University of Turin

Doctoral School in Life and Health Sciences

PhD Program in Complex Systems for Life Sciences

Integrated analysis of genomic, epigenomic and gene expression profiles in multiple datasets

Candidate: Dr. Clara Viberti

Tutor: Prof. Giuseppe Matullo

XXX Cycle

Academic years: 2014-2017

Index

Abstract

During my PhD work I was involved in several research projects integrating different omics data with the aim to identify susceptibility/early detection biomarkers and prognostic factors in different diseases.

In my PhD thesis I will focus on two cancer-related projects on bladder cancer (BC) and malignant pleural mesothelioma (MPM).

BC is the sixth most commonly diagnosed tumor worldwide. DNA repair pathways have been implicated in several cancer risk, including BC. The term DNA repair capacity (DRC) refers to the ability of a cell to protect the integrity of the genome and it has been observed that individuals with low DRC tend to accumulate more damage than those with a more efficient DRC.

We aimed at studying the relationship between DNA damage, DRC and BC risk or clinical outcome in 159 BC cases and 159 matched controls, integrating with gene expression and epigenetic profile data. We investigated γ-H2AX phosphorylation levels, micronuclei (MN) frequencies and DNA damage in peripheral blood mononuclear cells (PBMCs). We found higher MN levels in cases compared to controls (p=0.0004). On the other hand, we observed a significant association of γ-H2AX basal levels with risk of BC recurrence/progression (p=0.02). This suggests a protective effect of DNA damage signaling in terms of preventing BC recurrence or progression. Finally, we measured basal DNA damage by Comet assay and we found associations with patients' outcomes. DNA damage was also significantly lower in subjects with a longer survival time. Based on these data, we suggest that DNA damage levels in PBMCs of BC patients may potentially represent a prognostic marker associated with poor survival.

At present there are no suitable biomarkers for an early BC diagnosis or relapse/progression prognosis. To improve diagnostic accuracy and to overcome the disadvantages of current diagnostic strategies, the detection of biomarkers associated with BC in biofluids/surrogate tissues with non-invasive techniques could results in a more promising approach than analyzes of biopsies.

In order to identify BC biomarkers, we performed whole genome methylation and gene expression analyzes. Two metalloproteinases (*MMP23A* and *MMP23B*) resulted significantly under-expressed in BC cases compared to healthy controls (logFC=-0.23, p=0.01; logFC=-0.37, p=0.007, respectively) with their

expression levels significantly correlated with the relative CpGs methylation. Thus, we investigated MMP23B protein levels both in plasma and in urine samples, to confirm gene expression observations. Moreover, we investigated the expression levels of miRNAs targeting *MMP23B* gene, and we identified 5 of them differentially expressed in urine of BC patients and controls.

The second project I was involved in concerns malignant pleural mesothelioma, an aggressive tumor strongly associated with asbestos exposure. MPM patients are usually diagnosed at an advanced stage when current treatments have limited benefits. This aspect highlights the need of new non-invasive tests to monitor asbestos-exposed people for an early MPM diagnosis that might improve life expectancy.

We identified novel blood DNA methylation markers of MPM in 82 cases and 68 asbestos-exposed cancer-free controls (Training Set). Results were replicated in a Test Set including 81 MPM cases and 69 controls .

Considering the "top" differentially methylated signals, 7 single-CpGs and 5 genomic regions of coordinated methylation replicated with similar effect size in the Test Set ($p_{fdr}<0.05$), with the top hypomethylated single-CpG being in *FOXK1*.

In the Test set, the comparison of the area under the curve (AUC) of two models, including/excluding methylation, indicated a significant increase in case/control discrimination when considering DNA methylation together with asbestos exposure (AUC=0.81 *vs* AUC=0.89, p=0.0013). Overall, these results suggest that differential DNA methylation patterns of selected CpGs may aid in discriminating MPM cases from asbestos-exposed healthy subjects. If verified in prospectively collected samples, CpGs methylation may be useful to improve MPM risk estimation for subjects occupationally and/or environmentally exposed to asbestos and for early MPM detection.

1. Bladder cancer study

1.1. Introduction

Bladder cancer (BC) origins when cells of the urinary bladder start to grow and spread in an uncontrolled and abnormal way. The organization of the bladder in layers is very important to understand how the tumor develops and consequently what could be the best treatment.

Bladder tumors are divided into several types depending on the histological analysis .

- *Transitional cell (urothelial) carcinoma (TCC)* is the more frequent type, accounting for 90%-95% of all BC [1]. It begins in the urothelial cells found in the lining of the bladder. Since transitional cells also line other parts of the urinary tract (kidneys, ureters, or urethra), TCCs can also occur in these areas. TCCs are also divided into two subtypes, papillary and flat, based on the way they grow. Papillary TCCs grow in slender projections towards the hollow center of the bladder. Flat TCCs, instead, do not penetrate into the central part of the bladder; tumor of this subtype is known as a flat carcinoma in situ (CIS).
- *Squamous cell carcinoma* develops in the bladder lining in response to prolonged infection, irritation and inflammation. In certain parts of the Middle East and Africa this is the predominant form of BC and is associated with chronic infection caused by *Schistosoma* worm. This carcinoma accounts for about 3% of all BC. Almost all squamous cell carcinomas are invasive [2].
- *Adenocarcinoma* accounts for about 1% of all BCs and develops from glandular cells. Nearly all adenocarcinomas of the bladder are invasive [3].
- *Sarcoma* and *small cell cancer* arise, respectively, in muscle and neuroendocrine cells, but they are very rare [3].

BCs are often described by clinicians based on how far they have invaded into the wall of the bladder. *Non-muscle-invasive BC (NMIBC)*, also defined as superficial, are cancers still in the inner layer of cells (the transitional epithelium), accounting for approximately 75% of newly diagnosed BC patients. NMIBCs

frequently recur, but rarely progress to a muscle-invasive status. *Muscle-invasive BC (MIBC)*, or invasive BC grows deep into the muscle layer and represents around 25% of the newly diagnosed patients. Invasive cancers are more likely to spread and are harder to treat [1, 4].

Histopathological staging of bladder tumors is performed according to the 2009 Tumor, Node, Metastasis (TNM) classification system approved by the Union Internationale Contre le Cancer (UICC). Consequently to surgical removal and histopathological analysis, T-stage defines the extent of tumor invasion into the bladder wall, while N-stage describes the involvement of regional lymph nodes. M-stage reports the presence of distant metastasis (**Table 1**) [5].

Table 1 | **TNM classification of bladder cancer**

	T: Primary tumor
TX:	primary tumor cannot be assessed
T0:	no evidence of primary tumor
Ta:	noninvasive papillary carcinoma
CIS:	Carcinoma in situ, ‘‘flat tumor’’
T1:	tumor invades *lamina propria*, but not the detrusor muscle
T2:	tumor invades muscle under the connective tissue layer
	T2a: only superficial muscle affected (inner half)
	T2b: deep muscle penetrated (outer half)
T3:	tumor invades perivesical tissue
	T3a: microscopically
	T3b: macroscopically (extravesical visible mass)
T4:	tumor spreads beyond the bladder to the prostate, uterus, vagina, pelvic or abdominal wall
	T4a: tumor has spread to the prostate, uterus or vagina
	T4b: tumor has spread to the pelvic or abdominal wall
	N: Lymph nodes
NX:	regional lymph nodes cannot be assessed
N0:	no regional lymph nodes metastasis
N1:	metastasis in a single lymph node in the true pelvis
N2:	metastasis in multiple lymph nodes in the true pelvis
N3:	Metastasis in a common iliac lymph node(s)
	M: Distant metastasis
MX:	distant metastasis cannot assessed
M0:	no distant metastasis
M1	distant metastasis

The grading system is also very important in the classification of BC. In 2004, a new classification system for grading non-invasive urothelial bladder carcinomas was approved and published by the World Health

Organization (WHO) and the International Society of Urological Pathology [6]. In 2016, a third edition of the WHO classification was issued, accentuating the changes proposed by the 2004 over the 1973 classification [1]; the latter has 3 categories, based on how well the tumor is differentiated. Of note, the intermediate grade which was the subject of controversy in the 1973 WHO classification, was removed from the 2004 version. However, the use of both classifications is recommended for an appropriate stratification of BC patients in terms of risk of recurrence and progression [5, 6].

Finally, based on number of tumors, size, prior recurrence rate, T stage, grade and presence of CIS, the EORTC tables stratify patients into three groups by risk of recurrence and progression of non-muscle-invasive BC. This classification was developed with the purpose of helping in the selection of the most appropriate treatments, after performing transurethral bladder tumor resections [7, 8] (**Table 2**).

Table 2 | **Risk group stratification**

Risk Group	Definition
Low-risk tumors	Primary, solitary, Ta, G1/PUNLMP/LG, <3 cm, no CIS
Intermediate-risk tumors	All tumors not defined in the two adjacent categories (between the category of low and high risk)
High-Risk tumors	Any of the following: T1, HG/G3, CIS, Multiple and recurrent and large (>3 cm) Ta G1 G2 (all these conditions must be present)

PUNLMP: papillary urothelial neoplasms of low malignant potential
G1,G2,G3: 1973 grading system classification

1.2. Epidemiology of bladder cancer

Risk factors and susceptibility

As for the majority of the diseases there is also for BC a list of risk factors due to external exposures or to inherited genetic predispositions.

Tobacco smoking is recognized as the main risk factor for BC and active smokers have a 2.5 times higher risk than non-smokers [9]. In various populations, tobacco has been found to be responsible for about 50% of all BCs and 40% of all BC-related deaths [10]. Since a lag time of about 30 years is evident between tobacco exposure and diagnosis [11], nowadays BC incidence is higher in regions that had very high smoking prevalence in the '80s [10]. In particular, in Spain and in Italy, smoking rates in 1980 amounted to 44.4% and 44.3%, respectively; the highest BC incidence in men in these countries was around 37 per 100,000 in 2003 and 33 per 100,000 in 2007 [1].

Over the last decades, smoking prevalence has declined considerably in North America and many European countries: this may explain the drop in BC incidence and mortality observed in these regions. Unfortunately, in other less-developed regions of the world, such as Africa, the Middle East or Central and South America, the trend in tobacco consumption is opposite, fact probably due to the lack of awareness of smoking damages [1]. As a result, early assessment of smoking status is crucial in the management of BCs and cessation of smoking has been suggested to benefit BC outcomes, reducing the risk in both sexes [12].

Occupational exposure to carcinogens, namely aromatic amines, polycyclic aromatic and chlorinated hydrocarbons, is considered as the second most notable risk factor for BC. Approximately one fifth of all bladder tumors in the past have been suggested to be related to such exposure, mainly in industrial areas processing paint, rubber, petroleum products and dyes. Other workers with an increased risk of developing BC include painters, machinists, printers, hairdressers (because of the exposure to hair dyes), and truck drivers (exposure to diesel fumes). The great risk for BC-specific mortality occurs in electrical and chemical process workers [1, 12].

Exposure to high levels of arsenic in drinking water has been also recognized as a cause of BC. For instance, arsenic pollution was correlated with BC risk in Argentina, Chile and Bangladesh [4]. Although the current WHO maximum acceptable arsenic limit in water is 10 μg/L, under which it is not considered noxious, in a recent study Saint-Jacques and colleagues suggested an increased BC risk with exposure to drinking water arsenic levels within this limit. Indeed, the risk was 16% (2–5 μg/ L) and 18% (≥5 μg/L) greater than that of the referent group (< 2 μg/L) [13].

Since many carcinogens ingested via food are excreted into the urine, resulting in direct exposure of the urothelium, nutritional aspects have been considered relevant to BC tumorigenesis as for other cancers. In general, diets low in fruits and vegetables and high in meat consumption have been suggested to increase cancer risk; however, there is no significant evidence regarding BC onset or risk of recurrence/progression [14]. Furthermore, there are not clear associations between alcohol consumption and BC risk and epidemiological data are often confounded by other risk factors. In these sense, few evidence of possible positive association have been found just in men and smokers [15].

About gender, diverging trends by sex were observed in some countries: while the lower BC incidence is likely to represent historically lower smoking prevalence and less occupational exposures to carcinogens in women, the reasons for the higher mortality rate are still unclear [10].

Since BCs are associated with an acquired carcinogen exposure, the avoidance of carcinogens could significantly decrease the incidence of the diseases [4]. Perhaps the observed differences in incidence between men and women reflect in part the impact of such carcinogens. The importance of primary prevention (i.e. avoidance of causative agents, particularly smoking) needs to be stressed while secondary prevention (i.e. screening of the population for at-risk individuals) has been debated for high-risk populations [4].

Few data on the impact of race on BC incidence exist; African Americans show a lower age-standardized incidence (ASI) rate compared to white individuals in the United States (13 and 22 per 100,000, respectively), and blacks have been reported to be associated with reduced survival (5-year disease specific survival: whites, 82.8%; blacks, 70.2%; Hispanics, 80.7%; Asian/Pacific Islanders, 81.9%) [16]. Additionally,

other factors such as unemployment, low socio economic status, urban living, medical conditions have been found linked to BC mortality [1].

Starting from ten years ago, large genome-wide association studies (GWAS) have been conducted in order to inspect the role of genetic polymorphisms on BC risk of development or relapse. The first GWAS on BC has been published in 2008 by Kiemeney and colleagues. Based on 1,803 urinary BC cases and 34,336 controls, authors found two novel single nucleotide polymorphisms (SNPs) near *MYC* (rs9642880) and *TP63* (rs710521) genes associated with urinary BC risk [17]. Later, further GWAS and candidate gene studies identified and validated fifteen additional genomic regions that seem to play a major role in development of this complex disease [18-25].

Some evidence support a pronounced genetic predisposition to BC for genes involved in the metabolism of carcinogens. For example, inherited genetic factors such as *NAT2* variants and *GSTM1*-null genotypes are associated with increased risk, both for non-invasive and invasive BCs, especially in smokers [26-28].

Recently, Lukas and colleagues have revisited the relationship between BC occurrence and SNPs in genes encoding for xenobiotic metabolizing enzymes, finding the strongest associations with occupational BC for *GSTM1* and *UGT1A* variants, especially when co-occurring. Furthermore, the polymorphisms were not only associated with increased BC risk but also with shorter relapse-free survival [29]. Interestingly, other polymorphisms may statistically interact to cause an increased BC risk respect to individual variants. In another recent study, Selinski et al demonstrated that specific variant combinations of three to four variants (in *GSTM1, SLC14A1, UGT1A, MYC, NAT2, CCNE1* genes) seem to play a major role in different categories of smokers, supporting the concept that interaction analyzes and stratification regarding smoking habits and cancer invasiveness could be a promising future approach to uncover further susceptibility variants that are relevant for particular subgroups of BC patients [24].

Incidence and mortality

BC is the fifth most common malignancy in the general population and the thirteenth most common cause of cancer death worldwide [30]. Estimates for newly diagnosed BC cases in the United States for 2016

exceed 76,000 and for deaths attributable to BC exceed 16,000 [31]. According to GLOBOCAN estimates, about 429,793 new BC cases and 165,084 BC deaths occurred worldwide in 2012 [32].

BC incidence is high in elderly for both sexes, after the age of 60 and rarely before the age of 40 [33]. The median age at diagnosis is 69 years for men and 71 for women [34]

Incidence varies considerably between regions and countries: age-standardized index (ASI) is almost three times greater in more-developed countries (ASI = 9.5 per 100,000 population) than in less-developed countries (ASI = 3.3 per 100,000 population). The highest ASI is observed in European countries (the highest in Spain and the lowest in Finland) and North America. By contrast, the lowest rates were noted in Central and South America [10, 35]. Incidence rates are critically higher in men than in women, with a ratio of 3:1. In the European Union ASI rate is 27 per 100,000 in men and 6 per 100,000 in women [10].

BC prevalence and mortality are mainly determined by the initial tumor stage, success of treatment and other causes (competing) mortality. Non–muscle-invasive tumors have a high prevalence because their low progression rates allow many patients to survive for long time, while patients with muscle-invasive disease are at significantly higher risk of dying from their disease [36].

There is disagreement in BC future prediction: on one hand the incidence of BC has decreased in some registries over the years (probably a reflection of the decreased exposure to causative agents such as smoking and better occupational hygiene); on the other hand, according to the WHO, the number of BC cases and deaths are anticipated to almost double in the near future (as a consequence of life expectancy increasing by 3% since 1950) [30]. Another possible explanation of this unfavorable expectation might lie into the relative lack of improvement in BC treatment: Boormans and Zwarthoff found the most important reason for this scarcity in the low funding for BC research and highlighted that high-funded cancers are the most frequent in incidence, but not necessary also in mortality [37]. Moreover, BC incidence rates in some countries are very low because non-muscle-invasive tumors are not registered, consequently BC reserved funding might be affected negatively [37].

1.3. Symptoms, diagnosis and treatment

The most important symptom of BC is painless macroscopic hematuria, which is diagnosed in 10-20% of patients, while 2-5% have microscopic hematuria. BC can also be suspected if the patient has symptoms like impaired storage of urine or other causes of irritation, such as frequency, urgency and dysuria (pain during urination). The definitive evaluation and diagnosis is performed using cystoscopy. Any abnormal finding requires histological evaluation because benign conditions, such as inflammatory diseases, can mimic BC.

Histology can be obtained by transurethral biopsy or resection of the entire area. Cytology, the inspection of cells in the urine, is often performed as an additional measure to detect missed cancers [1].

In addition to cystoscopy, other tests are employed to diagnose BC, such as ultrasound, computed tomography (CT). or intravenous pyelogram (IVP):

Different types of treatment are available for patients with BC, depending on the stage of the cancer, the severity of the symptoms, and the overall health status of the subject. Currently, the main types of treatment for BC are surgery, intravesical therapy, chemotherapy and radiation therapy. Sometimes, more than one treatment might be used at the same time. Surgery, alone or with other treatments, is used in nearly all cases.

For early-stage or superficial BC, a transurethral resection (TUR) is the most common treatment. Unfortunately, even with successful treatment, BC often recurs in other parts of the organ. In patients with a long history of recurrent non-invasive low-grade tumors, the surgeon may sometimes just use fulguration to burn small tumors [5].

When BC is invasive, all or part of the bladder may need to be removed. This operation is called cystectomy. It could be partial, if the cancer has invaded the muscle but is not very large and only in one place, or radical, when the cancer is extended or is in more than one part of the bladder. [5]. Some patients may be given chemotherapy after surgery to kill any cancer cells that are left. Treatment given after surgery is called adjuvant therapy.

Intravesical therapy refers to the administration of the drug directly into the bladder. Bacillus Calmette-Guerin therapy (BCG) is the most effective intravesical immunotherapy for treating early-stage BCs.

Mitomycin and thiotepa, instead, are the most often used drugs for intravesical chemotherapy. The goals of intravesical treatments are to reduce the implantation of tumor cells after TUR, prevent tumor recurrence and reduce tumor progression [38].

Systemic chemotherapy is another type of treatment for BC, and it is delivered through the bloodstream; for metastatic BC patients, it can shrink the cancer, delay BC recurrence and potentially prolong life [39].

It remains to mention radiation therapy, that is also used for BC treatment often combined with chemotherapy for higher efficacy.

1.4. DNA repair capacity and bladder cancer

Every day a combination of endogenous and exogenous factors affects the stability of the human genome;. if not repaired or repaired incorrectly, these lesions can cause genome replication and transcription block, mutations or aberrations, that compromise cell or organism viability.

DNA, like any other molecule, is subject to chemical reactions. DNA damage may result from either intrinsic or extrinsic agents. In general, the vast majority of DNA modifications are endogenous, including spontaneous hydrolysis, alkylation, mismatches as well as insertions and deletions. Besides the numerous internal sources of damage, cellular DNA is also under constant attack from outer or environmental DNA-damaging agents. These include physical stresses, such as ultraviolet light (UV) from the sun, ionizing radiation, which can originate from both natural (e.g., cosmic and gamma radiation) and artificial sources (e.g., medical treatments, such as X-rays and radiotherapy). Other well-studied environmentally occurring DNA-damaging chemicals include amines commonly found in diet (e.g., N-nitrosoamines, heterocyclic amines polycyclic aromatic hydrocarbons) or produced in emissions such as cigarette smoke and vehicle exhaust [40].

It has been estimated that each human cell is subject to approximately 70,000 lesions per day that are more or less tolerable, with the possibility to be converted in more dangerous ones [41]. Given this huge daily amount of endogenous and exogenous DNA damage, it became evident that cells must have acquired a system of defense designed to restore genome integrity, namely DNA repair. The importance of DNA repair is evident from the large investment that cells make in DNA repair enzymes, that help in reducing DNA anomalies and thereby re-establish the structure and the sequence of DNA once damage has occurred [41]. For example, analysis of the genomes of bacteria and yeasts has revealed that several percent of the coding capacity of these organisms is devoted solely to DNA repair functions. The importance of DNA repair is also demonstrated by the increased rate of mutation that follows the inactivation of a DNA repair gene.

Recent studies on the consequences of a diminished capacity for DNA repair in humans have linked a variety of human diseases with decreased repair; in particular, deficiencies in the DNA repair system are strongly associated with the risk of cancer, neurodegenerative and inflammatory disorders and aging [42,

43]. Accumulation of DNA lesions in neurons is associated with ataxias together with Xeroderma pigmentosum (XP), Alzheimer's, Huntington's and Parkinson's diseases. For example, in Xeroderma pigmentosum the affected individuals have an extreme sensitivity to ultraviolet radiation because they are unable to repair certain DNA photoproducts. This repair defect results in an increased mutation rate that leads to serious skin lesions and an increased susceptibility to certain cancers [44].

DNA damage causes uncontrolled tumorous cell growth when oncogenes are activated or tumor suppressor genes inactivated. On the contrary, when DNA damage persists and interferes with replication or transcription could trigger cellular senescence or apoptosis, resulting in a consequent inactivation or elimination of the damaged cells and in tumorigenesis suppression [42, 45]. Moreover, also chemotherapeutics used in cancer therapy induce a plethora of DNA lesions to halt cancer cell proliferation or trigger the apoptotic demise of cancer cells [45].

A network of multiple mechanisms is required and essential to maintain genomic integrity and stability. The human DNA repair system consists, fundamentally, of five pathways in which are involved more than 130 genes: the nucleotide excision repair (NER), base excision repair (BER), mismatch repair (MMR) and double-strand break (DSB) repair, divided into non-homologous end-joining (NHEJ) and homologous recombination (HRe) pathways [46] (**Figure 1**). Each of these pathways is designed to resolve specific type of DNA damages. (For a complete list of genes involved in the DNA repair pathways see *https://www.mdanderson.org/documents/Labs/Wood-Laboratory/human-dna-repair-genes.html*).

BER and NER are multi-step ”cut-and-patch”’ damage excision mechanisms; in both, the damage is resected the original DNA sequence restored by a DNA polymerase that uses the undamaged strand as its template, and the remaining break into the double helix is sealed by DNA ligase that repair modified bases. The two pathways differ in the way in which the damage is removed from DNA. BER involves a battery of enzymes called DNA glycosylases that recognize a specific type of altered base in DNA and remove them. As BER removes base adducts from methylation, deamination, and hydroxylation, it is considered as the main guardian against DNA lesions caused by endogenous cellular metabolism.

NER, indeed, can repair a remarkably wide variety of structurally changes of the DNA double helix; such as “bulky lesions” which are recognized and removed by a DNA helicase enzyme. Since global genome NER repairs lesions in the entire genome, it is important for preventing mutations and consequently cancer [47]. Mismatch repair consists in a ”cut-and-patch”-like excision mechanism that does not excise damaged DNA but corrects replication mistakes such as single base-base mismatches and small insertion/deletion loops. Mismatch detection and removal involve re-synthesis of the new DNA strand from the mismatch. As MMR prevents accumulation of these mutagenic lesions it is mainly important in the prevention of cancer.

A potentially dangerous type of DNA damage occurs when both strands of the double helix are broken: DSBs are deleterious lesions that arise from ionizing radiation, free radicals, chemicals. If these lesions are left unrepaired, they quickly lead to the breakdown of chromosomes into smaller fragments. DSB consists of 2 main sub-pathways: in the first, HRe, the damage is repaired by exchanges of equivalent regions of DNA between homologous or sister chromosomes; while in the NHEJ specific enzymes juxtapose and ligate the broken ends without the use of a template [41].

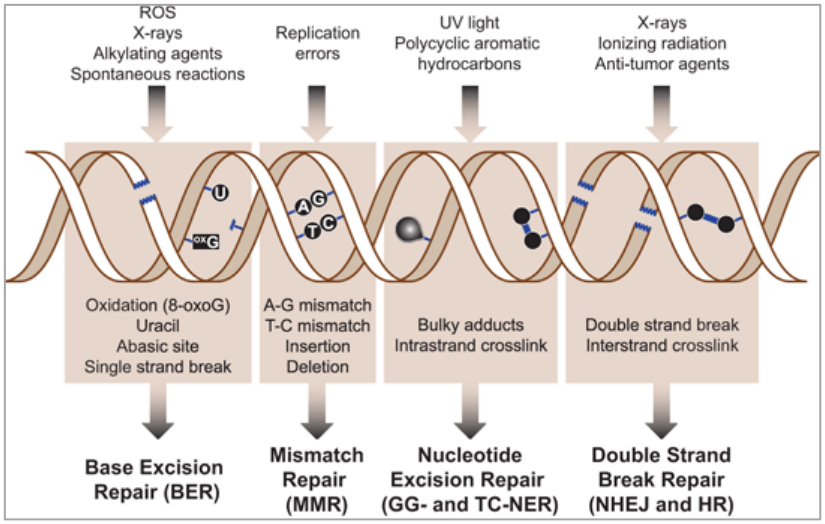


Figure 1 | **DNA damage and repair mechanisms**

Since genetic changes in DNA repair genes may impact individual variation in DNA repair capacity, a large number of SNPs in genes involved in the previously described pathways has been investigated, also in relation to BC risk or relapse/progression and survival.

Polymorphisms in BER genes potentially altering repair functions have been evaluated in a number of epidemiological studies in relation BC risk [48-51]. In 2007, Figueroa and colleagues conducted a large case-control study, demonstrating that SNPs in three BER genes, *OGG1*, *PARP1* and *POLB*, were significantly associated with BC risk. More precisely, subjects carrying the heterozygous or homozygous variant genotypes for rs125701 in *OGG1* were at significantly decreased BC risk compared to the common homozygous carriers, while a significant opposite trend was observed for heterozygous or homozygous variant carriers of rs1136410 in *PARP1* or rs3136717 in *POLB* [52]. More recently, Xie and collaborators showed an increased BC risk with an increasing number of adverse alleles in BER genes, highlighting the importance of using a multigenic approach to identify signatures of genetic variations as predictors of any cancer risk [53].

Genes involved in the NER pathway have been also deeply investigated with BC susceptibility; for example, it has been proved that individuals carrying variant alleles in *ERCC1* had a significant increase of BC risk, and the associations resulted even stronger in smokers, compared to never smokers. Authors obtained similar results for SNPs in *RAD23B*, *ERCC2* and *ERCC5* [54]. Also in this case there are evidence of increased BC risk for carriers of more than one variant allele: in particular, Wu et al. showed that smokers carrying four, five to six, and seven or more NER adverse alleles had a significant increased BC risk compared to the referent group (carriers of <4 adverse alleles) [55].

Moreover, also DSBs pathway genes may contribute to BC risk. In 2007, Figueroa and colleagues evaluated 39 DSB SNPs in seven candidate genes in a case-control study consisting of 1150 BC and 1149 controls. Authors observed a significant decreasing trend in BC risk for carriers of variants in *ZNF350* rs2278415, compared with individuals carrying the common homozygous genotype. Conversely, the variant allele in *XRCC4* rs1805377 confers a significant increase in BC risk. The association was stronger in never smokers rather than current smokers. The authors showed an additional reduced cancer risk for carriers of homozygote variants in three SNPs (rs10234749, rs6464268, rs3218373) in the promoter region of *XRCC2* [56].

1.5. Bladder cancer biomarkers

Bladder cancer is a real health problem worldwide because of its incidence, prevalence, high recurrence rate, with a long silent clinical evolution and advanced-stage detection; furthermore, BC is one of the most common malignancies and is considered to be the most expensive cancer in terms of lifetime cost per patient [57]. As a matter of fact, 70% rate of recurrence for NMIBC makes it a considerable strain to the healthcare system, due to the need for repeated monitoring by cystoscopies, while the enormous costs related to patients with metastatic disease are associated with late-stage and end-of-life care [57]. More importantly, current diagnostic techniques, such as cystoscopy and biopsies, are highly invasive and accompanied by undesirable side effects for patients.

Thus, there is a great need for new biomarkers detectable in non-invasive tissues, like blood or urine, that could help in early BC detection and might better characterize BC evolution, prognosis, survival rate, together with precise therapy to radically improve patients` outcomes. Moreover, investigating these biomarkers instead of target tissue samples could be convenient for the possibility of repeated measurements.

Several clinical and pathologic factors have been identified as predictors of tumor recurrence and progression, including stage, grade, size, multiplicity, presence of carcinoma *in situ* and previous recurrences [7].

In the last two decades, great effort has been made to evaluate the prognostic and predictive roles of several molecular markers in MIBC and in NMIBC, for which a precise risk stratification is urgently needed.

The investigation of molecules in minimally invasive biofluids such as blood, plasma or serum has open in the last years new perspectives for accurate prediction along with general applications for screening.

Among the non-invasive biomarkers there is DNA methylation. In 2011 Marsit and colleagues found 9 genes with an altered epigenetic profiles in blood associated with BC risk [58]. Interestingly, DNA methylation markers are not only capable of helping in BC diagnosis but also could efficiently differentiate between non-muscle and muscle invasiveness, progression and survival [59].

Significant markers in blood plasma for BC detection have been also identified, with an increased diagnostic value if combined together. For example, in 2006 Staack and collaborators evaluated the diagnostic value of various matrix metalloproteinases and their inhibitors (*MMPs* and *TIMPs*), showing that the highest sensitivities and specificities for a non-invasive BC detection were obtained by testing the markers in combination [60].

Several oncogenes (*TP63*, *EGFR* and *PIK3CA*) and tumor suppressor genes (*TP53*, *RB1* [61]and *PTEN* [62]) also play a role in the genetic pathogenesis of BC. For example, *EGFR* has been defined both as a prognostic BC biomarker (its overexpression has been associated to high-grade and high-stage BCs) and as independent predictor of BC survival and stage progression [61]. *PIK3CA* oncogene, instead, has been found associated to higher recurrence rates when in combination with *FGFR3* mutations [62].

Polymorphisms in *NAT2* and *GSTM1* genes (see Section 1.2) and other SNPs in candidate genes (*MYC*, *TP63*, *PSCA*, *TERT*, *CLPTM1L*, *FGFR3*, *TACC3*, *NAT2*, *CBX6*, *APOBEC3A*, *CCNE1* and *UGT1A*) have been associated with increased risk of BC, both for NMI and MI types, especially in relation to smoking.

In the last years, also SNPs in DNA repair genes [63, 64] have been identified as potential markers for BC risk estimation, prognosis and response to therapy, together with other kind of biomarkers such as leukocyte telomere length [65].

Urine biomarkers are preferred over other methods for BC surveillance and screening. There are several urine-based biomarkers that have been established due to their high specificity, such as *NMP22*, and there are other more potential molecular markers, which seem to have high specificity as well [66]. Some matrix metalloproteinases have been already identified in urine or blood of BC patients, in particular *MMP7* and *MMP9* have been detected in urine as trustworthy prognostic biomarkers associated with advanced BC stage or poor survival [67]. Also *HER2* oncogene was found to be highly expressed in BC; in addition, overexpression of HER2 protein was shown to correlate with increased tumor grade, survival, and incidence of metastatic disease [61].

Despite the huge number of identified BC biomarkers, there are still conflicting results among studies and a lack of validation of results among studies and across different populations. Due to these reasons, no

molecular markers have been included yet in the clinical practice. Only few examples, like p53, pRb, p21 and survivin, have proved to have a predictive value in studies with a large and homogeneous patient population undergoing standardized treatment [68].

2. Hypotheses and Aims

The term DNA repair capacity (DRC) refers to the ability of a cell to protect the integrity of the genome. It has been observed that individuals with low DRC tend to accumulate more damage than those with a more efficient DRC. This observed inter-individual variability is modulated by the genetic background, as well as differential genetic and epigenetic regulation.

The aim of the study was to investigate the relationship between DNA damage, DRC, and BC risk and clinical outcome in 159 BC cases and 159 matched controls, with an integration of data from whole genome gene expression, as well as genetic and epigenetic profile data.

In this context, we started with the evaluation of DNA damage and DRC in cryopreserved peripheral blood mononuclear cells (PBMC) by 3 different DNA repair assays: γ-H2AX phosphorylation, micronucleus and comet. Furthermore, in order to identify additional reliable biomarkers of BC risk, we examined more in depth the impact of differential gene expression and methylation in BC cases and controls.

3. Study population

The study population included newly diagnosed, histologically confirmed cases of BC registered at two urology departments of A.O.U. Città della Salute e della Scienza, in Turin (Italy), during the years 1994–2008. All subjects were men, aged 40-75 years, living in the Turin metropolitan area and were enrolled for a large hospital-based study on BC: the Turin Bladder Cancer Study (TBCS) [48, 63]. Before any treatment, a trained interviewer used a detailed questionnaire to conduct a face-to-face interview.

Controls were men recruited during the same period of cases in random fashion from patients treated at the same urology departments for non-neoplastic diseases (prostatic hyperplasia, cystitis and others) or from patients treated at the medical and surgical departments for hernias, vasculopathies, diabetes, heart failure, asthma or other benign diseases. Patients with any cancer, liver or renal diseases and smoking-related pathologies were excluded.

All subjects signed an informed consent form according to the Helsinki declaration. The design of the study was approved by the local Ethics Committees. Clinical information, including the type of therapy (e.g., Bacillus Calmette-Guerin (BCG), chemotherapy and radiotherapy) and risk category for NMIBC (based on the six most significant clinical and pathological factors: number of tumors, tumor size, prior recurrence rate, T category, presence of concurrent CIS, and tumor grade (WHO1973) as described in [5, 69]), was recorded through the perusal of clinical records in collaboration with urologists at A.O.U. Città della Salute e della Scienza, in Turin (Italy). Patients treated with one instillation of chemotherapy immediately after transurethral resection were not considered in the chemotherapy group. Clinical end-points (progression and recurrence) of BC patients were retrieved from their medical files. Progression was defined as the transition from non-muscle invasive status to invasive or metastatic disease, while recurrence was defined as a newly identified bladder tumor after a previous negative follow-up cystoscopy. Vital status was ascertained through linkage with the local demographic office and death certificates were retrieved to identify the specific causes of death.

4. DNA repair assays

In the next sections I report results published or in press in the following journals.

- Turinetto V, Pardini B, Allione A, Fiorito G, **Viberti C**, Guarrera S, Russo A, Anglesio S, Ruo Redda MG, Casetta G, Cucchiarale G, Destefanis P, Oderda M, Gontero P, Rolle L, Frea B, Vineis P, Sacerdote C, Giachino C, Matullo G, *H2AX phosphorylation level in peripheral blood mononuclear cells as an event-free survival predictor for bladder cancer.* Molecular carcinogenesis, 2016. **55**(11): p. 1833-18424.1.

- Pardini B*****, **Viberti C***, Naccarati A, Allione A, Oderda M, Critelli R, Preto M3, Zijno A, Cucchiarale G, Gontero P, Vineis P, Sacerdote C, Matullo G, *Increased micronucleus frequency in peripheral blood lymphocytes predicts the risk of bladder cancer.* British journal of cancer, 2017. **116**(2): p. 202-210.
- *** equally contributed**

- Allione A, Pardini B, **Viberti C**, Oderda M, Allasia M, Gontero P, Vineis P, Sacerdote C, Matullo G, *The prognostic value of basal DNA damage level in peripheral blood lymphocytes of patients affected by bladder cancer.* Urologic oncology: seminars and original investigations, 2018 Feb 6 pii: S1078-1439(18)30006-1.

4.1. Phosphorylation of the histone H2AX (γ-H2AX) assay

The maintenance of genome integrity is essential for human cells. The DNA damage response (DDR) pathway, which aims to restore the DNA to its original configuration, is a set of biochemical signaling and effector pathways that maintain genomic stability [70]. DSB lesions, i.e. complete breaks of the DNA double helix, may impair DNA replication, transcription or distribution of the genetic material to daughter cells. DSBs are difficult to repair and since broken DNA ends tend to dissociate, unrelated ends may re-join thus allowing for gross loss or amplification of genomic information, as well as chromosomal rearrangements. All these kinds of damage are commonly associated with the early stages of cellular transformation and tumorigenesis [71].

Among the different DNA DSB markers, the most well-characterized is phosphorylation of the histone H2AX (γ-H2AX). Within minutes after the induction of a DSB, several thousands of H2AX molecules are phosphorylated near the site of damage with a crucial role in recruiting and maintaining many factors involved in DNA DSB repair [72]. Nonetheless, γ-H2AX has also been suggested as a biomarker to predict patient response to specific radiotherapy treatments [73-75]. In the last years, γ-H2AX has become a widely

used marker for DSB damage in translational studies, including areas such as carcinogenesis, radiation research, drug development and clinical assessment of DNA targeted anticancer therapies [76-78].

BC has a typical etiology characterized by a multistep carcinogenesis, reflecting that multiple lesions in the DNA are required for tumor development. More specifically, exposure to carcinogens such as tobacco smoking, occupational exposure to aromatic amines, ionizing radiation (IR) and arsenic-contaminated drinking water causes different types of DNA damage, including DSBs [79]. For all these reasons, the identification of susceptibility to DSB may provide valuable information about individual cancer risk [80].

4.1.1. Material and Methods

Study population

The complete study population has been described in section 3.

Lymphocytes isolation and storing

Ten ml of heparinized venous blood were collected from all subjects. PBMCs were separated by centrifugation with FicollPaque PLUS (GE Healthcare, Milan, Italy) at 400 x g for 30 minutes at room temperature. After two washes in RPMI 1640, 1% fetal bovine serum (FBS), 25 mM EDTA (all from Gibco, Life Technologies, Paisley, UK), PBMCs were prepared for cryopreservation. They were suspended at 10x106 cells/ml in freezing medium (RPMI 1640, 50% FBS, 10% DMSO), aliquoted in cryovials and slowly frozen overnight at the rate of –1°C/minute in isopropyl alcohol to –80°C (Mr. Frosty containers, Nalgene, Roskilde, Denmark). Cryovials were then transferred into liquid nitrogen for long-term storage.

Case-control study

Cryopreserved PBMCs from 146 cases and 146 controls matched by gender, age (±1 year) and smoking habits, were selected from the TBCS collection and employed for the phenotypic assay to evaluate H2AX phosphorylation.

H2AX phosphorylation assay

PBMCs were thawed and allowed to recover overnight in RPMI 1640 (Gibco, Life Technologies) supplemented with 10% heat inactivated FBS (Sigma-Aldrich St Luis, MO, USA), 2mM L-Glutamine, 1% kanamycin, 1% sodium pyruvate, 1% nonessential amino acids, 0.1% β-mercaptoethanol (all from Gibco, Life Technologies). Cells were x-irradiated (2 Gy) using 6 MV opposed beams with a linear accelerator Synergy Platform (Elekta, Stockholm, Sweden). They were kept on ice for 1 h, and then returned to the incubator and harvested at 1 h and 3 h. Approximately 400,000 untreated and irradiated cells were collected, fixed in cold 70% ethanol and stored at −20°C for up to 2 weeks before analysis. Cells were washed in Tris-buffered saline (TBS) pH 7.4 and then rehydrated for 10 min at 4 °C in TBS containing 4% FBS and 0.1% Triton X-100 (TST) (Sigma–Aldrich Co) prior to staining with anti-γH2AX mAb (clone JBW301, Millipore, Billerica, MA, USA) diluted at 1:250 in TST, and incubated for 2 h at 37 °C. After two washes in TBS, they were suspended in PE-conjugated goat anti-mouse IgG1 (BD PharMingen, Becton Dickinson & Co., Franklin Lakes, NJ, USA) diluted at 1:100 in TST as a secondary antibody, and shaken for 1 h at room temperature in the dark. A minimum of 10,000 stained cells were acquired on a Cyan ADP flow cytometer (Beckman Coulter, Brea, CA, USA) and analyzed with Summit 4.3 software. In order to normalize the results, aliquots of PBMCs derived from the same subject were used at each irradiation and staining step as reference control and flow cytometer setting relative to PE-fluorescence were established on the reference control untreated sample. Each experimental point was evaluated in technical triplicate.

For the analysis, the mean of the reference control at each experimental point (untreated, 2 Gy 1 h and 2 Gy 3 h) was calculated considering all the irradiation experiments. Raw data were subsequently normalized to the average signal of the reference control of each specific irradiation (Average normalization procedure). Data corresponding to the amount of γ-H2AX at 1 h were expressed as ratio of median γ-H2AX level at 1 h to the baseline γ-H2AX level of the same non-irradiated sample. Data corresponding to the dephosphorylation of γ-H2AX after 3 h of recovery were expressed for each sample as percentage of median γ-H2AX fluorescence of irradiated cells at 3 h on median γ-H2AX fluorescence at 1 h.

Statistical analyzes

The patient characteristics were described as absolute frequencies for qualitative variables, and median, minimum and maximum values for quantitative variables. The analyzed variables were: basal damage in untreated cells, H2AX phosphorylation at 1 h (γ-H2AX ratio), and percentage of dephosphorylation after 3 h. Eleven samples encountered technical problems during the assay performance, therefore they were excluded from further analysis together with their relative matched counterpart. Fisher exact test was used to ensure no under/over-representation of cases or death in the excluded samples.

For each of the considered variables we tested whether they were normally distributed by the Shapiro-Wilk test. Since the variables were not normally distributed, non-parametric paired test (two-sided Wilcoxon Rank Sum test) was employed to verify differences between cases and controls.

Overall Survival (OS) in patients was evaluated as the time from diagnosis using the date of death (regardless of the cause) or the date of follow up termination as the end point. Event-free survival (EFS) was defined as the time from diagnosis to the occurrence of any event such as recurrence, progression or death, whichever came first. Investigated variables were considered as continuous. The survival curves for OS and EFS were derived by the Kaplan–Meier (KM) method and statistical significance was determined using the log-rank test (R Survival package). The relative risk of death or of recurrence/progression was estimated as hazard ratio (HR) using Cox regression. Multivariate survival analyzes were adjusted for matching variables, risk category (based on the six most significant clinical and pathological factors: number of tumors, tumor size, prior recurrence rate, T category, presence of concurrent CIS, and tumor grade (WHO1973) as described in [81, 82]), and therapy for the whole group of patients. The investigated variables that resulted significantly associated with OS or EFS were also considered as categorical. Several cut-off values for the categorization of continuous variables were considered, and the one maximizing the area under the ROC curve (AUC) was chosen as the best cut-off.

All the analyzes were performed with the open source R, version 3.0.3.

4.1.2. Results

Study population

The study included 146 males newly diagnosed, previously untreated BC patients and 146 control subjects, matched by age, gender and smoking habits. Details on epidemiological data, tumor histological classification and therapy are shown in **Table 3**.

During H2AX phosphorylation assay, technical problems were encountered for four controls and seven cases. Since no additional aliquots of cryopreserved PBMC were available, the experiments could not be repeated. Therefore, these samples were excluded from the analysis together with their relative matched samples. Finally, 135 matched couples were included in the case-control study and 139 cases were included in the overall and event-free survival analyzes.

Case-control study

H2AX phosphorylation assay results are reported in **Table 4**. We considered the basal H2AX phosphorylation as the basal individual level of DSB signaling, the H2AX phosphorylation after 1 h irradiation as the peak of DNA damage recognition and percentage of H2AX dephosphorylation after 3 h irradiation as the individual capacity to repair the induced damage. No significant differences were observed between cases and controls. Similarly, among all subgroups stratified for smoking habit, no significant differences were observed between cases and controls. There was only a significant difference in the percentage of H2AX dephosphorylation between cases and controls in smokers (p=0.05, **Table 5**); however, these data refer a very small population (n=22) and for this reason the result has to be considered with caution.

Table 3 | **H2AX assay: characteristics of the study population**

			BC cases*	Controls*
		N	146	146
	Gender	Male	146	146
		Female	0	0
	Age (years)	I quartile (≤ 57)	36	37
		II quartile (58-64)	38	34
		III quartile (65-68)	35	34
		IV quartile (≥69)	37	41
	Smoking	Never smokers	22	22
		Ex smokers	90	90
		Current smokers	34	34
Tumor classification	Tumor type**	Non Muscle Invasive	121	-
		Muscle invasive	13	-
		NA	12	
	Tumor stage	is	3	-
		a	75	-
		1	41	-
		>2	13	-
		X	2	-
		NA	12	-
	Grade[#]	High grade	60	-
		Non-High grade	56	-
		NA	30	
	Grade[#]	1	31	-
		2	56	-
		3	40	-
		NA	19	
	Risk[##]	High	50	-
		Intermediate	41	-
		Low	30	-
		NA	12	
Therapy	Bacillus Calmette-Guerin (BCG)		39	-
	Chemotherapy		24	-
	No therapy		69	
	NA		14	
Survival		Deceased (all causes)	24	-
		Deceased (for cancer)	10	-
		Progression***	4	-
		Recurrence***	49	-

*Numbers may not add up to 100% of available subjects because of missing information
** and [##] refer to [5]
# Cheng L. Cancer, 2000; 88(7):1513-6; Montironi R.and Lopez-Beltran A. Int J Surg Pathol, 2005, 13(2):143-53
*** one patient develop both progression and recurrence

Table 4 | **H2AX phosphorylation and dephosphorylation in PBMCs from BC cases and healthy controls** (Wilcoxon test)

	Basal H2AX phosphorylation			**H2AX phosphorylation at 1 h (ratio)**			**H2AX dephosphorylation at 3 h (%)**		
	Cases	**Controls**	**P-value**	**Cases**	**Controls**	**P-value**	**Cases**	**Controls**	**P-value**
N	135	135		135	135		135	135	
Median	5.96	6.10	0.07	9.16	9.06	0.17	17.19	17.09	0.19
Min	3.93	4.00		5.94	5.80		0	0	
Max	11.26	11.53		17.23	13.62		48.73	51.72	

Table 5 | **H2AX phosphorylation and dephosphorylation in PBMCs from BC cases and healthy controls stratified for smoking habit** (Wilcoxon test) (median values)

	Smokers			**Ex-smokers**			**Never smokers**		
	Cases	**Controls**	**P-value**	**Cases**	**Controls**	**P-value**	**Cases**	**Controls**	**P-value**
N	22	22		82	82		31	31	
Basal H2AX phosphorylation	5.68	6.1	0.22	6.17	6.27	0.27	5.85	5.78	0.43
H2AX phosphorylation at 1 h (ratio)	9.16	9.05	0.12	9.12	9.06	0.54	9.11	9.5	0.57
H2AX dephosphorylation at 3 h (%)	**17.64**	**16.15**	**0.05**	16.57	16.3	0.75	18.57	20.01	0.94

Significant results in bold

Overall and event-free survival analyzes

Results from the survival analysis are reported in **Table 6**. There was no association between OS in patients and γ-H2AX assay results. On the contrary, when analyzing the EFS, we found a significantly decreased risk of recurrence/progression in patients with an increased basal H2AX phosphorylation level both for all BC

patients and also when stratified only for NMIBC (HR 0.70, 95% CI 0.52-0.94, p=0.02 and HR 0.68, 95% CI 0.50-0.92, p=0.01, respectively).

We then repeated the analysis considering the significant variable as categorical (**Table 7**). Instead of separating samples according to the median value, we used the optimal cut-off value based on the AUC (R OptimalCutpoints package). A similar association between decreased risk of recurrence and basal H2AX phosphorylation above the calculated best cut-off was observed for all BC and NMIBC subjects: patients with high basal DSB signaling had a better EFS (HR 0.35, 95% CI 0.20-0.63, p=0.0005 and HR 0.32, 95% CI 0.17-0.62, p=0.0006, respectively; **Table 7**).

The results on EFS were confirmed also in the univariate KM analysis for the same model (log rank test p-value=0.004 and 0.002, respectively for all BC and NMIBC, for the basal H2AX phosphorylation; **Figure 2A** and **B**). For EFS, the best cut-offs obtained to perform the analyzes for the basal H2AX phosphorylation were 6.33 (considering all BC cases) and 6.29 (considering NMIBC cases).

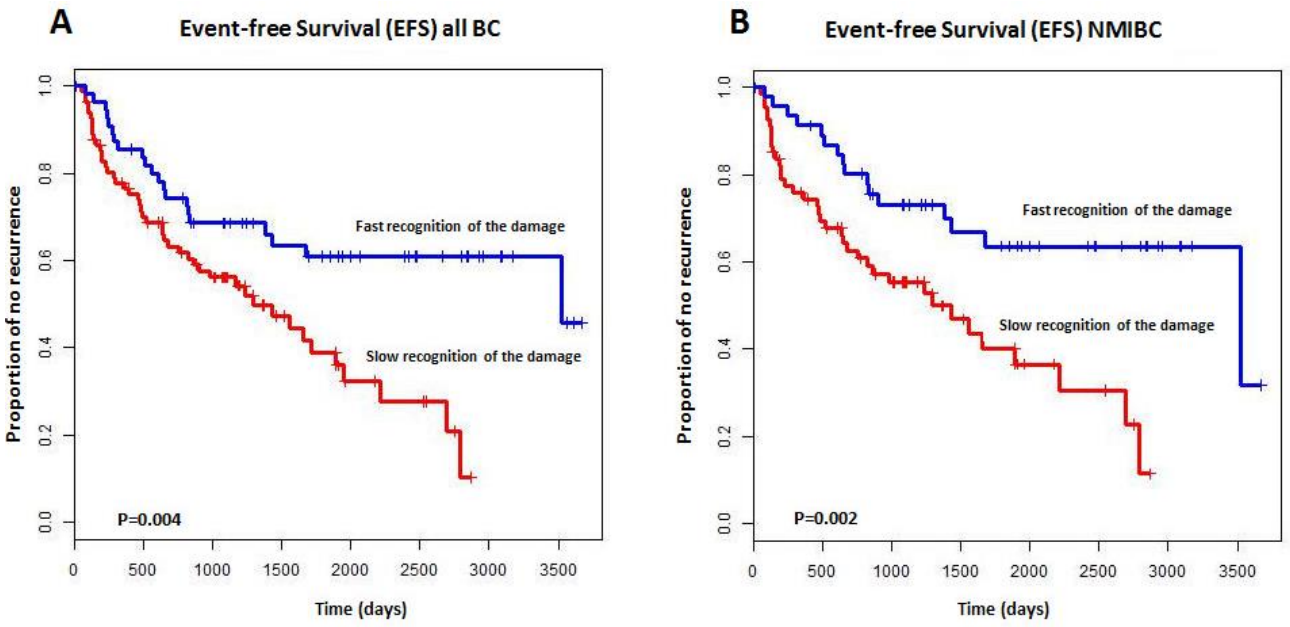


Figure 2 | **Kaplan–Meier event-free survival (EFS) curves according to the basal H2AX phosphorylation for all bladder cancer patients (n=139) (A) and for non-muscle invasive bladder cancer patients (n=124) (B)**. Patients were stratified into two categories: those with the basal H2AX phosphorylation above the best cut-off value correspond to the "Fast recognition of the damage" group (for A n=56; for B n=47; Median event-free time for A and B (MET)=9.65yr); those under the best cut-off value correspond to the "Slow recognition of the damage" group (for A n=83; for B n=67; MET for A and B=3.56yr)

Table 6 | **H2AX phosphorylation assay affecting Overall Survival (OS) and Event Free Survival (EFS) in BC patients and NMIBC patients** (Cox regression) (continuous variables)

	All BC patients*						**NMIBC patients****					
	Overall survival			**Event-free survival**			**Overall survival**			**Event-free survival**		
	HR	**95% CI**	**P-value**	**HR**	**95% CI**	**P-value**	**HR**	**95% CI**	**P-value**	**HR**	**95% CI**	**P-value**
Basal H2AX phosphorylation	1.1	0.66-1.88	0.69	**0.7**	**0.52-0.94**	**0.02**	1.4	0.71-2.69	0.34	**0.7**	**0.50-0.92**	**0.01**
H2AX phosphorylation at 1h (ratio)	1.1	0.77-1.44	0.75	1.1	0.93-1.25	0.32	0.9	0.58-1.47	0.74	1.1	0.91-1.25	0.41
H2AX dephosphorylation at 3h (%)	1	0.93-1.04	0.54	1	0.95-1.01	0.26	1	0.91-1.06	0.71	1	0.94-1.01	0.17

Significant results in bold

Adjusted for age, smoking, risk and therapy. HR= hazard ratio; 95% CI= 95% confidence interval

*Analyzes done on 134 cases for which all covariates were available

** Analyzes done on 121 cases for which all covariates were available

Table 7 | **H2AX phosphorylation assay affecting Event Free Survival (EFS) in BC cases and NMIBC stratified according to the optimal cut-off value based on the AUC**

	All BC patients*				NMIBC**			
	Best cut off	**HR**	**95% CI**	**P-value**	**Best cut off**	**HR**	**95% CI**	**P-value**
Basal H2AX phosphorylation	<6.33	Ref	-	-	>6.29	Ref	-	-
	≥6.33	**0.35**	**0.20-0.63**	**0.0005**	≤6.29	**0.32**	**0.17-0.62**	**0.0006**
H2AX phosphorylation at 1 h (ratio)	<9.35	Ref	-	-	>8.01	Ref	-	-
	≥9.35	1.67	0.97-2.88	0.06	≤8.01	1.79	0.88-3.67	0.11
H2AX dephosphorylation at 3 h (%)	<17.28	Ref	-	-	>15.73	Ref	-	-
	≥17.28	0.69	0.39-1.21	0.19	≤15.73	0.62	0.34-1.14	0.12

Significant results in bold
Adjusted for age, smoking, risk and therapy.
*Analyzes done on 134 cases for which all covariates were available
**Analyzes done on 121 cases for which all covariates were available

4.1.3. Discussion

In the present study we observed a significant association between H2AX basal phosphorylation level and BC recurrence risk. In particular, individuals with higher basal H2AX phosphorylation had a significantly decreased risk of recurrence, suggesting a protective effect of high basal DSB signaling in terms of preventing BC recurrences. One may speculate that individuals expressing high basal levels of γ-H2AX are capable of recruiting crucial DNA repair elements towards areas of DSBs and effectively activating the DNA repair pathway. Interestingly, a study on H2AX phosphorylation in non-invasive low grade urothelial carcinoma (LG-URC) of the bladder demonstrated a significant higher recurrence rate in γ-H2AX negative tumor tissues, suggesting a protective effect of γ-H2AX expression in terms of preventing LG-URC recurrences [83]. Surprisingly, we did not observe any association between the 1 h H2AX phosphorylation peak and recurrence risk, although a borderline association was evidenced (for all BC HR 1.67, 95% CI 0.97-2.88, p=0.06; for NMIBC HR 1.79, 95% CI 0.88-3.67, p=0.11).

Additionally, considering the H2AX dephosphorylation capacity after 3 h from IR, individuals with higher dephosphorylation capacity had a decreased risk of recurrence, although not reaching a statistical significance. These findings suggest a protective effect of a rapid response in recognizing (basal H2AX phosphorylation) and repairing (3 h H2AX dephosphorylation) DNA damage, possibly preventing BC recurrence.

Using the measured variables as categorical with a best cut-off analysis strategy, we obtained that the best cut-off to use was different from the median, allowing to hypothesize application of this newly determined value on other BC populations and to propose the H2AX phosphorylation assay for clinical validation on a large scale.

Interestingly, we obtained analogous results restricting our analysis to NMIBC patients. These findings are particularly relevant, considering that NMIBC is the most prevalent BC that recurs more frequently and often progresses to MIBC, a deadly disease.

As nearly half of patients with BC experiences recurrence, reliable predictors of this phenotype are needed to guide surveillance and treatment, thus the identification of γ-H2AX as a new molecular biomarker related to BC recurrence is particularly interesting.

In the last twenty years several tissue- and blood-based biomarkers have been discovered (reviewed in [84, 85]). Although none of these biomarkers have reached the clinical validation level, for some of them (e.g., p53, pRb, p21, and survivin) the predictive value has been proved in studies that included a homogeneous patient population treated with standard therapy, and, therefore, they are probably ready for clinical validation on a larger scale and could help clinicians to provide individualized prognosis and allow risk-stratified clinical decision.

DNA repair pathways may be also involved in the pathogenesis, progression, and treatment response of BC [86]. Many studies analyzed predictive value of different DNA repair factors by immunohistochemical tissue analysis, but no data are available on molecular biomarkers in PBMCs to predict recurrence risk. Besides the previously described work that analyzed H2AX phosphorylation in BC tissue [83], to the best of our knowledge there are no other studies investigating the role of γ-H2AX in the determination of BC survival

and/or clinical outcome. The present study was the first one employing an assay able to measure not only the peak of DNA damage recognition after irradiation, but also the individual capacity to recover and repair the induced damage. Another study analyzed protein expression of MRE11, RAD50, NBS1, ATM, and H2AX by immunohistochemistry in pre-treatment bladder tumor specimens from three different cohorts to evaluate if predictive markers could be identified [87]. Authors demonstrated that patients with increased MRE11 expression levels had longer survival in comparison to patients with lower values. It must be pointed out that previously published data regarding DNA damage factors related to BC referred mainly to immunohistochemical techniques on bladder tissues, while our results are based on the detection of a basal damage that could affect patient specific BC progression.

In the present study, we did not observe any differences between cases and controls using the H2AX phosphorylation assay. In contrast, other case-control studies reported differential γ-H2AX levels in PBMCs as a risk predictor for BC or other neoplastic diseases, such us lung cancer and esophageal adenocarcinoma, demonstrating higher IR-induced γ-H2AX levels in cases than in control [88-90]. However, there are some important differences between previous studies and the present one. First, we used unstimulated PBMCs, whereas in previous studies 72 h PHA-stimulated PBMCs were employed. In this sense, the different results could be ascribed to the different responses in quiescent vs proliferating cells. Second, we detected γ-H2AX by flow cytometry, while in other works a laser scanning cytometer-based immunocytochemical method was used; these two techniques could have a different sensitivity. Finally, controls were recruited by us in a random fashion from patients treated at the same urology department for benign diseases, mainly prostate hyperplasia and cystitis (all newly diagnosed), or from patients treated at the medical and surgical departments for hernias, vasculopathies, diabetes, heart failure, asthma, or other benign diseases. In contrast, in previous studies most of the controls were enrolled during work annual health check-up.

In conclusion, our data suggest that the H2AX phosphorylation assay applied on cryopreserved PBMCs can be considered as a possible useful molecular marker to identify BC patients with a higher risk of recurrence or progression. Our results support the inclusion of the DNA H2AX phosphorylation assay in future screening trials focused on high-risk cohorts in order to further assess its prognostic value in BC patients.

4.2. Cytokinesis-Block Micronucleus (CBMN) assay

Evidence that cytogenetic biomarkers are positively correlated with cancer risk has been strongly validated in both cohort and nested case-control studies, leading to the conclusion that chromosome aberrations are a relevant marker of cancer risk [91], which reflects the outcome of both the genotoxic effects of carcinogens and the genetic host susceptibility. The cytokinesis-block micronucleus (CBMN) assay is a relatively fast and easy technique extensively used in molecular epidemiology and cytogenetics. Its application is widely aimed at evaluating the presence and the extent of chromosomal damage in human populations exposed to genotoxic agents or bearing a susceptible genetic profile in several cell types, including newly divided PBMC [92]. The measurement of micronucleus (MN) frequency in PBMC has been also successfully applied to identify dietary and genetic factors that have a significant impact on genome stability. The high reliability and low cost of the CBMN technique has contributed to the worldwide success and adoption of this biomarker for *in vitro* and *in vivo* studies of genome damage [92].

MN originate from chromosome fragments or whole chromosomes that are not included in the main daughter nuclei during nuclear division. The MN formation in dividing cells derives from chromosome breakage due to unrepaired or misrepaired DNA lesions or chromosome malsegregation due to mitotic malfunction. These events may be induced by oxidative stress, exposure to clastogens or aneugens, genetic defects in cell cycle checkpoint and/or DNA repair genes, as well as deficiencies in nutrients required as cofactors in DNA metabolism and chromosome segregation machinery [93]. Therefore, the CBMN assay can be considered a possible test to evaluate the level of DNA damage in the cells.

The importance of patient's basal genetic background and consequently of his damage repair capacity in predicting BC recurrence is highlighted by different studies where genetic variants in DNA repair genes that could modify BC prognosis were identified, including polymorphisms in *XRCC1*, *XRCC4*, *XPF*, *XPD* and *XPC* [94-97].

In BC, it is highly improbable that a single marker can accurately segregate tumors into precise prognostic categories. Therefore, as already shown by different investigators, combination of independent,

complementary biomarkers may provide a more accurate prediction of outcome compared with any single markers [98, 99].

4.2.1. Material and Methods

Study population

The complete study population has been described in section 3.

Lymphocytes isolation and storing

The complete technique has been described in section 4.1.1.

Cytokinesis-Block Micronucleus (CBMN) assay

Cryopreserved PBMC from 158 cases (collected before treatment) and 158 controls matched by age (±1 year) and smoking habits, were selected from the TBCS collection and processed for the CBMN assay. Cell lymphocytes cultures were set up according to [7]. One aliquot of cryopreserved cell suspension was thawed quickly in a water bath at 37°C and suspended in 5 ml cold medium containing 50% FBS, 49% RPMI and 1% of dextrose. Cells were spun down by centrifugation (10 min at 1000 rpm) at 4°C, counted on a slide with a Burker chamber by Trypan Blue dye exclusion and suspended at 0.5 × 106 viable cells/ml in 0.75 ml RPMI 1640 containing 15% FBS and 2% phytohaemoagglutinin (PHA) (Sigma-Aldrich). For each blood sample at least 2 cultures were set up in round-bottomed tissue cultures tubes. For each experiment 7-8 samples were processed.

Lymphocyte cultures were incubated at 37°C with loose lids in a humidified atmosphere containing 5% CO2. Forty-four hours after PHA stimulation, cytochalasin B (Cyt B) (Sigma-Aldrich) was added at the final concentration of 4.5 µg/ml. Cells were collected onto slides by Shandon Cytospin® 3 (ThermoFisher) 72 h after stimulation, then fixed for 10 min in absolute methanol and stained in 5% Giemsa (Carlo Erba) solution. For each subject, 1000 binucleated cells were scored from untreated cultures (500 cells from each of two cultures). Scoring of binucleated cells with MN, nucleoplasmic bridges (NPB) and nuclear buds (NBUD) was performed by two independent scorers following the criteria developed by Fenech [93].

Genotyping and Quality Controls

DNA was isolated from PBMC with the Qiagen DNeasy blood and tissue kit (Qiagen) according to the manufacturer's protocol.

Whole-genome genotyping was performed on an Illumina HumanOmniExpressExome-8 v1.2 (containing 964193 SNPs) for 306 subjects. Genotypes were assessed by GenomeStudio V2011.1 (Illumina).

A cut-off for a genotyping call rate of 0.90 was set, leading to the exclusion of 3 study subjects. The Identity By Descent (IBD) estimation was used to check relatedness among subjects (R 3.0.3, SNPRelate package). Subjects with kinship coefficient >0.0625 (2 couples) were considered consanguineous and excluded (one per couple) from further analyzes. Population structure was investigated through Principal Component Analysis: two outlier subjects were found and then removed, leaving 299 subjects (152 cases and 147 controls).

SNPs having >5% missing genotypes (n = 7854), deviating from Hardy-Weinberg equilibrium (HWE) ($P<0.0001$, n = 378) or with minor allele frequency <5% (n=339031) were excluded from the analysis, for a final study dataset of 616454 SNPs. Over a total of 5412 SNPs in DNA repair genes present in the Illumina HumanOmniExpressExome-8 v1.2 BeadChip, 1457 passed the above described filters. Among them, 783 were selected after linkage disequilibrium (LD) analysis (R 3.0.3, SNPRelate package, LD threshold≥0.8) and analyzed for their distribution among cases and controls and for their potential association with CBMN frequencies.

Statistical analysis

All the analyzes were performed with the open source R (R 3.1.0). The influence of possible explicative variables on the frequency of MN, NBUD and NPB was evaluated by multiple regression analysis and/or non-parametric statistical hypothesis test, since variables did not follow a normal distribution (Shapiro-Wilk normality test).

The Wilcoxon rank sum test was applied to compare the distribution of MN, NBUD and NPB between groups. Correlation analysis and Kruskal-Wallis rank sum test were also performed to estimate associations among independent variables (smoking habits, age, batch effect, etc).

OS was evaluated for each patient calculating the time (in months) between the date of BC diagnosis and the date of death or follow up termination as the end point. EFS was calculated as the time (in months) between the date of BC diagnosis until date of relapse, death or censorship, whichever came first. The relative risk of death or of recurrence/progression against each of the 3 CBMN endpoints was estimated as hazard ratios (HR) using unadjusted Cox regression (R Survival package). MN, NBUD and NPB were all considered continuous or categorical; multivariate survival analyzes were adjusted for matching variables, risk category and therapy for the whole group of patients. The prognostic role of CBMN endpoints on survival was also evaluated using KM curves and log-rank test.

The distribution of genotypes of selected DNA repair variants among NMIBC cases and controls was evaluated in two steps, first by Chi-Square Test followed by regression analysis, adjusting for age and smoking habits.

The genotype influence on the frequency of micronucleated per 1000 binucleated cells was assessed by Kruskal-Wallis rank sum test only on NMIBC and in the same way as described above and then confirmed by regression analysis, adjusted for the same match variables.

Finally, different prediction models were built to evaluate the diagnostic power of investigated markers: 1) a model including age and smoking status (Model A); 2) Model A with the additional predictor "SNPs with different genotype frequencies among cases and controls" (Model B); 3) Model B with the inclusion of MN frequencies (Model C). Discrimination analysis was conducted comparing the area under ROC curve (AUC) of the three nested models by the DeLong's test (R pROC package).

4.2.2. Results

Study population

The study included PBMC from 158 cases and 158 controls matched by age and smoking habits. For 10 controls and six cases, technical problems were encountered (i.e. slides were not scorable or the number of counted binucleated cells was not sufficient). Since no additional aliquots of cryopreserved PBMC were

available these subjects were excluded from the analysis. Finally, 152 newly diagnosed, previously untreated males BC patients and 148 control subjects were included in the study. The characteristics of the whole cohort are summarized in **Table 8**.

Table 8| **CBMN assay: baseline characteristics of the BC patients and information at follow up**

Covariates	Controls (%)	Cases (%)
N	148	152
Non-muscle invasive BC	-	136 (89.5)
Muscle invasive BC	-	16 (10.5)
Age (years)		
Mean ± SD	62.8± 8.7	62.3 ± 7.9
Range	40.3-74.9	40.0-74.1
Smoking Status		
Never	24 (16.2)	21 (13.8)
Former	90 (60.8)	94 (61.8)
Current	34 (23.0)	37 (24.4)
T stage*		
Ta	-	84 (55.3)
T1	-	46 (30.3)
T2	-	14 (9.2)
T3	-	1 (0.7)
T4	-	1 (0.7)
Tis	-	3 (1.9)
Tx	-	3 (1.9)
Grading (1973)*		
G1	-	41 (27.0)
G2	-	64 (42.1)
G3	-	47 (30.9)
Grading (2004)*		
Non-high grade	-	77 (50.7)
High grade	-	75 (49.3)
Risk*		
Low-risk	-	37 (24.4)
Intermediate Risk	-	45 (29.6)
High-risk	-	54 (35.5)
Muscle invasive		16 (10.5)
Recurrence		
No	-	97 (63.8)
Yes	-	55 (36.2)
Number of recurrences		
1	-	35 (63.6)
2	-	15 (27.3)
≥3	-	5 (9.1)
Progression		
No	-	146 (96.1)
Yes	-	6 (3.9)
Therapy		
No	-	47 (30.9)
Yes	-	105 (69.1)
Cystectomy		
No	-	125 (82.2)
Yes	-	27 (17.8)
BC patient` status at follow up		
Alive	-	128 (84.2)
Died Of which for BC	-	24 (15.8) 11

*referred to [5, 100]

CBMN endpoints

The 3 CBMN endpoints analyzed, MN, NBUD and NPB, were not normally distributed (p=6.1e-09, p=3.4e-13 and p<2.2e-16, respectively), therefore non-parametric tests were employed to verify differences among groups.

Overall, cases had a mean MN ±Standard Deviation (SD) frequency of 9.51±4.73 compared with 7.73±3.91 of controls. The mean NBUD ±SD frequency in cases was 5.06±3.36 compared with 3.99±2.44 of controls. MN and NBUD frequencies were significantly higher in BC cases than in controls (Wilcoxon rank sum test, p=0.001 and p=0.006, respectively; **Table 9, Figure 3A** and **B**). When stratifying for the type of tumor and analyzing only NMIBC, the differences in MN and NBUD frequencies among cases and controls were even more significant (p=0.0005 and p=0.002, respectively; **Table 9, Figure 3C** and **D**). For MIBC the MN frequency trend was similar although results were not statistically significant. No differences in CBMN endpoints were observed after stratification for grade, tumor stage, risk class or number of recurrences (data not shown).

In a logistic regression model adjusted for age and smoking similar results were obtained for MN and NBUD frequencies (**Table 10**). In particular, for each increase of one unit in the MN frequency we observed a 1.11 increase of risk of developing BC (all cases versus controls) and a 1.12 for NMIBC cases. Similar results were observed when testing NBUD frequencies (1.14 and 1.16 increase of risk, respectively).

MN, NBUD and NPB significantly correlated each other but they did not correlate with smoking or age (data not shown). A significant decreased frequency of MN was found in current smokers when compared with never smokers both when considering all subjects (adjusted model for age and case-control, p=0.03) or only cases (adjusted model for age, p=0.05). BC cases showed higher MN frequencies in all subgroups with different smoking habits, though only former smokers showed significant differences in MN and NBUD levels between cases and controls (p=0.003 and p=0.009, respectively; **Figure 4A** and **B**). Similarly, when NMIBCs were compared with controls, there was an increased frequency of MN in former smokers (p=0.001; **Figure 4C**) and increased frequency of NBUD in current and former smokers (p=0.04 and p=0.004,

respectively; **Figure 4D**). No significant associations with survival or recurrence risk were observed for any of the CBMN endpoints analyzed (data not shown).

Table 9 | **Frequencies of analyzed endpoints by the CBMN assay in PBMC of BC cases and healthy controls**

		Controls	All BC Cases	p-value[b]	NMIBC	p-value[b]	MIBC	p-value[b]
Frequencies[a]	**N**	148	152		136		16	
MN	**Mean±SD**	7.73±3.91	9.51±4.73	**<0.01**	9.68±4.77	**<0.01**	8.06±4.24	0.91
NP	**Mean±SD**	0.88±0.90	0.98±1.03	0.61	0.99±1.03	0.55	0.88±1.09	0.85
NBUD	**Mean±SD**	3.99±2.44	5.06±3.36	**<0.01**	5.20±3.41	**<0.01**	3.87±2.69	0.65

Significant results in bold

[a]Mean frequencies for 1000 binucleated cells

[b]Wilcoxon Rank sum test

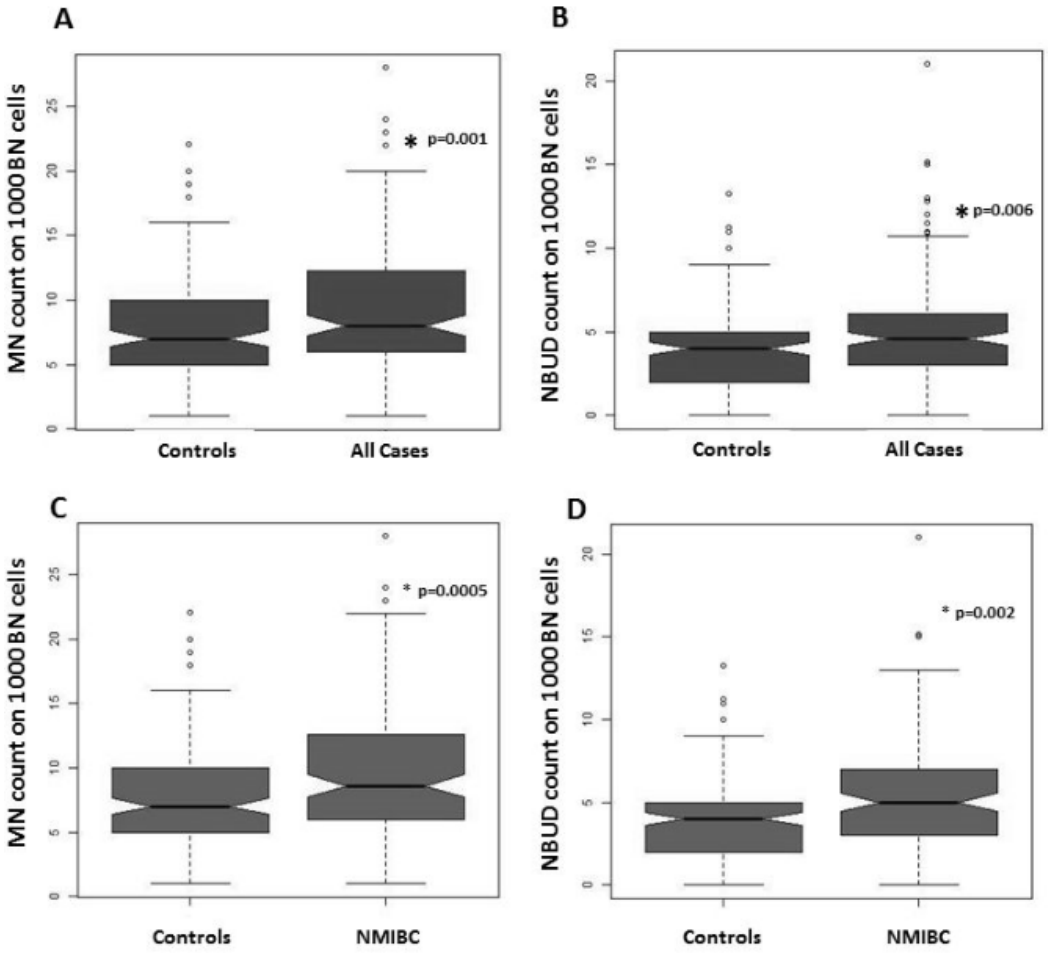


Figure 3 | **Box plot for comparison of MN and NBUD frequencies in different study groups.** MN frequencies in (A) BC cases vs controls and (C) in NMIBC cases vs controls; NBUD frequencies in (B) BC cases vs controls and (D) in NMIBC cases vs controls

Table 10 | **Logistic regression model of analyzed endpoints by the CBMN assay**. A) BC vs. controls; B) NMIBC vs. controls; C) MIBC vs. controls

		Coefficients	SE	P-value[a]
MN	BC vs. controls	0.104	0.029	**<0.01**
	NMIBC vs. controls	0.114	0.030	**<0.01**
	MIBC vs. controls	0.011	0.069	0.87
NPB	BC vs. controls	0.115	0.121	0.34
	NMIBC vs. controls	0.130	0.125	0.30
	MIBC vs. controls	-0.016	0.292	0.96
NBUD	BC vs. controls	0.132	0.043	**<0.01**
	NMIBC vs. controls	0.148	0.045	**<0.01**
	MIBC vs. controls	-0.037	0.112	0.74

Significant results in bold

[a]Adjusted for age and smoking

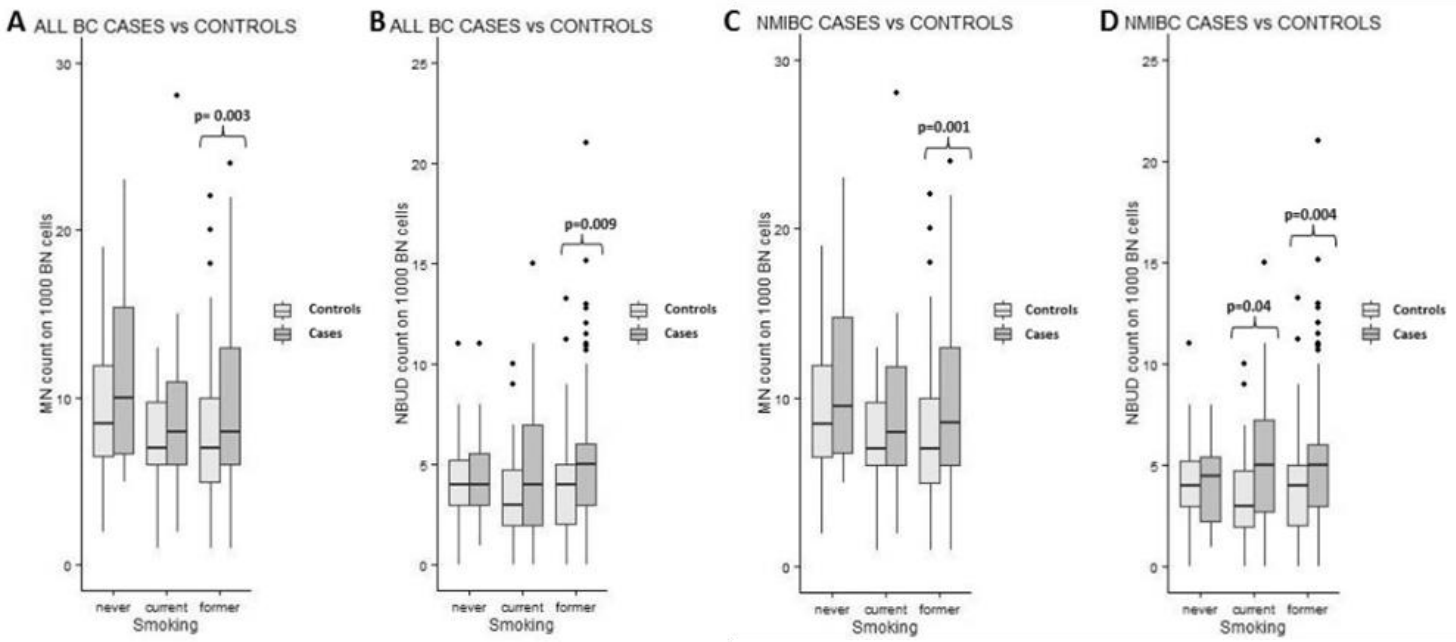


Figure 4 | **Comparison of MN and NBUD frequencies in cases and controls of different smoking status classes**

Genotype distribution in relation to case-control status and MN frequencies

A total of 783 SNPs in genes belonging to different pathways of DNA repair that remained after the quality control were analyzed only in the largest group of NMIBC patients (n=132) and in controls (n=147). Genotypes frequencies of eight SNPs were differentially distributed between NMIBC and controls ($p<0.01$; Chi-Square test) in a linear model adjusted for age and smoking habits (**Table 11**). In particular, an increased risk for NMIBC cases was observed in carriers of the AG or GG genotype for rs804267 in *NEIL2* (p=0.02 and p=0.001, respectively) when compared with the AA genotype. A similar trend was observed for carriers of the AC or CC genotypes for *NEIL3* rs6817959 (p=0.0006 and p=0.01, respectively) when compared with the AA genotype. Conversely, having one or two variant alleles for rs7311151 in *RAD52* appeared to be protective when collated to the reference genotype AA (p=0.004 for the heterozygous and p=0.03 for the homozygous variant GG). The same observations were derived for carriers of TC and CC genotypes for *FANCL* rs11903456 (p=0.03 for both genotypes when compared to the wild type).

Ten SNPs were associated with significantly different MN frequencies according to the genotype stratification in NMIBC patients ($p<0.01$; Kruskal-Wallis rank sum test; **Table 12**). Five of these SNPs were in genes involved in homologous recombination (HRe) (**Figure 5**). No differences in MN frequencies according to a particular genotype were observed among controls for these 10 SNPs (see **Table 12**). Interestingly, for *RAD51*B rs2257127, the MN frequencies in patients decreased according to the presence of 0, 1 or 2 variant alleles, while in controls this trend was the opposite, though significant only when comparing the genotype with 2 variant alleles versus the wild type. Similar considerations could be done for rs1866074 in *TDG*, but in this case MN frequency increased in relation to the number of variant alleles. Notably, the observed differences were confirmed when applying an additive model of inheritance of the minor allele (data not shown). In a linear model controlling for age and smoking status, NMIBC patients carrying the TC or the CC genotypes of *XPC* rs2470458 showed a significant decrease of 2.29 and 4.78 units of MN respectively when compared with the most frequent TT genotype (p=0.01 and p=0.03). An opposite trend was found for rs976080 in *ERCC8* (increase of 3.31 MN units for CA carriers when compared with the reference; p=0.002). For rs1866074 in *TDG*, there was an increase in the frequencies of MN in carriers of AG and GG genotypes

when compared with carriers of the reference genotype AA (increase of 2.43 and 4.05 units of MN; p=0.02 and p=0.001, respectively) (**Figure 6**).

In order to evaluate the improvement in the performance of NMIBC case prediction when including genotype of DNA repair genes and MN frequencies, we analyzed three nested models (Model A = including age and smoking status; model B = model A + rs804267+ rs6817959 + rs11903456 + rs7311151; Model C = Model B + MN frequencies). Comparing the AUC of the 3 models there was a statistically significant improvement in discrimination of NMIBC: in particular the AUC for model A was 0.55, the AUC for model B was 0.75 and 0.79 for model C. (Model A versus model B: DeLong's p = 2.801x10e-06; model B versus model C: DeLong's p = 0.02) (**Figure 7**)**.**

Table 11 | **SNPs in DNA repair genes significantly associated with non-muscle invasive BC (NMIBC) risk**

				NMIBC (N=132)			**Controls (N=141)**			
Gene	**ID rs**	**DNA repair Pathway**[1]	**Alleles (A/a)**	**AA**	**Aa**	**aa**	**AA**	**Aa**	**aa**	**P**[2]
RAD52	**rs7307680**	**HRe**	**[A/G]**	**103**	**24**	**5**	**83**	**53**	**5**	**0.002**
	rs7311151	**HRe**	**[A/G]**	**59**	**55**	**18**	**37**	**78**	**26**	**0.006**
NEIL3	**rs6817959**	**BER**	**[C/A]**	**27**	**81**	**24**	**57**	**62**	**20**	**0.001**
NEIL2	**rs804267**	**BER**	**[A/G]**	**39**	**67**	**26**	**68**	**61**	**12**	**0.002**
XRCC1	rs2854496	BER	[T/C]	79	49	4	107	27	7	0.004
	rs25487[3]	BER	[C/T]	68	56	8	61	53	27	0.005
FANCL	**rs11903456**	**Fanconi Anemia**	**[C/T]**	**90**	**41**	**1**	**74**	**59**	**8**	**0.007**
EXO1	rs4149852	DNA repair-related	[C/T]	126	5	1	121	20	0	0.007

In bold SNPs significant also in the linear model adjusted for age and smoking habits

[1] HRe=Homologous recombination; BER= Base excision repair

[2] Chi-Square test

[3] exm1475983 in the Illumina HumanOmniExpressExome-8 v1.2 Beadchip

Table 12 | **SNPs in DNA repair genes significantly associated with different MN frequencies according to the genotype stratification**

					NMIBC (N=132)						Controls (N=147)					
					Genotype distribution and MN frequency						Genotype distribution and MN frequency					
Gene	**rs ID**	**DNA repair pathway**	**Alleles (A/a)**	**P[2]**	**AA**	**MN‰±SD**	**Aa**	**MN‰±SD**	**aa**	**MN‰±SD**	**AA**	**MN‰±SD**	**Aa**	**MN‰±SD**	**aa**	**MN‰±SD**
PARP1	rs3754376[3]	HRe-related	[A/C]	**<0.001**	58	8.66±4.00	57	11.61±5.29	17	7.41±3.55	72	7.97±3.96	54	7.72±4.06	15	6.48±3.03
	rs1136410[4]	HRe-related	[A/G]	**0.003**	92	9.05±4.05	35	12.24±5.87	5	5.8±2.86	100	8.23±4.00	37	6.55±3.54	4	5.46±1.98
XRCC3	rs861539[5]	HRe	[A/G]	**0.004**	50	9.79±5.31	66	10.56±4.37	16	6.47±3.66	44	7.53±3.58	80	8.11±4.25	17	6.33±2.75
XPC	rs2470458[6]	NER	[T/C]	**0.004**	84	10.74±5.14	43	8.35±3.76	5	5.75±1.61	81	8.09±4.3	50	7.02±3.15	10	8.14±4.01
ERCC8	rs976080	NER	[C/A]	**0.002**	106	9.09±4.42	26	12.56±5.44	0	-	105	7.53±3.62	35	8.11±4.68	1	13±NA
TDG	rs1866074	BER	[A/G]	**0.003**	29	7.46±3.96	71	9.92±4.24	32	11.52±5.93	42	8.35±4.39	66	7.46±3.68	33	7.4±3.75
SHFM1	rs12704871	HRe	[T/C]	**0.005**	53	11.2±4.72	53	8.31±4.01	26	9.84±5.73	54	7.81±3.48	64	7.53±4.27	23	8.01±3.99
ENDOV	rs4491586	DNA repair-related	[T/C]	**0.007**	82	9.56±5.07	43	10.84±4.21	7	5.71±2.75	82	7.29±3.73	51	8.36±4.21	8	8±3.66
RAD51B	rs2257127	HRe	[A/G]	**0.008**	82	10.35±5.11	44	9.37±4.11	6	4.83±1.94	86	7.39±3.82	48	7.84±3.71	7	10.87±5.45
APLF	rs7419739	BER-related	[T/C]	**0.009**	28	7.87±4.4	79	10.69±4.93	25	9.02±4.27	39	7.42±3.18	61	8.21±3.86	41	7.25±4.58

Significant results in bold.
[2] Kruskal-Wallis adjusted for age and smoking
[3] in Linkage disequilibrium (LD) with rs3219090 and rs2666428
[4] in LD with rs3219104, rs907190 and rs1805411
[5] in LD with rs861536 and rs861531
[6] in LD with rs2228000

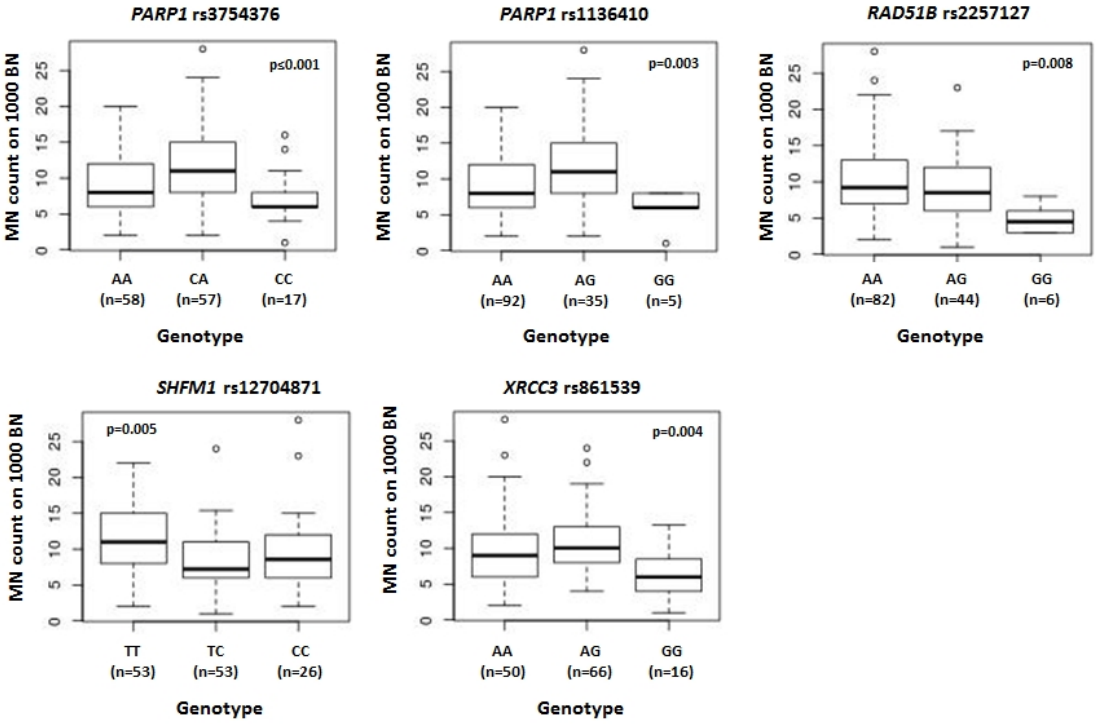


Figure 5 | **MN frequencies and genotype distribution**

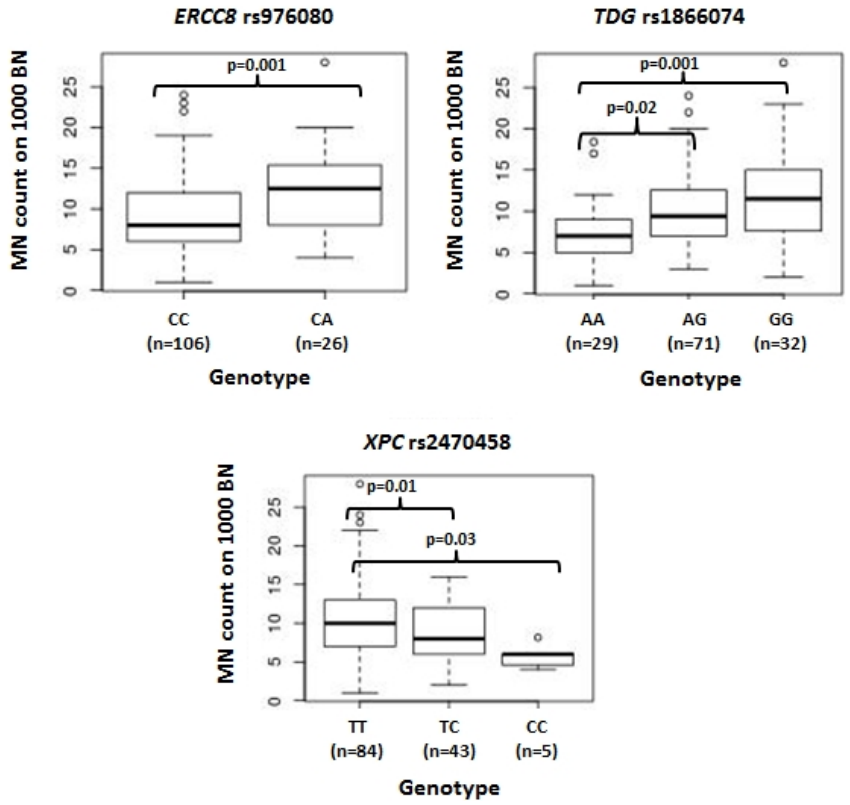


Figure 6 | **MN frequencies in NMIBC patients (n=132) according to genotype distribution of selected SNPs in DNA repair genes (Kruskal–Wallis P<0.01)**

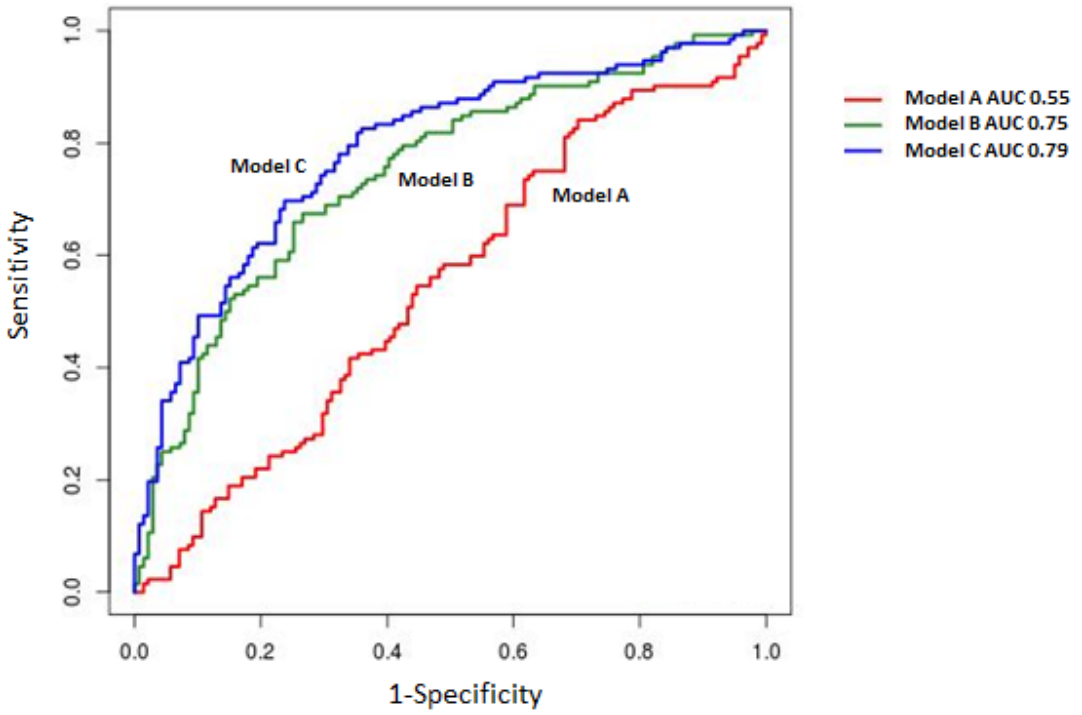


Figure 7 | **ROC curve analysis for discrimination of NMIBC.** Model A investigated the risk reached by adding age and smoking habit; model B included also the SNPs differentially distributed among cases and controls (model A+rs804267+ rs6817959+rs11903456+rs7311151); model C included model B and MN frequencies

4.2.3. Discussion

In the present study, the main result was an increased frequency of MN and NBUD in cryopreserved lymphocytes of BC cases assessed at diagnosis compared to controls, either considering all cases or, with an even stronger difference, the subgroup of NMIBC.

Several studies report an increased baseline frequency of MN and other CBMN endpoints in PBMC of cancer patients (reviewed by [101]), confirming the presence of a high genetic instability in cancer patients. The majority of studies reported the presence of increased MN frequencies in individuals with cancer or with pre-neoplastic lesions. The number of studies performed on PBMC is quite high (8 studies on breast, 6 on lung, 2 on gastro enteric cancers and 13 on other malignancies). However, all these studies were mostly based on a small sample size. The only exceptions are represented by the study of Duffaud et al. [102], performed on PBMC from 197 controls and 57 head-and-neck cancer patients, and the one of El-Zein et al.

[103], on lymphocytes from 139 lung cancer cases and 130 controls. Thus, our study is the first to consider BC patients, with an acceptable sample size.

It is generally accepted that events of genetic damage such as MN and NBUD may represent a reflection of misrepaired DNA breaks, dysregulation on telomere length as well as malfunctions in the mitotic machinery and DNA amplification [104]. An increase in MN frequency may be considered a biomarker of chromosome loss and/or breakage, while other anomalies such as NBUD are biomarkers of gene amplification and/or removal of unresolved DNA repair complexes [101]. In cancer, an increase of various manifestations of chromosomal instability is often observed. Assuming that the induction of chromosomal damage employs similar mechanisms in different tissues, the levels of damage measured in lymphocytes and other surrogate tissues may reflect those present in the cancer-prone tissues [101, 105].

An increase in MN frequency in PBMC of BC cases when compared with controls could be interpreted as an altered status in the DNA damage repair system or reflects an unknown past exposure. These lead to the accumulation of chromosomal damage promoting malignant transformation. However, there is no common agreement on that and an altered DNA repair system could be also a direct effect in the organism of the presence of the tumor [106]. A better understanding of the level of systemic DNA damage in cancer patients may help in a more precise use of therapeutic intervention. For example, a better definition for the minimum threshold of DNA/chromosomal damage that kills tumor cells directly or reactivates mechanisms of tumor-suppression can help in choosing a more effective chemo- or radiotherapy dosage [107].

Several factors may modulate the MN formation and frequency. Lifestyle habits (smoking, alcohol, vitamin intake) and host factors (age, gender) are among them, as well as genetic polymorphisms and exposure to specific mutagen agents [108]. Notably, we observed a significant decrease in MN frequency in current smokers when compared with non-smokers (both considering all subjects and only cases). This is in line with previous investigations on the impact of smoking on MN levels, in which the MN rates are higher in non-smokers than in smokers (reviewed by Nersesyan et al [109]). However, smoker or former smoker BC patients had still higher levels of chromosomal damage than controls and the observed effect of smoking could be due to a constitutively activated antioxidant response system or an active DNA repair.

Knowing a patient's basal genetic background and consequently the individual DNA repair capacity is important for predicting BC risk. Genetic variants in DNA repair genes that could modify BC risk and prognosis has already been identified in various studies, including polymorphisms in *XRCC1*, *XRCC4*, *XPF*, *XPD* and *XPC* [96, 110, 111]. Among the 8 SNPs significantly associated with NMIBC risk in the present study, *NEIL2* rs804267 has been associated in another study with risk of recurrence/progression in 421 NMIBC patients, particularly in those receiving BCG therapy [112]. Conversely, rs25487 in *XRCC1* was associated with decreased risk of BC in the present study and in a previous one performed on 227 Chinese BC patients and 260 controls [113]. The functional significance of *XPC* rs2470458 is still unclear. *XPC* rs2470458 has been indicated as an important and independent predictor for survival of melanoma patients [114]. In a previous study, the presence of the variant allele of rs2228000 (in LD with rs2470458) modulated expression levels of *XPC, XPB* and *XPD* genes in colorectal cancer patients [115]. In particular, rs2228000 is well-known to contribute to the susceptibility to BC in Caucasian populations [116]. Finally, BC patients carrying the GG genotype of rs1866074 resulted at decreased risk of disease progression [112]. Although the effect of a single SNP on BC risk may be limited, the combined effect of several SNPs in the same or different genes could be more significant.

MN formation derives from genetic defects in cell cycle checkpoints and/or DNA repair genes, therefore, genetic variations in such genes may also affect the levels of chromosome damage by ultimately modulating the individual DNA repair activity [117]. Specific genotypes of polymorphisms in DNA repair genes such as *XPD*, *XRCC1* and *XRCC3* were previously associated with increased MN frequencies in workers exposed to ionizing radiation [118]. In the present study, we evaluated MN frequencies in NMIBC cases in relation to a large number of polymorphisms in DNA repair genes, investigated at GWAS level. Among the 10 SNPs associated with different MN frequencies according to the genotype stratification, three of them (rs2470458 in *XPC*, rs1866074 in *TDG* and rs976080 in *ERCC8*) resulted in significant associations also in the linear model. For the polymorphism in *XPC* a decrease in the MN frequency was observed in variant allele carriers when compared with the reference genotypes. For *TDG* rs1866074, there was an increase in the frequencies of MN in carriers of AG and GG genotypes when compared with carriers of the reference

genotype AA. A similar observation was done for carriers of the CA genotype in *ERCC8* rs976080, although in the present study there were no carriers of the AA genotype, it seems that the presence of a single A allele for this SNP is sufficient to have a significant increase in the MN frequency. Interestingly, among the 10 SNPs altering MN frequency in NMIBC there was an overrepresentation of polymorphisms in genes involved in the HRe pathway (5 out of 10; *XRCC3* rs861539, *SHFM1* rs12704871, *RAD51B* rs2257127, *PARP1* rs3754376 and rs1136410), which is supposed to be the main DNA repair pathway involved in the control of chromosome stability [119], and whose subtle interindividual differences may be connected to altered MN frequencies.

In a cross-sectional study like the present one, the CBMN assay does not provide any insight as to whether the observed chromosomal damage is a cause or a consequence of some exposure/presence of the disease. Moreover, human lymphocytes are usually used for measuring genotoxic effects induced by various mutagens in human biomonitoring studies, but the sensitivity of MN endpoints may be limited by the repair of DNA damage during PHA-stimulated lymphocytes culture. On the other hand, previous studies have demonstrated that there are no differences in DNA damage and repair capacity between unstimulated and PHA-stimulated lymphocytes even after 72h of culture [120].

A recent review reported all the studies applying the MN assay in urine-derived cells (UDC) as a diagnostic tool for BC detection and prognosis [109]. To our knowledge, only few studies investigated if the MN assay could be applied as a diagnostic tool to identify individuals who have BC [121-123]. The currently available data (reviewed by [109]) suggested that MN assays in UDC may be potentially a suitable diagnostic tool for detection of BC and for the surveillance of the patients. The specificity and sensitivity of the assay on UDC has to be further validated in large studies. The available studies, in fact, are restricted to a small number of subjects (less than 50 subjects) highlighting that for large studies it could be difficult to process a large number of fresh urine samples. Although urine could be potentially a better specimen for the study of BC risk rather than PBMC, it should be considered that the MN assay applied to UDC presents still some limitations. UDC extraction needs in fact to be performed on fresh urine to ensure viable cells. Moreover, the proportion of live cells extracted from urine is very low so sometimes it is not possible to score enough

cells. In principle, the use of fresh blood provides significant practical advantages such as small amounts of blood required and good proliferation of cells. However, several drawbacks limit the usefulness of fresh blood in human cytogenetic surveys. In large studies, to process hundreds of samples in a short time can be logistically difficult and to arrange the collection of blood specimens over a long-time interval may result in seasonal variation and increased experimental variability [124]. These drawbacks can potentially be overcome using cryopreserved cells. The use of cryopreserved cells may allow a better planning of the study, with *a priori* selection of subjects of interest and the recovery of data in case of lost samples. Cryopreserved, unstimulated PBMC have been used previously in investigations on endogenous DNA damage and/or DNA repair capacity by means of the host reactivation [125] and comet assays [126]. In addition, PBMC can be successfully cryopreserved, stored and employed for MN analysis with cytokinesis-block bringing significant advantages in biomonitoring and for biomarkers research [124].

In conclusion, the observed increase in MN and NBUD frequencies in BC cases could be used to better estimate individual BC risk, including in the same model classical risk factors and DNA repair genetic susceptibility variants. Understanding the level of systemic chromosomal damage in cancer patients may also help in more precise therapeutic intervention.

4.3. Aphidicolin-block NER comet assay

It has been proposed that targeted inhibition of relevant DNA repair factors could sensitize tumors to therapy [127]. Therefore, information on DRC may represent a potential biomarker for cancer treatment [128]. In line with this assumption, data evaluating DRC in PBMCs from cancer patients suggested that DNA repair functional assays (e.g. comet assay) provide much more useful information than measuring expression of the DNA repair genes [129, 130].

The comet assay is a relatively simple, sensitive, rapid and inexpensive method that has already been employed in several DNA damage/repair clinical studies [130, 131]. Modified versions of the comet assay may specifically analyze DNA repair pathways activities or different levels/types of DNA damage. To address the question whether DNA damage and NER capacity could be used as a prognostic factor in BC, we investigated basal DNA damage and DRC in PBMCs from a cohort of BC patients and healthy controls using a modified comet assay able to evaluate NER activity [132, 133].

4.3.1. Material and Methods

Study population

The complete study population has been described in section 3.

Lymphocytes isolation and storing

The complete technique has been described in section 4.1.1.

Aphidicolin-block NER comet assay

Cryopreserved PBMCs were thawed quickly in a 37 °C water bath and suspended in 5 ml of cold medium containing 50% FBS, 49% RPMI 1640 and 1% dextrose. Treatment with aphidicolin (APC) or/and benzo[a]pyrene diol epoxide (BPDE) was performed as previously described [133]. Briefly, cells were spun down by centrifugation and treated with 2.5 μg/ml phytohaemoagglutinin (PHA) (Sigma-Aldrich Co, St Louis, MO). Twenty-four hours later samples were centrifuged, divided into 3 tubes and treated in the following ways: A –2.5 μg/ml APC (Sigma-Aldrich Co), 30 min at 37 °C, 5% CO_2; B – 0.5 μM BPDE (NCI

Chemical Carcinogen Reference Standards Repository, Midwest Research Institute, Kansas City, MO, USA), 2 h at 37 °C, 5% CO_2; C – pretreatment with 2.5 µg/ml APC (30 min) followed by 0.5 µM BPDE (2 h). At the end of the treatment, cells were centrifuged, and pelleted cells were processed for comet assay.

The NER comet assay was performed according to the methods previously described [132], with slight modifications [133]. After treatment, cells were mixed with low-melting-point agarose (0.75%; Sigma-Aldrich Co) and layered on 85 x 100 mm GelBond films (Lonza, Basel, Switzerland). Each GelBond film comprised eight 19 x 23 mm agarose gels. The GelBond films were immersed in lysis solution (2.5 M NaCl, 0.1 mM Na_2EDTA, 10 mM Tris, 1% Triton X-100, pH 10) for 1.5 hours at 4 °C, then placed in an electrophoresis tank for 40 minutes, submerged into electrophoresis buffer (0.3 M NaOH, 1 mM Na_2EDTA, approximately pH 13), and finally ran at 30 V (0.8 V/cm) and 300 mA for 20 min. After neutralization with 0.4 M Tris-HCl (pH 7.5), gels were fixed in ethanol and dried at room temperature. For scoring, slides were stained with YOYO-1 iodide (1 mM solution in DMSO, diluted 1:250 in PBS; Life Technologies Italia, Monza, Italy) and nuclei were visualized by a Leica fluorescence microscope at 40X magnification. Two gels of 50 nuclei for each sample treatment were scored with Comet IV software (Perceptive Instrument, Suffolk, UK). The median tail moment (TM) of 100 nuclei was used as a measure of DNA damage. The TM is defined as the product of the tail length and the fraction of total DNA in the tail (Tail moment=tail length x % of DNA in the tail). TM is calculated automatically by the computer software system as an average for the 50 cells selected for measurement. For each electrophoresis run, a human K562 erythroleukemia cell line was included as reference standard and used to normalize results [133]. Since in our previous study we demonstrated that APC-treated and untreated cells did not differ in DNA damage levels [133], in this study we directly used APC-treated cells endpoint as representative of basal DNA damage. DRC (NER capacity) was calculated for each subject as: TM $DNA_{APC+BPDE}$ – TM DNA_{BPDE} - TM DNA_{APC}. Basal DNA damage was evaluated from cells cultured with APC only [133].

Statistical analysis

Patient characteristics were described as absolute frequencies for qualitative variables, and with mean, standard deviation (SD) and range values for quantitative variables. The analyzed variables were basal DNA damage in APC-treated cells and DRC, obtained as described above.

The influence of possible explicative variables on the basal DNA damage or DRC was evaluated by multivariate analysis and/or non-parametric statistical hypothesis test, since variables did not follow a normal distribution (Shapiro–Wilk normality test). Wilcoxon or Kruskal-Wallis Rank Sum tests were applied to compare the distribution between groups of basal DNA damage or DRC.

OS was evaluated calculating the time (in years) between the date of BC diagnosis and the date of death or follow up termination as the endpoint for each patient. EFS was calculated as the time (in years) between the date of BC diagnosis until date of relapse, death or censorship, whichever came first. The relative risk of death or recurrence against each of the basal DNA damage or DRC endpoints was estimated as HR using Cox regression (R Survival package). Basal DNA damage or DRC were all considered as categorical (above/below the median value). Multivariate survival analyzes were adjusted for age, smoking, T stage and therapy variables for the whole group of patients. The prognostic role of DNA damage or DRC on survival was also evaluated using Kaplan Meier curves and log-rank test.

All the analyzes were performed with the open source R (version 3.2.4)

4.3.2. Results

Study population

The study included PBMCs from 146 cases and 155 controls. For 61 subjects technical problems were encountered in one or more of the three points (i.e., slides were not scorable), therefore basal DNA damage or DRC were not evaluable. Since no additional aliquots of cryopreserved PBMCs were available, these subjects were excluded from the analysis. Finally, 133 BC patients and 141 control subjects were included in the basal DNA damage analysis, while 121 BC patients and 119 controls subjects were included in the DRC analysis. The characteristics of the whole cohort are summarized in **Table 13**.

Table 13 | **NER Comet assay: characteristics of the study population**

Covariates	**Controls (%)**	**Cases (%)**
N	155	146
Non-muscle invasive BC	-	131 (89.7)
Muscle invasive BC	-	15 (10.3)
Age (years)		
Mean ± SD	62.46 ± 8.63	62 ± 7.86
Range	40.31 – 74.59	39.99 – 74.10
Smoking Status		
Never	25 (16.1)	23 (15.8)
Former	94 (60.7)	89 (60.9)
Current	36 (23.2)	34 (23.3)
T stage*		
Ta		78 (53.4)
T1		47 (32.2)
≥T2		15 (10.2)
Tis		3 (2.1)
Tx		3 (2.1)
Grading (WHO 1973)*		
G1		38 (26)
G2		60 (41.1)
G3		48 (32.9)
Grading (WHO 2004/2016)*		
Non-high grade		72 (49.3)
High grade		74 (50.7)
Risk*		
Low-risk		35 (24)
Intermediate Risk		41 (28.1)
High-risk		55 (37.7)
Muscle invasive		15 (10.2)
Recurrences		
No		94 (64.4)
Yes		52 (35.6)
Number of recurrences		
1		31 (59.6)
2		16 (30.8)
≥3		5 (9.6)
Progression		
No		140 (95.9)
Yes		6 (4.1)
Therapy		
No		43 (29.5)
Yes		103 (70.5)
Cystectomy		
No		119 (81.5)
Yes		27 (18.5)
BC patient' status at follow up		
Alive		122 (83.6)
Died		24 (16.4)

*referred to [5, 100]

Comet assay endpoints in relation to clinical outcomes

DRC was evaluated by a modified comet assay that measures the capacity of PBMCs to resolve DNA damage after *in vitro* challenging with BPDE in the presence or absence of APC, a potent and specific inhibitor of DNA polymerases α and σ [134]. No differences between cases and controls were observed by Wilcoxon Rank Sum test, neither considering only NMIBC or MIBC cases (**Table 14**). Same results were obtained with a generalized linear model, controlling for age and smoking habits.

Basal DNA damage levels and DRC correlation with grading, risk and recurrences were evaluated, either in all cases or NMIBC only (**Table 15**). While DRC did not show any significant correlation with clinical characteristics of tumors, basal DNA damage at diagnosis was significantly different between tumor grading (all BC: WHO 1973 G3 vs G1 adjusted linear regression p=0.02; WHO 2004/2016 adjusted linear regression p=0.04, **Figure 8A** and **B**, respectively; NMIBC: WHO 1973 G3 vs G1 adjusted linear regression p=0.01; WHO 2004/2016 adjusted linear regression p=0.03, **Figure 9A** and **B**, respectively). We also observed a significant positive trend of correlation between basal DNA damage and tumor grading (all BC: p=0.02; NMIBC: p=0.01), suggesting that DNA damage increases with invasive progression. Moreover, basal DNA damage was significantly different also between risk classes (all BC: High vs Low risk adjusted linear regression p=0.03, **Figure 8C**; NMIBC: High vs Low risk adjusted linear regression p=0.02, **Figure 9C**), with a significant positive trend in NMIBC with the increase of risk (p=0.02).

Table 14 | **Basal DNA damage and nucleotide excision DNA repair capacity in PBMCs from BC cases and healthy controls**

		Background DNA damage					
	Controls	**BC**	**p-value***	**NMIBC**	**p-value***	**MIBC**	**p-value***
N	141	133		119		14	
Mean[a] (SD)	0.21 (0.19)	0.19 (0.18)	0.64	0.19 (0.18)	0.53	0.22 (0.19)	0.66
		DNA Repair Capacity					
	Controls	**BC**	**p-value***	**NMIBC**	**p-value***	**MIBC**	**p-value***
N	119	121		108		13	
Mean[b] (SD)	12.28 (9.56)	14.15 (10.89)	0.21	14.01 (11.05)	0.29	15.34 (9.80)	0.25

[a] *Mean of comet tail moment*

[b] *Mean DRC calculated as described in Materials and methods section*

** Wilcoxon Rank Sum test*

Table 15 | **Basal DNA damage and nucleotide excision DNA repair capacity in PBMCs from all BC cases or NMIBC stratified for grade, risk and recurrence rate**

	ALL CASES	**DNA damage (mean±sd)**	**p***	**adj p****	**trend p*****	**ALL CASES**	**DRC (mean±sd)**	**p***	**adj p****	**trend p*****
Tumor Type	NMI (n=119) MI (n=14)	0.19 ± 0.18 0.22 ± 0.19	0.38	ref 0.62	nd	NMI (n=108) MI (n=13)	14 ± 11.1 15.3 ± 9.8	0.72	ref 0.99	nd
Grading (WHO 1973)	G1 (n=36) G2 (n=51) G3 (n=46)	0.13 ±0.11 0.2± 0.17 0.23 ±0.22	0.1	ref 0.08 **0.02**	**0.02**	G1 (n=33) G2 (n=45) G3 (n=43)	13.1 ±10.2 15.7± 12.1 13.3 ±10.1	0.67	ref 0.81 0.36	0.36
Grading (WHO 2004/2016)	Non high grade (n=65) High grade (n=68)	0.16 ±0.14 0.22± 0.20	0.11	ref **0.04**	nd	Non high grade (n=58) High grade (n=63)	14 ±11.1 14.3 ±10.8	0.89	ref 0.58	nd
Risk	Low (n=33) Intermediate (n=36) High (n=50) MI (n=14)	0.13±0.11 0.19±0.17 0.22±0.21 0.22±0.19	0.39	ref 0.25 **0.03** 0.16	0.11	Low (n=30) Intermediate (n=31) High (n=47) MI (n=13)	14.4±10.2 14.1±11.9 13.7±11.2 15.3±9.8	0.89	ref 0.6 0.26 0.62	0.55
Recurrences	No (n=88) Yes (n=45)	0.21 ±0.19 0.15 ±0.15	**0.03**	ref **0.03**[#]	nd	No (n=80) Yes (n=41)	14.5 ±10.9 13.5 ±10.9	0.67	ref 0.86[#]	nd
	NMIBC	**DNA damage (mean±sd)**	**p***	**adj p****	**trend p*****	**NMIBC**	**DRC (mean±sd)**	**p***	**adj p****	**trend p*****
Grading (WHO 1973)	G1 (n=36) G2 (n=50) G3 (n=33)	0.13± 0.11 0.20 ±0.17 0.23± 0.23	0.14	ref 0.09 **0.01**	**0.01**	G1 (n=33) G2 (n=44) G3 (n=31)	13.1± 10.2 15.5± 12.1 12.9± 10.5	0.69	ref 0.95 0.35	0.35
Grading (WHO 2004/2016)	Non high grade (n=65) High grade (n=54)	0.16± 0.14 0.22± 0.21	0.17	ref **0.03**	nd	Non high grade (n=58) High grade (n=50)	14± 11.1 14.1± 11.1	0.96	ref 0.56	nd
Risk	Low (n=33) Intermediate (n=36) High (n=50)	0.13 ±0.11 0.19 ±0.17 0.22 ±0.21	0.34	ref 0.24 **0.02**	**0.02**	Low (n=30) Intermediate (n=31) High (n=47)	14.4 ±10.2 14.1 ±11.9 13.7 ±11.2	0.81	ref 0.57 0.25	0.25
Recurrences	No (n=74) Yes (n=45)	0.21± 0.19 0.15 ±0.15	**0.04**	ref **0.01**[#]	nd	No (n=67) Yes (n=41)	14.3± 11.2 13.5 ±10.9	0.7	ref 0.90[#]	

* Wilcoxon Rank Sum or Kruskal-Wallis Rank test

** Generalized linear model adjusted for age and smoking habits (ref = reference group)

*** Trend test adjusted for age and smoking habits

[#] Generalized linear model adjusted for age, smoking habits and therapy

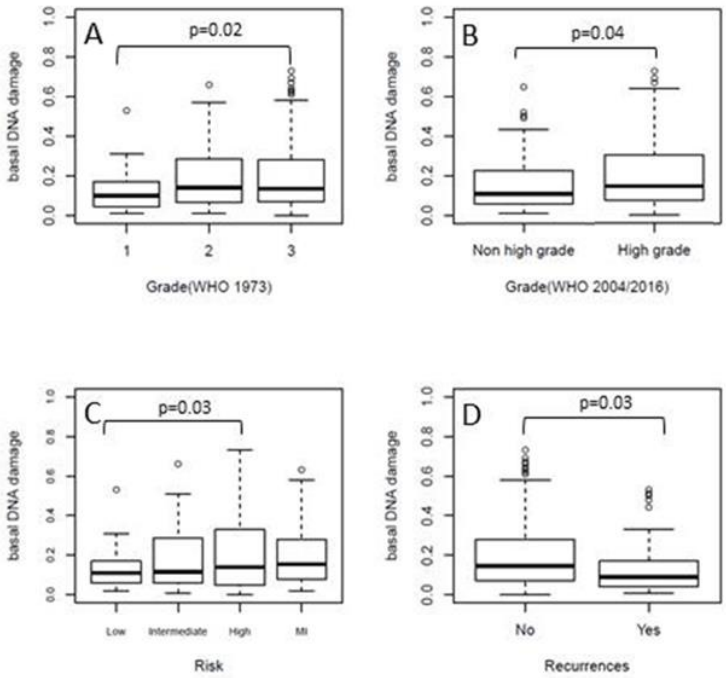


Figure 8 | **Basal DNA damage in all BC cases at diagnosis according to (A) tumor grading (WHO, 1973); (B) tumor grading (WHO 2004/2016); (C) risk classes; (D) presence of recurrences.** *P* values if significant comparisons are given in the figure (model adjusted for age and smoking status; also for therapy in (D)

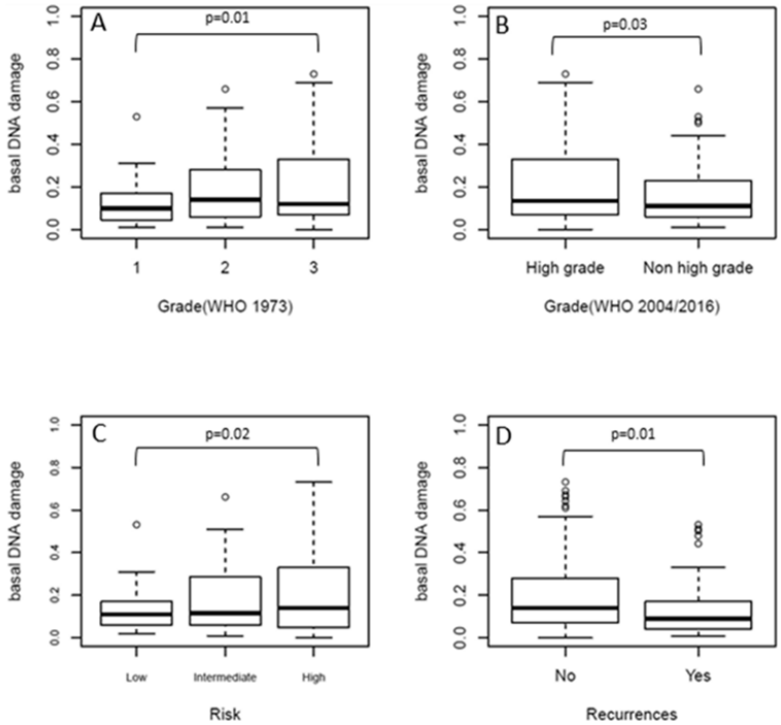


Figure 9 | **Basal DNA damage in NMIBC cases at diagnosis according to (A) tumor grading (WHO, 1973); (B) tumor grading (WHO 2004/2016); (C) risk classes; (D) presence of recurrences.** *P* values if significant comparisons are given in the figure (model adjusted for age and smoking status; also for therapy in (D)

Overall survival, recurrence rate and event-free survival analyzes

Results from the survival analysis are reported in **Table 16** and shown in **Figure 10**. Patients were stratified into two categories, above and under the median value of DNA damage or DRC, to calculate Kaplan–Meier OS curves.

Values above the median value of basal DNA damage corresponded to the "High DNA damage" group, those under the median value corresponded to the "Low DNA damage" group. There was a significantly decreased survival in patients with an increased basal DNA damage level both in all BC patients and also when stratified only for NMIBC (HR=3.7, 95%CI 1.3–10.6, p=0.02 and HR=4.4, 95%CI 1.1–17.3, p=0.03, respectively). There was no association between DRC and OS, neither between EFS and both comet assay endpoints (**Table 16**).

Patients that developed recurrences at follow-up showed a significantly lower basal DNA damage compared to patients that did not (all BC, adjusted linear regression p=0.03, **Figure 8D** and **Table 15**; NMIBC, adjusted linear regression p=0.01, **Figure 9D**). This suggests that patients having a higher basal DNA damage probably presented an induced increased DNA repair response resulting in a lower risk of recurrences.

Table 16 | **Basal DNA damage and nucleotide excision DNA repair capacity affecting overall survival (OS) and event-free survival (EFS) in all BC and NMIBC**

All BC	**Overall Survival**			**Event-Free Survival**		
	HR	**95% CI**	**p-value***	**HR**	**95% CI**	**p-value***
DNA damage[#]	**3.7**	**1.3-10.6**	**0.02**	1	0.6-1.7	0.9
DRC [#]	0.8	0.3-2.1	0.6	0.9	0.5-1.6	0.78

NMIBC	**Overall Survival**			**Event-Free Survival**		
	HR	**95% CI**	**p-value****	**HR**	**95% CI**	**p-value****
DNA damage[#]	**4.4**	**1.1-17.3**	**0.03**	0.9	0.5-1.5	0.58
DRC [#]	0.4	0.1-1.5	0.18	0.9	0.5-1.6	0.65

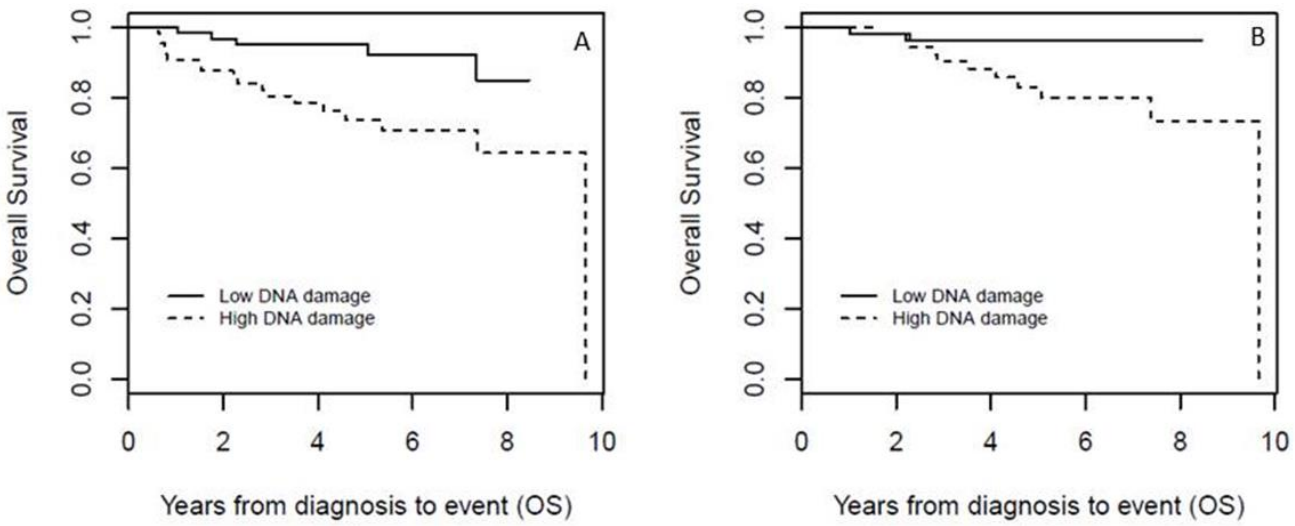


Figure 10 | **Kaplan-Meier OS curves according to basal DNA damage of all BC (A, *n*=133, *n* events=23) and NMIBC (B, *n*=119, *n* events=13).** Patients were stratified into 2 categories: those with DNA damage above the median value corresponding to "High DNA damage" (median OS time for A=4.11 years, *n* events=18; B=4.57 years, *n* events=10); those with DNA damage under the median value corresponding to "Low DNA damage" (median OS time for A=4.91 years, *n* events=5; B=4.81 years, *n* events=3)

4.3.3. Discussion

In the present study, we observed a significant increased basal DNA damage in PBMCs from BC patients when stratified for grade, risk or recurrence rate. Moreover, a lower DNA damage in patients at diagnosis was associated with longer survival time.

Morphologic and pathologic criteria (e.g. histology, stage, grade) used for conventional diagnosis of BC [135] have inadequate power to predict patient outcome precisely, and there remains significant variability in the prognosis of patients with similar characteristics [136]. Thus, new biomarkers that predict clinical behavior in patients with BC is needed especially regarding non-invasive tissues, like blood or urine [137].

Elevated DNA damage levels and suboptimal NER, the major DNA repair mechanism for repairing bulky DNA damage generated by most environmental factors [138], may be associated with BC, as reported in many studies [55, 139, 140]. We hypothesized that a quantification of the DNA damage and DRC might serve as BC risk and prognostic biomarkers. The modified comet assay used in this study allows in fact to measure the accumulation of DNA breaks, as incision events, by blocking repair synthesis, and is adequate to detect DRC inter-individual differences in the context of NER pathway [141].

Several studies identified SNPs in DNA repair genes associated with BC risk [142, 143] or response to therapy [63, 127, 144-146]. However, few of them could estimate the DNA damage and DRC in blood cells of BC patients. In a previous study, baseline levels of DNA damage measured by a similar assay were significantly higher in BC patients than in controls [147]. The main difference with our study consisted in the type of cells: while Schabath and colleagues quantified DNA damage on fresh whole blood , we measured DNA damage on cryopreserved PBMCs. Moreover, we compared BC cases with hospitalized controls: this could be an issue when measuring disease markers, as healthy controls from the general population should be a better comparison group.

In the present study, we found that the basal DNA damage was significantly associated with a worse patient prognosis. The DNA damage increase in PBMCs of high grade (WHO 2004/2016) or G3 (WHO 1973) BC, could be interpreted as an altered status in the DNA damage repair system or a consequence of an unknown past exposure. Tumor grade is an important predictor of cancer prognosis [148]. However, histopathological classifications are known for their inter- and intra-observer variability which may have profound limitations in prognosis [149]. Hence, basal DNA damage levels in PBMCs may serve as potential predictive biomarkers for better stratification of BC patients.

The comet assay used to measure DRC is specific for NER, but it detects a wide range of type of DNA damages at basal levels. Alkaline comet assay is capable of detecting DNA double-strand breaks, single-strand breaks, alkali-labile sites, DNA-DNA/DNA-protein cross-linking, and incomplete excision repair sites, with no specific assignment of the DNA repair pathways responsible for that. We hypothesize that increased DNA damage in patients with worse outcome are associated with alterations in different DNA repair pathways that subsequently may lead to carcinogenesis. However, the exact mechanism explaining this phenotype remains to be elucidated.

We are aware of some limitations of the present study. The potential use of the comet assay for DNA damage and repair activity associated with cancer was reviewed by McKenna et al. [150]. Recent studies showed that high levels of DNA damage in PBMCs were associated with different types of cancer, including breast [151, 152], carcinoma of the cervix [153], Hodgkin's disease [154], and esophageal cancer [155]. On

the other hand, these outcomes were not observed in other type of cancer such as lung [156] and prostate [157]. Noteworthy, the observed associations between DNA damage and disease established in case–control studies do not allow to conclude whether the elevated DNA damage is a cause or an effect of the disease. To establish causality, prospective studies need to be conducted. Finally, in our study we recruited only men affected by BC: we can only hypothesize that DNA damage could be a good prognostic biomarker also in BC women. Møller et al. [158] reported that gender is one of the factors that influence the level of DNA damage detected by the comet assay in biomonitoring occupational studies, confirming that further analyzes should be done in female patients affected by BC.

We have shown an association between DNA damage levels and BC patients' outcomes. The level of DNA damage thereby could provide valuable information to stratify patients for cancer treatments to decrease progression rate and improve patients' outcomes.

Our data suggest that basal DNA damage levels in BC patients may potentially represent an important prognostic marker associated with poor OS and after further validation could be used for better stratification of BC patients for therapy. Hence, reliable methods for detecting DNA damage levels in PBMCs in cancer patients may amplify the diagnostic and prognostic tool set.

5. Urinary BC biomarkers

In this section I report results recently submitted for publication.

> Allione A*, Pardini B*, Viberti C*, Giribaldi G, Turini S, Di Gaetano C, Guarrera S, Cordero F, Oderda M, Allasia M, Gontero P, Sacerdote C, Vineis P, Matullo G, *MMP23B expression and protein levels in blood and urine are associated with bladder cancer risk*. J Urol 2018, Under review.
> ***equally contributed**

5.1. MMP23B expression and protein levels in blood and urine

Several BC tissue gene profiles have been reported predicting a range of tumor characteristics and outcomes [159, 160]. However, to improve diagnostic accuracy and to overcome the disadvantages of current diagnostic strategies, the detection of biomarkers associated with BC in body fluids could results in a more promising approach than analyzes of biopsies [67].

Searching for cancer biomarkers in blood or urine offers several advantages over target tissue samples, such as minimal invasiveness, potential for repeated measurements and there relatively high tissue homogeneity. Various blood biomarkers have been investigated for cancer, and among them matrix metalloproteinase (MMP) family has been identified as a potential body fluids biomarkers in BC [159]. MMP family is composed by a group of enzymes that have the capacity to degrade the components of extracellular matrix [67, 161]. Although not generally used as a cancer diagnostic marker, MMPs expression has been widely studied as a potential prognostic marker for a number of malignancies [162].

MMP23 is a unique member of the MMP family: it is still unknown the physiological function of MMP23B, although there is evidence on its possible role in various cellular processes (Cathcart, J., Pulkoski-Gross, A. & Cao, J. Targeting Matrix Metalloproteinases in Cancer: Bringing New Life to Old Ideas. Genes Dis 2, 26-34, doi:10.1016/j.gendis.2014.12.002 (2015)).

In this study, we performed genome-wide gene expression analyzes and subsequent validation by qPCR in PBMCs of 66 BC patients and 70 controls. We observed a downregulation of *MMP23B* gene and its pseudogene *MMP23A* in BC when compared to controls. This result encouraged us to evaluate also

MMP23B protein levels, both in plasma and in urine samples, and to investigate the presence of microRNAs (miRNAs) targeting the *MMP23B* gene in urine of BC patients and controls.

5.1.1. Material and Methods

Study population

The complete study population has been described in section 3.

Sample collection

The isolation and storing of PBMCs has been described in section 4.1.1.

For all subjects, human plasma samples were obtained from 5-8 ml of EDTA blood centrifuged for 10 min at 1000 rpm. Plasma aliquots (about 200-300 µl each) were then stored at -80°C until use. Urine samples from each participant were collected in the morning, stored at 4° C until the processing consisting of centrifugation at 3,000g for 10 min. The urine supernatant aliquots were then transferred in tubes and stored at -80°C until use.

RNA isolation (gene expression and small RNA-Seq)

Total RNA was extracted from cryopreserved PBMCs from BC cases (collected before any treatment) and controls selected from the TBCS and collected as reported in [163, 164]. Total RNA was isolated using RNeasy extraction kit (Qiagen) following the manufacturer's protocol. The integrity of purified RNA was assessed using the RNA 6000 Nano Kit on an Agilent 2100 Bioanalyzer (Agilent Technologies) and RNA concentrations were determined using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies). Samples from BC cases and controls were selected according to the following QC requirements: RNA Integrity Number (RIN) ≥7, at least 300 ng RNA available.

Total RNA from urine was extracted with Urine microRNA Purification kit (Norgen biotek corp), following the manufacturer's standard protocol. RNA quality and quantity was verified according to the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines

(*http://miqe.gene-quantification.info/*). For all urine samples, RNA concentration was quantified by Qubit® 2.0 Fluorometer with Qubit® microRNA Assay Kit (Invitrogen).

Gene expression microarray

RNA from PBMCs was amplified, biotin-labeled, and hybridized on the Illumina Human BeadChip HT12 V4 whole genome expression array (66 BC cases and 70 controls, Discovery Set, **Table 17**). The Human HT-12 v4 Expression Beadchip is a genome-wide gene expression array that targets >31,000 annotated transcripts with >47,000 probes derived from the National Center for Biotechnology Information Reference Sequence (NCBI). All arrays were performed by the same operators in the same core facility. Synthesis of double-stranded cDNA, preparation and labeling of cRNA, hybridization to HumanHT-12 v4 BeadChip (Illumina, San Diego, CA), washing, and scanning were performed according to Illumina standard protocols. Slides were scanned using Illumina HiScan and the signal was analyzed and extracted using GenomeStudio software (Illumina, San Diego, CA).

Technical validation of candidate blood-based biomarkers

From the same samples set analyzed with the Illumina Human BeadChip, a subgroup (57 controls and 44 BC cases, Validation Set, **Table 17**) was included in the technical validation of the most differentially expressed genes by qPCR. One microgram of total RNA was used to synthesize complementary cDNA, using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems), according to the manufacturer's protocol. The TaqMan Gene Expression Master Mix and prevalidated TaqMan Gene Expression Assays (*MMP23A/MMP23B*: Hs00270380_m1; *MMP23B*: Hs04187882_g1) were used for qPCR (Applied Biosystems). TaqMan gene expression assays were designed, tested, and optimized to address all MIQE guidelines [165].

The Abelson murine leukemia viral oncogene homolog 1 (*ABL1*; Hs00245443_m1) and hypoxanthine phosphoribosyltransferase 1 (*HPRT1*;Hs02800695_m1) genes were used as reference genes for qPCR normalization. These reference genes are among the most stably expressed genes in PBMCs, suitable for

normalization in qPCR studies (*http://www.genecards.org/*). Moreover, they were identified by Applied Biosystems as appropriate genes for expression levels normalization in PBMCs. QPCR was performed in a 7900HT Real-Time PCR System (Applied Biosystems). Each sample was run in triplicate.

Replication in an independent set of samples

The same genes subject to the technical validation were further investigated in an independent set of samples (21 BC patients and 24 controls, Replication Set, **Table 17**). The qPCR methodology was applied as described above.

External replication in TCGA bladder tumor and normal samples

Results from gene expression on BC patients were compared with an open access dataset of BC individuals from The Cancer Genome Atlas (TCGA; *https://cancergenome.nih.gov/*) project. Level 2 data from BC publicly available in TCGA were employed as test set and analyzed by the Wanderer tool [166], an interactive viewer to explore DNA methylation and gene expression data in TCGA.

TCGA RNA-seq data were generated with Illumina HiSeq RNAseq, summarized by exons and genes and presented as log2-transformed RPKM values for each exon in two separate panels for a subset of normal and tumor tissue samples [167]. In total, RNA-Seq data from 267 BC tissues and 19 adjacent nonmalignant urothelial tissue samples were included as external validation.

Library preparation for small RNA-seq

Briefly, library preparation was performed with the NEBNext Multiplex Small RNA Library Prep Set for Illumina following the manufacturer`s protocol (Protocol E7330, New England BioLabs Inc., USA). For each library, 6 µL of RNA (min 35 ng of RNA) was used in all the experimental procedures as starting material. Each library was prepared with a unique indexed primer so that the libraries could all be pooled into one sequencing lane. Multiplex adaptor ligations, reverse transcription primer hybridization, reverse transcription reaction and the PCR amplification were processed according to the protocol for library

preparation (Protocol E7330, New England BioLabs Inc., USA). The obtained sequence libraries were subjected to the Illumina sequencing pipeline, passing through clonal cluster generation on a single-read flow cell (Illumina Inc., USA) by bridge amplification on the cBot (TruSeq SR Cluster Kit v3-cBOT-HS, Illumina Inc., USA) and 50 cycles sequencing-by-synthesis on the HiSeq 2000 (Illumina Inc., USA) (in collaboration with EMBL, Heidelberg, Germany).

Western Blot analysis

Twenty microliters of plasma (diluted 1:500) or urine supernatants were boiled at 99 °C for 10 min in reducing Laemmli sample buffer (Bio-Rad, Segrate, Italy), resolved as proteins by SDS-PAGE Mini-PROTEAN TGX Stain-Free Precast Gels 10% (Bio-Rad) and transferred to Mini format 0.2 um nitrocellulose membranes (Trans-Blot Turbo Transfer System, Bio-Rad). Blots were incubated overnight with anti-MMP23B antibody 1:500 (Abcam, ab155416) and subsequently with a mouse anti-rabbit IgG-HRP conjugate antibody 1:2000 (Santa Cruz Biotechnology, sc-2357) for one hour. Blots were developed by using enhanced chemiluminescence (Clarity Western ECL Blotting Substrate, Bio-Rad). Image acquisition was performed using ImageQuant LAS4000 and TL Version 7.0 software (GE Healthcare, Milan, Italy).

Enzyme-linked immunosorbent assays (ELISA)

The level of human MMP23B protein was quantified blinded in plasma and urine samples using ELISA kit (Cat# MBS2516225, MyBiosource, San Diego, California, USA). The assay was conducted according to the manufacturer's instructions. Curve fitting was accomplished by either linear or four-parameter logistic regression following manufacturer's instructions. The relatively constant production of creatinine, a non-enzymatic metabolite of creatine, makes urinary creatinine a useful tool for normalizing the levels of other molecules found in urine [168]. The concentration of urinary creatinine was measured using a commercially available enzymatic assay (Cat# KGE005 R & D Systems Inc., Minneapolis, MN, USA) according to the manufacturer's instruction. The measured creatinine concentrations in urine were added in generalized linear regression model to adjust statistical analyzes on urinary MMP23B differences between BC cases and

controls. Creatinine levels between BC cases and controls were not significantly different (mean BC = 114.02 mg/dl; mean controls = 100.80 mg/dl; adjusted linear regression p=0.09).

Statistical analysis

Raw data obtained from the Illumina Human HT12 arrays were quality-checked using Lumi [169], a Bioconductor package that performed the quality control, variance stabilize and normalize dataset. Expression values were log2 transformed.

Differential expression analysis (all patients or only NMI cases vs. controls) was performed using a Bioconductor package called Limma (lmFit and ebayes functions) [170]. The up- or downregulated genes associated with a Benjamini-Hochberg false discovery rate (FDR) p-value < 0.05 were considered for further analysis.

For qPCR analyzes (technical validation and replication in an independent set of samples), relative expression values were normalized to the reference genes (*HPRT1* and *ABL1*), and fold change values were calculated by using the 2-ΔΔCt method [171]. Student's t or Wilcoxon rank sum test was used (depending on Shapiro–Wilk normality test results) to compare the ΔCq (Cq, also named cycle threshold [Ct], is quantification cycle, as suggested by MIQE guidelines [165]) values between the two biological groups (BC vs. controls), and a p-value was calculated.

Same tests were employed to compare the distribution of *MMP23A/MMP23B* gene expression levels or MMP23B protein measurements in plasma and urine between groups. Multivariate analyzes were adjusted for age and smoking (and creatinine levels only for urine).

Overall survival (OS) was evaluated calculating the time (in years) between the date of BC diagnosis and the date of death or follow up termination as the endpoint for each patient. Event-free survival (EFS) was calculated as the time (in years) between the date of BC diagnosis until date of recurrence, progression or death, whichever came first. The relative risk of death or recurrence against urinary MMP23B was estimated as hazard ratios (HR) using Cox regression (R Survival package).

To identify miRNAs targeting *MMP23B* gene we used miRWalk 2.0 [172] which is a complete atlas of predicted and validated miRNA-target interactions. MiRWalk 2.0 shows miRNA-targets interactions information produced also by other established miRNA prediction programs; between them we chose RNA22 (*https://cm.jefferson.edu/rna22/*), miRanda (*http://www.microrna.org/microrna/home.do*) and TargetScan (*http://www.targetscan.org/vert_71/*) and we selected miRNA targeting *MMP23B* predicted by at least three of the used tools. We conducted differential expression analysis among BC cases and controls with DESeq2 Bioconductor package. Over/under expression was evaluated with Wald test and results were given as log2 fold changes for BC cases over controls. MiRNAs were considered significantly differentially expressed when FDR p-value < 0.05.

All the analyzes were performed with the open source R (version 3.2.4).

5.1.2. Results

Discovery set: gene expression microarray analysis

RNA samples from PBMCs of 66 BC cases and 70 controls were included in the Discovery set and analyzed with the Illumina Human BeadChip HT12 V4 whole genome expression array. Characteristics of the individuals included in this part of the study are reported in **Table 17**.

Overall, 20,096 genes resulted expressed (**Figure 11**). The differentially expressed genes among BC and controls (adj p-value <0.05) are reported in **Table 18**. The same analyzes were conducted selecting NMI BC samples with respect to the control samples (**Figure 11**; differentially expressed probes with adj p-value <0.05 in **Table 18**).

There was a significant decreased expression of the *MMP23B* gene and its relative pseudogene *MMP23A* in BC cases compared to controls (Log FC= -0.37, FDR adj p-value=0.02 and Log FC= -0.23, FDR adj p-value=0.03, respectively, adjusted for age, smoking and experimental batch). These differences also arose when only NMI BC were compared with controls (*MMP23B* Log FC= -0.39, FDR adj p-value=0.02; *MMP23A* Log FC= -0.24, FDR adj p-value=0.04; adjusted for age, smoking and experimental batch).

MMP23A and *MMP23B* are highly homologous. Alignment (BLAST) of ILMN_2193325 and ILMN_1790052 probes showed that they failed to be specific. Therefore, despite Illumina indication, probes could not distinguish between the two genes.

Table 17 | **Gene expression analysis: baseline characteristics of the BC patients and controls (Discovery, Validation and Replication sets)**

	Discovery Set		Validation Set		Replication Set	
Covariates	**Controls (%)**	**Cases (%)**	**Controls (%)**	**Cases (%)**	**Controls (%)**	**Cases (%)**
N	70	66	57	44	24	21
Non-muscle invasive BC	-	56 (84.8)	-	39 (88.6)	-	21 (100)
Muscle invasive BC	-	10 (15.2)	-	5 (11.3)	-	-
Age (years)						
Mean ± SD	61.4± 8.6	62.3 ± 7.8	62.3 ± 8.6	61.7 ± 7.7	63.2 ± 8.8	63.6 ± 7.5
Range	40.3-74.6	40.0-73.0	40.3-74.6	41.8-72.4	40.9-74.9	47.1-73.9
Smoking Status						
Never	10 (14.3)	11 (16.7)	9 (15.8)	6 (13.6)	7 (29.2)	3 (14.3)
Former	43 (61.4)	40 (60.6)	33 (57.9)	27 (61.4)	13 (54.2)	13 (61.9)
Current	17 (24.3)	15 (22.7)	14 (24.6)	11 (25)	4 (16.7)	5 (23.8)
T stage*						
Ta	-	34 (51.5)	-	25 (56.8)	-	9 (42.9)
T1	-	20 (30.3)	-	12 (27.3)	-	11 (52.4)
≥T2	-	10 (15.1)	-	5 (11.4)	-	0
Tis	-	2 (3.1)	-	1 (2.3)	-	0
Grading (1973)*						
G1	-	17 (25.8)	-	11 (25)	-	6 (28.6)
G2	-	23 (34.8)	-	16 (36.4)	-	9 (42.9)
G3	-	26 (39.4)	-	17 (38.6)	-	6 (28.5)
Grading (2004)*						
Non-high grade	-	30 (45.5)	-	21 (47.7)	-	10 (47.6)
High grade	-	36 (54.5)	-	23 (52.2)	-	11 (52.4)
Risk*						
Low-risk	-	17 (25.7)	-	12 (27.3)	-	6 (28.6)
Intermediate Risk	-	14 (21.2)	-	10 (22.7)	-	4 (19)
High-risk	-	25 (37.9)	-	16 (36.4)	-	11 (52.4)
Muscle invasive		10 (15.2)		2 (4.5)		0
Recurrence						
No	-	41 (62.1)	-	29 (65.9)	-	13 (61.9)
Yes	-	25 (37.9)	-	14 (31.8)	-	8 (38.1)
Number of recurrences						
1	-	14 (56.0)	-	7 (15.9)	-	4 (19)
2	-	8 (32.0)	-	6 (13.6)	-	3 (14.3)
≥3	-	3 (12.0)	-	1 (2.3)	-	1 (4.8)
Progression						
No	-	64 (97.0)	-	42 (95.5)	-	18 (85.7)
Yes	-	2 (3.0)	-	2 (4.5)	-	3 (14.3)
Therapy						
No	-	18 (27.3)	-	15 (34.1)	-	5 (23.8)
Yes	-	48 (72.7)	-	29 (65.9)	-	16 (76.2)
Cystectomy						
No	-	52 (78.8)	-	34 (77.3)	-	16 (76.2)
Yes	-	14 (21.2)	-	10 (22.7)	-	5 (23.8)
BC patient` status at follow up						
Alive	-	54 (81.8)	-	36 (81.8)	-	16 (76.2)
Died	-	12 (18.2)	-	8 (18.2)	-	5 (23.8)

*referred to [5, 100]

Table 18 | **Differentially expressed genes among all BC cases (*n*=66) VS controls (*n*=70) and NMIBC (*n*=56) VS controls (*n*=70).** Analyzes adjusted for age, smoking and experimental batch

All BC cases VS controls (n=70)							
Gene Name	**Probe ID**	**logFC**	**Average Expression**	**p-Value**	**Adj p-Value**	**mean cases**	**mean controls**
MMP23B	ILMN_2193325	-0.371	7.633	8.34E-07	0.0168	7.4448	7.8112
MMP23A	ILMN_1790052	-0.229	6.96	2.73E-06	0.0274	6.8397	7.0701

NMIBC cases VS controls							
Gene Name	**Probe ID**	**logFC**	**Average Expression**	**p-Value**	**Adj p-Value**	**mean NMIBC cases**	**mean controls**
MMP23B	ILMN_2193325	-0.387	7.64	1.09E-06	0.022	7.426	7.811
MMP23A	ILMN_1790052	-0.236	6.969	4.39E-06	0.044	6.838	7.07

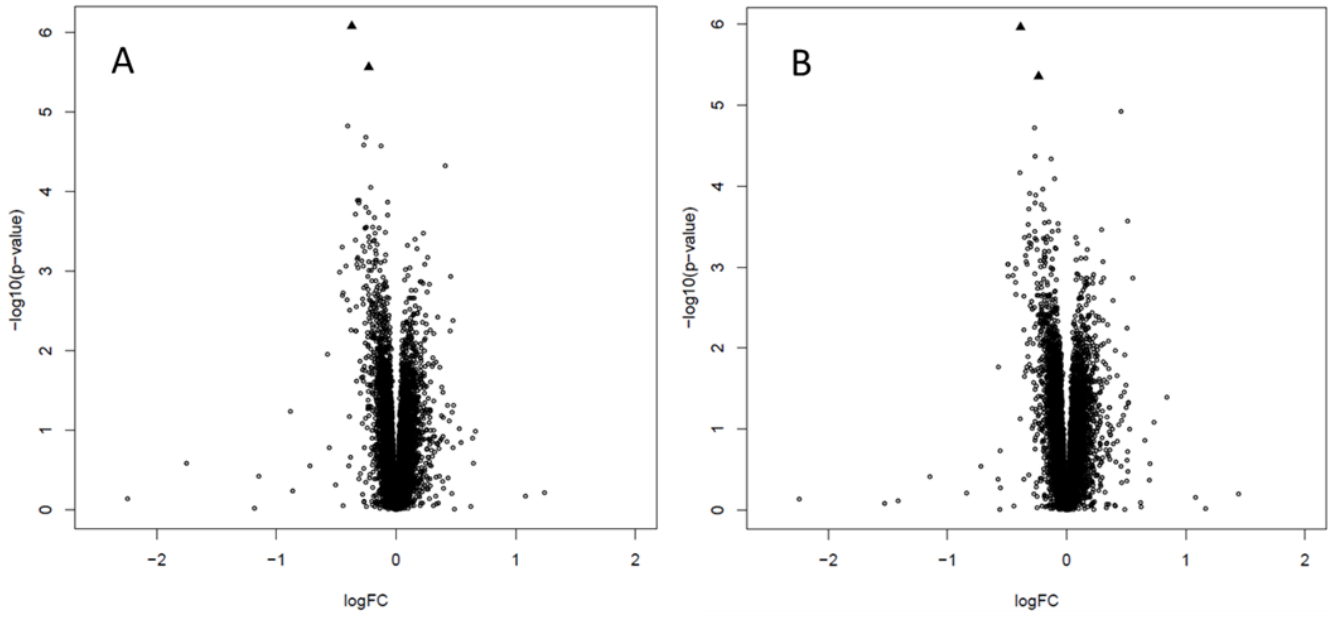


Figure 11 | **Volcano plots of expression levels of genes in the Illumina Human BeadChip HT12 V4 whole genome expression array in all BC (A) and in NMIBC patients only (B) versus controls.** For both plots, triangles are ILMN_2193325 and ILMN_1790052 probes for *MMP23A* and *MMP23B* genes, while squares are those genes non statistically significant but with an higher logFC

Validation and replication of MMP23A and MMP23B gene expression by qPCR

The differential expression levels of *MMP23A* and *MMP23B* genes were confirmed in the validation set (44 BC cases and 57 controls) by qPCR. Also in this case, despite TaqMan probe manufacturer's indications, the positions recognized by Hs00270380_m1 and Hs04187882_g1 probes fail to be specific and, due to the high homology between the two genes, the probes could not recognize between *MMP23A* and *MMP23B*.

Specifically, both genes resulted significantly down-regulated in BC cases when compared with controls (*MMP23* Hs00270380_m1 probe: t test p=0.014; adj p-value=0.02; *MMP23* Hs04187882_g1 probe: Wilcoxon p-value=0.005; adj p-value=0.012; adjusted for age and smoking, **Figure 12A** and **B**). In the replication set (21 BC cases and 24 controls), trends were similar but not statistically significant (*MMP23* Hs00270380_m1 probe: t test p=0.73; adj p-value=0.81; *MMP23* Hs04187882_g1 probe: t test p-value=0.30; adj p-value=0.29; adjusted for age and smoking, **Figure 12C** and **D**). Both in validation set and in replication set we observed a high correlation between the two *MMP23* probes (validation set Pearson p-value=$2.2x10^{-16}$, rho=0.858; replica set Pearson p-value=$2.2x10^{-16}$, rho=0.959).

When the analysis was performed in the total number of cases and controls (validation + replication sets, n=146, 65 BC cases and 81 controls), we observed a significant down-regulation of gene expression in cases compared to controls (*MMP23* Hs00270380_m1 probe: Wilcoxon p-value=0.037; adj p-value=0.15; *MMP23* Hs04187882_g1 probe: Wilcoxon p-value=0.006; adj p-value=0.03; adjusted for age and smoking; data not shown).

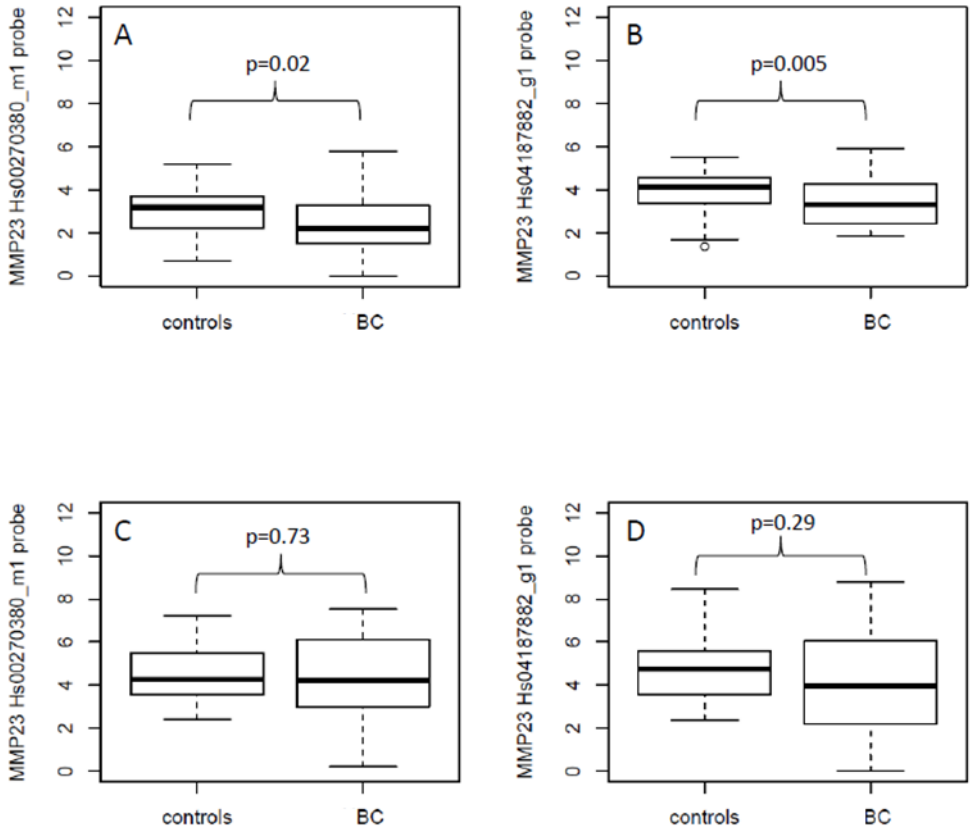


Figure 12 | **Expression levels of *MMP23* Hs00270380_m1 and *MMP23* Hs04187882_g1 probes in BC cases versus controls in the Validation (A-B) and Replication (C-D) sets.** P-values of significant comparisons are reported in the figure (multivariate analysis, model adjusted for age and smoking status)

MMP23A and MMP23B gene expression validation on TCGA

For RNA-seq, the TCGA dataset available on Wanderer tool consists of cancer tissues and 19 adjacent nonmalignant urothelial mucosa from 267 muscle invasive BC patients.

For all *MMP23B* and *MMP23A* detected transcripts there was a significant down-regulation in the tumor tissues compared to normal mucosa (**Table 19**; adj p-values ranging from 0.002 to 1.33×10^{-7} for *MMP23B* and from 4.42×10^{-7} to 1.42×10^{-7} for *MMP23A*). On the other hand, in a "per gene analysis", only *MMP23B* resulted significantly down-regulated in tumor tissues compared to normal mucosa (Wilcoxon p-value=8.61×10^{-9}; data not shown), being *MMP23A* not expressed in the majority of samples.

Table 19 | ***MMP23A* and *MMP23B* expression levels in MIBC tumor tissues and adjacent nonmalignant urothelial mucosa from TCGA database**

Gene name, ENSEMBL_geneID, ENSEMBL_transcriptID	chr	exon (chr: exon start-exon end: Strand)	Gene start	Gene end	Gene biotype	Normal tissues (n=19)		Tumor tissues (n=267)		wilcox_stat	pval	adj.pval
						Mean	SD	Mean	SD			
MMP23A, ENSG00000215914, ENST00000234610	chr1	chr1:1631378-1631511:+	1631369	1633249	pseudogene	1.946	0.617	1.001	0.738	4296	4.42E-07	4.42E-07
		chr1:1631786-1631953:+	1631369	1633249	pseudogene	3.568	0.795	2.087	1.066	4417	6.77E-08	1.42E-07
		chr1:1632800-1632926:+	1631369	1633249	pseudogene	4.164	1.095	2.408	1.224	4384	1.14E-07	1.52E-07
		chr1:1633047-1633247:+	1631369	1633249	pseudogene	3.589	0.977	1.946	1.135	4414	7.10E-08	1.42E-07
MMP23B, ENSG00000189409, ENST00000378675	chr1	chr1:1567560-1567753:+	1567474	1570639	protein_coding	0.286	0.326	0.138	0.253	3564	0.002	0.002
		chr1:1568162-1568292:+	1567474	1570639	protein_coding	2.377	1.015	1.020	0.813	4333	2.52E-07	3.03E-07
		chr1:1568360-1568734:+	1567474	1570639	protein_coding	3.424	1.089	1.754	1.030	4400	8.87E-08	1.33E-07
		chr1:1569127-1569478:+	1567474	1570639	protein_coding	4.030	1.021	2.176	1.161	4477	2.55E-08	1.33E-07
		chr1:1569581-1569707:+	1567474	1570639	protein_coding	4.900	1.073	2.971	1.313	4414	7.10E-08	1.33E-07
		chr1:1569827-1570029:+	1567474	1570639	protein_coding	4.020	1.029	2.262	1.176	4412	7.33E-08	1.33E-07

SD = standard deviation

MMP23B protein detection by Western Blot

The significant difference in *MMP23B* gene expression was further analyzed as protein expression by Western Blot analysis both in plasma and urine samples in 9 BC cases and 9 controls. The aim was to evaluate if MMP23B is present as secreted form in these biofluids, since this piece of information is still lacking[173]. Interestingly, we observed the presence of MMP23B protein in both biofluids, in both groups, with a molecular weight of about 50 kDa (**Figure 13**). As we did not perform a quantitative Western Blot, we could not speculate on differences between cases and controls. However, we could hypothesize that the amount of protein in plasma was higher than in urine, as plasma samples were diluted 1:500, while urine samples were run undiluted.

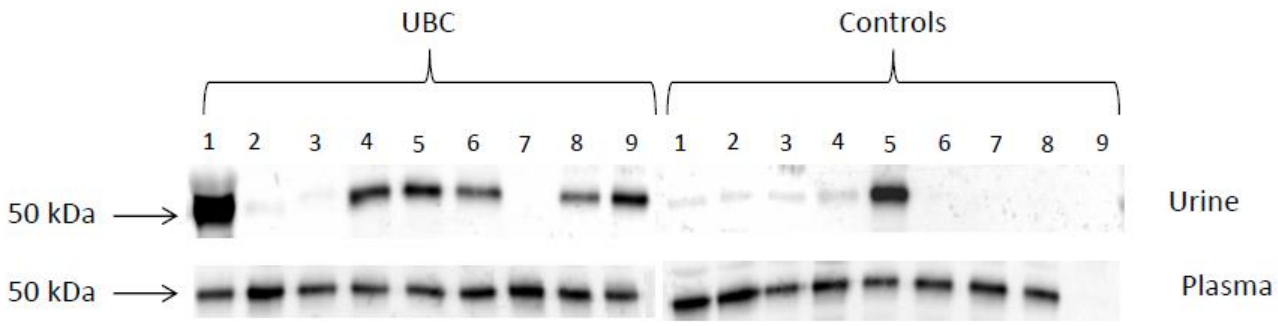


Figure 13 | **Western blot analysis on 9 BC cases and 9 controls in plasma and in urine samples**

MMP23B protein quantification by ELISA assay

We quantified MMP23B protein levels in plasma (49 controls and 53 BC) and urine (57 controls and 59 BC) samples from the same study (the description of subjects is included on **Table 20**). We observed an increase of MMP23B levels in BC cases respect to controls, both in plasma and in urine samples, with the difference statistically significant only in urine (plasma: mean BC=697.60 ng/ml, mean controls=543.80 ng/ml, p-value=0.17; urine: mean BC=1797.70 pg/ml, mean controls=1075.73 pg/ml, p-value=0.02, **Figure 14**). Similar results, although not statistically significant, were obtained comparing controls with NMI BC (plasma: mean NMI=701.32 ng/ml, mean controls=543.80 ng/ml, p-value=0.19; urine: mean NMI=1637.54

pg/ml, mean controls=1075.73 pg/ml, p-value=0.06, data not shown). The levels of MMP23B were higher in plasma than in urine, as already visualized by Western Blot experiments.

The correlation between plasma and urine measurements was not significant (n=42, Spearman correlation p-value=0.29, rho=0.167; data not shown). When we performed a general logistic regression analysis adjusting for smoking habits and age (and for creatinine levels for urine), we obtained significant differences between cases and controls in urine samples, both analyzing all cases and NMI BC only (All BC p-value=0.02; NMI p-value=0.01; data non shown).

OS and EFS were not affected by MMP23B protein levels in plasma or urine (data not shown).

Plasma MMP23B protein levels were not correlating with clinical characteristics of tumors. On the other hand, urinary MMP23B levels in cases were significantly different across tumor grading (all BC: WHO 1973 G3 vs G1 adj linear regression p-value <0.0001; G2 vs. G1 p-value=0.02, **Figure 15A**; WHO 2004/2016 adj linear regression p-value=0.03, **Figure 15B**; NMI: WHO 1973 G3 vs G1 ad linear regression p-value <0.0001; G2 vs. G1 p-value=0.02; WHO 2004/2016 adj linear regression p-value=0.17; data not shown). We also observed a significant positive trend of correlation between urinary MMP23B and tumor grading (WHO 1973 trend test, all BC: p-value<0.0001; NMI: p-value<0.0001). Finally, urinary MMP23B levels were significantly different also between risk classes (all BC: High vs Low risk adj linear regression p-value=0.0009, **Figure 15C**; NMI: High vs Low risk adj linear regression p-value=0.001; data not shown), with a significant positive trend in both all BC and NMI BC (trend test BC p-value<0.0001; NMI p-value=0.001).

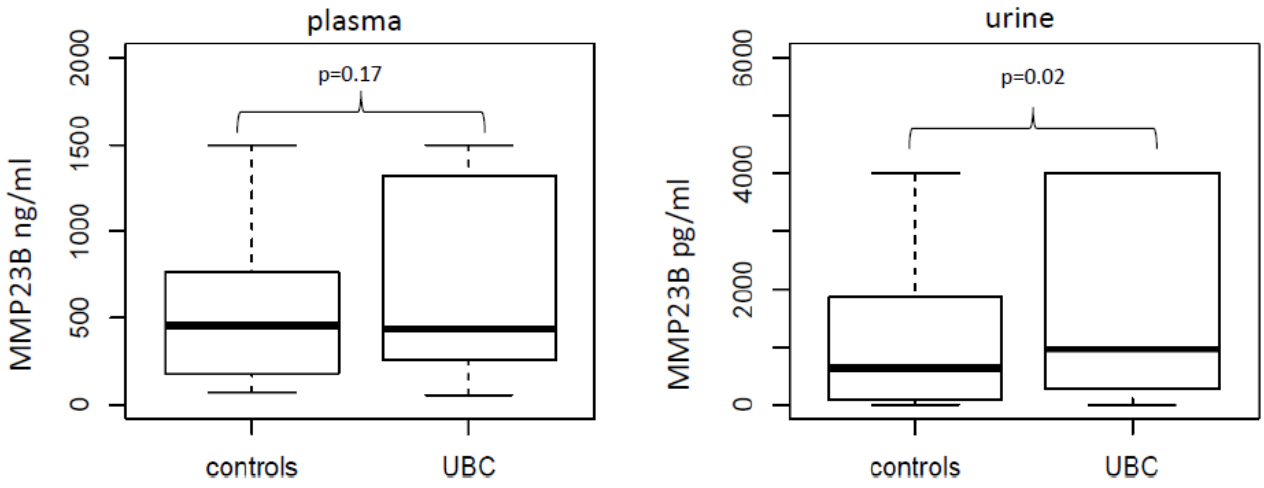


Figure 14 | **MMP23B protein levels in plasma and urine samples from BC cases and controls.** P-values of significant comparisons are given in the figure (multivariate analysis, model adjusted for age, smoking status, and for creatinine in urine)

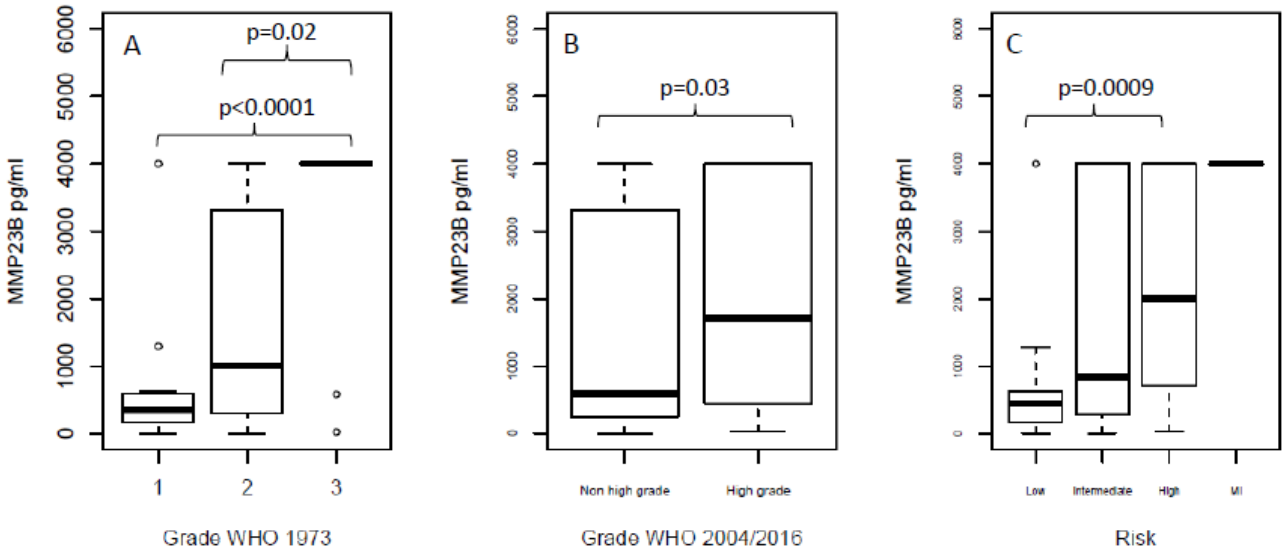


Figure 15 | **Urinary MMP23B levels in BC cases at diagnosis according to tumor grading WHO 1973 (A); tumor grading WHO 2004/2016 (B); and risk classes (C).** P-values of significant comparisons are reported in the figure (multivariate analysis, model adjusted for age, smoking status and creatinine)

Table 20 | **MMP23B protein quantification in plasma and urine samples: baseline characteristics of the BC patients and controls**

	Plasma		Urine	
Covariates	**Controls (%)**	**Cases (%)**	**Controls (%)**	**Cases (%)**
N	49	53	19	21
Non-muscle invasive BC	-	51	-	20 (95.2)
Muscle invasive BC	-	1	-	1 (4.8)
Age (years)				
Mean ± SD	65.1 ± 7.3	62.7 ± 7	65 ± 9.5	61.2 ± 7.5
Range	46.4 – 74.9	46.8 – 73.4	46.4 ± 74.9	46.8 ± 69.1
Smoking Status				
Never	7 (14.3)	5 (9.4)	2 (10.5)	2 (9.5)
Former	29 (59.2)	33 (62.3)	11 (57.9)	11 (52.4)
Current	13 (26.5)	15 (28.3)	6 (31.6)	8 (38.1)
T stage				
Ta		28 (52.8)		8 (38.1)
T1		19 (35.8)		11 (52.4)
≥T2		2 (3.8)		1 (4.8)
Tis		2 (3.8)		1 (4.8)
Grading (1973)				
G1		13 (24.5)		3 (14.3)
G2		20 (37.7)		10 (47.6)
G3		20 (37.7)		8 (38.1)
Grading (2004)				
Non-high grade		25 (47.2)		7 (33.3)
High grade		28 (52.8)		14 (66.7)
Risk				
Low-risk		14 (26.4)		3 (14.3)
Intermediate Risk		13 (24.5)		5 (23.8)
High-risk		23 (43.4)		12 (57.1)
Muscle invasive		2 (3.8)		1 (4.8)
Recurrence				
No		34 (64.2)		11 (52.4)
Yes		19 (35.8)		10 (47.6)
Number of recurrences				
1		8 (15.1)		3 (14.3)
2		7 (13.2)		5 (23.8)
≥3		4 (7.5)		2 (9.5)
Progression				
No		52 (98.1)		20 (95.2)
Yes		1 (1.9)		1 (4.8)
Therapy				
No		18 (34)		7 (33.3)
Yes		35 (66)		14 (66.6)
Cystectomy				
No		44 (83)		17 (81)
Yes		9 (17)		4 (19)
BC patient` status at follow up				
Alive		47 (88.7)		19 (90.5)
Died		6 (11.3)		2 (9.5)

*referred to [5, 100]

Analysis of urinary miRNAs targeting MMP23B gene

We investigated urinary miRNA profiles in association with BC and different clinic-pathological subtypes by next-generation sequencing (NGS). In total, 114 urine samples (66 BC cases and 48 controls) were sequenced. Among BC cases, 39 were G1+G2 and 17 G3. The analysis of the raw reads has led to the definition of the starting count matrix composed of 114 samples and 2560 miRNAs, of which 343 were significantly differentially expressed (i.e. FDR p-value < 0.05) among BC cases and controls.

With miRWalk 2.0 we searched all miRNAs targeting *MMP23B* and *MMP23A*. We did not find predicted miRNAs targeting *MMP23A,* while we found predicted miRNAs targeting *MMP23B* in the canonical binding site on 3`UTR of the gene, but also on 5`UTR and on gene promoter. We focused only on those miRNAs that were predicted at least from three tools over four (miRWalk 2.0, miRanda, RNA22, TargetScan). We obtained a list of 31, 8, and 330 predicted miRNAs binding sites in the 3`UTR, 5`UTR and gene promoter, respectively. When searching miRNAs targeting MMP23B among the differentially expressed in urine after RNASeq, we found that five urinary miRNAs targeting *MMP23B* 3`UTR were up-regulated in BC cases and controls (miR-93-5p, miR-106b-5p, miR-20a-5p, miR-17-5p and miR-3934-5p). Interestingly, 39 miRNAs targeting *MMP23B* in the promoter region resulted deregulated in the urine of cases compared to controls and one in particular, miR-34a-5p was also targeting the gene in the 5`UTR region (**Table 21**).

Table 21 | **miRNAs targeting *MMP23B* differentially expressed between BC cases and controls in urine.** Only miRNAs that were predicted from at least 3 tools over 4 used (miRWalk 2.0, miRanda, RNA22, TargetScan) are listed in the table. miRNAs are divided according to the region they are targeting on *MMP23B* gene

miRNA	Mean read counts	log2 Fold Change	Adjusted p-value (FDR)	*MMP23B* 3'UTR	*MMP23*B 5' UTR	*MMP23*B promoter
hsa-miR-93-5p	120.057	1.845	3.94E-12	x		
hsa-miR-17-5p	50.741	1.315	1.39E-06	x		
hsa-miR-20a-5p	131.101	1.237	2.81E-06	x		
hsa-miR-106b-5p	10.785	1.17	6.89E-06	x		
hsa-miR-3934-5p	4.023	0.899	2.24E-02	x		
hsa-miR-34a-5p	6.972	1.086	5.26E-03		x	x
hsa-miR-106b-3p	97.242	1.678	2.34E-13			x
hsa-miR-4792	131.192	2.378	5.22E-11			x
hsa-miR-483-5p	19.464	3.373	5.88E-10			x
hsa-miR-4651	1.535	-2.557	2.80E-05			x
hsa-miR-4492	70.869	1.573	2.93E-05			x
hsa-miR-1291	19.412	1.801	7.03E-05			x
hsa-miR-1301-3p	29.346	0.848	7.25E-05			x
hsa-miR-6760-5p	1.091	-2.329	8.70E-05			x
hsa-miR-30b-3p	7.466	-1.055	1.09E-04			x
hsa-miR-1275	1.545	-1.581	2.17E-04			x
hsa-miR-1471	0.987	-2.244	5.34E-04			x
hsa-miR-6127	0.678	-2.194	7.20E-04			x
hsa-miR-149-3p	2.372	-1.356	1.25E-03			x
hsa-miR-4443	6.431	-1.371	1.76E-03			x
hsa-miR-3180-3p	2.397	2.12	1.78E-03			x
hsa-miR-3180	2.313	2.059	2.97E-03			x
hsa-miR-1294	0.866	1.658	3.26E-03			x
hsa-miR-4740-3p	6.173	-1.323	3.28E-03			x
hsa-miR-3978	0.674	-2.57	3.60E-03			x
hsa-miR-1228-5p	7.034	-1.374	6.10E-03			x
hsa-miR-4459	90.928	-0.87	6.90E-03			x
hsa-miR-1908-3p	0.588	2.73	7.34E-03			x
hsa-miR-4694-3p	0.191	-3.033	NA			x
hsa-miR-1306-3p	3.14	1.147	1.28E-02			x
hsa-miR-4532	6.412	1.074	1.48E-02			x
hsa-miR-3191-3p	0.518	-1.754	1.54E-02			x
hsa-miR-4692	0.439	-2.033	1.59E-02			x
hsa-miR-5091	0.794	-1.836	1.91E-02			x
hsa-miR-548as-3p	0.156	-3.135	NA			x
hsa-miR-6806-5p	1.215	-1.955	2.47E-02			x
hsa-miR-1262	4.229	-0.768	2.55E-02			x
hsa-miR-6763-5p	0.555	-1.356	3.00E-02			x
hsa-miR-6786-3p	0.34	2.082	3.58E-02			x
hsa-miR-4472	1.002	-1.212	3.71E-02			x
hsa-miR-370-3p	4.769	-0.85	3.86E-02			x
hsa-miR-1827	6.758	-0.826	4.20E-02			x
hsa-miR-328-5p	0.976	-1.35	4.20E-02			x
hsa-miR-4486	0.258	-2.187	4.69E-02			x

5.1.3. Discussion

In the present study, we reported an altered expression levels of *MMP23B* gene transcripts in PBMCs of BC patients compared to control subjects. The present outcomes indicate a potential role of *MMP23B* and its pseudogene *MMP23A* as possible non-invasive BC biomarkers. We also analyzed MMP23B protein levels in plasma and urine from BC cases and we observed significantly higher levels of the protein in urine from cases respect to controls. Urinary MMP23B levels also correlated with risk and grade of tumor. These results are of particular importance as: 1) they are relevant in the field of non-invasive biomarker detection of diagnosis for bladder cancer; 2) they added piece of information on MMP23B characterization; 3) they demonstrated the secretion of MMP23B protein both at plasma and urine levels.

The identification of a biomarker for BC screening and early diagnosis has primary importance in improving survival and quality of patients' life. Urine cytology is currently the most commonly used non-invasive test for BC detection. However, this test is of limited value owing to its poor sensitivity, especially for low-grade lesions [174]. Cystoscopy-guided biopsy for histological evaluation can offer high diagnostic accuracy, but it is invasive and inconvenient for patients, limiting its use in general cancer screening. BC is among the most expensive cancers and poses a significant economic and social challenge because the high rate of recurrence necessitates continuous cystoscopic surveillance [175]. Hence, non-invasive and more sensitive molecular biomarkers are needed to improve current strategies for detection and monitoring of BC [176].

We performed a whole genome gene expression analysis to identify in PBMCs from BC patients and we identified an altered expression of *MMP23A* and *MMP23B* genes between cases and controls as unique significant signal, both in overall BC and in NMI BC. We tested gene expression in blood cells with the aim to find non-invasive BC biomarkers, since previous studies reporting gene expression profiling in BC were all performed on cancer tissue [177-181].

The two *MMP23* genes, referred to as *MMP23A* (pseudogene) and *MMP23B*, produce three alternatively spliced products each [182]. Unfortunately, due to the high homology between *MMP23A* and *MMP23B*, both microarray and qPCR probes used in the present study could not distinguish specifically between the 2 species.

It is still unknown the physiological function of MMP23B although there is evidence on its possible role in various cellular processes including bone matrix resorption and bone remodeling[183]. *MMP23B* gene is implicated in 1p36 deletion syndrome as it is located in the critical region for large, late-closing anterior fontanel [184].

TCGA provides a considerable resource for the molecular characterization of many tumor types, including BC [160, 167]. The decrease of *MMP23B* expression was confirmed in bladder cancer tumor tissue from TCGA, as well as in tumor prostate tissue from a previous study [185] (prostate cancer is sharing a common carcinogenic process with BC [186]). However, the TCGA bladder cohort includes only patients with advanced muscle invasive tumors, while our study population was mostly represented by samples from non-muscle invasive BC patients (>80%).

Since no additional aliquots of cryopreserved PBMCs were available, the measurement of levels of MMP23B protein in the blood cells was not possible. As consequence, we tested the presence of the protein in body fluids (plasma and urine) of BC patients and controls from the same cohort. We observed an increase of MMP23B in BC cases respect to controls, although the differences were statistically significant only in urine. Interestingly, a significant trend of MMP23 dosage was observed in BC urine samples in relation to the risk or grading.

The increasing MMP23B protein levels in plasma and urine seems to contrast with the altered gene expression levels found in PBMCs. However, Vogel and Marcotte showed that the correlation between mRNAs and protein levels can be as little as 40% depending on the system [187]. There are, in fact, many processes between transcription and translation that could affect correlation and protein stability is one of this. The half-life of different proteins can vary from minutes to days - whereas the degradation rate of mRNA would fall within a much tighter range (2-7 hrs for mammalian mRNAs vs 48hrs for protein [187]). Other important factors include the lower rate of mRNA transcription compared to protein translation in mammalian cells (2 mRNAs transcribed per hour - on average) versus dozens of proteins per mRNA per hour.

The lack of correlation between mRNA and protein level could be due also to a post-transcriptional crosstalk mediated by miRNAs competing for both *MMP23A* and *MMP23B* genes. Recently, it has been demonstrated that the interaction of the miRNA seed region with mRNA is not unidirectional. More specifically, the pool of mRNAs, transcribed pseudogenes, and other non-coding RNAs compete for the same pool of miRNAs thereby regulating their activity [188]. These competitive endogenous RNAs (ceRNAs) act as molecular sponges for miRNAs through their miRNA response elements (MRE), thereby de-repressing all target genes of the respective miRNA family [189]. To elucidate the possible miRNA/MMP23B crosstalk, we look for miRNA targeting *MMP23B* gene by dedicated tools. We identified several candidate miRNAs and among them we showed in urine a differential expression between BC cases and controls of 5 miRNAs targeting 3'UTR region of *MMP23B*. In particular, miR-93-5p, miR-106b-5p, miR-20a-5p, miR-17-5p and miR-3934-5p resulted upregulated in BC. It is interesting to notice that miR-93-5p, miR-106b-5p and miR-17-5p have been previously described as associated with bladder cancer. More specifically, miR-93-5p was found upregulated both in tumor tissue and blood cells from BC patients, but not in urine [190]. On the other hand, miR-106b resulted upregulated in BC preoperative cell-free urine compared with postoperative samples and its levels were associated with advanced tumor stage [191]. Finally, miR-17-5p was identified as one of the inflammasome-related miRNAs in BC urine samples and was significantly upregulated in high grade, high-risk and non-muscle invasive high grade BC patients [192].

The involvement in cancer of MMP23B was previously demonstrated. Moogk and collaborators showed an inverse association between primary melanoma, MMP23B protein expression in tumor tissue and the anti-tumor T cell response, suggesting this protein as a new immunotherapeutic target [193]. Although it has been observed that MMP23B promotes invasiveness in breast cancer cells [194], to our knowledge there is no other evidence of its role in the invasion of any cancer or even the detection in any cancer of altered levels of MMP23B in biofluids.

The only study that previously detected MMP23B in biofluids such as urine was performed by Chinello and colleagues[195]. Authors identified by MALDI-TOF profiling techniques several peptides whose urinary levels varied according to tumor size, stage and grading of renal cell cancer (RCC). In particular, MMP23B

urinary levels were lower in RCC patients compared to controls, have a negative alteration correlating with tumor size, but an increase in higher grade tumors. In this study, we detected for the first time MMP23B protein in both plasma and urine of BC. Pei et al. have described a secreted MMP23B in the form of a fully processed mature glycosylated enzyme of 66 kDa, significantly larger than the intracellular form at 56 kDa [173]. In the present study, we observed a protein of 50 kDa reflecting then more the previously observed intracellular form. Further biochemical analyzes are necessary to better understand the physiological and pathological roles of MMP23B secreted forms.

Other MMPs have been already identified in urine or blood of BC patients[196]. In particular MMP7 and MMP9 have been detected in urine from BC patients as reliable prognostic biomarkers associated with advanced stage disease or poor survival [67, 197-200]. MMPs, including MMP23B, could be used in combination with other biomarkers to increase accuracy. BC patients need a constant, lifelong surveillance after diagnosis, thus MMPs could be employed to allow the identification of patients at higher risk of recurrence [196].

We are aware of some limitations of the present study. First of all, we know that analytical performance of ELISA assay could be affected by urine complexity, that may hinders the efficient binding of a protein to its corresponding antibody used in the assay [201]. However, data obtained by Western Blot constitute a "visual" correlation with ELISA quantification, giving us a certain confidence on the real levels of urinary MMP23B measured by ELISA assay.

Urine are of particular interest in BC biomarker research thanks to their direct contact with the site of the tumor and because they may contain high concentrations of tumor-derived proteins. Moreover, urine is an ideal biofluid for biomarker research because it can be collected easily and non-invasively in different times and in high quantities with minimal risk for donors. The major disadvantages of spot urine samples are the variation in dilution effects, in sample volume variation and the rate of urine production. In an attempt to adjust for this variation, urinary creatinine concentration is most commonly used in a ratio format to normalize analytes quantification for specimen concentration [202, 203]. We adjusted for creatinine levels all the linear regression analyzes performed on urinary MMP23B protein.

This study revealed that the detection of urinary MMP23B might be a useful tool for BC diagnosis. This is the first study reporting on the presence of a secreted form of MMP23B in human biofluids (both in plasma and in urine). The functions of MMP23B and its expression regulation by miRNAs is intriguing both to understand new mechanisms in BC onset and progression and to identify a group of urinary, non-invasive biomarkers useful to create a molecular profile of BC at different stages and different risk to develop recurrences.

6. Malignant Pleural Mesothelioma study

6.1. Introduction

Malignant mesothelioma (MM) is an asbestos-related tumor that arises primarily from transformation of mesothelial cells that line the serosal surfaces of the pleural, peritoneal and pericardial cavities [204].

Although the most common form of mesothelioma arises in the pleura (80-85%), much more rarely can also originate in the peritoneum (10-15%) and tunica vaginalis (~5%) [205].

Interaction of asbestos fibers with mesothelial and inflammatory cells is thought to initiate prolonged cycles of tissue damage, repair, and local inflammation, which finally lead to carcinogenesis of MM with unknown mechanisms [206]. Malignant pleural mesothelioma (MPM) arises from the mesothelial cells lining the parietal pleura: asbestos fibers are inhaled with the breathing and migrate into the pleura either directly through the lung or via lymphatic/blood circulation [207].

There are three main MM histologic types. (American Cancer Society. Survival statistics for mesothelioma*, http://www.cancer.org/cancer/malignantmesothelioma/detailedguide/malignant-mesothelioma-survival-statistics (2016)*).

- *Epithelioid* mesotheliomas constitute roughly 60% of all cases. Cancer cells of this type are uniform and tend to stick together, which means that it takes longer for them to spread throughout the body; this is the reason why patients with a purely epithelioid diagnosis tend to have a better survival prognosis (12-27 months).
- *Sarcomatoid* type constitute around 15% of mesotheliomas; sarcomatoid cells appear elongated and because of their weak organization they spread more easily throughout the body. Patients with sarcomatoid histology have the worst survival (7-18 months).
- *Biphasic* mesotheliomas contain combinations of both epithelioid and sarcomatoid features and are termed biphasic-epithelioid (biphasicE) or biphasic-sarcomatoid (biphasicS), respectively, depending on whether epithelioid or sarcomatoid cells predominate. They make up the remaining

25% of mesotheliomas. Patients with this form of cancer will have varying life expectancy rates (8-12 months), because prognosis will depend on the ratio of epithelial and sarcomatoid cells.

Several staging systems for mesothelioma have been proposed over the years, almost exclusively dealing with pleural mesothelioma. There are currently no formal staging systems for the other types. The oldest staging system, introduced by Butchart *et al* in 1976 ([208]) was based on a simple description of the extent of the disease regardless of histologic subtype, size of the tumor, number of cancer cells, or level of cancer diffusion through the body. In 1993 Dr. Sugarbaker and colleagues developed a new staging system, namely Brigham Staging System, that primarily assessed the viability of surgical intervention and lymph node involvement. According to the Brigham system, Stage I mesothelioma tumors are still viable candidates for removal without nodal spread, while Stage II is resectable but there is lymph node involvement; however, by the time the mesothelioma reaches the chest wall, heart, diaphragm, or abdominal cavity (Stage III) or metastasizes into other areas of the body (Stage IV), surgery is no longer a valid option [209]. In 1995, the International Mesothelioma Interest Group (IMIG) proposed a TNM-based staging system that nowadays is the most widely used system for MPM. As described in section 1.1, the abbreviation TNM represents three different parts of a diagnosis: T describes tumor size and location, N describes the involvement of lymph nodes and M defines whether the tumors have metastasized. This system was approved by both the International Union Against Cancer (UICC) and the American Joint Committee on Cancer (AJCC)[209] (**Table 22**).

No grading system universally approved exists that adequately predicts MPM prognosis. However, in 2012 Kadota and colleagues proposed a grading system linked to survival for MPM of epithelial variant, based on nuclear atypia and mitotic activity. Using this information, researchers were able to stratify patients into three survival groups: grade I (median survival of 28 months), grade II (median survival of 14 months) and grade III (median survival of 5 months) [210].

Table 22 | **TNM classification of malignant pleural mesothelioma (MPM)**

Stage	T,N,M	Description
I	T1, N0, M0	Mesothelioma has developed in the pleura lining one side of the chest wall and may affect the lining near the breathing muscle (diaphragm) or the area between the lungs (mediastinum).
II	T2, N0, M0	Mesothelioma has developed in the pleura lining one side of the chest wall and may affect the lining near the breathing muscle (diaphragm) or the area between the lungs (mediastinum).
III	T3, N0–N2, M0	Mesothelioma has developed in the pleura lining one side of the chest wall and may affect the lining near the breathing muscle (diaphragm) or the area between the lungs (mediastinum).
	T1–T2, N1–N2, M0	The primary tumor mass has not spread far beyond the pleura, but mesothelioma cells have spread to lymph nodes near the tumor.
IV	T4	Mesothelioma has penetrated into the abdominal lining, the spine, the deep layers of the chest wall, the pleura on the other side of the chest, or any vital organ between the lungs.
	N3	Mesothelioma cells have spread to lymph nodes far from the primary tumor mass.
	M1	Mesothelioma tumors have developed in distant parts of the body.

6.2. Epidemiology of malignant pleural mesothelioma

Risk factors and susceptibility

Asbestos, that is the main risk factor for MPM development, refers to a group of six minerals (chrysotile, crocidolite, amosite, anthophyllite, tremolite, and actinolite) which are able to form very thin fibers [211]; nowadays all types of asbestos are considered causes of human malignant mesothelioma according both to the WHO [212] and the International Agency for Research on Cancer (IARC) report [213].

Since the early 1900s evidence about the carcinogenicity of asbestos has been reported [214]. The first epidemiological studies of cancer risk in relation to asbestos exposure date back to the 1950s; for example, in 1955 Doll and colleagues reported a more than 10-fold excess of lung cancer among British asbestos textile workers [215]. Wagner and co-workers instead, described mesothelioma of the pleura and peritoneum due to asbestos exposure in 1960 [216]. With these findings started the research that resulted in the in the banning of asbestos production and import in several European countries after 1970, and in the European Union as late as 2005 [217]

Asbestos exposure is typically work-related, therefore MPM is recognized as an occupational disease (accounting for more than 80% of all cases) and has a long latency period of up to 40 years. Moreover, past exposures to asbestos were more common in occupations with a prevalent male workforce, explaining the predominance of MPM in men (men to women ratio is 4:1), with a median age between 50 and 70 years old and a median survival time of 12 months [218]. In the last decades, the exposure history of MPM cases shifted from primary asbestos-workers (handling raw asbestos materials, such as miners) to asbestos-using professionals (such as electricians or plumbers) and in the second group the number of subjects at risk of MPM is higher [211, 219].

Environmental exposure to carcinogenic fibers that exist on the earth surface is another more uncontrollable risk factor; many MPM cases have been diagnosed among people living near natural asbestos resources (Turkey, Cyprus, Corsica) or in areas close to asbestos mines or factories. In Italy, in Casale Monferrato (Eternit factory), Broni and Bari (Fibronit factory), MPM was registered among

inhabitants and one third of cases (for both regions) had only one risk factor: living close to an asbestos cement factory; calculated ORs were higher for those living close to the factory and the fiber burden in the lungs of deceased cases was 10-fold that in those from other areas [220].

Furthermore, some spots in Greece and China have been found to host clusters of mesothelioma: young Greek women patients had been exposed to tremolite, used for white-washing the houses, while Chinese farmers had been exposed to crocidolite-containing soil [217].

Other sources, such as asbestos cement roofs or buildings were associated with a significant increase in the risk of MPM development. A significantly increased OR was observed when father or mother or spouse were occupationally exposed to asbestos [221]; having a family member (any category) occupationally exposed to asbestos doubles the risk of MPM, due to the presence of asbestos fibers in the clothes of the exposed worker.

Whereas that asbestos remains the only causal factor for MPM development, other cofactors have been suggested to play a role in MPM development, including synthetic materials, ionizing radiation, SV40 virus, whilst tobacco smoking has been excluded [219].

Moreover, GWA and EWA studies indicate that genetic and epigenetic susceptibility may have a role in determining susceptibility to MPM carcinogenesis. Our group published the first genome-wide association study (GWAS) on 392 Italian MPM cases and 379 controls with complete information on asbestos exposure. Most of the top-signals we identified were located in chromosomal regions reported to harbor aberrant alterations in mesothelioma, and cause an at most 2–3 fold increase in MPM risk [222].

In 2001, for the first time, familial forms of MPM have been reported in the Cappadocia region of Turkey, were a higher incidence of MPM has been observed in those families mainly exposed to erionite [223]

Moreover, studies of MPM among Wittenoom (Australia) former workers and residents showed that individuals with a first-degree relative with MPM had an approximately two-fold increased risk of MPM, even after adjusting for degree of asbestos exposure [224].

In 2011, Testa and collaborators linked a germline mutation in *BAP1* gene to MPM (and other cancers) predisposition [225]. Moreover, they hypothesized that mutations in *BAP1* may contribute to susceptibility

to MPM in asbestos-exposed individuals through a mechanism that involves a gene–environment interaction. Recently, we investigated the prevalence of germline variants in a panel of 94 cancer-predisposing genes in 93 sporadic MPM patients from the UPO study. Ten pathogenic truncating variants (PTVs) in *PALB2*, *BRCA1*, *FANCI*, *ATM*, *SLX4*, *BRCA2*, *FANCC*, *FANCF*, *PMS1* and *XPC* genes were identified, all of them involved in DNA repair pathways and mostly in homologous recombination repair. Patients carrying PTVs were 9.7% of the panel , and showed lower asbestos exposure than the other patients (p=0.0015). This suggests that inefficient repair of asbestos-induced DNA damage may increase MPM susceptibility [226].

Finally, alterations in epigenetic profiles have been found in tumor tissues and may provide important insights into the etiology and natural history of MPM. For example, a specific hypermethylation of *TMEM30B*, *KAZALD1* and *MAPK13* genes in MPM cell lines and tissue samples from Japanese patients was suggested as potential diagnostic marker [227]; moreover, methylation changes of the promoter region of nine candidate genes has been reported [228], as well as altered microRNA expression [229].

In conclusion, genetic and epigenetic alterations may be hallmarks of the intricate road by which asbestos induces MPM and should be more deeply investigated in order both to elucidate key pathways altered by asbestos that drive tumorigenesis and to find new potential therapeutic targets to develop innovative therapies.

Incidence and mortality

Asbestos exposure is the main risk factor for the development of malignant mesothelioma and the trend in incidence and mortality is parallel to asbestos use [230].

In 2004, WHO estimates that 107,000 global annual deaths were caused by mesothelioma, asbestos-related lung cancer and asbestosis from occupational exposures to asbestos, with other hundreds resulting from non-occupational exposures.

Consistent differences in incidences are reported from different countries worldwide; they vary from 7 per million (Japan) to 40 per million (Australia) inhabitants per year. In Europe, the incidence is 20 per million

[219]. Is reasonable to accept that these rates are highly dependent to the amount of asbestos removal, import, and industrialization.

Global incidence of MPM has risen firmly in the past years, and is predicted to reach a peak in 2020 in Europe [231]; precise numbers are difficult to determine as the disease is likely to be under reported in areas of low incidence. Nowadays, mesothelioma incidence is highest in the USA and UK although per capita, Australia and Italy also rank highly [207].

While in countries such as the USA, Australia, UK and the Nordic European countries, asbestos use levelled off during the 1960s and 1970s and then decreased, in Italy, Spain and France, asbestos imports gradually decreased since the 1980s only, the decline thus starting some 10–20 years afterwards [232]. So far, 55 countries have banned the use of asbestos, including Italy since 1992 (*http://ibasecretariat.org/alpha_ban_list.php*). In conclusion, although the Western world is moving towards a levelling-off of asbestos-related diseases incidence through the banning or constricting use of asbestos, the developing world is still greatly increasing the use of this toxic material. Major producers are Russia, Kazakhstan, China and Brazil, and continue to export asbestos especially to low- and middle-income countries that too often have weak or non-existent occupational and environmental regulations [233]. Unfortunately the ongoing and unregulated use of asbestos means that MPM will continue to represent a significant global health concern.

Italy has been a significant producer and user of asbestos [234]. Chrysotile was mined in Balangero (Piedmont) and Canadian/South Africa crocidolite were imported [234].

The first and largest Italian factory started in 1907 in Casale Monferrato (Piedmont), the second largest in Broni (Lombardy) in 1932, followed by many others in Liguria, Puglia Campania and Sicily [221].

Many asbestos industries were located in Piedmont. Since 1990, the Malignant Mesothelioma Registry has recorded newly diagnoses cases of MPM among residents [221].

The area of Casale Monferrato showed an extremely high incidence of MPM caused by occupational and environmental asbestos exposure from the asbestos cement factory, the Eternit, that was operational until 1986 [235].

The average annual incidence in the period 1990-2010 was 27.3 (per 100,000) among men and 15 (per 100,000) among women, about 10 times higher than the corresponding Italian incidence rates [236]. The asbestos cement factory was active in the town until 1986 and was the major source of asbestos pollution [235, 237]. Compared with the general population of Piedmont, the inhabitants of Casale Monferrato had been shown to have a higher risk to develop MPM (OR= 20.6) [238].

MPM incidence in the Casale Monferrato area has risen over the years up to an age-standardized rate of 35.5 (95% CI 28.7 to 42.4) cases per 100,000 per year in 2008–2012 among men, and 15.9 (95% CI 11.6 to 20.1) among women (definite diagnosis) [221].

6.3. Symptoms, diagnosis and treatment

Initial symptoms of MPM typically include shortness of breath (dyspnea), chest pain and persistent dry or raspy cough. In the early stages of disease breathlessness is usually due to a pleural effusion, found in the majority of patients [207]. Unfortunately, many of the more serious symptoms, such as painful breathing, coughing blood (hemoptysis), difficulty swallowing (dysphagia) together with weight loss and fatigue, are noticeable only when the cancer has reached its advanced stages, when treatment options are usually more limited.

MPM is challenging to diagnose because its symptoms are nonspecific, that is they are the same as for other common respiratory diseases. Hence, for a correct diagnosis it is of great importance to obtain a detailed occupational history. Most patients with MPM have a background of asbestos exposure and some may have had antecedent symptoms associated with benign asbestos-related diseases such as asbestosis or pleuritic; others may have radiological evidence of past asbestos exposure, such as pleural plaques [239].

The first step toward MPM diagnosis is taking imaging tests such as X-rays or computed tomography (CT) scan of chest and abdomen to show fluid or tumors around the lungs; also cytology can be used to detect malignant cells, but further imaging and blood tests are necessary to confirm MPM diagnosis [240].

Finally, tissue specimens must often be obtained with biopsy, to demonstrate that the lesion has a mesothelial phenotype and that it shows neoplastic invasion, as opposed to benign involvement of mesothelium as part of inflammatory processes [239]. Procedures used include blind percutaneous biopsy, fine needle aspiration (FNA) biopsy, imaging-guided core biopsy, video-assisted thoracoscopy (VAT)-guided biopsy and thoracotomy. Both CT and VAT biopsies have high sensitivity and low complication rates; CTs are suitable for patients with pleural or nodular lesions, while VATs are recommended for other patients with a pleural effusion but no mass lesion [239].

There is no curative treatment for MPM; moreover, as previously discussed, MPM is usually diagnosed at an advanced stage, therefore it generally responds poorly to classic anti-cancer treatments such as chemotherapy, radiotherapy and surgery, whether delivered separately or as part of multimodality treatment [207, 241].

Surgical treatment aims to extend the survival of patients who have evidence of remaining tumor [242].

Since the full resection of mesothelioma is impossible, it is accepted that surgery is aimed at local disease control, elimination of pleural effusion, release of lung trapped by the tumor and relief of pain from chest wall invasion [243]. These considerations make the surgery particularly applicable to epithelioid mesothelioma, due to its better diagnosis with respect to the sarcomatous or biphasic types [242].

Currently, patients with resectable disease can be treated with two procedures: extrapleural pneumonectomy (EPP) or extended pleurectomy/decortication (P/D). EEP is the more radical solution and involves resection of lung, parietal pleura, pericardium, diaphragm and regional lymph nodes [207]. Due to its complexity also deriving from post-surgery treatment, EPP is recommended only in the context of controlled clinical trials when progressive disease is observed after neoadjuvant chemotherapy [244]. P/D aims primarily at releasing the lung and the chest wall from constriction caused by the tumor, removing part of the visceral and parietal pleura [207]. Unlike EPP, P/D does not aim for macroscopic complete resection.

Since surgery cannot eradicate all residual microscopic disease, a multimodality approach is encouraged; trimodality treatment, the most used, combines chemotherapy, surgery and radiotherapy. Patients with good performance status, epithelioid histology and low burden of disease are most likely to benefit from trimodality therapy [245].

Chemotherapy with cisplatin and pemetrexed has become the standard of care therapy for unresectable MPM patient [245]. This combination improves response, survival, time to progression, pulmonary function, and disease-related symptoms. Other active agents include raltitrexed, gemcitabine, and vinorelbine [246]. Effectiveness of any of these drugs will depend on a variety of factors including the stage of the cancer and the patient's overall health.

Radiation therapy is principally administered as a palliative measure (tumor bulk reduction, symptoms placation, etc) or as an adjuvant to surgery and chemotherapy in the context of trimodality treatment, giving acceptable results [207]; despite the early experience demonstrated alarming toxicity to intensity

modulated radiation therapy (IMRT) and other novel radiation treatment techniques, recent reports have demonstrated that it can be delivered safely both in the setting of EPP or P/D [247].

6.4. Malignant pleural mesothelioma biomarkers

Immunohistochemical evaluation is essential for identifying MM histological subtypes [248]. However, no single marker has 100% sensitivity and specificity, so various monoclonal antibody panels must be investigated, of which at least 2 must be positive for mesothelioma in order to achieve a diagnosis.

According to current studies, the most valuable marker in differentiating epithelioid MPM from lung and breast adenocarcinoma is calretinin (CR), followed by cytokeratin 5 (CK5), podoplanin (PDP), thyroid transcription factor 1 (TTF-1) and Wilms' tumor-1 protein (WT1) [249]. Carcinoembryonic antigen (CEA) is very useful for distinguishing metastatic carcinoma, particularly of pulmonary origin, from mesothelioma [250]. When the tumor contains a sarcomatous component, it often needs to be distinguished from metastases; although some antibodies used for epithelioid mesothelioma are equally valid for sarcomatous types, some others are often needed for firm evidence, such as p63 and MOC 31 [251].

Since MPM diagnosis usually occurs at clinically advanced stages, an early MPM detection could allow new therapeutic approaches at a potentially treatable stage.

In recent years, increasing attention has been paid to the role of blood-based non-invasive methods as a useful diagnostic tool.

Soluble mesothelin-related peptide (SMRP) is a candidate marker for prognostic purpose [252]. The evidence of increased serum SMRP concentrations in individuals with past exposure to asbestos and in subjects with pleural asbestos-related diseases suggests a possible role of SMRP as a marker of asbestos exposure [253-255].

Circulating biomarkers, such as osteopontin, mesothelin-related peptides, SMRP and fibulin-3 have also been evaluated in MPM extensively. However, none of them is considered to be a reliable screening tool, since false-positive results are too frequent [256-260].

Demir and colleagues investigated serum biomarkers in three groups, composed of healthy subjects exposed or not to asbestos, and MPM patients. They found out SMRP and Thioredoxin (TRX) increased in a graded fashion from the controls to asbestos exposed and MPM groups [261].

The most promising biomarker is high mobility group box protein 1 (HMGB1) [262]. Several studies reported that levels of the hyper-acetylated form of HMGB1 in serum reliably distinguished malignant mesothelioma patients, asbestos-exposed individuals, and unexposed controls [263-266].

A new approach, based on the evidence that tumors generate a characteristic miRNA fingerprint in the cellular fraction of peripheral blood [267], has shown that miR-103 levels were able to discriminate MPM patients from asbestos-exposed subjects and healthy controls [268]. Combining miR-103 with mesothelin has improved diagnostic performance [269].

7. Hypotheses and aims

As previously discussed, MPM is usually diagnosed at a late stage, therefore it generally responds poorly to chemotherapy and radiotherapy, and surgery is rarely curative [241] . The availability of tools for monitoring high risk individuals aiming at an early diagnosis of MPM could improve life expectancy, hence, to further ameliorate specificity and sensitivity of diagnosis, research into the development of novel biological markers is urgently required. Along this line we sought whether white blood cells (WBCs) DNAm profiles might associate with MPM occurrence in subjects with asbestos exposure, as the tumor environment in the pleura may trigger molecular changes in perfusing WBCs.

8. Peripheral blood biomarkers and malignant pleural mesothelioma

Here I report some extracts of our work

> Guarrera S*, Viberti C*, Cugliari G, Allione A, Casalone E, Betti M, Ferrante D, Aspesi A, Casadio C, Grosso F, Libener R, Piccolini E, Mirabelli D, Dianzani I, Magnani C, Matullo G, *Peripheral blood DNA methylation as potential biomarker for Malignant Pleural Mesothelioma in asbestos-exposed subjects,* Submitted to Journal of Thoracic Oncology
> ***equally contributed**

8.1. DNA methylation and malignant pleural mesothelioma

As described in previous sections, for subjects who underwent environmental or occupational asbestos exposure in the past there is no preventive measure currently available, except for continuous monitoring aiming at an early MPM diagnosis, that might improve life expectancy.

So far, several serum biomarkers were tested for early MPM diagnosis and prognosis, with limited results [270]. In recent years, there is a growing interest in DNA methylation (DNAm) profile changes as possible diagnostic/prognostic biomarkers in cancer research [271]. DNAm is a common epigenetic modification consisting in the covalent bound of a methyl group to the cytosine in the dinucleotide CpG, and plays a key role in gene expression regulation.

DNAm is stable, yet it may be modified across lifetime by several factors including lifestyle, environmental exposures, ageing, and diseases [272] and is thus rewarded as an adaptive phenomenon potentially linking environmental factors and development of disease phenotypes. Being an early event in tumor development [273], aberrant DNAm has been suggested as a tool for early cancer detection and prognosis [274, 275]. DNAm changes are regarded as possible actors also in MPM development and progression [276]. Previous works support the potential use of DNAm profiles in peripheral blood as diagnostic biomarkers for several tumors, among which ovarian [277], bladder [58] and breast [278] cancers.

As of today, no sensitive screening method is in fact available to monitor asbestos exposed subjects, at high risk of developing MPM, aiming at an early diagnosis. Thus, the identification of reliable MPM diagnostic

biomarkers in peripheral blood or other easily accessible body fluids might provide the tool for detecting the disease at a more treatable stage.

8.1.1. Material and Methods

Study population

Study subjects belong to a wider ongoing collaborative study on MPM, which is actively enrolling MPM cases and healthy subjects in the municipalities of Casale Monferrato (Piedmont region, Italy), where a cement-asbestos factory was active for about 80 years until 1986 causing occupational and environmental asbestos exposure for dwellers, and Turin (Piedmont, Italy). Cases were enrolled in the main hospitals of the reference centers after histological confirmation of MPM diagnosis. Controls were randomly selected or from the local population (Casale Monferrato study) [279, 280], or among subjects not affected by neoplastic or respiratory conditions admitted to general medicine or urology units at the reference hospital (Turin study) [280]. Subjects were selected according to the following criteria: i) availability of good quality DNA at the time of the analyzes, ii) availability of quantitative assessment of asbestos exposure [281], and iii) asbestos exposure above the background level (as defined in [281]). Additional 13 MPM cases recruited in the main hospitals of the municipalities of Novara (Piedmont, Italy) (n=4), and Alessandria (Piedmont, Italy) (n=9) who complied with the above criteria were included in this study as well, and, for the purpose of descriptive statistics (**Table 23**), they were merged with Casale Monferrato cases.

For all subjects, occupational history and lifestyle habits information was collected through interviewer-administered questionnaires filled out at enrollment during a face-to-face interview. An occupational epidemiologist assessed intensity and duration of asbestos exposure and cumulative exposure for all the subjects, taking into account their occupational and residential history and other possible circumstances of asbestos exposure, like sharing home with an asbestos worker [281].

In total, 300 samples, 163 MPM cases and 137 non-MPM controls underwent DNAm analysis. Subjects were split in two analytical sets: a Training set (82 MPM cases and 68 healthy controls), used to primarily

identify DNAm differences between cases and controls, and a Test set (81 cases and 69 controls), used to confirm the discovery findings.

DNA-methylation analyzes

Illumina Human450K Methylation assay

Five-hundred nanograms of genomic DNA for each sample were bisulfite treated (EZ-96 DNA Methylation-Gold Kit, Zymo Research Corporation, Irvine, CA, USA) to convert unmethylated cytosine to uracil. In downstream analyzes the methylation status of a CpG site can thus be assessed by distinguishing between a C or T residue at the same position in the bisulfite treated DNA [282].

The Infinium HumanMethylation450 BeadChip (Illumina Inc., San Diego, CA, USA) was used to measure the methylation level of more than 485,000 individual CpG loci at a genome-wide resolution [283].

Cases and controls were randomly and blindly distributed across conversion plates and methylation BeadChips. Twelve samples were analyzed on each BeadChip. As a "position effect" was reported for Illumina Methylation BeadChips, each sample position on the BeadChip was completely random as well. We further verified the randomization of the position on each BeadChip was effective by checking for a position effect, and we found no occurrence of position effect (data not shown).

BeadChips were processed according to manufacturer protocols. Briefly, the enzymatically fragmented bisulfite converted DNA is hybridized on-chip to locus specific DNA oligomers. A subsequent single base extension reaction with Biotin- or DNP-labeled ddNTPs allows the discrimination between methylated (unconverted) or unmethylated (converted) cytosines at each CpG site. After fluorescent staining, the BeadChips were imaged with a dedicated scanner and fluorescent signals recorded. Data were inspected with the dedicated GenomeStudio software v2011.1 with Methylation module 1.9.0 (llumina Inc., San Diego, CA),and quality checked according to the built in quality controls, i.e. more than 200 control probes specifically designed to assess technical aspects of the bisulfite conversion process and of the BeadChip assay itself, including Bisulfite Conversion, Normalization (system background), Staining efficiency,

Extension efficiency, Hybridization efficiency, Target Removal (stripping after extension reaction), Specificity.

DNA methylation data preprocessing and quality control

Raw data were analyzed with the R package *methylumi.*

The average methylation value at each locus, or average "Beta-value", was computed as the ratio of the intensity of the methylated signal over the total signal (unmethylated + methylated) [284, 285]: Beta-values represent the percentage of methylation at each individual CpG locus, ranging from 0, no methylation, to 1, full methylation.

We excluded from the analyzes: i) single Beta-values with detection p-value≥0.01; ii) CpG loci with missing Beta-values in more than 20% of the assayed samples; iii) CpG loci detected by probes containing SNPs with MAF≥0.05 in the CEPH (Utah residents with ancestry from northern and western Europe, CEU) population; iv) samples with a global call rate≤95%. CpGs on chromosomes X and Y were excluded from the analysis. We further excluded from the analyzes 50,186 CpG according to Zhou *et al* [286].

Gene-expression analyzes

Illumina HumanHT-12 Gene Expression assay

Biotinilated cRNA was obtained from 300ng total RNA using the Illumina TotalPrep RNA Amplification Kit (Thermo Fisher Scientific Inc). Seven hundred and fifty nanograms of purified cRNA were hybridized to HumanHT-12 V4 BeadChip™ arrays (Illumina, San Diego, CA, USA) and streptavidin stained.

Gene expression levels of more than 47,000 transcripts targeting about 31,000 annotated genes were visualized using the GenomeStudio software v2011.1 with Gene Expression module 1.9.0 (llumina Inc., San Diego, CA). The raw data were quality checked according to the built in quality controls for hybridization, labeling, staining, background signal, expression level of housekeeping genes.

Batch effect and technical variability correction

To account for methylation assay variability and batch effects, we corrected all differential methylation analyzes for “control probes” Principal Components (PCs). Using PCs assessed by principal a component analysis (PCA) of the BeadChip’s built-in control probes as a correction factor for statistical analyzes of microarray data is a method that allows to take into account the technical variability of several steps in the DNA-methylation analysis, from the bisulfite conversion to BeadChip processing [287].

The first 10 PCs were thus included into the differential methylation analyzes to correct for technical variability and batch effect.

Population stratification correction

Population stratification and geographic origins of subjects may influence DNA methylation profiles [288]. To take into account this source of potential bias, we took advantage of the whole genome genotyping dataset from the same subjects [222]. The first PCs calculated on the basis of genome-wide genotyping were shown to correlate with different geographic origins of people [289].

The first 2 PCs were included into the differential methylation analyzes to correct for potential population stratification.

Statistical analyzes

All statistical analyzes were conducted using the open source software Rv3.4.0 [290].

Study subjects were randomly assigned to either a Training set or a Test set using the R function *split*, with the option of maintaining comparable proportions of cases and controls in the two groups. The two sets were compared to check for differences in the distribution of sex, age, center, exposure, WBCs estimated percentages, and the same comparison was done for descriptive purposes within each one of the two sets (**Table 23**): for comparisons between groups, we used the Wilcoxon Rank Sum test (W test) for continuous variables and the Chi-square test for categorical variables. All tests were two-tailed.

Epigenome-wide differential methylation between cases and controls was tested by regression models (see below) both at single CpG and at "regional" level. For all the regression analyzes, continuous variables such as age (years), and asbestos exposure doses (fibers/mL-years), were Rank transformed (R function *rntransform*) to remove skewness.

Since dysregulation of immunocompetent cells triggered by asbestos and possibly linked to MPM ethiopathogenesis was reported in literature, we did not adjust regression analyzes by WBCs composition, that would have removed methylation signals linked to WBC composition/regulation effects. However, we estimated WBC subtype percentages from genome-wide methylation data [291] for each subject to evaluate differences in WBCs profiles between cases and controls.

Potential population stratification was taken into account by including, in all the regression analyzes, the first two principal components (PCs) calculated on the basis of our previous genome-wide genotyping study [222]; batch and technical effects correction was performed as well adjusting for first ten "BeadChip control probes" PCs. All these variables were Rank transformed (R function *rntransform*) before including them as covariates into regression analyzes. For multiple comparisons tests, a Benjamini-Hochberg false discovery rate p-value (p_{fdr})≤0.05 was considered statistically significant.

Single CpG analysis

Differential methylation (DMe) at Single-CpGs between MPM cases and controls in the Training set was tested by generalized linear models analysis adjusted for asbestos exposure, gender, age, population stratification, and technical variability, as described above.

To ensure a power of the study greater than 90% (two-tailed test, 0.05 alpha error), only CpGs with regression effect sizes differences between cases and controls ≥|0.1|, that corresponds to a difference in methylation Beta-value (delta Beta-value) of about 0.1, were considered in the discovery phase on the Training set, and subsequently investigated in the Test set under the same analytical conditions. CpGs with effect sizes differences between cases and controls ≥|0.1| and p_{fdr}≤0.05 in both Training and Test sets were considered as replicating, and underwent gene set enrichment analysis (WEB-based GeneSeTAnaLysis

Toolkit, WebGestalt; *http://bioinfo.vanderbilt.edu/webgestalt*) to identify pathways potentially affected by MPM-related methylation changes.

Additionally, Test set subjects were clustered (unsupervised clustering analysis, R package *Rgplots*) according to their methylation levels of the differentially methylated single-CpGs (DM-CpGs) identified at the single CpGs analysis. For this analysis, samples with missing methylation Beta-values for any of the CpGs were removed.

Finally, in the Test set, we compared the area under the receiver operating characteristic curve (AUC) of two models by the DeLong test : Model 1 included age, sex and asbestos exposure; Model 2 was as Model 1 plus methylation levels of significant and replicating DM-CpGs: this was done to test the potential improvement in case-control discrimination when considering DMe information together with asbestos exposure information.

Regional differential methylation analysis

We used the A-clustering algorithm to define genomic regions characterized by a group of two or more neighboring CpGs (CpGs clusters) with correlated methylation levels . In the Training set, we tested the occurrence of differentially methylated regions (DMRs) between cases and controls by generalized estimating equations with the same adjustments as the single-CpG analysis, under the hypothesis that methylation changes in genomic regions may underlie potential functional changes linked to MPM. DMRs with effect size differences of at least 0.1 between cases and controls were investigated for occurrence of DMe in the Test set under the same analytical conditions. DMRs replicating in the Test set (i.e. effect sizes $\geq|0.1|$ and $p_{fdr}\leq 0.05$ in both Training and Test sets) underwent gene set enrichment analysis.

CpGs methylation vs gene expression

Our study envisaged the collection of DNA only, thus, due to unavailability of RNA, we were unable to assess gene expression levels for our mesothelioma cases and related controls. In order to assess a possible relationship between differentially methylated CpGs and gene expression variations of the corresponding

genes, we took advantage of other in-house studies for which whole transcriptome gene expression and DNAm data were available. Data from two sample sets were used: i) 72 healthy subjects included in a genotype-phenotype correlation study (from here on, Geno-Pheno) described elsewhere [141] for which methylation and gene expression were assessed on DNA and RNA from the same aliquot of peripheral blood mononuclear cells (PBMCs), and ii) 21 normal pleura surgical samples from donors that were subjected to thoracoscopy for conditions different from MPM, belonging to a previous study [280], for which whole genome methylation and whole transcriptome expression were recently assessed as a pilot study (data not shown). In the two studies, methylation levels were determined by the HumanMethylation450 BeadChip and quality controlled as described previously. Gene expression levels were determined by the Illumina HumanHT-12 v4 Expression BeadChip. Genes were considered expressed if their probe(s) intensity signal(s) was(were) at least twice the BeadChip background. The relationship between methylation and transcript levels of genes was investigated with a Pearson's correlation test.

8.1.2. Results

Descriptive statistics of the sample are reported in **Table 23**.

Training set and Test set did not differ by age, exposure, estimated WBCs percentages (W test), sex, center of recruitment, case-control distribution (Chi-square Test), except for a difference in B cells distribution in controls between Training set and Test set (W test, P=0.04).

Within each of the 2 sets, cases were exposed to significantly higher cumulative doses of asbestos than controls, and had statistically significant higher estimated levels of granulocytes and lower estimated levels of CD8T, CD4T, NK, B cells, whereas monocytes levels were significantly higher in cases only in the Test set (**Table 23**).

After quality controls on raw DNAm data, 389,147 CpG were used in the following analyzes.

Table 23 | **Sample characteristics and descriptive**

	TRAINING SET			TEST SET		
	Cases (N=82)	**Controls (N=68)**	***P*** [a]	**Cases (N=81)**	**Controls (N=69)**	***P*** [a]
Casale M. [N (%)]	51 (62.20)	51 (75.00)	ns	49 (60.49)	43 (62.32)	ns
Torino [N (%)]	31 (37.80)	17 (25.00)		32 (39.51)	26 (37.68)	
Males [N (%)]	60 (73.17)	43 (63.24)	ns	53 (65.43)	57 (82.61)	**0.03**
Females [N (%)]	22 (26.83)	25 (36.76)		28 (34.57)	12 (17.39)	
Mean Age [years, mean±sd]	68.46±11.41	65.16±10.07	**0.02**	66.71±10.67	64.02±10.00	ns
Exposure [mean±sd] [b]	34.8±165.39	5.11±12.46	**2.4×10^{-6}**	62.98±194.51	2.13±3.43	**1.5×10^{-8}**
CD8T % [mean±sd] [c]	0.03±0.05	0.06±0.05	**6.3×10^{-5}**	0.03±0.03	0.07±0.04	**4.0×10^{-10}**
CD4T % [mean±sd]	0.07±0.05	0.15±0.07	**2.2×10^{-10}**	0.08±0.05	0.14±0.06	**2.3×10^{-7}**
NK % [mean±sd]	0.05±0.05	0.08±0.05	**1.3×10^{-3}**	0.06±0.04	0.07±0.05	**0.04**
B cells % [mean±sd]	0.06±0.02	0.09±0.03	**4.1×10^{-10}**	0.06±0.03	0.08±0.03	**5.0×10^{-5}**
Monocytes % [mean±sd]	0.07±0.04	0.07±0.04	ns	0.08±0.05	0.07±0.03	**0.01**
Granulocytes % [mean±sd]	0.74±0.13	0.61±0.11	**2.0×10^{-10}**	0.73±0.11	0.62±0.11	**2.6×10^{-8}**

NOTE: Abbreviations: ns = non statistically significant (*P*>0.05)

[a] Chi-square test (two-sided, 2X2 contingency table) for Centre and Sex; Wilcoxon Rank Sum test (two-sided, Cases *vs* Controls) for Age, Exposure, WBC estimated percentages

[b] fibers/mL years

[c] Estimated percentages from methylation data (ref [291]) for all WBCs

Single-CpGs differential methylation analysis

In the Training set, 887 CpGs had delta Beta-values≥0.1 between cases and controls (i.e. effect size ≥|0.10|), 884 of them with p_{fdr}<0.05. The vast majority (n=871) were hypomethylated in cases while only 13 showed hypermethylation. Out of the 884 significantly differentially methylated CpGs identified in the Training set, 868 were significantly differentially methylated (p_{fdr} <0.05) also in the Test set with concordant and similar effect sizes: among hypomethylated CpGs (n=855), effect sizes ranged from -0.17 to -0.04 (mean±sd=-0.10±0.02); among the hypermethylated (n=13) effect sizes ranged from 0.06 to 0.10 (mean±sd=0.08±0.01). The 868 differentially methylated CpGs were annotated to 599 genes, that were analyzed for pathway enrichment. We found statistically significant enrichment for Neutrophil Degranulation, Innate Immune System and Immune System.

Appraised the huge number of hypomethylated signals (n=871) identified in the Training set, we focused on the more differentially hypomethylated posing a more stringent threshold at effect size ≤-0.15, obtaining a smaller group of 20 hypomethylated CpGs. The resulting DM-CpGs list (n=33) in the Training set, thus, included 20 hypomethylated CpGs with effect size ≤-0.15 and 13 hypermethylated with effect size ≥0.10. Checking for strictly replicating signals in the Test set, we ended up with 7 "top DM-CpGs", 3 hypomethylated CpGs with effect sizes ≤-0.15 in both Training and Test sets (genes *FOXK1*, *MYB*, and *TAF4*) and 4 CpGs hypermethylated with effect sizes ≥0.10 in both Training and Test sets (genes *CXCR6/FYCO1*, *TAP1*, *MORC2*, *LIME1*) (**Table 24**).

We then performed two unsupervised clustering analyzes on Test set samples to inspect the distribution of cases and controls according to their methylation levels at the DM-CpGs identified in the Training set: the first one (**Figure 16A**) included the top DM-CpGs, irrespective of strict replication criteria, i.e. all the 33 top DM-CpGs, identified in the Training set; the second one (**Figure 16B**) included only the 7 DM-CpGs whose effect sizes strictly replicated in the Test set, as described above. Cutting the sample dendrograms at the first node, in both analyzes the groups with the most hypomethylated/hypermethylated DM-CpGs included the highest proportion of cases as compared to the other group (**Figure 16A** and **B**).

The AUC comparison of 2 models showed a statistically significant improvement in discrimination between cases and controls when including methylation levels into the analysis (**Figure 17**).

Table 24 | **Differentially methylated CpGs in the Training set and Test set.** List of the top CpGs hypomethylated with effect size ≤-0.15 and hypermethylated with effect size ≥0.10 in both Training set and Test set (cases *vs* controls)

					TRAINING-SET					TEST-SET					
					Beta-values					Beta-values					
Probe ID	**Chr**	**Map position (GRCh37/hg19)**	**Gene Symbol**	**UCSC Refgene Group**	**Cases** Mean±sd	**Controls** Mean±sd	**Eff. size±se**	**p**	**p fdr EWAS**	**Cases** Mean±sd	**Controls** Mean±sd	**Eff. size±se**	**p**	**p fdr EWAS**	***P* fdr TARGETED**
CpGs hypomethylated in MPM cases															
cg04572930	7	4754834	*FOXK1*	Body	0.35±0.17	0.54±0.16	-0.18±0.03	5.19E-08	2.46E-05	0.37±0.19	0.51±0.16	-0.17±0.03	1.19E-06	1.75E-04	2.93E-06
cg04739200	6	135517046	*MYB*	Body	0.32±0.14	0.48±0.15	-0.16±0.03	5.94E-08	2.64E-05	0.30±0.14	0.48±0.17	-0.16±0.03	5.56E-07	1.05E-04	1.62E-06
cg01521397	20	60590872	*TAF4*	Body	0.33±0.16	0.50±0.16	-0.16±0.03	5.18E-07	8.58E-05	0.35±0.15	0.51±0.14	-0.15±0.03	2.09E-07	5.85E-05	8.33E-07
CpGs hypermethylated in MPM cases															
cg08450017	3	45984838	*CXCR; FYCO1*	TSS200 Body	0.76±0.11	0.63±0.13	0.13±0.02	1.04E-08	1.26E-05	0.73±0.13	0.63±0.11	0.10±0.02	1.39E-05	9.69E-04	2.31E-05
cg26033526	6	32819858	*TAP1*	Body	0.68±0.11	0.56±0.10	0.11±0.02	5.21E-08	2.46E-05	0.67±0.11	0.57±0.09	0.10±0.02	1.46E-06	2.00E-04	3.48E-06
cg23825480	22	31336785	*MORC2*	Body	0.78±0.10	0.66±0.10	0.10±0.02	1.28E-08	1.38E-05	0.76±0.11	0.65±0.13	0.10±0.02	2.20E-05	1.34E-03	3.43E-05
cg00446123	20	62367888	*LIME1*	TSS200	0.76±0.11	0.64±0.11	0.10±0.02	2.13E-07	5.20E-05	0.73±0.11	0.63±0.11	0.10±0.02	2.64E-06	3.03E-04	5.66E-06

Figure 16 | **Unsupervised clustering analyzes on Test set samples to inspect the distribution of cases and controls according to their methylation levels at the DM CpGs identified in the Training set**. Cluster A1, 12 cases 37 controls. Cluster A2, 41 cases 19 controls. Cluster B1, 33 cases, 6 controls. Cluster B2, 34 cases, 61 controls

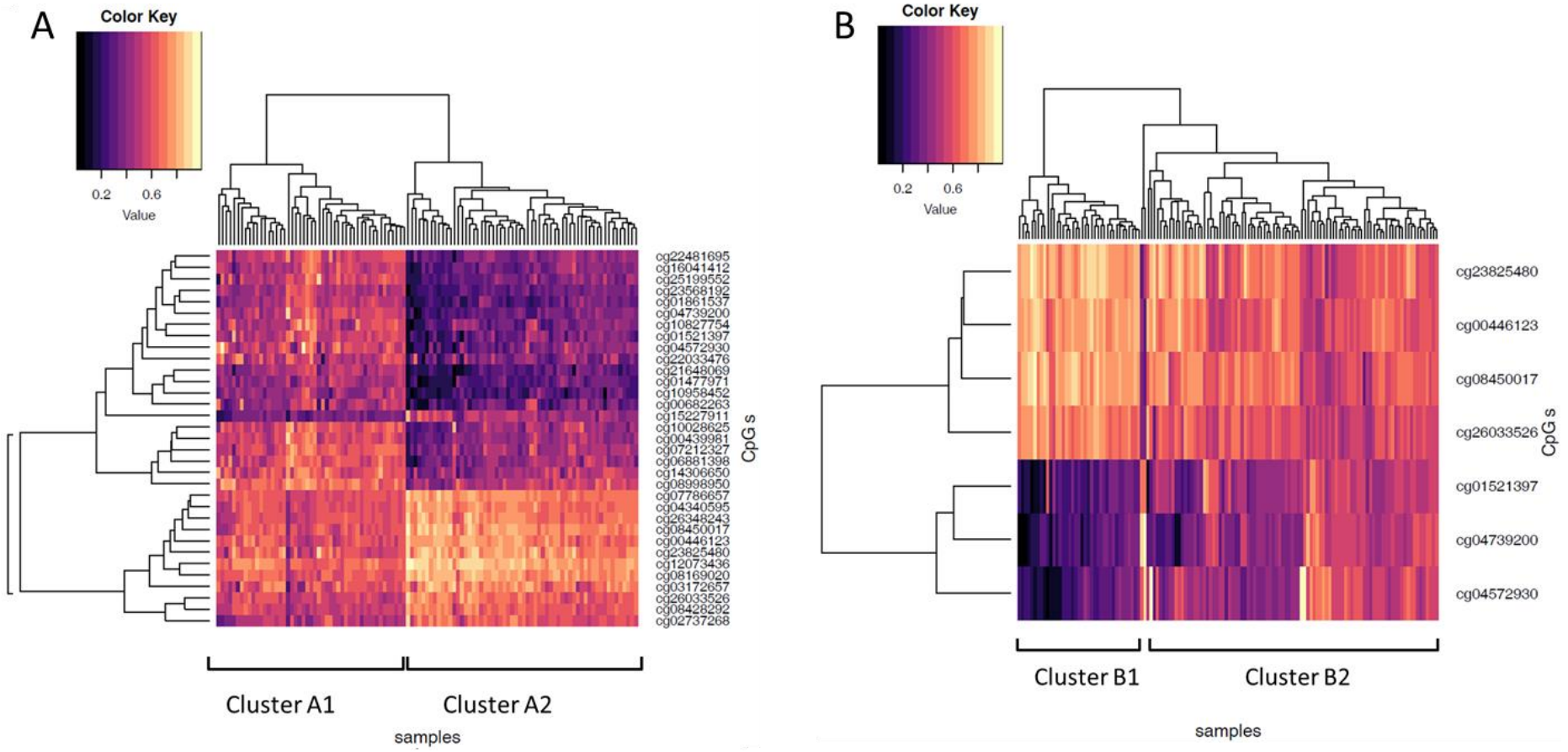
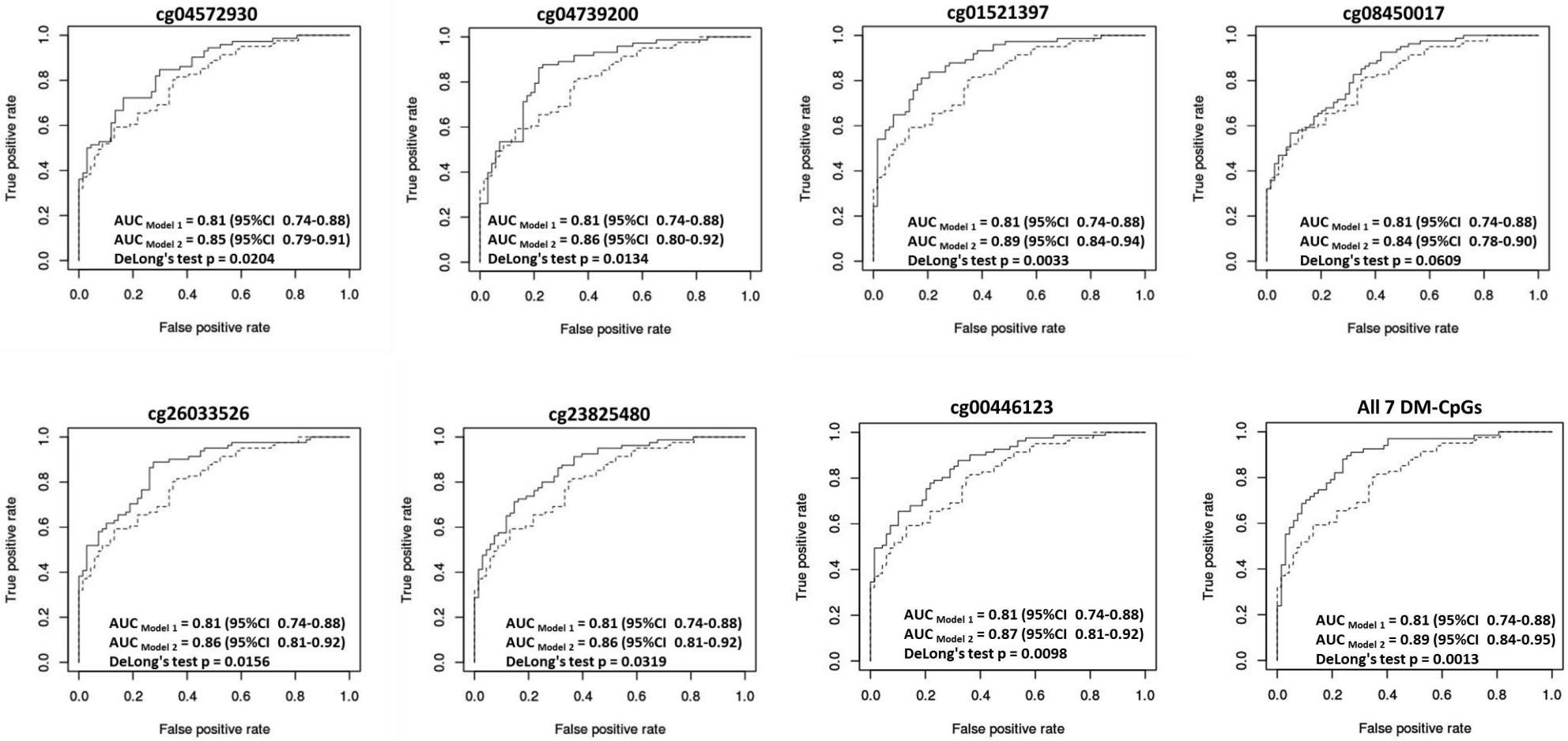


Figure 17 | **Receiver Operating Characteristic (ROC) curves on Test set**. Model 1 (dotted line): age, sex, asbestos exposure levels Model 2 (solid line): age, sex, asbestos exposure levels, CpGs methylation levels



Regional differential methylation analysis

The A-clustering algorithm identified 24,573 CpGs clusters with correlated methylation levels in the Training set and 23,676 in the Test set, that were tested for differential methylation between cases and controls. As a first step, we checked for clusters of differentially methylated CpGs around the 7 "top DM-CpGs" identified at the single-CpGs analysis, and found that 5 out of the 7 top DM-CpGs identified at the single CpGs analysis were indeed included in CpGs clusters differentially methylated between cases and controls (**Table 25**).

As a second step, across the 24,573 CpGs clusters identified in the Training set, we looked for DMRs between cases and controls with delta effect size ≥|0.10|, and found 35 DMRs hypomethylated (p_{fdr}<0.05) in Training set cases. Among the 35 regions, the broadest cluster included 9 CpGs in *FOXK1* gene around cg04572930, that was the most hypomethylated CpG identified at single-CpG analysis. No region hypermethylated above the established threshold (effect size ≥|0.10|) in cases with respect to controls was identified in the Training set.

All the 35 regions identified in the Training set were represented as clusters also in the Test set, and all of them showed statistically significant hypomethylation in cases (range[-0.12;-0.07], median=-0.10, p_{fdr}<0.05) (**Table 26**). Restricting the analysis to clusters with effect size ≤-0.10 and p_{fdr}<0.05 in both Training and Test sets (i.e. strictly replicating DMRs), 21 regions corresponding to 21 genes were identified. The 5 regions with more than 2 CpGs in cluster were noted as "top DMRs", and corresponded to the following genes: *FOXK1* (9 CpGs), *CSTA* (4 CpGs), *ZNF516* (3 CpGs), *TOLLIP* (3 CpGs), *TNFAIP6* (3 CpGs).

At a gene-set analysis on the 21 DMRs, no enriched category reached statistical significance. However, considering enriched categories that included more than 1 gene, we found the most represented pathways were related to immune systems processes (3 categories, 5 genes in multiple categories), cancer related pathways (5 categories, 7 genes in multiple categories), and developmental biology (1 category, 5 genes).

Table 25 | **DMRs around DM CpGs in the Training set and Test set**

CpG sites in cluster	Sites in Cluster	Map position 1st CpG in cluster (GRCh37/hg19)	Gene Symbol	UCSC Refgene Group	TRAINING-SET					TEST-SET				
					Average Beta-values		Eff. size±se	p	p_{fdr}	Average Beta-values		Eff. size±se	p	p_{fdr}
					Cases Mean±sd	Controls Mean±sd				Cases Mean±sd	Controls Mean±sd			
cg26548834; cg04261496; cg01509853; *cg26177213*; cg26136772; ***cg04572930***; cg15200418; *cg18276112*; *cg25344401*	9	Chr7:4752951	*FOXK1*	Body	0.41±0.18	0.52±0.16	-0.10±0.02	6.90E-11	2.17E-08	0.42±0.18	0.52±0.15	-0.10±0.01	7.24E-12	4.63E-09
cg26288715; cg27434890; **cg04739200**; cg18835596*	4	Chr6:135506834	*MYB*	Body	0.39±0.18	0.48±0.16	-0.09±0.02	2.20E-08	7.98E-07	0.31±0.15	0.44±0.16	-0.12±0.02	9.99E-10	7.53E-08
cg05705212; **cg08450017**; cg25226014; cg01178899	4	Chr3:45984743	*CXCR6* *FYCO1*	TSS1500 Body	0.87±0.10	0.80±0.13	0.06±0.01	6.15E-11	2.02E-08	0.86±0.11	0.80±0.13	0.05±0.01	7.15E-08	1.81E-06
cg23228341[a]; ***cg26033526***; cg01673307; cg24111025; cg02181920; cg06473288; cg17626301; cg26234900; cg10666909; cg08818207	10	Chr6:32818477	*TAP1*	Body	0.79±0.13	0.74±0.13	0.06±0.01	2.50E-08	8.75E-07	0.78±0.13	0.72±0.13	0.06±0.01	2.04E-08	6.85E-07
cg14977069; cg06653796; cg21201401; **cg00446123**; cg20513976; cg24631526; cg14396214; cg12413156	8	Chr20:62367698	*LIME1*	TSS1500	0.52±0.23	0.45±0.20	0.07±0.01	2.68E-09	1.89E-07	0.49±0.22	0.44±0.20	0.05±0.01	2.47E-05	2.10E-04

In bold the index CpG (i.e. among the 7 top DM-CpGs)
In italic underlined CpGs identified also at single-CpGs analysis (i.e. among the 868 DM-CpGs)
[a]this CpG is member of the cluster only in Training set

Table 26 | **Differentially methylated CpGs clusters (A-clustering regional analysis) in the Training set and Test set**

CpG sites in cluster	Nr. of Sites in Cluster	Map position of 1st CpG in cluster (GRCh37/hg19)	Gene Symbol	TRAINING-SET					TEST-SET				
				Average Cluster's Beta-values		Eff. size±se	p	p_{fdr}	Average Cluster's Beta-values		Eff. size±se	p	p_{fdr}
				Cases Mean±sd	Controls Mean±sd				Cases Mean±sd	Controls Mean±sd			
cg26548834*; cg04261496; cg01509853; *cg26177213*; cg26136772; *cg04572930*; cg15200418; *cg18276112*; *cg25344401	**9**	**Chr7:4752951**	***FOXK1***	**0.41±0.18**	**0.52±0.16**	**-0.10±0.02**	**6.90E-11**	**2.17E-08**	**0.42±0.18**	**0.52±0.15**	**-0.10±0.01**	**7.24E-12**	**4.63E-09**
cg26783127; *cg16732654; cg23334433*; *cg13618516*	4	Chr17:79128918	*AATK*	0.35±0.13	0.48±0.12	-0.12±0.02	2.00E-10	3.90E-08	0.37±0.14	0.46±0.11	-0.09±0.02	2.64E-06	3.31E-05
cg26928972*; *cg18618429*; *cg14664412*; *cg21932814	**4**	**Chr3:122043799**	***CSTA***	**0.29±0.12**	**0.41±0.12**	**-0.11±0.02**	**8.41E-09**	**4.12E-07**	**0.30±0.12**	**0.41±0.12**	**-0.10±0.02**	**9.04E-09**	**3.65E-07**
cg07047815***;** *cg22136363*; *cg18723409*; cg17566541	4	Chr11:1909656	*LSP1*	0.29±0.15	0.38±0.12	-0.11±0.01	4.77E-15	2.93E-11	0.30±0.15	0.37±0.12	-0.07±0.01	3.28E-07	5.97E-06
cg22858170***;** *cg00907842*; cg27606341; cg13703437	4	Chr5:39204386	*FYB*	0.39±0.25	0.47±0.20	-0.10±0.02	6.76E-11	2.16E-08	0.39±0.24	0.47±0.19	-0.07±0.01	1.25E-11	5.61E-09
cg20429104*; *cg12019801*; *cg00471371	**3**	**Chr18:74114570**	***ZNF516***	**0.31±0.12**	**0.43±0.12**	**-0.11±0.02**	**1.43E-08**	**5.94E-07**	**0.32±0.12**	**0.43±0.11**	**-0.10±0.02**	**4.15E-08**	**1.16E-06**
***cg23240927*; *cg07546654*; cg09566995**	**3**	**Chr11:1320497**	***TOLLIP***	**0.30±0.13**	**0.42±0.13**	**-0.11±0.02**	**1.20E-10**	**3.09E-08**	**0.29±0.11**	**0.41±0.11**	**-0.11±0.02**	**3.85E-12**	**2.68E-09**
cg01942558*; *cg09727050*; *cg03406844	**3**	**Chr2:152214027**	***TNFAIP6***	**0.35±0.14**	**0.47±0.13**	**-0.11±0.02**	**2.70E-07**	**5.37E-06**	**0.35±0.13**	**0.47±0.12**	**-0.12±0.02**	**1.71E-10**	**2.39E-08**
cg08253808; *cg05398700*; cg14315912	3	Chr14:102676957	*WDR20*	0.43±0.13	0.54±0.12	-0.11±0.02	3.96E-09	2.42E-07	0.44±0.13	0.53±0.11	-0.08±0.02	7.08E-08	1.79E-06
cg08359464*; *cg12405599	**2**	**Chr3:128370361**	***RPN1***	**0.36±0.14**	**0.50±0.14**	**-0.13±0.02**	**4.92E-10**	**6.47E-08**	**0.39±0.13**	**0.49±0.12**	**-0.11±0.02**	**1.14E-07**	**2.57E-06**
cg04319611*; *cg14016236	**2**	**Chr3:171784422**	***FNDC3B***	**0.25±0.12**	**0.38±0.14**	**-0.13±0.02**	**3.27E-09**	**2.15E-07**	**0.25±0.11**	**0.37±0.11**	**-0.11±0.02**	**1.74E-10**	**2.40E-08**
cg19069360*; *cg02929855	**2**	**Chr12:1922058**	***CACNA2D4***	**0.31±0.11**	**0.45±0.12**	**-0.12±0.02**	**9.40E-10**	**9.88E-08**	**0.32±0.11**	**0.44±0.10**	**-0.11±0.02**	**8.77E-11**	**1.58E-08**
cg01462353*; *cg05624376	**2**	**Chr2:169939873**	***DHRS9***	**0.30±0.11**	**0.43±0.11**	**-0.12±0.02**	**4.01E-11**	**1.47E-08**	**0.30±0.1**	**0.42±0.11**	**-0.11±0.02**	**9.69E-11**	**1.65E-08**
cg22820233*; *cg05252264	**2**	**Chr19:55385581**	***FCAR***	**0.25±0.11**	**0.38±0.12**	**-0.12±0.02**	**1.26E-09**	**1.15E-07**	**0.25±0.1**	**0.37±0.13**	**-0.11±0.02**	**1.63E-11**	**6.12E-09**
cg09694051*; *cg17419815	**2**	**Chr12:27181973**	***MED21***	**0.42±0.15**	**0.55±0.15**	**-0.11±0.03**	**2.59E-05**	**2.19E-04**	**0.43±0.14**	**0.53±0.12**	**-0.10±0.02**	**8.11E-06**	**8.37E-05**

cg24079727*; *cg04858164	**2**	**Chr15:57317980**	***TCF12***	**0.39±0.17**	**0.52±0.17**	**-0.11±0.03**	**1.03E-05**	**1.03E-04**	**0.40±0.17**	**0.51±0.15**	**-0.10±0.02**	**7.16E-08**	**1.81E-06**
***cg10958452*; cg17737314**	**2**	**Chr1:44114346**	***KDM4A***	**0.21±0.11**	**0.33±0.15**	**-0.11±0.02**	**2.70E-08**	**9.24E-07**	**0.23±0.13**	**0.33±0.15**	**-0.10±0.02**	**2.57E-06**	**3.24E-05**
cg24401262; *cg04552418*	**2**	**Chr1:31956405**		**0.27±0.12**	**0.40±0.13**	**-0.11±0.02**	**2.11E-09**	**1.60E-07**	**0.27±0.11**	**0.39±0.1**	**-0.11±0.02**	**3.98E-11**	**1.01E-08**
cg18095675*; *cg19005485	**2**	**Chr3:197272311**	***BDH1***	**0.50±0.14**	**0.61±0.12**	**-0.11±0.02**	**3.50E-06**	**4.25E-05**	**0.52±0.12**	**0.61±0.12**	**-0.10±0.02**	**4.74E-07**	**8.03E-06**
cg14575356; cg20760057	2	Chr6:130013903	*ARHGAP18*	0.38±0.13	0.51±0.13	-0.11±0.02	9.85E-08	2.42E-06	0.40±0.14	0.50±0.12	-0.09±0.02	3.91E-06	4.58E-05
***cg15891546*; cg09859659**	**2**	**Chr8:142179908**	***DENND3***	**0.46±0.12**	**0.58±0.11**	**-0.11±0.02**	**1.53E-10**	**3.47E-08**	**0.47±0.12**	**0.57±0.09**	**-0.10±0.02**	**4.99E-09**	**2.30E-07**
cg25811633*; *cg02520804	**2**	**Chr3:71626654**	***FOXP1***	**0.31±0.13**	**0.43±0.12**	**-0.11±0.02**	**7.90E-10**	**8.63E-08**	**0.31±0.12**	**0.43±0.12**	**-0.10±0.02**	**7.45E-11**	**1.42E-08**
cg16441068; cg14543285	2	Chr14:103112066	*RCOR1*	0.35±0.12	0.47±0.11	-0.11±0.02	1.92E-09	1.53E-07	0.38±0.12	0.46±0.10	-0.09±0.02	2.31E-07	4.49E-06
cg21249659*; *cg21685770	**2**	**Chr12:10324843**	***OLR1***	**0.32±0.11**	**0.44±0.10**	**-0.11±0.02**	**1.97E-10**	**3.87E-08**	**0.33±0.11**	**0.44±0.10**	**-0.10±0.02**	**2.42E-09**	**1.36E-07**
cg02019444*; *cg16452651	**2**	**Chr21:35016787**	***ITSN1***	**0.26±0.12**	**0.38±0.11**	**-0.11±0.02**	**1.87E-08**	**7.11E-07**	**0.26±0.10**	**0.38±0.11**	**-0.10±0.02**	**2.47E-11**	**7.60E-09**
cg15334250; cg27214730	2	Chr3:184297495	*EPHB3*	0.19±0.09	0.30±0.11	-0.11±0.02	4.09E-13	1.05E-09	0.19±0.08	0.29±0.10	-0.09±0.02	6.73E-10	5.84E-08
cg06465011; *cg26709988*	2	Chr16:84860871	*CRISPLD2*	0.31±0.10	0.42±0.09	-0.11±0.02	9.40E-10	9.88E-08	0.33±0.10	0.41±0.09	-0.09±0.02	1.31E-08	4.85E-07
***cg13381110*; cg16978268**	**2**	**Chr18:60646614**	***PHLPP1***	**0.35±0.15**	**0.47±0.17**	**-0.11±0.02**	**1.28E-09**	**1.16E-07**	**0.35±0.17**	**0.47±0.17**	**-0.12±0.02**	**2.44E-12**	**2.62E-09**
cg15361231; cg21815704	2	Chr1:193075191	*GLRX2*	0.36±0.14	0.48±0.13	-0.11±0.02	1.97E-06	2.68E-05	0.36±0.12	0.46±0.13	-0.09±0.02	4.27E-06	4.94E-05
***cg06373940*; cg14306819**	**2**	**Chr2:128052778**	***ERCC3***	**0.35±0.12**	**0.47±0.11**	**-0.11±0.02**	**4.64E-09**	**2.67E-07**	**0.36±0.11**	**0.48±0.10**	**-0.11±0.02**	**1.82E-11**	**6.43E-09**
***cg20357538*; cg08165960**	**2**	**Chr15:101777761**	***CHSY1***	**0.20±0.09**	**0.31±0.10**	**-0.10±0.02**	**4.17E-11**	**1.49E-08**	**0.21±0.09**	**0.31±0.09**	**-0.10±0.01**	**4.46E-11**	**1.06E-08**
cg04118119; cg05699739	2	Chr13:47371987	*ESD*	0.27±0.11	0.38±0.11	-0.10±0.01	4.33E-15	2.93E-11	0.28±0.11	0.36±0.12	-0.08±0.02	1.31E-07	2.89E-06
cg12237948; cg10206397	2	Chr20:2085157	*STK35*	0.45±0.12	0.56±0.12	-0.10±0.02	2.05E-08	7.59E-07	0.46±0.12	0.54±0.11	-0.08±0.02	6.98E-06	7.41E-05
cg17586302; cg24690314	2	Chr6:144013969	*PHACTR2*	0.44±0.15	0.56±0.13	-0.10±0.02	1.53E-08	6.15E-07	0.46±0.16	0.53±0.12	-0.09±0.02	3.39E-07	6.14E-06
cg05604874; *cg00660167*	2	Chr17:80200785		0.37±0.14	0.49±0.13	-0.10±0.02	1.51E-08	6.14E-07	0.40±0.13	0.48±0.13	-0.08±0.02	2.32E-05	2.00E-04

In bold, clusters with effect size ≤-0.10 in both Training and test sets

In italic underlined CpGs identified also at single CpGs analysis

* this CpG is member of the cluster only in test set

CpGs methylation vs Gene expression

The methylation levels of the 7 differentially methylated single-CpGs and of the 21 CpGs from the regional analysis were correlated with the corresponding gene-expression levels in 2 separate datasets: one comprising methylation and gene expression data from healthy subjects (Geno-Pheno) PBMCs', and the other one comprising methylation and gene expression data from healthy pleura (see methods).

In the Geno-Pheno dataset we found no significant correlation between methylation levels and transcript levels in PBMCs for any of the considered CpGs.

In the healthy pleura dataset, cg08450017 in *FYCO1*, and 8 out of 9 CpGs in the *FOXK1* cluster showed a statistical significant correlation between methylation levels and the corresponding gene transcript levels (**Table 27**).

Table 27 | **Correlation between DNA methylation and Gene Expression in the healthy pleura dataset**

Probe ID	Chromosome position (GRCh37/hg19)	UCSC Refgene Group	CpG identified in Analysis	Correlation rho	Correlation *P*	Average Beta value Mean±sd	Average Expr. value Mean±sd
FOXK1							
cg26548834	Chr7: 4752951	Body	Single-CpG/A-clustering	0.50	**0.02**	0.59±0.10	7.88±0.14
cg04261496	Chr7: 4753002	Body	A-clustering	0.43	**0.05**	0.50±0.10	
cg01509853	Chr7: 4754502	Body	A-clustering	0.51	**0.02**	0.87±0.04	
cg26177213	Chr7: 4754566	Body	Single-CpG/A-clustering	0.43	**0.05**	0.68±0.10	
cg26136772	Chr7: 4754681	Body	A-clustering	0.50	**0.02**	0.81±0.08	
cg04572930	Chr7: 4754834	Body	Single-CpG/A-clustering	0.50	**0.02**	0.73±0.14	
cg15200418	Chr7: 4755010	Body	A-clustering	0.47	**0.03**	0.56±0.09	
cg18276112	Chr7: 4755032	Body	Single-CpG/A-clustering	0.35	0.12	0.56±0.09	
cg25344401	Chr7: 4755415	Body	Single-CpG/A-clustering	0.58	**0.01**	0.75±0.11	
FYCO1							
cg08450017	Chr3: 45984838	Body	Single-CpG	0.43	**0.05**	0.81±0.08	8.22±0.20

In bold significant p-values (P≤0.05)

8.1.3. Discussion

Altered DNA methylation is frequently observed in cancer, and DNAm profiles of specific genes were already proposed as potential tools for cancer detection, risk prediction, and prognosis [274]. Concerning pleural mesothelioma, through methylation profiling of tumor biopsies and lung or pleura tissues Christensen *et al*. [292] were able to distinguish between MPM, adenocarcinoma and/or non-cancer lung, suggesting DNAm profiles as diagnostic markers. In our study we investigated WBCs DNAm profiles in a large series of MPM cases and controls, all of them with quantitative assessment of asbestos exposure, with the aim to identify molecular hallmarks of MPM in non-invasively collectible blood samples from asbestos-exposed subjects. Considering previous reports of asbestos related antigenic overstimulation and functional changes in WBCs putatively linked to MPM occurrence, we hypothesized this might be reflected in methylation changes in blood cells. Under this hypothesis, to retain methylation signals potentially linked to immune system deregulation that would otherwise be swept away, we did not include WBC estimated counts as covariate in our analyzes.

We identified signals of differential DNAm in WBCs from MPM cases as compared with controls by both single-CpG and regional analysis, and the findings were replicated in a second set of subjects sampled from the same area. The strongest DM-CpG hypomethylated signal in MPM cases is located in *FOXK1* (Forkhead Box K1), a transcription factor involved in development and metabolism, and Wnt signaling [293]. We found *FOXK1* differentially methylated also at regional analysis, with a 9 CpGs cluster hypomethylated in cases. The involvement of FOXK1 in tumor onset and progression was previously reported for colorectal and gastric cancers [294, 295]. Notably, *FOXK1* is located on chromosome 7p22.2, one of the 5 genomic regions that we previously identified as associated to MPM in a GWAS on Italian samples and that were further replicated in an independent Australian MPM sample [222]. Moreover, FOXK1 directly interacts with BAP1 (BRCA1-associated protein 1) [296], whose involvement in MPM has been already described [225, 297, 298].

Interestingly enough, the most hypermethylated single-CpG signal is located in *FYCO1* (FYVE And Coiled-Coil Domain Containing 1) gene, that encodes for a putative FOXK1 interactor within the human autophagy

network [299]. In the present study, CpGs in *FYCO1* and *FOXK1* are the only signals whose methylation levels showed significant positive correlation with gene expression levels, as measured in a set of healthy pleura samples.

Out of the 868 DM-CpGs (599 genes) and the 21 DMRs (21 genes), we found several signals in genes belonging to "Neutrophil Degranulation", "Innate Immune System", and "Immune System" Reactome pathways. This is not surprising, since we are studying DNAm in WBCs with the purpose to identify relevant changes potentially associated with processes leading to MPM onset.

Three out of the 5 hypermethylated DM-CpG have a role in the immune system, i.e. *TAP1* (Transporter 1, ATP Binding Cassette Subfamily B Member), *LIME1* (Lck Interacting Transmembrane Adaptor 1) and *CXCR6* (C-X-C Motif Chemokine Receptor 6), as well as 3 of the top DMRs, i.e. *CSTA* (Cystatin A), *TOLLIP* (Toll Interacting Protein) and *TNFAIP6* (TNF Alpha Induced Protein 6).

TAP1 is member of ATP-binding cassette transporters superfamily, involved in the transport of antigens from the cytoplasm to the endoplasmic reticulum followed by binding to MHC class I molecules. This process is frequently defective in cancer cells, and low to undetectable levels of TAP1 mRNA and/or protein have been reported in primary cells and cell lines from several tumors (reviewed in [300]).

LIME1 is a transmembrane adaptor protein involved in coupling immunoreceptors to intracellular signaling pathways, with regulatory functions in T cells, where its expression is up-regulated upon TCR stimulation [301].

CXCR6, a T cell chemokine receptor, and its ligand, the chemokine CXCL16, were recently described as markers and promoters of inflammation-associated cancers, such as prostate cancer, and were suggested as mediators of inflammation-related tumorigenesis through direct effects on cancer cell growth and by inducing the migration and proliferation of tumor-associated leukocytes [302]. The same authors reported that the expression of CXCR6/CXCL16 in both cancer cells and adjacent T cells correlated with prostate cancer progression.

CSTA is a cysteine protease inhibitor with a role in cell-cell adhesion, that was found deregulated in several cancers (reviewed in [303]).

TOLLIP (Toll Interacting Protein) and TNFAIP6 (TNF Alpha Induced Protein 6) are both involved in inflammation-related processes. TOLLIP, a ubiquitin-binding protein that interacts with several Toll-like receptor (TLR) signaling cascade components, has a role in inflammatory signaling and is involved in interleukin-1 receptor trafficking and in the turnover of IL1R-associated kinase [304].

TNFAIP6 has a role in the protease network associated with inflammation, as well as in extra cellular matrix stability and cell migration. Interestingly, TNFAIP6 was found upregulated in a matrix-free mesothelioma spheroid model [305] and in primary cultured human mesothelial cells exposed to asbestos [306].

Quite unexpectedly, since methylation was measured in DNA from WBCs, we also found several signals in genes involved in transcriptional regulation and epithelial–mesenchymal transition (EMT).

FOXK1 was reported to induce EMT in gastric cancer cells, and to play a crucial role in the metastatic progression of human gastric cancer [295], whereas MYB (MYB Proto-Oncogene) protein is required for *in vitro* expression of EMT-associated markers, invasion, and anchorage-independent growth of breast cancer cells [307]. Also CXCR6 is associated to cell migration and metastasis through EMT in osteosarcoma [308], colorectal [309] and gastric [310] cancers, where it was described as a prognostic factor.

MORC2 (MORC Family CW-Type Zinc Finger 2), that has a role in chromatin remodeling, regulation of DNA damage response and gene transcription, is possibly involved in EMT by indirectly inducing E-cadherin decrease, promoting breast cancer invasion and metastasis [311].

TAF4 (TATA-binding protein-associated factor 4), a component of the transcription factor IID (TFIID) complex whose binding to promoters initiates transcription, has a role in EMT through its hTAF4-TAFH domain activity [312]: in mice, loss of TAF4 leads to malignant transformation of chemically induced papilloma and the occurrence of invasive melanocytic tumors [313], whereas in humans TAF4 was found amplified and overexpressed in high-grade serous ovarian cancers [314].

Lastly, *ZNF516* (Zinc Finger Protein 516), that showed a 3 CpGs cluster hypomethylated in MPM cases, is a chromosomal instability suppressor gene that undergoes copy number loss in CIN+ colorectal cancer cells [315].

Even though unexpected, the finding of DMe in genes involved in EMT, transcriptional regulation, and genomic stability in DNA from whole blood, instead of tumor tissue, is quite intriguing: this is a suggestive finding, since the involvement of EMT in asbestos related diseases and mesothelioma was previously described [316, 317], as well as the induction of EMT by chrysotile asbestos through a mechanism involving a signaling pathway mediated by transforming growth factor beta (TGF-β) [318]. Moreover, several of the DMRs identified in our study belong to cancer related pathways, such as DNA Repair, Membrane Trafficking, Vesicle-mediated transport, Signaling by NGF, and Fatty acid, triacylglycerol, and ketone body metabolism. This is a suggestive finding, since DMe of several adjacent CpGs in a gene region could underlie functional changes that may be related to MPM onset.

Taken together, our findings of DMe in WBCs suggest that DNAm changes in the immune system components may play a role in the oncogenic process triggered by asbestos exposure. However, with our study design we could not assess the possible functional role of methylation changes in WBCs, nor how these changes are related to MPM onset. As a matter of fact, our investigation in the Geno-Pheno and healthy pleura datasets of a possible relationship between DMRs methylation and transcript levels showed a relationship between methylation changes and gene expression only for *FOXK1* and *FYCO1* in pleura tissue, whilst we found no such indication for the other genes. Nevertheless, we cannot exclude functional links between methylation changes and transcriptional regulation in WBCs from MPM patients and asbestos exposed subjects, that we were not able to assess due to the lack of RNA collection in our original sample collection, nor we can exclude long range regulation of other genes possibly involved in MPM onset and progression, that are worthy to be assessed in future studies.

One main limitation to the functional interpretation of our results is that all MPM subjects have already developed the disease at recruitment: thus, our findings likely reflect the disease status rather than being markers of the dynamic processes leading to MPM onset. Also the unavailability of MPM tissue from the same subjects poses major constrains to the functional interpretation of our findings.

Notwithstanding the above limitations, the discrimination between MPM cases and non-MPM asbestos exposed controls improved when DNAm levels were taken into account together with asbestos exposure

levels. The need for early diagnostic biomarkers is of particular importance for a disease such as MPM characterized by the accumulation and persistence of asbestos fibers in the lungs, leading to a long latency period before clear clinical signs of the tumor are detectable. As of today, early MPM detection is still unattainable. Serum biomarkers such as Mesothelin, megakaryocyte potentiating factors, Osteopontin, CYFRA21-1, and Fibulin-3 were tested for early MPM diagnosis with limited results [319] and, so far, neither radiologic nor biochemical screening studies have proven useful. When most MPM patients are diagnosed, therapeutic options such as surgery, radiation and chemotherapy provide only limited benefits. Although evidence of the carcinogenic risk for asbestos exposure led to banishment of asbestos itself and asbestos-containing materials in several countries, individuals who have already been exposed will remain at risk of MPM, whose incidence is just expected to peak in the next years in European countries. The non-invasive methylation-based biomarkers may contribute to early MPM detection, that could anticipate interventions at a potentially treatable stage. Thus, our results may pose the rationale for future studies, and should be thoroughly investigated in additional MPM samples, especially in prospective cohorts where the predictive value of the markers we identified could be further assessed.

9. Conclusions

Cancer is generally defined as a complex disease characterized by an abnormal cellular proliferation; cancer development involves a sequence of interactions between genomic and epigenomic makeup and environment that cannot occur without dysfunction in multiple systems, including DNA repair, apoptosis and immune functions [320]. Deciphering the mechanisms that drive tumorigenesis requires the analysis of multiple measurements of different types of data.

In the past, due to financial constraints and availability issues, very few studies investigated more than one "omic" technology in the same study, providing limited information about the etiology of oncogenesis and tumor progression [321]. In the last decades, thanks to technological advancements and reducing costs, researchers have been able to monitor different "omics" together in a high-throughput manner, leading to the realization of an increasing number of "multi-omics" studies.

In this context, data integration — the combination of different types of measurement— has become increasingly crucial in the study of human complex diseases, being a valuable computational tool in providing insights not achievable with a single methodology. For example, evaluating DNA methylation and mutational status as well as gene expression at the genome-wide level together with other biomarkers in the same study, have demonstrated the usefulness of an integrative approach to tackle complexity and deeper analyze clinical outcomes [322].

In this dissertation, "multi-omics" approaches were proposed to identify biomarkers of diagnosis and/or prognosis of bladder cancer and malignant pleural mesothelioma. In both cases, a combination of multiple- and different-platform data information was employed.

In the BC study, we started investigating multiple DNA repair assays, obtaining significant results for different outcomes. Briefly, our data suggest that γ-H2AX can be considered as a possible molecular biomarker to identify patients with a higher risk of BC recurrence, increased DNA damage levels detected through a modified Comet assay may potentially represent a prognostic marker associated with poor BC survival, whilst the introduction in a model of MN frequency and selected polymorphisms differentially distributed in cases and controls improved BC patient identification. Through whole gene expression

analysis we observed a lower expression of *MMP23B* in BCs compared to controls; moreover, we reported the first evidence of MMP23B secretion both in plasma and urine and we investigated the expression levels of miRNAs targeting *MMP23B* gene in urine of patients and controls. Our results indicate a potential role of this poorly characterized metalloproteinase in BC.

In the second study on MPM, we used a whole genome DNA methylation approach to investigate signatures of differential methylation in WBCs from MPM cases and healthy controls, all of them exposed to asbestos; we identify a panel of differentially methylated markers both at single-CpG and at genomic regions level. These signals were additionally correlated with the corresponding gene-expression levels in 2 separate datasets: we found statistical significant associations for CpGs in *FYCO* and *FOXK1* genes. Overall, our results suggest the potential use of DNA methylation profiles in blood to develop non-invasive tests for MPM detection in asbestos-exposed subjects.

Although further efforts are still needed, the results obtained from these approaches strongly suggest that integrating knowledge at different levels is a powerful tool for characterizing the mechanisms that underlie cancer.

References

1. Sanli, O., et al., *Bladder cancer.* Nature reviews. Disease primers, 2017. **3**: p. 17022.
2. Martin, J.W., et al., *Squamous cell carcinoma of the urinary bladder: Systematic review of clinical characteristics and therapeutic approaches.* Arab journal of urology, 2016. **14**(3): p. 183-91.
3. Willis, D. and A.M. Kamat, *Nonurothelial bladder cancer and rare variant histologies.* Hematology/oncology clinics of North America, 2015. **29**(2): p. 237-52, viii.
4. Burger, M., et al., *Epidemiology and risk factors of urothelial bladder cancer.* European urology, 2013. **63**(2): p. 234-41.
5. Babjuk, M., et al., *EAU Guidelines on Non-Muscle-invasive Urothelial Carcinoma of the Bladder: Update 2016.* European urology, 2017. **71**(3): p. 447-461.
6. May, M., et al., *Prognostic accuracy of individual uropathologists in noninvasive urinary bladder carcinoma: a multicentre study comparing the 1973 and 2004 World Health Organisation classifications.* European urology, 2010. **57**(5): p. 850-8.
7. Sylvester, R.J., et al., *Long-term efficacy results of EORTC genito-urinary group randomized phase 3 study 30911 comparing intravesical instillations of epirubicin, bacillus Calmette-Guerin, and bacillus Calmette-Guerin plus isoniazid in patients with intermediate- and high-risk stage Ta T1 urothelial carcinoma of the bladder.* European urology, 2010. **57**(5): p. 766-73.
8. Seo, K.W., et al., *The efficacy of the EORTC scoring system and risk tables for the prediction of recurrence and progression of non-muscle-invasive bladder cancer after intravesical bacillus calmette-guerin instillation.* Korean journal of urology, 2010. **51**(3): p. 165-70.
9. Cumberbatch, M.G., et al., *The Role of Tobacco Smoke in Bladder and Kidney Carcinogenesis: A Comparison of Exposures and Meta-analysis of Incidence and Mortality Risks.* European urology, 2016. **70**(3): p. 458-66.
10. Antoni, S., et al., *Bladder Cancer Incidence and Mortality: A Global Overview and Recent Trends.* European urology, 2017. **71**(1): p. 96-108.
11. Moolgavkar, S.H. and R.G. Stevens, *Smoking and cancers of bladder and pancreas: risks and temporal trends.* Journal of the National Cancer Institute, 1981. **67**(1): p. 15-23.
12. Letasiova, S., et al., *Bladder cancer, a review of the environmental risk factors.* Environmental health : a global access science source, 2012. **11 Suppl 1**: p. S11.
13. Saint-Jacques, N., et al., *Estimating the risk of bladder and kidney cancer from exposure to low-levels of arsenic in drinking water, Nova Scotia, Canada.* Environment international, 2018. **110**: p. 95-104.
14. Westhoff, E., et al., *Dietary patterns and risk of recurrence and progression in non-muscle-invasive bladder cancer.* International journal of cancer, 2017.
15. Botteri, E., et al., *Alcohol consumption and risk of urothelial cell bladder cancer in the European prospective investigation into cancer and nutrition cohort.* International journal of cancer, 2017. **141**(10): p. 1963-1970.
16. Yee, D.S., et al., *Ethnic differences in bladder cancer survival.* Urology, 2011. **78**(3): p. 544-9.
17. Kiemeney, L.A., et al., *Sequence variant on 8q24 confers susceptibility to urinary bladder cancer.* Nature genetics, 2008. **40**(11): p. 1307-12.
18. Selinski, S., *Genetic variants confer susceptibility to urinary bladder cancer: an updated list of confirmed polymorphisms.* EXCLI journal, 2012. **11**: p. 743-7.
19. Selinski, S., *Highlight report: Functional consequences of urinary bladder cancer risk variants.* EXCLI journal, 2013. **12**: p. 1017-9.
20. Selinski, S., *Urinary bladder cancer risk variants: recent findings and new challenges of GWAS and confirmatory studies.* Archives of toxicology, 2014. **88**(7): p. 1469-75.
21. Selinski, S., et al., *Refinement of the prediction of N-acetyltransferase 2 (NAT2) phenotypes with respect to enzyme activity and urinary bladder cancer risk.* Archives of toxicology, 2013. **87**(12): p. 2129-39.
22. Golka, K., et al., *Genetic variants in urinary bladder cancer: collective power of the "wimp SNPs".* Archives of toxicology, 2011. **85**(6): p. 539-54.
23. Dudek, A.M., et al., *Urinary bladder cancer susceptibility markers. What do we know about functional mechanisms?* International journal of molecular sciences, 2013. **14**(6): p. 12346-66.
24. Selinski, S., et al., *Identification and replication of the interplay of four genetic high-risk variants for urinary bladder cancer.* Carcinogenesis, 2017. **38**(12): p. 1167-1179.
25. Selinski, S., *Discovering urinary bladder cancer risk variants: Status quo after almost ten years of genome-wide association studies.* EXCLI journal, 2017. **16**: p. 1288-1296.
26. An, Y., et al., *Meta-analysis of the relationship between slow acetylation of N-acetyl transferase 2 and the risk of bladder cancer.* Genetics and molecular research : GMR, 2015. **14**(4): p. 16896-904.

27. Figueroa, J.D., et al., *Genome-wide interaction study of smoking and bladder cancer risk.* Carcinogenesis, 2014. **35**(8): p. 1737-44.
28. Garcia-Closas, M., et al., *NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses.* Lancet, 2005. **366**(9486): p. 649-59.
29. Lukas, C., et al., *Occupational bladder cancer: Polymorphisms of xenobiotic metabolizing enzymes, exposures, and prognosis.* Journal of toxicology and environmental health. Part A, 2017. **80**(7-8): p. 439-452.
30. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.* International journal of cancer, 2015. **136**(5): p. E359-86.
31. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2016.* CA: a cancer journal for clinicians, 2016. **66**(1): p. 7-30.
32. Mahdavifar, N., et al., *Epidemiology, Incidence and Mortality of Bladder Cancer and their Relationship with the Development Index in the World.* Asian Pacific journal of cancer prevention : APJCP, 2016. **17**(1): p. 381-6.
33. Gunlusoy, B., et al., *The potential effect of age on the natural behavior of bladder cancer: Does urothelial cell carcinoma progress differently in various age groups?* The Kaohsiung journal of medical sciences, 2016. **32**(5): p. 261-6.
34. Volanis, D., et al., *Environmental factors and genetic susceptibility promote urinary bladder cancer.* Toxicology letters, 2010. **193**(2): p. 131-7.
35. Wang, S.C., et al., *The gender difference and mortality-to-incidence ratio relate to health care disparities in bladder cancer: National estimates from 33 countries.* Scientific reports, 2017. **7**(1): p. 4360.
36. Kamel, M.H., et al., *Potential years of life lost due to urogenital cancer in the United States: trends from 1972 to 2006 based on data from the SEER database.* The Journal of urology, 2012. **187**(3): p. 868-71.
37. Boormans, J.L. and E.C. Zwarthoff, *Limited Funds for Bladder Cancer Research and What Can We Do About It.* Bladder cancer, 2016. **2**(1): p. 49-51.
38. Logan, C., M. Brown, and D. Hayne, *Intravesical therapies for bladder cancer - indications and limitations.* BJU international, 2012. **110 Suppl 4**: p. 12-21.
39. Alfred Witjes, J., et al., *Updated 2016 EAU Guidelines on Muscle-invasive and Metastatic Bladder Cancer.* European urology, 2017. **71**(3): p. 462-475.
40. Mathews, L., Cabarcas, S., Hurt, E., DNA repair of cancer stem cells, Springer Science+Business Media, 2013. DOI 10.1007/978-94-007-4590-2.
41. Tubbs, A. and A. Nussenzweig, *Endogenous DNA Damage as a Source of Genomic Instability in Cancer.* Cell, 2017. **168**(4): p. 644-656.
42. Jackson, S.P. and J. Bartek, *The DNA-damage response in human biology and disease.* Nature, 2009. **461**(7267): p. 1071-8.
43. Chatterjee, N. and G.C. Walker, *Mechanisms of DNA damage, repair, and mutagenesis.* Environmental and molecular mutagenesis, 2017. **58**(5): p. 235-263.
44. Alberts, B., Johnson, A., Lewis, J., et al., Molecular Biology of the Cell. 4th edition - DNA Repair. New York: Garland Science, 2002.
45. Torgovnick, A. and B. Schumacher, *DNA repair mechanisms in cancer development and therapy.* Frontiers in genetics, 2015. **6**: p. 157.
46. Jalal, S., J.N. Earley, and J.J. Turchi, *DNA repair: from genome maintenance to biomarker and therapeutic target.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2011. **17**(22): p. 6973-84.
47. Cleaver, J.E., *Diagnosis of Xeroderma Pigmentosum and Related DNA Repair-Deficient Cutaneous Diseases.* Current medical literature. Dermatology, 2008. **13**(2): p. 41-48.
48. Matullo, G., et al., *DNA repair gene polymorphisms, bulky DNA adducts in white blood cells and bladder cancer in a case-control study.* International journal of cancer, 2001. **92**(4): p. 562-7.
49. Andrew, A.S., et al., *Concordance of multiple analytical approaches demonstrates a complex relationship between DNA repair gene SNPs, smoking and bladder cancer susceptibility.* Carcinogenesis, 2006. **27**(5): p. 1030-7.
50. Sanyal, S., et al., *Polymorphisms in DNA repair and metabolic genes in bladder cancer.* Carcinogenesis, 2004. **25**(5): p. 729-34.
51. Shen, M., et al., *Polymorphisms of the DNA repair genes XRCC1, XRCC3, XPD, interaction with environmental exposures, and bladder cancer risk in a case-control study in northern Italy.* Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 2003. **12**(11 Pt 1): p. 1234-40.
52. Figueroa, J.D., et al., *Genetic variation in the base excision repair pathway and bladder cancer risk.* Human genetics, 2007. **121**(2): p. 233-42.

53. Xie, H., et al., *Genetic variations in base excision repair pathway and risk of bladder cancer: a case-control study in the United States.* Molecular carcinogenesis, 2015. **54**(1): p. 50-7.
54. Garcia-Closas, M., et al., *Genetic variation in the nucleotide excision repair pathway and bladder cancer risk.* Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 2006. **15**(3): p. 536-42.
55. Wu, X., et al., *Bladder cancer predisposition: a multigenic approach to DNA-repair and cell-cycle-control genes.* American journal of human genetics, 2006. **78**(3): p. 464-79.
56. Figueroa, J.D., et al., *Evaluation of genetic variation in the double-strand break repair pathway and bladder cancer risk.* Carcinogenesis, 2007. **28**(8): p. 1788-93.
57. Svatek, R.S., et al., *The economics of bladder cancer: costs and considerations of caring for this disease.* European urology, 2014. **66**(2): p. 253-62.
58. Marsit, C.J., et al., *DNA methylation array analysis identifies profiles of blood-derived DNA methylation associated with bladder cancer.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2011. **29**(9): p. 1133-9.
59. Friedrich, M.G., et al., *Prognostic relevance of methylation markers in patients with non-muscle invasive bladder carcinoma.* European journal of cancer, 2005. **41**(17): p. 2769-78.
60. Staack, A., et al., *Combined determination of plasma MMP2, MMP9, and TIMP1 improves the non-invasive detection of transitional cell carcinoma of the bladder.* BMC urology, 2006. **6**: p. 19.
61. Habuchi, T., et al., *Prognostic markers for bladder cancer: International Consensus Panel on bladder tumor markers.* Urology, 2005. **66**(6 Suppl 1): p. 64-74.
62. Sanguedolce, F., et al., *Predictive markers in bladder cancer: do we have molecular markers ready for clinical use?* Critical reviews in clinical laboratory sciences, 2014. **51**(5): p. 291-304.
63. Sacerdote, C., et al., *Polymorphisms in the XRCC1 gene modify survival of bladder cancer patients treated with chemotherapy.* International journal of cancer, 2013. **133**(8): p. 2004-9.
64. Stern, M.C., et al., *Polymorphisms in DNA repair genes, smoking, and bladder cancer risk: findings from the international consortium of bladder cancer.* Cancer research, 2009. **69**(17): p. 6857-64.
65. Russo, A., et al., *Shorter leukocyte telomere length is independently associated with poor survival in patients with bladder cancer.* Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 2014. **23**(11): p. 2439-46.
66. van Rhijn, B.W., H.G. van der Poel, and T.H. van der Kwast, *Urine markers for bladder cancer surveillance: a systematic review.* European urology, 2005. **47**(6): p. 736-48.
67. Ricci, S., D. Bruzzese, and D.I.C. A, *Evaluation of MMP-2, MMP-9, TIMP-1, TIMP-2, NGAL and MMP-9/NGAL complex in urine and sera from patients with bladder cancer.* Oncology letters, 2015. **10**(4): p. 2527-2532.
68. Matullo, G., A. Naccarati, and B. Pardini, *MicroRNA expression profiling in bladder cancer: the challenge of next-generation sequencing in tissues and biofluids.* International journal of cancer, 2016. **138**(10): p. 2334-45.
69. Witjes, J.A., et al., *EAU guidelines on muscle-invasive and metastatic bladder cancer: summary of the 2013 guidelines.* European urology, 2014. **65**(4): p. 778-92.
70. Kavanagh, J.N., et al., *DNA double strand break repair: a radiation perspective.* Antioxid Redox Signal, 2013. **18**(18): p. 2458-72.
71. Khanna, K.K. and S.P. Jackson, *DNA double-strand breaks: signaling, repair and the cancer connection.* Nat Genet, 2001. **27**(3): p. 247-54.
72. Kinner, A., et al., *Gamma-H2AX in recognition and signaling of DNA double-strand breaks in the context of chromatin.* Nucleic Acids Res, 2008. **36**(17): p. 5678-94.
73. Rothkamm, K. and M. Lobrich, *Evidence for a lack of DNA double-strand break repair in human cells exposed to very low x-ray doses.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(9): p. 5057-62.
74. Schmid, T.E., et al., *Differences in the kinetics of gamma-H2AX fluorescence decay after exposure to low and high LET radiation.* International journal of radiation biology, 2010. **86**(8): p. 682-91.
75. Ivashkevich, A., et al., *Use of the gamma-H2AX assay to monitor DNA damage and repair in translational cancer research.* Cancer letters, 2012. **327**(1-2): p. 123-33.
76. Sedelnikova, O.A., et al., *Histone H2AX in DNA damage and repair.* Cancer Biol Ther, 2003. **2**(3): p. 233-5.
77. Ivashkevich, A., et al., *Use of the gamma-H2AX assay to monitor DNA damage and repair in translational cancer research.* Cancer Lett, 2012. **327**(1-2): p. 123-33.
78. Valdiglesias, V., et al., *gammaH2AX as a marker of DNA double strand breaks and genomic instability in human population studies.* Mutat Res, 2013. **753**(1): p. 24-40.
79. Burger, M., et al., *Epidemiology and risk factors of urothelial bladder cancer.* Eur Urol, 2013. **63**(2): p. 234-41.

80. Bonner, W.M., et al., *GammaH2AX and cancer.* Nat Rev Cancer, 2008. **8**(12): p. 957-67.
81. Babjuk, M., et al., *EAU guidelines on non-muscle-invasive urothelial carcinoma of the bladder: update 2013.* Eur Urol, 2013. **64**(4): p. 639-53.
82. Witjes, J.A., et al., *EAU guidelines on muscle-invasive and metastatic bladder cancer: summary of the 2013 guidelines.* Eur Urol, 2014. **65**(4): p. 778-92.
83. Cheung, W.L., et al., *Phosphorylated H2AX in noninvasive low grade urothelial carcinoma of the bladder: correlation with tumor recurrence.* J Urol, 2009. **181**(3): p. 1387-92.
84. Xylinas, E., et al., *Blood- and tissue-based biomarkers for prediction of outcomes in urothelial carcinoma of the bladder.* Urol Oncol, 2014. **32**(3): p. 230-42.
85. Sanguedolce, F., et al., *Predictive markers in bladder cancer: Do we have molecular markers ready for clinical use?* Crit Rev Clin Lab Sci, 2014: p. 1-14.
86. Chang, D.W., J. Gu, and X. Wu, *Germline prognostic markers for urinary bladder cancer: obstacles and opportunities.* Urol Oncol, 2012. **30**(4): p. 524-32.
87. Choudhury, A., et al., *MRE11 expression is predictive of cause-specific survival following radical radiotherapy for muscle-invasive bladder cancer.* Cancer Res, 2010. **70**(18): p. 7017-26.
88. Fernandez, M.I., et al., *gamma-H2AX level in peripheral blood lymphocytes as a risk predictor for bladder cancer.* Carcinogenesis, 2013. **34**(11): p. 2543-7.
89. Xu, E., et al., *Risk assessment of esophageal adenocarcinoma using gamma-H2AX assay.* Cancer Epidemiol Biomarkers Prev, 2013. **22**(10): p. 1797-804.
90. He, Y., et al., *Ionizing radiation-induced gamma-H2AX activity in whole blood culture and the risk of lung cancer.* Cancer Epidemiol Biomarkers Prev, 2013. **22**(3): p. 443-51.
91. Bonassi, S., et al., *Chromosomal aberrations and risk of cancer in humans: an epidemiologic perspective.* Cytogenetic and genome research, 2004. **104**(1-4): p. 376-82.
92. Bonassi, S., et al., *An increased micronucleus frequency in peripheral blood lymphocytes predicts the risk of cancer in humans.* Carcinogenesis, 2007. **28**(3): p. 625-31.
93. Fenech, M., *Cytokinesis-block micronucleus cytome assay.* Nature protocols, 2007. **2**(5): p. 1084-104.
94. Mittal, R.D., et al., *XRCC1 codon 399 mutant allele: a risk factor for recurrence of urothelial bladder carcinoma in patients on BCG immunotherapy.* Cancer Biol Ther, 2008. **7**(5): p. 645-50.
95. Andrew, A.S., et al., *Genetic polymorphisms modify bladder cancer recurrence and survival in a USA population-based prognostic study.* BJU Int, 2014.
96. Wang, M., et al., *A novel XPF -357A>C polymorphism predicts risk and recurrence of bladder cancer.* Oncogene, 2010. **29**(13): p. 1920-8.
97. Sanyal, S., et al., *Polymorphisms in XPD, XPC and the risk of death in patients with urinary bladder neoplasms.* Acta Oncol, 2007. **46**(1): p. 31-41.
98. Shariat, S.F., et al., *Multiple biomarkers improve prediction of bladder cancer recurrence and mortality in patients undergoing cystectomy.* Cancer, 2008. **112**(2): p. 315-25.
99. Shariat, S.F., et al., *Combination of multiple molecular markers can improve prognostication in patients with locally advanced and lymph node positive bladder cancer.* J Urol, 2010. **183**(1): p. 68-75.
100. Babjuk, M., et al., *EAU guidelines on non-muscle-invasive urothelial carcinoma of the bladder: update 2013.* European urology, 2013. **64**(4): p. 639-53.
101. Bonassi, S., et al., *Micronuclei frequency in peripheral blood lymphocytes and cancer risk: evidence from human studies.* Mutagenesis, 2011. **26**(1): p. 93-100.
102. Duffaud, F., et al., *Micronucleated lymphocyte rates from head-and-neck cancer patients.* Mutation research, 1999. **439**(2): p. 259-66.
103. El-Zein, R.A., et al., *Cytokinesis-blocked micronucleus assay as a novel biomarker for lung cancer risk.* Cancer research, 2006. **66**(12): p. 6449-56.
104. Fenech, M., et al., *Molecular mechanisms of micronucleus, nucleoplasmic bridge and nuclear bud formation in mammalian and human cells.* Mutagenesis, 2011. **26**(1): p. 125-32.
105. Slyskova, J., et al., *Functional, genetic, and epigenetic aspects of base and nucleotide excision repair in colorectal carcinomas.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2012. **18**(21): p. 5878-87.
106. Vafa, O., et al., *c-Myc can induce DNA damage, increase reactive oxygen species, and mitigate p53 function: a mechanism for oncogene-induced genetic instability.* Molecular cell, 2002. **9**(5): p. 1031-44.
107. Khanna, A., *DNA Damage in Cancer Therapeutics: A Boon or a Curse?* Cancer research, 2015. **75**(11): p. 2133-2138.
108. Luzhna, L., P. Kathiria, and O. Kovalchuk, *Micronuclei in genotoxicity assessment: from genetics to epigenetics and beyond.* Frontiers in genetics, 2013. **4**: p. 131.

109. Nersesyan, A., et al., *Micronucleus assay with urine derived cells (UDC): a review of its application in human studies investigating genotoxin exposure and bladder cancer risk.* Mutation research. Reviews in mutation research, 2014. **762**: p. 37-51.
110. He, J., et al., *Associations of Lys939Gln and Ala499Val polymorphisms of the XPC gene with cancer susceptibility: a meta-analysis.* International journal of cancer, 2013. **133**(8): p. 1765-75.
111. Peng, Q., et al., *DNA repair gene XRCC3 polymorphisms and bladder cancer risk: a meta-analysis.* Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine, 2014. **35**(3): p. 1933-44.
112. Wei, H., et al., *Association of polymorphisms in oxidative stress genes with clinical outcomes for bladder cancer treated with Bacillus Calmette-Guerin.* PLoS One, 2012. **7**(6): p. e38533.
113. Zhu, G., et al., *Association of nineteen polymorphisms from seven DNA repair genes and the risk for bladder cancer in gansu province of China.* Oncotarget, 2016.
114. Li, C., et al., *Polymorphisms of nucleotide excision repair genes predict melanoma survival.* The Journal of investigative dermatology, 2013. **133**(7): p. 1813-21.
115. Slyskova, J., et al., *Differences in nucleotide excision repair capacity between newly diagnosed colorectal cancer patients and healthy controls.* Mutagenesis, 2012. **27**(4): p. 519-22.
116. Wang, Y., et al., *Association between CCND1 and XPC polymorphisms and bladder cancer risk: a meta-analysis based on 15 case-control studies.* Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine, 2014. **35**(4): p. 3155-65.
117. Iarmarcovai, G., et al., *Genetic polymorphisms and micronucleus formation: a review of the literature.* Mutation research, 2008. **658**(3): p. 215-33.
118. Angelini, S., et al., *Micronuclei in humans induced by exposure to low level of ionizing radiation: influence of polymorphisms in DNA repair genes.* Mutation research, 2005. **570**(1): p. 105-17.
119. Serra, H., et al., *Roles of XRCC2, RAD51B and RAD51D in RAD51-independent SSA recombination.* PLoS genetics, 2013. **9**(11): p. e1003971.
120. Bausinger, J. and G. Speit, *DNA repair capacity of cultured human lymphocytes exposed to mutagens measured by the comet assay and array expression analysis.* Mutagenesis, 2015. **30**(6): p. 811-20.
121. Arora, S.K., P. Dey, and U.N. Saikia, *Micronucleus in atypical urothelial cells.* Diagnostic cytopathology, 2010. **38**(11): p. 811-3.
122. Sharma, S., et al., *Micronucleus scoring in urine cytology specimen by DNA-specific stain.* Analytical and quantitative cytology and histology / the International Academy of Cytology [and] American Society of Cytology, 2012. **34**(1): p. 56-60.
123. de Castro Marcondes, J.P., et al., *Genetic instability persists in non-neoplastic urothelial cells from patients with a history of urothelial cell carcinoma.* PLoS One, 2014. **9**(1): p. e86162.
124. Zijno, A., F. Saini, and R. Crebelli, *Suitability of cryopreserved isolated lymphocytes for the analysis of micronuclei with the cytokinesis-block method.* Mutagenesis, 2007. **22**(5): p. 311-5.
125. D'Errico, M., et al., *Factors that influence the DNA repair capacity of normal and skin cancer-affected individuals.* Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 1999. **8**(6): p. 553-9.
126. Chang, J.L., et al., *DNA damage and repair measurements from cryopreserved lymphocytes without cell culture--a reproducible assay for intervention studies.* Environmental and molecular mutagenesis, 2006. **47**(7): p. 503-8.
127. Teo, M.Y., et al., *DNA Damage Response and Repair Gene Alterations Are Associated with Improved Survival in Patients with Platinum-Treated Advanced Urothelial Carcinoma.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2017.
128. Stover, E.H., et al., *Biomarkers of Response and Resistance to DNA Repair Targeted Therapies.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2016. **22**(23): p. 5651-5660.
129. Faust, F., et al., *The use of the alkaline comet assay with lymphocytes in human biomonitoring studies.* Mutation research, 2004. **566**(3): p. 209-29.
130. Galardi, F., et al., *Inter- and intra-tumoral heterogeneity in DNA damage evaluated by comet assay in early breast cancer patients.* Breast, 2012. **21**(3): p. 336-42.
131. Herrera, M., et al., *Differences in repair of DNA cross-links between lymphocytes and epithelial tumor cells from colon cancer patients measured in vitro with the comet assay.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2009. **15**(17): p. 5466-72.
132. Vande Loock, K., et al., *An aphidicolin-block nucleotide excision repair assay measuring DNA incision and repair capacity.* Mutagenesis, 2010. **25**(1): p. 25-32.
133. Allione, A., et al., *Validation of the nucleotide excision repair comet assay on cryopreserved PBMCs to measure inter-individual variation in DNA repair capacity.* Mutagenesis, 2013. **28**(1): p. 65-70.

134. Speit, G., P. Schutz, and H. Hoffmann, *Enhancement of genotoxic effects in the comet assay with human blood samples by aphidicolin.* Toxicology letters, 2004. **153**(3): p. 303-10.
135. Comperat, E. and J. Varinot, *Immunochemical and molecular assessment of urothelial neoplasms and aspects of the 2016 World Health Organization classification.* Histopathology, 2016. **69**(5): p. 717-726.
136. Zhao, M., X.L. He, and X.D. Teng, *Understanding the molecular pathogenesis and prognostics of bladder cancer: an overview.* Chinese journal of cancer research = Chung-kuo yen cheng yen chiu, 2016. **28**(1): p. 92-8.
137. Talukdar, S., et al., *Noninvasive approaches for detecting and monitoring bladder cancer.* Expert review of anticancer therapy, 2015. **15**(3): p. 283-94.
138. Neumann, A.S., E.M. Sturgis, and Q. Wei, *Nucleotide excision repair as a marker for susceptibility to tobacco-related cancers: a review of molecular epidemiological studies.* Molecular carcinogenesis, 2005. **42**(2): p. 65-92.
139. Qiu, J., et al., *Attenuated NER expressions of XPF and XPC associated with smoking are involved in the recurrence of bladder cancer.* PloS one, 2014. **9**(12): p. e115224.
140. Gu, J., et al., *Nucleotide excision repair gene polymorphisms and recurrence after treatment for superficial bladder cancer.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2005. **11**(4): p. 1408-15.
141. Allione, A., et al., *Inter-individual variation in nucleotide excision repair pathway is modulated by non-synonymous polymorphisms in ERCC4 and MBD4 genes.* Mutation research, 2013. **751-752**: p. 49-54.
142. Savina, N.V., et al., *The Cellular Response to Oxidatively Induced DNA Damage and Polymorphism of Some DNA Repair Genes Associated with Clinicopathological Features of Bladder Cancer.* Oxidative medicine and cellular longevity, 2016. **2016**: p. 5710403.
143. Michiels, S., et al., *Genetic polymorphisms in 85 DNA repair genes and bladder cancer risk.* Carcinogenesis, 2009. **30**(5): p. 763-8.
144. Mouw, K.W., *DNA Repair Pathway Alterations in Bladder Cancer.* Cancers, 2017. **9**(4).
145. Mullane, S.A., et al., *Expression Levels of DNA Damage Repair Proteins Are Associated With Overall Survival in Platinum-Treated Advanced Urothelial Carcinoma.* Clinical genitourinary cancer, 2016. **14**(4): p. 352-9.
146. Desai, N.B., et al., *Genomic characterization of response to chemoradiation in urothelial bladder cancer.* Cancer, 2016. **122**(23): p. 3715-3723.
147. Schabath, M.B., et al., *Genetic instability in bladder cancer assessed by the comet assay.* Journal of the National Cancer Institute, 2003. **95**(7): p. 540-7.
148. Sylvester, R.J., et al., *Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials.* European urology, 2006. **49**(3): p. 466-5; discussion 475-7.
149. van Rhijn, B.W., et al., *The pathologist's mean grade is constant and individualizes the prognostic value of bladder cancer grading.* European urology, 2010. **57**(6): p. 1052-7.
150. McKenna, D.J., S.R. McKeown, and V.J. McKelvey-Martin, *Potential use of the comet assay in the clinical management of cancer.* Mutagenesis, 2008. **23**(3): p. 183-90.
151. Santos, R.A., et al., *Basal levels of DNA damage detected by micronuclei and comet assays in untreated breast cancer patients and healthy women.* Clinical and experimental medicine, 2010. **10**(2): p. 87-92.
152. Synowiec, E., et al., *Association between DNA damage, DNA repair genes variability and clinical characteristics in breast cancer patients.* Mutation research, 2008. **648**(1-2): p. 65-72.
153. Gabelova, A., et al., *Radiosensitivity of peripheral blood lymphocytes from healthy donors and cervical cancer patients; the correspondence of in vitro data with the clinical outcome.* Neoplasma, 2008. **55**(3): p. 182-91.
154. Lorenzo, Y., et al., *Differential genetic and functional markers of second neoplasias in Hodgkin's disease patients.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2009. **15**(15): p. 4823-8.
155. Vasavi, M., et al., *Assessment of genetic damage in inflammatory, precancerous, and cancerous pathologies of the esophagus using the comet assay.* Genetic testing and molecular biomarkers, 2010. **14**(4): p. 477-82.
156. Sigurdson, A.J., et al., *Prospective analysis of DNA damage and repair markers of lung cancer risk from the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial.* Carcinogenesis, 2011. **32**(1): p. 69-73.
157. Kosti, O., et al., *DNA damage phenotype and prostate cancer risk.* Mutation research, 2011. **719**(1-2): p. 41-6.
158. Moller, P., et al., *The comet assay as a rapid test in biomonitoring occupational exposure to DNA-damaging agents and effect of confounding factors.* Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 2000. **9**(10): p. 1005-15.
159. Xylinas, E., et al., *Blood- and tissue-based biomarkers for prediction of outcomes in urothelial carcinoma of the bladder.* Urologic oncology, 2014. **32**(3): p. 230-42.

160. *Comprehensive molecular characterization of urothelial bladder carcinoma.* Nature, 2014. **507**(7492): p. 315-22.
161. Szarvas, T., et al., *Matrix metalloproteinases and their clinical relevance in urinary bladder cancer.* Nature reviews. Urology, 2011. **8**(5): p. 241-54.
162. Hadler-Olsen, E., et al., *Regulation of matrix metalloproteinase activity in health and disease.* The FEBS journal, 2011. **278**(1): p. 28-45.
163. Pardini, B., et al., *Increased micronucleus frequency in peripheral blood lymphocytes predicts the risk of bladder cancer.* British journal of cancer, 2017. **116**(2): p. 202-210.
164. Turinetto, V., et al., *H2AX phosphorylation level in peripheral blood mononuclear cells as an event-free survival predictor for bladder cancer.* Molecular carcinogenesis, 2016. **55**(11): p. 1833-1842.
165. Bustin, S.A., et al., *The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments.* Clinical chemistry, 2009. **55**(4): p. 611-22.
166. Diez-Villanueva, A., I. Mallona, and M.A. Peinado, *Wanderer, an interactive viewer to explore DNA methylation and gene expression data in human cancer.* Epigenetics & chromatin, 2015. **8**: p. 22.
167. Guo, Y., et al., *Large scale comparison of gene expression levels by microarrays and RNAseq using TCGA data.* PloS one, 2013. **8**(8): p. e71462.
168. Reid, C.N., et al., *Standardization of diagnostic biomarker concentrations in urine: the hematuria caveat.* PloS one, 2012. **7**(12): p. e53354.
169. Du, P., W.A. Kibbe, and S.M. Lin, *lumi: a pipeline for processing Illumina microarray.* Bioinformatics, 2008. **24**(13): p. 1547-8.
170. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic acids research, 2015. **43**(7): p. e47.
171. Schmittgen, T.D. and K.J. Livak, *Analyzing real-time PCR data by the comparative C(T) method.* Nature protocols, 2008. **3**(6): p. 1101-8.
172. Parveen, A., N. Gretz, and H. Dweep, *Obtaining miRNA-Target Interaction Information from miRWalk2.0.* Current protocols in bioinformatics, 2016. **55**: p. 12 15 1-12 15 27.
173. Pei, D., T. Kang, and H. Qi, *Cysteine array matrix metalloproteinase (CA-MMP)/MMP-23 is a type II transmembrane matrix metalloproteinase regulated by a single cleavage for both secretion and activation.* The Journal of biological chemistry, 2000. **275**(43): p. 33988-97.
174. Wild, P.J., et al., *Detection of urothelial bladder cancer cells in voided urine can be improved by a combination of cytology and standardized microsatellite analysis.* Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 2009. **18**(6): p. 1798-806.
175. Avritscher, E.B., et al., *Clinical model of lifetime cost of treating bladder cancer and associated complications.* Urology, 2006. **68**(3): p. 549-53.
176. Sapre, N., et al., *A urinary microRNA signature can predict the presence of bladder urothelial carcinoma in patients undergoing surveillance.* British journal of cancer, 2016. **114**(4): p. 454-62.
177. Zaravinos, A., et al., *Identification of common differentially expressed genes in urinary bladder cancer.* PloS one, 2011. **6**(4): p. e18135.
178. Zhang, S., et al., *Transcriptome profiling of a multiple recurrent muscle-invasive urothelial carcinoma of the bladder by deep sequencing.* PloS one, 2014. **9**(3): p. e91466.
179. Xiao, J. and Q. Yiqing, *Bioinformatics analysis of the gene expression profile in Bladder carcinoma.* Genetics and molecular biology, 2013. **36**(2): p. 287-91.
180. Descotes, F., et al., *Microarray gene expression profiling and analysis of bladder cancer supports the sub-classification of T1 tumours into T1a and T1b stages.* BJU international, 2014. **113**(2): p. 333-42.
181. Shen, Y., et al., *Differentially expressed genes and interacting pathways in bladder cancer revealed by bioinformatic analysis.* Molecular medicine reports, 2014. **10**(4): p. 1746-52.
182. Galea, C.A., et al., *Domain structure and function of matrix metalloprotease 23 (MMP23): role in potassium channel trafficking.* Cellular and molecular life sciences : CMLS, 2014. **71**(7): p. 1191-210.
183. Clancy, B.M., et al., *A gene expression profile for endochondral bone formation: oligonucleotide microarrays establish novel connections between known genes and BMP-2-induced bone formation in mouse quadriceps.* Bone, 2003. **33**(1): p. 46-63.
184. Heilstedt, H.A., et al., *Physical map of 1p36, placement of breakpoints in monosomy 1p36, and clinical characterization of the syndrome.* American journal of human genetics, 2003. **72**(5): p. 1200-12.
185. Riddick, A.C., et al., *Identification of degradome components associated with prostate cancer progression by expression analysis of human prostatic tissues.* British journal of cancer, 2005. **92**(12): p. 2171-80.
186. Singh, A., et al., *Higher than expected association of clinical prostate and bladder cancers.* The Journal of urology, 2008. **179**(5 Suppl): p. S2-5.

187. Vogel, C. and E.M. Marcotte, *Insights into the regulation of protein abundance from proteomic and transcriptomic analyses.* Nature reviews. Genetics, 2012. **13**(4): p. 227-32.
188. Salmena, L., et al., *A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?* Cell, 2011. **146**(3): p. 353-8.
189. Poliseno, L., et al., *A coding-independent function of gene and pseudogene mRNAs regulates tumour biology.* Nature, 2010. **465**(7301): p. 1033-8.
190. Armstrong, D.A., et al., *MicroRNA molecular profiling from matched tumor and bio-fluids in bladder cancer.* Molecular cancer, 2015. **14**: p. 194.
191. Zhou, X., et al., *Urinary cell-free microRNA-106b as a novel biomarker for detection of bladder cancer.* Medical oncology, 2014. **31**(10): p. 197.
192. Mearini, E., et al., *Expression of urinary miRNAs targeting NLRs inflammasomes in bladder cancer.* OncoTargets and therapy, 2017. **10**: p. 2665-2673.
193. Moogk, D., et al., *Melanoma expression of matrix metalloproteinase-23 is associated with blunted tumor immunity and poor responses to immunotherapy.* Journal of translational medicine, 2014. **12**: p. 342.
194. Hegedus, L., et al., *Additional MDA-MB-231 breast cancer cell matrix metalloproteinases promote invasiveness.* Journal of cellular physiology, 2008. **216**(2): p. 480-5.
195. Chinello, C., et al., *Tumor size, stage and grade alterations of urinary peptidome in RCC.* Journal of translational medicine, 2015. **13**: p. 332.
196. Rodriguez Faba, O., et al., *Matrix Metalloproteinases and Bladder Cancer: What is New?* ISRN urology, 2012. **2012**: p. 581539.
197. Offersen, B.V., et al., *Matrix metalloproteinase-9 measured in urine from bladder cancer patients is an independent prognostic marker of poor survival.* Acta oncologica, 2010. **49**(8): p. 1283-7.
198. Szarvas, T., et al., *Urinary matrix metalloproteinase-7 level is associated with the presence of metastasis in bladder cancer.* BJU international, 2011. **107**(7): p. 1069-73.
199. Eissa, S., et al., *Noninvasive diagnosis of bladder cancer by detection of matrix metalloproteinases (MMP-2 and MMP-9) and their inhibitor (TIMP-2) in urine.* European urology, 2007. **52**(5): p. 1388-96.
200. Di Carlo, A., et al., *Urinary gelatinase activities (matrix metalloproteinases 2 and 9) in human bladder tumors.* Oncology reports, 2006. **15**(5): p. 1321-6.
201. Chatziharalambous, D., et al., *Analytical Performance of ELISA Assays in Urine: One More Bottleneck towards Biomarker Validation and Clinical Implementation.* PloS one, 2016. **11**(2): p. e0149471.
202. Garde, A.H., et al., *Comparison of uncertainties related to standardization of urine samples with volume and creatinine concentration.* The Annals of occupational hygiene, 2004. **48**(2): p. 171-9.
203. Goldstein, S.L., *Urinary kidney injury biomarkers and urine creatinine normalization: a false premise or not?* Kidney international, 2010. **78**(5): p. 433-5.
204. Yap, T.A., et al., *Novel insights into mesothelioma biology and implications for therapy.* Nature reviews. Cancer, 2017. **17**(8): p. 475-488.
205. Shavelle, R., et al., *Life Expectancy in Pleural and Peritoneal Mesothelioma.* Lung cancer international, 2017. **2017**: p. 2782590.
206. Jean, D., et al., *Molecular changes in mesothelioma with an impact on prognosis and treatment.* Archives of pathology & laboratory medicine, 2012. **136**(3): p. 277-93.
207. Bibby, A.C., et al., *Malignant pleural mesothelioma: an update on investigation, diagnosis and treatment.* European respiratory review : an official journal of the European Respiratory Society, 2016. **25**(142): p. 472-486.
208. Butchart, E.G., et al., *Pleuropneumonectomy in the management of diffuse malignant mesothelioma of the pleura. Experience with 29 patients.* Thorax, 1976. **31**(1): p. 15-24.
209. Mott, F.E., *Mesothelioma: a review.* The Ochsner journal, 2012. **12**(1): p. 70-9.
210. Kadota, K., et al., *A nuclear grading system is a strong predictor of survival in epitheloid diffuse malignant pleural mesothelioma.* Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc, 2012. **25**(2): p. 260-71.
211. Scherpereel, A., et al., *Guidelines of the European Respiratory Society and the European Society of Thoracic Surgeons for the management of malignant pleural mesothelioma.* The European respiratory journal, 2010. **35**(3): p. 479-95.
212. World Health Organisation (WHO). Elimination of asbestos-related diseases. Geneva: WHO, 2006.
213. International Agency for Research on Cancer. IARC Monographs on Evaluation of Carcinogenic Risk to Humans: Volume 100c: Arsenic, Metals, Fibres, and Dust. WHO Press, 2012.
214. Marsili, D., et al., *Prevention of Asbestos-Related Disease in Countries Currently Using Asbestos.* International journal of environmental research and public health, 2016. **13**(5).

215. Doll, R., *Mortality from lung cancer in asbestos workers.* British journal of industrial medicine, 1955. **12**(2): p. 81-6.
216. Wagner, J.C., C.A. Sleggs, and P. Marchand, *Diffuse pleural mesothelioma and asbestos exposure in the North Western Cape Province.* Br J Ind Med, 1960. **17**: p. 260-71.
217. Roe, O.D. and G.M. Stella, *Malignant pleural mesothelioma: history, controversy and future of a manmade epidemic.* European respiratory review : an official journal of the European Respiratory Society, 2015. **24**(135): p. 115-31.
218. Odisio, E.G., et al., *Malignant Pleural Mesothelioma: Diagnosis, Staging, Pitfalls and Follow-up.* Seminars in ultrasound, CT, and MR, 2017. **38**(6): p. 559-570.
219. Geltner, C., et al., *Management of malignant pleural mesothelioma - part 1: epidemiology, diagnosis, and staging : Consensus of the Austrian Mesothelioma Interest Group (AMIG).* Wiener klinische Wochenschrift, 2016. **128**(17-18): p. 611-7.
220. Ferrante, D., et al., *Cancer mortality and incidence of mesothelioma in a cohort of wives of asbestos workers in Casale Monferrato, Italy.* Environmental health perspectives, 2007. **115**(10): p. 1401-5.
221. Mirabelli, D., et al., *Non-occupational exposure to asbestos and malignant mesothelioma in the Italian National Registry of Mesotheliomas.* Occup Environ Med, 2010. **67**(11): p. 792-4.
222. Matullo, G., et al., *Genetic variants associated with increased risk of malignant pleural mesothelioma: a genome-wide association study.* PloS one, 2013. **8**(4): p. e61253.
223. Roushdy-Hammady, I., et al., *Genetic-susceptibility factor and malignant mesothelioma in the Cappadocian region of Turkey.* Lancet, 2001. **357**(9254): p. 444-5.
224. de Klerk, N., et al., *Familial aggregation of malignant mesothelioma in former workers and residents of Wittenoom, Western Australia.* Int J Cancer, 2013. **132**(6): p. 1423-8.
225. Testa, J.R., et al., *Germline BAP1 mutations predispose to malignant mesothelioma.* Nature genetics, 2011. **43**(10): p. 1022-5.
226. Betti, M., et al., *Germline mutations in DNA repair genes predispose asbestos-exposed patients to malignant pleural mesothelioma.* Cancer letters, 2017. **405**: p. 38-45.
227. Goto, Y., et al., *Epigenetic profiles distinguish malignant pleural mesothelioma from lung adenocarcinoma.* Cancer research, 2009. **69**(23): p. 9073-82.
228. Fischer, J.R., et al., *Promoter methylation of RASSF1A, RARbeta and DAPK predict poor prognosis of patients with malignant mesothelioma.* Lung Cancer, 2006. **54**(1): p. 109-16.
229. Kubo, T., et al., *Epigenetic silencing of microRNA-34b/c plays an important role in the pathogenesis of malignant pleural mesothelioma.* Clin Cancer Res, 2011. **17**(15): p. 4965-74.
230. Tomasetti, M., et al., *Malignant mesothelioma: biology, diagnosis and therapeutic approaches.* Curr Mol Pharmacol, 2009. **2**(2): p. 190-206.
231. Robinson, B.M., *Malignant pleural mesothelioma: an epidemiological perspective.* Annals of cardiothoracic surgery, 2012. **1**(4): p. 491-6.
232. Le Stang, N., et al., *Evolution of pleural cancers and malignant pleural mesothelioma incidence in France between 1980 and 2005.* Int J Cancer, 2010. **126**(1): p. 232-8.
233. Frank, A.L. and T.K. Joshi, *The global spread of asbestos.* Ann Glob Health, 2014. **80**(4): p. 257-62.
234. Marinaccio, A., et al., *Analysis of latency time and its determinants in asbestos related malignant mesothelioma cases of the Italian register.* Eur J Cancer, 2007. **43**(18): p. 2722-8.
235. Magnani, C., et al., *Increased risk of malignant mesothelioma of the pleura after residential or domestic exposure to asbestos: a case-control study in Casale Monferrato, Italy.* Environ Health Perspect, 2001. **109**(9): p. 915-9.
236. Magnani, C., et al., *III Italian Consensus Conference on Malignant Mesothelioma of the Pleura. Epidemiology, Public Health and Occupational Medicine related issues.* Med Lav, 2015. **106**(5): p. 325-32.
237. Maule, M.M., et al., *Modeling mesothelioma risk associated with environmental asbestos exposure.* Environ Health Perspect, 2007. **115**(7): p. 1066-71.
238. Magnani, C., et al., *Pleural malignant mesothelioma and non-occupational exposure to asbestos in Casale Monferrato, Italy.* Occup Environ Med, 1995. **52**(6): p. 362-7.
239. van Zandwijk, N., et al., *Guidelines for the diagnosis and treatment of malignant pleural mesothelioma.* Journal of thoracic disease, 2013. **5**(6): p. E254-307.
240. Baas, P., et al., *Malignant pleural mesothelioma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Annals of oncology : official journal of the European Society for Medical Oncology, 2015. **26 Suppl 5**: p. v31-9.
241. Ando, H., et al., *Advanced therapeutic approach for the treatment of malignant pleural mesothelioma via the intrapleural administration of liposomal pemetrexed.* J Control Release, 2015. **220**(Pt A): p. 29-36.

242. Alexander, H.R., Jr., *Surgical treatment of malignant peritoneal mesothelioma: past, present, and future.* Ann Surg Oncol, 2010. **17**(1): p. 21-2.
243. Alexander, H.R., Jr., et al., *Treatment factors associated with long-term survival after cytoreductive surgery and regional chemotherapy for patients with malignant peritoneal mesothelioma.* Surgery, 2013. **153**(6): p. 779-86.
244. Friedberg, J.S., et al., *Extended Pleurectomy-Decortication-Based Treatment for Advanced Stage Epithelial Mesothelioma Yielding a Median Survival of Nearly Three Years.* Ann Thorac Surg, 2016.
245. Kondola, S., D. Manners, and A.K. Nowak, *Malignant pleural mesothelioma: an update on diagnosis and treatment options.* Therapeutic advances in respiratory disease, 2016. **10**(3): p. 275-88.
246. Ak, G., et al., *The effectiveness and safety of platinum-based pemetrexed and platinum-based gemcitabine treatment in patients with malignant pleural mesothelioma.* BMC Cancer, 2015. **15**: p. 510.
247. Rosenzweig, K.E. and P. Giraud, *Radiation therapy for malignant pleural mesothelioma.* Cancer radiotherapie : journal de la Societe francaise de radiotherapie oncologique, 2017. **21**(1): p. 73-76.
248. Gueugnon, F., et al., *Identification of novel markers for the diagnosis of malignant pleural mesothelioma.* Am J Pathol, 2011. **178**(3): p. 1033-42.
249. Panou, V., et al., *The established and future biomarkers of malignant pleural mesothelioma.* Cancer treatment reviews, 2015. **41**(6): p. 486-95.
250. Tan, D. and D.S. Zander, *Immunohistochemistry for assessment of pulmonary and pleural neoplasms: a review and update.* Int J Clin Exp Pathol, 2008. **1**(1): p. 19-31.
251. Kao, S.C., et al., *Estimation of an optimal chemotherapy utilization rate for malignant pleural mesothelioma: an evidence-based benchmark for cancer care.* Asia Pac J Clin Oncol, 2015. **11**(1): p. 85-92.
252. Grigoriu, B.D., et al., *Utility of osteopontin and serum mesothelin in malignant pleural mesothelioma diagnosis and prognosis assessment.* Clin Cancer Res, 2007. **13**(10): p. 2928-35.
253. Amati, M., et al., *Assessment of biomarkers in asbestos-exposed workers as indicators of cancer risk.* Mutat Res, 2008. **655**(1-2): p. 52-8.
254. Rodriguez Portal, J.A., et al., *Serum levels of soluble mesothelin-related peptides in malignant and nonmalignant asbestos-related pleural disease: relation with past asbestos exposure.* Cancer Epidemiol Biomarkers Prev, 2009. **18**(2): p. 646-50.
255. Filiberti, R., et al., *Individual predictors of increased serum mesothelin in asbestos-exposed workers.* Med Oncol, 2013. **30**(1): p. 422.
256. Filiberti, R., et al., *Is soluble mesothelin-related protein an upfront predictive marker of pleural mesothelioma? A prospective study on Italian workers exposed to asbestos.* Oncology, 2014. **86**(1): p. 33-43.
257. Roe, O.D., et al., *Mesothelin-related predictive and prognostic factors in malignant mesothelioma: a nested case-control study.* Lung Cancer, 2008. **61**(2): p. 235-43.
258. Park, E.K., P.S. Thomas, and D.H. Yates, *Biomarkers for early detection of mesothelioma in asbestos-exposed subjects.* Clin Chem Lab Med, 2010. **48**(11): p. 1673-4.
259. Park, E.K., et al., *Association of biomarker levels with severity of asbestos-related diseases.* Saf Health Work, 2012. **3**(1): p. 17-21.
260. Park, E.K., et al., *Blood-based early detection of malignant mesothelioma.* J Clin Oncol, 2009. **27**(1): p. 160; author reply 160.
261. Demir, M., et al., *Evaluation of New Biomarkers in the Prediction of Malignant Mesothelioma in Subjects with Environmental Asbestos Exposure.* Lung, 2016. **194**(3): p. 409-17.
262. Carbone, M. and H. Yang, *Molecular pathways: targeting mechanisms of asbestos and erionite carcinogenesis in mesothelioma.* Clin Cancer Res, 2012. **18**(3): p. 598-604.
263. Napolitano, A., et al., *HMGB1 and Its Hyperacetylated Isoform are Sensitive and Specific Serum Biomarkers to Detect Asbestos Exposure and to Identify Mesothelioma Patients.* Clin Cancer Res, 2016. **22**(12): p. 3087-96.
264. Marchetti, C., et al., *High mobility group box 1 is a novel substrate of dipeptidyl peptidase-IV.* Diabetologia, 2012. **55**(1): p. 236-44.
265. Baumann, F., et al., *Mesothelioma patients with germline BAP1 mutations have 7-fold improved long-term survival.* Carcinogenesis, 2015. **36**(1): p. 76-81.
266. Tabata, C., et al., *Serum HMGB1 as a prognostic marker for malignant pleural mesothelioma.* BMC Cancer, 2013. **13**: p. 205.
267. Hausler, S.F., et al., *Whole blood-derived miRNA profiles as potential new tools for ovarian cancer screening.* Br J Cancer, 2010. **103**(5): p. 693-700.
268. Weber, D.G., et al., *Identification of miRNA-103 in the cellular fraction of human peripheral blood as a potential biomarker for malignant mesothelioma--a pilot study.* PLoS One, 2012. **7**(1): p. e30221.
269. Weber, D.G., et al., *Combination of MiR-103a-3p and mesothelin improves the biomarker performance of malignant mesothelioma diagnosis.* PLoS One, 2014. **9**(12): p. e114483.

270. Creaney, J. and B.W.S. Robinson, *Malignant Mesothelioma Biomarkers: From Discovery to Use in Clinical Practice for Diagnosis, Monitoring, Screening, and Treatment.* Chest, 2017. **152**(1): p. 143-149.
271. Dong, Y., et al., *DNA methylation as an early diagnostic marker of cancer (Review).* Biomedical reports, 2014. **2**(3): p. 326-330.
272. Kanherkar, R.R., N. Bhatia-Dey, and A.B. Csoka, *Epigenetics across the human lifespan.* Front Cell Dev Biol, 2014. **2**: p. 49.
273. Klutstein, M., et al., *DNA Methylation in Cancer and Aging.* Cancer research, 2016.
274. Shivapurkar, N. and A.F. Gazdar, *DNA methylation based biomarkers in non-invasive cancer screening.* Current molecular medicine, 2010. **10**(2): p. 123-32.
275. Kagan, J., et al., *Towards Clinical Application of Methylated DNA Sequences as Cancer Biomarkers: A Joint NCI's EDRN and NIST Workshop on Standards, Methods, Assays, Reagents and Tools.* Cancer research, 2007. **67**(10): p. 4545-9.
276. Vandermeers, F., et al., *The role of epigenetics in malignant pleural mesothelioma.* Lung cancer, 2013. **81**(3): p. 311-8.
277. Teschendorff, A.E., et al., *An epigenetic signature in peripheral blood predicts active ovarian cancer.* PloS one, 2009. **4**(12): p. e8274.
278. Choi, J.Y., et al., *Association between global DNA hypomethylation in leukocytes and risk of breast cancer.* Carcinogenesis, 2009. **30**(11): p. 1889-97.
279. Dianzani, I., et al., *Polymorphisms in DNA repair genes as risk factors for asbestos-related malignant mesothelioma in a general population study.* Mutation research, 2006. **599**(1-2): p. 124-34.
280. Betti, M., et al., *XRCC1 and ERCC1 variants modify malignant mesothelioma risk: a case-control study.* Mutation research, 2011. **708**(1-2): p. 11-20.
281. Ferrante, D., et al., *Pleural mesothelioma and occupational and non-occupational asbestos exposure: a case-control study with quantitative risk assessment.* Occup Environ Med, 2016. **73**(3): p. 147-53.
282. Herman, J.G., et al., *Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(18): p. 9821-6.
283. Bibikova, M., et al., *High density DNA methylation array with single CpG site resolution.* Genomics, 2011. **98**(4): p. 288-95.
284. Bibikova, M., et al., *High-throughput DNA methylation profiling using universal bead arrays.* Genome research, 2006. **16**(3): p. 383-93.
285. Du, P., et al., *Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis.* BMC bioinformatics, 2010. **11**: p. 587.
286. Zhou, W., P.W. Laird, and H. Shen, *Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes.* Nucleic acids research, 2017. **45**(4): p. e22.
287. Lehne, B., et al., *A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies.* Genome Biol, 2015. **16**: p. 37.
288. Campanella, G., et al., *Epigenetic signatures of internal migration in Italy.* International journal of epidemiology, 2014.
289. Di Gaetano, C., et al., *An overview of the genetic structure within the Italian population from genome-wide data.* PloS one, 2012. **7**(9): p. e43759.
290. R-CoreTeam. *R: A language and environment for statistical computing. R Foundation for Statistical Computing.* 2014; Available from: Vienna, Austria, 2014 http://www.R-project.org/.
291. Horvath, S. and A.J. Levine, *HIV-1 Infection Accelerates Age According to the Epigenetic Clock.* The Journal of infectious diseases, 2015. **212**(10): p. 1563-73.
292. Christensen, B.C., et al., *Differentiation of lung adenocarcinoma, pleural mesothelioma, and nonmalignant pulmonary tissues using DNA methylation profiles.* Cancer research, 2009. **69**(15): p. 6315-21.
293. Wang, W., et al., *FOXKs promote Wnt/beta-catenin signaling by translocating DVL into the nucleus.* Developmental cell, 2015. **32**(6): p. 707-18.
294. Wu, M., et al., *FOXK1 interaction with FHL2 promotes proliferation, invasion and metastasis in colorectal cancer.* Oncogenesis, 2016. **5**(11): p. e271.
295. Peng, Y., et al., *Direct regulation of FOXK1 by C-jun promotes proliferation, invasion and metastasis in gastric cancer cells.* Cell death & disease, 2016. **7**(11): p. e2480.
296. Yu, H., et al., *The ubiquitin carboxyl hydrolase BAP1 forms a ternary complex with YY1 and HCF-1 and is a critical regulator of gene expression.* Molecular and cellular biology, 2010. **30**(21): p. 5071-85.
297. Betti, M., et al., *CDKN2A and BAP1 germline mutations predispose to melanoma and mesothelioma.* Cancer Lett, 2016. **378**(2): p. 120-30.
298. Betti, M., et al., *Inference on germline BAP1 mutations and asbestos exposure from the analysis of familial and sporadic mesothelioma in a high-risk area.* Genes, chromosomes & cancer, 2015. **54**(1): p. 51-62.

299. Behrends, C., et al., *Network organization of the human autophagy system.* Nature, 2010. **466**(7302): p. 68-76.
300. Leone, P., et al., *MHC class I antigen processing and presenting machinery: organization, function, and defects in tumor cells.* Journal of the National Cancer Institute, 2013. **105**(16): p. 1172-87.
301. Hur, E.M., et al., *LIME, a novel transmembrane adaptor protein, associates with p56lck and mediates T cell activation.* The Journal of experimental medicine, 2003. **198**(10): p. 1463-73.
302. Darash-Yahana, M., et al., *The chemokine CXCL16 and its receptor, CXCR6, as markers and promoters of inflammation-associated cancers.* PloS one, 2009. **4**(8): p. e6695.
303. Keppler, D., *Towards novel anti-cancer strategies based on cystatin function.* Cancer letters, 2006. **235**(2): p. 159-176.
304. Burns, K., et al., *Tollip, a new component of the IL-1RI pathway, links IRAK to the IL-1 receptor.* Nature cell biology, 2000. **2**(6): p. 346-51.
305. Kim, H., Y. Phung, and M. Ho, *Changes in global gene expression associated with 3D structure of tumors: an ex vivo matrix-free mesothelioma spheroid model.* PloS one, 2012. **7**(6): p. e39556.
306. Qi, F., et al., *Continuous exposure to chrysotile asbestos can cause transformation of human mesothelial cells via HMGB1 and TNF-alpha signaling.* The American journal of pathology, 2013. **183**(5): p. 1654-66.
307. Cesi, V., et al., *TGFbeta-induced c-Myb affects the expression of EMT-associated genes and promotes invasion of ER+ breast cancer cells.* Cell cycle, 2011. **10**(23): p. 4149-61.
308. Ma, Y., X. Xu, and M. Luo, *CXCR6 promotes tumor cell proliferation and metastasis in osteosarcoma through the Akt pathway.* Cellular immunology, 2017. **311**: p. 80-85.
309. Matsushita, K., et al., *Soluble CXCL16 in preoperative serum is a novel prognostic marker and predicts recurrence of liver metastases in colorectal cancer patients.* Annals of surgical oncology, 2012. **19 Suppl 3**: p. S518-27.
310. Jin, J.J., et al., *CXCR6 predicts poor prognosis in gastric cancer and promotes tumor metastasis through epithelial-mesenchymal transition.* Oncology reports, 2017. **37**(6): p. 3279-3286.
311. Liao, X.H., et al., *Chromatin remodeling protein MORC2 promotes breast cancer invasion and metastasis through a PRD domain-mediated interaction with CTNND1.* Oncotarget, 2017.
312. Kazantseva, J., et al., *Targeted alternative splicing of TAF4: a new strategy for cell reprogramming.* Scientific reports, 2016. **6**: p. 30852.
313. Fadloun, A., et al., *The TFIID subunit TAF4 regulates keratinocyte proliferation and has cell-autonomous and non-cell-autonomous tumour suppressor activity in mouse epidermis.* Development, 2007. **134**(16): p. 2947-58.
314. Ribeiro, J.R., et al., *Targeting TBP-Associated Factors in Ovarian Cancer.* Frontiers in oncology, 2014. **4**: p. 45.
315. Burrell, R.A., et al., *Replication stress links structural and numerical cancer chromosomal instability.* Nature, 2013. **494**(7438): p. 492-496.
316. Casarsa, C., et al., *Epithelial-to-mesenchymal transition, cell polarity and stemness-associated features in malignant pleural mesothelioma.* Cancer letters, 2011. **302**(2): p. 136-43.
317. Iwanami, T., et al., *Clinical significance of epithelial-mesenchymal transition-associated markers in malignant pleural mesothelioma.* Oncology, 2014. **86**(2): p. 109-16.
318. Gulino, G.R., et al., *Effects of Chrysotile Exposure in Human Bronchial Epithelial Cells: Insights into the Pathogenic Mechanisms of Asbestos-Related Diseases.* Environmental health perspectives, 2016. **124**(6): p. 776-84.
319. Wolff, H., et al., *Asbestos, asbestosis, and cancer, the Helsinki criteria for diagnosis and attribution 2014: recommendations.* Scand J Work Environ Health, 2015. **41**(1): p. 5-15.
320. Knox, S.S., *From 'omics' to complex disease: a systems biology approach to gene-environment interactions in cancer.* Cancer cell international, 2010. **10**: p. 11.
321. Zhang, W., F. Li, and L. Nie, *Integrating multiple 'omics' analysis for microbial biology: application and methodologies.* Microbiology, 2010. **156**(Pt 2): p. 287-301.
322. Kristensen, V.N., et al., *Principles and methods of integrative genomic analyses in cancer.* Nature reviews. Cancer, 2014. **14**(5): p. 299-313.