Gene-level consequences of new cancer-specific chromosomal rearrangements

Kaja Beate Nyquist
Oslo, September 2013

Thesis for the degree of Philosophiae Doctor (PhD)
Gene-level consequences of new cancer-specific chromosomal rearrangements

Kaja Beate Nyquist

University of Oslo

2013
My first love in science was the stars and planets, and for some years I dreamt about becoming an astronaut and exploring the universe above. In cancer research I found a starry sky in the FISH microscope, where bright signals in different colors resembled the stars I had dreamt about as a child. Even though we now know more about the components inside the body, sometimes it feels as if we have a whole universe within where only few stars have been fully characterized.
# Table of Contents

Table of Contents ............................................................................................................ 5  
Acknowledgements ......................................................................................................... 7  
List of papers................................................................................................................... 9  
Introduction ................................................................................................................... 11  
  Cancer and cytogenetics–from mysterious lumps to chromosomal barcodes ...... 11  
  The BCR-ABL story................................................................................................ 13  
  Targeted cancer therapy ......................................................................................... 14  
  Cytogenetic nomenclature...................................................................................... 15  
  Nomenclature of abnormal chromosomes .......................................................... 16  
  Consequences of balanced chromosomal translocations .................................... 17  
  Fusion genes ........................................................................................................... 18  
  Gene deregulation ................................................................................................. 20  
  Rearrangements without functional consequences ............................................. 20  
  Submicroscopic rearrangements around breakpoints .......................................... 21  
  Acute leukemias and sarcomas ............................................................................ 21  
Methods ....................................................................................................................... 25  
  Fluorescence in situ hybridization ...................................................................... 25  
  FISH with locus-specific probes derived from bacterial artificial chromosomes.. 26  
  Metaphase FISH ................................................................................................... 28  
  Interphase FISH .................................................................................................... 29  
  Molecular methods for identifying and characterizing breakpoint regions ....... 29  
  Long-distance PCR ............................................................................................... 30  
  Rapid amplification of cDNA ends...................................................................... 31  
  Quantitative PCR ................................................................................................. 31
Acknowledgements

The work presented in this thesis was carried out at the Section for Cancer Cytogenetics, Institute for Medical Informatics at the Norwegian Radium Hospital in Oslo. The studies were financed by grants from the South-Eastern Regional Health Authority and the Norwegian Cancer Society.

First of all, I would like express my gratitude to my supervisor Francesca Micci for her guidance and support through these years. You have always being enthusiastic and encouraging me in my work, even when we met hurdles along the way. I am grateful for the thoughtfulness you have had for me, both professionally and privately.

I am also very grateful to my co-supervisor professor Sverre Heim for introducing me to the world of cancer cytogenetics. Your immerse knowledge in the field has been a true inspiration, and you have opened a word of fuzzy black and white creatures I didn’t know existed. Thank you for your patience and constructive advice in the writing process, your contribution has been invaluable.

A special thanks to Jim Thorsen. I am grateful for your patience trying to teach me anything from how to optimize the PCR cycle to make figures with vector-based graphics. Your time and effort has been truly appreciated. Ioannis Panagopoulos who joined us in 2012, also deserves a note. Your impressive knowledge and experience in the field of fusion genes have been inspirational to me, thank you for sharing your knowledge.

I am grateful to all of my co-authors for each of their individual contributions. I want to mention in particular Bodil Bjerkehagen and Anne Tierens at the Pathology Department. Thank you for always being helpful and for taking the time to advise me in a busy clinical setting.

To all the girls at the lab; thank you for a pleasant working atmosphere and interesting discussions during lunch. A special thanks to Lisbeth Haugom and Anne Mette Eng Eibak. Thank you for teaching me the lab procedures, and for always being helpful.
Thank you to my former colleagues in the Ph.D forum- for expanding my view of the PhD experience and cancer research in general.

A warm thanks to my colleagues at the “stipendiat-office” Thank you for sharing frustrations and achievements, for inspiring conversations, for “English Wednesday” and “Bun-Friday”. It has been a pleasure to share office with you, and I know I will miss you.

Finally I want to thank my family and friends for taking interest in my work. To Karsten and Edla for being the best family I could dream of. A special though to my deceased mother in law Gabriella who put my view on cancer research perspective, and who I know would have loved to see this work finalized.

Kaja B Nyquist

Oslo, September 2013
List of papers

Paper I

Kaja B. Nyquist, Jim Thorsen, Bernward Zeller, Alf Haaland, Gunhild Trøen, Sverre Heim, Francesca Micci:

Identification of the TAF15-ZNF384 fusion gene in two new cases of acute lymphoblastic leukemia with a t(12;17)(p13;q12).

Cancer Genetics (2011), 204(3):147-152.

Paper II

Kaja B. Nyquist, Ioannis Panagopoulos, Jim Thorsen, Roberta Roberto, Hilde S. Wik, Anne Tierens, Sverre Heim, Francesca Micci:

t(12;13)(q14;q31) leading to HMGA2 upregulation in acute myeloid leukaemia.


Paper III


Whole-transcriptome sequencing identifies novel IRF2BP2-CDX1 fusion gene brought about by translocation t(1;5)(q42;q32) in mesenchymal chondrosarcoma.


Paper IV

Francesca Micci, Jim Thorsen, Ioannis Panagopoulos, Kaja B. Nyquist, Bernward Zeller, Anne Tierens, Sverre Heim:

High-throughput sequencing identifies an NFIA/CBFA2T3 fusion gene in acute erythroid leukemia with t(1;16)(p31;q24).

**Introduction**

*Cancer and cytogenetics—from mysterious lumps to chromosomal barcodes*

One of the first known descriptions of cancer is found in a papyrus which scientists have dated to approximately 2500 BC (Mukherjee, 2011). It is believed to be the teachings of the Egyptian physician Imhotep and describes patients with various medical conditions. One description tells about a woman with a hard, bulging mass of the breast, without fever or discharging fluids. This is most likely a description of a woman with breast cancer. In addition to written reports, solid evidence of cancer has been detected in mummified bodies with bone tumors (Mukherjee, 2011). These ancient findings imply the coexistence of man and neoplasia for as long as we have been here.

Throughout history, likewise, many theories of the origin of cancer have been proposed. Hippokrates (460-370 BC), referred to by many as the father of Western medicine, adhered to the dominant theory of his time believing that the body had four fluids or humors (black bile, yellow bile, phlegm, and blood) and that cancer was caused by an imbalance affecting predominantly the black bile (American Cancer Society, 2009). This humoral theory more or less held sway until the mid nineteenth century when the German pathologist Rudolph Virchow established cellular pathology as the new paradigm of disease, and especially cancerous diseases. One of Virchow’s assistants, David Hansemann, studied mitoses in several human tumor types and documented mitotic asymmetries in most of them; this he realized might point to a mechanism of carcinogenesis. This line of thinking matured into the somatic mutation theory of cancer suggested by Theodor Boveri in 1914. His theory was based on studies of fertilized sea-urchin eggs, originally enough, but led to our current standard understanding of the principles of tumorigenesis (Balmain, 2001). Boveri observed that eggs fertilized with two sperm cells sometimes developed an unequal distribution of chromosomes in the daughter cells. Some of these cells died, but some lived on with an abnormal appearance. Boveri therefore suggested that individual chromosomes must possess different abilities, and that some abnormal combinations of
chromosomes could create abnormal growth. In retrospect, his assumptions fit surprisingly well with the theory of proto-oncogenes and tumor suppressor genes discovered over 70 years later (Balmain, 2001). It is today generally accepted that cancer is a genetic disease caused by DNA sequence changes (Stratton et al, 2009).

The next major contribution to cancer genetics came in 1960 when Nowell and Hungerford identified a minute chromosome, called the Philadelphia chromosome, in the bone marrow cells of patients suffering from chronic myeloid leukemia (CML) (Nowell and Hungerford, 1960b). This was the first specific chromosome abnormality to be detected in cancer. Cytogenetic analysis experienced a boom in the 1970s after the development of banding protocols for human chromosomes (Caspersson et al, 1968). Until then chromosomes had been arranged according to their size in 23 matching pairs and grouped after the location of their centromere (Trask, 2002). The new staining techniques revealed dark and light bands across the arms of the chromosomes, and these soon became the bands of reference for classifying normal and aberrant chromosomes. Modern cancer cytogenetics is an important tool for an initial screening of the cancer genome, both in diagnostics and research. Cytogenetic studies have greatly improved our understanding of carcinogenesis of especially hematological and mesenchymal neoplasias (Heim and Mitelman, 2009).
The *BCR-ABL* story

![Diagram showing the formation of the BCR-ABL fusion gene](image)

*Figure 1:* The formation of the *BCR-ABL* fusion gene is caused by a translocation of the long arm of chromosome 9 to chromosome 22. Below the schematic presentations is a photo of the normal chromosome 22 and the smaller “Philadelphia” chromosome.

Nowell and Hungerford’s finding, the so called “Philadelphia chromosome” in patients with CML (Nowell and Hungerford, 1960b) (Figure 1), turned out to be one of the most seminal in the history of cancer genetics. This chromosome, named after the city in which it was discovered, is specific for chronic myeloid leukemia. In the early 1970s, it was discovered that it resulted from a translocation between chromosomes 9 and 22 (Rowley, 1973, Goldman, 2010). On chromosome 9, the *ABL1* gene was involved (de Klein *et al*, 1982). In 1984, the breakpoint on chromosome 22 was shown to occur within a 5.8 Mb area called the breakpoint cluster region (today the *BCR* gene) (Groffen *et al*, 1984). The fusion of these two genes results in a qualitatively new gene that encodes an aberrant tyrosine kinase which is crucial for CML development (Lugo *et al*, 1990, Deininger *et al*, 2000). Based on this knowledge,
Druker and colleagues searched for compounds that could specifically bind and inhibit the *BCR-ABL1* protein (Druker *et al*, 1996). Their finding, a tyrosine kinase inhibitor (TKI) called imatinib, has since revolutionized treatment of CML patients (Druker, 2002). Although imatinib does not cure the patients of their disease, it gives long-term remissions without significant side effects in the majority of patients, and it is today the first choice of treatment for newly diagnosed patients with chronic phase CML (Baccarani *et al*, 2006).

**Targeted cancer therapy**

Targeted therapy aims to interfere with a molecular target essential for cancer cell growth and survival. In contrast to classical chemotherapy, targeted therapy aims to kill cancer cells more selectively and thus avoid the adverse side effects and general toxicity of cytostatics (Sawyers, 2004). A potential target for therapy should be (1) necessary and sufficient for cancer cell survival, (2) expressed and active in the tumor, (3) inhibiting effector molecules for which no possible circumventions in the implied pathway is possible (Borden *et al*, 2003). Inhibition with imatinib of the mutant kinase chimeric protein BCR-ABL1 in CML is the most successful example of this approach, even though resistance ultimately will develop (see above). However, in most other cancers where TKIs have been tested, patients develop resistance within a shorter time frame (Sierra *et al*, 2010). Efforts are now being made to try to understand the mechanisms underlying resistance both *in vitro* and *in vivo*. 
Cytogenetic nomenclature

Tjio and Levan detected in the mid 1950s (Tjio and Levan, 1956) the correct human chromosome number of 46 (44 autosomes and 2 sex chromosomes) after it had been believed for many years that 48 was the correct number. In the years after, research in this field expanded and a common system for cytogenetic nomenclature was needed. The first generally accepted cytogenetic nomenclature was established at a meeting in Denver in 1960, and this became the forerunner of today’s International System for Cytogenetic Nomenclature (ISCN) (Shaffer et al, 2009). Cytogenetic nomenclature is based upon the chromosomal banding pattern and the occurrence of other specific, distinct morphological hallmarks. Chromosomes are arranged by their size and the location of the centromeres. Accordingly, chromosome 1 is the largest chromosome, chromosome 2 the second largest and so on, with one exception, that chromosome 21 is smaller than chromosome 22 (Shaffer et al, 2009). The chromosomal bands appear by the differential staining throughout the chromosomes and are defined as “a part of a chromosome that is clearly distinguishable from its adjacent segments by appearing darker or lighter with one or more banding techniques”(Shaffer et al, 2009). The shorter chromosome arm is designated the “p” arm, while the longer arm is called the “q” arm. The chromosomal location 2p13 therefore designates chromosome 2, the short arm, region 1 within that arm, and band 3 within that region (Figure 2).

![Graphic presentation of a chromosome](image)

**Figure 2:** Graphic presentation of a chromosome. See text for details. From: Essential Haematology, 6th Edn. © Hoffbrand A.V. and Moss P.A.H., Blackwell Publishing, 2011 (with permission)
Nomenclature of abnormal chromosomes

Chromosomal aberrations are divided in two groups after their nature: structural and numerical. Structural aberrations are rearrangements of one or more chromosomes, whereas numerical aberrations refer to copy number changes (Heim and Mitelman, 2009). The most common structural aberrations are translocations, deletions, and inversions. They can be balanced, with no net gain or loss of genetic material, or unbalanced, implying unequal exchange of genetic material. Several other abnormalities can be seen and a specific terminology is used for karyotypic descriptions. A list of cytogenetic abbreviations can be found in Table 1. Only clonal chromosomal abnormalities are reported in karyotypic descriptions. A clone is defined as a group of cells stemming from a single progenitor cell, and at least two cells with the same structural aberration must be detected in order to infer the existence of a clone when doing cytogenetic analyses (Shaffer et al, 2009). If the aberration is numerical, that is loss or gain of a whole chromosome, two cells are needed for a supernumerary chromosome to be considered clonal, and three cells if the chromosome is lost. Clonally related cells share one or more aberrations, but additional abnormalities can be detected as a part of clonal evolution.
### Table 1: Cytogenetic abbreviations relevant for this thesis

<table>
<thead>
<tr>
<th>Rearrangement</th>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addition</td>
<td>add</td>
<td>Addition of material of unknown chromosomal origin</td>
</tr>
<tr>
<td>Composite</td>
<td>cp</td>
<td>Composite karyotype describes all clonally occurring abnormalities in a cytogenetically heterogeneous tumor</td>
</tr>
<tr>
<td>Deletion</td>
<td>del</td>
<td>Deletion (terminal or interstitial); loss of chromosomal material seemingly including or not including the tip of the chromosome arm</td>
</tr>
<tr>
<td>Derivative</td>
<td>der</td>
<td>Structurally rearranged chromosome</td>
</tr>
<tr>
<td>Marker</td>
<td>mar</td>
<td>Abnormal chromosome that cannot be identified</td>
</tr>
<tr>
<td>Translocation</td>
<td>t</td>
<td>Interchanges of material between two non-homologous chromosomes</td>
</tr>
</tbody>
</table>

**Consequences of balanced chromosomal translocations**

Balanced chromosomal translocations involve swapping of material between two chromosomes with no net gain or loss of chromosomal material. Two main pathogenetic consequences may in the cancer context follow from such translocations: either two genes are fused with each other and a qualitatively new hybrid or fusion gene is formed, or one gene loses or gains regulatory elements leading to its aberrant expression (Mitelman et al, 2007) (Figure 3). Translocations can also be present without any known consequences. Some of these are probably non-recurrent changes seen in unstable cancer genomes, but others might involve as yet undiscovered mechanisms leading to growth advantages for the cancer cell.
**Figure 3:** The two main consequences of chromosomal translocations are illustrated; the formation of a fusion gene and the dysregulation of a normal gene. The bars indicate exons while the circle illustrates a regulatory element.

**Fusion genes**
Fusion genes are hybrid genes formed by the fusing together of two previously separate genes. If joined properly, the new gene may give rise to a new chimeric protein with novel and oncogenic properties different from those of the native proteins. Based on cytogenetic studies, it has been estimated that fusion genes account for about 20% of cancer morbidity (Mitelman *et al.*, 2007), but this may well be only the tip of the iceberg. Most known fusion genes are found in leukemias, lymphomas and mesenchymal tumors, while fewer have so far been found in the more common epithelial cancers. This does not necessarily mean that fusion genes in fact are more common in these rare tumor types, it may rather reflect the number of cytogenetically analyzed cases or the ease with which fusion genes are found in different settings (Mitelman *et al.*, 2004). Studies relying on new investigative technologies are now revealing fusion genes on a much larger scale than was previously acknowledged,
especially in epithelial cancers (Edwards, 2010). The number of known fusion genes in cancer has increased from about 100 in the year 2000 (Mitelman, 2000) to 1603 as of August 14, 2013 (Mitelman et al, 2013). Recurrent fusion genes have now been detected also in common epithelial cancers such as carcinoma of the prostate (Tomlins et al, 2005), lung (Soda et al, 2007), breast (Stephens et al, 2009), and colon (Bass et al, 2011). Some of the cases analyzed revealed multiple fusion genes within the same tumor, with hitherto unknown oncogenic potential (Maher et al, 2009). In mesenchymal tumors, a fusion gene is often found in a high proportion of cases within each diagnostic entity. By comparison, the fusion genes identified so far in carcinomas have been found in a much lower percentage of cases.

Several aspects make the fusion genes of cancer interesting to study and some of them will be discussed below. For many of the well-described fusion genes there is convincing evidence for an important role in cancer cell survival and initiation (Mitelman et al, 2007). Identification of new fusion genes may therefore locate new players in the complex process of carcinogenesis. In addition, the detection in a tumor sample of a specific fusion gene can help resolve diagnostic dilemmas. An example is the finding of FUS-DDIT3 in a lipomatous tumor. This means that the tumor is a myxoid liposarcoma resulting from a t(12;16)(q13;p11); over 80% of myxoid liposarcomas harbor this translocation (Fletcher et al, 2013). This is not a unique example; in fact, most fusions are specific for certain cancer types. Certain gene fusions occur across several cancer types. ETV6-NTRK1 is such a fusion gene, which has been detected in secretory breast cancer (Tognon et al, 2002), congenital fibrosarcoma (Knezevich et al, 1998), and acute myeloid leukemia (Eguchi et al, 1999). Phenotypic tumor classifications will therefore always be needed. Furthermore, and maybe most interestingly, the fusion genes may represent therapeutic targets unique to the tumor in question. The main problem with current chemotherapeutics is their lack of specificity. The drugs kill all dividing cells, causing severe side-effects to the patient and reducing the dosage the patients can tolerate. Again, the imatinib story demonstrates the success of this approach. Imatinib treatment has increased survival and is generally well tolerated by CML patients (Druker, 2008).
Gene deregulation

If a chromosomal rearrangement does not lead to the formation of a fusion gene in a cancer, chances are good that it juxtaposes regulatory sequences from one gene with the transcribing parts of the other gene, which usually is a proto-oncogene, and therefore gives rise to the latter gene’s activation. As the DNA structure is three-dimensional, also promoters far from the translocation breakpoint can affect gene transcription. One well-known example of gene deregulation following a translocation is the juxtaposition of the proto-oncogene MYC, from 8q24, to the vicinity of an immunoglobulin gene. This deregulation is typical of Burkitt lymphoma (Hecht and Aster, 2000) and leads to constitutive overexpression of MYC. In other cases, genes at translocation breakpoints may be inactivated by the chromosomal rearrangement. This seems to be rare, and the functional consequences are often difficult to predict (Mitelman et al, 2007), however the result could be loss of tumor suppressor function. The affected genes can either be translated into truncated transcripts or be degraded as a consequence of the cellular quality control system for example by a mechanism called nonsense-mediated decay that recognizes preterm stop-codons (Chang et al, 2007). The result would be an aberrant protein with unknown functional consequences, in the former case, and lack of the protein in the latter.

The human miRNA genes, which are small, non-coding RNA molecules with influence on transcriptional and post-transcriptional regulation of gene expression, are non-randomly located at hot spots for chromosomal rearrangements (Calin et al, 2004). Translocations in these areas may lead to deregulation of protein coding genes in a manner similar to that seen in lymphomas (Calin and Croce, 2007). Also, genes can contain binding sites for miRNAs with repressive functions. Disruption of these sites through chromosomal translocations may lead to overexpression of the affected genes. An example of this mechanism is found in translocations involving 12q13-15 affecting the HMGA2 gene (Mayr et al, 2007) (paper II).

Rearrangements without functional consequences

Rearrangements may also occur in non-coding areas of DNA with no functional gene alteration resulting. The rarity of reports describing this phenomenon may be due to
publication bias, i.e., journals are less likely to publish negative results. The possibility is well illustrated by individuals carrying a balanced translocation with apparently normal phenotypes. Approximately 1 of 1000 newborns are estimated to carry such a translocation (Ferlin et al., 2007), and because all transcribed genes are intact, no functional consequence is apparent. The translocation might first be detected when an affected individual turns out to be infertile or have spontaneous abortions because of a higher likelihood of genomically unbalanced offspring.

**Submicroscopic rearrangements around breakpoints**

Not all seemingly balanced translocations turn out to be exactly balanced when analyzed at the base pair level. Deletions (Sinclair et al., 2000), inversions (Desmaze et al., 1997, Micci et al., 2006), and duplications (Howarth et al., 2011) in and around the translocation breakpoints may accompany what looks cytogenetically as balanced changes, both in malignant diseases and in constitutional translocations. This possibility should be kept in mind when searching for fusion genes; it could account for otherwise difficult-to-explain findings made during the study. In addition, the true prevalence of this phenomenon might be underestimated as such a rearrangement makes the characterization of the breakpoint more difficult.

**Acute leukemias and sarcomas**

Hematological and mesenchymal malignancies were the targets of studies leading up to this thesis. Both originate from the mesoderm and in both types of neoplasia, balanced translocations leading to fusion genes are often identified (Mitelman et al., 2007).

The term leukemia (from Greek leucos: white, and haima: blood) was first used by the German pathologist Rudolf Virchow in 1847 (Kampen, 2012). He believed the disease to be caused by too many white blood cells. Today, leukemia is known to be a clonal neoplastic disease of hematopoietic cells. By way of operational definition, acute leukemia is said to be present when the bone marrow contains >20 percent immature cells or blasts. Depending on the leukemic cell lineage commitment, the acute leukemias are divided into acute lymphoblastic leukemia (ALL) if the cells resemble
immature lymphatic cells (in which case they are assumed to originate from cells of lymphoid origin) and, correspondingly, acute myeloid leukemia (AML) if made up of immature myeloid cells (Swerdlow, 2008). The term “acute” in contrast to “chronic” originally referred to the rapid course of the acute form of the disease; in current times, it reflects the immature nature of the cells of the acute leukemias. Acute leukemia can be fatal within days to weeks if not adequately treated. Signs of leukemia are caused by the replacement of the normal bone marrow cells by a malignant leukemic clone, and include manifestations brought about by anemia, thrombocytopenia, and neutropenia. As a consequence, patients with leukemia usually present with fatigue, bone pain, headache, infections, bleedings or other unspecific signs and symptoms.

Leukemias have in later decades been classified according to the FAB (French-American-British) system first launched in 1976 (Bennett et al., 1976) which is based on morphological and cytochemical leukemic cell features. Since 2001, the World Health Organization (WHO) classification has largely superseded previous classifications (Swerdlow, 2001). In addition to cell morphology, the WHO system includes immunophenotype and genetic characteristics in the classification.

ALL is divided in two main subgroups, those expressing B cell and those expressing T cell features, whereas AML is divided in six subgroups depending on the maturation level and the predominantly affected cell type. While ALL is primarily a disease of children with most cases being diagnosed in patients under 18 years of age (Swerdlow, 2008), AML is much more common in adults with a median age at diagnosis of about 70 years (Smith et al., 2011).

Known risk factors for the development of leukemias, especially the myeloid ones, include exposure to organic solvents such as benzene (Galbraith et al., 2010) and previous treatment with chemotherapeutic drugs (Czader and Orazi, 2009), but in the majority of cases no etiologic factor can be identified (Zeeb and Blettner, 1998).

When treating patients with acute leukemia, the first goal is to obtain a complete remission using chemotoxic agents. Whether this can be accomplished depends on how well the patient tolerates the induction treatment, their performance status at
diagnosis, co-morbidities, and biological properties of the neoplastic disorder. Cytogenetic abnormalities, along with other criteria, determine the risk classification and therefore indirectly influence the further treatment in each group (Vardiman et al., 2009). In the medium and high risk groups, hematopoietic cell transplantation during first remission is an option for young patients with good performance status, but the treatment is dependent of the existence of a HLA-matched donor and associated with several adverse side effects.

Sarcomas (from Greek sarx: flesh, and oma: tumor) are tumors replicating connective tissue architectural features and which presumably originate from cells of mesenchymal origin. They are anatomically subdivided into bone and soft tissue tumors (Fletcher et al., 2013). Sarcomas occur in all age-groups and usually present as a hard, painless swellings. They account for only 1% of human malignant neoplasms (Fletcher et al., 2013). Sarcomas can be further subdivided into over 70 different histological subtypes according to the WHO (Fletcher et al., 2013). Another way of classifying sarcomas, in principle independently of their histological appearance and tissue of origin, is based on their cytogenetic changes (Taylor et al., 2011). Tumors in the first group harbor simple chromosomal rearrangements or near-diploid karyotypes while those of the other group typically display complex karyotypes with both numerical and structural rearrangements. In sarcomas with simple karyotypes (about a third of all sarcomas), fusion genes are often detected (Mitelman, 2000). Some of these fusion genes are tightly linked to a histological disease entity, as for example Ewing sarcoma. In this malignancy, the majority of cases display a t(11;22) leading to a fusion between the EWSR1 and the FLI1 genes (Delattre et al., 1992, Fletcher et al., 2013). Some sarcomas harbor instead activating point mutations in key genes, with gastrointestinal stromal tumors (GISTs) as the most thoroughly studied example. The majority of GISTs have a mutated tyrosine kinase gene which renders them susceptible to TKI treatment (Hirota et al., 1998, Heinrich et al., 2003).

Most sarcomas arise without any known etiology, but some genetic syndromes such as the Li-Fraumeni syndrome (Li et al., 1988) and neurofibromatosis type I (Ferner and
Gutmann, 2002) predispose to sarcoma development, as does previous radiation therapy (Bjerkehagen et al, 2008).

Limb-sparing surgery can be sufficient for cure in some patients with small tumors and no metastasis at diagnosis, but for most patients, multimodal approaches with combinations of chemotherapy and radiotherapy together with surgery are offered (Fletcher et al, 2013). Due to the rarity of sarcomas and especially tumors of individual subgroups, even multicenter trials spend years to include enough patients to conduct randomized trials for evaluation of new treatments. Therefore, it is difficult to obtain well-founded protocols. In research settings, it is possible to try new drugs based on a theoretical rationale for example in terminal ill patients. The first GIST patient treated with imatinib (Joensuu et al, 2001) demonstrated the potential success of this approach. A shift from a classification based on histological appearance to a genetic signature in these rare cancers may facilitate the introduction of new drugs.
Methods

The techniques described here are the main methods used in the studies behind this thesis. All methods have advantages as well as limitations, and some of them will be discussed in the following, together with a brief overview of the theoretical background underlying each method.

Fluorescence in situ hybridization

Fluorescence in situ hybridization (FISH) is a procedure based on the principle that complimentary DNA or RNA sequences hybridize to each other. The first hybridization using radioactively labeled DNA probes was described in the late 1960s (Pardue and Gall, 1969). Further development of hybridization techniques came in the early 1980s, after the discovery of the usefulness of fluorescent tags for labeling. These dyes are safer and simpler to handle, furthermore they can be stored for longer periods (Trask, 2002). These advantages led to several new applications for the method: characterization of chromosomal rearrangements, detection of microdeletions, and prenatal diagnosis of common congenital aneuploidies (Fan, 2002).

Three main types of FISH probes exist: locus-specific, whole chromosome painting, and repetitive sequence probes. Locus-specific probes hybridize to a defined area of interest in the genome. Whole chromosome painting probes contain a pool of DNA fragments all hybridizing on the same chromosome so that it appears fluorescent and can be easily identified in a metaphase. The repetitive sequence probes map to an amplified sequence in the genome. An example of the latter is the centromeric probe targeting pericentromeric sequences. They are well suited for detection of numerical aberrations either in interphase nuclei or if metaphase quality is poor (Moyzis et al, 1987).

The resolution of the experimental FISH approaches rapidly improved. By 1985 the first human gene, thyroglobulin, was mapped to a chromosome using cosmid clones of about 20 kb (Landegent et al, 1985). Improvements in software, hardware, probe technology, and reduced costs had a great impact on implementing the method, and today probes for FISH can detect areas down to 1 kb (Speicher and Carter, 2005).
One of the fascinating technological developments within the family of FISH methods is the so-called multicolor FISH. It is a whole genome screening technique using human chromosome painting probes to distinguish individual chromosomes. The spectral signature of each probe identifies the 24 different human chromosomes. Diverse imaging systems are available as, by way of example, M-FISH (Speicher et al, 1996) and Spectral karyotyping (Schrock et al, 1996). The method enables a relatively simple way of analyzing complex karyotypic rearrangements (Lee et al, 2001, Trask, 2002).

Another application of the FISH techniques came in 1992 through studies made by the group of Kallioniemi (Kallioniemi et al, 1992). The method is called comparative genomic hybridization (CGH) and is based on a competitive hybridization of tumor and normal DNA, labeled with different colors (green and red, respectively). The labeled DNA is then hybridized to normal metaphases. By measuring the green-to-red ratio it is possible to identify imbalances present in the tumor genome. Regions gained/amplified in cancer will appear green, whereas regions lost will appear red. The resolution of this chromosomal-based approach is about 10 Mb. Even higher resolution is obtained using array-based CGH. The first arrays had a resolution of about 40 kb (Pinkel et al, 1998). In this variant of competitive hybridization, tumor and normal DNA targets well-characterized probes attached to a glass slide. Later technological improvements made it possible to reduce the probe size and increase the probe number so that today all known exon sequences can be evaluated in one single hybridization experiment (Kapur et al, 2007). The progress of FISH methods elegantly shows the transition from classical cytogenetics to molecular analyses.

**FISH with locus-specific probes derived from bacterial artificial chromosomes**

DNA constructs of various sizes can be transfected to bacteria that function as stable vectors. Bacterial artificial chromosomes (BACs) are such constructs of unique sequences of about 150-300 kb in size (Shizuya et al, 1992). The BAC clones commercially available are either end-sequenced or fully sequenced. Clones mapping to the region of interest can easily be identified using different databases such as the UCSC Genome Browser (available at [http://genome.ucsc.edu/](http://genome.ucsc.edu/)). The BACs used in this
thesis were retrieved from the RPCI-11 library (Osoegawa et al, 2001) and the CalTech human BAC library (https://bacpac.chori.org/). As several genes can map within one BAC clone, the discrimination among them can be performed by FISH using smaller fosmid clones (Kim et al, 1992). After culturing of BAC and/or fosmid clones on an agar plate including antibiotics (the host bacteria have an antibiotic-resistant insert), the plasmids can be extracted by lysis of the bacterial cell membrane and separation of proteins and bacterial DNA. To label the extracted DNA, we used nick translation (Rigby et al, 1977), a technique whereby a fraction of the dNTPs in the DNA target strand is substituted with fluorescently marked nucleotides using DNase and DNA polymerase 1. Denaturation of the slide and probe can be done either by heat or using chemicals such as formamide. After denaturation of both probe and target DNA on the slide, the probe will anneal to its target sequence because of its presence in higher numbers. After incubation in a humid chamber, a series of washes remove unspecific hybridization and slides are counterstained, typically using 4′,6-diamidino-2-phenylindole (DAPI). The result of the hybridization is evaluated using a fluorescent microscope (Figure 4).
**Methods**

**Figure 4:** Overview of the FISH workflow

**Metaphase FISH**
The location of locus-specific FISH probes can be determined looking at only one metaphase. In a normal diploid cell, two signals will appear using a locus-specific probe. If a chromosomal break occurs in the sequence covered by the probe, three signals will be detected, one from the normal chromosome and one from each of the derivative chromosomes corresponding to a splitting of the second signal. In this manner, candidate genes involved in translocations can be detected by FISH. However, if the break occurs close to the end of the probe sequence, the extra signal might be too small to be detected. Metaphase FISH requires cell culturing and is dependent of the growth of malignant cells *in vitro*. If an aberration is recognized by karyotyping and material is sparse, it is possible to destain the very same slide and utilize it for FISH (Teixeira *et al*, 2000, Micci *et al*, 2002). Repeated hybridizations to the same slide testing different BAC probes are also possible (Epstein *et al*, 1995).
**Interphase FISH**

FISH experiments performed on interphase cells do not require cell culturing which may at times be a major advantage. FISH can then be performed on paraffin embedded material (Werner *et al.*, 1997). Interphase FISH (IP-FISH) is often used for the detection of numerical aberrations as well as amplifications of suspected targets (Speicher and Carter, 2005). Commercially available kits exist for leukemias and solid tumors that can detect the most common aberrations. Because many cells are examined, this method can detect low numbers of genetically distinct subclones (Raimondi, 2000, Teixeira, 2002). But since the technique may be prone to false positive and/or false negative results, it is recommended to count a large number of nuclei and to test the specificity of the probe on a normal slide. In our laboratory, signals from 200 nuclei are counted by two independent observers and an abnormal signal pattern present in >10 % of cells is considered significant (Brandal *et al.*, 2004). The probe design can reduce the number of false positive signals, for example by designing a double fusion probe. IP-FISH is, however, not suitable for breakpoint detection because of its inability to determine the precise probe localization.

Some limitations in the use of FISH on cancer cell preparations must be pointed out. Unspecific hybridization to both intrachromosomal and extrachromosomal areas may in some instances resemble specific hybridization and confuse the researcher. This limitation must be kept in mind especially when interpreting IP-FISH results. To overcome this problem, the procedure should be optimized and control tests performed in order to get reliable results. Prior to hybridization experiments, the target slides must be treated to remove proteins and other components that could interfere with probe hybridization.

**Molecular methods for identifying and characterizing breakpoint regions**

The minimal size of a FISH probe in regular fluorescence microscopy is about 40 kb. This resolution level is sufficient to identify genes involved in a rearrangement, but does not shed light on which parts of the genes are involved. To analyze in detail the breakpoint position, one has to resort to molecular methods. All methods described below are, to some extent, based on the polymerase chain reaction (PCR), first
described in a clinical setting by Saiki (Saiki et al., 1985) after its development by Kary Mullis (Bartlett and Stirling, 2003). The wide implementation of this method has revolutionized analyses in molecular biology. The PCR reaction is a method for amplifying double stranded DNA using sequential heating and cooling steps. Key elements in the reaction are primers, DNA polymerase, and nucleotides. Selective and repeated amplification is enabled by the sequence specific primers and the tolerance of the polymerase enzyme of high temperatures. The procedure continues in an exponential fashion using the newly formed DNA as template. The two used PCR primers are oriented in opposite directions and typically amplify a fragment of about 500-600 bp. In cases with a low number of target templates, a modified version of the PCR reaction called *nested* PCR can be applied. In this protocol, the sensitivity of the reaction is increased by running a second PCR with primers placed within the ones previously used. The procedure can be helpful in detecting rare mutations in a heterogeneous cell population. PCR based methods are extremely sensitive, therefore great care must be taken to avoid DNA contamination. PCR reactions searching for rare fusion genes are, however, less likely to produce false positives because of the rarity of the fusion genes we look for. Nonspecific binding of the primers to DNA or RNA can yield unspecific bands on PCR. To avoid this problem, careful primer design is important. The primers should among other factors be unique in the genome, should not tend to form secondary structures and the primer set should have a similar melting point (Sambrook and Russel, 2001).

**Long-distance PCR**
Under standard PCR reaction conditions, fragments up to two kb are amplified (Sambrook and Russel, 2001). Using modifications of the PCR protocol, genomic areas of up to 35 kb can be amplified (Barnes, 1994). To enable the synthesis of such long products, optimization of extension time, primer design, buffers, template quality, and the performance of the polymerase is necessary to obtain a good result. Given two candidate genes likely to take part in a translocation, long-distance PCR (LD-PCR) can be a good procedure to detect fusion genes, run with genomic DNA as the template. Primers are designed for each gene/exon and, if successful, the fusion gene is
amplified. If no bands occur, no information is obtained about the location of the rearrangement. Therefore, this method requires a clear idea as to where the breakpoints are situated.

**Rapid amplification of cDNA ends**

Rapid amplification of cDNA ends (RACE-PCR) can be applied if RNA is available and one of the fusion partners is identified. This method enables amplification of the 3’ end of an mRNA transcript, and can thereby identify the second, previously unknown fusion partner. Using this method, only 23-25 base pairs of known sequence are necessary to run a PCR amplifying also the 3’ end of the potential fusion partner. The method takes advantage of the construction of human mRNAs. These consist of the coding sequence followed by a poly-A tail (Scotto-Lavino *et al.*, 2006a). Briefly, cDNA sequence is generated by reverse transcription of mRNA. The primer used targets the poly A tail and is universal for all mRNAs. In a second round of PCR, a gene–specific primer mapping to the gene of interest is used to amplify the target sequence. A nested primer located further downstream can give more specific products in a new PCR. The obtained PCR product is then sequenced and alignment with the reference genome identifies the unknown sequence. 5’end RACE is somewhat more complicated and requires modifications of the protocol because the Poly A tail cannot be used as template (Sambrook and Russell, 2001, Scotto-Lavino *et al.*, 2006b). As the poly A tail is often destroyed during RNA degradation, RACE PCR requires RNA of good quality.

**Quantitative PCR**

In order to establish the expression level of an mRNA transcript, quantitative PCR (qPCR) is today the most frequently used and generally preferred method among researchers/molecular laboratories. The procedure is composed of three steps: (1) the conversion of RNA to cDNA using reverse transcriptase, (2) amplification of the cDNA template in a PCR reaction, and (3) detection of amplified products in real-time (Gibson *et al.*, 1996). The method differs from conventional PCR by monitoring the amplification process whereas in conventional PCR only the end result is analyzed. There are currently two ways to follow the amplification process, both of which
measure increases in fluorescence. The TaqMan method (Livak et al, 1995) utilizes specifically labeled primer probes that, when binding to their target sequence, emit a fluorescence signal. The increase in target sequences as the PCR progresses results in intensified emission. The other systems use a more unselective probe that binds all double stranded DNA (for example, SYBR Green®), and a rise in DNA results in increased emission. Unselective amplification, for example from primer-dimer formation, is more problematic with this approach.

Possible pitfalls to qPCR are manifold and must be avoided from the very beginning, even before the sample is loaded on the qPCR instrument. Incorrect handling of the tumor tissue after it is removed from the human body can greatly impact the expression levels of the mRNAs. The RNA should be of good quality and, if more samples are compared, should be of similar quality. Unequal quantities of tumor cells in each sample may affect the expression levels of cancer specific genes. The method for RNA quality measurement should be equal for all samples as estimates of quality can differ from method to method. Furthermore, care must be taken when designing primers for an experiment. Traditionally, primers overlapping on one or two exons have been used as an expression level for the whole gene (Bustin et al, 2009). This can be problematic as most human genes can be alternatively spliced (de la Grange et al, 2007). Also, using sequence from only one exon may detect contaminating DNA. It is therefore recommended to use primers overlapping at least 2 exons (Bustin et al, 2009). Choosing controls for normalization is a challenging task. For a reference gene to be suitable for normalization, it should be expressed at a similar level across all samples analyzed, and must not be affected by the experimental conditions. The last step in the qPCR is data analysis, which is also very important. Spurious differences can occur if researchers use different algorithms to interpret the data. Two main methods are used: absolute and relative quantification.

Sequencing
Since the 1970s, Sanger sequencing (Sanger et al, 1977) has been used to determine the order of bases in a given DNA fragment. The method consists of three main steps. First, the DNA fragment must be amplified, either by cloning into a high-copy-number
plasmid, or with primers in a PCR reaction. Then, cycles of template denaturation, primer annealing, and primer extension are carried out. This process resembles the PCR reaction apart for one important step: in addition to normal nucleotides, fluorescently labeled dideoxynucleotides are added to the reaction. Incorporation of these analogues results in a premature stop of the amplification process, and the resulting reaction mix will contain fragments of various lengths. By electrophoretic separation of the single-stranded end products, the sequence can be determined based on its emission spectrum. Sanger sequencing is applicable with DNA sequences up to about 1000 bases.

Since 2005, “next generation sequencing” has emerged and provides a faster and cheaper sequencing technology than the Sanger method. Advances in technology include improvements of the initial cloning step and simultaneous detection of numerous DNA fragments (Shendure and Ji, 2008). The technology is still evolving, and promising methods based on microchips reduce costs further. One of the main challenges using these new technologies is processing of the enormous amount of data generated (Shendure and Ji, 2008). This issue does not only apply to analysis of the data and storage capacity, but one also has to protect patient confidentiality in a safe way. In addition, the newer technologies yield shorter sequence reads, with lower accuracy than Sanger sequencing. Several applications for next-generation sequencing have emerged. Planning a given study, the approach chosen should reflect its aim. Sequencing platforms exist for genome-wide DNA studies, RNA studies, and targeted areas of the genome (Shendure and Lieberman Aiden, 2012). “High throughput”, “long-read” and “bench top” instruments allow the researcher to choose between different costs and quality of the data obtained.

**Transcriptome sequencing**

The transcriptome is the complete set of transcripts of a cell, and RNA sequencing (RNA-seq) is a technique that analyzes the transcriptome through deep-sequencing (Wang et al, 2009). The first step in RNA sequencing after isolation of total RNA is construction of a cDNA library using mRNA as a template. mRNA must be isolated from the other RNA components (such as rRNA, tRNA siRNA, and miRNA).
Following purification, mRNA is fragmented and converted to single stranded cDNA. After second strand synthesis, the fragments are ligated using specific adapters enabling PCR amplification to a cDNA library. This library can then be sequenced using next generation sequencing.

After high quality reads have been obtained from sequencing of the cDNA, the reads must be aligned to a reference sequence for analysis and interpretation. Different algorithms for analysis have been developed for various purposes. We have used two algorithms for fusion gene detection in the studies described in this thesis, deFuse (McPherson et al., 2011) and FusionMap (Ge et al., 2011), but several others are available (Carrara et al., 2013). As the cost of conducting RNA-seq experiments is declining, the method quickly becomes more widely applicable.

Several challenges and limitations yet to be solved must be kept in mind when conducting and interpreting data from RNA-seq experiments. The RNA obtained reflects a defined moment in the cell cycle and is tissue specific. The data will represent RNA from all the cells analyzed, not just the tumor parenchyma cells. Also, no consensus exists as to which tissue is appropriate as “normal” tissue for comparison. In addition, all the steps in the experimental procedure are susceptible to errors.
Figure 5: Fusion gene discovery from RNA-seq datasets. In principle, three different possibilities for alignment exist when the reads obtained by sequencing are compared to the reference genome: mapping reads (1), partly aligned reads (2) and non-mapping reads (3). Potential fusion genes are found in the second group, but additional filtering and validation are necessary for fusion gene detection. Modified from Maher et al (2009)
Aims

General aims
The study of chromosome changes in cancer cells has since the introduction of banding techniques in the 1970s (Caspersson et al, 1968, Smeets, 2004) provided many insights into the pathogenetic processes underlying tumorigenesis. Through persistent characterization of the breakpoints of cancer-specific chromosomal rearrangements, numerous new genes involved in human malignancies have been identified (Mitelman et al, 2007). At times, the complexity of acquired chromosomal changes is almost overwhelming, especially tumors of epithelial origin tend to display complex karyotypes (Heim and Mitelman, 2009). However, using the screening technique provided by cytogenetics, one can occasionally identify relatively simple but nonrandomly occurring chromosomal changes which are likely to influence cancer development. Further characterization of these changes using molecular cytogenetic and strictly molecular methods may reveal further molecular details. The translational strategy combining research efforts applied to acquired chromosomal rearrangements detected during diagnostic analyses of neoplastic cells has proven successful in detecting new cancer-specific chromosomal aberrations. From the regular diagnostic services provided by the Section for Cancer Cytogenetics, we have exclusive access to a large number of karyotypically characterized tumor samples of different origin, samples that show a variety of chromosomal aberrations.

The aim of the present thesis was to examine in detail such balanced chromosomal rearrangements detected in neoplastic cells and further characterize their consequences at the genomic level. Our aim was not to screen all patients within a disease entity, but rather to describe the pathogenetic events resulting from a few selected chromosomal aberrations that probably define a small subset of patients, and thereby elucidate an alternative way to cancer development. For that purpose we applied FISH and PCR based methods as well as transcriptome sequencing in the last two papers. These methods were used in different combinations to gain a complete overview of the genomic rearrangements and to identify their products at the RNA level. By means of
this multimodal approach, we could circumvent some of the limitations particular to each technique.

Today, most tumor classification systems are based on the morphologic appearance of the tumor cells and their tissue organization as it appears through the microscope. We hope that our findings will help identify additional diagnostic markers that can distinguish tumors based on their acquired mutations, as such a classification may reflect important tumor characteristics better than do their histological appearance (Brandal et al, 2010). Eventually, our hope is that a more profound understanding of disease-causing mutations will reveal new targets for specific treatments.

Additional specific aims and the theoretical rationale for each case studied are described in the paragraph dealing with each paper.
Specific aims

Paper I
Cytogenetic findings are crucial both for the diagnosis and prognosis of ALL patients (Group Francais de Cytogenetique Hematologique, 1996, Faderl et al, 1998, Harrison, 2009). The great majority of patients with this disease have cytogenetic abnormalities in their bone marrow cells, with detailed analysis that includes IP-FISH examination having demonstrated aberrations in up to 90% of newly diagnosed ALL cases (Harrison et al, 2005, Moorman et al, 2007). Up to 30% of ALLs have cytogenetic changes leading to fusion genes (Mitelman et al, 2007) with the type of fusion varying across age groups. In childhood ALL, the most common rearrangements are hyperdiploidy, followed by a t(12;21) recombining the genes ETV6 and RUNX1 (Moorman, 2012).

Prognosis in ALL is highly dependent on the leukemic karyotype at disease debut. An example is the group of patients with a high-hyperdiploid karyotype which corresponds to a favorable risk profile with nearly all patients surviving if treated adequately (Harrison, 2011). In contrast, patients with a low-hypodiploid karyotype have a far worse prognosis, one study measured event-free survival after three years of follow-up to only 29% in this group (Harrison et al, 2004). Examples of prognostic differences also exist among patients with structural rearrangements; patients with t(12;21) have an overall favorable prognosis while t(9;22) is associated with a poor prognosis necessitating allogeneic stem-cell transplantation in the first complete remission for these patients (Pui and Evans, 2006). Several other groups based on recurrent aberrations are acknowledged, but because of few reported cases, many of the cytogenetic aberrations in ALL are of unknown prognostic significance.

A 19-year-old patient diagnosed with B-precursor ALL was treated at our hospital. The cytogenetic analysis of his bone marrow aspirate revealed a 12;17-translocation as the sole karyotypic abnormality in the leukemic cells. The t(12;17)(p13;q12) is rare and of uncertain prognostic significance, but the reported data suggest a relatively good prognosis. Prior to our study, only 25 patients with this translocation had been reported, and in six of these, a fusion gene TAF15-ZNF384 had been identified.
Aims

(Martini et al, 2002, La Starza et al, 2005). Our aim was to investigate in molecular detail the consequences of the t(12;17). Were the same genes involved in our case? If so, was the breakpoint identical at the molecular level, i.e., were the same exons or introns involved? Was the fusion gene translated to an RNA transcript? Were there any cryptic rearrangements involved? Furthermore, we identified another ALL patient in our archive with a complex karyotype which included a similar t(12;17). We therefore examined also this second translocation addressing the questions mentioned above.

Paper II

AML is the most thoroughly studied malignancy in terms of cytogenetic aberrations with 16,998 cases (as of August 2013) reported in the Mitelman Database (Mitelman et al, 2013). The number of identified fusion genes in this disease is high and increasing. Roughly 60 % of patients with de novo AML display cytogenetic changes detectable by chromosome banding methods (Grimwade et al, 2010). In one series of young adult patients, the most common aberrations were trisomy 8, monosomy 7, t(8;21), and t(15;17) (Grimwade et al, 2010). As in ALL, the frequency of the aberrations depends on which age groups are examined. One of the characteristics of AML is that it is a heterogeneous disease. Numerous balanced translocations in addition to the ones mentioned above have been detected, only some of which carry a known diagnostic or prognostic impact (Heim and Mitelman, 2009). In addition to recognized cytogenetic subgroups, a panel of genes is found to be associated with prognosis if altered (for overview, see Gianfelici et al (Gianfelici et al, 2012)).

A 28-year-old patient with AML was admitted to his local hospital, and a bone marrow aspirate was sent to our laboratory for cytogenetic examination. The G-banding analysis revealed a t(12;13)(q14;q31) as the sole karyotypic change. We wanted to characterize the breakpoints in detail and determine the underlying gene-level changes. Which were the genes involved? Was there a fusion gene resulting from the rearrangement or is a deregulation of gene(s) the result of the translocation? What was the translocation product(s) at the RNA level? Was one of the well known cancer-related genes MDM2 and/or HMGA2, both of which map to 12q14, involved? Or were
genes mapping to the 13q31 area crucially involved, for example the \textit{GPC5} gene known to be amplified in selected cases of rhabdomyosarcoma (Williamson \textit{et al}, 2007)?

\textbf{Paper III}

Mesenchymal chondrosarcoma (MC) is a rare subtype of chondrosarcoma (<3\% of primary chondrosarcomas) (Fletcher \textit{et al}, 2013). Most patients are middle-aged and most tumors arise in the craniofacial bones, the ribs or the vertebrae. A biphasic pattern with areas of round primitive mesenchymal cells interrupted by chondroid elements characterizes the tumor histologically (Nakashima \textit{et al}, 1986).

Knowledge about the pathogenetic factors responsible for MC is limited, but a fusion gene identified in a majority of investigated cases was recently identified (Wang \textit{et al}, 2012). The fusion joins the DNA binding domain of the \textit{HEY1} gene to the C-terminal transcriptional activation domain of \textit{NCOA2}, and it is the fourth fusion gene involving \textit{NCOA2} identified in human cancer (Wang \textit{et al}, 2012). The fusion was identified using a genome-wide exon based expression array and no karyotypic information about the tumors was provided. The authors suggested the rearrangement to be caused by a submicroscopic deletion on chromosome 8.

No consistent cytogenetic findings were described in the ten karyotyped cases of MC in the Mitelman database (Mitelman \textit{et al}, 2013). Most karyotypes were complex. In two tumors, a Robertsonian translocation between chromosomes 13 and 21 was detected in addition to loss of all or a portion of chromosomes 8 and 20 and gain of all or a portion of chromosome 12 (Naumann \textit{et al}, 2002). Only two previous cases with a simple karyotype were reported, one with trisomy 8 as the sole aberration (47,XX,+8) (Gatter \textit{et al}, 2005), the other with a balanced t(4;19) (46,XY,t(4;19)(q35;q13)) (Richkind \textit{et al}, 1996).

We analyzed tumor cells from a 63-year-old woman with an MC harboring a 1;5-translocation as the sole aberration identified at the cytogenetic level. Since the translocation was present in all cells cytogenetically analyzed, we karyotyped also cells from a peripheral blood sample to rule out the possibility of a rare constitutional
Aims

variant. Once we had established that the 1;5-translocation was indeed tumor specific, we decided to characterize this new aberration at the molecular level. Did the MC under examination perhaps harbor a submicroscopic deletion of chromosome 8 leading to the HEY1-NCOA2, the fusion gene published as typical of MC while we were working on our case? If so, would the putative deletion lead to the same fusion that had already been reported or would it result in a variant? If not, did the present tumor rearrangement lead to a totally different fusion gene? If a novel fusion existed, what was its result at the molecular level? What was the prevalence of this/these fusion genes in other MCs?

Paper IV

Pure erythroid leukemia is characterized by neoplastic proliferation of cells resembling proerythroblasts (Swerdlow, 2008). The disease is extremely rare and is recognized as a diagnostic entity together with erythroleukemia (erythroid/myeloid) under the term “acute erythroid leukemias” in the WHO classification system (Swerdlow, 2008). Differential diagnosis against megakaryoblastic leukemia may be challenging, especially in cases of pure erythroid leukemia without erythroid maturation. There is sometimes gradual progression from myelodysplasia with erythroid dysplasia to pure erythroid leukemia, and a diagnostic marker is needed (Domingo-Claros et al., 2002). The patients often follow a rapid clinical course even with the best current treatment.

Previous cytogenetic studies of patients with erythroleukemia have displayed numerical and structural aberrations resembling those seen in AML in general, especially trisomy 8 and structural changes leading to loss of material from chromosomes 5 and 7 (Mitelman et al., 2013).

An 18-month-old boy was admitted to our hospital and diagnosed with pure erythroid leukemia. G-banding analysis showed a karyotype interpreted as 46,XY,der(1)t(1;1)(p31;q21),t(1;16)(p31;q24). Extensive FISH studies revealed that the CBFA2T3 gene was rearranged in the 1;16-translocation (Micci et al., 2011). However, because the material available for molecular analysis was degraded, we were not able to determine the leukemogenic mechanism whether it be the generation of a fusion gene with CBFA2T3 as one of the participants or loss of tumor suppressor
activity, in which case \textit{KANK1} or \textit{L1TD1} could be the target since both these genes were found to be homozygously deleted. Furthermore, it looked as if the 1;16-translocation is recurrent in pure erythroid leukemia since two additional cases have been reported with the same rearrangement (Koller \textit{et al}., 1989, Castaneda \textit{et al}., 1991).

A few months after the first publication (Micci \textit{et al}., 2011), we got hold of an additional 2 ml of coagulated blood with 3\% leukemic cells from the patient. Since the technology in the field of sequencing had developed so fast, we decided to extract RNA from the sample and try high-throughput sequencing analysis.

Our aim was now to characterize the t(1;16) at the molecular level. Would it be possible to identify the putative fusion gene using high throughput RNA sequencing working from a specimen with such a small number of abnormal cells? Was there a fusion gene between chromosomes 1 and 16? If so, which were the fusion partners, and could they be identified by PCR? What would be the possible mechanisms by which the chimeric protein executes its oncogenic potential? Since the FISH studies had identified a more complex karyotype, described as 46,XY,der(1)t(1;1)(p31;q21), del(1)(p11p31),der(16)t(1;16)(p31;q24), we were also wondering if the other described chromosomal breakpoints in the karyotype generated fusion genes.
Results in brief

Paper I
The first description of the translocation found in this case was t(12;17)(p13;q21), and although the most likely fusion gene was the known but rare TAF15-ZNF384 resulting from a t(12;17)(p13;q12), there were also several other potential candidate genes. ETV6 is located on 12p13 and is one of the most promiscuous gene partners in leukemia. To date, >30 different ETV6 partner genes have been identified (De Braekeleer et al, 2012, de Braekeleer et al, 2013). We used a commercially available ETV6/RUNX1 dual-color translocation probe to rule out the involvement of ETV6. We sat up a FISH-BAC assay to narrow down the breakpoints further. BAC RP11-151M4 overlapping eight genes including the ZNF384 gene gave a split signal when hybridized to patient metaphase cells. We then used fosmid clones to determine the breakpoint with greater precision. This latter approach did not succeed, probably due to repeated hybridizations to the same slide, and gave too weak signals to interpret. Meanwhile, we identified a splitting BAC overlapping the TAF15 gene on 17q12. Our results so far pointed to the fusion gene TAF15-ZNF384 and the karyotype was reinterpreted as t(12;17)(p13;q12). Both BACs covered an area of about 150 kb. This is a large area to analyze at the molecular level and would require many PCR reactions since we did not know where the precise breakpoint was. We assumed that the breakpoints were the same as those detected by Martini et al. (2002), and therefore use the same primers. We successfully amplified a TAF15-ZNF384 fragment.

We then wanted to investigate whether the TAF15-ZNF384 was present also in the second ALL patient identified through a search of our patient archives. FISH and PCR experiments confirmed the presence of a TAF15-ZNF384 fusion from DNA extracted from the leukemic cells also in that case.

Sanger sequencing of the PCR products from the two cases demonstrated that the break in TAF15 had occurred within intron nine in the first patient and within intron six in the second patient. In the ZNF384 gene, the break occurred within intron two in
both cases. Only in patient 1 was RNA available, and the transcript could be confirmed.

**Paper II**
To characterize the breakpoint position of the t(12;13) we began with FISH using specific BAC clones. Since we had no putative fusion genes for this case, we scattered the probes in chromosomal bands 12q14-q21, an area of approximately 20 Mb (UCSC Genome Browser (Kent *et al.*, 2002). In total, 36 BAC probes were tested before two splitting probes (one on the derivative chromosome 12 and one on the derivative chromosome 13) were identified. On chromosomal band 12q14 a split signal was seen using a BAC which overlaps the *HMGA2* gene. Fosmid clones were used to narrow down the breakpoint, and we were able to map it to the 3’end of the gene. The breakpoint on chromosomal band 13q31 was found to be in a region where no known genes are located. A fusion gene was therefore an unlikely consequence of this translocation. Attempts to amplify the breakpoint region on genomic DNA were unsuccessful, so in order to describe the precise rearrangement, we ran a RACE PCR amplifying the 5’end of *HMGA2*. We identified a truncated form of *HMGA2* by sequencing the resulting transcript. To demonstrate that the truncated transcript had consequences also at the RNA and protein level, we wanted to quantify the expression of *HMGA2* by qPCR and immunohistochemistry. *HMGA2* expression was considerable increased compared to a control group of AML patients assed by qPCR. Increased expression of HMGA2 was also demonstrated at the protein level, as we by immunohistochemistry saw specific staining of the blast cell nuclei.

**Paper III**
Having identified an MC patient with a tumor-specific t(1;5), we searched for additional patients within the Sarcoma Registry at the Norwegian Radium Hospital to find patients with the same diagnosis. From 1986 to 2010 only 3 such patients were identified, underlining the rarity of these tumors. We got hold of frozen material from all three tumors, and run a reverse transcriptase PCR aiming to amplify the *HEY1-NCOA2* fusion using the same primers as used by Wang *et al* (2012). A PCR fragment
was amplified in all but the t(1;5) case, and subsequent Sanger sequencing confirmed the presence of a \textit{HEY1-NCOA2} transcript in the other tumors.

The splitting BAC on 5q33 overlapped four genes. Using fosmid clones we narrowed the area further, to the genes \textit{PDGFR\beta} and \textit{CDX1}. Transcriptome sequencing was performed and the “deFuse” algorithm (McPherson \textit{et al}, 2011) was used to align sequences and predict potential fusion genes.

The list of predicted fusion genes exceeded 100. We focused on putative fusions between chromosomes 1 and 5, as we knew these two chromosomes were rearranged. A putative \textit{IRF2BP2-CDX1} transcript topped the list of potential fusions when ranked by split count, i.e., the number of reads supporting the prediction. The \textit{IRF2BP2-CDX1} was validated using primers outside the predicted transcript, and was also detected on genomic DNA extracted from the tumor. The fusion was not identified in any of the other described cases.

\textbf{Paper IV}

The blood sample from the patient harboring the t(1;16) contained 3% leukemic cells. RNA was extracted and sent for sequencing. A total of 107 million reads were obtained. The filtered data were run through the Fusion Map software (Ge \textit{et al}, 2011), and a list with about 500 putative fusion genes was identified. As number ten in this list ordered by the number of reads supporting the predicted fusion, a fusion between the genes \textit{CBFA2T3} and \textit{NFIA} was identified. These genes map to chromosome bands 16q24 and 1p31, respectively, which fits well with the described karyotype and the previous FISH results. The \textit{NFIA/CBFA2T3} fusion was verified by PCR and subsequent Sanger sequencing. As the FISH studies had revealed that the t(1;16) was in fact more complex than originally thought, \textit{in silico} analysis was performed of the sequences that involved chromosome 1. All predicted transcripts contained repetitive sequences or sequences with a high sequence identity to other genes, making them unlikely fusion partners.
The fusion gene gave an open reading frame and was predicted to encode a protein consisting of 811 amino acids, combining 208 amino-acid residues from NFIA and 603 residues from CBFA2T3.
Discussion

Methodological considerations
In this section, I shall discuss the study design of the subprojects that form the basis for this thesis in addition to some aspects of the methods applied. General aspects of each of the utilized method were discussed in the Introduction.

Study design
In all the studies presented in this thesis, only one or a few patient samples were examined. The selected cases were those with rare, cancer-specific cytogenetic aberrations, all identified through our diagnostic service. It is unquestionable that case reports and small series cannot replace well-designed, statistically-oriented studies when focusing on a particular therapy or the efficiency of a certain diagnostic test. However, they can be equally important when aimed in the right direction (Vandenbroucke, 2001). Case studies have historically provided significant advances also in cancer research. The famous $BCR-ABL$ story started with the identification of a minute chromosome in cells from two patients with chronic leukemia (Nowell and Hungerford, 1960a). Nowadays, single-patient studies are the first to be presented when a new technology is available, but still too expensive and time-consuming for the general research community. One such study was published in 2008 when the entire genome of a cytogenetically normal AML patient was sequenced (Ley et al, 2008). Unusual effects of a drug be it adverse or beneficial, are also often first reported as a case report. In other words, case studies are hypothesis-generating for forthcoming studies in the field, but cannot have statistical value or be directly implemented in evidence-based medicine.

We used G-banding of chromosomal preparations as an initial screening technique to identify clonal chromosomal aberrations. Cytogenetic analysis is a robust technique that has survived the changing trends in cancer research since the correct chromosome number was identified in 1956 (Tjio and Levan, 1956). It offers an overview of the cell genome which can be directly visualized down the microscope without the need for
complex algorithms. Some limitations of the method must be kept in mind: dividing cells are required for cytogenetic analysis, and some tumor cells do not divide in vitro, resulting in a false negative result (i.e., a normal karyotype). This has turned out to apply mostly to tumors of epithelial origin (Edwards, 2010). The number of aberrant cells represents the neoplastic cells’ ability to divide in vitro rather than in vivo, and so analysis of G-banded chromosomes is not ideal to assess the relative size of a given tumor or leukemia clone. Also, chromosome morphology may be too poor to identify the karyotypic abnormalities (Mitelman et al, 2007). The cases singled out for scrutiny in this thesis had cells that were able to divide in vitro. Cases with balanced translocations (with the exception of the case studied in paper IV and of one subproject for which a complete molecular characterization was not obtained; see Table 2) were chosen because these changes most probably are important for the neoplastic transformation.

**Working with limited material**

Studying cases identified through the diagnostic service had some drawbacks. Material available for analysis was often sparse and had not been optimally collected for research purposes. We had no means of controlling how much tissue was left for analysis or for how long the tissue was stored at room temperature before freezing (something that affects the quality of extracted nucleic acids). The following example demonstrates this limitation: in paper II, where good quality RNA was available from the malignant cells, we were able to run a 3’RACE PCR after the FISH results indicated involvement of the *HMGA2* gene. In contrast, the same technique was not feasible in the study leading to paper III due to degraded RNA.

After karyotyping, we applied FISH-based methods whenever feasible. To meet the challenge of sparse material, we tried to re-use previously G-banded metaphase spreads for FISH analysis, an approach that has been successfully applied in our laboratory before (Micci et al, 2002). The advantage is that the identity of the chromosomes studied is known beforehand, something that makes the interpretation of weak signals easier. Slides can thus be used for several rounds of hybridization. The identification of the fusion gene *TAF15-ZNF384* in paper I was made possible by this
approach when only two abnormal metaphases were available for analysis. However, more material for analysis would have been necessary had we not suspected that these two genes were the ones involved. The downside of re-using material is the harshness of the destaining protocol which wears down the biological material. Sequential hybridizations first remove proteins, but repeated hybridizations also reduce the DNA content of the chromosomes, making signal intensity weaker. Repeated hybridizations may also increase unspecific hybridization because of remaining debris from previous procedures.

Even with these efforts to save and re-use material, four additional projects planned to be a part of this thesis had to be discontinued because of lack of material before a complete molecular characterization of the cytogenetic aberration was achieved. (Table 2 provides an overview). Three of these cases were sarcomas characterized by apparently balanced translocations. The main problem working with these cases was that so few metaphases were available for FISH studies. One example was the case of a chondrosarcoma showing a 5;10-translocation as the sole karyotypic change. In the 20 slides dropped only four abnormal metaphases were identified. Neither in this case nor in any of the other sarcoma cases was the RNA quality good enough to perform transcriptome sequencing.

<table>
<thead>
<tr>
<th>Malignancy</th>
<th>Chromosomal aberration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chondrosarcoma</td>
<td>t(5;10)(q13;q26)</td>
</tr>
<tr>
<td>Liposarcoma</td>
<td>t(3;9)(p25;q22)</td>
</tr>
<tr>
<td>Sarcoma</td>
<td>t(2;17)(q23;q23)</td>
</tr>
<tr>
<td>Leukemia</td>
<td>der(9)t(9;12)(p23;q13)*</td>
</tr>
</tbody>
</table>

Table 2: Projects that had to be ended because of lacking material.*Part of complex karyotype.
Locus-specific FISH

Locus-specific FISH was used in papers I-III. This technique was crucial for our experiments before we implemented transcriptome sequencing as a tool. To determine a breakpoint position, clear and recognizable signals are necessary, as the staining used with FISH (DAPI) only displays the chromosome size and breakpoint position. From time to time we experienced background hybridization to be a problem with locus-specific FISH, especially when using old and destained slides (Figure 6). There were three main issues or problems with the FISH protocol: unspecific hybridization of the probe to the chromosomes, insufficient pretreatment of the slides, and poor probe quality. Addressing the first issue, we assessed the necessary amount of Cot-DNA. Cot-DNA is added to FISH experiments because it hybridizes to repetitive sequences in the genome, preventing false signals caused by unspecific binding (Fan, 2002). However, changing the concentration of Cot-DNA did not alter background signals. Next, we modified the pre and post hybridization washes, but with no improvement of results. Third, the prepared BAC probes were evaluated. When extracted from the bacterial cell, the BAC DNA is mixed with bacterial DNA (which exceeds BAC DNA by far), and we applied a different protocol to improve the BAC DNA output. Unfortunately, no consistent results favoring one or the other modification were obtained and troubleshooting of such a comprehensive protocol turned out to be a difficult task. Most likely, physical and/or chemical properties of the probe in combination with the slides contributed to the background signals.

Figure 6: Hybridization of this probe resulted in both chromosomal and extrachromosomal signals, making interpretation difficult
Discussion

**PCR-based methods**

We used PCR-based methods to characterize the breakpoint area defined by FISH. Our choice of method in each case was highly dependent on the amount of available material.

The genes involved in chromosomal rearrangements may contain large introns. Trying to amplify a hybrid gene from genomic DNA can therefore result in long PCR fragments, sometimes longer than the capacity of the enzymes used for the PCR reaction. We met this challenge by optimizing the PCR protocol for amplifying long fragments (LD-PCR). This variant of PCR is more susceptible to false positive PCR products because of its long extension time. Therefore, if a band appeared as a result of LD-PCR on genomic DNA, we ran more reactions involving neighboring exons. Whenever possible we preferred RNA as a template for RT-PCR. The “trial and error” approach we used when we had identified two putative genes does not work if unexpected submicroscopic rearrangements are present at the breakpoint position. As an example, the \textit{IRF2BP2-CDX1} fusion gene (paper III) would not have been easy to identify using this approach.

We used qPCR in paper II to assess the expression of \textit{HMGA2}. As discussed in the introduction, there are many pitfalls to qPCR, one being that it is difficult to select an appropriate control group. Not much is known about \textit{HMGA2} in hematological malignancies, but it seems that neoplasia-associated rearrangements involving chromosomal bands 12q13-15 lead to inappropriate expression of the protein (Fusco and Fedele, 2007). We chose to include AML patients with a common cytogenetic aberration, trisomy 8 (Heim and Mitelman, 2009) but without any visible changes of chromosome 12 and a “normal” bone marrow sample as a control group. We figured that comparison only with bone marrow from healthy donors could give an incorrect estimate of the gene expression as several genes may be aberrantly expressed in AML in general. We identified two AML patients with trisomy 8 as the only cytogenetic aberration, and one patient with monosomy 7 in addition to trisomy 8. A larger control group would have given us a more robust result, but the set up for this study limited us in this regard. As an internal control we chose the \textit{GUSB} gene, a commonly used gene
for endogenous control in qPCR studies, and one of the genes suggested by Beillard and colleagues as suitable for detecting minimal residual disease in leukemia (Beillard et al, 2003).

**Identifying fusion genes by transcriptome sequencing**

The traditional way to identify fusion genes, the use of locus-specific FISH followed by PCR-based methods, was applied in papers I and II. The drawback of this approach is that it is labor-intensive and requires good quality metaphases in sufficient number. In papers III and IV we used RNA sequencing in combination with G-banding. This rather new method for fusion gene detection allowed us to target our analysis to the chromosomal aberration described cytogenetically. Identification of fusion genes in RNA-seq datasets was first reported in 2009 (Maher et al, 2009). Tanas and colleagues were the first group to implement cytogenetic methods together with transcriptome sequencing, thus identifying a fusion gene in epithelioid hemangioendothelioma (a rare vascular sarcoma) (Tanas et al, 2011). Analysis of transcriptome data can be challenging, but we avoid some problems by combining cytogenetics and sequencing. Several algorithms have been proposed for fusion gene detection in RNA-seq datasets, all with different biases that must be taken into consideration (Carrara et al, 2013). Some algorithms apply a more stringent filtering than others, aiming to reduce the number of false positive predictions. The risk with this approach is that also true transcripts may be eliminated. FusionMap (applied in paper IV) uses less stringent filtering than deFuse (paper III) and requires less computational recourses. When a putative fusion region is already identified through cytogenetic analysis, the problem with false positive predictions becomes less challenging.

Detection of fusion genes resulting from cryptic genomic aberrations is now possible because of advances in genomic technologies. Through analysis of data from array-based studies, several investigators have reported novel and recurrent gene fusions (Tomlins et al, 2005, Wang et al, 2012, Giacomini et al, 2013). Although such studies are said to be unbiased, informed choices must nonetheless be made when analyzing the data. The large datasets generated compel researchers to filter the data in some
manner, and often genes not previously linked to cancer are excluded from further analysis, risking elimination of pathogenically important genes.

These different methods for identifying altered genes complement each other. The HEY1-NCOA2 recently detected in MCs would probably not have been identified cytogenetically, as it results from an interstitial deletion, while the t(1;5) if present in a subset of MC seems to be too rare to be identified in studies of only a few patients; it is probably a rare aberration in a rare cancer.

**Biological considerations**

In this thesis, three chromosomal rearrangements in acute leukemia and one rearrangement in an MC were investigated in molecular detail. The aims of the studies were descriptive, i.e., we wanted to characterize the rearranged gene(s) at the genome level. We also assessed the consequences at the RNA level, whenever possible.

**Why study rare cancers?**

Despite the fact that most research efforts are directed towards common malignancies such as cancers of the prostate, breast, lung, and colon (Cancer Registry of Norway, 2013), we decided to study samples from some rare tumor types. We see several reasons why also rare cancers may be worth studying. First, in some perspectives the division between rare and common cancers is artificial. Molecular profiling of common tumors have demonstrated that they are the phenotypic outcome of numerous and different pathogenetic processes (Munoz and Kurzrock, 2012). Based on the notion that there are fewer pathways that lead to the development of rare cancers (Braiteh and Kurzrock, 2007), rare cancers may be more homogenous, and therefore easier to study. In fact, most of the successful therapies recently developed are directed towards rare cancers (Braiteh and Kurzrock, 2007). Occasionally, discoveries in rare cancer types are useful also in other, more common cancers. One example is found in a subgroup of lung cancers; activation of the ALK gene through a gene fusion was first identified in anaplastic large-cell lymphoma (Kutok and Aster, 2002). The same fusion gene has subsequently been identified in about 4% of non small cell lung carcinomas.
(Soda et al, 2007), and targeted therapy inhibiting ALK is now implemented in the treatment of these patients (Kwak et al, 2010).

**Why study balanced translocations?**
In the “cytogenetic dictionary”, chromosomal aberrations are divided into primary and secondary aberrations, in addition to a third type, called cytogenetic noise (Heim and Mitelman, 1989). The primary aberrations are thought to be the tumor initiating events, i.e., necessary for cancer development. These aberrations are often found alone and are correlated with a distinct tumor type. Secondary aberrations, on the other hand, rarely exist alone and are believed to occur in cancer cells already harboring primary aberrations. They are important in tumor progression, and are probably selected in a Darwinian fashion. Cytogenetic noise is a term for non-clonal aberrations that occur because of genomic instability (Heim and Mitelman, 1989).

The division into “drivers” and “passengers” is another way of categorizing mutations in cancer used frequently today (Stratton et al, 2009). “Driver” mutations are those thought to be causally implicated in oncogenesis. Passenger mutations are not believed to confer a growth advantage on the cell carrying them, and therefore do not contribute to cancer development (Stratton et al, 2009).

Translation between these two set of terms is not always easy. Cytogenetic aberrations may refer to loss or gain of a whole or parts of a chromosome, while the drivers and passengers more often refer to single base substitutions within genes. On other occasions the translation is more straightforward. Balanced translocations, for example, are regarded as primary cytogenetic aberrations when identified as a sole cytogenetic change, and the fusion genes that such aberrations often form are “drivers” of the oncogenic process.

We work from the premise that the simple chromosome aberrations such as those examined in this thesis represent primary abnormalities and/or driver mutations, something that makes them all the more interesting to study. Although we cannot be certain about their contribution to oncogenesis, several considerations, some described below, support this hypothesis.
Genes involved in transcriptional regulation and nucleic acid binding are overrepresented among genes involved in human malignancies (Furney et al., 2006). Fusion gene partners are, according to recent studies, either genes involved in transcriptional control or they act as tyrosine kinases (Mitelman et al., 2004). All seven genes identified through the chromosomal translocations studied by us as part of this thesis influence the transcription process. ZNF384, CDX1, HMGA2 and NFIA are all transcription factors and bind directly to DNA (Vaquerizas et al., 2009). TAF15 interacts with transcription factor TBP, IRF2BP2 interacts with transcription factor IRF, and CBFA2T3 acts as a transcriptional repressor (Kochetkova et al., 2002).

The fusion transcripts found in leukemias and sarcomas are often composed of parts of genes involved in gene regulation or chromatin-modification (Edwards, 2010). The fact that all the four fusion genes identified (TAF15-ZNF384, HEY1-NCOA2, IRF2BP2-CDX1, and NFIA-CBFA2T3) join two gene-regulating genes therefore supports a functional role in oncogenesis. Also, structural similarities to already well-known fusion genes strengthen the probability of a role in cancer development. This is exemplified by the TAF15 gene, first identified in 2002 (Martini et al., 2002). TAF15 belongs to the TET protein family. It is structurally similar to another, more famous member of the same family, specifically the Ewing sarcoma protein known from the childhood malignancy carrying the same name (Tan and Manley, 2009). Not surprisingly, EWSR1 was demonstrated to form pair with ZNF384 in other leukemia patients (Martini et al., 2002). Such parallels can also be drawn for the CDX1 gene that forms the 3’ end of the IRF2BP2-CDX1 gene, although less powerful. CDX2, closely related to CDX1, has been detected as a part of a fusion gene in leukemia (Chase et al., 1999). Structural similarities also exist between NFIA-CBFA2T3 and other, more well-known fusion genes, namely the RUNX1-CBFA2T3 resulting from a t(16;21) (Gamou et al., 1998). Two of the fusion genes identified in this thesis are already known to be recurrent (TAF15-ZNF384, HEY1-NCOA2), and this supports their role in oncogenesis. As mentioned above, the NFIA-CBFA2T3 fusion is also likely to be recurrent, since a description of two patients with erythroleukemia and a t(1;16)(p31;q24) in their
karyotype was reported in the Mitelman database (Koller et al, 1989, Castaneda et al, 1991).

In paper II we investigated a novel t(12;13) identified in a patient with AML. Instead of finding a new fusion gene as we had expected, we mapped the breakpoint in 12q13 to inside the \textit{HMGA2} gene. No gene resides at the breakpoint position identified on chromosome 13q. This mechanism of gene alteration is not the typical consequence of a balanced translocation as translocations usually cause either a fusion gene or deregulation of one gene through juxtaposition with regulatory sequences from another gene (see Introduction). Deregulation of \textit{HMGA2} through 12q13~15 rearrangements is nevertheless well-known in benign mesenchymal tumors, lipomas in particular (Fusco and Fedele, 2007). The growth benefit of these tumors is suggested to be caused by a loss of a repressive miRNA targeting the 3` end of \textit{HMGA2} (Mayr et al, 2007). When overexpression of \textit{HMGA2} and the closely related \textit{HMGA1} is detected in malignant tumors, it seems to be correlated with a highly malignant phenotype and reduced survival (Fusco and Fedele, 2007). Knowledge about \textit{HMGA2}`s role in hematological malignancies is limited, but the mechanism of deregulation seems to be similar to that observed in the case reported by us; \textit{HMGA2} is activated by a translocation involving 12q, and no expressed fusion genes are identified. Hematological malignancies of different types have been associated with \textit{HMGA2} deregulation. (Table 3 provides an overview of the reported cases).

Do the identified fusion genes define distinct disease subgroups?

Fusion genes are in principal tumor type specific, and therefore constitute a useful tool in diagnostics (Mitelman et al, 2007). This recognition makes fusion genes important for pathological diagnosis in a range of hematological and mesenchymal tumors.
However, for most tumors, such disease specific biomarkers are absent. The MCs were until recently such a tumor entity lacking a distinctive marker and the diagnosis depended solely on histological characteristics and immunostaining (Fletcher et al, 2013). The HEY1-NCOA2 fusion gene (Wang et al, 2012), identified while we were working on paper III, now comes across as such a new and specific marker. So far, 29 cases have been analyzed, with the identification of the gene HEY1-NCOA2 in 21 of 29 (72%) of the tested patients (Nakayama et al, 2012, Wang et al, 2012) and paper III. IRF2BP2-CDX1 is the second fusion gene to be identified in this tumor type and may constitute an alternative pathogenetic pathway. However, since it so far is identified in only one tumor, the IRF2BP2-CDX1 may be unique to these tumor cells. Also, when more than one fusion gene is identified in a sarcoma subtype, they usually activate the same oncogenic pathway (Tan and Manley, 2009). The rarity of MCs makes it difficult to investigate more tumors and thus clarify these issues.

Translocations involving the ZNF384 gene on 12p13 (paper I) seem to define a subgroup of ALL patients (La Starza et al, 2005). In addition to the t(12;17), ZNF384 is also involved in a t(12;22) and a t(12;19). The clinical characteristics of the patients reported by us in paper I are in accordance with previously reported cases. Patients are usually young adults with an initial good response to chemotherapy (La Starza et al, 2005). However, the total number of reported cases is too small to conclude as to the prognostic impact of these rearrangements.

The clinical characteristics of leukemia patients characterized cytogenetically by the t(1;16) examined in paper IV seem to correspond to a distinct clinical subtype. Three patients, including the one reported in paper IV, have so far been identified with a t(1;16)(p13;q24) and leukemic blasts of the erythroid lineage. They were all young boys and the outcome was poor.

Is one fusion gene enough to generate a malignant cell clone?
The oncogenic potential of the genes identified by us, relative to other possible mutations in the tumors examined, is difficult to assess. Most available data imply that more than one mutation is required to generate a malignancy (Mitelman et al, 2007).
On the other hand, *in vitro* studies of cells transfected with *EWSR1-FLI1*, the dominating fusion gene in Ewing sarcoma, showed the development of a tumor-like phenotype in spite of the fact that only a single oncogenic event had occurred (Riggi *et al*, 2005). The oncogenic contribution of fusion genes can also be studied in transgenic mouse models. Introduction of the *FUS-DDIT3* fusion gene specific for liposarcoma into the mouse genome resulted in the development of liposarcomas resembling those seen in humans (Perez-Losada *et al*, 2000). This fusion gene seems to require a susceptible cell type to cause tumor growth; although the chimeric protein was identified in several tissues, only from adipocytes did tumors develop. The *BCR-ABL* fusion transcript of CML also seems to be sufficient to cause disease; however, additional genetic events are strongly correlated with disease progression (Deininger *et al*, 2000). Data also suggest that a fusion gene has to be present “at the wrong place, at the wrong time” in order to cause disease; *BCR-ABL* transcripts can be detected at low levels in the blood from normal individuals (Bose *et al*, 1998).

In all the leukemia cases described in papers I, II and IV (except for the additional patient identified in our database), a scan for mutations of known leukemia-associated genes was performed at the Department of Pathology at our hospital. This was done as part of the diagnostic routine. In papers I and II, alterations of the *FLT3* gene were identified. Point mutations in this gene are associated with a poor prognosis in AML patients, and usually co-exist with other oncogenic mutations (Kottaridis *et al*, 2001, Gilliland and Griffin, 2002). Newer data obtained from sequencing studies demonstrate that a typical solid tumor harbors 33-66 mutations expected to alter protein expression, while the corresponding number for pediatric tumors and leukemias is 9.6 mutations (Vogelstein *et al*, 2013). Therefore, it is entirely possible, perhaps even likely, that the malignancies studied by us harbored also other, not recognized, mutations.

**Consequences of gene alteration on protein expression**

If a fusion gene/gene mutation is identified, it can in theory be translated into a protein with new and oncogenic capabilities, or it can result in deregulation of one of the implied genes, usually overexpression (Mitelman *et al*, 2007). Most genes
characterized in this thesis were examined both at the DNA (FISH, PCR on genomic DNA) and the RNA (RT-PCR, qPCR and RNA-seq) level. In papers II and III, we also wanted to assess whether the altered genes influenced protein expression. In paper II, immunohistochemistry targeting *HMGA2* demonstrated specific staining in blast nuclei. *HMGA2* is not normally expressed in adult tissues (Fusco and Fedele, 2007) and detection by immunohistochemistry therefore implies abnormal expression of this protein. In paper III, CDX1 protein expression would be an indirect assessment for the translation of the *IRF2BP2-CDX1* fusion gene. CDX1 antibodies were commercially available, but repeated control experiments resulted in unspecific staining. Therefore, we were not able to evaluate if a fusion protein was formed using immunohistochemistry. Some assumptions about gene translation can still be made; using the “Open Reading Frame Finder” at NCBI both the novel fusion transcripts *IRF2BP2-CDX1* and *NFIA-CBFA2T3* were predicted to give an open reading frame (http://www.ncbi.nlm.nih.gov/projects/gorf/), supporting the notion that these fusion genes are translated into a protein.

The *TAF15-ZNF384* gene has been functionally evaluated before; the fusion showed transforming properties in a cell assay (Martini *et al*, 2002). From time to time, however, a balanced translocation can result in a truncated mRNA transcript, either due to a preterm stop codon in the coding sequence or a missense mutation resulting in an mRNA with altered protein-generating capabilities. This consequence can also be of pathogenetic importance if affecting a tumor suppressor gene. A wide range of methods exists to assess the functional significance of gene fusions. Some of these are exemplified in Table 4. Cell assays would be a natural starting point if one were to examine the fusion genes identified by us at the protein level.
Table 4: Examples of methods for evaluation of the oncogenic contribution of a fusion gene

<table>
<thead>
<tr>
<th>Setting</th>
<th>Study subject</th>
<th>End point</th>
<th>Requirements</th>
<th>Example study</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>In vitro</em></td>
<td>Cell lines</td>
<td>Cell viability</td>
<td>Tumor cells/cell line expression the fusion gene in question</td>
<td>Giacomini <em>et al.</em>, (2013)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cell proliferation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cell invasion</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cell transforming properties</td>
<td></td>
<td>Maritini <em>et al.</em>, (2002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Disease-specific survival</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>In vivo</em></td>
<td>Human beings</td>
<td>Progression free survival</td>
<td>Patients harboring the fusion gene in question and a small molecule inhibitor targeting it</td>
<td>Kwak <em>et al.</em>, (2010)</td>
</tr>
<tr>
<td></td>
<td>(Phase I study)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The structure of breakpoints and the fusion genes they cause

We have described in detail the breakpoints, the involved genes, and the fusion transcripts identified. By similar descriptions also in future cases, it can be clarified which parts of the involved genes account for the oncogenic potential. This knowledge can be essential if a potential interfering molecule is identified. One fusion gene can also exist in slightly different versions, i.e., involving different exons from the involved genes. This may again lead to fusion proteins with different transcriptional activities that again affect tumor phenotype, progression and prognosis (Zoubek *et al.*, 1996, de Alava *et al.*, 1998, Gonzalez *et al.*, 2007). The *TAF15-ZNF384* fusion had been identified in two different versions before, one containing exons 1-9 of *TAF15* and the other exons 1-6, both fused to exon 3 of *ZNF384* (Martini *et al.*, 2002). We confirmed these two fusion gene variants that exons 1-6 of *TAF15* are sufficient for development of the disease. Potential prognostic differences caused by these two
Discussion

*TAF15-ZNF384* transcripts are impossible to assess due to the low number of patients reported.

The breakpoint region of the *HEY1-NCOA2* fusion was identical in our study (paper III) and all other analyzed cases (Wang *et al.*, 2012). Future analyses of the *IRF2BP2-CDX1* and *NFI-CBFA2T3* genes will reveal if different variants exist for these fusions.

In paper III, we expected that a balanced translocation had occurred between chromosomes 1 and 5. However, due to the orientation of the involved genes (*IRF2BP2* and *CDX1*), some additional rearrangement must have taken place. We tried to investigate this issue further by locus-specific FISH overlapping the breakpoint position on chromosome band 1q42. We validated *IRF2BP2* as one of the fusion partners and expected a BAC probe overlapping *IRF2BP2* to yield a split signal. Instead, we saw two signals located on chromosome 1. We concluded that a submicroscopic rearrangement must have taken place, but were not able to characterize the rearrangement further.
Conclusions and future perspectives

In the papers presented in this thesis we have described the molecular consequences of four different cytogenetic aberrations in malignant cells. This has resulted in better characterization of two infrequent aberrations whose gene-level consequences were already known, and the detection of two new cancer-specific fusion genes.

The aim of cancer research should ultimately be to reduce the mortality and morbidity associated with cancer. There are many ways in which this goal can be attained. Our work may first have an impact diagnostically, as numerous genetic aberrations have been demonstrated to be disease specific (Mitelman et al., 2007). Improved diagnostics would contribute to better cancer treatment, by selection of the right patients for the right treatment. Stratifying patients on the basis of genetic aberrations is already implemented for the most common hematological neoplasias, and has resulted in differences in treatment protocols based on the known prognostic impact of certain aberrations (Swerdlow, 2008).

Our group has for many years studied such changes and the research continues. With the recent introduction of transcriptome sequencing, the group has so far published four articles in addition to papers III and IV using transcriptome sequencing to identify fusion genes (Panagopoulos et al., 2012, Panagopoulos et al., 2013a, Panagopoulos et al., 2013b, Panagopoulos et al., 2013c). Even though these studies for the most part were of single cases, we expect to broaden them in the future. To generate more knowledge and understanding, gene fusions should be examined in larger cohorts of patients with malignant diseases, preferably with the same cytogenetic characteristics, to determine their prevalence and specificity as biomarkers. For HMGA2, the gene we found to be deregulated in AML in paper II, the expression could be measured in a panel of AML patients with rearrangements of chromosome region 12q1.

The oncogenic potential of the involved genes can be investigated by functional studies as outlined in Table 4. Research in our laboratory is currently focused on describing the consequences of altered chromosomes rather than doing functional
Conclusions and future perspectives

analysis, so by publishing our results we hope that other researchers with more cell biological expertise will take the findings further.

Today, fusion gene detection in the clinical diagnostic setting is based on FISH (mostly interphase FISH) or RT-PCR. Either as an array, or using targeted next generation sequencing, it will soon be feasible to detect all known gene fusions in cancer in a single experiment, also applicable in a diagnostic setting (Lovf et al, 2011, Lovf et al, 2013). As outlined in the discussion, all variants of disease specific rearrangements are important to investigate, as subtypes of them may impact prognosis. An updated database including all these chimeric transcripts is necessary for such an approach, and much effort has already been made to catalogue chimeric transcripts in databases (Frenkel-Morgenstern et al, 2013, Mitelman et al, 2013).

In hematologic and lymphatic malignancies, monitoring of minimal residual disease can be performed by PCR analysis amplifying fusion genes (van der Velden et al, 2003). This technique has been successfully used for detection of BCR-ABL fusion gene transcripts in CML (Bagg, 2002). Recent studies have demonstrated that this method is applicable also in solid tumors. By next generation sequencing, Leary and co-workers identified tumor-specific rearrangements in breast and colorectal cancer patients and were able to assess disease status and drug responsiveness by monitoring these disease-specific gene rearrangements (Leary et al, 2010). Although still too expensive in most diagnostic settings, this approach demonstrates another potential benefit of studying acquired mutations in cancer, independently of the oncogenic properties they may possess.

A novel fusion gene such as IRF2BP2-CDX1 (paper III) would probably be a good individual biomarker for the patient in whom it was identified, regardless of the limited knowledge of its function. A nascent relapse would most probable be detectable by PCR targeting this fusion gene. If such fusion gene detection could be implemented in a clinical setting, it would probably save both recourses associated with conducting and interpreting the evaluations and the adverse effects linked to recurrent radiological evaluations. However, disease monitoring using “individual”
fusion genes requires that the mutations in question are in fact “primary” or “drivers” and that they are present in both the primary tumor and at relapse. A recent paper reported that among genes found to be mutated, only ~30% were found throughout the tumor parenchyma (Gerlinger et al., 2012). One approach to meet this issue is to investigate a panel of mutated genes, where one or more most probably will be present in all tumor cells.

In order to obtain a better pathogenetic understanding of a specific tumor entity, a deeper investigation of one or a few well-characterized tumors using several methods (cytogenetics, RNA sequencing, gene expression profiling etc.) may reveal more about the pathogenesis than using one single approach applied to many, probably molecular distinct tumours (Liu, 2013). We hope our present and future studies of malignancies at different resolution levels will contribute to this development.

A final note to the therapeutic potential of chimeric proteins caused by chromosomal rearrangements will round off this work. First, a protein product has to be important for tumor cell survival in order to make a good target for treatment. Second, it is easier to target a protein that results in a gain-of-function and has enzymatic activity, than a protein with numerous weak interactions (Vogelstein et al., 2013). Although development of treatments directed against fusion genes may seem like a challenging task, we must be patient in our search. Four decades passed from the identification of the Philadelphia chromosome to imatinib was approved for therapy against CML, but it was definitely worth the wait (Druker, 2008).
List of abbreviations

ALL    acute lymphoblastic leukemia
AML    acute myeloid leukemia
BAC    bacterial artificial chromosome
cDNA   complementary DNA
CGH    comparative genomic hybridization
CML    chronic myeloid leukemia
DAPI   4',6-diamidino-2-phenylindole
DNA    deoxyribonucleic acid
dNTP   deoxyribonucleotide
ESTs   expressed sequence tags
FAB    French American British
FISH   fluorescence in situ hybridization
GISTs  gastrointestinal stromal tumors
HLA    human leukocyte antigen
IP-FISH interphase FISH
ISCN   International System for Human Cytogenetic Nomenclature
LD-PCR long distance PCR
MC     mesenchymal chondrosarcoma
M-FISH multiplex FISH
miRNA  micro RNA
mRNA   messenger RNA
NCBI   National Center for Biotechnology Information
PCR    polymerase chain reaction
qPCR   quantitative PCR (Real-time PCR)
RACE-PCR rapid amplification of cDNA ends
RNA    ribonucleic acid
RNA-seq RNA sequencing
rRNA   ribosomal RNA
RT-PCR reverse transcription
siRNA  small interfering
TKI    tyrosine kinase inhibitor
tRNA   transfer RNA
UCSC   University of California, Santa Cruz
WHO    World Health Organization
Gene symbols

**ABL1**  c-abl oncogene 1, non-receptor tyrosine kinase

**ALK**  anaplastic lymphoma receptor tyrosine kinase

**BCR**  breakpoint cluster region

**CBFA2T3**  core-binding factor, runt domain, alpha subunit 2; translocated to, 3

**CDX1**  caudal type homeobox 1

**CDX2**  caudal type homeobox 2

**DDIT3**  DNA-damage-inducible transcript 3

**ETV6**  ets variant 6

**EWSR1**  EWS RNA-binding protein 1

**FLI1**  Fli-1 proto-oncogene, ETS transcription factor

**FLT3**  fms-related tyrosine kinase 3

**FUS**  fused in sarcoma

**GPC5**  glypican 5

**GUSB**  glucuronidase, beta

**HEY1**  hairy/enhancer-of-split related with YRPW motif 1

**HMGA2**  high mobility group AT-hook 2

**HMGA1**  high mobility group AT-hook 1

**IRF2**  interferon regulatory factor 2

**IRF2BP2**  interferon regulatory factor 2 binding protein 2

**JAK2**  Janus kinase 2

**KANK1**  KN motif and ankyrin repeat domains 1

**KIT**  v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog

**L1TD1**  LINE-1 type transposase domain containing 1

**MDM2**  MDM2 oncogene, E3 ubiquitin protein ligase

**MYC**  v-myc avian myelocytomatosis viral oncogene homolog

**NCOA2**  nuclear receptor coactivator 2

**NFIA**  nuclear factor I/A

**NTRK1**  neurotrophic tyrosine kinase, receptor, type 1

**PDGFRβ**  platelet-derived growth factor receptor, beta polypeptide

**PIGA**  phosphatidylinositol glycan anchor biosynthesis, class A

**RUNX1**  runt-related transcription factor 1

**TAF15**  TAF15 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 68kDa

**TBP**  TATA box binding protein

**TNK2**  TRAF2 and NCK interacting kinase

**ZNF384**  zinc finger protein 384
References


References


Cancer Registry of Norway (2013) Fakta om kreft [http://www.kreftregisteret.no/no/Generelt/Fakta-om-kreft-test/]


References


References


References


Identification of the TAF15-ZNF384 fusion gene in two new cases of acute lymphoblastic leukemia with a t(12;17)(p13;q12).

Cancer Genetics (2011) 204(3):147-152.
Paper II

Kaja B. Nyquist, Ioannis Panagopoulos, Jim Thorsen, Roberta Roberto, Hilde S. Wik, Anne Tierens, Sverre Heim, Francesca Micci:

t(12;13)(q14;q31) leading to HMGAI2 upregulation in acute myeloid leukaemia.

Kaja B. Nyquist, Ioannis Panagopoulos, Jim Thorsen, Lisbeth Haugom, Ludmila Gorunova, Bodil Bjerkehagen, Alexander Fosså, Marianne Guriby, Torfinn Nome, Ragnhild A. Lothe, Rolf I. Skotheim, Sverre Heim, Francesca Micci:

Whole-transcriptome sequencing identifies novel \textit{IRF2BP2-CDX1} fusion gene brought about by translocation t(1;5)(q42;q32) in mesenchymal chondrosarcoma.

Whole-Transcriptome Sequencing Identifies Novel IRF2BP2-CDX1 Fusion Gene Brought about by Translocation t(1;5)(q42;q32) in Mesenchymal Chondrosarcoma

Kaja B. Nyquist1,2,3, Ioannis Panagopoulos1,2, Jim Thorsen1,2, Lisbeth Haugom1,2, Ludmila Gorunova1,2, Bodil Bjerkehagen4, Alexander Fossa5, Marianne Guriby2,6, Torfinn Nome2,6, Ragnhild A. Lothe2,6, Rolf I. Skotheim2,6, Sverre Heim1,2,3, Francesca Micci1,2

Abstract

Mesenchymal chondrosarcomas (MCs) account for 3–10% of primary chondrosarcomas. The cytogenetic literature includes only ten such tumours with karyotypic information and no specific aberrations have been identified. Using a purely molecular genetic approach a HEY1-NCOA2 fusion gene was recently detected in 10 of 15 investigated MCs. The fusion probably arises through intrachromosomal rearrangement of chromosome arm 8 q. We report a new case of MC showing a t(1;5)(q42;q32) as the sole karyotypic aberration. Through FISH and whole transcriptome sequencing analysis we found a novel fusion between the IRF2BP2 gene and the transcription factor CDX1 gene arising from the translocation. The IRF2BP2-CDX1 gene fusion has not formerly been described in human neoplasia. In our hospital’s archives three more cases of MC were found, and we examined them looking for the supposedly more common HEY1-NCOA2 fusion, finding it in all three tumours but not in the case showing t(1;5) and IRF2BP2-CDX1 gene fusion. This demonstrates that genetic heterogeneity exists in mesenchymal chondrosarcoma.

Introduction

The classification of sarcomas describes over 50 different histological subtypes [1]. In approximately 20% of them, recurrent balanced translocations leading to formation of fusion genes were identified [2]. Fusion genes provide diagnostic and sometimes prognostic information on the tumours they characterize and knowledge about them could ultimately lead to new targeted therapies [3].

Mesenchymal chondrosarcomas (MCs) are rare tumours that account for 3–10% of primary chondrosarcomas [1]. Their typical histological appearance includes a biphasic pattern with areas of round primitive mesenchymal cells interrupted by chondroid elements [4]. Most cases are diagnosed in the second and third decade of life and the prognosis is mostly poor, with a 5-year survival rate of about 50% [5]. Adequate surgery is the gold standard for treatment of localized disease [6] and the role of chemotherapy and radiotherapy remains poorly defined [7,8].

According to the Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer [9], only ten MCs have been karyotyped and no consistent cytogenetic findings have been described. Recently, however, using a genome-wide exon-resolution expression screen, a fusion between the hairy/enhancer-of-split related with YRPW motif 1 (HEY1) gene and the nuclear receptor coactivator 2 (NCOA2) gene was detected in 10 out of 15 analysed MCs (67%) [10]. Both genes are located on the long arm of chromosome 8 and so the fusion presumably results from an intrachromosomal rearrangement, probably a deletion (~9.6 Mb according to the UCSC browser, assembly 2009).

We report the finding of a balanced t(1;5)(q42q32) as the sole karyotypic abnormality in an MC. The translocation led to a new fusion between the interferon regulatory factor 2 binding protein 2 (IRF2BP2) and the caudal type homeobox 1 (CDX1) gene. Based on the recent report by Wang et al (2012) [10], we also examined archival material from another three MCs we had access to for the presence of the HEY1-NCOA2 gene fusion, finding it in all three.
Materials and Methods

Patient Samples
Patient 1 was a 63-year-old female in whom a solitary tumour mass was detected in the right cerebral hemisphere in August 2007. Examination of biopsy material revealed the tumour to be a diffuse large B-cell lymphoma of activated B-cell subtype. Cytogenetic analysis of this tumour was unsuccessful. Detailed work-up for other manifestations of lymphoma was negative, compatible with a diagnosis of primary central nervous system lymphoma (PCNSL). However, a tumour in the left iliac muscle was detected, 3 cm in largest diameter. Biopsies revealed a spindle cell tumour of uncertain malignant potential. The patient received chemotherapy for PCNSL according to Albreu et al. [11] including high-dose methotrexate and high-dose cytarabine. Evaluation after 7 courses of chemotherapy confirmed complete remission of her PCNSL. There was no change in size of the tumour in the left iliac muscle and in June 2008 a wide excision of it was performed. A detailed work-up of the tumour specimen revealed a small cell and chondromatous tumour diagnosed as a mesenchymal chondrosarcoma (Figure 1A). Focal infiltrative growth and necroses were present. Because of narrow margins, postoperative radiotherapy 2 Gy x 25 was given. A CNS recurrence of her lymphoma was detected in November 2011, and the patient has received radiation therapy. She remains without sign of recurrence of the MC at the time of writing.

The Norwegian Radium Hospital (NRH) is the largest referral centre for Norwegian patients with bone and soft tissue tumours covering a population of 2.6 million. To identify additional patients with a diagnosis of MC, a database search was performed for cases with this disease. Three additional patients (patients 2–4) were identified (see Table 1 for clinical details and Figure 1A for histological image).

Ethics Statement
Written informed consent was obtained from patients 1 and 3. In the latter case, one of the parents consented on the patient’s behalf. Frozen tissue from deceased patients (patients 2 and 4) was retrieved from The Radium Hospital biobank (project nr S-0747a approved by the Regional Ethics Committee and The Directory of Health (Helsedirektorat) in 2008). Specific permission to perform RNA analysis/sequencing was obtained from patient 1 after approval by the Regional Ethics Committee for Medical and Health Research Ethics South-East (REC number: 2010/1389A). The entire study was also approved by the institutional review board at the Norwegian Radium Hospital. All data were analyzed anonymously.

G-banding and karyotyping
Fresh tissue from a representative area of the tumour (patient 1) was received and analysed as part of our diagnostic routine. The samples were disaggregated mechanically and enzymatically with collagenase II (Worthington, Freehold, NJ, USA). The resulting cells were cultured and harvested using standard cytogenetic techniques [12]. Chromosome preparations were G-banded with Wright stain. The karyotype was written according to The International System for Human Cytogenetic Nomenclature (ISCN; 2009 guidelines [13]). Phosphoethamglutamin (VWR, Oslo, Norway)–stimulated leucocytes were obtained from peripheral blood to determine the patient’s constitutional karyotype.

Fluorescence in Situ Hybridization Analysis
Fluorescence in situ hybridization (FISH) was performed using probes from bacterial artificial chromosomes (BACs). BACs and fosmid clones flanking and covering the breakpoint positions were selected using the Human Genome Browser at the University of California web site (Feb.2009/release: hg19, http://genome.ucsc.edu/). The selected clones (see Table S1 for detailed information) were purchased from Life Technologies (Carlsbad, CA, USA) or the BACPAC Resource Center (Oakland, CA, USA).

Bacteria were cultured in selective media according to the manufacturer’s recommendation. DNA was extracted using High Pure Plasmid Isolation kit (Roche Applied Science, Penzberg, Germany). DNA labelling was done in a nick translation reaction and the synthesized probes were hybridized to previously G-banded slides. All procedures were performed as previously described [14]. The slides were counterstained with 4’,6-diamidino-2-phenylindole (DAPI). The analysis was done using a CytoVision system (Applied Imaging, Newcastle, UK). All probes were tested for their correct location on normal metaphase spreads prior to use.

Material for Molecular Analysis
Representative samples of tumour tissue were frozen and stored at −80°C after surgery. DNA was isolated using Genomic-tip (Qiagen, Hilden, Germany) to obtain pure high molecular weight DNA. RNA was extracted from tumour tissue using the Trizol reagent (Life Technologies) with a homogenizer (Omni THQ Digital Tissue Homogenizer, Kennesaw, GA, USA). The RNA quality was evaluated using the Experion Automated Electrophoresis System (Bio-Rad Laboratories, Hercules, CA, USA). cDNA synthesis was done using the iScript kit and random primers (Bio-Rad Laboratories). All procedures were done according to the manufacturers’ recommendations.

High-throughput Paired-end RNA-sequencing
Sequencing was performed according to the TruSeq paired-end RNA-sequencing protocols from Illumina for Solexa sequencing on a Genome Analyzer Ix with paired end module (Illumina Inc., San Diego, CA, USA). 3.5 μg total RNA was used as starting material for library construction, using the TruSeq RNA Sample Preparation Kit v2 where the steps include poly-A mRNA isolation, fragmentation, and cDNA synthesis before adapters are ligated to the products and amplified to a final cDNA library. Shearing to about 150 bp fragments was achieved using divalent cations under elevated temperature. Approximately 38 million clusters were generated by the TruSeq PE Cluster Kit v2 on the Illumina cBot Cluster Generation System, and 76 base pairs were sequenced, from each side of the fragments, using reagents from the TruSeq SBS Kit v3 (all kits from Illumina).

Gene Fusion Prediction
The Illumina software pipeline was used for processing of image data into raw sequencing data (SCS 2.9 and Casava 1.8.2), and only sequence reads marked as “passed filtering” were used in the downstream data analysis. A total of 91 million reads were obtained. We utilized the fusion discovery software deFuse (version 0.4.3) [15], with Ensembl release 65 reference genome (hg19) and gene models, RepeatMasker, EST, and spliced EST annotations downloaded from the University of California Santa Cruz Table Browser (http://genome.ucsc.edu/), accessed May 2012). UniGene clusters were downloaded from National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/, accessed May 2012) to assist in locating potential gene fusions. Three spanning reads and two split reads were required to call sequence reads a gene fusion.
PCR and Sequencing

Primers used in PCR were designed with the FastPCR software [16]. The full list of applied primers is given in Table 2. The primers used for detection of the HEY1-NCOA2 fusion were identical to the primers used by Wang et al [10]. cDNA PCR was run using 2 µl cDNA in a 25 µl PCR reaction using TaKaRa Ex Taq Hot Start (Takara Bio Inc, Shiga, Japan). The PCR conditions were as follows: 98°C for 7 sec, 68°C for 2 min after a 1 min initial denaturation at 98°C. 34 cycles were run. Amplified products were cloned using the TOPO TA cloning kit (Life Technologies). Selected products were sent for Sanger sequencing (GATC Biotech, Konstanz, Germany) and obtained sequences were analysed using BLAST (Basic Local Alignment Search Tool, www.ncbi.nlm.nih.gov/BLAST/). All cases were tested for expression of a zinc-finger gene suppressor of zeste 12 homolog (Drosophila) (SUZ12) to assess RNA quality.

Long Distance PCR

PCR experiments on genomic DNA were performed using ~100 ng DNA as template in 25 µl PCR reactions using TaKaRa LA Taq following the manufacturer’s recommendations for LD-PCR: 30 cycles of 98°C for 10 sec (denaturation) followed by 68°C for 15 min (annealing and extension; Takara Bio Inc). PCR products were purified using GeneJET PCR purification kit (Fermentas GmbH, St. Leon-Rot, Germany) and sent for Sanger sequencing (GATC Biotech).

Table 1. Patient characteristics.

<table>
<thead>
<tr>
<th>Patient number</th>
<th>Sex/age at diagnosis</th>
<th>Histological diagnosis</th>
<th>Location of primary tumour</th>
<th>Tumour material analysed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F/63</td>
<td>Mesenchymal CS</td>
<td>Left iliacus muscle (soft tissue tumour)</td>
<td>Primary tumour</td>
</tr>
<tr>
<td>2</td>
<td>F/38</td>
<td>Mesenchymal CS</td>
<td>Pelvic bone (bone tumour)</td>
<td>Metastasis</td>
</tr>
<tr>
<td>3</td>
<td>F/12</td>
<td>Mesenchymal CS</td>
<td>Vertebral (bone tumour)</td>
<td>Metastasis</td>
</tr>
<tr>
<td>4</td>
<td>F/39</td>
<td>Mesenchymal CS</td>
<td>Vertebral and right thigh (bone and soft tissue tumour)</td>
<td>Metastasis</td>
</tr>
</tbody>
</table>

F = female.

doi:10.1371/journal.pone.0049705.t001
Results

The cytogenetic analysis of the only tumour (patient 1) from which we got a fresh sample revealed a balanced t(1;5)(q42;q32) as the sole abnormality in all cells analysed (Figure 2A). Analysis of peripheral blood leukocytes displayed a normal female karyotype as the sole abnormality in all cells analysed (Figure 2A). Analysis of peripheral blood leukocytes displayed a normal female karyotype ruling out the possibility of a constitutional aberration.

All cases (i.e., the tumour carrying a t(1;5) as well as the three archival tumours) were tested for the HEY1-NCOD2 fusion which was recently identified in 10 of 15 investigated MCs [10]. A PCR product of approximately 300 bp was amplified from tumours 2, 3, and 4 but not from tumour 1. Subsequent sequencing analysis confirmed the fusion between HEY1 exon 4 and NCOD2 exon 13 in cases 2–4, identical to the one previously described [10] (Figures 1B and 1C). As no PCR product was amplified in tumour 1, i.e., the one showing the t(1;5), we assumed that a new fusion gene was generated by the 1;5-rearrangement. To better characterize the breakpoint on the rearranged chromosome 5, a series of selected BAC clones mapping to the involved bands were hybridized to metaphase plates. Clone CTC-802J2 mapping on chromosome 1. However, no signal was seen on chromosome 5 as expected if the breakpoint was in the large intron 1 of IRF2BP2. Signals hybridized BAC clones overlapping the breakpoint. The breakpoint position was further narrowed down using fosmid clones with clone G248P81640F4 giving a split signal, mapping the breakpoint to a genomic area between the 3’ end of the platelet derived growth factor receptor β gene (PDGFRβ) and the large intron 1 of CDX1.

Since the breakpoint region as delimited by FISH was large, probably involving one of two genes, we decided to investigate the translocation in detail using a whole transcriptome sequencing approach focusing on potential fusion transcripts between chromosomes 1 and 5. The deFuse algorithm [13], designed for fusion gene discovery in paired-end RNA sequence data sets, gave us a list of 92 putative fusions in the tumour transcriptome that we filtered to 85 after removing isoforms of the same fusions (Table S2). An IRF2BP2-CDX1 transcript involving two coding regions yielded the highest split count value (number of split reads supporting the fusion) of all the predicted fusions and was confirmed in the tumour RNA. Both fusion transcripts were found to be “in frame” and are predicted to encode proteins of 466 and 450 amino acids, respectively, before being terminated by a stop codon. The reciprocal fusion between CDX1-IRF2BP2 did not yield any products by cDNA PCR (primer combinations CDX1-214F+IRF2BP2-1248R and CDX1-369F+IRF2BP2-1172R, Table 2). The presence of the IRF2BP2-CDX1 fusion gene was tested for in specimens from tumour 2–4 using primer combinations IRF2BP2-926F and CDX1-771R (Table 2). None of the specimens showed such fusion.

Next, we wanted to identify the genomic breakpoints of the IRF2BP2-CDX1 fusion. Using primers IRF2BP2-926F and CDX1-26970R (Table 2) we managed to amplify a product of about 800 bp which by sequencing was shown to contain the breakpoint, i.e., sequencing analysis confirmed the predicted positions of the genomic breakpoints. On chromosome 1 the breakpoint was in intron 1 of IRF2BP2 (chr1:2534743757 bp), whereas on chromosome 5 it was in the large intron 1 of CDX1 (chr5:149551799 bp).

To investigate the involvement of the IRF2BP2 gene also by FISH, we hybridized BAC clones overlapping IRF2BP2 to metaphases obtained from the cultured cells (Table S1). Signals were detected on the normal chromosome 1 and the derivative chromosome 1. However, no signal was seen on chromosome 3 as it would be expected if the IRF2BP2-CDX1 fusion had resulted from a simple balanced translocation. These findings thus indicate that a more complex rearrangement had taken place, possibly including inversions at the breakpoint.

Discussion

The cytogenetic knowledge on mesenchymal chondrosarcomas is limited to ten cases [9]. We report here a solitary t(1;5)(q42;q32) in a case of MC. The translocation led to recombination of the IRF2BP2 and CDX1 genes.

This is the first time an IRF2BP2-CDX1 fusion has been detected in human neoplasia. CDX1 belongs to the homeobox gene family [18]. These genes share a homeobox domain that encodes a DNA binding protein functioning as a transcription

Table 2. List of primers.

<table>
<thead>
<tr>
<th>Primers</th>
<th>Sequence (5’ to 3’)</th>
<th>Tested fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEY1_F1</td>
<td>CGAGTGTTGAAAGGAGATGG</td>
<td>HEY1-NCOD2</td>
</tr>
<tr>
<td>NCOA2_E13-R3</td>
<td>AGTTGGGCTTGCAATGGGA</td>
<td>HEY1-NCOD2</td>
</tr>
<tr>
<td>CDX1-214F</td>
<td>CGCCAGTACCCCGACCTTCGCCA</td>
<td>CDX1-IRF2BP2</td>
</tr>
<tr>
<td>CDX1-369F</td>
<td>ATTCGGGCCCTCCAGCATTTA</td>
<td>CDX1-IRF2BP2</td>
</tr>
<tr>
<td>CDX1-659R</td>
<td>GTCGAGGAGGCCAGATGTCAGACTCA</td>
<td>IRF2BP2-CDX1</td>
</tr>
<tr>
<td>CDX1-771R</td>
<td>TGTATCCTGTGGCAGCTGC</td>
<td>IRF2BP2-CDX1</td>
</tr>
<tr>
<td>CDX1-26970R</td>
<td>GTCTCGGCTCCCCTCTGTGAGTGTGCT</td>
<td>IRF2BP2-CDX1</td>
</tr>
<tr>
<td>IRF2BP2-895F</td>
<td>CAAGAGCGCGGTCGGTGAGA</td>
<td>IRF2BP2-CDX1</td>
</tr>
<tr>
<td>IRF2BP2-926F</td>
<td>GTCAACAGAGCCAAAGCGGCGG</td>
<td>IRF2BP2-CDX1</td>
</tr>
<tr>
<td>IRF2BP2-1172R</td>
<td>CTGGACCGCTCCTGATTGAGGA</td>
<td>CDX1-IRF2BP2</td>
</tr>
<tr>
<td>IRF2BP2-1248R</td>
<td>GTGTTTCGCTGGTGAAAGGTCGG</td>
<td>CDX1-IRF2BP2</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0049705.t002
factor [19]. In particular, CDX1 is an upstream regulator of Hox-gene expression [20] that has been implicated in malignancies such as leukaemias [21,22]. In adults, CDX1 expression is restricted to intestinal epithelium [23–25] and aberrant expression has been linked to intestinal cancer [24,26–28]. No fusion gene involving CDX1 has so far been described as opposed to another member of the Cdx family, CDX2. CDX2 is overexpressed in both lymphoid and myeloid leukaemias [29–31] and a fusion gene resulting from a balanced t(12;13) leading to an ETV6-CDX2 fusion was detected in a patient with acute myeloid leukaemia [32].

The first exon of IRF2BP2 forms the 5`end of the IRF2BP2-CDX1 fusion. IRF2BP2 normally exists in two isoforms resulting from alternative splicing of the gene [17]. Both variants contain a Zinc finger motif at their N-terminus possibly binding DNA [17].

Although no direct link to cancer has been described for this gene, IRF2BP2 interacts with partners that are involved in cancer as for example the tumour suppressor gene TP53 [33] and the oncogene IRF2 [17]. IRF2BP2 also acts as a co-repressor of IRF2, inhibiting the expression of interferon-responsive genes. Recently also NFAT1, which encodes a transcription factor involved in the cell cycle, differentiation, and apoptosis, was shown to be repressed by IRF2BP2 [34]. According to BioGPS [35], IRF2BP2 is expressed in a variety of human tissues [36].

Two PCR products were obtained by cDNA PCR investigations for the IRF2BP2-CDX1 fusion. The difference between the two products was by sequencing shown to be caused by the alternative splice variants of IRF2BP2. Both sequences were shown to be in frame, with the largest transcript predicted to encode a 466 amino acid protein and the smaller encoding 450 amino acids. The
biological implications of the predicted fusion protein IRF2BP2-
CDX1 can only be speculated upon, but as both fusion partners
are involved in transcriptional regulation, a protein disturbing
DNA transcription is likely. The IRF2BP2-CDX1 fusion is thus
suggested to take part in MC tumorigenesis and/or progression.

MCs are rare, and at our institution only five patients (four
included in this study) received this diagnosis during the last 25
years. We identified the IRF2BP2-CDX1 in three of these tumours
(patients 2, 3, and 4), confirming that this fusion gene is common
in MC. In a majority of the well-known translocation-related
sarcomas such as myxoid liposarcoma and low-grade fibromyxoid
sarcoma, more than one defining fusion variant has been detected.
Often one fusion variant is more common than the others [2,3].
IRF2BP2-CDX1 could be such an additional fusion variant
identified in a subset of MCs, but only analysis of larger series of
tumours can determine the prevalence of the IRF2BP2-CDX1.

Of possible interest is the fact that the three HEV1-NCOA2-
positive MCs all had tumour manifestations detected in bone,
whereas the MC showing the t(1;5) and IRF2BP2-CDX1 fusion
originated from soft tissue. Although most common in bone, one-
fifth to one-third of MCs do arise in soft tissue [4]. The tissue of
manifestation was not reported in the MCs where the HEV1-
NCOA2 was first described [10]. Given the rarity of these tumours,
only future surveys of larger groups of patients can clarify if there
is a correlation between the tissue the tumour affects and the type
of fusion gene present. This study demonstrates the feasibility and
indeed advantage of using karyotyping and molecular cytogenetic
methods together with transcriptome sequencing to identify fusion
genes caused by chromosomal rearrangements. Traditionally,
chromosome walking using BACs or equivalent probes has been
used to narrow down the breakpoint regions followed by PCR.

References

Classification of Tumours. Pathology and Genetics of Tumours of Soft Tissue
and Bone. Lyon: IARCPress.


Fletcher C, Unni K, Mertens F, editors. WHO classification of solid and soft
tissue tumours. Lyon: IARCPress.


(2007) Mesenchymal chondrosarcoma of soft tissues and bone in children,
adolescents, and young adults: experiences of the CWS and COSS study groups.
Cancer 112: 127–139.

chondrosarcoma. An analysis of patients treated at a single


of novel co-expressor molecules for Interferon Regulatory Factor-2. Nuclease
Res 31: 3067–3076.

of all human homeobox genes. BMC Biol 5: 17.


an algorithm for gene fusion discovery in tumor RNA-seq data. PLoS Comput
Biol 7: e1001139.


Supporting Information

Table S1  BAC probes used for FISH experiments in case 1.
(XLS)

Table S2 List of fusions suggested by the deFuse
algorithm. (XLS)

Acknowledgments

The authors would like to thank Kirsten Sundby-Hall for help handling
register data. Researchers interested in more data are encouraged to
contact the authors directly.

Author Contributions

Conceived and designed the experiments: FM SH RAL RIS. Performed
the experiments: KBN IP JT LH LG MG TN. Analyzed the data: KBN IP
JT LG MG RIS. Wrote the paper: KBN IP JT LG TN RIS FM SH BB AF. Provided essential diagnostic information: BB AF. Revised the
manuscript critically: FM RAL SH.

IRF2BP2-CDX1 Fusion in Mesenchymal Chondrosarcoma
Paper IV

Francesca Micci, Jim Thorsen, Ioannis Panagopoulos, Kaja B. Nyquist, Bernward Zeller, Anne Tierens, Sverre Heim:

High-throughput sequencing identifies an \textit{NFIA/CBFA2T3} fusion gene in acute erythroid leukemia with \textit{t}(1;16)(p31;q24).

High-throughput sequencing identifies an NFIA/CBFA2T3 fusion gene in acute erythroid leukemia with t(1;16)(p31;q24)

Leukemia (2013) 27, 980–982; doi:10.1038/leu.2012.266

In a previous publication of ours\(^1\) we showed the involvement of the myeloid translocation gene-related protein 2 gene (CBFA2T3) in a case of acute erythroid leukemia with the translocation t(1;16)(p31;q24). Because of lack of material available for analysis, we could not with certainty determine the leukemogenic mechanism, whether it be generation of a fusion gene with CBFA2T3 as one of the partners or loss of tumor suppressor activity, in which case genes KANK1 and L1TD1, both homozygously lost, might be of the essence.

Some months after our article was published we got hold of 2 ml coagulated blood with 3% of abnormal cells from the same patient. As sequencing technology has also developed very fast lately, we decided to extract RNA from the sample and try to use it for high-throughput sequence analysis. The RNA was extracted and its quality checked by the Experion automated electrophoresis system (Bio-Rad Laboratories, Hercules, CA, USA). A total of 3 µg of RNA was sent for high-throughput pair-end RNA-sequencing to the Norwegian Sequencing Center at the Ulleval Hospital (http://www.sequencing.uio.no/). The Illumina software pipeline was used to process image data into raw sequencing data and only sequence reads marked as ‘passed filtering’ were used in the downstream data analysis. A total of 107 million reads were obtained. The FASTQC software was used for quality control of the raw sequence data (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). We used the fusion discovery software FusionMap (release date 16-04-2012)\(^2\) and the pre-built Human B37 and RefGene from the FusionMap website (http://www.omicsoft.com/fusionmap/). A list of over 500 possible fusion genes was obtained. A specific fusion involving the CBFA2T3 gene, which maps to chromosome band 16q24, and the nuclear factor I/A (NFIA) gene, which maps to chromosome band 1p31, was identified as number 10 in the list (seed count-rank 48; Supplementary Table 1). The involvement of the NFIA gene fits well with the fluorescence in situ hybridization (FISH) data on chromosome 1 previously obtained and published\(^3\). The presence of the NFIA/CBFA2T3 fusion was verified by PCR using the NFIA-956F and CBFA2T3–622R primer combination (Supplementary Table 2). A specific PCR product of about 500 bp was identified and directly sequenced (Figure 1a).

The specific fusion occurs between exon 6 of the NFIA gene (accession number NM_001134673.3) and exon 3 of the CBFA2T3 gene (accession number NM_005187.5; Figure 1b).

The NFIA/CBFA2T3 fusion gave an open reading frame and is expected to lead to a chimeric protein containing 208 amino-acid residues from NFIA (according to NP_001128145.1) and 603 residues from CBFA2T3 (according to NP_005178.4). The predicted fusion protein should thus consist of 811 amino acids (Figure 1c).

The NFIA gene encodes a member of the NFI family of transcription factors (http://genome.ucsc.edu). Interestingly, it has been found that NFIA exhibits a marked lineage-specific expression pattern in normal human hematopoiesis; it is upregulated in the erythroid lineage but fully suppressed in granulocytopoiesis.\(^4\) It has been shown that in early hematopoiesis, the NFIA expression level acts as a factor channeling hematopoietic progenitor cells into either the erythroid or granulocytopoietic lineage.\(^5\) The NFI proteins have a DNA-binding and dimerization domain in their N-terminal half, which contains four cysteine residues, and a transactivation and repression domain in their C-terminal half.\(^6\) It was found involved in an NFIA/EHF chimeric fusion in one breast cancer cell line out of 24 breast tumors analyzed (nine cell lines and 15 primary tumors).\(^7\) However, its role as either a passenger event or a direct, albeit infrequent, contributor to breast cancer development remains uncertain.

CBFA2T3 encodes an ETO myeloid translocation gene family protein, which interacts with DNA-bound transcription factors and recruits a variety of corepressors to facilitate transcriptional repression.\(^8–10\) The t(16;21)(q24;q22) translocation is one of the less common karyotypic abnormalities specifically associated with acute myeloid leukemia (AML). The translocation produces a chimeric gene made up of the 5’-region of the runt-related transcription factor 1 (RUNX1) gene fused to the 3’-region of CBFA2T3 (Figure 1d). In AMLs with either t(8;21) or t(16;21), the transcription factor RUNX1 is juxtaposed to one of the zinc finger nuclear proteins CBFA2T1 and CBFA2T3, respectively, resulting in transcriptional repression of RUNX1 target genes.\(^11\) Lately, its involvement as a partner in fusion genes was underlined by the identification of a IGH/CFBA2T3 fusion in a case of Burkitt lymphoma and a diffuse large B-cell lymphoma.\(^12\) This gene is also a putative breast tumor suppressor.\(^13,14\) Interestingly, CBFA2T3 is downregulated during erythroid differentiation, and it has been
suggested to have a repressive role in early, as well as late human erythroid differentiation. Hildebrand et al. demonstrated that the nuclear protein ETO (eight-twenty-one, a family to which also CBFA2T1, CBFA2T2 and CBFA2T3 belong) does not show reduced repressor activity even if it lacks the first 236 amino acids. As in the present fusion the altered CBFA2T3 protein lacks only the first 50 amino acids, we assume that its repressor activity is still retained (Figure 1e). More specifically, we hypothesize a pathogenetic parallel between AML showing a t(8;21) or t(16;21) and the present erythroleukemia with the 1;16-translocation with transcriptional repression of the NFIA target genes in the present case.

As the karyotype was described as 46,XY,der(1)t(1;1)(p31;q21),-del(1)(p11p31),der(16)t(1;16)(p31;q24), that is, presented additional rearrangement besides the 1;16-translocation, we decided to screen the list of possible fusion genes in search of genes located in karyotypic breakpoints to see if those were involved in fusions as well. We identified four possible fusions (seed count-rank >12) where one of the genes mapped to a breakpoint position on chromosome 1. An analysis of the hypothetical fusions using the BLAST program (http://blast.ncbi.nlm.nih.gov/Blast.cgi) showed in one of the two genes high-sequence identity with several genes and/or numerous repetitive sequences (for example, SINE). Hence, the reality of the putative fusions was seriously called into question and no further investigations were undertaken.

In addition to the present case, two more cases of erythroleukemia showing a t(1;16)(p31;q24) in their karyotype can be found in the Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer. All three patients (including ours) were very young children, and clinical outcome was poor. We assume that a NFIA/CBFA2T3 fusion existed also in these leukemias, but no evidence is at hand to corroborate or falsify this assumption.

In summary, we describe the first fusion gene identified in acute erythroleukemia. Knowledge of its specific functions in the neoplastic context is still incomplete, but pathogenetic similarities with other leukemic fusion genes are readily discernible. As for other leukemias characterized genetically by fusion genes, one may assume that the detailed pathogenetic knowledge now emerging may eventually form a starting point from which therapeutic attempts may begin.

Figure 1. Detection of the NFIA/CBFA2T3 fusion. (a) Gel picture showing the amplified fragment. Lane 1: ladder, lane 2: PCR product obtained with primer combination NFIA-924 F and CBFA2T3–622 R, lane 3: product of the NESTED-PCR obtained with primer combination NFIA-956 F and CBFA2T3–398 R. (b) Partial chromatogram showing the junction of the NFIA and CBFA2T3 genes. (c) Deduced amino acid sequence of the fusion transcript. (d) Schematic overview of the breakpoint region of the NFIA and CBFA2T3 genes. The exons are not in scale. Arrows point to primer positions. (e) Schematic overview of the position of the different domains of the NFIA and CBFA2T3 proteins and the NFIA/CBFA2T3 chimeric protein, according to ensembl (http://www.ensembl.org/index.html).
CONFLICT OF INTEREST
The authors declare no conflict of interest.

ACKNOWLEDGEMENTS
This study was approved by the Regional Ethics Committee (REK number: S-07474a) and the institutional review board. The study was supported by grants from the Norwegian Cancer Society and the South-Eastern Norway Regional Health Authority. The sequencing service was provided by the Norwegian High-Throughput Sequencing Center, a national technology platform supported by the ‘Functional Genomics’ and ‘Infrastructure’ programs of the Research Council of Norway and the South-Eastern Regional Health Authorities.

F Micci1,2, J Thorsen1,2, I Panagopoulos1,2, KB Nyquist1,2, B Zeller3, A Tierens4 and S Heim1,2,5
1Section for Cancer Cytogenetics, Institute for Medical Informatics, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway;
2Centre for Cancer Biomedicine, University of Oslo, Oslo, Norway;
3Department of Pediatrics, Rikshospitalet, Oslo University Hospital, Oslo, Norway;
4Department of Pathology, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway and
5Faculty of Medicine, University of Oslo, Oslo, Norway
E-mail: francesca.micci@labmed.uio.no

REFERENCES

Supplementary Information accompanies the paper on the Leukemia website (http://www.nature.com/leu)