Extended evaluation of differential chromatin interaction detection analysis using simulated 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. To better control for existing differences in the real Hi-C data, simulated Hi-C datasets were used. The data was simulated using the `hic_simulate` function. For the simulated matrices, the default values were used (see `?hic_simulate`) except for changing the size to be a 250 x 250 matrix. The effect of varying fold changes (1.5 by default) and varying number of controlled changes (500 by default) is investigated.

Fold changes are applied to one of the datasets by upregulating 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 downregulated 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	112	22	87	89	68
false positive	1060	1440	1090	1130	1140
true negative	24800	29400	24100	24800	24700
false negative	388	478	396	411	432
Total	26400	31400	25700	26400	26400
TPR	0.224	0.044	0.18	0.178	0.136
SPC	0.959	0.953	0.957	0.957	0.956
F1	0.972	0.968	0.97	0.97	0.969
AUC	0.749	0.503	0.667	0.669	0.624
AUC 20%	0.0768	0.0201	0.0614	0.0608	0.0471
FDR	0.904	0.985	0.926	0.927	0.944
Accuracy	0.945	0.939	0.942	0.942	0.94
Precision	0.0959	0.0151	0.0737	0.0733	0.0561
FPR	0.0408	0.0466	0.0435	0.0435	0.0442
FNR	0.776	0.956	0.82	0.822	0.864
FOR	0.0154	0.016	0.0162	0.0163	0.0172
NPV	0.985	0.984	0.984	0.984	0.983

2.0 Fold change

	loess	chromoR	ice	kr	scn
true positive	287	33	200	201	143
false positive	890	1430	985	1020	1070
true negative	25100	29400	24300	24900	24900
false negative	213	467	289	299	357
Total	26500	31400	25700	26500	26500
TPR	0.574	0.066	0.409	0.402	0.286
SPC	0.966	0.954	0.961	0.961	0.959
F1	0.978	0.969	0.974	0.974	0.972
AUC	0.912	0.516	0.785	0.785	0.727
AUC 20%	0.141	0.0251	0.101	0.1	0.0806
FDR	0.756	0.977	0.831	0.835	0.882
Accuracy	0.958	0.94	0.95	0.95	0.946
Precision	0.244	0.0226	0.169	0.165	0.118
FPR	0.0343	0.0463	0.039	0.0391	0.0412
FNR	0.426	0.934	0.591	0.598	0.714
FOR	0.00843	0.0156	0.0118	0.0118	0.0141
NPV	0.992	0.984	0.988	0.988	0.986

4.0 Fold change

	loess	chromoR	ice	kr	scn
true positive	483	56	421	432	402
false positive	700	1400	754	783	807
true negative	25200	29500	24200	25100	25100
false negative	17	444	66	68	98
Total	26400	31400	25500	26400	26400
TPR	0.966	0.112	0.864	0.864	0.804
SPC	0.973	0.955	0.97	0.97	0.969
F1	0.986	0.97	0.983	0.983	0.982
AUC	0.993	0.488	0.968	0.967	0.958
AUC 20%	0.194	0.0232	0.179	0.179	0.171
FDR	0.592	0.961	0.642	0.644	0.667
Accuracy	0.973	0.941	0.968	0.968	0.966
Precision	0.408	0.0386	0.358	0.356	0.333
FPR	0.0271	0.0452	0.0302	0.0303	0.0312
FNR	0.034	0.888	0.136	0.136	0.196
FOR	0.000675	0.0148	0.00272	0.0027	0.00389
NPV	0.999	0.985	0.997	0.997	0.996

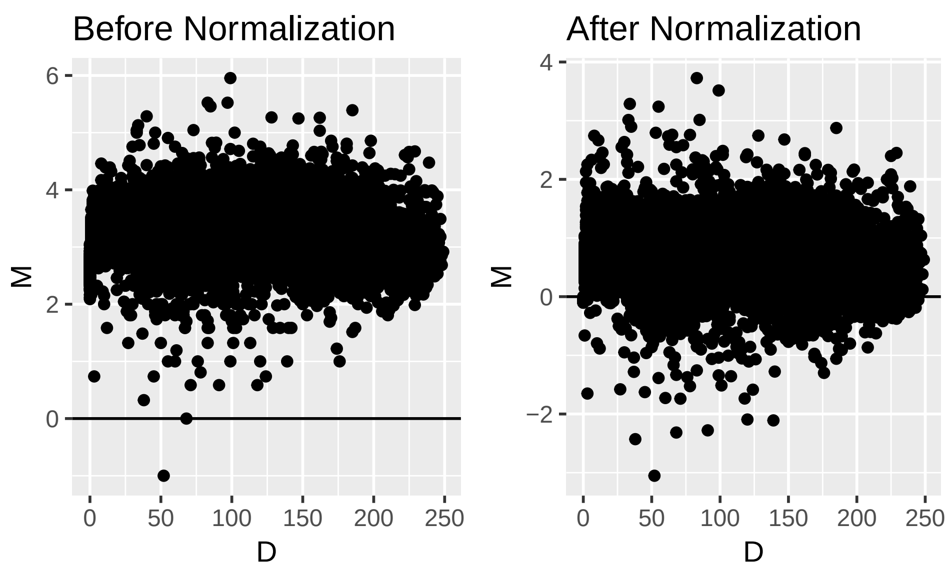
10.0 fold change

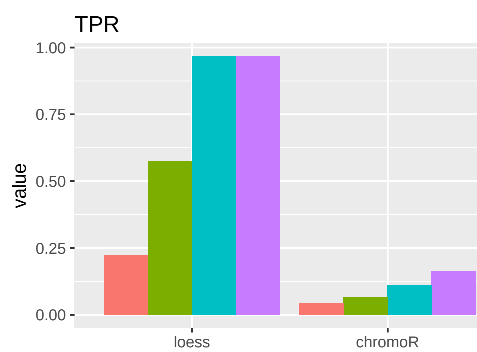
	loess	chromoR	ice	kr	scn
true positive	483	82	475	484	476
false positive	710	1390	722	735	746
true negative	25300	29500	24700	25200	25200
false negative	17	418	17	16	24

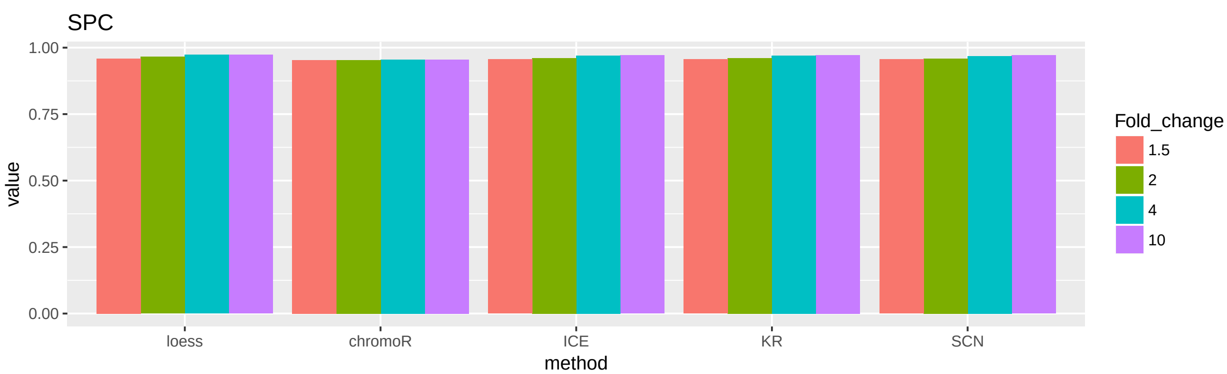
	loess	chromoR	ice	kr	scn
Total	26500	31400	25900	26500	26500
TPR	0.966	0.164	0.965	0.968	0.952
SPC	0.973	0.955	0.972	0.972	0.971
F1	0.986	0.97	0.985	0.985	0.985
AUC	0.995	0.543	0.994	0.994	0.99
AUC 20%	0.195	0.0424	0.194	0.194	0.192
FDR	0.595	0.944	0.603	0.603	0.61
Accuracy	0.973	0.942	0.971	0.972	0.971
Precision	0.405	0.0557	0.397	0.397	0.39
FPR	0.0273	0.045	0.0284	0.0283	0.0287
FNR	0.034	0.836	0.0346	0.032	0.048
FOR	0.000672	0.014	0.000688	0.000633	0.00095
NPV	0.999	0.986	0.999	0.999	0.999

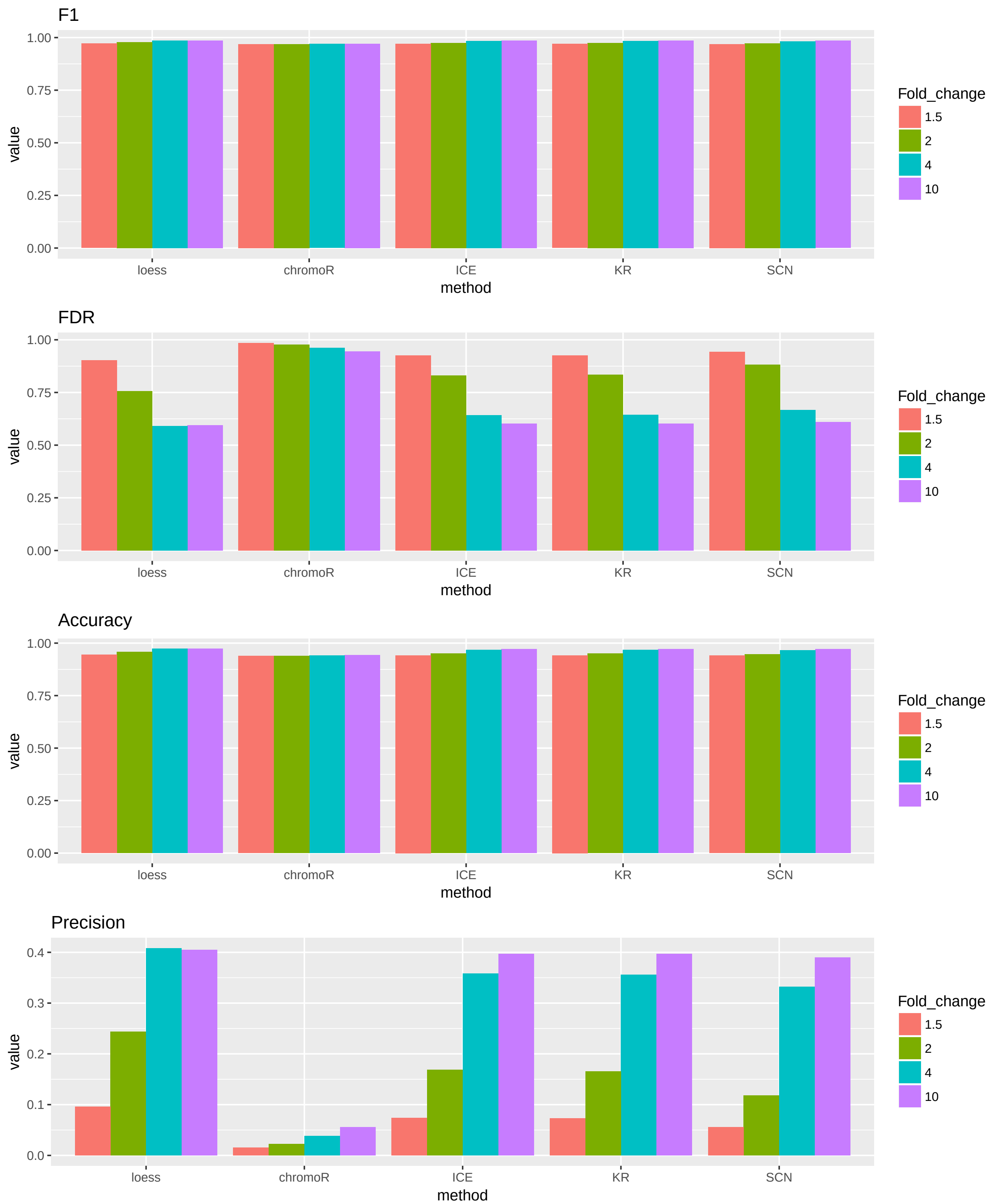
Bar plots

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









Summary

The tables show that the most true differences are detected after joint `loess` normalization compared to the other normalization techniques. `loess` also has the lowest number of false positives among the normalization techniques. With `loess`, differences at smaller fold changes (1.5 and 2) are able to be detected more reliably compared to the other methods and this superiority continues to the higher fold changes though the other methods tend to make up some ground. `ChromoR` perofrmed the worst of the normalization techniques tested while `KR` and `SCN` tended to perform better.

The effect of introduced number of changes

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

1 change

	loess	chromoR	ice	kr	scn
true positive	0	0	0	0	0
false positive	1160	1460	1200	1220	1220
true negative	25300	29900	24900	25300	25300
false negative	1	1	1	1	1
Total	26500	31400	26100	26500	26500
TPR	0	0	0	0	0
SPC	0.956	0.953	0.954	0.954	0.954
F1	0.978	0.976	0.977	0.976	0.976
AUC	0.876	0.788	0.819	0.83	0.828
AUC 20%	0.0756	0	0.0189	0.0301	0.0283
FDR	1	1	1	1	1
Accuracy	0.956	0.953	0.954	0.954	0.954
Precision	0	0	0	0	0
FPR	0.0437	0.0465	0.0458	0.0461	0.046
FNR	1	1	1	1	1
FOR	3.95e-05	3.34e-05	4.01e-05	3.96e-05	3.96e-05
NPV	1	1	1	1	1

100 changes

	loess	chromoR	ice	kr	scn
true positive	23	3	16	18	14
false positive	1140	1450	1140	1190	1190
true negative	25000	29800	24300	25000	25000
false negative	77	97	82	82	86
Total	26300	31400	25600	26300	26300
TPR	0.23	0.03	0.163	0.18	0.14
SPC	0.956	0.954	0.955	0.955	0.954
F1	0.976	0.975	0.975	0.975	0.975
AUC	0.735	0.537	0.63	0.63	0.578
AUC 20%	0.0723	0.026	0.0464	0.0479	0.0401
FDR	0.98	0.998	0.986	0.985	0.988
Accuracy	0.954	0.951	0.952	0.952	0.951
Precision	0.0197	0.00207	0.0138	0.0149	0.0116
FPR	0.0436	0.0463	0.0449	0.0455	0.0455
FNR	0.77	0.97	0.837	0.82	0.86
FOR	0.00306	0.00324	0.00336	0.00327	0.00343
NPV	0.997	0.997	0.997	0.997	0.997

200 changes

	loess	chromoR	ice	kr	scn
true positive	45	9	37	43	41
false positive	1130	1460	1140	1170	1170
true negative	25100	29700	24200	25100	25100
false negative	155	191	154	157	159
Total	26400	31400	25500	26400	26400
TPR	0.225	0.045	0.194	0.215	0.205
SPC	0.957	0.953	0.955	0.955	0.955
F1	0.975	0.973	0.974	0.974	0.974
AUC	0.763	0.501	0.677	0.681	0.652
AUC 20%	0.0785	0.0225	0.0612	0.0623	0.0573
FDR	0.962	0.994	0.969	0.965	0.966
Accuracy	0.951	0.947	0.949	0.95	0.95
Precision	0.0383	0.00612	0.0315	0.0354	0.0339
FPR	0.043	0.0469	0.045	0.0447	0.0445
FNR	0.775	0.955	0.806	0.785	0.795
FOR	0.00613	0.00639	0.00632	0.00622	0.0063
NPV	0.994	0.994	0.994	0.994	0.994

1000 changes

	loess	chromoR	ice	kr	scn
true positive	235	48	163	168	134
false positive	940	1420	1010	1030	1080
true negative	24400	28900	23600	24300	24200
false negative	765	952	802	832	866
Total	26300	31400	25600	26300	26300
TPR	0.235	0.048	0.169	0.168	0.134
SPC	0.963	0.953	0.959	0.959	0.957
F1	0.966	0.961	0.963	0.963	0.961
AUC	0.735	0.511	0.639	0.636	0.599
AUC 20%	0.0716	0.0233	0.0542	0.0543	0.0439
FDR	0.8	0.967	0.861	0.86	0.89
Accuracy	0.935	0.925	0.929	0.929	0.926
Precision	0.2	0.0328	0.139	0.14	0.11
FPR	0.0371	0.0466	0.0411	0.0408	0.0428
FNR	0.765	0.952	0.831	0.832	0.866
FOR	0.0304	0.0318	0.0328	0.0331	0.0345
NPV	0.97	0.968	0.967	0.967	0.966

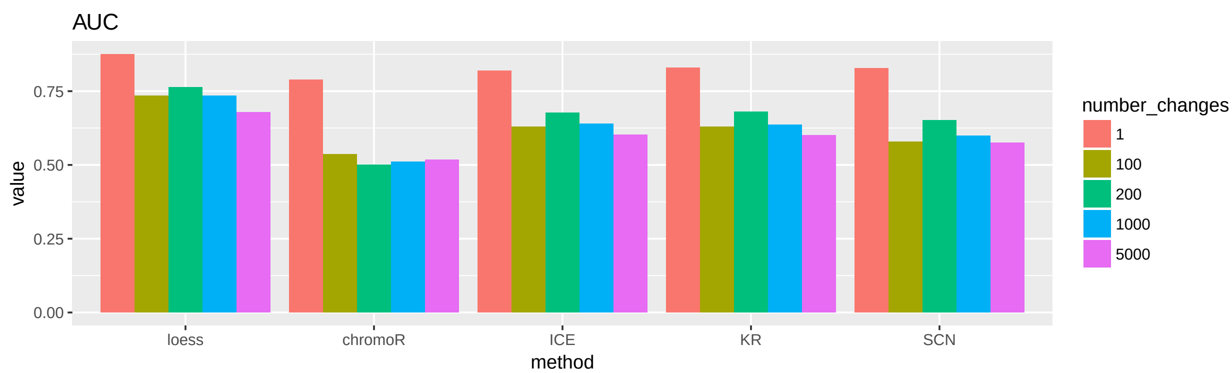
5000 changes

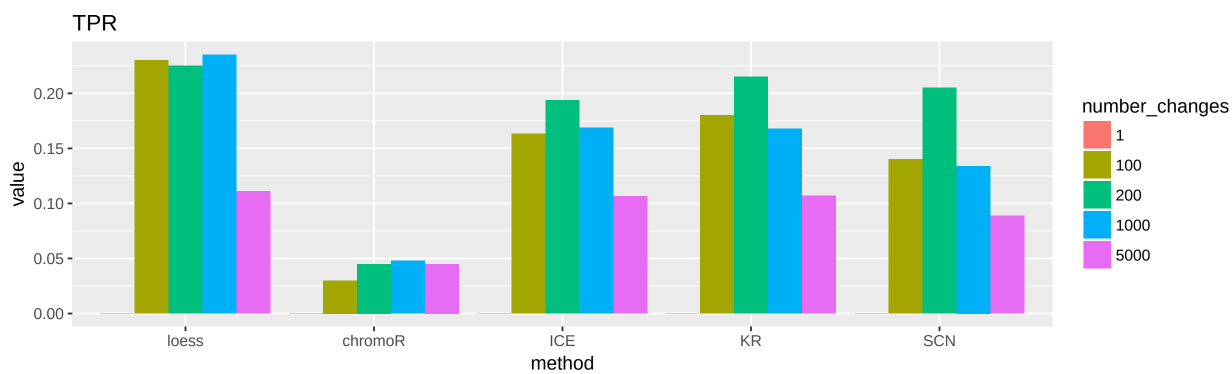
	loess	chromoR	ice	kr	scn
true positive	555	225	519	534	445
false positive	630	1240	671	687	762
true negative	20900	25100	20400	20800	20700
false negative	4440	4770	4360	4470	4560

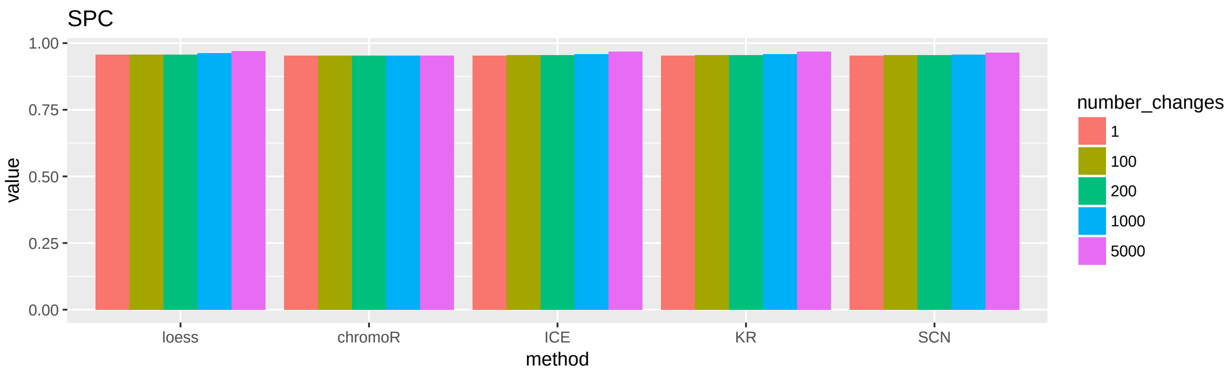
	loess	chromoR	ice	kr	scn
Total	26500	31400	26000	26500	26500
TPR	0.111	0.045	0.106	0.107	0.089
SPC	0.971	0.953	0.968	0.968	0.965
F1	0.892	0.893	0.89	0.89	0.886
AUC	0.679	0.518	0.603	0.601	0.575
AUC 20%	0.0544	0.0231	0.0448	0.0442	0.0372
FDR	0.532	0.846	0.564	0.563	0.631
Accuracy	0.809	0.808	0.806	0.806	0.799
Precision	0.468	0.154	0.436	0.437	0.369
FPR	0.0293	0.047	0.0318	0.0319	0.0354
FNR	0.889	0.955	0.894	0.893	0.911
FOR	0.176	0.16	0.176	0.177	0.18
NPV	0.824	0.84	0.824	0.823	0.82

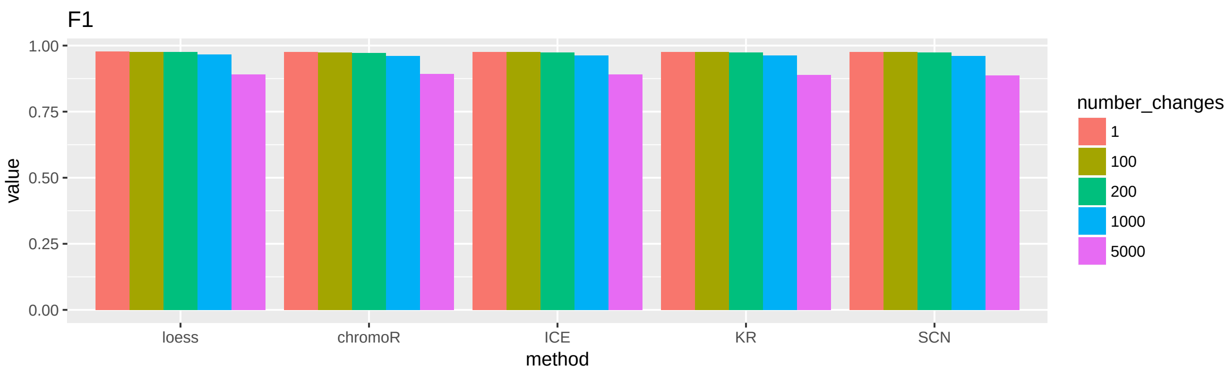
Bar plots

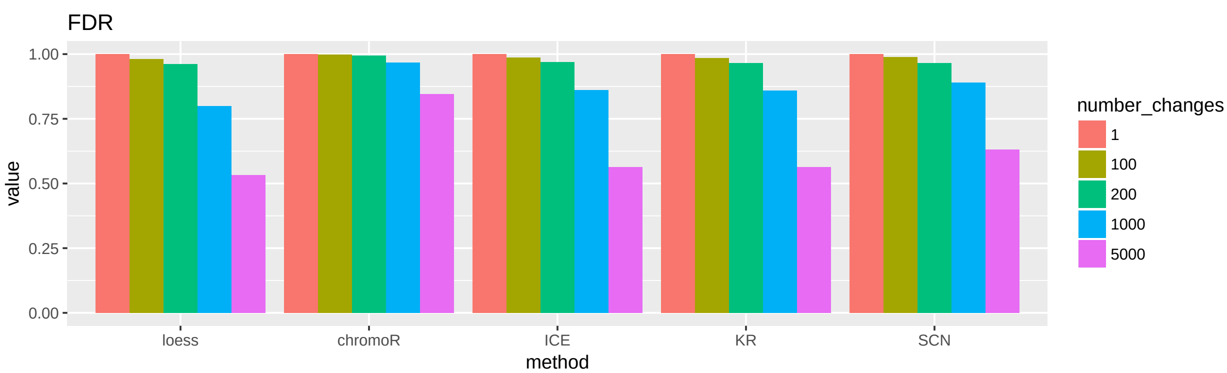
Below are bar plots showing comparisons of the different normalization methods over the varying numbers of changes at a fixed fold change for selected metrics.

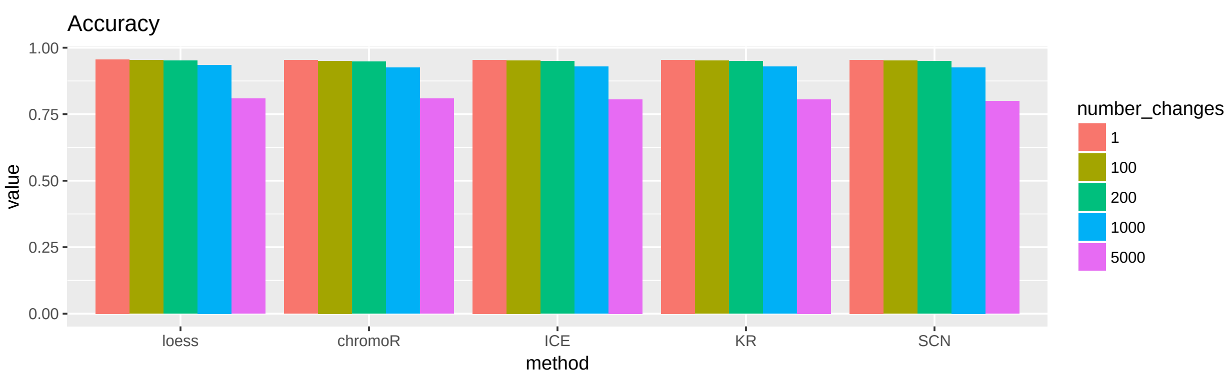


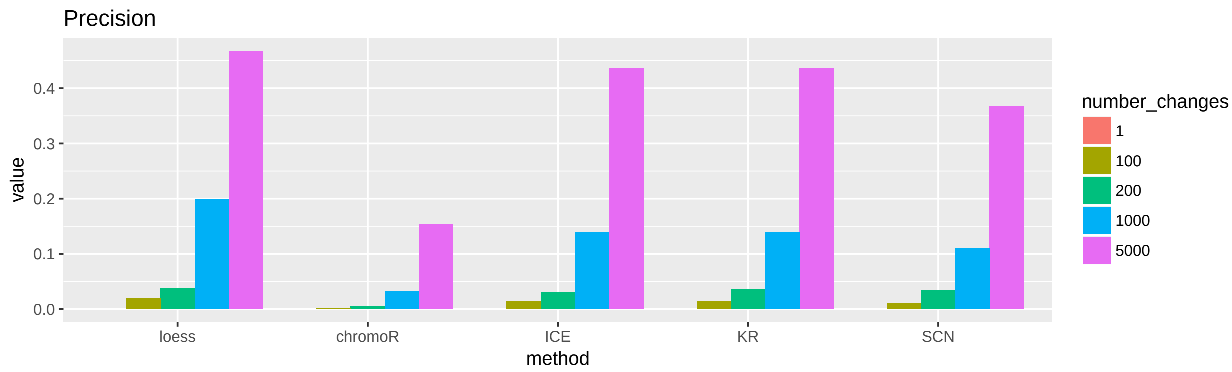



SPC
1.00
0.75
0.50
0.25
0.00
value
loess
chromoR
ICE
KR
SCN
method
number_changes
1
100
200
1000
5000


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


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


Accuracy
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

When only a single true difference was introduced at a 1.5 fold change all methods failed to detect it. At all other numbers of introduced differences checked, `loess` was able to detect the most while also maintaining the lowest number of false positives. `KR` and `SCN` again performed better than `ICE` and `ChromoR` at detecting the true differences.