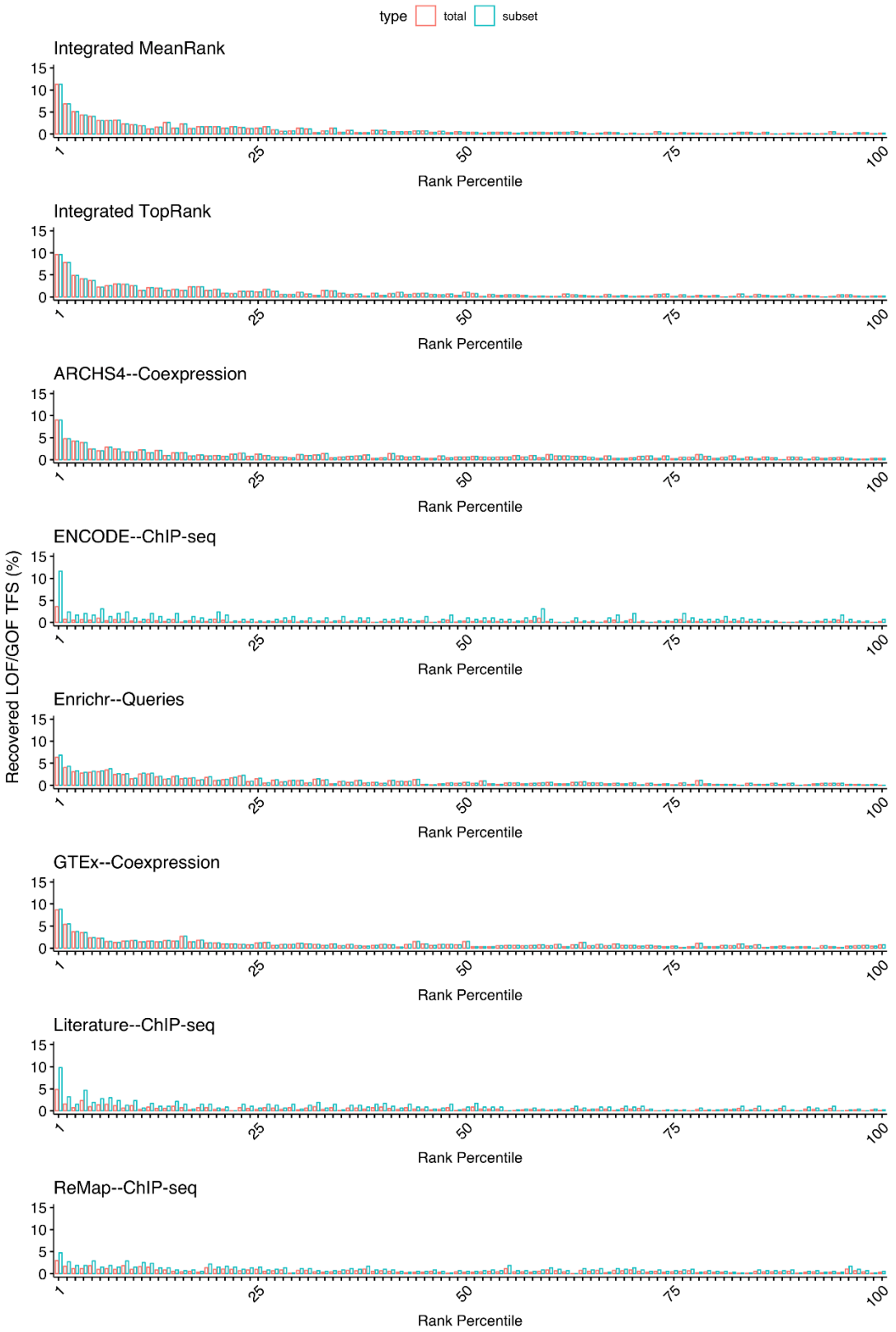
Supplementary Figures for:

ChEA3: Transcription Factor Enrichment Analysis by Orthogonal Omics Integration

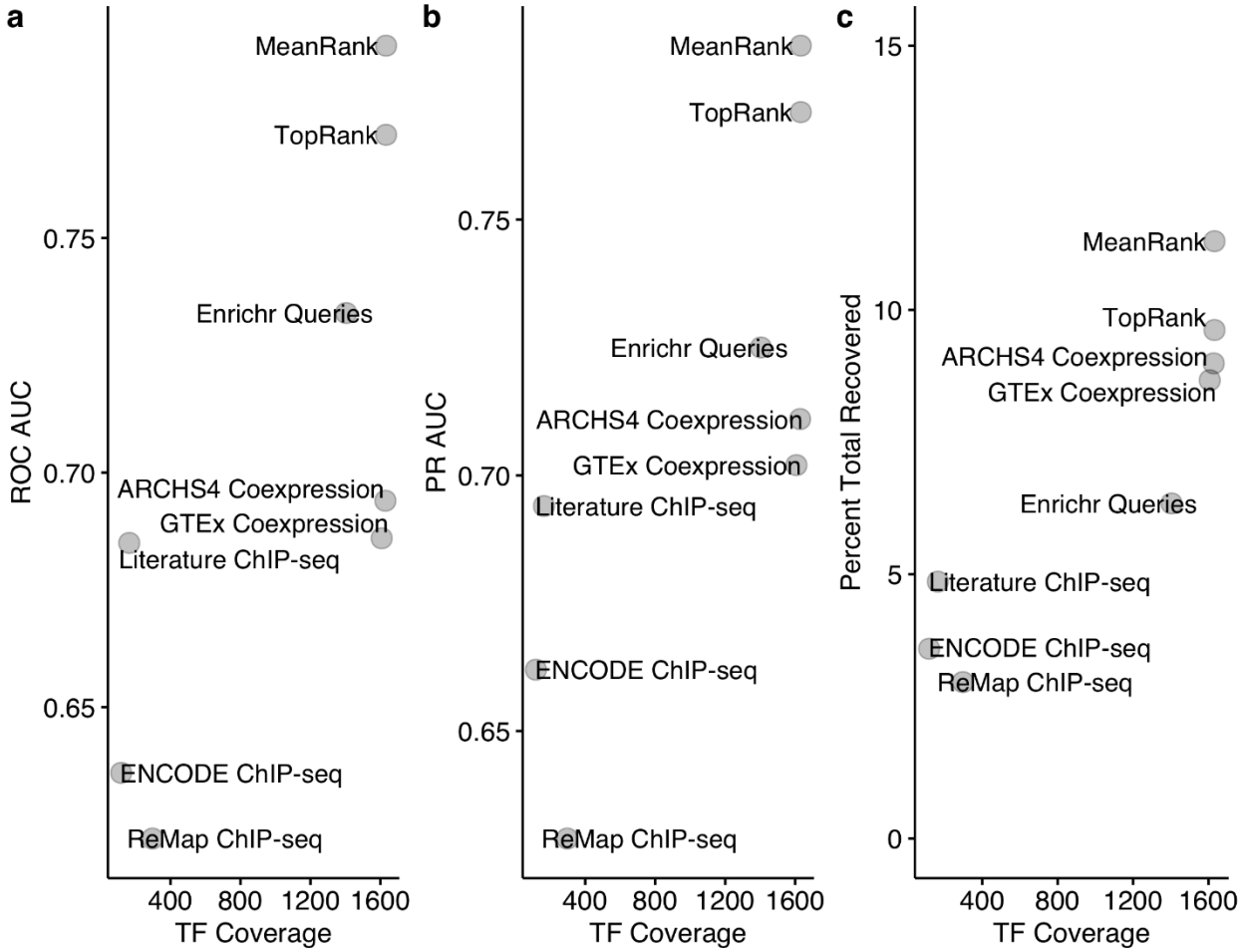
Alexandra B. Keenan (alexandra.keenan@icahn.mssm.edu)[1], Denis Torre (denis.torre@mssm.edu)[1], Alexander Lachmann (alexander.lachmann@mssm.edu)[1], Ariel K. Leong (akl1@stanford.edu)1, Megan L. Wojciechowicz (megan.wojciechowicz@mssm.edu)[1], Vivian Utti (vu27@cornell.edu)[1], Kathleen M. Jagodnik (kathleen.jagodnik@mssm.edu)[1], Eryk Kropiwnicki (eryk.kroiwnicki@icahn.mssm.edu)[1], Zichen Wang (zichen.wang@mssm.edu)[1], Avi Ma'ayan (avi.maayan@mssm.edu)[1]

[1]Department of Pharmacological Sciences, Mount Sinai Center for Bioinformatics, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1603, New York, NY 10029 USA

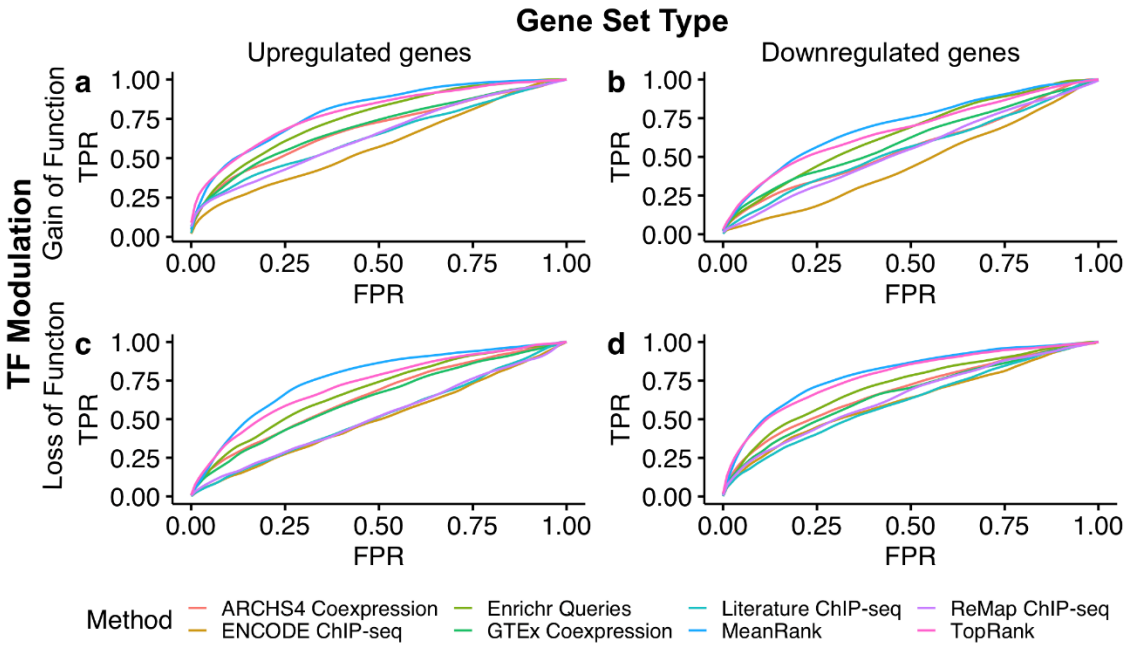
*To whom correspondence should be addressed. Email: avi.maayan@mssm.edu; Tel: +212 241 1153



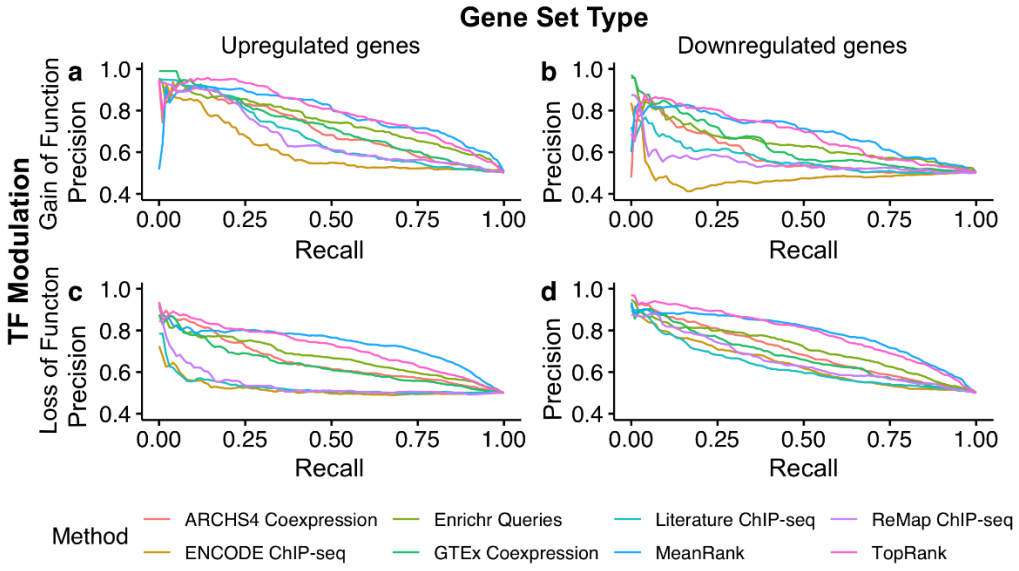
Supplementary Figure 1. Performance of the ChEA3 libraries and integration techniques in recovering the perturbed TFs from 946 TF LOF and GOF experiments from the TFpertGEOupdn benchmarking dataset. The red bars indicate the fraction of the 946 experiments in TFpertGEOupdn where the upstream perturbed TF was recovered for each rank percentile. The blue bars indicate the rank percentile distribution of recovery of the subset of TFpertGEOupdn where the perturbed TFs are covered by the library.



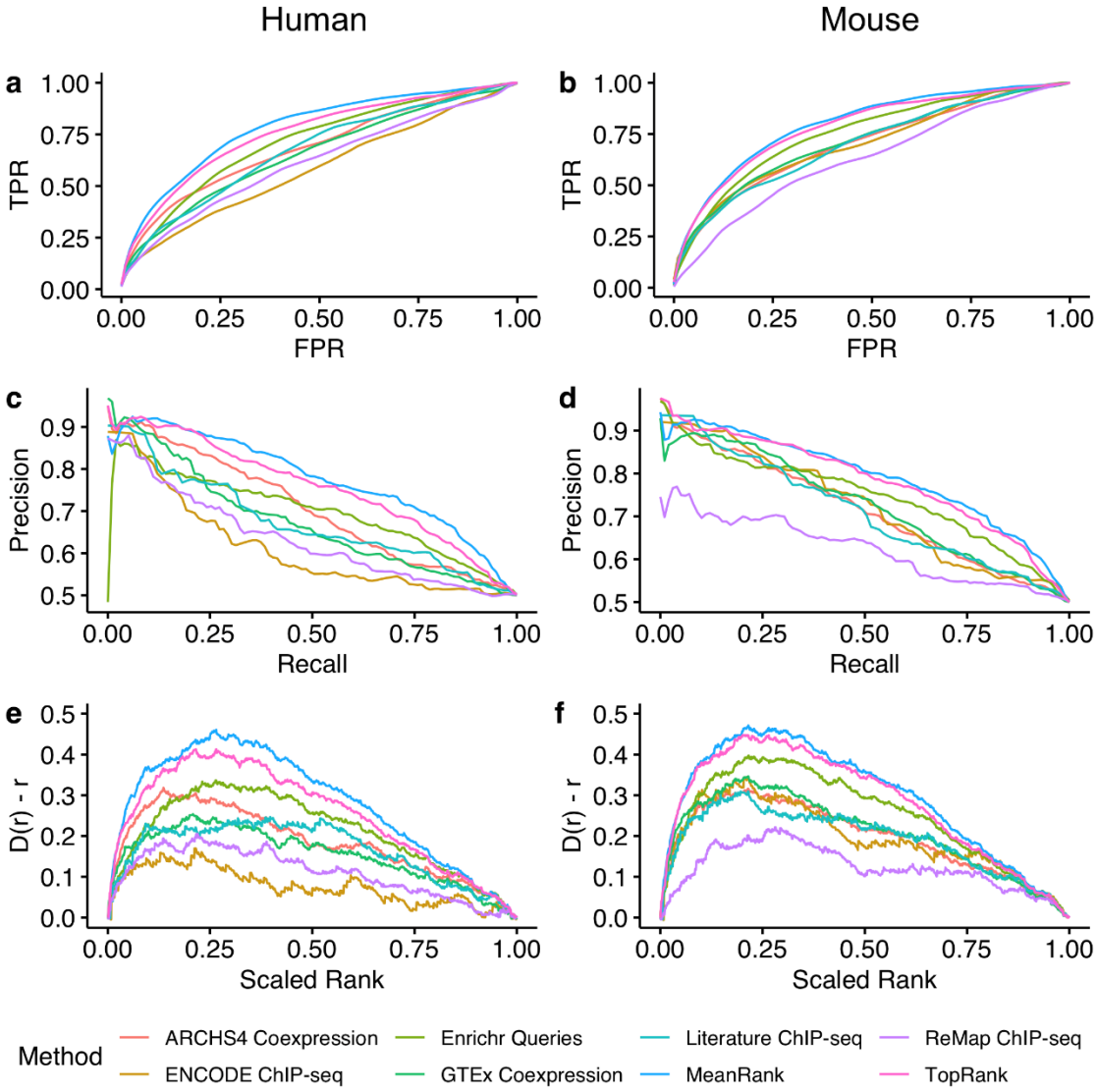
Supplementary Figure 2. ChEA3 library performance on the TFpertGEOupdn benchmarking dataset compared to library TF coverage. a) AUROC compared to the number of unique TFs covered by the library. b) AUPR compared to the number of unique TFs covered by the library. c) The fraction of the TFpertGEO perturbed TFs recovered by the top 1 percentile of ranks.



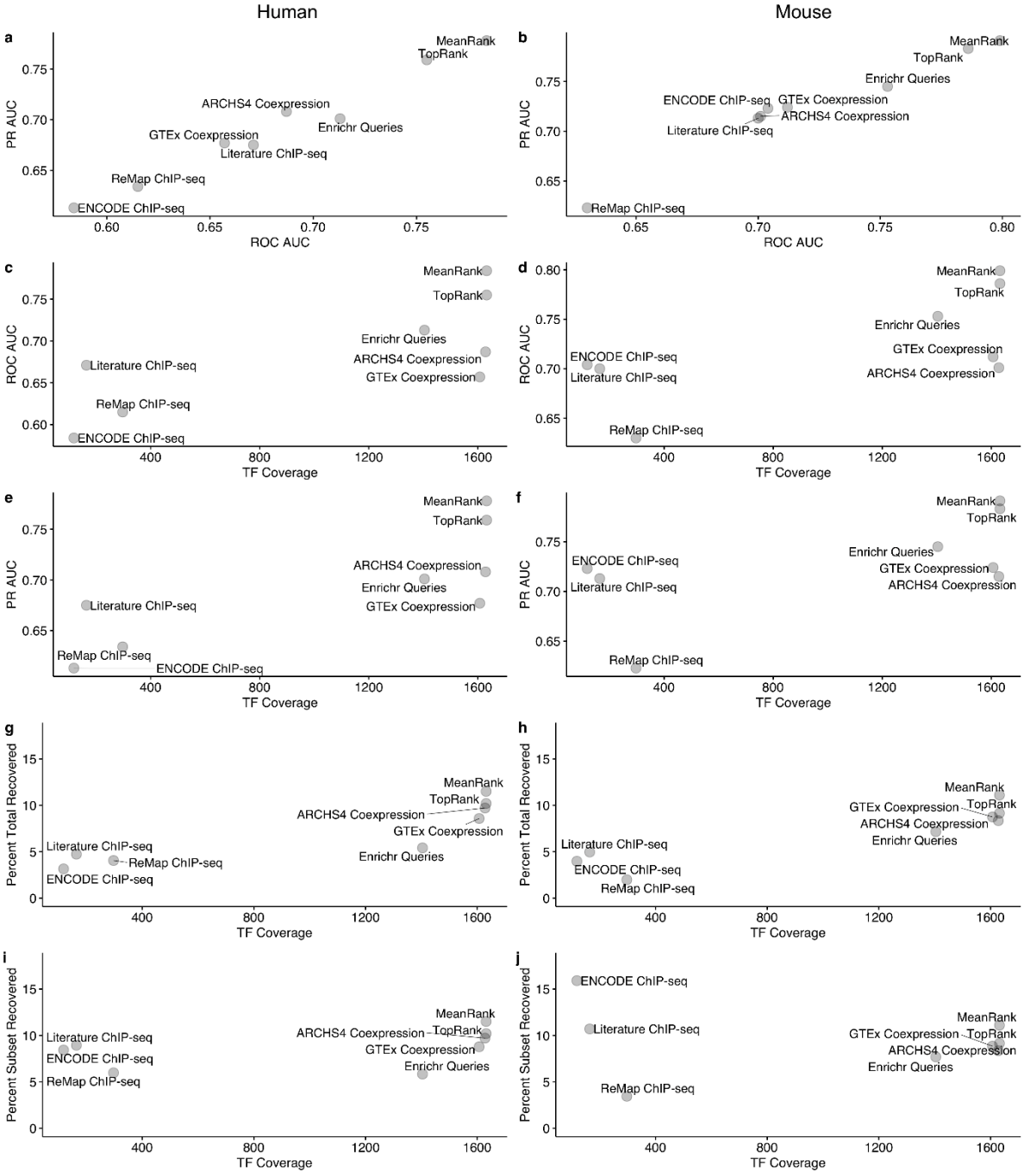
Supplementary Figure 3. Effect of input type on ChEA3 performance. ROC curves describing performance for a) TF overexpression or chemical activation experiments from TFpertGEOup; b) TF overexpression or chemical activation experiments from TFpertGEOdn; c) TF knockdown, knockout or chemical inactivation experiments from TFpertGEOup; and d) TF knockdown, knockout or chemical inactivation experiments from TFpertGEOdn.



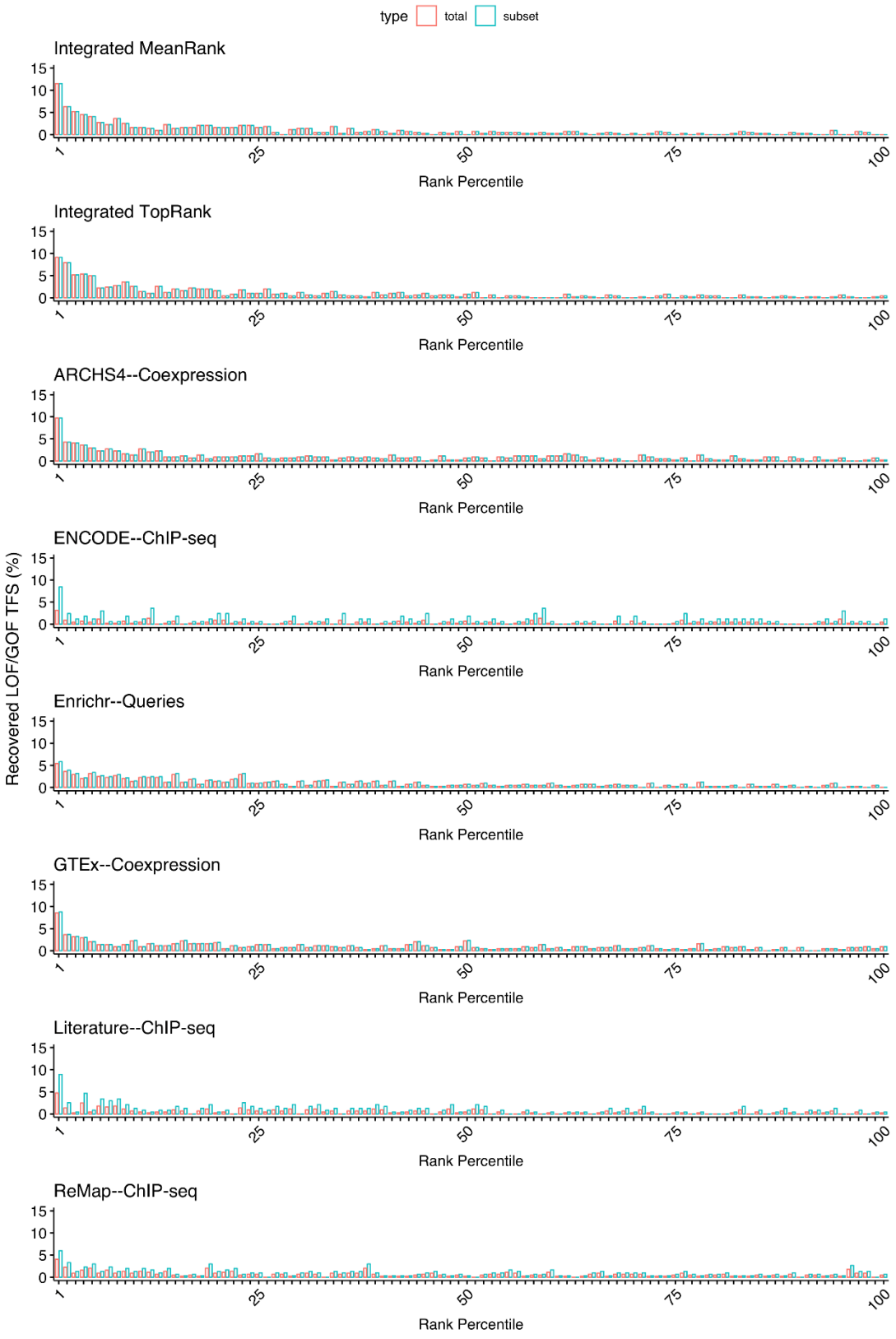
Supplementary Figure 4. Effect of input type on ChEA3 performance. PR curves describing performance for a) TF overexpression or chemical activation experiments from TFpertGEOup; b) TF overexpression or chemical activation experiments from TFpertGEOdn; c) TF knockdown, knockout or chemical inactivation experiments from TFpertGEOup; and d) TF knockdown, knockout or chemical inactivation experiments from TFpertGEOdn.



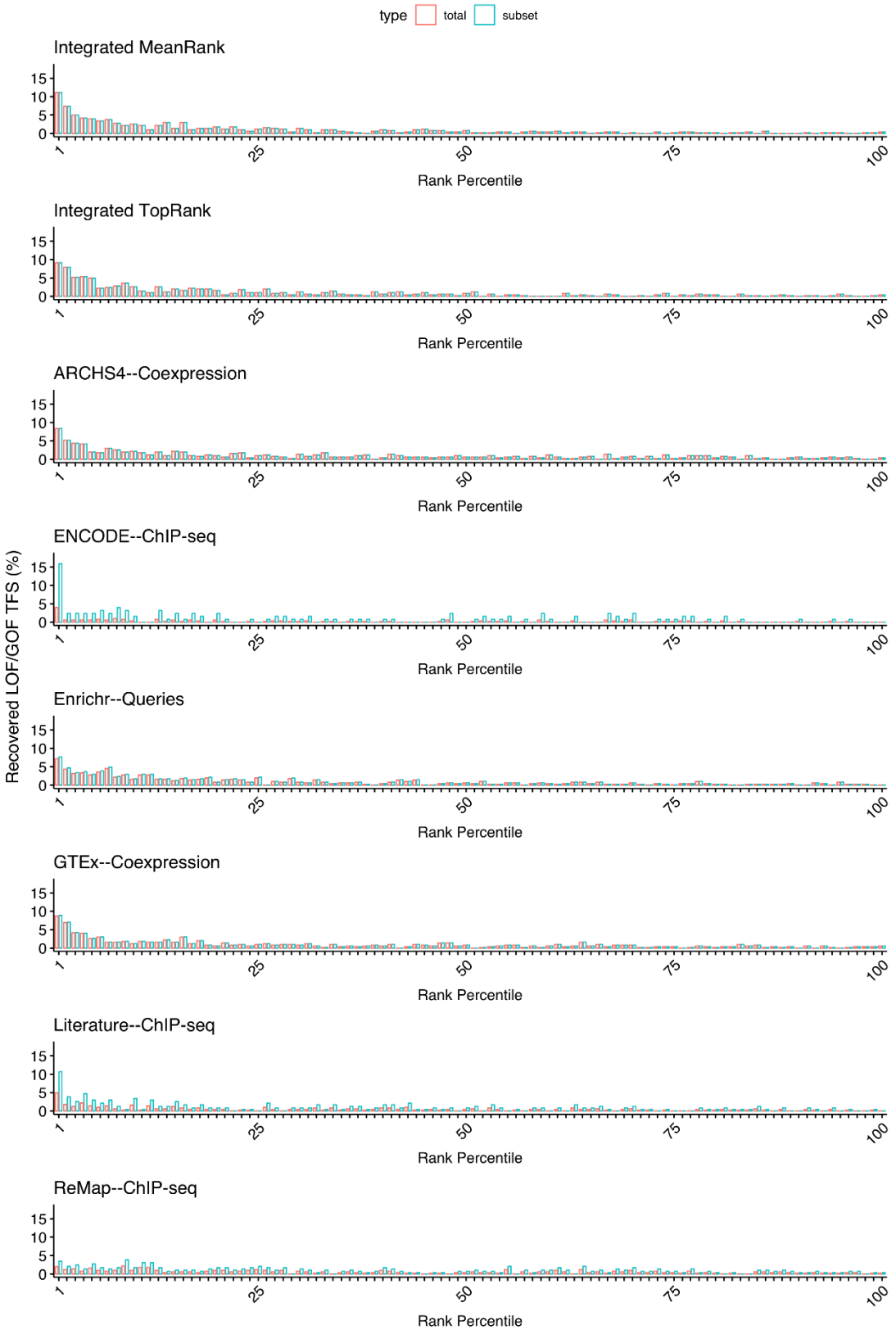
Supplementary Figure 5. Effect of human and mouse query types on ChEA3 performance. a) Composite ROC curves generated from 5,000 bootstrapped curves demonstrating performance on the hsTFpertGEOupdn benchmarking dataset. b) Composite ROC curves generated from 5,000 bootstrapped curves demonstrating performance on the mmTFpertGEOupdn benchmarking dataset. c) Composite PR curves generated from 5,000 bootstrapped curves demonstrating performance on the hsTFpertGEOupdn benchmarking dataset. d) Composite PR curves generated from 5,000 bootstrapped curves demonstrating performance on the mmTFpertGEOupdn benchmarking dataset. e) The deviation of the cumulative distribution from uniform of the scaled rankings of each perturbed TF in the hsTFpertGEOupdn. Anderson Darling test of f) the deviation of the cumulative distribution from uniform of the scaled rankings of each perturbed TF in the mmTFpertGEOupdn.



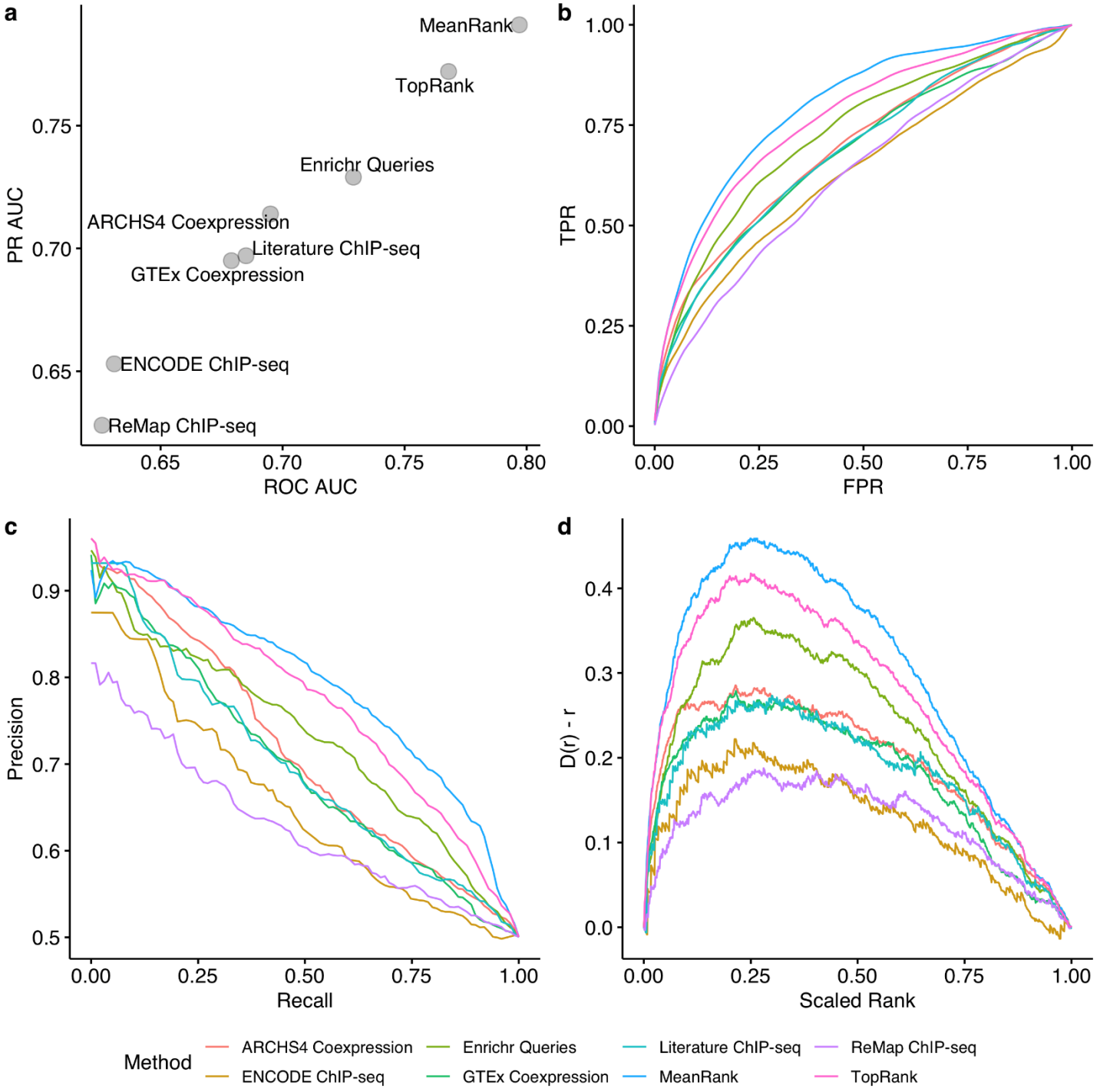
Supplementary Figure 6. Effect of human and mouse query types on ChEA3 performance. a) Mean AUROC and mean AUPR over 5,000 bootstrapped ROC and PR curves from the hsTFpertGEOupdn benchmarking dataset. b) Mean AUROC and mean AUPR over 5,000 bootstrapped ROC and PR curves from the mmTFpertGEOupdn benchmarking dataset. c) Mean AUROC compared to TF coverage from the hsTFpertGEOupdn benchmarking dataset. d) Mean AUROC compared to TF coverage from the mmTFpertGEOupdn benchmarking dataset. e) Mean AUPR compared to TF coverage from the hsTFpertGEOupdn benchmarking dataset. f) Mean AUPR compared to TF coverage from the mmTFpertGEOupdn benchmarking dataset. g) Percent of the 443 TFs perturbed in hsTFpertGEOupdn that were recovered in the top 1% of ranks. h) Percent of the 503 TFs perturbed in mmTFpertGEOupdn recovered in the top 1% of ranks. i) Percent of the subset of TF experiments in hsTFpertGEOupdn where the perturbed TF is covered by the library that were recovered in the top 1% of ranks. j) Percent of the subset of TF experiments in mmTFpertGEOupdn where the perturbed TF is covered by the library that were recovered in the top 1% of ranks.



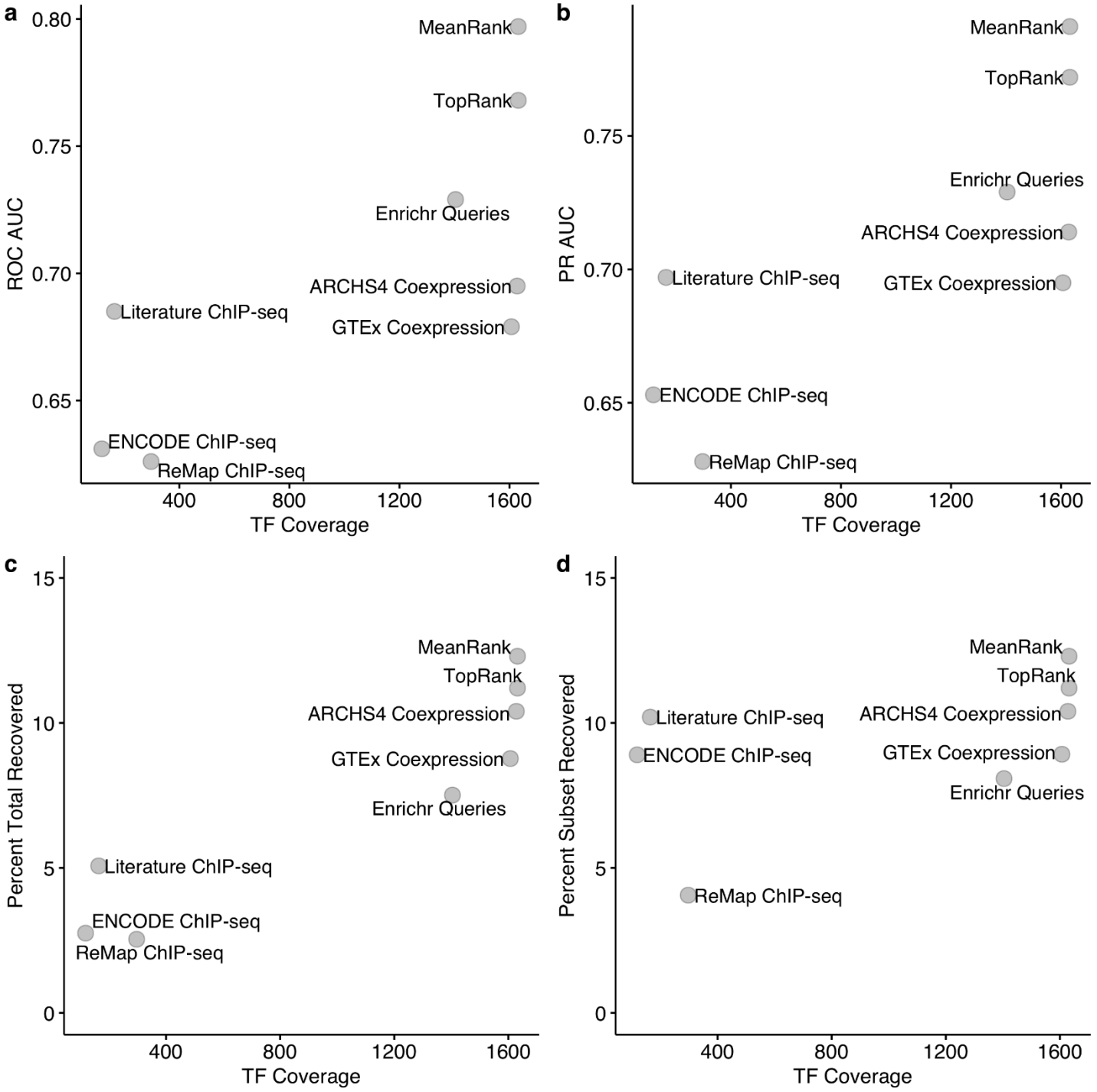
Supplementary Figure 7. Performance of the ChEA3 libraries and integration techniques in recovering the perturbed TFs from 443 TF LOF/GOF experiments from the hsTFpertGEOupdn benchmarking dataset. The red bars indicate the fraction of the 443 experiments in hsTFpertGEOupdn where the upstream perturbed TF was recovered for each rank percentile. The blue bars indicate the rank percentile distribution of recovery of the subset of hsTFpertGEOupdn where the perturbed TFs are covered by the library.



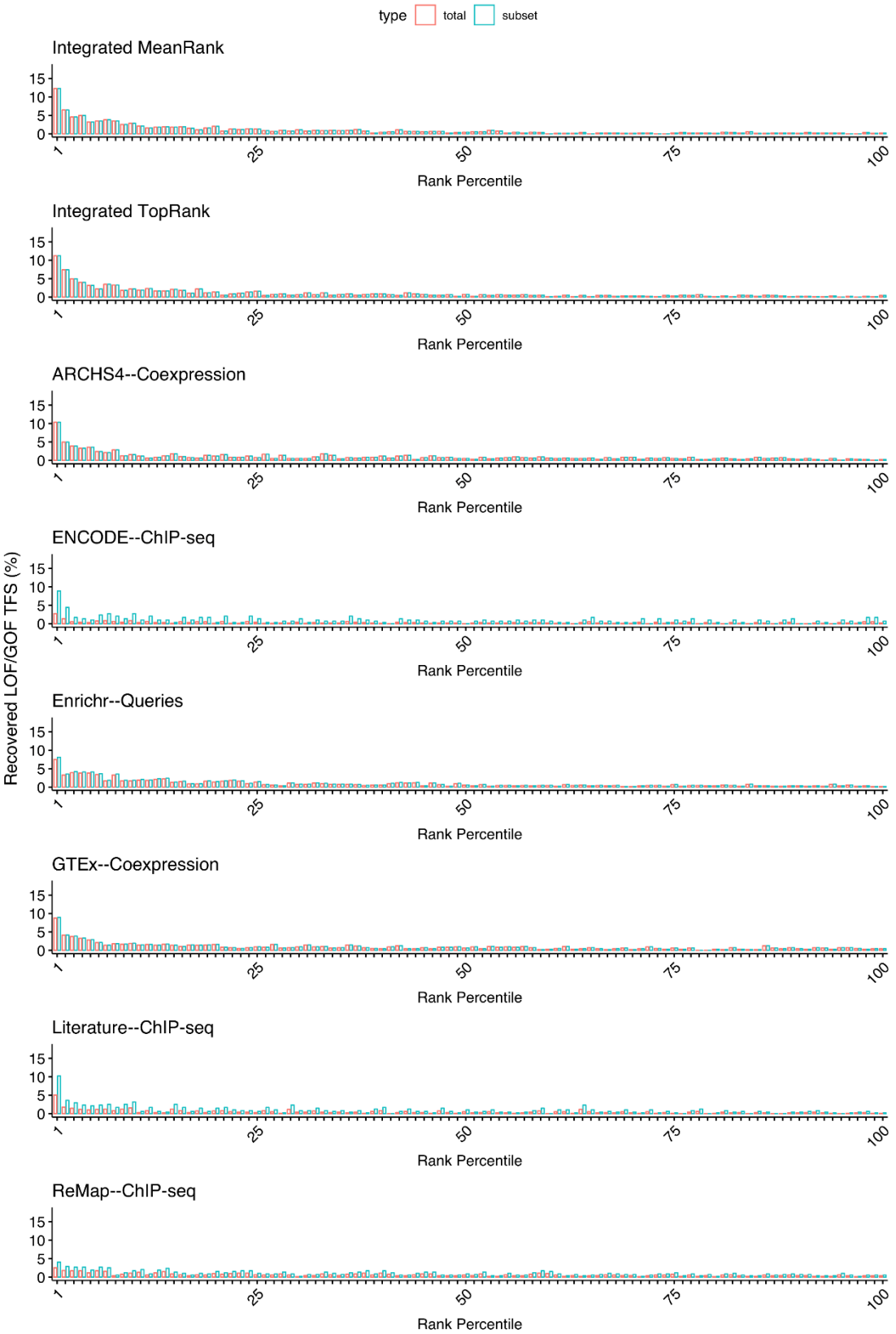
Supplementary Figure 8. Performance of the ChEA3 libraries and integration techniques in recovering the perturbed TFs from 503 TF LOF/GOF experiments from the mmTFpertGEOupdn benchmarking dataset. The red bars indicate the fraction of the 503 experiments in mmTFpertGEOupdn where the upstream perturbed TF was recovered for each rank percentile. The blue bars indicate the rank percentile distribution of the recovery of the subset of mmTFpertGEOupdn where the perturbed TFs are covered by the library.



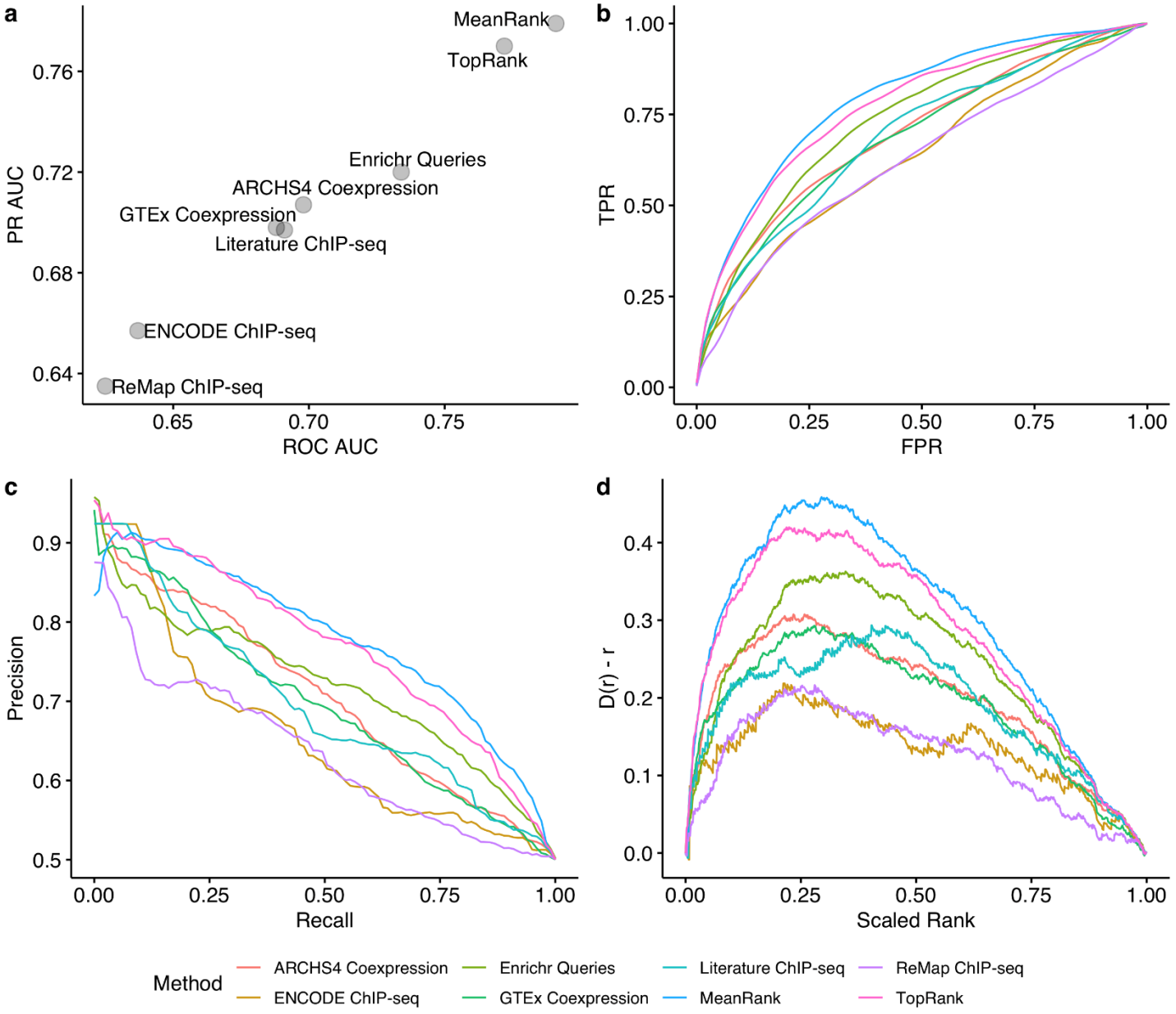
Supplementary Figure 9. Performance of the ChEA3 libraries and integration techniques in recovering the perturbed TFs from 946 TF LOF and GOF experiments from the TFpertGEO200 benchmark dataset. a) Mean ROC AUC and mean PR AUC over 5,000 bootstrapped ROC and PR curves; b) Composite ROC curves generated from 5,000 bootstrapped curves; c) Composite PR curves generated from 5,000 bootstrapped curves; d) The deviation of the cumulative distribution from uniform of the scaled rankings of each perturbed TF in the benchmarking dataset. Anderson-Darling test of uniformity: MeanRank p = 6.34 x 10^{-7}; TopRank p = 6.34 x 10^{-7}; ARCHS4 p = 6.34 x 10^{-7}; ENCODE p = 2.06 x 10^{-6}; Enrichr Queries p = 6.82 x 10^{-7}; GTEx p = 6.45 x 10^{-7}; Literature ChIP-seq p = 1.28 x 10^{-6}; ReMap p = 1.02 x 10^{-6}.



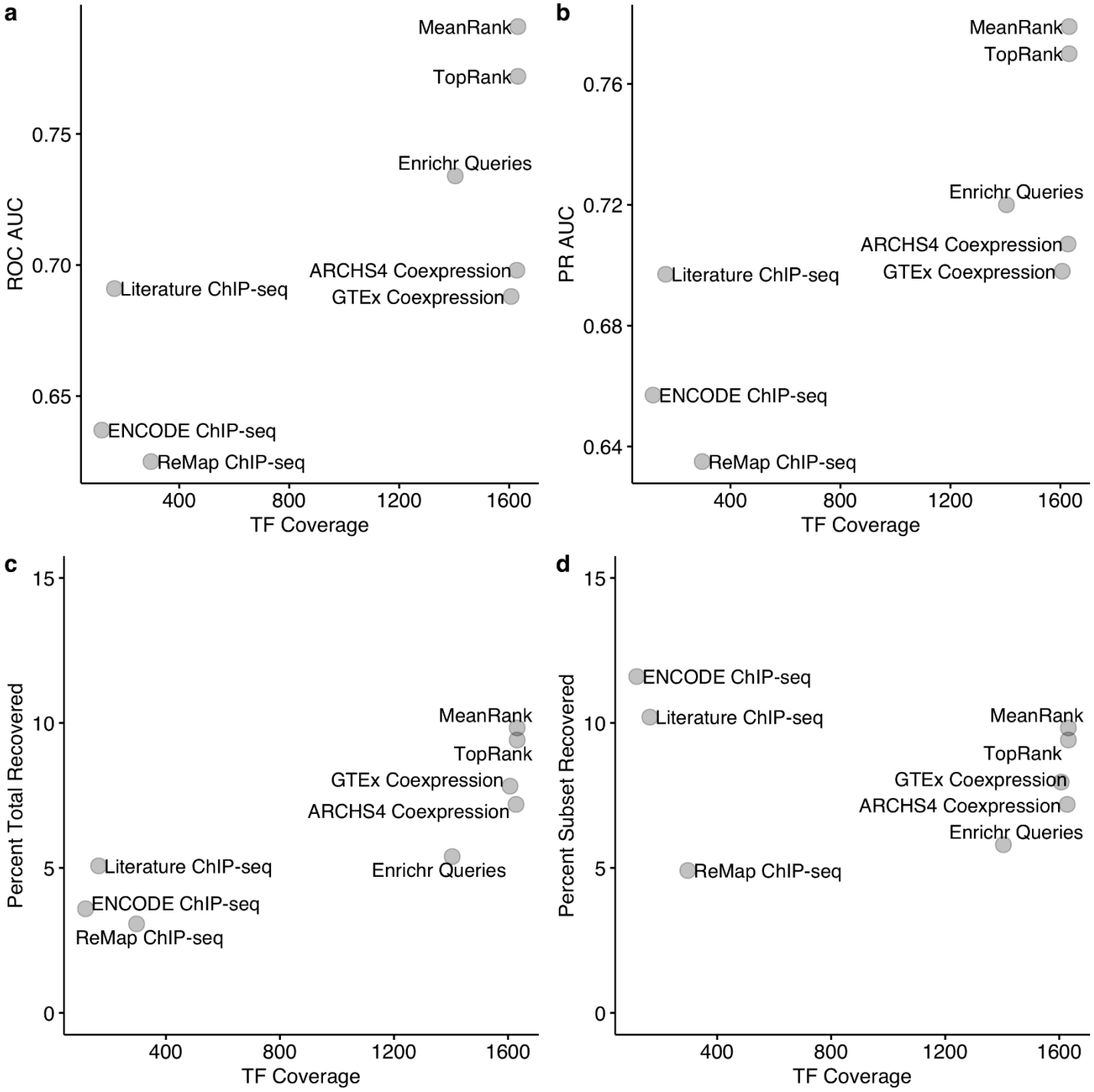
Supplementary Figure 10. Performance of the ChEA3 libraries and integration techniques in recovering the perturbed TFs from 946 TF LOF and GOF experiments from the TFpertGEO200 benchmark dataset compared to ChEA3 TF coverage. a) Mean ROC AUC over 5,000 bootstrapped ROC curves compared to library TF coverage. b) Mean PR AUC over 5,000 bootstrapped ROC curves compared to library TF coverage. c) The percent of the LOF/GOF TFs in TFpertGEO200 recovered in the top 1 percentile of ranks compared to the library TF coverage. d) The percent of the LOF/GOF TFs in the subset of TFpertGEO200 that are covered by the library that were recovered in the top 1 percentile of ranks compared to the library TF coverage.



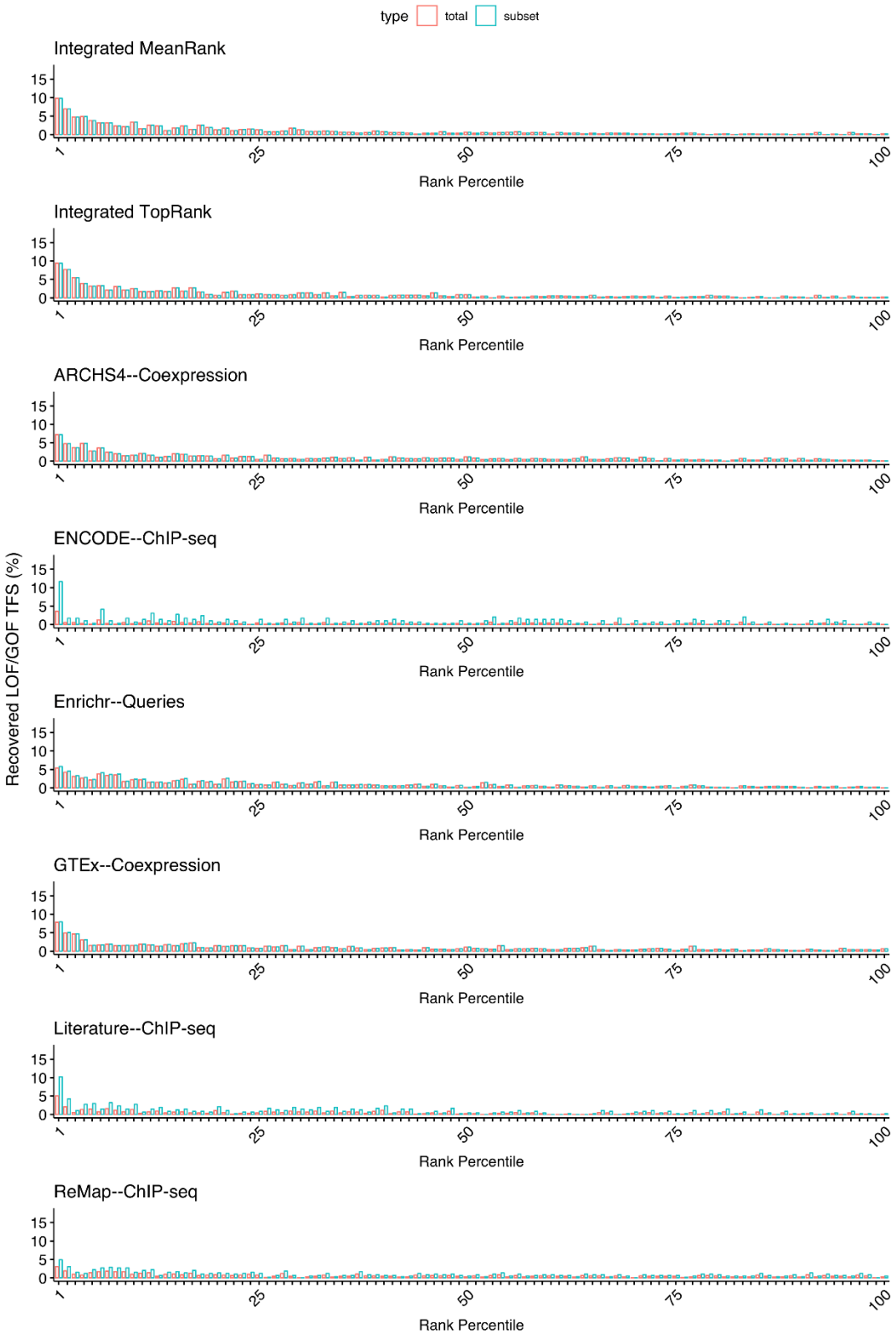
Supplementary Figure 11. Performance of the ChEA3 libraries and integration techniques in recovering the perturbed TFs from 946 TF LOF and GOF experiments from the TFpertGEO200 benchmarking dataset. The red bars indicate the fraction of the 946 experiments in TFpertGEO200 where the upstream perturbed TF was recovered for each rank percentile. The blue bars indicate the rank percentile distribution of recovery of the subset of TFpertGEO200 where the perturbed TFs are covered by the library.



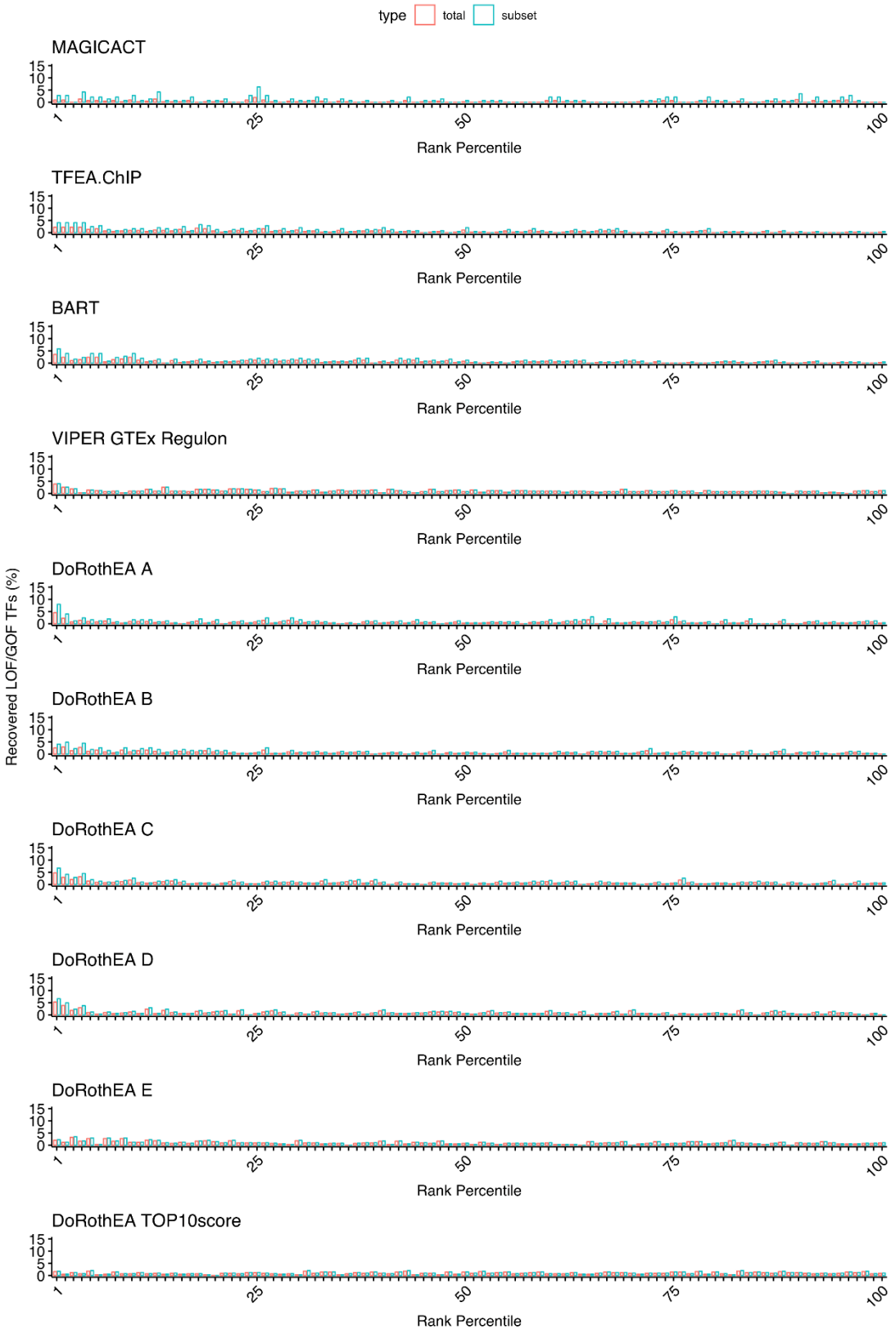
Supplementary Figure 12. Performance of the ChEA3 libraries and integration techniques in recovering the perturbed TFs from 946 TF LOF and GOF experiments from the TFpertGEO1000 benchmark dataset. a) Mean ROC AUC and mean PR AUC over 5,000 bootstrapped ROC and PR curves; b) Composite ROC curves generated from 5,000 bootstrapped curves; c) Composite PR curves generated from 5,000 bootstrapped curves; d) The deviation of the cumulative distribution from uniform of the scaled rankings of each perturbed TF in the benchmarking dataset. Anderson-Darling test of uniformity: MeanRank p = 6.34 x 10^{-7}; TopRank p = 6.34 x 10^{-7}; ARCHS4 p = 6.34 x 10^{-7}; ENCODE p =2.06 x 10^{-6}; Enrichr Queries p = 6.83 x 10^{-7}; GTEx p = 6.45 x 10^{-7}; Literature ChIP-seq p = 1.27 x 10^{-6}; ReMap p = 1.02 x 10^{-6}.



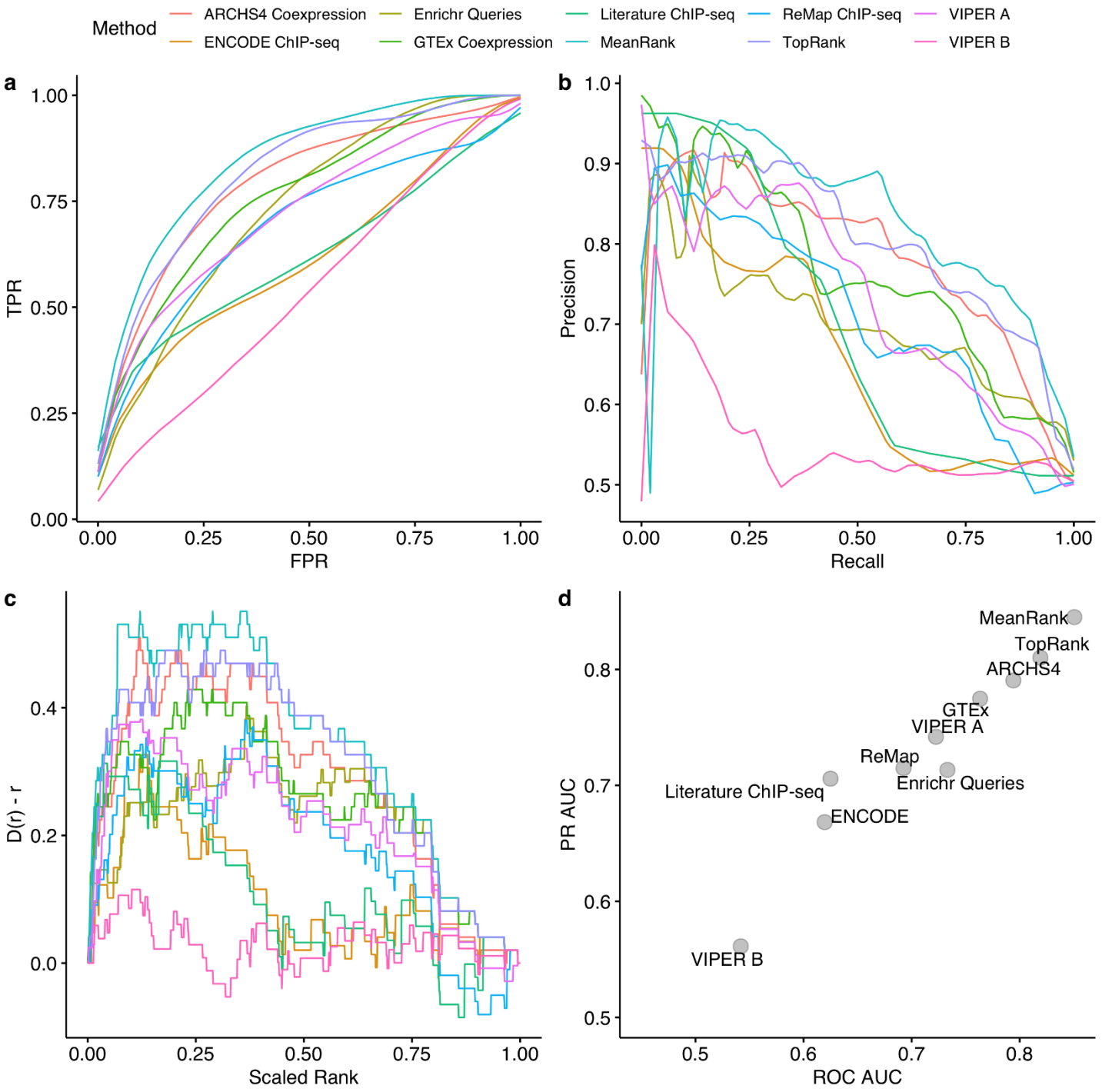
Supplementary Figure 13. Performance of the ChEA3 libraries and integration techniques in recovering the perturbed TFs from 946 TF LOF and GOF experiments from the TFpertGEO1000 benchmark dataset compared to ChEA3 TF coverage. a) Mean ROC AUC over 5,000 bootstrapped ROC curves compared to library TF coverage. b) Mean PR AUC over 5,000 bootstrapped ROC curves compared to library TF coverage. c) The percent of the LOF/GOF TFs in TFpertGEO1000 recovered in the top 1 percentile of ranks compared to the library TF coverage. d) The percent of the LOF/GOF TFs in the subset of TFpertGEO1000 that are covered by the library that were recovered in the top 1 percentile of ranks compared to the library TF coverage.



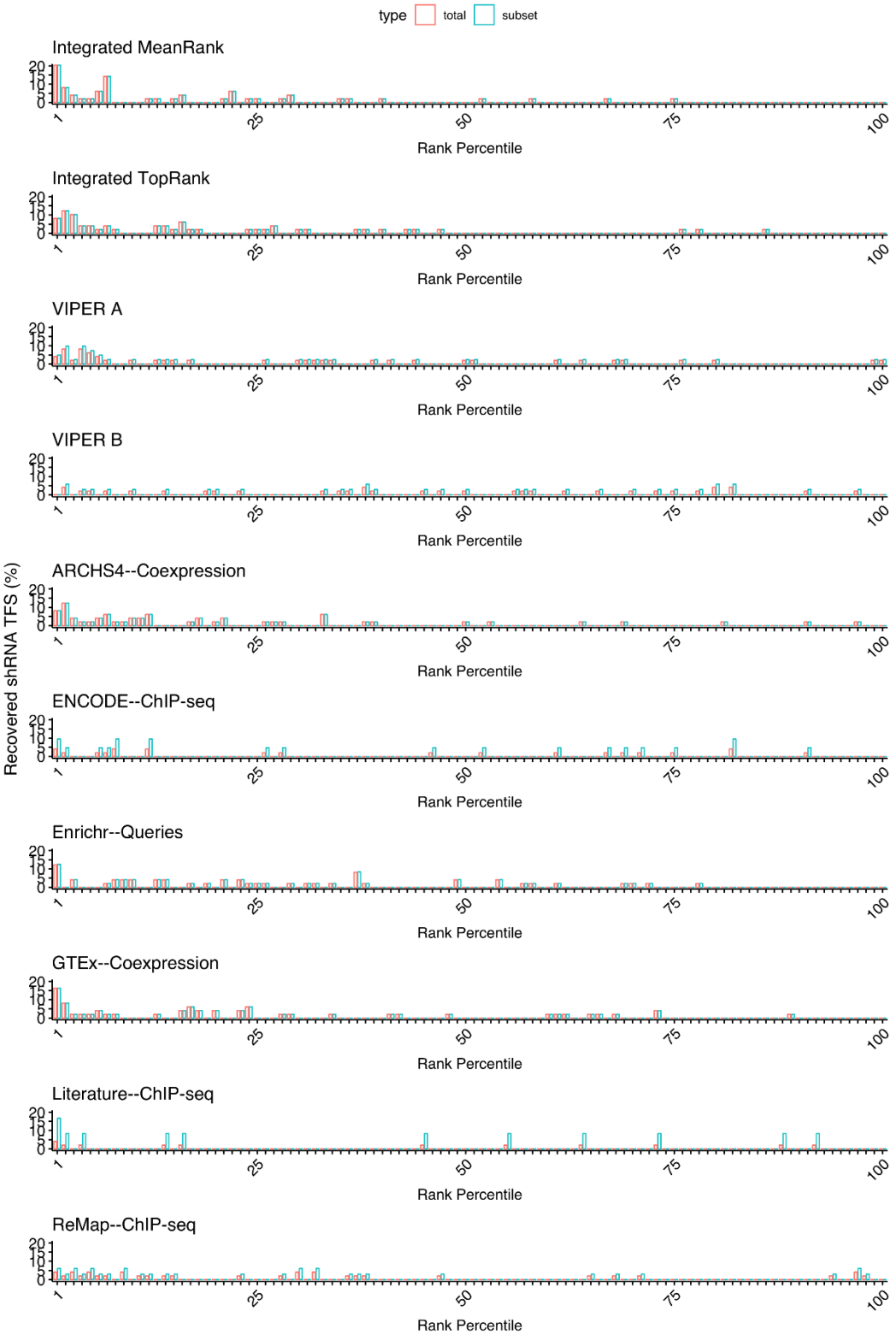
Supplementary Figure 14. Performance of the ChEA3 libraries and integration techniques in recovering the perturbed TFs from 946 TF LOF and GOF experiments from the TFpertGEO1000 benchmarking dataset. The red bars indicate the fraction of the 946 experiments in TFpertGEO1000 where the upstream perturbed TF was recovered for each rank percentile. The blue bars indicate the rank percentile distribution of recovery of the subset of TFpertGEO1000 where the perturbed TFs are covered by the library.



Supplementary Figure 15. Performance of other available tools in recovering the perturbed TFs from the 443 TF LOF/GOF experiments from the hsTFpertGEOupdn benchmarking dataset. The red bars indicate the fraction of the 443 experiments in hsTFpertGEOupdn where the upstream perturbed TF was recovered for each rank percentile. The blue bars indicate the rank percentile distribution of recovery of the subset of hsTFpertGEOupdn where the perturbed TFs are covered by the tool.



Supplementary Figure 16. Performance of VIPER as compared to ChEA3 in recovering the perturbed TFs from the 49 TF shRNA experiments from the Cusanovich dataset. ChEA3 libraries were queried with gene sets from setCusanovich while the VIPER regulons were queried with full signatures from sigCusanovich. VIPER A refers to the the ARACNe-AP generated regulon from Cusanovich expression samples, and VIPER B refers to the published B-cell VIPER regulon. a) Composite ROC curves generated from 5,000 bootstrapped curves demonstrating performance on the Cusanovich benchmarking dataset. b) Composite PR curves generated from 5,000 bootstrapped curves demonstrating performance on the Cusanovich benchmarking dataset. c) The deviation of the cumulative distribution from uniform of the scaled rankings of each perturbed TF in the Cusanovich benchmarking dataset. d) Mean AUROC and mean AUPR from 5,000 bootstrapped ROC and 5,000 bootstrapped PR curves.



Supplementary Figure 17. Performance of VIPER as compared to ChEA3 in recovering the perturbed TFs from the 49 TF shRNA experiments from the Cusanovich dataset. ChEA3 libraries were queried with gene sets from setCusanovich and the VIPER regulons were queried with full signatures from sigCusanovich. VIPER A refers to the the ARACNe-AP generated regulon from Cusanovich expression samples, and VIPER B refers to the published B-cell VIPER regulon. The red bars indicate the fraction of the 49 experiments in Cusanovich where the upstream perturbed TF was recovered for each rank percentile. The blue bars indicate the rank percentile distribution of recovery of the subset of Cusanovich where the perturbed TFs are covered by the tool.