Extended evaluation of differential chromatin interaction detection analysis using real Hi-C data

John Stansfield, Mikhail Dozmorov

Contents

Introduction

To evaluate the performance of different normalization methods on the detection of chromatin interaction differences, controlled changes were used. Real Hi-C data is used: GM12878, Chromosome 1, MboI vs. DpnII restriction enzymes, 1Mb resolution. The dimensions of the chromatin interaction matrices are 250 x 250. The performance of the joint (`loess`) and individual (`ChromoR`, Iterative Correction and Eigenvector decomposition - `ICE`, Knight-Ruiz - `KR`, Sequential Component Normalization - `SCN`) normalization methods at varying fold changes (1.5 by default) and varying number of controlled changes (500 by default) is investigated. The data were globally rescaled to have the same total count of interaction frequencies.

Fold changes are applied to one of the datasets by up-regulating the selected IF if the difference between the datasets is positive. If the difference between the datasets at that point is negative the IF is down-regulated by the specified fold change. This method of making changes ensures that the fold change specified is actually realized on the MD plot.

The effect of fold change

1.5 Fold change

	loess	chromoR	ice	kr	scn
true positive	211	14	11	13	7
false positive	1000	1440	1150	1200	1200
true negative	24900	29400	23800	24700	24700
false negative	289	486	467	487	493
Total	26400	31400	25400	26400	26400
TPR	0.422	0.028	0.023	0.026	0.014
SPC	0.961	0.953	0.954	0.954	0.954
F1	0.975	0.968	0.967	0.967	0.967
AUC	0.855	0.541	0.766	0.752	0.701
AUC 20%	0.115	0.0216	0.0786	0.0761	0.057
FDR	0.826	0.99	0.991	0.989	0.994
Accuracy	0.951	0.939	0.936	0.936	0.936
Precision	0.174	0.00964	0.00944	0.0107	0.0058
FPR	0.0388	0.0466	0.0463	0.0463	0.0463
FNR	0.578	0.972	0.977	0.974	0.986
FOR	0.0115	0.0162	0.0193	0.0193	0.0196
NPV	0.989	0.984	0.981	0.981	0.98

2.0 Fold change

	loess	chromoR	ice	kr	scn
true positive	369	9	39	42	50
false positive	856	1450	1130	1180	1160
true negative	25000	29400	23800	24700	24700
false negative	131	491	440	458	450
Total	26400	31400	25400	26400	26400
TPR	0.738	0.018	0.0814	0.084	0.1
SPC	0.967	0.953	0.955	0.954	0.955
F1	0.981	0.968	0.968	0.968	0.968
AUC	0.963	0.582	0.57	0.576	0.598
AUC 20%	0.169	0.0256	0.0317	0.0329	0.0355
FDR	0.699	0.994	0.967	0.966	0.959
Accuracy	0.963	0.938	0.938	0.938	0.939
Precision	0.301	0.00617	0.0334	0.0344	0.0413
FPR	0.0331	0.0469	0.0452	0.0455	0.0449
FNR	0.262	0.982	0.919	0.916	0.9
FOR	0.0052	0.0164	0.0181	0.0182	0.0179
NPV	0.995	0.984	0.982	0.982	0.982

4.0 Fold change

	loess	chromoR	ice	kr	scn
true positive	477	82	386	397	340
false positive	735	1380	792	827	867
true negative	25200	29500	24200	25100	25000
false negative	23	418	93	103	160
Total	26400	31400	25400	26400	26400
TPR	0.954	0.164	0.806	0.794	0.68
SPC	0.972	0.955	0.968	0.968	0.967
F1	0.985	0.971	0.982	0.982	0.98
AUC	0.992	0.508	0.972	0.966	0.95
AUC 20%	0.192	0.0211	0.177	0.173	0.163
FDR	0.606	0.944	0.672	0.676	0.718
Accuracy	0.971	0.943	0.965	0.965	0.961
Precision	0.394	0.0563	0.328	0.324	0.282
FPR	0.0284	0.0445	0.0317	0.0319	0.0335
FNR	0.046	0.836	0.194	0.206	0.32
FOR	0.000913	0.014	0.00384	0.00409	0.00635
NPV	0.999	0.986	0.996	0.996	0.994

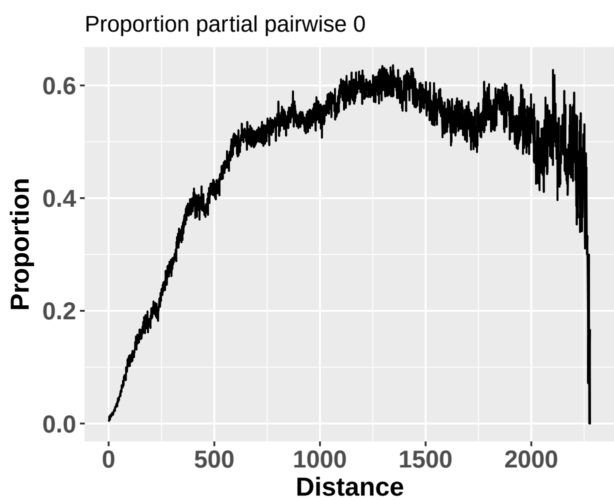
10.0 fold change

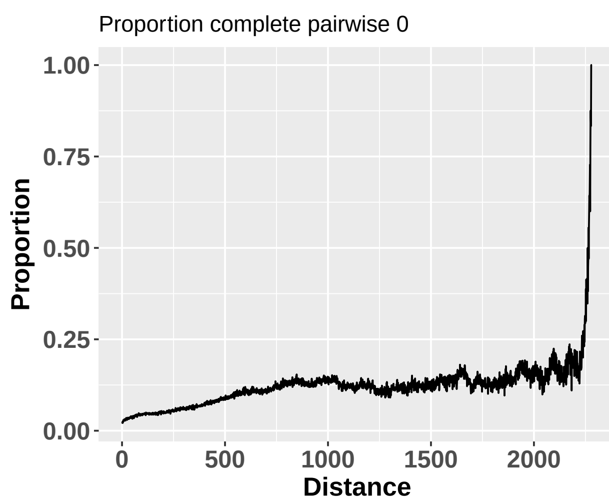
	loess	chromoR	ice	kr	scn
true positive	492	364	471	490	480
false positive	726	1090	698	730	728
true negative	25200	29800	24200	25200	25200
false negative	8	136	10	10	20

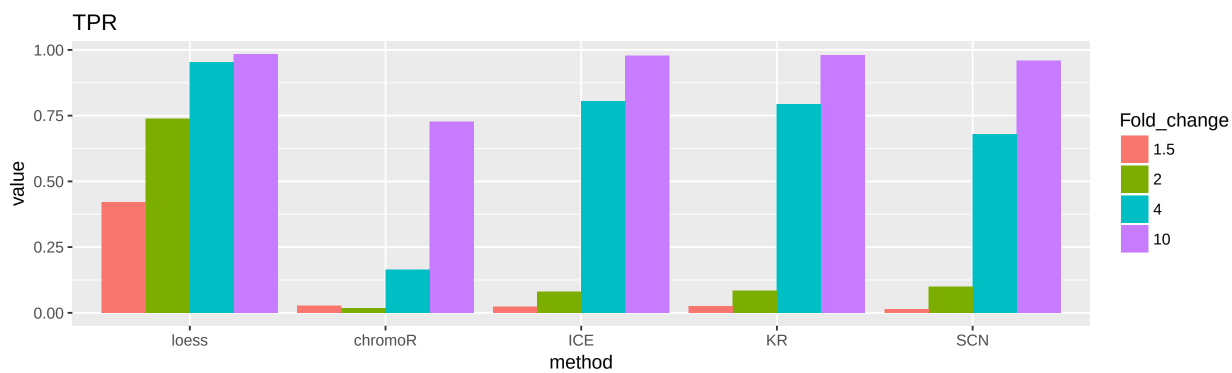
	loess	chromoR	ice	kr	scn
Total	26400	31400	25400	26400	26400
TPR	0.984	0.728	0.979	0.98	0.96
SPC	0.972	0.965	0.972	0.972	0.972
F1	0.986	0.98	0.986	0.986	0.985
AUC	0.995	0.864	0.994	0.994	0.987
AUC 20%	0.195	0.148	0.195	0.195	0.189
FDR	0.596	0.749	0.597	0.598	0.603
Accuracy	0.972	0.961	0.972	0.972	0.972
Precision	0.404	0.251	0.403	0.402	0.397
FPR	0.028	0.0352	0.028	0.0282	0.0281
FNR	0.016	0.272	0.0208	0.02	0.04
FOR	0.000318	0.00454	0.000412	0.000397	0.000794
NPV	1	0.995	1	1	0.999

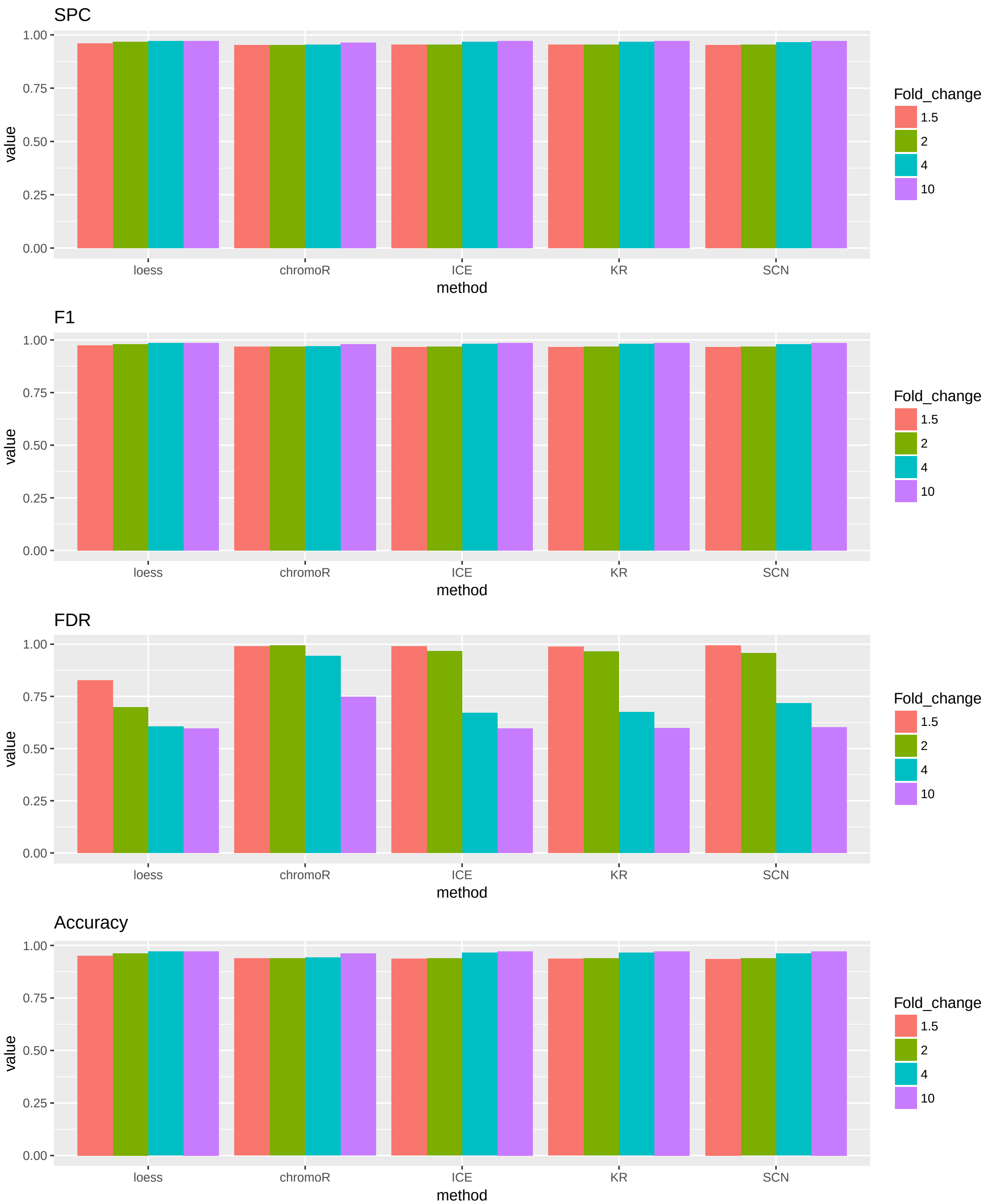
Bar plots

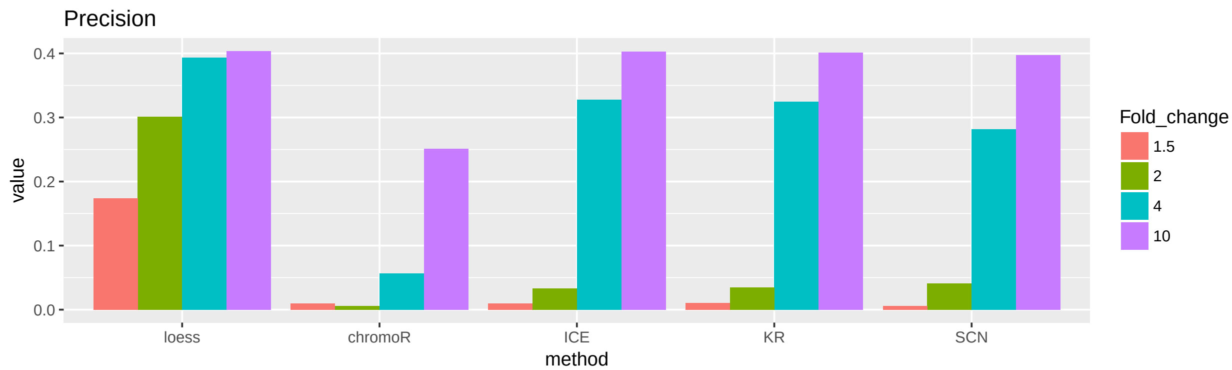
The bar plots show comparisons of the effect of different fold changes using fixed numbers of controlled changes on different performance metrics.








SPC
F1
FDR
Accuracy
value
1.00
0.75
0.50
0.25
0.00
loess
chromoR
ICE
KR
SCN
method
Fold_change
1.5
2
4
10



Summary

For the real data with changes added, `loess` was once again able to detect the most true changes compared to the other normalization methods. `loess` was clearly superior at the lower fold changes (1.5 and 2) with much higher TPRs compared to the other methods with TPRs below 10%. `loess` also had the lowest number of false negatives compared to the other methods. The gap in detection once again began to close at the higher fold changes between `loess`, `KR`, `SCN`, and `ICE` however `loess` still was slightly better. `ChromoR` once again proved to be the worst normalization method and had the lowest detection rates.

The effect of introduced number of changes

A given number of interaction frequencies were increased or decreased to produce a 1.5-fold change.

1 change

	loess	chromoR	ice	kr	scn
true positive	1	0	0	0	0
false positive	1200	1460	1180	1220	1220
true negative	25200	29900	24200	25200	25200
false negative	0	1	1	1	1
Total	26400	31400	25400	26400	26400
TPR	1	0	0	0	0
SPC	0.955	0.953	0.954	0.954	0.954
F1	0.977	0.976	0.976	0.976	0.976
AUC	0.981	0.54	0.996	0.996	0.77
AUC 20%	0.181	0	0.196	0.196	0
FDR	0.999	1	1	1	1
Accuracy	0.955	0.953	0.954	0.954	0.954
Precision	0.000833	0	0	0	0
FPR	0.0455	0.0466	0.0463	0.0462	0.0463
FNR	0	1	1	1	1
FOR	0	3.34e-05	4.12e-05	3.97e-05	3.97e-05
NPV	1	1	1	1	1

100 changes

	loess	chromoR	ice	kr	scn
true positive	51	3	4	4	4
false positive	1170	1460	1160	1220	1210
true negative	25100	29800	24200	25100	25100
false negative	49	97	95	96	96
Total	26400	31400	25400	26400	26400
TPR	0.51	0.03	0.0404	0.04	0.04
SPC	0.956	0.953	0.954	0.954	0.954
F1	0.976	0.975	0.975	0.975	0.975
AUC	0.877	0.542	0.772	0.766	0.72
AUC 20%	0.122	0.0208	0.0907	0.089	0.0734
FDR	0.958	0.998	0.997	0.997	0.997
Accuracy	0.954	0.95	0.951	0.95	0.95
Precision	0.0419	0.00205	0.00344	0.00328	0.00329
FPR	0.0444	0.0467	0.0457	0.0462	0.0461
FNR	0.49	0.97	0.96	0.96	0.96
FOR	0.00195	0.00324	0.00392	0.00381	0.00381
NPV	0.998	0.997	0.996	0.996	0.996

200 changes

	loess	chromoR	ice	kr	scn
true positive	93	2	2	2	3
false positive	1110	1460	1160	1210	1210
true negative	25100	29700	24100	25000	25000
false negative	107	198	189	198	197
Total	26400	31400	25400	26400	26400
TPR	0.465	0.01	0.0105	0.01	0.015
SPC	0.958	0.953	0.954	0.954	0.954
F1	0.976	0.973	0.973	0.973	0.973
AUC	0.87	0.547	0.755	0.738	0.699
AUC 20%	0.118	0.0241	0.0733	0.0706	0.0635
FDR	0.923	0.999	0.998	0.998	0.998
Accuracy	0.954	0.947	0.947	0.947	0.947
Precision	0.0773	0.00136	0.00173	0.00165	0.00247
FPR	0.0424	0.047	0.0458	0.0463	0.0462
FNR	0.535	0.99	0.99	0.99	0.985
FOR	0.00425	0.00662	0.00779	0.00786	0.00782
NPV	0.996	0.993	0.992	0.992	0.992

1000 changes

	loess	chromoR	ice	kr	scn
true positive	340	33	7	11	18
false positive	869	1430	1160	1190	1200
true negative	24500	28900	23300	24200	24200
false negative	660	967	953	989	982
Total	26400	31400	25400	26400	26400
TPR	0.34	0.033	0.00729	0.011	0.018

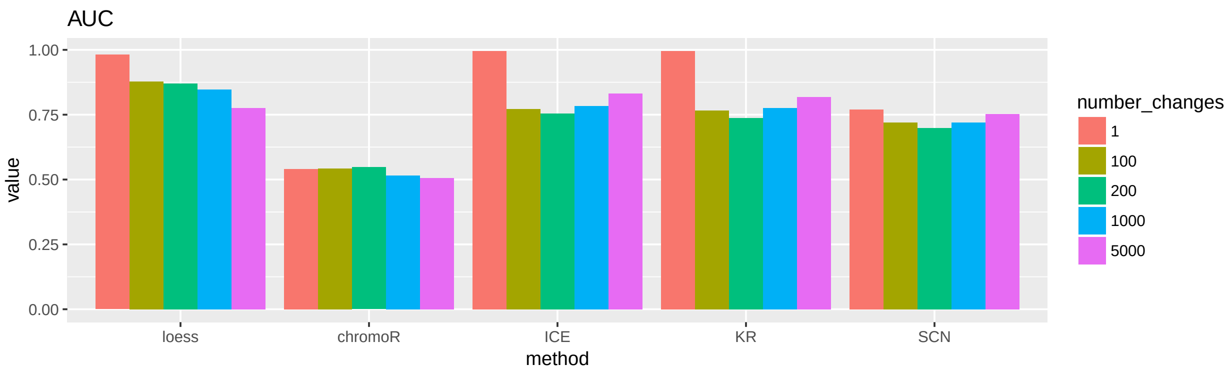
	loess	chromoR	ice	kr	scn
SPC	0.966	0.953	0.953	0.953	0.953
F1	0.97	0.96	0.957	0.957	0.957
AUC	0.847	0.515	0.784	0.776	0.72
AUC 20%	0.11	0.0161	0.0825	0.0805	0.0616
FDR	0.719	0.977	0.994	0.991	0.985
Accuracy	0.942	0.924	0.917	0.917	0.918
Precision	0.281	0.0225	0.006	0.00914	0.0148
FPR	0.0342	0.0472	0.0474	0.047	0.047
FNR	0.66	0.967	0.993	0.989	0.982
FOR	0.0262	0.0323	0.0393	0.0393	0.039
NPV	0.974	0.968	0.961	0.961	0.961

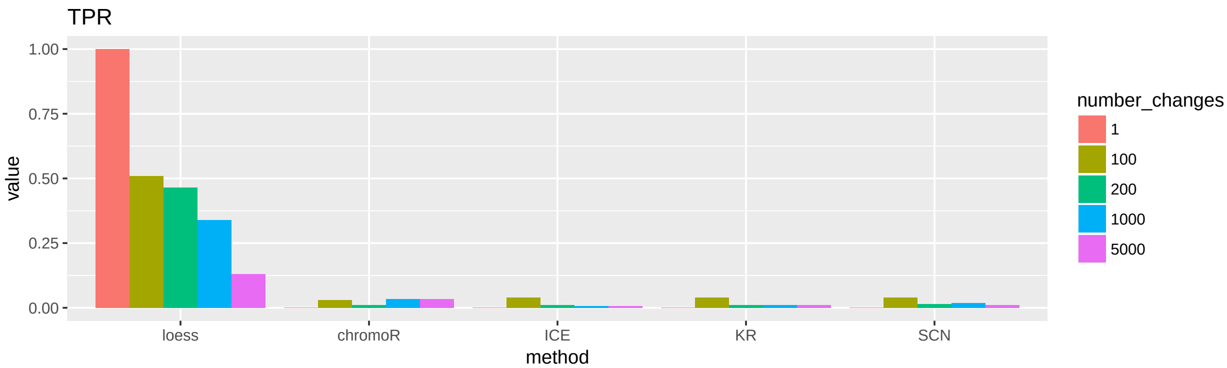
5000 changes

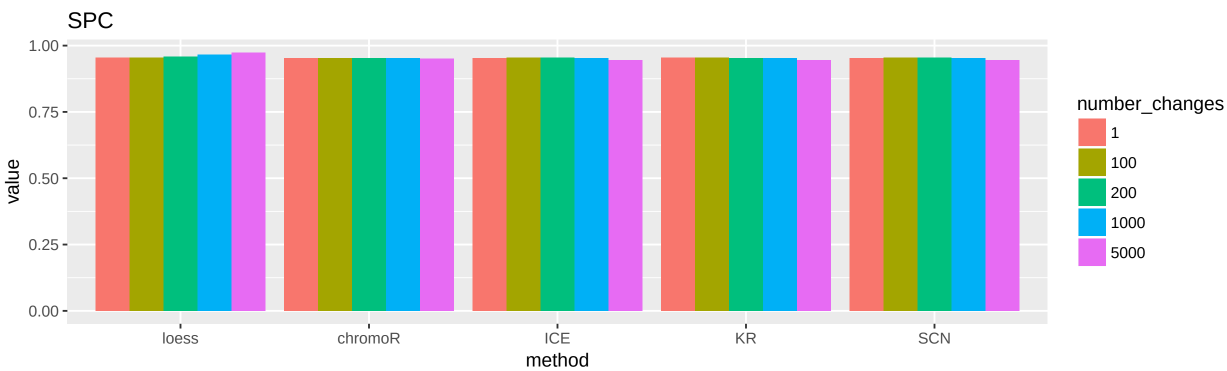
	loess	chromoR	ice	kr	scn
true positive	657	165	29	56	56
false positive	557	1290	1130	1170	1160
true negative	20800	25100	19500	20200	20200
false negative	4340	4840	4780	4940	4940
Total	26400	31400	25400	26400	26400
TPR	0.131	0.033	0.00603	0.0112	0.0112
SPC	0.974	0.951	0.945	0.945	0.946
F1	0.895	0.891	0.868	0.869	0.869
AUC	0.775	0.507	0.832	0.819	0.752
AUC 20%	0.0795	0.0152	0.0977	0.0945	0.0669
FDR	0.459	0.887	0.975	0.954	0.954
Accuracy	0.814	0.805	0.767	0.769	0.769
Precision	0.541	0.113	0.025	0.0458	0.0459
FPR	0.026	0.0491	0.055	0.0545	0.0544
FNR	0.869	0.967	0.994	0.989	0.989
FOR	0.172	0.162	0.197	0.196	0.196
NPV	0.828	0.838	0.803	0.804	0.804

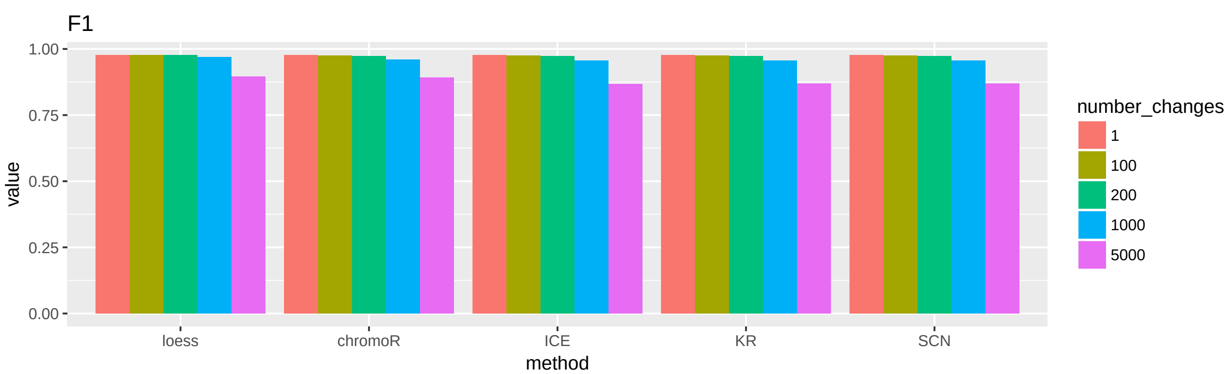
Bar plots

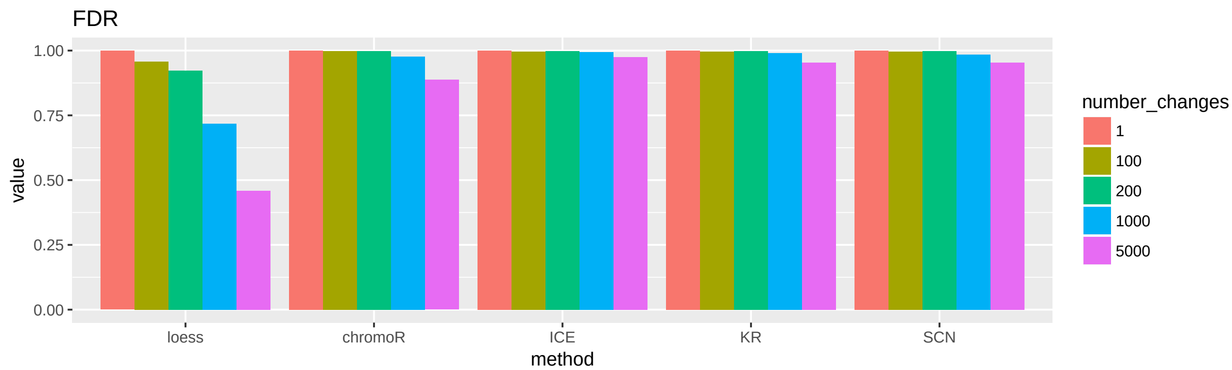
Below are bar plots showing comparisons of the different normalization methods over the varying numbers of changes at a fixed fold change for selected metrics.

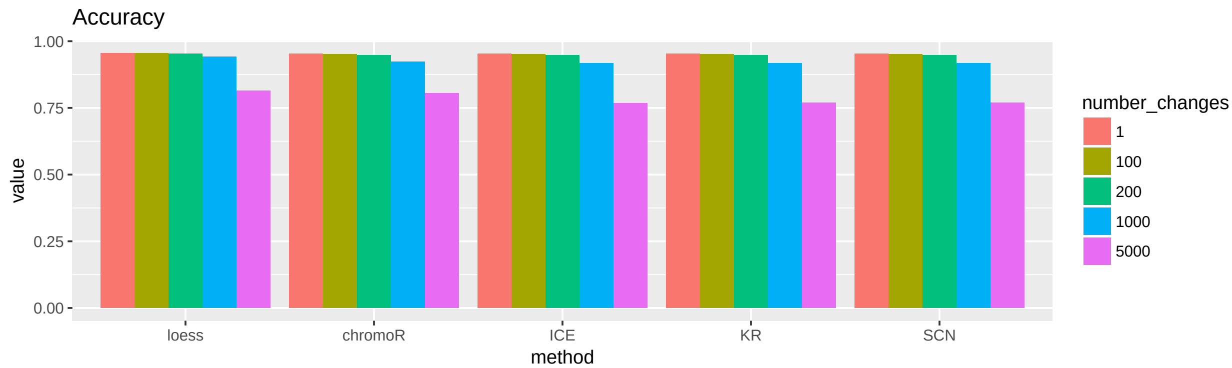

AUC
1.00
0.75
0.50
0.25
0.00
value
loess
chromoR
ICE
KR
SCN
method
number_changes
1
100
200
1000
5000

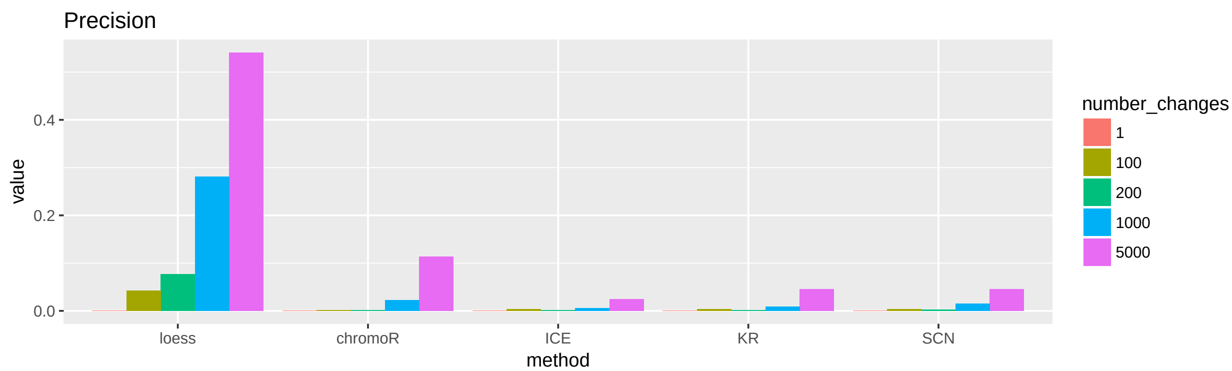

TPR
1.00
0.75
0.50
0.25
0.00
value
loess
chromoR
ICE
KR
SCN
method
number_changes
1
100
200
1000
5000


SPC
1.00
0.75
0.50
0.25
0.00
value
loess
chromoR
ICE
KR
SCN
method
number_changes
1
100
200
1000
5000


F1
1.00
0.75
0.50
0.25
0.00
value
loess
chromoR
ICE
KR
SCN
method
number_changes
1
100
200
1000
5000







Summary

`loess` was able to detect a single true change added at a 1.5 fold change. `loess` also detected the most changes that were added for all levels changes made compared to the other methods. `loess` again had the lowest numbers of false negatives compared to the other methods. Interestingly, `ChromoR` seemed to detect more differences when a large number of changes were made (1000 and 5000) compared to `KR`, `SCN`, and `ICE`.