Integrative Genomic Analysis of Pediatric Osteosarcoma

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Sebastian Tobias Ribi

aus Basel-Stadt

Basel, 2017

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Nancy Hynes
Prof. Dr. Daniel Baumhoer
Prof. Dr. Karl Heinimann

Basel, den 21. Februar 2017

Prof. Dr. Martin Spiess
Dekan

The mind is not a vessel to be filled,
but a fire to be kindled.
— Plutarch

Für meine Familie…

Acknowledgements

First and foremost, I would like to thank Prof. Daniel Baumhoer for supervising my work at the University Hospital Basel and for giving me the opportunity to work on such an exciting and multi-faceted project. His trust, oversight, and granted freedom allowed me to explore various aspects of cancer genomics and to grow as a researcher. The same gratitude I want to express for Dr. Axel Hillmer who was supervising my work during the two years at the Genome Institute of Singapore. It should not been taken for granted that the organization and transition within such a joint-PhD project goes as seamless and smooth as it did. I also owe my deepest gratitude to Prof. Gernot Jundt, who was essential in making this project possible, as well as my sister Tanja and Prof. Marius Kränzlin, for laying the foundation of the collaboration. I am also indepted to my many colleagues who supported me during the last four years and with whom I was also lucky enough to build good friendships. I want to especially thank Dr. Michal Kovac for sharing his bioinformatics knowledge with me and for the many facsinating scientific discussions over coffee. Further I am very grateful to fellow nerd, brasseur, and 'the facilitator' Gabriel, whith whom I shared many scientific and technical discussions and who is never too busy to help out in any way possible. I am also thankful for swiss-brasseur Lucas, comrade-in-arms Freddy, and Dr. Lisa for their humor, support and countless merry hours. I further want to thank Dr. Christian Ruiz for his leadership of the Molecular Pathology Unit, and his scientific input, as well as the whole molecular pathology research lab, Ivana, the Davids, Darius, Arthur, Tommy, Charlotte, Valeria, Salvatore, Cristina, Nadia, Tanja, Lara, Vincent, Luigi, and Martin. Also many thanks to Martin, Venkatesh, Katharina, Nicole and Wenjin, who all provided a lively and friendly office atmoshpere. I want to extend the same gratitude to the CTSO4-laboratory in Singapore, Yee Yen, Joanna, Joanne, Alexis, Yao Fei, Audrey, Elaine, Eleanor, Rahul, Asif, and Tushangi. Their friendship, and the helpful and lively environment of the lab made me feel very welcome from the first day and made my stay in Singapore pass by way too quickly. I want to thank them for sharing their wisdom of science and Asian food.
Finally, all words are not enough to express how grateful I am to my family — Christoph, Belinda and Tanja — for their support and for being there during this long journey.

Basel, February 2016 S. R.

Summary

Osteosarcoma (OS) is the most common malignant tumor of bone, affecting predominantly children and young adults. Even though modern treatment regimens including pre- and postoperative multidrug chemotherapy and surgical resection, have additional improved 5 year survival rates to currently 60-70%, no significant improvements have been achieved in the past decades. Furthermore, the systemic chemotherapy applied lacks specificity and can lead to severe adverse effects. New targeted cancer treatment approaches, informed by genomic analysis, aim to exploit molecular properties specific to the neoplastic cells. However, the genomic landscape of OS is complex and is characterized by chromosomal instability, which has historically confounded driver gene discovery. While few and small previous genomic studies exist, a higher sample number and state of the art analysis methods can contribute to a more complete picture of recurring and driving events. To this end, we first characterized the genomes of OS using high-throughput sequencing and single nucleotide polymorphism arrays. We found features of *TP53* intron 1 rearrangements suggesting a highly specific mechanism correlated with transcription. Screening of 288 OS and 1'090 other tumor types revealed evidence for *TP53* rearrangements in 46 (16%) OS, while none were detected in residual tumors, indicating high specificity to OS. Further, we identified a *TP53* intron 1 rearrangement as the causal aberration in a four-generation Li-Fraumeni syndrome family and showed that this rearrangement can at least partially explain the diagnostic gap of formerly classified “TP53 wild-type” Li-Fraumeni syndrome. We then sequenced the exomes of 31 OS and integrated the findings with the corresponding sequencing and single nucleotide polymorphism array data from a replication set of 92 tumors. We identified 14 main driver genes, including some which were formerly unknown in the context of OS. More than 80% of analyzed tumors furthermore exhibited a specific combination of single-base substitutions, LOH, or large-scale genome instability signatures which are characteristic of BRCA1/2-deficient tumors. These BRCA-like traits might be of therapeutic potential since they could render the tumor cells susceptible to PARP inhibitor treatment, which might constitute a novel therapeutic strategy to support current chemotherapy regimens. Last, we screened 337 OS patients for germline alterations and identified *bona fide* pathogenic mutations in the *RET* proto-oncogene in 2% of cases. These mutations appear to combine functional kinase activity with dysfunctional ligand binding and expose affected individuals to an increased risk of developing OS when compared to the normal population (odds ratio 9.12). Our findings add OS to the spectrum of *RET* associated diseases and highlight *RET* as a potential target for multi-targeted tyrosine kinase inhibitors.

Contents

List of Figures

List of Tables

List of Abbreviations

ACMG American College of Medical Genetics and Genomics
AML acute myeloid leukemia
AP alkaline phosphatase
APE1 apurinic endonuclease 1
APL acute promyelocytic leukemia
ATA American Thyroid Association
ATRA all-trans retinoic acid
BAC bacterial artificial chromosome
BAM binary Alignment/Map format
BER base excision repair
BPC Biopathology Center
cDNA complementary DNA
ChIP-seq chromatin immunoprecipitation-sequencing
CML chronic myeloid leukaemia
CNA copy number alteration
COG Childrens Oncology Group
COSMIC Catalogue of Somatic Mutations in Cancer
COSS Cooperative German-Austrian-Swiss Osteosarcoma Study Group
CT computed tomography
DFS disease-free survival
DHFR dihydrofolate reductase
DNA desoxryribonucleic acid
DSB double-stranded DNA break
ERCC excision repair associated
ExAC Exome Aggregation Consortium
FDA Food and Drug Administration
FFPE formalin fixed paraffin embedded
FGF fibroblast growth factor
FISH fluorescence in situ hybridization
FMTC familial medullary thyroid carcinoma
FoSTeS fork stalling and template switching
GLF glial cell line-derived neurotrophic factor
HRR homologous recombination repair
HSCR Hirschsprung's Disease
IHC immunohistochemistry
IL-2 interleukin-2
InDel insertion/deletion
LFS Li-Fraumeni syndrome
LINE long interspersed nuclear elements
LOH loss of heterozygocity
LST large-scale copy number transitions
MEN2 multiple endocrine neoplasia type 2
MLPA multiplex ligation-dependent probe amplification
MMBIR microhomology-mediated break-induced replication
MRI magnetic resonance imaging
mRNA messenger RNA
MSC mesenchymal stem cell
MTC medullary thyroid carcinoma
NaAc sodium acetate
NAHR nonallelic homologous recombination
NCBI National Center for Biotechnology Information
NCI National Cancer Institute
NGS next-generation sequencing
NHEJ non-homologous end joining
NSCLC non-small-cell lung cancer
OS osteosarcoma
PAGE polyacrylamide gel
PARP poly ADP ribose polymerase
PC-1 plasma-cell membrane glycoprotein 1

PCR polymerase chain reaction
PET paired-end tag
PTEN phosphatase and tensin homolog
RFC reduced folate carrier
RNA ribonucleic acid
RT radiotherapy
RT-qPCR real-time quantitative reverse transcription PCR
RTK receptor tyrosine kinase
SAM Sequence Alignment/Map format
SINE short interspersed nuclear elements
siRNA small interfering RNA
SNP single nucleotide polymorphism
SNV single nucleotide variation
SSB single-stranded DNA break
SV structural variation
TGF-beta transforming growth factor beta
TKI tyrosine kinase inhibitor
TMA tissue microarray
TSGA The Cancer Genome Atlas
VEGF vascular endothelial growth factor
VUS variants of unknown significance

Introduction

1.1 Osteosarcoma

Osteosarcoma (OS) is the most common primary malignancy of bone and mainly affects children and young adults. OS can occur in any bone, but is most frequently observed in the long bones around the knee (distal femur, proximal tibia, proximal fibula, 71% of all OS) and in the proximal humerus (12%) [1], where it generally arises near the metaphyseal growth plates. OS in the axial skeleton (skull, ribs, vertebrae, and pelvis) is rare and associated with higher age and poor prognosis [1]. Sarcomas are neoplasms which originate from cells of mesenchymal lineage. Besides OS, sarcomas can be divided into other subclasses, such as chondrosarcoma (consisting of cartilage producing cells), Ewing sarcoma, and soft tissue sarcomas. OS, chondrosarcoma, and Ewing sarcoma together make up 75% of all malignant bone tumors [2].

1.1.1 Epidemiology

Based on data collected by the American National Cancer Institute (NCI) from 1992 to 2004, the incidence rate of OS in children and adolescents is reported as 4.7 per million population per year [3]. OS contributes with 35% to the incidence of primary malignant bone tumors, followed by chondrosarcoma (25%) and Ewing sarcoma (16%) [4]. Overall, OS is the eight most frequent tumor in children and adolescents [5]. However, OS rarely occurs in very young children below the age of 5, which make up only 2% of described patients in the Cooperative German-Austrian-Swiss Osteosarcoma Study Group (COSS) [6]. Children of African American, Asian or Hispanic ethnicity are more often affected than Caucasians [7]. The age at OS diagnosis shows a bimodal distribution with a first peak in the second decade

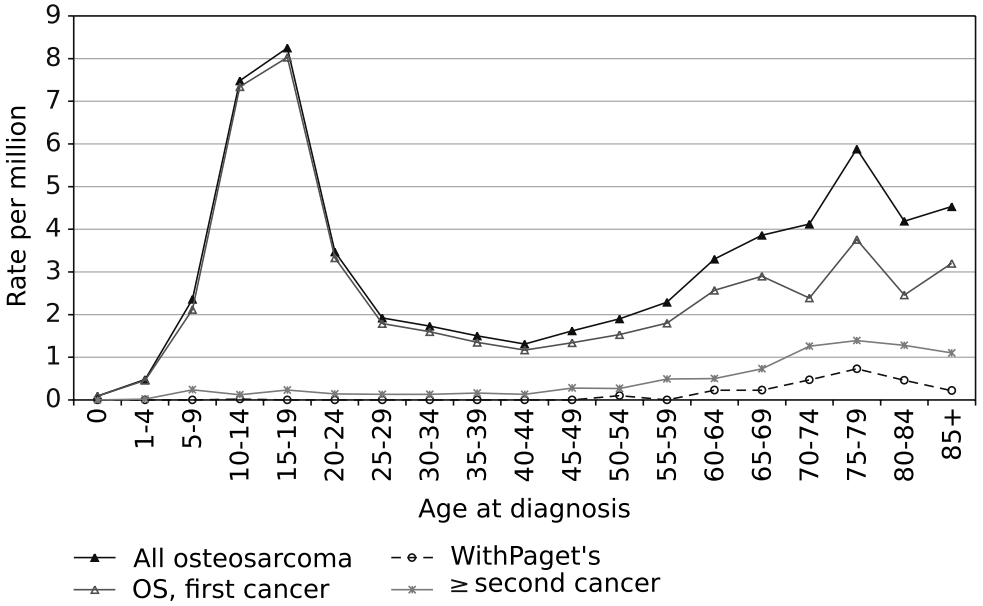


Figure 1 – Incidence of osteosarcoma per million population. Data was derived from the Surveillance, Epidemiology, and End Results (SEER) program on the US population. Modified from Savage *et al.* [9].

of life and a second smaller peak in the seventh and eighth decade (see figure 1 on page 2). In elderly individuals a higher incidence of OS in combination with Paget disease is observed. Likewise, secondary OS is more frequent in older persons, at least partially due to previous radiation treatment [8].

1.1.2 Etiology

Compared to other malignant tumors, the etiology of OS is not well understood. In accordance with the age distribution, skeletal growth has been found to be a host factor for OS. The disease mainly occurs at the metaphyses of the femur and tibia, where the growth plates are located. The first peak of the bimodal age distribution at the onset of puberty coincides with the adolescent growth spurt. This growth spurt is earlier in girls, which is also reflected in an earlier first incidence peak when compared to boys [10, 11]. In canines, giant breeds are more at risk to develop OS than small-sized breeds [12]. Conflicting reports exist on the connection between overall height and OS incidence [13, 14]. The only described exogenous risk factor is exposure to ionizing radiation, both as a result of a therapeutic intervention or environmental exposure. Of all pediatric cancers, patients with Ewing sarcoma are reported to be at highest risk to develop OS as a second primary neoplasm, due to the high irradiation

doses used in the treatment of these patients [15]. However, radiation induced OS appears to be rare and develops after a long interval of 10-20 years [16]. Accordingly, environmental radiation likely plays no major role in the etiology of pediatric OS. Paget disease was identified as a risk factor for secondary OS development [17].

1.1.3 Cell of Origin in Osteosarcoma

Unlike for other neoplasms, such as the basal subtype of breast and prostate cancer for which the cell of origin was experimentally identified as a luminal and basal progenitor cell [18], respectively, the cell of origin in OS has not been identified conclusively. Accumulating results from lineage-specific inducible mouse models have shown that various cell types of the osteogenic lineage can give rise to OS, when the archetypical cancer driver genes *TP53* and *RB1* are targeted with a cell type specific Cre-Lox knockout system. Mesenchymal stem cell (MSC) which attained osteoblast-specific lineage commitment seem to possess higher sarcomagenic potential than less committed cells; $TP53^{-/-}/RB^{-/-}$ mutations in mesenchymal progenitors induced OS with an incidence of 30-60% and in osteoblast precursors in 50-100% of mice [19]. Accordingly, bone marrow derived MSCs with a knockout of *TP53* and *RB1* require some degree of osteogenic differentiation to induce OS, otherwise leiomyosarcoma-like tumors, which usually arise from smooth muscle tissue, were observed [20]. The unsolved question of the originating cell in OS has several implications for research and drug development. Differential gene expression or methylation analysis of cancer cells, for example, require the assumption of a normal cell for comparison and can therefore be difficult to perform in OS studies.

1.1.4 The Clinical Picture of Osteosarcoma

Clinical Presentation

Patients with OS usually experience pain, often during the night, which may increase in intensity when the growing tumor is penetrating the bone cortex (hard outer layer). Due to the coincidence with the adolescent growth spurt, these symptoms might be mistaken as “growing pains” [21], which can lead to a delay in diagnosis. These circumstances can increase the chance of developing metastases before the treatment can be initiated. Physical examination often shows local swelling and occasionally pathologic bone fracture.

Diagnosis

The diagnosis is initially suspected clinically and further substantiated by imaging, such as plain x-ray radiographs, as well as computed tomography (CT), and magnetic resonance imaging (MRI) scans. Radiographs can reveal patterns of aggressive growth, such as an elevated periosteum which gets dislocated from the bone surface by the growing tumor (Codman triangle) or periosteal spikes extending into the surrounding soft tissue (sunburst appearance) [22]. MRI is usually performed to investigate the exact position and extent of the tumor, which is crucial for planning of subsequent biopsy and surgery. Also, other aggressive features such as soft tissue infiltration and skip lesions can be observed. CT of the thorax can reveal possible macroscopic lung metastases [23]. In some cases, increases in alkaline phosphatase (AP) and lactic dehydrogenase can be found [24]. These measurements are, however, of limited diagnostic value since they are also observed in other skeletal diseases, but can correlate with adverse outcome in OS [23].

Tumor Biopsies

To confirm the diagnosis of OS, a biopsy and histopathologic assessment is required. The necessary material can be collected either by an open or core needle biopsy. While the latter method has the advantages of minimal invasiveness, lower costs, and reduced morbidity, core needle biopsies can only capture small amounts of tissue that might not be sufficient to make the diagnosis and generally leave no material for subsequent scientific studies. This makes pretherapeutic OS samples a scarce and precious resource for biomedical studies.

The Histology of Osteosarcoma

OS belongs to the group of osteogenic tumors, which by definition are neoplasms that produce an osteoid or bony matrix [25]. A general categorization of OS can be made according to the site where the tumor arises in the bone and includes intramedullary or central (arising inside the medullary canal) and surface-related types (arising from the bone surface or periosteum) [22]. Histopathology allows the further classification into several subgroups, based on the features of the neoplastic cells and the amount and type of osteoid or other matrix they produce (see figure 2 on page 5). Intramedullary OS, which constitute 80-90% of all OS, include the conventional, telangiectatic, and small cell subtypes [23, 26, 27]. The tumor cells produce bony, cartilaginous, or fibrous matrix and based on the predominant type of extracellular matrix and cell types, conventional OS is divided into osteoblastic, chondroblastic, and fibroblastic subtypes [22]. Unlike conventional OS, the small cell subtype

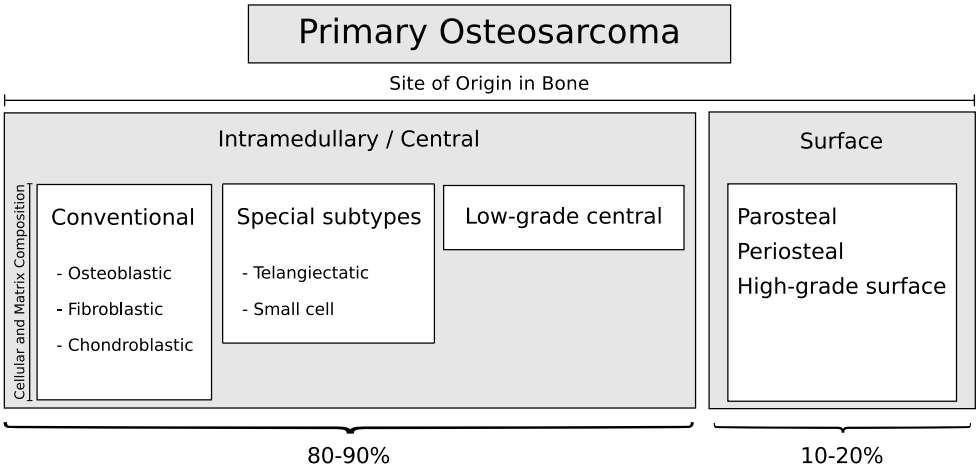


Figure 2 – Classification of primary osteosarcoma. The most common subtypes of OS are shown. General classifications can be made based on the site of origin within the bone. Further histologic subclassifications are defined according to the predominant cellular component and composition of the produced bone matrix.

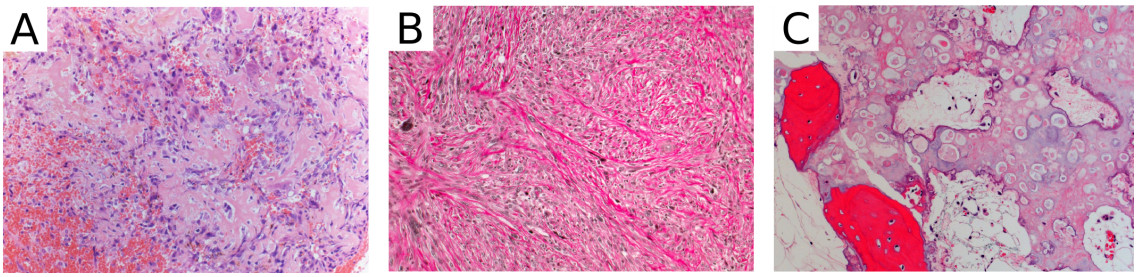


Figure 3 – Subtypes of conventional high-grade osteosarcoma (hematoxylin-eosin staining). **A,** osteoblastic subtype. **B,** fibroblastic subtype. **C,** chondroblastic subtype.

is rare and accounts for only 1-2% of cases. It is characterized by atypically small nuclei with scant amounts of cytoplasm. Small cell OS has a slightly worse prognosis than conventional OS [25]. Histologic differences notwithstanding, current clinical protocols treat all subtypes of high-grade OS identically (see section 1.2.2 on page 6).

Histological Grading

In OS, low- and high-grade tumors based on cellularity, variations in size and shape of nuclei (nuclear pleomorphism), number of mitoses and distinct growth patterns can be distinguished. Some authors also use a three- (G1-G3) or even four-tiered (G1-G4) grading system but the clinically relevant decision has to be made between low- and high-grade. As mentioned above, 80-90% of OS represent conventional or surface associated high-grade lesions. The residual types include low-grade central (G1), parosteal (G1) and periosteal (G2) tumors.

1.1.5 Metastasizing Potential

At the time of diagnosis, 20% of OS cases show detectable metastases, most frequently in the lungs (85-90%) or other bones (8-10%) [28, 29, 30]. Regional metastases affecting the same bone or situated across an adjacent joint are designated as skip lesions [21]. The remaining 80% of cases are considered to have localized disease, although it is know from historical observations (before the introduction of chemotherapy) that 80-90% of these patients die of subsequent systemic spread nevertheless and despite immediate tumor resection. OS is therefore considered to represent a systemic disease already at the time of diagnosis [31]. Even with current treatment protocols, 30-40% of patients with localized OS develop metastases or local recurrences, most of which occur within the first 2-3 years [32].

1.2 Treatment Strategies for Osteosarcoma

1.2.1 The Rise of Chemotherapy

Until the 1970s, the treatment options for OS were limited mainly to amputation, which was associated with very poor survival of 5-20% [33]. Chemotherapy was introduced to OS treatment in the mid 1970s, when doxorubicin and methotrexate were shown to be effective in OS and a combination treatment using the two agents was used in some clinics [34, 35]. While initially controversial, a large multi-institutional trial later showed that the use of bleomycin, cyclophosphamide, dactinomycin, methotrexate, doxorubicin and cisplatin as postoperative multi-agent chemotherapy increased 5-year survival rate to 60-70% [36, 37], whereas patients treated with amputation only showed the historically reported 20% survival [24].

1.2.2 Current Treatment Approaches for Localized Osteosarcoma

Current treatment protocols for localized, high-grade OS are based on surgical resection of the primary tumor and systemic pre- and postoperative multidrug chemotherapy. With this strategy, long-term disease-free survival (DFS) can be achieved in 60-70% of pediatric cases [38] (see figure 4 on page 7). Preoperative (neoadjuvant) chemotherapy can eradicate micrometastases [21] and help to predict treatment response. The response to chemotherapy is analyzed by investigating a complete section through the largest tumor diameter histologically by assessing the percentage of viable tumor. It is still regarded as the gold-standard in prognostic prediction and distinguishes good and poor responders defined as having less or more than 10% viable tumor residues [39]. Low-grade OS are generally only resected and no chemotherapy is applied. Commonly combined agents are methotrexate, doxorubicin,

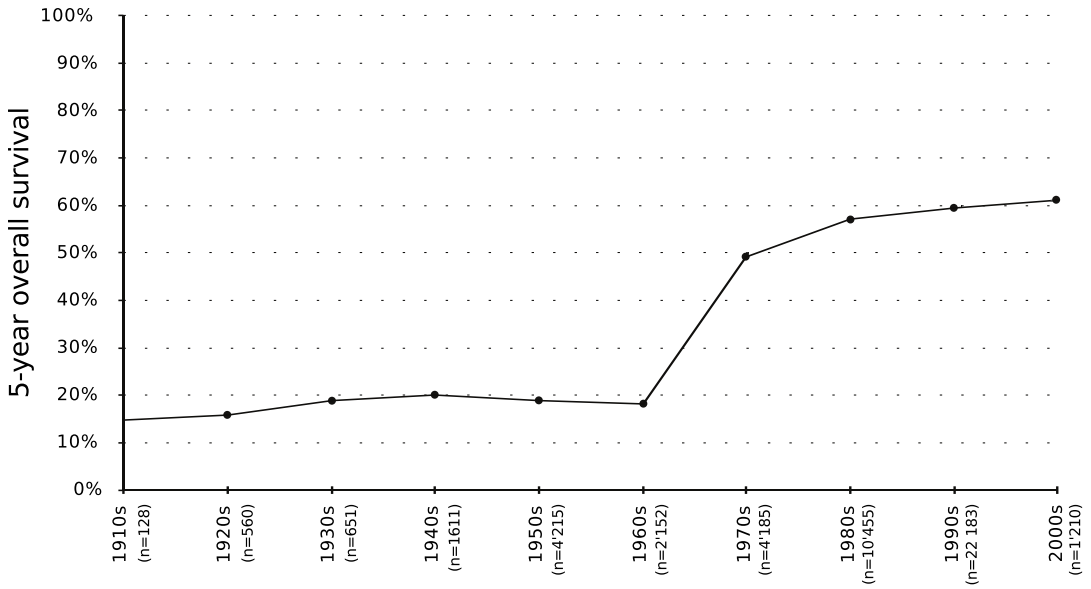


Figure 4 – 5-year overall survival of localized osteosarcoma. A meta-analysis of 5-year survival derived from 264 OS patient series from various clinical institutions over the time span from 1920s to the 2000s is depicted. The numbers of included patients per decade are indicated in parenthesis (47'227 in total). A steep rise in 5-year survival coincides with the introduction of doxorubicin and methotrexate as standard treatment in the 1970s and 1980s. Since then, no statistically significant improvement in survival was recorded. Figure adapted from Allison *et al.* [44].

and cisplatin (MAP regimen). MAP is sometimes extended with ifosfamide and/or etoposide (IE), however, this does not seem to improve patient survival, but increases hematologic toxicity [40]. The recent EURAMOS-1 trial further showed that even patients who responded poorly to preoperative MAP treatment did not benefit when MAP was extended with IE during the adjuvant treatment phase [41, 42]. Reduction of the treatment regimen to two drugs significantly reduces the rate of good responders [43]. Today, surgical resection aims to resect the complete tumor with a wide margin to avoid local recurrence, while conserving the limb. Amputation is always the last considered option due to its invasiveness and the psychological and cosmetic consequences [27].

1.2.3 Current Treatment Approaches for Metastatic Osteosarcoma

In contrast to localized OS, the success in treating (overt) metastatic OS remains limited. Only a few studies that include OS patients with metastatic disease have been reported and patients across different treatment facilities often do not receive consistent treatment. Frequently, patients who qualify for surgical resection are treated with the same chemotherapeutic

protocols as localized OS patients. Resection of lung metastases can prolong survival [45] and is therefore performed whenever possible. However, even complete resection in combination with chemotherapy is generally associated with a poor prognosis and two-thirds of patients relapse due to the development of new lung metastases [46], resulting in a current 5-year survival rate of less than 20% [38].

1.2.4 Prognostic Factors for Survival

Response to preoperative chemotherapy is an important prognostic factor for localized OS, as was demonstrated in a retrospective study assessing 5-year event-free survival of 789 patients [43]. Patients with OS are classified into good and bad responders to preoperative chemotherapy by histologic examination of necrosis in multiple tissue samples of the resected primary tumor. Good responders show necrosis attributable to the chemotherapy of more than 90% of their mass [47]. 63% of patients showed good response to various chemotherapy regimens and 37% responded poorly. Interestingly, the fraction of good responders seems to be significantly correlated to the histologic subtype of OS (63.9% for osteoblastic tumors: 50.6% for chondroblastic tumors, 25% for small-cell tumors, and 61.7% for fibroblastic tumors) [43]. Poor responders have a 5-year overall survival rate of 45-55%, which is substantially lower than the survival rate of 60-70% of good responders [38, 41]. An other study which analyzed an American cohort of 2'849 patients with high-grade OS of all types found that metastatic disease at presentation was an additional strong prognostic factor for decreased survival. Furthermore, axial tumor location, large tumor size of more than 10 cm as well as patient patient age over 25 and male sex were identified to confer poor prognosis [48].

1.3 From Chemotherapy to Targeted Cancer Therapies

The introduction of chemotherapy has greatly improved the survival of patients diagnosed with OS. Nonetheless, unlike other cancers, treatment of OS has not seen any major improvement for several decades (see figure 4 on page 7). Moreover, the principle of systemic chemotherapy bears some intrinsic disadvantages which will be discussed below.

1.3.1 Chemotherapy Lacks Specificity and Results in Adverse Effects

Cytotoxic chemotherapy regimens, like the ones currently used for the treatment of patients with OS, have the disadvantage of being non-specific and can lead to severe side effects. The cause of this becomes clear when looking into the mechanisms with which the used agents achieve their therapeutic effect. Generally, cytotoxic drugs inhibit one of the steps in the process of cell reproduction. This is supposed to affect primarily cancer cells, since one of the hallmarks of cancer is the ability to maintain chronic proliferation [49] but is still toxic for normal cells which also undergo cell division. Methotrexate is a small molecule which inhibits dihydrofolate reductase (DHFR) by competing with folate. Folate is the substrate of DHFR and serves as a cofactor in the *de novo* synthesis of the thymine nucleotides, purine and pyrimidine bases. In these synthesis processes folate is oxidized. DHFR subsequently reduces folate and therefore enables another synthesis cycle. When the activity of DHFR is blocked by methotrexate no thymine nucleotides, purine and pyrimidine bases can be produced and therefore desoxryribonucleic acid (DNA) and ribonucleic acid (RNA) synthesis in the cell are halted, resulting in cell death [50, 51].

Cisplatin introduces DNA damage and interferes with DNA replication. Inside the cell the concentration of chloride is lower than in the extracellular fluid, which leads to a slow replacement of the cisplatin chloride ligands with water. This results in highly reactive aquated cisplatin which covalently cross-links DNA and leads to single-strand DNA breaks and eventually apoptosis [52].

Doxorubicin is a topoisomerase II inhibitor, which can also intercalate with DNA. The resulting enzyme-doxorubicin-DNA complex blocks topoisomerase progression and therefore blocks DNA replication and induces DNA damage, which eventually leads to cell death [53]. The three mechanisms of action described above target mainly DNA replication and synthesis; processes which are most frequently active in - but importantly are not restricted to - fast dividing neoplastic cells, as they do evidently also occur in normal cells. The lack of specificity to cancerous cells can cause a plethora of adverse effects. This is further amplified by the properties of bone tissue. Due to the poor permeability and low blood flow, the chemotherapeutics, which are usually delivered as systemic infusions, need to be administered in relatively high doses to reach a sufficient concentration in the tumor, which can lead to severe toxicity [54]. Toxicity data of 599 OS patients enrolled in the recent EURAMOS-1 trial reported adverse events to chemotherapy in 95% of cases [41].

1.3.2 Resistance to Chemotherapy

As in other malignant tumors, chemotherapy resistance is responsible for the majority of treatment failures in OS [55]. Drug resistance can be either intrinsic or acquired during the course of treatment. The subpopulation of cancer cells which has acquired a mutation leading to increased resistance might possess a growth advantage over non-resistant cells. Several examples of mutations conferring chemotherapy resistance have been described in OS. Firstly, chemoresistance can be achieved by decreasing the intracellular accumulation of the drug. Methotrexate for instance passes through the cell membrane by the reduced folate carrier (RFC). Acquired mutations as well as reduced expression of RFC are associated with Methotrexate resistance and metastases in OS patients [56, 57]. Another way in which neoplasms can decrease intracellular drug concentration is through non-specific removal of the compounds by the membrane efflux pump P-glycoprotein (P-GP), which is encoded by the multidrug-resistance gene *MDR1*. Cell line experiments have shown that high levels of P-GP might be responsible for doxorubicin resistance but attempts to use high P-GP levels as a predictor for survival in OS patients remain inconclusive [58]. Other reported resistance mechanisms are based on enhanced repair of the DNA damage which is inflicted by therapeutics such as cisplatin. Apurinic endonuclease 1 (APE1) is a key player in the base excision repair (BER) pathway and high *APE1* expression levels in OS patients are correlated with reduced survival times [59]. Similar observations have been made with members of the excision repair associated (ERCC) set of proteins which are involved in the nucleotide excision repair pathway; several polymorphisms in *ERCC* genes, on the other hand, are associated with a good response to cisplatin treatment [60].

1.3.3 Improvements of Existing Chemotherapy Drugs

Different strategies have been tried to tackle the problems of adverse effects and resistance to chemotherapy. Existing chemotherapeutics and their mode of delivery can be modified to increase specific targeting of neoplastic cells. A recent innovation is the use of drug nanocarriers and liposomes which, based on surface moieties recognizing specific tumor antigens, the particle size or geometry, can be designed to introduce tropism to tumor cells [61]. While these approaches show encouraging results in animal studies [62] and nanoparticle-based formulations have been approved for breast and ovarian cancers [63, 64], they have not had any impact on OS treatment yet.

Attempts are also made to modify existing chemotherapeutics to circumvent chemoresistance mechanisms found in OS. In one approach, doxorubicin is directly delivered to mitochondria, where it can lead to mitochondrial impairment and cell death. To this end, the drug is conjugated to a hydrophobic peptide with tropism to the mitochondria [65]. This limits

the time the compound remains in the cytoplasm and within reach of the P-GP membrane efflux pump, and therefore increases the drug efficacy. While cell line results have been promising, these insights have also not translated into clinical use, yet. A last notable avenue was explored by a recent animal study, in which the antitumor immune response modulator interleukin-2 (IL-2) was delivered to tumor cells via overexpression in the attenuated and facultative anaerobic *Salmonella typhimurium* strain and showed antitumoral activity against primary OS and lung metastases [66]. *Salmonella* bacteria preferentially replicate in tumors due to their affinity for a hypoxic or anaerobic environment. IL-2 can enhance an antitumoral immune response and has long been used in the treatment of metastatic melanoma, but toxicity is limiting the dose which can be administered systemically. Again, despite promising animal studies no clinical utility has been reported so far.

1.3.4 Principles and Notable Examples of Targeted Cancer Treatment

In contrast to systemic chemotherapy, a targeted approach to cancer treatment aims to target individual molecular properties which are specific to the neoplastic cells. With the advent of molecular analysis methods and the burgeoning of the related fields of genomics and proteomics, new tools came to the disposal of cancer research. These methods allow the characterization of tumors beyond classical pathology, which is largely based on morphologic observations. Ideally, by finding the molecular Achilles' heel of a tumor, drugs can be developed which specifically exploit this property of the tumor cells while sparing normal cells.

One point of attack for targeted cancer therapeutics are pathways which are deregulated in neoplasms. Often, this dysregulation leads to sustained proliferative signaling, a fundamental hallmark of cancer [49]. For example, growth factors and other ligands normally bind to receptor tyrosine kinase (RTK) located on the cell membrane, which in turn activate diverging intracellular signalling pathways. These pathways eventually control vital processes such as cell cycle progression, cell survival, and energy metabolism. Cancer driving mutations in members of these signalling pathways can lead to constitutive activation and thus continuous proliferation. Thus not surprisingly, many drugs have been developed which try to counteract activating mutations of oncogenes or the resulting pathway activation. Conversely, deactivating mutations in tumor suppressor genes are not as easily actionable as oncogenes. However, due to regulatory feedback loops, loss of tumor suppressor gene function can lead to an overall activation of signalling pathways, which then in turn can potentially be targeted by an inhibitor.

Already before the rise of genomics and next-generation sequencing (NGS) a few breakthroughs have been made in the area of targeted cancer treatment. In 1985, all-trans retinoic acid (ATRA) was reported to be effective in targeting the PML-RARA fusion protein, a driver in acute promyelocytic leukemia (APL) [67]. This fusion protein is the result of a balanced reciprocal translocation between the chromosomes 15 and 17 and is present in 70-90% of APL cases [68]. Not unlike in OS, the treatment success of APL with chemotherapy alone had reached a plateau, whereby only 35 to 45% of patients could be cured. Introduction of ATRA in combination with chemotherapy as the first-line treatment increased 6-year DFS to 86 % [67]. Another big success story is the development of imatinib for the treatment of chronic myeloid leukaemia (CML). In 98% of CML patients, the disease is driven by the archetypical *BCR-ABL1* gene fusion, which eventuated from a reciprocal translocation involving chromosomes 22 and 9. As a result of this genomic rearrangement, the non-receptor tyrosine kinase ABL1 is retained in the cytoplasm and is constitutively activated. This leads to the dysregulation of the MAPK, JAK-STAT and PI3K pathways and subsequent uncontrolled cell division. Based on this knowledge, a rational drug discovery program lead to the development of the first therapeutically used kinase inhibitor imatinib and its approval for the treatment of CML [69]. As compared to the previous regimen consisting of interferon-α and chemotherapy, targeted treatment with imatinib increased 5-year survival rates from 30% to 89% and still remains the first-line treatment. The indications have since been extended to primary acute lymphoblastic leukaemia, where 20% of patients harbor the BCR-ABL1 fusion and therefore are responsive to imatinib, and to gastrointestinal stromal tumors (GIST) due to the ability of imatinib to also inhibit the c-KIT tyrosine kinase [70]. A third prominent example of a targeted cancer drug is the tyrosine kinase inhibitor (TKI) vandetanib. While originally developed as a selective inhibitor of the vascular endothelial growth factor (VEGF) receptor-2, the drug was later shown to be an effective multi-kinase inhibitor which is also active against the EGF receptor, the RET oncoprotein and other kinases [71, 72]. As a consequence, vandetanib was approved for the treatment of late-stage medullary thyroid carcinoma (MTC) patients ineligible for resection. Also for targeted cancer therapy, resistance to treatment is a major problem. Common examples are imatinib and the other tyrosine kinase inhibitors which are currently used in the treatment of CML. Resistance is arising primarily due to point mutations in the drug target, which in CML is the kinase domain of the constitutively active ABL1. These mutations interfere with the drug-protein interaction. For ABL1 alone, more than 100 different resistance mutations have been cataloged so far [73]. This resistance can either be intrinsic or acquired during treatment. Various studies have shown that more than 10% of patients do not show any response to imatinib at all, suggesting that resistant cells already existed prior to treatment. Even though more seldom, ATRA treatment for APL also

occasionaly fails, due to *PML* and *RARA* mutations and both intrinsic and acquired resistance has been reported. Taken together, targeted cancer treatment is a promising paradigm, ideally mediating its toxic effect specifically to an individual neoplasm. However, this specificity can also be prone to developing resistance, e.g. due to secondary mutations.

1.3.5 Translational Research Enables Targeted Treatments

Becoming apparent from the examples described above, molecular classification and analysis of tumors precedes the development and implementation of targeted treatment strategies. This is even more emphasized by our growing understanding of intra- and inter-tumoral heterogeneity which mandates in-depth analysis of as many cancer subtypes as possible. Recent advances in the field of massively-parallel NGS along with the computational resources and algorithms needed for storing and interpreting the copious amounts of data facilitate the genomic characterization of tumors and the identification of genomic aberrations which lead to tumorigenesis. Since the completion of the human genome project in 2001 [74] much effort was put into cancer genomics and with projects such as The Cancer Genome Atlas (TSGA) [75] and the Catalogue of Somatic Mutations in Cancer (COSMIC) [76] they start to bear first fruits. Cancer genomics now play an integral part in finding and selecting molecular targets for novel and innovative anticancer drugs [77]. The work presented here aims to leverage current genomic approaches to characterize OS and to identify new actionable mutations with potential clinical utility.

1.3.6 Genomic Aberrations and Available Analysis Methods

In this thesis, several methodologies for the analysis of genomic aberrations have been used and a brief overview over these techniques will be given in this section. Generally, genomic aberrations can be classified into single nucleotide variations, InDels and structural variations (see figure 5 on page 16). Structural variation (SV) include copy number neutral (= balanced) rearrangements such as inversions and translocations, and unbalanced copy number alteration (CNA), such as deletions, insertions, tandem duplications, and unbalanced translocations. Recurring SV have been implicated in various diseases and phenotypes. Prominent examples are a 20 kb deletion upstream of the *IRGM* gene in Crohn's disease [78] or the recurrent 11;22 chromosomal translocation between the *EWS* and *FLI1* genes in Ewing sarcoma [79].

Before the development of NGS and DNA arrays, the available methods for analyzing single nucleotide variation (SNV) and structural changes were very limited. SNV and InDels were detectable with conventional Sanger sequencing, which is still used in routine diagnostics since it is reliable, easy to use and cost effective. Because of the principle of the technique

which only allows the probing of one short genomic region at a time, Sanger sequencing is not scalable to exploratory genomic studies, where multiple genes or complete tumor genomes are to be analyzed. Gross structural anomalies were detected by karyotyping and chromosome banding [80]. With the rather low resolution achieved by these approaches only gross numerical aberrations such as trisomy 21 [81] or subchromosomal SV of several Mb in size can be detected. More sensitivity can be achieved by fluorescence in situ hybridization (FISH). Rather than relying on banding patterns to visually identify chromosomal regions, FISH uses fluorescently labeled DNA probes designed specifically against the locus of interest. While offering more resolution than classical DNA banding, the resolution of FISH is still limited by the size of the bacterial artificial chromosome (BAC)-probes which is usually between 100-300 kb, as well as by optical limitations of microscopy. Small insertions and deletions in the kb range, for example are difficult to identify. Still, low costs and high throughput opportunities make FISH a useful screening tool. A big technological leap for chromosome diagnostics came with the introduction of array-based comparative genomic hybridization (aCGH) in the early 1990s [82, 83]. This technique measures hybridization ratios between test samples and normal DNA, from which CNA can be inferred [84]. A limitation of CGH-based methods lies in the fact that they are not able to detect copy-number neutral regions of loss of heterozygosity. To detect these, single nucleotide polymorphism (SNP) -based arrays can be used. In SNP arrays not only the amount of genetic material from a given locus is assessed, but also the genotype of predefined SNP probes distributed across the genome. Newer iterations of commercially available SNP arrays several millions of probes (e.g. 1.9 million copy number + 750'000 SNP probes in Cytoscan HD), whereby SNP are positioned at specific sites in the genome and the non-polymorphic copy number probes fill the gaps to allow high density coverage. Current arrays have a probe spacing of ~880 bp in most genes and 400-660 bp in known disease-related genes. Typically, SNP arrays can be used to confidently call CNA or loss of heterozygocity (LOH)-segments of ~50 kb.

With the advent of massively parallel NGS, a new tool became available for the study of both SV and SNV. The basic principle of NGS shall not be explained here as it has been done abundantly in the literature [85]. Of note, however, are conceptual varieties in regard to SV detection. With classical single-end sequencing, CNA can be inferred using read depth, which after normalizing for biases introduced by genomic context such as GC-content, reflects the abundance of a given genomic site in the original sample material. Single-end reads are also used to identify copy-number neutral SV. In this case, two parts of the same sequencing read can map discordantly to the reference genome, and orientation and positioning of the mapping allow for conclusions about the underlying rearrangements to be drawn. With paired-end sequencing a similar principle applies but with each read of a pair being located at the ends of the sequenced fragment [86]. This allows for a better mappability, due to

generally longer read length when compared to split reads as well as for higher physical coverage. This advantage is even further increased in so-called mate-pair or DNA-paired-end tag (PET)-sequencing, in which the sequenced pairs originate from genomic sites which can be spaced by several kb (see figure 12 on page 132). This increases physical coverage and, importantly, is able to cross large repeats which are present at many SV breakpoints but are notoriously hard to map to a unique genomic location. Yet another big advancement in SV analysis, and sequencing in general, is currently being introduced in the form of long-read technologies, such as single-molecule real-time sequencing by PacBio [87] and nanopore based systems by Oxford Nanopore [88]. With promised average read-lengths of over 15 kb and 5 kb, respectively, both systems are able to span long and complex SV can be spanned in one read.

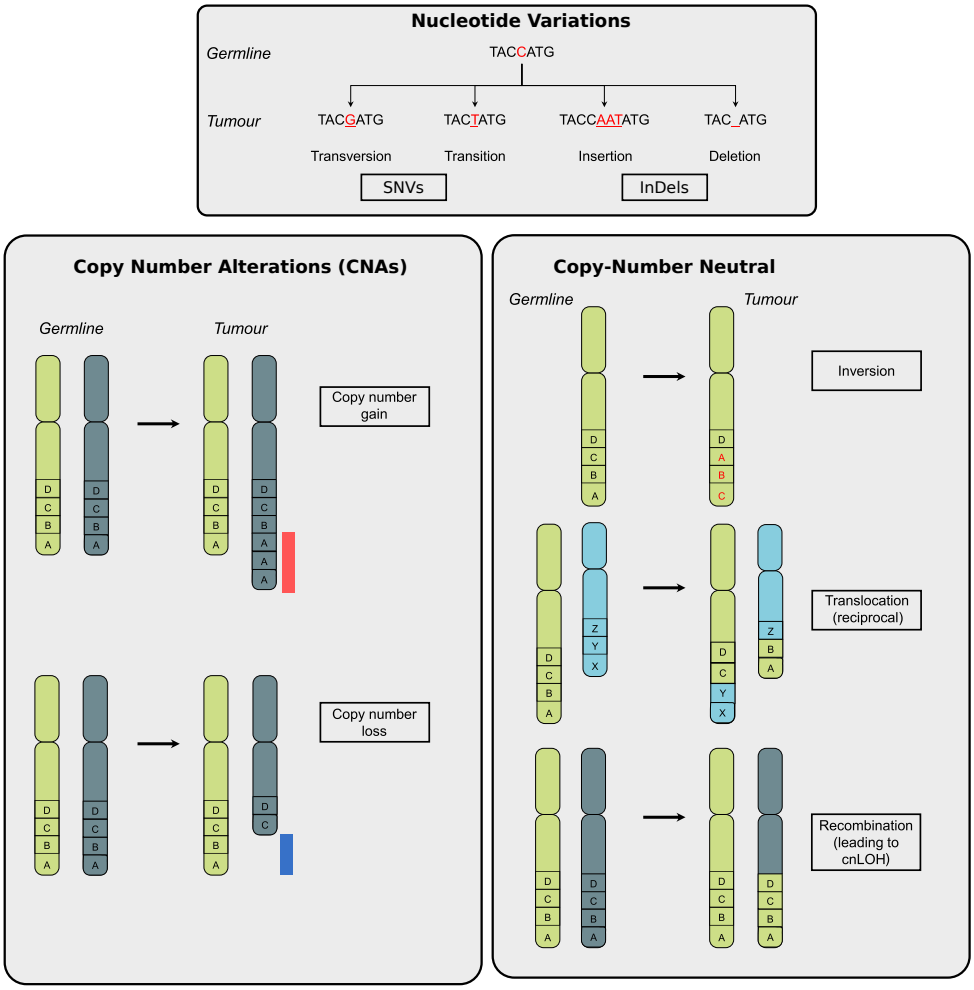


Figure 5 – Genomic aberrations in cancer. Three classes of genomic aberrations found in cancer genomes are depicted: small nucleotide variations, structural copy number alterations (bottom left box), and copy number neutral structural variations (bottom right box). SNV are of three basic types: substitutions (transversions and transitions); insertions; and deletions (InDels). Copy Number Gains: The two homologous chromosomes are shown with a gain (blue bar) of two additional copies of region A on the paternal chromosome. Copy number losses: Regions A and B on the paternal chromosome are shown with a red bar highlighting the deleted region. The three most common types of genomic structural variations (SV) are shown. Inversions on the same chromosome results in a change of the orientation of DNA sequences. The inverted regions in the tumor are highlighted with red letters. Translocations can be reciprocal or nonreciprocal and typically occur between nonhomologous chromosomes (depicted in green and blue). Recombination between sister chromatids can result in copy number neutral LOH. A, adenine; C, cytosine; cnLOH, copy number-neutral loss of heterozygosity; G, guanine; LOH, loss of heterozygosity; T, thymine. Adapted from Watkins *et al* [89].

1.4 Current Knowledge of the Genomic Landscape of Osteosarcoma

As highlighted in the previous section, knowledge about genomic aberrations can not only help in understanding the etiology and pathogenesis of a disease but can also be used to derive targeted therapies. Several studies have therefore begun to unravel the genome of OS in the past [90, 91]. However, many of these studies are based on cell lines or mouse models, or include only a limited number of clinical samples. Additionally, early studies could not benefit from the technological advancements in sequencing and array technologies and therefore were more constrained in their scope than the genomic cancer studies of today. Nevertheless, important observations about OS driver genes, chromosomal instability and other aspects were accrued.

1.4.1 Sporadic Osteosarcoma

In contrast to other sarcoma subtypes, no canonical genomic rearrangement or genetic mutation has been identified for OS. Nonetheless, causal roles of the tumor suppressor genes *TP53* and *RB1* have been demonstrated in some tumors already some decades ago and were later confirmed in several species [92, 93, 94]. Aberrations in these driver genes are thought to induce chromosomal instability (see figure 6 on page 18). RB1 blocks the $G1_1$/S phase transition in the cell cycle by binding to E2F family transcription factors, thereby preventing cell cycle progression. During mitosis, the E2F/RB1 interaction is reversed through phosphorylation of RB1 by CDK4. Loss of function mutations in *RB1* therefore remove this cell cycle checkpoint and can lead to increased proliferation. CDK4 in turn is inhibited by p16, a protein which is encoded by the *CDKN2A* gene. The archetypical tumor suppressor p53 is a transcription factor and a critical regulator of DNA damage response, cell cycle progression and apoptosis [95]. In the absence of stress signals, p53 protein is kept at low levels through continuous degradation by MDM2. The p53 pathway can be activated by ARF, which can sequester MDM2 and therefore prevent the degradation of p53. ARF, like the above-mentioned p16, is encoded by *CDKN2A*. Recent studies have shown that *TP53* mutations are present in OS with high frequency and can be caused by both SNV and SV [96, 97].

That being said, the clinical utility of *TP53* and *RB1* mutations to date are limited. Beyond OS, *TP53* is recognized as one of the most mutated genes in human malignancies [98] and germline mutations in *TP53* can lead to the Li-Fraumeni cancer predisposition syndrome. Therefore, drugs which are able to restore the p53 pathway are an active area of research. While compounds like APR-246 show promising preclinical results in restoring p53 transcriptional

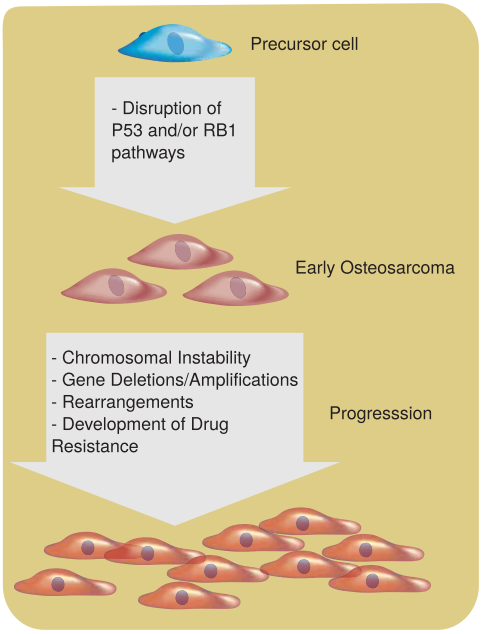


Figure 6 – Classical Model of osteosarcoma pathogenesis. In the classical view of tumor initiation and progression in OS, mutations in the P53 or RB1 tumor suppressors in an undetermined cell of origin leads to chromosomal instability. The resulting deletions, amplifications and rearrangements contribute to tumor progression and development of drug resistance. P53, Cellular tumor antigen P53; RB1, Retinoblastoma. Adapted from Chou *et al* [55].

activity of selected missense mutations, no drugs to counter other aberrations like deletions, truncating mutations or disrupting rearrangements are currently available [99]. Strategies for wildtype *TP53* reconstitution in tumors (gene therapy) via introduction of an intact complementary DNA (cDNA) copy of the gene for example using viral vectors are being tested, but have not yet lead to clinical success. Other genomic studies have identified additional OS driver candidates. *ATRX*, which is involved in telomere maintenance and chromatin remodelling, has been reported as mutated in 10– 50% of tumors; the oncogene *MDM2* which suppresses *TP53* is amplified in 4-5% [96, 97]; and *MYC*-amplifications were described in 15% [97], amongst others. Further, deletions of the phosphatase and tensin homolog (PTEN) tumor suppressor were found in 44% of cases [100]. For a comprehensive list of additional driver genes which have been reported with lower frequency, the interested reader may refer to a review by *Rickel et al.* [90].

The emerged consensus of the existing studies paints the picture of a highly complex and rearranged genome. OS is generally characterised by high levels of genomic instability [91]. Genomic instability can be divided into several types: Microsatellite instability, which is attributed to deficiency in DNA mismatch repair and which has mainly been described in

colorectal- and bladder cancer, results in the accumulation of SNV and InDels as a consequence of replication errors [101]. The second type is chromosomal instability and represents an increase in the rate of genomic rearrangements as well as deletions and gains [102]. Whereas OS generally appear to be microsatellite stable, chromosomal instability has been observed in most genomic studies and is thought to be primarily induced by p53 deficiency [90, 103]. Moreover, ploidies of haploid to hexaploid have been described in OS [104]. Two of the most recent genomic studies with comprehensive cohorts by Chen *et al.* and Perry *et al.* underline the importance of chromosomal instability in OS and report an average of 317 and a median of 230 somatic rearrangements per tumor, respectively. Chen and colleagues further described catastrophic genomic events in form of chromothripsis in several OS cases. In chromothripsis, a cell acquires tens to thousands of rearrangements in a one-off event, as a result of chromosome shattering and error-prone repair [105]. Interestingly, like "classical" chromosomal instability, chromothripsis has been linked to *TP53* mutations [106].

1.4.2 Familial Osteosarcoma

Most cases of OS are sporadic, yet there are several known heritable syndromes which predispose patients to develop OS. Germline mutations in the *TP53* tumor suppressor cause Li-Fraumeni syndrome (LFS), a rare autosomal dominant disorder with a wide spectrum of early-onset tumors, including OS, soft tissue sarcomas, breast cancer, and tumors of the central nervous system. The majority of the causal alterations are missense mutations which are mostly located in exons 3 to 11, but rarer rearrangements (mostly deletions) have been described [107]. Germline mutations in the cell cycle regulator *RB1* have also been shown to increase the incidence of OS significantly [108], as well as mutations in the RecQ protein-like 4 helicase, which causes Rothmund-Thomson syndrome and leads to development of OS in up to 32% of affected individuals [109]. *BLM* and *WRN* are two other members of the RecQ family in which germline mutations lead to syndromes associated with OS.

Results

2.1 Rationale and Aims of the Thesis

The work presented in this thesis aimed to leverage current genomic approaches to better characterize OS and to identify new actionable mutations of potential clinical utility. A better understanding of the genomic basis of the disease is crucial to improve the prognosis of patients and may lead to more individualized treatment approaches. The number of patients with refractory or recurrent disease remains high and underlines the clear need for innovative and novel therapies. By analyzing a large set of OS samples using state of the art techniques for exome-, whole genome-, and targeted panel-sequencing, as well as copy number arrays, we therefore aimed to contribute to the expanding knowledge of OS genomics. Even though few studies on OS genomics have been reported, many are based on cell lines or mouse models or include only a limited number of tumor samples. Since OS is a rare disease and commonly used biopsy methods often result in insufficient sample material for research, the extensive pretherapeutic sample collection of the Basel Bone Tumor Reference Center represents a highly valuable resource for this undertaking.

2.1.1 Publications and Manuscripts Arising from this Work

In **Publication No 1** ("*TP53* intron 1 hotspot rearrangements are specific to sporadic osteosarcoma and can cause Li-Fraumeni syndrome", page 23) we started to disentangle the chaotic genomic landscape of four pretherapeutic OS. Using whole-genome DNA-PET sequencing, we found rearrangements with breakpoints in intron 1 of the tumor suppressor gene *TP53* as a recurrent aberration.Using FISH analyses, we demonstrated such rearrangements to be specific to a subset of OS and not to occur in an extended cohort of 1'090 bone-producing and bone-unrelated tumors. We further showed that intron 1 rearrangements can cause Li-Fraumeni syndrome in a four-generation family with previously unidentified *TP53* mutations.

In **Publication No 2** ("Exome sequencing of osteosarcoma reveals mutation signatures reminiscent of BRCA deficiency", page 63) we analyzed SNVs and CNAs of a large cohort consisting of 123 pretherapeutic OS samples using next-generation sequencing and SNP-based copy number arrays. By combining the data we found that 47% of tumors could be explained by clonal *TP53* and *RB1* aberrations and an additional 40% by clonal aberrations in 12 other well-described cancer-driver or -susceptibility genes. We further showed that the majority of analyzed cases revealed genomic signatures similar to BRCA-deficient tumors (so-called BRCAness), indicating homologous recombination repair deficiency which might be therapeutically exploitable. In fact, PARP-inhibitors targeting this impairment in double strand repair are currently tested in several other malignant tumors. At least in OS cell lines also showing BRCAness, we could demonstrate PARP-inhibitors to significantly reduce cell viability.

In **Manuscript No 3** ("RET Germline Mutations and Susceptibility to Osteosarcoma", page 75) we picked up on our observations of recurrent *TP53* rearrangements in Li-Fraumeni syndrome and further investigated the role of germline alterations in the context of inherited OS, beyond the known susceptibility genes *TP53, RB1, BLM, WRN*, and *RECQL4*. Based on the findings in an index case with a pathogenic germline variant in the *RET* tyrosine kinase and a family history of medullary thyroid cancer, we analyzed whole-genome, exome, and targeted sequencing data of 336 OS patients. We discovered seven patients with *bona fide* pathogenic *RET* germline variants and four patients with germline variants of unknown clinical significance. We compared mutation burden, somatic mutation signatures and copy number profiles and showed that the genomic landscapes of these tumors resemble previously described OS. When comparing these findings to the frequency in the population, we could show that patients carrying disease-associated germline *RET* mutations are at a significantly increased risk of developing OS. Taken together, these data suggests the inclusion of OS to the spectrum of *RET*-associated disorders and encourages the consideration of tyrosine kinase inhibitors as a treatment option in appropriate cases.

2.2 Publication No 1

***TP53* intron 1 hotspot rearrangements are specific to sporadic osteosarcoma and can cause Li-Fraumeni syndrome.**

Ribi S*, Baumhoer D*, Lee K*, Edison, Teo AS, Madan B, Zhang K, Kohlmann WK, Yao F, Lee WH, Hoi Q, Cai S, Woo XY, Tan P, Jundt G, Smida J, Nathrath M, Sung WK, Schiffman JD, Virshup DM, Hillmer AM.

***These authors have contributed equally to this work**

Oncotarget. 2015 Apr 10;6(10):7727-40.

www.impactjournals.com/oncotarget/ Oncotarget, Vol. 6, No. 10

TP53 intron 1 hotspot rearrangements are specific to sporadic osteosarcoma and can cause Li-Fraumeni syndrome

Sebastian Ribi[1,*], Daniel Baumhoer[2,3,*], Kristy Lee[4,*], Edison[5], Audrey S.M. Teo[1], Babita Madan[5], Kang Zhang[6], Wendy K. Kohlmann[7], Fei Yao[1], Wah Heng Lee[8], Qiangze Hoi[8], Shaojiang Cai[8], Xing Yi Woo[9], Patrick Tan[1,5,10], Gernot Jundt[2], Jan Smida[3,11], Michaela Nathrath[3,11], Wing-Kin Sung[8,12], Joshua D. Schiffman[4], David M. Virshup[5], Axel M. Hillmer[1]

[1]Cancer Therapeutics & Stratified Oncology, Genome Institute of Singapore, Singapore 138672, Singapore

[2]Bone Tumor Reference Center at the Institute of Pathology, University Hospital Basel, CH-4003 Basel, Switzerland

[3]Clinical Cooperation Group Osteosarcoma, Helmholtz Zentrum Muenchen, German Research Center for Environmental Health, 85764 Neuherberg, Germany

[4]Department of Pediatrics and Oncological Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT 84112, USA

[5]Duke-NUS Graduate Medical School Singapore, Singapore 169857, Singapore

[6]Institute for Genomic Medicine, UC San Diego, La Jolla, CA 92830, USA

[7]Huntsman Cancer Institute, University of Utah Health Care, Utah, UT 84112, USA

[8]Computational & Systems Biology, Genome Institute of Singapore, Singapore 138672, Singapore

[9]Personal Genomics Solutions, Genome Institute of Singapore, Singapore 138672, Singapore

[10]Cancer Science Institute of Singapore, National University of Singapore, Singapore 117599, Singapore

[11]Department of Pediatrics and Wilhelm Sander Sarcoma Treatment Unit, Technische Universität München and Pediatric Oncology Center, 81675 Munich, Germany

[12]School of Computing, National University of Singapore, Singapore 117417, Singapore

*These authors have contributed equally to this work

Correspondence to:
David M. Virshup, **e-mail:** david.virshup@duke-nus.edu.sg
Axel M. Hillmer, **e-mail:** hillmer@gis.a-star.edu.sg

Keywords: *TP53*, Li-Fraumeni syndrome, osteosarcoma, cancer genomics, structural variations

Received: November 10, 2014 ***Accepted:*** January 08, 2015 ***Published:*** February 25, 2015

ABSTRACT

Somatic mutations of *TP53* are among the most common in cancer and germline mutations of *TP53* (usually missense) can cause Li-Fraumeni syndrome (LFS). Recently, recurrent genomic rearrangements in intron 1 of *TP53* have been described in osteosarcoma (OS), a highly malignant neoplasm of bone belonging to the spectrum of LFS tumors. Using whole-genome sequencing of OS, we found features of *TP53* intron 1 rearrangements suggesting a unique mechanism correlated with transcription. Screening of 288 OS and 1,090 tumors of other types revealed evidence for *TP53* rearrangements in 46 (16%) OS, while none were detected in other tumor types, indicating this rearrangement to be highly specific to OS. We revisited a four-generation LFS family where no *TP53* mutation had been identified and found a 445 kb inversion spanning from the *TP53* intron 1 towards the centromere. The inversion segregated with tumors in the LFS family. Cancers in this family had loss of heterozygosity, retaining the rearranged allele and resulting in *TP53* expression loss. In conclusion, intron 1 rearrangements cause p53-driven malignancies by both germline and somatic mechanisms and provide an important mechanism of *TP53* inactivation in LFS, which might in part explain the diagnostic gap of formerly classified "*TP53* wild-type" LFS.

INTRODUCTION

Germline mutations in the *TP53* tumor suppressor gene cause Li-Fraumeni syndrome (LFS), an autosomal dominantly inherited predisposition syndrome to various cancers, including osteosarcoma (OS) [1, 2]. *TP53* coding mutations can be identified in 70% of classic LFS families [3] leaving a significant proportion of LFS cases with an unknown genetic basis. The vast majority of *TP53* mutations in LFS, OS and other tumors are point mutations dominated by missense mutations [4]. Larger germline deletions encompassing the entire *TP53* gene together with neighboring genes have been described to correlate with developmental delay [5]. Partial deletions of *TP53* have been found to be associated with LFS suggesting that the partial loss of *TP53* has a different functional outcome than the entire deletion of the gene [5]. Some genomic structural variations (SVs) have been described that can affect *TP53* function. These SVs are either deletions, which were identified by PCR based methods or comparative genome hybridization, that affect the *TP53* gene in LFS patients [5, 6], or rearrangements in intron 1 of *TP53* which initially have been identified by Southern blot in OS [7–9]. Recently, whole-genome sequencing of tumors from 32 OS patients showed cancer-specific *TP53* rearrangements in > 50% of patients [10].

p53 is a DNA-damage response protein [11] and its inactivation could be expected to result in further genomic instability [12]. Mutations of *TP53* are among the most common defects associated with human cancer in general. Given the large number of *TP53* point mutations which have been identified in the majority of cancer types, it is surprising that *TP53* intron 1 rearrangements have only been found in OS [7–10]. Since exome sequencing does not allow the identification of copy number neutral genome rearrangements with intergenic or intronic breakpoints, it is possible that *TP53* intron 1 rearrangements have been missed in many studies. In addition, the suggested specificity of *TP53* intron 1 rearrangements for OS is based on screens of a limited number of samples [7–9]. Further, it seems possible that *TP53* intron 1 rearrangements do not only contribute to sporadic OS but also to LFS, which is driven by germline *TP53* mutations. In the present study, we analyze the nature of *TP53* intron 1 rearrangements, screen the to date largest collection of OS and other tumor types for such rearrangements, describe the identification of a *TP53* intron 1 disrupting germline inversion in a four generation LFS family and characterize this locus and *TP53* activity in tumors of this family.

RESULTS

Characterization of recurrent rearrangement points in intron 1 of *TP53*

We analyzed the genome structures of four pre-therapeutic OS using DNA paired-end tag sequencing (DNA-PET), a genome-wide mate-pair sequencing approach [13–15] and predicted 434, 289, 348 and 420 SVs, respectively, to be somatically acquired (Supplementary Tables S1–S6, Figures S1A and S1B, S2 and S3A, S3B and S3C). We identified seven breakpoints within a small region of intron 1 of *TP53* in three OS tumors (Figure 1, Figure S4 and Supplementary Table S7) and the fourth (AJF) had a 94 kb deletion that included the entire *TP53* gene as well as neighboring genes (Figure 1A and 1B). Tumor YZH showed a balanced translocation between *TP53* intron 1 and chromosome 1. The sequence of the breakpoints showed the presence of the same 555 bp and 293 bp of the *TP53* and chromosome 1 loci, respectively, on both sides of the translocations (Supplementary Figure S5A and S5B). Tumor PZP had a 12.5 kb inverted insertion originating from chromosome 6 containing *ENPP1* exons 19 to 25 including the stop codon (Supplementary Figure S6A and S6B). In addition, the *TP53* intronic sequences on both sides of the insertion overlapped by 59 bp suggesting that a similar mechanism was responsible for the translocations in both YZH and PZP. Tumor KRD had complex inter-chromosomal translocations with the three different partner chromosomes 1, 5 and 6 (Figure 1B) implying that these are three independent events. At least one event had to be non-clonal meaning that two or three independent clones with structural rearrangements in *TP53* intron 1 underlie this tumor. The translocation breakpoints in intron 1 of *TP53* with chromosomes 1 and 6 were only 45 bp apart with an overlap of 46 bp of the intron 1 sequence. The overlap and orientations were compatible with one event of similar mechanism as for tumors YZH and PZP. In contrast, the DNA-PET mapping regions of the chromosome 5 translocation suggest that this rearrangement occurred on the other allele of *TP53* or in an independent clone (Figure 1B).

Centromeric of the breakpoint cluster region (2.5 kb of its center towards exon 1 of *TP53*) data of the Encyclopedia of DNA Elements (ENCODE) [16] show strong signals of open chromatin and active enhancers. It seems possible that the open chromatin state and/or the active transcription of *TP53* contribute to the rearrangement mechanism. Six of the seven breakpoints were located within long interspersed elements (LINE), and the seventh breakpoint was within a short interspersed element (SINE). While the breakpoint partner sites do not show enrichment for LINE or SINE sequences, it is striking that five of the seven partner breakpoints also have strong signals of open chromatin within a region of 10 kb (Supplementary Figures S7A and S7B, and S8 to S12).

In two of the three tumors with *TP53* intron 1 rearrangements, the breakpoint locations predicted gene fusions forming *TP53-ENPP1-TP53* and *SUCO-TP53* (Supplementary Table S7). Interestingly, both fusion gene partners are involved in bone biology. *ENPP1* has been shown to be a key regulator of ossification [17].

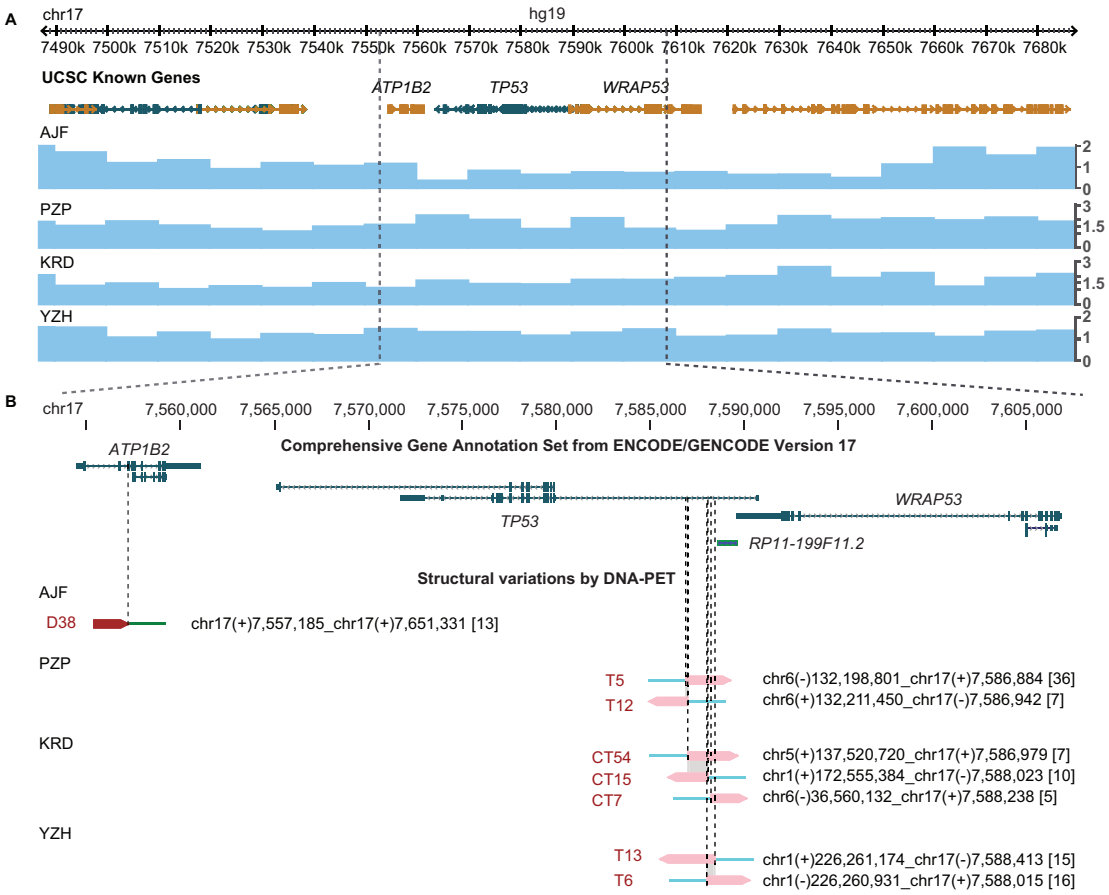


Figure 1: Translocation hotspot in intron 1 of *TP53* in OS samples. (A) Genes derived from the UCSC known genes database [43] (top) and copy number information derived from DNA-PET sequencing data of four OS samples (blue tracks, bottom) are shown in the Genome Browser. Genes transcribed from the plus strand are represented in gold, genes transcribed from the minus strand are represented in green. Boxes indicate exons, barbed lines indicate introns. The *TP53* locus for patients PZP, KRD, and YZH has a copy number of two while patient AJF shows loss of one copy. **(B)** Enlargement of gene (top) and breakpoint (bottom) view of the *TP53* locus. GENCODE transcripts with unresolved problems have been excluded. Note that *TP53* is transcribed on the minus strand (from right to left). Mapping regions of DNA-PET sequence tags which represent a rearrangement are shown as dark red (5'-tags) and pink (3'-tags) arrow heads with the predicted breakpoint being located at the tip of the dark red and the base of the pink arrow heads (dashed lines). SV identifiers are in red letters, predicted breakpoint locations and connections are indicated for each rearrangement in black letters. Numbers in squared brackets indicate number of PETs which connect the two genomic regions of a SV (dPET cluster size). Shaded in gray are stretches of identical sequences for both breakpoint sides.

SUCO (SUN domain containing ossification factor) is an essential regulator of postnatal osteoblast maturation [18]. Furthermore, we found both genes expressed in a collection of in-house established OS cell lines (unpublished data). The expression of the rearranged genes in bone supports the idea that active expression might mechanistically contribute to the translocations.

Somatic *TP53* rearrangements are a frequent phenomenon specific for OS

We designed a break-apart FISH test using probes flanking the *TP53* gene (Figure 2A) and investigated a series of 215 pre-therapeutic OS samples arranged on a tissue microarray (TMA). We found 11% (23 out of 215) of the cases to have rearrangements at the *TP53* locus (FISH break-apart positive; Figure 2B). Of note, in all 23 FISH positive cases, both alleles showed the break-apart signal. However, FISH positive patients did not differ from negative patients in terms of overall-survival ($p = 0.6$), event-free survival ($p = 0.7$), occurrence of metastases, or response to neoadjuvant chemotherapy (Table 1 and Supplementary Tables S8 and S9). Rearrangements at this locus nevertheless appear to be a recurrent finding in OS. To test whether the *TP53* rearrangement also occurred in other bone-forming tumors that sometimes can be difficult

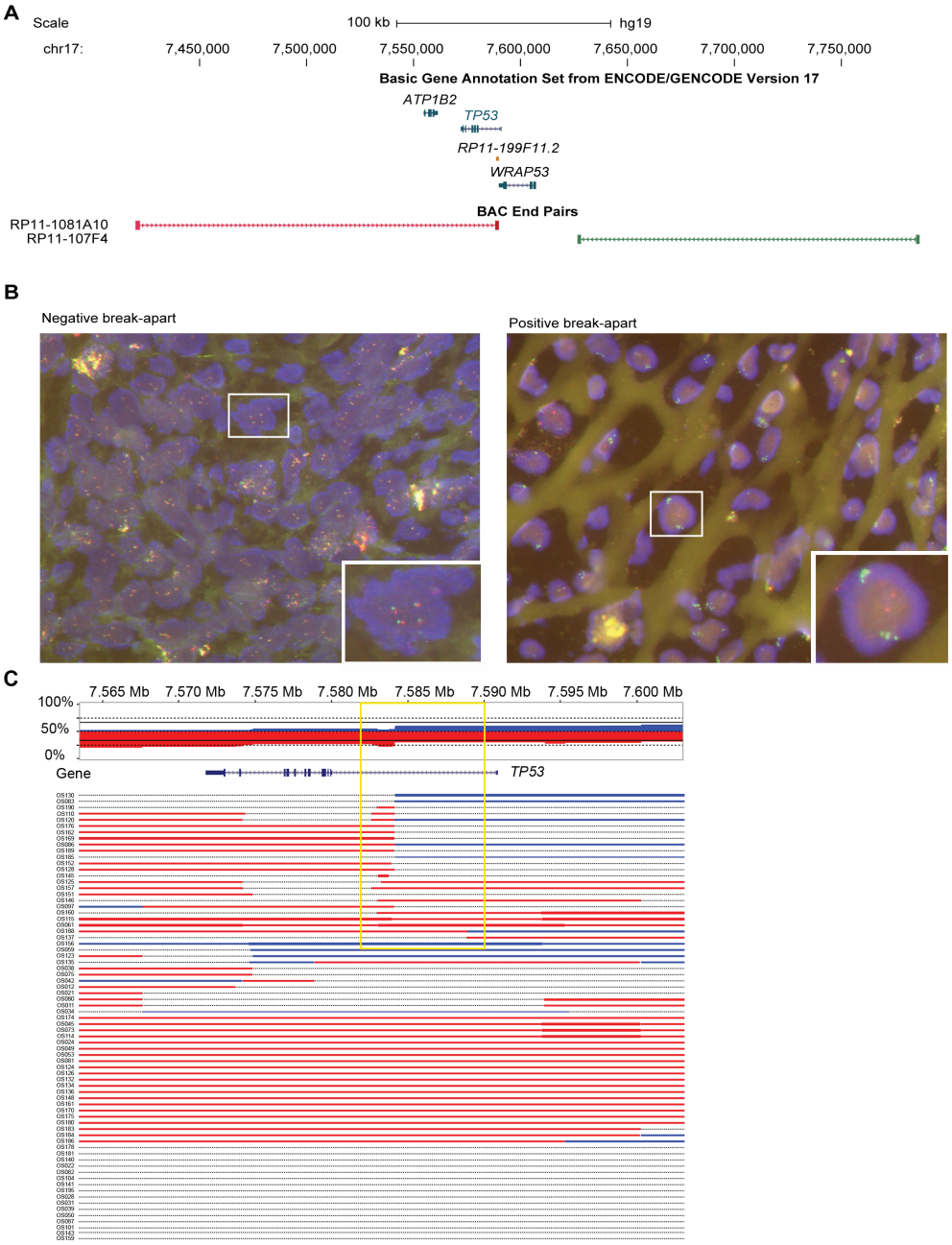


Figure 2: Translocation hotspot in intron 1 of *TP53* in OS samples. (A) Location of BAC clones which have been selected for FISH relative to *TP53* and immediate neighbouring genes. Other genes of the track have been deleted for clarity. Color code matches fluorophore of FISH analysis shown in B. **(B)** Examples of a negative and positive break-apart signal of two color FISH which has been used to screen 267 formalin fixed and paraffin embedded (FFPE) OS samples and 141 other bone-forming tumors. **(C)** Copy number overview of 73 OS tumors at the *TP53* locus based on CytoScan array analysis. Top panel shows the cumulative copy number across all samples with red indicating loss and blue gain in copy number. Lower panel shows the copy number gains and losses for each of the 73 OS tumors individually. Note the changes in copy number within intron 1 of *TP53* detectable in 23 cases (yellow box).

Table 1: Characteristics of OS patients

	OS TMA		OS fresh frozen	
	All	FISH positive (out of 215 evaluable)	All	Intron 1 CN change
Gender	254/267 (95.1%)	23/23 (100%)	73/73 (100%)	24/24 (100%)
male	129	9	39	13
female	125	14	34	11
Age at diagnosis	258/267 (96.6%)	23/23 (100%)	73/73 (100%)	24/24 (100%)
average	24.3 years	17 years	17.3 years	18.6 years
median	17 years	15 years	15 years	15.5 years
range	4–88 years	6–48 years	3–56 years	5–56 years
Observation period	259/267 (97%)	23/23 (100%)	72/73 (98.6%)	24/24 (100%)
average	61.6 months	46.6 months	73.9 months	64.6 months
median	35 months	24 months	66.5 months	64 months
range	0–287 months	0–179 months	0–205 months	2–196 months
Response to neoadjuvant treatment	180/267 (67.4%)	19/23 (83%)	66/73 (90.4%)	21/24 (87.5%)
good (< 10% viable tumor)	102	14	35	10
poor (≥ 10% viable tumor)	78	5	31	11
Metastases	267/267 (100%)	23/23 (100%)	65/73 (89%)	19/24 (79.2%)
yes	101	9	39	10
no	166	14	26	9
Survival	259/267 (97%)	23/23 (100%)	72/73 (98.6%)	24/24 (100%)
alive	174	18	55	16
deceased	85	5	17	8
TP53 immunhisto-chemistry	212/267 (79.4%)	19/23 (83%)		
negative	170	15		
positive	42	4		
Location		23/23 (100%)		24/24 (100%)
femur		14		12
tibia		5		8
jaws		2		-
humerus		1		1
fibula		1		1
other		-		2

OS TMA series / OS Fresh Frozen series: number of evaluable cases / total number of cases (percentage) CN, copy number

to distinguish histologically from OS in small biopsies, we analyzed another series of 124 bone-forming tumors and tumor-like lesions using our FISH assay. None of these cases showed evidence of *TP53* rearrangement. To further exclude *TP53* intron 1 rearrangements in other tumor types we used our FISH assay to analyze an additional 966 tumors on a TMA (Supplementary Tables S10 and S11). None of the 966 tumors showed a break-apart signal suggesting the somatic *TP53* intron 1 rearrangements represent a specific finding in OS.

To further validate our findings by another platform, we analyzed an independent set of 73 pre-therapeutic fresh-frozen OS samples for copy number alterations (CNAs) using CytoScan® high density arrays. We found that 74% (54 out of 73) of the OS samples had alterations affecting the *TP53* gene. Amongst these alterations, 23 showed transition points into copy number losses ($n = 17$) or gains ($n = 3$) or abrupt transitions from losses into gains ($n = 3$) in intron 1 of the gene (32% of total samples, 23 out of 73, Figure 2C). Again, the rearrangements did not correlate with any clinico-pathological parameters.

TP53 intron 1 rearrangement in a family with LFS

Since mutations of *TP53* are associated with LFS it seemed possible that *TP53* intron 1 rearrangements could constitute a previously underappreciated category of alterations that can cause LFS. We revisited an LFS family with 12 affected members with cancer across four generations in which we previously had been unable to identify a coding *TP53* mutation or a co-segregating, potentially damaging and disease-causing alteration based on exome sequencing of patients P1, P2 and P13 (Figure 3A and data not shown). Copy number analysis of DNA from blood of patients P1 and P13 by single nucleotide polymorphism (SNP) array revealed evidence for an approximately 2.5 kb deletion in intron 1 of *TP53* including exon 1 (Supplementary Figure S13 and Table S12). However, we were not able to amplify the fusion point of a deletion by PCR. We then used a custom sequence capture assay for targeted paired-end sequencing of the *TP53* locus to search for a rearrangement point. We identified a 445 kb inversion spanning from the breakpoint cluster region in intron 1 of *TP53* towards the centromere (upstream of *TP53*) with a loss of 2,275 bp of intron 1/exon 1 of *TP53* (Figure 4A and 4B). The lost sequence was in agreement with the deletion identified by SNP arrays. In the adult family members tested, we found this specific rearrangement co-segregating with the disease, implicating the rearrangement as the causative alteration (Figure 3B and 3C).

Twelve different protein isoforms of *TP53* (p53, p53β, −γ, Δ40p53α, −β, −γ, Δ133p53α, −β, −γ, Δ160p53α, −β, −γ) can be generated by alternative splicing, alternative promoter usage, and alternative initiation sites of translation which have different functional properties (reviewed in [19, 20]). A weak promoter is located just upstream of exon 1, a strong promoter 1 kb downstream of exon 1 and a third promoter at exon 5 [19, 21]. Transcripts for p53 (full length) and the Δ40p53 isoforms contain the non-coding exon 1. The intron 1 rearrangements disconnect the exon 1 of *TP53* and the two first promoters from the remaining gene body. To investigate the impact of *TP53* rearrangements on *TP53* expression, we obtained RNA from blood of LFS patients H2 and P13, OS lung metastasis of H2 and a cell line derived from the lung adenocarcinoma of P13.

We performed quantitative reverse transcription polymerase chain reactions (qRT-PCRs) targeting transcripts encoding for the twelve described p53 isoforms and for isoforms Δ133α, −β, −γ and Δ160α, −β, −γ (*TP53* delta), respectively, and found a reduction of transcripts by 23–53% for the blood of H2 and P13 where the rearrangement was in a heterozygous state, and a reduction by 89–100% for the OS lung metastasis of H2 and a cell line derived from the lung adenocarcinoma of P13 (Figure 5). This implies that the rearrangements result in a loss of *TP53* transcription and function rather than a switch to reported transcripts lacking exon 1.

To test for deletions of the second allele in tumors, copy number analysis using OncoScan FFPE Express (Affymetrix, Inc.) were performed on tumor samples of H2 (two OS lung metastases that developed six months apart), P1 (undifferentiated pleomorphic sarcoma) and P13 (lung adenocarcinoma and meningioma). Loss of heterozygosity (LOH) at the *TP53* locus occurred in all five investigated tumors (Figure 6A and Supplementary Figures S14 and S15). Semi quantitative genomic PCR for LFS-BP2 showed a stronger signal for the LFS breakpoint than for the non-rearranged allele suggesting a duplication of the rearranged allele and loss of the non-rearranged allele with the weak PCR signal derived from stroma contamination (Figure 6B). Our findings suggest LOH as a frequent mechanism for *TP53* inactivation after an initial intron 1 rearrangement.

DISCUSSION

Rearrangement hotspot in *TP53* intron 1

More than 20 years ago, rearrangements in *TP53* in OS were identified by Southern blot [7–9]. Although based on a small number of samples, these rearrangements were thought to be specific for OS. Recently, *TP53* rearrangements have been rediscovered in OS in more than 50% of 32 OS by whole-genome sequencing at a higher resolution [10]. In our study, we also found *TP53* intron 1 rearrangements to be the most recurrent focal rearrangement point in four OS samples. We found evidence for *TP53* rearrangements in 11% (23 out of 215) of OS samples by FISH and copy number changes in intron 1 of *TP53* in 32% (23 out of 73) of OS samples by CytoScan arrays. Importantly, there were

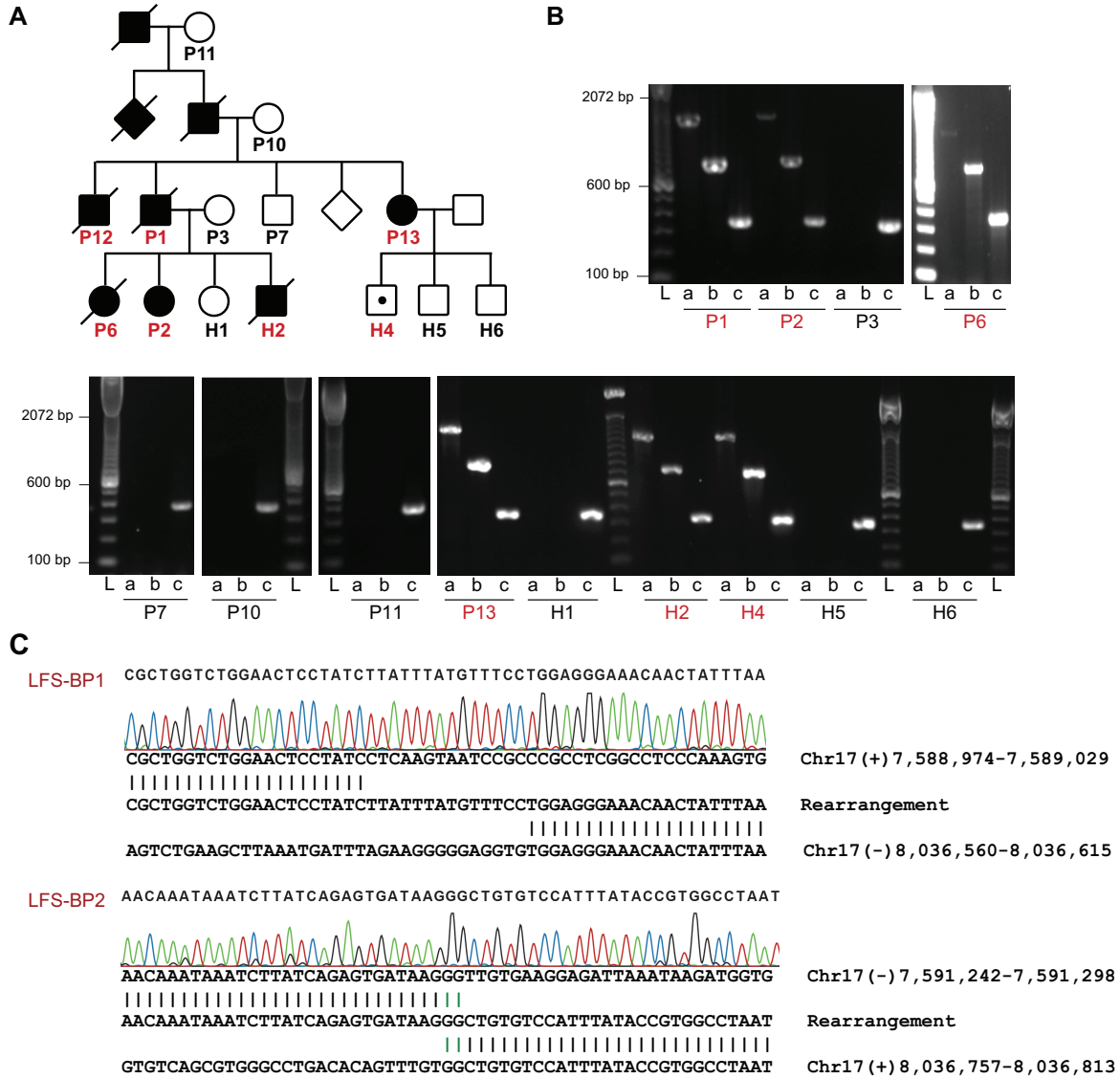


Figure 3: *TP53* intron 1 rearrangement in a family with LFS. (A) Pedigree of a family with LFS. Squares and circles represent males and females, respectively. Filled symbols indicate individuals with early onset cancer. Symbols with a diagonal line indicate that the individual was deceased at the time of recruitment. Diamonds represent several additional family members whose gender should not be disclosed. Individuals of whom DNA samples have been obtained are numbered with individuals tested positive for *TP53* exon 1 deletion by MLPA in red. *TP53* rearrangement carrier without cancer at the age of 10 years is indicated by a white symbol with a black dot. **(B)** PCR analysis of positive control and *TP53* intron 1 rearrangement points (LFS-BP1 and LFS-BP2) of family members of whom high quality DNA was available. DNA quality of P12 did not allow PCR amplification, including positive control (not shown). a, LFS breakpoint 1 (BP1); b, LFS breakpoint 2 (BP2); c, positive control amplicon at the RNA polymerase II locus (POLR2A); sample IDs correspond to A with affected individuals in red. L, ladder. **(C)** Sanger sequencing of rearrangement points LFS-BP1 and LFS-BP2 at the *TP53* locus of P2. Micro homologies between the two participating break point regions are illustrated by green vertical lines. Genomic coordinates are based on NCBI build 37.

no rearrangements found in other bone-forming tumors (0 out of 124) by FISH analysis. Due to the low resolution of the bacterial artificial chromosome (BAC) probes of the FISH experiments which were 169 kb and 159 kb in size, respectively, the 23 detected rearrangements might include translocations outside intron 1 of *TP53*. However, we did not observe other rearrangements which could result in the 'break-apart' FISH signal in the whole-genome DNA-PET data of the four sequenced OS samples. Similarly, Chen *et al.* also found the majority of breaks in intron 1 [10], suggesting that most of the rearrangements detected by FISH were due to intron 1 rearrangements. In addition,

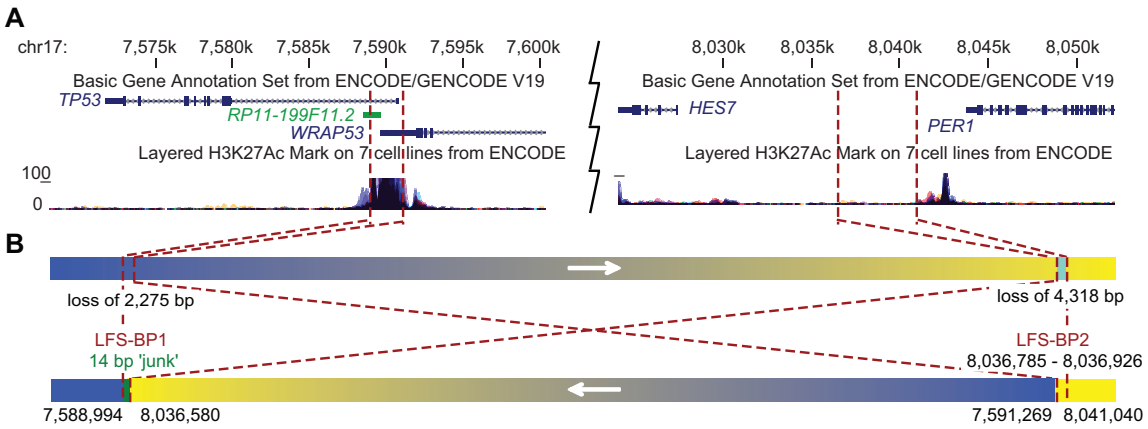


Figure 4: Inversion of 445 kb affecting intron 1 of *TP53* in a family with LFS. (A) University of California, Santa Cruz (UCSC) genome browser view [43] of the *TP53* locus with representative *TP53* and *WRAP53* transcripts, and H3K27Ac binding information of the ENCODE project [44]. Breakpoint locations are indicated by red dashed lines. **(B)** Schematic view of a 445 kb inversion affecting *TP53* in LFS family. The telomeric rearrangement points of the inversion are located in the breakpoint cluster region in intron 1 of *TP53*. Arrows indicate genomic orientations pointing towards higher coordinates. Blue, *TP53* locus; yellow, *HES7*/*PER1* locus; chromosome 17 coordinates of breakpoints are indicated below.

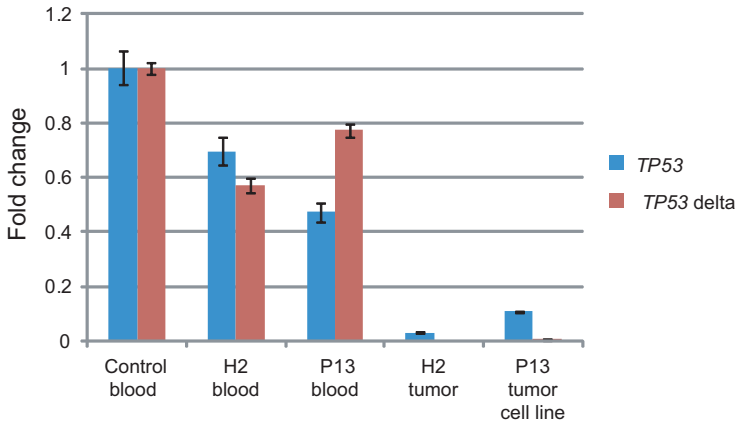


Figure 5: Tumors of LFS family show impaired *TP53* transcription. RNA was extracted from blood from one unrelated control individual, *TP53* intron 1 rearrangement carriers of the LFS family H2 and P13, OS lung metastasis of H2 and a lung adenocarcinoma derived cell line of P13. qRT-PCRs targeting all twelve *TP53* isoforms (*TP53*) and short transcripts encoding for the N-terminus lacking isoforms Δ133p53 and Δ160p53 (*TP53* delta) were performed in triplicates. Quantification cycle values (Cq) were normalized to *GAPDH* expression and are shown as fold change relative to control (y-axis). Error bars represent standard deviations. One representative of two experiments.

the FISH analysis did not have the resolution to identify small insertions as discovered in tumor PZP or deletions of similar sizes, thus the actual frequency of *TP53* intron 1 rearrangements might be underestimated. Further, we found *TP53* intron 1 rearrangements in OS by CytoScan analysis which has the resolution to locate the breakpoints to the first intron of *TP53*. The different frequencies of identified *TP53* rearrangements by FISH and CytoScan arrays might be explained by the lack of sensitivity of the FISH assay for the identification of small rearrangements. The differences of *TP53* intron 1 rearrangement frequencies between our study and the report of Chen and colleagues may be due to different analysis protocols or sampling biases.

TP53 intron 1 rearrangement mechanism

Three main types of mechanisms for genome rearrangements have been established: homologous recombination, replication-based mechanisms, and non-replicative non-homologous repair [22, 23]. The *TP53* intron 1 locus does not show significant sequence similarity with the seven translocation partner sites,

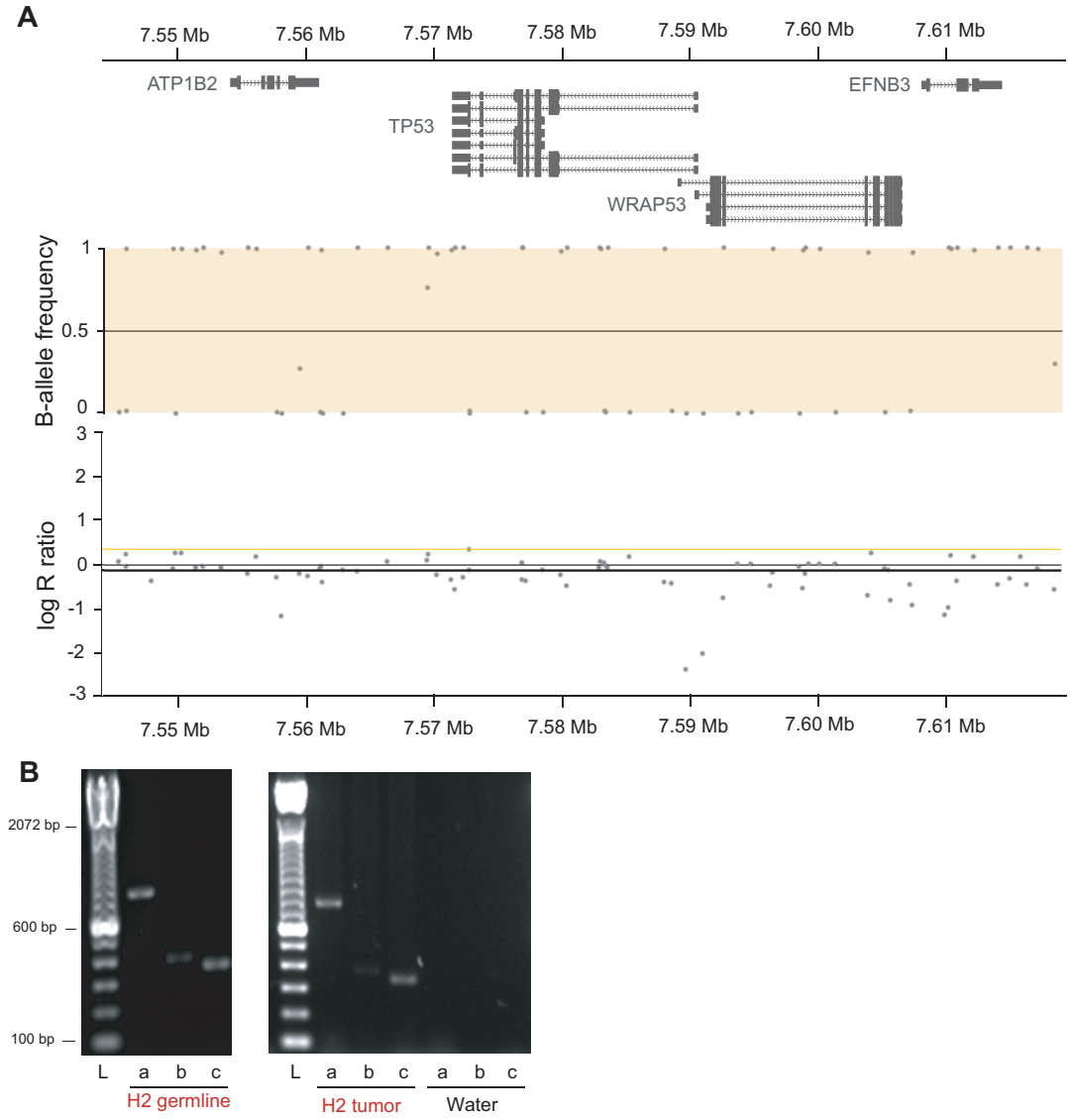


Figure 6: Tumors of LFS family show LOH at the *TP53* locus. (A) OncoScan array derived copy number and allele frequency plots of the *TP53* locus of OS lung metastasis of H2 show copy number neutral LOH. Chromosome 17 at 7.544 Mb to 7.618 Mb is shown with gene track (top), copy number (middle) and allele frequency (bottom). Y-axis of copy number is copy number ratio compared to reference (normal); y-axis of allele frequency shows homozygous allele calls for all single nucleotide polymorphisms in the window with values close to 0 and 1, respectively. **(B)** Semi quantitative PCR of LFS-BP2 (a), the wild type allele at the BP2 locus (b) and RNA polymerase II (positive control, c) of H2 germline DNA (30 PCR cycles) and H2 OS lung metastasis (28 PCR cycles) is shown. L, ladder.

arguing against a homology driven rearrangement mechanism. A replication coupled mechanism seems to be unlikely as well since a) it would require two replication fork invasions occurring in parallel at the same locus, b) it would involve two non-homologous chromosomes and replication-based mechanisms such as fork stalling and template switching (FoSTeS), which to the best of our knowledge have been described only for intra-chromosomal rearrangements, and c) we did not find evidence for rearrangement points which 'switch back' to the original *TP53* chromosome 17. The rearrangement points of non-replicative non-homologous repair, with the subgroups non-homologous end joining (NHEJ) and alternative end joining, are characterized by the absence of large homology but the presence of micro-homology of a few base pairs, blunt end joining or the insertion of small stretches of 'junk' DNA of unknown origin.

The *TP53* intron 1 region does not seem to be compatible with these common mechanisms of genome rearrangements or might be classified as a subcategory of NHEJ with sequence duplication. What we found in the three OS tumors with sequenced *TP53* intron 1 rearrangement points is different since both sides of the balanced intron 1 rearrangements contain a stretch of identical sequence of 46 bp to 555 bp. We therefore suggest a mechanism where a double strand break with long single stranded DNA overhangs occurs (Supplementary Figure S16A to S16G). The single stranded DNA overhangs might get filled in by the DNA repair machinery allowing blunt end fusions with other double strand breaks. Remarkably, all three OS tumors in which we initially found the *TP53* rearrangements share this unique feature of breakpoint sequences suggesting that they were caused by the same mechanism. The break point regions of the *TP53* locus and the translocation partner sites show a pattern of general active chromatin marks and gene expression in OS and/or bone tissue. The open state of the chromatin and/or the active expression of the respective genes may lead or contribute to the formation of the *TP53* intron 1 specific rearrangements.

Somatic SVs in *TP53* intron 1 are specific to OS

We assayed a comprehensive tumor collection of more than ten tissue types with a total of 1,090 non-OS tumors and 215 OS and show that the mechanistic event of the somatic rearrangements in *TP53* is highly specific to OS. In contrast, our findings in the LFS family indicate that *TP53* intron 1 can also occur in the germline and once such a SV is present as a germline alteration it can give rise to not only OS but multiple types of cancer, including adenocarcinoma, meningioma, astrocytoma, colon cancer, basal cell and squamous cell carcinoma of the skin. Interestingly, none of these tumor types were positive in our FISH TMA assay (n = 54, n = 14, n = 11, n = 62, n = 23 and n = 10, respectively) indicating that the mechanistic occurrence of *TP53* intron 1 rearrangements in the soma is specific for the osteoblast lineage, the believed cell of origin for OS [24]; however, the pro-cancer effect of such rearrangements supports tumor growth in many other tissues. Similarly, $TP53^{-/-}$ and $TP53^{+/-}$ knockout mice develop not only OS but also lymphoma, carcinoma and testicular cancer [25]. Again, our FISH analysis of tumors of these types did not show evidence for somatic *TP53* intron 1 rearrangements (n = 33, n = 566, n = 33, respectively). It therefore seems plausible that the OS specificity of the *TP53* intron 1 SVs is based on the mechanism of the rearrangement that leads it to occur only in osteoblast lineage with observable frequency, rather than OS being the specific consequence of the alteration. Cell lineage specific DNA replication properties or transcriptional processes might be responsible for the remarkable specificity of the described rearrangements.

Other genes at the rearrangement hotspot

It remains possible that the intron 1 rearrangements affect other genes at this locus which could result in pro-malignant effects. *RP11–199F11.2* (Hp53int1 or D17S2179E) is a non-spliced, probably non-coding transcript in intron 1 of *TP53* which is transcribed in the same orientation as *TP53* (Figure 1B). It has been identified by a targeted cDNA library screen but its function is unknown [26]. Further, exon 1 of *TP53* overlaps with *WRAP53*, a gene which is oriented in antisense relative to *TP53*, on the plus strand of chromosome 17, that has been found to upregulate *TP53* transcripts. Downregulation of *WRAP53* is reported to lead to significant suppression of p53 induction in response to DNA damage [27]. *WRAP53* protein has been characterized as an essential protein for the localization and processing of nuclear ribonucleoproteins [28, 29].

Germline SVs in *TP53* intron 1 can cause LFS

We found germline rearrangements in the *TP53* intron 1 hotspot likely to be causative for LFS in a large family. Our findings are in agreement with the second hit model for tumor suppressor genes where all 23 FISH positive OS showed two break-apart signals and the LFS family had the rearrangement in a heterozygous state in the germline followed by LOH in all tumors. Our qRT-PCR experiments suggest that the overall transcription of *TP53* is lost upon intron 1 rearrangement but it remains possible that in certain tissue/tumor contexts or depending on regulatory elements in the rearrangement partner sites, such rearrangements could cause a shift towards the expression of the shorter isoforms Δ133p53 and Δ160p53. In contrast to p53, Δ133p53 is defective in promoting apoptosis [21]. The comparison of *TP53* transcripts in breast tumor versus normal tissues revealed obvious differences with Δ133p53 being expressed in 24 out of 30 breast tumors but not in 8 normal breast tissue samples [21]. This is supported by a clinical study which found Δ133p53 isoforms to be abnormally expressed in renal cancer, suggesting that they play a role in carcinogenesis [30]. Further, Δ133p53α inhibits p53-mediated replicative senescence, promotes cellular proliferation of normal human fibroblasts by inhibiting p21 expression, and represses the expression of miR-34a to regulate p53-mediated senescence [31]. Cotransfection experiments indicated that Δ133p53 has a dominant negative effect on the proapoptotic properties of p53 [21]. p53 forms tetramers to execute its function as a transcription factor. Coimmunoprecipitation experiments showed that Δ133p53 forms a complex with p53 and it is likely that this interaction mediates a negative effect on p53 function of the intact allele by interfering with the tetramer structure [19].

TP53 coding mutations are identified in 70% of classic LFS families [3]. Germline *TP53* intron 1 rearrangements might be present in some of the remaining 30% of families, suggesting that LFS and Li-Fraumeni-like

(LFL) families should be tested for such rearrangements. This could be tested using FISH and *TP53* capture-based sequencing approaches for break point identification as in our study. Only a few cases have been reported where genomic structural alterations, usually deletions, cause cancer syndromes [32–39] and one study where a Robertsonian translocation results in a highly increased risk for childhood acute lymphoblastic leukemia with focal chromosome 21 amplification [40]. We present here a very rare phenomenon of a rearrangement hotspot which can give rise to both somatic rearrangements as well as a germline cancer syndrome. It seems possible that the open chromatin state together with components of the transcription machinery and characteristic LINE sequences underlie the fragility of this locus. The disruption of the gene structure results in loss of p53 function and thereby promotes cancer. In conclusion, intron 1 of *TP53* represents a rare case of a tumor type dependent somatic rearrangement hotspot that can also acquire germline SVs causing a Mendelian inherited cancer syndrome.

METHODS

OS patient samples for DNA-PET sequencing

DNA of four treatment naive OS tumors and paired normal blood were obtained from the Biopathology Center (BPC) of the Children's Oncology Group (COG), a cooperative group that includes medical centers in the United States, Canada, Mexico, Australia, New Zealand and selected countries in Europe. Informed consent of the participating patients or legal representatives has been obtained and approval of the respective institutional ethics review boards has been granted.

DNA-PET libraries construction, sequencing, mapping and data analysis

DNA-PET library construction from 1 to 4 kb fragments of genomic DNA, sequencing, mapping and data analysis were performed as described in [13] with refined bioinformatics filtering as described in [41]. High throughput sequencing by 2 x 35 bp or 2 x 50 bp was performed on SOLiD sequencers (v3plus and v4, respectively) according to the manufacturer's recommendation (Life Technologies). The short reads were aligned to the NCBI human reference genome build 37 (hg19) using Bioscope (Life Technologies).

Custom sequence capture and breakpoint identification in a LFS family by paired-end sequencing

The *TP53*-containing region chr17:7,520,000–7,680,000 (NCBI build 37) was defined for a custom sequence capture (SeqCap EZ Choice, Roche NimbleGen Inc.). Repetitive regions are usually excluded for sequence capture assays. Since most of the observed breakpoints in OS were in LINE sequences, we forced to include repetitive sequences of the intron 1 region of *TP53* (chr17:7,579,941–7,590,694). Illumina sequencing library was constructed and capturing was performed according to the manufacturer's recommendations and the library was sequenced on an Illumina HiSeq 2000 by 2x50 bases. Reads were mapped to the human reference genome (NCBI build 37) by BWA and read-pairs were filtered of which one read mapped to the *TP53* target region and the paired read mapped outside of that region. Reads of similar mapping pattern were clustered together as described earlier [13] and prioritized for validation by PCR and Sanger sequencing based on cluster size (number of paired-reads with similar mapping patterns). Identified *TP53* breakpoints were screened by PCR in all LFS family members of whom DNA was available. Validation PCRs were conducted using Jumpstart REDAccuTaq LA DNA Polymerase (Sigma-Aldrich) following manufacturer's instructions with 120 ng DNA template. Where sample was insufficient, the genomic DNA was amplified using REPLI-g Mini Kit (Qiagen) and 3 ul of 1:20 diluted amplification product were used as template for validation PCRs. PCRs were performed using the following primer pairs: CTCAAAAGGCC ATCAAAAGG and GTATGGTGGCCTGTTCCTGT (LFS-BP1), GGCTGCTGGGAGTTGTAGTC and AGCT ATCTTGACCCCACACG (LFS-BP2), CCCGAATA GCTGGGATTACA and GCAAGTGCAC AGGAAGATGA (LFS-BP1-wt), GGAGGAATCC TGCATTGTGT and CAGGCTTCAGACCTGTCTCC (LFS-BP2-wt), GCTGCTGGACGTGAGTATGA and AGTTCCAACAATGGCTACCG (positive control primers for POLR2A), and the following cycling conditions: 3 min at 94°C, (20 s at 94°C, 30 s at 58°C, 2 min at 68°C)x15, (20 s at 94°C, 30 s at 55°C, 2 min at 68°C)x20, 5 min at 68°C.

Tissue samples and patient's characteristics for fluorescence *in situ* hybridization (FISH) and CytoScan screen

All tissue samples were obtained from the archives of the Bone Tumor Reference Center at the University Hospital Basel and the Clinical Cooperation Group Osteosarcoma at the Helmholtz Zentrum Muenchen. Only specimens from patients without prior treatment were included in the study. FISH analyses were performed on nine TMAs comprising OS samples of 267 patients, samples from other bone-forming tumors or tumor-like lesions of 144 patients, and a collection of 1,163 tumors of various types and normal tissue samples. For the CytoScan HD Arrays (Affymetrix, CA, USA) genomic DNA from fresh frozen samples of 73 independent OS patients was used. Full patient's characteristics are presented in Table 1 and Supplementary Tables S8, S10 and S11.

Dual-color break-apart FISH

BAC clones RP11–1081A10 and RP11–107F4 (BACPAC, Children's Hospital Oakland Research Institute, USA) were nick-translated (Abbott Laboratories, USA) and labeled with fluorescent dUTPs (Enzo Life Sciences, USA) resulting in red labeling of RP11–1081A10 and green labeling of RP11–107F4. The *in situ* hybridisation was performed according to routine protocols. Tumors were considered to have a rearrangement in the 5' region of *TP53* when at least 10% of cells showed clearly separated green and red hybridisation signals (= FISH positive/break-apart).

CytoScan HD arrays

Genome-wide CytoScan HD Arrays (Affymetrix, CA, USA) were performed according to the manufacturer's instructions using 250 ng of genomic DNA from each tumor sample. To evaluate copy number variations (CNV), data was processed using the Nexus Copy Number software (Version 7.0, BioDiscovery, CA, USA) with integrated algorithms for segmentation and normalization.

Clinical characteristics of a four-generation LFS family

All members of the LFS family who contributed samples to this study or legal representatives gave informed consent and are indicated with numbers in Figure 3A. The study was approved by the respective ethical review boards. P1 was diagnosed with colon cancer at age of 22 years, an oligodendroglioma at 34 years, and an undifferentiated pleomorphic sarcoma at 40 years, and basal cell carcinoma and squamous cell carcinoma of the skin, P2 a phyllodes tumor of the breast at 12 years, P6 a rhabdomyosarcoma at 33 months, P12 an OS at 19 years, H2 an OS at 13 years, P13 bilateral ductal carcinoma *in situ* at 32 and 33 years, meningioma at 38 years, and adenocarcinoma of the lung at 40 years. Family members P1, P7, P13, P12, H1, H2, H4, H5 and H6 were analyzed by multiplex ligation-dependent probe amplification (MLPA) (MRC-Holland) according to the manufacturer's recommendation to test for deletions in *TP53* with 2 probe sets in exon 1, and one probe set each in exons 2 – 9, 11, and 12. Both probe sets for exon 1 with probe pair TGTAGCTGCTGGGCTCCGGGGACACT and TTGCGTTCGGGCTGGGAGCGTGCTTTCCACGA (exon 1) and CCATTTCCTTTGCTTCCTCCGGCA and GGCGGATTACTTGCCCTTACTTGTCATGGCGACTGT CCAG (5' of exon 1) indicated a heterozygous deletion of exon 1 for P1, P13, P12, H2 and H4 but not for P7, H1, H5 and H6. The presence of the deletion co-segregated with affection status except for H4 who was tested positive but had no cancer at the age of 10 years (of note, H4 is being followed with repeat imaging due to suspicious lesion in the brain that could represent the development of an early tumor).

qRT-PCR of *TP53* in samples of LFS family

RNA was isolated from PAXgene Blood RNA Tubes using the PAXgene Blood miRNA kit (Qiagen). RNA from fresh frozen tumor, H2, and RNA from a tumor derived cell line, P13, were isolated using RNAeasy Mini kit (Qiagen). The one step qRT-PCR was carried out using QuantiTect Probe RT-PCR kit (Qiagen), with a total of 50 ng RNA per PCR. The TaqMan primer/probe set (LifeTechnologies) was used for *TP53* full length (HS01034249) and the PrimeTime primer/probe set (Integrated DNA Technologies Inc.) for transcripts encoding for Δ133p53 and Δ160p53 were designed using sequences from Marcel and colleagues [42]. qRT-PCRs were performed on a Bio-Rad CFX device. Cq values were normalized against *GAPDH* as ΔCq and displayed relative to normal blood control as ΔΔCq as fold-change ($2^{-\Delta\Delta Cq}$).

OncoScan FFPE Express (Molecular Inversion Probe) array

Samples were processed using the Affymetrix OncoScan FFPE Express kit according to the manufacturer's instructions using 80 ng of genomic DNA from each tumor sample. To evaluate copy number variations (CNV) data was processed using the Nexus Copy Number software (Version 7.5, BioDiscovery).

Sequencing data

Sequences have been submitted to the Short Read Archive (http://trace.ncbi.nlm.nih.gov/Traces/sra/) at the National Center for Biotechnology Information (NCBI) with the study reference number PRJNA244486. LFS array data can be accessed at Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) with the study ID GSE64293.

ACKNOWLEDGMENTS

We are grateful for all patients who contributed tissue samples to this study. We thank Ioana Cutcutache and Steve Rozen for bioinformatics analysis of LFS family data, Kyle D. Proffitt for help on tissue culture and molecular biology work, Joanne Hui Ping Lim for help on FISH probe labeling, Dawn Poh Sum Choi, See Ting Leong, and Say Chuan Neo for SOLiD sequencing, Jo Chin Thing Ong, Ming Lai Ivan Chua, and Yen Ling Lee for Illumina sequencing, and Joanna Hui Juan Tan, Chee Seng Chan, Pramila N. Ariyaratne, Pauline Jieqi Chen, Justin Jeyakani for bioinformatics assistance.

FINANCIAL SUPPORT

This work was supported by the Agency for Science Technology and Research (A*STAR). DMV

is supported by the Singapore Translational Research Investigator Award of the National Medical Research Council. SR is supported by the A*STAR Research Attachment Programme. DB and GJ are supported by the Foundation for the Preservation of the Basel Bone Tumor Reference Center. KL is supported by the Soccer for Hope Foundation. JS and MN are members of the Translational Sarcoma Research Network supported by the Bundesministerium für Bildung und Forschung (FKZ01GM0870). JDS is supported by the Edward B. Clark, MD Chair in Pediatric Research and the Primary Children's Hospital (PCH) Pediatric Cancer Program. LFS experiments supported by funds in conjunction with Soccer for Hope Foundation and grant P30 CA042014 awarded to Huntsman Cancer Institute.

CONFLICT OF INTEREST

All authors declare that they have no conflict of interest

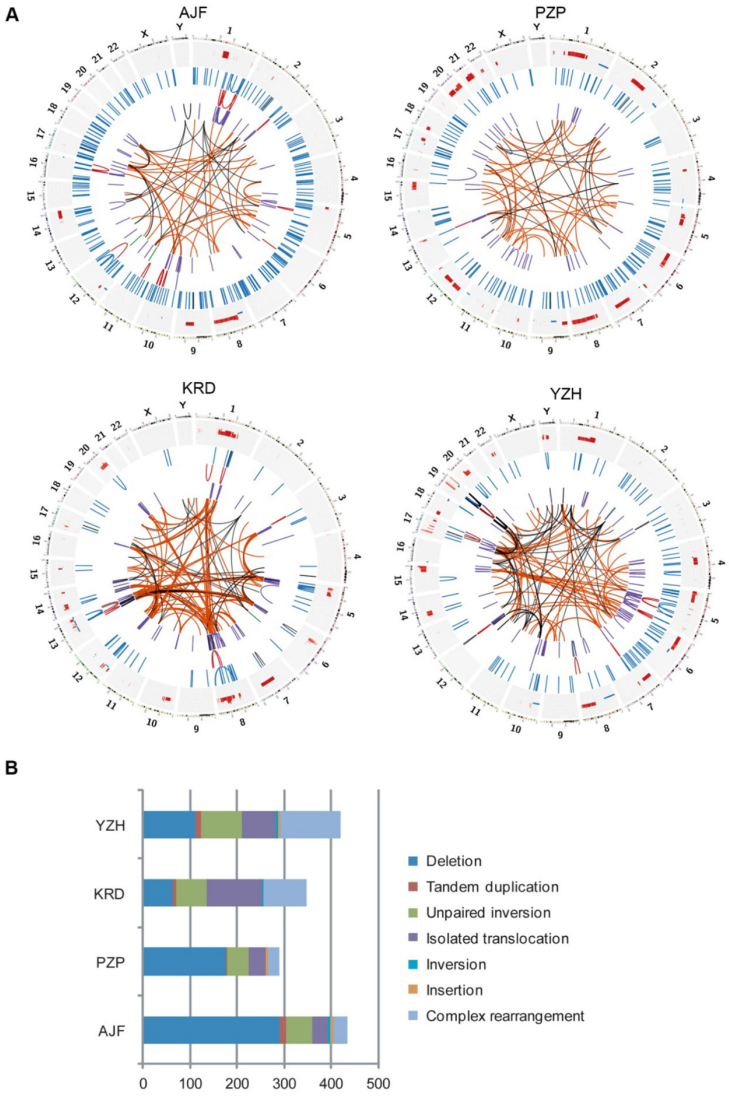
REFERENCES

1. Malkin D. Li-Fraumeni syndrome. In: Fletcher C, Bridge J, Hogendoorn P and Mertens F, eds. WHO Classification of Tumours of Tumours of Soft Tissue and Bone. (Lyon, France: IARC Press), 2013; pp. 379–381.
2. McBride KA, Ballinger ML, Killick E, Kirk J, Tattersall MH, Eeles RA, Thomas DM, Mitchell G. Li-Fraumeni syndrome: cancer risk assessment and clinical management. Nat Rev Clin Oncol. 2014; 11:260–271.
3. Varley JM, McGown G, Thorncroft M, Santibanez-Koref MF, Kelsey AM, Tricker KJ, Evans DG, Birch JM. Germ-line mutations of TP53 in Li-Fraumeni families: an extended study of 39 families. Cancer Res. 1997; 57:3245–3252.
4. Petitjean A, Achatz MI, Borresen-Dale AL, Hainaut P, Olivier M. TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. Oncogene. 2007; 26:2157–2165.
5. Shlien A, Baskin B, Achatz MI, Stavropoulos DJ, Nichols KE, Hudgins L, Morel CF, Adam MP, Zhukova N, Rotin L, Novokmet A, Druker H, Shago M, Ray PN, Hainaut P, Malkin D. A common molecular mechanism underlies two phenotypically distinct 17p13.1 microdeletion syndromes. Am J Hum Genet. 2010; 87:631–642.
6. Bougeard G, Brugieres L, Chompret A, Gesta P, Charbonnier F, Valent A, Martin C, Raux G, Feunteun J, Bressac-de Paillerets B, Frebourg T. Screening for TP53 rearrangements in families with the Li-Fraumeni syndrome reveals a complete deletion of the TP53 gene. Oncogene. 2003; 22:840–846.
7. Masuda H, Miller C, Koeffler HP, Battifora H, Cline MJ. Rearrangement of the p53 gene in human osteogenic sarcomas. Proc Natl Acad Sci U S A. 1987; 84:7716–7719.
8. Miller CW, Aslo A, Tsay C, Slamon D, Ishizaki K, Toguchida J, Yamamuro T, Lampkin B, Koeffler HP. Frequency and structure of p53 rearrangements in human osteosarcoma. Cancer Res. 1990; 50:7950–7954.
9. Chandar N, Billig B, McMaster J, Novak J. Inactivation of p53 gene in human and murine osteosarcoma cells. Br J Cancer. 1992; 65:208–214.
10. Chen X, Bahrami A, Pappo A, Easton J, Dalton J, Hedlund E, Ellison D, Shurtleff S, Wu G, Wei L, Parker M, Rusch M, Nagahawatte P, Wu J, Mao S, Boggs K, et al. Recurrent Somatic Structural Variations Contribute to Tumorigenesis in Pediatric Osteosarcoma. Cell Rep. 2014; 7:104–112.
11. Kuerbitz SJ, Plunkett BS, Walsh WV, Kastan MB. Wild-type p53 is a cell cycle checkpoint determinant following irradiation. Proc Natl Acad Sci U S A. 1992; 89:7491–7495.
12. Hanel W, Moll UM. Links between mutant p53 and genomic instability. Journal of cellular biochemistry. 2012; 113:433–439.
13. Hillmer AM, Yao F, Inaki K, Lee WH, Ariyaratne PN, Teo AS, Woo XY, Zhang Z, Zhao H, Ukil L, Chen JP, Zhu F, So JB, Salto-Tellez M, Poh WT, Zawack KF, et al. Comprehensive long-span paired-end-tag mapping reveals characteristic patterns of structural variations in epithelial cancer genomes. Genome Res. 2011; 21:665–675.
14. Ng KP, Hillmer AM, Chuah CT, Juan WC, Ko TK, Teo AS, Ariyaratne PN, Takahashi N, Sawada K, Fei Y, Soh S, Lee WH, Huang JW, Allen JC Jr., Woo XY, Nagarajan N, et al. A common BIM deletion polymorphism mediates intrinsic resistance and inferior responses to tyrosine kinase inhibitors in cancer. Nat Med. 2012; 18:521–528.
15. Yao F, Ariyaratne PN, Hillmer AM, Lee WH, Li G, Teo AS, Woo XY, Zhang Z, Chen JP, Poh WT, Zawack KF, Chan CS, Leong ST, Neo SC, Choi PS, Gao S, et al. Long span DNA paired-end-tag (DNA-PET) sequencing strategy for the interrogation of genomic structural mutations and fusion-point-guided reconstruction of amplicons. PLoS One. 2012; 7:e46152.
16. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, et al. Landscape of transcription in human cells. Nature. 2012; 489:101–108.
17. Hessle L, Johnson KA, Anderson HC, Narisawa S, Sali A, Goding JW, Terkeltaub R, Millan JL. Tissue-nonspecific alkaline phosphatase and plasma cell membrane glycoprotein-1 are central antagonistic regulators of bone mineralization. Proc Natl Acad Sci U S A. 2002; 99:9445–9449.
18. Sohaskey ML, Jiang Y, Zhao JJ, Mohr A, Roemer F, Harland RM. Osteopotentia regulates osteoblast maturation, bone formation, and skeletal integrity in mice. J Cell Biol. 2010; 189:511–525.
19. Khoury MP, Bourdon JC. The isoforms of the p53 protein. Cold Spring Harb Perspect Biol. 2010; 2:a000927.

20. Khoury MP, Bourdon JC. p53 Isoforms: An Intracellular Microprocessor? Genes Cancer. 2011; 2:453–465.
21. Bourdon JC, Fernandes K, Murray-Zmijewski F, Liu G, Diot A, Xirodimas DP, Saville MK, Lane DP. p53 isoforms can regulate p53 transcriptional activity. Genes Dev. 2005; 19:2122–2137.
22. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. Pathogenetics. 2008; 1:4.
23. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. Nat Rev Genet. 2009; 10:551–564.
24. Mutsaers AJ, Walkley CR. Cells of origin in osteosarcoma: mesenchymal stem cells or osteoblast committed cells? Bone. 2014; 62:56–63.
25. Blackburn AC, Jerry DJ. Knockout and transgenic mice of Trp53: what have we learned about p53 in breast cancer? Breast Cancer Res. 2002; 4:101–111.
26. Reisman D, Balint e, Loging WT, Rotter V, Almon E. A novel transcript encoded within the 10-kb first intron of the human p53 tumor suppressor gene (D17S2179E) is induced during differentiation of myeloid leukemia cells. Genomics. 1996; 38:364–370.
27. Mahmoudi S, Henriksson S, Corcoran M, Mendez-Vidal C, Wiman KG, Farnebo M. Wrap53, a natural p53 antisense transcript required for p53 induction upon DNA damage. Mol Cell. 2009; 33:462–471.
28. Tycowski KT, Shu MD, Kukoyi A, Steitz JA. A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles. Mol Cell. 2009; 34:47–57.
29. Venteicher AS, Abreu EB, Meng Z, McCann KE, Terns RM, Veenstra TD, Terns MP, Artandi SE. A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. Science. 2009; 323:644–648.
30. Song W, Huo SW, Lu JJ, Liu Z, Fang XL, Jin XB, Yuan MZ. Expression of p53 isoforms in renal cell carcinoma. Chin Med J (Engl). 2009; 122:921–926.
31. Fujita K, Mondal AM, Horikawa I, Nguyen GH, Kumamoto K, Sohn JJ, Bowman ED, Mathe EA, Schetter AJ, Pine SR, Ji H, Vojtesek B, Bourdon JC, Lane DP, Harris CC. p53 isoforms Delta133p53 and p53beta are endogenous regulators of replicative cellular senescence. Nat Cell Biol. 2009; 11:1135–1142.
32. Andersson U, Wibom C, Cederquist K, Aradottir S, Borg A, Armstrong GN, Shete S, Lau CC, Bainbridge MN, Claus EB, Barnholtz-Sloan J, Lai R, Il'yasova D, Houlston RS, Schildkraut J, Bernstein JL, et al. Germline rearrangements in families with strong family history of glioma and malignant melanoma, colon, and breast cancer. Neuro Oncol. 2014; 16:1333–1340.
33. Gylling A, Ridanpaa M, Vierimaa O, Aittomaki K, Avela K, Kaariainen H, Laivuori H, Poyhonen M, Sallinen SL, Wallgren-Pettersson C, Jarvinen HJ, Mecklin JP, Peltomaki P. Large genomic rearrangements and germline epimutations in Lynch syndrome. Int J Cancer. 2009; 124:2333–2340.
34. Judkins T, Rosenthal E, Arnell C, Burbidge LA, Geary W, Barrus T, Schoenberger J, Trost J, Wenstrup RJ, Roa BB. Clinical significance of large rearrangements in BRCA1 and BRCA2. Cancer. 2012; 118:5210–5216.
35. Lee MP, Hu RJ, Johnson LA, Feinberg AP. Human KVLQT1 gene shows tissue-specific imprinting and encompasses Beckwith-Wiedemann syndrome chromosomal rearrangements. Nat Genet. 1997; 15:181–185.
36. Plaschke J, Ruschoff J, Schackert HK. Genomic rearrangements of hMSH6 contribute to the genetic predisposition in suspected hereditary non-polyposis colorectal cancer syndrome. J Med Genet. 2003; 40:597–600.
37. Duraturo F, Cavallo A, Liccardo R, Cudia B, De Rosa M, Diana G, Izzo P. Contribution of large genomic rearrangements in Italian Lynch syndrome patients: characterization of a novel alu-mediated deletion. Biomed Res Int. 2013; 2013:219897.
38. Tutlewska K, Lubinski J, Kurzawski G. Germline deletions in the EPCAM gene as a cause of Lynch syndrome - literature review. Hered Cancer Clin Pract. 2013; 11:9.
39. Andersson U, Wibom C, Cederquist K, Aradottir S, Borg Å, Armstrong GN, Shete S, Lau CC, Bainbridge MN, Claus EB, Barnholtz-Sloan J, Lai R, Il'yasova D, Houlston RS, Schildkraut J, Bernstein JL, et al. Germline rearrangements in families with strong family history of glioma and malignant melanoma, colon, and breast cancer. Neuro-Oncology. 2014; 16:1333–1340.
40. Li Y, Schwab C, Ryan SL, Papaemmanuil E, Robinson HM, Jacobs P, Moorman AV, Dyer S, Borrow J, Griffiths M, Heerema NA, Carroll AJ, Talley P, Bown N, Telford N, Ross FM, et al. Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia. Nature. 2014; 508:98–102.
41. Nagarajan N, Bertrand D, Hillmer AM, Zang ZJ, Yao F, Jacques PE, Teo AS, Cutcutache I, Zhang Z, Lee WH, Sia YY, Gao S, Ariyaratne PN, Ho A, Woo XY, Veeravali L, et al. Whole-genome reconstruction and mutational signatures in gastric cancer. Genome Biol. 2012; 13:R115.
42. Marcel V, Perrier S, Aoubala M, Ageorges S, Groves MJ, Diot A, Fernandes K, Tauro S, Bourdon JC. Delta160p53 is a novel N-terminal p53 isoform encoded by Delta133p53 transcript. FEBS Lett. 2010; 584:4463–4468.
43. Rhead B, Karolchik D, Kuhn RM, Hinrichs AS, Zweig AS, Fujita PA, Diekhans M, Smith KE, Rosenbloom KR, Raney BJ, Pohl A, Pheasant M, Meyer LR, Learned K, Hsu F, Hillman-Jackson J, et al. The UCSC Genome Browser database: update. Nucleic Acids Res. 2010; 38:D613–619.
44. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74.

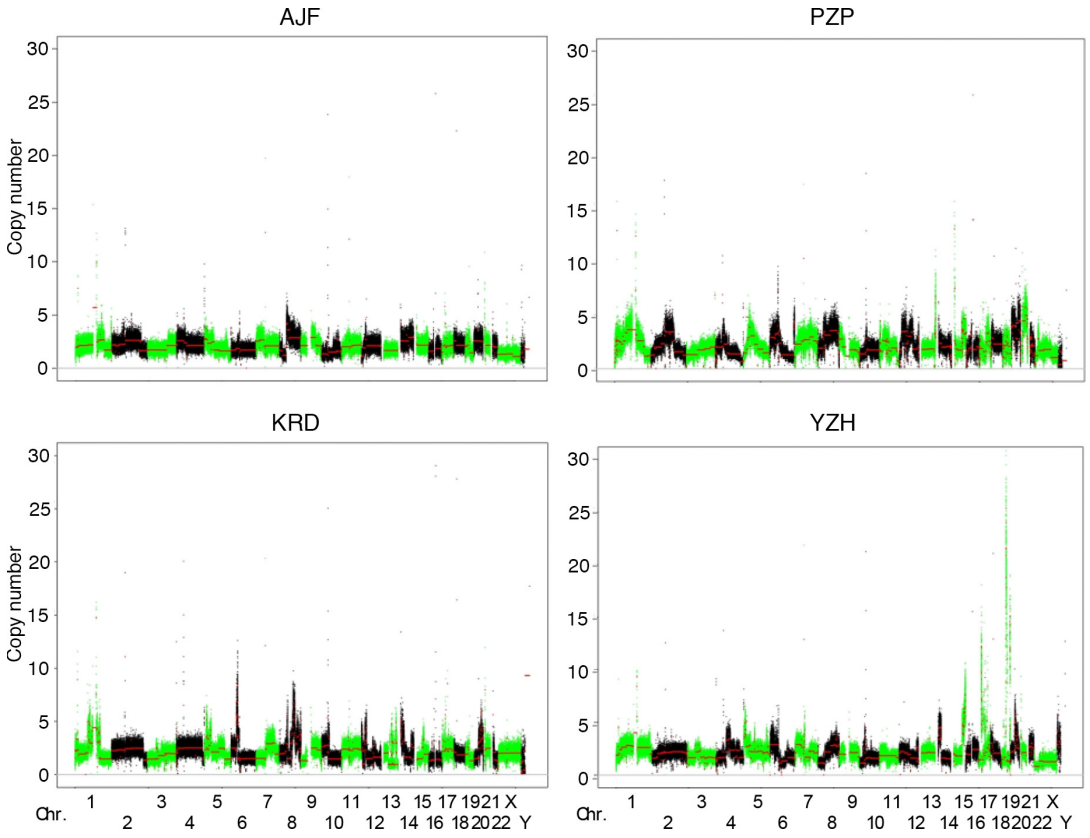
www.impactjournals.com/oncotarget/ Oncotarget, Supplementary Materials 2015

SUPPLEMENTARY FIGURES AND TABLES



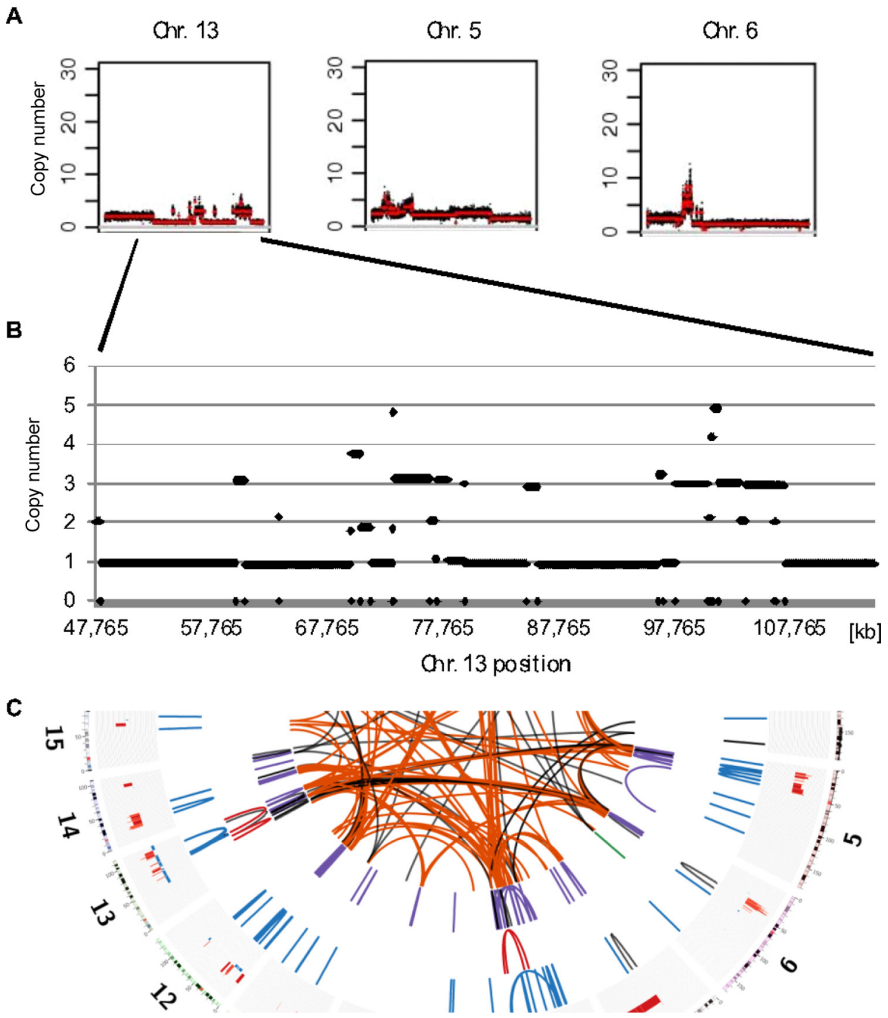
Supplementary Figure S1: Genome rearrangements of four primary OS analyzed by DNA-PET sequencing. **(A)** Genomic SVs enriched for somatic events identified by genome-wide mate-pair sequencing of long DNA fragments of four OS tumors are displayed by Circos [1]. The human genome is arranged in a circle with the chromosome number and the cytogenetic bands displayed in the outer ring, the somatic copy number alterations in the next inner concentric ring (red, amplifications; blue, deletions), followed by deletions, tandem duplications, inversions/unpaired inversions, and in the center of the circle inter-chromosomal rearrangements. Isolated and complex inter-chromosomal rearrangements are represented by orange and black lines, respectively. **(B)** Distribution of SV categories which are enriched for somatic events in the four OS tumors shown in A, X-axis indicates absolute numbers of color coded SV types as described in the legend.

www.impactjournals.com/oncotarget/ Oncotarget, Supplementary Materials 2015

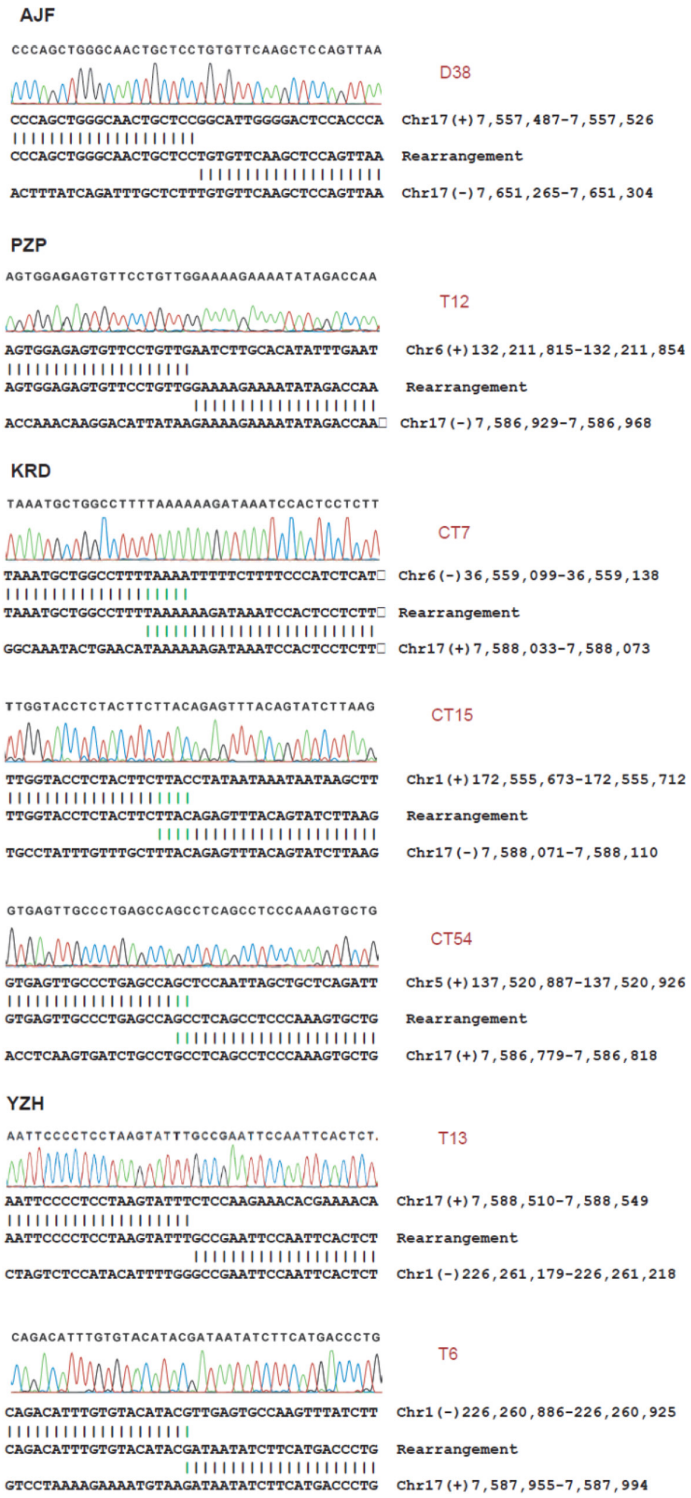


Supplementary Figure S2: Copy number profile of four primary OS derived from DNA-PET sequencing. Chromosomes are aligned on the x-axis and copy number values are on the y-axis. All four OS display a large number of copy number alterations.

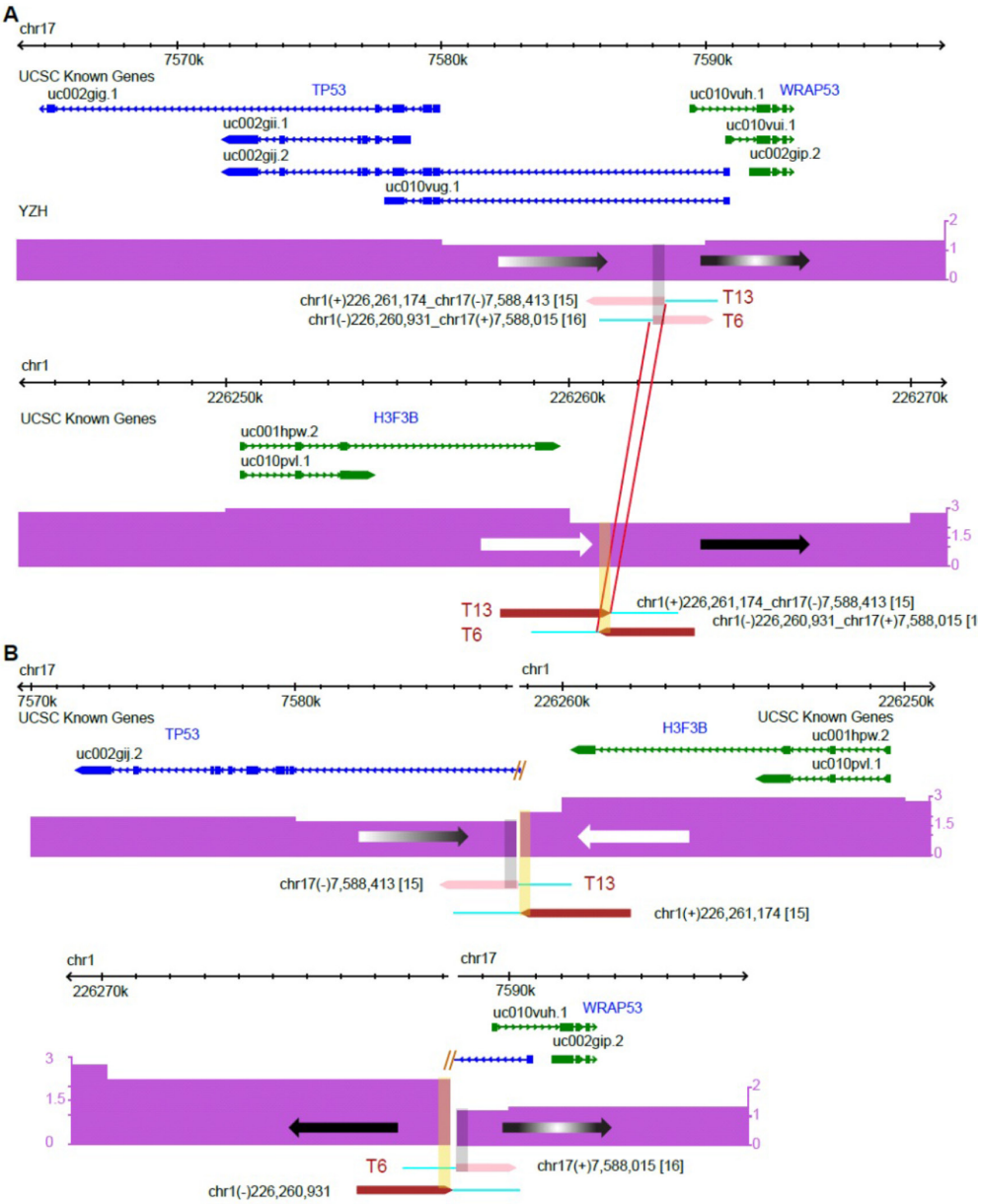
www.impactjournals.com/oncotarget/ Oncotarget, Supplementary Materials 2015



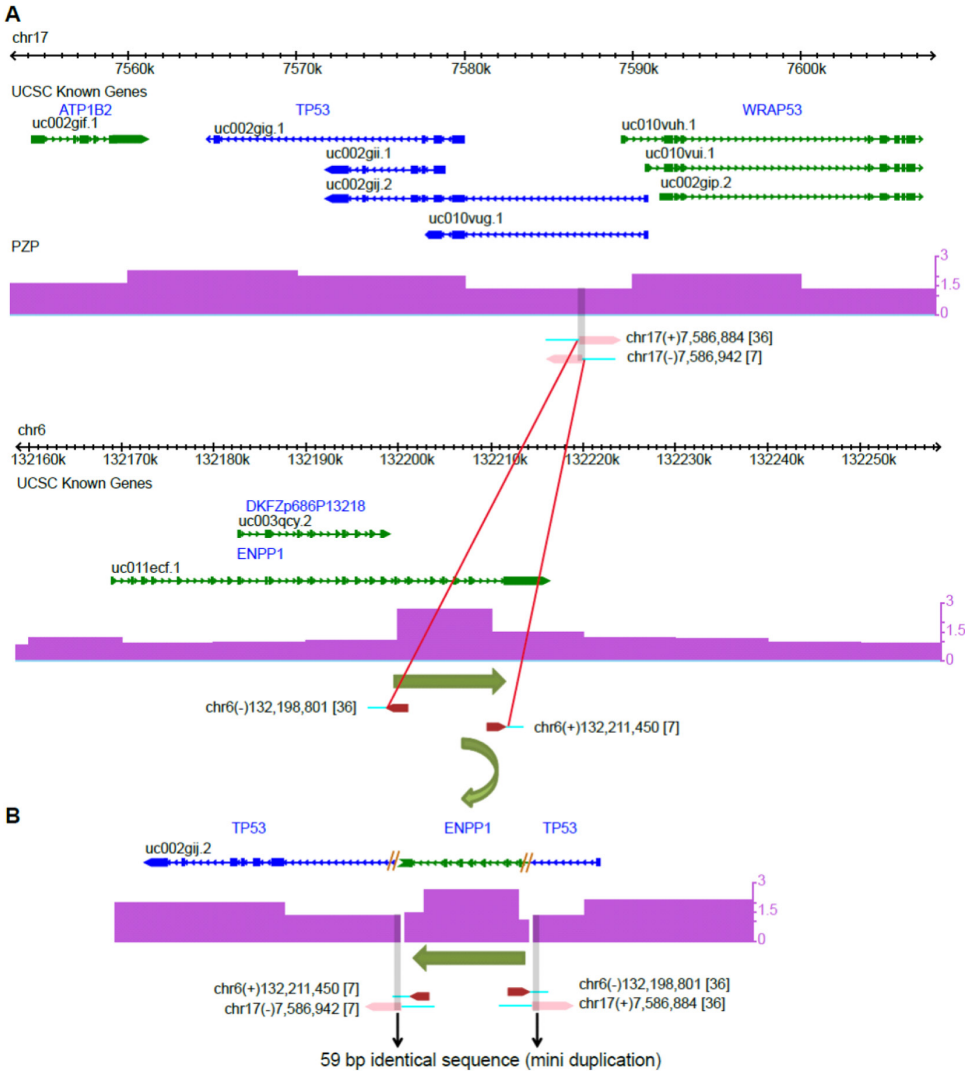
Supplementary Figure S3: Evidence for chromothripsis in OS tumor KRD. (A) Copy number profiles of chromosomes 13, 5 and 6 derived from concordant mapping paired-end tag sequencing data relative to simulation control. Chromosomes are aligned on the x-axes and copy number values are on the y-axes. Black dots indicate copy number based on 10 kb windows. Red lines are smoothened copy number across neighboring windows. **(B)** Enlarged copy number profile of chromosome 13 positions 47.7 Mb to 115.1 Mb. Data points are based on 10 kb windows. **(C)** Capture of genomic Circos plot [1] including chromosomes 13, 5 and 6. Information displayed in the concentric rings is as follows from outside to inside: chromosome number; cytogenetic banding, copy number with gains in red and losses in blue; deletions; tandem duplications; inversions/unpaired inversions; inter-chromosomal rearrangements. Color coding for inter-chromosomal SVs is indicated in orange and inter-chromosomal SVs in complex regions in black.



Supplementary Figure S4: Validation of *TP53* rearrangements by PCR and Sanger sequencing. Patient IDs are indicated in bold capital letters. Bases matching between the rearrangement and the reference sequences of the participating regions are indicated by vertical lines. Micro homologies between the two participating break point regions are illustrated by green vertical lines. Break point IDs are shown in red.

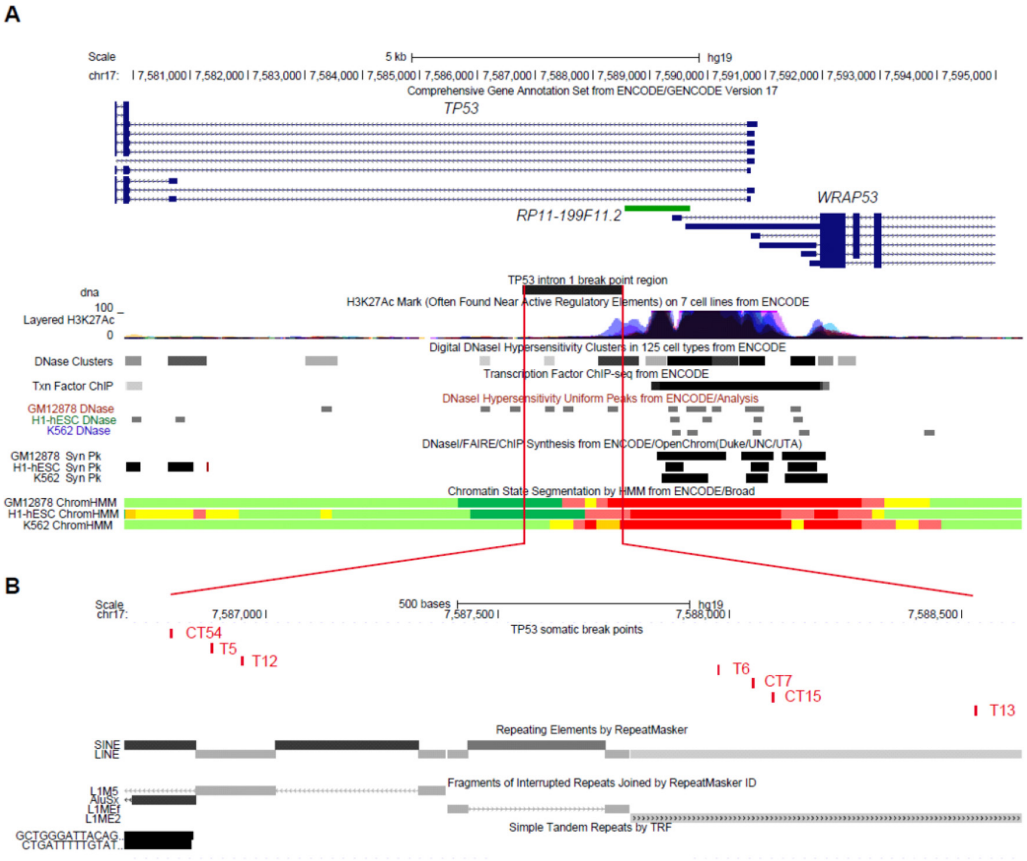


Supplementary Figure S5: Balanced translocation between intron 1 of *TP53* and a locus on chromosome 1 in OS patient YZH. (A) Chromosome 17 and 1 loci of the human reference genome (hg19) which show the PET mapping regions indicating the balanced translocation between the two loci. Genes showing the major splice variants derived from the UCSC known genes database [2] are shown on top and smoothened copy number information derived from DNA-PET sequencing data is shown at the bottom (purple tracks) in the Genome Browser. Genes transcribed from the plus strand are represented in green, genes transcribed from the minus strand are represented in blue. Boxes indicate exons, barbed lines indicate introns. Mapping regions of 5' and 3' PET clusters and read orientations are indicated by dark red and pink arrow heads, respectively, with turquois lines indicating the side of breakpoint. Gray and yellow shading indicate sequences of 555 bp and 293 bp which are shared by both sides of the balanced translocation product for chromosome 17 and 1, respectively. Red lines indicate connectivity between the two chromosomes. Note that DNA-PET coordinates differ from the exact break point coordinates which have been determined by PCR and Sanger sequencing and which are shown in Table S7. **(B)** Translocation outcome as indicated by DNA-PET truncating *TP53* and duplicating gray and yellow shaded sequences.

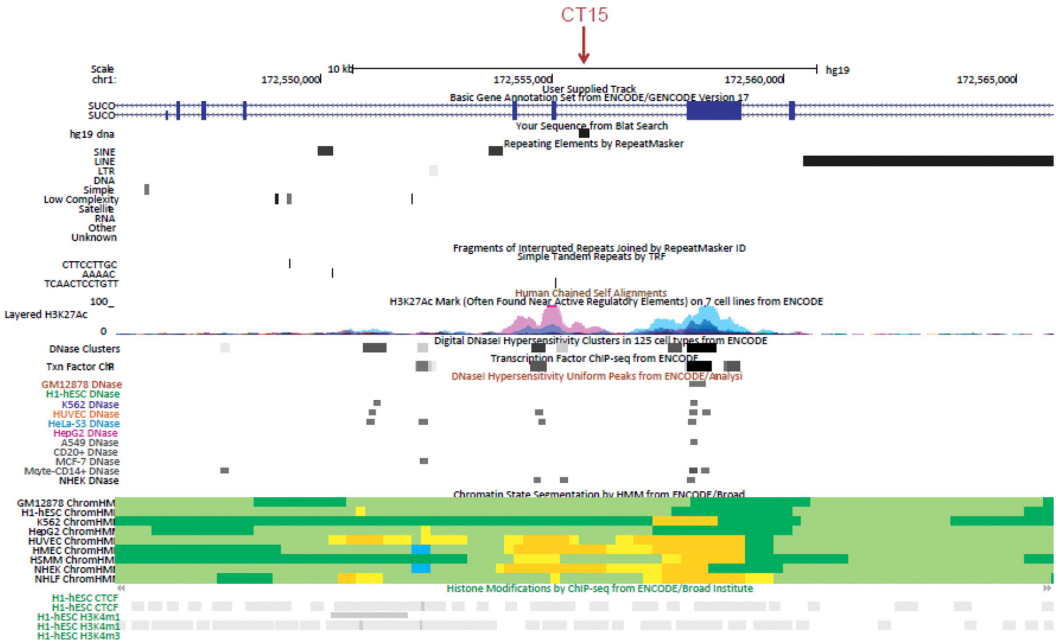


Supplementary Figure S6: Inverted insertion of 12.5 kb of chromosome 6 into intron 1 of *TP53* in OS patient PZP. Identification of an inverted insertion originating from chromosome 6 containing *ENPP1* exon 19 to the 5' part of exon 25 into intron 1 of *TP53* by DNA-PET. For more detailed track information, see legend of Supplementary Figure S5, for exact break point coordinates, see Table S7. **(A)** Chromosome 17 and 6 loci of the human reference genome (hg19) which show the PET mapping regions indicating the inverted insertion. **(B)** Insertion outcome as indicated by DNA-PET creating a potential fusion transcript between exon 1 of *TP53* (right blue box) and exon 19 to the first part of exon 25 of *ENPP1*. Since exon 1 of *TP53* is non-coding, and the inserted part of *ENPP1* exon 25 contains a stop codon, the formation of a fusion protein is unlikely.

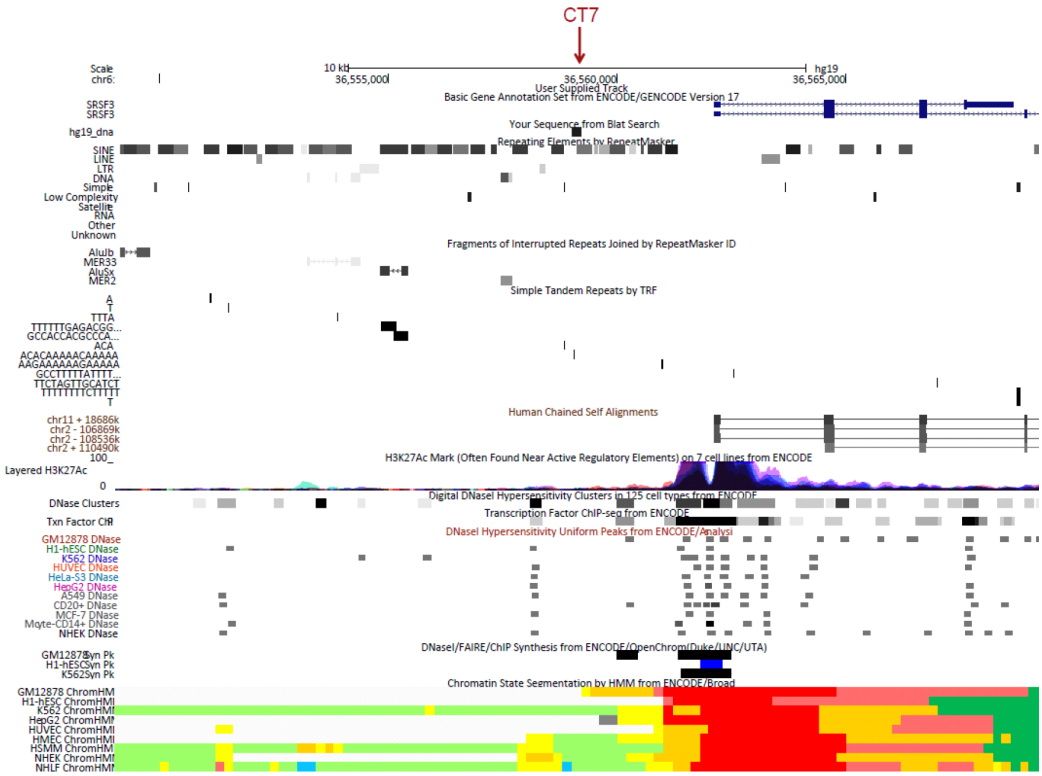
www.impactjournals.com/oncotarget/ Oncotarget, Supplementary Materials 2015



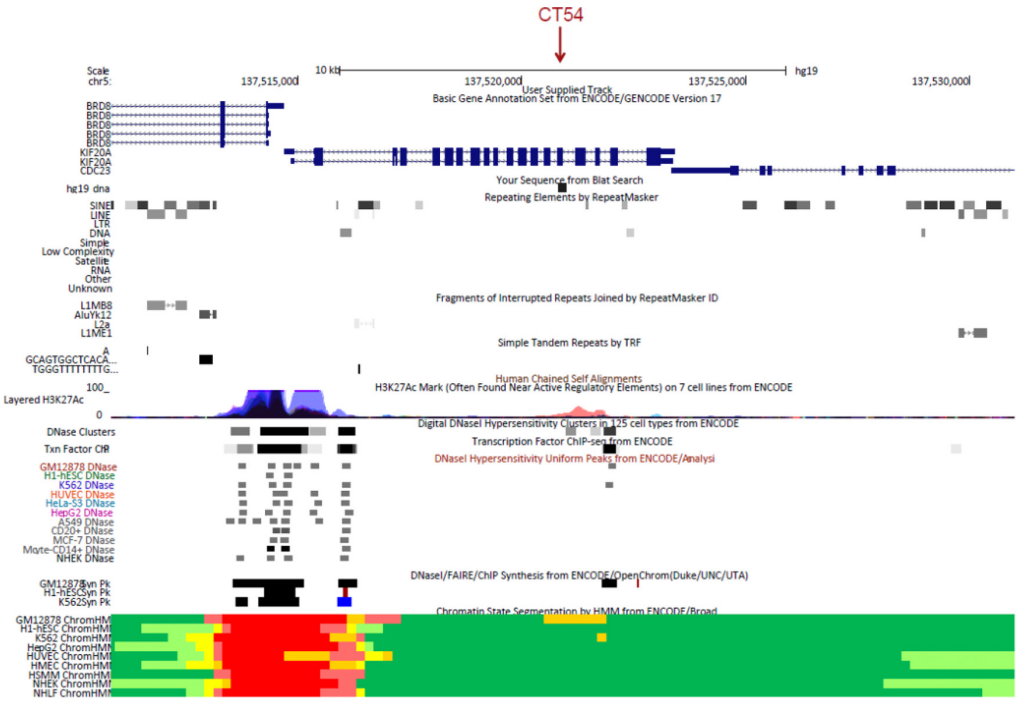
Supplementary Figure S7: *TP53* break point cluster region is located near sites of active chromatin and DNA breaks occur in LINE sequences. UCSC genome browser view (http://genome.ucsc.edu/; [2]) of the *TP53* intron 1 locus. **(A)** ENCODE epigenetic information is derived from the analysis of commonly used cell lines [3]. Tracks from top to bottom are: gene annotation after exclusion of 'problematic' transcripts; user track indicating the break point cluster region; histone 3 lysine 27 acetylation (H3K27ac) chromatin immunoprecipitation-sequencing (ChIP-seq) of seven cell lines as a superimposed layered track suggests active regulatory sites; DNaseI hypersensitive sites derived from ChIP-seq and a uniform peak analysis represent open chromatin; integrated active chromatin information prediction of DNaseI hypersensitivity, formaldehyde-assisted isolation of regulatory elements (FAIRE), and ChIP-seq of three cell lines; chromatin state segmentation analysis: dark red = active promoter, light red = weak promoter, orange = strong enhancer, yellow = weak/poised enhancer, dark green = transcriptional transition/elongation, light green = weak transcription. **(B)** *TP53* intron 1 break points are indicated by red vertical lines and IDs. Below are UCSC track information on repetitive DNA sequences.



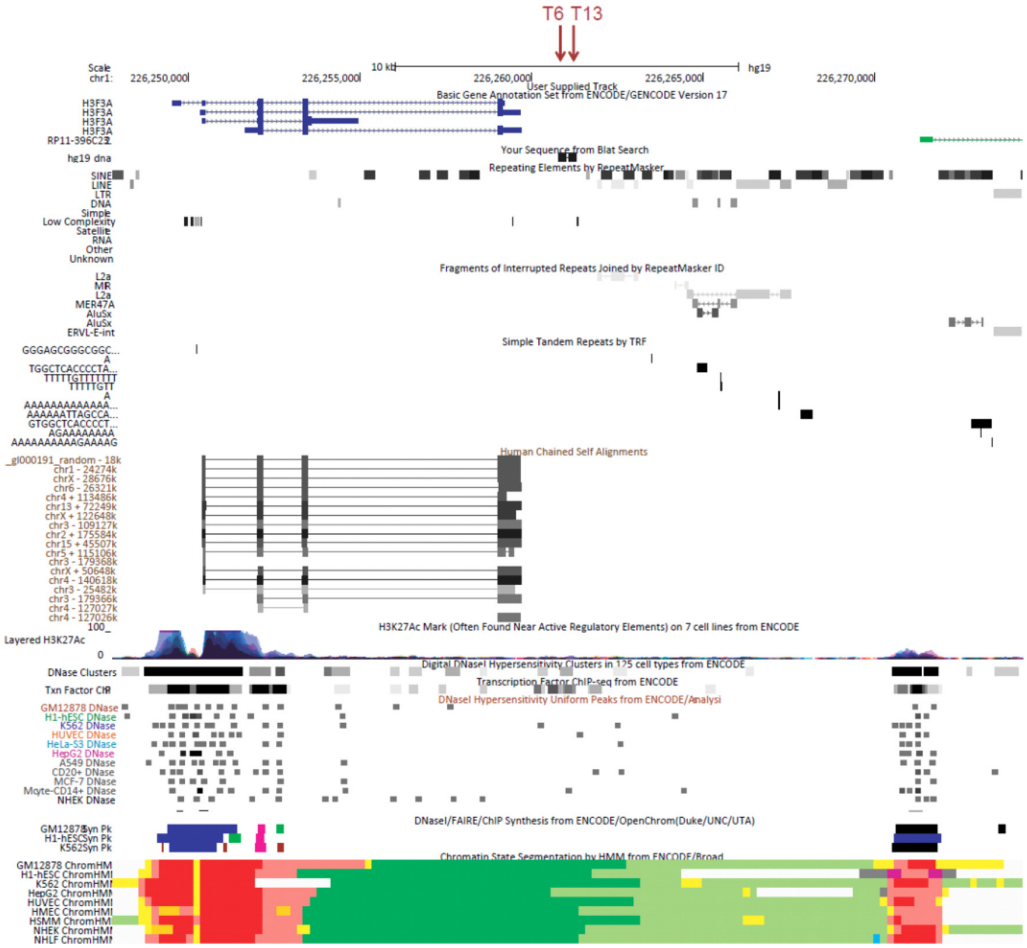
Supplementary Figure S8: Regulatory information of translocation partner site on chromosome 1 of *TP53* translocation in OS patient KRD. UCSC genome browser view of repetitive sequences and ENCODE regulatory information. Break point location is indicated by red arrow. For track description, see legend of Supplementary Figure S7.



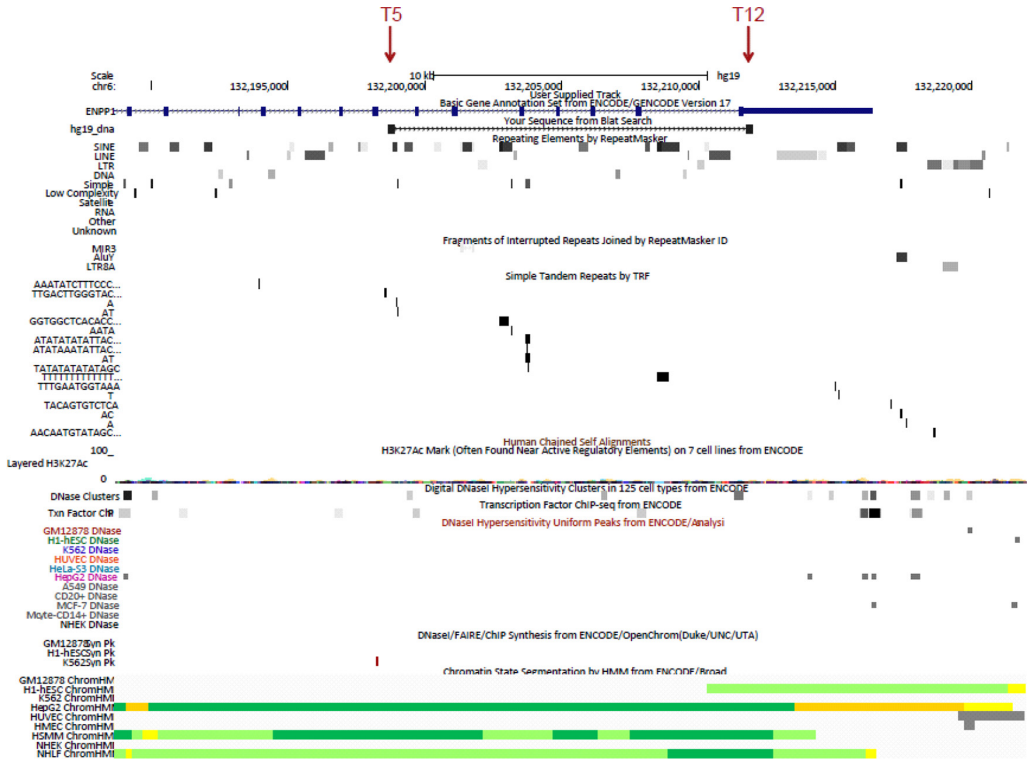
Supplementary Figure S9: Regulatory information of translocation partner site on chromosome 6 of *TP53* translocation in OS patient KRD. UCSC genome browser view of repetitive sequences and ENCODE regulatory information. Break point location is indicated by red arrow. For track description, see legend of Supplementary Figure S7.



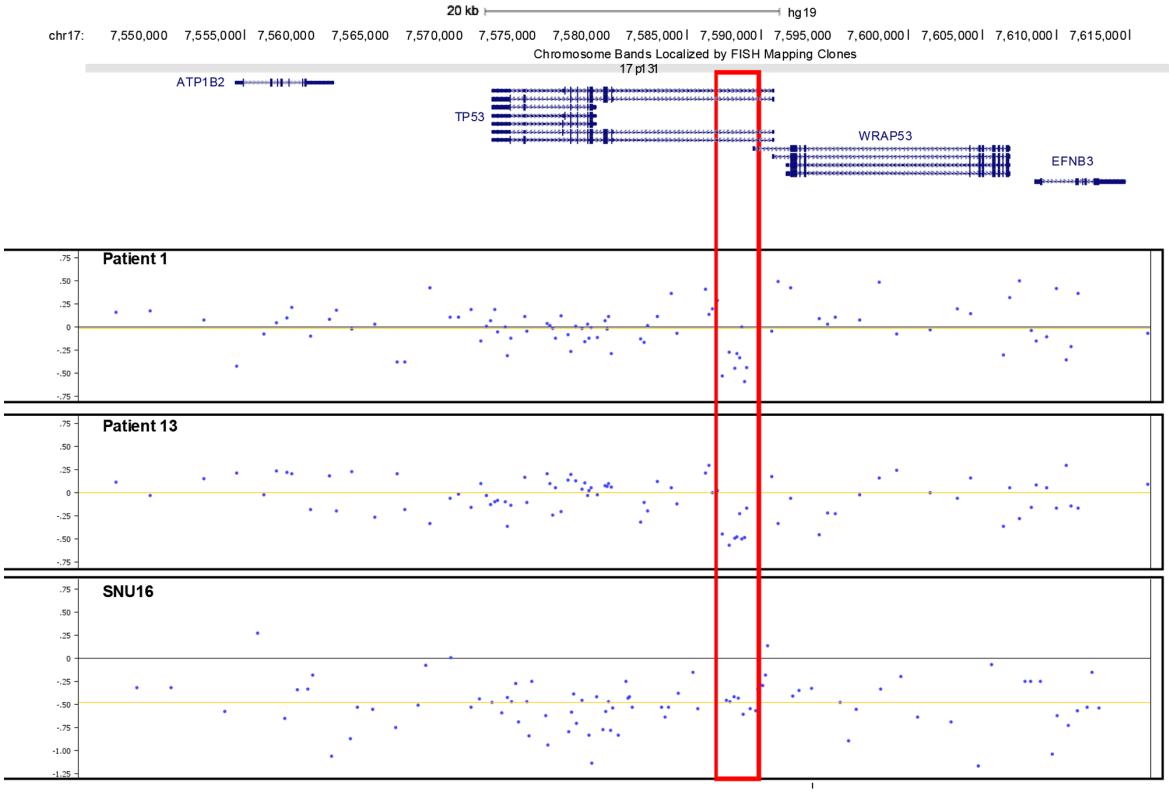
Supplementary Figure S10: Regulatory information of translocation partner site on chromosome 5 of *TP53* translocation in OS patient KRD. UCSC genome browser view of repetitive sequences and ENCODE regulatory information. Break point location is indicated by red arrow. For track description, see legend of Supplementary Figure S7.



Supplementary Figure S11: Regulatory information of partner site on chromosome 1 of *TP53* balanced translocation in OS patient YZH. UCSC genome browser view of repetitive sequences and ENCODE regulatory information. Break point locations are indicated by red arrows. For track description, see legend of Supplementary Figure S7.

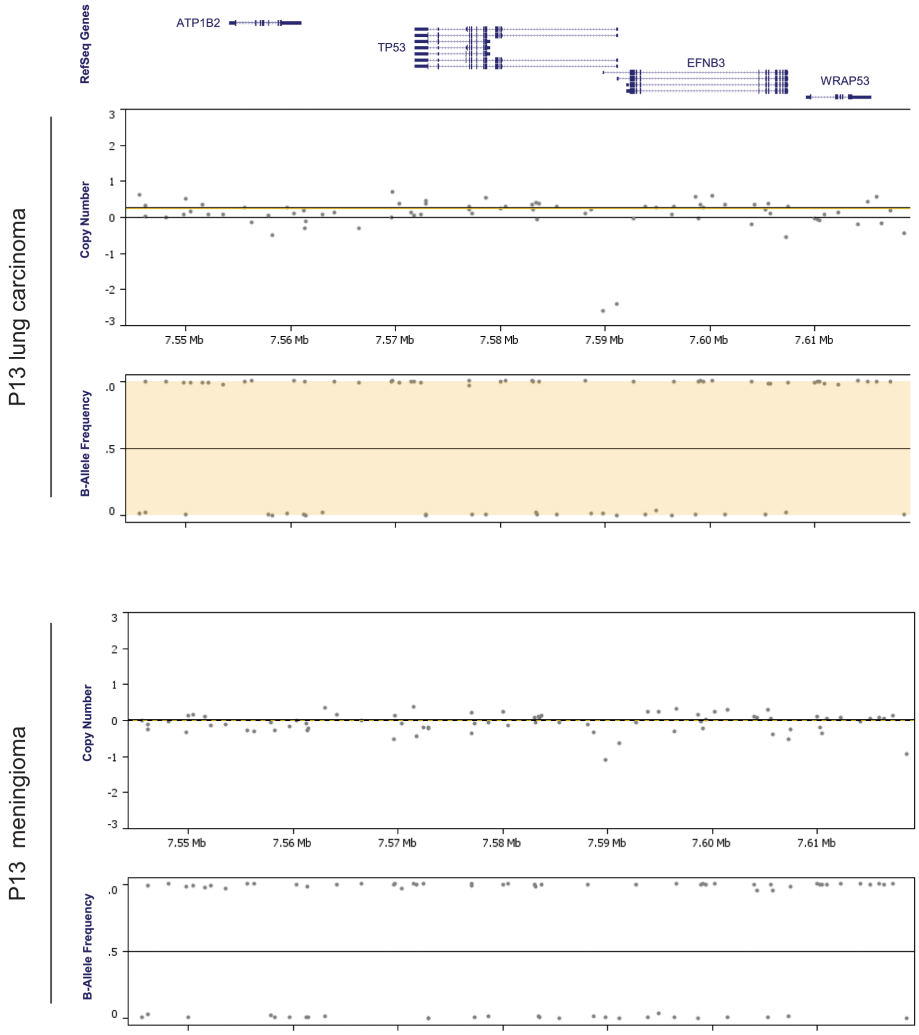


Supplementary Figure S12: Regulatory information of genomic region on chromosome 6 which got inserted into *TP53* in OS patient PZP. UCSC genome browser view of repetitive sequences and ENCODE regulatory information. Break points of the 12.5 kb segment which got inserted into intron 1 of *TP53* are indicated by red arrows. For track description, see legend of Supplementary Figure S7.



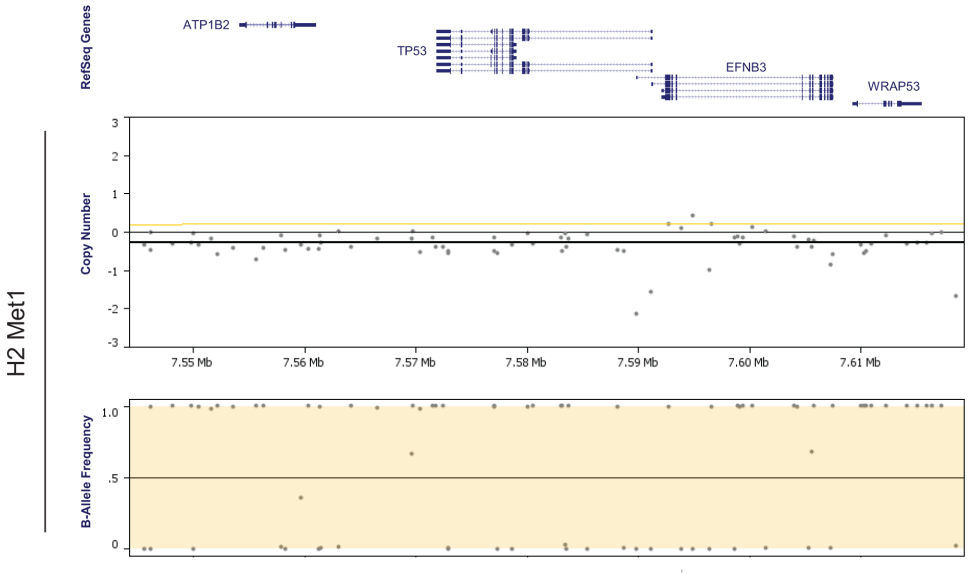
Supplementary Figure S13: Copy number analysis of the *TP53* locus of LFS family members 1 and 13 by CytoScan array. CytoScan array derived copy number analysis of LFS patients 1 and 13 and as copy number loss control gastric cancer cell line SNU16. Gene content of chr17:7,544,275..7,619,275 (NCBI genome build 37) is shown on top, log2 ratios of genomic marker intensities (y-axis) are displayed across the locus (x-axis). SNU16 has a heterozygous deletion of the entire locus (log2 ratio ca. 0.5) and intensities of seven markers of patients 1 and 13 show lower intensities (log2 ratio ca. 0.5, red box) which was not called by Chromosome Analysis Suite version CytoB-N1.2.2.271 (r4615; Affymetrix).

www.impactjournals.com/oncotarget/ Oncotarget, Supplementary Materials 2015



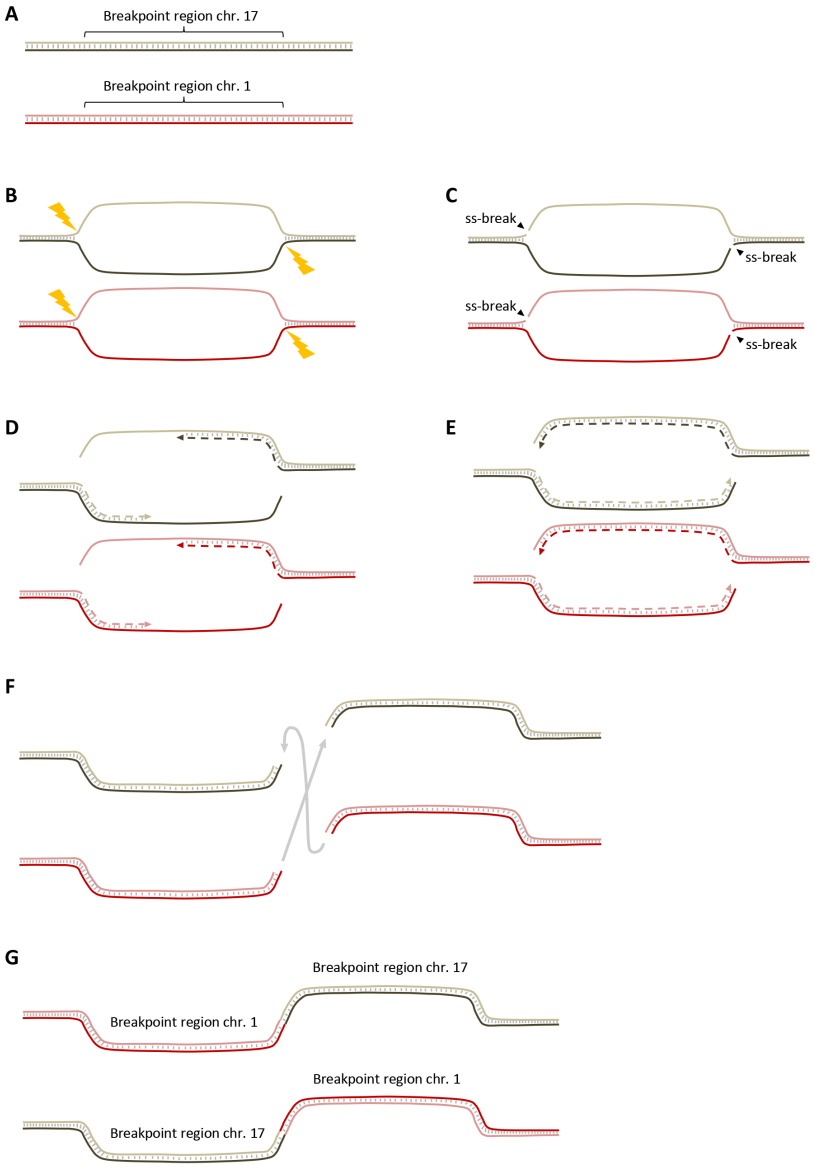
Supplementary Figure S14: Copy number analysis of the *TP53* locus of two tumors of LFS family member P13 by OncoScan array. OncoScan array derived copy number analysis of lung carcinoma and meningioma of LFS patient P13. Gene content of chr17:7,544,000..7,618,000 (NCBI genome build 37) is shown on top, log2 ratios of genomic marker intensities (y-axis) are displayed across the locus (x-axis). Copy number ratio relative to control indicates the presence of two chromosome copies and B allele frequency shows only homozygous variant calls with frequencies close to 1 or 0, respectively, indicating LOH.

www.impactjournals.com/oncotarget/ Oncotarget, Supplementary Materials 2015



Supplementary Figure S15: Copy number analysis of the *TP53* locus of metastasis of LFS family member H2 by OncoScan array. OncoScan array display is as described for Supplementary Figure S14.

www.impactjournals.com/oncotarget/ Oncotarget, Supplementary Materials 2015



Supplementary Figure S16: Hypothetical *TP53* intron 1 rearrangement mechanism in OS. Schematic representation of a hypothetical rearrangement mechanism that can explain the duplication of 46 bp to 555 bp flanking the *TP53* intron 1 rearrangement points. We suggest that at a site of intense transcription, a large 'bubble' of two single DNA strands or an accumulation of more than one, maybe stalled, transcription bubbles occur. DNA breaks on both strands at different positions in the bubble(s) followed by end repair and NHEJ with other chromosomal partner sites. **(A)** to **(G)** chronological order of events. Single DNA strands are represented by light/dark green and red lines, respectively. DNA double strand pairing is represented by short gray lines. After opening of the DNA double strand, two single strand breaks occur that are not complementary to each other, i.e. that are at different sites of the bubble, resulting in a double strand break with about 50 to 550 bases of single strand 5' overhangs. Complementary strands are filled in 5' > 3' by the DNA repair machinery (dashed lines) resulting in a duplication of the single stranded segments. This process can take place at two different loci in the same cell (and same transcriptional hub) and can result in the fusion of different genomic regions by NHEJ **(F)**. **(G)** As a result, both reciprocal fusion products contain the same flanking sequences. ss-break, single strand DNA break. Of note, we observed only in patient YZH that also the non-*TP53* locus (chr. 1) shows duplication of break point-flanking sequences suggesting that the non-*TP53* translocation partner site can have no or short single strand overhangs.

www.impactjournals.com/oncotarget/ Oncotarget, Supplementary Materials 2015

Supplementary Table S1: Sequencing statistics of four OS primary tumors by DNA-PET using the SOLiD platform

Library	Sample	Total beads	Both tags mapped	Span [bp]	Within span	Within span (NR)[1]	Redundancy	Coverage[2]	dPETs[3]	dPET clusters (size 2+)[4]
DHO004HG19	AJF	172,897,363	92,577,911	1510-2570	89,073,088	66,594,042	1.34	46.85	1,383,636	2,278
DHO005HG19	PZP	275,786,428	102,187,549	1910-3060	86,873,155	62,826,184	1.38	53.78	13,893,715	3,811
DHO006DHO008HG19	KRD	286,596,700	168,109,213	660-3980	146,976,367	42,328,532	3.47	23.13	9,695,122	2,733
DHO007DHO009HG19	YZH	378,703,399	167,518,527	1140-4030	147,613,747	60,005,658	2.46	44.98	10,336,682	3,707

[1)]non-redundant paired reads; exclusion of identical PCR amplicons
[2)]physical genome-wide coverage
[3)]discordant mapping paired-end tags
[4)]clusters of dPETs of which their tag one reads and tag 2 reads map to the same regions (regions A and B, respectively), connecting the same two regions

Supplementary Table S2: Statistics of filtering of dPET cluster artifacts before identification of SVs

dPET clusters of size ≥3 before artifact filtering	4737
Filter clusters with superclustersize >40 and UniqueClusterSize <6	218
Blast1 score >2000	265
Blast1 alignment type EC	140
MaxSelfChainLeftCov >80	240
MaxSelfChainRightCov >80	6
Match with simulated library	226
Exclude Overlap	1
dPET clusters of size ≥3 after artifact filtering	3641

For explanation of individual steps see **Materials and Methods** and (Hillmer, Yao et al. 2011; Ng, Hillmer et al. 2012).

Supplementary Table S3: Statistics of germline and somatic SVs identified in four OS tumors

	DHO004HG19	DHO005HG19	DHO006DHO008HG19	DHO007DHO009HG19
	AJF	PZP	KRD	YZH
Deletion	773	603	358	532
Tandem duplication	20	3	11	15
Unpaired inversion	87	72	92	110
Isolated translocation	49	53	129	80
Inversion	33	30	31	39
Insertion	18	30	17	27
Complex rearrangement	68	57	111	193
Total	1,048	848	749	996

www.impactjournals.com/oncotarget/ Oncotarget, Supplementary Materials 2015

Supplementary Table S4: Statistics of somatic SVs identified in four OS tumors

	DHO004HG19	DHO005HG19	DHO006DHO008HG19	DHO007DHO009HG19
	AJF	PZP	KRD	YZH
Deletion	289	177	63	112
Tandem duplication	15	1	7	10
Unpaired inversion	56	46	64	89
Isolated translocation	34	36	119	72
Inversion	3	0	2	4
Insertion	9	6	2	4
Complex rearrangement	28	23	91	129
Total	434	289	348	420

Supplementary Table S5: Statistics of genes affected by somatic SVs of four OS primary tumors

Type of gene alteration	Number of genes
Genes in deletions (<1Mb)	39
Genes in tandem duplications (<1Mb)	2
Genes spanning SVs within one intron	181
Genes spanning SVs affecting coding sequences	31
Fusion genes	67
5' truncated genes	265
3' truncated genes	329

Supplementary Table S6: Somatic SVs of four OS primary tumors (Supplementary Data Set)

www.impactjournals.com/oncotarget/ Oncotarget, Supplementary Materials 2015

Supplementary Table S7: Break point coordinates of *TP53* affecting rearrangements in four OS tumors

Sample	SV ID	Strand left	Validated break left	Strand right	Validated break right	Structural variation	Truncated gene left (strand)	Truncated gene right (strand)	Predicted fusion gene or mode of truncation
AJF	D38	+	chr17:7,557,506	+	chr17:7,651,285	deletion	*ATP1B2*(+)	*DNAH2*(+)	*ATP1B2-DNAH2*
KRD	CT54	+	chr5:137,520,904	+	chr17:7,586,797	complex inter-chr.	*KIF20A*(+)	*TP53*(–)	2 x 3' truncation
KRD	CT7	–	chr6:36,559,119	+	chr17:7,588,049	complex inter-chr.	–	*TP53*(–)	3' truncation
KRD	CT15	+	chr1:172,555,688	–	chr17:7,588,094	complex inter-chr.	*SUCO*(+)	*TP53*(–)	*SUCO-TP53*
PZP	T5	–	chr6:132,198,801[1)]	+	chr17:7,586,884[1)]	insertion inter-chr.	*ENPP1*(+)	*TP53*(–)	*TP53-ENPP1*[2)]
PZP	T12	+	chr6:132,211,834	–	chr17:7,586,948	insertion inter-chr.	*ENPP1*(+)	*TP53*(–)	*ENPP1-TP53*[2)]
YZH	T6	–	chr1:226,260,906	+	chr17:7,587,974	isolated translocation	–	*TP53*(–)	3' truncation
YZH	T13	+	chr1:226,261,198	–	chr17:7,588,529	isolated translocation	–	*TP53*(–)	5' truncation

[1)]DNA-PET mapping coordinates; rearrangement point could not be amplified by PCR
[2)]One fusion transcript based on an insertion of a part of *ENPP1* into *TP53* (see Figure S5)

Supplementary Table S8: Bone-forming tumors or tumor-like lesions other than OS

Diagnosis	*n*	Diagnosis	*n*
Fibrous dysplasia	42/50	Fibrocartilaginous mesenchymoma	4/4
Aneurysmal bone cyst	14/17	Nora lesion	2/2
Osteoid osteoma	14/15	Osteofibrous dysplasia	2/2
Ossifying fibroma	13/15	Osseous dysplasia	2/2
Reactive bone formation	12/13	Desmoplastic fibroma	2/2
Osteoblastoma	8/8	Adamantinoma	1/1
Myositis ossificans	7/9	Non-ossifying fibroma	1/1

n = number of evaluable cases/total number of cases

www.impactjournals.com/oncotarget/ Oncotarget, Supplementary Materials 2015

Supplementary Table S9: Overall survival of TMA OS patients stratified for *TP53* FISH signal and protein expression

Overall survival	5 years	10 years
TP53 FISH		
FISH negative (*n* = 192)	61.20%	58.70%
FISH positive (*n* = 23)	64.90%	64.90%
TP53 immunhistochemistry		
IHC negative (*n* = 170)	59.80%	57.10%
IHC positive (*n* = 42)	59.20%	54.30%

Supplementary Table S10: TMA content and result of *TP53* FISH analysis

Tissue type	*n*
Adrenal gland, adenoma	4/4
Brain, astrocytoma	11/12
Brain, glioblastoma	15/17
Brain, Meningeoma	14/15
Brain, normal	1/1
Brain, oligodendroglioma	5/5
Breast, ductal carcinoma	18/20
Breast, lobular carcinoma	16/18
Breast, medullary carcinoma	21/24
Breast, mucinous carcinoma	7/8
Breast, normal	1/1
Breast, tubular carcinoma	8/9
Cervix, *in situ* carcinoma	10/11
Colon, adenoma	41/46
Colon, carcinoma	21/24
Colon, normal	0/0
Endometrium, endometrioid carcinoma	12/13
Endometrium, normal	1/1
Endometrium, serous carcinoma	11/11
Esophagus, adenocarcinoma	3/3
Esophagus, normal	2/2
Esophagus, small cell carcinoma	1/1
Esophagus, squamous cell carcinoma	7/8
Fat tissue, normal	0/0
Gall bladder, carcinoma	12/13
Gall bladder, normal	5/6
Heart, normal	4/4

(*Continued*)

Tissue type	*n*
Kidney, chromophobe carcinoma	2/3
Kidney, clear cell carcinoma	24/27
Kidney, normal	3/3
Kidney, oncocytoma	8/8
Kidney, papillary carcinoma	11/12
Larynx, carcinoma	13/14
Liver, hepatocellular carcinoma	28(32
Liver, normal	5/5
Lung, adenocarcinoma	54/61
Lung, large cell carcinoma	15/17
Lung, normal	4/5
Lung, small cell carcinoma	15/15
Lung, squamous cell carcinoma	30/32
Lymph node, Hodgkin Lymphoma	15/18
Lymph node, non Hodgkin Lymphoma	18/19
Lymph node, normal	2/2
Mesothelioma	13/14
Myometrium, normal	2/2
Myometrium, myoma	18/22
Nerve, neurofibroma	12/14
Nerve, schwannoma	8/9
Oral cavity, carcinoma	18/21
Oral cavity, normal	5/5
Ovary, endometrioid carcinoma	18/20
Ovary, mucinous carcinoma	3/3
Ovary, normal	0/0
Ovary, serous carcinoma	17/18
Pancreas, adenocarcinoma	14/15
Pancreas, normal tissue	4/6
Paraganglioma	5/6
Parathyroid, adenoma	17/21
Parathyroid, normal	0/0
Pheochromozytoma	10/11
Prostate, adenocarcinoma	33/37
Prostate, normal	9/9
Salivary gland, cylindroma	11/11
Salivary gland, normal	6/6

(*Continued*)

www.impactjournals.com/oncotarget/ Oncotarget, Supplementary Materials 2015

Tissue type	*n*
Salivary gland, pleomorphic adenoma	16/18
Salivary gland, Warthin tumor	7/8
Skeletal muscle, normal	2/3
Skin, basal cell carcinoma	23/27
Skin, histiocytoma	9/10
Skin, Kaposi sarcoma	8/9
Skin, melanoma	32/36
Skin, naevus	22/24
Skin, normal	4/4
Skin, squamous cell carcinoma	10/10
Small intestine carcinoma	8/9
Small intestine, normal	0/0
Soft tissue, giant cell tumor of tendon sheath	10/11
Soft tissue, hemangioma, capillary type	6/6
Soft tissue, leiomyosarcoma	13/15
Soft tissue, lipoma	5/5
Soft tissue, liposarcoma	6/6
Soft tissue, pleomorphic sarcoma	9/10
Stomach, carcinoma, diffuse type	7/8
Stomach, carcinoma, intestinal type	14/17
Stomach, normal	4/5
Testis, non-seminomatous carcinoma	16/20
Testis, normal	3/3
Testis, seminoma	14/14
Thymus, normal	1/1
Thymus, thymoma	8/8
Thyroid, adenoma	11/12
Thyroid, follicular carcinoma	12/13
Thyroid, normal	1/1
Thyroid, papillary carcinoma	5/5
Urinary bladder, invasive carcinoma	10/11
Urinary bladder, non-invasive carcinoma	10/10
Urinary bladder, normal	15/16
Vulva, squamous cell carcinoma	8/8
Total tumor	**966/1072**
Total normal	**84/91**
Total tumor + normal	**1,050/1,163**

n = number of evaluable cases/total number of cases

www.impactjournals.com/oncotarget/ Oncotarget, Supplementary Materials 2015

Supplementary Table S11: Summary of TMA *TP53* FISH analysis

Kind of tissue	*n*
Normal tissue	84/91
Brain tumors	45/49
Breast tumors	70/79
Gynecologic tumors	97/106
Genitourinary tumors	128/142
Others	118/131
Gastrointestinal tumors	208/234
Lung tumors	114/125
Lymphoid tumors	33/37
Skin tumors	104/116
Soft tissue tumors	49/53
Total	**1,050/1,163**

n = number of evaluable cases/total number of cases

Supplementary Table S12: Copy number variations of two affected LFS family members based on CytoScan array analysis

Patient	CN[1] State	Type	Chr	Min[2]	Max[2]	Size (kb)	Marker Count	Confidence	Genes	Comment for CNV shared between patient 1 and 2
1	0	Loss	2	41,239,606	41,250,106	10.5	5	0.9954		
1	1	Loss	2	168,172,736	168,180,626	7.89	11	0.9297		
13	1	Loss	2	168,172,736	168,180,626	7.89	11	0.9623		
1	1	Loss	3	160,772,598	160,778,822	6.224	9	0.9388	PPM1L	
13	3	Gain	4	132,780,784	132,894,269	113.485	45	0.8855		
1	3	Gain	5	32,108,414	32,162,756	54.342	53	0.9247	PDZD2, GOLPH3	CNV site in DGV
13	3	Gain	5	32,108,414	32,166,970	58.556	55	0.9186	PDZD2, GOLPH3	CNV site in DGV
13	1	Loss	7	3,833,336	3,852,657	19.321	14	0.9512	SDK1	
13	0	Loss	8	39,234,303	39,357,501	123.198	8	1.0000	ADAM5P, ADAM3A	
13	1	Loss	8	84,367,320	84,380,551	13.231	11	0.9519		
1	1	Loss	8	96,076,737	96,098,249	21.512	14	0.9106	MIR3150	CNV site in DGV
13	1	Loss	8	96,076,737	96,098,249	21.512	14	0.9538	MIR3150	CNV site in DGV
13	3	Gain	12	32,005,944	32,062,001	56.057	40	0.9289		
1	1	Loss	12	69,023,587	69,032,212	8.625	9	0.9250	RAP1B	
13	3	Gain	14	106,022,513	107,051,759	1029.25	23	0.9158	KIAA0125, ADAM6, NCRNA00226, NCRNA00221	

(*Continued*)

www.impactjournals.com/oncotarget/ Oncotarget, Supplementary Materials 2015

Patient	CN[1] State	Type	Chr	Min[2]	Max[2]	Size (kb)	Marker Count	Confidence	Genes	Comment for CNV shared between patient 1 and 2
13	1	Loss	15	20,175,623	22,504,198	2328.58	48	0.8801	HERC2P3, GOLGA6L6, GOLGA8C, BCL8, POTEB, NF1P1, LOC646214, CXADRP2, LOC727924, OR4M2, OR4N4, OR4N3P	CNV site in DGV
1	1	Loss	15	22,300,190	22,504,198	204.008	38	0.8682	LOC727924, OR4M2, OR4N4, OR4N3P	CNV site in DGV
13	3	Gain	16	34,455,753	34,762,298	306.545	84	0.9135	LOC283914, LOC146481, LOC100130700	
13	1	Loss	17	28,303,998	28,317,131	13.133	18	0.9198	EFCAB5	
1	1	Loss	18	1,722,426	1,838,901	116.475	110	0.9158		
13	3	Gain	18	51,345,688	51,423,648	77.96	57	0.9415		
13	3	Gain	19	20,833,215	20,987,762	154.547	123	0.9063	ZNF626	
13	1	Loss	19	28,294,543	28,318,358	23.815	15	0.9048		
13	3	Gain	19	58,364,597	58,383,708	19.111	20	0.9187	ZNF587, ZNF814	
1	1	Loss	20	32,806,979	32,822,219	15.24	9	0.9050		
13	3	Gain	22	22,861,993	23,358,013	496.02	65	0.9360	ZNF280B, ZNF280A, PRAME, LOC648691, POM121L1P, GGTLC2, MIR650, IGLL5	
1	2	Gain	X	2,706,943	2,713,139	6.196	2	0.2191	XG	
1	2	Gain	X	58,094,436	62,058,620	3964.18	19	0.2466		
1	4	Gain	X	150,192,436	150,195,253	2.817	8	0.9315		

[1]Copy number
[2]Position based on NCBI human reference Build 37

REFERENCES

1. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Research. 2009; 19:1639–1645.
2. Rhead B, Karolchik D, Kuhn RM, Hinrichs AS, Zweig AS, Fujita P a, Diekhans M, Smith KE, Rosenbloom KR, Raney BJ, Pohl A, Pheasant M, Meyer LR, et al. The UCSC Genome Browser database: update 2010. Nucleic Acids Research. 2010; 38:D613–619.
3. Djebali S, Davis C, Merkel A, Dobin A. Landscape of transcription in human cells. Nature. 2012; 489:101–108.

2.3 Publication No 2

Exome sequencing of osteosarcoma reveals mutation signatures reminiscent of BRCA deficiency.

Kovac M*, Blattmann C*, **Ribi S***, Smida J, Mueller NS, Engert F, Castro-Giner F, Weischenfeldt J, Kovacova M, Krieg A, Andreou D, Tunn PU, Dürr HR, Rechl H, Schaser KD, Melcher I, Burdach S, Kulozik A, Specht K, Heinimann K, Fulda S, Bielack S, Jundt G, Tomlinson I, Korbel JO, Nathrath M, Baumhoer D.
***These authors contributed equally to this work**

Nat Commun. 2015 Dec 3;6:8940. doi: 10.1038/ncomms9940.

nature COMMUNICATIONS

ARTICLE

Received 30 Apr 2015 | Accepted 19 Oct 2015 | Published 3 Dec 2015

DOI: 10.1038/ncomms9940 OPEN

Exome sequencing of osteosarcoma reveals mutation signatures reminiscent of BRCA deficiency

Michal Kovac[1,2,*], Claudia Blattmann[3,4,*], Sebastian Ribi[1,*], Jan Smida[5,6], Nikola S. Mueller[7], Florian Engert[8,9,10], Francesc Castro-Giner[2], Joachim Weischenfeldt[11], Monika Kovacova[12], Andreas Krieg[13], Dimosthenis Andreou[14,†], Per-Ulf Tunn[14], Hans Roland Dürr[15], Hans Rechl[16], Klaus-Dieter Schaser[17], Ingo Melcher[18], Stefan Burdach[6], Andreas Kulozik[4], Katja Specht[19], Karl Heinimann[20], Simone Fulda[8,9,10], Stefan Bielack[3], Gernot Jundt[1], Ian Tomlinson[2], Jan O. Korbel[11], Michaela Nathrath[5,6,21,**] & Daniel Baumhoer[1,**]

Osteosarcomas are aggressive bone tumours with a high degree of genetic heterogeneity, which has historically complicated driver gene discovery. Here we sequence exomes of 31 tumours and decipher their evolutionary landscape by inferring clonality of the individual mutation events. Exome findings are interpreted in the context of mutation and SNP array data from a replication set of 92 tumours. We identify 14 genes as the main drivers, of which some were formerly unknown in the context of osteosarcoma. None of the drivers is clearly responsible for the majority of tumours and even *TP53* mutations are frequently mapped into subclones. However, >80% of osteosarcomas exhibit a specific combination of single-base substitutions, LOH, or large-scale genome instability signatures characteristic of BRCA1/2-deficient tumours. Our findings imply that multiple oncogenic pathways drive chromosomal instability during osteosarcoma evolution and result in the acquisition of BRCA-like traits, which could be therapeutically exploited.

[1] Bone Tumour Reference Center at the Institute of Pathology, University Hospital Basel, Schönbeinstrasse 40, 4031 Basel, Switzerland. [2] The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK. [3] Pediatrics 5 (Oncology, Hematology, Immunology), Klinikum Stuttgart Olgahospital, Kriegsbergstrasse 62, 70174 Stuttgart, Germany. [4] Department of Pediatric Hematology, Oncology, Immunology and Pulmology, University of Heidelberg, Im Neuenheimer Feld 430, 69120 Heidelberg, Germany. [5] Institute of Radiation Biology, Clinical Cooperation Group Osteosarcoma, Helmholtz Zentrum München, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany. [6] Pediatric Oncology Center, Department of Pediatrics, Technische Universität München and Comprehensive Cancer Center, Kölner Platz 1, 80804 Munich, Germany. [7] Institute of Computational Biology, Helmholtz Zentrum München, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany. [8] Institute for Experimental Cancer Research in Pediatrics, Goethe-University, Komturstrasse 3a, 60528 Frankfurt, Germany. [9] German Cancer Consortium (DKTK), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. [10] German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. [11] European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Meyerhofstrasse 1, 69117 Heidelberg, Germany. [12] The Institute of Mathematics and Physics, Faculty of Mechanical Engineering, Slovak University of Technology, 84248 Bratislava, Slovak Republic. [13] Orthopaedic Department, Basel University Childrens Hospital (UKBB), Spitalstrasse 33, 4056 Basel, Switzerland. [14] Department of Orthopedic Oncology, Sarcoma Center Berlin-Brandenburg, HELIOS Klinikum Berlin-Buch, Schwanebecker Chaussee 50, 13125 Berlin, Germany. [15] Department of Orthopedic Surgery, Ludwig-Maximilians-University Munich, Campus Grosshadern, Marchioninistrasse 15, 81377 Munich, Germany. [16] Clinic and Policlinic of Orthopedics and Sports Orthopedics, Technische Universität München, Ismaninger Strasse 22, 81675 Munich, Germany. [17] Department of Orthopaedics and Trauma Surgery, University Hospital Dresden, Fetscherstrasse 74, 01307 Dresden, Germany. [18] Center for Musculoskeletal Surgery, Charité—University Medicine Berlin, Campus Virchow Klinikum, Augustenburger Platz 1, 13353 Berlin, Germany. [19] Institute of Pathology, Technische Universität München, Trogerstrasse 18, 81675 Munich, Germany. [20] Medical Genetics, University Hospital Basel, Burgfelderstrasse 101, 4055 Basel, Switzerland. [21] Department of Pediatric Oncology, Klinikum Kassel, Mönchebergstrasse 41-43, 34125 Kassel, Germany. * These authors contributed equally to this work. ** These authors jointly supervised this work. † Present addresses: Department of General Orthopedics and Tumour Orthopedics, Münster University Hospital, Albert-Schweitzer-Campus 1, 48149 Münster, Germany (D.A.). Correspondence and requests for materials should be addressed to D.B. (email: dbaumhoer@mac.com).

Osteosarcomas (OS) are primary malignant tumours of bone with complex karyotypes showing abundant structural and numerical aberrations. Rapid tumour progression and early metastatic spread are the rational for multimodal treatment approaches that can achieve long-term survival in about 60% of patients. Effective treatment options are still lacking for the remaining 40% of patients suffering from refractory or recurrent disease, however[1]. A partial explanation for treatment failure might lie in different aetiologies responsible for the structural aberrations marking the onset of the disease and resulting in a variety of mutations in genes and pathways of which only few are targetable. OS arise owing to mutations in the *TP53* tumour suppressor gene[2–4] and a plethora of other cancer drivers, as for example *RB1* (refs 2–4), *ATRX*[2,3], *DLG23* (ref. 2), *RUNX2* (ref. 5), *WRN*[6,7], *RECQL4* (refs 5,7,8), *CDKN2A/B*[9], *BLM*[7], *PTEN*[3] and other PI3K/Akt/mTOR pathway members' each of which contribute only to a small proportion of patients. As a consequence, widely applicable therapeutic targets have not been identified so far.

The introduction of high-throughput sequencing of cancer genomes has expanded a list of potential OS driver genes, but disappointingly failed to provide important leads that could improve patients's care. In fact, data published so far only confirmed that the intricate nature of these tumours develops through cancer heterogeneity promoting the accumulation of chromosomal aberrations, kataegis and chromothripsis[2–4]. Marked inter- and intratumoural heterogeneity seems to be linked to the observed differences in treatment efficacy and clinical outcome of the respective patients[10], but the particular sets of mutations underlying these patterns, the clonal frequencies of driver mutations and how different (sub-)populations of tumour cells relate functionally to each other, remains largely unknown. In this study, we search for driver gene mutations in 123 OS and attempt to reduce the complexity of their genomes by applying abstraction analyses supplemented by clonal mapping and ordering. Finding common traits of different cell lineages within and between tumours seems to constitute a first step towards developing more individualized and effective treatment strategies in the future.

Results

Mutation processes moulding the exomes of 31 OS. Our discovery set for whole-exome sequencing comprised 31 OS, which were sequenced at a median depth of ~150× (range 131–370) alongside with paired constitutional DNA from peripheral blood (Table 1). All tumour samples were derived from pre-therapeutic biopsies. Making use of Stampy and Platypus programs for mapping and variant calling, the number of somatic base substitutions (single-nucleotide variants (SNVs)) and indels called with high confidence ranged from 7 to 3,153 (median = 66.5) and 17 to 529 (median = 33.5) per exome, respectively. Of these, a median of 21 SNV changes (range 4–174) and seven small indels (range 2–210) were potentially functional within protein-coding regions. The SNV spectrum for each OS exome is shown in Fig. 1a and Supplementary Figs 1 and 2. With two exceptions, the C:G > T:A changes were the most common changes followed by C:G > A:T and T:A > C:G changes. The mutation burden of OS was similar to the ageing signature of most tumour types from the pan-cancer analysis[11], however, a decomposition of mutation spectra with a non-negative matrix factorization algorithm revealed a pattern similar to the signatures 3 and 5 (Fig. 1b and Supplementary Fig. 3). A subset of tumours of this class is known to acquire a characteristic pattern of substitution mutations including kataegis and the presence of signature 3 was strongly associated with *BRCA1* and *BRCA2* mutations within breast and pancreatic cancer types[11,12].

Driver mutations impair cell control of genome integrity. For the discovery of somatic driver mutations we restricted our analysis to protein-coding regions. A minimum variant allele frequency of 0.05 was set. We then prioritized genes for further investigation by filtering mutations to exclude all SNVs with moderate or benign predicted functional effects (Sorting Intolerant From Tolerant (SIFT) score = < 0.05 and Mutation Taster and PolyPhen2 > 0.8). All protein-truncating and splice-site mutations were retained. We then identified genes that were mutated in two or more tumours and inspected all variant reads in the Integrated Genome Viewer to exclude any mutations at sites of evidently poor quality. Eventually, 26 genes (Supplementary Data 1) including some that had already been reported in the context of OS (for example, *ARTX*, *SFPQ*, *FGFRL1* and *RB1*) remained[2–4,13].

Since patients with Werner, Li–Fraumeni, Rothmund–Thomson and Bloom syndromes have a comparably higher risk of OS[8,14,15], we examined our sequencing data for germline mutations in the *WRN*, *TP53*, *RECQL4* and *BLM* genes. We identified two germline *TP53* mutations (OS-241 and OS-228), one germline *WRN* mutation in addition to loss of heterozygosity (LOH) around the locus (OS-230) and two rare germline *RECQL4* variants with unknown significance (OS-227 and OS-238). We then applied the same prioritization scheme to the remainder of genes with germline variants. Prioritized cancer drivers not previously reported in OS included *RET*, *MUTYH*, *NUMA1*, *FANCA*, *BRCA2* and *ATM*. We chose these genes for further investigation, based on the presence of at least two clearly pathogenic mutations and/or co-segregation of mutations with cancer in the respective families. Furthermore, these genes were excellent functional candidates.

We validated mutations detected in 388 genes (Supplementary Data 2) by Ion Torrent sequencing in the discovery set of tumours and then undertook replication testing of 30 genes (Supplementary Data 3) in a set of additional 92 unpaired OS. The analysis of merged variant calls from 123 tumours yielded a total of 22, 10 and 11 *TP53*, *RB1* and *ATRX* mutations, respectively (Supplementary Table 1). Three tumours acquired somatic *RET* mutations. One patient, OS-250, carried a germline *RET* mutation, which at the time of diagnosis had not yet manifested by multiple endocrine neoplasia type 2 but the mutation co-segregated with breast cancer and rhabdomyosarcoma in two first-degree relatives. Two *de novo* germline *RET* mutations previously associated with late-onset multiple endocrine neoplasia type 2 were identified in patients OS-224 and OS-242. We also identified seven *MUTYH* mutations (germline or somatic) and four, four, three and one mutations in the *FANCA*, *MDC1*, *NUMA1* and *PTEN* genes, respectively. Somatic missense mutations in *MDC1* and *NUMA1* genes affected conserved residues of protein domains encoding nuclear localization signal and/or interacting with other proteins of BRCA complex, for example CHEK2. *MUTYH* mutations were equally distributed between an endo-nuclease domain and a NUDIX-type hydrolase domain. There were also two somatic *WRN* mutations affecting a DNA-binding domain and three germline mutations in the *ATM* gene, two of which have been linked to breast cancer susceptibility before[16,17].

The Intogen[18] pathway analysis reassuringly identified *ATRX* and *RB1* as the main drivers (Supplementary Data 4). The Intogen list of drivers did not include *TP53* since there was only one mutation in the discovery set of tumours, but *TP53* and three other genes (*RB1*, *ATRX* and *ATM*) were notably reported by Intogen as drivers in the pan-cancer analysis[19]. Exploring gene ontology of the remaining genes we found that many, if not all genes, have been functionally related to DNA damage repair,

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms9940 ARTICLE

Table 1 | Clinicopathological and summary mutation data for each OS.

ID	Age	Histology	Site	SNV	Indels	Dn:Ds	Ts:Tv	%SCNA	Clonality
OS-046	10	Mixed	Femur	91	24	2.8	1.76	67.45	3
OS-059	13	Osteoblastic	Fibula	50	29	1.18	1.38	67.1	4
OS-061	12	Chondroblastic	Femur	106	36	1.8	1.65	67.5	7
OS-063	17	Teleangiectatic	Humerus	127	48	1.13	1.75	64.3	6
OS-065	15	Osteoblastic	Tibia	78	34	1.75	1.43	44.2	3
OS-079	45	Osteoblastic	Femur	48	20	2.35	1.17	56.2	2
OS-224	11	Osteoblastic	Tibia	310	397	1.62	1.64	57.1	14
OS-225	14	Osteoblastic	Tibia	38	25	1	1.53	50.8	2
OS-226	11	Osteoblastic	Tibia	52	36	1.77	1.36	57.8	6
OS-227	19	Osteoblastic	Femur	29	32	3	1.41	0.2	3
OS-228	13	Chondroblastic	Femur	48	24	2	1.28	59.3	4
OS-230	18	Chondroblastic	Tibia	192	185	1.5	1.23	62.5	11
OS-231	13	Chondroblastic	Tibia	195	246	1.39	1.18	81.8	8
OS-232	21	Extraosseous	Femur	606	529	2.59	1.18	61.1	16
OS-234	17	Osteoblastic	Humerus	46	20	3.8	1.29	60.5	4
OS-235	14	Osteoblastic	Tibia	52	23	1.7	0.79	76.2	5
OS-236	19	Chondroblastic	Femur	46	37	2.6	1.29	59.4	5
OS-237	14	Chondroblastic	Femur	29	28	1.33	1.89	44.9	5
OS-238	17	Osteoblastic	Tibia	63	33	2	2.49	38.3	6
OS-240	15	Osteoblastic	Femur	32	19	3.75	1.66	15.6	6
OS-241	5	Osteoblastic	Fibula	7	19	-	0.4	NA	2
OS-242	14	Osteoblastic	Tibia	35	17	7.33	0.94	26.4	4
OS-250	14	Osteoblastic	Femur	70	35	1.25	0.79	20.5	4
OS-251	9	Osteoblastic	Femur	106	38	1.64	0.89	0.2	9
OS-252	16	Chondroblastic	Femur	253	33	0.82	2.37	34.7	4
OS-253	6	Chondroblastic	Tibia	71	47	3.33	0.82	56.7	3
OS-253M	6	Chondroblastic	Lung	167	56	2.15	0.77	70.1	12
OS-254	27	Mixed	Fibula	33	20	1.25	2.14	30.1	1
OS-255	12	Chondroblastic	Tibia	3,153	459	2.56	2.01	33.9	7
OS-256	17	Osteoblastic	Tibia	50	44	2	1.78	55.6	4
OS-257	12	Mixed	Tibia	1,993	345	1.35	1.92	17.6	4
OS-258	20	Unknown	Humerus	104	29	2.22	1.66	29.4	7

NA, not applicable
SNV denotes high-quality somatic single-nucleotide variants; Dn:Ds denotes the ratio of non-synonymous to synonymous exonic mutations; Ts:Tv denotes a ratio of somatic exonic transitions to transversions; % genome changed by SCNA, per cent of genome affected by copy-number changes; clonality denotes a number of clones estimated by PyClone algorithm.

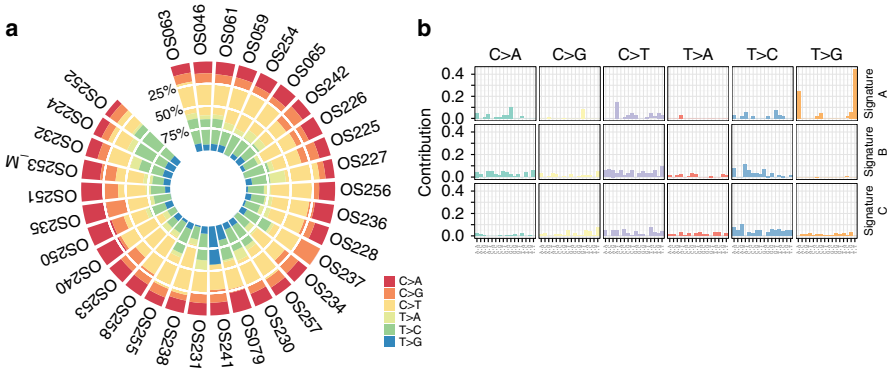


Figure 1 | Somatic SNV spectra and mutation signatures. (**a**,**b**) Data are derived from exomes of each tumour. Note that OS with very few somatic SNVs are included for sake of completeness.

chromosomal segregation and cell cycle control of genome integrity[8,20–24].

Structural and copy-number alterations. We next turned our attention to somatic copy-number alterations (SCNA; Fig. 2a,b), which affected 0.2 (OS-251) to 87% (OS-038) of OS genomes. The median size of a SCNA called with high-sensitivity settings was 4.7 Mb and a typical genome contained 69 of such events. Analysis of single-nucleotide polymorphism arrays using dedicated data-processing methods identified large-scale genomic instability[25] and LOH[26,27] signatures similar to that of breast and ovarian cancers with BRCA1/2 inactivation in 91% (112/123; Fig. 2c) and 78% (96/123; Fig. 2d) of OS, respectively.

The most frequent SCNA events involved gains of the long arm of chromosome 8 (75%). It has been plausibly suggested that gains of 8q target *MYC* in OS, but chromosome 8 gains in our set of tumours involved very long segments without evidence of targeting any gene specifically. Similarly, other frequent gains of 1p (55%), 1q (53%), 5p (46%), 6p (56%), 17p (66%) and 18p (33%) involved large regions. Large deletions were almost as

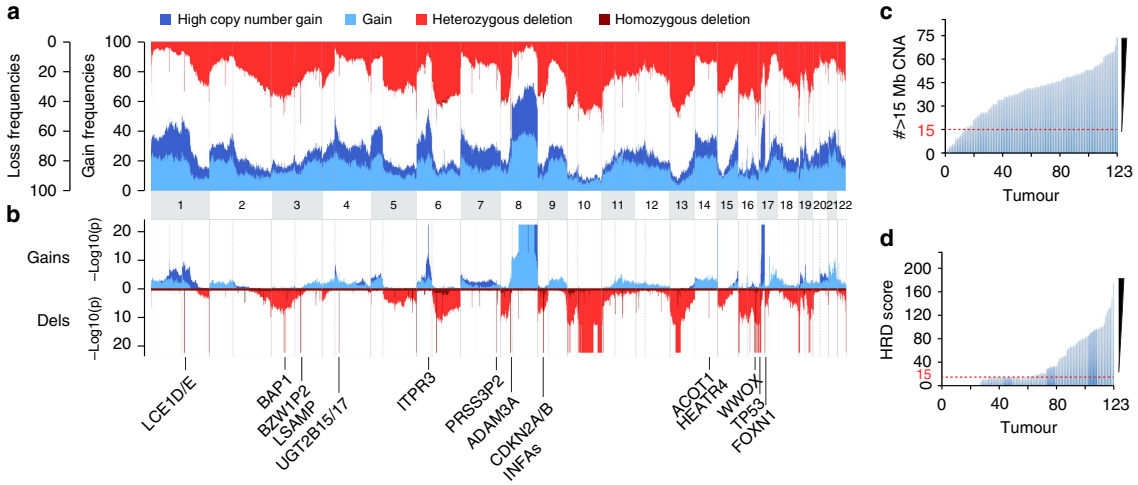


Figure 2 | Copy-number variation analysis. (**a**) Copy-number profile of 123 OS genomes. (**b**) Significantly over-represented copy-number alterations. The most important over-represented SCNA loci are highlighted. Blue, gains; red, losses. (**c**) The number of copy-number alterations fulfilling specific criteria of a BRCA-like phenotype shown for each tumour (threshold line in red). (**d**) Homologous recombination deficiency (HRD) score shown for each tumour. The threshold for considering a tumour to be BRCA deficient was set to 15 (red dashed line).

common as gains, the most frequent involving chromosomes 3 (50%), 6q (45%), 5q (40%), 8p (43%), 10 (56%), 13 (50%), 16 (64%), 17 (47%), 18q (34%) and 19 (54%). Deletions of these chromosomes almost invariably included main OS drivers including *TP53*, *BRCA2*, *RET*, *WRN*, *FANCA* and *RB1* (Fisher exact test, $P<0.01$; Supplementary Fig. 4). We also found good evidence supporting SCNA mutations in the *BRCA1* gene itself (26%) and in members of the homologous recombination repair pathway—*BAP1* (38%), *PTEN* (50%) and *PALB2* (43%)—in which exome sequencing did not identify any mutation.

We specifically searched for small (<1 Mb) and focal SNCAs that might represent oncogene amplifications or tumour suppressor deletions. After filtering out common variants, 20,758 regions were identified although only 80 were found to be recurrent (defined as having frequencies >15%; Supplementary Fig. 5 and Supplementary Data 5). Two focal SCNA involved gene loci with a strong *a priori* importance in OS[2,9], including a deletion of *CDKN2A/2B* located at chromosome 9p21 (15%) and a deletion of *DLG2* at chromosome 11q14 (24%). Other SCNA were detected in known fragile sites (WWOX, for example), deeply intragenic regions as well as microRNAs and genes with no prior association to cancer.

We then assessed the significance of called SCNA regions by using a random sampling model similar to the GISTIC analysis (Supplementary Figs 6 and 7). After filtering for SCNAs with adjusted *P* values <0.01, 92, 412 and 322 amplifications (copy number >4), gains (copy number = 3) and deletions (copy number <2) remained (Supplementary Data 6), respectively. *TP53*, *RB1*, *PTEN*, *RUNX2*, *BAP1* and *CDKN2A/B* genes were reassuringly among the over-represented SCNA regions, as well as several new loci including a homozygous loss of interferon genes on chromosome 9 (chr9: 21,248,029–21,408,261) and an amplification of a 10-gene locus on chromosome 1 (chr1: 59,762,489–62,907,413). We then re-examined each candidate SCNA events for association with overall and/or disease-free survival using the Cox regression model. Five loci (including *CDKN2A/B* and *TP53*) were associated with an adverse outcome of patients (Supplementary Fig. 8). Three SCNA loci were associated with earlier onset of disease (Supplementary Fig. 9).

BRCAness as an emergent property of OS. Our hypothesis of 'BRCAness' does not replace the existing view of OS as a monoclonal expansion of one initial *TP53* mutant cell, but rather offers an explanation of how the vulnerability to chromosomal breakage may be sustained parallel to or in the absence of *TP53* mutations. Using PyClone[28] estimates of cellular frequencies we found that *TP53* and *RB1* driver mutations were clonal events in 58 tumours (47%; Fig. 3). The remaining tumours ($n=65$, 53%) showed good evidence of the acquisition of genomic instability through alternative pathways—for example, by *MUTYH* mutations[20]—well before *TP53*, *RB1* mutations and/or BRCA-like traits emerged. We envision that OS at this stage also acquired a propensity to shatter into subpopulations of cancer cells. Given enough time, different cancer cell lineages are likely to acquire private SCNAs in the *BRCA1/2* genes (and in their 67 binding partners) simply by chance, and by the time the disease emerges, different clonal populations will have already gained partial or full deficiency in homologous recombination repair. We find this view consistent with evidence that, on average, a typical OS carries 17 SCNA mutations in *BRCA* genes and their core binding partners (Fig. 4a and Supplementary Data 7), mutations in different 'BRCA' genes can be functionally equivalent (for example, *PALB2*, *CHEK2*, *PTEN* and *ATM* mutations result in chromosomal instability analogous to BRCA1/2 mutations[29–33]) and OS are polyclonal (Table 1).

We therefore estimated the relative importance of each of the 69 'BRCA' genes from the clonality of their SCNA mutations. For Fig. 4b we used a clustering method based on Euclidian distances; the individual position of each gene in the plot is the product of clonal and subclonal frequencies of its SCNA events. It is not surprising that the most frequently mutated and thus outlying genes were *TP53*, *RB1*, *BRCA2* and *PTEN*, and that *TP53* and *RB1* became even more outlying when SNV changes were included. In fact, *TP53* and *RB1* were the only two genes mutated with somatic SNV/indel mutations at a frequency >3%. Putative cancer drivers with a propensity to carry clonal SCNAs were an android receptor gene, the RNA gene *XIST*, the BRCA1 binding partners *RBBP7* and *NCOA2*, a BRCC3 metaloprotease unit of the BRISC complex and CSTF2 that prevents inappropriate RNA processing at sites of DNA repair.

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms9940 ARTICLE

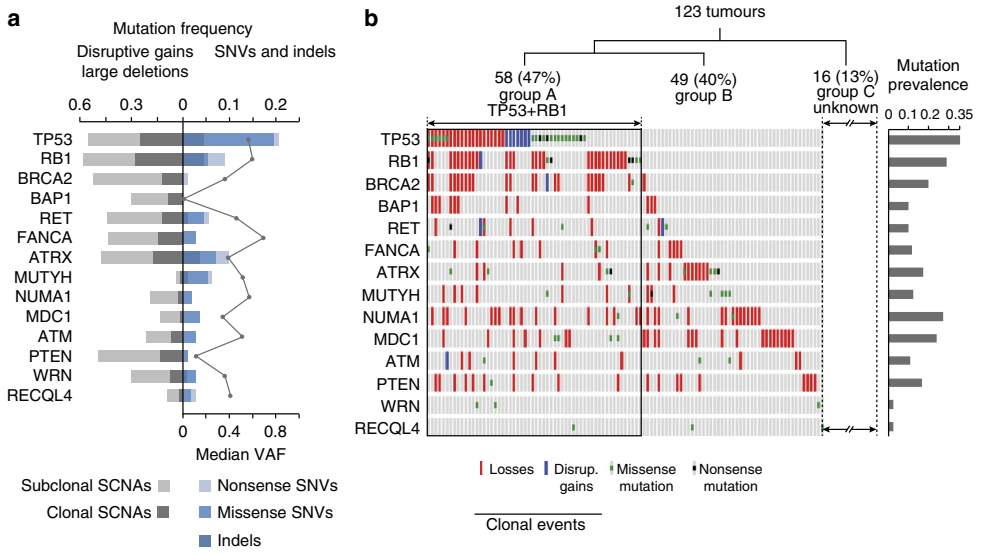


Figure 3 | Mutations burden of OS driver genes. (**a**) Mutation profiles of selected OS driver genes, split by copy-number alterations and single-nucleotide changes/indels. The median variant allelic frequency (VAF) is plotted for each gene. (**b**) Distribution of selected somatic SNVs with predicted pathogenic effects and indels across cancers. Note that only clonal copy-number losses and disruptive gains are shown here. Blue, disruptive gains (that is, one breakpoint within a gene); red, losses.

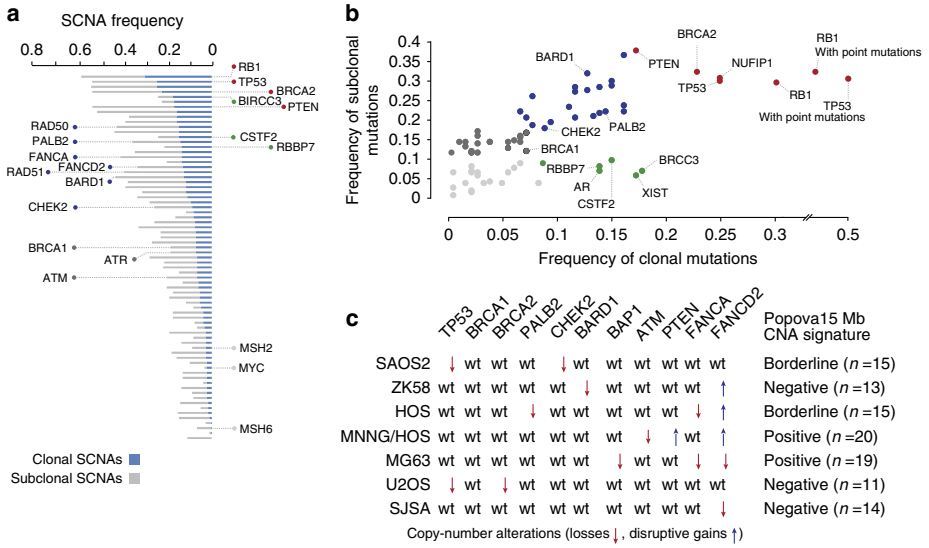


Figure 4 | Copy-number alterations in BRCA1/2 and binding partners. (**a**,**b**) Frequencies of clonal and subclonal SCNA mutations. (**c**) SCNA mutational profiles of selected OS cell lines.

Efficacy of talazoparib inhibition in OS cell lines. Currently there is no available OS cell line, which is fully deficient in *BRCA1* or *BRCA2*. Although this may constitute a bottleneck for current experimental approaches, some OS cell lines can carry mutations in other genes of the homologous recombination pathway leading to defects functionally analogous to *BRCA1* and *BRCA2* mutations (Fig. 4c). We tested three cell lines, one being a double mutant in *PTEN* and *ATM* (MNNG/HOS) and two with heterozygous SCNA mutations in the checkpoint kinase 2 *CHEK2* (SAOS2) and *FANCD2* gene (SJSA-1), respectively. SJSA-1 and SAOS2 cells were included in the study mainly for comparative purposes since the heterozygous mutations did not seem sufficient to result in homologous recombination repair deficiency. Indeed, *in vitro* tests showed only a limited response of SAOS2 and SJSA cells to a standalone 72-h treatment with the phase-3 poly ADP ribose polymerase (PARP) inhibitor talazoparib (Fig. 5). By contrast, MNNG/HOS cells showed good evidence of IC_{50} after standalone treatment, and even a better response when talazoparib was combined with the alkylating agent temozolomide and the topoisomerase I inhibitor SN-38. The reason for pursuing combined treatment strategies was that the cytotoxicity of talazoparib results from the availability of

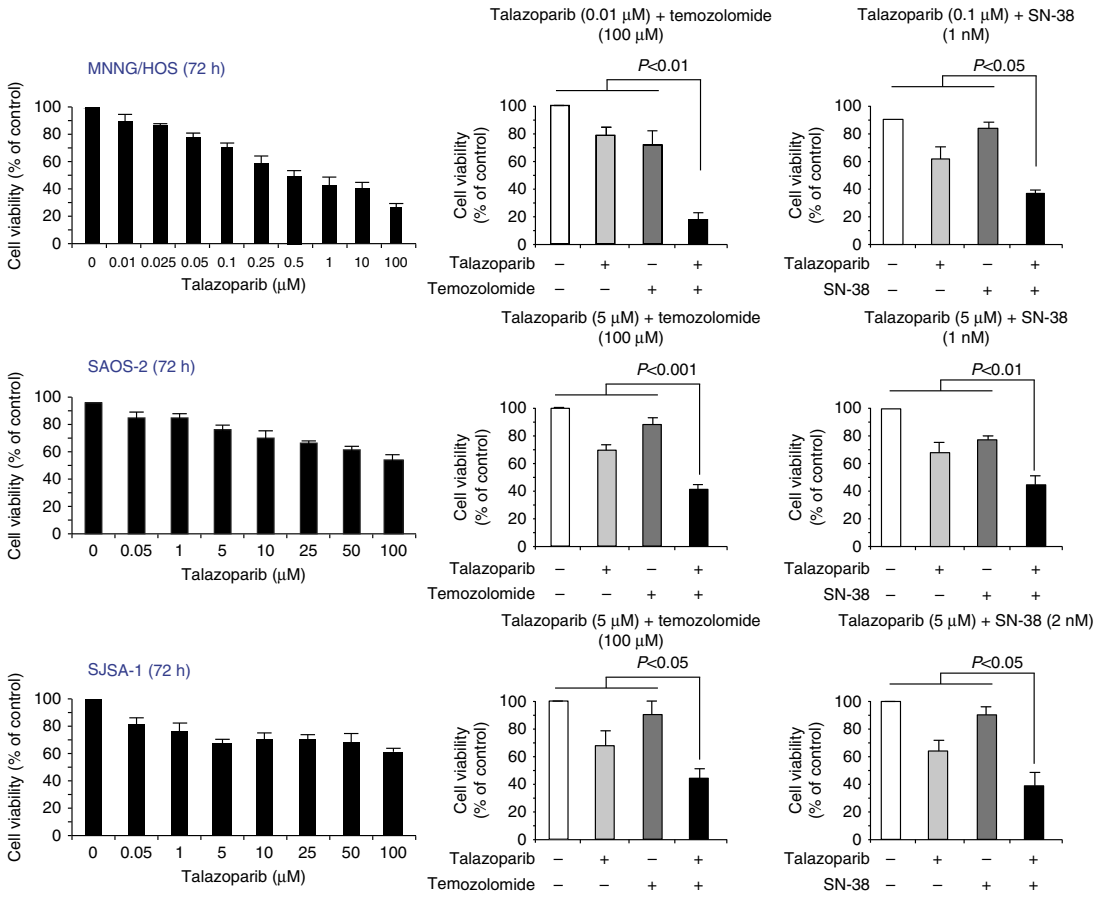


Figure 5 | Cell viability assays. MNNG, SAOS2 and SJSA-1 OS cells were treated for 72 h with indicated talazoparib, temozolomide and SN-38 concentrations. Cell viability was assessed by a MTT assay and is expressed as percentage of untreated cells. Data are shown as mean ± s.d. of three independent experiments, each performed in triplicates. MTT, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide.

single-stranded DNA breaks, which temozolomide and SN-38 are particularly effective in creating[34]. Talazoparib concentrations of up to 5 μM used in our experiments have been considered therapeutically achievable concentrations *in vivo* according to a recent study of Hopkins *et al.*[35].

Discussion

Cataloguing mutations from 123 tumours has yielded several new insights into the underlying mutational mechanisms in OS. We identified clearly monoclonal mutation events in *TP53* or *RB1* in 47% of cases, whereas additional 40% of tumours could be explained by invoking mutations in alternative driver genes. Eight of these genes—*BRCA2*, *BAP1*, *RET*, *MUTYH*, *ATM*, *PTEN*, *WRN* and *RECQL4*—are well-known Mendelian cancer drivers, whilst the four remaining genes (*ATRX*, *FANCA*, *NUMA1* and *MDC1*) have been reported in the context of cancer susceptibility[2,22–24]. In total, 87% of OS could be explained by any of the 14 driver genes. Our findings thus supplement the 'traditional' *TP53*-centred model of OS evolution such that molecular pathways functionally similar to *TP53* can drive its initial phases by conferring to genome instability. Tumours without clonal *TP53* mutations may still acquire subclonal *TP53* mutations later which adds ~25% to the total mutation prevalence. This is important to realize when interpreting findings of other studies as, for example, Chen *et al.*[2] reported structural and SNV mutations in *TP53* in 30 of 34 genome-sequenced tumours. In a second study, whole-exome sequencing of 53 paediatric OS yielded a 75% prevalence of *TP53* mutations[3]. The difference between these and our estimate might in part result from our inability to identify rearrangements within the first intron of *TP53* (which we estimated at 15% frequency[36]), but other aspects being equal, we conclude that not all previously reported *TP53* mutations are clonal and thus represent the triggering event. Instead we envision a situation, in which multiple oncogenic pathways drive chromosomal aneuploidy and instability in early stages of OS evolution.

Mutational landscapes of two tumours without TP53 mutations (Fig. 6) shall illustrate our hypothesis. The first tumour (OS-046) does likely acquired a homozygous mutation in *NUMA1*, resulting in mitotic segregation errors that at some point led to the chromoplexy of chromosomes 2, 8 and 17 driven by inter-chromosomal exchanges. The evidence supporting *NUMA1* mutation as a triggering event in OS-046 stems from the fact that all somatic SNV mutations were monoclonal, which would be hardly possible if they were sequential to the chromosomal instability. Similarly, an early evolutionary history of the second tumour (OS-059) could have been driven either by a germline *MUTYH* mutation or by a somatic *RB1* mutation. Mutations in both genes were found hemizygous in a tumour,

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms9940 ARTICLE

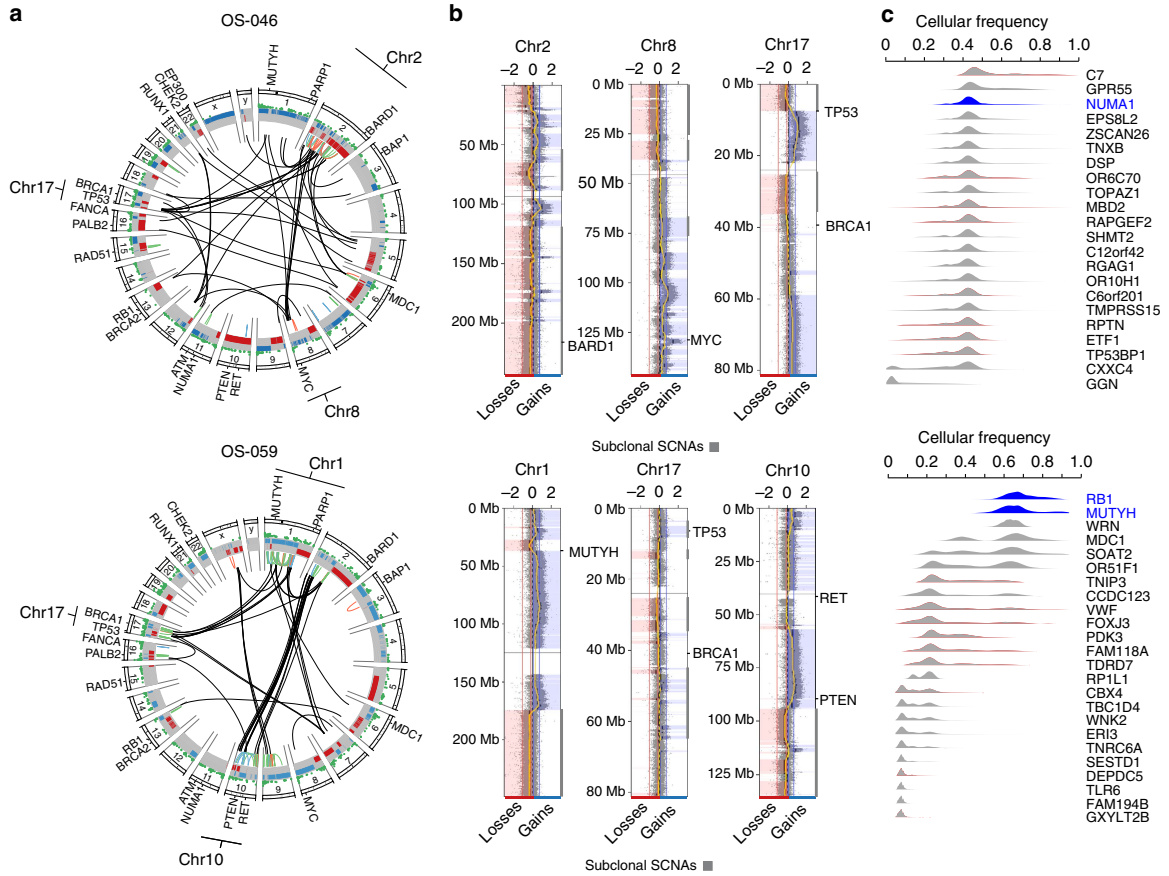


Figure 6 | Comprehensive analysis of two tumours. (**a**) A genome-wide plot depicting mutational burden, SCNAs and intra- and inter-chromosomal structural variants. Large structural rearrangements have been detected from whole-genome mate-pair sequencing data as described in Supplementary Note 1 and Supplementary Methods. (**b**) Copy-number profile of selected chromosomes targeted by chromoplexy. Note that chromoplexy likely overlaps with additional structural events including chromothripsis-like events and breakage–bridge fusions. (**c**) Posterior cellular frequencies of somatic exome mutations. Main driver genes are in blue.

both mutations were clonal events and the functions of both genes are related to chromosomal instability[2,20]. However, a parsimonious explanation would still favour a single somatic hit in the *MUTYH* gene over two somatic hits in *RB1*.

In a recent study Nik-Zainal *et al.*[12] described a distinct profile (which they termed profile D) of substitutions and deletions in breast cancer with *BRCA1/2* mutations, which later was reported by Alexandrov *et al.*[11] as signature 3 in a pan-cancer analysis. With two outliers (OS-241 and OS-079) we found this signature (as a combination of signatures B and C; Fig. 1b) in 29/31 (94%) tumours from which exome-sequencing data were available. This seems particularly interesting because early evolution of OS has historically been attributed to *TP53* and *RB1* driver mutations, but the root molecular cause of its genomic complexity (in later phases) remains unclear. Seeking to validate this finding we turned to whole-genome copy-number profiles of 123 OS and specifically looked for characteristic large-scale copy-number changes that also constitute the hallmark of *BRCA1/2*-mutated breast cancer[12]. Surprisingly, 84% of the analysed tumours fulfilled these criteria, suggesting that exome findings are likely to be correct and that various degrees of BRCAness are acquired throughout OS evolution. A recent study in ovarian cancer developed a LOH-based homologous recombination deficiency score, which reflects the number of subchromosomal LOH segments with a size exceeding 15 Mbp. Intriguingly, a high homologous recombination deficiency-LOH score, as seen in 78% of OS assayed in our study, was shown to correlate with deficiency in homologous repair and a positive response to a combination treatment including a PARP inhibitor in breast cancer[27].

Putative deficiency in homologous repair and other similarities between OS and tumours with BRCA-like phenotype indicate a specific weakness that could be therapeutically exploited. Specifically, the inhibition of PARPs contributing to DNA damage repair was shown to induce cell cycle arrest and apoptosis in *BRCA1*-, *BRCA2*- and *PALB2*-deficient breast cancers[29,37]. Furthermore, cancer cells with mutations in the *ATM* pathway members and *PTEN* gene were also shown to be sensitive to PARP inhibitors, thus raising the possibility that a 'synthetic lethality' approach could also be effective in OS[30,33].

We sought to test this possibility on selected OS cell lines (MNNG/HOS, SAOS2, SJSA-1, ZK58, HOS, MG-63 and U2-OS) but none of them carried bi-allelic BRCA1/2 mutations. We found, however, a good response of MNNG/HOS cells carrying a disruptive gain in the *PTEN* gene and a deletion of the *ATM* gene to a standalone treatment with the phase-3 PARP inhibitor

talazoparib. Combined with the alkylating agent temozolomide or the topoisomerase I inhibitor SN-38, 0.1 and 5 μM concentrations of talazoparib led to an even greater decrease of MNNG/HOS cell viability, respectively. Furthermore, a good response of HOS and MG-63 cells to the phase-2 PARP inhibitor olaparip was recently reported by Smith *et al.*[29].

Using OS cell line experiments impose a few limitations to our study, however. Perhaps the most striking one stems from the paucity of OS cell lines truly deficient in BRCA1/2. This may be surprising, but it should be noted that OS cell lines arise through monoclonal expansions of one cancer cell surviving immortalization, which may not necessarily be the most common cell type in the tumour it is derived from. The molecular portrait of OS presented here prerequisites the existence of intratumour heterogeneity, which is in fact lost *in vitro*. The polyclonal nature of OS also implies that BRCA-like traits of OS genomes might result not only from *BRCA1/2* mutations alone but also from mutations in other genes of the homologous recombination pathway, each of which confers to additional biological properties. These limitations do not invalidate our findings or conclusions, but suggest that, for example, responsiveness of individual tumours to PARP inhibitors will not only depend on the proportion of cancer cells with BRCA-like features but also on the specific mutations and genes that bring about homologous recombination deficiency.

In conclusion, the presented data support a BRCA-like phenotype as a unifying trait of OS independently of which oncogenic pathway drives tumorigenesis. We have shown that at least 14 different genes underlie the disease, including Mendelian cancer drivers for which OS has not been reported to belong within their phenotypic repertoire. More importantly, the effect of mutations within different genes and pathways seem to complement each other and result in a specific signature characteristic for BRCA1/2-deficient tumours. Our findings warrant further testing of PARP inhibitors in experimental settings and eventually sequencing of individual tumours for therapeutically targetable driver gene mutations.

Methods

Sample description. The discovery set comprised 31 previously untreated OS samples, paired with peripheral blood or normal tissue. A paired primary tumour and a metastasis was sequenced in one patient. Informed consent was obtained from all patients and ethical approval was obtained from the local ethical committee of Heidelberg (project no. S-327/2011) for the analysis of anonymized samples and for the purpose of driver gene discovery. All tumour samples were re-evaluated by an experienced bone pathologist and confirmed the diagnosis and a tumour content >70% per sample. Genomic DNA was extracted from each tumour and paired blood sample using standard methods. The replication set comprised 92 unpaired frozen pre-therapeutic tumour biopsies, which were subjected to the DNA extraction the same way as the 31 tumours of the discovery set.

Exome sequencing. Exome capture was performed using the Agilent SureSelect kits (version 4). Samples were quantified using the Qubit system (Invitrogen) and sequencing libraries constructed from 1 μg DNA. Samples were sequenced using the Illumina HiSeq 2000 platform as paired 100-bp reads with Chemistry version 3.0, with the aim of target coverage of 100 × for the blood DNA and 200 × for the tumours. After removal of PCR duplicates using Picard, reads were mapped with Stampy version 1.0.12 onto the Human Reference Genome (GRCh37d5/hg19). SNVs and small indels were called with Platypus version 0.5 using the tumour-normal pairs of bam files together to ensure comparable calls at every locus. Variants were only called if they were assigned sufficiently high posterior probability (phred score >20). We removed the allele bias filter to increase sensitivity. Finally, for selected variants, we made sure that the automatic call matched the data by expert visual inspection of the mapped reads onto the reference genome using read direction colouring on top of the standard integrated genomic viewer scheme.

Annotation was performed using ANNOVAR using hg19 reference genome and 2014 versions of standard databases and functional prediction programs. We excluded duplicated genomic regions (>90% homology) from the analysis and variants within regions with low mapability scores. Variants were annotated with ANNOVAR Refseq gene model using dbSNP(135); 1,000 genomes project allele frequencies (October 2014), 6,500 exome-sequencing project allele frequencies, University of Santa Cruiz (UCSC) segmental duplication scores and UCSC 46 species conservation scores and predictions of functional importance from SIFT, PolyPhen2 and Mutation Taster.

For filtering of variant calls for analysis, calls were first compared between matched constitutional and tumour samples to identify somatic mutations. For the analysis of mutation burden and mutation spectra, we applied the following exclusion filters to somatic variants: (i) presence in a segmental duplication region or a region with mapability score <0.5; (ii) variant present in any read from paired normal sample; (iii) fewer than 10 reads in total at the variant site in the normal sample; (iv) fewer than 8 reads in total in the tumour; (v) fewer 3 three variants in the tumour; variant allele frequency <5% in the tumour; and (vi) presence of variant in public databases (Exome Variant Server, 1,000 genomes project and Complete Genomics 69 reference genomes) at a frequency >2%. Variants identified in constitutional DNA from any of the other local, non-cancer sequencing project (for example, 29 million variants across 284 samples from the Oxford-Illumina WGS500 consortium) were discarded as being more likely due to systematic error in our pipeline than genuine somatic mutations.

Ion Torrent technical replication. Technical replication of mutations identified by exome sequencing of 31 OS was carried out by Ion Torrent sequencing using a custom Ion AmpliSeq 388 gene panel (gene list and genomic coordinates available as Supplementary Data 2). Approximately 100 ng DNA was used for sequencing. We sequenced to an average depth of 300 × and used the Ion Reporter software for variant calling. Variant calls were annotated using the ANNOVAR software and the same databases as for exome-sequencing data. Variants present in the exomes were assessed alongside the equivalent Ion Torrent data. In addition, visual inspection with the integrated genomic viewer genome browser was required for variant calls with posterior quality scores of 20–30. For specific *RET* mutations we tested additional family members by Sanger sequencing. In addition, we used the Ion AmpliSeq 388 gene panel to validate mutations from 15 randomly picked samples that underwent Illumina MiSeq sequencing.

Targeted Illumina re-sequencing. Biological replication was carried out using 1 μg of DNA from 92 additional tumours and Illumina MiSeq re-sequencing to an average depth of 480 × . The custom Nimblegen target enrichment kit was used to capture all exons of 30 genes (that is, 13 main OS drivers and their 17 interaction partners; Supplementary Data 2) and a ~20-kb region encompassing the *TP53* gene. An average coverage of the target region was 92%. Read mapping, filtering and variant calling was done with the Stampy-Platypus pipeline using the same setting as for the discovery set of 31 exomes.

Mutation spectra and signatures. For each tumour, somatic SNV frequencies were calculated and the deviation of their spectra from the background (determined across all samples) was assessed using χ^2-test (df = 5). The trinucleotide spectrum was normalized according to 3-mer frequencies in the reference exome (Agilent SureSelect (version 4)). Somatic mutation signatures were inferred using the R package[38], in which a mutation spectrum was decomposed with a non-negative matrix factorization algorithm. The decomposition was performed for prior sets of 2, 3 and 4. The optimal number of signatures ($r = 3$) was manually chosen based on the maximum differentiation between the signatures. The signature analysis was repeated 10 times with the same results obtained after each run.

Affymetrix CytoscanHD array data analysis. Affymetrix CytoscanHD arrays were processed with ChaS 2.1 and Nexus 7.5 software such that SCNA events larger than 50 kb with a minimum support of 21 probes were considered for subsequent analysis. Gene annotations were extracted using Ensembl REST API GRCh37 assembly. We assessed statistical significance of called SCNA regions by an algorithm similar to the GISTIC[39] analysis, which identifies SCNA with frequency higher than expected by chance. Concretely, for each chromosome SCNA events were resampled to generate uniform distribution and then significance was assessed across all positions in 1-kb windows by deriving Benjamini–Hochberg adjusted *P* values (*q* values). Each candidate SCNA was subjected to survival analysis with subsequent multiple testing correction using Benjamini–Hochberg *q* values <0.05. For event-free survival and overall survival we used Cox proportional hazards regression model.

***In vitro* cell line experiments.** OS cell lines were grown in RPMI 1640 medium (SJSA-1), minimum essential medium supplemented with 10% fetal calf serum (MNNG/HOS) or in McCoy's 5A medium supplemented with 15% fetal calf serum. Talazoparib was obtained from Selleckchem (Munich, Germany). Control tests for mycoplasm contamination were carried out routinely (two times a month) using a PCR-based and commercially available detection kit according to the manufacturer's protocol (VenorGem, Minerva Biolabs, Berlin, Germany). Furthermore, authentication of cell lines has been carried out by genotyping using microsatellite markers (Supplementary Data 8) to exclude cross-contaminations. Cell viability was assessed by a 3-(4,5-dimethylthiazol-2-yl)-2,5-

diphenyltetrazolium bromide assay according to the manufacturer's instructions (Roche Diagnostics, Mannheim, Germany). Statistical significance was assessed by Student's *t*-test (two-tailed distribution, two sample, unequal variance).

References

1. Bielack, S. S. *et al.* Prognostic factors in high-grade osteosarcoma of the extremities or trunk: an analysis of 1,702 patients treated on neoadjuvant cooperative osteosarcoma study group protocols. *J. Clin. Oncol.* **20,** 776–790 (2002).
2. Chen, X. *et al.* Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. *Cell Rep.* **7,** 104–112 (2014).
3. Perry, J. A. *et al.* Complementary genomic approaches highlight the PI3K/mTOR pathway as a common vulnerability in osteosarcoma. *Proc. Natl Acad. Sci. USA* **111,** E5564–E5573 (2014).
4. Joseph, C. G. *et al.* Exomic analysis of myxoid liposarcomas, synovial sarcomas, and osteosarcomas. *Genes Chromosomes Cancer* **53,** 15–24 (2014).
5. Martin, J. W. *et al.* Digital expression profiling identifies RUNX2, CDC5L, MDM2, RECQL4, and CDK4 as potential predictive biomarkers for neo-adjuvant chemotherapy response in paediatric osteosarcoma. *PLoS ONE* **9,** e95843 (2014).
6. Murata, K. *et al.* A case of Werner's syndrome associated with osteosarcoma. *J. Dermatol.* **26,** 682–686 (1999).
7. Lu, L., Jin, W., Liu, H. & Wang, L. L. RECQ DNA helicases and osteosarcoma. *Adv. Exp. Med. Biol.* **804,** 129–145 (2014).
8. Wang, L. L. *et al.* Association between osteosarcoma and deleterious mutations in the RECQL4 gene in Rothmund-Thomson syndrome. *J. Natl Cancer Inst.* **95,** 669–674 (2003).
9. Mohseny, A. B. *et al.* Small deletions but not methylation underlie CDKN2A/p16 loss of expression in conventional osteosarcoma. *Genes Chromosomes Cancer* **49,** 1095–1103 (2010).
10. Scott, M. C. *et al.* Molecular subtypes of osteosarcoma identified by reducing tumour heterogeneity through an interspecies comparative approach. *Bone* **49,** 356–367 (2011).
11. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).
12. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149,** 979–993 (2012).
13. Trueb, B., Zhuang, L., Taeschler, S. & Wiedemann, M. Characterization of FGFRL1, a novel fibroblast growth factor (FGF) receptor preferentially expressed in skeletal tissues. *J. Biol. Chem.* **278,** 33857–33865 (2003).
14. Mohaghegh, P. & Hickson, I. D. DNA helicase deficiencies associated with cancer predisposition and premature ageing disorders. *Hum. Mol. Genet.* **10,** 741–746 (2001).
15. Fuchs, B. & Pritchard, D. J. Etiology of osteosarcoma. *Clin. Orthop. Relat. Res.* 40–52 (2002).
16. Tavtigian, S. V. *et al.* Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. *Am. J. Hum. Genet.* **85,** 427–446 (2009).
17. Johnson, N. *et al.* Counting potentially functional variants in BRCA1, BRCA2 and ATM predicts breast cancer susceptibility. *Hum. Mol. Genet.* **16,** 1051–1057 (2007).
18. Gundem, G. *et al.* IntOGen: integration and data mining of multidimensional oncogenomic data. *Nat. Methods* **7,** 92–93 (2010).
19. Gonzalez-Perez, A. *et al.* IntOGen-mutations identifies cancer drivers across tumour types. *Nat. Methods* **10,** 1081–1082 (2013).
20. Cardoso, J. *et al.* Chromosomal instability in MYH- and APC-mutant adenomatous polyps. *Cancer Res.* **66,** 2514–2519 (2006).
21. Yin, Y. & Shen, W. H. PTEN: a new guardian of the genome. *Oncogene* **27,** 5443–5453 (2008).
22. Lee, J.-H. *et al.* MicroRNA-22 suppresses DNA repair and promotes genomic instability through targeting of MDC1. *Cancer Res.* **75,** 1298–1310 (2015).
23. Vidi, P.-A. *et al.* NuMA promotes homologous recombination repair by regulating the accumulation of the ISWI ATPase SNF2h at DNA breaks. *Nucleic Acids Res.* **42,** 6365–6379 (2014).
24. Walden, H. & Deans, A. J. The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder. *Annu. Rev. Biophys.* **43,** 257–278 (2014).
25. Popova, T. *et al.* Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. *Cancer Res.* **72,** 5454–5462 (2012).
26. Abkevich, V. *et al.* Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *Br. J. Cancer* **107,** 1776–1782 (2012).
27. Telli, M. L. *et al.* Phase II study of gemcitabine, carboplatin, and iniparib as neoadjuvant therapy for triple-negative and BRCA1/2 mutation-associated breast cancer with assessment of a tumour-based measure of genomic instability: PrECOG 0105. *J. Clin. Oncol.* **33,** 1895–1901 (2015).
28. Roth, A. *et al.* PyClone: statistical inference of clonal population structure in cancer. *Nat. Methods* **11,** 396–398 (2014).
29. Smith, M. A. *et al.* Initial testing (stage 1) of the PARP inhibitor BMN 673 by the pediatric preclinical testing program: PALB2 mutation predicts exceptional in vivo response to BMN 673. *Pediatr. Blood Cancer* **62,** 91–98 (2015).
30. Mendes-Pereira, A. M. *et al.* Synthetic lethal targeting of PTEN mutant cells with PARP inhibitors. *EMBO Mol. Med.* **1,** 315–322 (2009).
31. McCabe, N. *et al.* Deficiency in the repair of DNA damage by homologous recombination and sensitivity to poly(ADP-ribose) polymerase inhibition. *Cancer Res.* **66,** 8109–8115 (2006).
32. Hay, T. *et al.* Efficient deletion of normal Brca2-deficient intestinal epithelium by poly(ADP-ribose) polymerase inhibition models potential prophylactic therapy. *Cancer Res.* **65,** 10145–10148 (2005).
33. Weston, V. J. *et al.* The PARP inhibitor olaparib induces significant killing of ATM-deficient lymphoid tumour cells in vitro and in vivo. *Blood* **116,** 4578–4587 (2010).
34. Smith, M. A. *et al.* Synergistic activity of PARP inhibition by talazoparib (BMN 673) with temozolomide in pediatric cancer models in the pediatric preclinical testing program. *Clin. Cancer Res.* **21,** 819–832 (2015).
35. Hopkins, T. A. *et al.* Mechanistic dissection of PARP1 trapping and the impact on in vivo tolerability and efficacy of PARP inhibitors. *Mol. Cancer Res.*, doi:10.1158/1541-7786.MCR-15-0191-T (2015).
36. Ribi, S. *et al.* TP53 intron 1 hotspot rearrangements are specific to sporadic osteosarcoma and can cause Li-Fraumeni syndrome. *Oncotarget* **6,** 7727–7740 (2015).
37. Thompson, L. H. Recognition, signaling, and repair of DNA double-strand breaks produced by ionizing radiation in mammalian cells: the molecular choreography. *Mutat. Res.* **751,** 158–246 (2012).
38. Gehring, J.-S., Fischer, B., Lawrence, M. & Huber, W. SomaticSignatures: inferring mutational signatures from single-nucleotide variants. *Bioinformatics*, doi:10.1093/bioinformatics/btv408 (2015).
39. Beroukhim, R. *et al.* Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc. Natl Acad. Sci. USA* **104,** 20007–20012 (2007).

Acknowledgements

Mi.K., S.R., G.J. and D.B. were supported by the foundation of the Basel Bone Tumour Reference Centre. Mi.K., S.R. and D.B. were furthermore supported by the Gertrude von Meissner-Stiftung. D.B. was additionally supported by the Wilhelm-Sander-Stiftung. Mi.K. was supported by the University of Basel (Förderung exzellenter Nachwuchs-forschender). S.Bi. received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under the project ENCCA, grant agreement HEALTH-F2-2011-261474. J.S. and M.N. were supported by the BMBF (TranSarNet, FKZ 01GM0870).

Author contributions

M.N. and D.B. initiated the study; Mi.K., C.B., M.N. and D.B. designed the study; C.B., A.K., D.A., P.-U.T., R.D., H.R., K.-D.S., I.M., S.Bu., A.K., K.S., S.Bi., G.J. and M.N. collected OS samples and the corresponding clinical data; J.W. and J.O.K. performed exome and whole-genome sequencing, Mi.K. carried out targeted sequencing and J.S. performed SNP array analysis; F.E. and S.F. conducted the cell line experiments; Mi.K., S.R., N.S.M., F.C.-G., J.W., Mo.K. and I.T. evaluated and analysed sequencing and SNP data; Mi.K. and Mo.K. carried out clonal modelling techniques; Mi.K. and D.B. wrote the manuscript and all other authors contributed to it.

Additional information

Accession codes: Exome-sequencing data for osteosarcoma samples have been deposited in the European Nucleotide Archive under the study accession number: PRJEB11430 and the secondary study accession number: ERP012816. Affymetrix CytoscanHD data are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-3998.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Kovac, M. *et al.* Exome sequencing of osteosarcoma reveals mutation signatures reminiscent of BRCA deficiency. *Nat. Commun.* 6:8940 doi: 10.1038/ncomms9940 (2015).

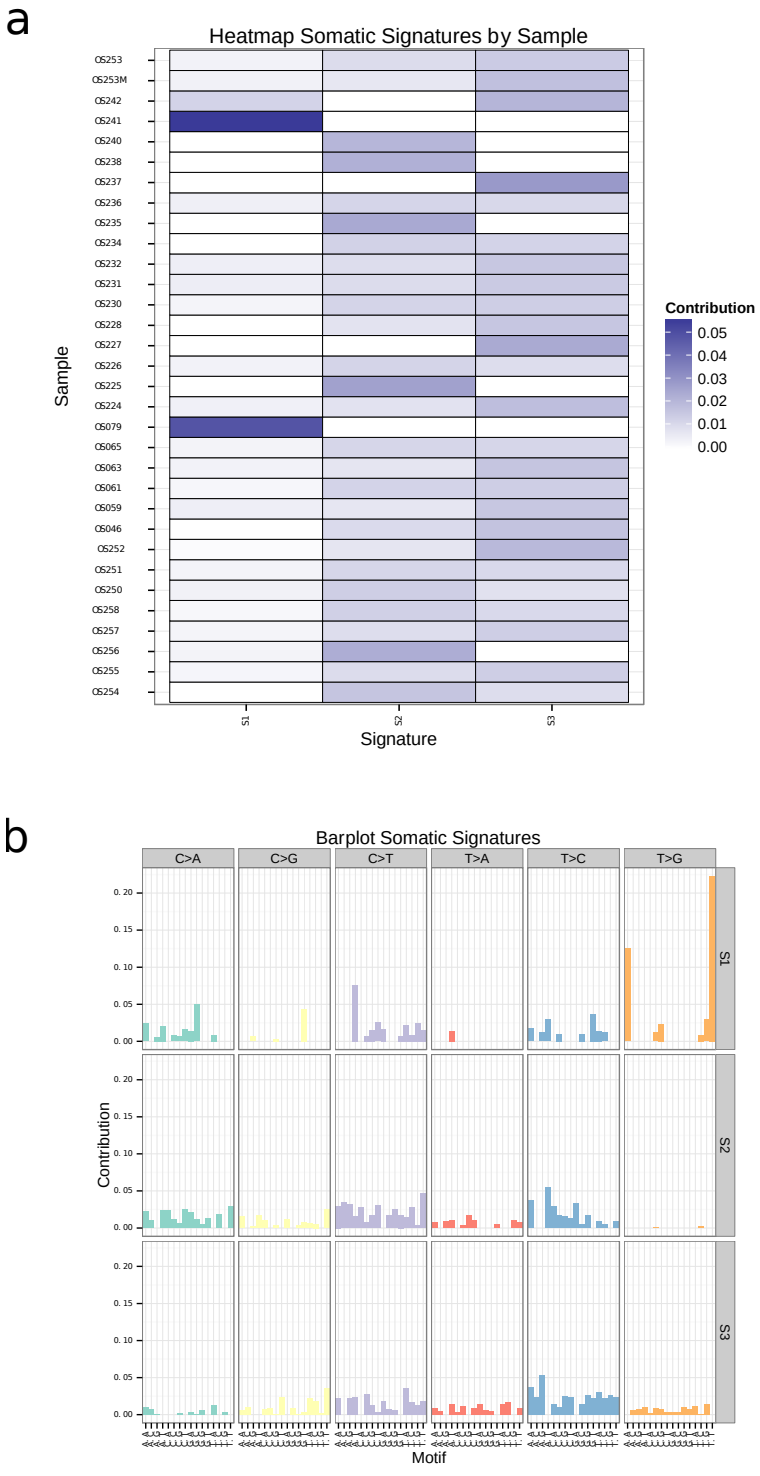


Figure S1 – Mutation signatures of 31 exome-sequenced osteosarcomas. **(a)** Contribution of each signature to filtered SNV mutation burden of each tumor. **(b)** mutations within a trinucleotide context split by a signature.

2.4 Submitted Manuscript No 3

RET Germline Mutations and Susceptibility to Osteosarcoma

Michal Kovac*, **Sebastian Ribi***, Claudia Blattmann, Eva Roth, Monika Kovacova, Andreas Kulozik, Wolfgang Hartmann, Stefan Bielack, Mandy Ballinger, David Thomas, Michaela Nathrath, Karl Heinimann*, and Daniel Baumhoer*
***These authors contributed equally to this work**

Manuscript submitted

Note to the reader: The following section includes a manuscript as it was submitted for peer-review. The abbreviation *SV* is used for *sequence variant* and therefore deviates from the rest of this thesis, where it stands for *structural variant*.

Germline RET Mutations and Susceptibility to Osteosarcoma

Michal Kovac[1#], Sebastian Ribi[1#], Claudia Blattmann[2,3], Eva Roth[3], Monika Kovacova[4], Andreas Kulozik[3], Wolfgang Hartmann[5], Stefan Bielack[2], Mandy Ballinger[6], David Thomas[6], Michaela Nathrath[7-8], Karl Heinimann[9*], and Daniel Baumhoer[1*]

1: Bone Tumour Reference Center at the Institute of Pathology, University Hospital Basel and University of Basel, Schoenbeinstrasse 40, 4031 Basel, Switzerland

2: Pediatrics 5 (Oncology, Hematology, Immunology), Klinikum Stuttgart Olgahospital, Kriegsbergstrasse 62, 70174 Stuttgart, Germany

3: Department of Pediatric Hematology, Oncology, Immunology and Pulmology, University of Heidelberg, Im Neuenheimer Feld 430, 69120 Heidelberg, Germany

4: The Institute of Mathematics and Physics, Faculty of Mechanical Engineering, Slovak University of Technology, 84248 Bratislava, Slovak Republic

5: Gerhard-Domagk-Institut of Pathology, University Hospital Münster, Albert-Schweitzer-Campus 1, D17, 48149 Münster, Germany

6: The Kinghorn Cancer Centre and Cancer Division, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia

7: Department of Pediatric Oncology, Klinikum Kassel, Moenchebergstrasse 41-43, 34125 Kassel, Germany

8: Pediatric Oncology Center, Department of Pediatrics, Technische Universität München and Comprehensive Cancer Center, Kölner Platz 1, 80804 Munich, Germany

9: Medical Genetics and Research Group Human Genomics, University Hospital Basel and University of Basel, Schoenbeinstrasse 40, 4031 Basel, Switzerland

# The authors equally contributed to this work

* The authors jointly supervised this work

2.4. RET Germline Mutations and Susceptibility to Osteosarcoma

Osteosarcoma is the most common primary tumor of the bone for which the identification of the underlying genetic alterations has remained difficult. Applying an integrative genomics approach in 337 patients we identified previously unknown germline *bona fide* pathogenic mutations in the RET proto-oncogene in 2% of cases, which is a proportion similar to TP53-related osteosarcomas. These RET mutations appear to couple functional kinase activity to dysfunctional ligand binding in the same molecule and act in cooperation with (epi)genetic alterations compromising cell cycle regulation and homologous recombination repair. Our findings indicate that patients carrying pathogenic RET germline mutations are at an increased risk of developing osteosarcoma (odds ratio 9.12) and highlight RET as a potential target for multi-targeted tyrosine kinase inhibitors.

The RET proto-oncogene (REarranged during Transfection) encodes a single-pass transmembrane receptor tyrosine kinase for the glial cell line-derived neurotrophic factor family of ligands. Its pivotal role in medullary thyroid cancer (MTC) development is well-established: while RET activating mutations are present in approximately 43% of sporadic MTCs[1], germline alterations can be identified in 95 to 98% of patients with the autosomal dominant endocrine tumor syndrome MEN2 (multiple endocrine neoplasia type 2)[2]. Although more than 100 RET variants have been reported to date, most recurrent variants are limited to a number of key residues in the extracellular and kinase domains. Genotype-phenotype correlations allowed an estimate of the risk of developing MTC and other RET-associated disorders, including pheochromocytoma and parathyroid adenoma[3].

Studies using phosphoproteomic screening proposed that the activation of receptor tyrosine kinases including RET contributes to the development of metastatic osteosarcoma[4]. Since confirmation of these results awaits the identification of the underlying pathogenic somatic or germline alterations, we reasoned that RET mutations could have contributed to the development of an osteosarcoma in a 34-years-old patient (family ID: OS 263). The patient was diagnosed with MTC (age 30, Fig. 1) and carried a germline *bona fide* pathogenic RET mutation, p.Ser891Ala, which confers a moderate risk of MTC[3], and which was also present in the patient's mother who was diagnosed with MTC at age 66. Using an integrative analysis of the patient's osteosarcoma we identified loss of the RET wild-type

allele resulting from a structural rearrangement of chromosome 10, together with somatic (epi-) mutations in cell cycle regulators RB1, ATRX and TP53 (Fig. 1).

We therefore hypothesized that RET mutations might define a molecular subtype of osteosarcoma and examined protein-coding sequence variants (SV) in the germline of additional 336 affected individuals of European, Asian and American ancestry (our patients n=248, public data n=89, Supplementary Fig. 1). Using a two-pronged strategy – searching for functionally important variants and filtering using consensus mutation classification criteria[5] – six additional patients were identified to carry SVs encoding known *bona fide* pathogenic RET mutations (p.Val292Met, p.Gly321Arg, p.Asn361Lys, p.Lys603Glu, p.Cys620Arg, and p.Cys620Ser). Four more patients carried RET missense SVs of uncertain clinical significance (p.Pro218Arg, p.Phe381Leu, p.Leu389Phe, and p.Tyr1038Ala). Pathogenic RET germline mutations in osteosarcoma patients have all been considered moderate in terms of risk level for MTC[3]. Six of seven mutations and three of four SVs with uncertain clinical significance mapped to the extracellular domain of the protein, especially to the cadherin-like (CLD) and cysteine-rich (CR) domains. Complete clinical and genetic information of the eleven patients are depicted in the Supplementary Table 1. None of these patients had a personal history of MTC and family history was uneventful.

RET mutations mapped predominantly to the extracellular domains of the protein, often at or next to the sites where both gain-of-function and loss-of-function mutations occur. Specifically, in two of our patients we identified dual-type Janus mutations at the cysteine 620 residues which have been associated with both MEN2 and Hirschsprung's disease. Like the two mutually incompatible faces of the Roman god Janus, these mutations lead to proteins incapable of responding to the glial cell line-derived neurotrophic factor family of ligands, but can also stimulate cell proliferation as typical MEN2 mutations do[6]. Similarly, the studies of CLD domain mutations confirmed that a likely pathogenic mechanism is coupling functional kinase activity to dysfunctional ligand binding in the same molecule[7].

2.4. RET Germline Mutations and Susceptibility to Osteosarcoma

To compare the prevalence of germline RET mutations in osteosarcoma patients to the population, we analyzed exome sequencing data compiled from 53105 individuals included in the Exome Aggregation Consortium (ExAC). We excluded data from individuals with cancer who have been included in Cancer Genome Atlas studies. The cumulative odds of any American Thyroid Association-classified MEN2 type A or familial MTC-predisposing RET mutation in osteosarcoma patients were significantly higher than the odds in the ExAc population (odds ratio 9.12, 95% CI, 4.26 to 19.54, P<0.0001, ExAC sequence variant database version 3.1).

We then examined chemotherapy-naïve tumor biopsies from eight carriers of germline SVs in RET for somatic alterations. A second hit by loss of heterozygosity involving the germline wild-type RET allele was found in three (37.5%) out of eight osteosarcomas (Supplementary Fig. 2). No additional somatic RET mutations were detected. In some respect genomes of the RET-mutant tumors were similar to those of previously described osteosarcomas; tumors were microsatellite-stable, but had a comparably higher mutation burden, chromosomal instability with preponderance to deletions, chromothrypsis, kataegis and frequent rearrangements/fusion genes at loci harboring the *MYC*, *ERBB2*, *FGFR1-4*, *TP53* and *RB1* genes (Fig. 2, Supplementary Fig. 3 and 4, Supplementary Table 2). Signature analysis confirmed the acquisition of BRCA-like traits also in RET-mutant osteosarcoma. Unlike previously-published clonal structures of osteosarcoma genomes, most somatic driver mutations, indels, fusion genes and chromosomal copy-number changes in RET-mutant osteosarcomas were either heterozygous or subclonal events.

Based on our findings, osteosarcoma must be considered a part of the tumor spectrum of RET-associated disorders. Consideration of RET mutation testing in osteosarcoma patients, and conversely, awareness of an increased risk to develop osteosarcoma in patients with RET-associated disorders could enable early identification of tumor predisposition-related disease and at-risk family members. When added to the list of genetic aberrations in tyrosine kinase genes in osteosarcoma[4,8], mutations in RET emphasize the critical role of cell cycle dysregulation in osteosarcoma development and highlight the potential benefits of the use of multi-targeted tyrosine kinase inhibitors with anti-RET activity.

Acknowledgements

The results published here in part based upon data generated by the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative managed by the NCI. The data used for this analysis are available under dbGaP accession number phs000468 and was retrieved on 28th September 2016. Information about TARGET can be found at http://ocg.cancer.gov/programs/target. We also thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532/Z/09/Z) for generation of sequencing data. M.K.., S.R. and D.B. were supported by the Gertrude von Meissner-Stiftung, the Gedächtnis-Stiftung Susy Rückert and the Stiftung für krebskranke Kinder Regio Basiliensis.

Author Contribution

M. Kovac, S.R and M. Kovacova processed omics datasets and performed bioinformatics analyses. S.R. and E.R. performed laboratory experiments. C.B. and D.B. supervised laboratory experiments and K.H. supervised genetic analyses. C.B., A.K., W.H., S.B., M.B., D.T. and M.N. collected data and/or provided patient materials. M. Kovac, K.H. and D.B wrote the manuscript.

Online Methods

Patients, Samples and Data Types

Osteosarcoma patients comprise four cohorts of which the first three – Swiss, German and Australian patients – include 248 individuals, whilst exomes of 89 patients have been retrieved from the TARGET dataset. Annotations of patients, data types and the cohorts are depicted on Fig. 1, Supplementary Fig. 1 and Supplementary Table 1. Informed consent was obtained from all patients and ethical approval was obtained from the local ethical committee of Heidelberg (ref. S-327/2011) and Ethikkommission beider Basel (ref. EK: 274/12) for analysis of anonymized samples and for the purpose of driver gene discovery. The Australian patients were enrolled under the umbrella of the International Kindred Sarcoma Consortium[9].

DNA and RNA extraction

DNA and RNA was extracted from fresh-frozen tumor samples using a TissueLyser (Qiagen) and the AllPrep DNA/RNA Micro or Mini Kit (Qiagen) according to the manufacturer's instructions. RNA was incubated with RNAse-Free DNase (Qiagen) according to the manufacturer's instructions. Constitutive DNA was extracted from whole blood using QIAamp DNA blood mini kit (Qiagen). Quality of both RNA and DNA was assessed by Nanodrop or 2100 Bioanalyzer (Agilent), respectively.

Exome sequencing

Exome capture was performed using the Agilent SureSelect kit version 4. Samples were quantified using the Qubit system (Invitrogen) and sequencing libraries constructed from 0.5-1 µg DNA. Samples were sequenced using the Illumina HiSeq 2000 or 4000 platform as paired 100-bp reads with the aim of a minimum target coverage of 150x. After removal of PCR duplicates using Picard, reads were mapped with BWA version 0.7 and Stampy version 1.0.23 onto the Human Reference Genome (GRCh37d5/hg19).

Whole-genome sequencing

DNA samples were quantified using the Qubit system (Invitrogen) and sequencing libraries constructed from 0.5µg of DNA. Samples were sequenced using the Illumina HiSeq 2000 or 4000 platform as paired 100-bp reads with Chemistry version 3.0, with the aim of a minimum target coverage of 15x. Quality control, removal of PCR duplicates and Stampy/BWA mapping was done using the bioinformatics pipeline used for exome sequencing.

Gene-Panel sequencing

Target sequencing was carried out using 1µg of DNA and Illumina MiSeq resequencing to an average depth of 480x. The custom Nimblegen enrichment array was constructed to capture coding exons of 30 genes and ~20kb region encompassing the TP53 gene. Am average coverage of the target region was 92%. Read mapping, filtering and variant calling was done with the Stampy-Platypus pipeline using the same settings as for exome sequencing.

RNA sequencing

Ribo-depleted fraction of the RNA samples was isolated from the total RNA before conversion to cDNA. The cDNA was end-repaired, A-tailed and adapter-ligated. The prepared libraries were size-selected, quality-controlled before pair-end sequencing over a lane of a flow cell.

Variant calling

SVs and small indels were called with Platypus version 0.8.1 using the tumour-normal pairs of bam files together to ensure comparable calls at every locus. Where available, reads from whole genomes and exomes were pooled. Variants were only called if they were assigned sufficiently high posterior probability (phred score >20). We removed the allele bias filter to increase sensitivity. For selected variants, we made sure that the automatic call matched the data by expert visual inspection of the mapped reads onto the reference genome using read direction colouring on top of the standard integrated genomic viewer scheme.

Annotation was performed using ANNOVAR (release date version 2016-02-01) with the hg19 reference genome and 2015 versions of standard databases and functional prediction programs. We excluded duplicated genomic regions (>90% homology) from the analysis and variants within regions with low mapping scores. Variants were annotated with ANNOVAR Refseq gene model using dbSNP(135); 1,000 genomes project allele frequencies (August 2015), 6,500 exome-sequencing project allele frequencies, The ExAC database of variant frequencies, University of Santa Cruiz (UCSC) segmental duplication scores and UCSC 46 species conservation scores and predictions of functional importance from SIFT, PolyPhen2 and Mutation Taster.

Germline and somatic mutations were identified by comparison of matched constitutional and tumor samples. For the analysis of mutation burden and mutation spectra, we applied the following exclusion filters to somatic variants: (i) presence in a segmental duplication region or a region with mapability score <0.5; (ii) variant present in any read from paired normal sample; (iii) fewer than 10 reads in total at the variant site in the normal exome; (iv) fewer than 8 reads in total in the tumor exome; (v) fewer three variants in the tumour exome; variant allele frequency <5% in the tumour; and (vi) presence of variant in public databases (Exome Variant Server, 1,000 genomes project and Complete Genomics 69 reference genomes) at a frequency >1%. Variants identified in constitutional DNA from any of the other local, non-cancer sequencing project (for example, 29 million variants across 284 samples from the Oxford-Illumina WGS500 consortium) were discarded as being more likely due to systematic error in our pipeline than genuine somatic mutations.

Structural rearrangements

Structural variants were predicted from whole-genome sequencing data using breakdancer version 1.4.5 and the Human Reference Genome GRCh37d. Candidate rearrangements which were present in a normal control genome, or mapping to any decoy chromosome or having less than 6 supporting reads were filtered out.

Mutation spectra and signatures

For each tumor, somatic sequence variant frequencies were calculated and the deviation of their spectra from the background (determined across all samples) was assessed using Chi square-test (df=5). The trinucleotide spectrum was normalized according to 3-mer frequencies in the reference exome (Agilent SureSelect (version 4)). Somatic mutation signatures were inferred using the R package38, in which a mutation spectrum was decomposed with a non-negative matrix factorization algorithm. The decomposition was performed for prior sets of 2 to 6. The optimal number of signatures (r=6) was manually chosen based on the maximum differentiation between the signatures. The signature analysis was repeated 10 times with the same results obtained after each run.

Methylation

250 ng of DNA were used on an Infinium MethylationEPIC BeadChip (Illumina, USA). Differential methylation analysis between tumors was performed in R using the Chip Analysis Methylation Pipeline for Illumina HumanMethylation450 and EPIC chips (version 1.11.1). The data was normalized by beta-mixture quantile normalization and methylation variable positions were segmented into biologically relevant differentially methylated regions using the Probe Lasso algorithm.

Fusion gene prediction

For each sample, 45 million sequencing reads were aligned to the Human Reference Genome (hg19) using TopHat version 2.0.10. Fusion gene predictions were performed with fusioncatcher software, version 0.99a and ericscript version 0.5.5, using ensembl_v84.

Pyclone Analysis

For each tumor, pyClone[12] probabilistic model was used to convert allelic count data of successfully validated somatic mutations into cellular frequencies, representing the fraction of cells in the tumour population containing a specific set of mutations.

Data Availability

Exome/genome-sequencing data have been deposited in the European Nucleotide Archive under the study accession number: PRJEB11430 and the secondary study accession number: ERP012816. Affymetrix CytoscanHD data are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-3998. Sequencing data from the International Sarcoma Kindred Consortium are not publically available. Sequencing data of the TARGET initiative is available for controlled access in the database for Genotypes and Phenotypes (https://dbgap.ncbi.nlm.nih.gov).

References

1. Elisei, R., *et al.* Prognostic significance of somatic RET oncogene mutations in sporadic medullary thyroid cancer: a 10-year follow-up study. *J Clin Endocrinol Metab* **93**, 682-687 (2008).
2. Eng, C., *et al.* The relationship between specific RET proto-oncogene mutations and disease phenotype in multiple endocrine neoplasia type 2. International RET mutation consortium analysis. *JAMA* **276**, 1575-1579 (1996).
3. Wells, S.A., Jr., *et al.* Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid* **25**, 567-610 (2015).
4. Rettew, A.N., Getty, P.J. & Greenfield, E.M. Receptor tyrosine kinases in osteosarcoma: not just the usual suspects. *Adv Exp Med Biol* **804**, 47-66 (2014).
5. Richards, S., *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-424 (2015).
6. Arighi, E., *et al.* Biological effects of the dual phenotypic Janus mutation of ret cosegregating with both multiple endocrine neoplasia type 2 and Hirschsprung's disease. *Mol Endocrinol* **18**, 1004-1017 (2004).
7. Anders, J., Kjar, S. & Ibanez, C.F. Molecular modeling of the extracellular domain of the RET receptor tyrosine kinase reveals multiple cadherin-like domains and a calcium-binding site. *J Biol Chem* **276**, 35808-35817 (2001).
8. Dani, N., *et al.* The MET oncogene transforms human primary bone-derived cells into osteosarcomas by targeting committed osteo-progenitors. *J Bone Miner Res* **27**, 1322-1334 (2012).
9. Ballinger, M.L., *et al.* Monogenic and polygenic determinants of sarcoma risk: an international genetic study. *Lancet Oncol* **17**, 1261-1271 (2016).
10. Goodman, K.M., *et al.* RET recognition of GDNF-GFRalpha1 ligand by a composite binding site promotes membrane-proximal self-association. *Cell Rep* **8**, 1894-1904 (2014).

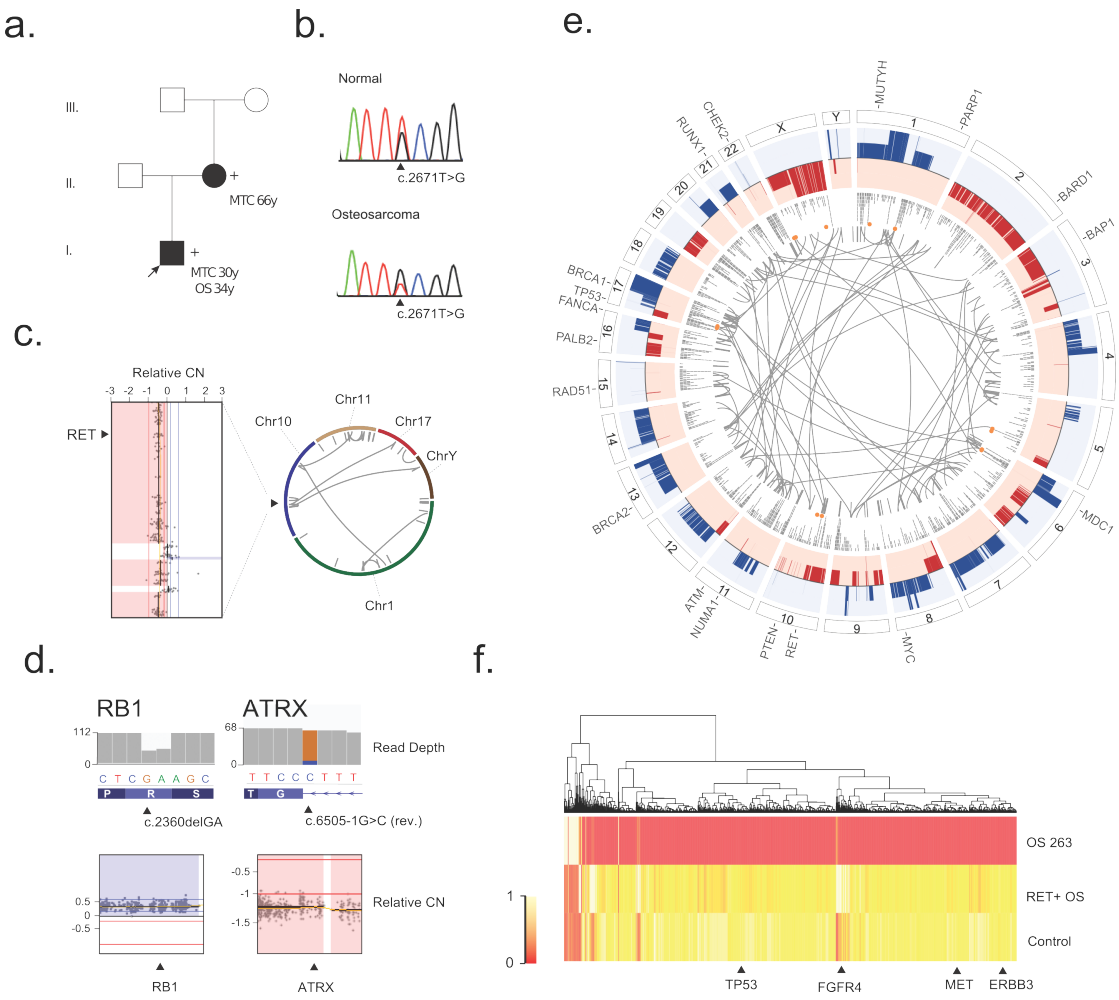


Figure 1 – The germline RET mutation c.2671T>G and the genetic profile of the patient's osteosarcoma. **(a)** Multiple thyroid cancer (MTC) in the history of the family 263. The index patient with osteosarcoma (OS) is indicated by an arrow. **(b)** Sanger sequencing chromatograms from normal and tumor tissue depict the mutation and **(c)** the loss of the wild-type allele on the chromosome 10. Inter- and intra-chromosomal rearrangements affecting chromosome 10 are shown on the right. **(d)** Somatic driver mutations in *RB1* and *ATRX* affected by non-neutral copy-number loss of heterozygosity. **(e)** Mutation burden of the tumor. The outermost layer displays chromosomal orientation and the locations of osteosarcoma driver genes. The next inner layers depict chromosomal losses, gains and SV densities, respectively. Putative sites of kataegis are highlighted in orange, chromosomal gains are in blue and losses in red. The two innermost tracks depict inter- and intra-chromosomal rearrangements. **(f)** Methylome profile of the patient's osteosarcoma comprises 1701 genes and is compared to 28 *RET*-wild type osteosarcomas and constitutional DNA of respective patients. The analysis was carried out in triplicates and only genes with P values <0.00001 are shown.

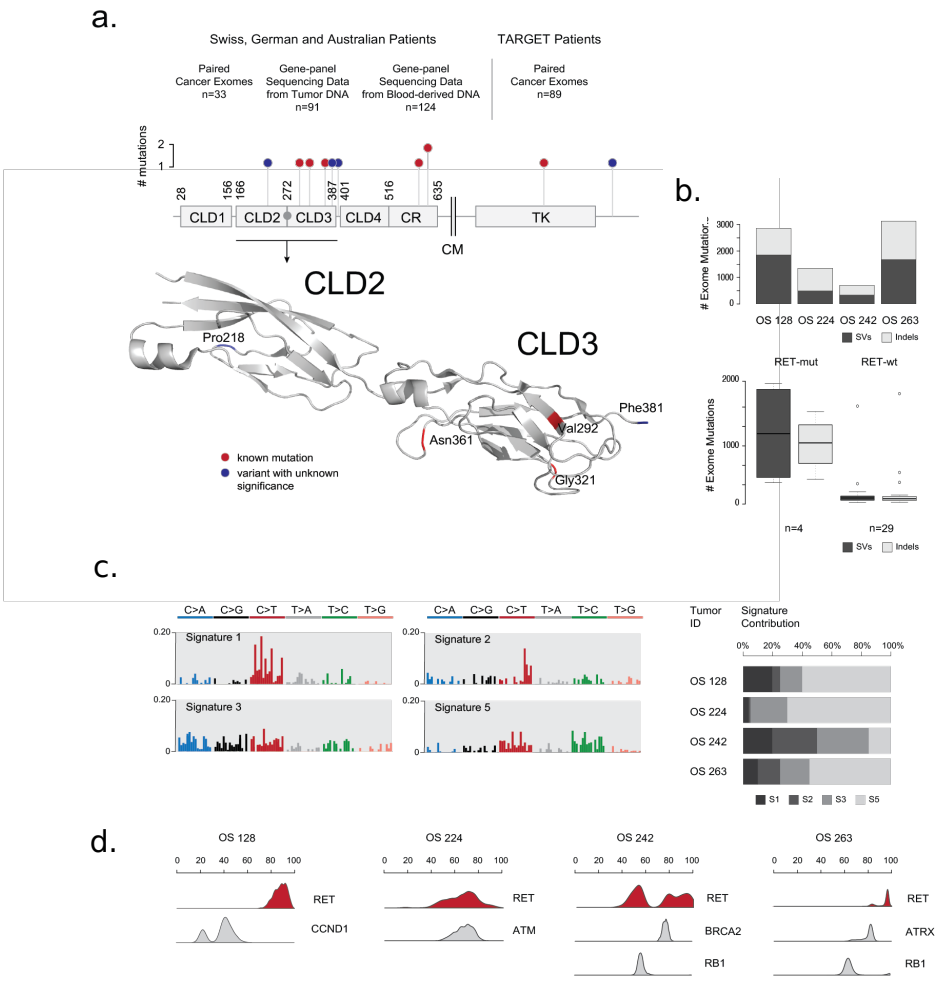


Figure 2 – Germline RET sequence variants identified in 337 osteosarcoma patients. **(a)** Variants are shown as lollipop structures with type indicated by color. Mutations affecting cadherin-like domains (CLD) 2 and 3 were mapped onto the protein domain structure 4UX8 (DOI: 10.2210/pdb4ux8/pdb) which was constructed by Goodman et al.[1]5 **(b)** Mutation burden of *RET*-mutant and *RET*-wild type osteosarcomas. **(c)** Signatures and signature contributions of mutation processes in four osteosarcoma genomes with *RET* mutations. Signature numbers are referring to the COSMIC Mutational Signatures Catalogue [110]. **(d)** Posterior clonal frequencies of the osteosarcoma driver mutations with *RET* mutations being highlighted.

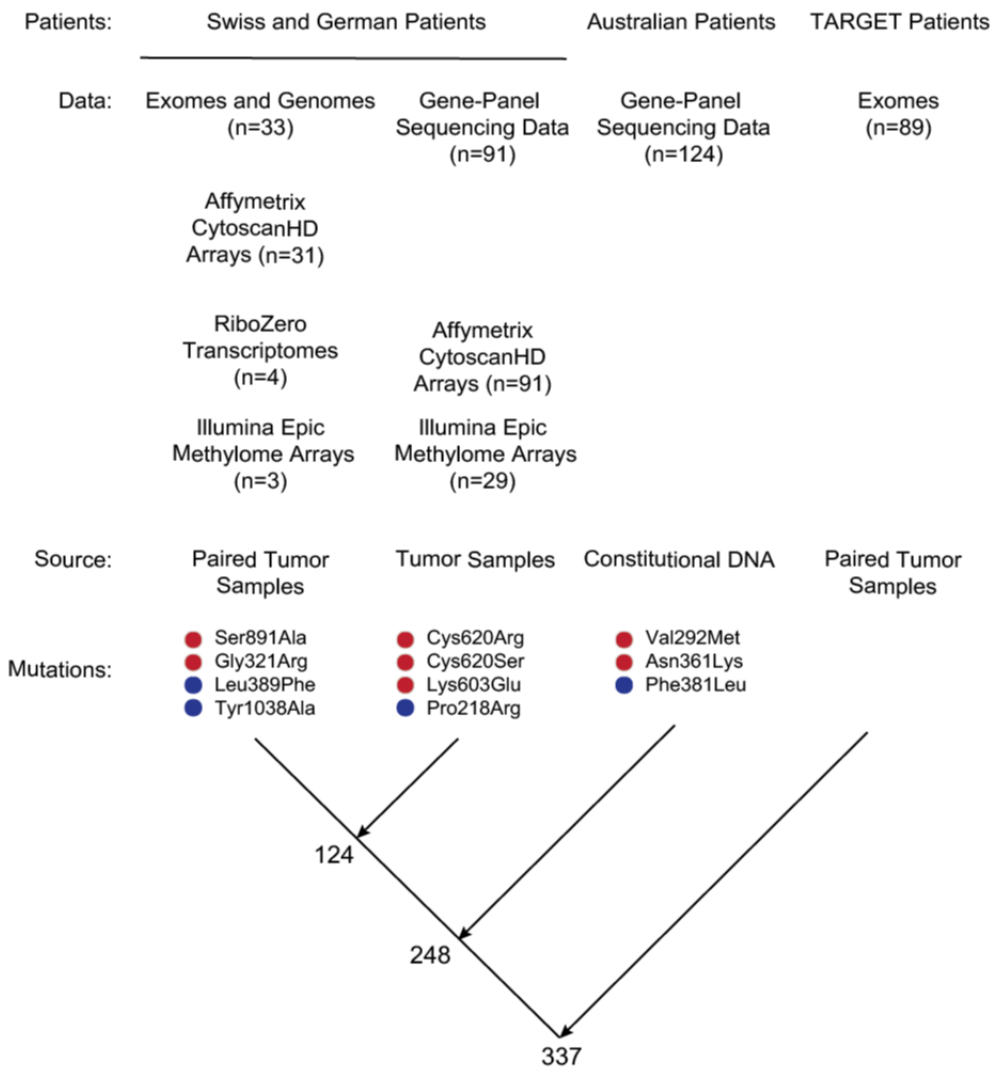


Figure S1 – Patients and Cohorts. Data from 248 own osteosarcoma patients and 89 patients from the TARGET initiative was analyzed. Data types, which include Cytoscan HD copy number arrays, RiboZero transcriptomes, Illumina Epic methylation arrays, whole-genome/exome sequencing data and targeted sequencing, and the respective DNA and RNA sources are indicated for each patient subset. *RET* mutations which are classified as pathogenic (red) and *RET* variants with unknown clinical significance (blue) are listed.

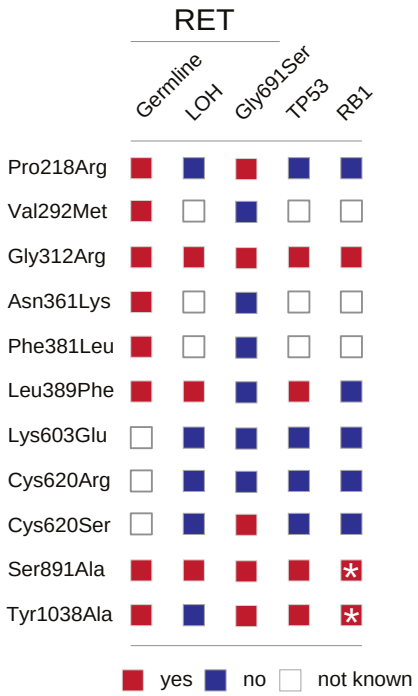


Figure S2 – Overview of *RET*-mutated Osteosarcomas. Information about the presence of LOH at the *RET* locus, the presence of the enhancer *RET* mutation p.Gly691Ser and driver mutations in the *TP53* and *RB1* genes. Mutational events include disruptive gains, chromosomal losses and point mutations with known pathogenic effect (*).

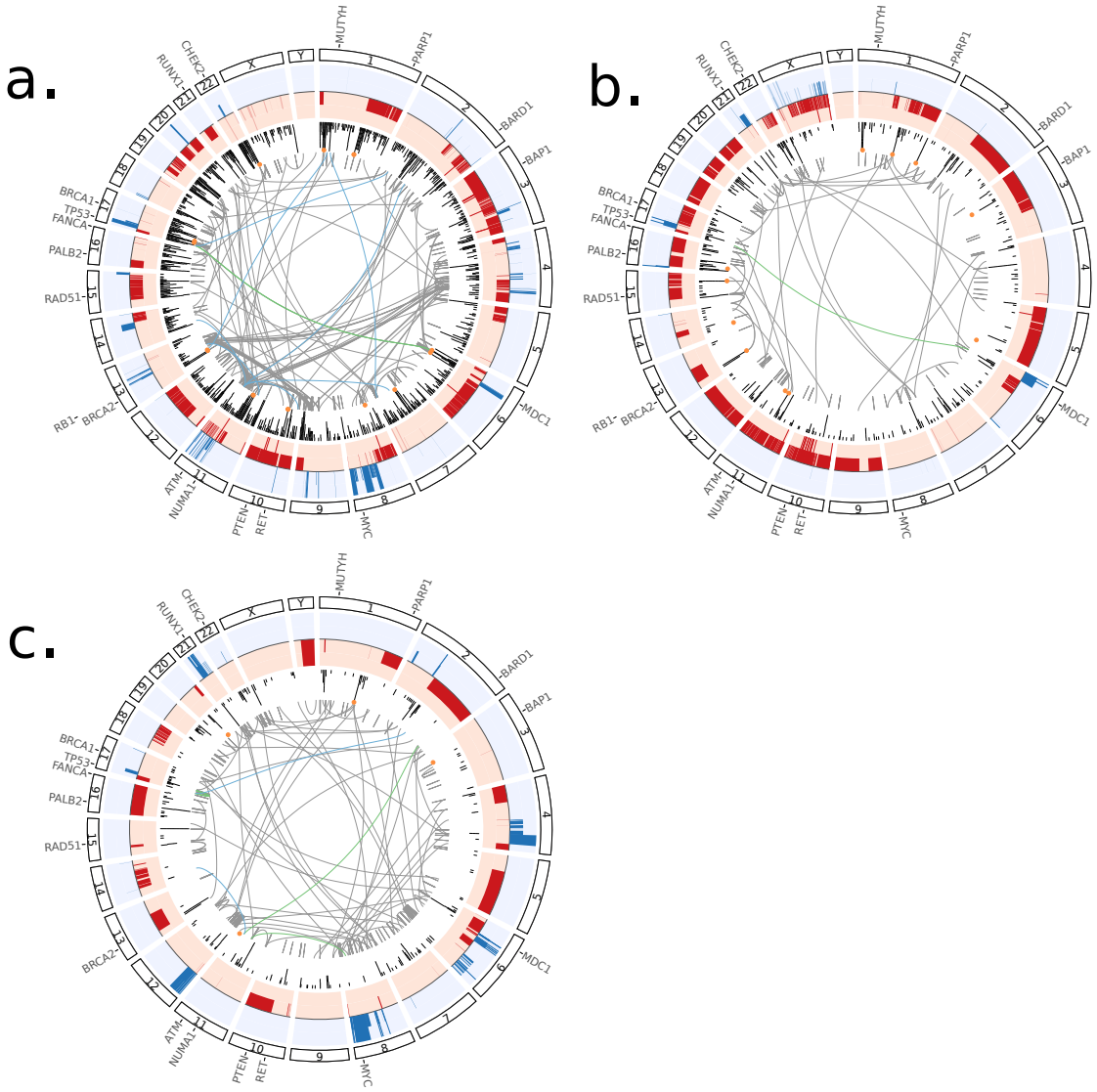


Figure S3 – Comprehensive Genomic Analysis of Osteosarcomas. Mutation burden of the tumors 128 (a.), 224 (b.) and 242 (c.) depicted as Circos plots. The outermost layer displays chromosomal orientation and the locations of osteosarcoma driver genes. The next inner layers depict chromosomal losses, gains and SV densities, respectively. Putative sites of kataegis are highlighted in orange, chromosomal gains are in blue and losses in red. The two innermost tracks depict inter- and intra-chromosomal rearrangements (grey) and fusion genes (blue and green).

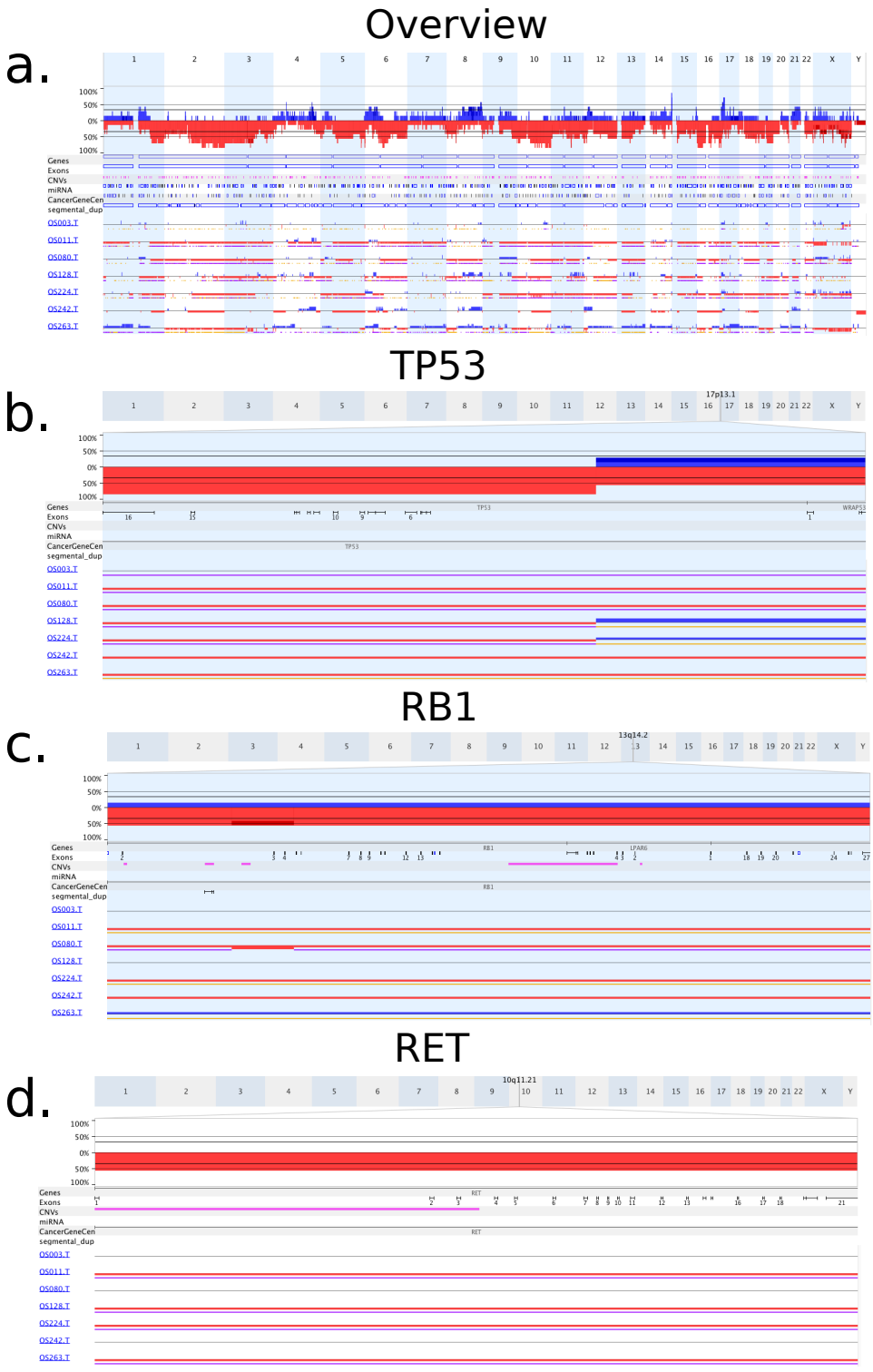


Figure S4 – Copy-Number Profile of *RET*-mutated Osteosarcomas. Copy number profiles as calculated from whole-genome sequencing data by Nexus Copynumber are shown for the whole genome (a.); the *TP53* tumor suppressor gene (b.); *RB1* Retinoblastoma 1 (c.); and the *RET* proto-oncogene. Gains: blue, losses: red, loss-of-heterozygosity: yellow, allelic imbalance: purple

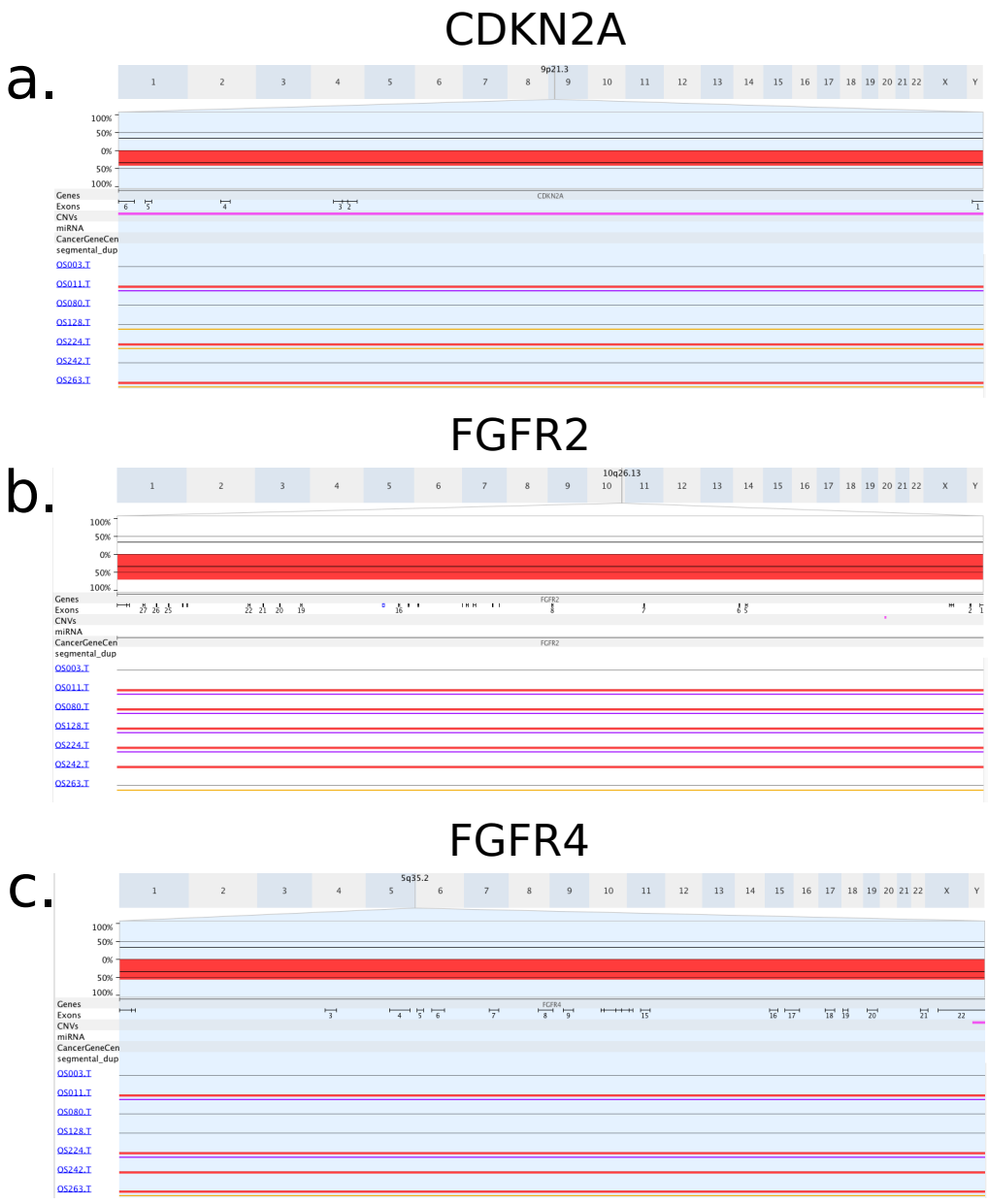


Figure S5 – Copy-Number Profile of *RET*-mutated Osteosarcomas. Copy number profiles as calculated from whole-genome sequencing data by Nexus Copynumber are shown for *CDKN2A* Cyclin Dependent Kinase Inhibitor 2A (a.); *FGFR2* Fibroblast Growth Factor Receptor 2; *FGFR4* Fibroblast Growth Factor Receptor 4 (c.). Gains: blue, losses: red, loss-of-heterozygosity: yellow, allelic imbalance: purple

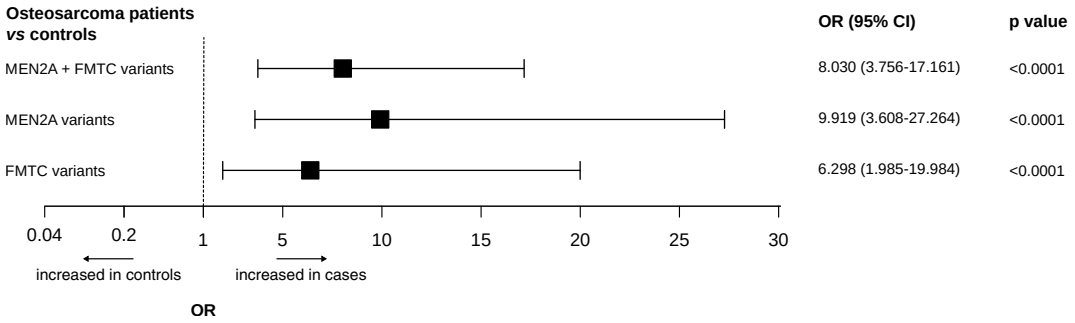


Figure S6 – Cumulative odds ratio of *RET* mutations in the osteosarcoma cohort versus normal population. The prevalence of *RET* germline mutations in the osteosarcoma cohort was compared to the ExAC population comprising of 53105 individuals (excluding individuals from TCGA studies). Cumulative odds ratios were calculated for mutations grouped according to the American Thyroid Association. MEN: multiple endocrine neoplasia type 2A, FMTC: familial medullary thyroid carcinoma. Error bars represent upper and lower 95% confidence intervals. P values were calculated using Fisher's exact test.

Patient					Mutation				Clinical Phenotype		
ID	Gender	Age	Ethnicity	FH	DNA	Protein	VC	Domain	Tumor Histology	Location	Last Follow-up
OS 011	M	16	Caucasian	None	c.653C>G	p.Pro218Arg	VUS	CLD2	Osteoblastic OS	Pelvis	Died, 18 years
ISKS	F	28	Chinese	None	c. 874G>A	p.Val292Met	PM	CLD3	OS, NOS	Pelvis	NA
OS 224	F	11	Caucasian	None	c.961G>A	p.Gly321Arg	PM	CLD3	Fibroblastic OS	Tibia	Alive, 17 years
ISKS	F	29	Caucasian	None	c.1083C>A	p.Asn361Lys	PM	CLD3	Periosteal OS	NA	NA
ISKS	M	31	South Asian	None	c.1141T>C	p.Phe381Leu	VUS	CLD3	Chondroblastic OS	Pelvis	NA
OS 128	F	11	Caucasian	None	c.1165C>T	p.Leu389Phe	VUS	-	Fibroblastic OS	Tibia	Alive, 20 years
OS 003	F	15	Caucasian	None	c.1807A>G	p.Lys603Glu	PM	CRD	Osteoblastic OS	Femur	Alive, 22 years
OS 080	F	15	Caucasian	None	c.1858T>C	p.Cys620Arg	PM	CRD	Osteoblastic OS	Femur	Alive, 21 years
OS 053	M	15	Caucasian	None	c.1858T>A	p.Cys620Ser	PM	CRD	Osteoblastic OS	Femur	Alive, 22 years
OS 263	M	34	Caucasian	MEN2	c.2671T>G	p.Ser891Ala	PM	TK	Small cell OS	Femur	Alive, 34 years
OS 242	M	14	Caucasian	None	c.3112A>G	p.Tyr1038Ala	VUS	-	Small cell OS	Tibia	Alive, 19 years

Table S1 – Osteosarcoma patients with germline RET mutations Osteosarcoma patients with pathogenic sequence variants or variants of unknown significance in *RET*. Abbreviations: FH: family history, VC: variant classification, F: female, M: male, CLD: cadherin-like domain, PM: pathogenic mutation, NOS: not otherwise specified, CRD: cysteine-rich domain, TK: tyrosine kinase domain, VUS: variant with unknown significance

Discussion and Outlook

3.1 *TP53* Intron 1 Hotspot Rearrangements are Specific to Sporadic Osteosarcoma and Can Cause Li-Fraumeni Syndrome

A hallmark of OS is the high amount of chromosomal instability resulting in marked intertumoral heterogeneity and a high abundance of structural variations. In an attempt to characterize this genomic chaos, we sequenced four pretherapeutic tumors with whole-genome DNA-PET, a method which is sensitive for copy number neutral- and copy number changing-rearrangements (see section 1.3.6 on page 13). We identified between 289 and 434 somatic SV per tumor, confirming the expected complexity. We found *TP53* to be the most recurrently affected gene: One tumor had a deletion of the whole gene, one tumor showed a balanced translocation between *TP53* and chromosome 1, one tumor had an inverted insertion from chromosome 6, and the last tumor had multiple complex inter-chromosomal translocations with chromosomes 1, 5, and 6 serving as partner sites. Complex rearrangements of this sort entail multiple breakpoints in the affected gene. Using a specialized bioinformatics pipeline and Sanger sequencing for validation, we reconstructed the total of seven *TP53* breakpoints in the tumors at base pair level and observed a clustering in intron 1. However, it is important to note that, while the *TP53* breakpoints were recurrently located within this hotspot region, the rearrangement partners differed from tumor to tumor which complicates the detection of such events. We further found *TP53* intron 1 rearrangements in the shape of a 445 kb inversion and a 2.5 kb deletion co-segregating with cancer in a four-generation LFS family.

3.1.1 What Is the Frequency of *TP53* Intron 1 SVs in Osteosarcoma?

From the four pretherapeutic primary OS we sequenced with DNA-PET (discovery cohort), all cases showed rearrangements involving the *TP53* gene and all *TP53*-internal breakpoints were located in intron 1. We validated our findings in two independent cohorts of primary pretherapeutic OS with FISH and SNP-based copy number arrays. First, FISH on a tissue microarray (TMA) of formalin fixed paraffin embedded (FFPE) samples revealed rearrangements in 11% of cases (23 of 215). In the second cohort we found rearrangement breakpoints in intron 1 as indicated by copy number transitions in 32% cases (23 of 73). Notably, the *TP53* locus was affected by copy number changes in 54 of 73 (74%) cases of this cohort. Therefore, intron 1 rearrangements, while not the only type of event, do contribute substantially to *TP53*-altering events. During the time of publishing these results, Chen and colleagues also reported on recurrent intron 1 rearrangements in 45% of their OS cohort as identified by whole-genome sequencing [96]. The differences in the observed frequencies can be explained by varying sensitivities and limitations of the techniques used. FISH has a relatively low resolution due to the size of the BAC clones (159 and 169 kb in our assay) and because of optical limitations of fluorescence microscopy. Copy number arrays have a higher resolution but fail to detect copy number neutral events, such as inversions or balanced translocations. This limitation is of importance, since three out of four DNA-PET sequenced tumors carried copy number neutral intron 1 SV, and the inversion in the Li-Fraumeni family only resulted in a small deletion which is close to the detection limit of copy number arrays and undetectable by FISH. It is therefore possible that the actual frequency of intron 1 rearrangements is higher than predicted by our analyses. Even though sequencing does provide base pair-resolution, difficulties can arise in genomically complex loci or in regions with sequence repeats. As discussed in section 3.1.2, intron 1 is rich in repetitive elements and all breakpoints we observed were located in such regions. Therefore, we used DNA-PET sequencing for our initial discovery cohort, which is is particularly designed to span such problematic regions, hence avoiding mapping artifacts and increasing the sensitivity for SV detection in repeat regions. Chen and colleagues sequenced their discovery cohort with a conventional paired-end approach, which might explain the higher intron 1 SV frequency in our discovery cohort. Intrinsic cohort properties or a selection bias could also explain inter-study differences: we only included pretherapeutic biopsy samples, while the sample set of Chen and colleagues also included metastases. Taken together, our data suggest that *TP53* intron 1 rearrangements are the most recurrent rearrangements in OS and occur with a frequency of at least 11-32%. This highlights the frequently overlooked role of copy number neutral structural variants in tumorigenesis, alongside the more studied SNV and copy number alterations.

3.1.2 The Breakpoint Sequence Context of *TP53* Intron 1 SVs May Not Be Compatible With Common Rearrangement Mechanisms

DNA-PET followed by Sanger sequencing allows the identification of rearrangement breakpoints and its sequence context at base pair resolution. All seven *TP53* breakpoints we observed were located in similar repetitive elements; six breakpoints in long interspersed nuclear elements (LINE) repeats and one in short interspersed nuclear elements (SINE) repeats. It has been reported that repetitive DNA sequences contribute to rearrangement susceptibility [111]. Such repetitive sequences make up ~50% of the human genome [74] and consist of transposable DNA elements, including the aforementioned LINE and SINE, simple sequence, tandem repeats, and low-copy repeats, such as segmental duplications. Repeat-dependent rearrangements seem to mainly result from a mechanism termed nonallelic homologous recombination (NAHR) [111]. In this process, paralogous sequences are paired and crossover, resulting in deletions, inversions and duplications. While most NAHR mediated events are intrachromosomal, interchromosomal NAHR does exist [112, 113]. Contrarily to our observations, NAHR is mainly implicated in recurrent rearrangements where size, genomic content, and rearrangement partner sites are identical between unrelated individuals [114] and the current understanding of NAHR does not explain the breakpoint-flanking sequence duplications we observed.

Other characteristics of the *TP53* intron 1 rearrangements we described also do not seem to be fully compatible with known mechanisms of structural variations. They include a) low sequence homology between intron 1 and the partner sites, b) occurrence of interchromosomal and complex rearrangements, c) recurrence of *TP53* intron 1 but varying partner locations, and d) duplication of the breakpoint-flanking sequence of up to 555 bp. For instance, homologous recombination based mechanisms require several hundred basepairs of uninterrupted sequence identity between rearrangement partners [115]. In microhomology-mediated break-induced replication (MMBIR) or fork stalling and template switching (FoSTeS) mechanisms, a nick on the template strand during replication causes stalling or collapse of the replication fork, which is later resolved by DNA template switching based on microhomology. MMBIR has been proposed to be underlying most non-recurrent disease-causing structural variants [116]. Interchromosomal rearrangements between distinct replication forks due to these template switching mechanisms do occur and multiple template switches can further result in complex rearrangements [117]. While some of our observed breakpoints include a few basepairs of homology, these regions are smaller than the 5-25 bp typically required for MMBIR [118] and breakpoint sequence duplication is not common in MMBIR.

Another known rearrangement mechanism is the non-replicative non-homologous end joining (NHEJ), which plays an important role in the immune system's V(D)J recombination, where it generates antibody diversity. In NHEJ, the ends of double-stranded DNA break (DSB) are first modified by addition or deletion of bases to make them compatible, and subsequently ligated [119]. These modifications leave a "scar" in the DNA sequence at the site of rejoining. A particularly well studied disease where NHEJ is pathogenetically involved is dystrophinopathy, caused by rearrangements in the *DMD* dystrophin gene. The DMD protein is crucial for normal muscle function and structure and therefore its loss causes muscle weakness and the loss of walking ability. Introns 45-50 of *DMD* contain rearrangement hotspots, similar to intron 1 in *TP53*. In contrast to the complex rearrangement events we observed in OS, in the *DMD* gene, NHEJ predominantly causes deletions [120], with breakpoint-surrounding sequence duplications of up to 25 bp [120]. Another similarity between *DMD* and *TP53* intron 1 rearrangements is the enrichment of repetitive elements such as LINE and SINE. Nevertheless, the larger length of breakpoint sequence duplications and the complex nature of the rearrangements we observed do not fit the classical understanding of NHEJ.

3.1.3 The *TP53* Intron 1 Rearrangement Mechanism May Be Transcription Dependent

Data from the Encyclopedia of DNA Elements (ENCODE) based on DNAse hypersensitivity sequencing and chromatin immunoprecipitation-sequencing (ChIP-seq) histone acetylation experiments [121] show strong signals of open chromatin and active enhancers, respectively, 2.5 kb upstream of the breakpoint cluster region, as well as within 10 kb of six out of seven partner breakpoints in the sequenced tumors and in the LFS family. It is therefore tempting to speculate that the rearrangement mechanism we observed is based on active transcription of *TP53* and the rearrangement partners. Supporting this hypothesis is the fact that our preliminary RNA sequencing data (unpublished) finds the two genes overlapping the breakpoints in sporadic OS, *SUCO* and *ENPP1*, to be expressed in primary OS cell lines. The gene adjacent to the partner breakpoint in the LFS is *HES7*. Strikingly, all three genes are involved in bone development. *HES7* is a member of the Notch pathway and mutations can lead to spondylocostal dysostosis, an axial skeleton growth disorder [122]. *ENPP1* encodes the plasma-cell membrane glycoprotein 1, and its expression regulates bone mineralization [123]. Plasma-cell membrane glycoprotein 1 (PC-1) expression in turn is regulated by the growth factors transforming growth factor beta (TGF-beta) and fibroblast growth factor (FGF) [124, 125], which are abundantly expressed in osteoblasts [126]. In mouse, *Suco*, also

known as Osteopotentia (*Opt*), has been found to be crucial for osteoblast function and bone formation [127]. Since OS prevalence is correlated to the pubertal growth spurt, it seems likely that these genes are actively transcribed at the time of tumorigenesis and therefore may contribute mechanistically to *TP53* intron 1 rearrangements.

Active transcription as a contributor to genomic rearrangements has been observed in a genome-wide analysis of SV breakpoint distributions in some cancer types [128]. A hypothetical, transcription dependent rearrangement mechanism is depicted and discussed in detail in supplementary figure S16 of the original publication [129]. In short, during active transcription of a genomic locus, the respective template DNA sequence is unwound by the RNA polymerase elongation complex, forming a transcription bubble [130]. If two single-stranded DNA break (SSB) occur on different strands and positions in the bubble, a DSB would result, with a stretch of unpaired overhangs. The DNA repair machinery may fill in the complementary nucleotides, producing blunt ends which can be fused together with other DSBs. During the interphase, sites of active transcription are located in specialized compartments [131]. The close proximity between transcribed strands might promote false re-assembly of blunt ends, leading to rearrangements. While explaining our observations, to confirm this hypothesis more mechanistic insights, for example about the size of transcription bubbles at the affected loci and the relationship to the duplicated regions, would be necessary. Also, the ligation mechanism which fuses the two blunt ends remains to be elucidated.

3.1.4 What Are the Functional Consequences of the Intron 1 Rearrangements in the Four Sequenced Tumors?

Genomic rearrangements can result in fusion proteins. The BCR-ABL fusion caused by a translocation in chronic myeloid leukemia is the most prominent example of this phenomenon [132]. It is thus tempting to speculate that the recurrent *TP53* rearrangements produce gene fusions, in which an oncogene is put under the control of the *TP53* promoters. However, our data from the limited number of five cases does not support this hypothesis. In two of the four initial tumors sequenced by DNA-PET, the *TP53*-affecting arrangements resulted in gene fusions. One case putatively produces a *TP53-ENPP1-TP53* fusion transcript, which is lacking a start codon and which is truncated by the stop codon in *ENPP*1 exon 25 and therefore is unlikely to translate into a protein. The second case results in a *SUCO-TP53* fusion, in which the exons 1-15 of *SUCO* are fused upstream of the *TP53* exons 2-12. The chimeric structure of the fusion protein and the membrane localization of SUCO may prevent the physiological function of the p53 protein. Due to the lack of RNA from these patient samples, we were not able to investigate the expression levels of the predicted fusion transcripts.

DNA sequencing data for the remaining cases do not suggest gene fusions at the *TP53* locus. Rather, the rearrangements result in partial or complete loss of *TP53* expression, as is illustrated by our data of a four generation LFS family. There, the inversion ranging 445 kb upstream from intron 1 towards the centromere disconnects two of the three known promoters from the *TP53* gene body. The remaining third promoter located in intron 4, is left intact. This internal promoter is driving the expression of the shortened *TP53* isoforms Δ133p53 and Δ160p53. These isoforms are conserved across different species and in *TP53* homologues, which further underlines their biological relevance. Interestingly, these isoforms are reported to lack *TP53*'s typical proapoptotic activity [133]. Expression of the Δ133p53α isoform is deregulated in several cancers, including colorectal carcinoma, breast cancer, acute myeloid leukemia (AML), and melanoma; Δ133p53β has been shown to increase the risk of cancer recurrence and death in breast cancer patients [134, 135, 136]. Similarly, Δ160p53 isoforms promote enhanced cell survival, proliferation, invasion and adhesion *in vitro* [137]. Taking this into consideration, it would seem plausible that the intron 1 rearrangements contribute to OS tumorigenesis by shifting expression towards the Δ133p53 and Δ160p53 isoforms due to the removal of the promoters which generate the full-length protein. However, our real-time quantitative reverse transcription PCR (RT-qPCR) experiments in the affected LFS family revealed a 23-53% reduction of overall *TP53* expression in the blood, where the rearrangements were present in a heterozygous state, and 89-100% reduction in the paired tumors with LOH at that locus. This suggests that the rearrangements' functional consequence is not an isoform shift but loss of *TP53* expression which, considering the gene's function, will likely give rise to further genomic instability, in accordance with the classical model of osteosarcomagenesis (see figure 6 on page 18). Given the relatively small number of tumors in which we have characterized the functional consequences of *TP53* intron 1 rearrangements, we cannot rule out that an isoform shift contributes to tumorigenesis in other OS cases.

Other functional consequences need to be considered. Intron 1 of *TP53* overlaps with the transcripts of *WRAP53* and *HP53INT* (also known as *RP11-199F11.2*). *WRAP53* encodes three different transcripts by alternative transcriptional start site usage. $WRAP53\alpha$ is an antisense transcript to *TP53* and is required to maintain elevated endogenous p53 messenger RNA (mRNA) levels in OS cell lines, likely by stabilizing the mRNA through a $WRAP53\alpha/p53$ RNA interaction. The transcript is also necessary to increase p53 protein levels in response to DNA damage [138]. $WRAP53\beta$ encodes the WRAP53β protein, which acts as a scaffold during the assembly of double-stranded DNA break repair complex components, including 53BP1, BRCA1, and RAD51 [138, 139]. Loss of $WRAP53\beta$ might therefore impair the accurate repair of DNA lesions by homologous recombination repair (HRR). The function of $WRAP53\gamma$ is still elusive. In our described LFS family, the small 2'275 bp deletion in intron 1 contains the

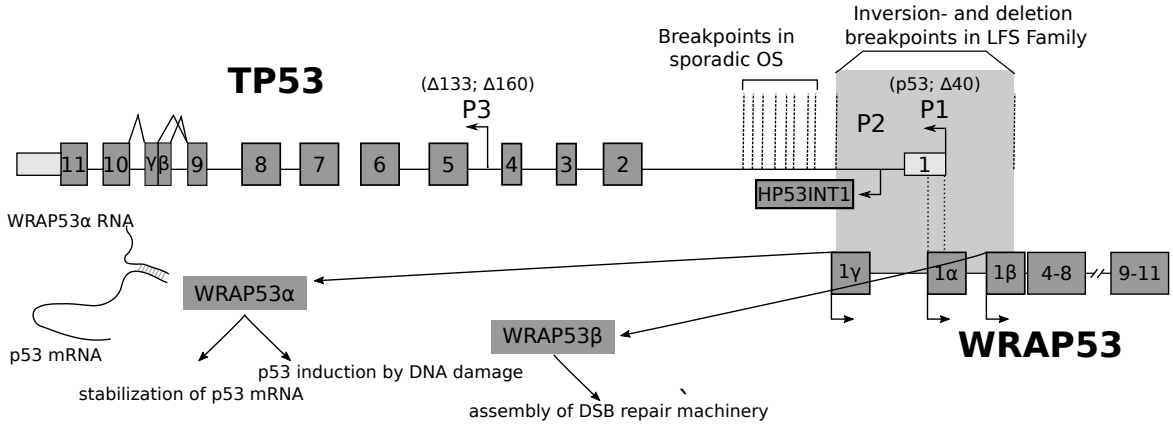


Figure 7 – Rearrangement breakpoints and overlapping transcripts in *TP53* intron 1. Clustering breakpoints identified by DNA-PET and targeted sequencing in four sporadic OS cases and a LFS family are displayed (dashed lines) in context of the *TP53* gene structure, whereby the non-coding exon 1 and parts of exon 11 are colored in lighter gray. Promoter 3 (P3) which drives expression of the full length p53 protein and Δ40p53 isoform and promoter 1 (P1) driving Δ133p53 and Δ160p53 isoforms are shown. Promoter 2 (P2) is driving expression of Hp53Int1, a transcript of unknown function. The loss of genetic material occurring in parallel to the 450kb inversion in the LFS family is depicted as gray shaded region. Observed rearrangements with intron 1 breakpoints disconnect the main promoter P1 from the gene body and either delete parts of the *WRAP53* transcript or potentially disrupt its regulatory elements acting *in cis*. Functions of the antisense transcript *WRAP*53α and the scaffold protein *WRAP*53β are indicated. Figure adapted from Henriksson *et al.* [140] and Marcel *et al* [141].

transcription start sites of both *WRAP*53β and *WRAP*53γ. The intron 1 breakpoints of the OS tumors from the discovery cohort are also located in close proximity to the promoters of *WRAP53* and *HP53INT1* and these rearrangements might therefore also displace the transcripts from their regulative elements (see figure 7 on page 103). It is thus possible that reduced *WRAP*53β levels contribute to tumorigenesis in sporadic OS and the LFS family, by further reducing *TP53* levels and its ability to induce DNA repair.

3.1.5 LOH as a Common Second Hit After *TP53* Intron 1 Rearrangements

A striking observation is the occurrence of copy number neutral LOH at the *TP53*-locus with concomitant loss of the non-rearranged allele in all five tumors we analyzed in the LFS family. Copy number neutral LOH are a common form of LOH, as was demonstrated in neurofibromatosis, glioblastoma, and hematolymphoproliferative malignancies [142, 143]. They are not detectable with classical cytogenetic methods and thus their importance in pathogenesis might have been widely overlooked until recently, when SNP-arrays and high-

throughput sequencing became more widespread. Copy number neutral LOH in tumor cells can change gene expression by duplication or deletion of epigenetic marks. If acquired after a mutational event, copy number neutral LOH can transform the mutation from a heterozygous to a homozygous state.

In our LFS family, the acquired LOH acted as a second hit, in accordance with Knudson's two-hit hypothesis of cancer evolution [144], by doubling the rearranged allele, while the wild-type allele is lost. This is in concert with our *TP53* expression data, which show a loss of expression in the tumors. Also, in the validation cohort of sporadic OS which we assayed with FISH, all of the 23 rearrangement-positive cases overwhelmingly showed two break-apart signals and no wildtype-like fusion signal. The absence of any pathogenic SNV in *TP53* as a second hits in all five LFS tumors, as well as in the four tumors sequenced by DNA-PET seems surprising, given the fact that most mutations inactivating the gene in LFS are missense mutations. This disparity may be explained by the generally high amount of chromosomal instability which is typical for OS. Further, intron 1 rearrangements can affect *WRAP53* transcripts. Loss of *WRAP53*α and β isoforms have been demonstrated to reduce p53 protein expression and DNA repair capabilities [138], and thus may further increase the likelihood of LOH as a second hit. In our copy number arrays, 75% of cases which show copy number transitions in *TP53* intron 1 also show copy number neutral or copy number changing LOH at the locus. Both observations support the idea that LOH after *TP53* rearrangements represent a common second hit in OS. Furthermore, several studies in mice have demonstrated that $TP53^{+/-}$strains show a different tumor spectrum (mainly osteosarcomas) than $TP53^{-/-}$mice, which mostly develop lymphomas, indicating that inactivation of the second *TP53* allele in OS is not restricted to humans. Remarkably, OS in $TP53^{+/-}$mice also show LOH at that locus [145].

Several mechanisms have been proposed to underlie somatically acquired copy-number neutral LOH events in cancer. Non-disjunction during mitosis can result in copy number neutral LOH affecting whole chromosomes. Subchromosomal events are thought to arise due to mitotic recombination between identical low copy repeats [146] or by a repair attempt of DSB induced deletions, whereby the lost region is restored using the remaining allele as a template. Subchromosomal copy number neutral LOH events have been found in OS by other studies and a significant increase in these events correlates with poor response to neoadjuvant chemotherapy [147]. Rearrangements and LOH in the *TP53* gene have been found in 40% of a recent sequencing study, supporting our observation of LOH as a frequent second hit [96].

3.1.6 Specificity of *TP53* Rearrangements in Osteosarcoma and Li-Fraumeni Syndrome

Using a customized *TP53* break-apart FISH with probes flanking the whole gene and ~35 kb of the upstream region, we demonstrated a rearrangement frequency of 11% in OS, but found no events in other bone-forming tumors, such as fibrous dysplasia, aneurysmal bone cyst and ossifying fibroma (n = 124). We also found no positive cases in any other tumor types unrelated to bone, including gastrointestinal-, genitourinary-, lung- or skin tumors (n = 966). While cases with *TP53* intron 1 rearrangements have been described before, our work represents the largest screen of OS and other tumor types for such rearrangements, and their striking specificity to a subset of OS was previously unknown. The reason for this specificity is currently not understood and requires further functional *in vitro* and *in vivo* studies. Nonetheless, our analysis of the sequence context of intron 1 rearrangements allows the hypothesis that the rearrangement mechanism is dependent on active transcription (see section 3.1.3). Interestingly, the genes located at or adjacent to the rearrangement partner sites are involved in bone biology and are expressed in OS cell lines. It is therefore possible that tissue-specific expression patterns trigger the intron 1 rearrangement mechanism with a higher frequency in bone than in other tissues, and replication stress caused by intensive bone-restructuring in the pubertal growth phase might further increase the possibility of its occurrence. It is well established that transcriptional activity can be fundamentally different between cell types. Cell lineage specific DNA replication properties, transcriptional processes and chromatin conformations might therefore ultimately be responsible for the remarkable specificity of the described rearrangements.

The LFS-family carrying a germline *TP53* intron 1 rearrangement did not only develop OS but multiple types of cancer, including adenocarcinoma, meningioma, astrocytoma, colon cancer, basal cell and squamous cell carcinoma of the skin. This does not directly contradict our observed specificity in sporadic tumors; it remains possible that, while rare outside of bone, this type of rearrangement occurs in the germline. The oncogenic effect, which in this family we demonstrated to be generic loss of *TP53* expression, evidently contributes to tumorigenesis in many other tissues.

3.1.7 The Role of Intron 1

Another remarkable feature of the *TP53* breakpoints we identified in sporadic OS and LFS is their clustering into a ~4'500 bp region in intron 1. While this first intron is the largest intron of the gene, the accumulation of breakpoints at this particular locus and the lack of breakpoints in other regions still is striking. We analyzed breakpoint sequence contexts and found direct, inverted, complementary, and mirror repeats within 50 bp of each breakpoint

in intron 1. These repeats can lead to the formation of non-canonical DNA conformations such as cruciforms and left-handed Z-DNA. These structures have been shown to impede DNA replication and in consequence can result in breaks [148]. However, repeats of this kind are also present in other regions of the gene where we do not observe any breaks. Their contribution to the accumulation of breakpoints in intron 1 therefore remains unclear. In section 3.1.3 we hypothesized that the rearrangement mechanism is transcription dependent. Compared to the other parts of the gene, intron 1 of *TP53* is located closest to the transcriptional regulators and two promoters. Depending on the transcription dynamics in the cell, the likelihood of a DNA break occurring at a given site might be a function of its distance to transcription initiation sites, leading to accumulation of breaks in the first intron. Intron 4 also contains a promoter. However, this third promoter is thought to be weak and the short variants whose expression it drives are not expressed at high levels in normal cells, which could explain the lack of rearrangements we see at this position. Besides the internal promoter, *TP53* intron 1 overlaps with the transcripts of *WRAP53* and *HP53INT1*. An alternative start site coding for the *WRAP53*γ splice variant is located just upstream of the breakpoint cluster and in addition, the promoter of *Hp53Int1* is located in intron 1. The expression of these two transcripts and the resulting open chromatin could contribute to the instability of the locus.

3.1.8 Clinical Utility

The eventual clinical consequences of our findings remain to be elucidated. Our data does not support any role of *TP53* intron 1 rearrangements as a prognostic or predictive marker; no statistically significant correlation was observed to clinicopathological markers, including survival, occurrence of metastases, or response to neoadjuvant chemotherapy. This can be explained by the fact that the main functional consequence of the rearrangements are a loss of *TP53* expression, which is frequently seen in OS as a result of large deletions encompassing the whole gene, and would thus result in comparable disease phenotypes (see section 3.1.4). In small biopsies, OS can be difficult to distinguish from histologic mimics, such as osteoblastoma, aneurysmal bone cysts or undifferentiated pleomorphic sarcoma, particularly in core needle biopsies [26]. Robust molecular OS specific biomarkers are still lacking. The remarkable specificity of intron 1 rearrangements we describe for a subset of OS could therefore be of diagnostic use in difficult cases. The diagnosis of Ewing sarcoma, for instance, is based on rearrangements of *EWSR1* besides pure morphology, but in contrast to OS, the frequency of these events is substantially higher (roughly 85%) [149]. Although *EWSR1* does not exclusively fuse with *FLI1* in Ewing sarcoma, the most common rearrangement partners can be detected by breakpoint-polymerase chain reaction (PCR), which is a lot less complex than DNA-PET sequencing. Alternatively, FISH is a technique which is well established in

most diagnostic institutions. However, the lacking sensitivity for some copy number neutral events might make its use in detecting *TP53* intron 1 rearrangements difficult in a routine diagnostic setting. In many pathology departments, tumors are increasingly analyzed by gene panel- or exome sequencing, which are, however, not sensitive towards all types of genomic rearrangements. In the future, more sensitive sequencing approaches are likely to be routinely used in diagnostic settings, which might render intron 1 rearrangements a useful marker nevertheless.

Besides sporadic OS, intron 1 rearrangements might have diagnostic implications in LFS. The classical LFS diagnosis is based on cancer occurrence patterns in the family, since only in 70% of cases the causal mutations in *TP53* can be identified. Current clinical routine testing in LFS patients includes Sanger sequencing of the *TP53* exons for discovery of point mutations and small InDels, as well as multiplex ligation-dependent probe amplification (MLPA) for copy number analysis. These methods are not suitable to detect copy number neutral rearrangements or rearrangements with only very small regional copy number changes, as we found in OS and the LFS family. Awareness of genomic rearrangements as a contributing factor to *TP53* aberrations in the 30% of LFS patients with unidentified mutations might lead to a more thorough and complete molecular diagnosis. This might prove invaluable for identifying at-risk individuals and for providing the appropriate genetic counselling. Another important clinical consequence of identifying LFS cases is based on the observation that 30% of germline *TP53* mutation carriers develop secondary tumors in the radiation field of a primary tumor [107]. If more LFS patients can be identified as such, appropriate radiation therapy decisions could be made to reduce the risk of secondary tumors. Due to the nature of the observed rearrangements, which only possess one recurrent rearrangement partner in intron 1, a comprehensive and straight-forward screen with PCR or MLPA is not possible. DNA-PET or custom capture sequencing to identify rearrangements like we used in our studies is labor and cost intensive and is not validated for clinical settings. That being said, novel long-read sequencing methods such as PacBio's SMRT and Oxford Nanopore are being developed, which generate continuous sequence information of fragments of up to 20 kb [150]. Such long reads have the potential to greatly facilitate SV breakpoint identification and validation, since they are able to span repeats and complex breakpoint regions. Alternative methods combining tiling-PCR with long read sequencing have already been proposed to leverage these technological advances for identifying breakpoints of SV with unknown rearrangement partners [151].

3.2 Exome Sequencing of Osteosarcoma Reveals Mutation Signatures Reminiscent of BRCA Deficiency

We characterized (SNV) and copy number states of a large OS cohort of 123 pretherapeutic samples, using next-generation sequencing and SNP-based copy number arrays. We initially sequenced the exomes of a discovery set of 31 OS biopsies with matched normal samples. After stringent filtering based on quality metrics and functional annotation, we found a median of 21 and seven somatically acquired and potentially pathogenic SNV and InDels per tumor, respectively. Based on these data, we further sequenced a replication cohort of 92 additional tumors using targeted panel-sequencing of 13 initial candidate genes and their 17 interaction partners (see table 3 on page 141). Combining the data, we found SNV in the previously reported OS driver genes *TP53, RB1*, and *ATRX* in 18%, 8%, and 9%. Further recurrently mutated genes included *MUTYH* and *RET*. We then analyzed copy number alterations in the 123 tumors and found highly altered genomes. Similar to SNV, recurring clonally disrupted driver candidates (either by losses or by disruptive gains) included *TP53*, *RB1*, and *ATRX*. Taken together, exome sequencing and copy-number array data showed that no monoclonal and thus potentially triggering event in a single gene seemed to be responsible for causing the majority of OS cases. However, 47% of tumors could be explained by *TP53* and *RB1* mutations, and clonal mutations in twelve other well-known cancer-driver or -susceptibility genes could explain additional 40% of tumors.

Interestingly, most tumors showed CNA in *BRCA1/BRCA2, PTEN, PALB2, ATM, CHEK2*, and in other genes involved in homologous recombination repair (HRR), a pattern of mutations similar to those seen in BRCA-deficient breast and ovarian cancers (so-called “BRCAness”). Of note, the majority of these CNA were subclonal. Prompted by these findings we hypothesized that BRCAness could represent a common functional trait in OS with potential therapeutic implications, since HRR deficiency is known to be specifically targetable by poly ADP ribose polymerase (PARP) inhibitors. We therefore compared the mutational signatures from our exome sequencing cohort with those reported in a pan-cancer analysis and found that 94% of cases (29/31) carried a signature which has been associated with *BRCA1/BRCA2* mutations in breast cancer. Furthermore, we found characteristic copy-number patterns as described in *BRCA1/BRCA2* mutated breast cancer in ~91% of OS, and 78% of OS fulfilled the criteria of a homologous recombination deficiency score based on sub-chromosomal LOH segments, which have been shown to correlate with homologous repair deficiency and a positive response to PARP inhibitors. First *in-vitro* experiments using the commercially available OS-cell line MG-63 showed a marked reduction in cell viability in response to the PARP inhibitor talazoparib.

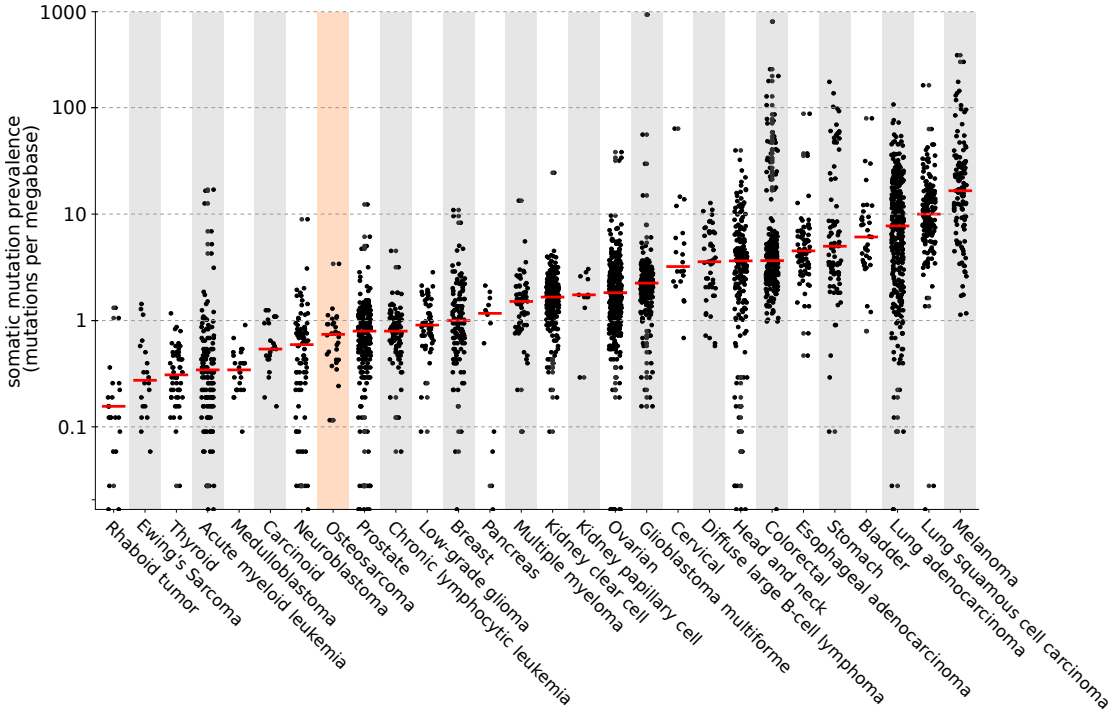


Figure 8 – Prevalence of somatic mutations across human cancer types. Tumor samples are represented as dots, red horizontal lines are the median numbers of mutations in the respective cancer type. The vertical axis (log scaled) shows the number of non-silent mutations per megabase. OS samples are highlighted with orange background. Figure and all data points for non-OS samples adapted from Lawrence *et al* [152].

3.2.1 What Can We Learn from Exome-Sequencing of 31 Osteosarcomas?

Osteosarcoma Is Characterized by a Modest SNV Burden

Exome sequencing of 31 OS with paired normal DNA revealed a median of 66.5 somatically acquired SNV (equal to 1.7 SNV/Mb), of which 28.5 were non-silent (0.72 SNV/ Mb). When these numbers are put into context of a pan-cancer study which involved 3'083 tumors of 27 subtypes, OS presents itself as a subtype with a modest SNV burden [152] (see figure 8 on page 109). While our bioinformatics pipeline differs from the pan-cancer study and therefore some numeric bias might have been introduced, the mutation numbers can still serve as a rough indicator for mutational load. Cancers induced by carcinogens such as tobacco and UV light show a ~10x higher SNV burden than tumors from our OS cohort, yet the observed SNV burden is still ~3x higher than that of other pediatric tumors (Ewing sarcoma: 0.28 SNV/Mb; AML: 0.33 SNV/Mb) and is comparable to that of prostate cancer (0.73 SNV/Mb).

Besides carcinogens, genomic instability e.g. due to microsatellite instability can lead to a high SNV burden. In microsatellite instability, genetic hypermutability results as a consequence of defective DNA mismatch repair and is commonly observed in colorectal and gastric cancers. Recently, microsatellite instability has gained importance as a biomarker for cancer immunotherapy: high mutational burden might translate into a bigger repertoire of neoantigens presented by the tumor cells, which can increase the discriminatory power of the immune system between healthy cells and neoplasms and thus improve response to immunotherapy [153]. Inhibition of the PD-1/PD-L1 immune checkpoint blockade, which in homeostasis is required to avoid autoimmune responses, is a promising targeted treatment strategy and antibody-based drugs for metastatic melanoma, non-small-cell lung cancer, and bladder cancer have already been approved. While our sequencing data shows some outliers with high SNV burden, in general OS seems to be microsatellite-stable and from this point of view shows no molecular rationale for immunotherapy approaches.

Osteosarcomas Show Somatic Mutation Signatures Associated with BRCA-Deficiency

Besides for screening of individual SNV, whole-genome and exome-sequencing data can be used to search for patterns in the mutational catalog of a tumor, which can reveal information about the underlying mutational processes. Prominent examples of such processes are reflected in specific DNA changes such as G>T/C>A transversions induced by tobacco carcinogens and C>T/CC>TT transitions formed after exposure to UV light [154, 155]. In signature analysis, these nucleotide changes are extended to mutation motifs, which include the surrounding trinucleotide sequence context. Each mutational motif and its frequency as found across all samples are shaped into a matrix. Mathematical non-negative matrix decomposition then attempts to break down the ensemble of mutations from each sample into a fixed number of specific mutational signatures: The two resulting matrices describe individual signatures and their contribution to the mutation burden present in each sample, respectively [156]. These signatures can metaphorically be thought of as stamps, consisting of mutation motifs that occur together and are likely the result of a distinct mutational process. Large sequencing efforts have led to the compilation and curation of mutational signatures and their proposed aetiologies [110, 157]. Selected signatures and their occurrence in cancer subtypes are illustrated in figure 9 on page 118. Using this approach, we found signature 3 to be present in 94% of OS exomes (29 of 31), which has previously been described in breast, ovarian, and pancreatic cancers. In the pan-cancer analysis this signature was furthermore associated with failure of DNA double-strand break repair by homologous recombination and was correlated with germline and somatic BRCA1 and BRCA2 mutations [110].

3.2.2 What Can We Learn from Copy Number SNP Arrays in 123 Osteosarcomas?

The Typical Osteosarcoma Genome is Highly Affected by CNAs

0.2 - 87% of the genomes of 123 OS samples were affected by CNA (median 51%). This is considerably more than in other cancer types and is in line with the historical association of OS with complex genomes. A study in which 3'131 tumors from 26 histologic subtypes were analyzed for somatic CNA revealed that on average 31% of the genome was amplified or deleted in a typical cancer sample and less than 0.5% in normal samples [158].

Osteosarcomas Are Not Driven by Focal Deletions or Amplifications of Tumor Suppressors or Oncogenes

Genomically unstable tumors often show focal CNA which affect genes of importance for tumorigenesis. Such small CNA are usually the result of selection for deletions of tumor suppressor genes or amplification of oncogenes during tumor evolution. Common examples are amplifications of *MYC* or *MDM2*, and deletion of *ATM* and *NOTCH1*. We observed copy number gains at the *MYC* 8p locus in 78% of cases, but these changes were mostly large segmental events. Nonetheless, in our dataset we found 80 focally altered regions which are recurrent (here defined as occurring in $\geq$15% of cases). Of these, only 2 encompassed genes with known association to OS: *DLG2* at locus 11q14 was altered with focal deletions in 24% of cases and recently reported to be mutated by genomic rearrangements in OS in 53% [96]. *DLG2* is a known tumor suppressor in *Drosophila* [159], but its role in human cancer is still unknown. We further found the paralogous cyclin-dependent kinase inhibitors *CDKN2A/2B* at locus 9p21 to be focally deleted. *CDKN2A* encodes the two proteins p16 and ARF. P16 is an inhibitor of CDK4-dependent inactivation of RB1 and loss of p16 can therefore phenocopy *RB1* loss [160]. Taken together, even though our data supports the common characteristic of a highly unstable genome, we do not find good evidence for recurrent focal CNA affecting cancer driver genes in the majority of the tumors.

Recurrently Disrupted Genes Are Involved in Homologous Recombination Repair

When looking into genes which are recurrently disrupted by large non-focal CNA (copy number losses or disruptive gains) we found the classical OS drivers *TP53* and *RB1* to be frequently affected. This is in line with previous results from our group and others [129, 96], which are based on smaller cohort sizes. Interestingly, most other recurrently disrupted genes

are functionally tied to the HRR pathway: *BRCA2*, (47%) *FANCA* (51%) , *BRCA1* (26%), *BAP1* (31%), *PTEN* (50%), and *PALB2* (43%). Of note, the last three genes were exclusively mutated by CNA, and not by SNV. Functional and potential clinical consequences are further discussed in sections 3.2.4 and 3.2.5.

3.2.3 Integration of Sequencing and copy Number Data: Clonal Mutations in 14 Driver Genes Can Explain the Majority of Osteosarcomas

To gain a more complete picture of putative driving events in the OS genomes, we combined sequencing and copy number data after excluding subclonal aberrations, since they are more likely to be passenger mutations present only in a subset of cancer cells, or have emerged after the tumor developed polyploidy and therefore after the initiation of genomic instability. We then assembled pathogenic mutations and disruptive copy number changes in genes with strong *a priori* implication in cancer. Using this approach we found that 47% of cases can be explained by *TP53* and/or *RB1* aberrations. Notably, the vast majority of *RB1* aberrations are based on genomic rearrangements manifested as CNA, whereas *TP53* showed both CNA and SNV. This is in line with the traditional view which historically attributed early OS tumorigenesis to *TP53* and *RB1* driver mutations [91].

Another 40% of cases could be explained by aberrations in the additional 12 genes *BRCA2, BAP1, RET, MUTYH, ATM, PTEN, WRN, MDC1, FANCA, ATRX, NUMA1*, and *RECQL4* (see Figure 3 of the published results in section 2.3 for a full overview) [161]. *RECQL4* encodes a RecQ DNA helicase and its loss of function has previously been found to predispose to OS as part of the tumor spectrum of Rothmund-Thomson syndrome [162]. *ATRX* regulates chromatin remodelling as a member of a protein complex and was recently found to be a prognostic marker in neuroblastoma [163]. *NUMA1* is a nuclear matrix protein involved in microtubule linking during mitosis and SNV in the gene are associated with increased risk for breast cancer [164], yet its role in OS is unknown. When exploring the gene ontology of the remaining driver genes we found that they are all functionally associated with genome integrity and HRR. The potential implications of this observation for therapy is discussed in section 3.2.5. In total, 87% of the 123 tumors from our cohort carried aberrations in the above 14 genes, whereby the majority are copy-number driven, except for *RECQL4* and *WRN*, which exclusively had missense mutations. Taken together, our findings suggest that, besides the known OS drivers *TP53* and *RB1*, multiple oncogenic pathways can initiate chromosomal instability in early stages of tumor development and thus complement the traditional model of OS tumorigenesis (see figure 6 on page 18).

3.2.4 BRCAness as a Unifying Trait of Osteosarcoma

The observation of an accumulation of mutations in key genes of the HRR pathway has been made also in other cancers. *BRCA1/BRCA2* mutations are important tumor susceptibility markers in familial breast cancer and lead to extreme levels of genomic instability in these tumors, characterized by chromosomal rearrangements and CNA. BRCA2 binds to the DNA recombinase RAD51 and guides it to damaged DNA, so HRR can be initiated; BRCA1 is controlling the signal transduction processes modulating HRR [165]. HRR uses DNA recombination to repair DNA double-strand breaks with high-fidelity. If HRR is impaired, cells need to rely on more destructive DNA repair pathways to repair DSB, such as NHEJ, which can result in genomic rearrangements. It was later discovered that mutations in *BRCA1/BRCA2* as well as other modulators of HRR do not only cause familial breast cancer, but are also widespread in sporadic tumors, where they promote similar complex genomes [166]. This led to the definition of BRCAness as a phenocopy of *BRCA1* or *BRCA2* mutations; it describes the situation in which an HRR defect exists in a tumor, in the absence of a germline *BRCA1* or *BRCA2* mutation [167]. Soon after its discovery, BRCAness gained special interest, since it is linked to increased sensitivity to platinum salts and targeted agents such as PARP inhibitors (this is discussed further in section 3.2.5 on page 115). We found several indicators for BRCAness in the 123 OS cases, which shall be discussed in the following.

Aberrations in HRR Genes

As discussed above, BRCAness by definition requires defects in genes of the HRR pathway. In our cohort of 123 OS, the accumulation of aberrations (mainly by CNA) in HRR associated genes is striking: a typical OS in our cohort carried 17 aberrations in these genes. Of note is that most of these events were subclonal; it is likely that this is a consequence of intratumor heterogeneity and that independent tumor subclones acquired different aberrations which affect the HRR pathway.

Somatic Mutation Signatures

Somatic mutation signatures can infer mutational processes (see section 3.2.1 on page section 110). A large pan-cancer study of 7'052 tumors identified a signature which was strongly associated with *BRCA1* and *BRCA2* mutations [168]. This signature, commonly referred to as signature 3, is characterized by a relatively even contribution of all 96 sequence motifs (see figure 9 on page 118). A whole-genome sequencing study of ovarian cancer later also found signature 3 in tumors with mutations in other HRR pathway components [169]. Our exome sequencing data showed this signature in 94% (29/31) of cases, which included tumors

where we could not identify any *BRCA1/BRCA2* mutations, but where other HRR genes were mutated. The process which leads to signature 3 specific base substitutions as a consequence of HRR deficiency is not well understood. It is thought that it might be a result of an attempt to compensate HRR deficiency by increased recruitment of error-prone polymerases to bypass DNA lesions [170].

A Specific LOH-Pattern

Abkevich and colleagues demonstrated that in ovarian cancer and 57 cancer cell lines, mutations in the *BRCA1/BRCA2* genes lead to an accumulation of LOH regions, which are sub-chromosomal, but greater than 15 Mb in length. Based on this, the group created a so-called HRD score (homologous recombination deficiency score), which is defined as the number of such LOH regions per genome. Similar to mutation signatures, they could show that an increased HRD-score is not only observable in *BRCA1/BRCA2* mutated tumors, but also when other HRR-related genes are affected [171]. Consequently, the HRD-score was proposed to serve as a marker for BRCAness. Several studies have shown that an HRD-score of $\geq$10 can predict treatment response to PARP inhibitors, a class of drugs which specifically target cells with HRR deficiency. This threshold is also reflective of HRR deficiency in tumors of the esophagus, lung, and prostate [172, 89]. It is therefore a valid assumption that the HRD-score is applicable to multiple cancer types. When we applied the analysis to our cohort, we found that 78% qualify as HRR deficient. However, some questions about the transferability of the HRD-threshold remain. When correlating the score to PARP inhibitor response in triple-negative breast cancer, Telli and colleagues derived the above threshold of $\geq$10 from the 10th percentile of the HRD score distribution in a set of 260 ovarian and breast cancer tumors, with aberrations in the *BRCA1/BRCA2* genes, or promoter methylation of the *BRCA1* or *RAD51C* gene. They then validated their prediction of BRCAness by measuring the response of the respective patients to PARP inhibitor treatments. At the current point we are not able validate the threshold for OS, since no appreciable number of patients has been treated with PARP inhibitors, but the HRD-score might still hold predictive value in OS nonetheless.

Large-scale Copy Number Transitions

Another predictor of BRCAness is based on tumor genome complexity and was established by Popova *et al.*, who analyzed large-scale copy number transitions (LST) in *BRCA1*-mutated basal-like breast carcinomas [173]. The group demonstrated that tumors with *BRCA1* deficiency are enriched in LST of $\geq$10 Mb in size. They were able to predict *BRCA1* inactivation by defining a threshold of 15 and 20 of such LST per near-diploid and near-tetraploid tumor

genomes, respectively, with high sensitivity. Ploidy is usually measured by flow cytometry and analysis of DNA content per cell. Since our study setup does not provide these data, we applied both threshold values to our cohort. This results in a prediction of 91% (threshold = 15) and 73% (threshold = 20) of BRCAness positive cases. As with the LOH based HRD-score, we cannot exclude tumor type specific threshold variations between breast tumors and OS, since no PARP inhibitor treatment data of OS patients are available at this point.

3.2.5 BRCAness as a Possible Targeted Treatment Strategy for Osteosarcoma

The Rise of PARP Inhibitors

In the past few years, inhibitors of PARP have drawn a lot of interest as a targeted treatment for tumors with HRR pathway deficiency. Amongst the six known primary DNA repair pathways, HRR is responsible to resolve DSB with high fidelity. Conversely, PARP-1 plays an important and rate-limiting role in SSB repair, where it identifies the site of DNA lesions and recruits the repair complexes of the BER pathway, including the APE1 endonuclease [174]. SSBs are common in any cell and can be caused by free oxygen radicals originating from normal metabolism, UV light, and replication errors. Several drugs have been developed which inhibit PARP-mediated SSB repair, either by impairing PARP-1's DSB sensing role in BER, or by trapping the PARP-1 protein bound to the site of DNA damage. When the replication machinery encounters an unrepaired SSB or a trapped PARP-DNA complex, DSB will arise. In normal cells, these are repaired by HRR. However, in cancer cells which do not have a working HRR pathway, the accumulating DSB will be repaired by low-fidelity NHEJ, which will eventually result in extreme chromosomal instability and cell death [175]. This constellation, where simultaneous inhibition of two pathways leads to cell death and inhibition of only one of these does not, is called synthetic lethality (see figure 10 on page 119).

The clinical utility of PARP inhibitors has been demonstrated in several studies [177], and the PARP inhibitor olaparib has received Food and Drug Administration (FDA) approval for the treatment of refractory *BRCA*-mutated ovarian cancer. Other drugs, such as veliparib and niraparib are currently in Phase III evaluation for treatment of breast cancer and recurrent solid tumors [178]. Importantly, a landmark study in prostate cancer has shown that sensitivity to PARP inhibitors is not restricted to *BRCA1/BRCA2* deficient tumors, but also extends to tumors with aberrations in HRR-related genes [179]. To this date, no OS-specific trial with PARP inhibitor has been conducted. However, we think that our findings of BRCAness as an abundant feature of OS provide the needed rationale for evaluating PARP inhibitors for the treatment of OS.

3.2.6 PARP Inhibitors and Osteosarcoma

Prompted by the results presented in this work, the sensitivity of several OS cell lines towards the PARP inhibitor talazoparib was tested [180]. The study demonstrated that OS cell lines were indeed sensitive to talazoparib and that the response correlated with their mutation signatures and the resulting degree of BRCAness. Also, the response correlated with aberrations in HRR-related genes, such as *PTEN, ATM, BAP1, BARD1* or *FANCA*, and a high HRD-score. Furthermore, another group recently reported good response of the OS cell lines HOS and MG-63 cells to olaparib [181].

Despite these first preliminary results in cell lines, open questions about the potential efficacy of PARP inhibitor in OS remain. First, our data suggests that OS is a tumor with considerable intratumor heterogeneity. This polyclonal nature is lost or altered in cell lines, where immortalization and prolonged culturing can lead to a selection of subclones, which were not necessarily the most prevalent in the tumor they are derived from. Furthermore, in prostate cancer patients which were treated with olaparib, most responders carried biallelic aberrations in BRCAness related genes [179]. In our data of 123 OS, mutations in HRR-genes were mainly heterozygous or subclonal CNA. It is conceivable that heterozygous mutations in multiple HRR-genes can synergistically add up to HRR deficiency. Another possibility is that CNA are present in homozygous states, but only in a subset of tumor cells due to intratumor heterogeneity. The methods used in our study are not able to discriminate between the two scenarios. To gain a more detailed picture, methods like multiregion sequencing or even single cell sequencing would need to be employed. Nonetheless, the majority of primary OS are sensitive to cisplatin treatment, indicating that the DNA repair mechanisms are impaired. If this property persists in recurrent and metastatic tumors, PARP inhibitors might represent an additional approach to exploit this weakness.

Indications for “compound-BRCAness“ were also found in a prostate cancer study, where several responsive patients only harbored heterozygous mutations in their sequenced BRCAness genes [179]. Ultimately, response to PARP inhibitor in OS will likely be governed by the nature of mutations, the players of HRR which are mutated, and the subclonal structure which underlies the intratumoral heterogeneity. Multiple PARP inhibitor companion diagnostic assays have been developed based on the genomic markers discussed above. However, so far only sequencing of *BRCA1/BRCA2* has received FDA approval. It remains to be elucidated how well these assays predict PARP inhibitor response to “compound-BRCAness“. Since PARP inhibitors are thought to target HRR-deficient cells independent of which BRCAness genes are mutated, a functional biomarker for HRR-activity would ultimately hold the most predictive value. An immunohistochemical marker surveying the ability of cells to accumulate RAD51 in the nucleus as response to DNA damage has been proposed [167]. A challenge

for the application of such assays in research and clinics is that in the basal state almost no RAD51 accumulates in the nucleus; DNA-damage needs to be induced by drugs or radiation first in order to assay the HRR response. An avenue to explore is would be to perform immunohistochemistry (IHC) of RAD51 after *ex vivo* DNA damage induction, e.g. in primary tumor cell cultures [182].

Challenges of PARP Inhibitor Treatment

Several mechanisms for resistance to PARP inhibitor have been identified [183]. Secondary mutations which reverse the pathogenic effect of SNV in the HRR-gene by restoring the open reading frame are found in ovarian cancers, but the overall frequency of this phenomenon is not clear. Since most HRR-affecting aberrations in our cohort are CNA, secondary SNV are not likely to be a major contributor to resistance, but compensation of heterozygous losses by uniparental disomy cannot be excluded. Moreover, loss of *TP53BP1* and *REV7* (also known as *MAD2L2*) can partially restore HRR in *BRCA1*- or *ATM*-deficient cells, but not *BRCA2*-mutant cell lines [184]. In our OS cohort, 35% and 8% show large-scale heterozygous- or subclonal losses of *TP53BP1* and *REV7*, respectively. However, in our samples BRCAness seems to stem mainly from aberrations in HRR-genes other than *BRCA1*, and of the ten cases with disrupted *ATM*, five have additional deletions in other HRR genes. The contribution of these known PARP inhibitor resistance mechanisms in OS might therefore be minor.

Another concern is the efficacy of PARP inhibitors in the treatment of metastatic or recurrent OS, since novel treatment approaches are generally tested in advanced disease stages first (see section 1.2.3 on page 7). A functional HRR-machinery is required for the repair of DNA damage induced by many currently used DNA-damaging chemotherapy agents and consequently, refractory tumors which survived the current standard-of-care MAP-regimen might only consist of subclones that are devoid of BRCAness properties or have acquired resistance to PARP inhibitors [167, 183]. Since the study presented here is based on pretreatment tumors, we cannot derive any information about the clonal composition of refractory tumors. However, a study to investigate this issue is currently in progress. Regardless of the outcome, PARP inhibitor treatment might still be a potential maintenance therapy candidate to support current chemotherapy regimens, since hyperadditive effects between PARP inhibition and cisplatin chemotherapy have been demonstrated in several cancers [185]. Therefore a reduction of chemotherapy dosages might be possible, which could greatly reduce side effects.

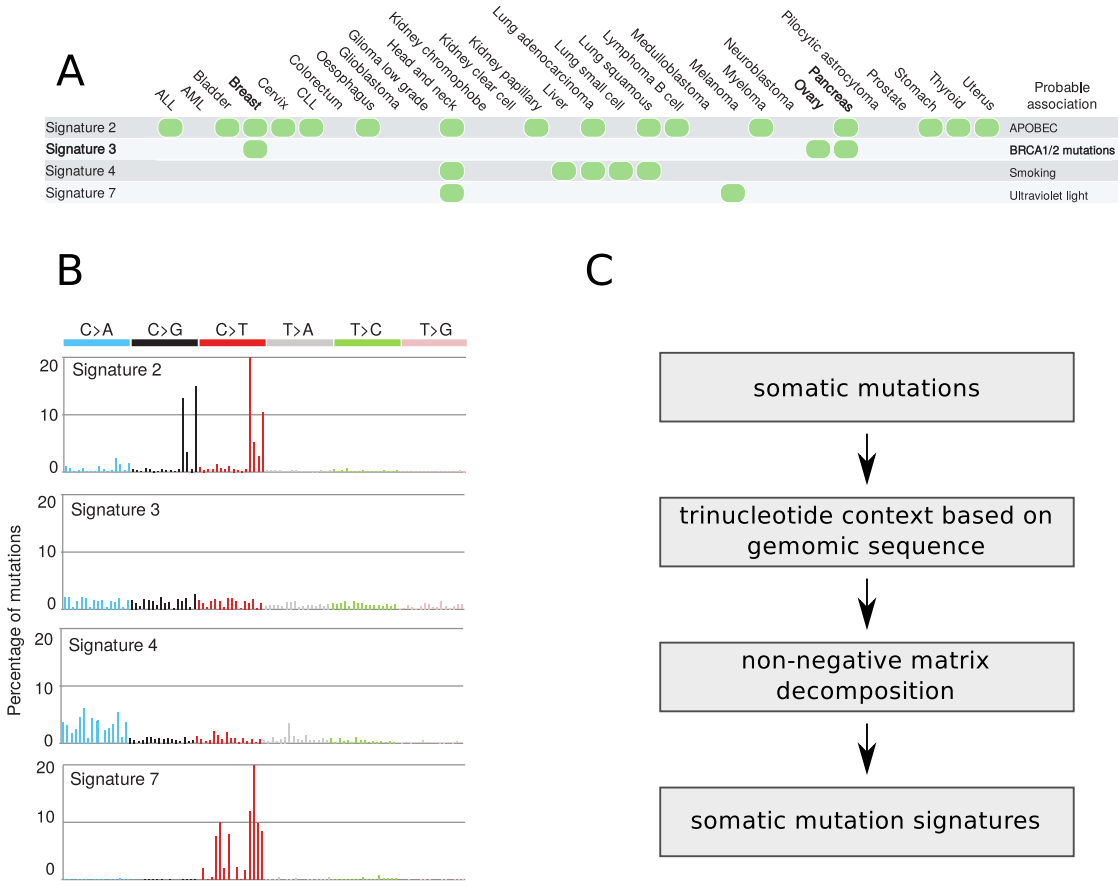


Figure 9 – Validated somatic mutation signatures and their presence in human cancer genomes. **A,** The presence of mutational signatures across human cancer types is depicted, based on a pan-cancer data set of 7'042 tumors. Probable mutation mechanisms are indicated. **B,** Each signature is displayed according to the 96 mutation motifs including the trinucleotide context of the mutated base. The contributions of substitutions to the mutational signatures are displayed in different colors. The mutation types are on the horizontal axes, whereas vertical axes depict the percentage of mutations attributed to a specific mutation type. All mutational signatures are displayed on the basis of the trinucleotide frequency of the human genome. **C,** The workflow from somatic mutations to mutation signatures is depicted. See text for more details. Adapted from Nik-Zainal *et al* [157].

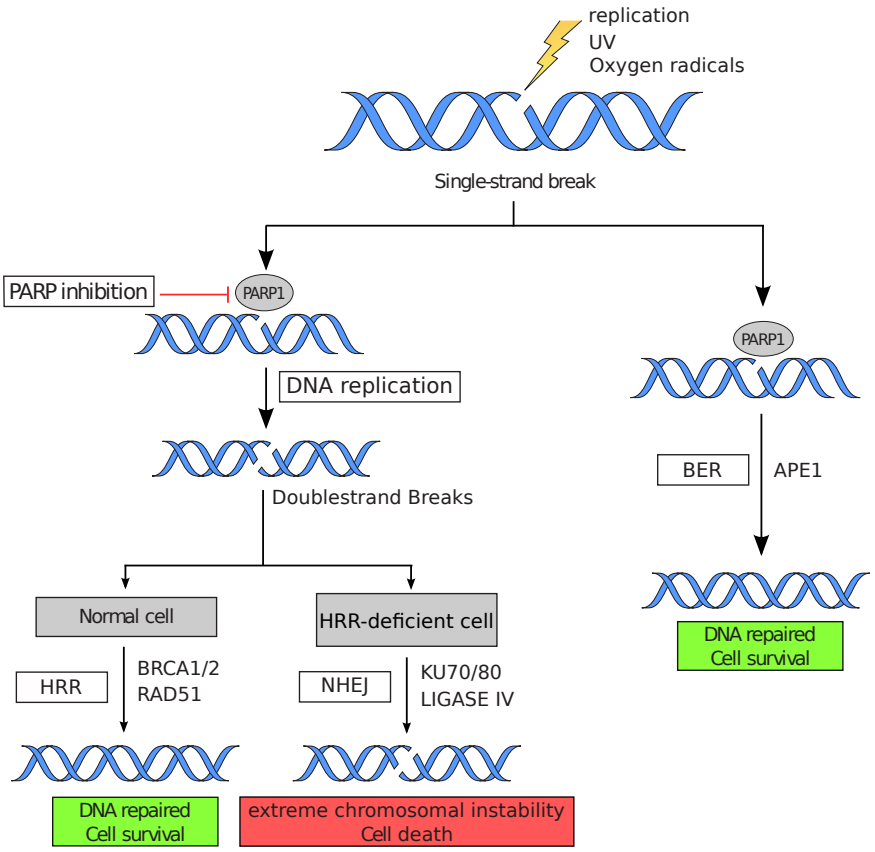


Figure 10 – Mechanism of PARP inhibition. The role of PARP inhibitors in synthetic lethality. Enzymes of the PARP family play a key part in the repair of DNA damage. Single-strand breaks occur frequently during each cell cycle and can be repaired. In particular, PARP1 binds to single-strand breaks in DNA and recruits other enzymes to repair the DNA damage through the base-excision repair pathway. Failure to repair single-strand breaks can result in double-strand breaks during DNA replication; thus, PARP inhibition can induce further DNA damage. In normal cells, double-strand breaks can be repaired through homologous repair. BRCA1, BRCA2 and RAD51 are key components of these homologous recombination repair pathways and therefore, *BRCA*-mutant tumors are inherently deficient in DNA repair. This vulnerability forms the foundation for selective targeted synthetic lethal therapy with PARP inhibitors in patients with BRCA-mutant breast cancer. The DNA damage that occurs after inhibition of PARP activity cannot be adequately repaired and the alternative, low-fidelity NHEJ repair pathway is engaged. This will eventually result in chromosomal instability, cell-cycle arrest, and subsequent apoptotic cell death. As DNA-repair processes remain intact in noncancerous cells, which generally retain at least one functional copy of both *BRCA1* and *BRCA2*, PARP inhibition is hypothesized to selectively kill cancer cells, sparing normal tissue. Abbreviation: UV; ultraviolet irradiation; PARP, poly(ADP-ribose) polymerase; HRR, homologous recombination repair; NHEJ; non-homologous end-joining, BER, base excision repair, APE1, Apurinic/apyrimidinic endonuclease 1. Adapted from Sonnenblick *et al* [176].

3.3 RET Germline Mutations and Susceptibility to Osteosarcoma

In an osteosarcoma cohort of 337 patients which we analyzed for germline mutations, we found seven cases with *bona fide* pathogenic *RET* mutations and four cases with variants of unknown clinical significance. By comparing these findings to the mutation frequency in the normal population as compiled by the Exome Aggregation Consortium (ExAC) consortium [186], we demonstrated that the cumulative odds of any American Thyroid Association (ATA)-classified multiple endocrine neoplasia type 2 (MEN2) type A or familial MTC-predisposing *RET* mutation were significantly higher in the OS cohort (odds ratio 9.12, 95% CI, 4.26 to 19.54, P<0.0001). *RET* germline mutations have been previously indicated in cancer development, yet the connection to OS is, to the best of my knowledge, a new discovery.

3.3.1 What Is the Role of *RET* in Normal Development?

RET (rearranged during transfection) is a proto-oncogene, which encodes a single-pass transmembrane receptor tyrosine kinase [187]. The RET protein consists of an extracellular domain for ligand-binding, a hydrophobic transmembrane region, an intracellular tyrosine kinase domain for downstream signalling [188], and a cadherin domain, supposedly for cell-cell interactions (see figure 11 on page 127) [189]. The protein is required for normal development of several cell types and processes, such as kidney induction, spermatogonial stem cell maintenance, neural crest cell migration, and development and maintenance of the central and peripheral nervous system [190]. Recently, RET was also implicated in the survival, expansion and function of haematopoietic stem cells [191]. High conservation between species from *Drosophila melanogaster* to primates underlines the relevance of RET in these developmental processes. The ligands of RET belong to the group of glial cell line-derived neurotrophic factor family ligands and promote the survival of certain neuron populations during normal development. Abnormally high GFL expression has been correlated with invasiveness and progression of pancreatic adenocarcinoma and oral cancer [192, 193]. Binding of the ligand to RET requires a co-receptor from a glial cell line-derived neurotrophic factor (GLF) receptor-α family. Varying affinity between members of the ligand and co-receptor families, as well as spatio-temporal expression patterns allow specific RET activation during development and homeostasis of mature tissues [194]. Upon binding of the co-receptor/ligand complex, the extracellular cysteines mediate the formation of RET dimers. Subsequently, specific tyrosine residues in the cytoplasmic domain are autophosphorylated which can then activate a plethora of downstream signaling cascades [190]. Multiple docking sites for different signalling complexes at phosphotyrosines pY1062, pY1096 and others have been described to stimulate RAS-MAPK, JAK-STAT3, and PI3K-AKT pathways [195, 196, 197] in response to RET activation.

3.3.2 What is the Role of *RET* in Cancer?

RET as a Somatic Cancer Driver

The transforming properties of *RET* were originally discovered in early experiments in which DNA from human lymphoma cells was transfected into the mouse NIH-3T3 fibroblast cell line, upon which *RET* was rearranged and activated [187]. Since RTK are involved in many physiological processes, its role in tumorigenesis and tumor progression has been further studied and demonstrated in many cancers. In humans, *RET* is located on chromosome 10q11.2 and consists of 53,325 bp and 21 exons. Early studies have shown that somatic mutations in the gene can be the driving events in thyroid cancer development. In sporadic medullary thyroid carcinoma, 30-50% of cases carry activating *RET* mutations [198]. The most common mutation is the M918T methionine to threonine substitution in the kinase domain, which can lead to decreased autoinhibition and increase of kinase activity and ATP binding [199]. More recently, recurrent *RET* rearrangements have been described as oncogenic drivers in papillary thyroid carcinoma [200], as well as in a smaller subset of non-small-cell lung cancer (KIF5B-RET) [201]. Other frequently mutated RTK include ABL and ALK in hematological cancers, and EGFR and HER-2 in non-small-cell lung cancer (NSCLC) and breast cancer, respectively [202].

RET in Familial Cancer Syndromes

Germline mutations in *RET* with a loss of function effect can lead to a congenital disorder called Hirschsprung's Disease (HSCR) disease, which is characterized by the absence of enteric ganglions in the gut and blocked peristalsis (often leading to a congenital megacolon). Causal alterations are dispersed over the whole gene but are mostly nonsense and missense mutations, which lead to loss of RET signaling [203, 204]. These HSCR mutations have been found to constitutively induce apoptosis in *RET* expressing cells, which can explain the absence of the affected ganglion cells in the disease phenotype.

Gain-of-function mutations in *RET* can lead to multiple endocrine neoplasia type 2, which has emerged as a paradigm for heritable activated oncogenes in cancer predisposition syndromes. MEN2 is an autosomal dominant familial syndrome and predisposes patients to a spectrum of endocrine tumors [205, 206]. The syndrome's genetic basis lies exclusively in germline mutations in the *RET* oncogene. While these are mainly heterozygous missense variations, a whole repertoire of these mutations has been described which leads to the distinction of three MEN2 subtypes: MEN2A, MEN2B and familial medullary thyroid carcinoma (FMTC). These subtypes, however, can show a wide variability in their phenotypes. MEN2A, the most common subtype, manifests itself by MTC and often includes development of pheochromocytoma, a

tumor of the adrenal gland. MEN2A mutations are mainly affecting extracellular cysteine residues. Gain-of-function mutations in codons 609, 611, 618, 620, and 634 together make up 98% of cases. Functional studies have shown that these mutations lead to a ligand-independent dimerization and therefore constitutive activation and downstream-signalling of RET [207]. MEN2B is characterized by mutations in the kinase domain (mostly M918T), causing receptor autophosphorylation, which can drive MTC formation in mice [208]. In addition, pheochromocytoma as well as developmental irregularities can occur which, together with the generally early onset in infancy or early childhood, makes MEN2B the clinically most severe subtype. Mutations causing FMTC can either affect the extracellular cysteines or the kinase domain. Current data and guidelines by the ATA hint towards a phenotypic continuum between MEN2A and FMTC [209]. Modifying variants have also been described. When occurring in *cis*, such mutations can amplify deleterious mutations and further complicate unequivocal genotype-phenotype correlations [210].

The distinction between HSRC-causing loss of function variants and MEN2-causing gain-of-function variants is not always trivial. So-called Janus mutations can lead to both phenotypes by constitutively activating RET but simultaneously hindering protein maturation and reducing the receptor expression on the cell membrane [211]. A publicly available and comprehensive list of mutations is maintained by the University of Utah and can aid in genotype/phenotype correlations, for both MEN2 and HSCR [212].

3.3.3 What Is the Potential Role of *RET* in Osteosarcoma?

As discussed previously, the function of *RET* as an oncogene is well known, yet a link of *RET* germline mutations to OS has, to the best of my knowledge, not been previously described. Generally, the role of RET in OS has not yet been extensively studied and thus not a lot is known about its involvement in osteosarcomagenesis. A phosphoproteomic screening has shown that RET and four other RTK are expressed and activated in the metastatic human OS cell lines 143B and LM7. Subsequent functional genomic screens using small interfering RNA (siRNA) demonstrated that RET contributes to the motility and colony forming phenotype of LM7 cells [213]. However, the underlying mutation leading to its activation has not been identified and the clinical significance and prevalence of these *in vitro* findings have not been followed up.

We identified seven cases with *bona fide* pathogenic *RET* mutations in our analyzed cohort of 337 OS patients. Like in MEN2, most of these mutations are located in the extracellular domain of the protein (six of seven). Of the four additional *RET* variants with unknown clinical significance, three mapped to the extracellular domain (see table S1 on page 95). This extracellular domain of RET consists of a cysteine-rich region, which is located adjacent

to the transmembrane portion of the protein, and a cadherin-like domain, which in turn is made up of four cadherin-homology repeats [214]. Mutations in the RET cadherin-like domains are common in HSCR and are thought to affect folding and membrane transport of the protein [215]. Nevertheless, the functional consequences of these mutations in cancer are not fully understood.

The cysteine-rich region dictates protein conformation and ligand binding. Amino acid substitutions of these cysteines lead to unpaired cysteine residues which can form intermolecular disulfide-bonds and therefore GLF ligand-independent RET homodimers. This dimerization results in constitutive RET signalling [190]. Alterations of this kind are found in virtually all MEN2A cases and are also frequently found in FMTC. RET signalling is organized around phosphotyrosines in the intracellular domain. Upon ligand binding and dimerization of RET receptors, these residues are autophosphorylated and allow direct interaction with signalling molecules such as SRC or STAT, or a broad range of adaptor proteins. This leads to the activation of various downstream pathways, including RAS-MAPK, JAK-STAT, and PI3K-AKT pathways. These cascades can promote cell growth, proliferation, survival and differentiation [190]. Alongside the pathogenic variants and variants of unknown significance (VUS) in *RET*, we identified a G691S polymorphism in 5 of the 11 cases (see figure S2 on page 90). Functional studies have shown that *RET*-G691S leads to an increased oncogenic activity of the S891A mutation and a trend towards earlier onset in patients. It has therefore been proposed that in MEN the status of this polymorphism should be evaluated for more appropriate follow-up decisions [216]. The validity of this observation in OS remains to be confirmed.

Particularly the activity of the PI3K/AKT pathway has been experimentally linked to the MEN2 phenotype and it was shown in cell culture studies that the oncogenic potential of the typical C634R MEN2 mutation is dependent on PI3K/AKT activation [217]. Interestingly, high activity of the PI3K/AKT pathway has been implicated in genomic instability, which constitutes a hallmark of osteosarcoma genomes. Phosphoinositide 3-kinase family kinases (PI3K) and their target AKT get activated by RTK. Activated AKT protects cells from apoptosis during the G2/M checkpoint transition in normal cell cycles. Since the G2 checkpoint normally prevents cells with DNA damage from undergoing mitosis, dysregulation of AKT can result in toleration and accumulation of DNA damage [218]. Findings in glioblastoma suggest that AKT signalling might also actively promote genomic instability by interfering with the recruitment of BRCA1, RAD51 and other DNA damage response proteins to the site of damage, and thereby repressing the accurate repair of DNA lesions by HRR. In this case, DNA breaks will be preferentially repaired by the low-fidelity NHEJ mechanism, which can result in rearrangements [219]. Importantly, negative regulation of the PI3K/AKT pathway is mainly achieved by *PTEN* in homeostasis [220] . This tumor suppressor gene is the second-most mutated gene after *P53* in solid tumor types and we have identified *PTEN* as a potential driver

mutation in 16% of our sequenced osteosarcoma cases. An other study found *PTEN* deleted in 43% of OS cases [100]. Furthermore, a transposon-based screen demonstrated that *PTEN, AKT2* and other PI3K/AKT pathway members can drive OS tumorigenesis in mice [221]. In addition to deletions of *PTEN*, constitutive activation of *RET* as implied by the mutations in our cohort might therefore represent an alternative mutational trajectory towards PI3K/AKT activation and accumulation of genomic lesions in a subset of OS. This hypothesis is supported by our observation that germline *RET*-mutated tumors result in genomically similar tumors to sporadic OS in terms of mutation burden, mutation signatures and methylated regions. Besides promoting genomic instability, typical *RET* MEN2 mutations have also been shown to increase proliferation in cell line experiments [222]. It is thus conceivable that *RET* mutations can contribute to OS tumorigenesis on both, an antiapoptotic level, i.e. by rendering the cell unresponsive to DNA damage during the acquisition of genomic aberrations, and repression of HRR as well as by directly increasing cell proliferation rates. Other modes of action also seem possible. Activated RET is known to induce overexpression of the cell cycle protein *c-MET*, which in turn has been demonstrated to transform human osteoblasts into an OS phenotype [223, 224]. The exact mechanism with which *RET* mutations contribute to osteosarcoma tumorigenesis remains to be elucidated. Experiments to establish direct functional evidence between our observed RET mutations and PI3K/AKT activation in OS still need to be conducted. However, in either case constitutively active RET signalling seems to be emerging as a common denominator with clinical potential.

3.3.4 How Can The New Knowledge About Osteosarcoma as a *RET*-associated Disease Benefit the Clinics?

By analyzing 337 osteosarcoma patients we identified seven patients with *bona fide* pathogenic germline variants and additional four patients with *RET* germline variants of currently unknown clinical significance, which corresponds to a frequency of 2% or 3.3%, respectively. Even though this can unarguably be perceived as a rare finding in a rare tumor, our results may still hold value for diagnosis and treatment of osteosarcoma, which shall be discussed hereafter.

Early Detection of At-Risk Patients

The frequency of germline *RET* mutations observed by us is similar to the frequency of pathogenic germline *TP53*-mutations, as determined in a large study of 765 osteosarcoma cases (3.1 %) [225], as well as to the frequency of germline *TP53* mutations in early-onset breast cancer patients (2%) [226]. For both cases, routine genetic testing has recently been recommended [227]. Given the similar prevalence between these studies and our data and genetic testing for *RET* mutations already being well established in clinics for MTC, it might be warranted in OS patients as well, independent of family history.

Furthermore, the pathogenic *RET* germline mutations we described in the OS patients have all been categorized as moderate by the ATA in respect of risk for developing MTC; for carriers of these mutations thyroidectomy in childhood or young adulthood is advised [228]. It is therefore possible that affected OS patients, if untreated, will develop thyroid carcinomas later in life. With the improved survival outcome in OS due to introduction of chemotherapy in the 1970s, it can be expected that more patients will eventually reach the age to develop variable late-onset MTC caused by moderate risk *RET* mutations as we found in OS. Early identification of these patients will allow adequate surveillance and treatment and enable predictive testing of at-risk family members.

RET as a Therapeutic Target

RET is a tyrosine kinase and as such can be targeted with small molecule inhibitors. Current RET-targeting drugs are compounds which were initially developed as inhibitors for other tyrosine kinases but which, due to high sequence homology in the functionally important ATP-binding and kinase motifs, are effective in decreasing RET activity as well (multikinase TKI). So far, no RET specific inhibitors have been approved for clinical use. Even though clinical data for unspecific inhibitors do show prolonged progression-free survival of metastatic or unresectable MTC, their cross-reaction with other tyrosine kinases, such as the VEGF receptor, can lead to a regression of normal blood vessels and associated thyroid dysfunction [229]. These adverse effects can prevent long-term treatment [198]. With non-small-cell lung carcinoma recently being added to the portfolio of *RET* related cancers [230], agents which target RET more specifically and efficiently are likely to emerge in the near future [231].

Since our data suggest that a subset of OS is caused by pathogenic germline *RET* mutations, targeting of RET is a potential treatment option which should be investigated in more detail. Even though no direct link between *RET* and OS existed so far, preliminary experiments with TKI have been done, prompted by the observation that several RTK are activated in OS. A study with seven OS cell lines and mouse xenografts suggest that the TKI sorafenib

exerts apoptotic and antitumor activity [232]. Interestingly, the murine xenograft models also showed a reduction in number and size of lung metastases upon sorafenib treatment. A dose-escalation study which assessed the efficacy of regorafenib, another multikinase RET inhibitor, against refractory advanced solid tumors showed preliminary evidence of antitumor activity in an osteosarcoma patient [233]. However, trials involving more patients and the integration of their *RET* mutation profile are needed to accrue more solid evidence. The different proposed contributions of *RET*-mutations to tumorigenesis, and varying degrees of intratumor heterogeneity might perturb therapeutic effects of RET inhibition. If constitutively activated RET functions as an enabler of early genomic instability by inactivating DNA-damage induced apoptosis and HRR repression, then treatment of OS with TKI at the stage of recurrence or metastatic spread may be belated since the tumor has already accumulated its transforming genomic aberrations. On the other hand, in tumors where dysfunctional RET signalling does not only induce chromosomal instability but also contributes to cell proliferation and metastatic spread at a later stage, targeting of RET might represent a new avenue for OS therapy.

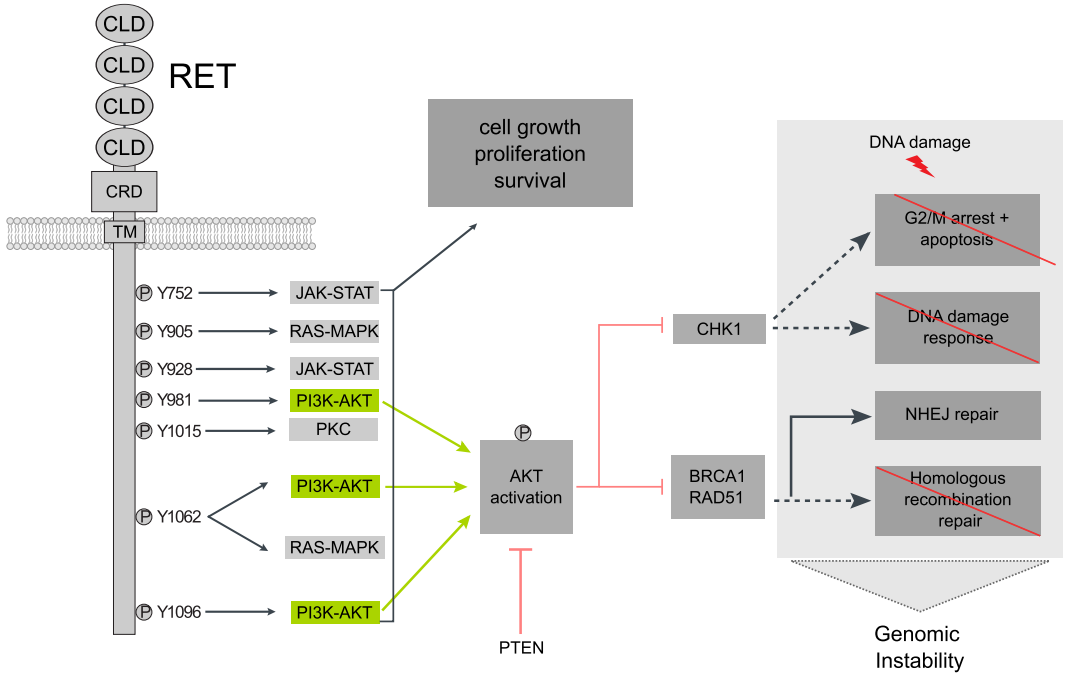


Figure 11 – The RET protein and its connection to the PI3K-AKT signalling cascade. RET signalling, either induced by ligand binding or activating mutations, is organized around phosphotyrosines in the intracellular domain. These signaling hubs can interact with various downstream pathways such as RAS-MAPK, JAK-STAT and PI3K-AKT. Phosphoinositide 3 -kinase family kinases (PI3K) and their target AKT can be activated by various receptor tyrosine kinases (RTK), including RET. Activation of the PI3K/AKT pathway by phosphorylation of AKT can protect cells from apoptosis during G2/M checkpoint transitions, which can lead to accumulation and toleration of DNA damage. AKT signalling can promote genomic instability by interfering with the recruitment of BRCA1, RAD51 and other DNA damage response proteins to the site of damage, thereby repressing the accurate repair of DNA lesions by HRR. See text for more details. CLD; cadherin-like domain, TM; transmembrane domain, CRD; cysteine-rich domain. Encircled Ps depict phosphorylation sites.

3.4 General Conclusion and Outlook

In this work, we combined state of the art genomic analysis methods with an extensive collection of pretherapeutic samples to gain a better understanding of the genomic landscape of OS, which could lead to therapeutic approaches that complement current treatment protocols. To this end we integrated data from SNP-based copy number arrays, gene panel-, exome, and whole-genome sequencing as well as the specialized DNA-PET method for detection of structural variations.

We identified recurrent rearrangements with breakpoints in intron 1 of *TP53* to be highly specific for OS. We demonstrated that these rearrangements lead to abrogated expression of *TP53* and that they represent the genetic basis for LFS in a family. Even though these findings do not offer a direct therapeutic target, they may hold value as a diagnostic marker for both OS and LFS, especially since in the future advanced long-read based sequencing techniques are likely to be routinely used in diagnostics. Additional functional studies are needed to elucidate the indicated transcription-dependent rearrangement mechanism we propose.

We further identified 12 driver genes which supplement the classical, *TP53* and *RB1* centric model of osteosarcomagenesis. Mutational signature analyses suggest that the majority of OS in our cohort are deficient in HRR ("BRCAness"). PARP inhibitors, which specifically target HRR deficiency show promising treatment results in other malignant tumors and we show that PARP inhibition significantly reduces viability in OS cell lines with BRCAness. Nevertheless, the high intratumoral heterogeneity of OS complicates a clear correlation between the mutations we identified in bulk tumors and the measured BRCAness. Multi-region or single cell sequencing (once the latter technique has matured more) could help in establishing clearer prognostic correlations. Also, the analyzed cohort consisted of pretherapeutic samples. Thus, the degree of BRCAness and the accompanying predicted PARP inhibitor sensitivity of refractory and metastatic tumors remain to be elucidated; an ongoing project is addressing this limitation of the current data.

Based on the findings in an index case with a pathogenic germline *RET*, we analyzed sequencing data 337 of OS patients and discovered predisposing *RET* mutations, which occur with a comparable frequency to germline *TP53* mutations in OS and breast cancer. There are ongoing collaborative efforts to create mouse models that reflect *RET* mutated OS. Therefore, future functional studies will need to shed light on the consequences *RET* mutations have on downstream signaling pathways such as PI3K/AKT. Such models will further allow the assessment of the possibilities of TKI in the treatment of OS. The cost of sequencing has

dramatically decreased in the past and will continue to do so in the future. This will enable individual pathology departments as well as cooperative research consortia to sequence a high number of OS and matching normal samples. It is therefore likely that more predisposing germline mutations of similar frequency than the ones we describe in *RET* will surface.

Materials and Methods

The methods described in this chapter have previously been published as parts of the presented publications [129, 161]. For the reader's convenience and clarity, some of the methods are addressed again and extended below.

4.1 DNA-PET Sequencing of Osteosarcoma

4.1.1 OS Patient Samples for DNA-PET Sequencing

For initial analysis of recurrent structural variations in OS, DNA of four pretreatment OS tumors and paired normal blood was obtained from the Biopathology Center (BPC) of the Childrens Oncology Group (COG). The COG cooperative group includes medical centers in the United States, Canada, Mexico, Australia, New Zealand and countries in Europe.

4.1.2 DNA-PET Library Construction and Sequencing

DNA-PET library construction from 1 to 4 kb fragments of genomic DNA was performed for OS DNA samples from the Children's oncology group, where enough DNA was available. Library construction was adapted from the protocol described by Hillmer *et al* [234]. 5-10 μg of genomic tumor DNA was randomly fragmented using HydroShear (Genomic Solutions, USA) and end-repaired using the DNA End-Repair kit (Epicentre Biotechnologies, USA). Long Mate Pair LMP CAP adaptors (Applied Biosystems, USA) were ligated to both ends. This adapter is missing the 5'phosphate on the shorter oligonucleotide and therefore the ligation product carries one nick in each strand. DNA constructs were size-separated on an agarose gel and 1-4 Kb fragments were selected and extracted (yielding in 150-500 ng) for further ligation with the biotinylated internal adapter. By a nick translation using DNA polymerase I, the nick

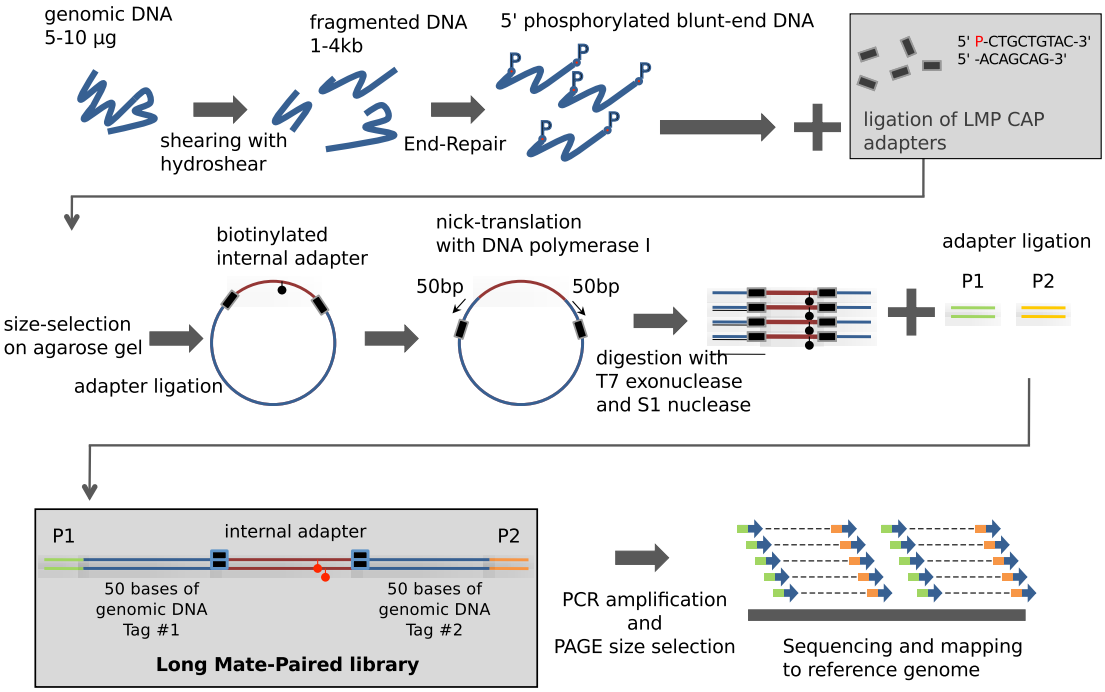


Figure 12 – DNA-PET sequencing library preparation. 5-10 μg genomic DNA (blue) was fragmented to a target size of 1-4kb and end-repaired. Long Mate Pair LMP CAP adaptors, which are missing a 5'phosphate (upper gray box), are ligated. After size-selection on an agarose gel the biotinylated internal adapter (red fragment) is ligated. The nick in the ligation product is pushed inwards using DNA polymerase I and The DNA-PET construct is released by T7 exonuclease and S1 nuclease digestion. The sequencing adapters P1 and P2 (green and yellow, respectively) are ligated and the final library (lower gray box) is PCR amplified, size-selected using polyacrylamide gel (PAGE) and sequenced. The reads are mapped to the human genome and discordant reads analyzed using a dedicated bioinformatics pipeline to identify SV candidates. Figure adapted from UC Santa Cruz [235].

was pushed from the LMP CAP adaptor in 5'direction into the genomic DNA fragment. The construct was digested by T7 exonuclease and S1 nuclease resulting in a cut at the nick position and release of the DNA PET constructs. SOLiD sequencing adaptors (Applied Biosystems, USA) were ligated to the PET constructs, which were then amplified for 15 cycles by PCR using the SOLiD PCR primers (Applied Biosystems, USA) and the final 225-425 bp libraries was extracted from a polyacrylamide gel. High throughput sequencing was performed on SOLiD sequencers (v3plus and v4) according to the manufacturer's recommendation (Applied Biosystems, USA). The resulting PET reads were mapped to the National Center for Biotechnology Information (NCBI) human reference genome build 37 (hg19).

4.1.3 Bioinformatics Workflow for SV Detection

Sequencing, mapping and data analysis were performed as described by Hillmer *et al* [234], with refined bioinformatics filtering as described by Nagarajan *et al* [236]. The short reads were aligned to the NCBI human reference genome build 37 (hg19) using Bioscope (Life Technologies).

Since not enough constitutional DNA was available to generate DNA-PET libraries, potential somatic SV were filtered out by subtracting any SV identified by paired-end sequencing in an additional 29 unrelated normal individuals (20 individuals analysed by DNA-PET and 9 individuals analyzed by other paired-end sequencing protocols. Sequence context and -homology of rearrangement partners were assessed by expert visual inspection to exclude false positive results. Somatic SV calls were validated by PCR and Sanger sequencing. Candidate SV selected for further analysis were validated to be somatic by breakpoint PCR in the paired normal sample.

4.2 Dual-Color Break-Apart FISH Assay for Detection of *TP53* Rearrangements In Osteosarcoma

The BAC clones RP11-1081A10 and RP11-107F4 (BACPAC, Children's Hospital Oakland Research Institute, USA) were selected for the *TP53* break-apart assay (see table 1 on page 133 for details). 1.5 μg of each BAC clone was labelled by nick translation (Nick Translation Kit, Abbott Laboratories, USA) for 15 mins at 15°C according to the manufacturer's instructions, using red or green fluorescently labelled dUTP nucleotides, respectively (Enzo Life Sciences, Switzerland). Labeled probes were precipitated with salmon sperm DNA and 3M sodium acetate (NaAc) (10% v/v) at -20°C and resuspended in nuclease-free H_2O. The probes were heated for 4 hours at 37°C to ensure complete dissolving. The in situ hybridization was performed according to routine protocols. Tumors were considered to have a rearrangement in *TP53* when at least 10% of cells showed clearly separated green and red hybridization signals (= FISH positive/break-apart).

BAC Clone	Labelling	Start Coordinate	End Coordinate
RP11-1081A10	red dUTP	chr17:7420240	chr17:7589389
RP11-107F4	green dUTP	chr17:7627095	chr17:7785964

Table 1 – Fluorophore labelling and genomic coordinates (NCBI human reference genome build 37 / hg19) of BAC clones as used for break-apart FISH for detection of *TP53* rearrangements.

4.3 Methylation- and Copy Number Arrays of Osteosarcoma

4.3.1 Copy Number Arrays

Genome-wide CytoScan HD Arrays (Affymetrix, CA, USA) were performed according to the manufacturer's instructions using 250 ng of genomic DNA from each tumor sample. To evaluate copy number variations, data was processed using the Nexus Copy Number software (Version 7.0, BioDiscovery, USA) with integrated algorithms for segmentation and normalization. Copy number alterations were called using the SNP-FASST2 algorithm with a significance threshold of $1.0E^{-14}$, a maximum contiguous probe spacing of 1 Mbp, and a minimum number of 3 probes per segment. For LOH regions, a minimum length of 500 KB was required.

4.3.2 Methylation Arrays and Differential Methylation Analysis

250 ng of DNA were used on an Infinium Methylation EPIC BeadChip (Illumina, USA). Differential methylation analysis between tumors was performed in R using the Chip Analysis Methylation Pipeline for Illumina HumanMethylation450 and EPIC chips (ChAMP version 1.11.1) [237]. The data was normalized by beta-mixture quantile normalization (BMIQ) and methylation variable positions were segmented into biologically relevant differentially methylated regions using the Probe Lasso algorithm [238].

4.4 *TP53* Breakpoint Identification in a LFS Family

4.4.1 Custom Sequence Capture for *TP53* Breakpoint Identification In a LFS Family by Paired-End Sequencing

The *TP53*-containing region chr17:7,520,000 - 7,680,000 (NCBI build 37) was defined for a custom sequence capture (SeqCap EZ Choice, Roche NimbleGen Inc.). Repetitive regions are usually excluded for sequence capture assays but since most of the observed breakpoints in OS were in LINE sequences, we forced the inclusion of the repetitive sequences of the intron 1 region of *TP53* (chr17:7,579,941 - 7,590,694). Illumina sequencing libraries were constructed and capturing was performed according to the manufacturer's recommendations and the library was sequenced on an Illumina HiSeq 2000 by 2x50 bases. Reads were mapped to the human reference genome (NCBI build 37) by BWA and read-pairs where one read

Breakpoint Name	Sequence Forward	Sequence Reverse
LFS-BP1	CTCAAAAGGCCATCAAAAGG	GTATGGTGGCCTGTTCCTGT
LFS-BP2	GGCTGCTGGGAGTTGTAGTC	AGCTATCTTGACCCCACACG
LFS-BP1-wt	CCCGAATAGCTGGGATTACA	GCAAGTGCACAGGAAGATGA
LFS-BP2-wt	GGAGGAATCCTGCATTGTGT	CAGGCTTCAGACCTGTCTCC
POLR2A positive control	GCTGCTGGACGTGAGTATGA	AGTTCCAACAATGGCTACCG

Table 2 – PCR primer pairs as used for validation of *TP53* breakpoint candidates in a Li-Fraumeni Family.

mapped to the *TP53* target region and the paired read mapped outside of that region were identified by custom scripts. Reads of similar mapping patterns were clustered together as described earlier in [234] and prioritized for validation by PCR and Sanger sequencing based on cluster size (number of paired-reads with similar mapping patterns).

4.4.2 Sanger Sequencing Validation of *TP53* Rearrangements In a LFS Family

Identified *TP53* breakpoints were screened by PCR in all LFS family members of whom DNA was available. Validation PCR with 120 ng DNA template were conducted using Jumpstart REDAccuTaq LA DNA Polymerase (Sigma-Aldrich) following manufacturer's instructions. Where sample was insufficient, the genomic DNA was amplified using REPLI-g Mini Kit (Qiagen) and 3 μl diluted amplification product were used as template for validation PCR. PCR were performed using the following cycling conditions: 3 min at 94°C, (20 s at 94°C, 30 s at 58°C, 2 min at 68°C)x15, (20 s at 94°C , 30 s at 55°C , 2 min at 68°C)x20, 5 min at 68°C. For primer sequences see table 2 on page 135.

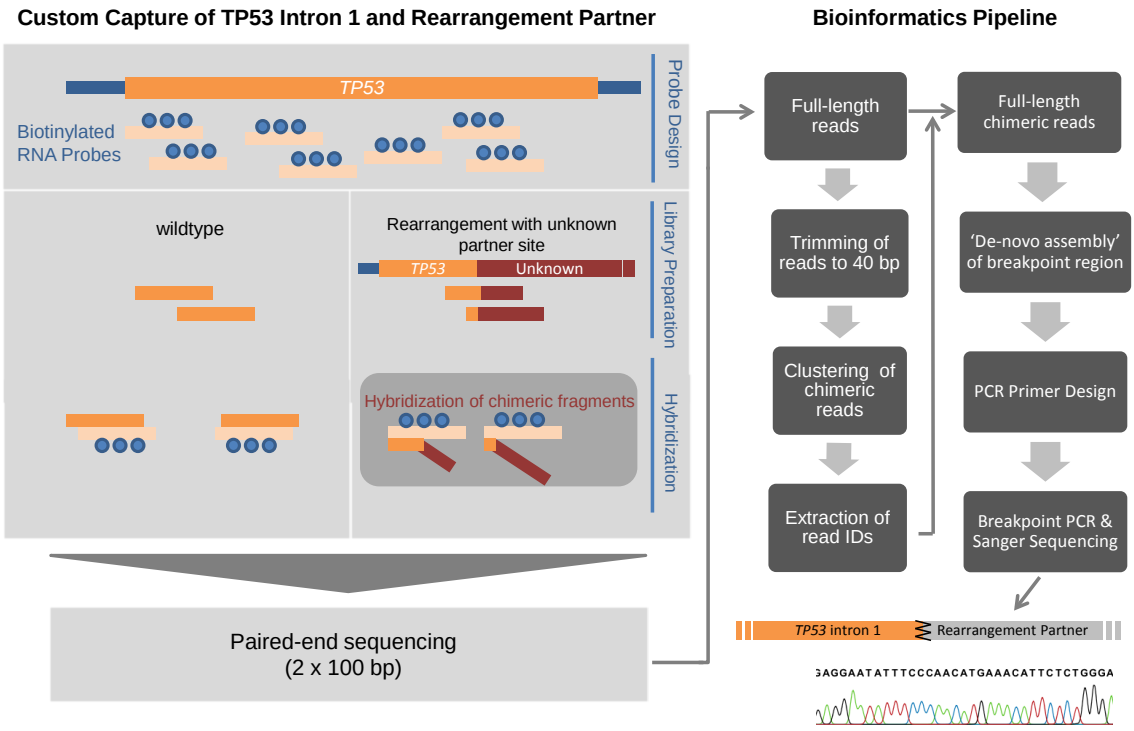


Figure 13 – Workflow of Custom Sequence Capture and Bioinformatics Pipeline for *TP53* intron 1 breakpoint identification in a LFS Family. Custom capture of the *TP53* region and its rearrangement partners, based on biotinylated RNA probes, is depicted on the left. Captured fragments, including wildtype *TP53* and chimeric fragments are sequenced using conventional paired-end sequencing. The custom-built bioinformatics workflow is displayed on the right. Full-length reads were trimmed to 40bp to increase mapping probability on chimeric regions, and clustered according to their mapping patterns using an in-house developed pipeline. Read IDs contributing to candidate SV are extracted and the full-length read sequence retrieved. Breakpoint contexts are recreated using de-novo assembly. Validation was carried out using breakpoint PCR and subsequent Sanger sequencing.

4.5 RT-qPCR Expression Analysis of TP53 in LFS Family Samples

RNA was isolated from PAXgene Blood RNA Tubes using the PAXgene Blood miRNA kit (Qiagen). RNA from fresh frozen tumor H2, and RNA from a tumor derived cell line, P13, were isolated using RNAeasy Mini kit (Qiagen). The one step qRT-PCR was carried out using QuantiTect Probe radiotherapy (RT)-PCR kit (Qiagen), with a total of 50 ng RNA per PCR. The TaqMan primer/probe set (Life Technologies) was used for *TP53* full length (HS01034249) and the PrimeTime primer/probe set (Integrated DNA Technologies Inc.) for transcripts encoding for Δ133p53 and Δ160p53 were designed using sequences from Marcel and colleagues [141]. qRT-PCR were performed on a Bio-Rad CFX device. Cq values were normalized against GAPDH as ΔCq and displayed relative to normal blood control as ΔΔCq as fold-change ($2^{-\Delta\Delta Cq}$).

4.6 Whole-Exome and Whole-Genome Sequencing of Osteosarcoma

4.6.1 DNA and RNA Extractions

DNA and RNA was extracted from fresh-frozen tumor samples using a TissueLyser (Qiagen, Germany) and the AllPrep DNA/RNA Micro or Mini Kit (Qiagen, Germany) according to the manufacturer's instructions. RNA was incubated with RNAse-Free DNase (Qiagen, Germany) according to the manufacturer's instructions. Constitutive DNA was extracted from whole blood using QIAamp DNA blood mini kit (Qiagen Germany). DNA was eluted in elution buffer (Qiagen, Germany) and RNA was eluted in nuclease-free H_2O (Ambion / Thermo Fisher Scientific, USA). Quality of both RNA and DNA was assessed by Nanodrop and 2100 Bioanalyzer (Agilent Technologies, USA).

4.6.2 Library Preparation for Illumina Whole-Exome Sequencing

Exome enrichment was performed using the SureSelect target capture kit (version 4, Agilent Technologies, USA). Samples were quantified using the Qubit system (Invitrogen, USA) and sequencing libraries constructed from 1μg DNA. Samples were sequenced using the Illumina HiSeq 2000 platform as paired 100-bp reads with Chemistry version 3.0, with the aim of target coverage of 100x for the blood DNA and 200x for the tumors.

4.6.3 Bioinformatics Pipeline for Illumina Whole-Exome Sequencing of 31 OS Exomes

Bioinformatic processing of Illumina sequencing data was performed on a Linux server running Ubuntu, or a cluster system running CentOS with the Univa Grid Engine scheduler. The below workflow is visualized in figure 14 on page 140.

Initial quality control of base-called fastq-files was performed using FastQC [239] to assess base quality, sequencing adapter contamination, and kmer enrichment, amongst other indicators. If necessary CutAdapt [240] was used to trim off low quality bases or to remove adapters. Reads were mapped with Stampy [241] version 1.0.12 onto the Human Reference Genome (GRCh37d5/hg19) and PCR duplicates were removed using Picard. Resulting Sequence Alignment/Map format (SAM) files were converted to the binary binary Alignment/Map format (BAM) format and indexed using SAMtools [242]. SNV and small insertion/deletion (InDel) were called with the Platypus variant caller [243] version 0.5, using the bam files of tumor-normal pairs together to ensure comparable calls at every locus. Variants were only called if they were assigned a sufficiently high posterior probability (phred score >20). Variants with an assigned PASS quality flag were used for further analysed. Variants called with an allele bias flag were also included to increase sensitivity for potentially subclonal mutations. Finally, for selected variants, we made sure that the automatic call matched the data by expert visual inspection of the mapped reads onto the reference genome using the standard integrated genomic viewer [244] and its read direction colouring scheme. Somatic mutations were identified by comparison of matched constitutional and tumor samples.

Variant annotation for both somatic and germline variants was performed by ANNOVAR [245], using hg19 reference genome and 2014 versions of standard databases and functional prediction programs. To reduce false-positive calls, we excluded duplicated genomic regions (>90% homology) from the analysis and variants within regions with low mappability scores. Variants were annotated with ANNOVAR Refseq gene model using dbSNP(135); 1,000 genomes project allele frequencies (October 2014), 6,500 exome-sequencing project allele frequencies, University of Santa Cruz (UCSC) segmental duplication scores and UCSC 46 species conservation scores and predictions of functional importance from SIFT, PolyPhen2 and MutationTaster.

For the analysis of mutation burden and mutation spectra, we applied the following exclusion filters to somatic variants: (i) presence in a segmental duplication region or a region with mapability score <0.5; (ii) variant is present in any read from the paired normal sample; (iii) fewer than 10 reads in total at the variant site in the normal sample; (iv) fewer than 8 reads in total in the tumor; (v) fewer than three variants in the tumor; variant allele frequency <5%

in the tumor; and (vi) presence of variant in public databases (Exome Variant Server, 1,000 genomes project and Complete Genomics 69 reference genomes) at a frequency of >2%. Variants identified in constitutional DNA from any of the other local, non-cancer sequencing projects (for example, 29 million variants across 284 samples from the Oxford-Illumina WGS500 consortium) were discarded as being more likely based on a systematic error in our sequencing pipeline than on a genuine somatic mutation.

4.7 Bioinformatics Pipeline for Illumina Whole Exome Sequencing of OS Cases with RET Germline Mutations

The pipeline for Whole Exome Sequencing was as described in section 4.6.3 on page 138, with the following deviations: Reads were mapped to the human hs37d5 reference genome using BWA [246] version 0.7.12-r1039. Where available, reads from whole genomes and exomes were pooled for Platypus SNV and small InDel calling. Annotation was performed using ANNOVAR (release date version 2016-02-01) with the hg19 reference genome and 2015 versions of standard databases and functional prediction programs. In addition to databases described in section 4.6.3 on page 138, ANNOVAR was used to also annotate the variants with the ExAC database of variant frequencies.

4.7.1 Mutation Spectra, Somatic Signatures and Analysis of Mutation Burden

For each tumor with exome sequencing or whole-genome sequencing data available, mutation burden in terms of somatic SNV frequencies were calculated and the deviation of their spectra from the background (determined across all samples) was assessed using χ^2-test (df=5). The trinucleotide spectrum was normalized according to 3-mer frequencies in the reference exome (Agilent SureSelect version 4). Somatic mutation signatures were inferred using the R package SomaticSignatures [247, 248], in which a mutation spectrum was decomposed with a non-negative matrix factorization algorithm. The decomposition was performed for prior sets of 2, 3 ,4, 5, 6 and 10. The optimal number of signatures was manually chosen based on the maximum differentiation between the signatures. The signature analysis was repeated 10 times with the same results obtained after each run.

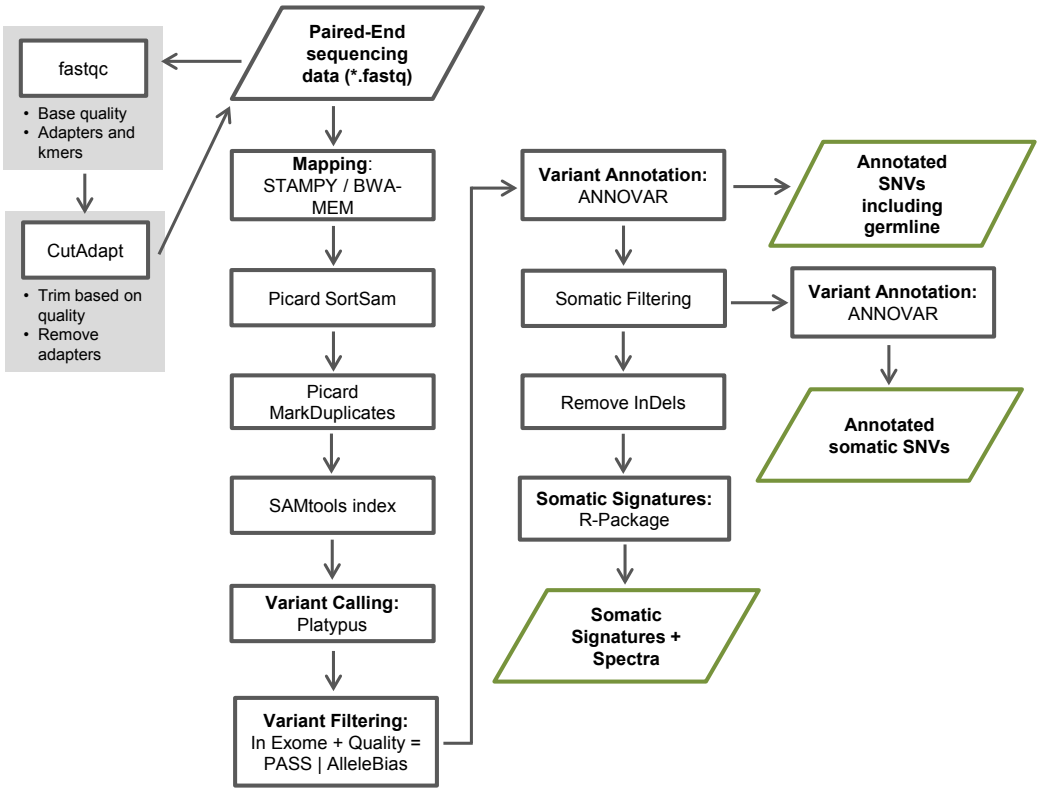


Figure 14 – Bioinformatics pipeline workflow for quality control, mutation calling and annotation of Illumina exome-sequencing data. Input data (gray trapezoid) in form of base-called fastq files were quality checked using fastqc and trimmed by CutAdapt based on base quality or adapter sequences present in the reads. Reads were mapped to the Human Reference Genome (GRCh37d5/hg19) by Stampy and BWA-MEM, PCR duplicates removed and the bam files indexed. Platypus variant caller was used to identify SNV and small InDel which were assigned a PASS or AllelBias quality flag. Variant annotation for both germline and somatic variants was done with ANNOVAR. Somatic signatures and spectra were calculated using R.

4.7.2 Illumina Targeted Re-Sequencing

Biological replication was carried out using 1 μg of DNA from 92 additional tumors and Illumina MiSeq re-sequencing to an average depth of 480x. The custom Nimblegen target enrichment kit was used to capture all exons of 30 genes (that is, 13 main OS drivers and their 17 interaction partners; see table 3 on page 141), and a ~20-kb region encompassing the *TP53* gene Repetitive elements and other sequence features of the desired capture region allowed a probe coverage of 92%. Read mapping, filtering and variant calling was done with the Stampy-Platypus pipeline using the same setting as described above for Whole Exome Sequencing.

Gene Symbol	Description	Chromosome	Position
APC	APC, WNT signaling pathway regulator	chr5	112043201-112181936
ATRX	ATRX, chromatin remodeler	chrX	76760355-77041755
CCND1	cyclin D1	chr11	69455872-69469242
CDKN1A	cyclin dependent kinase inhibitor 1A	chr6	36644236-36655116
CDKN2A	cyclin dependent kinase inhibitor 2A	chr9	21967750-21994490
EP300	E1A binding protein p300	chr22	41488613-41576081
ERBB4	erb-b2 receptor tyrosine kinase 4	chr2	212240441-213403352
FANCA	Fanconi anemia complementation group A	chr16	89803958-89883065
FAS	Fas cell surface death receptor	chr10	90750315-90776818
LIMCH1	LIM and calponin homology domains 1	chr4	41361623-41702061
MDC1	mediator of DNA damage checkpoint 1	chr6	30667583-30685458
MUTYH	mutY DNA glycosylase	chr1	45794913-45806142
MYCN	v-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog	chr2	16080559-16087129
NUMA1	nuclear mitotic apparatus protein 1	chr11	71713909-71791739
PPP2R3B	protein phosphatase 2 regulatory subunit B"beta	chrX	294667-347690
PTEN	phosphatase and tensin homolog	chr10	89623194-89731687
PTENP1	phosphatase and tensin homolog pseudogene 1	chr9	33673501-33677418
PTPRZ1	protein tyrosine phosphatase, receptor type Z1	chr7	121513158-121702090
RB1	RB transcriptional corepressor 1	chr13	48877882-49056026
RECQL4	RecQ like helicase 4	chr8	145736666-145743210
RET	ret proto-oncogene	chr10	43572516-43625797
RREB1	ras responsive element binding protein 1	chr6	7107829-7252213
RUNX1	runt related transcription factor 1	chr21	36160097-36421595
RUNX2	runt related transcription factor 2	chr6	45296053-45518819
RUNX3	runt related transcription factor 3	chr1	25226001-25291501
SFPQ	splicing factor proline and glutamine rich	chr1	35641978-35658746
STAT3	signal transducer and activator of transcription 3	chr17	40465342-40540513
TP53	tumor protein p53	chr17	7571719-7590868
TTN	titin	chr2	179390716-179672150
ZNF276	zinc finger protein 276	chr16	89786775-89807332

Table 3 – Genes included in the custom target capture sequencing for biological replication of osteosarcoma sequencing. Based on results from an exome sequencing of 31 OS cases the genes listed above were used for re-sequencing of a replication cohort consisting of 92 additional tumors. Sequencing is further described on page 140. Genomic coordinates given are referring to the Human Reference Genome (hg19).

4.7.3 Illumina Whole Genome Sequencing of OS Cases with RET Germline Mutations

DNA samples were quantified using the Qubit system (Invitrogen, USA) and sequencing libraries constructed from 0.5 μg of DNA. Samples were sequenced using the Illumina HiSeq 2000 or 4000 platform as paired-end 100-bp reads with Chemistry version 3.0. A minimum target coverage of 15x was aimed for. Quality control, removal of PCR duplicates and Stampy/BWA mapping was done using the bioinformatics pipeline used for exome sequencing which was described in section 4.6.3 on page 138.

4.7.4 Classification of *RET* Mutations

Mutation Classification by ATA Guidelines

The correlation between mutations and HSCR/MEN can be complex. Therefore, the American thyroid association has released detailed treatment guidelines for thyroid cancer based on RET mutations [209]. Described mutations are classified into the risk levels ATA-A (lowest risk) to ATA-D (highest risk). Independent of symptoms, prophylactic thyroidectomy is generally recommended before the age of 5 for carriers of mutations of class ATA-A to ATA-C, and before age 1 for carriers of class ATA-D mutations (mostly MEN2B).

Mutation Classification Based on American College of Medical Genetics and Genomics (ACMG)

The ACMG has developed standards for the interpretation of germline sequence variants [249], which are widely used by geneticists in clinical settings. The ACMG recommends the classification of variants into the groups 'pathogenic', 'likely pathogenic', uncertain significance', 'likely benign', and 'benign'. This classification is based on a broad spectrum of possible evidence, such as population data, *in silico* predictions, functional data, segregation data in affected families, and *de novo* occurrence. These criteria are combined according to a matrix which then leads to the classification of a variant into the above groups. Depending on the nature and strength of the observations, the evidence can further be classified into supporting, moderate, strong, or very strong evidence for pathogenicity or benignity. So shall, for example, the fact that a suspected disease causing variant is found *de novo* in a patient be classified as 'moderate pathogenic' evidence, until maternity and paternity is confirmed,

upon which the evidence can be reclassified as 'strong pathogenic' It should be noted that the ACMG guidelines have been developed for the interpretation of germline variants. Somatic mutations, as found in cancer cells, introduce several additional layers of complexity, such as tumor heterogeneity and -purity of the sample.

4.7.5 Structural Variant Detection From Conventional Whole Genome Sequencing

Structural variants were predicted from whole-genome sequencing data using Breakdancer [250] version 1.4.5 and the Human Reference Genome GRCh37d5. Candidate rearrangements which were present in a normal control genome, or mapping to any decoy chromosome or having less than 6 supporting reads were removed.

4.8 RNA-Sequencing of Osteosarcoma

4.8.1 Library Preparation and Sequencing

Ribosomal RNA was depleted from total RNA using the Ribo-Zero rRNA Removal Kit (Illumina, USA), and subsequently converted to cDNA. The cDNA was end-repaired, A-tailed and adapter-ligated. The resulting libraries were size-selected and quality was assessed. The finalized libraries were paired-end sequenced with 75-bp reads on an Illumina HiSeq 4000 over one lane of a flow cell.

4.8.2 Fusion Gene Prediction from RNA Sequencing Data

Base-called fastq files were quality-assessed by FastQC [239]. For each sample, 45 million reads were used for analysis. Fusion gene predictions were performed with Fusioncatcher software [251] version 0.99a with the ensembl v84 database and EricScript [252] version 0.5.5 with the provided ensembl database for human transcripts. To reduce false-positives, candidates were only included for further analysis if they were predicted by both prediction tools, or if at least one of the rearrangement partners included a *bona fide* OS driver gene or a tyrosine kinase.

References

[1] Stefan S Bielack, Beate Kempf-Bielack, Günter Delling, G Ulrich Exner, Silke Flege, Knut Helmke, Rainer Kotz, Mechthild Salzer-Kuntschik, Matthias Werner, Winfried Winkelmann, et al. Prognostic factors in high-grade osteosarcoma of the extremities or trunk: an analysis of 1,702 patients treated on neoadjuvant cooperative osteosarcoma study group protocols. *Journal of Clinical Oncology*, 20(3):776–790, 2002.

[2] Timothy A Damron, William G Ward, and Andrew Stewart. Osteosarcoma, chondrosarcoma, and ewing's sarcoma: National cancer data base report. *Clinical orthopaedics and related research*, 459:40–47, 2007.

[3] IL Christopher. National cancer institutes surveillance epidemiology and end results (seer) data analysis from nine population-based us cancer registries. *JAMA*, 289:1421–1424, 2003.

[4] Christopher DM Fletcher, K Krishnan Unni, and FE Mertens. World health organisation classification of tumors. *Pathology and genetics of tumours of soft tissue and bone. Lyon: IARC*, 2002.

[5] American Cancer Society. American cancer society: Cancer facts and figures 2012, 2012.

[6] Leo Kager, Andreas Zoubek, Martin Dominkus, Susanna Lang, Nicole Bodmer, Gernot Jundt, Thomas Klingebiel, Heribert Jürgens, Helmut Gadner, and Stefan Bielack. Osteosarcoma in very young children. *Cancer*, 116(22):5316–5324, 2010.

[7] Leo Mascarenhas, Stuart Siegel, Logan Spector, Carola Arndt, Dominic Femino, and Marcio Malogolowkin. Malignant bone tumors. *Cancer epidemiology in older adolescents and young adults*, 15(29):1975–2000, 2006.

[8] Lisa Mirabello, Rebecca J Troisi, and Sharon A Savage. International osteosarcoma incidence patterns in children and adolescents, middle ages and elderly persons. *International Journal of Cancer*, 125(1):229–234, 2009.

[9] Sharon A Savage and Lisa Mirabello. Using epidemiology and genomics to understand osteosarcoma etiology. *Sarcoma*, 2011, 2011.

[10] James G Gurney, Andrine R Swensen, Marc Bulterys, et al. Malignant bone tumors. *Cancer incidence and survival among children and adolescents: United States SEER Program*, 1995(1999):99–110, 1975.

[11] CHG Price. Primary bone-forming tumours and their relationship to skeletal growth. *Bone & Joint Journal*, 40(3):574–593, 1958.

[12] RA Tjalma. Canine bone sarcoma: estimation of relative risk as a function of body size. *Journal of the National Cancer Institute*, 36(6):1137–1150, 1966.

[13] Simon J Cotterill, Charlotte M Wright, Mark S Pearce, and Alan W Craft. Stature of young people with malignant bone tumors. *Pediatric blood & cancer*, 42(1):59–63, 2004.

[14] R Troisi, MN Masters, K Joshipura, C Douglass, BF Cole, and RN Hoover. Perinatal factors, growth and development, and osteosarcoma risk. *British journal of cancer*, 95(11):1603–1607, 2006.

[15] PD Inskip, LAG Ries, RJ Cohen, and RE Curtis. New malignancies following childhood cancer. *New malignancies among cancer survivors: SEER cancer registries*, 2000:465–482, 1973.

[16] Alessandra Longhi, Enza Barbieri, Nicola Fabbri, Michele Macchiagodena, Leonardo Favale, Cataldo Lippo, Nicola Salducca, and Gaetano Bacci. Radiation-induced osteosarcoma arising 20 years after the treatment of ewing's sarcoma. *Tumori*, 89(5):569–572, 2002.

[17] Giulia Ottaviani and Norman Jaffe. The epidemiology of osteosarcoma. In *Pediatric and adolescent osteosarcoma*, pages 3–13. Springer, 2009.

[18] Jane E Visvader. Cells of origin in cancer. *Nature*, 469(7330):314–322, 2011.

[19] Ander Abarrategi, Juan Tornin, Lucia Martinez-Cruzado, Ashley Hamilton, Enrique Martinez-Campos, Juan P Rodrigo, M González, Nicola Baldini, Javier Garcia-Castro, and Rene Rodriguez. Osteosarcoma: Cells-of-origin, cancer stem cells, and targeted therapies. *Stem Cells International*, 2016, 2016.

[20] R Rubio, I Gutierrez-Aranda, AI Sáez-Castillo, A Labarga, M Rosu-Myles, Sara González-García, María Luisa Toribio, P Menendez, and R Rodriguez. The differentiation stage of p53-rb-deficient bone marrow mesenchymal stem cells imposes the phenotype of in vivo sarcoma development. *Oncogene*, 32(41):4970–4980, 2013.

[21] James C Wittig, Jacob Bickels, Dennis Priebat, James Jelinek, Kristen Kellar-Graney, Barry Shmookler, and Martin M Malawer. Osteosarcoma: a multidisciplinary approach to diagnosis and treatment. *American family physician*, 65(6), 2002.

[22] Ahmad Oryan, Soodeh Alidadi, and Ali Moshiri. Osteosarcoma: Current concepts, challenges and future directions. *Current Orthopaedic Practice*, 26(2):181–198, 2015.

[23] S Bielack, D Carrle, PG Casali, ESMO Guidelines Working Group, et al. Osteosarcoma: Esmo clinical recommendations for diagnosis, treatment and follow-up. *Annals of Oncology*, 20(suppl 4):iv137–iv139, 2009.

[24] Michael P Link, Allen M Goorin, Marc Horowitz, William H Meyer, Jean Belasco, Alan Baker, Alberto Ayala, and Jonathan Shuster. Adjuvant chemotherapy of high-grade osteosarcoma of the extremity: Updated results of the multi-institutional osteosarcoma study. *Clinical orthopaedics and related research*, 270:8–14, 1991.

[25] Christopher DM Fletcher, World Health Organization, et al. *WHO classification of tumours of soft tissue and bone:[this book reflects the views of a working group that convened for a consensus and editorial meeting at the University of Zurich, Switzerland, 18-20 April 2012]*. Internat. Agency for Research on Cancer, 2013.

[26] Piero Picci. Osteosarcoma (osteogenic sarcoma). *Orphanet journal of rare diseases*, 2(1):1, 2007.

[27] Nando Ferreira and Leonard Charles Marais. Osteosarcoma presentation stages at a tumour unit in south africa. *SAMJ: South African Medical Journal*, 102(8):673–676, 2012.

[28] Neyssa Marina, Mark Gebhardt, Lisa Teot, and Richard Gorlick. Biology and therapeutic advances for pediatric osteosarcoma. *The oncologist*, 9(4):422–441, 2004.

[29] Alexander J Chou, David S Geller, and Richard Gorlick. Therapy for osteosarcoma. *Pediatric Drugs*, 10(5):315–327, 2008.

[30] Gaetano Bacci, Alessandra Longhi, Franco Bertoni, Antonio Briccoli, Michela Versari, Elettra Pignotti, and Piero Picci. Bone metastases in osteosarcoma patients treated with neoadjuvant or adjuvant chemotherapy the rizzoli experience in 52 patients. *Acta orthopaedica*, 77(6):938–943, 2006.

[31] Michael A Friedman and Stephen K Carter. The therapy of osteogenic sarcoma: current status and thoughts for the future. *Journal of surgical oncology*, 4(5-6):482–510, 1972.

[32] Beate Kempf-Bielack, Stefan S Bielack, Heribert Jürgens, Detlev Branscheid, Wolfgang E Berdel, G Ulrich Exner, Ulrich Göbel, Knut Helmke, Gernot Jundt, Hartmut Kabisch, et al. Osteosarcoma relapse after combined modality therapy: an analysis of unselected patients in the cooperative osteosarcoma study group (coss). *Journal of Clinical Oncology*, 23(3):559–568, 2005.

[33] Norman Jaffe, Oyvind S Bruland, and Stefan Bielack. *Pediatric and adolescent osteosarcoma*, volume 152. Springer Science & Business Media, 2010.

[34] Allen M Goorin, Antonio Perez-Atayde, Mark Gebhardt, Janet W Andersen, Robert H Wilkinson, MJ Delorey, Hugh Watts, Michael Link, Norman Jaffe, and E Frei. Weekly high-dose methotrexate and doxorubicin for osteosarcoma: the dana-farber cancer institute/the children's hospital–study iii. *Journal of Clinical Oncology*, 5(8):1178–1184, 1987.

[35] AM Goorin, M Delorey, RD Gelber, A Perez-Atayde, M Gebhardt, JW Andersen, et al. The dana-farber cancer institute/the childrens hospital adjuvant chemotherapy trials for osteosarcoma: three sequential studies. In *Cancer Treat Symp*, volume 3, pages 155–159, 1985.

[36] William F Taylor, John C Ivins, Douglas J Pritchard, David C Dahlin, Gerald S Gilchrist, and John H Edmonson. Trends and variability in survival among patients with osteosarcoma: a 7-year update. In *Mayo Clinic Proceedings*, volume 60, pages 91–104. Elsevier, 1985.

[37] Michael P Link, Allen M Goorin, Angela W Miser, Alexander A Green, Charles B Pratt, Jean B Belasco, Jon Pritchard, James S Malpas, Alan R Baker, John A Kirkpatrick, et al. The effect of adjuvant chemotherapy on relapse-free survival in patients with osteosarcoma of the extremity. *New England Journal of Medicine*, 314(25):1600–1606, 1986.

[38] Dorothe Carrle and Stefan S Bielack. Current strategies of chemotherapy in osteosarcoma. *International orthopaedics*, 30(6):445–451, 2006.

[39] M Salzer-Kuntschik, G Delling, G Beron, and R Sigmund. Morphological grades of regression in osteosarcoma after polychemotherapy - study coss 80. *Journal of cancer research and clinical oncology*, 106(1):21–24, 1983.

[40] Stefano Ferrari, Pietro Ruggieri, Graziella Cefalo, Angela Tamburini, Rodolfo Capanna, Franca Fagioli, Alessandro Comandone, Rossella Bertulli, Gianni Bisogno, Emanuela Palmerini, et al. Neoadjuvant chemotherapy with methotrexate, cisplatin, and doxorubicin with or without ifosfamide in nonmetastatic osteosarcoma of the extremity: an italian sarcoma group trial isg/os-1. *Journal of clinical oncology*, pages JCO–2011, 2012.

[41] Neyssa M Marina, Sigbjørn Smeland, Stefan S Bielack, Mark Bernstein, Gordana Jovic, Mark D Krailo, Jane M Hook, Carola Arndt, Henk van den Berg, Bernadette Brennan, et al. Comparison of mapie versus map in patients with a poor response to preoperative chemotherapy for newly diagnosed high-grade osteosarcoma (euramos-1): an open-label, international, randomised controlled trial. *The Lancet Oncology*, 17(10):1396–1408, 2016.

[42] Stefan S Bielack, Sigbjørn Smeland, Jeremy S Whelan, Neyssa Marina, Gordana Jovic, Jane M Hook, Mark D Krailo, Mark Gebhardt, Zsuzsanna Pápai, James Meyer, et al. Methotrexate, doxorubicin, and cisplatin (map) plus maintenance pegylated interferon alfa-2b versus map alone in patients with resectable high-grade osteosarcoma and good histologic response to preoperative map: first results of the euramos-1 good response randomized controlled trial. *Journal of Clinical Oncology*, 33(20):2279–2287, 2015.

[43] Gaetano Bacci, Alessandra Longhi, Michela Versari, Mario Mercuri, Antonio Briccoli, and Piero Picci. Prognostic factors for osteosarcoma of the extremity treated with neoadjuvant chemotherapy. *Cancer*, 106(5):1154–1161, 2006.

[44] Daniel C Allison, Scott C Carney, Elke R Ahlmann, Andrew Hendifar, Sant Chawla, Alex Fedenko, Constance Angeles, and Lawrence R Menendez. A meta-analysis of osteosarcoma outcomes in the modern medical era. *Sarcoma*, 2012, 2012.

[45] Gaetano Bacci, Mario Mercuri, Antonio Briccoli, Stefano Ferrari, Franco Bertoni, Davide Donati, Carlo Monti, Antonio Zanoni, Cristiana Forni, and Marco Manfrini. Osteogenic sarcoma of the extremity with detectable lung metastases at presentation. *Cancer*, 79(2):245–254, 1997.

[46] V Nataraj, S Rastogi, SA Khan, MC Sharma, S Agarwala, S Vishnubhatla, and S Bakhshi. Prognosticating metastatic osteosarcoma treated with uniform chemotherapy protocol without high dose methotrexate and delayed metastasectomy: a single center experience of 102 patients. *Clinical and Translational Oncology*, pages 1–8, 2016.

[47] Gerald Rosen, B Caparros, AG Huvos, C Kosloff, A Nirenberg, A Cacavio, RC Marcove, JM Lane, B Mehta, and Ch Urban. Preoperative chemotherapy for osteogenic sarcoma. *Cancer*, 49(122):1–1, 1982.

[48] Kyle R Duchman, Yubo Gao, and Benjamin J Miller. Prognostic factors for survival in patients with high-grade osteosarcoma using the surveillance, epidemiology, and end results (seer) program database. *Cancer epidemiology*, 39(4):593–599, 2015.

[49] Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer: the next generation. *cell*, 144(5):646–674, 2011.

[50] David S Goodsell. The molecular perspective: methotrexate. *The Oncologist*, 4(4):340–341, 1999.

[51] PT Ravi Rajagopalan, Zhiquan Zhang, Lynn McCourt, Mary Dwyer, Stephen J Benkovic, and Gordon G Hammes. Interaction of dihydrofolate reductase with methotrexate: ensemble and single-molecule kinetics. *Proceedings of the National Academy of Sciences*, 99(21):13481–13486, 2002.

[52] E Reed. Platinum-dna adduct, nucleotide excision repair and platinum based anti-cancer chemotherapy. *Cancer treatment reviews*, 24(5):331–344, 1998.

[53] Sui Zhang, Xiaobing Liu, Tasneem Bawa-Khalfe, Long-Sheng Lu, Yi Lisa Lyu, Leroy F Liu, and Edward TH Yeh. Identification of the molecular basis of doxorubicin-induced cardiotoxicity. *Nature medicine*, 18(11):1639–1642, 2012.

[54] Cheng-jun Li, Xiao-zhou Liu, Lei Zhang, Long-bang Chen, Xin Shi, Su-jia Wu, and Jian-ning Zhao. Advances in bone-targeted drug delivery systems for neoadjuvant chemotherapy for osteosarcoma. *Orthopaedic Surgery*, 8(2):105–110, 2016.

[55] Alexander J Chou and Richard Gorlick. Chemotherapy resistance in osteosarcoma: current challenges and future directions. *Expert review of anticancer therapy*, 6(7):1075–1085, 2006.

[56] Ana Patiño-García, Marta Zalacaín, Lucía Marrodán, Mikel San-Julián, and Luis Sierrasesúmaga. Methotrexate in pediatric osteosarcoma: response and toxicity in relation to genetic polymorphisms and dihydrofolate reductase and reduced folate carrier 1 expression. *The Journal of pediatrics*, 154(5):688–693, 2009.

[57] Wayne F Flintoff, Heather Sadlish, Richard Gorlick, Rui Yang, and Frederick MR Williams. Functional analysis of altered reduced folate carrier sequence changes identified in osteosarcomas. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1690(2):110–117, 2004.

[58] Hongtao He, Jiangdong Ni, and Jun Huang. Molecular mechanisms of chemoresistance in osteosarcoma (review). *Oncology letters*, 7(5):1352–1362, 2014.

[59] Dong Wang, Meihua Luo, and Mark R Kelley. Human apurinic endonuclease 1 (ape1) expression and prognostic significance in osteosarcoma: enhanced sensitivity of osteosarcoma to dna damaging agents using silencing rna ape1 expression inhibition. *Molecular cancer therapeutics*, 3(6):679–686, 2004.

[60] D Caronia, A Patino-Garcia, RL Milne, M Zalacain-Diez, G Pita, MR Alonso, LT Moreno, L Sierrasesumaga-Ariznabarreta, J Benitez, and A Gonzalez-Neira. Common variations in ercc2 are associated with response to cisplatin chemotherapy and clinical outcome in osteosarcoma patients. *The pharmacogenomics journal*, 9(5):347–353, 2009.

[61] Edgar Pérez-Herrero and Alberto Fernández-Medarde. Advanced targeted therapies in cancer: Drug nanocarriers, the future of chemotherapy. *European journal of pharmaceutics and biopharmaceutics*, 93:52–79, 2015.

[62] Rong Xu, Guodong Zhang, Junhua Mai, Xiaoyong Deng, Victor Segura-Ibarra, Suhong Wu, Jianliang Shen, Haoran Liu, Zhenhua Hu, Lingxiao Chen, et al. An injectable nanoparticle generator enhances delivery of cancer therapeutics. *Nature biotechnology*, 2016.

[63] Yechezkel Chezy Barenholz. Doxil® - the first fda-approved nano-drug: lessons learned. *Journal of controlled release*, 160(2):117–134, 2012.

[64] Neha Gupta, Hassan Hatoum, and Grace K Dy. First line treatment of advanced non-small-cell lung cancer–specific focus on albumin bound paclitaxel. *International journal of nanomedicine*, 9:209, 2014.

[65] Ilaria Buondonno, Elena Gazzano, Sae Rin Jean, Valentina Audrito, Joanna Kopecka, Marilù Fanelli, Iris C Salaroglio, Costanzo Costamagna, Ilaria Roato, Eleonora Mungo, et al. Mitochondria-targeted doxorubicin: a new therapeutic strategy against doxorubicin-resistant osteosarcoma. *Molecular Cancer Therapeutics*, 2016.

[66] Katsuhiro Hayashi, Ming Zhao, Kensuke Yamauchi, Norio Yamamoto, Hiroyuki Tsuchiya, Katsuro Tomita, Hiroyuki Kishimoto, Michael Bouvet, and Robert M Hoffman. Systemic targeting of primary bone tumor and lung metastasis of high-grade osteosarcoma in nude mice with a tumor-selective strain of salmonella typhymurium. *Cell Cycle*, 8(6):870–875, 2009.

[67] Zhen-Yi Wang and Zhu Chen. Acute promyelocytic leukemia: from highly fatal to highly curable. *Blood*, 111(5):2505–2515, 2008.

[68] Myriam Alcalay, Daniela Zangrilli, Pier Paolo Pandolfi, Letizia Longo, Amedea Mencarelli, angelo Giacomucci, Mariano Rocchi, Andrea Biondi, Alessandro Rambaldi, and F Lo Coco. Translocation breakpoint of acute promyelocytic leukemia lies within the retinoic acid receptor alpha locus. *Proceedings of the National Academy of Sciences*, 88(5):1977–1981, 1991.

[69] Renaud Capdeville, Elisabeth Buchdunger, Juerg Zimmermann, and Alex Matter. Glivec (sti571, imatinib), a rationally developed, targeted anticancer drug. *Nature reviews Drug discovery*, 1(7):493–502, 2002.

[70] Michael W Schmitt, Lawrence A Loeb, and Jesse J Salk. The influence of subclonal resistance mutations on targeted cancer therapy. *Nature reviews Clinical oncology*, 13(6):335–347, 2016.

[71] Francesca Carlomagno, Donata Vitagliano, Teresa Guida, Fortunato Ciardiello, Giampaolo Tortora, Giancarlo Vecchio, Anderson J Ryan, Gabriella Fontanini, Alfredo Fusco, and Massimo Santoro. Zd6474, an orally available inhibitor of kdr tyrosine kinase activity, efficiently blocks oncogenic ret kinases. *Cancer research*, 62(24):7284–7290, 2002.

[72] Stephen R Wedge, Donald J Ogilvie, Michael Dukes, Jane Kendrew, Rosemary Chester, Janet A Jackson, Sarah J Boffey, Paula J Valentine, Jon O Curwen, Helen L Musgrove, et al. Zd6474 inhibits vascular endothelial growth factor signaling, angiogenesis, and tumor growth following oral administration. *Cancer research*, 62(16):4645–4655, 2002.

[73] Simona Soverini, Susan Branford, Franck E Nicolini, Moshe Talpaz, Michael WN Deininger, Giovanni Martinelli, Martin C Müller, Jerald P Radich, and Neil P Shah. Implications of bcr-abl1 kinase domain-mediated resistance in chronic myeloid leukemia. *Leukemia research*, 38(1):10–20, 2014.

[74] Eric S Lander, Lauren M Linton, Bruce Birren, Chad Nusbaum, Michael C Zody, Jennifer Baldwin, Keri Devon, Ken Dewar, Michael Doyle, William FitzHugh, et al. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, 2001.

[75] John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Mills Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, Joshua M Stuart, Cancer Genome Atlas Research Network, et al. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120, 2013.

[76] Simon A Forbes, David Beare, Prasad Gunasekaran, Kenric Leung, Nidhi Bindal, Harry Boutselakis, Minjie Ding, Sally Bamford, Charlotte Cole, Sari Ward, et al. Cosmic: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic acids research*, 43(D1):D805–D811, 2015.

[77] Alex Matter. Bridging academic science and clinical research in the search for novel targeted anti-cancer agents. *Cancer biology & medicine*, 12(4):316, 2015.

[78] Steven A McCarroll, Alan Huett, Petric Kuballa, Shannon D Chilewski, Aimee Landry, Philippe Goyette, Michael C Zody, Jennifer L Hall, Steven R Brant, Judy H Cho, et al. Deletion polymorphism upstream of irgm associated with altered irgm expression and crohn's disease. *Nature genetics*, 40(9):1107–1112, 2008.

[79] William A May, Mikhail L Gishizky, Stephen L Lessnick, Lynn B Lunsford, Brian C Lewis, Olivier Delattre, Jessica Zucman, Gilles Thomas, and Christopher T Denny. Ewing sarcoma 11; 22 translocation produces a chimeric transcription factor that requires the dna-binding domain encoded by fli1 for transformation. *Proceedings of the National Academy of Sciences*, 90(12):5752–5756, 1993.

[80] Jérôme Lejeune, R TURPIN, Roland Berger, Jacques Vialatte, P Seringe, J LAFOURCADE, and M BOESWILLWALD. Genetique-trois cas de deletion partielle du bras court dun chromosome 5. *COMPTES RENDUS HEBDOMADAIRES DES SEANCES DE L ACADEMIE DES SCIENCES*, 257(21):3098, 1963.

[81] Jerome Lejeune, Martha Gautier, and Raymond Turpin. Study of somatic chromosomes from 9 mongoloid children. *Comptes rendus hebdomadaires des seances de l'Academie des sciences*, 248(11):1721, 1959.

[82] Sabina Solinas-Toldo, Stefan Lampel, Stephan Stilgenbauer, Jeremy Nickolenko, Axel Benner, Hartmut Döhner, Thomas Cremer, and Peter Lichter. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes, chromosomes and cancer*, 20(4):399–407, 1997.

[83] Daniel Pinkel, Richard Segraves, Damir Sudar, Steven Clark, Ian Poole, David Kowbel, Colin Collins, Wen-Lin Kuo, Chira Chen, Ye Zhai, et al. High resolution analysis of dna copy number variation using comparative genomic hybridization to microarrays. *Nature genetics*, 20(2):207–211, 1998.

[84] Cameron Brennan, Yunyu Zhang, Christopher Leo, Bin Feng, Craig Cauwels, Andrew J Aguirre, Minjung Kim, Alexei Protopopov, and Lynda Chin. High-resolution global profiling of genomic alterations with long oligonucleotide microarray. *Cancer research*, 64(14):4744–4748, 2004.

[85] Jason M Rizzo and Michael J Buck. Key principles and clinical applications of "next-generation" dna sequencing. *Cancer Prevention Research*, 5(7):887–900, 2012.

[86] Eray Tuzun, Andrew J Sharp, Jeffrey A Bailey, Rajinder Kaul, V Anne Morrison, Lisa M Pertz, Eric Haugen, Hillary Hayden, Donna Albertson, Daniel Pinkel, et al. Fine-scale structural variation of the human genome. *Nature genetics*, 37(7):727–732, 2005.

[87] Eric E Schadt, Steve Turner, and Andrew Kasarskis. A window into third-generation sequencing. *Human molecular genetics*, 19(R2):R227–R240, 2010.

[88] Alexander S Mikheyev and Mandy MY Tin. A first look at the oxford nanopore minion sequencer. *Molecular ecology resources*, 14(6):1097–1102, 2014.

[89] Johnathan A Watkins, Sheeba Irshad, Anita Grigoriadis, and Andrew NJ Tutt. Genomic scars as biomarkers of homologous recombination deficiency and drug response in breast and ovarian cancers. *Breast Cancer Research*, 16(3):1, 2014.

[90] Kirby Rickel, Fang Fang, and Jianning Tao. Molecular genetics of osteosarcoma. *Bone*, 2016.

[91] Jeff W Martin, Jeremy A Squire, and Maria Zielenska. The genetics of osteosarcoma. *Sarcoma*, 2012, 2012.

[92] Carl W Miller, Abdulkarim Aslo, Alvina Won, Michael Tan, Beatrice Lampkin, and H Phillip Koefflar. Alterations of thep53, rb andmdm2 genes in osteosarcoms. *Journal of cancer research and clinical oncology*, 122(9):559–565, 1996.

[93] Lawrence A Donehower, Michele Harvey, Betty L Slagle, Mark J McArthuri, and Charles A Montgomery Jr. Mice deficient for p53 are developmentally normal but susceptible to spontaneous. *Nature*, 356:19, 1992.

[94] A Saalfrank, KP Janssen, M Ravon, K Flisikowski, S Eser, K Steiger, T Flisikowska, P Müller-Fliedner, É Schulze, C Brönner, et al. A porcine model of osteosarcoma. *Oncogenesis*, 5(3):e210, 2016.

[95] Sandra L Harris and Arnold J Levine. The p53 pathway: positive and negative feedback loops. *Oncogene*, 24(17):2899–2908, 2005.

[96] Xiang Chen, Armita Bahrami, Alberto Pappo, John Easton, James Dalton, Erin Hedlund, David Ellison, Sheila Shurtleff, Gang Wu, Lei Wei, et al. Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. *Cell reports*, 7(1):104–112, 2014.

[97] Jennifer A Perry, Adam Kiezun, Peter Tonzi, Eliezer M Van Allen, Scott L Carter, Sylvan C Baca, Glenn S Cowley, Ami S Bhatt, Esther Rheinbay, Chandra Sekhar Pedamallu, et al. Complementary genomic approaches highlight the pi3k/mtor pathway as a common vulnerability in osteosarcoma. *Proceedings of the National Academy of Sciences*, 111(51):E5564–E5573, 2014.

[98] Thierry Soussi and Klas G Wiman. Shaping genetic alterations in human cancer: the p53 mutation paradigm. *Cancer cell*, 12(4):303–312, 2007.

[99] Fiamma Mantovani, Dawid Walerych, and Giannino Del Sal. Targeting mutant p53 in cancer: a long road to precision therapy. *The FEBS Journal*, 2016.

[100] Serena S Freeman, Steven W Allen, Ramapriya Ganti, Jianrong Wu, Jing Ma, Xiaoping Su, Geoff Neale, Jeffrey S Dome, Najat C Daw, and Joseph D Khoury. Copy number gains in egfr and copy number losses in pten are common events in osteosarcoma tumors. *Cancer*, 113(6):1453–1461, 2008.

[101] Daniel J Weisenberger, Kimberly D Siegmund, Mihaela Campan, Joanne Young, Tiffany I Long, Mark A Faasse, Gyeong Hoon Kang, Martin Widschwendter, Deborah Weener, Daniel Buchanan, et al. Cpg island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with braf mutation in colorectal cancer. *Nature genetics*, 38(7):787–793, 2006.

[102] Christoph Lengauer, Kenneth W Kinzler, and Bert Vogelstein. Genetic instabilities in human cancers. *Nature*, 396(6712):643–649, 1998.

[103] Michael Overholtzer, Pulivarthi H Rao, Reyna Favis, Xin-Yan Lu, Michael B Elowitz, Francis Barany, Marc Ladanyi, Richard Gorlick, and Arnold J Levine. The presence of p53 mutations in human osteosarcomas correlates with high levels of genomic instability. *Proceedings of the National Academy of Sciences*, 100(20):11547–11552, 2003.

[104] Avery A Sandberg and Julia A Bridge. Updates on the cytogenetics and molecular genetics of bone and soft tissue tumors: osteosarcoma and related tumors. *Cancer genetics and cytogenetics*, 145(1):1–30, 2003.

[105] Philip J Stephens, Chris D Greenman, Beiyuan Fu, Fengtang Yang, Graham R Bignell, Laura J Mudie, Erin D Pleasance, King Wai Lau, David Beare, Lucy A Stebbings, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*, 144(1):27–40, 2011.

[106] Tobias Rausch, David TW Jones, Marc Zapatka, Adrian M Stütz, Thomas Zichner, Joachim Weischenfeldt, Natalie Jäger, Marc Remke, David Shih, Paul A Northcott, et al. Genome sequencing of pediatric medulloblastoma links catastrophic dna rearrangements with tp53 mutations. *Cell*, 148(1):59–71, 2012.

[107] Gaëlle Bougeard, Mariette Renaux-Petel, Jean-Michel Flaman, Camille Charbonnier, Pierre Fermey, Muriel Belotti, Marion Gauthier-Villars, Dominique Stoppa-Lyonnet, Emilie Consolino, Laurence Brugières, et al. Revisiting li-fraumeni syndrome from tp53 mutation carriers. *Journal of Clinical Oncology*, 33(21):2345–2352, 2015.

[108] Marc F Hansen, Alex Koufos, Brenda L Gallie, Robert A Phillips, Oystein Fodstad, A Brøgger, Tobias Gedde-Dahl, and Webster K Cavenee. Osteosarcoma and retinoblastoma: a shared chromosomal mechanism revealing recessive predisposition. *Proceedings of the National Academy of Sciences*, 82(18):6216–6220, 1985.

[109] Maya Kansara, Michele W Teng, Mark J Smyth, and David M Thomas. Translational biology of osteosarcoma. *Nature Reviews Cancer*, 14(11):722–735, 2014.

[110] Cosmic: Signatures of mutational processes in human cancer. http://cancer.sanger.ac.uk/cosmic/signatures. Accessed: 2016-12-28.

[111] Claudia MB Carvalho and James R Lupski. Mechanisms underlying structural variant formation in genomic disorders. *Nature Reviews Genetics*, 17(4):224–238, 2016.

[112] Zhishuo Ou, Paweł Stankiewicz, Zhilian Xia, Amy M Breman, Brian Dawson, Joanna Wiszniewska, Przemyslaw Szafranski, M Lance Cooper, Mitchell Rao, Lina Shao, et al. Observation and prediction of recurrent human translocations mediated by nahr between nonhomologous chromosomes. *Genome research*, 21(1):33–46, 2011.

[113] Pawel Stankiewicz and James R Lupski. Genome architecture, rearrangements and genomic disorders. *TRENDS in Genetics*, 18(2):74–82, 2002.

[114] Piotr Dittwald, Tomasz Gambin, Przemyslaw Szafranski, Jian Li, Stephen Amato, Michael Y Divon, Lisa Ximena Rodríguez Rojas, Lindsay E Elton, Daryl A Scott, Christian P Schaaf, et al. Nahr-mediated copy-number variants in a clinical population: mechanistic insights into both genomic disorders and mendelizing traits. *Genome research*, 23(9):1395–1409, 2013.

[115] AS Waldman and Robert Mike Liskay. Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. *Molecular and cellular biology*, 8(12):5350–5357, 1988.

[116] PJ Hastings, Grzegorz Ira, and James R Lupski. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS genet*, 5(1):e1000327, 2009.

[117] Claudia MB Carvalho, Melissa B Ramocki, Davut Pehlivan, Luis M Franco, Claudia Gonzaga-Jauregui, Ping Fang, Alanna McCall, Eniko Karman Pivnick, Stacy Hines-Dowell, Laurie H Seaver, et al. Inverted genomic segments and complex triplication rearrangements are mediated by inverted repeats in the human genome. *Nature genetics*, 43(11):1074–1081, 2011.

[118] Mitch McVey and Sang Eun Lee. Mmej repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends in Genetics*, 24(11):529–538, 2008.

[119] Eric Weterings and Dik C van Gent. The mechanism of non-homologous end-joining: a synopsis of synapsis. *DNA repair*, 3(11):1425–1435, 2004.

[120] Luisa Toffolatti, Barbara Cardazzo, Carlo Nobile, Gian Antonio Danieli, Francesca Gualandi, Francesco Muntoni, Steve Abbs, Patrizia Zanetti, Corrado Angelini, Alessandra Ferlini, et al. Investigating the mechanism of chromosomal deletion: characterization of 39 deletion breakpoints in introns 47 and 48 of the human dystrophin gene. *Genomics*, 80(5):523–530, 2002.

[121] Sarah Djebali, Carrie A Davis, Angelika Merkel, Alex Dobin, Timo Lassmann, Ali Mortazavi, Andrea Tanzer, Julien Lagarde, Wei Lin, Felix Schlesinger, et al. Landscape of transcription in human cells. *Nature*, 489(7414):101–108, 2012.

[122] Duncan B Sparrow, David Sillence, Merridee A Wouters, Peter D Turnpenny, and Sally L Dunwoodie. Two novel missense mutations in hairy-and-enhancer-of-split-7 in a family with spondylocostal dysostosis. *European Journal of Human Genetics*, 18(6):674–679, 2010.

[123] Lovisa Hessle, Kristen A Johnson, H Clarke Anderson, Sonoko Narisawa, Adnan Sali, James W Goding, Robert Terkeltaub, and José Luis Millán. Tissue-nonspecific alkaline phosphatase and plasma cell membrane glycoprotein-1 are central antagonistic regulators of bone mineralization. *Proceedings of the National Academy of Sciences*, 99(14):9445–9449, 2002.

[124] R Huang, M Rosenbach, R Vaughn, D Provvedini, N Rebbe, S Hickman, J Goding, and R Terkeltaub. Expression of the murine plasma cell nucleotide pyrophosphohydrolase pc-1 is shared by human liver, bone, and cartilage cells. regulation of pc-1 expression in osteosarcoma cells by transforming growth factor-beta. *Journal of Clinical Investigation*, 94(2):560, 1994.

[125] Joell L Solan, Leonard J Deftos, James W Coding, and Robert A Terkeltaub. Expression of the nucleoside triphosphate pyrophosphohydrolase pc-1 is induced by basic fibroblast growth factor (bfgf) and modulated by activation of the protein kinase a and c pathways in osteoblast-like osteosarcoma cells. *Journal of Bone and Mineral Research*, 11(2):183–192, 1996.

[126] RUTH K GLOBUS, JEAN PLOUET, and DENIS GOSPODAROWICZ. Cultured bovine bone cells synthesize basic fibroblast growth factor and store it in their extracellular matrix. *Endocrinology*, 124(3):1539–1547, 1989.

[127] Michael L Sohaskey, Yebin Jiang, Jenny J Zhao, Andreas Mohr, Frank Roemer, and Richard M Harland. Osteopotentia regulates osteoblast maturation, bone formation, and skeletal integrity in mice. *The Journal of cell biology*, 189(3):511–525, 2010.

[128] Yotam Drier, Michael S Lawrence, Scott L Carter, Chip Stewart, Stacey B Gabriel, Eric S Lander, Matthew Meyerson, Rameen Beroukhim, and Gad Getz. Somatic rearrangements across cancer reveal classes of samples with distinct patterns of dna breakage and rearrangement-induced hypermutability. *Genome research*, 23(2):228–235, 2013.

[129] Sebastian Ribi, Daniel Baumhoer, Kristy Lee, Audrey SM Teo, Babita Madan, Kang Zhang, Wendy K Kohlmann, Fei Yao, Wah Heng Lee, Qiangze Hoi, et al. Tp53 intron 1 hotspot rearrangements are specific to sporadic osteosarcoma and can cause li-fraumeni syndrome. *Oncotarget*, 6(10):7727–7740, 2015.

[130] Mahadeb Pal, Alfred S Ponticelli, and Donal S Luse. The role of the transcription bubble and tfiib in promoter clearance by rna polymerase ii. *Molecular cell*, 19(1):101–110, 2005.

[131] Christian Lanctôt, Thierry Cheutin, Marion Cremer, Giacomo Cavalli, and Thomas Cremer. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nature Reviews Genetics*, 8(2):104–115, 2007.

[132] C Nowell. The minute chromosome (ph 1) in chronic granulocytic leukemia. *Annals of Hematology*, 8(2):65–66, 1962.

[133] Jean-Christophe Bourdon, Kenneth Fernandes, Fiona Murray-Zmijewski, Geng Liu, Alexandra Diot, Dimitris P Xirodimas, Mark K Saville, and David P Lane. p53 isoforms can regulate p53 transcriptional activity. *Genes & development*, 19(18):2122–2137, 2005.

[134] Gilles Gadea, Nikola Arsic, Kenneth Fernandes, Alexandra Diot, Sébastien M Joruiz, Samer Abdallah, Valerie Meuray, Stéphanie Vinot, Christelle Anguille, Judit Remenyi, et al. Tp53 drives invasion through expression of its δ133p53β variant. *eLife*, 5:e14734, 2016.

[135] Marie P Khoury and Jean-Christophe Bourdon. p53 isoforms an intracellular microprocessor? *Genes & cancer*, 2(4):453–465, 2011.

[136] Marie P Khoury and Jean-Christophe Bourdon. The isoforms of the p53 protein. *Cold Spring Harbor perspectives in biology*, 2(3):a000927, 2010.

[137] Marco M Candeias, Masatoshi Hagiwara, and Michiyuki Matsuda. Cancer-specific mutations in p53 induce the translation of δ160p53 promoting tumorigenesis. *EMBO reports*, page e201541956, 2016.

[138] Salah Mahmoudi, Sofia Henriksson, Martin Corcoran, Cristina Méndez-Vidal, Klas G Wiman, and Marianne Farnebo. Wrap53, a natural p53 antisense transcript required for p53 induction upon dna damage. *Molecular cell*, 33(4):462–471, 2009.

[139] Sofia Henriksson, Hanif Rassoolzadeh, Elisabeth Hedström, Christos Coucoravas, Alexander Julner, Michael Goldstein, Gabriela Imreh, Boris Zhivotovsky, Michael B Kastan, Thomas Helleday, et al. The scaffold protein wrap53β orchestrates the ubiquitin response critical for dna double-strand break repair. *Genes & development*, 28(24):2726–2738, 2014.

[140] Sofia Henriksson and Marianne Farnebo. On the road with wrap53β: guardian of cajal bodies and genome integrity. *Frontiers in genetics*, 6:91–91, 2014.

[141] Virginie Marcel, Stéphane Perrier, Mustapha Aoubala, Sylvain Ageorges, Michael J Groves, Alexandra Diot, Kenneth Fernandes, Sudhir Tauro, and Jean-Christophe Bourdon. δ160p53 is a novel n-terminal p53 isoform encoded by δ133p53 transcript. *FEBS letters*, 584(21):4463–4468, 2010.

[142] Carles Garcia-Linares, Juana Fernández-Rodríguez, Ernest Terribas, Jaume Mercadé, Eva Pros, Llúcia Benito, Yolanda Benavente, Gabriel Capella, Anna Ravella, Ignacio Blanco, et al. Dissecting loss of heterozygosity (loh) in neurofibromatosis type 1-associated neurofibromas: Importance of copy neutral loh. *Human mutation*, 32(1):78–90, 2011.

[143] Christine O'Keefe, Michael A McDevitt, and Jaroslaw P Maciejewski. Copy neutral loss of heterozygosity: a novel chromosomal lesion in myeloid malignancies. *Blood*, 115(14):2731–2739, 2010.

[144] CO Nordling. A new theory on the cancer-inducing mechanism. *British journal of cancer*, 7(1):68, 1953.

[145] Sundaresan Venkatachalam, Stuart D Tyner, Curtis R Pickering, Scott Boley, Leslie Recio, John E French, and Lawrence A Donehower. Is p53 haploinsufficient for tumor suppression? implications for the p53+/-mouse model in carcinogenicity testing. *Toxicologic pathology*, 29(1 suppl):147–154, 2001.

[146] Pablo Lapunzina and David Monk. The consequences of uniparental disomy and copy number neutral loss-of-heterozygosity during human development and cancer. *Biology of the Cell*, 103(7):303–317, 2011.

[147] Jan Smida, Daniel Baumhoer, Michael Rosemann, Axel Walch, Stefan Bielack, Christopher Poremba, Klaus Remberger, Eberhard Korsching, Wolfram Scheurlen, Christian Dierkes, et al. Genomic alterations and allelic imbalances are strong prognostic predictors in osteosarcoma. *Clinical Cancer Research*, 16(16):4256–4267, 2010.

[148] Blake C Ballif, Scott G Sulpizio, Richard M Lloyd, Sara L Minier, Aaron Theisen, Bassem A Bejjani, and Lisa G Shaffer. The clinical utility of enhanced subtelomeric coverage in array cgh. *American Journal of Medical Genetics Part A*, 143(16):1850–1857, 2007.

[149] Xing Dai, Wei Ma, Xijing He, and Rajiv Kumar Jha. Review of therapeutic strategies for osteosarcoma, chondrosarcoma, and ewing's sarcoma. *Medical Science Monitor*, 17(8):RA177–RA190, 2011.

[150] Anthony Rhoads and Kin Fai Au. Pacbio sequencing and its applications. *Genomics, proteomics & bioinformatics*, 13(5):278–289, 2015.

[151] Anand Patel, Richard Schwab, Yu-Tsueng Liu, and Vineet Bafna. Amplification and thrifty single-molecule sequencing of recurrent somatic structural variations. *Genome research*, 24(2):318–328, 2014.

[152] Michael S Lawrence, Petar Stojanov, Paz Polak, Gregory V Kryukov, Kristian Cibulskis, Andrey Sivachenko, Scott L Carter, Chip Stewart, Craig H Mermel, Steven A Roberts, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*, 499(7457):214–218, 2013.

[153] Dung T Le, Jennifer N Uram, Hao Wang, Bjarne R Bartlett, Holly Kemberling, Aleksandra D Eyring, Andrew D Skora, Brandon S Luber, Nilofer S Azad, Dan Laheru, et al. Pd-1 blockade in tumors with mismatch-repair deficiency. *New England Journal of Medicine*, 372(26):2509–2520, 2015.

[154] Gerd P Pfeifer, Mikhail F Denissenko, Magali Olivier, Natalia Tretyakova, Stephen S Hecht, and Pierre Hainaut. Tobacco smoke carcinogens, dna damage and p 53 mutations in smoking-associated cancers. *Oncogene*, 21(48):7435–7451, 2002.

[155] Gerd P Pfeifer, Young-Hyun You, and Ahmad Besaratinia. Mutations induced by ultraviolet light. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 571(1):19–31, 2005.

[156] Bioconductor:inferring somatic signatures from single nucleotide variant calls. https://bioconductor.org/packages/release/bioc/vignettes/SomaticSignatures/inst/doc/SomaticSignatures-vignette.html. Accessed: 2017-01-08.

[157] Serena Nik-Zainal, Ludmil B Alexandrov, David C Wedge, Peter Van Loo, Christopher D Greenman, Keiran Raine, David Jones, Jonathan Hinton, John Marshall, Lucy A Stebbings, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell*, 149(5):979–993, 2012.

[158] Rameen Beroukhim, Craig H Mermel, Dale Porter, Guo Wei, Soumya Raychaudhuri, Jerry Donovan, Jordi Barretina, Jesse S Boehm, Jennifer Dobson, Mitsuyoshi Urashima, et al. The landscape of somatic copy-number alteration across human cancers. *Nature*, 463(7283):899–905, 2010.

[159] Sylvie Mazoyer, Simon A Gayther, Maria A Nagai, Simon A Smith, Alison Dunning, Elizabeth J van Rensburg, Hans Albertsen, Ray White, and Bruce AJ Ponder. A gene (dlg2) located at 17q12-q21 encodes a new homologue of the drosophila tumor suppressor dlg-a. *Genomics*, 28(1):25–31, 1995.

[160] James J Morrow and Chand Khanna. Osteosarcoma genetics and epigenetics: Emerging biology and candidate therapies. *Critical Reviews™ in Oncogenesis*, 20(3-4), 2015.

[161] Michal Kovac, Claudia Blattmann, Sebastian Ribi, Jan Smida, Nikola S Mueller, Florian Engert, Francesc Castro-Giner, Joachim Weischenfeldt, Monika Kovacova, Andreas Krieg, et al. Exome sequencing of osteosarcoma reveals mutation signatures reminiscent of brca deficiency. *Nature communications*, 6, 2015.

[162] Lisa L Wang, Anu Gannavarapu, Claudia A Kozinetz, Moise L Levy, Richard A Lewis, Murali M Chintagumpala, Ramon Ruiz-Maldanado, Jose Contreras-Ruiz, Christopher Cunniff, Robert P Erickson, et al. Association between osteosarcoma and deleterious mutations in the recql4 gene in rothmund–thomson syndrome. *Journal of the National Cancer Institute*, 95(9):669–674, 2003.

[163] Nai-Kong V Cheung, Jinghui Zhang, Charles Lu, Matthew Parker, Armita Bahrami, Satish K Tickoo, Adriana Heguy, Alberto S Pappo, Sara Federico, James Dalton, et al. Association of age at diagnosis and genetic mutations in patients with neuroblastoma. *Jama*, 307(10):1062–1071, 2012.

[164] Janet E Olson, X Wang, VS Pankratz, ZS Fredericksen, Celine M Vachon, RA Vierkant, James R Cerhan, and Fergus J Couch. Centrosome-related genes, genetic variation, and risk of breast cancer. *Breast cancer research and treatment*, 125(1):221–228, 2011.

[165] Rohit Prakash, Yu Zhang, Weiran Feng, and Maria Jasin. Homologous recombination and human health: the roles of brca1, brca2, and associated proteins. *Cold Spring Harbor perspectives in biology*, 7(4):a016600, 2015.

[166] Nicholas Turner, Andrew Tutt, and Alan Ashworth. Hallmarks of'brcaness' in sporadic cancers. *Nature Reviews Cancer*, 4(10):814–819, 2004.

[167] Christopher J Lord and Alan Ashworth. Brcaness revisited. *Nature Reviews Cancer*, 2016.

[168] Ludmil B Alexandrov, Serena Nik-Zainal, David C Wedge, Samuel AJR Aparicio, Sam Behjati, Andrew V Biankin, Graham R Bignell, Niccolò Bolli, Ake Borg, Anne-Lise Børresen-Dale, et al. Signatures of mutational processes in human cancer. *Nature*, 500(7463):415–421, 2013.

[169] Ann-Marie Patch, Elizabeth L Christie, Dariush Etemadmoghadam, Dale W Garsed, Joshy George, Sian Fereday, Katia Nones, Prue Cowin, Kathryn Alsop, Peter J Bailey, et al. Whole–genome characterization of chemoresistant ovarian cancer. *Nature*, 521(7553):489–494, 2015.

[170] Thomas Helleday, Saeed Eshtad, and Serena Nik-Zainal. Mechanisms underlying mutational signatures in human cancers. *Nature Reviews Genetics*, 15(9):585–598, 2014.

[171] V Abkevich, KM Timms, BT Hennessy, J Potter, MS Carey, LA Meyer, K Smith-McCune, R Broaddus, KH Lu, J Chen, et al. Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *British journal of cancer*, 107(10):1776–1782, 2012.

[172] Melinda L Telli, Kristin C Jensen, Shaveta Vinayak, Allison W Kurian, Jafi A Lipson, Patrick J Flaherty, Kirsten Timms, Victor Abkevich, Elizabeth A Schackmann, Irene L Wapnir, et al. Phase ii study of gemcitabine, carboplatin, and iniparib as neoadjuvant therapy for triple-negative and brca1/2 mutation–associated breast cancer with assessment of a tumor-based measure of genomic instability: Precog 0105. *Journal of Clinical Oncology*, 33(17):1895–1901, 2015.

[173] Tatiana Popova, Elodie Manié, Guillaume Rieunier, Virginie Caux-Moncoutier, Carole Tirapo, Thierry Dubois, Olivier Delattre, Brigitte Sigal-Zafrani, Marc Bollet, Michel Longy, et al. Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with brca1/2 inactivation. *Cancer research*, 72(21):5454–5462, 2012.

[174] László Virág and Csaba Szabó. The therapeutic potential of poly (adp-ribose) polymerase inhibitors. *Pharmacological reviews*, 54(3):375–429, 2002.

[175] Michèle Rouleau, Anand Patel, Michael J Hendzel, Scott H Kaufmann, and Guy G Poirier. Parp inhibition: Parp1 and beyond. *Nature Reviews Cancer*, 10(4):293–301, 2010.

[176] Amir Sonnenblick, Evandro De Azambuja, Hatem A Azim Jr, and Martine Piccart. An update on parp inhibitors - moving to the adjuvant setting. *Nature reviews Clinical oncology*, 12(1):27–41, 2015.

[177] Mark J O'Connor. Targeting the dna damage response in cancer. *Molecular cell*, 60(4):547–560, 2015.

[178] Aniello Cerrato, Francesco Morra, and Angela Celetti. Use of poly adp-ribose polymerase [parp] inhibitors in cancer cells bearing ddr defects: the rationale for their inclusion in the clinic. *Journal of Experimental & Clinical Cancer Research*, 35(1):179, 2016.

[179] Joaquin Mateo, Suzanne Carreira, Shahneen Sandhu, Susana Miranda, Helen Mossop, Raquel Perez-Lopez, Daniel Nava Rodrigues, Dan Robinson, Aurelius Omlin, Nina Tunariu, et al. Dna-repair defects and olaparib in metastatic prostate cancer. *New England Journal of Medicine*, 373(18):1697–1708, 2015.

[180] Florian Engert, Michal Kovac, Daniel Baumhoer, Michaela Nathrath, and Simone Fulda. Osteosarcoma cells with genetic signatures of brcaness are susceptible to the parp inhibitor talazoparib alone or in combination with chemotherapeutics. *Oncotarget*, 5, 2016.

[181] Malcolm A Smith, Oliver A Hampton, C Patrick Reynolds, Min H Kang, John M Maris, Richard Gorlick, E Anders Kolb, Richard Lock, Hernan Carol, Stephen T Keir, et al. Initial testing (stage 1) of the parp inhibitor bmn 673 by the pediatric preclinical testing program: Palb2 mutation predicts exceptional in vivo response to bmn 673. *Pediatric blood & cancer*, 62(1):91–98, 2015.

[182] LO Rachel, Aiste McCormick, Asima Mukhopadhyay, Laura C Woodhouse, Madeleine Moat, Anna Grundy, Michelle Dixon, Angelika Kaufman, San Soohoo, Ahmed Elattar, et al. The use of ovarian cancer cells from patients undergoing surgery to generate primary cultures capable of undergoing functional analysis. *PloS one*, 9(3):e90604, 2014.

[183] Christopher J Lord and Alan Ashworth. Mechanisms of resistance to therapies targeting brca-mutant cancers. *Nature medicine*, 19(11):1381–1388, 2013.

[184] Guotai Xu, J Ross Chapman, Inger Brandsma, Jingsong Yuan, Martin Mistrik, Peter Bouwman, Jirina Bartkova, Ewa Gogola, Daniël Warmerdam, Marco Barazas, et al. Rev7 counteracts dna double-strand break resection and affects parp inhibition. *Nature*, 521(7553):541–544, 2015.

[185] L Galluzzi, E Vacchelli, J Michels, P Garcia, O Kepp, L Senovilla, I Vitale, and G Kroemer. Effects of vitamin b6 metabolism on oncogenesis, tumor progression and therapeutic responses. *Oncogene*, 32(42):4995–5004, 2013.

[186] Monkol Lek, Konrad Karczewski, Eric Minikel, Kaitlin Samocha, Eric Banks, Timothy Fennell, Anne O'Donnell-Luria, James Ware, Andrew Hill, Beryl Cummings, et al. Analysis of protein-coding genetic variation in 60,706 humans. *BioRxiv*, page 030338, 2016.

[187] Masahide Takahashi, Jerome Ritz, and Geoffrey M Cooper. Activation of a novel human transforming gene, ret, by dna rearrangement. *Cell*, 42(2):581–588, 1985.

[188] MASAHIDE Takahashi and GEOFFREY M Cooper. ret transforming gene encodes a fusion protein homologous to tyrosine kinases. *Molecular and cellular biology*, 7(4):1378–1385, 1987.

[189] Masatoshi Takeichi et al. Cadherin cell adhesion receptors as a morphogenetic regulator. *Science*, 251(5000):1451–1455, 1991.

[190] Lois M Mulligan. Ret revisited: expanding the oncogenic portfolio. *Nature Reviews Cancer*, 14(3):173–186, 2014.

[191] Diogo Fonseca-Pereira, Sílvia Arroz-Madeira, Mariana Rodrigues-Campos, Inês AM Barbosa, Rita G Domingues, Teresa Bento, Afonso RM Almeida, Hélder Ribeiro, Alexandre J Potocnik, Hideki Enomoto, et al. The neurotrophic factor receptor ret drives haematopoietic stem cell survival and function. *Nature*, 514(7520):98–101, 2014.

[192] Haiji Bo Li Gao, Yang Wang, Jing Zhang, and Minghua Zhu. Neurotrophic factor artemin promotes invasiveness and neurotrophic function of pancreatic adenocarcinoma in vivo and in vitro. *Pancreas*, 44(1):134, 2015.

[193] Takao Baba, Yosuke Sakamoto, Atsushi Kasamatsu, Yasuyuki Minakawa, Satoshi Yokota, Morihiro Higo, Hidetaka Yokoe, Katsunori Ogawara, Masashi Shiiba, Hideki Tanzawa, et al. Persephin: A potential key component in human oral cancer progression through the ret receptor tyrosine kinase-mitogen-activated protein kinase signaling pathway. *Molecular carcinogenesis*, 54(8):608–617, 2015.

[194] Matti S Airaksinen and Mart Saarma. The gdnf family: signalling, biological functions and therapeutic value. *Nature Reviews Neuroscience*, 3(5):383–394, 2002.

[195] Hironori Hayashi, Masatoshi Ichihara, Toshihide Iwashita, Hideki Murakami, Yohei Shimono, Kumi Kawai, Kei Kurokawa, Yoshiki Murakumo, Tsuneo Imai, Hiroomi Funahashi, et al. Characterization of intracellular signals via tyrosine 1062 in ret activated by glial cell line-derived neurotrophic factor. *Oncogene*, 19(39):4469–4475, 2000.

[196] Gabriella De Vita, Rosa Marina Melillo, Francesca Carlomagno, Roberta Visconti, Maria Domenica Castellone, Alfonso Bellacosa, Marc Billaud, Alfredo Fusco, Philip N Tsichlis, and Massimo Santoro. Tyrosine 1062 of ret-men2a mediates activation of akt (protein kinase b) and mitogen-activated protein kinase pathways leading to pc12 cell survival. *Cancer research*, 60(14):3727–3731, 2000.

[197] Jan Jacob Schuringa, Katarzyna Wojtachnio, Werner Hagens, Edo Vellenga, CH Buys, Robert Hofstra, and Wiebe Kruijer. Men2a-ret-induced cellular transformation by activation of stat3. *Oncogene*, 20(38):5350–5358, 2001.

[198] Maria Grazia Borrello, Elena Ardini, Laura D Locati, Angela Greco, Lisa Licitra, and Marco A Pierotti. Ret inhibition: implications in cancer therapy. *Expert opinion on therapeutic targets*, 17(4):403–419, 2013.

[199] Taranjit S Gujral, Vinay K Singh, Zongchao Jia, and Lois M Mulligan. Molecular mechanisms of ret receptor–mediated oncogenesis in multiple endocrine neoplasia 2b. *Cancer research*, 66(22):10741–10749, 2006.

[200] A Fusco, M Grieco, M Santoro, MT Berlingieri, S Pilotti, MA Pierotti, G Della Porta, and G Vecchio. A new oncogene in human thyroid papillary carcinomas and their lymph-nodal metastases. *Nature*, 328(6126):170–172, 1987.

[201] Young Seok Ju, Won-Chul Lee, Jong-Yeon Shin, Seungbok Lee, Thomas Bleazard, Jae-Kyung Won, Young Tae Kim, Jong-Il Kim, Jin-Hyoung Kang, and Jeong-Sun Seo. A transforming kif5b and ret gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. *Genome research*, 22(3):436–445, 2012.

[202] Daniela S Krause and Richard A Van Etten. Tyrosine kinases as targets for cancer therapy. *New England Journal of Medicine*, 353(2):172–187, 2005.

[203] AS Brooks, BA Oostra, and RMW Hofstra. Studying the genetics of hirschsprung's disease: unraveling an oligogenic disorder. *Clinical genetics*, 67(1):6–14, 2005.

[204] Lisa M Ellerby, Abigail S Hackam, Stephanie S Propp, H Michael Ellerby, Shahrooz Rabizadeh, Neil R Cashman, Mark A Trifiro, Leonard Pinsky, Cheryl L Wellington, Guy S Salvesen, et al. Kennedy's disease. *Journal of neurochemistry*, 72(1):185–195, 1999.

[205] John H Sipple. The association of pheochromocytoma with carcinoma of the thyroid gland. *The American Journal of Medicine*, 31(1):163–166, 1961.

[206] Stephen J Marx. Molecular genetics of multiple endocrine neoplasia types 1 and 2. *Nature Reviews Cancer*, 5(5):367–375, 2005.

[207] Naoya Asai, Toshihide Iwashita, Mutsushi Matsuyama, and Masahide Takahashi. Mechanism of activation of the ret proto-oncogene by multiple endocrine neoplasia 2a mutations. *Molecular and Cellular Biology*, 15(3):1613–1619, 1995.

[208] Dennis S Acton, Danny Velthuyzen, Cornelis JM Lips, and Jo WM Höppener. Multiple endocrine neoplasia type 2b mutation in human ret oncogene induces medullary thyroid carcinoma in transgenic mice. *Multistep tumorigenesis of medullary thyroid carcinoma in man and transgenic mouse models*, page 51, 2000.

[209] Richard T Kloos, Charis Eng, Douglas B Evans, Gary L Francis, Robert F Gagel, Hossein Gharib, Jeffrey F Moley, Furio Pacini, Matthew D Ringel, Martin Schlumberger, et al. Medullary thyroid cancer: management guidelines of the american thyroid association. *Thyroid*, 19(6):565–612, 2009.

[210] Kanako-Tanase Nakao, Takeshi Usui, Mayumi Ikeda, Yusuke Mori, Tetsuro Yamamoto, Sachiko-Tsukamoto Kawashima, Kazutaka Nanba, Akiko Yuno, Tamiko Tamanaha, Tetsuya Tagami, et al. Novel tandem germline ret proto-oncogene mutations in a patient with multiple endocrine neoplasia type 2b: Report of a case and a literature review of tandem ret mutations with in silico analysis. *Head & neck*, 35(12):E363–E368, 2013.

[211] Masahide Takahashi, Toshihide Iwashita, Massimo Santoro, Stanislas Lyonnet, Gilbert M Lenoir, and Marc Billaud. Co-segregation of men2 and hirschsprung's disease: the same mutation of ret with both gain and loss-of-function? *Human mutation*, 13(4):331, 1999.

[212] Rebecca L Margraf, David K Crockett, Patti MF Krautscheid, Ryan Seamons, Fernanda RO Calderon, Carl T Wittwer, and Rong Mao. Multiple endocrine neoplasia type 2 ret protooncogene database: Repository of men2-associated ret sequence variation and reference for genotype/phenotype correlations. *Human mutation*, 30(4):548–556, 2009.

[213] AN Rettew, ED Young, DC Lev, ES Kleinerman, FW Abdul-Karim, PJ Getty, and EM Greenfield. Multiple receptor tyrosine kinases promote the in vitro phenotype of metastatic human osteosarcoma cell lines. *Oncogenesis*, 1(11):e34, 2012.

[214] Jonas Anders, Svend Kjær, and Carlos F Ibáñez. Molecular modeling of the extracellular domain of the ret receptor tyrosine kinase reveals multiple cadherin-like domains and a calcium-binding site. *Journal of Biological Chemistry*, 276(38):35808–35817, 2001.

[215] Svend Kjær and Carlos F Ibáñez. Intrinsic susceptibility to misfolding of a hot-spot for hirschsprung disease mutations in the ectodomain of ret. *Human molecular genetics*, 12(17):2133–2144, 2003.

[216] Carla Colombo, Emanuela Minna, Maria Grazia Rizzetti, Paola Romeo, Daniele Lecis, Luca Persani, Piera Mondellini, Marco A Pierotti, Angela Greco, Laura Fugazzola, et al. The modifier role of ret-g691s polymorphism in hereditary medullary thyroid carcinoma: functional characterization and expression/penetrance studies. *Orphanet journal of rare diseases*, 10(1):1, 2015.

[217] Carine Segouffin-Cariou and Marc Billaud. Transforming ability of men2a-ret requires activation of the phosphatidylinositol 3-kinase/akt signaling pathway. *Journal of Biological Chemistry*, 275(5):3568–3576, 2000.

[218] Naihan Xu, Yuanzhi Lao, Yaou Zhang, and David A Gillespie. Akt: a double-edged sword in cell proliferation and genome stability. *Journal of oncology*, 2012, 2012.

[219] Sarah E Golding, Rhiannon N Morgan, Bret R Adams, Amy J Hawkins, Lawrence F Povirk, and Kristoffer Valerie. Pro-survival akt and erk signaling from egfr and mutant egfrviii enhances dna double-strand break repair in human glioma cells. *Cancer biology & therapy*, 8(8):730–738, 2009.

[220] Isabelle Sansal and William R Sellers. The biology and clinical relevance of the pten tumor suppressor pathway. *Journal of Clinical Oncology*, 22(14):2954–2963, 2004.

[221] Branden S Moriarity, George M Otto, Eric P Rahrmann, Susan K Rathe, Natalie K Wolf, Madison T Weg, Luke A Manlove, Rebecca S LaRue, Nuri A Temiz, Sam D Molyneux, et al. A sleeping beauty forward genetic screen identifies new genes and pathways driving osteosarcoma development and metastasis. *Nature genetics*, 47(6):615–624, 2015.

[222] Elena Arighi, Anna Popsueva, Debora Degl'Innocenti, Maria Grazia Borrello, Cristiana Carniti, Nina M Perala, Marco A Pierotti, and Hannu Sariola. Biological effects of the dual phenotypic janus mutation of ret cosegregating with both multiple endocrine neoplasia type 2 and hirschsprung's disease. *Molecular Endocrinology*, 18(4):1004–1017, 2004.

[223] Mircea Ivan, Jane A Bond, Maria Prat, Paolo Maria Comoglio, and David Wynford-Thomas. Activated ras and ret oncogenes induce over-expression of c-met (hepatocyte growth factor receptor) in human thyroid epithelial cells. *Oncogene*, 14(20):2417–2423, 1997.

[224] Nadia Dani, Martina Olivero, Katia Mareschi, Marjan Maria van Duist, Silvia Miretti, Sara Cuvertino, Salvatore Patanè, Raffaele Calogero, Riccardo Ferracini, Katia Scotlandi, et al. The met oncogene transforms human primary bone-derived cells into osteosarcomas by targeting committed osteo-progenitors. *Journal of Bone and Mineral Research*, 27(6):1322–1334, 2012.

[225] Lisa Mirabello, Meredith Yeager, Phuong L Mai, Julie M Gastier-Foster, Richard Gorlick, Chand Khanna, Ana Patiño-Garcia, Luis Sierrasesúmaga, Fernando Lecanda, Irene L Andrulis, et al. Germline tp53 variants and susceptibility to osteosarcoma. *Journal of the National Cancer Institute*, 107(7):djv101, 2015.

[226] Dirce Maria Carraro, Maria Aparecida Azevedo Koike Folgueira, Bianca Cristina Garcia Lisboa, Eloisa Helena Ribeiro Olivieri, Ana Cristina Vitorino Krepischi, Alex Fiorini de Carvalho, Louise Danielle de Carvalho Mota, Renato David Puga, Maria do Socorro Maciel, Rodrigo Augusto Depieri Michelli, et al. Comprehensive analysis of brca1, brca2 and tp53 germline mutation and tumor characterization: a portrait of early-onset breast cancer in brazil. *PloS one*, 8(3):e57581, 2013.

[227] Maria Isabel Waddington Achatz and Patricia Ashton-Prolla. Genetic counseling for tp53 germline mutations. In *p53 in the Clinics*, pages 327–343. Springer, 2013.

[228] Samuel A Wells Jr, Sylvia L Asa, Henning Dralle, Rossella Elisei, Douglas B Evans, Robert F Gagel, Nancy Lee, Andreas Machens, Jeffrey F Moley, Furio Pacini, et al. Revised american thyroid association guidelines for the management of medullary thyroid carcinoma: the american thyroid association guidelines task force on medullary thyroid carcinoma. *Thyroid*, 25(6):567–610, 2015.

[229] T Kamba and DM McDonald. Mechanisms of adverse effects of anti-vegf therapy for cancer. *British journal of cancer*, 96(12):1788–1795, 2007.

[230] Doron Lipson, Marzia Capelletti, Roman Yelensky, Geoff Otto, Alex Parker, Mirna Jarosz, John A Curran, Sohail Balasubramanian, Troy Bloom, Kristina W Brennan, et al. Identification of new alk and ret gene fusions from colorectal and lung cancer biopsies. *Nature medicine*, 18(3):382–384, 2012.

[231] Luca Mologni, Carlo Gambacorti-Passerini, Peter Goekjian, and Leonardo Scapozza. Ret kinase inhibitors: a review of recent patents (2012-2015). *Expert Opinion on Therapeutic Patents*, 2016.

[232] Ymera Pignochino, Giovanni Grignani, Giuliana Cavalloni, Manuela Motta, Marta Tapparo, Stefania Bruno, Alessia Bottos, Loretta Gammaitoni, Giorgia Migliardi, Giovanni Camussi, et al. Sorafenib blocks tumour growth, angiogenesis and metastatic potential in preclinical models of osteosarcoma through a mechanism potentially involving the inhibition of erk1/2, mcl-1 and ezrin pathways. *Molecular cancer*, 8(1):1, 2009.

[233] Klaus Mross, Annette Frost, Simone Steinbild, Susanne Hedbom, Martin Büchert, Ulrike Fasol, Clemens Unger, Jörn Krätzschmar, Roland Heinig, Oliver Boix, et al. A phase i dose–escalation study of regorafenib (bay 73–4506), an inhibitor of oncogenic, angiogenic, and stromal kinases, in patients with advanced solid tumors. *Clinical Cancer Research*, 18(9):2658–2667, 2012.

[234] Axel M Hillmer, Fei Yao, Koichiro Inaki, Wah Heng Lee, Pramila N Ariyaratne, Audrey SM Teo, Xing Yi Woo, Zhenshui Zhang, Hao Zhao, Leena Ukil, et al. Comprehensive long-span paired-end-tag mapping reveals characteristic patterns of structural variations in epithelial cancer genomes. *Genome research*, 21(5):665–675, 2011.

[235] Online course material for module bme215: Applied gene technology. https://classes.soe.ucsc.edu/bme215/Spring09/PPT/BME%20215-6.pdf. Accessed: 2016-12-21.

[236] Niranjan Nagarajan, Denis Bertrand, Axel M Hillmer, Zhi Jiang Zang, Fei Yao, Pierre-Étienne Jacques, Audrey SM Teo, Ioana Cutcutache, Zhenshui Zhang, Wah Heng Lee, et al. Whole-genome reconstruction and mutational signatures in gastric cancer. *Genome biology*, 13(12):1, 2012.

[237] Tiffany J Morris, Lee M Butcher, Andrew Feber, Andrew E Teschendorff, Ankur R Chakravarthy, Tomasz K Wojdacz, and Stephan Beck. Champ: 450k chip analysis methylation pipeline. *Bioinformatics*, 30(3):428–430, 2014.

[238] Lee M Butcher and Stephan Beck. Probe lasso: a novel method to rope in differentially methylated regions with 450k dna methylation data. *Methods*, 72:21–28, 2015.

[239] Fastqc: A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed: 2016-12-12.

[240] Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17(1):pp–10, 2011.

[241] Gerton Lunter and Martin Goodson. Stampy: a statistical algorithm for sensitive and fast mapping of illumina sequence reads. *Genome research*, 21(6):936–939, 2011.

[242] Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, et al. The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–2079, 2009.

[243] Andy Rimmer, Hang Phan, Iain Mathieson, Zamin Iqbal, Stephen RF Twigg, Andrew OM Wilkie, Gil McVean, Gerton Lunter, WGS500 Consortium, et al. Integrating mapping-, assembly-and haplotype-based approaches for calling variants in clinical sequencing applications. *Nature genetics*, 46(8):912–918, 2014.

[244] Helga Thorvaldsdóttir, James T Robinson, and Jill P Mesirov. Integrative genomics viewer (igv): high-performance genomics data visualization and exploration. *Briefings in bioinformatics*, 14(2):178–192, 2013.

[245] Kai Wang, Mingyao Li, and Hakon Hakonarson. Annovar: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*, 38(16):e164–e164, 2010.

[246] Heng Li and Richard Durbin. Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics*, 25(14):1754–1760, 2009.

[247] Julian S Gehring, Bernd Fischer, Michael Lawrence, and Wolfgang Huber. Somaticsignatures: inferring mutational signatures from single-nucleotide variants. *Bioinformatics*, page btv408, 2015.

[248] R Core Team. R: A language and environment for statistical computing. r foundation for statistical computing, vienna, austria. 2013, 2014.

[249] Sue Richards, Nazneen Aziz, Sherri Bale, David Bick, Soma Das, Julie Gastier-Foster, Wayne W Grody, Madhuri Hegde, Elaine Lyon, Elaine Spector, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the american college of medical genetics and genomics and the association for molecular pathology. *Genetics in Medicine*, 17(5):405–423, 2015.

[250] Xian Fan, Travis E Abbott, David Larson, and Ken Chen. Breakdancer: Identification of genomic structural variation from paired-end read mapping. *Current Protocols in Bioinformatics*, pages 15–6, 2014.

[251] Daniel Nicorici, Mihaela Satalan, Henrik Edgren, Sara Kangaspeska, Astrid Murumagi, Olli Kallioniemi, Sami Virtanen, and Olavi Kilkku. Fusioncatcher-a tool for finding somatic fusion genes in paired-end rna-sequencing data. *bioRxiv*, page 011650, 2014.

[252] Matteo Benelli, Chiara Pescucci, Giuseppina Marseglia, Marco Severgnini, Francesca Torricelli, and Alberto Magi. Discovering chimeric transcripts in paired-end rna-seq data by using ericscript. *Bioinformatics*, 28(24):3232–3239, 2012.