SUPPLEMENTARY FIGURES

Supplementary Figure 1: Biological replicates of nascent strands isolated from mouse and *Drosophila* cells and confirmation of the microarray data by qPCR

(A) Box plots showing that mouse chromosome 11 is comparable to the entire mouse genome concerning replication timing (left panel) and transcription activity (right panel). **(B)** Alignment of four entirely independent biological replicates of microarray data for P19 cells, and **(C)** representative Scatter Plots with computation of the Pearson correlation (R^2) of two biological replicates. **(D)** Alignment of two entirely independent biological replicates of microarray data for *Drosophila* Kc cells and **(E)** representative Scatter Plots of two biological replicates. **(F)** Nascent strand (NS) preparations from mouse cells were validated using a known origin by qPCR with different sets of primers localized along the *Myc* gene. **(G)** Microarray data for mouse cells were confirmed at the *Hoxb* locus by qPCR with different sets of primers localized along the locus.

Supplementary Figure 2: Confirmation of microarray data by qPCR

(A) qPCR confirmation of the *Histone* gene repeat origins in *Drosophila* Kc cells. Different sets of primers localized along various loci of the *Drosophila* (**B)** or mouse **(C**) genome were used for qPCR measurements of nascent strand enrichment.

Supplementary Figure 3: Association of replication origins with ORC

Immunoprecipitation of chromatin associated with ORC2 was carried out in P19 cells as described in Methods. **(A)** DNA fragments were analyzed by microarrays and validated by qPCR at the *Myc* gene. **(B)** Alignment of origins and ORC2 sites on a representative region in P19 cells.

Supplementary Figure 4: Common origins in the three mouse cell lines

(A) The percentage of replication origins' overlap in the different mouse cells is shown. **(B)** Common origins in the three mouse cell lines. Origins conserved between two mouse cell lines where compared to each other. The proportion of conserved origins between ES and P19 cells was significantly higher than between ES and MEF or MEF and P19 cells.

Supplementary Figure 5: Distribution of origins along genes

(A) Intragenic or intergenic distribution of origins. **(B)** Origins are enriched at gene promoters and exon sequences (*=$p<0.001$) compared to randomized data sets (dashed white boxes) in P19 (left panel) and MEF cells (right panel).

Supplementary Figure 6: Origins in MEF and P19 cells are frequently associated with TSS and CGI

Patterns of NS strength at TSS in P19 **(A)** and MEF cells **(B). (C)** Association of origins with TSS which contain or not CGI in P19 **(C)** and MEF **(D)** cells. **(E)** Origins found by microarrays are highly associated with CGI in MEF **(E)** and P19 cells **(F).** The percentage of the CGI-origin association is also shown.

Supplementary Figure 7: Analysis of bimodal origins located at TSS in mouse cells.

The NS profile of individual TSS associated with an origin was examined. Examples of NS profiles for each class of TSS are shown. TSS were scored as bimodal if the $\log_2$-ratio increased both upstream and downstream of the TSS.

Supplementary Figure 8: Nucleotide asymmetry of origins in mouse cells

Nucleotide composition along a 3 kb region (A-green, T-red, G-black and C-blue) centered on the origin peaks in mouse ES **(A)**, P19 **(B)** and MEF **(C)** cells. In these cells lines, an asymmetric distribution of G/T versus A/C is observed, like in *Drosophila* Kc cells.

Supplementary Figure 9: Hierarchical organization of origins in *Drosophila* Kc cells and mouse ES cells

The grey profile is the distribution of inter-origin distances obtained by DNA combing of ES **(A)** and Kc **(B)** cells. The red line represents the simulated distribution of inter-origin distances according to each model. The 'Flexible Replicon' model is the only to yield a simulated distribution of inter-origin distances that is statistically indistinguishable from that obtained from DNA combing data for these cell lines.

Supplementary Figure 10: Characterization of the models of origin organization in metazoans

(A) Distribution of fork speed (measured by DNA combing) in mouse and *Drosophila* cell lines (see Methods). **(B)** Simulated origin firing efficiency in the 'Increasing origin efficiency' model for ES cells. Note the increase in firing efficiency as replication takes place. Similar profiles were obtained for MEF and Kc cells. **(C)** Dendrogram illustrating how origin clusters (replicons) were defined in the 'Flexible Replicon' model. Origins were grouped based on their closeness along the chromosome. Clusters were defined by cutting the tree at a specific height. Shown are three cluster generations based on different height (h) cuts. Clustered origins are highlighted. **(D-E)** Clusters were exhaustively generated by cutting the tree every 1000 steps. For every cluster generation, origin firing was performed (100 simulations). The simulated inter-origin distribution was compared with the DNA combing data and a p-value was calculated with the Kolmogorov-

Smirnov test. The p-values were plotted in function of the cutting height. A cubic smoothing spline function was applied to the data (grey curve). The significance value (p=0.05) is indicated. The minimal (h_{min}), optimal (h_{opt}) (where the simulated inter-origin distribution is not statistically different from the DNA combing data) and maximal (h_{max}) cutting heights are highlighted. **(F)** Statistics on the clusters generated in the Flexible Replicon model. The average number of origins/cluster, length of clusters and the inter-cluster distance are indicated for ES, MEF and Kc cells for the optimal cutting height (h_{opt}). The values in brackets are for the h_{min} and h_{max} clusters.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Description of the genomic features

Gene databases were Flybase (for *Drosophila*) and RefSeq (for *Mus Musculus*). CGI were defined as a DNA region of at least 200 pb with a GC content greater than 60% and the (observed CpG/expected CpG) ratio equal to or greater than 0.6 (classical definition of a CpG island, (Gardiner-Garden and Frommer 1987). *Drosophila* HP1 binding sites were determined from DamID data (de Wit et al. 2007). Replication timing data for Kc and ES cells were from (Schwaiger et al. 2009) and (Hiratani et al. 2008) respectively. Divergent transcription start sites (TSS) used for ES cells were described in (Sailo et al. 2008).

Nascent Strand-ChIP Data Analyses

Microarray Design

Drosophila melanogaster samples were hybridized using 2.1M Nimblegen microarrays (Design ID 6262). These tiling arrays contain in total 2,164,511 oligonucleotide probes representing the non-repetitive regions of the *Drosophila melanogaster* genome (chromosome 2L, 2R, 3L, 3R, 4 and X; Flybase release 4.3).

To analyze the data, 1,807,015 oligonucleotide probes were selected (909,279 for the top strand and 897,736 for the bottom strand) with an average length of 50 bp for oligonucleotides and for inter-oligo spacing. All the processed data were generated using the BDGP/Flybase release 4 of the *Drosophila melanogaster* genome assembly (UCSC dm2, April 2004).

Mouse samples were hybridized using the Nimblegen 389K tiling arrays (Design ID 4095) which cover 60.4 Mb of non-repetitive DNA sequences in chromosome 11 (56.6-117 Mb). In total, 385,496 probes were analyzed with an average coverage of one 50 bp-probe each 100 bp. All the

processed data were generated using the UCSC mm8 (NCBI Build 36, February 2006) of the *Mus musculus* genome assembly.

Correlation between biological replicates

The degree of correlation between biological replicates was evaluated using a scatter plot and computing the Pearson's Correlation Coefficient (R^2).

Data normalization and determination of significant probes

Experimental (Cy5) and control (Cy3) signal intensities quantified and provided by Nimblegen were converted into $\log_2$-ratios ($\log_2$ (Cy5/Cy3)). The Lowess normalization method was applied to eliminate intensity-dependent variations in dye bias (Yang et al. 2002). A sliding median window with a length of 5 oligonucleotide probes was used to smooth the signal. Mode (m) and s (median absolute deviation) of normalized $\log_2$-ratios were computed. Assuming that the normal distribution (specified by m and s) covered the entire background noise (non-significant signals), for each probe, one p-value was computed by applying the false discovery rate (FDR) correction (Benjamin and Hochberg 1995). Two biological independent samples for *Drosophila*, four independent samples for P19 cells, three independent samples from ES and three for MEF cells were used. Normalized $\log_2$-ratios of replicate samples were combined by averaging the values at the corresponding genomic positions and the corrected p-values were combined using a Chi-Square distribution (Fisher 1932). Thus, one probe was denoted as significant if the combined p-value was lower than 5% (level of significance).

Origin definition

The minimum size of purified NS is 0.5 kb. Thus, potential Oris should be at least 1 kb (2 X 0.5 kb for a bidirectional origin). We defined Oris as regions that have at least one significant probe ($p<0.05$ with FDR correction) in an area containing a minimum of 10 consecutive positive probes (showing NS enrichment with a $\log_2$-ratio>0). For *Drosophila* cells, two significant probes

(because they are twice denser in *Drosophila* than in mouse Chips) and at least ten consecutive positive probes should be detected. If two enriched regions were separated by <1 kb, they were merged into one. These conditions were used to minimize false positive events by excluding over-hybridization signals of single probes or small regions, and to score as Oris only sites with multiple consecutive positive values.

Comparative analysis of Oris and genome features

For each profile (*Drosophila* and mouse cells), 1000 bootstrap samples of random Oris were generated. Random Oris contained the same number of Oris with the same length, but each origin segment was randomly picked in the chromosome region with the condition that the segments did not overlap. For each profile and each studied genome feature (CGI or CGI-like, TSS, etc …), one permutation test with theoretical expectation under a null hypothesis was performed from the 1000 random Oris to compute the statistical significance of the Ori positions relative to the studied genome feature.

NS signal strength around specific features

For each profile (*Drosophila* and mouse cells), specific feature positions (TSS, middle of CGI or CGI-like) were taken as 'Local center' (Lcent). For each nucleotide position around every Lcent (Lcent - 5 kb to Lcent + 5 kb), p-values (previously calculated) were retrieved. P-values were merged in a matrix (rows representing the nucleotide coordinate/position and columns representing Lcent). The strand was also considered. Thus, for TSS from the minus strand, nucleotide positions and associated p-values were reversed. To obtain only one overall p-value distribution around the set of Lcent, p-values were combined using a Chi-Square distribution (Fisher 1932). To visualize the combined p-value distribution around the specific features, results were plotted using the transformation '-log(p-value)' and labeled as 'NS signal strength'.

Analysis of Bimodal TSS in mouse cells

For both upstream and downstream regions (TSS -2 kb to TSS + 2 kb) of each TSS overlapping one Ori, the highest '-log(p-values)' (noted '$-\log(\text{p-value})_{upstream}$ and '$-\log(\text{p-value})_{downstream}$') were retrieved. In the same way, the lowest '-log(p-value)' (noted '$-\log(\text{p-value})_{middle}$') around each TSS (TSS -0.1 kb to TSS + 0.1 kb), was collected. Note that, high '-log(p-values)' corresponds to high $\log_2$-ratio of NS/total genomic DNA. In this analysis, the orientation of TSS was considered.

Four classes were created:

- If the '-log(p-value)' increased both upstream and downstream of the feature, the TSS was scored as bimodal.

  More precisely, this category corresponds to TSS in which:

  $-\log(\text{p-value})_{upstream} > -\log(\text{p-value})_{middle}$ and $-\log(\text{p-value})_{downstream} > -\log(\text{p-value})_{middle}$
- If the '-log(p-value)' increased only upstream of the feature, the TSS was scored as unimodal with NS enrichment at the 5' of the feature. More precisely, this category corresponds to TSS in which:

  $-\log(\text{p-value})_{upstream} > -\log(\text{p-value})_{middle}$ and $-\log(\text{p-value})_{downstream} \leq -\log(\text{p-value})_{middle}$
- If the '-log(p-value)' increased only downstream of the feature, the TSS was scored as unimodal TSS with NS enrichment at the 3' of the feature. More precisely, this category corresponds to TSS in which:

  $-\log(\text{p-value})_{upstream} \leq -\log(\text{p-value})_{middle}$ and $-\log(\text{p-value})_{downstream} > -\log(\text{p-value})_{middle}$
- Otherwise, TSS was associated with one Ori exhibiting a more symmetrical NS profile around the TSS.

Sequence distribution around specific regions

For each profile (*Drosophila* and mouse cells), the sequence distribution was centered on the middle of the CGI or CGI-like regions associated with Oris and taken as the 'Local center' (Lcent). The 3-kb sequence around each Lcent (Lcent – 1.5 kb to Lcent + 1.5 kb) was retrieved. The resulting sequences were merged in a matrix (rows representing nucleotide coordinate/position and columns representing Lcent). The number and the percentage of A/T/C/G nucleotides were computed. Results were plotted using a sliding mean window to fit the signal.

The same analysis was performed to represent the sequence distribution centered on the probe with maximum intensity for Oris.

Organization of Oris

Computer simulations were performed to model Ori organization. For each model (Random, Increasing Firing Efficiency and Flexible Replicon), inter-origin distances from 100 simulations were calculated. Importantly, the firing density (e. g. the number of activated Oris/Mb) was identical to the density observed in DNA combing experiments. The simulated inter-origin distribution was compared with the inter-origin distribution of DNA combing data by calculating the p-value with the Kolmogorov-Smirnov test (Massey 1951). A high p-value (p>0.05) indicates that the two distributions cannot be considered as statistically different. The different models were evaluated as follow.

1) *Random Ori firing model*

   In the random model, Oris are fired in a purely stochastic manner. In this model, firing efficiency is supposed to be constant. The inter-origin distances for each of the 100 simulations were calculated.

2) *Increasing Ori efficiency model*

This model is based on the hypothesis that Ori firing efficiency increases during S-phase progression (Rhind 2006). Also, the advancing replication fork passively suppresses replicated Ori regions. During each cycle (simulation of time), one Ori is randomly selected. The resulting bidirectional replication fork was simulated using the mean fork speed obtained from DNA combing experiments. The duration of each cycle was optimized to achieve a firing efficiency identical to the one of the single DNA molecule experiments. The model stops when the entire DNA is replicated. Inter-origin distances between fired origins from 100 simulations were collected. The simulated Ori firing efficiency was also calculated.

3) Flexible Replicon Model

This hierarchical clustering model is based on the hypothesis that Oris are functionally grouped and that activation of one Ori suppresses the firing of other Oris within the same group. In this model, Ori firing is randomly selected and the firing efficiency is supposed constant. The steps to obtain groups of Oris, called clusters, are described below.

First, Oris were classified using hierarchical cluster analysis with Euclidean distance as the distance metric to determine how the similarity of two elements was calculated, and average linkage clustering to determine the distance between sets of observations (Brian et al. 2001). Then, by cutting the dendrogram at different heights different clusters were defined. For each height cut, 100 simulations were collected and the distribution of inter-origin distances was compared with DNA combing data. The range of selected height cuts corresponded to heights where the p-values were the highest ($p>0.05$).

Precisely, the retained clusters of Oris were obtained by cutting the dendrogram at the optimum height of 26,560 bp for *Drosophila,* 66,374 bp for mouse ES cells and 71,032 bp for

mouse MEF cells. For each profile, cluster characteristics (inter-cluster distance, cluster length, etc.) were calculated at the optimum height.

Density of origins and other genome features

Density analysis was used to compare specific data distribution along the genome at large scale. The coordinates of the specific regions and the genome positions were retrieved. Each nucleotide inside specific regions was flagged as 1 (if belonging to one Ori) and 0 (if not belonging to one Ori). A sliding window was used to compute the frequency of data per window.

For each profile (*Drosophila* and mouse cells), the window size was based on the optimal height cut from the hierarchical cluster of Oris.

Conserved regions

To test whether Oris were in conserved regions, conservation scores were downloaded for alignments of 14 insect genomes with the *Drosophila melanogaster* genome and 16 vertebrate genomes with the *Mus Musculus* genome from the UCSC Website. Conservation data were divided in two groups called "inside origins" and "outside origins". The Wilcoxon-Mann-Whitney (Mann and Whitney 1947) test was used to determine whether the conservation scores between the two groups were significantly different.

Comparison of the Ori repertoires in mouse cells

Considering as reference the Oris from P19 cells, two proportions of “common Oris” were calculated: “common Oris” between P19 and ES cells and “common Oris” between P19 and MEF cells. Then, the difference between these two proportions (Newcombe 1998) was tested by

computation of the p-value (p). A p-value <0.05 indicates that the two proportions are significantly different. The same analysis was carried out considering as reference the Oris from ES and then from MEF cells as well.

Comparison of Ori coverage in early and late replication timing regions

For each profile of mouse cells, Ori coverage in early and late replication timing domains was calculated. To compare the coverage values, a test of difference between the two groups was performed (Newcombe 1998).

Software

All Nascent Strand-ChIP data analyses were carried out using the software R, version 2.11.1 (www.R-project.org) (R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0).

Cell Cycle Analysis

For cell cycle analysis a Beckman Coulter flow cytometer was used. Cells were fixed with 70% ethanol in PBS at -20C° for at least 20 min. After one wash in PBS, cells were incubated in propidium iodide (PI, Sigma) at room temperature for 30 min before treatment with DNase-free RNase A (Sigma).

DNA Combing

Cells were sequentially labeled with iodo-deoxyuridine (IdU) and chloro-deoxyuridine (CldU). Asynchronous cell populations were first labeled with 40 mM IdU for 20 minutes and then with

40 mM CldU for another 20 minutes, without intermediate wash. Cells were then washed with 1x phosphate-buffered saline (PBS), trypsinized, pooled, counted and 100 000 cells were resuspended in 100 ml of 1x PBS with 1% low-melting agarose in order to make agarose plugs with imbedded cells. Plugs were incubated in 0.5 ml 0.5 M EDTA with 1% N-lauryl-sarcosyl and 1 mg/ml proteinase K and incubated at 50°C for 2 days (fresh solution added after the first day). Complete removal of digested proteins and other degradation products was performed by washing the plugs in 0.5M EDTA and TE buffer several times. Protein-free DNA plugs were then stored in 0.5 M EDTA at 4°C or used immediately for combing. Agarose plugs were stained with YOYO-1 fluorescent dye (Molecular Probes) in TE buffer for 2 h, washed with TE buffer, resuspended in 100 µl of TE buffer and melted at 65°C for 15 minutes. The solution was maintained at 42°C for 15 minutes and treated overnight with agarase (New England Biolabs). After digestion, 4 ml of 50 mM MES (2-(*N*-morpholino)ethanesulfonic acid, pH 5.7) were added very gently to the DNA solution and then combing of DNA fibers on silanized cover slips was performed as described (Michalet et al. 1997). Combed DNA was denatured in 1N NaOH for 20 minutes and washed several times in PBS. After denaturing, silanized cover slips with DNA were blocked with 1% BSA in PBS, 0.1% Triton X100. Immunodetection was done with antibodies diluted in PBS, 0.1% TritonX100, 1% BSA and incubated at 37°C in a humid chamber for 30 min. Each step of incubation with antibodies was followed by extensive washes with PBS. Immunodetection was with a mouse anti-BrdU antibody (1/50 dilution, Becton Dickinson) and a rat anti-BrdU antibody (1/25 dilution, Sera Lab) that recognize the IdU and CldU tracks, respectively, goat anti-rat antibody coupled to Alexa 488 (1/50 dilution, Molecular Probes), goat anti-mouse IgG1 coupled to Alexa 546 (1/50 dilution, Molecular Probes), anti-ssDNA antibody (1/100 dilution, Chemicon) and goat anti-mouse IgG2a coupled to Alexa 647 (1/50 dilution,

Molecular Probes). Cover slips were mounted with 20 µl of Prolong Gold Antifade (Molecular Probes), dried at room temperature for 2 hr and processed for image acquisition using a fully motorized Leica DM6000B microscope equipped with a CoolSNAP HQ CDD camera and controlled by MetaMorph (Roper Scientific). Images were acquired with a 40x objective: 1 pixel was equal to 340 bp. Inter-origin distances were measured manually using MetaMorph. Statistical analysis of inter-origin distances was performed with Prism 5.0 (GraphPad).

ORC2 ChIP on Chip and qPCR analysis.

Briefly, approximately 1.5×10^8 P19 cells were treated with 100 ng/ul nocodazole for three hours and seeded after shaking off, followed by three washes with PBS. After 30 min (cells in G2\M phase by flow cytometry analysis), cells were cross-linked by adding fresh 0.5% paraformaldehyde solution to the medium at 37°C for 15 minutes. Paraformaldehyde was neutralized by adding 250 mM glycine at room temperature for 10 min. Cells were washed twice with 1× phosphate-buffered saline (PBS), scraped off the plates, and nuclei were isolated with NE buffer (50 mM HEPES at pH 7.6, 350 mM sucrose, 0.1% Tween20, 5 mM MgCl2, 1 mM EDTA and protease inhibitors). After centrifugation, nuclei were lysed in 1 ml RIPA buffer (50 mM HEPES at pH 7.6, 150 mM NaCl, 0.3% SDS, 0.5% NaDoc, 1% TritonX100 and protease inhibitors) and sonicated into fragments ranging from 300 to 1000 bp using the Bioruptor (Diagenode). The chromatin solution was clarified by centrifugation at 15 000g at 4°C for 5 min. The supernatant was pre-cleared with 50 µl of Dynabeads protein A for 2 h at 4°C. Pre-cleared chromatin was separated in two fractions and incubated at 4°C overnight with 50 µl of Dynabeads protein A, blocked with 0.05% bovine serum albumin/PBS and pre-incubated with 30 µg of ORC2 antibody (home-made with recombinant mouse ORC2) or with 30 µg of pre-immune

antibody (from the same rabbit used for generating the ORC2 antibody, but before injection) for 2 h. After extensive washing with RIPA buffer, cross-linking of each immune complex was reversed by incubation of the eluate at 65 °C in 50 mM Tris pH 8, 1% SDS, 10 mM EDTA overnight. After digestion with RNaseA at 37°C for 1 h and proteinase K at 50°C for 2 h, DNA was purified by phenol–chloroform extraction and precipitated with ethanol. The amount of DNA in the immunoprecipitates and in the input was quantified by real-time PCR with primers localized along the *Myc* gene and promoter. ChIP data are reported as the percentage of the total input that was immunoprecipitated. Quantitative PCR was performed on a Roche LightCycler 480 machine using LightCycler® 480 SYBR Green I Master (Roche). DNA from immunoprecipitates was amplified using the WGAII kit (Sigma). Amplification products were purified with NucleoSpin columns (Machrey Nagel). Hybridization, washing and scanning of microarrays were done by the Nimblegen Service Laboratory. For this experiment, the Nimblegen 389K tiling arrays (Design ID 4095) were used. ChIP on chip signals were analyzed in the same manner as the data from hybridization with nascent strands (see above).

Supplementary References

Benjamin Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *JRSSB* 57: 289-300.

Brian S, Everitt, Landau S, Leese M. 2001. Cluster Analysis 4th Edition. *Oxford University Press*.

de Wit E, Greil F, van Steensel B. 2007. High-resolution mapping reveals links of HP1 with active and inactive chromatin components. *PLoS Genet* 3(3): e38.

Fisher RA. 1932. Statistical Methods for Research Workers. *Oliver and Boyd* London.

Gardiner-Garden M, Frommer M. 1987. CpG islands in vertebrate genomes. *J Mol Biol* 196(2): 261-282.

Hiratani I, Ryba T, Itoh M, Yokochi T, Schwaiger M, Chang CW, Lyou Y, Townes TM, Schubeler D, Gilbert DM. 2008. Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol* 6(10): e245.

Mann HB, Whitney DR. 1947. On a test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics* 18(1): 50-60.

Massey FJ. 1951. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association* 46(N° 253): 68-78.

Michalet X, Ekong R, Fougerousse F, Rousseaux S, Schurra C, Hornigold N, van Slegtenhorst M, Wolfe J, Povey S, Beckmann JS et al. 1997. Dynamic molecular combing: stretching the whole human genome for high-resolution studies. *Science* 277(5331): 1518-1523.

Newcombe RG. 1998. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Stat Med* 17(8): 873-890.

Rhind N. 2006. DNA replication timing: random thoughts about origin firing. *Nat Cell Biol* 8(12): 1313-1316.

Schwaiger M, Stadler MB, Bell O, Kohler H, Oakeley EJ, Schubeler D. 2009. Chromatin state marks cell-type- and gender-specific replication of the Drosophila genome. *Genes Dev* 23(5): 589-601.

Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP. 2002. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 30(4): e15.