Identification of translocations in single disseminated cancer cells

DISSERTATION ZUR ERLANGUNG DES DOKTOGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.) DER FAKULTÄT FÜR BIOLOGIE UND VORKINISCHE MEDIZIN DER UNIVERSITÄT REGENSBURG

vorgelegt von

Giancarlo Feliciello

aus

Ottaviano, Italien

im Jahr

2015

Das Promotionsgesuch wurde eingereicht am:

13.02.2015

Die Arbeit wurde angeleitet von:

Prof. Dr. Christoph A. Klein

Unterschrift:

Index of contents

List of abbreviations

Description	Abbreviation
Array Comparative Genomic Hybridization	aCGH
Base pair	bp
Circulating Tumor Cells	CTCs
Deoxynucleotide Triphosphates	dNTP
Deoxyribonucleic acid	DNA
Depletion Product	DP
Disseminated Cancer Cells	DCCs
Double Strand DNA	dsDNA
Genomic Integrity Index	GII
Kilobase	kb
Megabase	Mb
Microgram	µg
Microliter	µl
Micromolar	µM
Milliliter	ml
Millimolar	mM
Minimal Residual Disease	MRD
Molar	M
Multiplex Fluorescence In-Situ Hybridization	M-FISH
Nanogram	ng
Next Generation Sequencing	NGS
Phenol Emulsion Reassociation Technique	PERT
Polymerase Chain Reaction	PCR
Quantitative PCR	qPCR
Representational Difference Analysis	RDA
Single Strand DNA	ssDNA
Spectral Karyotyping	SKY
Whole Genome Amplification	WGA
Whole Genome Sequencing	WGS

1 Introduction

1.1 Cancer: general remarks

Cancer is the leading cause of death worldwide with 14.1 million new cancer cases and 8.2 million cancer deaths in 2012 (Ferlay et al., 2015). The complex and dynamic nature of cancer development makes difficult to unravel the fundamental key mechanisms that could prevent the patients from dying. One of the most enigmatic aspects of cancer pathogenesis is the complex multi-step process of metastasis starting from primary tumor. Metastasis is responsible for as much as 90% of cancer-associated mortality (Sleeman and Steeg, 2010). Despite the relative inefficiency of the metastatic process (Wong et al., 2001), the absolute magnitude of metastatic spread is enormous and represents the primary associated cause of cancer morbidity and mortality (Chaffer and Weinberg, 2011).

1.2 Molecular biology of primary tumor

The initiation of cancer needs the acquisition of some fundamental characteristics that constitute the major distinguishing features of cancer cells. The first trait a cancer cell has to gain is the ability to sustain chronic proliferation (Hanahan and Weinberg, 2011). In cancer cells this is possibly achieved by the continuous release of mitogenic signals, as growth factor ligands, binding the family of receptor tyrosine kinases (RTKs) localized at the plasma membrane, inducing the kinase activity of the receptor and activating signaling pathways that promote mitogenesis (Lemmon and Schlessinger, 2010). Also, activating mutations in mitogen-activated protein (MAP)-kinase pathway lead to cell survival and proliferation in cancer cells (Dhillon et al., 2007). Parallel to the acquisition of sustained proliferation, cancer cells have to avoid cellular programs that could suppress sustained growth. This growth-control role is usually carried out by the prototypical tumor suppressor such as retinoblastoma (RB) and tumor protein p53 (TP53). In particular RB integrates inhibitory signals coming from the intracellular and extracellular environments and regulates whether or not the cell should enter the division cycle (Burkhart and Sage, 2008). TP53 acts mainly by collecting the input derived from the intracellular stress or DNA damage and arrests the cell cycle till the repair of DNA, or it can trigger apoptosis (Meek, 2009).

Once cancer cells acquire unlimited proliferative capacity, they have to resist mechanisms of cell death such as apoptosis. In cancer cells, an imbalance between pro-

apoptotic (BAX, BID, BAD, BAK) and anti-apoptotic (BCL-X_L and BCL-2) molecules, in favor of the latter, leads to the evasion of apoptosis (Lessene et al., 2008).

At this point the cancer cells are ready to form macroscopic tumors but, to succeed, they need to develop unlimited replicative potential. A pivotal role in this regard is performed by the telomeres providing the capability for unlimited proliferation (Blasco, 2005). During the course of multistep tumor progression, clones of cancer cells at early stages often go through telomere loss-induced crisis due to their inability to express significant levels of telomerase (Kawai et al., 2007). This transient situation actually fosters tumor progression because telomere crisis produces significant chromosomal instability, increasing the occurrence of genetic alterations, which favor neoplastic transformation (Artandi and DePinho, 2010).

Furthermore, the continuous growth of the primary tumor increases the demand of nutrients and oxygen necessary for survival. In this regard, a mechanism called "angiogenic switch", leading to the formation of neovasculature (Hanahan and Folkman, 1996), is fundamental and seems to be a very early step in the development of the invasive disease as it has been observed in histological analyses of pre-malignant, non-invasive lesions and dysplasia (Raica et al., 2009). Vascular Endothelial Growth Factor-A (VEGF-A) is a prototypic pro-angiogenic factor, responsible for the formation of new blood vessels. It is up regulated by hypoxia and oncogenic signaling, key feature of the tumor (Carmeliet, 2005). An important property that cancer cells need to acquire is the activation of invasion mechanisms enabling the penetration into the surrounding tissues. The switch from a polarized, epithelial phenotype to a highly motile fibroblastoid or mesenchymal phenotype, called "epithelial-mesenchymal transition" (EMT), seems to be a fundamental requisite for the invasion and dissemination (Thiery et al., 2009).

1.3 Models of metastatic progression and metastatic cascade

For a long time the linear progression has been accepted as the only model of metastasis suggesting that the metastatic outgrowth is the final step of the primary tumor development (Fearon and Vogelstein, 1990). However, during the years the linear progression model has been placed side by side with the parallel progression model (Klein, 2009) (Figure 1).

According to the linear progression model, primary tumor cells undergo successive rounds of mutation and selection giving rise to a biologically heterogeneous cellular

population in which a subset of malignant clones have accumulated genetic alterations, necessary for metastasis (Vogelstein et al., 1988; Cairns, 1975). The metastatic capabilities may be developed at the primary site as a by-product of the selective pressures, or may further evolve after the tumor cells reached the secondary organs. Nevertheless, studies with the final aim to quantify human cancer growth rates concluded that the metastatic process must start before the clinical symptoms appear or the primary tumor is diagnosed (Collins et al., 1956; Friberg and Mattson, 1997).

In the parallel progression model tumor cells may disseminate very early, colonize multiple secondary sites at different times and ultimately accumulate genetic changes independently from those incurred by the primary tumor (Klein, 2009). This concept presupposes greater genetic disparity between the primary tumors and metastatic founders due to the fact that the early disseminated cancer cells (DCCs) are exposed to high selective pressure that is site-specific (Stoecklein and Klein, 2010). Moreover, the seeding of cancer cells can happen at different times to different sites and the outgrowth is determined by niche-adaptation of the DCCs (Klein, 2009). Early dissemination can be deduced from disease course and patient-derived data (Engel et al., 2003; Klein, 2009; Klein and Holzel, 2006). Moreover, recent reports provide evidence that support early dissemination of metastatic tumor cells in transgenic mouse models of breast cancer (Eyles et al., 2010; Husemann et al., 2008; Rhim et al., 2012). Corroboration for early tumor disseminations has also been found in pre-invasive lesions of patients with in situ carcinomas (Banys et al., 2012; Sanger et al., 2011). These means that, even before clinical detection of a primary tumor, cancer cells can invade the adjacent structures, from where they travel through lymphatic (lymphogeneous) and blood vessels (hematogeneous) as circulating tumor cells (CTCs). CTCs colonize distant organ sites, becoming DCCs, and eventually form microscopic deposits (micrometastasis), which may remain dormant, but ultimately leading to an overt metastatic disease (Aguirre-Ghiso, 2007).

Whether or not dissemination occurs late or early, the process by which cancer cells escape from the primary tumor to begin the metastatic cascade can be summarized in the following sequence of steps: 1) local invasion of the host stroma. Thin-walled venules, such as lymphatic channels, offer very little resistance to penetration by tumor cells and provide the most common route for tumor cell entry into the circulation (intravasation) (Fidler, 2003). 2) Detachment and embolization of single tumor cells or aggregates with most circulating tumor cells being rapidly destroyed. After the tumor cells have survived the circulation, they

become trapped in the capillary beds of distant organs by adhering either to capillary endothelial cells or to subendothelial basement membrane that might be exposed. 3) Extravasation from the bloodstream. 4) Survival and proliferation within the organ parenchyma which completes the metastatic process (Fidler, 2003; Valastyan and Weinberg, 2011).

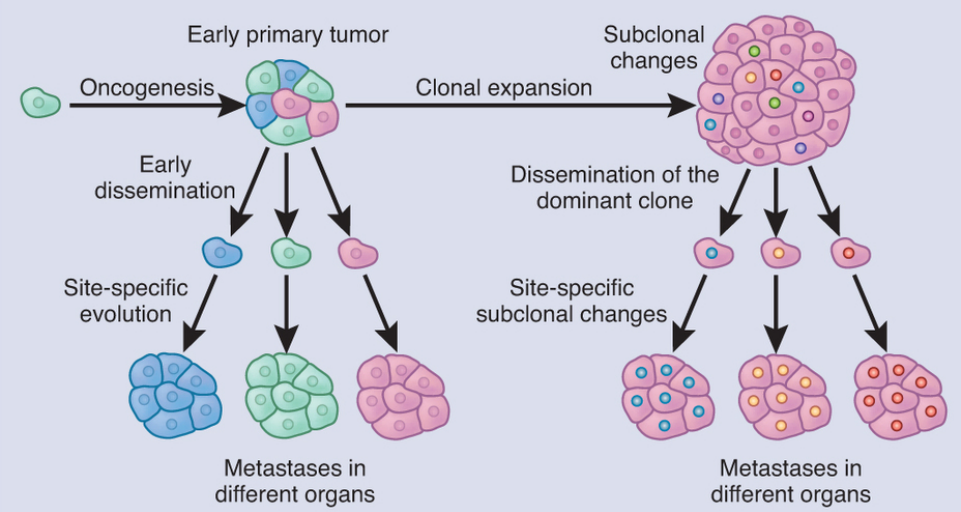


Figure 1: Metastasis progression models. Left, in the early dissemination (parallel progression model) different tumor cell clones are seeded in parallel to different organs. Right, in the late dissemination (linear progression model) tumor cells undergo clonal selection, during which the advantageous clones expand and dominate over the others, with additional subclonal and mutational changes occurring within the clonal populations. When disseminated, the dominant clone seed and colonize different organs. Figure adapted from (Wan et al., 2013).

1.3.1 Metastatic dormancy and reactivation at ectopic sites

Once cancer cells have found new soil, they need to adapt to foreign tissue microenvironments in order to grow from DCCs to micrometastases, and ultimately into macroscopic tumors. The adaptation to the metastatic niche can take a long time. In some cancers such as breast cancer and melanoma, macroscopic metastases may erupt decades after the primary tumor has been surgically removed. This means that these metastatic cells can pass through a state of dormancy before acquiring full metastatic growth (Aguirre-Ghiso, 2007). According to the current knowledge, the mechanism of dormancy can be summarized as a maladaptation of the DCCs to the new microenvironment because they are deprived of appropriate adhesive and signaling interactions (Liu et al., 2002; Shibue and Weinberg, 2009). So DCCs undergo dormancy as a consequence of intrinsic defects or in response to inhibitory signals originating in the stroma of target organs. Other systemic signals that appear to promote dormancy of micrometastic lesions are usually blocking neoangiogenesis (Hanahan and Folkman, 1996; Nyberg et al., 2005). During this lag time, DCCs need to overcome the

adaptation problem by activating other mechanisms like enhanced survival signaling (Chen et al., 2011; Gao et al., 2012; Zhang et al., 2009). After a variable time, a small minority of dormant DCCs undergoes reactivation in order to give rise to micrometastases. A large fraction of micrometastases regress because they fail to establish a tolerant microenvironment, further contributing to the inefficiency of colonization (Luzzi et al., 1998). However, small fractions of micrometastases that have adapted to the microenvironment grow into macroscopic lesions (Eyles et al., 2010).

Systemic and local niche signaling can drive the reactivation of metastatic pathways. Carcinoma cells can establish an accommodating niche in the target organ even before the seeding. In fact primary tumors release systemic factors that shape the local microenvironment of the niche and other factors that promotes angiogenesis (Psaila and Lyden, 2009). Organ-specific reactivators originating in the parenchyma of target organ, such as Coco, induce solitary breast cancer cells to undergo reactivation at lung metastatic sites (Gao et al., 2012). Additional signals, such as the expression of Tenascin C (TNC) can participate in the reactivation of micrometastatic lesions facilitating the activation of WNT and NOTCH pathways (Malanchi et al., 2012; Oskarsson et al., 2011). The metastatic outgrowth relies also on the recruitment of other cells of the tumor microenvironment like endothelial progenitor cells (Png et al., 2012), myeloid cells (Kim et al., 2009) and stromal fibroblasts (Elkabets et al., 2011) suggesting that neoangiogenesis, inflammation, and fibrosis foster this process (Joyce and Pollard, 2009).

1.4 Prognostic significance CTCs/DCCs as surrogate of minimal residual disease

The intrinsic metastatic potential of CTCs/DCCs makes them suitable surrogates of minimal residual disease (MRD), which is clinically not detectable (Riethdorf and Pantel, 2010). The presence of MRD may influence patient's prognosis despite successful tumor excision and completed adjuvant therapy. CTCs arrived quickly in clinical studies because of the easy accessibility of patient's blood samples. Numerous studies, mainly based on the enumeration of CTCs, have shown the prognostic significance of CTCs correlating the number of detected cells with the poor overall survival (OS) and disease free survival (DFS) in different types of cancers (Lianidou et al., 2014). Also for DCCs, relevance for patient prognosis has been indicated in relevant studies correlating their presence in bone marrow or lymph node with relapse and poor OS (Braun et al., 2000; Hosch et al., 2000; Kubuschok et al., 1999; Passlick et al., 1994; Passlick et al., 1999; Pierga et al., 2004; Thorban et al., 1996; Vashist et al.,

2012; Weckermann et al., 2001). Despite the numerous clinical data indicating that these cells are involved in the metastatic process, the genomic characterization of CTCs/DCCs is still incomplete and needs supporting data.

1.4.1 Available genomic data of the CTCs/DCCs

Genomic profiling studies have been performed on CTCs (Fabbri et al., 2013; Gasch et al., 2013; Heitzer et al., 2013; Ulmer et al., 2004), although most of the data derive from patients with metastatic disease (M1-stage) or from pooled CTCs (Magbanua et al., 2013; Magbanua et al., 2012). In these studies detection of late-stage genomic alterations is evident. Only few studies have pinpointed the attention to a deeper genomic characterization of single early-stage DCCs (Klein et al., 2002; Schardt et al., 2005; Schmidt-Kittler et al., 2003; Weckermann et al., 2009). DCCs isolated from patients without systemic tumor spread (M0-stage) are genetically different from patients with tumor spread (M1-stage). DCCs from M0-stage have a significantly low number of chromosomal changes than DCCs from M1 stage. For example, point mutations characteristic of the primary tumor are mainly absent (Klein et al., 2002) and typical chromosomal aberrations are often not present (Schmidt-Kittler et al., 2003; Weckermann et al., 2009). Additionally, there are no evidences of telomere crisis or loss/gains of whole chromosomes (Schmidt-Kittler et al., 2003). Instead there are general genetic instability and subchromosomal DNA changes (gain/losses) (Schardt et al., 2005). DCCs isolated from the bone marrow or lymph nodes of patients without manifest metastasis (M0-stage), do not share much chromosomal aberration and present heterogeneous genomes (Klein et al., 2002; Schardt et al., 2005; Schmidt-Kittler et al., 2003; Stoecklein et al., 2008). This finding is true also for DCCs isolated from the same patient (Klein et al., 2002; Schmidt-Kittler et al., 2003). On the contrary, when the metastatic disease spreads, the genome of the DCCs becomes stable and more similar between individual cells suggesting the expansion of an aggressive clone (Klein et al., 2002; Weckermann et al., 2009). Moreover, the genomes of DCCs in patients with metastatic disease are often similar, but not identical, to the predominant clone found in the primary tumors (Heitzer et al., 2013; Klein et al., 2002; Schmidt-Kittler et al., 2003). These data suggest that DCCs deriving from patients without overt metastasis disseminate in a genetically immature state without the typical genomic changes (Klein, 2013). Genomic instability accumulated outside the primary tumor, during years or decades, may play a role in the process of metastatic reactivation.

1.5 Genomic instability in cancer cells

Genomic instability is a hallmark of cancer that leads to an increase of genetic alterations enabling the acquisition of additional capabilities for tumorigenesis and progression. Genomic instability generates a variety of genetic alterations ranging from single nucleotide differences to large-scale changes at the chromosomal levels (Lengauer et al., 1998). The establishment of genomic instability is probably the initiating event as it has been shown that it is present in all stages of cancer, from precancerous lesions, even before *TP53* mutations are acquired (Bartkova et al., 2005; Gorgoulis et al., 2005), to advanced cancers. The most prevalent form of genomic instability, observed in over 90% of all malignancies, is chromosomal instability (CIN), and it is detected throughout the entire neoplastic transformation process, from premalignant lesions to metastatic lesions (Gagos and Irminger-Finger, 2005). In particular, chromosomal instability is occurring during the conversion from ductal hyperplasia to in situ carcinoma, that means before the cancer becomes morphologically invasive (Chin et al., 2004). CIN refers to alterations of segments of chromosomes, or whole chromosomes, in terms of their structure or number, including also amplifications, deletions, translocations, insertions and inversions (Geigl et al., 2008). So far it has been proposed that chromosomal imbalances are among the earliest changes of pre-cancerous lesions and are associated with hyperproliferation in breast cancer (Allred et al., 2001; Buerger et al., 1999; Waldman et al., 2000).

1.6 Gene fusion in hematological and solid cancers

Chromosomal aberrations, in particular translocations and their corresponding gene fusions, have an important role in the initial steps of tumorigenesis. At present 2276 gene fusions have been identified in both hematological disorders and solid tumors (http://cgap.nci.nih.gov/Chromosomes/Mitelman).

Historically, fusion genes have been associated only with hematological malignancies because the first discoveries have been made in these groups of neoplastic disorders. For example, the first translocation with oncogenic properties *BCR-ABL1* t(9;22)(q34;q11) was identified in chronic myelogenous leukemia (CML) (Rowley, 1973). This rearrangement forms a chimeric protein with increased tyrosine kinase activity and abnormal cellular localization. Around the same time, the *IGH-MYC* t(8;14)(q24;q32) translocation in Burkitt lymphoma was identified (Zech et al., 1976). This translocation links the immunoglobulin

heavy chain enhancer to the *MYC* oncogene leading to the enhanced activation (Mitelman et al., 2007; Rabbitts, 2009). Since then, dozens of fusions have been identified in hematologic cancers as the *IGH-BCL2* t(14;18)(q32;q21) that drives the follicular lymphoma (Vaandrager et al., 2000) or the gene fusions *PML-RARA* t(15;17)(q22;q12) in acute promyelocytic leukemia (APL) and *RUNX1-RUNX1T1* t(8;21)(q22;q22) in acute myeloid leukemia (AML) (Mitelman et al., 2007; Taki and Taniwaki, 2006).

Until recently, fusion genes caused by recurrent chromosome aberration were thought to be a feature of hematological and sarcoma cancers, but not of carcinomas, where other mechanisms such as deletion and point mutation were thought to be more important, and recurrent chromosomal aberrations were rarely seen. Thyroid carcinoma was the first epithelial tumor in which a gene fusion, *CCDC6-RET* inv(10)(q11.2;q21), was detected (Pierotti et al., 1992). Rearranged-during-transfection gene (*RET*) encodes for a receptor tyrosine kinase. The juxtaposition of the C-terminal region of the RET protein with an N-terminal portion of another protein leads to constitutive activation of the RET kinase (Mulligan, 2014). Together with this, other 14 gene fusions, 9 of which involve *RET* gene, have been identified in various subset of thyroid carcinoma with 40% of thyroid carcinomas harboring one of these chimeric genes (Pierotti, 2001).

TMPRSS2-ERG gene fusion has been shown to be the most common gene rearrangement found in 40-80% of prostate cancer cases (Soller et al., 2006; Tomlins et al., 2005). The fusion of these genes leads to the production of *ERG* under the control of the androgen sensitive promoter elements of *TMPRSS2*. This allows for a situation in which androgen-bound androgen receptor can bind these regions of *TMPRSS2*, resulting in the overexpression of *ERG*. This over-production of *ERG* can then exert its effects by binding target gene promoter regions, which results in their activation or inhibition, and the generation of a neoplastic phenotype (Tomlins et al., 2005).

The fusion gene *EML4-ALK* identified in non-small-cell lung cancer (NSCLC) is an aberrant fusion gene that encodes a cytoplasmic chimeric protein with constitutive kinase activity (Soda et al., 2007). *EML4-ALK* is uncommon, occurring in 2 to 7% of all non–small-cell lung cancers (Rikova et al., 2007). In general, while some fusions can be really specific by occurring between two invariant genes (i.e. *BCR-ABL*), it can also happen that one gene can be combined with many others, for example *MLL*, *RET*, *ALK* and the *ETS* family (Mitelman et al., 2004).

1.7 Mechanisms leading to the formation of chromosome translocations

The formation of a translocation requires three basic steps: firstly, the occurrence of multiple DNA double strand breaks on distinct chromosomes, secondly, the physical association of the broken ends, and finally the rejoining of the broken partner chromosomes (Roukos et al., 2013). Biologically, most, if not all, chromosome aberrations arise after DNA double-strand breaks (DSBs) that may arise spontaneously through replication errors, exogenous stress such as ionizing radiation and chemotherapeutic agents, or from scheduled breaks induced during development of the adaptive immune system, such as V(D)J recombination and immunoglobulin gene class-switch recombination (Aplan, 2006; Povirk, 2006). Additionally, complex damages that induce the collapse of the replication fork (Arnaudeau et al., 2001) are also responsible to generate DSBs. The main mechanisms joining the DSBs are homologous recombination (HR) and non-homologous end-joining (NHEJ). HR is mainly active during the S and G2 phases, in normal cells, at which a homologous sister chromatid is available. On the contrary, NHEJ is continually functional through the cell cycle (Branzei and Foiani, 2010; Shrivastav et al., 2008) without the need for template DNA. The characteristic of NHEJ is the absence of sequence homology at the end of DSBs that can result in either precise end-joining, insertion between breakpoints, or deletions (Lieber, 2010). In a subclass of NHEJ, the micro-homology-mediated end-joining (MMEJ), the process of re-joining can be facilitated by regions of micro-homologies between ssDNA exposed at the ends of the DSBs (Lieber, 2010) and it is responsible for structural mutations in primary cancers and cell lines (Chiarle et al., 2011; Hakim et al., 2012; Klein et al., 2011).

An important factor that affects the translocation outcome and plays a role for the occurrence of gene fusions is the spatial proximity of the involved genes within the nucleus (Cremer et al., 2006). In fact, cytogenetic studies have pointed to a strong correlation between spatial proximity of chromosomes or genes and their translocation frequencies by showing that proximal genome sites are more likely to form translocations than distal ones (Meaburn et al., 2007). For example, the spatial proximity of the *MYC* gene, relative to its possible translocation partners *IGH, IGK* and *IGL* in Burkitt's lymphoma, directly correlates with the observed frequency of these translocations in patients (Roix et al., 2003). Similar examples are *BCR-ABL* in chronic myeloid leukemia (CML) (Lukasova et al., 1997), *TPR-NTRK1* and *RET-ELE1* in papillary thyroid carcinoma (Nikiforov et al., 1999; Roccato et al., 2005). Factors that predispose genomic regions to breakage and translocations can be different. DNA sequence features as well as chromatin properties may facilitate breakage susceptibility of

genome regions (Figure 2).

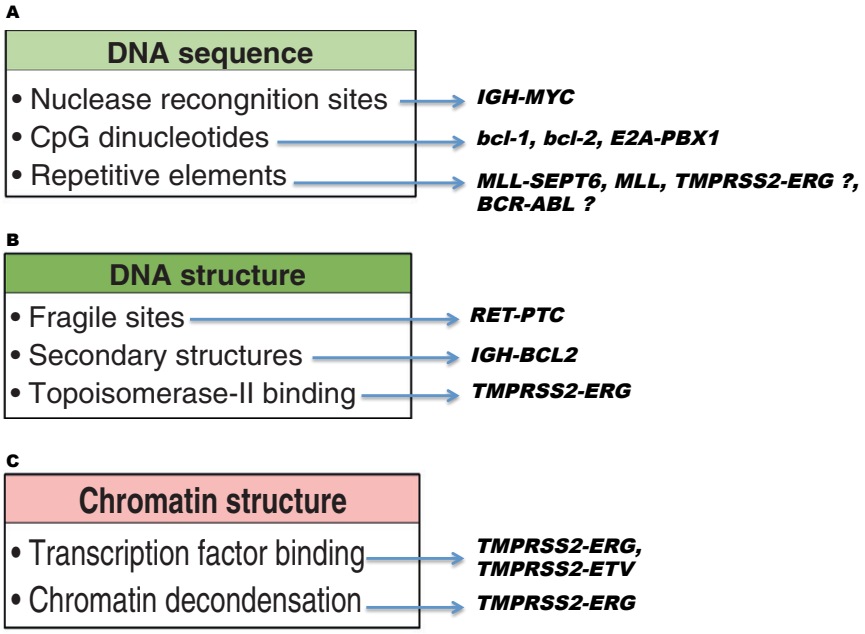


Figure 2: DNA and chromatin features involved in breakage susceptibility and examples of fusion genes formation.

1.7.1 DNA sequence features as susceptibility sites of breakage

DNA sequences, such as nuclease recognition sites, are recognized by endogenous nucleases and lead to the formation of DSBs and translocations. For example, in germinal center B cells, activation-induced cytidine deaminase (AID) recognizes a single-strand sequence motif during the transcription of regions involved in somatic hypermutation and class-switch recombination and promotes DSBs to generate antibody diversity (Stavnezer et al., 2008). However, misrecognition of non-Ig targets can generate translocations as observed in B cell lymphomas for *IGH-MYC* (Okazaki et al., 2007; Robbiani et al., 2008) (Figure 2A). It has been shown that in prostate cancer, AID is co-recruited with liganded androgen receptor (AR) to AR-binding DNA sequences, sensitizing them to DSB breaks and leading to the formation of translocations involving *TMPRSS2*, *ERG* and *ETV1* (Lin et al., 2009).

1.7.2 CpG islands and repetitive elements

CpG islands represent also candidates regions prone to breakage. In fact, CpG dinucleotides are present in 40–70% of *bcl-2* and *bcl-1* breakpoints in pro-B and pre-B lymphocytes suggesting that CpG islands facilitate the breakage (Tsai et al., 2008) (Figure 2A).

Repetitive elements comprise half of the human genome and recent studies suggest

that repeats such as LINE-1 and Alu elements are major contributors to structural variation (Beck et al., 2010; Ewing and Kazazian, 2010; Huang et al., 2010) (Figure 2A). These can join either with other Alu elements or with non-Alu sequences. Alu–Alu intra-chromosomal recombination was seen in mixed lineage leukemia (*MLL*) gene, involved in acute myeloid leukemia (Hess, 2004; So et al., 1997). In the Philadelphia chromosome, *BCR* and *ABL1* genes contain Alu elements near their breakpoint regions, which may possibly lead to the occurrence of Alu-mediated DNA recombination (Jeffs et al., 2001). Furthermore, fine mapping of the deletion breakpoints located within the *ERG* and *TMPRSS2* loci, followed by sequencing, revealed the presence of consensus sequences homologous to the human Alu-Sq and Alu-Sp subfamily (Liu et al., 2006). The presence of these consensus sequences within intronic regions correlated with the presence of the fusion gene and may be a factor contributing to the deletion at 21q22.2-3, resulting in the fusion gene.

1.7.3 Common fragile sites and DNA structures

Common fragile sites (CFS) have also been linked to translocations (Arlt et al., 2006) (Figure 2B). CFSs are regions of chromosomes containing gaps and constrictions in metaphase under partial replication stress, and these regions have been shown to be prone to breakage (Ozeri-Galai et al., 2012). A potential link between fragile sites and translocation formation comes from the observation that exposure of thyroid cells to chemicals that induce fragile sites promotes *RET-PTC* translocations (Gandhi et al., 2010). Some common sequence features have been identified responsible for the emergence of CFSs. CFSs are enriched in strings of AT-dinucleotide repeats that give these regions high DNA helix flexibility and the ability to form stable non-B DNA secondary structures, which may inhibit DNA replication (Inagaki et al., 2009). In fact, translocations have been postulated to form at AT palindromic sequences through a mechanism involving cruciform DNA structures that may be prone to breakage (Kurahashi et al., 2010). Indeed, computational analysis of five translocation genes (*CBFB*, *HMGA1*, *LAMA4*, *MLL*, and *AFF4*) revealed significantly high AT content (Burrow et al., 2009).

More direct evidence for DNA secondary structures in breakage and translocations was the discovery that the major breakpoint region of *bcl-2* adopts a stable non-B DNA structure that is targeted by recombination-activating genes (RAGs) in a sequence-independent manner (Raghavan et al., 2004). By containing stable regions of single-strands, this DNA structure promotes RAG-mediated cleavage of the *bcl-2* locus and formation of the

translocation *IGH-BCL2* t(14;18)(q32;q21) in follicular lymphoma (Raghavan et al., 2004). This secondary structure may be a 'G-quadruplex,' a four-stranded DNA structure that can spontaneously form in G-rich sequences (Katapadi et al., 2012) (Figure 2B).

Topological features of DNA may also contribute to breakage susceptibility. Topoisomerase II (TOP2) generates a transient DSB to regulate under-winding and over-winding of DNA, for example in mitotic chromosomes and in replication, and also during transcription (Felix et al., 2006; Ju et al., 2006). The TOP2 beta isoform has been shown to associate with androgen receptor upon transcriptional activation and to trigger DSBs at *TMPRSS2* and *ERG* breakpoints in prostate cancer (Haffner et al., 2010) (Figure 2B).

1.7.4 Chromatin structure and translocations

Various aspects of chromatin may play a role in chromosome breakage susceptibility and translocations (Figure 2C). DSBs occur primarily in transcriptionally active regions. Breakpoints in, or near, transcriptionally active genome regions have been documented in anaplastic large cell lymphoma (Mathas et al., 2009). Similarly, liganded androgen receptor, a potent transcriptional activator, binds near the breakpoints of *TMPRSS2*, *ERG*, and *ETV*, which are involved in translocations in prostate cancer and under genotoxic stress, induces translocations (Lin et al., 2009). This means that chromatin remodeling and binding of transcription factors may predispose genomic regions to breakage and translocations. Histones, that modulate transcription, are also potential candidates in DSB susceptibility and translocation mechanisms. Genome-wide mapping of a set of histone modifications in seven primary human prostate cancers has indicated possible enrichment of active chromatin marks, H3K4me3, H3K36me3, and acetylated H3, over the *TMPRSS2-ERG* translocation region (Berger et al., 2011).

1.8 Gene fusion as primary events of chromosomal instability in early tumorigenesis and potential relevance for CTCs/DCCs

Translocations play an undisputed role in the initial steps of carcinogenesis and it is estimated that they are causal in 20% of cancers. There is striking evidence in support of their role the in the initiation of cancer. First, they are usually closely correlated with specific tumor phenotypes (Borden et al., 2003; Johansson et al., 2004). Moreover, it has been shown, predominantly in hematological malignancies, that successful treatment is associated to a

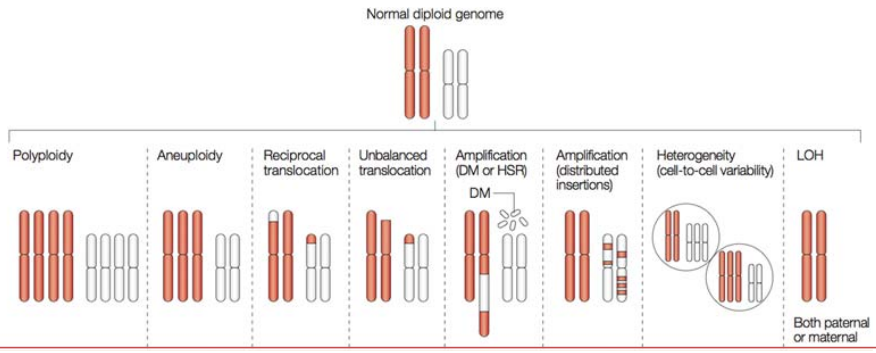
decrease or eradication of the disease-associated gene fusion (Deininger et al., 2005; Kern et al., 2005). Supporting studies in experimental animal models have shown that gene fusion constructs generally give rise to neoplastic disorders of the same type as those seen in sporadic human neoplasms that carry the same gene fusion (Rego et al., 2006). Finally, silencing fusion transcripts *in vitro* leads to the reversal of tumorigenicity, decreased proliferation and/or differentiation (Thomas et al., 2006). The identification of balanced structural chromosome changes is therefore very important, and the breakpoints involved could indicate the location of genes relevant for early metastatic progression.

Based on these assumptions it will not be surprising that CTCs/DCCs, representing early metastatic cells, would contain gene fusions as some of the initial genomic aberration necessary for the migration and survival at ectopic site. For example, it has been shown that *TMPRSS2-ERG* is an early event necessary for the transition from prostatic intraepithelial neoplasia (PIN), a precursor lesion of prostate cancer, to invasive prostate cancer (Perner et al., 2007). In addition, transgenic studies have reported that mouse prostate with *TMPRSS2–ERG* gene fusion alone develops PIN (Tomlins et al., 2008), but when cooperating with other oncogenic pathways, such as *PTEN* deletion or androgen receptor (AR) overexpression, it leads to the development of prostatic adenocarcinoma (King et al., 2009; Zong et al., 2009). So *TMPRSS2-ERG*, although alone is not responsible for cancer progression, can be considered a major driver of tumorigenesis by regulating and cooperating with other mechanisms (Sun et al., 2008; Wang et al., 2008; Yu et al., 2010).

1.8.1 Technical challenges to detect gene fusions in early CTCs/DCCs

Methods to identify chromosomal changes, which target different kind of aberrations, are available. In general, their applicability to single cells, such as for the deeper genomic characterization of CTCs/DCCs, is possible but still very challenging. In order to identify recurrent and rare gene fusions with a role in early metastasis and to use them as a target to decrease, or even eradicate, MRD after the curative surgery, novel methods are needed. Figure 3 outlines some of the current methods used to identify genomic abnormalities. As can be seen, while some chromosomal changes can be easily assessed from the use of different technologies, genes fusions as balanced translocations are detected only with techniques such as spectral karyotyping (SKY) and multicolor FISH (M-FISH), but still with technical limitations. Even worse, karyotyping of solid tumors is more challenging due to poor chromosome morphology, and the karyotypes are often so complex that they cannot be

characterized completely (Mitelman et al., 2004). In fact, due to the complexity of combinatorial labeling used to color the chromosomes, unknown cryptic translocations may be very difficult to detect, and even the identification of known balanced rearrangements depends critically on the fluorochrome combination in the chromosomes involved in rearrangements (Kearney, 2006). Moreover the labor intensity required and the low resolution makes them not suitable for the screening of chromosomal rearrangements. It is now clear, with the advent of the next generation sequencing (NGS) that the lack of recurrent gene fusions in epithelial tumors stems from the difficulties of performing cytogenetic analysis and that the actual proportion of malignancies with recurrent rearrangements may be as high in epithelial tumors as in hematological malignancies (Mitelman et al., 2004). In fact, the discovery of *TMPRSS2-ERG* lead to the conclusion that, considering the high incidence of prostate cancer and the high frequency of the fusion, this is the most frequent gene rearrangement described (Tomlins et al., 2005). The advent of NGS technologies has started to unravel a complex universe of chromosomal translocations and to show that some are unique to individual tumors and even different from patient to patient. So only whole genome sequencing (WGS) approaches have the power to detect all the somatic mutations in a cell but the costs are still very high to make it affordable and applicable for routine screening.



Technique	Detection							
Banding	+	+	+	+	+	-	+	-
M-Fish/SKY	+	+	+	+	+	+	+	-
(Array)CGH	-	+	-	+	+	+	(+)	-
SNP-array	-	+	-	+	+	+	(+)	+
Digital Karyotyping	-	+	-	+	+	+	-	+
Whole Genome Sequencing	+	+	+	+	+	+	+	+

Figure 3: Comparison of techniques for identifying chromosomal abnormalities. A '+' indicates that an approach is suited for identifying the chromosomal rearrangement, a '-' indicates that the aberration would be missed. A (+) indicates that several experiments are needed. Adapted and modified from Speicher MR, Nat Rev Genet, 2005

1.9 Aim of the work

The aim of this dissertation was to establish a method for the identification of novel genomic breakpoints in single-cell genomes. Available techniques used to identify chromosomal translocations, as karyotyping by G-banding showed to have a breakpoint resolution in the magnitudes of 5 to 15Mb. Interphase (and metaphase) FISH can only detect known genetic aberrations, providing that the specific probe is available. Therefore, FISH cannot serve as a screening test for chromosomal rearrangements since most FISH techniques can only detect known imbalances. SKY (and M-FISH) can detect multiple karyotype abnormalities simultaneously, but both techniques are dependent on combined fluorochrome probes. In addition, these techniques can be labor intensive. To overcome these limitations, a method to allow the isolation and enrichment of fusion breakpoints in single-cell genomes had to be developed.

To identify genomic breakpoints at nucleotide resolution, the method should meet the following criteria: (1) applicability to single-cell Ampli1 whole genome amplification (WGA) products, (2) reduction in complexity by subdividing the WGA in distinct genomic populations according to determined nucleotide combinations, (3) enrichment of breakpoints contained in the subdivided genomic populations by subtractive hybridization technique, (4) monitoring the enrichment of genomic breakpoints by means of quantitative real time PCR, (5) high sensitivity enabling detection of low copies genomic breakpoints, (6) high specificity minimizing the rates of false-positive chimeric sequences.

2 Materials and Methods

2.1 Materials

Table 1: Chemicals

Product	**Company**
4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES) ($C_8H_{18}N_2O_4S$) pH 8.0	Applichem GmbH
Agarose, for routine use	Sigma-Aldrich
Ammonium Acetate 3M pH 5.2 (CH_3COONH_4)	Merck KgaA
Bromphenolblu ($C_{19}H_{10}Br_4O_5S$)	Fluka (Biochemika)
Chloroform 99.8% pure ($CHCl_3$)	Sigma-Aldrich
Deionized Formamide (CH_3NO)	Sigma-Aldrich
Ethanol (C_2H_6O) 99,9 %	J. T. Baker
Ethidiumbromid ($C_{21}H_{20}BrN_3$)	Fluka BioChemika
Ethylenediaminetetraacetic acid (EDTA-$C_{10}H_{16}N_2O_8$)	Sigma-Aldrich
Hank's buffered salt solution, with $NaHCO_3$ 10X (HBSS)	Biochrom AG
Igepal CA-630 (Octylphenoxypolyethoxyethanol) (C_2H_4O)n$C_{14}H_{22}O$	Sigma-Aldrich
Isopropyl-beta-D-thiogalactopyranoside (IPTG) ($C_9H_{18}O_5S$)	Biomol GmbH
Linear Polyacrylamide Carrier (LPA) 5mg/ml	Ambion, Inc.
Magnesium Acetate BioUltra, 1M in H_2O ($Mg(CH_3COO)_2$)	Sigma-Aldrich
Magnesiumchlorid ($MgCl_2$)	Carl Roth GmbH+Co
Nonidet P40 ($C_{18}H_{30}O3$) (NP40)	Sigma-Aldrich
PCR-H_2O (LiChrosolv for Chromatography, DNAse-RNAse free)	Millipore-Merck
PCR-Oil	Sigma-Aldrich
Percoll	GE Healthcare Life Sciences
Potassium Acetate (CH_3COOK)	Sigma
Potassium Chloride (KCl)	Merck KgaA
Potassium dihydrogenphosphat (KH_2PO_4)	Merck KgaA
Roti-Phenol/Chloroform/Isoamyl alcohol 25:24:1 pH 7.5 -8.0	Carl Roth GmbH&Co
RPMI (Roswell Park Memorial Institute Medium) medium 1640	Gibco BRL

Sodium Chloride (NaCl)	J. T. Baker
Sodium dihydrogenphosphat ($Na_2H_2PO_4$) Monohydrat	Merck KgaA
Sodium dodecyl sulfate (SDS) ($C_{12}H_{25}O_4SNa$)	Serva
Tris EDTA-Buffer 1X - pH 8,0	AppliChem GmbH
Tris-HCl pH 8,8 ($C_4H_{11}NO_3$)	Applichem GmbH
Trizma Acetate BioUltra, >99.0% ($NH_2C(CH_2OH)_3 \cdot CH_3COOH$)	Sigma-Aldrich
TWEEN® 20 ($C_{58}H_{114}O_{26}$)	Sigma-Aldrich
X-Gal (5-bromo-4-chloro-3-indolyl-ß-D-galactoside) ($C_{14}H_{15}BrClNO_6$)	Roche
Xylencyanol ($C_{25}H_{27}N_2NaO_6S_2$)	Merck KgaA

Table 2: Consumables for cell culture and cloning

Product	**Company**
Ampicillin Na-Salt ($C_{16}H_{18}N_3O_4SNa$)	Biomol GmbH
Bovine Serum Albumin, bioreagent 20 mg/ml, Protease Free (for picking)	Sigma-Aldrich
Fetal Bovine Serum (FBS) Sera Plus	Pan-Biotech
K562 leukemia cell line	Sigma-Aldrich
LAB-Tek Chamber slides, 8 well	NUNC GmbH
NEB 10-beta Competent E. Coli (High Efficiency)	New England Biolabs
Penicillin (100x) ($C_{16}H_{18}N_2O_5S$)	PAA Laboratories
pGEM®-T Easy Vector System I	Promega
Pepton (Select Pepton 140)	Invitrogen
Petri dish (90 mm)	Greiner Bio-One
Qiagen Plasmid Midi Kit	Qiagen
RPMI (Roswell Park Memorial Institute)	Gibco
Streptomycin (100X) ($C_{21}H_{39}N_7O_{12}$)	PAA Laboratories
Trypton/Pepton from Casein	Carl Roth GmbH+Co
Yeast extract (Select Yeast Extract)	Invitrogen

Table 3: Consumables for Whole Genome Amplification (WGA), Fractioning, Subtractive Hybridization and Quality Control

Product	**Company**
Adenosine-5'-Triphosphate (ATP-10mM)	Roche
Agilent High Sensitivity DNA Kit	Agilent
Bovine Serum Albumin (BSA) 20 mg/ml	Roche
dATP, dTTP, dCTP, dGTP (Sequencing Grade)	GE Healthcare Life Sciences
DNA Molecular Weight Marker 2-Log	New England Biolabs
Dynabeads® M-280 Streptavidin	Invitrogen
Expand Long Template PCR System	Roche
LightCycler FastStart DNA Master SYBR Green I	Roche
MAXYMum Recovery™ PCR Tubes	Axygen Scientific
Microcon YM-30	Millipore
Nanolink Streptavidin Magnetic Beads	Solulink, Inc.
Phase Lock Gel	5' Prime
Pipets (P2, P20, P200, P1000)	Gilson
QIAprep Miniprep Kit	Qiagen
QIAquick Gel Extraction Kit	Qiagen
Qubit dsDNA Reagents and Kits	Invitrogen
Trimmer-2 cDNA normalization kit	Evrogen
UltraPure Herring Sperm DNA (10 mg/ml)	Invitrogen
Yeast tRNA (10mg/ml)	Invitrogen
Wizard® SV PCR Clean-Up System	Promega

Table 4: Buffers and solutions

Buffers and solutions	**Composition**
Binding Buffer (BP2X) for Dynal Streptavidin-Beads	2 M NaCl, 20 mM Tris-HCl (pH 8,0), 2 mM EDTA (pH 8,0)
Binding Buffer (BP1X) for Dynal Streptavidin-Beads	1 M NaCl, 10 mM Tris-HCl (pH 8,0), 1 mM EDTA (pH 8,0)
HEPES-Buffer	140 mM NaCl, 1,5 mM Na_2HPO_4, 50 mM HEPES. Sterile filtered (0,22 µm).
Hybridization Buffer	50 mM HEPES, 0.2 mM EDTA, pH 8.9

Nucleic Acid Binding and Wash Buffer (BW)	50 mM Tris-Hcl, 150 mM NaCl, 0,05% Tween 20, pH 8.0
One-Phor-All Buffer plus (OPA+)	100mM $NH_2C(CH_2OH)_3 \cdot CH_3COOH$ pH 7.5, 100mM $Mg(CH_3COO)_2$, 500mM CH_3COOK
Phosphate Buffered Saline (PBS1X)	8,5 mM Na_2HPO_4, 2 mM KH_2PO_4, 150 mM NaCl, pH 7,4
PCR-Buffer with dNTPs 10x	500 mM KCl, 100 mM Tris HCl (pH 8,5), 10 mM $MgCl_2$, each 1 mM dATP, dTTP, dCTP, dGTP
Tris-Solution	50 mM Tris/HCL (pH 8,9)
Washing Buffer 1 50 ($WP_1$50)	0,1% SDS in HEPES-Buffer, 50% Formamide (deionized) in WP1
Washing Buffer 2 50 ($WP_2$50)	0,05% Igepal in HEPES-Buffer, 50% Formamide (deionized) in WP2

Table 5: Enzymes

Product	Company
BspTI (AflII) (10 U/µl) plus Buffer O (10X)	Thermo Fisher Scientific
DNA Polymerase Mix (5 U/µl) Expand Long Template PCR System	Roche
DraI (10 U/µl) plus Buffer Tango (10X)	Thermo Fisher Scientific
Exonuclease I (20U/µl) plus ExoI Buffer (10X)	Thermo Fisher Scientific
KspAI (HpaI) (10 U/µl) plus Buffer B (10X)	Thermo Fisher Scientific
Mse I (50 U/µl)	New England Biolabs
Mung-Bean Nuclease I (10 U/µl) plus MBN Buffer (10X)	New England Biolabs
Proteinase K (14-22 mg/ml)	Roche
Quick Ligase (50U/µl) - Quick Ligation Kit	New England Biolabs
T4 DNA Ligase (≥ 5U/µl)	Roche
Taq DNA Polymerase (5 U/µl)	Roche
TruI (50U/µl) plus Buffer R (10X)	Thermo Fisher Scientific
Vsp I (AseI) (10 U/µl) plus Buffer O	Thermo Fisher Scientific
XmnI (20U/µl) plus Buffer 4	New England Biolabs

Table 6: Equipment

Product	**Company**
Electrophoresis Chamber	Biorad
MJ Research PCR Thermal Cycler	MJ Research
LightCycler 480 Instrument II 96 wells	Roche

Magnetic Rack (MPC-9600)	Invitrogen
Nanodrop 2000	Thermo Fisher Scientific
Qubit 2.0 Fluorometer	Invitrogen
Roller SRT6	Stuart Scientific
Table Centrifuge 5417C	Eppendorf
Termomixer	Eppendorf

Table 7: Oligonucleotides

Oligonucleotide*	**Sequence 5´→3´**	**Modification***
23R´C	TCG CAC TCG ACT ATC CTC TGG TT	
23R´G	GAG CAA ATG ACC ATG CTA GCA CC	
23R'Am	AAT TCG TCA ATC AGC TGC GCT TG	
23R'Tm	TAG TCT CGG CTC AAC GTC AGG AT	
26RAn	AAT TCG TCA ATC AGC TGC GCT TGT TT	5´Bio*
26RC	CTA TCG CAC TCG ACT ATC CTC TGG TT	5´Bio*
26RGn	TCT GAG CAA ATG ACC ATG CAG CAC C	5´Bio*
26RTn	GTA TAG TCT CGG CTC AAC GTC AGG AT	5´Bio*
ddMse11	TAA CTG ACA G	3´ddC
J13A	TAA TGT CTG AGA C	3´ddC
J13C	TAA GGT CTG AGA C	3´ddC
J13G	TAA CGT CTG AGA C	3´ddC
J13T	TAA AGT CTG AGA C	3´ddC
J20n	TAG GTT CCA GCG TCT CAG AC	
J21	AGT CTC CAG GCT CTC ACC GCT	
J24A	TGT AGG TTC CAG CGT CTC AGA CAT T	
J24C	TGT AGG TTC CAG CGT CTC AGA CCT T	
J24G	TGT AGG TTC CAG CGT CTC AGA CGT T	
J24T	TGT AGG TTC CAG CGT CTC AGA CTT T	
Lib1	AGT GGG ATT CCT GCT GTC AGT	
R12C	AAC CAG AGG ATA	3´ddC
R13An	AAA CAA GCG CAG C	3´ddC
R13Gn	T TA AGG TGC TAG C	3´ddC

R14Tn	TAA TCC TGA CGT TG	3´ddC
R21	GAC CTC GAC TAT CCA GTG ACT	

*: All oligonucleotides were HPLC purified.

**: Oligonucleotides are also used without modifications for the generation of the A/C, A/G, A/T, C/G, C/T, and G/T-fractions.

Table 8: Adaptors

Adaptor*	**Long Oligonucleotide**	**Short Oligonucleotide**
26RA	26RA	R12A
26RAn	26RAn	R13An
26RC	26RC	R12C
26RG	26RG	R15G
26RGn	26RG	R13Gn
26RT	26RT	R13T
26RTn	26RTn	R14Tn
J21	J21	J12
J24An	J24An	J13An
J24Cn	J24Cn	J13Cn
J24Gn	J24Gn	J13Gn
J24Tn	J24Tn	J13Tn
Lib1	Lib1	ddMse11
R21	R21	R12

Table 9: Oligonucleotides for specific MseI-fragment control PCR

Gene/Marker	**D***	**Fragment** [bp]**	**Position [GRCh38]**	**Primer**	**Sequence 5´->3´**	**T*** [°C]**	**PCR-Product [bp]**
D5S2117	AA	1376	133701437-133701583	5´	CCA GGT GAG AAC CTA GTC AG	58	147
				3´	ACT GAG TCC TCC AAC CAT GG		
BCR	AA	751	23210945-23211187	5´	GAC CAT TCA GTG GTG GTT AG	61	243
				3´	GCT CTA AGA ACT CTG TCA CC		
BRCA1	AA	594	43052727-43052877	5´	GGA TGG CCT TTT AGA AAG TGG	60	151
				3´	ACA CAG ACT TGT CCT ACT GCC		
PDPR	AA	568	70160956-70161099	5´	AAT CTT CTG TGG CCC ACA GT	55	144

				3´	GAC CCT TAT TGA CTA CAC TGG		
TP53	AA	558	7673515-7673888	5´	AGG ACC TGA TTT CCT TACTGC	58	374
				3´	GAG GTC CCA AGA CTT AGT AC		
LPO	AA	316	58242980-58267994	5´	CAG TTA GCC AAG ATA ATG CC	58	193
				3´	CTC ATA AGG GGA TAA AGG AC		
TRAK1	AA	247	42212037-42212250	5´	GGG AGA ACA ATG AAT TGG CTC	55	214
				3´	GAC CAG ACC ACC ATG GAT AC		
TRAK1	AA	128	42210614-42210726	5´	TAG AAC CAG GAA AGG CTC AG	55	113
				3´	CAC TTT GAA CAC ATG TGT AAG C		
BCR	TT	1936	23268437-23268849	5´	CGT GGA CAA CTA CGG AGT TG	61	413
				3´	TCA GCC TCA GGA CTC TTG TG		
D16S3066	TT	536	73295993-73296182	5´	GCT GTT AAT ATG AAA CAA TTG CC	58	190
				3´	GGG GTC TAA TGG TTC AGC C		
BCR	TT	508	23218452-23218686	5´	CAC TAG TTG TCC CTG CTC AG	58	235
				3´	GTC CAGTCA CTT GTG G		
CLASP2	TT	294	33510459-33510653	5´	CTG GCT GCA ATC AAA ATG CAA	58	195
				3´	GAA GGC TTG ATC ATG CCA GT		
D6S314	TT	277	139693305- 139693514	5´	TAA TCC ACT TCC TGA CCT AG	60	210
				3´	CTT CCC AGG TGA TTC TCA TG		
COG4	TT	261	70512286-70512502	5´	ACA CAA AGT GGC ATG GAC CT	55	217
				3´	AAC CCA GCA GTG GGA AGA TC		
TFRC	TT	242	196069661-196069791	5´	GCA TGA AAG ACA TAG GAG GC	58	131
				3´	CAG GCA ACC CAA GTA AGA GAT		
TOP2B	TT	209	25629781-25629957	5´	GAG ACT TGC CCG AAG ATG GT	58	177
				3´	CTA CTT GTA GGT AGG CTG CTA		
ABL1	TT	173	130779351-130779475	5´	CAT CGT TCA TGATGG CAA GG	58	125
				3´	GAA ATG TAC TGC GTT ATG CC		
D5S500	TT	159	138511343-138511474	5´	CTT TTT ACA TTT TTG GTA CCT TG	58	132
				3´	GTA TCG GTG AAA TGC AAC TA		
ABL1	TT	157	130749086-130749227	5´	TTA TAC TAG ATC TTG AGC ACC	60	142
				3´	CCC TCA TAA CTA CAA GAA CTC		
BCR	CC	1233	23297396-23297719	5´	TTC GGT CCT TGC AGC AGA TC	61	324
				3´	GTT GGG CAG AAT CTG AAC CTC		
ABL1	CC	291	130732130-130732401	5´	AGC TAC TTA GTG AAT AACG CAG	58	273

				3´	GAT CAG GTT GAG TAA ATG TCC		
ABL1	CC	203	130715659-130715806	5´	GTG ATG ATC ATG CAC CTT AC	59	148
				3´	CAC AGG GAA CTA CAC TGC		
BCR	GG	1026	23309463- 23309759	5´	ACC GTC ATC GCC ATG AAT GG	60	297
				3´	CCA GGT CCC TGA GTG GAG		
BCR	GG	944	23295308-23295667	5´	CGG TGC CTG CAT GTA TCT TG	61	360
				3´	ATT AGC CTG GGA GGA GAC C		
BCR	GG	715	23272238-23272706	5´	TGT CAG GAA TTG TCA GTC AC	60	470
				3´	AGC AGA CCT CTT AGC TGC TC		
D16S3019	GG	605	66095305-66095505	5´	CAA CTC ATT CCC TGT GTG AC	57	220
				3´	AAC CAA GTG GGT TAG GTC AG		
BCR	GG	367	23236242-23236571	5´	CTG GTC TCA GGT GAC TCT GA	58	329
				3´	TAT CTT GGC TAG TAG GTG TGC		
D5S592	GG	348	119766170-119765980	5´	GTC AAC AAA GTA ATG TAA AGA CAG	58	191
				3´	TGG AGT GGA GAG CGT CTC AG		
BCR	GG	142	23238278-23238415	5´	AGA ATG GAA GAG TTC CTG GC	61	138
				3´	ATG TTC CAG GTG TGG GCA G		
BCR	AT	790	23173512- 23174076	5´	TGC AGG GGT ATG ATC ATA GC	64	583
				3´	CTT CTC ATG CCT CAA AGC C		
BCR	AT	701	23224613- 23224958	5´	AGC TCA CGC CTG TAA TCC C	66	346
				3´	ATA AGA ACA CTT GCT CCA GCC		
BCR	AT	600	23178127-23178700	5´	CCA TTG AGG GCT GGG TGC A	64	358
				3´	GAG AAC ATT CCT GCC CTG GA		
D5S299	AT	550	102311503-102311661	5´	GCT ATT CTC TCA GGA TCT TG	58	159
				3´	GTA AGC CAG GAC AAG ATG ACA G		
UBP1	TA	394	33393184- 33393363	5´	TCA GGT TTA CAG ACA GGG TC	55	180
				3´	CCC AAC TCT CTG AAA ATG ATT C		
SRP19	AT	350	112877927-112878051	5´	ACT CAC TCT AGT GAT AAA TCG GG	60	110
				3´	AGC AGA TAA GAC AGT ATT ACT AGT T		
D5S816	AT	307	135965886-135966063	5´	TTG CCA CTG AAA ATC ATA TCC	58	178
				3´	CAG TGT CCC AGA CTC AGA C		
D16S3040	AT	302	79617710-79617849	5´	CTG CAA CAA GAA AGA TAC TCC	59	140
				3´	AGT GCC TCA CAG GCT GCC		
APIG1	AT	220	71734412-71734630	5´	TAA AGT TCT GAA CCC TCA GAA G	56	219

				3´	TAA GAT GCC ACT AAG CCA CCT		
ZKSCAN7	CA	648	44555220-44555476	5´	CTA ACG GGC AGT AGA GTG TC	61	257
				3´	CAC GTC TGC CCT GTT CAT CG		
CDH3	CA	637	68679828-68680048	5´	AGC TTC TGC TCT CAG AGT CA	57	221
				3´	CAT GAG GAT GAT GCC ATC TAC		
RH27788	AC	576	23166313-23166701	5´	CAT AAG GGT GTG AAG AAC TG	60	389
				3´	CTG TGA GAC TAG CCG TGC A		
D6S453	AC	479	141060084-14106022	5´	ACC TGG CTG AGG TAG GAC T	58	140
				3´	ACT ATG GCT ACT AAT TGT GAC T		
AGTR1	AC	279	148742490-148742594	5´	CAG GAG ATG AGA GTT CCA GA	60	105
				3´	GTA CCA GGT GCA AGT GTA GC		
D6S1633	AG	1317	156653724-156653993	5´	CTC ATG GAG CTT ATA GCC TG	59	270
				3´	TGT TCC TTC TGG CTA GCA TG		
ULK4	AG	682	41455584-41455709	5´	AGA CTT CTG GGC CTG GAA C	59	126
				3´	GAG GTA GAA TGC TTC CTC AG		
CDH1	AG	486	68823418-68823596	5´	GAA GCC AAA GAT GGC CTT AG	55	179
				3´	CCA GAA TGG CAG GAA TTT GC		
ZFHX3	AG	480	72793791-72793934	5´	CAT GAC TTG AGG ACC TTC AG	56	144
				3´	AAC CTC AAG CCT TGC AGA TC		
D5S399	AG	485	135991437 - 135991552	5´	ACG GAA CTT CAA TTG ACC TC	58	178
				3´	GCA GGC TGT GGC CTC AAC		
RTP3	AG	394	46499543-46499844	5´	GTG TCC AGG GTT CTC CTA G	58	302
				3´	GCA GCC ATG TGC TTC CTG C		
ANO10	AG	370	43549502-43549691	5´	CCT GTT TGC ATT GTG GTG AC	55	190
				3´	CAG TAT CCA TAT TAC GTT TCT C		
D17S800	AG	333	40900195-40900347	5´	CTT ATG GTC TCA TCC ATC AGG	61	153
				3´	GAC AGA AAG ATG GAT AAG ACA AG		
TP53	TC	1032	7674852-7675312	5´	TGC TGC CGT GTT CCA GTT GC	64	461
				3´	ACC AGA CCT CAG GCG GCT CA		
ENTPD3	CT	651	40427301-40427590	5´	TGG CTA CAT GCT CAG CCT GA	61	290
				3´	GGC AGC CAC TTC ACC TGC AT		
ZFHX3	TC	576	72798142-72798309	5´	CTG ACA TCC AAC AGC TTT ATG G	56	170
				3´	GTA CTG ACC CAG AGT CAC TG		
CDH3	CT	552	68684785-68684903	5´	CAG AAC AAA CGT TGG CCA TG	59	119

				3´	ACA CTG CAG AAG ACC CTG AC		
D5S1360	TC	513	145518073-145518211	5´	ACA AAC AAA ACC AAG AGT GC	55	139
				3´	TGG CTC ATG TAT CCC TAT GT		
D6S308	CT	262	140935578-140935791	5´	GCC TTG GAG AGA AAT TCA CGT A	61	214
				3´	GGC CTC ATA GTC TAA TCA CTG		
OPN1SW	TC	283	128773490-128773566	5´	GAC TCT ACC CAG GTT TCT AG	60	76
				3´	CTG TGC TTA CCA AAG GCT TC		
BCR	GT	1351	23240763-23241440	5´	CTG TCA CTA CTG TCC ATC TC	62	679
				3´	GCA AAG GTG CTC AGC CAG		
BCR	GT	1068	23270842-23271141	5´	ACA GCC TAG GCT ATT CAC TT	58	300
				3´	ACA AAA TCA CGC CTG GCT G		
BCR	GT	881	23253602-23253981	5´	GAG TGA GCT CAT GTT CAT CC	62	379
				3´	CCG CTT TAG TGG ACT CCA G		
BCR	GT	756	23211851-23212218	5´	GTG GAG CTG TTC TCA CTC AG	62	368
				3´	CAT GTG GCT CGA TGG CTT CT		
BCR	TG	670	23211862-23212054	5´	TGC CAG TTC AGA CCC TTA TG	58	192
				3´	CTC ACT CAG TTT GTC CTC AG		
D5S615	TG	641	125827596-125827822	5´	GGT AAA CCC TCA AGC AGT C	57	227
				3´	AAC CAG TTT CTT ATT ATA AGC C		
APC	TG	618	112840411-112840747	5´	CAA GGA AAC CAA GTC AGC TGC	57	330
				3´	GCT TTA TTG TCA TCC AAT TCA G		
MYRIP	TG	600	40162571-40162744	5´	GCA TTA TTG TGG GAC CTC ATG	57	174
				3´	CAA AAA GGC TTC CTC CTG GCA		
FAM117A	TG	260	49763020-49763196	5´	CAC GTT GGC TGC AAT GCC C	60	176
				3´	CTG TCT CAT TCC AAC CAT CG		
CCK	GT	231	42258050-42258159	5´	ACT ACA TGG GCT GGA TGG AT	60	109
				3´	TCT GGG TTG GGA GGT TGC T		
ZNF19	GT	216	71474062-71474192	5´	GAA TTG GGT AGG GAG ATT CCA	56	131
				3´	ACA CTG CAG AGA TCT CTC AG		
BCR	GC	1071	23223449-23223707	5´	TTG TAG ATG CGG ATG CTG ACT CC	60	258
				3´	TCA CCT GGA CCC TTG CCA ACT		
ZNF23	GC	417	71461493-71461783	5´	CCA AGT ACT GTA CTT TAG GTA G	55	290
				3´	GCT CCT GAC ATA GGG TCA TG		
TRAK1	GC	355	42209657-42209884	5´	TCG TCC AGG TCA ACA TCC AG	56	228

				3´	AGG GGT TCA CGC TAA GCA CT		
ULK4	GC	298	41566092-41566235	5´	ACT TTG CAT GGT ATT ACC CAG	55	145
				3´	ACC TTG TGC AAA GAA TGA		
HERBB2	GC	226	39698666-39698798	5´	GCT CAG CAC ATG GAA GCA AG	58	133
				3´	GGA TCA AAG GCA CCT ATC AG		
D5S471	GC	188	119713478-119713583	5´	GTT TTC ACA CAT TTT CCC AGC	60	106
				3´	GTT ACA ACA AAT AGC AAC AGC		
BCR/ABL	GG	201	t(9;22)	5´III	AGG AAG GAC TCA TCG GGC AG	58	122
				3´	TCT GTT TGG GTA TGG AAG CTG		

*: Differentiating base (genomic nucleotide immediately following the MseI restriction site); ** MseI-Fragment length based on human reference genome (hg19); ***: Annealing-Temperature.

Table 10: Oligonucleotides for Ampli1 WGA quality control

Gene/Marker	**Fragment** [bp]**	**Primer**	**Sequence 5´->3´**	**T*** [°C]**	**PCR-Product [bp]**
D5S2117	1376	5´	CCA GGT GAG AAC CTA GTC AG	58	147
		3´	ACT GAG TCC TCC AAC CAT GG		
TP53	1374	5´	GAA GCG TCT CAT GCT GGA TC	58	301
		3´	CAG CCC AAC CCT TGT CCT TA		
CK19	1146	5’	GAA GAT CCG CGA CTG GTA C	58	614
		3´	TTC ATG CTC AGC TGT GAC TG		
KRAS	192	5’	ATA AGG CCT GCT GAA AAT GAC	58	91
		3’	CTG AAT TAG CTG TAT CGT CAA GG		

** MseI-Fragment length based on human reference genome (hg19); ***: Annealing-Temperature.

2.2 Methods

2.2.1 Sample preparation

2.2.1.1 Cell culture: K562 leukaemia cell line

1-2x10^6 K562 cells were isolated from a suspension culture in RPMI 1640 medium (Gibco, Carlsbad, CA) supplemented with 10% Fetal Bovine Sera (FBS) (Pan-Biotech GmbH, Aidenbach, DE) and 200g/ml of penicillin and streptomycin (PAA Laboratories GmbH, Pasching, Austria). The cells were collected and centrifuged at 4°C for 5 minutes and 500g; the pellet was washed once with 1X Phosphate Buffered Saline Solution (PBS), centrifuged again and resuspended in 1-2 ml of 1X PBS.

2.2.1.2 Isolation of peripheral blood lymphocytes (PBL)

Peripheral blood was collected from a group of healthy donors, directly transferred to a 50ml Falcon tube and filled up to a final volume of 50ml with Hanks Buffered Salt Solution (HBSS) (Biochrom AG, Berlin, DE). Blood samples were centrifuged at 4°C for 10 minutes at 170-200xg, the supernatant was removed and cells washed again with HBSS followed by an additional centrifugation step at 4°C for 10 minutes at 170-200xg. The supernatant is removed and the blood filled up to 9ml with HBSS. The solution is poured slowly on 6ml of Percoll 60% (GE Healthcare, Buckinghamshire, UK) in a 15 ml Falcon tube. The sample is centrifuged at 4°C for 20 minutes at 1000xg, the interphase collected, transferred to a new 50ml falcon tube and filled up with 1X PBS to a final volume of 50 ml. The sample is centrifuged again at 4°C for 10 minutes at 500xg, the supernatant removed and cells resuspended in 1-2 ml of 1X PBS.

2.2.1.3 Isolation of pools and single cells

Ten single lymphocyte cells (Drivers), a pool of five K562 cells (positive controls), and nine K562 single cells (Testers) were isolated by micromanipulation at an inverse microscope. Cells were plated at a density of 250.000 cells/0.8 cm^2 in a volume of 200µl on a 8 well chamber slide (NUNC-Thermo, Rochester, NY). Single cells were aspirated into a glass pipette of 30µm diameter containing FBS and transferred to a new slide. After confirming that only a single cell had been transferred, this cell was finally picked in 1µl of 1X PBS into the PCR reaction tube.

2.3 Whole Genome Amplification (WGA) of single cells and cell pools

1. The single cells and cell pools, contained in 1µl of 1X PBS, were subjected to proteinase K digestion by adding 2µl of the following mix:

Proteinase K digestion mix (10 reactions)	[µl]
OPA+ Buffer (10X)	5
Tween 10%	1,3
Igepal 10%	1,3
Proteinase K (10mg/ml)	2,6
H_2O	12,7
Incubation: 10 h at 42°C Inactivation: 10 min at 80°C	

2. Restriction digestion with endonuclease MseI (New England Biolabs-NEB, Ipswich, MA) was performed by direct addition of 0,5µl of the following MseI digestion mix to the Proteinase K mixture of Step 1.

MseI digestion mix (1 reaction)	[µl]
MseI (50U/µl)	0,25
H_2O	0,25
Incubation: 3 h at 37°C Inactivation: 20 min at 80°C	

3. The Lib1-adaptor is formed by pre-annealing of Lib1 primer and ddMse11:

Pre-annealing Reaction mix (1 reaction)	[µl]
OPA+ Buffer (10X)	0,5
Lib1 primer (100 µM)	0,5
ddMse11 (100 µM)	0,5
H_2O	1,5

Annealing is started at temperature of 65°C and the shifted down to 15°C with a ramp of 1°C/minute.

4. At 15°C, 1µl of ATP 10mM (Roche) and 1µl T4 DNA Ligase (Roche) are added to the pre-annealing reaction:

Ligation Reaction mix (1 reaction)	[µl]
Pre-annealing reaction	3
ATP (10mM)	1
T4 DNA Ligase (≥ 5U/µl)	1

5. Five µl of the ligation reaction mix is added to each MseI-digested sample and ligation carried out over night at 15°C.

6. For PCR amplification of the ligated sample, the following WGA amplification mix was prepared:

WGA amplification mix (1 reaction)	[µl]
PCR Buffer 1 (10X)	3
dNTPs (10mM)	2
DNA Polymerase Mix (5U/µl)	1
H_2O	35

7. Forty µl of the WGA amplification mix is added to each sample and the reaction directly transferred to a thermocycler:

WGA Protocol						
Temperature	1 Cycle	14 Cycles	8 Cycles	**Temperature**	22 Cycles	1 Cycle
94°C		40 sec	40 sec	**94°C**	40 sec	—
57°C		30 sec	30 sec + 1°C/Cycle	**65°C**	30 sec	—
68°C	3 min	1 min 30 sec + 1 sec/Cycle	1 min 45 sec + 1 sec/Cycle	**68°C**	1 min 53 sec + 1 sec/Cycle	3 min 40 sec

2.3.1 Quality control of the WGA library

2.3.1.1 Ampli1 WGA quality control PCR

The quality of each WGA library is assessed by performing a multiplex control PCR assay detecting the presence of four marker genes (Table 10: Oligonucleotides for Ampli1 WGA quality control). Only WGA libraries positive for at least 3 markers are subjected to further processing.

Ampli1 WGA Control PCR mix	[µl]
PCR Buffer (10X), Tube 2 (Green)	1
Primer mix (4 primer pairs) (3 µM each)*	1
dNTPs (10 mM)	0,2
BSA (20mg/ml)	0,2
FastStart Taq Polymerase (5U/µl)	0,1
WGA Template	1
H_2O	6,5

*primers mix: 6µl of each primer (100µM) in a final volume of 200µl

Temperature	1 Cycle	17 Cycles	19 Cycles	1 Cycle
95°C	4 min	30 sec	30 sec	30 sec
58°C	-	30 sec	20 sec	30 sec
72°C	-	20 sec	20 sec	2 min

PCR products are separated on a 1,5% agarose gel for 45 minutes at 160V in 1X TBE.

2.3.1.2 Assessment of the WGA MseI-fragment length profile

One µl of 1:10 dilution of each amplified WGA library was assessed for the length distribution of the MseI-fragment on an Agilent DNA High Sensitivity Kit (Agilent Technologies, Santa Clara, CA) according to the instructions of the manufacturer.

2.3.1.3 Specific MseI-fragment control PCR

A control PCR with different primer pairs amplifying markers specifically residing in discrete MseI-fragment (Table 9: Oligonucleotides for specific MseI-fragment control PCR) is performed to check the presence of ten MseI-fragment populations, each one representing the ten possible nucleotide combinations presents in the original WGA library.

Temperature	1 Cycle	17 Cycles	19 Cycles	1 Cycle
94°C	2 min	30 sec	30 sec	30 sec
55 − 64°C*	30 sec	30 sec	20 sec	30 sec
72°C	2 min	20 sec	20 sec	2 min

*: T_{anneal} of the primers pair

Specific MseI-fragment control PCR Mix	[µl]
PCR-Buffer (10X) with dNTPs (each1 mM)	1
5´-Primer 8 µM	0,5
3´-Primer 8 µM	0,5
BSA (20mg/ml)	0,25
Taq DNA Polymerase 5 U/µl	0,15
H_2O	7,25
DNA sample 1:10	0,5

The PCR products were separated for 45 minutes at 160V on a 1,5% agarose gel in 1X TBE.

2.3.1.4 Double strand DNA quantification of the WGA library

The double strand DNA (dsDNA) concentration of the WGA was measured with Qubit dsDNA Reagents and Kit – high sensitivity assay (Invitrogen, Carlsbad, CA) following the instructions of the provider.

2.3.1.5 Development of an absolute qPCR assay to determine the enrichment/loss of the BCR/ABL breakpoint-MseI-fragment

A calibration curve was designed and assayed on the single cell K562 WGA libraries in order to follow the trend of the BCR-ABL breakpoint-MseI-fragment during all the critical steps of the WGA library modification, genome fractioning and subtraction. The BCR/ABL breakpoint-MseI-fragment was amplified from the cell line K562 with the following primers: BCR/ABL forward 5' CTGGATTTAAGCAGAGTTCA 3' and BCR/ABL reverse 5' GAAAATCCTTAAGGGTATTTCTG 3'. The amplified BCR/ABL breakpoint-MseI-fragment was cloned in pGEM-T Easy plasmids according to the manufacturer instructions (Promega, Fitchburg, WI). NEB10 beta competent cells (New England Biolabs) were transformed and grown according to the instructions of the manufacturer. The plasmid isolation was performed using Qiagen Midi Kit (QIAGEN, Hilden, Germany). The isolated plasmid DNA was quantified using Nanodrop (Thermo Scientific) and linearized with 100U of the restriction enzyme XmnI (New England Biolabs), purified using QIAquick PCR Purification Kit (Qiagen) and quantified with Qubit dsDNA Reagents and Kits (Invitrogen). A 10-fold serial dilution of the plasmid standard DNA was prepared in molecular biology pure water, containing 100-1.000.000.000 copies per 5 µl. The following formulas were used to calculate the copy numbers for the standard curve:

1. Calculate the mass of a single plasmid molecule:

   $m = (n) * (1.096*e\text{-}^{21}\ g/bp)$

   where n is the plamid sizes (bp), m is the mass, $e\text{-}21=*10^{-21}$

2. Calculate the mass of plasmid containing the copy numbers of interest:

   Mass of plasmid DNA needed = copy number of interest * mass of single plasmid

A triplicate of each dilution was amplified with primers K562 5' and 3' using LightCycler 480 SYBR Green I Master 2X (Roche Diagnostic GmbH) according to the manufacturer. Transformations were performed using a calibration curve, because Ct values are proportionally related to the $\log^2$ of the copy number of marker.

Absolute qPCR mix	[µl]
Syber Green I Master mix (2X)	10
5' Primer (10 µM)	1
3' Primer (10 µM)	1
H_2O	3
DNA Template	5ng

qPCR	**Activation**		**Amplification**		**Melting Curve**		**Cooling**
Temperature	1 Cycle	**Temperature**	40 Cycle	**Temperature**	1 Cycle	**Temperature**	1 Cycle
95°C	5 min	**95°C**	10 sec	**95°C**	5 sec	—	
—	30 sec	**60°C**	20 sec	**65°C**	1 min	—	
—	—	**72°C**	20 sec	**95°C**	—	**40°C**	10 sec

2.4 WGA library modification for genome fractioning

2.4.1 Re-amplification of primary WGA

One µl of primary WGA generated from pooled lymphocytes and single K562 cells, respectively, were used as starting material for a limited number of re-amplification cycles. Eight reactions per sample were prepared in order to minimize reaction bias and generate sufficient amounts of high-quality WGA samples for the fractioning process.

Re-amplification mix	[µl]
PCR-Buffer 1 (10X)	5
Lib1 Primer (10 µM)	5
dNTPs (10 mM)	2
BSA (20mg/ml)	1,25
DNA Polymerase Mix (5U/µl)	1
Template	1
H_2O	35,0

Re-amplification protocol		
Temperature	1 Cycle	9 Cycles
94°C	1 min	30 sec
60°C	30 sec	30 sec
68°C	2 min	2 min + 20 sec/Cycle

All reactions were combined and the PCR product purified using Agencourt AmpureXP (Beckman Coulter, Brea, CA) according to the manufacturer. DsDNA concentration was measured with Qubit dsDNA Reagents and Kits-Broad range assay (Invitrogen, Carlsbad, CA) following the instruction of the provider.

All necessary quality controls have been applied as described in paragraph 2.3.1.2 – 2.3.1.5.

2.4.2 Removal of the Lib1-adaptor

The re-amplified PCR product was digested with the enzyme MseI (NEB) to release the Lib1-adaptor:

Lib1-adaptor digestion mix	[µl]
OPA+ Buffer (10X)	5
MseI (50U/µl)	1
5µg of re-amplified product	X
Add H_2O to final volume	50
Incubation: 3h at 37°C Inactivation: 20min at 65°C	

The reaction was purified to remove the Lib1-adaptor and dsDNA concentration measured as described in paragraph 2.4.1

All necessary quality controls have been applied as described in paragraph 2.3.1.2 – 2.3.1.5.

2.4.3 Preparation, Ligation of J24 (A,C,G,T) adaptors and J20 secondary PCR

After removal of the primary Lib1-adaptor, different restriction recognition sites for six-cutter enzymes were implemented in the MseI-fragments by the ligation of secondary J24 (A,C,G,T) adaptors. These adaptors differ in the single nucleotide preceding the reformed MseI recognition site. J24-adaptors are created by combining the corresponding J24 (A,C,G,T) primers and J13 (A,C,G,T) oligonucleotides in equimolar ratios:

J24 (A,C,G,T) adaptor annealing mix	[µl]
OPA+ Buffer (10X)	10
J24 (A,C,G,T) (100µM)	45
J13 (A,C,G,T) (100µM)	45

Adaptor annealing is carried out by incubation at 95°C for one minute followed by cooling to 15°C at a rate of 1°C per minute. The adaptors J24A, J24T, J24C and J24G were mixed prior to ligation in a 1:1:1:1 ratio to ensure random ligation to MseI-fragments.

J24 (A,C,G,T) ligation mix	[µl]
Quick ligation buffer 2X	20
J24 (A,C,G,T) Mix (45 µM)	8
Quick Ligase (50U/µl)	1
MseI digested DNA 400ng	X
Add H_2O to final volume	40

The ligation reaction was carried out for 15 minutes at room temperature. Eight parallel reactions of the ligated DNA product were amplified to generate sufficient dsDNA amounts for the subsequent steps. The amplification performed with the J20 primer facilitates the uniform amplification of the J24 (A,C,G,T) adaptor-ligated MseI-fragments.

J20n-PCR mix	[µl]
PCR-Buffer 1 (10X)	5
J20n Primer (10 µM)	5
dNTPs (10 mM)	2
BSA (20mg/ml)	1,25
DNA Polymerase Mix (5U/µl)	1
MseI-digested DNA	5
H_2O	30,75

J20n-PCR protocol				
Temperature	1 Cycle	8 Cycles	5 Cycles	1 Cycle
94°C	—	40 sec	40 sec	—
57°C	—	30 sec	30 sec (+ 1°C/Cycle)	—
68°C	3 min	1 min 30 sec+ 1 sec/Cycle	1 min 45 sec + 1 sec/Cycle	3 min 40 sec

The amplified J24 (A,C,G,T) adaptor-ligated MseI-fragments were combined, purified and dsDNA concentration measured as described in paragraph 2.4.1.

All necessary quality controls have been applied as described in paragraph 2.3.1.2 – 2.3.1.5 on the J20 PCR secondary amplification.

2.4.4 Preparation of the specific fractioning-adaptors

Depending on the six-cutter restriction site intended to reform at the end of the original MseI-fragment, site-specific fractioning-adaptors (26R A,C,G,T and Bio26R A,C,G,T) are selected according to Table 11. Depending on their use in the protocol to select original MseI-fragments harbouring different or identical reformed six cutter restriction sites at their opposite ends, 26R (A,C,G,T) adaptors are applied either in a biotinylated (26R A,C,G,T) or non-biotinylated form (Bio26R A,C,G,T).

Table 11: Adaptors selection according to the restriction site reformed at the end of the MseI-fragment

Enzyme	Recognition sequence	Fractioning-adaptor	Sequence 5´→3´
DraI	TTT´AAA	Bio26RA	AAT TCG TCA ATC AGC TGC GCT TGT TT
VspI	AT´TAAT	Bio26RT	GTA TAG TCT CGG CTC AAC GTC AGG AT
KspAI	GTT´AAC	Bio26RC	CTA TCG CAC TCG ACT ATC CTC TGG TT
BspTI	C´TTAAG	Bio26RG	TCT GAG CAA ATG ACC ATG CAG CAC C
DraI; VspI		26RA; Bio26RT	
DraI; KspAI		26RA; Bio26RC	
DraI; BspTI		26RA; Bio26RG	
VspI; KspAI		26RT; Bio26RC	
VspI; BspTI		26RT; Bio26RG	
BspTI; KspAI		26RG; Bio26RC	

These third type of adaptors are created by pre-annealing two long oligonucleotides, 26R (A,C,G,T) and Bio26R (A,C,G,T), respectively, in a 1:1 ratio to the short oligonucleotide R13 (A,C,G,T).

Bio26RN and 26RN primer-adaptor annealing mix	[µl]
OPA+ Buffer (10X)	10
Bio-26R (A,C,G,T) or 26R (A,C,G,T) (100µM)	45
R13 (A,C,G,T) (100µM)	45

Both oligonucleotides are incubated at 95°C for one minute and cooled down to 15°C at a rate of 1°C per minute.

2.4.5 Coupling of the Bio26R (A,C,G,T) fractioning-adaptors to magnetic streptavidin conjugated beads

Five µl of Nanolink magnetic streptavidin-conjugated beads (Solulink, Inc., San Diego, CA) are washed two times with 200µl of Nucleic Acid Binding and Wash Buffer (BW), resuspended in 200µl of BW buffer in low binding PCR tubes (Axygen). Ten µl of Bio26R (A,C,G,T) adaptors are added to the beads and roll-incubated for 15 minutes at room temperature. After incubation the streptavidin-conjugated fractioning-adaptor complex was

washed two times with 200µl of BW buffer and resuspended in 200µl of BW buffer. By this way a streptavidin conjugated-adaptor complex is pre-formed and ready for the use in the next step of the protocol.

2.5 Genome Fractioning of the modified WGA-libraries

2.5.1 Restriction digestion of the amplified J24 adaptor-ligated MseI fragments

After secondary amplification with a J20n primer, amplified J24 adaptor-ligated MseI-fragments were digested with six-cutter enzymes DraI (**T***TTAA*A), VspI (**A***TTAA*T), KspAI (**G***TTAA*C) and BspTI (**C***TTAA*G) alone or in combination of two different enzymes. (Thermo Fisher Scientific, Inc.). The digestion is enabled by the incorporation of J24 (A,C,G,T) adaptor (in bold character the single nucleotide preceding the re-formed MseI recognition site), which, in conjunction with the genomic nucleotide immediately following the MseI restriction site (in underlined character the differentiating base) specifies the restriction sites of the selected enzymes. Amplified J24 adaptor-ligated MseI-fragments are digested separately by each of the four selected restriction enzymes:

1st Specific restriction digestion mix	[µl]
Fast Digestion Buffer (10X)	10
Fast Digestion Enzyme	10
J24 adaptor-coupled MseI fragments 10µg	X
Add H_2O to final volume	100
Incubation: 15min at 37°C Inactivation: 20min at 65°C	

After restriction, fragments are purified and dsDNA concentration measured as described in paragraph 2.4.1.

All necessary quality controls have been applied as described in paragraph 2.3.1.2 – 2.3.1.5.

2.5.2 Selection of the MseI-fragment fractions depending on the reformed six-cutter restriction sites

The basic protocol to select MseI-fragment fractions carrying identical (simple fractions) or different (mixed fractions) six-cutter restriction sites at their opposite ends is highly similar except that for the mixed fractions an additional ligation/restriction cycle has to be included. By these means the different ends of a mixed fraction are coupled with different site-specific adaptors.

2.5.2.1 Ligation of the non-biotinylated fractioning-adaptor to the digested DNA

Digested DNA, as generated in paragraph 2.5.1, is ligated with the non-biotinylated form of each of the four specific adaptors, according to the ends generated from the digesting enzyme.

26R (A,C,G,T) Adaptor Ligation mix	[µl]
OPA+ Buffer (10X)	10
ATP (10mM)	10
T4 DNA ligase (5U/µl)	10
26RNn	30
5 µg digested DNA	X
H_2O to a final volume	100

The ligation reaction is purified as described in paragraph 2.4.1 and eluted in 41µl of PCR grade water. The ligated adaptor is fully converted to a double-stranded form by a fill-in reaction incubated for 20 minutes at 72°C.

Fill-in reaction mix	[µl]
PCR-Buffer 1 (10X)	5
dNTPs (10mM)	2
BSA (20mg/ml)	1,25
Taq DNA Polymerase (5U/µl)	1
Ligated DNA sample	41

The fill-in reaction is purified as described in paragraph 2.4.1, eluted in 80µl of PCR grade water, and the dsDNA concentration measured as described in paragraph 2.4.1.

2.5.2.2 Second specific restriction digestion for the mixed fractions

The DNA generated in paragraph 2.5.2.1 was digested with a second specific restriction enzyme:

2nd Specific restriction digestion mix	[µl]
Specific Fast Digest Buffer (10X)	10
Specific Fast Digest Enzyme	10
DNA Sample ca. 5µg	50
H_2O to a final volume	100
Incubation: 15min at 37°C Inactivation: 20min at 65°C	

The digestion product was purified as described in paragraph 2.4.1, eluted in 50µl of water and the dsDNA concentration measured as described in paragraph 2.4.1.

All necessary quality controls have been applied as described in paragraph 2.3.1.2 – 2.3.1.5.

2.5.3 Ligation of the MseI-fragments to the streptavidin-conjugated fractioning-adaptor complex

After removal of the resuspension buffer from the streptavidin-conjugated fractioning-adaptor complex, either 600ng of DNA MseI-fragments for simple fractions or 1250ng for mixed fractions (generated, respectively, in section 2.5.1 and 2.5.2.2) were added and roll-incubated together in the presence of a ligase for 30 minutes at room temperature.

Bio26R (A,C,G,T) adaptor ligation mix	[µl]
Quick ligation buffer 2X	20
Quick Ligase (50U/µl)	1
Digested DNA: simple fractions 600ng, mixed fraction:1250ng	X
Streptavidin-conjugated fractioning-adaptor complex add H_2O to final volume	40

After the ligation reaction, the streptavidin-conjugated fractioning-adaptor complex ligated to the DNA were washed three times with 200µl of BW buffer and finally resuspended in 200µl of BW buffer.

2.5.4 First fractioning PCRs

2.5.4.1 Simple fractions

For simple fractions one biotinylated adapter and, on consequence, one amplification primer is needed. Before PCR initiation, the bead resuspension buffer is replaced by the PCR reagents and a fill-in reaction for the adaptors is performed for 5 minutes at 68°C. After incubation, the enrichment primer is added and the 1st genome fractioning PCR directly performed on the beads:

1st Fractioning PCR Simple Fractions (AA, TT, CC, GG) mix	[µl]
PCR-Buffer 1 (10X)	5
23R´1 (10 µM)	5 (after 5 min at 68°C)
23R´2 (10 µM)	—
dNTPs (10 mM)	2
BSA (20mg/ml)	1
DNA Polymerase Mix	1
Template on beads	—
H_2O	35,75

1st Fractioning PCR Simple Fractions protocol		
Temperature	14 Cycles	1 Cycle
94°C	40 sec	—
60°C	30 sec	—
68°C	1 min 40 sec+ 1 sec/Cycle	3 min 40 sec

2.5.4.2 Mixed Fractions

As the mixed fractions have been ligated to different specific fractioning-adaptors (one biotinylated adaptor and one non-biotinylated adaptor), two specific amplification primers are needed for the fractioning PCR. Before PCR initiation, the bead resuspension buffer is replaced by the PCR reagents and the 1st Fractioning PCR directly performed on the beads.

1st Fractioning PCR Mixed Fractions (AG, AT, AC, TG, TC, CG) mix	[µl]
PCR-Buffer 1 (10X)	5
23R´1 (10 µM)	5
23R´2 (10 µM)	5
dNTPs (10 mM)	2
BSA (20mg/ml)	1,25
DNA Polymerase Mix (5U/µl)	1
Template on beads	—
H_2O	35,75

1st Fractioning PCR Mixed Fractions Protocol		
Temperature	14 Cycles	1 Cycle
94°C	40 sec	—
60°C	30 sec	—
68°C	1 min 40 sec+ 1 sec/Cycle	3 min 40 sec

2.5.5 ssDNA nuclease digestion and 2nd Fractioning PCR

PCR reactions were transferred to a magnetic stand. After the liquid becomes clear, the supernatant is removed from the magnetic streptavidin-conjugate beads, transferred to a new microvessel and digested at 37°C for one hour with ExonucleaseI. By these means excess of primers and single stranded DNA is degraded.

ExonucleaseI digestion mix	[µl]
1st Fractioning PCR	50
ExoI Buffer 10X	6
ExonucleaseI	4

The enzyme is heat-inactivated at 80°C for 20 minutes. A 2nd fractioning PCR is performed for simple and mixed fractions in an identical manner.

2nd Fractioning PCR mix	[µl]
PCR-Buffer 1 (10X)	5
23R´1 (10 µM)	5
23R´2 (10 µM)	0-5
dNTPs (10 mM)	2,0 [µl]
BSA (20mg/ml)	1,25 [µl]
DNA Polymerase Mix	1 [µl]
Template 1st Fractioning	5,0 [µl]
H_2O	35,75 [µl]

2nd Fractioning PCR protocol				
Temperature	14 Cycles	8 Cycles	1 Cycle	1 Cycle
94°C	40 sec	40 sec	40 sec	—
60°C	30 sec	30 sec	30 sec	—
68°C	1 min 30 sec + 1 sec/Cycle	1 min 45 sec + 1 sec/Cycle	1 min 53 sec	3 min 40 sec

All necessary quality controls have been applied as described in paragraph 2.3.1.2 – 2.3.1.5.

2.6 Normalization of the genome MseI-fractions (from paragraph 2.5.5)

Driver and Tester DNAs are normalized with Trimmer-2 cDNA normalization Kit (Evrogen, Moscow, RUS). For normalization, 1200ng of DNA are used according to the instruction of the manufacturer, with minor modifications: the number of PCR cycles for amplification of the normalized DNA sample was shifted to 9-11-13-15 cycles for optimization of PCR parameters and primers and annealing temperature adapted to the WGA protocol.

All necessary quality controls have been applied as described in paragraph 2.3.1.2 – 2.3.1.5.

2.7 Subtractive Hybridization of the genome MseI-fractions

2.7.1 Re-amplification of the genome MseI-fractions

The normalized genome MseI-fractions were re-amplified to generate sufficient high-quality dsDNA amounts needed for subtractive hybridization, e.g. 48 reactions were prepared for the Driver (lymphocyte pool) and one reaction for the Tester (K562 single cell) using a starting amount of 5 ng of the normalized 2nd Fractioning PCR. The reagents and PCR program have been the same used in paragraph 2.5.5 with the exception that maximum 11 cycles have been performed.

Adaptors used for the fractioning, are removed by digestion directly after DNA amplification:

TruI digestion mix	[µl]
Buffer R 10X	50
TruI (50U/µl)	6
ExonucleaseI	4

The digestion reaction was incubated for 3 hours at 65°C.

The dsDNA product was purified and concentration of dsDNA measured as described in paragraph 2.4.1.

2.7.2 Ligation of a specific biotinylated subtractive-adaptor (BioR21)

Tester DNA is ligated to a specific biotinylated-adaptor, BioR21, for selective amplification after the first round of subtraction. The Driver will not be ligated to any adaptor.

BioR21 ligation mix	[µl]
OPA+ Buffer (10X)	10
BioR21 (100 µM)	20
ATP (10 mM)	10
T4 DNA Ligase (5 U/µl)	10
Tester Sample 3µg	X
H_2O	To 100

The ligation reaction is incubated for 2 hours at 16C°. After purification the dsDNA concentration is measured as described in paragraph 2.4.1.

2.7.3 Driver and Tester precipitation before the subtractive hybridization

For the 1st round of subtractive hybridization (S.H.) 500ng of Tester, ligated with BioR21 subtractive adaptors, is used. The adapter-ligated Tester and the Driver are mixed together, extracted with Phase Lock Gel (5 PRIME GmbH, Hilden, DE) in phenol-chloroform-isoamylalchol according to the manufacturer and precipitated with 10 µl of linear polyacrylamide carrier (LPA) 5mg/ml, salt (ammonium acetate, 3M, pH 5.2) for 1/10 of the volume, and 2,25% (v/v) of ethanol, overnight -20°C. For the sample recovery after precipitation, a centrifugation step of 45 minutes at 4°C and 14.000 rpm was performed; the pellet has been washed one time with ethanol 70% for 10 minutes at 22°C and 14.000 rpm.

2.7.4 First round of subtraction: generation of the depletion product 1 (DP1)

The pellet was resuspended in 6 µl of 1X hybridization buffer (50mM HEPES, 0.2mM EDTA), incubated 5 minutes at 50°C to fully resuspend the DNA, transferred into a low-binding PCR tube (Axygen) and covered with 35µl of Mineral oil (Sigma-Aldrich, St. Louis, MO). After a10 minute denaturation step at 95°C, 1 µl of sodium chloride (NaCl) 5M was added and subtractive hybridization was performed for 18-24 hours at 68°C.

2.7.5 Positive selection of the re-hybridized Tester/Tester sequences with magnetic streptavidin- conjugated beads

The reaction is diluted with 44 µl of 1X TE. Five µl of Nanolink Streptavidin Magnetic beads (Solulink Inc., San Diego, California) are washed two times with 200µl of BW buffer, resuspended in 200µl of BW buffer in low binding PCR tubes (Axygen), mixed with the sample and agitated for 30 minutes at 40C°. The particles were washed three times in 200µl BW buffer at room temperature.

2.7.6 First enrichment PCR and ssDNA nuclease digestion

The first enrichment PCR was started after removal of the washing buffer by direct addition of the PCR reagents to the bead-ligated DNA sample.

1st Enrichment PCR Mix	[µl]
PCR-Buffer 1 (10X)	5
R21 (10 µM)	5 (after 5 min at 68°C)
dNTPs (10 mM)	2
BSA (20mg/ml)	1,25
DNA Polymerase Mix (5U/µl)	1
Template on beads	—
H_2O	35,75

1st Enrichment PCR Protocol			
Temperature	1 Cycle	10 Cycles	1 Cycle
94°C	—	30 s	—
60°C	—	30 s	—
68°C	5 min	2 min	10 min

The supernatant of the reaction was isolated on a magnet, purified with Agencourt AmpureXP and digested with MBN (New England Biolabs) to remove single stranded DNA.

Mung Bean Nuclease Digestion	[µl]
MBN Buffer (10X)	4
Mung Bean Nuclease (10U/µl)	3
DNA Template	33

The digestion was performed for 30 minutes at 30°C, then inactivated by adding 160µl of Tris-HCl pH 8.9 and incubating at 98°C for further 5 minutes.

2.7.7 Second enrichment PCR

Enrichment PCR2 was performed to amplify the depletion product 1 (DP1). Eight reactions per sample were prepared:

2nd Enrichment PCR mix	[µl]
PCR-Buffer 1 (10X)	5
R21 (10 µM)	5
dNTPs (10 mM)	2
BSA (20mg/ml)	1,25
DNA Polymerase Mix (5U/µl)	1
Template	25
H_2O	10,75

2nd Enrichment PCR protocol				
Temperature	1 Cycle	Temperature	26 Cycle	1 Cycle
94°C	1 min	94°C	30 s	—
		60°C	30 s	—
		68°C	2 min	10 min

The product was purified with QIAquick PCR Purification Kit (Qiagen) and quantified with Qubit dsDNA Reagents and Kits (Invitrogen). Absolute qPCR quantification was performed to determine the copy number of BCR/ABL per nanogram of DNA as previously described in paragraph 2.3.1.5.

At least three rounds of subtractive hybridization were performed and DP1, DP2 and DP3 products separated on a 1.5% agarose gel for 45 minutes at 160V in 1X TBE.

Absolute qPCR quantification was performed to determine the copy number of BCR/ABL breakpoint-MseI-fragment per nanogram of DNA as previously described in paragraph 2.3.1.5.

2.8 Normalization of the first depletion product (DP1)

The first depletion product is normalized with Trimmer-2 cDNA normalization Kit as described in paragraph 2.6.

Absolute qPCR quantification was performed to determine the copy number of BCR/ABL breakpoint-MseI-fragment per nanogram of DNA as previously described in paragraph 2.3.1.5.

2.8.1 Second round of subtraction: generation of the depletion product (DP2)

The normalized DP1 is directly used for a second round of depletion and enrichment. The R21 adaptor is not removed but it does not carry a biotinylated residue, as previous amplifications were carried out with non-biotinylated R21 primer (enrichment PCR and normalization PCR). The second round of hybridization is performed with 500ng of normalized DP1 (Tester) and 40µg of Driver leaving the Driver/Tester ratio unaltered compared to the first subtraction round. The subtraction is carried as previously described in paragraphs 2.7.2-2.7.3-2.7.4. After the subtraction, the re-hybridized DNA is digested directly with Mung Bean Nuclease (NEB):

Mung Bean Nuclease Digestion	[µl]
MBN Buffer (10X)	4
Mung Bean Nuclease (10U/µl)	3
Re-hybridized DNA	6
H_2O	27

The enzyme is inactivated as described in the paragraph 2.7.6 and the digested DNA re-amplified with a single amplification step as described in paragraph 2.7.7.

The PCR product was purified and dsDNA quantified as described in paragraph 2.7.7.

Absolute qPCR quantification was performed to determine the copy number of BCR/ABL breakpoint-MseI-fragment per nanogram of DNA as previously described in paragraph 2.3.1.5.

2.8.2 Third round of subtraction: generation of the depletion product (DP3)

The third round of subtraction is performed as in paragraph 2.8.1 with the exception that the DP2 product is not normalized.

Absolute qPCR quantification was performed to determine the copy number of BCR/ABL breakpoint-MseI-fragment per nanogram of DNA as previously described in paragraph 2.3.1.5.

3. Theoretical considerations of method design

3.1 Design of a method for breakpoint identification in single tumour cells and definition of relevant quality controls

Systematic enrichment and isolation of molecular breakpoints in single tumour cells would allow breakpoint identification down to the nucleotide level without the need to sequence whole genomes of tumour cells. Such a methodological approach can only be built upon a deterministic whole genome amplification (WGA) method ensuring little or no patient-to-patient and cell-to-cell variation whilst it is not applicable to stochastic WGA approaches. Ampli1 WGA meets these requirements as it is based on the MseI-restriction of a single cell genome, ligation of adaptors and exponential amplification of the ligated MseI-fragments. To implement a breakpoint identification method on the MseI-fragment population of a single cell, a strategy to reduce the sequence complexity of the human genome whilst preventing the simultaneous loss of genetic information has to be developed. With that goal in mind a step termed “Genome Fractioning” aimed at reducing the complexity of the genome has been developed. As Ampli1 WGA relies on the MseI-restriction (recognition/restriction site TTAA) of the genome each created MseI-fragment will have defined nucleotides next to its genomic MseI restriction sites: such a nucleotide is defined as the “differentiating base”. As a consequence, there are ten different types of nucleotide combinations forming the ends of MseI-fragments: (a) either simple combinations if the same nucleotides occur next to the genomic MseI restriction site (A/A, C/C, G/G, T/T) or (b) mixed combinations if the genomic nucleotides next to the MseI restriction sites are different (A/C, A/G, A/T, C/G, C/T, G/T) (Figure 4).

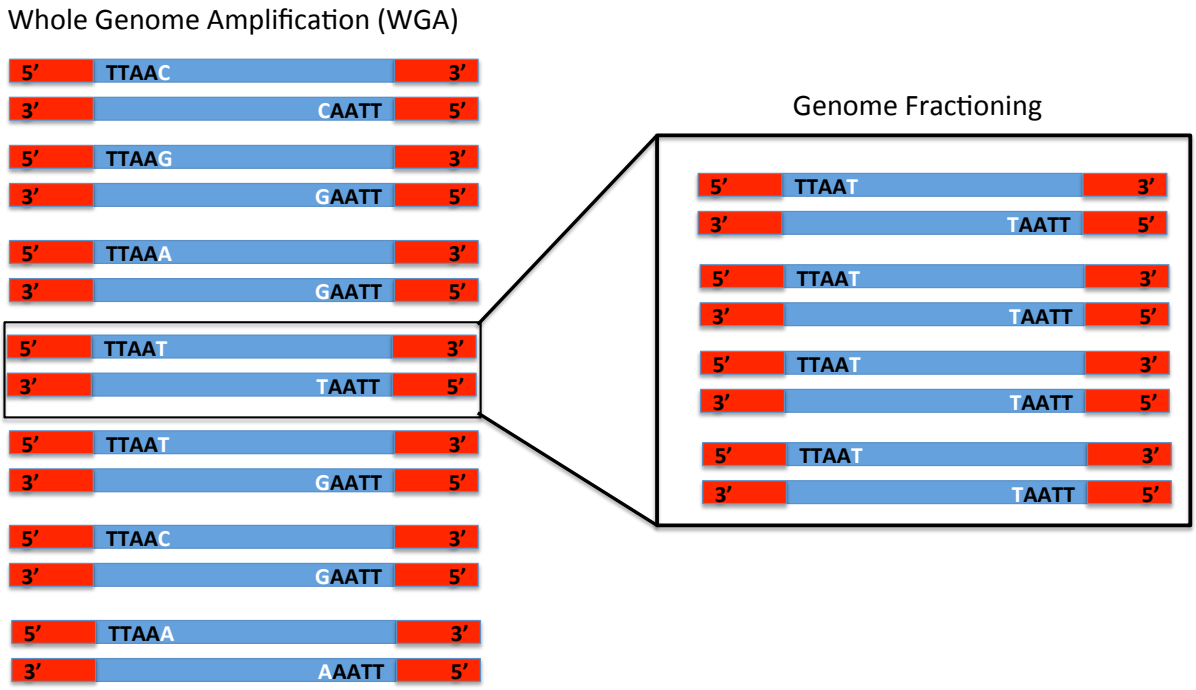


Figure 4: The WGA contains a myriad of MseI-fragments of different lengths with various combinations of differentiating bases. The final aim of the genome fractioning is the selective isolation of genomic MseI-fragments characterized by the identical differentiating bases. The final genomic fractioning contains just MseI-fragments of different base pair length with the same base after the MseI restriction site (in this case the nucleotide T, representing the T/T fraction). For simplicity not all the nucleotide combinations have been designed in the picture.

Theoretically, this leads to a 10-fold reduction of complexity of the original MseI-fragment population and consequently to the selective isolation of a distinct MseI-fragment population (MseI-fraction) characterized by their nucleotide combination of differentiating bases. The ten resulting discrete MseI-fractions of lower complexity can be subsequently modified and subjected to subtractive hybridization vs. germline single cell genomes, allowing for PCR-based amplification of breakpoint-carrying MseI-fragments (Figure 5).

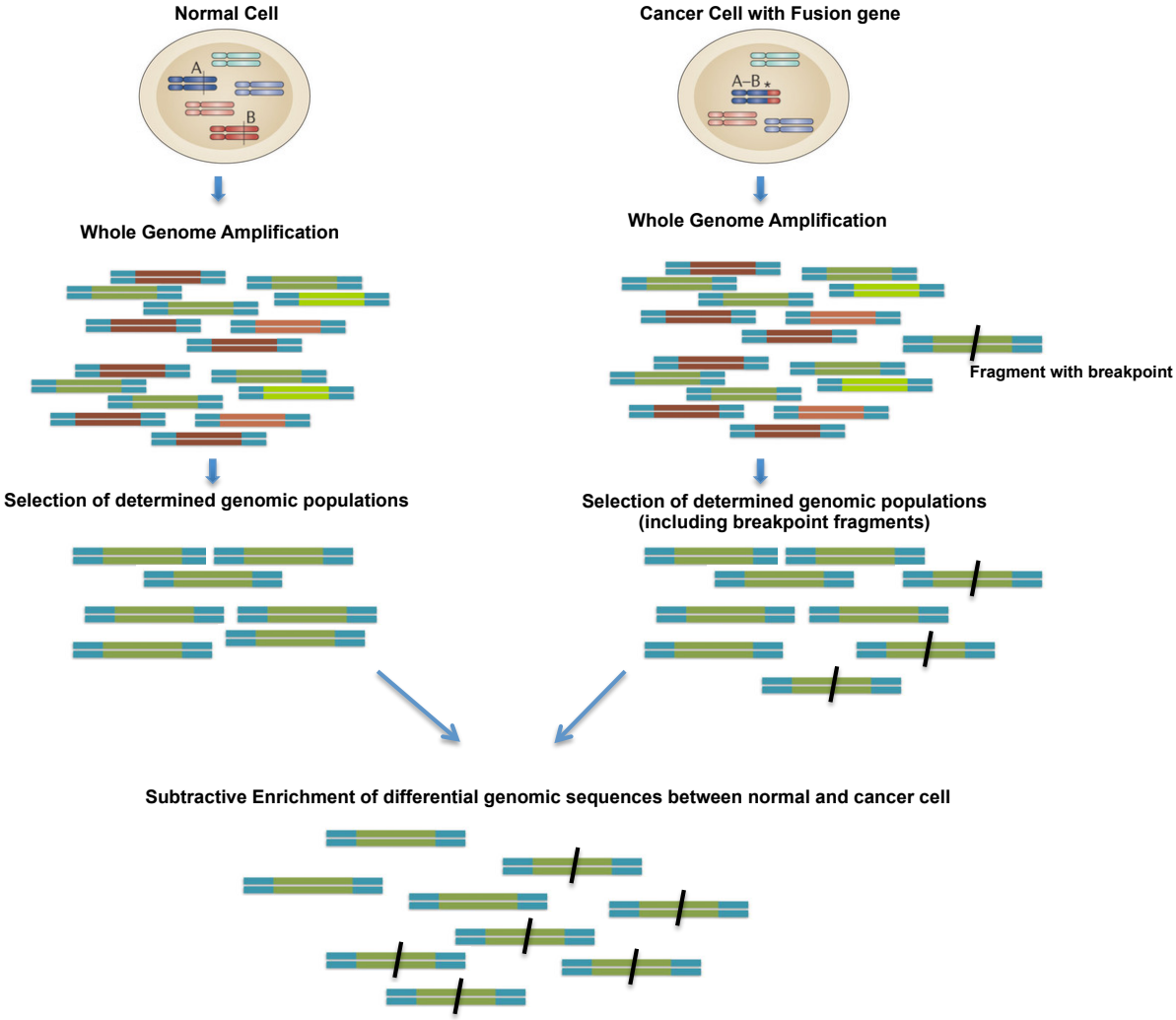


Figure 5: Overview of the method. The whole genome of a normal single cell and a tumor cell is amplified. Discrete genomic MseI-fragment populations are isolated, including genomic Mse-fragments containing a breakpoint sequence; the selection of discrete populations of the genome implies a reduction in complexity compared to the initial whole genome. The subtractive hybridization between the selected genomic population of the normal and cancer cell will enrich differential sequences, such as genomic breakpoints.

To monitor the successful implementation of the method, the leukemia cell line K562 harboring a well-characterized BCR/ABL fusion breakpoint contained in a specific MseI-fragment containing the G/G nucleotides was selected. High quality single cell K562 WGAs will help to evaluate the efficiency of each experimental step during the setting up of the breakpoint identification method by assessing the proper quality controls and by controlling the enrichment of the desired fusion sequence (Figure 6).

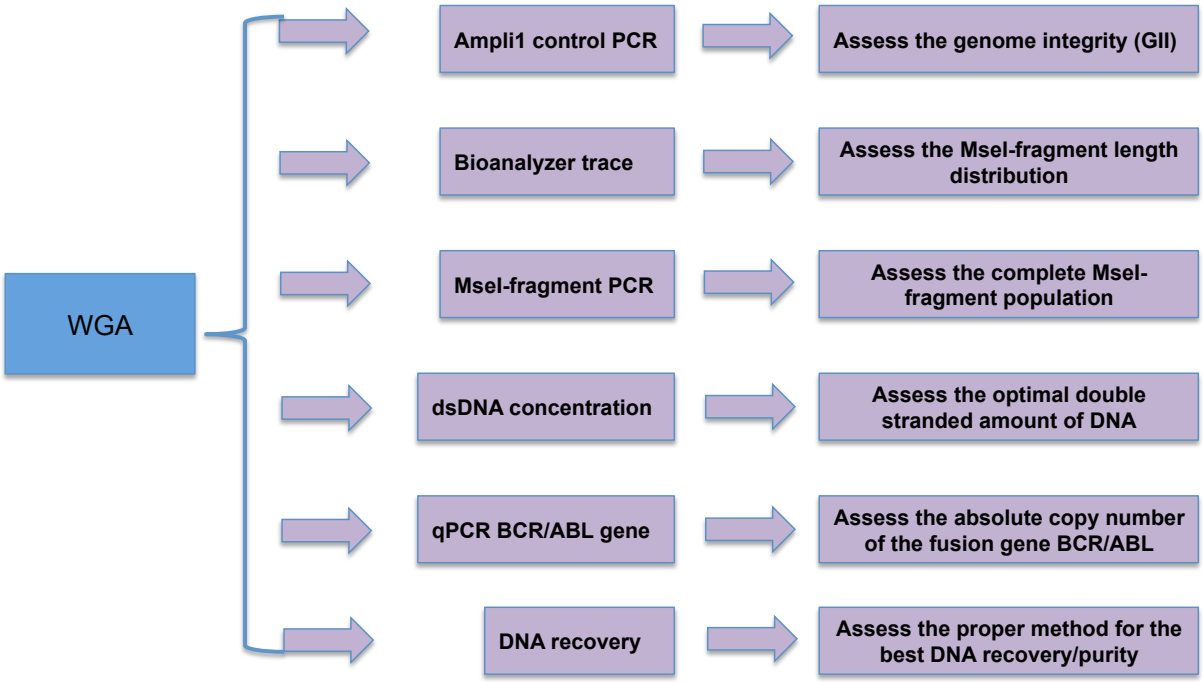


Figure 6: Quality control implemented for the original WGA library necessary for the proper assessment of the method.

3.1.1 Ampli1 quality control PCR to assess the quality of the WGA

The use of a high quality WGA is fundamental for the establishment of the method. Therefore the PCR-based Ampli1 quality control assay (Polzer et al., 2014), which enables the identification of high-quality amplified WGA samples, has been applied. Assay primers have been designed to amplify the four marker amplicons CK19, TP53 exon2/3, D5S2117, KRAS residing on MseI-fragments of different length classes. The WGA samples of highest quality samples can be identified by the presence of all four tested markers. Samples showing a lower number of positive markers were not used during method implementation.

3.1.2 Assessment of the MseI-fragments length distribution of high and low-quality WGA

As additional quality control the MseI-fragment length distribution of WGAs of diverse quality has been determined. By this means the MseI-fragment profile of high quality WGAs would be considered as standard benchmark for all the succeeding steps. In this way the changes of the MseI-fragment profile can be monitored during the development of the protocol. This control also enables to detect the unspecific enrichment of repetitive sequences as prominent PCR products at an early stage of method development. As the method consists of different steps of adaptor ligations and amplifications, it is not unlikely that MseI-

fragments representing highly repetitive sequences of the genome are selected and over-represented. This quality control also enables to ensure homogeneous MseI-fragment size distributions of the library samples.

3.1.3 Specific MseI-fragment control PCR to identify the ten discrete populations in the original WGA

High-quality WGA samples contain theoretically the complete population of the MseI-fragments composed of the ten different genomic fractions as defined by their differentiating base combinations. The concept of the differentiating base in the MseI-fragments allows the design of primer pairs amplifying markers specifically residing in discrete MseI-fragment. By systematically exploiting this concept, a total of 77 primer pairs have been designed in order to identify 77 different MseI-fragments (fragment length from 128bp to 1936bp), located in different genomic loci and representing the ten different fractions (Figure 7).

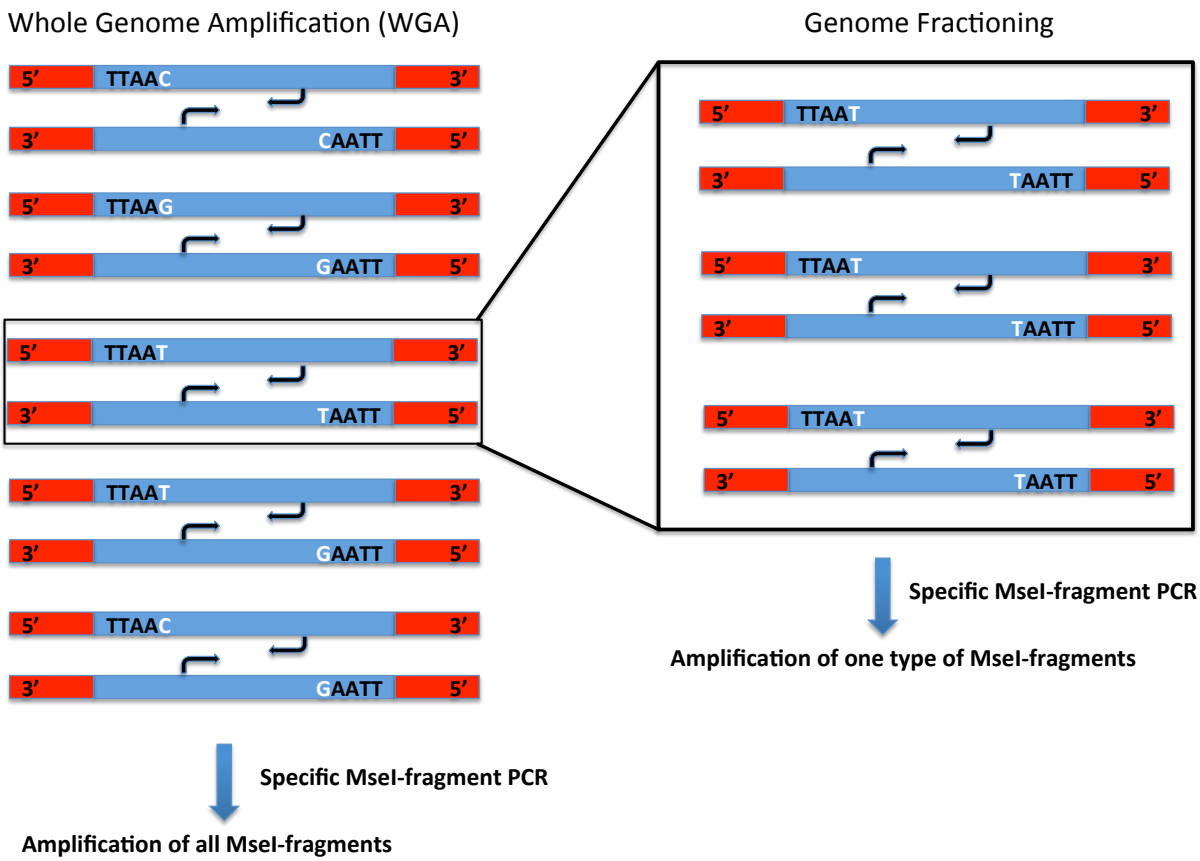


Figure 7: Specific MseI-fragment control PCR. Primer pairs (dark blue arrows) have been designed on the MseI-fragments in between the differentiating bases for each possible nucleotide combination. The specific MseI-fragment control PCR performed on WGA will give the amplification of all the combinations, while after the genomic fractioning we will have the amplification of one specific MseI-fragments population.

The established specific MseI-fragment control PCR allows controlling the reliable genome fractioning and isolation of the selected MseI-fragments during the development of the procedure for the genome complexity reduction. It is expected that all the control MseI-fragments representative for each of the ten nucleotide combinations to be preserved from the WGA up to the final selective isolation step; by these means accidental losses of the control MseI-fragments can be assessed at each step of the protocol. A genome fractioning is defined successful if the control amplicon contained within the MseI-fragments harboring the expected differentiating bases is amplified with the specific primers (Table 12-successful genome fractioning-pure). A genome fractioning is not successful if the control amplicon contained within the MseI-fragments harboring the expected differentiating bases is not amplified (Table 12-unsuccessful genome fractioning-negative), or if the isolated genome MseI-fraction will include amplified controls, indicating the presence of MseI-fragments harboring unexpected differentiating bases (Table 12-unsuccessful genome fractioning-unpure).

Table 12: Schematic representation of the specific MseI-fragments control PCR performed on the WGA and specifically isolated MseI-fragments population.

Specific primers for:	WGA	Successful Genome Fractioning (Pure)	Unsuccessful Genome Fractioning (Negative)	Unsuccessful Genome Fractioning (Unpure)
MseI-fragments A/A	+	-	-	-
MseI-fragments C/C	+	-	-	+
MseI-fragments G/G	+	-	-	-
MseI-fragments T/T	+	+	-	+
MseI-fragments A/C	+	-	-	-
MseI-fragments A/G	+	-	-	-
MseI-fragments A/T	+	-	-	+
MseI-fragments C/G	+	-	-	+
MseI-fragments C/T	+	-	-	-
MseI-fragments G/T	+	-	-	-

3.1.4 Precise assessment of the double strand DNA concentration of the WGA

As fully double strand DNA (dsDNA) enables proper functioning of the critical steps such as enzymatic digestions and adaptor ligations during method development, in the first place, an ultrasensitive method based on the use of a fluorescent intercalating dye specific for dsDNA was used. Using this assay the initial dsDNA amount of the WGA could be assessed and the optimal dsDNA amount for consecutive steps of the protocol determined.

3.1.5 Absolute qPCR quantification of the MseI-fragment carrying the BCR/ABL breakpoint in WGA amplified K562 single cells

To establish an assay for absolute quantification of the MseI-fragment carrying the BCR/ABL fusion gene breakpoint (BCR/ABL breakpoint-MseI-fragment), the initial copy numbers of the BCR/ABL translocation in amplified single K562 cells was assessed. This qPCR assay allows monitoring the degree of enrichment of the BCR/ABL breakpoint-MseI-fragment during each step of method development thereby optimizing and defining critical checkpoints for the complete process.

3.1.6 Choice of the purification system to define optimal parameters for the DNA recovery

The fidelity of the method used for the isolation/purification of DNA is of critical importance when considering the success of subsequent protocol establishment. It is required that the DNA purification procedure results in high yields of high quality DNA. High DNA recovery is a fundamental pre-requisite to avoid losses of informative MseI-fragments in order to ensure a broader representative population to be analyzed.

3.2 Concept of the WGA library modification

The genome fractioning to reduce the complexity of the WGA, in order to select specific MseI-fragment populations, requires some preceding steps involving the modification of the original Ampli1 WGA library. First, the WGA representation is re-amplified and the original adaptors are removed. Secondary adaptors will then be ligated to create a restriction site for a six-cutter enzyme that will generate specific ends for a tertiary type of adaptor, needed for the specific MseI-fragments selection before the final enrichment step.

3.2.1 Re-amplification of the MseI-fragments population and removal of the primary Lib1-adaptor

In order to generate sufficient amounts of high-quality double-stranded MseI-fragments for the successive steps of the protocol, the whole genome amplified MseI-fragment population needs to be re-amplified for a limited number of cycles. After the re-amplification, the

primary Lib1-adaptors need to be removed by digesting the re-amplified WGA representation with the restriction enzyme MseI (Figure 8). It is important at this step to ensure complete double strandedness of the re-amplified MseI-fragments to achieve optimal digestion of the primary Lib1-adaptors thereby providing sufficient dsDNA amounts for the consecutive steps of the protocol.

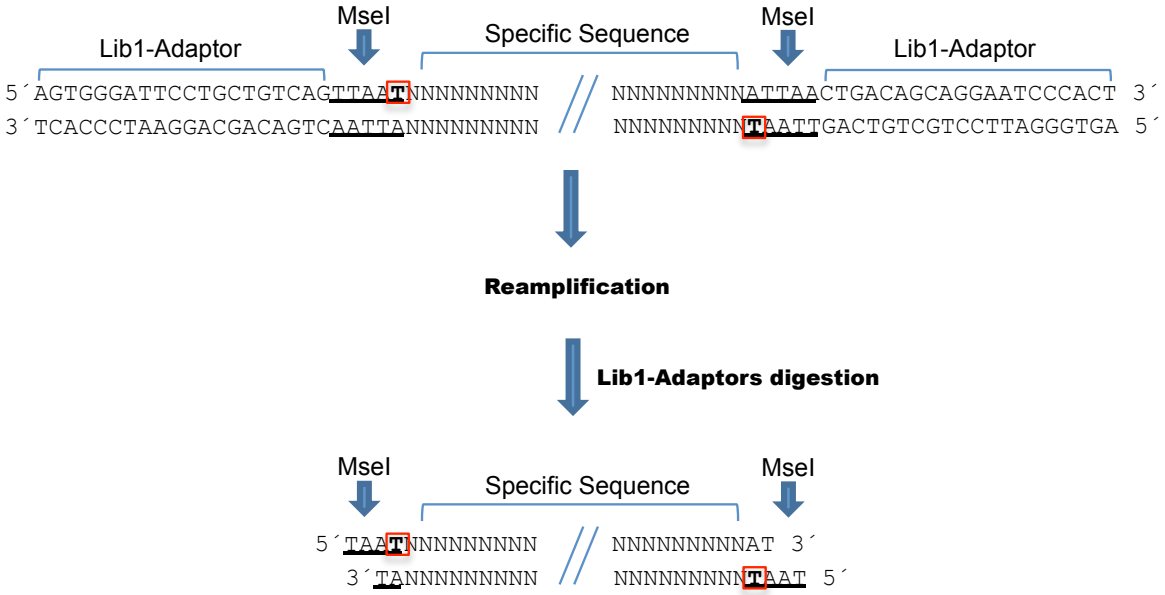


Figure 8: The WGA needs to be re-amplified to obtain sufficient amounts of DNA. After the re-amplification, the Lib1-adaptor is removed by digestion with MseI enzyme. In the red box are indicated, as example, the differentiating nucleotides after the MseI restriction site.

3.2.2 Introduction of the secondary adaptors

After removal of the original Lib1-adaptors used for the generation of the WGA, it is necessary to implement a step to mark each MseI-fragment of the original MseI-fragments population for subsequent fractioning. This is achieved by introducing secondary adaptors (J24) only differing at their terminal nucleotide position. Using these adaptors in equimolar ratios, each MseI-fragment terminus will randomly ligate to one of the four adaptors with the same probability. The function of the J24 (A,C,G,T) adaptors is to create a restriction site for a six-cutter enzyme that will digest the recognition sequence according to the reformed site at the end of the original MseI-fragment (Figure 9). This means that a restriction site is generated every fourth ligation, i.e. four different possibilities of ligation will form four different recognition sequences for specific six-cutter enzymes. This process is independent from the distribution of the differentiating bases in the genome. Whenever the secondary adaptor ligation does not lead to the formation of a palindromic six-cutter enzyme restriction

site at the end of the original MseI-fragment, the MseI-fragment will be excluded for further use. The probability that a reformed restriction site is created results from the percentage of successfully digested and removed Lib1-adaptors for the percentage of successfully ligated J24 (A,C,G,T) adaptors, that means the adaptors in which the terminal nucleotide forms a palindromic sequence with the differentiating base. At 100% efficiency of MseI digestion and J24 (A,C,G,T) ligation, one fragment has 6.25% of possibilities to be digested by the specific six-cutter. This implies that the amount of DNA used for the digestion is considerable high and necessarily in a double strand form.

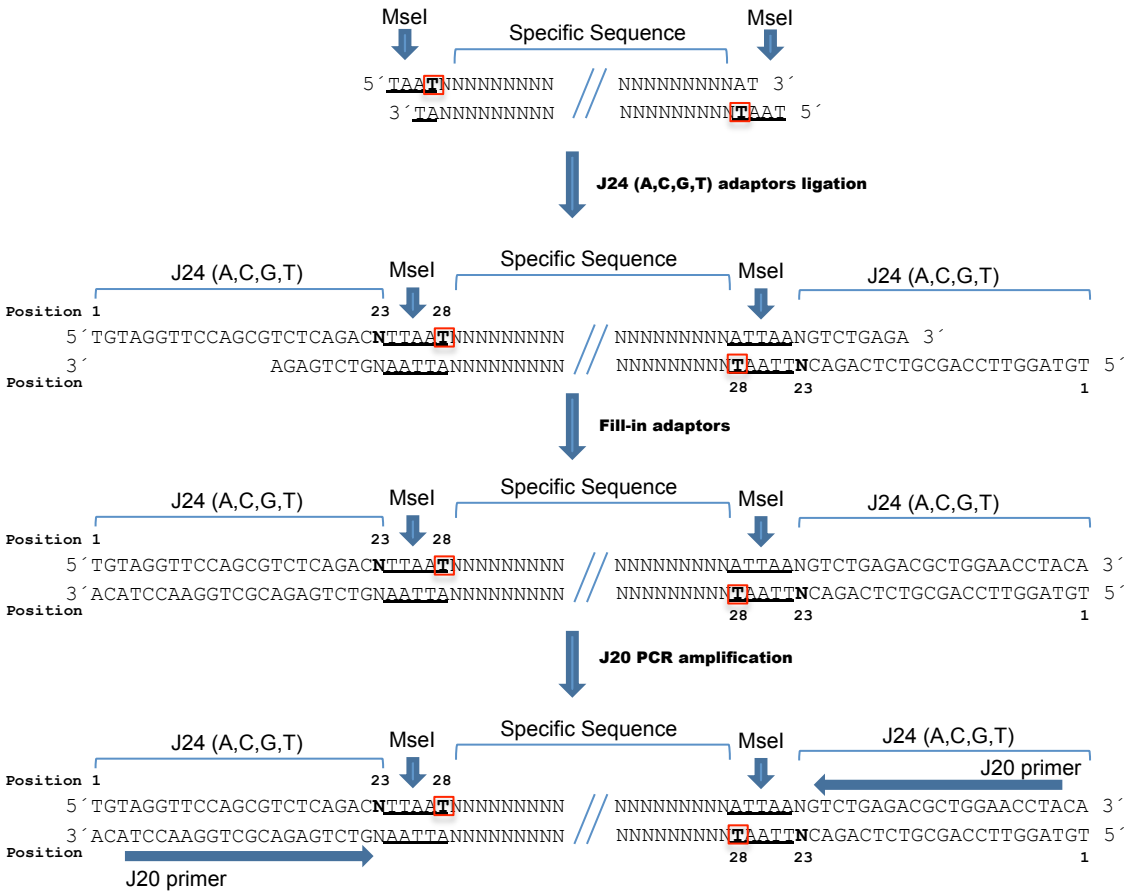


Figure 9: The J24 (A,C,G,T) adaptors are designed to create a restriction site for four different six-cutter enzymes, according to the N-terminal base of the adaptor (base at position 23) that will ligate, and to the differentiating base after the MseI restriction site (base at position 28). After the J24 (A,C,G,T) adaptor ligation, the J24 adaptor ligated MseI-fragments are shortly re-amplified with J20n primer in order to obtain sufficient amount of DNA for the specific restriction digestion. The MseI digestion site is underlined, differentiating nucleotide are indicated in the red boxes, J20 primers as long blu arrows.

After J24 (A,C,G,T) adaptor ligation, adaptor-ligated MseI-fragments are further amplified to enrich all the successfully reformed MseI-fragments. For this purpose, a primer (J20) complementary to the unique part of the J24 adaptors is used for the amplification (Figure 9). The end of this process enriches all the original MseI-fragments according to the new

reformed six cutter restriction sites. For optimal performance, the ligation efficiency of the four different J24 (A,C,G,T) adaptors should be ideally the same and, moreover, the secondary amplification needs to be optimized to obtain sufficient amounts of high quality dsDNA for the specific digestion.

3.3 Molecular properties of the Genome Fractioning procedure (Complexity Reduction Step)

The WGA library modification comprises series of steps transforming the MseI-fragment population in a condition that enables the reduction of the human genome complexity ("Genome Fractioning"). This genome fractioning procedure consists mainly of three basic steps: the specific restriction digestion, the selection of the digested MseI-fragments with specific fractioning-adaptors, and the selective enrichment PCR (fractioning PCR).

3.3.1 Specific restriction digestion to select specific MseI-fragments

The amplified J20 adaptor-ligated MseI-fragments have to be specifically selected by enzymatic digestion with individual six-cutters enzymes according to the reformed restriction site at the end of the original genomic MseI-fragments. As previously mentioned, according to the ligation of the nucleotide at position 23, a new restriction site will be created for one of the four different six-cutter enzymes: DraI (^{23}**T**TTAA**A**^{28}), VspI (^{23}**A**TTAA**T**^{28}), KspAI (^{23}**G**TTAA**C**^{28}) and BspTI (^{23}**C**TTAA**G**^{28}) (Figure 10). Only the digestion with specific enzymes will allow to select a pre-determined population of MseI-fragments: the digestion with one enzyme will select MseI-fragments harboring simple nucleotide combinations, whereas the digestion with two enzymes will select MseI-fragments harboring mixed nucleotide combinations. For instance, if the isolation of original MseI-fragments harboring on both sides the differentiating base T (simple fraction; ^{23}**A**TTAA**T**^{28}) is intended, the VspI enzyme is used to cut the J20 re-amplified MseI-fragment population (Figure 10, top line), or if the isolation of the original MseI-fragments harboring on both sides the differentiating base A (simple fraction; ^{23}**T**TTAA**A**^{28}) is intended, the DraI enzyme is used to cut the J20 re-amplified MseI-fragment population (Figure 10, second line). However, if the isolation of original MseI-fragments harboring on both sides the two different differentiating bases T/A (mixed fraction; ^{23}**A**TTAA**T**^{28} and ^{23}**T**TTAA**A**^{28}) is intended, two restriction enzymes (VspI

and DraI) have to be used to used to cut the J20 re-amplified MseI-fragment population (Figure 10, fifth line).

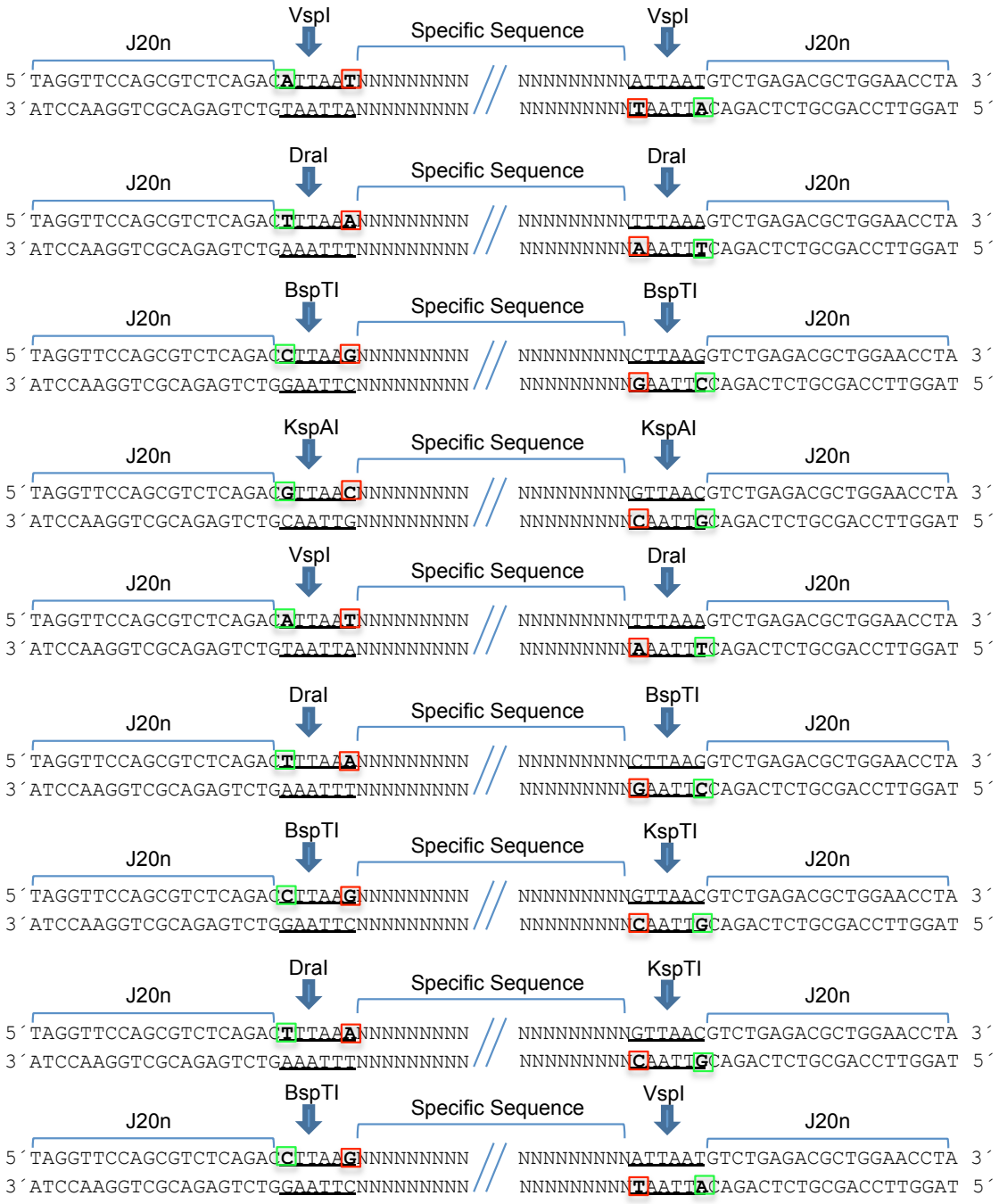


Figure 10: Specific digestion. The amplified J20 adaptor-ligated MseI-fragments are digested with the specific six-cutter enzymes in order to pre-select the sequences that are going to be enriched in the genome fractioning step. The restriction site for each enzyme is represented. Either one or two restriction enzymes can be used to select the original MseI-fragments according to the base ligated before the MseI site (green box) and the differentiating nucleotide present after the MseI site (red box).

3.3.2 Use of the specific fractioning-adaptors coupled to magnetic streptavidin-conjugated beads

The fractioning-adaptors 26R (A,C,G,T) and Bio26R (A,C,G,T) are specifically designed to ligate to the original MseI-fragments according to the ends generated by the digestion of the six-cutter restriction enzymes at the reformed site after the J24 (A,C,G,T) adaptor ligation. Depending on the original MseI-fragment harboring the same or different reformed six-cutter restriction site at the opposite ends, either a biotinylated or a non-biotinylated version of the 26R adaptor has to be used. To be precise, one biotinylated fractioning-adaptor will be used for the MseI-fragments that harbor the same restriction site on the opposite ends (simple fraction), and one biotinylated and one non-biotinylated fractioning-adaptor will be used for the MseI-fragments that harbor different restriction sites on the opposite ends (mixed fraction) (Table 11). The use of biotinylated fractioning-adaptors coupled to magnetic streptavidin-conjugated beads is essential to positively select those MseI-fragments, which have been correctly ligated after the specific digestion. They are used either for the simple or the mixed fractions as part of a streptavidin-conjugated fractioning-adaptor complex generated by pre-incubation with magnetic streptavidin-conjugated beads. By this means unbound fractioning-adaptors can be easily removed through consecutive washes of the magnetic streptavidin-conjugated beads.

3.3.3 Molecular mechanisms involved in the generation of the simple fractions

Streptavidin-conjugated fractioning-adaptor complexes containing the Bio26R (A,C,G,T) fractioning-adaptor are ligated to specifically digested MseI-fragments, e.g. with VspI (Figure 11). The streptavidin-conjugated fractioning-adaptor complex is incubated with the digested MseI-fragments to allow binding of the fractioning-adaptors to the respective DNA ends. MseI-fragments that bound to the streptavidin-conjugated fractioning-adaptor complex are then isolated with the help of a magnet, the fractioning-adaptors are completed with a fill-in reaction and the DNA is amplified by the 1st fractioning PCR, consisting of a limited number of cycles, with one amplification primer complementary to the biotinylated fractioning-adaptors (Figure 11).

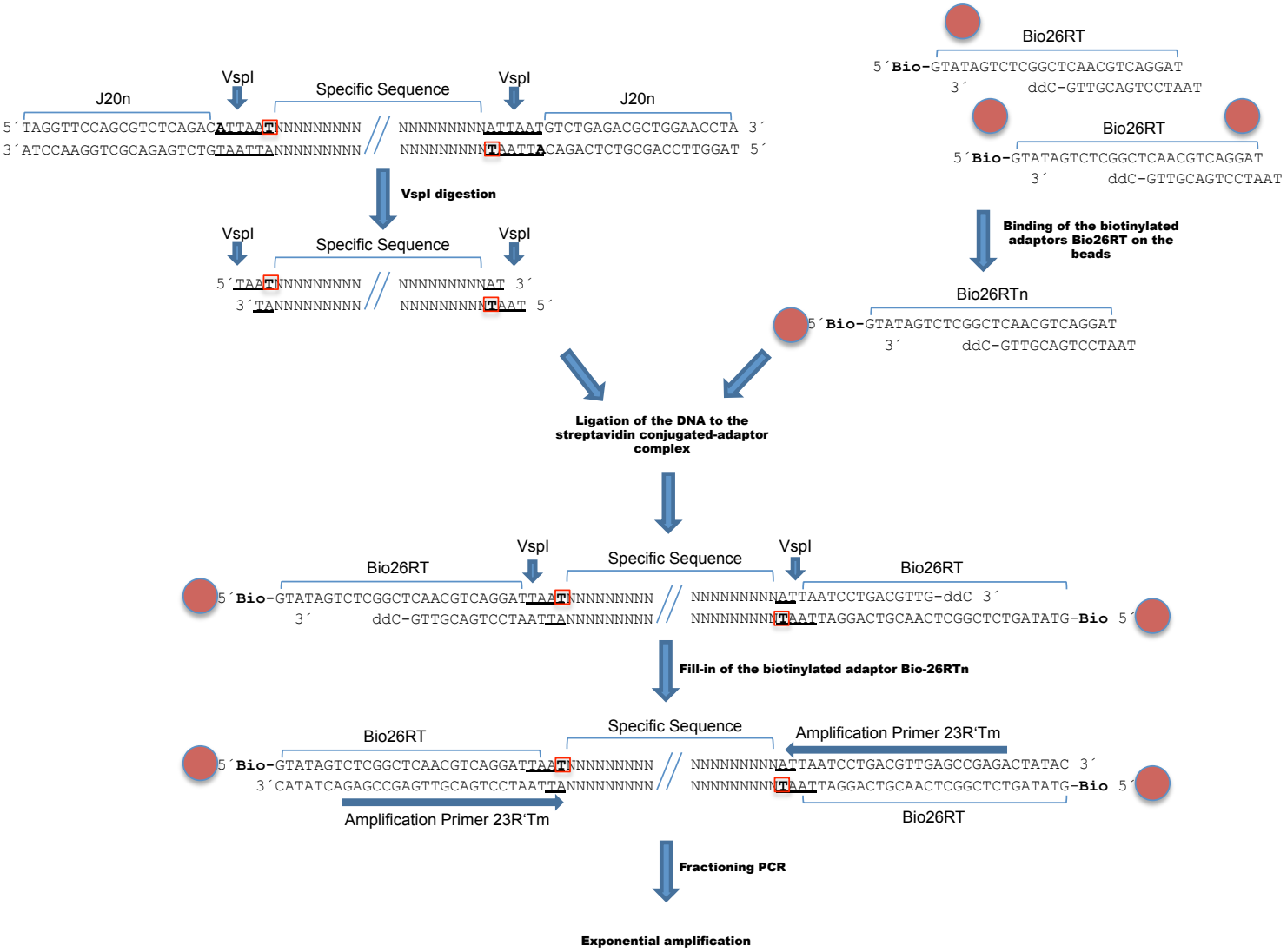


Figure 11: Selection and enrichment of the simple MseI-fragments. In the case of simple MseI-fragments the same ends will be generated on both sides of the DNA sequence: a single biotinylated fraction-adaptor (Bio26RT) will be used to ligate the ends; the biotinylated fractioning-adaptors are incubated with magnetic streptavidin-conjugated beads (brown circles) to allow the binding of the Bio26RT fractioning-adaptor to the beads; the DNA molecules are incubated with the streptavidin-conjugated fractioning-adaptor complex to allow the ligation of the DNA to the fractioning-adaptors. The DNA molecules that have incorporated the fractioning-adaptors will be positively selected. After the ligation the fractioning-adaptors are completed by filled-in reaction to allow the PCR with specific primers (long blue arrow). The differentiating nucleotides are indicated in the red box.

The 1st fractioning PCR is followed by a single strand nuclease digestion step to remove single strand DNA (ssDNA) molecules originating from the linear amplification of MseI-fragments that just bound a single adaptor, e.g. in the process of isolating the MseI-fragments carrying a T as a differentiating base at both fragment ends (harboring a VspI restriction site on both sides), MseI-fragments presenting the mixed differentiating base combinations A/T, T/G or C/T at their fragment ends (harboring a VspI restriction site only on one side) will also be ligated with the biotinylated fractioning-adaptor specific for the VspI restriction site, positively selected and linearly amplified (Figure 12). Following the removal of such linear

amplification products by Exonuclease I digestion, the 2nd fractioning PCR is performed to exponentially amplify the dsDNA molecules carrying the same adaptor on both fragment ends.

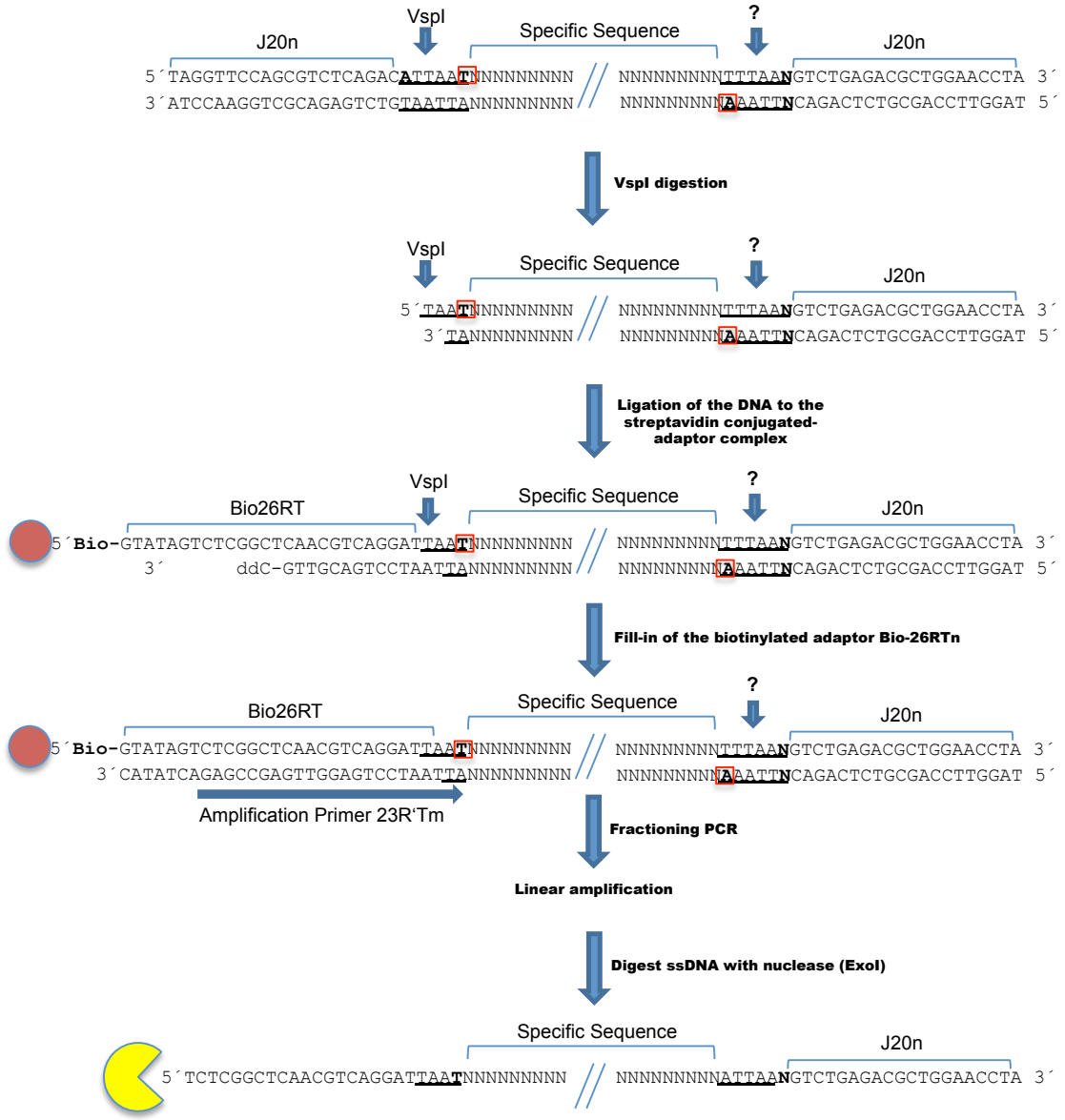


Figure 12: Mechanism of linear amplification. MseI-fragments harboring T/A, T/G, or T/C as differentiating bases are cut by the enzyme, e.g. VspI, only on one DNA terminus, so they will bind the fractioning-adaptors streptavidin-conjugated bead complex. The other DNA terminus will remain undigested and will preserve the sequence of the J20n primer. During the PCR, the amplification primer will prime only on one binding site generating always a 5'-3' ssDNA molecule. The ssDNA generated is digested with Exonuclease I (yellow circle) to reduce the background of adaptor-ligated MseI-fragments that have not been exponentially amplified. Other mechanisms that could generate linear amplified DNA are also incomplete digestion of one side of the MseI-fragment or, again, failed ligation of the specific fractioning-adaptor that bounds only on one side instead of two. The differentiating nucleotides are indicated in the red box.

3.3.4 Molecular mechanisms involved in the generation of the mixed fractions

Concerning the generation of the mixed fractions, the concept is very similar to the one applied to the simple fractions except that two six-cutters are used to generate two different ends on the MseI-fragment. Accordingly, two different site-specific fractioning-adaptors are ligated, one biotinylated and one not-biotinylated (Figure 13). J20 adaptor-ligated MseI-fragments are digested with the first specific enzyme, for example DraI, and the non-biotinylated fractioning-adaptor (26RA) is ligated to the cleaved end followed by a fill-in reaction to create a fully double-strand fragment end. After digestion of the J20 adaptor-ligated MseI-fragments with the second specific enzyme, e.g. VspI, cleaved fragments are incubated with the pre-formed streptavidin-conjugated fractioning-adaptor complex (with Bio26RT) to allow for ligation and magnetic capture of correctly ligated MseI-fragments. The biotinylated fractioning-adaptor ligated to the second restriction site of the MseI-fragment is not filled-in to avoid the exponential amplification of MseI-fragments that are digested from the same enzyme and ligated by the same biotinylated fractioning-adaptors (e.g. MseI-fragments harboring the T/T nucleotides; Figure 11). Also, MseI-fragments harboring T/C and T/G nucleotides that are digested on one side from the enzyme VspI and ligated to the biotinylated fractioning-adaptors will be prevented from linear amplification due to the lack of the amplification primer binding site. A 1st fractioning PCR is necessary to enrich the MseI-fragments that have been successfully ligated with two different site-specific fractioning-adaptors. Similar to the simple fractions a ssDNA digestion step is implemented before the 2nd fractioning PCR.

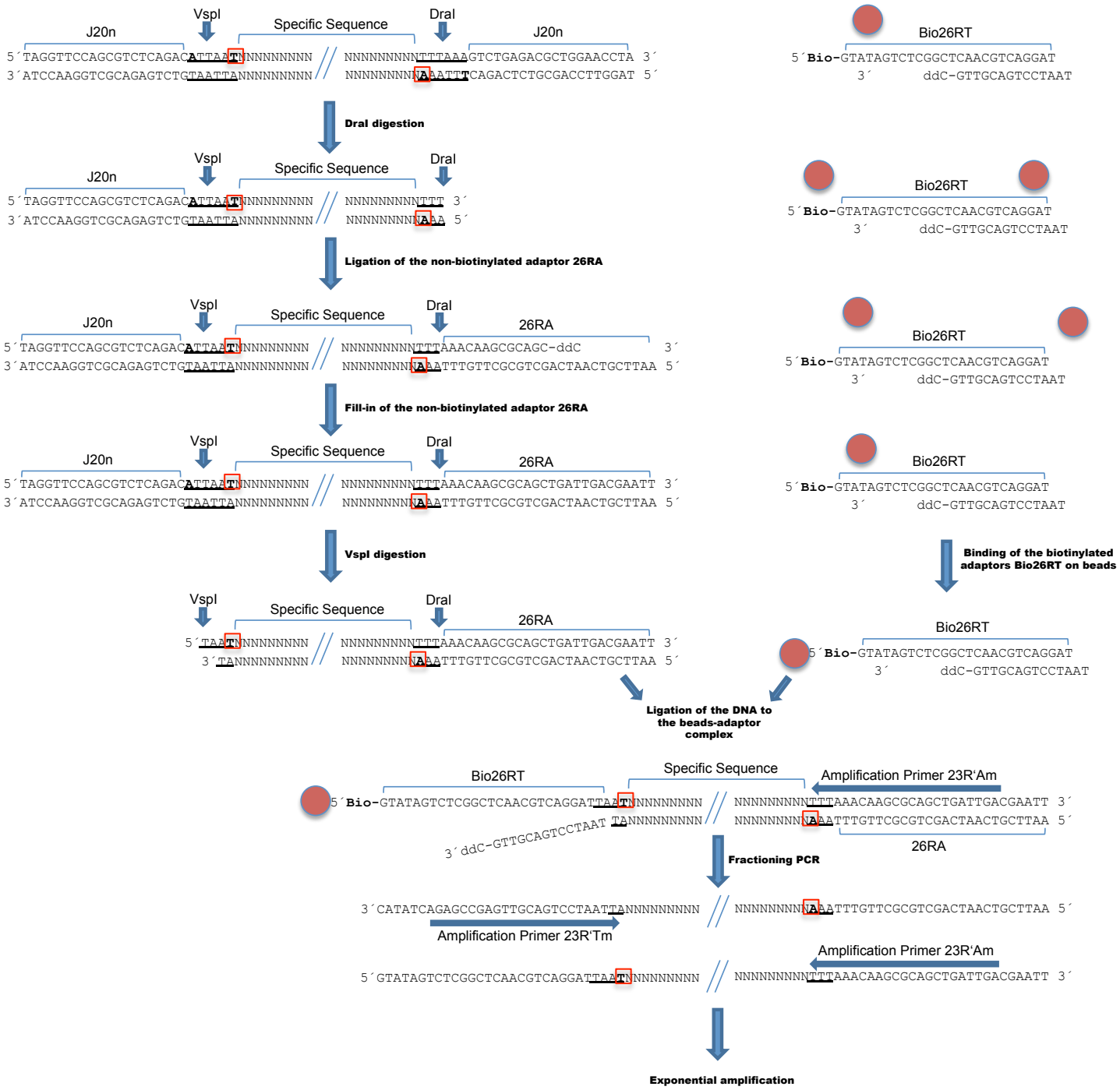


Figure 13: Generation of the mixed fractions. Bio26RT adaptors are pre-bound on the magnetic streptavidin-conjugated beads represented as brown circles. Amplification primers 23R'Tm and 23R'Am are represented as long blue arrow. The differentiating nucleotides are indicated in the red box.

3.4 Subtractive DNA Hybridization as a tool to enrich genomic breakpoints in cancer genomes of single cells

Subtractive DNA hybridization is a technique allowing for the enrichment of sequences differing between two closely related cDNA or genomic DNA samples. In subtractive hybridization techniques, one DNA population (the "Driver") is hybridized in excess with a second DNA population (the "Tester"), which is similar but not identical to the first DNA population. The differences between these two DNA populations (present in the Tester, but not in the Driver) are the "target" sequences for isolation. Typically, the Tester DNA is digested with a restriction endonuclease, ligated to adaptors and mixed with an n-fold excess of digested Driver DNA. The DNA mixture is then denatured and allowed to re-associate to form three types of hybrids: Tester-Tester and Driver-Driver homohybrids and Tester-Driver heterohybrids. The chance that a non-target Tester sequence hybridizes with its Tester complement is substantially smaller than the chance that it hybridizes with its complement from the Driver DNA population. The target sequence is present only in the Tester and will therefore always hybridize with its complement from the Tester DNA population. The excess of Driver DNA is thus used to drive the non-target Tester sequences from the pool of Tester DNA. In order to reach a proper DNA re-hybridization, the conditions for the re-association have to be almost optimal. The choice and optimization of the subtractive DNA hybridization procedure are important to ensure the maximum enrichment of differential MseI-fragments, in this case the best enrichment of the BCR/ABL breakpoint-MseI-fragment sequence. In order to yield optimal results, the re-amplification step to generate sufficient amounts of genome MseI-fraction must be optimized. Moreover, the maximum binding capacity of the magnetic streptavidin-conjugated beads for biotinylated MseI-fragment sequences has been investigated to evaluate if a reverse approach consisting in the use of large amount of biotinylated Driver DNA for the subtraction would be possible. Also different methods for the re-hybridization of nucleic acids can be tested (Figure 14).

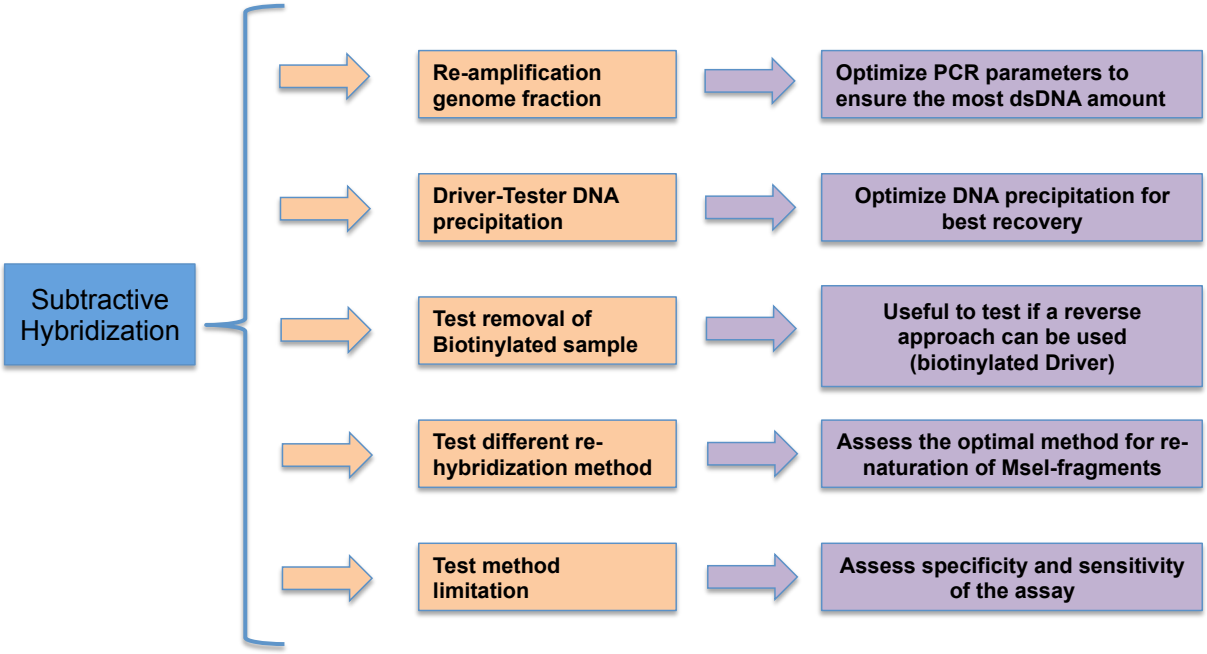


Figure 14: Overview of the optimization steps necessary to establish the process of subtractive hybridization.

3.4.1 Re-amplification of Driver and Tester samples to obtain sufficient DNA for subtraction

In order to yield enough DNA amounts for the different rounds of subtraction the Driver and Tester DNA must be re-amplified. Moreover, it is important that the re-amplification is optimized to maximize the amount of dsDNA. This will allow the proper assessment of the Driver to Tester ratio necessary for the optimal enrichment of the BCR/ABL breakpoint-MseI-fragment.

3.4.2 DNA precipitation of Driver and Tester

To ensure fast and proper renaturation, high amount of Driver DNA has to be mixed with limited amounts of Tester DNA before the re-hybridization in small volumes. The precipitation of the mixed DNA is necessary for this step, as well as an almost complete recovery of the mixed DNA sample to ensure that the Driver to Tester ratio is unchanged in the following re-hybridization. Classical phenol-chloroform-isoamylalcohol extraction and ethanol precipitation can be used but also commercial systems excluding contamination of the organic phase and yielding maximum recovery of DNA are available.

3.4.3 Subtractive approaches with biotinylated Driver or Tester DNA

Subtractive DNA hybridization approaches can either use a biotinylated Driver DNA sample to completely remove background sequences derived from selection of Tester-Driver heterohybrids, or a biotinylated Tester DNA sample to positively select Tester-Tester specific homohybrids. One problem that occurs by using the latter approach is the simultaneous selection of the Tester-Driver heterohybrids that are linearly amplified in the enrichment PCR. These sequences are mostly, but not always completely, removed by the help of an ssDNA nuclease. If an approach with biotinylated Drive DNA would prove to be successful, any selection and linear amplification of the Tester-Driver heterohybrids could be excluded, and also the intermediate nuclease step to remove ssDNA could be excluded.

3.4.4 Subtractive hybridization using different re-hybridization methods

Different methods available from literature can be used to re-hybridize DNA. The choice of a method to ensure the best enrichment of differential MseI-fragments between a Driver and Tester sample can be assessed considering different factors. First the method has to prove that after the re-hybridization most of the DNA is again double stranded and second that the level of enrichment of differential MseI-fragments, in this case BCR/ABL breakpoint-MseI-fragment, is considerable.

3.4.5 Specificity and sensitivity of the subtractive hybridization method

The selected method has to provide an appropriate specificity and sensitivity. This can be investigated by spiking known amounts of differential MseI-fragments in a Driver sample that underwent genome fractioning in order to simulate the process of subtraction hybridization. A specific method will provide the complete banding pattern of re-hybridized and enriched MseI-fragments without any other fragment deriving from improper hybridization. Additionally if the method is sensitive all the MseI-fragments spiked-in at very low amounts will be detectable after re-hybridization followed by endpoint PCR.

3.5 First subtraction: generation of the Depletion product (DP1)

Figure 15 depicts the exact subtractive DNA hybridization approach used in this experimental setting. Briefly, the enriched genome MseI-fraction is re-amplified using a limited-cycle PCR to generate sufficient amounts of DNA for subtractive DNA hybridization. Many reactions are generally performed in order to generate the necessary quantities of the Driver DNA sample (lymphocyte pool), while only one reaction is sufficient to generate the necessary amount of Tester DNA sample (K562 single cell). The Driver and Tester DNA are digested with MseI enzyme to remove the fractioning-adaptor remnants. Prior to the first round of subtraction the Tester DNA is ligated with biotinylated subtractive-adaptor (BioR21-adaptor) in order to select, after the re-hybridization, Tester-Tester homohybrids and the Tester-Driver heterohybrids by the use of magnetic streptavidin-conjugated beads. A large excess of Driver and limited amount of Tester are mixed together and heat-denatured. The re-hybridization of the samples is happening in aqueous solution in a time frame of 20-24 hours. The re-hybridized DNA sequences are mixed together with magnetic streptavidin-conjugated beads to allow the binding of the Tester-Tester homohybrids and Tester-Driver heterohybrids while Driver-Driver will not bind on the beads. The BioR21-adaptor is completed by a fill-in reaction, and the DNA sequences immobilized on the beads are amplified by a first enrichment PCR with a non-biotinylated primer (R21 primer) complementary to the BioR21-adaptor, followed by mung bean nuclease digestion to remove ssDNA molecules originating from linear amplification of Tester-Driver heterohybrids. This step is followed by a second enrichment PCR resulting in the exclusive exponential amplification of the Tester-Tester homohybrids carrying the BioR21-adaptor on both fragment ends. Subsequently the first depletion product (DP1) is normalized in order to minimize the presence of cross-annealed repetitive sequences in the re-hybridized sample thereby counteracting the excessive accumulation of such chimeric sequences in successive steps of the protocol.

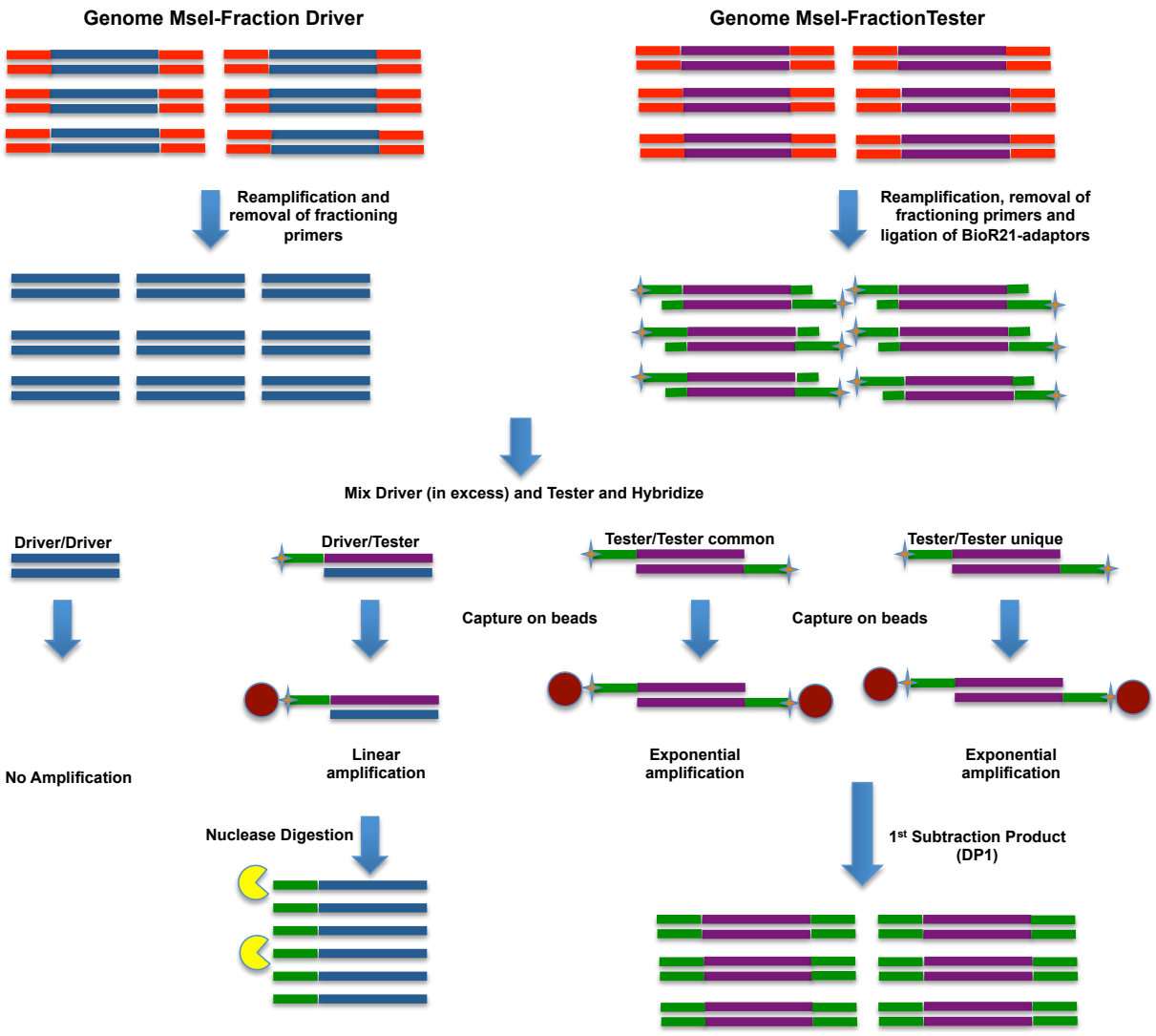


Figure 15: Generation of Depletion Product 1 (DP1). Fractioning-adaptors, removed after genome fractioning, are indicated in red. Biotinylated subtractive-adaptors (BioR21-adaptor), ligated only to the Tester, are indicated in green. Magnetic streptavidin-conjugated beads necessary to positively select Tester-Tester and Tester-Driver sequences are represented as brown circles. ssDNA nuclease necessary for the removal of linearly amplified sequences is represented as yellow circle. NB: Tester/Tester common sequences represent DNA sequences that are also present in the Driver sample but that are not completely subtracted. Tester/Tester unique sequences are exclusively present in the Tester sample.

3.6 Generation of the successive depletion products

The normalized first depletion product (DP1) is then used for a second round of subtractive DNA hybridization. The lack of a BioR21-adaptor (as far as, enrichment amplification and amplification of the normalized DP1 product is carried out with the non-biotinylated primer R21) does not allow for the second round of subtractive DNA hybridization the use of magnetic streptavidin-conjugated beads for selective enrichment of the re-hybridized Tester-Tester homohybrids. Therefore a mung bean nuclease digestion is performed immediately after the re-hybridization of Tester and Driver DNA sample. By this means unpaired ssDNA

sequences and overhanging subtractive-adaptor ends of Tester-Driver heterohybrids are removed thereby excluding those sequences from linear amplification (Figure 16).

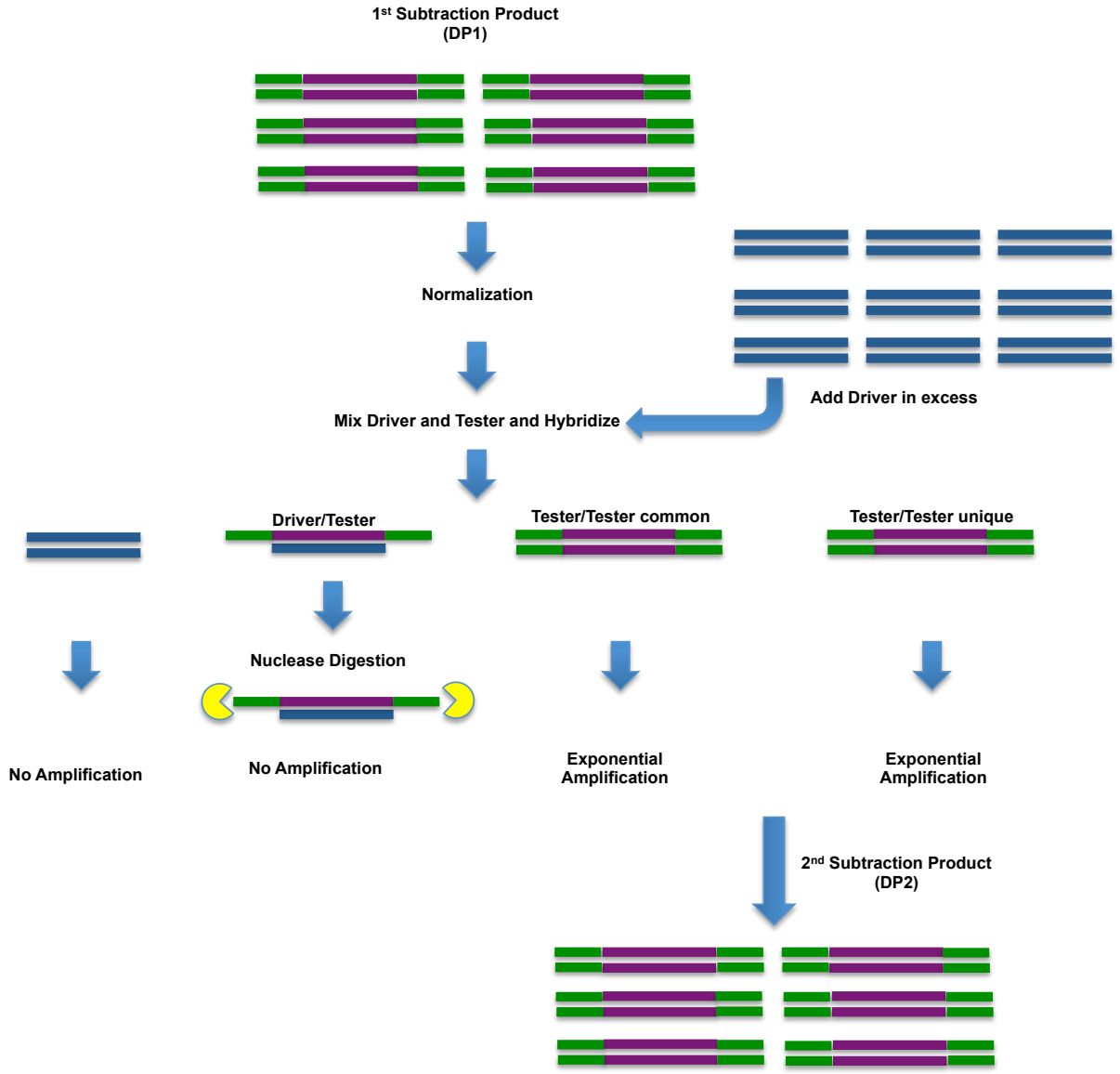


Figure 16: Generation of depletion product 2 (DP2). After the first subtraction round, the DP1 is normalized to decrease selected repetitive sequences and it is re-hybridized again to an excess of Driver. The DP1 preserve the same primer binding site (R21, in green) but without biotinylation motif. After the hybridization Driver-Driver sequences will not be amplified as well as Tester-Driver sequences in which the primer binding sites will be digested from the mung bean nuclease (yellow circle). Tester-Tester sequences will be again exponentially amplified. NB: Tester/Tester common sequences represent DNA sequences that are also present in the Driver sample but that are not completely subtracted. Tester/Tester unique sequences are exclusively present in the Tester sample.

4. Results

4.1. Establishment of the quality controls on WGA amplified K562 single cells

4.1.1 PCR-based Ampli1 WGA quality control assay

It is important to ensure that only WGA samples of high-quality are used during the process of method development (Method Design Paragraph 3.1.1). Therefore it is necessary to control the comprehensive amplification of a single cell genome by testing the presence of the four amplicons CK19, TP53, D5S2117, and KRAS (Table 13) (Figure 17).

Table 13: Ampli1 QC amplicons used to check the DNA integrity of single cell WGA samples. Amplicon name, length of the WGA amplified MseI-fragment and length of the PCR fragment are indicated.

Primer ID	Lenght of the MseI-Fragment	Lenght of the PCR fragment
D5S2117	1376 bp	140 bp
TP53	1374 bp	301 bp
CK19	1146 bp	614 bp
KRAS	192 bp	91 bp

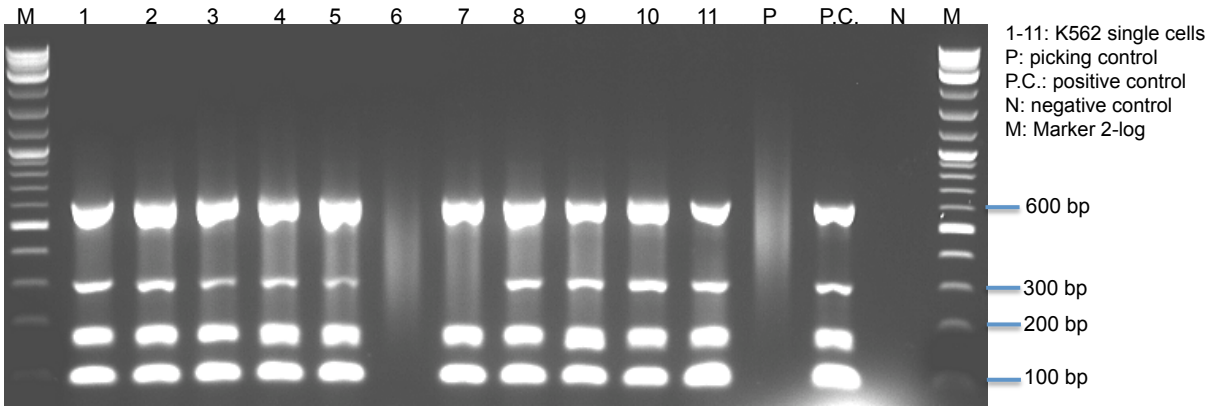


Figure 17: Multiplex quality control PCR assay of isolated and amplified single cell genomes of the K562 cell line. A positivity rate of 4/4 and 3/4 markers is representative of very good quality DNA. PCR positive control is composed of a pool of unrelated peripheral blood lymphocytes. Please note that almost all cells show a very good WGA quality except for cell number 6. The lack of any detectable MseI-fragments indicates the loss of the cell during the isolation.

A cell with good quality DNA is identified by the presence of three or four of the four tested amplicons. Not successfully picked/amplified cells will not produce any amplification products. This allows to determine the Genome Integrity Index (GII0-4) for each amplified single cell genome based on the number of detected PCR bands.

4.1.2 Whole genome amplified-DNA size distribution

The determination of the size distribution of initial high-quality WGA libraries allows to assess the fragment length profile of the amplified MseI-fragments in such libraries and therefore provides an independent indicator for genome integrity (Method Design Paragraph 3.1.2). From observation, the experimental size distribution of the MseI-fragment population ranges from ~250 to ~3000bp (Figure 18).

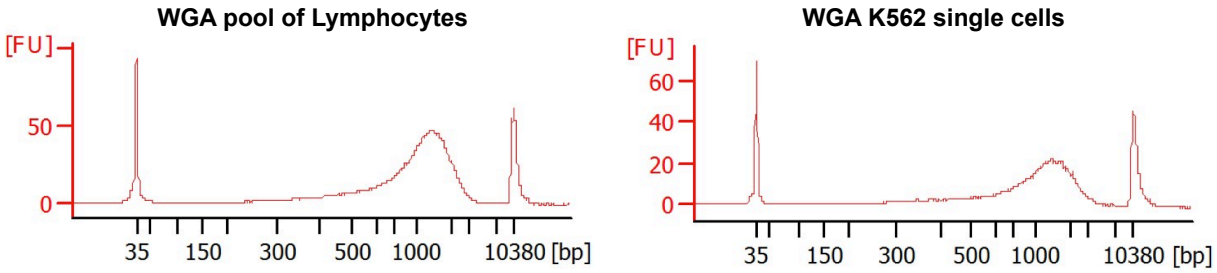


Figure 18: Examples of bioanalyzer profile of the WGA library for the lymphocytes pool and single K562 cells. The MseI-fragments distribution is between ~250 to ~3000bp. The X-axis represents the product size in base pairs (bp) and the Y-axis is the arbitrary fluorescence intensity (FU).

Low quality WGA showed a different profile compared to high quality WGA. As can be observed low quality WGAs show low amplification level (from FU data) and ambiguous MseI-fragments length distribution reflecting maybe the not optimal outcome of WGA amplification (Figure 19).

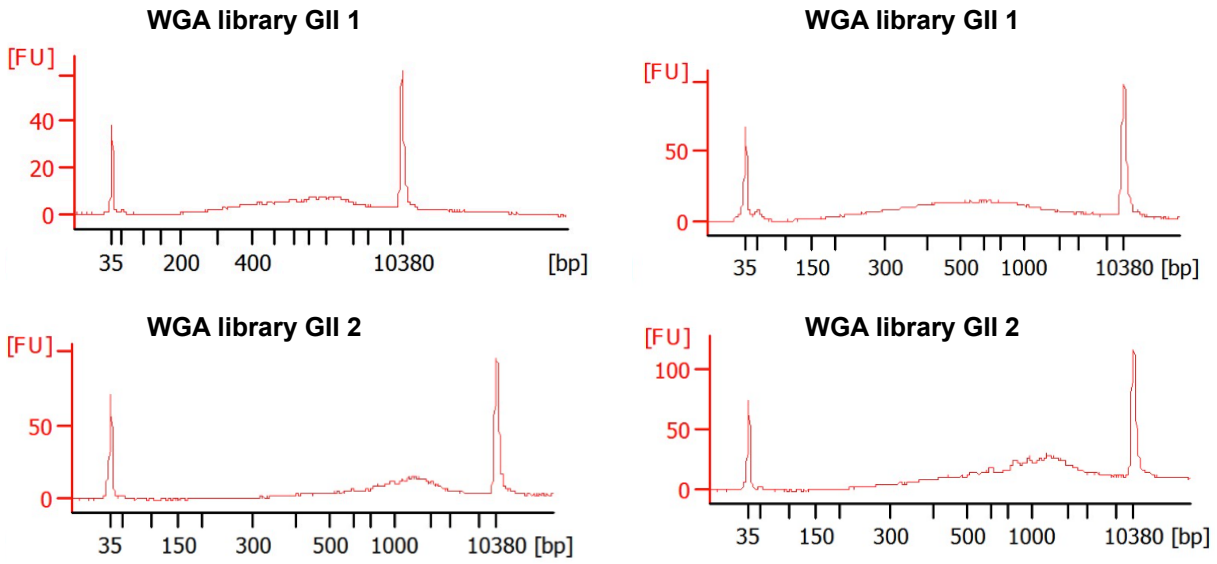
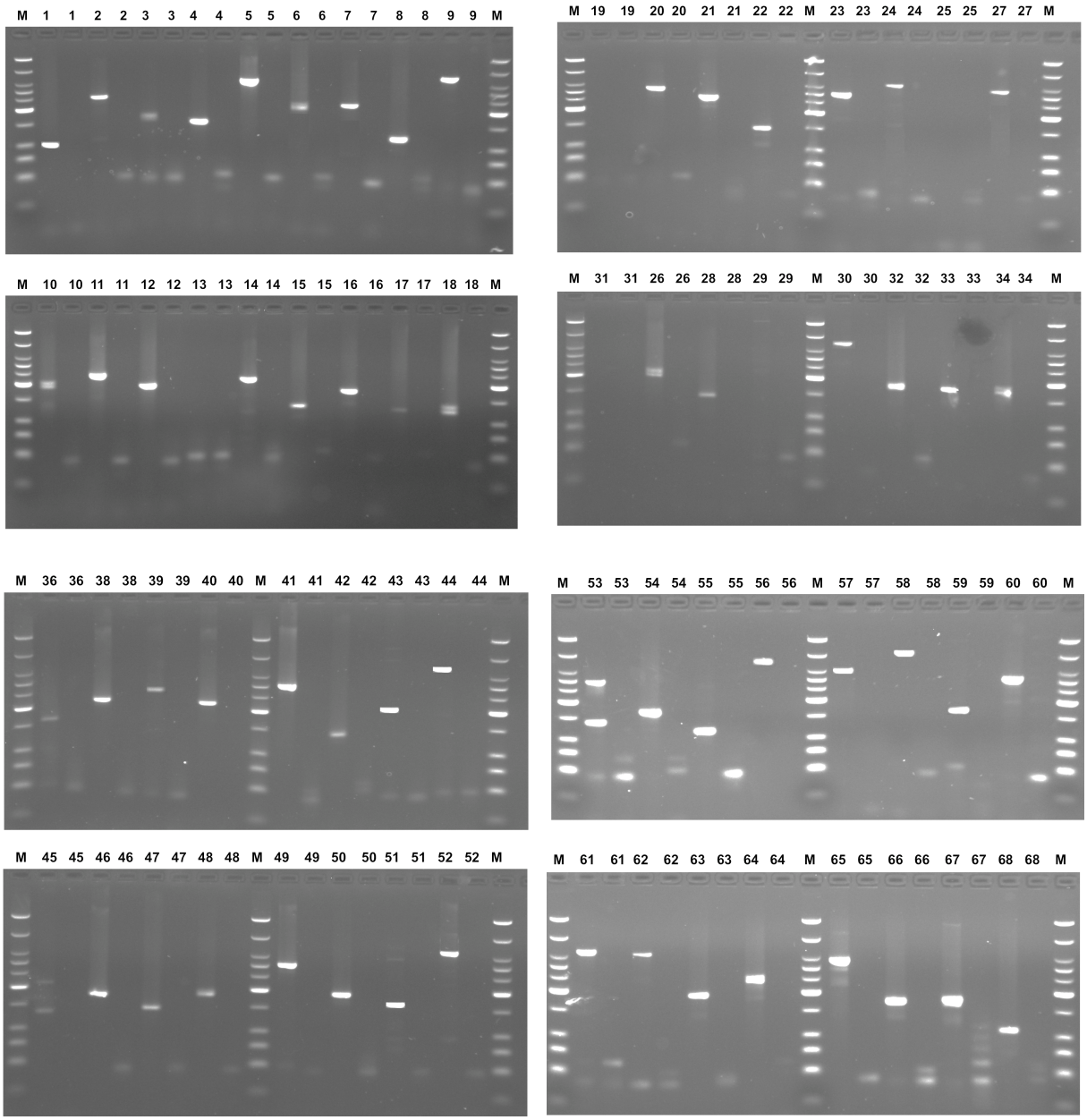
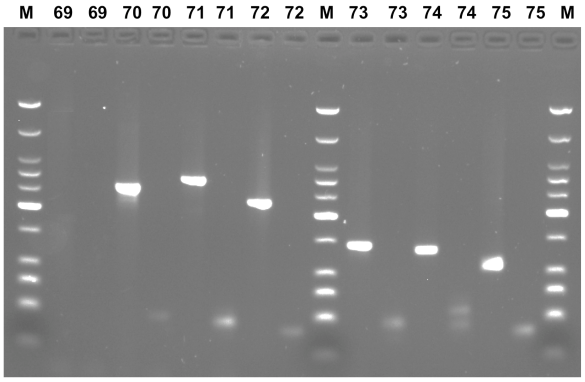


Figure 19: Examples of bioanalyzer profiles of low quality WGA libraries with GII 1-GII 2. The X-axis represents the product size in base pairs (bp) and the Y-axis is the arbitrary fluorescence intensity (FU).

4.1.3 Specific MseI-fragment PCR-based quality control assays

To ensure the proper and orderly separation of the ten sub-genomic MseI-fragments populations, it was necessary the establishment of a specific MseI-fragment PCR-based quality control assay enabling the detection of genomic MseI-fragments per each differentiating base combination (Method Design Paragraph 3.1.3). This assay allows gaining a more comprehensive view on the completeness of the genomic MseI-representation in the primary WGA sample (Figure 20).





Nr.	Primer ID	Differentiating base	MseI-fragment lenght (bp)	PCR-Product (bp)	Nr.	Primer ID	Differentiating base	MseI-fragment lenght (bp)	PCR-Product (bp)	Nr.	Primer ID	Differentiating base	MseI-fragment lenght (bp)	PCR-Product (bp)
1	D5S2117	A/A	1376	147	26	D16S3019	G/G	605	220	51	D17S800	AG/GA	333	153
2	BCR	A/A	751	243	27	BCR	G/G	367	329	52	TP53	TC/CT	1032	461
3	BRCA1	A/A	594	151	28	D5S592	G/G	348	191	53	ENTPD3	TC/CT	651	290
4	PDPR	A/A	568	144	29	BCR	G/G	142	138	54	ZFHX3	TC/CT	576	170
5	TP53	A/A	558	374	30	BCR	AT/TA	790	583	55	CDH3	TC/CT	552	119
6	LPO	A/A	316	193	31	BCR	AT/TA	701	346	56	D5S1360	TC/CT	513	139
7	TRAK1	A/A	247	214	32	BCR	AT/TA	600	358	57	D6S308	TC/CT	262	214
8	TRAK1	A/A	128	113	33	D5S299	AT/TA	550	159	58	OPN1SW	TC/CT	283	76
9	BCR	T/T	1936	413	34	UBP1	AT/TA	394	180	59	BCR	TG/GT	1351	679
10	D16S3066	T/T	536	190	35	SRP19	AT/TA	350	110	60	BCR	TG/GT	1068	300
11	BCR	T/T	508	235	36	D5S816	AT/TA	307	178	61	BCR	TG/GT	881	379
12	CLASP2	T/T	294	195	37	D16S3040	AT/TA	302	140	62	BCR	TG/GT	756	368
13	D6S314	T/T	277	210	38	APIG1	AT/TA	220	219	63	BCR	TG/GT	670	192
14	COG4	T/T	261	217	39	ZKSCAN7	CA/AC	648	257	64	D5S615	TG/GT	641	227
15	TFRC	T/T	242	131	40	CDH3	CA/AC	637	221	65	APC	TG/GT	618	330
16	TOP2B	T/T	209	177	41	RH27788	CA/AC	576	389	66	MYRIP	TG/GT	600	174
17	ABL1	T/T	173	125	42	D6S453	CA/AC	479	140	67	FAM117A	TG/GT	260	176
18	D5S500	T/T	159	132	43	AGTR1	CA/AC	279	105	68	CCK	TG/GT	231	109
19	ABL1	T/T	157	142	44	D6S1633	AG/GA	1317	270	69	ZNF19	TG/GT	216	131
20	BCR	C/C	1233	324	45	ULK4	AG/GA	682	126	70	BCR	GC/CG	1071	258
21	ABL1	C/C	291	273	46	CDH1	AG/GA	486	179	71	ZNF23	GC/CG	417	290
22	ABL1	C/C	203	148	47	ZFHX3	AG/GA	480	144	72	TRAK1	GC/CG	355	228
23	BCR	G/G	1026	297	48	D5S399	AG/GA	485	178	73	ULK4	GC/CG	298	145
24	BCR	G/G	944	360	49	RTP3	AG/GA	394	302	74	HERBB2	GC/CG	226	133
25	BCR	G/G	715	470	50	ANO10	AG/GA	370	190	75	D5S471	GC/CG	188	106

Figure 20: Specific MseI-fragment control PCR of the WGA with specific primers amplifying different MseI-fragments according to the nucleotide combination after the MseI restriction sequence. All the ten combinations of MseI-fragments are amplified in the WGA. The table summarizes the MseI-fragments and the differentiating bases where primers are located, the length of the MseI-fragments, and the PCR amplicons length. The markers are numbered and represent the amplification tested on single cell and negative control. Markers 13, 19, 25, 29, 31, 69 did not perform successfully on single cells. M: low molecular weight DNA marker (NEB).

4.1.4 Quantification of dsDNA

In order to ensure that sufficient quantities of dsDNA are available before the start of method development, the dsDNA amount of eleven K562 single cells and one pool of ten lymphocytes, which underwent Ampli1 WGA, are comparatively measured (Method Design Paragraph 3.1.4). With the exception of K562 single cell 6, all tested single cell samples showed similar results (Table 14). The higher amount of generated dsDNA in the cell pool is a result of the higher initial DNA quantity.

Table 14: dsDNA measurement of the WGA amplified K562 single cells and of the lymphocytes pool.

Sample	DNA concentration	DNA total amount
K562 single cell 1	20 ng/µl	1000 ng
K562 single cell 2	21,3 ng/µl	1065 ng
K562 single cell 3	21,6 ng/µl	1080 ng
K562 single cell 4	22,1 ng/µl	1105 ng
K562 single cell 5	19,8 ng/µl	990 ng
K562 single cell 6	0,003 ng/µl	0,15 ng
K562 single cell 7	18,6 ng/µl	930 ng
K562 single cell 8	20,8 ng/µl	1040 ng
K562 single cell 9	21 ng/µl	1050 ng
K562 single cell 10	19,1 ng/µl	955 ng
K562 single cell 11	18,4 ng/µl	920 ng
Lymphocytes pool	54,6 ng/µl	2730 ng

4.1.5 WGA and BCR/ABL qPCR

The well-characterized genomic breakpoint of BCR/ABL fusion gene in the K562 cell line has been used as a monitoring system to trace the degree of enrichment of BCR/ABL breakpoint-MseI-fragment in a single cell genome (Figure 21) (Method Design Paragraph 3.1.5).

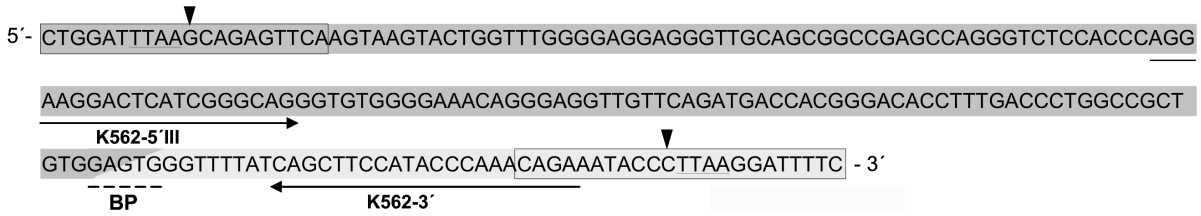


Figure 21: Chromosomal breakpoint (BP) of the K562 leukemia cell line. In dark gray is the sequence of the BCR gene, in light gray the sequence of the ABL gene. Control primers for the fragment are K562-5'III and K562-3'. Small black arrows point the differentiating bases of the MseI fragment in which the fusion gene is located, in this case an MseI-fragment of 200bp with G nucleotide as differentiating bases.

Absolute plasmid DNA standards, containing the BCR/ABL breakpoint-MseI-fragment in the range from 1000 to 1.000.000.000 copies, have been used to develop a calibration curve (Figure 22).

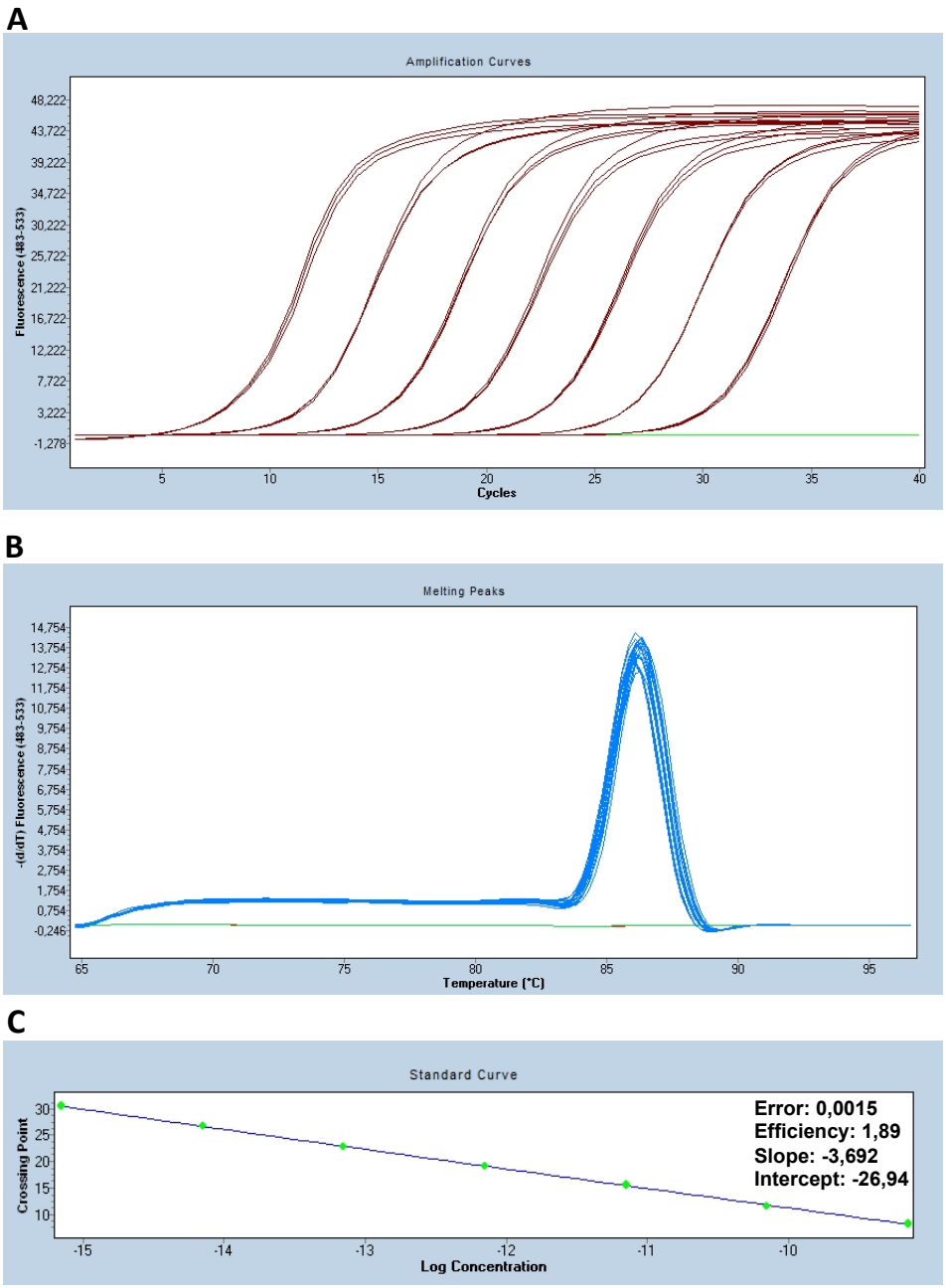


Figure 22: A) amplification curves, B) amplicon specific melting peak and C) standard curve for the BCR/ABL breakpoint-MseI-fragment. The standard curve ranges from a copy number of 1000 till 1.000.000.000 in 10-fold dilutions. Efficiency, error, intercept and slope are all in the acceptable range.

Plasmid calibration ranges were used to quantify the number of copies of the BCR/ABL breakpoint-MseI-fragment in K562 cells before and after Ampli1 WGA. Lymphocyte WGA pools were included in the assay to exclude any unspecific BCR/ABL amplification. The Figure 23 summarizes the amplification and melting curves of the whole genome amplified single cells proving the specificity of the amplified product.

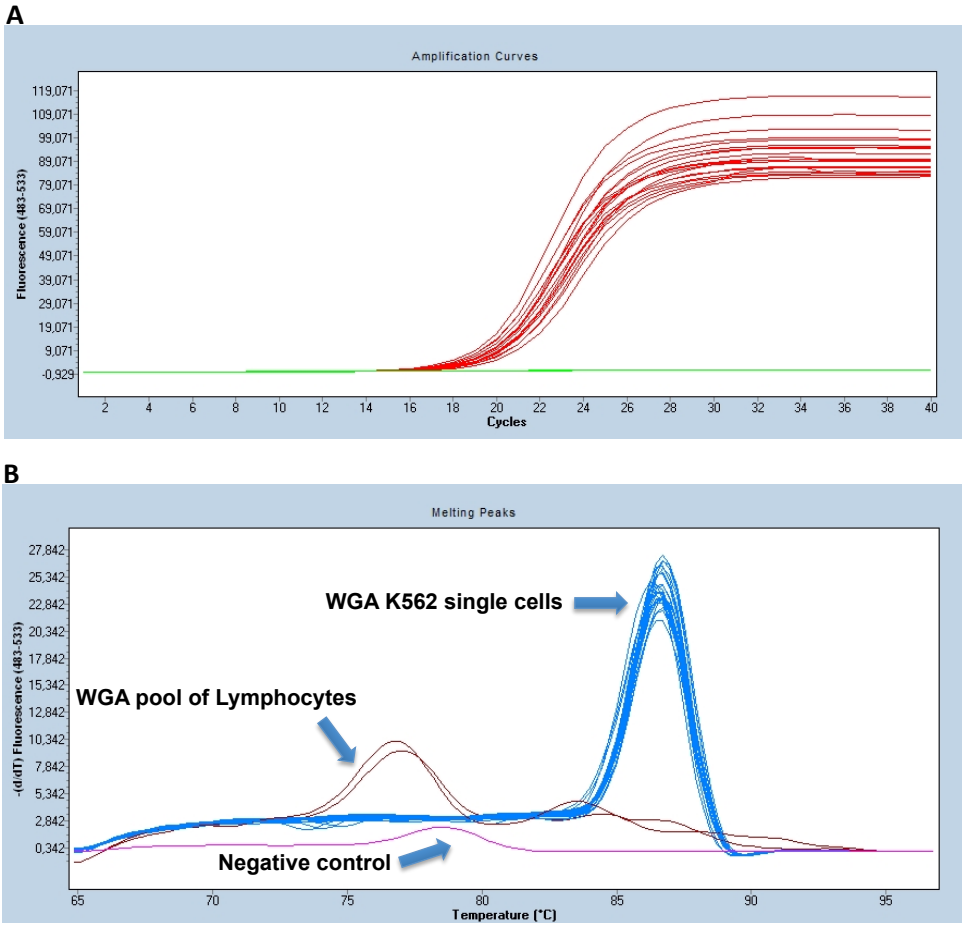


Figure 23: A) amplification curves of the WGA from single K562 cells for the BCR/ABL breakpoint-MseI-fragment, B) melting peak. The lymphocyte pool WGA is negative for the fusion gene.

All single cells showed similar copy numbers for the BCR/ABL breakpoint-MseI-fragment (Table 15). The copy numbers in the genomic DNA isolated from the K562 cells is also quite similar, indicating that the WGA did not significantly alter the gene ratio, at least for the sequence of interest.

Table 15: BCR/ABL breakpoint-MseI-fragments copies per nanogram of DNA for each amplified K562 single cell and unamplified genomic DNA calculated after absolute qPCR.

Sample	BCR/ABL copies/ng of DNA
K562 single cell 1	12215
K562 single cell 2	11230
K562 single cell 3	13001
K562 single cell 4	12214
K562 single cell 5	11475
K562 single cell 6	-
K562 single cell 7	10129
K562 single cell 8	11214
K562 single cell 9	11765
K562 single cell 10	10781
K562 single cell 11	12601
K562 genomic DNA unamplified	10229

4.1.6 Optimization of the DNA recovery with different purification systems

To select the best-working purification system under the given conditions (Method Design Paragraph 3.1.6), it was necessary to compare the following products with respect to purity and amount of the recovered dsDNA: Microcon YM30 (Millipore), Wizard SV PCR Clean-Up System (Promega Corporation), Qiagen PCR Purification Kit (Qiagen), AmpureXP purification beads (Beckman Coulter), DNA Clean & Concentrator 25 (Zymo Research Corporation, Irvine, CA). Starting amounts of 5µg dsDNA derived either from PCR amplification or restriction/ligation reactions were purified, and quantity and quality of the recovered dsDNA assessed by Qubit and Nanodrop measurement. Results of the use of the different purification systems are summarized in Table 16. The purification of PCR reactions gave similar results in terms of DNA recovery between the different systems: the Microcon column, mainly used to concentrate PCR reactions, gave high recovery but un-pure products with a 260/280 ratio very low, indicating the presence of many contaminants, even after repetitive washing steps of the column. The other purification method gave high recovery and pure samples with best results achieved with the use of AmpureXP beads. The purification of restriction and ligation reactions showed to perform differently probably due to the buffer composition: Wizard and Zymokit purification system showed a loss of around 40-50% of the purified product. The others did not look affected from the change of buffer composition. In these two purification systems the optimization of the pH of the binding buffer necessary for the correct binding of the DNA to the silica column was not possible, so they were excluded from the use in the protocol. Based on these findings, the choice has fallen on Qiagen PCR Purification Kit and AmpureXP: the former was used when multiple reactions of the samples were set up, and was not possible to use the AmpureXP due to the big amount of beads required for the purification.

Table 16: Five clean-up systems have been tested in terms of recovery rate and purity.

Product	**Microcon YM30**	**Wizard**	**Qiaquick**	**Zymo**	**AmpureXP**
Technology	Spin Procedure-Size exclusion	Spin Procedure	Spin Procedure	Spin Procedure	Solide-Phase-Reverse Immobilization
Type of matrix	Cellulose	Silica Membrane	Silica Membrane	Silica Membrane	Magnetic Beads
Time needed	25-30 min	5 min	5 min	5 min	10 min
Range Fragment sizes	137 bp - /	100 bp – 10 Kb	100 bp – 10 Kb	50 bp – 23Kb	100 bp – 10 Kb
Capacity per prep	N.A	40µg	10µg	25µg	N.A
DNA Quality	A260/260: 0,6-0,9	A260/260: 1,78	A260/260: 1,8	A260/260: 1,79	A260/260: 1,86
Recovery Performance:					
PCR (5µg)	4,8µg	4,4µg	4,7µg	4,7µg	**4,9µg**
Restriction/Ligation (5µg)	4,7µg	3,7µg	4,4µg	2,6µg	**4,85µg**

The human reference genome (hg19) has been digested in-silico with MseI to determine which percentage of each fraction would be lost during purification as a result of different molecular weight cut-off levels. The base pair distribution of each of the ten nucleotide combinations of MseI-fragments is summarized in Table 17.

Table 17: In-silico digestion of the human genome (hg19) with MseI enzyme. In the table are shown the base pair distribution of the ten MseI-fragments in percentage.

[%] per fraction		fractions									
	MseI-fragment length	A/A	T/T	C/C	G/G	AC/CA	AT/TA	AG/GA	CT/TC	CG/GC	GT/TG
Fragments Excluded	**0-5**	5,87	5,87	6,80	6,79	1,75	5,38	2,11	2,12	1,31	1,73
	6-10	6,11	6,10	4,33	4,38	5,47	5,98	5,65	5,67	4,80	5,52
	11-50	28,99	29,02	24,89	24,87	27,51	30,02	27,42	27,40	26,16	27,46
	51-70	9,04	9,00	8,53	8,60	9,28	9,17	9,39	9,40	8,97	9,34
		50,01	**49,99**	**44,55**	**44,64**	**44,01**	**50,55**	**44,57**	**44,59**	**41,24**	**44,05**
Fragments Included	**71-100**	10,33	10,33	10,19	10,19	10,95	10,83	11,50	11,48	10,74	10,91
	101-150	12,38	12,38	11,82	11,85	12,64	11,43	12,57	12,64	13,34	12,64
	151-300	16,21	16,22	19,07	19,03	18,22	16,16	17,68	17,63	18,52	18,23
	301-600	9,31	9,32	11,03	10,98	11,15	9,15	10,78	10,78	12,41	11,17
	601-1200	1,77	1,77	3,35	3,32	3,03	1,88	2,89	2,89	3,75	3,00
		49,99	**50,01**	**55,45**	**55,36**	**55,99**	**49,45**	**55,43**	**55,41**	**58,76**	**55,95**

Around 50% of the genomic MseI-fragments are in the range of 5-50bp and can therefore not be purified by available purification systems. Although the Zymokit would allow the recovery of fragments down to 50bp, the molecular weight cut off was set to a level of 100bp. The isolation of MseI-fragments in between 50-100bp would have raised the inherent problem of potential carry-over of primer-dimer and adaptor concatamer. MseI-fragments below 100bp

representative of 40-50% of the DNA material would be in theory excluded from the analysis, so the theoretical percentage of MseI-fragments that can be further analysed is between 50-60%.

4.2 WGA library modification: protocol establishment

WGA modifications required for method implementation involves re-amplification of the original WGA, the Lib1-adaptor removal and the introduction of secondary adaptors enabling secondary PCR. The establishment of the respective quality control assay has been an essential pre-requisite to ensure the proper development of the protocol and, in turn, to define quality checkpoints for each step of the procedure (Figure 24).

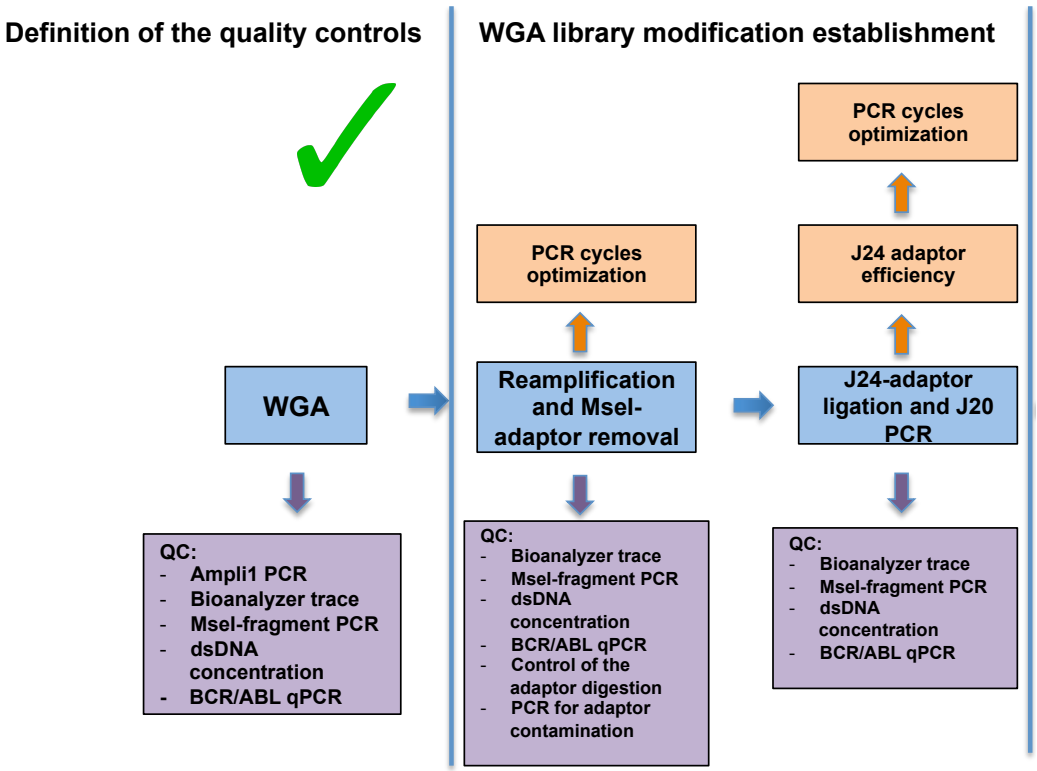


Figure 24: The diagram shows optimizations (orange boxes) and quality controls (violet boxes) applied at each step of the protocol development of the WGA library modification.

4.2.1 Re-amplification of the genomic MseI-population and Lib1-adaptor removal: optimization and quality controls

In order to obtain sufficient amounts of high-quality dsDNA the optimal number of PCR cycles for MseI-fragments re-amplification has been determined (Method Design Paragraph

3.2.1). The process of PCR cycle optimization is of fundamental importance as reaching a plateau effect would result in the generation of unwanted ssDNA by-product unusable in dsDNA-dependent enzymatic reactions. After conducting limited cycle PCRs with 10-12-15-17 cycles for re-amplification of the primary WGA product, the amount of dsDNA has been determined. After the 12th PCR cycle the dsDNA amount started to decrease (Table 18).

Table 18: Summary of the PCR cycles optimization. In bold are the optimal numbers of cycles and the corresponding amount of dsDNA obtained.

PCR Program	10 cycles	**12 cycles**	15 cycles	17 cycles	19 cycles
Re-amplification of the WGA	82ng/µl	**108ng/µl**	91ng/µl	83ng/µl	64ng/µl

Based on these results, the number of PCR cycles for re-amplification of the WGA product of the K562 single cells and the pool of lymphocytes (Table 19) has been limited to a total number of 12.

Table 19: DsDNA amounts obtained from the re-amplification of the samples after twelve PCR cycles.

Re-amplification of the WGA	10 cycles	**12 cycles**	15 cycles	17 cycles	19 cycles
Re-amplification WGA Lymphocytes pool	---	**114 ng/µl**	---	---	---
Re-amplification WGA K562 single cell	---	**102 ng/µl**	---	---	---

Following re-amplification and purification of the primary WGA, the size distribution of the MseI-fragment representation is controlled with the DNA High Sensitivity Chip on the Agilent Bioanalyzer. The fragment size distribution of the re-amplified and original WGA is still similar, although a small shift toward smaller fragment is observed, ranging from ~150 to ~3000bp (Fig. 25; Fig. 18) showing that limited re-amplification does not change drastically the main amplicons length composition.

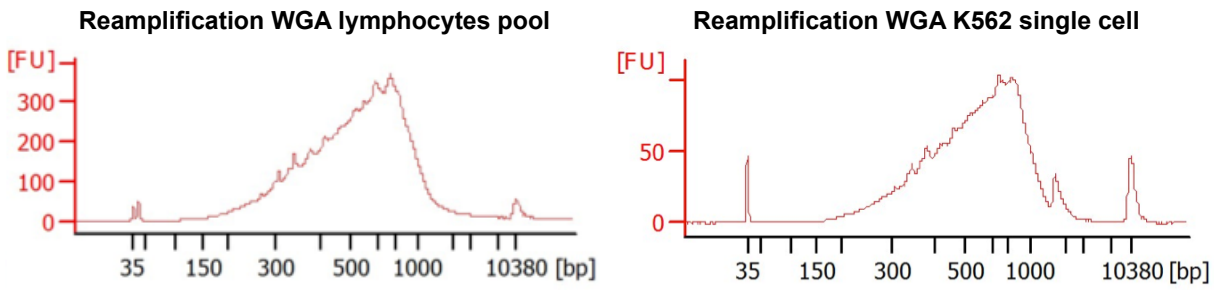


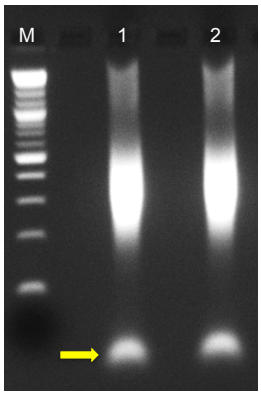
Figure 25: Examples traces of WGA re-amplification from lymphocyte pool and K562 single cell to control the size distribution and to avoid the selection of PCR amplicons. The X-axis represents the product size in base pairs (bp) and the Y-axis is the arbitrary fluorescence intensity (FU).

The absolute quantification using the BCR/ABL standard curve assay was applied to quantify the copy number of the BCR/ABL breakpoint-MseI-fragment in the K562 single cell after re-amplification. The re-amplification of the WGA did not increase the copies of BCR/ABL breakpoint-MseI-fragments as would be expected from a further PCR amplification. In table are summarized the qPCR results (Table 20).

Table 20: BCR/ABL breakpoint-MseI-fragments copies/ng of DNA calculated after the WGA re-amplification of the K562 single cell.

Sample for Absolute qPCR	BCR/ABL breakpoint-MseI-fragments copies/ng of DNA
Re-amplification K562 single cell	1597

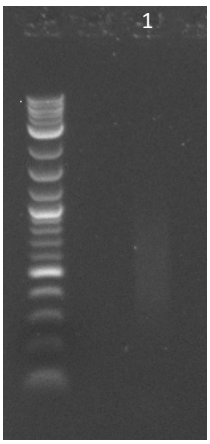
Subsequently, the re-amplified WGA was digested to release the original Lib1-adaptor and the digestion has been controlled on agarose gel to prove the effective removal of the Lib1-adaptor (Figure 26).



1: Lib1-adaptor removal Lymphocytes pool
2: Lib1-adaptor removal K562 single cell
M: Marker 2-log

Figure 26: Lib1-adaptor removal from re-amplified lymphocyte pool and K562 single cell. Digested and un-purified DNA has been controlled on agarose gel. The yellow arrow indicates the digested adaptor.

Moreover, to exclude adaptor contamination or minimal adaptor carry-over, 0,5 µl of the digested sample has been re-amplified with the Lib1 primer (Figure 27).



1: Re-amplification of Lib1-adaptor removed sample with Lib1-primer
M: 2-log molecular marker

Figure 27: Re-amplification of Lib1-adaptor removed sample with Lib1-primer to exclude adaptor carry-over. As can be observed a very light smear barely visible.

After controlling the proper digestion, the samples are purified for complete adaptor removal and dsDNA amount measured by Qubit/PicoGreen – Broad Range assay (Table 21).

Table 21: DsDNA concentration obtained after the Lib1-adaptor removal.

Samples for Lib1-adaptor removal	dsDNA concentration
Lib1-adaptor removal Lymphocytes pool	124 ng/µl
Lib1-adaptor removal K562 single cell	112 ng/µl

Additionally, the purified digestion product was analysed with the DNA High Sensitivity Chip on the Agilent Bioanalyzer to control the size distribution, which remains invariant compared to the re-amplification step (Figure 28).

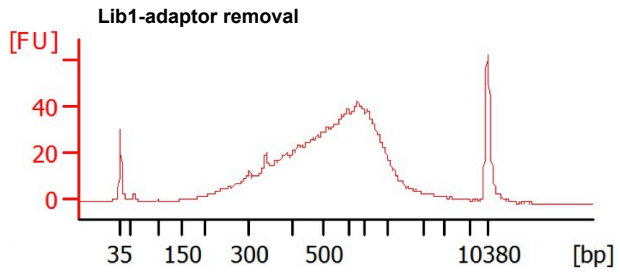


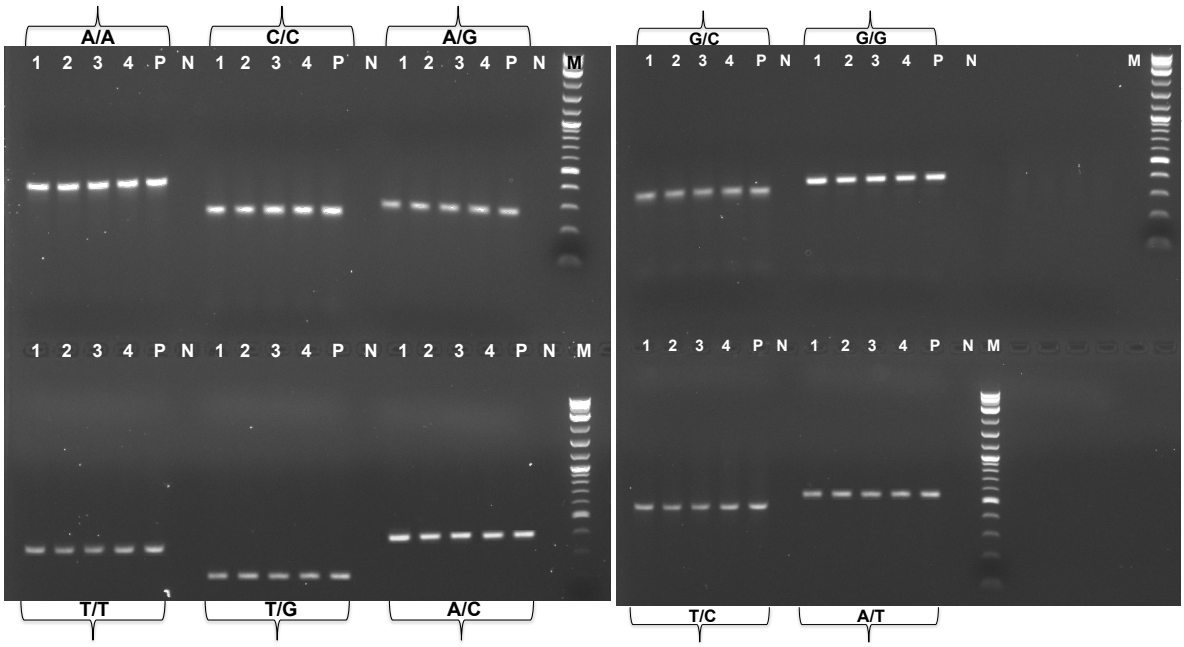
Figure 28: Examples trace of Lib1-adaptor removal after re-amplification. The X-axis represents the product size in base pairs (bp) and the Y-axis is the arbitrary fluorescence intensity (FU).

The established absolute quantification using the BCR/ABL standard curve assay qPCR was also used here to quantify the copy number of the BCR/ABL breakpoint-MseI-fragment in the K562 single cell after the removal of the Lib1-adaptor (Table 22). The digestion with MseI and the purification did not affect significantly the copy numbers of the BCR/ABL breakpoint-MseI-fragment compared to the re-amplification.

Table 22: BCR/ABL breakpoint-MseI-fragments copies/ng of DNA calculated after the digestion of the Lib1-adaptor.

Sample for Absolute qPCR	BCR/ABL breakpoint-MseI-fragments copies/ng of DNA
Lib1-adaptor removal K562 single cell	1325

The specific MseI-fragments PCR-based quality control was assayed on re-amplified and purified Lib1-adaptor digested lymphocyte pool and K562 single cell samples. Loss of MseI-fragments was not observed in any of the tested samples indicating that during these two steps all ten MseI-fragments combinations are preserved (Figure 29).



1: Lymphocyte pool – Re-amplification
2: K562 single cell – Re-amplification
3: Lymphocyte pool – Lib1 removal
4: K562 single cell – Lib1 removal
P: Positive control
N: Negative control
M: 2-log molecular marker

Primer ID	Differentiating Base	Lenght of the MseI Fragment	Lenght of the PCR fragment
TP53	A/A	558 bp	374 bp
ABL1	C/C	291 bp	273 bp
RTP3	AG/GA	394 bp	302 bp
BCR	T/T	1936 bp	413 bp
BCR	TG/GT	670 bp	192 bp
RH27788	AC/CA	576 bp	389 bp
BCR	GC/CG	1071 bp	258 bp
BCR	G/G	944 bp	390 bp
TP53	TC/CT	1032 bp	461 bp
BCR	AT/TA	790 bp	583 bp

Figure 29: Specific MseI-fragment control PCR. After re-amplification and Lib1-adaptor removal step, control PCR was performed and was positive in all of the samples. The specific MseI-fragment amplified is indicated in curly brackets at the edge of the gel. In the small table the control primers used for this step are listed.

4.2.2 Ligation of the secondary adaptors and secondary PCR: optimization and quality controls

After successfully implementing primary WGA re-amplification and Lib1-adaptor removal, the secondary adaptor ligation and amplification has been established (Method Design Paragraph 3.2.2). To determine the ligation efficiency of each of the four J24 adaptors, each adaptor was ligated individually to the sample, and then the dsDNA amount assessed for each reaction after J20-mediated re-amplification of the ligated samples (Table 23). In each of the four reactions, a similar amount of dsDNA was measured after re-amplification.

Table 23: dsDNA amount obtained after the ligation of each single adaptor with different terminal nucleotides to the DNA deprived of the Lib1-adaptors.

J24 adaptor used in combination to 100ng of MseI-digested DNA	dsDNA concentration
J24 adaptor-terminal nucleotide A (2µM)	62,0ng/µl
J24 adaptor-terminal nucleotide C (2µM)	63,4ng/µl
J24 adaptor-terminal nucleotide G (2µM)	61,7ng/µl
J24 adaptor-terminal nucleotide T (2µM)	62,6ng/µl

The MseI-fragment length distribution of each reaction product showed to be of similar length, except that the ligation and amplification of J24 adaptors with terminal nucleotide C and G gave a slight shift towards the amplification of bigger fragments. J24 adaptors with terminal nucleotide A and T gave an amplification peak at 481bp while J24 adaptors with terminal nucleotide C and G gave a peak at 485bp, so a total shift of 50bp (Figure 30).

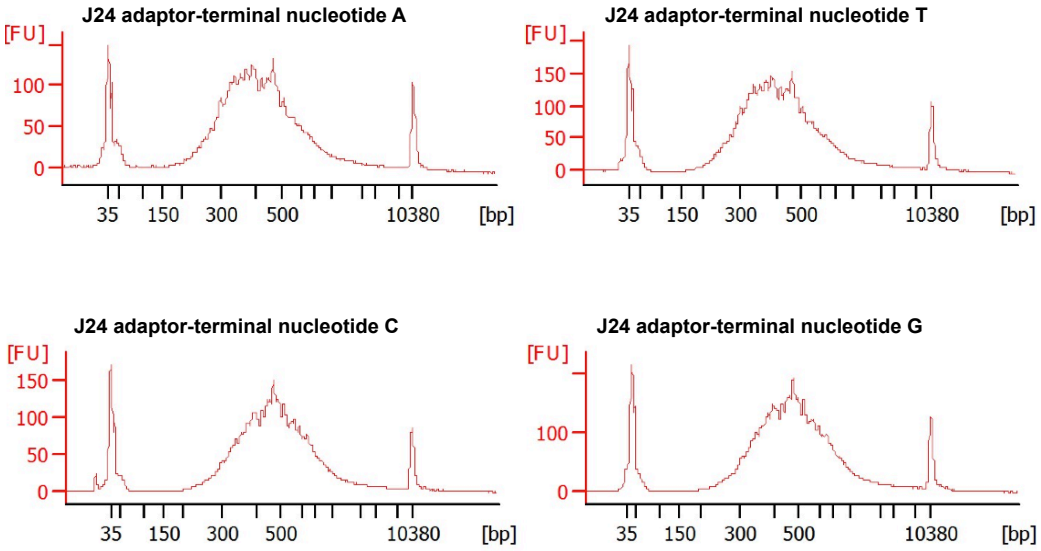


Figure 30: Bioanalyzer profile derived from the singular ligation and PCR amplification of the J24 adaptors with different terminal nucleotides. The amplification profile is very similar between them; a slight base pair shift towards bigger fragments can be noted in the amplification with the J24 adaptors with terminal nucleotide C and G. The X-axis represents the product size in base pairs (bp) and the Y-axis is the arbitrary fluorescence intensity (FU).

After the ligation of the secondary adaptors, the PCR cycles for the J20 PCR have been optimized (Table 24).

Table 24: Summary of the PCR cycles optimization for the J20 secondary PCR. In bold are the optimal numbers of cycles and the corresponding amount of dsDNA obtained.

PCR Program	10 cycles	12 cycles	**15 cycles**	17 cycles	19 cycles
J20 PCR	45ng/µl	51ng/µl	**67ng/µl**	59ng/µl	48ng/µl

Lymphocyte pool and K562 single cell have been amplified for 15 cycles followed by purification and dsDNA quantification (Table 25).

Table 25: dsDNA concentration obtained after the J20 secondary PCR for the secondary adaptor-ligated lymphocyte pool and K562 single cell.

J20 secondary PCR	10 cycles	12 cycles	**15 cycles**	17 cycles	19 cycles
J20 secondary PCR Lymphocytes pool	---	---	**74ng/µl**	---	---
J20 secondary PCR K562 single cells	---	---	**61ng/µl**	---	---

Subsequently, the size distribution of all secondary PCRs were analysed with the DNA High Sensitivity Chip on the Agilent Bioanalyzer. The J20 primary PCR products show a slight different profile ranging from smaller fragments (100bp) to bigger ones (3000bp) (Figure 31).

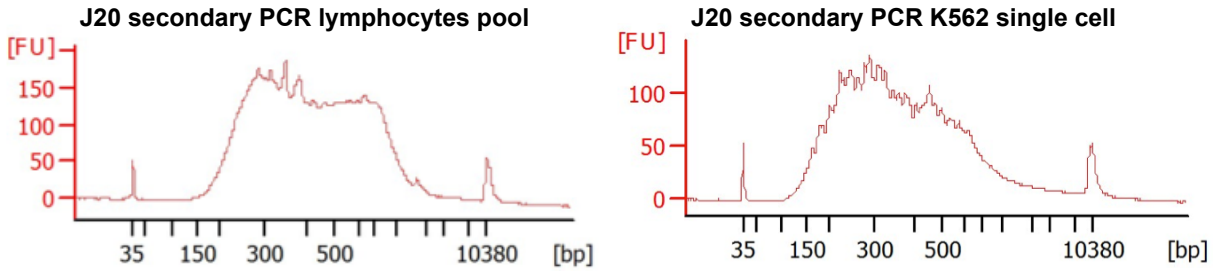


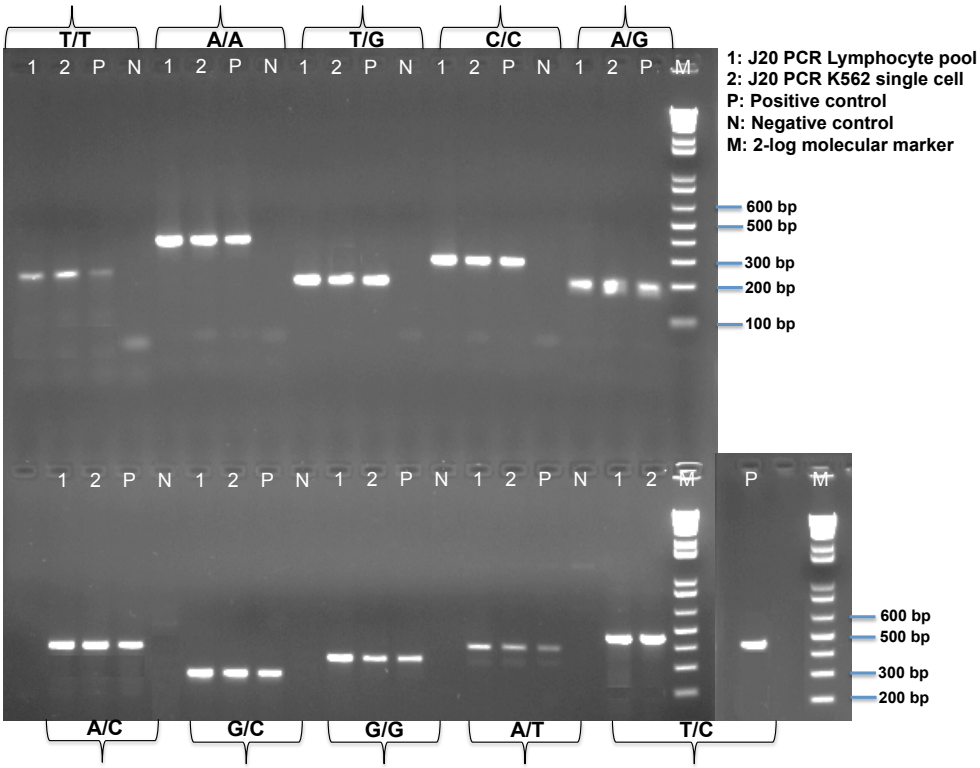
Figure 31: Examples traces of J20 secondary PCR derived from the samples to control the size distribution. The J20 PCR ranges from 100-3000bp. The X-axis represents the product size in base pairs (bp) and the Y-axis is the arbitrary fluorescence intensity (FU).

The absolute quantification qPCR assay was applied to quantify the copy numbers of the BCR/ABL breakpoint-MseI-fragment in the K562 single cells after the J20 secondary re-amplification PCR (Table 26). The copies/ng increased compared to the previous steps due to the effect of a secondary amplification.

Table 26: BCR/ABL breakpoint-MseI-fragments copies/ng of DNA calculated after the J20 secondary PCR.

Sample for Absolute qPCR	BCR/ABL breakpoint-MseI-fragments copies/ng of DNA
J20 secondary PCR K562 single cell	3463

Identical to Results Paragraph 4.2.1, the fraction-specific PCR-based quality controls were assayed on amplified and purified secondary re-amplification products of lymphocyte pool and K562 single cell samples. As experienced in the previous chapter, in none of the tested samples loss of MseI-fragments was noted indicating that during these two steps all ten MseI-fragments combinations are preserved (Figure 32).



Primer ID	Differentiating Base	Lenght of the MseI Fragment	Lenght of the PCR fragment
TP53	A/A	558 bp	374 bp
ABL1	C/C	291 bp	273 bp
D5S399	AG/GA	489 bp	178-201 bp
BCR	T/T	1936 bp	413 bp
BCR	TG/GT	670 bp	192 bp
RH27788	AC/CA	576 bp	389 bp
BCR	GC/CG	1071 bp	258 bp
BCR	G/G	944 bp	390 bp
TP53	TC/CT	1032 bp	461 bp
BCR	AT/TA	790 bp	583 bp

Figure 32: Specific MseI-fragment control PCR for the secondary amplification J20 PCR. The samples show the amplification for all the control primer pairs. The specific MseI-fragment amplified is indicated in curly brackets at the edge of the gel. In the small table the control primer used for this step are listed.

4.3 Genome fractioning: specific selection and enrichment of MseI-fragments

The major challenge of the genome fractioning is to efficiently separate each of the ten MseI-fragments populations thereby minimizing cross-contamination of undesired other fragments (Method Design Paragraph 3.3). The BCR/ABL breakpoint-MseI-fragment is contained within the MseI-fragment population characterized by G/G differentiating bases. This MseI-fragment population is used as the experimental test. Various attempts have been made to optimize the procedure and to shorten the length of the protocol. Figure 33 summarizes the different approaches and the quality controls applied to assess the successful implementation of the genome fractioning procedure.

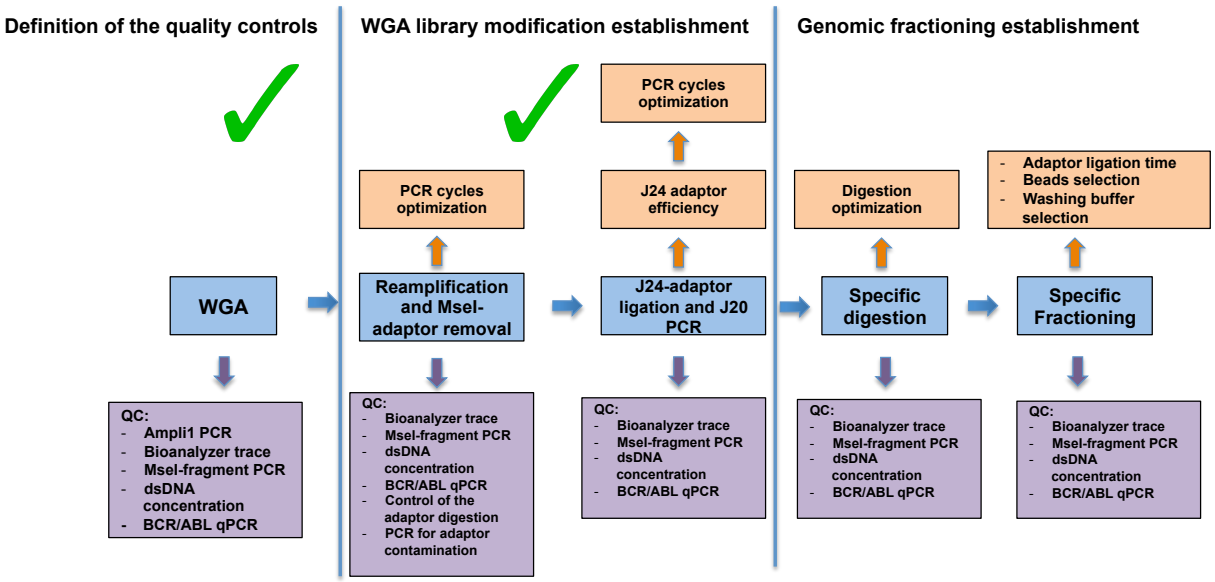


Figure 33: Diagram showing which optimizations (orange boxes) and quality controls (violet boxes) have been applied at each step of the protocol development of the genome fractioning.

4.3.1 Optimization of the genome fractioning procedure: specific digestion of MseI-fragments

Three main protocols have been investigated to optimize the genome fractioning (Figure 34) (Method Design Paragraph 3.3.1). At first, the specific digestion has been optimized to perform at high efficiency in the shortest time. Protocols A and B use a digestion time of 1 hour while protocol C uses only 15 minutes for the digestion.

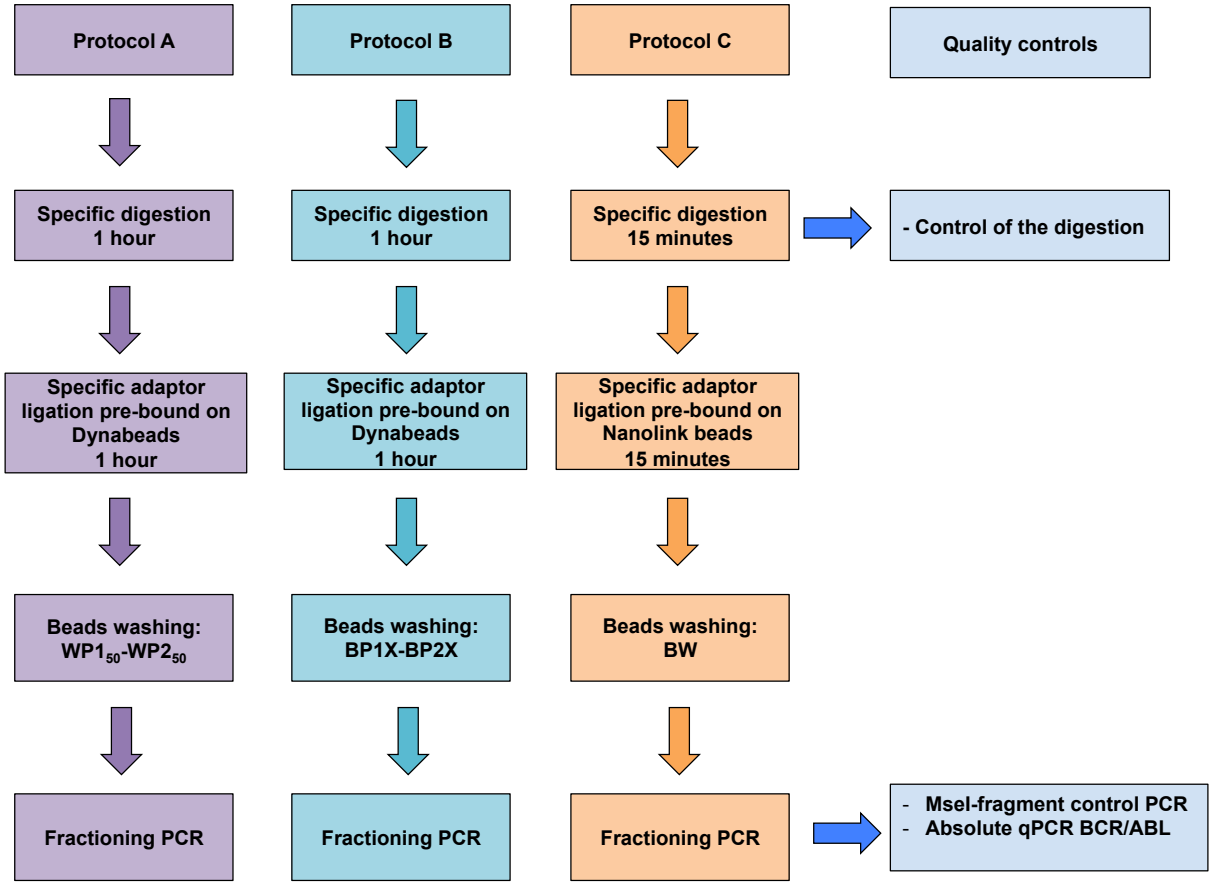


Figure 34: Different protocols tested for the establishment and optimization of the genome fractioning procedure.

The specific isolation of MseI-fragments constituting the G/G-fraction requires a BspTI digestion. Digested products were separated on an agarose gel to visualize the released J20-adaptors in all tested samples. The digestion was shown to be successful under both conditions (Figure 35).

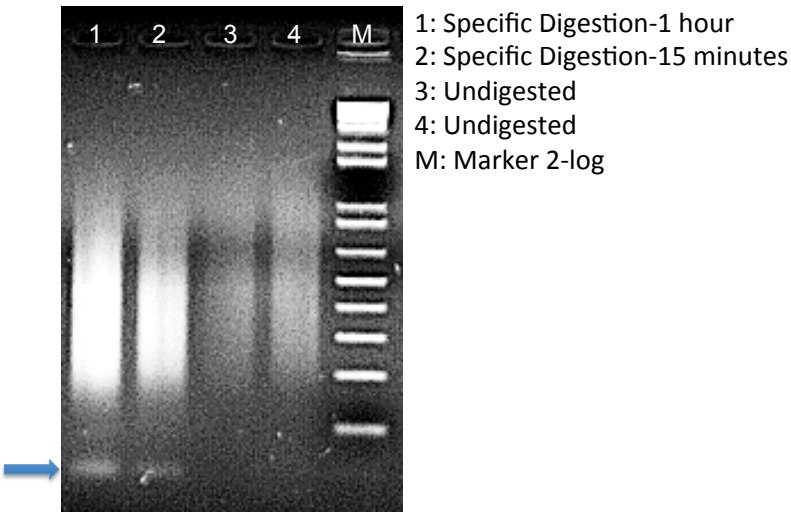


Figure 35: Control of the specific digestion on agarose gel 3%. The digested J20-adaptors are indicated from the blue arrow.

A control PCR to exclude carry-over of the MseI-fragments harbouring the J20-adaptor in the next step of the protocol has not been performed. In fact, at this step not all the J20-amplified MseI-fragments are digested by the specific enzyme, but only those harbouring the correct reformed restriction site. This means that most of the J20-amplified MseI-fragments will remain undigested.

4.3.2 Optimization of the genome fractioning procedure: beads selection, fractioning-adaptor ligation and washing buffer conditions

The ligation of the specific fractioning-adaptor to the digested MseI-fragments has been evaluated in order to shorten the incubation time (Method Design Paragraph 3.3.2). Two types of magnetic streptavidin-conjugated beads have been tested, the Dynabeads M-280 (Invitrogen, Inc.) and the Nanolink beads (Solulink Inc.). Biotinylated fractioning-adaptors pre-bound either to streptavidin-conjugated Dynabeads M-280 or Nanolink beads have been incubated with the specifically digested MseI-fragments generated in the previous Paragraph 4.3.1 (Figure 34).

The washing steps for the magnetic streptavidin-conjugated beads have been optimized in order to remove the majority of unbound and unspecific bound MseI-fragments whilst not applying too stringent washing procedure leading to a loss of specifically bound MseI-fragments. In protocol A very stringent washing steps are used: Dynabeads M-280 were rinsed with washing buffers containing 50% formamide, $WP1_{50}$-$WP2_{50}$, at high temperatures (50°C) four times for 4 minutes. In protocol B, one wash with BP2X and two washes with BP1X were used at room temperature for washing of Dynabeads M-280. In protocol C three washing steps at room temperature with BW buffer have been used to rinse the Nanolink beads (Figure 34). The final outcome of these different practical approaches is directly related to the results of the specific MseI-fragment PCR and qPCR for BCR/ABL (Result Paragraph 4.3.3).

4.3.3 Optimization of the genome fractioning procedure: generation of simple genomic fractions and quality controls

After implementation of the various washing procedures, the best-suited protocol was identified on the basis of purity of the generated fractions i.e. absence of cross-contaminating MseI-fragments after specific MseI-fragment control PCR, and the obtained enrichment of the

BCR/ABL breakpoint-MseI-fragment observed after absolute quantification by qPCR (Method Design Paragraph 3.3.3).

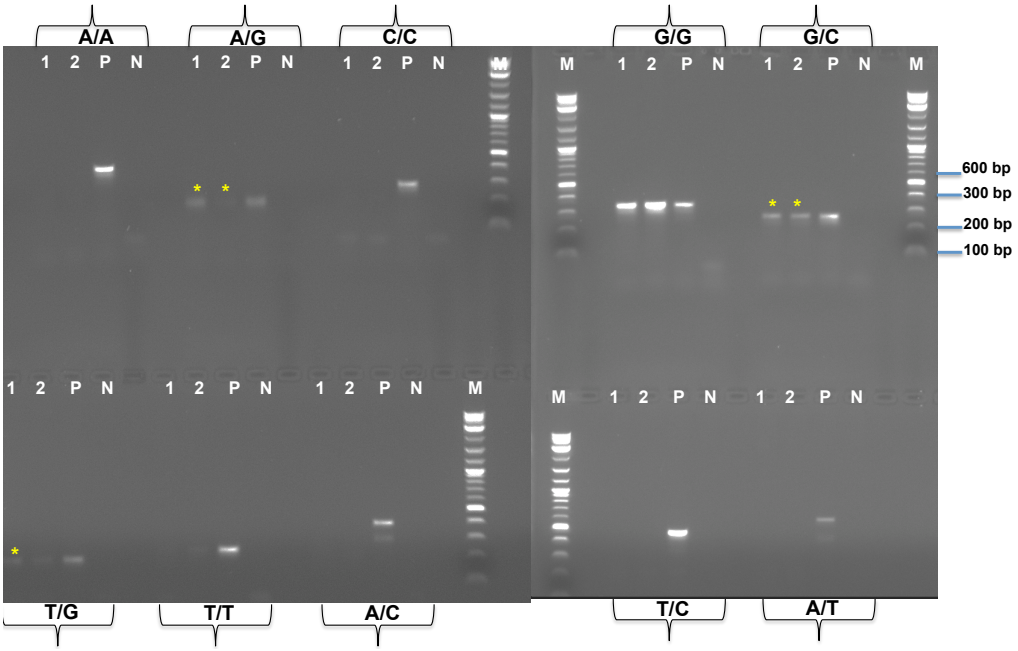
The specific MseI-fragment control PCR performed to ensure the proper MseI-fragment isolation for each experimental setting showed different outcome. The enrichment of the MseI-fragment population containing the differentiating bases G/G shows the lowest background of unwanted MseI-fragments amplified with the use of Nanolink beads and BW buffers (Figure 36 C) compared to if Dynabeads M-280 would be used (Figure 36 A-B). Under stringent washing conditions (Figure 36 A) the Dynabeads M-280 showed similar, but not considerably better results to the Nanolink beads. If Dynabeads M-280 are washed with BP2X-BP1X (Figure 36 B), they show a high background of unwanted amplified MseI-fragments as can be seen from their amplification after the specific genome fractioning.

The absolute quantification BCR/ABL after the genome fractioning with the use of Protocol A showed a decrease in BCR/ABL breakpoint-MseI-fragment copy numbers in the final genome MseI-fraction product compared to Protocol B and Protocol C. Protocol B and Protocol C performed equally good in terms of BCR/ABL breakpoint-MseI-fragment enrichment. Based on these results Protocol C is the most suitable to perform the genome fractioning of the K562 single cells (Table 27).

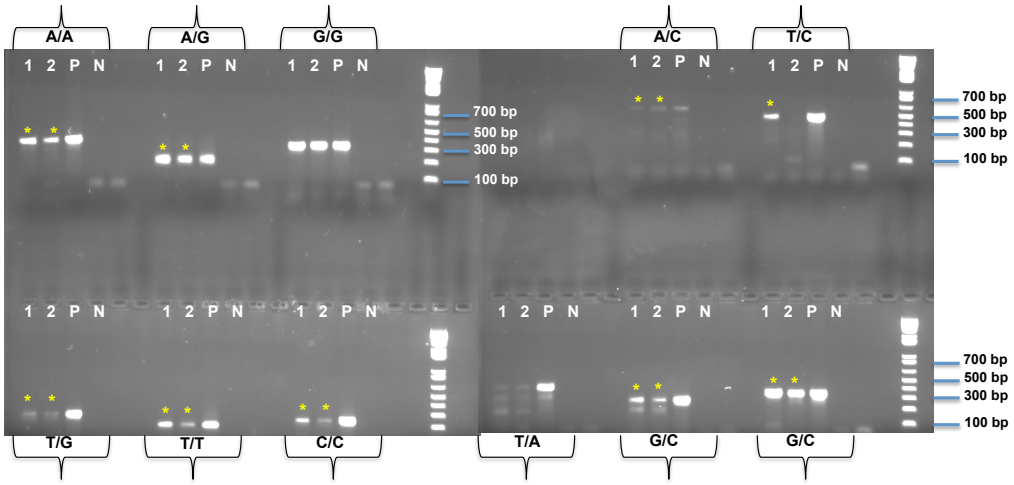
Table 27: Selection of the suitable protocol for genome fractioning based on the BCR/ABL breakpoint-MseI-fragment enrichment and purity of the specific MseI-control PCR.

Protocol	BCR/ABL breakpoint-MseI-fragment copies/ng	Purity of the MseI-fraction Specific MseI-fragment PCR	Optimal protocol
Protocol A	**32.165**	++	✖
Protocol B	**54.125**	-	✖
Protocol C	**52.935**	+++	✔

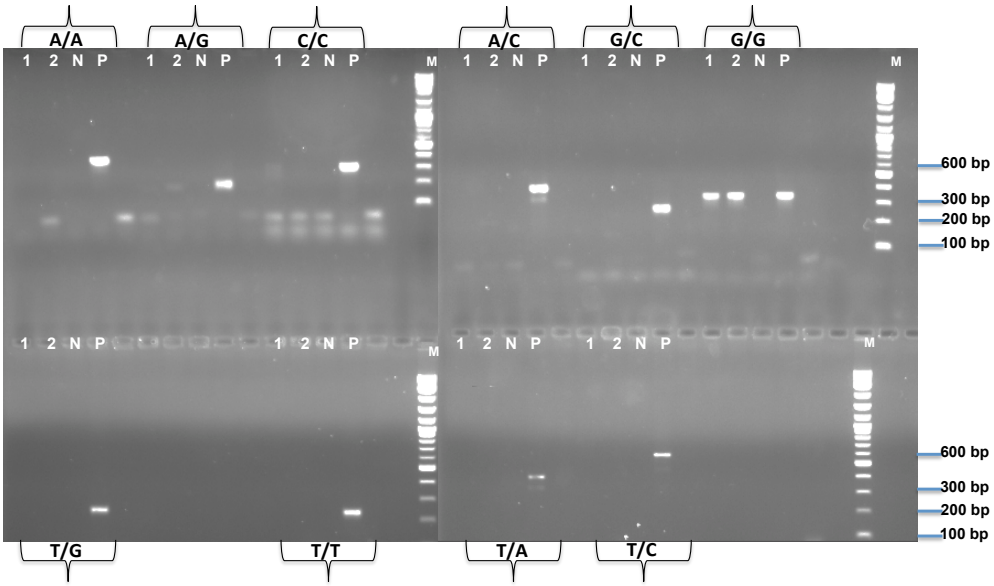
A) Genomic fractioning using Protocol A



B) Genomic fractioning using Protocol B



C) Genomic fractioning using Protocol C



Primer ID	Differentiating Base	Lenght of the MseI Fragment	Lenght of the PCR fragment
TP53	A/A	558 bp	374 bp
ABL1	C/C	291 bp	273 bp
D5S399	AG/GA	485 bp	178 bp
RTP3	AG/GA	394 bp	302 bp
BCR	T/T	508 bp	235 bp
BCR	TG/GT	670 bp	192 bp
RH27788	AC/CA	576 bp	389 bp
BCR	GC/CG	1071 bp	258 bp
ZNF23	GC/CG	417 bp	290 bp
BCR	G/G	715 bp	470 bp
TP53	TC/CT	1032 bp	461 bp
BCR	AT/TA	790 bp	583 bp
BCR	AT/TA	600 bp	358 bp

Figure 36: Specific MseI-fragment control PCR after genome fractioning using three different protocols. The isolated MseI-fragments carry a G/G nucleotide as differentiating bases. (A) Genome fractioning using protocol A; (B) genome fractioning using protocol B; (C) genome fractioning using protocol C. 1: genome fractioning of the lymphocyte pool, 2: genome fractioning of the K562 single cell; P: positive control; N: negative control. Yellow asterisks are indicating the MseI-fragments that are amplified after the specific genome fractioning, representing "contaminating" fragments. These are coming from the amplification of randomly bound fragments on the magnetic streptavidin-conjugated beads that are not properly washed away. The specific MseI-fragment amplified is indicated in curly brackets at the edge of the gel. In the table the control primers used are indicated.

After selecting Protocol C as the standard protocol for genome fractioning, quality controls have been applied to identify and define thresholds for specific steps of genome fractioning.

The specific digestion with BspTI of the J20 secondary PCR from lymphocytes pool and K562 single cells (from Paragraph 4.3.1) was quantified to identify the exact amount of dsDNA at this step of method development (Table 28).

Table 28: dsDNA concentration obtained after the specific enzyme digestion.

Samples for specific digestion	dsDNA concentration
Specific digestion Lymphocyte pool	141ng/µl
Specific digestion K562 single cell	135ng/µl

Furthermore, the size distribution of the specific digestion with BspTI enzyme, for G/G-MseI-fraction selection, was determined on the Agilent Bioanalyzer (Figure 37). The MseI-fragments length distribution at this step range from ~100bp till ~3000bp and it is very similar to the J20 secondary. Generally, the size distribution of the specific digestions did not indicate a visible or significant size shift after J20-adaptors removal, although the digestion was successful (Figure 35).

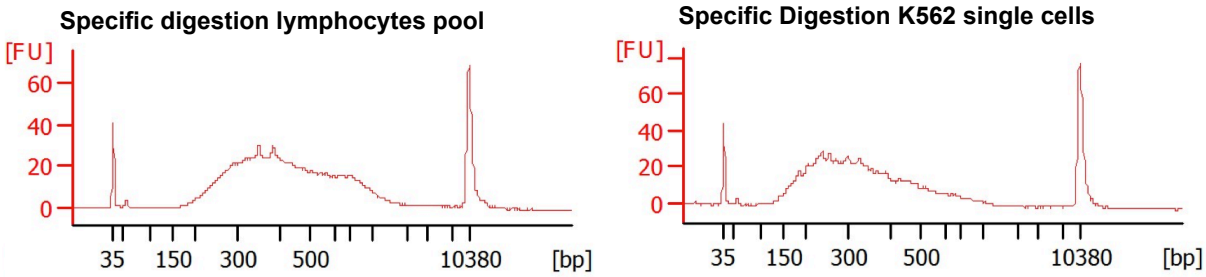


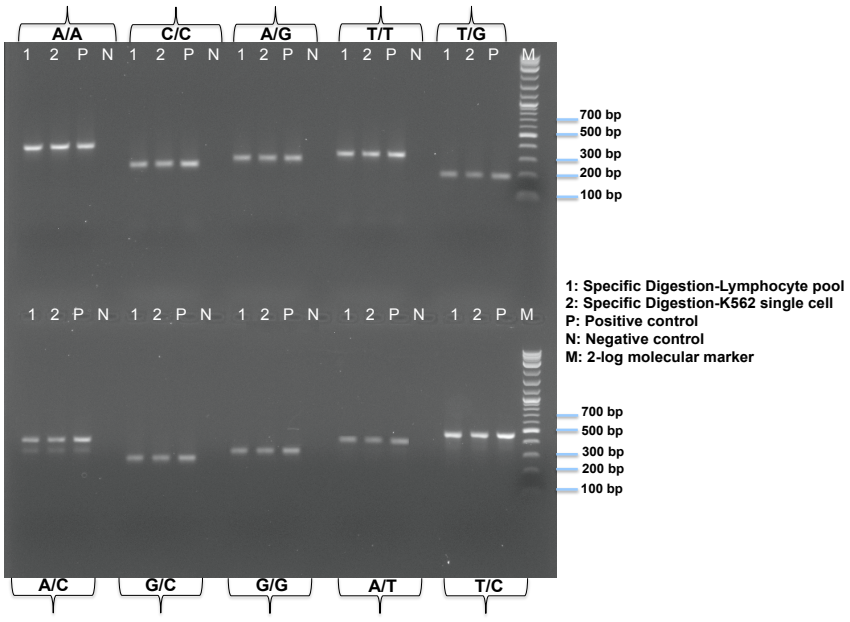
Figure 37: Example of bioanalyzer traces of specific digestion with BspTI enzyme derived from the analysed samples to control the size distribution. The MseI-fragments length distributions at this step range from ~100bp till ~3000bp. The X-axis represents the product size in base pairs (bp) and the Y-axis is the arbitrary fluorescence intensity (FU)

Finally, the absolute BCR/ABL qPCR assay was performed to determine the number of BCR/ABL breakpoint-MseI-fragment copies (Table 29). The copy numbers do not differ significantly from the previous step meaning that the specific digestion and purification do not induce a further loss of the BCR/ABL breakpoint-MseI-fragment.

Table 29: BCR/ABL breakpoint MseI-fragment copies/ng of DNA calculated after the BspTI specific digestion of the K562 single cell.

Sample for Absolute qPCR	BCR/ABL breakpoint-MseI-fragment copies/ng of DNA
Specific digestion K562 single cell	3299

The specific digestion represents the last step of the protocol in which all ten MseI-fragments populations are still represented in one sample: after this step each of the ten MseI-fragment populations will be specifically selected and individually enriched. The quality of the separation process is assayed with the specific MseI-fragments control PCRs to ensure the purity and unbiased representation of all ten MseI-fragment populations (Figure 38).



Primer ID	Differentiating Base	Lenght of the MseI Fragment	Lenght of the PCR fragment
TP53	A/A	558 bp	374 bp
ABL1	C/C	291 bp	273 bp
RTP3	AG/GA	394 bp	302 bp
BCR	T/T	508 bp	235 bp
BCR	TG/GT	670 bp	192 bp
RH27788	AC/CA	576 bp	389 bp
ZNF23	GC/CG	417 bp	290 bp
BCR	G/G	715 bp	470 bp
TP53	TC/CT	1032 bp	461 bp
BCR	AT/TA	600 bp	358 bp

Figure 38: Specific MseI-fragment control PCR assayed for the digestion with BspTI. The samples show the amplification for all the control primer pairs. The specific MseI-fragment amplified is indicated in curly brackets at the edge of the gel. In the table the control primers used are listed.

The final genome MseI-fraction, representing all the MseI-fragments containing the G/G nucleotide (Figure 36 C), has been quantified to assess the dsDNA concentration (Table 30).

Table 30: dsDNA concentration obtained at the end of the genome fractioning procedure.

Samples for Genome Fractioning	dsDNA concentration
Genome MseI-fraction Lymphocyte pool not-normalized	144ng/µl
Genome MseI-fraction K562 single cell not-normalized	151ng/µl

Moreover, the size distribution of the final genome MseI-fraction has been analysed on the Agilent Bioanalyzer. The digital profile generated from the lymphocyte pool and the K562 single cell showed that the genome fractioning could lead to a selective enrichment of certain amplicon sequences. This effect was more pronounced in the genome fractioning of the single K562 cell compared to the pool of lymphocytes (Figure 39).

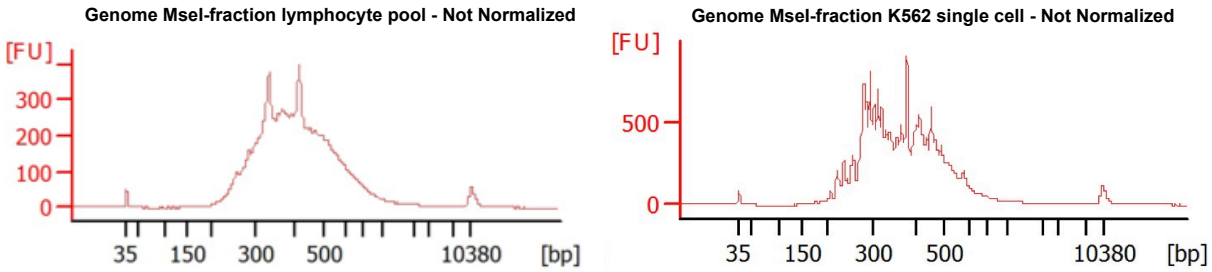


Figure 39: Examples of bioanalyzer profiles of the genome fractioning for the lymphocyte pool and the K562 single cell. The profiles show a MseI-fragment length distribution in between ~130 and ~1000bp with peaks selection mainly at 350-450bp.

To solve the problem related to the sequence selection at the end of the genome fractioning, the specifically enriched genome MseI-fraction population was normalized to obtain a homogenous distribution of the MseI-fragments length. The normalization showed to be very efficient in removing the over-represented DNA sequences (Figure 40).

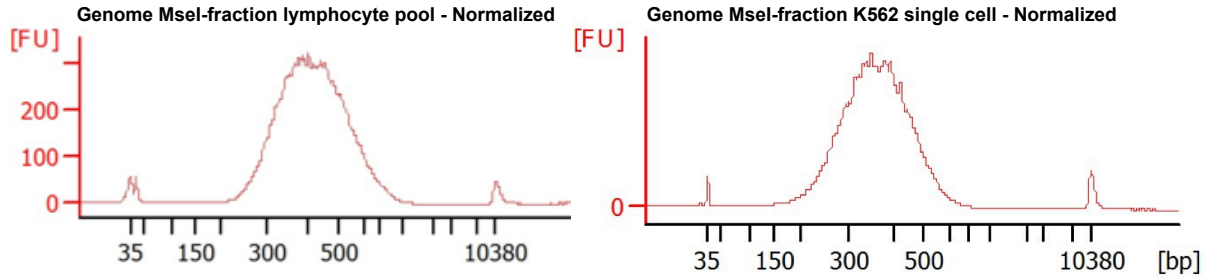


Figure 40: Examples of bioanalyzer traces of the normalized genome fractioning for the lymphocyte pool and the K562 single cell. The profile shows a homogenous distribution of the MseI-fragments. The normalization suppresses amplicons that are selected during the isolation/enrichment of the MseI-GG fraction, re-establishing a homogenous distribution in between ~130 and ~1000bp.

The dsDNA amount has been also quantified for the normalized genome MseI-fraction (Table 31).

Table 31: dsDNA concentration obtained after the normalization of the genome MseI-fraction.

Samples for Genome Fractioning	dsDNA concentration
Genome MseI-fraction Lymphocyte pool Normalized	195ng/µl
Genome MseI-fraction K562 single cell Normalized	167ng/µl

The absolute fold enrichment of the BCR/ABL breakpoint-MseI-fragment at each step of the developed method was determined, starting from the original WGA product until the specifically enriched MseI-fraction population. Direct comparison of the absolute copy numbers should enable to disclose critical steps affecting gain or loss of the desired sequence (Figure 41). The WGA of a single cell does not dramatically change the representation of the BCR/ABL breakpoint-MseI-fragment compared to the genome of a pool of cells, as the unamplified genomic DNA of the K562 cells and the WGA of single K562 single cells do not present significant differences in terms of absolute copy numbers. The WGA re-amplification of the K562 single cell did not show the expected PCR-based enrichment, instead a loss of the breakpoint MseI-fragment was noted. The removal of the Lib1-adaptor from the re-amplified sample showed a BCR/ABL breakpoint-MseI-fragment copy numbers similar to the re-amplified samples, proving the reliability of the purification method in recovering most of the MseI-fragments after the digestion. The secondary J20 PCR enriches again slightly the copy numbers of the BCR/ABL breakpoint-MseI-fragment, which subsequently remains roughly unchanged in the specific digestion for the MseI-fragments selection. The genome fractioning shows a significant enrichment of the breakpoint BCR/ABL breakpoint-MseI-fragment indicating that complexity reduction is the essential step in selecting a determined population of MseI-fragments. Compared to the initial WGA the enrichment after the genome fractioning

is 4.3 fold. If the genome MseI-fraction is normalized, there is a further substantial increase of copy numbers of the BCR/ABL breakpoint-MseI-fragment thus bringing the overall enrichment up to 13 fold (Figure 41).

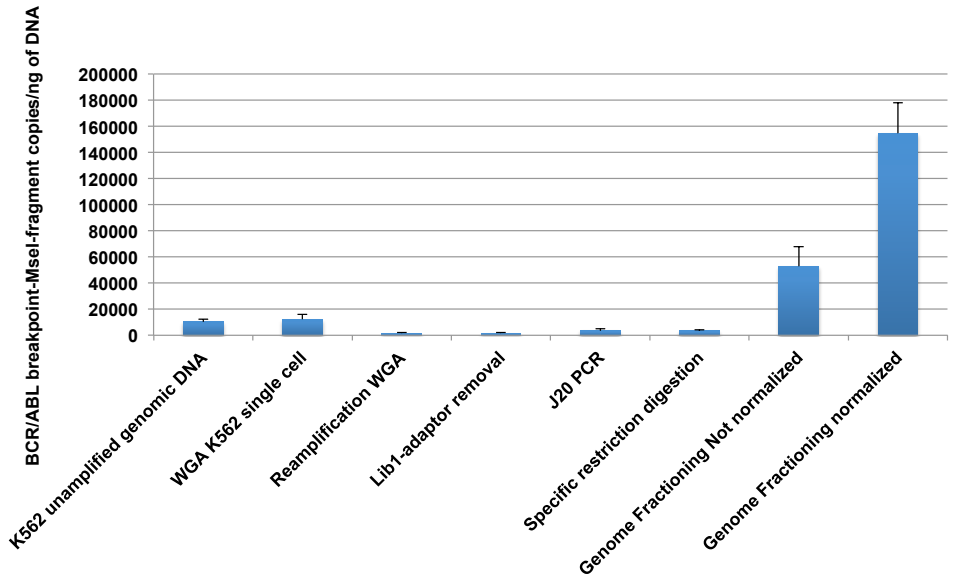
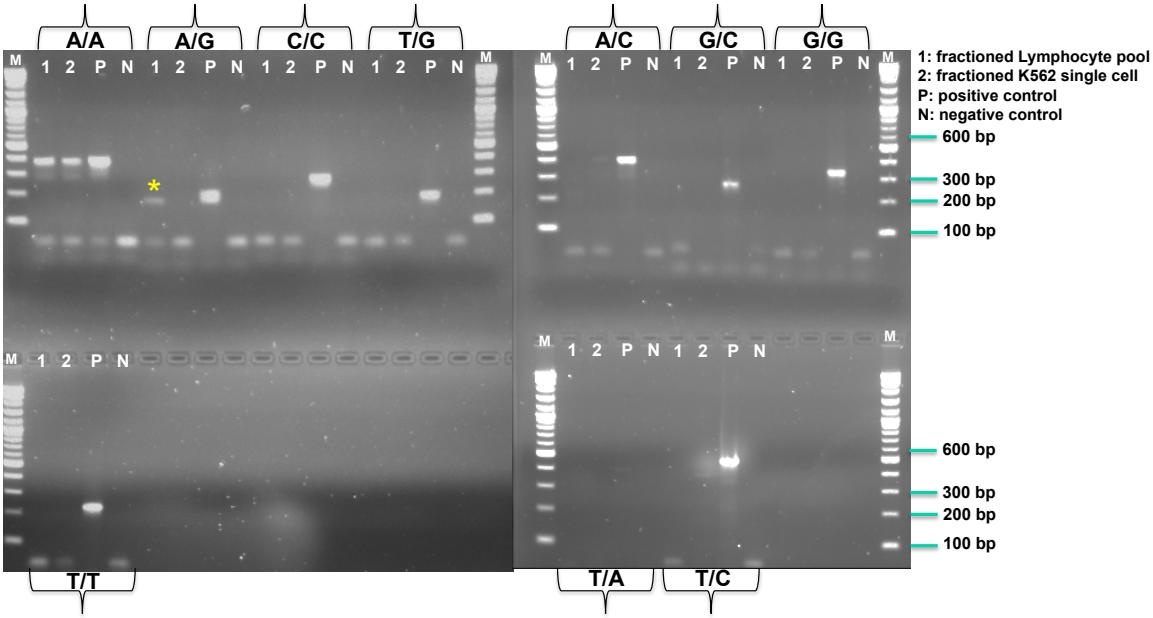


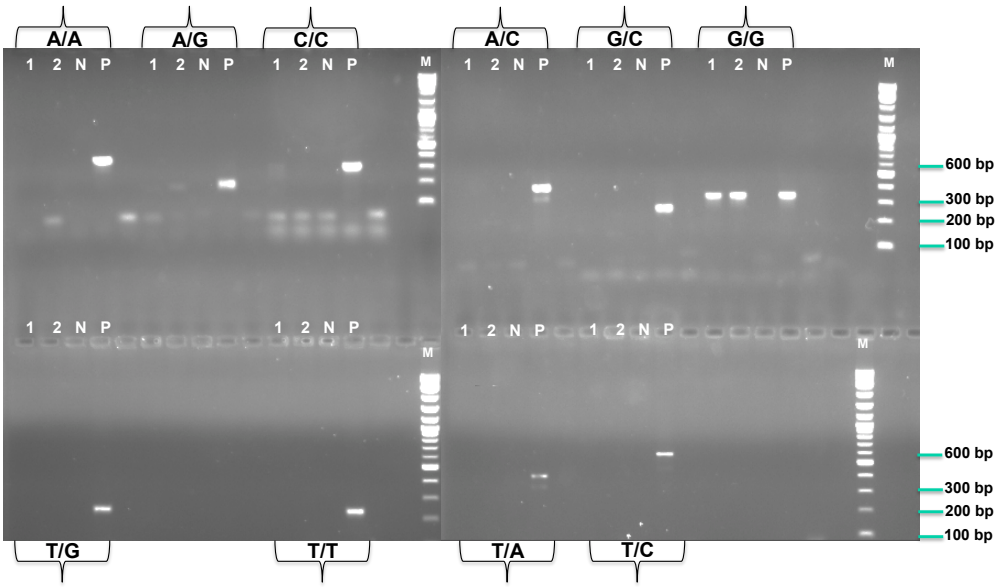
Figure 41: Overview of the K562 genome fractioning. On the Y-axis, are shown the copies of BCR/ABL breakpoint-MseI-fragment per nanogram of DNA of the K562 single cell sample and, on the X-axis, the step of the protocol controlled.

Based on the obtained results, the genome fractioning protocol was applied to the K562 single cell and the pool of lymphocytes to test the selective enrichment of the different MseI-fragments present in the genome. Figure 42 summarizes some examples of genomic fractioning for the selective isolation of MseI-fragments harboring simple nucleotide combinations assessed by applying the specific MseI-fragment control PCR. Figure 42 (A) shows the selective amplification of the genome MseI-fraction carrying the A/A nucleotide combination as differentiating base. Figure 42 (B) shows the isolation of the genome MseI-fraction harboring G/G nucleotide combination as differentiating bases. Figure 42 (C) shows the isolation of the genome MseI-fraction carrying the T/T nucleotide combination. The genomic fractioning is quite pure for each MseI-fraction isolated with minimal cross-contamination of other MseI-fragments.

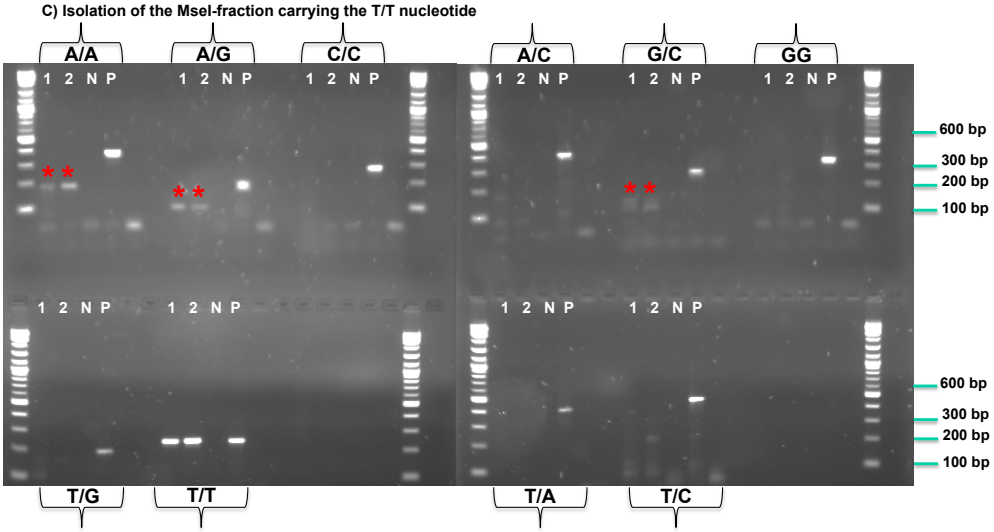
A) Isolation of the MseI-fraction carrying the A/A nucleotide



B) Isolation of the MseI-fraction carrying the G/G nucleotide



C) Isolation of the MseI-fraction carrying the T/T nucleotide

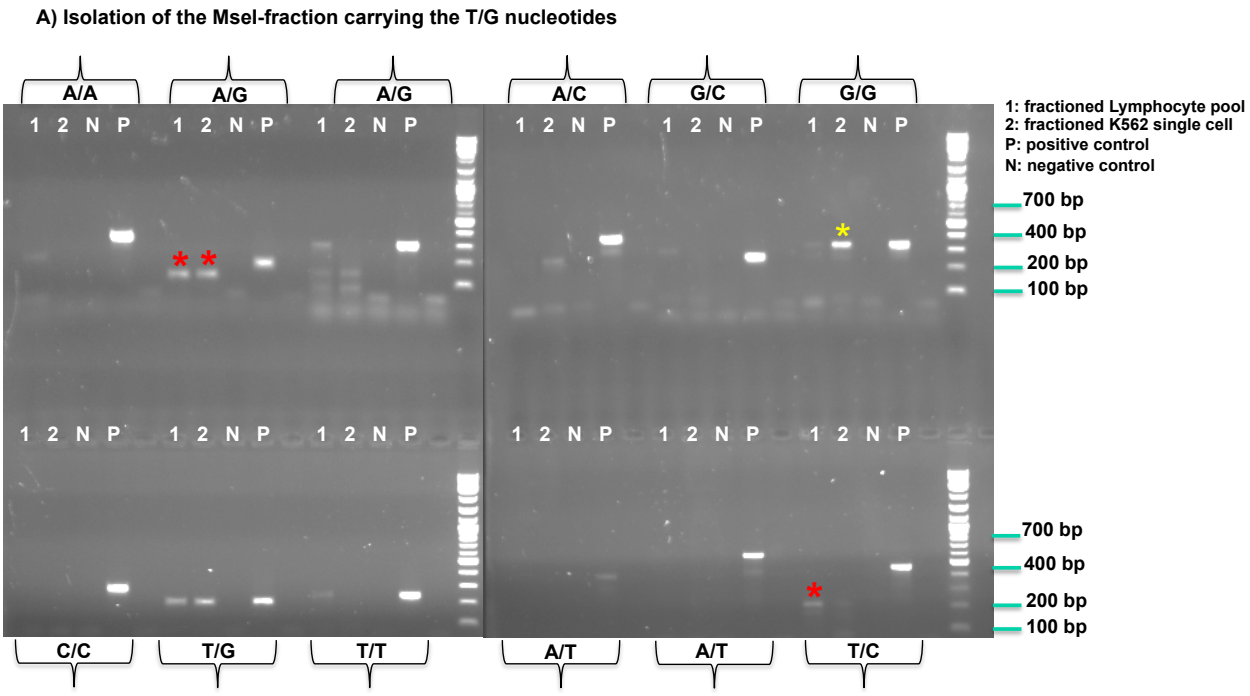


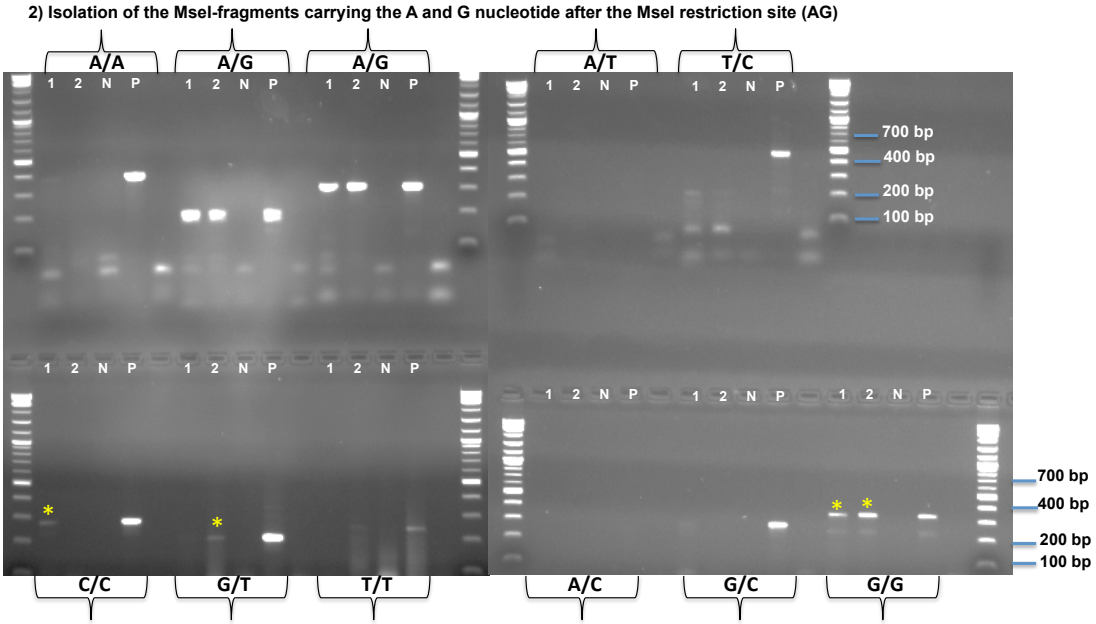
Primer ID	Differentiating Base	Lenght of the MseI Fragment	Lenght of the PCR fragment
TP53	A/A	558 bp	374 bp
ABL1	C/C	291 bp	273 bp
D5S399	AG/GA	485 bp	178 bp
RTP3	AG/GA	394 bp	302 bp
BCR	T/T	508 bp	235 bp
BCR	TG/GT	670 bp	192 bp
RH27788	AC/CA	576 bp	389 bp
BCR	GC/CG	1071 bp	258 bp
ZNF23	GC/CG	417 bp	290 bp
BCR	G/G	715 bp	470 bp
TP53	TC/CT	1032 bp	461 bp
BCR	AT/TA	790 bp	583 bp
BCR	AT/TA	600 bp	358 bp

Figure 42: Specific MseI-fragment control PCR after the genome fractioning of the lymphocyte pool and K562 single cells. (1) Isolation of the genome MseI-fraction carrying the A/A nucleotide after the MseI restriction site: only MseI-fragments with the same differentiating base A after the MseI recognition sites are enriched. Sometimes, small levels of "contamination" from other fragments can happen. For example in the fractioned genome of lymphocyte pool there is a tiny level of amplification of MseI-fragments with two different differentiating bases, in this case A/G (yellow asterisk); (2) genome MseI-fraction containing the G/G nucleotide, where is also located the BCR/ABL breakpoint-MseI-fragment of the cell line, is also quite clean; (3) genome MseI-fraction with the same differentiating base T/T after the MseI recognition sites are enriched; it is pure, except for some unspecific amplification (red asterisks, not corresponding to the size of the positive control). The specific MseI-fragment amplified is indicated in curly brackets at the edge of the gel. In the table the control primers used are indicated.

4.3.4. Examples of genome fractioning for mixed fractions

The genome fractioning for the isolation and enrichment of MseI-fragments harboring two different nucleotides after the MseI restriction sites (two differentiating bases) has been performed applying the same protocol optimized for the simple MseI-fragments (Method Design Paragraph 3.3.4). Figure 43 depicts two examples of genome fractioning controlled by the use of specific MseI-fragment control PCR.





Primer ID	Differentiating Base	Lenght of the MseI Fragment	Lenght of the PCR fragment
TP53	A/A	558 bp	374 bp
ABL1	C/C	291 bp	273 bp
D5S399	AG/GA	485 bp	178 bp
RTP3	AG/GA	394 bp	302 bp
BCR	T/T	508 bp	235 bp
BCR	TG/GT	670 bp	192 bp
RH27788	AC/CA	576 bp	389 bp
BCR	GC/CG	1071 bp	258 bp
ZNF23	GC/CG	417 bp	290 bp
BCR	G/G	715 bp	470 bp
TP53	TC/CT	1032 bp	461 bp
BCR	AT/TA	790 bp	583 bp
BCR	AT/TA	600 bp	358 bp

Figure 43: Specific MseI-fragment PCR after the genome fractioning of the lymphocyte pool and K562 single cells for MseI-fragments carrying different nucleotides. (1) Selection and enrichment of the MseI-fragments carrying the T/G nucleotides after the MseI restriction site. (2) Selection and enrichment of the MseI-fragments carrying the A/G nucleotide after the MseI restriction site. Yellow asterisks are denoting the contamination from different MseI-fragments that are amplified after the specific enrichment of a single MseI-fragments population; the red asterisks are indicating unspecific amplification of MseI-fragments not corresponding to the specific loci amplified. The specific MseI-fragment amplified is indicated in curly brackets at the edge of the gel. In the table the control primers used are indicated.

4.3.5. Summary of the WGA, WGA library modification and genome fractioning procedure applied to the K562-Lymphocyte pool

Here it is shown a summary table of the whole procedure necessary to generate enriched genome MseI-fraction and the quality controls necessary for the correct control of the method (Table 32).

Table 32: Summary of the quality controls applied at each step of protocol development. The qPCR for BCR/ABL breakpoint MseI-fragment has been applied only to K562 cells. N/A: not assessed.

Sample	Ampli1 WGA control PCR	Bioanalyzer Profile	Specific MseI-fragment PCR (amplified markers)	qPCR BCR/ABL breakpoint-MseI-fragment
WGA	4 markers	250-3000 bp	75 markers out of 75 markers	12.215 copies/ng of DNA
Re-amplification	N/A	150-3000 bp	10 markers out of 75 markers	1.597 copies/ng of DNA
Lib1-adaptor removal	N/A	150-3000 bp	10 markers out of 10 markers	1.325 copies/ng of DNA
J20 PCR	N/A	100-3000 bp	10 markers out of 10 markers	3.463 copies/ng of DNA
Specific digestion	N/A	100-3000 bp	10 markers out of 10 markers	3.299 copies/ng of DNA
Genome MseI-fraction not normalized	N/A	130-1000 bp	1 marker out of 10 markers	52.934 copies/ng of DNA
Genome MseI-fraction normalized	N/A	130-1000 bp	1 marker out of 10 markers	154.833 copies/ng of DNA

The GII of the WGA has been assessed with Ampli1 WGA control PCR to identify samples of high quality. Each step of the protocol has been controlled for the MseI-fragment length distribution to monitor possible changes of the original WGA fragment length profile. As observed on the bioanalyzer, the MseI-fragment population remains almost invariant until the generation of the genome MseI-fraction that sets the size distribution to 130-1000bp. Specific MseI-fragment PCR was performed to assess the completeness of the different MseI-fragment population during the protocol development. A total of 75 markers covering all ten different nucleotide combinations of MseI-fragments were established on the initial WGA population, however, for the successive steps of the protocol 10 markers, one for each representative MseI-fragment, have been used. The specific control MseI-fragment PCR has shown that the complete population of MseI-fragments is preserved until the digestion with the specific enzyme necessary to pre-select the MseI-fragment population. The final genome MseI-fraction shows the amplification of a single marker for a single MseI-fragment population. The assessment of the BCR/ABL breakpoint-MseI-fragment step after step indicates that the genome fractioning procedure, in addition to select specific MseI-fragment populations, pre-enriches the breakpoint.

4.4 Evaluation of the optimal Subtractive DNA Hybridization procedure

4.4.1 PCR parameters for the re-amplification of the genome MseI-fraction

The genome fractioning containing the enriched MseI-fraction need to be re-amplified to obtain sufficient amounts of Driver DNA (derived from lymphocyte pool) and Tester DNA (derived from K562 single cell) necessary for subtractive DNA hybridization (Method Design Paragraph 3.4.1). The re-amplification program has been optimized at 12 cycles (Table 33).

Table 33: Optimization of the re-amplification protocol for the genome fractioning. dsDNA amount obtained after different PCR cycles are indicated.

PCR Program	10 cycles	**12 cycles**	15 cycles	17 cycles	19 cycles
Re-Amplification Genome MseI-Fraction	123ng/µl	**132ng/µl**	127ng/µl	117ng/µl	98ng/µl

4.4.2 DNA precipitation before the subtractive hybridization

For the joint precipitation of the Driver and Tester DNA samples prior to the actual subtraction the classical phenol-chloroform extraction (Method Design Paragraph 3.4.2) was replaced by the use of the Phase Lock Gel Column (5 Prime, Inc.). Table 34 compares the DNA recovery results of the two extraction methods after the precipitation of 40µg of DNA (the amount usually used for the precipitation of the Driver before the subtraction).

Table 34: Optimization of the phenol-chloroform extraction protocol. Comparisons between the classical phenol-chloroform extraction and the commercial Phase Lock Gel. As can be seen after the phenol-chloroform extraction of 40µg of DNA with Phase Lock Gel almost the all DNA could be recovered, while with the classical method of extraction there was a higher risk of DNA loss.

40µg of DNA	Phase Lock Gel	Classic Phenol-chloroform
Recovery (µg)	38.5 µg	25µg

4.4.3 Testing the binding of high amounts of biotinylated MseI-fragments

In order to evaluate a subtractive hybridization approach using high amounts of biotinylated Driver DNA, it is necessary to test the complete removal of the biotinylated genome MseI-fraction by the use of magnetic streptavidin-conjugated beads (Method Design Paragraph 3.4.3). In order to fulfill the objective, the biotinylation of the Driver DNA sample is necessary and is achieved by the use of biotinylated fractioning primers and a dNTPs mixture containing biotin-14-dCTP. The experimental steps are summarized in Table 35.

Table 35: Generation of biotinylated DNA Driver. The genome MseI-fraction has been re-amplified with biotinylated primers and biotin-14-dCTP nucleotides according to the optimized PCR program. The product has been purified and quantified.

	Experimental details	Volume
PCR Reaction	Genomic fractioned DNA from the Driver sample (100ng/µl)	1µl
	Buffer 1 (10X)	5µl
	10X dNTP Mixture: [1 mM biotin-14-dCTP, 1 mM dCTP, 2mM dATP, 2 mM dGTP, 2 mM dTTP in 10 mM Tris-HCl (pH 7.5), 1 mM Na2EDTA]	5µl
	BSA 20mg/ml	1,25µl
	Specific fractioning primers (10µM)	5µl
	Polymerase Mix	1µl
	H2O	31,75µl
Purification	AmpureXP beads used according to the manufacturer	1,8X
Quantification	Qubit dsDNA assay used according to the manufacturer	

Ten micrograms of the biotinylated Driver DNA sample have been incubated with different amounts of magnetic streptavidin-conjugated beads to investigate the removal efficiency of biotinylated Driver molecules. Subsequently, the unbound dsDNA amount of the Driver DNA sample that remained in solution has been quantified (Table 36). Irrespective of the input amount of magnetic streptavidin-conjugated beads, it was always observed an incomplete removal of the biotinylated Driver DNA: up to 25% is not binding to the magnetic streptavidin-conjugated beads. Due to the substantial difficulties in the removal of high amount of biotinylated MseI-fragments it was excluded the use of biotinylated Driver DNA for the subtractive hybridization.

Table 36: Amounts of magnetic streptavidin-conjugated beads used to test the efficient removal of biotinylated Driver.

Streptavidin beads capture	Amount of beads used for the capture of *10µg of biotinylated DNA*	Bound Driver	Unbound Driver
	5µl	4.2µg	5.8µg
	10µl	5.3µg	4.7µg
	20µl	6.4µg	3.6µg
	30µl	6.8µg	3.2µg
	40µl	7.1µg	2.9µg
	50µl	7.7µg	2.3µg
	60µl	7.5µg	2.5µg
	100µl	7.6µg	2.4µg

4.4.4 Comparison of different methods for the re-hybridization of nucleic acids

Subtractive DNA hybridization performed in not optimal conditions can be a limiting factor for the proper enrichment of differential sequences, in this case of the BCR/ABL breakpoint-MseI-fragment (Method Design Paragraph 3.4.4). Therefore, it became necessary to test different re-hybridization methods with the aim of achieving the best re-association. In order to choose a subtraction method that could be better adapted to the established genome fractioning approach, four different experimental systems were set up:

a) Phenol emulsion re-association technique (PERT) allowing the re-association of DNA in a water-salt solution containing phenol 9% by using thermal cycler oscillation between 65° and 25° C,

b) Phenol emulsion re-association technique (PERT) allowing the re-association of DNA by using strong agitation of the DNA in a water-salt solution containing phenol 9%,

c) Re-association in small volume solution with 50% formamide,

d) Re-association n small volume solution with high salt concentration (1M).

The precise experimental settings of the four methods are summarized in the Table 37.

Table 37: Summary of the experimental conditions used for testing the different re-hybridization methods.

Reassociation Method	Experimental details
a) PERT thermal cycler	➢ Resuspend DNA in 45µl of Tris-HCl 10mM ➢ Heat the sample at 95°C for 10 min to denature the DNA duplexes ➢ Add 45µl of 2M NaCl (the acqueous component should be mixed first) and 10µl Ultrapure phenol buffer 90% (added last). ➢ Quickly chill the reaction on ice to form an emulsion in the PERT reaction ➢ Place the reaction in a thermocycler and cycle 15 sec at 65°C and 15 sec at 25°C for 24 hours
b) PERT agitation	➢ Resuspend DNA in 45µl of Tris-HCl 10mM ➢ Heat the sample at 95°C for 10 min to denature the DNA duplexes ➢ Add 45µl of 2M NaCl (the acqueous component should be mixed first) and 10µl Ultrapure phenol buffer 90% (added last). ➢ Quickly chill the reaction on ice to form an emulsion in the PERT reaction ➢ Vortex with intensive stirring (ca. 2000 oscillations/min) for 24 hours at room temperature to maintain the emulsion
c) In solution-formamide	➢ Resuspend DNA in 4,5µl of high grade water ➢ Add 10,5µl of hybridization buffer (10mM Tris-Hcl, pH 8.5, 50mM HEPES pH 8.9, 50% deionized formamide) ➢ Heat the sample at 95°C for 2 min to denature the DNA duplexes ➢ Add 2 µl of NaCl 5M ➢ Incubate in a thermocycler at 47°C for 24 hours
d) In solution-high salt	➢ Resuspend DNA in 6 µl of hybridization buffer (50 mM HEPES, 0.2 mM EDTA, pH 8.9) ➢ Cover the hybridization mixture with one drop of mineral oil ➢ Heat the sample at 95°C for 10 min to denature the DNA duplexes ➢ Add 1 µl of 5M NaCl under the oil ➢ Incubate in a thermocycler at 80°C with a decrease of 1°C per minute till 68°C for 24 hours

In order to assess the re-hybridization of the four approaches, known DNA amounts of genome MseI-fraction dsDNA have been spiked with known copy numbers of the BCR/ABL

breakpoint-MseI-fragment ligated to the specific BioR21-adaptor, used in subtractive DNA hybridization (Table 38).

Table 38: Genome MseI-fraction amount and copies number of BCR/ABL breakpoint-MseI-fragment spiked-in used for the testing the re-hybridization.

Genome MseI-Fraction DNA	Copy number of BCR/ABL breakpoint-MseI-fragment	BCR/ABL breakpoint-MseI-fragment amount
40 µg	4.000.000	1040 fg

The copy numbers of the BCR/ABL breakpoint-MseI-fragment have been calculated according the following formula: number of copies = (amount (ng) * $6.022x10^{23}$) / (length (bp) * $1x10^{9}$ * 650 Daltons). The length of the BCR/ABL breakpoint-MseI-fragment including the ligated BioR21-adaptor is 242bp. 4.000.000 copies spiked into 40µg of the fractioned genomic dsDNA sample represents 0,000000026% of the total dsDNA amount (Figure 44). This corresponds to far less (~3.5 fold) than the DNA amount of the BCR/ABL breakpoint-MseI-fragments at the onset of the subtractive DNA hybridization, as determined by the BCR/ABL absolute qPCR assay on the enriched genome MseI-fraction at the end of the genome fractioning (Figure 41).

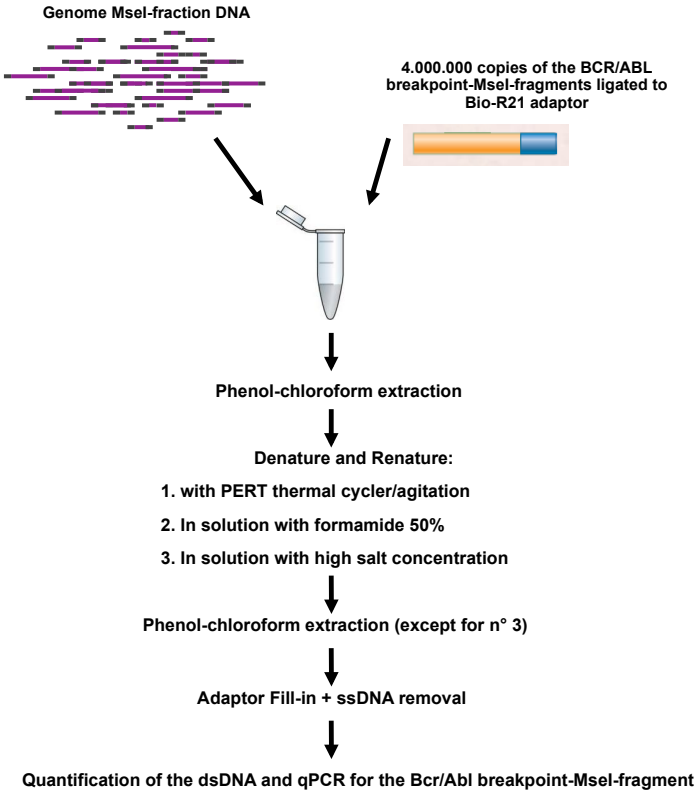


Figure 44: Experimental overview depicting the four approaches to test the best re-hybridization method.

After the spiking, the DNA has been extracted by phenol-chloroform and the recovered DNA re-hybridized with the four different techniques. The re-hybridized DNA has been extracted again to remove contaminants from the hybridization mixture such as the phenol contained in

the PERT and the 50% formamide in the small volume assay. The BioR21-adaptors have then been filled-in to render the MseI-fragments sequences amplifiable in the BCR/ABL qPCR assay and then the sample was treated with mung bean nuclease to remove non-hybridized ssDNA molecules. Enrichment PCR was not carried out to eliminate any subsequent PCR-introduced shifts in the representation of the BCR/ABL breakpoint-MseI-fragment in the re-hybridized samples. The quantified amount of dsDNA after the re-hybridization indicated that PERT in thermal cycler, PERT in agitation and re-hybridization in high-salt solution gave similar results with a maximum DNA re-association rate of 85-90%. The re-association in solution with formamide gave the lowest re-hybridization of dsDNA of around 50% (Figure 45).

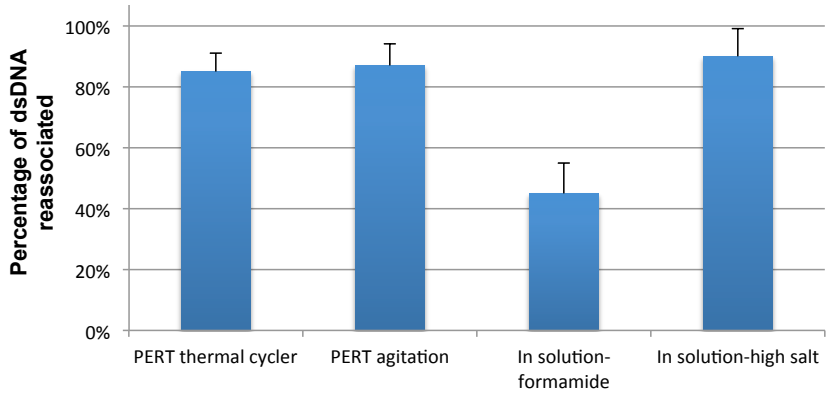


Figure 45: Percentage of dsDNA that re-associate after using four different re-hybridization methods.

The absolute BCR/ABL qPCR quantification assay was performed to determine the amount of re-hybridized BCR/ABL breakpoint-MseI-fragments, either with the R21 primer complementary to the BioR21-adaptor and with a primer pair specifically designed to amplify the breakpoint BCR/ABL sequence, the K562 breakpoint-primer. To ensure that no single-strand BCR/ABL breakpoint-MseI-fragment will be the target for amplification in the qPCR assay, a ssDNA removal step with mung bean nuclease following the re-hybridization was performed. The absolute quantification of the re-associated BCR/ABL breakpoint-MseI-fragment indicated similar re-association rates (Figure 46) irrespective of whether the R21 primer or the specific K562 breakpoint-primer pair was used (with slightly higher value for the K562 breakpoint-primer). The re-hybridization in small volume solution with 50% formamide showed 20-25% rate of re-association for the BCR/ABL breakpoint-MseI-fragment (Figure 46), while PERT in thermal cycler/agitation and in high-salt solution showed the highest re-association rate of about 85%.

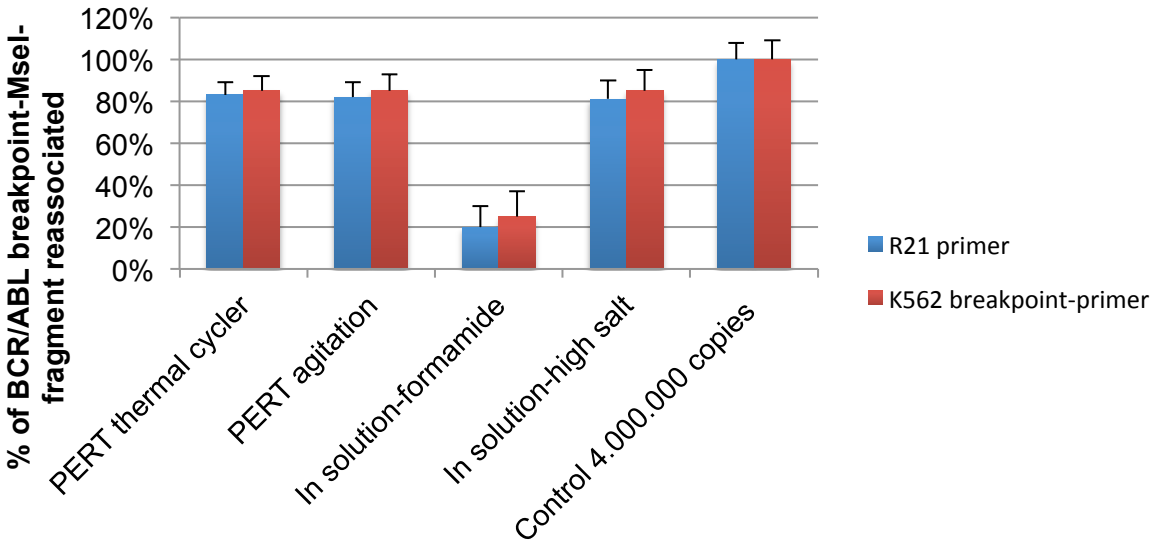


Figure 46: Percentage of BCR/ABL breakpoint-MseI-fragment that re-associate after using four different re-hybridization methods. The Cp cycle derived from the amplification of 4.000.000 copies of BCR/ABL breakpoint-MseI-fragment has been set as the default value for 100% re-association of the fragment.

The amplification products generated with the specific K562 breakpoint-primer and with the R21 primer were investigated by agarose gel electrophoresis. The amplification with the specific K562 breakpoint-primer showed specific amplicons for all four hybridization methods (Figure 47 A). The amplification with R21 primer showed specific amplicons too except for the re-hybridization with 50% formamide where unspecific background amplification was observed. Regarding the re-hybridization in high-salt solution the amplification with R21 primer showed to generate very intense amplicons maybe reflecting a higher degree of dsDNA re-association, although there was no obvious correlation to the calculated degree of re-hybridized total DNA (Figure 47 B). Additionally, in order to prove that these results are solely based on the proper re-hybridization of the specific BCR/ABL breakpoint-MseI-fragment sequences, an identical experiment of re-hybridization was performed without carrying out the fill-in reaction for the BioR21-adaptor. As expected, no specific amplification product for the BCR/ABL MseI-fragment was visible after PCR with R21 primer, but only the smear of re-hybridized MseI-fragments DNA (Figure 47 C).

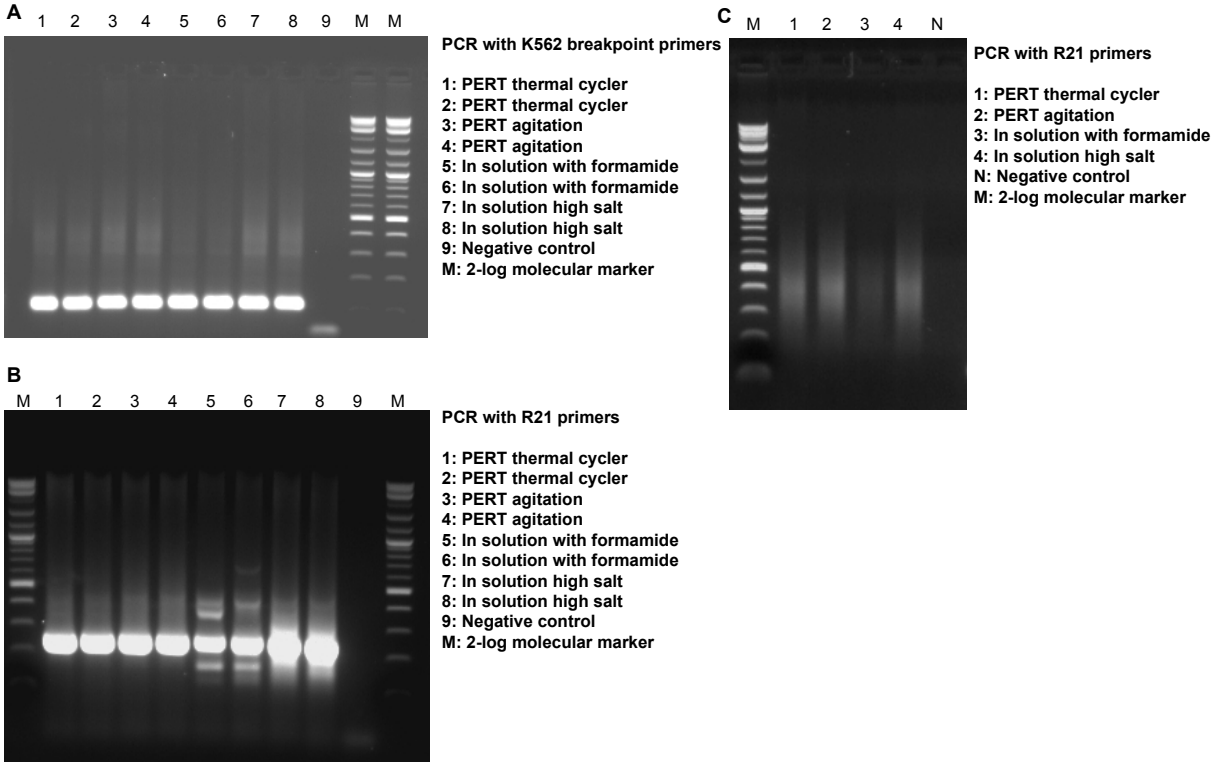


Figure 47: Control PCR performed after the re-hybridization with the different methods; A) PCR for the BCR/ABL breakpoint-MseI-fragment using the K562 breakpoint-primers showing the specific amplicon size; B) PCR for the BCR/ABL breakpoint-MseI-fragment using the R21 primers showing the specific amplicon size. The PCR with the R21 primer gave specific amplification product except for the re-association in solution with 50% formamide, in which unspecific fragments were amplified; C) PCR of the negative control for the re-hybridization using the R21 primers. The negative control consists in re-associated DNA in which the fill-in of the BioR21-adaptor has not been performed after the re-hybridization. As can be seen there is no amplification of the BCR/ABL breakpoint MseI-fragment by using the R21 primers.

The agarose gel electrophoresis results were also confirmed from the melting curve analysis for the amplification with the R21 primer (Figure 48 A) and the specific K562 breakpoint-primer (Figure 48 B).

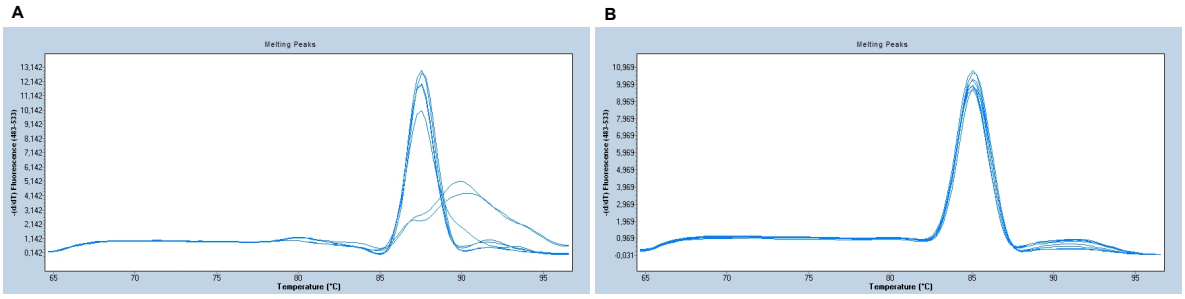


Figure 48: Melting peak analysis indicating the specificity of the amplification of the BCR/ABL breakpoint-MseI-fragment after the re-hybridization. (A) Melting peak generated after the amplification with R21 primers: can be observed unspecific amplification of the BCR/ABL MseI-fragment after re-hybridization in buffer containing formamide. (B) Melting peak generated after the amplification with K562 breakpoint-primers.

Based on the obtained results the following considerations were decisive in selecting the best-suited method: the re-association in small volume solution with 50% formamide was excluded

because of the lowest re-association rate. Of the remaining three methods presenting similar results, the classical re-hybridization in solution with high-salt concentration was preferred for two reasons: general avoidance of the use of phenol and the exclusion of the additional phenol extraction step after the re-hybridization to prevent further loss of DNA material.

4.4.5 Proof of principle of the re-hybridization and detection limit of the technique

In order to show that the selected re-hybridization method is effective in identifying multiple differential MseI-fragment sequences in low copies, a pGEM-T Easy vector containing the BCR/ABL breakpoint-MseI-fragment has been spiked in the genome MseI-fraction of a Driver sample derived from the normal lymphocytes pool (Method Design Paragraph 3.4.5). Before spiking to the genome MseI-fraction, the plasmid was digested with MseI enzyme, thereby producing a known pattern of fragments, including the BCR/ABL breakpoint-MseI-fragment. The MseI-fragments generated from the digestion were ligated to the BioR21-adaptor. The experimental workflow is illustrated in Figure 49.

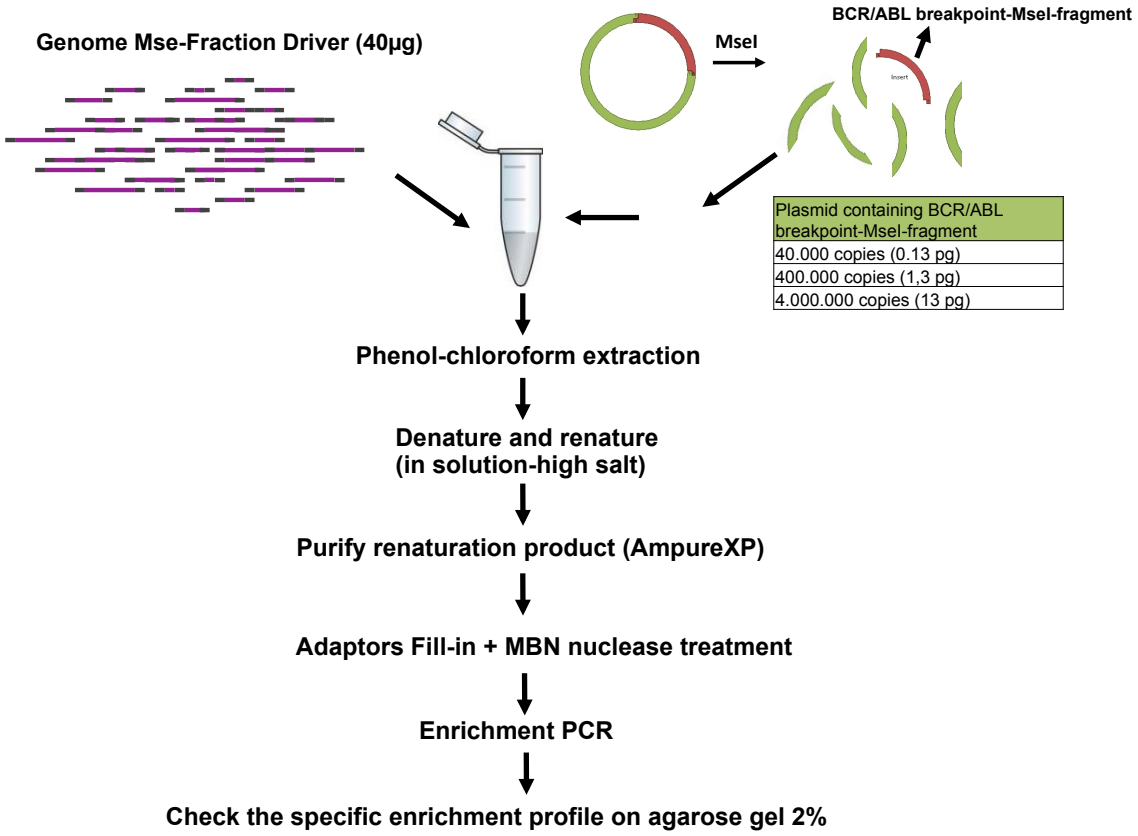


Figure 49: Experimental setting used to show the enrichment of differential MseI-fragments spiked in a genome MseI-fraction of Driver DNA.

After the spiking, the DNA sample is extracted with phenol-chloroform-isoamylalcohol and precipitated in ethanol over night at -20°C. The recovered pellet is re-suspended in the hybridization buffer, denatured and re-hybridized according to the selected protocol. The re-associated sample is subjected to enrichment PCR with R21 primer and the amplification product is analysed by agarose gel electrophoresis (Figure 50). The fragment patterns visible after the enrichment PCR is identical to the fragment patterns of the MseI-digested plasmid, including the insert BCR/ABL breakpoint-MseI-fragment. Fragments below 200bp were barely represented, maybe due to a preferential renaturation or amplifications of amplicons between 200bp-1000bp. Moreover, it was not evident any unspecific amplified MseI-fragments indicating a possible improper re-hybridization. This clearly shows that the selected method of re-hybridization is useful in terms of the identification of differential sequences present at small quantities.

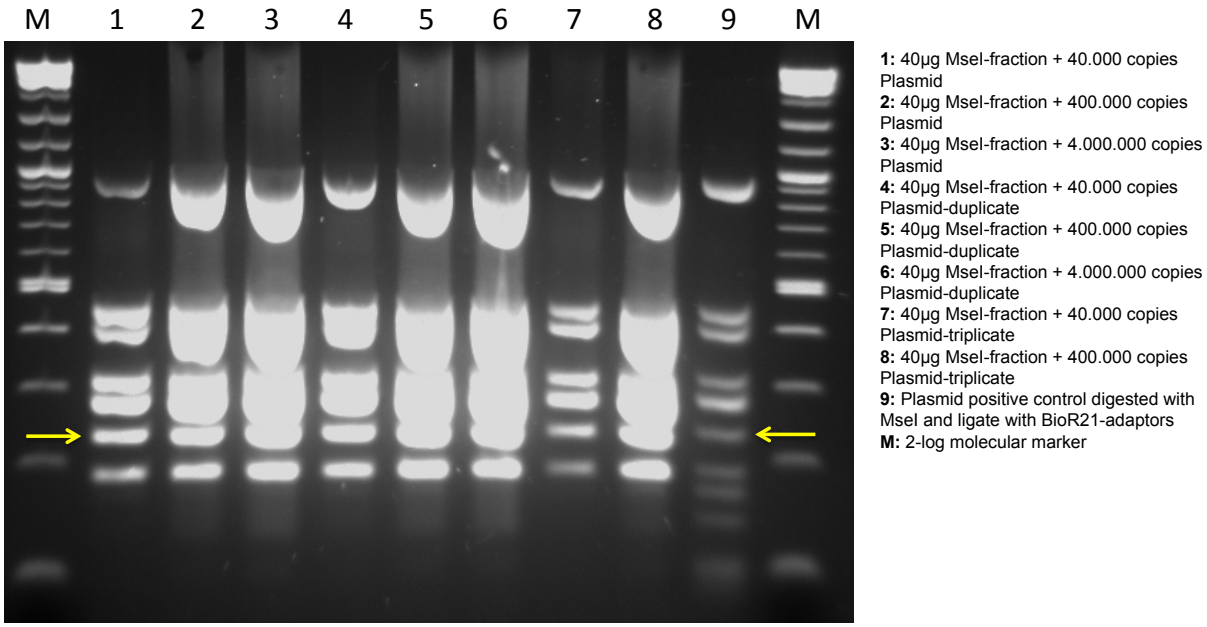


Figure 50: Re-hybridization and enrichment of the MseI-fragments, derived from the MseI digestion of the pGEM-T Easy vector containing the BCR/ABL breakpoint-MseI-fragment as insert, spiked in 40µg of genome MseI-fraction DNA. The DNA re-naturation of the plasmid is successful and shows the enrichment of almost all the MseI-fragments, including the BCR/ABL breakpoint-MseI-fragment (yellow arrows). Lanes 2, 3, 5, 6, 8 are overloaded due to the higher amount of plasmid spiked-in; lanes 1, 4, 7 represents clear distinction of the fragments.

4.5 Subtractive DNA hybridization of the genome fractioned K562 cell: depletion product 1 (DP1)

The first round of subtractive hybridization (Method Design Paragraph 3.5) has been performed with different starting amounts of Tester DNA, from 5ng to 1µg, in order to see when the maximum level of enrichment is reached, that means a further increase of the initial Tester amounts would not result in an increase of the BCR/ABL breakpoint-MseI-fragment copy numbers (overview of the experimental design in Figure 51). For simplicity, the DPs will always be named after the amount of Tester DNA (for the first round) and amount of DP (for the subsequent round) that has been used, the subtraction round where they have been generated and if the normalization has been applied or not.

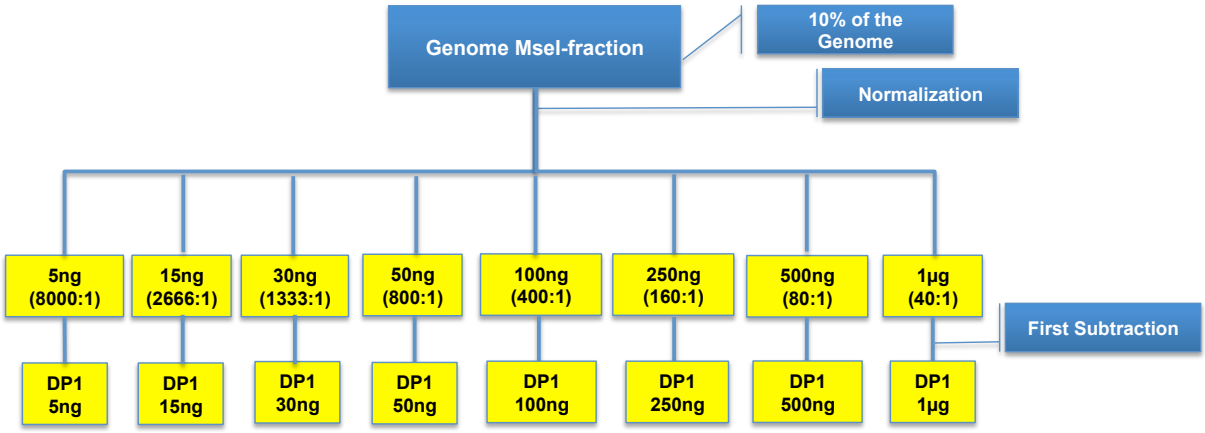


Figure 51: Overview of the subtraction scheme for generating the DP1. Different amount of normalized genome MseI-fraction of the Tester have been mixed with an excess of genome MseI-fraction of the Driver (40µg). The DP1s are the results of the subtraction of different amount of Tester to evaluate which of the titrations give the best enrichment of the BCR/ABL breakpoint-MseI-fragment. The Driver to tester ratio is specified.

After the first round of subtraction the absolute copy numbers of the BCR/ABL breakpoint-MseI-fragment was assessed (Figure 52). The Tester DNA amount of 50ng has been subtracted in duplicate (two separate experiments) to evaluate the reproducibility of the assay: very similar results in the enrichment of the BCR/ABL breakpoint-MseI-fragment copy numbers were noted for the 50ng duplicate (with a variability between the two experiment of 10%). The increasing starting amounts of Tester DNA lead to a substantial progressive enrichment of the BCR/ABL breakpoint-MseI-fragment. The enrichment is not anymore substantial between 500ng and 1µg of input amount of Tester DNA. For this reason, it appears that the maximum enrichment for DP1 can be obtained using Tester DNA amounts in the range of 500ng-1µg (22 fold), setting a favourable kinetic enrichment at Driver to tester ratio of 80:1-40:1.

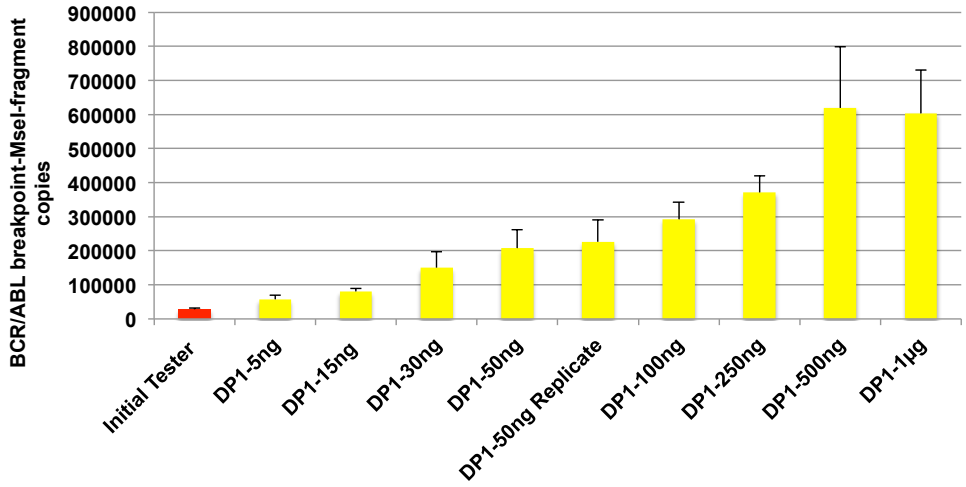


Figure 52: Quantification of the BCR/ABL breakpoint-MseI-fragment absolute copies/ng of DNA in DP1. There is an increased enrichment of the BCR/ABL breakpoint-MseI-fragment by increasing the input amounts of Tester. The enrichment of the first depletion product seems to reach the saturation point between 500ng and 1µg of initial Tester amounts. The total fold enrichment of the DP1 reaches around 22 fold compared to the initial Tester.

DP1 products were separated by agarose gel electrophoresis to inspect the effect of unwanted selective enrichment of cross-annealed repetitive sequences in the first round of subtractive DNA hybridization (Figure 53). The presence of some PCR bands not corresponding to the size of the BCR/ABL MseI-fragment warrants a further improvement of the specificity of subtractive DNA hybridization.

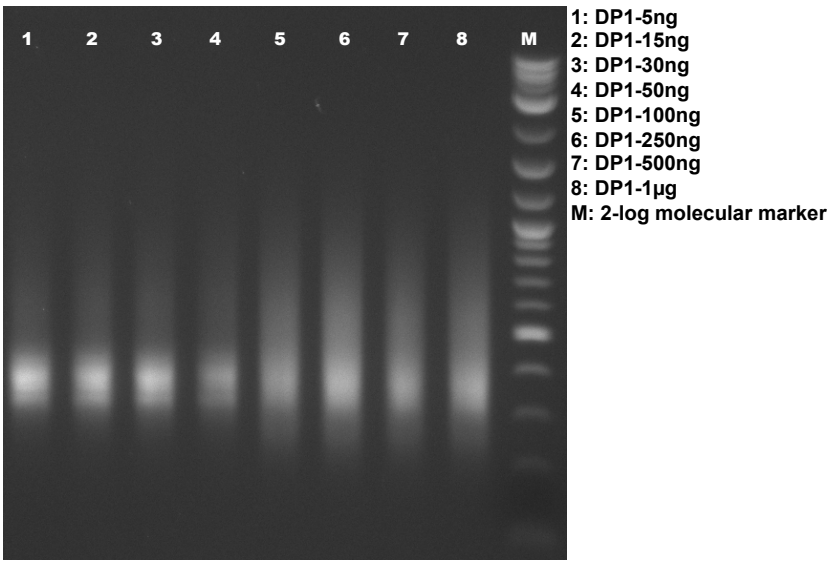


Figure 53: Agarose gel electrophoresis of the different generated DP1s. Amplicons selection can be observed.

In order to suppress the unwanted selected sequences, the DP1-500ng and DP1-1µg were normalized. The normalization of the DP1 suppresses the over-amplified MseI-fragments (Figure 54 A). Additionally, the BCR/ABL qPCR assay was performed on the DP1-500ng-normalized and DP1-1µg-normalized to assess the effect of the normalization on the

enrichment of the BCR/ABL breakpoint-MseI-fragment. It has been clear that the normalization of the DP1-500ng and DP1-1µg contributed to a further enrichment of the BCR/ABL breakpoint-MseI-fragment, respectively, of 3,0 fold and 1,8 fold compared to the respectively not-normalized DP1s (Figure 54 B). The normalization of the DP1-500ng showed to be more effective than the normalization of the DP1-1µg. The total achievable enrichment after the first subtraction round of 500ng of Tester, including the normalization, was 64 fold, while for the subtraction of 1µg of Tester was 39 fold.

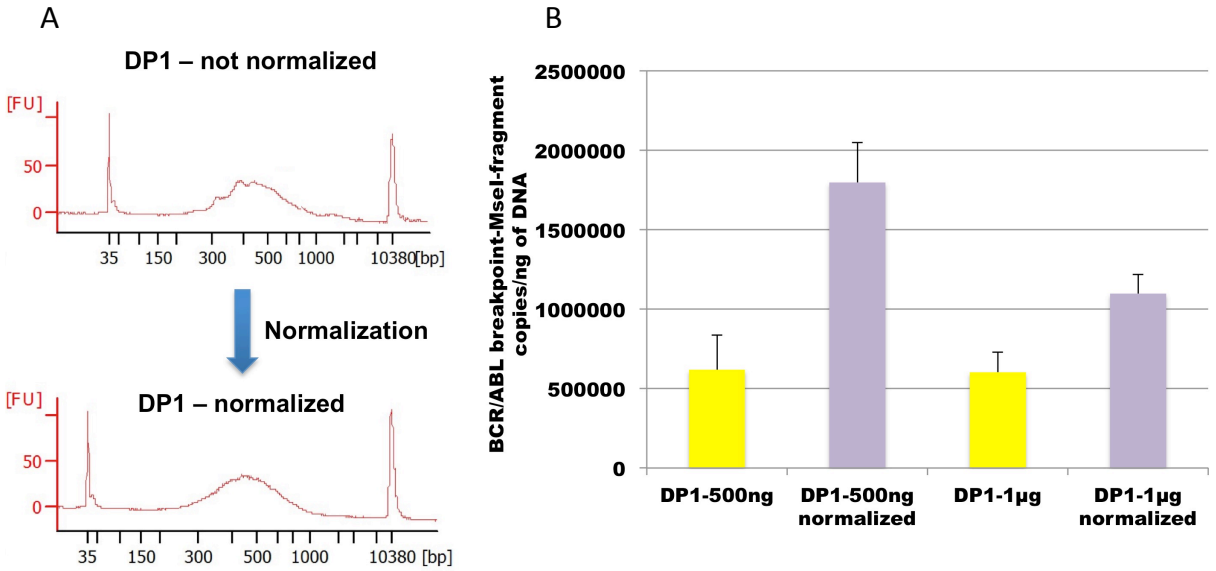


Figure 54: A) Example of normalization of DP1. The subtracted Tester shows an irregular profile due to the selection of MseI-fragments that are preferentially amplified (maybe belonging to repetitive elements of the genome). The normalization of the DP1 re-establishes a homogeneous distribution of the MseI-fragments. B) qPCR assay for the BCR/ABL breakpoint MseI-fragment. The normalization further increases the copies of BCR/ABL breakpoint MseI-fragment.

4.6 Optimization of the second round of subtractive DNA hybridization: depletion product 2 (DP2)

To perform the second round of subtractive hybridization two different Driver to Tester ratios have been tested using the DNA of the normalized DP1-500ng and DP1-1µg (Figure 55):

1. Driver to Tester ratio of 800:1 and 80:1 obtained by subtracting, respectively, 50ng and 500ng of the DP1-500ng-normalized sample;
2. Driver to Tester ratio of 400:1 and 40:1 obtained by subtracting, respectively, 100ng and 1µg of the DP1-1µg-normalized sample.

The higher ratios (800:1 and 400:1) will assess if an increase of the Driver to Tester ratio would favour the kinetic enrichment of the BCR/ABL breakpoint-MseI-fragment, while the lower ratios (80:1 and 40:1), previously used, if are still optimal to achieve a continuous enrichment.

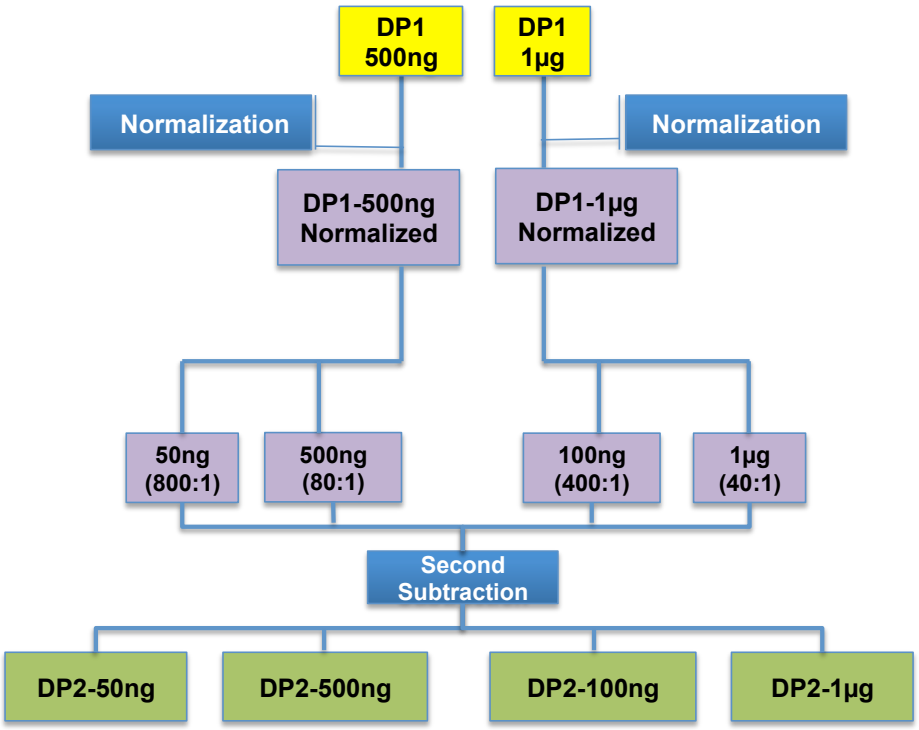


Figure 55: Scheme for the generation of DP2 product. The DP1-500ng and DP1-1µg have been normalized and subtracted in order to generate DP2s. For the normalized DP1-500ng, 50ng and 500ng of DNA have been used for the subtraction; for the normalized DP1-1µg, 100ng and 1µg of DNA amounts have been subtracted.

The generated DP2s were assayed by qPCR for the BCR/ABL breakpoint-MseI-fragment. The DP2-50ng did not show any enrichment compared to the DP1-500ng-normalized, while the DP2-500ng showed 1,6 fold enrichment compared to the DP1-500ng-normalized (Figure 56 A). The DP2-100ng and DP2-1µg did not show any enrichment in comparison to the DP1-1µg-normalized (Figure 56 A). Therefore, either the experimental normalization of DP1-500ng either the constant ratio Tester-Driver 1:80 were considered as essential factors in the process of selective enrichment of BCR/ABL breakpoint-MseI-fragment sequences. In addition, the generated DP2s were separated by agarose gel electrophoresis to inspect the MseI-fragment distribution and the eventual presence of selected sequences. The smear appeared to be very homogeneous (Figure 56 B).

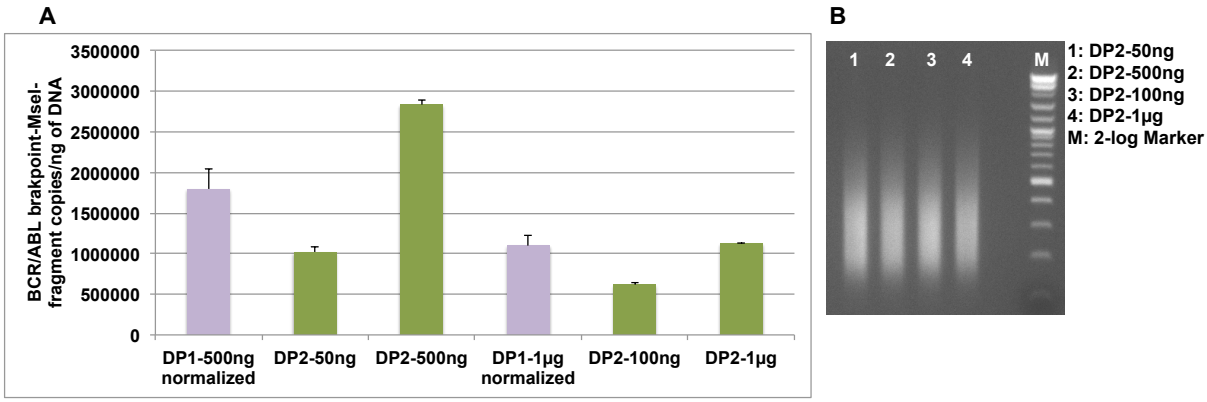


Figure 56: DP2s analysis after second round of subtractive hybridization. Only the DP2-500ng showed to slightly enrich the BCR/ABL breakpoint-MseI-fragment. The subtraction of 100ng and 1µg deriving from DP1-1µg-normalized did not succeed in the enrichment as well as the subtraction of 50ng DNA deriving from the DP1-500ng-normalized.

Subsequently, the DP2-500ng, that showed the better enrichment, has been normalized to assess if an additional normalization step would lead to an increased copy numbers of the BCR/ABL breakpoint-MseI-fragment, as previously observed for the normalization of the DP1. However, the normalization of the DP2-500ng did not bring to any increase in the copy number of the BCR/ABL breakpoint-MseI-fragment indication that a further normalization is not required (Figure 57).

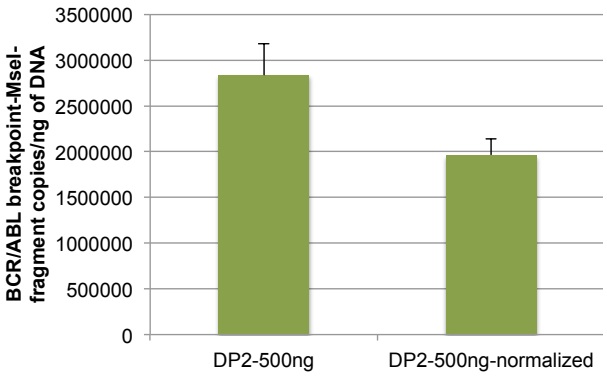


Figure 57: Normalization of the DP2-500ng showing loss of the BCR/ABL breakpoint-MseI-fragment.

4.6.1 Optimization of the 3rd round of Subtractive DNA hybridization: absolute quantification of depletion product 3 (DP3) and agarose gel inspection

For the third round of subtractive hybridization the Driver to Tester ratio is maintained at 80:1 as this ratio was showed to preserve the enrichment of the BCR/ABL breakpoint-MseI-fragment. The subtraction of 500ng of DNA, deriving either from DP2-500ng-normalized either from DP2-500ng not-normalized, was performed (Figure 58).

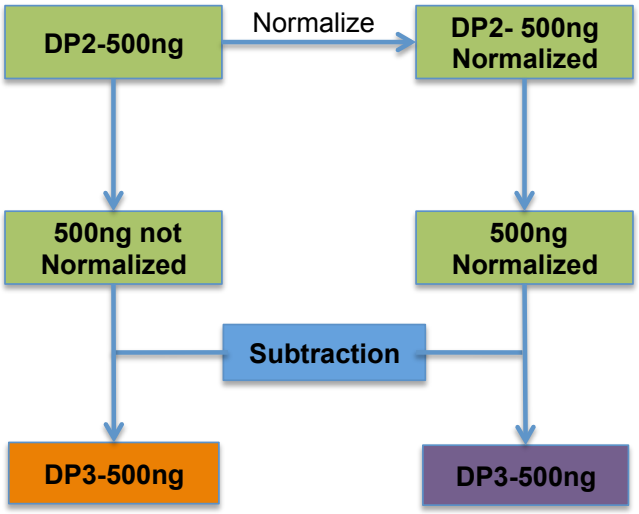


Figure 58: Scheme for the generation of DP3. Part of the DP2-500ng has been normalized, part it has been left un-normalized. After, 500ng of normalized and 500ng of not normalized have been subtracted.

The generated DP3s were assayed by qPCR. The subtraction of the DP2-500ng-normalized did show a loss in the copy numbers of the BCR/ABL breakpoint-MseI-fragment confirming the previous results (Figure 57) and indicating that a further normalization does not improve the enrichment (violet column, Figure 59 A). On the contrary, the subtraction of the DP2-500ng-not-normalized showed a minor increase of enrichment (1,23 fold) (orange column, Figure 59 A). The DP3 products have been observed by agarose gel electrophoresis and no evident differences in the intensity and size distribution of the MseI-fragments were noted (Figure 59 B).

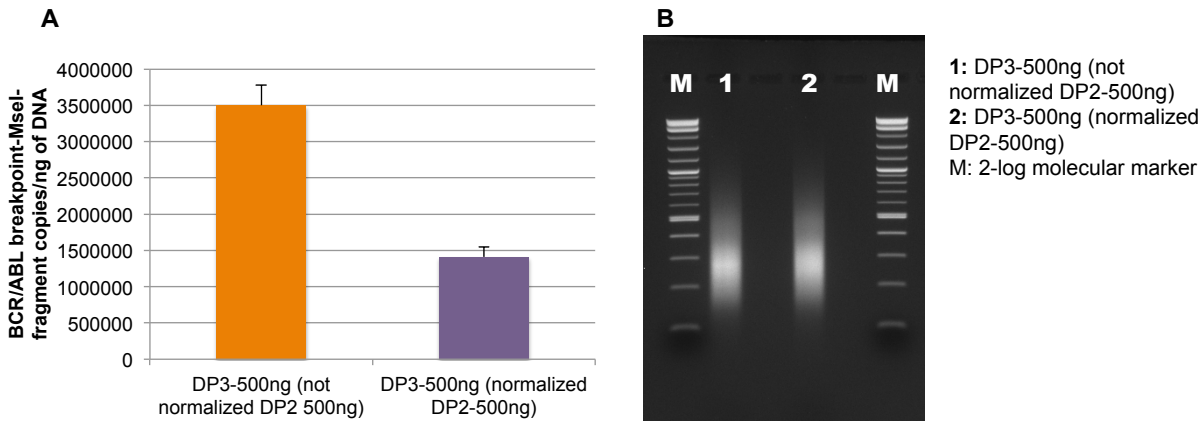


Figure 59: A) subtraction of 500ng DNA of normalized and not normalized DP2-500ng. The fold enrichment of the BCR/ABL breakpoint-MseI-fragment is 1,23 fold if the subtraction of a DP2-500ng-not normalized is performed (orange column). The subtraction of 500ng of normalized DP2-500ng does not increase the copy numbers of the BCR/ABL breakpoint-MseI-fragment (violet column). B) agarose gel inspection of the DP3 deriving from the subtraction of a normalized and non-normalized DP2.

4.7 Summary of the subtractive hybridization performed on the K562 single cell

The best final approach for subtractive hybridization used in order to reach the highest enrichment of MseI-fragments carrying a fusion breakpoint, as in this specific case of the BCR/ABL breakpoint-MseI-fragment, is summarized in the following scheme (Figure 60). The final observed enrichment of an MseI-fragment carrying a breakpoint sequence is ≈ 1664 fold compared to the initial WGA of a single tumor cell.

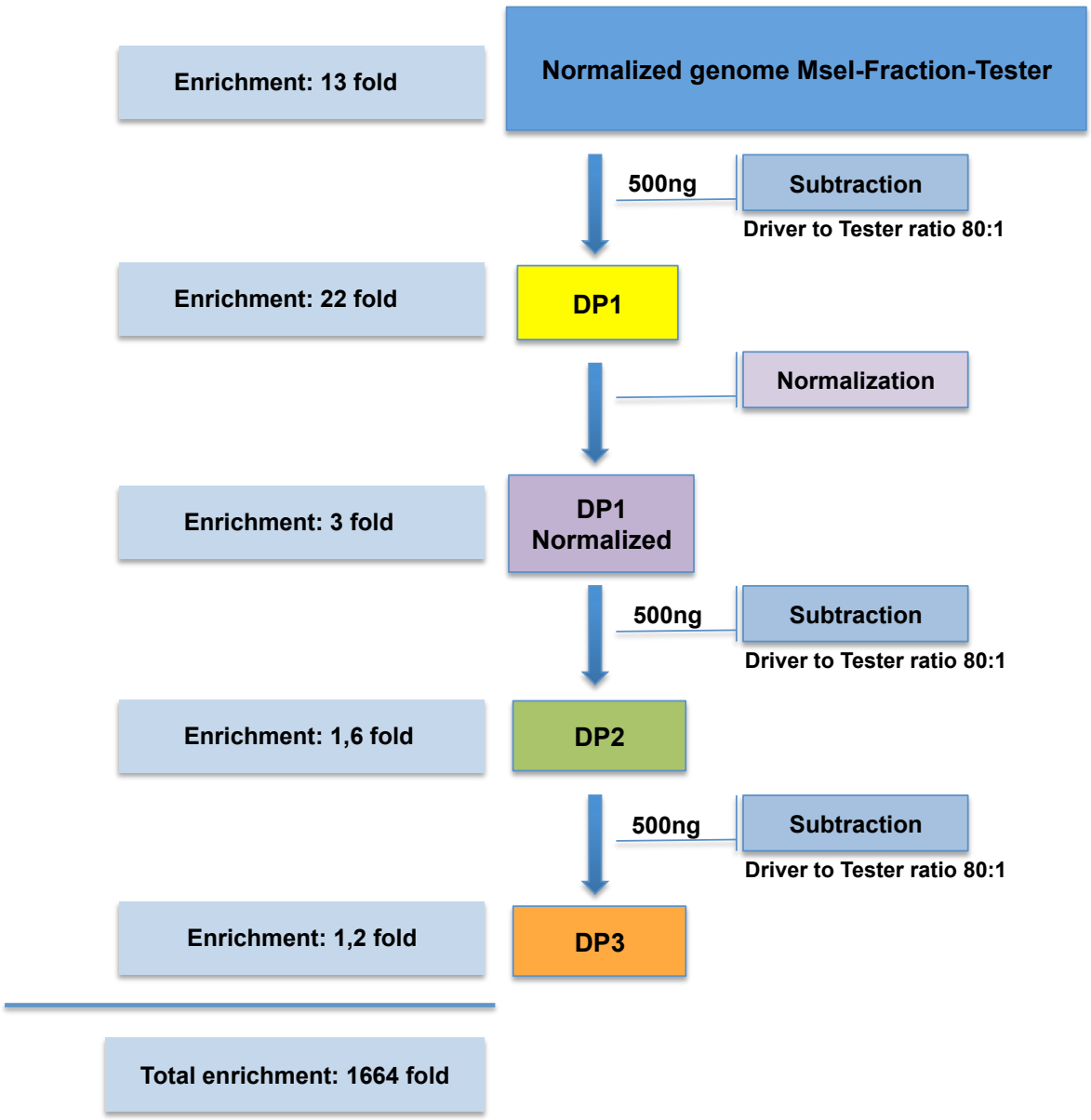


Figure 60: optimal approach to achieve substantial enrichment of MseI-fragments carrying breakpoint sequences starting from the WGA of a single tumor cell.

5. Discussion

5.1 Establishment of a method for the enrichment of genomic breakpoint in single tumor cells

This study describes a method for the enrichment of unknown breakpoint sequences in single tumor cells. The method can be applied to high complex genomes, for example from human cells, in order to find rearranged genomic sequences that could play a role in the early dissemination of cancer cells or for monitoring of minimal residual cancer. In the approach presented here, the critical factors to take in consideration are first, the capability of controlling the selective isolation of sub-genomic DNA-fragment populations in order to maximize the genome complexity reduction but at the same time, minimizing the loss of genomic information. Second, the applicability of a subtractive hybridization method suitable for the enrichment of fusion breakpoints within the selected sub-genomic DNA-fragments population. Additionally, it will be considered the future possibility to combine the novel method with automated PCR systems and NGS technologies, in order to increase the potential to detect unknown genomic breakpoints.

5.2 Importance to reduce genome complexity before subtractive hybridization

The concept of generating “representations” of the genome has been applied for the first time by Lisitsyn et al. in the representational difference analysis (RDA) in order to raise the efficiency of subtractive hybridization (Lisitsyn et al., 1993). Applications of RDA and subtractive hybridization at genomic level showed their potential use for the isolation of polymorphic markers in eukaryotic genomes (Lisitsyn et al., 1994; Nadezhdin et al., 2001), for the detection of genetic losses and amplifications in tumors (Kibel et al., 1999; Kibel et al., 1998; Lisitsyn et al., 1995; Schutte et al., 1995; Zeschnigk et al., 1999), for the identification of mouse Y-specific repeats by whole chromosome RDA (Navin et al., 1996), for the isolation of species-specific loci (Buzdin et al., 2003; Buzdin et al., 2002). Moreover, RDA has been used for filling in extended gaps in large-scale sequencing projects (Frohme et al., 2001). However, it is well documented from experiments (Clapp et al., 1993; Wieland et al., 1990) and from computational modeling (Cho and Park, 1998; Ermolaeva et al., 1996; Ermolaeva and Sverdlov, 1996; Milner et al., 1995) that the complexity of the starting genome has an influence on the performances of the subtractive hybridization. A genome complexity over $5x10^8$bp (far less than the size of the human genome) become already a

limiting factor in the kinetic of the hybridization and, consequently, of the enrichment of target sequences (Milner et al., 1995). The high complexity results in insufficient re-association rate, and in the consequential risk of reducing the enrichment of differential genomic sequences present in low abundance in cancer cells. Therefore genome complexity must be reduced before subtractive hybridization. The classical RDA approaches reduces the complexity of the genome by the use of restriction enzymes: for example the digestion with BamHI results in an estimated 55-fold complexity reduction, while with BglII and HindIII the complexity reduction is, respectively 13 and 8-fold (Bishop et al., 1983). In the specific case discussed in this study, the digestion with MseI generates 19,045,730 millions MseI-fragments. Of these MseI-fragments, 8,151,761 millions have a base pair distribution in between 100-1000bp that represents the optimal size for re-hybridization. This results in an estimated 2.3-fold complexity reduction. The 8.151.761 MseI-fragments will represent 43% of the total genome. On one side this approach will allow to preserve most of the genome information enhancing the success for the identification of genomic breakpoints. On the other side the complexity reduction achieved will not be sufficient for a proper enrichment in subtractive hybridization.

5.2.1 Reducing the complexity of MseI genome representations by "Genome Fractioning"

The reduced genome complexity obtained with MseI needs to be counteracted by the establishment of a process called "Genome Fractioning". The applicability of this approach is clearly linked to the deterministic nature of the Ampli1 WGA (Klein et al., 1999). For Ampli1 WGA, the MseI enzyme cuts the genome at the restriction-site TTAA; such approach will allow the generation of discrete "representations" of MseI-fragments that can be sorted according to the nucleotides next to its genomic MseI restriction site (Figure 4). Ten MseI-fragments populations, representing each different nucleotide combinations (either simple either mixed – Theoretical method design paragraph 3.1), can be separately assayed and will, by large, represent most of the genome. Therefore, a highly reproducible method of generating MseI-fragments representations with reduced complexity for comparison of tumor and normal cells has been possible due to the selective isolation and to the equal enrichment of MseI-fragment populations. It is clear that the genome fractioning approach is not thought for the application to single cell amplification methods that employ random priming such as degenerate-oligonucleotide-primed PCR (DOP) (Telenius et al., 1992), primer extension preamplification (PEP) (Zhang et al., 1992), multiple annealing and looping based

amplification cycles (MALBAC) (Zong et al., 2012) or multiple strand displacement (Hou et al., 2012). Also, the fractioning approach overcomes the major drawback of the classical RDA, as established from Lisitsyn et al. The use of a rare cutter enzyme (BamHI, BglII or HindIII), to reduce the genome complexity, and of the PCR used to amplify the genomic representation ensures a random complexity reduction. It is random because PCR introduces biases by amplifying fragments of different size with different efficiency, which is further confounded by a systematic bias for sequences with different nucleotide GC content. So the final selected amplicons represent only a very small portion comprising 2-10% of the initial fragment diversity. Using the here presented genome fractioning approach it is possible to perform a non-random, deterministic complexity reduction of the genome without excluding 90-98% of the initial representation. The established fractioning procedure reduces the genome complexity by a 10-fold factor, due to the fact that 10 discrete MseI-fragment populations can be exclusively isolated. This means that each selected population will theoretically represent 4,3% of the initial MseI-fragments diversity (8.151.761 millions), rendering each of them appropriate for a proper subtractive hybridization. Taken together, the ten generated MseI-fragment populations (43%) will include in the analysis a portion of the genome that is 4 to 20 times bigger compared to the classical RDA approach.

5.2.2 Major quality controls necessary to define a successful genome fractioning

In order to set up a proper method for the complexity reduction, the process of genome fractioning has been controlled by the use of different quality assays that can be established on the WGA of single cells and applied to amplified DNA samples.

First, the MseI-fragment length distribution has been assessed during the development of the protocol to ensure consistent homogeneous MseI-fragment distribution (Table 32). The MseI-fragment length does not change significantly during the different steps preceding the completion of the genome fractioning; at this point the size distribution of the MseI-fragments shortens to 100-1000bp. This effect is due to the magnetic streptavidin-conjugated beads, which bind with decreased efficiency long DNA fragments because of steric hindrance effect. Nevertheless, the binding of MseI-fragments in between 100-1000bp ensures the optimal size selection necessary for the DNA re-hybridization during the subtraction process.

Second, the process of isolation of discrete MseI-fragment population has been followed step by step by the use of specific MseI-fragment control PCRs. The intention of this quality control PCR-based assay was to show that the ten different genomic fractions can be

properly separated and that minimal or no cross-contamination of different MseI-fractions happens during the genome fractioning. Moreover, this assay showed that it is possible to assess the completeness of the genomic MseI-representation before the selection of a single MseI-fraction (Table 32). The successful selection of the specific MseI-fragments population is achieved by the use of the specific six-cutters enzyme and of the biotinylated fractioning-adaptors. This approach with specific biotinylated fractioning-adaptors ensure that MseI-fragments belonging to the same population are selected avoiding the co-amplification of undesired MseI-fragments. In this regard, it has been shown that also the choice of the magnetic streptavidin-conjugated beads and the related washing procedure can influence the outcome of the genome fractioning. Dynabeads and Nanolink magnetic streptavidin conjugated beads were compared for the generation of the specific MseI-fractions and showed to perform differently, with the first type of beads giving a higher background of un-wanted MseI-fragments (Table 27). This can be an effect of the physical properties of the magnetic streptavidin-conjugated beads. In fact, due to the bigger diameter of the Dynabeads (2.8μm) the chance of unspecific binding of not-biotinylated MseI-fragments is higher compared to the Nanolink beads that have a smaller diameter (heterogeneous population 0.15-0.77μm) but high binding capacity. Such property leads to lower non-specific binding of not-biotinylated MseI-fragments besides to lower costs, because a considerably fewer amount of beads is needed.

Third, the establishment of the quantitative PCR assay has been fundamental to assess the copy numbers of the BCR/ABL breakpoint-MseI-fragment during the setup of the genome fractioning protocol in order to define steps of loss/gain of the fusion breakpoint. By absolute quantitative PCR means, it has been shown that the genome fractioning, over to reduce the complexity of the genome, has also a role on the selective enrichment of the determined MseI-fraction (Figure 41). In fact, the generated genome fraction is enriched for the BCR/ABL breakpoint-MseI-fragment by a factor of 4,3 fold compared to the initial abundance in the WGA.

A problem observed is that genome fractioning can lead to the over-amplification of selected MseI-fragments. This effect it is, as expected, more pronounced on the genome fractioning of single cell compared to pools of cells. These over-amplified MseI-fragments maybe represent repetitive abundant sequences of the human genome as well as preferentially amplified sequences due to the consecutive steps of restriction digestion, adaptors ligation and PCR amplification. Nevertheless, the normalization of the genome MseI-fraction has shown

that it is possible to suppress these sequences and to re-establish a homogeneous representation of the MseI-fragments (Figure 40). Additionally, the suppression of these sequences increases the representation of MseI-fragments present in lower amount, as shown from the further enrichment of the BCR/ABL breakpoint-MseI-fragment by a factor of 3 fold compared to the not-normalized genome MseI-fraction (Figure 41).

5.3.3 Putting into perspective: automation of the genome fractioning procedure

The genome fractioning procedure includes many handling steps that lead to a time-consuming procedure and the introduction of technical error. Even small errors in the accuracy of DNA sample preparation can translate into huge differences after amplification. Of course, it is not impossible to think that this problem can be solved by the use of the many liquid-handling systems present on the market (Gaisford, 2012). These robots enable high reproducibility and accuracy of low-volume dispensing, which is difficult to achieve using manual methods. Additionally, the establishment of the complete procedure on a microfluidic system, where all steps of the genome fractioning protocol are executed in a sequential way, may be possible. Here handling could be reduced and throughput of the analyzed samples could be increased. The use of such systems would, of course, enhance also the possibility to enrich rare differential sequences. Indeed, many microfluidic systems use droplet digital PCR. This type of PCR has many advantages over the conventional PCR especially because it allows the efficient amplification of genomic libraries and other complex mixtures of genes avoiding the generation of chimeric fragments by recombination between homologous regions of DNA (Meyerhans et al., 1990). A variety of PCR protocols have been proposed to minimize these problems, most of which rely on high template concentrations and low numbers of PCR cycles (Polz and Cavanaugh, 1998; Qiu et al., 2001). In the presented study, to minimize the bias in PCR, the number of cycles has been lowered, avoiding the over-amplification of the DNA, the generation of ssDNA and consequent formation of chimeric sequences. However, an approach as droplet PCR would allow the amplification of complex DNA mixtures by segregating template fragments in the minute aqueous droplets of the emulsion and amplified by PCR separately. This would prevent the formation of chimeric products while enabling the use of small amounts of template DNA and high numbers of PCR cycles (Williams et al., 2006). To summarize, the genome fractioning procedure represents, to the current knowledge, the only approach to reduce the complexity of the genome whilst preserving most of the genomic information.

5.4 Applicability of subtractive hybridization to genome fractioning for the enrichment of MseI-fragments carrying fusion breakpoint

This study provides a proof-of-principle that genome fractioning, starting from WGA of a single cell, can be coupled to subtractive hybridization. Subtractive hybridization showed to be very effective in the enrichment of MseI-fragments carrying breakpoints, such as the BCR/ABL fusion investigated in this work. In fact, the total enrichment achieved after three rounds of subtraction was 1664 fold (Figure 60). The presented method performed similar or even better than reported examples of subtractive hybridization applied on complex mammalian genomes. For example, Wieland et al. after three rounds of subtraction obtained between 100- and 700-fold enrichment for bacteriophage λ DNA sequences spiked in human placenta DNA (Wieland et al., 1990). Similarly, Hansen-Hagge et al., in a spiking experiment with λ DNA sequences, observed during the first three rounds of subtractive hybridization a 100-fold enrichment of the λ fragment (Hansen-Hagge et al., 2001). Another approach for the subtraction of high complex genome is the in-gel competitive re-association method, in which the re-hybridization of Driver and Tester samples happens during electrophoresis. Yokota et al. isolated tissue specific changes in repetitive DNA sequences in rats with an enrichment of 3000 fold after two rounds of hybridization (Yokota et al., 1989; Yokota and Oishi, 1990), although such level of enrichment is not surprising due to the fact that it is related to repetitive element of the genome. Sasaki et al. used the same approach for the enrichment of DNA sequences of papillomavirus in cervical cancer with a final 560 fold enrichment after two cycles of in-gel re-association (Sasaki et al., 1994). It is relevant that the method presented in this thesis has been established and tested on whole genome amplified single K562 cells carrying a fusion breakpoint. So, these results show a real observed enrichment of a breakpoint harbored in the genome of the cell, avoiding all the errors related to spiking experiments with external DNA sequences.

5.4.1 Comparison of different subtractive hybridization approaches.

Subtractive hybridization suffers from technical limitations depending on genome complexity, despite the many efforts to improve it for genome-wide comparison. Limitations in the re-hybridization of complex genomes can be partially solved by experimental factors, such as the use of special techniques as solvent exclusion (Barr and Emanuel, 1990) or phenol emulsion re-association technique (PERT) (Bruzel and Cheung, 2006; Goldar and Sikorav, 2004; Kohne et al., 1977; Laman et al., 2001; Miller and Riblet, 1995; Travis and Sutcliffe, 1988;

Wieder and Wetmur, 1982) that enhance the re-hybridization rate by several fold. In a variation of classical PERT, thermal cycler oscillations between 25°C and 65C°, during which the emulsion is formed and broken, resulted in a high rate of reassociation ranging from 100% for lambda digested DNA to only 16% of the human genome DNA (Bruzel and Cheung, 2006). In the presented study, different hybridization methods were compared for the application on WGA of single cells: the classical hybridization in high-salt solution at 68°C, hybridization with formamide at 42°C, and PERT in agitation or with thermal cycling oscillations. In this proof-of-principle experiment, known amount of BCR/ABL breakpoint-MseI-fragments were spiked in a genome fraction and the re-association rate was controlled by quantifying the dsDNA total amount and by detecting the re-hybridized copies of BCR/ABL breakpoint-MseI-fragment. The re-hybridization in high-salt solution at 68°C and PERT using agitation or thermal cycling gave similar results with a percentage of re-association of the total dsDNA and of the BCR/ABL breakpoint-MseI-fragment of about 85%. The number of BCR/ABL breakpoint-MseI-fragments spiked constituted 0,000000026% of the total DNA suggesting that it is possible to detect very low copies of fusion breakpoint. Additionally, multiple different MseI-fragments, presents at even lower amount than before mentioned, can re-associate specifically confirming the sensitivity and specificity of the established protocol (Figure 50). PERT has been claimed to be more suitable for complex genomes, due to the higher re-association rates, when compared to classical subtractive hybridization in high salt solution, but based on our evidence both protocols performed equally well. Probably, the reduction in sequence complexity achieved by genome fractioning enables equal performance, while in high complex genomic mixtures PERT may represent the method of choice.

5.4.2 Importance of Driver to Tester ratio for optimal enrichment

Driver to Tester ratio is one of the factors that can influence subtractive hybridization. According to the literature, high amount of Driver sample were used, but the Driver to Tester ratio was kept to 80:1 because this approach enabled a constant level of enrichment for the BCR/ABL breakpoint-MseI-fragment after three subtraction rounds. Indeed, I found that the rule “the higher the Driver to Tester ratio the higher the enrichment” is not always true for all the subtractive approaches and depends on the experimental scheme (Cho and Park, 1998). For the double-strand Tester selection used in this work, in which only re-hybridized double strand Tester is selected, I noted an optimum Driver to Tester ratio between 1:50-1:100.

5.4.3 Use of biotinylated Driver for subtractive hybridization

The uses of a biotinylated Tester or Driver DNA is an additional factor influencing the subtractive hybridization. In the first case Tester/Tester homohybrids are selected by the use of magnetic streptavidin-conjugated beads, and in the second case background Driver/Tester heterohybrids are removed from the solution. In this study, I noticed that after biotinylation of Driver DNA with biotinylated primers and biotin-14-dCTP, the removal with streptavidin-conjugated magnetic beads was incomplete and even up to 25% of the Driver remained unbound. This is probably due to the incomplete biotinylation or the incomplete capture from the streptavidin magnetic beads. Some groups have used approaches with biotinylated Driver to remove Driver/Tester hybrids (Sive and St John, 1988; Straus and Ausubel, 1990; Sun et al., 1992) in which the driver was photobiotinylated, and removed trough the addition of streptavidin. Barr and Emanuel reported that even after two rounds of photobiotinylation around 1% of the Driver DNA remains unbiotinylated (Barr and Emanuel, 1990). Under these conditions, the incomplete removal of biotinylated driver affects drastically the efficiency of the subtraction (Cho and Park, 1998). For this reason, I have preferred an approach using biotinylated Tester DNA for the first round of subtractive hybridization followed by a classical approach without biotinylation in the next two rounds.

5.4.5 The role of repetitive sequences in subtractive hybridization

Subtractive hybridization is hampered by the presence of repetitive elements in the mammalian genome that constitute up to two-thirds of genome sequences (de Koning et al., 2011; Lander et al., 2001; Venter et al., 2001). These repetitive sequences re-associate much faster than unique sequences (Milner et al., 1995) and subtracted genomic libraries can be greatly enriched in the number of these sequences (Rubin et al., 1993). Moreover the presence of repetitive sequences in high complex genome can lead to cross-annealing between them in regions of partial annealing (Hames and Higgins, 1985) and, consequently, to the formation of chimeric sequences that alter the DNA analysis. These chimeric sequences can constitute up to 40-60% of the DNA subtracted library (Chalaya et al., 2004). The chimeric sequences are directly correlated to the complexity of the genome that is analyzed and are the main problem that makes the subtractive hybridization less applicable to mammalian genomes. An attempt to solve the problem generated by repetitive elements is the use of Cot DNA, commercially available and enriched in genomic repeats that can be used to “capture” these sequences. I did not use any competitor DNA because it has been shown that breakpoints are often located in

or near repetitive elements of the genome (Deininger and Batzer, 1999; Kolomietz et al., 2002; Wei et al., 2003) and at enrichment sites of microhomologous sequences (Kato et al., 2012; Lawson et al., 2011). This could have reduced the possibility to identify such fusion sequences. Additionally, Chalaya et al 2004 showed that the use of competitor DNA to 100 fold excess during subtractive hybridization decreases the number of clones containing genomic repeats only from 93% to 76-78%, so chances for success using this approach were deemed low. Instead, the same research group showed in their mispaired DNA rejection (MDR) method that mung bean nuclease digests loops formed from improper hybridized DNA and reduces chimera formation from 60% to 4% (Chalaya et al 2004). Repetitive elements share regions of homology that, although not identical, can form imperfect heteroduplexes. For this reason, I used mung bean nuclease to reduce the improperly mispaired DNA duplexes. Additionally, linearly amplified Driver/Tester hybrids can be digested by mung bean nuclease. Hansen-Hagge et al. addressed the problem of repetitive sequences and artificial repeat-mediated fusion sequences by the use of a highly specific ligation reaction and purification of such ligation products with streptavidin-coupled magnetic beads (Hansen-Hagge et al., 2001). In this way they could achieve a 100-fold enrichment of the sequences of interest in the first three rounds of subtraction that declined subsequently, due to the fact that some other sequences start to dominate. This was also observed in the current study, when after the second round of subtraction only a negligible enrichment of the BCR/ABL breakpoint-MseI-fragment was achieved. To circumvent the problem, the normalization of the DP1 helped to suppress the sequences that were over-represented after the first round of subtraction. This increased the amount of detectable BCR/ABL breakpoint-MseI-fragment in qPCR assay by a factor of 3 folds (Figure 54 B).

5.4.6 Improvement of subtractive hybridization by sample normalization

The normalization of the first depletion product (DP1) enabled to further enrich the target sequence despite the presence of repetitive sequences. Normalization enabled to slightly increase the enrichment of the fusion BCR/ABL breakpoint-MseI-fragment until the third round of subtractive hybridization. However the normalization and subtraction of the second depletion product (DP2) did not improve the enrichment of the BCR/ABL breakpoint-MseI-fragment. This evidence indicates that an equalized MseI-fragment library has been reached, meaning that the abundance of the BCR/ABL breakpoint MseI-fragment is similar to other MseI-fragments. The normalization results presented here are supported by the only data available regarding normalization of genomic DNA: normalization proved to be very useful

for the evaluation of the sequencing data of the human genome due to the reduction of repetitive elements (Shagina et al., 2010). In that study, normalization reduced the number of repetitive elements from 40% to 25%, the most affected sequences being Alu, LINE L1P, ERV-K, and ERV1 repeats, as well as satellite sequences.

5.5 Advantages of the method in the era of next generation sequencing

Until recent years, the detection of new structural rearrangements has been very challenging, but the unprecedented increase in the throughput of DNA sequencing driven by next-generation technologies now allows efficient analysis of complete genomes and transcriptomes (Fuller et al., 2009; Metzker, 2010). Nevertheless, it is not yet feasible to sequence large numbers of complex genomes in their entirety because the cost and time necessary are still too great. To obtain 30-fold coverage of a human genome (90 Gb in total) would currently require several sequencing runs and would cost several thousand euros. In addition to the price demands, such a project would request laboratory time and funding and would place a substantial load on a research center's informatics infrastructure. The presented method aims also to support the discovery of genomic rearrangements by a combined approach with next generation sequencing technologies. The developed protocol in which it is possible to selectively divide genomic sub-regions and to enrich, through subtractive hybridization, differential genomic sequences characteristic of cancer cells, could reduce significantly costs and efforts, and the resulting informatics data would be considerably less difficult to analyze compared with whole-genome sequencing. Strategies for direct selection of genomic regions were already developed in anticipation of the introduction of NGS (Bashiardes et al., 2005; Dahl et al., 2005) but only limited regions of interest of the genome could be enriched. The presented method allows the selection (through the genome fractioning) of a large part of the genome and the enrichment (through the subtraction) of sequences present uniquely in cancer cells. Each subtracted MseI-fractions, mostly enriched in differential DNA sequences, could be bar-coded and pooled allowing to analyze multiple genomic MseI-fractions in a single sequencing run. For example, for the investigated genomic MseI-fraction, containing the BCR/ABL breakpoint-MseI-fragment, the enrichment level achieved (1664 fold) will allow to use a low sequencing coverage sufficient for the detection of the translocation. To summarize, the method could provide a cost-effective approach for the discovery of structural variation in single cancer cells that can be evaluated for the development of potential targeted therapies or of cancer vaccines (neo-antigens).

Additionally, the method can be used to isolate defined sub-genomic populations for experiments of targeted re-sequencing or custom mini-genome sequencing projects.

5.6 A method to identify potential biomarkers for monitoring minimal residual disease

One application of the method in the near future is the genomic characterization of early DCCs/CTCs. Genome characterization of the DCCs/CTCs will allow identifying stable genomic rearrangements usable as novel biomarkers to monitor MRD, as it is generally performed for chronic myeloid leukemia (CML), acute lymphocytic leukemia (ALL) or acute myeloid leukemia (AML). Substantial evidence suggests that translocations represent essential and early steps in the initiation of carcinogenesis and that gene fusion, such as *TMPRSS2-ERG,* are early events necessary for the transition from precursor lesion of prostate cancer to an invasive phenotype. This renders fusion genes appealing candidates to monitor disease state and burden. Until now, applicable targets for MRD detection have been more difficult to determine in carcinomas and the monitoring of MRD is much less advanced compared to leukemia and lymphoma. Monitoring the clonal expansion or depletion of cell populations harboring specific unique DNA sequences, could constitute an ideal method for real-time tracking of the efficacy of systemic adjuvant therapy in individual patients during follow-up. Manifest metastatic disease is usually resistant to current therapies. Therefore, manifestation of metastasis could be prevented if biomarkers were available during minimal residual disease, assessing disease activity and enabling timely intervention. After completion of standard adjuvant chemotherapies the identification of patients at increased risk for relapse, based on the detection level of biomarkers, could represent an application of high clinical relevance, as these patients might benefit from an additional treatment with new drugs (Figure 61).

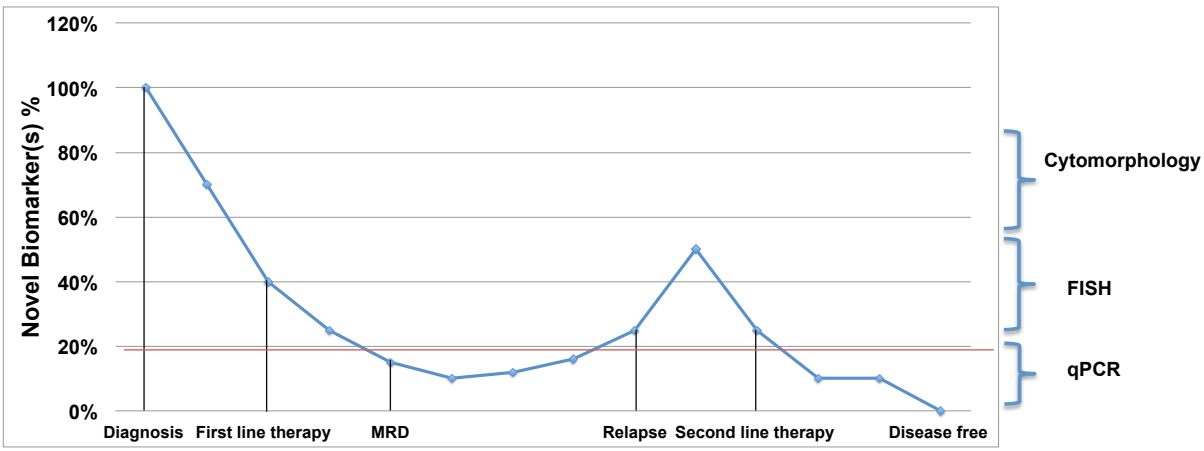


Figure 61: hypothetical scenario for monitoring MRD. qPCR can detect cancer DNA sequences with higher sensitivity compared to classical cytogenetic approaches constituting a valid mean to monitor MRD. The red line indicates an arbitrary threshold for the detection of MRD.

Of course, for such a routine monitoring approach peripheral blood analyses are more convenient for patients than invasive BM sampling. Additionally, 'liquid biopsies' are an attractive alternative option because DNA fragments released from cancer cells, cell-free tumor DNA (cfDNA), into the blood can be used to generate molecular profiles of tumors (Diaz and Bardelli, 2014). Circulating tumor DNA fragments contain identical genetic defects to those in the tumors themselves and virtually all cancer-related molecular alterations can be detected in cfDNA, such as somatic point mutations, loss of heterozygosity (LOH), gene copy number changes and DNA methylation changes, including translocations. Therefore, detection of tumor-specific sequences such translocation breakpoints may enable PCR-based disease-activity monitoring and emergence of drug resistance.

5.6.1 Development of therapeutic drugs targeting the clonal evolution

Another possible application of the method, but more challenging, is the characterization of the early genomic changes in DCCs/CTCs that can be used as potential targets for therapeutic drugs. The time lag between the surgical removal of the primary tumor and the manifestation of macroscopic metastasis years to decades later, suggests ongoing adaptation and survival of these cells in metastatic niches, before acquiring a metastatic growth. Of note, early changes, if selected and preserved, are likely to be shared among all tumor cells and may be exploited for the prevention of metachronous metastases. Furthermore, such studies may uncover functionally important changes driving clonal evolution. A way to monitor genome evolution of DCCs/CTCs could be done by the isolation of these single cells from patients at different time-points, such as at the diagnosis/surgery of the primary tumor, during the course of adjuvant chemotherapy and by multiple samplings during the metastasis-free disease. This might allow identification of changes that are stable and shared over time and therefore represent potential drug targets (Figure 62).

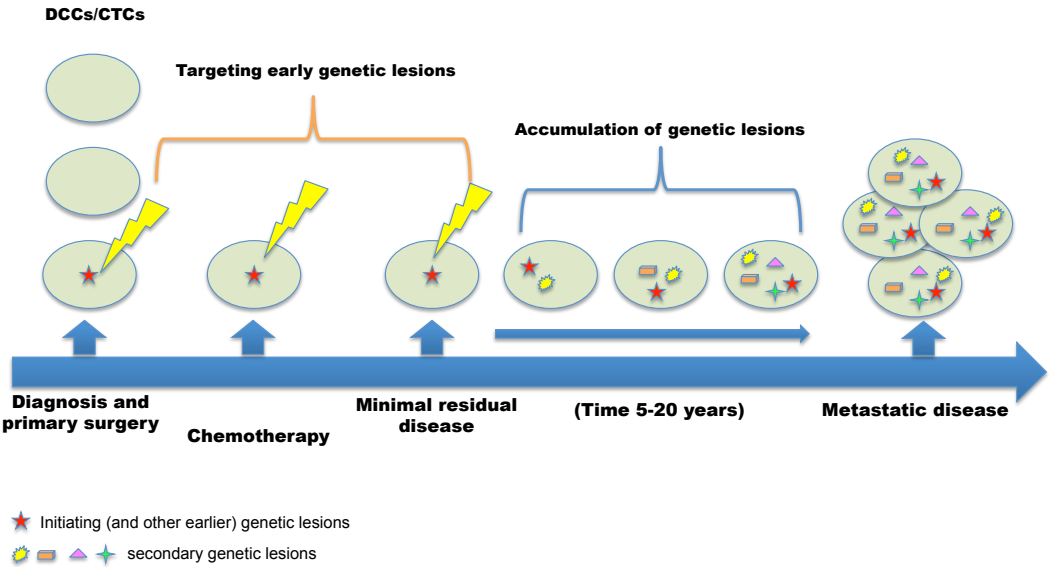


Figure 62: monitoring of the metastatic disease progression. Monitoring DCCs/CTCs evolution is possible by multiple sampling of these cells at diagnosis and at different time point of the treatment and follow-up (blue time line). Early and stable genomic changes important for early dissemination, homing and survival can be identified at the time of the diagnosis. Such genomic rearrangements can constitute potential targets for therapies necessary for the eradication of the MRD.

For this, a recent interest in a novel class of cancer immunogens has arisen. Neoantigens arise as a direct consequence of somatic mutations within tumor cells, therefore they are very specific to the tumor, but also unique to the patient (Fritsch et al., 2014; Hacohen et al., 2013; Litterman et al., 2013; Wolchok and Chan, 2014). T-cells can recognize these antigens that are presented via the MHC on the surface of human tumor cells and thereby mediate cancer regression (Kvistborg et al., 2013). Fusion genes are promising candidates as neoantigen, because fusion proteins created by chromosomal translocations in tumors can generate neoantigenic determinants at the breakpoint, which are unique to the tumor cells (Worley et al., 2001). It has been shown already that *BCR-ABL* fusion gene that arises in chronic myeloid leukaemia (CML) is a neoantigen. Peptides derived from the BCR-ABL fusion junction may therefore be immunogenic, if appropriately presented to the immune system (Cai et al., 2012; Jain et al., 2009; Rajasagi et al., 2014; Scheich et al., 2007; Wong and Chatterjee, 2005). Ideally, a personalized neoantigen approach could be applied early in disease evolution and immunologic intervention may effectively eliminate incipient disease cells such as DCCs/CTCs.

6. Summary

Metastasis is the major cause of cancer-related deaths. Occult cancer cells when found in the bone marrow of patients with carcinomas have been termed disseminated cancer cells (DCCs) and when found in the blood circulating tumor cells (CTCs). CTCs and DCCs may comprise metastasis-initiating cells and surrogate markers for minimal residual disease (MRD) after curative surgery of the primary tumor. Numerous studies have shown that the presence of DCCs in bone marrow or CTCs in blood detected at time of curative surgery or during the follow-up of the patient put the patients at risk for disease progression and death. Therefore, there is a medical need to better characterize these cells and identify early oncogenic changes that drive metastatic progression. Fusion genes have been shown to be among the very early genomic changes necessary for cancer progression, as demonstrated for the *TMPRSS2-ERG* rearrangement in prostate cancer, but the identification of these structural variations has been challenging until now, especially in carcinomas, due to the lack of methods able to provide high resolution.

Therefore, this work aimed to develop a method for the enrichment of breakpoint sequences that can be applied to DCCs and CTCs. Since these cells are extremely rare the method should enable fusion gene identification in single cells. The technique was established as downstream application of a deterministic method to amplify the genome of single cells. The method was based on the reduction of genome complexity and subtractive hybridization. Reduction of complexity was achieved by portioning the genome into ten fractions based on deterministic fragment selection. Several approaches of subtractive hybridization were subsequently compared and applied. In order to establish the technique, single cells derived from the K562 cell line (Tester), harboring the *BCR-ABL* gene fusion, and a pool of unrelated lymphocytes (Driver) were used. The enrichment of the breakpoint fusion was monitored by quantitative PCR which demonstrated an enrichment of 1664-fold, sufficient to detect fusion sequences by low coverage sequencing.

Thus combination of genome fractioning with subtractive hybridization is a valid approach for the enrichment of breakpoint sequences and could be applied for the genetic characterization of CTCs and DCCs. Future applications of the method might include monitoring cancer specific sequences during the course of therapy of solid cancers, as it is routinely performed for the *BCR-ABL* fusion gene to monitor therapy response and disease remission, or the identification of novel therapy targets, such as neo-antigens.

7. References

Aguirre-Ghiso, J.A. (2007). Models, mechanisms and clinical evidence for cancer dormancy. Nature reviews Cancer *7*, 834-846.

Allred, D.C., Mohsin, S.K., and Fuqua, S.A. (2001). Histological and biological evolution of human premalignant breast disease. Endocrine-related cancer *8*, 47-61.

Aplan, P.D. (2006). Causes of oncogenic chromosomal translocation. Trends in genetics : TIG *22*, 46-55.

Arlt, M.F., Durkin, S.G., Ragland, R.L., and Glover, T.W. (2006). Common fragile sites as targets for chromosome rearrangements. DNA repair *5*, 1126-1135.

Arnaudeau, C., Lundin, C., and Helleday, T. (2001). DNA double-strand breaks associated with replication forks are predominantly repaired by homologous recombination involving an exchange mechanism in mammalian cells. Journal of molecular biology *307*, 1235-1245.

Artandi, S.E., and DePinho, R.A. (2010). Telomeres and telomerase in cancer. Carcinogenesis *31*, 9-18.

Banys, M., Gruber, I., Krawczyk, N., Becker, S., Kurth, R., Wallwiener, D., Jakubowska, J., Hoffmann, J., Rothmund, R., Staebler, A.*, et al.* (2012). Hematogenous and lymphatic tumor cell dissemination may be detected in patients diagnosed with ductal carcinoma in situ of the breast. Breast cancer research and treatment *131*, 801-808.

Barr, F.G., and Emanuel, B.S. (1990). Application of a subtraction hybridization technique involving photoactivatable biotin and organic extraction to solution hybridization analysis of genomic DNA. Analytical biochemistry *186*, 369-373.

Bartkova, J., Horejsi, Z., Koed, K., Kramer, A., Tort, F., Zieger, K., Guldberg, P., Sehested, M., Nesland, J.M., Lukas, C.*, et al.* (2005). DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. Nature *434*, 864-870.

Bashiardes, S., Veile, R., Helms, C., Mardis, E.R., Bowcock, A.M., and Lovett, M. (2005). Direct genomic selection. Nature methods *2*, 63-69.

Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M., and Moran, J.V. (2010). LINE-1 retrotransposition activity in human genomes. Cell *141*, 1159-1170.

Berger, M.F., Lawrence, M.S., Demichelis, F., Drier, Y., Cibulskis, K., Sivachenko, A.Y., Sboner, A., Esgueva, R., Pflueger, D., Sougnez, C.*, et al.* (2011). The genomic complexity of primary human prostate cancer. Nature *470*, 214-220.

Bishop, D.T., Williamson, J.A., and Skolnick, M.H. (1983). A model for restriction fragment length distributions. American journal of human genetics *35*, 795-815.

Blasco, M.A. (2005). Telomeres and human disease: ageing, cancer and beyond. Nature reviews Genetics *6*, 611-622.

Borden, E.C., Baker, L.H., Bell, R.S., Bramwell, V., Demetri, G.D., Eisenberg, B.L., Fletcher, C.D., Fletcher, J.A., Ladanyi, M., Meltzer, P.*, et al.* (2003). Soft tissue sarcomas of adults: state of the translational science. Clinical cancer research : an official journal of the American Association for Cancer Research *9*, 1941-1956.

Branzei, D., and Foiani, M. (2010). Maintaining genome stability at the replication fork. Nature reviews Molecular cell biology *11*, 208-219.

Braun, S., Pantel, K., Muller, P., Janni, W., Hepp, F., Kentenich, C.R., Gastroph, S., Wischnik, A., Dimpfl, T., Kindermann, G.*, et al.* (2000). Cytokeratin-positive cells in the bone marrow and survival of patients with stage I, II, or III breast cancer. The New England journal of medicine *342*, 525-533.

Bruzel, A., and Cheung, V.G. (2006). DNA reassociation using oscillating phenol emulsions. Genomics *87*, 286-289.

Buerger, H., Otterbach, F., Simon, R., Poremba, C., Diallo, R., Decker, T., Riethdorf, L., Brinkschmidt, C., Dockhorn-Dworniczak, B., and Boecker, W. (1999). Comparative genomic hybridization of ductal carcinoma in situ of the breast-evidence of multiple genetic pathways. The Journal of pathology *187*, 396-402.

Burkhart, D.L., and Sage, J. (2008). Cellular mechanisms of tumour suppression by the retinoblastoma gene. Nature reviews Cancer *8*, 671-682.

Burrow, A.A., Williams, L.E., Pierce, L.C., and Wang, Y.H. (2009). Over half of breakpoints in gene pairs involved in cancer-specific recurrent translocations are mapped to human chromosomal fragile sites. BMC genomics *10*, 59.

Buzdin, A., Ustyugova, S., Gogvadze, E., Lebedev, Y., Hunsmann, G., and Sverdlov, E. (2003). Genome-wide targeted search for human specific and polymorphic L1 integrations. Human genetics *112*, 527-533.

Buzdin, A., Ustyugova, S., Gogvadze, E., Vinogradova, T., Lebedev, Y., and Sverdlov, E. (2002). A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3' terminus of l1. Genomics *80*, 402-406.

Cai, A., Keskin, D.B., DeLuca, D.S., Alonso, A., Zhang, W., Zhang, G.L., Hammond, N.N., Nardi, V., Stone, R.M., Neuberg, D.*, et al.* (2012). Mutated BCR-ABL generates immunogenic T-cell epitopes in CML patients. Clinical cancer research : an official journal of the American Association for Cancer Research *18*, 5761-5772.

Cairns, J. (1975). Mutation selection and the natural history of cancer. Nature *255*, 197-200.

Carmeliet, P. (2005). VEGF as a key mediator of angiogenesis in cancer. Oncology *69 Suppl 3*, 4-10.

Chaffer, C.L., and Weinberg, R.A. (2011). A perspective on cancer cell metastasis. Science *331*, 1559-1564.
Chalaya, T., Gogvadze, E., Buzdin, A., Kovalskaya, E., and Sverdlov, E.D. (2004). Improving specificity of DNA hybridization-based methods. Nucleic acids research *32*, e130.

Chen, Q., Zhang, X.H., and Massague, J. (2011). Macrophage binding to receptor VCAM-1 transmits survival signals in breast cancer cells that invade the lungs. Cancer cell *20*, 538-549.

Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.J., Myers, D.R., Choi, V.W., Compagno, M., Malkin, D.J.*, et al.* (2011). Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. Cell *147*, 107-119.

Chin, K., de Solorzano, C.O., Knowles, D., Jones, A., Chou, W., Rodriguez, E.G., Kuo, W.L., Ljung, B.M., Chew, K., Myambo, K.*, et al.* (2004). In situ analyses of genome instability in breast cancer. Nature genetics *36*, 984-988.

Cho, T.J., and Park, S.S. (1998). A simulation of subtractive hybridization. Nucleic acids research *26*, 1440-1448.

Clapp, J.P., McKee, R.A., Allen-Williams, L., Hopley, J.G., and Slater, R.J. (1993). Genomic subtractive hybridization to isolate species-specific DNA sequences in insects. Insect molecular biology *1*, 133-138.

Collins, V.P., Loeffler, R.K., and Tivey, H. (1956). Observations on growth rates of human tumors. The American journal of roentgenology, radium therapy, and nuclear medicine *76*, 988-1000.

Cremer, T., Cremer, M., Dietzel, S., Muller, S., Solovei, I., and Fakan, S. (2006). Chromosome territories--a functional nuclear landscape. Current opinion in cell biology *18*, 307-316.

Dahl, F., Gullberg, M., Stenberg, J., Landegren, U., and Nilsson, M. (2005). Multiplex amplification enabled by selective circularization of large sets of genomic DNA fragments. Nucleic acids research *33*, e71.

de Koning, A.P., Gu, W., Castoe, T.A., Batzer, M.A., and Pollock, D.D. (2011). Repetitive elements may comprise over two-thirds of the human genome. PLoS genetics *7*, e1002384.

Deininger, M., Buchdunger, E., and Druker, B.J. (2005). The development of imatinib as a therapeutic agent for chronic myeloid leukemia. Blood *105*, 2640-2653.

Deininger, P.L., and Batzer, M.A. (1999). Alu repeats and human disease. Molecular genetics and metabolism *67*, 183-193.

Dhillon, A.S., Hagan, S., Rath, O., and Kolch, W. (2007). MAP kinase signalling pathways in cancer. Oncogene *26*, 3279-3290.

Diaz, L.A., Jr., and Bardelli, A. (2014). Liquid biopsies: genotyping circulating tumor DNA. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *32*, 579-586.

Elkabets, M., Gifford, A.M., Scheel, C., Nilsson, B., Reinhardt, F., Bray, M.A., Carpenter, A.E., Jirstrom, K., Magnusson, K., Ebert, B.L.*, et al.* (2011). Human tumors instigate granulin-expressing hematopoietic cells that promote malignancy by activating stromal fibroblasts in mice. The Journal of clinical investigation *121*, 784-799.

Engel, J., Eckel, R., Kerr, J., Schmidt, M., Furstenberger, G., Richter, R., Sauer, H., Senn, H.J., and Holzel, D. (2003). The process of metastasisation for breast cancer. European journal of cancer *39*, 1794-1806.

Ermolaeva, O.D., Lukyanov, S.A., and Sverdlov, E.D. (1996). The mathematical model of subtractive hybridization and its practical application. Proceedings / International Conference on Intelligent Systems for Molecular Biology ; ISMB International Conference on Intelligent Systems for Molecular Biology *4*, 52-58.

Ermolaeva, O.D., and Sverdlov, E.D. (1996). Subtractive hybridization, a technique for extraction of DNA sequences distinguishing two closely related genomes: critical analysis. Genetic analysis : biomolecular engineering *13*, 49-58.

Ewing, A.D., and Kazazian, H.H., Jr. (2010). High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. Genome research *20*, 1262-1270.

Eyles, J., Puaux, A.L., Wang, X., Toh, B., Prakash, C., Hong, M., Tan, T.G., Zheng, L., Ong, L.C., Jin, Y.*, et al.* (2010). Tumor cells disseminate early, but immunosurveillance limits metastatic outgrowth, in a mouse model of melanoma. The Journal of clinical investigation *120*, 2030-2039.

Fabbri, F., Carloni, S., Zoli, W., Ulivi, P., Gallerani, G., Fici, P., Chiadini, E., Passardi, A., Frassineti, G.L., Ragazzini, A.*, et al.* (2013). Detection and recovery of circulating colon cancer cells using a dielectrophoresis-based device: KRAS mutation status in pure CTCs. Cancer letters *335*, 225-231.

Fearon, E.R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. Cell *61*, 759-767.

Felix, C.A., Kolaris, C.P., and Osheroff, N. (2006). Topoisomerase II and the etiology of chromosomal translocations. DNA repair *5*, 1093-1108.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. International journal of cancer Journal international du cancer *136*, E359-386.

Fidler, I.J. (2003). The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. Nature reviews Cancer *3*, 453-458.

Friberg, S., and Mattson, S. (1997). On the growth rates of human malignant tumors: implications for medical decision making. Journal of surgical oncology *65*, 284-297.

Fritsch, E.F., Hacohen, N., and Wu, C.J. (2014). Personal neoantigen cancer vaccines: The momentum builds. Oncoimmunology *3*, e29311.

Frohme, M., Camargo, A.A., Czink, C., Matsukuma, A.Y., Simpson, A.J., Hoheisel, J.D., and Verjovski-Almeida, S. (2001). Directed gap closure in large-scale sequencing projects. Genome research *11*, 901-903.

Fuller, C.W., Middendorf, L.R., Benner, S.A., Church, G.M., Harris, T., Huang, X., Jovanovich, S.B., Nelson, J.R., Schloss, J.A., Schwartz, D.C.*, et al.* (2009). The challenges of sequencing by synthesis. Nature biotechnology *27*, 1013-1023.

Gagos, S., and Irminger-Finger, I. (2005). Chromosome instability in neoplasia: chaotic roots to continuous growth. The international journal of biochemistry & cell biology *37*, 1014-1033.

Gaisford, W. (2012). Robotic liquid handling and automation in epigenetics. Journal of laboratory automation *17*, 327-329.

Gandhi, M., Dillon, L.W., Pramanik, S., Nikiforov, Y.E., and Wang, Y.H. (2010). DNA breaks at fragile sites generate oncogenic RET/PTC rearrangements in human thyroid cells. Oncogene *29*, 2272-2280.

Gao, H., Chakraborty, G., Lee-Lim, A.P., Mo, Q., Decker, M., Vonica, A., Shen, R., Brogi, E., Brivanlou, A.H., and Giancotti, F.G. (2012). The BMP inhibitor Coco reactivates breast cancer cells at lung metastatic sites. Cell *150*, 764-779.

Gasch, C., Bauernhofer, T., Pichler, M., Langer-Freitag, S., Reeh, M., Seifert, A.M., Mauermann, O., Izbicki, J.R., Pantel, K., and Riethdorf, S. (2013). Heterogeneity of epidermal growth factor receptor status and mutations of KRAS/PIK3CA in circulating tumor cells of patients with colorectal cancer. Clinical chemistry *59*, 252-260.

Geigl, J.B., Obenauf, A.C., Schwarzbraun, T., and Speicher, M.R. (2008). Defining 'chromosomal instability'. Trends in genetics : TIG *24*, 64-69.

Goldar, A., and Sikorav, J.L. (2004). DNA renaturation at the water-phenol interface. The European physical journal E, Soft matter *14*, 211-239.

Gorgoulis, V.G., Vassiliou, L.V., Karakaidos, P., Zacharatos, P., Kotsinas, A., Liloglou, T., Venere, M., Ditullio, R.A., Jr., Kastrinakis, N.G., Levy, B.*, et al.* (2005). Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. Nature *434*, 907-913.

Hacohen, N., Fritsch, E.F., Carter, T.A., Lander, E.S., and Wu, C.J. (2013). Getting personal with neoantigen-based therapeutic cancer vaccines. Cancer immunology research *1*, 11-15.

Haffner, M.C., Aryee, M.J., Toubaji, A., Esopi, D.M., Albadine, R., Gurel, B., Isaacs, W.B., Bova, G.S., Liu, W., Xu, J.*, et al.* (2010). Androgen-induced TOP2B-mediated double-strand breaks and prostate cancer gene rearrangements. Nature genetics *42*, 668-675.

Hakim, O., Resch, W., Yamane, A., Klein, I., Kieffer-Kwon, K.R., Jankovic, M., Oliveira, T., Bothmer, A., Voss, T.C., Ansarah-Sobrinho, C.*, et al.* (2012). DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. Nature *484*, 69-74.

Hames, B.D., and Higgins, S.J. (1985). Nucleic acid hybridisation : a practical approach (Oxford: IRL).

Hanahan, D., and Folkman, J. (1996). Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. Cell *86*, 353-364.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674.

Hansen-Hagge, T.E., Trefzer, U., zu Reventlow, A.S., Kaltoft, K., and Sterry, W. (2001). Identification of sample-specific sequences in mammalian cDNA and genomic DNA by the novel ligation-mediated subtraction (Limes). Nucleic acids research *29*, E20.

Heitzer, E., Auer, M., Gasch, C., Pichler, M., Ulz, P., Hoffmann, E.M., Lax, S., Waldispuehl-Geigl, J., Mauermann, O., Lackner, C.*, et al.* (2013). Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing. Cancer research *73*, 2965-2975.

Hess, J.L. (2004). MLL: a histone methyltransferase disrupted in leukemia. Trends in molecular medicine *10*, 500-507.

Hosch, S., Kraus, J., Scheunemann, P., Izbicki, J.R., Schneider, C., Schumacher, U., Witter, K., Speicher, M.R., and Pantel, K. (2000). Malignant potential and cytogenetic characteristics of occult disseminated tumor cells in esophageal cancer. Cancer research *60*, 6836-6840.

Hou, Y., Song, L., Zhu, P., Zhang, B., Tao, Y., Xu, X., Li, F., Wu, K., Liang, J., Shao, D.*, et al.* (2012). Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. Cell *148*, 873-885.

Huang, C.R., Schneider, A.M., Lu, Y., Niranjan, T., Shen, P., Robinson, M.A., Steranka, J.P., Valle, D., Civin, C.I., Wang, T.*, et al.* (2010). Mobile interspersed repeats are major structural variants in the human genome. Cell *141*, 1171-1182.

Husemann, Y., Geigl, J.B., Schubert, F., Musiani, P., Meyer, M., Burghart, E., Forni, G., Eils, R., Fehm, T., Riethmuller, G.*, et al.* (2008). Systemic spread is an early step in breast cancer. Cancer cell *13*, 58-68.

Inagaki, H., Ohye, T., Kogo, H., Kato, T., Bolor, H., Taniguchi, M., Shaikh, T.H., Emanuel, B.S., and Kurahashi, H. (2009). Chromosomal instability mediated by non-B DNA: cruciform conformation and not DNA sequence is responsible for recurrent translocation in humans. Genome research *19*, 191-198.

Jain, N., Reuben, J.M., Kantarjian, H., Li, C., Gao, H., Lee, B.N., Cohen, E.N., Ebarb, T., Scheinberg, D.A., and Cortes, J. (2009). Synthetic tumor-specific breakpoint peptide vaccine in patients with chronic myeloid leukemia and minimal residual disease: a phase 2 trial. Cancer *115*, 3924-3934.

Jeffs, A.R., Wells, E., and Morris, C.M. (2001). Nonrandom distribution of interspersed repeat elements in the BCR and ABL1 genes and its relation to breakpoint cluster regions. Genes, chromosomes & cancer *32*, 144-154.

Johansson, B., Mertens, F., and Mitelman, F. (2004). Clinical and biological importance of cytogenetic abnormalities in childhood and adult acute lymphoblastic leukemia. Annals of medicine *36*, 492-503.

Joyce, J.A., and Pollard, J.W. (2009). Microenvironmental regulation of metastasis. Nature reviews Cancer *9*, 239-252.

Ju, B.G., Lunyak, V.V., Perissi, V., Garcia-Bassets, I., Rose, D.W., Glass, C.K., and Rosenfeld, M.G. (2006). A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription. Science *312*, 1798-1802.

Junttila, M.R., and Evan, G.I. (2009). p53--a Jack of all trades but master of none. Nature reviews Cancer *9*, 821-829.

Katapadi, V.K., Nambiar, M., and Raghavan, S.C. (2012). Potential G-quadruplex formation at breakpoint regions of chromosomal translocations in cancer may explain their fragility. Genomics *100*, 72-80.

Kato, T., Kurahashi, H., and Emanuel, B.S. (2012). Chromosomal translocations and palindromic AT-rich repeats. Current opinion in genetics & development *22*, 221-228.

Kawai, T., Hiroi, S., Nakanishi, K., and Meeker, A.K. (2007). Telomere length and telomerase expression in atypical adenomatous hyperplasia and small bronchioloalveolar carcinoma of the lung. American journal of clinical pathology *127*, 254-262.

Kearney, L. (2006). Multiplex-FISH (M-FISH): technique, developments and applications. Cytogenetic and genome research *114*, 189-198.

Kern, W., Schoch, C., Haferlach, T., and Schnittger, S. (2005). Monitoring of minimal residual disease in acute myeloid leukemia. Critical reviews in oncology/hematology *56*, 283-309.

Kibel, A.S., Freije, D., Isaacs, W.B., and Bova, G.S. (1999). Deletion mapping at 12p12-13 in metastatic prostate cancer. Genes, chromosomes & cancer *25*, 270-276.

Kibel, A.S., Schutte, M., Kern, S.E., Isaacs, W.B., and Bova, G.S. (1998). Identification of 12p as a region of frequent deletion in advanced prostate cancer. Cancer research *58*, 5652-5655.

Kim, S., Takahashi, H., Lin, W.W., Descargues, P., Grivennikov, S., Kim, Y., Luo, J.L., and Karin, M. (2009). Carcinoma-produced factors activate myeloid cells through TLR2 to stimulate metastasis. Nature *457*, 102-106.

King, J.C., Xu, J., Wongvipat, J., Hieronymus, H., Carver, B.S., Leung, D.H., Taylor, B.S., Sander, C., Cardiff, R.D., Couto, S.S*., et al.* (2009). Cooperativity of TMPRSS2-ERG with PI3-kinase pathway activation in prostate oncogenesis. Nature genetics *41*, 524-526.
Klein, C.A. (2009). Parallel progression of primary tumours and metastases. Nature reviews Cancer *9*, 302-312.

Klein, C.A. (2013). Selection and adaptation during metastatic cancer progression. Nature *501*, 365-372.

Klein, C.A., Blankenstein, T.J., Schmidt-Kittler, O., Petronio, M., Polzer, B., Stoecklein, N.H., and Riethmuller, G. (2002). Genetic heterogeneity of single disseminated tumour cells in minimal residual cancer. Lancet *360*, 683-689.

Klein, C.A., and Holzel, D. (2006). Systemic cancer progression and tumor dormancy: mathematical models meet single cell genomics. Cell cycle *5*, 1788-1798.

Klein, C.A., Schmidt-Kittler, O., Schardt, J.A., Pantel, K., Speicher, M.R., and Riethmuller, G. (1999). Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. Proceedings of the National Academy of Sciences of the United States of America *96*, 4494-4499.

Klein, I.A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D.F*., et al.* (2011). Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. Cell *147*, 95-106.

Kohne, D.E., Levison, S.A., and Byers, M.J. (1977). Room temperature method for increasing the rate of DNA reassociation by many thousandfold: the phenol emulsion reassociation technique. Biochemistry *16*, 5329-5341.

Kolomietz, E., Meyn, M.S., Pandita, A., and Squire, J.A. (2002). The role of Alu repeat clusters as mediators of recurrent chromosomal aberrations in tumors. Genes, chromosomes & cancer *35*, 97-112.

Kubuschok, B., Passlick, B., Izbicki, J.R., Thetter, O., and Pantel, K. (1999). Disseminated tumor cells in lymph nodes as a determinant for survival in surgically resected non-small-cell lung cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *17*, 19-24.

Kurahashi, H., Inagaki, H., Ohye, T., Kogo, H., Tsutsumi, M., Kato, T., Tong, M., and Emanuel, B.S. (2010). The constitutional t(11;22): implications for a novel mechanism responsible for gross chromosomal rearrangements. Clinical genetics *78*, 299-309.

Kvistborg, P., van Buuren, M.M., and Schumacher, T.N. (2013). Human cancer regression antigens. Current opinion in immunology *25*, 284-290.

Laman, A.G., Kurjukov, S.G., Bulgakova, E.V., Anikeeva, N.N., and Brovko, F.A. (2001). Subtractive hybridization of biotinylated DNA in phenol emulsion. Journal of biochemical and biophysical methods *50*, 43-52.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lawson, A.R., Hindley, G.F., Forshew, T., Tatevossian, R.G., Jamie, G.A., Kelly, G.P., Neale, G.A., Ma, J., Jones, T.A., Ellison, D.W.*, et al.* (2011). RAF gene fusion breakpoints in pediatric brain tumors are characterized by significant enrichment of sequence microhomology. Genome research *21*, 505-514.

Lemmon, M.A., and Schlessinger, J. (2010). Cell signaling by receptor tyrosine kinases. Cell *141*, 1117-1134.

Lengauer, C., Kinzler, K.W., and Vogelstein, B. (1998). Genetic instabilities in human cancers. Nature *396*, 643-649.

Lessene, G., Czabotar, P.E., and Colman, P.M. (2008). BCL-2 family antagonists for cancer therapy. Nature reviews Drug discovery *7*, 989-1000.

Lianidou, E.S., Strati, A., and Markou, A. (2014). Circulating tumor cells as promising novel biomarkers in solid cancers. Critical reviews in clinical laboratory sciences *51*, 160-171.

Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annual review of biochemistry *79*, 181-211.

Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B.G., Ohgi, K., Zhang, J., Rose, D.W., Fu, X.D., Glass, C.K.*, et al.* (2009). Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell *139*, 1069-1083.

Lisitsyn, N., Lisitsyn, N., and Wigler, M. (1993). Cloning the differences between two complex genomes. Science *259*, 946-951.

Lisitsyn, N.A., Lisitsina, N.M., Dalbagni, G., Barker, P., Sanchez, C.A., Gnarra, J., Linehan, W.M., Reid, B.J., and Wigler, M.H. (1995). Comparative genomic analysis of tumors: detection of DNA losses and amplification. Proceedings of the National Academy of Sciences of the United States of America *92*, 151-155.

Lisitsyn, N.A., Segre, J.A., Kusumi, K., Lisitsyn, N.M., Nadeau, J.H., Frankel, W.N., Wigler, M.H., and Lander, E.S. (1994). Direct isolation of polymorphic markers linked to a trait by genetically directed representational difference analysis. Nature genetics *6*, 57-63.

Litterman, A.J., Zellmer, D.M., Grinnen, K.L., Hunt, M.A., Dudek, A.Z., Salazar, A.M., and Ohlfest, J.R. (2013). Profound impairment of adaptive immune responses by alkylating chemotherapy. Journal of immunology *190*, 6259-6268.

Liu, D., Aguirre Ghiso, J., Estrada, Y., and Ossowski, L. (2002). EGFR is a transducer of the urokinase receptor initiated signal that is required for in vivo growth of a human carcinoma. Cancer cell *1*, 445-457.

Liu, W., Chang, B., Sauvageot, J., Dimitrov, L., Gielzak, M., Li, T., Yan, G., Sun, J., Sun, J., Adams, T.S.*, et al.* (2006). Comprehensive assessment of DNA copy number alterations in human prostate cancers using Affymetrix 100K SNP mapping array. Genes, chromosomes & cancer *45*, 1018-1032.

Lukasova, E., Kozubek, S., Kozubek, M., Kjeronska, J., Ryznar, L., Horakova, J., Krahulcova, E., and Horneck, G. (1997). Localisation and distance between ABL and BCR genes in interphase nuclei of bone marrow cells of control donors and patients with chronic myeloid leukaemia. Human genetics *100*, 525-535.

Luzzi, K.J., MacDonald, I.C., Schmidt, E.E., Kerkvliet, N., Morris, V.L., Chambers, A.F., and Groom, A.C. (1998). Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. The American journal of pathology *153*, 865-873.

Magbanua, M.J., Sosa, E.V., Roy, R., Eisenbud, L.E., Scott, J.H., Olshen, A., Pinkel, D., Rugo, H.S., and Park, J.W. (2013). Genomic profiling of isolated circulating tumor cells from metastatic breast cancer patients. Cancer research *73*, 30-40.

Magbanua, M.J., Sosa, E.V., Scott, J.H., Simko, J., Collins, C., Pinkel, D., Ryan, C.J., and Park, J.W. (2012). Isolation and genomic analysis of circulating tumor cells from castration resistant metastatic prostate cancer. BMC cancer *12*, 78.

Malanchi, I., Santamaria-Martinez, A., Susanto, E., Peng, H., Lehr, H.A., Delaloye, J.F., and Huelsken, J. (2012). Interactions between cancer stem cells and their niche govern metastatic colonization. Nature *481*, 85-89.

Mathas, S., Kreher, S., Meaburn, K.J., Johrens, K., Lamprecht, B., Assaf, C., Sterry, W., Kadin, M.E., Daibata, M., Joos, S.*, et al.* (2009). Gene deregulation and spatial genome reorganization near breakpoints prior to formation of translocations in anaplastic large cell lymphoma. Proceedings of the National Academy of Sciences of the United States of America *106*, 5831-5836.

Meaburn, K.J., Misteli, T., and Soutoglou, E. (2007). Spatial genome organization in the formation of chromosomal translocations. Seminars in cancer biology *17*, 80-90.

Meek, D.W. (2009). Tumour suppression by p53: a role for the DNA damage response? Nature reviews Cancer *9*, 714-723.

Metzker, M.L. (2010). Sequencing technologies - the next generation. Nature reviews Genetics *11*, 31-46.

Meyerhans, A., Vartanian, J.P., and Wain-Hobson, S. (1990). DNA recombination during PCR. Nucleic acids research *18*, 1687-1691.

Miller, R.D., and Riblet, R. (1995). Improved phenol emulsion DNA reassociation technique (PERT) using thermal cycling. Nucleic acids research *23*, 2339-2340.

Milner, J.J., Cecchini, E., and Dominy, P.J. (1995). A kinetic model for subtractive hybridization. Nucleic acids research *23*, 176-187.

Mitelman, F., Johansson, B., and Mertens, F. (2004). Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. Nature genetics *36*, 331-334.

Mitelman, F., Johansson, B., and Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation. Nature reviews Cancer *7*, 233-245.
Mulligan, L.M. (2014). RET revisited: expanding the oncogenic portfolio. Nature reviews Cancer *14*, 173-186.

Nadezhdin, E.V., Lebedev, Y.B., Glazkova, D.V., Bornholdt, D., Arman, I.P., Grzeschik, K.H., Hunsmann, G., and Sverdlov, E.D. (2001). Identification of paralogous HERV-K LTRs on human chromosomes 3, 4, 7 and 11 in regions containing clusters of olfactory receptor genes. Molecular genetics and genomics : MGG *265*, 820-825.

Navin, A., Prekeris, R., Lisitsyn, N.A., Sonti, M.M., Grieco, D.A., Narayanswami, S., Lander, E.S., and Simpson, E.M. (1996). Mouse Y-specific repeats isolated by whole chromosome representational difference analysis. Genomics *36*, 349-353.

Nikiforov, Y.E., Koshoffer, A., Nikiforova, M., Stringer, J., and Fagin, J.A. (1999). Chromosomal breakpoint positions suggest a direct role for radiation in inducing illegitimate recombination between the ELE1 and RET genes in radiation-induced thyroid carcinomas. Oncogene *18*, 6330-6334.

Nyberg, P., Xie, L., and Kalluri, R. (2005). Endogenous inhibitors of angiogenesis. Cancer research *65*, 3967-3979.

Okazaki, I.M., Kotani, A., and Honjo, T. (2007). Role of AID in tumorigenesis. Advances in immunology *94*, 245-273.

Oskarsson, T., Acharyya, S., Zhang, X.H., Vanharanta, S., Tavazoie, S.F., Morris, P.G., Downey, R.J., Manova-Todorova, K., Brogi, E., and Massague, J. (2011). Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs. Nature medicine *17*, 867-874.

Ozeri-Galai, E., Bester, A.C., and Kerem, B. (2012). The complex basis underlying common fragile site instability in cancer. Trends in genetics : TIG *28*, 295-302.

Passlick, B., Izbicki, J.R., Kubuschok, B., Nathrath, W., Thetter, O., Pichlmeier, U., Schweiberer, L., Riethmuller, G., and Pantel, K. (1994). Immunohistochemical assessment of individual tumor cells in lymph nodes of patients with non-small-cell lung cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *12*, 1827-1832.

Passlick, B., Kubuschok, B., Izbicki, J.R., Thetter, O., and Pantel, K. (1999). Isolated tumor cells in bone marrow predict reduced survival in node-negative non-small cell lung cancer. The Annals of thoracic surgery *68*, 2053-2058.

Perner, S., Mosquera, J.M., Demichelis, F., Hofer, M.D., Paris, P.L., Simko, J., Collins, C., Bismar, T.A., Chinnaiyan, A.M., De Marzo, A.M., *et al.* (2007). TMPRSS2-ERG fusion prostate cancer: an early molecular event associated with invasion. The American journal of surgical pathology *31*, 882-888.

Pierga, J.Y., Bonneton, C., Vincent-Salomon, A., de Cremoux, P., Nos, C., Blin, N., Pouillart, P., Thiery, J.P., and Magdelenat, H. (2004). Clinical significance of immunocytochemical detection of tumor cells using digital microscopy in peripheral blood and bone marrow of breast cancer patients. Clinical cancer research : an official journal of the American Association for Cancer Research *10*, 1392-1400.

Pierotti, M.A. (2001). Chromosomal rearrangements in thyroid carcinomas: a recombination or death dilemma. Cancer letters *166*, 1-7.

Pierotti, M.A., Santoro, M., Jenkins, R.B., Sozzi, G., Bongarzone, I., Grieco, M., Monzini, N., Miozzo, M., Herrmann, M.A., Fusco, A., *et al.* (1992). Characterization of an inversion on the long arm of chromosome 10 juxtaposing D10S170 and RET and creating the oncogenic sequence RET/PTC. Proceedings of the National Academy of Sciences of the United States of America *89*, 1616-1620.

Png, K.J., Halberg, N., Yoshida, M., and Tavazoie, S.F. (2012). A microRNA regulon that mediates endothelial recruitment and metastasis by cancer cells. Nature *481*, 190-194.

Polz, M.F., and Cavanaugh, C.M. (1998). Bias in template-to-product ratios in multitemplate PCR. Applied and environmental microbiology *64*, 3724-3730.

Polzer, B., Medoro, G., Pasch, S., Fontana, F., Zorzino, L., Pestka, A., Andergassen, U., Meier-Stiegen, F., Czyz, Z.T., Alberter, B., *et al.* (2014). Molecular profiling of single circulating tumor cells with diagnostic intention. EMBO molecular medicine *6*, 1371-1386.

Povirk, L.F. (2006). Biochemical mechanisms of chromosomal translocations resulting from DNA double-strand breaks. DNA repair *5*, 1199-1212.

Psaila, B., and Lyden, D. (2009). The metastatic niche: adapting the foreign soil. Nature reviews Cancer *9*, 285-293.

Qiu, X., Wu, L., Huang, H., McDonel, P.E., Palumbo, A.V., Tiedje, J.M., and Zhou, J. (2001). Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. Applied and environmental microbiology *67*, 880-887.

Rabbitts, T.H. (2009). Commonality but diversity in cancer gene fusions. Cell *137*, 391-395.

Raghavan, S.C., Swanson, P.C., Wu, X., Hsieh, C.L., and Lieber, M.R. (2004). A non-B-DNA structure at the Bcl-2 major breakpoint region is cleaved by the RAG complex. Nature *428*, 88-93.

Raica, M., Cimpean, A.M., and Ribatti, D. (2009). Angiogenesis in pre-malignant conditions. European journal of cancer *45*, 1924-1934.

Rajasagi, M., Shukla, S.A., Fritsch, E.F., Keskin, D.B., DeLuca, D., Carmona, E., Zhang, W., Sougnez, C., Cibulskis, K., Sidney, J.*, et al.* (2014). Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. Blood *124*, 453-462.

Rego, E.M., Ruggero, D., Tribioli, C., Cattoretti, G., Kogan, S., Redner, R.L., and Pandolfi, P.P. (2006). Leukemia with distinct phenotypes in transgenic mice expressing PML/RAR alpha, PLZF/RAR alpha or NPM/RAR alpha. Oncogene *25*, 1974-1979.

Rhim, A.D., Mirek, E.T., Aiello, N.M., Maitra, A., Bailey, J.M., McAllister, F., Reichert, M., Beatty, G.L., Rustgi, A.K., Vonderheide, R.H.*, et al.* (2012). EMT and dissemination precede pancreatic tumor formation. Cell *148*, 349-361.

Riethdorf, S., and Pantel, K. (2010). Advancing personalized cancer therapy by detection and characterization of circulating carcinoma cells. Annals of the New York Academy of Sciences *1210*, 66-77.

Rikova, K., Guo, A., Zeng, Q., Possemato, A., Yu, J., Haack, H., Nardone, J., Lee, K., Reeves, C., Li, Y.*, et al.* (2007). Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. Cell *131*, 1190-1203.

Robbiani, D.F., Bothmer, A., Callen, E., Reina-San-Martin, B., Dorsett, Y., Difilippantonio, S., Bolland, D.J., Chen, H.T., Corcoran, A.E., Nussenzweig, A.*, et al.* (2008). AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. Cell *135*, 1028-1038.

Roccato, E., Bressan, P., Sabatella, G., Rumio, C., Vizzotto, L., Pierotti, M.A., and Greco, A. (2005). Proximity of TPR and NTRK1 rearranging loci in human thyrocytes. Cancer research *65*, 2572-2576.

Roix, J.J., McQueen, P.G., Munson, P.J., Parada, L.A., and Misteli, T. (2003). Spatial proximity of translocation-prone gene loci in human lymphomas. Nature genetics *34*, 287-291.

Roukos, V., Burman, B., and Misteli, T. (2013). The cellular etiology of chromosome translocations. Current opinion in cell biology *25*, 357-364.

Rowley, J.D. (1973). Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature *243*, 290-293.

Rubin, C.M., Leeflang, E.P., Rinehart, F.P., and Schmid, C.W. (1993). Paucity of novel short interspersed repetitive element (SINE) families in human DNA and isolation of a novel MER repeat. Genomics *18*, 322-328.

Sanger, N., Effenberger, K.E., Riethdorf, S., Van Haasteren, V., Gauwerky, J., Wiegratz, I., Strebhardt, K., Kaufmann, M., and Pantel, K. (2011). Disseminated tumor cells in the bone marrow of patients with ductal carcinoma in situ. International journal of cancer Journal international du cancer *129*, 2522-2526.

Sasaki, H., Nomura, S., Akiyama, N., Takahashi, A., Sugimura, T., Oishi, M., and Terada, M. (1994). Highly efficient method for obtaining a subtracted genomic DNA library by the modified in-gel competitive reassociation method. Cancer research *54*, 5821-5823.

Schardt, J.A., Meyer, M., Hartmann, C.H., Schubert, F., Schmidt-Kittler, O., Fuhrmann, C., Polzer, B., Petronio, M., Eils, R., and Klein, C.A. (2005). Genomic analysis of single cytokeratin-positive cells from bone marrow reveals early mutational events in breast cancer. Cancer cell *8*, 227-239.

Scheich, F., Duyster, J., Peschel, C., and Bernhard, H. (2007). The immunogenicity of Bcr-Abl expressing dendritic cells is dependent on the Bcr-Abl kinase activity and dominated by Bcr-Abl regulated antigens. Blood *110*, 2556-2560.

Schmidt-Kittler, O., Ragg, T., Daskalakis, A., Granzow, M., Ahr, A., Blankenstein, T.J., Kaufmann, M., Diebold, J., Arnholdt, H., Muller, P.*, et al.* (2003). From latent disseminated cells to overt metastasis: genetic analysis of systemic breast cancer progression. Proceedings of the National Academy of Sciences of the United States of America *100*, 7737-7742.

Schutte, M., da Costa, L.T., Hahn, S.A., Moskaluk, C., Hoque, A.T., Rozenblum, E., Weinstein, C.L., Bittner, M., Meltzer, P.S., Trent, J.M.*, et al.* (1995). Identification by representational difference analysis of a homozygous deletion in pancreatic carcinoma that lies within the BRCA2 region. Proceedings of the National Academy of Sciences of the United States of America *92*, 5950-5954.

Shagina, I., Bogdanova, E., Mamedov, I.Z., Lebedev, Y., Lukyanov, S., and Shagin, D. (2010). Normalization of genomic DNA using duplex-specific nuclease. BioTechniques *48*, 455-459.

Shibue, T., and Weinberg, R.A. (2009). Integrin beta1-focal adhesion kinase signaling directs the proliferation of metastatic cancer cells disseminated in the lungs. Proceedings of the National Academy of Sciences of the United States of America *106*, 10290-10295.

Shrivastav, M., De Haro, L.P., and Nickoloff, J.A. (2008). Regulation of DNA double-strand break repair pathway choice. Cell research *18*, 134-147.

Sive, H.L., and St John, T. (1988). A simple subtractive hybridization technique employing photoactivatable biotin and phenol extraction. Nucleic acids research *16*, 10937.

Sleeman, J., and Steeg, P.S. (2010). Cancer metastasis as a therapeutic target. European journal of cancer *46*, 1177-1180.

So, C.W., Ma, Z.G., Price, C.M., Dong, S., Chen, S.J., Gu, L.J., So, C.K., Wiedemann, L.M., and Chan, L.C. (1997). MLL self fusion mediated by Alu repeat homologous recombination and prognosis of AML-M4/M5 subtypes. Cancer research *57*, 117-122.

Soda, M., Choi, Y.L., Enomoto, M., Takada, S., Yamashita, Y., Ishikawa, S., Fujiwara, S., Watanabe, H., Kurashina, K., Hatanaka, H.*, et al.* (2007). Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature *448*, 561-566.

Soller, M.J., Isaksson, M., Elfving, P., Soller, W., Lundgren, R., and Panagopoulos, I. (2006). Confirmation of the high frequency of the TMPRSS2/ERG fusion gene in prostate cancer. Genes, chromosomes & cancer *45*, 717-719.

Stavnezer, J., Guikema, J.E., and Schrader, C.E. (2008). Mechanism and regulation of class switch recombination. Annual review of immunology *26*, 261-292.

Stoecklein, N.H., Hosch, S.B., Bezler, M., Stern, F., Hartmann, C.H., Vay, C., Siegmund, A., Scheunemann, P., Schurr, P., Knoefel, W.T.*, et al.* (2008). Direct genetic analysis of single disseminated cancer cells for prediction of outcome and therapy selection in esophageal cancer. Cancer cell *13*, 441-453.

Stoecklein, N.H., and Klein, C.A. (2010). Genetic disparity between primary tumours, disseminated tumour cells, and manifest metastasis. International journal of cancer Journal international du cancer *126*, 589-598.

Straus, D., and Ausubel, F.M. (1990). Genomic subtraction for cloning DNA corresponding to deletion mutations. Proceedings of the National Academy of Sciences of the United States of America *87*, 1889-1893.

Sun, C., Dobi, A., Mohamed, A., Li, H., Thangapazham, R.L., Furusato, B., Shaheduzzaman, S., Tan, S.H., Vaidyanathan, G., Whitman, E.*, et al.* (2008). TMPRSS2-ERG fusion, a common genomic alteration in prostate cancer activates C-MYC and abrogates prostate epithelial differentiation. Oncogene *27*, 5348-5353.

Sun, T., Goodman, H.M., and Ausubel, F.M. (1992). Cloning the Arabidopsis GA1 Locus by Genomic Subtraction. The Plant cell *4*, 119-128.

Taki, T., and Taniwaki, M. (2006). Chromosomal translocations in cancer and their relevance for therapy. Current opinion in oncology *18*, 62-68.

Telenius, H., Carter, N.P., Bebb, C.E., Nordenskjold, M., Ponder, B.A., and Tunnacliffe, A. (1992). Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. Genomics *13*, 718-725.

Thiery, J.P., Acloque, H., Huang, R.Y., and Nieto, M.A. (2009). Epithelial-mesenchymal transitions in development and disease. Cell *139*, 871-890.

Thomas, M., Greil, J., and Heidenreich, O. (2006). Targeting leukemic fusion proteins with small interfering RNAs: recent advances and therapeutic potentials. Acta pharmacologica Sinica *27*, 273-281.

Thorban, S., Roder, J.D., Nekarda, H., Funk, A., Siewert, J.R., and Pantel, K. (1996). Immunocytochemical detection of disseminated tumor cells in the bone marrow of patients with esophageal carcinoma. Journal of the National Cancer Institute *88*, 1222-1227.

Tomlins, S.A., Laxman, B., Varambally, S., Cao, X., Yu, J., Helgeson, B.E., Cao, Q., Prensner, J.R., Rubin, M.A., Shah, R.B.*, et al.* (2008). Role of the TMPRSS2-ERG gene fusion in prostate cancer. Neoplasia *10*, 177-188.

Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R.*, et al.* (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science *310*, 644-648.

Travis, G.H., and Sutcliffe, J.G. (1988). Phenol emulsion-enhanced DNA-driven subtractive cDNA cloning: isolation of low-abundance monkey cortex-specific mRNAs. Proceedings of the National Academy of Sciences of the United States of America *85*, 1696-1700.

Tsai, A.G., Lu, H., Raghavan, S.C., Muschen, M., Hsieh, C.L., and Lieber, M.R. (2008). Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. Cell *135*, 1130-1142.

Ulmer, A., Schmidt-Kittler, O., Fischer, J., Ellwanger, U., Rassner, G., Riethmuller, G., Fierlbeck, G., and Klein, C.A. (2004). Immunomagnetic enrichment, genomic characterization, and prognostic impact of circulating melanoma cells. Clinical cancer research : an official journal of the American Association for Cancer Research *10*, 531-537.

Vaandrager, J.W., Schuuring, E., Philippo, K., and Kluin, P.M. (2000). V(D)J recombinase-mediated transposition of the BCL2 gene to the IGH locus in follicular lymphoma. Blood *96*, 1947-1952.

Valastyan, S., and Weinberg, R.A. (2011). Tumor metastasis: molecular insights and evolving paradigms. Cell *147*, 275-292.

Vashist, Y.K., Effenberger, K.E., Vettorazzi, E., Riethdorf, S., Yekebas, E.F., Izbicki, J.R., and Pantel, K. (2012). Disseminated tumor cells in bone marrow and the natural course of resected esophageal cancer. Annals of surgery *255*, 1105-1112.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A.*, et al.* (2001). The sequence of the human genome. Science *291*, 1304-1351.

Vogelstein, B., Fearon, E.R., Hamilton, S.R., Kern, S.E., Preisinger, A.C., Leppert, M., Nakamura, Y., White, R., Smits, A.M., and Bos, J.L. (1988). Genetic alterations during colorectal-tumor development. The New England journal of medicine *319*, 525-532.

Waldman, F.M., DeVries, S., Chew, K.L., Moore, D.H., 2nd, Kerlikowske, K., and Ljung, B.M. (2000). Chromosomal alterations in ductal carcinomas in situ and their in situ recurrences. Journal of the National Cancer Institute *92*, 313-320.

Wan, L., Pantel, K., and Kang, Y. (2013). Tumor metastasis: moving new biological insights into the clinic. Nature medicine *19*, 1450-1464.

Wang, J., Lu, Y., Wang, J., Koch, A.E., Zhang, J., and Taichman, R.S. (2008). CXCR6 induces prostate cancer progression by the AKT/mammalian target of rapamycin signaling pathway. Cancer research *68*, 10367-10376.

Weckermann, D., Muller, P., Wawroschek, F., Harzmann, R., Riethmuller, G., and Schlimok, G. (2001). Disseminated cytokeratin positive tumor cells in the bone marrow of patients with prostate cancer: detection and prognostic value. The Journal of urology *166*, 699-703.

Weckermann, D., Polzer, B., Ragg, T., Blana, A., Schlimok, G., Arnholdt, H., Bertz, S., Harzmann, R., and Klein, C.A. (2009). Perioperative activation of disseminated tumor cells in bone marrow of patients with prostate cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *27*, 1549-1556.

Wei, Y., Sun, M., Nilsson, G., Dwight, T., Xie, Y., Wang, J., Hou, Y., Larsson, O., Larsson, C., and Zhu, X. (2003). Characteristic sequence motifs located at the genomic breakpoints of the translocation t(X;18) in synovial sarcomas. Oncogene *22*, 2215-2222.

Wieder, R., and Wetmur, J.G. (1982). Factors affecting the kinetics of DNA reassociation in phenol-water emulsion at high DNA concentrations. Biopolymers *21*, 665-677.

Wieland, I., Bolger, G., Asouline, G., and Wigler, M. (1990). A method for difference cloning: gene amplification following subtractive hybridization. Proceedings of the National Academy of Sciences of the United States of America *87*, 2720-2724.

Williams, R., Peisajovich, S.G., Miller, O.J., Magdassi, S., Tawfik, D.S., and Griffiths, A.D. (2006). Amplification of complex gene libraries by emulsion PCR. Nature methods *3*, 545-550.

Wolchok, J.D., and Chan, T.A. (2014). Cancer: Antitumour immunity gets a boost. Nature *515*, 496-498.

Wong, C.W., Lee, A., Shientag, L., Yu, J., Dong, Y., Kao, G., Al-Mehdi, A.B., Bernhard, E.J., and Muschel, R.J. (2001). Apoptosis: an early event in metastatic inefficiency. Cancer research *61*, 333-338.

Wong, K.K., Jr., and Chatterjee, S. (2005). Vaccine development for chronic myelogenous leukaemia. Lancet *365*, 631-632.

Worley, B.S., van den Broeke, L.T., Goletz, T.J., Pendleton, C.D., Daschbach, E.M., Thomas, E.K., Marincola, F.M., Helman, L.J., and Berzofsky, J.A. (2001). Antigenicity of fusion proteins from sarcoma-associated chromosomal translocations. Cancer research *61*, 6868-6875.

Yokota, H., Iwasaki, T., Takahashi, M., and Oishi, M. (1989). A tissue-specific change in repetitive DNA in rats. Proceedings of the National Academy of Sciences of the United States of America *86*, 9233-9237.

Yokota, H., and Oishi, M. (1990). Differential cloning of genomic DNA: cloning of DNA with an altered primary structure by in-gel competitive reassociation. Proceedings of the National Academy of Sciences of the United States of America *87*, 6398-6402.

Yu, J., Yu, J., Mani, R.S., Cao, Q., Brenner, C.J., Cao, X., Wang, X., Wu, L., Li, J., Hu, M.*, et al.* (2010). An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. Cancer cell *17*, 443-454.

Zech, L., Haglund, U., Nilsson, K., and Klein, G. (1976). Characteristic chromosomal abnormalities in biopsies and lymphoid-cell lines from patients with Burkitt and non-Burkitt lymphomas. International journal of cancer Journal international du cancer *17*, 47-56.

Zeschnigk, M., Horsthemke, B., and Lohmann, D. (1999). Detection of homozygous deletions in tumors by hybridization of representational difference analysis (RDA) products to chromosome-specific YAC clone arrays. Nucleic acids research *27*, e30.

Zhang, L., Cui, X., Schmitt, K., Hubert, R., Navidi, W., and Arnheim, N. (1992). Whole genome amplification from a single cell: implications for genetic analysis. Proceedings of the National Academy of Sciences of the United States of America *89*, 5847-5851.

Zhang, X.H., Wang, Q., Gerald, W., Hudis, C.A., Norton, L., Smid, M., Foekens, J.A., and Massague, J. (2009). Latent bone metastasis in breast cancer tied to Src-dependent survival signals. Cancer cell *16*, 67-78.

Zong, C., Lu, S., Chapman, A.R., and Xie, X.S. (2012). Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science *338*, 1622-1626.

Zong, Y., Xin, L., Goldstein, A.S., Lawson, D.A., Teitell, M.A., and Witte, O.N. (2009). ETS family transcription factors collaborate with alternative signaling pathways to induce carcinoma from adult murine prostate cells. Proceedings of the National Academy of Sciences of the United States of America *106*, 12465-12470.

8 Acknowledgments

I would like to thank all the people who contributed in some way to the work described in this thesis. First and foremost, I thank my academic advisor, Prof. Dr. Christoph Klein, for accepting me into his group. His truly scientist intuition has made him as a constant oasis of ideas and passions in science, which exceptionally inspire and enrich my growth as a researcher and scientist. I am indebted to him more than he knows.

Additionally, I would like to thank my mentors Prof. Dr. Gernot Längst and Prof. Dr. Jürgen Ruland for the interest in my work, and the members of the committee Prof Dr. Thomas Dresselhaus and Prof. Dr. Gunter Meister.

I thank all my fellow lab-mates and technicians for their precious support and the serene working atmosphere. I gained a lot from them, through their personal and scholarly interactions, their suggestions at various points of my research program.

I thank Dr. Stefan Kirsch, Dr. Miodrag Gužvić and Dr. Elisabeth Doblinger for carefully reading and for giving comments and suggestions that have been helpful to improve the thesis.

Last but not least, I thank my family and friends for their invaluable support in every possible way.

9. Curriculum Vitae

Giancarlo Feliciello

Date of birth 12.08.1982

Place of birth: Ottaviano, Italy

EDUCATIONAL TRAINING

01.2008 - 02.2015

PhD training

Executed at the Chair of Experimental Medicine and Therapy Research, University Clinic of Regensburg

Topic: Identification of genomic breakpoints in single tumor cells

Supervised by: Prof. Dr. Christoph A. Klein

01.2007 - 05.2008

Post Master degree training

Executed at Bayer CropScience Frankfurt am Main

Topic: Genetic analysis of sugar beet with herbicide resistance, mutant isolation in oilseed rape

Supervised by: Dr. Günter Donn

09.2004 – 10.2007

Master Degree Training

Executed at Ce.In.Ge.-Center for Genetic Engineering, Naples

Topic: Genetic analysis of macrodeletions in phenylalanine hydroxylase gene

Supervised by: Prof Dr. Aurora Daniele

EDUCATION

10.2001 – 03.2007

University of Naples “Federico II”, Naples

Master Degree in Medical Biotechnology with honours

09.1996 – 07.2001

High school for Math and Science, “E. Pascal”, Pompei

High School Diploma, mark 80/100