EMBO reports

scientific report

High-throughput mapping of origins of replication in human cells

Isabelle Lucas[1], Aparna Palakodeti[1], Yanwen Jiang[1], David J. Young[1], Nan Jiang[2], Anthony A. Fernald[1] & Michelle M. Le Beau[1+]

[1]Department of Medicine, Section of Hematology/Oncology, The University of Chicago, Chicago, Illinois, USA, and [2]NimbleGen Systems Inc, Madison, Wisconsin, USA

Mapping origins of replication has been challenging in higher eukaryotes. We have developed a rapid, genome-wide method to map origins of replication in asynchronous human cells by combining the nascent strand abundance assay with a highly tiled microarray platform, and we validated the technique by two independent assays. We applied this method to analyse the enrichment of nascent DNA in three 50-kb regions containing known origins of replication in the *MYC*, lamin B2 (*LMNB2*) and haemoglobin β (*HBB*) genes, a 200-kb region containing the rare fragile site, *FRAXA*, and a 1,075-kb region on chromosome 22; we detected most of the known origins and also 28 new origins. Surprisingly, the 28 new origins were small in size and located predominantly within genes. Our study also showed a strong correlation between origin replication timing and chromatin acetylation.

Keywords: genome-wide origin mapping; high-density microarray; human cells

EMBO reports (2007) **8**, 770–777. doi:10.1038/sj.embor.7401026

INTRODUCTION

Mapping the origins of replication in higher eukaryotes has proved to be a challenge because of the size and complexity of their genomes, and the absence of sequence specificity at the origins. Several techniques have been developed to map origins in eukaryotes based on the detection of events present only at the origin, such as the presence of short, newly replicated single-stranded DNA (Giacca *et al*, 1994; reviewed by Hamlin & Dijkwel, 1995). However, major limitations of each of these techniques—for example, the fact that they are labour-intensive and can be applied only to small regions of the genome—have limited their widespread application. Thus, after two decades of intense efforts, only a small number of origins have been mapped in metazoans (reviewed by Gilbert, 2001; Machida *et al*, 2005).

The availability of the genome sequence of several species, together with the development of DNA microarray technology, has facilitated the analysis of replication dynamics and the mapping of origins at a genome-wide level. However, the available methods are indirect assays, they involve cell-cycle synchronization, which might not be feasible in all cell types and requires the use of replication inhibitors that might alter origin usage (Anglana *et al*, 2003), or they require PCR amplification steps that might introduce bias, and most are labour-intensive and time-consuming (Mesner *et al*, 2006; reviewed by MacAlpine & Bell, 2005). By combining the nascent strand assay with a tiled microarray platform, we have developed a rapid, non-PCR based, high-throughput approach to map active origins in asynchronous human cells (Fig 1). We also developed a peak finding method that is applicable to any microarray-based genome-wide mapping study in which discrete regions are expected. We used our origin-mapping method to analyse three 50-kb regions, *MYC*, lamin B2 (*LMNB2*) and haemoglobin β (*HBB*), which contain known origins (Vassilev & Johnson, 1990; Kitsberg *et al*, 1993; Trivedi *et al*, 1998; Abdurashidova *et al*, 2000; Kamath & Leffak, 2001), a 200-kb region containing the *FRAXA* rare fragile site (Hansen *et al*, 1993) and a 1,075-kb region of chromosome 22 for which the replication timing has been studied in detail (White *et al*, 2004). We validated our results by two independent assays, thereby showing that this new genome-wide origin-mapping approach is robust.

RESULTS AND DISCUSSION

Assaying hybridization quality

To map the origins of replication at a genome-wide level, we designed a highly tiled microarray platform covering both the forward (FWD) and reverse (REV) DNA strands of five human regions (supplementary Table 1 online). For statistical purposes, each probe was duplicated on the array to create four data sets: FWD1, FWD2, REV1 and REV2. To summarize our array design, 93% of the five regions were covered at the highest density with a 20-nt overlap between probes; for the remaining 7%, the gaps were 507 ± 864 nt in length (range 31–9,817 nt).

[1]Department of Medicine, Section of Hematology/Oncology, The University of Chicago, 5841 South Maryland Avenue, MC2115, Chicago, Illinois 60637, USA
[2]NimbleGen Systems Inc, 1 Science Court, Madison, Wisconsin 53711, USA
[+]Corresponding author. Tel: +1 773 702 0795; Fax: +1 773 702 9311;
E-mail: mlebeau@medicine.bsd.uchicago.edu

Received 13 December 2006; revised 18 May 2007; accepted 4 June 2007; published online 13 July 2007

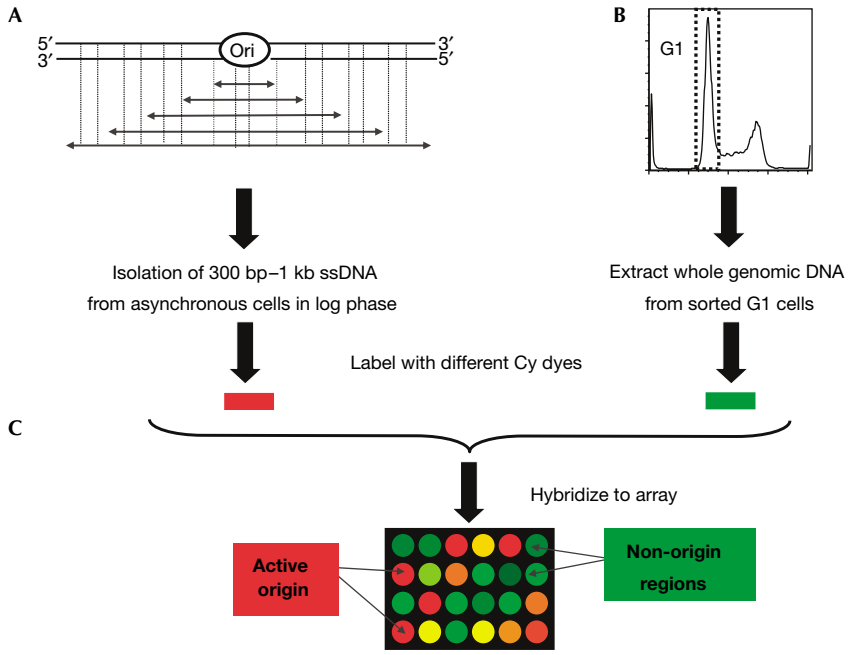

Fig 1 | Experimental design for mapping active origins of replication using microarrays. (**A**) Short single-strand DNA was isolated by neutral sucrose gradient centrifugation of denatured whole genomic DNA. (**B**) Reference genomic DNA was obtained by isolating whole genomic DNA from sorted G1 cells. (**C**) The nascent strand and G1 DNAs were labelled with the dyes Cy-3 and Cy-5, respectively, combined and hybridized to the microarray. A red spot on the microarray indicates that this DNA fragment is enriched in the nascent strand sample in comparison to the G1 DNA sample, suggesting the presence of an active origin. Conversely, a green spot indicates that this DNA is under-represented in the nascent strand sample, suggesting that the locus does not contain an active origin. Ori, origin of replication; ssDNA, single-stranded DNA.

We used a *Z*-score transformation to assess the reproducibility of hybridization across each array. This analysis allowed us to identify the probes for which the quality of the data was not sufficient, that is, poor quality of local hybridization or inconsistent probe behaviour. Whenever the *Z*-score value of a probe and its duplicate was above the chosen cutoff of 3, the two probes were excluded from the analysis. We observed a similar distribution of *Z*-score values between the four sets of data for the five regions (see supplementary Fig 1A online), indicating that both strands behaved in a similar manner. Finally, approximately 98% of the data satisfied our standards of reproducibility and were included in subsequent analyses (supplementary Fig 2B online).

Development of the origin peak finding method

To identify peaks representing potential origins, we used a 'sliding window' method, which was based on three parameters: the size of the window; the cutoff of the $\log_2$ ratio—that is, nascent DNA over G1 DNA hybridization signals; and the threshold percentage of probes within the window that were above the $\log_2$ ratio cutoff. To assess the likelihood of any peak as representing a true origin, we estimated the false-positive rate (FPR) for each peak that was identified. Peaks with an FPR of more than 10% were considered to be false-positives and were not included in the remaining analyses. To enhance the confidence level of the origin-finding method, we systematically increased the threshold percentage of probes within the selected sliding window of 625 nt from 4 out of 20 (20%) to 20 out of 20 (100%) probes (Fig 2). Finally, a peak must be present, with an FPR of 10% or less in at least three out of the four data sets analysed at any cutoff between 40% and 100% to be designated as a putative origin.

We analysed the distribution of peaks with an FPR of 10% or less and present in one, two, three or four of the data sets, for each of the thresholds used in the analysis (40–100%; supplementary Fig 2 online). Almost 60% of the peaks were present in all four data sets at the 40% threshold in two independent nascent strand isolations and hybridisations; we did not detect any strand bias for the peaks detected in fewer than three data sets.

The next step was to merge the data obtained for the four data sets for each percentage of qualifying probes (40–100%) into a set of shared putative origins. Given the size of the nascent strand fragments used for the experiment (300–1,000 nt), we elected to merge any two peaks located less than 2 kb apart into a single-origin region. We used this '2-kb-rule' to determine the position of the origins for each data set, and then for the shared peaks between the four data sets of an experiment (supplementary Fig 3 online).

Application of the method to multiple data sets

By analysing the origin-mapping results from two experiments (Fig 3; supplementary Fig 4 online), we identified 85% of the known origins in the *MYC*, *LMNB2* and *HBB* regions, and also several new origins that are located outside the previously studied regions. As predicted, we mapped an origin adjacent to *FRAXA*, which was independently identified while our manuscript was in preparation (Gray *et al*, 2007). The specificity of our method was confirmed further by the fact that we were unable to detect the known origins in hybridizations using shorter nascent fragments

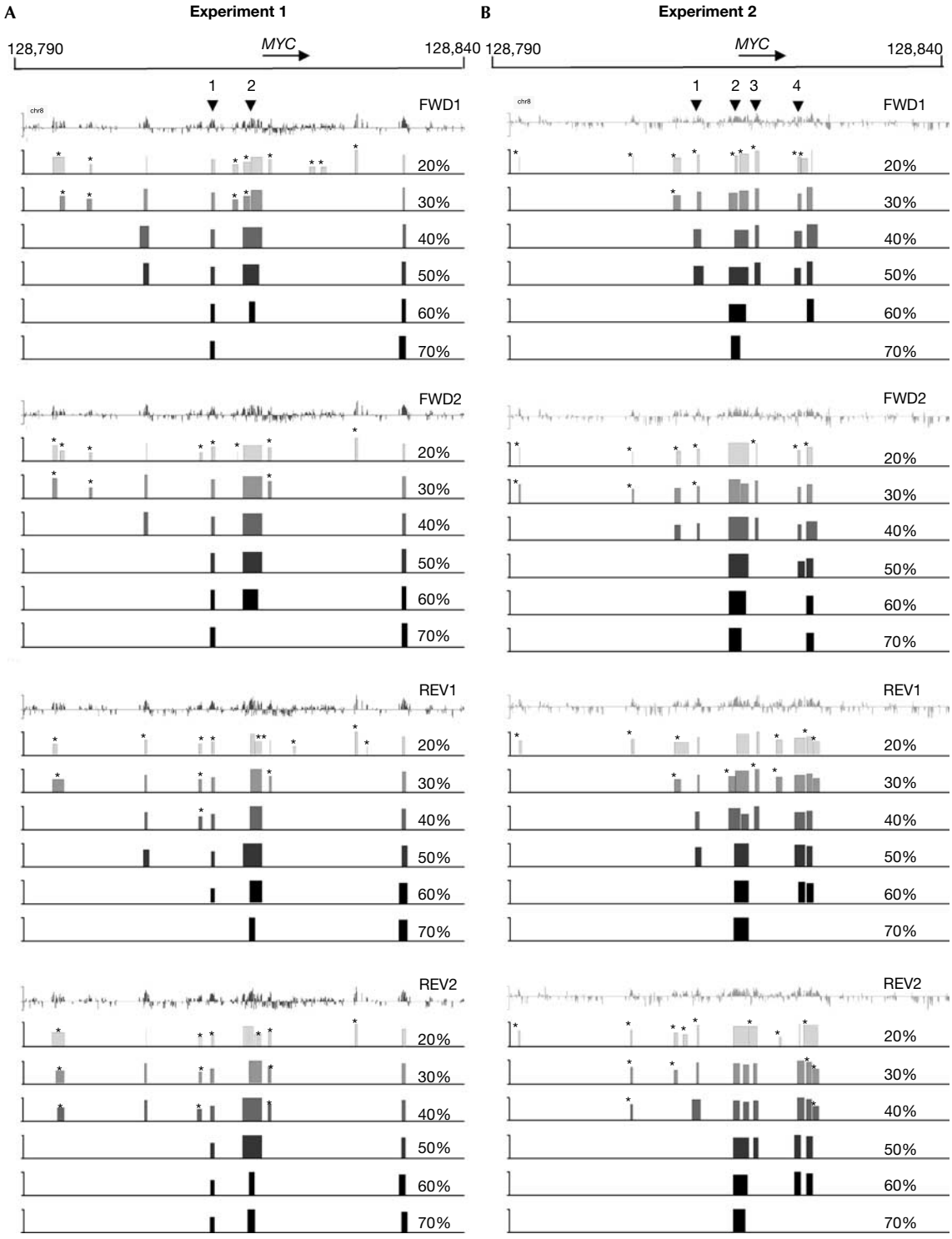

Fig 2 | Origin of replication peak finding method. The analysis is shown for all four data sets of two independent experiments: (**A**) experiment 1 and (**B**) experiment 2 for the *MYC* region. The first line of each data set corresponds to the *Z*-score cleaned data; the next six tracks show the location of the peaks found at each percentage of qualifying probes (20–70%) within the sliding window. The asterisks indicate the peaks with an FPR of more than 10%. The height of each peak corresponds to the maximum $\log_2$ ratio value of the qualifying probes. The arrowheads indicate the location of known origins of replication: (1) STS-EP11, (2) *MYC*, (3) STS-I and (4) STS-J-M (Trivedi *et al*, 1998). FPR, false-positive rate. FWD, forward; REV, reverse.

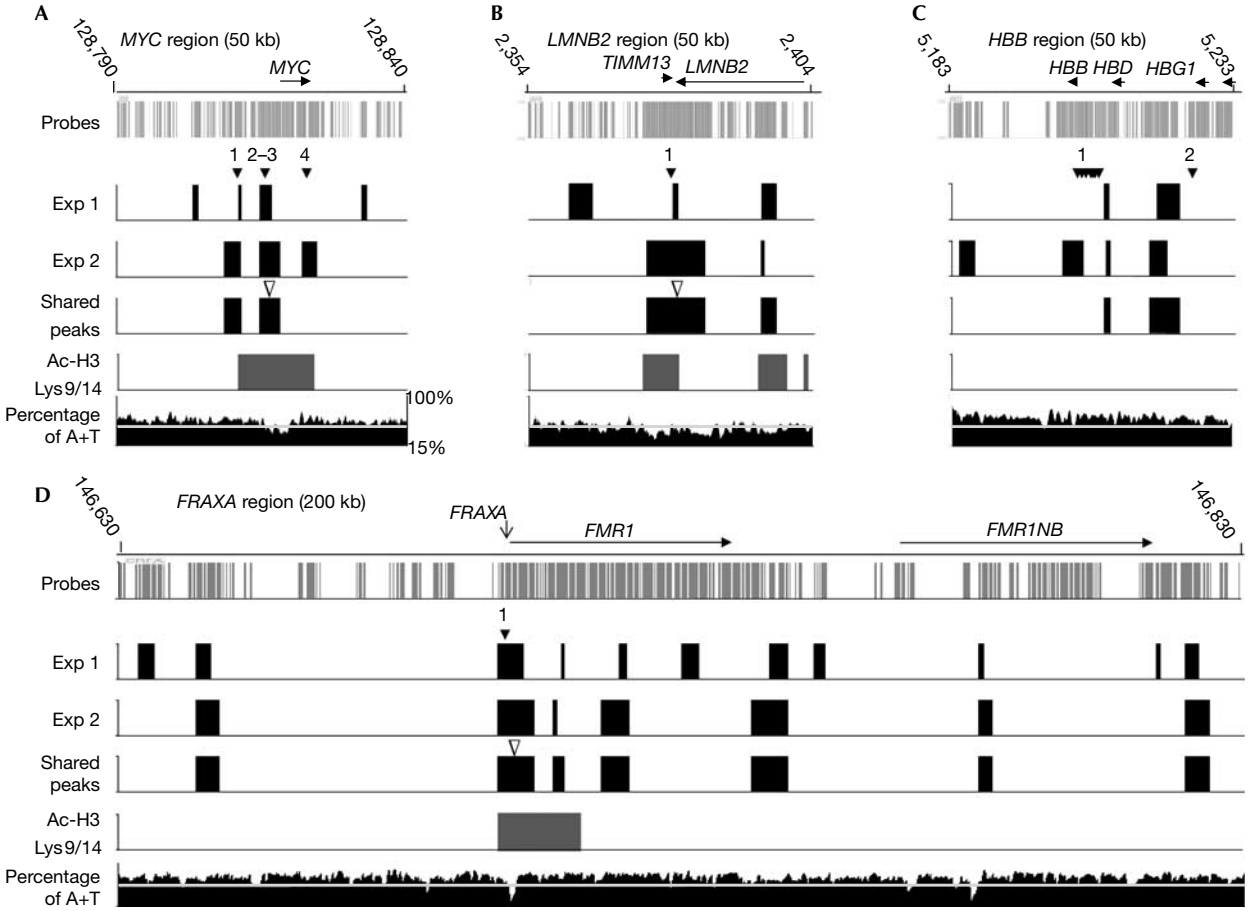

Fig 3 | Origin of replication, chromatin acetylation and A+T content maps of the *MYC*, *LMNB2*, *HBB* and *FRAXA* regions. The origins identified in experiments 1 and 2, and those shared by both experiments, as well as the Ac-H3 Lys 9/14 peaks shared by both chromatin immunoprecipitation experiments are shown. The filled arrowheads indicate the positions of known origins of replication: (**A**) (1) STS-EP11, (2) *MYC*, (3) STS-I and (4) STS-J-M for the *MYC* region; (**B**) (1) the only known origin for the *LMNB2* region; (**C**) (1) and (2) the two mapped origins for the *HBB* region; and (**D**) (1) the only known origin for the *FRAXA* region. The open arrowheads indicate the peaks that were examined using real-time PCR analyses (Fig 4). The A+T content is plotted as a 500 bp sliding window along each region, and a horizontal white line indicates the 50% level.

(smallest fragments less than 100 bp), which are highly enriched for Okazaki fragments and are therefore not specific to origins of replication (supplementary Fig 5 online).

The origins identified in this analysis were distributed into two groups (supplementary Table 2 online). The first category comprises origins that were detected in both experiments; within the 1.4 Mb of the genome examined, we identified a total of 28 new origins in this category. On average, 69±22% of all origins were common in both experiments (Table 1). Origins in this group might correspond to the most efficient origins and/or origins that are less sensitive to experimental variations. The second category includes origins that were detected in only one experiment. As both new and known origins were observed in this group, it is unlikely that all of these origins are false-positives (supplementary Fig 5 online). Surprisingly, none of the putative origins mapped in a 377-kb portion of the chromosome 22 region was found in both experiments. This result could be explained by experimental variations as we observed up to 20% variation when the same DNA samples were used for replicate hybridizations (data not shown) and/or by variations in origin usage, which were influenced by culture conditions. In this regard, it has been proposed that the human genome might contain many potential origins (Machida *et al*, 2005). Thus, within a given region of the genome, the efficiency of origin usage might be variable among cells within a population and also between populations. To evaluate the robustness of our method, we mapped origins within six more regions (1.4–5.4 Mb) of the human genome covering a total of approximately 20 Mb on five different chromosomes. In two independent experiments, we observed 70±7% (61–85%) origin conservation (data not shown). We observed variability similar to the region on chromosome 22. Within four of the six regions examined, specifically, we identified nine comparable regions, ranging from 106 to 357 kb, which represented approximately 9% of the 20 Mb analysed—criteria for inclusion were more than 100 kb with at least two origins per experiment that were not shared. These results indicate that the local variability observed on chromosome 22, and the other genomic regions that we have examined, is more likely to be the result of a biological

Table 1 | Summary of the origin mapping results for the peaks shared in the two independent experiments

Regions	*MYC*	*LMNB2*	*HBB*	*FRAXA*	Chromosome 22
Number of origins	2	2	2	7	19
Size region (kb)	50	50	50	200	1,075
Average distance (kb) between two origins (range)	6	16	10	29±18 (8–54)	56±87 (5–382) 37±33 (5–104)*
Average origin size (kb) (range)	3.2±0.4 (3.0–3.6)	6.5±5.3 (2.7–10.2)	3.3±3.0 (1.2–5.4)	3.8±1.6 (2.4–6.4)	4.0±3.0 (0.9–13.8)
Percentage of origins in shared:					
Experiment 1 (%)	50	67	100	64	51
Experiment 2 (%)	67	100	50	100	45
Average percentage A+T at:					
Newly mapped origin sequences (%)	53±8	37±1	60±0	61±3	50±6
Previously known origins (%)	56±7	57	65±0	35	—
Percentage of early, mid or late firing origins co-localized with H3 Lys 9/14 acetylation					
Early (%)	50	100	50	—	100
Mid (%)	—	—	—	—	0
Late (%)	—	—	—	29	9

*Average distance between adjacent origins for the two domains centromeric or telomeric of the central 377-kb region (29,146–29,523 kb on the chromosome, see supplementary Fig 4 online).

process rather than a technical limitation. Finally, because we applied very stringent criteria, it is possible that origins detected in one experiment could be just below the cutoff criteria in another experiment. To evaluate this possibility, we examined two origins that were detected in some, but not all, experiments using real-time PCR quantification of nascent strand DNA and confirmed that they represented authentic origins (data not shown).

We analysed the average distance from centre to centre of adjacent origins for all five regions combined (supplementary Fig 6A online). With the exception of the 377-kb domain on chromosome 22 mentioned above, 75% of the shared origins were 50 kb or less apart. Our results are compatible with the estimates for inter-origin distance obtained by analysis of autoradiographs of DNA fibres (reviewed by Hand, 1978) and by recent DNA-combing studies (Anglana *et al*, 2003; Norio *et al*, 2005).

It has been proposed that broad initiation zones will be observed in large intergenic regions and in non-transcribed genes, whereas small intergenic regions will contain circumscribed origins, and transcribed regions and/or regions with a nonpermissive chromatin will contain no active origins (reviewed by Machida *et al*, 2005). However, we did not identify large replication zones (approximately 50 kb), and 95%, 90% and 69% of the mapped origins for experiments 1, 2 and the shared peaks, respectively, were 5 kb or less (supplementary Fig 6B online). Furthermore, we found that approximately 80% of the shared origins identified within the region of chromosome 22 analysed are within genes. The mapping of origins and the analysis of the transcription level within other large regions of the genome will be required to determine whether there is a general trend between origin location and the position of genes.

As it has been suggested that origins of replication in higher eukaryotes might be determined by chromatin elements such as A+T-rich DNA rather than specific sequences, we analysed the A+T content of the origins mapped in this study (Table 1; Fig 3; supplementary Fig 4 online; data not shown). We were unable to identify any clear A+T-rich common element or strong sequence similarity among the origins; however, we could not exclude the possibility that some sequence elements, localized near or at the replication start sites, which would only be shown by fine mapping, have a role in origin selection and activity (Ina *et al*, 2001).

Confirmation of origins by two independent assays

We evaluated our origin-mapping method by real-time PCR analysis of the abundance of nascent strand DNA and chromatin immunoprecipitation (ChIP) of two proteins of the prereplication complex: origin recognition complex (ORC6) and minichromosome maintenance deficient 5 (MCM5). By using these two independent assays, we confirmed the known origin of *LMNB2*, and all three new origins within the *FRAXA* and chromosome 22 regions (Fig 4). For all four origins examined, ORC6 binding correlated with the results of nascent strand abundance. We found that MCM5 can have a dispersed binding pattern as observed at the known *FMR1* origin, most probably owing to the spreading of MCM proteins from the origin, which has been observed previously (Todorovic *et al*, 2005). We also confirmed the presence of the known *MYC* origin by real-time PCR analysis of nascent strand abundance (data not shown).

Correlation of origins and chromatin acetylation

As it has been proposed that chromatin acetylation has a role in origin selection, activity and timing of firing (reviewed by Zhou

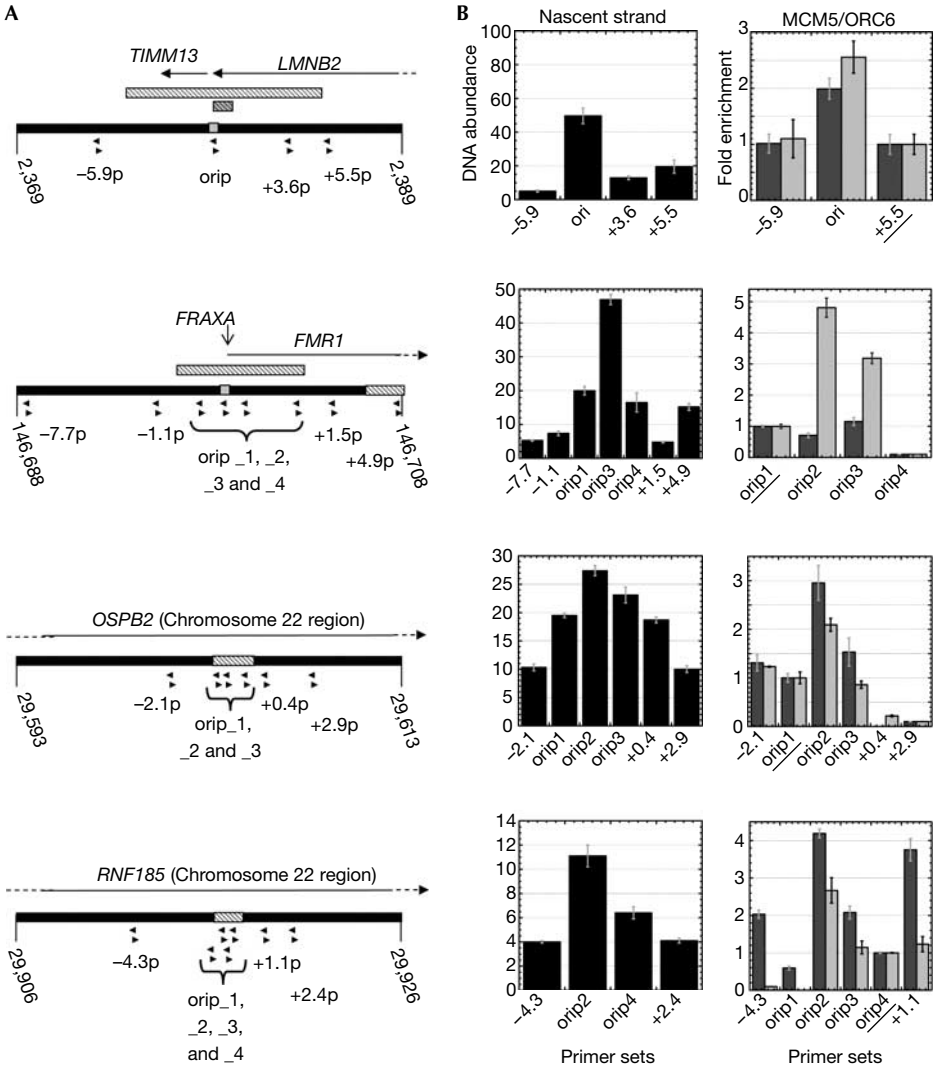

Fig 4 | Origin confirmation by real-time PCR analysis of nascent strand abundance and chromatin immunoprecipitation assays. (**A**) Position of the real-time PCR primer sets on a 20-kb segment of the *LMNB2*, *FRAXA* and chromosome 22 regions arrayed. The grey bars and the hatched bars indicate the arrayed position of the known and predicted origins, respectively. (**B**) Real-time PCR quantification of nascent strand abundance and DNA immunoprecipitated with MCM5 (dark grey) or ORC6 (light grey) antibodies. For each region, the ChIP results are presented as the fold enrichment over the value of the underlined primer set. Both assays were performed independently twice and yielded similar results (data not shown). ChIP, chromatin immunoprecipitation. MCM5, minichromosome maintenance deficient 5; ORC6, origin recognition complex 6; ori, origin of replication.

et al, 2005), we carried out a ChIP-on-chip analysis, combining ChIP of histone H3 acetylated on Lys 9/14 and our microarray platform (supplementary Figs 7 and 8 online). Comparison of the location of origins and Ac-H3 Lys 9/14 peaks with the available replication timing data (Fig 3; supplementary Fig 4 online; Table 1) showed a strong correlation between early origin firing and chromatin acetylation, as previously shown in yeast (Vogelauer *et al*, 2002).

METHODS

Cell culture. Human Epstein–Barr virus-transformed lymphoblastoid cells (cell line 11365) with a normal karyotype (46,XX) were cultured in RPMI1640 medium with 1% HEPES, 100 U/ml penicillin, 100 μg/ml streptomycin and 10% fetal bovine serum (all reagents from Invitrogen, Carlsbad, CA, USA) at 37 °C in a 5% CO_2/95% air atmosphere.

Isolation of nascent DNA. Nascent DNA was isolated according to the method described by Giacca *et al* (1994), with the following modifications. The sucrose gradient fractions corresponding to the 300 bp–1 kb size range were pooled and precipitated by adding 1/10 volume of 3 M sodium acetate (pH 6.0) and 2.5 volumes of cold ethanol (6–8 μg of nascent DNA was obtained from 100×10^6 cells). The size range of the DNA was verified by gel electrophoresis.

Isolation of genomic DNA from sorted G1 cells. Cells (50×10^6) in logarithmic growth phase were fixed at a final concentration of 0.5×10^6 cells/ml in 30% PBS/70% ethanol and stored at −20 °C for at least 30 min. The cells were washed twice with cold PBS

($MgCl_2$/$CaCl_2$-free; Sigma, St Louis, MO, USA), resuspended at a final concentration of 1×10^6 cells/ml in RNaseA solution (0.1 mg/ml in PBS, Sigma) and incubated at 37 °C for 30 min. Cells were washed twice with cold PBS, resuspended at a final concentration of 1×10^6 cells/ml in propidium iodide (PI) solution (0.1 mg/ml in PBS, Sigma) and incubated on ice in the dark for at least 30 min. PI-stained cells were analysed and sorted on a MoFlo cytometer (Dako Colorado Inc, Fort Collins, CO, USA) equipped with a Coherent Innova 90-5 argon laser emitting at 488 nm. DNA-bound Pl emission was collected at the FL2 position through a 580/30 band pass filter. Cells were gated on a pulse width versus forward scatter plot to discriminate against aggregates. The sorted G1 cells were incubated at 65 °C after adding EDTA (5 mM), Tris (pH 8.0, 10 mM), SDS (0.1%) and proteinase K (200 μg/ml, Roche, Basel, Switzerland), followed by phenol–chloroform extraction. The aqueous phase was dialysed against water at 24 °C for 4 h. The DNA (approximately 6 μg) was precipitated by adding 1/25 volume of 5 M NaCl and 2.5 volumes of cold ethanol.

Chromatin immunoprecipitation. ChIP was carried out according to the protocol provided by Upstate Inc (Millipore, Billerica, MA, USA; http://www.upstate.com) with the following modifications. Crosslinking of the protein–DNA complexes was carried out by incubating $5–10 \times 10^6$ log-phase cells with 1% formaldehyde at 37 °C for 15 min and stopped by the addition of glycine (0.125 M, 5 min, 24 °C). Cells were washed twice with ice-cold PBS and lysed with 500 μl of NP-40 buffer (5 μM PIPES (pH 8.0), 85 μM KCl, 0.5% NP-40 and 1% protease inhibitor cocktail; Calbiochem, EMD Chemicals Inc, San Diego, CA, USA). Nuclei were lysed with 200 μl of SDS buffer (0.05 M Tris–HCl (pH 8.1), 10 μM EDTA, 1% SDS and 1% protease inhibitor cocktail), and DNA was sheared. Before the pre-clearing step with salmon sperm DNA–protein A or G agarose beads (32 μg; Upstate), 1% of the nuclear lysate was set aside as the input sample. The following antibodies were used for ChIP: acetylated histone H3 Lys 9/14 (06-599, Upstate, 1:132 dilution), MCM5 (ab17967, AbCam (Cambridge, MA, USA), 1:125) and ORC6 (05-938, Upstate, 1:500). The antibody–protein–DNA complexes were retrieved by the addition of 24 μg of salmon sperm DNA–protein A or G agarose beads.

Microarray technique. All regions were tiled with 50-mer probes avoiding ambiguous bases by using repeat masked data obtained from UCSC (http://genome.ucsc.edu; supplementary Table 1 online). The minimum and median tiling intervals were 30 nt (start-to-start, 20-nt overlap) for the five regions. The chromosome location of the probes is shown in Fig 3 (*MYC*, *LMNB2*, *HBB* and *FRAXA)* and in supplementary Fig 4 online (chromosome 22 region). The probes were then randomly placed on the array in duplicate using ArrayScribe array design software (NimbleGen Systems Inc, Madison, WI, USA). The arrays were constructed by NimbleGen using maskless array synthesis technology, with approximately 385,000 oligonucleotides synthesized by photo-lithography per array (Singh-Gasson *et al*, 1999). The labelling of DNA, microarray hybridization and scanning were carried out by the NimbleGen Service Laboratory as described previously (Selzer *et al*, 2005). Briefly, 1 μg of each DNA sample was labelled using random nonamers. Thirty-fold amplification was typically achieved and 13 μg of each DNA sample was used for the hybridization. Log_2 ratios (Equation (1)) were computed for each sample pair and scaled using the Tukey bi-weight mean (Mosteller and Tukey, 1977) using NimbleScan 2.1:

$$\log_2\left(\frac{\text{Nascent strand hybridization signal}}{\text{G1 DNA hybridization signal}}\right) \quad (1)$$

***Z*-score analysis.** The absolute difference ($|X|$) (Equation (2)) between the linear converted data obtained for a probe and its duplicate was normalized by applying a *Z*-score transformation, using the mean (μ) and the standard deviation (δ) of all of the absolute differences calculated across the array (Equation (3)). The absolute difference was used to avoid favouring the results of one probe over its duplicate:

$$|X| = |\text{linear ratio probe} - \text{linear ratio duplicated probe}| \quad (2)$$

$$Z = \frac{(|X|) - \mu}{\delta} \quad (3)$$

Origin peak finding method and FPR analysis. During the peak finding process, the sliding window was moved across each data set. We tested a range of sizes for the sliding window and, based on the detection of the known origins, we selected a window size of 625 nt. Within the window, a probe was considered to 'qualify' if its log_2 ratio was above a set cutoff. A peak was identified when the number of qualifying probes was above a set percentage of the probes within the window (from 20% or 4 out of 20 probes to 100% or 20 out of 20 probes). The start position of the first qualifying probe and the end position of the last qualifying probe were set as the start and end positions of the peak. This procedure was repeated using a series of log_2 ratio cutoff values, r_1, r_2, … r_i, where $r_1 > r_2 > \cdots > r_i$, and r_1 equals the maximum of the log_2 values across a data set. To minimize the effects of outliers, we used a 'hypothetical maximum', which was calculated for each strand of each region, and corresponded to the mean of the log_2 ratio values plus six standard deviations.

To estimate the FPR of each peak, we randomized the data, strand-by-strand, for each region and estimated the number of peaks found by chance at each log_2 ratio cutoff value r_i. The process was repeated 20 times and averaged. The FPR was defined as the ratio of the average number of peaks found using the log_2 ratio cutoff value r_i for the randomized and the non-randomized data (Equation (4)). This FPR value was assigned to the peaks that were present at the log_2 ratio cutoff value r_i, but absent at the value r_{i-1}:

$$\text{FPR} = \frac{\text{Number of randomized data peaks}}{\text{Number of non-randomized data peaks}} \quad (4)$$

Data merging methods. The merging of peaks was carried out in two steps: first strand- by strand and then for the four data sets for each region. Any two peaks 2 kb or less apart on the same strand were merged. The most left and the most right positions of the two peaks were used to define the merged-peak coordinates. Only the peaks present in at least three out of four data sets of an experiment were retained as shared peaks. The merged peaks obtained between the four data sets for each percentage of qualifying probes (40–100%) were then combined. The programs developed for *Z*-score analysis and peak finding are available on request.

Real-time PCR. PCR reactions were carried out in triplicate in a LightCycler® PCR machine using LightCycler® capillary tubes

(Roche) and the RT^2 real-time PCR assay mix (Superarray), following the manufacturer's protocol. The abundance of nascent strand was estimated using a six-point standard curve obtained with a fivefold serial dilution of genomic DNA. The ChIP sample quantifications were calculated relative to the input sample and are presented as the fold enrichment over a given marker for each region analysed. PCR reactions were verified by melting curve analysis and gel electrophoresis. The PCR conditions were set as follows: an initial 15 min denaturation step at 95 °C, followed by 40 cycles including a denaturation step of 15 s at 95 °C and an amplification step of 60 s at the temperature specific for each primer set (supplementary Table 3 online), and finally a melting cycle going up to 95 °C with a slope of 0.1 °C/s.

Data deposition. The data discussed in this publication have been deposited in National Center for Biotechnology Information's Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE7889.

Supplementary information is available at *EMBO reports* online (http://www.emboreports.org).

ACKNOWLEDGEMENTS

We thank L. Godley, L. Rothman-Denes, D. Bishop, S. Kron and members of the Le Beau laboratory for helpful discussions. We are grateful to D. Gilbert for critical reading of the manuscript. This work was supported by PHS grant CA41644 (M.M.L.).

REFERENCES

Abdurashidova G, Deganuto M, Klima R, Riva S, Biamonti G, Giacca M, Falaschi A (2000) Start sites of bidirectional DNA synthesis at the human lamin B2 origin. *Science* **287:** 2023–2026

Anglana M, Apiou F, Bensimon A, Debatisse M (2003) Dynamics of DNA replication in mammalian somatic cells: nucleotide pool modulates origin choice and interorigin spacing. *Cell* **114:** 385–394

Giacca M *et al* (1994) Fine mapping of a replication origin of human DNA. *Proc Natl Acad Sci USA* **91:** 7119–7123

Gilbert DM (2001) Making sense of eukaryotic DNA replication origins. *Science* **294:** 96–100

Gray SJ, Gerhardt J, Doerfler W, Small LE, Fanning E (2007) An origin of DNA replication in the promoter region of the human fragile X mental retardation (FMR1) gene. *Mol Cell Biol* **27:** 426–437

Hamlin JL, Dijkwel PA (1995) On the nature of replication origins in higher eukaryotes. *Curr Opin Genet Dev* **5:** 153–161

Hand R (1978) Eucaryotic DNA: organization of the genome for replication. *Cell* **15:** 317–325

Hansen RS, Canfield TK, Lamb MM, Gartler SM, Laird CD (1993) Association of fragile X syndrome with delayed replication of the FMR1 gene. *Cell* **73:** 1403–1409

Ina S, Sasaki T, Yokota Y, Shinomiya T (2001) A broad replication origin of *Drosophila melanogaster*, oriDα, consists of AT-rich multiple discrete initiation sites. *Chromosoma* **109:** 551–564

Kamath S, Leffak M (2001) Multiple sites of replication initiation in the human β-globin gene locus. *Nucleic Acids Res* **29:** 809–817

Kitsberg D, Selig S, Keshet I, Cedar H (1993) Replication structure of the human β-globin gene domain. *Nature* **366:** 588–590

MacAlpine DM, Bell SP (2005) A genomic view of eukaryotic DNA replication. *Chromosome Res* **13:** 309–326

Machida YJ, Hamlin JL, Dutta A (2005) Right place, right time, and only once: replication initiation in metazoans. *Cell* **123:** 13–24

Mesner LD, Crawford EL, Hamlin JL (2006) Isolating apparently pure libraries of replication origins from complex genomes. *Mol Cell* **21:** 719–726

Mosteller F, Tukey JW (1977) *Data Analysis and Regression: A Second Course in Statistics*. Boston, MA, USA: Addison-Wesley

Norio P, Kosiyatrakul S, Yang Q, Guan Z, Brown NM, Thomas S, Riblet R, Schildkraut CL (2005) Progressive activation of DNA replication initiation in large domains of the immunoglobulin heavy chain locus during B cell development. *Mol Cell* **20:** 575–587

Selzer RR, Richmond TA, Pofahl NJ, Green RD, Eis PS, Nair P, Brothman AR, Stallings RL (2005) Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. *Genes Chromosomes Cancer* **44:** 305–319

Singh-Gasson S, Green RD, Yue Y, Nelson C, Blattner F, Sussman MR, Cerrina F (1999) Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat Biotechnol* **17:** 974–978

Todorovic V, Giadrossi S, Pelizon C, Mendoza-Maldonado R, Masai H, Giacca M (2005) Human origins of DNA replication selected from a library of nascent DNA. *Mol Cell* **19:** 567–575

Trivedi A, Waltz SE, Kamath S, Leffak M (1998) Multiple initiations in the c-*myc* replication origin independent of chromosomal location. *DNA Cell Biol* **17:** 885–896

Vassilev L, Johnson EM (1990) An initiation zone of chromosomal DNA replication located upstream of the c-*myc* gene in proliferating HeLa cells. *Mol Cell Biol* **10:** 4899–4904

Vogelauer M, Rubbi L, Lucas I, Brewer BJ, Grunstein M (2002) Histone acetylation regulates the time of replication origin firing. *Mol Cell* **10:** 1223–1233

White EJ *et al* (2004) DNA replication-timing analysis of human chromosome 22 at high resolution and different developmental states. *Proc Natl Acad Sci USA* **101:** 17771–17776

Zhou J, Chau C, Deng Z, Stedman W, Lieberman PM (2005) Epigenetic control of replication origins. *Cell Cycle* **4:** 889–892