Fluorescence *in situ* hybridization and chromosomal organization of the human Sirtuin 7 gene

SUSANNE VOELTER-MAHLKNECHT[2], STEPHAN LETZEL[2] and ULRICH MAHLKNECHT[1]

[1]Department of Hematology/Oncology, University of Heidelberg Medical Center, Im Neuenheimer Feld 410, D-69120 Heidelberg; [2]Department of Occupational, Social and Environmental Medicine, University of Mainz, Obere Zahlbacher Str. 67, D-55131 Mainz, Germany

Abstract. Sirtuin 7 (SIRT7) is a member of the sirtuin family of protein deacetylases and is, therefore, a derivative of yeast Silent information regulator 2 (SIR2). SIR2 and its mammalian orthologs play an important role in epigenetic gene silencing, DNA recombination, cellular differentiation and metabolism, and the regulation of aging. In contrast to most sirtuins, SIRT7 does not exert characteristic NAD$^+$-dependent deacetylase activity. We have isolated and characterized the human *Sirt7* genomic sequence, which spans a region of 6.2 kb and which has one single genomic locus. Determination of the exon/intron splice junctions found the full-length SIRT7 protein to consist of 10 exons ranging in size from 71 bp (exon 4) to 237 bp (exon 7). The human *Sirt7* open reading frame encodes a 400-aa protein with a predictive molecular weight of 44.9 kDa and an isoelectric point of 9.80. Characterization of the 5' flanking genomic region, which precedes the *Sirt7* open reading frame, revealed a TATA- and CCAAT-box less promoter that lacks CpG islands. A number of AML-1 and GATA-x transcription factor binding sites were found, which remain to be further evaluated experimentally. Fluorescence *in situ* hybridization analysis localized the human *Sirt7* gene to chromosome 17q25.3; a region which is frequently affected by chromosomal alterations in acute leukemias and lymphomas. Human SIRT7 appears to be most predominantly expressed in the blood and in CD33$^+$ myeloid bone marrow precursor cells, while the lowest levels are found in the ovaries and skeletal muscle. Functional characteristics of SIRT7 are essentially unknown at present and remain to be further elucidated.

Correspondence to: Dr U. Mahlknecht, University of Heidelberg Medical Center, Department of Hematology/Oncology, Im Neuenheimer Feld 410, D-69120 Heidelberg, Germany
E-mail: Ulrich.Mahlknecht@med.uni-heidelberg.de

Introduction

Mammalian histone deacetylases (HDACs) are grouped into four categories, of which three contain non-sirtuin HDACs which include the yeast histone deacetylases, RPD3 (class I HDACs), HDA1 (class II HDACs) and the more recently described HDAC11-related enzymes (class IV HDACs), while one category is composed of the sirtuin protein deacetylases (class III HDACs), which are orthologs to the yeast SIR2 protein. Mammalian SIRT7, which is strongly related to SIRT6 (1), is only distantly homologous to human SIRT1, which is most closely related to *S. cerevisiae* SIR2. The currently known seven human sirtuins have been further subgrouped into four distinct phylogenetic classes: SIRT1, SIRT2, SIRT3 (subclass 1), SIRT4 (subclass 2), SIRT5 (subclass 3) and, finally, SIRT6 and SIRT7 (subclass 4, Fig. 1, Table II) (1,2). Derivatives of the yeast SIR2 histone deacetylase typically share a common catalytic domain, which is highly conserved in organisms ranging from bacteria to humans and which is composed of two distinct motifs that bind NAD$^+$ and the acetyl-lysine substrate, respectively (3,4). The yeast silent information regulator 2 protein (SIR2) is a NAD$^+$-dependent histone deacetylase which hydrolyzes one molecule of NAD$^+$ for every lysine residue that is deacetylated (5). The deacetylase activity of the sirtuins is closely related to their phosphoribosyltransferase activity and requires the presence of highly specific conserved amino-acid residues within the catalytic core of the protein, which are not contained in the mammalian SIRT4, SIRT6 and SIRT7 proteins, which are therefore lacking enzymatic deacetylase activity (2,6-10). SIRT7 is a broadly expressed protein, which is associated with heterochromatic regions and nucleoli, respectively (11), and which is predominantly found in the cell nucleus (56.5%) and in smaller proportions in the cytoplasm (26.1%), mitochondria (13.0%) and cytoskeleton (4.3%) (12). The yeast Sir2 protein, as well as its mammalian derivatives, has been shown to directly modify chromatin and silence transcription (7,13,14), to modulate the meiotic checkpoint (15) and, as a probable anti-aging effect, to increase genomic stability and suppress rDNA recombination (7,16,17). While yeast SIR2 exclusively targets histone proteins, mammalian SIRT1 has a large and growing list of targets, such as p53 and forkhead transcription factors, which are mammalian homologs of the Daf-16 protein, a key regulator within the insulin signaling pathway (7,18).
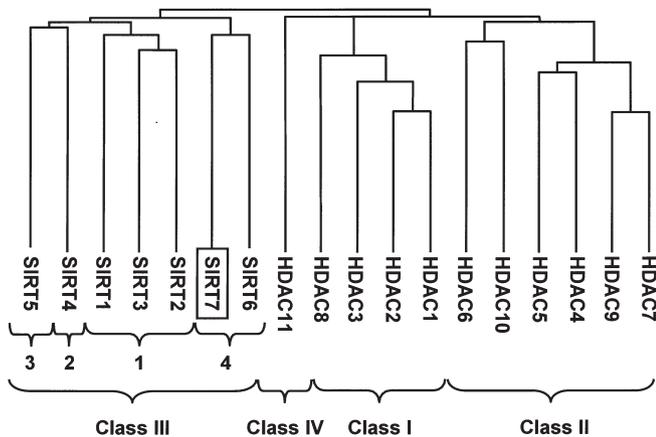
Figure 1. The position of human SIRT7 among the human orthologs for yeast RPD3-, HDA1-, SIR2- and HDAC11-related protein families of histone deacetylases. Accession numbers of the sequences used in this tree: human HDAC1 (GenPept Q13547), human HDAC2 (GenPept Q92769), human HDAC3 (GenPept O15379), human HDAC8 (GenPept AAF73428), human HDAC4 (GenPept AAD29046), human HDAC5 (GenPept AAD29047), human HDAC6 (GenPept AAD29048), human HDAC7 (GenPept AAF04254), human HDAC9 (GenPept AAK66821), human HDAC10 (GenPept AAL30513), human HDAC11 (GenPept NP_079103), human SIRT1 (GenPept AAD40849), human SIRT2 (GenPept AAD40850), human SIRT3 (GenPept AAD40851), human SIRT4 (GenPept AAD40852), human SIRT5 (GenPept AAD40853), human SIRT6 (GenPept AAF43432) and human SIRT7 (GenPept AAF43431).

Calorie restriction is known to induce a metabolic switch that increases the NAD/NADH ratio and/or decreases the levels of nicotinamide (vitamin B3), which is known to be a yeast SIR2 inhibitor, and finally increases rDNA stability and, consequently, longevity (7,19). Mammalian SIRT1 binds, deacetylates and reduces the activity of a number of transcription factors *in vivo*, including MyoD, p53 and FOXO3A, thereby affecting cell differentiation and survival under stress (18,20-25).

In contrast to SIRT1, only little information is currently available on human SIRT7. Human SIRT7 appears to be most predominantly expressed in the blood and CD33+ myeloid bone marrow precursor cells, while the lowest levels are found in the ovaries and skeletal muscle. The functional characteristics of SIRT7 are essentially unknown at present and remain to be further elucidated. The further functional characterization of mammalian SIRT7 may help to further elucidate its potential role in the mediation of stress resistance, anti-apoptosis, anti-aging and as a modulator of endocrine changes. In the study presented herein, we report the chromosomal localization and genomic organization of the human *Sirt7* gene.

Materials and methods

Identification of the human Sirt7 cDNA. A homology search of the EST database at NCBI (National Center for Biotechnology Information) yielded eight positive cDNA clones, of which GenBank clone NM_016538 was obtained from the Reference Center of the German Human Genome Project (RZPD, Berlin, Germany). The authenticity of its insert was confirmed by DNA cycle sequencing.

Identification of BAC genomic clone RZPDB737E052163D. The human *Sirt7* genomic clone was obtained from an arrayed BAC genomic library (Human Genomic Set - RZPD 1.0) after *in silico* screening with *Sirt7* cDNA (GenBank clone NM_016538), which was shown to contain full-length human *Sirt7* cDNA. BAC clone *RZPDB737E052163D* was identified to contain inserts with an average size of approximately 120 kb in vector pBACe3.6, which included the human *Sirt7* genomic sequence. BAC genomic DNA was prepared according to published protocols (26) and the *Sirt7* insert was confirmed by cycle sequencing (27).

Instrumental methods. Dye terminator cycle sequencing was performed using the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA polymerase (Perkin-Elmer, Branchburg, NJ) and analyzed using an ABI PRISM 310 Genetic Analyzer which utilizes four-color sequencing chemistry.

PCR methods. The Sirt7 sequence was partially sequenced by primer walking on both strands using a direct sequencing strategy (27). Sequencing reactions were performed using 0.6 μg cDNA and 20-30 mer oligonucleotide primers (Thermo Electon, Dreieich, Germany). Sequencing reactions were set up in a volume of 20 μl containing 10 pmol of the sequencing primer, 4 μl BigDye Terminator cycle sequencing ready reaction mix (Perkin-Elmer, Norwalk, CT), DNA as indicated and ddH$_2$O added up to a final volume of 20 μl. The thermal cycling profile for the sequencing of the cDNA-clones was as follows: denaturation at 95˚C for 30 sec, annealing at 50˚C for 15 sec, extension at 60˚C for 4 min (25 cycles), and storage at 4˚C.

Chromosomal localization by fluorescence in situ hybridization (FISH). Standard chromosome preparations were used from a human lymphoblastoid cell line. In order to remove excess cytoplasm, slides were treated with pepsin (0.5 mg/ml in 0.01 M HCl, pH 2.0) at 37˚C for 40 min and then washed for 2x10 min in 1X PBS and 1x10 min in 1X PBS/50 mM MgCl$_2$ at room temperature. BAC DNA was labeled by standard nick translation procedure. Digoxigenin (Roche Diagnostics) was used as labeled dUTP at the concentration of 40 μM. Probe length was analyzed on a 1% agarose gel. The probe showed the optimal average length of about 300 bp after nick translation. Approximately 50 ng DNA was pooled with 2 μg cot-1 in 10 μl hybridization buffer (50% formamide, 2X SSC, 10% dextran sulfate). The DNA was applied to chromosomes fixed on a slide, mounted with a cover slip and sealed with rubber cement. Probe DNA and chromosomes were denatured at 72˚C for 3 min. Hybridization was overnight at 37˚C in a wet chamber. After hybridization, the cover slip was carefully removed and the slide was washed in 2X SSC for 8 min. Slides were then incubated at 72˚C in 0.4X SSC/ 0.1% Tween for 1 min. The slide was then washed briefly in 2X SSC at room temperature and stained in DAPI (4',6-diamidino-2-phenylindole) for 10 min. For microscopy, the slide was mounted in antifade solution (Vectashield). *In situ* hybridization signals were analyzed on a Zeiss Axioplan II microscope. Each image plain (blue and orange) was recorded separately using a b/w CCD camera. Chromosomes and FISH

Table I. Exon/intron splice-junctions of the human *Sirt7* gene.

Exon no.	Exon size	5'-Splice-donor	Intron no.	Intron size	3'-Splice-acceptor
1	93	CCTCCGCCAG**gt**acgccgccgc	1	71	gcctgcccgc**ag**GTGTCGCGCA
2	138	GCAGGAGGAG**gc**gagttccgcg	2	110	gcgtcttggc**ag**GTGTGCGACG
3	105	AATCAGCACG**gt**agggagggag	3	1.933	ctcgacatgc**ag**GCAGCGTCTA
4	71	GAAGCGTTAG**gt**aagcgggcca	4	98	tctgcctggc**ag**TGCTGCCGAC
5	73	GCAGAAGCTG**gt**aagagccctg	5	737	tctgccttcc**ag**GTGCAGCATG
6	99	GTACATTGAA**gt**gagcagtcct	6	73	tcttgcctct**ag**GTCTGTACCT
7	237	CAGCCTGAAG**gt**acgtgccgat	7	126	ctgcttccac**ag**GTTCTAAAGA
8	81	GAACCTGCAG**gt**aactcgggtg	8	206	ccctgtcctc**ag**TGGACCCCGA
9	107	CCTATAGCAG**gt**gagtgagccg	9	1.159	ccttttggc**ag**GTGGCAGGAT
10	199				

Exon sequences are given in uppercase and intron sequences are given in lowercase letters. The sizes of the single exons and introns are indicated. Consensus splice donor and splice acceptor sequences are given in bold and are underlined.

Table II. Sequence identity and similarity among class III sirtuin-HDACs.

	Human SIRT1	Human SIRT2	Human SIRT3	Human SIRT4	Human SIRT5	Human SIRT6	Human SIRT7	Yeast SIR2
Human SIRT1	-	42	40	30	28	22	23	40
Human SIRT2	65	-	50	26	27	27	25	31
Human SIRT3	63	66	-	28	31	28	28	35
Human SIRT4	47	43	43	-	27	28	28	25
Human SIRT5	43	44	43	46	-	21	24	26
Human SIRT6	39	44	40	43	36	-	42	23
Human SIRT7	39	42	40	45	37	56	-	21
Yeast SIR2	56	47	49	44	41	38	38	-

The indicated numbers represent the percentage of sequence identity and similarity from pairwise sequence comparisons.

signals were then displayed in false colors and images merged on the computer. Camera control, image capture and merging were performed using SmartCapture X software (Digital Scientific, Cambridge, UK).

Sequence analysis and computer database searches. DNA sequence analysis was performed using the HUSAR (Heidelberg Unix Sequence Analysis Resources) server hosted by the Biocomputing Service Group at the German Cancer Research Center (DKFZ, Heidelberg) and the UniGene and LocusLink programs at the National Center for Biotechnology Information (NCBI). Sequence comparisons were performed with the BLAST algorithm of the GenBank and EMBL databases (28). Protein similarity scores were calculated from fast alignments generated by the method of Wilbur and Lipman using the CLUSTAL W Multiple Alignment Program Version 1.7 (Figs. 1 and 5, Table III) (29). Protein motifs were identified online at the ExPASy (Expert Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics (SIB) using the program, PROSITE, and double-checked using

the MotifFinder program hosted by the GenomeNet WWW server at the Institute for Chemical Research, Kyoto University (Japan), but still remain to be experimentally confirmed. Potential transcription factor binding sites were identified using the TRANSFAC program, which is part of the GenomeNet Computation Service, which is hosted by the Bioinformatics Center at the Institute for Chemical Research at the Kyoto University. Sequence similarities were calculated using GAP software, which considers all possible alignments and gap positions between two sequences and creates a global alignment that maximizes the number of matched residues and minimizes the number and size of gaps on the HUSAR server (30). Repetitive elements were identified on the Repeat Masker Server at the University of Washington and CpG elements were searched using the CPG software hosted by the European Bioinformatics Institute (EMBL outstation) (Figs. 2 and 3).

Phylogenetic analysis. Phylogenetic trees were constructed from known human class I through class IV histone deacetylase sequences, which were obtained from a protein sequence

Table III. Amino-acid sequence alignment of human SIRT7 catalytic domain from different species.

ClustalW Colors mark similarities in protein sequences. Black background: >75% of nucleotides of a column are identical. Grey: more than half of the amino acids of a column are identical or belong to one of the strong groups (amino acids with strong similarities). The sirtuin catalytic domain is indicated by a horizontal box. Accession numbers of the sequences used in this tree: Anopheles gambiae (GenPept XP_321936), Apis mellifera (GenPept XP_397168), Arabidopsis thaliana (GenPept AAP68206), Bos Taurus (GenPept AAF56851), Canis familiaris (GenPept XP_581984), Dario rerio (GenPept XP_540490), Dario rerio (GenPept XP_698800), Drosophila melanogaster (GenPept AAF56851), Gallus gallus (GenPept NP_001026277), Homo sapiens (GenPept NP_057622), Mus musculus (GenPept Q8BKJ9), Neurospora crassa (GenPept XP_328330), Oryza sativa (GenPept XP_471492), Pan troglodytes (GenPept XP_511750), Rattus norvegicus (GenPept XP_221204), Strongylocentrotus purpuratus (GenPept XP_782193), Tetraodon nigroviridans (GenPept CAG06028), Xenopus laevis (GenPept AAH84656), Zea mays (GenPept AAK67144), yeast SIR2 (GenPept P06700).

```
taagtatccatgaagaaaacgtcatgatctgcctcggagtggggagaaacaggccctccttgtccttgtgcagccccactttttgtcctggctgtgggtggccatccatttt
gagcacctctgccaggctgcagacctggaggcgaaggagtgtggaagaggaggtgtggaagaggagggtaaggtcgacgcgtgctgaatgggaggtgcaggtcctcggtcggtggcttggcagaag
aggcttctggagtttcgaggccatacctcctttcaagagcttttggatttgggcgacttcctgggtctaggggagtgttcctgtccctgtcaggtcagcctgcgggagcc
actccacccctcttcccttcccctggatttcctgacatagatggtgtgttcatgtgcaggtgtgagtgtgctgtccttctgatctctgttgatcccgctttcctactgaga
gggtggtggagagggtggtggtctcttctcacctcagtttttctgacacagcgaatcgagggctttggccaaggccaggtctctcaaacttgtccgccaagtaaaccatgat
atatgtcttccttccacctttaactttgccccaactcctcttttatcctaacatggacgctcagatttcagaggaataccccagcagggacaggtcagctctcaccaggagag
tgaggccctggctgctgctcggggtggcactggtgttgctggctcagctcttggtgcatgtttcagatatgtgtgtacacttctgactggtgatcagggccggcgccacgg
ccctgagcggctgcagcctgctttgggacgatggtcaggtcccatgttgtccctgctgcctccctctggagtggccacttggccaaagcgcccatccacatctggaaagtcg
```

AML-1

```
Gagccgcagcctgctgcaccgtttccgctctgggatcttgtgagttacttttttaggct**gtggt**ttggtgaaaggaaccaacacattaacgattttccccccagaagccact
gaataattcttttttggtgtattttgccttcctgttggctgtctggccttggagcaggtcgggagagtagagtacataatcagtgccacacagaggcagggcgtgtctgaga
gactaatgcctgtcctgcctctgggctgttgctgggtgtgtaccatacctgctgggccagtcccacacatttccaccctgccctggagcagcagctttgatacatggt
gtccttgagtctgagctagagtgtgtccccgataaactgtgagacctccctgttcttggtgaccccagttgggctttggccctccagcaacctttgtgtcccagccctgcc
cttcttccaccccctcaatctgctagtccctgaagcctttaaccaaacgggagtgggcacagaaagccttctcctggcaacaggacaagggctgcccgcgtgtccagccct
gttgctctcctggcgctgagaggtgggtccaagcagagttgatcagtccctgcctgccctgcctaggtctagccaggggaggttgtgtgtgtgtgggagaagaccccgggatct
```

GATA-3

```
Gctggggtgggatggaggttatta**aagatctaa**gtgaggagggggctgggggtttggatgcggggtgaggcccgagcgctcacacttcgtgtagggcgggcaggggcctgacct
```

C/EBPalpha

```
cttgcaggggcagtggccttggtgccctacccctgccctgcgcagtatt**tattgctaaatta**ttgtccaggagggggcagcactgggcctggccccccgggtatttattgctgt
acatagtgtatgtttgtgatatataaggtttttctttattttgtatatgatcaataaaccttttaaagagactggcaggtgcttctttggggcgacgcggcgggcggagccgg
```

SP1   NRF-2

```
aaggggtgggt**ggggggcggggcc**catcttgcgcatgcgcgcctgagcgcggcctgccgtgtgaggcggaa**gcggaagag**caggtctccaggggagcg
```

```
M    A    A    G    G    L    S    R    S    E    R    K    A    A    E    R    V    R    R    L    R    E    E    Q    Q    R    E    R
ATG  GCA  GCC  GGG  GGT  CTG  AGC  CGC  TCC  GAG  CGC  AAA  GCG  GCG  GAG  CGG  GTC  CGG  AGG  TTG  CGG  GAG  GAG  CAG  CAG  AGG  GAG  CGC

L    R    Q         71 bp      V    S    R    I    L    R    K    A    A    A    E    R    S    A    E    E    G    R    L    L    A    E
CTC  CGC  CAG   ...Intron1...  GTG  TCG  CGC  ATC  CTG  AGG  AAG  GCG  GCG  GCG  GAG  CGC  AGC  GCC  GAG  GAG  GGC  CGG  CTG  CTG  GCC  GAG

S    A    D    L    V    T    E    L    Q    G    R    S    R    R    R    E    G    L    K    R    R    Q    E    E       110 bp      V
AGC  GCG  GAC  CTG  GTA  ACG  GAG  CTG  CAG  GGC  CGG  AGC  CGG  CGG  CGG  CGC  GAG  GGC  CTG  AAG  CGG  CGG  CAG  GAG  GAG  ...Intron2...  GTG

C    D    D    P    E    E    L    R    G    K    V    R    E    L    A    S    A    V    R    N    A    K    Y    L    V    V    Y    T
TGC  GAC  GAC  CCG  GAG  GAG  CTG  CGG  GGG  AAG  GTC  CGG  GAG  CTG  GCC  AGC  GCC  GTC  CGG  AAC  GCC  AAA  TAC  TTG  GTC  GTC  TAC  ACA

G    A    G    I    S    T      1,933 bp     A    A    S    I    P    D    Y    R    G    P    N    G    V    W    T    L    L    Q    K
GGC  GCG  GGA  ATC  AGC  ACG   ...Intron3...  GCA  GCG  TCT  ATC  CCA  GAC  TAC  CGG  GGC  CCT  AAT  GGA  GTG  TGG  ACA  CTG  CTT  CAG  AAA

G    R    S    V    S     98 bp      A    A    D    L    S    E    A    E    P    T    L    T    H    M    S    I    T    R    L    H
GGG  AGA  AGC  GTT  AG  ...Intron4...  T  GCT  GCC  GAC  CTG  AGC  GAG  GCC  GAG  CCA  ACC  CTC  ACC  CAC  ATG  AGC  ATC  ACC  CGT  CTG  CAT

E    Q    K    L     737 bp      V    Q    H    V    V    S    Q    N    C    D    G    L    H    L    R    S    G    L    P    R    T
GAG  CAG  AAG  CTG  ...Intron5...  GTG  CAG  CAT  GTG  GTG  TCT  CAG  AAC  TGT  GAC  GGG  CTC  CAC  CTG  AGG  AGT  GGG  CTG  CCG  CGC  ACG

A    I    S    E    L    H    G    N    M    Y    I    E      73 bp      V    C    T    S    C    V    P    N    R    E    Y    V    R
GCC  ATC  TCC  GAG  CTC  CAC  GGG  AAC  ATG  TAC  ATT  GAA   ...Intron6...  GTC  TGT  ACC  TCC  TGC  GTT  CCC  AAC  AGG  GAG  TAC  GTG  CGG

V    F    D    V    T    E    R    T    A    L    H    R    H    Q    T    G    R    T    C    H    K    C    G    T    Q    L    R    D
GTG  TTC  GAT  GTG  ACG  GAG  CGC  ACT  GCC  CTC  CAC  AGA  CAC  CAG  ACA  GGC  CGG  ACC  TGC  CAC  AAG  TGT  GGG  ACC  CAG  CTG  CGG  GAC

T    I    V    H    F    G    E    R    G    T    L    G    Q    P    L    N    W    E    A    A    T    E    A    A    S    R    A    D
ACC  ATT  GTG  CAC  TTT  GGG  GAG  AGG  GGG  ACG  TTG  GGG  CAG  CCT  CTG  AAC  TGG  GAA  GCG  GCG  ACC  GAG  GCT  GCC  AGC  AGA  GCA  GAC

T    I    L    C    L    G    S    S    L    K     126 bp      V    L    K    K    Y    P    R    L    W    C    M    T    K    P    P
ACC  ATC  CTG  TGT  CTA  GGG  TCC  AGC  CTG  AAG  ...Intron7...  GTT  CTA  AAG  AAG  TAC  CCA  CGC  CTC  TGG  TGC  ATG  ACC  AAG  CCC  CCT

S    R    R    P    K    L    Y    I    V    N    L    Q     206 bp      W    T    P    K    D    D    W    A    A    L    K    L    H
AGC  CGG  CGG  CCG  AAG  CTT  TAC  ATC  GTG  AAC  CTG  CAG  ...Intron8...  TGG  ACC  CCG  AAG  GAT  GAC  TGG  GCT  GCC  CTG  AAG  CTA  CAT

G    K    C    D    D    V    M    R    L    L    M    A    E    L    G    L    E    I    P    A    Y    S    R     1,159 bp     W    Q
GGG  AAG  TGT  GAT  GAC  GTC  ATG  CGG  CTC  CTC  ATG  GCC  GAG  CTG  GGC  TTG  GAG  ATC  CCC  GCC  TAT  AGC  AG  ...Intron9...  G  TGG  CAG

D    P    I    F    S    L    A    T    P    L    R    A    G    E    E    G    S    H    S    R    K    S    L    C    R    S    R    E
GAT  CCC  ATT  TTC  TCA  CTG  GCG  ACT  CCC  CTG  CGT  GCT  GGT  GAA  GAA  GGC  AGC  CAC  AGT  CGG  AAG  TCG  CTG  TGC  AGA  AGC  AGA  GAG

E    A    P    P    G    D    R    G    A    P    L    S    S    A    P    I    L    G    G    W    F    G    R    G    C    T    K    R
GAG  GCC  CCG  CCT  GGG  GAC  CGG  GGT  GCA  CCG  CTT  AGC  TCG  GCC  CCC  ATC  CTA  GGG  GGC  TGG  TTT  GGC  AGG  GGC  TGC  ACA  AAA  CGC

T    K    R    K    K    V    T    *
ACA  AAA  AGG  AAG  AAA  GTG  ACG  TAA
```

```
tcacgtgctcgatgaagaacagttggcactttgcagatggccagtgtcacggtgaaggctgggttgcccccacgggtctagggagaacgaactctttggggatgacattttc
accgtgacatttttagccatttgtccttgaggaagccccttgcactgctgcggttgtaccctgatacggcctggccatcgaggacacctgcccatccggcctctgtgtcaag
aggtggcagccgcacctttctgtgagaacggaactcgggttatttcagccccggcctgcagagtggaagcgcccagcggcctttcctcgctcaccaggccagtctcagggcc
tcaccgtatttctactactacttaatgaaaaagtgtgaactttatagaatcctctctgtactggatgtgcggcagaggggtggctccgagcctcggctctatgcagacctttt
ttatttct|attaaa|cgtttctgcactggc
```

Figure 2. The complete sequence of *Sirt7* cDNA together with the predicted amino-acid sequence is shown with the location of each intron with respect to the cDNA sequence. Two kb of the 5' upstream promoter region are indicated 5' upstream of the translational start codon. Putative transcription factor binding sites are underlined in bold italics. The translational start (ATG) and stop codons (TAA) are underlined, the polyadenylation signal (ATTAAA) is boxed.

similarity search with the human SIRT7 protein using the BLAST 2.0 program at the NCBI database (Non-redundant GenBank CDS: translations+PDB+SwissProt+SPupdate+PIR). Progressive multiple sequence alignments were performed using the CLUSTAL W Multiple Alignment Program Version 1.7 (31). Trees were calculated and drawn using PATH (Phylogenetic Analysis Task in HUSAR) software, which

estimates and realizes phylogenies by executing the three main phylogenetic methods: distance, parsimony and maximum likelihood, and which is hosted by the HUSAR server from the Biocomputing Service Group at the German Cancer Research Center (Fig. 1). For Fig. 5, trees were calculated and drawn using PileUp software, which computes a multiple sequence alignment using a simplification of the progressive alignment

human *Sirt7* gene (total 6,193 bp)

Figure 3. Genomic organization of the human *Sirt7* gene is shown. Repetitive sequences (here SINEs) are indicated. The 3' untranslated flanking region of exon 10 contains STS-marker STS-stSG31459. The sirtuin catalytic domain overlaps the SIRT7 protein region that is encoded by exons 3 through 8.

method of Feng and Doolittle (32) and which can plot a dendrogram that shows the clustering relationships used to determine the order of the pairwise alignments that together create the final multiple sequence alignment. The distance along the horizontal axis is proportional to the difference between sequences; the distance along the vertical axis has no significance.

Results

Identification and cloning of cDNAs encoding human Sirt7. Homology searches of the dbEST at NCBI (National Center for Biotechnology Information) (28) for the *Sirt7* cDNA sequence (1,2) yielded eight positive cDNA clones: GenBank NM_016538 (1,718 bp), AF233395 (1,718 bp), AK002027 (2,511 bp), AK094326 (1,883 bp), AK131437 (2,914 bp), AK131484 (3,445 bp), AL137626 (1,205 bp) and BC017305 (1,713 bp), of which GenBank clone NM_016538 was obtained from RZPD. The authenticity of its insert was confirmed by DNA cycle sequencing. Sequences flanking the 5'- and 3'-ends of the *Sirt7* open reading frame were identified from the *Sirt7* human genomic clone, BAC *RZPDB 737E052163D*. Characterization of the 5' flanking genomic region, which precedes the *Sirt7* open reading frame, revealed a TATA- and CCAAT-box less promoter and a number of putative optimal transcription factor binding sites for AML-1, GATA, C/EBP-α and SP1. Their biological relevance, however, still awaits to be investigated experimentally. CpG islands have not been found. The 1,718-bp human *Sirt7* mRNA encodes a 400-aa protein with a predictive molecular weight of 44.9 kDa and an isoelectric point of 9.80. Fluorescence *in situ* hybridization analysis localized the human *Sirt7* gene to chromosome 17q25.3. Translational stop codons in all reading frames precede the human *Sirt7* open reading frame. The 3' flanking region was shown to contain the eukaryotic polyadenylation

consensus signal, ATTAAA, 456 bp downstream of the termination of translation signal TAA (Fig. 2) (33).

Identification and characterization of the human Sirt7 genomic locus. The human *Sirt7* genomic clone was obtained from an arrayed BAC genomic library (Human Genomic Set - RZPD 1.0) after *in silico* screening with *Sirt7* cDNA (GenBank clone NM_016538), which was shown to contain full-length human *Sirt7* cDNA. BAC clone *RZPDB737 E052163D* was identified to contain inserts with an average size of approximately 120 kb in the 11.6-kb vector, pBACe3.6, which included the human *Sirt7* genomic sequence. BAC genomic DNA was prepared according to published protocols (26) and the Sirt7 insert was confirmed by cycle sequencing (27). Genomic sequence comparison analyses with the BLAST algorithm helped us with the identification of human chromosome 17 genomic contig RP11-1182P23 (GenBank AC137723), which has been sequenced by the Whitehead Institute/MIT Center for Genome Research (project L28713), is part of the third release of the finished human reference genome, and was assembled from individual clone sequences by the Human Genome Sequencing Consortium together with NCBI. We have used this sequence for the determination of *Sirt7* introns and exon/intron boundaries (Table I). The human *Sirt7* gene spans a region of 6,193 bp. Determination of the exon/intron splice junctions established that the *Sirt7* gene is encoded by 10 exons ranging in size from 71 bp (exon 4) to 237 bp (exon 7). Particularly, within intron 3 we have identified an accumulation of interspersed repetitive elements (SINEs, short interspersed nuclear elements). Additionally, we have identified an internal STS-marker, *stSG31459*, within the untranslated proportion of exon 10 of *Sirt7* between the *Sirt7* translational termination signal (TAA) and the polyadenylation consensus signal (ATTAAA). The sirtuin catalytic domain, which is highly conserved in all members of mammalian
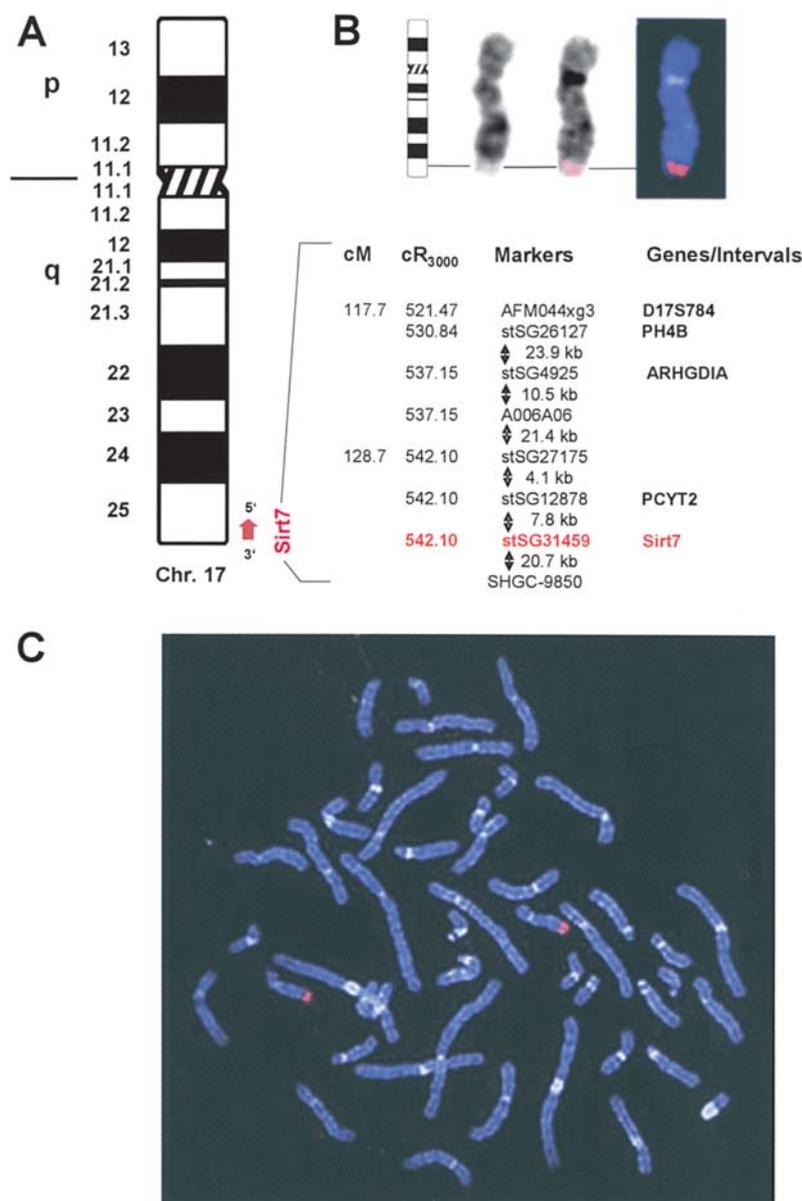
Figure 4. Chromosomal mapping of the human *Sirt7* gene. Lower right panel: fluorescence *in situ* hybridization of BAC clone *RZPDB737E052163D* to human chromosome 17q25.3. (A) Chromosome 17 idiogram according to the International System for Cytogenetic Nomenclature (ISCN 1995), which illustrates the chromosomal position of BAC clone *RZPDB737E052163D* within the interval D17S784 to SHGC-985. Human *Sirt7* is closely neighbored by the PCYT2 gene, which is located 7.8 kb towards the chromosome 17 centromere, and SHGC-9850, which is located 20.7 kb towards the chromosome 17 telomere. The chromosomal orientation of *Sirt7* is shown (red arrow). (B) From left to right: next to the chromosome 17 idiogram, pictures of a DAPI-stained chromosome 17, together with the same chromosome carrying the BAC hybridization signal are illustrated. (C) Fluorescence *in situ* hybridization of BAC clone *RZPDB 737E052163D* to human chromosome 17q25.3.

sirtuins that have been described so far, as well as in their Sir2 yeast ancestor protein, is found between amino-acid residues 107 and 278, i.e. within exons 3 and 8 of the protein (Fig. 3).

Sirt7 is a single copy gene. Both sequencing and results obtained by electronic PCR of BAC clone *RZPDB737 E052163D* identified STS marker STS-stSG31459 at the very 3'-end of the *Sirt7* genomic sequence. Our fluorescence *in situ* hybridization studies localized *Sirt7* to chromosome 17q25.3. These data, together with the results obtained by electronic PCR and the already known location of STS marker *stSG31459*, indicated one single site of hybridization of *Sirt7* on human

metaphase chromosomes and its specific localization on chromosome 17q25.3 (Fig. 4).

Sirt7 expression analyses. In silico expression profile analyses have been carried out using the UniGene EST profile viewer, which is hosted by the NCBI homepage and which suggested the strongest expression of *Sirt7* in the pancreas, followed by testicular tissue, stomach cells and peripheral blood cells on the basis of an analysis of EST counts. Additional analyses were carried out with help of the Human GeneAtlas gene expression database, which is hosted by the Genomics Institute of the Novartis Research Foundation (GNF) and which showed the strongest accumulation of human SIRT7 in the blood and in
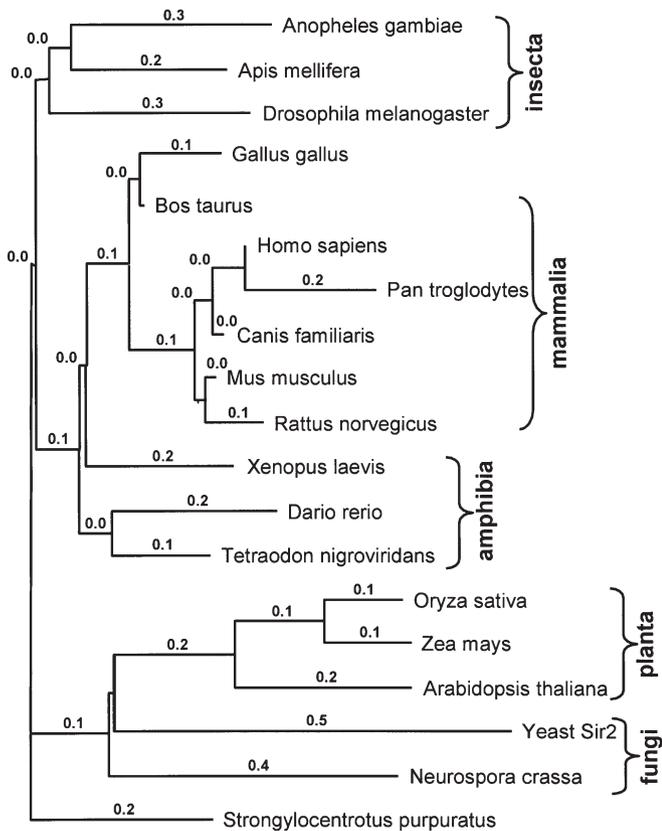
Figure 5. SIRT7 phylogenetic tree. This dendrogram depicts the sequence relatedness of the human SIRT7 protein with the SIRT7 proteins from different species. The GenPept accession numbers correspond to the ones that have also been used for the multiple sequence alignment as shown in Table III.

CD33⁺ myeloid bone marrow precursor cells, while the lowest levels were found in the ovaries and skeletal muscle (34-36).

Phylogenetic analysis and pairwise sequence comparisons. We have screened the expressed sequence tag database (NCBI) with the human SIRT7 protein sequence and identified several yeast and human histone deacetylases which shared a significant degree of sequence identity with SIRT7, indicating a high degree of phylogenetic conservation of protein structure and associated function throughout evolution. A consensus evolutionary tree was obtained using an alignment of yeast SIR2 with a selection of different mammalian and non-mammalian SIRT7 homologs (Fig. 5). In addition, a dendro-gram was calculated from class I through class IV human sirtuin and non-sirtuin HDACs (Fig. 1). Obviously, the sirtuin family of HDACs (class III) does not reveal significant sequence homology with the three classes of non-sirtuin HDACs. The tree was constructed after bootstrapping and clearly identifies four families of human histone deacetylases, with HDAC1, HDAC2, HDAC3 and HDAC8 being members of the yeast RPD3 family of histone deacetylases (so-called 'mammalian class I histone deacetylases'); HDAC4, HDAC5, HDAC6, HDAC7, HDAC9 and HDAC10 being members of the yeast HDA1 family of histone deacetylases ('mammalian class II histone deacetylases'); and SIRT1 through SIRT7 being homologs of the yeast SIR2 protein ('mammalian class

III histone deacetylases'), while HDAC11 is so far the only member of a distinct group of class IV HDACs (Fig. 1).

Discussion

The human *Sirt7* gene encodes a member of the sirtuin family of proteins, which are generally referred to as class III NAD⁺-dependent histone deacetylases on the basis of their homo-logy to the yeast SIR2 protein (5). Their sirtuin deacetylase activity is closely related to their phosphoribosyltransferase activity and requires the presence of specifically conserved amino-acid residues within the sirtuin catalytic core, which are missing in the SIRT7, SIRT6 and SIRT4 proteins and which are, therefore, lacking enzymatic deacetylase activity (2,6-10). The members of the sirtuin family are grouped into four subclasses with SIRT7 being a class 4 sirtuin member (Fig. 1). For most of the currently known human sirtuins, a function has not yet been determined. In yeast, however, sirtuin proteins are known to regulate epigenetic gene silencing and to suppress recombination of rDNA. In addition to their deacetylating activity, the human sirtuins may function as intracellular regulatory proteins with mono-ADP-ribosyl-transferase activity (2). Human SIRT7 appears to be most predominantly expressed in the blood and in CD33⁺ myeloid bone marrow precursor cells, while the lowest levels are found in the ovaries and skeletal muscle (34-37). The protein has been predicted to be predominantly found in the cell nucleus and to a lesser extent in the cytoplasm, in mitochondria and in the cytoskeleton (12).

In the present study, we report the identification, cloning and mapping of *Sirt7* on the genomic level. Human *Sirt7* is a single-copy gene that spans a region of approximately 6.2 kb. It is composed of 10 exons (Fig. 3, Table I) ranging in size from 71 bp (exon 4) to 237 bp (exon 7) and reveals an accumulation of interspersed SINEs (*Alu* repeats) predominantly within intron 3 (38). The SIR2 family domain is highly conserved within all members of mammalian sirtuin proteins that have been described so far and is located within exons 3 through 8 (Fig. 3). The 5' upstream *Sirt7* promoter region revealed a TATA- and CCAAT-box less promoter that lacks CpG islands (Fig. 2). Human *Sirt7* mRNA encodes a 400-aa protein with a predictive molecular weight of 44.9 kDa. Fluorescence *in situ* hybridization analysis in conjunction with electronic PCR localized the human *Sirt7* gene to the subband of chromosome 17q25.3 (Fig. 4); a region which is frequently affected by chromosomal alterations in acute leukemias and lymphomas (Fig. 6). These data were retrieved from the Cancer Genome Anatomy Project (CGAP) database at the National Cancer Institute (39).

It is currently not clear to what extent chromosomal abnormalities that involve the chromosome 17q25.3 chromo-somal region influence SIRT7-mediated functional effects. It is, however, evident that a number of sirtuin proteins are located within chromosomal regions that are particularly prone to chromosomal breaks. In such cases, gains and losses of chromosomal material may affect the availability of functionally active sirtuin proteins, which in turn disturbs the tightly controlled intracellular equilibrium of protein acetylation and/or ADP ribosylation, respectively (40). Protein acetylation modifiers are therefore gaining increasing attention as potential
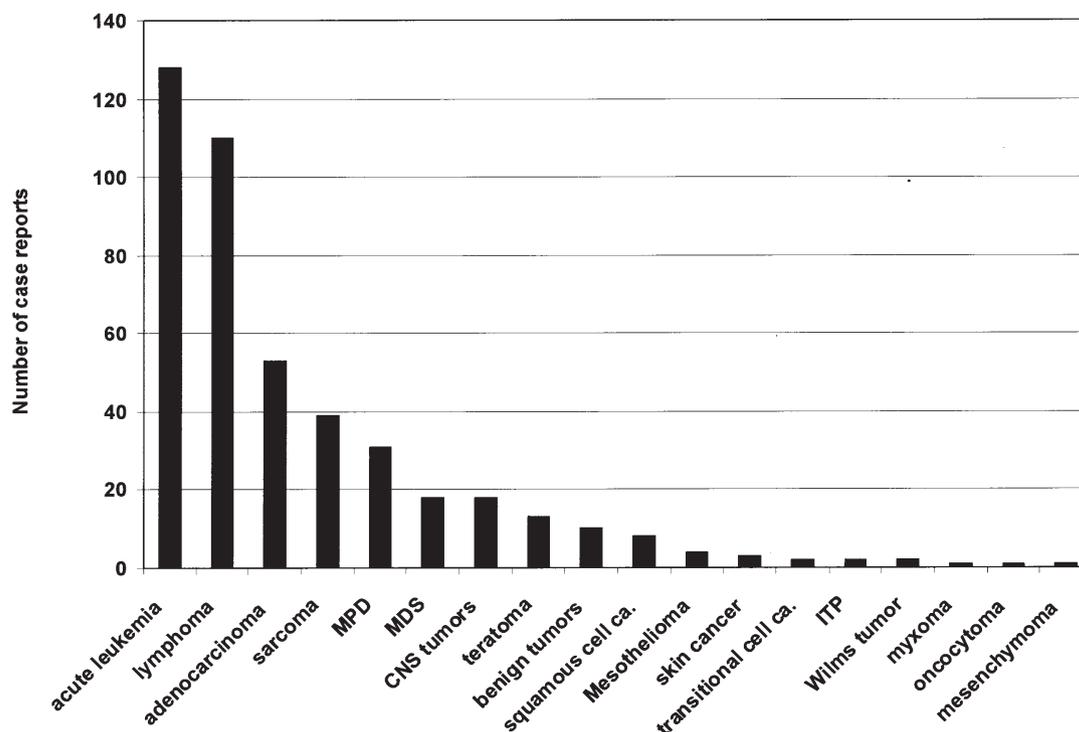
Figure 6. The chromosome region, 17q25.3, is a section which has been found to be involved in various malignancies. Out of 444 case reports which were retrieved from the Cancer Genome Anatomy Project (CGAP) database at the National Cancer Institute (45), the most frequent mutations and chromosomal alterations were found in acute leukemias (128/444, with 88 AML, 40 ALL), lymphoma (110/444, with 9 CLL, 6 mycosis fungoides, 2 Burkitt lymphomas, 2 Hodgkin's diseases, 25 multiple myelomas, 4 T-cell lymphomas, 61 B-cell lymphomas, 1 gastrointestinal stromal tumor), adeno-carcinomas (53/444), sarcomas (39/444 with 1 Ewing sarcoma, 1 liposarcoma, 33 soft tissue sarcomas, 4 osteosarcomas), myeloproliferative diseases (31/444, with 30 CML, 1 osteomyelofibrosis), myelodysplastic syndromes (18/444), central nervous system tumors (18/444, with 3 meningiomas, 7 astrocytoma, 3 primitive neuroectodermal tumors, 5 neuroblastomas), 13/444 teratomas, benign tumors (10/444 with 4 lipomas, 1 papilloma, 3 adenomas, 2 leiomyomas), 8/444 squamous cell carcinomas, mesotheliomas (4/444), skin cancer (3/444 with 1 basal cell carcinoma, 2 malignant melanomas), transitional cell carcinoma (2/444), idiopathic thrombocytopaenic purpura (2/444), Wilms tumor (2/444), myxoma (1/444), oncocytoma (1/444), and mesenchymoma (1/444).

targets in the treatment of cancer. Relaxation of the chromatin fiber facilitates transcription and is regulated by two competing enzymatic activities, histone acetyltransferases (HATs) and histone deacetylases (HDACs), which modify the acetylation state of histone proteins and other promoter-bound transcription factors. While HATs, which are frequently part of multisubunit coactivator complexes, lead to the relaxation of chromatin structure and transcriptional activation, HDACs tend to associate with multisubunit corepressor complexes, which results in chromatin condensation and the transcriptional repression of specific target genes.

Unfortunately it is currently not possible to assess to what extent human SIRT7 plays a role in the pathogenesis of hematological malignancies and acute myeloid leukemia. It is, however, evident that SIRT7 contains multiple repetitive elements at the genomic level, which makes the region particularly prone to chromosomal breaks, while it is located within a chromosomal region that is known to be frequently part of chromosomal alterations in acute leukemia. In the context of such chromosomal modifications that may involve SIRT7, the SIRT7 protein could potentially be either missing, dysfunctional or exhibit its enzymatic activity at wrong times in the wrong places and therefore contribute to an imbalance of the intracellular acetylation status and to the development of disease. The further characterization of the functional role of human SIRT7 is therefore likely to become an exciting endeavor.

Acknowledgements

References

1. Frye RA: Phylogenetic classification of prokaryotic and eukaryotic Sir2-like proteins. Biochem Biophys Res Commun 273: 793-798, 2000.
2. Frye RA: Characterization of five human cDNAs with homology to the yeast SIR2 gene: Sir2-like proteins (sirtuins) metabolize NAD and may have protein ADP-ribosyltransferase activity. Biochem Biophys Res Commun 260: 273-279, 1999.
3. Brachmann CB, Sherman JM, Devine SE, Cameron EE, Pillus L and Boeke JD: The SIR2 gene family, conserved from bacteria to humans, functions in silencing, cell cycle progression and chromosome stability. Genes Dev 9: 2888-2902, 1995.
4. Voelter-Mahlknecht S, Ho AD and Mahlknecht U: FISH-mapping and genomic organization of the NAD-dependent histone deacetylase gene, Sirtuin 2 (Sirt2). Int J Oncol 27: 1187-1196, 2005.
5. Imai S, Armstrong CM, Kaeberlein M and Guarente L: Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. Nature 403: 795-800, 2000.
6. Liszt G, Ford E, Kurtev M and Guarente L: Mouse Sir2 homolog SIRT6 is a nuclear ADP-ribosyltransferase. J Biol Chem 280: 21313-21320, 2005.
7. Blander G and Guarente L: The Sir2 family of protein deacetylases. Annu Rev Biochem 73: 417-435, 2004.
8. Tanny JC, Dowd GJ, Huang J, Hilz H and Moazed D: An enzymatic activity in the yeast Sir2 protein that is essential for gene silencing. Cell 99: 735-745, 1999.

9. Tanny JC and Moazed D: Coupling of histone deacetylation to NAD breakdown by the yeast silencing protein Sir2: Evidence for acetyl transfer from substrate to an NAD breakdown product. Proc Natl Acad Sci USA 98: 415-420, 2001.
10. Tanner KG, Landry J, Sternglanz R and Denu JM: Silent information regulator 2 family of NAD-dependent histone/protein deacetylases generates a unique product, 1-O-acetyl-ADP-ribose. Proc Natl Acad Sci USA 97: 14178-14182, 2000.
11. Michishita E, Park JY, Burneskis JM, Barrett JC and Horikawa I: Evolutionarily conserved and non-conserved cellular localizations and functions of human SIRT proteins. Mol Biol Cell 16: 4623-4635, 2005.
12. Dryden SC, Nahhas FA, Nowak JE, Goustin AS and Tainsky MA: Role for human SIRT2 NAD-dependent deacetylase activity in control of mitotic exit in the cell cycle. Mol Cell Biol 23: 3173-3185, 2003.
13. Vaquero A, Scher M, Lee D, Erdjument-Bromage H, Tempst P and Reinberg D: Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin. Mol Cell 16: 93-105, 2004.
14. Fritze CE, Verschueren K, Strich R and Easton Esposito R: Direct evidence for SIR2 modulation of chromatin structure in yeast rDNA. EMBO J 16: 6495-6509, 1997.
15. San-Segundo PA and Roeder GS: Pch2 links chromatin silencing to meiotic checkpoint control. Cell 97: 313-324, 1999.
16. Kaeberlein M, McVey M and Guarente L: The SIR2/3/4 complex and SIR2 alone promote longevity in Saccharomyces cerevisiae by two different mechanisms. Genes Dev 13: 2570-2580, 1999.
17. Shore D, Squire M and Nasmyth KA: Characterization of two genes required for the position-effect control of yeast mating-type genes. EMBO J 3: 2817-2823, 1984.
18. Brunet A, Sweeney LB, Sturgill JF, *et al*: Stress-dependent regulation of FOXO transcription factors by the SIRT1 de-acetylase. Science 303: 2011-2015, 2004.
19. Lin SJ, Kaeberlein M, Andalis AA, *et al*: Calorie restriction extends Saccharomyces cerevisiae lifespan by increasing respiration. Nature 418: 344-348, 2002.
20. North BJ and Verdin E: Sirtuins: Sir2-related NAD-dependent protein deacetylases. Genome Biol 5: 224.1-7, 2004.
21. Motta MC, Divecha N, Lemieux M, *et al*: Mammalian SIRT1 represses forkhead transcription factors. Cell 116: 551-563, 2004.
22. Luo J, Nikolaev AY, Imai S, *et al*: Negative control of p53 by Sir2alpha promotes cell survival under stress. Cell 107: 137-148, 2001.
23. Vaziri H, Dessain SK, Ng Eaton E, *et al*: hSIR2(SIRT1) functions as an NAD-dependent p53 deacetylase. Cell 107: 149-159, 2001.
24. Nemoto S, Fergusson MM and Finkel T: Nutrient availability regulates SIRT1 through a forkhead-dependent pathway. Science 306: 2105-2108, 2004.
25. Cohen HY, Miller C, Bitterman KJ, *et al*: Calorie restriction promotes mammalian cell survival by inducing the SIRT1 de-acetylase. Science 305: 390-392, 2004.
26. Birnboim HC and Doly J: A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7: 1513-1523, 1979.
27. Mahlknecht U, Hoelzer D and Bucala R: Sequencing of genomic DNA. Biotechniques 27: 406-408, 1999.
28. Altschul SF, Madden TL, Schaffer AA, *et al*: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402, 1997.
29. Wilbur WJ and Lipman DJ: Rapid similarity searches of nucleic acid and protein data banks. Proc Natl Acad Sci USA 80: 726-730, 1983.
30. Needleman SB and Wunsch CD: A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48: 443-453, 1970.
31. Thompson JD, Higgins DG and Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680, 1994.
32. Feng DF and Doolittle RF: Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J Mol Evol 25: 351-360, 1987.
33. Fitzgerald M and Shenk T: The sequence 5'-AAUAAA-3' forms parts of the recognition site for polyadenylation of late SV40 mRNAs. Cell 24: 251-260, 1981.
34. Su AI, Wiltshire T, Batalov S, *et al*: A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci USA 101: 6062-6067, 2004.
35. Su AI, Cooke MP, Ching KA, *et al*: Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci USA 99: 4465-4470, 2002.
36. Walker JR, Su AI, Self DW, *et al*: Applications of a rat multiple tissue gene expression data set. Genome Res 14: 742-749, 2004.
37. Yang YH, Chen YH, Zhang CY, Nimmakayalu MA, Ward DC and Weissman S: Cloning and characterization of two mouse genes with homology to the yeast Sir2 gene. Genomics 69: 355-369, 2000.
38. Singer MF, Thayer RE, Grimaldi G, Lerman MI and Fanning TG: Homology between the KpnI primate and BamH1 (M1F-1) rodent families of long interspersed repeated sequences. Nucleic Acids Res 11: 5739-5745, 1983.
39. Mitelman F, Mertens F and Johansson B: A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. Nat Genet: 417-474, 1997.
40. Mahlknecht U, Ottmann OG and Hoelzer D: When the band begins to play: Histone acetylation caught in the crossfire of gene control. Mol Carcinog 27: 268-271, 2000.