


A



 Workflow

Welcome

DNA replication is a highly regulated process involved in maintaining of genome stability. It relies on a spatio-temporal program determining where and when replication starts along the genome. The replication timing (RT) program is finely tuned and reflects the high order organization of the metazoan genome. It is clear that this program evolves with cell differentiation and cell fate. Moreover aberrant DNA RT has been reported in many diseases including cancer.

The protocol developed to study the RT of a given cell type is well established. We have improved the method permitting analysis of RT of the entire genome in mammalian cells. We implemented two bioinformatic tools, START-R (Simple Tool for Analysis of Replication Timing with R) and START-R viewer. START-R written in R language, makes it possible to analyse the data from microarrays with five steps: normalization, smoothing, TTR (Timing Transition Region) and CTR (Constant Timing Region) detection and finally differential analysis. For each step, several options are available and selectable by the user. At the end, with a classical computer configuration, the analysis takes few minutes. START-R viewer is a tool that allows to visualize the data generated by START-R as does a genome browser. These softwares have a user-friendly interface requiring only a "simple click" and can be run under Windows, Mac OS X and Linux systems without virtual machine. The script will remain open to the scientific community so that the software can always be improved or developed.

[Visit website !](#)

This web application will allow you to enter the necessary informations for the analysis to run correctly. All the steps are organized in workflow. Each input is checked to avoid incorrect analyses.

Next

B



 Workflow **Initialisation**

Initialisation

Organism

Two types of organisms can be studied in START-R

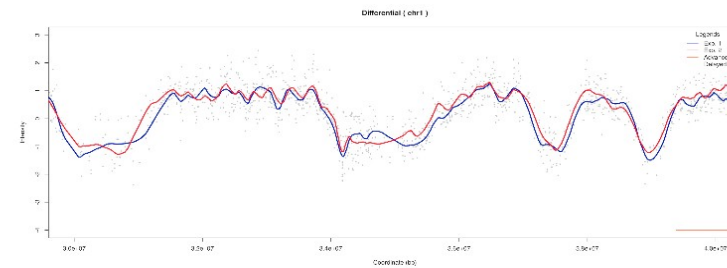
- Human (hg18)
 No centromere
 Others (Human > hg19 or other species)



Differential

Detect differences between 2 experiments

- Yes
 No



Prev.

Next



C

Import data

In this step, you set up and select the files to import

General parameters

Column name of green signal :

gProcessedSignal

Column name of red signal :

rProcessedSignal

Early fraction

- Cy3
- Cy5

Late fraction

- Cy3
- Cy5

Select file

Experiment 1

Replica 1

Browse... US92003687_252206079783_

Upload complete

Header

- Yes
- No

Separator

- Comma
- Semicolon
- Tab

Quote

- None
- Double Quote
- Single Quote

Number of microarray information lines to skip

9

Display

- Head
- More (100 lines)
- No

Preview

Show 25 entries

Search:

FEATURES	FeatureNum	Row	Col	SubTypeMask	SubTypeName	Start	Sequence	ProbeUID	ControlType	ProbeName	GeneName	SystematicN
DATA	1	1	1	0		0		0	1	HSCGHBrightCorner	HSCGHBrightCorner	HSCGHBrightC
DATA	2	1	2	0		0		1	1	DarkCorner2	DarkCorner2	DarkCorner2
DATA	3	1	3	0		0		1	1	DarkCorner2	DarkCorner2	DarkCorner2
DATA	4	1	4	0		0		1	1	DarkCorner2	DarkCorner2	DarkCorner2
DATA	5	1	5	0		0		1	1	DarkCorner2	DarkCorner2	DarkCorner2
DATA	6	1	6	0		0		1	1	DarkCorner2	DarkCorner2	DarkCorner2

Showing 1 to 6 of 6 entries

Previous 1 Next

Replica 2

Browse... No file selected

Display

- Head
- More (100 lines)
- No

Header

- Yes
- No

Separator

- Comma
- Semicolon
- Tab

Quote

- None
- Double Quote
- Single Quote

Number of microarray information lines to skip

9

Preview

D



Workflow

Initialisation

Import data

Normalisation

Normalisation

In this step, you can choose the types of normalizations you want to apply to your data.

Intra array

Correction of global intensities in a microarray.

Methods

- loess
- control
- composite
- printiploess
- median
- none
- robustspline

Description

The loess normalization methods ("loess", "printiploess" and "composite") were proposed by Yang et al (2001, 2002). Smyth and Speed (2003) review these methods and describe how the methods are implemented in the limma package, including choices of tuning parameters. More information on the loess control parameters span and iterations can be found under loessFit.

Inter replica

Standardization of replicates to make them comparable.

Methods

- scale
- quantile
- none
- cyclicloess

Description

Scale (method="scale") scales the columns to have the same median.

Inter experience

With START-R, it is possible to compare two experiences. However, from one microarray to another, the quantity of material deposited or the quality of marking may differ. In order to make the most accurate comparisons possible, a standardization between the data from the two experiments is proposed in order to limit this type of experimental bias.

Methods

- standardization
- without normalization
- unitization

Description

(x-mean)/sd

Prev.

Next



Workflow

Initialisation


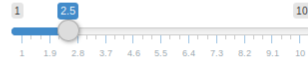
Import data

Normalisation

Analysis

Analysis Parameters

In this step, you can set up the different steps of the analysis.


Steps	Objective	Method	Method description	Additional options
<input checked="" type="checkbox"/> Smooth	The smoothing step will generate a curve that will try, depending on the algorithm used, to be the most representative of this cloud of points. Once again, START-R allows the user to choose several smoothing functions.	Loess	Fit a polynomial surface determined by one or more numerical predictors, using local fitting.	<p>Span : The parameter alpha which controls the degree of smoothing. Fitting is done locally. That is, for the fit at point x, the fit is made using points in a neighbourhood of x, weighted by their distance from x (with differences in 'parametric' variables being ignored when computing the distance). The size of the neighbourhood is controlled by alpha (set by span or <code>enp.target</code>). For $\alpha < 1$, the neighbourhood includes proportion alpha of the points, and these have tricubic weighting (proportional to $(1 - (\text{dist}/\text{maxdist})^3)^3$). For $\alpha > 1$, all points are used, with the 'maximum distance' assumed to be $\alpha^{1/p}$ times the actual maximum distance for p explanatory variables.</p> 
<input checked="" type="checkbox"/> TTR	The curves obtained, after smoothing from DNA chip data, show 'flat' areas that replicate either early or late, called CTRs for Constant Timing Regions. Between an early and a late zone, there is a transition zone called TTR for Timing Transition Region.			
<input checked="" type="checkbox"/> Segmentation	Detection of Constant Timing Regions with TTR information			<p>Standard deviation: The number of SDs between means to keep a split</p> 
<input checked="" type="checkbox"/> Fusion	Combine result of TTR and CTR			

Prev.

Next

F



-  Workflow
- Initialisation
- Import data
- Normalisation
- Analysis
- Outputs**

Outputs

File Outputs (positions,intensity,...)

- .bed
- .txt
- both

Graphical outputs

- Yes
- No

Prev.

Next

Summary

General informations

Organism : Human
Differential : No

Selected Files

Experience 1

Replica 1

File name : US92003687_252206079783_S01_CGH_1105_Oct12_1_1.txt

Replica 2

File name :

Experience 2

Replica 1

File name :

Replica 2

File name :

Normalisation

Intra array : loess
Inter replica : scale
Inter experience : n1

Analysis

Smooth

Realiazed? : TRUE
Method : Loess
Span : 500000

TTR

Realiazed? : TRUE

Segmentation

Realiazed? : TRUE
SD : 2.5

Differential

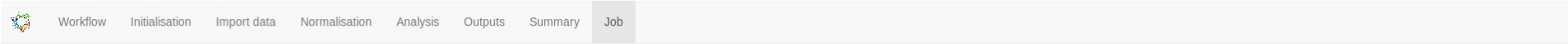
Realiazed? : FALSE
Method : Mean method
PVT : 0.05
Method adjustment PV : holm
Empirical threshold: 0.45
Overlap: 30
Windows size: 60

Outputs

File format : both
Graphical : Yes

[Prev.](#)[Validate](#)

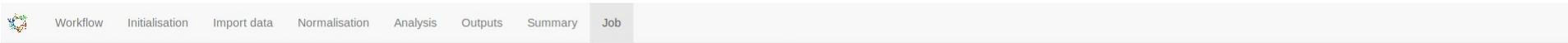
H



Click to run analysis



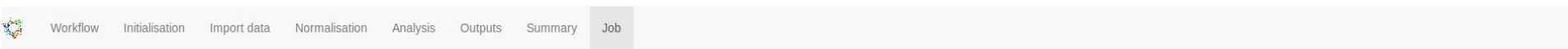
I



Processing



J



The analysis was successful!



K



Choose CSV File

Browse... chr9.SRV

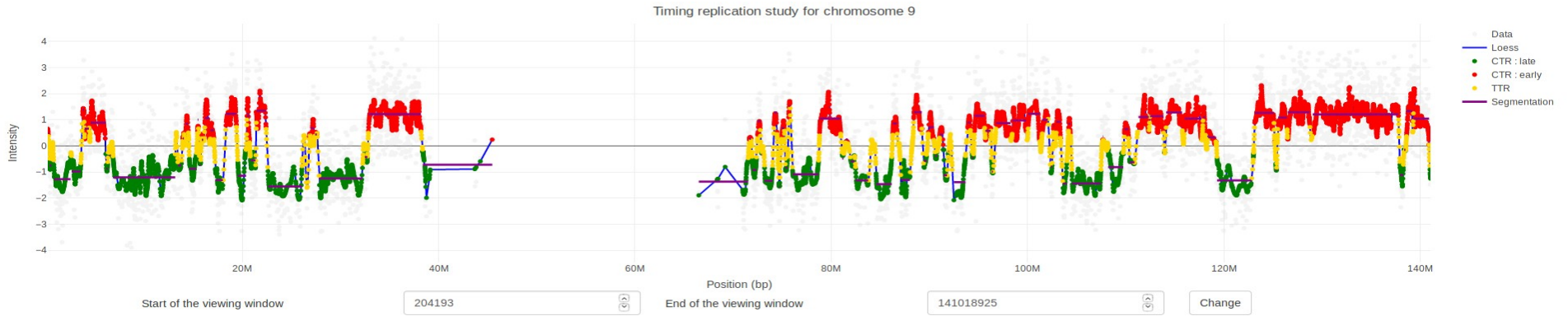
Upload complete

Load data

Data informations

Organism : Human
Chromosome : chromosome 9
Method differential : Mean method
of advanced : 0
of delayed : 0

Smooth method : Loess
Normalisation intra : loess
Normalisation inter replicas : scale
Normalisation inter experiences : standardization



Customization

Data #CCCCCC

Smooth #0000FF

CTR : late #00FF00

CTR : early #FF0000

TTR #FFD700

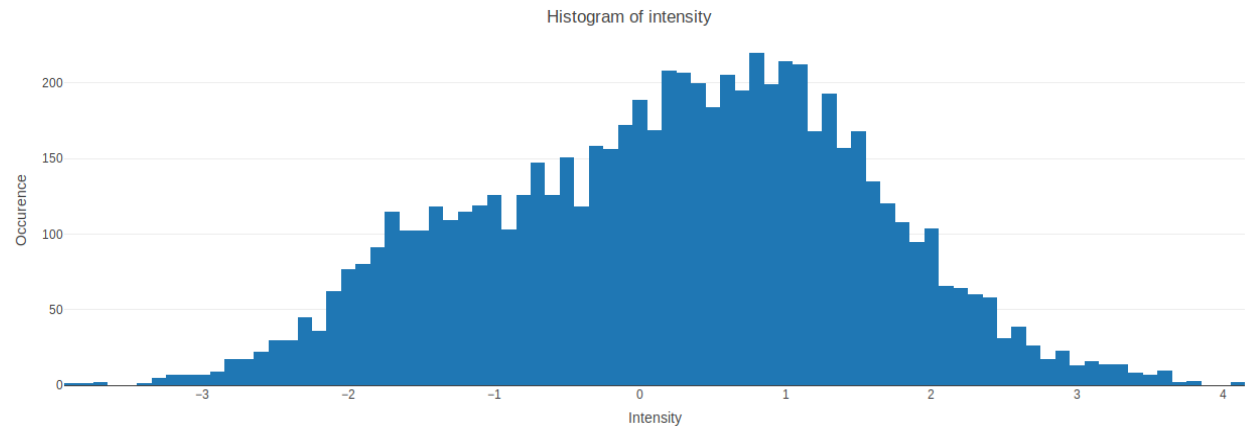
Segmentation #8B008B

Change

Select general information to plot:

- None
- Data intensity
- CTR/TTR boxplots

General informations about chromosome 9



Supplementary Figure S1: Screenshots showing the web-based user-friendly interface of START-R Analyzer and START-R Viewer.

A) START-R Analyzer homepage.

START-R Analyzer is designed to analyze RT data with 5 steps: normalisation, smoothing, TTR and CTR detections and differential analysis. Several options are available and can be selected by the user for each step. All the steps are organized in workflow with a “Next” button allowing the user to access the next page. Each input is checked to avoid incorrect analyses. A link to the website containing precisions on the START-R Analyzer protocol and START-R Viewer use is available in this homepage (<https://thomasdenecker.github.io/START-R/>).

B) Organism and differential analysis selection.

The initialisation step consists in choosing the analyzed organism which can be either Human (for microarray data obtained with the hg18 genome assembly), or No centromere (for example for Drosophila genomic data) or Others (for human data obtained with a genome assembly higher than hg19 or for other species). We have to note that it's also possible to upload a file containing the centromeres coordinates for non mammalian organisms. The user has then to choose to perform a differential analysis or not. “Prev” and “Next” buttons allow the user to move forward or backward within START-R Analyzer.

C) Upload of intensity files.

The user has first to set up the general parameters for the Early and Late replicating fractions, by choosing the column names of the files for the green and red signals and attributing the Cy3 or Cy5 dye to the Early or Late fractions. The selected files for the analyses are imported and should be in txt format. A format compatible with START-R Analyzer can be obtained for NGS data by using a script available on GitHub in the “supplement script” file (`convert_bamcoverage_file.R`). The user can choose several options concerning the imported files as the presence of a header, the kind of separator, the presence of a quote in the file, the number of information lines to skip before launching the analysis and the display of a part of the loaded files.

D) Selection of intra-array, inter-replicate and inter-experiment normalisation algorithms.

The next step is the choice of the normalisation's algorithms that the user wants to apply to the data. For each method available for each step of the normalisation (intra array, inter replica and inter experience), a short explanation of the method will be displayed on the screen when checked by the user.

E and F) Selection of differential analysis settings and output files.

The set up of the analysis parameters is next chosen by the user. For each step, the objective of the analysis is described beside. The smoothing method is chosen with a drop-down menu and a short description of the method is displayed on a dialog box. An additional option allows to fix the span to control the degree of smoothing. The second step is the TTR detection which is followed by the segmentation step to detect the CTR. The user can then adjust the standard deviation if wanted. The last step is the fusion that will combine the results of the TTR and CTR detections.

The file and graphical outputs are then selected by the user.

G) Validation of the analysis summary.

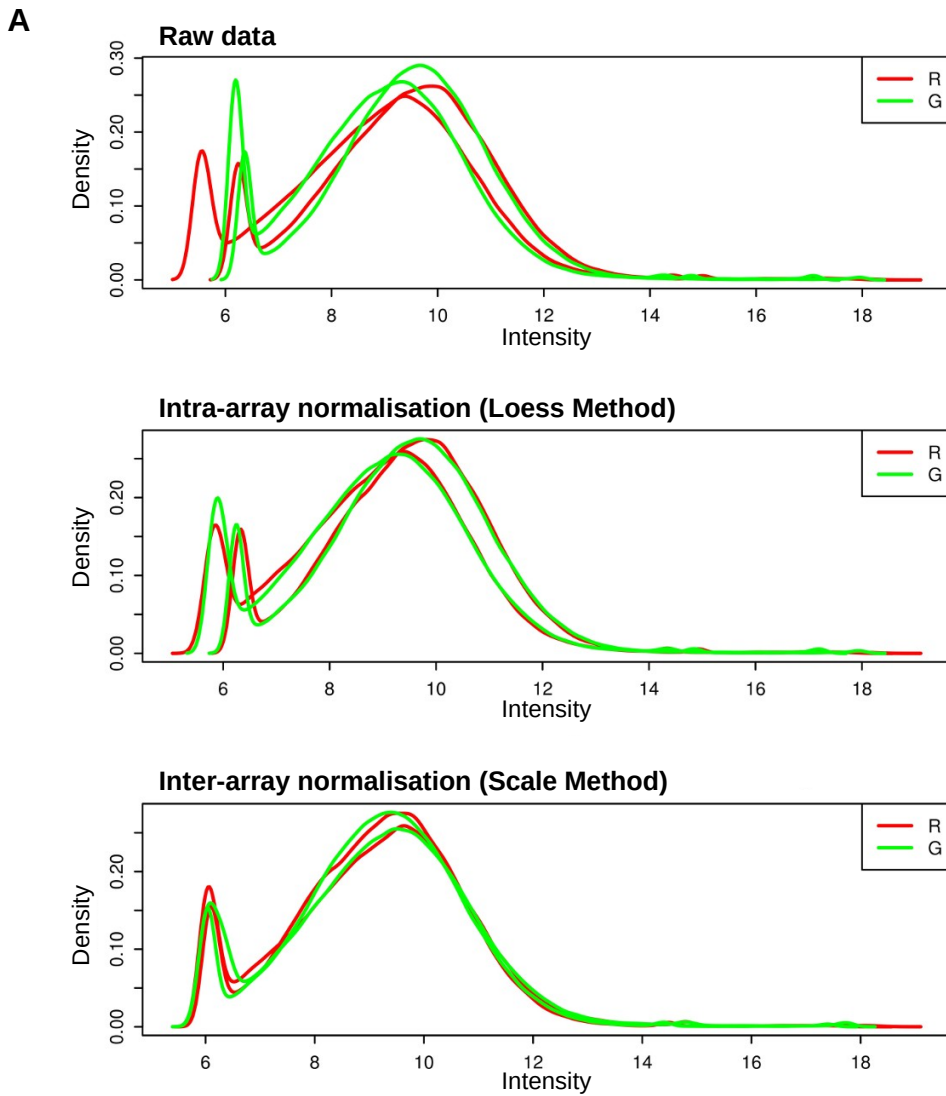
The user can click to the “Validate” button after checking the summary of the START-R analysis. The “Prev” button allows to get backward inside the software if needed.

H, I and J) The analysis running sections. The user just has to click on the “Run” button to launch the analysis, which will be processed in around 15 minutes. A green check mark will appear on the screen when the analysis is performed. When an analysis is done with START-R analyzer, all results for this analysis are available in the folder START-R_analyzer/Outputs. The folder name begins with the date of analysis and successive numbers (the first analysis will be named 20200511_1, the second analysis 20200511_2 ...). A summary of the analysis can be found in the codebook.txt file. This software offers several types of outputs to evaluate the results. Text outputs are in bed format. The BED format consists of one line per feature, each containing 3-12 columns of data, plus optional track definition lines. The first three fields in each feature line are required: chrom (name of the chromosome or scaffold), chromStart and chromEnd (start and end position of the feature in standard chromosomal coordinates). With bed files outputs, START-R results can be integrated in Galaxy workflow to perform further analyses. Galaxy is an open source web-based platform for data intensive biomedical research.

K) START-R Viewer interface allowing the START-R Analyzer files to be uploaded and the RT profile to be visualized for each chromosome with general information about each chromosome.

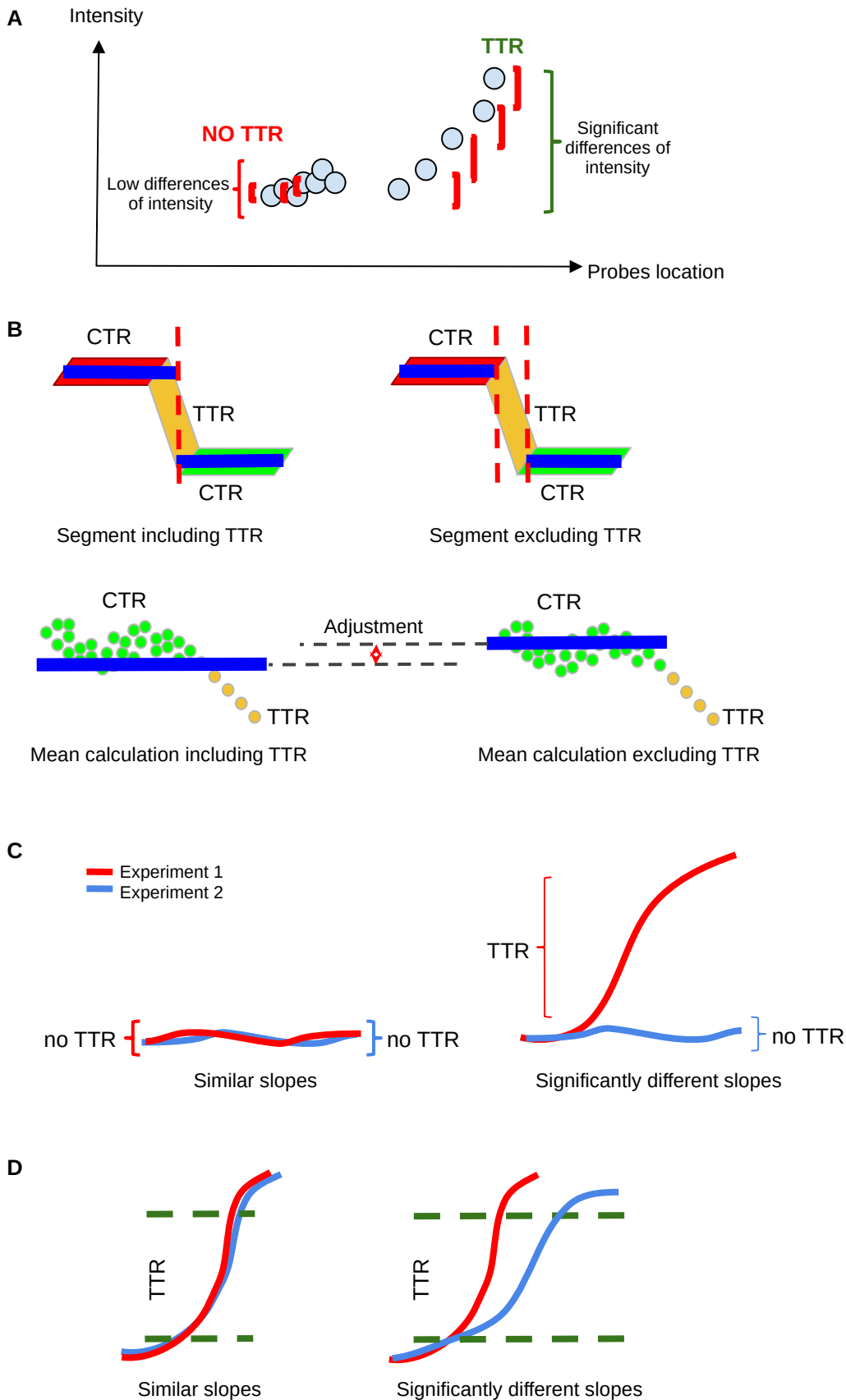
START-R Viewer allows the visualization of the analyzed data for each chromosome as a genome browser. After the analysis, the user can find different informations in the Differential file or in the Experience 1 file from the output folder, depending of the type of analysis performed by START-R (differential analysis or not). Files showing the analysis for each chromosome are available in pdf format. The user can also find a file containing differential percentage for each chromosome and a file named differential_position in txt format showing for each RT changing region its location on a specific chromosome with its associated chromosome coordinates chromStart, chromEnd and its Status (advanced or delayed). A viewer folder containing the results for each chromosome allows to upload files in Javascript and SRV format for a visualization on START-R viewer. A summary of the analyses done with START-R Analyzer is displayed nearby the load data button. The RT profile of the selected chromosome is displayed in a dynamic chart showing the distribution of RT intensities along the chromosome, with positives intensities displaying early replicating regions and negative ones displaying late replicating regions. Early and late CTRs appear in red and green, respectively and TTRs in yellow. Segments corresponding to regions of constant timing are shown in purple. Grey spots indicate data from the RT experiment and the blue line depicts the smooth method used for the analysis. START-R viewer offers a dynamic exploration for each chromosome with the Plotly library. This dynamic chart gives the user the possibility to zoom in the chart by selecting a shorter window on the chart as in a genome browser or by indicating the chromosome coordinates in the viewing windows and clicking on the "Change" button. The user can also select the curves displayed in the chart directly by clicking on the chart or by selecting the curves in the Customization dialog box and eventually take pictures of the chart.

General information about the chromosome can be chosen in the "Select general information to plot" dialog box, like the histogram of intensity, showing the occurrence versus the intensity of the RT signal. We can note that the intensity of the signal is mostly distributed between -2 and +2, corresponding to late and early replicating regions, respectively. The user can also display the box plot of CTR/TTR intensity.



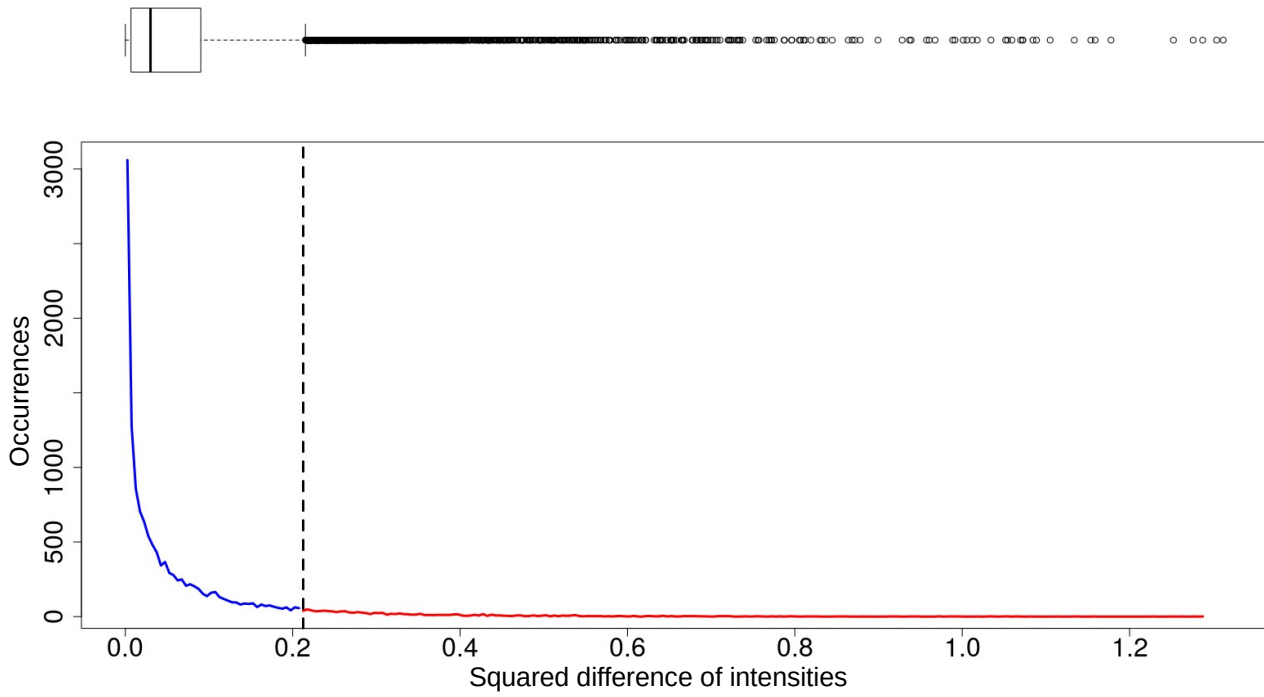
Supplementary Figure S2: Normalisation and smoothing of the chip data.

A) Normalisation of the chip data. Graphs depicting the distribution of log ratio intensities of Cy3 and Cy5 shown in red (R) and green (G), for 2 replicas before (raw data) and after intra-array normalisation with the Loess method, and inter-array normalisation with the scale method. B) Smoothing of the normalized data obtained by chip. Smoothing was applied on the normalised data to generate the most representative curve of the distribution of dots using either the Loess method based on linear regression or the Triangular method using a moving mean with overlapping windows.



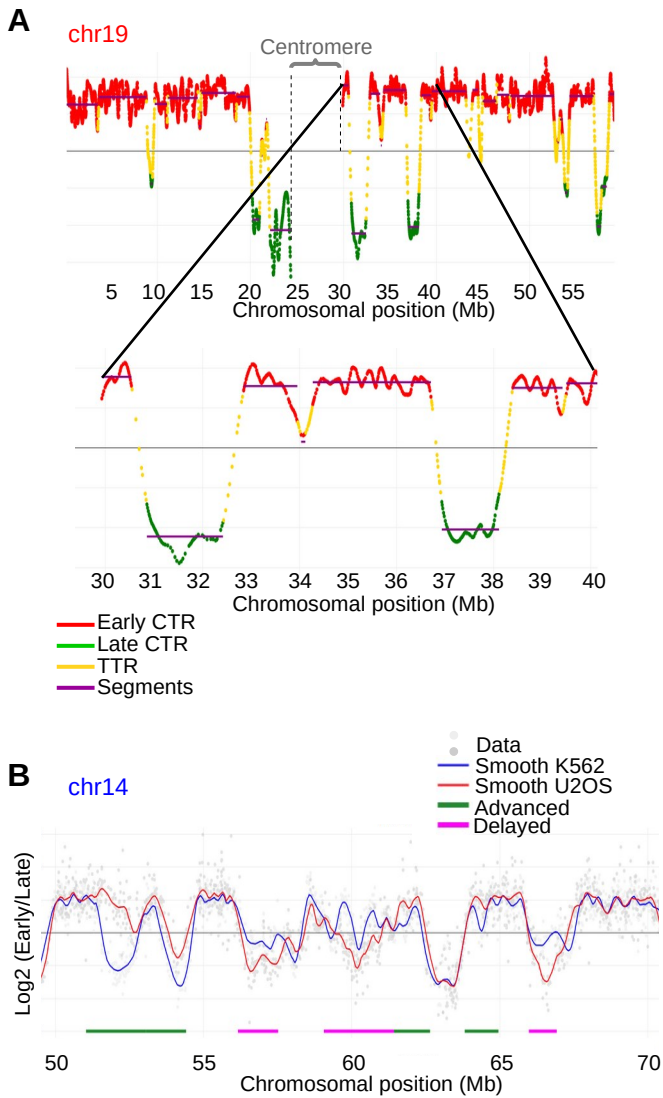
Supplementary Figure S3: TTR detection.

A) Diagram showing that a TTR is defined by significant differences of intensity of at least 3 consecutive adjacent probes. B) Schemes representing how CTRs are precisely defined by subtraction of the TTR. The calculation of the mean for the CTR region, excluding the TTR, allows an adjustment for the segment value and a more precise characterization of this domain. C) Drawings illustrating the comparison of the slopes for two experiments (red and blue lines) to detect the presence of TTR D) Diagrams illustrating differences between the slopes of two TTRs for two experiments.



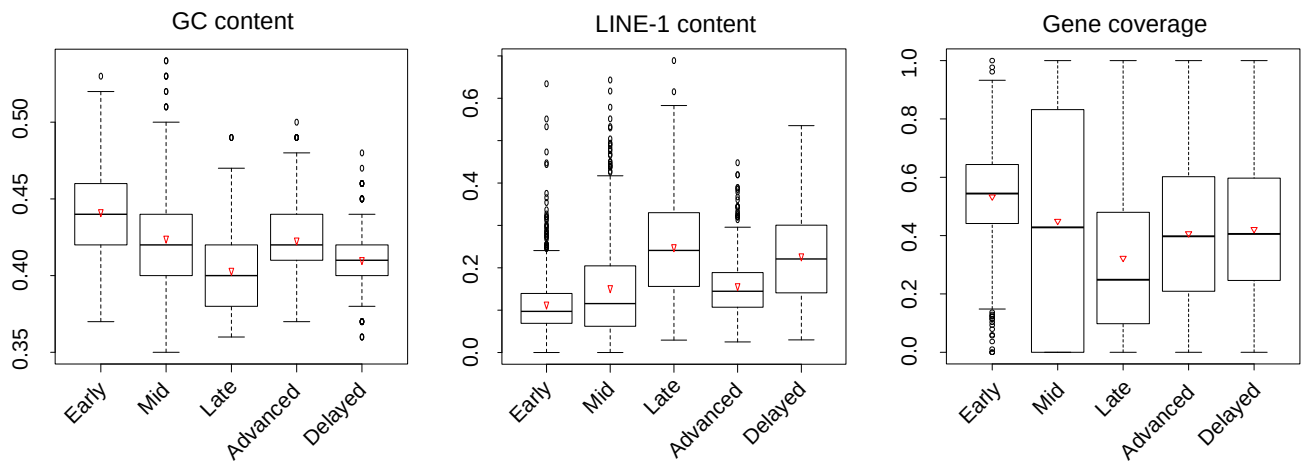
Supplementary Figure S4: Boxplot of squared difference of intensities computed by the Euclidean method.

Squared differences of the $\log_2(\text{Early/Late})$ for the same probes for two different experiments are plotted in a boxplot. The outliers (open circles) are significantly different and correspond to the red line in the graph depicting the squared difference of intensities versus their occurrences in the boxplot.



Supplementary Figure S5 : Examples of data generated with our RT protocol visualized by START-R-Viewer.

A) START-R Viewer allows visualizing RT data with many features. The top panel displays the distribution of early and late constant timing regions (CTR, in red and green, respectively) and of transition timing regions (TTR, in yellow) on a portion of human chromosome 19. Segments corresponding to regions of constant timing are shown in purple. Chromosome 19 centromere is indicated by grey dashed lines and a curly bracket. The bottom panel displays a zoom of a smaller region of chromosome 19 where timing profile can be seen through the zoom option of START-R-Viewer. B) Differential analyses are done on a portion of human chromosome 14 comparing RT profiles of two cell lines: K562 in blue and U2OS in red. Advanced (green) and Delayed (pink) regions are identified with START-R Analyzer using the mean comparison analysis with the Holm's p-value correction and a limit corrected p-value of 0.05. Light grey and grey spots indicate data from both RT experiments.



Supplementary Figure S6: Genomic characteristics of regions harboring different replication timing programs.

Using START-R Analyzer with the standard options, 2,066 CTRs were detected in the genome and 910 regions showed a different replication timing between D3esc and D3npc9 cells. A portion of chromosome 7 is depicted in Figure 2. START-R Analyzer automatically generates BED files that can be imported into a GALAXY session in order to generate complementary results by using the “Coverage of a set of intervals on second set of intervals” software. Boxplots illustrating differences in GC content, LINE-1 content and gene coverage between Early, Mid and Late replicating regions are shown. The two other categories show the characteristics of Advanced and Delayed regions. For each category, the mean value is indicated by an open red triangle. The band at the middle of the box indicates the median value. The bottom and top of the box are the 25th and 75th percentiles. Bottom and top whiskers represent the limits with exclusion of outliers (open circles). Advanced regions showed a GC/LINE-1 density and gene coverage similar to constant early regions, while Delayed regions showed GC/LINE-1 density and gene coverage similar to constant late regions. The genomic regions detected with the START-R suite showed the same specific molecular signatures as the ones previously described in Gilbert’s laboratory. Unfortunately, we cannot compare the regions that we have identified with START-R with those previously discovered, since the article of Hiratani *et al* does not mention the genome coordinates.