UNIVERSITÀ DEGLI STUDI DI MILANO
FACOLTÀ DI MEDICINA E CHIRURGIA

Dipartimento Di Biotecnologie Mediche E Medicina Traslazionale

PhD course in Experimental Medicine and Medical Biotechnologies
Immunology and Hemato-Oncology
PhD cohort: XXXI

MONITORING MINIMAL RESIDUAL DISEASE IN MULTIPLE MYELOMA PATIENTS BY NEXT-GENERATION SEQUENCING APPROACHES

Med/15 - Malattie del Sangue

Giulia Biancon
R11335

Tutors: Prof. Paolo Corradini
Dr. Cristiana Carniti

PhD program coordinator: Prof. Massimo Locati

A.A. 2017-2018

SUMMARY

ABBREVIATIONS

ABC, Activated B-Cell-like
ABVD, Adriamycin/Bleomycin/Vinblastine/Dacarbazine
Ag, Antigen
ASCT, Autologous Stem Cell Transplantation
ASO-qPCR, Allele-Specific Oligonucleotide - Quantitative Polymerase Chain Reaction
BCR, B-Cell Receptor
BL, Burkitt Lymphoma
BM, Bone Marrow
BMB, Bone Marrow Biopsy
B-NHL, B-cell Non-Hodgkin Lymphoma
bp, base pair
BSA, Bovine Serum Albumin
CA, Chromosomal Abnormalities
CDR, Complementarity-Determining Region
cfDNA, cell-free Deoxyribonucleic Acid
cHL, classical HL
CLL, Chronic Lymphocytic Leukemia
CR, Complete Response
CSR, Class-Switch Recombination
CT, Computed Tomography
CTC, Circulating Tumor Cell
dNTP, deoxynucleotide
ddNTP, dideoxynucleotide
ddPCR, droplet digital Polymerase Chain Reaction
DLBCL, Diffuse Large B-Cell Lymphoma
DSB, Double Strand Break
DSS, Durie-Salmon Staging system
EBER, Epstein-Barr virus Encoded RNA
EBV, Epstein-Barr Virus
ECOG, Eastern Cooperative Oncology Group
EOT, End Of Treatment
EtOH, Ethanol
FAM, 6-carboxy Fluorescein
FDG, [^{18}F] 2-Fluoro-2-Deoxy-D-Glucose
FFPE, Formalin-Fixed Paraffin-Embedded
FISH, Fluorescent In Situ Hybridization
FL, Follicular Lymphoma
FR, Framework Region
FSC, Forward Scatter
FU, Fluorescence Unit
GC, Germinal Center
GCB, Germinal Center B-Cell-like
gDNA, genomic Deoxyriboucleic Acid
HCL, Hairy Cell Leukemia
HL, Hodgkin Lymphoma
HRS, Hodgkin and Reed-Sternberg cells
IG, Immunoglobulin
IGH, Immunoglobulin Heavy Chain
IGL, Immunoglobulin Light Chain
IMiDs, Immunomodulatory Drugs
IMWG, International Myeloma Working Group
IPI, International Prognostic Index
IPS, International Prognostic Score
IRB, Institutional Review Board
ISS, International Staging System
LDH, Lactate Dehydrogenase
MALT, Mucosa-Associated Lymphoid Tissue
MCL, Mantle Cell Lymphoma
MFC, Multiparameter Flow Cytometry
MGUS, Monoclonal Gammopathy of Undetermined Significance
MM, Multiple Myeloma
MRD, Minimal Residual Disease
MRI, Magnetic Resonance Imaging
MZ, Marginal Zone
MZL, Marginal Zone Lymphoma
NGS, Next-Generation Sequencing
NHEJ, Non-Homologous DNA End-Joining
NHL, Non-Hodgkin Lymphoma

NLPHL, Nodular Lymphocyte-Predominant HL
NMZL, Nodal Marginal Zone Lymphoma
NTC, No-Template Control
OS, Overall Survival
PB, Peripheral Blood
PBS, Phosphate Buffered Saline
PC, Plasma Cell
PET, Positron Emission Tomography
PFS, Progression-Free Survival
PD, Progressive Disease
PGM, Personal Genome Machine
PIs, Proteasome Inhibitors
PR, Partial Response
QA, Quality Assessment
RAG, Recombination Activating Gene
RCD, lenalidomide(Revlimid)/ Cyclophosphamide/Dexamethasone
R-CHOP, Rituximab-Cyclophosphamide/Doxorubicin/Vincristine/Prednisone
R-ISS, Revised-International Staging System
RSS, Recombination Signal Sequence
RT, Room Temperature
SD, Stable Disease
sFLC, serum Free-Light Chain
SH, Somatic Hypermutation
SMM, Smoldering Multiple Myeloma
SMZL, Splenic Marginal Zone Lymphoma
SPEP, Serum Protein Electrophoresis
SSC, Side Scatter
SUVmax, highest Standardized Uptake Volume
TAMRA, 6-carboxy-Tetramethyl Rhodamine
tcDNA, tumor cell Deoxyribonucleic Acid
TdT, Terminal deoxynucleotidyl Transferase
TTP, Time To Progression
VCD, bortezomib(Velcade)/ Cyclophosphamide/Dexamethasone
V(D)J, Variable (Diversity) Joining regions
VGPR, Very Good Partial Response
WHO, World Health Organization

LIST OF FIGURES AND TABLES

ABSTRACT

Novel treatments for multiple myeloma (MM) have increased rates of complete response (CR) raising interest in more accurate methods to evaluate residual disease. Cell-free tumor DNA (cfDNA) analysis could represent a minimally invasive approach complementary to multiparameter flow cytometry (MFC) and molecular methods on bone marrow (BM) aspirates. A sequencing approach using the Ion Torrent Personal Genome Machine was applied to identify clonal immunoglobulin heavy chain (IGH) gene rearrangements in tumor plasma cells (PCs) and in serial plasma samples of 25 MM patients receiving second-line therapy. The same clonal IGH rearrangement identified in tumor PCs was detected in paired plasma samples and levels of IGH cfDNA correlated with outcome and mirrored tumor dynamics evaluated using conventional laboratory parameters. In addition, IGH cfDNA levels reflected the number of PCs enumerated by MFC immunophenotyping even in the CR context. Minimal residual disease (MRD)-negative patients by MFC were characterized by low frequencies of tumor clonotypes in cfDNA and longer survival. Despite the limited sample size, results showed that our developed deep-sequencing workflow is feasible for the identification and the monitoring of IGH gene rearrangements in tumor samples and in serial plasma samples of patients affected by MM. Moreover, since our method is based on a straightforward pipeline tailored for the analysis of the monoclonal B-cell expansion, we decided to evaluate the possibility of applying our developed NGS workflow in the non-invasive monitoring of other mature B-cell neoplasms such as diffuse large B-cell lymphomas (DLBCLs) and Hodgkin lymphomas (HLs) that have very rarely circulating tumor cells, but release tumor DNA in the bloodstream. Profiling the IGH repertoire of 26 newly diagnosed DLBCL patients we demonstrated that our NGS workflow allows the identification of clonal IGH rearrangements in archival formalin-fixed paraffin-embedded (FFPE) tissue biopsies and paired cfDNA samples. IGH cfDNA frequencies reflected levels of DLBCL burden detected by imaging modalities at diagnosis and during the course of treatment. In HL cases (n=20), data obtained applying a reverse approach and starting with the analysis of cfDNA to overcome the limitations imposed by the paucity of tumor cells in tissue biopsies, suggested that methods to select pure HL tumor cells are

mandatory to exclude inflammatory contamination and to identify clonal IGH rearrangements in HL. Overall, results I have obtained during my PhD project confirm the clinical utility of IGH cfDNA in the evaluation of tumor burden and response to treatment in mature B-cell malignancies, and support the application of the developed NGS workflow in large prospective studies that will validate the feasibility in the clinical setting and the prognostic value of disease monitoring through the quantitative assessment of plasma clonotypic IGH gene rearrangements.

INTRODUCTION

1. The immune system

The function of the human immune system is primarily to protect against foreign microorganisms. However, it also has a function in maintaining tissue homeostasis and in cancer development. The immune system is composed of two distinct compartments, mediating an early innate immune response and an adaptive late immune response.[1] Macrophages, dendritic cells and neutrophils of innate immunity react very rapidly to infections recognizing particular molecules that are common to many pathogens but are absent in the host, and play a crucial role in the initiation and consequent direction of adaptive immune responses. The B and T lymphocytes of the adaptive immunity embody the function of adaptation, providing a more versatile means of defense, and the function of memory, allowing increased defensive capabilities against subsequent reinfection.[2]

The immune system consists of: primary lymphoid tissues, bone marrow (BM) and thymus, where lymphocytes are generated and mature; and secondary lymphoid tissues, including lymph nodes, spleen and mucosal lymphoid tissues, where adaptive responses are induced and regulated.[3] The microarchitecture of secondary lymphoid tissues consists of B- and T-cell zones, comprising the different subsets of B and T lymphocytes, respectively. These organized anatomical niches, together with the chemokine milieu, allow cellular interactions necessary for antigen recognition and lymphocyte activation.[4]

2. B-cell differentiation

B lymphocytes derive from BM-resident hematopoietic stem cells. In order to recognize the enormous variety of pathogens, these cells require the production of the antigen-specific B-cell receptor (BCR). The BCR is composed of two identical immunoglobulin heavy chains (IGH) and two identical immunoglobulin light chains (IGL), containing a variable antigen-binding domain fused to a constant domain. The variable domain of IGH is encoded by variable (VH), diversity (DH), and

joining (JH) regions, whereas the variable domain of IGL is encoded only by variable (VL) and joining (JL) regions.[5] The immunoglobulin gene remodeling involves three genetic mechanisms: i) V(D)J recombination, ii) somatic hypermutation, and iii) class-switch recombination.[6]

2.1. V(D)J recombination

During early B-cell development in the bone marrow, the IGH gene rearrangement precedes IGL recombination (VJ only), and DH to JH joining precedes VH to DJH joining.[7] The IGH gene is located on chromosome 14q32.3 covering approximately 1250 kb. The constant domain contains ordered clusters of exons encoding different classes of IGH: IGM, IGD, IGG, IGA, IGE. In the variable domain there are about 50 functional VH segments, which can be grouped according to their homology in seven VH families, 27 DH segments and 6 JH segments (Figure 1).[8]

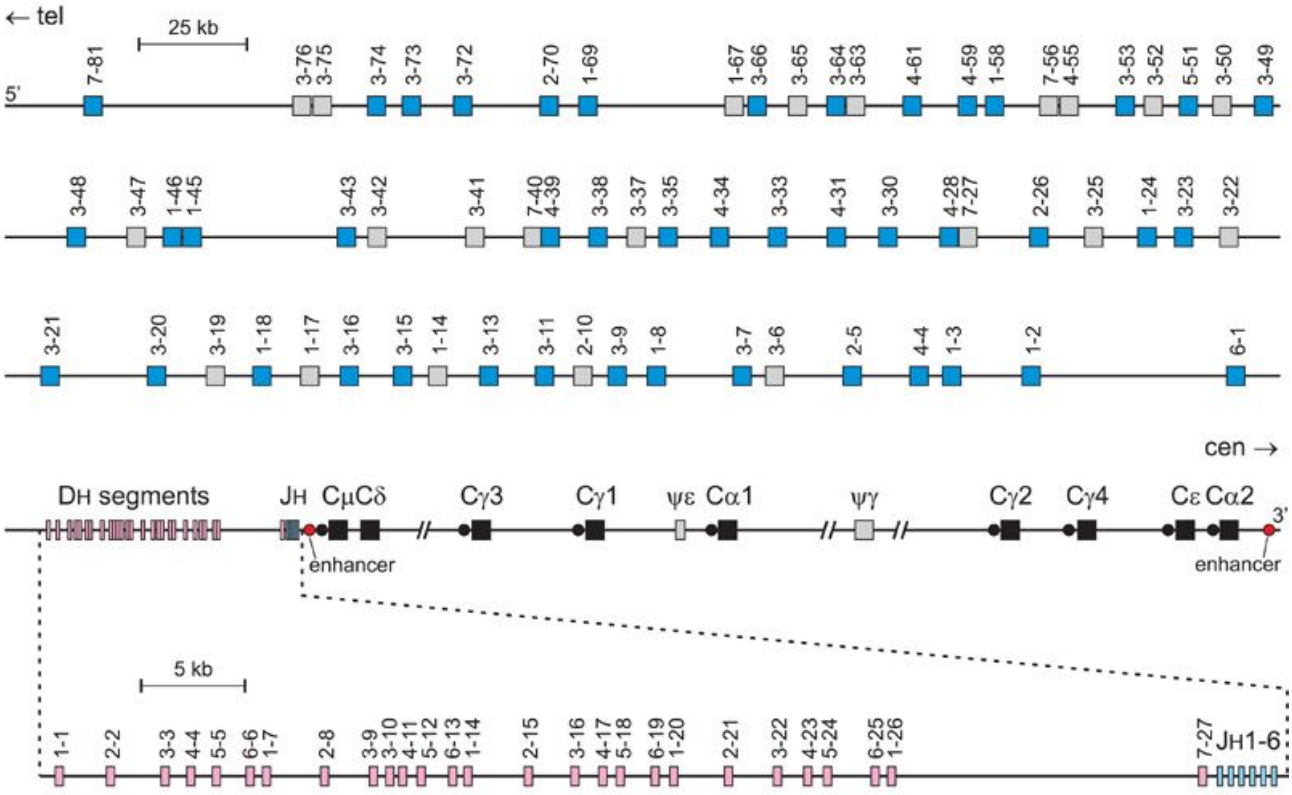


Figure 1. Immunoglobulin heavy chain (IGH) gene complex. Schematic diagram of IGH locus located adjacent to the telomere of chromosome 14 (tel, telomere; cen, centromere). The locus is composed of: VH gene segments (rearrangeable functional VH are denoted by blue bars, rearrangeable pseudogenes by gray bars); DH gene segments (pink bars); JH gene segments (light-blue bars); constant domain gene segments (black bars); enhancers for transcription (red dots). Figure exported from[9].

The mechanism for V(D)J recombination involves the DNA cleavage mediated by the recombination activating genes (RAGs), and the imprecise joining process carried out primarily by ubiquitously expressed non-homologous DNA end-joining (NHEJ) proteins.[10]

The lymphoid-specific RAG-1 and RAG-2 endonucleases introduce DNA double-strand breaks (DSBs) at defined recombination signal sequences (RSSs) that flank 3′ of VH gene segments, 5′ of JH segments and both sides of DH segments (Figure 2).

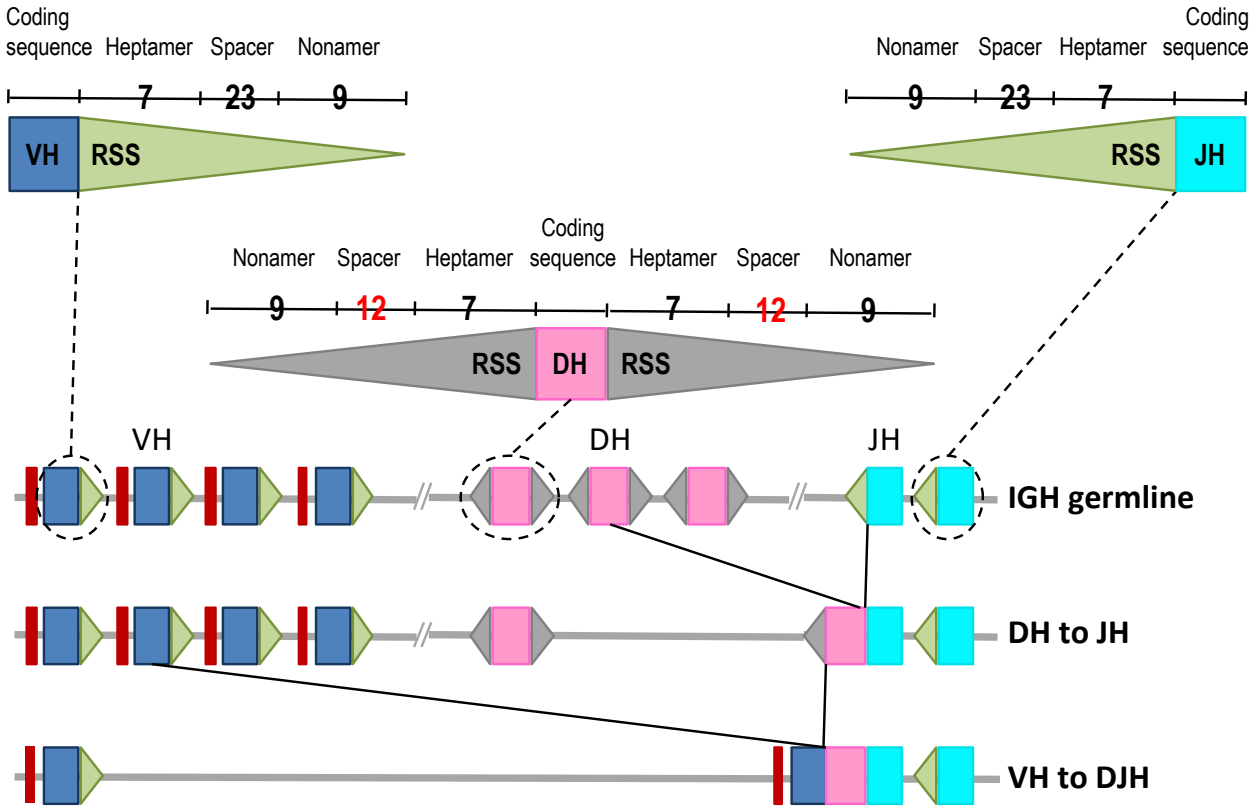


Figure 2. Recombination signal sequences (RSSs) and 12/23 rule. Blue, pink and light-blue rectangles represent VH, DH, and JH gene segments, respectively, with red leader regions upstream each VH segment; gray and green triangles represent 12RSS and 23RSS, respectively; the text above indicates the RSS elements with the respective length. Starting from IGH germline configuration, IGH DJ joining creates a partial rearrangement, then the subsequent junction of a VH segment gives rise to a complete, functional IGH rearrangement.

RSSs are composed of a conserved palindromic heptamer and an AT-rich nonamer separated by non-conserved 12 bp or 23 bp spacers. The specific length of spacers, corresponding to one or two turns of the DNA double helix, brings the heptamer and nonamer sequences to the same side of the DNA helix, where they can be bound by RAG-1 and RAG-2. V(D)J recombination occurs only between two gene segments flanked respectively by RSSs that contain 12 and 23 bp spacers,

referred to as the 12/23 rule.[11] Thus, first a DH segment is joined to a JH segment, and then a VH segment is joined to a DH segment, but VH and JH segments cannot be joined directly as both are flanked by 23 bp spacers (Figure 2).

RAG proteins bind to each other, bringing together the segments to be joined, and cleave RSSs to create two hairpin extremities (Figure 3).

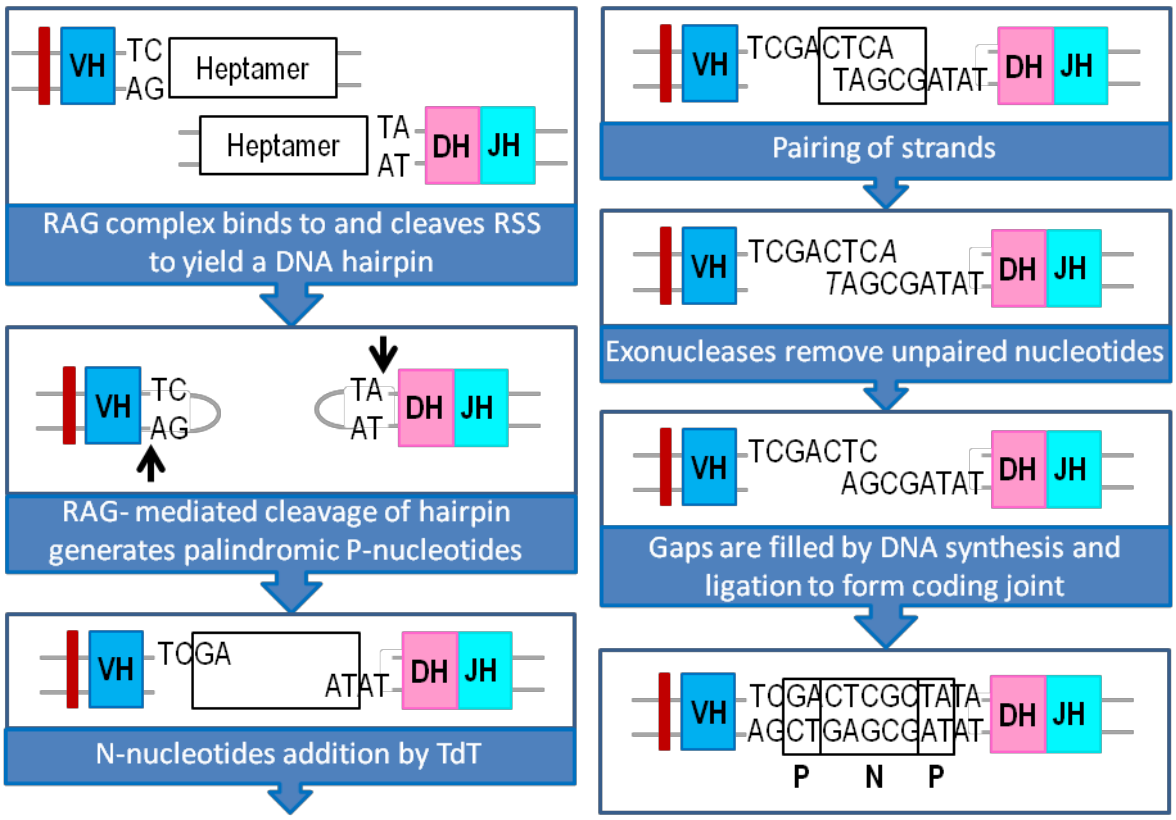


Figure 3. Junctional diversity during V(D)J recombination. Overview of addition and subtraction of nucleotides at the joints between the different gene segments during V(D)J recombination. The combinatorial diversity, given by all the different possible combinations of gene segments, together with the junctional diversity, could give rise in theory to approximately $3.5x10^6$ different antibody molecules.

Then, through the action of additional essential proteins, such as Ku70:Ku80, Artemis and activated DNA-dependent protein kinases, the DNA hairpins are randomly cleaved to yield single-stranded DNA ends.[12] The subsequent joining process is imprecise: due to the intrinsic error prone property of NHEJ, the coding joints frequently gain and/or loose nucleotides. The nucleotide gain results from: i) the asymmetrical Artemis cleavage that creates one longer DNA strand containing DNA palindromes derived from originally complementary nucleotides (P-nucleotide addition to the

shorter DNA strand); ii) the template-independent activity of the terminal deoxynucleotidyl transferase (TdT) that randomly adds nucleotides to the ends of the single-stranded segments (N-nucleotide addition) (Figure 3). The ends of the joined segments can be trimmed back by exonucleases.[13]

The entire IGH variable domain contains three complementarity-determining regions (CDR1, CDR2 and CDR3) interspersed between four framework regions (FR1, FR2, FR3 and FR4) (Figure 4). While CDR1 and CDR2 are encoded within the VH region and therefore the variability in the repertoire is correlated with the VH gene segment usage only, the CDR3 region is the most variable since it is the result of the unique recombination and joining of VH, DH and JH segments.[14] CDR3 is indeed the major determinant of the antigen-binding site specificity and its unique nucleotide sequence represents the molecular marker or "fingerprint" of each B-cell.[15]

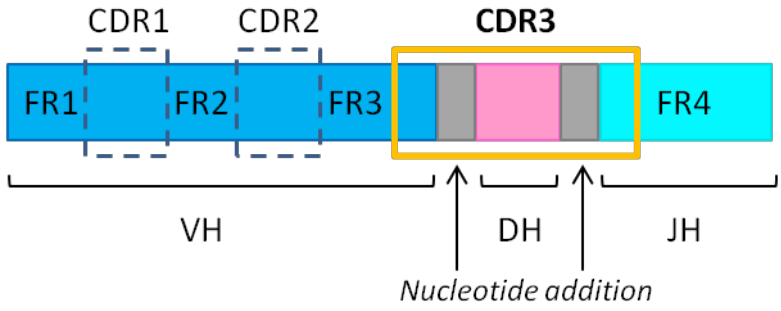


Figure 4. Antigen-binding sites of immunoglobulin heavy chains. Functional variable domain of each immunoglobulin heavy chain is composed of four conserved structural framework regions (FR) and three more diverse complementarity determining regions (CDR). CDR3 (yellow box) is the most diverse as it spans VH (blue rectangle), DH (pink rectangle) and JH (light-blue rectangle) gene segments and contains random nucleotides (gray rectangles) added by V(D)J recombination process.

Throughout B-cell differentiation in the BM, first in the early pro-B-cell the DH-JH recombination produces an incomplete rearrangement. Joining of a VH segment to the DJH completes the late pro-B-cell stage (Figura 5). If the complete VH-DH-JH recombination results productive, the process of allelic exclusion ensures that the other IGH allele is generally excluded from further recombination attempts.[16]

Pro-B cells become large pre-B-cells when the two identical IGM heavy chain proteins are transiently exposed on the cell surface with surrogate light chains constituting the pre-BCR.

Signals from the pre-BCR orchestrate the proliferation and subsequent developmental transition to the small pre-B-cell stage, where IGH return intracellular and starts the IGL VJ recombination intracellularly: initial attempts occur at the IG kappa (IGκ) locus on chromosome 2p11.2 and if a functional IGκ rearrangement is not achieved (mainly by non-in-frame rearrangements) the IG lambda (IGλ) locus on chromosome 22q11.2 undergoes recombination.[17] Assembly of a functional IGH-IGL protein complex on the cell surface allows B-cells to escape apoptosis and proceed with maturation. Naïve immature (IGM+/IGD-) yet immunocompetent B-cells expressing a fully formed BCR, leave the BM and circulate through secondary lymphoid organs. There, in the outer marginal zone (MZ), B-cells further differentiate into naïve mature B-cells co-expressing IGM and IGD as result of IGH alternative splicing (Figure 5).[18]

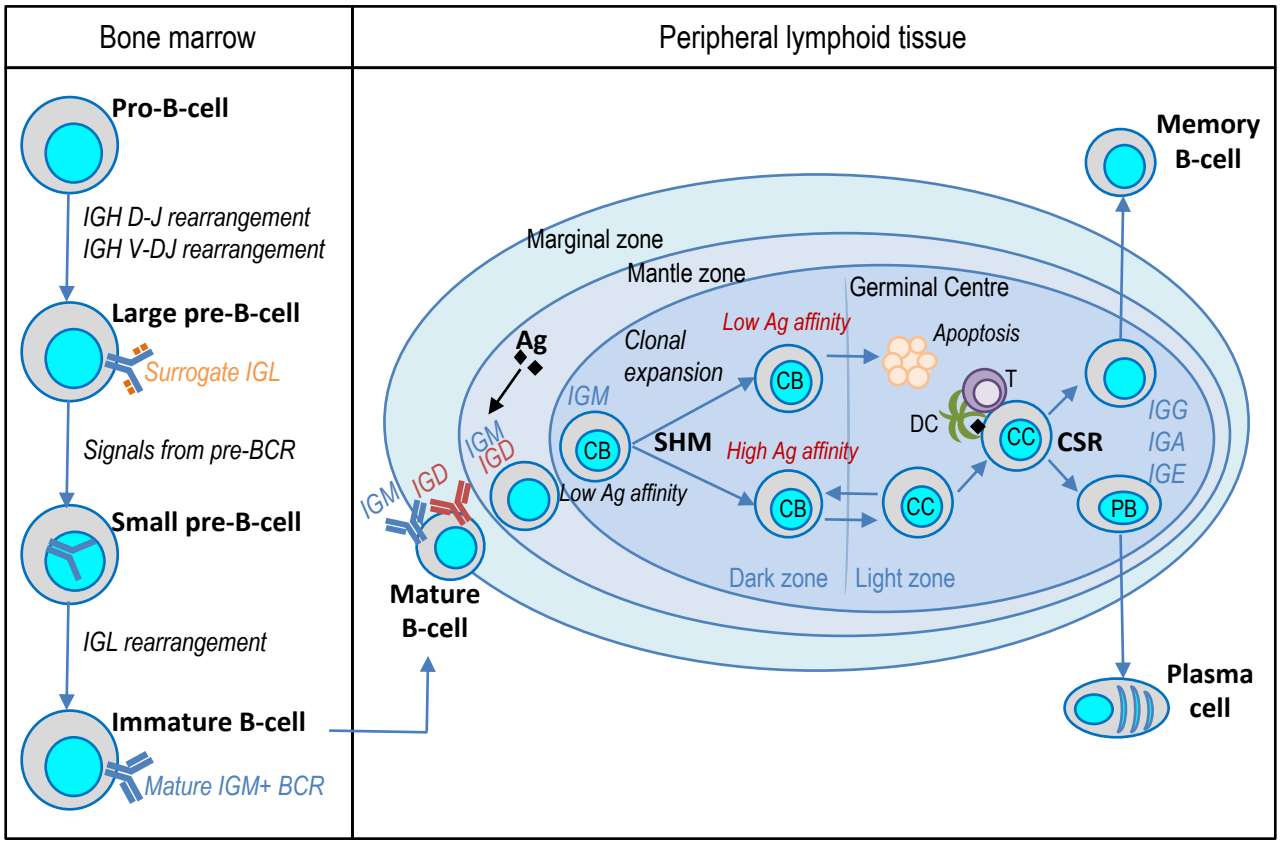


Figure 5. B-cell differentiation and activation. B-cell development commences in the bone marrow whereby precursor B-cells rearrange the immunoglobulin (IG) heavy chain (IGH) and light chain (IGL) genes to generate a B-cell receptor (BCR). On entering the periphery, B-cells congregate in lymphoid tissue wherein antigen-dependent B-cell development takes place. Naïve B-cells on encountering antigen become activated and generate germinal centers in which the processes of somatic hypermutation (SHM) and class switch recombination (CSR) take place in the dark and light zones, respectively. Ag, antigen; CB, centroblast; CC, centrocyte; DC, dendritic cell; T, T-cell; PB, plasmablast.

2.2. Somatic hypermutation and class-switch recombination

The activation of BCR takes place in lymph nodes, where naïve mature B-cells moves from MZ, to mantle zone, and then to the T-cell-rich region, aggregating into primary follicles and forming germinal centers (GCs), which are sites of sustained B-cell proliferation and differentiation.[19]

Within GCs mature B-cells encounter their cognate antigen and are stimulated to further modified immunoglobulin genes to tailor the humoral immune response. Specifically, GC B-cells undergo two major molecular events: somatic hypermutation (SHM) and class-switch recombination (CSR), both mediated by activation-induced cytosine deaminase (AID).[20]

SHM introduces point mutations in the variable domain of rearranged immunoglobulin genes, increasing the affinity for the antigen. CSR is the mechanism that replaces the IGM heavy chain constant region with the constant region for another isotype (IGG, IGA or less frequently IGE), in order to change the effector function preserving the unique variable region. Once this last process has been completed, activated B-cells leave the germinal center to become plasma cells or memory B-cells (Figure 5).[21]

When the regulation of B-cell differentiation and activation is disrupted, lymphomas and leukemias can occur. The processes that normally create immunoglobulin diversity might be misdirected, resulting in oncogenic chromosomal translocations that block differentiation, prevent apoptosis and/or promote proliferation.[22]

3. Mature B-cell neoplasms

Neoplasms of B-cell origin are exceedingly more common than disorders arising from T-cells. Moreover, mature neoplasms account for more than 90% of all lymphoid neoplasms and according to the World Health Organization (WHO) are classified in: chronic lymphocytic leukemia (CLL); Hodgkin (HL) and non-Hodgkin lymphomas (NHL); and plasma cell disorders mostly represented by multiple myeloma (MM).[23] Mature B-cell malignancies represent a heterogeneous group of distinct entities that are classified on the basis of morphology, immunophenotype and clinical

behavior.[24] The recent genetic and molecular studies are increasing the understanding of the mechanisms underlying the clinical and biological diversity of these tumors and are providing valuable information for new therapeutic opportunities.[25] Mature B-cell malignancies arise from normal B lymphocytes arrested at a specific stage of differentiation (Figure 6).[26] Naïve mature MZ B-cells are the cells of origin of CLL and marginal zone lymphoma (MZL). Mantle cell lymphoma (MCL) derived from B-cells moved in the mantle zone from the MZ. The cycles of SHM, proliferation, and clonal selection of B lymphocytes in the GCs are potentially oncogenic, thus GC B-cells give rise to diverse common lymphoma types: Hodgkin lymphoma (HL), diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FL) and Burkitt lymphoma (BL).[27] Memory B-cell is the so called cell-of-origin of CLL displaying immunoglobulin genes that are somatically mutated. Plasma cell is the cell-of-origin of multiple myeloma (MM).

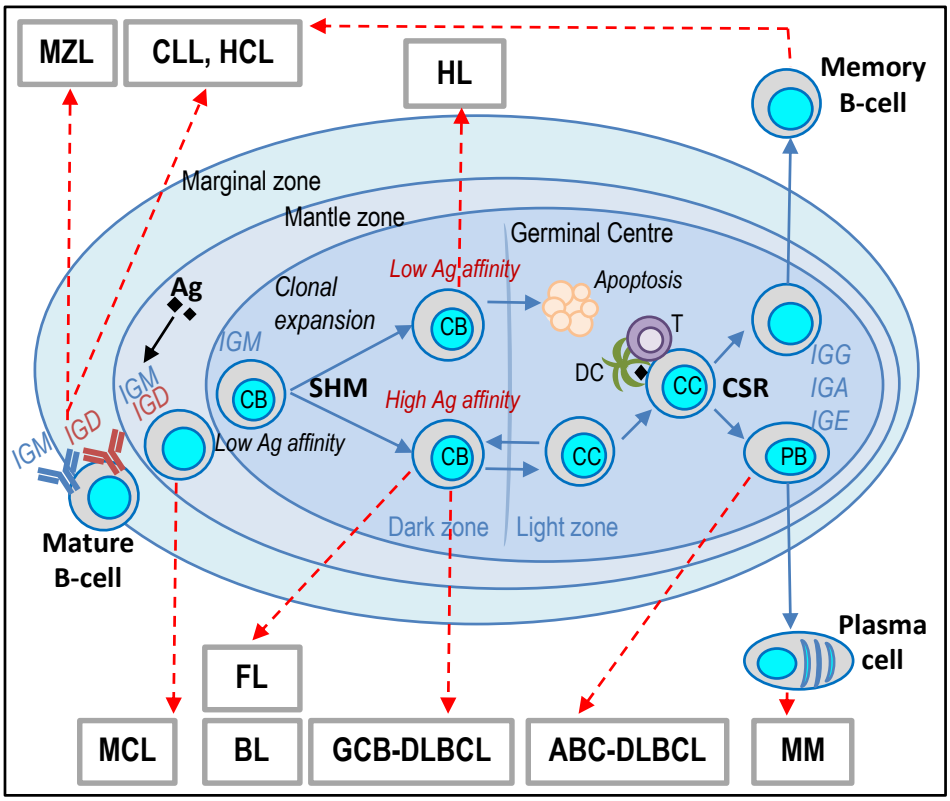


Figure 6. Cellular origin of mature B-cell malignancies. The gray blocks illustrate at which stage of B-cell differentiation mature B-cell tumors are thought to originate according to the V-gene analysis. SHM, somatic hypermutation; CSR, class-switch recombination; Ag, antigen; CB, centroblast; CC, centrocyte; DC, dendritic cell; T, T-cell; PB, plasmablast; MZL, marginal zone lymphoma; MCL, mantle cell lymphoma; HL, Hodgkin lymphoma; FL, follicular lymphoma; BL, Burkitt lymphoma; DLBCL, diffuse large B-cell lymphoma (GCB, germinal center B-cell-like; ABC, activated B-cell-like); CLL, chronic lymphocytic leukemia; HCL, hairy cell leukemia; MM, multiple myeloma. Only major categories are included.

Since mature B-cell neoplasms derive from the monoclonal expansion of a B-cell undergoing unique immunoglobulin gene remodeling, V-gene analysis yields valuable information about the stage of differentiation of malignant B-cells.[28] Moreover, the CDR3 joining region as the final product of V(D)J recombination, represents a tailor-made patient-specific marker to monitor the disease burden at diagnosis and during treatment.[29]

3.1. Hodgkin lymphoma

HL is one of the most frequent lymphomas in the Western world, with an annual incidence of about 3 per 100.000 persons.[30] Although the cause of HL is unknown, evidence of Epstein-Barr virus (EBV) infection is found in 20-50% of North American and European cases.[31] Presentation typically includes a painless lymph node mass, often affecting the cervical chain, and about 40% of patients suffer from constitutional B-symptoms (recurrent fevers, night sweats, loss of weight).[32] Splenic or bone marrow involvement can also occur in a minority of cases. Based on differences in morphology and immunophenotype of tumor cells, the WHO classification[23] recognizes two distinct disease entities: the classical HL (cHL) and the rare nodular lymphocyte-predominant HL (NLPHL). Classical type constitutes approximately 90% of cases, with a bimodal incidence that peaks in young adults (20s-30s) and over the fifth decade of life, and is further divided into four histological subgroups: nodular sclerosis (60-80% of cases), mixed cellularity, lymphocyte-rich and lymphocyte-depletion.[33]

Tumor cells in HL are very rare and usually accounting for only about 0.1%-2% of cells in the tissue, and are surrounded by a microenvironment rich in T lymphocytes, macrophages, neutrophils, eosinophils and plasma cells.[32] In classical HL, the malignant cells are referred to as Hodgkin and Reed-Sternberg (HRS) cells, given the histological presence of mononuclear Hodgkin (H) and bi- to multi-nuclear Reed-Sternberg (RS) cells.[34] Despite HRS cells derive from clonal B-cells in the germinal centers, typical GC B-cell surface markers (CD19, CD79) and transcription factors (OCT-2, BOB.1 and PU.1) are down-regulated or completely absent and CD20 is expressed in only 20% of cases. In contrast, HRS cells universally express CD30 and 75-85% of cases express CD15.[35] The origin of HRS cells from mature B-cells was clarified

by the demonstration that they carry clonal and somatically mutated IG variable genes[36]. However, in about 25% of cHL cases, V genes show loss of function mutations, including nonsense mutations; GC B-cells acquiring such mutations normally rapidly undergo apoptosis. Thus, the critical steps in HL pathogenesis most likely happen in the GC enabling the crippled HRS cell precursors to escape apoptosis.[37] The reprogramming and loss of the B-cell phenotype in HRS cells likely occur due to genomic alterations that alter important signaling pathways, including NOTCH-1, JAK/STAT and NF-κB.[38]

An accurate assessment of the stage of disease in patients with HL is critical for the selection of the appropriate therapy. Patients with HL are staged according to the Ann Arbor system, including Cotswolds modifications, based on: the number of involved sites (stage I or II), whether the involved lymph nodes are on one or both sides of the diaphragm (stages I-II or III-IV), whether the sites of involvement are bulky (> 10 cm in diameter) (suffix "X"), whether there is contiguous extranodal involvement (suffix "E") or disseminated extranodal disease (stage IV), and whether B symptoms are present (suffix "B") or not (suffix "A").[39] Stages IA, IB, and IIA without bulky disease define patients with early-stage HL, while stages III, IV, I and II with bulky disease and IIB are considered advanced stages.[40] Prognostic factors for early-stage HL include: age ≥ 50 years, the presence of a large mediastinal mass, an elevated sedimentation rate, extranodal involvement or involvement of multiple nodal sites. In contrast, prognostication in advanced-stage patients is defined by the International Prognostic Score (IPS)[41] that includes the following seven variables: age ≥ 45 years, presence of stage IV disease, male sex, white blood count ≥ 15.000 cells/μL, lymphocyte count < 600 cells/μl, albumin < 4.0 g/dL, hemoglobin < 10.5 g/dL. Patients with four or more of these factors have inferior outcome compared to low-risk patients (IPS score: 0-3). [^{18}F]2-fluoro-2-deoxy-D-glucose (FDG) positron emission tomography (PET) has emerged as an important tool in the HL staging and response assessment during and after treatment, demonstrating a greater sensitivity compared to the computed tomography (CT) through the development of the highly reproducible Deauville 5-point scale for reporting results.[39,42] Achievement of a negative interim PET after 2 cycles of ABVD (adriamycin(doxorubicin), bleomycin, vinblastine, dacarbazine) chemotherapy (PET-2), is

highly predictive of long-term progression-free survival.[42]

Four cycles of ABVD combination chemotherapy followed by involved-field irradiation is the standard treatment for patients with early-stage HL, with a 5-year OS exceeding 90%. Prognosis is also very good in advanced-stage disease, reaching an overall survival of approximately 75% after 6 cycles of ABVD chemotherapy followed by involved-filed irradiation. However, roughly 10% of patients with early-stage disease and 20-25% of patients with advanced disease are refractory to, or relapse after, initial treatment.[43]

3.2. Non-Hodgkin lymphomas

3.2.1 Diffuse large B-cell lymphoma

DLBCL is the most common B-cell NHL (B-NHL), representing approximately 30-40% of all newly lymphoma cases worldwide.[44] DLBCL can affect people of any age, but it occurs mostly in older people (average age at the time of diagnosis: mid-60s). It usually starts as a fast-growing aggressive mass in single or multiple, nodal or extranodal (chest, abdomen, bones, or even brain or spinal cord) sites.[45] The most common type of DLBCL, designated as not otherwise specified, represents 80-85% of all cases. The remaining 15-20% includes several rare variants of DLBCL segregated out in the WHO classification because of a propensity to affect distinct sites (e.g. primary DLBCL of the central nervous system, primary cutaneous DLBCL, leg type, intravascular large B-cell lymphoma).[46]

DLBCL not otherwise specified (referred to henceforth as DLBCL) is a heterogeneous entity in terms of clinical presentation, genetic findings, response to therapy and prognosis. Investigating efforts to define patient subsets associated with favorable or poor prognosis are focused on: clinical factors recognized in the International Prognostic Index (IPI), gene expression profiles, and molecular factors as characterized by immunohistochemistry and cytogenetics.[47]

The IPI is a clinical tool used in risk stratification of lymphoma patients incorporates 5 factors: 1) age > 60; 2) Ann Arbor stage III or IV; 3) elevated lactate dehydrogenase (LDH) levels; 4) Eastern Cooperative Oncology Group (ECOG) performance status ≥ 2; and 5) sites of involvement ≥ 2. Newly diagnosed patients with DLBCL are classified into 4 different risk groups based on the sum

of reported risk factors: low (IPI score = 0-1), low-intermediate (IPI score = 2), high-intermediate (IPI score = 3), high (IPI score = 4).[44]

Gene expression profiling studies allowed the identification of two main prognostic subtypes based on the cell-of-origin (COO): i) germinal center B-cell-like (GCB) DLBCL deriving from germinal center lymphoid cells; and ii) activated B-cell-like (ABC) DLBCL arising from B-cells at a plasmablastic stage.[48] Pathogenetic hallmarks of GCB-DLBCL are the t(14;18) translocation, mutations of the histone methyltransferase EZH2 and PTEN deletion. On the other side, ABC-DLBCL is characterized by the constitutive activation of the NF-κB signaling pathway with recurring mutations in CD79A/B, CARD11 and MYD88,[49] and exhibits an inferior survival following R-CHOP(rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone)-like chemotherapy. Concurrent aberrancies of MYC, BCL2 and BCL6, that play a fundamental role in the DLBCL pathogenesis ultimately leading to the expansion of the malignant clone, are also considered molecular prognostic factors: double-expressor DLBCL patients, characterized by the co-overexpression of MYC and BCL2 (detected by immunohistochemistry), and double-hit DLBCL patients, demonstrating MYC rearrangement in addition to BCL2 and/or BCL6 rearrangements (detected by cytogenetic techniques), are associated with poor clinical outcome.[50]

Response assessment is based on Lugano criteria,[39] that include PET-CT as the cornerstone in lymphoma management at initial staging, during treatment after 2 to 4 cycles of chemotherapy, and at the end of treatment (EOT) to evaluate treatment response.

Although clinical/molecular factors and the sensitive PET scanning help in predicting outcome, DLBCL remains a clinically heterogeneous disease.[51] Early-stage DLBCLs are considered curable with standard R-CHOP-like combined therapies, with 60-70% of 5-year overall survival rates; however, a significant number of DLBCL patients do not achieve a complete response (CR), and about 30-40% of patients ultimately relapse.[52]

3.2.2. Marginal zone lymphoma

MZL is a group of indolent NHLs that arise from marginal zone B-cells present in lymph nodes and in extranodal tissues, such as spleen and mucosal lymphoid tissues. These neoplasms are

commonly associated with chronic antigenic stimulation as a result of either infection or autoimmune diseases (e.g., Sjögren syndrome).[53] MZLs comprise about 5-17% of all B-NHLs in adults. According to the involved sites and characteristic molecular findings, the WHO classification singles out three different MZL subtypes: splenic marginal zone lymphoma (SMZL), nodal marginal zone lymphoma (NMZL) and extranodal marginal zone B-cell lymphoma, also called mucosa-associated lymphoid tissue (MALT) lymphoma. These subgroups share common immunophenotypic markers, including CD19, CD20 and CD22, but are different in their biology and behavior.[54] MALT lymphoma is the most common entity, accounting for approximately 70% of all MZLs and is the third most common B-NHL. *Helicobacter pylori* infection is implicated in the pathogenesis of MALT and remission can be achieved with eradication of the bacteria in many patients, although 30-40% of cases do not respond to antibiotic therapy.[53] Owing to the rarity of MZLs, universally accepted therapeutic guidelines do not exist. Surgical resection, radiotherapy and chemotherapy with alkylating agents constitute standard approaches.[55]

3.2.3 Mantle cell lymphoma

MCL constitutes 5-7% of all B-NHLs and predominates in males (male:female ratio ~2-7:1) with advanced age (median age: 60 years).[56] MCL is characterized by an abnormal proliferation of mature B lymphocytes, which probably derive from naïve B-cells expressing CD5. The genetic hallmark is the t(11;14)(q13;q32) translocation that juxtaposes the proto-oncogene CCND1 at chromosome 11 to the IGH locus at chromosome 14. The translocation leads to the constitutive expression of cyclin D1, which has an important pathogenetic role accelerating the progression of lymphoma cells through the G1/S checkpoint of the cell cycle.[57] Other biologic features such as a high Ki-67 proliferation index or p53 mutations and p16 deletions, are closely related to the more aggressive MCL subtypes such as the blastoid variants.[57]

Although MCL typically responds to frontline chemotherapy, it remains incurable with standard approaches. The median overall survival of patients with MCL is 3-4 years. Complete remission of disease with conventional therapy is obtained in 20-80% of patients, but almost all will relapse.[58]

3.2.4 Follicular lymphoma

FL is the most frequent indolent NHL and the second most common form of NHL overall. The median age at diagnosis is 60 years. FL recapitulates normal GC B-cells of secondary lymphoid follicles. Approximately 85% of patients with FL have the t(14;18) translocation which results in the over-expression of the BCL-2 protein, a member of a family of proteins that block apoptosis.[59] FL often presents with painless lymphadenopathy and grows more slowly compared to other forms of NHLs such as DLBCLs and MCLs. The neoplastic cells consist of a mixture of centrocytes (small to medium sized cells) and centroblasts (large cells). The clinical aggressiveness of the tumor increases with increasing numbers of centroblasts. Most FLs are advanced at the time of diagnosis, with involvement of bone marrow (60-70% of cases), spleen (40%) and less commonly other extranodal sites.[52] There is no known cure and historically the median survival of patients presenting with advanced stage follicular lymphoma is 10 years, although the introduction of monoclonal antibody treatment has increased the survival rates. Histological transformation from FL to DLBCL occurs in 10-70% of cases, and is associated with poor prognosis.[60]

3.2.5 Burkitt lymphoma

Burkitt lymphoma (BL) is a highly aggressive B-NHL associated with translocations involving the c-myc oncogene on chromosome 8.[61] In WHO classification,[23] three clinical variants of BL are described: endemic, sporadic, and immunodeficiency-associated type. Endemic BL refers to cases occurring in African children (4-7 years old) accounting for most childhood malignancies, and typically involves jaw and facial bones. Sporadic BL occurs worldwide with a frequent involvement of abdomen area. Immunodeficiency-associated BL occurs mainly in HIV-infected patients, but also occurs in allograft recipients and individuals with congenital immunodeficiencies.[62] Patients with any of the three clinical variants are at risk for spread to the BM and central nervous system. BL has high proliferation index, which makes prompt diagnosis and initiation of therapy very important to increase chances of survival. Sporadic and immunodeficiency-associated BLs are historically associated with poor prognosis, whereas an excellent response to chemotherapy could be obtained with endemic type, eventually attaining up to 80% long-term survival.[61]

4. Multiple myeloma

Multiple myeloma (MM) is a neoplastic plasma cell (PC) disorder characterized by clonal proliferation of malignant PCs in the bone marrow. This uncontrolled PC proliferation results in abnormal IGH (IGG/IGA/IGD) and/or IGL (kappa, lambda) production that can be detected in blood or urine (M-component).[63] To fulfill the definition of multiple myeloma, end-organ damage must be demonstrated: hypercalcemia, renal failure, anemia, and osteolytic bone lesions, commonly referred to as CRAB features.[64]

MM accounts for 10% of all hematologic tumors representing the second most common hematological malignancy after NHLs. In Western countries, the annual age-adjusted incidence is 5 cases per 100.000 with a median age at diagnosis of approximately 66-70 years.[65]

In 2014, the International Myeloma Working Group (IMWG) revised the MM diagnostic criteria to improve early diagnosis of CRAB-negative cases.[64]

For the diagnosis of MM, both criteria must be met:

- clonal bone marrow plasma cells ≥ 10%;
- any one or more of the following myeloma defining events:
    - evidence of end-organ damage (CRAB criteria):
        - hypercalcemia: serum calcium ≥ 11 mg/dL
        - renal insufficiency: serum creatinine > 2 mg/dL
        - anemia: hemoglobin value ≤ 10 g/dL or 2 g below the routine patient values
        - bone lesions: one or more osteolytic lesions on skeletal radiography or PET-CT;
    - presence of serum and/or urinary M-component (serum free light chain ratio ≥100), except in patients with true non-secretory MM (2-3% of all cases);[66]
    - clonal bone marrow plasma cells ≥ 60%;
    - >1 focal bone lesion (≥ 5 mm) at magnetic resonance imaging (MRI).

4.1 Multiple myeloma oncogenesis

MM is a genetically complex and heterogeneous disease resulting from multiple genomic events that lead to tumor development and progression. Multiple myeloma is part of a range of disorders

referred to as monoclonal gammopathies and evolves from a silent pre-malignant condition clinically recognized as monoclonal gammopathy of undetermined significance (MGUS).

MGUS becomes increasingly common with age, being present in 3% and 5% of patients older than 50 years and 70 years, respectively. MGUS converts to overt MM with a rate of approximately 1% each year.[67]

Smoldering MM (SMM) is an intermediate asymptomatic stage between MGUS and MM and is associated with a higher risk of progression of approximately 10% per year.[68]

The progression to MM is thought to involve a multistep transformation process, occurring as a consequence of the sequential acquisition of genetic hits[69] (Figure 7).

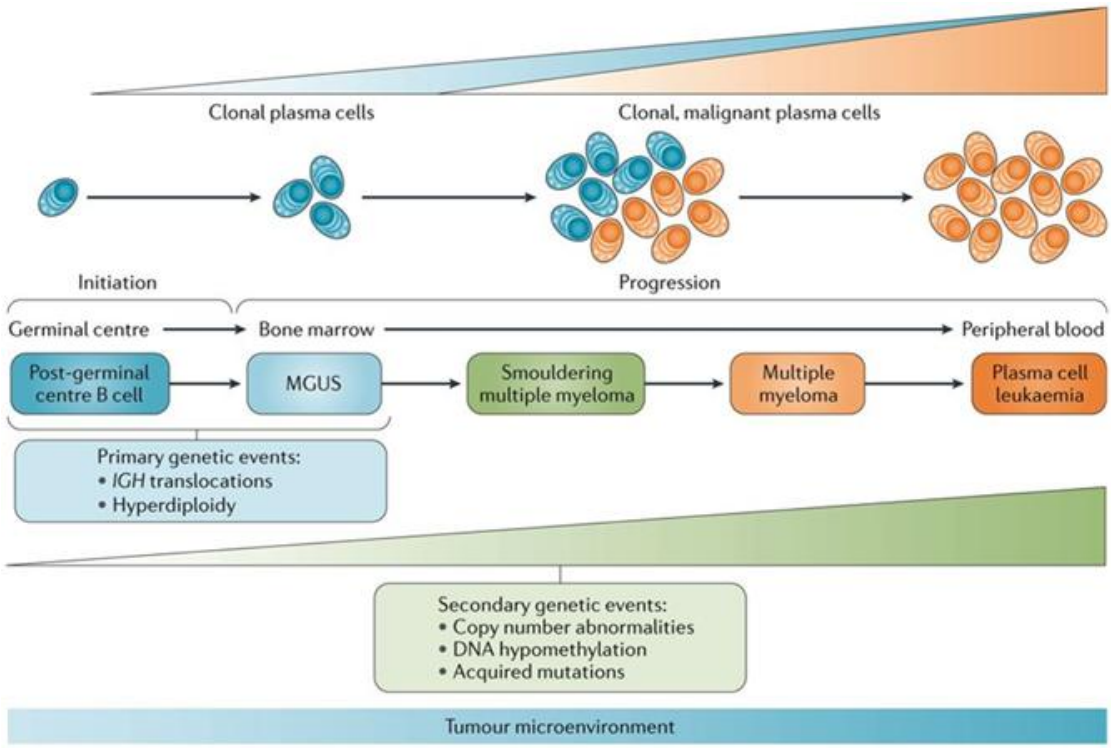


Figure 7. Initiation and progression of multiple myeloma. IGH, immunoglobulin heavy chain locus; MGUS, monoclonal gammopathy of undetermined significance. Multiple myeloma can progress to bone marrow-independent diseases, such as extramedullary myeloma and plasma cell leukaemia. Figure exported from[69].

Myeloma PCs are the neoplastic counterpart of long-lived, post-germinal center PCs, a subset of PCs characterized by long-term (months to years) survival within the bone marrow. However, based on the analysis of IGH variable region of malignant PCs, the first oncogenic events appear to occur in the germinal center, facilitated by the processes of SHM and CSR, which are, by

nature, mutation prone. [27]

Primary genetic events include chromosomal translocations involving the IGH locus and aneuploidy (with hyperdiploidy as the most frequent entity). The observation that patients with premalignant PC dyscrasia (MGUS and SMM) also carry these initial variants, suggests that they are necessary, but not sufficient, in MM pathogenesis. Therefore, secondary events that provide a fitness advantage to a particular subclone over others, are required for tumor progression.

Copy number abnormalities, DNA hypomethylation and acquired mutations are likely to be late oncogenic events occurring in the BM microenvironment after the founder cancer clone is completely differentiated into a long-lived PC (Figure 7).[70]

Primary translocations, mediated by errors in IGH remodeling process, juxtapose one or more oncogenes and one IG enhancer leading to the consequent up-regulation of partner genes: CCND1 and CCND3 [encoding for cyclin D1 and D3; t(11;14) and t(6;14), respectively], MMSET/FGFR3 [intracellular signaling pathway; t(4;14)], and MAF [transcription factor; t(14;16)]. Hyperdiploid karyotype accounts for 50-60% of cases and is characterized by trisomies involving chromosome 3, 5, 7, 9, 11, 15, 19, or 21. A small proportion of patients have both trisomies and IGH translocations.[71] Later oncogenic events in MM pathogenesis include: other copy number variations, such as gain of chromosome 1q, loss of chromosome 1p (del1p) and loss of chromosome 17p (del17p); translocations involving MYC; somatic mutations affecting MAPK pathway (KRAS, NRAS, BRAF), NFκB (CYLD, TRAF3), RNA metabolism (DIS3, FAM46C), DNA-repair (TP53, ATM, ATR) and plasma cell differentiation (IRF4, PRDM1).[71]

Both primary and secondary cytogenetic abnormalities can influence disease course, response to therapy, and prognosis.

4.2 Diagnostic tests

The diagnostic work-up of multiple myeloma comprises history and physical examination, biochemistry screen, cytogenetic analysis of BM, and radiological investigation to detect bone changes. MRI and PET can be used for clarification if required.[72] Table 1 summarizes the minimal diagnostic and prognostic tests reported by the International Myeloma Workshop Consensus Panel

3, that are necessary for the initial investigation of patients with suspected MM.[73]

Table 1. Laboratory tests for multiple myeloma. Adapted from[73].

Complete blood count and differential; peripheral blood smear
Chemistry screen, including calcium and creatinine
Serum protein electrophoresis (SPEP), immunofixation
Nephelometric quantification of serum immunoglobulins
Routine urinalysis, 24-hour urine collection for electrophoresis and immunofixation
Bone marrow aspirate and/or biopsy
Cytogenetics (metaphase karyotype and FISH)
Radiologic skeletal bone survey, including spine, pelvis, skull, humeri, and femurs
Serum β_2-microglobulin and lactate dehydrogenase (LDH)
Measurement of serum-free light chains (sFLC)

4.3 Staging and risk stratification

Although median survival is approximately 5-7 years, the main variations in outcome are related to: host factors, tumor burden (stage), biology (cytogenetic abnormalities), and response to therapy.[74] Tumor burden in MM has traditionally been assessed using the Durie-Salmon Staging (DSS) system and the International Staging System (ISS). DSS system was historically used to determine stage of MM and is based on assessment of total M-component in serum or urine, hemoglobin value, serum calcium level and the number of osteolytic lesions. Currently the cornerstone of MM staging is the ISS score, which incorporates serum albumin and β2-microglobulin levels.[75]

In recent years, as already demonstrated for acute leukemia and lymphomas, the detection of chromosomal abnormalities (CA) by interphase fluorescent in situ hybridization (FISH) was introduced as these lesions represent key determinants of myeloma biology and prognosis. Deletion of chromosome 17p13, translocations t(4;14)(p16;q32) and t(14;16)(q32;q23) define the high-risk disease associated with poor prognosis.[76] Serum LDH is another relevant biomarker in MM.[77] In order to improve the risk stratification of the ISS and to bring genetic determinants of

myeloma one step closer to clinical utility, the IMWG has recently proposed a revised ISS (R-ISS) by adding high-risk CA and serum LDH.[78] R-ISS identifies three risk groups:

- R-ISS I, including ISS stage I (serum β2-microglobulin level < 3.5 mg/L and serum albumin level ≥ 3.5 g/dL), no high-risk CA, normal LDH level (less than the upper limit of normal range);
- R-ISS III, including ISS stage III (serum β2-microglobulin level ≥ 5.5 mg/L) and high-risk CA [del(17p) and/or t(4;14) and/or t(14;16)] or high LDH level;
- R-ISS II, not stage I or III.

4.4 Treatment

An enhanced understanding of MM disease mechanisms has led to the development of new targeted treatments. There was a paradigm shift in the disease management of MM moving from melphalan-prednisone combinations with median survival of 2-3 years, to high-dose chemotherapy and autologous stem cell transplantation (ASCT) with a median survival of 4-5 years, until a median survival of more than 7 years has been achieved with the introduction of immunomodulators (IMiDs) and proteasome inhibitors (PIs).[74]

The immunomodulatory agents (thalidomide-derivatives such as lenalidomide and pomalidomide) represent a class of drugs with a pleiotropic mechanism of action, including the modulation of host immune responses, the disruption of microenvironment interactions, and the direct induction of PC apoptosis (Figure 8).[79] IMiDs induce cell-surface expression of positive co-stimulatory molecules on partially activated T-cells, which are down-regulated in MM. T-cell co-stimulation is associated with up-regulation of cytokines, such as IL-2 and IFN-γ, which mediate the activation of T- and natural killer (NK)-cells inducing myeloma cell apoptosis.[80] IMiDs also act by disrupting the stromal support within the bone marrow. Specifically, IMiDs reduce levels of IL-6 inhibiting TNF-α expression, and thereby, the expression of adhesion molecules on both MM cells and BM stromal cells. This situation leads to the down-regulation of the transcription factor NK-kB and the subsequent block of myeloma growth factors. Finally, IMiDs have also direct apoptotic effects through the activation of Caspase 8 and the G0/G1 cell cycle arrest.[80]

PIs, bortezomib and carfilzomib, act against NF-kB blocking the proteasome activity. NF-kB is constitutively presents in the cytosol and inactivated by the IkB inhibitor. When phosphorylated, IkB is targeted for degradation by the 26S proteasome and allows the translocation of NF-kB into the nucleus. PIs block the degradation of IkB, thereby inhibiting the related NF-kB activation and impairing one of the major anti-apoptotic mechanisms of tumor clones (Figure 8).[81]

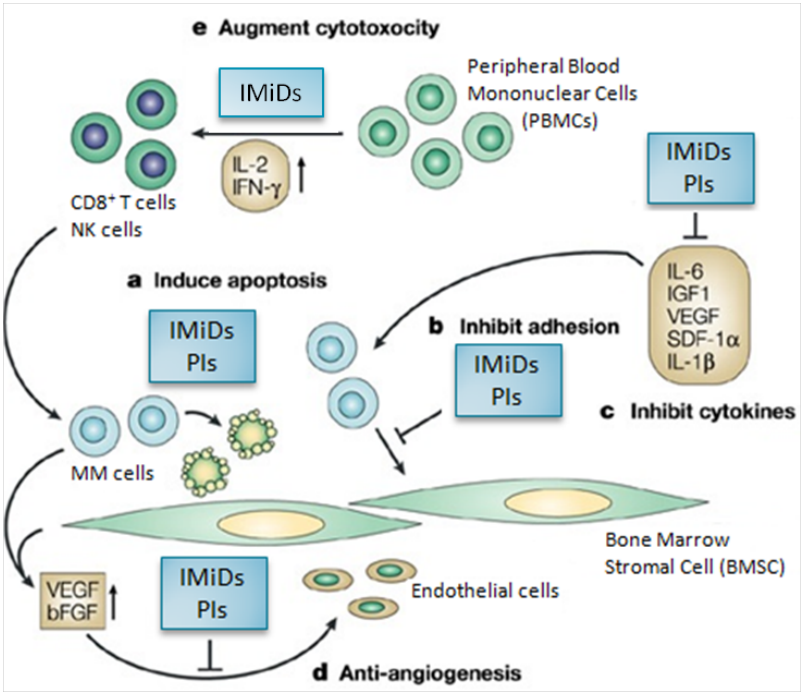


Figure 8. Novel targeted therapies in multiple myeloma treatment. Mechanisms of action (a-e) of immunomodulatory drugs (IMiDs) and proteasome inhibitors (PIs). Figure adapted from[81].

4.5 Clinical course

The International Myeloma Working Group, together with diagnostic criteria, has also defined the MM response criteria.[82] Response is broken down into a number of categories that include partial response (PR), very good partial response (VGPR), complete response (CR), and stable disease (SD) when recorded parameters are outside the range of other subcategories.[83] These responses are mainly based on the changes in M-component levels: PR is based on a reduction of 50% in serum M-component and to 24-h urinary M-component <200 mg; VGPR is determined by a reduction of 90% or greater in serum M-component plus urine M-component level <100 mg per 24 h; for CR the following three criteria must be met:

- negative immunofixation on serum and urine;
- disappearance of any soft tissue plasmacytomas;
- ≤ 5% plasma cells in the bone marrow.

Despite achievement of progressively higher CR rates with the introduction of novel and more effective therapies, the majority of patients will inevitably relapse and myeloma is still considered an incurable disease.[74] The two main relapse subcategories defined by the IMWG are progressive disease (PD) and relapse from CR.[82] PD is based on an increase of 25% from initially documented value in any of the following parameters: serum M-component, urine M-component, serum free light chains (if serum or urine M-component are not detectable at baseline). In addition, the percentage of plasma cells in the bone marrow must be ≥ 10%.

Relapse from CR includes any one or more of the following criteria:

- reappearance of serum or urine M-component by immunofixation or electrophoresis;
- development of ≥ 5% plasma cells in the bone marrow;
- appearance of any other sign of progression (e.g., new plasmacytoma, lytic bone lesion, or hypercalcemia).

MM is clinically characterized by serial cycles of response, remission and relapse. The depth and duration of response following each relapse are generally diminished until the development of drug resistance, which eventually results in refractory disease (Figure 9).[84]

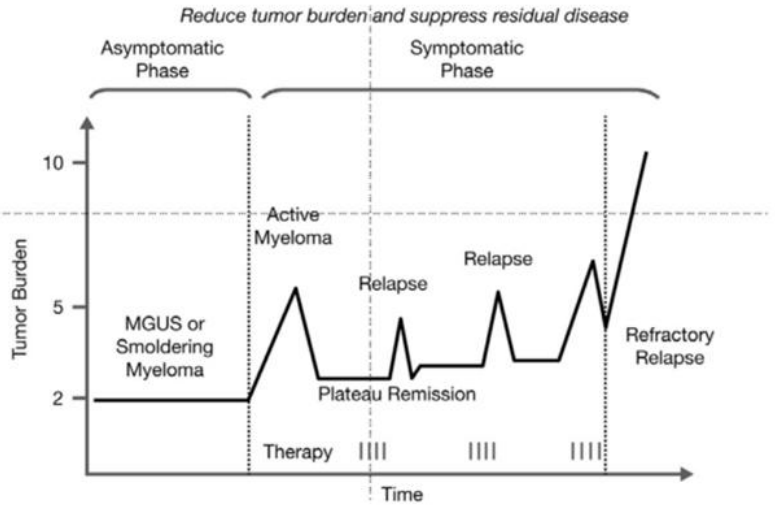


Figure 9. Characteristic pattern of remission and relapse following conventional chemotherapy in multiple myeloma. MGUS, monoclonal gammopathy of undetermined significance. Figure exported from[84].

This pattern suggests the presence of residual disease after treatment, even in presence of an apparent CR. Therefore the development of reproducible highly sensitive assays for detecting minimal residual disease (MRD) become essential to assess treatment effectiveness and to define the value of MRD in predicting for progression-free (PFS) and overall survival (OS).[85]

5. Minimal residual disease in multiple myeloma

The term MRD was introduced when techniques to measure tumor burden were rather insensitive by today's standards.[86] Historically, residual disease in MM patients was detected by morphologic examination of the bone marrow using light microscopy or by serum/urine protein electrophoresis. However, as technology has improved, we are now able to detect very small populations of myeloma cells that cannot be seen by light microscopy.[87] Today, MRD is generally referred to the persistence of low levels of disease in the order of one tumor cell in ≥ 10^5 normal cells.

Several studies using multiparameter flow cytometry (MFC), allele-specific oligonucleotide quantitative polymerase chain reaction (ASO-qPCR), or more recently next-generation sequencing (NGS) techniques have shown that the presence of residual tumor cells after therapy is associated with poor prognosis.[88]

5.1. Multiparameter flow cytometry and quantitative PCR

As it has occurred in other lymphoid malignancies over the past decade, MFC and ASO-qPCR have emerged as the most attractive, appropriate, and sensitive approaches to detect MRD.

MFC immunophenotyping is the current gold standard in the clinical management of MM, both for diagnosis and MRD evaluation.[89] MFC is a worldwide used technique that allows: i) identification and characterization of PCs using multiple parameters simultaneously; ii) combined detection of cell surface and intracellular antigens; iii) automated analysis of a large numbers of cells (> 10^7) within a relatively short period of time; iv) the monitoring of disease in virtually every MM patient without requiring patient-specific markers.[90] The European Myeloma Network guidelines for MFC have established consensus monoclonal antibody combinations for an accurate identification of

malignant BM PCs.[91] Consistent with this, specific PC identification markers (CD38 and CD138) plus discriminatory markers to distinguish clonal and normal PCs need to be used. These antibodies include those that target antigens which are absent (CD19 and CD45), weakly expressed (CD27 and CD81) and strongly positive (CD56 and CD117) in neoplastic PCs. Additionally, cytoplasmic IGL restriction (κ or λ) is helpful to confirm aberrant immunophenotype. The 2-tube 8-color MFC method based on consensus antibodies described above and developed by the EuroFlow group[91] is the current reference approach for the definition of immunophenotypic CR.[92] However, a number of variables might influence results and sensitivity rates. The maximum detection sensitivity of MFC ranges from 0.0005% to 0.02%, based on: i) the number of BM cells analyzed (events); ii) the number of abnormal plasma cells needed to define the presence of MRD; iii) the number of antigens studied (and therefore also the lab available cytometric technology).[93]

Table 2. Current practice for myeloma MRD assessment in different institutions. *Maximum possible sensitivity determined by dividing the minimum number of abnormal plasma cells needs to call MRD by the number of events acquired in MRD testing. Adapted from[93].

Events acquired in MRD testing, *n*	**Abnormal plasma cells needed to call MRD, *n***	**Maximum possible sensitivity, %***	**Antigens studied, *n***
3.000.000 - 4.000.000	20	0.0005	12
1.800.000	50	0.003	9
1.000.000	50	0.005	8
500.000	25	0.005	6
300.000 - 500.000	30-35	0.006	8
250.000 - 500.000	50	0.01	10
100.000	20	0.02	10

Assessment of MRD can also be achieved using PCR-based technologies that identify patient-specific MM cells. Every B lymphocyte, after V(D)J recombination process, carries a unique CDR3 sequence, which serves as a marker for tumor in lymphoma and myeloma patients.[94] Since IGL rearrangement involves only V and J regions, the IGL CDR3 is less diverse than the IGH CDR3,

that involving two different joining sites, D/J and V/DJ, determines a highly variable specific region.

Molecular MRD detection was initially based on qualitative ASO-PCR approaches using patient-specific primers generated from the rearranged IGH V(D)J region.[95] Subsequently, the real-time quantitative ASO-PCR (ASO-qPCR) has emerged as a sensitive and specific method for the accurate quantification of residual disease. ASO-qPCR includes two major steps. Starting from diagnostic BM sample, IGH gene rearrangements are amplified by qualitative PCR using consensus primers for genomic DNA,[9,96] and PCR bands are analyzed by traditional Sanger sequencing to verify the presence of clonal PCs and to identify patient-specific IGH CDR3 joining region. Then, real time qPCR is performed using ASO primers or probes, which are designed to bind to the unique IGH CDR3 sequence.[94] ASO-qPCR of IGH junctional regions is highly specific and sensitive (sensitivity down to 10^{-5}),[97] fully standardized by EuroMRD,[98] can be perform on stored samples, and the result is very simple to analyze.

Recently, another quantitative PCR is emerging for MRD detection, the droplet-based digital PCR (ddPCR).[99] Based on sample partitioning, ddPCR increases the probability to detect rare DNA molecules and allows their absolute quantification without using standard curves. On the other hand, the first steps of CDR3 identification and primers/probe design are the same of ASO-qPCR.

Both MFC and quantitative PCR have some disadvantages. MFC requires fresh samples, and hemodiluted BM aspirates, together with the possible (even if rare in the case of MM) “antigen shifts” in neoplastic cells after treatment, may lead to false-negative results.[100] ASO-qPCR is time-consuming and custom patient-specific primers or probes are needed. Additionally, the SHM process can introduce mutations in DNA sequence that hamper PCR primer annealing resulting in possible false-negative results.[101]

5.2. Next-generation sequencing

Recently, there has been significant interest in NGS for detection of MRD in MM patient, and data supporting its use continue to emerge.[102]

Sanger method has been the mainstay of DNA sequencing for decades. Along with the post-

genome era, genetic studies demanded high-throughput and low cost sequencing technology. Therefore, 'first-generation' Sanger sequencing was replaced by next-generation sequencing.[103] NGS technologies allow genomic analysis through parallel sequencing of thousands to millions of DNA molecules in a relatively short period of time. Roche 454 was the first commercially available next-generation sequencer. Later other major NGS platforms were released, such as ABI SOLiD, Illumina Genome Analyzer, Life Technologies (now Thermo Fisher Scientific) Ion Torrent.[104]

Similar to ASO-qPCR, MRD assessment by NGS is a molecular method based on the amplification of IGH, but also IGL, gene rearrangements as tumor biomarkers.[105] Regardless of the NGS platform, the shared NGS workflow consists of the following steps[106]: i) DNA extraction from diagnostic fresh or stored samples; ii) IG gene rearrangements amplification by PCR using consensus primers (locus specific primers designed for IGH or IGL instead of patient-specific primers); iii) adapters' and tag sequences' ligation to the PCR products to allow the simultaneous analysis of different samples ("high-throughput"); iv) re-amplification of barcoded PCR products (libraries) by PCR using primers for adapters; v) sequencing of IG libraries (10^4-10^6 times; depth or "coverage"); vi) evaluation of the frequencies of clonal rearrangements through intense bioinformatic analysis (Figure 10).

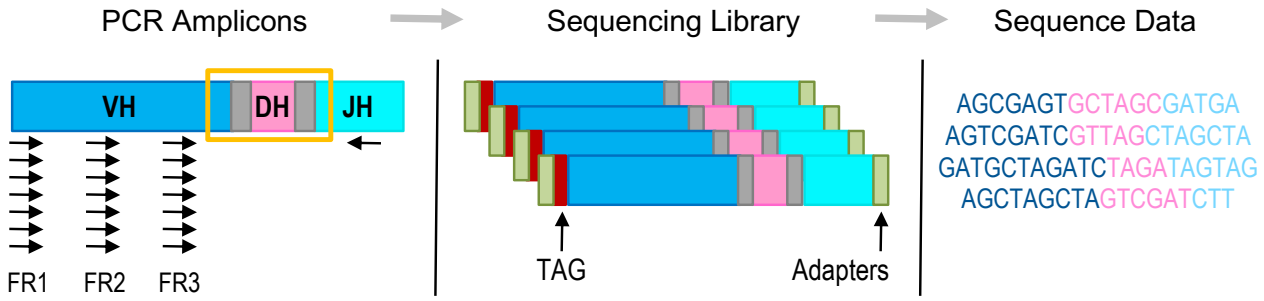


Figure 10. Next-generation sequencing workflow for IGH analysis. Schematic of two-stage PCR strategy (first PCR for targeted sequencing with reverse JH primer and forward VH primers complementary to FR1, FR2 or FR3; second PCR for library amplification) and sequencing output.

The presence of at least two identical sequencing reads, sharing the same V(D)J segment usage and the same CDR3 nucleotide sequence, defines a clonotype.[107] Any clonotype with a frequency of more than 5% (arbitrary cut-off)[108] in diagnostic samples identifies the condition of clonality, i.e. the presence of at least a clonal IG rearrangement. Frequencies of tumor clonotypes are

subsequently calculated in follow-up samples to quantify MRD levels during and after treatment.[94] MRD detection by NGS platforms provides an attractive alternative to MFC and ASO-qPCR. NGS technology is highly sensitive, reaching a potential sensitivity of 10^{-6} (Table 3). Compared with MFC, MRD assessment by NGS can be performed using stored samples. Compared with ASO-qPCR, NGS obviates the need for patient-specific primers or probes increasing applicability, reducing overall times for target detection and allowing the simultaneous analysis of all different IG gene rearrangements (Table 3). However, NGS-based MRD detection has some limitations: i) there is no international standardization regarding experimental design and data interpretation; ii) mainly two commercial platforms (LymphoSIGHT, now ClonoSEQ, and ImmunoSEQ, Adaptive Biotechnologies, Seattle, WA) are used increasing costs and turn-around time.[109]

Table 3. Features of currently available methods for MRD monitoring in MM. Adapted from[101].

	MFC (≥ 8-color)	**ASO-qPCR**	**NGS**
Applicability	~100%	60% to 70%	~90%
Availability in individual laboratories	High	Intermediate	Limited
Diagnostic sample	Important but not mandatory	Mandatory	Mandatory
Time	2-3 h	≥5 d (follow-up), 3-4 wk (target identification)	≥7 d
Cost per sample	~350 USD	500 USD (follow-up), 1500 USD at diagnosis (target identification)	~700 USD
Sensitivity	10^{-5} to 10^{-6}	10^{-5} to 10^{-6}	10^{-6}
Quantitative	Yes (directly; high accuracy)	Yes	Yes
Fresh sample	Needed (<36 h)	Not needed	Not needed
Patchy sample	Impacts	Impacts	Impacts
Global cell characterization	Yes	No	No
Standardization	Ongoing (EuroFlow/IMF)	Yes, since 15 y (EuroMRD)	Not reported
Clinical benefit associated with MRD-negative status	Improvements in PFS and OS	Improvements in PFS and OS	Improvements in PFS and OS

NGS has also been applied to investigate the clinical impact of MRD in MM. Martinez-Lopez et al.[108] have sequenced IGH gene rearrangements of MM patients treated front-line with novel therapies, and have demonstrated that different levels of MRD define different prognostic groups (Figure 11). MRD-negativity (frequencies $\leq 10^{-5}$) is significantly associated with longer time to progression (TTP; median 80 months *vs* 31 months; P<0.0001) and better OS (median not reached *vs* 81 months; P=0.02). When restricted to patients in CR, TTP median was 131 *vs* 35 months with P<0.0009.

Similar results were obtained by Avet-Loiseau et al.,[110] showing the correlation between NGS MRD status and PFS in MM patients before and after lenalidomide maintenance. Using a cut-off at 10^{-6}, patients with clonal IGH frequencies below 10^{-6} at pre-maintenance presented a 3-year PFS of 83% *vs* 53% for patients with frequencies $>10^{-6}$. At post-maintenance, these percentages were 90% *vs* 59%, respectively. Among patients achieving CR, the 3-year PFS was 87% and 63% at pre-maintenance, 92% and 64% at post-maintenance.

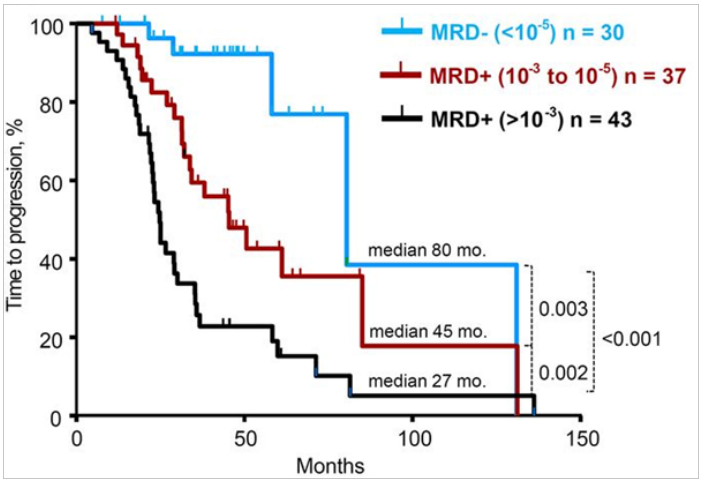


Figure 11. Time to progression of series stratified according to different MRD levels $>10^{-3}$ *vs* 10^{-3} to 10^{-5} *vs* $<10^{-5}$. Figure exported from[108].

6. Liquid biopsy in mature B-cell neoplasms

Biopsy of peripheral blood (PB), the so-called liquid biopsy, is emerging as a non-invasive strategy for disease monitoring and the advent of highly sensitive NGS technologies has fostered its use.[111]

Blood contains two types of cancer-derived materials that are susceptible to detailed molecular analysis: circulating tumor cells (CTCs) and circulating tumor cell-free DNA (cfDNA).[112]

Most hematological malignancies, such as leukemias, are circulating cancer; hence CTCs are often accessible in the blood at times of overt disease. Other malignancies, including lymphomas and myeloma, are mainly restricted to specific tissue compartments; therefore, CTCs shed from primary or metastatic tumor deposits are rare. Here, cfDNA can be used for non-invasive tumor monitoring.[113]

CTCs include all the structural and genetic components of a lymphoma and can be analyzed for recurrent translocations, such as t(14;18) [IGH-BCL2] and t(11;14) [IGH-CCND1] translocations, by PCR-based method; however, CTCs mirror circulating disease only.[112] On the other hand, cfDNA can be studied by NGS to reveal the presence of tumor V(D)J rearrangements and disease-specific mutations (Figure 12).[114,115] Because V(D)J rearrangements represent tumor- and patient-specific biomarkers, NGS of V(D)J cfDNA allows the direct measurement of tumor dynamics, including MRD levels. Mutation panels are instead more useful for tumor profiling (e.g., drug resistance mechanisms) and heterogeneity evaluation (e.g., clonal evolution).[116]

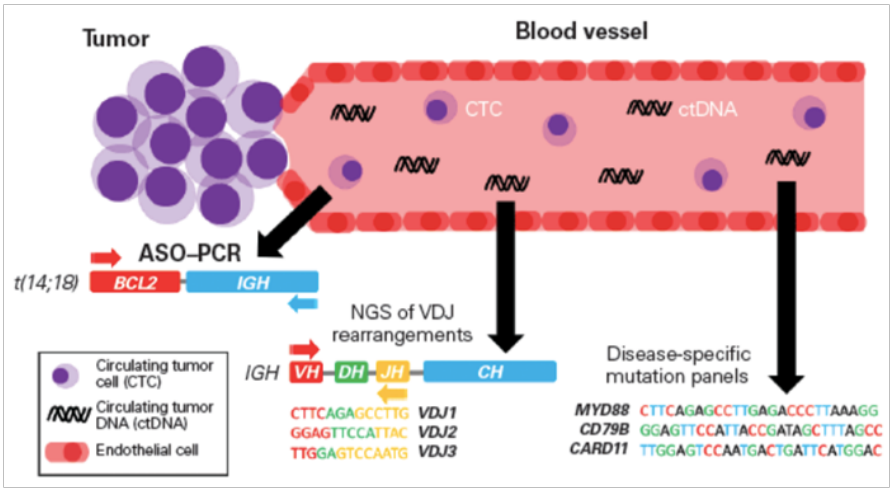


Figure 12. Molecular monitoring of circulating tumor biomarkers in B-cell malignancies. Figure exported from[112].

All cells, including tumor cells and healthy cells, shed fragments of DNA, called cfDNA, into the bloodstream.[117] The mechanisms whereby tumor cells release DNA into the blood are not well understood: DNA may enter the circulatory system via secretion from viable tumor cells, from phagocytes post-engulfment of tumor cells, or as result of tumor cell apoptosis and necrosis.[118] The rapid growth of a tumor is indeed associated with increased cellular turnover and number of apoptotic and necrotic cells, resulting in the accumulation of cellular debris released into the

circulation as cfDNA.[119] Assessment of the size distribution of cfDNA seems to confirm the apoptosis/necrosis mechanism as the main driver of release, since fragments' length corresponds to the size of DNA in one or multiple nucleoprotein complexes (Figure 13). Cell-free DNA can be sampled from either serum or plasma.[120] Serum generally contains 2-24 times more of cfDNA than plasma; however, this is likely due to the contamination by genomic DNA that originates from white blood cell lysis during clotting. Therefore, plasma is the preferred source of cfDNA.[121] Besides PB, cfDNA can be detected in other body fluids, such as saliva, synovial fluid and urine.[122]

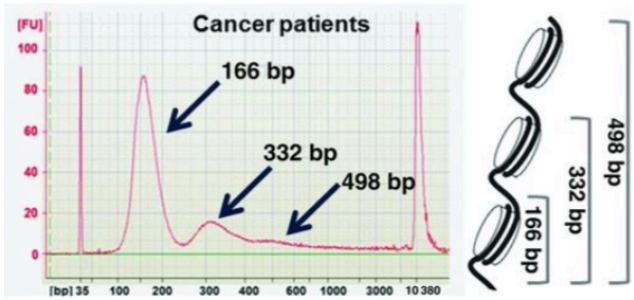


Figure 13. Size distribution of plasma cfDNA. Samples showed an enrichment in 166 bp fragments, corresponding to a single nucleosomal DNA. Longer fragments, corresponding to DNA on multiple nucleosomes, are also present event if in lower amount**.** X-axis: length of fragments. Y-axis: concentration of fragments with a specific length. Figure exported from[120].

In hematological neoplasms, tumor cfDNA analysis by NGS has important clinical applications.[109] In patients with curable lymphomas, such as early-stage HL and DLBCL, the eradication of all disease is necessary for cure. The current response criteria for lymphomas rely on CT and PET imaging, but they are fundamentally limited by radiation risks, costs, lack of tumor specificity, and inability to detect disease at the molecular level; in order to overcome these issues, today the liquid biopsy represents a sensitive and minimally invasive approach for disease monitoring.[123]

Two important studies have demonstrated the prognostic impact of tumor V(D)J cfDNA quantification in DLBCL.[114,124] Deep-sequencing of V(D)J genes in the tissue biopsy provides unique DNA sequences that are markers of clonality. The same V(D)J sequences can be detected in cfDNA at diagnosis as complementary biomarkers of tumor load, and during treatment as complementary biomarker of MRD level, reducing the detection limit beyond the capabilities of imaging techniques (Figure 14). Roschewski et al.[124] have indeed shown that: i) tumor V(D)J cfDNA levels correlate with clinical markers of tumor burden such as International Prognostic Index

and LDH (P<0.0001); ii) patients who clear tumor cfDNA after two cycles are significantly more likely to be progression-free at 5 years compared with patients who remain positive (80.2% *vs* 41.7%, P<0.0001; iii) surveillance monitoring of V(D)J cfDNA detected recurrence a median of 3.5 months prior to CT scans.

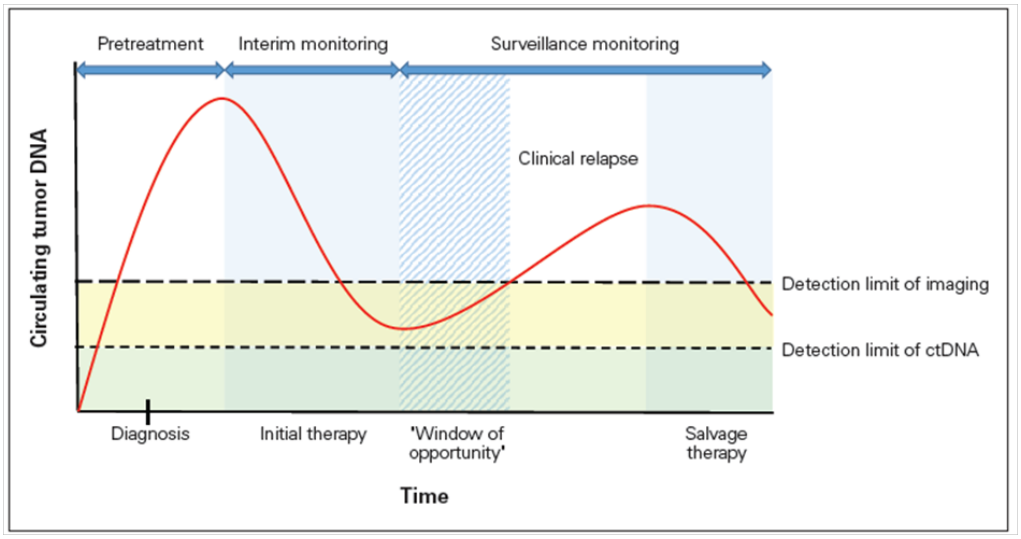


Figure 14. Disease monitoring by V(D)J cfDNA sequencing. Tumor cfDNA correlates with tumor kinetics and the possibility of frequent samplings creates a "window of opportunity" for early intervention with minimal tumor burden. Figure exported from[112].

In patients with indolent lymphomas, instead, where molecular remission generally do not indicate cure, Sarkozy et al.[125] have analyzed pre-treatment levels of circulating IG rearrangements in patients affected by FL, and have demonstrated that patients with higher levels are associated with shorter PFS (median 15.7 months *vs* median not reached; P=0.027).

In the MM context, few very recent studies seem to confirm the clinical relevance of cfDNA. Vij et al.[126] have employed the commercial LymphoSIGHT method (Sequenta, Inc., San Francisco, CA) to demonstrate a clear correlation (R^2 = 0.57) between myeloma clone levels in paired BM and PB plasma samples. Oberle and colleagues,[127] instead, have suggested an association between the presence of V(D)J cfDNA and response to therapy, showing that clonal V(D)J rearrangement levels remain positive in cfDNA and CTCs in 91% of non-responder patients and in 41% of patients that respond to treatment. However, these preliminary even if highly interesting data are not sufficient to assess the potential of cfDNA for MRD monitoring as only few patients reached CR and the treatment and monitoring strategies were not homogeneous.

AIMS OF THE PROJECT

1. Multiple myeloma

MM is the second most common hematologic neoplasm characterized by clonal proliferation of malignant BM PCs [Introduction, Paragraph 4]. Although the introduction of novel therapies such as IMiDs and PIs [Introduction, Paragraph 4.4] has increased rates of CR, most patients eventually relapse [Introduction, Paragraph 4.5] raising interest in more accurate method to evaluate residual disease. Response criteria now include MRD assessement by MFC, ASO-qPCR, and more recently NGS techniques [Introduction, Paragraph 5]. High-throughput sequencing methodologies provide an attractive alternative to MFC and to complex and time-consuming ASO-qPCR [Introduction, Paragraph 5.1]. MRD analysis by NGS mainly relies on the amplification and sequencing of the highly variable IGH gene rearrangements [Introduction, Paragraph 2.1], using consensus primers with a potential sensitivity of 10^{-6} [Introduction, Paragraph 5.2]. However, disease monitoring is conventionally performed on BM aspirates, which are invasive and therefore not feasible for a frequent MRD assessment during the course of treatment. Furthermore, given the patchy pattern of BM infiltration typically observed in MM, information acquired from a single BM aspirate provides only a snapshot of the tumor and might fail to reflect its heterogeneity [Introduction, Paragraph 5.2, Table 3, “Patch samples” row]. In order to overcome these issues, the NGS analysis of IGH rearrangements in PB samples, the so called liquid biopsy, is emerging as a complementary non-invasive approach for disease monitoring [Introduction, Paragraph 6]; nevertheless, its clinical utility is still under investigation and few data are available in MM context. My PhD project was designed to:

i) develop a novel in-house deep-sequencing method for the analysis of IGH gene rearrangements starting from tumor cells and cfDNA extracted from plasma samples;
ii) identify the tumor-associated IGH clonotype in BM tumor PCs of relapsed MM patients, and consequently in paired PB samples;
iii) assess the prognostic implication of clonotypic IGH cfDNA levels at baseline and during therapy.

2. Diffuse large B-cell lymphoma and Hodgkin lymphoma

DLBCL [Introduction, Paragraph 3.2.1] and HL [Introduction, Paragraph 3.1] are the two most frequent lymphomas in the Western world and derive from clonal expansion of GC B-cells. Although considered potentially curable tumors, ~50% of DLBCLs and ~20% of HLs are not cured by current treatments. Radiographic imaging modalities such as CT and PET are currently used for the assessment of therapeutic response although associated with limited specificity, low sensitivity and radiation exposure [Introduction, Paragraph 6].

The deep-sequencing of cfDNA encoding the highly specific tumor-associated IGH gene rearrangements can represent a non-invasive tool for MRD monitoring in DLBCL, and possibly also in HL.

Clinical implications of IGH cfDNA have been already extensively studied in DLBCLs [Introduction, Paragraph 6], but most of these studies have been conducted using commercial NGS platforms available at centralized laboratories, increasing the turn-around time and limiting clinical applicability.

On the other side, the application of a liquid biopsy approach for disease monitoring in HL patients might help to overcome the tissue biopsy limitations due to the low representation (<5%) of HRS malignant cells.

Additional aims of my PhD project were therefore to apply the NGS workflow for the analysis of the IGH gene rearrangements I have developed to:

i) identify the tumor-associated IGH clonotype in DLBCL patients using formalin-fixed paraffin-embedded (FFPE) samples and consequently monitor its levels in PB samples;
ii) analyze the clonal IGH gene rearrangements in PB samples of HL patients at the time of diagnosis, and then verify the presence of dominant rearrangements in diagnostic FFPE samples.

MATERIALS AND METHODS

1. Sample collection

1.1. Feasibility studies

For the evaluation of sensitivity, we employed Namalwa (human Burkitt lymphoma; ACC 24) and JVM-2 (human chronic B-cell leukemia; ACC 12) cell lines from DSMZ (Braunschweig, Germany).
In order to set up the experimental conditions for plasma analysis, PB of 10 healthy donors was collected in both K2-EDTA tubes (BD Vacutainer, Franklin Lakes, NJ) and Cell-Free DNA BCT tubes (Streck, Inc., Omaha, NE). PB samples in K2-EDTA tubes were processed within 3 hours. PB samples in Cell-Free DNA BCT tubes were processed after 3, 24, 48, 72 hours and one week.
Plasma from PB samples was obtained with a first centrifugation at 1500g for 10 minutes and a second high-speed centrifugation at 16000g for 10 minutes at 4°C. Plasma samples were stored at -80°C until extraction.[121]

1.2. Multiple myeloma

We conducted a prospective randomized phase III trial in MM patients at first relapse comparing the activity of bortezomib(Velcade)/cyclophosphamide/dexamethasone (VCD) *vs* lenalidomide(Revlimid)/cyclophosphamide/dexamethasone (RCD) as second-line therapy. Within this study, we analyzed 25 patients who achieved CR during the course of therapy (IMWG guidelines)[82] and had longitudinal biological samples collected for disease monitoring. The study was approved by our Institutional Review Board (IRB) and patients provided informed consent.
BM aspirates and PB samples collected in K2-EDTA tubes were obtained during routine clinical evaluations at study entry, after 3, 6, 9 cycles of therapy and at follow-up time points (Figure 15a).

1.3. Diffuse large B-cell lymphoma and Hodgkin lymphoma

We analyzed pre-treatment FFPE lymph node biopsies and PB samples of 26 patients affected by DLBCL and of 20 patients affected by HL. FFPE samples were obtained from tissue biopsies

collected during standard diagnostic procedures, with no additional discomfort for patients. For DLBCL patients, K2-EDTA PB samples were collected at diagnosis, during treatment after 2 and 4 cycles of R-CHOP-based chemotherapy, at the end of treatment (after 6 cycles), and during follow-up (Figure 15b). For HL patients, K2-EDTA PB samples were collected at diagnosis, during ABVD treatment (after 2 cycles for early-stage HL, after 2 and 4 cycles for the advance-stage HL), at the EOT (after 4 or 6 cycles for respectively early- or advance-stage), and during follow-up (Figure 15c). Imaging scans were acquired at baseline, at interim and at the end on treatment (Figure 15b-c). Patients provided written consent in accordance with local IRB requirements (INT128-16).

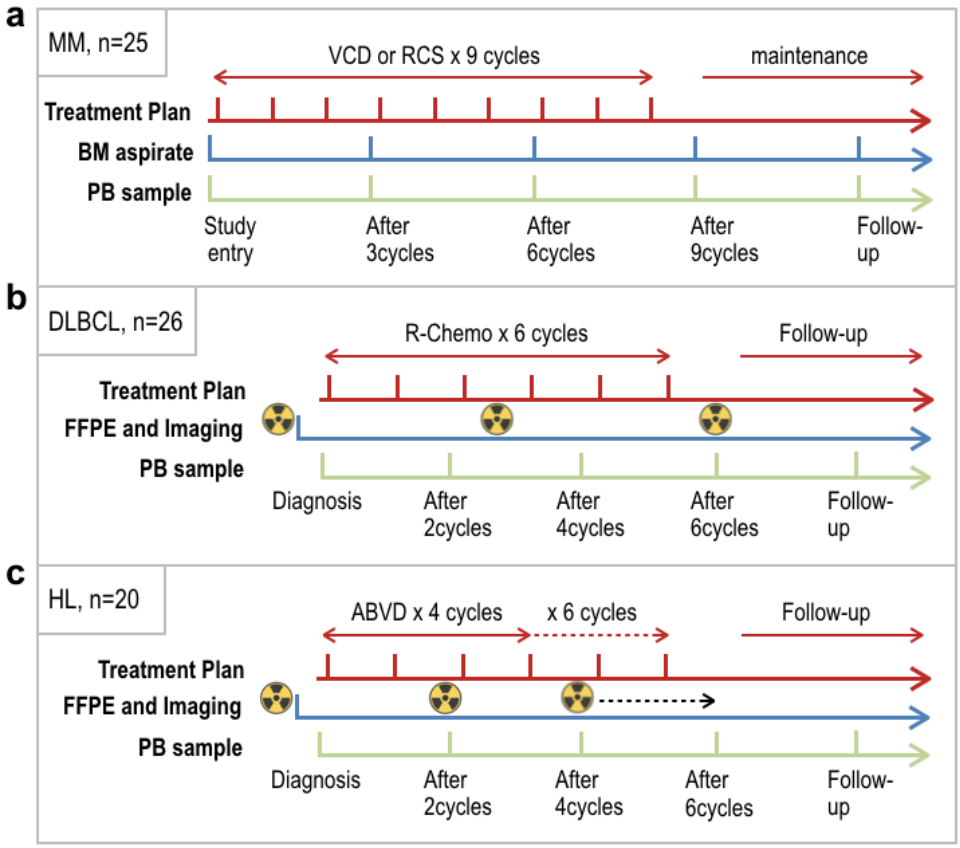


Figure 15. Tissue biopsy and PB sampling in patients affected by (a) multiple myeloma (MM), (b) diffuse large B-cell lymphoma (DLBCL), and (c) Hodgkin lymphoma (HL). Red arrows indicate treatment schedule, blue and green arrows indicate respectively tumor biopsies and PB samplings. In panel c, dashed arrows indicate 2 additional treatment cycles for advanced stage HL, corresponding to 1 addition PB sample, and EOT imaging after 6 instead of 4 cycles. BM, bone marrow; PB; peripheral blood; FFPE, formalin-fixed paraffin-embedded.
VCD, bortezomib(Velcade)/cyclophosphamide/dexamethasone.
RCD, lenalidomide(Revlimid)/cyclophosphamide/dexamethasone.
R-Chemo, rituximab-cyclophosphamide/doxorubicin/vincristine/prednisone-based chemotherapy.
ABVD, adriamycin(doxorubicin)/bleomycin/vinblastine/dacarbazine.

2. Isolation of myeloma plasma cells

2.1. Multiparameter flow cytometry

MFC is a technique that allows the analysis of physical and/or biochemical features of monodisperse cell suspensions; MFC is based on the refraction (scattering) of laser beam that passes through the sample, inside a laminar flow, and the light signal coming from any excited fluorochromes contained in the sample.[128]

The instrument used to perform flow cytometric analyses, called flow cytometer, has the following components (Figure 16):

i) a fluidic system, that forces cells to run one after the other thanks to the different pressure existing between the sample stream and the surrounding sheath stream (hydrodynamic focusing);

ii) an optical system, that generates and detects light signals. The light is either scattered or absorbed when it strikes a cell: light scatter is dependent on the internal structure of the cell (Side Scatter, SSC) and on its size and shape (Forward Scatter, FSC); light of appropriate wavelength is absorbed by fluorochrome-labeled antibodies and re-emitted as fluorescence signal allowing the identification of the different antigenic determinants;

iii) an electronic system, that converts optical signals firstly into proportional electrical pulses using photodiodes and photomultipliers, and then into digital signals for data visualization through a computer system;

iv) a software that allows the data analysis through the creation of histograms (fluorescence intensity on x-axis; number of cells expressing the antigen on y-axis) or of two-dimensional representations, also called dot plots, displaying the distribution of the various populations based on two or more parameters.

In order to avoid fluorescence artifacts, two steps are mandatory: the application of a compensation matrix and the doublet exclusion.

Despite the complexity of the optical system and the accurate choice of fluorochromes, it is possible that the emission wavelength of a fluorochrome overlaps the wavelength of another fluorochrome due to the spectrum amplitudes; it is therefore necessary to use flow cytometer built-

in functions that, using single stained controls, are able to calculate and subtract the spillover of one fluorochrome from the others creating a compensation matrix to apply before each acquisition.[129] Once the acquisition has been completed, it is necessary to exclude the doublets. If a fluorescence-positive and a fluorescence-negative cells pass through the laser concurrently (doublet), only the positive pulse will be produced leading to a false positive result. Doublets are characterized by double area and width whilst the height is roughly the same. Therefore disproportions between height, width and area can be used to identify and exclude doublets.[130]

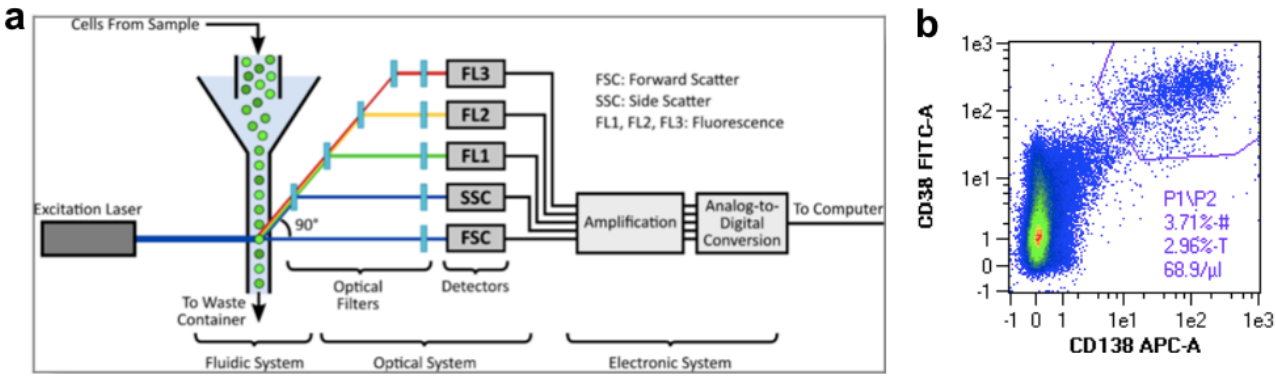


Figure 16. The working parts of the flow cytometer. Overview of (a) the three main components of a flow cytometer - the fluidics, optics and electronics - that work together to provide a complete system of (b) cell analysis.

Phenotypic aberrancies of MM PCs in BM aspirates were detected using MACSQuant Analyzer (Miltenyi Biotec, Gladbach, Germany) flow cytometer, equipped with three lasers (Violet 405nm, Blue 488nm, Red 635nm) and able to record eight fluorescence channels. Data were analyzed using MACSQuantify Software version 2.6 (Miltenyi Biotec) and FlowJo version 10.2 (FlowJo LLC, Ashland, OR).

2.2. Immunomagnetic separation

Cell immunomagnetic separation (IMS) is a technique for isolating specific cell populations based on antigen-specific antibodies coupled to superparamagnetic beads. Different systems are available for IMS: a column-based technology that relies on nano-sized superparamagnetic beads, and a column-free technology using micro-sized superparamagnetic beads. Moreover, there are two general strategies for IMS: positive separation, and negative separation.[131]

Positive separation means that a particular target cell type is magnetically labeled: during incubation with a cell suspension, the antibody/bead complex binds to cells expressing the corresponding antigen; when the cell suspension is placed into a magnetic field, only magnetically labeled cells are retained, while unlabeled cells are flushed out as negative fraction; after a washing step, the sample is removed from the magnetic field allowing the elution of target cells (Figure 17). On the other hand, negative separation is useful when antigen for a specific cell type is unknown: superparamagnetic beads are functionalized with a cocktail of antibodies that specifically bind to all other cell types in the solution except for the target cells; when sample passes through the magnetic field, all other cell types are retained whereas the remaining solution is enriched for the target cells.

Positive selection can be performed by direct or indirect magnetic labeling. In the case of direct magnetic labeling (Figure 17), the antibody/bead complex binds directly to antigens on the cell surface. For an indirect labeling approach, primary antibodies that bind to cell surface antigens and beads that bind to the primary antibodies are added to the cells in a two-step procedure.

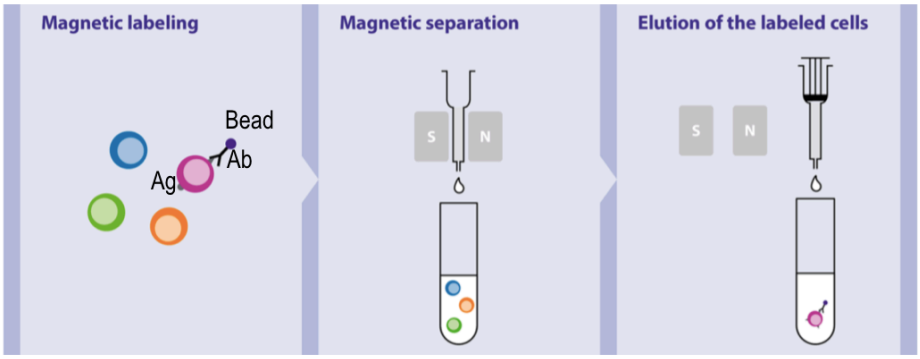


Figure 17. Positive immunomagnetic selection of a target cell. Ag, antigen; Ab, antibody; magnetic system is indicated by S/N poles.

For the separation of myeloma PCs, we used the AutoMACS ProSeparator (Miltenyi Biotec) applying a direct positive selection with sensitive double-column program ("Posselds"), that allows the isolation of rare cells with low antigen expression.

2.3. Protocol for flow cytometric MM immunophenotyping

PCs from a median of 8 mL (range: 7-14 mL) of BM aspirates were stained according to the

European Myeloma Network guidelines[91] using 8-color monoclonal antibody combinations (Table 4). The target for collection was >500000 cellular events.

Protocol:

1. Aliquot 100 µL of BM aspirate into 3 different 1.5 mL microcentrifuge tubes.
2. Add 50 µL of the first antibody combination (CD38, CD138, CD45, CD19, CD56, CD27, CD117 and CD81) into tube 1; 50 µL of the second antibody combination (CD38, CD138, CD20, CD117, CD28, CD27, CD45 and CD81) into tube 2; tube 3 remains unstained for the acquisition of physical parameters (FSC and SSC).
3. Mix well and incubate tube 1 and tube 2 at 4°C for 20 minutes in the dark.
4. Lyse red blood cells of all the three tubes with 1 mL of 1x lysis solution (NH_4Cl 0.15M, $KHCO_3$ 10nM and Na_4EDTA 1nM, pH 7.2-7.4).
5. Incubate at room temperature (RT) for 10 minutes in the dark.
6. Centrifuge at 5000g for 5 minutes.
7. Wash cells by adding 1 mL Running Buffer [phosphate buffered saline (PBS) + 0.5% bovine serum albumin (BSA)] and centrifuge at 5000g for 5 minutes.
8. Resuspend cells in 200 µL Running Buffer for flow cytometric analysis on MACSQuant Analyzer.

Table 4. List of monoclonal antibodies used for flow cytometric analysis of BM myeloma plasma cells. [1]Miltenyi Biotec, [2]Biolegend. "-" indicates negative expression; "+" indicates positive expression.

Antigen	Fluorochrome	Clone	Function	Expression in MM PCs
CD19[1]	PE	LT19	Type I transmembrane glycoprotein of 95 kDa that belongs to the IG superfamily. CD19 is expressed on B-cells throughout most stages of B-cell differentiation.	CD19 -
CD20[2]	VioBlue	LT20	Non-glycosylated transmembrane protein of 33-37 kDa that is expressed from the pre-B-cell stage to the B-cell lymphoblast stage. The antigen is not found on early B-cell progenitors or PCs.	CD20 +/-

CD27[1]	FITC	M-T271	Member of the tumor necrosis factor receptor family. CD27 is expressed at differing levels on memory B-cells, on a fraction of PCs, and on naïve and memory T-cells, but is not expressed on naïve B-cells or effector T-cells.	CD27 dim or -
CD28[1]	PE Vio770	15E8	Type I transmembrane protein, which is highly expressed on $CD3^+$ thymocytes, on most peripheral thymocytes and on PCs. Binding of CD28 to its ligands CD80 or CD86 costimulates T-cell effector function and T-cell-dependent antibody production *in vitro* and *in vivo*.	CD28 +
CD38[2]	BV500	HIT2	Single-chain type II transmembrane glycoprotein with enzymatic activity. It is present on the majority of hematopoietic cells, prevalent during early differentiation and activation processes. PCs express CD38 brightly. This antigen is considered a disease marker for human leukemias and myeloma.	CD38 +
CD45[1]	VioBlue APC Vio770	5B1	Member of tyrosine phosphatase family, CD45 is a type I transmembrane protein required for T- and B-cell activation; it is involved in cell growth, differentiation, cell cycle, and oncogenic transformation.	CD45 -
CD56[1]	PE Vio770 PE Cy7	REA196 MEM-188	Glycoprotein of the IG superfamily, also known as neural cell adhesion molecule (NCAM), which is expressed in practically all resting and activated NK cells and on a minor subset of $CD3^+$ T-cells.	CD56 +
CD81[1]	PerCP Vio700	REA513	Multi-pass membrane protein, member of the tetraspanin family, also known as TAPA-1. The 26 kDa cell surface protein is involved in adhesion, activation, proliferation, and differentiation of B-cells, T-cells, and other cells.	CD81 -
CD117[1]	PE APC Vio770	A3C6E2	Also known as c-kit, is a 145 kDa receptor tyrosine kinase involved in stem cell signaling, activation, and proliferation.	CD117 +
CD138[1]	APC	44F9	Transmembrane heparan sulfate proteoglycan macromolecule also known as syndecan-1. The CD138 molecule interacts with extracellular matrix proteins, cell surface molecules, and other soluble proteins. It is expressed on normal and malignant human PCs and on basolateral surfaces of endothelial cells.	CD138 +

2.4. Protocol for CD138$^+$ MM plasma cells' separation

Once verified, through fluorochrome-labeled antigen-specific antibodies (Table 4), the presence of a distinct population of myeloma PCs, the total volume of BM aspirate into a 15mL centrifuge tube was subjected to processing in order to isolate CD138$^+$ MM PCs for downstream analyses.

Protocol:

1. Centrifuge at 300g for 10 minutes to separate and discard plasma.
2. Lyse red blood cells 1:10 with 1x lysis solution.
3. Incubate at RT for 10 minutes.
4. After lysis, centrifuge immediately at 300g for 10 minutes and aspirate supernatant completely.
5. Wash cells by adding 10 mL Running Buffer and centrifuge at 300g for 10 minutes.
6. Resuspend cells in 10 mL Running Buffer and filter with 30 μm filter to remove aggregates.
7. Determine cell number with Türk solution (Crystal violet dye in 6% acetic acid) on a Bürker chamber (Sigma-Aldrich, Zwijndrecht, the Nederlands).
8. Centrifuge at 300g for 10 minutes and aspirate supernatant completely.
9. Resuspend pellet in 80 μL cold (4°C) Running Buffer per 2x10^7 total cells.
10. Add 20 μL CD138 MicroBeads (Miltenyi Biotec) per 2x10^7 total cells.
11. Mix well and incubate at 4°C for 15 minutes.
12. Wash cells by adding 10 mL cold Running Buffer, centrifuge at 300g for 10 minutes and aspirate supernatant completely.
13. Resuspend in 2 mL cold Running Buffer and proceed to immunomagnetic separation of CD138$^+$ MM PCs on AutoMACS ProSeparator.

Isolated CD138$^+$ cells can be subsequently phenotypically re-analyzed for purity evaluation, using the following protocol:

1. Add 10 μL anti-CD138 and anti-CD38 antibodies to 100 μL of positive fraction.
2. Mix well and incubate at 4°C for 20 minutes in the dark.
3. Centrifuge at 5000g for 5 minutes.
4. Wash cells by adding 1 mL Running Buffer and centrifuge at 5000g for 5 minutes.
5. Resuspend in 200 μL Running Buffer for MFC analysis (usually CD138$^+$ cells are ≥98% pure).

3. DNA extraction

3.1. Genomic DNA from cells

Genomic DNA (gDNA) was extracted from 10^6 Namalwa and JVM-2 cell, and from 10^6 isolated BM $CD138^+$ PCs, by silica membrane technology using NucleoSpin Tissue kit (Macherey-Nagel GmbH &Co., Düren, Germany).

Protocol:

1. Minimize RNA contamination: resuspend cells (into a 1.5 mL microcentrifuge tube) in 100 µL PBS, add 5 µL RNase A (100 mg/mL) and incubate at RT for 5 minutes.
2. Lyse sample: add 200 µL Pre-Lysis Buffer T1, 25 µL Proteinase K solution, 200 µL Lysis Buffer B3, and incubate the sample at 70°C for 10 minutes.
3. Adjust DNA binding conditions: add 210 µL Ethanol (EtOH) 100% to the sample and vortex.
4. Bind DNA: place one NucleoSpin Tissue Column into a 2 mL collection tube, apply the sample to the column, centrifuge at 11000g for 1 minute, discard the flow-through, and place the column back into the collection tube.
5. First wash of silica membrane: add 500 µL Buffer BW, centrifuge at 11000g for 1 minute, discard the flow-through, and place the column back into the collection tube.
6. Second wash of silica membrane: add 600 µL Buffer B5, centrifuge at 11000g for 1 minute, discard the flow-through, and place the column back into the collection tube.
7. Dry silica membrane: centrifuge at 11000g for 1 minute to remove residual EtOH.
8. Elute highly pure DNA: place the column into a 1.5 mL microcentrifuge tube, add 70 µL pre-warmed (70°C) Elution Buffer, incubate at RT for 1 minute, centrifuge at 11000g for 1 minute.
9. Store at 4°C.

3.2. Genomic DNA from FFPE samples

For DLBCL and HL patients, gDNA was extracted from archival pre-treatment FFPE samples (three to four 3 µm tissue curls processed by Hematopathology Unit). We decided to use the

GeneRead DNA FFPE kit (Qiagen, Hilden, Germany) following an adjusted protocol to increase DNA purity and concentration.

Stage I, Deparaffinization and Lysis:

1. Resuspend curls (into a 1.5 mL microcentrifuge tube) in 1 mL of xylene.
2. Incubate at 56°C for 4 minutes (vortex briefly after 2 minutes).
3. Centrifuge at 13000rpm for 3 minutes.
4. Aspirate supernatant completely without disturbing pellet.
5. Repeat steps 1-4.
6. Resuspend pellet in 1 mL of EtOH 100%.
7. Incubate at RT for 4 minutes.
8. Centrifuge at 13000rpm for 3 minutes.
9. Aspirate supernatant completely without disturbing pellet.
10. Repeat steps 6-9.
11. Incubate at 56°C for 10 minutes (or until complete EtOH evaporation avoiding to completely dry out the pellet).
12. Add 55 µL RNase-free water, 25 µL Buffer FTB, and 20 µL Proteinase K.
13. Vortex and briefly centrifuge the sample.
14. Incubate at 56°C overnight.

Stage II, Artifacts' Digestion:

15. Next morning, incubate at 90°C for 1 hour to reverse formalin cross-links (leave the sample at RT after the 56°C incubation in step 14 until the heating block has reached 90°C).
16. Briefly centrifuge the tube to remove drops from inside the lid.
17. Add 115 µL RNase-free water, and mix.
18. Add 35 µL UNG to reduce deaminated cytosine bases that leads to a C-T conversion during sequencing reaction.
19. Incubate at 50°C for 1 hour.

20. Briefly centrifuge the tube to remove drops from inside the lid.

Stage III, DNA binding to silica membrane column:

21. Add 2 µL RNase A (100 mg/mL), mix, and incubate at RT for 2 min.

22. Add 250 µL Buffer AL, and mix thoroughly by vortexing. Then add 250 µL EtOH 100% and mix again thoroughly by vortexing. Centrifuge briefly to remove drops from lid.

23. Transfer 700 µL of lysate to the QIAamp MinElute column (into a 2 mL collection tube), and centrifuge at 13000rpm for 1 minute. Discard the flow-through and reuse the collection tube.

24. Repeat step 23 until the complete lysate is used.

25. Add 500 µL Buffer AW1 to the spin column, and centrifuge at 13000rpm for 1 minute. Discard the flow-through and reuse the collection tube.

26. Add 500 µL Buffer AW2 to the spin column, and centrifuge at 13000rpm for 1 minute. Discard the flow-through and reuse the collection tube.

27. Add 250 µL EtOH 100% to the spin column, and centrifuge at 13000rpm for 1 minute. Discard the flow-through and the collection tube.

28. Place the spin column into a new 2 mL collection tube, and centrifuge at 13000rpm for 1 minute to remove any residual liquid.

29. Place the column into a clean 1.5 mL microcentrifuge tube, carefully open the lid and apply 25 µL Buffer ATE to the center of the membrane.

30. Incubate at RT for 5 minutes.

31. Centrifuge at 13000rpm for 1 minute.

32. Store at -20°C.

3.3. Cell-free DNA

Cell-free DNA was extracted from 1-3 mL of plasma using QIAmp Circulating Nucleic Acid kit (Qiagen) following protocol for processing ≤ 3 mL of serum or plasma.

The procedure comprises 4 steps (lyse, bind, wash, elute) and is carried out using silica membrane columns on a vacuum manifold (Figure 18). Green text denotes reagent volumes for 1

mL of plasma, blue text for 2 mL; red text for 3 mL.

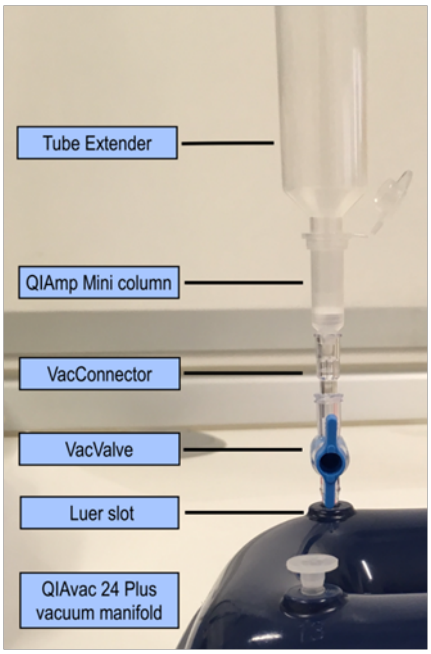


Figure 18. Setting up the QIAvac 24 Plus vacuum manifold with QIAamp Mini columns using VacValves, VacConnectors, and Tube Extenders.

Protocol:

1. Pipet 100 µL, 200 µL, or 300 µL of QIAGEN Proteinase K into a 15 mL centrifuge tube.
2. Add 1 mL, 2 mL, or 3 mL of plasma (after equilibration to RT).
3. Add 0.8 mL, 1.6 mL, or 2.4 mL Buffer ACL (containing 1.0 µg carrier RNA). Mix by pulse-vortexing for 30 seconds.
4. Incubate at 60°C for 30 minutes.
5. Add 1.8 mL, 3.6 mL, or 5.4 mL Buffer ACB to the lysate in the tube. Mix by pulse-vortexing for 30 seconds.
6. Incubate for 5 minutes on ice.
7. Insert the QIAamp Mini column (stored at 4°C) into the VacConnector on the QIAvac 24 Plus. Insert a 20 mL tube extender into the open QIAamp Mini column (Figure 18).
8. Apply the lysate-Buffer ACB mixture from step 6 into the tube extender. Switch on the vacuum pump. When all of mixture have been drawn through the columns completely, switch off the vacuum pump and release the pressure to 0 mbar. Remove and discard the tube extender.
9. Apply 600 µL Buffer ACW1 to the column. Leave the lid of the column open, and switch on the

vacuum pump. After all of Buffer ACW1 has been drawn through the column, switch off the vacuum pump and release the pressure to 0 mbar.

10. Apply 750 µL Buffer ACW2 to the column. Leave the lid of the column open, and switch on the vacuum pump. After all of Buffer ACW2 has been drawn through the column, switch off the vacuum pump and release the pressure to 0 mbar.

11. Apply 750 µL EtOH 100% to the column. Leave the lid of the column open, and switch on the vacuum pump. After all of EtOH has been drawn through the column, switch off the vacuum pump and release the pressure to 0 mbar.

12. Close the lid of the column and remove it from the vacuum manifold. Place the column into a clean 2 mL collection tube, and centrifuge at 14000rpm for 3 minutes.

13. Place the column into a new 2 ml collection tube. Open the lid, and incubate at 56°C for 10 minutes to dry the membrane completely.

14. Place the column into a clean 1.5 mL microcentrifuge tube, apply 50 µL Buffer AVE to the center of the membrane, and incubate at RT for 3 minutes.

15. Centrifuge at 14000rpm for 1 minute to elute the nucleic acid.

16. Store cfDNA at -20°C.

4. Quantity and quality assessment of extracted DNA

4.1. Quantity analysis

Extracted gDNA and cfDNA were quantified on the bench-top fluorimeter Qubit 2.0 (Thermo Fisher Scientific, Waltham, MA) using the high sensitive kit (Qubit dsDNA HS Assay Kit, Thermo Fisher Scientific). DNA quantification is based on the fluorescent signal emitted by a fluorescent dye only when specifically bound to the double-stranded DNA. The unbound dye is characterized by extremely low fluorescence; upon binding, the amount of fluorescence signal is directly proportional to the concentration of DNA in the solution; the fluorimeter detectes the signal and calculates the DNA concentration comparing sample fluorescence intensity to the fluorescence emitted by DNA standards of known concentration.[132]

Protocol:

1. Prepare a Working Solution composed by 199 µL HS Buffer and 1µL HS Reagent (dye), then mix by vortexing for 2-3 seconds.
2. Add 198 µL of Working Solution to 2 µL of sample, and vortex for 2-3 seconds without creating bubbles.
3. Incubate at RT for 2 minutes in the dark.
4. Read sample at Qubit 2.0 according to instrument instructions.

4.2. Quality analysis

In order to verify that fragments' size of extracted DNA, mainly gDNA from FFPE and cfDNA, was compatible with downstream analyses, we used the Agilent 2100 Bioanalyzer (Agilent Technologies, Böblingen, Germany), a chip-based capillary electrophoresis machine. The chip accommodates sample wells, gel wells, a well for an external standard (ladder), micro-channels for well interconnection, and 16 electrodes (one for each well).[133] During chip preparation, the micro-channels are filled with a sieving polymer and a fluorescence dye. Charged DNA molecules are electrophoretically driven by a voltage gradient and separated by size. Moreover, dye molecules intercalating into DNA emit a fluorescent signal. Using ladder with fragments of known sizes and concentrations, a standard curve is plotted, and the sample fragments' size and concentration are calculated according to the migration time and the fluorescence intensity, respectively. The drift effect that may occur during chip run is compensate by another standard within each well containing the "lower" and "upper" markers.

Stage I, Gel-Dye Mix Preparation:

1. Allow High Sensitivity DNA dye concentrate and High Sensitivity DNA gel matrix to equilibrate to RT for 30 minutes.
2. Add 15 µL High Sensitivity DNA dye concentrate to a High Sensitivity DNA gel matrix vial.
3. Vortex solution well and spin down. Transfer to spin filter.
4. Centrifuge at 2250g for 15 minutes. Store at 4°C in the dark and use within 6 weeks.

Stage II, Chip Loading:

5. Allow the gel-dye mix to equilibrate to RT for 30 minutes.

6. Put a new High Sensitivity DNA chip on the chip priming station.

7. Pipette 9 µL of gel-dye mix in the well marked with an encircled G (Figure 19).

8. Make sure that the syringe plunger is positioned at 1 mL, close the chip priming station, and press plunger until it is held by the clip.

9. Wait for exactly 60 seconds then release clip.

10. Wait for 5 seconds, then slowly pull back the plunger to the 1 mL position.

11. Open the chip priming station and pipette 9 µL of gel-dye mix in the G wells (Figure 19).

12. Pipette 5 µL of internal marker in all sample and ladder wells.

13. Pipette 1 µL of High Sensitivity DNA ladder in the well marked with dsDNA symbol (Figure 19).

14. In each of the 11 sample wells (Figure 19), pipette 1 µL of sample (used wells) or 1 µL of marker (unused wells).

15. Put the chip horizontally in the adapter and vortex at 2000rpm for 1 minute.

16. Run the chip in the Agilent 2100 Bioanalyzer within 5 minutes following instrument instructions.

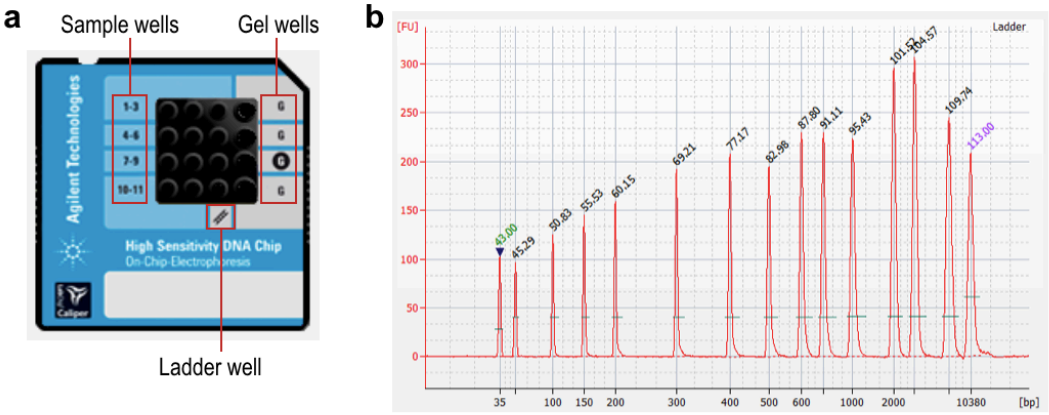
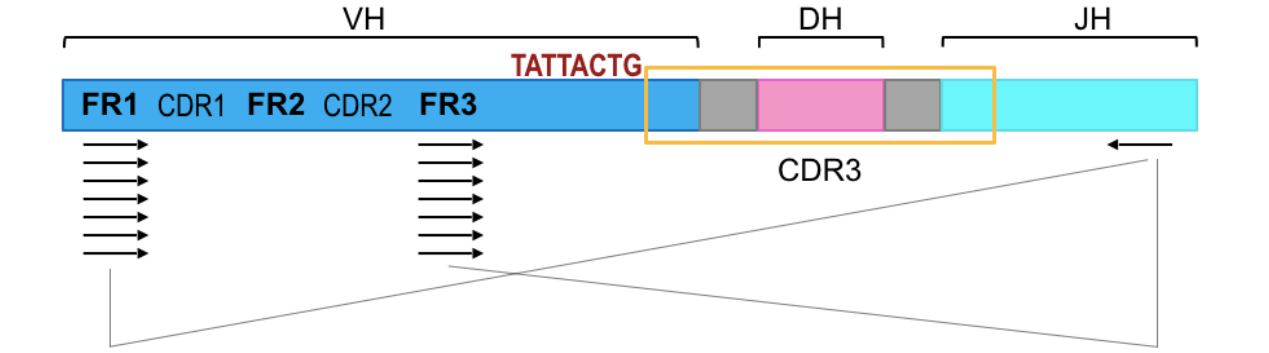


Figure 19. Agilent High Sensitivity DNA kit for DNA quality assessment. (a) High Sensitivity DNA chip containing 11 sample wells, 4 gel wells and 1 ladder well, as indicated by red boxes. (b) Electropherogram visualization of High Sensitivity DNA Ladder: the first peak marked with green text corresponds to the lower marker, the next 13 peaks correspond to ladder fragments of known sizes and concentrations, the last peak marked with violet text corresponds to the upper marker; x-axis indicates the size (bp) based on the migration time, y-axis indicates the fluorescence units (FU) for DNA quantification.

5. IGH gene rearrangements' amplification

Amplification of IGH gene rearrangements, in both gDNA and cfDNA samples, was performed by multiplex PCR using a consensus reverse JH primer and in the forward direction a mix of degenerate VH family-specific primers complementary to FR1 or FR3 (Figure 20). Each of these two independent primer sets is able to amplify all VH families[96] without any significant amplification bias,[134] and has been already validated by the BIOMED-2 Concerted Action for the application in multiplex assays.[9]



Primer	Sequence
VH1D	5'-CCTCAGTGAAGGTCTCCTGCAAGG-3'
VH2D	5'-TCCTGCGCTGGTGAAAGCCACACA-3'
VH3D	5'-GGTCCCTGAGACTCTCCTGTGCA-3'
VH4aD	5'-TCGGAGACCCTGTCCCTCACCTGCA-3'
VH4bD	5'-CGCTGTCTCTGGTTACTCCATCAG-3'
VH5D	5'-GAAAAAGCCCGGGGAGTCTCTGAA-3'
VH6D	5'-CCTGTGCCATCTCCGGGGACAGTG-3'
JHD	5'-ACCTGAGGAGACGGTGACCAGGGT-3'

Primer	Sequence
VH1-FR3	5'-TGGAGCTGAGCAGCCTGAGATCTGA-3'
VH2-FR3	5'-CAATGACCAACATGGACCCTGTGGA-3'
VH3-FR3	5'-TCTGCAAATGAACAGCCTGAGAGCC-3'
VH4-FR3	5'-GAGCTCTGTGACCGCCGCGGACACG-3'
VH5-FR3	5'-CAGCACCGCCTACCTGCAGTGGAGC-3'
VH6-FR3	5'-GTTCTCCCTGCAGCTGAACTCTGTG-3'
VH7-FR3	5'-CAGCACGGCATATCTGCAGATCAG-3'
JHD	5'-ACCTGAGGAGACGGTGACCAGGGT-3'

Figure 20. PCR analysis of IGH gene rearrangements. Schematic diagram of rearranged IGH locus with the used multiplex PCR tubes containing consensus reverse JH primer (JHD) and one of the two VH family-specific primer sets (on the left, forward primers complementary to FR1; on the right, forward primers complementary to FR3). Red text indicates a highly conservative sequence preceding the highly variable CDR3.

The gDNA PCR reactions were performed in 50 µL with the following components:

- 5 µL of dNTPs 2 mM
- 5 µL of Buffer 10X
- 5 µL of $MgCl_2$ (FR1: 20 mM; FR3: 15 mM)

- 1 µL of each primer 10 µM (seven different forward FR1 or FR3 primers and one reverse JH primer)
- 0.2 µL of Taq Gold 5U/µL, and
- 1-26.8 µL of gDNA: 500 ng for Namalwa and JVM-2 cell lines, 500 ng for BM $CD138^+$ cells from MM patients; 40 ng for FFPE samples from DLBCL and HL patients.

The cfDNA PCR reactions were performed in 85 µL with the following components:

- 8.5 µL of dNTPs 2 mM
- 8.5 µL of Buffer 10X
- 8.5 µL of $MgCl_2$ (FR1: 20 mM; FR3: 15 mM)
- 1 µL of each primer 10 µM (seven different forward FR1 or FR3 primers and one reverse JH primer)
- 0.2 µL of Taq Gold 5U/µL, and
- 45 µL of cfDNA (all amount of extracted cfDNA, regardless of obtained concentration).

PCR conditions were 96°C for 10 minutes followed by 35 cycles of 96°C for 30 seconds, 62°C for 30 seconds, 72°C for 30 seconds, and a final extension at 72°C (FR1: 10 minutes; FR3: 5 minutes) with cooling to 4°C. PCRs were performed on a ProFlex thermal cycler (Thermo Fisher Scientific).

PCR products were controlled for quantity and quality on Agilent 2100 Bioanalyzer using the High Sensitivity DNA kit (Agilent Technologies).

6. Ion Torrent next-generation sequencing

Ion Torrent (Thermo Fisher Scientific) sequencing technology employs a semiconductor chip that directly translates chemically encoded information (A, C, G, T) into digital information (0, 1) for base calling, increasing speed and scalability and reducing costs compared to other fluorescence-based NGS platforms. Each Ion chip contains millions of wells with corresponding Ion sensors that

allow parallel detection of multiple sequencing reactions.[135]

Ion Torrent sequencing workflow (Figure 21) includes: i) library preparation; ii) template preparation; iii) sequencing.

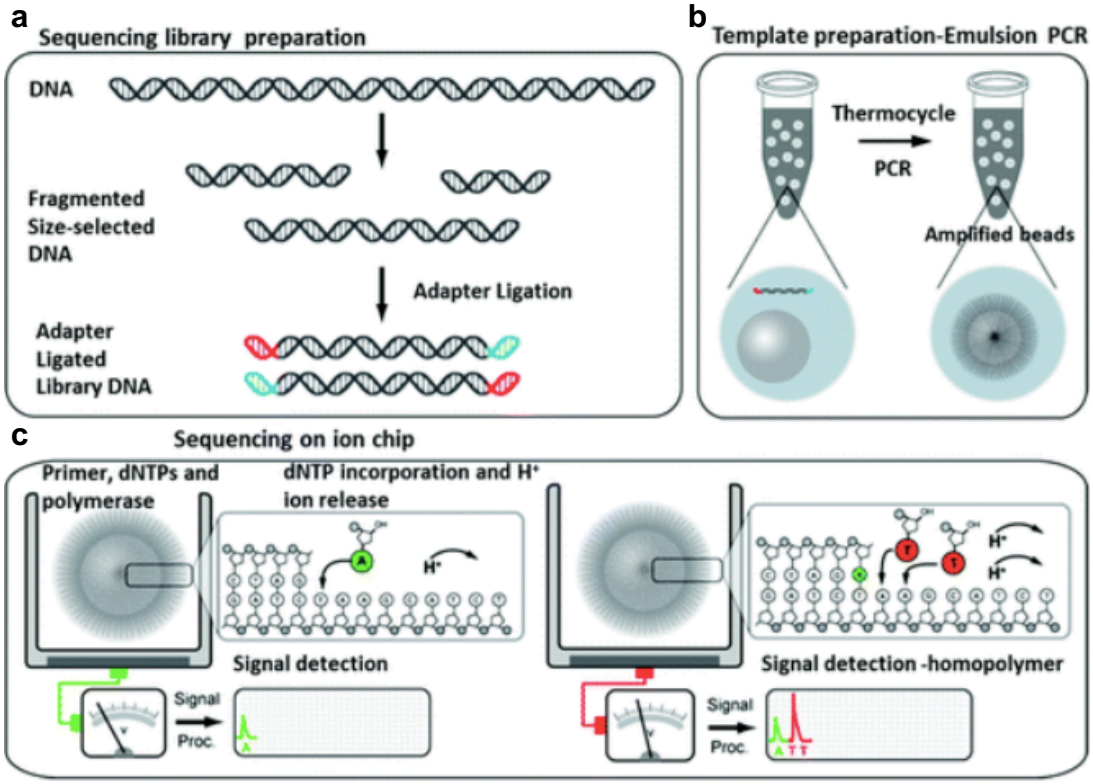


Figure 21. Ion Torrent sequencing wokflow (a-c).

The process starts with the preparation of libraries through the ligation of specific adapters to size-specific PCR amplicons for platform processing and patients' identification (Figure 21a). Each library fragment is then clonally amplified on the surface of a DNA capture bead compartmentalized in a water-in-oil droplet (emulsion PCR) (Figure 21b); after amplification, the enrichment process selects the template-positive beads only. The semiconductor chip is loaded with these beads, each depositing into a well, and placed on the sequencer where it is sequentially flooded with one of the four DNA nucleotides (A, C, G, or T); whenever a nucleotide is incorporated into the growing complementary strand, a positively charged hydrogen (H^+) ion is naturally released, changing the pH of the well solution. A sensitive layer beneath the well measures this pH change (ΔpH) and converts it to voltage; the voltage change (ΔV) is recorded by an Ion sensor indicating that the nucleotide was incorporated and the base is called. If there are two identical

nucleotides next to each other on the DNA template, the voltage is double, and the chip records two identical nucleotides (Figure 21c). If the next nucleotide that floods the chip is not a match, no H^+ ion is released, no voltage change is recorded, and no base is called.

After sequencing, data are automatically transferred to the Torrent Server where processing algorithms produce the DNA sequences associated with individual reads for downstream analyses.

6.1. Library preparation

Aadapter ligation and barcoding were performed using Ion Plus Fragment Library kit and Ion Xpress Barcode Adapters 1-16/17-32 kits (Thermo Fisher Scientific), following the protocol to prepare libraries without fragmentation. Adapter-ligated libraries were purified using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA) according to the library size (FR1 >300 bp; FR3 <200 bp).

Stage I, Purification of Multiplex PCR Amplicons:

1. Resuspend the Agencourt AMPure XP beads, and allow the suspension to come to RT (~30 minutes).
2. Add 1.8X sample volume of beads (90 µL for gDNA, 153 µL for cfDNA) in each PCR tube, pipet up and down to thoroughly mix and incubate at RT for 5 minutes.
3. Place PCR tubes on the DynaMag-2 magnetic rack (Thermo Fisher Scientific) for 5 minutes or until the solution clears. Remove and discard the supernatant from each tube without disturbing the bead pellet.
4. Without removing the samples from the magnet, dispense 150 µL of freshly prepared EtOH 70% into each tube. Incubate the samples at RT for 30 seconds allowing the movement of beads through EtOH. After the solution clears, remove and discard the supernatant without disturbing the pellet.
5. Repeat step 4 for a second wash.
6. To remove residual EtOH, keep the sample on the magnet and carefully remove any remaining supernatant with a 20µL pipettor without disturbing the pellet.

7. Keeping the sample on the magnet, air-dry the beads at RT for 3-5 minutes (ensure that the pellet does not dry out completely).
8. Remove tubes from the magnet, and add 15 µL of Nuclease-free Water directly to each bead pellet to disperse the beads. Pipet the mixture up and down five times to mix thoroughly, and incubate at RT for 1 minute.
9. Place tubes on the magnet again for 5 minutes. After the solution clears, transfer the supernatant containing the purified amplicons to new tubes without disturbing the pellet.
10. Analyze 1 µL of each tube using Agilent 2100 Bioanalyzer instrument and Agilent High Sensitivity DNA Kit to determine quality and quantity.

[Optional] Store purified amplicons at -20°C.

Stage II, End Repair:

11. Prepare 10-100 ng of the purified amplicons in a total volume of 79 µL with Nuclease-free Water.
12. Mix by pipetting in a 1.5mL LoBind microcentrifuge tubes (Eppendorf, Hamburg, Germany):

Component	**Volume**
Purified amplicons, 10-100 ng	79 µL
5X End Repair Buffer	20 µL
End Repair Enzyme	1 µL
Total	*100 µL*

13. Incubate the end repair reaction at RT for 20 minutes.

Stage III, Purification of End-Repaired DNA:

14. Add 180 µL of Agencourt AMPure XP beads (1.8X sample volume) to 1.5mL LoBind tubes, pipet up and down to thoroughly mix and incubate at RT for 5 minutes.
15. Place PCR tubes on the DynaMag-2 magnetic rack (Thermo Fisher Scientific) for 5 minutes or until the solution clears. Remove and discard the supernatant from each tube without disturbing the bead pellet.

16. Without removing the samples from the magnet, dispense 500 µL of freshly prepared EtOH 70% into each tube. Incubate the samples at RT for 30 seconds allowing the movement of beads through EtOH. After the solution clears, remove and discard supernatant without disturbing pellet.

17. Repeat step 16 for a second wash.

18. To remove residual EtOH, keep the sample on the magnet and carefully remove any remaining supernatant with a 20µL pipettor without disturbing the pellet.

19. Keeping the sample on the magnet, air-dry the beads at RT for 3-5 minutes (ensure that the pellet does not dry out completely).

20. Remove tubes from the magnet, and add 28 µL of Low TE directly to each bead pellet to disperse the beads. Pipet the mixture up and down at least five times to mix thoroughly, and incubate at RT for 1 minute.

21. Place tubes on the magnet again for 5 minutes. After the solution clears, transfer the supernatant containing the eluted DNA to new 1.5mL LoBind tubes without disturbing the pellet.

[Optional] Store at -20°C.

Stage IV, Adapter Ligation and Nick Repair:

22. In a 0.2-mL PCR tube, combine the reagents as indicated in the table, and mix well by pipetting up and down.

Component	**Volume**
DNA	25 µL
10X Ligase Buffer	10 µL
Ion P1 Adapter	2 µL
Ion Xpress Barcode X[1]	2 µL
dNTP Mix	2 µL
Nuclease-free Water	49 µL
DNA Ligase	2 µL
Nick Repair Polymerase	8 µL
Total	*100 µL*

[1] X = barcode chosen for patient identification.

23. Place the tube in the thermal cycler (ProFlex) and run the following program:

Temperature	**Volume**
25°C	15 min
72°C	5 min
4°C	up to 1h[1]

[1] Continue directly to the purification step.

24. Transfer the entire reaction mixtures to new 1.5mL LoBind tubes for next cleanup step.

Stage V, Purification of the adapter-ligated libraries:

25. Add 100 µL (1X FR1 library volume) or 150 µL (1.5X FR3 library volume) of Agencourt AMPure XP beads to 1.5mL LoBind tubes, pipet up and down to thoroughly mix and incubate at RT for 5 minutes.

26. Place PCR tubes on the DynaMag-2 magnetic rack (Thermo Fisher Scientific) for 5 minutes or until the solution clears. Remove and discard the supernatant from each tube without disturbing the bead pellet.

27. Without removing the samples from the magnet, dispense 500 µL of freshly prepared EtOH 70% into each tube. Incubate the samples at RT for 30 seconds allowing the movement of beads through EtOH. After the solution clears, remove and discard the supernatant without disturbing the pellet.

28. Repeat step 27 for a second wash.

29. To remove residual EtOH, keep the sample on the magnet and carefully remove any remaining supernatant with a 20µL pipettor without disturbing the pellet.

30. Keeping the sample on the magnet, air-dry the beads at RT for 3-5 minutes (ensure that the pellet does not dry out completely).

31. Remove tubes from the magnet, and add 23 µL of Low TE directly to each bead pellet to disperse the beads. Pipet the mixture up and down at least five times to mix thoroughly, and incubate at RT for 1 minute.

32. Place tubes on the magnet again for 5 minutes. After the solution clears, transfer the

supernatant containing the eluted DNA to new 1.5mL LoBind tubes without disturbing the pellet.

33. Store libraries at -20°C.

6.2. Template preparation

Libraries were pooled in equimolar concentrations and template preparation was performed using Ion Chef Hi-Q View kit (Thermo Fisher Scientific) following manufacturer's instructions. Ion Chef system requires two library pools for each run and two different runs for libraries of different length (≥ 300 bp or < 300 bp), therefore FR1 and FR3 pools were subjected to separate runs.

Stage I, Pool barcoded libraries:

1. Analyze 1 µL of each barcoded library using Qubit 2.0 instrument and Qubit dsDNA HS Assay Kit to determine library concentration in ng/mL.
2. FR1 libraries: pool 10 µL of each library 10 ng/mL (50 pM); FR3 libraries: pool 10 µL of each library 5 ng/mL (20 pM).

Stage II, Ion Chef Run:

3. Allow Ion PGM Hi-Q View Chef Reagents cartridge to warm to RT at least 45 minutes.
4. Create a Planned Run selecting the 400bp-protocol for FR1 libraries, or the 200bp-protocol for FR3 libraries.
5. Pipet 15 µL of each pool to the bottom of the appropriate Ion Chef Library Sample Tube, add 10 µL of Calibration Standard panel (Thermo Fisher Scientific) for a greater accuracy of base calling, vortex, and centrifuge for 2-3 seconds.
6. Load the Ion Chef instrument following manufacturer's instructions and start the run.

6.3. Sequencing

Libraries were sequenced on Ion Personal Genome Machine (PGM) Sequencer with Ion PGM Hi-Q View Sequencing kit using Ion 316 or 318 Chips v2 BC (Thermo Fisher Scientific). We performed a targeted sequencing using GRCh37/hg19 (Homo sapiens) as reference genome, and

IGH locus BED file downloaded from UCSC Genome Browser (https://www.genome.ucsc.edu) as target region. Reproducibility of Ion Torrent sequencing workflow was tested by running gDNA samples in duplicate with a different barcode in the same run and with the same barcode on a separate run. Cell-free DNA libraries, instead, given the limited availability, were sequenced in duplicate in two different runs using the same barcode. Namalwa cell line was used as internal positive control in each run.

When the Ion Chef run is complete, proceed immediately with sequencing steps:

1. Initialize the Ion PGM following manufacturer's instructions.
2. Create a Planned Run selecting the 400bp-protocol for FR1 libraries, or the 200bp-protocol for FR3 libraries.
3. Remove chips from the Ion Chef.
4. Load one chip into the Ion PGM, and store the second chip at 4°C. For 400bp-read libraries (FR1), the chip can be stored for up to 24 hours; for 200bp-read libraries (FR3), the chip can be stored for up to 8 hours.
5. Touch CHEF to select the instrument used to prepare the template.
6. Touch Chip Check to perform the first chip check.
7. Follow the on-screen instructions and start the sequencing run.
8. Twenty minutes before the end of the first run, allow the second chip to warm to RT in the dark.
9. When first run is complete, sequence the second chip.

7. Bioinformatic workflow

Using the Torrent Suite software (version 5.0.2 with default parameters), raw reads were trimmed for low-quality 3' ends and barcodes demultiplexed, thus obtaining the raw sequencing data (FASTQ format) produced by the Ion Torrent PGM (Figure 22). Quality assessment (QA) and reads filtering were performed using open-source tools (FastX-toolkit, Cutadapt) in a custom bash (Unix Shell) script. Specifically, raw sequencing reads were analyzed and filtered for base quality (median Phred score > 30, minimum quality score 20) and sequence length (FR1: > 255 bp, FR3: >70 bp) in order to get rid of incomplete reads or aspecific amplification products. Sequences were

retained only if both forward and reverse primers could be identified, and then primer sequences were trimmed from the reads (Figure 22).

Filtered reads were converted to FASTA format and aligned against IMGT germline database (IMGT, Montpellier, France) using IMGT/HighV-QUEST online tool (http://imgt.org/HighV-QUEST/login.action), freely available upon registration, with default parameters (Figure 22).[136]

Aligned reads were parsed using the open-source VDJtools software[137] version 1.1.1. In detail, mapped reads were aggregated into clonotypes based on their CDR3 nucleotide sequence and the same IGH V(D)J gene segment usage. Frequency-based correction was performed with default parameters (maximum number of mismatches allowed between clonotypes being compared = 2; child-to-parent clonotype size ratio threshold under which child clonotype is considered erroneous = 0.05) to eliminate erroneous clonotypes. Corrected samples were stored as a clonotype abundance tables for the subsequent analyses (Figures 22-23).

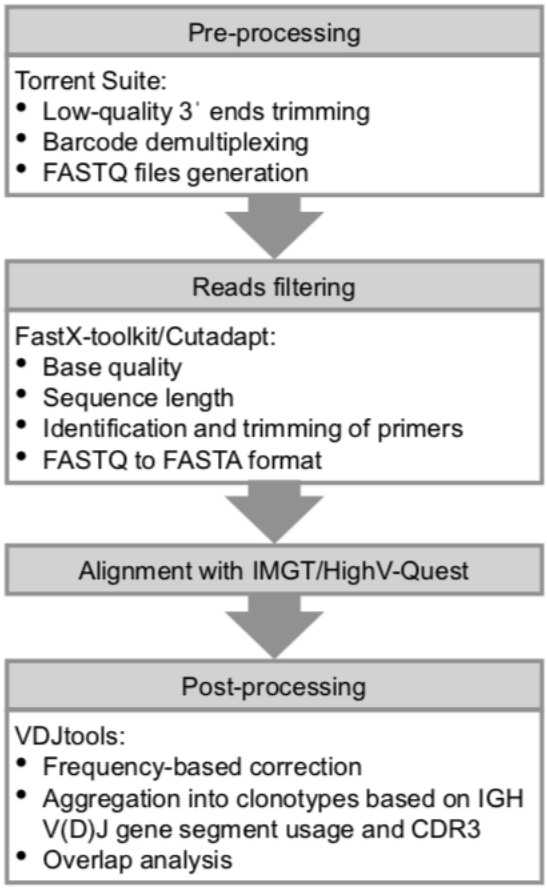


Figure 22. Outline of the bioinformatic workflow we developed for the analysis of Ion Torrent PGM sequencing data.

Post-processing results were also obtained using open-source MiXCR tool[138] version 2.1.12 in a custom bash script. FASTA files containing filtered reads were aligned against IMGT germline database using downloaded IMGT external library; then clones were assembled based on their CDR3 nucleotide sequence and specified clustering parameters to reproduce VDJtools conditions: maximum 2 mismatches between parent and direct child clones (by default); probability of a single nucleotide mutation in clonal sequence which is PCR or sequencing error (corresponding to child-to-parent ratio threshold) = $5e^{-2}$ instead of $1e^{-3}$ default value. IGH clones were eventually exported into text files.

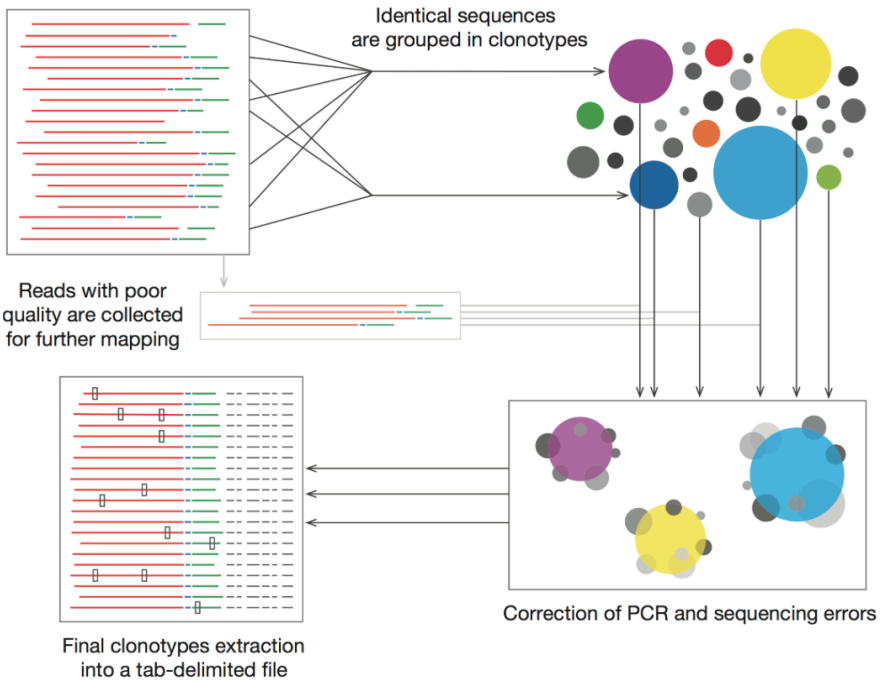


Figure 23. Pipeline for post-processing analysis of BCR repertoire sequencing data.

Bash scripts for NGS reads processing are reported in Appendix A.

8. Sanger sequencing

The current gold standard for NGS data validation is the Sanger sequencing, the “first-generation” DNA sequencing platform also known as “chain-termination” technique. Sanger sequencing is based on the use of dideoxynucleotides (ddNTPs), chemical analogues of the DNA deoxynucleotides (dNTPs) lacking the 3' hydroxyl group that is required to form a bond with the 5'

phosphate of the next dNTP, and therefore for the DNA synthesis. Mixing fluorescent-labeled ddNTPs into a DNA extension reaction, at a fraction (approximately 100-fold lower) of the concentration of standard dNTPs, results in the production of DNA strands of each possible length, as the ddNTPs randomly incorporated into the growing strand halt the further progression. The resulting DNA fragments are heat denatured and separated by size using gel electrophoresis. A laser within the automated sequencer detects the fluorescent signal (ddNTPs are labeled with fluorescent dyes emitting light at different wavelength) and translates it into chromatogram peaks representing the DNA nucleotide sequence (Figure 24).[103]

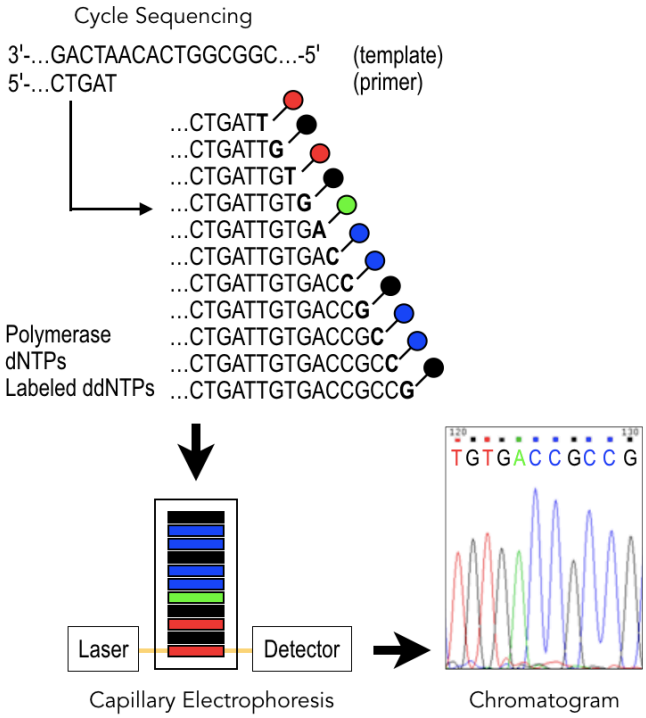


Figure 24. Sanger sequencing method. In each sequencing reaction cycle, amplified products are terminated by the incorporation of fluorescently-labeled dideoxynucleotides (ddNTPs, represented by red, black, green and blue circles), which generates a ladder of differently-size products. These products are subjected to capillary electrophoretic separation, and the four-color detection system translates it to sequence chromatogram.

To validate the consistency of the Ion Torrent NGS data, we carried out traditional Sanger sequencing of PCR products obtained using the same consensus sense primers complementary to FR1 or FR3 and the antisense JH primer.[139] 5 µL of each PCR product, corresponding to a PCR band in a 2% agarose gel, were purified using 5 µL ExoSAP-IT PCR Product Cleanup Reagent (Thermo Fisher Scientific), and sent with the diluted VH family-specific primer (2.5 µL primer + 2.5

µL Nuclease-free Water) to GATC Biotech center (Köln, Germany) to carry out Sanger sequencing. Sequences were then visualized with Chromas Lite version 2.1.1 (Technelysium, South Brisbane, Australia), and the single obtained sequence was analyzed using IMGT/V-QUEST online tool (http://imgt.org/IMGT_vquest/share/textes), with default parameters.[140]

9. Droplet Digital PCR

Droplet digital PCR experiments were performed on the QX200 Droplet Digital PCR System (BioRad, Hercules, CA), based on water-oil emulsion droplet technology. The workflow consists of three steps: PCR sample is partitioned into 20.000 water-in-oil uniform droplets with target and background DNA randomly distributed (Figure 25a); each droplet is then subjected to TaqMan-based PCR amplification determining an enrichment in target DNA fluorescent signal (Figure 25b); after amplification (end-point measurement), droplets stream single file through the reader: positive droplets, exhibiting increased fluorescence compared to negative droplets (Figure 25c), are read as "1", and the target DNA concentration is calculated according to the Poisson distribution.

The sample partitioning, together with the end-point detection, allows the absolute and highly sensitive quantification of low-abundant target DNA.

For my PhD project, droplet digital PCR reactions were conducted using the ddPCR Mutation Detection Assays kit (BioRad), following a specific protocol for IGH rearrangement analysis. Reaction mixtures (21 µL) were prepared using:

- 11 µL ddPCR Supermix (no dUTP) 2X (BioRad);
- 1 µL target forward primer 18 µM: VH family-specific primer complementary to FR1 (Figure 20);
- 1 µL target antisense probe 5 µM: VH family-specific probe complementary to FR2 or FR3 (VH1P 5'-CTGCTCAGCTCCATGTAGGCTGTGC-3'; VH3P 5'-CCCTTCCCTGGAGCCTGGCGGA-3'; VH4P 5'-ACCCGTCCCTCAAGAAGTCGAGT-3'), labeled at the 5' end with reporter dye 6-carboxy fluorescein (FAM) and at the 3' end with quencher activity 6-carboxy-tetramethyl rhodamine (TAMRA);[141,142]
- 1 µL target reverse primer 18 µM: CDR3-specific primer designed using Primer 3 Input online tool (http://www.primer3.ut.ee);

- 1 µL albumin (HEX) primers/probe (BioRad) as reference gene in setting assays only;
- 6-7 µL gDNA (250 ng for Namalwa cell line and for leukocytes extracted from 5 healthy donors as negative controls).

Samples, tested at least in duplicates (Namalwa cells and healthy donors in 6-fold), were then loaded into the QX200 Droplet Generator (BioRad) together with 70 µL droplet oil. Generated droplets (approximately 40 µL) were transferred into a 96-well PCR plate, heat sealed (175°C) with foil, and amplified in a conventional calibrated thermal cycler (ProFlex). Thermal cycling conditions consisted of: enzyme activation at 95°C for 10 minutes, 40 cycles of a two steps thermal profile (ramping rate 2°C/sec) of denaturation at 94°C for 30 seconds and annealing/extension at 60°C for 1 minute, and a final enzyme inactivation step at 98°C for 10 minutes. The plate containing the droplet amplicons was subsequently loaded into the QX200 Droplet Reader (BioRad) and the QuantaSoft software (BioRad) was used to begin the reading process ("ABS" program) and to convert the raw data into a number (copies/µL).

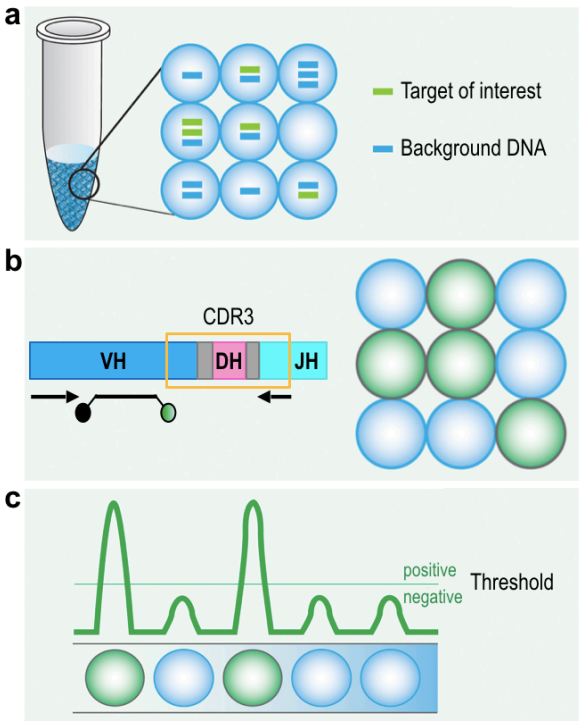


Figure 25. Droplet Digital PCR workflow to detect patient-specific IGH rearrangements. (a) Partitioning: sample is fractionated into 20.000 water-in-oil droplets. (b) Droplet amplification: TaqMan-based PCR using VH family-specific forward primer, CDR3-specific reverse primer, and VH family-specific antisense labeled probe (FAM reporter at the 5' end). (c) Detection: end-point fluorescence quantification of positive *vs* negative droplets.

A sample was considered positive but not quantifiable if no reproducible numbers of droplets were observed (at least one) in the replicates and if the number of positive droplets was ≤ 3. A sample was considered negative if no positive droplets were observed or if positive droplets were below the background. A sample was considered positive and quantifiable if the previous conditions were achieved. Replicates' data were added up.

10. Statistical analysis

Repertoire overlap was analyzed with VDJtools and MiXCR. Fisher's exact test, Mann-Whitney test, Kaplan-Meier survival analysis, Log-rank test, Pearson's correlation test (r coefficient) and One-way ANOVA with post-hoc Tukey HSD test were performed in R version 3.3.2 (R Development Core Team, http://www.r-project.org)[143] and in GraphPad Prism version 5.0 (GraphPad Software, San Diego, California). Figure plots were realized using R and Microsoft Office Excel 2007 (Microsoft Corp., Redmond, WA). For both statistics and graphic visualization we used the following packages: *data.table*, *plyr*, *dplyr*, *reshape2*, *Biostrings*, *ape, survival*, *stats*, *ggplot2*, *grid*, *scales*, *RColorBrewer* (loaded from CRAN repository or installed from Bioconductor, http://www.bioconductor.org). Venny (http://www.bioinfogp.cnb.csic.es/tools/venny) was used to generate Venn diagrams. P-values (P) <0.05 were considered statistically significant (standard progression: *P<0.05, **P<0.01, and ***P<0.001).

RESULTS AIM I

1. Feasibility studies

1.1. Sensitivity estimation of developed IGH sequencing workflow

To assess the sensitivity of our sequencing approach, we studied samples generated by serial 10-fold dilutions of Namalwa and JVM-2 gDNA, bearing a known clonal IGH gene rearrangement (Namalwa: IGHV4-4/IGHD3-3/IGHJ4 5'-TGTGCGAGAAGGAATTACGATTTTTGGAGTGGTGGCGACGGGCCATTTGACTACTGG-3'; JVM-2: IGHV3-9/IGHD6-19/IGHJ4 5'-TATGCAAAAGATATGCAGGGAGCAGTGGCAGGGGGCGTTGACTACTGG -3'), into healthy controls' polyclonal gDNA. We could identify and quantify the clonal sequences at dilutions as low as 10^{-5}, corresponding to a sensitivity of at least 0.001% (2 Namalwa specific sequencing reads over a total of 60624 filtered reads, 0.003299%; 3 JVM-2 specific sequencing reads over a total of 71387 filtered reads, 0.0042%; Table 5).

Table 5. Sensitivity of developed Ion Torrent PGM-based sequencing workflow in detecting clonal IGH gene rearrangements. Titration of Namalwa and JVM-2 cell line genomic DNA (starting from 500ng) into healthy donors' genomic DNA. Experiments were performed in duplicate.

Dilution	Frequency of Namalwa clonotype	Frequency of JVM-2 clonotype
10^{-1}	0.091531	0.0858
10^{-2}	0.015972	0.009666
10^{-3}	0.004907	0.001121
10^{-4}	0.0005	0.000322
10^{-5}	0.000033	0.000042
10^{-6}	0	0

1.2. Cell-free DNA quantity and quality: experimental setup

1.2.1. Peripheral blood sampling conditions

In order to verify whether the plasma samples stored within the clinical trial for MM patients were

suitable for cfDNA extraction and amplification, we compared the cfDNA yield when using plasma obtained from K2-EDTA tubes or from Cell-Free DNA BCT tubes, containing a preservative agent that has been demonstrated to minimize cell lysis and cfDNA degradation for up to 14 days at RT.[144] Peripheral blood samples from 10 healthy donors were collected into both tubes and processed for plasma separation within 3, 24, 48, 72 hours and 1 week; plasma samples were then stored at -80°C for at least one week. Of note, when PB samples are processed within three hours (compatible with what was done for the samples stored within the MM clinical trial) both Cell-Free DNA BCT tubes and K2-EDTA tubes showed similar performances with only slightly higher concentrations of cfDNA from 1 mL plasma samples collected in Cell-Free DNA BCT tubes (median: 8.2 ng *vs* 7.5 ng with K2-EDTA tubes; P=0.3457, Mann-Whitney test; Figure 26a). On the other hand, we could clearly confirm that cfDNA stability is guaranteed only with Cell-Free DNA BCT tube when samples are processed after 3 hours (median: 7.1 ng after 24 hours, 6.8 ng after 48 hours, 4.8 ng after 72 hours, 3.6 ng after one week), making them the best option for multi-center clinical studies where samples are often centralized.

1.2.2. Plasma samples' input volumes

We tested different plasma volumes for cfDNA extraction in a 0.5-3 mL range. The median cfDNA quantity extracted from healthy donors' PB samples was 4.6 ng from 0.5 mL of plasma, 7.5 ng from 1 mL, 13.0 ng from 1.5 mL and 20.4 ng from 3 mL. These results indicate that as expected, higher plasma volumes yield increased cfDNA quantities, even though we did not observe a linear correlation probably due to a saturation effect of the silica membrane. Nonetheless, from quantities as low as 0.5 mL, it is still possible to obtain sufficient cfDNA for NGS applications (Figure 26b).

1.2.3. Plasma samples' storage conditions

We also wondered whether plasma storage conditions might impact cfDNA recovery. No differences (P=0.7635, One-way ANOVA with post-hoc Tukey HSD test) were observed between cfDNA extracted from fresh plasma samples (within 4-6 hours after processing) and cfDNA extracted from frozen samples (median cfDNA quantity from 1 mL of plasma: 7.8 ng after immediate extraction, 7.5 ng after short-term storage of less than 6 months, 7.4 ng after > 6 months; Figure 26c).

1.2.4. Quality assessment of extracted cfDNA

Fragments' size distribution of cfDNA extracted from EDTA plasma samples was evaluated by capillary electrophoresis (Agilent 2100 Bioanalyzer). Samples showed an enrichment in short-fragment DNA (about 150-180 bp), along with smaller peaks corresponding to fragments of larger size (in excess of 300 bp). Moreover, no peaks >800 bp were detected, suggesting the absence of gDNA contamination (Figure 26d).

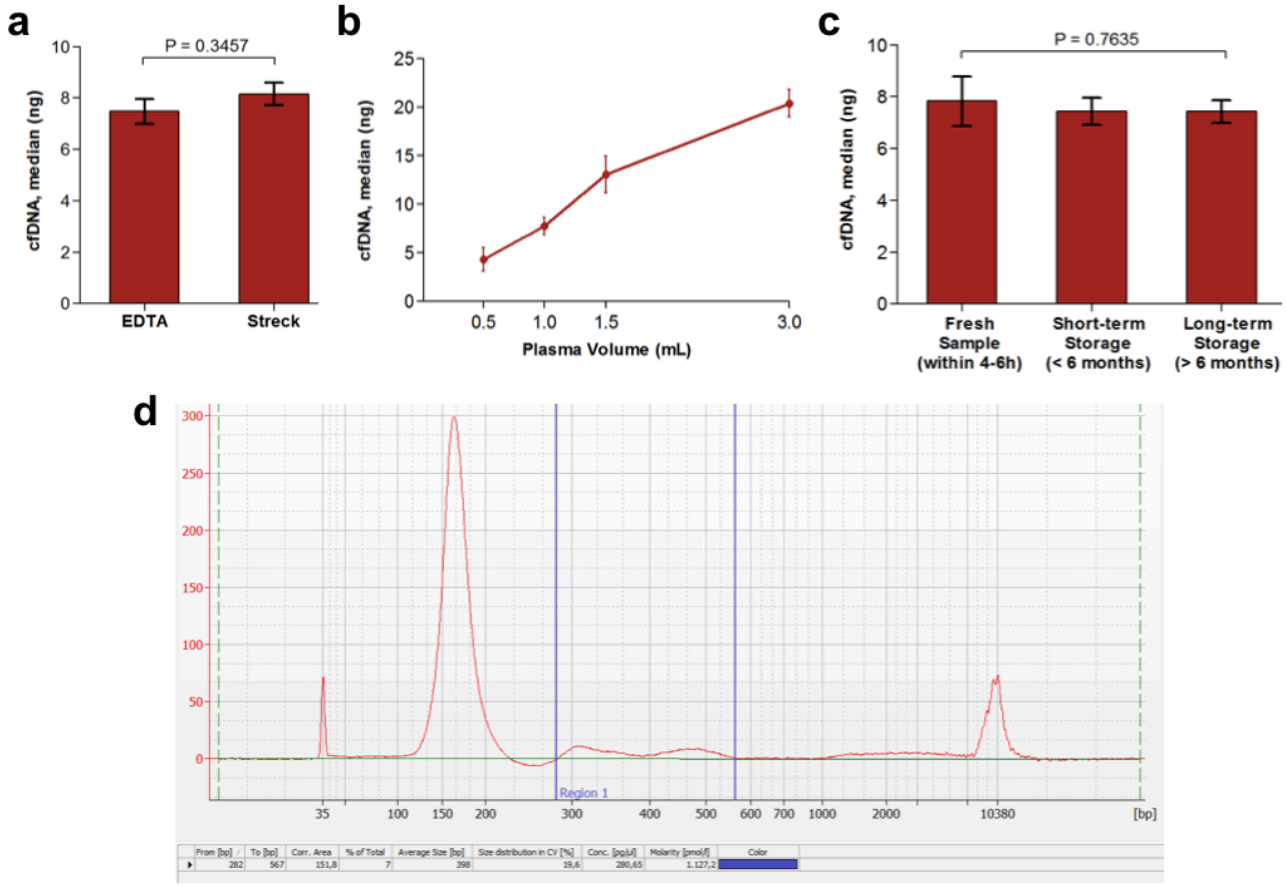


Figure 26. Pre-analytical factors for cfDNA analysis. For a-c panels, on y-axis are plotted the median cfDNA quantities (ng), while on the x-axis are plotted: (a) the comparison between K2-EDTA and Cell-free DNA BCT tubes for PB drawing (Mann-Whitney test); (b) different plasma volumes; (c) different plasma storage (One-way ANOVA test). For panel (d), the electropherogram reports fluorescence intensity (in fluorescence units) on y-axis, fragments' size (in base pair, bp) for quality assessment of extracted cfDNA on x-axis; the left peak represents the lower marker, the right peak the upper marker.

Therefore, considering the minimum quantity of PCR products required for library preparation [Materials and Methods, Paragraph 6.1], in the present study the available stored volumes of plasma (1-3 mL) collected into K2-EDTA tubes, were considered appropriate for cfDNA extraction and downstream IGH amplification with both FR1 and FR3 primer sets.

2. Multiple myeloma

2.1. Patient characteristics

Clinical features of the 25 patients at their first relapse are summarized in Table 6. Previous treatments included mainly bortezomib (9 patients, 36%) and high-dose melphalan (16 patients, 64%). At accrual time point, patients had variable levels of $CD138^+$ PCs in the BM (median: 8.8%, range: 0.05-80%) but no circulating $CD138^+$ PCs in the PB.

Table 6. Baseline characteristics of MM study cohort.

Parameters	
Total patients, *n* (%)	25 (100%)
Median Age, years (range)	65 (41-75)
Sex, *n* (%)	
Male	16 (64%)
Female	9 (36%)
MM subtype, *n* (%)	
IgG	13 (52%)
IgA	8 (32%)
Light chain	4 (16%)
Laboratory findings	
Median BM infiltration on trephine biopsy, % (range)	40 (10-95)
Median BM infiltration by flow cytometry, % (range)	8.8 (0.05-80)
Median BM infiltration by NGS, % (range)	50.1 (1.4-86.7)
Median serum M-protein, g/dL (range)	2.60 (0.09-9.79)
Median serum FLC ratio (range)	33 (0.01-344)
Median serum LDH, U/L (range)	294 (228-433)
Cytogenetic abnormalities, *n* (%)	
t4;14	1 (4%)
t14;16	2 (8%)
17p-	2 (8%)
1q+	4 (16%)
More than 1	6 (24%)
None	5 (20%)
NA	5 (20%)
MM staging, *n* (%)	
ISS I	10 (40%)
ISS II	6 (24%)
ISS III	9 (36%)
Median time from diagnosis, months (range)	51 (22-117)
Median follow-up, months (range)	17 (2-82)

MM, multiple myeloma; Ig, immunoglobulin; BM, bone marrow; NGS, next-generation sequencing; FLC, free light chain; LDH, lactate dehydrogenase; ISS, international staging system.

2.2. Identification and characterization of rearranged clonal IGH by Ion Torrent PGM sequencing in MM PCs

To identify the dominant tumor clonotype, we amplified by multiplex PCR rearranged IGH V(D)J loci in genomic tumor cell DNA (tcDNA) derived from isolated BM CD138$^+$ PCs using validated degenerate PCR primers complementary to IGH FR1 and the consensus IGHJ primer.[9,96] The amplification generated products of the expected sizes that were sequenced on the Ion PGM. Sequencing yielded an average of 101455 reads (±79377 SD) that were subjected to QA and filtering (Figure 22). A mean of 47755 (±32787 SD) sequences (47% of total reads ±32% SD) passed quality filters and were used for subsequent downstream analyses.

To assess clonality in our samples, we determined the percentage of reads having the same V(D)J usage and a unique CDR3 nucleotide sequence. The most represented clone had a median frequency of 50.1% (range: 1.4%-86.7%). Setting a frequency threshold > 5% to define a clone as a tumor clonotype[108,123], our sequencing method successfully identified a myeloma clonotype in 22 out of 25 MM patients (88%), indicating a specificity similar to the one reported in previously described NGS assays for clonal IGH identification (91%).[108,145]

In all 22 patients, the V(D)J identity and the CDR3 sequence of the tumor clonotype identified by NGS corresponded to that inferred with traditional Sanger sequencing approach. Moreover, in the remaining 3 patients for whom the NGS approach identified IGH sequences at frequencies under the validated threshold, Sanger sequencing highlighted a polyclonal IGH repertoire (Table 7).

Table 7. Genomic tumor cell DNA (BM CD138$^+$ PCs) sequencing stats.

Sample	% most represented IGH clone	CDR3 sequence	Sanger seq
MM_1	57.5%	5'-TGTGCACACAGCGAGGTCGGTGGTTATTACTACCCTAAACCCCTTGACTTCTGG-3'	Confirmed clonal
MM_2	80.1%	5'-TGTGTACGTAGGGGGCAGCCTTCAATAGTCGCCGATGGAGGCATCTACTTTGACCACTGG-3'	Confirmed clonal
MM_3	76.4%	5'-TGTGCGAGAGTTTCGGGGGGTATAGCAGTGGCTGCGATCGACCGACCCATATTATACTACTACGGTATGGACGTCTGG-3'	Confirmed clonal

MM_4	1.4%	No IGH tumor clonotype (<5%)	Confirmed polyclonal
MM_5	69.4%	5'-TGTGCGAGAGGTGTCGTGAGTGGTAGCTGCTACTTTCGAAATTGGTTCGACCCCTGG-3'	Confirmed clonal
MM_6	8.0%	5'-TGTGTGAAGTCTCTCGGGGCTTCTTGGCACTACGCTATGGACGTCTGG-3'	Confirmed clonal
MM_7	63.1%	5'-TGTGCGAGAGATAAGATAGGAGCAGCAGCTGGTAGTTGGTTCGACCCCTGG-3'	Confirmed clonal
MM_8	42.7%	5'-TGTGCGAGAGATTTAGGGGACGCTATGGACGTCTGG-3'	Confirmed clonal
MM_9	47.6%	5'-TGTGCGAGAGTCACACGAGGGTACTACTTTGACTACTGG-3'	Confirmed clonal
MM_10	26.6%	5'-TGTGCGAGAGGCCTGGGTAACGGAGCAGCTGCCCAGGAAACTCACCTCGTCTGGTTCGACCCCTGG-3'	Confirmed clonal
MM_11	81.4%	5'-TGTGCGAAAGATCATAACGAGTGGGAGCTGAGACGATCCGGGGACTGG-3'	Confirmed clonal
MM_12	85.8%	5'-TGTGCGATGGACCGAACTGCAACGGAGGGGCTCGACCCCTGG-3'	Confirmed clonal
MM_13	86.7%	5'-TGTGCGAGACATTCTGGAACAGTGGCTGGTATCTTTGACAACTGG-3'	Confirmed clonal
MM_14	55.3%	5'-TGTACCACCTGGGCGACCGCAGTGTCTGGGCGACTG-3'	Confirmed clonal
MM_15	50.1%	5'-TGTGCAAAAGACGGGGGGTATAGCAGTGGCTGGGCCCAAGAGGGCTTGGACTACTGG-3'	Confirmed clonal
MM_16	2.7%	No IGH tumor clonotype (<5%)	Confirmed polyclonal
MM_17	21.0%	5'-TGTGCGAAACTGCAGGGGCATTACTATGATAGTAGTGGTTATCCGAACTGG-3'	Confirmed clonal
MM_18	51.4%	5'-TGTGCGAGTTCTATTGTAGTAGTACCACCGGGCGTCTGG-3'	Confirmed clonal
MM_19	9.1%	5'-TGTGCGAGGGATCGTGATGGCAGTGGCTGGTCCTTTGATTACTGG-3'	Confirmed clonal
MM_20	13.0%	5'-TGTGCGAGAGCGGCGTCGGCAGCAGCTGGTACGGAGGGGTTGTTCGACCCCTGG-3'	Confirmed clonal
MM_21	62.3%	5'-TGTGCGACTATAGCAGTGGCTGGTCCCTACTGGTACTTCGATCTCTGG-3'	Confirmed clonal
MM_22	9.8%	5'-TGTGCGGCGGGCCTGGAACCCCGCTACTGG-3'	Confirmed clonal
MM_23	74.9%	5'-TGTGCGACCGCCCCAATACCGGACTCGGGAAGACTTGACAACTGG-3'	Confirmed clonal

MM_24	4.1%	No IGH tumor clonotype (<5%)	Confirmed polyclonal
MM_25	17.8%	5'-TGTGCGAGAACGACCAGCAGCTCCTTTGGCGACTGG-3'	Confirmed clonal

In addition, consistent with previously published data,[146] in all 22 patients the IGH V(D)J gene segment usage of the tumor clonotype was non-random: IGHV3, IGHV4, IGHV1 and IGHV2 were the VH-families expressed at the highest frequencies (Figure 27a), and a preferential expression of VH3-23 (18%, 4/22 patients), VH1-69, VH2-5, VH3-30, VH3-48 and VH3-7 (each at 9%) was observed (Figure 27b). IGHD6-19 and IGHD6-13 predominated (Figure 27c) consistently with what seen in other B-cell malignancies.[147] Among the IGHJ subgroups (Figure 27d), IGHJ4 was statistically over-represented (45% observed *vs* 16.7% expected assuming a random choice for JH; P<0.001, Fisher's exact test), while IGHJ2 and IGHJ3 were significantly under-represented (5%; P<0.001).

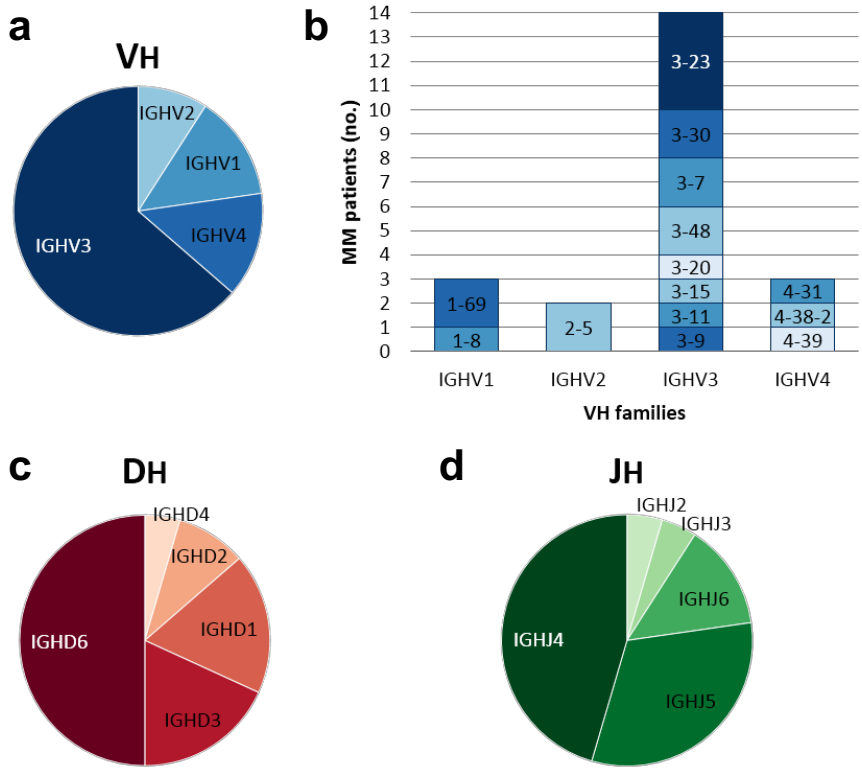


Figure 27. IGH V(D)J repertoire in multiple myeloma patients. The broadness of each section of the pie charts represents the percentage of VH/DH/JH family usage for the respective IGH genes. The IMGT/High-VQUEST online tool (http://imgt.org/HighV-QUEST/login.action) was used to identify the IGH V(D)J families and genes used in the clonal rearrangements. (a) IGHV families. (b) IGHV gene segment usage. (c) IGHD families. (d) IGHJ families.

Consistent results were obtained when amplicons were generated using the FR3 primer set (Figure 20, right table), sequenced on the Ion Torrent PGM and analyzed with our custom bioinformatic workflow using a tailored length filtering. We confirmed the absence of a tumor clonotype in the same 3 patients excluding, overall, a primer annealing bias under the conditions here reported.

2.3. Sequencing of the IGH gene rearrangements in plasma samples: performance and concordance of tcDNA and cfDNA results

Matched genomic tcDNA and plasma cfDNA samples obtained at study entry were available for all 22 patients with a defined myeloma clonotype. Extracted cfDNA was amplified by multiplex PCR using a median of 20.3 ng (range: 8.6-45.9 ng) with the same primers used for tcDNA (the consensus JH primer and the FR1 primer set; Figure 20, left table). IGH cfDNA sequencing yielded an average of 210056 (±161520 SD) reads, with a mean of 77188 (±55058 SD) sequences passing quality filters (37% of total reads *vs* 47% on tcDNA) and available for subsequent analyses. In MM plasma samples, the percentage of predominant clones ranged from 1.2% to 18% (Figure 28a). In comparison, in control plasma samples from healthy donors (n=10), several different clones were present at low percentages and no clonally related sequences were observed (mean: 0.006%, range:0.002-1%; Figure 28b).

To compare the recombinant repertoire in plasma of MM patients and donors, we created Circos plots (VDJtools) showing the relative prominence of each VH-JH recombination within the repertoire of each subject (Figures 28c-d). These plots revealed that V genes are paired with many other J genes in control samples, whereas MM samples show some few but broad connections indicating the predominance of one VH-JH combination (one clonotype) in patients among other clones present at lower frequencies. The same clustering pattern was also obtained analyzing the nucleotide sequence similarity between V(D)J aligned reads of each MM subject and of each healthy donor; patients' heatmaps, generated by *hierarchical clustering* function in R, showed few large-size defined clusters indicating a clearly BCR clonality status compared to the high number of small-size distant clusters in healthy donor plasma samples (Figure 28e-f).

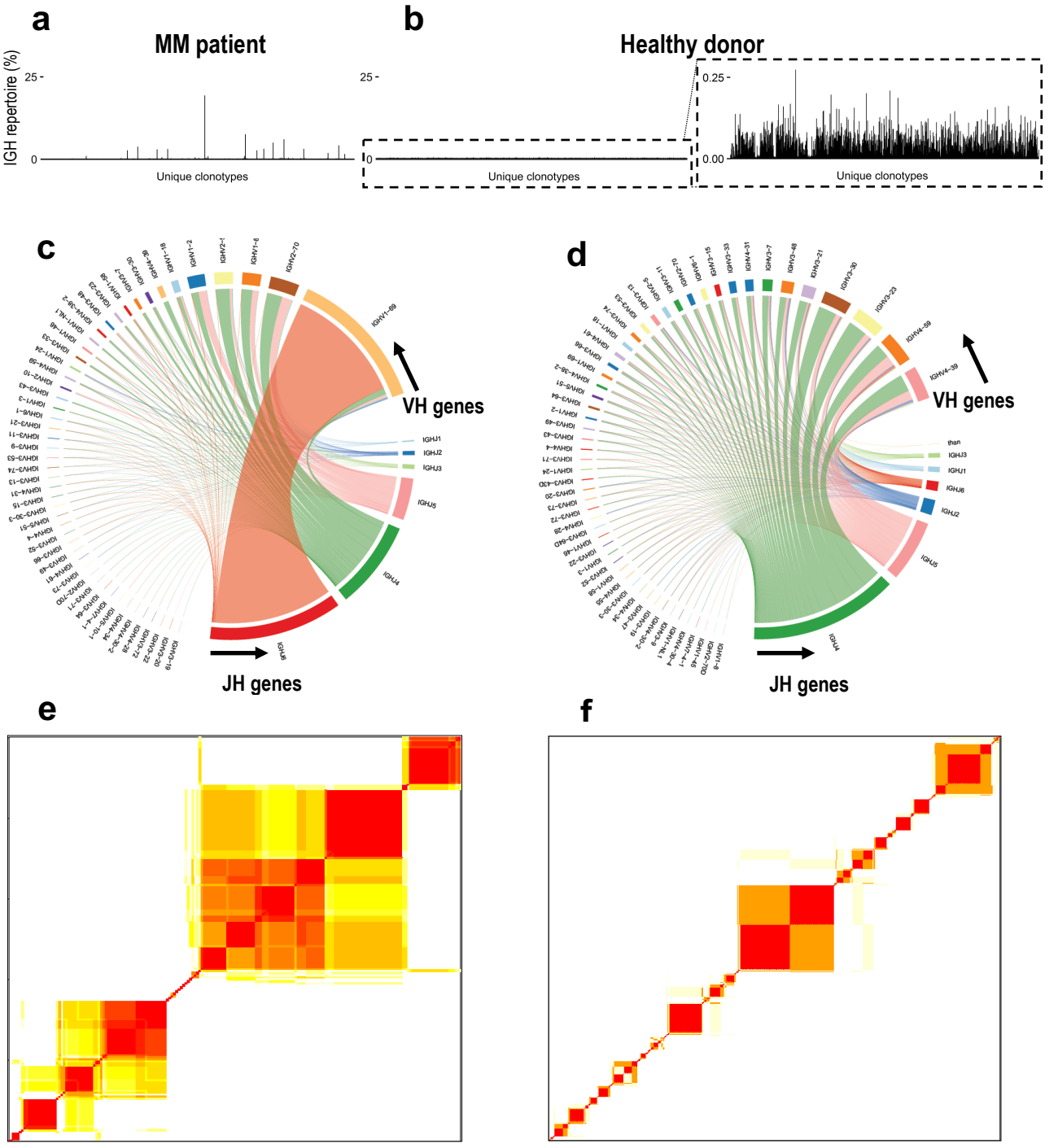


Figure 28. Sequencing of the IGH rearrangements in cfDNA of a representative MM patient and a healthy donor. Histogram of plasma clonotype frequencies detected in (a) a MM patient and (b) a healthy donor with a zoom-in on lower frequencies (dotted box); unique clonotype frequencies are plotted on the y-axis for both the MM subject and the healthy donor (as a percentage of the total reads analyzed). V-J junction circos plot for (c) a MM patient and for (d) a healthy donor; arcs correspond to different VH and JH segments, scaled to their frequency in sample; ribbons represent VH-JH pairings and their broadness corresponds to the number of IGH sequences that exhibit this gene combination. (e-f) Heatmaps of matrices of Hamming distance between complete V(D)J nucleotide sequences in (e) MM and in (f) healthy donor plasma sample; rows and columns are ordered according to the gene clusters found by hierarchical clustering using *dist.dna* and *hclust* functions in *ape* R package. Color spectrum ranges linearly from red to white for nucleotide distances 0-10%. Differences greater than 10% are white.

We then searched for the tumor clonotype identified in cellular NA in plasma samples. In all 22 cases (100%) the rearranged tumor-associated IGH sequence was present in plasma (Figure 29a) at percentages ranging from 1.2% to 18% of total filtered circulating IGH DNA reads (median frequency of 4.7%). As reported above for tcDNA, results were confirmed using FR3 primer set. Additionally, we detected clonotypes (10%±7%, mean±SD) at high frequencies that are shared between paired plasma and tumor samples, whereas other clonotypes are exclusive of each compartment (Figure 29b). This suggests that plasma cfDNA can accurately mirror the profiles of the most abundant clonotypes and, furthermore, that plasma cfDNA may reflect myeloma burden in bone marrow but also in extramedullary sites.

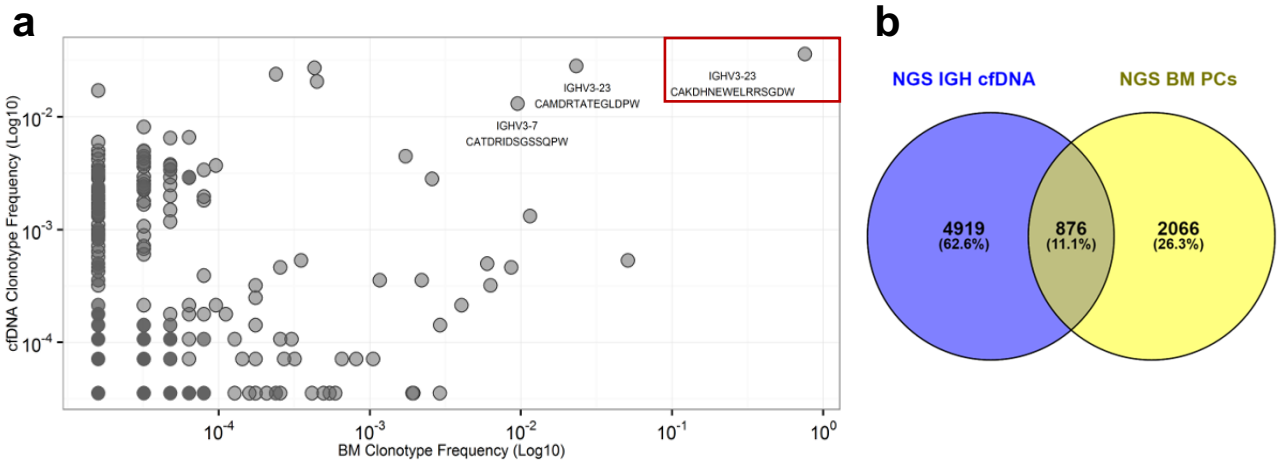


Figure 29. Overlap analysis of the clonotypes detected by NGS in tcDNA and cfDNA. (a) Scatterplot of the clonotype frequencies (log10 scale) observed in a representative MM patient (BM clonotypes on the x-axis and cfDNA on the y-axis). The majority of clonotypes are detected at low frequencies, while the tumor-associated clonotype (red box) can be identified as the one at the highest frequency in both compartments (plasma and BM tumor sample). Next to the top 3 most frequent shared clonotypes, are indicated the VH gene (IGHVx-x) and the related CDR3 amminoacid sequence. (b) Venn diagram showing the number (x% calculated on the total clonotypes) of unique and shared clonotypes between plasma (blue circle) and tumor (yellow circle) compartment.

2.4. Prognostic implication of clonotypic IGH cfDNA levels in myeloma patients

As there are different ways to determine the PCs percentage in the BM [counting cells on trephine bone marrow biopsies (BMB), or on an aspirate smears] and these methods often give discrepant

results,[148] we speculated that the analysis of clonal IGH in cfDNA could be used to help estimating tumor levels, overcoming BM sampling and analysis limitations. PCs on BMB, present in all patients before therapy at a median frequency of 40% of BM leukocytes (range: 10-95%), did not correlate with clonotypic cfDNA frequencies (r=0.098, P=0.6724). Similarly, no significant correlation was found between M-protein levels evaluated by SPEP and clonotypic cfDNA frequencies (r=-0.018, P=0.9504). IGH cfDNA levels and sFLC ratios show a moderate positive correlation but not significant (r=0.4567, P=0.2552) (Figure 30a). Likewise, no correlation was found between BMB PCs and SPEP levels (r=0.289, P=0.295), BMB PCs levels and sFLC ratios (r=0.148, P=0.7793), SPEP levels and sFLC ratios (r=-0.032, P=0.8465), supporting the idea that MM is a complex disease and none of these laboratory parameters alone accurately describes tumor levels in a single patient. Of interest, disease levels at baseline evaluated by NGS on plasma samples compare favorably to BM PCs infiltration by MFC (r=0.713, P=0.0002) and to NGS performed using tcDNA (r=0.45, P=0.0354) (Figure 30b).

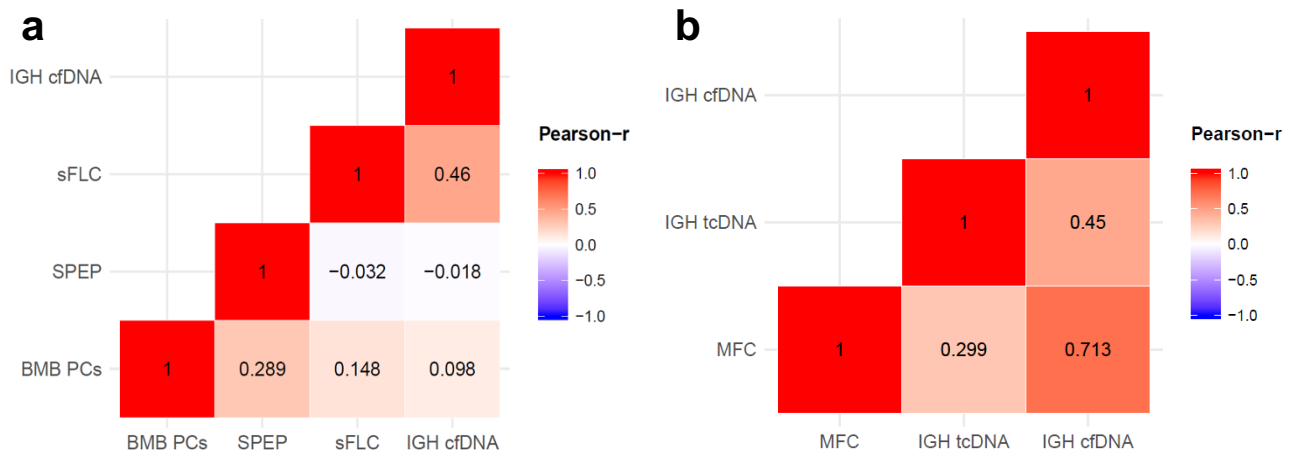


Figure 30. Correlation of IGH cfDNA levels and other laboratory parameters. Heatmaps show the level of correlation calculated by Pearson's correlation test using a colorimetric scale (red = perfect positive correlation, white = no correlation, blue = perfect negative correlation). In the middle of each square is reported the exact value of Pearson's correlation coefficient (r) obtained for each comparison. (a) Correlation of IGH cfDNA levels and IMWG defined clinical indices. BMB PCs, bone marrow plasma cells on trephine biopsy; SPEP, M-protein evaluated by serum electrophoresis; sFLC, serum free light chain ratio; IGH cfDNA, immunoglobulin heavy chain cell-free DNA. (b) Correlation of IGH cfDNA and the 2 high-sensitivity techniques available for disease quantification in MM. MFC, multiparameter flow cytometry; IGH tcDNA, immunoglobulin heavy chain tumor cell DNA.

We next sought to determine whether cfDNA analysis might facilitate the early identification of clinically relevant risk groups among the cohort of myeloma patients enrolled in the study. Levels of the tumor-associated clonotype in cfDNA distinguish between groups of patients with different prognosis with the median percentage of the tumor-associated clonotype in cfDNA as the cut-off value (4.7% of total reads). Patients with levels ≥4.7% (n=12) of the tumor-associated IGH sequence prior to therapy had significantly inferior PFS (median values estimated: 268 *vs* 990 days; HR=3.507, P=0.04988, Log-rank test; Figure 31) than patients with levels <4.7% (n=10).

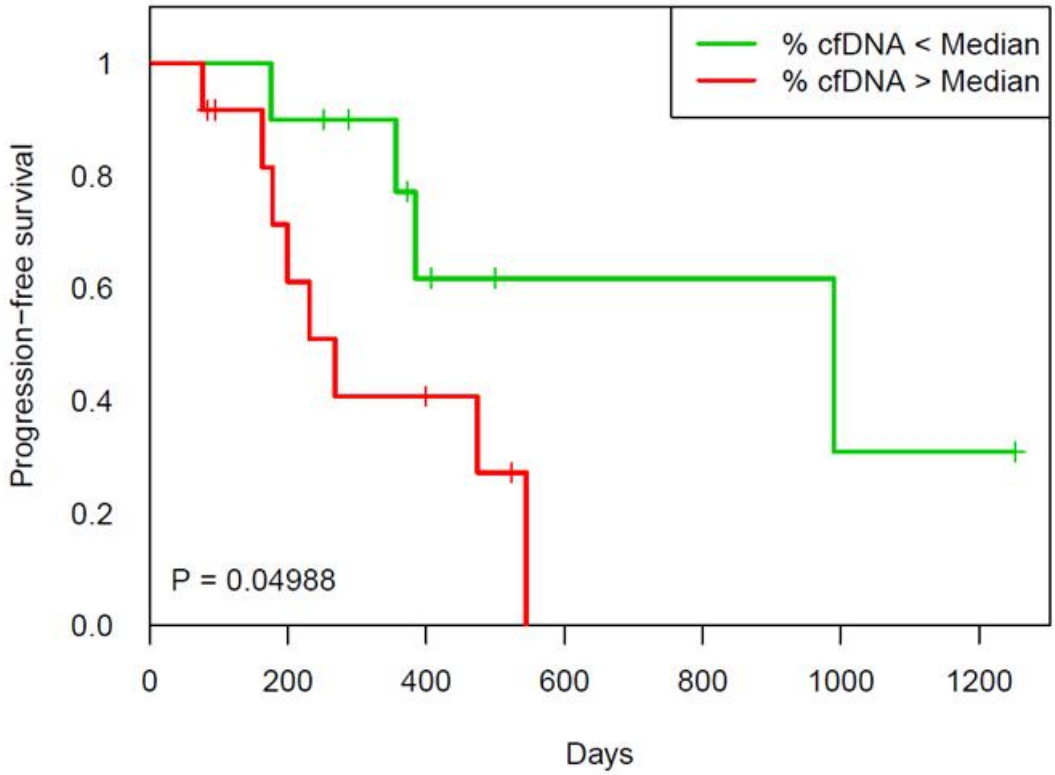


Figure 31. Correlation of IGH cfDNA levels at baseline with progression-free survival. Patients with frequency percentages <4.7% (median value): green line, n=10; patients with frequencies ≥4.7%: red line, n=12. HR=3.507 by *CoxPH model* in R, P=0.04988 by Log-rank test.

When considering baseline features such as BMB PCs, SPEP levels and sFLC ratios, no significant association was found with survival. Similarly, none of the variables known to impact clinical outcome in MM patients such as age, International Staging System (ISS) status or the Revised ISS (R-ISS),[78] which includes serum lactate dehydrogenase and high-risk cytogenetic abnormalities (t4;14, t14;16, 17p- and 1q+),[149] is significantly correlated with PFS or OS (Table 8).

Table 8. Prognostic impact of all variables at study entry time point. Exp(coef) is calculated by Cox proportional hazards model in R; Exp(coef) >1 indicates increased risk of relapse/death, Exp(coef) <1 indicates decreased risk, Exp(coef) =1 indicates no correlation.

Variable	Correlation with PFS		Correlation with OS	
	Exp(coef)	P-value	Exp(coef)	P-value
BMB PCs (over or equal to/under the median)	1.645	0.4164 (ns)	1.067	0.9439 (ns)
SPEP (over or equal to/under the median)	1.466	0.5928 (ns)	1.183	0.9055 (ns)
sFLC (over or equal to/under the median)	1.044	0.9533 (ns)	1.175	0.7901 (ns)
Age (over or equal to/under the median)	0.8737	0.8301 (ns)	0.2844	0.2324 (ns)
ISS (first stage or worse)	0.7151	0.5698 (ns)	2.58	0.3803 (ns)
LDH (over or equal to/under the median)	0.3558	0.3834 (ns)	2.067	0.4795 (ns)
Cytogenetic abnormalities (presence/absence of at least 1 abnorm.)	0.7589	0.7365 (ns)	0.4452	0.0499 (*)
IGH cfDNA (over/under the median)	**3.507**	**0.04988 (*)**	1.619	0.5964 (ns)

On the other hand, as expected given the high concordance of cfDNA NGS results and MFC (Figure 30b), higher numbers of PCs enumerated by MFC (median percentage as cut-off value, 8.8%) are associated with poorer PFS (median values estimated: 268 *vs* 545 days; HR=3.745, P=0.04655, Log-rank test; Figure 32a), whereas no significant association was found between NGS results on tcDNA and survival (Figure 32b).

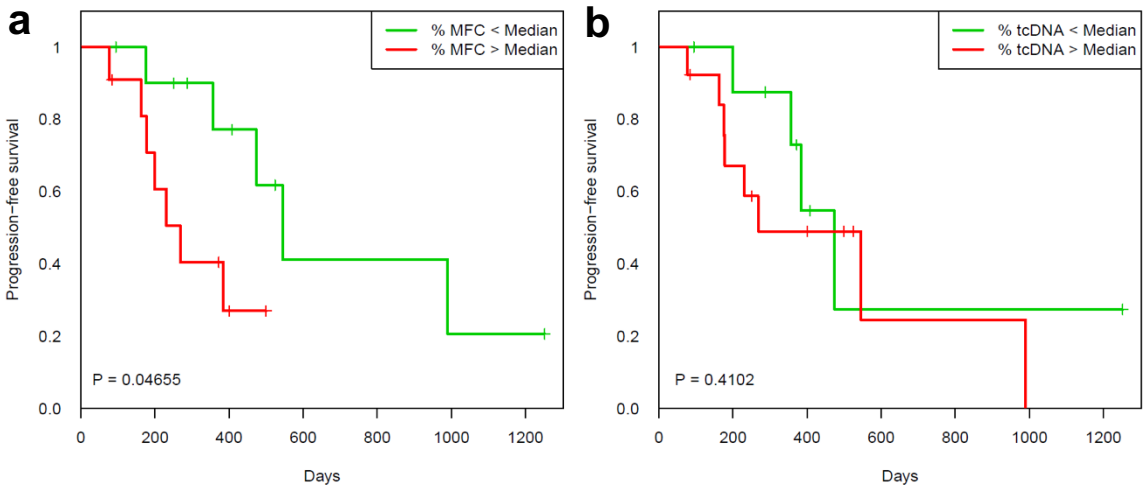


Figure 32. Correlation of the 2 high-sensitivity techniques available for disease quantification in MM with progression-free survival at baseline time point. (a) Correlation of MFC with PFS. Patients with percentages of BM tumor PCs detected by MFC <8.8% (median value): green line; patients with percentage ≥8.8%: red line. HR=3.745 by *CoxPH model* in R, P=0.04655 by Log-rank test. (b) Correlation of NGS on tumor cell DNA (tcDNA) with PFS. Patients with percentages of IGH tcDNA detected by NGS <50.1% (median value): green line; patients with percentage ≥50.1%: red line. HR=1.654 by *CoxPH model* in R, P=0.4102 by Log-rank test.

2.5. Disease monitoring by clonotypic IGH cfDNA sequencing

We next hypothesized that longitudinal analysis of clonal IGH cfDNA could help understanding tumor dynamics over time. We analyzed 70 plasma samples of the selected 22 MM patients. IGH cfDNA sequencing revealed that the tumor-associated clonotype could be tracked over time in plasma samples. The levels of the clonotypic IGH sequences in plasma (Figure 33a, b and c; last row) reflected tumor dynamics evaluated using the IMWG criteria (BM PCs, SPEP levels and sFLC ratio) (Figure 33a, b and c; first, second and third row).

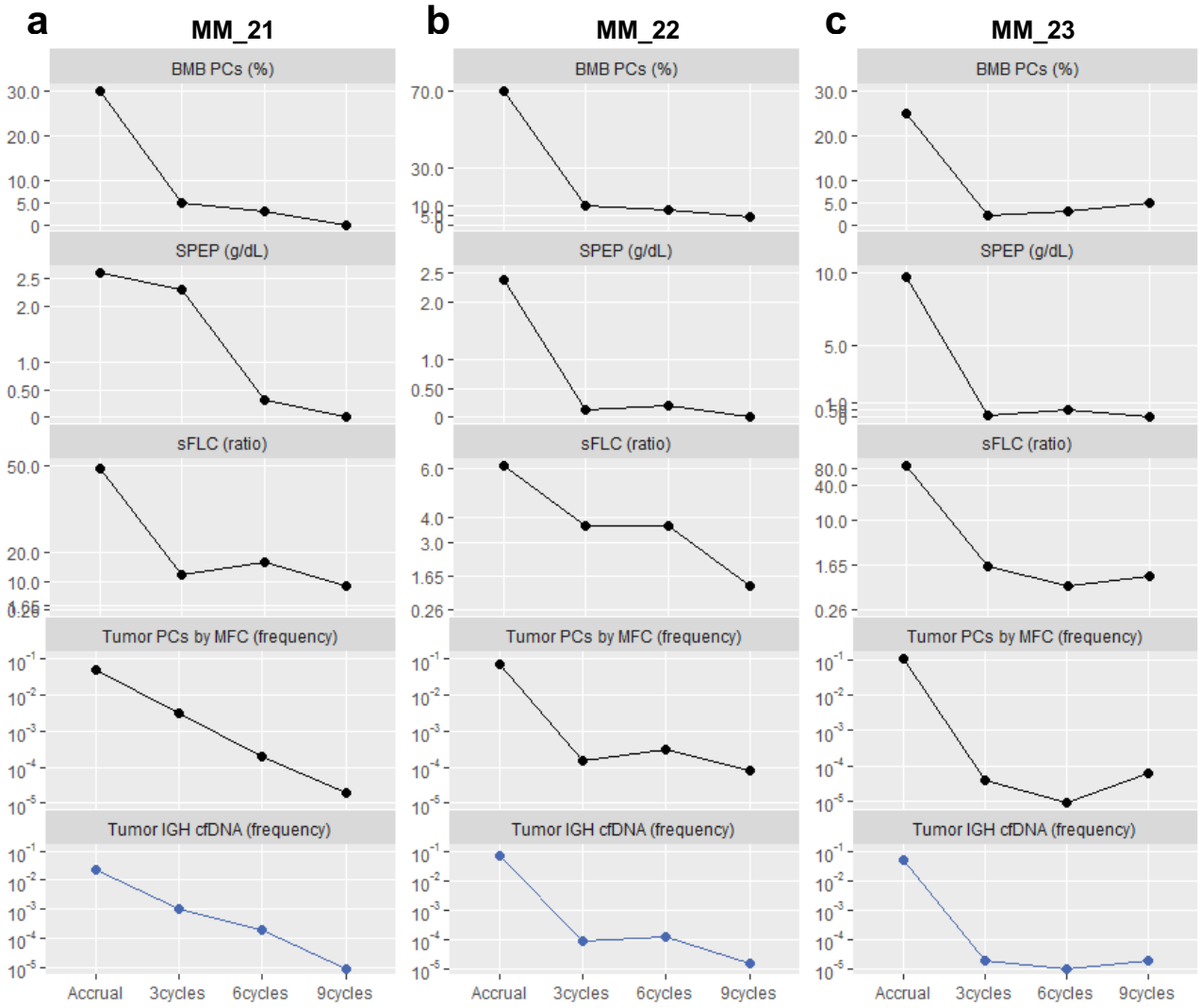


Figure 33. Quantification of IGH cfDNA levels in relation to MM clinical indices. Following a top-down disposition, y-axes of line plots are respectively related to: percentage of bone marrow PCs on trephine biopsies (BMB PCs, %, linear scale); M-protein concentrations evaluated by serum electrophoresis (SPEP, g/dL, linear scale); serum free light chain (sFLC ratio, linear scale for panels A and B/log2 scale for panel C, normal range: 0.26-1.65); bone marrow PCs enumerated by multiparameter flow cytometry (tumor PCs by MFC, frequency, log10 scale); plasma IGH cfDNA levels (tumor IGH cfDNA, frequency, log10 scale). Time points are labeled on x-axis. (a) Case MM_21; (b) Case MM_22; (c) Case MM_23.

Additionally, the phenotypic aberrancies detected in PCs at study entry for each patient (based on: CD38/CD56/CD19/CD45, CD38/CD27/CD45/CD28, CD20/CD81/CD38/CD117) were used as patient-specific markers for residual disease assessment after therapy on consecutive samples. Cell-free DNA levels reflected the number of PCs enumerated by MFC immunophenotyping (Figure 33a, b and c; forth row).

Three representative cases of the correlation between IGH cfDNA and standard clinical indices in myeloma monitoring are reported below (Figure 33).

In the representative case MM_21 (Figure 33a), plasma IGH cfDNA levels decreased over the course of treatment, paralleling the decline of BMB PCs and SPEP levels which became undetectable after 9 cycles of treatment. This is consistent with MFC trend and, on the other hand, contrasted with sFLC ratio that is stable during the course of treatment. In patient MM_22 (Figure 33b), plasma IGH cfDNA levels decreased rapidly after the initial cycles of therapy, stabilized from cycle 3 to 6 and then declined again after the 6th cycle. These dynamic changes were reflected by SPEP, sFLC and MFC. BMB PCs levels instead, steadly decrease during the course of treatment.

In case MM_23 (Figure 33c), levels of IGH cfDNA decline after 3 cycles of therapy, become stable and present a slight increase after 9 cycles, mirroring specifically the trend of sFLC and MFC, and also of BMB PCs. In this case SPEP levels contrasted as they became undetectable at the end of therapy.

In the context of CR time points (n=22), MFC was used for MRD monitoring. At least $1x10^6$ cellular events were collected and residual cells detected at a median frequency of 0.00065 (range: 0.00000668-0.027) tumor cells out of the total events analyzed (range: 1014888-1946108). MRD-negativity was defined when <50 aberrant-phenotype PCs were detected,[93] thus 6 patients achieved MRD-negativity during the timeframe of our analysis. When analyzing cfDNA IGH frequencies in these 22 patients at CR time points (median: 0.0000395, range: 0.00000756-0.037), a high level of correlation was found between cfDNA NGS and MFC data (r=0.5831, P=0.0044, Pearson's correlation test; Figure 34a). Accordingly, PFS was significantly prolonged (***P<0.001)

for the 6 patients who achieved MRD-negativity displaying frequencies of the clonotypic cfDNA rearrangement lower than 10^{-5} (PFS 714±327 days, mean±SD, for clonotypic cfDNA frequencies $<10^{-5}$, n=6; 325±75 days for frequencies $\geq 10^{-5}$-$\leq 10^{-4}$, n=9; 143±59 days for frequencies $>10^{-4}$, n=7; Figure 34b). In these cases, to increase theoretical sensitivity, NGS of cfDNA samples was performed with a lower level of plexing (maximum 8 samples on a Ion 318 Chip v2 BC) to obtain at least $5x10^5$ reads per sample.

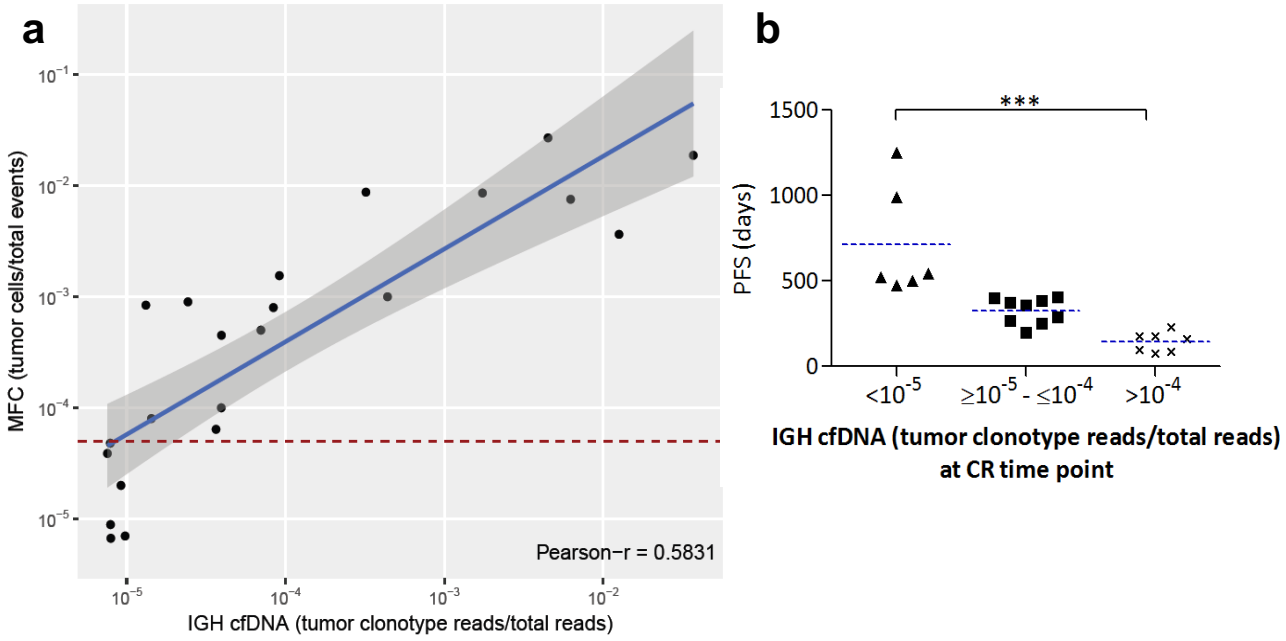


Figure 34. Comparison between IGH cfDNA sequencing and MFC data. (a) Scatter plot showing correlation between IGH cfDNA sequencing (IGH cfDNA, log10 x-axis) and 8-color flow cytometry (MFC, log10 y-axis) in assessing residual disease in the selected 22 patients (black dots). Blue line: best fit line obtained using the linear model method in R; shaded area: 95% CI of the best fit line. Pearson's correlation test, applied in R with *linear model* method and *cor.test* function, was used to quantify the strength of the relationship (Pearson-r = 0.5831; P=0.0044). Red dashed line set the sensitivity threshold for MFC (frequency $<5x10^{-5}$). (b) Group analysis shows that levels of disease detected by IGH cfDNA sequencing (categorized as frequencies $<10^{-5}$, black triangles, n=6; frequencies $\geq 10^{-5}$-$\leq 10^{-4}$, black squares, n=9; frequencies $>10^{-4}$, black crosses, n=7; x-axis) are significantly correlated with PFS (***P<0.001, One-way ANOVA with post-hoc Tukey HSD test). Dashed blue lines: mean value of each category.

RESULTS AIM II

1. Diffuse large B-cell lymphoma

1.1. Patient characteristics and imaging parameters

Twenty-six newly diagnosed DLBCL patients were prospectively enrolled and followed during R-CHOP or R-CHOP-like treatment. The majority (65%) of patients presented with Ann Arbor stage III or IV but encompassed variable IPI scores as detailed in Table 9.

Table 9. Baseline characteristics of DLBCL study cohort.

Parameters	
Total patients, *n* (%)	26 (100%)
Median Age, years (range)	61 (30-84)
Sex, *n* (%) Male Female	 15 (58%) 11 (42%)
COO classification, *n* (%) GCB Non-GCB	 16 (62%) 10 (38%)
Molecular prognostic factors MYC+ expression, *n* (%) BCL2+ expression, *n* (%) BCL6+ expression, *n* (%)	 13 (50%) 16 (62%) 15 (58%)
SUVmax, median (range)	20.6 (6.2-42.4)
Elevated serum LDH, *n* (%)	9 (35%)
ECOG performance status ≥ 2, *n* (%)	3 (12%)
Extranodal sites ≥ 2, *n* (%)	5 (19%)
Ann Arbor staging, *n* (%) Stage I-II Stage III-IV	 9 (35%) 17 (65%)
IPI score, *n* (%) 0-1 / 2-3 / 4-5	 11 (42%) / 12 (46%) / 3 (12%)
Treatment Response, *n* (%) CR / PR / PD	 16 (62%) / 3 (12%) / 7 (27%)

COO, cell-of-origin; GCB, germinal center B-cell-like DLBCL; SUVmax, highest standardized uptake volume; LDH, lactate dehydrogenase; ECOG, Eastern Cooperative Oncology Group; IPI, International Prognostic Index; CR/PR, complete and partial response; PD, progressive disease.

Recruited patients did not harbor morphological/immunophenotypic evidence of PB involvement of $CD20^+$ DLBCL cells.

Following the Lugano guidelines for staging and response assessment in lymphomas,[39] all DLBCL patients underwent whole-body CT and/or FDG-PET scan at baseline, at interim during treatment, and after the last cycle of chemotherapy for response assessment. Imaging parameters considered for clinical correlation with longitudinal plasma samples were: i) variation in the volume of tumor lesions detected by CT; ii) for PET-avid histologies, the highest standardized uptake volume (SUVmax) measured in the tumor sites.

1.2. Quality assessment of gDNA extracted from FFPE samples and FR1 primer amplification

The process of fixation and embedding, as well as long-term storage, can result in severe degradation of nucleic acids in tissue biopsies. The fixation process causes cross-links between nucleic acids and proteins, as well as DNA bases' modifications that lead to NGS artifacts.[150] FFPE gDNA samples were therefore initially tested for quality. As reported in 4 representative patients in Figure 35, we observed a different fragmentation pattern, probably due to the type of tissue biopsy (surgical excision biopsy, or needle-core and endoscopic biopsies) and the time of storage.

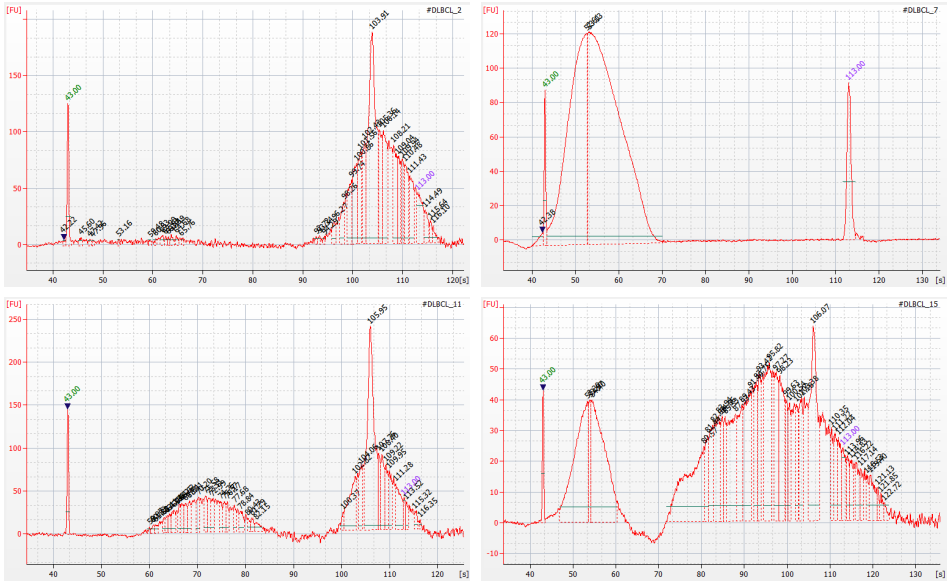


Figure 35. Quality assessment of FFPE gDNA. Electropherograms produced by Agilent 2100 Bioanalyzer representing fragments' size of 4 representative DLBCL samples; x-axes: migration time (seconds, s) or corresponding fragments' size (bp), y-axes: fluorescence units (FU).

For 5 gDNA samples characterized by high quality (fragments' size > 200 bp) and high quantity (concentration > 20 ng/µl) we decided to proceed with subsequent multiplex PCR amplification using FR1 primer set as forward primers. Applying the 400bp protocol for library preparation and sequencing, we obtained an average of 188461 raw reads (±51602 SD) but, after length filtering, only approximately 20% (range: 13.1%-24.2%) of reads was available for post-processing analysis, thus confirming the negative impact of the fixation process.

In order to standardize and optimize sequencing output, starting from i) archival FFPE samples without any enrichment step for the selection of tumor cells and ii) low quantity and quality of extracted gDNA, we decided to use the FR3 primer set (Figure 20, right table) in order to amplify and sequence (200bp protocol) shorter fragments from both diagnostic tissue samples (gDNA: 40 ng) and plasma cfDNA samples (16.6-292.5 ng).

1.3. Identification of clonal IGH rearrangements in FFPE samples and in paired plasma samples of DLBCL patients

A malignant IGH clonotype (frequency threshold >5%)[108] was successfully identified from baseline FFPE samples at frequencies ranging from 6.3% to 69% (median: 10.9%) in 58% (15/26) of DLBCL patients (Table 10; Figure 36, left plot); the level of specificity was consistent with results reported in previously NGS studies working on gDNA isolated from archival FFPE samples as input DNA.[114] Seven out of 26 (27%) turned out to be polyclonal (Table 10) with different clones present at similar low frequencies (<5%; Figure 36, right plot). Of note, in 2 of these 7 polyclonal patients, only the ~10% of reads passed filter for the presence of forward VH primer (from 216898 to 28437 reads, from 325281 to 31602 reads), suggesting the possibility of an incomplete DH-JH rearrangement as clonal marker.

Additionally, in 4 DLBCL patients (15%) a low number of sequences corresponding to the IGH locus (on-target reads, range: 0.27%-3.12%) was found (Table 10). Probably due to the non-specific annealing between degraded DNA fragments during library construction, soft clipped bases can be added at the 5′ or 3′ end of FFPE sequencing reads.[151] Despite the sufficient

sequencing yields, these soft clipped reads lower the percentage of on-target reads because a read that overlaps only the flanking region of a target reference, due to poor matching of 5' or 3' bases, is not counted (off-target).

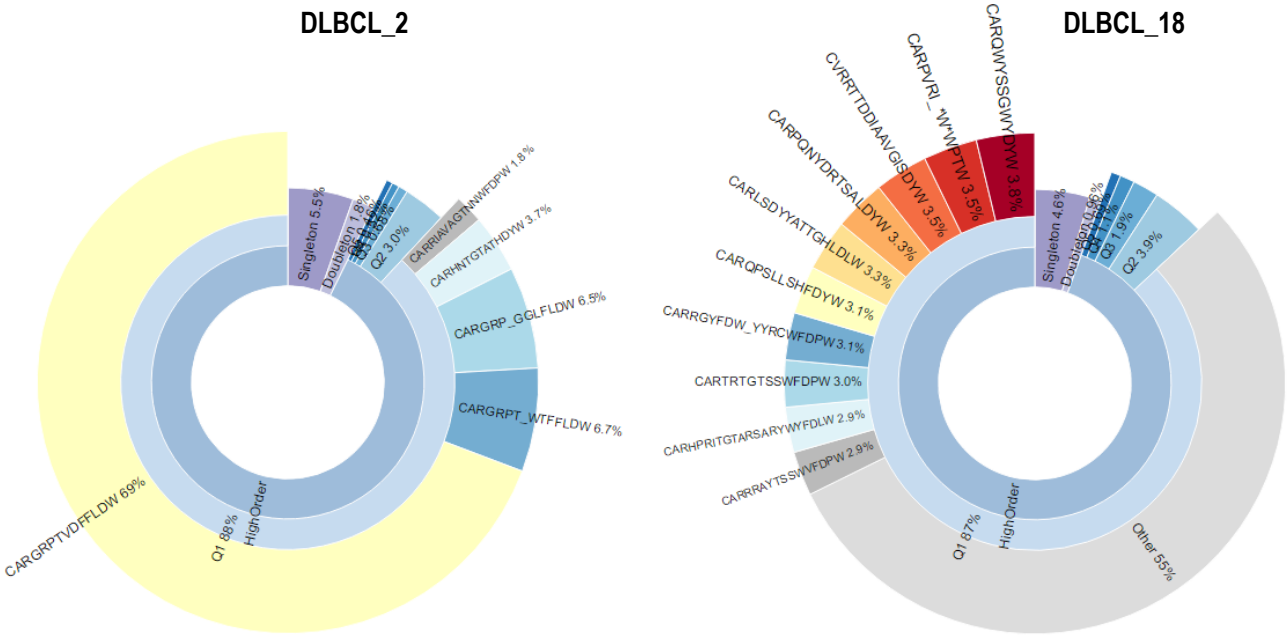


Figure 36. IGH clonality status in FFPE samples of patients affected by DLBCL. Sunburst charts of 2 representative patients generated by VDJtools software; high order clonotypes (represented by at least 3 identical reads) are divided into five quantiles (top 20% of high order clonotypes and so on), and the amino acidic sequence of top ten clonotypes of first quantile are explicitly shown (counterclockwise direction); the size of each segment is proportional to clonotype frequency. DLBCL_2: patient with a dominant malignant clonotype (yellow segment); DLBCL_18: patient with many clonotypes at low non-malignant frequencies (red segment is the most frequent).

For DLBCL patients with a tumor-derived IGH rearrangement successfully identified, we then analyzed paired baseline PB samples. The same tumor clonotype was identified in cfDNA of 12 out of 15 patients (80%) at percentages ranging from 1.4% to 74.8% (median: 5.6%) (Table 10).

Table 10. IGH sequencing results from baseline gDNA and cfDNA of DLBCL patients.

Sample	CDR3 sequence	Frequency in gDNA	Frequency in cfDNA
DLBCL_1	No IGH tumor clonotype (<5%)	2.8%	-
DLBCL_2	5'-TGTGCGCGAGGCCGCCCGACGGTGGACTTTTTTCTAGACTGG-3'	69.0%	67%
DLBCL_3	Off-target reads	-	-

DLBCL_4	No IGH tumor clonotype (<5%)	3.4%	-
DLBCL_5	5'-TGTGCGAGAGGTTCGGGTTTGACTGGTTTTAATAAAGGAGCGGACTACTGG-3'	7.3%	4.5%
DLBCL_6	5'-TGTGCGAGAGTTGGAGTAGCGTACTATGATAGTAGTGGTTATTACGGGACGTACTTTGACTATTGG-3'	53.0%	45.3%
DLBCL_7	5'-TGTGCGAGATACAAATTTTGGACTGGAGATTACAGTTTTGACTCTTGG-3'	7.5%	5.5%
DLBCL_8	5'-TGTGCGAGAGGCGGGCTTACCGGGTATCGCAGCAGCTGGTACTTCTTTGACTTCTGG-3'	6.3%	Not detected
DLBCL_9	5'-TGTGCACTTATGGGGCAACAGATGATACTAGACTACTGG-3'	61.0%	74.6%
DLBCL_10	No IGH tumor clonotype (<5%)	1.8%	-
DLBCL_11	5'-TGTGCCGGGGCGATCGGCAGTGGCTGGAACTGGTACTTCGATCTCTGG-3'	8.5%	5.7%
DLBCL_12	No IGH tumor clonotype (<5%)	2.3%	-
DLBCL_13	5'-TGTGCGAGACTTGAGGGGCTATTTACTGACTACGGTGACTACAACTGGTTCGACCCCCTG-3'	6.5%	Not detected
DLBCL_14	5'-TGTGCGAGAGATTCGGAGAACTATGGACACAACTGGTTTGACCCCTGG-3'	21.0%	3.1%
DLBCL_15	5'-TGTGCGAGACGATCTGTAAGAGGAAGCGTTGACAACTGG-3'	65.4%	74.8%
DLBCL_16	No IGH tumor clonotype (<5%)	2.6%	-
DLBCL_17	Off-target reads	-	-
DLBCL_18	No IGH tumor clonotype (<5%)	3.8%	-
DLBCL_19	Off-target reads	-	-
DLBCL_20	5'-TGTGCGAGGGGGACGGTGGCTACGCGTCAGTTTGACTACTGG-3'	8.2%	1.6%
DLBCL_21	5'-TGTGCAAGAGGGGACGGGTATAGCAGTGGCTGGTACTACTTTGACTACTGG-3'	10.9%	2.1%
DLBCL_22	No IGH tumor clonotype (<5%)	2.7%	-
DLBCL_23	Off-target reads	-	-
DLBCL_24	5'-TGTGCGAGACGATCTGTAAGAGGAAGCGTTGACAACTGG-3'	56.0%	54.0%
DLBCL_25	5'-TGTGCGAGAGATTCCTGGAGTAGTCGGGACTTCTATGGCTACTGG-3'	7.1%	Not detected
DLBCL_26	5'-TGTGCGAGAGATCGAAACGGTATGGACGTCTGG-3'	18.2%	3.4%

Specifically, as reported in Table 10, the highest IGH cfDNA frequencies (range: 45.3%-74.8%) were detected in patients characterized by a marked monoclonal IGH expansion in the tissue sample (range: 53%-69%). Pearson's correlation test has indeed demonstrated a strong positive correlation (Figure 37) between frequencies of clonal IGH rearrangements obtained from FFPE gDNA samples and frequencies obtained from plasma cfDNA samples (r=0.9754, P<0.001).

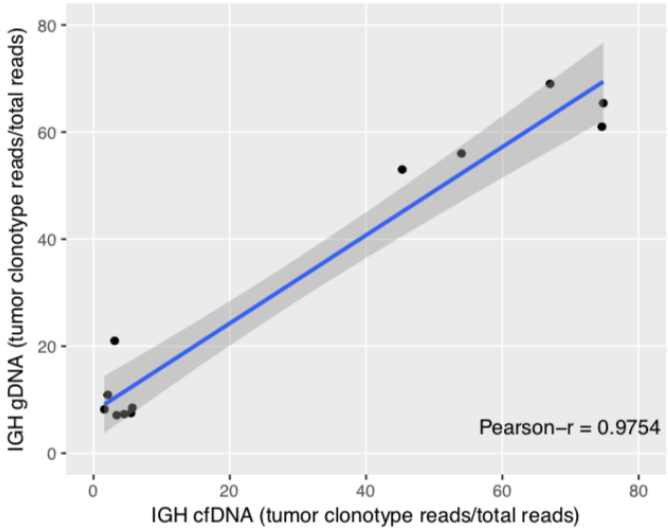


Figure 37. Comparison between IGH cfDNA and gDNA sequencing at baseline. Scatter plot (n=12) showing correlation between IGH cfDNA sequencing (x-axis) and IGH gDNA sequencing (y-axis). Blue line: best fit line obtained using the linear model method in R; shaded area: 95% CI of the best fit line. Pearson's correlation test, applied in R with *linear model* method and *cor.test* function, was used to quantify the strength of the relationship (Pearson-r).

1.4. Longitudinal cfDNA analysis for disease monitoring

To determine if cfDNA sequencing can be used to monitor a DLBCL patient's response to systemic therapy, we analyzed a total of 24 serial plasma samples from 6 selected patients. Three mL of stored plasma for each time points were used in order to increase the input cfDNA and therefore the sensitivity of our workflow. Specifically, for each patient we decided to investigate the levels of circulating tumor-associated IGH clonotype in the following four time points: at baseline (diagnosis), during R-CHOP-based chemotherapy (after 2 and after 4 cycles), and at the end of treatment (EOT, after 6 cycles). Two representative examples (DLBCL_9, Figure 38; DLBLCL_5, Figure 39) of how changes detected in IGH cfDNA reflect changes in tumor burden with a high level of sensibility and specificity are highlighted below.

DLBCL_9 patient, diagnosed with an IPI score of 3, presented a diffuse gastric wall thickening

particularly evident in the pyloric antrum and in the stomach body, as documented by the CT scans collected at baseline (Figure 38, red arrow and box). PET scans (not shown for low resolution) indicated a SUVmax of 25. When profiling IGH gene rearrangements in this patient's samples collected at diagnosis, an expansion of the IGHV4-34 was detected in the tissue biopsy (frequency: 61%) and consistently in the cfDNA sample (frequency: 74.6%). After 2 cycles of R-CHOP therapy, the level of the clonotypic IGH rearrangement in cfDNA decreased abruptly to 0.2%, and interim imaging after completion of the third cycle confirmed a clear reduction in gastric walls. Additional cfDNA assessment after 4 cycles showed an increase in the frequency of the clonotypic IGH rearrangement to 6.3%, preceding the evidence of disease progression detected by imaging routinely performed at the end of the 6 cycles of treatment (recurrence of wall thickening and SUVmax = 21.5). NGS results of paired cfDNA at EOT displayed a marked increase in circulating IGH levels reaching 52.3% of frequency.

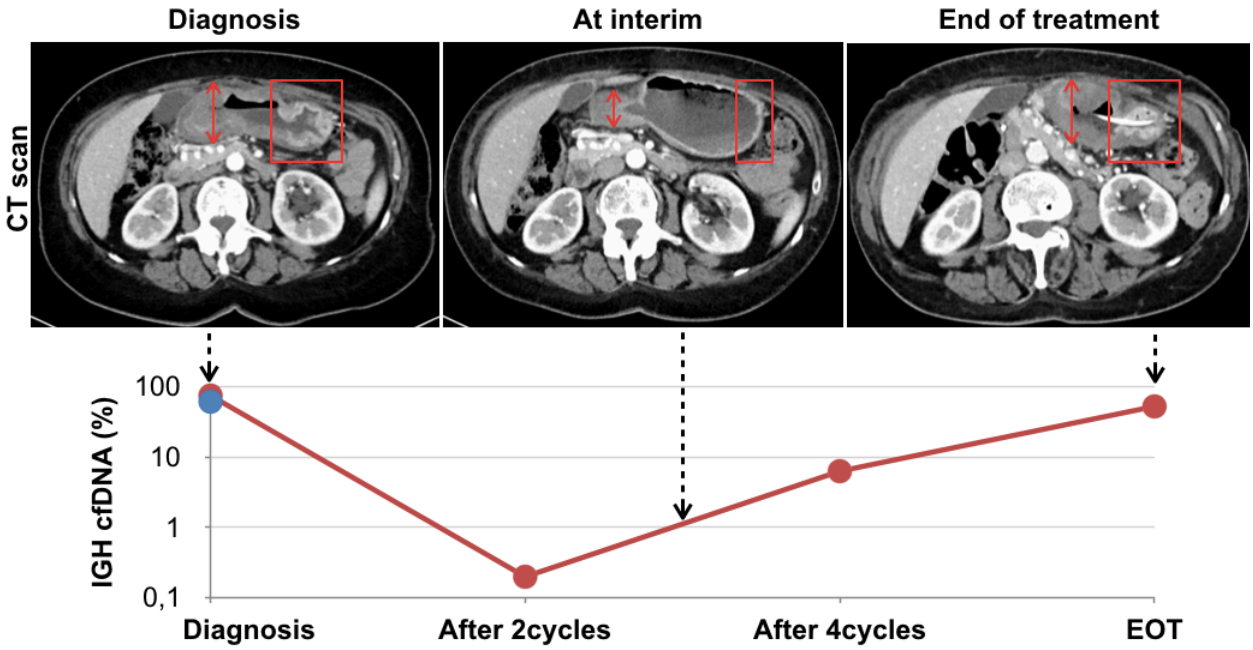


Figure 38. IGH cfDNA surveillance for disease progression. Above are reported CT scans of patient DLBCL_9 prior, during and after treatment, with red arrow and box indicating patient specific disease-related changes in gastric wall. Line plot below shows the corresponding changes (red dots) in tumor-associated IGH cfDNA clonotype (log10 y-axis) at 4 different time points (x-axis). Blue dot: frequency of malignant IGH clonotype identified in FFPE gDNA at diagnosis.

The major disadvantage of PET-CT imaging is the low specificity: it is often difficult to differentiate inflammation from tumor at the site of initial disease detection after treatment.[152] In these cases with ambiguous imaging findings, that would normally be followed-up by close repeated imaging

scans, cfDNA analysis could provide the opportunity for non-invasive monitoring of disease levels. DLBCL_5 patient was initially diagnosed with a bulky tumor mass (SUVmax = 34.5) at the level of superior mediastinum detected by PET-CT scans (Figure 39, red boxes). IGH rearrangements prior to therapy were detected at comparable frequencies on both tissue biopsy and cfDNA (IGH gDNA: 7.3% and IGH cfDNA: 4.5%; full blue and red dot, respectively, in Figure 39). After 2 ABVD cycles, the malignant circulating IGH rearrangement became undetectable and, as demonstrated by IGH cfDNA sequencing after 4 and 6 cycles, patient ended treatment schedule remaining in a stable molecular remission (empty dots in Figure 39). Despite negative plasma sequencing output, interim (after 3 cycles) and EOT (after 6 cycles) PET-CT exams reported the presence of a possible residual disease at the mediastinal level (SUVmax of 4.7 and 3.5, respectively; red arrows and gray dots in Figure 39). The persistent residual disease detected by imaging techniques was subsequently proven to be a false-positive result when subjected to biopsy and subsequent histological examination. Immunohistochemistry has indeed showed an inflammatory condition instead of a neoplastic signal (reactive $CD3^+$ T-cells, $CD56^+$ NK-cells, $CD20^-$ B-cells).

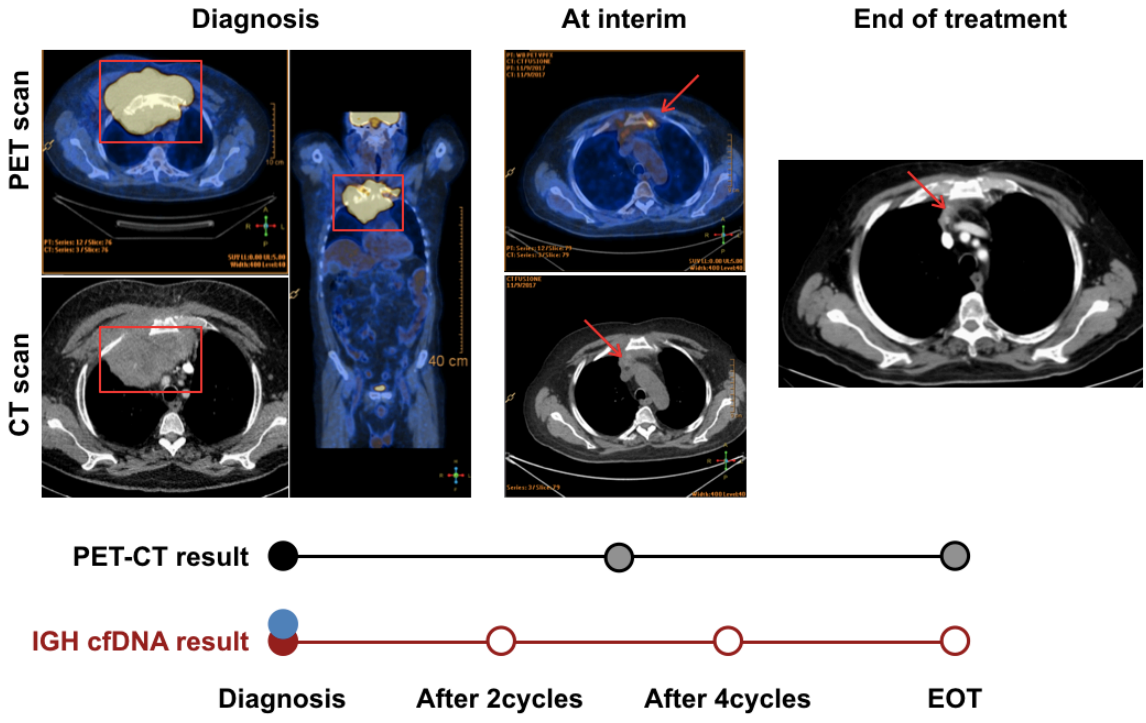


Figure 39. IGH cfDNA analysis can clarify uncertain imaging results. Above are reported PET-CT scans of patient DLBCL_5 prior, during and after treatment; red box indicating the mediastinal tumor mass, red arrows indicating imaging-positive signal of uncertain significance as reported by the Department on Nuclear Medicine of our Institute. Below are reported the line plots representing PET-CT or IGH cfDNA results (full dot = positive result; empty dot = negative result). Blue dot represents the identification of malignant IGH clonotype in FFPE gDNA at diagnosis.

2. Hodgkin lymphoma

2.1. Patient characteristics and imaging parameters

To assess the feasibility of analyzing clonal IGH rearrangements in HL, we prospectively collected PB samples and tissue biopsies of 20 cHL patients. Baseline features of patients are described in Table 11. All patients, except one characterized by mixed cellularity histology, were classified as nodular sclerosis subtype (95%). Early- (55%) and advanced-stage (45%) HL patients were respectively treated with 4 or 6 cycles of ABVD chemotherapy followed by involved-field irradiation.

Table 11. Baseline characteristics of HL study cohort.

Parameters	
Total patients, *n* (%)	20 (100%)
Age, *n* (%) < 45 ≥ 45	 12 (60%) 8 (40%)
Sex, *n* (%) Male Female	 11 (55%) 9 (45%)
Stage stratification, *n* (%) Early (IA, IB, or IIA without bulky disease) Advanced (III, IV, I or II with bulky disease, or IIB)	 11 (55%) 9 (45%)
IPS score, *n* (%) Low, 0-3 High, ≥ 4	 16 (80%) 4 (20%)
SUVmax, median (range)	12.2 (6.4-19.9)
PET-2, *n* (%) Positive Negative Not assessable	 6 (30%) 12 (60%) 2 (10%)
EBER, *n* (%) Positive Negative Not assessable	 4 (20%) 11 (55%) 5 (25%)
Treatment Response, *n* (%) CR / PR / PD	 13 (62%) / 4 (20%) / 3 (15%)
Median follow-up, months (range)	8 (6-21)

IPS, International Prognostic Score; SUVmax, highest standardized uptake volume; PET-2, positron emission tomography after 2 cycles of chemotherapy; EBER, Epstein-Barr virus encoded RNA; CR/PR, complete and partial response; PD, progressive disease.

After 2 cycles of treatment, all patients underwent an interim PET-CT scan to assess residual disease which is predictive of survival. According to the Lugano criteria,[39] PET-2 was considered negative if the Deauville score reported by the Department of Nuclear Medicine was < 3 for early-stage HL or < 4 for advanced-stage HL.

2.2. Identification of clonal IGH rearrangements in baseline plasma cfDNA of HL patients

Given the high level of correlation between clonotypic IGH frequencies in FFPE gDNA and in paired plasma cfDNA samples demonstrated in DLBCL patients (Figure 37), and in order to overcome tissue biopsy limitations due to the low representation (<5%) of HRS malignant cells [Introduction, Paragraph 3.1], for HL patients we decided to use a reverse approach: plasma cfDNA collected at diagnosis was used to identify the clonal IGH rearrangement that was then searched in the tissue biopsy.

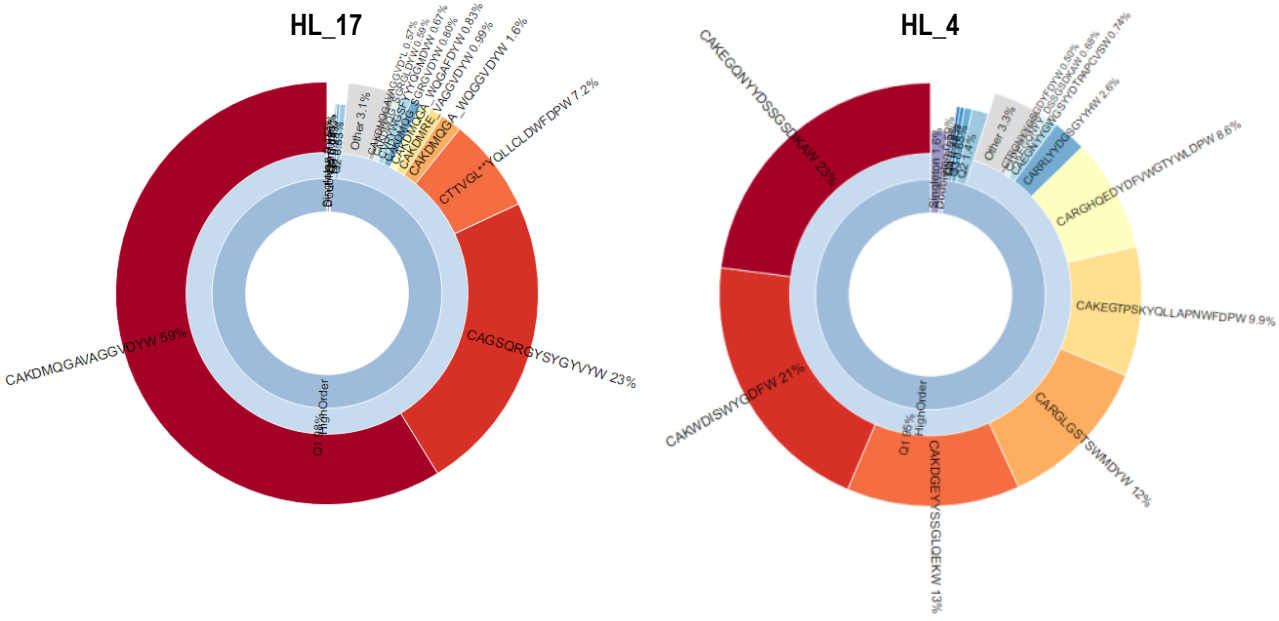


Figure 40. IGH clonality status in plasma cfDNA samples of patients affected by HL. In the representative patient HL_17 on the left, the most represented clonotype (dark red segment) is present at 59%, while the frequency of the second clonotype (red segment) is 23%; therefore, the CDR3 sequence reported in the dark red segment was considered the clonal IGH rearrangement of patient HL_17. On the contrary, in patient HL_4 on the right, the frequency of the most frequent clonotype is over the 5% but not twice that of the second clonotype (23% *vs* 21%); patient HL_4 was therefore excluded from downstream analysis. Sunburst charts were generated by VDJtools.

While for tissue biopsies there is a validated frequency threshold of 5% to define a clone as a

tumor clonotype,[108] no criteria are reported for defining clonality status in cfDNA. Therefore, we decided to apply the tissue biopsy threshold in cfDNA analysis assigning the presence of a clonal IGH gene rearrangement in patients for whom a specific IGH CDR3 sequence was detected at a frequency over the 5%. Additionally, to be considered tumor-specific, this sequence had to be present at a frequency at least twice that of the second most represented sequence (Figure 40).

Using these established parameters, our sequencing method (200bp protocol) identified a circulating tumor clonotype in 8 out of 20 HL patients at frequencies ranging from 17.4% to 66% (Table 12) and with a specificity level (40%) similar to those obtained in DLBCL patients (58%). Of note, 6 patients with a clonal IGH rearrangement (75%) had advanced-stage HL. The high frequencies observed in 5 of these advanced-stage patients (range 59%-66%) could suggest a correlation between the level of tumor burden detected by the IGH cfDNA sequencing and those assessed by standard clinical parameters for Hodgkin disease at diagnosis.

Table 12. Cell-free DNA sequencing results for HL patients with a tumor IGH clonotype.

Sample	% most represented IGH clone	CDR3 sequence
HL_1	19.3%	5'-TGTGCGAGAGATCGAAACGGTATGGACGTCTGG-3'
HL_7	63.0%	5' TGTGCGAGAGGCGGGGAGTATAGCAGCAGCTGGTATACCTAC TACTACTACGGTATGGACGTCTGG-3'
HL_9	61.7%	5' -TGTACCACAGATACCGACTTTTACCATATTGTAGTAGTACCAG CTGCTATCTGG-3'
HL_12	62.0%	5'-TGTGCGAGAGTACAGTGGCTACCAATAAACCACTTTGACTACT GG-3'
HL_14	17.4%	5'-TGTGCGAGAGTCCGGTATAACTGGAAGGGCTACTTTTACTGG-3'
HL_16	66.0%	5'-TGTGCGAGAGATTTGATGGGGATGGCTAGAGGCTACTGG-3'
HL_17	59.0%	5' TGTGCAAAAGATATGCAGGGAGCAGTGGCAGGGGGCGTTGAC TACTGG-3'
HL_19	23.5%	5'TGTGCAAAAGATATTGGGACTACAGTAACTTTCTACGGTATGGA CGTCTGG-3'

We then searched for the clonotypic IGH rearrangement identified in cfDNA in FFPE gDNA samples. Given the presence of rare tumor HRS cells in tissue biopsies, to increase the level of

sensitivity, the 8 gDNA libraries were sequenced in a 318 chip to obtain at least $5x10^5$ reads per sample. After quality and length filtering, we obtained an average of 284307 reads (±61775 SD), but neither the identified tumor IGH clonotype nor another clonal IGH gene rearrangements (frequency > 5%) were detected.

2.3. Droplet digital PCR for detection and quantification of clonal IGH rearrangements in gDNA samples

Before excluding the possible application of a reverse approach to identify clonal IGH rearrangements, we decided to use an alternative highly sensitive molecular test to assess the presence of the tumor-specific clonotype in tissue biopsies. FFPE gDNA samples were analyzed with a ddPCR approach that, based on the partitioning of the sample, allows the detection of rare variants with the same probability of wild-type variants, and therefore could point out a specific HL-associated IGH rearrangement over the normal B-cell repertoire.

To set the sensitivity of the ddPCR method, we initially performed six 10-fold dilutions of Namalwa cell line gDNA, bearing a specific clonal rearrangement [Results Aim I, Paragraph 1.1], in a costant background (250 ng) of gDNA extracted from leukocytes of 5 healthy donors. Namalwa gDNA within water-in-oil droplets was then amplified with VH4 family-specific forward primer and FAM-labeled probe [Materials and Methods, Figure 20 and Paragraph 9] and a specific reverse primer complementary to the Namalwa CDR3 sequence in order to detect the clonal rearrangement specific of the cell line over the normal B-cell background. As reported in Figure 41, the lowest concentration level of Namalwa-specific IGH rearrangement detected by ddPCR was equivalent to 10^{-5} dilution, consistent with the sensitivity reached with our developed NGS workflow [Table 5] and with previously published results on the application of ddPCR for the analysis of IGH rearrangements in hematological malignancies.[99] Specificity was confirmed by the absence of positive droplets for the Namalwa IGH rearrangement in healthy donor (PolyHD in Figure 41) and in no-template control wells (NTC in Figure 41).

The method has also shown a high degree of linearity: 2129 copies/µL (mean) of Namalwa-specific IGH rearrangement were detected in 250 ng of undiluted cell line samples, 192.4 copies/µL at the

10^{-1} dilution, 28.6 at the 10^{-2} dilution, 3.7 at the 10^{-3} dilution, 0.48 at the 10^{-4} dilution, and a mean of 0.09 copies/µL at the latest detectable dilution of 10^{-5}.

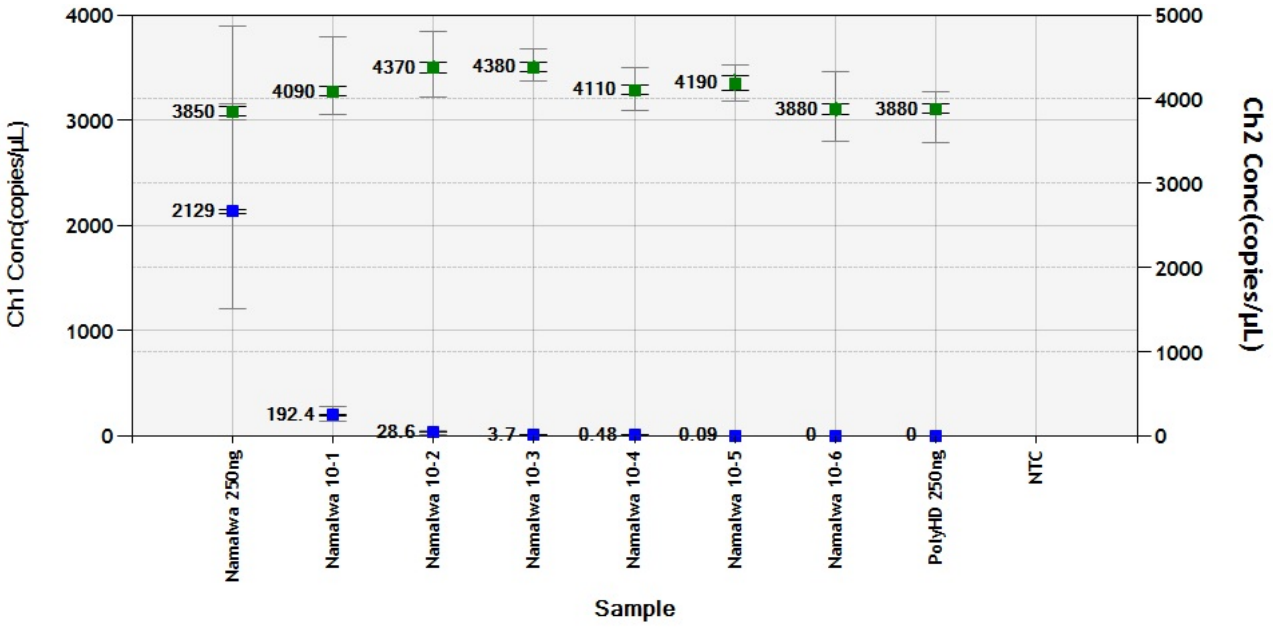


Figure 41. Preliminary experiment to assess sensitivity of ddPCR in IGH rearrangement detection. Concentration plot shows merged 6-fold replicate wells across Namalwa cell line titration series from undiluted sample to 10^{-6} dilution in a constant background of healthy donors' leukocytes (PolyHD). Copies/µL indicated on y-axes are counted in 2 different channel: channel 1 (Ch1) for FAM-labeled IGHV probe (left y-axis), channel 2 (Ch2) for HEX-labeled albumin probe (right y-axis). The blue markers indicate the copies of Namalwa-specific clonal IGH rearrangement/µL; the green markers indicate the copies of albumin reference gene/µL. All error bars are generated by the QuantaSoft software and represent a 95% confidence interval. NTC indicates the no-template negative controls.

We then moved to the analysis of matched gDNA extracted from the tissue biopsy of each HL patient with a circulating IGH marker (n=8). Using patient-specific reverse primers complementary to the clonal CDR3 identified in cfDNA and applying the same experimental conditions used for the preliminary assay, we performed a ddPCR run using 8 different reagent mixes to assess the presence of identified clonal rearrangements in FFPE gDNA samples. In each patient well, no positive droplets were called by QuantaSoft software, despite optimal results for positive controls. These results, consistent with those already obtained applying our developed NGS workflow, tend to exclude the association of the circulating clonal rearrangements detected in plasma samples with HRS tumor cells. The clonal expansion observed in plasma compartment could be related to the immune response to viral antigens (3/8 patients are EBER positive and 1 presented a cytomegalovirus infection), but further studied are needed to test this hypothesis.

DISCUSSION AIM I

Cell-free DNA is emerging as a non-invasive disease biomarker in solid tumors and lymphomas.[111,113] However, its clinical utility is still under investigation and few data are available in MM context [Introduction, Paragraph 6]. Here we provide the first report describing the clinical significance of detecting and monitoring cfDNA in patients with relapsed/refractory MM using an IGH deep-sequencing method. We extend previous findings[127] by demonstrating that this approach allows: (i) the identification of the tumor-associate IGH clonotype in BM tumor PCs and consequently in plasma samples; (ii) the assessment of tumor levels at study entry and during therapy from plasma samples; (iii) the analysis of residual disease using cfDNA with results comparable to standard MFC.

The clonally rearranged IGH genes of mature B-cell malignancies offer specific somatic DNA sequences that can serve as molecular markers for tumor cells [Introduction, Paragraph 3]. We used standardized primers developed by us[96,153] and by the BIOMED-2 concerted action[9] to amplify all IGH sequences in patients' tumor samples. Specifically, we decided to use FR1 primers that allow a more comprehensive characterization of the VH segment involved in the IGH V(D)J recombination process [Introduction, Paragraph 2], providing additional information regarding the IGH mutational status (SHM process) and the phylogenetic relationships among different clonotypes. Libraries were sequenced on the Ion Torrent PGM bench-top system and data were analyzed using a custom bioinformatic workflow. Based on freely available online tools and open-source tools in bash scripts, our pipeline was developed to discard aspecific or incomplete amplification products and retain only high-quality sequences as well as complete V(D)J gene rearrangements for clonotypes' aggregation and post-processing analysis.

Clonotypic IGH sequences could be determined from baseline tumor cells in 22 out of 25 patients, indicating a specificity (88%) similar to that reported using commercial sequencing approaches.[108,145] Consistent results were obtained with traditional Sanger sequencing to

confirm the clonality status, and using a different primer set (FR3) with the same NGS system to exclude amplification bias. Moreover, the IGH V(D)J gene segment usage in our cohort of MM patients, resembling that seen by other studies,[146,147] supports the idea that the approach we have used can be applied for identification, characterization and quantification of myeloma-associated IGH clonotype in tumor samples.

Consistently, in all patients with the characterized IGH gene rearrangement in tumor samples, our NGS approach was able to detect the same rearrangement in plasma samples. Here, a predominant clonotype was detected among other clones likely representative of normal B-cells. In contrast, in healthy donors no clonally related sequences were observed, most likely reflecting the enormous diversity of IGH V(D)J combinations generated during B-cell development. Our results therefore suggest that PCR amplification and identification of tumor IGH cfDNA is technically feasible even investigating V(D)J rearrangements approximately 300 bp long.

A significant advantage of cfDNA analysis in MM patients care would be the ability to measure tumor burden avoiding invasive BM biopsies. Quantification of BM PCs is crucial for diagnosis and prognostication, as re-emphasized in the revised IMWG guidelines.[64] However, the degree of BM infiltration detected by conventional morphology or immunohistochemistry may vary significantly not only among, but also within the same patient. This has been attributed to the patchy pattern of PCs infiltration, a factor that could also explain the inconsistency of PCs counts on BM aspirates as a prognostic factor.[90] MFC and innovative molecular methods are currently used as highly-sensitive techniques to determine tumor levels, but they are still limited by invasive biopsies and suboptimal results due to sampling variability [Introduction, Paragraph 5.2, Table 3, “Patch samples” row]. In the present study, we demonstrate the clinical implication of the detection of cfDNA: MM patients with a high cfDNA level of the tumor-associated IGH sequence before treatment have a shorter PFS than the others. This suggests that cfDNA levels might reflect tumor burden and represent potentially relevant prognostic biomarkers at study entry.

Molecular disease monitoring using cfDNA has a clear clinical relevance even during therapy. In fact our findings reveal the ability of cfDNA analysis to track tumor kinetics as already shown in lymphoma,[124] and suggest that cfDNA might complement the longitudinal assessment of BM PCs

counts, serum monoclonal proteins or sFLC concentration for response evaluation.

In the MRD context, the current gold standard involves immunophenotyping using flow cytometry on BM samples to detect the residual MM cells bearing an aberrant phenotype [Introduction, Paragraph 5.1]. In our study, results obtained analyzing MRD by MFC using consensus antibody panels [Materials and Methods, Paragraph 2.3] showed complete concordance with cfDNA analysis in all cases. Of note, patients defined as MRD-negative by MFC ($<5x10^{-5}$ BM PCs) were also characterized by very low frequencies of the tumor-associated clonotype in plasma samples ($<10^{-5}$) and had a significantly longer PFS than patients with $>10^{-5}$ BM PCs.

Several studies have shown that NGS of the IGH represents an effective and highly promising tool for MRD detection in BM samples of lymphoma and MM patients with sensitivity that may approach 10^{-6} or better.[106,145] Sensitivity is a critical aspect in MRD detection and is limited by the number of input cells or DNA. However, the sensitivity that can be obtained using cfDNA is still unknown as the origin of cfDNA has yet to be fully understood. In this study, our sequencing approach on cfDNA allowed us to analyze at least $5x10^5$ reads among which the clonal rearrangement was searched for. Increasing the amount of cfDNA will in the future serve to further improve the theoretical sensitivity. Another reasonable option, given the cfDNA enrichment in short fragments (approximately 160 bp, the length of DNA on one nucleosome), would be to focus on shorter IGH fragments using the FR3 primer set. Additionally, it still remains to be determined whether enhanced sensitivity will provide supplementary clinical information in the context of current therapy protocols for MM.[154]

Despite the limited sample size, results demonstrate that our developed deep-sequencing workflow is feasible for the identification and the monitoring of IGH gene rearrangements in tumor samples and in serial plasma samples of patients affected by MM. Moreover, since our method is based on a straightforward pipeline tailored for the analysis of the monoclonal B-cell expansion, it can be implemented in any laboratory with NGS capabilities with a short turn-around time, and can be exploitable for the study of circulating tumor rearrangements in other B-cell malignancies.

DISCUSSION AIM II

To evaluate the possibility of applying our developed NGS workflow in the non-invasive monitoring of other mature B-cell neoplasms, we decided to focus on DLBCL and HL.

Early-stage DLBCL and HL are considered curable with combined chemotherapy, resulting in more than 60% of DLBCL and 75-80% of HL patients surviving for at least 5 years after diagnosis [Introduction, Paragraph 3.1 and 3.2.1]. Given these excellent outcomes, minimizing the adverse effects of treatment, while maintaining effective disease control, is critical. In addition, it is critical to identify the subset of patients that will develop refractory or relapsed disease, which can be fatal. In both scenarios, sensitive tools that allow disease outcome prediction and monitoring as well as individualization of therapy are highly desirable to avoid under- or over-treatment.

Approaches for initial staging, response evaluation and surveillance monitoring in DLBCL and HL have historically relied upon serial imaging assessments, using CT and PET-CT.[39] Following the dynamics of lymphoma response during the early stages of chemotherapy has a clear clinical value, as demonstrated by the role of interim PET after 2 cycles (PET-2) of chemotherapy in directing treatment intensification in case of inadequate responses in HL. Nonetheless, imaging has limited specificity and sensitivity in detecting small volume disease as highlighted by the ~20-25% of HL patients with PET-negative remission that ultimately relapse.

While imaging in lymphoma represents a surrogate of the presence of tumor cells, the availability of molecular markers of disease would allow a direct and more accurate identification of residual lymphoma cells during treatment and follow-up. The recent advent of highly sensitive and specific NGS techniques provided the opportunity to measure the level of MRD analyzing the circulating tumor cfDNA in peripheral blood of those hematological malignancies such as DLBCL and HL that have very rarely circulating tumor cells, but release tumor DNA in the bloodstream.[155] The analysis of cfDNA relies on the fact that any B-cell malignancy will have a unique IGH V(D)J sequence, created through genetic rearrangements [Introduction, Paragraph 2]. If the tumor IGH clonotype

can be identified, this sequence can be used to identify and quantify tumor cfDNA by NGS.

The clinical utility of the NGS analysis of clonal IGH gene rearrangements in cfDNA has been already documented in DLBCL patients,[114,124] although the reported results are generated retrospectively using commercial NGS platforms (Adaptive Biotechnologies) that limit the routine clinical application due to elevated costs and turn-around time. Moreover, in the majority of cases, input gDNA is extracted from freshly isolated and immediately frozen tissue biopsies, a type of specimen usually not collected during standard diagnostic procedures determining additional discomfort for patients.

We therefore decided to profile the IGH repertoire in a cohort of 26 newly diagnosed DLBCL patients and we were able to demonstrate that: i) our custom sequencing workflow can be successfully applied to the analysis of IGH clonal rearrangements present in pre-treatment FFPE tumor biopsies and paired cfDNA in DLBCL patients; ii) IGH cfDNA levels reflect DLBCL tumor burden at diagnosis and can be used for the non-invasive disease monitoring with results in line with those obtained using standard imaging techniques.

Applying the short fragments-based NGS protocol, we successfully identified the presence of a clonal IGH rearrangement in the 58% of FFPE diagnostic samples (frequency range: 6.3-69%), without applying any initial enrichment steps for the selection of tumor cells. This specificity rate could be implemented at the amplification level adding primers complementary to the IGHD region to detect not only complete V(D)J rearrangements but also incomplete DJ rearrangements.

Once the tumor-associated IGH rearrangement was identified in pre-treatment biopsies, then matched cfDNA extracted from plasma samples were analyzed. The same tumor clonotype was detected in 80% of patients (frequency range: 1.4-74.8%), consistent with previously reported results.[114] Of note, we observed an extremely high level of correlation between the frequencies of clonal IGH rearrangements detected in FFPE gDNA and in paired plasma cfDNA samples ($r=0.9754$, $P<0.001$; Pearson's correlation test). Although the clonotype was identified in cfDNA of the majority of patients, we cannot rule-out the possibility that an increased in input volumes of

plasma samples would result in improved detection probability in cfDNA.

The longitudinal analysis of plasma samples (n=24) of 6 DLBCL patients obtained during the course of treatment, revealed that levels of circulating tumor-associated IGH gene rearrangements reflect changes in tumor burden as detected by imaging modalities, including escalating levels at progressive disease and declining or undetectable levels at complete response (Figure 38-39).

These data support the hypothesis that the methodology and the analysis workflow we have pioneered to characterize the IGH repertoire from gDNA and cfDNA can be used to identify and monitor the tumor-associated V(D)J rearrangement in different B-cell malignancies. This implies that the deep-sequencing of IGH cfDNA has the potential to individualize treatment in order to avoid over-treatment and to tailor intensification strategies on specific patients needs.

However, whether IGH NGS combined with traditional biopsy and imaging can improve outcome has yet to be determined. The reported preliminary results I have obtained with my work, prompted the design of a large prospective observational study that will validate the feasibility in the clinical setting of disease monitoring through the quantitative assessment of plasma clonotypic IGH gene rearrangements by NGS in 240 DLBCL patients at diagnosis who will undergo conventional treatment and standard disease assessment by imaging. This study also includes the analysis of DLBCL-specific mutations (mutNGS)[156] in cfDNA that will be conducted at Humanitas Research Hospital, Milano. The possibility to monitor treatment response using NGS methods that profile overtime the mutational landscape of DLBCLs, could allow to detect the appearance of new clones related to chemoresistance or increased aggressiveness of disease justifying a closer monitoring for progression. This study will provide the bases for addressing a number of unmet medical needs in the field of lymphoma management by: i) validating in a larger prospective clinical trial the utility of the innovative NGS-based tool we have developed in measuring residual disease in DLBCLs; ii) defining the prognostic role of MRD assessment using cfDNA in DLBCL patients; iii) validating cfDNA genotyping to identify actionable genomic biomarkers and to track treatment resistance.

The encouraging data obtained in the training cohort of DLBCL patients prompted us to hypothesize that the identification of tumor-specific IGH rearrangements by NGS on cfDNA might be applied to HL in order to non-invasively assess response kinetics and possibly residual disease, providing the opportunity for early stratification of patients based on depth of response and risk of relapse, overcoming the discomfort and exposure to radiation of repeated imaging.

Despite remaining elusive for decades, the clonal B-cell nature of HRS cells has been demonstrated by light chain restriction on immunohistochemistry,[36] and later by molecular techniques showing clonal DNA rearrangements of the immunoglobulin heavy and/or light chain locus.[37] However, no single case expresses intact immunoglobulins due to the profoundly deranged transcriptional program of HRS cells and the presence of crippling mutations in a fraction of cells [Introduction, Paragraph 3.1]. Furthermore, the paucity of tumor cells in biopsy specimens mandate that molecular studies are performed in microdissected or otherwise purified HRS cells, thus negating the routine use of this technique in the clinic. Since those early days though, the advent of NGS techniques has changed the perspective in the field, allowing the identification of unique DNA sequences from as low as one cell in a million and from cfDNA.[157]

We have therefore decided to explore the feasibility of identifying the tumor-specific IGH gene rearrangements in gDNA from HL tissue biopsies and in matched plasma cfDNA using the NGS approach we have developed. In order to overcome the limitations imposed by the paucity of HL tumor cells dispersed in an inflammatory milieu that characterizes HL biopsies, we have decided to completely reverse the approach used from MM and DLBCL and to start with the analysis of cfDNA. Even when assessing different approaches, the IGH sequences present at high frequencies in the analyzed plasma samples were undetectable in tumor sections. Although it is tempting to speculate that the IGH sequences so well represented in plasma samples are related to infections rather than reflecting tumor presence, we need to considered that we have not yet implemented any method to select pure HRS cells to try to identify the clonal rearrangements.

The results we have obtained to now suggest that DNA extraction from unmanipulated tissue sections, that allows contamination from the admixing reactive cell infiltrate, does not allow to identify any clonal IGH rearrangement in HL. In future we will need to assess the feasibility of gDNA extraction and IGH amplification from 1, 5, 10, or 50 pooled tissue cores cut from areas containing aggregates of HRS cells to reduce the number of contaminating cells. Additionally, HRS cells retrieved from fresh-frozen or FFPE samples through laser capture microdissection or DEPArray cell-isolation system could be used as baits to study clonal IGH rearrangements in HL patients. Although we recognize that these efforts are time-consuming and that results are not guaranteed, this step is necessary to understand the nature of the IGH sequences present at high frequencies in plasma samples. It is clear that if performed using cfDNA, the study of the clonal IGH levels could provide an easy and accessible method less invasive and possibly equally accurate than PET to detect HL presence irrespective of its localization. Since the majority of patients affected by HL will ultimately be cured of their disease, this approach could allow the identification of low- and high-risk patient to minimize long-term toxicity and maximize cure rates.

Recently the feasibility of using plasma cfDNA as an easily accessible source of tumor cfDNA for HL genotyping has been documented.[115] This analysis based on specimens from 80 newly diagnosed and 32 refractory HL patients highlighted that most of the mutations discovered in cfDNA were also identified in paired tumor DNA from tissue biopsies. The authors identified STAT6 as the most frequently mutated gene (~40% of cases) and longitudinal cfDNA profiling identified treatment-dependent patterns of clonal evolution in patients relapsing after chemotherapy and patients receiving immunotherapy. Analysis of cfDNA clearance during front-line therapy supports the use of this NGS mutational analysis on cfDNA to track disease and to early identify refractory HL patients. These data provide the proof of concept that cfDNA may serve as novel precision medicine biomarker in HL but we believe that the incorporation of a clonal marker of HL disease (such as the IGH rearrangement) with mutational NGS analysis and imaging would result in improved ability to monitor disease outcome and evolution.

CONCLUSION

Overall, the data presented in this PhD thesis, even if preliminary, support the hypothesis that the experimental and the bioinformatic workflow we have pioneered to characterize the IGH repertoire from gDNA and cfDNA can be used to identify and monitor the tumor-associated V(D)J rearrangements in mature B-cell malignancies. Future larger prospective studies will address if the non-invasive IGH cfDNA sequencing is of predictive importance and therefore of additional value to the established protein- or imaging-based approaches in MM and in B-cell lymphomas.

REFERENCES

1. Bremnes RM, Al-Shibli K, Donnem T, Sirera R, Al-Saad S, Andersen S, Stenvold H, Camps C, Busund LT. The role of tumor-infiltrating immune cells and chronic inflammation at the tumor site on cancer development, progression, and prognosis: Emphasis on non-small cell lung cancer. J Thorac Oncol, 2011, 6:824–33
2. Cronkite DA, Strutt TM. The Regulation of Inflammation by Innate and Adaptive Lymphocytes. J Immunol Res, 2018, 2018:1–14
3. Ruddle NH, Akirav EM. Secondary Lymphoid Organs: Responding to Genetic and Environmental Cues in Ontogeny and the Immune Response. J Immunol, 2009, 183:2205–12
4. Ng YH, Chalasani G. Role of secondary lymphoid tissues in primary and memory T-cell responses to a transplanted organ. Transplant Rev, 2010, 24:32–41
5. Schroeder HWJ, Cavacini L. Structure and Function of Immunoglobulins (author manuscript). J Allergy Clin Immunol, 2010, 125:S41–52
6. Hwang JK, Alt FW, Yeap LS. Related Mechanisms of Antibody Somatic Hypermutation and Class Switch Recombination. Microbiol Spectr, 2015, 3:1–35
7. González D, van der Burg M, García-Sanz R, Fenton JA, Langerak AW, González M, van Dongen JJ, San Miguel JF, Morgan GJ. Immunoglobulin gene rearrangements and the pathogenesis of multiple myeloma. Blood, 2007, 110:3112–22
8. Li A, Rue M, Zhou J, Wang H, Goldwasser MA, Neuberg D, Dalton V, Zuckerman D, Lyons C, Silverman LB, Sallan SE, Gribben JG. Utilization of Ig heavy chain variable, diversity, and joining gene segments in children with B-lineage acute lymphoblastic leukemia: Implications for the mechanisms of VDJ recombination and for pathogenesis. Blood, 2004, 103:4602–9
9. van Dongen JJM, Langerak AW, Brüggemann M, Evans PAS, Hummel M, Lavender FL, Delabesse E, Davi F, Schuuring E, García-Sanz R, van Krieken JHJM, Droese J, González D, Bastard C, White HE, Spaargaren M, González M, Parreira A, Smith JL, Morgan GJ, Kneba M, Macintyre EA. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: Report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia, 2003, 17:2257–317
10. Roth DB. V(D)J Recombination: Mechanism, Errors, and Fidelity. Microbiol Spectr, 2014:40
11. Sadofsky MJ. The RAG proteins in V(D)J recombination: more than just a nuclease, Nucleic Acids Res, 2001, 29:1399–409
12. Bassing CH, Swat W, Alt FW. The mechanism and regulation of chromosomal V(D)J recombination. Cell, 2002, 109:45–55
13. Nguyen TV, Pawlikowska P, Firlej V, Rosselli F, Aoufouchi S. V(D)J recombination process

and the Pre-B to immature B-cells transition are altered in Fanca-/-mice. Sci Rep, 2016, 6:1–11

14. Dunn-Walters D, Townsend C, Sinclair E, Stewart A. Immunoglobulin gene analysis as a tool for investigating human immune responses. Immunol Rev, 2018, 284:132–47
15. Bagg A, Braziel RM, Arber DA, Bijwaard KE, Chu AY. Immunoglobulin heavy chain gene analysis in lymphomas: A multi-center study demonstrating the heterogeneity of performance of polymerase chain reaction assays. J Mol Diagnostics, 2002, 4:81–9
16. Vettermann C, Schlissel MS. Allelic exclusion of immunoglobulin genes: Models and mechanisms. Immunol Rev, 2010, 237:22–42
17. Mårtensson IL, Almqvist N, Grimsholm O, Bernardi AI. The pre-B cell receptor checkpoint. FEBS Lett, 2010, 584:2572–9
18. Koues OI, Oltz EM, Payton JE. Short-Circuiting Gene Regulatory Networks : Origins of B Cell Lymphoma. Trends Genet, 2015, 31:720–31
19. Malcolm TIM, Hodson DJ, Macintyre EA, Turner SD. Challenging perspectives on the cellular origins of lymphoma. Open Biol, 2016, 6
20. Qiao Q, Wang L, Meng FL, Hwang JK, Alt FW, Wu H. AID Recognizes Structured DNA for Class Switch Recombination. Mol Cell, 2017, 67:361–373.e4
21. De Silva NS, Klein U. Dynamics of B cells in germinal centres. Nat Rev Immunol, 2015, 15:137–48
22. Shaffer AL, Rosenwald A, Staudt LM. Lymphoid malignancies: The dark side of B-cell differentiation. Nat Rev Immunol, 2002, 2:920–32
23. Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H TJ. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. Lyon, France: IARC Press, 2008
24. Fisher SG, Fisher RI. The epidemiology of non-Hodgkin's lymphoma. Oncogene, 2004, 23:6524–34
25. Campo E. Pathology and classification of aggressive mature B-cell lymphomas. Hematol Oncol, 2017, 35:80–3
26. Rickert RC. New insights into pre-BCR and BCR signalling with relevance to B cell malignancies. Nat Rev Immunol, 2013, 13:578–91
27. Küppers R. Mechanisms of B-cell lymphoma pathogenesis. Nat Rev Cancer, 2005, 5
28. Mullighan CG. Genome sequencing of lymphoid malignancies. Blood, 2013, 122:3899–908
29. Walsh S, Rosenquist R. Immunoglobulin gene analysis of mature B-cell malignancies. Med Oncol, 2005, 22:327–41
30. Thomas RK, Re D, Zander T, Wolf J, Diehl V. Epidemiology and etiology of Hodgkin's lymphoma. Ann Oncol, 2002, 13:147–52
31. Vockerodt M, Cader FZ, Shannon-Lowe C, Murray P. Epstein-Barr virus and the origin of hodgkin lymphoma. Chin J Cancer, 2014, 33:591–7
32. Fields P, Wrench D. Hodgkin lymphoma. Medicine, 2017, 45:305–10

33. Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, Advani R, Ghielmini M, Salles GA, Zelenetz AD, Jaffe ES. The 2016 revision of the World Health Organization classi fi cation of lymphoid neoplasms. Blood, 2016, 127:2375–91
34. Küppers R, Hansmann ML. The Hodgkin and Reed/Sternberg cell. Int J Biochem Cell Biol, 2005, 37:511–7
35. Schwering I, Bra A, Klein U, Jungnickel B, Tinguely M, Diehl V, Hansmann M, Dalla-favera R, Rajewsky K, Ku R. Loss of the B-lineage – specific gene expression program in Hodgkin and Reed-Sternberg cells of Hodgkin lymphoma. Blood, 2003, 101:1505–12
36. Kuppers R, Rajewsky K, Zhao M, Simons G, Laumann R, Fischer R, Hansmann ML. Hodgkin disease: Hodgkin and Reed-Sternberg cells picked from histological sections show clonal immunoglobulin gene rearrangements and appear to be derived from B cells at various stages of development. Proc Natl Acad Sci, 1994, 91:10962–6
37. Marafioti T, Hummel M, Foss HD, Laumen H, Korbjuhn P, Anagnostopoulos I, Lammert H, Demel G, Theil J, Wirth T, Stein H. Hodgkin and reed-sternberg cells represent an expansion of a single clone originating from a germinal center B-cell with functional immunoglobulin gene rearrangements but defective immunoglobulin transcription. Blood, 2000, 95:1443–50
38. Meti N, Esfahani K, Johnson NA. The role of immune checkpoint inhibitors in classical hodgkin lymphoma. Cancers (Basel), 2018, 10
39. Cheson BD, Fisher RI, Barrington SF, Cavalli F, Schwartz LH, Zucca E, Lister TA. Recommendations for initial evaluation, staging, and response assessment of hodgkin and non-hodgkin lymphoma: The lugano classification. J Clin Oncol, 2014, 32:3059–67
40. Allen PB, Gordon LI. Frontline Therapy for Classical Hodgkin Lymphoma by Stage and Prognostic Factors. Clin Med Insights Oncol, 2017, 11
41. Hasenclever D, Diehl V. A prognostic score for advanced Hodgkin's disease. International Prognostic Factors Project on Advanced Hodgkin's Disease. N Engl J Med, 1998, 339:1506-14
42. Evens AM, Kostakoglu L. The role of FDG-PET in defining prognosis of Hodgkin lymphoma for early-stage disease. Hematology Am Soc Hematol Educ Program, 2014, 135–43
43. Townsend W, Linch D. Hodgkin's lymphoma in adults. Lancet, 2012, 380:836–47
44. Sehn LH, Gascoyne RD. Diffuse large B-cell lymphoma: optimizing outcome in the context of clinical and biologic heterogeneity. Blood, 2015, 125:22–32
45. Mey U, Hitz F, Lohri A, Pederiva S, Taverna C, Tzankov A, Meier O, Yeow K, Renner C. Diagnosis and treatment of diffuse large B-cell lymphoma. Swiss Med Wkly, 2012, 142:1–15
46. Menon M, Pittaluga S, Jaffe E. The Histological and Biological Spectrum of Diffuse Large B-cell Lymphoma in the WHO Classification. Cancer J, 2012, 18:411–20
47. Lenz G, Staudt LM. Aggressive Lymphomas. N Engl J Med, 2010, 362:1417–29
48. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H,

Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature, 2000, 403:503–11

49. Lim MS, Elenitoba-Johnson KSJ. Precision medicine for diffuse large B-cell lymphoma. Clin Cancer Res, 2016, 22:2829–31
50. Li S, Young KH, Medeiros LJ. Diffuse large B-cell lymphoma. Pathology, 2018, 50:74–87
51. Shah HJ, Keraliya AR, Jagannathan JP, Tirumani SH, Lele VR, Dipiro PJ. Diffuse large B-cell lymphoma in the era of precision oncology: How imaging is helpful. Korean J Radiol, 2017, 18:54–70
52. Shankland KR, Armitage JO, Hancock BW, Park W. Non-Hodgkin lymphoma. Lancet, 2012, 380:848–57
53. Zinzani PL. The many faces of marginal zone lymphoma. Hematol Am Soc Hematol Educ Progr, 2012, 2012:426–32
54. Joshi M, Sheikh H, Abbi K, Sharma K, Tulchinsky M, Epner E, Long S. Marginal zone lymphoma: Old, new, targeted, and epigenetic therapies. Ther Adv Hematol, 2012, 3:275–90
55. Ferreri AJM, Zucca E. Marginal-zone lymphoma. Crit Rev Oncol Hematol, 2007, 63:245–56
56. Jares P, Colomer D, Campo E. Genetic and molecular pathogenesis of mantle cell lymphoma: Perspectives for new targeted therapeutics. Nat Rev Cancer, 2007, 7:750–62
57. Vose JM. Mantle cell lymphoma: 2017 update on diagnosis, risk-stratification, and clinical management. Am J Hematol, 2017, 92:806–13
58. Pérez-Galán P, Dreyling M, Wiestner A. Mantle cell lymphoma : biology, pathogenesis, and the molecular basis of treatment in the genomic era. Blood, 2011, 117:26–39
59. Freedman A. Follicular lymphoma: 2018 update on diagnosis and management. Am J Hematol, 2018, 93:296–305
60. Ma S. Risk Factors of Follicular Lymphoma. Expert Opin Med Diagn, 2012, 6:323–33
61. LaCasce CJ and A. How I treat Burkitt lymphoma in adults. Clin Oncol, 2010, 124:97–101
62. Ferry JA. Burkitt's Lymphoma: Clinicopathologic Features and Differential Diagnosis, 2006:375–83
63. Palumbo A, Anderson K. Multiple Myeloma. New Engl J Med, 2011, 364:1046-60
64. Rajkumar SV, Dimopoulos MA, Palumbo A, Blade J, Merlini G, Mateos MV, Kumar S, Hillengass J, Kastritis E, Richardson P, Landgren O, Paiva B, Dispenzieri A, Weiss B, LeLeu X, Zweegman S, Lonial S, Rosinol L, Zamagni E, Jagannath S, Sezer O, Kristinsson SY, Caers J, Usmani SZ, Lahuerta JJ, Johnsen HE, Beksac M, Cavo M, Goldschmidt H, Terpos E, Kyle RA, Anderson KC, Durie BGM, Miguel JFS. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. Lancet Oncol, 2014, 15:e538–48
65. Kazandjian D, Branch LM. Multiple myeloma epidemiology and survival, a unique

malignancy. Semin Oncol, 2017, 43:676–81

66. Chawla SS, Kumar SK, Dispenzieri A, Greenberg AJ, Larson DR, Kyle RA, Lacy MQ, Gertz MA, Rajkumar SV. Clinical course and prognosis of non-secretory multiple myeloma. Eur J Haematol, 2015, 95:57–64
67. Bladé J. Monoclonal Gammopathy of Undetermined Significance. Mayo Clin Proc Innov Qual Outcomes, 2017, 1:161–9
68. Kyle RA, Remstein ED, Therneau TM, Dispenzieri A, Kurtin PJ, Hodnefield JM, Larson DR, Plevak MF, Jelinek DF, Fonseca R, Melton LJ, Rajkumar SV. Clinical Course and Prognosis of Smoldering (Asymptomatic) Multiple Myeloma. N Engl J Med, 2007, 356:2582–90
69. Kumar SK, Rajkumar V, Kyle RA, Van Duin M, Sonneveld P, Mateos MV, Gay F, Anderson KC. Multiple myeloma. Nat Rev Dis Prim, 2017, 3:1–20
70. Kuehl WM, Bergsagel PL. Multiple myeloma: Evolving genetic events and host interactions. Nat Rev Cancer, 2002, 2:175–87
71. Bolli N, Biancon G, Moarii M, Gimondi S, Li Y, de Philippis C, Maura F, Sathiaseelan V, Tai YT, Mudie L, O'Meara S, Raine K, Teague JW, Butler AP, Carniti C, Gerstung M, Bagratuni T, Kastritis E, Dimopoulos M, Corradini P, Anderson KC, Moreau P, Minvielle S, Campbell PJ, Papaemmanuil E, Avet-Loiseau H, Munshi NC. Analysis of the genomic landscape of multiple myeloma highlights novel prognostic markers and disease subgroups. Leukemia, 2018:1–13
72. Gerecke C, Fuhrmann S, Strifler S, Schmidt-Hieber M, Einsele H, Knop S. The Diagnosis and Treatment of Multiple Myeloma. Dtsch Aerzteblatt Online, 2016:470–7
73. Dimopoulos M, Kyle R, Fermand JP, Rajkumar SV, San Miguel J, Chanan-Khan A, Ludwig H, Joshua D, Mehta J, Gertz M, Avet-Loiseau H, Beksaç M, Anderson KC, Moreau P, Singhal S, Goldschmidt H, Boccadoro M, Kumar S, Giralt S, Munshi NC, Jagannath S. Consensus recommendations for standard investigative workup: Report of the International Myeloma Workshop Consensus Panel 3. Blood, 2011, 117:4701–5
74. Rajkumar SV, Kumar S. Multiple Myeloma: Diagnosis and Treatment. Mayo Clin Proc, 2016, 91:101–19
75. Greipp PR, Miguel JS, Dune BGM, Crowley JJ, Barlogie B, Bladé J, Boccadoro M, Child JA, Harousseau JL, Kyle RA, Lahuerta JJ, Ludwig H, Morgan G, Powles R, Shimizu K, Shustik C, Sonneveld P, Tosi P, Turesson I, Westin J. International staging system for multiple myeloma. J Clin Oncol, 2005, 23:3412–20
76. Sonneveld P, Avet-Loiseau H, Lonial S, Usmani S, Siegel D, Anderson KC, Chng WJ, Moreau P, Attal M, Kyle RA, Caers J, Hillengass J, Miguel JS, Van De Donk NWCJ, Einsele H, Bladé J, Durie BGM, Goldschmidt H, Mateos MV, Palumbo A, Orlowski R. Treatment of multiple myeloma with high-risk cytogenetics: A consensus of the International Myeloma Working Group. Blood, 2016, 127:2955–62
77. Terpos E, Katodritou E, Roussou M, Pouli A, Michalis E, Delimpasi S, Parcharidou A,

Kartasis Z, Zomas A, Symeonidis A, Viniou NA, Anagnostopoulos N, Economopoulos T, Zervas K, Dimopoulos MA. High serum lactate dehydrogenase adds prognostic value to the international myeloma staging system even in the era of novel agents. Eur J Haematol, 2010, 85:114–9

78. Palumbo A, Avet-Loiseau H, Oliva S, Lokhorst HM, Goldschmidt H, Rosinol L, Richardson P, Caltagirone S, Lahuerta JJ, Facon T, Bringhen S, Gay F, Attal M, Passera R, Spencer A, Offidani M, Kumar S, Musto P, Lonial S, Petrucci MT, Orlowski RZ, Zamagni E, Morgan G, Dimopoulos MA, Durie BGM, Anderson KC, Sonneveld P, Miguel JS, Cavo M, Rajkumar SV, Moreau P. Revised international staging system for multiple myeloma: A report from international myeloma working group. J Clin Oncol, 2015, 33:2863–9
79. Sedlarikova L, Kubiczkova L, Sevcikova S, Hajek R. Mechanism of immunomodulatory drugs in multiple myeloma. Leuk Res, 2012, 36:1218–24
80. Noonan K, Colson K. Immunomodulatory Agents and Proteasome Inhibitors in the Treatment of Multiple Myeloma. Semin Oncol Nurs, 2017, 33:279–91
81. Hideshima T, Anderson KC. Molecular mechanisms of novel therapeutic approaches for multiple myeloma. Nat Rev Cancer, 2002, 2:927–37
82. Durie BGM, Harousseau J-L, Miguel JS, Bladé J, Barlogie B, Anderson K, Gertz M, Dimopoulos M, Westin J, Sonneveld P, Ludwig H, Gahrton G, Beksac M, Crowley J, Belch A, Boccadaro M, Cavo M, Turesson I, Joshua D, Vesole D, Kyle R, Alexanian R, Tricot G, Attal M, Merlini G, Powles R, Richardson P, Shimizu K, Tosi P, Morgan G, Rajkumar S V, International Myeloma Working Group. International uniform response criteria for multiple myeloma. Leukemia, 2006, 20:1467–73
83. Kyle RA, Rajkumar SV. Criteria for diagnosis, staging, risk stratification and response assessment of multiple myeloma. Leukemia, 2009, 23:3–9
84. Borrello I. Can we change the disease biology of multiple myeloma? Leuk Res, 2012, 36:S3–12
85. Nishihori T, Song J, Shain KH. Minimal Residual Disease Assessment in the Context of Multiple Myeloma Treatment. Curr Hematol Malig Rep, 2016, 11:118–26
86. Goldman JM, Gale RP. What does MRD in leukemia really mean? Leukemia, 2014, 28:1131
87. Baughn LB, Linden MA. Standardization of Minimal Residual Disease Testing in Multiple Myeloma. J Appl Lab Med An AACC Publ, 2017, 2:118–22
88. Mailankody S, Korde N, Lesokhin AM, Lendvai N, Hassoun H, Stetler-Stevenson M, Landgren O. Minimal residual disease in multiple myeloma: bringing the bench to the bedside. Nat Rev Clin Oncol, 2015, 12:1–10
89. Rawstron AC, Child JA, de Tute RM, Davies FE, Gregory WM, Bell SE, Szubert AJ, Navarro-Coy N, Drayson MT, Feyler S, Ross FM, Cook G, Jackson GH, Morgan GJ, Owen RG. Minimal residual disease assessed by multiparameter flow cytometry in multiple myeloma: impact on outcome in the Medical Research Council Myeloma IX Study. J Clin

Oncol, 2013, 31:2540–7

90. Paiva B, Vidriales MB, Pérez JJ, Mateo G, Montalbán MA, Mateos MV, Bladé J, Lahuerta JJ, Orfao A, San Miguel JF. Multiparameter flow cytometry quantification of bone marrow plasma cells at diagnosis provides more prognostic information than morphological assessment in myeloma patients. Haematologica, 2009, 94:1599–602
91. van Dongen JJM, Lhermitte L, Böttcher S, Almeida J, Van Der Velden VHJ, Flores-Montero J, Rawstron A, Asnafi V, Lécrevisse Q, Lucio P, Mejstrikova E, Szczepaski T, Kalina T, De Tute R, Brüggemann M, Sedek L, Cullen M, Langerak AW, Mendonça A, MacIntyre E, Martin-Ayuso M, Hrusak O, Vidriales MB, Orfao A. EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes. Leukemia, 2012, 26:1908–75
92. Kumar S, Paiva B, Anderson KC, Durie B, Landgren O, Moreau P, Munshi N, Lonial S, Bladé J, Mateos MV, Dimopoulos M, Kastritis E, Boccadoro M, Orlowski R, Goldschmidt H, Spencer A, Hou J, Chng WJ, Usmani SZ, Zamagni E, Shimizu K, Jagannath S, Johnsen HE, Terpos E, Reiman A, Kyle RA, Sonneveld P, Richardson PG, McCarthy P, Ludwig H, Chen W, Cavo M, Harousseau JL, Lentzsch S, Hillengass J, et al. International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. Lancet Oncol, 2016, 17:e328–46
93. Flanders A, Stetler-Stevenson M, Landgren O. Minimal residual disease testing in multiple myeloma by flow cytometry: major heterogeneity. Blood, 2013, 122:1088-89
94. Bai Y, Orfao A, Chim CS. Molecular detection of minimal residual disease in multiple myeloma. Br J Haematol, 2018, 181:11–26
95. Puig N, Sarasquete ME, Balanzategui A, Martínez J, Paiva B, García H, Fumero S, Jiménez C, Alcoceba M, Chillón MC, Sebastián E, Marín L, Montalbán MA, Mateos M V., Oriol A, Palomera L, De La Rubia J, Vidriales MB, Bladé J, Lahuerta JJ, González M, Miguel JFS, García-Sanz R. Critical evaluation of ASO RQ-PCR for minimal residual disease evaluation in multiple myeloma. A comparative analysis with flow cytometry. Leukemia, 2014, 28:391–7
96. Voena C, Ladetto M, Astolfi M, Provan D, Gribben JG, Boccadoro M, Pileri A, Corradini P. A novel nested-PCR strategy for the detection of rearranged immunoglobulin heavy-chain genes in B cell tumors. Leukemia, 1997, 11:1793–8
97. Ladetto M, Brüggemann M, Monitillo L, Ferrero S, Pepin F, Drandi D, Barbero D, Palumbo A, Passera R, Boccadoro M, Ritgen M, Gökbuget N, Zheng J, Carlton V, Trautmann H, Faham M, Pott C. Next-generation sequencing and real-time quantitative PCR for minimal residual disease detection in B-cell disorders. Leukemia, 2014, 28:1299–307
98. van der Velden VHJ, Cazzaniga G, Schrauder A, Hancock J, Bader P, Panzer-Grumayer ER, Flohr T, Sutton R, Cave H, Madsen HO, Cayuela JM, Trka J, Eckert C, Foroni L, zur Stadt U, Beldjord K, Raff T, van der Schoot CE, van Dongen JJM. Analysis of minimal residual disease by Ig/TCR gene rearrangements: Guidelines for interpretation of real-time

quantitative PCR data. Leukemia, 2007, 21:604–11

99. Della Starza I, Nunes V, Cavalli M, De Novi LA, Ilari C, Apicella V, Vitale A, Testi AM, Del Giudice I, Chiaretti S, Foà R, Guarini A. Comparative analysis between RQ-PCR and digital-droplet-PCR of immunoglobulin/T-cell receptor gene rearrangements to monitor minimal residual disease in acute lymphoblastic leukaemia. Br J Haematol, 2016, 174:541–9
100. Singh C. Utility of Flow Cytometry to Classify Abnormal Plasma Cell Populations in Marrow Samples Collected from Patients with Putative Plasma Cell Neoplasms. Open J Blood Dis, 2012, 2:39–45
101. Paiva B, van Dongen JJM, Orfao A. New criteria for response assessment: role of minimal residual disease in multiple myeloma. Blood, 2015, 125:3059–68
102. Martinez-Lopez J, Sanchez-Vega B, Barrio S, Cuenca I, Ruiz-Heredia Y, Alonso R, Rapado I, Marin C, Cedena M-T, Paiva B, Puig N, Mateos M-V, Ayala R, Hernández M-T, Jimenez C, Rosiñol L, Martínez R, Teruel A-I, Gutiérrez N, Martin-Ramos M-L, Oriol A, Bargay J, Bladé J, San-Miguel J, Garcia-Sanz R, Lahuerta J-J. Analytical and clinical validation of a novel in-house deep sequencing method for minimal residual disease monitoring in a phase II trial for multiple myeloma. Leukemia, 2017:1–4
103. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol, 2008, 26:1135–45
104. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet, 2010, 11:31–46
105. Avet-Loiseau H. Minimal Residual Disease by Next-Generation Sequencing: Pros and Cons. Asco Educ B, 2016:425–30
106. Takamatsu H. Comparison of Minimal Residual Disease Detection by Multiparameter Flow Cytometry, ASO-qPCR, Droplet Digital PCR, and Deep Sequencing in Patients with Multiple Myeloma Who Underwent Autologous Stem Cell Transplantation. J Clin Med, 2017, 6:91
107. Yaari G, Kleinstein SH. Practical guidelines for B-cell receptor repertoire sequencing analysis. Genome Med, 2015, 7:1–14
108. Martinez-Lopez J, Lahuerta JJ, Pepin F, González M, Barrio S, Ayala R, Puig N, Montalban MA, Paiva B, Weng L, Jiménez C, Sopena M, Moorhead M, Cedena T, Rapado I, Victoria Mateos M, Rosiñol L, Oriol A, Blanchard MJ, Martínez R, Bladé J, Miguel JS, Faham M, García-Sanz R. Prognostic value of deep sequencing method for minimal residual disease detection in multiple myeloma. Blood, 2014, 123:3073–9
109. Langerak AW, Brüggemann M, Davi F, Darzentas N, van Dongen JJM, Gonzalez D, Cazzaniga G, Giudicelli V, Lefranc M-P, Giraud M, Macintyre EA, Hummel M, Pott C, Groenen PJTA, Stamatopoulos K. High-Throughput Immunogenetics for Clinical and Research Applications in Immunohematology: Potential and Challenges. J Immunol, 2017, 198:3765–74
110. Herve Avet-Loiseau, Jill Corre, Valerie Lauwers-Cances, Marie-Lorraine Chretien, Nelly Robillard, Xavier Leleu, Cyrille Hulin, Catherine Gentil, Bertrand Arnulf, Karim Belhadj,

Sabine Brechignac, Laurent Garderet, Lionel Karlin, Gerald Marit, Lotfi Benbou MA and NCM. Evaluation of minimal residual disease (MRD) by next generation sequencing (NGS) is highly predictive of progression free survival in the IFM/DFCI 2009 Trial. Blood, 2015

111. Bohers E, Viailly PJ, Dubois S, Bertrand P, Maingonnat C, Mareschal S, Ruminy P, Picquenot JM, Bastard C, Desmots F, Fest T, Leroy K, Tilly H, Jardin F. Somatic mutations of cell-free circulating DNA detected by next-generation sequencing reflect the genetic changes in both germinal center B-cell-like and activated B-cell-like diffuse large B-cell lymphomas at the time of diagnosis. Haematologica, 2015, 100:e280–4
112. Melani C, Roschewski M. Molecular Monitoring of Cell-Free Circulating Tumor DNA in Non-Hodgkin Lymphoma. Oncology (Williston Park), 2016, 30:731-8
113. Hocking J, Mithraprabhu S, Kalff A, Spencer A. Liquid biopsies for liquid tumors: emerging potential of circulating free nucleic acid evaluation for the management of hematologic malignancies. Cancer Biol Med, 2016, 13:215–25
114. Kurtz DM, Green MR, Bratman S V, Scherer F, Liu CL, Kunder CA, Takahashi K, Glover C, Keane C, Kihira S, Visser B, Callahan J, Kong KA, Faham M, Corbelli KS, Miklos D, Advani RH, Levy R, Hicks RJ, Hertzberg M, Ohgami RS, Gandhi MK, Diehn M, Alizadeh AA. Noninvasive monitoring of diffuse large B-cell lymphoma by immunoglobulin high-throughput sequencing. Am Soc Hematol, 2015, 125:3679–88
115. Spina V, Bruscaggin A, Cuccaro A, Martini M, Trani M Di, Forestieri G, Manzoni M, Condoluci A, Arribas A, Terzi-Di-Bergamo L, Locatelli SL, Cupelli E, Ceriani L, Moccia AA, Stathis A, Nassi L, Deambrogi C, Diop F, Guidetti F, Cocomazzi A, Annunziata S, Rufini V, Giordano A, Neri A, Boldorini R, Gerber B, Bertoni F, Ghielmini M, Stüssi G, Santoro A, Cavalli F, Zucca E, Larocca LM, Gaidano G, Hohaus S, Carlo-Stella C, Rossi D. Circulating tumor DNA reveals genetics, clonal evolution, and residual disease in classical Hodgkin lymphoma. Blood, 2018, 131:2413–25
116. Roschewski M, Staudt LM, Wilson WH. Dynamic monitoring of circulating tumor DNA in non-Hodgkin lymphoma. Blood Spotlight, 2016, 127:3127–33
117. Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: Monitoring cancer-genetics in the blood. Nat Rev Clin Oncol, 2013, 10:472–84
118. Diaz LA, Bardelli A. Liquid biopsies: Genotyping circulating tumor DNA. J Clin Oncol, 2014, 32:579–86
119. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch R, Knippers R. DNA Fragments in the Blood Plasma of Cancer Patients : Quantitations and Evidence for Their Origin from Apoptotic and Necrotic Cells DNA Fragments in the Blood Plasma of Cancer Patients : Quantitations and Evidence for Their Origin from Apoptotic and Necr. Cancer Res, 2001, 61:1659–65
120. Heitzer E, Ulz P, Geigl JB. Circulating tumor DNA as a liquid biopsy for cancer. Clin Chem, 2015, 61:112–23

121. El Messaoudi S, Rolet F, Mouliere F, Thierry AR. Circulating cell free DNA: Preanalytical considerations. Clin Chim Acta, 2013, 424:222–30
122. Bryzgunova OE, Skvortsova TE, Kolesnikova E V., Starikov A V., Rykova EY, Vlassov V V., Laktionov PP. Isolation and comparative study of cell-free nucleic acids from human urine. Ann N Y Acad Sci, 2006, 1075:334–40
123. Herrera AF, Kim HT, Kong KA, Faham M, Sun H, Sohani AR, Alyea EP, Carlton VE, Chen Y Bin, Cutler CS, Ho VT, Koreth J, Kotwaliwale C, Nikiforow S, Ritz J, Rodig SJ, Soiffer RJ, Antin JH, Armand P. Next-generation sequencing-based detection of circulating tumour DNA After allogeneic stem cell transplantation for lymphoma. Br J Haematol, 2016, 175:841–50
124. Roschewski M, Dunleavy K, Pittaluga S, Moorhead M, Pepin F, Kong K, Shovlin M, Jaffe ES, Staudt LM, Lai C, Steinberg SM, Chen CC, Zheng J, Willis TD, Faham M, Wilson WH. Circulating tumour DNA and CT monitoring in patients with untreated diffuse large B-cell lymphoma: a correlative biomarker study. Lancet Oncol, 2015, 16:541–9
125. Sarkozy C, Huet S, Carlton VEH, Fabiani B, Delmer A, Jardin F, Delfau-Larue M-H, Hacini M, Ribrag V, Guidez S, Faham M, Salles G. The prognostic value of clonal heterogeneity and quantitative assessment of plasma circulating clonal IG-VDJ sequences at diagnosis in patients with follicular lymphoma. Oncotarget, 2017, 8:1–10
126. Vij R, Mazumder A, Klinger M, O'dea D, Paasch J, Martin T, Weng L, Park J, Fiala M, Faham M, Wolf J. Deep sequencing reveals myeloma cells in peripheral blood in majority of multiple myeloma patients. Clin Lymphoma, Myeloma Leuk, 2014, 14:131–139.e1
127. Oberle A, Brandt A, Voigtlaender M, Thiele B, Radloff J, Schulenkorf A, Alawi M, Akyüz N, März M, Ford CT, Krohn-Grimberghe A, Binder M. Monitoring multiple myeloma by next-generation sequencing of V(D)J rearrangements from circulating myeloma cells and cell-free myeloma DNA. Haematologica, 2017, 102:1105–11
128. Brown M, Wittwer C. Flow cytometry: Principles and clinical applications in hematology. Clin Chem, 2000, 46:1221–9
129. Roederer M. Compensation in flow cytometry. Curr Protoc Cytom, 2002
130. Wersto RP, Chrest FJ, Leary JF, Morris C, Stetler-Stevenson MA, Gabrielson E. Doublet discrimination in DNA cell-cycle analysis. Commun Clin Cytom, 2001, 46:296–306
131. Plouffe B, Murthy S, Lewis L. Fundamentals and application of magnetic particles in cell isolation and enrichment: a review. Rep Prog Phys, 2015
132. Kumar D, Panigrahi MK, Suryavanshi M, Mehta A, Saikia KK. Quantification of DNA extracted from formalin fixed paraffin-embeded tissue comparison of three techniques: Effect on PCR efficiency. J Clin Diagnostic Res, 2016, 10:3–5
133. Lu CY, Tso DJ, Yang T, Jong YJ, Wei YH. Detection of DNA mutations associated with mitochondrial diseases by Agilent 2100 bioanalyzer. Clin Chim Acta, 2002, 318:97–105
134. Bashford-Rogers RJM, Palser AL, Huntly BJ, Rance R, Vassiliou GS, Follows GA, Kellam P.

Network properties derived from deep sequencing of human b-cell receptor repertoires delineate b-cell populations. Genome Res, 2013, 23:1874–84

135. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature, 2011, 475:348–52
136. Alamyar E, Giudicelli V, Li S, Duroux P, Lefranc M-P. IMGT/HighV-QUEST: The IGMGT web portal for immunoglobulin (Ig) or antibody and T cell receptor (TC) analyses from NGS high throughout and deep sequencing. Immunome Res, 2012, 8:1–15
137. Shugay M, Bagaev DV, Turchaninova MA, Bolotin DA, Britanova OV, Putintseva EV, Pogorelyy MV, Nazarov VI, Zvyagin IV, Kirgizova VI, Kirgizov KI, Skorobogatova EV, Chudakov DM. VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. PLoS Comput Biol, 2015, 11:1–16
138. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva E V., Chudakov DM. MiXCR: Software for comprehensive adaptive immunity profiling. Nat Methods, 2015, 12:380–1
139. Corradini P, Cavo M, Lokhorst H, Martinelli G, Terragna C, Majolino I, Valagussa P, Boccadoro M, Samson D, Bacigalupo A, Russell N, Montefusco V, Voena C, Gahrton G. Molecular remission after myeloablative allogeneic stem cell transplantation predicts a better relapse-free survival in patients with multiple myeloma. Blood, 2003, 102:1927–9
140. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucleic Acids Res, 2008, 36:503–8
141. Ladetto M, Donovan JW, Harig S, Weller E, Trojan A, Poor C, Schlossman R, Anderson KC, Gribben JG. Real-Time Polymerase Chain Reaction of Immunoglobulin Rearrangements for Quantitative Evaluation of Minimal Residual Disease in Multiple Myeloma. Biol Blood Marrow Transplant, 2000, 6:241–53
142. Donovan JW, Ladetto M, Zou G, Neuberg D, Poor C, Bowers D, Gribben JG. Immunoglobulin heavy-chain consensus probes for real-time PCR quantification of residual disease in acute lymphoblastic leukemia Immunoglobulin heavy-chain consensus probes for real-time PCR quantification of residual disease in acute lymphoblastic leukem. Blood, 2000, 95:2651–8
143. R Development Core Team R. R: A Language and Environment for Statistical Computing. R Found Stat Comput, 2011, 1:409
144. Medina Diaz I, Nocon A, Mehnert DH, Fredebohm J, Diehl F, Holtrup F. Performance of Streck cfDNA Blood Collection Tubes for Liquid Biopsy Testing. PLoS One, 2016:1–18

145. Shin S, Hwang IS, Kim J, Lee KA, Lee ST, Choi JR. Detection of immunoglobulin heavy chain gene clonality by next-generation sequencing for minimal residual disease monitoring in B-lymphoblastic leukemia. Ann Lab Med, 2017, 37:331–5
146. Ferrero S, Capello D, Svaldi M, Boi M, Gatti D, Drandi D, Rossi D, Barbiero S, Mantoan B, Mantella E, Zanni M, Ghione P, Larocca A, Passera R, Bertoni F, Gattei V, Forconi F, Laurenti L, del Poeta G, Marasca R, Cortelazzo S, Gaidano G, Palumbo A, Boccadoro M, Ladetto M. Multiple myeloma shows no intra-disease clustering of immunoglobulin heavy chain genes. Haematologica, 2012, 97:849–53
147. González D, González M, Balanzategui A, Sarasquete ME, López-Pérez R, Chillón MC, García-Sanz R, San Miguel JF. Molecular characteristics and gene segment usage in IGH gene rearrangements in multiple myeloma. Haematologica, 2005, 90:906–13
148. Lee N, Moon SY, Lee J, Park H-K, Kong S-Y, Bang S-M, Lee JH, Yoon S-S, Lee DS. Discrepancies between the percentage of plasma cells in bone marrow aspiration and BM biopsy: Impact on the revised IMWG diagnostic criteria of multiple myeloma. Blood Cancer J, 2017, 7:e530
149. Sonneveld P, Avet-Loiseau H, Lonial S, Usmani S, Siegel D, Anderson KC, Chng WJ, Moreau P, Attal M, Kyle RA, Caers J, Hillengass J, Miguel JS, Van De Donk NWCJ, Einsele H, Bladé J, Durie BGM, Goldschmidt H, Mateos MV, Palumbo A, Orlowski R. Treatment of multiple myeloma with high-risk cytogenetics: A consensus of the International Myeloma Working Group. Blood, 2016, 127:2955–62
150. Ivanov M, Laktionov K, Breder V, Chernenko P, Novikova E, Telysheva E, Musienko S, Baranova A, Mileyko V. Towards standardization of next-generation sequencing of FFPE samples for clinical oncology: Intrinsic obstacles and possible solutions. J Transl Med, 2017, 15:1–13
151. Oh E, Choi YL, Kwon MJ, Kim RN, Kim YJ, Song JY, Jung KS, Shin YK. Comparison of Accuracy of Whole-Exome Sequencing with Formalin-Fixed Paraffin-Embedded and Fresh Frozen Tissue Samples. PLoS One, 2015, 10:e0144162
152. Han HS, Escalón MP, Hsiao B, Serafini A, Lossos IS. High incidence of false-positive PET scans in patients with aggressive non-Hodgkin's lymphoma treated with rituximab-containing regimens. Ann Oncol, 2009, 20:309–18
153. Farina L, Carniti C, Dodero A, Vendramin A, Raganato A, Spina F, Patriarca F, Narni F, Benedetti F, Olivieri A, Corradini P. Qualitative and quantitative polymerase chain reaction monitoring of minimal residual disease in relapsed chronic lymphocytic leukemia: early assessment can predict long-term outcome after reduced intensity allogeneic transplantation. Haematologica, 2009, 94:654–62
154. Moreau P, Zamagni E. MRD in multiple myeloma: More questions than answers? Blood Cancer J, 2017, 7:7–10
155. Mancuso P, Calleri A, Antoniotti P, Quarna J, Pruneri G, Bertolini F. If it is in the marrow , is

it also in the blood？An analysis of 1 , 000 paired samples from patients with B-cell non-Hodgkin lymphoma. BMC Cancer, 2010, 10:644

156. Scherer F, Kurtz DM, Newman AM, Stehr H, Craig AFM, Esfahani MS, Lovejoy AF, Chabon JJ, Klass DM, Liu CL, Zhou L, Glover C, Visser BC, Poultsides GA, Advani RH, Maeda LS, Gupta NK, Levy R, Ohgami RS, Kunder CA, Diehn M, Alizadeh AA. Distinct biological subtypes and patterns of genome evolution in lymphoma revealed by circulating tumor DNA. Sci Transl Med, 2016, 8:364ra155

157. Oki Y. Detection of classical Hodgkin lymphoma specific sequence in peripheral blood using a next-generation sequencing approach. Br J Haematol, 2015, 169:689–93

APPENDIX A. Bash scripts for NGS reads processing

```
$ cd /home/user/quality
#!/bin/bash
for file in *.fastq
do
##FastX-toolkit command script##
#FASTX Statistics to analyze the median quality score of raw reads#
/home/user/fastx_toolkit/bin/fastx_quality_stats -Q 33 -i $file -o
/home/user/stats_files/${file}_raw_stats.txt
#FASTQ Quality Filter to set the minimum quality score to keep#
/home/user/fastx_toolkit/bin/fastq_quality_filter -q 20 -p 50 -Q 33 -i $file -o
/home/user/quality/${file}_qualityfilt.fastq -v > /home/user/reports/${file}_quality_filter_report.txt
##Cutadapt command script##
#Length filtering: FR1 -m 255, FR3 -m 70#
/home/user/cutadapt-1.9.1/bin/cutadapt -m 255 /home/user/quality/${file}_qualityfilt.fastq -o
/home/user/quality/${file}_length.fastq > /home/user/reports/${file}_length_filter_report.txt
#Primers identification and trimming#
/home/user/cutadapt-1.9.1/bin/cutadapt -b file:/home/user/primer_seq/rev_primers.fasta -n 2
/home/user/quality/${file}_length.fastq -o /home/user/quality/${file}_output_rev.fastq --discard-
untrimmed > /home/user/reports/${file}_REV_report.txt
/home/user/cutadapt-1.9.1/bin/cutadapt -b file:/home/user/primer_seq/fw_primers.fasta -n 2
/home/user/quality/${file}_output_rev.fastq -o /home/user/quality/${file}_output_fw.fastq --discard-
untrimmed > /home/user/reports/${file}_FW_report.txt
rm /home/user/quality/${file}_output_rev.fastq
##FastX-toolkit command script##
#FASTX Statistics to analyze the median quality score of filtered reads#
/home/user/fastx_toolkit/bin/fastx_quality_stats -Q 33 -i /home/user/quality/${file}_output_fw.fastq -
o /home/user/stats_files/${file}_QC_stats.txt
done
```

```
$ cd /home/user/fastq_fasta
#!/bin/bash
for file in *.fastq
do
##FastX-toolkit command script##
#FASTQ-to-FASTA for conversion to FASTA format#
/home/user/fastx_toolkit/bin/fastq_to_fasta -n -Q 33 -i $file -o ${file}.fasta
```

```
/home/user/fastx_toolkit/bin/fasta_formatter -i ${file}.fasta -o fasta_QC/${file}.fasta
rm ${file}.fasta
done

$ cd /home/user/vdjtools_analysis
#!/bin/bash
for file in *.txt
do
##VDJtools command script for the aggregation into clonotypes##
#Convert to prepare datasets, input files: 3_Nt-sequences from IMGT/HighV-QUEST#
java -jar /home/user/vdjtools-1.0.8/vdjtools.jar Convert -S imgthighvquest $file
/home/user/vdjtools_analysis/converted/${file}_converted.txt
done
$ cd /home/user/vdjtools_analysis/converted
#!/bin/bash
for file in *.txt
do
#Correct to perform a frequency-based erroneous clonotypes correction#
java -jar /home/labtmo/vdjtools-1.0.8/vdjtools.jar Correct $file
/home/user/vdjtools_analysis/corrected/${file}_corrected.txt
done

$ cd /home/user/mixcr
#!/bin/bash
for file in *.fasta
do
##MiXCR command script for the aggregation into clonotypes##
#Alignment to IMGT germline database#
java -Xmx4g -Xms3g -jar /home/labtmo/mixcr-2.1.5/mixcr.jar align --library imgt $file
alignments.vdjca
#Clone assembly performing frequency-based correction#
java -Xmx4g -Xms3g -jar /home/labtmo/mixcr-2.1.5/mixcr.jar assemble
-OcloneClusteringParameters.searchParameters=twoMismatchesOrIndels
-OclusteringFilter.specificMutationProbability=5E-2 alignments.vdjca clones.clns
#Exporting IGH clones to tab-delimited file#
java -Xmx4g -Xms3g -jar /home/labtmo/mixcr-2.1.5/mixcr.jar exportClones --chains IGH
clones.clns clones.txt
done
```

APPENDIX B. Published papers

1. **Biancon G**, Gimondi S, Vendramin A, Carniti C, Corradini P. *Non-invasive molecular monitoring in multiple myeloma patients using cell-free tumor DNA: a pilot study*. J Mol Diagn. 2018 Nov;20(6):859-870. doi: 10.1016/j.jmoldx.2018.07.006. Epub 2018 Aug 28. PubMed PMID: 30165206.

2. Bolli N, **Biancon G**, Moarii M, Gimondi S, Li Y, de Philippis C, Maura F, Sathiaseelan V, Tai YT, Mudie L, O'Meara S, Raine K, Teague JW, Butler AP, Carniti C, Gerstung M, Bagratuni T, Kastritis E, Dimopoulos M, Corradini P, Anderson KC, Moreau P, Minvielle S, Campbell PJ, Papaemmanuil E, Avet-Loiseau H, Munshi NC. *Analysis of the genomic landscape of multiple myeloma highlights novel prognostic markers and disease subgroups*. Leukemia. 2018 Dec;32(12):2604-2616. doi: 10.1038/s41375-018-0037-9. Epub 2018 May 22. PubMed PMID: 29789651; PubMed Central PMCID: PMC6286326.

The Journal of Molecular Diagnostics, Vol. 20, No. 6, November 2018

ELSEVIER

the Journal of Molecular Diagnostics
jmd.amjpathol.org

Noninvasive Molecular Monitoring in Multiple Myeloma Patients Using Cell-Free Tumor DNA

A Pilot Study

Giulia Biancon,*† Silvia Gimondi,*† Antonio Vendramin,† Cristiana Carniti,* and Paolo Corradini*†

From the Department of Medical Oncology and Hematology, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano; and the Department of Hematology,† Università degli Studi di Milano, Milano, Italy*

Accepted for publication July 9, 2018.

Address correspondence to Cristiana Carniti, Ph.D., Department of Medical Oncology and Hematology, Fondazione IRCCS Istituto Nazionale dei Tumori, via Venezian 1, Milano 20133, Italy. E-mail: cristiana.carniti@istitutotumori.mi.it.

Novel treatments for multiple myeloma (MM) have increased rates of complete response, raising interest in more accurate methods to evaluate residual disease. Cell-free tumor DNA (cfDNA) analysis may represent a minimally invasive approach complementary to multiparameter flow cytometry (MFC) and molecular methods on bone marrow aspirates. A sequencing approach using the Ion Torrent Personal Genome Machine was applied to identify clonal *IGH* gene rearrangements in tumor plasma cells (PCs) and in serial plasma samples of 25 patients with MM receiving second-line therapy. The same clonal *IGH* rearrangement identified in tumor PCs was detected in paired plasma samples, and levels of *IGH* cfDNA correlated with outcome and mirrored tumor dynamics evaluated using conventional laboratory parameters. In addition, *IGH* cfDNA levels reflected the number of PCs enumerated by MFC immunophenotyping even in the complete response context. Patients determined by MFC to be free of minimal residual disease were characterized by low frequencies of tumor clonotypes in cfDNA and longer survival. This pilot study supports the clinical applicability of the noninvasive monitoring of tumor levels in plasma samples of patients with MM by *IGH* sequencing. *(J Mol Diagn 2018, 20: 859—870; https://doi.org/10.1016/j.jmoldx.2018.07.006)*

The treatment of multiple myeloma (MM) has changed during the past decade with the introduction of several classes of new effective drugs that have greatly improved response rates,[1,2] highlighting the need for more accurate methods of residual disease assessment. Several studies using multiparameter flow cytometry (MFC) or allele-specific oligonucleotide quantitative PCR have found that the presence of residual tumor cells after therapy is associated with shorter progression-free survival (PFS).[3] More recently, next-generation sequencing (NGS) techniques on bone marrow (BM) aspirates have been used.[4] Minimum residual disease (MRD) negativity by NGS of the *IGH* gene rearrangements is significantly associated with longer time to progression and better overall survival in patients treated with front-line novel therapies. Similar results were observed using MFC, although authors report the presence of a subset of patients who tested positive by MFC and negative by sequencing that probably reflects the patchy pattern of BM infiltration typically observed in MM.[4] This, together with extramedullary disease, represents a potential pitfall common to all techniques that use BM aspirates that may be nonrepresentative of the disease infiltration. The use of alternative methods for disease assessment, such as the monitoring of circulating cell-free tumor DNA (cfDNA), could improve the estimation of the risk of progression.[5] Ongoing studies are examining tumor cfDNA as a sensitive measure of small amounts of residual cells in lymphoma and myeloma.[6–8] The sensitivity of cfDNA analysis is still unknown, but the development of peripheral blood (PB)—based disease monitoring approaches should be

Supported by Associazione Italiana contro le Leucemie-linfomi e mieloma, Associazione Italiana Ricerca Cancro, and Fondazione Adiuvare.
G.B. and S.G. contributed equally to this work.
C.C. and P.C. contributed equally to this work as senior authors.
Disclosures: None declared.

https://doi.org/10.1016/j.jmoldx.2018.07.006

a goal because it would allow for serial sampling without repeated BM biopsies (BMBs) in MM. In the present study, we describe a deep-sequencing method that allows identifying and quantifying residual tumor burden in patients with MM from plasma samples. The method was applied to a cohort of myeloma samples collected prospectively within a clinical trial.

Materials and Methods

Patients, Treatment Plan, and Sample Collection

We conducted a prospective randomized phase 3 trial in MM at first relapse comparing the activity of bortezomib, cyclophosphamide, and dexamethasone versus lenalidomide, cyclophosphamide, and dexamethasone as second-line therapy. Within this study, 25 patients who achieved complete response during therapy (International Myeloma Working Group guidelines)[9] and had longitudinal biological samples collected for disease monitoring were analyzed. Our institutional review board approved this study (INT 57/10), and patients provided informed consent.

BM and PB samples were obtained during routine clinical evaluations at study entry; after 3, 6, and 9 cycles of therapy; and at follow-up time points. Plasma was obtained processing PB samples collected in K2-EDTA tubes (BD Vacutainer, Becton Dickinson, Franklin Lakes, NJ) within 3 hours, with a first centrifugation at 1500 × *g* for 10 minutes and a second high-speed centrifugation at 16,000 × *g* for 10 minutes at 4°C. Plasma samples were stored at −80°C until extraction.[10] PB of 10 healthy donors was also collected. Namalwa (human Burkitt lymphoma; ACC 24) and JVM-2 (human chronic B-cell leukemia; ACC 12) cell lines were from DSMZ (Braunschweig, Germany).

Flow Cytometry and Immunomagnetic Separation

BM nucleated cells were isolated from a median of 8 mL (range, 7 to 14 mL) of BM aspirates after red blood cells lysis with a hypotonic solution (NH_4Cl, 1.5 mol/L; $KHCo_3$, 100 nmol/L; and Na_4EDTA, 10 nmol/L; pH 7.2 to 7.4). Plasma cells (PCs) were stained according to the European Myeloma Network guidelines[11] using eight-color monoclonal antibody combinations (Table 1) on a MACSQuant Analyzer (Miltenyi Biotec, Gladbach, Germany). Data were analyzed using MACSQuantify software version 2.6 (Miltenyi Biotec) and FlowJo software version 10.2 (FlowJo LLC, Ashland, OR). The target for collection was >500,000 cellular events in each tube. An immunomagnetic bead−based strategy was used to isolate BM $CD138^+$ PCs on the AutoMACS ProSeparator (Miltenyi Biotec).

Genomic and Cell-Free DNA Extraction

Genomic DNA (gDNA) was extracted using Nucleospin Tissue kit (Macherey-Nagel GmbH & Co., Düren, Germany). cfDNA was extracted from 1 to 3 mL of plasma and eluted in 50 μL of Buffer AVE using a QIAamp Circulating Nucleic Acid Kit (Qiagen, Hilden, Germany) following manufacturer's instructions. Extracted cfDNA was quantified on a high-sensitivity benchtop fluorometer (Qubit 2.0, Thermo Fisher Scientific, Waltham, MA). The size distribution of the fragments was tested by capillary electrophoresis (Agilent 2100 Bioanalyzer, Agilent Technologies, Böblingen, Germany). cfDNA was stored at −20°C until amplification and library preparation.

Table 1 List of Monoclonal Antibodies Used for Flow Cytometric Analysis

Antigen	Fluorochrome	Clone
CD19*	PE	LT19
CD20†	VioBlue	LT20
CD27*	FITC	M-T271
CD28*	PE Vio770	15E8
CD38†	BV500	HIT2
CD45*	VioBlue	5B1
CD45*	APC Vio770	5B1
CD56*	PE Vio770	REA196
CD56†	PE Cy7	MEM-188
CD81*	PerCP Vio700	REA513
CD117*	PE	A3C6E2
CD117*	APC Vio770	A3C6E2
CD138*	APC	44F9

*Miltenyi Biotec, Gladbach, Germany.
†Biolegend, San Diego, CA.
APC, allophycocyanin; FITC, fluorescein isothiocyanate; PE, phycoerythrin; PerCP, peridinin chlorophyll protein complex.

Amplification, Library Preparation, and Sequencing of *IGH* Gene Rearrangements

Amplification was performed by multiplex PCR using 500 ng of gDNA or a median of 20.3 ng of cfDNA (range, 8.6 to 45.9 ng) according to the BIOMED-2 concerted action with a consensus reverse *IGHJ* primer and in the forward direction a mix of framework region 1 (FR1) or a framework region 3 (FR3) *IGHV* family−specific primers[12,13] (Table 2). The gDNA PCR reactions were performed in 50 μL with the following components: 5 μL of dNTPs 2 mmol/L, 5 μL of buffer 10×, 5 μL of $MgCl_2$ (FR1: 20 mmol/L; FR3: 15 mmol/L), 1 μL of each primer 10 μmol/L (seven to FR1 or FR3 and one to *IGHJ*), 0.2 μL of Taq Gold 5 U/μL, and 1 to 26.8 μL of gDNA. The cfDNA PCR reactions were performed in 85 μL with the following components: 8.5 μL of dNTPs 2 mmol/L, 8.5 μL of buffer 10×, 8.5 μL of $MgCl_2$ (FR1: 20 mmol/L; FR3: 15 mmol/L), 1 μL of each primer 10 μmol/L (seven to FR1 or FR3 and one to *IGHJ*), 0.2 μL of Taq Gold 5 U/μL, and 45 μL of cfDNA. PCR conditions were 96°C for 10 minutes followed by 35 cycles of 96°C for 30 seconds, 62°C for 30 seconds, 72°C for 30 seconds, and a final extension at 72°C (FR1: 10 minutes; FR3: 5 minutes) with cooling to 4°C.

Table 2 Primers Used for *IGH* Gene Rearrangements' Amplification

Primer	Sequence
VH1D	5′-CCTCAGTGAAGGTCTCCTGCAAGG-3′
VH2D	5′-TCCTGCGCTGGTGAAAGCCACACA-3′
VH3D	5′-GGTCCCTGAGACTCTCCTGTGCA-3′
VH4aD	5′-TCGGAGACCCTGTCCCTCACCTGCA-3′
VH4bD	5′-CGCTGTCTCTGGTTACTCCATCAG-3′
VH5D	5′-GAAAAAGCCCGGGGAGTCTCTGAA-3′
VH6D	5′-CCTGTGCCATCTCCGGGGACAGTG-3′
VH1-FR3	5′-TGGAGCTGAGCAGCCTGAGATCTGA-3′
VH2-FR3	5′-CAATGACCAACATGGACCCTGTGGA-3′
VH3-FR3	5′-TCTGCAAATGAACAGCCTGAGAGCC-3′
VH4-FR3	5′-GAGCTCTGTGACCGCCGCGGACACG-3′
VH5-FR3	5′-CAGCACCGCCTACCTGCAGTGGAGC-3′
VH6-FR3	5′-GTTCTCCCTGCAGCTGAACTCTGTG-3′
VH7-FR3	5′-CAGCACGGCATATCTGCAGATCAG-3′
JHD	5′-ACCTGAGGAGACGGTGACCAGGGT-3′

Forward and reverse primers were pooled together in a multiplex PCR reaction. The expected amplicon size is in the range of 310 to 360 bp for framework region 1 (FR1) primer set and 100 to 170 bp for framework region 3 (FR3) primer set.

PCR products were controlled for quantity and quality on an Agilent 2100 Bioanalyzer using the high-sensitivity DNA kit (Agilent Technologies). Adapter ligation and barcoding were performed using Ion Plus Fragment Library kit and Ion Xpress Barcode Adapters 1-16/17-32 kits (Thermo Fisher Scientific) following the manufacturer's instructions. Barcoded libraries were purified using AMPure beads (Beckman Coulter, Brea, CA) according to the library size as reported in the manufacturer's instructions (FR1: 1×; FR3: 1.5×). Equimolar concentrations of each library were pooled and sequenced with a calibration standard panel (Thermo Fisher Scientific) on an Ion Personal Genome Machine (PGM) system with Ion PGM Hi-Q View Chef kit and Ion PGM Hi-Q View Sequencing kit using Ion 318 Chips v2 BC (Thermo Fisher Scientific). Reproducibility of the high-throughput sequencing method using the PGM was tested by running gDNA samples in duplicate with a different barcode in the same run and with the same barcode on a separate run. As for cfDNA, instead, given the limited availability, libraries were sequenced in duplicate in two runs using the same barcode. Namalwa cell line was used as internal positive control in each run.

Bioinformatics Workflow

Using the Torrent Suite software version 5.0.2 with default parameters (Thermo Fisher Scientific), raw reads were trimmed for low-quality 3′ ends and barcodes demultiplexed, thus obtaining the raw sequencing data (FASTQ format) produced by the Ion Torrent PGM. Quality assessment and reads filtering were performed using open-source tools (FastX-toolkit, Cutadapt) in a custom bash (Unix Shell) script (Supplemental Script S1). Specifically, raw sequencing reads were analyzed and filtered for base quality (median Phred score, >30; minimum quality score, 20) and sequence length (FR1: >255 bp, FR3: >70 bp) to get rid of incomplete reads or unspecific amplification products. Sequences were retained only if both forward and reverse primers could be identified, and then primer sequences were trimmed from the reads.

Filtered reads were converted to FASTA format (Supplemental Script S1) and aligned against IMGT germline database (IMGT, Montpellier, France) using the IMGT/HighV-QUEST online tool (*http://imgt.org/HighV-QUEST/login.action*, freely available on registration, last accessed October 11, 2017) with default parameters.[14]

Aligned reads were parsed using the open source VDJtools software (*https://github.com/mikessh/vdjtools*)[15] version 1.1.1. In detail, mapped reads were aggregated into clonotypes based on their complementary-determining region 3 (CDR3) nucleotide sequence and the same IGH variable, diversity, and joining [*IGH V(D)J*] gene segment use. Frequency-based correction (Supplemental Script S1) was performed with default parameters (maximum number of mismatches allowed between clonotypes being compared, 2; child-to-parent clonotype size ratio threshold under which child clonotype is considered erroneous, 0.05) to eliminate erroneous clonotypes. Corrected samples were stored as a clonotype abundance tables for the subsequent analyses.

Statistical Analysis

Repertoire overlap was analyzed with VDJtools. Fisher's exact test, *U*-test, Kaplan-Meier analysis, log-rank test, and one-way analysis of variance with posthoc Tukey honestly significant difference test were performed in R version 3.3.2 (R Development Core Team; *http://www.r-project.org*)[16] and GraphPad Prism version 5.0 (GraphPad Software, San Diego, CA). Correlation analyses were conducted using the Pearson's correlation coefficient (r). Venny version 2.1.0 (*http://bioinfogp.cnb.csic.es/tools/venny/index.html*) was used to generate Venn diagrams. $P < 0.05$ was considered statistically significant.

Results

Patient Characteristics

Clinical features of the 25 patients at their first relapse are summarized in Table 3. Previous treatments included mainly bortezomib [nine patients (36%)] and high-dose melphalan [16 patients (64%)]. Median follow-up for the cohort is 17 months (range, 2 to 82 months). At accrual time point, patients had variable levels of $CD138^+$ PCs in the BM (median, 8.8%; range, 0.05% to 80%) but no circulating $CD138^+$ PCs in the PB.

Table 3 Baseline Characteristics of the Study Cohort

Characteristic	Finding (n = 25)
Age in years, median (range)	65 (41–75)
Sex, *n* (%)	
Male	16 (64)
Female	9 (36)
MM subtype, *n* (%)	
IgG	13 (52)
IgA	8 (32)
Light chain	4 (16)
Laboratory findings	
BM infiltration on trephine biopsy, median (range), %	40 (10–95)
BM infiltration by flow cytometry, median (range), %	8.8 (0.05–80)
BM infiltration by NGS, median (range), %	50.1 (1.4–86.7)
Serum M protein, median (range), g/dL	2.60 (0.09–9.79)
Serum FLC ratio, median (range)	33 (0.01–344)
Serum LDH, median (range), U/L	294 (228–433)
Cytogenetic abnormalities, *n* (%)	
t4; 14	1 (4)
t14; 16	2 (8)
17p-	2 (8)
1q+	4 (16)
>1	6 (24)
None	5 (20)
NA	5 (20)
MM staging, *n* (%)	
ISS I	10 (40)
ISS II	6 (24)
ISS III	9 (36)
Time in months from diagnosis, median (range)	51 (22–117)
Follow-up, median (range), months	17 (2–82)

BM, bone marrow; FLC, free light chain; ISS, International Staging System; LDH, lactate dehydrogenase; MM, multiple myeloma; NA, not applicable; NGS, next-generation sequencing.

Identification and Characterization of Rearranged Clonal IGH by Ion PGM Sequencing in MM PCs

To identify the dominant tumor clonotype, rearranged *IGH V(D)J* loci in genomic tumor cell DNA (tcDNA) derived from isolated BM CD138$^+$ PCs were amplified by multiplex PCR using validated degenerate PCR primers complementary to *IGH* FR1 and the consensus *IGHJ* primer.[12,13,17] These oligonucleotides do not cause significant PCR amplification bias in multiplex PCR settings, allowing the correct representation of all the VH gene families.[18,19] The amplification generated products of expected sizes that were sequenced on Ion PGM. Sequencing yielded a means ± SD of 101,455 (79,377) reads that were subjected to quality assessment and filtering (Supplemental Figure S1). A means ± SD of 47,755 ± 32,787 sequences (47% ± 32% of total reads) passed quality filters and were used for subsequent downstream analyses. To assess clonality in the samples, the percentage of reads having the same *V(D)J* use and a unique CDR3 nucleotide sequence was determined. The most represented clone had a median frequency of 50.1% (range, 1.4% to 86.7%). Setting a frequency threshold >5% to define a clone as a tumor clonotype,[4,20] the sequencing method successfully identified a myeloma clonotype in 22 of 25 patients with MM (88%), indicating a specificity similar to the one reported in previously described NGS assays for clonal *IGH* identification (91%).[4,21]

In addition, consistent with previously published data,[22] in all 22 patients the *IGH V(D)J* gene segment use of the tumor clonotype was nonrandom: *IGHV3*, *IGHV4*, *IGHV1*, and *IGHV2* were the VH families expressed at the highest frequencies (Supplemental Figure S2A), and a preferential expression of VH3-23 [4 of 22 patients (18%)], VH1-69, VH2-5, VH3-30, VH3-48, and VH3-7 (each at 9%) was observed (Supplemental Figure S2B). *IGHD6-19* and *IGHD6-13* predominated (Supplemental Figure S2C), consistent with what seen in other B-cell malignant tumors.[23] Among the *IGHJ* subgroups (Supplemental Figure S2D), *IGHJ4* was statistically overrepresented (45% observed versus 16.7% expected, assuming a random choice for JH; $P < 0.001$, Fisher's exact test), whereas *IGHJ2* and *IGHJ3* were significantly underrepresented (5%; $P < 0.001$).

Validation Studies

Validation studies included i) comparison with traditional sequencing, ii) amplification with a different set of primers, and iii) evaluation of sensitivity. To validate the consistency of the NGS data, traditional Sanger sequencing of PCR products obtained using the same consensus sense primers complementary to FR1 and an antisense primer derived from the JH region at the 3′ end was performed.[17] Sequences were visualized with Chromas Lite version 2.1.1 (Technelysium, South Brisbane, Australia), and a single consensus sequence was generated, which was analyzed using IMGT/V-QUEST online tool (*http://imgt.org/IMGT_vquest/share/textes*, last accessed November 3, 2017).[24] In all 22 patients, the *V(D)J* identity and the CDR3 sequence of the tumor clonotype identified by NGS corresponded to that inferred with traditional Sanger sequencing approach. Moreover, in the remaining three patients for whom the NGS approach identified *IGH* sequences at frequencies under the validated threshold,[20] Sanger sequencing highlighted a polyclonal *IGH* repertoire (Table 4).

Consistent results were obtained when amplicons were generated using the FR3 primer set,[13] sequenced on the Ion PGM following manufacturer's instruction and analyzed with our custom bioinformatics workflow using a tailored length filtering. The absence of a tumor clonotype was

Table 4 Genomic Tumor Cell DNA (Bone Marrow CD138$^+$ Plasma Cells) Sequencing Statistics

Sample	Reads, *n*	Reads after filtering, *n*	Most represented *IGH* clone, %	CDR3 sequence	Sanger sequencing
MM_1	106,941	61,651	57.5	5′-TGTGCACACAGCGAGGTCGGTGGTTATTACTAC-CCTAAACCCCTTGACTTCTGG-3′	Confirmed clonal
MM_2	122,011	68,574	80.1	5′-TGTGTACGTAGGGGGCAGCCTTCAATAGTCGCCG-ATGGAGGCATCTACTTTGACCACTGG-3′	Confirmed clonal
MM_3	112,528	63,862	76.4	5′-TGTGCGAGAGTTTCGGGGGGTATAGCAGTGGCTG-CGATCGACCGACCCATATTATACTACTACGGTATG-GACGTCTGG-3′	Confirmed clonal
MM_4	113,414	64,861	1.4	No *IGH* tumor clonotype (<5%)	Confirmed polyclonal
MM_5	50,607	16,530	69.4	5′-TGTGCGAGAGGTGTCGTGAGTGGTAGCTGCTACT-TTCGAAATTGGTTCGACCCCTGG-3′	Confirmed clonal
MM_6	81,978	35,170	8	5′-TGTGTGAAGTCTCTCGGGGCTTCTTGGCACTACG-CTATGGACGTCTGG-3′	Confirmed clonal
MM_7	68,208	22,751	63.1	5′-TGTGCGAGAGATAAGATAGGAGCAGCAGCTGGTA-GTTGGTTCGACCCCTGG-3′	Confirmed clonal
MM_8	44,827	14,832	42.7	5′-TGTGCGAGAGATTTAGGGGACGCTATGGACGTCTGG-3′	Confirmed clonal
MM_9	22,078	14,968	47.6	5′-TGTGCGAGAGTCACACGAGGGTACTACTTTGACT-ACTGG-3′	Confirmed clonal
MM_10	166,011	48,873	26.6	5′-TGTGCGAGAGGCCTGGGTAACGGAGCAGCTGCCC-AGGAAACTCACCTCGTCTGGTTCGACCCCTGG-3′	Confirmed clonal
MM_11	161,260	70,487	81.4	5′-TGTGCGAAAGATCATAACGAGTGGGAGCTGAGAC-GATCCGGGGACTGG-3′	Confirmed clonal
MM_12	101,841	32,718	85.8	5′-TGTGCGATGGACCGAACTGCAACGGAGGGGCTCG-ACCCCTGG-3′	Confirmed clonal
MM_13	113,960	54,211	86.7	5′-TGTGCGAGACATTCTGGAACAGTGGCTGGTATCT-TTGACAACTGG-3′	Confirmed clonal
MM_14	89,243	40,015	55.3	5′-TGTACCACCTGGGCGACCGCAGTGTCTGGGCGACTG-3′	Confirmed clonal
MM_15	190,832	76,913	50.1	5′-TGTGCAAAAGACGGGGGGTATAGCAGTGGCTGGG-CCCAAGAGGGCTTGGACTACTGG-3′	Confirmed clonal
MM_16	48,521	20,387	2.7	No *IGH* tumor clonotype (<5%)	Confirmed polyclonal
MM_17	37,374	17,433	21.0	5′-TGTGCGAAACTGCAGGGGCATTACTATGATAGTA-GTGGTTATCCGAACTGG-3′	Confirmed clonal
MM_18	49,426	21,320	51.4	5′-TGTGCGAGTTCTATTGTAGTAGTACCACCGGGC-GTCTGG-3′	Confirmed clonal
MM_19	36,952	14,994	9.1	5′-TGTGCGAGGGATCGTGATGGCAGTGGCTGGTCCT-TTGATTACTGG-3′	Confirmed clonal
MM_20	117,138	67,269	13.0	5′-TGTGCGAGAGCGGCGTCGGCAGCAGCTGGTACG-GAGGGGTTGTTCGACCCCTGG-3′	Confirmed clonal
MM_21	174,692	80,542	62.3	5′-TGTGCGACTATAGCAGTGGCTGGTCCCTACTGGT-ACTTCGATCTCTGG-3′	Confirmed clonal
MM_22	133,781	80,193	9.8	5′-TGTGCGGCGGGCCTGGAACCCCGCTACTGG-3′	Confirmed clonal
MM_23	159,121	71,957	74.9	5′-TGTGCGACCGCCCCAATACCGGACTCGGGAAGAC-TTGACAACTGG-3′	Confirmed clonal
MM_24	66,296	23,777	4.1	No *IGH* tumor clonotype (<5%)	Confirmed polyclonal
MM_25	167,322	69,586	17.8	5′-TGTGCGAGAACGACCAGCAGCTCCTTTGGCGACTGG-3′	Confirmed clonal

CDR3, complementary-determing region 3; MM, multiple myeloma.

confirmed in the same three patients excluding, overall, a primer amplification bias under the conditions here reported.

To assess the sensitivity of the sequencing approach, samples generated by serial 10-fold dilutions of DNA extracted from the two cell lines Namalwa and JVM-2 bearing a known clonal IGH rearrangement, into healthy controls' polyclonal DNA were studied. The clonal sequences at dilutions as low as 10^{-5}, corresponding to a sensitivity of at least 0.001% [two Namalwa-specific sequence reads for a total of 60,624 filtered reads (0.003299%) and three JVM-2–specific sequencing reads for a total of 71,387 filtered reads (0.0042%)] were identified and quantified (Supplemental Table S1).

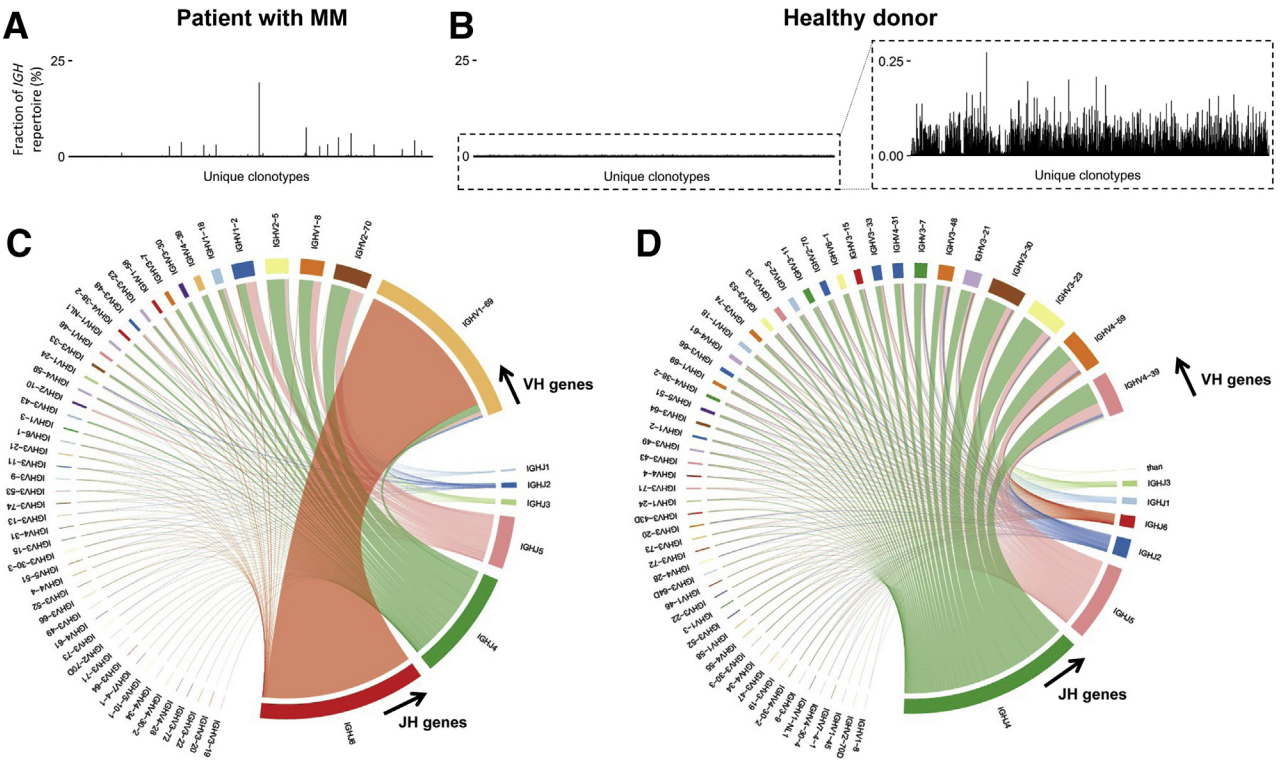


Figure 1 Sequencing of the *IGH* rearrangements in cell-free DNA of a representative patient with multiple myeloma (MM) and a healthy donor. Histogram of plasma clonotype frequencies detected in a patient with MM (**A**) and a healthy donor (**B**) with a zoom-in on lower frequencies (dotted box); unique clonotype frequencies are plotted on the *y* axis for both the patient with MM and the healthy donor (as a percentage of the total reads analyzed). V-J junction circos plot for a patient with MM (**C**) and a healthy donor (**D**). Arcs correspond to different VH and JH segments, scaled to their frequency in sample. Ribbons represent VH-JH pairings, and their broadness corresponds to the number of *IGH* sequences that exhibit this gene combination. JH, junction region; VH, variable region of the heavy chain.

cfDNA Extraction and Amplification: Experimental Setup

cfDNA was extracted with QIAamp Circulating Nucleic Acid kit (Qiagen) starting from different volumes of plasma (0.5, 1, 1.5, 3 mL). A linear correlation between plasma volume and cfDNA concentration was not observed, probably because of a saturation effect of the silica membrane. The median cfDNA quantity extracted from healthy donors' PB samples was indeed 4.6 ng from 0.5 mL of plasma, 7.5 ng from 1 mL, 13.0 ng from 1.5 mL, and 20.4 ng from 3 mL.

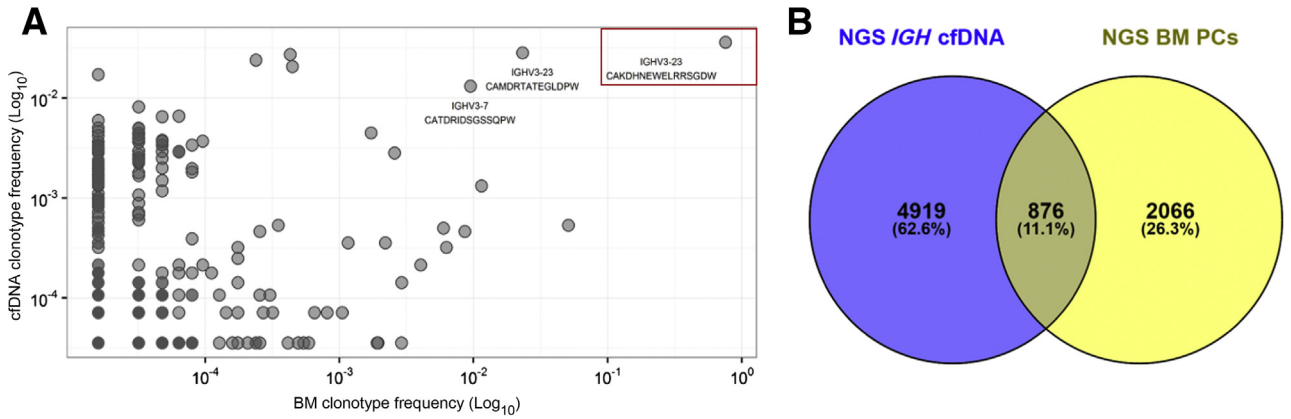


Figure 2 Overlap analysis of the clonotypes detected by next-generation sequencing (NGS) in tumor cell DNA and cell-free DNA (cfDNA). **A:** Scatterplot of the clonotype frequencies (log_{10} scale) observed in a representative patient with multiple myeloma (MM) [bone marrow (BM) clonotypes on the *x* axis and cfDNA on the *y* axis]. Most clonotypes are detected at low frequencies, whereas the tumor-associated clonotype (**red box**) can be identified as the one at the highest frequency in both compartments (plasma and BM tumor sample). Next to the top 3 most frequent shared clonotypes are the VH gene (*IGHV*x-x) and the related complementary-determining region 3 amino acid sequence. **B:** Venn diagram showing the number (x% calculated on the total clonotypes) of unique and shared clonotypes between the plasma (blue circle) and tumor (yellow circle) compartments. The Venn diagram was generated using Venny (*http://bioinfogp.cnb.csic.es/tools/venny/index.html*). PCs, plasma cells.

To verify whether the plasma samples stored within the clinical trial for patients with MM at first relapse were suitable for cfDNA extraction and amplification, the quality and quantity of cfDNA extracted from plasma into K2-EDTA tubes (BD Vacutainer) or in Cell-Free DNA BCT tubes (Streck, Inc., Omaha, NE) that contained a cell-stabilizing agent that prevents cell lysis were compared.

PB from 10 healthy donors was collected in both tubes and processed as described above within 3 hours for K2-EDTA tubes or within 3, 24, 48, 72 hours, and 1 week for Cell-Free DNA BCT tubes. Plasma samples were then stored at −80°C for at least 1 week. Of note, when plasma samples are processed within 3 hours (compatible with what was done for the samples stored within the MM clinical trial) both cfDNA BCT tubes and K2-EDTA tubes had similar performances, with only slightly higher concentrations of cfDNA from plasma samples collected in cfDNA BCT tubes (mean, 8.2 versus 7.5 ng with K2-EDTA tubes). It could clearly be confirmed that cfDNA stability is guaranteed only with the cfDNA BCT tube when samples are processed after 3 hours (median, 7.1 ng after 24 hours, 6.8 ng after 48 hours, 4.8 ng after 72 hours, 3.6 ng after 1 week), making them the best option for multicenter clinical studies in which samples are often centralized.

It was also studied whether plasma storage conditions might affect cfDNA recovery. No difference was observed when processing fresh or frozen samples (median cfDNA quantity from 1 mL of plasma: 7.8 ng after immediate processing, 7.5 ng after short-term storage of <6 months, 7.4 ng after >6 months).

Therefore, considering the minimum quantity of PCR products required for library preparation [10 to 100 ng to obtain 40 to 60 pM of final library (Ion Plus Fragment library kit and Ion PGM Hi-Q View Chef Kit protocols)], in the present study, the available 1 to 3 mL of plasma collected in K2-EDTA tubes was considered sufficient for extraction and downstream amplification.

Sequencing of the *IGH* Gene Rearrangements in Plasma Samples: Performance and Concordance of tcDNA and cfDNA Results

Matched genomic tcDNA and plasma cfDNA samples obtained at study entry were available for all 22 patients with a defined myeloma clonotype. *IGH* cfDNA sequencing yielded a means ± SD of 21,0056 (161,520) reads, with a means ± SD of 77,188 (55,058) sequences passing quality filters (37% of total reads versus 47% on tcDNA) and available for subsequent analyses. In MM plasma samples, the percentage of predominant clones ranged from 1.2% to 18% (Figure 1A). In comparison, in control plasma samples from healthy donors ($n = 10$), several different clones were present at low percentages, and no clonally related sequences were observed (mean, 0.006%; range, 0.002% to 1%) (Figure 1B).

To compare the recombinant repertoire in plasma of patients with MM and donors, Circos plots (VDJtools) showing the relative prominence of each VH-JH recombination within the repertoire of each subject were generated (Figure 1, C and D). These plots revealed that V genes are paired with many other J genes in control samples, whereas MM samples show few but broad connections, indicating the predominance of one VH-JH combination (one clonotype) in patients among other clones present at lower frequencies.

The tumor clonotype identified in cellular DNA in plasma samples was then searched. In all 22 cases (100%), the rearranged tumor-associated *IGH* sequence was present in plasma (Figure 2A) at percentages ranging from 1.2% to 18% of total filtered circulating *IGH* DNA reads (median frequency, 4.7%). As reported above for tcDNA, results were confirmed using an FR3 primer set. In addition, clonotypes (means ± SD, 10% ± 7%) were detected at high frequencies that are shared between paired plasma and tumor samples, whereas other clonotypes were found to be exclusive of each compartment (Figure 2B). Therefore, plasma cfDNA can accurately mirror the profiles of the most abundant clonotypes; furthermore, plasma cfDNA may reflect myeloma burden in BM but also in extramedullary sites.

Prognostic Implication of Clonotypic IGH cfDNA Levels in Patients with MM

Because there are different ways to determine the percentage of PCs in the BM (counting cells on trephine BMBs or on aspirate smears) and these methods often give discrepant results,[25] it was speculated that the analysis of clonal *IGH* in cfDNA may be used to help estimating tumor levels, overcoming BM sampling and analysis limitations. PCs on BMB were present in all patients before therapy at a median frequency of 40% of BM leukocytes (range, 10% to 95%): these frequencies did not correlate with clonotypic cfDNA frequencies ($r = 0.098$, $P = 0.6724$). Similarly, no significant correlation was found between M-protein levels evaluated by serum electrophoresis (SPEP) and clonotypic cfDNA frequencies ($r = -0.018$, $P = 0.9504$). *IGH* cfDNA levels and serum-free light chain (sFLC) ratios show a moderate positive but not significant correlation ($r = 0.4567$, $P = 0.2552$) (Supplemental Figure S3A). Likewise, no correlation was found between BMB PCs and SPEP levels ($r = 0.289$, $P = 0.295$), BMB PCs levels and sFLC ratios ($r = 0.148$, $P = 0.7793$), and SPEP levels and sFLC ratios ($r = -0.032$, $P = 0.8465$), supporting the idea that MM is a complex disease and none of these laboratory parameters alone accurately describes tumor levels in a single patient. Of interest, disease levels at baseline evaluated by NGS on plasma samples compare favorably to BM PCs infiltration by MFC ($r = 0.713$, $P = 0.0002$) and to NGS performed using tcDNA ($r = 0.45$, $P = 0.0354$) (Supplemental Figure S3B).

It was next determined whether cfDNA analysis might facilitate the early identification of clinically relevant risk

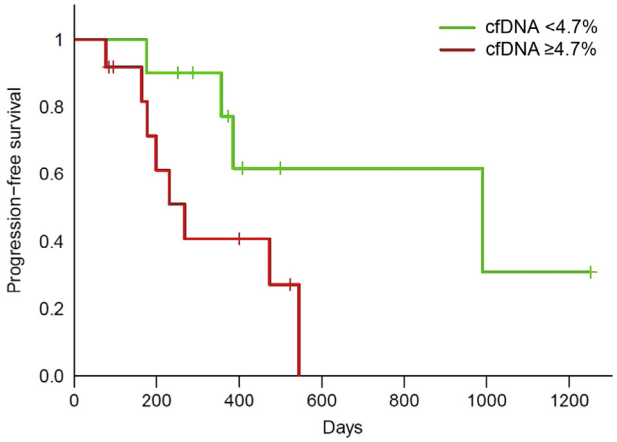


Figure 3 Correlation of *IGH* cell-free DNA (cfDNA) levels at baseline with progression-free survival. Data are expressed as estimated median values. $n = 10$ patients with frequencies <4.7%; $n = 12$ patients with frequencies ≥4.7%. Hazard ratio = 3.507 (Cox proportional hazards regression model). $P = 0.04988$ (log-rank test).

groups among the cohort of patients with MM enrolled in the study. Levels of the tumor-associated clonotype in cfDNA distinguish between groups of patients with different prognoses, with the median percentage of the tumor-associated clonotype in cfDNA as the cutoff value (4.7% of total reads). Patients with levels ≥4.7% ($n = 12$) of the tumor-associated *IGH* sequence before therapy had significantly inferior PFS [estimated median values, 268 versus 990 days; hazard ratio (HR) = 3.507, $P = 0.04988$, log-rank test] (Figure 3) than patients with levels <4.7% ($n = 10$). When considering baseline features, such as BMB PCs, SPEP levels, and sFLC ratios, no significant association was found with survival. Similarly, none of the variables known to affect clinical outcome in patients with MM, such as age, International Staging System status, or the Revised International Staging System status,[26] which includes serum lactate dehydrogenase and high-risk cytogenetic abnormalities (t4; 14, t14; 16, 17p- and 1q+),[27] significantly correlated with PFS or overall survival (Supplemental Table S2). On the other hand, as expected given the high concordance of cfDNA NGS results and MFC (Supplemental Figure S3B), higher numbers of PCs enumerated by MFC (median percentage as cutoff value, 8.8%) are associated with poorer PFS (estimated median values, 268 versus 545 days; HR = 3.745, $P = 0.04655$, log-rank test) (Supplemental Figure S4A), whereas no significant association was found between NGS results on tcDNA and survival (Supplemental Figure S4B).

Disease Monitoring by Clonotypic *IGH* cfDNA Sequencing

It was next hypothesized that longitudinal analysis of clonal *IGH* cfDNA may help understand tumor dynamics over time. Seventy plasma samples of the selected 22 patients with MM were analyzed. *IGH* cfDNA sequencing revealed that the tumor-associated clonotype could be tracked over time in plasma samples. The levels of the clonotypic *IGH* sequences in plasma reflected tumor dynamics evaluated using the International Myeloma Working Group criteria (BM PCs, SPEP levels, and sFLC ratios) (Figure 4). In addition, the phenotypic aberrancies detected in PCs at study entry for each patient (based on CD38/CD56/CD19/CD45, CD38/CD27/CD45/CD28, CD20/CD81/CD38/CD117) were used as patient-specific probes for residual disease assessment after therapy on consecutive samples. Cell-free DNA levels reflected the number of PCs enumerated by MFC immunophenotyping (Figure 4).

In the context of complete response time points ($n = 22$), MFC was used for MRD monitoring. At least 1×10^6 cellular events were collected and residual cells detected at a median frequency of 0.00065 (range, 0.00000668 to 0.027) tumor cells of the total events analyzed (range, 10,14,888 to 19,46,108). MRD negativity was defined when <50 aberrant-phenotype PCs were detected[28]; thus, six patients achieved MRD negativity during the timeframe of our analysis. When analyzing cfDNA *IGH* frequencies in these 22 patients at complete response time points (median, 0.0000395; range, 0.00000756 to 0.037), a high level of correlation was found between cfDNA NGS and MFC data ($r = 0.5831$, $P = 0.0044$, Pearson's correlation test) (Figure 5A). Accordingly, PFS was significantly prolonged ($P < 0.001$) for the six patients who achieved MRD negativity, displaying frequencies of the clonotypic cfDNA rearrangement $<10^{-5}$ (for clonotypic cfDNA frequencies $<10^{-5}$: $n = 6$, means ± SD PFS, 714 ± 327 days; for clonotypic cfDNA frequencies $\geq 10^{-5}$-$\leq 10^{-4}$: $n = 9$, 325 ± 75 days; for clonotypic cfDNA frequencies $>10^{-4}$, $n = 7$, 143 ± 59 days) (Figure 5B). In these cases, to increase theoretical sensitivity, NGS of cfDNA samples was performed with a lower level of plexing (maximum of eight samples on an Ion 318 Chip v2 BC) to obtain at least 5×10^5 reads per sample.

Discussion

cfDNA is emerging as a noninvasive disease biomarker in solid tumors and lymphomas.[29,30] However, its clinical utility is still under investigation, and few data are available in the context of MM. We provide the first report describing the clinical significance of detecting and monitoring cfDNA in patients with relapsed or refractory MM using an *IGH* deep-sequencing method. We extend previous findings[8] by demonstrating that this approach allows i) the identification of the tumor-associate *IGH* clonotype in BM tumor PCs and consequently in plasma samples, ii) the assessment of tumor levels at study entry and during therapy from plasma samples, and iii) the analysis of residual disease using cfDNA with results comparable to standard MFC.

The clonally rearranged *IGH* genes of mature B-cell malignant tumors offer specific somatic DNA sequences

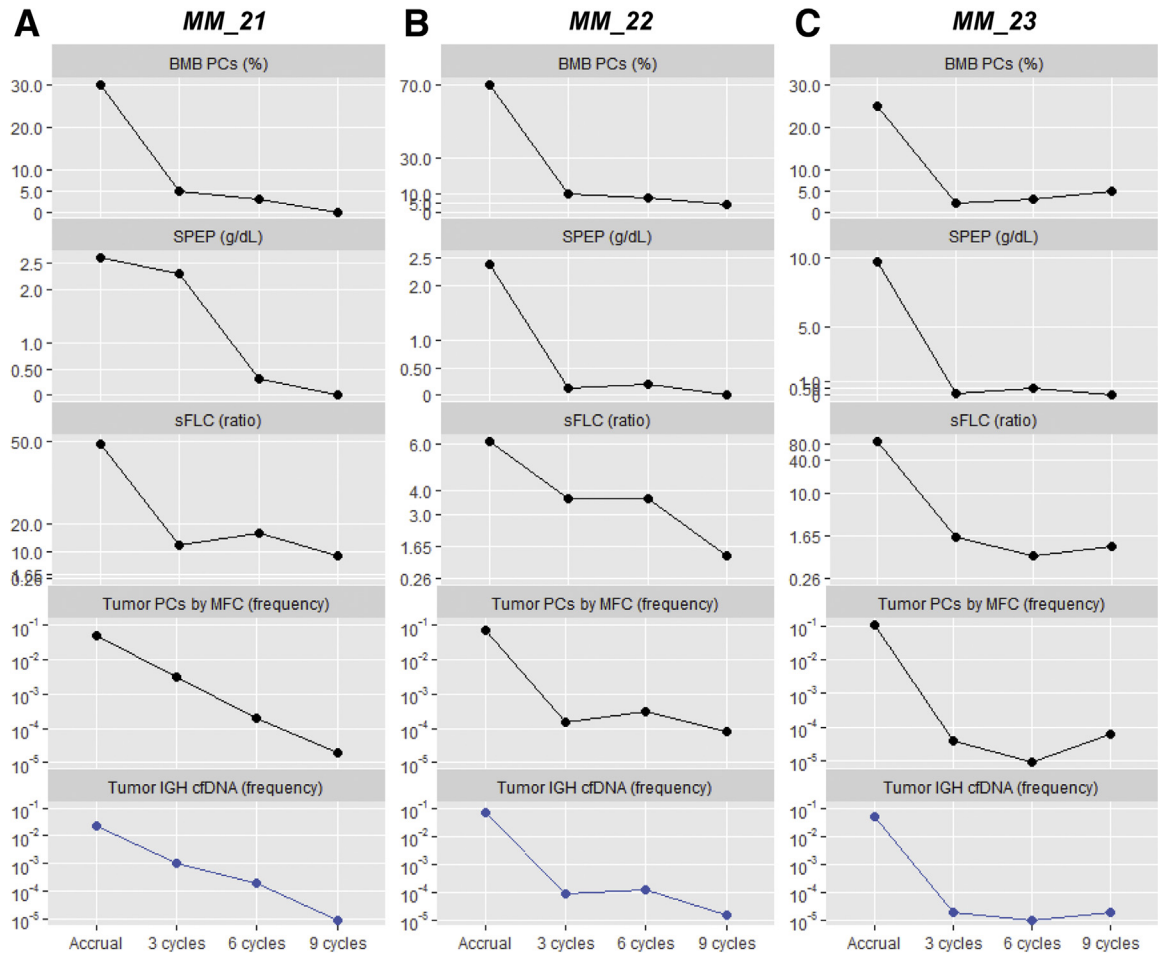


Figure 4 Quantification of *IGH* cell-free DNA (cfDNA) levels in relation to multiple myeloma (MM) clinical indexes. Following a top-down disposition, the *y* axes of the line plots are respectively related to percentage of bone marrow plasma cells on trephine biopsies (BMB PCs) (linear scale), M-protein concentrations evaluated by serum electrophoresis (SPEP) (linear scale), serum-free light chain (sFLC ratio) (linear scale for **A** and **B**, $\log_2$ scale for **C**; reference range, 0.26 to 1.65); bone marrow PCs enumerated by multiparameter flow cytometry (MFC) (tumor PCs by MFC, frequency, $\log_{10}$ scale); and plasma *IGH* cfDNA levels (tumor IGH cfDNA, frequency, log10 scale). Time points are labeled on the *x* axis. **A:** Case MM_21: plasma *IGH* cfDNA levels decrease during treatment, paralleling the decrease of BMB PCs and SPEP levels, which become undetectable after nine cycles of treatment. This finding is consistent with the MFC trend and contrasts with the sFLC ratio, which is stable during treatment. **B:** Case MM_22: Plasma *IGH* cfDNA levels decrease rapidly after the initial cycles of therapy, stabilize from cycles 3 to 6, and then decrease again after the sixth cycle. These dynamic changes are reflected by SPEP, sFLC, and MFC. BMB PCs levels instead steadily decrease during treatment. **C:** Case MM_23: Levels of *IGH* cfDNA decrease after three cycles of therapy and become stable and present a slight increase after nine cycles, mirroring specifically the trend of sFLC and MFC and also of BMB PCs. In this case, the SPEP levels contrast as they become undetectable at the end of therapy.

that can serve as molecular markers for tumor cells. Standardized primers developed by us[12,31] and by the BIOMED-2 concerted action[13] were used to amplify all *IGH* sequences in patients' tumor samples. Libraries were sequenced on an Ion PGM bench-top system and data analyzed with a set of web-based specific bioinformatic tools openly available. *IGH V(D)J* gene segment use in our cohort of patients with MM resembles that seen by other studies[22,23] and supports the idea that the approach used can be applied for identification, characterization, and quantification of myeloma-associated *IGH* clonotype in tumor samples.

Clonotypic *IGH* sequences could be determined from baseline tumor cells in 88% of patients with MM at first relapse, and results were consistent with those obtained by PCR and Sanger sequencing, indicating a specificity similar to that reported using commercial sequencing approaches.[4] Consistently, in all patients with the characterized *IGH* gene rearrangement in tumor samples, the NGS approach was able to detect the same rearrangement in plasma samples. A predominant clonotype was detected among other clones likely representative of normal B-cells. In contrast, in healthy donors no clonally related sequences were observed, most likely reflecting the enormous diversity of *IGH V(D)J* combinations generated during B-cell development.

A significant advantage of cfDNA analysis in the care of patients with MM would be the ability to measure tumor burden by avoiding invasive BMBs. Quantification of BM PCs is crucial for diagnosis and prognostication, as reemphasized in the revised International Myeloma Working

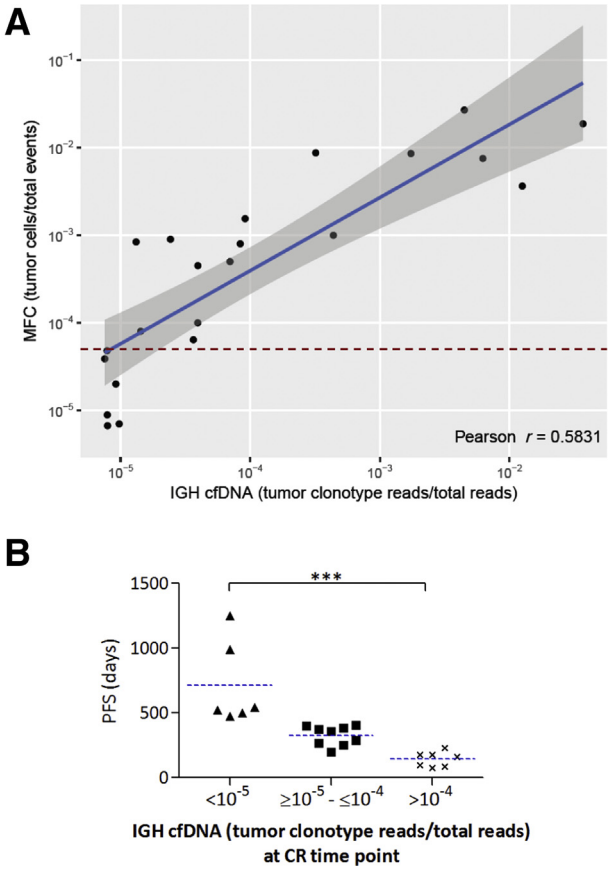


Figure 5 Comparison between *IGH* cell-free DNA (cfDNA) sequencing and multiparameter flow cytometry (MFC) data. **A:** Scatterplot showing correlation between *IGH* cfDNA sequencing (IGH cfDNA, $\log_{10}$ on the *x* axis) and eight-color flow cytometry (MFC, $\log_{10}$ on the *y* axis) in assessing residual disease in the selected 22 patients (black dots). The frequencies of *IGH* cfDNA (*x* axis) are expressed as the number of patient-specific tumor clonotype sequencing reads over the total sequencing filtered reads. The frequencies of bone marrow (BM) multiple myeloma (MM) plasma cells (PCs) detected by MFC are expressed as the number of cells bearing the aberrant immune-phenotype (detected at study entry) over the total collected events. **Blue line** represents the best-fit line obtained using the linear method; shaded area indicates the 95% CI of the best-fit line. Pearson's correlation test was used to quantify the strength of the association ($P = 0.0044$). **Red dashed line** set the sensitivity threshold for MFC (frequency $<5 \times 10^{-5}$). **B:** Group analysis shows that levels of disease detected by *IGH* cfDNA sequencing are significantly correlated with progression-free survival (PFS). **Dashed blue lines** represent the mean value of each category. The strength of the correlation (*P*) was calculated using one-way analysis of variance with the posthoc Tukey honestly significant difference test. $n = 6$ patients with frequencies $<10^{-5}$; $n = 9$ patients with frequencies $\geq 10^{-5}$-$\leq 10^{-4}$; $n = 7$ patients with frequencies $>10^{-4}$. ***$P < 0.001$. CR, complete response.

Group guidelines.[32] However, the degree of BM infiltration by conventional morphologic or immunohistochemical analysis may vary significantly not only among patients but also within the same patient. This has been attributed to the patchy pattern of PCs infiltration, a factor that could also explain the inconsistency of PC counts on BM aspirates as a prognostic factor.[33] In the present study, we demonstrate the clinical implication of the detection of cfDNA: patients with MM with a high cfDNA level of the tumor-associated *IGH* sequence before treatment have a shorter PFS than the others. This finding suggests that cfDNA levels might indeed reflect tumor burden and represent potentially relevant prognostic biomarkers at study entry. This strategy could complement current methods used to determine tumor levels that are limited by invasive biopsies and suboptimal results attributable to sampling variability.

Molecular disease monitoring using cfDNA has a clear clinical relevance even during therapy and might also complement the longitudinal assessment of BM PCs counts, serum monoclonal proteins, or sFLC concentration for response evaluation. In fact, our findings reveal the ability of cfDNA analysis to track tumor kinetics as already shown in lymphoma[5] and suggest that cfDNA may be further useful in cases where disease is predominantly confined to the tissues (eg, solitary plasmacytomas, extramedullary and nonsecretory myeloma).

The fact that cfDNA levels well mirror results obtained with MFC prompted us to evaluate whether such analysis may be relevant even in the MRD context. The current gold standard for MRD monitoring in MM involves immunophenotyping using flow cytometry on BM samples to detect the residual MM cells bearing an aberrant phenotype.[34,35] The use of a PB-based assay for MRD analysis has numerous advantages over any method relying on invasive BMBs.[36] In our study, results obtained analyzing MRD by MFC on BM samples using consensus antibody panels[37] showed complete concordance with cfDNA analysis in all cases. Of note, patients defined as MRD negative by MFC ($<5 \times 10^{-5}$ BM PCs) were also characterized by very low frequencies of the tumor-associated clonotype in plasma samples ($<10^{-5}$) and had a significantly longer PFS than patients with $>10^{-5}$ BM PCs.

Several studies have shown that NGS of the *IGH* represents an effective and highly promising tool for MRD detection in BM samples of patients with lymphoma and MM with sensitivity that may approach $\geq 10^{-6}$.[4,37] Sensitivity is a critical aspect in MRD detection and is limited by the number of input cells or DNA. However, the sensitivity that can be obtained using cfDNA is still unknown because the origin of cfDNA has yet to be fully understood. In this study, our sequencing approach on cfDNA allowed us to analyze at least 5×10^5 reads among which the clonal rearrangement was searched for. Increasing the amount of cfDNA will in the future serve to further improve the theoretical sensitivity. Another reasonable option would be to focus on shorter *IGH* fragments using the FR3 primer set. In cfDNA, because the most abundant DNA fragments are 160 bp long (length of DNA on one nucleosome), one could imagine that shorter amplificates (100 to 170 bp) would be more suitable and that such an approach would reach higher sensitivity than restricting the analysis to less abundant longer DNA molecules. On the other hand, as comparable results were obtained using FR3 primers in cfDNA of patients and healthy donors, this study was conducted

analyzing FR1 amplified fragments that would potentially provide additional sequence information relevant to monitor clonal dynamics. A formal comparison between sensitivity of FR1- and FR3-based strategies on cfDNA is still lacking. In addition, it still remains to be determined whether enhanced sensitivity will provide supplementary clinical information in the context of current therapy protocols for MM.[38]

In summary, results of this pilot study support the clinical applicability and utility of quantifying tumor levels by our deep sequencing of *IGH* gene rearrangements present in plasma of patients with MM. This method can be implemented in any laboratory with NGS capability, can be applied to most patients with MM with a short turnaround time, and can be exploitable for the study of MRD.

Acknowledgments

We thank Anisa Bermema, Vittorio Montefusco, and Niccolò Bolli for their assistance in data acquisition and many useful discussions on this research.

C.C. and P. C. conceived and designed the study; G.B., S.G., and A.V. performed experiments and collected data; G.B., S.G., C.C., and P.C. analyzed and interpreted data; GB, S.G., C.C., and P.C. wrote and edited the manuscript.

Supplemental Data

Supplemental material for this article can be found at *https://doi.org/10.1016/j.jmoldx.2018.07.006*.

References

1. Kumar SK, Rajkumar SV, Dispenzieri A, Lacy MQ, Hayman SR, Buadi FK, Zeldenrust SR, Dingli D, Russell SJ, Lust JA, Greipp PR, Kyle RA, Gertz MA: Improved survival in multiple myeloma and the impact of novel therapies. Blood 2008, 111:2516–2520
2. Anderson KC: The 39th David A. Karnofsky lecture: bench-to-bedside translation of targeted therapies in multiple myeloma. J Clin Oncol 2012, 30:445–452
3. Mailankody S, Korde N, Lesokhin AM, Lendvai N, Hassoun H, Stetler-Stevenson M, Landgren O: Minimal residual disease in multiple myeloma: bringing the bench to the bedside. Nat Rev Clin Oncol 2015, 12:1–10
4. Martinez-Lopez J, Lahuerta JJ, Pepin F, González M, Barrio S, Ayala R, Puig N, Montalban MA, Paiva B, Weng L, Jiménez C, Sopena M, Moorhead M, Cedena T, Rapado I, Victoria Mateos M, Rosiñol L, Oriol A, Blanchard MJ, Martínez R, Bladé J, Miguel JS, Faham M, García-Sanz R: Prognostic value of deep sequencing method for minimal residual disease detection in multiple myeloma. Blood 2014, 123:3073–3079
5. Roschewski M, Dunleavy K, Pittaluga S, Moorhead M, Pepin F, Kong K, Shovlin M, Jaffe ES, Staudt LM, Lai C, Steinberg SM, Chen CC, Zheng J, Willis TD, Faham M, Wilson WH: Circulating tumour DNA and CT monitoring in patients with untreated diffuse large B-cell lymphoma: a correlative biomarker study. Lancet Oncol 2015, 16:541–549
6. Roschewski M, Staudt LM, Wilson WH: Dynamic monitoring of circulating tumor DNA in non-Hodgkin lymphoma. Blood 2016, 127:3127–3133
7. Scherer F, Kurtz DM, Newman AM, Stehr H, Craig AFM, Esfahani MS, Lovejoy AF, Chabon JJ, Klass DM, Liu CL, Zhou L, Glover C, Visser BC, Poultsides GA, Advani RH, Maeda LS, Gupta NK, Levy R, Ohgami RS, Kunder CA, Diehn M, Alizadeh AA: Distinct biological subtypes and patterns of genome evolution in lymphoma revealed by circulating tumor DNA. Sci Transl Med 2016, 8:364ra155
8. Oberle A, Brandt A, Voigtlaender M, Thiele B, Radloff J, Schulenkorf A, Alawi M, Akyüz N, März M, Ford CT, Krohn-Grimberghe A, Binder M: Monitoring multiple myeloma by next-generation sequencing of V(D)J rearrangements from circulating myeloma cells and cell-free myeloma DNA. Haematologica 2017, 102:1105–1111
9. Durie BGM, Harousseau J-L, Miguel JS, Bladé J, Barlogie B, Anderson K, Gertz M, Dimopoulos M, Westin J, Sonneveld P, Ludwig H, Gahrton G, Beksac M, Crowley J, Belch A, Boccadaro M, Cavo M, Turesson I, Joshua D, Vesole D, Kyle R, Alexanian R, Tricot G, Attal M, Merlini G, Powles R, Richardson P, Shimizu K, Tosi P, Morgan G, Rajkumar SV; International Myeloma Working Group: International uniform response criteria for multiple myeloma. Leukemia 2006, 20:1467–1473
10. El Messaoudi S, Rolet F, Mouliere F, Thierry AR: Circulating cell free DNA: preanalytical considerations. Clin Chim Acta 2013, 424:222–230
11. van Dongen JJM, Lhermitte L, Böttcher S, Almeida J, Van Der Velden VHJ, Flores-Montero J, Rawstron A, Asnafi V, Lécrevisse Q, Lucio P, Mejstrikova E, Szczepaski T, Kalina T, De Tute R, Brüggemann M, Sedek L, Cullen M, Langerak AW, Mendonça A, MacIntyre E, Martin-Ayuso M, Hrusak O, Vidriales MB, Orfao A: EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes. Leukemia 2012, 26:1908–1975
12. Voena C, Ladetto M, Astolfi M, Provan D, Gribben JG, Boccadoro M, Pileri A, Corradini P: A novel nested-PCR strategy for the detection of rearranged immunoglobulin heavy-chain genes in B cell tumors. Leukemia 1997, 11:1793–1798
13. van Dongen JJM, Langerak AW, Brüggemann M, Evans PAS, Hummel M, Lavender FL, Delabesse E, Davi F, Schuuring E, García-Sanz R, van Krieken JHJM, Droese J, González D, Bastard C, White HE, Spaargaren M, González M, Parreira A, Smith JL, Morgan GJ, Kneba M, Macintyre EA: Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia 2003, 17:2257–2317
14. Alamyar E, Giudicelli V, Li S, Duroux P, Lefranc M-P: IMGT/HighV-QUEST: the IGMGT web portal for immunoglobulin (Ig) or antibody and T cell receptor (TC) analyses from NGS high throughout and deep sequencing. Immunome Res 2012, 8:1–15
15. Shugay M, Bagaev DV, Turchaninova MA, Bolotin DA, Britanova OV, Putintseva EV, Pogorelyy MV, Nazarov VI, Zvyagin IV, Kirgizova VI, Kirgizov KI, Skorobogatova EV, Chudakov DM: VDJtools: unifying post-analysis of T cell receptor repertoires. PLoS Comput Biol 2015, 11:1–16
16. R Development Core Team R: a language and environment for statistical computing. R Found Stat Comput 2011, 1:409
17. Corradini P, Cavo M, Lokhorst H, Martinelli G, Terragna C, Majolino I, Valagussa P, Boccadoro M, Samson D, Bacigalupo A, Russell N, Montefusco V, Voena C, Gahrton G: Molecular remission after myeloablative allogeneic stem cell transplantation predicts a better relapse-free survival in patients with multiple myeloma. Blood 2003, 102:1927–1929
18. Bashford-Rogers RJM, Palser AL, Huntly BJ, Rance R, Vassiliou GS, Follows GA, Kellam P: Network properties derived from deep sequencing of human b-cell receptor repertoires delineate B-cell populations. Genome Res 2013, 23:1874–1884

19. McClure R, Mai M, McClure S: High-throughput sequencing using the Ion Torrent Personal Genome Machine for clinical evaluation of somatic hypermutation status in chronic lymphocytic leukemia. J Mol Diagn 2015, 17:145−154
20. Herrera AF, Kim HT, Kong KA, Faham M, Sun H, Sohani AR, Alyea EP, Carlton VE, Chen YB, Cutler CS, Ho VT, Koreth J, Kotwaliwale C, Nikiforow S, Ritz J, Rodig SJ, Soiffer RJ, Antin JH, Armand P: Next-generation sequencing-based detection of circulating tumour DNA After allogeneic stem cell transplantation for lymphoma. Br J Haematol 2016, 175:841−850
21. Shin S, Hwang IS, Kim J, Lee KA, Lee ST, Choi JR: Detection of immunoglobulin heavy chain gene clonality by next-generation sequencing for minimal residual disease monitoring in B-lymphoblastic leukemia. Ann Lab Med 2017, 37:331−335
22. Ferrero S, Capello D, Svaldi M, Boi M, Gatti D, Drandi D, Rossi D, Barbiero S, Mantoan B, Mantella E, Zanni M, Ghione P, Larocca A, Passera R, Bertoni F, Gattei V, Forconi F, Laurenti L, del Poeta G, Marasca R, Cortelazzo S, Gaidano G, Palumbo A, Boccadoro M, Ladetto M: Multiple myeloma shows no intra-disease clustering of immunoglobulin heavy chain genes. Haematologica 2012, 97:849−853
23. González D, González M, Balanzategui A, Sarasquete ME, López-Pérez R, Chillón MC, García-Sanz R, San Miguel JF: Molecular characteristics and gene segment usage in IGH gene rearrangements in multiple myeloma. Haematologica 2005, 90:906−913
24. Brochet X, Lefranc MP, Giudicelli V: IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucleic Acids Res 2008, 36:503−508
25. Lee N, Moon SY, Lee J, Park H-K, Kong S-Y, Bang S-M, Lee JH, Yoon S-S, Lee DS: Discrepancies between the percentage of plasma cells in bone marrow aspiration and BM biopsy: impact on the revised IMWG diagnostic criteria of multiple myeloma. Blood Cancer J 2017, 7:e530
26. Palumbo A, Avet-Loiseau H, Oliva S, Lokhorst HM, Goldschmidt H, Rosinol L, Richardson P, Caltagirone S, Lahuerta JJ, Facon T, Bringhen S, Gay F, Attal M, Passera R, Spencer A, Offidani M, Kumar S, Musto P, Lonial S, Petrucci MT, Orlowski RZ, Zamagni E, Morgan G, Dimopoulos MA, Durie BGM, Anderson KC, Sonneveld P, Miguel JS, Cavo M, Rajkumar SV, Moreau P: Revised international staging system for multiple myeloma: a report from international myeloma working group. J Clin Oncol 2015, 33:2863−2869
27. Sonneveld P, Avet-Loiseau H, Lonial S, Usmani S, Siegel D, Anderson KC, Chng WJ, Moreau P, Attal M, Kyle RA, Caers J, Hillengass J, Miguel JS, Van De Donk NWCJ, Einsele H, Bladé J, Durie BGM, Goldschmidt H, Mateos MV, Palumbo A, Orlowski R: Treatment of multiple myeloma with high-risk cytogenetics: a consensus of the International Myeloma Working Group. Blood 2016, 127:2955−2962
28. Flanders A, Stetler-Stevenson M, Landgren O: Minimal residual disease testing in multiple myeloma by flow cytometry: major determogeneity. Blood 2013, 122:1088−1089
29. Bohers E, Viailly PJ, Dubois S, Bertrand P, Maingonnat C, Mareschal S, Ruminy P, Picquenot JM, Bastard C, Desmots F, Fest T, Leroy K, Tilly H, Jardin F: Somatic mutations of cell-free circulating DNA detected by next-generation sequencing reflect the genetic changes in both germinal center B-cell-like and activated B-cell-like diffuse large B-cell lymphomas at the time of diagnosis. Haematologica 2015, 100:e280−e284
30. Hocking J, Mithraprabhu S, Kalff A, Spencer A: Liquid biopsies for liquid tumors: emerging potential of circulating free nucleic acid evaluation for the management of hematologic malignancies. Cancer Biol Med 2016, 13:215−225
31. Farina L, Carniti C, Dodero A, Vendramin A, Raganato A, Spina F, Patriarca F, Narni F, Benedetti F, Olivieri A, Corradini P: Qualitative and quantitative polymerase chain reaction monitoring of minimal residual disease in relapsed chronic lymphocytic leukemia: early assessment can predict long-term outcome after reduced intensity allogeneic transplantation. Haematologica 2009, 94:654−662
32. Rajkumar SV, Dimopoulos MA, Palumbo A, Blade J, Merlini G, Mateos MV, Kumar S, Hillengass J, Kastritis E, Richardson P, Landgren O, Paiva B, Dispenzieri A, Weiss B, LeLeu X, Zweegman S, Lonial S, Rosinol L, Zamagni E, Jagannath S, Sezer O, Kristinsson SY, Caers J, Usmani SZ, Lahuerta JJ, Johnsen HE, Beksac M, Cavo M, Goldschmidt H, Terpos E, Kyle RA, Anderson KC, Durie BGM, Miguel JFS: International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. Lancet Oncol 2014, 15:e538−e548
33. Paiva B, Vidriales MB, Pérez JJ, Mateo G, Montalbán MA, Mateos MV, Bladé J, Lahuerta JJ, Orfao A, San Miguel JF: Multiparameter flow cytometry quantification of bone marrow plasma cells at diagnosis provides more prognostic information than morphological assessment in myeloma patients. Haematologica 2009, 94:1599−1602
34. Rawstron AC, Child JA, de Tute RM, Davies FE, Gregory WM, Bell SE, Szubert AJ, Navarro-Coy N, Drayson MT, Feyler S, Ross FM, Cook G, Jackson GH, Morgan GJ, Owen RG: Minimal residual disease assessed by multiparameter flow cytometry in multiple myeloma: impact on outcome in the Medical Research Council Myeloma IX Study. J Clin Oncol 2013, 31:2540−2547
35. Paiva B, van Dongen JJM, Orfao A: New criteria for response assessment: role of minimal residual disease in multiple myeloma. Blood 2015, 125:3059−3068
36. Flores-Montero J, Sanoja-Flores L, Paiva B, Puig N, García-Sánchez O, Böttcher S, Van Der Velden VHJ, Pérez-Morán JJ, Vidriales MB, García-Sanz R, Jimenez C, González M, Martínez-López J, Corral-Mateos A, Grigore GE, Fluxá R, Pontes R, Caetano J, Sedek L, Del Cañizo MC, Bladé J, Lahuerta JJ, Aguilar C, Bárez A, García-Mateo A, Labrador J, Leoz P, Aguilera-Sanz C, San-Miguel J, Mateos MV, Durie B, Van Dongen JJM, Orfao A: Next generation flow for highly sensitive and standardized detection of minimal residual disease in multiple myeloma. Leukemia 2017, 31:2094−2103
37. Nishihori T, Song J, Shain KH: Minimal residual disease assessment in the context of multiple myeloma treatment. Curr Hematol Malig Rep 2016, 11:118−126
38. Moreau P, Zamagni E: MRD in multiple myeloma: more questions than answers? Blood Cancer J 2017, 7:7−10

Leukemia (2018) 32:2604–2616
https://doi.org/10.1038/s41375-018-0037-9

ARTICLE

Multiple myeloma gammopathies

Analysis of the genomic landscape of multiple myeloma highlights novel prognostic markers and disease subgroups

Niccolo Bolli[1,2,3] · Giulia Biancon[1] · Matahi Moarii[4] · Silvia Gimondi[1] · Yilong Li[3] · Chiara de Philippis[1,2] · Francesco Maura[1] · Vijitha Sathiaseelan[3] · Yu-Tzu Tai[5] · Laura Mudie[3] · Sarah O'Meara[3] · Keiran Raine[3] · Jon W. Teague[3] · Adam P. Butler[3] · Cristiana Carniti[2] · Moritz Gerstung[6] · Tina Bagratuni[7] · Efstathios Kastritis[7] · Meletios Dimopoulos[7] · Paolo Corradini[1,2] · Kenneth C. Anderson[5] · Philippe Moreau[8] · Stephane Minvielle[8,9] · Peter J. Campbell[3] · Elli Papaemmanuil[4] · Herve Avet-Loiseau[10,11] · Nikhil C. Munshi[5]

Received: 18 August 2017 / Revised: 28 October 2017 / Accepted: 10 November 2017 / Published online: 22 May 2018

Abstract

In multiple myeloma, next-generation sequencing (NGS) has expanded our knowledge of genomic lesions, and highlighted a dynamic and heterogeneous composition of the tumor. Here we used NGS to characterize the genomic landscape of 418 multiple myeloma cases at diagnosis and correlate this with prognosis and classification. Translocations and copy number abnormalities (CNAs) had a preponderant contribution over gene mutations in defining the genotype and prognosis of each case. Known and novel independent prognostic markers were identified in our cohort of proteasome inhibitor and immunomodulatory drug-treated patients with long follow-up, including events with context-specific prognostic value, such as deletions of the *PRDM1* gene. Taking advantage of the comprehensive genomic annotation of each case, we used innovative statistical approaches to identify potential novel myeloma subgroups. We observed clusters of patients stratified based on the overall number of mutations and number/type of CNAs, with distinct effects on survival, suggesting that extended genotype of multiple myeloma at diagnosis may lead to improved disease classification and prognostication.

Key points

1. Next-generation sequencing allows analysis of the integrated spectrum of gene mutations, aneuploidies and IGH translocations in multiple myeloma.
2. Karyotypic events have a stronger impact on prognosis than mutations, but extended genotyping shows novel prognostic categories.

Electronic supplementary material The online version of this article (https://doi.org/10.1038/s41375-018-0037-9) contains supplementary material, which is available to authorized users.

✉ Nikhil C. Munshi
Nikhil_Munshi@dfci.harvard.edu

[1] Department of Oncology and Onco-Hematology, University of Milan, Milan, Italy

[2] Department of Medical Oncology and Hematology, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy

[3] Cancer Genome Project, Wellcome Trust Sanger Institute, Cambridge, UK

[4] Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[5] Harvard Medical School, LeBow Institute for Myeloma Therapeutics and Jerome Lipper Center for Multiple Myeloma Research, Dana-Farber Cancer Institute, Boston, MA, USA

[6] European Bioinformatics Institute, Computational and Cancer Biology, Cambridge, UK

[7] Department of Clinical Therapeutics, National and Kapodistrian University of Athens, Athens, Greece

[8] Department of Hematology, University Hospital Hôtel-Dieu, Nantes, France

[9] CRCINA, INSERM, CNRS, Université d'Angers, Université de Nantes, Nantes, France

[10] Institute Universitaire du Cancer de Toulouse Oncopole, Toulouse, France

[11] University Hospital, Toulouse, France

Introduction

Multiple myeloma (MM) derives from the neoplastic transformation and proliferation of a post-germinal center B-cell. Karyotypic events are the main drivers of early stages of transformation [1–3], and most MM cases are either hyperdiploid (HDMM) or harbor translocations of the immunoglobulin heavy chain (IGH) locus [4]. In addition, several additional recurrent translocations and copy number abnormalities (CNAs) can be found [5], although at diagnosis, their analysis is usually limited to events that have an established prognostic role and may guide treatment: del17p, t(4;14) and t(14;16) [6, 7]. Gene mutations are thought to be secondary events associated with tumor progression rather than initiation [2, 8]. Several next-generation sequencing (NGS) studies to date have expanded our knowledge on the spectrum of gene mutations in MM [8–13]. However, such studies have also highlighted substantial heterogeneity and low recurrence rates as compared to other hematological malignancies. Furthermore, the integrated spectrum of gene mutations, rearrangements and CNAs of each case is composed of alterations that may belong to different subclones, often in dynamic and differential evolution during the various stages of disease [14]. NGS technologies are ideally suited to return information on those driver events, whose analysis and correlation with clinical and laboratory features of the patient can impact prognosis and disease classification. Initial efforts based on whole exome sequencing (WES) have shown that integration of certain genomic lesions into standard risk models can improve prognostication in MM [13]. However, the full potential of NGS studies has not been exploited so far. Custom target pulldown (TPD) has significant advantages over WES in that it allows analysis of a limited fraction of the genome, thus reducing the cost and complexity of downstream analysis per sample. The success of this methodology in detecting gene mutations at diagnosis or relapse has been illustrated in myeloid malignancies [15–18] and MM [19–21]. We have previously shown that TPD panels can be designed to interrogate the integrated spectrum of gene mutations, CNAs and IGH translocations in MM, thus providing a potential one-stop platform for prognostication of newly diagnosed MM cases [22]. Here we applied our custom TPD to a large cohort of multiple myeloma samples at diagnosis to understand MM genomic and clinical interrelationships. We report our unbiased analysis of a large set of MM driver events, leading to the identification of novel prognostic factors and significant interactions between genomic events, and suggesting that analysis of the extended myeloma genotype in larger cohorts can identify novel subgroups that are biologically and clinically distinct.

Methods

Sequencing and identification of mutations

In total 10 ng of genomic DNA from CD138+ bone marrow cells were subjected to whole-genome amplification. Patient-level tags were added and samples pooled before custom target enrichment and sequencing, which was performed on Illumina HiSeq2000 machines on a paired-end 75 bp protocol. Reads were aligned with BWA-mem and variants called with in-house algorithms [23, 24] using a previously described pipeline to filter out probable artifacts and germline variants, and to rank somatic variants based on their likely oncogenic potential [10, 15].

CNAs, IGH rearrangements and VAF adjustment

Coverage data were used to identify regions of aneuploidies, after normalization to diploid samples. This was performed at the whole-chromosome level, and from there down to cytogenetic bands and gene loci. Hyperdiploid samples were defined by a gain in at least two of the following chromosomes: 3, 5, 7, 9, 11, 15, 19, 21. Translocations were called as previously described [22]. Gene-level CNA information was also used to adjust the variant allelic frequency (VAF) of each variant, to estimate the number of cells bearing a given variant, as previously described [16].

Statistical analysis

Pairwise association studies were performed using Fisher test corrected for multiple hypotheses testing. For survival analysis, both progression-free (PFS) and overall survival (OS) were used as end-points, and log-rank tests were used for univariate hypothesis tests after correction for multiple hypothesis testing. For multivariate survival analyses, sparse Cox regression was performed on the full set of driver events. Further information on methodology can be found in the online supplement.

Results

Patients and sequencing metrics of the study

We used TPD to sequence unmatched genomic DNA from 373 MM patients at diagnosis, and we added 45 patients from a previously published WES study [10] for a total of 418 (Table 1). Mean age was 56.6 years. Most patients (76.3%) received bortezomib-based induction treatment followed by autologous hemopoietic stem cell transplant. First-line bortezomib-treated patients had a younger median

age (55.6 vs 59.8 years, $p = 1 \times 10^{-5}$, Student's *t*-test). Median follow-up was 5.4 years. At the time of analysis, 85.9% of patients relapsed, and 48.3% died.

We sequenced an average of 1.46 Gb per sample to a target region depth of 118.9× (Supplementary Figure 1). The average on-target efficiency was 34.3%, including the IGH locus that shows extensive deletions in plasma cells, leading to underestimation of the actual on-target efficiency [22]. Although coverage was lower than our previous WES study (Supplementary Figure 2A), the distribution of gene mutations did not show any significant differences (Supplementary Figure 2B), suggesting that the overall performance of our TPD was comparable to the previous WES study.

Analysis of gene mutations in multiple myeloma

After excluding artifacts and likely germline single-nucleotide polyorphisms, we identified 2269 high-confidence, likely somatic variants in 215 of the 246 genes included in the design in 412 out of the 418 patients analyzed (Supplementary Table S1). We then compared the expected pattern of mutations in each given gene derived from the literature—typically inactivating mutations for tumor suppressor genes and hot-spot mutations for oncogenes—to the pattern observed in our series to triage each variant into "oncogenic", "possible oncogenic", or "unknown" classes (see Supplemental Methods for further details). Looking at oncogenic mutations only, we found 695 variants in 106 genes in 342 patients (Fig. 1a). Numbers increased to 1250 variants in 177 genes in 395 patients with the addition of possible oncogenic variants. At least one oncogenic or possible oncogenic variant was found in 94.5% of patients, with a median of two per patient. The list of most commonly mutated genes closely recapitulated previously published data (Fig. 1b, top) [9–13]. Clustered missense mutations were prevalent in known driver oncogenes such as *KRAS*, *NRAS* and *BRAF*, whereas an excess of truncating mutations was found in known tumor suppressors such as *TP53*, *FAM46C*, *SP140* (Fig. 1b, bottom). Most oncogenic mutations (63%) were accounted for by nine of the top driver genes previously identified, i.e., *KRAS, NRAS, TP53, FAM46C, BRAF, DIS3, TRAF3, SP140, IRF4* (Fig. 1b). A mutation in at least one of these nine genes was found in 64% of patients. Conversely, many genes showed a large excess of variants of unknown or possible oncogenic potential. Among these, large genes like *FAT1, FAT3, FAT4, DNAH9, DNAH11, PCLO*, whose role in myeloma pathogenesis remains unclear [10, 13, 25]. Last, we found sporadic oncogenic mutations with potential clinical impact in *CRBN* and *IKZF1*, previously associated with resistance to immunomodulatory drug [19, 26] in <1% of patients each. Our mixed confirmation/discovery effort did therefore not identify novel genes mutated at a significant recurrence rate in MM, but at the same time we showed how the long tail of uncommonly mutated genes contributed a significant fraction of the heterogeneous genomic landscape of MM. Given the possible inclusion of artifacts and/or germline variants in mutations of "unknown" class, we only considered variants of oncogenic or possible oncogenic classes as driver mutations for subsequent analysis.

Table 1 Overall clinical features of cases included in the study

Variable	Baseline distribution in cohort
Sample	
Sample size for sequencing	418
Sample size for outcome data	418
Median (range) follow-up	5.4 (0.1–11.5) years
Demographics	
Sex	
Male	243 (58.1%)
Female	171 (40.9%)
NA	4 (1%)
Age, mean ± SD	56.6 ± 8.4 years
Biochemical	
β2 microglobulin, mean ± SD	7.8 ± 9.6 (mg/L)
MM staging	
ISS I	150 (35.9%)
ISS II	102 (24.4%)
ISS III	86 (20.6 %)
NA	80 (19.1%)
Treatment	
VD-HDM	266 (63.6%)
VTD-HDM	53 (12.7%)
VAD-HDM	67 (16.0%)
MPV/KMP	5 (1.2%)
MPT	10 (2.4%)
RD	5 (1.2%)
SMM	1 (0.2%)
NA	11 (2.6%)
Outcome	
Relapse	359/418 (85.9%)
Death	202/418 (48.3%)

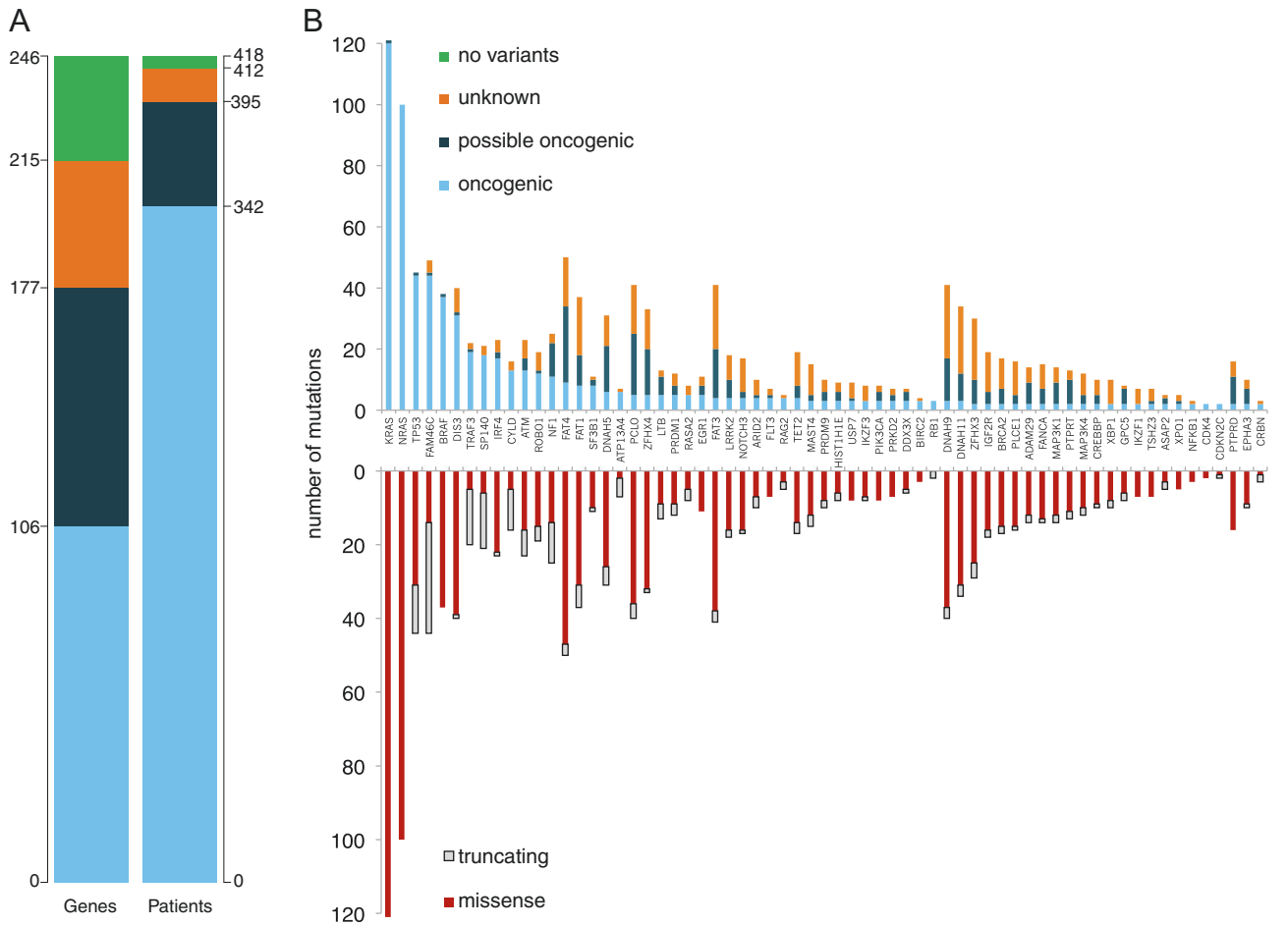


Fig. 1 Absolute number of mutations in the study: **a** Left: absolute numbers of genes with at least one mutation of the specified class found in the study. Right: Number of patients carrying at least one mutation of the specified class. Oncogenic variants: light blue. Possible oncogenic variants: dark blue. Variants of unknown significance: orange. No variants identified: green. **b** Top: stacked bar chart of mutations in the study, limited to genes with >2 oncogenic mutations, broken down by mutation class. Bottom: For the same genes, missense variants are in red, truncating variants are in gray

Overall, 56% of patients displayed activation of the RAS/MAPK pathway, as clustered missense mutations of *KRAS*, *NRAS*, *BRAF*, and truncating mutations of *NF1* and *RASA2*. The NF-kB pathway was recurrently hit as well, with mutations—mostly inactivating—of TRAF3,CYLD and LTB in 12% of patients. Last, mutations or deletions of genes broadly implicated in the DNA damage response (*TP53*, *ATM*, *ATR*, *BRCA2*) were observed in 22% of patients. Although the mutational spectrum was heterogeneous and most genes showed low recurrence rates, frequent involvement of these pathways confirms their functional relevance in MM pathogenesis.

Identification of copy number changes and IGH translocations allows analysis of the integrated genomic landscape of myeloma

With most MM patients displaying prognostically relevant structural abnormalities by karyotype or fluorescence in situ hybridization (FISH) analysis, we sought to derive this information from our NGS data [22]. Recurrent IGH translocations were identified in 135 out of 418 patients. FISH validation for t(4;14) on 119 of them showed our approach had 91% sensitivity and 98% specificity (Supplementary Table S2). We identified 57% of patients as HDMM and 32% as IGH-translocated (Fig. 2a). 9% of patients harbored both an hyperdiploid karyotype and an IGH translocation, and t(4:14) was the most prevalent in this subgroup [27]. We then identified segmental chromosomal aneuploidies with prognostic significance in 61% of patients, at rates comparable to previous studies based on SNP arrays [28]. Furthermore, we identified gene-level gains and losses. Many gene amplifications were concordant with whole-chromosome trisomies in hyperdiploid cases, such as *CCND1* in chromosome 11, prompting exclusion of these events from further analysis. The most frequent gene deletions instead consisted of known tumor suppressors, and for *TP53*, *CYLD*, *SNX7* these coexisted so

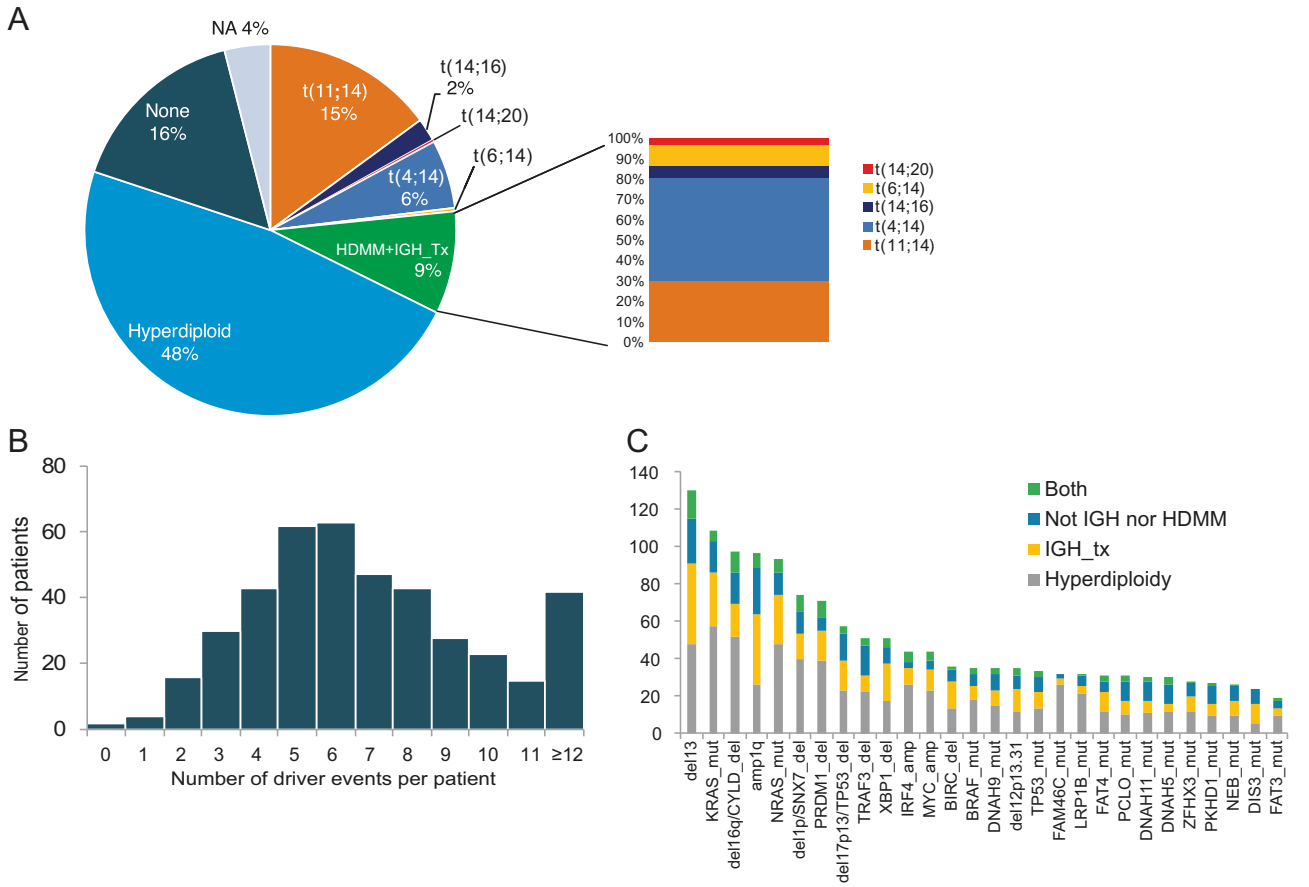


Fig. 2 The genomic landscape of multiple myeloma: **a** Pie chart of the breakdown of samples in the study, classified based on the presumed founder cytogenetic abnormalities: hyperdiploidy and IGH translocations. HDMM hyperdiploid multiple myeloma, NA not available for analysis, None no cytogenetic event identified. IGH_Tx translocation involving the IGH region. **b** Bar chart of the prevalence of each class of alteration (copy number abnormalities -CNA-, karyotype, gene-level copy number abnormalities, gene mutations) in samples in the study. **c** Stacked bar chart of the most frequent genomic events of any class, broken down by cytogenetic group

frequently with larger segmental chromosomal deletions that we aggregated loss of the gene locus and loss of the chromosome segment as one variable for further analysis. After these adjustments, genuine gene-level copy number alterations were found in 51% of patients.

Considering recurrent translocations and aneuploidies, deletions of tumor suppressor genes, amplification of oncogenes, and mutations pertaining to "oncogenic" or "possible oncogenic" classes, at least one such driver event was present in >99% of patients (Supplementary Figure 3A). Overall, a median of 6 events were present in each patient (Fig. 2b). *KRAS* and *NRAS* were the only point mutations present in the 15 most frequent driver events, highlighting how karyotypic events dominate the integrated genomic landscape of MM (Fig. 2c). In addition, we interestingly found frequent deletions in genes required for plasma cell development such as *XBP1* and *PRDM1* (Fig. 2c) that also show recurrent mutations, often truncating, suggesting they may represent novel tumor suppressors in MM (Supplementary Table S1). Hyperdiploid cases showed a small but significantly higher average number of mutations than cases with IGH translocations (2.75 vs 2.5, chi-square test $p = 0.002$) (Supplementary Figure 3B), suggesting that while additional genetic events are required for disease progression, the evolutionary trajectory of each myeloma case may vary based on the initiating event.

Predictive and prognostic value of individual genomic events

We asked whether genomic features were correlated with clinico-laboratoristic features. Using linear models, we found only two associations. 17p deletions presented a modest but significant ($p = 0.001$, likelihood ratio test) correlation with increasing age. Unlike myeloid disorders [15, 16] and clonal hemopoiesis [29–33], here *TP53* mutations were not correlated with age (Supplementary Figure 4A, B), suggesting a different evolutionary trajectory. Furthermore, 1q amplification was significantly correlated with higher beta-2 microglobulin levels ($p = 0.003$,

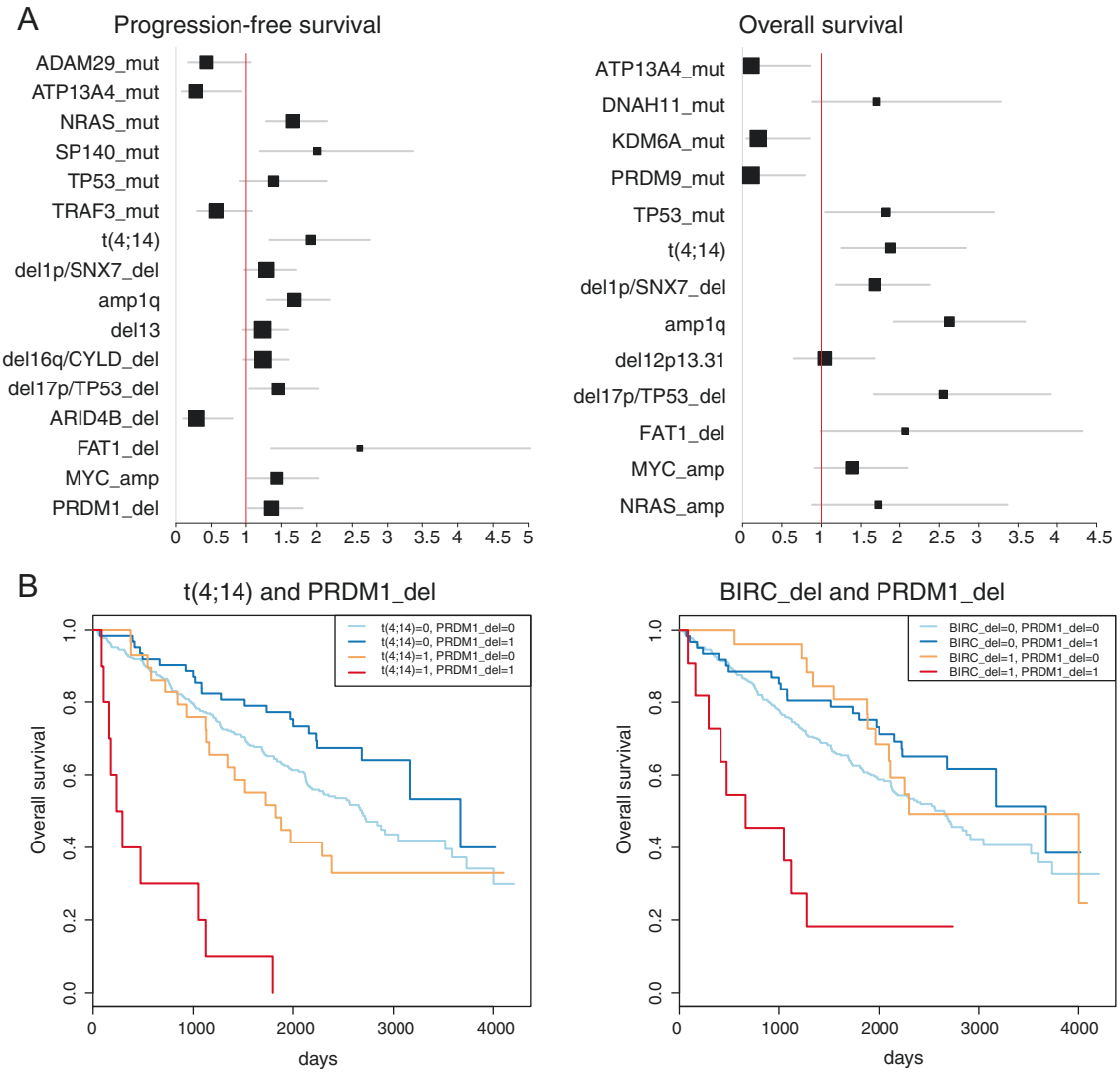


Fig. 3 Independent variants affecting survival: **a** Forest plot of variables implicated in progression-free survival (left) and overall survival (right). Hazard ratio on the *X* axis, values <1 confer better prognosis, values >1 confer worse prognosis. For each variable, the confidence interval is a horizontal gray bar and the hazard ratio—referenced to the *X*-axis—is represented by a black box. **b** Kaplan–Meier survival curves for two significant instances of interaction found in the study: prognosis for overall survival is significantly worse in case of coexistence of *PRDM1* deletions and t(4;14) (left) or *PRDM1* deletions and *BIRC2/3* deletions (right) than with either variable alone

Supplementary Figure 4C). Among disease-independent predictors of survival, older age was associated with poorer survival as previously reported (Cox $p = 0{,}019$, Supplementary Figure 4D) [34]. First line proteasome inhibitor (PI) treatment was associated with improved PFS, and with a trend towards improved overall survival (OS, Supplementary Figure 4E), whose impact was more difficult to evaluate given the availability of PI-based salvage treatment in the non-PI induction cohort. Finally, we found no significant association between genomic lesions and survival for PI-treated patients, including no significant benefit from bortezomib in t(4;14) [7]. The low frequency of most driver events, and the likely modest size of the effect of a given treatment on a driver event, warrant larger cohorts are analyzed to answer this question in a comprehensive way.

We performed a univariate Cox analysis to infer the prognostic value of each genomic feature for PFS and OS (Supplementary Table S3). We found 12 variants significantly associated with poorer PFS and/or OS (Supplementary Figure 5). Interestingly, most gene mutations were not relevant for prognosis. The only mutated gene with a clear prognostic impact on both PFS and OS was *TP53*, while *DNAH11* mutations conferred worse OS only (Log-rank test, $p = 6.9 \times 10^{-4}$ and 7.3×10^{-3} for OS, respectively). Looking at gene-level gains and losses, we found 5 events conferring shorter OS, including losses of TP53/17p, CYLD/16q, *FAT1*, and amplifications of *MYC* and *NRAS*. We could confirm the negative impact on survival of t(4;14) [4] and that of regions of recurrent CNAs [28] (that were selected for their prognostic impact in the first place).

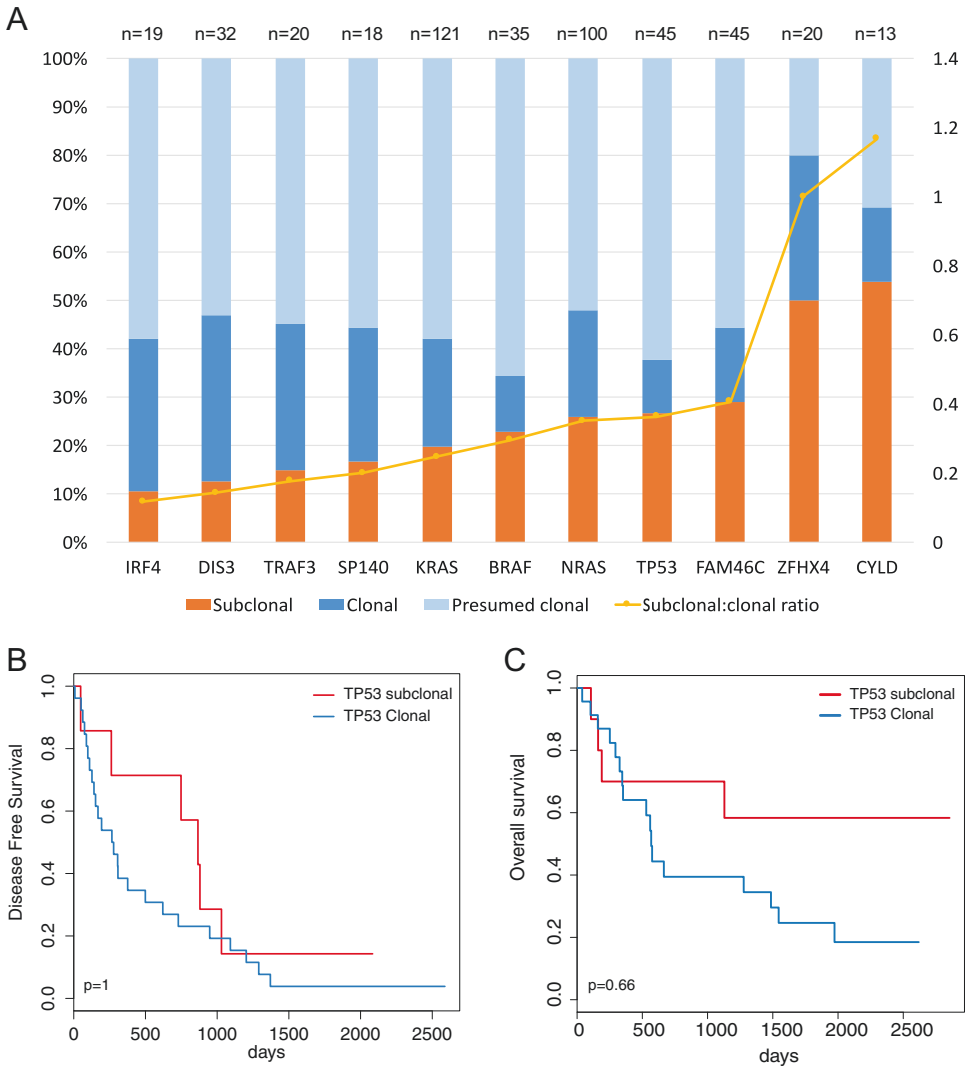


Fig. 4 Subclonal mutations: **a** Stacked bar chart of the clonality status of the top 11 mutated genes in our cohort: bars refer to the left *Y*-axis and plotted is the proportion of variants that are subclonal (orange), clonal (dark blue) and presumed clonal (light blue)—i.e., cases where confidence intervals of aVAF are all overlapping and low tumor purity could lead to over-estimation of clonality. The yellow line refers to the right *Y*-axis and represents, for each gene, the ratio between subclonal and clonal variants, i.e., higher values correspond to genes with more subclonal variants. (**b**)Kaplan–Meier plot of progression-free survival in patients with subclonal (red) or clonal (blue) TP53 mutations with *p*-values from univariate analysis (log-rank test) adjusted for multiple hypotheses testing. **c** Kaplan–Meier plot of overall survival in patients with subclonal (red) or clonal (blue) TP53 mutations with *p*-values from univariate analysis (log-rank test) adjusted for multiple hypotheses testing

Despite this selection bias, the lack of prognostic value of most gene mutations and their overall lower level of recurrence compared to cytogenetic changes underscores the relevance of karyotypic events in shaping the clinical behavior of MM.

Independent prognostic value of driver genomic lesions and their interaction

We next performed a multivariate sparse Cox regression on the full set of genomic variables. For PFS, 18 features had non-zero coefficients, with 9 of them being significant in a second multivariate Cox regression restricted to the first shortlist of 18 variables. Among these, mutations in *ATP13A4* and deletions in *ARID4B* had a favorable impact, whereas mutations in *SP140* and *NRAS*, t(4;14), amp(1q), del(17p13) and deletions of *FAT1* and *PRDM1* had a negative impact (Fig. 3a, left). For the OS regression, 13 features had non-zero coefficients with 7 of them being significant in the restricted analysis: mutations in *ATP13A4, KDM6A* and *PRDM9* had a favorable impact, while t(4;14), amp(1q), del(17p13), del(1p) had a negative impact (Fig. 3a, right). Finally, because our cohort was heterogeneous in demographics and treatment received, we used multivariate linear regression models to exclude a confounding effect of clinical features on the prognostic value of each of these genomic variables, confirming they can be considered as independent prognostic events (Supplementary Figure 6).

Subsequently, we systematically looked for significant interactions between driver events, i.e. combinations where the prognostic impact on OS of one is significantly altered if another is co-occurring. Despite the large set of variables, we found only two significant interactions: cases bearing both t(4;14) and *PRDM1* deletion had a dismal median OS of 265 days (HR 9.05, $p = 3 \times 10^{-4}$ for interaction, Fig. 3b, left); *BIRC2/3* deletion and *PRDM1* deletion conferred a median OS of 666 days (HR 6.3, $p = 0.04$) (Fig. 3b, right).

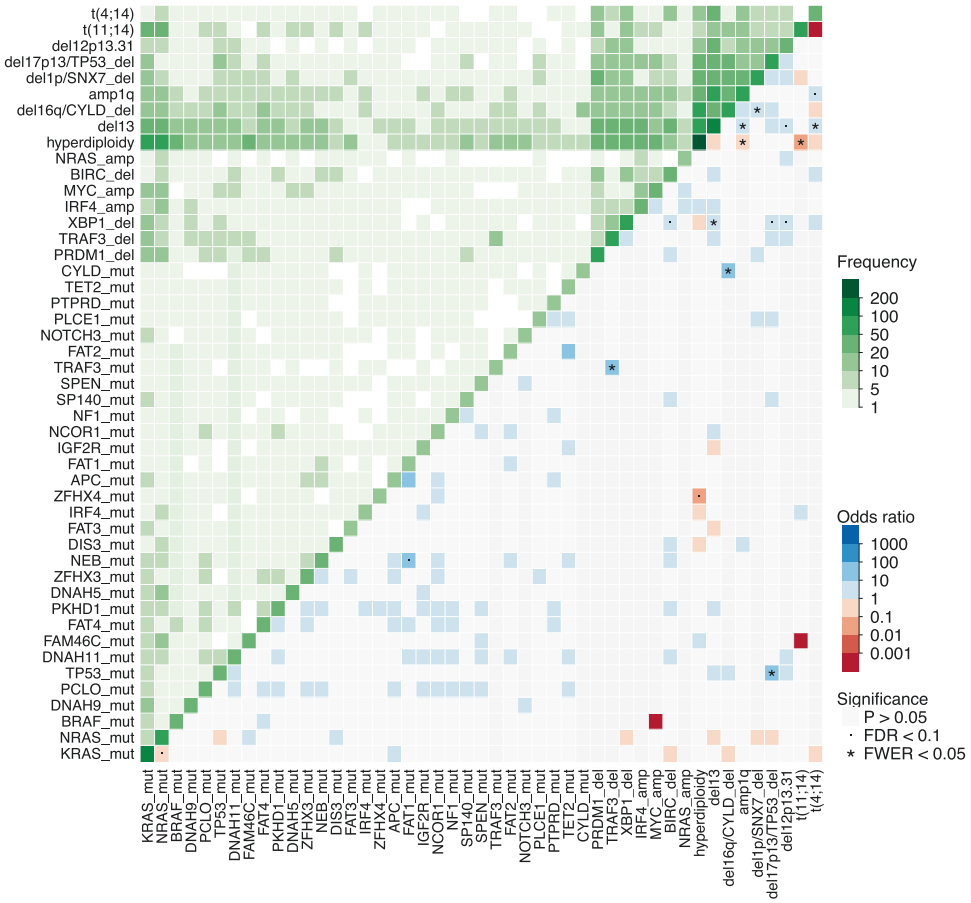


Fig. 5 Pairwise association between variables: Heatmap showing pairwise analysis of occurrence of the most frequent genomic events in MM. The same variable is plotted in the *X* and *Y* axis, and the intensity of color in the leading diagonal indicated the frequency of the variable in the dataset. In the upper triangle, intensity of green indicates the frequency of co-occurrence of any two variables. In the lower triangle, associations are colored by odds ratio: non-significant ones are in gray, while significant ones are in blue if co-occurring, and red if mutually exclusive (p-value < 0.05, fisher test corrected for multiple hypothesis testing). Events with false-discovery rate <0.1 are marked with a dot and events with family-wise error rate <0.05 are marked with a star. While only 47 variables are shown, statistical significance is computed on the full dataset of 192 variables shown in Supplementary Figure 10

When analyzed in isolation, *PRDM1* deletions had no effect (Supplementary Table S3), implying that the prognostic value of driver oncogenic events in MM may differ based on their genomic context.

We also looked for combinations of variables that could be relevant for OS, even if not significant for interaction, across the whole dataset. We first analyzed the association of *TP53* mutations and chr17p deletions, two events that are known to co-occur [10], and the prognosis was slightly worse for the combination (20/418 patients), in a trend suggestive of an additive effect (Supplementary Figure 7A). However, the shortest OS was that of t(4;14) and *TP53* mutations (Supplementary Figure 7B), conferring a dismal 228 day median overall survival to patients bearing both events. HDMM patients bearing an IGH translocation showed a worse OS compared to those without (Supplementary Figure 7C). On the same lines, OS was negatively affected by the total number of driver oncogenic events, irrespective of their nature, in a stepwise decline (Supplementary Figure 7D), mainly supported by karyotypic events (Supplementary Figures 7E-F). This reinforces the notion that myeloma evolution toward more aggressive disease is driven by acquisition of additional driver events.

Clonal and subclonal driver mutations

Adjusted VAF (aVAF) can be used to infer the fraction of cells carrying each mutation, so to time its order of acquisition [15, 16]. Subclonal mutations in known driver genes such as *KRAS*, *NRAS*, *TP53* were identified in 19.8%, 26% and 26.7% of cases, respectively (examples in Supplementary Figure 8A). When ranked based on the fraction of subclonal occurrences, all genes had evidence of both clonal and subclonal mutations, but *CYLD* and *ZFHX4* had a

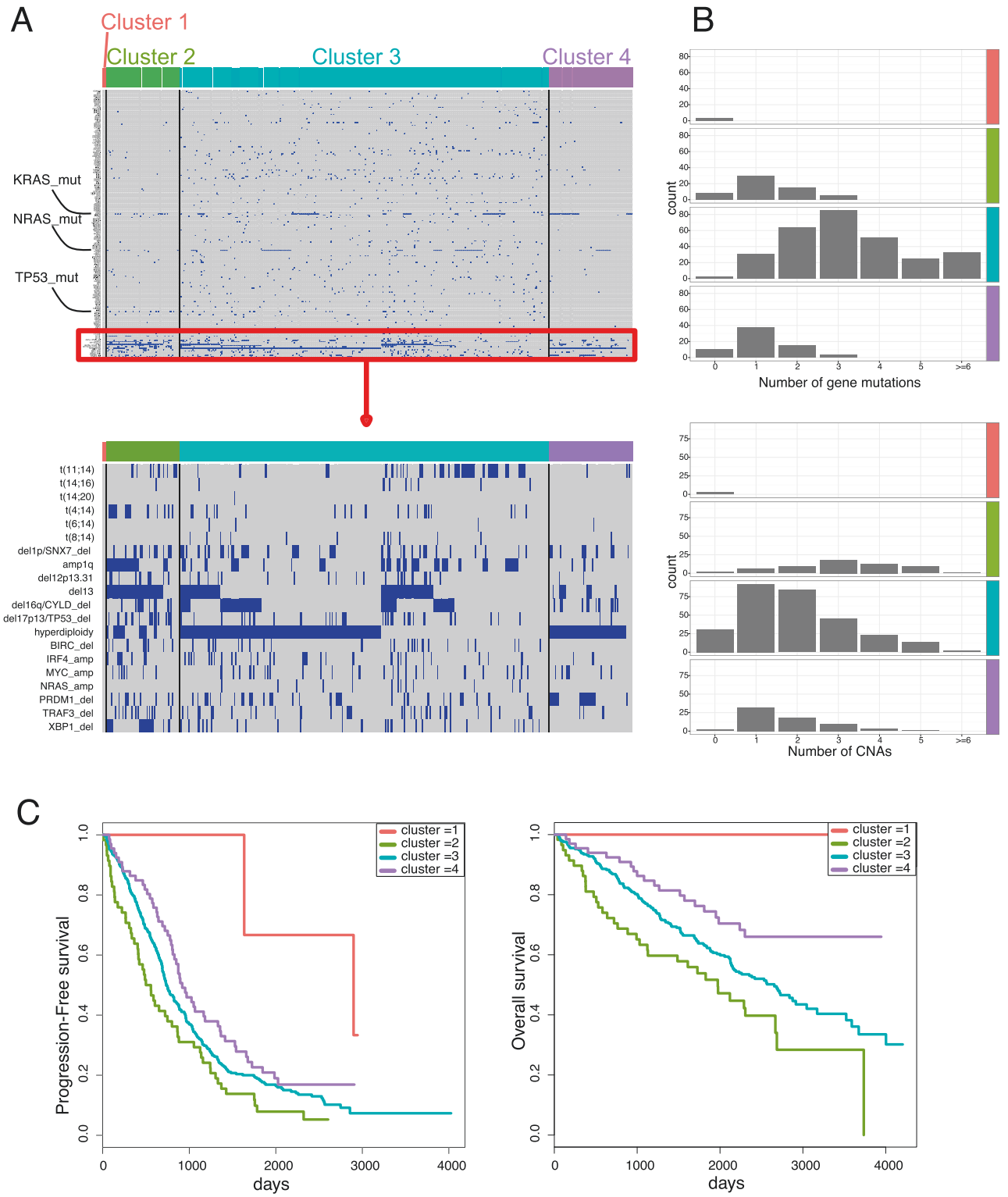


Fig. 6 Clustering of myeloma samples based on genomic landscape: **a** Bayesian Dirichlet clustering process of the 418 MM cases (in columns) based on genomic variables (in rows, top panel), with verticals black lines showing separation between identified clusters. A zoomed in view for the karyotypic abnormalities is provided in the bottom panel. Cluster 1 is composed of three patients where no driver event could be identified. **b** For each of the four clusters, the histogram of the distribution of gene mutations (top) and that of the CNAs (bottom) are provided. **c** Progression-free (left) and overall survival (right) Kaplan–Meier analysis of the four clusters

somewhat higher percentage of subclonal variants, suggesting these mutations may preferentially be acquired later in myeloma evolution (Fig. 4a). *TP53* was also more frequently subclonal, but was nevertheless clonal in 11% of myeloma cases. Conversely, *IRF4* and *DIS3* mutations were most likely to be clonal and therefore acquired earlier in evolution. We then looked at recurrent precedencies in pairwise analysis of mutated genes, and again found no recurrent pattern suggesting that mutations of driver genes, in our sample size, do not follow preferential evolutionary trajectories (Supplementary Table S4). Interestingly, we found that multiple mutated alleles of driver gene mutations -up to 5 for *TP53*- could be found at the subclonal level in a significant fraction of patients (Supplementary Figure 8B), again highlighting the heterogeneous subclonal structure of MM due to convergent evolution [35, 36].

The impact of clonal status on the prognostic value of gene mutations is unclear for most cancers and has never been studied in myeloma. Interestingly, we found that the clonality status of mutations did not influence survival, with the exception of a trend towards improved OS for TP53 subclonal mutations (Fig. 4b, c, Supplementary Figure 9).

Patterns of co-occurring and mutually exclusive events

Patterns of co-occurrence of driver events were systematically analyzed to look for potential cooperativity in pairwise associations and functional redundancy for mutually exclusive variables. In the list of 197 driver events we identified 843 significant interactions at $p < 0.05$ (Fisher Test), 15 at false-discovery rate <0.1 and 9 at family-wise error rate <0.05 out of a total 19306 possible interactions (Supplementary Figure 10, most frequent events in Fig. 5). Among, significant pairwise interactions, some were previously described, such as the positive correlation between chr13 and known poor prognostic factors such as t(4;14), amp1q, del12p13.31, del17p13, *TP53* mutations. *KRAS* and *NRAS* mutations were mutually exclusive between each other but not with *BRAF* mutations. We found a general pattern where mutations in tumor suppressor genes co-occurred with deletion of the wild-type allele in a typical double-hit manner for *TP53*, *CYLD*, and *TRAF3* mutations, supporting that these bi-allelic events not only underlie relapse but may be present at diagnosis as well [37]. Hyperdiploid and IGH-translocated cases showed significant differences in the spectrum of associated genomic events. Chr13 deletions, chr1q amplifications, *DIS3* and *ZFHX4* mutations were less frequent in HDMM, which in turn was enriched for *FAM46C* mutations. *XBP1* deletions clustered with deletions of NF-kB pathway genes *TRAF3* and *BIRC2/3*, and with chromosomal events such as chr12p, chr13, chr17p deletions and t(4;14). Finally, t(11;14) and t(4;14) showed a diverse spectrum of associated driver lesions, with t(4;14) cases in particular being associated with chr13 deletions and chr1q amplifications.

We then extended this analysis to uncover preferential trajectories of cancer evolution highlighting different associations of hotspot mutations within genes (Supplementary Figure 11). We found a tendency for hotspot mutations in *KRAS*, *NRAS*, *BRAF*, *IRF4* to show a differential pattern of co-occurrence, but none with a significant false-discovery rate.

Genomic structure suggests new myeloma subtypes with clinical impact

The identification of preferential patterns of associations or mutual exclusivity of driver lesions, prompted us to explore whether new myeloma subtypes could be identified based on their extended genotype using a previously described Bayesian model [16]. We could reliably identify four clusters (Fig. 6a) of MM cases with good posterior probability of class assignment (Supplementary Figure 12Ai–Aiii), albeit lower than acute myeloid leukemia, which carry a simpler genomic structure [16]. Aside from a small cluster 1 (3 patients) where no driver events could be identified, most hyperdiploid and IGH-translocated cases clustered together in the large cluster 3 (291 patients, 69.6%). We then identified two clusters, 2 and 4, composed of 58 (13.9%) and 66 (15.8%) patients, respectively. Both clusters were characterized by a significantly lower number of mutations but showed opposing features otherwise. Cluster 2 was enriched for IGH translocations, had the highest number of CNAs, was enriched for amp(1q), del(13), del(17p), deletions of *BIRC2/3* and *XBP1* and carried more *TP53* mutations (Fig. 6b, Supplementary Figure 12B). On the contrary, cluster 4 was mostly composed of hyperdiploid cases and showed fewest CNAs and mutations overall (Fig. 6b, Supplementary Figure 12B). There was no association between treatment (PI, ASCT) and subgroups, while cluster 4 was characterized by a slightly lower age (Supplementary Figure 12C).

Prognosis was different for both PFS (Fig. 6c, left) and OS (Fig. 6c, right) across the clusters. Cluster 2 showed the worse median OS—1973 days—whereas cluster 4 was associated with longer survival (median OS not reached), again showing concordance between the number of CNAs and prognosis.

Discussion

We provide the first large-scale TPD effort in MM to comprehensively describe its driver genomic landscape and

how this impacts on clinical behavior. Compared to WES, we reduced the amount of input DNA, the complexity of downstream analysis and forwent the need for a matched sample, to move this technology one step closer to the clinic. Our approach was robust and conservative, and even larger studies may allow re-classification of events discarded in our list as non-driver.

Gene mutations were frequent in our cohort; however, we found a preponderant contribution of the few highly recurrently mutated genes that were already described. Nevertheless in individual cases, rare gene mutations may impact treatment decisions. For example, *CRBN*, *IKZF1* or *IKZF3* mutations (totaling a combined 1.6% of patients in our cohort) would predict resistance to immunomodulatory drug [26], whereas *XBP1* deletions could predict bortezomib resistance [38], suggesting future clinical-grade targeted platforms must include rare driver events that have predictive value for treatment.

Confirming previous reports, the prognostic yield of gene mutations was rather low in our large cohort as compared to chromosomal events. PFS and OS were both impacted only by *TP53* mutations and by rare mutations in *ATP13A4*, an ion transporter previously associated with developmental disorders [39] but whose role in MM was never reported before. We did not confirm a prognostic role of genes such as *IRF4*, *EGR1* and *ZFXH4* [13], possibly because of front-line PI treatment and longer follow-up of our cohort (63 months here vs 25 in ref. [13]). We identified subclonal mutations for all recurrent driver genes, showing that lesions classically associated with disease progression can spontaneously arise before diagnosis. As these lesions can be positively selected by treatment, sensitive methods must be employed for their identification at diagnosis to achieve an accurate prognostication. Some events had a strong, context-dependent, prognostic value, such as deletions of *PRDM1* coupled with either t(4;14) or *BIRC2/3* deletions. These interactions may be infrequent but are highly instructive and clinically relevant, supporting the use of extended gene panels in prognostic studies. However, to identify more such instances, much larger sample sizes will be needed given the heterogeneous genome of MM and the expected modest effect of most interactions. In the future, global risk calculators that incorporate extended clinical, demographic, laboratoristic and genomic variables will be able to address these points [40]. *BIRC2/3* genes deletions have been described in myeloma [41, 42], whereas *PRDM1* has only recently been identified as a tumor suppressor based on recurrent truncating mutations [11]. Truncations and deletions in *PRDM1* were nevertheless previously described in lymphomas [43, 44], where they are associated with poor prognosis [45]. Deletions of the 6q21 chromosome band containing the *PRDM1* locus are known to be recurrent in MM [46]. The targeted nature of our study, however, did not allow us to characterize the size of such deletions and thus we cannot exclude that other tumor suppressors deleted in 6q contributed to this observation.

Overall, driver events had an additive effect on prognosis, showing that progressive accumulation of abnormalities correlates with clinical aggressiveness, almost independently of the nature of such events. Therefore, our findings extend previous initial observations on modulation of prognosis by additive negative prognostic factors in FISH [47], and that of aneuploidies in the prognosis of HDMM or IGH translocations [48, 49].

Current MM classification based on IGH translocations or hyperdiploidy is reliable and biologically relevant, as it is based on known early driver events. Here, we report on the first attempt toward a genomic classification of myeloma using innovative clustering algorithms based on the extended genotype of each patient. Allowing for the heterogeneity of the disease, we could identify disease subgroups characterized by a different spectrum of translocations, CNAs and gene mutations, as well as different contribution by each class of events. A cluster of hyperdiploid cases characterized by the fewest gene mutations and CNAs showed a relatively good prognosis, whereas non-hyperdiploid cases carrying multiple segmental chromosomal aneuploidies and fewer gene mutations carried a worse prognosis. Again, specific gene mutations did not contribute much to the clustering. Similar to our analysis, a report integrating MM expression profiles with mutational data to identify subgroups with biological and prognostic values showed little contribution of gene mutations to the clustering [50]. However, knowledge of the mutational status of a large number of genes allowed us to perform a prognostically relevant clustering of cases based on the overall mutational burden, supporting the value of analysis of infrequently mutated genes. Lastly, a recent study also reported on the dissection of hyperdiploid cases with different prognosis based on specific additional cytogenetic lesions, supporting the validity of extended genotyping for prognostication of MM [51].

The increasing availability of novel drugs and the better understanding of pathways involved in MM progression prompts the need for rationalized treatment approaches and mandates that high-risk disease is diagnosed with accuracy [52,53]. Probably, future genomic studies will provide not only improved prognostication, but also predictive factors of response and actionable mutations that will help in treatment choices. The methodology and results described here represent an important advance that can accelerate the introduction of genomics in the clinical approach to MM.

Acknowledgements This work was supported by a grant from NIH PO1-155258 (NCM, PJC, KCA, HAL, and SM), RO1-124929 (NCM) P50-100007 (KCA, NCM) and by the Wellcome Trust (grant reference 077012/Z/05/Z). NB was funded by a MFAG Grant No. 17658 from the Associazione Italiana Ricerca sul Cancro (AIRC). FM was supported by the Associazione Italiana Leucemie e Linfomi (AIL) and by the Società Italiana di Ematologia Sperimentale (SIES). PJC is personally funded through a Wellcome Trust senior clinical research fellowship. KCA is an American Cancer Society clinical research professor.

Author contributions NB designed the study, collected and analyzed the data and wrote the paper; GB, MM, SG, YL, CdP, FM, TB, MG analyzed data. Y-TT, EK, MD, PM, SM provided samples. VS, LM, SO'M, KR, JWT, APB, CC processed samples and raw sequencing data. PC, KCA, PJC, EP analyzed data and critically revised the manuscript. HA-L and NCM conceived the project, provided samples, interpreted data.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Palumbo A, Anderson K. Multiple myeloma. N Engl J Med. 2011;364:1046–60.
2. Morgan GJ, Walker BA, Davies FE. The genetic architecture of multiple myeloma. Nat Rev Cancer. 2012;12:335–48.
3. Corre J, Munshi N, Avet-Loiseau H. Genetics of multiple myeloma: another heterogeneity level? Blood. 2015;125:1870–6.
4. Fonseca R, Leif Bergsagel PL, Drach J, Shaughnessy J, Gutierrez N, Stewart AK, et al. International Myeloma Working Group molecular classification of multiple myeloma: spotlight review. Leukemia. 2009;23:2210–21.
5. Zingone A, Kuehl WM. Pathogenesis of monoclonal gammopathy of undetermined significance and progression to multiple myeloma. Semin Hematol. 2011;48:4–12.
6. Palumbo A, Avet-Loiseau H, Oliva S, Lokhorst HM, Goldschmidt H, Rosiñol L, et al. Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. J Clin Oncol. 2015;33:2863–9.
7. Avet-Loiseau H, LeLeu X, Roussel M, Moreau P, Guerin-Charbonnel C, Caillot D, et al. Bortezomib plus dexamethasone induction improves outcome of patients with t(4;14) myeloma but not outcome of patients with del(17p). J Clin Oncol. 2010;28:4630–4.
8. Walker BA, Walker BA, Wardell CP, Wardell CP, Melchor L, Melchor L, et al. Intraclonal heterogeneity is a critical early event in the development of myeloma and precedes the development of clinical symptoms. Leukemia. 2014;28:384–90.
9. Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, Schinzel AC, et al. Initial genome sequencing and analysis of multiple myeloma. Nature. 2011;471:467–72.
10. Bolli N, Avet-Loiseau H, Wedge DC, Van Loo P, Alexandrov LB, Martincorena I, et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. Nat Commun. 2014;5:2997.
11. Lohr JG, Stojanov P, Carter SL, Cruz-Gordillo P, Lawrence MS, Auclair D, et al. Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. Cancer Cell. 2014;25:91–101.
12. Walker BA, Wardell CP, Melchor L, Hulkki S, Potter NE, Johnson DC, et al. Intraclonal heterogeneity and distinct molecular mechanisms characterize the development of t(4;14) and t(11;14) myeloma. Blood. 2012;120:1077–86.
13. Walker BA, Boyle EM, Wardell CP, Murison A, Begum DB, Dahir NM, et al. Mutational spectrum, copy number changes, and outcome: results of a sequencing study of patients with newly diagnosed myeloma. J Clin Oncol. 2015;33:3911–20.
14. Manier S, Salem KZ, Park J, Landau DA, Getz G, Ghobrial IM. Genomic complexity of multiple myeloma and its clinical implications. Nat Rev Clin Oncol. 2017;14:100–13.
15. Papaemmanuil E, Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood. 2013;122:3616–27.
16. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic classification and prognosis in acute myeloid leukemia. N Engl J Med. 2016;374:2209–21.
17. Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia. 2014;28:241–7.
18. Bolli N, Manes N, McKerrell T, Chi J, Park N, Gundem G, et al. Characterization of gene mutations and copy number changes in acute myeloid leukemia using a rapid target enrichment protocol. Haematologica. 2015;100:214–22.
19. Kortüm KM, Mai EK, Hanafiah NH, Shi C-X, Zhu YX, Bruins L, et al. Targeted sequencing of refractory myeloma reveals a high incidence of mutations in CRBN and Ras pathway genes. Blood. 2016;128:1226–33.
20. Kortüm KM, Langer C, Monge J, Bruins L, Egan JB, Zhu YX, et al. Targeted sequencing using a 47 gene multiple myeloma mutation panel (M(3) P) in -17p high risk disease. Br J Haematol. 2015;168:507–10.
21. Kortuem KM, Braggio E, Bruins L, Barrio S, Shi CS, Zhu YX, et al. Panel sequencing for clinically oriented variant screening and copy number detection in 142 untreated multiple myeloma patients. Blood Cancer J. 2016;6:e397.
22. Bolli N, Li Y, Sathiaseelan V, Raine K, Jones D, Ganly P, et al. A DNA target-enrichment approach to detect mutations, copy number changes and immunoglobulin translocations in multiple myeloma. Blood Cancer J. 2016;6:e467.
23. Jones D, Raine KM, Davies H, Tarpey PS, Butler AP, Teague JW, et al. cgpCaVEManWrapper: simple execution of CaVEMan in order to detect somatic single nucleotide variants in NGS Data. Curr Protoc Bioinf. 2016;56:15.10. 1–15.10.18.
24. Raine KM, Hinton J, Butler AP, Teague JW, Davies H, Tarpey P, et al. cgpPindel: identifying somatically acquired insertion and deletion events from paired end sequencing. Curr Protoc Bioinf. 2015;52:15.7.1–12.
25. Lohr JG, Lohr JG, Stojanov P, Stojanov P, Lawrence MS, Lawrence MS, et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by

whole-exome sequencing. Proc Natl Acad Sci USA. 2012;109:3879–84.
26. Egan JB, Kortuem KM, Kurdoglu A, Izatt T, Aldrich J, Reiman R, et al. Extramedullary myeloma whole genome sequencing reveals novel mutations in Cereblon, proteasome subunit G2 and the glucocorticoid receptor in multi drug resistant disease. Br J Haematol. 2013;161:748–51.
27. Fonseca R, Debes-Marun CS, Picken EB, Dewald GW, Bryant SC, Winkler JM, et al. The recurrent IgH translocations are highly associated with nonhyperdiploid variant multiple myeloma. Blood. 2003;102:2562–7.
28. Avet-Loiseau H, Li C, Magrangeas F, Gouraud W, Charbonnel C, Harousseau J-L, et al. Prognostic significance of copy-number alterations in multiple myeloma. J Clin Oncol. 2009;27:4585–90.
29. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med. 2014;371:2488–98.
30. Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med. 2014;371:2477–87.
31. McKerrell T, Park N, Moreno T, Grove CS, Ponstingl H, Stephens J, et al. Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. Cell Rep. 2015;10:1239–45.
32. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat Med. 2014;20:1472–8.
33. Busque L, Patel JP, Figueroa ME, Vasanthakumar A, Provost S, Hamilou Z, et al. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. Nat Genet. 2012;44:1179–81.
34. Ludwig H, Bolejack V, Crowley J, Bladé J, Miguel JS, Kyle RA, et al. Survival and years of life lost in different age cohorts of patients with multiple myeloma. J Clin Oncol. 2010;28:1599–605.
35. Melchor L, Brioli A, Wardell CP, Murison A, Potter NE, Kaiser MF, et al. Single-cell genetic analysis reveals the composition of initiating clones and phylogenetic patterns of branching and parallel evolution in myeloma. Leukemia. 2014;28:1705–15.
36. Raab MS, Lehners N, Xu J, Ho AD, Schirmacher P, Goldschmidt H, et al. Spatially divergent clonal evolution in multiple myeloma: overcoming resistance to BRAF inhibition. Blood. 2016;127:2155–7.
37. Weinhold N, Ashby C, Rasche L, Chavan SS, Stein C, Stephens OW, et al. Clonal selection and double hit events involving tumor suppressor genes underlie relapse from chemotherapy: myeloma as a model. Blood. 2016;128:1735–44.
38. Soriano GP, Besse L, Li N, Kraus M, Besse A, Meeuwenoord N, et al. Proteasome inhibitor-adapted myeloma cells are largely independent from proteasome activity and show complex proteomic changes, in particular in redox and energy metabolism. Leukemia. 2016;30:2198–207.
39. Kwasnicka-Crawford DA, Carson AR, Roberts W, Summers AM, Rehnström K, Järvelä I, et al. Characterization of a novel cation transporter ATPase gene (ATP13A4) interrupted by 3q25-q29 inversion in an individual with language delay. Genomics. 2005;86:182–94.
40. Gerstung M, Papaemmanuil E, Martincorena I, Bullinger L, Gaidzik VI, Paschka P, et al. Precision oncology for acute myeloid leukemia using a knowledge bank approach. Nat Genet. 2017;49:332–40.
41. Keats JJ, Fonseca R, Chesi M, Schop R, Baker A, Chng W-J, et al. Promiscuous mutations activate the noncanonical NF-kappaB pathway in multiple myeloma. Cancer Cell. 2007;12:131–44.
42. Annunziata CM, Davis RE, Demchenko Y, Bellamy W, Gabrea A, Zhan F, et al. Frequent engagement of the classical and alternative NF-kappaB pathways by diverse genetic abnormalities in multiple myeloma. Cancer Cell. 2007;12:115–30.
43. Pasqualucci L, Compagno M, Houldsworth J, Monti S, Grunn A, Nandula SV, et al. Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma. J Exp Med. 2006;203:311–7.
44. Tam W, Gomez M, Chadburn A, Lee JW, Chan WC, Knowles DM. Mutational analysis of PRDM1 indicates a tumor-suppressor role in diffuse large B-cell lymphomas. Blood. 2006;107:4090–100.
45. Xia Y, Xu-Monette ZY, Tzankov A, Li X, Manyam GC, Murty V, et al. Loss of PRDM1/BLIMP-1 function contributes to poor prognosis of activated B-cell-like diffuse large B-cell lymphoma. Leukemia. 2016;31:625–36.
46. Cigudosa JC, Rao PH, Calasanz MJ, Odero MD, Michaeli J, Jhanwar SC, et al. Characterization of nonrandom chromosomal gains and losses in multiple myeloma by comparative genomic hybridization. Blood. 1998;91:3007–10.
47. Boyd KD, Ross FM, Chiecchio L, Dagrada GP, Konn ZJ, Tapper WJ, et al. A novel prognostic model in myeloma based on co-segregating adverse FISH lesions and the ISS: analysis of patients treated in the MRC Myeloma IX trial. Leukemia. 2012;26:349–55.
48. Pawlyn C, Melchor L, Murison A, Wardell CP, Brioli A, Boyle EM, et al. Coexistent hyperdiploidy does not abrogate poor prognosis in myeloma with adverse cytogenetics and may precede IGH translocations. Blood. 2015;125:831–40.
49. Chretien ML, Corre J, Lauwers-Cances V, Magrangeas F, Cleynen A, Yon E, et al. Understanding the role of hyperdiploidy in myeloma prognosis: which trisomies really matter? Blood. 2015;126:2713–9.
50. Laganà A, Perumal D, Melnekoff D, Readhead B, Kidd BA, Leshchenko V, et al. Integrative network analysis identifies novel drivers of pathogenesis and progression in newly diagnosed multiple myeloma. Leukemia. 2017. https://doi.org/10.1038/leu.2017.197
51. Shah V, Sherborne AL, Walker BA, Johnson DC, Boyle EM, Ellis S, et al. Prediction of outcome in newly diagnosed myeloma: a meta-analysis of the molecular profiles of 1905 trial patients. Leukemia. 2017;32:102–10. https://doi.org/10.1038/leu.2017.179
52. Usmani SZ, Rodriguez-Otero P, Bhutani M, Mateos MV, Miguel JS. Defining and treating high-risk multiple myeloma. Leukemia. 2015;29:2119–25.
53. Maura F, Petljak M, Lionetti M, Cifola I, Liang W, Pinatel E. et al. Biological and prognostic impact of APOBEC-induced mutations in the spectrum of plasma cell dyscrasias and multiple myeloma cell lines. Leukemia. 2017. ePub 6 December 2017.