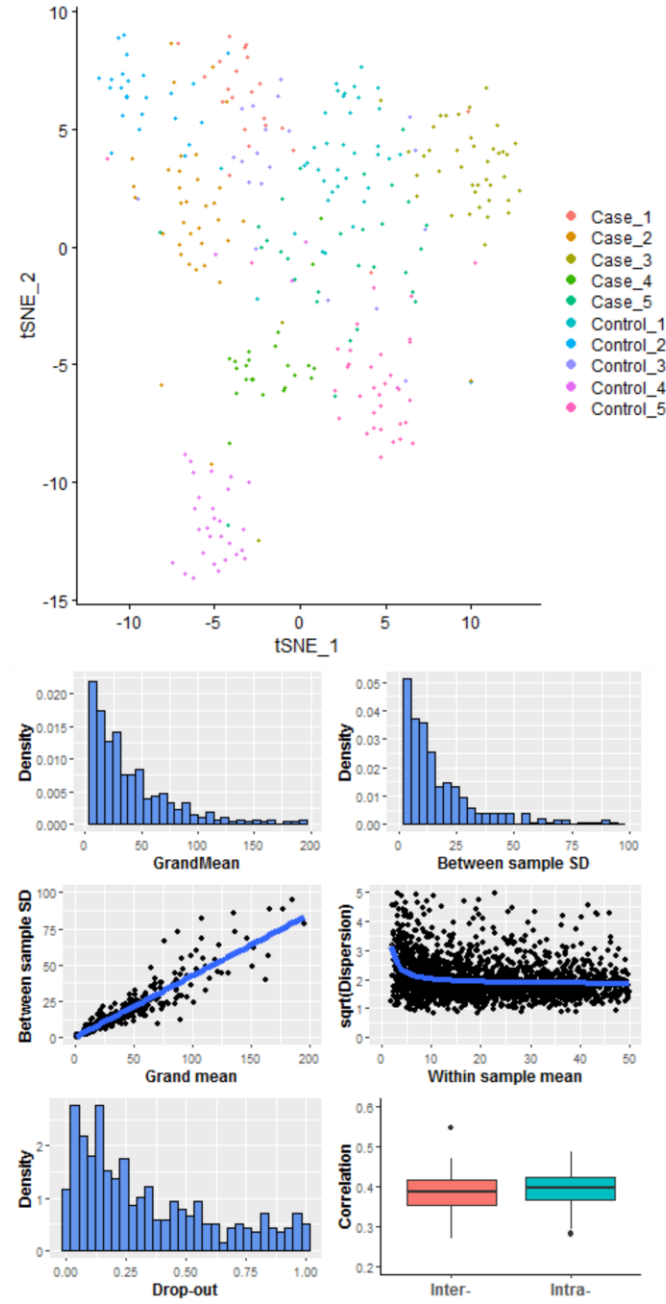


Supplementary Figure 2: Similarities between simulated data and real data. tSNE plots and plots of various parameters demonstrate similarities between simulated and real data. All axes, except tSNE plot axes, are held constant across the two sets of panels. Parameters were estimated on pancreatic ductal cells with 500 genes correlated with a Spearman's coefficient < 0.25 . The simulated data consist of 500 independently simulated genes. Top panels: tSNE plots of the data. The next panel below: histograms of the grand mean and between-sample standard deviation. Middle set of panels: relationships between the grand mean and the between-sample standard deviation and relationships between intra-individual means and dispersion. Bottom set of panels: dropout results and the boxplot demonstrate distributions of intra-individual (blue) and inter-individual (red) correlations. Differences between inter- and intra-individual correlations are less exaggerated in our simulated data. The center line represents the median. The lower and upper box limits represent the 25% and 75% quantiles, respectively. The whiskers extend to the largest observation within the box limit plus or minus one and a half multiplied by the interquartile range.

Real data - Pancreatic beta cells (10 samples, ~30 cells per sample)

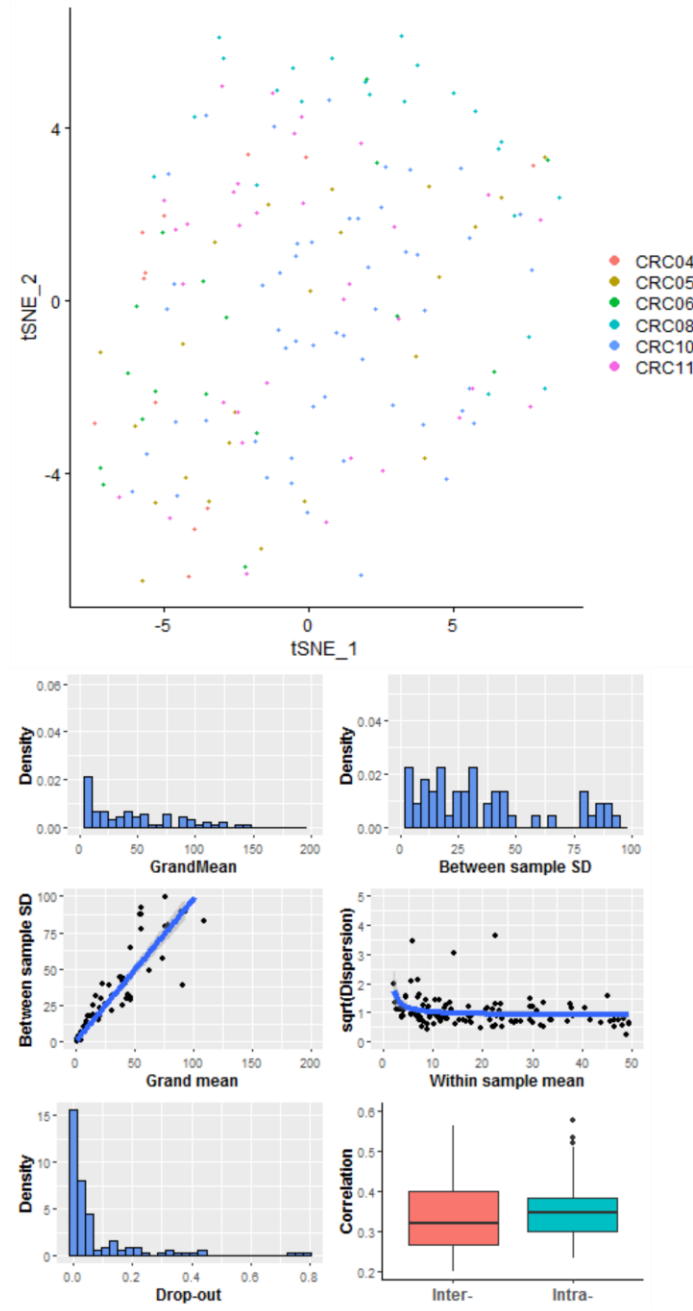


Simulated data (10 samples, ~30 cells per sample)

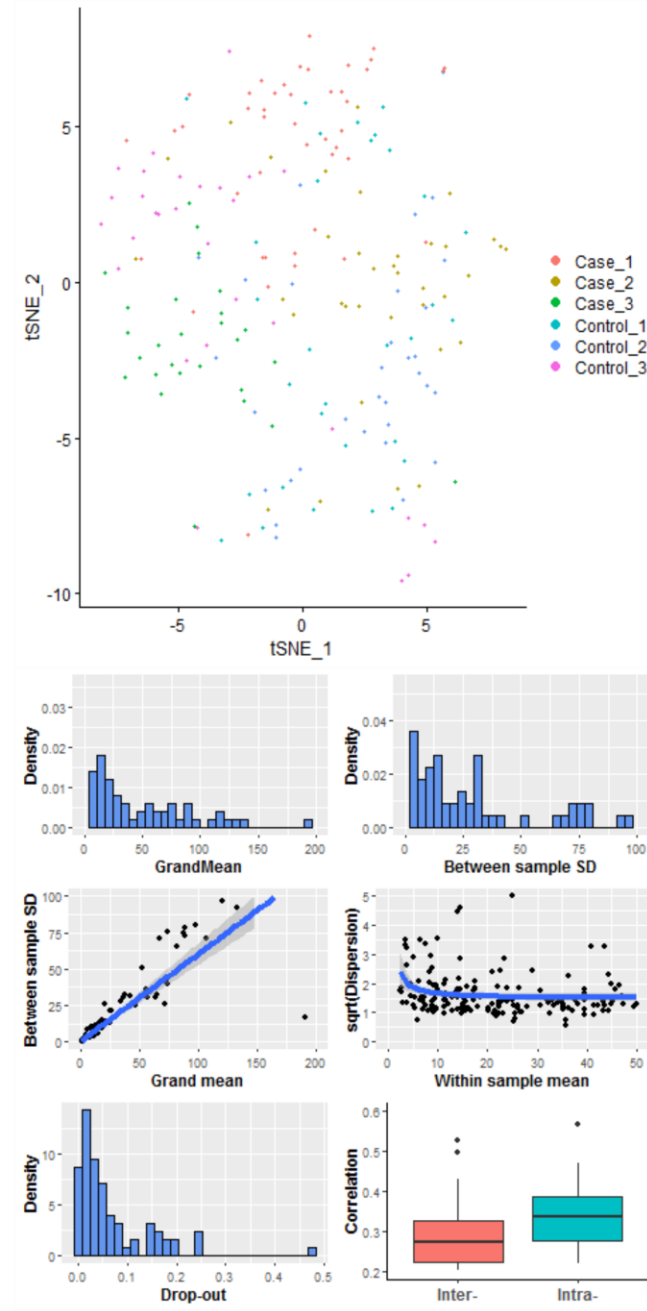


Supplementary Fig. 3: Similarities between simulated data and real data. tSNE plots and plots of various parameters demonstrate similarities between simulated and real data. All axes, except tSNE plot axes, are held constant across the two sets of panels. Parameters were estimated on pancreatic beta cells with 500 genes correlated with a Spearman's coefficient < 0.25 . The simulated data consist of 500 independently simulated genes. Top panels: tSNE plots of the data. The next panel below: histograms of the grand mean and between-sample standard deviation. Middle set of panels: relationships between the grand mean and the between-sample standard deviation and relationships between intra-individual means and dispersion. Bottom set of panels: dropout results and the boxplot demonstrate distributions of intra-individual (blue) and inter-individual (red) correlations. Differences between inter- and intra-individual correlations are less exaggerated in our simulated data. The center line represents the median. The lower and upper box limits represent the 25% and 75% quantiles, respectively. The whiskers extend to the largest observation within the box limit plus or minus one and a half multiplied by the interquartile range.

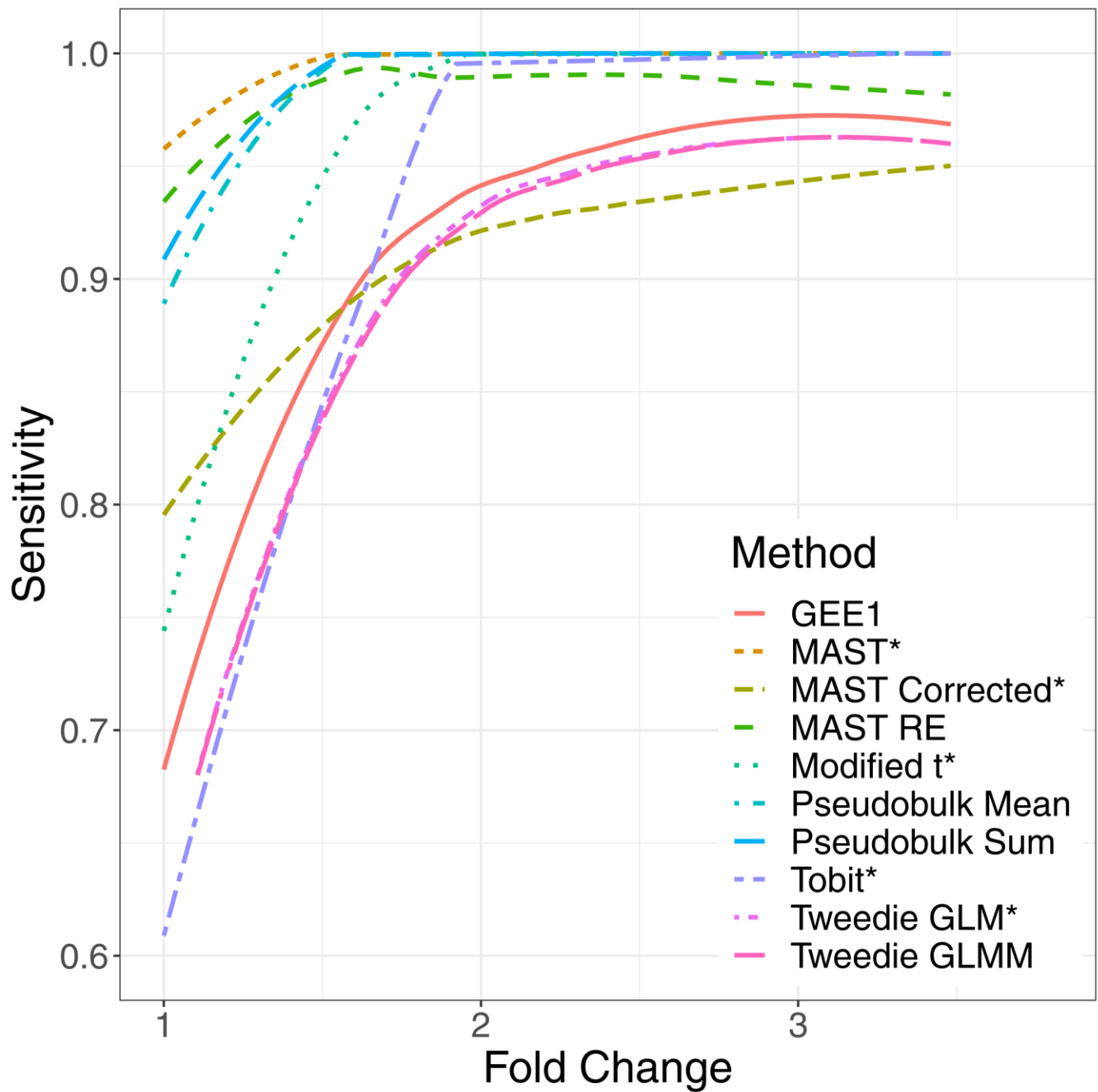
Real data - Mucosal cells (6 samples, ~30 cells per sample)



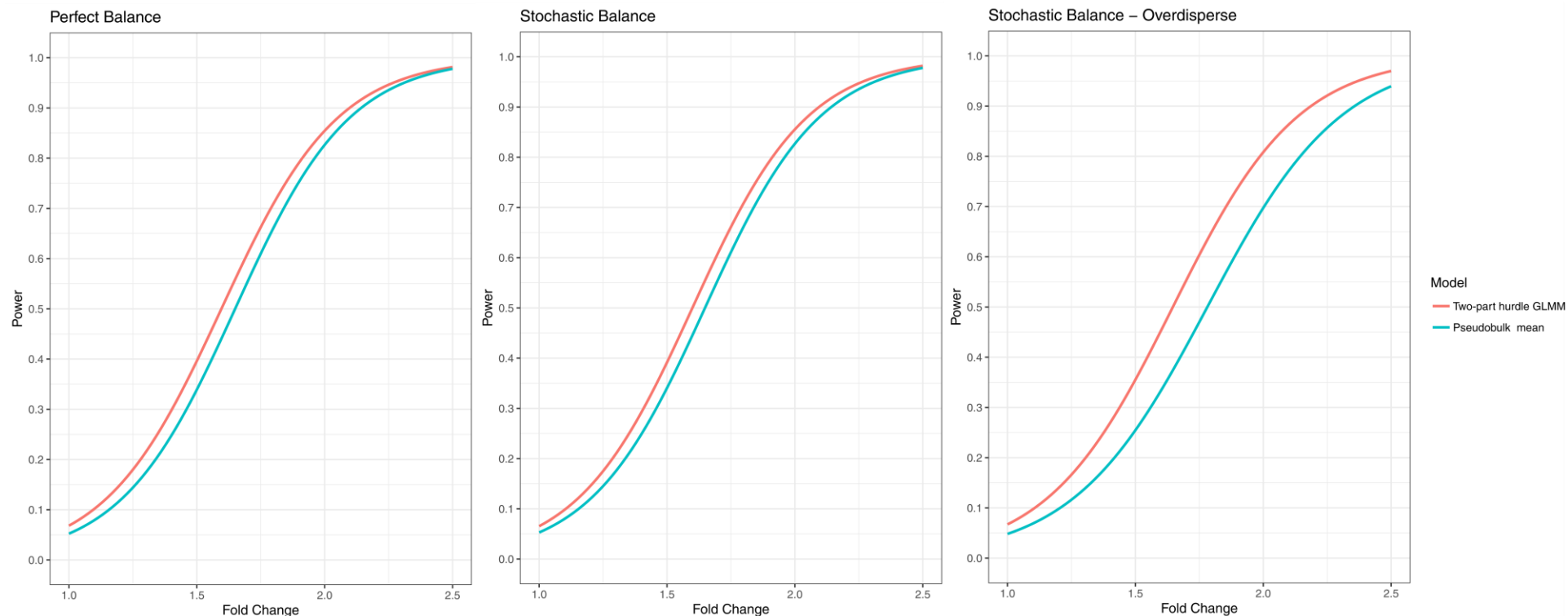
Simulated data (6 samples, ~30 cells per sample)



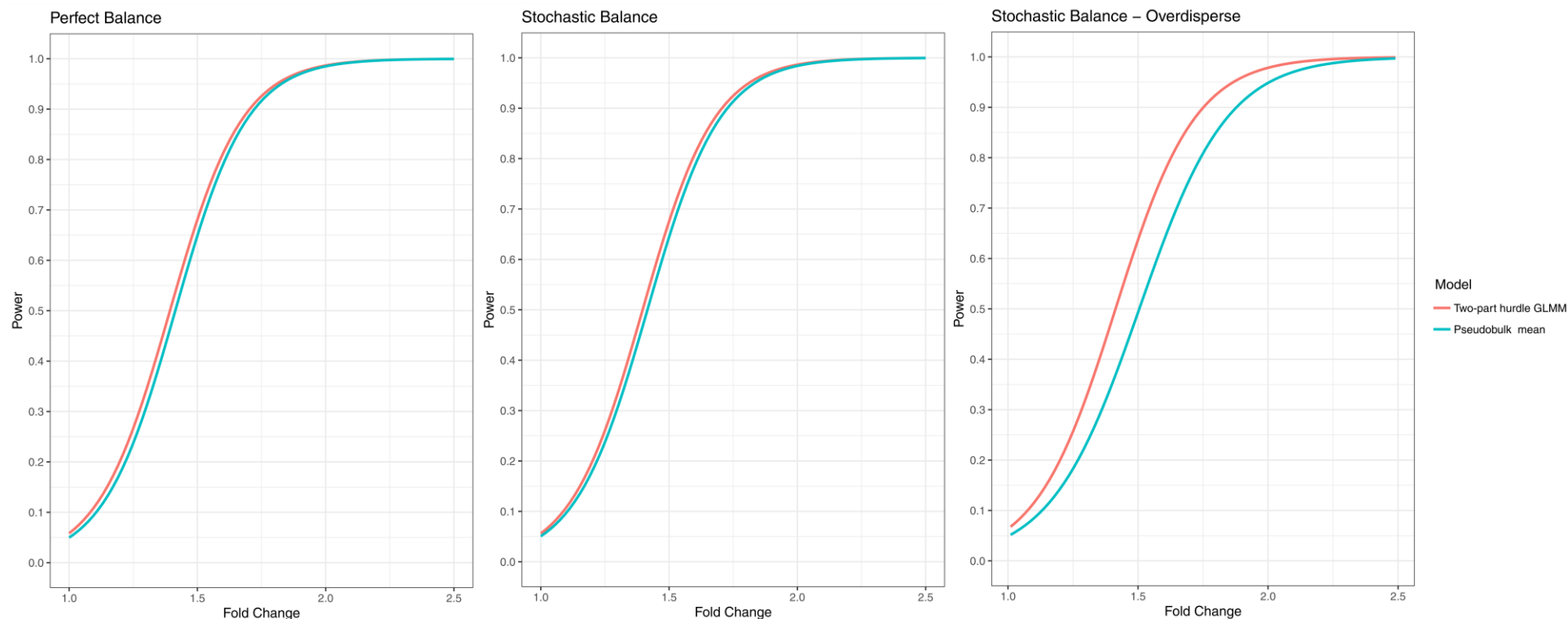
Supplementary Fig. 4: Similarities between simulated data and real data. tSNE plots and plots of various parameters demonstrate similarities between simulated and real data. All axes, except tSNE plot axes, are held constant across the two sets of panels. Parameters were estimated on normal mucosal cells with 215 genes correlated with a Spearman's coefficient < 0.25 . The simulated data consist of 215 independently simulated genes. Top panels: tSNE plots of the data. The next panel below: histograms of the grand mean and between-sample standard deviation. Middle set of panels: relationships between the grand mean and the between-sample standard deviation and relationships between intra-individual means and dispersion. Bottom set of panels: dropout results and the boxplot demonstrate distributions of intra-individual (blue) and inter-individual (red) correlations. Differences between inter- and intra-individual correlations are less exaggerated in our simulated data. The center line represents the median. The lower and upper box limits represent the 25% and 75% quantiles, respectively. The whiskers extend to the largest observation within the box limit plus or minus one and a half multiplied by the interquartile range.



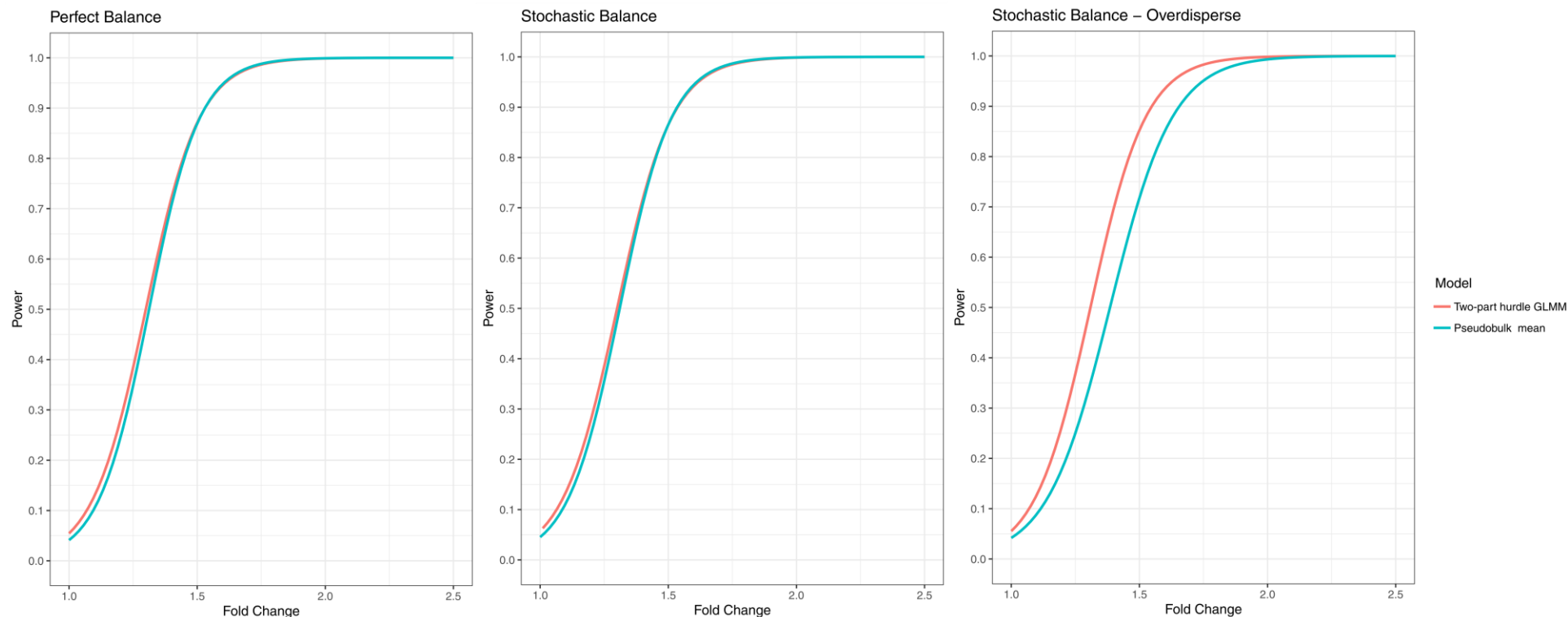
Supplementary Fig. 5: Sensitivity at varying fold changes for each method. Estimates of sensitivity for each method are provided for increasing values of fold change. MAST was implemented without random effects, MAST RE denotes MAST was implemented with random effects, MAST Corrected denotes data were batch-corrected for individual prior to analysis without using individual as a random effect, GLM denotes generalized linear model, and GLMM denotes generalized linear mixed-effects model. Tweedie model as implemented in 'glmmTMB', GEE1 as implemented in 'geepack', Pseudobulk approaches averaged or summed across cells within an individual and were implemented in DESeq2, Modified t as implemented in ROTS, and Tobit as implemented in Monocle. *These methods exhibit inflated type 1 error rates and therefore inflated sensitivity.



Supplementary Fig. 6: Power comparisons between the two-hurdle model with a random effect for individual and the pseudo-bulk mean approach. Power curves for a two-hurdle model with a random effect for individual (red lines) and a pseudo-bulk mean approach (blue lines) when there are 20 individuals per treatment group. Left: differences in power between the two approaches when each individual has exactly 100 cells. Middle: differences in power between the two approaches when the number of cells for each individual is drawn from a Poisson distribution with an expected value of 100 cells. Right: Differences in power between the two approaches when the number of cells for each individual is drawn from a negative binomial distribution with an expected value of 100 cells and a dispersion parameter equal to one.



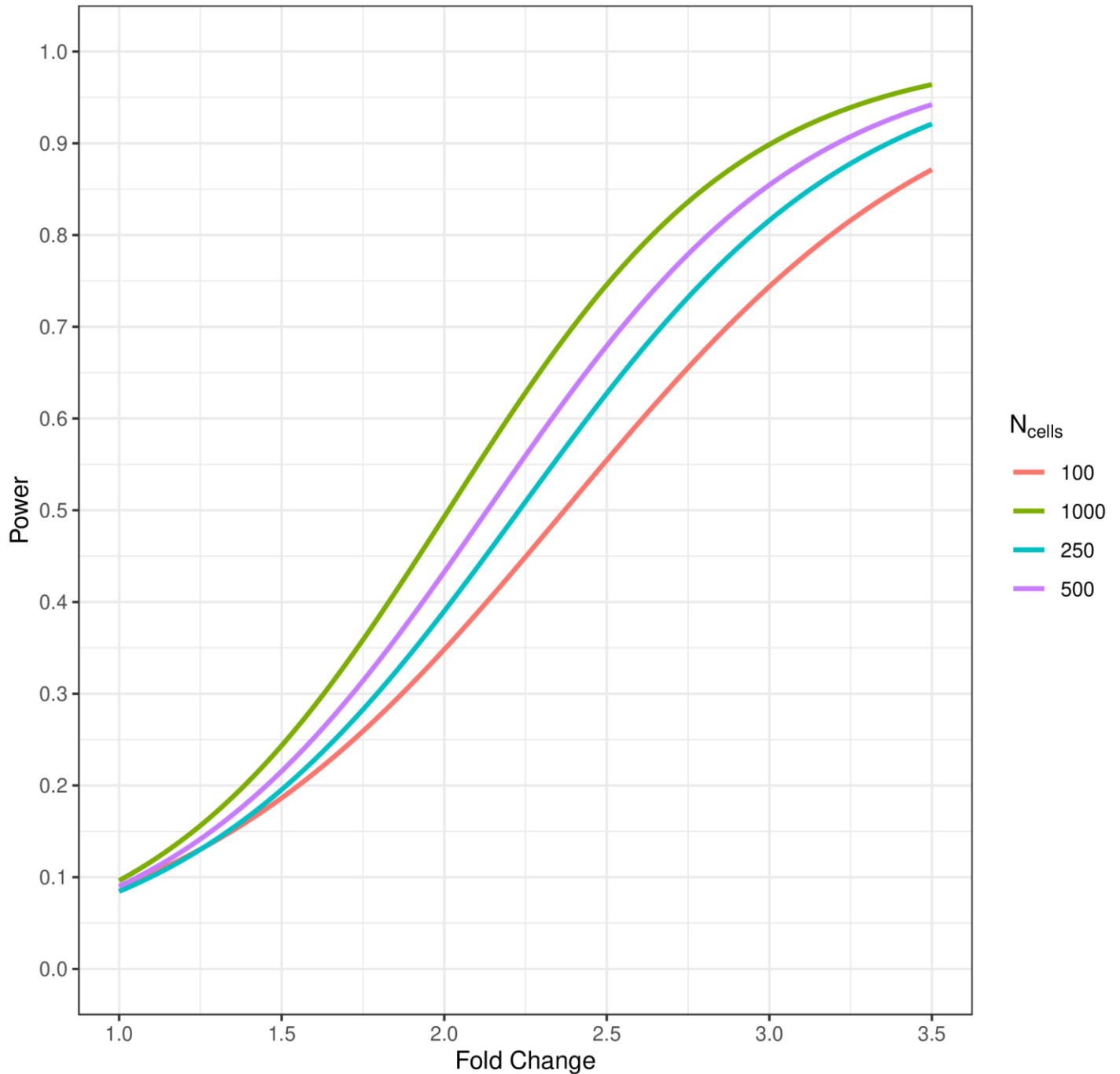
Supplementary Fig. 7: Power comparisons between the two-hurdle model with a random effect for individual and the pseudo-bulk mean approach. Power curves for a two-hurdle model with a random effect for individual (red lines) and a pseudo-bulk mean approach (blue lines) when there are 40 individuals per treatment group. Left panel: Differences in power between the two approaches when each individual has exactly 100 cells. Middle: Differences in power between the two approaches when the number of cells for each individual is drawn from a Poisson distribution with an expected value of 100 cells. Right: Difference in power between the two approaches when the number of cells for each individual is drawn from a negative binomial distribution with an expected value of 100 cells and a dispersion parameter equal to one.



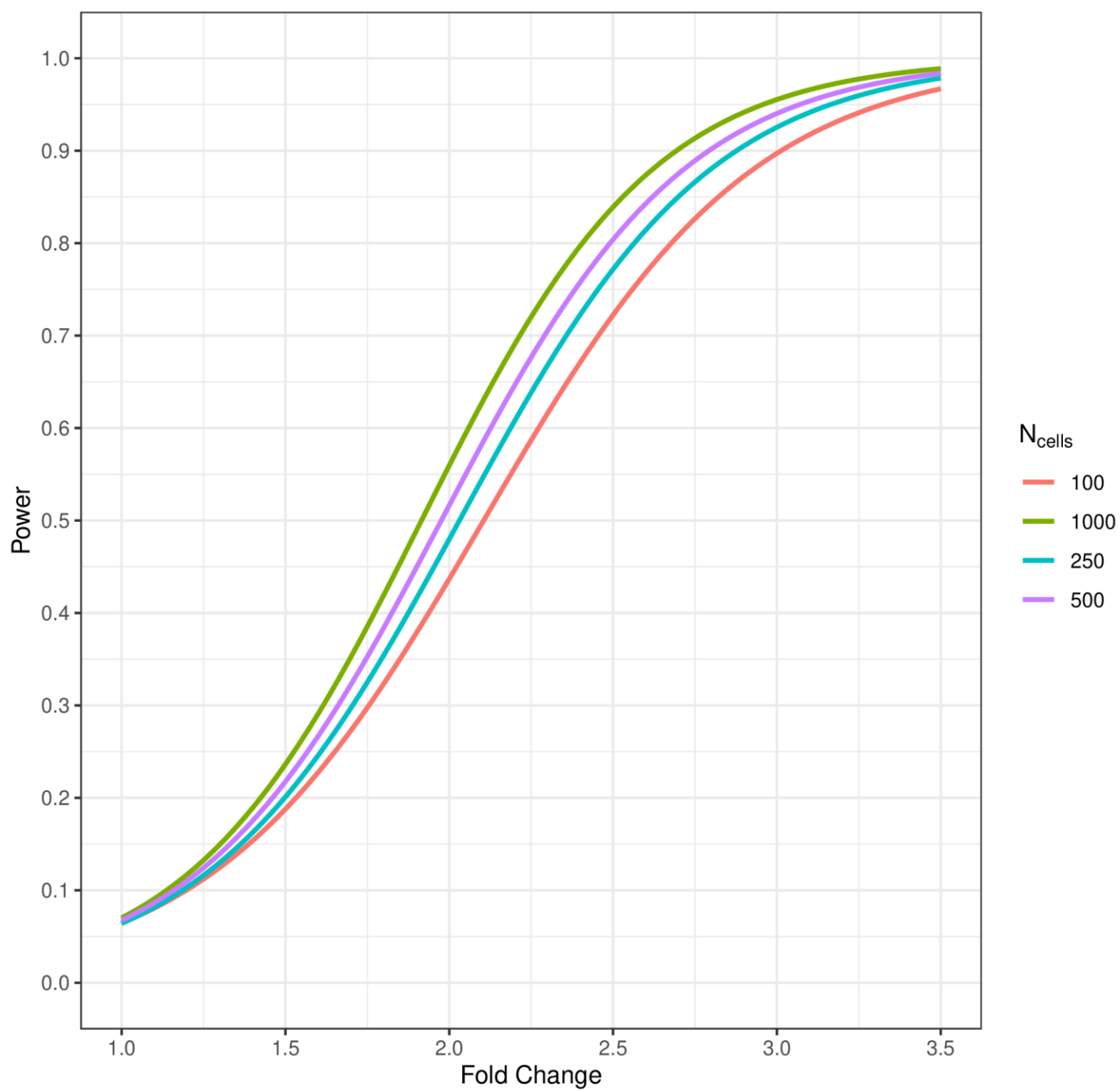
Supplementary Fig. 8: Power comparisons between the two-hurdle model with a random effect for individual and the pseudo-bulk mean approach. Power curves for a two-hurdle model with a random effect for individual (red lines) and a pseudo-bulk mean approach (blue lines) when there are 60 individuals per treatment group. Left panel: Differences in power between the two approaches when each individual has exactly 100 cells. Middle: Differences in power between the two approaches when the number of cells for each individual is drawn from a Poisson distribution with an expected value of 100 cells. Right: Difference in power between the two approaches when the number of cells for each individual is drawn from a negative binomial distribution with an expected value of 100 cells and a dispersion parameter equal to one.

Supplementary Fig. 9: Power calculations using MAST with a random effect for the individual. Power curves for MAST using a random effect to account for intra-individual correlation. Curves are computed for 100 (red line), 250, Blue line) 500 (purple line), and 1,000 (red line) cells per individual with $\alpha = 0.05$. The numbers of individuals per group range from 3 to 100 and are listed above each plot.

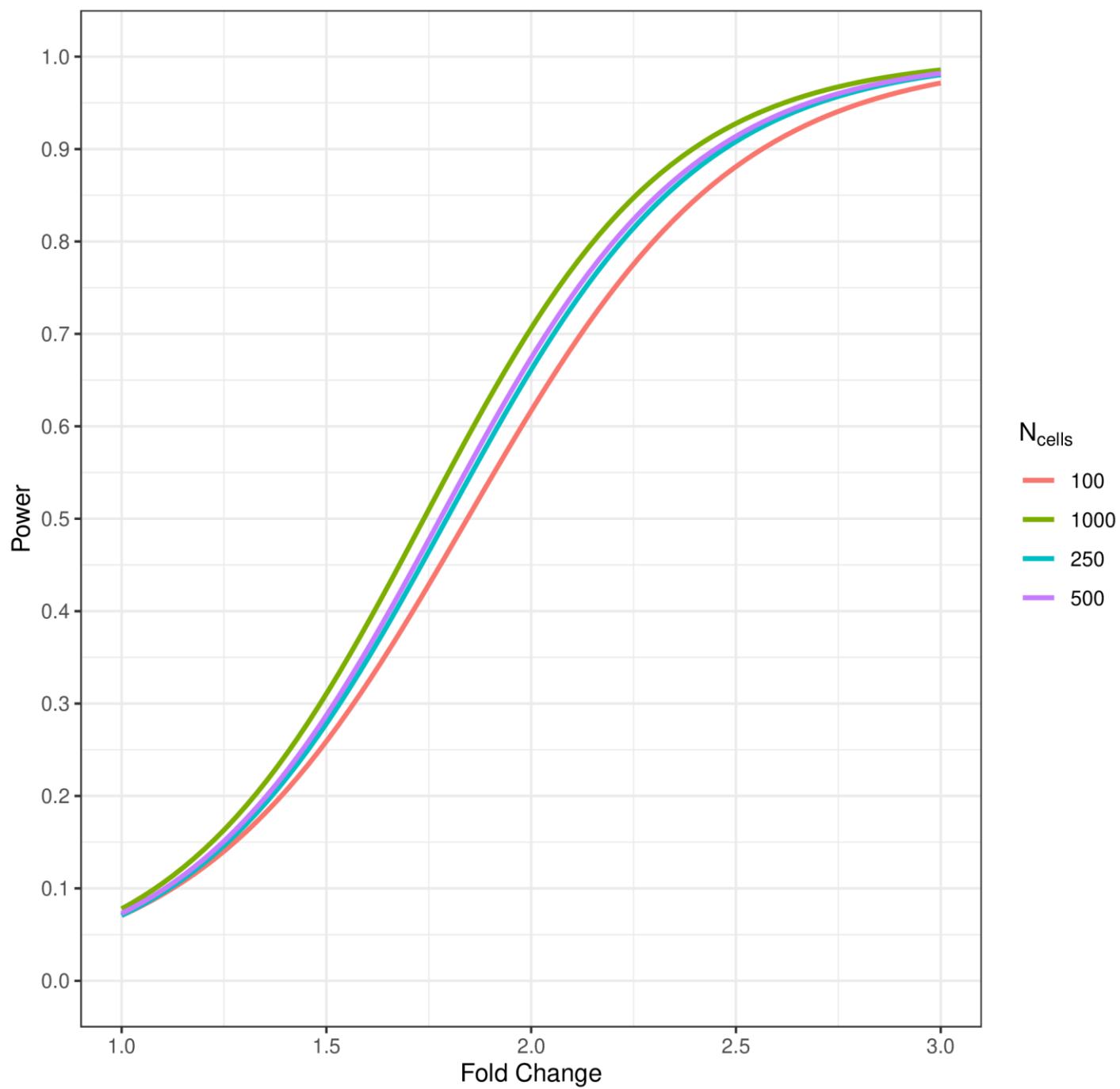
3 Individuals per Group



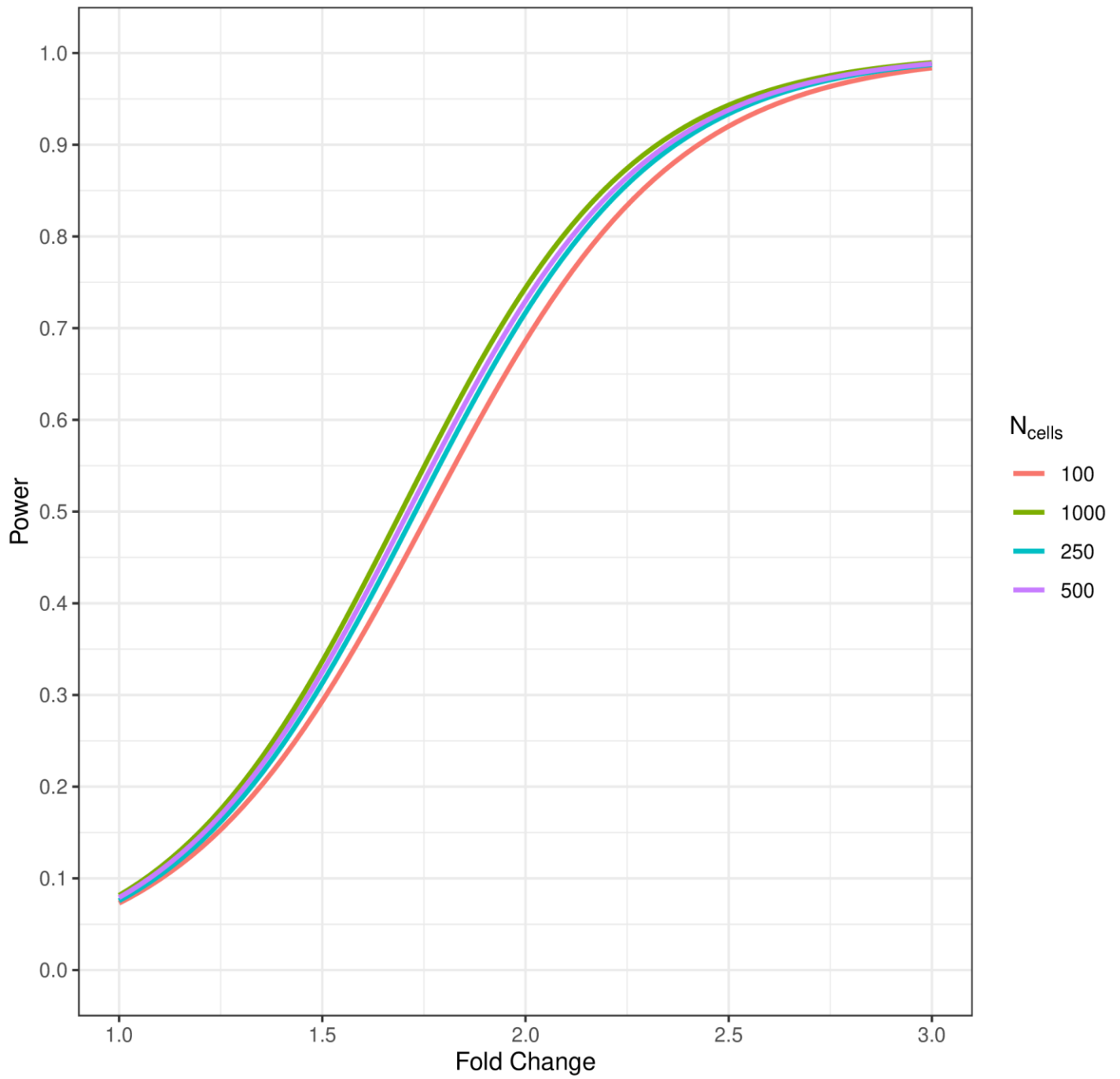
5 Individuals per Group



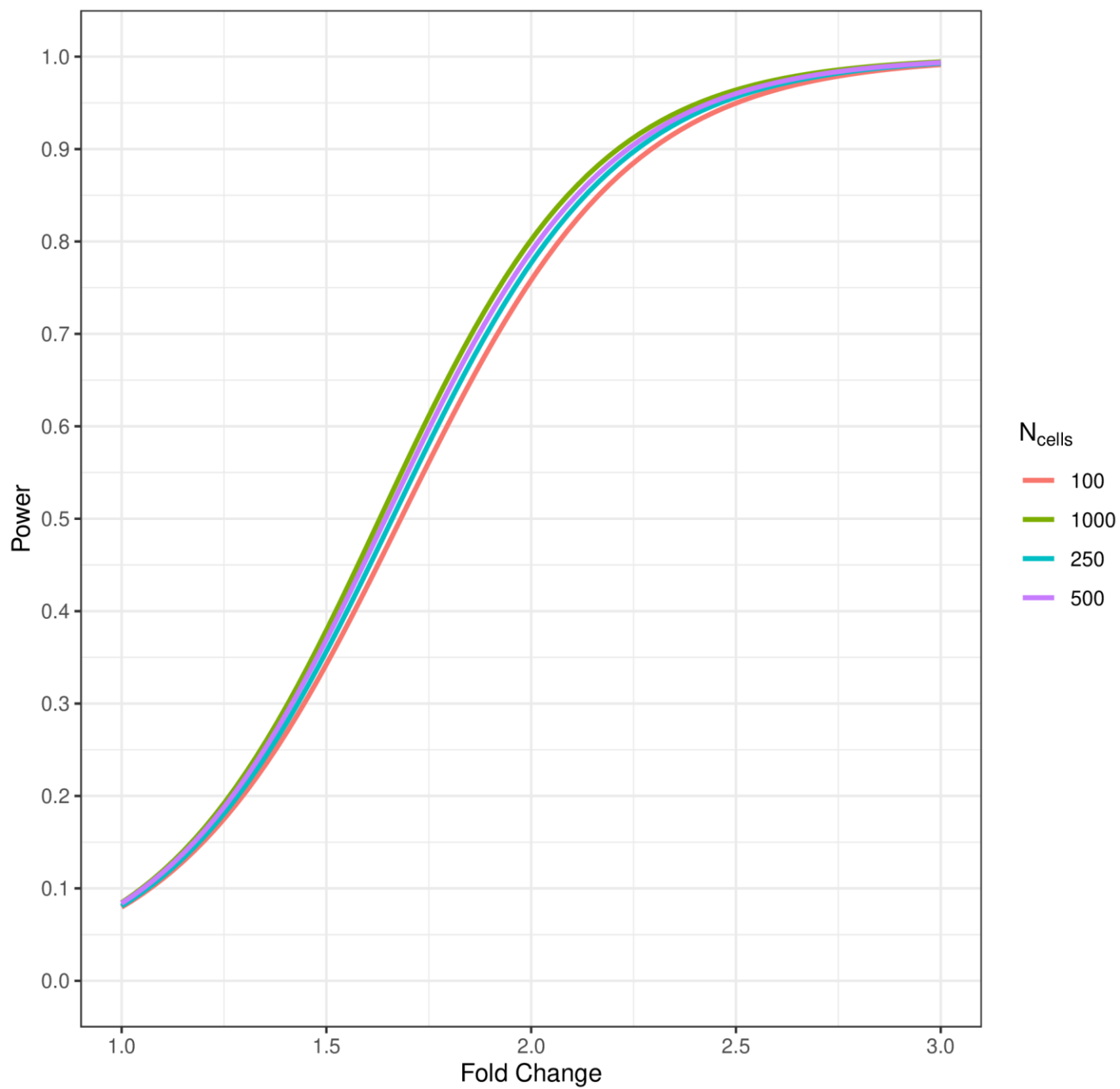
10 Individuals per Group



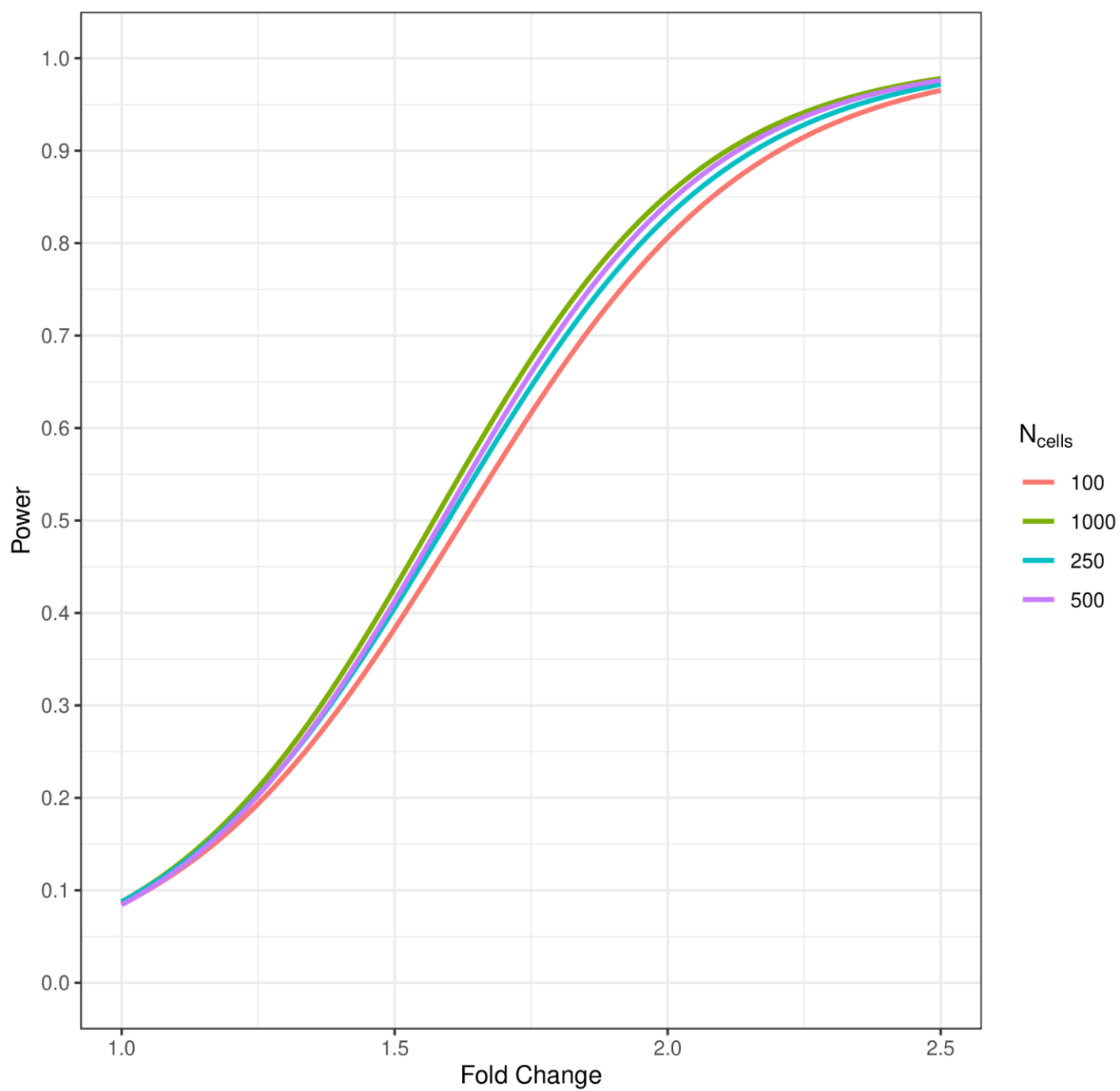
12 Individuals per Group



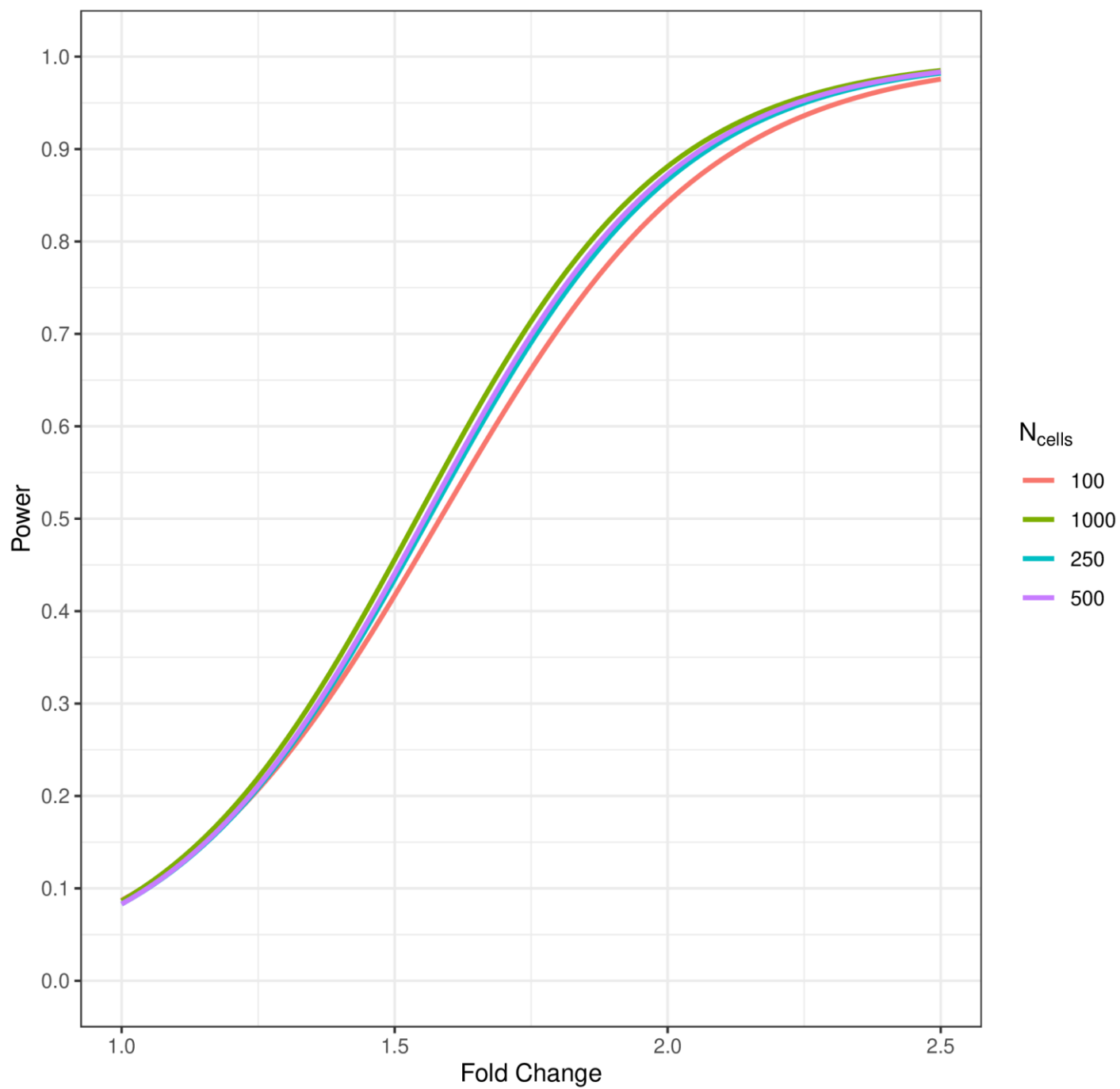
15 Individuals per Group



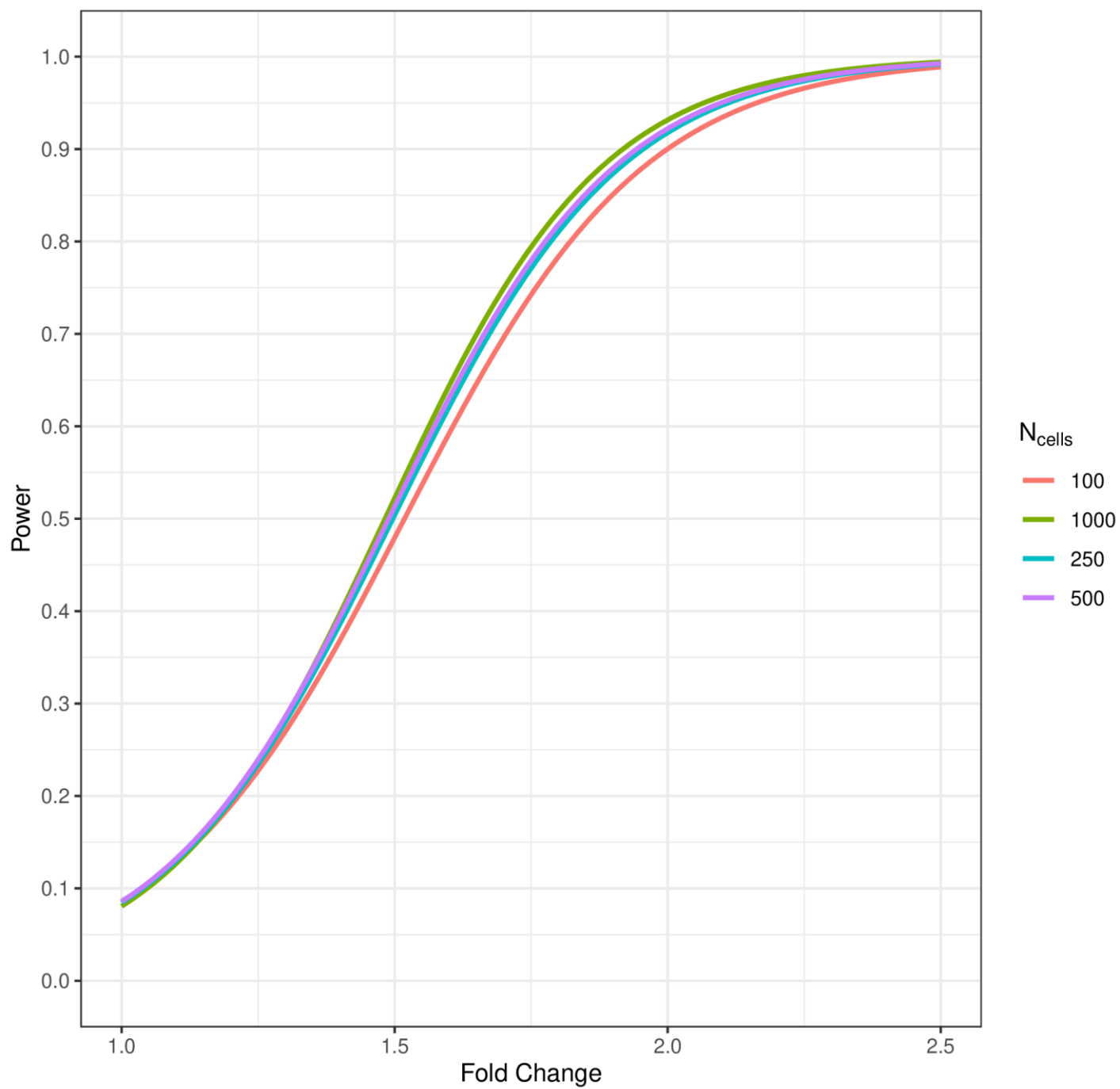
18 Individuals per Group



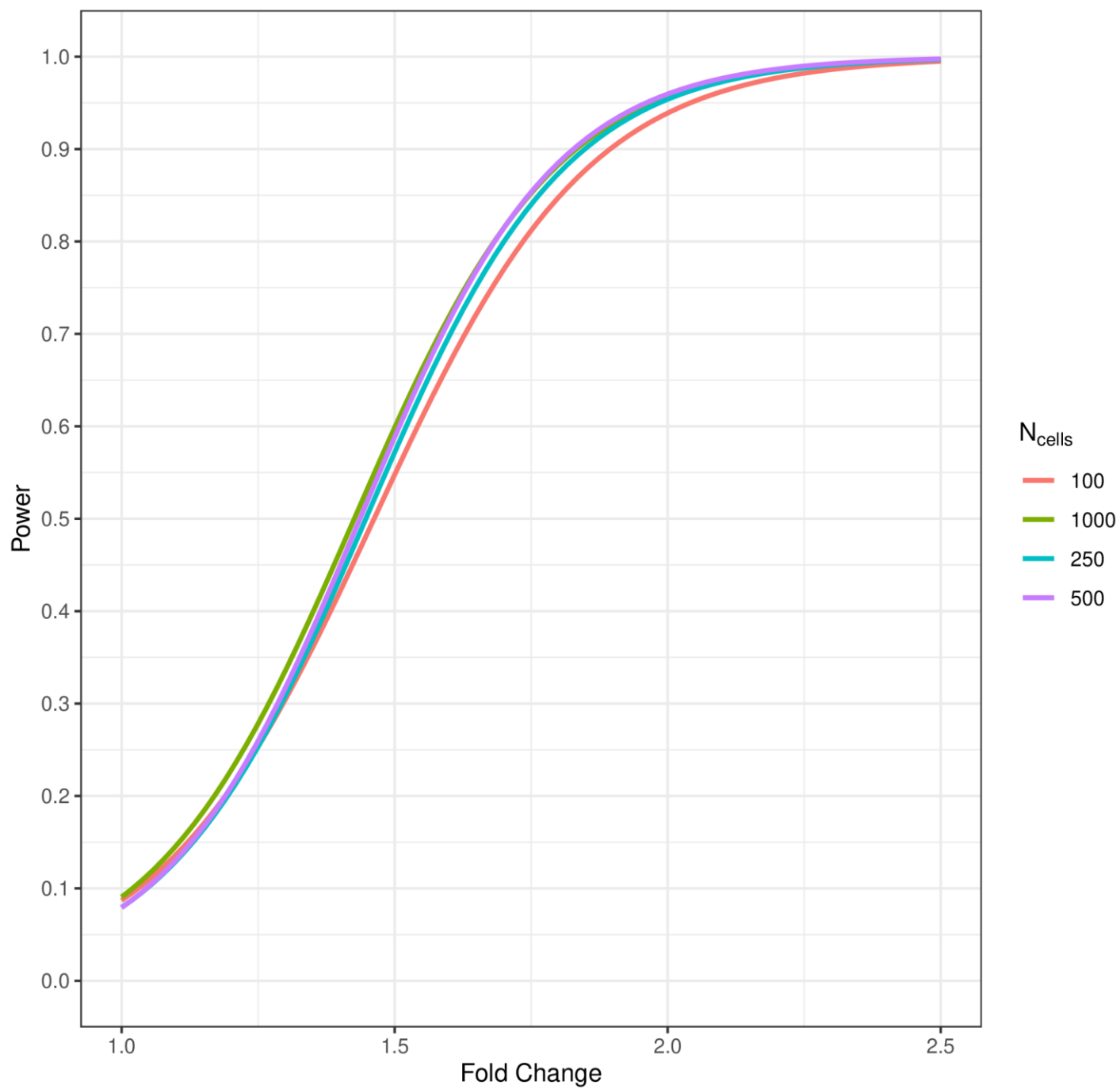
20 Individuals per Group



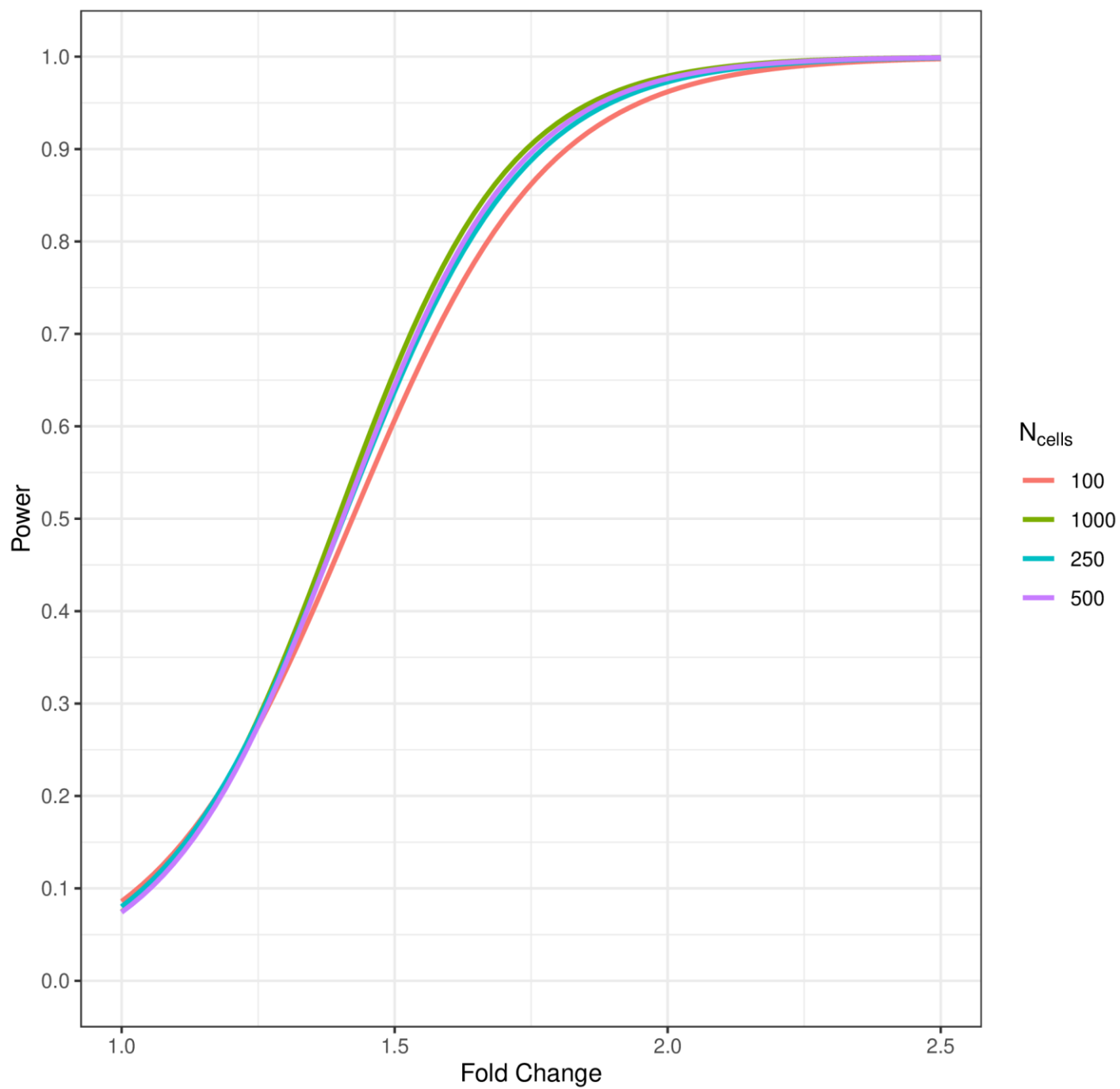
25 Individuals per Group



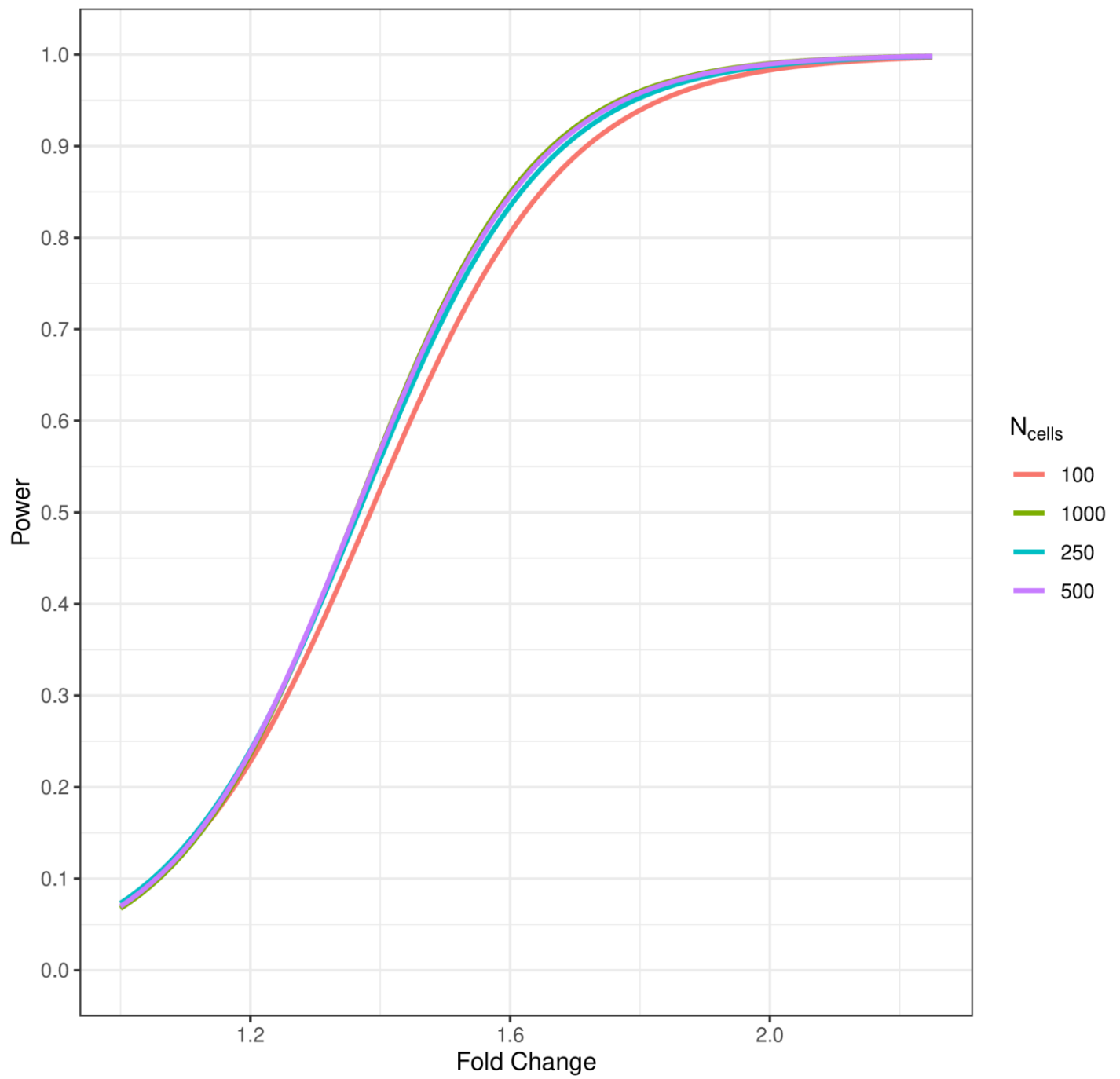
30 Individuals per Group



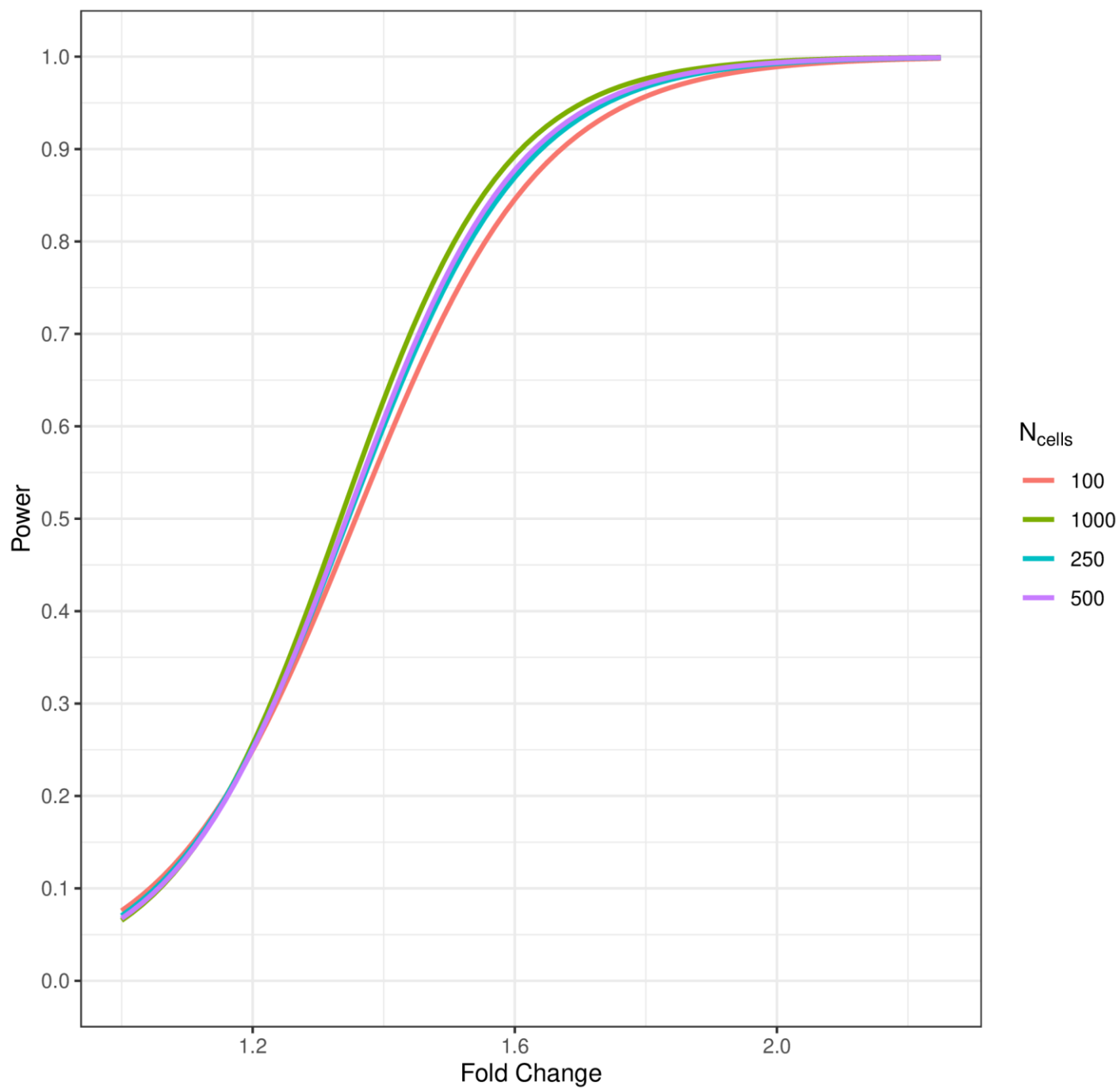
35 Individuals per Group



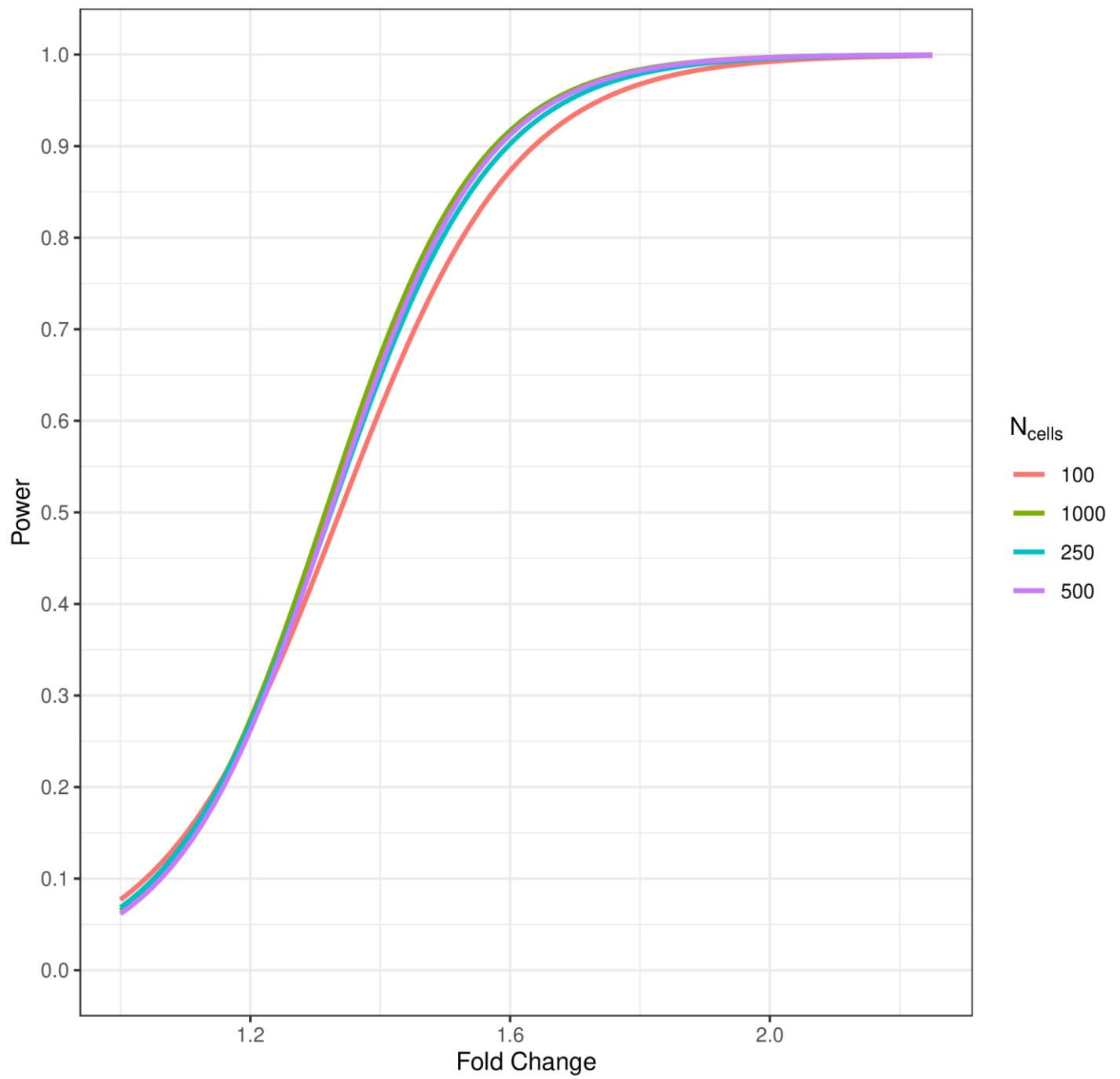
40 Individuals per Group



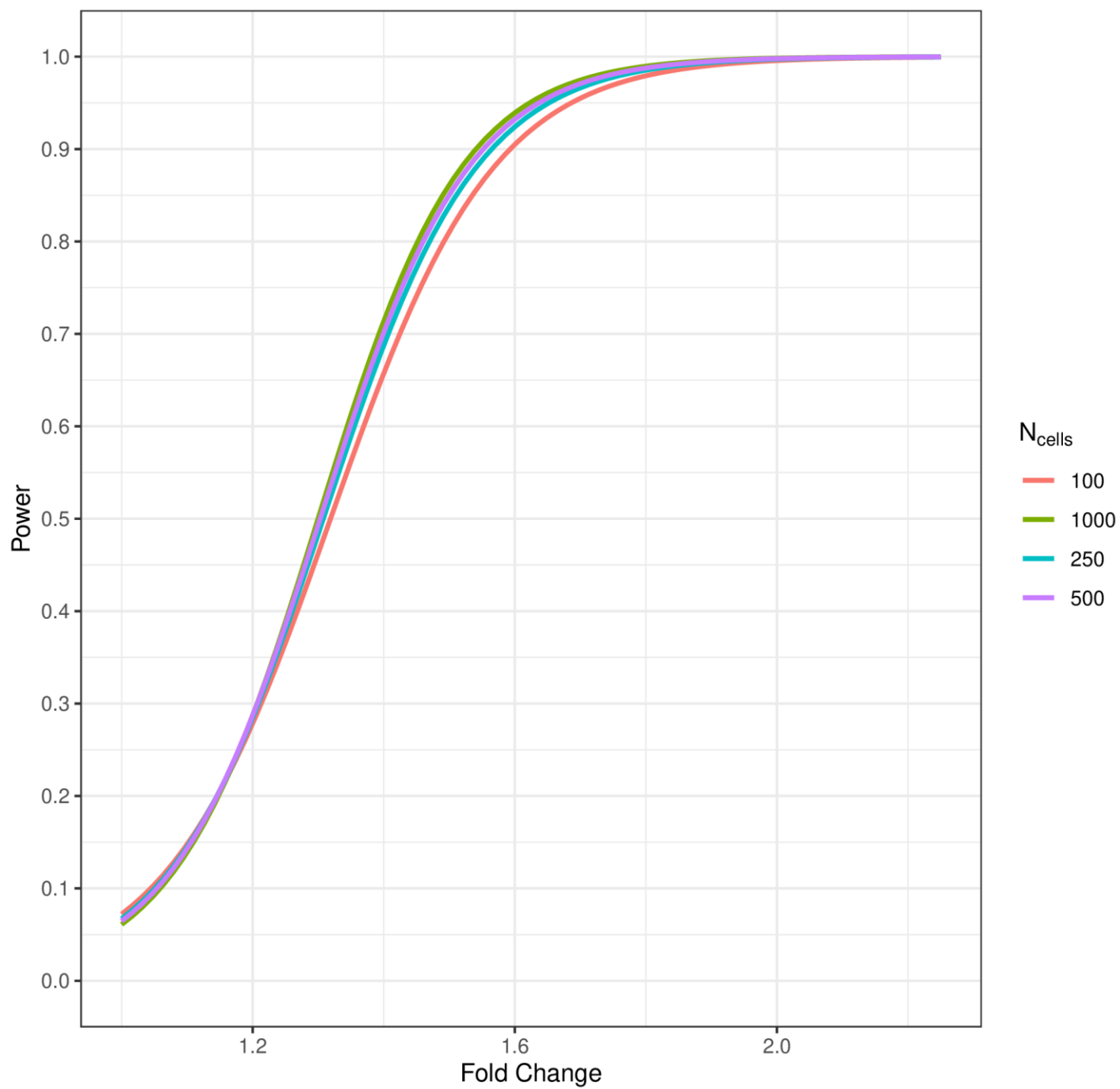
45 Individuals per Group



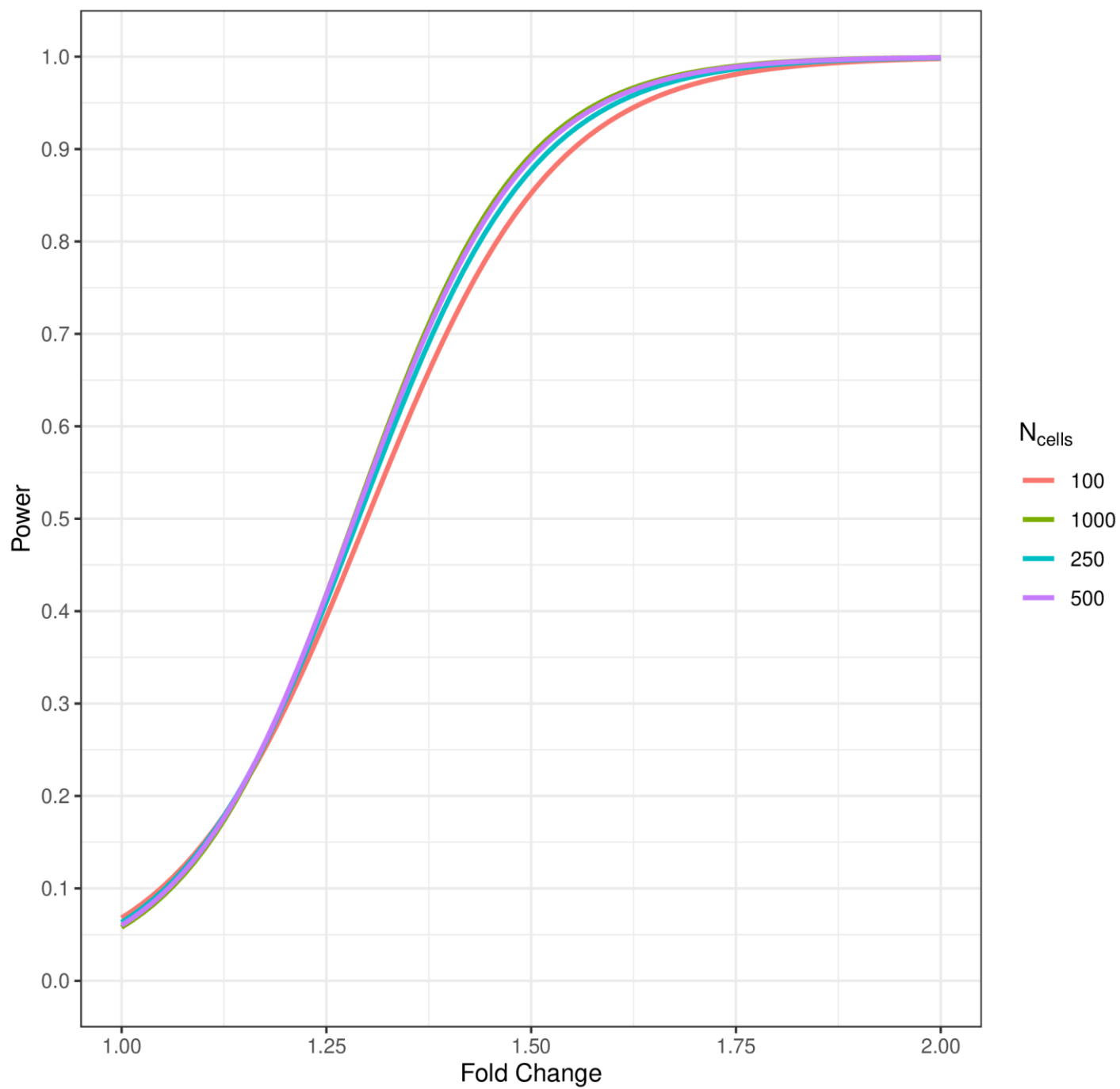
50 Individuals per Group



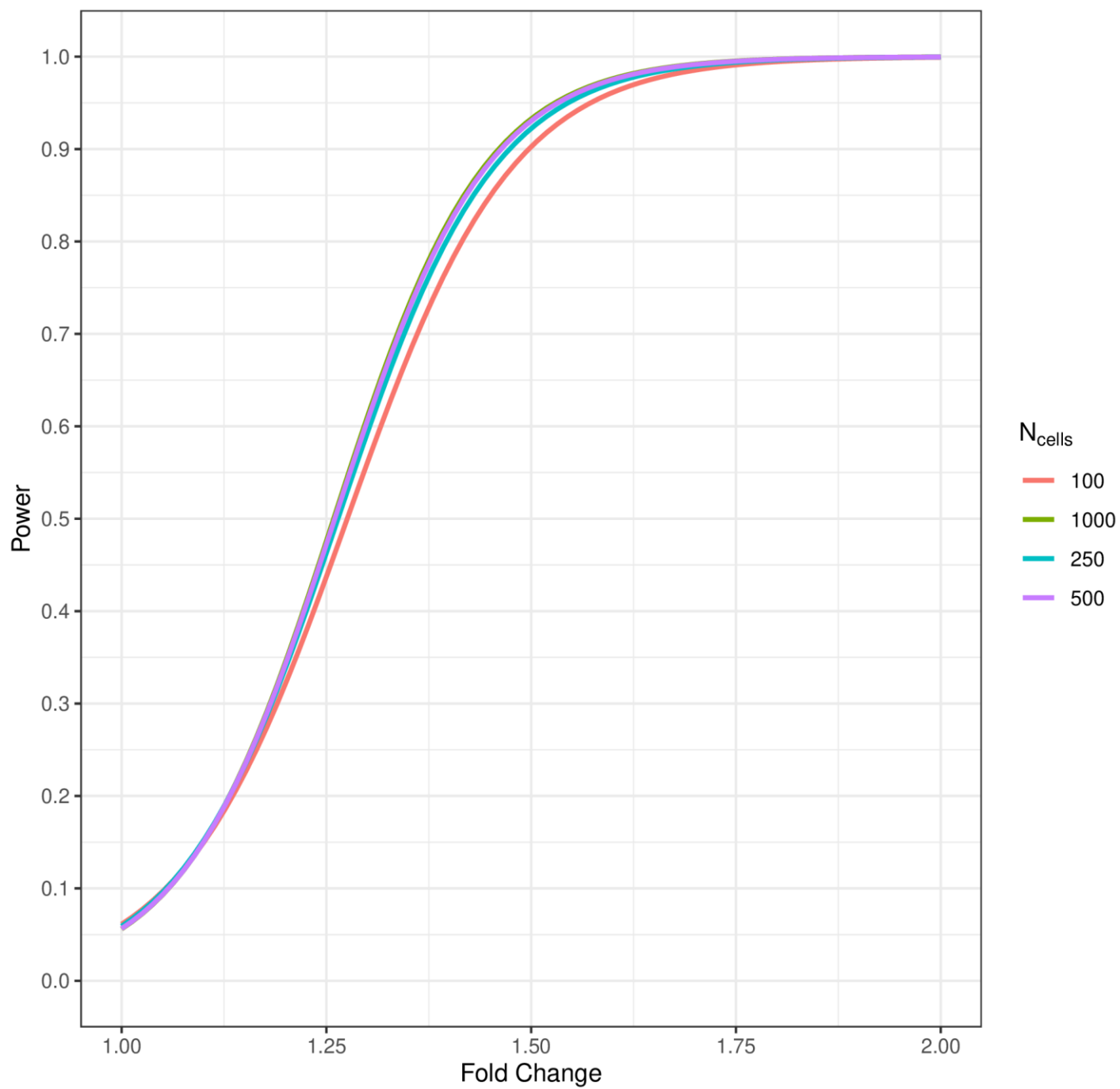
55 Individuals per Group



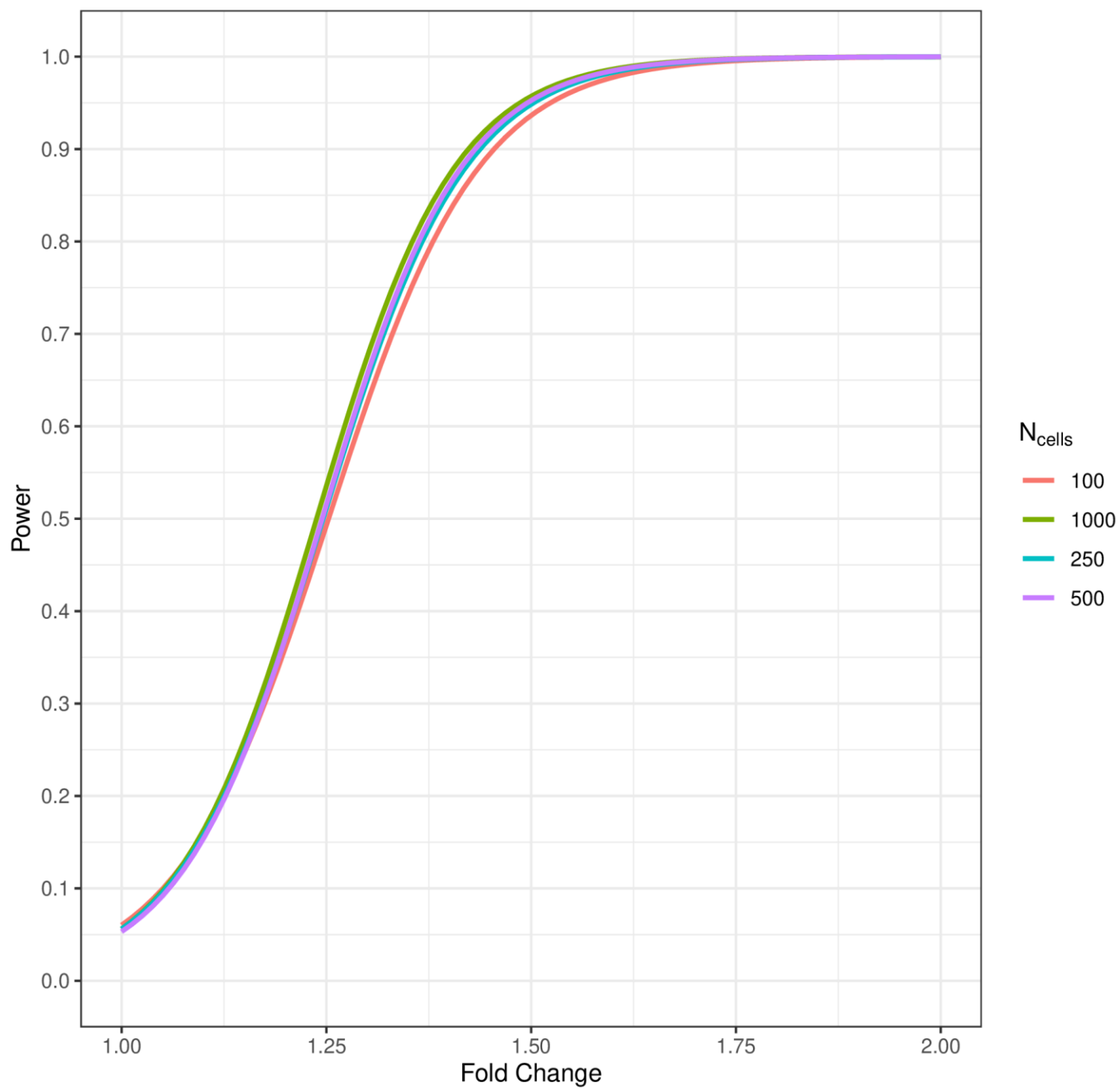
60 Individuals per Group



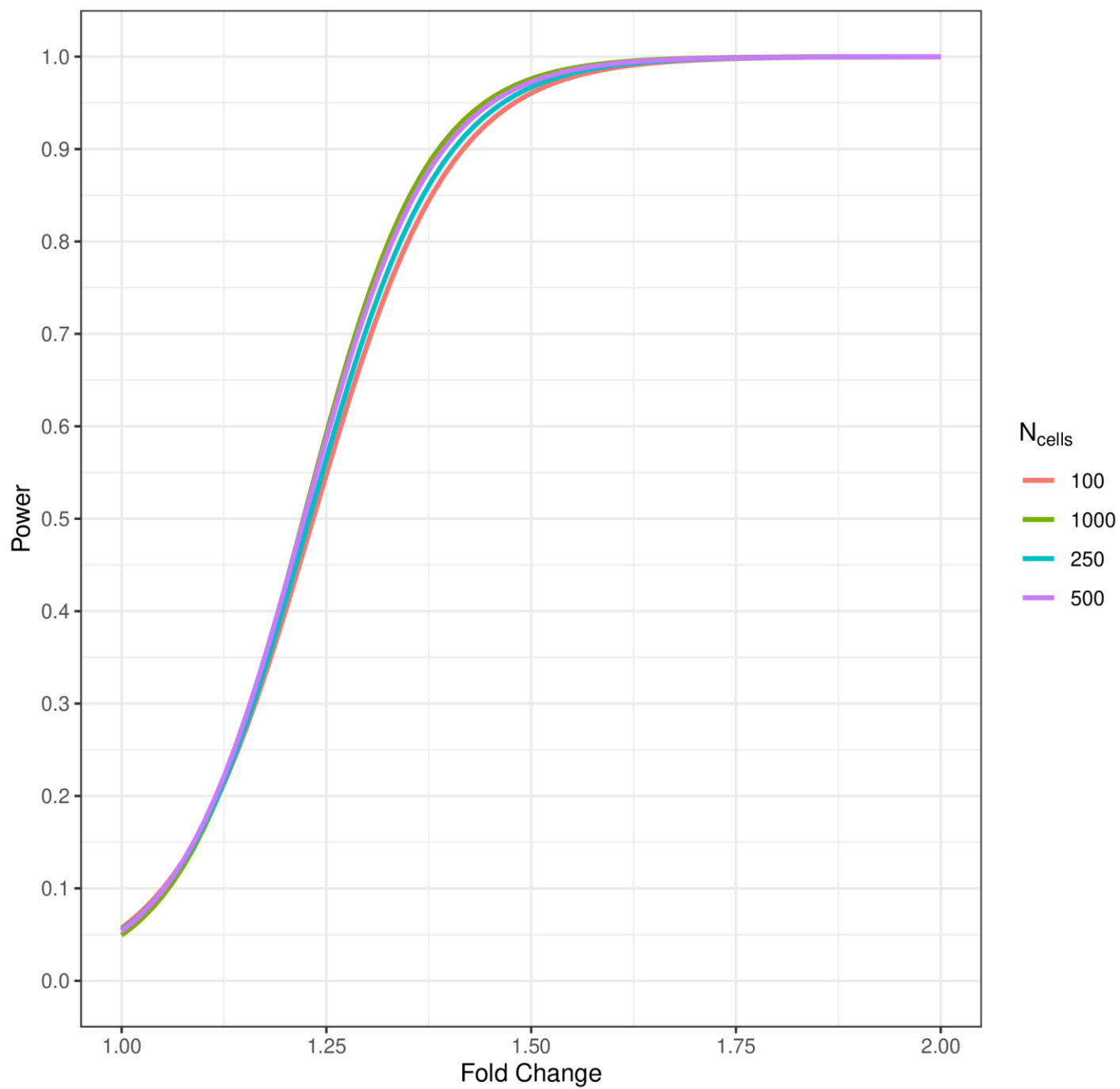
70 Individuals per Group



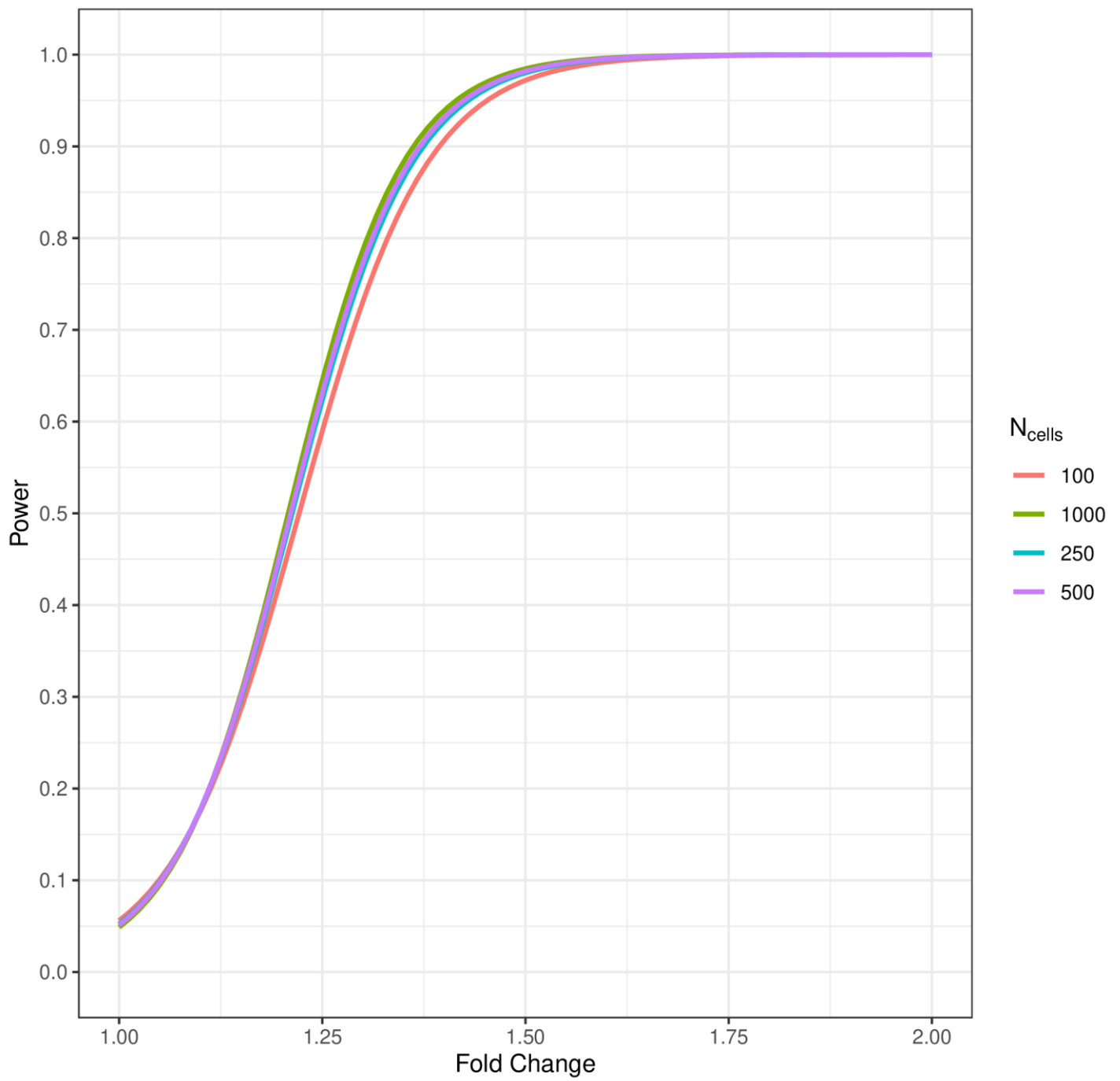
80 Individuals per Group



90 Individuals per Group

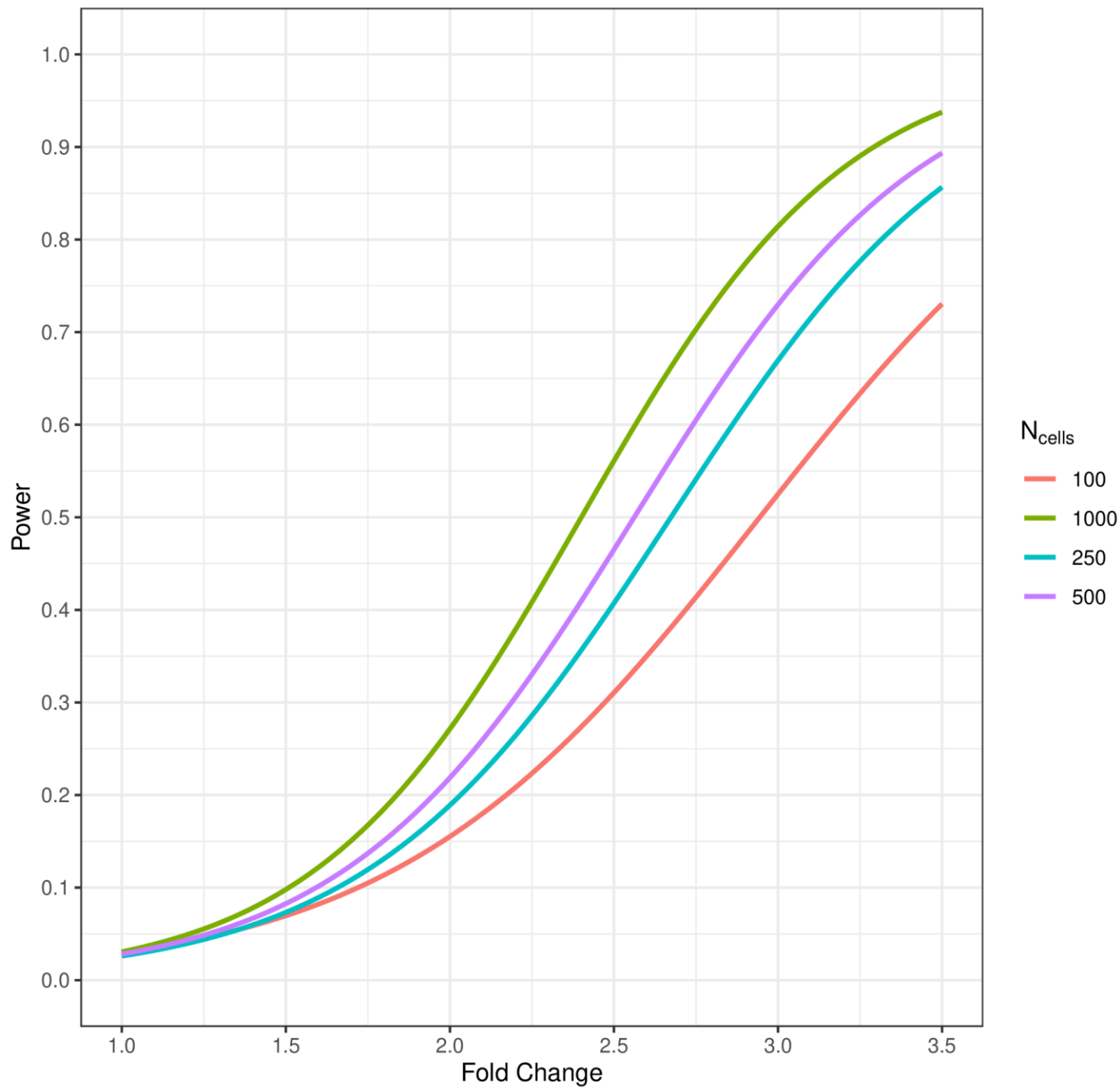


100 Individuals per Group

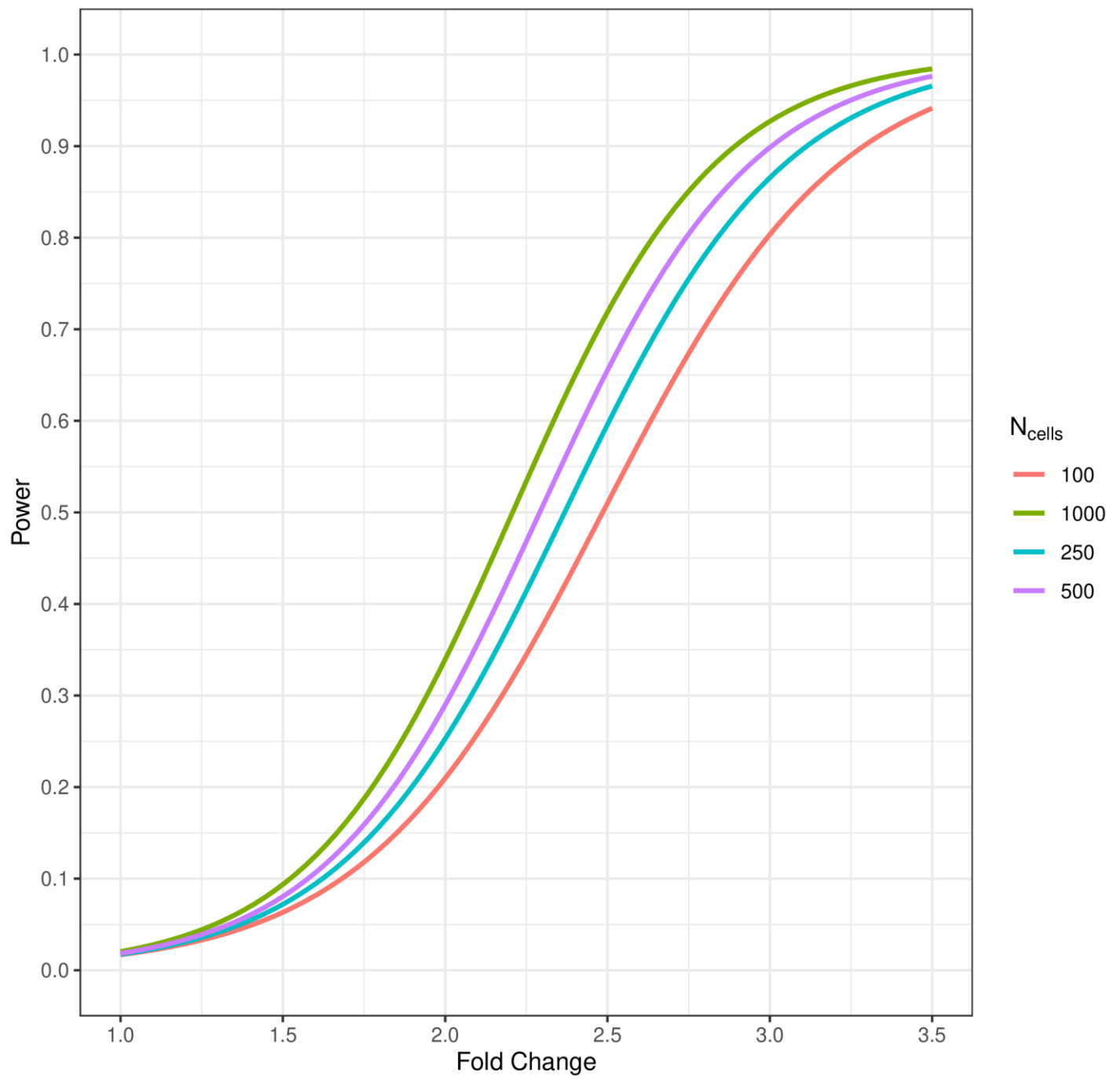


Supplementary Fig. 10: Power calculations using MAST with a random effect for individual. Power curves for MAST using a random effect to account for intra-individual correlation. Curves are computed for 100, 250, 500, and 1,000 cells per individual using an $\alpha = 0.01$. The number of individuals per group ranges from 3 to 100 and is listed above each plot.

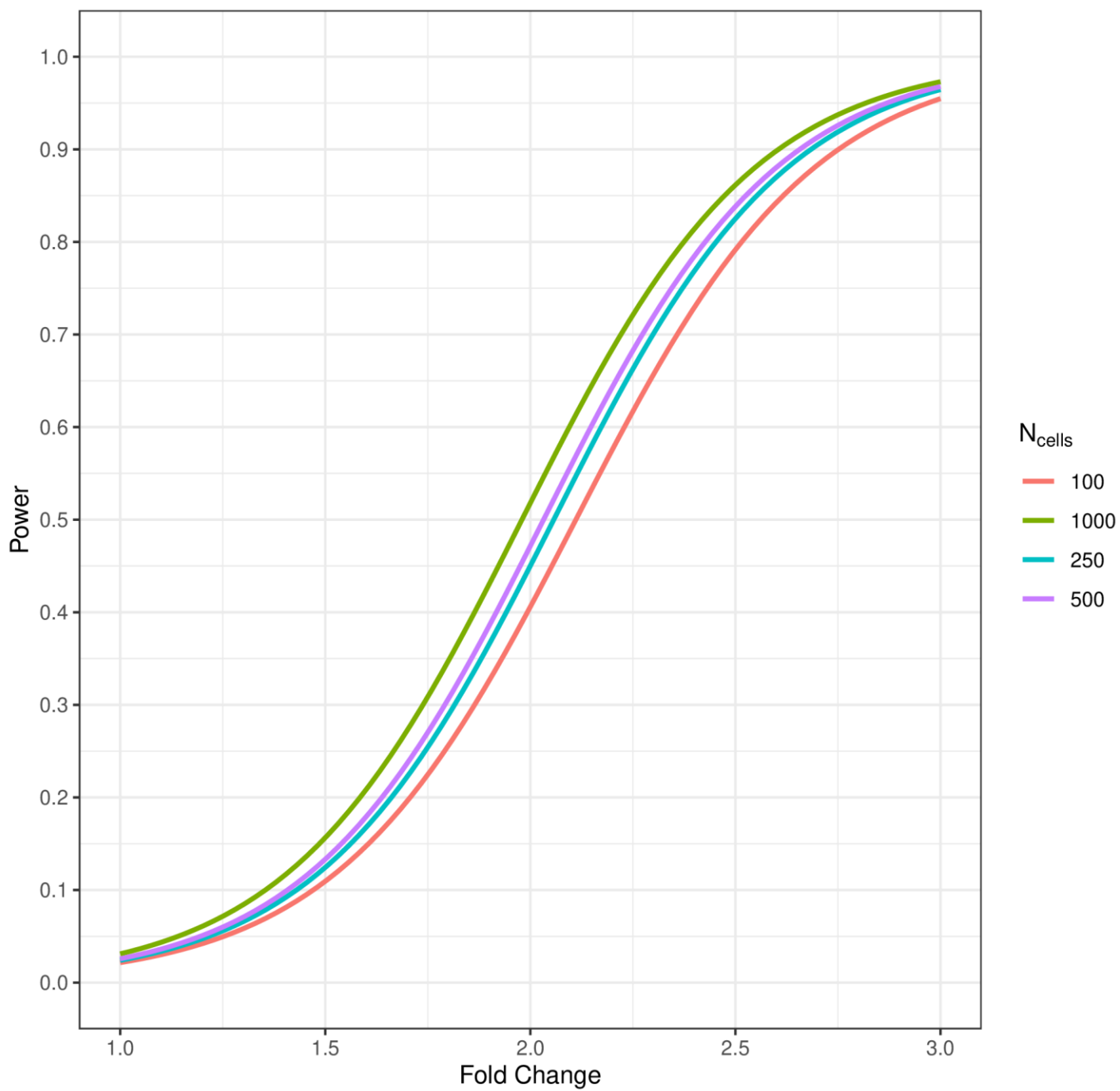
3 Individuals per Group



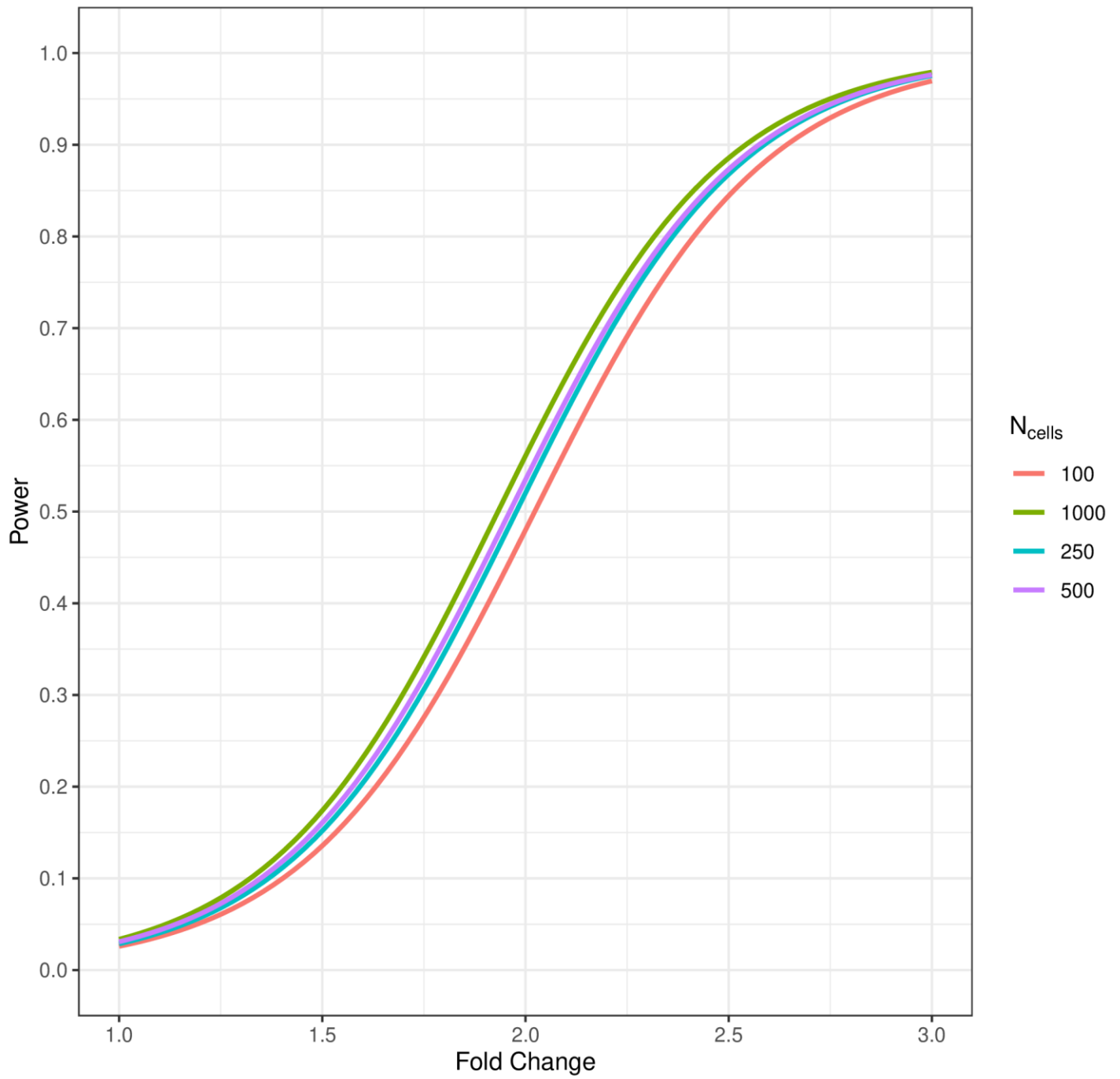
5 Individuals per Group



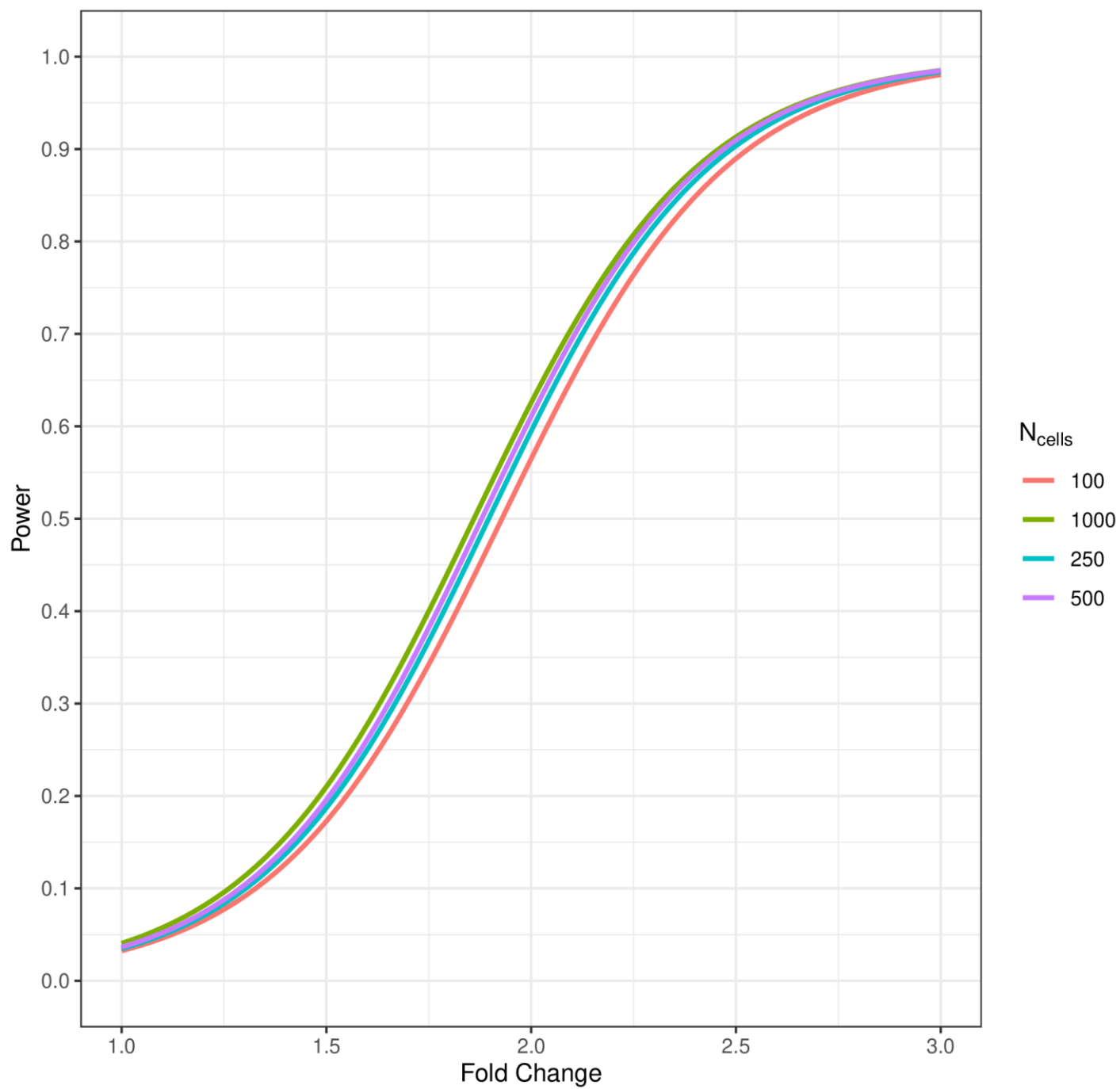
10 Individuals per Group



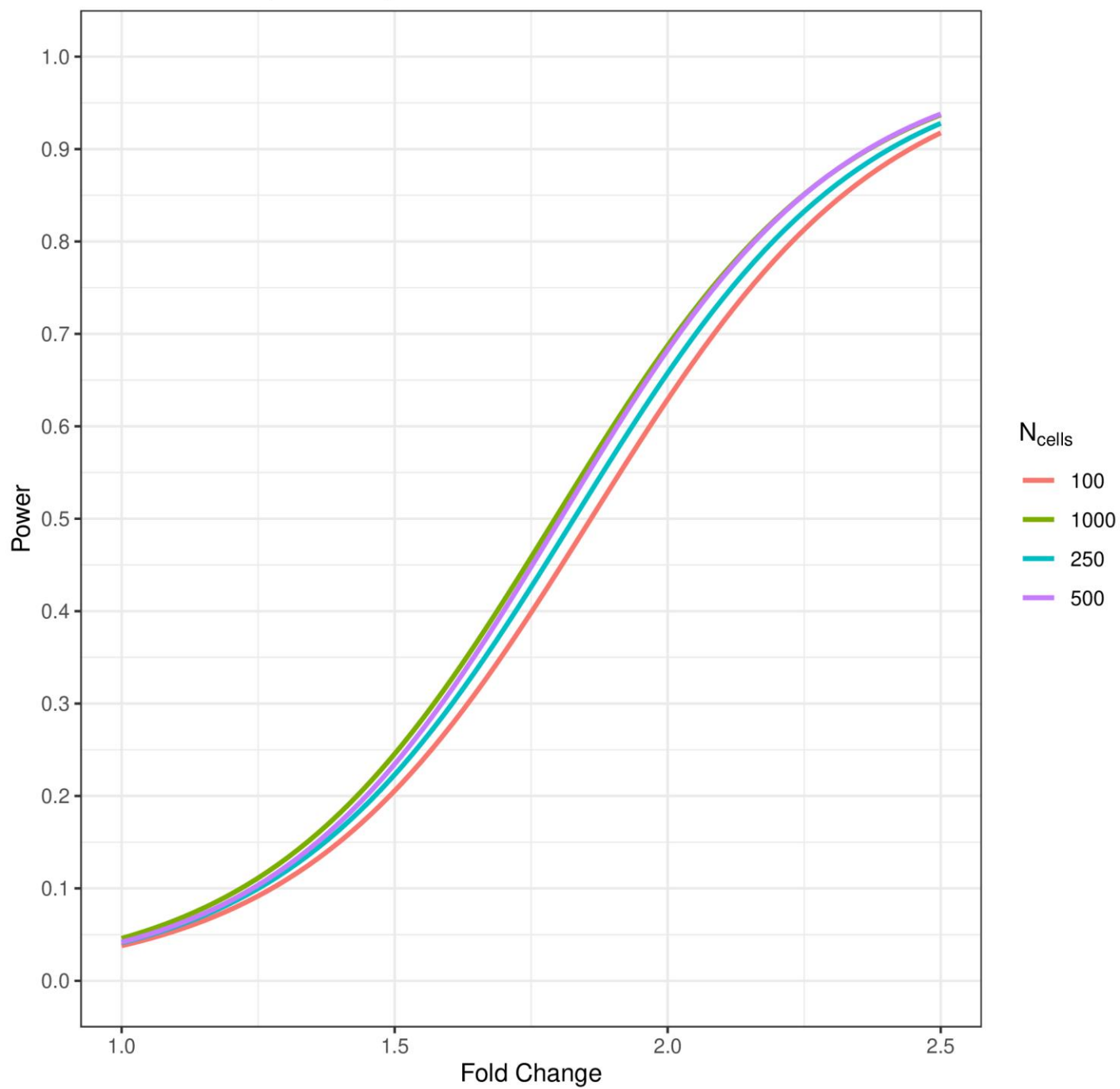
12 Individuals per Group



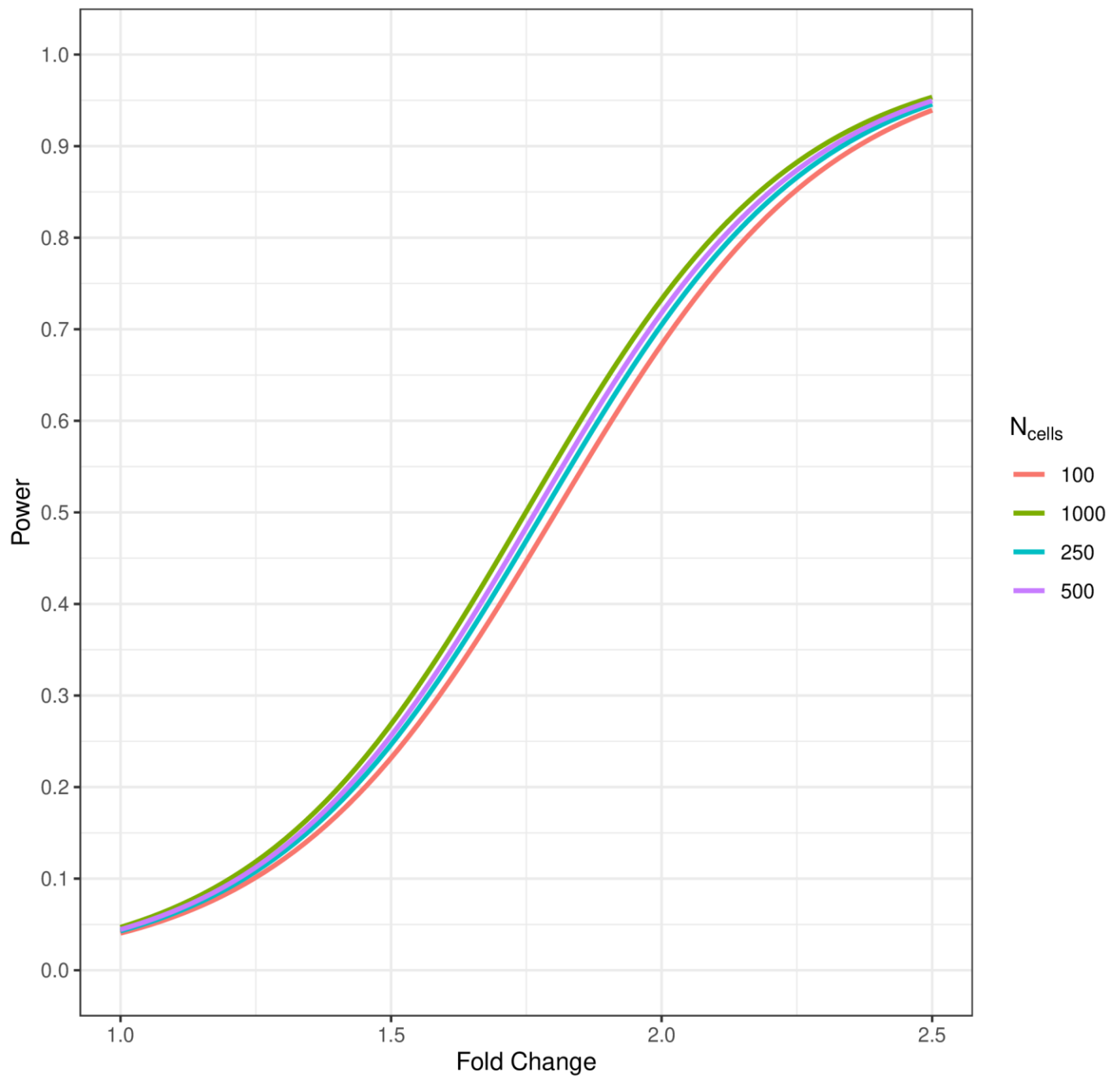
15 Individuals per Group



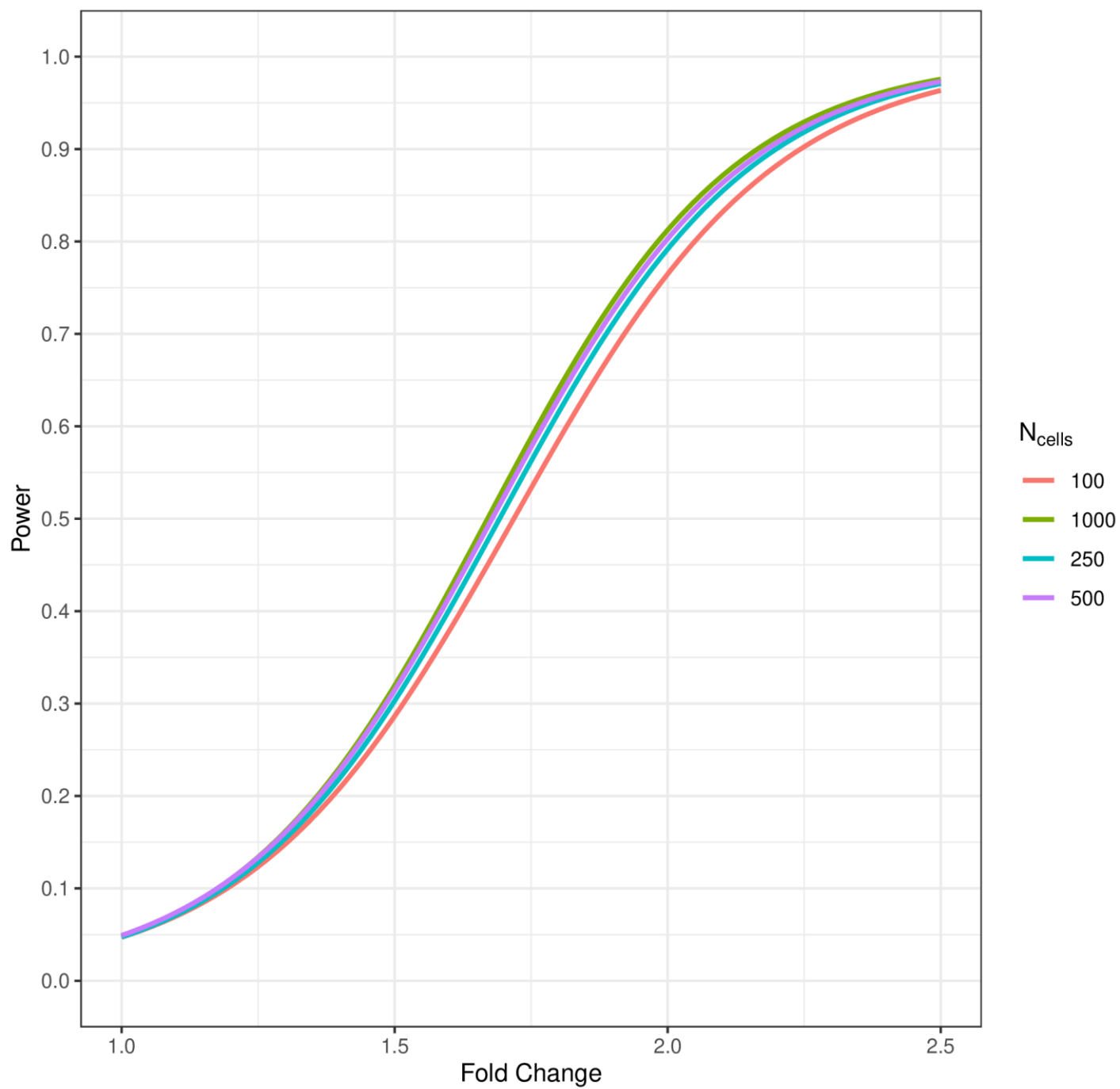
18 Individuals per Group



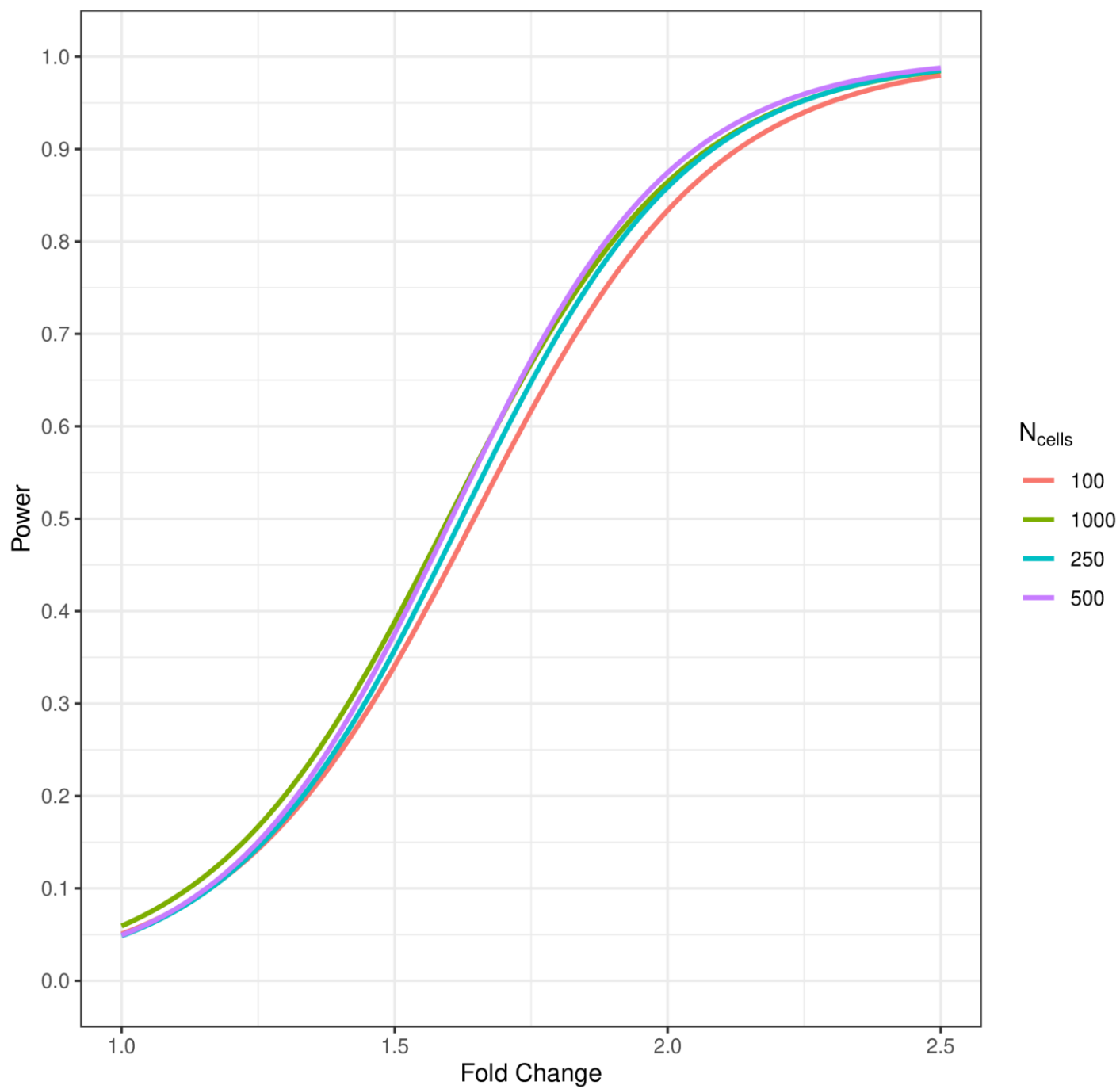
20 Individuals per Group



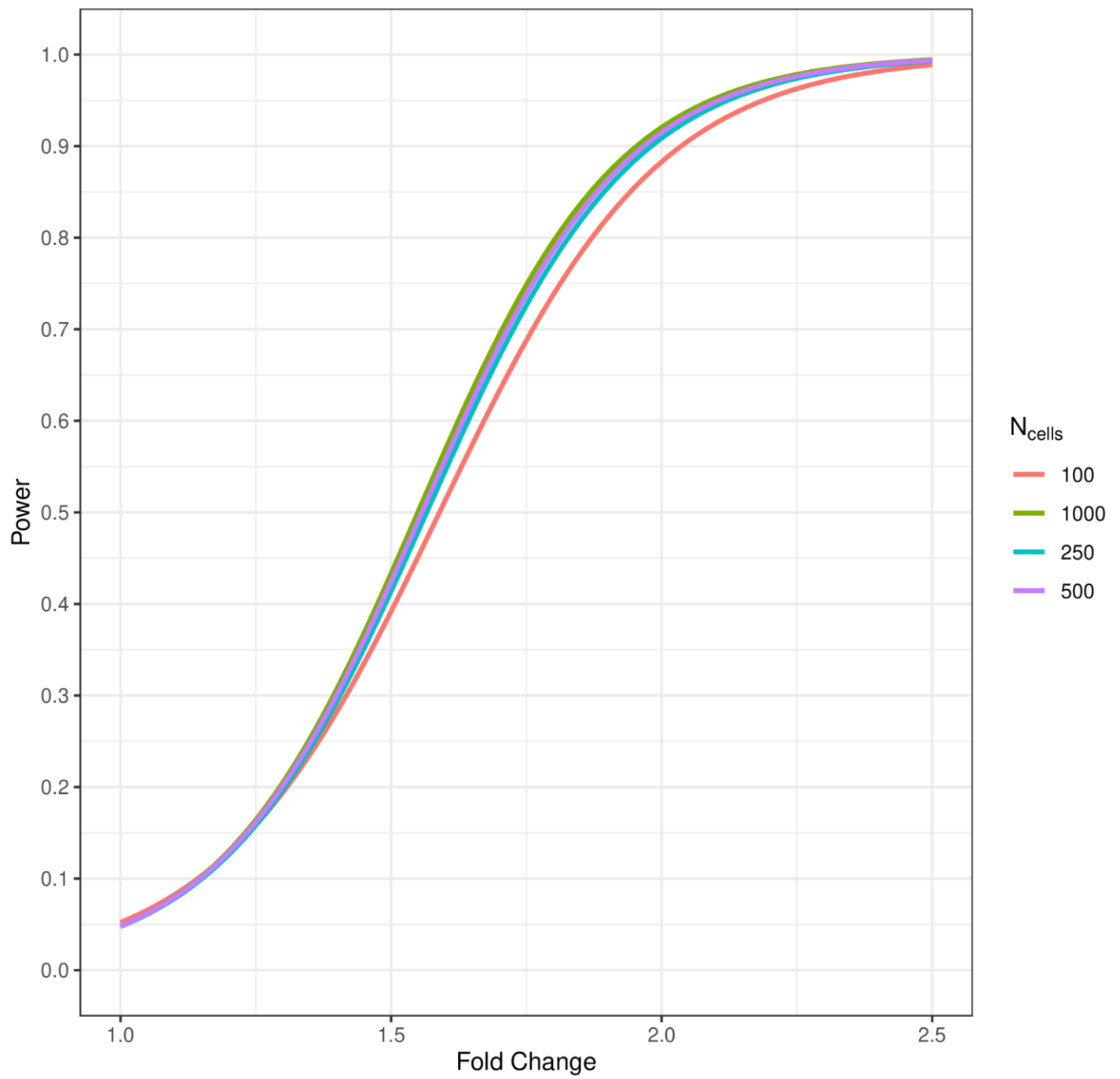
25 Individuals per Group



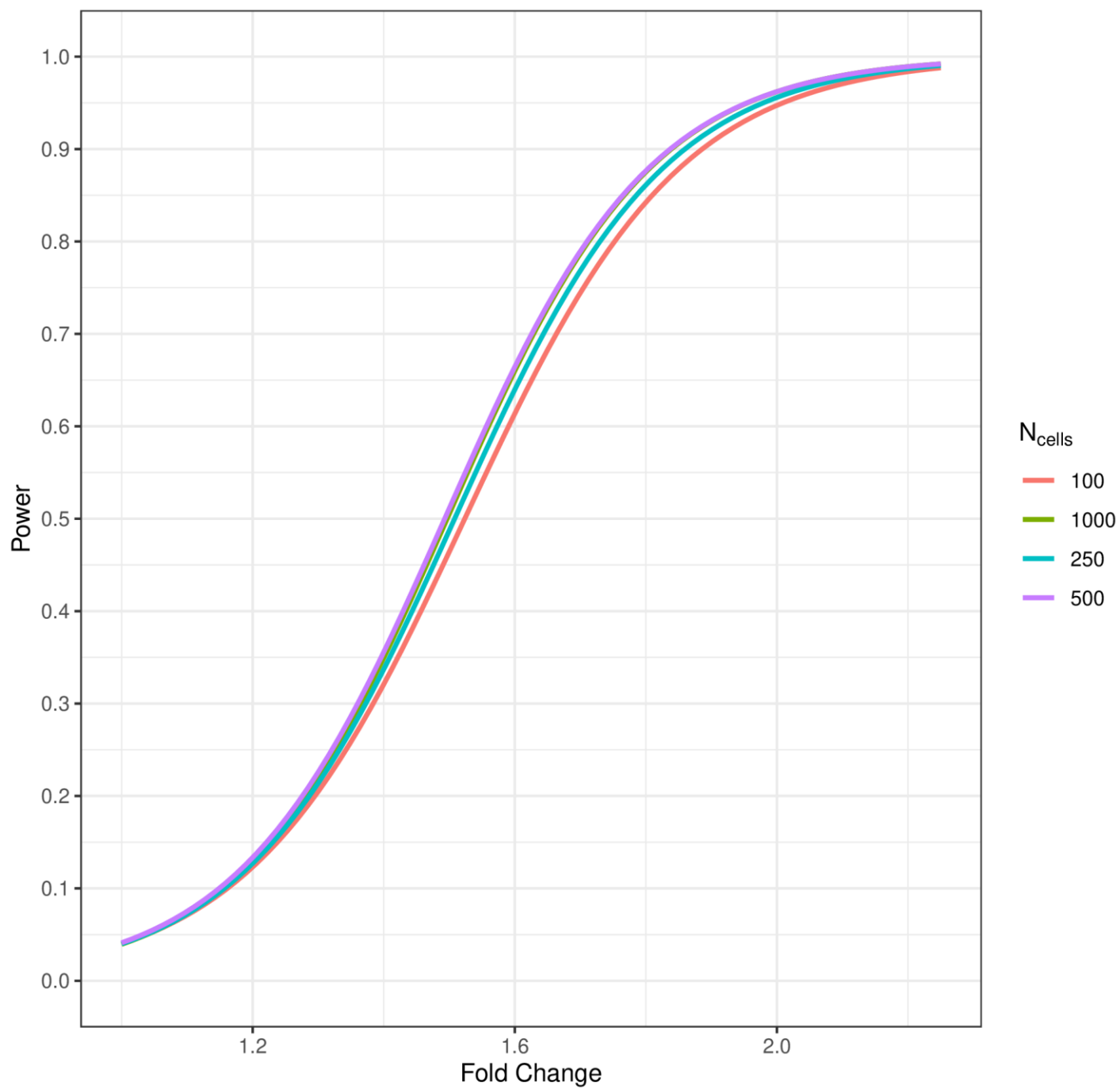
30 Individuals per Group



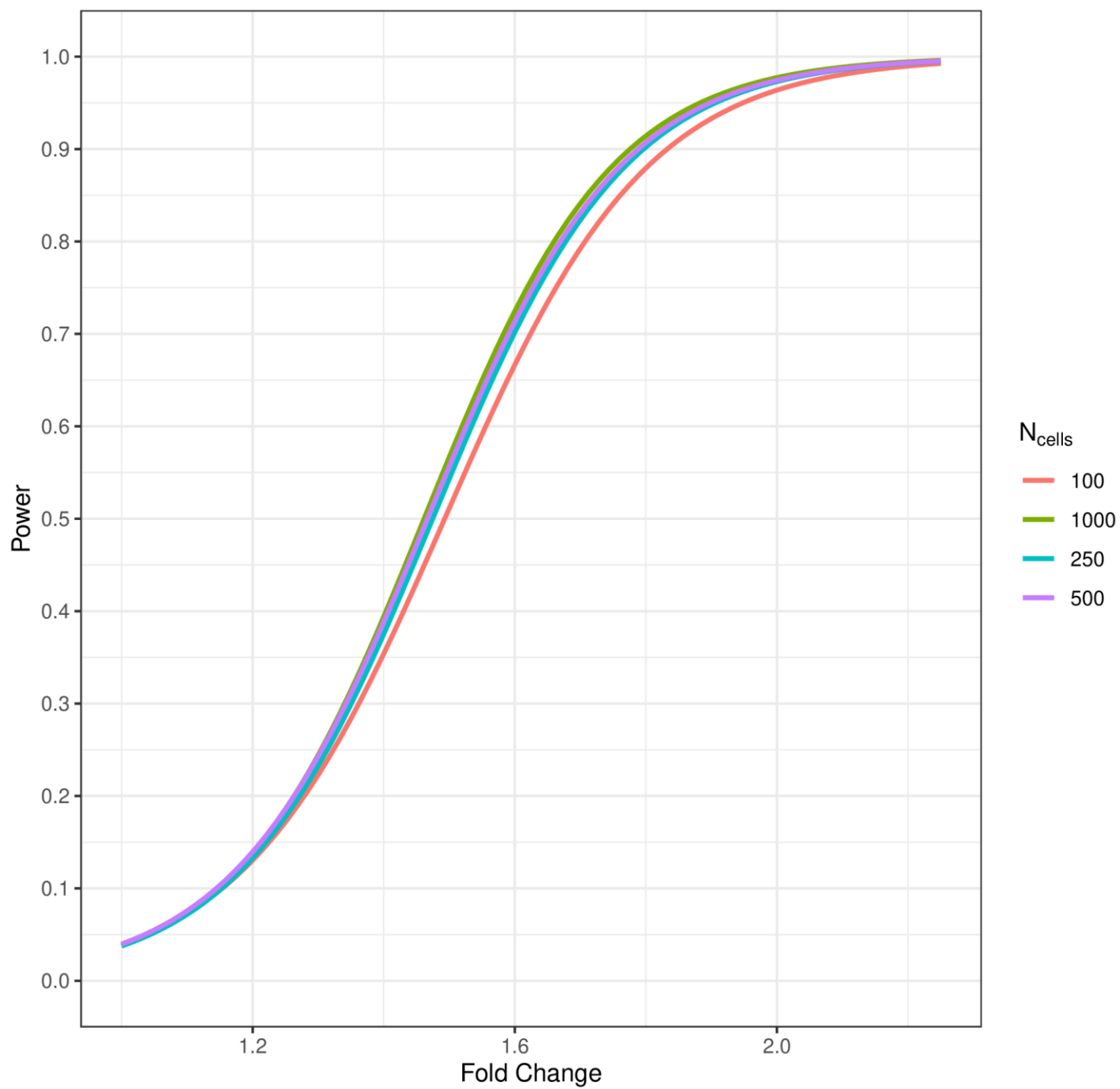
35 Individuals per Group



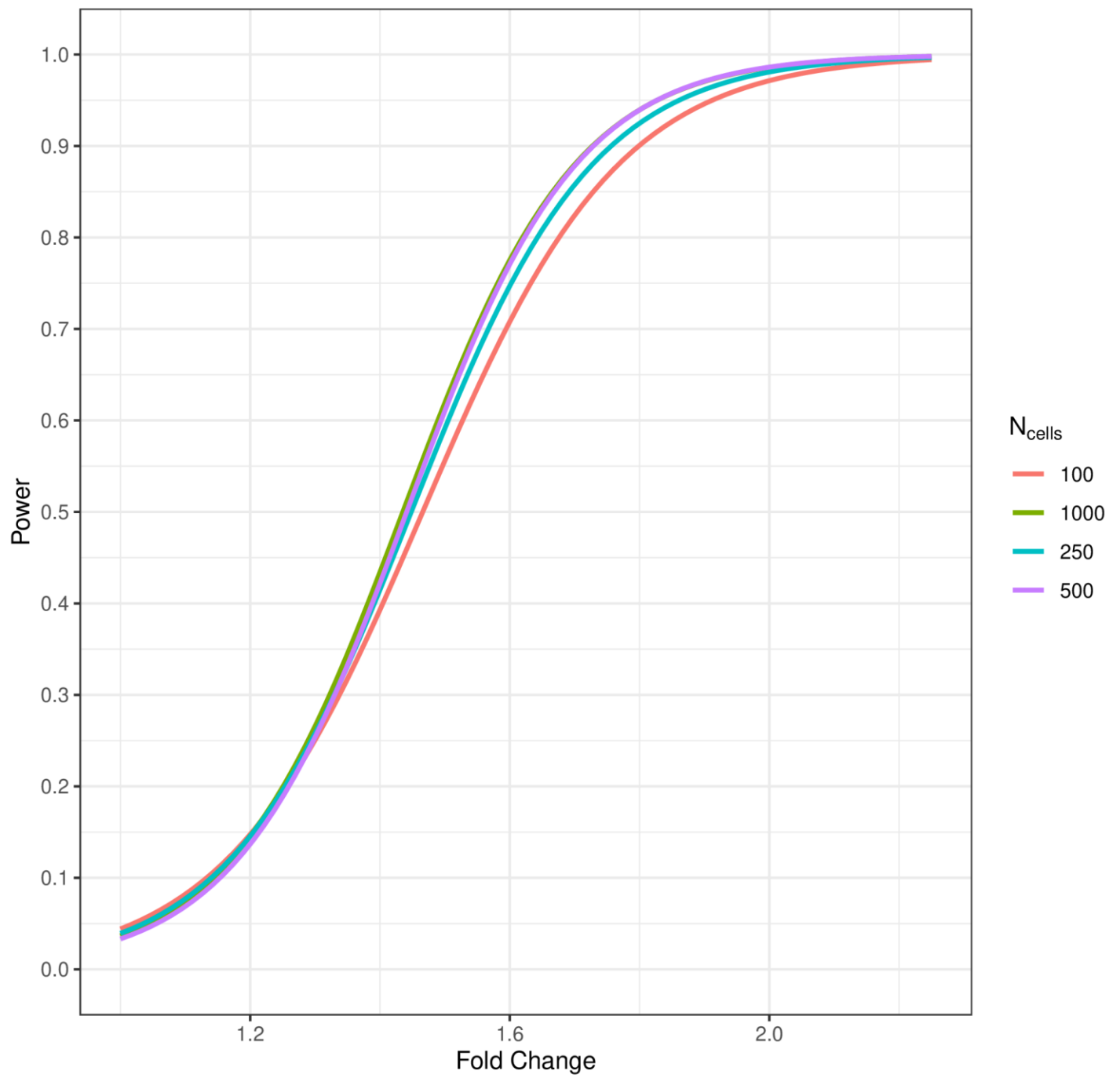
40 Individuals per Group



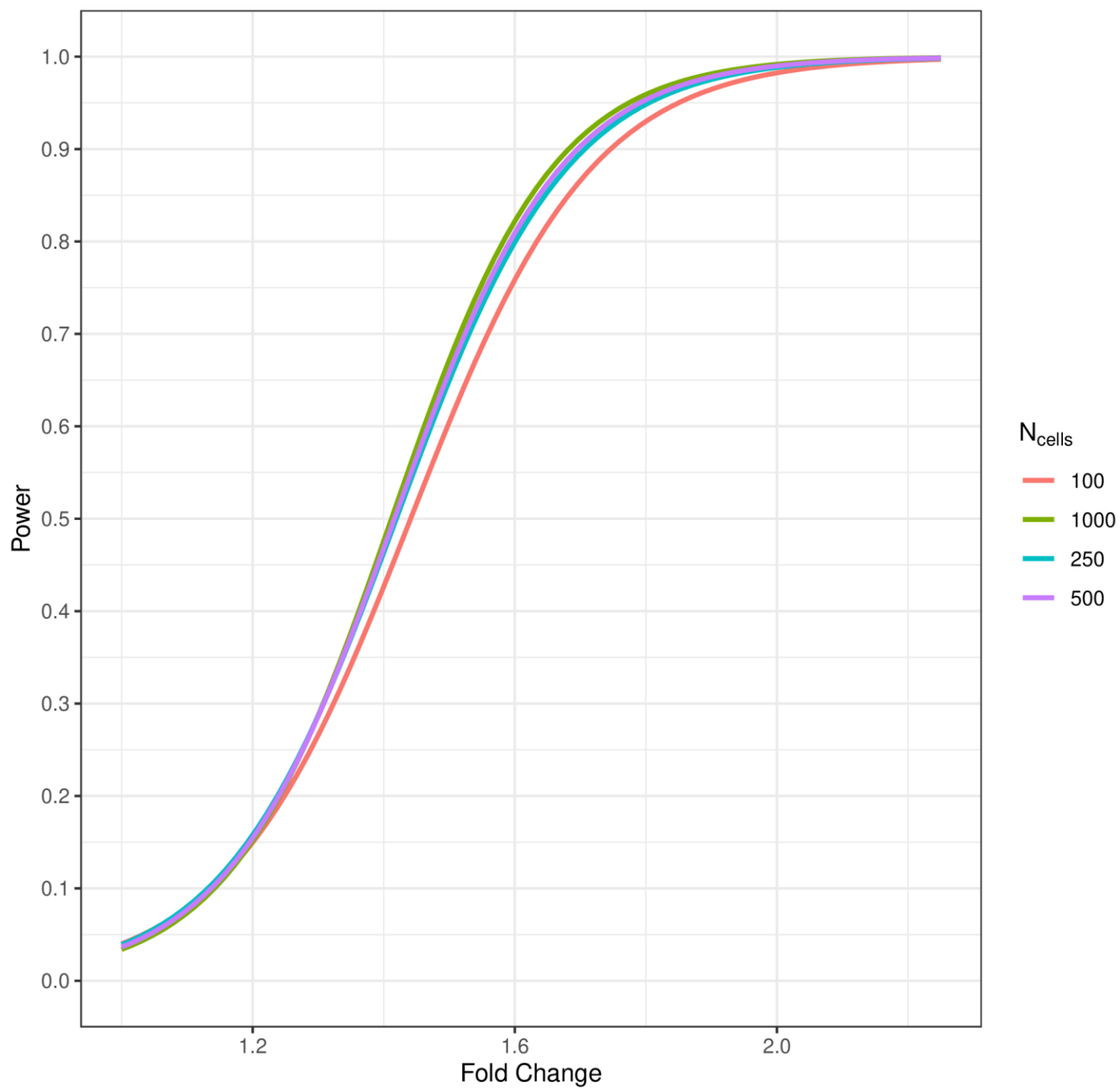
45 Individuals per Group



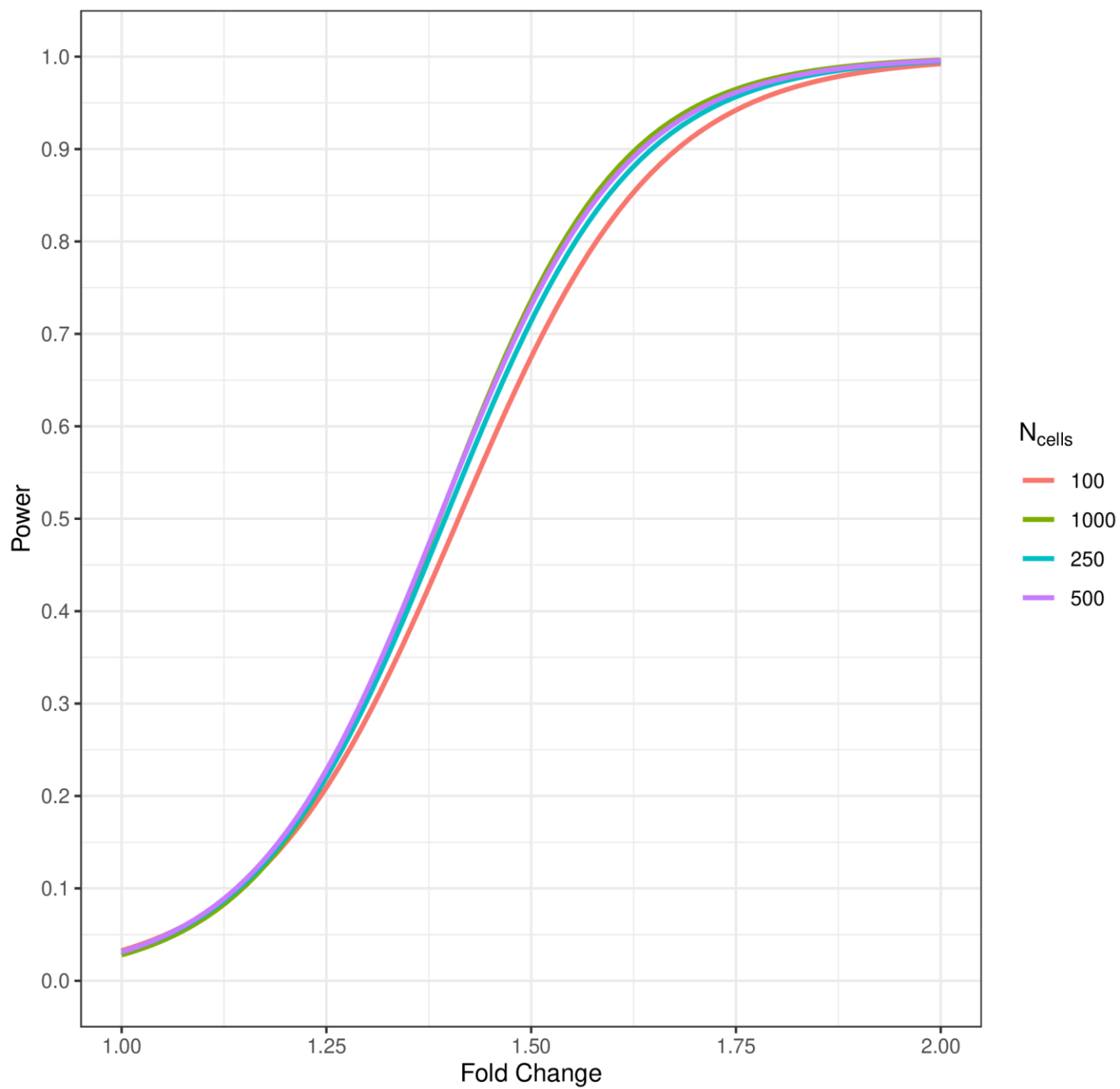
50 Individuals per Group



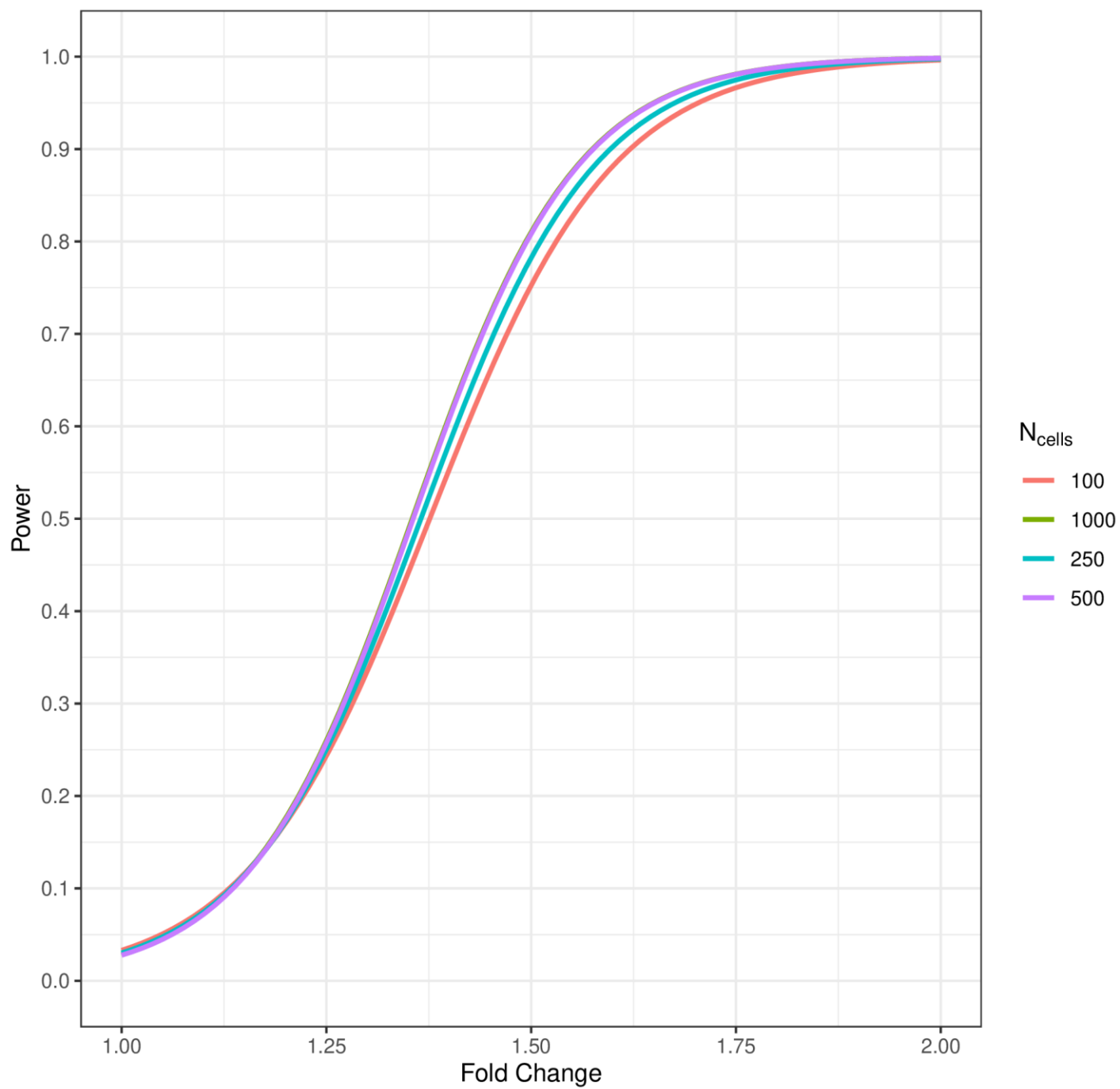
55 Individuals per Group



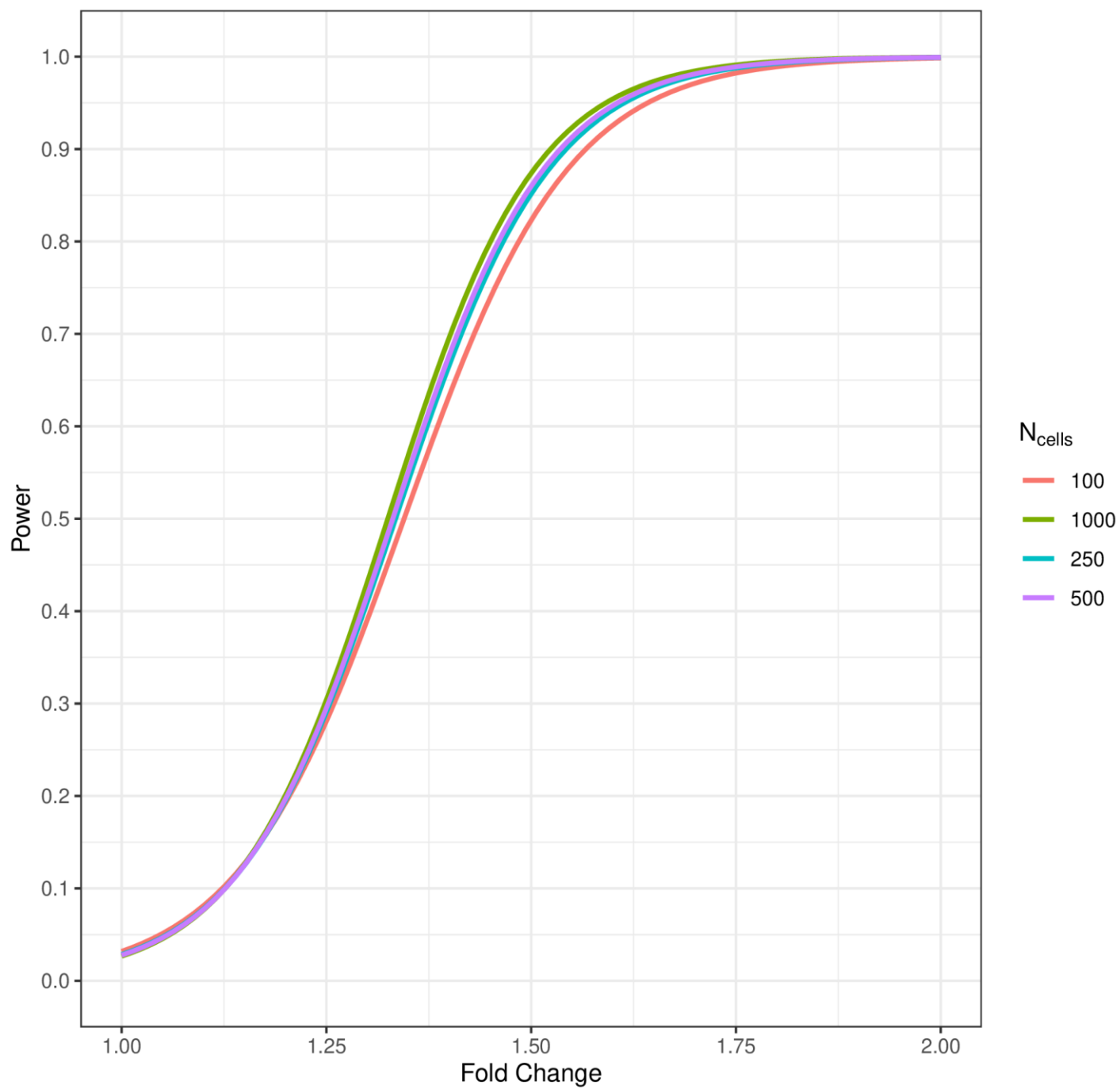
60 Individuals per Group



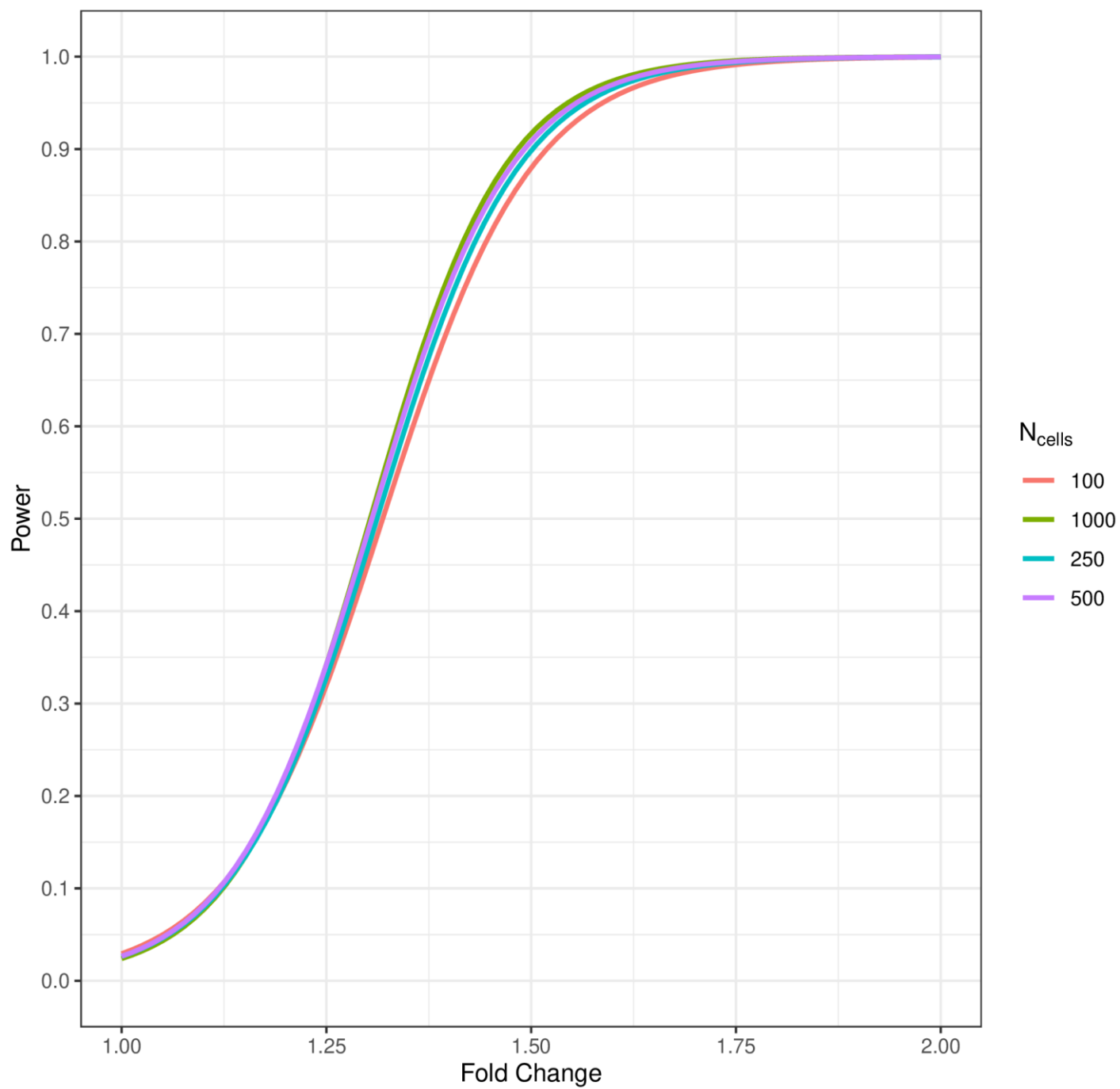
70 Individuals per Group



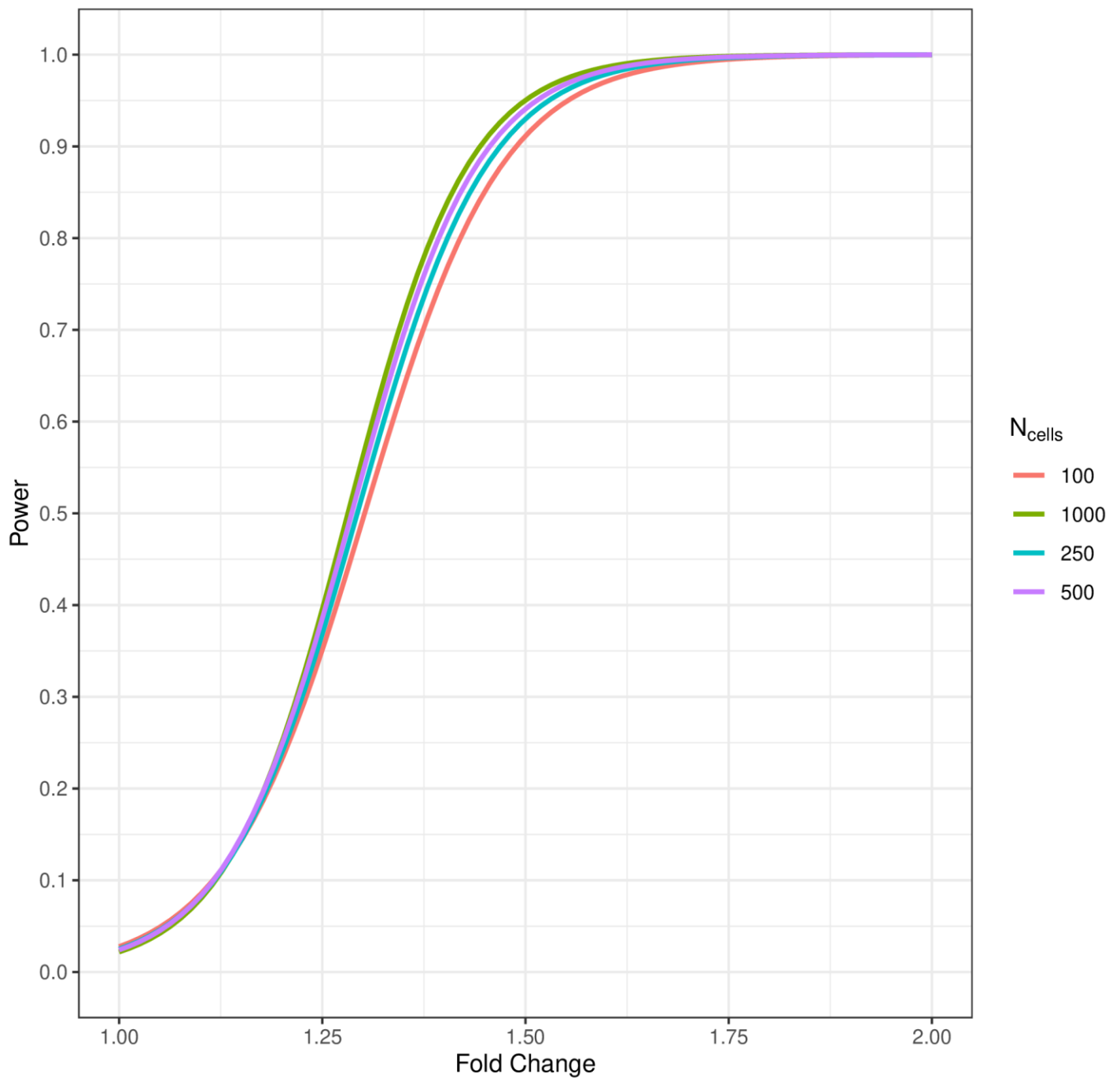
80 Individuals per Group



90 Individuals per Group

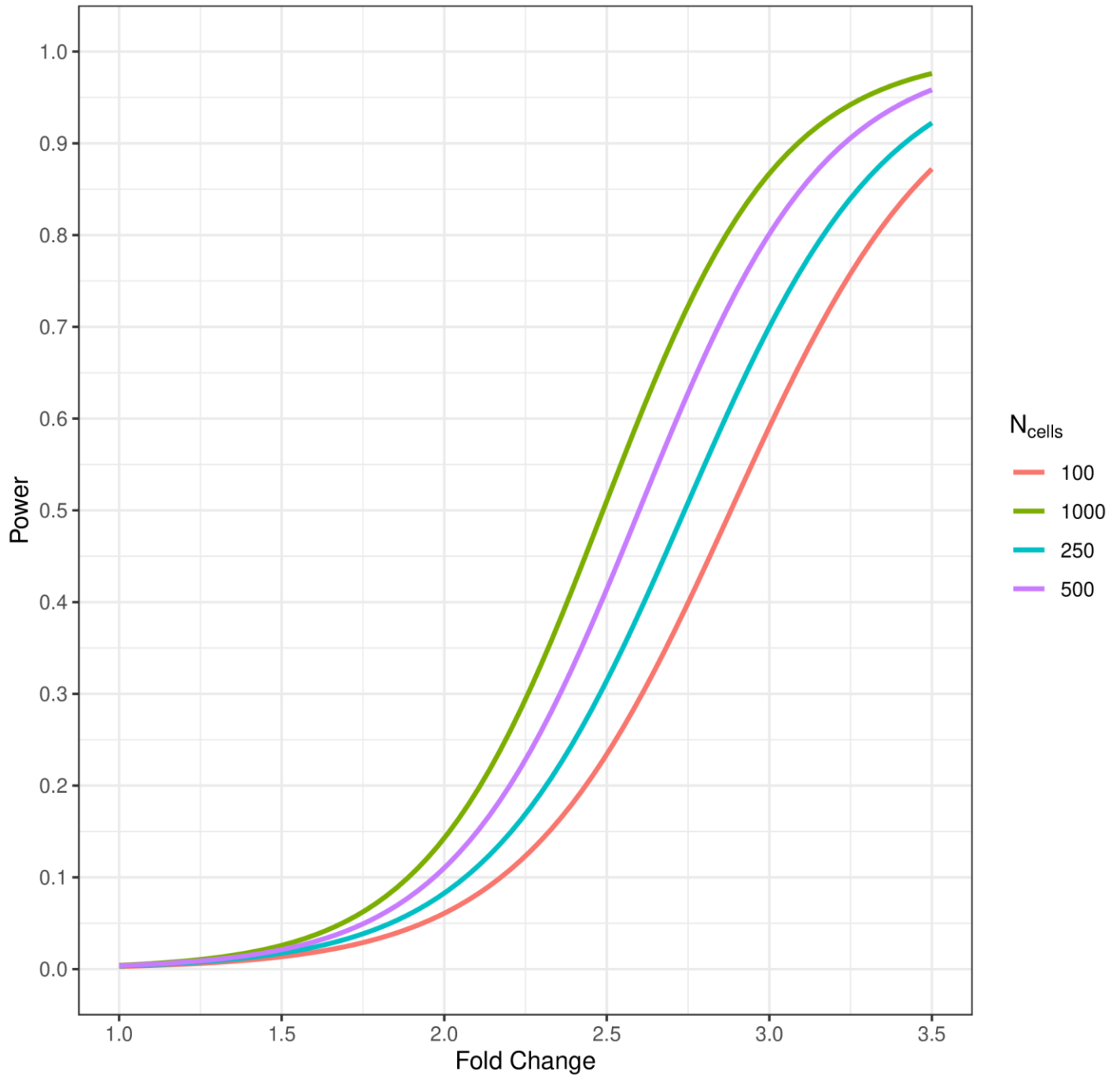


100 Individuals per Group

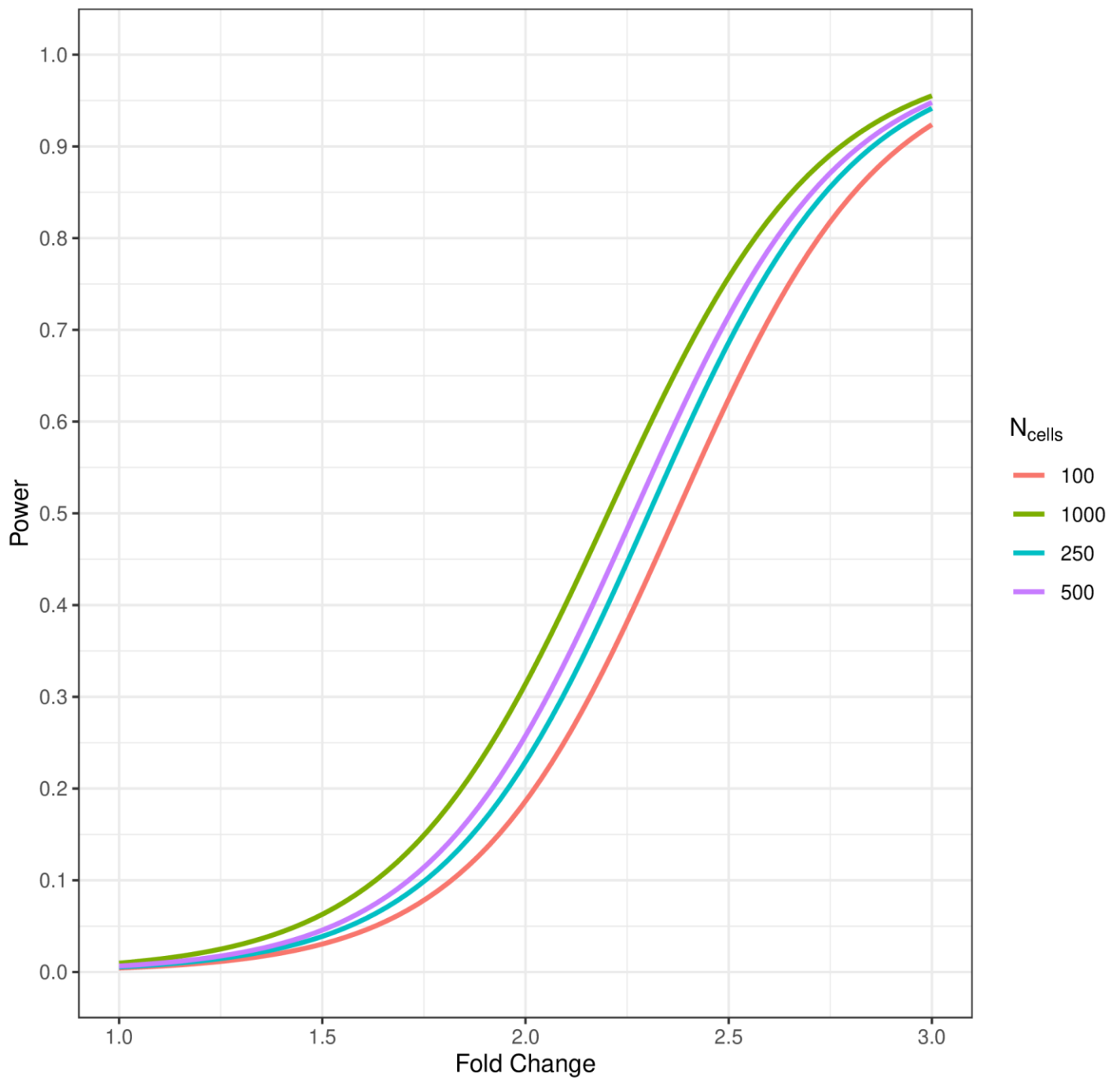


Supplementary Fig. 11: Power calculations using MAST with a random effect for the individual. Power curves for MAST using a random effect to account for intra-individual correlation. Curves are computed for 100 (red line), 250, Blue line) 500 (purple line), and 1,000 (red line) cells per individual with $\alpha = 0.001$. The numbers of individuals per group range from 5 to 100 and are listed above each plot.

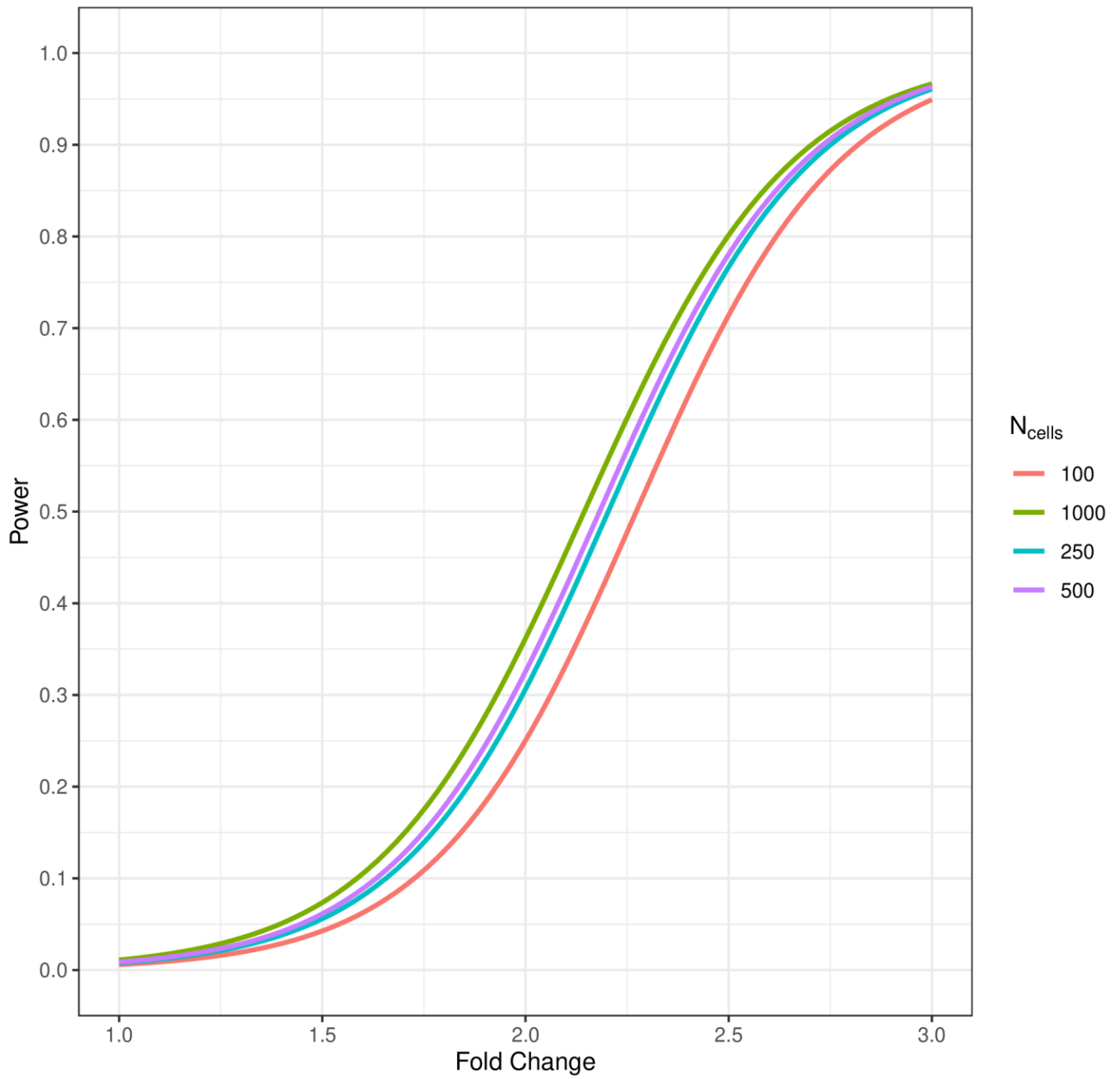
5 Individuals per Group



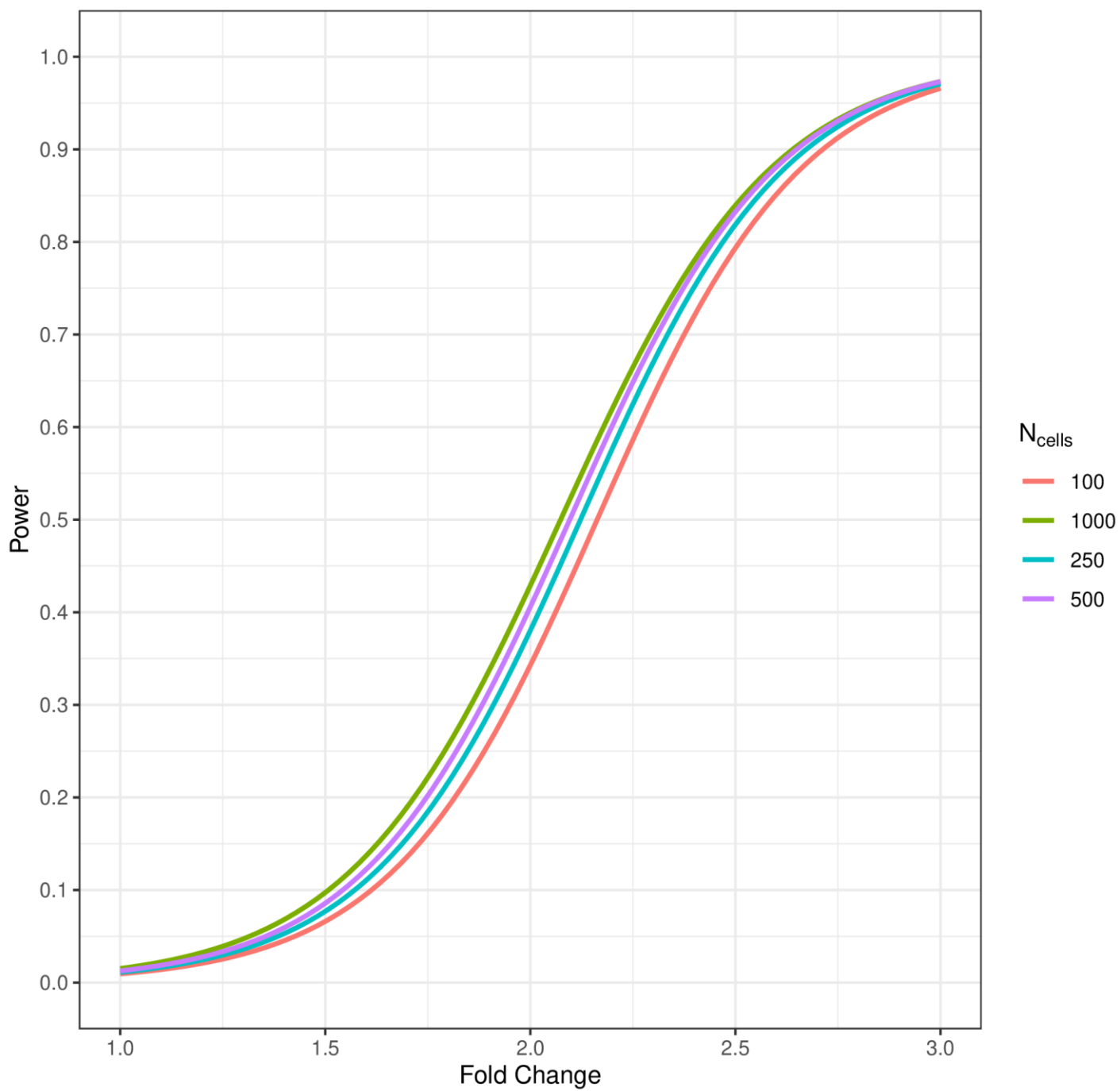
10 Individuals per Group



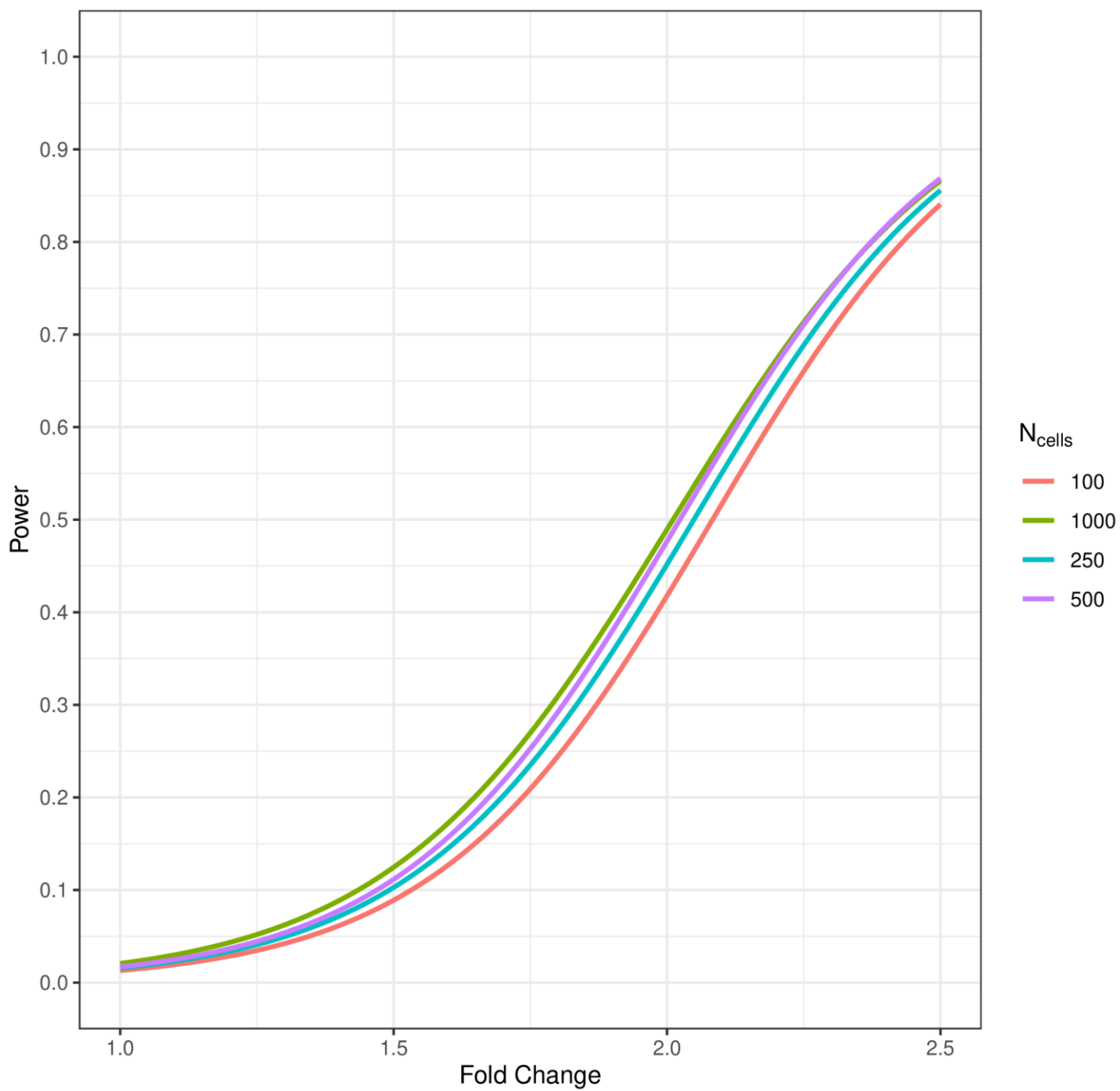
12 Individuals per Group



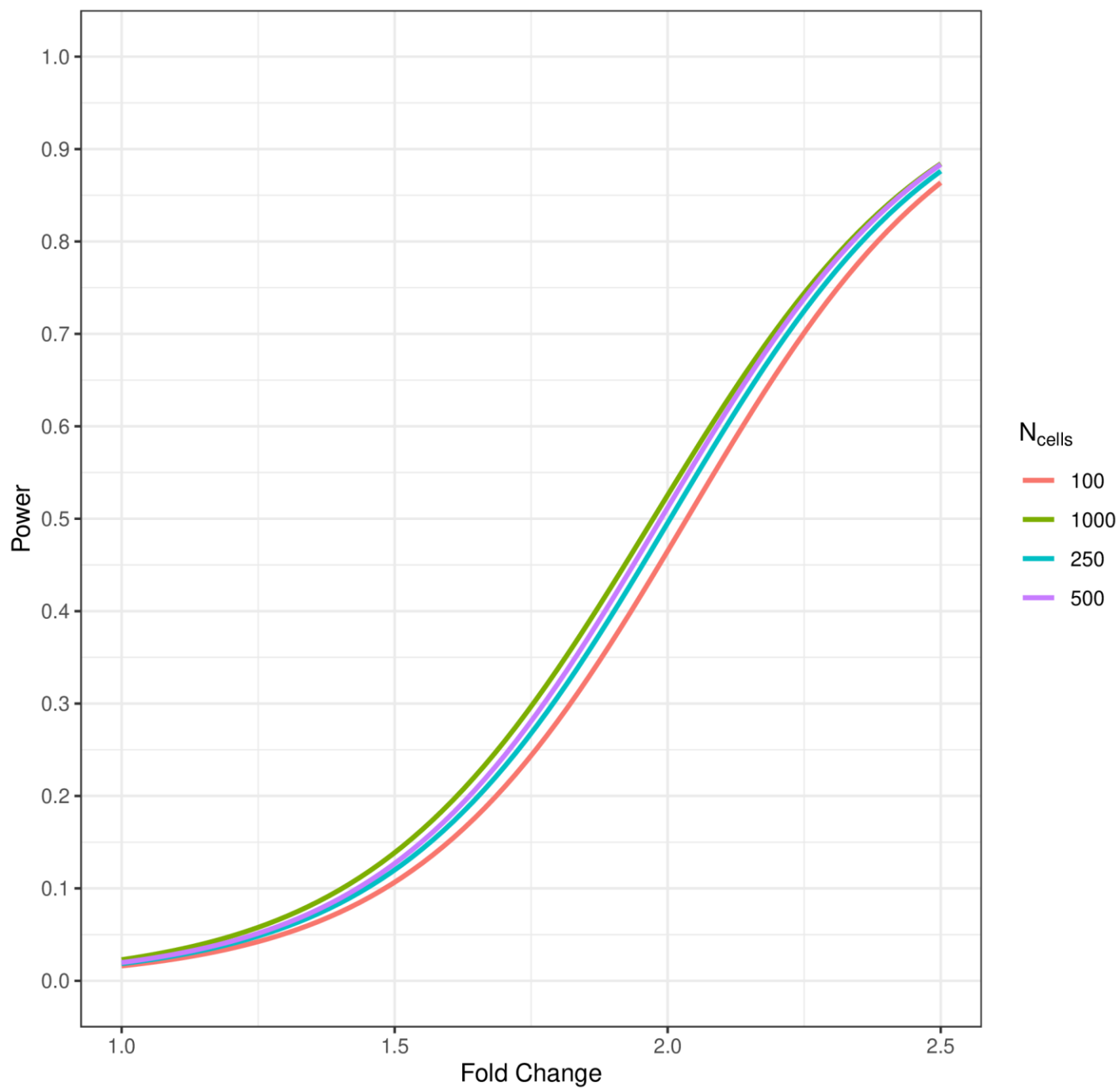
15 Individuals per Group



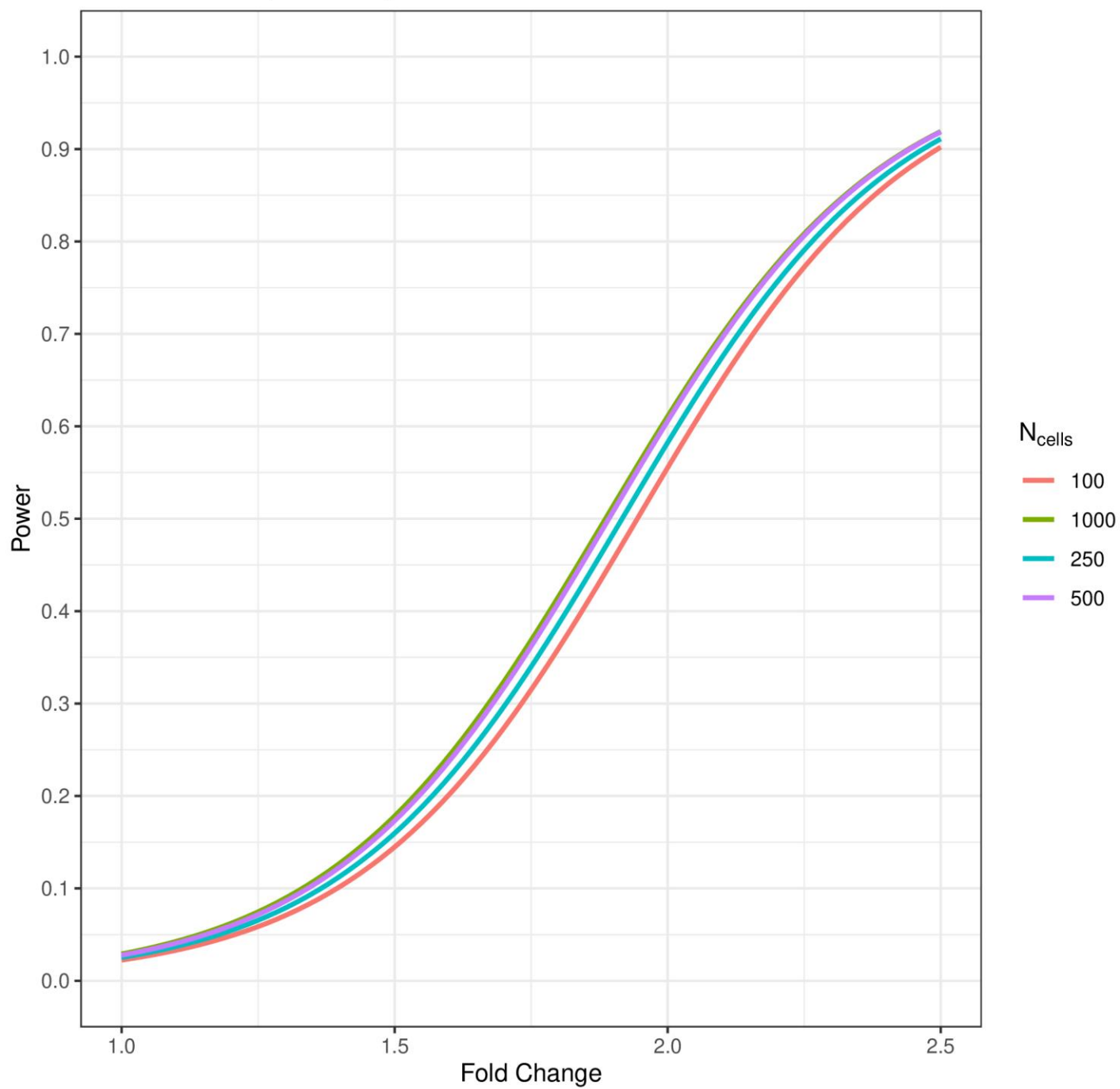
18 Individuals per Group



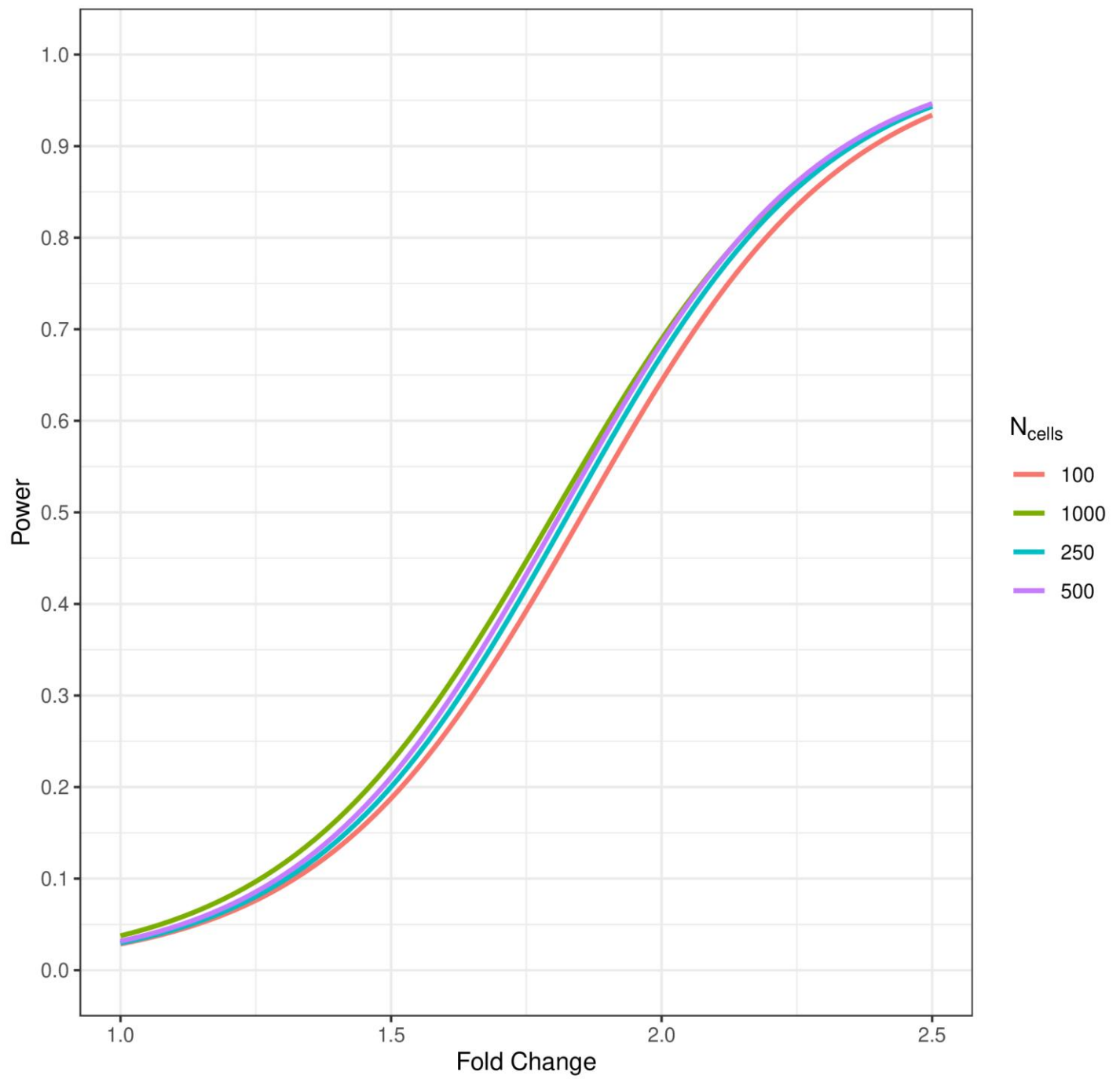
20 Individuals per Group



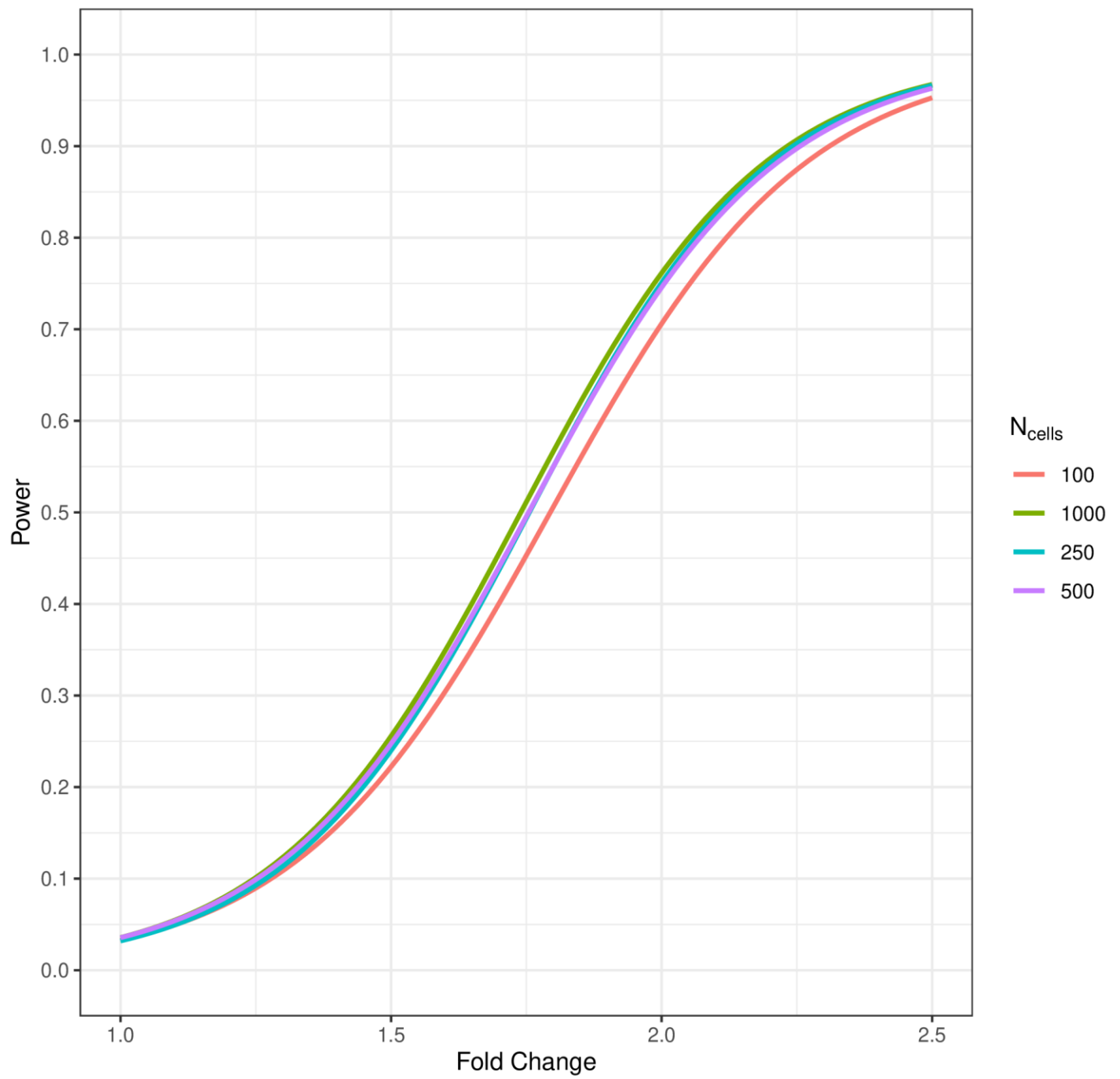
25 Individuals per Group



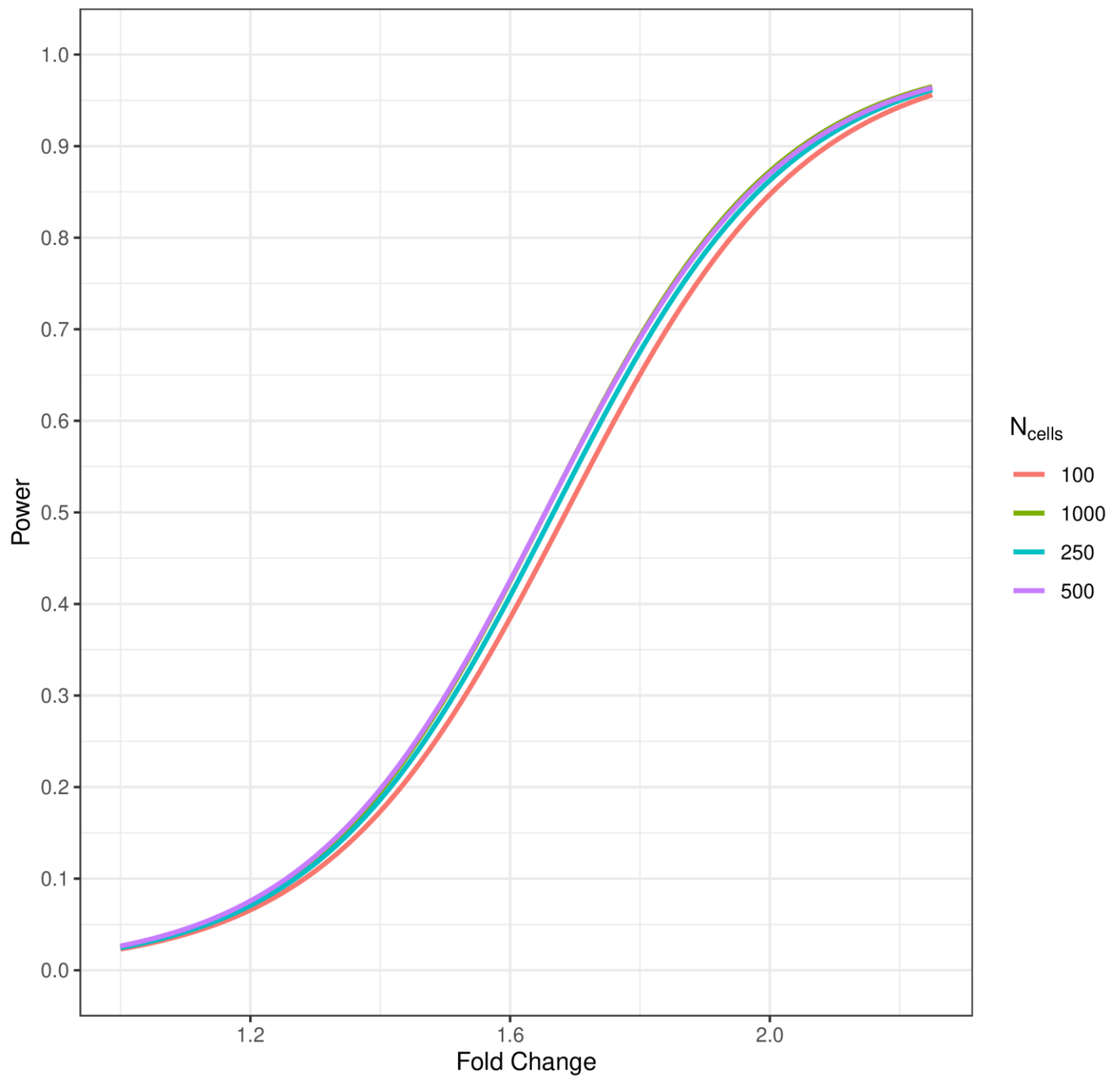
30 Individuals per Group



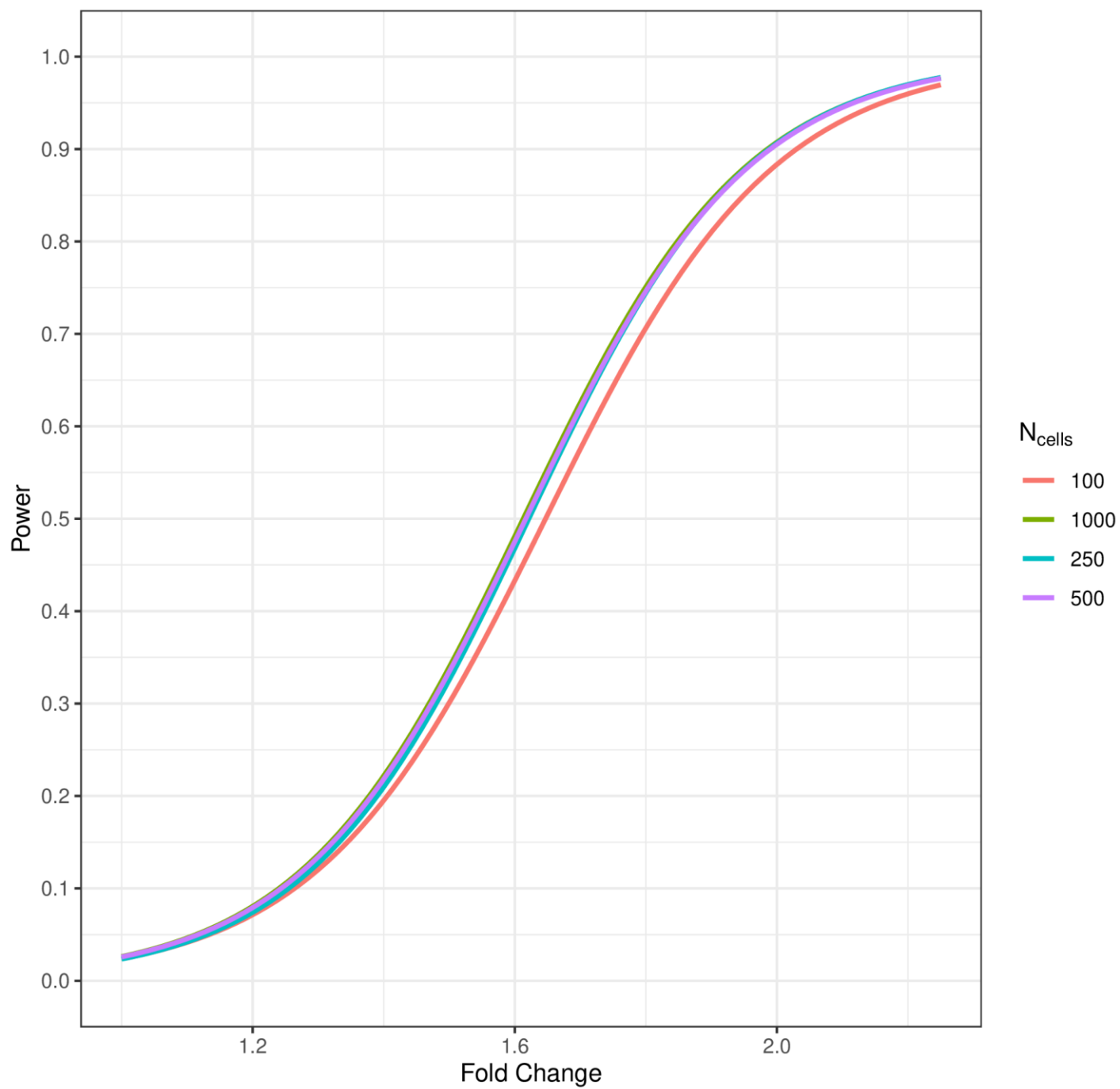
35 Individuals per Group



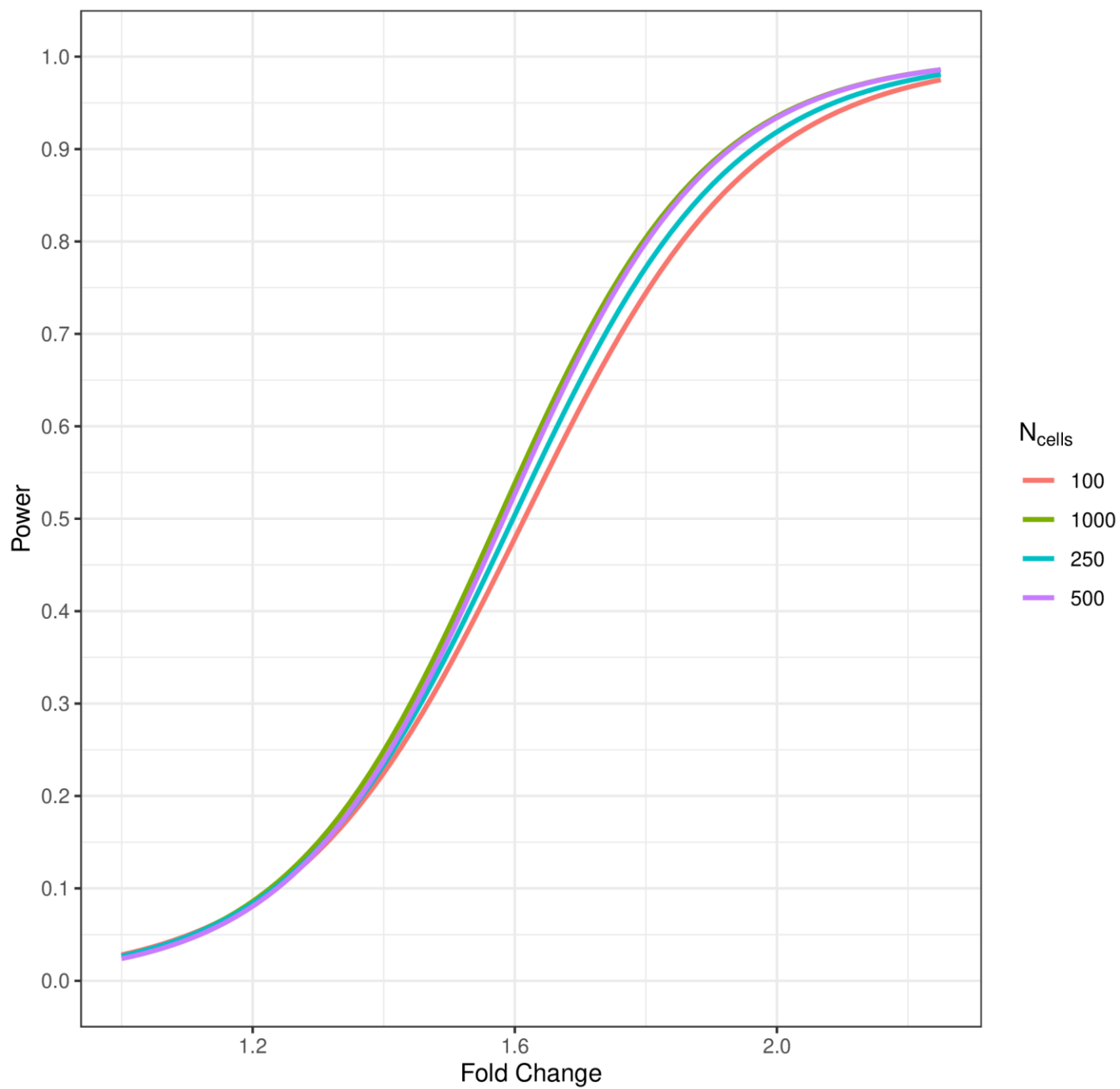
40 Individuals per Group



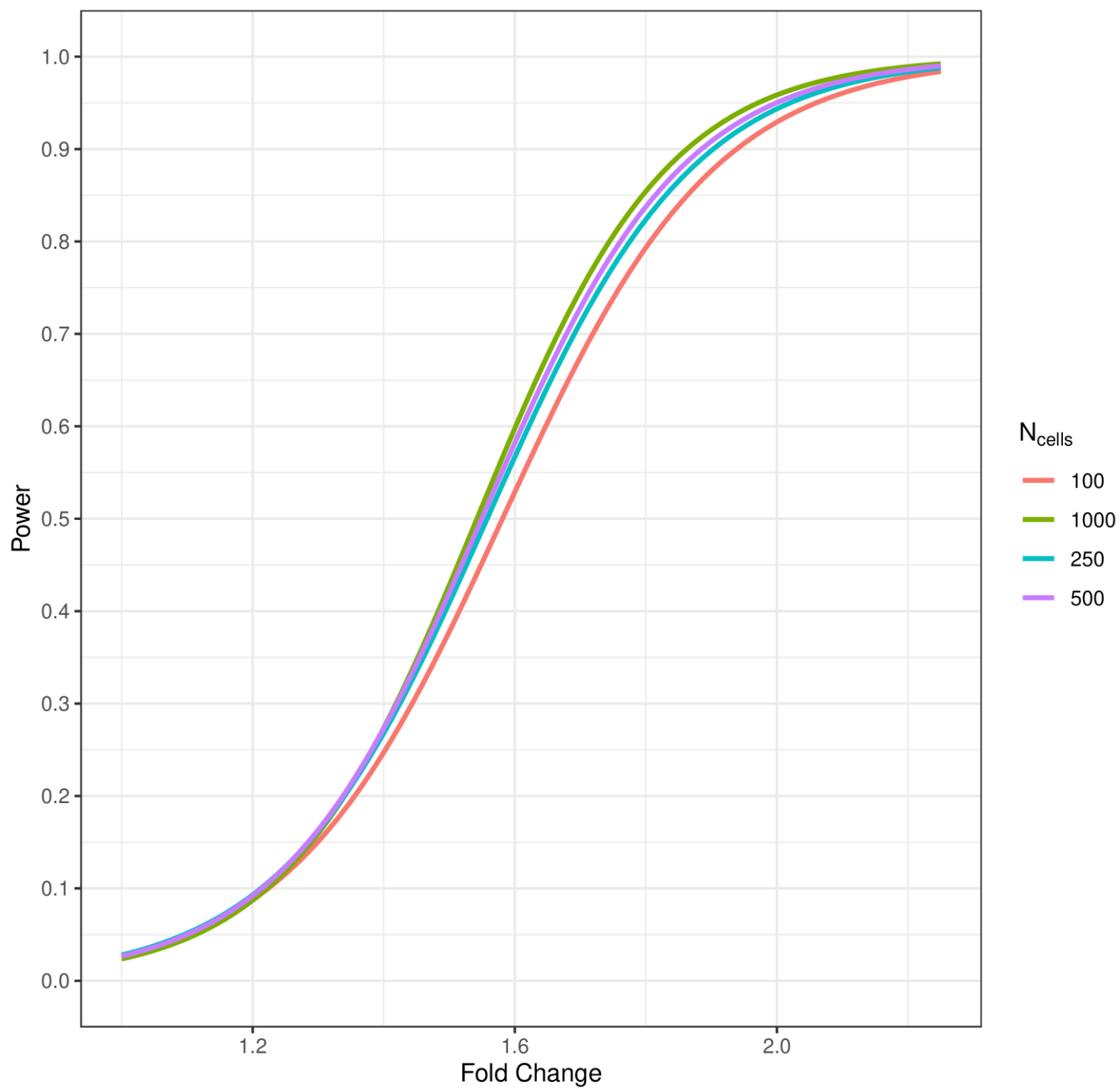
45 Individuals per Group



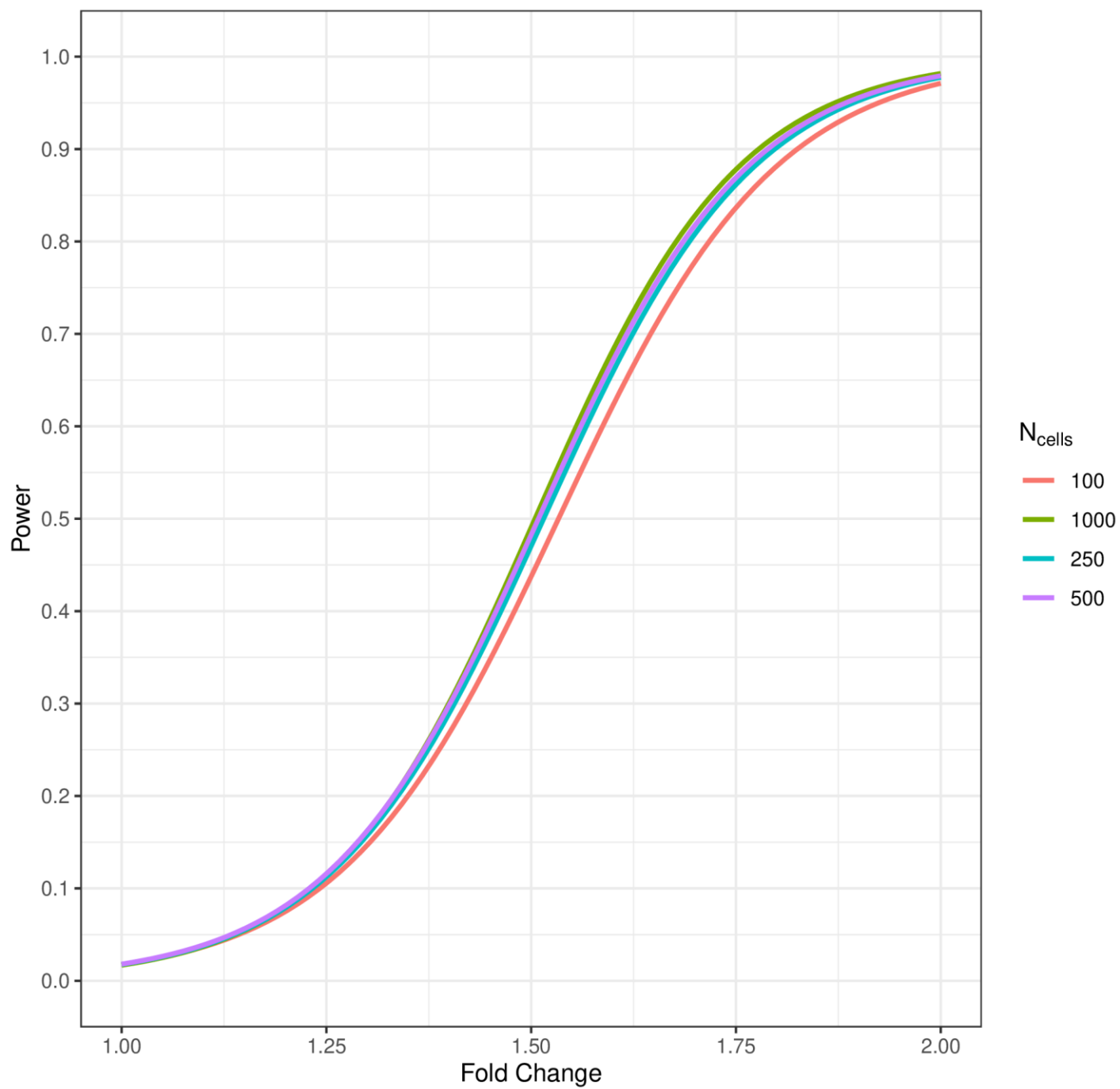
50 Individuals per Group



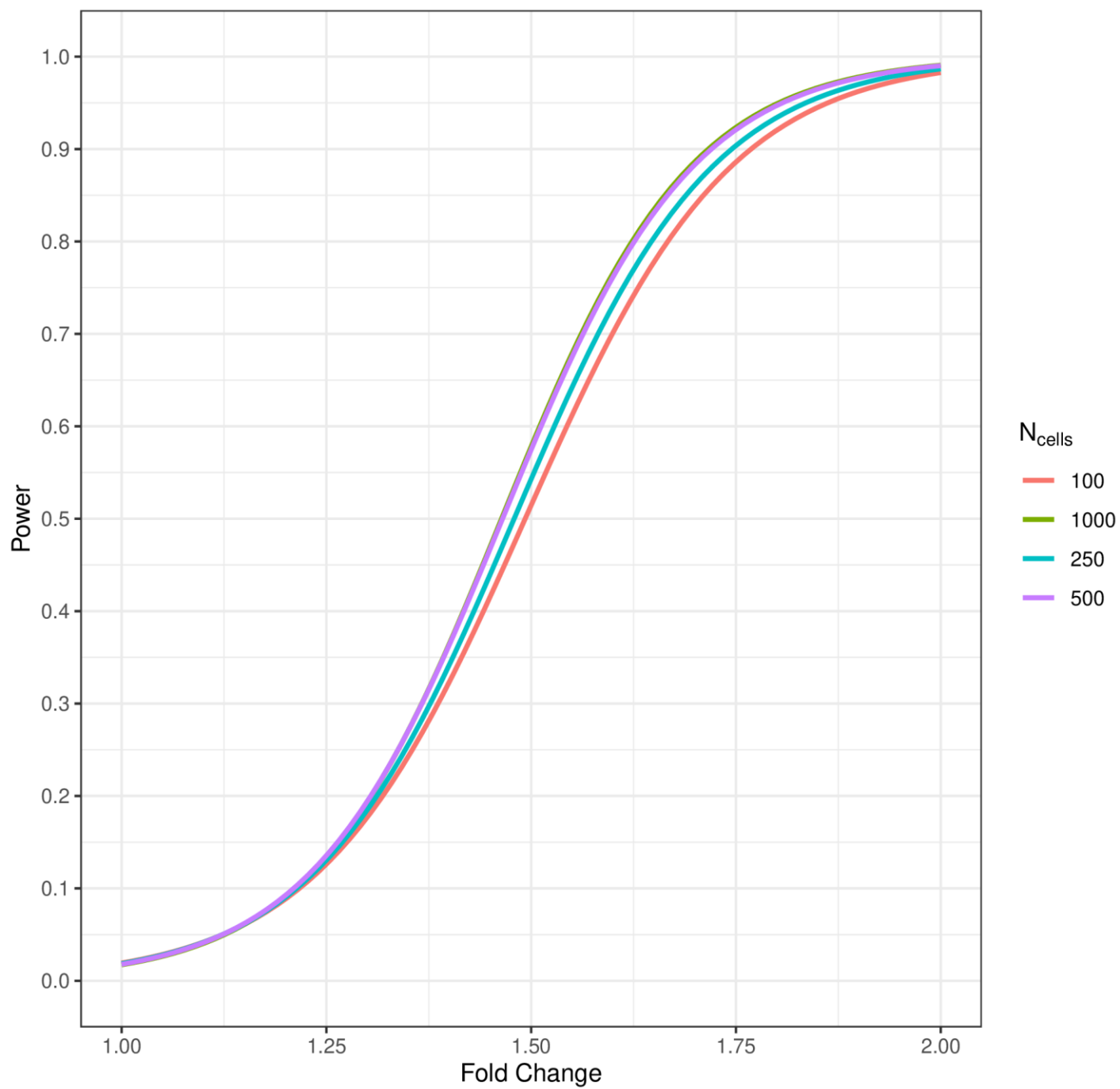
55 Individuals per Group



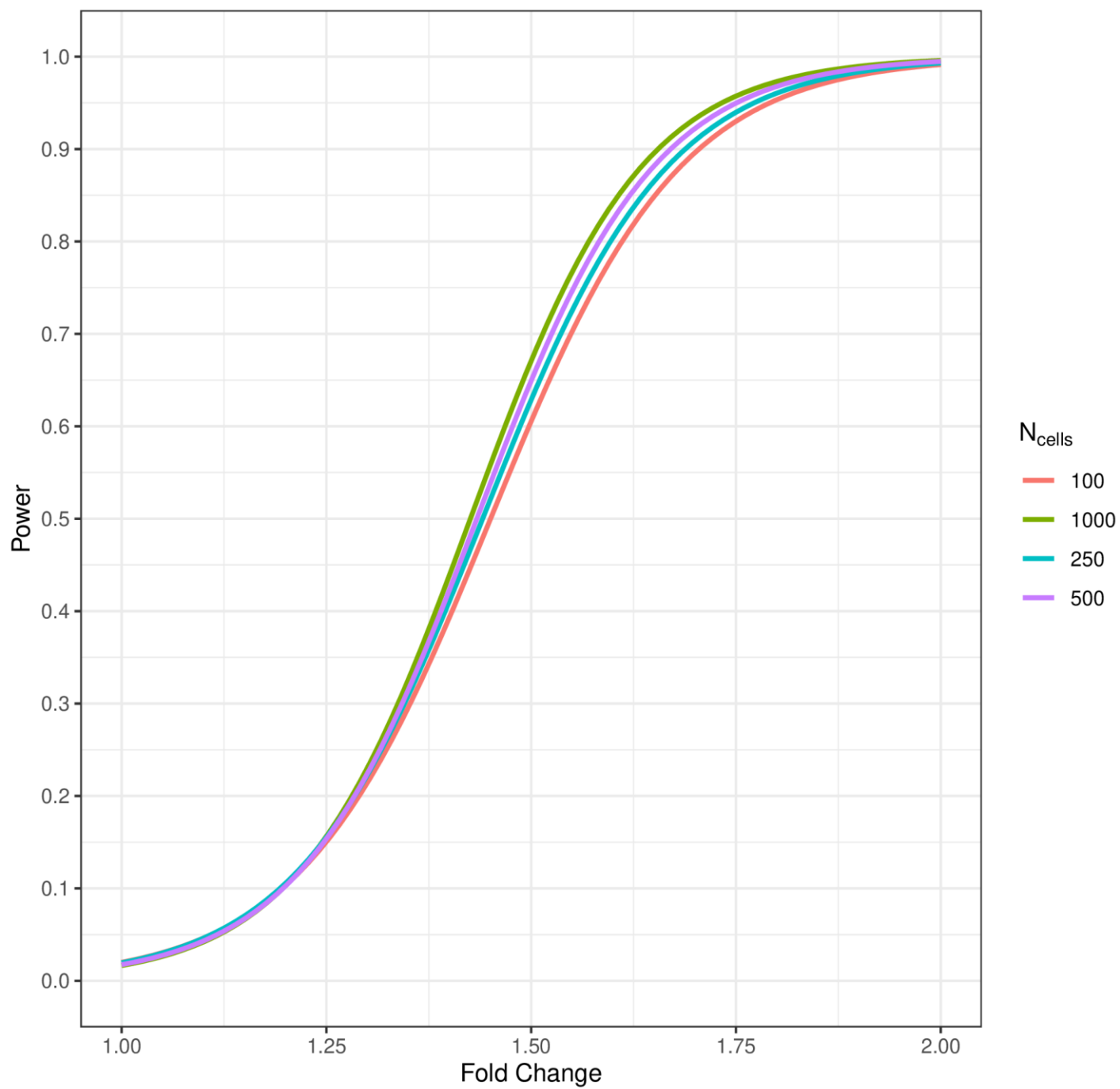
60 Individuals per Group



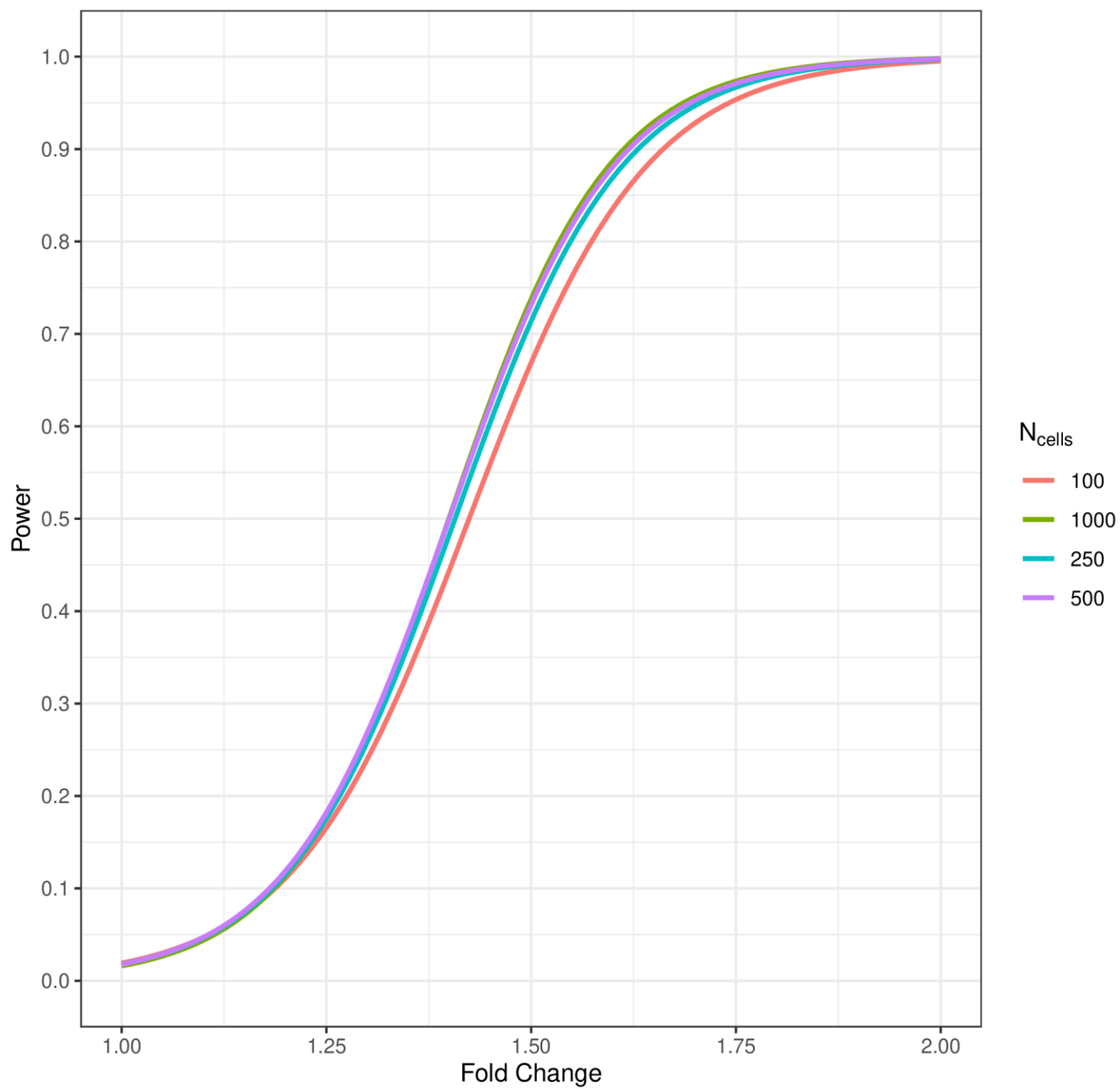
70 Individuals per Group



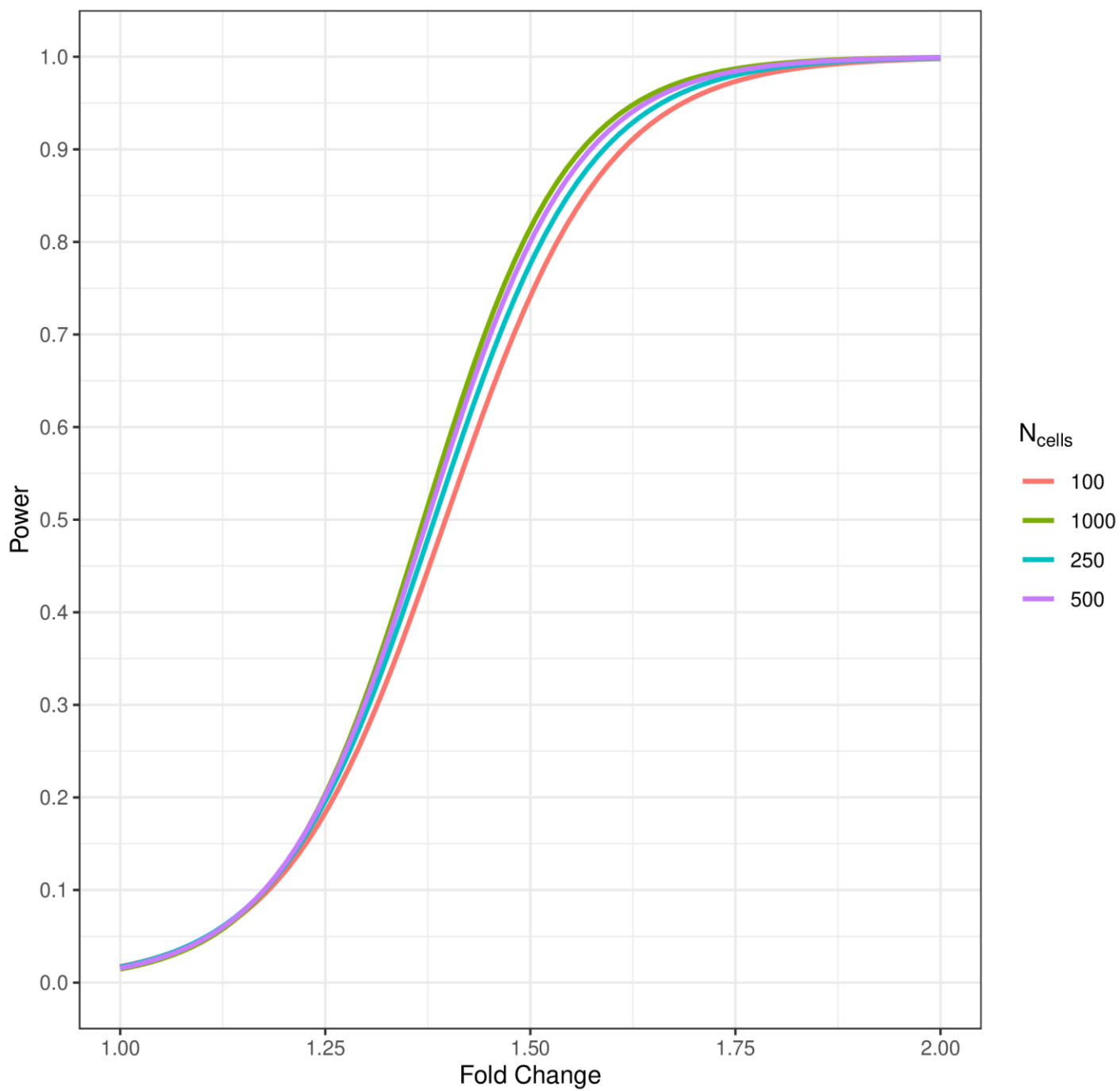
80 Individuals per Group



90 Individuals per Group



100 Individuals per Group



Supplementary Table 1: Type I error rates and their associated confidence intervals (95%) for some currently applied tools in single-cell analyses. Type I error rates of 10 different methods under 20 different conditions and a significance threshold of $p < 0.05$. Hypothesis tests were two-sided with no correction for multiple comparisons – as each gene was simulated and analyzed independently. 250,000 iterations were computed to obtain an error rate for each method.

N _{ind}	N _{cells}	Two-part Hurdle			Tweedie		GEE1	Pseudobulk		Tobit	Modified <i>t</i>
		Default	Corrected	RE	GLMM	GLM		Mean	Sum		
5	50	0.561 (0.559-0.563)	0.637 (0.635-0.639)	0.069 (0.068-0.07)	0.082 (0.081-0.083)	0.34 (0.338-0.342)	0.114 (0.113-0.116)	0.023 (0.023-0.024)	0.035 (0.034-0.035)	0.353 (0.351-0.355)	0.4 (0.398-0.402)
	100	0.677 (0.676-0.679)	0.719 (0.717-0.72)	0.064 (0.063-0.065)	0.084 (0.083-0.086)	0.463 (0.461-0.465)	0.11 (0.109-0.111)	0.022 (0.021-0.022)	0.032 (0.031-0.033)	0.471 (0.469-0.473)	0.51 (0.508-0.511)
	250	0.798 (0.796-0.799)	0.778 (0.776-0.78)	0.066 (0.065-0.067)	0.083 (0.082-0.084)	0.609 (0.607-0.611)	0.103 (0.102-0.104)	0.023 (0.023-0.024)	0.028 (0.027-0.029)	0.628 (0.626-0.63)	0.644 (0.642-0.646)
	500	0.862 (0.861-0.863)	0.803 (0.801-0.805)	0.065 (0.064-0.066)	0.081 (0.08-0.082)	0.705 (0.704-0.707)	0.104 (0.103-0.105)	0.023 (0.022-0.024)	0.026 (0.026-0.027)	0.725 (0.723-0.727)	0.718 (0.716-0.72)
	500	0.862 (0.861-0.863)	0.803 (0.801-0.805)	0.065 (0.064-0.066)	0.081 (0.08-0.082)	0.705 (0.704-0.707)	0.104 (0.103-0.105)	0.023 (0.022-0.024)	0.026 (0.026-0.027)	0.725 (0.723-0.727)	0.718 (0.716-0.72)
10	50	0.563 (0.561-0.565)	0.611 (0.609-0.613)	0.055 (0.054-0.056)	0.064 (0.063-0.065)	0.35 (0.348-0.352)	0.076 (0.075-0.077)	0.024 (0.023-0.024)	0.021 (0.021-0.022)	0.345 (0.343-0.347)	0.397 (0.395-0.398)
	100	0.689 (0.687-0.69)	0.718 (0.716-0.719)	0.053 (0.052-0.054)	0.065 (0.064-0.066)	0.462 (0.46-0.464)	0.077 (0.076-0.078)	0.024 (0.024-0.025)	0.02 (0.019-0.02)	0.47 (0.468-0.472)	0.502 (0.5-0.504)
	250	0.81 (0.808-0.811)	0.793 (0.792-0.795)	0.049 (0.048-0.05)	0.064 (0.063-0.064)	0.61 (0.609-0.612)	0.074 (0.073-0.075)	0.022 (0.022-0.023)	0.019 (0.018-0.019)	0.624 (0.622-0.626)	0.635 (0.633-0.637)
	500	0.875 (0.874-0.876)	0.827 (0.826-0.829)	0.049 (0.048-0.05)	0.061 (0.06-0.062)	0.705 (0.703-0.707)	0.073 (0.072-0.074)	0.021 (0.02-0.022)	0.018 (0.017-0.018)	0.722 (0.72-0.724)	0.717 (0.715-0.719)
	500	0.875 (0.874-0.876)	0.827 (0.826-0.829)	0.049 (0.048-0.05)	0.061 (0.06-0.062)	0.705 (0.703-0.707)	0.073 (0.072-0.074)	0.021 (0.02-0.022)	0.018 (0.017-0.018)	0.722 (0.72-0.724)	0.717 (0.715-0.719)
20	50	0.562 (0.56-0.564)	0.606 (0.604-0.607)	0.051 (0.05-0.052)	0.056 (0.055-0.057)	0.344 (0.342-0.346)	0.063 (0.062-0.064)	0.024 (0.023-0.025)	0.016 (0.016-0.017)	0.343 (0.342-0.345)	0.393 (0.391-0.395)
	100	0.687 (0.685-0.688)	0.705 (0.703-0.706)	0.048 (0.048-0.049)	0.056 (0.056-0.057)	0.459 (0.457-0.461)	0.064 (0.063-0.065)	0.024 (0.024-0.025)	0.014 (0.013-0.014)	0.466 (0.464-0.468)	0.503 (0.501-0.505)
	250	0.817 (0.815-0.818)	0.805 (0.804-0.807)	0.042 (0.041-0.043)	0.058 (0.057-0.059)	0.61 (0.608-0.612)	0.06 (0.059-0.061)	0.022 (0.022-0.023)	0.011 (0.011-0.012)	0.619 (0.618-0.621)	0.637 (0.636-0.639)
	500	0.884 (0.883-0.885)	0.844 (0.843-0.846)	0.042 (0.041-0.043)	0.055 (0.054-0.056)	0.705 (0.703-0.707)	0.062 (0.061-0.063)	0.021 (0.02-0.021)	0.01 (0.01-0.011)	0.72 (0.718-0.722)	0.716 (0.714-0.717)
	500	0.884 (0.883-0.885)	0.844 (0.843-0.846)	0.042 (0.041-0.043)	0.055 (0.054-0.056)	0.705 (0.703-0.707)	0.062 (0.061-0.063)	0.021 (0.02-0.021)	0.01 (0.01-0.011)	0.72 (0.718-0.722)	0.716 (0.714-0.717)
30	50	0.563 (0.561-0.565)	0.604 (0.602-0.606)	0.053 (0.052-0.054)	0.054 (0.053-0.055)	0.341 (0.34-0.343)	0.058 (0.058-0.059)	0.025 (0.024-0.026)	0.013 (0.013-0.014)	0.344 (0.342-0.346)	0.395 (0.393-0.397)
	100	0.691 (0.689-0.693)	0.698 (0.696-0.699)	0.049 (0.048-0.05)	0.056 (0.055-0.056)	0.463 (0.461-0.465)	0.058 (0.057-0.059)	0.025 (0.024-0.025)	0.012 (0.011-0.012)	0.469 (0.467-0.471)	0.504 (0.502-0.506)
	250	0.818 (0.816-0.819)	0.803 (0.801-0.804)	0.044 (0.044-0.045)	0.055 (0.054-0.056)	0.608 (0.606-0.61)	0.057 (0.056-0.058)	0.022 (0.022-0.023)	0.01 (0.01-0.01)	0.624 (0.622-0.626)	0.636 (0.634-0.638)
	500	0.886 (0.885-0.888)	0.853 (0.852-0.854)	0.041 (0.04-0.042)	0.055 (0.054-0.056)	0.707 (0.705-0.709)	0.058 (0.057-0.059)	0.022 (0.021-0.022)	0.009 (0.008-0.009)	0.719 (0.717-0.721)	0.706 (0.704-0.708)
	500	0.886 (0.885-0.888)	0.853 (0.852-0.854)	0.041 (0.04-0.042)	0.055 (0.054-0.056)	0.707 (0.705-0.709)	0.058 (0.057-0.059)	0.022 (0.021-0.022)	0.009 (0.008-0.009)	0.719 (0.717-0.721)	0.706 (0.704-0.708)
40	50	0.561 (0.559-0.563)	0.602 (0.6-0.603)	0.051 (0.05-0.052)	0.054 (0.053-0.055)	0.345 (0.343-0.347)	0.055 (0.054-0.055)	0.025 (0.025-0.026)	0.013 (0.013-0.014)	0.34 (0.338-0.342)	0.393 (0.391-0.394)
	100	0.689 (0.687-0.691)	0.699 (0.697-0.701)	0.049 (0.048-0.05)	0.053 (0.052-0.054)	0.455 (0.453-0.457)	0.055 (0.054-0.056)	0.026 (0.025-0.027)	0.012 (0.011-0.012)	0.467 (0.465-0.469)	0.502 (0.5-0.504)
	250	0.82 (0.818-0.821)	0.803 (0.801-0.804)	0.044 (0.043-0.045)	0.053 (0.052-0.054)	0.607 (0.605-0.609)	0.053 (0.052-0.054)	0.022 (0.022-0.023)	0.01 (0.009-0.01)	0.622 (0.62-0.624)	0.639 (0.637-0.641)
	500	0.888 (0.887-0.89)	0.856 (0.855-0.857)	0.042 (0.041-0.043)	0.053 (0.052-0.054)	0.704 (0.702-0.706)	0.054 (0.053-0.055)	0.022 (0.022-0.023)	0.008 (0.008-0.008)	0.721 (0.719-0.723)	0.713 (0.711-0.715)
	500	0.888 (0.887-0.89)	0.856 (0.855-0.857)	0.042 (0.041-0.043)	0.053 (0.052-0.054)	0.704 (0.702-0.706)	0.054 (0.053-0.055)	0.022 (0.022-0.023)	0.008 (0.008-0.008)	0.721 (0.719-0.723)	0.713 (0.711-0.715)

*Default denotes MAST was implemented without random effects, RE denotes random effects, Corrected denotes data were batch-corrected for individual prior to analysis without using individual as a random effect, GLM denotes generalized linear model, and GLMM denotes generalized linear mixed-effects model.

**Two-part Hurdle model as implemented in MAST, Tweedie distribution as implemented in 'glmmTMB', GEE1 as implemented in 'geepack', Pseudobulk averaged or summed across cells within an individual and was implemented in DESeq2, Modified *t* as implemented in ROTS, and Tobit as implemented in Monocle.

Supplementary Table 2: Type I error rates for some currently applied tools in single-cell analyses. Type I error rates of 10 different methods under 20 different conditions and a significance threshold of $p < 0.01$. Hypothesis tests were two-sided with no correction for multiple comparisons – as each gene was simulated and analyzed independently. 250,000 iterations were computed to obtain an error rate for each method.

N _{ind}	N _{cells}	Two-part Hurdle			Tweedie		GEE1	Pseudobulk			Modified <i>t</i>
		Default	Corrected	RE	GLMM	GLM		Mean	Sum	Tobit	
5	50	0.447	0.567	0.016	0.028	0.232	0.046	0.005	0.011	0.233	0.286
	100	0.584	0.674	0.015	0.029	0.351	0.042	0.004	0.009	0.353	0.397
	250	0.736	0.748	0.015	0.026	0.516	0.037	0.005	0.008	0.528	0.547
	500	0.820	0.776	0.015	0.024	0.626	0.037	0.005	0.007	0.645	0.630
10	50	0.447	0.532	0.010	0.016	0.234	0.022	0.004	0.005	0.229	0.282
	100	0.593	0.659	0.010	0.015	0.348	0.020	0.005	0.004	0.350	0.390
	250	0.744	0.763	0.009	0.013	0.512	0.019	0.004	0.004	0.524	0.538
	500	0.830	0.805	0.009	0.012	0.626	0.018	0.004	0.004	0.642	0.628
20	50	0.447	0.528	0.009	0.012	0.230	0.015	0.004	0.003	0.225	0.277
	100	0.591	0.642	0.008	0.012	0.345	0.014	0.004	0.002	0.342	0.391
	250	0.752	0.767	0.007	0.011	0.512	0.013	0.004	0.002	0.517	0.539
	500	0.838	0.820	0.007	0.010	0.624	0.012	0.003	0.001	0.639	0.628
30	50	0.447	0.527	0.010	0.011	0.227	0.013	0.004	0.002	0.223	0.278
	100	0.593	0.637	0.009	0.011	0.348	0.013	0.004	0.001	0.348	0.391
	250	0.752	0.759	0.007	0.010	0.512	0.012	0.004	0.001	0.523	0.541
	500	0.842	0.827	0.007	0.010	0.626	0.013	0.003	0.001	0.637	0.615
40	50	0.445	0.525	0.009	0.011	0.230	0.012	0.004	0.002	0.221	0.277
	100	0.591	0.638	0.009	0.011	0.342	0.012	0.004	0.001	0.345	0.390
	250	0.751	0.758	0.007	0.010	0.508	0.009	0.004	0.001	0.519	0.541
	500	0.844	0.828	0.006	0.009	0.622	0.011	0.004	0.001	0.639	0.624

*Default denotes MAST was implemented without random effects, RE denotes random effects, Corrected denotes data were batch-corrected for individual prior to analysis without using individual as a random effect, GLM denotes generalized linear model, and GLMM denotes generalized linear mixed-effects model.

**Two-part Hurdle model as implemented in MAST, Tweedie distribution as implemented in ‘glmmTMB’, GEE1 as implemented in ‘geepack’, Pseudobulk averaged or summed across cells within an individual and was implemented in DESeq2, Modified *t* as implemented in ROTS, and Tobit as implemented in Monocle.

Supplementary Table 3: Type I error rates for some currently applied tools in single-cell analyses. Type I error rates of 10 different methods under 20 different conditions and a significance threshold of $p < 0.001$. Hypothesis tests were two-sided with no correction for multiple comparisons – as each gene was simulated and analyzed independently. 250,000 iterations were computed to obtain an error rate for each method.

N _{ind}	N _{cells}	Two-part Hurdle			Tweedie		GEE1	Pseudobulk		Tobit	Modified <i>t</i>
		Default	Corrected	RE	GLMM	GLM		Mean	Sum		
5	50	3.40x10 ⁻¹	4.84x10 ⁻¹	2.27x10 ⁻³	8.79x10 ⁻³	1.48x10 ⁻¹	1.65x10 ⁻²	8.43x10 ⁻⁴	3.53x10 ⁻³	1.39x10 ⁻¹	1.94x10 ⁻¹
	100	4.87x10 ⁻¹	6.22x10 ⁻¹	1.84x10 ⁻³	7.98x10 ⁻³	2.53x10 ⁻¹	1.46x10 ⁻²	4.96x10 ⁻⁴	2.66x10 ⁻³	2.45x10 ⁻¹	2.99x10 ⁻¹
	250	6.66x10 ⁻¹	7.18x10 ⁻¹	1.96x10 ⁻³	7.55x10 ⁻³	4.23x10 ⁻¹	1.20x10 ⁻²	7.36x10 ⁻⁴	1.76x10 ⁻³	4.25x10 ⁻¹	4.52x10 ⁻¹
	500	7.71x10 ⁻¹	7.52x10 ⁻¹	2.19x10 ⁻³	6.85x10 ⁻³	5.46x10 ⁻¹	1.22x10 ⁻²	6.97x10 ⁻⁴	1.56x10 ⁻³	5.59x10 ⁻¹	5.40x10 ⁻¹
10	50	3.39x10 ⁻¹	4.53x10 ⁻¹	1.03x10 ⁻³	2.81x10 ⁻³	1.46x10 ⁻¹	4.81x10 ⁻³	5.10x10 ⁻⁴	1.35x10 ⁻³	1.36x10 ⁻¹	1.88x10 ⁻¹
	100	4.92x10 ⁻¹	5.87x10 ⁻¹	9.39x10 ⁻⁴	2.58x10 ⁻³	2.49x10 ⁻¹	3.92x10 ⁻³	6.48x10 ⁻⁴	8.32x10 ⁻⁴	2.41x10 ⁻¹	2.93x10 ⁻¹
	250	6.71x10 ⁻¹	7.29x10 ⁻¹	8.41x10 ⁻⁴	1.97x10 ⁻³	4.18x10 ⁻¹	3.57x10 ⁻³	3.41x10 ⁻⁴	6.46x10 ⁻⁴	4.21x10 ⁻¹	4.41x10 ⁻¹
	500	7.78x10 ⁻¹	7.82x10 ⁻¹	9.53x10 ⁻⁴	1.68x10 ⁻³	5.42x10 ⁻¹	2.86x10 ⁻³	3.14x10 ⁻⁴	6.46x10 ⁻⁴	5.54x10 ⁻¹	5.39x10 ⁻¹
20	50	3.39x10 ⁻¹	4.50x10 ⁻¹	9.28x10 ⁻⁴	1.47x10 ⁻³	1.43x10 ⁻¹	2.11x10 ⁻³	3.79x10 ⁻⁴	6.48x10 ⁻⁴	1.34x10 ⁻¹	1.83x10 ⁻¹
	100	4.92x10 ⁻¹	5.78x10 ⁻¹	6.90x10 ⁻⁴	1.35x10 ⁻³	2.48x10 ⁻¹	2.02x10 ⁻³	5.05x10 ⁻⁴	3.11x10 ⁻⁴	2.36x10 ⁻¹	2.90x10 ⁻¹
	250	6.76x10 ⁻¹	7.18x10 ⁻¹	5.88x10 ⁻⁴	9.21x10 ⁻⁴	4.17x10 ⁻¹	1.31x10 ⁻³	3.53x10 ⁻⁴	1.94x10 ⁻⁴	4.15x10 ⁻¹	4.45x10 ⁻¹
	500	7.84x10 ⁻¹	7.93x10 ⁻¹	7.33x10 ⁻⁴	9.09x10 ⁻⁴	5.40x10 ⁻¹	1.15x10 ⁻³	2.18x10 ⁻⁴	1.29x10 ⁻⁴	5.51x10 ⁻¹	5.37x10 ⁻¹
30	50	3.37x10 ⁻¹	4.46x10 ⁻¹	7.98x10 ⁻⁴	8.64x10 ⁻⁴	1.39x10 ⁻¹	1.49x10 ⁻³	3.58x10 ⁻⁴	2.27x10 ⁻⁴	1.31x10 ⁻¹	1.86x10 ⁻¹
	100	4.92x10 ⁻¹	5.70x10 ⁻¹	5.93x10 ⁻⁴	8.62x10 ⁻⁴	2.47x10 ⁻¹	1.22x10 ⁻³	4.04x10 ⁻⁴	1.55x10 ⁻⁴	2.40x10 ⁻¹	2.91x10 ⁻¹
	250	6.76x10 ⁻¹	7.06x10 ⁻¹	7.01x10 ⁻⁴	9.36x10 ⁻⁴	4.16x10 ⁻¹	9.74x10 ⁻⁴	3.66x10 ⁻⁴	8.40x10 ⁻⁵	4.21x10 ⁻¹	4.44x10 ⁻¹
	500	7.88x10 ⁻¹	7.96x10 ⁻¹	6.14x10 ⁻⁴	8.30x10 ⁻⁴	5.42x10 ⁻¹	1.64x10 ⁻³	2.46x10 ⁻⁴	2.58x10 ⁻⁵	5.50x10 ⁻¹	5.23x10 ⁻¹
40	50	3.37x10 ⁻¹	4.45x10 ⁻¹	7.15x10 ⁻⁴	1.19x10 ⁻³	1.42x10 ⁻¹	1.54x10 ⁻³	3.90x10 ⁻⁴	1.42x10 ⁻⁴	1.30x10 ⁻¹	1.84x10 ⁻¹
	100	4.90x10 ⁻¹	5.71x10 ⁻¹	8.31x10 ⁻⁴	1.20x10 ⁻³	2.43x10 ⁻¹	1.20x10 ⁻³	3.46x10 ⁻⁴	1.36x10 ⁻⁴	2.37x10 ⁻¹	2.92x10 ⁻¹
	250	6.76x10 ⁻¹	7.08x10 ⁻¹	4.94x10 ⁻⁴	8.58x10 ⁻⁴	4.13x10 ⁻¹	1.20x10 ⁻³	2.84x10 ⁻⁴	9.04x10 ⁻⁵	4.14x10 ⁻¹	4.44x10 ⁻¹
	500	7.90x10 ⁻¹	7.90x10 ⁻¹	4.46x10 ⁻⁴	5.21x10 ⁻⁴	5.40x10 ⁻¹	8.39x10 ⁻⁴	3.40x10 ⁻⁴	3.90x10 ⁻⁵	5.51x10 ⁻¹	5.33x10 ⁻¹

*Default denotes MAST was implemented without random effects, RE denotes random effects, Corrected denotes data were batch-corrected for individual prior to analysis without using individual as a random effect, GLM denotes generalized linear model, and GLMM denotes generalized linear mixed-effects model.

**Two-part Hurdle model as implemented in MAST, Tweedie distribution as implemented in 'glmmTMB', GEE1 as implemented in 'geepack', Pseudobulk averaged or summed across cells within an individual and was implemented in DESeq2, Modified *t* as implemented in ROTS, and Tobit as implemented in Monocle.

Supplementary Table 4: Type I error rates for some currently applied tools in single-cell analyses. Type I error rates of 10 different methods under 20 different conditions and a significance threshold of $p < 0.0001$. Hypothesis tests were two-sided with no correction for multiple comparisons – as each gene was simulated and analyzed independently. 250,000 iterations were computed to obtain an error rate for each method.

N _{ind}	N _{cells}	Two-part Hurdle			Tweedie		GEE1	Pseudobulk			
		Default	Corrected	RE	GLMM	GLM		Mean	Sum	Tobit	Modified <i>t</i>
5	50	2.68x10 ⁻¹	4.18x10 ⁻¹	4.35x10 ⁻⁴	3.68x10 ⁻³	9.93x10 ⁻²	7.71x10 ⁻³	1.67x10 ⁻⁴	1.76x10 ⁻³	8.79x10 ⁻²	1.40x10 ⁻¹
	100	4.16x10 ⁻¹	5.74x10 ⁻¹	2.65x10 ⁻⁴	2.79x10 ⁻³	1.95x10 ⁻¹	6.28x10 ⁻³	8.72x10 ⁻⁵	1.15x10 ⁻³	1.79x10 ⁻¹	2.36x10 ⁻¹
	250	6.11x10 ⁻¹	6.94x10 ⁻¹	3.44x10 ⁻⁴	3.18x10 ⁻³	3.55x10 ⁻¹	4.83x10 ⁻³	1.91x10 ⁻⁴	6.55x10 ⁻⁴	3.53x10 ⁻¹	3.86x10 ⁻¹
	500	7.28x10 ⁻¹	7.34x10 ⁻¹	3.60x10 ⁻⁴	2.48x10 ⁻³	4.85x10 ⁻¹	4.98x10 ⁻³	1.41x10 ⁻⁴	5.61x10 ⁻⁴	4.95x10 ⁻¹	4.75x10 ⁻¹
10	50	2.67x10 ⁻¹	3.96x10 ⁻¹	1.12x10 ⁻⁴	7.80x10 ⁻⁴	9.80x10 ⁻²	1.23x10 ⁻³	6.45x10 ⁻⁵	7.84x10 ⁻⁴	8.69x10 ⁻²	1.35x10 ⁻¹
	100	4.20x10 ⁻¹	5.31x10 ⁻¹	9.76x10 ⁻⁵	6.11x10 ⁻⁴	1.88x10 ⁻¹	9.31x10 ⁻⁴	9.31x10 ⁻⁵	3.97x10 ⁻⁴	1.74x10 ⁻¹	2.30x10 ⁻¹
	250	6.11x10 ⁻¹	6.97x10 ⁻¹	9.89x10 ⁻⁵	3.94x10 ⁻⁴	3.51x10 ⁻¹	9.24x10 ⁻⁴	8.00x10 ⁻⁵	2.38x10 ⁻⁴	3.46x10 ⁻¹	3.74x10 ⁻¹
	500	7.34x10 ⁻¹	7.65x10 ⁻¹	6.78x10 ⁻⁵	4.12x10 ⁻⁴	4.79x10 ⁻¹	6.09x10 ⁻⁴	5.18x10 ⁻⁵	1.81x10 ⁻⁴	4.88x10 ⁻¹	4.72x10 ⁻¹
20	50	2.67x10 ⁻¹	3.91x10 ⁻¹	5.75x10 ⁻⁵	2.00x10 ⁻⁴	9.65x10 ⁻²	3.29x10 ⁻⁴	5.59x10 ⁻⁵	4.28x10 ⁻⁴	8.44x10 ⁻²	1.32x10 ⁻¹
	100	4.19x10 ⁻¹	5.28x10 ⁻¹	4.19x10 ⁻⁵	1.58x10 ⁻⁴	1.85x10 ⁻¹	3.00x10 ⁻⁴	8.52x10 ⁻⁵	1.50x10 ⁻⁴	1.69x10 ⁻¹	2.26x10 ⁻¹
	250	6.16x10 ⁻¹	6.73x10 ⁻¹	6.82x10 ⁻⁵	1.95x10 ⁻⁴	3.49x10 ⁻¹	1.51x10 ⁻⁴	3.62x10 ⁻⁵	4.53x10 ⁻⁵	3.41x10 ⁻¹	3.77x10 ⁻¹
	500	7.39x10 ⁻¹	7.68x10 ⁻¹	6.81x10 ⁻⁵	1.15x10 ⁻⁴	4.77x10 ⁻¹	1.49x10 ⁻⁴	2.87x10 ⁻⁵	2.59x10 ⁻⁵	4.83x10 ⁻¹	4.73x10 ⁻¹
30	50	2.67x10 ⁻¹	3.87x10 ⁻¹	1.10x10 ⁻⁴	1.14x10 ⁻⁴	9.27x10 ⁻²	1.64x10 ⁻⁴	4.21x10 ⁻⁵	1.10x10 ⁻⁴	8.35x10 ⁻²	1.33x10 ⁻¹
	100	4.18x10 ⁻¹	5.20x10 ⁻¹	6.85x10 ⁻⁵	1.14x10 ⁻⁴	1.85x10 ⁻¹	1.49x10 ⁻⁴	9.20x10 ⁻⁵	8.44x10 ⁻⁵	1.74x10 ⁻¹	2.26x10 ⁻¹
	250	6.15x10 ⁻¹	6.66x10 ⁻¹	1.10x10 ⁻⁴	1.24x10 ⁻⁴	3.49x10 ⁻¹	7.95x10 ⁻⁵	1.44x10 ⁻⁵	1.94x10 ⁻⁵	3.45x10 ⁻¹	3.77x10 ⁻¹
	500	7.42x10 ⁻¹	7.65x10 ⁻¹	2.73x10 ⁻⁵	2.34x10 ⁻⁵	4.80x10 ⁻¹	-	2.88x10 ⁻⁵	6.46x10 ⁻⁶	4.82x10 ⁻¹	4.57x10 ⁻¹
40	50	2.64x10 ⁻¹	3.84x10 ⁻¹	1.48x10 ⁻⁴	1.53x10 ⁻⁴	9.25x10 ⁻²	1.51x10 ⁻⁴	4.18x10 ⁻⁵	7.78x10 ⁻⁵	8.06x10 ⁻²	1.31x10 ⁻¹
	100	4.17x10 ⁻¹	5.19x10 ⁻¹	5.38x10 ⁻⁵	1.76x10 ⁻⁴	1.82x10 ⁻¹	1.89x10 ⁻⁴	3.53x10 ⁻⁵	3.21x10 ⁻⁵	1.71x10 ⁻¹	2.29x10 ⁻¹
	250	6.16x10 ⁻¹	6.70x10 ⁻¹	6.87x10 ⁻⁵	8.00x10 ⁻⁵	3.45x10 ⁻¹	-	2.83x10 ⁻⁵	1.30x10 ⁻⁵	3.40x10 ⁻¹	3.76x10 ⁻¹
	500	7.44x10 ⁻¹	7.58x10 ⁻¹	4.15x10 ⁻⁵	1.26x10 ⁻⁴	4.78x10 ⁻¹	-	4.36x10 ⁻⁵	1.32x10 ⁻⁵	4.85x10 ⁻¹	4.66x10 ⁻¹

*Default denotes MAST was implemented without random effects, RE denotes random effects, Corrected denotes data were batch-corrected for individual prior to analysis without using individual as a random effect, GLM denotes generalized linear model, and GLMM denotes generalized linear mixed-effects model.

**Two-part Hurdle model as implemented in MAST, Tweedie distribution as implemented in ‘glmmTMB’, GEE1 as implemented in ‘geepack’, Pseudobulk averaged or summed across cells within an individual and was implemented in DESeq2, Modified *t* as implemented in ROTS, and Tobit as implemented in Monocle.

Supplementary Table 5: Rank-order correlations of some currently applied tools in single-cell analysis. Using a dataset of 5,000 independently simulated genes with varying degrees of fold-changes, p-values were estimated with each method. Hypothesis tests were two-sided with no correction for multiple comparisons – as each gene was simulated and analyzed independently. P-values were ranked alongside the simulated $-\log(\text{fold-change})$ values and Spearman's rank-correlation coefficients were computed between each method. Methods that properly account for within-sample correlation show the strongest correlation with the absolute value of $-\log(\text{fold-change})$.

Spearman's rank-order correlation coefficient		Two-part hurdle				Tweedie		Pseudobulk			Modified	
		-logFC	Default	Corrected	RE	GLM	GLMM	Sum	Mean	GEE1	Tobit	t
-logFC		1	0.88	0.46	0.9	0.81	0.82	0.87	0.88	0.82	0.49	0.84
Two-part hurdle	Default	0.88	1	0.42	1	0.89	0.9	0.94	0.93	0.9	0.44	0.89
	Corrected	0.46	0.42	1	0.5	0.47	0.49	0.42	0.41	0.49	0.2	0.34
	RE	0.9	0.96	0.46	1	0.87	0.89	0.95	0.93	0.89	0.4	0.86
Tweedie	GLM	0.81	0.89	0.47	0.9	1	0.99	0.87	0.86	0.98	0.4	0.8
	GLMM	0.82	0.9	0.49	0.9	0.99	1	0.9	0.88	0.99	0.38	0.81
Pseudobulk	Sum	0.87	0.94	0.42	1	0.87	0.9	1	0.98	0.9	0.4	0.91
	Mean	0.88	0.93	0.41	0.9	0.86	0.88	0.98	1	0.88	0.47	0.96
GEE1		0.82	0.9	0.49	0.9	0.98	0.99	0.9	0.88	1	0.39	0.82
Tobit		0.49	0.44	0.2	0.4	0.4	0.38	0.4	0.47	0.39	1	0.54
Modified t		0.84	0.89	0.34	0.9	0.8	0.81	0.91	0.96	0.82	0.54	1

*Default denotes MAST was implemented without random effects, RE denotes random effects, Corrected denotes data were batch-corrected for individual prior to analysis without using individual as a random effect, GLM denotes generalized linear model, and GLMM denotes generalized linear mixed-effects model.

**Two-part Hurdle model as implemented in MAST, Tweedie distribution as implemented in 'glmmTMB', GEE1 as implemented in 'geepack', Pseudobulk averaged or summed across cells within an individual and was implemented in DESeq2, Modified t as implemented in ROTS, and Tobit as implemented in Monocle.

Supplementary Table 6: Run times (in seconds) of some tools applied in these analyses. Run times for each method were estimated for 50 cells per individual and 1,000 genes per run computed to obtain an error rate for each method.

N _{ind}	Two-part Hurdle			Tweedie			Pseudobulk		Tobit	Modified t
	Default	Corrected	RE	GLMM	GLM	GEE1	Mean	Sum		
5	6.16	6.94	114.84	333.54	79.18	23.40	5.62	5.69	8.06	23.85
10	7.95	8.50	211.89	589.32	98.72	44.38	13.50	7.63	10.06	45.70
20	12.17	12.85	294.17	1372.14	240.44	78.31	8.91	10.19	12.43	82.08
30	15.61	17.19	474.62	1673.15	294.81	108.83	10.68	12.39	14.41	139.08

*Default denotes MAST was implemented without random effects, RE denotes random effects, Corrected denotes data were batch-corrected for individual prior to analysis without using individual as a random effect, GLM denotes generalized linear model, and GLMM denotes generalized linear mixed-effects model.

**Two-part Hurdle model as implemented in MAST, Tweedie distribution as implemented in ‘glmmTMB’, GEE1 as implemented in ‘geepack’, Pseudobulk averaged or summed across cells within an individual and was implemented in DESeq2, Modified t as implemented in ROTS, and Tobit as implemented in Monocle.