European Journal of Human Genetics (2012) 20, 185–191
© 2012 Macmillan Publishers Limited All rights reserved 1018-4813/12
www.nature.com/ejhg

ARTICLE

Epigenetic regulation of the X-chromosomal macrosatellite repeat encoding for the cancer/testis gene *CT47*

Judit Balog[1], Dan Miller[2], Elena Sanchez-Curtailles[1], Jose Carbo-Marques[1], Gregory Block[2], Marco Potman[1], Peter de Knijff[1], Richard JLF Lemmers[1], Stephen J Tapscott[3] and Silvère M van der Maarel*[,1]

Macrosatellite repeats (MSRs) present an extreme example of copy number variation, yet their epigenetic regulation in normal and malignant cells is largely understudied. The CT47 cancer/testis antigen located on human Xq24 is organized as an array of 4.8 kb large units. *CT47* is expressed in the testis and in certain types of cancer, but not in non-malignant somatic tissue. We used CT47 as a model to study a possible correlation between copy number variation, epigenetic regulation and transcription originating from MSRs in normal and malignant cells. In lymphoblastoid cell lines and primary fibroblasts, *CT47* expression was absent, consistent with the observed heterochromatic structure and DNA hypermethylation of the *CT47* promoter. Heterochromatinization of CT47 occurs early during development as human embryonic stem cells show high levels of DNA methylation and repressive chromatin modifications in the absence of *CT47* expression. In small-cell lung carcinoma cell lines with low levels of *CT47* transcripts, we observed reduced levels of histone 3 lysine 9 trimethylation (H3K9me3) and trimethylated lysine 27 of histone H3 (H3K27me3) without concomitant increase in euchromatic histone modifications. DNA methylation levels in the promoter region of *CT47* are also significantly reduced in these cells. This supports a model in which during oncogenic transformation, there is a relative loss of repressive chromatin markers resulting in leaky expression of *CT47*. We conclude that some MSRs, like CT47 and the autosomal MSRs TAF11-Like, PRR20, ZAV and D4Z4, the latter being involved in facioscapulohumeral muscular dystrophy, seem to be governed by common regulatory mechanisms with their abundant expression mostly being restricted to the germ line.
European Journal of Human Genetics (2012) **20,** 185–191; doi:10.1038/ejhg.2011.150; published online 3 August 2011

Keywords: macrosatellite repeats; cancer/testis antigen; copy number variation; epigenetics; D4Z4; FSHD

INTRODUCTION

Approximately half of the human genome is composed of repetitive DNA and a significant proportion of this DNA is arranged as tandemly repeated units. It is now widely accepted that these elements are an integral part of the genome with structural and regulatory functionalities. Macrosatellite repeats (MSRs) are a specific class of repeated DNA defined by a minimum unit size of 100 bp. Typically, MSR unit sizes are much larger and MSRs usually span hundreds of kilobases of DNA. MSRs can be protein-coding or non-coding, and are often subtelomerically or centromerically located. MSRs have been implicated to be involved in human disease,[1] including cancer,[2] and epigenetic regulation of the genome, such as X inactivation.[3] Recently, a genome-wide survey revealed that MSRs are more abundant than previously anticipated, are distributed over the entire genome and are organized in a higher order structure.[4] Owing to their size variability, MSRs thus represent an important source of copy number variation (CNV) that may contribute to the missing heritability observed in some genome-wide association and CNV studies.

The structure and function of MSRs is largely understudied. Probably the best studied MSR is D4Z4, because contractions of this MSR are associated with facioscapulohumeral muscular dystrophy (FSHD).[1,5] Other MSRs that have been studied in various details include DXZ4 on the X chromosome, which was recently proposed to have a role in X-chromosome inactivation;[3] RS447, which encodes for a deubiquitinating enzyme;[6] and RNU2, encoding a small nuclear RNA,[7] TAF11-Like,[8,9] which was recently suggested to be involved in a psychiatric disorder, and very recently SST1, PRR20 and ZAV.[9]

D4Z4 is characterized by bidirectional transcriptional activity, high DNA methylation and the presence of both heterochromatic histone modifications such as histone 3 lysine 9 trimethylation (H3K9me3) as well as the transcriptionally permissive histone modification such as dimethylated lysine 4 of histone H3 (H3K4me2). Upon contraction, as seen in FSHD, D4Z4 becomes hypomethylated,[10] shows selective loss of H3K9me3, with secondary losses of heterochromatin protein 1γ and the cohesin complex,[11] thus leading to a more open chromatin structure. DXZ4 has chromatin features very similar to D4Z4, but shows opposite chromatin configurations on the active (Xa) and inactive (Xi) X chromosome. On Xa, DXZ4 is organized as constitutive heterochromatin with high levels of H3K9me3, while on the Xi, DXZ4 adopts a more euchromatic structure with high levels of H3K4me2 and H3K9ac. Also, DXZ4 is bidirectionally transcribed with discrete differences in the transcriptional landscape on Xa and Xi.[3]

[1]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; [2]Department of Pediatrics, University of Washington, Seattle, WA, USA; [3]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
*Correspondence: Professor SM van der Maarel, Department of Human Genetics, Leiden University Medical Center, PO Box 9600, Leiden AL 2333, The Netherlands. Tel: +31 71 526 9480; Fax: +31 71 526 8285; E-mail: maarel@lumc.nl
Received 8 February 2011; revised 5 July 2011; accepted 6 July 2011; published online 3 August 2011

From these studies, the picture emerges that CNV, expression profile and chromatin structures are important attributes of MSRs, which when disturbed may have consequences for genome integrity and function, as has recently been demonstrated in *Saccharomyces cerevisiae* where large rDNA repeats are required for the maintenance of genome integrity.[12]

Cancer/testis antigens (CTAs) represent a class of genes that are typically only expressed in germ cells of the testis, fetal ovary and placenta, and sometimes in the brain, and which also show expression in various tumor types.[13,14] Epigenetic changes in cancer typically involve hypermethylation of CpG islands and hypomethylation of repetitive DNA elements.[15,16] Hypermethylation can lead to transcriptional silencing of tumor suppressor genes while the effect of hypomethylation in cancer is less clear, but thought to contribute to genomic instability. In accordance, it has been established that many CTAs show expression in malignancies, which is associated with promoter demethylation.[17,18] However, by and large, the chromatin structure of CTAs under normal and malignant conditions has not been studied in great detail.

The cancer/testis gene *CT47* on Xq24 is arranged as an MSR with up to 17 copies of a putative protein coding gene ordered head to tail.[19] Being a CTA and an MSR, in this study we set out to examine the array length and chromatin structure of this MSR in *CT47*-expressing and -non-expressing cells.

MATERIALS AND METHODS

Cell lines and culturing conditions

HapMap cell lines and primary fetal lung fibroblast WI38 were purchased from Coriell Cell Repositories (Camden, NJ, USA) and cultured according to the supplier's instruction. Small-cell lung carcinoma (SCLC) cell lines were purchased from LGC Standards GmbH (LGC Promochem, Teddington, UK) and cultured as described;[20] human embryonic stem cells (hESCs) WA9 and WA14[21] were purchased from WiCell Research Institute (Madison, WI, USA). Cells were cultured according to a published protocol.[22] hESC and human embryonic body (hEB) samples were analyzed for the presence of the differentiation markers *NANOG*, *OCT4* and *XIST* as described,[23] and showed that ESCs had differentiated properly.

Primers

All primers used in the study are described in Supplementary Table S1.

Characterization of the CT47 array

CT47 array size was determined by pulsed field gel electrophoresis (PFGE) after digestion of agarose-embedded DNA with *Eco*RI. Hybridization was performed with radioactively labeled CT47 probe recognizing the CT47 unit sequence (NT_011786.15), generated by PCR.

Quantitative methylation analysis of bisulfite-converted DNA

Genomic DNA was treated with sodium bisulfite (EZ DNA Methylation Kit; Zymo Research, Irvine, CA, USA; no. D5002). Bisulfite-converted DNA was PCR amplified in a final volume of 25 μl containing 1× PCR buffer supplied with the enzyme, 25 mmol of each dNTP, 10 pmol of bis CT47 primer pair and 0.1 units of HotStar Taq DNA polymerase (Qiagen, Venlo, The Netherlands; no. 203203). Primers specific for the CT47 promoter region were designed with MethPrime to amplify the genomic region 270 base pairs (bp) upstream of exon 1 of CT47. PCR cycling conditions were as follows: 15 min denaturation step at 94°C, followed by 45 cycles of 94°C for 20 s, 55°C for 30 s and 72°C for 60 s, and a final step of 72°C for 5 min. PCR products were sequenced and the level of methylation of individual CpGs was measured with the ESME software (Epigenomics AG, Berlin, Germany).[24]

Chromatin immunoprecipitation

Chromatin was prepared from cells fixed with 1% formaldehyde and then quenched with 125 mM glycine. The chromatin immunoprecipitation (ChIP) procedure was carried out according to a published protocol.[25] Antibodies against H3K4me2 (no. 07-030), H3K9me3 (no. 17-625) and trimethylated lysine 27 of histone H3 (H3K27me3) (no. 17-622) were purchased from Millipore (Billerica, MA, USA). EZH2 antibody was purchased from Active Motif (no. 39103) and the histone H3 antibody (La Hulpe, Belgium; no. ab1791) from Abcam (Cambridge, UK).

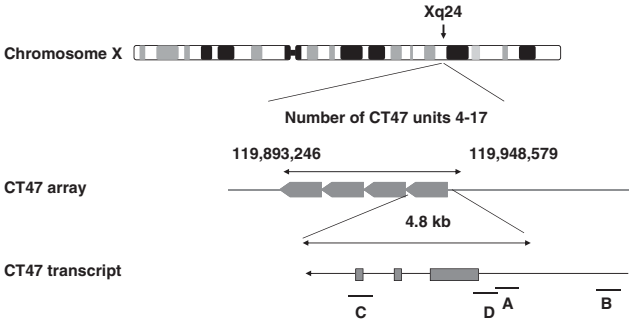

Figure 1 Schematic representation of CT47 macrosatellite array. The CT47 repeat array is localized on Xq24 at coordinates chrX: 119 893 246–119 948 579 according to hg18 assembly. The size of the array varies between 4 and 17 units, each unit being 4.8 kb in size. Chromatin markers were studied in the promoter region of the *CT47* gene encoded by each unit (exons are indicated in gray boxes). Genomic locations where relative abundance of different histone modifications were quantified are represented by the black line; A represents ChIP primer set for *CT47* promoter, B represents distal primer set located 3.7 kb from the array, C represents ChIP primer set for *CT47* exon 3, and D represents the bisulfate primer set.

To determine the level of nonspecific signals, we used IgG, which was supplied with the antibodies for H3K9me3 and H3K27me3. Immunoprecipitation with histone antibodies and IgG was carried out with a chromatin containing 3 μg DNA. Immunoprecipitation with EZH2 antibody and IgG was carried out with a chromatin containing 6 μg DNA. The amount of immunopurified DNA was determined by quantitative PCR with specific primer pairs at three different sites in CT47: promoter primer pair (labeled A in Figure 1), exon 3 primer pair (labeled C in Figure 1) and distal primer pair (labeled D in Figure 1). All primers were designed by primer3 program and optimized for minimizing sequence variations between individual repeat units. Primers were tested for specificity on a monochromosomal cell panel. A published primer pair was used for the *GAPDH* promoter.[26] Relative abundance of histone modifications was calculated in the following way: enrichment values obtained with IgG were subtracted from the enrichment values from specific antibodies. Normalization for CNV was carried out by correcting the enrichment values from the ChIP by the input values. Input values were obtained by determining the total amount of DNA used in the ChIP by using the specific primer pairs. These results are presented in the figures as relative enrichment to input. Samples were tested by ChIP in two independent experiments and every immunopurified DNA was quantified twice independently.

cDNA synthesis and quantification of CT47 mRNA

Human testis total RNA was purchased from Ambion (Austin, TX, USA; no. AM7072). Total RNA was isolated from LCL cell lines, hESCs, hEBs from lines WA9 and WA14, and SCLC lines with Qiasol reagent (Qiagen; no. 79306). RNA was purified using the miRNeasy kit (Qiagen; no. 217004). cDNA synthesis was carried out with Omniscript Reverse transcription kit (Qiagen; no. 205113) using 500 ng total RNA according to the manufacturer's instructions. cDNA synthesis for all samples was carried out in parallel under identical conditions and repeated twice. cDNA was quantified by using specific intron spanning primers for *CT47* using SRBR Green (BioRad, Veenendaal, The Netherlands; no. 170-8882) mix and measurements were performed with MyiQ Single-Color Real-Time PCR Detection System, and analyzed with the Bio-Rad iQ5 program. Samples were measured twice in duplicate.

Statistical analyses

Statistical analysis of DNA methylation levels of different samples was carried out by Mann–Whitney *U*-test.

RESULTS

No correlation between CT47 unit number and chromatin structure inside the CT47 array

HapMap cell lines were sized for the CT47 array, located at Xq24, by PFGE analysis of *Eco*RI-digested genomic DNA and using a probe recognizing the repeat (Supplementary Figure S1). The CT47 repeat sizes of selected cell lines used in this study are listed in Table 1. Taking advantage of the hemizygosity of CT47 in men, to study the repeat-size dependency of the CT47 chromatin structure, four male samples were chosen with short CT47 arrays (4–7 repeat units) and two with long arrays (12 or 14 units). Female samples were selected based on their homozygosity for the CT47 repeat size. The relative abundance of H3K4me2, H3K27me3 and H3K9me3 in *CT47* promoter region, exon 3 of the *CT47* gene and 3700 bp distal of the array (Figure 1) was determined by quantitative ChIP analysis. H3K4me2 is associated with an open chromatin structure, while H3K27me3 marks transcriptionally silenced regions, and H3K9me3 is specific for heterochromatin. The *GAPDH* promoter region with an open chromatin structure was included as a control (Supplementary Figure S2). The most abundant histone modification at the *CT47* promoter in the male cell lines is H3K9me3, suggesting that the chromatin is highly condensed in this region (Figure 2a). We also detected high levels of H3K27me3 and low levels of H3K4me2, further supporting a closed chromatin structure. Cell lines with short or long CT47 repeat arrays showed high inter-individual variation in H3K9me3 and H3K27me3 levels without clear correlation with repeat length. Distal to the array, the relative abundance of H3K27me3 is higher and H3K9me3 levels are lower than in the *CT47* promoter, whereas H3K4me2 levels are similarly low (Figure 2b). There is no correlation between CT47 array length and chromatin structure distal to the array.

Similar to the male samples, high levels of H3K9me3 and H3K27me3 and low levels of H3K4me2 were detected in the female cell lines with similarly sized long or short CT47 repeats on both alleles (Figure 2c). The H3K9me3 levels were lower than or close to the H3K27me3 levels, whereas in male samples H3K9me3 was the most abundant modification. Distal to the CT47 array, the relative abundance of these histone modifications is similar as within the array in female samples (Figure 2d). We also included a female primary fibroblast cell line (5–6 units) to exclude that immortalization by EBV has significant effects on the chromatin structure of CT47. We did not observe a difference in chromatin structure between female immortalized LCLs and female primary fibroblasts, excluding a large effect of the immortalization process. Thus, in normal somatic cells, the CT47 repeat contains repressive chromatin markers, and within the tested size range, there is no correlation between CT47 repeat array size and chromatin structure.

Table 1 List of cell lines included in the study

HapMap identifier	*Lab identifier*	*Cell type*	*Gender*	*CT47 unit number*
GM18980	237	LCL	Female	5, 5
GM12044	166	LCL	Female	10, 10
GM10831	138	LCL	Female	11, 8
	WI38	Primary fetal lung fibroblast	Female	5, 6
GM18561	248	LCL	Male	4
GM12144	185	LCL	Male	5
GM10830	135	LCL	Male	6
GM19204	334	LCL	Male	7
GM18953	274	LCL	Male	12
GM18504	385	LCL	Male	14
	WA9	hESC	Female	ND
	WA14	hESC	Male	ND
	NCI-H69	SCLC	Male	9, 9, 5
	NCI-H82	SCLC	Male	10, 8
	NCI-H187	SCLC	Male	8, 8, 4
	H1607	SCLC	Male	ND

Abbreviation: ND, not determined.

CT47 is heterochromatic in hESCs and EBs

Male WA14 and female WA9 human hESC cell lines were studied for CT47 chromatin structure in the pluripotent state and when differentiated into hEBs. This showed that H3K9me3 is already abundantly present on this locus in hESCs and hEBs (Figure 3). H3K4me2 and H3K27me3 levels were equally low in these samples, which is different from LCLs or primary fibroblasts, where H3K27me3 was far more abundant than H3K4me2. We did not observe major changes in the relative abundance of the different histone markers during differentiation. We conclude that the *CT47* promoter region is highly heterochromatic early in development and remains heterochromatic during early differentiation, suggesting that at these stages *CT47* expression does not change.

SCLC cells are expressing CT47 RNA

We selected four commercially available SCLC cell lines for determining *CT47* expression in relation to chromatin structure. Two of them were already reported to express *CT47* (NCI-H82 and NCI-H187),[19] while the two other SCLC cell lines were not studied yet (NCI-H69 and H1607). Human testis total RNA was included as a control. By quantitative RT-PCR we detected *CT47* expression in all SCLC cell lines but not in hESCs. hEBs or in LCLs and fibroblast cell cultures studied. In good agreement with previously published data, SCLC cell lines only express 0.1-3% of *CT47* levels compared to testis (Figure 4a).[19]

Chromatin structure of CT47 array in SCLCs

We studied the role in transcription of the chromatin structure of the *CT47* promoter region in all 4 SCLC cell lines. Figure 4c shows that the relative abundance of H3K4me2 is lower in SCLCs compared to LCL 385. However, H3K9me3 levels are dramatically reduced in all SCLC cell lines. Of all SCLCs, the highest levels of H3K9me3 are observed in sample NCI-H82, which expresses low levels of *CT47*. SCLC cell lines NCI-H69, NCI-H187 and H1607 show similar low levels of H3K9me3. H3K27me3 levels do not show a reduction but an increase, except for NCI-H69. H3K27me3 is generated by the catalytic subunit of the PRC2 complex, the histone methyl transferase EZH2. Transcriptionally silenced regions often show the simultaneous presence of H3K27me3 and EZH2. We therefore also tested the presence of EZH2 on the *CT47* promoter in LCLs and SCLCs, and observed that EZH2 levels are strongly reduced in SCLCs when compared with LCLs.

The chromatin structure of the SCLCs was also analyzed in the third exon of the *CT47* gene and 3700 bp distal to the CT47 array (Figure 4b). Chromatin markers showed a very similar picture in the third exon as in the promoter region, except sample NCI-H82, where we detected H3K9me3 levels comparable to the LCL sample. We detected high levels of H3K9me3 and H3K27me3 compared to H3K4me2 in the LCL 385. Interestingly in this region, H3K27me3 was more abundant than H3K9me3. SCLCs showed a general loss of repressive chromatin markers in this region. EZH2 levels corresponding

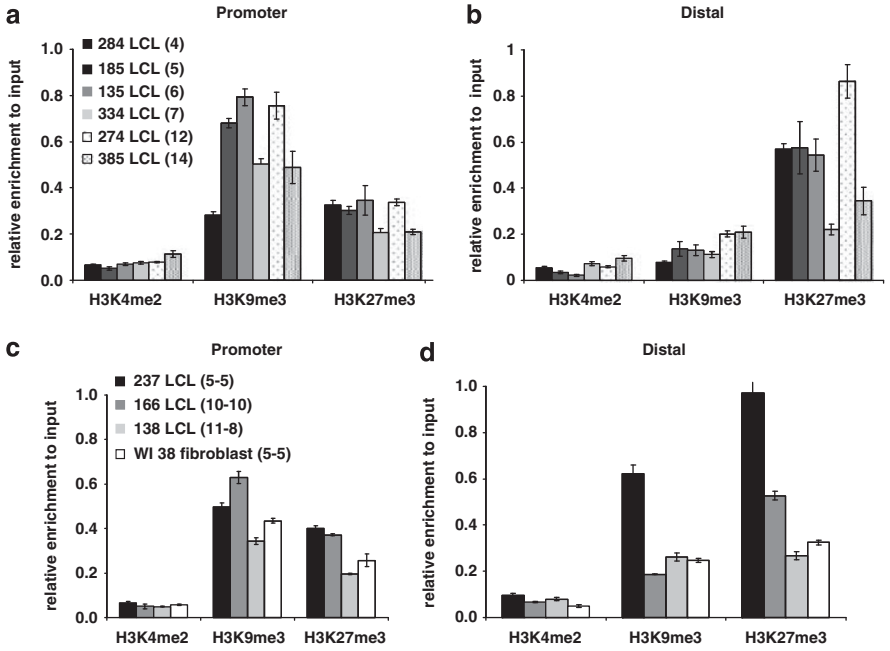

Figure 2 Relative abundance of histone modifications at the *CT47* promoter (**a**) male, (**c**) female and distal region (**b**) in male and female LCL samples from the HapMap panel. Numbers within brackets are corresponding to the number of CT47 units in the sample. (**a**) The euchromatic H3K4me2 mark has similar low levels in all samples. H3K27me3 is considered a transcriptional repressive chromatin mark, and is also very similar in samples 248, 185 and 135 with shorter CT47 arrays and shows more variation in samples with longer arrays. The heterochromatic mark H3K9me3 shows highest individual variation, independent of the array length. There is no correlation between chromatin structure and array length. (**b**) The histone modification profile at 3701 bp distal of the CT47 array, coordinates chrX: 119 952 280–119 952 414 (hg18). H3K27me3 levels are higher than H3K9me3 and H3K4me2 levels are similarly low as in the promoter region of *CT47*. Relative abundance of different histone modifications at the *CT47* promoter (**c**) and distal region (**d**) in female samples with low CT47 unit number (237) and with high CT47 unit number (166 and 138) on both X chromosomes. WI38 is a primary fetal lung fibroblast and was used in the study to show that immortalization does not lead to changes in chromatin structure of CT47 in LCLs. Histone profiles in female samples are similar to male samples in both regions studied, showing a transcriptionally repressive state marked by high levels of H3K9me3 in the promoter region and by high H3K27me3 in the distal region with low H3K4me2 levels. There is no correlation between array size and chromatin structure in these region in female samples. The histone modification landscape of primary fibroblast cell line WI38 is very similar to the LCL samples.

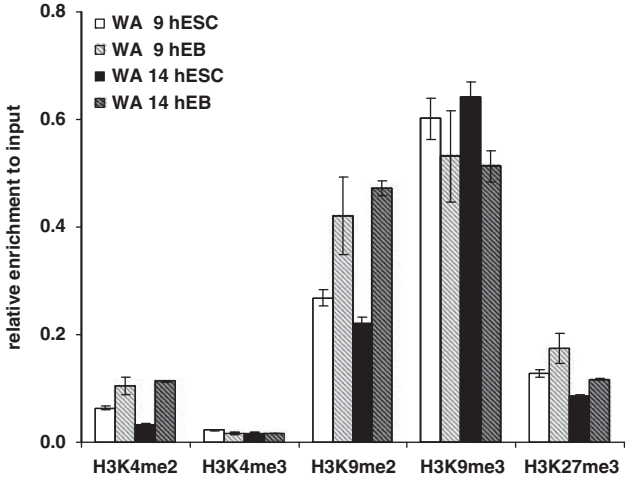

Figure 3 Relative abundance of different histone modifications at CT47 in human male (WA14) and female (WA9) pluripotent hESCs and in differentiated hEBs. The relative abundance of H3K4me2, associated with transcriptionally permissive chromatin, is increasing during differentiation in male and female samples. The repressive chromatin mark H3K9me3 is detected at high levels at pluripotent state and slightly decreases after differentiation. H3K27me3, associated with transcriptional repression, is detected in pluripotent state and increasing during differentiation.

to the H3K27me3 levels were high in LCL 385 and lower in the SCLCs. Thus, in LCLs the CT47 promoter region contains both the transcriptional repressive histone modification H3K27me3 together with EZH2, and the heterochromatic histone modification H3K9me3. In SCLCs, the heterochromatic H3K9me3 marker is dramatically reduced and it seems that the epigenetic regulation of transcription through the PRC2 complex is disturbed: while EZH2 levels are decreased in SCLCs, the histone marker generated by PRC2, H3K27me3 is not reduced. Thus, we observed that transcriptional activity inversely correlates with H3K9me3 levels and EZH2 occupation, rather than directly with H3K4me2 levels.

DNA methylation of *CT47* promoter region in LCLs, embryonic stem cells and SCLCs

Expression of CTAs in different tumor samples was previously shown to correlate with DNA hypomethylation.[18,27,28] Using bisulfite sequencing, we measured the level of DNA methylation in the promoter region of *CT47*, from 275 bp upstream from the first exon, in LCLs and SCLCs (Figure 4e), as well as in hESCs and hEBs (Supplementary Figure S3). The promoter region of *CT47* contains seven CpGs. Quantitative methylation analysis of these seven CpGs showed that the promoter region of *CT47* is highly methylated in LCLs, hESCs and hEBs. A significant difference was detected in DNA methylation between LCLs and SCLCs, with SCLCs showing decreased methylation levels at all seven CpGs ($P<0.01$).

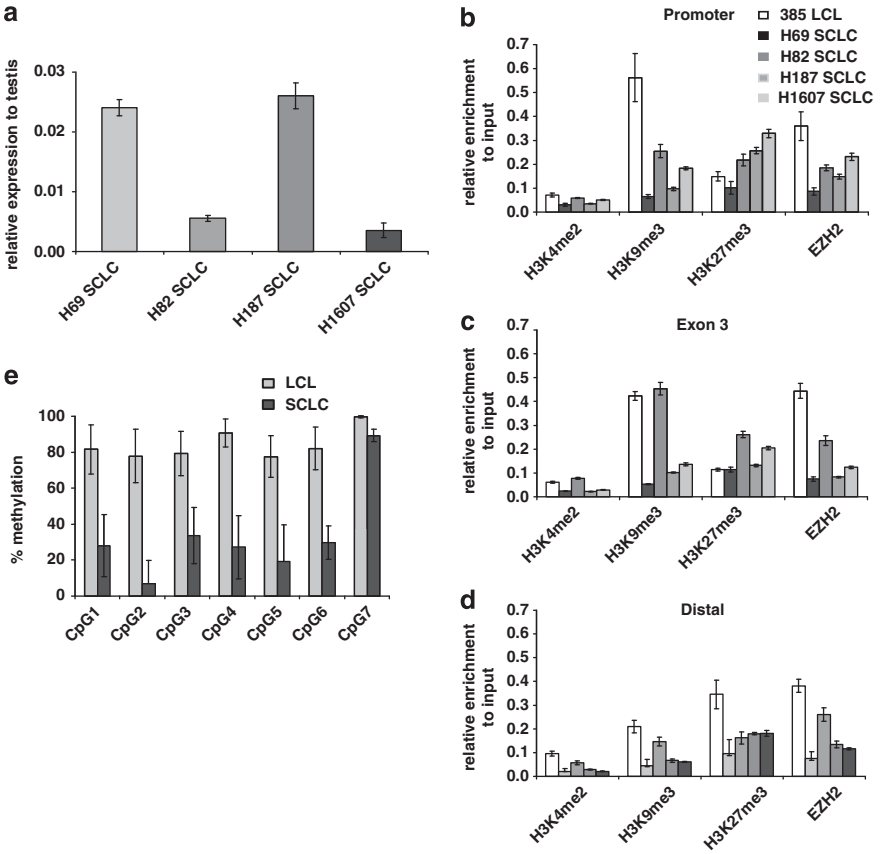

Figure 4 (**a**) Expression levels of *CT47* mRNA in different SCLC cell lines relative to the expression in human testis. Commercially available human total testis RNA and RNA isolated from SCLC lines were used for cDNA synthesis under identical conditions. *CT47* expression levels are normalized to the expression levels measured in the testis. Different SCLC lines have different, but low levels of *CT47* expression compared to the testis. Relative abundance of histone modifications and EZH2 at the *CT47* promoter (**b**), exon 3 (**c**) and the distal region in SCLC cell lines (**d**). SCLC cell lines show individual variation in the abundance of different histone modifications. Generally, a loss of the repressive chromatin mark H3K9me3 can be observed in SCLCs. H3K27me3 levels are higher in SCLCs at the promoter and exon 3 region than in LCLs, but lower at the distal region. The relative abundance of the PRC2 component EZH2 responsible for generating H2K27me3 is dramatically reduced in SCLCs compared to LCLs at all region studied. (**e**) DNA methylation levels at CpGs located in the *CT47* promoter region in LCL and SCLC samples. The methylation level of seven different CpGs, next to the transcriptional start site of *CT47*, was determined by bisulfite sequencing and quantified by ESME program. There is a significant difference between the methylation level of LCLs and SCLCs at every CpG tested ($P<0.01$).

DISCUSSION

On human Xq24, highly homologous copies of CT47 are organized into an MSR of tandemly arranged units of each 4.8 kb in size. This polymorphic MSR varies from 4 to 17 copies in length (manuscript in preparation and Chen *et al*[19]). As *CT47* has a very specific expression profile, it is a good model to study the tissue-specific epigenetic regulation of MSRs. CTAs form a large family of protein coding sequences with specific expression pattern: they are expressed in the testis and certain types of cancers. Our study presents the first detailed analysis of the chromatin structure of a CTA in non-malignant and malignant cells and during early stages of development. Although biological processes leading to cancer should preferably studied on human tumor material, the established SCLCs represent a valuable and readily available tool for studying genetic[29] and epigenetic changes[30] in cancer.

Our expression data are in a good agreement with earlier studies showing that *CT47* expression is restricted to the testis and malignant cells,[19] albeit that the expression in SCLCs is only a fraction of that observed in the testis. We cannot exclude the possibility that only a small proportion of SCLC cells in culture express *CT47*, analogous to the presence of small subpopulation of solid tumors.[2]

We did not observe the expression of *CT47* in hESC cells and hEBs. The CTAs MAGE-A, GAGE and NY-ESO-1 were previously shown not to be expressed in hESCs,[31] whereas MAGE-A2 and MAGE-A6 expression could be detected in 20-day-old hEBs. Although we cannot exclude the possibility that the absence of *CT47* transcript in our hESC and hEB samples is the result of the epigenetic fluidity of human hESCs,[23,32,33] our results suggest that *CT47* transcription is induced at later stages of development, possibly between the pluripotent stem cell and primordial germ cell stages, from which mature germ cells are developing.

Chromatin analysis of non-expressing cells showed high levels of heterochromatic histone modifications H3K27me3 and H3K9me3, and relatively low levels of the euchromatic histone modification H3K4me2 in the promoter region of *CT47*. We could not detect a correlation between CT47 array length and its chromatin structure in analogy with FSHD, where contractions only below a threshold of 11 units result in D4Z4 chromatin relaxation.[11] Rather, we observed a high variation in the abundance of different histone modifications in samples with the same array lengths (Figures 2 and 3). As D4Z4 shows similar variability in levels of histone modifications with repeat lengths >11 units, it is likely that 4 units of CT47 is sufficient to ensure a

heterochromatic chromatin environment in non-germline cells and opens up the possibility that each MSR needs a minimal, but for each MSR unique, number of repeat units to establish its appropriate chromatin structure.

We observed a relative higher abundance of H3K27me3 to H3K9me3 in female samples. The observation that females have relatively higher levels of H3K27me3 than males suggests that repression of CT47 is differently established on Xa and Xi. Indeed, Xi in female samples can be more densely populated with H3K27me3 than Xa.[34] Nevertheless, this is different from the X-chromosomal MSR DXZ4, showing that two MSR on Xq can behave very differently in chromatin structure on the Xa and Xi.

A unique feature of the pluripotent state is the so-called bivalent chromatin state defined by the presence of permissive (H3K4me3) and repressive (H3K27me3) histone modifications.[35] Bivalent states are found at promoter regions of genes expressed at low levels in hESCs that are associated with differentiation. We detected identical low levels of H3K4me2, H3K4me3 and H3K27me3 in hESCs and hEBs at the *CT47* promoter region, suggesting that the product of this gene is not necessary for early development.

Repetitive elements were identified as targets of PRC1 and PRC2 complexes in the pluripotent state.[36] Our results show that the MSR CT47 has, like D4Z4,[11] a heterochromatic structure already at early stages of development, marked by abundant H3K9me3 levels, and that this chromatin structure does not change during differentiation. The high levels of H3K9me3 suggest that repression of MSR is mainly governed by SUV39H1 and SUV39H2 and not so much by PRC complexes.

SCLC cell lines, which do express *CT47*, show a disturbed chromatin structure of this MSR. In addition, male cell lines NCI-H69 and NCI-H187 harbor three arrays and NCI-H82 two arrays of CT47 MSR, suggesting that they have become aneuploid for the X chromosome. We detected significantly lower levels of H3K9me3 in these cancer cell lines without a concomitant relative increase in the abundance of the transcriptionally permissive histone modification H3K4me2. This might explain the relatively low levels of *CT47* expression in these cells. Rather than transcriptional reactivation, SCLCs seem to show a relative loss of transcriptional repression, resulting in leaky expression of *CT47*. Owing to technical limitations, we can currently not exclude the possibility that individual units within an array are differently organized.

In contrast to H3K9me3, H3K27me3 levels are not significantly lower in SCLCs, whereas EZH2 levels are significantly decreased in these samples. This shows that SCLC cells have serious defects in coordinated epigenetic regulation of this MSR, consistent with reports of genome-wide epigenetic changes in other malignancies.[16]

CTA gene expression is often correlated with DNA promoter hypomethylation. We observed that all of the CpG dinucleotides examined in the promoter region of CT47 show a significant decrease in DNA methylation in SCLCs compared with the otherwise high methylation levels in non-malignant cells. Epigenetic factors like DNA methylation and histone modifications are acting in a coordinated way to establish normal development. Our study reveals that transcriptional regulation/derepression of *CT47* is dominated by the loss of repressive markers such as DNA methylation and H3K9me3: SCLCs with significantly lower levels of DNA methylation and H3K9me3 at the promoter region are expressing *CT47*, despite the absence of open chromatin markers.

We observed striking commonalities between the autosomal MSR D4Z4 and the X-chromosomal MSR CT47. In normal somatic cells, hESCs and hEBs, both MSRs are highly methylated and have high levels of repressive chromatin markers. In the adult body, *DUX4*,[37] encoded by D4Z4, and *CT47* transcripts are highly abundant in the testis and ovary. In disease conditions, like FSHD or cancer, there is a loss of DNA methylation and repressive chromatin markers, but no increase in permissive chromatin modifications, leading to low or leaky expression levels of the MSR-encoded genes. Like *CT47* expression in SCLCs, evidence was presented that *DUX4* is expressed in malignant adult and fetal human rhabdomyosarcoma cell lines,[38] and therefore *DUX4* also fits the definition of a CTA. The shared features of epigenetic regulation and expression patterns of CT47 and D4Z4 suggest that other MSRs may also be regulated similarly. A recent publication shows that indeed the family of MSRs exclusively expressed in the germ line and testis is expanding and includes new members like TAF11-Like, PRR20 and ZAV.[9] Understanding the epigenetic organization of MSRs in health and disease, which cover a high portion of the human genome, may therefore significantly advance our understanding of the epigenetic map of the human genome and of germline biology.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by institutional funds from the LUMC, the Netherlands Organization for Scientific Research (NWO 917.56.338); Breakthrough Project Grant by the Netherlands Genomics Initiative (NWO 93.51.8001); and the National Institutes of Health (P01NS069539). We thank Peter Thijssen for critical reading of the manuscript and Yvonne Krom with helping the DNA methylation analysis.

1 Wijmenga C, Hewitt JE, Sandkuijl LA *et al*: Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet* 1992; **2**: 26–30.
2 Costa FF, Le BK, Brodin B: Concise review: cancer/testis antigens, stem cells, and cancer. *Stem Cells* 2007; **25**: 707–711.
3 Chadwick BP: DXZ4 chromatin adopts an opposing conformation to that of the surrounding chromosome and acquires a novel inactive X-specific role involving CTCF and antisense transcripts. *Genome Res* 2008; **18**: 1259–1269.
4 Warburton PE, Hasson D, Guillem F, Lescale C, Jin X, Abrusan G: Analysis of the largest tandemly repeated DNA families in the human genome. *BMC Genom* 2008; **9**: 533.
5 van Deutekom JC, Wijmenga C, van Tienhoven EA *et al*: FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum Mol Genet* 1993; **2**: 2037–2042.
6 Saitoh Y, Miyamoto N, Okada T *et al*: The RS447 human megasatellite tandem repetitive sequence encodes a novel deubiquitinating enzyme with a functional promoter. *Genomics* 2000; **67**: 291–300.
7 Liao D, Pavelitz T, Kidd JR, Kidd KK, Weiner AM: Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the RNU2 locus) involves rapid intrachromosomal homogenization and rare interchromosomal gene conversion. *EMBO J* 1997; **16**: 588–598.
8 Bruce HA, Sachs N, Rudnicki DD *et al*: Long tandem repeats as a form of genomic copy number variation: structure and length polymorphism of a chromosome 5p repeat in control and schizophrenia populations. *Psychiatr Genet* 2009; **19**: 64–71.
9 Tremblay DC, Alexander Jr G, Moseley S, Chadwick BP: Expression, tandem repeat copy number variation and stability of four macrosatellite arrays in the human genome. *BMC Genom* 2010; **11**: 632.
10 van Overveld PG, Lemmers RJ, Sandkuijl LA *et al*: Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat Genet* 2003; **35**: 315–317.
11 Zeng W, de Greef JC, Chen YY *et al*: Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). *PLoS Genet* 2009; **5**: e1000559.
12 Bairwa NK, Mohanty BK, Stemenova R, Curcio MJ, Bastia D: The intra-S phase checkpoint protein Tof1 collaborates with the helicase Rrm3 and the F-box protein Dia2 to maintain genome stability in Saccharomyces cerevisiae. *J Biol Chem* 2010; **286**: 2445–2454.
13 Caballero OL, Chen YT: Cancer/testis (CT) antigens: potential targets for immunotherapy. *Cancer Sci* 2009; **100**: 2014–2021.
14 Hofmann O, Caballero OL, Stevenson BJ: Genome-wide analysis of cancer/testis gene expression. *Proc Natl Acad Sci USA* 2008; **105**: 20422–20427.

15 Pfeifer GP, Rauch TA: DNA methylation patterns in lung carcinomas. *Semin Cancer Biol* 2009; **19**: 181–187.
16 De Smet C, Loriot A: DNA hypomethylation in cancer: epigenetic scars of a neoplastic journey. *Epigenetics* 2010; **5**: 206–213.
17 Luetkens T, Schafhausen P, Uhlich F *et al*: Expression, epigenetic regulation, and humoral immunogenicity of cancer-testis antigens in chronic myeloid leukemia. *Leuk Res* 2010; **34**: 1647–1655.
18 Xiao L, Wang Y, Zhou Y *et al*: Identification of a novel human cancer/testis gene MAEL that is regulated by DNA methylation. *Mol Biol Rep* 2010; **37**: 2355–2360.
19 Chen YT, Iseli C, Venditti CA, Old LJ, Simpson AJ, Jongeneel CV: Identification of a new cancer/testis gene family, CT47, among expressed multicopy genes on the human X chromosome. *Genes Chromosomes Cancer* 2006; **45**: 392–400.
20 Carney DN, Gazdar AF, Bepler G *et al*: Establishment and identification of small cell lung cancer cell lines having classic and variant features. *Cancer Res* 1985; **45**: 2913–2923.
21 Thomson JA, Itskovitz-Eldor J, Shapiro SS *et al*: Embryonic stem cell lines derived from human blastocysts. *Science* 1998; **282**: 1145–1147.
22 Ware CB, Nelson AM, Blau CA: A comparison of NIH-approved human ESC lines. *Stem Cells* 2006; **24**: 2677–2684.
23 Silva SS, Rowntree RK, Mekhoubad S, Lee JT: X-chromosome inactivation and epigenetic fluidity in human embryonic stem cells. *Proc Natl Acad Sci USA* 2008; **105**: 4820–4825.
24 Lewin J, Schmitt AO, Adorjan P, Hildmann T, Piepenbrock C: Quantitative DNA methylation analysis based on four-dye trace data from direct sequencing of PCR amplificates. *Bioinformatics* 2004; **20**: 3005–3012.
25 Nelson JD, Denisenko O, Bomsztyk K: Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nat Protoc* 2010; **1**: 179–185.
26 Barrand S, Andersen IS, Collas P: Promoter-exon relationship of H3 lysine 9, 27, 36 and 79 methylation on pluripotency-associated genes. *Biochem Biophys Res Commun* 2010; **401**: 611–617.
27 Glazer CA, Smith IM, Ochs MF *et al*: Integrative discovery of epigenetically derepressed cancer testis antigens in NSCLC. *PLoS One* 2009; **4**: e8189.
28 Bai G, Liu Y, Zhang H, Su D, Tao D, Yang Y, Ma Y, Zhang S: Promoter demethylation mediates the expression of ZNF645, a novel cancer/testis gene. *BMB Rep* 2010; **43**: 400–406.
29 Singh A, Greninger P, Rhodes D, Koopman L, Violette S, Bardeesy N, Settleman J: A gene expression signature associated with 'K-Ras addiction' reveals regulators of EMT and tumor cell survival. *Cancer Cell* 2009; **15**: 489–500.
30 Coe BP, Lee EH, Chi B *et al*: Gain of a region on 7p22.3, containing MAD1L1, is the most frequent event in small-cell lung cancer cell lines. *Genes Chromosomes Cancer* 2006; **45**: 11–19.
31 Gjerstorff MF, Harkness L, Kassem M *et al*: Distinct GAGE and MAGE-A expression during early human development indicate specific roles in lineage differentiation. *Hum Reprod* 2008; **23**: 2194–2201.
32 Hall LL, Byron M, Butler J *et al*: X-inactivation reveals epigenetic anomalies in most hESC but identifies sublines that initiate as expected. *J Cell Physiol* 2008; **216**: 445–452.
33 Shen Y, Matsuno Y, Fouse SD *et al*: X-inactivation in female human embryonic stem cells is in a nonrandom pattern and prone to epigenetic alterations. *Proc Natl Acad Sci USA* 2008; **105**: 4709–4714.
34 Marks H, Chow JC, Denissov S *et al*: High-resolution analysis of epigenetic changes associated with X inactivation. *Genome Res* 2009; **19**: 1361–1373.
35 Bernstein BE, Mikkelsen TS, Xie X *et al*: A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 2006; **125**: 315–326.
36 Leeb M, Pasini D, Novatchkova M, Jaritz M, Helin K, Wutz A: Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev* 2010; **24**: 265–276.
37 Snider L, Geng LN, Lemmers RJ *et al*: Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS Genet* 2010; **6**: e1001181.
38 Kowaljow V, Marcowycz A, Ansseau E *et al*: The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscul Disord* 2007; **17**: 611–623.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)