

# Schmid Lab ChIP Microarray Pipeline

October 4, 2013

## **Contents**

### **List of Figures**

# 1 Initialization

This document contains the output of the Schmid lab's ChIP-chip microarray pipeline, an R/Bioconductor based platform for ChIP microarray prepossessing. This pipeline is adapted from the generic microarray version written by Nicholas Gillum (which was originally adapted from CARMAweb) and modified for ChIP-chip data. For questions regarding this program, email peter.tonner@duke.edu.

```
> config <- yaml.load(  
+ "experiment: 0258_none_densityLoess_max100_final  
+  
+ multicore: yes  
+  
+ slides:  
+   slide01:  
+     name: 0258_noH202  
+     file: 252681910034_S02_ChIP-v1_95_May07_1_1.txt  
+     IP: cy5  
+     condition: noH202  
+   slide02:  
+     name: 0258_noH202_swap  
+     file: 252681910034_S02_ChIP-v1_95_May07_1_2.txt  
+     IP: cy3  
+     condition: noH202  
+   slide03:  
+     name: 0258_+H202_10min  
+     file: 252681910035_S01_ChIP-v1_95_May07_1_1.txt  
+     IP: cy5  
+     condition: wH202_10m  
+   slide04:  
+     name: 0258_+H202_10min_swap  
+     file: 252681910035_S01_ChIP-v1_95_May07_1_2.txt  
+     IP: cy3  
+     condition: wH202_10m  
+   slide05:  
+     name: 0258_+H202_20min  
+     file: 252681910036_S01_ChIP-v1_95_May07_1_1.txt  
+     IP: cy5  
+     condition: wH202_20m  
+   slide06:  
+     name: 0258_+H202_20min_swap  
+     file: 252681910036_S01_ChIP-v1_95_May07_1_2.txt  
+     IP: cy3  
+     condition: wH202_20m  
+   slide07:  
+     name: 0258_+H202_60min  
+     file: 252681910037_S01_ChIP-v1_95_May07_1_1.txt  
+     IP: cy5  
+     condition: wH202_60m  
+   slide08:  
+     name: 0258_+H202_60min_swap  
+     file: 252681910037_S01_ChIP-v1_95_May07_1_2.txt  
+     IP: cy3  
+     condition: wH202_60m  
+   slide09:
```

```

+
+   name: 0258_22
+   file: US22502698_252681910022_S01_ChIP-v1_95_May07_1_2.txt
+   IP: cy5
+   condition: 0258_22
+
slide10:
+
+   name: 0258_22_swap
+   file: US22502698_252681910022_S01_ChIP-v1_95_May07_1_1.txt
+   IP: cy3
+   condition: 0258_22
+
slide11:
+
+   name: 0258_32
+   file: US22502698_252681910032_S01_ChIP-v1_95_May07_1_1.txt
+   IP: cy5
+   condition: 0258_32
+
slide12:
+
+   name: 0258_32_swap
+   file: US22502698_252681910032_S01_ChIP-v1_95_May07_1_2.txt
+   IP: cy3
+   condition: 0258_32
+
slide13:
+
+   name: 0258_33
+   file: US22502698_252681910033_S01_ChIP-v1_95_May07_1_1.txt
+   IP: cy5
+   condition: 0258_33
+
slide14:
+
+   name: 0258_33_swap
+   file: US22502698_252681910033_S01_ChIP-v1_95_May07_1_2.txt
+   IP: cy3
+   condition: 0258_33
+
slide17:
+
+   name: 0258_11
+   file: US22502698_252681910011_S01_ChIP-v1_95_May07_1_2.txt
+   IP: cy5
+   condition: 0258_11
+
slide18:
+
+   name: 0258_11_swap
+   file: US22502698_252681910011_S01_ChIP-v1_95_May07_1_1.txt
+   IP: cy3
+   condition: 0258_11
+
slide19:
+
+   name: 0258_21
+   file: US22502698_252681910021_S01_ChIP-v1_95_May07_1_1.txt
+   IP: cy3
+   condition: 0258_21
+
slide20:
+
+   name: 0258_21_swap
+   file: US22502698_252681910021_S01_ChIP-v1_95_May07_1_2.txt
+   IP: cy5
+   condition: 0258_21
+
slide21:
+
+   name: 0258_13
+   file: US22502698_252681910013_S01_ChIP-v1_95_May07_1_2.txt
+   IP: cy5

```

```

+   condition: 0258_13
+ slide22:
+   name: 0258_13_swap
+   file: US22502698_252681910013_S01_ChIP-v1_95_May07_1_1.txt
+   IP: cy3
+   condition: 0258_13
+
+
+ algorithms:
+   bgCorrection: none
+   bgOffset: 50
+   normexp.method: mle
+   withinArray: density.loess
+   betweenArray: quantile
+   median: yes
+   loess_cutoff: 0.4
+   kernel_size: 300
+
+ medichi:
+   max.steps: 100
+   fit.res: 10
+   n.boot: 10
+   boot.sample.opt: residual
+   kernel: kernel.halo.hires
+   coords.file: ../data/Chr_coords.txt
+
+ files:
+   library: ./lib
+   metadata: ./metadata autogenerated.txt
+   rawDir: ./raw
+   tmpDir: ./analysis
+   outDir: ./output
+
+ quiet:
+   backgroundCorrection: no
+   withinArrayNormalization: no
+
+ dixon:
+   cutoff: 0.01
+
+ bgText:
+   main: >
+     The following code draws a plot of the raw data for each microarray.
+     Specifically, it draws an MA plot of the regulation values (M values,
+     differential expression) of all the genes against their average
+     expression (A). The red and green lines in the plot represent the
+     mean and median, respectively. The turquoise line in the plot
+     represents the lowess fit line.
+   verbose: >
+     The following MA plots show the data after background correction.
+   quiet: >
+     These diagnostic plots have been suppressed. Please edit the
+     config.yaml file and set quiet:backgroundCorrection to no to

```

```

+      re-enable them.
+
+ wanText:
+   vsn: The vsn algorithm does not require within array normalization.
+   other: >
+     In this section, we normalize within arrays using the xxx algorithm.
+   verbose: >
+     The following MA plots show the data after with array normalization.
+   quiet: >
+     These diagnostic plots have been suppressed. Please edit the config.yaml
+     file and set quiet:withArrayNormalization to no to re-enable them.
+ ")
> source(paste(config$files$library,"/pipeline_utils.R",sep=""))
> source(paste(config$files$library,"/latex_utils.R",sep=""))
> source(paste(config$files$library,"/chipchip.normalize.R",sep=""))
> source(paste(config$files$library,"/medichi.utils.R",sep=""))

Loading MeDiChI, version 0.4.0 (Tue Oct 12 11:20:57 2010)
MeDiChI utils. (c) David J Reiss, ISB.
Please email dreiss@systemsbiology.org for questions or comments.
Loading function 'get.strongest.hits'
Loading function 'get.data'
Loading function 'compare.chip.hits'
Loading function 'get.genes.hit'
Loading function 'chip.hits.to.sif'
Loading function 'plot.significant.peaks'
Loading function 'fraction.in.coding.rgns'
Loading function 'medichi.clone.files'
Loading function 'medichi.to.genome.browser' (thanks, Tie!)

> config$files$metadata = paste(config$experiment,"targets.txt",sep="_")
> writeMetadata(config$slides, config$files$metadata)
> if(config$multicore){
+   library(multicore)
+   library(foreach)
+   library(doMC); registerDoMC()
+   library(MeDiChIMod)
+ } else{
+   library(MeDiChI)
+ }

Loading MeDiChI, version 0.4.0 (Tue Oct 12 11:20:57 2010)

> config$files$expDir = paste(config$files$outDir,"/",config$experiment,sep="")
> #make output directory if needed
> if (!file.exists(config$files$expDir)){
+   cat(paste("Making directory", config$files$expDir, " to save output.\n"))
+   dir.create(config$files$expDir)
+ }

Making directory ./output/0258_none_densityLoess_max100_final to save output.

This is the analysis for the 0258.none.densityLoess.max100.final experiments.

```

### **0258\_none\_densityLoess\_max100\_final Raw Data**

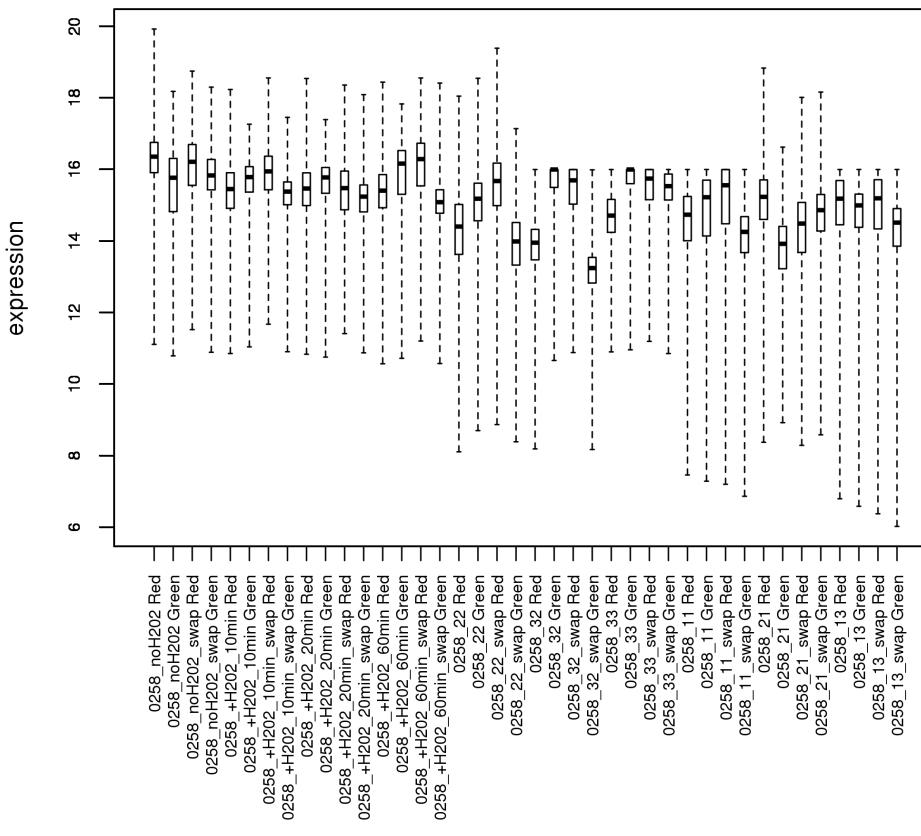


Figure 1: Boxplots of the signal intensities of each signal channel of the microarrays. Raw data before any correction or normalization.

## 2 Background correction

The following code draws a plot of the raw data for each microarray. Specifically, it draws an MA plot of the regulation values (M values, differential expression) of all the genes against their average expression (A). The red and green lines in the plot represent the mean and median, respectively. The turquoise line in the plot represents the lowess fit line.

```
> Slides.raw.bg <- backgroundCorrect(Slides.raw, method=config$algorithms$bgCorrection, offset = config$  
> #No signal less than 0 possible  
> Slides.raw.bg$R[Slides.raw.bg$R<1e-9] = 1e-9  
> Slides.raw.bg$G[Slides.raw.bg$G<1e-9] = 1e-9  
>
```

**0258\_none\_densityLoess\_max100\_final background corrected data**

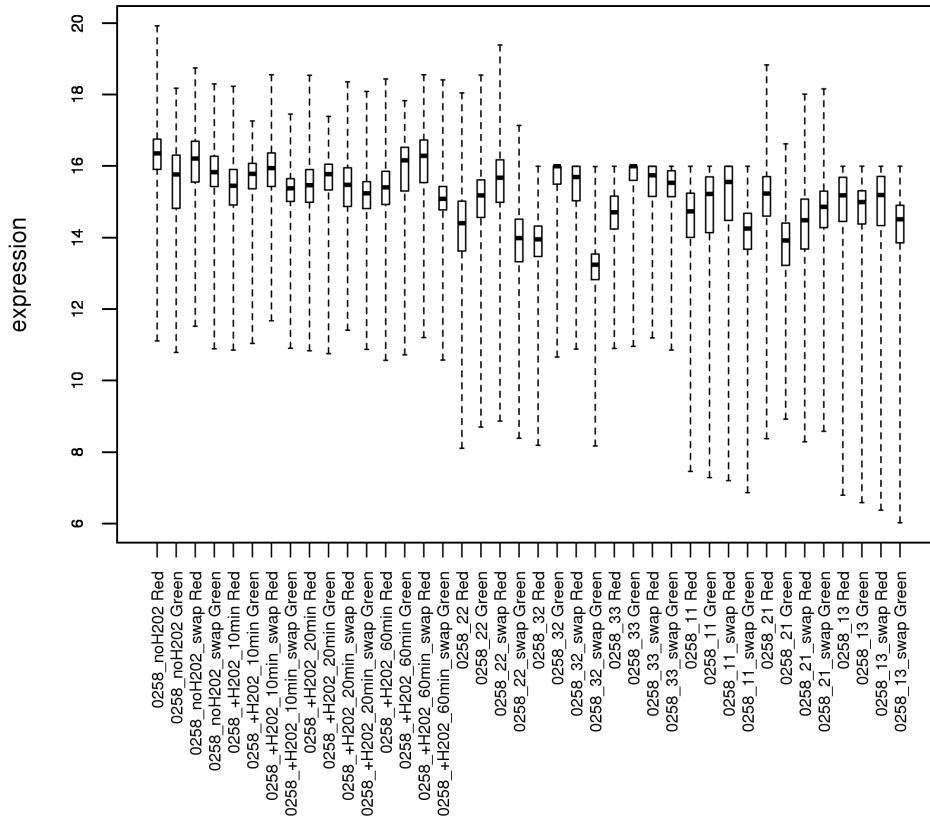


Figure 2: Boxplots of the signal intensities of each signal channel of the microarrays. Raw data after background correction.

### 3 Within array normalization

In this section, we normalize within arrays using the density.loess algorithm.

**0258\_none\_densityLoess\_max100\_final within array normalized data**

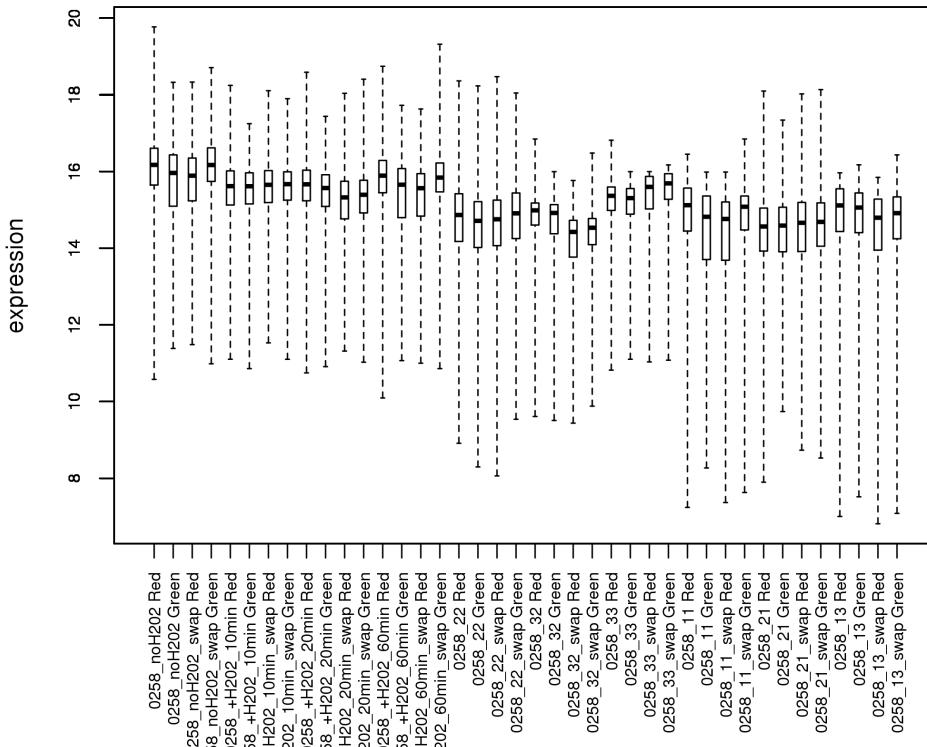
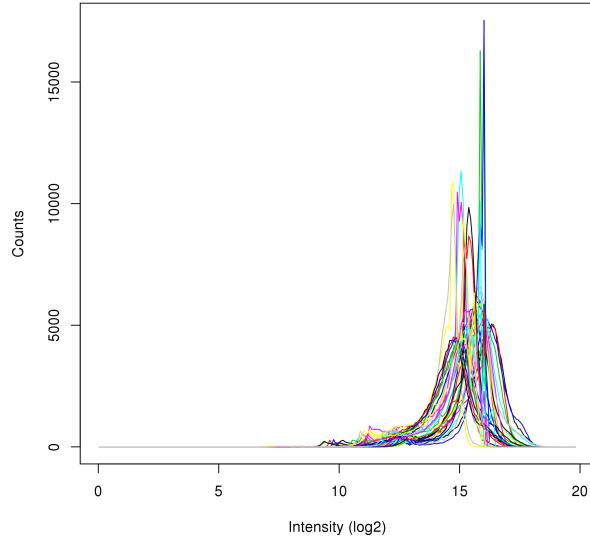


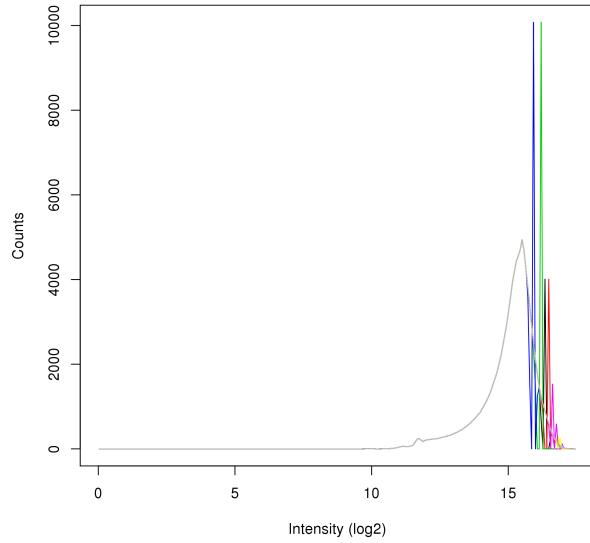
Figure 3: Boxplots of the signal intensities of each signal channel of the microarrays after within array normalization.

## 4 Between array normalization

In this section, first draw a diagnostic histogram of the within array normalized data. In the case of the vsn algorithm, it draws a histogram of the raw data. We then normalize between arrays using the quantile algorithm and draw a set of diagnostic MA plots.



(a) Before between array normalization

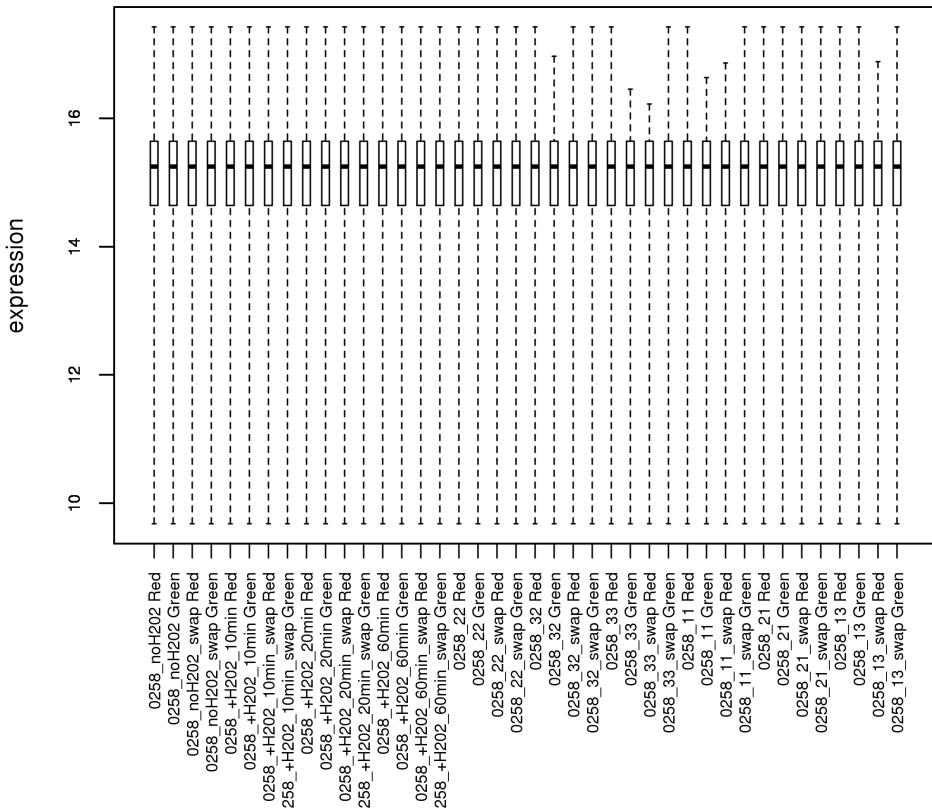


(b) After between array normalization

Figure 4: Histogram of all arrays within this experiment before and after the between array normalization

The quality of the within array normalization process can also be assessed using a boxplot.

### **0258\_none\_densityLoess\_max100\_final between array normalized data**



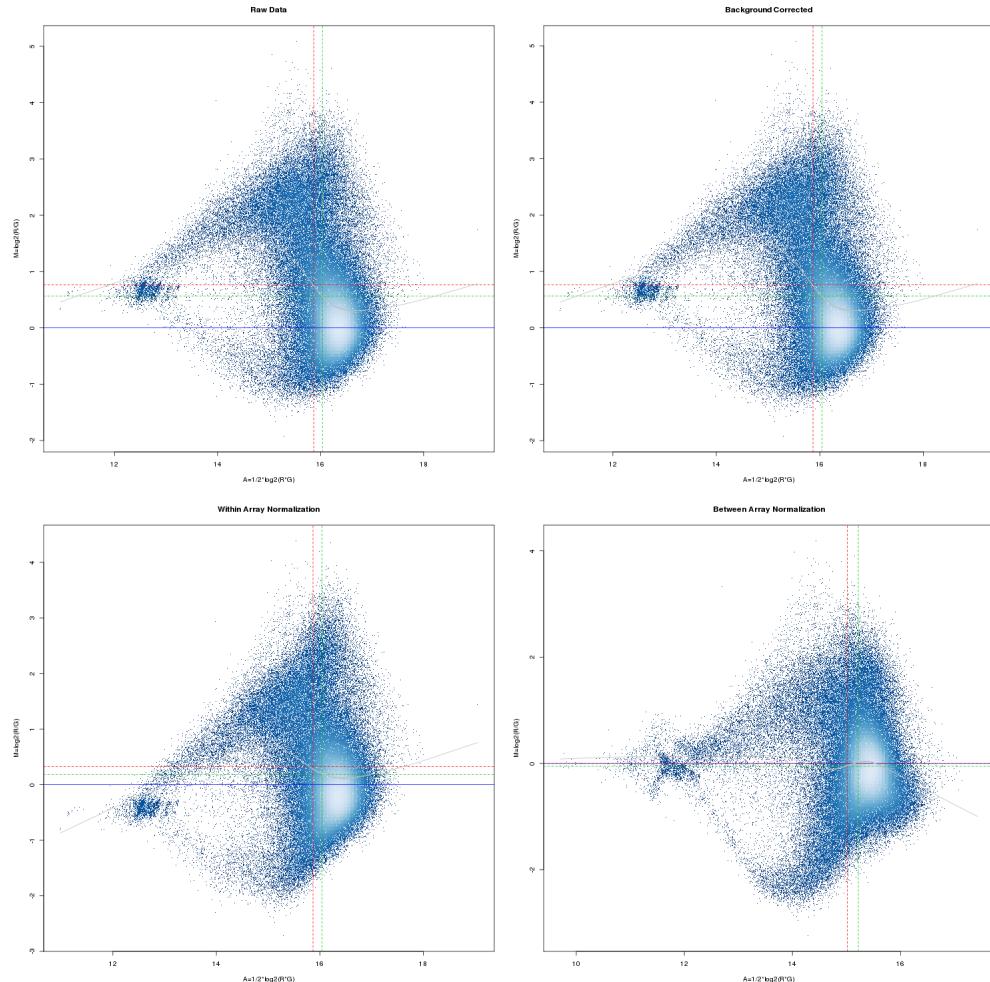
## 5 Save data

Normalized data saved to ./output.

## 6 Experiment Results

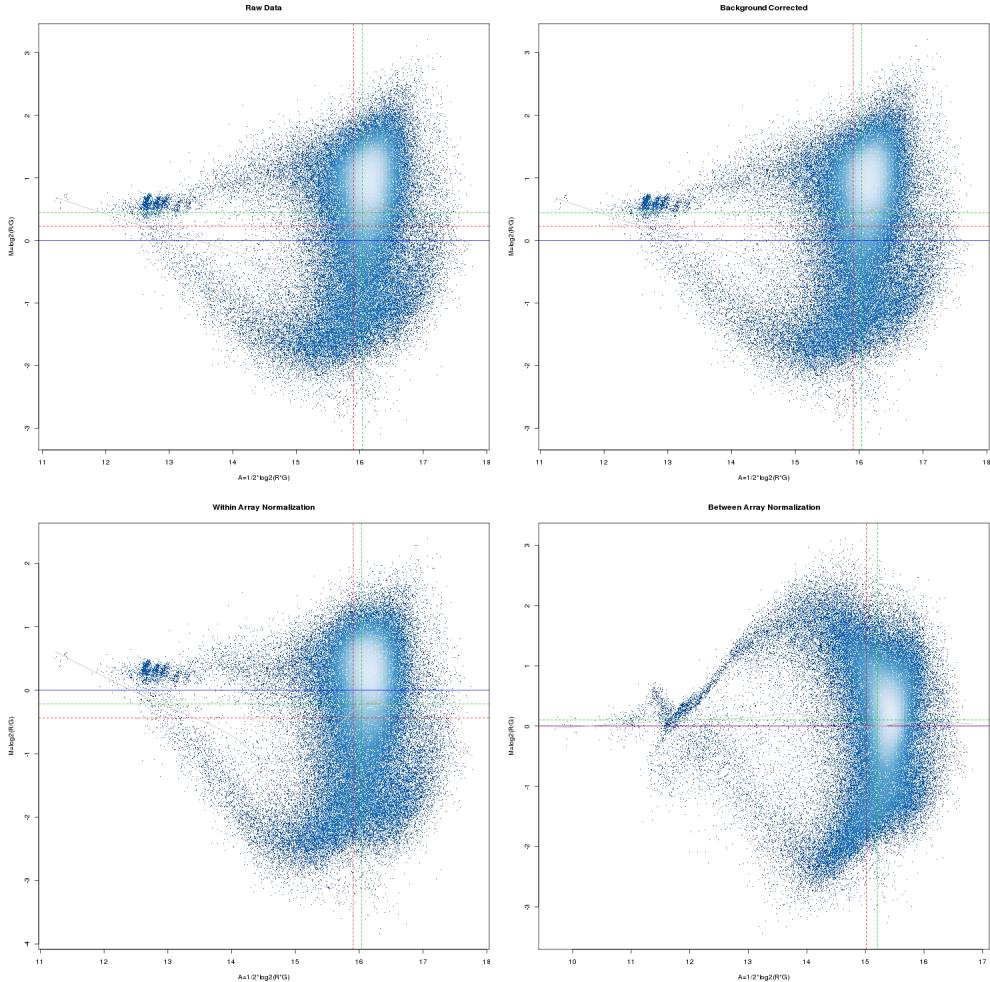
The results for each individual experiment will now be shown; raw data, background correction, within array normalization, and between array normalization.

### 6.1 0258\_noH2O2 Results



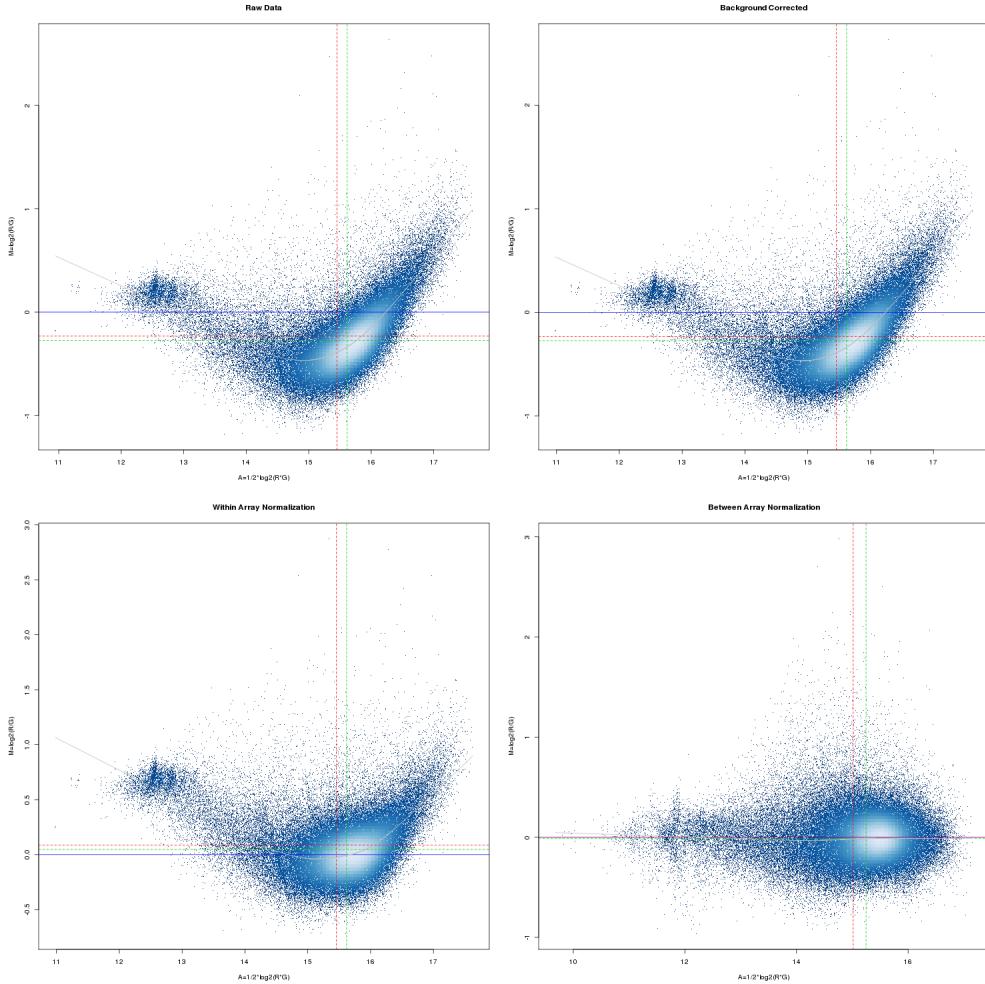
Results of all processing steps; background correction within array normalization, and between array normalization.

## 6.2 0258\_noH2O2\_swap Results



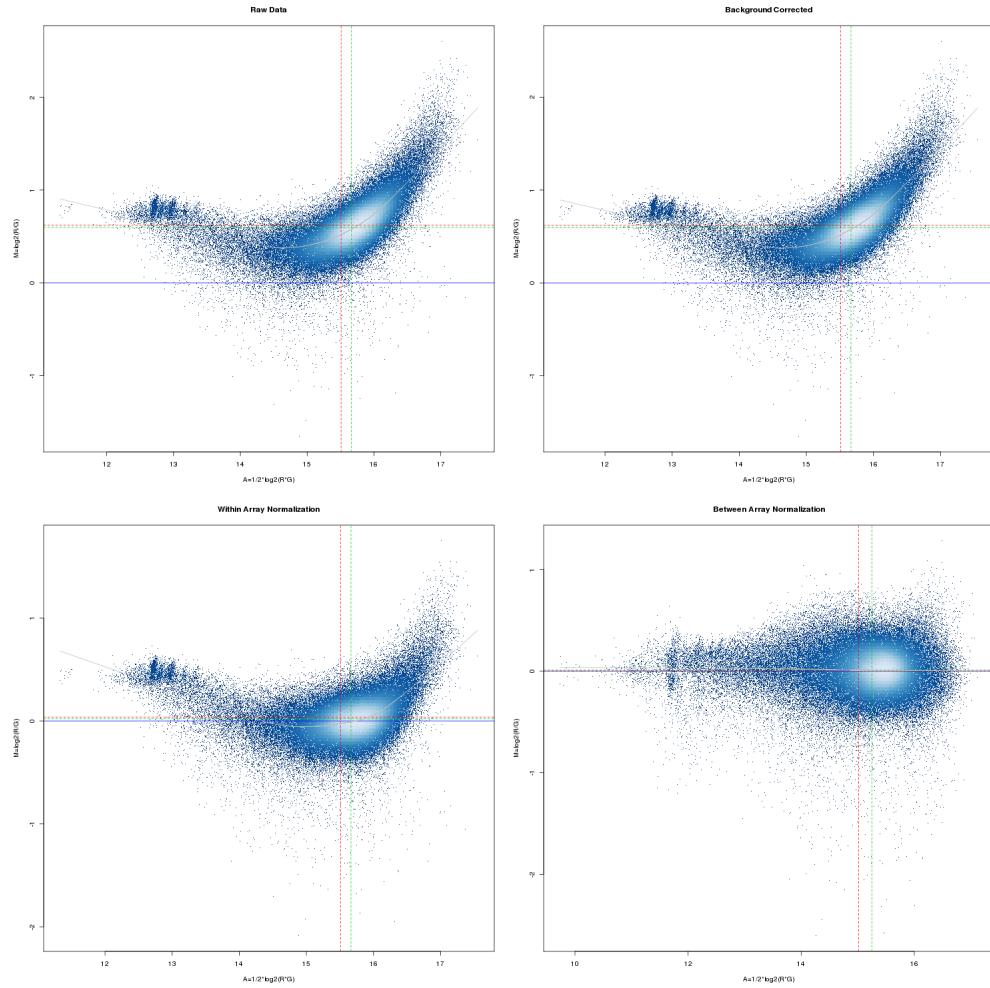
Results of all processing steps; background correction within array normalization, and between array normalization.

### 6.3 0258\_+H2O2\_10min Results



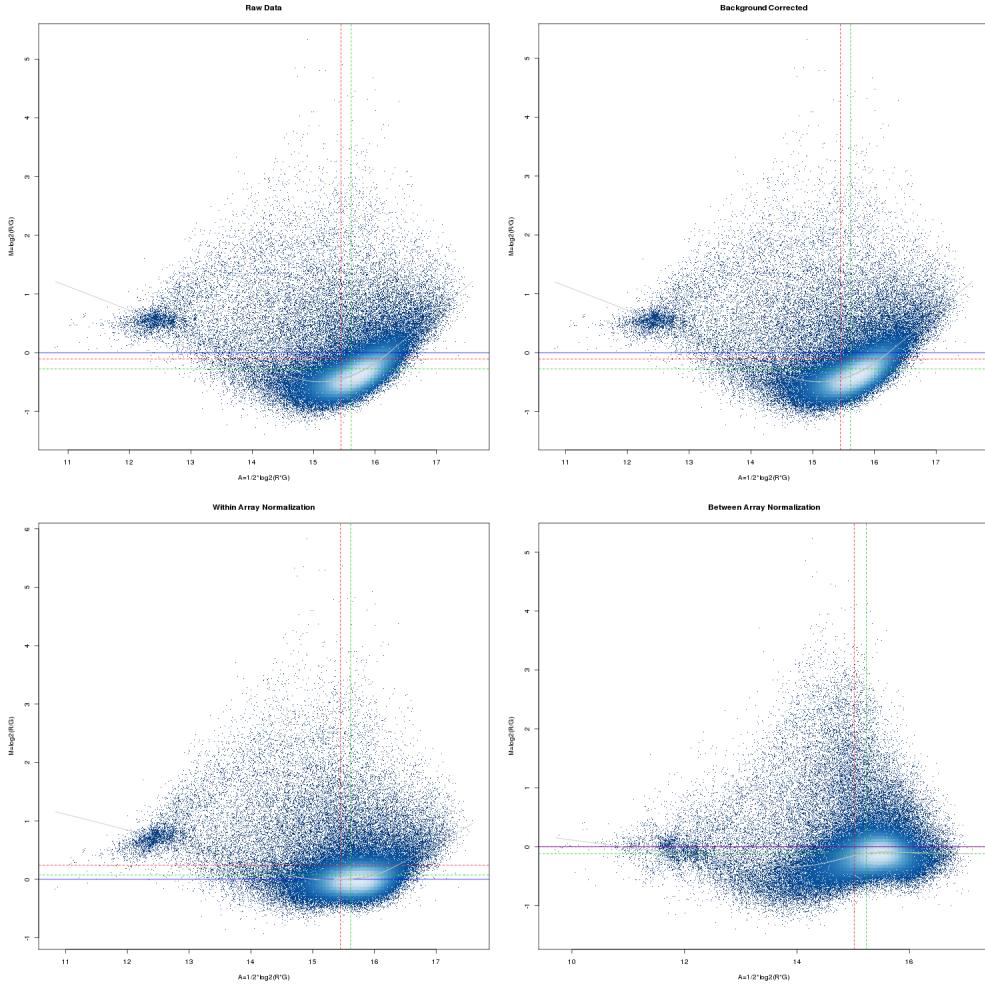
Results of all processing steps; background correction within array normalization, and between array normalization.

## 6.4 0258\_+H2O2\_10min\_swap Results



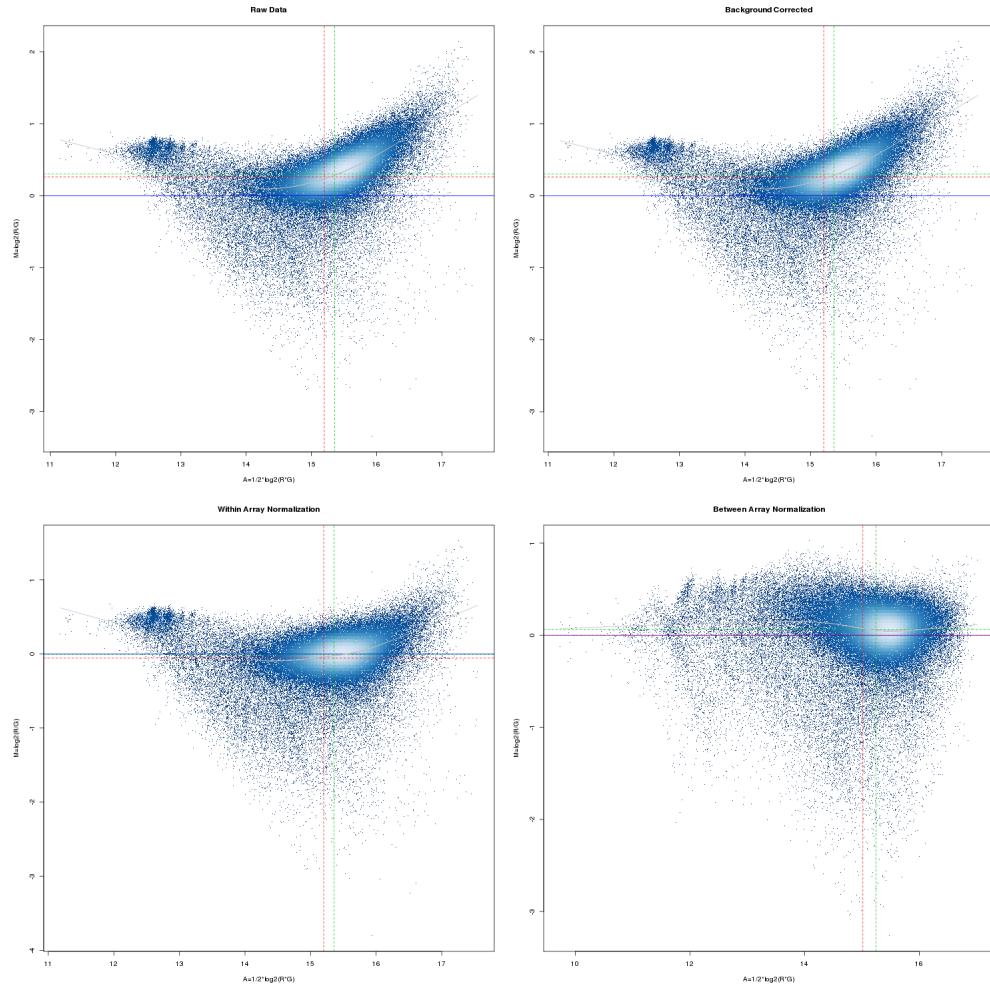
Results of all processing steps; background correction within array normalization, and between array normalization.

## 6.5 0258\_+H2O2\_20min Results



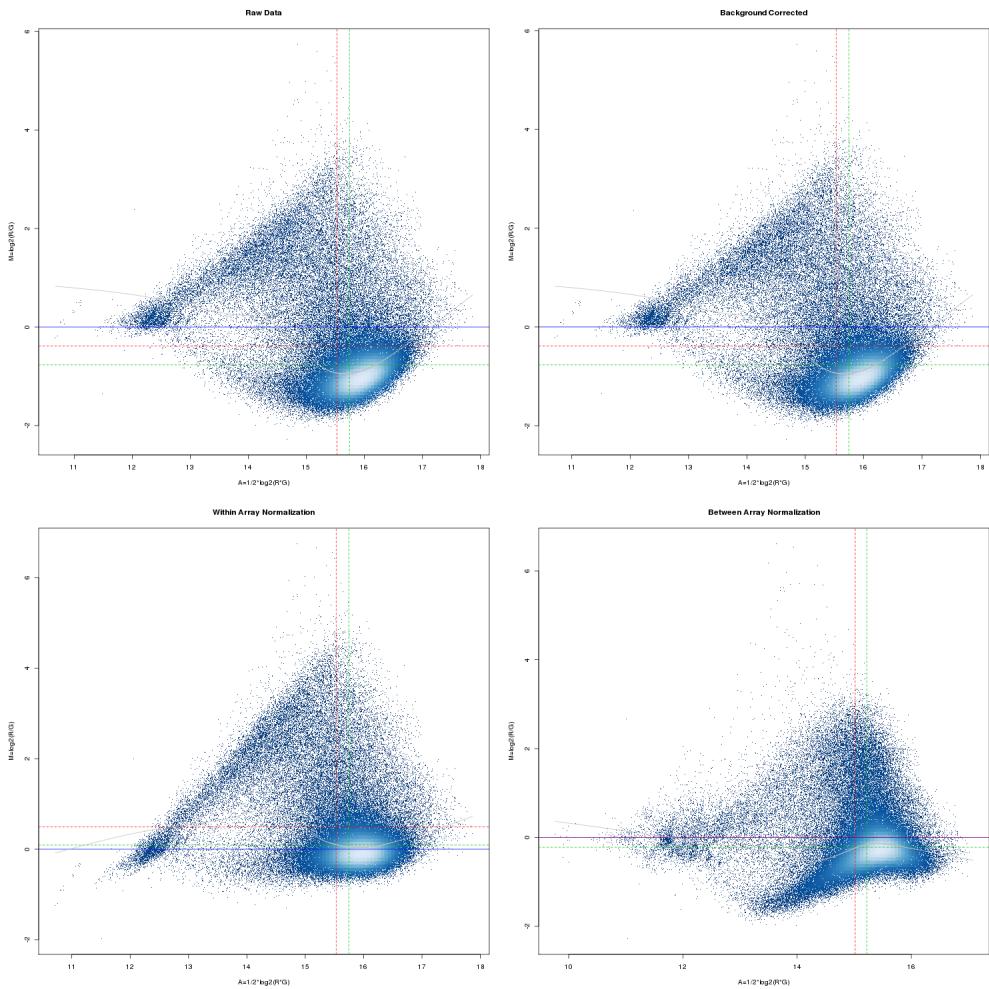
Results of all processing steps; background correction within array normalization, and between array normalization.

## 6.6 0258\_+H2O2\_20min\_swap Results



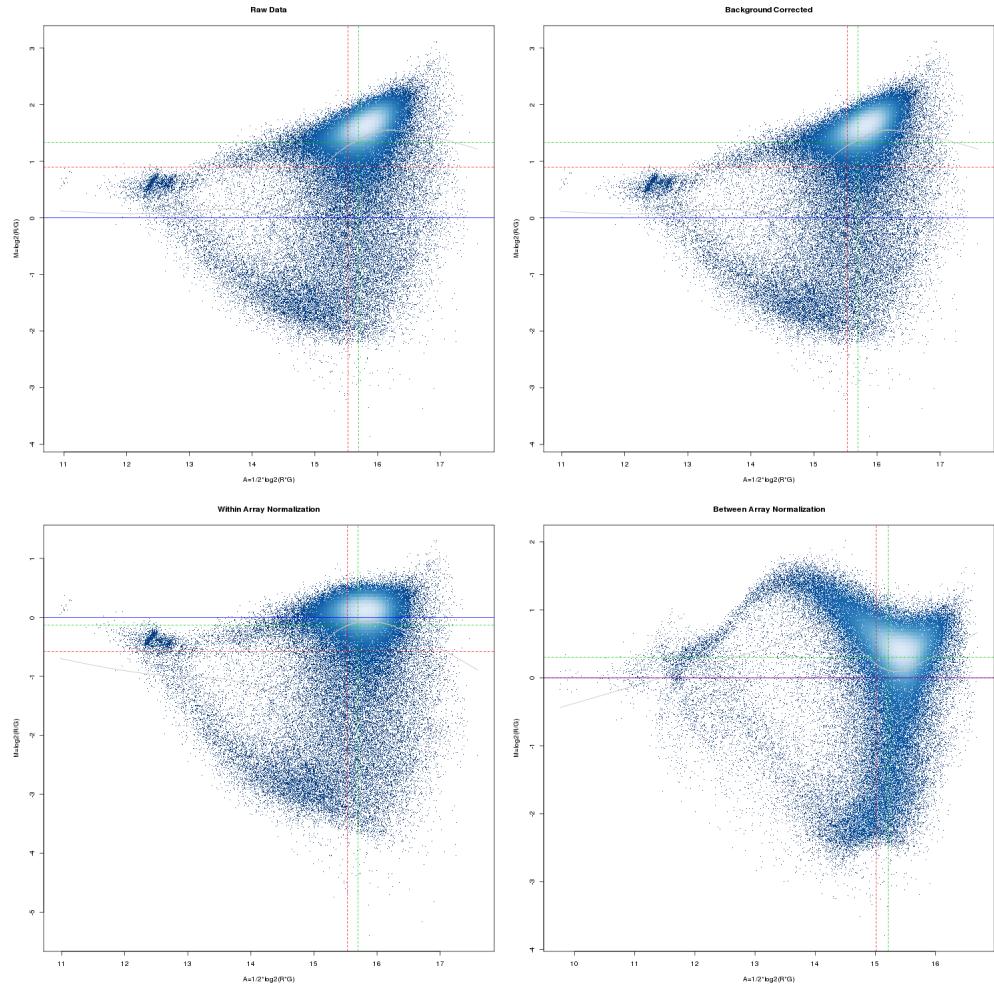
Results of all processing steps; background correction within array normalization, and between array normalization.

## 6.7 0258\_+H2O2\_60min Results



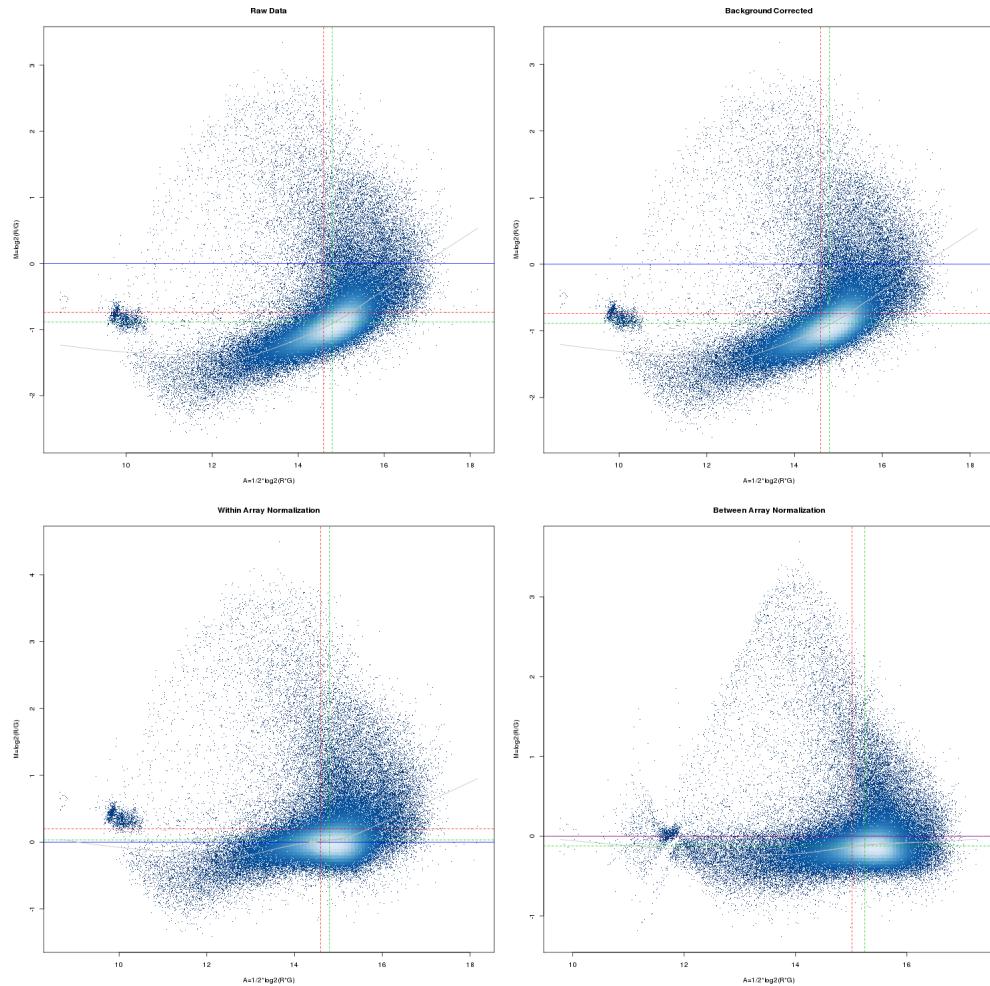
Results of all processing steps; background correction within array normalization, and between array normalization.

## 6.8 0258\_+H2O2\_60min\_swap Results



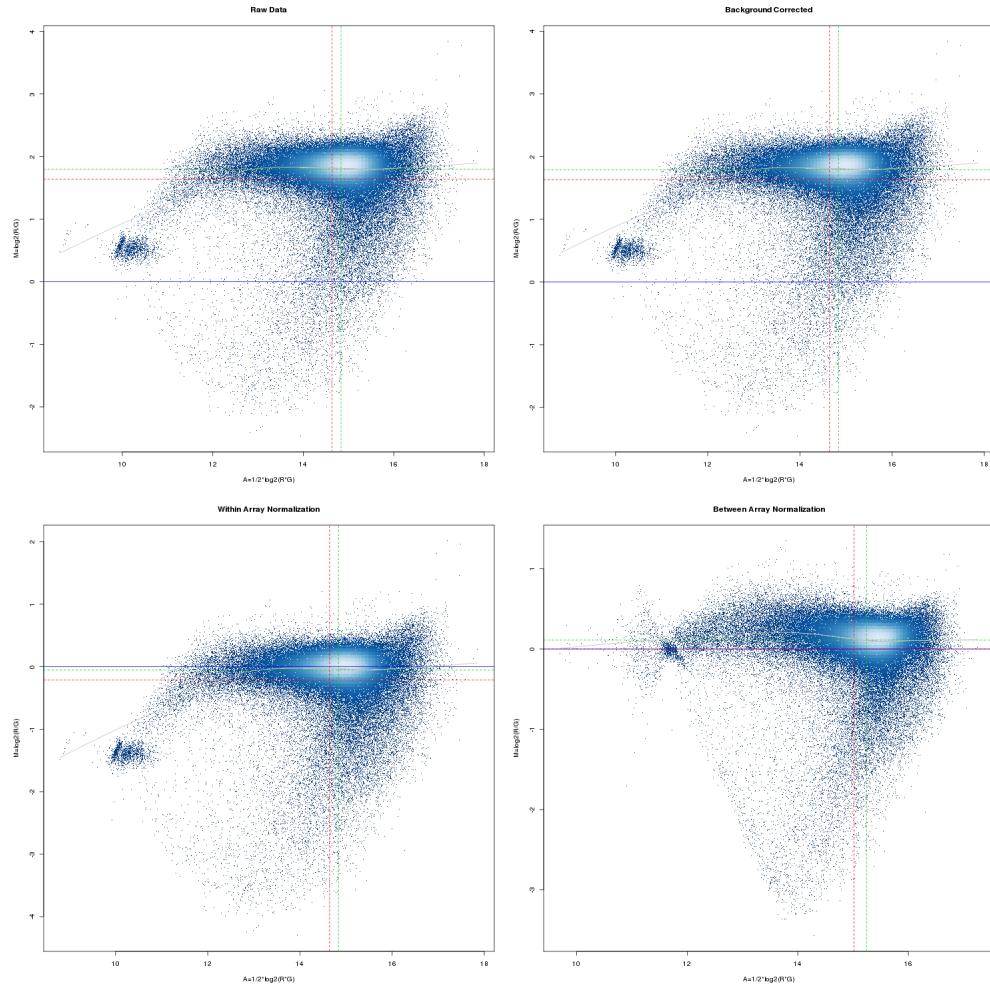
Results of all processing steps; background correction within array normalization, and between array normalization.

## 6.9 0258\_22 Results



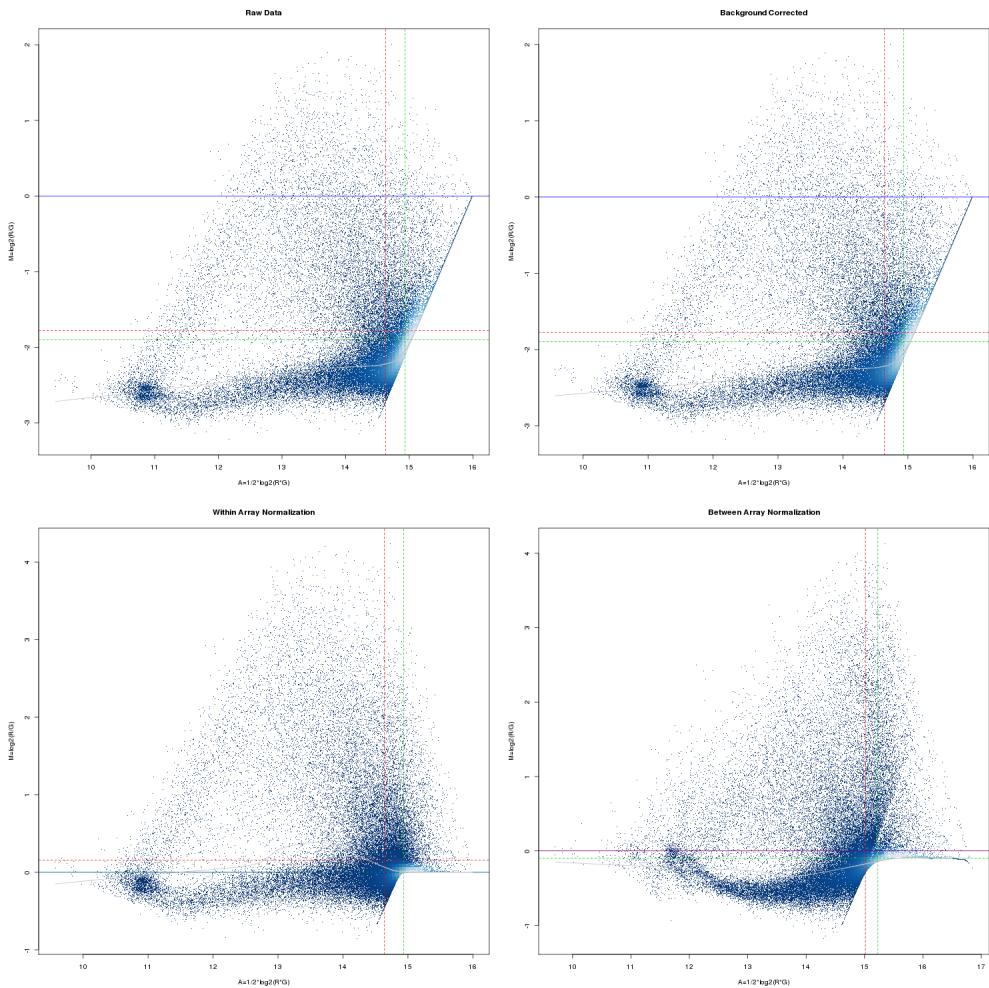
Results of all processing steps; background correction within array normalization, and between array normalization.

## 6.10 0258\_22\_swap Results



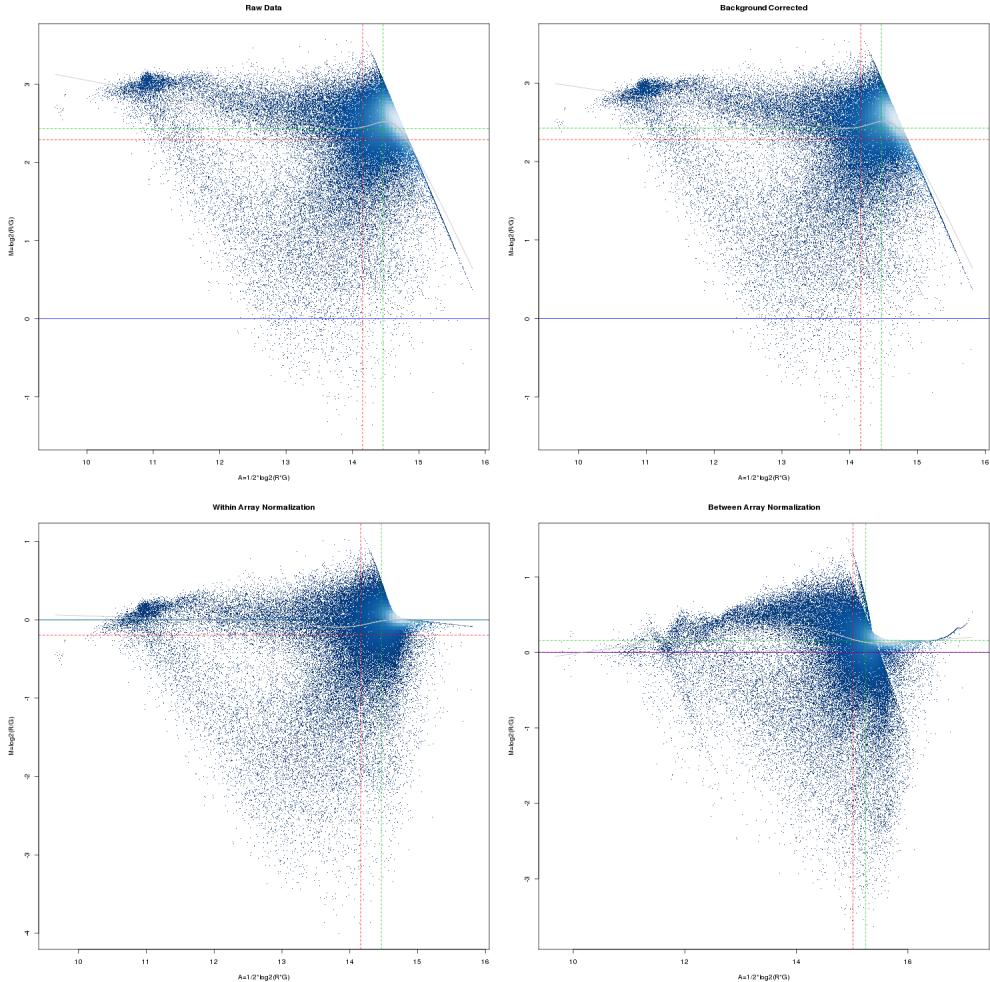
Results of all processing steps; background correction within array normalization, and between array normalization.

## 6.11 0258\_32 Results



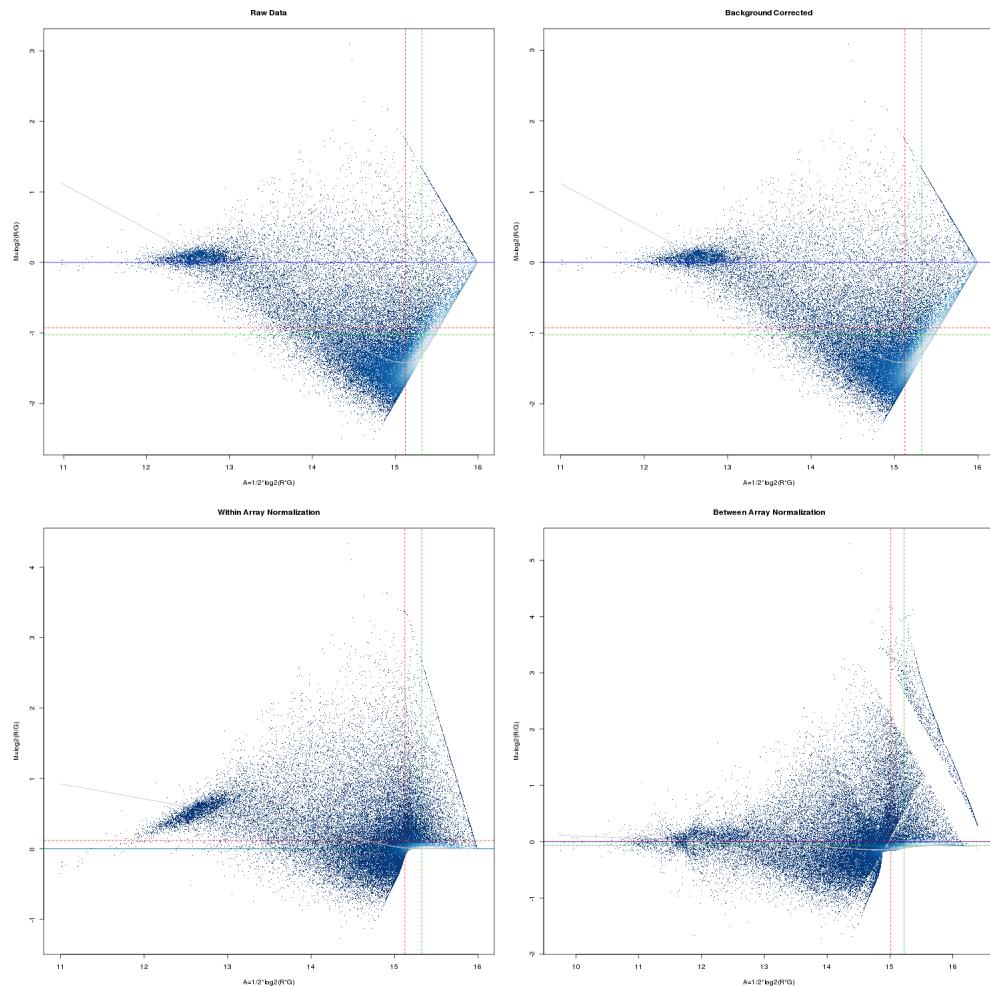
Results of all processing steps; background correction within array normalization, and between array normalization.

## 6.12 0258\_32\_swap Results



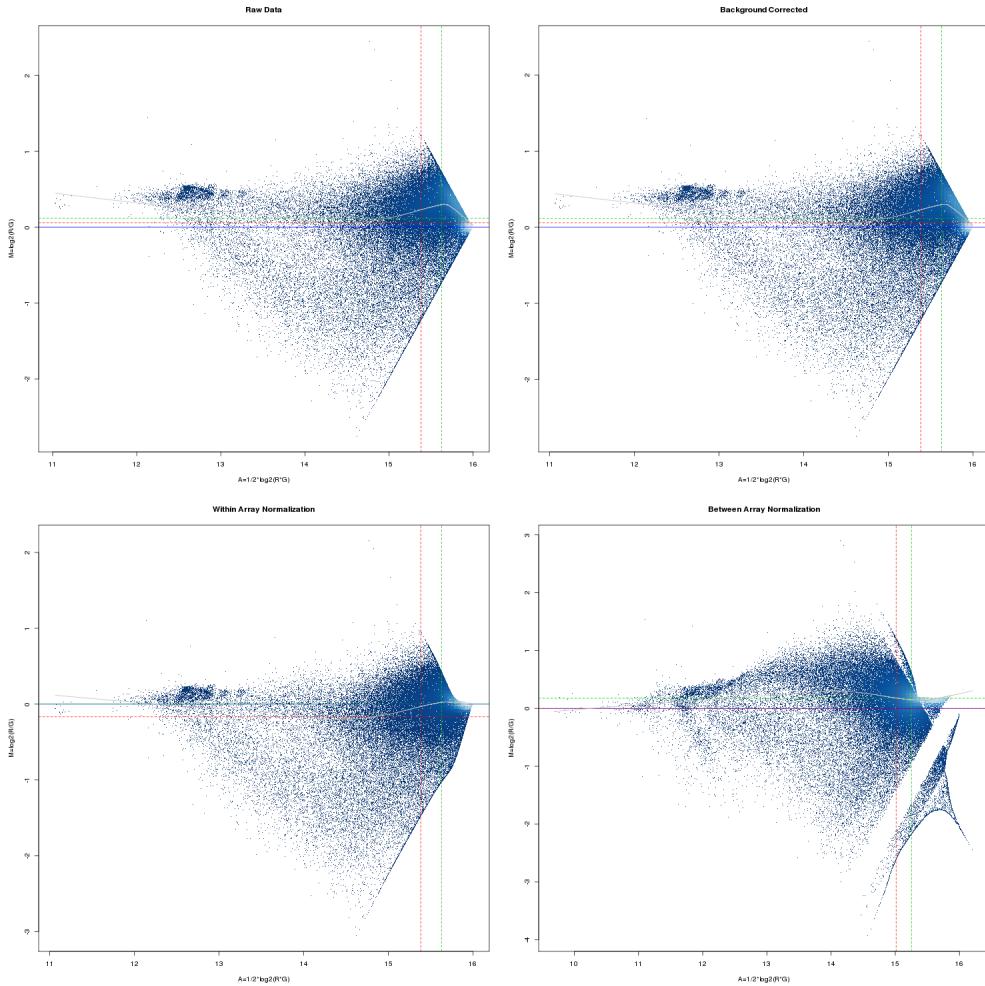
Results of all processing steps; background correction within array normalization, and between array normalization.

## 6.13 0258\_33 Results



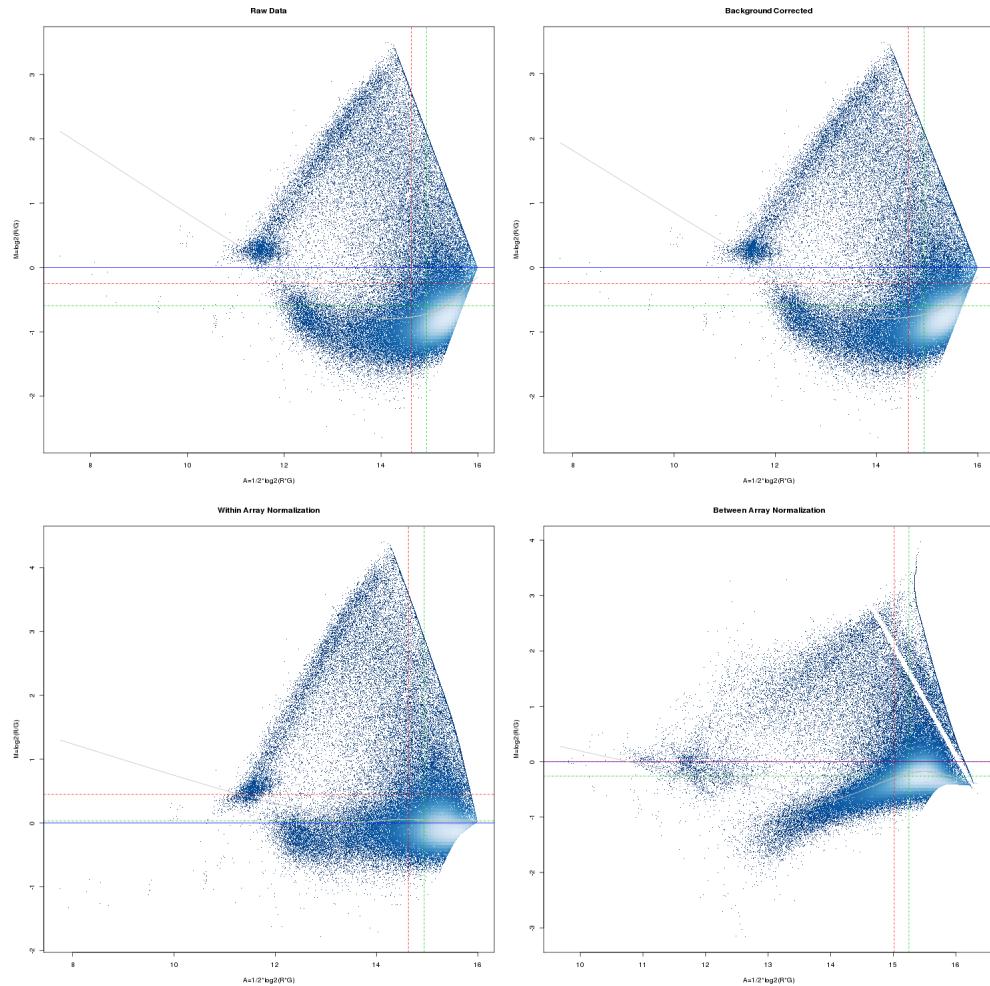
Results of all processing steps; background correction within array normalization, and between array normalization.

## 6.14 0258\_33\_swap Results



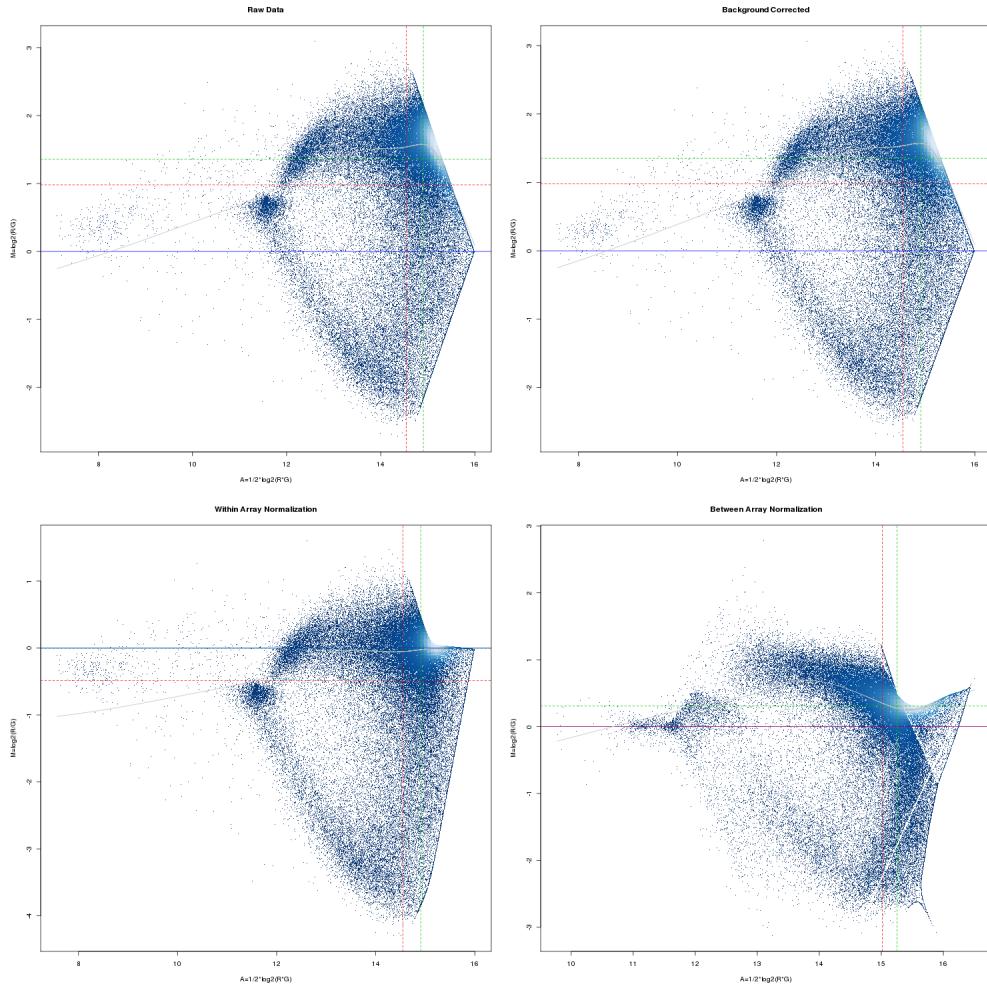
Results of all processing steps; background correction within array normalization, and between array normalization.

## 6.15 0258\_11 Results



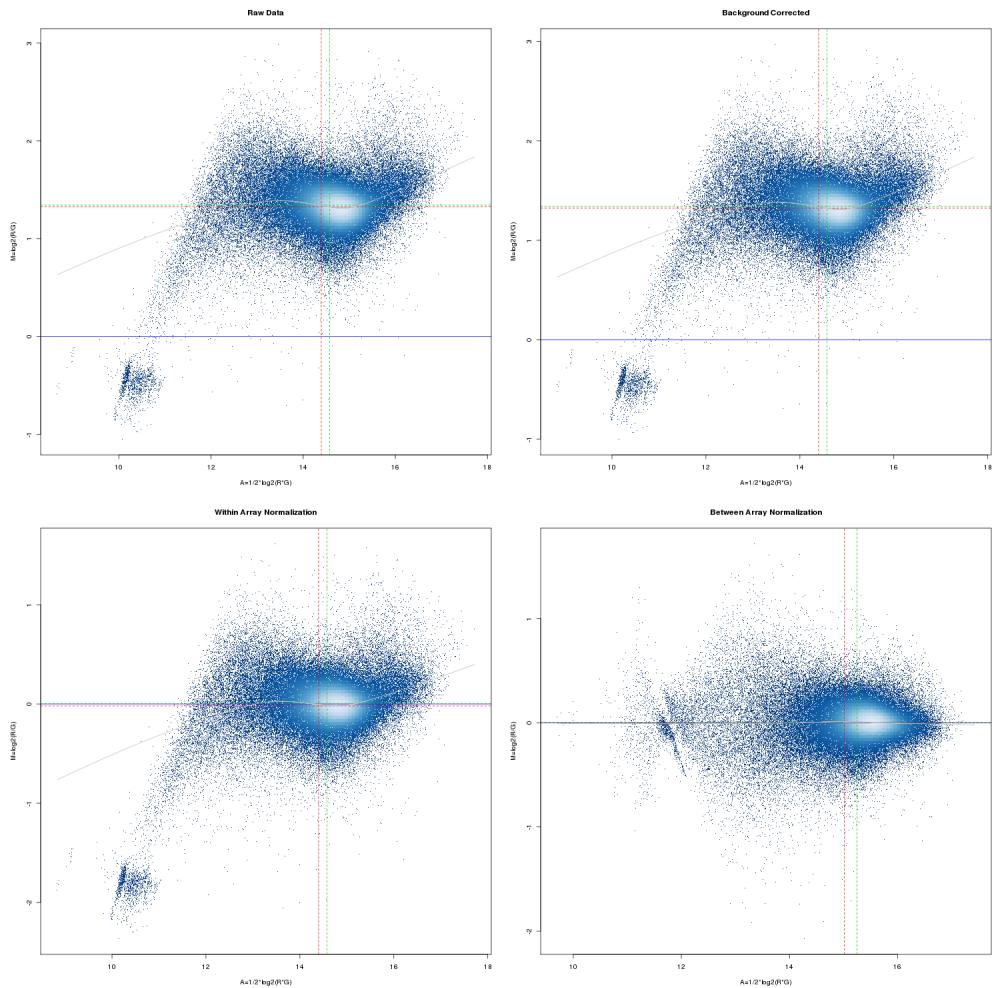
Results of all processing steps; background correction within array normalization, and between array normalization.

## 6.16 0258\_11\_swap Results



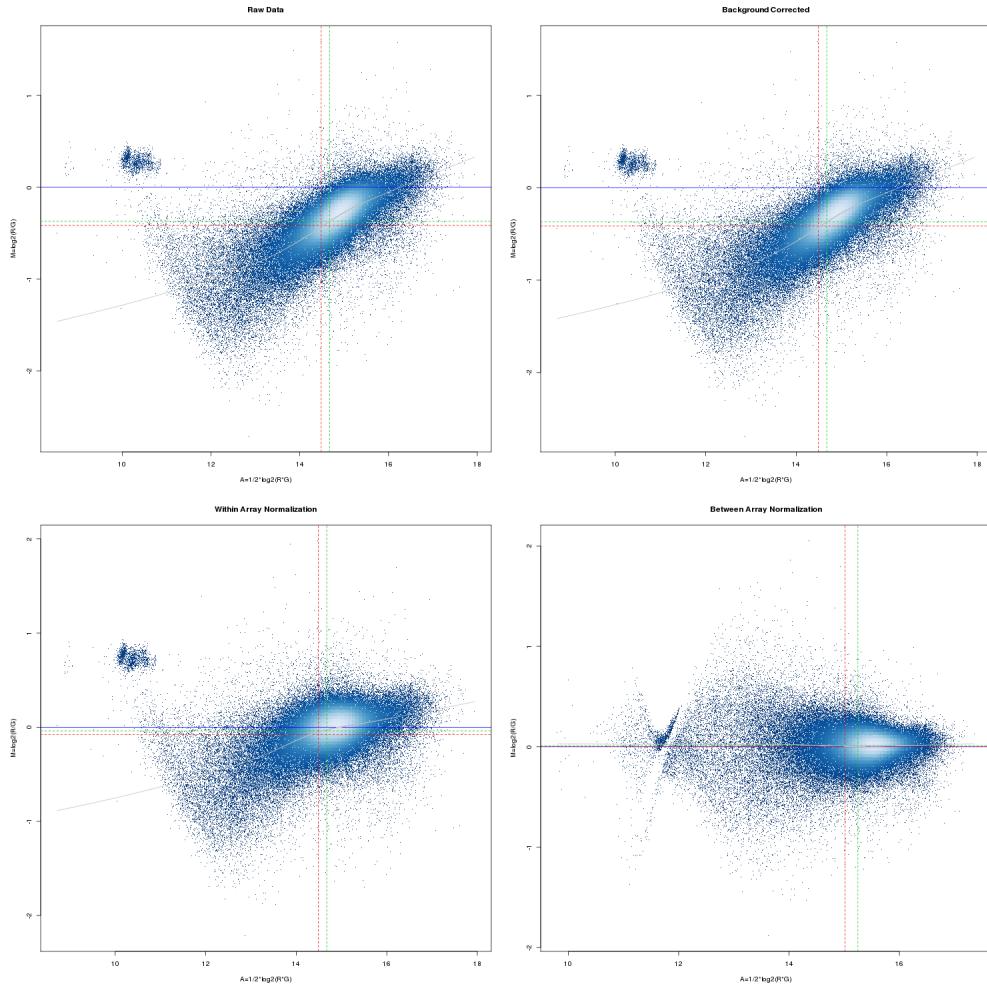
Results of all processing steps; background correction within array normalization, and between array normalization.

## 6.17 0258\_21 Results



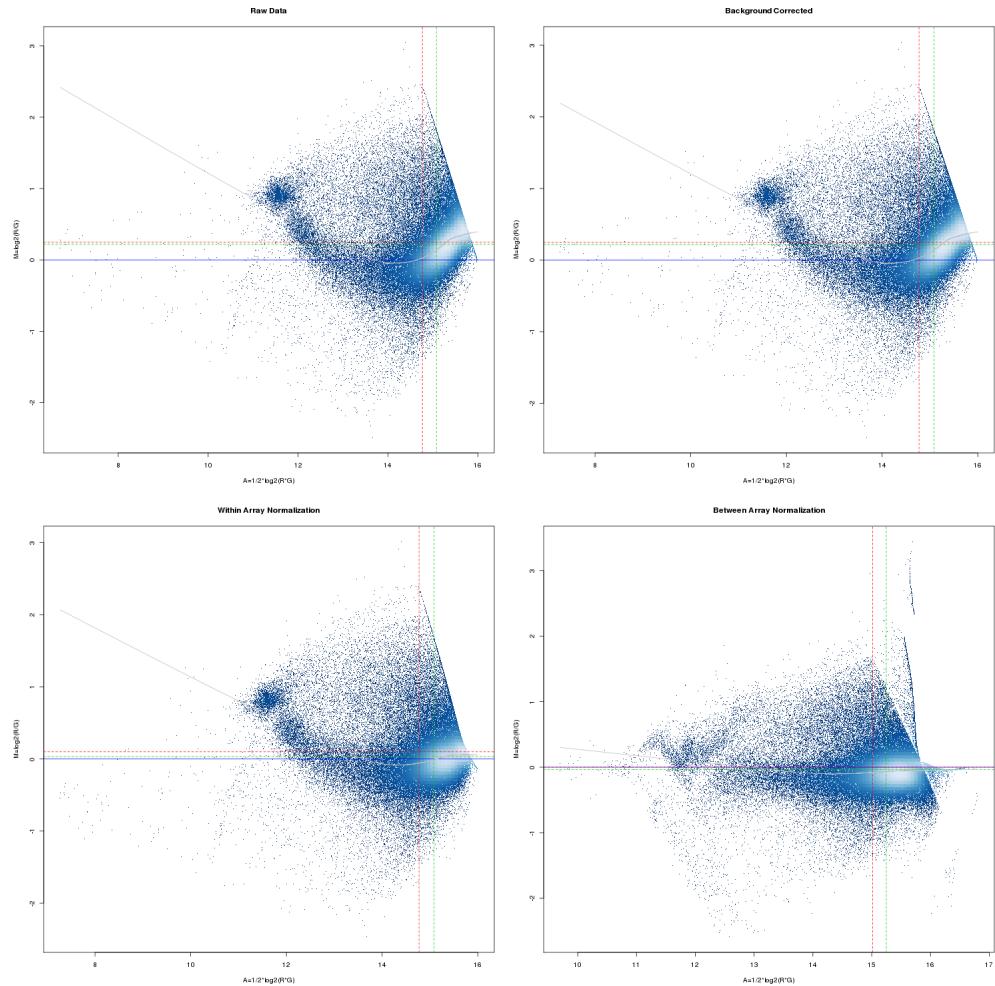
Results of all processing steps; background correction within array normalization, and between array normalization.

## 6.18 0258\_21\_swap Results



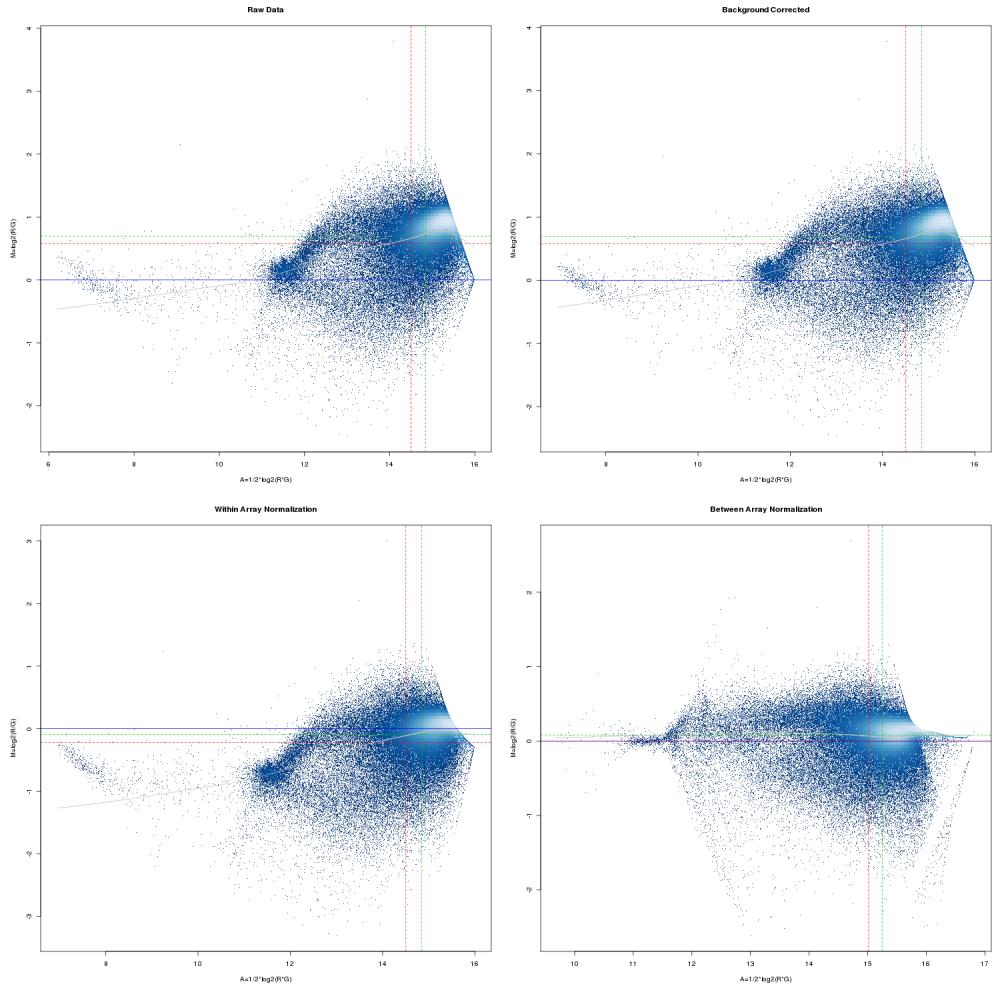
Results of all processing steps; background correction within array normalization, and between array normalization.

## 6.19 0258\_13 Results



Results of all processing steps; background correction within array normalization, and between array normalization.

## 6.20 0258\_13\_swap Results

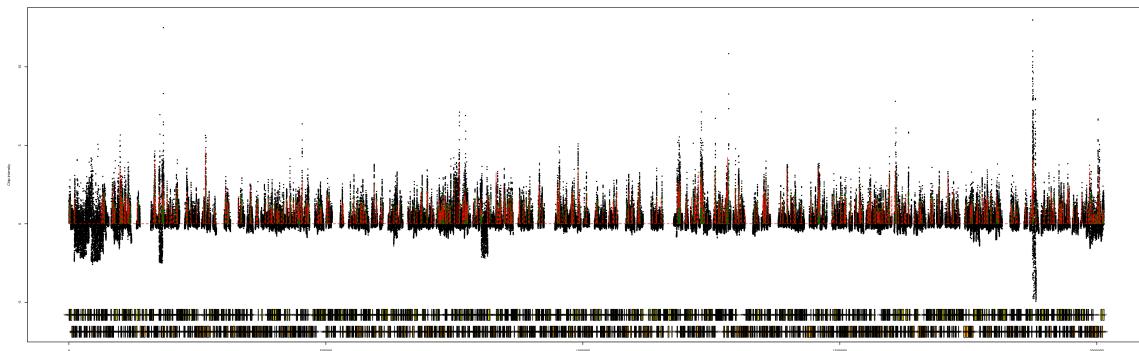


Results of all processing steps; background correction within array normalization, and between array normalization.

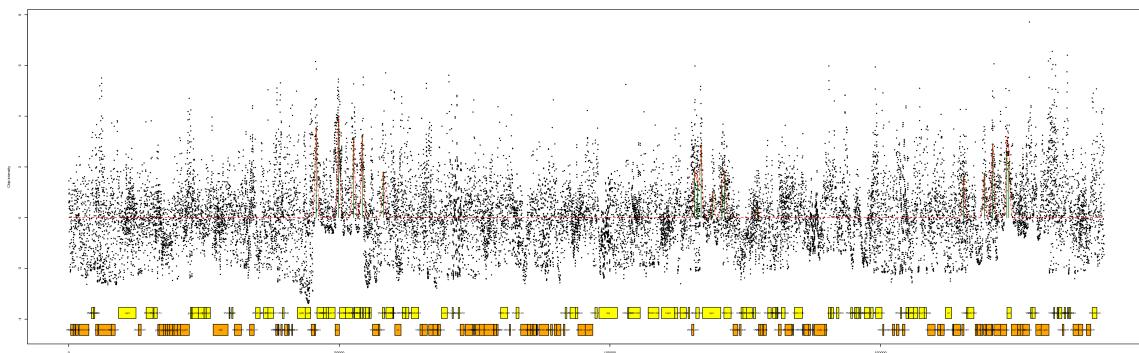
## 7 MeDiChI Peak Finding

Run MeDiChI with paramters max.steps=100, fit.res=10, n.boot=10, boot.sample.opt=residual.

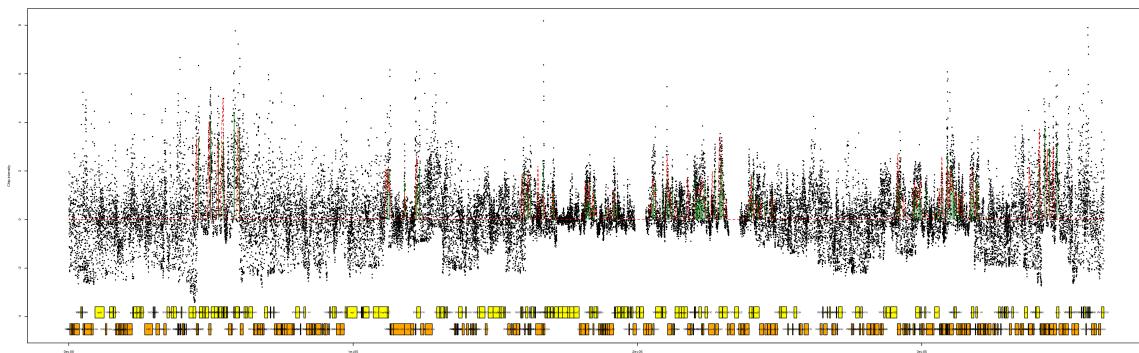
### 7.1 noH<sub>2</sub>O<sub>2</sub>



MeDiChI peaks of noH<sub>2</sub>O<sub>2</sub> on the main chromosome

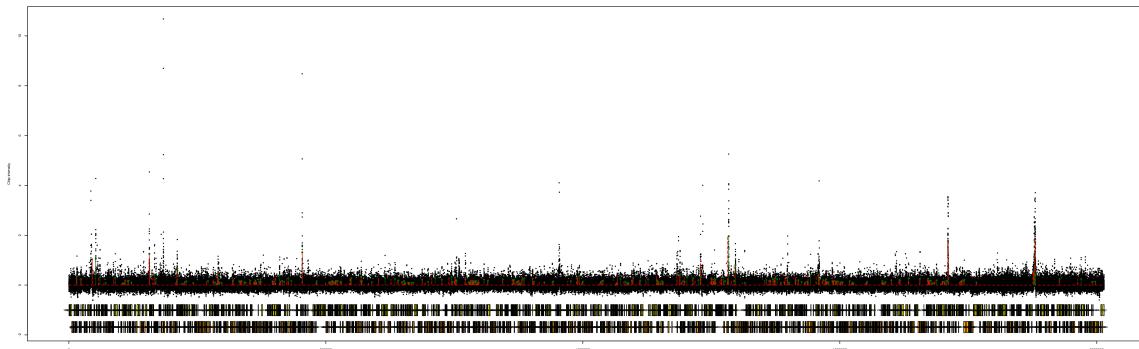


MeDiChI peaks of noH<sub>2</sub>O<sub>2</sub> on pNRC100

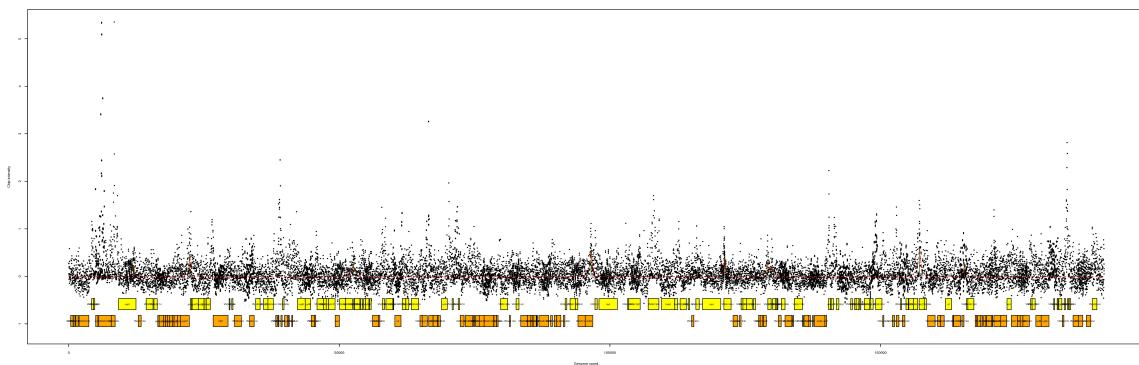


MeDiChI peaks of noH<sub>2</sub>O<sub>2</sub> on pNRC200

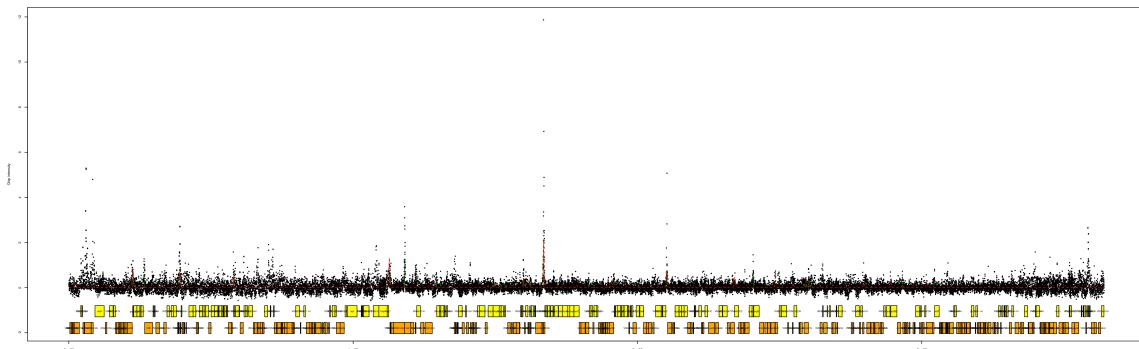
## 7.2 wH2O2\_10m



MeDiChI peaks of wH2O2\_10m on the main chromosome

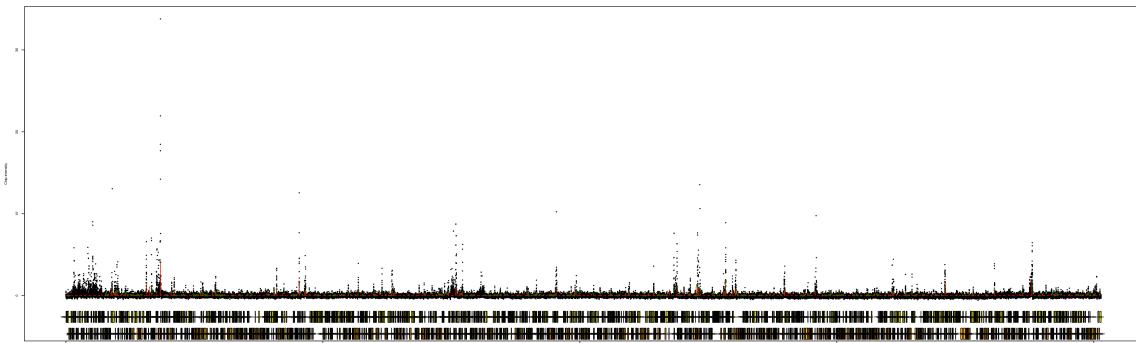


MeDiChI peaks of wH2O2\_10m on pNRC100

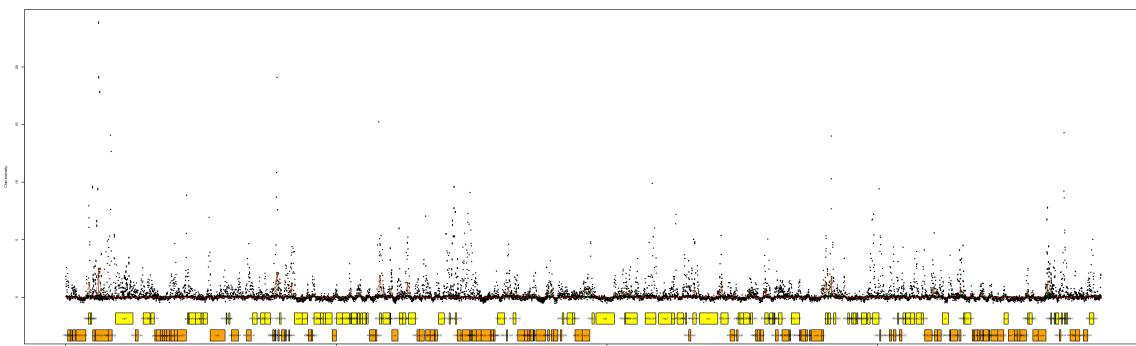


MeDiChI peaks of wH2O2\_10m on pNRC200

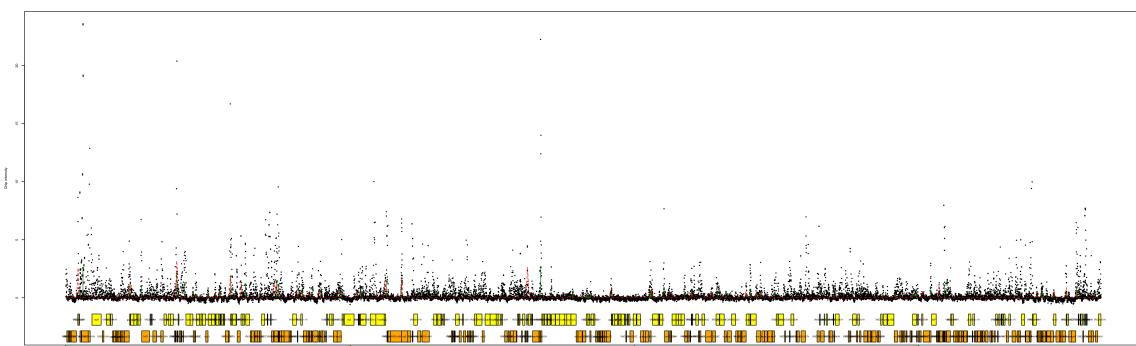
### 7.3 wH<sub>2</sub>O<sub>2</sub>\_20m



MeDiChI peaks of wH<sub>2</sub>O<sub>2</sub>\_20m on the main chromosome

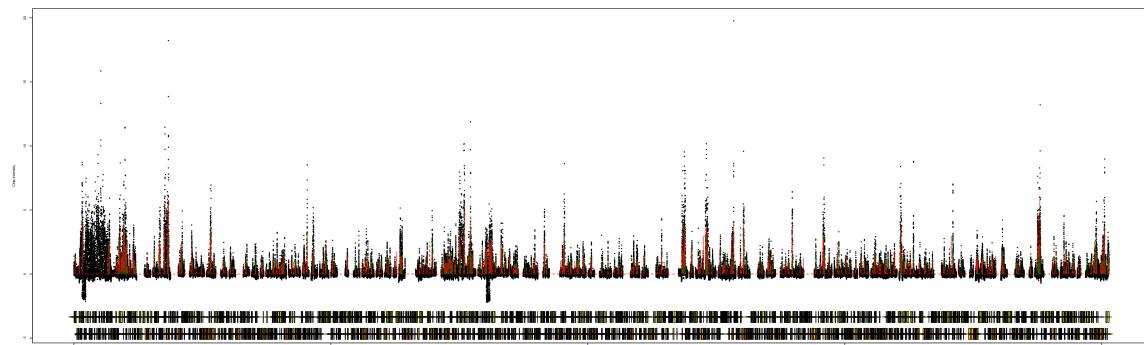


MeDiChI peaks of wH<sub>2</sub>O<sub>2</sub>\_20m on pNRC100

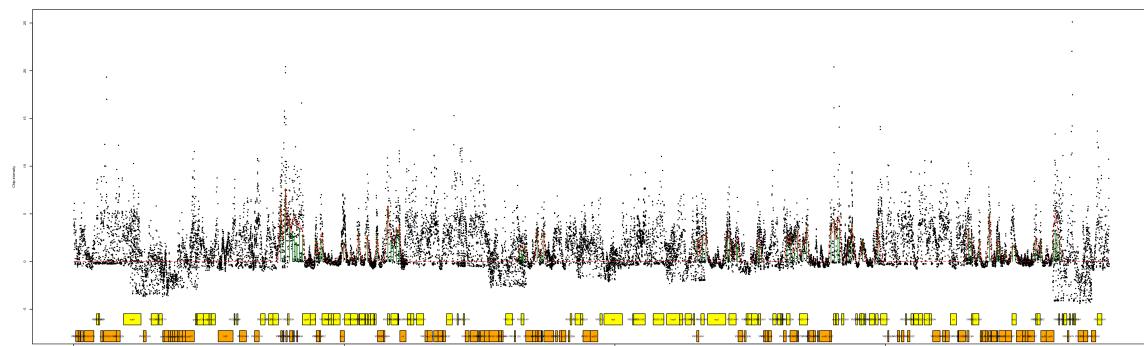


MeDiChI peaks of wH<sub>2</sub>O<sub>2</sub>\_20m on pNRC200

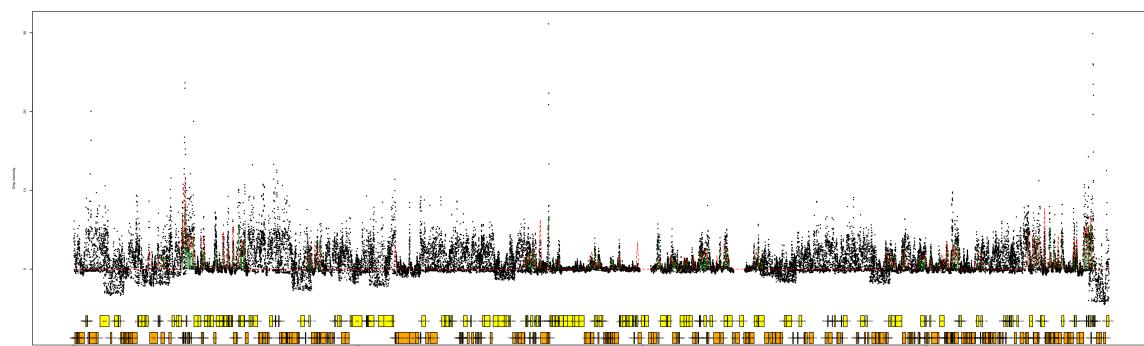
## 7.4 wH2O2\_60m



MeDiChI peaks of wH2O2\_60m on the main chromosome

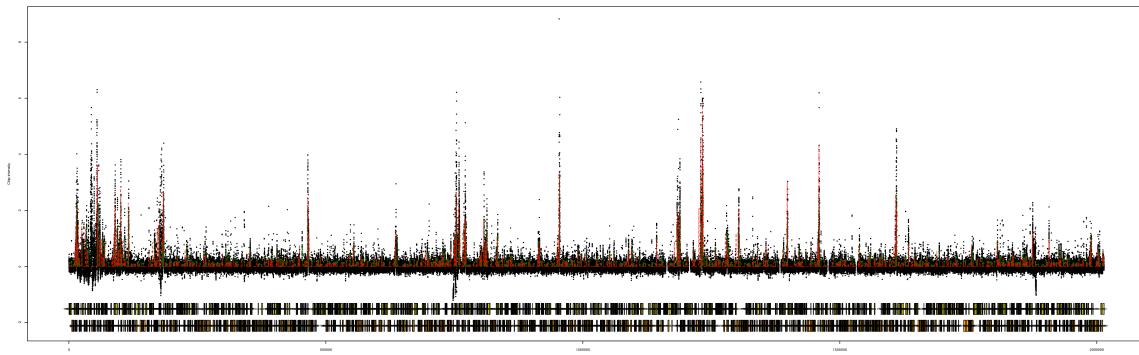


MeDiChI peaks of wH2O2\_60m on pNRC100

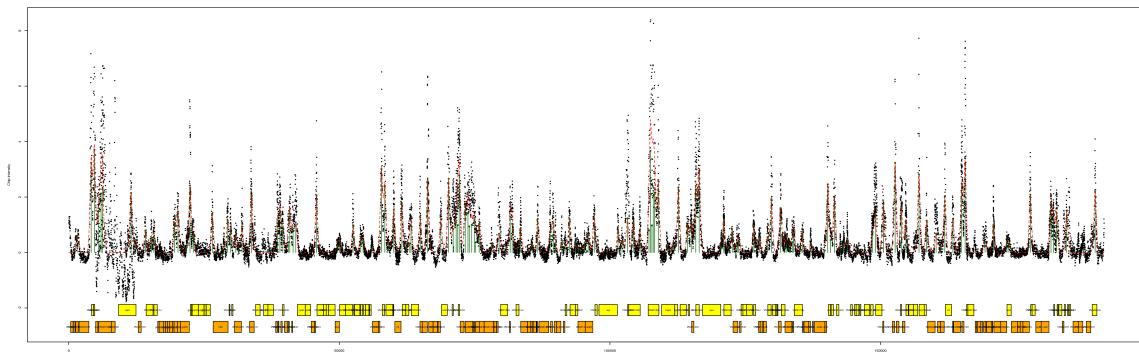


MeDiChI peaks of wH2O2\_60m on pNRC200

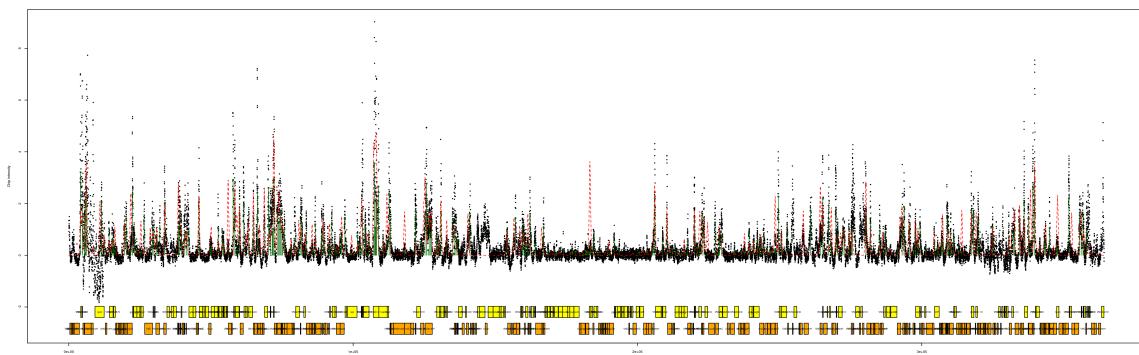
## 7.5 0258\_22



MeDiChI peaks of 0258\_22 on the main chromosome

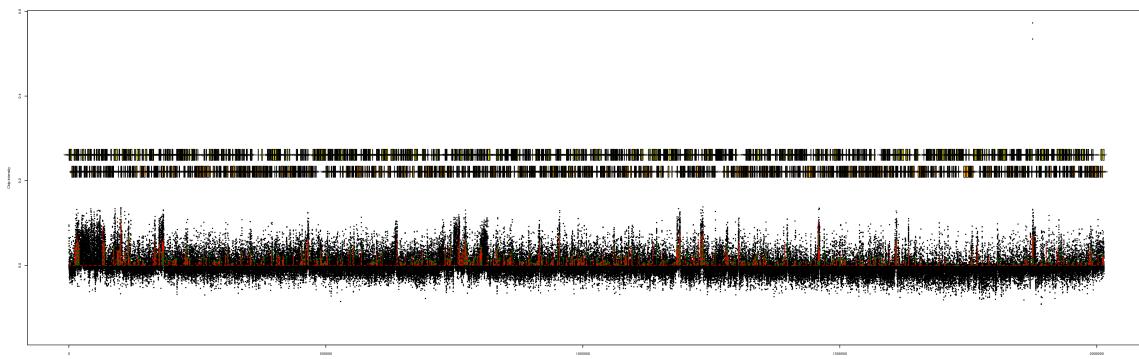


MeDiChI peaks of 0258\_22 on pNRC100

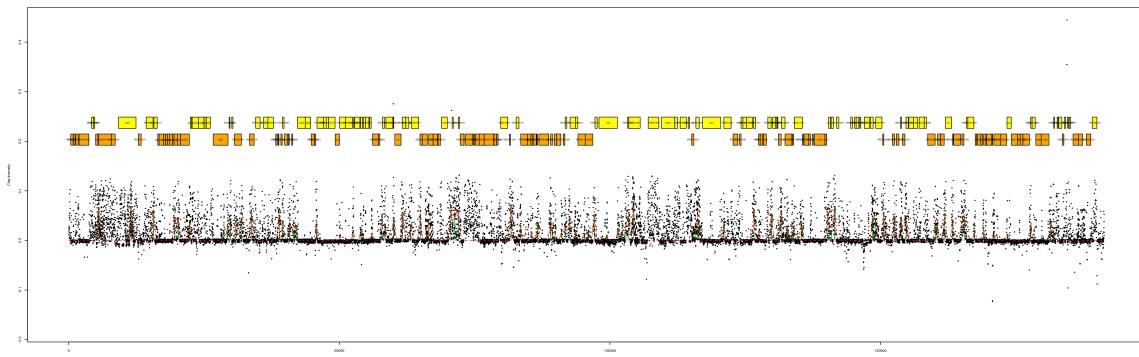


MeDiChI peaks of 0258\_22 on pNRC200

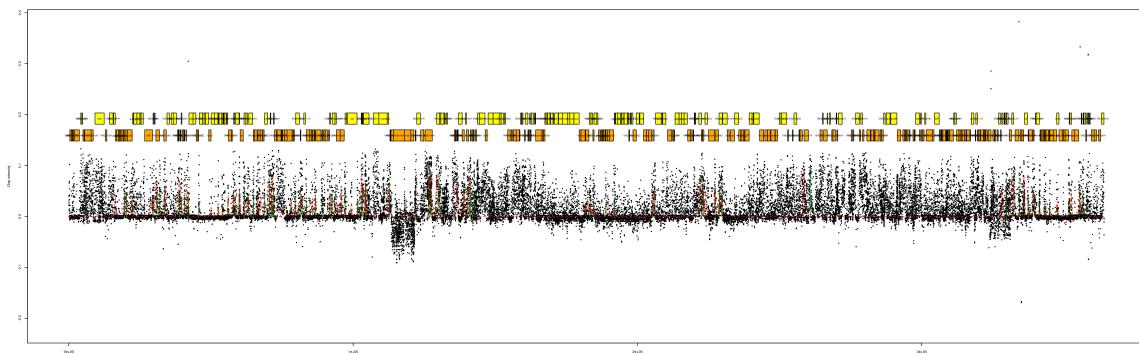
## 7.6 0258\_32



MeDiChI peaks of 0258\_32 on the main chromosome

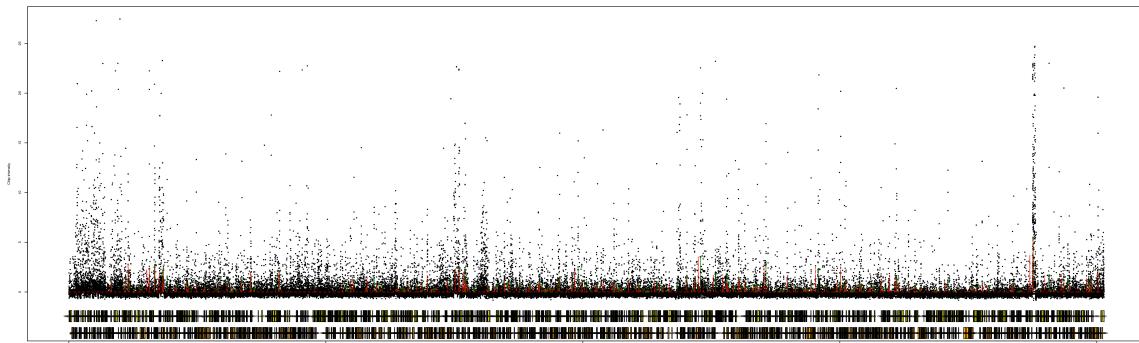


MeDiChI peaks of 0258\_32 on pNRC100

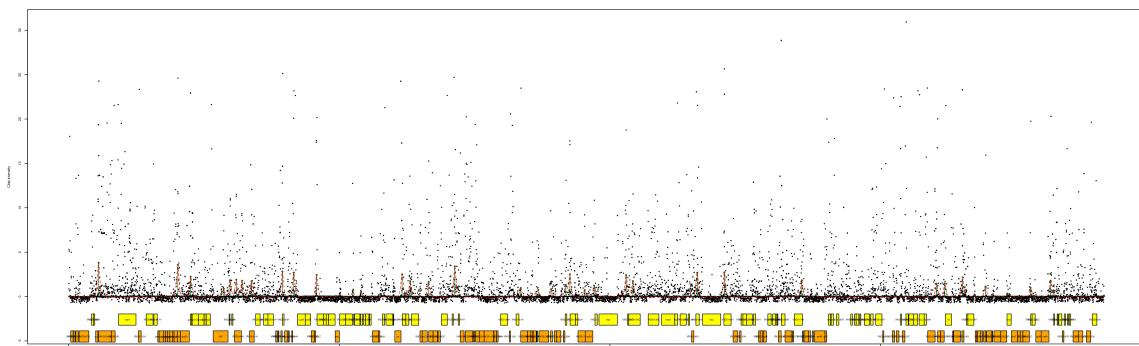


MeDiChI peaks of 0258\_32 on pNRC200

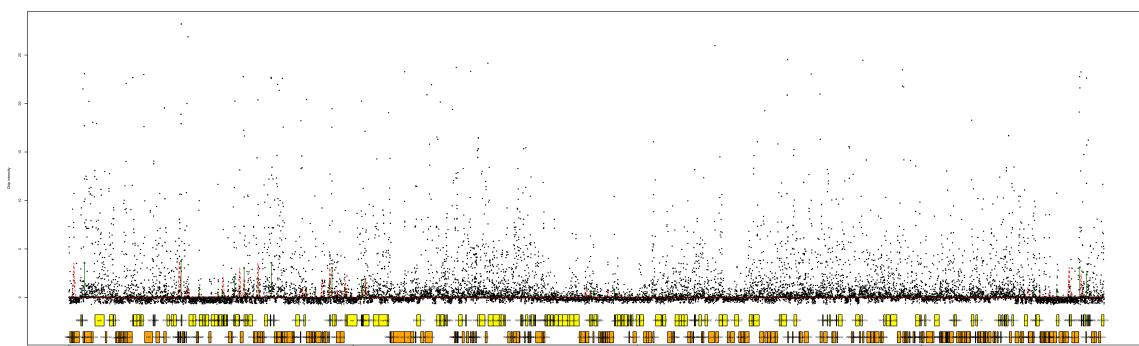
## 7.7 0258\_33



MeDiChI peaks of 0258\_33 on the main chromosome

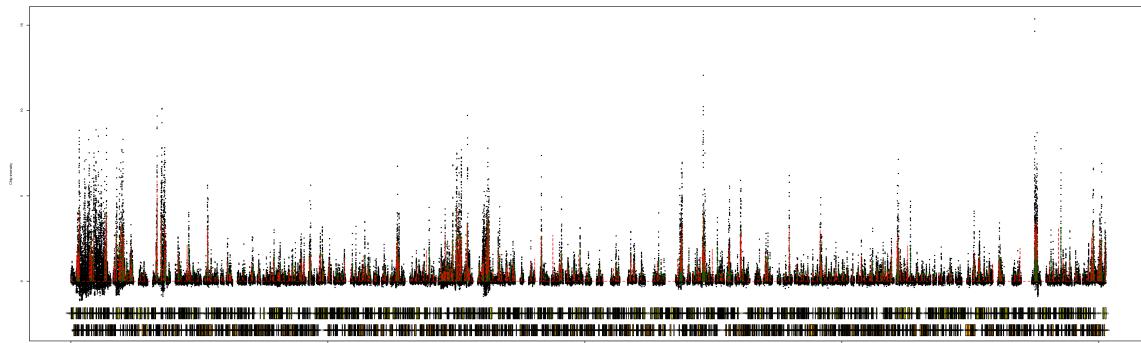


MeDiChI peaks of 0258\_33 on pNRC100

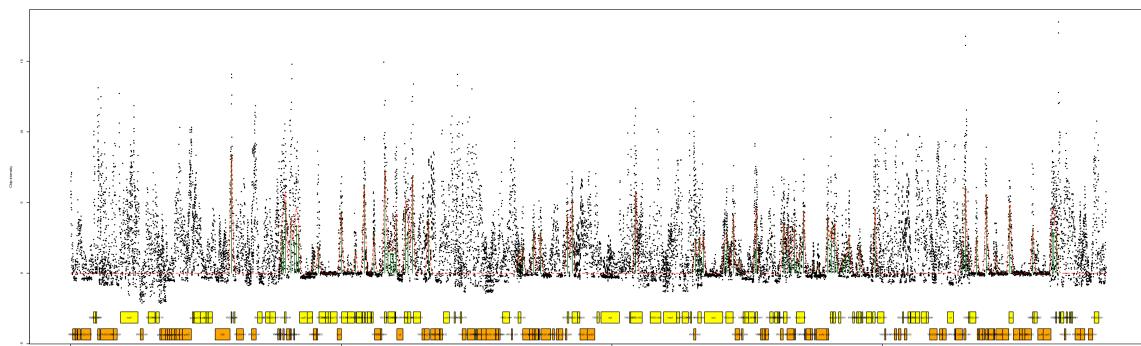


MeDiChI peaks of 0258\_33 on pNRC200

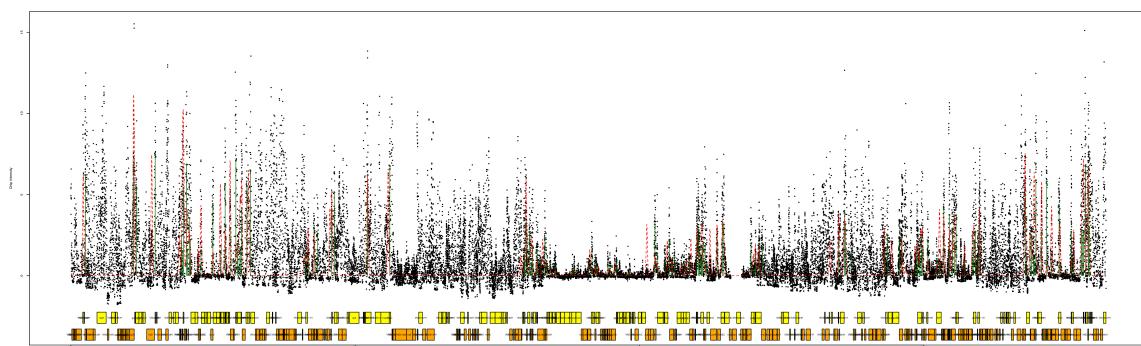
## 7.8 0258\_11



MeDiChI peaks of 0258\_11 on the main chromosome

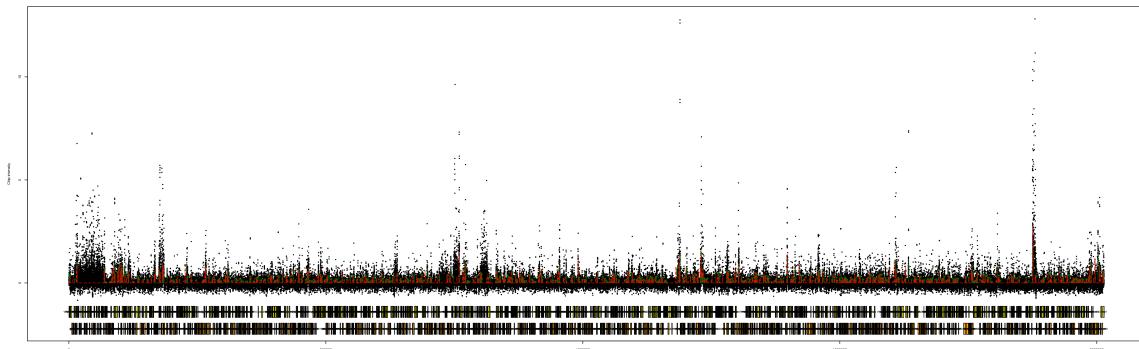


MeDiChI peaks of 0258\_11 on pNRC100

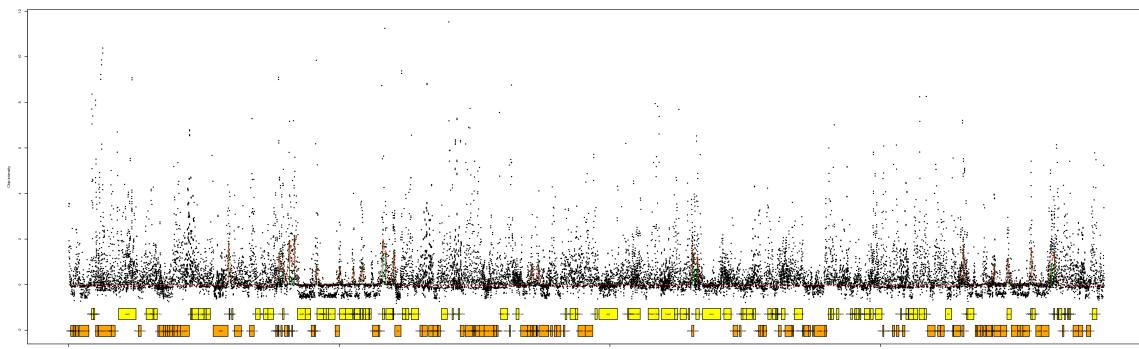


MeDiChI peaks of 0258\_11 on pNRC200

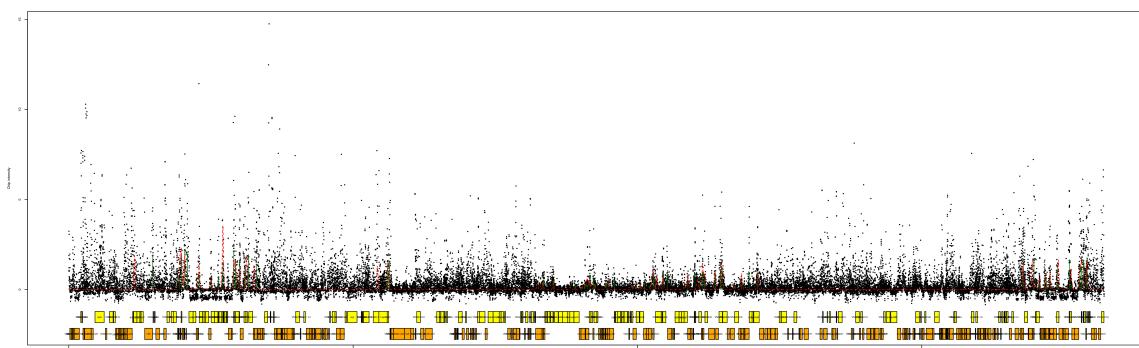
## 7.9 0258\_13



MeDiChI peaks of 0258\_13 on the main chromosome



MeDiChI peaks of 0258\_13 on pNRC100



MeDiChI peaks of 0258\_13 on pNRC200

## 8 Completion

Time to generate document: 49.92624