Supplementary Material to "Functional Genomics via Multiscale Analysis: Application to Gene Expression and ChIP-on-chip Data"

GILAD LERMAN,* JOSEPH MCQUOWN, ALEXANDRE BLAIS, BRIAN D. DYNLACHT, GUANGLIANG CHEN, BUD MISHRA

November 7, 2006

---

*To whom correspondence should be addressed. E-mail: lerman@umn.edu
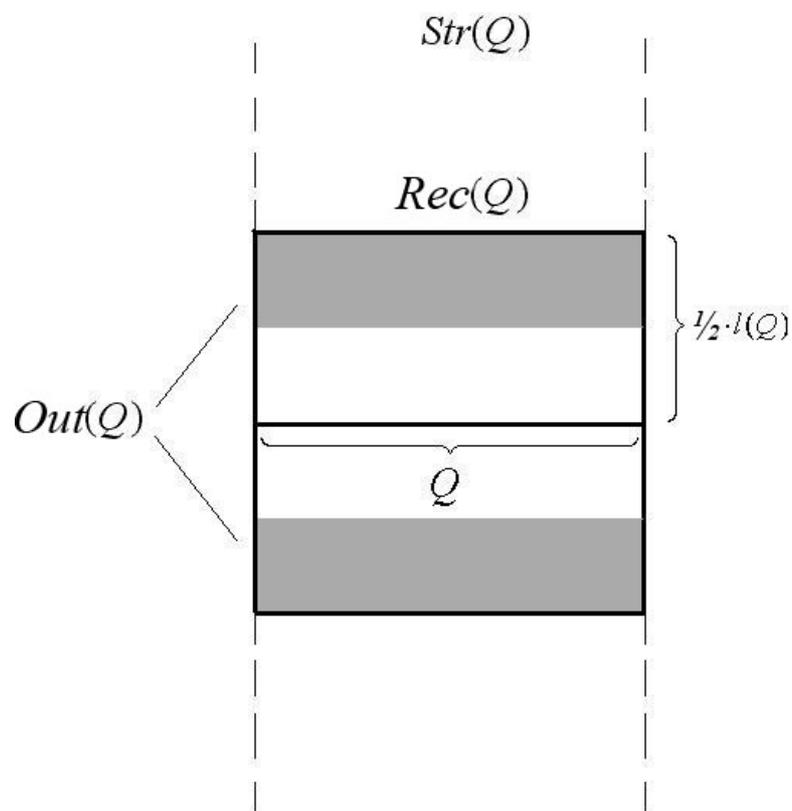
Figure S1: Demonstration of the different regions associated with the interval Q.
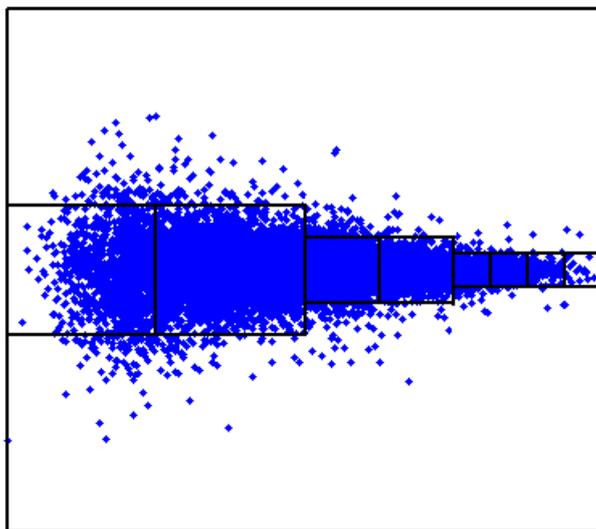
Figure S2: Rectangles of the form $Rec(Q)$ associated with stopping intervals Q for a synthetic data along a line ($\alpha_0 = 0.075$, $n_0 = 10$).
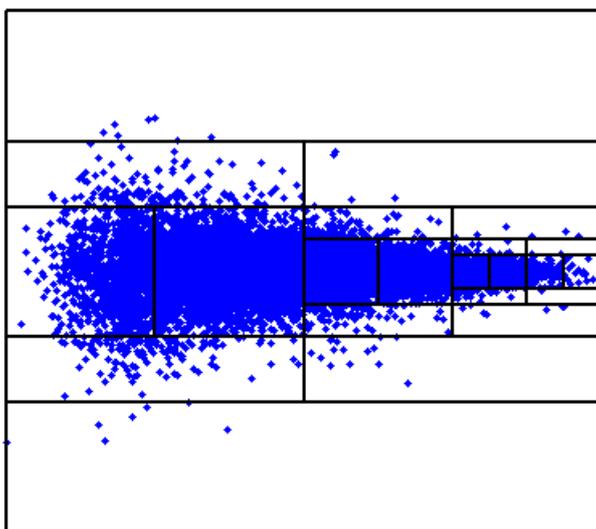
Figure S3: Rectangles associated with all dyadic intervals containing (and including) stopping intervals (same data and parameters as above).
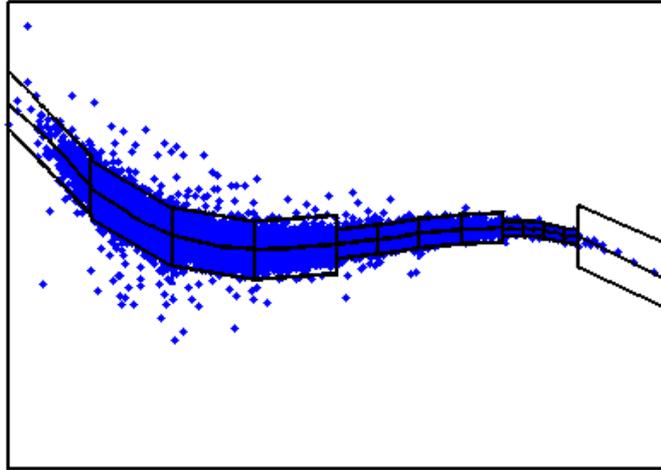
Figure S4: Sheared regions ($Rec(Q)$) associated with stopping intervals for a synthetic data along a curve ($\alpha_0 = 0.075$, $n_0 = 10$).
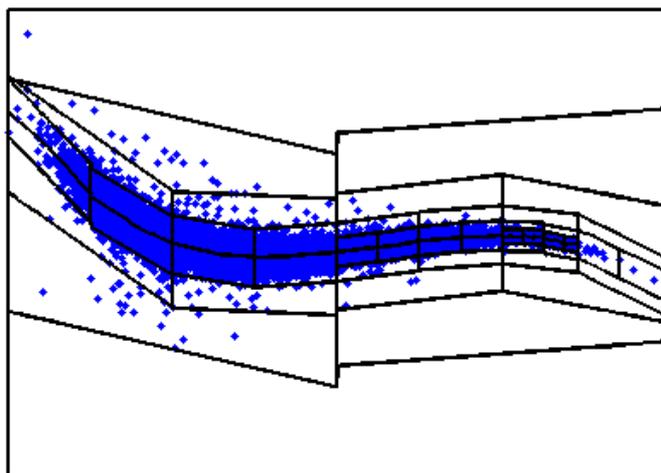
Figure S5: Sheared regions associated with all dyadic intervals containing (and including) stopping intervals (same data and parameters as above).
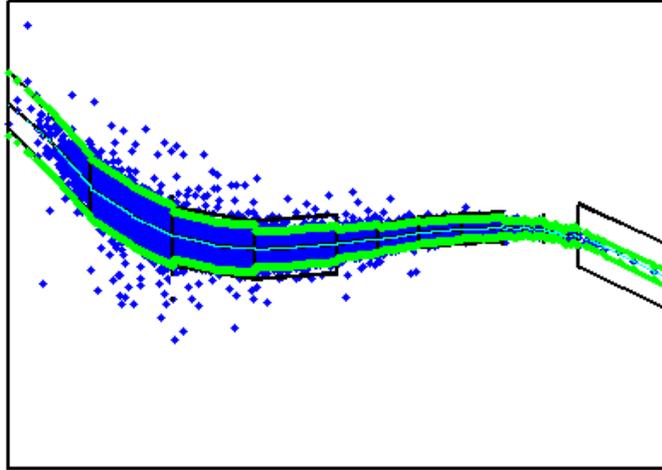
Figure S6: The strip $\widetilde{S}$ in green, normalizing curve C in light blue and stopping regions in black (same data and parameters as above).
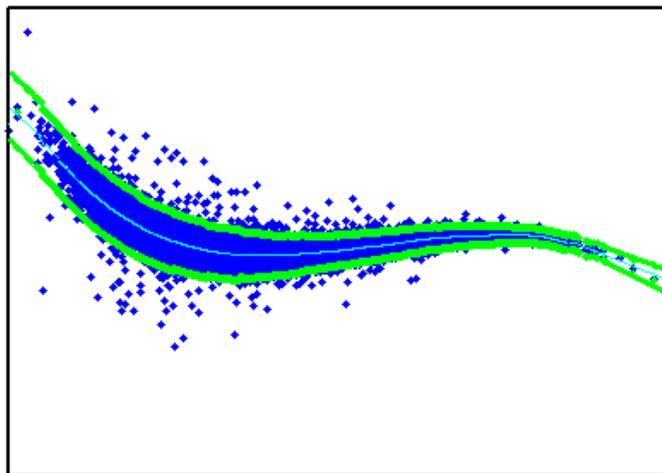
Figure S7: The Strip $\widetilde{S}$ (green) and normalizing curve C (light blue) averaged according to three uniformly shifted grids (same data and parameters as above).
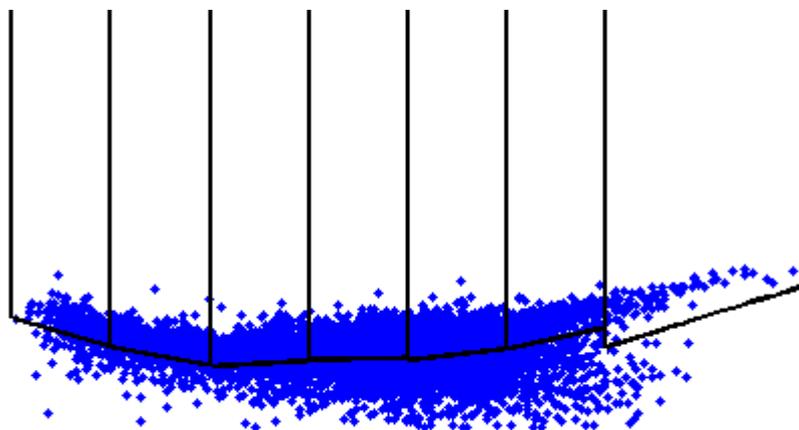
Figure S8: Non-symmetric regions ($Rec(Q)$) associated with stopping intervals for a ChIP-on-chip data ($\alpha_0 = 0.4$, $n_0 = 20$).
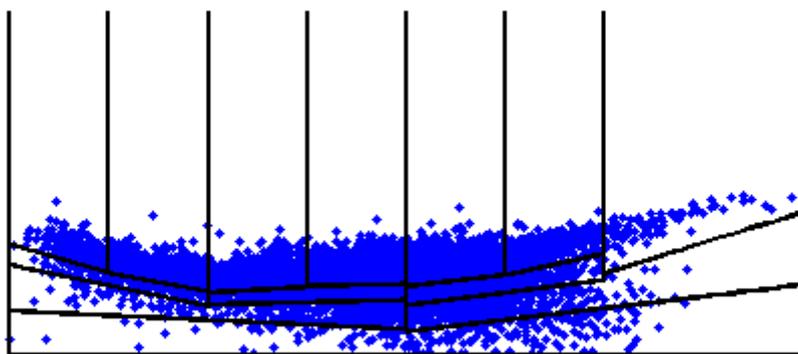
Figure S9: Nonsymmetric regions for ChIP-on-chip data associated with all dyadic intervals containing (and including) stopping intervals (same data and parameters as above).
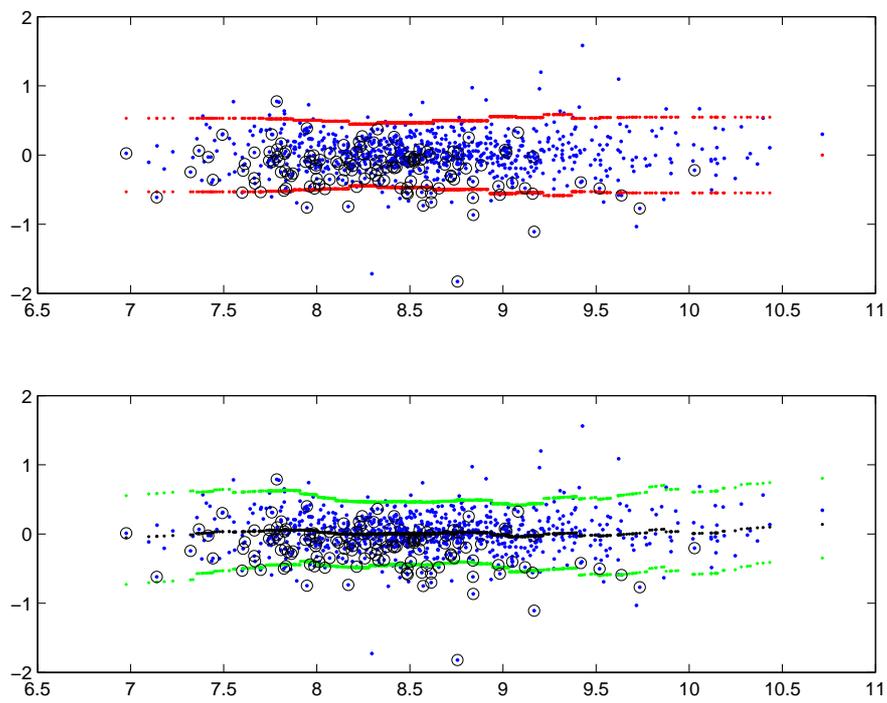
Figure S10: The separating strips for the data of slide 804. pSOL1 genes are circled. On top, SNN-Lerm strip (in red), where the data is normalized by SNN-Lerm (we used a preparsed pseudocode for SNN-Lerm and thus we could not present the results on original data). On bottom, MSC strip (in green) around normalizing curve (in black).
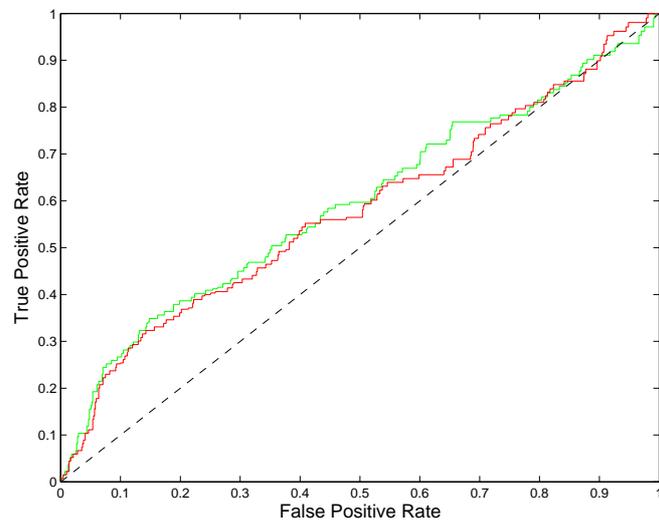
Figure S11: ROC curves for slide 804; MSC in green, SNN-Lerm in red. Recall that we define TPR as the fraction of pSOL1 genes identified as *down-regulated* and FPR as the fraction of non-pSOL1 genes identified as *differentially expressed*. Consequently, the curves are lower than the usual case where TPR is the fraction of pSOL1 genes identified as differentially expressed genes.
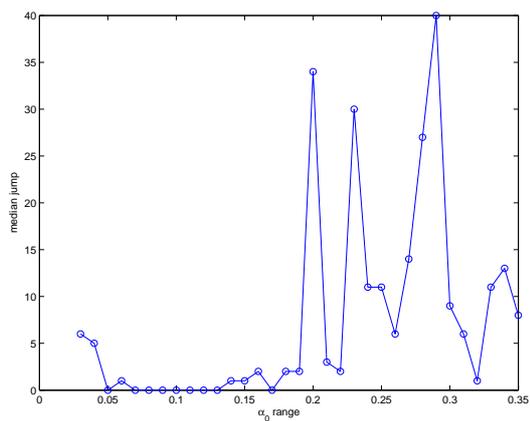
(a) Replicate A.
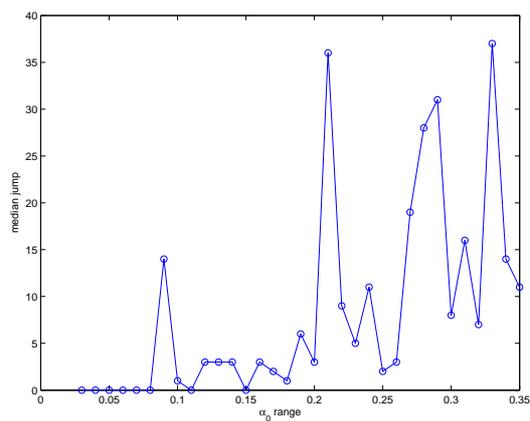
(b) Replicate B.
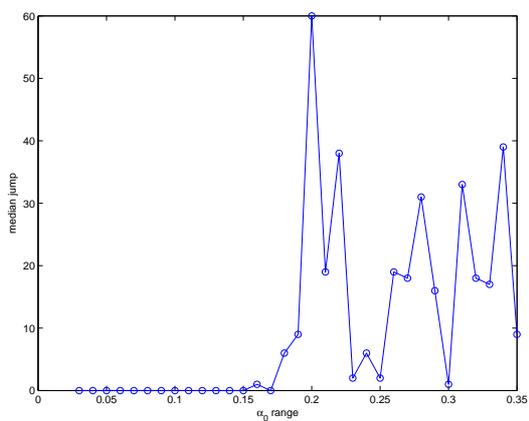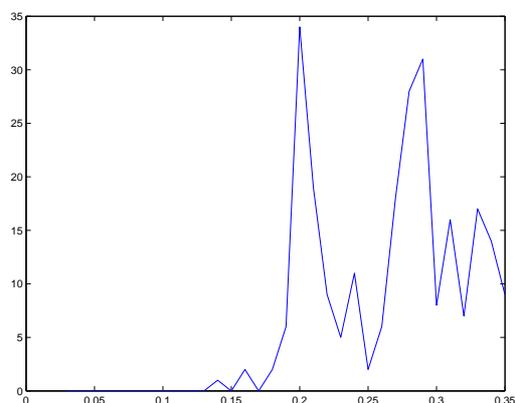
(c) Replicate C.

(d) Median of Jump of three replicates.

Figure S12: Jumps in detected number of outliers of our main ChIP-on-chip data (the transcription factor is myogenin) for regular MSC. More precisely, we fix FDR values between 0.05 and 0.1 and find the differences (jumps) between detected number of outliers by MSC at a given α_0 and $\alpha_0 + 0.01$. The MSC is applied here with initial transformation so that x-axis coincides with main principal axis. The first three figures demonstrate those jumps for the three replicates and the last one is the median value among all three. We apply the algorithm with the value of α_0 corresponding to the first noticeable jump. For replicates A and C we have thus chosen $\alpha_0 = .2$ and for replicate B we have assigned $\alpha_0 = .21$.

(a) Replicate A.

(b) Replicate B.

(c) Replicate C.

(d) Median of Jump of three replicates.
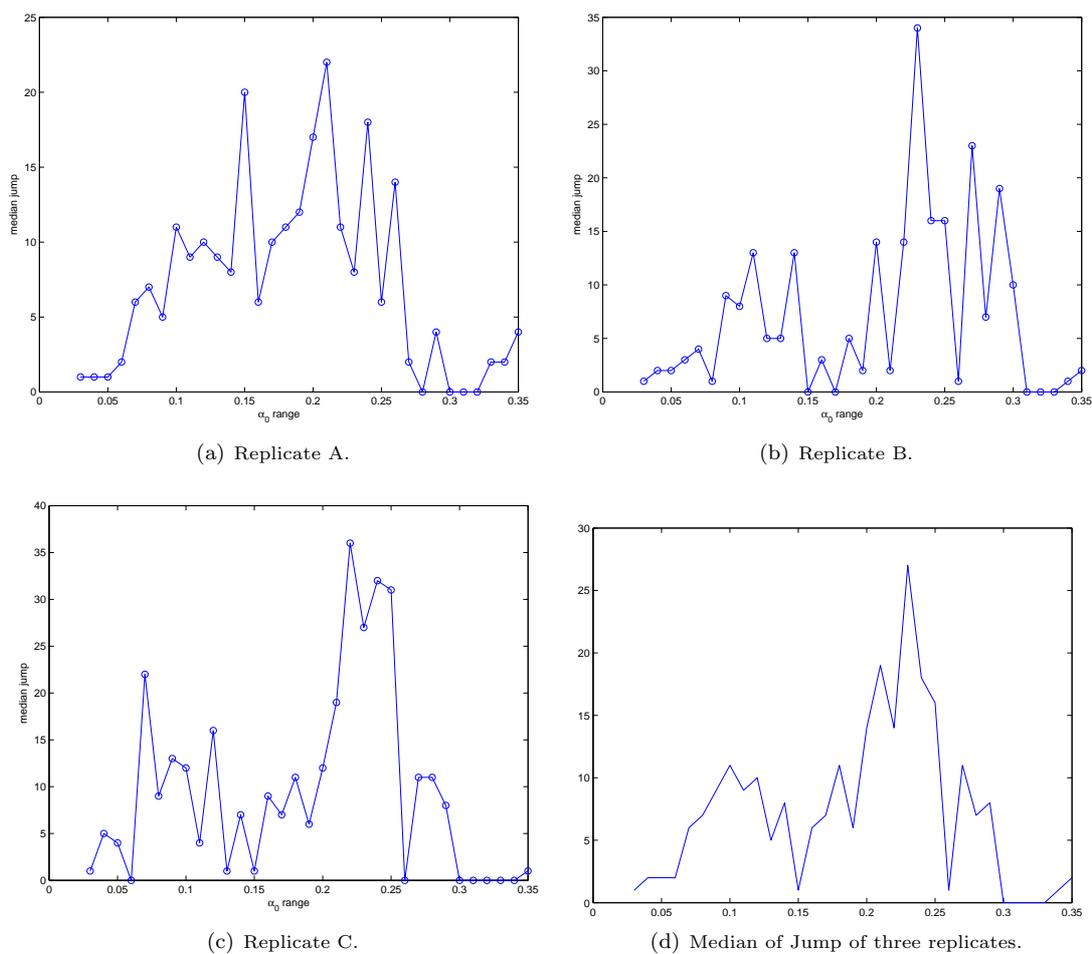
Figure S13: Jumps in detected number of outliers of our main ChIP-on-chip data (the transcription factor is myogenin) for MSC without initial transformation. More precisely, we fix FDR values between 0.05 and 0.1 and find the differences (jumps) between detected number of outliers by MSC at a given α_0 and $\alpha_0 + 0.01$. The MSC is applied here to given data in MA coordinates without initial transformation so that x-axis coincides with main principal axis. The first three figures demonstrate those jumps for the three replicates and the last one is the median value among all three. It shows that the first jump for all three replicates has to be searched around $\alpha_0 = 0.1$. By observing the three replicates we choose the specific values of α_0 (according to first jumps) as follows: $\alpha_0 = 0.1$ for replicate A, $\alpha_0 = 0.11$ for replicate B, $\alpha_0 = 0.07$ for replicate C.
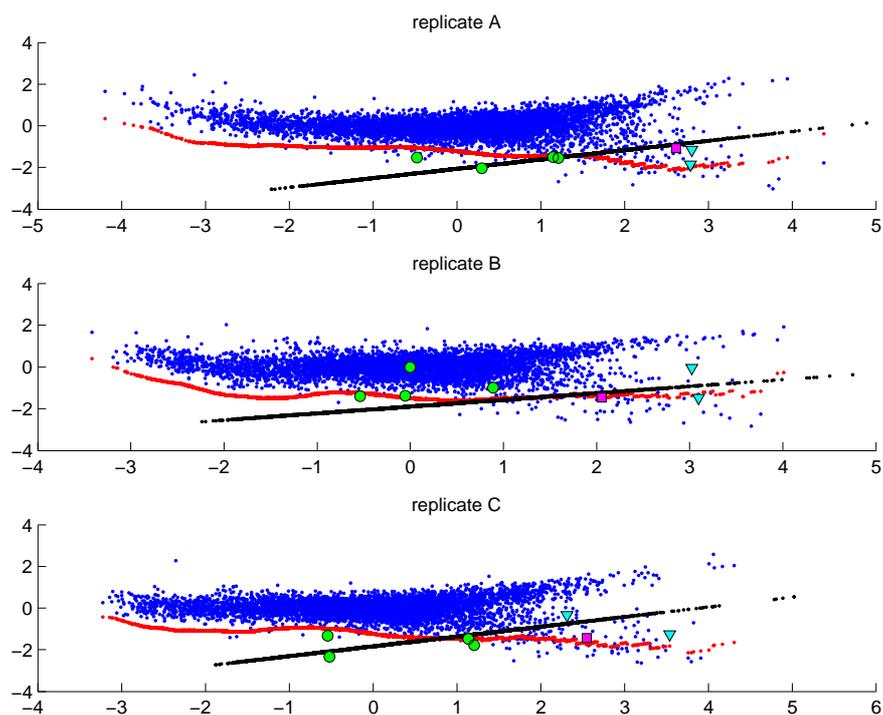
Figure S14: Comparison of BR with MSC on the principal axis for the three replicates of our data. That is, MSC is applied with initial shift and rotation of the data on its principal axes. The axes in the figure are the principal axes; BR line has been computed line with respect to MA coordinates. We have identified the enriched points of MSC in at least two replicates while applying FDR level of 0.1 for each replicate; We identified the same number of enriched points with a weighted BR score. Circles reveal enriched spots that the MSC algorithm distinguished and the binding ratio method failed to distinguish over all 3 replicates, while squares reveal enriched points identified by BR and not by MSC. Triangles reveal points which are not enriched and were identified by BR as enriched, unlike MSC. There were no spots that MSC failed to identify as not enriched, while BR did not. The threshold given by the MSC is indicated by a straight line, while that of MSC is the thicker red curve.

Figure S15: MA coordinates of Intensities for three replicates of Myogenin ChIP-on-chip data. We have circled the point corresponding to the promoter Cacng1. It is not enriched but detected by all algorithm as enriched with low false positive rates. MSC without initial transformation detects this point earlier than regular MSC and thus results in lower area under the ROC curve. If α_0 is detected according to our method the differences are insignificant. The differences increase with higher values of α_0.
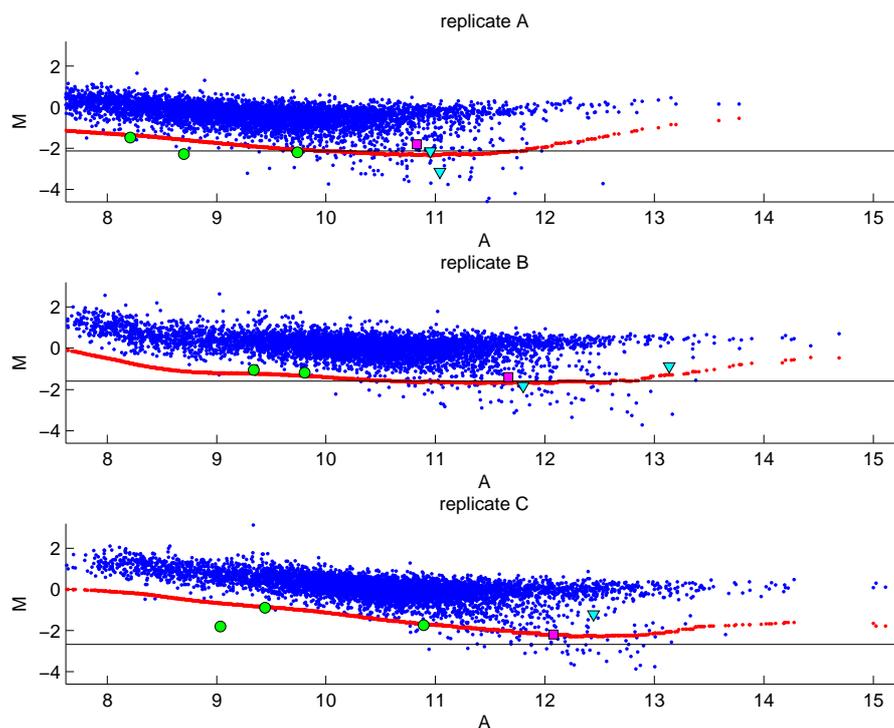
Figure S16: Comparison of BR with MSC, applied **without initial shift and rotation** onto principal axis, for the three replicates of our data, while **Cacng1 is excluded**. We have identified the enriched points of MSC in at least two replicates while applying FDR level of 0.1 for each replicate; We identified the same number of enriched points with a weighted BR score. Circles reveal enriched spots that the MSC algorithm distinguished and the binding ratio method failed to distinguish over all 3 replicates, while squares reveal enriched points identified by BR and not by MSC. Triangles reveal points which are not enriched and were identified by BR as enriched, unlike MSC. There were no spots that MSC failed to identify as not enriched, while BR did not. The threshold given by the MSC is indicated by a straight line, while that of MSC is the thicker red curve.

Figure S17: Comparison of different normalization curves for three replicates of Myogenin ChIP-on-chip data. In black, regular MSC (initial transformation onto principal axis) with α_0 chosen by the jump curves; in light blue MSC without initial transformation with α_0 chosen by the jump curves; in red LOESS.
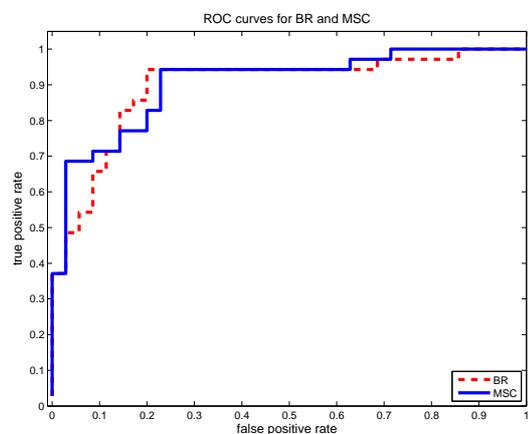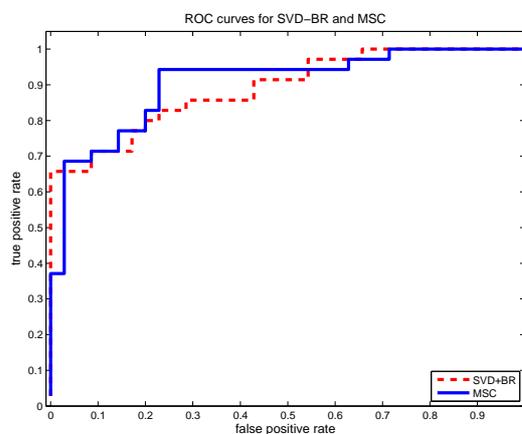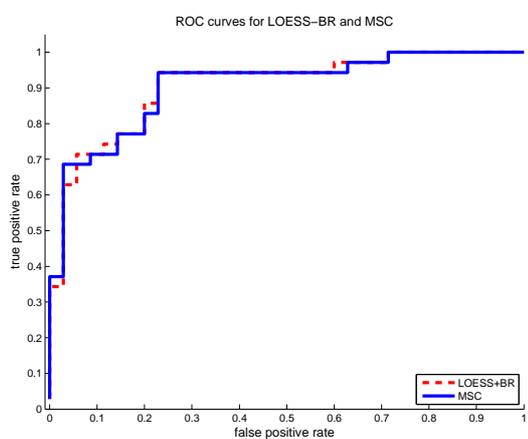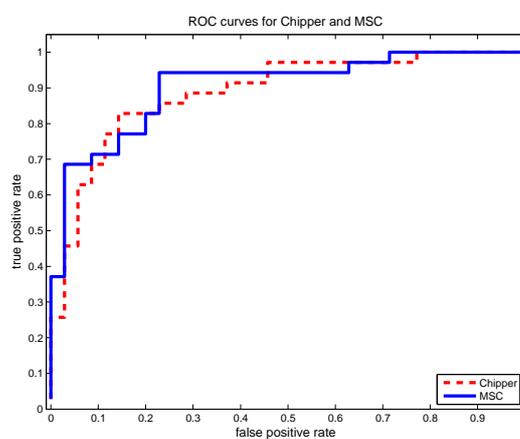
Figure S18: ROC Curves comparing each one of the methods: BR, BR with LOESS, BR with PCA and Chipper with MSC without initial rotation. The MSC curve (without initial rotation) is described by a solid line and the other curves by dashed lines.
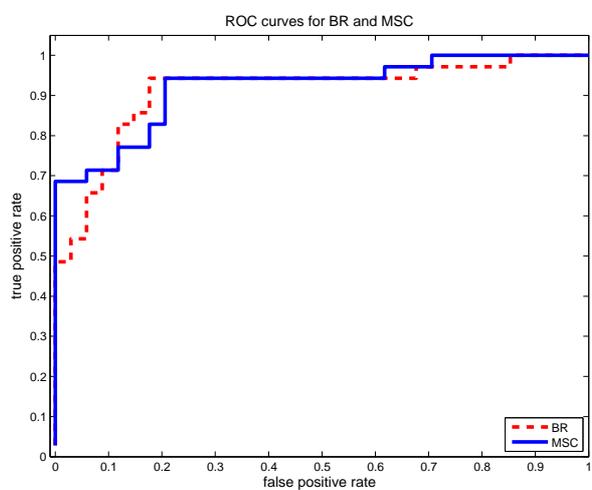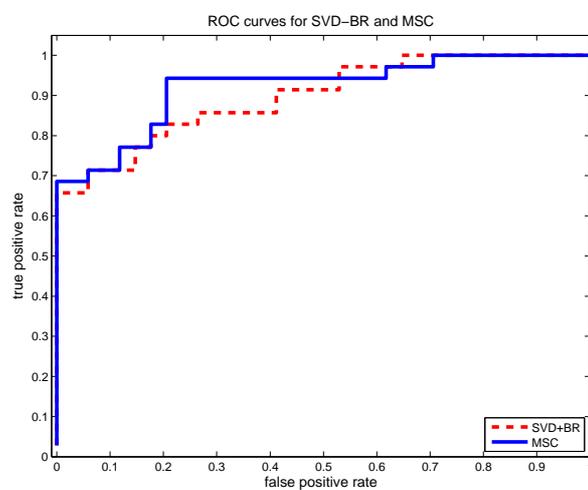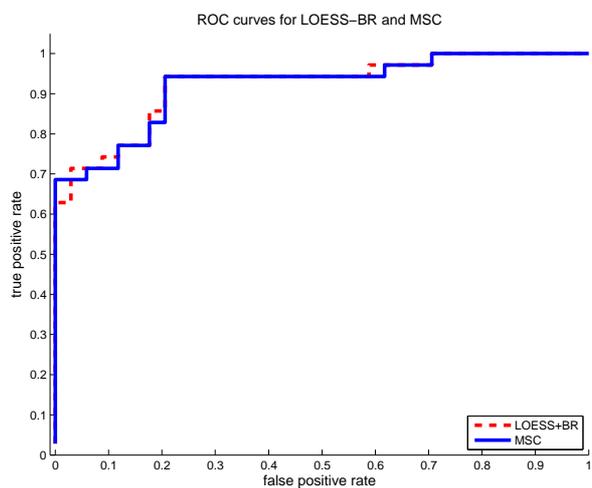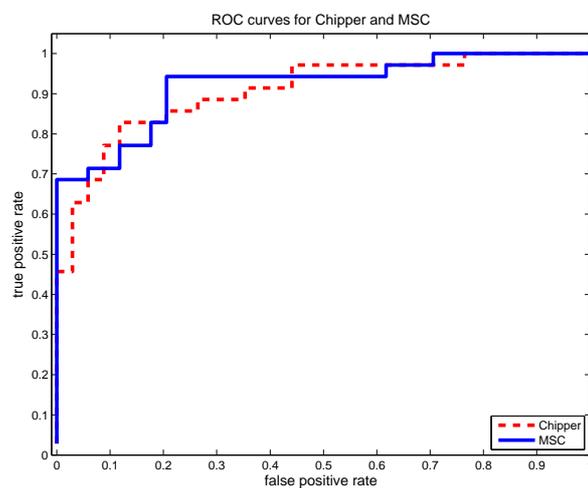
Figure S19: ROC Curves comparing each one of the methods: BR, BR with LOESS, BR with PCA and Chipper with MSC without initial rotation, while excluding the promoter Cacng1. The MSC curve (without initial rotation) is described by a solid line and the other curves by dashed lines.
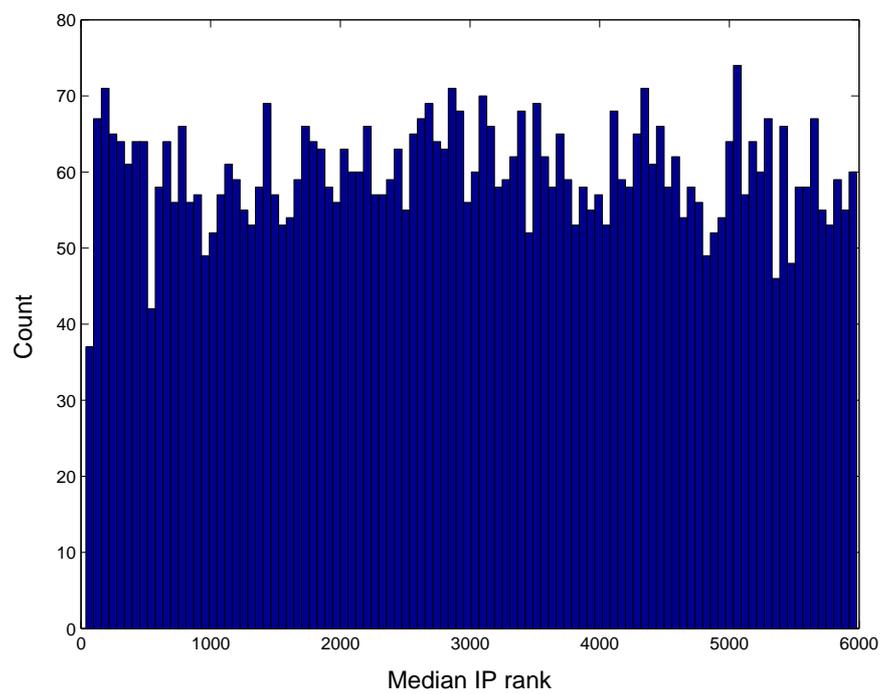
Figure S20: Median Rank histogram for 3 Replicates of previous ChIP-on-chip data.
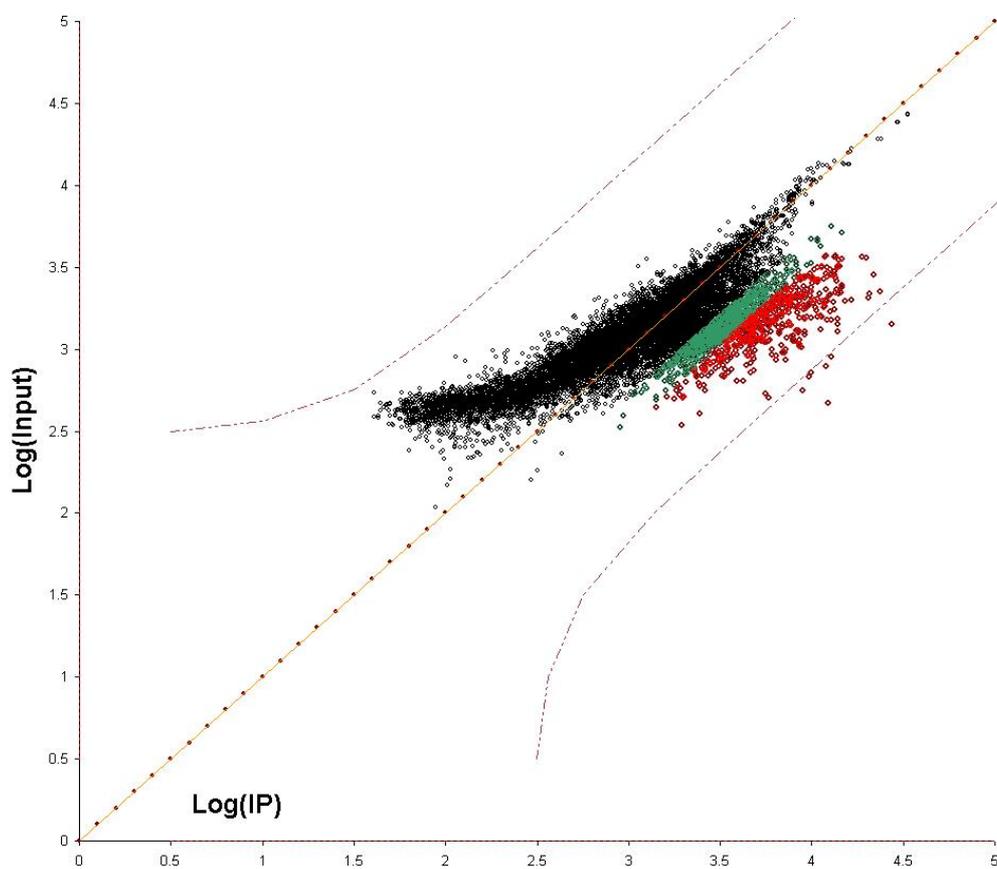
Figure S21: Rosetta model applied to previous data. Data is presented on log intensities' coordinates and $y = x$ is the normalizing line. The wrong shape of the strip is due to both wrong normalization but also wrong parametric assumptions of the Rosetta model. Spots with binding ratio greater than 3 are in green and greater than 2 are in red.

(a) Replicate A.

(b) Replicate B.

(c) Replicate C.

(d) Median of Jump of three replicates.
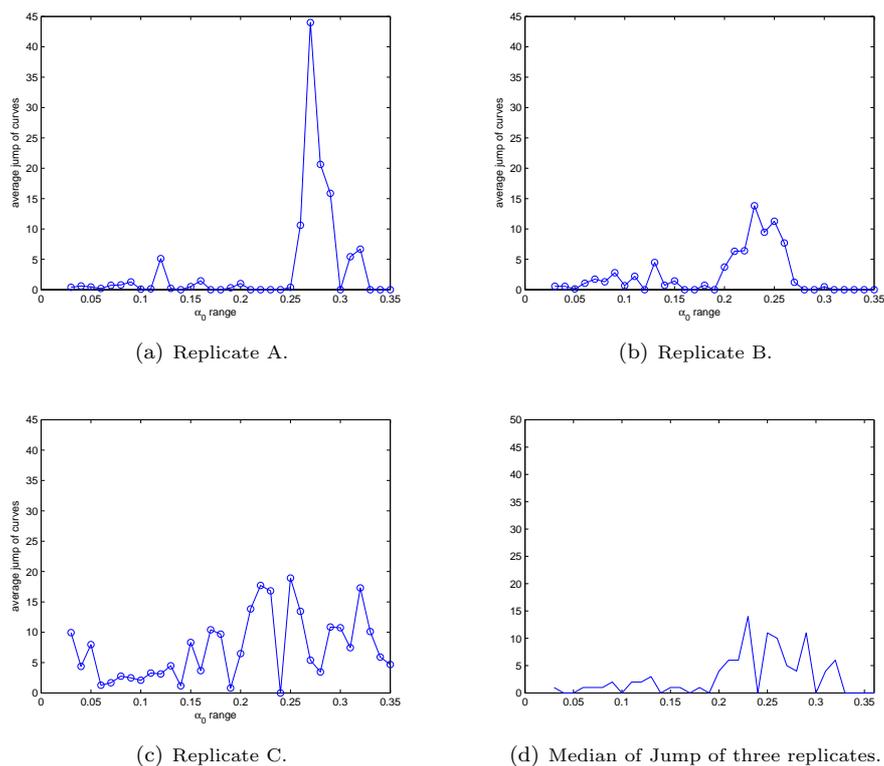
Figure S22: Jumps in detected number of outliers of ChIP-on-chip data with antibodies recognizing MyoD in myotubes, while applying regular MSC. More precisely, we fix FDR values between 0.05 and 0.1 and find the differences (jumps) between detected number of outliers by MSC at a given α_0 and $\alpha_0 + 0.01$. The MSC is applied here with initial transformation so that x-axis coincides with main principal axis. The first three figures demonstrate those jumps for the three replicates and the last one is the median value among all three. We apply the algorithm with the value of α_0 corresponding to the first noticeable jump. We have chosen $\alpha_0 = 0.27$ for replicate A, $\alpha_0 = 0.23$ for replicate B and $\alpha_0 = 0.22$ for replicate C.

(a) Replicate A.

(b) Replicate B.

(c) Replicate C.

(d) Median of Jump of three replicates.
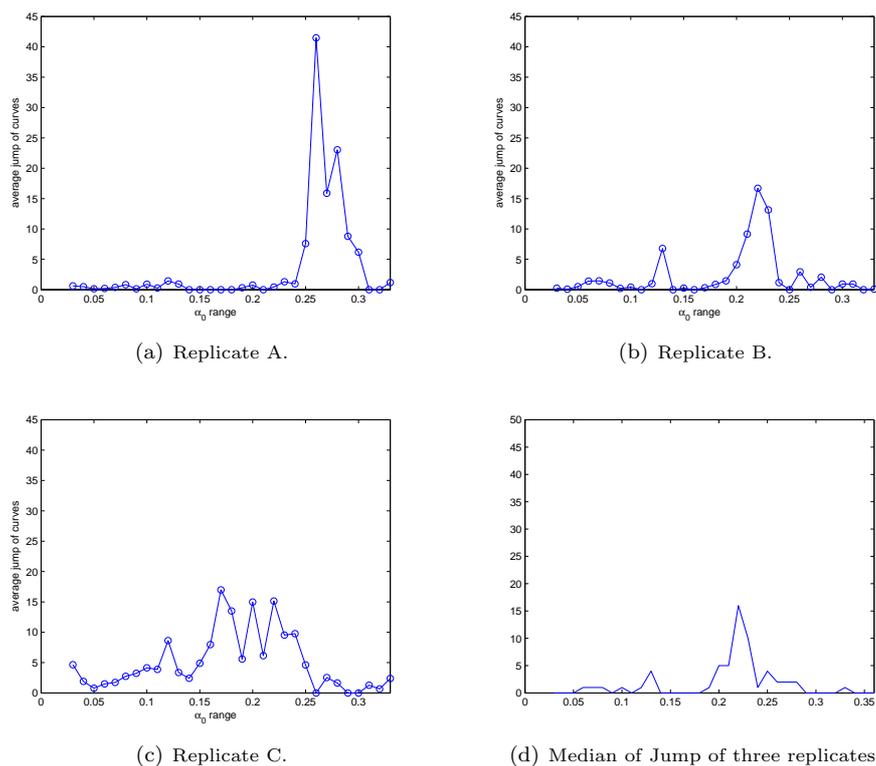
Figure S23: Jumps in detected number of outliers of ChIP-on-chip data with antibodies recognizing MyoD in myotubes, while applying MSC without initial transform. More precisely, we fix FDR values between 0.05 and 0.1 and find the differences (jumps) between detected number of outliers by MSC (without initial shift and rotation onto main principal axis) at a given α_0 and $\alpha_0 + 0.01$. The first three figures demonstrate those jumps for the three replicates and the last one is the median value among all three. We apply the algorithm with the value of α_0 corresponding to the first noticeable jump. We have chosen $\alpha_0 = 0.26$ for replicate A, $\alpha_0 = 0.22$ for replicate B and $\alpha_0 = 0.17$ for replicate C.

(a) Replicate A.

(b) Replicate B.

(c) Replicate C.

(d) Median of Jump of three replicates.
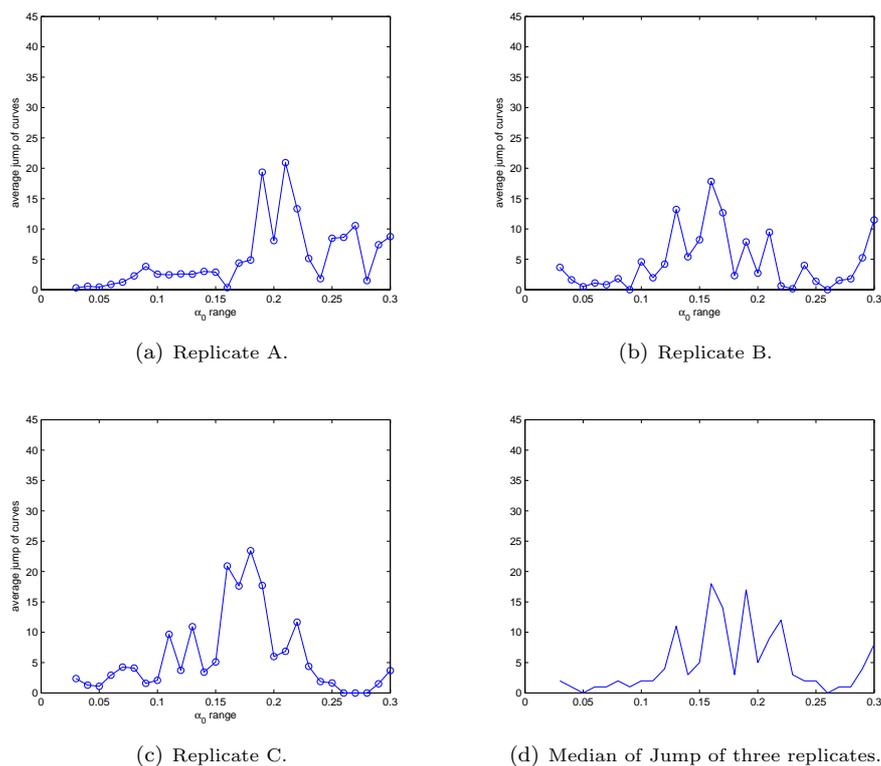
Figure S24: Jumps in detected number of outliers of ChIP-on-chip data with antibodies recognizing MyoD in growing cells, while applying regular MSC. More precisely, we fix FDR values between 0.05 and 0.1 and find the differences (jumps) between detected number of outliers by MSC at a given α_0 and $\alpha_0 + 0.01$. The MSC is applied here with initial transformation so that x-axis coincides with main principal axis. The first three figures demonstrate those jumps for the three replicates and the last one is the median value among all three. We apply the algorithm with the value of α_0 corresponding to the first noticeable jump. We have chosen $\alpha_0 = 0.19$ for replicate A, $\alpha_0 = 0.16$ for replicate B and $\alpha_0 = 0.16$ for replicate C.

(a) Replicate A.

(b) Replicate B.

(c) Replicate C.

(d) Median of Jump of three replicates.
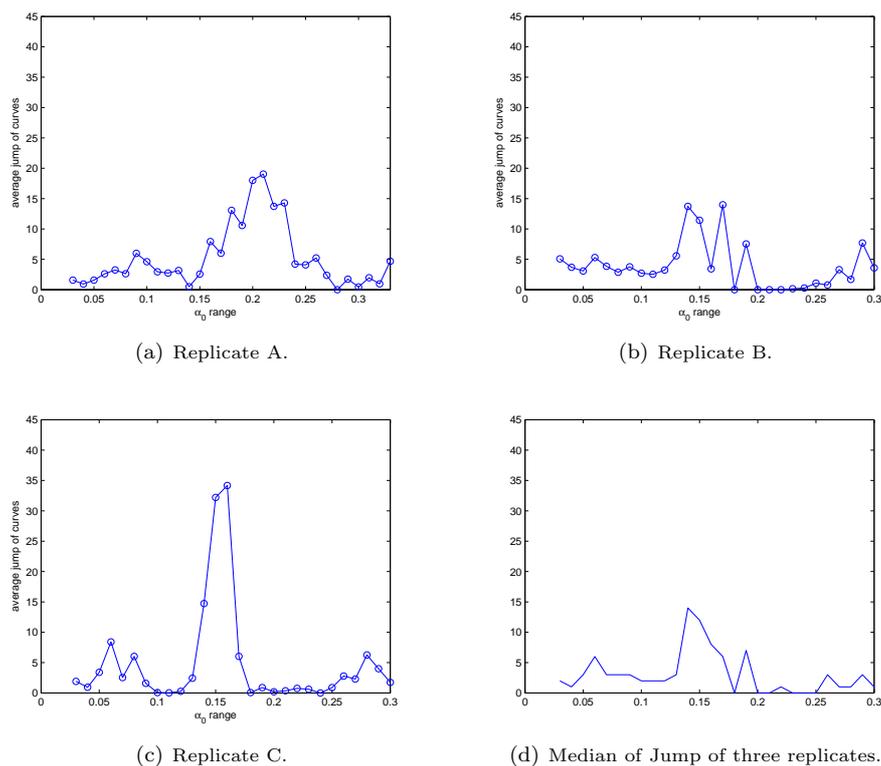
Figure S25: Jumps in detected number of outliers of ChIP-on-chip data with antibodies recognizing MyoD in growing cells, while applying MSC without initial transform. More precisely, we fix FDR values between 0.05 and 0.1 and find the differences (jumps) between detected number of outliers by MSC (without initial shift and rotation onto main principal axis) at a given α_0 and $\alpha_0 + 0.01$. The first three figures demonstrate those jumps for the three replicates and the last one is the median value among all three. We apply the algorithm with the value of α_0 corresponding to the first noticeable jump. We have chosen $\alpha_0 = 0.21$ for replicate A, $\alpha_0 = 0.14$ for replicate B and $\alpha_0 = 0.16$ for replicate C.

23
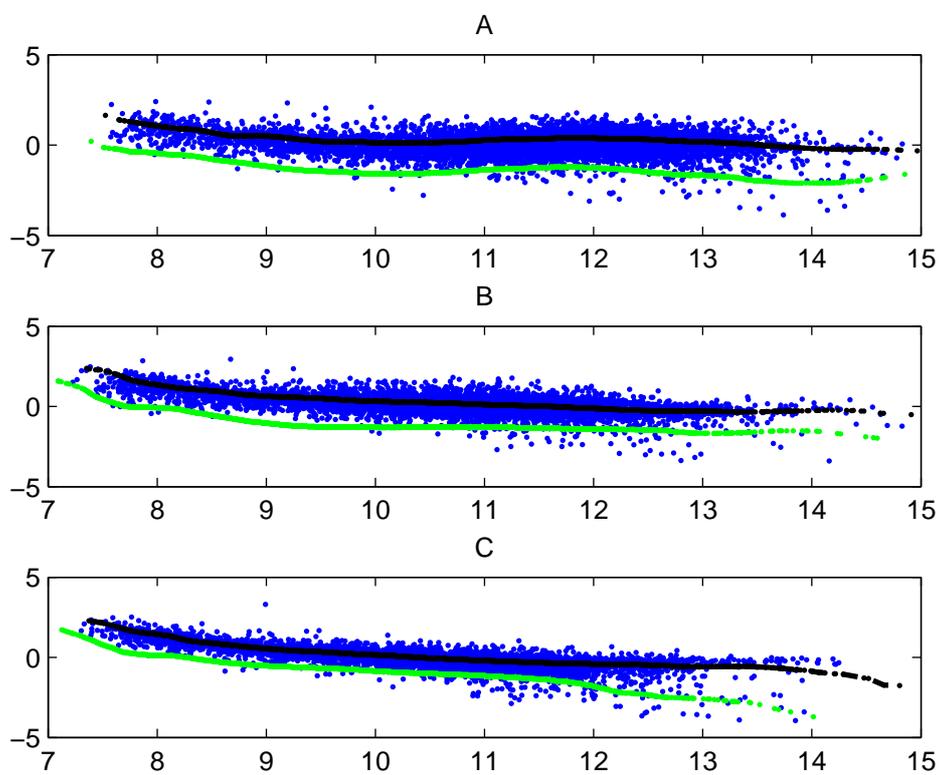
Figure S26: Curve and Strip of regular MSC (initial transformation onto principal axis) for ChIP-on-chip data with antibodies recognizing MyoD in myotubes.
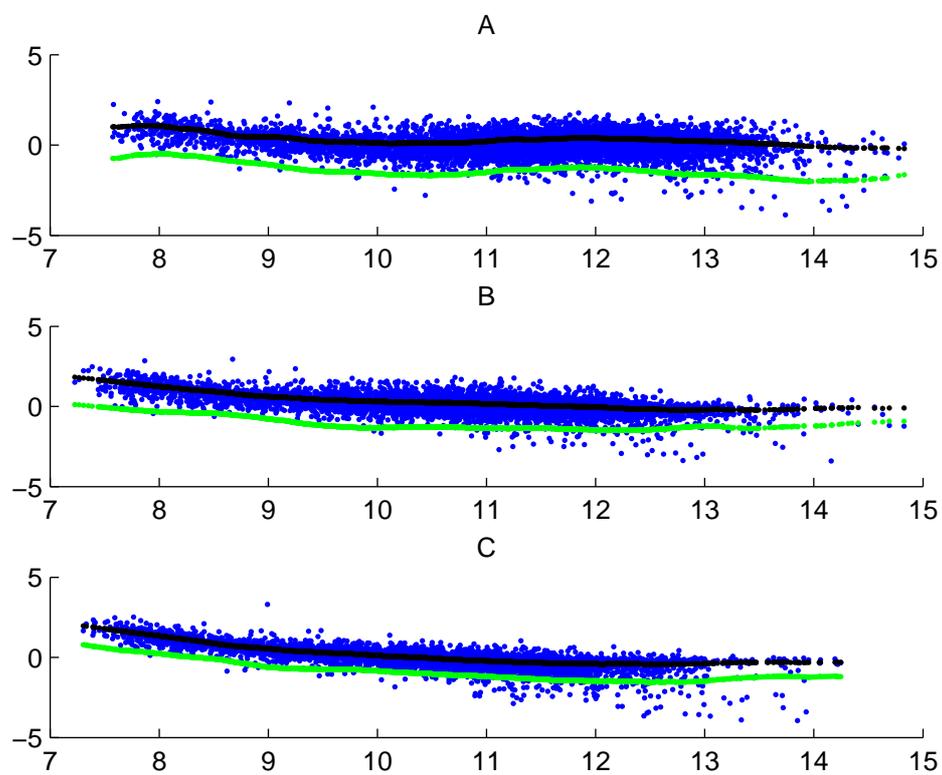
Figure S27: Curve and Strip of MSC without initial transformation (onto principal axis) for ChIP-on-chip data with antibodies recognizing MyoD in myotubes.
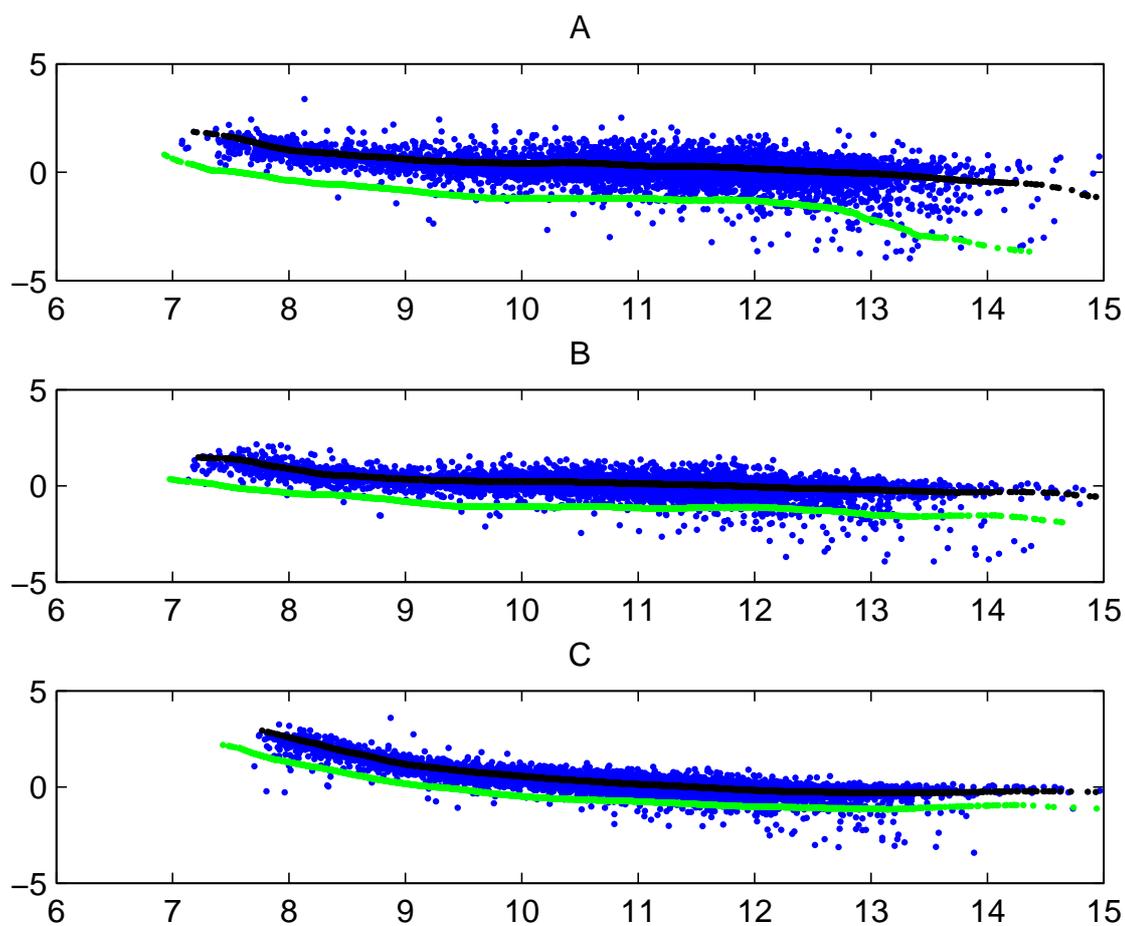
Figure S28: Curve and Strip of regular MSC (initial transformation onto principal axis) for ChIP-on-chip data with antibodies recognizing MyoD in growing cells.
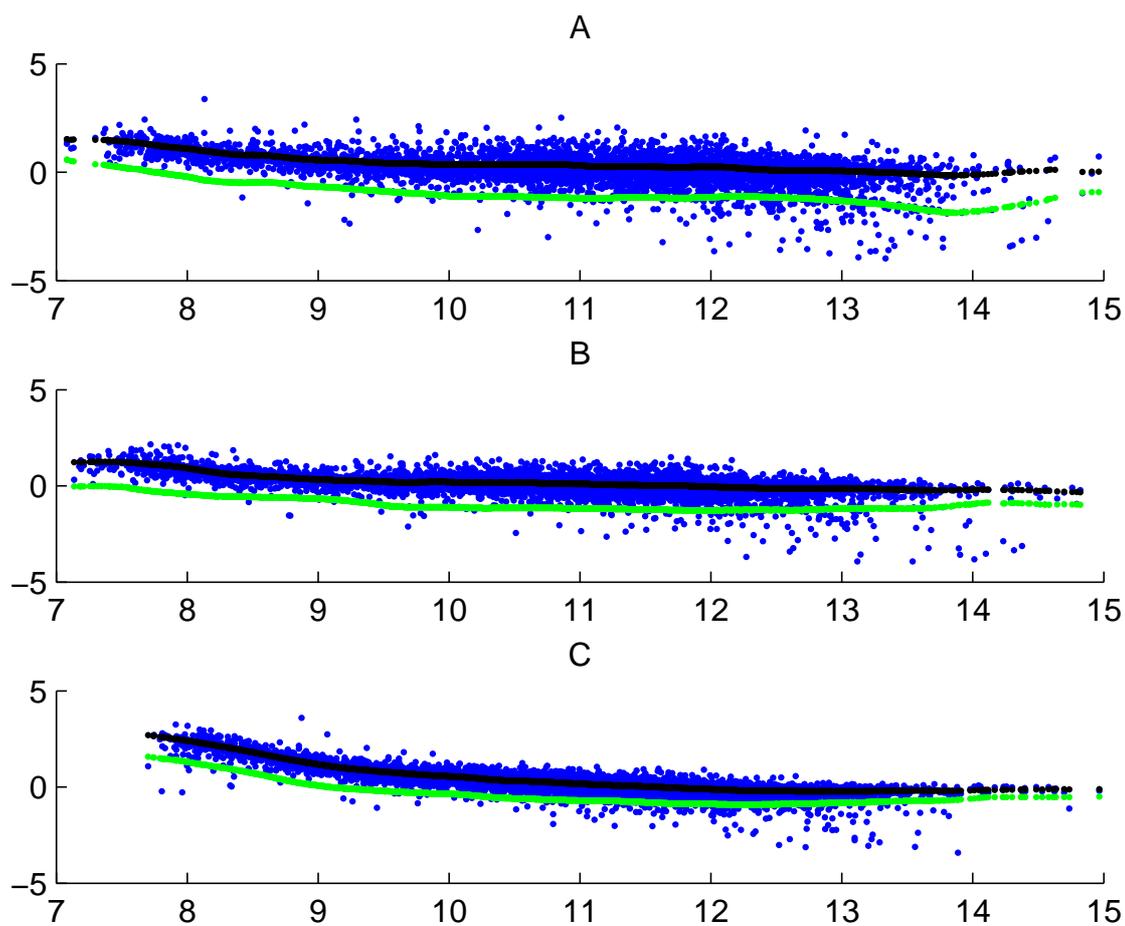
Figure S29: Curve and Strip of MSC without initial transformation (onto principal axis) for ChIP-on-chip data with antibodies recognizing MyoD in growing cells.
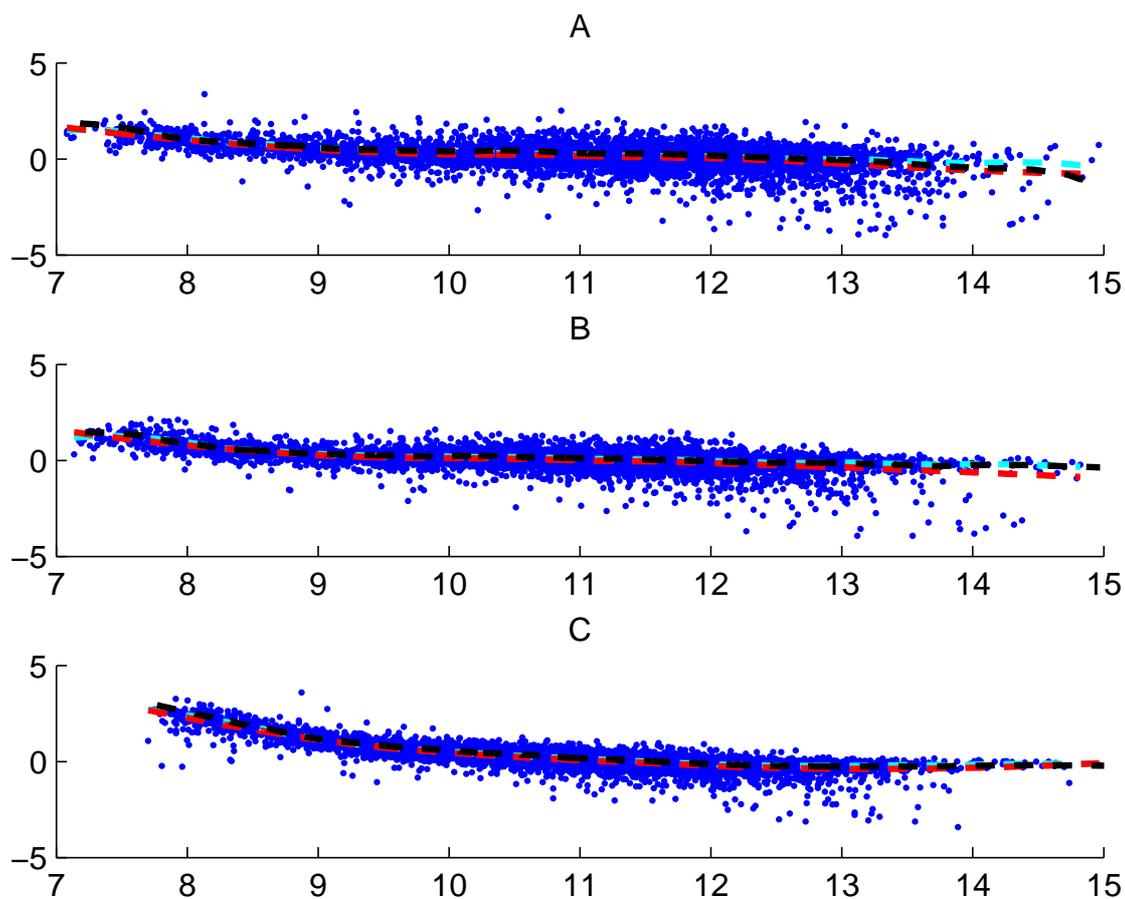
Figure S30: Normalizing curves of LOESS (red curve, span = 0.4), regular MSC (black curve, with initial transformation onto principal axis) and MSC without initial transformation (light-blue curve) for ChIP-on-chip data with antibodies recognizing MyoD in growing cells.
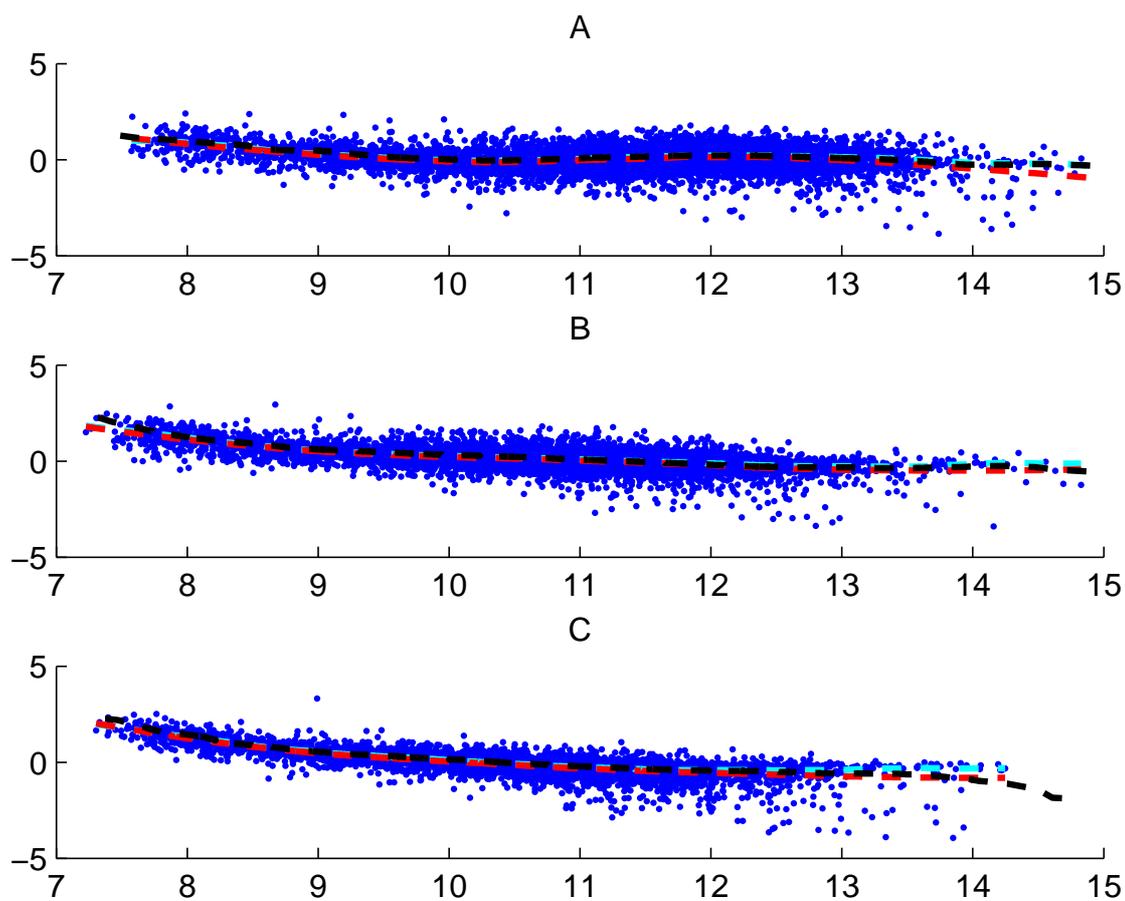
Figure S31: Normalizing curves of LOESS (red curve, span = 0.4), regular MSC(black curve, with initial transformation onto principal axis) and MSC without initial transformation (light-blue curve) for ChIP-on-chip data with antibodies recognizing MyoD in myotubes.

α_0	0.05	0.1	0.15	0.2	0.21	0.22	0.23	0.24	0.25	0.3	0.35	0.4	0.45	0.5
Area	0.8988	0.8996	0.8963	0.9086	0.9135	0.9143	0.9151	0.9118	0.9127	0.9086	0.9086	0.9069	0.9029	0.9020

Table S1: Different Area under ROC for different α_0s, when applying regular MSC (with initial transformation)

α_0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Area	0.9086	0.9053	0.898	0.8939	0.8947	0.8857	0.8833	0.8816	0.8873	0.8906

Table S2: Different Area under ROC for different α_0s, when MSC is applied without an initial shift and rotation onto principal axis

α_0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Area	0.9218	0.9235	0.9193	0.9193	0.9202	0.9109	0.9067	0.9034	0.9084	0.9118

Table S3: Different Area under ROC for different α_0s, when MSC is applied without an initial shift and rotation onto principal axis and when excluding the promoter Cacng1 from test set

Identification for regular MSC (with initial transformation):

α_0	0.05	0.1	0.15	0.2	α_0^*	0.21	0.22	0.23	0.24	0.25	0.3	0.35	0.4	0.45	0.5
MS TP	0	1	2	18	22	23	23	24	24	24	28	31	33	33	34
MS FP	0	0	0	0	0	1	1	1	1	1	7	7	10	12	19
MS FN	35	34	33	17	13	12	12	11	11	11	7	4	2	2	1
MS TN	35	35	35	35	35	34	34	34	34	34	28	28	25	23	16
MS TP rate	0	0.0286	0.0571	0.5143	0.6286	0.6571	0.6571	0.6857	0.6857	0.6857	0.8	0.8857	0.9429	0.9429	0.9714
MS FP rate	0	0	0	0	0	0.0286	0.0286	0.0286	0.0286	0.0286	0.2	0.2	0.2857	0.3429	0.5429

Identification for BR (same % as MSC above):

α_0	0.05	0.1	0.15	0.2	α_0^*	0.21	0.22	0.23	0.24	0.25	0.3	0.35	0.4	0.45	0.5
BR TP	0	2	3	16	19	21	22	23	23	25	32	33	33	33	33
BR FP	0	0	0	1	2	3	3	3	4	5	7	9	11	17	23
BR FN	35	33	32	19	16	14	13	12	12	10	3	2	2	2	2
BR TN	35	35	35	34	33	32	32	32	31	30	28	26	24	18	12
BR TP rate	0	0.0571	0.0857	0.4571	0.5429	0.6	0.6286	0.6571	0.6571	0.7143	0.9143	0.9429	0.9429	0.9429	0.9429
BR FP rate	0	0	0	0.0286	0.0571	0.0857	0.0857	0.0857	0.1143	0.1429	0.2	0.2571	0.3143	0.4857	0.6571

Identification for BR w.r.t. principal axis (same % as MSC above):

α_0	0.05	0.1	0.15	0.2	α_0^*	0.21	0.22	0.23	0.24	0.25	0.3	0.35	0.4	0.45	0.5
BR pca TP	0	3	5	18	22	23	23	23	23	25	27	28	29	30	32
BR pca FP	0	0	0	0	0	0	1	3	3	3	7	7	8	11	18
BR pca FN	35	32	30	17	13	12	12	12	12	10	8	7	6	5	3
BR pca TN	35	35	35	35	35	35	34	32	32	32	28	28	27	24	17
BR pca TP rate	0	0.0857	0.1429	0.5143	0.6286	0.6571	0.6571	0.6571	0.6571	0.7143	0.7714	0.8	0.8286	0.8571	0.9143
BR pca FP rate	0	0	0	0	0	0	0.0286	0.0857	0.0857	0.0857	0.2	0.2	0.2286	0.3143	0.5143

Identification for BR with initial LOESS normalization (same % as MSC above):

α_0	0.05	0.1	0.15	0.2	α_0^*	0.21	0.22	0.23	0.24	0.25	0.3	0.35	0.4	0.45	0.5
Loess BR TP	0	1	3	17	18	23	24	24	25	25	30	33	33	33	33
Loess BR FP	0	0	0	1	2	2	2	2	3	4	8	8	9	14	18
Loess BR FN	35	34	32	18	17	12	11	11	10	10	5	2	2	2	2
Loess BR TN	35	35	35	34	34	33	33	33	32	31	27	27	26	21	17
Loess BR TP rate	0	0.0286	0.0857	0.4857	0.0571	0.6571	0.6857	0.6857	0.7143	0.7143	0.8571	0.9429	0.9429	0.9429	0.9429
Loess BR FP rate	0	0	0	0.0286	0.0286	0.0571	0.0571	0.0571	0.0857	0.1143	0.2286	0.2286	0.2571	0.4	0.5143

Identification for Chipper (same % as MSC above):

α_0	0.05	0.1	0.15	0.2	α_0^*	0.21	0.22	0.23	0.24	0.25	0.3	0.35	0.4	0.45	0.5
Chipper TP	0	2	4	16	20	21	23	24	24	29	30	30	31	34	
Chipper FP	0	0	0	2	2	2	3	3	3	6	9	9	12	17	
Chipper FN	35	33	31	19	15	14	12	11	11	6	5	5	4	1	
Chipper TN	35	35	35	33	33	33	32	32	32	29	26	26	23	18	
Chipper TP rate	0	0.0571	0.1143	0.4571	0.5143	0.5714	0.6	0.6571	0.6857	0.6857	0.8286	0.8571	0.8571	0.8857	0.9714
Chipper FP rate	0	0	0	0.0571	0.0571	0.0571	0.0571	0.0857	0.0857	0.0857	0.1714	0.2571	0.2571	0.3429	0.4857

Table S4: True positives (TP), false positives (FP), false negatives (FN) and true negatives (TN) out of 35 enriched and 35 unenriched confirmed targets for 5 instances: 1). regular MSC (with initial shift and rotation onto principal axis) while maintaining FDR level of 0.1. The parameter α_0^* represents the parameters chosen for the three replicates by our jump method. That is 0.2 for replicate A, 0.21 for replicate B and 0.2 for replicate C. 2). BR, while maintaining same percentage of identified targets as first instance. 3). BR with respect to the principal axis of the data, while maintaining same percentage of identified targets as first instance. 4). BR with initial LOESS normalization, while maintaining same percentage of identified targets as first instance. 5). Chipper, while maintaining same percentage of identified targets as first instance.

Identification for MSC without initial transformation:

α_0	0.05	$\alpha_0^\star$	0.1	0.15	0.2	0.21	0.22	0.23	0.24	0.25	0.3	0.35	0.4	0.45	0.5
MS TP	18	24	24	28	31	31	32	32	32	32	33	33	33	33	33
MS FP	1	1	1	3	7	7	9	10	10	10	10	10	11	19	21
MS FN	17	11	11	7	4	4	3	3	3	3	2	2	2	2	2
MS TN	34	34	34	32	28	28	26	25	25	25	25	25	24	16	14
MS TP rate	0.5143	0.6857	0.6857	0.8	0.8857	0.8857	0.9143	0.9143	0.9143	0.9143	0.9429	0.9429	0.9429	0.9429	0.9429
MS FP rate	0.0286	0.0286	0.0286	0.0857	0.2	0.2	0.2571	0.2857	0.2857	0.2857	0.2857	0.2857	0.3143	0.5429	0.6

Identification for BR (same % as MSC above):

α_0	0.05	$\alpha_0^\star$	0.1	0.15	0.2	0.21	0.22	0.23	0.24	0.25	0.3	0.35	0.4	0.45	0.5
BR TP	16	22	22	26	30	31	31	33	33	33	33	33	33	33	33
BR FP	1	3	3	5	7	7	7	7	9	9	10	10	12	20	23
BR FN	19	13	13	9	5	4	4	2	2	2	2	2	2	2	2
BR TN	34	32	32	30	28	28	28	28	26	26	25	25	23	15	12
BR TP rate	0.4571	0.6286	0.6286	0.7429	0.8571	0.8857	0.8857	0.9429	0.9429	0.9429	0.9429	0.9429	0.9429	0.9429	0.9429
BR FP rate	0.0286	0.0857	0.0857	0.1429	0.2	0.2	0.2	0.2	0.2571	0.2571	0.2857	0.2857	0.3429	0.5714	0.6571

Identification for BR w.r.t. principal axis (same % as MSC above):

α_0	0.05	$\alpha_0^\star$	0.1	0.15	0.2	0.21	0.22	0.23	0.24	0.25	0.3	0.35	0.4	0.45	0.5
BR pca TP	18	23	23	25	27	27	27	27	28	28	28	28	29	30	32
BR pca FP	0	1	1	4	6	7	7	7	7	7	8	8	9	15	19
BR pca FN	17	12	12	10	8	8	8	8	7	7	7	7	6	5	3
BR pca TN	35	34	34	31	29	28	28	28	28	28	27	27	26	20	16
BR pca TP rate	0.5143	0.6571	0.6571	0.7143	0.7714	0.7714	0.7714	0.7714	0.8	0.8	0.8	0.8	0.8286	0.8571	0.9143
BR pca FP rate	0	0.0286	0.0286	0.1143	0.1714	0.2	0.2	0.2	0.2	0.2	0.2286	0.2286	0.2571	0.4286	0.5429

Identification for BR with LOESS normalization (same % as MSC above):

α_0	0.05	$\alpha_0^\star$	0.1	0.15	0.2	0.21	0.22	0.23	0.24	0.25	0.3	0.35	0.4	0.45	0.5
Loess BR TP	17	24	24	26	29	30	30	30	31	32	33	33	33	33	33
Loess BR FP	1	2	2	5	7	7	8	8	8	8	8	8	9	17	18
Loess BR FN	18	11	11	9	6	5	5	5	4	3	2	2	2	2	2
Loess BR TN	34	33	33	30	28	28	27	27	27	27	27	27	26	18	17
Loess BR TP rate	0.4857	0.6857	0.6857	0.7429	0.8286	0.8571	0.8571	0.8571	0.8857	0.9143	0.9429	0.9429	0.9429	0.9429	0.9429
Loess BR FP rate	0.0286	0.1429	0.0571	0.1429	0.2	0.2	0.2286	0.2286	0.2286	0.2286	0.2286	0.2286	0.2571	0.4857	0.5143

Identification for Chipper (same % as MSC above):

α_0	0.05	$\alpha_0^\star$	0.1	0.15	0.2	0.21	0.22	0.23	0.24	0.25	0.3	0.35	0.4	0.45	0.5
Chipper TP	16	20	20	24	27	29	29	29	29	29	30	30	30	32	34
Chipper FP	2	2	2	4	5	5	6	7	7	7	9	9	10	15	17
Chipper FN	19	15	15	11	8	6	6	6	6	6	5	5	5	3	1
Chipper TN	33	33	33	31	30	30	29	28	28	28	26	26	25	20	18
Chipper TP rate	0.4571	0.5714	0.5714	0.6857	0.7714	0.8286	0.8286	0.8286	0.8286	0.8286	0.8571	0.8571	0.8571	0.9143	0.9714
Chipper FP rate	0.0571	0.0571	0.0571	0.1143	0.1429	0.1429	0.1714	0.2	0.2	0.2	0.2571	0.2571	0.2857	0.4286	0.4857

Table S5: True positives (TP), false positives (FP), false negatives (FN) and true negatives (TN) out of 35 enriched and 35 unenriched confirmed targets for 5 instances: 1). MSC without initial shift and rotation while maintaining FDR level of 0.1. The parameter $\alpha_0^\star$ represents the parameters chosen for the three replicates by our jump method. That is 0.1 for replicate A, 0.11 for replicate B and 0.07 for replicate C. 2). BR, while maintaining same percentage of identified targets as first instance. 3). BR with respect to the principal axis of the data, while maintaining same percentage of identified targets as first instance. 4). BR with initial LOESS normalization, while maintaining same percentage of identified targets as first instance. 5). Chipper, while maintaining same percentage of identified targets as first instance.