



Original article

# Structuring osteosarcoma knowledge: an osteosarcoma-gene association database based on literature mining and manual annotation

Kathrin Poos<sup>1</sup>, Jan Smida<sup>2,3</sup>, Michaela Nathrath<sup>2,3</sup>, Doris Maugg<sup>2,3</sup>, Daniel Baumhoer<sup>2,4</sup>, Anna Neumann<sup>1</sup> and Eberhard Korsching<sup>1,\*</sup>

<sup>1</sup>Institute of Bioinformatics, University of Münster, Münster, Germany, <sup>2</sup>Clinical Cooperation Group Osteosarcoma, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany, <sup>3</sup>Children's Cancer Research Center and Department of Pediatrics, Klinikum rechts der Isar, Technische Universität München, 81664 Munich, Germany and <sup>4</sup>Bone Tumor Reference Center at the Institute of Pathology, University Hospital Basel, Basel, Switzerland

\*To whom correspondence should be addressed. Tel: +49 251 8350651; Fax: +49 251 8353005; Email: korschi@uni-muenster.de.

Citation details: Poos,K., Smida,J., Nathrath,M., *et al.* Structuring osteosarcoma knowledge: an osteosarcoma-gene association database based on literature mining and manual annotation. *Database* (2014) Vol. 2014: article ID bau042; doi:10.1093/database/bau042

Received 8 November 2013; Revised 4 March 2014; Accepted 23 April 2014

## Abstract

Osteosarcoma (OS) is the most common primary bone cancer exhibiting high genomic instability. This genomic instability affects multiple genes and microRNAs to a varying extent depending on patient and tumor subtype. Massive research is ongoing to identify genes including their gene products and microRNAs that correlate with disease progression and might be used as biomarkers for OS. However, the genomic complexity hampers the identification of reliable biomarkers. Up to now, clinico-pathological factors are the key determinants to guide prognosis and therapeutic treatments. Each day, new studies about OS are published and complicate the acquisition of information to support biomarker discovery and therapeutic improvements. Thus, it is necessary to provide a structured and annotated view on the current OS knowledge that is quick and easily accessible to researchers of the field. Therefore, we developed a publicly available database and Web interface that serves as resource for OS-associated genes and microRNAs. Genes and microRNAs were collected using an automated dictionary-based gene recognition procedure followed by manual review and annotation by experts of the field. In total, 911 genes and 81 microRNAs related to 1331 PubMed abstracts were collected (last update: 29 October 2013). Users can evaluate genes and microRNAs according to their potential prognostic and therapeutic impact, the experimental procedures, the sample types, the biological contexts and microRNA target gene interactions. Additionally, a pathway enrichment analysis of the collected genes highlights different aspects of OS progression. OS requires

pathways commonly deregulated in cancer but also features OS-specific alterations like deregulated osteoclast differentiation. To our knowledge, this is the first effort of an OS database containing manual reviewed and annotated up-to-date OS knowledge. It might be a useful resource especially for the bone tumor research community, as specific information about genes or microRNAs is quick and easily accessible. Hence, this platform can support the ongoing OS research and biomarker discovery.

**Database URL:** <http://osteosarcoma-db.uni-muenster.de>

## Introduction

Osteosarcoma (OS) the most common primary malignant tumor of bone frequently affects children and young adolescents (1). It is a complex disease with manifold numerical and structural genomic alterations affecting multiple genes to a varying extent (2). Patients without clinical signs of systematic spread show 5-year survival rates of 60–80% (3), whereas patients with metastasis at diagnosis exhibit 5-year survival rates of 20–30%. Since 1980, the prognosis of patients has more or less stagnated and no significant therapy improvements have been achieved (4).

Massive research in the field of OS is ongoing to assess the prognostic and therapeutic impact of possible biomarkers and altered molecular pathways. For instance, several studies detected frequent genomic alterations of the tumor suppressor genes TP53 and RB1 in OS and correlated these findings with disease outcome (5–7). Other studies identified p-glycoprotein and ezrin that influence the response to chemotherapy and metastatic spread, respectively (8). Recently, attention has been paid to the value of small non-coding microRNAs in the pathogenesis of OS, e.g. the miR-17~92 cluster (9, 10) and miR-9-5p (11, 12). MicroRNAs represent interesting biomarkers for OS, as they are able to simultaneously regulate hundreds of target genes and several molecular pathways (13). However, the prognostic and therapeutic significance neither for distinct genes including their gene products nor for microRNAs has been determined in controlled clinical studies yet (3). The key prognostic determinants are still clinico-pathological factors and include tumor stage (14), patient age, tumor size and location and the response to neoadjuvant chemotherapy (15). Consequently, all patients are treated with multiagent chemotherapy irrespective of its individual efficacy (16). Moreover, new studies about OS are continuously published and complicate the acquisition of information for specific research purposes and questions.

To support the efforts in OS research and biomarker discovery, we constructed the Osteosarcoma Database. It provides a structured and review-like overview on current OS knowledge with the possibility to rank and sort

the literature according to various parameters, including therapeutic and prognostic value of specific genes and microRNAs and the type of samples used. Information of genes and microRNAs in OS was collected by automated literature mining and manual review and annotation of PubMed abstracts. This information was further enriched by determining microRNA–target gene interactions (MTIs) of all collected candidates related to OS.

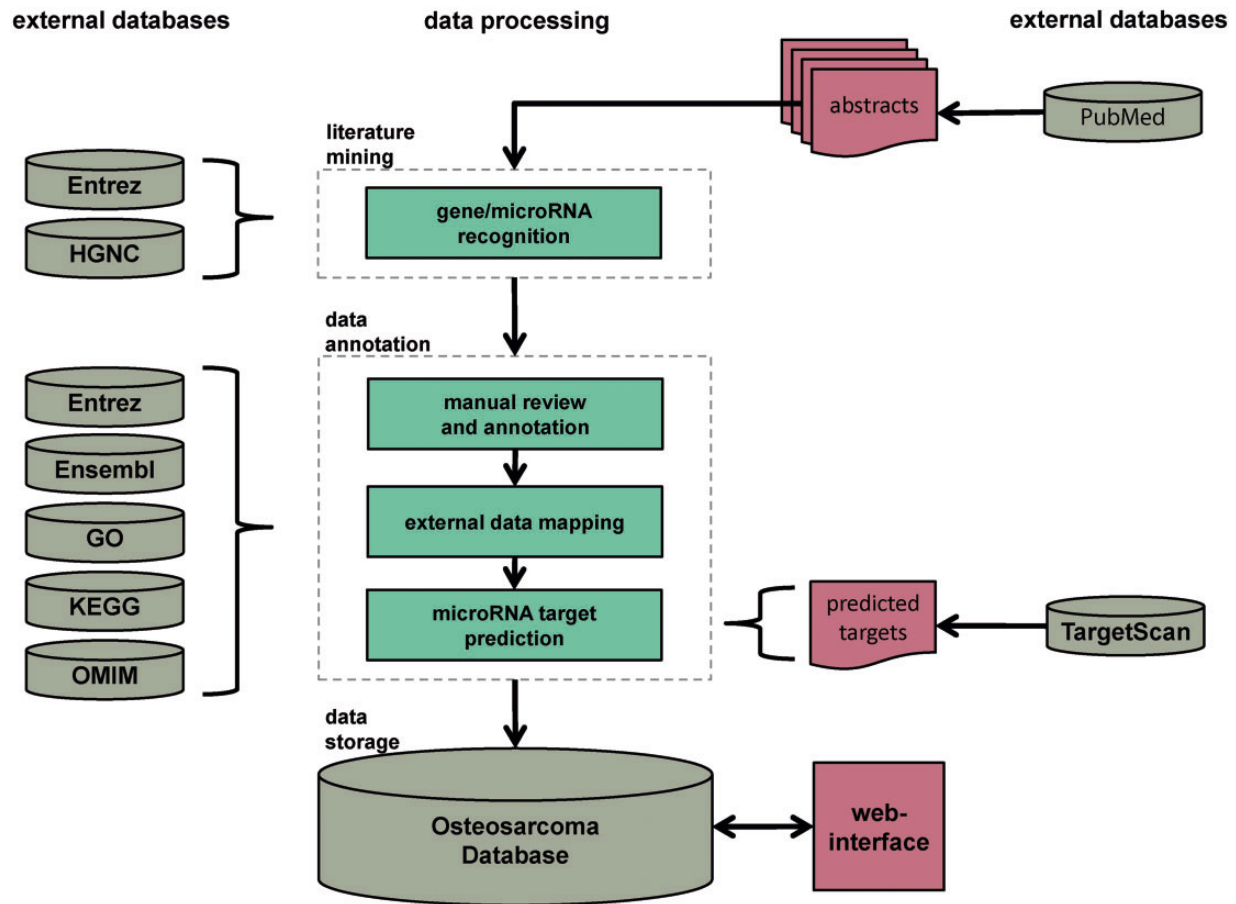
## Database Construction

The Osteosarcoma Database aims to provide a high-quality collection of genes and microRNAs implicated in the pathogenesis of OS, reviewed by experts of the field. The data collection and processing steps are illustrated in [Figure 1](#). The workflow comprised three major steps: automated dictionary-based gene and microRNA recognition, manual review and annotation and data storage. The pipeline was based on PubMed abstracts that contained the keywords ‘osteosarcoma\*’ or ‘osteogenic+sarcoma\*’ in their titles and/or abstracts. They were downloaded with the R package XML (17) via NCBI’s E-utilities. Only abstracts written in English and involving human data or specimens were considered. The last download of abstracts was executed on 29 October 2013. In total, 9908 PubMed abstracts were obtained and served as initial corpus for further processing.

### Dictionary-based gene and microRNA recognition

To reduce the time-consuming process of manual review and annotation, a dictionary-based gene and microRNA recognition was performed on the initial corpus of abstracts.

The dictionary of human genes was compiled from the Human Genome Organisation (HUGO) gene nomenclature committee (18) and the National Center for Biotechnology Information (NCBI) Entrez gene database (19). Official symbols, aliases, synonyms, descriptions, names and database accessions of all genes were combined to generate the gene dictionary with the Entrez geneid as unique identifier. The gene dictionary was extended by



**Figure 1.** Database construction pipeline. The database construction is performed in three major steps: automated dictionary-based literature mining, data review and annotation by reviewers and external data sources and data storage in a MySQL relational database with Web interface. The whole pipeline is based on PubMed-derived abstracts related to OS research.

textual variants of genes (e.g. IL6, IL 6 or IL-6) to be as complete as possible. Ambiguous synonyms and frequent English words according to the stop words function of the R package *tm* (20) were excluded to avoid inaccurate gene recognitions. In case of microRNAs, regular expressions like ‘mir’, ‘miR’, ‘MIR’, ‘miRNA’ and ‘microRNA’ were used for entity recognition. The miRBase (21) accessions of mature microRNA sequences served as unique identifiers.

Genes included in the dictionary were identified in the initial corpus of abstracts by string matching and the microRNAs by regular expressions using the R package *tm* (20). Abstracts without any gene or microRNA occurrence were excluded from further processing, e.g. abstracts of epidemiologic studies. The remaining abstracts were manually reviewed and annotated according to their functional role in the OS.

### Manual review and annotation

During the manual review and annotation step, the reviewers verified the specific genes and microRNAs recognized in the

abstracts. Additionally, information about experimental settings, the biological context and therapeutic and prognostic impact was marked. The experimental settings comprised the experimental procedure, name of cell lines and kind of samples. Abstracts dealing with human OS cell lines but describing anything but OS biology were excluded.

To provide as much information as possible, we mapped OS-related genes and microRNAs to external databases like NCBI Entrez gene (19), Ensembl (22), Online Mendelian Inheritance in Man (OMIM) (23), Gene Ontology (24), Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway (25) and miRBase (21). Furthermore, the OS-related literature derived from PubMed (26) was linked to each gene and microRNA entry.

As microRNA regulation has become a major subject of OS research, we determined possible MTIs between OS-related genes and microRNAs. Predicted microRNA targets were computed by running the local perl scripts *targetscan\_60.pl* and *targetscan\_61\_context\_scores.pl* that were downloaded from the TargetScan Web site (<http://www.targetscan.org/>) (27). Mature microRNA sequences

were gained from miRBase release 20 (21). To obtain high-efficacy targets, we excluded target predictions with a context score  $> -0.1$  (27).

### Data storage

To store and access the collected information on OS-related genes including their gene products and microRNAs, we implemented a database and a user-friendly Web interface. The Osteosarcoma Database is a MySQL relational database. The database scheme is illustrated in [Supplementary Figure S1](#). To easily access OS-related genes and microRNAs, users can search and browse via a Web interface at <http://osteosarcoma-db.uni-muenster.de>. It is built on PHP and JavaScript. For interactive data visualization, we applied tagcanvas (<http://www.goat1000.com/tagcanvas.php>) and cytoscapeweb (28). Alternatively, users can download the Osteosarcoma Database sql file to perform their own queries. The download link is provided at <http://osteosarcoma-db.uni-muenster.de/download.php>.

### Database Description

The Osteosarcoma Database allows retrieving information of candidate genes including their gene products and microRNAs associated with the pathogenesis of OS to support their individual research purposes. Beside gene and microRNA information derived from external databases, manual annotations of OS-related abstracts are provided. Annotations include the number of abstracts focusing on the specific genes with their gene products and microRNAs, the experimental procedures conducted in distinct studies, the potential therapeutic and prognostic value of genes and microRNAs, the specific data types and the biological context investigated. Additionally, regulatory MTIs between collected microRNAs and genes were added. Currently, the database contains 911 genes including their gene products and 81 microRNAs associated with osteosarcoma biology according to 1331 abstracts. Between these microRNAs and genes, we determined 6305 regulatory MTIs due to TargetScan 6 (27).

The database can be searched using the Web interface (<http://osteosarcoma-db.uni-muenster.de>) with two possible input forms depending on the user's research focus. For gene search, Entrez geneids and official gene symbols are accepted. MicroRNAs require miRBase accessions or names of mature microRNA sequences. A search for word components is also possible. After submitting the query, suggestions of genes or microRNAs are presented matching the search term. Users can select their requested entry and the results page is displayed.

The main results page lists general information of the requested gene or microRNA. Underscored entries provide links to respective external databases. Below the general gene or microRNA information, a table marks the abstracts describing the gene's or microRNA's involvement in the pathogenesis of OS. The abstracts can be filtered according to potential therapeutic and prognostic value and according to tumor samples. Further annotation of experimental settings and biological contexts is provided for download using the export button on top of the table. To note, even if the selection of abstracts was initially based on gene names, we also included experiments involving their gene products such as immunohistochemistry and western blots. However, gene symbols are used as unique identifiers for each gene and/or gene product. Moreover, regulatory MTIs of a specific query are accessible via the MTI button on top of the results page. This button directs the user to predicted microRNA target gene networks. For microRNAs, all target genes are visualized, and for genes, the microRNAs that regulate the respective genes are presented. The network can be explored by zooming in and out or drag and drop nodes. Below the network, details of TargetScan predictions are given. [Figure 2](#) illustrates the main results page and the MTI network using the example of the gene CDKN1A.

Alternatively, the user can browse collected genes, microRNAs and abstracts stored in the database. The last column of all browse tables provides a link to the main results page of the respective gene or microRNA. To visually explore genes including gene products frequently mentioned in OS-related literature, a tagcloud of the top genes was implemented. Just genes mentioned in at least five PubMed abstracts are visualized as top genes. By clicking on gene names, the user is again directed to the main results page for the specific gene.

If we miss specific genes or publications about osteosarcoma, users are welcome to suggest them to us via a contact form, and we are pleased to add them to the database. A graphical guide through the Osteosarcoma Database is available for download on the database Web site at <http://osteosarcoma-db.uni-muenster.de/php/tutorial.pdf>.

### Discussion and Future Directions

The ongoing research to detect genes or pathways frequently altered in OS and the search for new therapeutic and prognostic procedures is hampered by the genetic complexity of OS. It becomes even more complicated because of the ever increasing literature about studies of OS that make literature research highly time-consuming. Therefore, it is necessary to structure the existing knowledge of genes and microRNAs associated with OS.

**Osteosarcoma Database**

Home  
Search ← ①  
Browse genes  
Browse miRNAs  
Browse abstracts  
Top genes  
Contact  
Getting help  
Download

### Results for CDKN1A

General Information

Gene ID	1026
Gene Symbol	CDKN1A
Gene Name	cyclin-dependent kinase inhibitor 1A (p21, Cip1)
Gene Type	protein-coding
Cytoband	6p21.2
Entrez ID	ENSG00000124762
miR regulators	10 ← ⑥
Omit ID	116829
GO	GO:000075: cell cycle checkpoint GO:000079: regulation of cyclin-dependent protein kinase activity GO:000089: G1 phase of mitotic cell cycle

Pubmed abstracts associated with CDKN1A

PMID	Title	Tumor	Value
7545620	Retroviral vector-mediated gene transfer of antisense cyclin G1 (CYCG1) inhibits proliferation of human osteogenic sarcoma cells	no	no
8616638	Apoptotic response to oncogenic stimuli cooperative and antagonistic interactions between c-myc and the growth suppressor p53	no	no
9154817	Glucocorticoid receptor-mediated cell cycle arrest is achieved through distinct cell-specific transcriptional regulatory mechanisms	no	no
10239697	Expression of G1 phase regulators in MG-63 osteosarcoma cell line	no	no
10496094	p53-dependent apoptosis or growth arrest induced by different forms of radiation in U2OS cells: p21WAF1/CIP1 repression in UV induced cells	no	no
18931932	Elevated expression of Runx2 as a key parameter in the etiology of osteosarcoma	no	no
19159633	Berberine induces p53-dependent cell cycle arrest and apoptosis of human osteosarcoma cells by inflicting DNA damage	no	no
19455146	Inhibition of Notch pathway prevents osteosarcoma growth by cell cycle regulation	no	no
19559722	Reveratrol induces senescence-like growth inhibition of U-2 OS cells associated with the instability of telomeric DNA and upregulation of BRCA1	no	no

Showing 1 to 25 of 43 entries

③

④

⑤

⑥

⑦

⑧

MicroRNA regulators of target gene CDKN1A

The graph below represents the microRNA regulators of CDKN1A determined by TargetScan6. The gene is represented in red color and its respective microRNA regulators in green. You can drag and drop the network's nodes to take a closer look. Below the graph you can find a table of context scores illustrating the strength of each microRNA target gene regulation.

Detailed prediction results of TargetScan 6 for CDKN1A

Accession	Name	Site Type	Context score
MIHAT0000070	hsa-miR-17	7mer-m8	-0.107
MIHAT0000075	hsa-miR-20a	7mer-m8	-0.107
MIHAT0000082	hsa-miR-63	7mer-m8	-0.107
MIHAT0000089	hsa-miR-193a	7mer-m8	-0.107
MIHAT0000225	hsa-miR-130a	7mer-m8	-0.141
MIHAT0000226	hsa-miR-132	7mer-m8	-0.132
MIHAT0000280	hsa-miR-100b	7mer-m8	-0.107
MIHAT0000281	hsa-miR-130b	7mer-m8	-0.160
MIHAT0001413	hsa-miR-20b	7mer-m8	-0.107
MIHAT0004959	hsa-miR-301b	7mer-m8	-0.141

Showing 1 to 10 of 10 entries

© 2014 Cancer & Complex Systems Research Group, Institute of Bioinformatics, Westphalian Wilhelms-University Münster · [www.ccsr.uni-muenster.de](http://www.ccsr.uni-muenster.de)

**Figure 2.** Screenshot of the CDKN1A results page. The database screenshots show the main results page of a gene search and the corresponding MTI network using the example of CDKN1A. (1) The search menu enables the user to search for a gene or microRNA query. (2) Submitting the query delivers the results page for the specific query that shows general information derived from external databases and abstracts associated with the query. (3) The table of abstracts can be browsed using pagination buttons and (4) filtered according to type of samples, potential prognostic and/or therapeutic value or text search within the titles. (5) To receive more manual annotations like experimental settings, biological context and information about the abstracts, an export button is provided. (6 + 7) The MTI network visually illustrates the possible regulatory relationships of the user's query. A detailed description of the prediction results is given in the table below. (8) Again, users are able to export the table and receive additional information like UTR coordinates and so on.

On that account, we developed the Osteosarcoma Database to supply a review of the current state of OS research and made this information easily accessible to researchers.

### Pathway enrichment analysis on osteosarcoma-related genes

To evaluate the content of the Osteosarcoma Database regarding its functional association to cancer, we performed a KEGG pathway enrichment analysis. All Entrez genes in the human genome were used as a background set. The hypergeometric test was computed to find significantly overrepresented categories (false discovery rate < 0.05). The top 20 enriched pathways are listed in Table 1.

The enrichment results show that the collected OS genes are overrepresented in cancer-related pathways. This

indicates that in OS, many well-known oncogenes (e.g. MYC) and tumor suppressor genes (e.g. TP53 and PTEN) are altered. Furthermore, the TGF $\beta$  signaling pathway is discussed for its contribution to tumor suppression and progression, (29) and the terms apoptosis, cell cycle and focal adhesion represent key signaling pathways in cancer (hallmarks of cancer) (30). Interestingly, we also detected the osteoclast differentiation pathway. In a normal bone, there is a precisely regulated balance between osteoclastic and osteoblastic activity. In OS, this critical balance might be interrupted (31). Taken together, these results indicate OS to require pathways commonly deregulated in cancer as well as to feature OS-specific alterations comprising deregulated osteoclast differentiation.

All properties of OS mentioned earlier are included in the Osteosarcoma Database in terms of OS-related genes, supporting the quality of this collection.

**Table 1.** KEGG pathway enrichment analysis

ID	KEGG pathway	Number of genes	Number of genes in pathway	P-value	FDR <sup>a</sup>
hsa05200	Pathways in cancer	158	327	$5.74 \times 10^{-48}$	$1.11 \times 10^{-45}$
hsa05215	Prostate cancer	59	89	$8.83 \times 10^{-28}$	$8.57 \times 10^{-26}$
hsa05219	Bladder cancer	33	42	$9.62 \times 10^{-20}$	$6.22 \times 10^{-18}$
hsa05212	Pancreatic cancer	44	70	$1.30 \times 10^{-19}$	$6.31 \times 10^{-18}$
hsa04510	Focal adhesion	82	200	$4.58 \times 10^{-19}$	$1.78 \times 10^{-17}$
hsa05222	Small-cell lung cancer	46	85	$8.39 \times 10^{-17}$	$2.62 \times 10^{-15}$
hsa05220	Chronic myeloid leukemia	42	73	$9.45 \times 10^{-17}$	$2.62 \times 10^{-15}$
hsa05210	Colorectal cancer	38	62	$1.46 \times 10^{-16}$	$3.55 \times 10^{-15}$
hsa04110	Cell cycle	58	128	$3.74 \times 10^{-16}$	$8.07 \times 10^{-15}$
hsa04350	TGF-beta signaling pathway	44	85	$3.55 \times 10^{-15}$	$6.89 \times 10^{-14}$
hsa05223	Non-small-cell lung cancer	33	54	$1.72 \times 10^{-14}$	$3.04 \times 10^{-13}$
hsa04115	p53 signaling pathway	38	69	$2.01 \times 10^{-14}$	$3.25 \times 10^{-13}$
hsa04210	Apoptosis	44	89	$3.12 \times 10^{-14}$	$4.66 \times 10^{-13}$
hsa05214	Glioma	36	65	$7.85 \times 10^{-14}$	$1.09 \times 10^{-12}$
hsa05213	Endometrial cancer	31	52	$2.86 \times 10^{-13}$	$3.70 \times 10^{-12}$
hsa05218	Melanoma	37	71	$4.46 \times 10^{-13}$	$5.41 \times 10^{-12}$
hsa05142	Chagas' disease (American trypanosomiasis)	46	104	$1.37 \times 10^{-11}$	$1.57 \times 10^{-11}$
hsa05221	Acute myeloid leukemia	31	58	$1.53 \times 10^{-11}$	$1.65 \times 10^{-10}$
hsa04380	Osteoclast differentiation	50	128	$4.04 \times 10^{-11}$	$4.12 \times 10^{-10}$
hsa04012	ErbB signaling pathway	39	87	$4.39 \times 10^{-11}$	$4.26 \times 10^{-10}$

The table shows the results of the hypergeometric test of KEGG pathways.

<sup>a</sup>FDR, false discovery rate.

### Prognostic or therapeutic value of genes and microRNAs in osteosarcoma

The ultimate aim of OS research is to understand the molecular mechanism underlying OS biology that would imply the discovery of innovative prognostic and/or predictive biomarkers. The Osteosarcoma Database provides a table that lists the prognostic and/or therapeutic value of genes or microRNAs in corresponding PubMed abstracts. This table can be ranked according to genes or microRNAs with possible impact. Table 2 presents genes and microRNAs that might serve as potential biomarkers in OS. Only genes proposed as candidate markers in at least five studies are listed. As microRNA research is still a young field of research, we list all microRNAs with potential prognostic and predictive impact.

Alkaline phosphatase (ALPL) and lactate dehydrogenase (LDHA) are the only accepted biomarkers with prognostic significance, detectable in the peripheral blood. Concentrations correlate with tumor burden and an adverse outcome (32, 33). Nevertheless, the remaining genes and microRNAs are equally promising candidate markers. For instance, the genes including their gene products EZR and VEGFA are significantly correlated with metastatic spread (8, 34), and the ABCB1 gene coding for the p-glycoprotein seems to be associated with multiple-drug-resistance (8). Additionally, the table shows two members of the microRNA family microRNA-34. These family members are well-characterized tumor suppressors in

many cancers and activate TP53 regulated pathways. This microRNA family was extensively tested for its therapeutic use in several tumors and might be the first microRNA family to reach the clinic (35).

Up to now, the prognostic prediction or therapeutic stratification of OS is not based on biomarkers. However, the table suggests many promising candidates that should be further investigated and sometime enter clinical studies.

### Osteosarcoma-related microRNA target gene regulation

Much attention has been focused on microRNAs in the pathogenesis of OS as a new tool for assisting prognosis or therapy. They function through multiple pathways simultaneously, which is in accordance with the perspective on cancer as a disease affecting the whole cellular system. For the collected data, we determined potential MTIs by using TargetScan 6 (27). All microRNAs affecting the largest number of genes ( $\geq 100$  targets) are shown in Table 3. Again, members of the microRNA family microRNA-34 are listed in the table. They regulate the highest number of target genes collected in the Osteosarcoma Database supporting a crucial role in OS as well as in other cancer types. Further, the remaining microRNAs are also known to function as tumor suppressors or oncomirs, e.g. the microRNA families microRNA-29 and -15. Both families have several members involved in various cancer subtypes (36, 37).

**Table 2.** Most frequent genes and microRNAs with potential therapeutic/prognostic impact

ID	Symbol/Name	Number of abstracts
7157	TP53	26
7422	VEGFA	24
5243	ABCB1	20
2064	ERBB2	14
4193	MDM2	14
5925	RB1	14
7430	EZR	12
249	ALPL	9
1029	CDKN2A	9
632	BGLAP	8
1019	CDK4	8
4609	MYC	7
6678	SPARC	7
595	CCND1	6
4313	MMP2	6
4318	MMP9	6
5743	PTGS2	6
1956	EGFR	5
2353	FOS	5
3939	LDHA	5
4233	MET	5
4288	MKI67	5
MIMAT0000076	hsa-miR-21-5p	2
MIMAT0000092	hsa-miR-92a-3p	1
MIMAT0000232	hsa-miR-199a-3p	1
MIMAT0000267	hsa-miR-210-3p	1
MIMAT0000426	hsa-miR-132-3p	1
MIMAT0000435	hsa-miR-143-3p	1
MIMAT0000447	hsa-miR-134-5p	1
MIMAT0000459	hsa-miR-193a-3p	1
MIMAT0000686 <sup>a</sup>	hsa-miR-34c-5p	1
MIMAT0000689	hsa-miR-99b-5p	1
MIMAT0000737	hsa-miR-382-5p	1
MIMAT0001339	hsa-miR-422a	1
MIMAT0004676 <sup>a</sup>	hsa-miR-34b-3p	1

The table lists the number of OS-related abstracts of the most frequently mentioned genes and microRNAs associated with any possible prognostic or therapeutic value. The ID column lists Entrez geneids for genes and miRBase accessions for microRNAs.

<sup>a</sup>miR-34 family.

As already mentioned, microRNA research is a young field and not much is known about their function in OS. Thus, we provide detailed and up-to-date networks about possible MTIs to researchers for hypothesis generation and testing of individual models.

### Future directions

Currently, the Osteosarcoma Database focuses on genes including their gene products and microRNAs associated with OS development and progression. However, the OS is

**Table 3.** Top OS-related microRNAs

ID	Name	MTI <sup>a</sup>
MIMAT0000255 <sup>b</sup>	hsa-miR-34a-5p	139
MIMAT0000686 <sup>b</sup>	hsa-miR-34c-5p	138
MIMAT0000271	hsa-miR-214-3p	128
MIMAT0000430	hsa-miR-138-5p	127
MIMAT0000080	hsa-miR-24-3p	126
MIMAT0000068 <sup>c</sup>	hsa-miR-15a-5p	122
MIMAT0000417 <sup>c</sup>	hsa-miR-15b-5p	121
MIMAT0000100 <sup>d</sup>	hsa-miR-29b-3p	119
MIMAT0002820 <sup>c</sup>	hsa-miR-497-5p	118
MIMAT0000084	hsa-miR-27a-3p	117
MIMAT0000086 <sup>d</sup>	hsa-miR-29a-3p	117
MIMAT0000461 <sup>c</sup>	hsa-miR-195-5p	117
MIMAT0000069 <sup>c</sup>	hsa-miR-16-5p	116
MIMAT0000763	hsa-miR-338-3p	116
MIMAT0000231	hsa-miR-199a-5p	110
MIMAT0000423	hsa-miR-125b-5p	106
MIMAT0000261	hsa-miR-183-5p	100
MIMAT0000691	hsa-miR-130b-3p	100

The table illustrates the microRNAs regulating most of the genes in the Osteosarcoma Database. All microRNAs regulating  $\geq 100$  targets are denoted. The ID column lists miRBase accessions for mature microRNAs.

<sup>a</sup>MTI, microRNA–target gene interaction.

<sup>b</sup>miR-34 family.

<sup>c</sup>miR-15 family.

<sup>d</sup>miR-29 family.

a complex tumor with a huge amount of genomic instability that influences the expression and function of several genes and microRNAs. Hence, genomic alterations need to be added in future versions. We plan to include already known genomic positions marking regions of copy number variations, allelic imbalances and translocations, as it has been shown that structural chromosomal alterations could be used to predict prognosis at diagnosis (2). Moreover, observations of genome-wide changes from next-generation sequencing studies might further obtain new insights into OS biology and must be added as soon as they are available.

We plan to update the database biannually to provide state-of-the-art knowledge and keep track of improvements in the field. We hope that the Osteosarcoma Database will serve as a platform for information and hypothesis generation for the research community that helps to uncover the complexity of OS.

### Supplementary data

Supplementary Data are available at *Database* Online.

### Acknowledgements

K.P. and E.K. designed the study. K.P. implemented the Osteosarcoma Database and wrote the article. K.P., J.S., M.N., D.M., DB, A.N. and E.K. manually reviewed and annotated the PubMed abstracts and participated in discussions and preparation

of the manuscript. The authors thank Christian Ehrhart from anderthalb.com for support in Web design.

## FUNDING

This work was funded by the Translational Sarcoma Research Network [FKZ 01GM0870 to JS, MN, DM and FKZ 01GM0869 to KP, EK], the European TRANSCAN I consortium - PROspective VALidation of Biomarkers in Ewing Sarcoma for personalized translational medicine both supported by the BMBF [FZK 01KT1310 to K.P.E.K.] and the Foundation for the Preservation of the Basel Bone Tumor Center [to DB]. The authors acknowledge support by Deutsche Forschungsgemeinschaft and Open Access Publication Fund of the University of Münster. Funding for open access charge: Deutsche Forschungsgemeinschaft (DFG) and Open Access Publication Fund of the University of Münster (WWU).

*Conflict of interest.* None declared.

## References

- Picci, P. (2007) Osteosarcoma (osteogenic sarcoma). *Orphanet J. Rare Dis.*, 2, 6.
- Smida, J., Baumhoer, D., Rosemann, M., et al. (2010) Genomic alterations and allelic imbalances are strong prognostic predictors in osteosarcoma. *Clin. Cancer Res.*, 16, 4256–4267.
- Marina, N., Gebhardt, M., Teot, L. et al. (2004) Biology and therapeutic advances for pediatric osteosarcoma. *Oncologist*, 9, 422–441.
- Allison, D.C., Carney, S.C., Ahlmann, E.R. et al. (2012) A meta-analysis of osteosarcoma outcomes in the modern medical era. *Sarcoma*, 2012, 704872.
- Patiño-García, A., Piñero, E.S., Díez, M.Z. et al. (2003) Genetic and epigenetic alterations of the cell cycle regulators and tumor suppressor genes in pediatric osteosarcomas. *J. Pediatr. Hematol. Oncol.*, 25, 362–367.
- Tsuchiya, T., Sekine, K., Hinohara, S. et al. (2000) Analysis of the p16INK4, p14ARF, p15, TP53, and MDM2 genes and their prognostic implications in osteosarcoma and Ewing sarcoma. *Cancer Genet. Cytogenet.*, 120, 91–98.
- Wadayama, B., Toguchida, J., Shimizu, T. et al. (1994) Mutation spectrum of the retinoblastoma gene in osteosarcomas. *Cancer Res.*, 54, 3042–3048.
- Kong, C. and Hansen, M.F. (2009) Biomarkers in Osteosarcoma. *Expert Opin. Med. Diagn.*, 3, 13–23.
- Baumhoer, D., Zillmer, S., Unger, K. et al. (2012) MicroRNA profiling with correlation to gene expression revealed the oncogenic miR-17-92 cluster to be up-regulated in Osteosarcoma. *Cancer Genet.*, 205, 212–219.
- Huang, G., Nishimoto, K., Zhou, Z. et al. (2012) miR-20a encoded by the miR-17-92 cluster increases the metastatic potential of osteosarcoma cells by regulating Fas expression. *Cancer Res.*, 72, 908–916.
- Namløs, H.M., Meza-Zepeda, L.A., Barøy, T. et al. (2012) Modulation of the osteosarcoma expression phenotype by MicroRNAs. *PLoS One*, 7, e48086.
- Poos, K., Smida, J., Nathrath, M. et al. (2013) How MicroRNA and transcription factor co-regulatory networks affect osteosarcoma cell proliferation. *PLoS Comput. Biol.*, 9, e1003210.
- Davis, A.M., Bell, R.S. and Goodwin, P.J. (1994) Prognostic factors in osteosarcoma: a critical review. *J. Clin. Oncol.*, 12, 423–431.
- Enneking, W.F. (1986) A system of staging musculoskeletal neoplasms. *Clin. Orthop. Relat. Res.*, 9–24.
- Clark, J.C.M., Dass, C.R. and Choong, P.F.M. (2008) A review of clinical and molecular prognostic factors in osteosarcoma. *J. Cancer Res. Clin. Oncol.*, 134, 281–297.
- Ta, H.T., Dass, C.R., Choong, P.F.M. et al. (2009) Osteosarcoma treatment: state of the art. *Cancer Metastasis Rev.*, 28, 247–263.
- Lang, D.T. (2013) XML: tools for parsing and generating XML within R and S-Plus. R package version 3.96-1.1.
- Gray, K.A., Daugherty, L.C., Gordon, S.M. et al. (2013) Genenames.org: the HGNC resources in 2013. *Nucleic Acids Res.*, 41, D545–D552.
- Maglott, D., Ostell, J., Pruitt, K.D. et al. (2007) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, 35, D26–D31.
- Feinerer, I., Hornik, K. and Meyer, D. (2008) Text Mining Infrastructure in R. *J. Stat. Software*, 25, 1–54.
- Kozomara, A. and Griffiths-Jones, S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, 39, D152–D157.
- Flicek, P., Amode, M.R., Barrell, D. et al. (2011) Ensembl 2011. *Nucleic Acids Res.*, 39, D800–D806.
- McKusick, V.A. (1998) *Mendelian Inheritance in Man. A Catalog of Human Genes and Genetic Disorders*. Johns Hopkins University Press, Baltimore.
- Gene Ontology Consortium. (2010) The Gene Ontology in 2010: extensions and refinements. *Nucleic Acids Res.*, 38, D331–D335.
- Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, 28, 27–30.
- McEntyre, J. and Lipman, D. (2001) PubMed: bridging the information gap. *CMAJ*, 164, 1317–1319.
- Friedman, R.C., Farh, K.K.H., Burge, C.B. et al. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, 19, 92–105.
- Lopes, C.T., Franz, M., Kazi, F. et al. (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, 26, 2347–2348.
- Derynck, R., Akhurst, R.J. and Balmain, A. (2001) TGF-beta signaling in tumor suppression and cancer progression. *Nat. Genet.*, 29, 117–129.
- Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, 144, 646–674.
- Akiyama, T., Dass, C.R. and Choong, P.F.M. (2008) Novel therapeutic strategy for osteosarcoma targeting osteoclast differentiation, bone-resorbing activity, and apoptosis pathway. *Mol. Cancer Ther.*, 7, 3461–3469.
- Bacci, G., Longhi, A., Ferrari, S. et al. (2004) Prognostic significance of serum lactate dehydrogenase in osteosarcoma of the extremity: experience at Rizzoli on 1421 patients treated over the last 30 years. *Tumori*, 90, 478–484.
- Han, J., Yong, B., Luo, C. et al. (2012) High serum alkaline phosphatase cooperating with MMP-9 predicts metastasis and poor prognosis in patients with primary osteosarcoma in Southern China. *World J. Surg. Oncol.*, 10, 37.



34. Kaya,M., Wada,T., Akatsuka,T. *et al.* (2000) Vascular endothelial growth factor expression in untreated osteosarcoma is predictive of pulmonary metastasis and poor prognosis. *Clin. Cancer Res.*, 6, 572–577.
35. Bader,A.G. (2012) miR-34 - a microRNA replacement therapy is headed to the clinic. *Front. Genet.*, 3, 120.
36. Aqeilan,R.I., Calin,G.A. and Croce,C.M. (2010) miR-15a and miR-16-1 in cancer: discovery, function and future perspectives. *Cell Death Differ.*, 17, 215–220.
37. Wang,Y., Zhang,X., Li,H. *et al.* (2013) The role of miRNA-29 family in cancer. *Eur. J. Cell Biol.*, 92, 123–128.