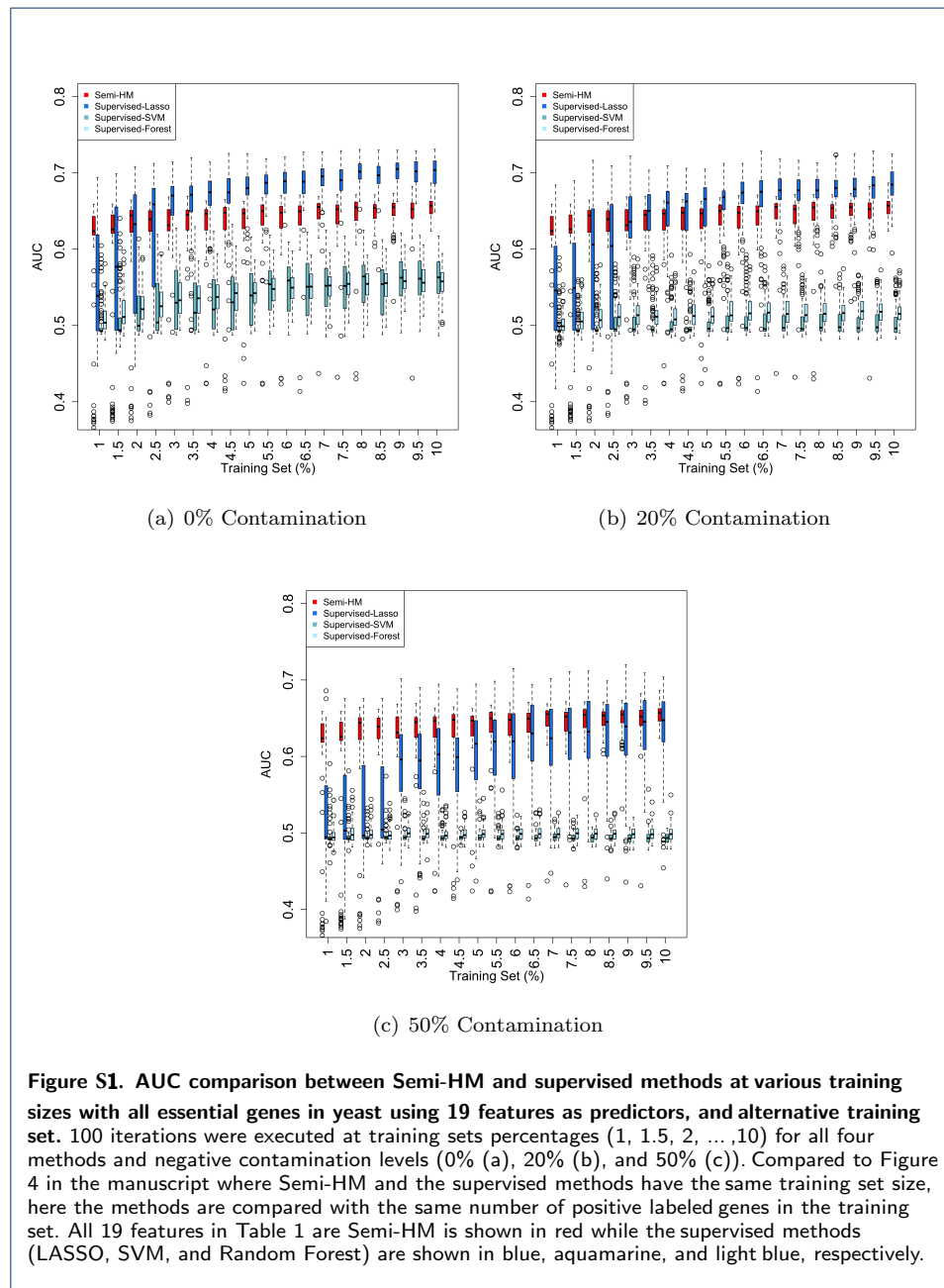
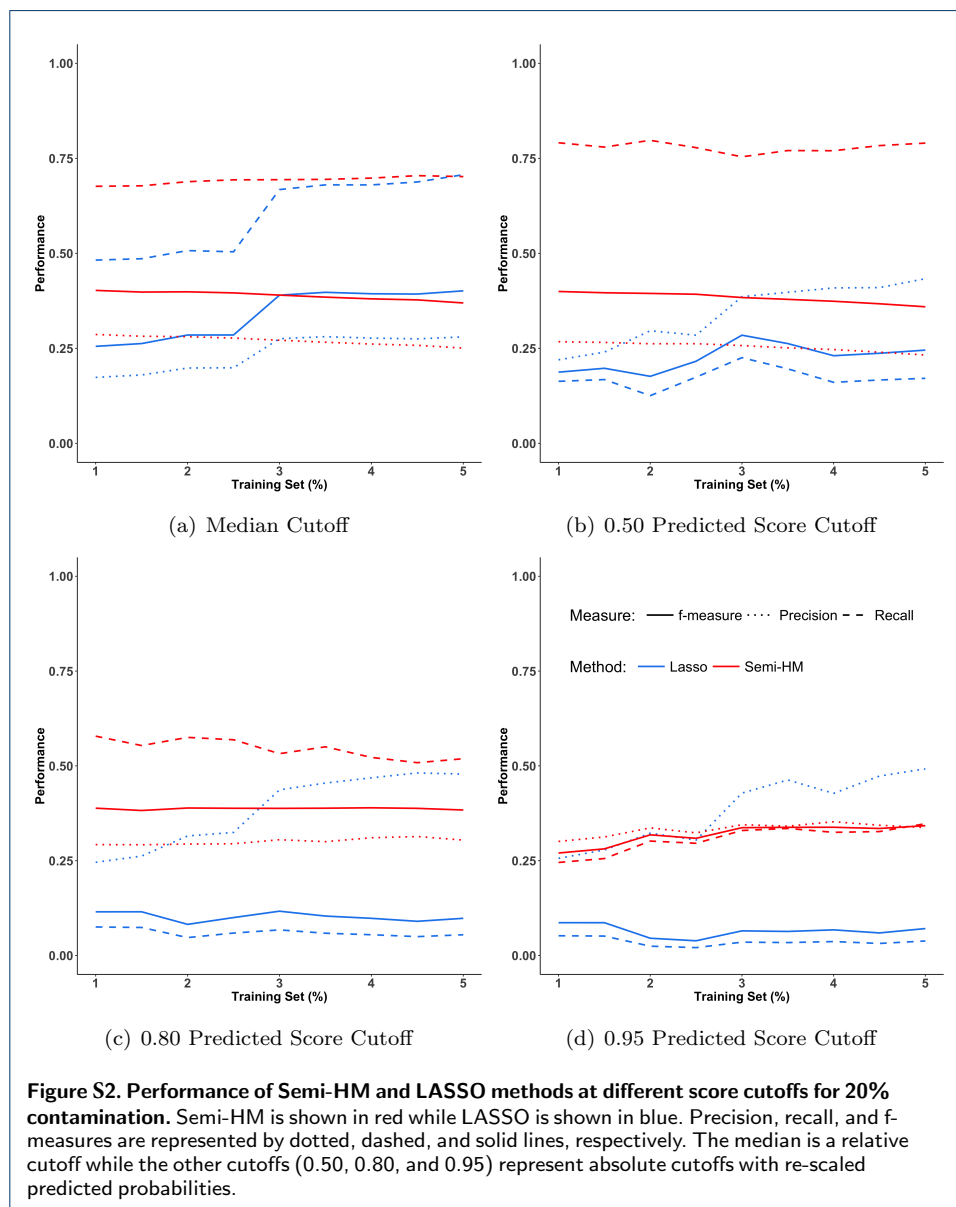
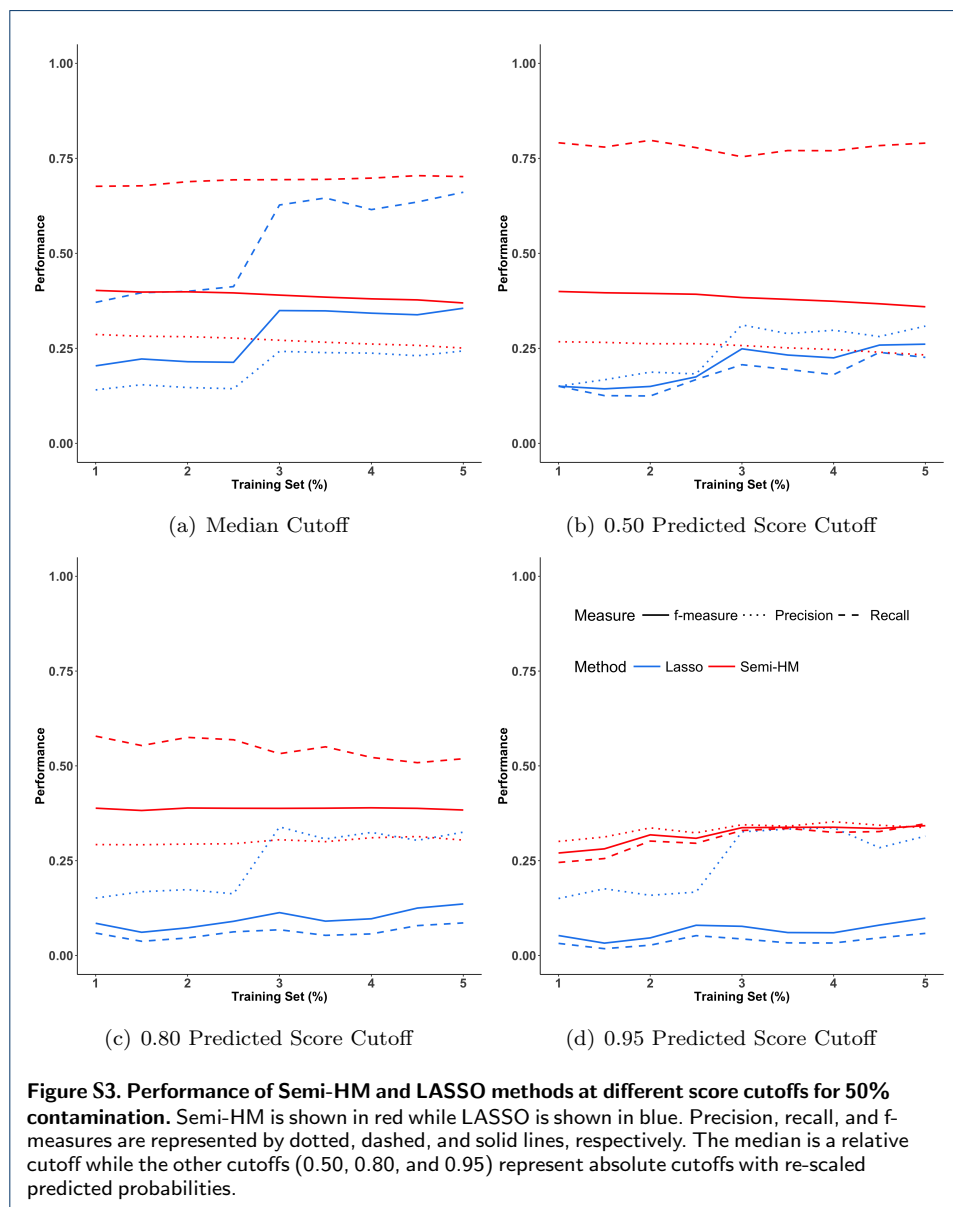


Supplementary File 2: Supplementary Figures and Tables







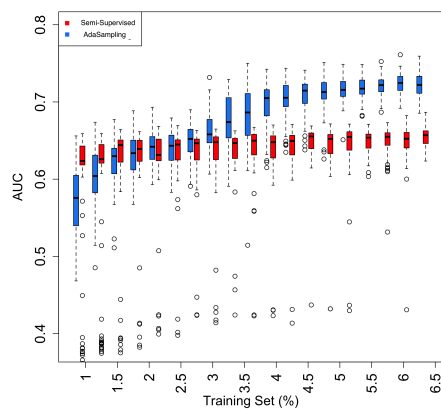


Figure S4. AUC comparison between Semi-HM compared to AdaSampling at various training sizes. 100 iterations were executed at training sets percentages (1, 1.5, 2, ..., 10) for both methods, using the 19 features described in the text. Semi-HM is shown in red while AdaSampling (using the logistic method) is shown in blue. Results are for the case with 600 additional randomly selected genes were considered unlabeled for training.

PANTHER GO-Slim Biological Process	Ref	Obs	Exp	FC	FDR
RNA metabolic process (GO:0016070)	550	101	16.12	6.27	7.24E-51
nucleobase-containing compound metabolic process (GO:0006139)	1025	117	30.04	3.89	1.11E-41
nitrogen compound metabolic process (GO:0006807)	1102	109	32.3	3.38	2.01E-32
primary metabolic process (GO:0044238)	1787	128	52.37	2.44	1.11E-26
metabolic process (GO:0008152)	2118	136	62.07	2.19	9.76E-25
mRNA processing (GO:0006397)	108	35	3.17	11.06	2.29E-22
transcription, DNA-dependent (GO:0006351)	225	45	6.59	6.82	1.02E-21
transcription from RNA polymerase II promoter (GO:0006366)	160	37	4.69	7.89	1.59E-19
Unclassified (UNCLASSIFIED)	3804	48	111.48	0.43	3.25E-18
cellular process (GO:0009987)	2424	129	71.04	1.82	3.03E-15
RNA splicing, via transesterification reactions (GO:0000375)	72	22	2.11	10.43	1.17E-13
mRNA splicing, via spliceosome (GO:0000398)	72	22	2.11	10.43	1.27E-13
regulation of transcription from RNA polymerase II promoter (GO:0006357)	105	23	3.08	7.47	1.06E-11
biosynthetic process (GO:0009058)	819	61	24	2.54	1.07E-10
DNA metabolic process (GO:0006259)	168	26	4.92	5.28	3.03E-10
cellular component biogenesis (GO:0044085)	433	38	12.69	2.99	2.98E-08
DNA replication (GO:0006260)	67	14	1.96	7.13	5.00E-07
mRNA polyadenylation (GO:0006378)	14	8	0.41	19.5	8.13E-07
mRNA 3'-end processing (GO:0031124)	22	9	0.64	13.96	1.09E-06
DNA repair (GO:0006281)	96	15	2.81	5.33	4.06E-06
transcription initiation from RNA polymerase II promoter (GO:0006367)	23	8	0.67	11.87	1.38E-05
rRNA metabolic process (GO:0016072)	110	15	3.22	4.65	1.66E-05
cellular component organization or biogenesis (GO:0071840)	831	49	24.35	2.01	1.66E-05
chromatin organization (GO:0006325)	96	12	2.81	4.27	3.69E-04
protein acetylation (GO:0006473)	29	7	0.85	8.24	3.88E-04
transcription elongation from RNA polymerase II promoter (GO:0006368)	20	6	0.59	10.24	4.67E-04

Table S1. Panther Enrichment Analysis for True Positives. Each row is a significant Panther pathway from the GO-Slim Biological Process category. Only pathways with fold change (FC) > 2 and false discovery rate (FDR) < 1.00E-05 are displayed. Column Ref indicates the number of genes in the pathway in the reference genome, Obs is the observed number of genes in this pathway for the gene list, and Exp is the expected number of genes in this pathway for the gene list based on the frequency in the genome.

PANTHER GO-Slim Biological Process	Ref	Obs	Exp	FC	FDR
RNA metabolic process (GO:0016070)	550	96	40.5	2.37	1.47E-12
protein localization (GO:0008104)	224	50	16.5	3.03	1.60E-09
cellular component biogenesis (GO:0044085)	433	72	31.89	2.26	2.15E-08
protein transport (GO:0015031)	272	52	20.03	2.6	8.82E-08
intracellular protein transport (GO:0006886)	261	50	19.22	2.6	1.68E-07
protein targeting (GO:0006605)	110	29	8.1	3.58	5.46E-07
tRNA metabolic process (GO:0006399)	80	23	5.89	3.9	3.37E-06
cell cycle (GO:0007049)	229	42	16.86	2.49	4.25E-06
rRNA metabolic process (GO:0016072)	110	27	8.1	3.33	4.48E-06
mitosis (GO:0007067)	94	24	6.92	3.47	9.24E-06

Table S2. Panther Enrichment Analysis for False Negatives. Each row is a significant Panther pathway from the GO-Slim Biological Process category. Only pathways with fold change (FC) > 2 and false discovery rate (FDR) < 1.00E-05 are displayed. Column Ref indicates the number of genes in the pathway in the reference genome, Obs is the observed number of genes in this pathway for the gene list, and Exp is the expected number of genes in this pathway for the gene list based on the frequency in the genome.