powsimR: Power analysis for bulk and single cell RNA-seq experiments

Supplementary File 2

by

Beate Vieth[1], Christoph Ziegenhain[1], Swati Parekh[1], Wolfgang Enard[1] and Ines Hellmann[1]

[1]Anthropology & Human Genomics, Department of Biology II,
Ludwig-Maximilians University, Munich, Germany

1 Single Cell RNA-sequencing Datasets

We analyzed RNA-sequencing data of 8 published studies that utilized 9 different RNA-seq library preparation methods (Table S1). One of the major differences between the methods is the use of Unique Molecular Identifiers (UMIs) that allow for confident removal of PCR-duplicates (Grün et al. (2014); Ziegenhain et al. (2017)). For all datasets, we evaluated the fit of 5 different distributions, namely the Poisson, negative binomial (NB), zero-inflated negative binomial (ZINB) and Poisson (ZIP) and Beta-Poisson (BP). For the vast majority the NB would be the distribution of choice. This is especially true for the UMI-methods: Here no zero-inflation is needed for modeling the gene expression distribution. On the contrary, also a simple Poisson often provides the best fit.

	Study	Species	Cell type	Cell classification	No. cells	Library preparation	UMI	Coverage
1	Buettner et al. 2015	Mouse	Embryonic stem cells	G1 cell cycle stage	96	SmartSeq/C1	no	full length
2	Buettner et al. 2015	Mouse	Embryonic stem cells	G2M cell cycle stage	96	SmartSeq/C1	no	full length
3	Buettner et al. 2015	Mouse	Embryonic stem cells	S cell cycle stage	96	SmartSeq/C1	no	full length
4	Islam et al. 2014	Mouse	Embryonic stem cells		96	STRT-UMI	yes	end sequencing
5	Islam et al. 2011	Mouse	Embryonic stem cells		48	STRT	no	end sequencing
6	Islam et al. 2011	Mouse	Mouse embryonic fibroblast		48	STRT	no	end sequencing
7	Kolodziejczk et al. 2015	Mouse	Embryonic stem cells	alternative 2i media + LIF	194	SmartSeq/C1	no	full length
8	Kolodziejczk et al. 2015	Mouse	Embryonic stem cells	serum + LIF	242	SmartSeq/C1	no	full length
9	Kolodziejczk et al. 2015	Mouse	Embryonic stem cells	standard 2i media + LIF	433	Smart-seq/C1	no	full length
10	Pollen et al. 2014	Human	Primary epidermal keratinocytes		80	SmartSeq/C1	no	full length
11	Pollen et al 2014	Human	Induced pluripotent stem cells		48	SmartSeq/C1	no	full length
12	Pollen et al. 2014	Human	Cultured primary human neurons		32	SmartSeq/C1	no	full length
13	Pollen et al. 2014	Human	HCC1954 breast cancer cells		44	SmartSeq/C1	no	full length
14	Pollen et al. 2014	Human	HCC1954 B lymphoblastoid cells		34	SmartSeq/C1	no	full length
15	Pollen et al. 2014	Human	HL-60 human promyelocytic leukemia cells		108	SmartSeq/C1	no	full length
16	Pollen et al. 2014	Human	K-562 myelogenous leukemia cells		178	SmartSeq/C1	no	full length
17	Pollen et al. 2014	Human	Neural progenitor cells obtained by differentiation of iPS line		30	SmartSeq/C1	no	full length
18	Pollen et al. 2014	Human	Primary human neurons		16	SmartSeq/C1	no	full length
19	Pollen et al. 2014	Human	BJ Human Fibroblasts	early passage, p6	74	Smart-seq/C1	no	full length
20	Soumillon et al. 2014	Human	adipose-derived stem cells	1 day post-differentiation	6197	SCRB-seq	yes	end sequencing
21	Soumillon et al. 2014	Human	adipose-derived stem cells	2 days post-differentiation	1599	SCRB-seq	yes	end sequencing
22	Soumillon et al. 2014	Human	adipose-derived stem cells	3 days post-differentiation	2068	SCRB-seq	yes	end sequencing
23	Zheng et al. 2017	Human	Peripheral Blood Mononuclear Cells	CD19+ B Cells	10085	10XGenomics	yes	end sequencing
24	Zheng et al. 2017	Human	Peripheral Blood Mononuclear Cells	CD14+ Monocytes	2612	10XGenomics	yes	end sequencing
25	Zheng et al. 2017	Human	Peripheral Blood Mononuclear Cells	CD34+ Cells	9232	10XGenomics	yes	end sequencing
26	Zheng et al. 2017	Human	Peripheral Blood Mononuclear Cells	CD4+ T Helper Cells	11213	10XGenomics	yes	end sequencing
27	Zheng et al. 2017	Human	Peripheral Blood Mononuclear Cells	CD56+ NK Cells	8385	10XGenomics	yes	end sequencing
28	Zheng et al. 2017	Human	Peripheral Blood Mononuclear Cells	CD8+ Cytotoxic T Cells	10209	10XGenomics	yes	end sequencing
29	Zheng et al. 2017	Human	Peripheral Blood Mononuclear Cells	CD4+/CD45RO+ Memory T Cells	10224	10XGenomics	yes	end sequencing
30	Zheng et al. 2017	Human	Peripheral Blood Mononuclear Cells	CD8+/CD45RA+ Naive Cytotoxic T Cells	11953	10XGenomics	yes	end sequencing
31	Zheng et al. 2017	Human	Peripheral Blood Mononuclear Cells	CD4+/CD45RA+/CD25- Naive T Cells	10479	10XGenomics	yes	end sequencing
32	Zheng et al. 2017	Human	Peripheral Blood Mononuclear Cells	CD4+/CD25+ Regulatory T Cells	10263	10XGenomics	yes	end sequencing
33	Ziegenhain et al. 2017	Mouse	Embryonic stem cells		71	CEL-seq2	yes	end sequencing
34	Ziegenhain et al. 2017	Mouse	Embryonic stem cells		76	Drop-seq	yes	end sequencing
35	Ziegenhain et al. 2017	Mouse	Embryonic stem cells		65	MARS-seq	yes	end sequencing
36	Ziegenhain et al. 2017	Mouse	Embryonic stem cells		84	SCRB-seq	yes	end sequencing
37	Ziegenhain et al. 2017	Mouse	Embryonic stem cells		287	Smart-seq/C1	no	full length
38	Ziegenhain et al. 2017	Mouse	Embryonic stem cells		157	Smart-seq2	no	full length

Table S1: Key properties of the single cell RNA-seq experiments for distribution evaluation.

2 Distributional Fitting per Dataset

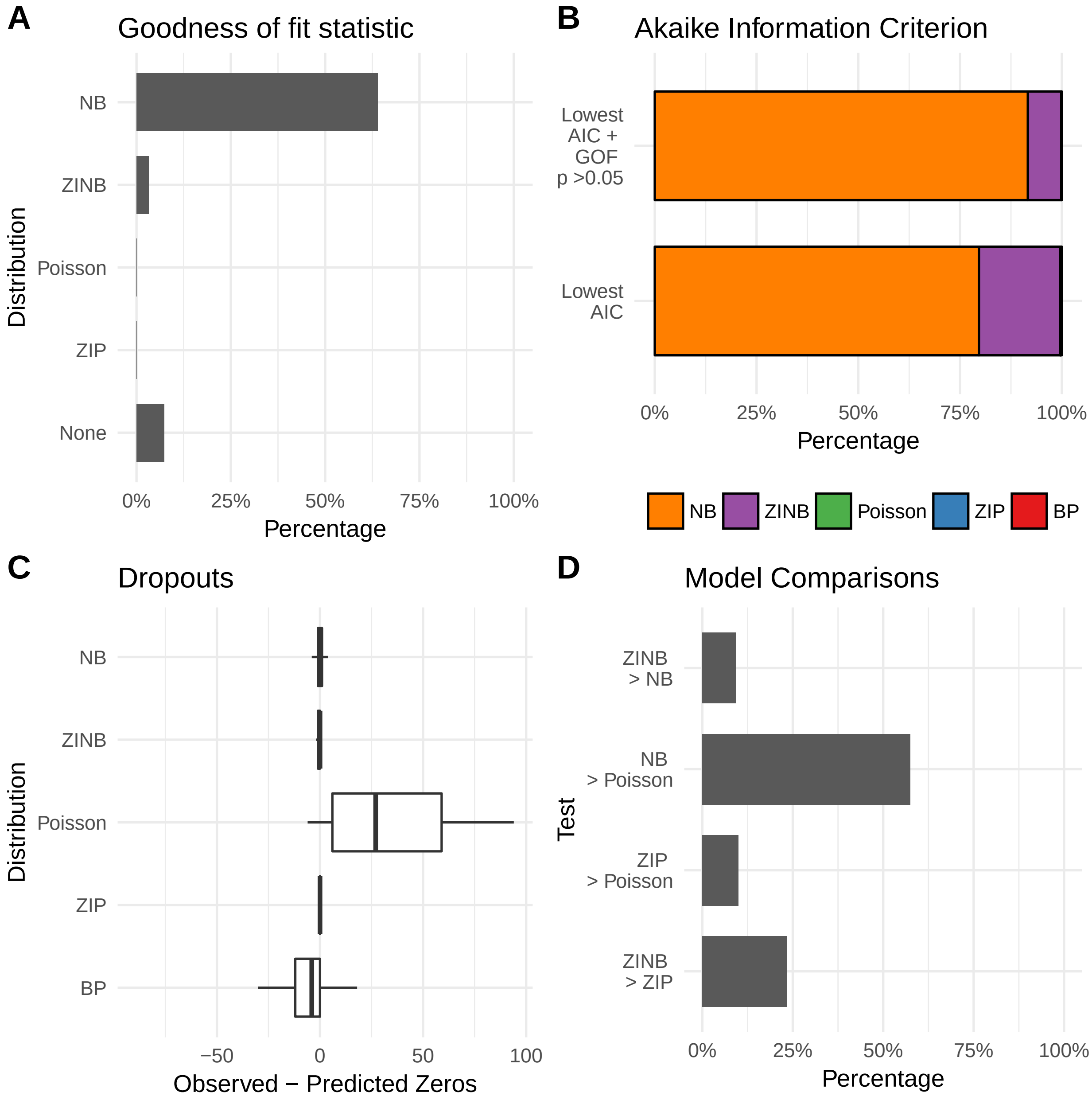

Figure S1: Buettner et al. 2015: Embryonic stem cells G1 cell cycle stage (SmartSeq/C1). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

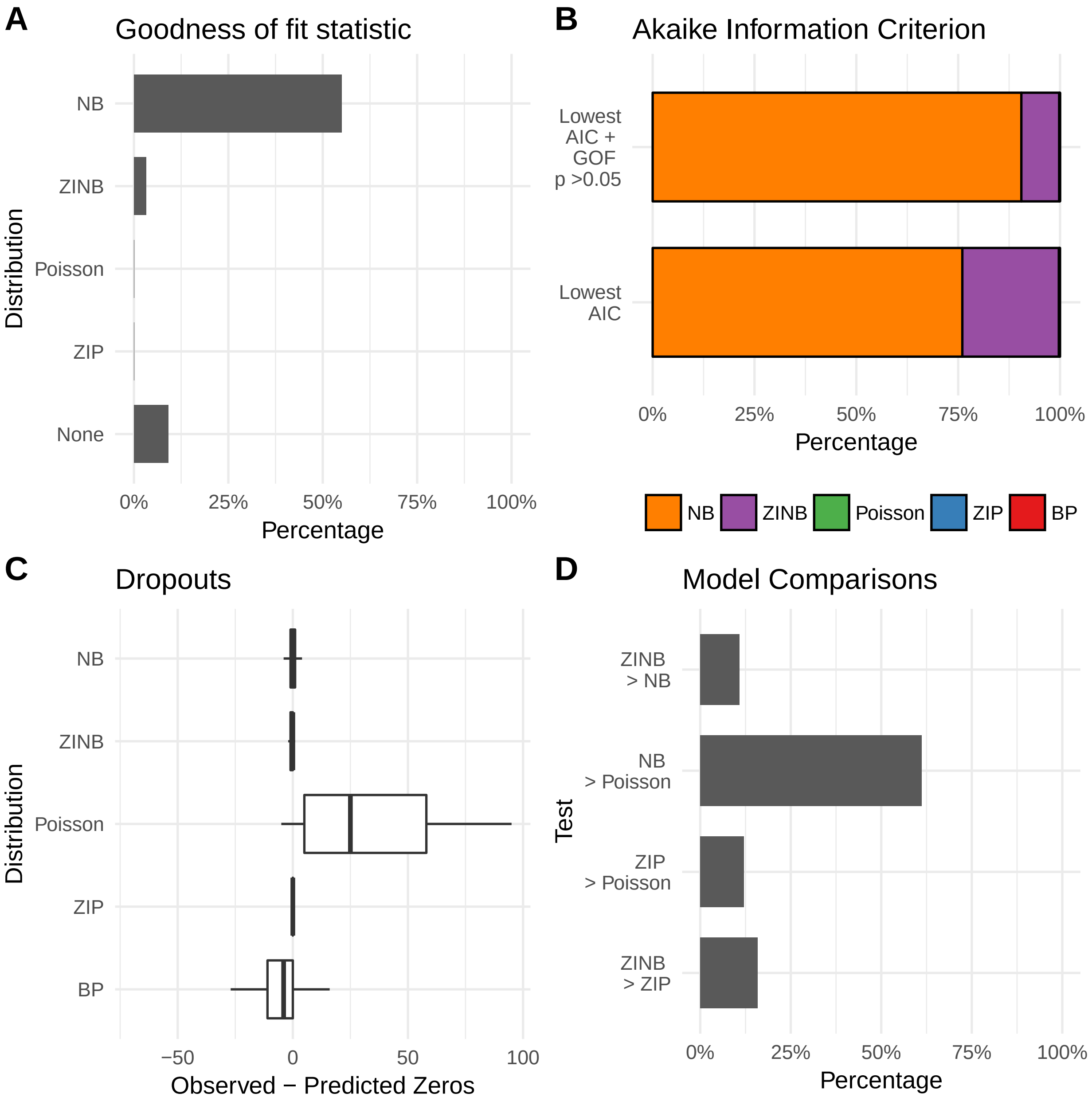

Figure S2: Buettner et al. 2015: Embryonic stem cells G2M cell cycle stage (SmartSeq/C1). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

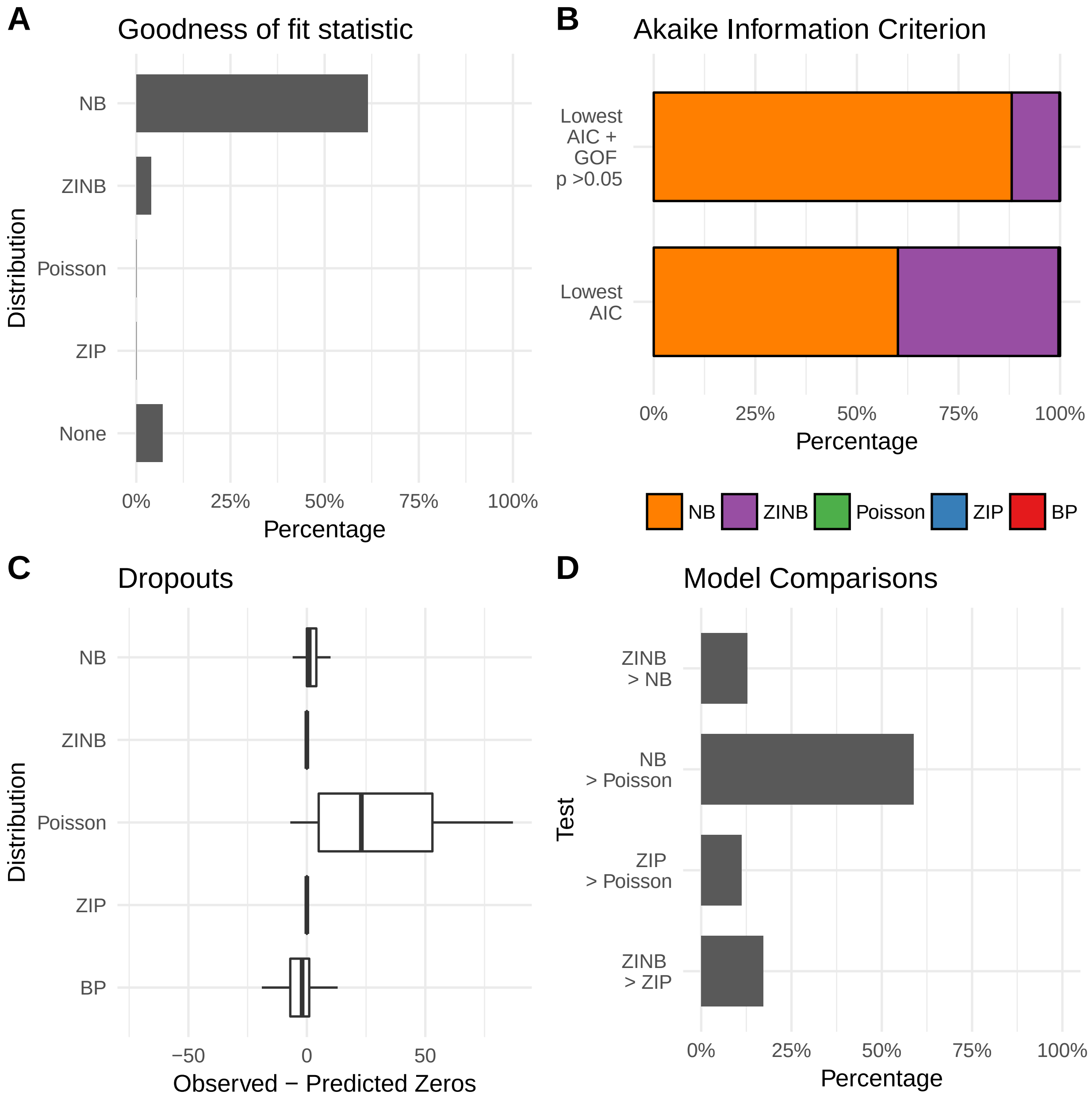

Figure S3: Buettner et al. 2015: Embryonic stem cells S cell cycle stage (SmartSeq/C1). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

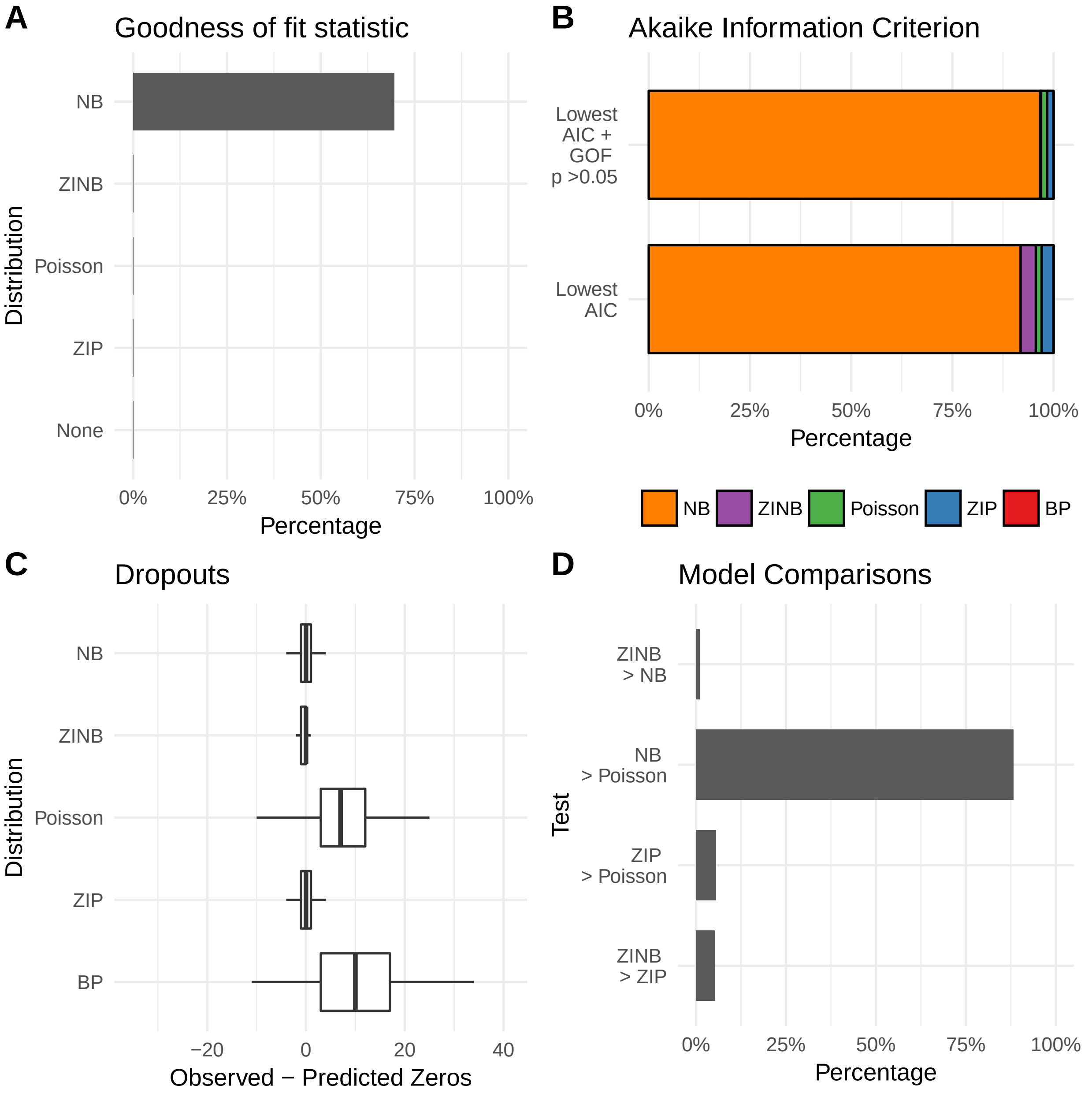

Figure S4: Islam et al. 2014: Embryonic stem cells (STRT-UMI). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

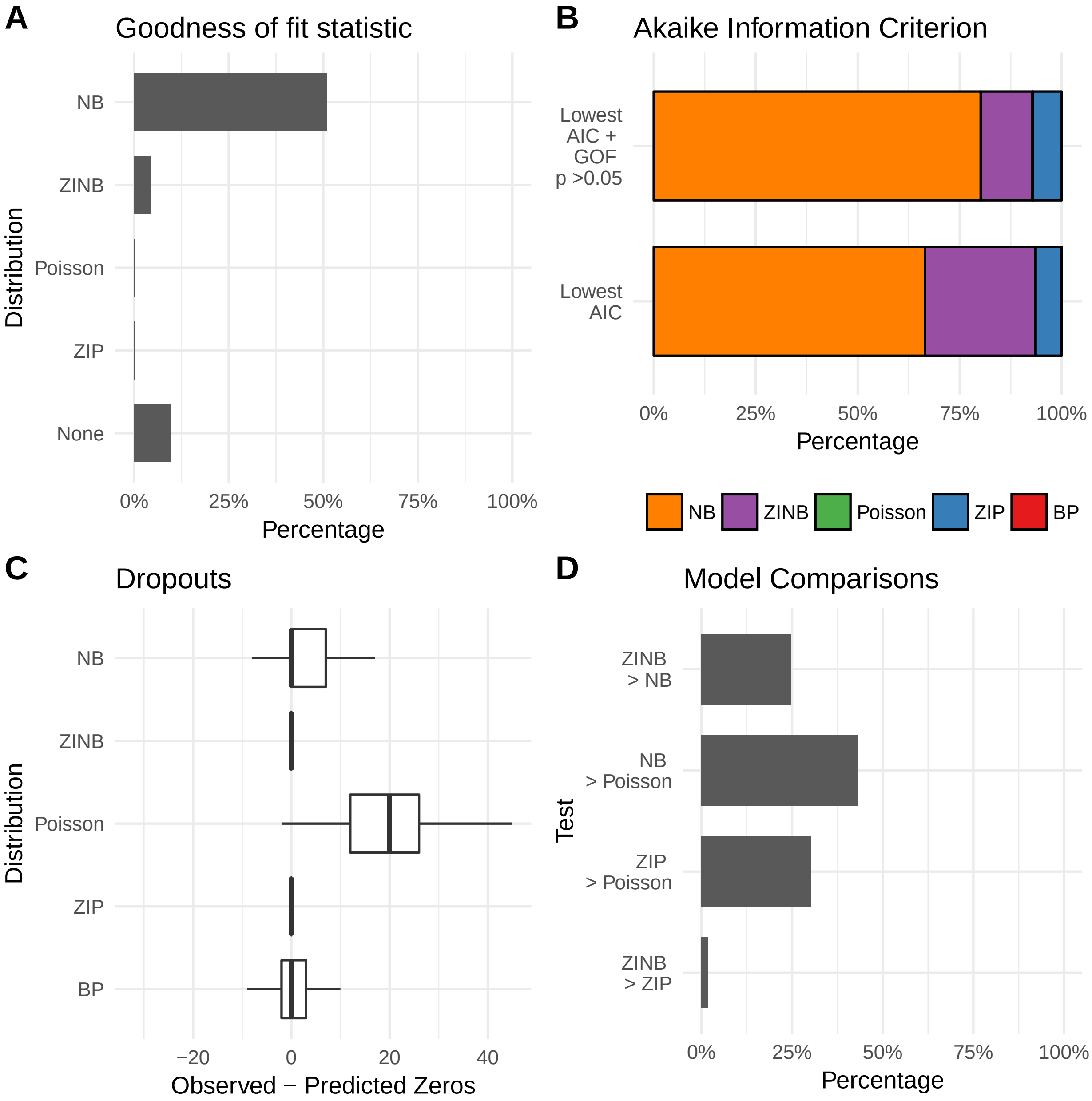

Figure S5: Islam et al. 2011: Embryonic stem cells (STRT). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

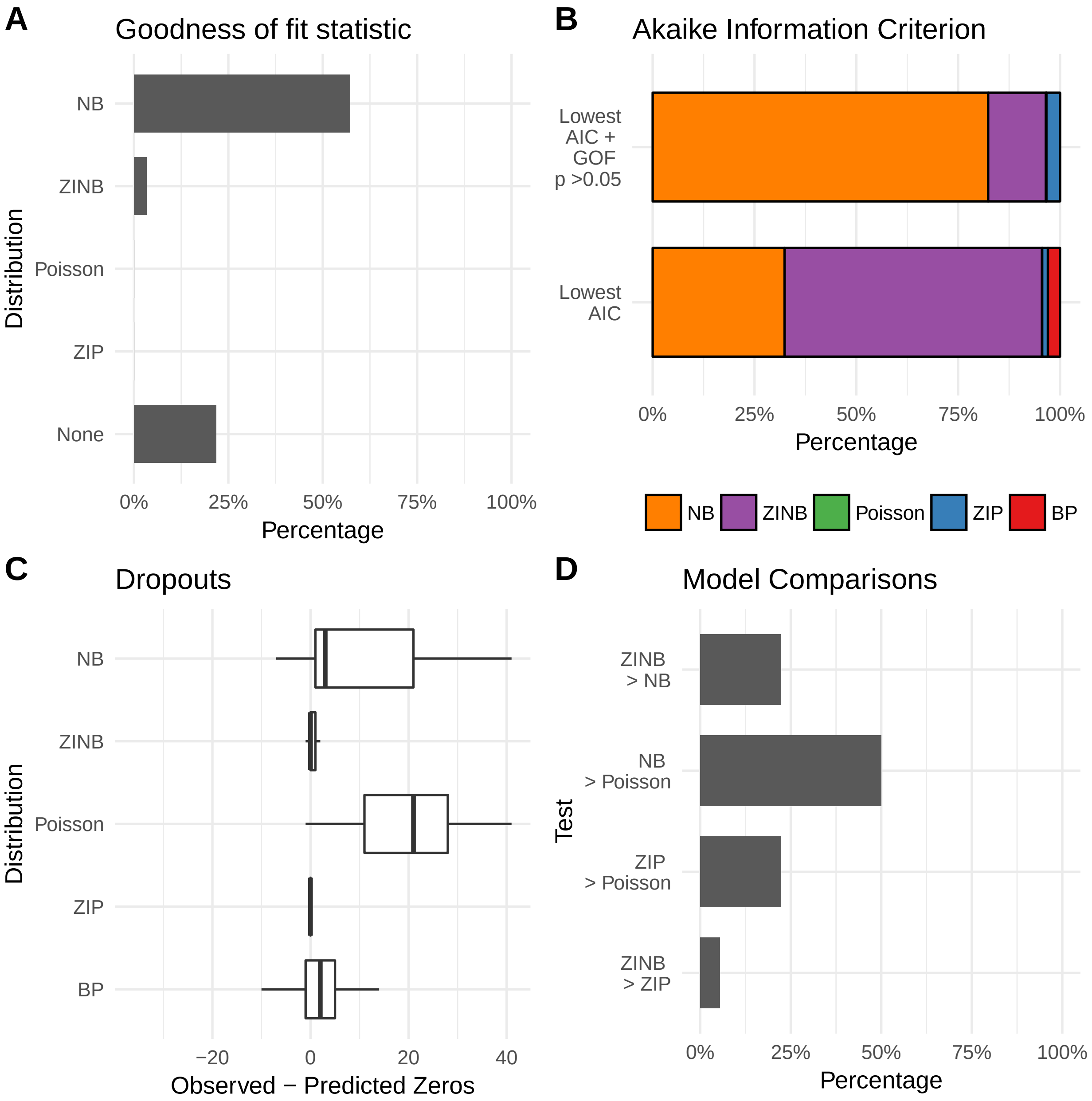

Figure S6: Islam et al. 2011: Mouse embryonic fibroblast (STRT). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

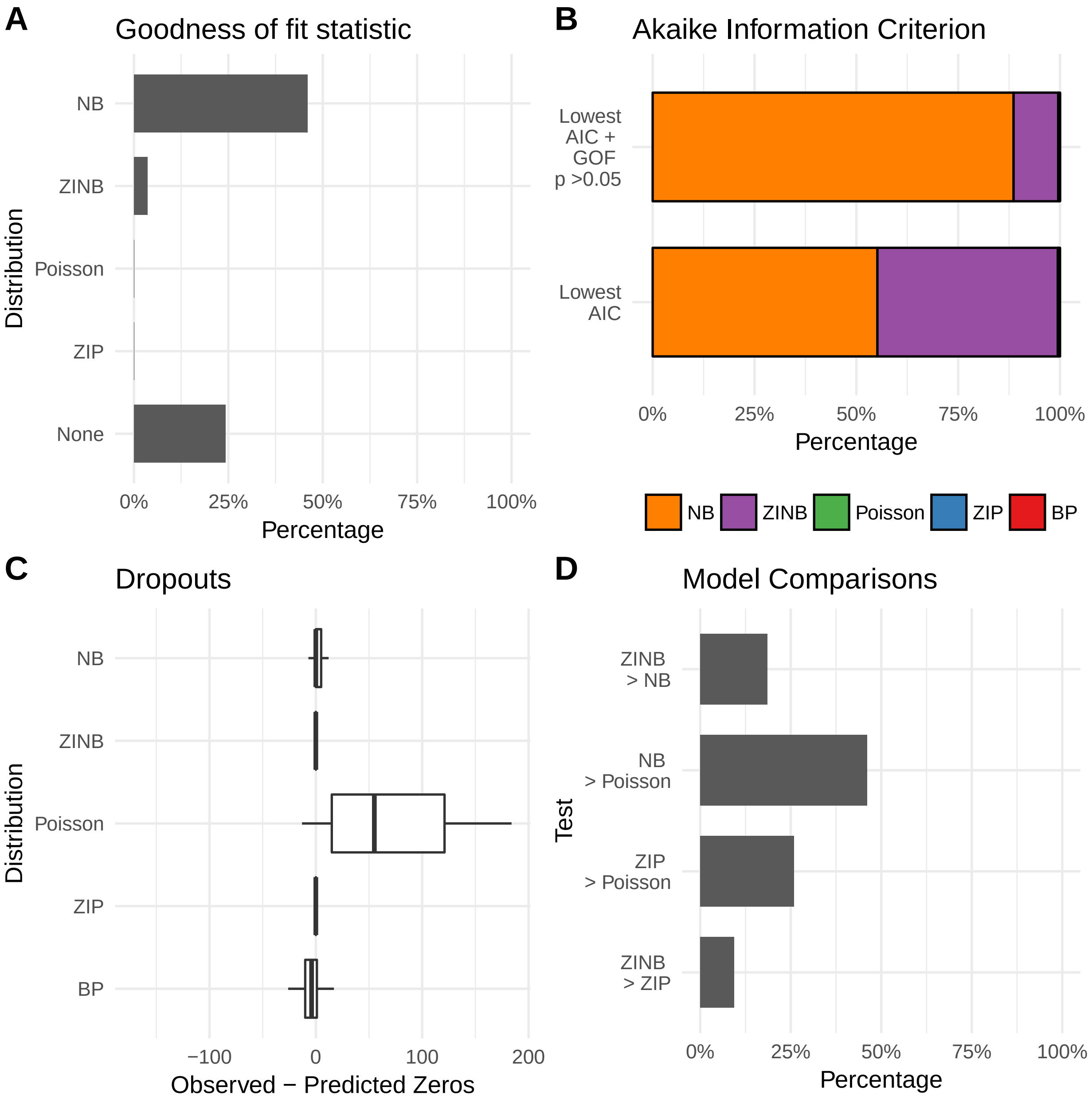

Figure S7: Kolodziejczk et al. 2015: Embryonic stem cells alternative 2i media + LIF (SmartSeq/C1). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

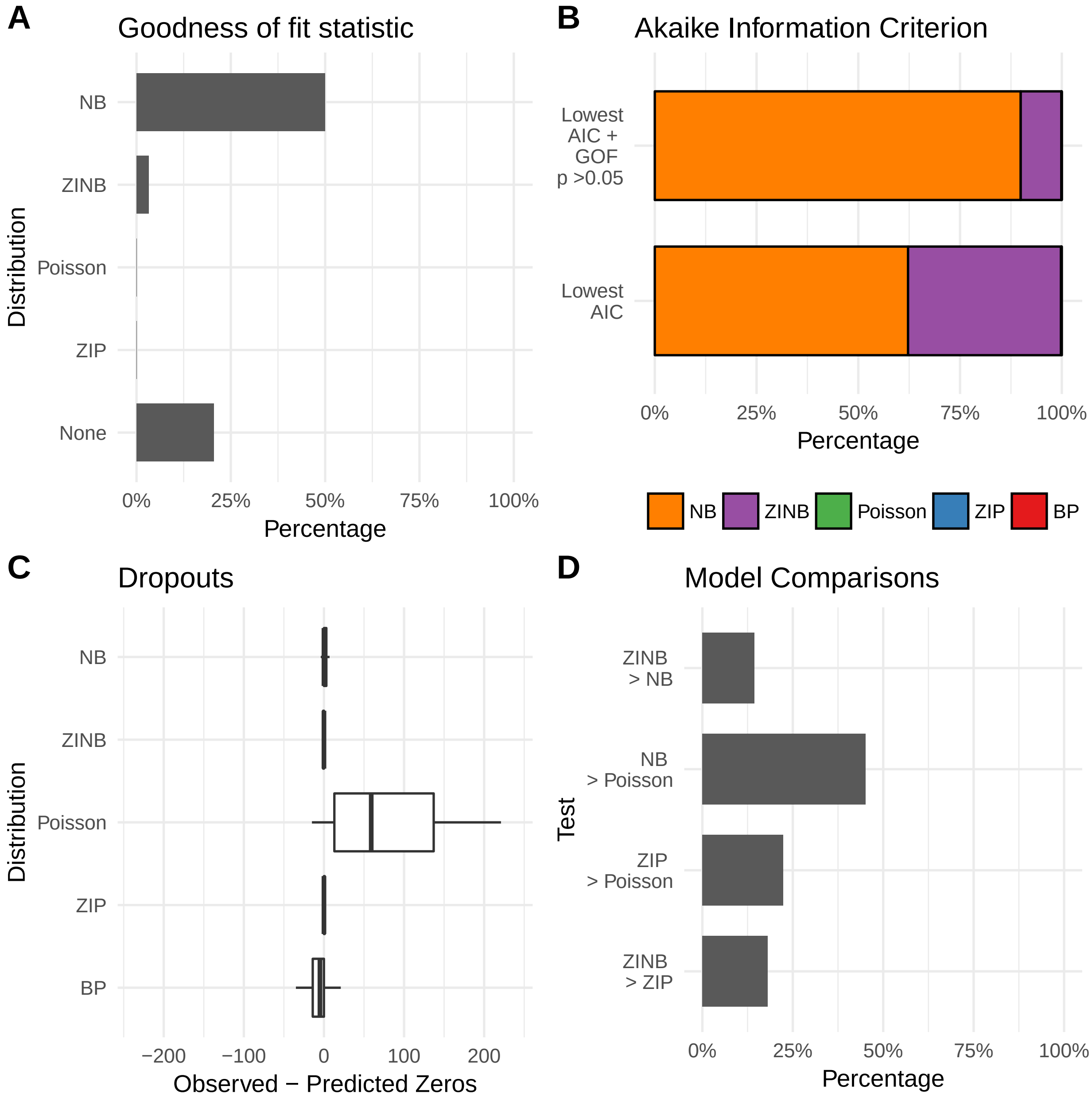

Figure S8: Kolodziejczk et al. 2015: Embryonic stem cells serum + LIF (SmartSeq/C1). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

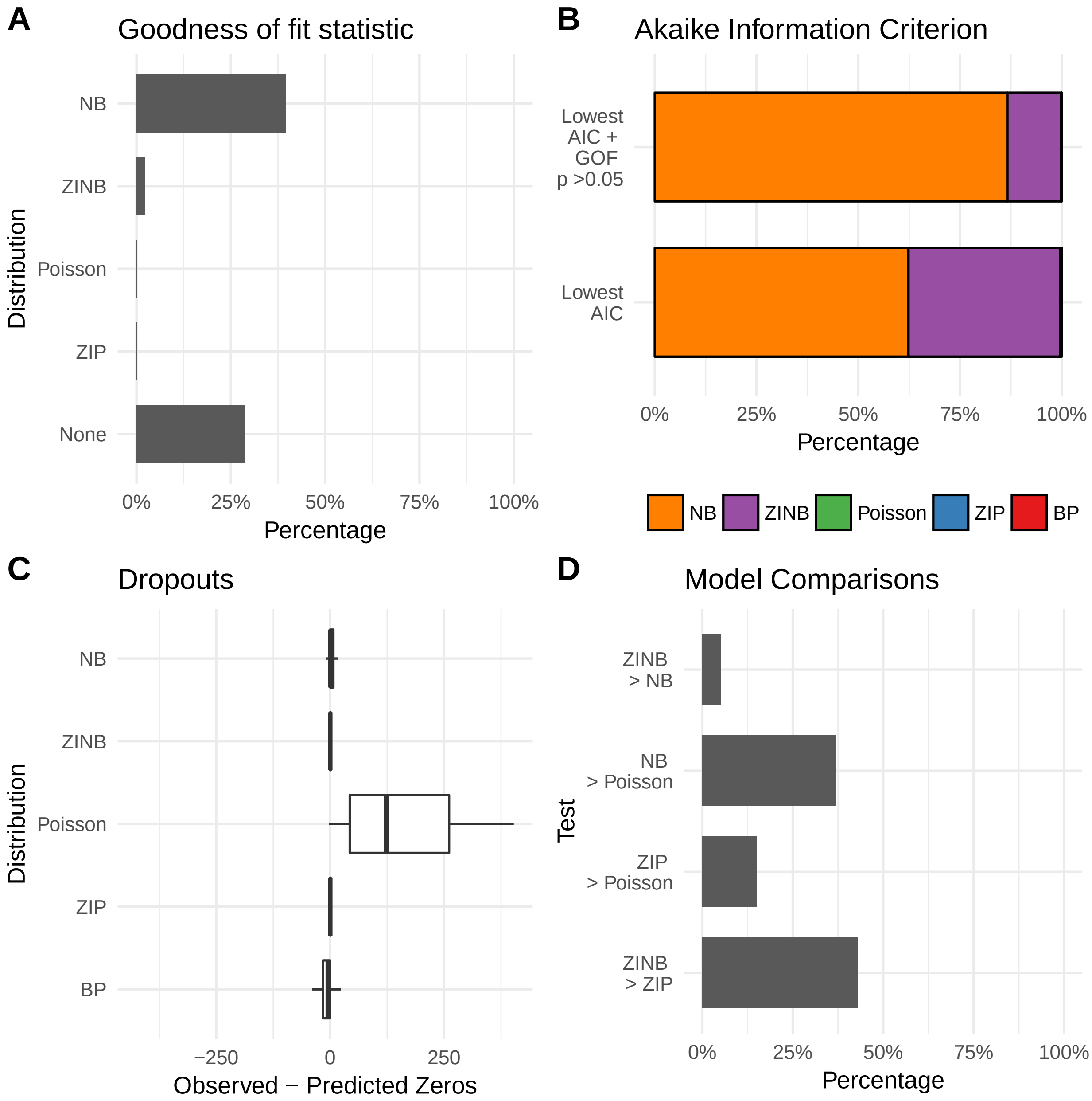

Figure S9: Kolodziejczk et al. 2015: Embryonic stem cells standard 2i media + LIF (Smart-seq/C1). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

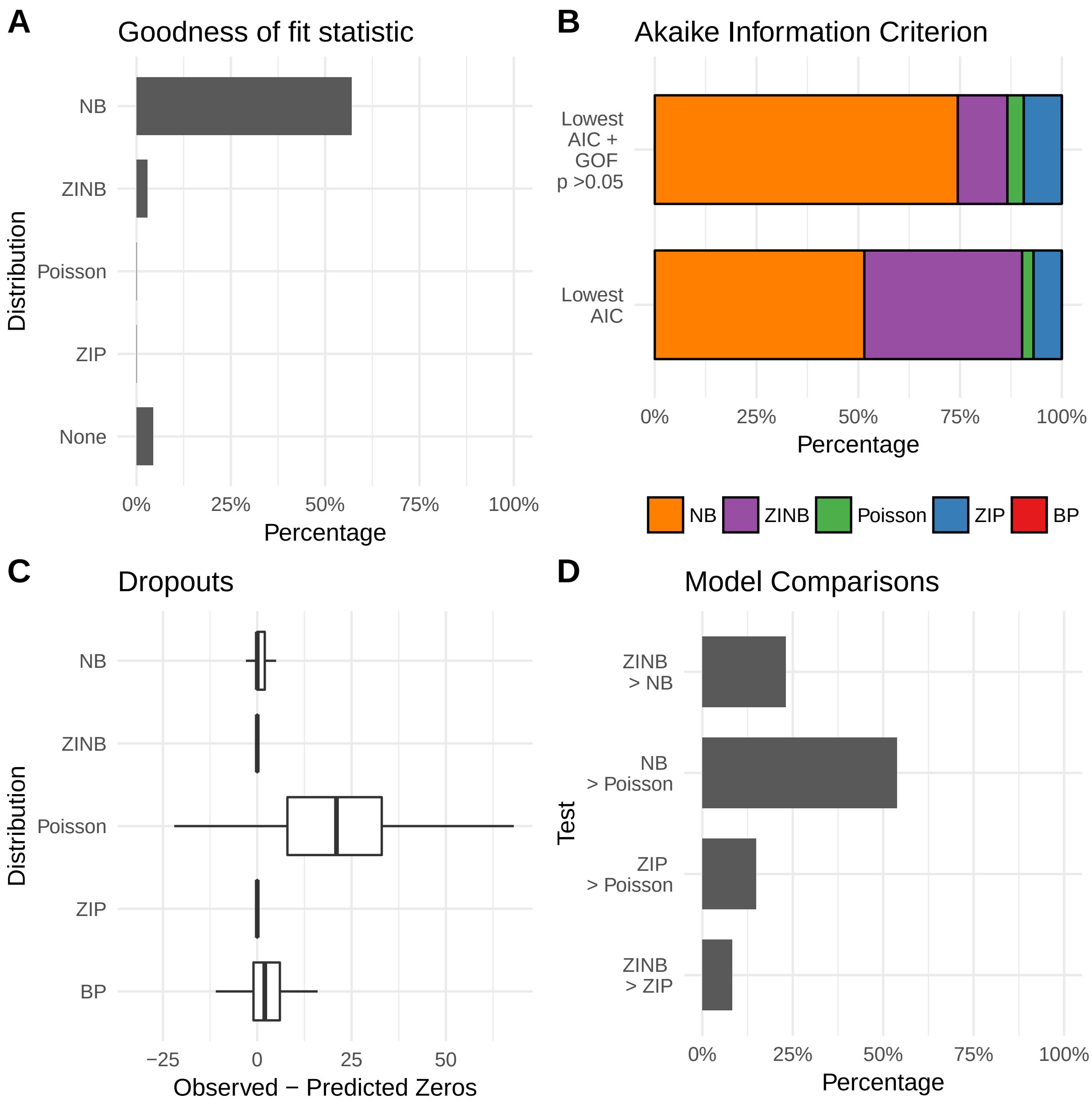

Figure S10: Pollen et al. 2014: Primary epidermal keratinocytes (SmartSeq/C1). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

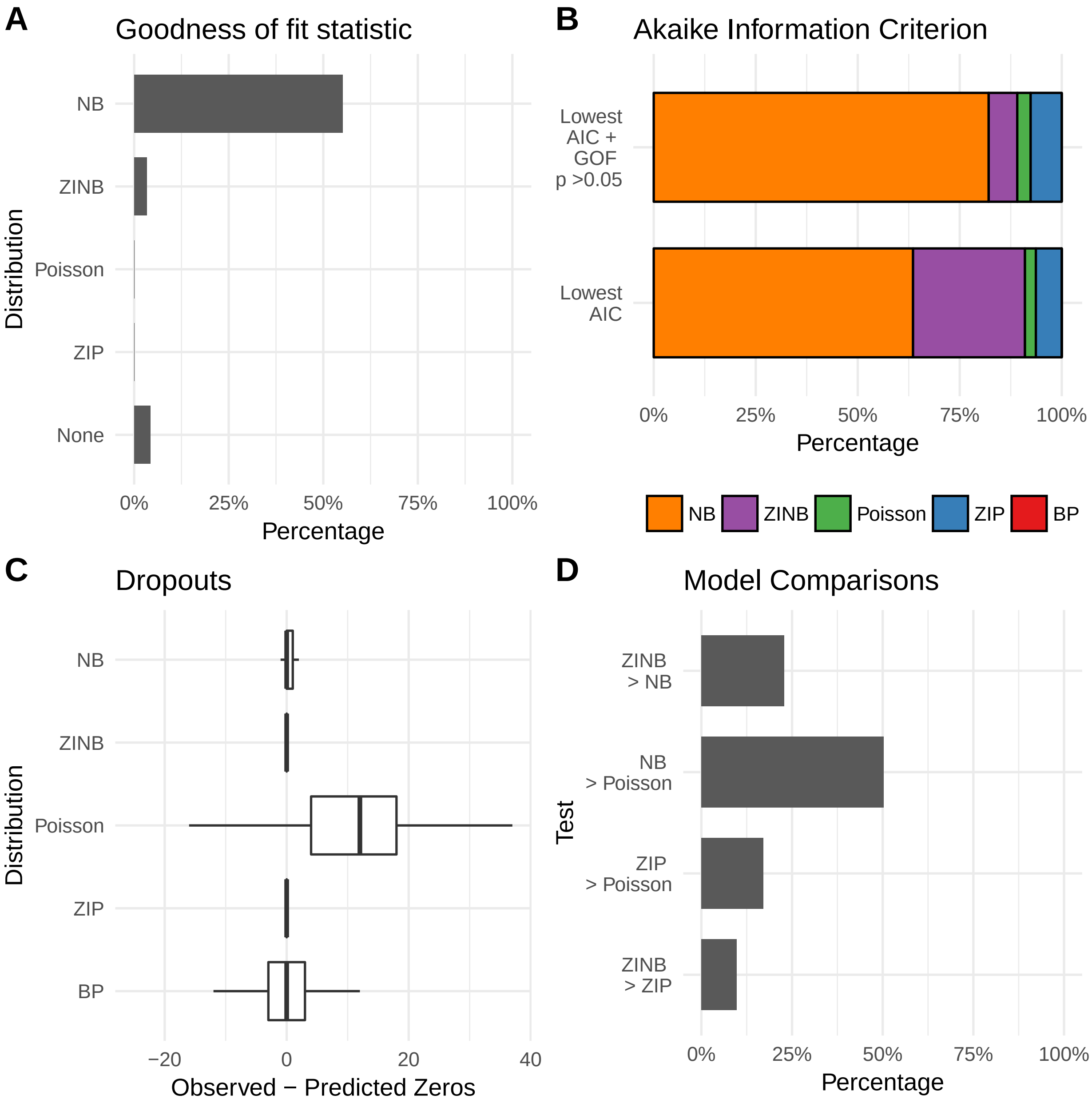

Figure S11: Pollen et al 2014: Induced pluripotent stem cells (SmartSeq/C1). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

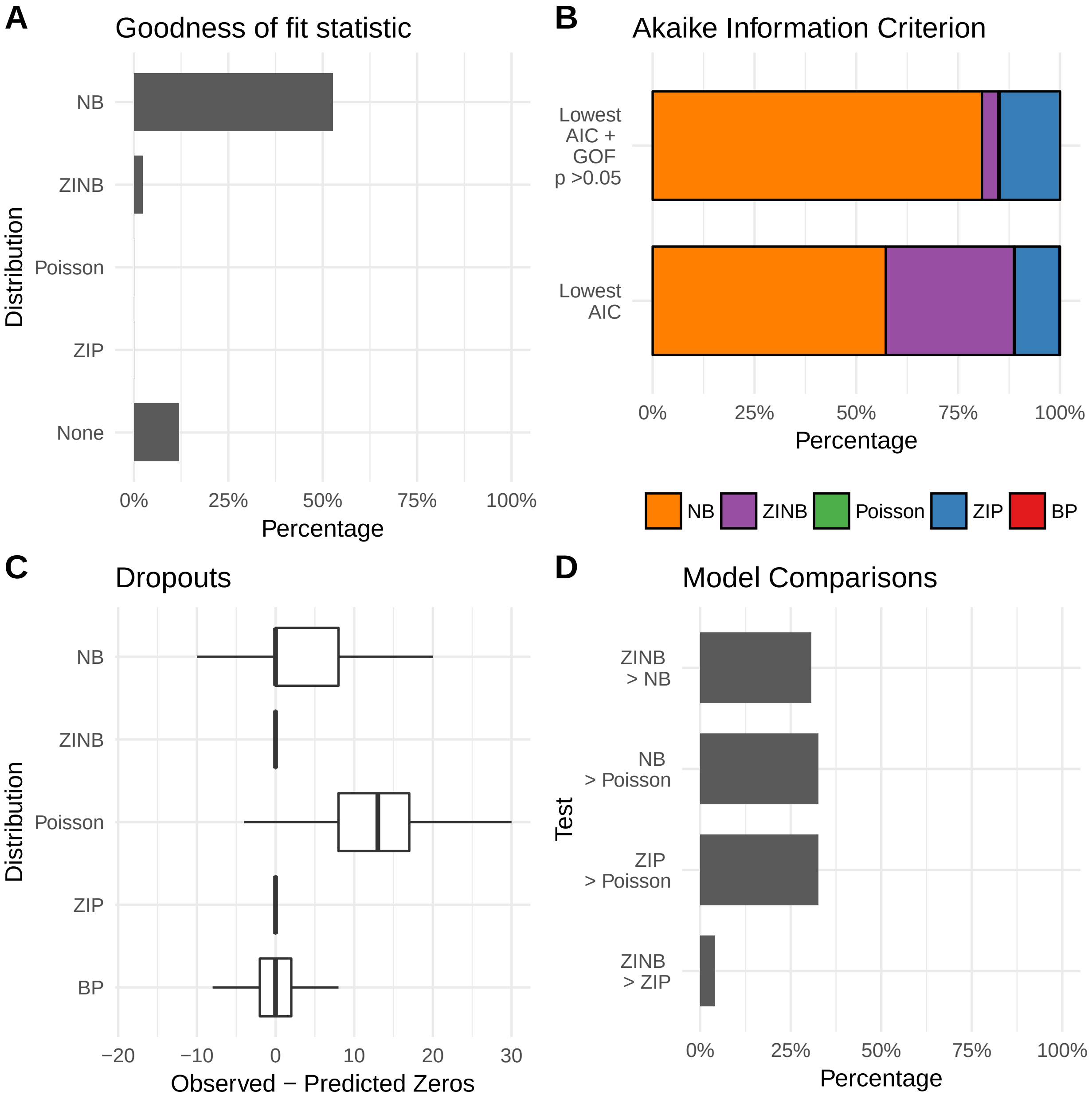

Figure S12: Pollen et al. 2014: Cultured primary human neurons (SmartSeq/C1). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

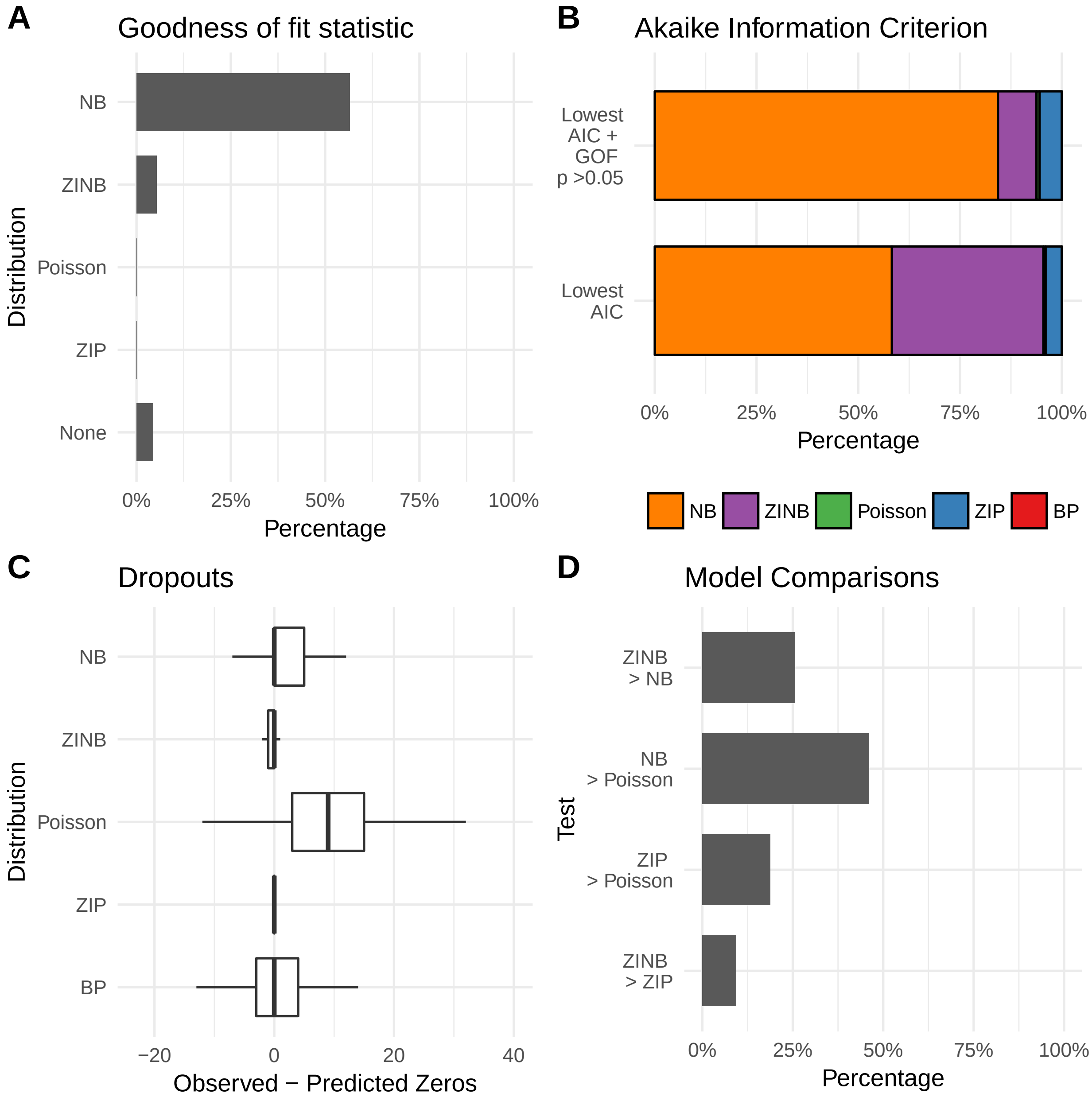

Figure S13: Pollen et al. 2014: HCC1954 breast cancer cells (SmartSeq/C1). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

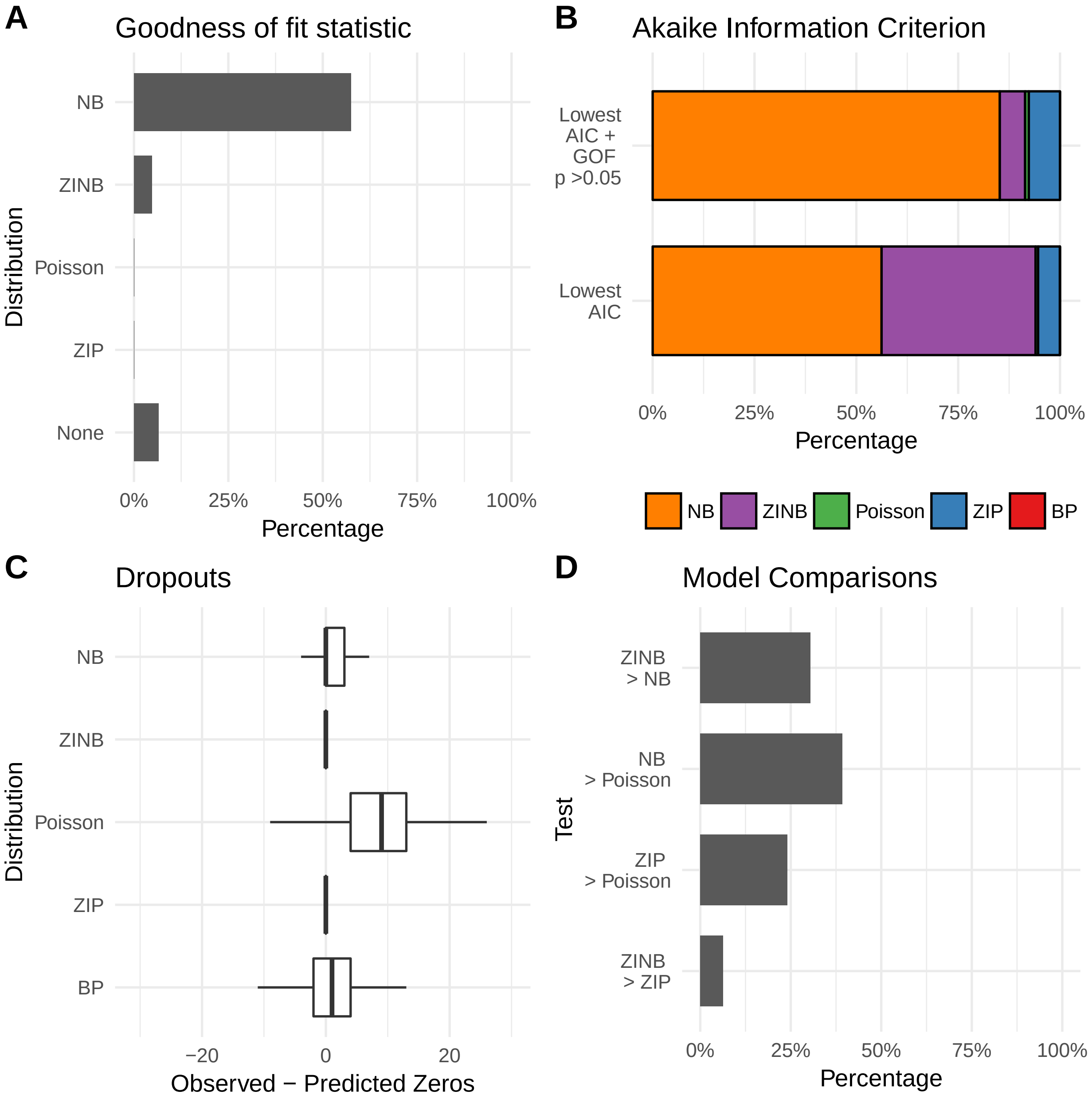

Figure S14: Pollen et al. 2014: HCC1954 B lymphoblastoid cells (SmartSeq/C1). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

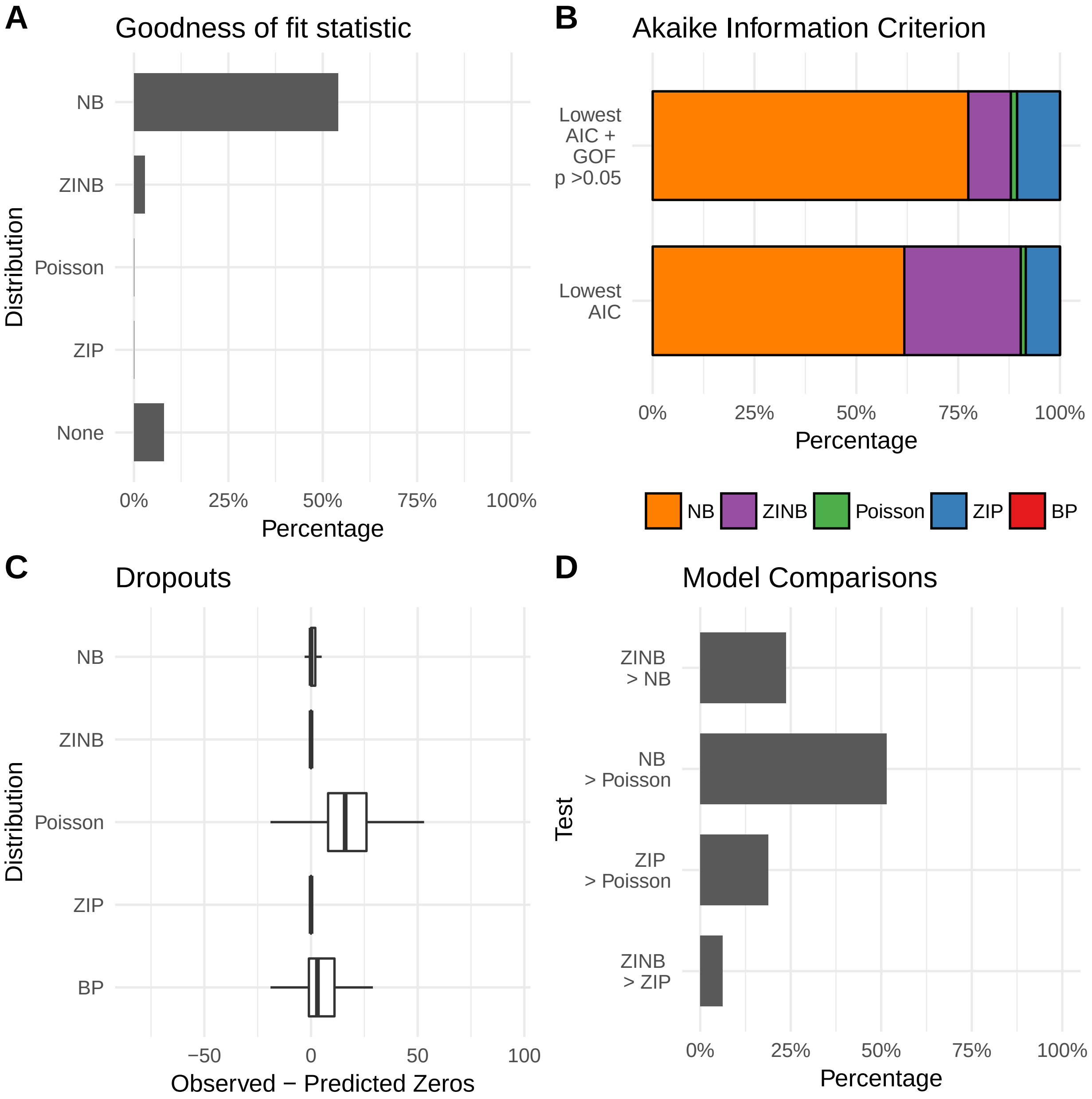

Figure S15: Pollen et al. 2014: HL-60 human promyelocytic leukemia cells (SmartSeq/C1). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

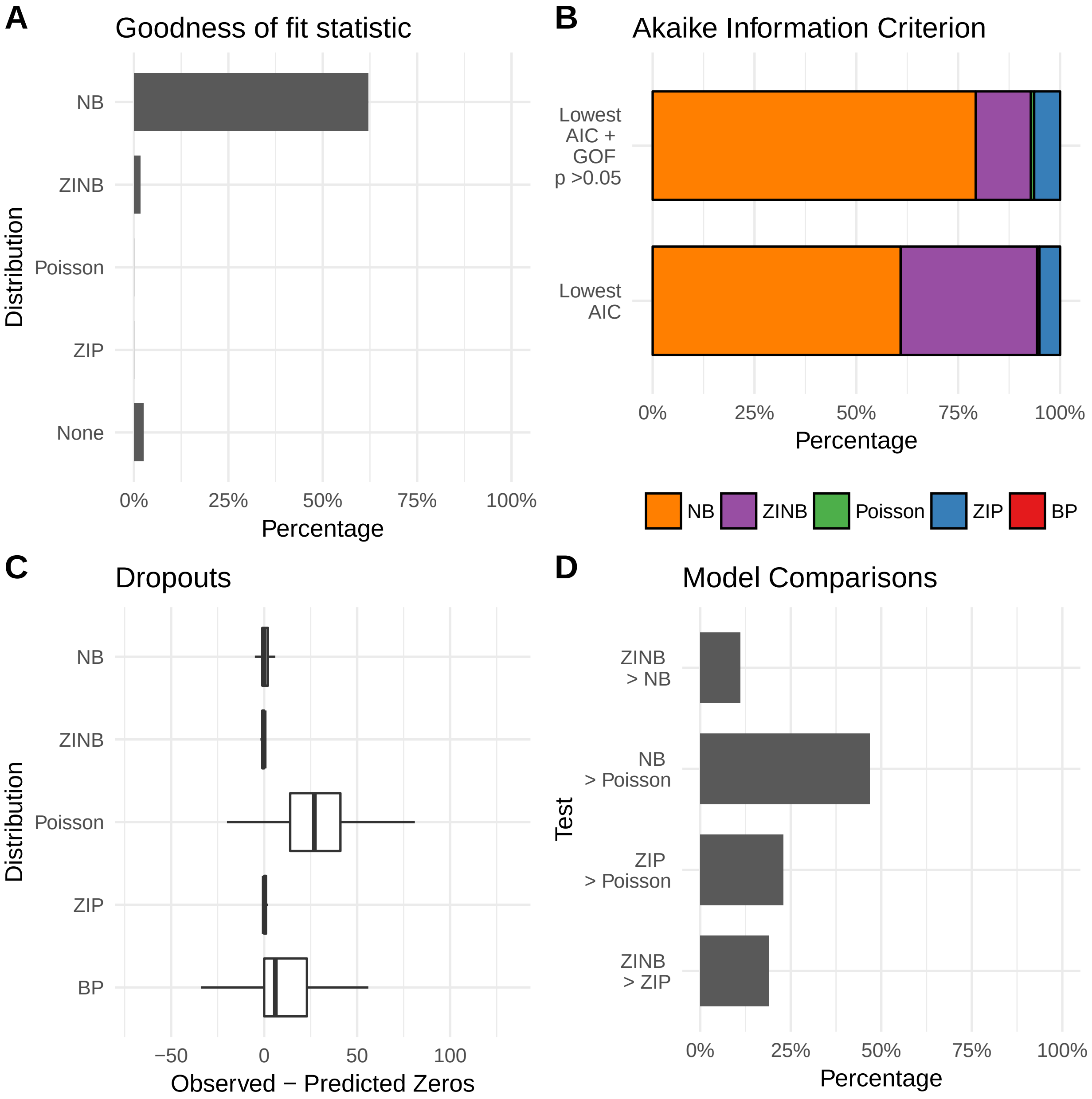

Figure S16: Pollen et al. 2014: K-562 myelogenous leukemia cells (SmartSeq/C1). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

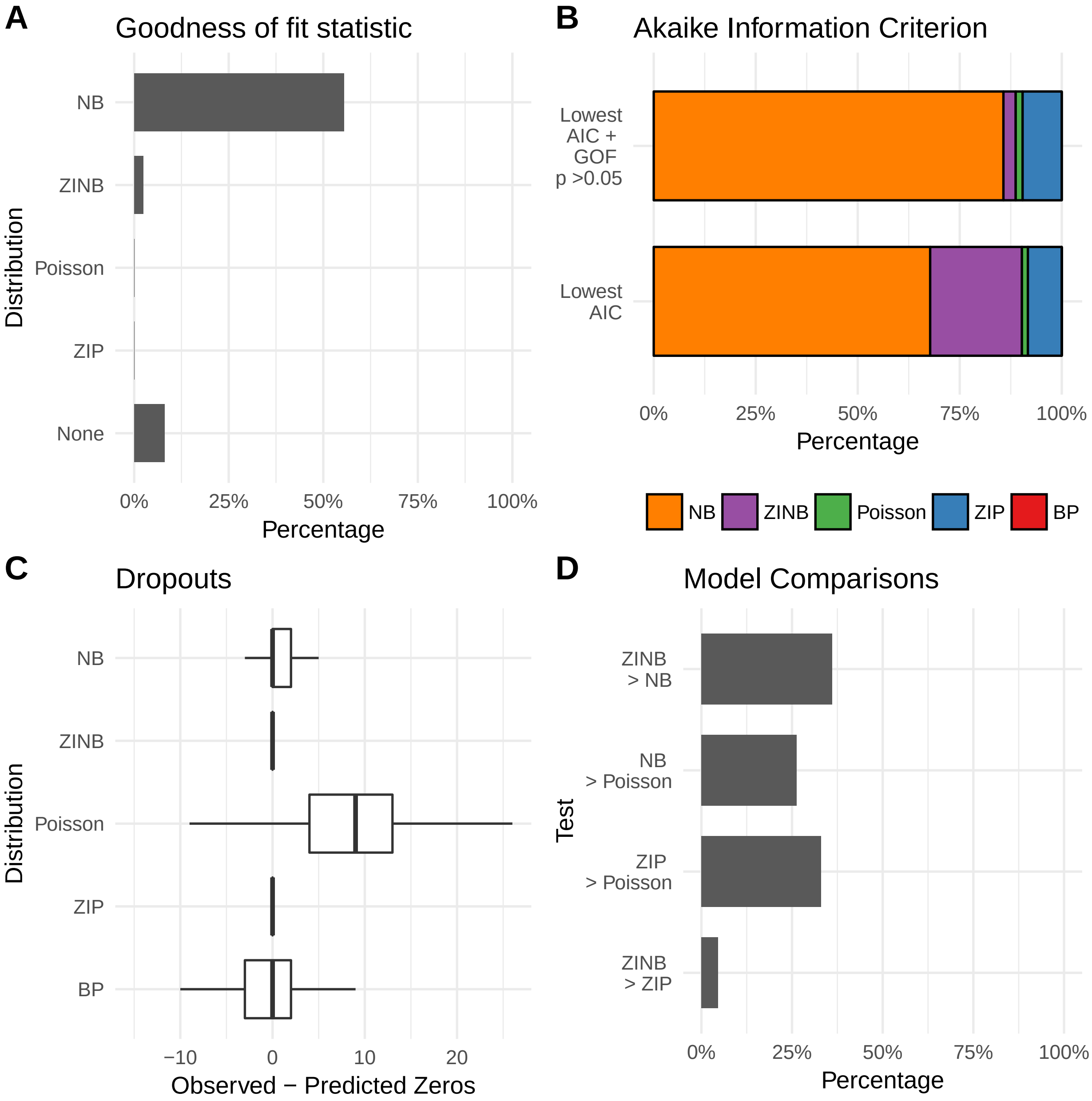

Figure S17: Pollen et al. 2014: Neural progenitor cells obtained by differentiation of iPS line (Smart-Seq/C1). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

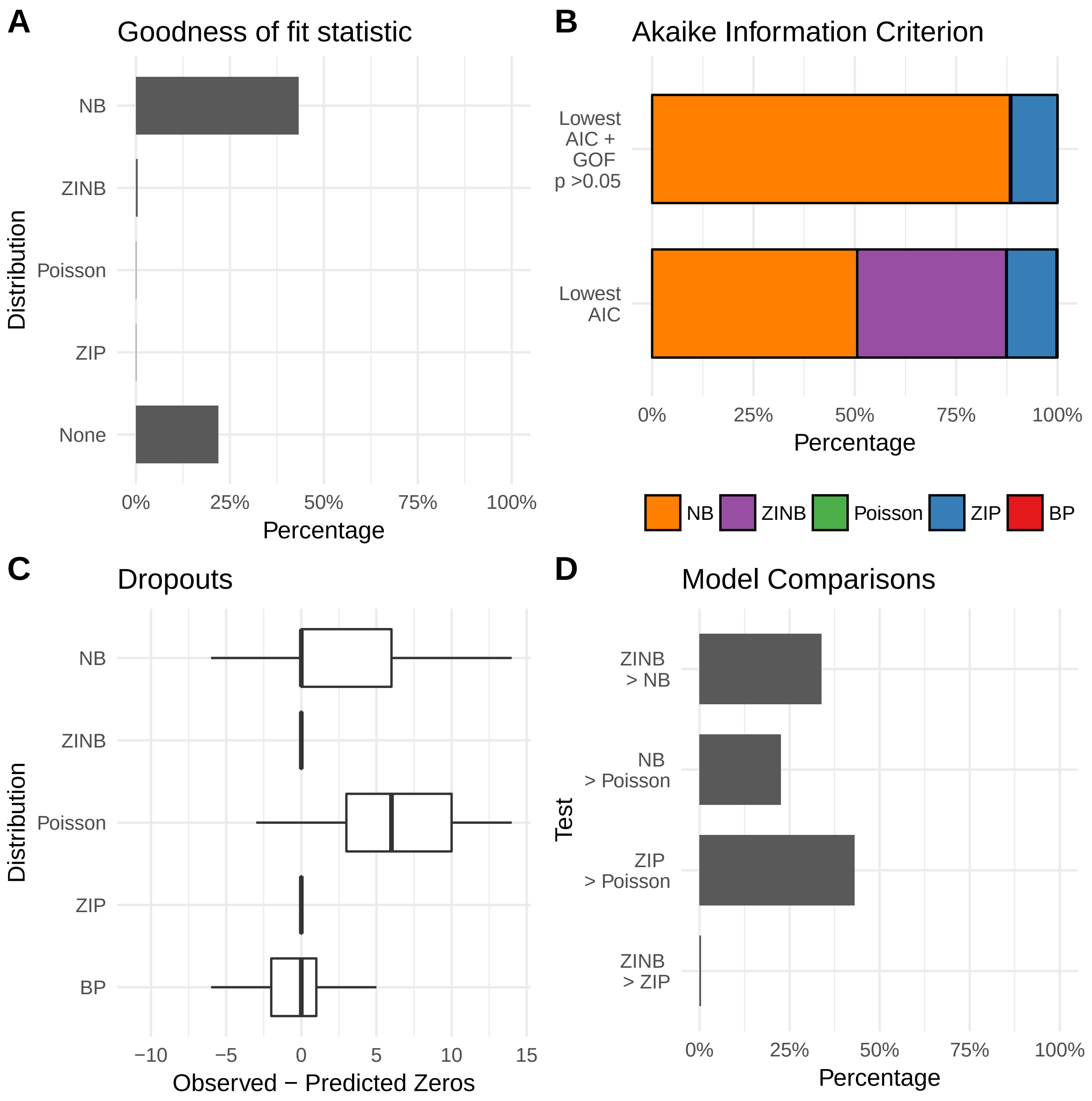

Figure S18: Pollen et al. 2014: Primary human neurons (SmartSeq/C1). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

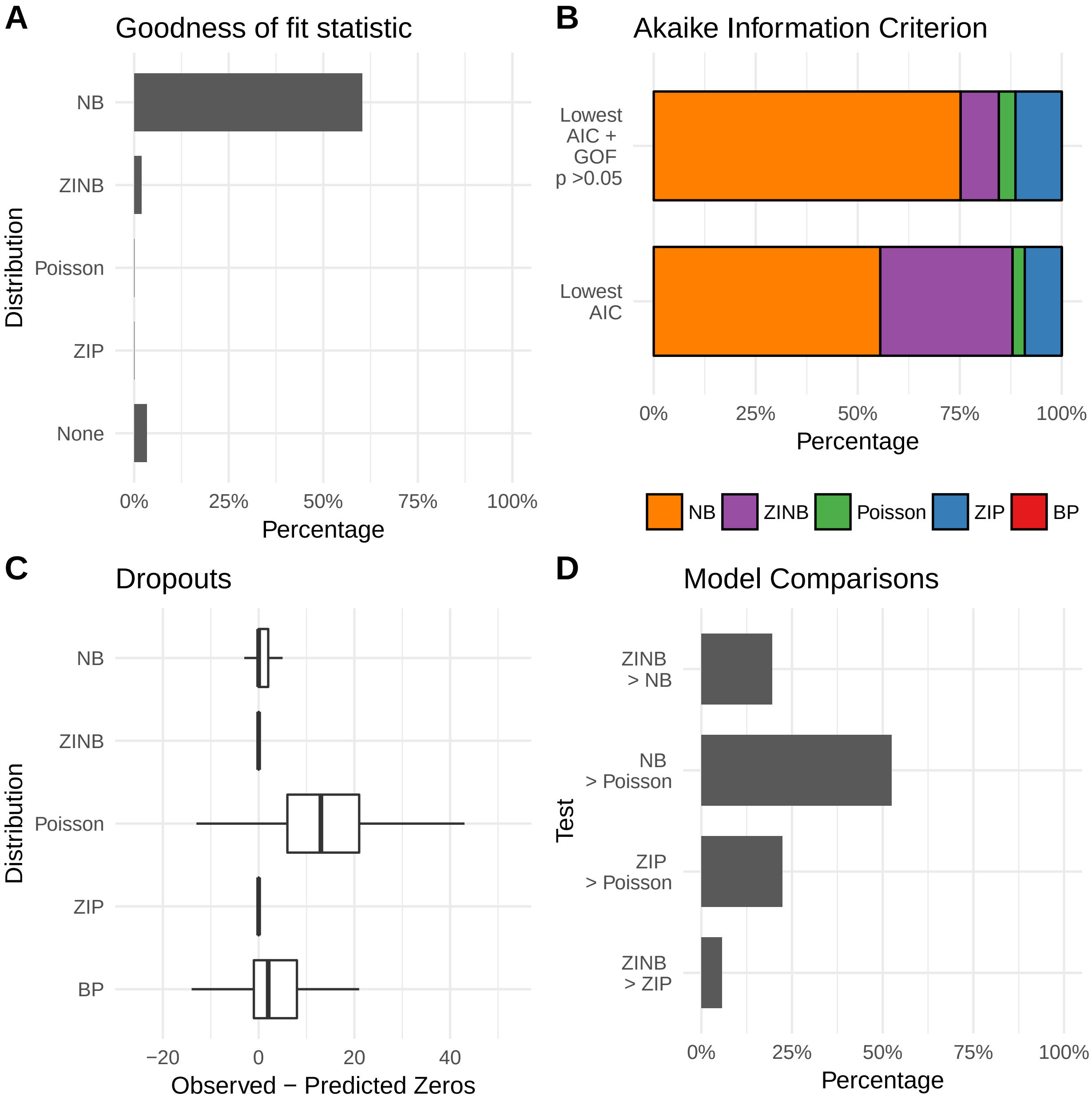

Figure S19: Pollen et al. 2014: BJ Human Fibroblasts early passage, p6 (Smart-seq/C1). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

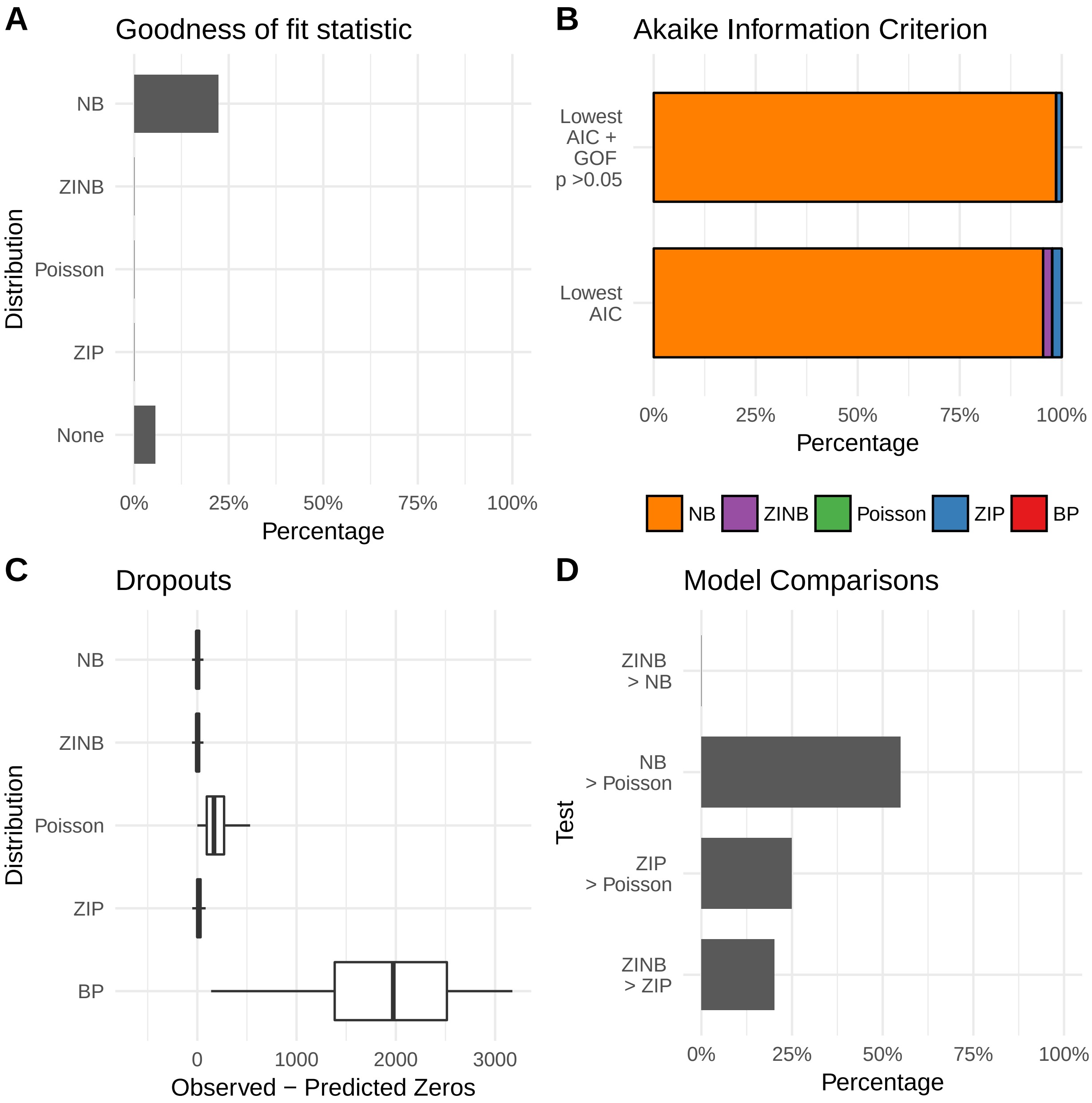

Figure S20: Soumillon et al. 2014: adipose-derived stem cells 1 day post-differentiation (SCRB-seq). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

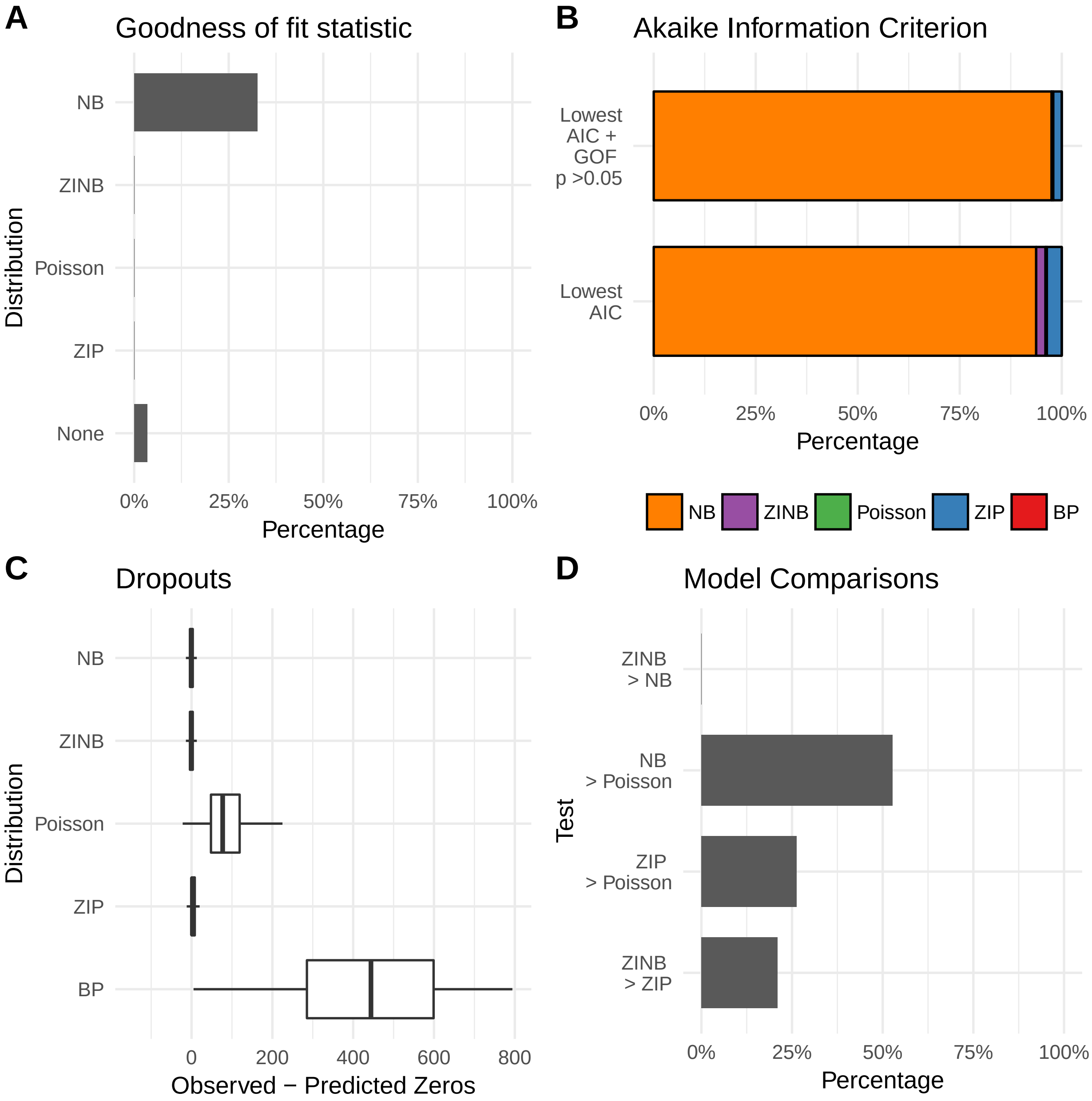

Figure S21: Soumillon et al. 2014: adipose-derived stem cells 2 days post-differentiation (SCRB-seq). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

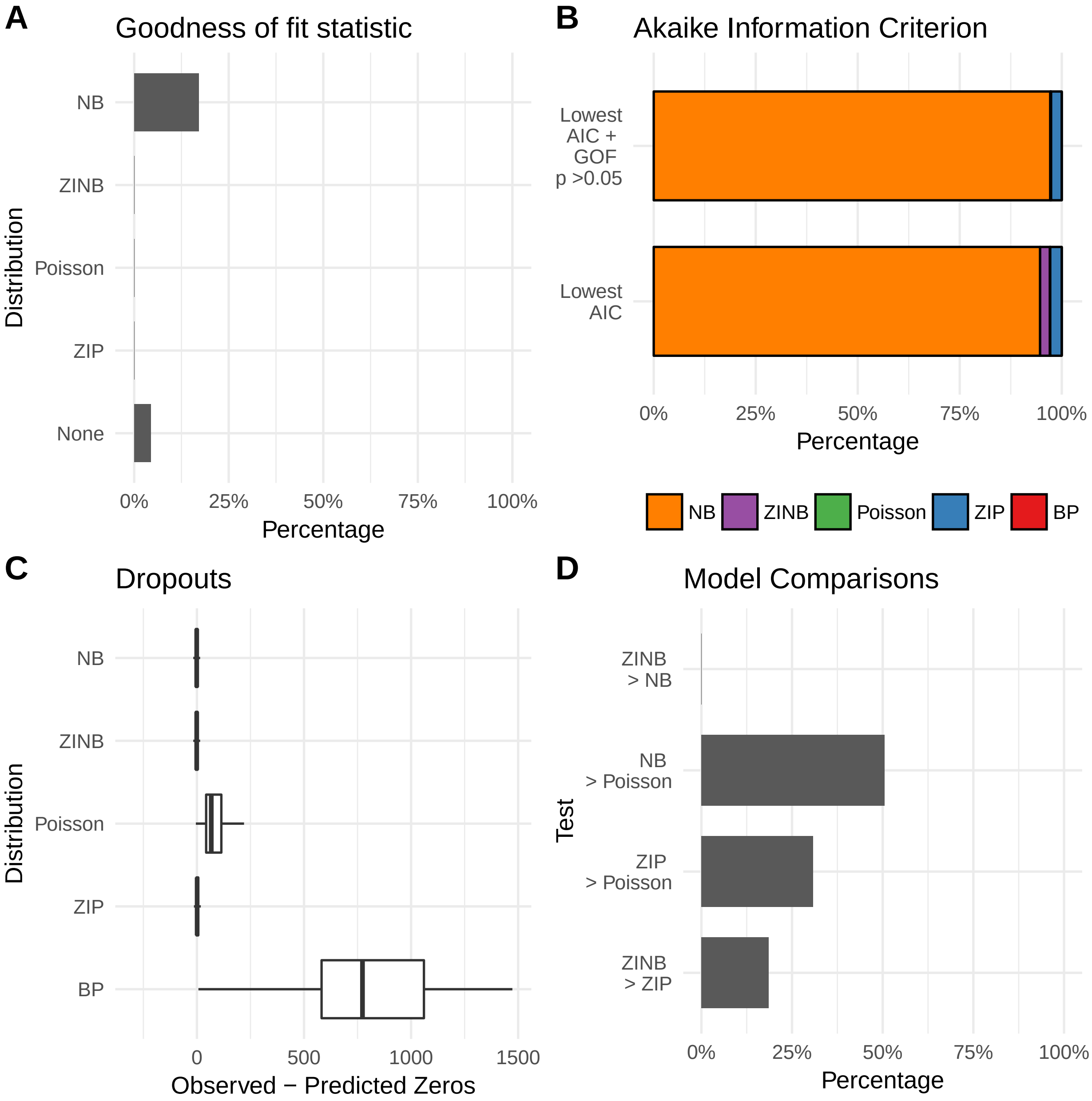

Figure S22: Soumillon et al. 2014: adipose-derived stem cells 3 days post-differentiation (SCRB-seq). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

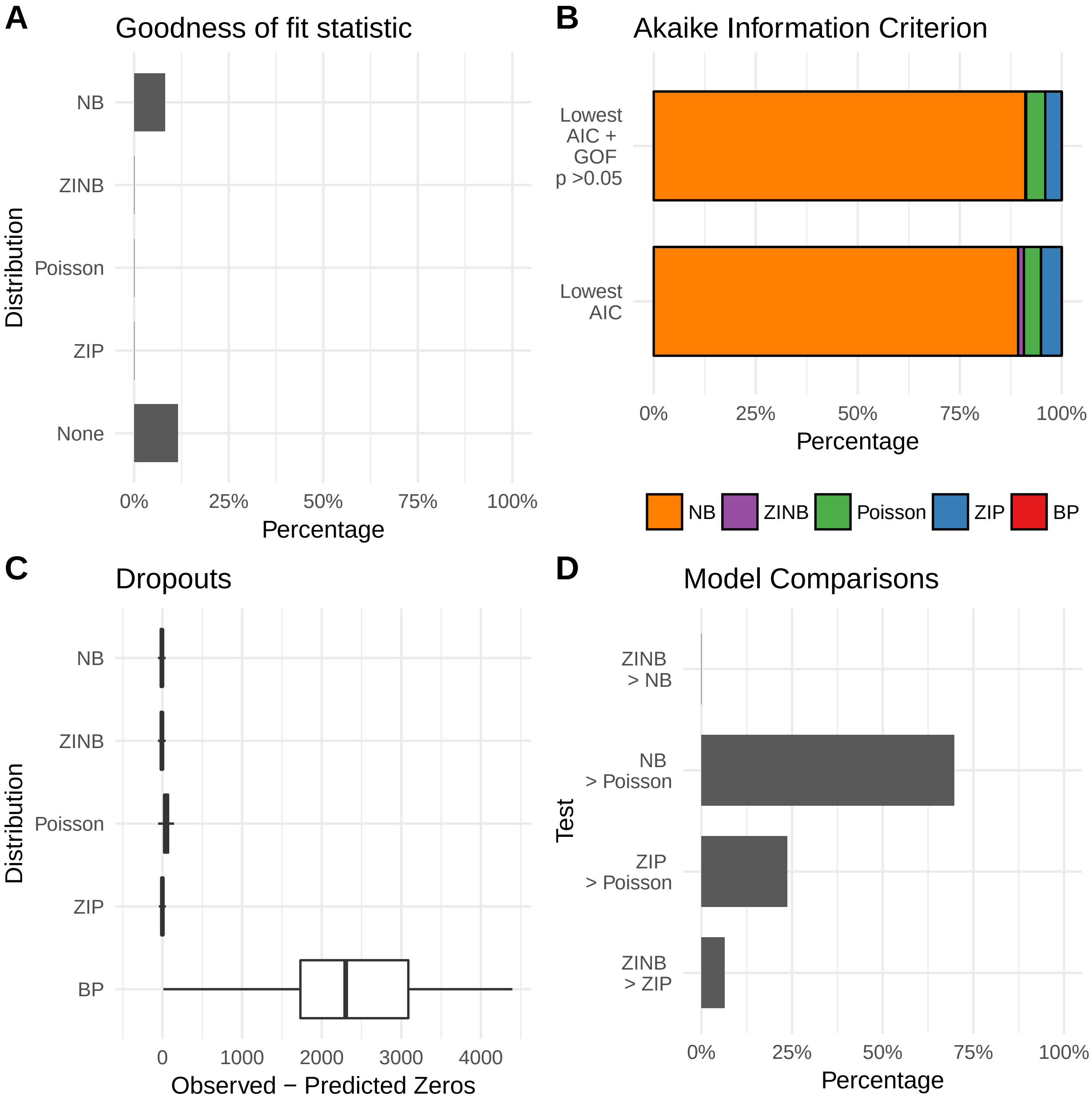

Figure S23: Zheng et al. 2017: Peripheral Blood Mononuclear Cells CD19+ B Cells (10XGenomics). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

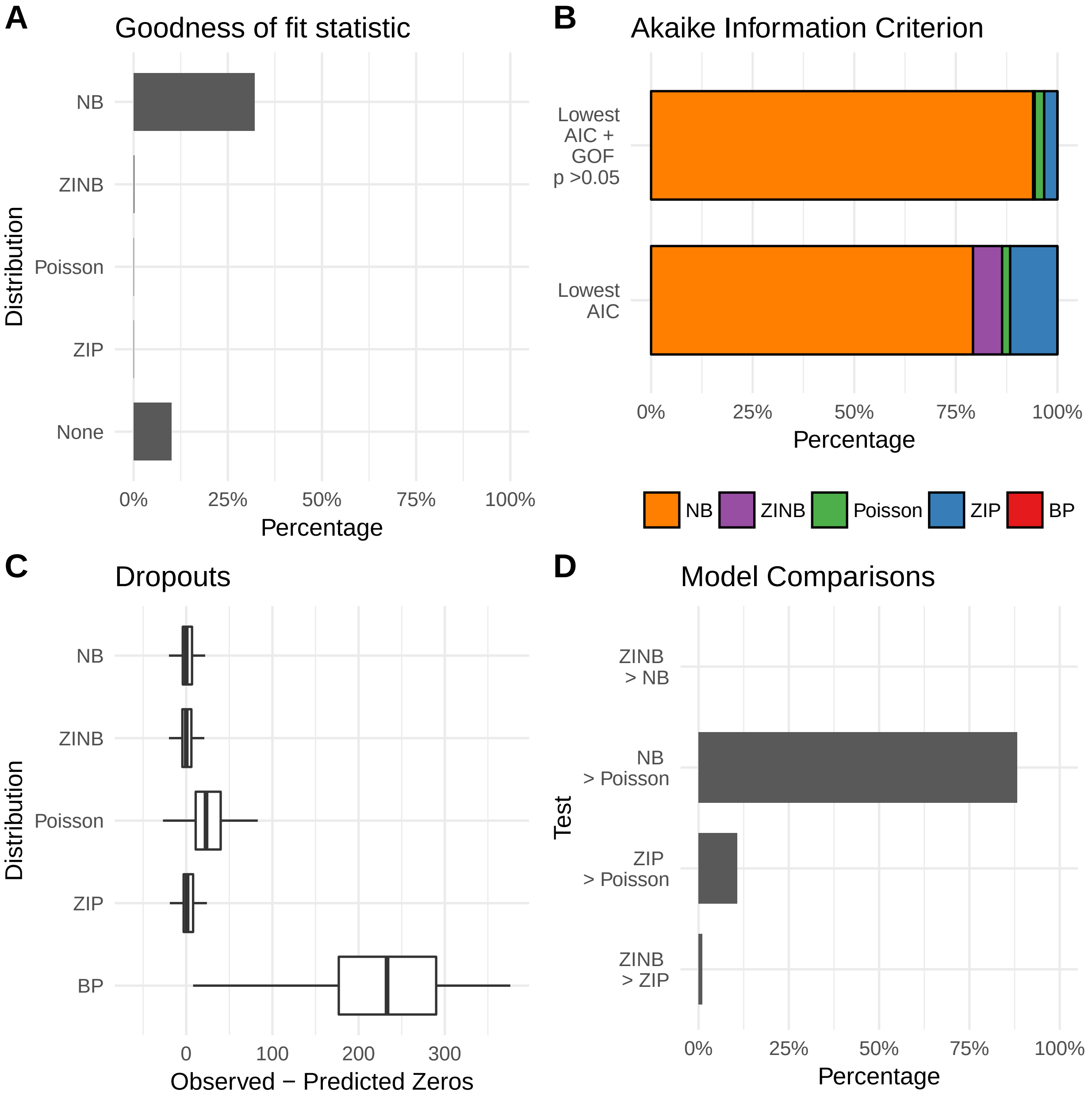

Figure S24: Zheng et al. 2017: Peripheral Blood Mononuclear Cells CD14+ Monocytes (10XGenomics). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

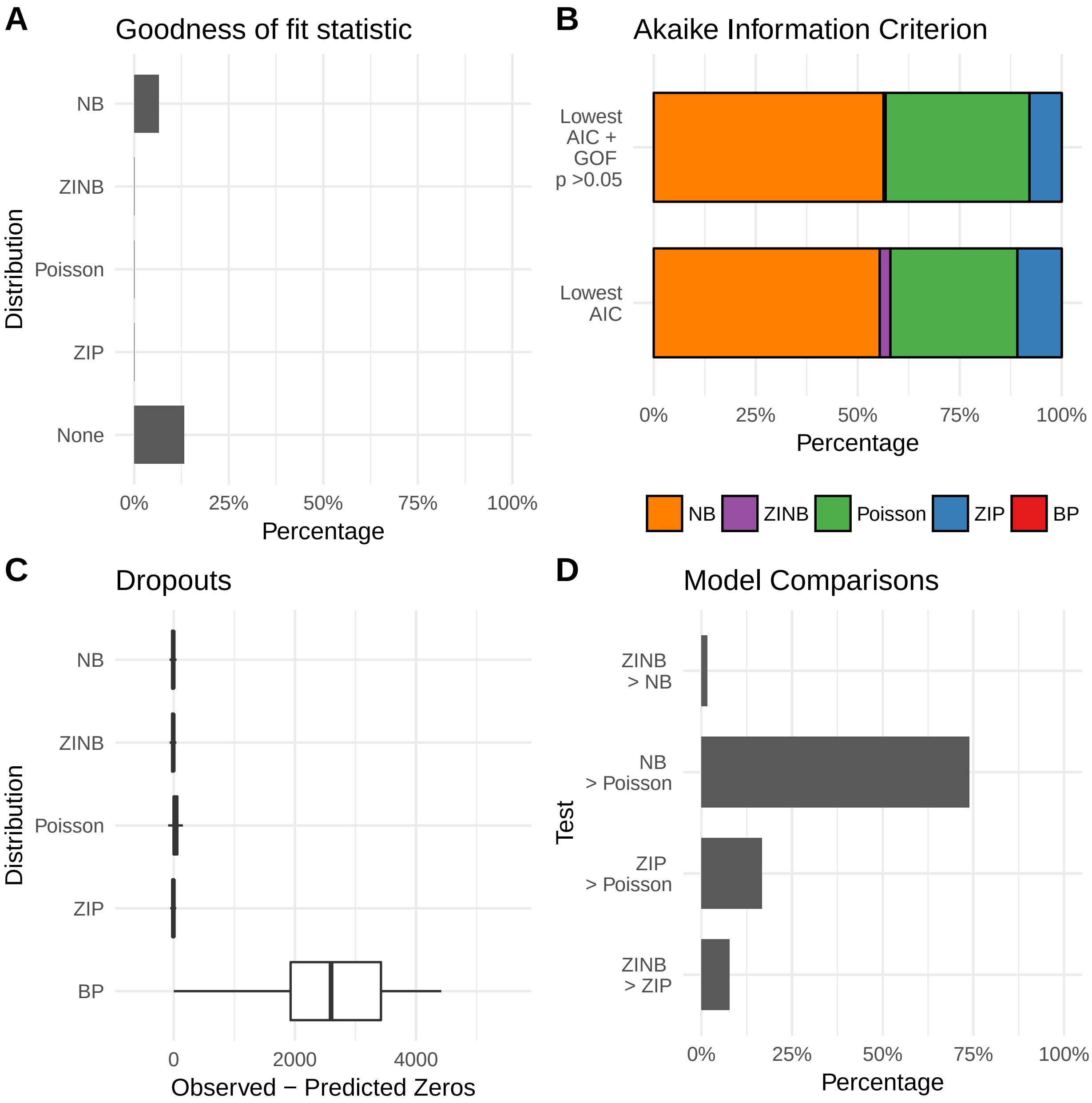

Figure S25: Zheng et al. 2017: Peripheral Blood Mononuclear Cells CD34+ Cells (10XGenomics). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

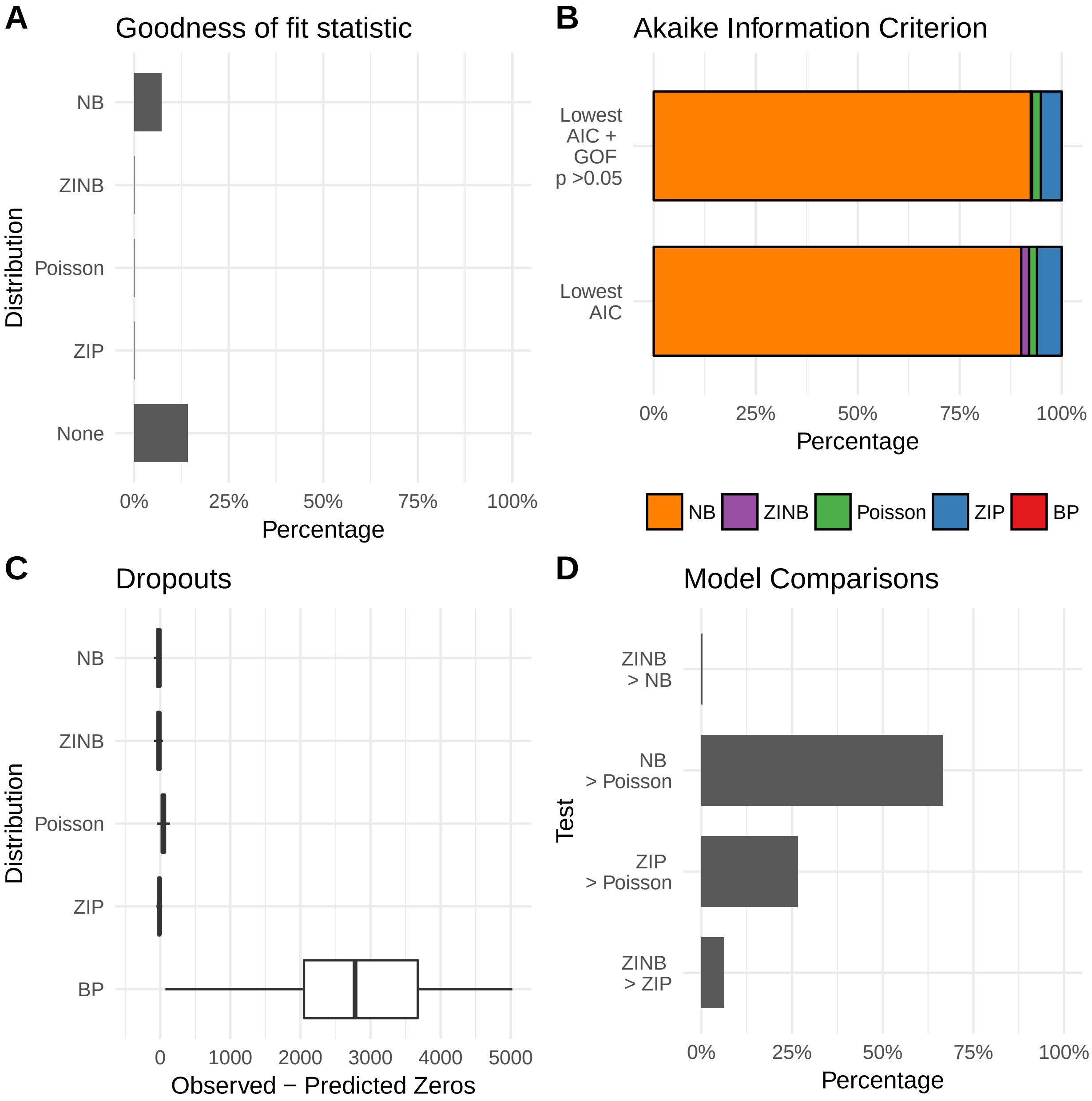

Figure S26: Zheng et al. 2017: Peripheral Blood Mononuclear Cells CD4+ T Helper Cells (10XGenomics). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

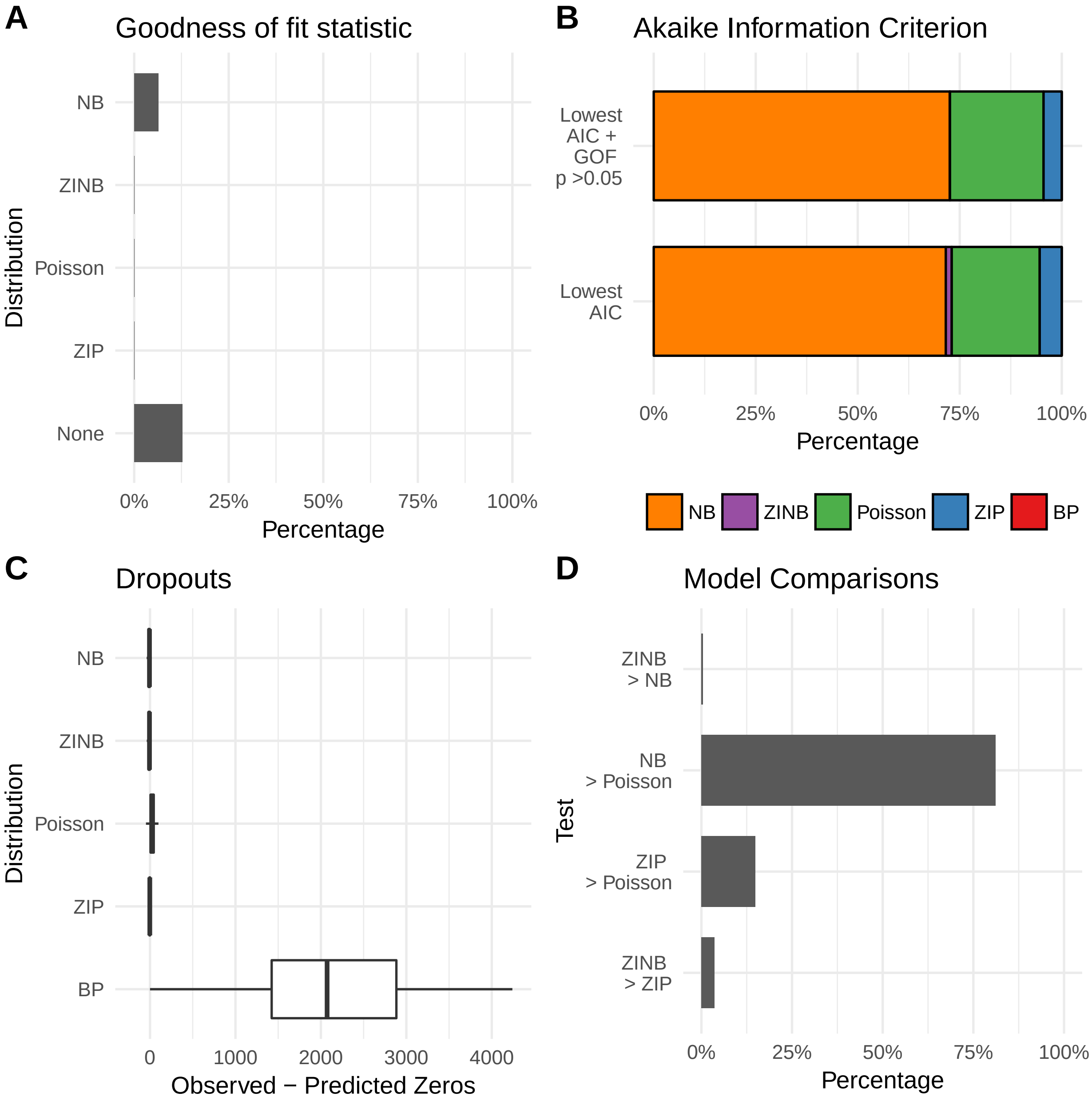

Figure S27: Zheng et al. 2017: Peripheral Blood Mononuclear Cells CD56+ NK Cells (10XGenomics). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

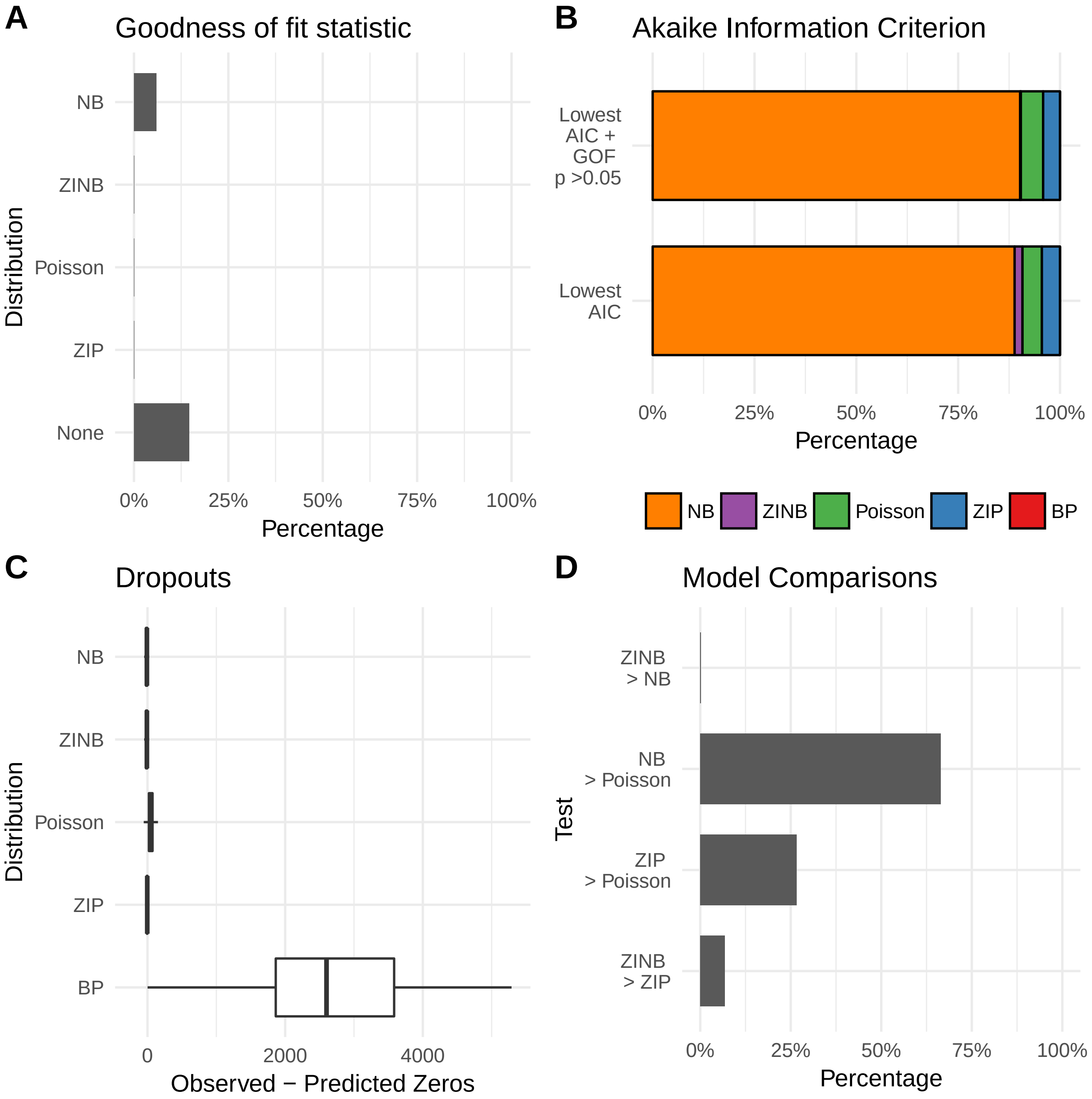

Figure S28: Zheng et al. 2017: Peripheral Blood Mononuclear Cells CD8+ Cytotoxic T Cells (10XGenomics). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

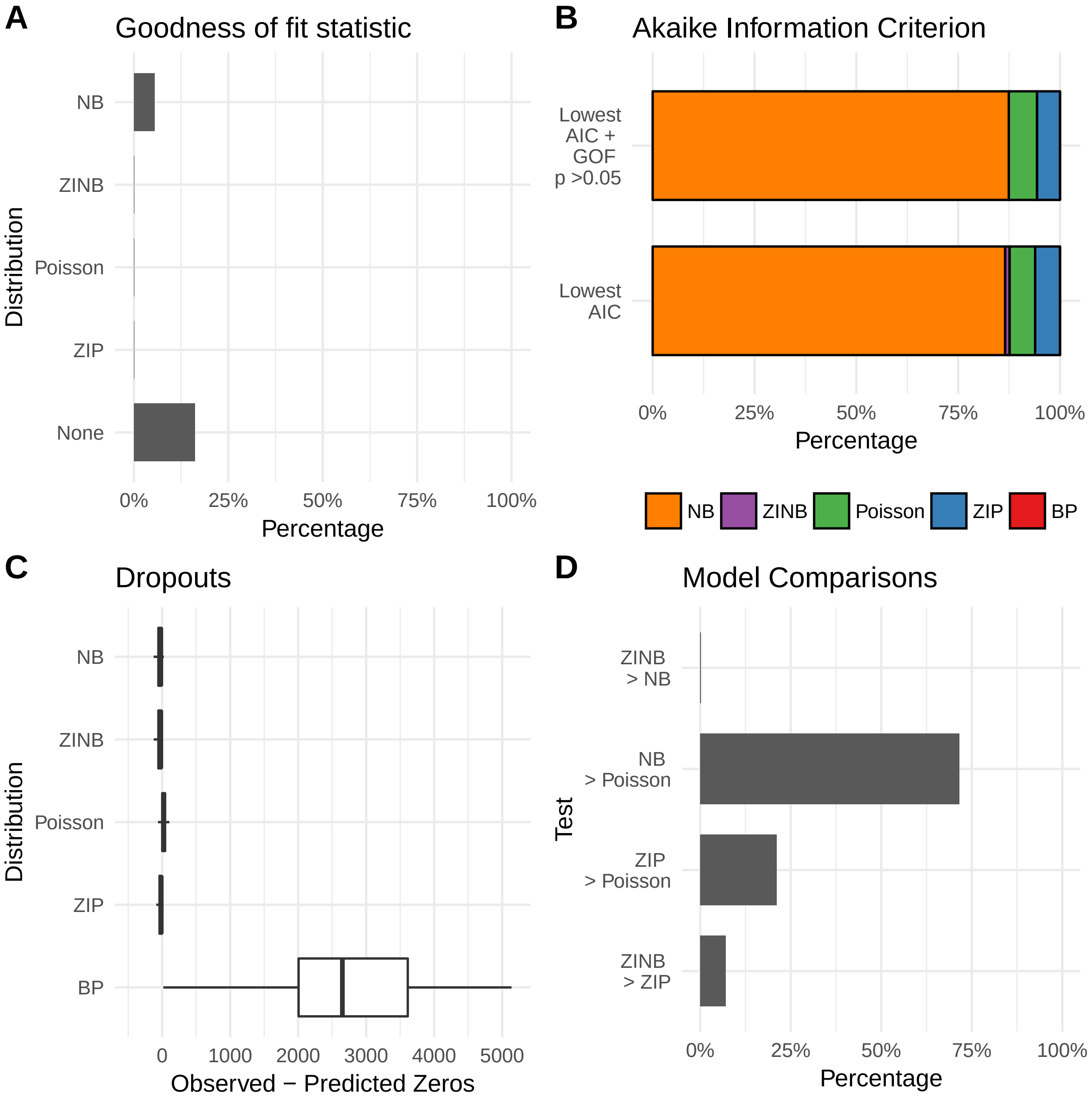

Figure S29: Zheng et al. 2017: Peripheral Blood Mononuclear Cells CD4+/CD45RO+ Memory T Cells (10XGenomics). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

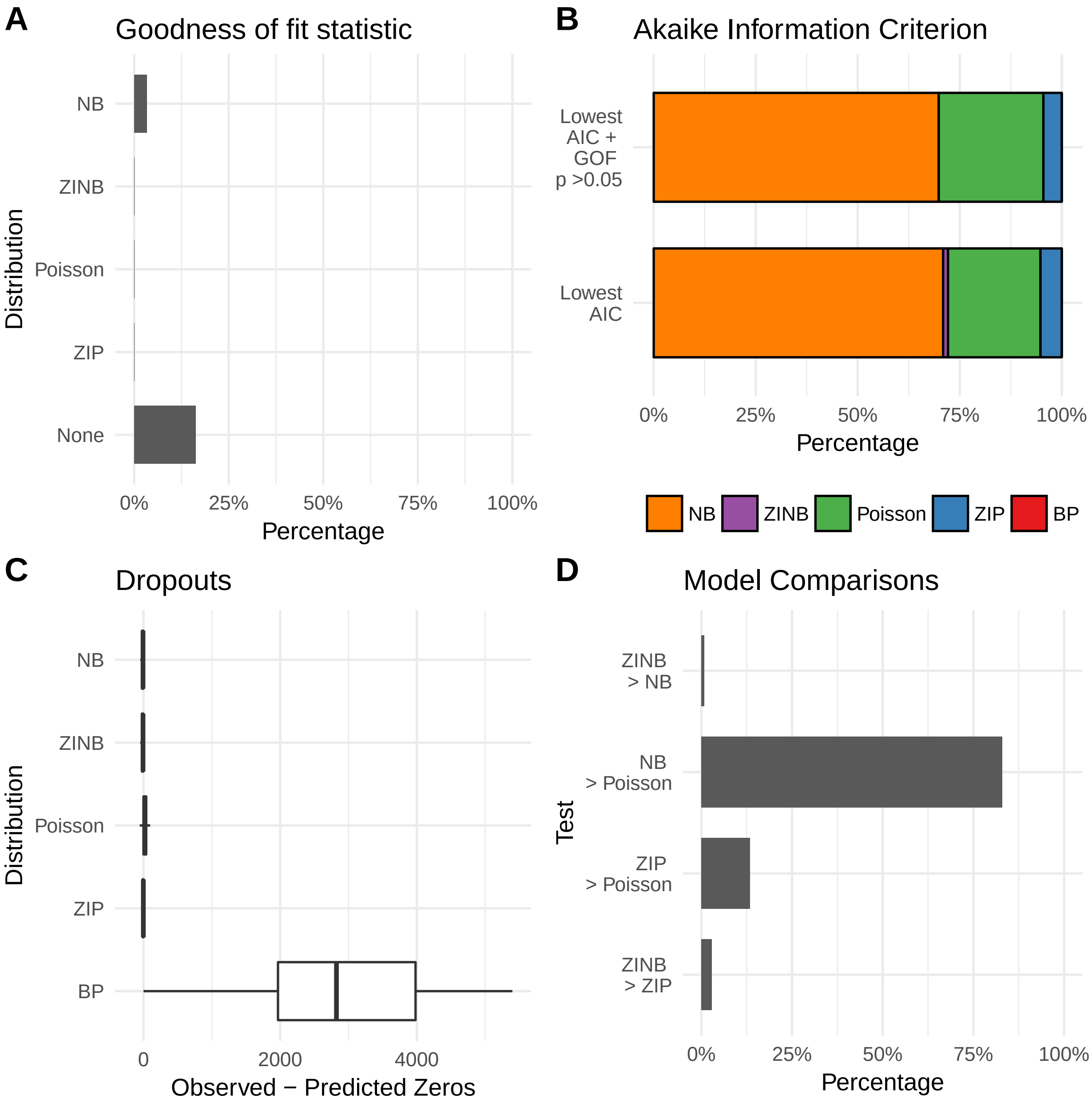

Figure S30: Zheng et al. 2017: Peripheral Blood Mononuclear Cells CD8+/CD45RA+ Naive Cytotoxic T Cells (10XGenomics). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

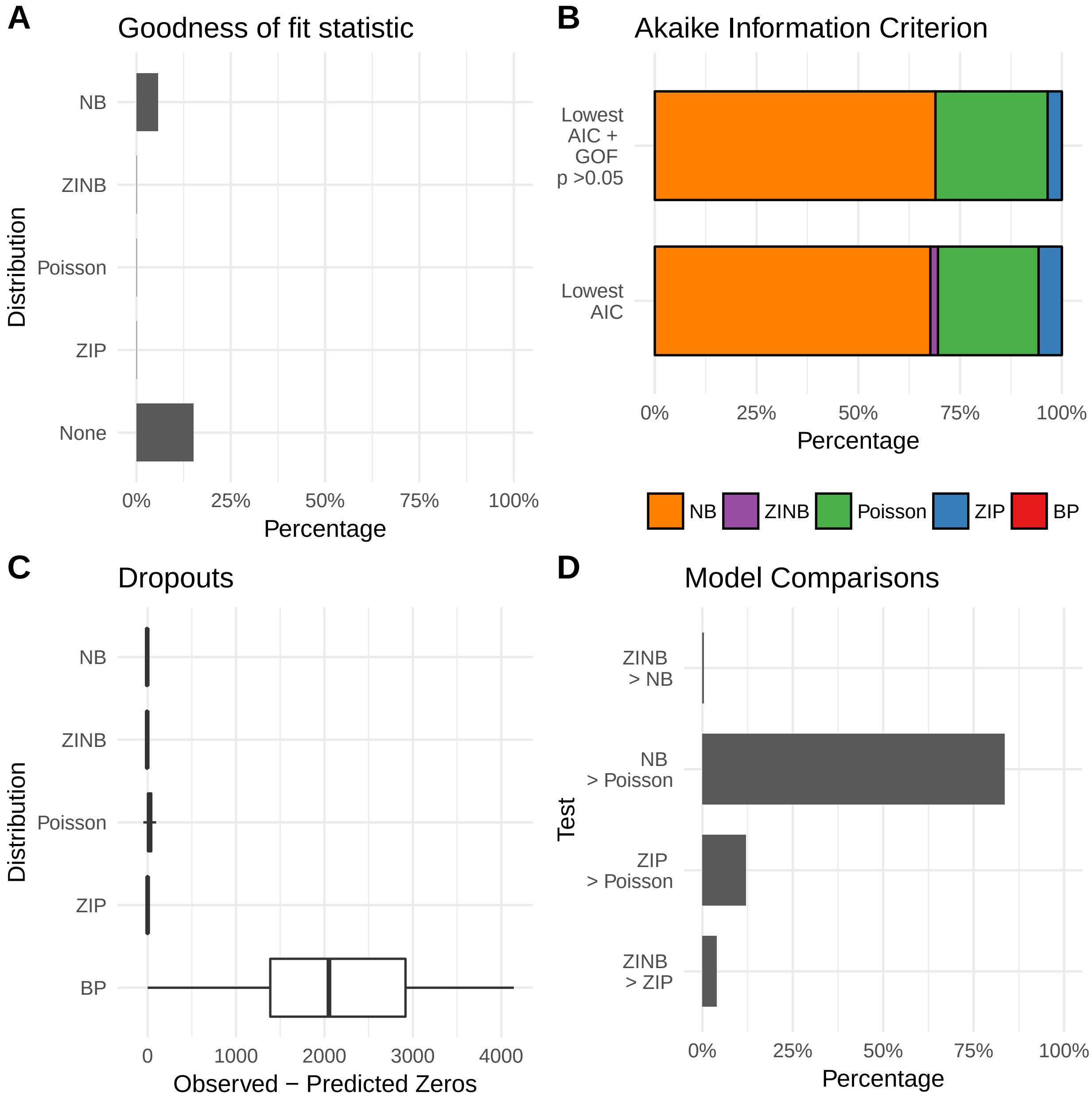

Figure S31: Zheng et al. 2017: Peripheral Blood Mononuclear Cells CD4+/CD45RA+/CD25- Naive T Cells (10XGenomics). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

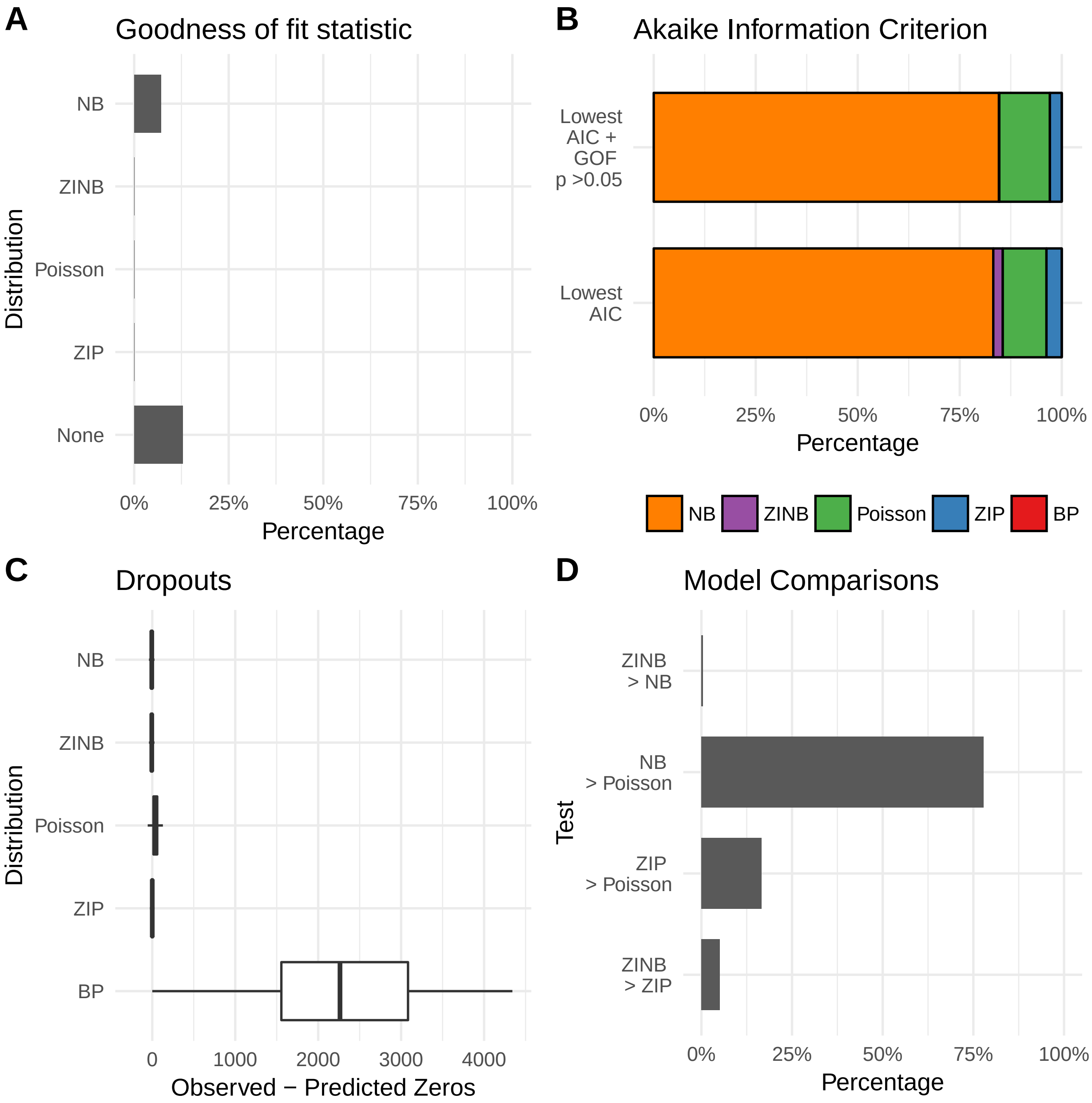

Figure S32: Zheng et al. 2017: Peripheral Blood Mononuclear Cells CD4+/CD25+ Regulatory T Cells (10XGenomics). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

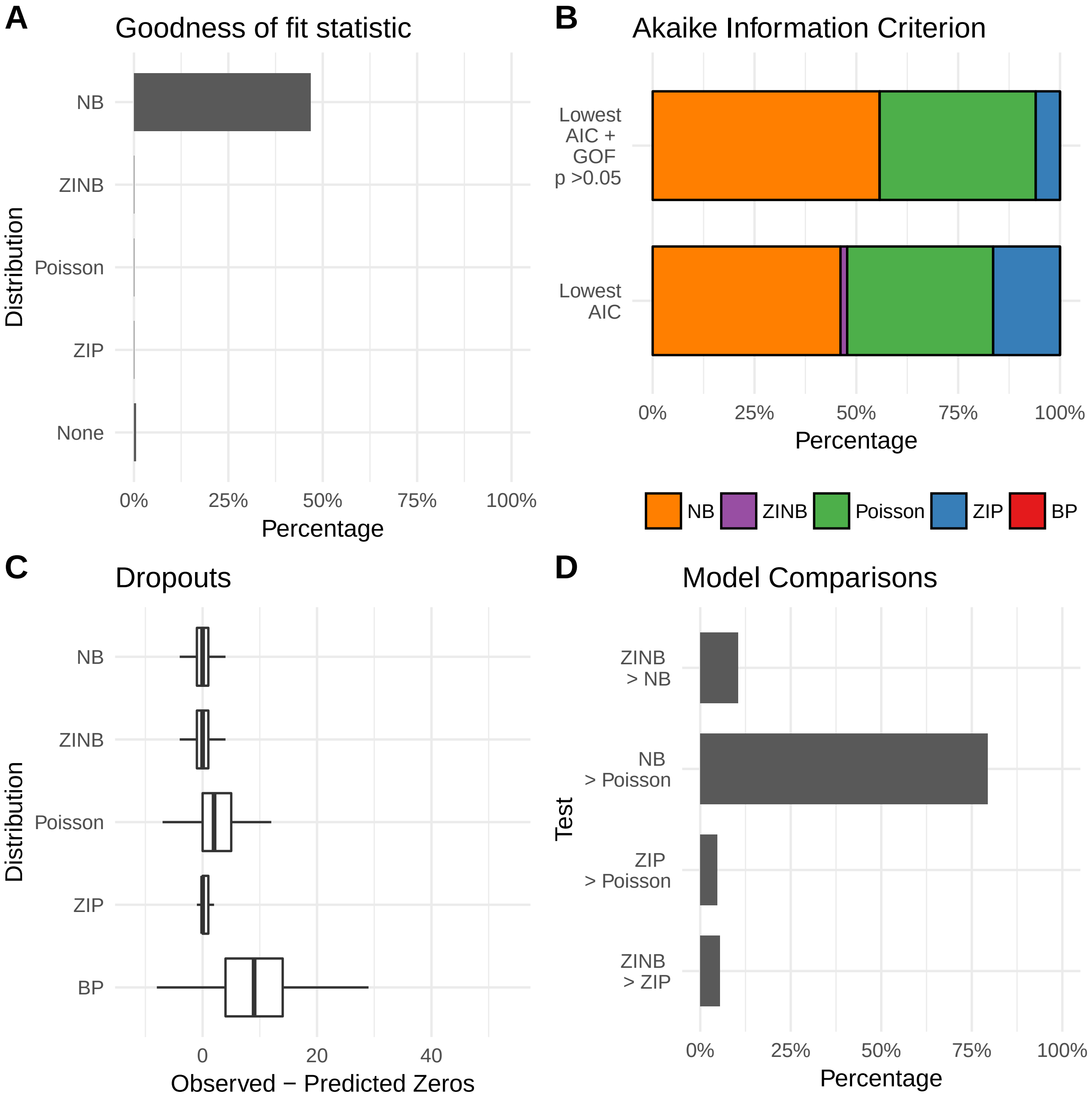

Figure S33: Ziegenhain et al. 2017: Embryonic stem cells (CEL-seq2). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

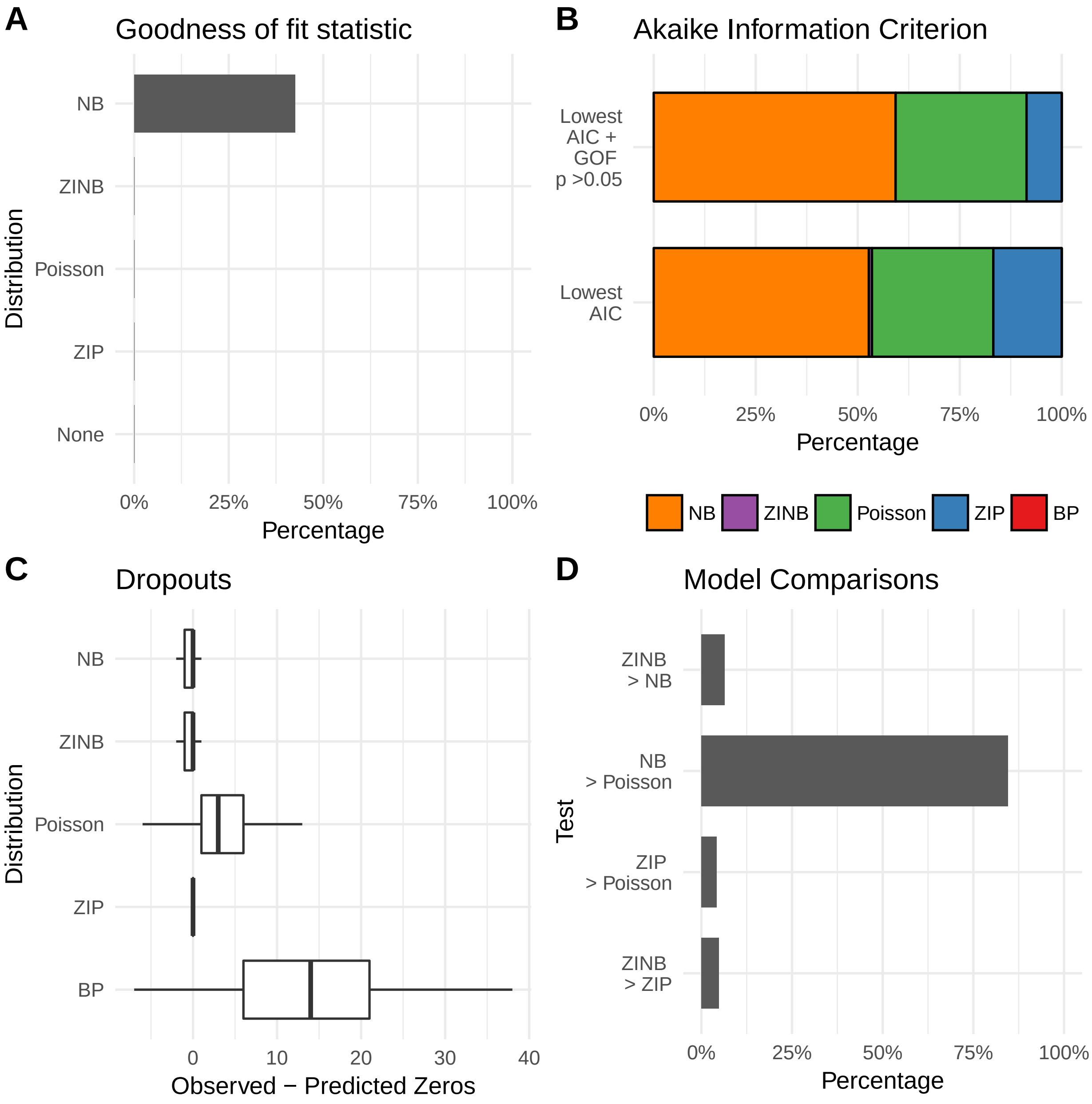

Figure S34: Ziegenhain et al. 2017: Embryonic stem cells (Drop-seq). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

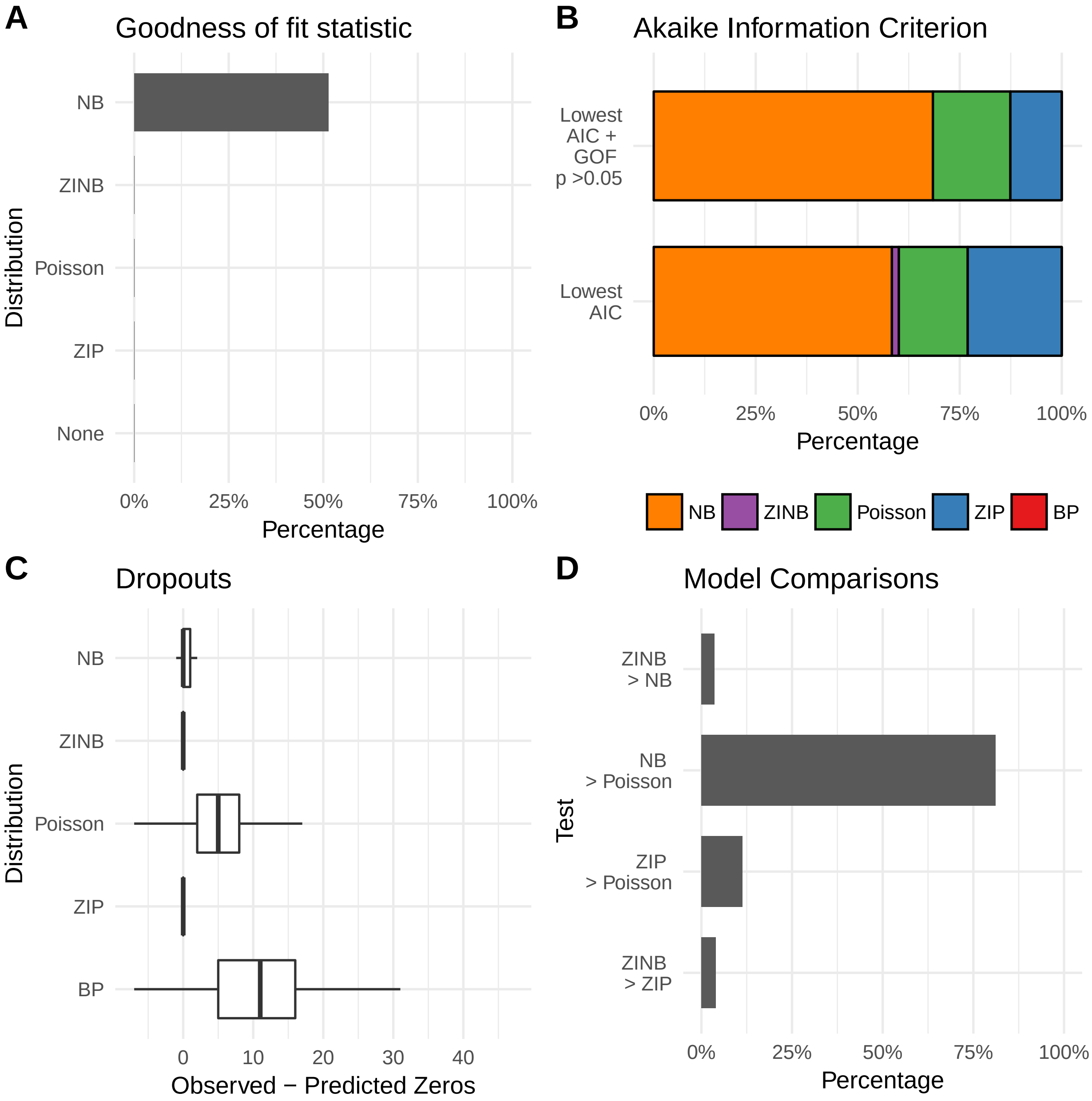

Figure S35: Ziegenhain et al. 2017: Embryonic stem cells (MARS-seq). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

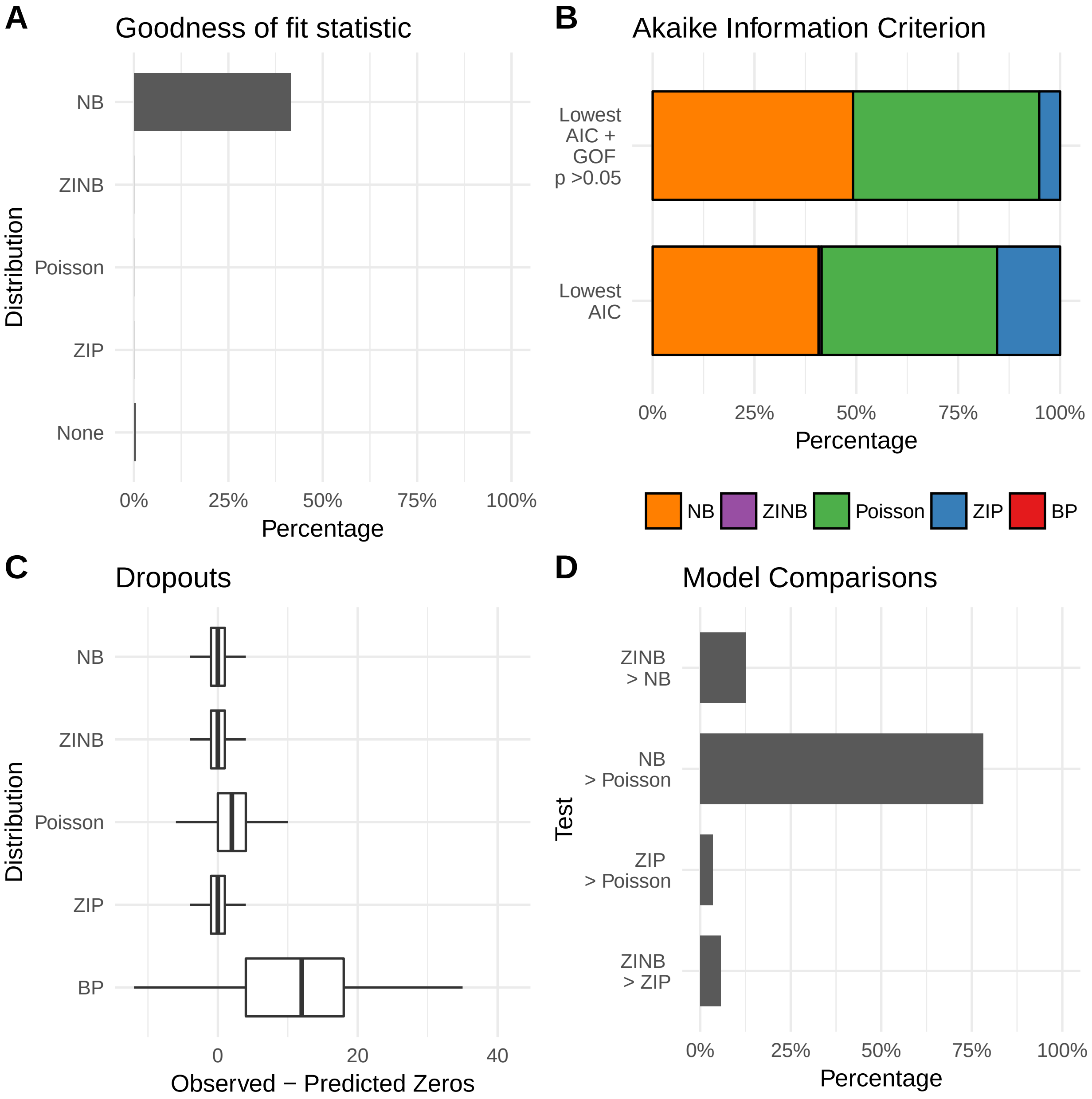

Figure S36: Ziegenhain et al. 2017: Embryonic stem cells (SCRB-seq). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

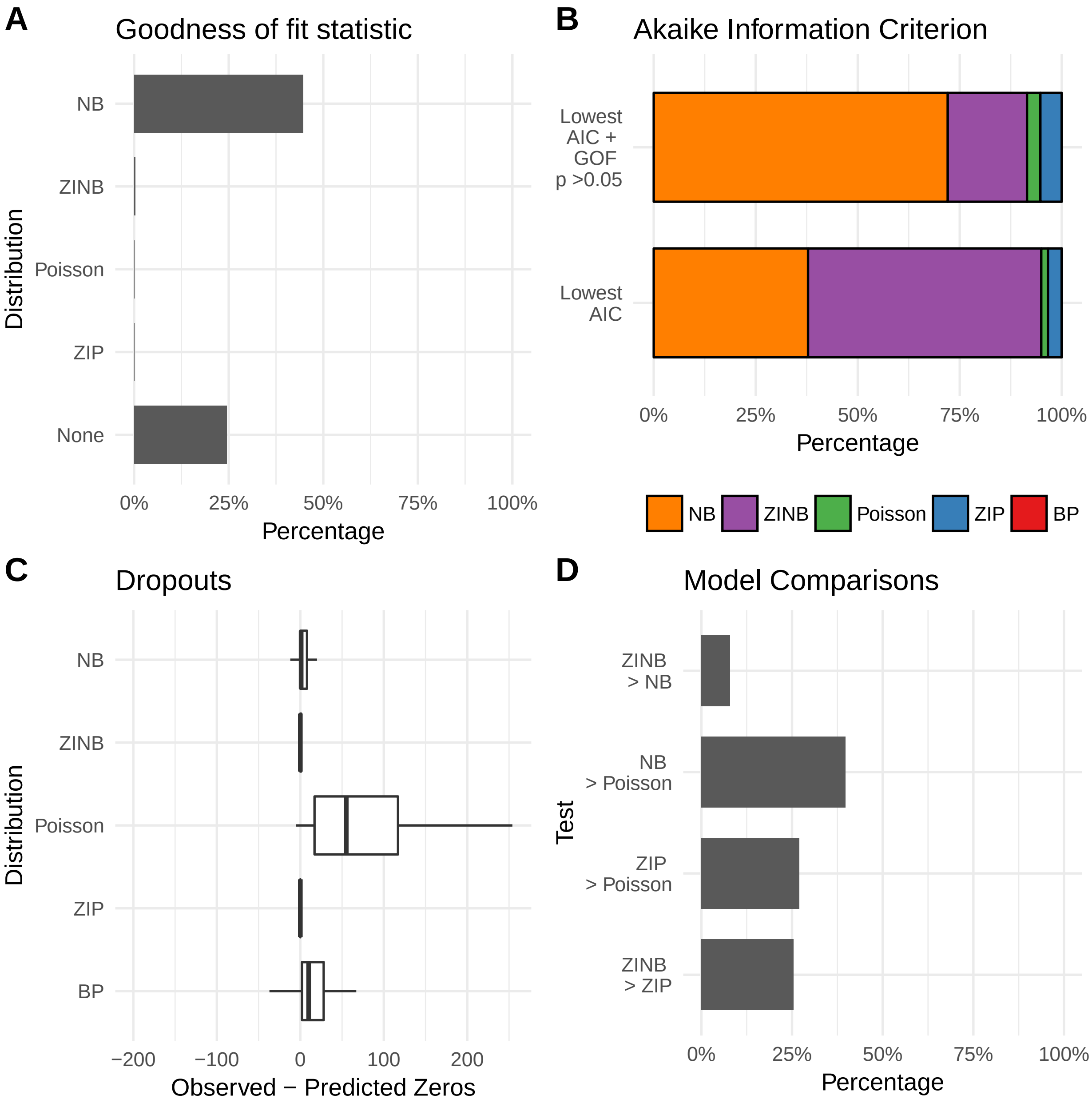

Figure S37: Ziegenhain et al. 2017: Embryonic stem cells (Smart-seq/C1). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

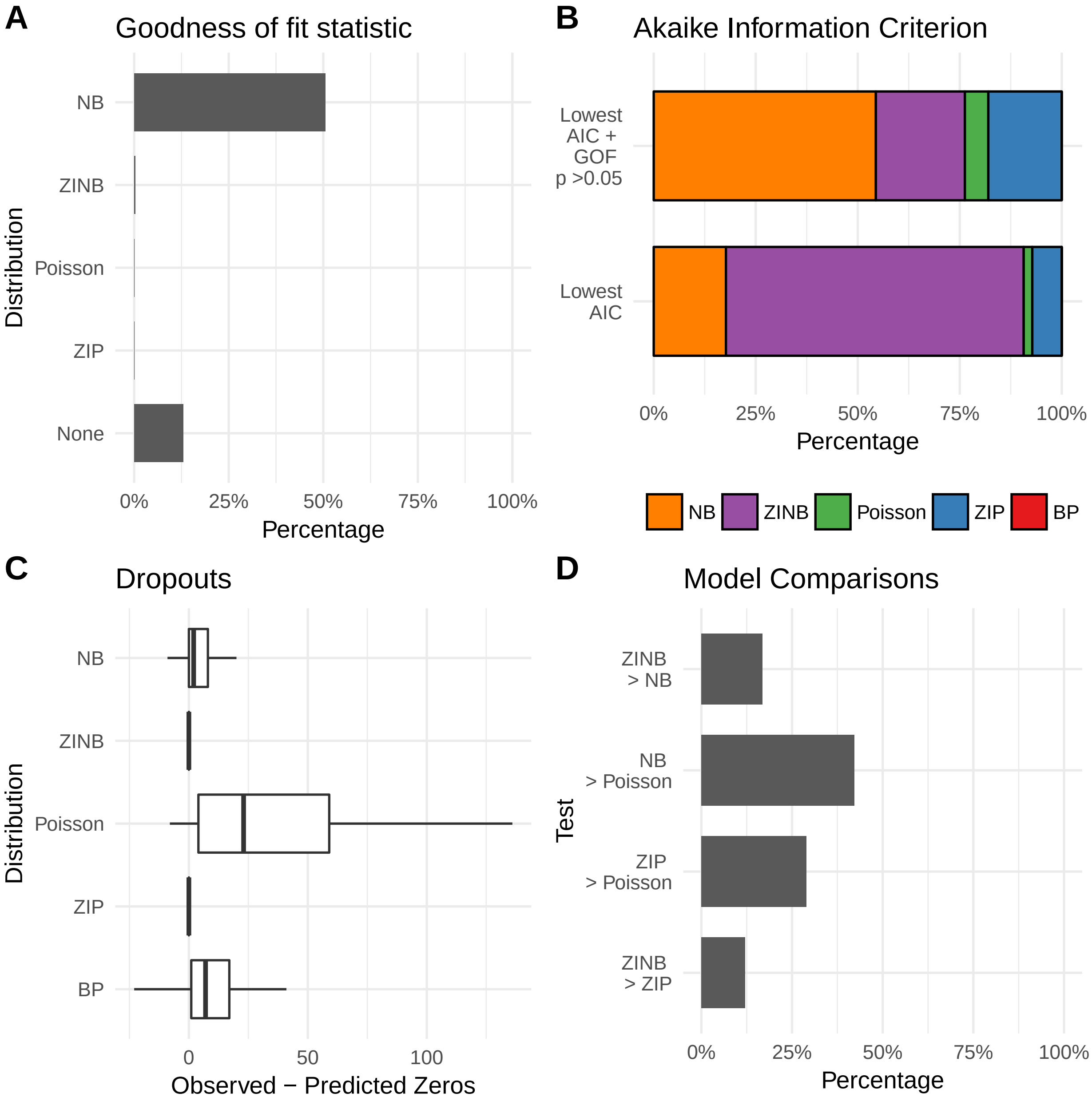

Figure S38: Ziegenhain et al. 2017: Embryonic stem cells (Smart-seq2). A) Goodness-of-fit of the model assessed with a chi-square test based on residual deviance and degrees of freedom. B) Akaike Information Criterion per gene: Model with the lowest AIC. Model with the lowest AIC and passed goodness-of-fit statistic test. C) Observed versus predicted dropouts per model and gene plotted without outliers. D) Model Assessment based on Likelihood Ratio Test for nested models and Vuong Test for non-nested models. NB = Negative binomial; ZIP = Zero-inflated Poisson; ZINB = Zero-inflated negative binomial; BP = Beta-Poisson.

References

Buettner, F., Natarajan, K. N., Casale, F. P., Proserpio, V., Scialdone, A., Theis, F. J., Teichmann, S. A., Marioni, J. C., and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.*, advance online publication.

Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. *Nat. Methods*, 11(6):637–640.

Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.-B., Lönnerberg, P., and Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.*, 21(7):1160–1167.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods*, 11(2):163–166.

Kolodziejczyk, A. A., Kim, J. K., Tsang, J. C. H., Ilicic, T., Henriksson, J., Natarajan, K. N., Tuck, A. C., Gao, X., Bühler, M., Liu, P., Marioni, J. C., and Teichmann, S. A. (2015). Single cell RNA-Sequencing of pluripotent states unlocks modular transcriptional variation. *Cell Stem Cell*, 17(4):471–485.

Pollen, A. A., Nowakowski, T. J., Shuga, J., Wang, X., Leyrat, A. A., Lui, J. H., Li, N., Szpankowski, L., Fowler, B., Chen, P., Ramalingam, N., Sun, G., Thu, M., Norris, M., Lebofsky, R., Toppani, D., Kemp, Ii, D. W., Wong, M., Clerkson, B., Jones, B. N., Wu, S., Knutsson, L., Alvarado, B., Wang, J., Weaver, L. S., May, A. P., Jones, R. C., Unger, M. A., Kriegstein, A. R., and West, J. A. A. (2014). Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.*, 32(10):1053–1058.

Soumillon, M., Cacchiarelli, D., Semrau, S., Oudenaarden, A. v., and Mikkelsen, T. S. (2014). Characterization of directed differentiation by high-throughput single-cell RNA-Seq. *bioRxiv*, page 003236.

Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., Bharadwaj, R., Wong, A., Ness, K. D., Beppu, L. W., Deeg, H. J., McFarland, C., Loeb, K. R., Valente, W. J., Ericson, N. G., Stevens, E. A., Radich, J. P., Mikkelsen, T. S., Hindson, B. J., and Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.*, 8:14049.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative analysis of Single-Cell RNA sequencing methods. *Mol. Cell*, 65(4):631–643.e4.