CalMaTe Supplementary Note: Assessment of TumorBoost and CalMaTe based on tumor/normal pair TCGA-23-1027 in the Affymetrix GenomeWideSNP_6 data set TCGA,OV,CRMAv2

Maria Ortiz-Estevez, Ander Aramburu, Henrik Bengtsson, Pierre Neuvial, Angel Rubio

March 17, 2012

Contents

1 Introduction

In order to formally evaluate the influence of CalMaTe on signal to noise ratio, we have used receiver operating characteristic (ROC) analysis on several known change points. ROC analysis was performed as described in Bengtsson *et al.* (2009, 2010). We refer to these papers for a thorough description of this report. Here, we explain in detail how this evaluation is carried out a specific tumor sample: TCGA-23-1027. We chose the same change points as those used for the evaluation of the TumorBoost method (Bengtsson *et al.*, 2010) in order to facilitate comparison between CalMaTe and TumorBoost, and interpretation of the results. The evaluation is done at set of change points corresponding to common types of copy number state transitions, and at one region with no change point (negative control) as summarized in Table 1.1 (also Bengtsson *et al.* (2010, Table 1)).

1.1 Data

1.1.1 Data set

The evaluation in this report is based on the tumor/normal pair for individual TCGA-23-1027 in the data set TCGA,OV,CRMAv2.

1.1.2 List of change points

For this data set, we have selected a few regions for which one can safely assume that there exists a single copy number change point. By definition, each change point separates two sets of genomic loci such that the true Total Copy Number (TCN) and Decrease in Heterozygosity (DH) is the same within one set of loci but differs between the two sets. These regions were selected by visual inspection. For each region we chose a large enough safety margin to make our evaluation independent of the uncertainty on the true location of the change point.

Chr	Region	Change point	Margin	Before	After
2	108-140	124	0.5	normal (1,1)	gain (1,2)
2	125-157	141	0.5	gain (1,2)	copy neutral LOH (0,2)
10	80-109	94	0.5	normal (1,1)	deletion (0,1)
10	106.5-113.5	110	0.5	deletion (0,1)	copy neutral LOH (0,2)
2	55-75	65	0.5	normal (1,1)	gain (1,2)

Table 1.1: Regions in TCGA-23-1027 used for the evaluation and that each contain a single changepoint. The last region is a negative control. All positions and lengths are in units of Mb.

1.2 Methods

1.2.1 Preprocessing method

The data was generated on the Affymetrix GenomeWideSNP_6 chip type. There is one CEL file per hybridization. Each CEL file was preprocessed separately using CRMAv2 (Bengtsson *et al.*, 2009), without relying neither on reference samples nor prior estimates.

1.2.2 Postprocessing methods

The result of the above preprocessing method are estimates of total copy numbers (TCNs) and allele B fractions (BAFs). The (TCN,BAF) signals are then postprocessed by TumorBoost as well as CalMaTe. Because TumorBoost models each tumor-normal pair independently, it can be applied to the studied pair

without the need for reference samples (Bengtsson *et al.*, 2010). The CalMaTe method is a multi-array method, which estimates SNP effects from a set of reference samples, and then correct all samples based on these estimates. For details on the reference set used, see descriptions elsewhere.

1.2.3 Identifying heterozygous SNPs

When evaluating the power to detect a change point using the BAFs, we identify the heterozyous SNPs from the matched normal to calculate the Decrease of Heterozygousity (DH) signals. We use a naive genotyping algorithm (Bengtsson *et al.*, 2010) to identify the heterozygous SNPs. For each method, the genotypes are called based on the BAFs in generates.

1.2.4 Stratification on genotype confidence scores

In Bengtsson *et al.* (2010) it was shown that it is possible to slightly improve the discrimination power in DH by filtering out SNPs for which the genotype confidence scores are low. In this report we choose not do to this, i.e. all heterozygous SNPs are used.

1.2.5 Evaluation

Each change point (except for the negative control) corresponds to a real change in both TCN and BAF, between what we call a "negative" state (defined as the closest to TCN=2) and a "positive" state (the other one). We focus on a genomic region surrounding the change point. Consider a threshold value τ. For BAF signals, we calculate the Decrease in Heterozygosity (DH) defined for each heterozygous SNP j as $\mathrm{DH}_j = 2|\mathrm{BAF}_j - 1/2|$. DH is appropriate for evaluation purposes because contrary for BAF its distribution only has one mode, as discussed e.g. in Bengtsson *et al.* (2010). A SNP j is then classified as "positive" if $DH_j \geq \tau$, and as "negative" otherwise, and we report the true-positive rate (TPR) and the true-negative rate (TNR) in the region. Repeating this for each possible value of τ, we build a ROC curve for DH by plotting TPR against TNR in the genomic region. A ROC curve can be built along similar lines for TCN, where a locus j is classified as "positive" if $\mathrm{TCN}_j \geq \tau$ if the positive state has greater TCN than the negative state, and $\mathrm{TCN}_j \leq \tau$ otherwise. Using this strategy, we plot a TCN and a DH ROC curve for each method to be compared, e.g. genomic signals before and after CalMaTe. Specifically, we compare the following sets of data:

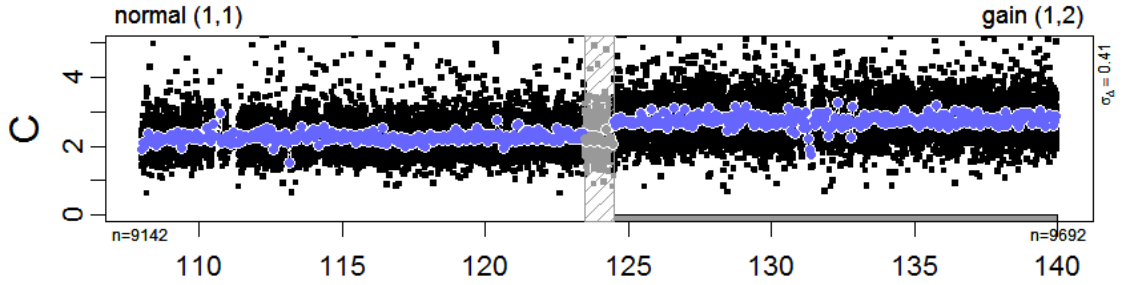
1. **"CRMAv2"**: CRMAv2 preprocessed signals without postprocessing.
2. **"CRMAv2,TumorBoost"**: CRMAv2 preprocessed signals with TumorBoost postprocessing.
3. **"CRMAv2,CalMaTe"**: CRMAv2 preprocessed signals with CalMaTe postprocessing.

The assessment is based on ROC analysis at the full resolution as well as various smoothed resolution with bin sizes $h = \{5, 10, 25, 50\}$ kb.

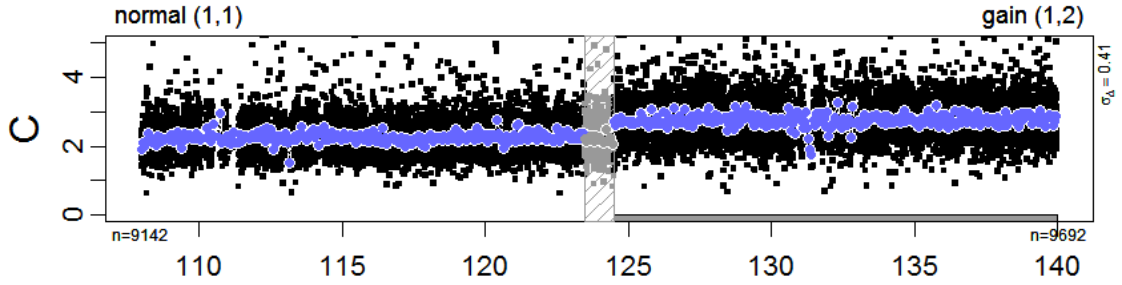
2 Region: TCGA-23-1027:Chr2@108-140

2.1 TCN and DH tracks

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

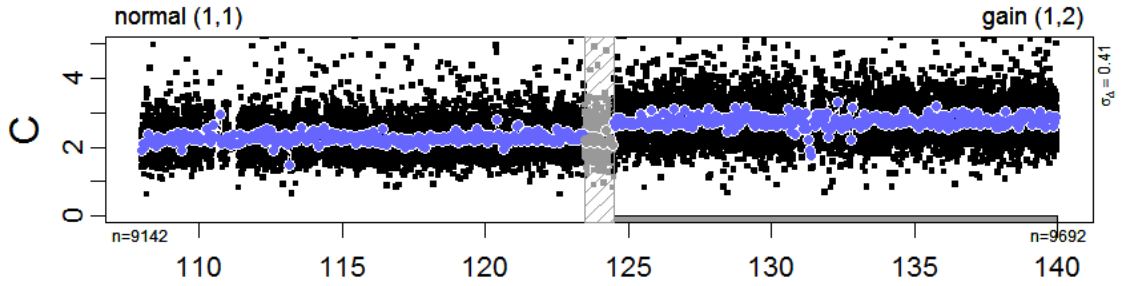
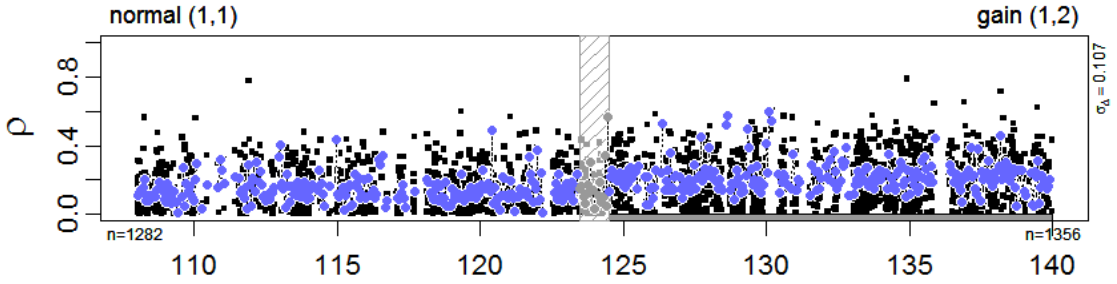
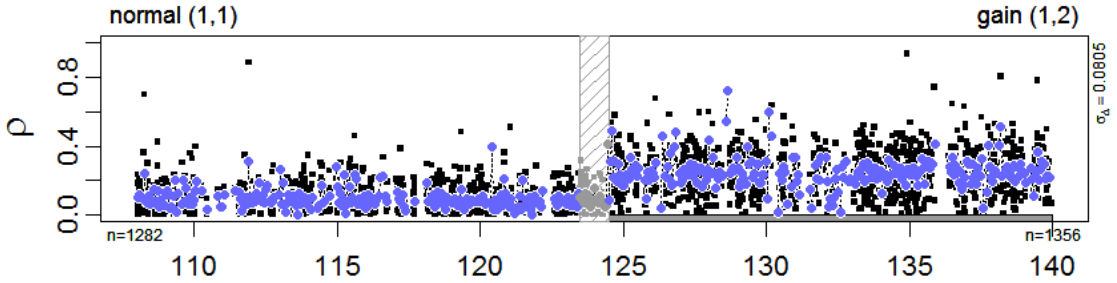


Figure 2.1: TCNs of the different methods for region TCGA-23-1027:Chr2@108-140,cp=124+/-0.5,s=0/1. The sets of loci to the left and the right belong to two different copy-number states, separated by a change points. The exact location of the change point is unknown, except that it is located within safety margin (shaded with diagonal lines). One of the copy-number states is referred to as the "negative" state ('normal (1,1)') and the other as the "positive" state ('gain (1,2)'), where the latter are highlighted with a solid bar beneath. Legend: Full resolution TCNs are displayed in black and smoothed (bin width 50 kb) ones in blue. Loci within the safety margin are displayed in gray and excluded from all analysis.

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

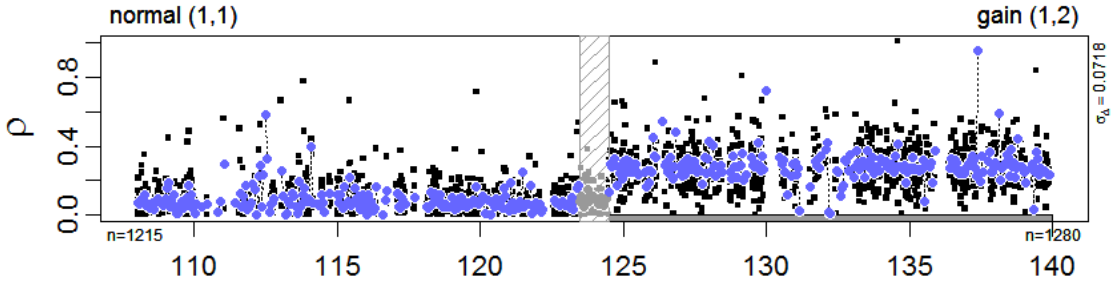


Figure 2.2: DHs of the different methods for region TCGA-23-1027:Chr2@108-140,cp=124+/-0.5,s=0/1. DHs are only defined for SNPs that are genotyped to be heterozygous based on data from the corresponding method. The sets of loci to the left and the right belong to two different copy-number states, separated by a change points. The exact location of the change point is unknown, except that it is located within safety margin (shaded with diagonal lines). One of the copy-number states is referred to as the "negative" state ('normal (1,1)') and the other as the "positive" state ('gain (1,2)'), where the latter are highlighted with a solid bar beneath. Legend: Full resolution DHs are displayed in black and smoothed (bin width 50 kb) ones in blue. Loci within the safety margin are displayed in gray and excluded from all analysis.

2.2 TCN and BAF signal density plots

(a) normal (1,1). (b) gain (1,2).

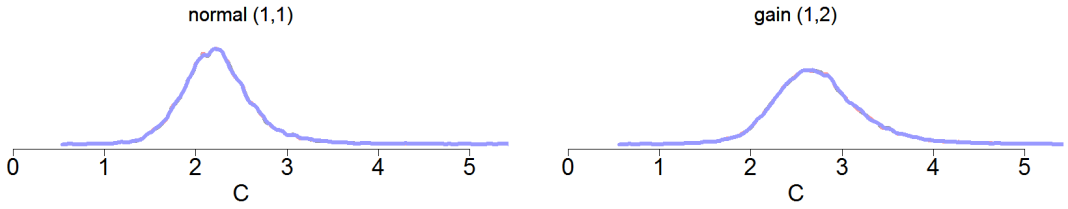


Figure 2.3: Densities of TCNs for region TCGA-23-1027:Chr2@108-140,cp=124+/-0.5,s=0/1. Legend: 'CRMAv2' (dashed gray), 'CRMAv2,TumorBoost' (dotted light red), and 'CRMAv2,CalMaTe' (solid light blue).

(a) normal (1,1). (b) gain (1,2).

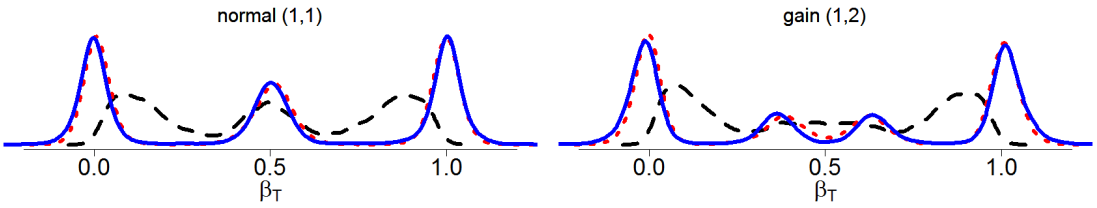
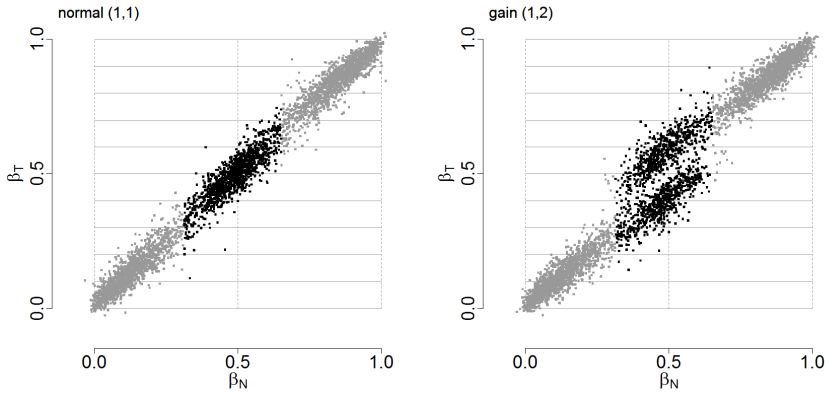


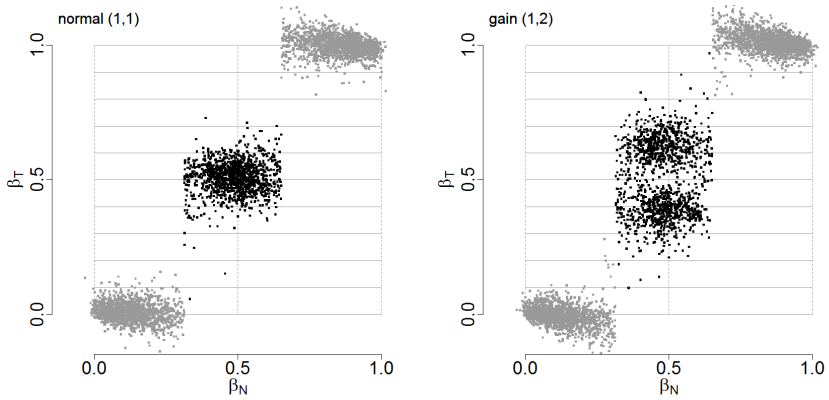
Figure 2.4: Densities of BAFs for region TCGA-23-1027:Chr2@108-140,cp=124+/-0.5,s=0/1. Legend: 'CRMAv2' (dashed black), 'CRMAv2,TumorBoost' (dotted red), and 'CRMAv2,CalMaTe' (solid blue).

2.3 Tumor versus normal BAF plots

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

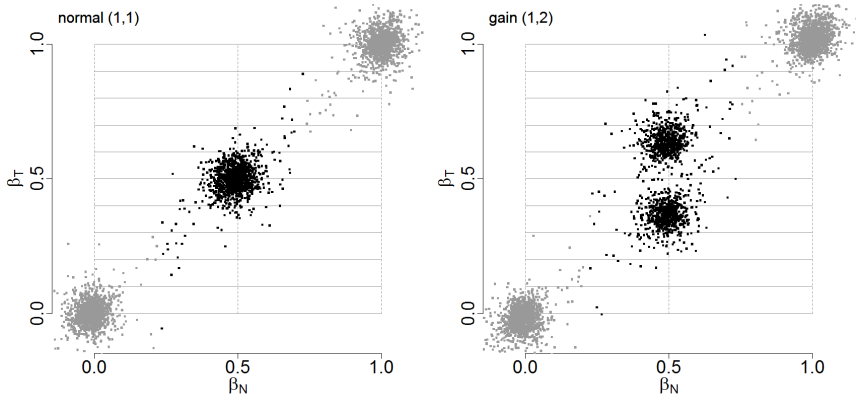
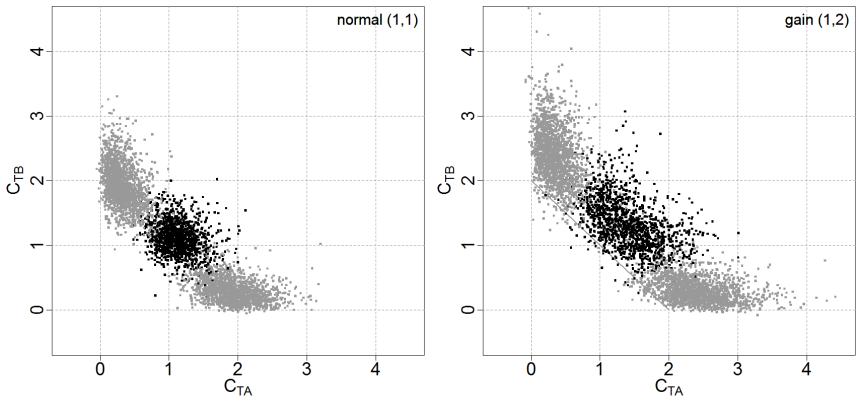


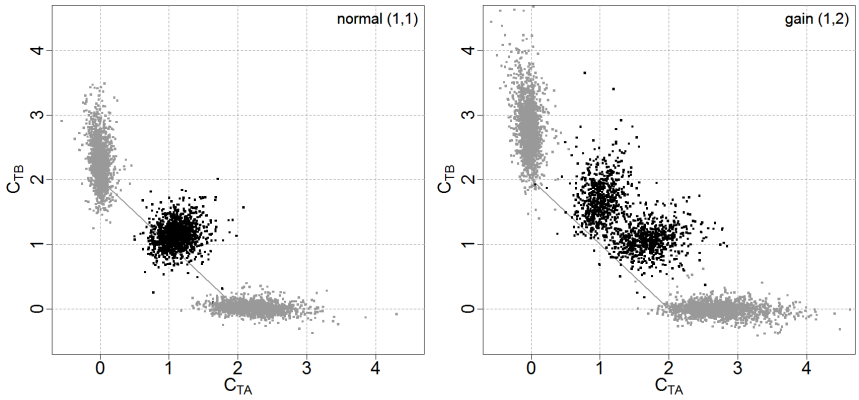
Figure 2.5: Tumor versus normal BAFs of 'CRMAv2', 'CRMAv2,TumorBoost', and 'CRMAv2,CalMaTe' for region TCGA-23-1027:Chr2@108-140,cp=124+/-0.5,s=0/1. Legend: SNPs genotyped (based on data from the corresponding method) as being heterozygous (homozygous) are displayed in black (gray).

2.4 ASCN plots

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

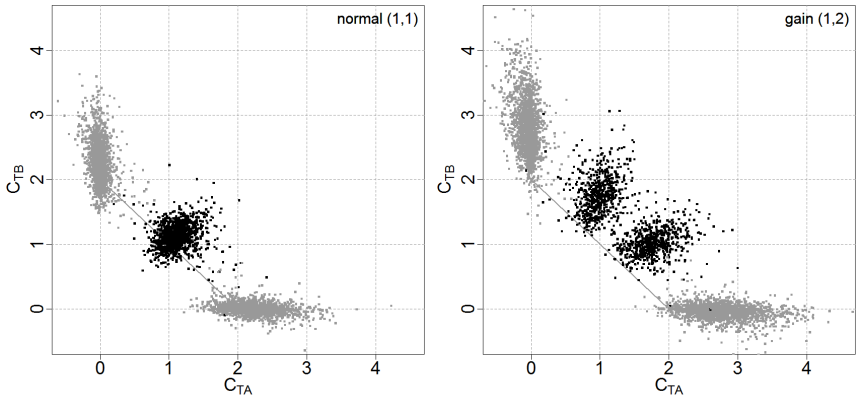
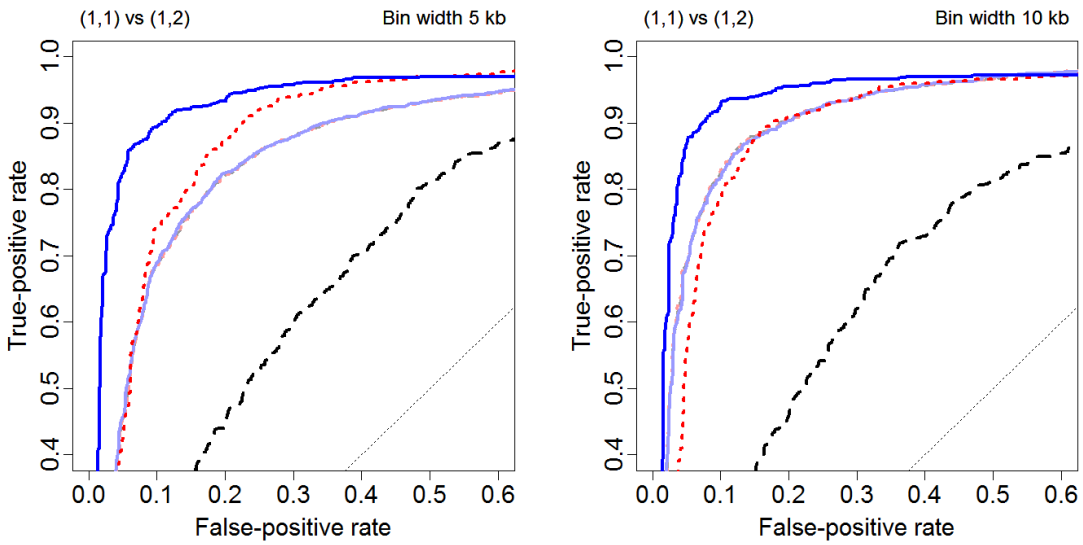


Figure 2.6: Tumor ASCNs (allele-specific copy numbers) of 'CRMAv2', 'CRMAv2,TumorBoost', and 'CRMAv2,CalMaTe' for region TCGA-23-1027:Chr2@108-140,cp=124+/-0.5,s=0/1. Legend: SNPs called heterozygous (homozygous) by the corresponing method are displayed in black (gray).

2.5 ROC curves

(a) Bin width 5 kb.

(b) Bin width 10 kb.



(c) Bin width 25 kb.

(d) Bin width 50 kb.

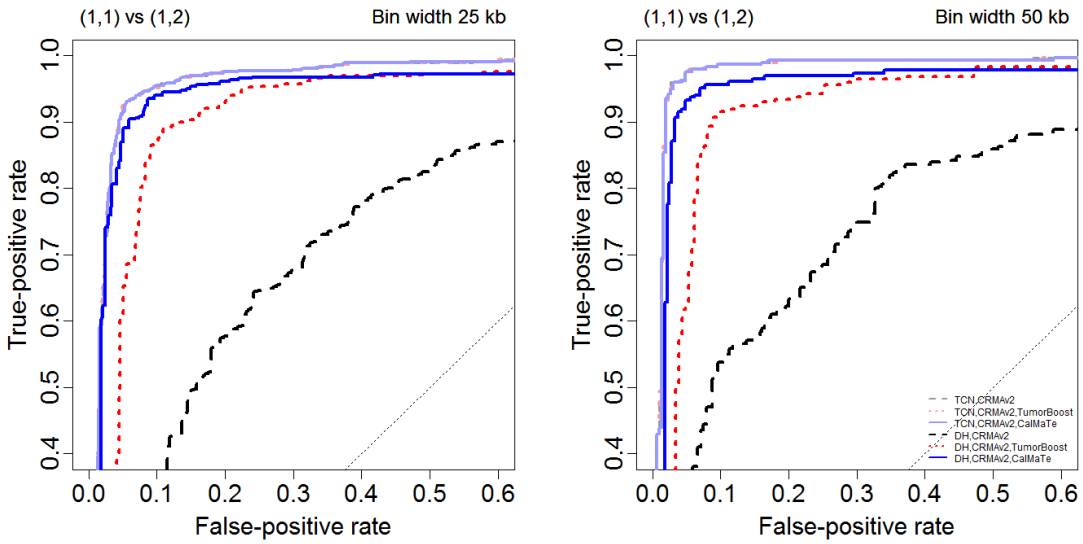
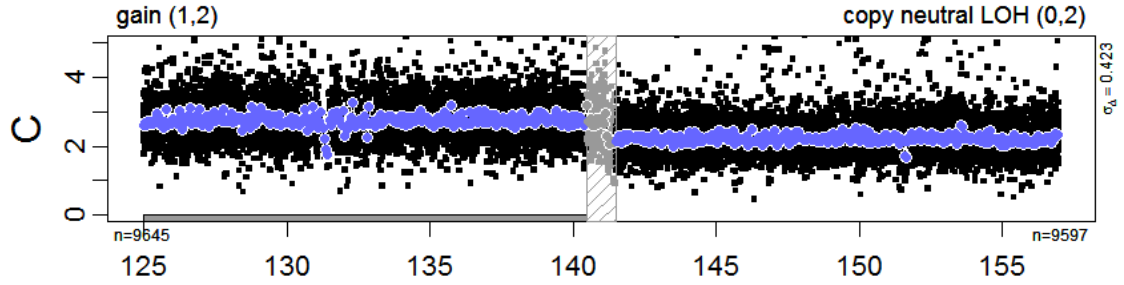


Figure 2.7: ROC curves for detecting the change point based on TCNs and DHs for each preprocessing method at different amounts of smoothing ("resolutions") for region TCGA-23-1027:Chr2@108-140,cp=124+/-0.5,s=0/1. Smoothing was done by averaging (using the mean() function) in non-overlapping bins. Legend: The TCN and DH curves for the same method are depicted with the same line type and color with the difference that the TCN curve use a lighter version of color for 'CRMAv2' (dashed black), 'CRMAv2,TumorBoost' (dotted red), and 'CRMAv2,CalMaTe' (solid blue).

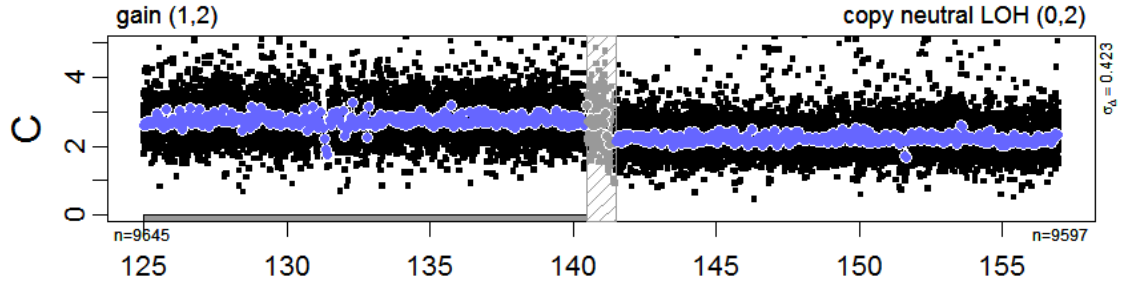
3 Region: TCGA-23-1027:Chr2@125.0-157.0

3.1 TCN and DH tracks

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

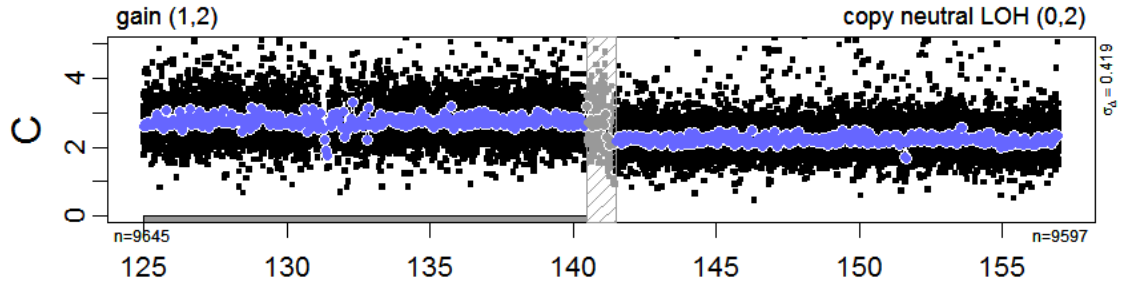
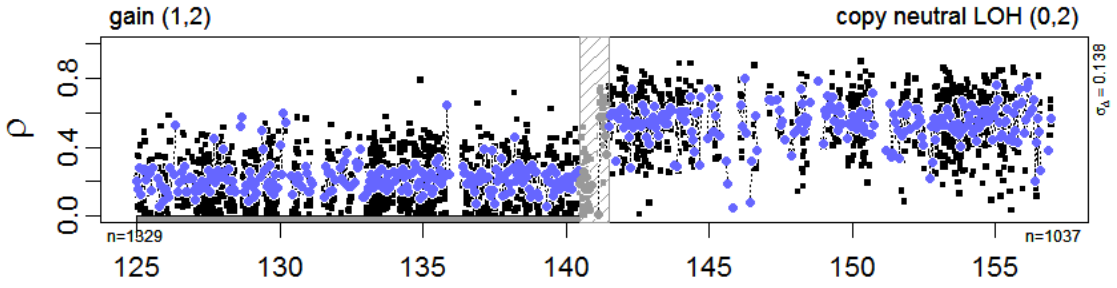
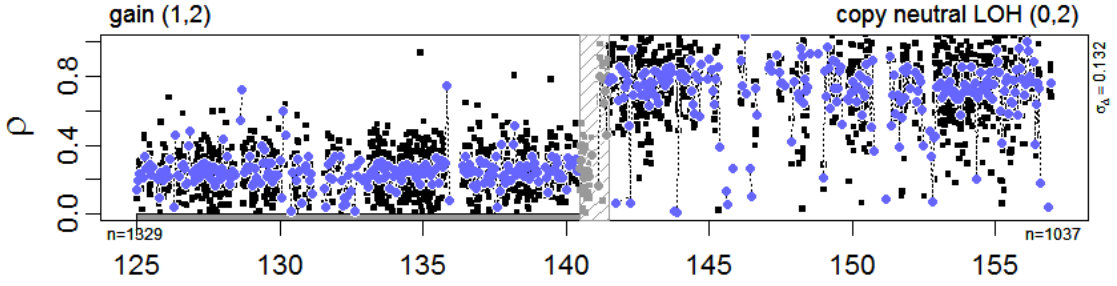


Figure 3.1: TCNs of the different methods for region TCGA-23-1027:Chr2@125.0-157.0,cp=141.0+/-0.5,s=1/3. The sets of loci to the left and the right belong to two different copy-number states, separated by a change points. The exact location of the change point is unknown, except that it is located within safety margin (shaded with diagonal lines). One of the copy-number states is referred to as the "negative" state ('gain (1,2)') and the other as the "positive" state ('copy neutral LOH (0,2)'), where the latter are highlighted with a solid bar beneath. Legend: Full resolution TCNs are displayed in black and smoothed (bin width 50 kb) ones in blue. Loci within the safety margin are displayed in gray and excluded from all analysis.

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

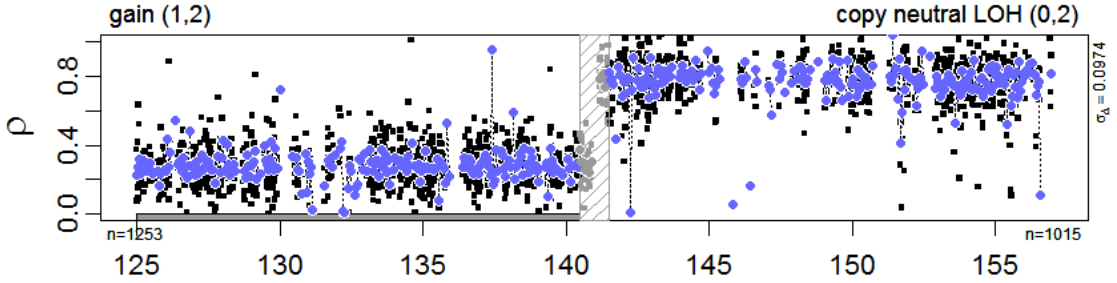


Figure 3.2: DHs of the different methods for region TCGA-23-1027:Chr2@125.0-157.0,cp=141.0+/-0.5,s=1/3. DHs are only defined for SNPs that are genotyped to be heterozygous based on data from the corresponding method. The sets of loci to the left and the right belong to two different copy-number states, separated by a change points. The exact location of the change point is unknown, except that it is located within safety margin (shaded with diagonal lines). One of the copy-number states is referred to as the "negative" state ('gain (1,2)') and the other as the "positive" state ('copy neutral LOH (0,2)'), where the latter are highlighted with a solid bar beneath. Legend: Full resolution DHs are displayed in black and smoothed (bin width 50 kb) ones in blue. Loci within the safety margin are displayed in gray and excluded from all analysis.

3.2 TCN and BAF signal density plots

(a) gain (1,2). (b) copy neutral LOH (0,2).

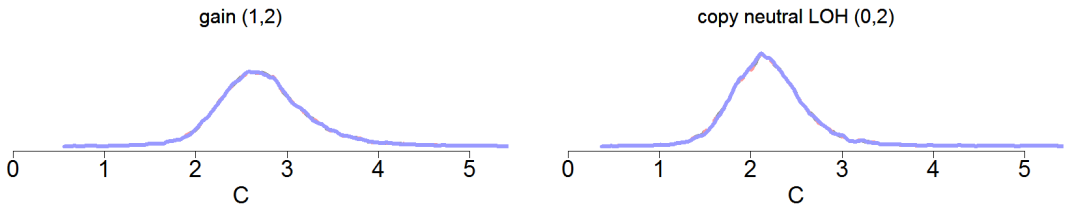


Figure 3.3: Densities of TCNs for region TCGA-23-1027:Chr2@125.0-157.0,cp=141.0+/-0.5,s=1/3. Legend: 'CRMAv2' (dashed gray), 'CRMAv2,TumorBoost' (dotted light red), and 'CRMAv2,CalMaTe' (solid light blue).

(a) gain (1,2). (b) copy neutral LOH (0,2).

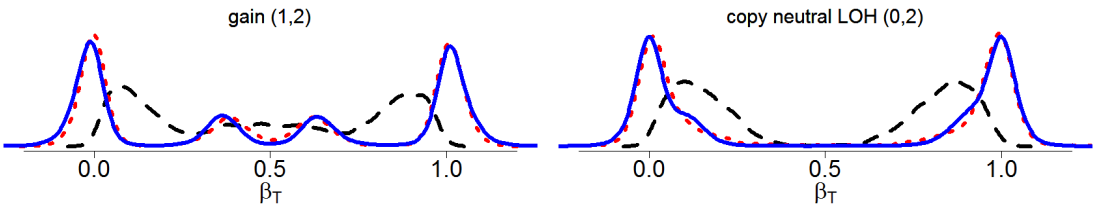
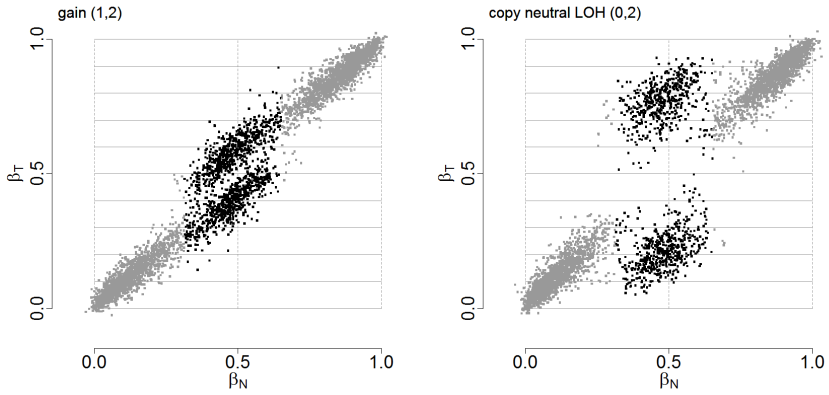


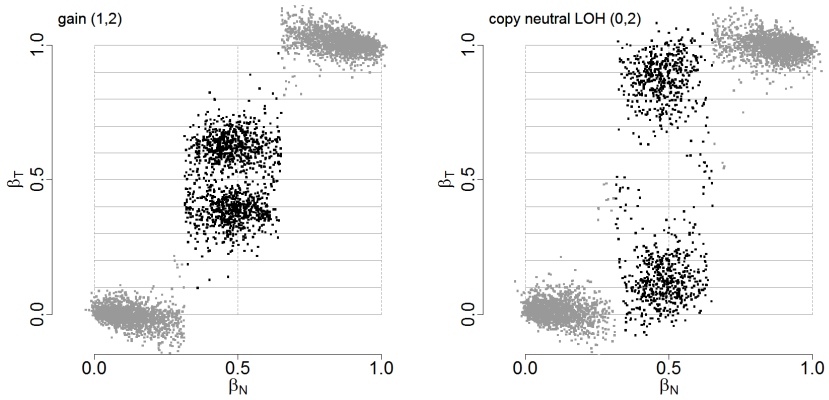
Figure 3.4: Densities of BAFs for region TCGA-23-1027:Chr2@125.0-157.0,cp=141.0+/-0.5,s=1/3. Legend: 'CRMAv2' (dashed black), 'CRMAv2,TumorBoost' (dotted red), and 'CRMAv2,CalMaTe' (solid blue).

3.3 Tumor versus normal BAF plots

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

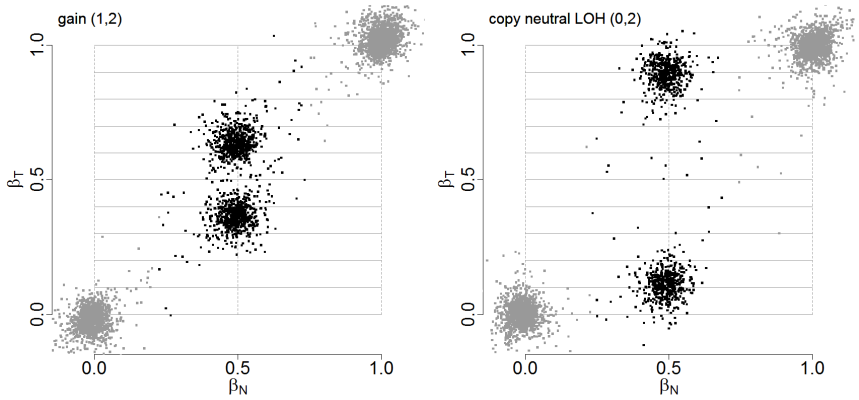
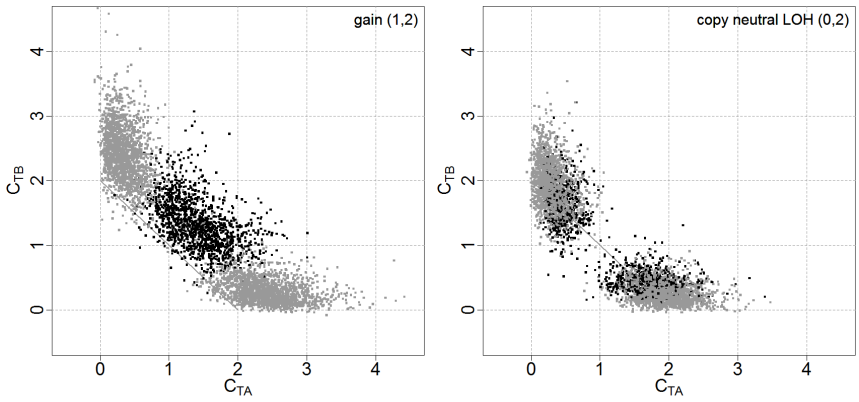


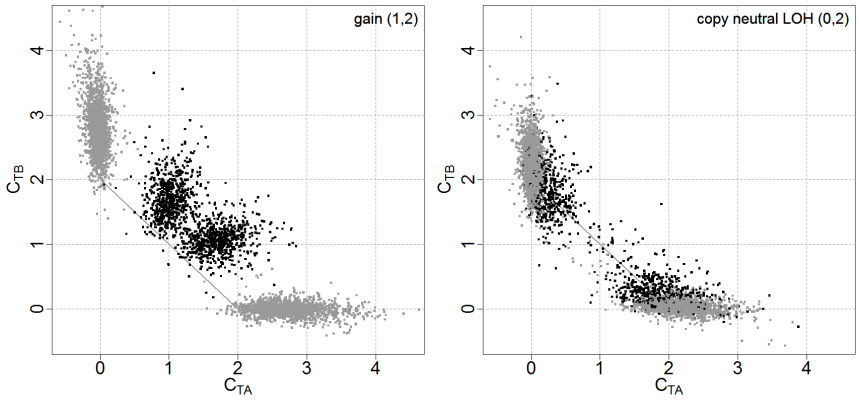
Figure 3.5: Tumor versus normal BAFs of 'CRMAv2', 'CRMAv2,TumorBoost', and 'CRMAv2,CalMaTe' for region TCGA-23-1027:Chr2@125.0-157.0,cp=141.0+/-0.5,s=1/3. Legend: SNPs genotyped (based on data from the corresponding method) as being heterozygous (homozygous) are displayed in black (gray).

3.4 ASCN plots

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

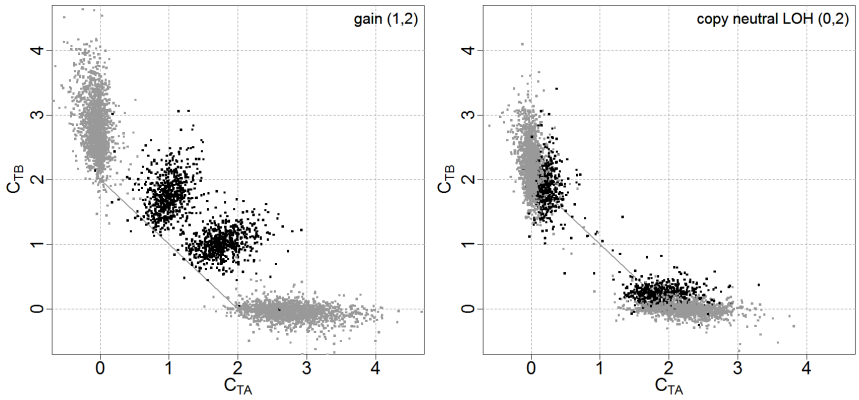
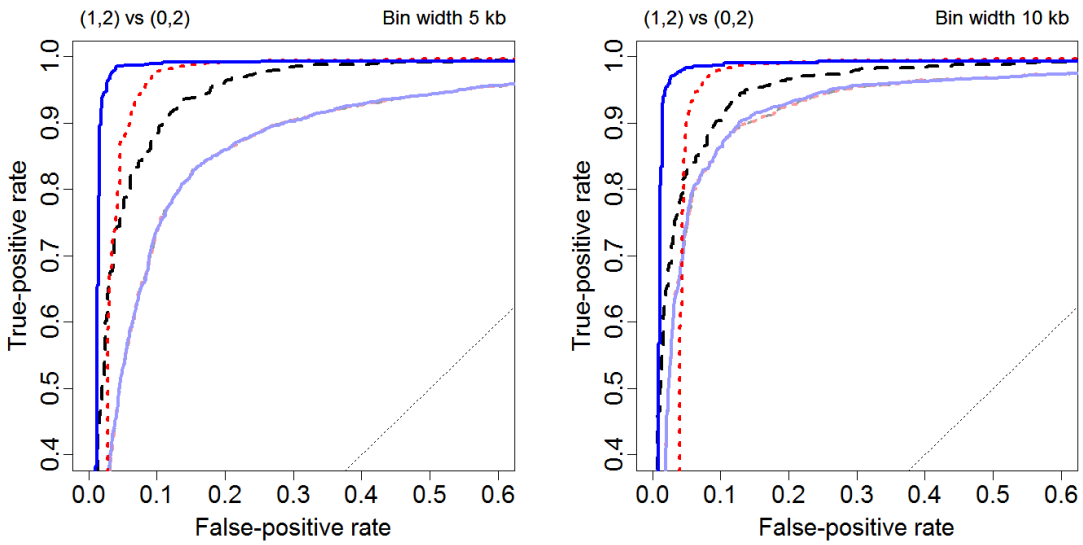


Figure 3.6: Tumor ASCNs (allele-specific copy numbers) of 'CRMAv2', 'CRMAv2,TumorBoost', and 'CRMAv2,CalMaTe' for region TCGA-23-1027:Chr2@125.0-157.0,cp=141.0+/-0.5,s=1/3. Legend: SNPs called heterozygous (homozygous) by the corresponing method are displayed in black (gray).

3.5 ROC curves

(a) Bin width 5 kb.

(b) Bin width 10 kb.



(c) Bin width 25 kb.

(d) Bin width 50 kb.

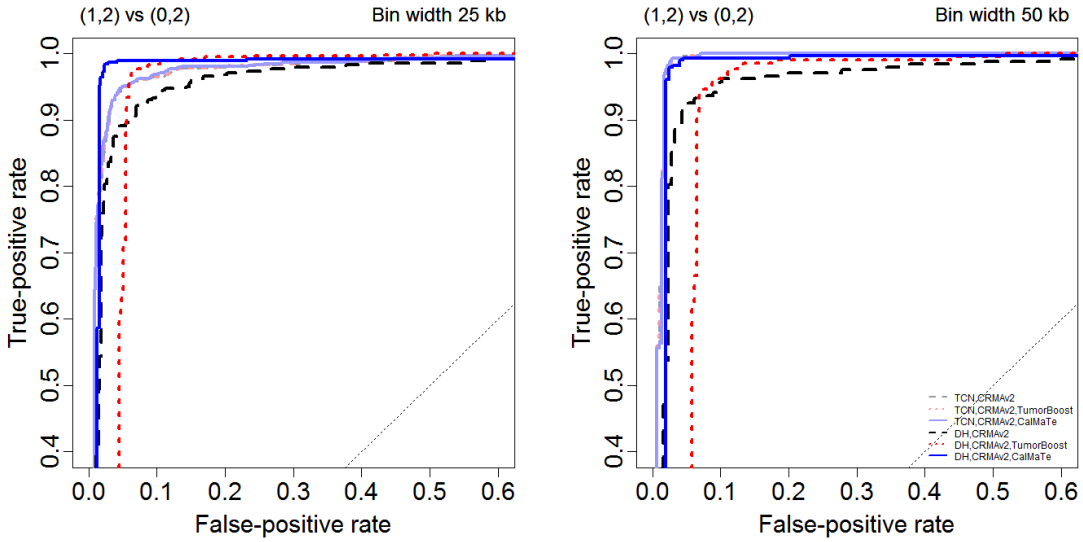
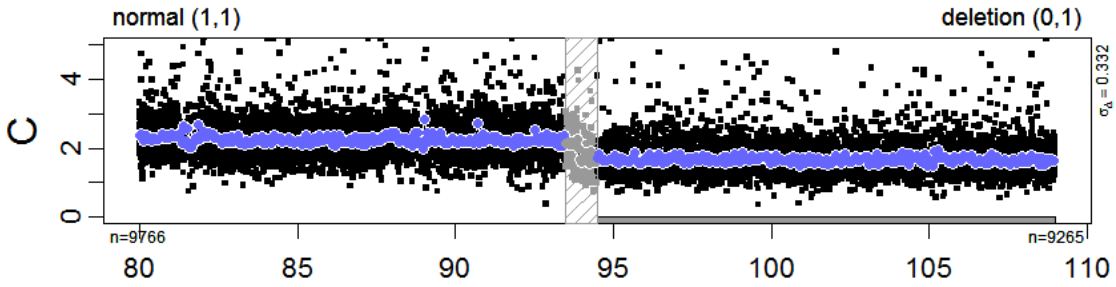


Figure 3.7: ROC curves for detecting the change point based on TCNs and DHs for each preprocessing method at different amounts of smoothing ("resolutions") for region TCGA-23-1027:Chr2@125.0-157.0,cp=141.0+/-0.5,s=1/3. Smoothing was done by averaging (using the mean() function) in non-overlapping bins. Legend: The TCN and DH curves for the same method are depicted with the same line type and color with the difference that the TCN curve use a lighter version of color for 'CRMAv2' (dashed black), 'CRMAv2,TumorBoost' (dotted red), and 'CRMAv2,CalMaTe' (solid blue).

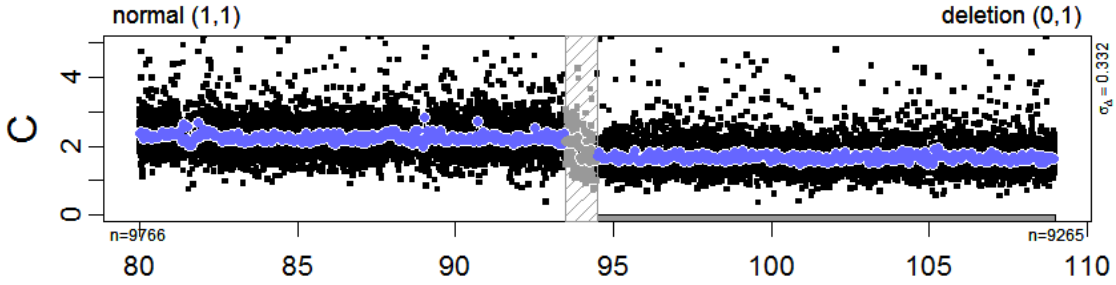
4 Region: TCGA-23-1027:Chr10@80-109

4.1 TCN and DH tracks

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

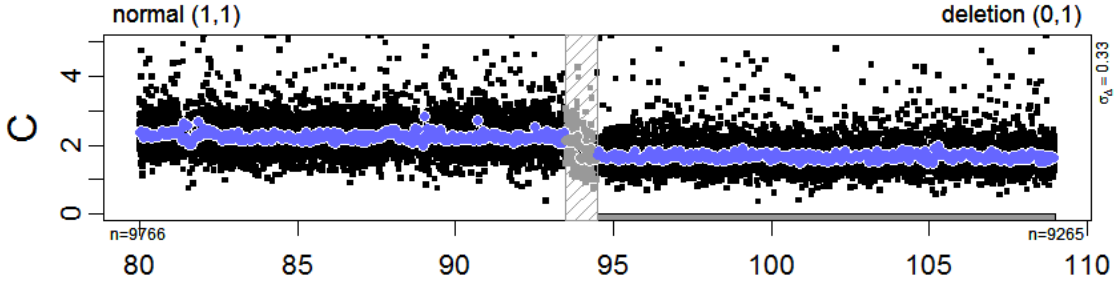
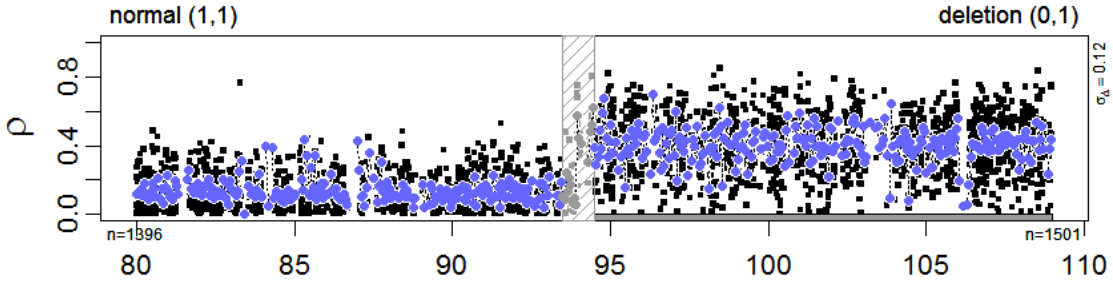
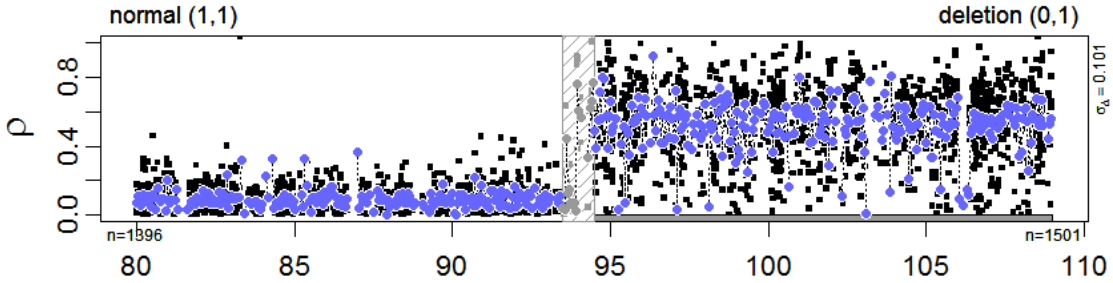


Figure 4.1: TCNs of the different methods for region TCGA-23-1027:Chr10@80-109,cp=94+/-0.5,s=0/2. The sets of loci to the left and the right belong to two different copy-number states, separated by a change points. The exact location of the change point is unknown, except that it is located within safety margin (shaded with diagonal lines). One of the copy-number states is referred to as the "negative" state ('normal (1,1)') and the other as the "positive" state ('deletion (0,1)'), where the latter are highlighted with a solid bar beneath. Legend: Full resolution TCNs are displayed in black and smoothed (bin width 50 kb) ones in blue. Loci within the safety margin are displayed in gray and excluded from all analysis.

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

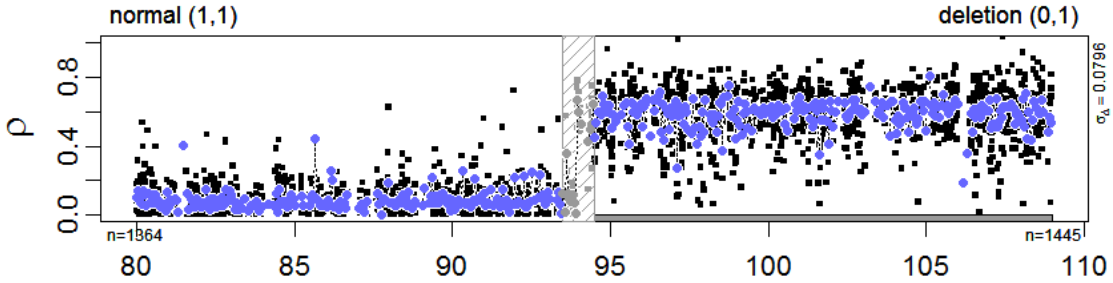


Figure 4.2: DHs of the different methods for region TCGA-23-1027:Chr10@80-109,cp=94+/-0.5,s=0/2. DHs are only defined for SNPs that are genotyped to be heterozygous based on data from the corresponding method. The sets of loci to the left and the right belong to two different copy-number states, separated by a change points. The exact location of the change point is unknown, except that it is located within safety margin (shaded with diagonal lines). One of the copy-number states is referred to as the "negative" state ('normal (1,1)') and the other as the "positive" state ('deletion (0,1)'), where the latter are highlighted with a solid bar beneath. Legend: Full resolution DHs are displayed in black and smoothed (bin width 50 kb) ones in blue. Loci within the safety margin are displayed in gray and excluded from all analysis.

4.2 TCN and BAF signal density plots

(a) normal (1,1). (b) deletion (0,1).

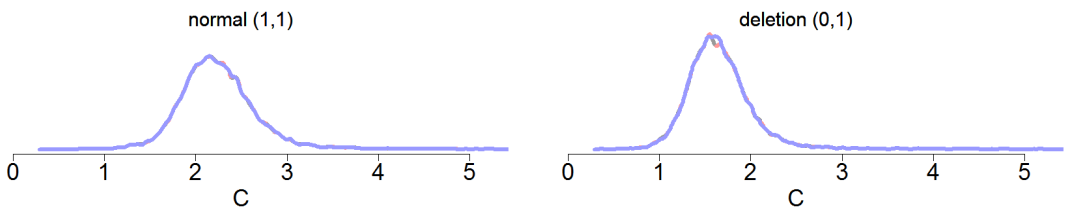


Figure 4.3: Densities of TCNs for region TCGA-23-1027:Chr10@80-109,cp=94+/-0.5,s=0/2. Legend: 'CRMAv2' (dashed gray), 'CRMAv2,TumorBoost' (dotted light red), and 'CRMAv2,CalMaTe' (solid light blue).

(a) normal (1,1). (b) deletion (0,1).

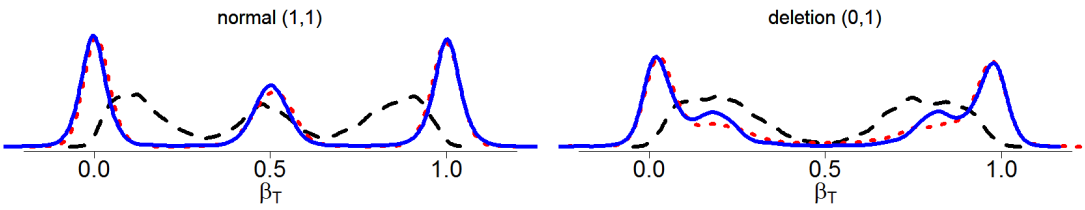
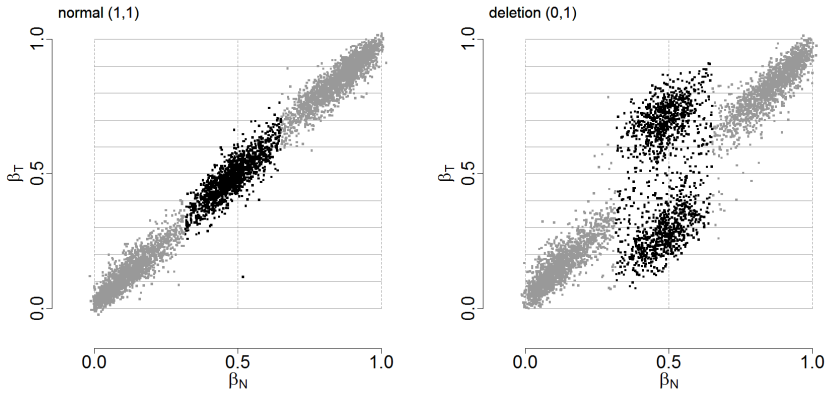


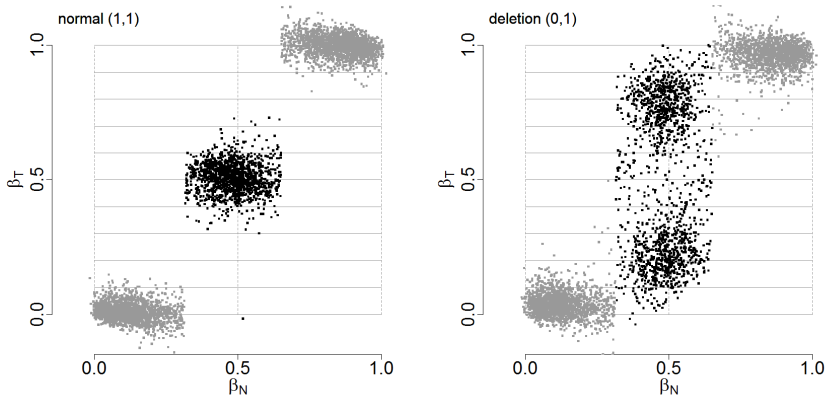
Figure 4.4: Densities of BAFs for region TCGA-23-1027:Chr10@80-109,cp=94+/-0.5,s=0/2. Legend: 'CRMAv2' (dashed black), 'CRMAv2,TumorBoost' (dotted red), and 'CRMAv2,CalMaTe' (solid blue).

4.3 Tumor versus normal BAF plots

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

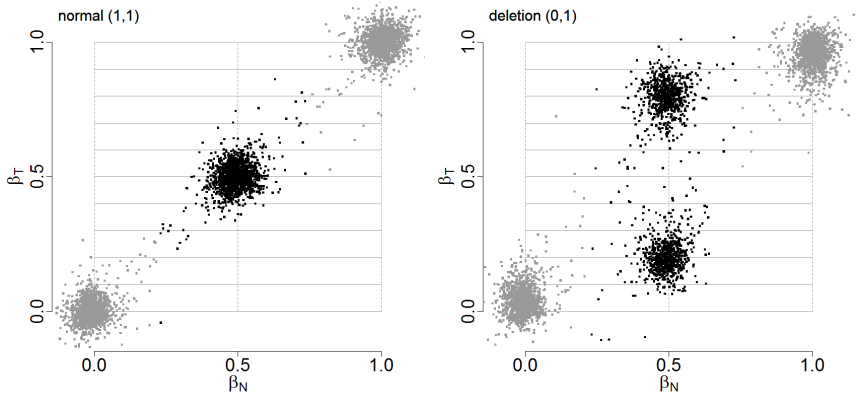
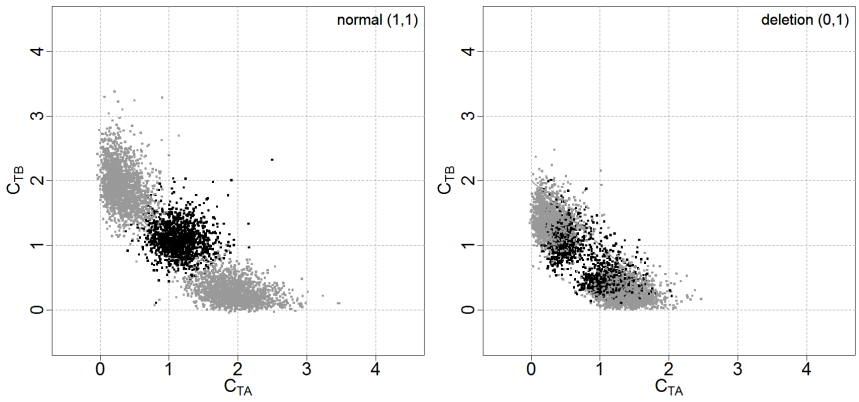


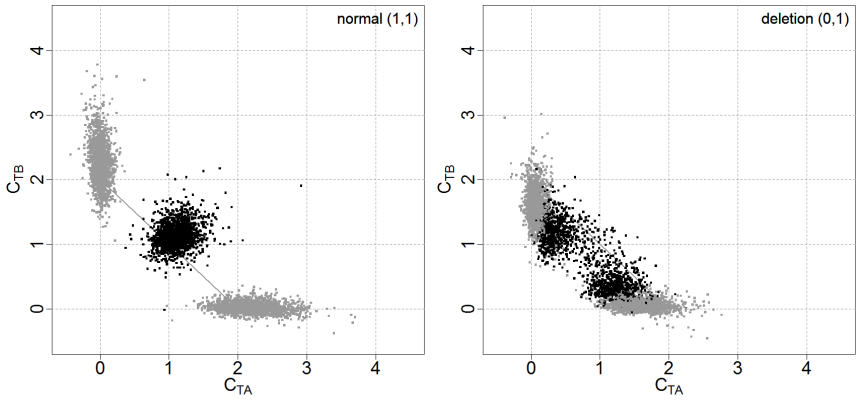
Figure 4.5: Tumor versus normal BAFs of 'CRMAv2', 'CRMAv2,TumorBoost', and 'CRMAv2,CalMaTe' for region TCGA-23-1027:Chr10@80-109,cp=94+/-0.5,s=0/2. Legend: SNPs genotyped (based on data from the corresponding method) as being heterozygous (homozygous) are displayed in black (gray).

4.4 ASCN plots

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

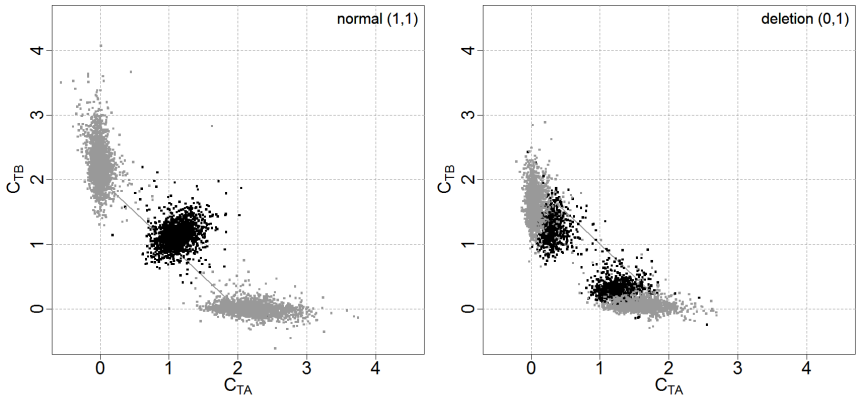
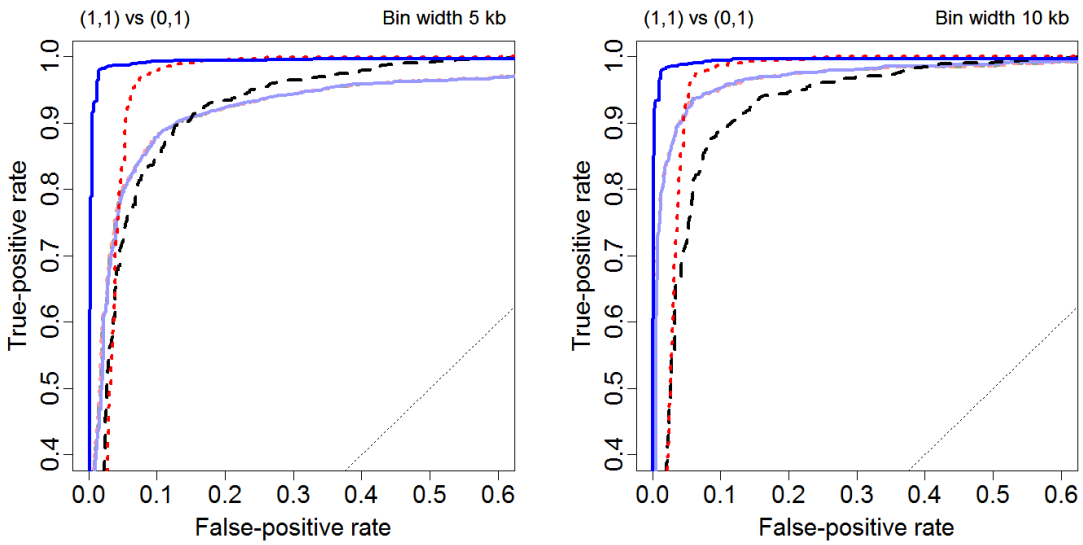


Figure 4.6: Tumor ASCNs (allele-specific copy numbers) of 'CRMAv2', 'CRMAv2,TumorBoost', and 'CRMAv2,CalMaTe' for region TCGA-23-1027:Chr10@80-109,cp=94+/-0.5,s=0/2. Legend: SNPs called heterozygous (homozygous) by the corresponing method are displayed in black (gray).

4.5 ROC curves

(a) Bin width 5 kb.

(b) Bin width 10 kb.



(c) Bin width 25 kb.

(d) Bin width 50 kb.

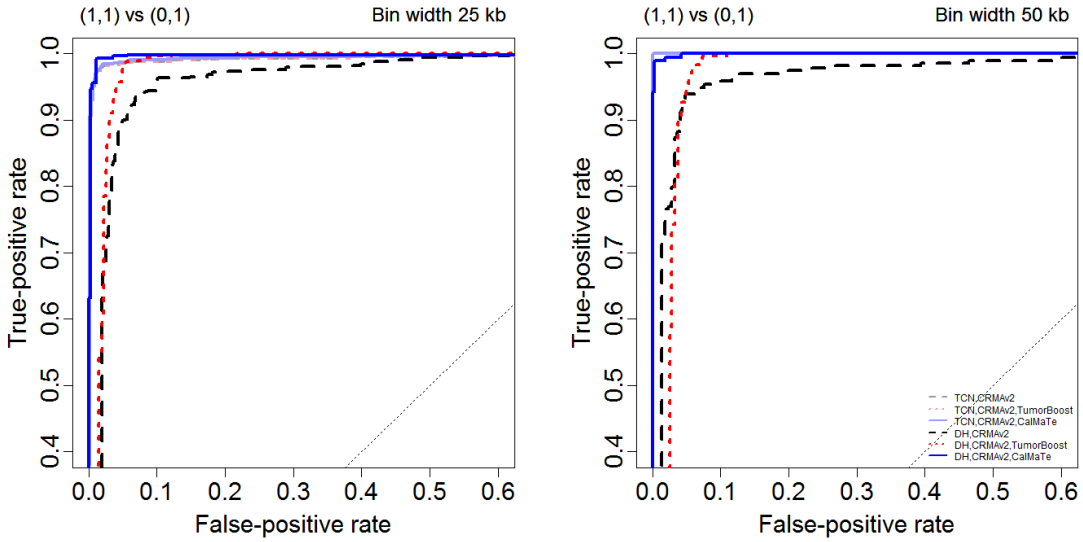
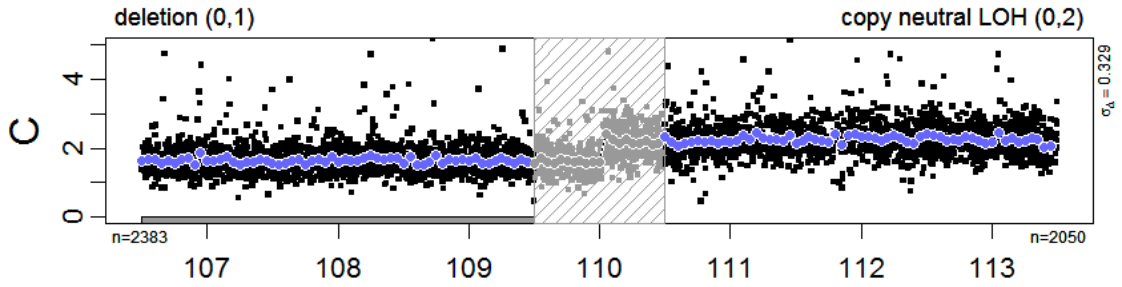


Figure 4.7: ROC curves for detecting the change point based on TCNs and DHs for each preprocessing method at different amounts of smoothing ("resolutions") for region TCGA-23-1027:Chr10@80-109,cp=94+/-0.5,s=0/2. Smoothing was done by averaging (using the mean() function) in non-overlapping bins. Legend: The TCN and DH curves for the same method are depicted with the same line type and color with the difference that the TCN curve use a lighter version of color for 'CRMAv2' (dashed black), 'CRMAv2,TumorBoost' (dotted red), and 'CRMAv2,CalMaTe' (solid blue).

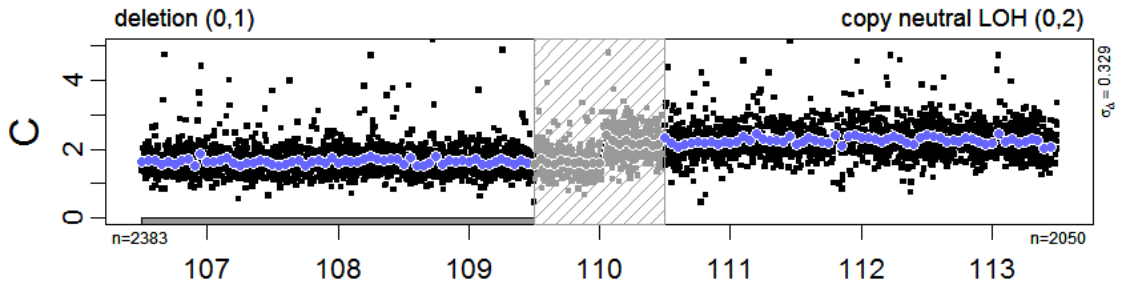
5 Region: TCGA-23-1027:Chr10@106.5-113.5

5.1 TCN and DH tracks

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

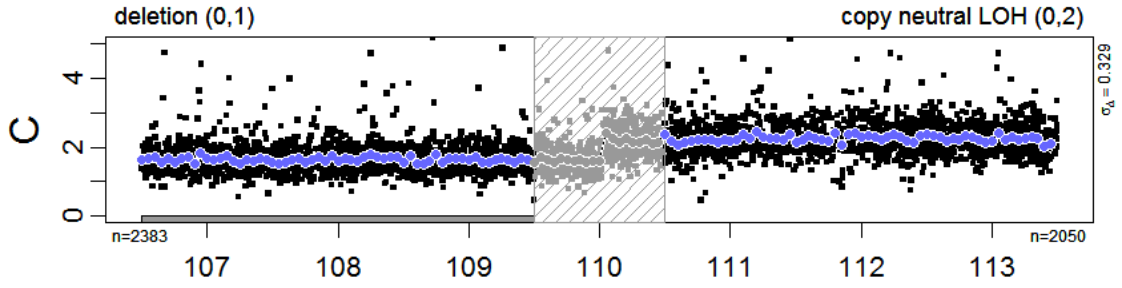
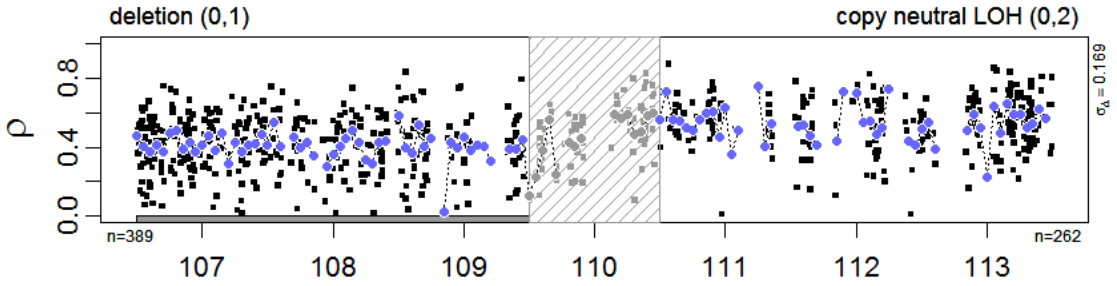
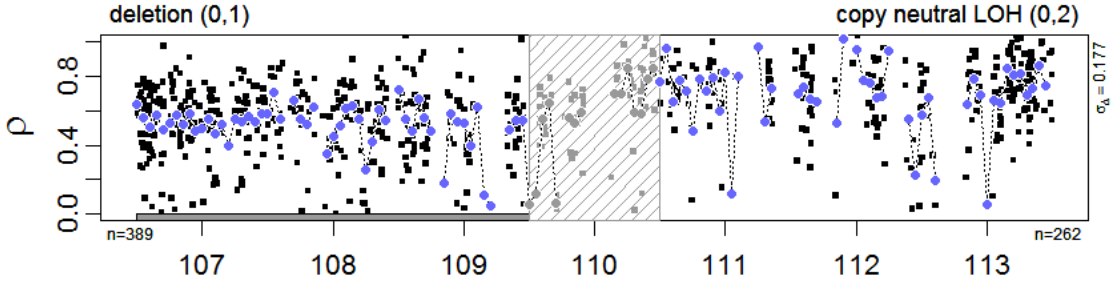


Figure 5.1: TCNs of the different methods for region TCGA-23-1027:Chr10@106.5-113.5,cp=110+/-0.5,s=2/3. The sets of loci to the left and the right belong to two different copy-number states, separated by a change points. The exact location of the change point is unknown, except that it is located within safety margin (shaded with diagonal lines). One of the copy-number states is referred to as the "negative" state ('deletion (0,1)') and the other as the "positive" state ('copy neutral LOH (0,2)'), where the latter are highlighted with a solid bar beneath. Legend: Full resolution TCNs are displayed in black and smoothed (bin width 50 kb) ones in blue. Loci within the safety margin are displayed in gray and excluded from all analysis.

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

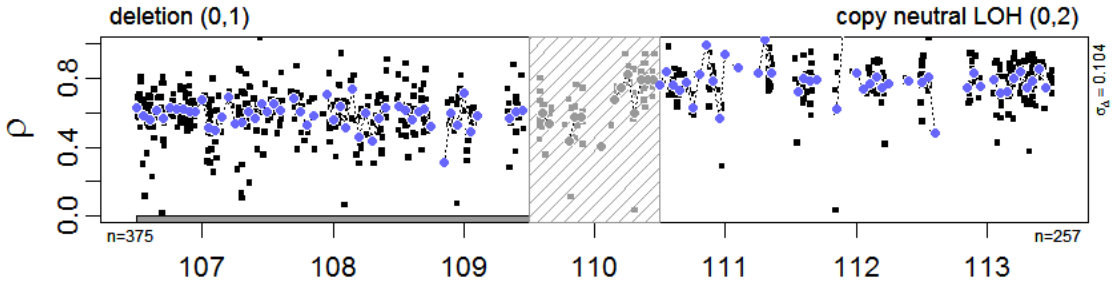


Figure 5.2: DHs of the different methods for region TCGA-23-1027:Chr10@106.5-113.5,cp=110+/-0.5,s=2/3. DHs are only defined for SNPs that are genotyped to be heterozygous based on data from the corresponding method. The sets of loci to the left and the right belong to two different copy-number states, separated by a change points. The exact location of the change point is unknown, except that it is located within safety margin (shaded with diagonal lines). One of the copy-number states is referred to as the "negative" state ('deletion (0,1)') and the other as the "positive" state ('copy neutral LOH (0,2)'), where the latter are highlighted with a solid bar beneath. Legend: Full resolution DHs are displayed in black and smoothed (bin width 50 kb) ones in blue. Loci within the safety margin are displayed in gray and excluded from all analysis.

5.2 TCN and BAF signal density plots

(a) deletion (0,1). (b) copy neutral LOH (0,2).

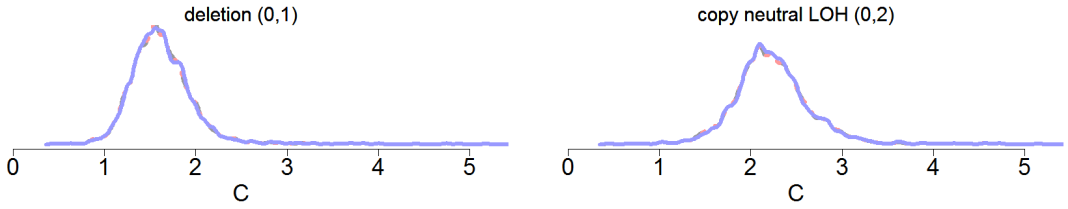


Figure 5.3: Densities of TCNs for region TCGA-23-1027:Chr10@106.5-113.5,cp=110+/-0.5,s=2/3. Legend: 'CRMAv2' (dashed gray), 'CRMAv2,TumorBoost' (dotted light red), and 'CRMAv2,CalMaTe' (solid light blue).

(a) deletion (0,1). (b) copy neutral LOH (0,2).

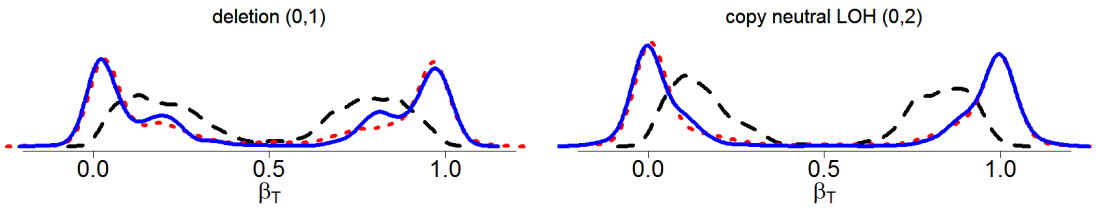
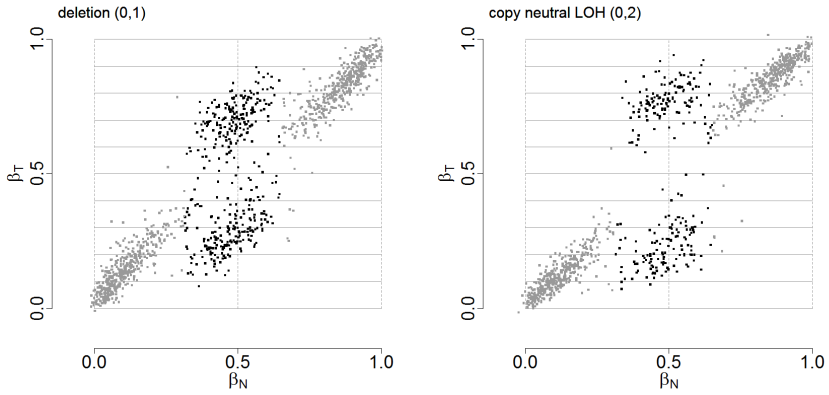


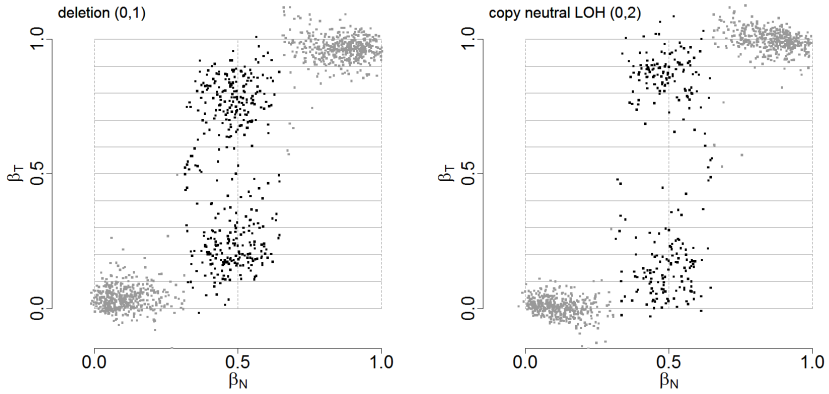
Figure 5.4: Densities of BAFs for region TCGA-23-1027:Chr10@106.5-113.5,cp=110+/-0.5,s=2/3. Legend: 'CRMAv2' (dashed black), 'CRMAv2,TumorBoost' (dotted red), and 'CRMAv2,CalMaTe' (solid blue).

5.3 Tumor versus normal BAF plots

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

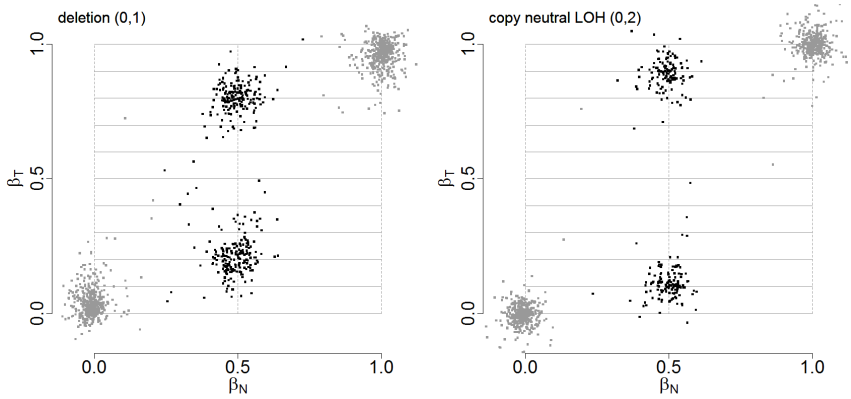
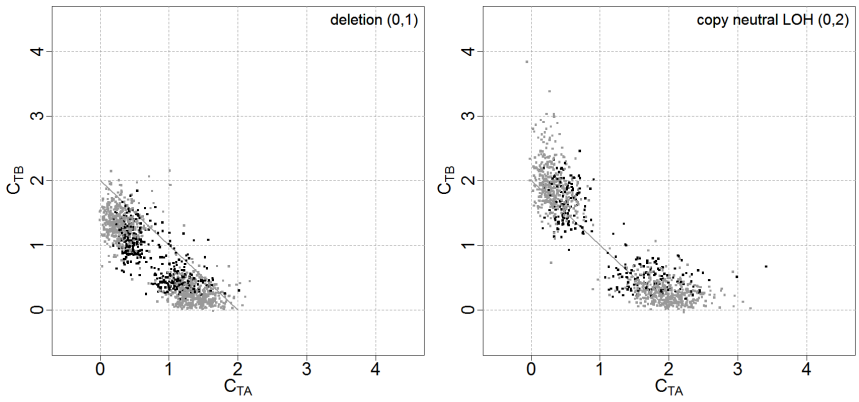


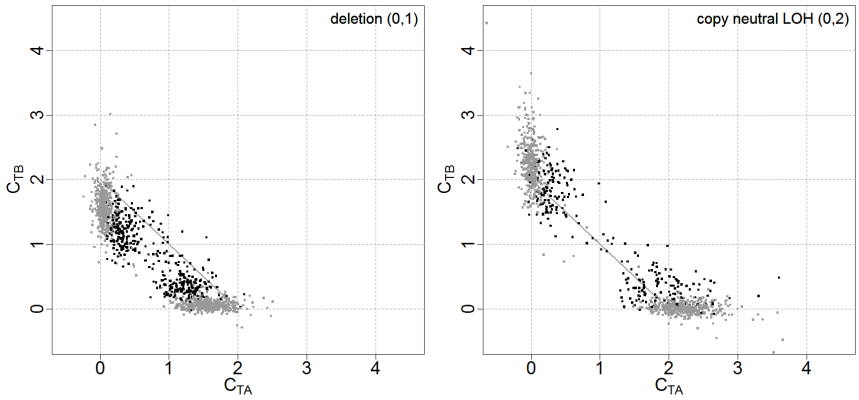
Figure 5.5: Tumor versus normal BAFs of 'CRMAv2', 'CRMAv2,TumorBoost', and 'CRMAv2,CalMaTe' for region TCGA-23-1027:Chr10@106.5-113.5,cp=110+/-0.5,s=2/3. Legend: SNPs genotyped (based on data from the corresponding method) as being heterozygous (homozygous) are displayed in black (gray).

5.4 ASCN plots

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

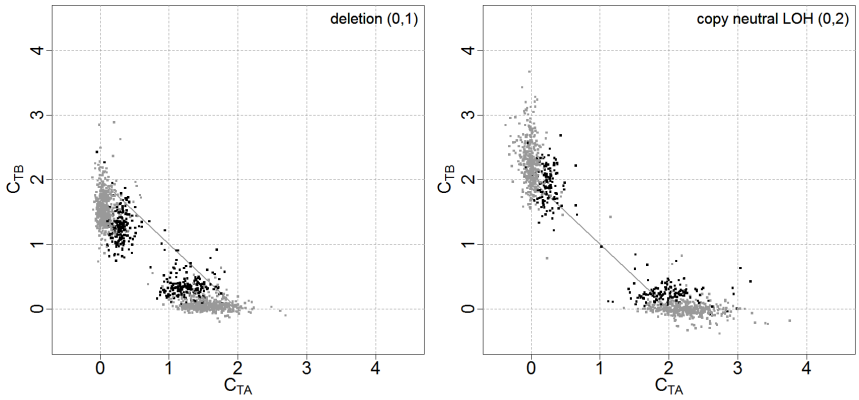
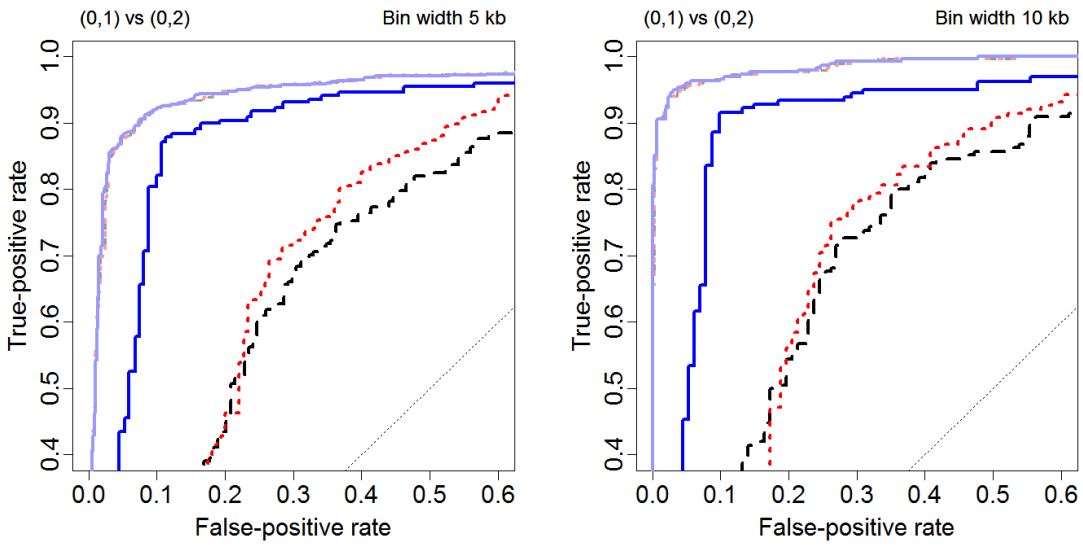


Figure 5.6: Tumor ASCNs (allele-specific copy numbers) of 'CRMAv2', 'CRMAv2,TumorBoost', and 'CRMAv2,CalMaTe' for region TCGA-23-1027:Chr10@106.5-113.5,cp=110+/-0.5,s=2/3. Legend: SNPs called heterozygous (homozygous) by the corresponing method are displayed in black (gray).

5.5 ROC curves

(a) Bin width 5 kb.

(b) Bin width 10 kb.



(c) Bin width 25 kb.

(d) Bin width 50 kb.

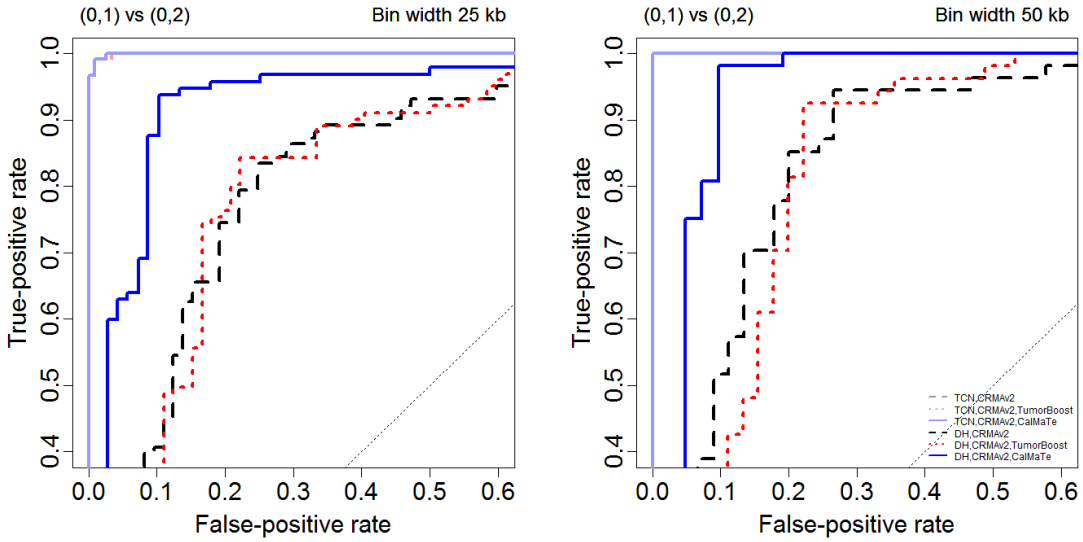
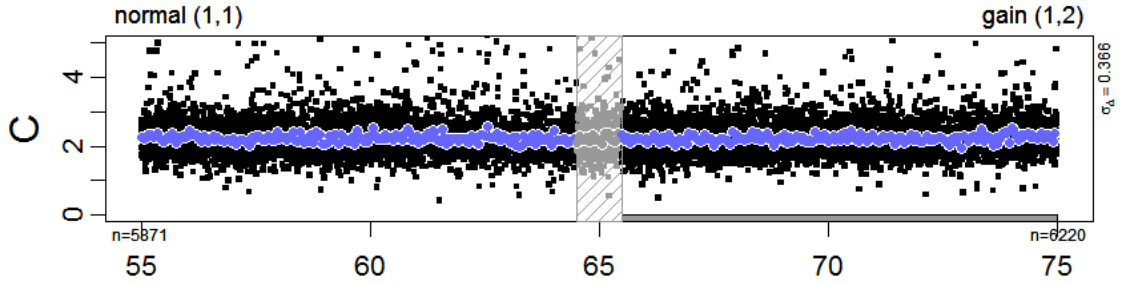


Figure 5.7: ROC curves for detecting the change point based on TCNs and DHs for each preprocessing method at different amounts of smoothing ("resolutions") for region TCGA-23-1027:Chr10@106.5-113.5,cp=110+/-0.5,s=2/3. Smoothing was done by averaging (using the mean() function) in non-overlapping bins. Legend: The TCN and DH curves for the same method are depicted with the same line type and color with the difference that the TCN curve use a lighter version of color for 'CRMAv2' (dashed black), 'CRMAv2,TumorBoost' (dotted red), and 'CRMAv2,CalMaTe' (solid blue).

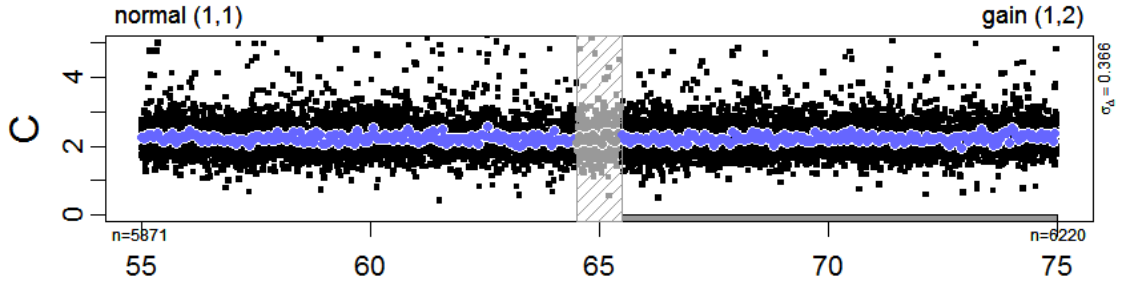
6 Region: TCGA-23-1027:Chr2@55-75.0

6.1 TCN and DH tracks

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

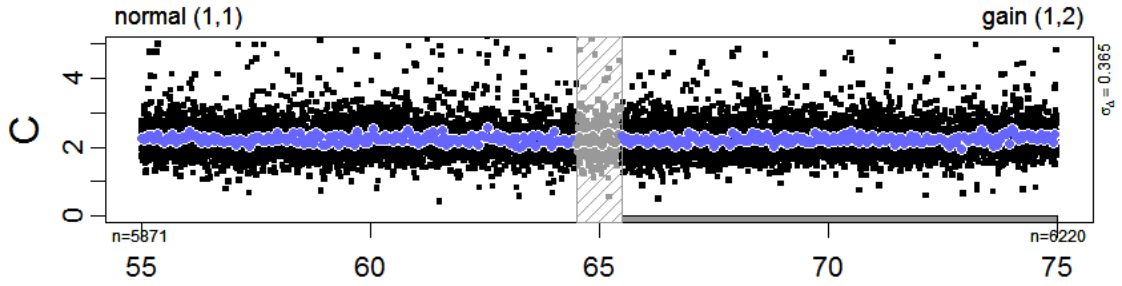
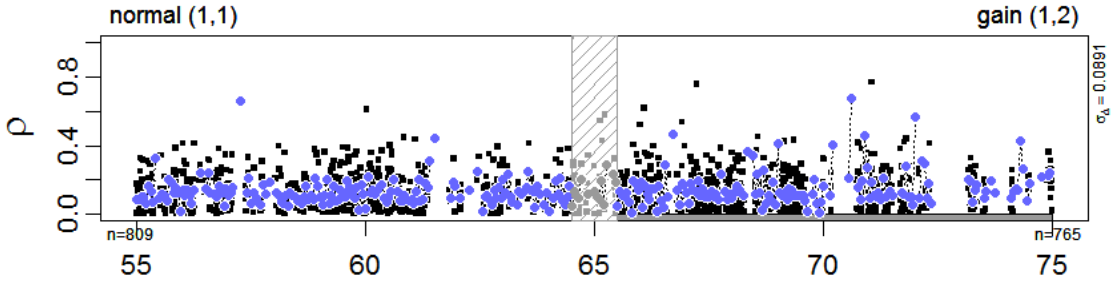
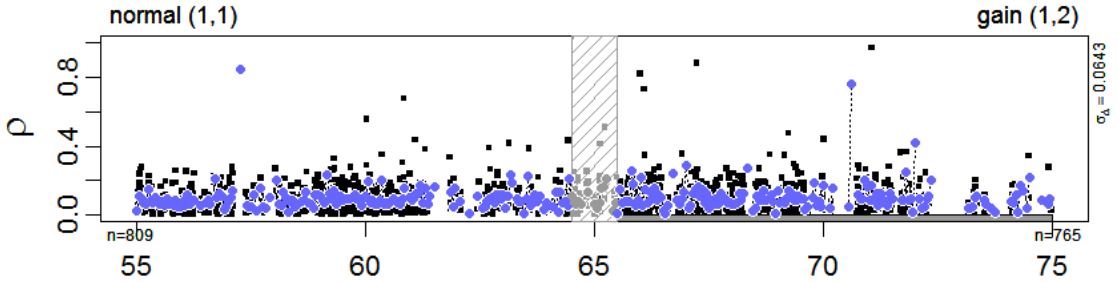


Figure 6.1: TCNs of the different methods for region TCGA-23-1027:Chr2@55-75.0,cp=65.0+/-0.5,s=0/1. The sets of loci to the left and the right belong to two different copy-number states, separated by a change points. The exact location of the change point is unknown, except that it is located within safety margin (shaded with diagonal lines). One of the copy-number states is referred to as the "negative" state ('normal (1,1)') and the other as the "positive" state ('gain (1,2)'), where the latter are highlighted with a solid bar beneath. Legend: Full resolution TCNs are displayed in black and smoothed (bin width 50 kb) ones in blue. Loci within the safety margin are displayed in gray and excluded from all analysis.

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

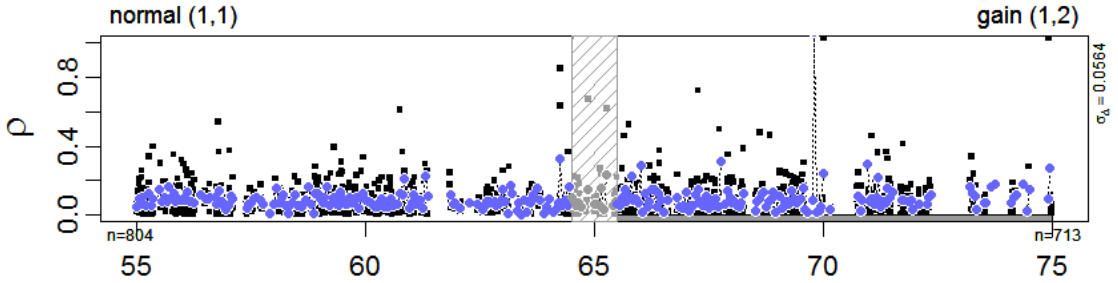


Figure 6.2: DHs of the different methods for region TCGA-23-1027:Chr2@55-75.0,cp=65.0+/-0.5,s=0/1. DHs are only defined for SNPs that are genotyped to be heterozygous based on data from the corresponding method. The sets of loci to the left and the right belong to two different copy-number states, separated by a change points. The exact location of the change point is unknown, except that it is located within safety margin (shaded with diagonal lines). One of the copy-number states is referred to as the "negative" state ('normal (1,1)') and the other as the "positive" state ('gain (1,2)'), where the latter are highlighted with a solid bar beneath. Legend: Full resolution DHs are displayed in black and smoothed (bin width 50 kb) ones in blue. Loci within the safety margin are displayed in gray and excluded from all analysis.

6.2 TCN and BAF signal density plots

(a) normal (1,1). (b) gain (1,2).

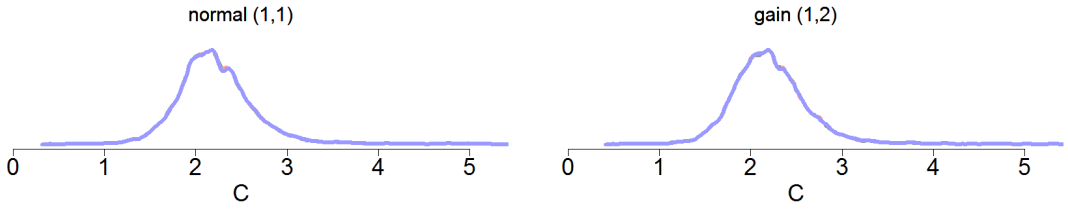


Figure 6.3: Densities of TCNs for region TCGA-23-1027:Chr2@55-75.0,cp=65.0+/-0.5,s=0/1. Legend: 'CRMAv2' (dashed gray), 'CRMAv2,TumorBoost' (dotted light red), and 'CRMAv2,CalMaTe' (solid light blue).

(a) normal (1,1). (b) gain (1,2).

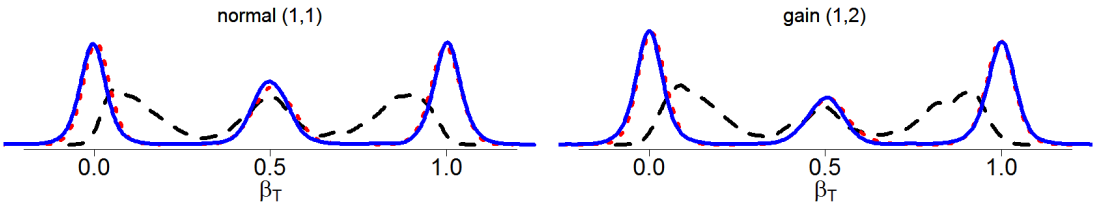
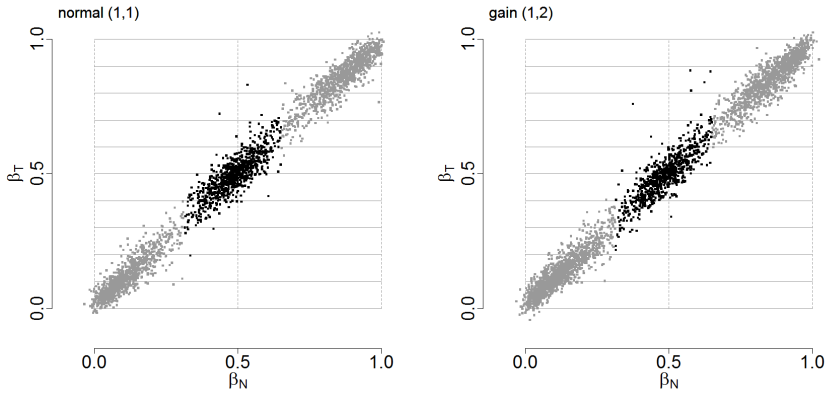


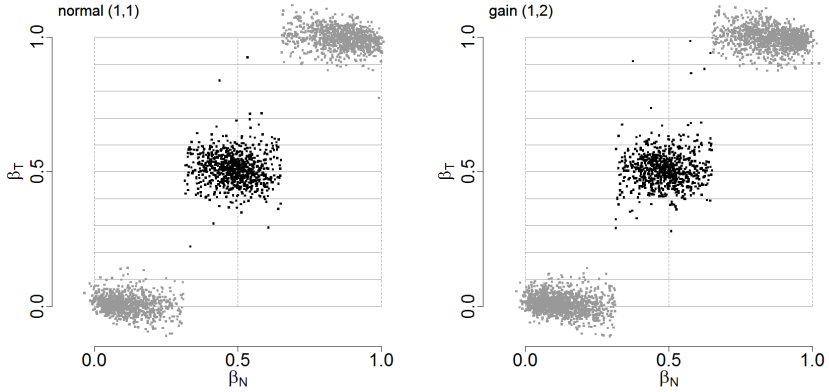
Figure 6.4: Densities of BAFs for region TCGA-23-1027:Chr2@55-75.0,cp=65.0+/-0.5,s=0/1. Legend: 'CRMAv2' (dashed black), 'CRMAv2,TumorBoost' (dotted red), and 'CRMAv2,CalMaTe' (solid blue).

6.3 Tumor versus normal BAF plots

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

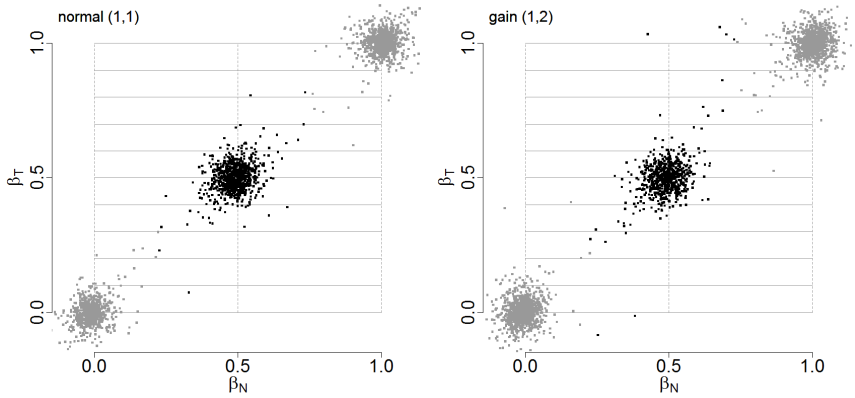
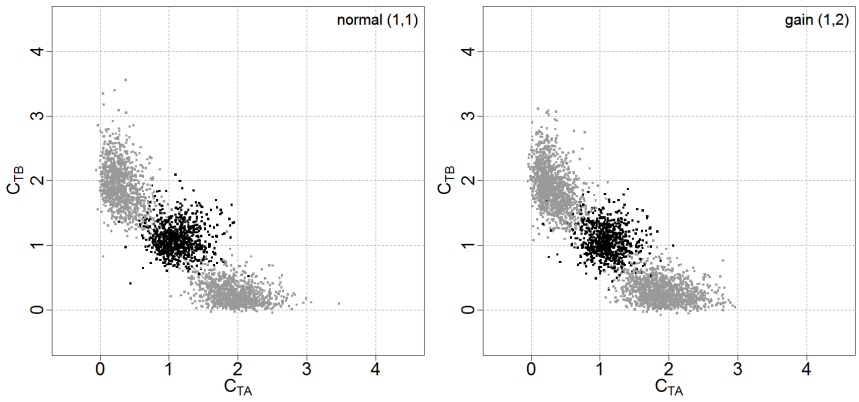


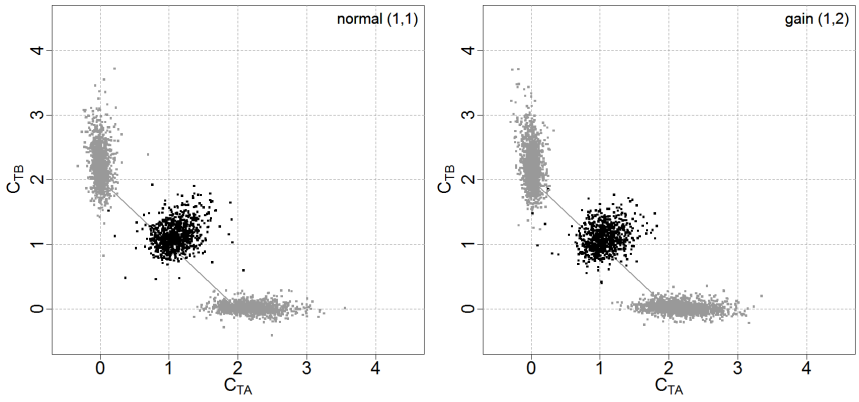
Figure 6.5: Tumor versus normal BAFs of 'CRMAv2', 'CRMAv2,TumorBoost', and 'CRMAv2,CalMaTe' for region TCGA-23-1027:Chr2@55-75.0,cp=65.0+/-0.5,s=0/1. Legend: SNPs genotyped (based on data from the corresponding method) as being heterozygous (homozygous) are displayed in black (gray).

6.4 ASCN plots

(a) CRMAv2.



(b) CRMAv2,TumorBoost.



(c) CRMAv2,CalMaTe.

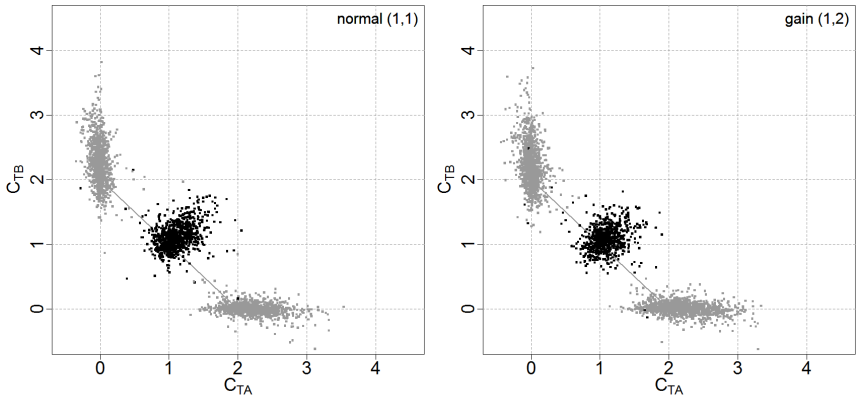
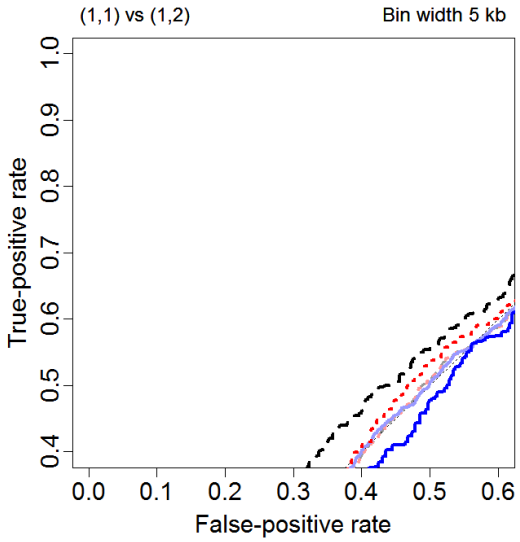


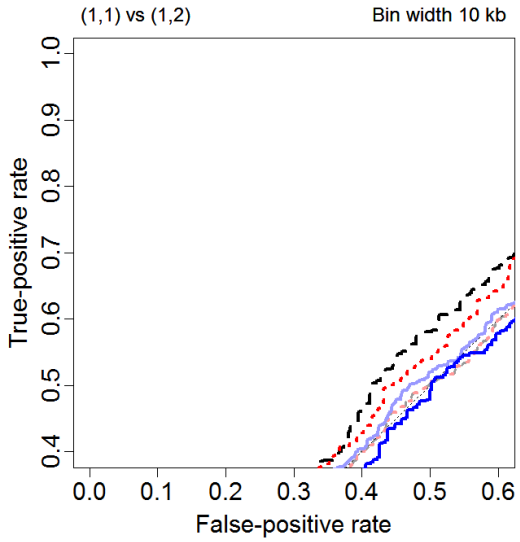
Figure 6.6: Tumor ASCNs (allele-specific copy numbers) of 'CRMAv2', 'CRMAv2,TumorBoost', and 'CRMAv2,CalMaTe' for region TCGA-23-1027:Chr2@55-75.0,cp=65.0+/-0.5,s=0/1. Legend: SNPs called heterozygous (homozygous) by the corresponing method are displayed in black (gray).

6.5 ROC curves

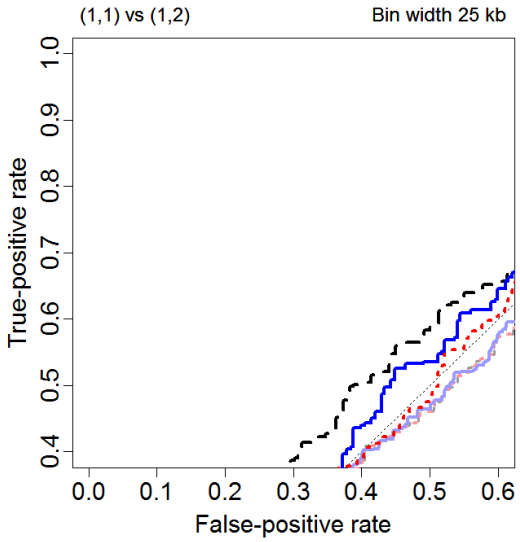
(a) Bin width 5 kb.



(b) Bin width 10 kb.



(c) Bin width 25 kb.



(d) Bin width 50 kb.

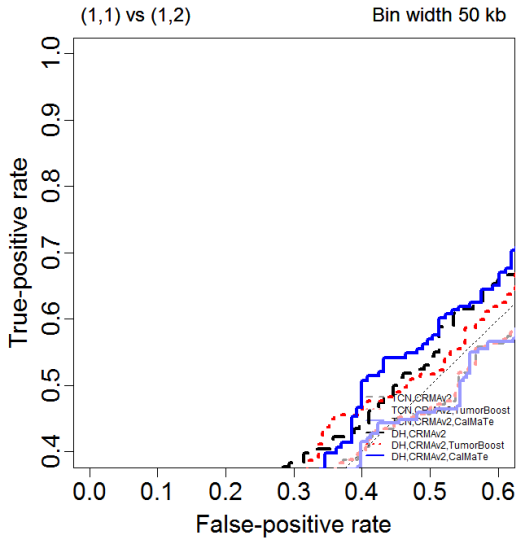


Figure 6.7: ROC curves for detecting the change point based on TCNs and DHs for each preprocessing method at different amounts of smoothing ("resolutions") for region TCGA-23-1027:Chr2@55-75.0,cp=65.0+/-0.5,s=0/1. Smoothing was done by averaging (using the mean() function) in non-overlapping bins. Legend: The TCN and DH curves for the same method are depicted with the same line type and color with the difference that the TCN curve use a lighter version of color for 'CRMAv2' (dashed black), 'CRMAv2,TumorBoost' (dotted red), and 'CRMAv2,CalMaTe' (solid blue).

Bibliography

Bengtsson, H., Wirapati, P., and Speed, T. P. (2009). A single-array preprocessing method for estimating full-resolution raw copy numbers from all Affymetrix genotyping arrays including GenomeWideSNP 5 & 6. *Bioinformatics*, **25**(17), 2149–2156.

Bengtsson, H., Neuvial, P., and Speed, T. P. (2010). TumorBoost: Normalization of allele-specific tumor copy numbers from a single pair of tumor-normal genotyping microarrays. *BMC Bioinformatics*, **11**(1), 245.

A Data

A.1 Paired tumor-normal data

$CRMAv2
PairedPSCNData:
Name: TCGA−23−1027
Number of rows: 1881415
Number of columns: 11
Columns: chromosome [numeric], x [numeric], CT [numeric], CN [numeric], betaT [numeric], betaN [numeric], isSNP [logical], muN [numeric], isHet [logical], C∗ [numeric], rho∗ [numeric]
RAM: 100.51MB
Chromosomes: 1−25 [25]
Number of loci: 1881415

$‘CRMAv2,TumorBoost‘
PairedPSCNData:
Name: TCGA−23−1027
Number of rows: 1881415
Number of columns: 11
Columns: chromosome [numeric], x [numeric], CT [numeric], CN [numeric], betaT [numeric], betaN [numeric], isSNP [logical], muN [numeric], isHet [logical], C∗ [numeric], rho∗ [numeric]
RAM: 114.86MB
Chromosomes: 1−25 [25]
Number of loci: 1881415

$‘CRMAv2,CalMaTe‘
PairedPSCNData:
Name: TCGA−23−1027
Number of rows: 1881415
Number of columns: 11
Columns: chromosome [numeric], x [numeric], CT [numeric], CN [numeric], betaT [numeric], betaN [numeric], isSNP [logical], muN [numeric], isHet [logical], C∗ [numeric], rho∗ [numeric]
RAM: 100.51MB
Chromosomes: 1−25 [25]
Number of loci: 1881415

B Session information

This report was automatically generated using the R.rsp package.

- R version 2.14.2 Patched (2012-02-29 r58590), `x86_64-pc-mingw32`
- Locale: `LC_COLLATE=English_United States.1252`, `LC_CTYPE=English_United States.1252`, `LC_MONETARY=English_United States.1252`, `LC_NUMERIC=C`, `LC_TIME=English_United States.1252`
- Base packages: base, datasets, graphics, grDevices, methods, stats, utils
- Other packages: aroma.cn.eval 0.3.1, aroma.core 2.4.13, aroma.light 1.23.1, digest 0.5.1, MASS 7.3-17, matrixStats 0.4.4, R.cache 0.6.1, R.filesets 1.1.5, R.menu 0.2.0, R.methodsS3 1.2.3, R.oo 1.9.2, R.rsp 0.7.4, R.utils 1.12.1, RColorBrewer 1.0-5, xtable 1.7-0
- Loaded via a namespace (and not attached): affxparser 1.27.4, tools 2.14.2