NATHALIA QUINTERO RUIZ

PERFIL DE MUTAGENICIDADE INDUZIDO PELA LUZ UVA E UVB EM CÉLULAS DE PACIENTES COM XERODERMA PIGMENTOSUM GRUPO C

Tese apresentada ao Programa de Pós-Graduação em Microbiologia do Instituto de Ciência Biomédicas da Universidade de São Paulo, para obtenção do Título de Doutor em Ciências.

São Paulo
2020

NATHALIA QUINTERO RUIZ

PERFIL DE MUTAGENICIDADE INDUZIDO PELA LUZ UVA E UVB EM CÉLULAS DE PACIENTES COM XERODERMA PIGMENTOSUM GRUPO C

Tese apresentada ao Programa de Pós-Graduação em Microbiologia do Instituto de Ciência Biomédicas da Universidade de São Paulo, para obtenção do Título de Doutor em Ciências.

Área de concentração: Microbiologia
Orientador: Prof. Dr. Carlos Frederico Martins Menck

Versão original

São Paulo
2020

NATHALIA QUINTERO RUIZ

MUTAGENICITY PROFILE INDUCED BY UVA AND UVB LIGHT IN CELLS FROM PATIENTS WITH XERODERMA PIGMENTOSUM GROUP C

Thesis presented to the Graduate Program in Microbiology at the Institute of Biomedical Science of the University of São Paulo, to obtain the title of Doctor of Science.

Concentration area: Microbiology
Advisor: Prof. Dr. Carlos Frederico Martins Menck

Original version

São Paulo
2020

CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
Serviço de Biblioteca e informação Biomédica
do Instituto de Ciências Biomédicas da Universidade de São Paulo

Ficha Catalográfica elaborada pelo(a) autor(a)

Quintero Ruiz, Nathalia
Mutagenicity profile induced by UVA and UVB light in cells from patients with xeroderma pigmentosum group C / Nathalia Quintero Ruiz; orientador Carlos Frederico Martins Menck. -- São Paulo, 2020.
141 p.

Tese (Doutorado)) -- Universidade de São Paulo, Instituto de Ciências Biomédicas.

1. Ultraviolet radiation. 2. Xeroderma pigmentos, XP-C. 3. Mutagenesis. 4. Exome. 5. Mutational signature. I. Martins Menck, Carlos Frederico, orientador. II. Título.

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Candidato(a): Nathalia QuinteroRuiz

Titulo da Dissertação/Tese: Mutagenicity profile induced by UVA and UVB light in cells from patients with xeroderma pigmentosum group C

Orientador: Carlos Frederico Martins Menck

A Comissão Julgadora dos trabalhos de Defesa da Dissertação de Mestrado/Tese de Doutorado, em sessão publica realizada a **...../...../.....**, considerou o(a) candidato(a):

( ) Aprovado(a) ( ) Reprovado(a)

Examinador(a): Assinatura: .....
Nome: .....
Instituição: .....

Examinador(a): Assinatura: .....
Nome: .....
Instituição: .....

Examinador(a): Assinatura: .....
Nome: .....
Instituição: .....

Presidente: Assinatura: .....
Nome: .....
Instituição: .....

ICB USP

Cidade Universitária "Armando de Salles Oliveira", Butantã, São Paulo, SP · Av. Professor Lineu Prestes, 2415 - ICB III - 05508 000
Comissão de Ética em Pesquisa - Telefone (11) 3091-7733 - e-mail: cep@icb.usp.br

CERTIFICADO DE ISENÇÃO

Certificamos que o Protocolo CEP-ICB nº **785/2015** referente ao projeto intitulado: "***Perfil da mutagenicidade causado por luz UVA e UVB sobre células de pacientes com xeroderma pigmentosum grupo C***" sob a responsabilidade de ***Nathalia Quintero Ruiz*** e orientação do(a) Prof.(a) Dr.(a) ***Carlos Frederico Martins Menck,*** do Departamento de Microbiologia, foi analisado pela **CEUA** - Comissão de Ética no Uso de Animais e pela **CEPSH** - Comissão de Ética em Pesquisa com Seres Humanos, tendo sido deliberado que o referido projeto não utilizará animais que estejam sob a égide da Lei nº 11.794, de 8 de outubro de 2008, nem envolverá procedimentos regulados pela Resolução CONEP nº 466 de 2012.

São Paulo, 14 de dezembro de 2015

Prof. Dr. **Anderson de Sá Nunes**
Coordenador CEUA ICB/USP

Prof. Dr. **Paolo M. A. Zanotto**
Coordenador CEPSH ICB/USP

Aos meus pais que sempre acreditam em mim.
Ao Edu que não soltou minha mão.

AGRADECIMENTOS

Agradeço imensamente a COLCIENCIAS, pelo financiamento para eu conseguir fazer este doutorado. Ao Departamento de Microbiologia do Instituto de Ciências Biomédicas da Universidade de São Paulo pela infraestrutura que possibilitou a realização desta tese, por seus professores e funcionários, por proporcionar excelentes condições de formação profissional e pessoal. Especialmente à Gisele que não sei como mas lembra de cada um dos alunos da pós-graduação e facilita sempre nossa vida. Do mesmo modo agradeço à FAPESP e à CAPES pelo financiamento do projeto ao qual fiquei vinculada. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Quero agradecer especialmente ao meu orientador, Carlos Menck, por ter aceitado o desafio de receber uma pessoa aleatória que mandou um e-mail pedindo uma oportunidade em seu laboratório. Desde o primeiro contato que tivemos percebi que não tinha errado na minha escolha, você além de ser um pesquisador extraordinário é uma pessoa maravilhosa que sempre se preocupa com as necessidades particulares de cada aluno. Espero que você não tenha se arrependido de me receber, rs.

Tenho muitas pessoas a quem eu quero agradecer e espero não esquecer de ninguém, mas se isso acontecer me perdoem, pois vocês já me conhecem e sou assim mesmo, esquecida e enrolada. Passaram-se um pouco mais de quatro anos desde que decidi arriscar tudo o que eu tinha e sair da minha zona de conforto para entrar nesta montanha russa que é fazer um doutorado longe de casa. Muitas pessoas acharam que eu era louca, que eu não ia conseguir pelo meu jeito “manteiga derretida”, que não valeria a pena o sacrifício. Confesso que em vários momentos eu também pensei desta forma, mas hoje eu tenho certeza que valeu a pena. Estou aqui fechando este capítulo e percebendo que eu sou uma pessoa muito afortunada, neste período além de fazer um doutorado, fiz amizades dentro e fora do laboratório, construí uma vida e sou muito feliz.

Particularmente quero agradecer todos os membros do Laboratório de Reparo de DNA, só tenho carinho e gratidão por cada um de vocês, que de uma forma ou outra viveram comigo este processo. A convivência no laboratório foi muito boa e como a Lu disse “o laboratório é a cara do chefe”, não tinha como este laboratório ser

ruim. Veri, você é o espirito do lab, sem você nada funcionaria igual, obrigada pelo suporte em todos os sentidos!

À mulherada do lab(!!!), por terem me acolhido com todo o carinho e me levado indistintamente para o bar ou para tomar café com um docinho, mesmo sem eu entender uma palavra das 10 conversas paralelas que rolavam, vocês me deram suporte sempre e com vocês aprendi o que é sororidade; Ale, Ligia, Nati, Gi, Marcelula, Clariss, Livia, Cá, Lu, Val, Ju, Pilar e Maira.

Nati, minha mãe linda, muito obrigada por me ensinar com tanta paciência, por me ajudar sempre que eu precisei, por me dar uma irmãzinha fofa como a Maju, por todas as caronas e principalmente por me mostrar que o bullying deve vir de casa para criar filhos fortes. Assim, também quero agradecer a Livia por me emprestar seus pais maravilhosos, me senti muito amada na sua família!. Obrigada André, meu amigo querido por me escutar durante todas minhas crises e por ter sempre tempo para ler meus textos e me ajudar a desenrolar várias ideias bloqueadas. Obrigada Fabi pela companhia neste último mês em casa e claro pela ajuda com o trabalho chato de formatação da tese, rs. Obrigada Nati, Tiago e Cá pelas conversas, sempre foi enriquecedor falar com vocês do trabalho, e claro de muitos outros assuntos.

Como eu falei também tem muitas pessoas fora do lab que fizeram minha estadia em SP muito mais legal, Adrian e Adriana, Cesar, Susana, Jimena, Isa, Agustin, Stefy, Sarita, Renato, Pedro, Didi, Daniel e Renata. Sem vocês não teria sido tão bom.

Finalmente, mas não menos importante eu quero agradecer as pessoas mais especiais para mim, minha família. Obrigada ao meu companheiro de vida, Edu, porque com você minha vida é mais leve, mais gostosa, mais feliz. Obrigada aos meu pais e meu irmão que a vida toda tem me enchido de amor e me apoiado em todas as minhas decisões. Eu sei que para vocês também não tem sido fácil, eu amo vocês. Tem uma palavra linda em português que a gente não tem no espanhol, SAUDADE, que define o que eu sinto todos os dias quando penso em vocês

“Navegar é preciso, viver não é preciso”

Fernando Pessoa

ABSTRACT

QUINTERO-RUIZ, N. **Mutagenicity profile induced by UVA and UVB light in cells from patients with xeroderma pigmentosum group C**. 2020. 141 p. Ph.D.Tesis (Microbiology) – Instituto de Ciências Biomédicas, Universidade de São Paulo, 2020.

Sunlight radiation is the main etiologic agent of skin cancer, mainly due to its ultraviolet (UV) component that damages the DNA molecule by inducing pyrimidine dimers (*i.e.* CPD, 6-4PP) and oxidized bases. If unrepaired, these lesions generate mainly mutations of the C>T and C>A types, respectively. Nucleotide excision repair (NER) is the main repair pathway to protect cells from UV-induced damage. Defects in NER proteins result in a clinical phenotype denominated xeroderma pigmentosum (XP) that is characterized by high incidence of skin cancer in body areas exposed to sunlight. UVA and UVB wavelengths are considered biologically relevant, as they are part of the sunlight spectrum that reaches the Earth's surface. UVA light is the most prevalent radiation (95%) and penetrates deep into the skin; however, it has low energy and is poorly absorbed by DNA. In contrast, UVB light is more energetic and directly absorbed by DNA, although it corresponds to a small fraction of the solar spectrum (5%) and does not penetrate deeply in the skin. The aim of this work was to perform a comparative study of the UVA and UVB mutagenic capacity using a novel strategy to detect mutations by whole exome sequencing (WES) of randomly selected clones. The XP-C cell line (NER deficient; mutated in *XPC* gene), and its isogenic cell line (complemented with wild-type *XPC* gene; denominated COMP) were used. Results indicate that XP-C cells exhibit a dose-dependent decrease in survival rates when exposed to UVA or UVB, compared to COMP cells. Also, both UV wavelengths increased apoptosis levels (sub-G1) and genotoxic stress (γH2AX). Moreover, only UVB causes a partial S/G2 arrest in XP-C cell cycle. The WES data showed a significant increase in mutagenesis of UVA and UVB-irradiated cells, as expected. Remarkably, the main point mutations induced by both irradiations were C>T transitions in the C<u>C</u>N and T<u>C</u>N tri-nucleotide contexts, which are potential sites for pyrimidine dimer formation. These transitions preferentially occurs at the 3' base of the 5'T<u>C</u> dimer and the CC>TT tandem mutations were also enriched. Additionally, the T>N base substitutions were significantly increased by UVB, however, it was not possible to relate this increase to the pyrimidine dimer context. On the other hand, C>A transversions were the second most frequent induced mutation by both UV wavelengths on XP-C cells, which can be related to oxidized bases. Finally, the mutagenic spectrum generated in UVA and UVB irradiated XP-C cells were very similar and the mutation signatures of the COSMIC of skin cancer (SBS7 group) were recovered. These results indicate that in cells without NER repair, a unique dose of UVA or UVB contribute to generate the typical spectrum of mutation found in skin cancer tumors and its mutagenic spectrum is similar.

Keywords: Ultraviolet radiation, xeroderma pigmentosum, XP-C, mutagenesis, exome, NGS, mutational signature, nucleotide excision repair.

RESUMO

QUINTERO-RUIZ, N. **Perfil de mutagenicidade induzido pela luz UVA e UVB em células de pacientes com xeroderma pigmentosum grupo C**. 2020. 141 f. Tese (Doutorado em Microbiologia) - Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2020.

A radiação solar é o principal agente etiológico no câncer de pele, principalmente devido ao seu componente ultravioleta (UV), que danifica a molécula de DNA pela indução de dímeros de pirimidina (CPD e 6-4PP) e bases oxidadas. Esses tipos de lesões são principalmente removidos pela via de reparo por excisão de nucleotídeos (do inglês, NER) e, quando não reparados, geram principalmente mutações nos tipos C>T e C>A, respectivamente. Defeitos nas proteínas de NER resultam em um fenótipo clínico denominado xeroderma pigmentosum (XP), que apresenta uma alta incidência de câncer de pele em áreas corporais expostas à luz solar. Tanto UVA como UVB são considerados biologicamente relevantes, pois fazem parte do espectro da luz solar que atinge a superfície da Terra. Luz UVA é a radiação mais prevalente (95%) e penetra profundamente na pele, no entanto, possui baixa energia e é pouco absorvido pelo DNA. Por outro lado, luz UVB é mais energético e diretamente absorvido pelo DNA, contudo, corresponde a uma pequena fração do espectro solar (5%) e não penetra profundamente na pele. O objetivo deste trabalho foi realizar um estudo comparativo da capacidade mutagênica de irradiação UVA e UVB, utilizando uma nova estratégia para detectar mutações por sequenciamento total do exoma (do inglês WES) de clones selecionados aleatoriamente. A linhagem celular XP-C (deficiente em NER e mutada no gene *XPC*) e sua linhagem celular isogênica (complementada com o gene *XPC* selvagem; denominada COMP) foram usadas como modelo. Os resultados indicam que as células XP-C exibem uma diminuição dose-dependente nas taxas de sobrevivência quando expostas a UVA e UVB, em comparação com células COMP. Além disso, ambos os comprimentos de onda UV aumentaram os níveis de apoptose (sub-G1) e estresse genotóxico (γH2AX). Ainda, somente UVB causou uma parada parcial no ciclo celular em S/G2. Os dados do WES mostraram um aumento significativo da mutagênese nas células irradiadas por UVA e UVB, conforme o esperado. Surpreendentemente, a principal mutação induzida por ambos os tipos de luz UV foi a transição C>T no contexto de tri-nucleotídeo C<u>C</u>N e T<u>C</u>N, locais potenciais para a formação de dímeros de pirimidina. Essa transição ocorre preferencialmente na base 3' do dímero 5'T<u>C</u> e mutações em tandem CC>TT também foram enriquecidas. Além disso, substituições de base T>N foram aumentadas significativamente pela UVB, no entanto, não foi possível relacionar esse aumento ao contexto do dímero de pirimidina. Por outro lado, a transversão C>A foi a segunda mutação mais frequente induzida por ambos os comprimentos de onda UV nas células XP-C, e pode ser devido a bases oxidadas. Finalmente, o espectro mutagênico gerado nas células XP-C irradiadas por UVA e UVB foi muito semelhante e as assinaturas de mutação para câncer de pele do COSMIC (grupo SBS7) foram recuperadas. Esses resultados indicam que em células sem reparo de NER, uma única dose de UVA ou UVB contribui para gerar o espectro típico de mutação encontrado em tumores de câncer de pele e o espectro mutagênico entre elas é semelhante.

Palavras-chave: Radiação ultravioleta, xeroderma pigmentosum, XP-C, mutagênese, exoma, NGS, assinatura mutacional, reparo por excisão de nucleotídeos.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

5mC	5-methylcytocine
5hmC	5-hydroxymethylcytosine
6-4PPs	Pyrimidine-pyrimidone (6,4) photoproducts
8oxoG	8-oxoguanine
A	Adenine
BER	Base excision repair
C	Cytosine
CI	Mitochondrial respiratory complex I
CII	Mitochondrial respiratory complex II
CHO	Chinese hamster ovary cell
COMP	XPC complemented cells
CSA	Cockayne syndrome A protein
CSB	Cockayne syndrome B protein
CPDs	Cyclobutane pyrimidine dimers
DDB	Damaged DNA binding complex
Del	Deletion
DNA	Deoxyribonucleic acid
ERCC1	DNA excision repair protein ERCC-1
ES	Embryonic stem cells
fsX	Frame shift
G	Guanine
GGR	Global genome repair sub-pathway
HR23B	UV excision repair protein RAD23 homolog B
NADPH	Nicotinamide adenine dinucleotide phosphate hydrogen
NER	Nucleotide excision repair
NGS	Next generation sequencing technology
Pol eta	DNA polymerase η
T	Thymine
TCR	Transcription-coupled repair sub-pathway
TFIIH	Transcription factor IIH
TLS	Translesion synthesis
OGG1	glycosylases

PCNA	Proliferating cell nuclear antigen
Pol eta	DNA polymerase η
RAD4	DNA repair protein RAD4
RAD23	UV excision repair protein RAD23
ROS	Reactive oxygen species
RNA	Ribonucleic acid
RPA	Replication protein A
RFC	Replication factor C small subunit
SNVs	Single nucleotide variants
ssDNA	single - stranded DNA
TDG	glycosylases
UV	Ultraviolet
UVA	Ultraviolet radiation type A or long wave (315-400 nm)
UVB	Ultraviolet radiation type B or medium wave (280-315 nm)
UVC	Ultraviolet radiation type C or short wave (280-315 nm)
X	stop codon
XP	xeroderma pigmentosum
XPA	xeroderma pigmentosum complementation group A protein, DNA repair protein complementing XP-A cells
XPB	xeroderma pigmentosum complementation group B protein, General transcription and DNA repair factor IIH helicase subunit XPB
XPC	xeroderma pigmentosum complementation group C protein, DNA repair protein complementing XP-C cells
XPD	xeroderma pigmentosum complementation group D protein, Excision repair protein ERCC2/XPD
XPE	xeroderma pigmentosum complementation group E protein, DNA damage-binding protein 2
XPF	xeroderma pigmentosum complementation group F protein, DNA repair endonuclease XPF, DNA excision repair protein ERCC-4
XPG	xeroderma pigmentosum complementation group G protein, DNA repair protein complementing XP-G cells, DNA excision repair protein ERCC-5
XP-V	xeroderma pigmentosum variant group, DNA polymerase eta
WES	Whole exome sequencing

SUMMARY

CHAPTER 1 – Introduction and Objectives

1.1 Introduction

1.1.1 Ultraviolet light effects

The ultraviolet light (UV) was one of the first environmental mutagen to be discovered. Since all living organisms are constantly exposed to UV via the sunlight, this discovery generated great interest in the scientific community. Experiments using UV light as a mutagen began in the 1930s, with works in Drosophila and maize. Years later, by using microorganisms as a model it was clearly demonstrated that the maximum absorption spectrum of nucleic acids match with the maximum harmful (mutagenic and killing), effects of UV radiation (DeMarini et al. 2020). At present it is already clearly established the relation of UV light to the development of skin cancer, the most common type of cancer around the world (de Gruijl et al. 1993, Armstrong & Kricker 2001, Matsumura & Ananthaswamy 2002). Therefore, in 2009 the UV radiation was included in the list of carcinogenic substances for humans, group 1 (El Ghissassi et al. 2009).

Based on its spectral range and therefore on the energy it contains, the UV radiation can be divided in three main types: i) UVA, ultraviolet radiation type A or long wave (315-400 nm); ii) UVB, ultraviolet radiation type B or medium wave (280-315 nm) and; iii) UVC, ultraviolet radiation type C or short wave (200-280 nm). However, the relative importance of UV types in terms of exposure it is not directly proportional to their energy, since they are differently filtered by the stratospheric ozone layer that protects our planet, which only filters efficiently wavelengths below 300 nm. Thus, while ozone layer is able to filter completely (~100%) or almost completely (~95%) the UVC and UVB wavelengths, respectively. It is not able to filter efficiently the UVA (~ 5%). Although UVB light corresponds to a small fraction of the sunlight spectrum, this wavelength is much more energetic than UVA and it can penetrate the skin to the epidermis. Meanwhile, despite being less energetic, UVA light is epidemiologically more important as it corresponds to 95% of UV from sunlight and it penetrates more deeply in the skin, reaching the dermis (Anderson & Parrish 1981, Frederick et al. 1989, Cadet et al. 2005, Matsumura & Ananthaswamy 2002, Schuch et al. 2012).

The energy contained in the UV light is able to induce a variety of specific structural changes in the DNA double helix by direct or indirect means (Pfeifer et al. 2005). Direct excitation of the molecule mainly causes cyclobutane pyrimidine dimers

(CPDs) and pyrimidine (6,4) pyrimidone (6-4PPs) photoproducts, both generated at dipyrimidine sites and cause large distortions. In CPDs, the adjacent pyrimidines form a cyclobutane ring through a covalent bond between C5 and C6 carbon atoms of the nitrogenous bases (thymine, cytosine or 5-methylcytosine). In the case of 6-4PP, a non-cyclic covalent bond occurs between the C4 and C6 atoms (Rastogi et al. 2010). Although CPDs are formed in a higher proportion than 6-4PPs, their repair is slower, probably because 6-4PPs causes a stronger distortion, making it more easily detected. Therefore, it is believed that CPDs are mainly responsible for the mutations induced by UV light in mammals (Matsumura & Ananthaswamy 2004, Pfeifer et al. 2005). Unrepaired CPDs and 6-4PPs mainly generate transitions of the C>T type, as well as CC>TT tandem transitions and both are known as UVC and UVB signature mutations (Ziegler et al. 1993, Ikehata & Ono 2011).

Indirect damage into the DNA is generated through photosensitization reactions, which can involve oxygen or other photosensitizer molecules, as NADPH, riboflavin and porphyrin, or even the DNA itself. These result in oxidized bases such as 8-oxoguanine (8oxoG) and thymine glycol (Ravanat et al. 2001, Pfeifer et al. 2005, Rastogi et al. 2010, Sage et al. 2012, Yagura et al. 2017). Apparently, due to the lower reduction potential of guanine, the 8oxoG is the most abundant oxidized base generated; however, it is also rapidly and efficiently repaired (Steenken & Jovanovic 1997). When unrepaired, 8oxoG generates mainly transversions of the G>T (C>A) type, caused by the wrong pairing with an adenine, or A>C transversions, by the erroneous incorporation of oxidized guanine opposite to an adenine (Epe 1991, Cheng et al. 1992, Ikehata & Ono 2011).

Direct damage into the DNA molecule has been traditionally related to UVB wavelength due to its higher energy, while indirect damage has been related to UVA wavelength, because direct absorption of UVA energy by DNA is low (Sutherland & Griffin 1981, Mouret et al. 2006, Nichols & Katiyar 2010). Therefore, the C>T transitions resulting from unrepaired pyrimidine dimers are related to the UVB portion of the solar UV radiation spectrum and the C>A transversions caused by oxidized guanine are attributed mainly to the UVA portion. However, it is controversial whether the mutations caused by UVA are just a consequence of oxidative generated damage, since different studies presented evidence of the CPD and even 6-4PP induction after UVA treatment (Douki et al. 2003, Rochette et al. 2003, Mouret et al. 2006, Schuch et al. 2009, Nichols & Katiyar 2010, Ikehata & Ono 2011, Cortat et al. 2013). Thus, the C>T transitions

have been proposed as a characteristic signature of UV radiation in general, and not specifically of UVB and UVC wavelength (Pfeifer et al. 2005, Sage et al. 2012). This information clearly reveal that the mechanisms by which sunlight induces mutations and skin cancer remains a matter of debate (Sage et al. 2012, Rünger et al. 2012).

1.1.2 Nucleotide Excision Repair (NER): an important and versatile pathway to remove photoproducts from the genome

Genomic DNA is a highly reactive molecule due to its structure, so it is naturally subject to suffer damage, which are chemical alterations of the double helix that challenge the genomic stability and can compromise both replication and transcription. It is estimated that more than 20,000 lesions are induced daily in the DNA as a result of endogenous cell metabolism (Friedberg et al. 2006). Also, physical and chemical exogenous agents such as UV from sunlight, smoke, pollution, and natural or artificial drugs can interact and induce specific damage. Thus, cells have several repair mechanisms to deal with the different types of DNA lesions and try to maintain genomic stability.

Nucleotide excision repair (NER) is a highly conserved and versatile DNA repair pathway, that recognizes and removes a wide range of DNA damage that cause structural distortions in the double helix. NER removes fragments containing lesions in several coordinated steps involving specialized proteins. The first step is damage recognition, followed by unwinding of the double helix around 30 bp at the site of the lesion. Then, cleavage occurs on both sides near the lesion and the damaged chain is excised. After this, the gap is filled by polymerases using the intact chain as a template and ends with a ligation (Sancar 1996, Laat et al. 1999, Menck & Munford 2014).

In eukaryotes, damage recognition can occur by two sub-pathways depending on its locale. DNA damage throughout the genome is recognized by the global genome repair (GGR) sub-pathway. In this case, the primary damage detector is the XPC/HR23B protein complex. The XPE protein is also involved to the damaged DNA binding (DDB) complex and help to improve the efficiency of recognition and removal of some types of lesions, as CPD. On the other hand, DNA lesions in the transcribed strand of actively expressed genes are recognized by the transcription-coupled repair (TCR) sub-pathway. In this case, CSA and CSB proteins recognize lesions that cause blockage of RNA polymerase II during transcription, recruiting NER proteins to

damaged DNA. After damage recognition, the following steps are the same for both sub-pathways (Gillet & Schärer 2006, Menck & Munford 2014).

Once the lesion is recognized, the double helix is unwinded by the XPB (3´-5´) and XPD (5´-3´) helicases, which are part of the transcription factor IIH (TFIIH), a multiprotein complex. Then, the replication protein A (RPA) complex and XPA protein bind to protect the DNA single strand and ensure the stability and proper assembly of proteins involved in repair. A region of around 30 bp containing the lesion is cleaved by the ERCC1-XPF and XPG endonucleases, at the 5´ and the 3´ of the lesion, respectively. The next step is to fill the gap by the coordinated action of the DNA polymerases delta (δ) or epsilon (ε), with the help of several accessory factors of replication (as the RPA, RFC and PCNA proteins) by using the complementary strand as a template and the 3'-hydroxyl extremity generated by the ERCC1-XPF as a primer. Finally, the ligation of the 3´ recently synthesized fragment with the 5´ extremity (generated by XPG) is performed by the DNA ligase I or III and the repair process is completed in an error free manner (A general scheme of the NER mechanism is summarized in figure 1.1.) (Gillet & Schärer 2006, Menck & Munford 2014).

1.1.3 XPC protein

The main function of the XPC protein is to detect a variety of DNA lesions throughout the GGR sub-pathway of NER (Sugasawa et al. 1998). During this recognition process the double helix is thermodynamically destabilized, which is essential for the subsequent recruitment of NER factors in the pre-incision complex (Sugasawa et al. 1998, Min & Pavletich 2007).

The *XPC* gene encodes a protein containing 940 amino acid residues with a predicted mass of 125 kDa, it is located on the short arm of chromosome 3 at position 25.1 (3p25.1, NCBI gene ID:7508) (Legerski et al. 1994). The XPC protein was cloned by Legerski & Peterson in 1992 and then purified by Masutani and co-workers in 1994. Different research revealed that this protein shares limited similarity with the homologous product of the yeast DNA repair gene *RAD4*, and form a complex with the centrin 2 protein and the human homolog of the yeast ubiquitin-domain repair factor RAD23, designated HR23B (58KDa). The Centrin 2 is not essential in the NER repair pathway, but potentiates the damage recognition by XPC protein, while the HR23B is essential in this repair mechanism since it promotes the XPC protein stability. The XPC protein also has high affinity for ssDNA, is highly hydrophilic and has acid and basic

domains, suggesting a possible interaction with chromosomal proteins and DNA. *In vitro*, it was established that the C-terminal portion of the XPC protein is responsible for the interactions with DNA, RAD23B, Centrin 2 and TFIIH, while the N-terminal portion of the protein interacts with XPA (Legerski & Peterson, 1992, Masutani et al. 1994, Ng et al. 2003, Bunick et al. 2006, Nishi et al. 2005).

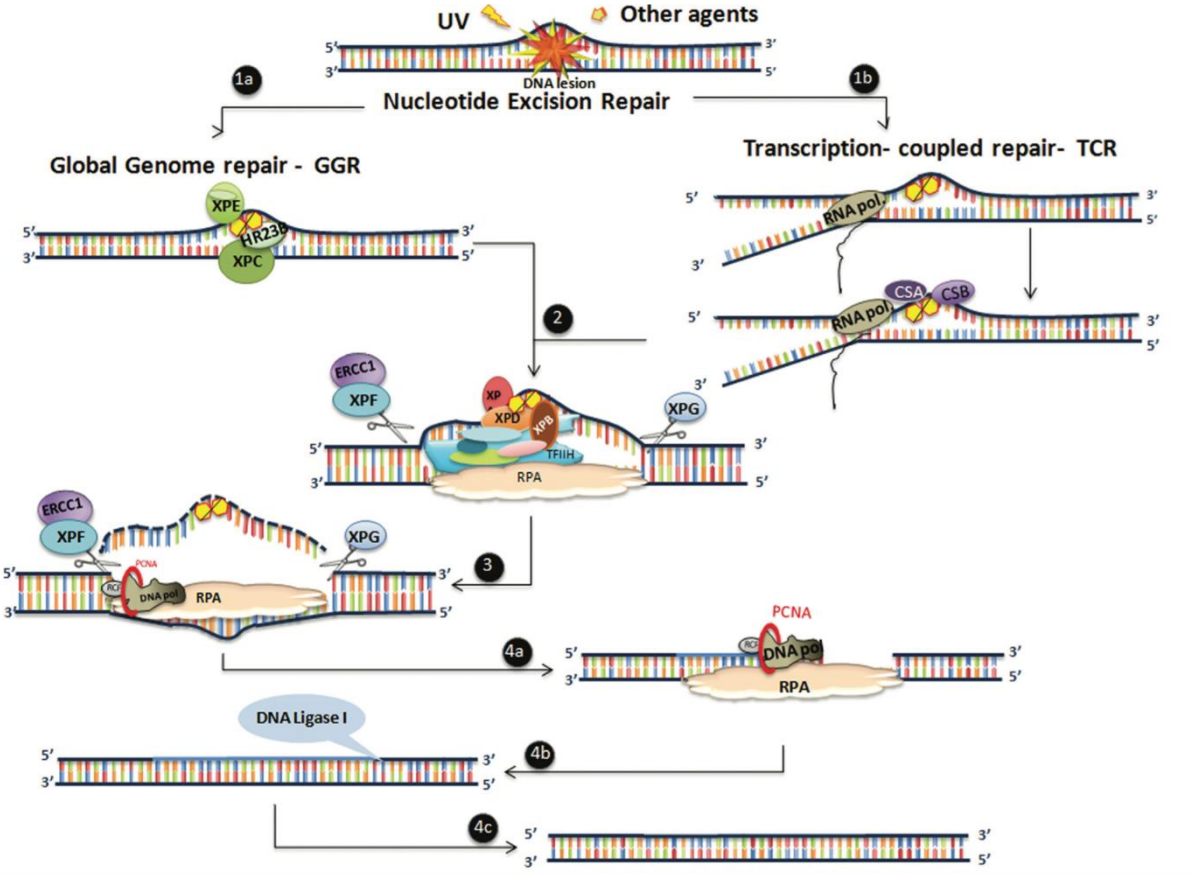


Figure 1.1. Nucleotide excision repair (NER) pathway: Schematic representation of the NER steps for DNA lesion removal in an error-free manner: 1) damage recognition can occur by the Global Genome Repair (GGR) sub-pathway or the Transcription Coupled Repair (TCR) sub-pathway, depending on the damage localization. 2) opening of the double strand of DNA, 3) excision of the region containing the lesion, 4) synthesis of a new error-free strand (figure from Menck & Munford 2014).

Several studies presented evidence of additional roles for the XPC-HR23B protein complex, either as auxiliary protein in repair routes other than NER or in cellular processes not related to DNA repair. For example, XPC-HR23B complex can act as a co-factor in the base excision repair (BER) pathway, since it interacts with the human glycosylases OGG1 and TDG, which participate in the repair of 8oxoG and G/T mismatches, respectively (Shimizu et al. 2003, D´Errico et al. 2006). On the other hand, XPC protein also could act as transcriptional co-activator at the promoters of

inducible genes in the absence of exogenous genotoxic attack and on embryonic stem (ES) cells (Le May et al. 2010, Fong et al. 2011). Additionally, XPC protein also seems to have an important role in cellular metabolic processes, since in his absence the balance between the mitochondrial respiratory complex I and II (CI and CII) is altered: the CI activity is diminished as long as the CII activity is increased (Mori et al. 2017).

For the year 2010, forty-five different mutations had been characterized in the XPC gene sequence of xeroderma pigmentosum (XP) patients (for more details see Li et al. 1993, Cleaver 1999, Soufir et al. 2010). The study model used in this work is the human skin fibroblast XP4PA-SV40 (Daya-Grosjean et al. 1987). This cell line is homozygous for the most common mutation reported, a deletion of two nucleotides (TG) in exon 9 of the *XPC* gene (c.1643_1644delTG) that leads to a frameshift mutation and a premature stop codon (p.Val548AlafsX572), resulting in an inactive protein (Li et al. 1993, Ben Rekaya et al. 2009, Soufir et al. 2010). Thus, the XP4PA-SV40 cell line is deficient on the GGR sub-pathway of the NER repair mechanism. As a result, unrepaired lesions can block the DNA replication and transcription processes, leading to fork collapse, DNA breaks and genomic instability, which may result in cell death and increased mutagenesis, what may result in the development of cancer (Setlow 1974, Pfeifer et al. 2005, Sugasawa 2008, Ikehata & Ono 2011, Menck & Munford 2014, Bowden et al. 2015).

1.1.4 Xeroderma pigmentosum (XP), a syndrome related to defects in the NER repair pathway

The cellular risk of accumulate potentially oncogenic mutations depends on the frequency with which DNA lesions occur and the ability of cells to deal with such lesions. Thus, to safeguard genomic integrity it is necessary to maintain a balance between damage induction and the efficient functioning of the multiple DNA repair pathways, damage tolerance processes and cell cycle checkpoints (Kawanishi et al. 2001; Laval et al. 1998; Peltomäki 2001, Giglia-Mari et al. 2011). In this context, the importance of cellular responses to DNA damage is exemplified by various syndromes caused by mutations in genes that participate in any of the response pathways. Some of the clinical features include neurological abnormalities, premature aging and predisposition to cancer, among others (For detailed information, please refer to: Hoeijmakers 2009, Ghosal & Chen 2013, Bukowska & Karwowski 2018). Defects in the NER system, for example, can lead to a phenotypic diversity of disorders such as

XP, trichothiodystrophy (TTD) and Cockayne syndrome (CS) (Menck & Munford 2014, Bukowska & Karwowski 2018).

XP is a rare autosomal recessive and hereditary disease that promotes high sunlight sensitivity, especially by the ultraviolet (UV) spectrum. Individuals who suffer from XP have strong pigmentation, skin dryness, ten thousand more chance to develop skin cancer, ten times more chance to develop internal tumors and a short lifetime in relation to the unaffected population. Additionally, in approximately 25-30% of cases, patients also exhibit progressive neurological abnormalities and in 40% of cases ophthalmologic pathologies (Taylor 1994, Sugasawa 2008, Cleaver et al. 2009, Menck & Munford 2014, Bowden et al. 2015, Bukowska & Karwowski 2018). This syndrome was originally documented by the dermatologists Ferdinand von Hebra and Moriz Kaposi in 1874, but only in 1968 the molecular defects were identified (Cleaver 1968). By means of cell fusion experiments, seven complementation groups that affect genes encoding proteins (XPA to XPG) associated to the NER pathway were established. In addition, there is a variant group (XPV), that is NER proficient but deficient in DNA polymerase η (pol eta), involved with translesion synthesis (TLS) of DNA damage (De Weerd-Kastelein et al. 1972, Cleaver et al. 2009, DiGiovanna et al. 2012).

XP patients have been identified around the world and in all ethnic groups. The XP estimated incidence in population range from 1:450,000 in Western Europe, 1: 250,000 in USA and 1: 20,000 In Japan and North Africa (Ben Rekaya et al. 2009, Menck & Munford 2014). In Brazil, there are not official estimates, however, in the northwest region of Goiás (municipality of Faina, in the village of Araras), one of the highest frequencies of XP in the world is found, 1:410 inhabitants of Faina (Munford and Castro et al. 2017). This high frequency of the disease is explained by a high rate of consanguineous marriage and endogamy on some of those populations (Ben Rekaya et al. 2009, Munford and Castro et al. 2017). The complementation group C of XP patients (XPC) appears to be the most prevalent type of classical XP in the world. The most common clinical features include strong photosensitivity, early mortality mainly due to a high accumulation of skin cancers in body-exposed areas to sunlight (including basal cell carcinoma, squamous cell carcinoma, and malignant melanoma) and ophthalmological sings. The vast majority of patients do not have neurologic abnormalities so early diagnosis and full protection from sun-exposure are crucial (Soufir et al. 2010).

As previously mentioned the *XPC* gene most common mutation c.1643_1644delTG (p.Val548AlafsX572) result in an incomplete protein unable to recognize cellular DNA damage on the non-transcribed regions of the genome (Li et al. 1993, Soufir et al. 2010), so XPC patients are specifically deficient in GGR, since they are unable to recognize the damage generated in the DNA. This mutation was mainly identified in XPC patients with severe clinical XP symptoms and is the most prevalent type in Western and Southern Europe and North Africa (Ben Rekaya et al. 2009, Soufir et al. 2010). Locals for which haplotype analysis suggest a founder effect for this mutation, estimated in 50 generations or 1250 years (Ben Rekaya et al. 2009, Soufir et al. 2010). This mutation was also detected on USA, Honduras and Brazil (Khan et al. 2006, Leite et al. 2009, Soufir et al. 2010, Santiago et al. 2020).

1.1.5 Analysis of somatic mutation patterns in cancer genomes: from fingerprints to Mutational Signatures

Cancer is an aggressive and silent disease that causes about one in six deaths worldwide, being the second leading cause of death after cardiovascular diseases (American Cancer Society 2018). Efforts and resources have focused on prevention, early diagnosis and to improve treatment. However, its basic biology has not been still fully understood (Greaves 2015). Until now, it has been well established that the accumulation of mutations is related to cancer development and can contribute to the proliferation and survival of cancer cells (Tomasetti et al. 2015).

Several studies have attempted to connect specific agents with particular molecular mutations that occur in the carcinogenic processes (or carcinogenesis) (Vogelstein & Kinzler 1992). The first studies were performed by analyzing somatic mutations in single oncogenes or in tumor suppressor genes, and revealed that a mutational process generated by a particular agent leads to specific molecular fingerprints. The first link was suggested for the G>T mutation at the codon 249 of the p53 gene of hepatocellular carcinomas from patients exposed to aflatoxin B1, but which is rarely found in tumors of other organs (Vogelstein & Kinzler 1992). However, although this strategy revealed valuable information that allowed to clarify some questions related to mutagenesis by establishing fingerprints for diverse mutagens, it is not enough to understand the complexity of the final catalog of mutations observed, since multiple mutational processes could be involved during the tumor development (Petljak and Alexandrov 2016).

Then, the analysis of somatic mutation patterns found in the cancer genome was extended from a single gene to a group of genes. However, it was with the development and widespread use of the next generation sequencing (NGS) technology that the analyses of the exome and whole-genome sequences of tumors became possible. Several studies using this high-throughput technology confirmed patterns previously obtained in pioneering studies within p53 and, naturally, also revealed new patterns of characteristic mutations in different types of cancer, indicating that diverse mutational processes operate on different tissues (Rubin & Green 2009, Petljak and Alexandrov 2016).

A large amount of cancer genomics data has been generated in the last decade by NGS, giving access to the ‘mutational record’ of a cancer genome. However, in addition to accessing and the ability to read the "mutational record", it is necessary to be able to organize and understand these data, which was made possible by the development of advanced mathematical models and computational tools (Petljak and Alexandrov 2016).

The concept of mutational signatures emerged in 2012 and refers to patterns generated by the mutational processes that cells during the tumor development suffer/endure, defined by the mechanisms of DNA damage (originated both from endogenous sources and exogenous factors) and DNA repair involved. Thus, the diversity of somatic mutations on cancers can be explained by one or more mutational signatures depending on the amount of mutational processes, and the strength and the duration of exposure to each one (Nik-Zainal et al. 2012). The mathematical algorithm used to extract mutational signatures, as well as the systematic computational framework that can be freely used to establish them, were published in 2013 (Alexandrov et al. 2013a).

The method developed by Alexandrov (2013a) is based on a matrix factorization algorithm and takes into consideration the six possible types of single base substitutions in a trinucleotide context (*i.e.* the mutated base and the immediately 5’and 3’ sequence context, generating 96 different possible combinations). A total of 30 independent mutational signatures were established by applying this method and after analyzing 12.000 cancer genomic data of 40 different cancer types (Alexandrov et al. 2013b). Recently, this number was updated to 67 signatures (Alexandrov et al. 2020). The patterns of these mutational signatures and information about them, including prevalence in different cancer types and possible etiology, can be found in the

catalogue of somatic mutations in cancer, COSMIC, database (available at: https://cancer.sanger.ac.uk/cosmic/signatures/SBS/). The mutational signatures analysis has been established as a very useful analytical tool that constitutes a breakthrough in the cancer research, since it is used to identify novel mutational signatures and to study the processes involved in different cancers and patients (Bayati et al. 2020).

1.2 Objectives

1.2.1 General objective

- Establish and compare the mutagenicity profile, type and frequency of mutations, induced by UVA and UVB light in cells from XP-C patients.

1.2.2 Specific objectives

- Verify human cell survival after UVA and UVB irradiation;
- Determine the effects of UVA and UVB irradiation on the cell cycle profile of XP-C cells using flow cytometry;
- Evaluate the induction of mutations in XP-C cells submitted to different doses of UVA and UVB irradiation;
- Identify the type and frequency of mutations in XP-C cells using exome sequencing of the isolated clones of irradiated cells.

CHAPTER 2 - Mutagenicity profile induced by UVB light in human xeroderma pigmentosum group C cells

Nathalia Quintero-Ruiz, Camila Corradi, Natalia Cestari Moreno, Tiago Antonio de Souza, Clarissa Ribeiro Reily Rocha, Carlos Frederico Martins Menck*.

DNA repair Laboratory, Department of Microbiology, Institute of Biomedical Sciences, São Paulo University - USP, São Paulo, SP.

*Corresponding author: **E-mail:** cfmmenck@usp.br

2.1 Abstract

UVB wavelength is the most energetic ultraviolet (UV) component of the solar spectra reaching the Earth´s surface. It causes DNA structural damage, such as pyrimidine dimers and oxidized bases, which may generate genome mutations, the major causes of skin cancer. Nucleotide excision repair (NER) is one of the main pathways for the removal of DNA lesions and genome protection against sunlight. In this work, we investigated the mutation spectra induced by UVB in NER deficient and proficient human skin fibroblasts, using the whole exome sequencing (WES) strategy. As a model, we used a XP-C cell line, unable to recognize lesions, and its complemented isogenic control (COMP). Mutations were detected after exome sequencing of randomly selected clones of these cell lines, irradiated or not with UVB light. As expected, the mutation data show a significant mutagenesis increase in irradiated XP-C cells, mainly C>T transitions, but also CC>TT and C>A base substitutions. Remarkably, the C>T and C>A mutations occur mainly in a dipyrimidine sequence context, with preference to the second base. Most mutations were found in the 5´TC dimer. Additionally, although T>N mutations were also significantly increased, it was not possible to relate them directly to pyrimidine dimers. Finally, the mutation spectrum generated in irradiated XP-C cells display a mutational signature related to cutaneous skin cancer, the signature 7 of the COSMIC catalogue. These results indicate that in cells without nucleotide excision repair, a unique dose of UVB contributes to generate the typical spectrum of mutation found in skin cancer tumors. Eventually, the data may be used for comparison with the mutational profiles of skin tumors obtained from XP patients, as well as they may help to understand the mutational process in non-affected individuals.

Key-words: Ultraviolet radiation, Xeroderma Pigmentosum, XP-C, mutagenesis, exome, NGS, mutational signature, nucleotide excision repair.

2.2 Introduction

Sunlight is the major source of ultraviolet (UV) radiation, a ubiquitous carcinogenic physical agent that is harmful to DNA, especially at wavelengths below 300 nm (Ikehata & Ono 2011). UV radiation has been extensively related to skin cancer development, the most common type of cancer in the human population (de Gruijl et al. 1993, Armstrong & Kricker 2001, Matsumura & Ananthaswamy 2002). Therefore, since 2009, it was classified as carcinogenic to humans, group 1 (El Ghissassi et al. 2009). UV radiation is divided into three spectral ranges based on the wavelengths: UVA (315-400 nm), UVB (280-315 nm), and UVC (200-280 nm) (Matsumura & Ananthaswamy 2002). The stratospheric ozone layer blocks all UVC and most part of UVB wavelengths, while UVA crosses the atmosphere with little attenuation (Williamson et al. 2014). Therefore, only UVA and UVB lights are considered biologically relevant, as they are part of the sunlight spectrum that reaches the Earth's surface in a proportion of 95% and 5%, respectively (Taylor 1994, Schuch et al. 2012).

Although proportionally UVB corresponds to a small fraction of the UV spectrum, it is much more energetic than UVA and has potent biological effects because it is directly absorbed by the DNA molecule, causing direct genetic damage (Cadet et al. 2005). Importantly, UVB has been directly related to non-melanoma skin cancers (Pfeifer et al. 2005, Ratushny et al. 2012). Direct excitation of DNA molecule by UVB irradiation mainly causes cyclobutane pyrimidine dimers (CPDs) and pyrimidine (6,4) pyrimidone (6-4PP) photoproducts, both generated at dipyrimidine sites (Rastogi et al. 2010). When unrepaired, CPD and 6-4PP generate basically transitions of the C>T type, as well as CC>TT tandem base substitutions (Ziegler et al. 1993). Thus, these two types of base substitution are known as UV signature mutations (Ikehata & Ono 2011). Similarly, UVB light can cause indirect damage through photosensitization reactions, generating oxidized bases such as 8-oxo-deoxyguanine (8oxo-dG) and thymine glycol (Ravanat et al. 2001, Pfeifer et al. 2005, Rastogi et al. 2010, Sage et al. 2012, Yagura et al. 2017). When unrepaired, 8oxo-dG generates mainly transversions of the G>T (or C>A) type, caused by the wrong pairing with an adenine, or A>C transversions, by the erroneous incorporation of 8oxo-dG opposite to an adenine (Epe 1991, Cheng et al. 1992, Ikehata & Ono 2011).

Cells have a diversity of repair mechanisms to deal with different types of DNA lesions. CPDs and 6-4PP cause structural distortions on DNA and are mainly removed by the nucleotide excision repair (NER) (Sancar 1996, Menck & Munford 2014). This is a versatile pathway that recognizes DNA damage through two sub-pathways depending on its location. The XPC/HR23B protein complex recognizes DNA lesions throughout the genome, in the Global Genome Repair (GGR) sub-pathway. On the other hand, DNA lesions in the transcribed strand of actively expressed genes are recognized by the transcription-coupled repair (TCR) sub-pathway, with DNA damage recognition involving CSA and CSB proteins. After damage recognition, the subsequent steps are similar for GGR and TCR (Menck & Munford 2014). Unrepaired lesions can block the DNA replication and transcription processes, leading to fork collapse, DNA breaks and genomic instability. These processes result in cell death and increased mutagenesis, the latter intrinsically related to cancer development (Setlow 1974, Pfeifer et al. 2005, Sugasawa 2008, Ikehata & Ono 2011, Menck & Munford 2014, Bowden et al. 2015).

Deficiency in one of some proteins (XPA to XPG) involved in the NER repair pathway has been linked to the xeroderma pigmentosum (XP) syndrome, a rare autosomal recessive hereditary disease that demonstrates, dramatically, the straight relationship between unrepaired DNA damage, accumulation of mutations and cancer (Cleaver 1968, Ravanat et al. 2001, Menck & Munford 2014). In addition, there is a variant group (XPV), deficient in DNA polymerase η (Pol eta), responsible for translesion synthesis (TLS) (Lehmann et al. 1975, Munford et al. 2017). Individuals who suffer from XP have high sunlight sensitivity with ten thousand times increased risk to develop skin cancer, ten times more risk to develop internal tumors and a short lifetime in relation to unaffected population (Taylor 1994, Sugasawa 2008, Cleaver et al. 2009, Menck & Munford 2014, Bowden et al. 2015).

In this work, we selected as a model a XPC protein deficient cell line, lacking GGR, and thus DNA lesion removal is reduced to approximately 20-30% of normal levels, and an isogenic counterpart to study UVB induced mutagenicity (Cleaver 1986). Importantly, XP-C patients carrying exactly this mutated XPC allele are highly prevalent in Europe, USA and North Africa (Stary & Sarasin 2002, Soufir et al. 2010), and probably also in Brazil (Leite et al. 2009, LP Castro, personal communication). Base substitutions in the exome of clones obtained from both cell lines, differing only by their repair (GGR) capacity, were detected as single nucleotide variants (SNVs),

that were further compared to pre-established motifs and sequence context was analyzed by the tri-nucleotide classification proposed by Alexandrov and collaborators (2013a). Then, the mutation spectra were compared with the pre-established signatures in the catalogue of somatic mutations in cancer, COSMIC, database (available at: https://cancer.sanger.ac.uk/cosmic/signatures/SBS/). Thus, this research aims to contribute to the understanding of the importance of DNA repair proteins in the removal of UVB-induced damage and skin carcinogenesis.

2.3 Material and Methods

2.3.1 Cell lineages

The XP4PA (XP-C) cell line immortalized by SV40 virus (Daya-Grosjean et al. 1987) was used. This is a NER deficient human skin fibroblast that carries a ΔTG mutation in exon 9 of the XPC gene (c1643-1644delTG), which translates in an inactive truncated protein (p.Val548AlafsX25) (Li et al. 1993, Soufir et al. 2010). As isogenic control, the same cell line was complemented with XPC gene by a lentivirus-based strategy, COMP cells. The XP-C cells were kindly provided by Drs. Alain Sarasin and Anne Stary (Institut Gustave Roussy, France).

2.3.2 Culture conditions

Cells were grown on Dulbecco's Modified Eagle Medium High Glucose (DMEM, LGC Biotechnologies, Cotia, SP, Brazil) supplemented with 10% fetal bovine serum (FBS, Cultilab, Campinas, SP, Brazil) and 1% antibiotic/antimycotic solution (0.1 mg/ml penicillin, 0.1 mg/ml streptomycin and 0.25 mg/ml fungizone) (Life Technologies, Carlsbad, CA, USA). Cultures were maintained at 37°C in a humid atmosphere and 5% of CO2. The cells grew adhered to the surface of plastic bottles and were picked up in periods of 3 to 4 days, as they reached confluence.

2.3.3 Clonal expansion

The cloning strategy used was previously established in our laboratory and allows the detection of mutations induced on the first round of replication (Moreno et al. 2020). Here, an additional cloning step was included before the selection of clones for exome sequencing in order to obtain a more homogeneous population, and consequently, a more accurate detection of mutational signatures induced by UVB light (Supplementary figure S2.1). Briefly, cells were plated in a very low confluence, 500

cells in 100 mm Petri dishes, the next day cells were submitted to UVB irradiation and allowed to grow by two weeks for the formation of isolated colonies. Individual clones were randomly selected and transferred manually to 96 multiwell plates, 1 colony per well. Upon reaching confluence, cells were expanded consecutively into larger plates; this was done successively until obtaining two 100 mm Petri dishes (approximately 45 days in total), always renewing the medium every 3 to 4 days. One plate was used for genomic DNA extraction using the Blood and Cell Culture DNA Mini Kit (Qiagen, Hilden, Nordrhein-Westfalen, Germany. Cat number: 13323), the other plate was frozen in liquid nitrogen.

2.3.4 Irradiation of cells with UVB light

All UV exposures were performed under controlled conditions, on adherent cells and in a time window between 16 to 24 h after plating. Once the culture medium was discarded, cells were washed with PBS and irradiated in PBS. Non-irradiated control plates were maintained in the dark during UV irradiation and submitted to the same process. Irradiation was performed using a Philips UVB Vilber Loumart T 15M 15 W lamp in an average dose rate of 3.4 $J/m^2/s$, associated with a polycarbonate plate to eliminate contamination with UVC light. The intensity of lamp emission was measured at each experiment with a VLX 3 W UV radiometer (Vilber Lourmat, Torcy, France) and the exposure time adjusted, so the cells received the desired dose of UVB light.

2.3.5 Clonogenic assay

Differences in the viability of cells were measured by selecting cells that have maintained the ability to form colonies, after treatment (Franken et al. 2006, Rafehi et al. 2011). Cells were plated at a high dilution ($1.5x10^3$ cells in 60 mm diameter plates), 16 h later were irradiated or not and maintained for two weeks in DMEM medium with medium replacement every 4 or 5 days, until the formation of isolated colonies with at least 30 cells each. The colonies were then fixed with 10% formaldehyde, stained with 1% violet crystal and counted. The survival percentage was estimated in relation to the non-irradiated control group.

2.3.6 Cell cycle, γH2Ax phosphorylation and sub-G1 content

These endpoints were investigated by using flow cytometry, as previously described (Quinet et al. 2014). In general, $1x10^5$ cells were plated in 60 mm diameter

plates, irradiated the next day and harvested 24 or 72 h later. Both detached, dying cells, and trypsinized adherent cells were collected and centrifuged. The cell pellet was fixed with formaldehyde 1% on ice, washed with PBS, resuspended in cold ethanol 70% and stored at -20°C for at least 24 h. Samples were then centrifuged and permeabilized with PBS-T-BSA buffer (0.2% Triton X- 100, 1% bovine serum albumin, BSA, Sigma–Aldrich, San Luis, Missouri, USA, in PBS) and incubated overnight at 4ºC with the antibody for γH2AX (05-636 Millipore, at 1:500 dilution). Thereafter, samples were washed with PBS-T-BSA buffer and incubated in the dark for 1 h at room temperature with the secondary antibody (anti-mouse fluorescein iso-thiocyanate (FITC) antibody, Sigma–Aldrich, at 1:200 dilution). Samples were washed again and resuspended in propidium iodide (PI) solution (200 µg/mL RNAse A, Invitrogen-Life Technologies, 20 µg/ml PI, 0.1% (v/v) Triton X-100 in filtered PBS) and stirred for at least 30 min at room temperature, avoiding the incidence of direct light. Finally, samples were washed with filtered PBS, centrifuged and resuspended in filtered PBS.

2.3.7 DNA photoproduct quantification

The average number of CPD photoproducts generated by a UVB lamp was quantified from the relative amounts of supercoiled (FI) and circle (FII) plasmid DNA forms through densitometry analysis (ImageJ 1.51r, National Institutes of Health, USA. http://imagej.nih.gov/ij, Schneider et al. 2012); after treatment with the T4 bacteriophage endonuclease V (T4-endo V, that mainly recognizes CPD lesions) and separation by 0.8% agarose gel electrophoresis. The pCMut plasmid (200 ng per sample) was pre-incubated with T4-endo V (produced in this laboratory) at 37ºC for 60 min in a final volume of 20 µL. The enzyme was previously tested up to saturation and was used in amounts where no nonspecific cleavage was observed. The number of enzyme sensitive sites per Kbp of plasmid DNA was calculated, assuming a Poisson distribution, by the following equation: X = -ln (1.4 x FI/1.4 x FI + FII)/1.8; where FI and FII represents the intensity of fluorescence measured in the supercoiled and relaxed DNA bands, respectively; 1.4 is a factor employed for correcting the increased fluorescence of ethidium bromide when bound to the relaxed form compared to the supercoiled form, and 1.8 is the pCMUT vector size in Kbp (Schuch et al. 2009).

2.3.8 Slot-Blot assay

Direct detection of pyrimidine dimers (CPD) formation after UVB irradiation was

measured immunologically by slot-blot assay as previously described (Schuch et al. 2009, Yu et al. 2015). Briefly, $1x10^6$ cells were plated in 100 mm diameter plates, irradiated 16 to 24 h later and collected at different recovery times after irradiation. Total genomic DNA was extracted and quantified by spectrophotometry (NanoDropTM 1000, Thermo Fisher Scientific, Waltham, Massachusetts, USA). The DNA was denatured using NaOH (Merck, Darmstadt, Germany) and EDTA (Sigma-Aldrich) at final concentrations of 0.4 M and 10 mM, respectively, and boiled for 10 min at 70ºC. Denatured samples were neutralized with equal volume of cold ammonium acetate 2 M (Sigma-Aldrich), and fixed with ammonium acetate 1 M onto a pre-wetted (in ammonium acetate 1 M) positive nitrocellulose membrane (Amercham Hybond N+, GE Healtcare, Chicago, USA), in successive steps using a vacuum pump and a Slot-Blot apparatus (Omniphor, San Jose, CA, USA), at the rate of 200 ng per well. Subsequently, the membrane was incubated in 5xSSC buffer (750 mM NaCl, 75 mM sodium citrate, Sigma-Aldrich) for 15 min, dried at room temperature, baked for 2 h at 80°C and blocked overnight at 4ºC in 5% (w/v) non-fat milk prepared in PBS-T (0.1% Tween 20) under constant shaking. Thereafter, the membrane is well washed in PBS-T and incubated for 2 h with the monoclonal primary antibody TDM-2 (anti- CPD, Cosmo Bio Co., Ltd, Japan at 1:2000) at room temperature. After successive washes, the membrane was incubated for 2 h with the secondary antibody (Horseradish peroxidase-conjugated anti-mouse-HRP, GE Healthcare, USA at 1:2000 dilution). Finally, after successive washes with PBS-T and a PBS the membrane was revealed with a chemiluminescence reagent (LuminataTM Forte Western HRP substrate, Millipore Corporation, Billerica, MA, USA) that generates a light signal when oxidized by the enzyme HRP (conjugated to the secondary antibody). The membrane was analyzed using a chemiluminescence detector system (Alliance Q9, Uvitec, Cambridge, England, UK). The percentage of CPDs was estimated in relation to UV-irradiated cells by bands analysis using ImageJ 1.51r (National Institutes of Health, USA. http://imagej.nih.gov/ij, Schneider et al. 2012).

2.3.9 Whole exome sequencing and analyses

Library preparation and exon capture was performed using the Agilent's SureSelect QXT Human All Exon v6 kit (Agilent Technologies, Santa Clara, California, USA), starting from 5 µg of genomic DNA and following the manufacturer's specifications. Libraries were sequenced in paired-end mode on the Illumina NextSeq

or HiSeq platforms (Illumina, San Diego, California, USA), for details see Supplementary table S2.1. The data analysis was performed based on Genome Analysis Toolkit (GATK) best Practices for exome variant calling (DePristo et al. 2011, Van der Auwera et al. 2013). Reads were mapped to the UCSC hg19/ucsc.hg19 reference genome using BWA (Li & Durbin 2010). Duplicates were marked using Picard, the remaining reads were local realigned around indels, and BaseRecalibrator from GATK toolkit was used to correct sequencing errors and experimental artifacts in base quality scores. Subsequently, the identification of raw variants relative to the reference genome was performed by HaplotipeCaller using the default parameters mode, this caller is based in a strategy of local de-novo assembly of haplotypes in an active region. Additionally, we refine the call set of variants to reduce false positives by variant quality score recalibration (VQSR). The annotation of SNVs was made by ANNOVAR (Wang et al. 2010) using RefSeq (O´leary et al. 2016), and in order to restrict the analysis to regions enriched in exome sequencing, an additional filter for SNVs annotated as “exonic” or “splicing” was applied. We perform a call of exclusive variants for each clone following the strategy proposed by Moreno et al. (2020) and including a pre-cloning stage, as described above. For this, comparisons between all clones of the experiment (clones irradiated or not with UVB light) were performed using VCFTools (Danecek et al. 2011). As a general rule, variants detected in two or more independent clones were discarded in order to reduce the false positive SNVs. The exploratory analysis of the exclusive variants and the general patterns of nucleotide changes, as well as its associated motif sequences were performed using WOLAND (https://github.com/tiagoantonio/woland, Souza et al. in preparation). The somatic spectra analyses of filtered exclusive point mutations in a tri-nucleotide context as well as the mutational signature analyses were performed using the SomaticSignatures (Gehring et al. 2015), an implementation of the NMR method in R (Nik-Zainal et al. 2012, Alexandrov et al. 2013a, 2013b). Finally, the deconstructSigs package was used to evaluate the published signatures from Sanger COSMIC in our data (Rosenthal et al. 2016). The statistics of alignment are detailed in Supplementary table S2.2.

Also an independent experiment using a lower dose of UVB was performed employing another commercial human exome capture platform, the Illumina’s Nextera Rapid Capture Exome protocol (San Diego, California, USA), the amplification and sequencing of this library was performed in paired-end mode (2x150 bp) using the Nextseq platform (Illumina) and the data analysis was performed as described above.

2.3.9 Statistical analyses

To infer group differences under the null hypothesis that mean value of the experimental group is not greater than the mean value of the control group we employed a non-parametric permutation test. The mean values of both groups generated using Monte-Carlo simulated data from a total of 4999 permutations, were compared and the p-values calculated. Considering a one-sided test, we determine P-values <0.001 as highly statistically significant (***), 0.01 < P-values as high statistically significant (**), P-values < 0.05 as statistically significant (*) and P-values > 0.05 as non-significant (ns).

2.4 Results

2.4.1 Clone selection and cell response to UVB treatment

This work was performed with isogenic cell lines and selected clones for each cell line, to obtain a homogeneous population that allows a more accurate identification of mutations generated by UVB irradiation and discard polymorphism. All the results were obtained using these clones, identified as XP-C and COMP.

Initially, as these were novel cloned cell lines, they were characterized for their responses to UVB irradiation. The sensitivity of XP-C and COMP cell lines upon UVB light was evaluated by the survival clonogenic assay (Figure 2.1a), and as expected, XP-C cells were clearly more sensitive, when compared to control cells. From these results, the UVB-irradiation dose of 120 J/m^2 was chosen for mutagenesis experiments, taking into account a cell survival rate near to 20% for XP-C cells, since this threshold increases the possibility of detecting mutations as well as allows the recovery of colonies required. The effects of UVB light on induction of apoptosis (sub-G1), genotoxic stress (γH2AX) and cell cycle progression were evaluated by flow cytometry. An increase in genotoxic stress (Figure 2.1b) in XP-C cells was observed by the increase of H2AX phosphorylation levels. These cells also showed a significant increase in apoptosis levels (Figure 2.1c), as detected by the sub-G1 content, that indicates nuclei fragmentation. Finally, UVB light causes partial S/G2 arrest in XP-C cell cycle (Figure 2.1d), consistent with previous observations (Quinet et al. 2014, 2016).

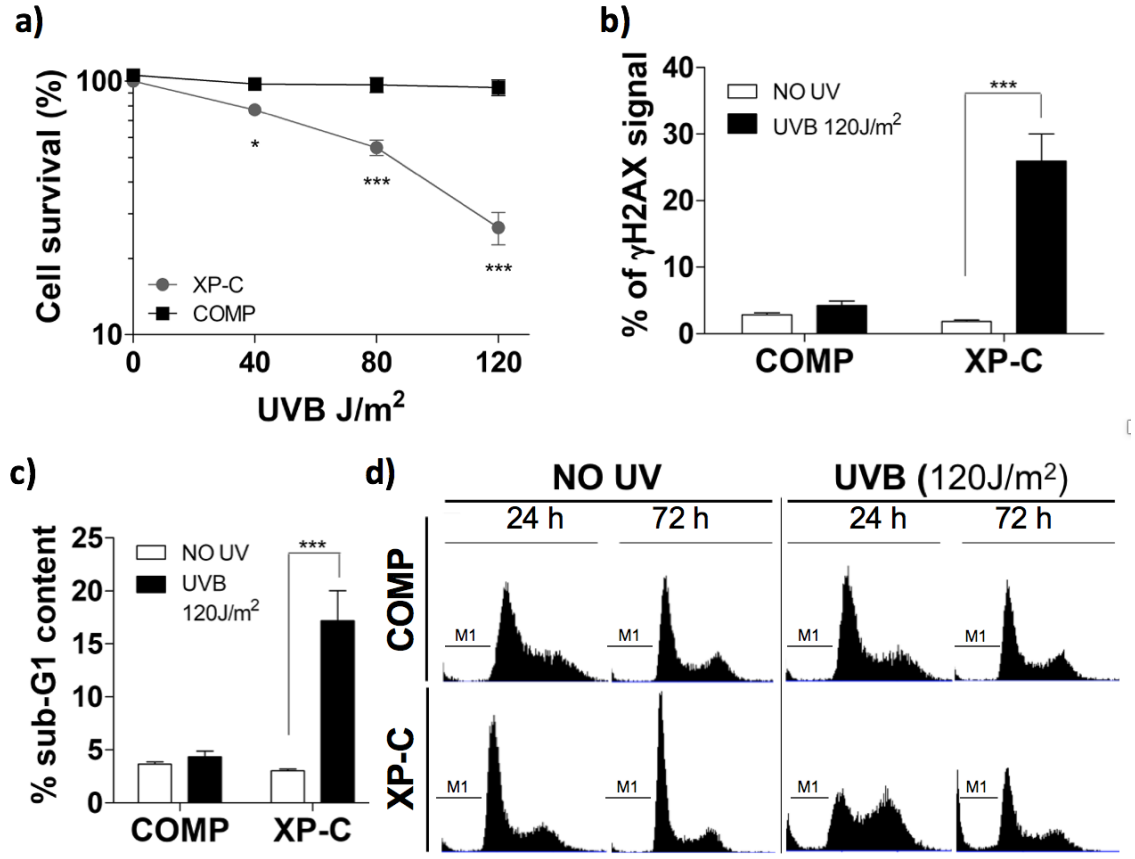


Figure 2.1 Response of the cell lines to UVB irradiation. a) Clonogenic assay. Dose response curves of cells exposed to UVB light; **b)** phosphorylation of H2AX histone (percentage) 24 h after UVB light exposure **c)** sub-G1 content (indicative of apoptosis) induced 72 h after UVB irradiation; **d)** cell cycle response to UVB irradiation: representative histograms of cell cycle progression after 24 and 72 h of irradiation. M1 indicates sub-G1 cell population.

CPD lesions resulting from exposure to UVB radiation were also measured from a combination of enzymatic and immunological quantification approaches. First, we investigated the induction of lesions in the pCMut plasmid DNA irradiated with increasing doses of UVB, and treated with T4-endo V repair enzyme that recognizes and nicks at the CPD lesions. To quantify the number of CPDs, pCMut samples were analyzed by electrophoresis migration in agarose gels to discriminate the relative amounts of super-coiled (FI), from open-circle relaxed DNA (FII), carrying nicks caused as consequence of single-strand breaks (SSB) or enzymatic treatment (Supplementary figure S2.2a). The CPD levels increased dose dependently, exhibiting a direct correlation between the dose and the amount of detecting CPDs (R^2=0.998). This calibration curve reveals that plasmid DNA molecules irradiated 120 J/m^2, generate approximately 141 CPDs per 10^6 bp, while few or no direct breaks, SSB, where detected after increasing doses of UVB treatment. The Slot-Blot assay showed

that at different doses of irradiation the amount of CPDs detected in the assay is proportional to the amount of DNA used. Also, it was evident that using the same amount of DNA and two doses of UV-B light, 40 and 120 J/m^2, the amount of CPDs detected *in vitro* and in the cellular genome is equivalent (Supplementary figure S2.2b). Thus, we can consider that the number of CPDs generated by UVB light *in vitro* in pCMut plasmid is equivalent to those generated *in vivo* when irradiating cell culture. Considering these data, 120 J/m^2 of UVB generates around 8.46x10^5 CPDs per diploid cell.

2.4.2 UVB light induced point mutations in NER deficient fibroblast

To explore the type and frequency of mutations generated by UVB light the cloned populations of XP-C and COMP cell lines were plated in a very low confluence, irradiated or not with 120 J/m^2 of UVB, and subjected to a new round of clonal expansion process. Clones of each condition were randomly selected for whole-exome sequencing as follows; twelve clones of XP-C, six non irradiated and six irradiated, as well as ten clones of COMP, five non irradiated and five irradiated, for a total of 22 clones in the experiment, for details see Supplementary Table S2.1.

UVB light irradiation increased the mutation frequency (single nucleotide variants - SNVs) in XP-C cells, while for the control cells, no significant increase was detected (Table 2.1, Figure 2.2a). Interestingly, just the absence of XPC protein led to a statistically significant increase of the mutation frequency in non-irradiated cells. In both UVB-irradiated cell lines, there were more transitions than transversions, as expected (Figure 2.2b). These results highlight the mutagenic capacity of UVB light. In fact, in an independent experiment, performed with the Nextera rapid capture exome kit, and a dose of 40 J/m^2 of UVB light, a significant increase of point mutations in XP-C clones was also observed (Supplementary figure S2.3).

Table 2.1 UVB light increased the spontaneous mutation rate per base per replication of XP-C cells.

Cell line	Treatment	XPC protein	Capture platform	SNVs per clone
COMP	NO UV	WT	Agilent´s SureSelect QXT ≈38 Mb	36
	UVB 120 J/M^2	WT		32
XP-C	NO UV	Deficient		58
	UVB 120 J/M^2	Deficient		307

SNVs per clone: mean of the Single Nucleotide Variants (SNVs) for sequenced exome of 5 or 6 independent clones.

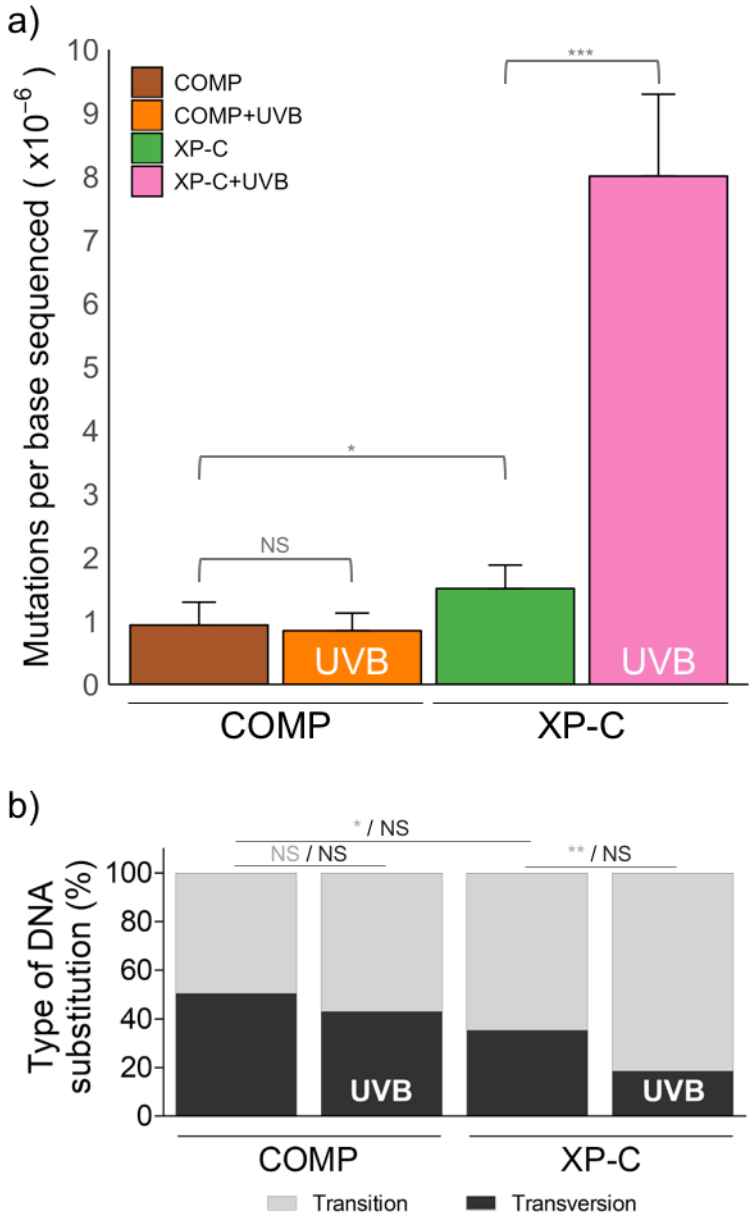


Figure 2.2 UVB light increases point mutations in NER deficient fibroblast human cells. **a)** Unique point mutations within exons and splicing sites per sequenced base (n = 5 or 6 clones per group), after UVB irradiation. The data represent mean and standard deviation (SD). **b)** Contribution of transitions and transversions to total mutagenesis detected, UVB-irradiated cells (120 J/m^2) are indicated. Statistically significant differences were calculated by non-parametric permutation tests: $P < 0.001$ (***), $P < 0.01$ (**), $P < 0.05$ (*), $P > 0.05$ (NS).

2.4.3 C>T transition is the main type of point mutation induced by UVB light

To understand how UVB light influence the increase in mutagenesis, we explored the changes generated in the six possible types of base substitutions: C:G>A:T, C:G>G:C, C:G>T:A, T:A>A:T, T:A<C:G and T:A>G:C; for which we will refer by the pyrimidine of the canonic Watson-Crick base pair. Basically, this wavelength

increased all types of base substitutions (except for C>G) in XP-C cells when compared to non-irradiated cells; being significant for the C>T and T>C transitions, as well as the C>A, T>A and T>G transversions, with an increased rate of 8.1, 1.6, 3.8, 2.9 and 2.9 times, respectively (Figure 2.3). This result makes it clear how UVB light preferentially generates mutations of the C>T type, as previously reported (Ikehata and Ono 2011). Curiously, the increment in the basal mutation rate of XPC protein deficient cells compared to its isogenic control seems to be explained just by an increase in the C>T (2.3 times) transitions. Again, a similar result was obtained when used a lower dose of UVB light (Supplementary figure S2.4).

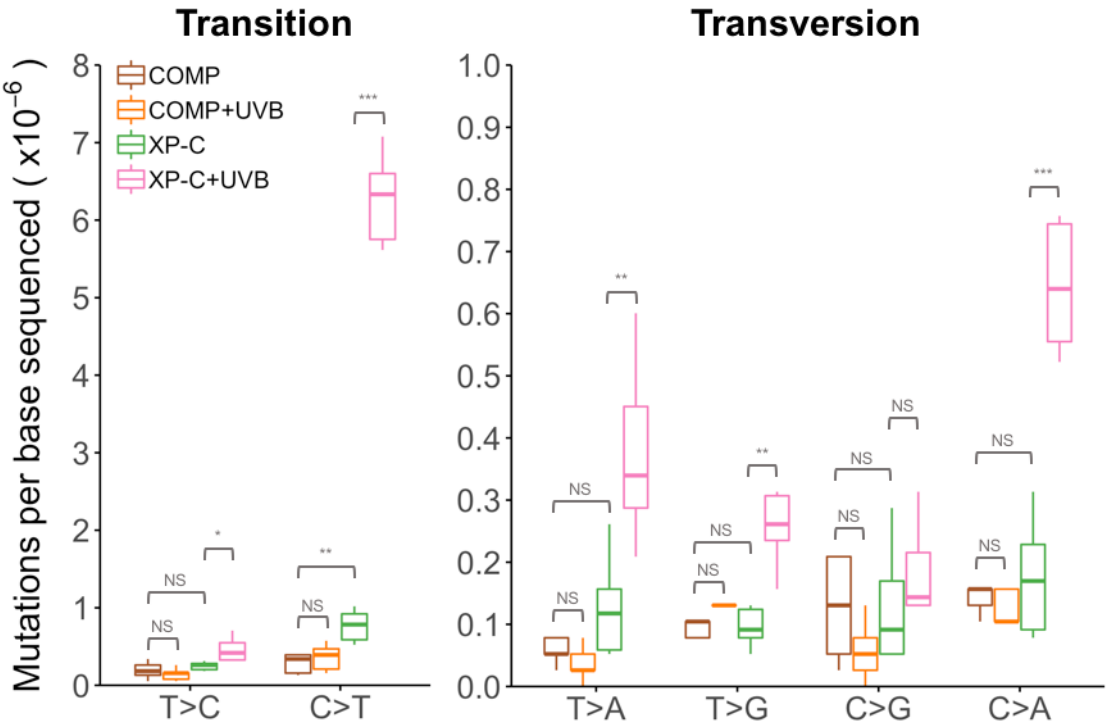


Figure 2.3 Analyses of the point mutations induced, according to the nature of the base substitution type: transitions or transversions. The data represent the mean and standard deviation (SD) of exclusive substitutions found within exonic and splicing sites in the sequenced clones for each condition. Cells were irradiated with 0 or 120 J/m^2 of UVB light. Statistically significant differences were calculated by non-parametric permutation tests: P < 0.001 (***), P < 0.01 (**), P < 0.05 (*), P > 0.05 (NS).

Additionally, the analysis of the somatic spectra of point nucleotide mutations in the tri-nucleotide contexts was also performed. This analysis includes the mutated base and the immediately 5´and 3´ bases information, generating 96 possible context sequences for substitutions (Alexandrov et al. 2013a), revealing preferences for a particular DNA sequence where mutations are more frequent (Figure 2.4). Non-irradiated COMP clones (COMP, figure 2.4a) have a homogeneous spectrum between the 96 possible tri-nucleotide combinations, without preference for a specific sequence

context. Although there is no significant increase in the frequency of any type of base substitutions in COMP cells after UVB irradiation (COMP+UVB), a small variation in the mutation spectra can be evidenced in this analysis, when compared to non irradiated clones (Figure 2.4b), particularly, for dipyrimidine sites in C>T (CCY), T>C (CTG), C>A (CCG) and T>G (CTC) base substitutions. On the other hand, non irradiated XPC clones (XP-C) have a similar homogeneous spectrum as observed for COMP clones without large changes between both cell lines, except for a flat, but clear, increase in the C>T point mutations, without particular preference for a context sequence was observed (Figure 2.4c). However, for XPC clones irradiated with UVB light (XP-C+UVB) there are clear increases, particularly evidenced on the C>T transition, occurring mainly in CCN, TCN, ACY and GCY tri-nucleotide context (Figure 2.4d). Interestingly, all of these are sequences correspond to dipyrimidine sites, where pyrimidine dimers could have been generated during UVB irradiation. For the other types of base substitutions, a context of preference for the occurrence of these mutations was not detected, since the spectrum is flat among all possible combinations of tri-nucleotides. Maybe, the huge difference in the proportion of C>T mutations with respect to the others base point changes, ends up masking these effects.

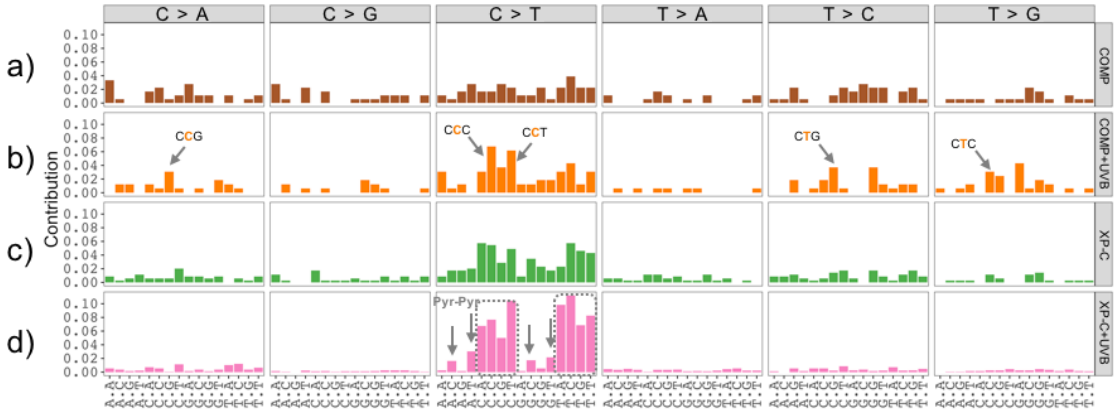


Figure 2.4 Somatic mutation spectra of irradiated or non-irradiated cells. Each group of clones was merged according to treatments, this was used to characterize the somatic spectra of trinucleotide motifs using the SomaticSignatures package, an implementation in R. **a)** Non irradiated COMP cell line clones; **b)** UVB irradiated COMP cell line clones; **c)** Non irradiated XP-C cell line clones; **d)** UVB irradiated XP-C cell line clones. Trinucleotide contribution was calculated by its frequency. The most frequent mutated trinucleotides were highlighted by arrows. Irradiation dose: 120 J/m^2 of UVB light.

2.4.4 UVB light generates C>T point mutations mainly on the second base of potential pyrimidine dimers sites

C>T mutations were further investigated. First, we considered the occurrence

of tandem CC>TT transitions, since the C>T and CC>TT mutations are known as UVC and UVB signature mutations (Ikehata & Ono 2011). Effectively, in UVB irradiated XPC clones a 7.4 fold increase was detected in the frequency of CC>TT mutations when compared with the non-irradiated ones. However, this increment was not observed in irradiated clones of the control cell line (Figure 2.5a). Then, in order to establish the contribution of pyrimidine dimers on mutations caused by UVB light, an analysis of the sequence contexts where point mutations occurred was performed. For non-irradiated XPC and COMP clones about 70-85% of mutations occur within the dipyrimidine motif, YY. On XP-C clones, this percentage increases significantly to 97.6% after UVB irradiation. This means that only 2.4% of C>T transitions in UVB irradiated XP-C clones are out of possible sites of pyrimidine dimers. Whereas for irradiated control clones, COMP, approximately 22% of these mutations are outside of these sites (Figure 2.5b). Interestingly, none of the potential CG sites (where C methylation is potentially increased) appears to be preferentially mutated by UVB irradiation (Figure 2.5b). Interestingly, for irradiated XP-C cells, the C>T mutations were basically found in the second pyrimidine of a potential dimer (Figure 2.5c), and mainly at the TC dimer, but not at CC. However, in control irradiated clones (COMP+UVB) and in non-irradiated XP-C clones the C>T mutations occur preferentially in the CC dimer (Figure 2.5c).

Additionally, a graphical representation (logo) of the context sequences where the C>T mutations occur was generated grouping clones by cell line and treatment (Figure 2.6). For this purpose, we used the pLogo generator, which scales the residue heights relative to their statistical significance and use exome as background (available at: https://plogo.uconn.edu/, O´Shea et al. 2013). In non-irradiated COMP cells, no bias was observed in the sequence context. However, for non-irradiated XP-C cells a small but statistically significant enrichment bias of the C base, at the -1 and +3 positions, was observed. On the other hand, in UVB-irradiated COMP cells a bias for the C base at the -1 position was evidenced, what matches the previous analysis. However, for the UVB-irradiated XP-C clones a more complex sequence context for the mutations were observed: a clear increase in the preference for T nucleotide, followed by C, at the -1, -2 and +1 positions. Also, a preference for C nucleotide at -3 and +3 positions was observed. Thus, a bias to a pyrimidine-rich sequence context (CYYCYNC) was clearly evidenced for the C>T mutations in UVB-irradiated XP-C cells.

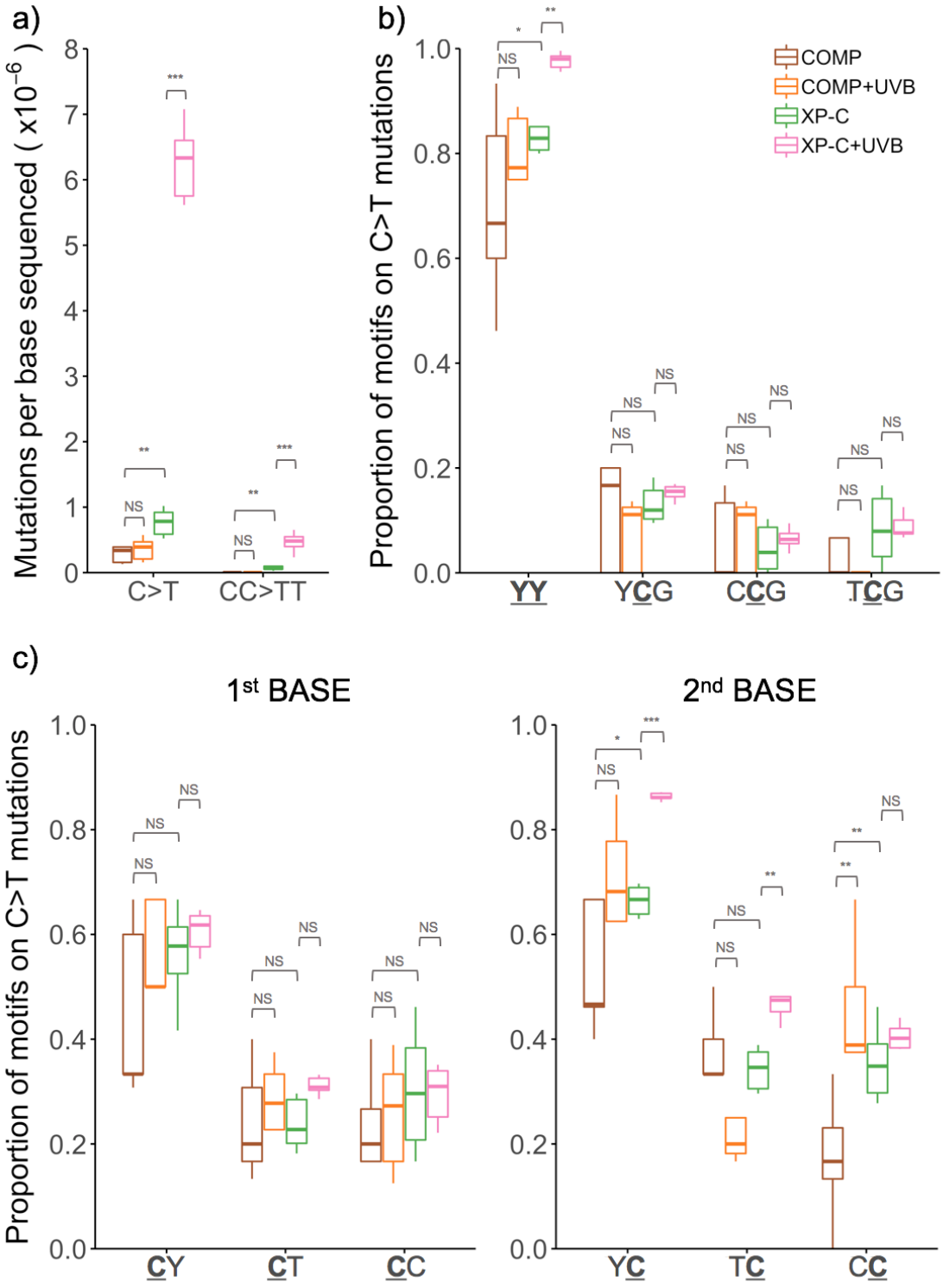


Figure 2.5 Nucleotide changes and mutagen-associated motif patterns in C>T point mutations. a) Exploratory analyses of C>T point mutation and CC>TT tandem mutation after UVB exposure. **b)** Analyses of C>T mutations occurring at dipyrimidine sites (YY) and at dipyrimidine sites with a potentially methylated cytosine (YCG). **c)** Analyse of C>T mutations at the first or the second pyrimidine of a dipyrimidine site. Irradiation dose: UVB 120 J/m^2. The IUPAC code was used to represent the nucleotides; R = A or G; Y = C or T and N = A, C, G or T. Statistically significant differences were calculated by non-parametric permutation tests: P < 0.001 (***), P < 0.01 (**), P < 0.05 (*), P > 0.05 (NS).

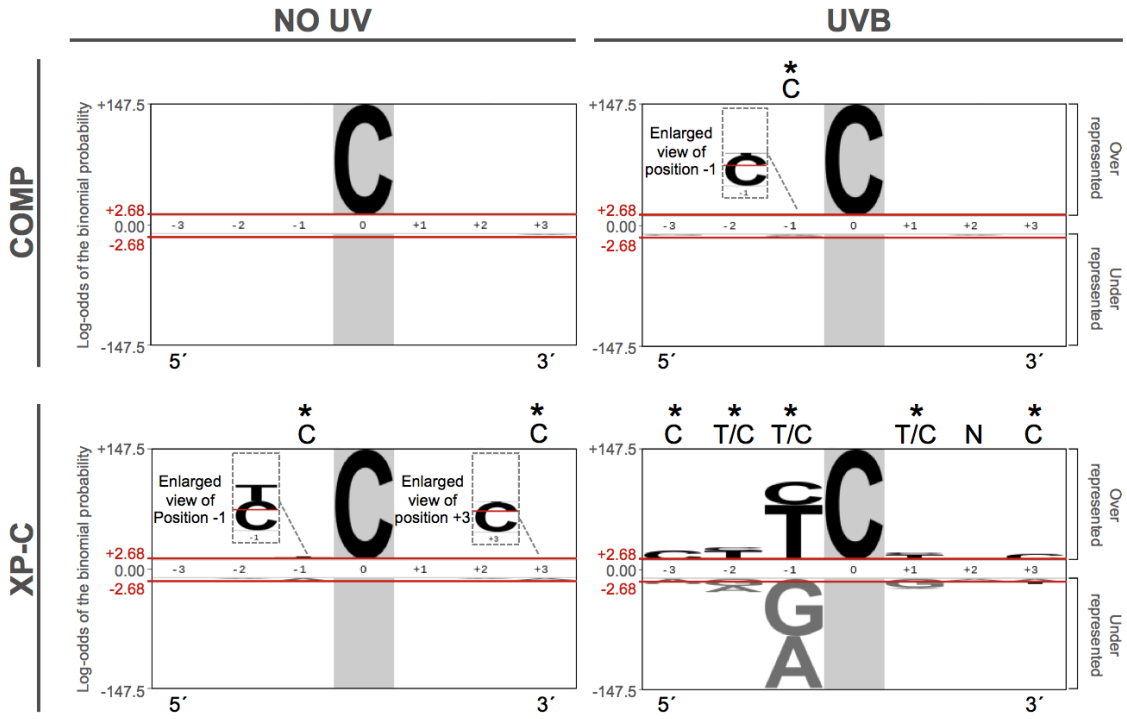


Figure 2.6. Sequence contexts for the C>T mutations. Sequence context logo of all C>T mutations grouped by cell line and treatment, adjusted by strand to mutations at C. The -pLogo- web software (https://plogo.uconn.edu/) was used to generate the graphic representation of probability logo. The log-odds of the binomial probability for each nucleotide at each position with respect to the human exome (used as background) are illustrated. The pLogo scales the nucleotide residue heights proportional to their statistical significance, and not to their frequency, over- and underrepresented nucleotides are drawn above and below the x-axis, respectively. The "fixed" position (that allows the use of conditional probabilities) is highlighted by a vertical grey bar. Red horizontal bars denote the threshold of the Bonferroni-corrected statistical significance values (p=0.05), significant enrichment of a particular nucleotide in a specific position is highlighted (*) at the top of the logo and the closer the nucleotides are drawn to the x-axis, the higher is its statistical significance. Purine nucleotides are presented in gray and pyrimidine nucleotides in black. Number of aligned foreground (n(fg)) sequences used to generate the image: COMP (n(fg)=54, COMP+UVB (n(fg)=69, XP-C (n(fg)=176, XP-C+UVB (n(fg)=1393. Number of background (n(bg)) 4096 sequences used.

Finally, to check if those potential lesions had any bias for transcription strands, we performed a comparison of C>T mutations occurring inside the YY motif that were found on the transcribed strand (T) related to the untranscribed (UT) strand, taking into account all sequenced clones. For UVB irradiated XP-C cells, fewer mutations were clearly observed on the transcribed strand (Figure 2.7), a pattern that is maintained when the analysis is performed by each individual clone (Supplementary Figure S2.5a). We also calculated the transcriptional strand bias score (SC), that measures

the ratio between mutational motifs in concordance (motifs detected in the same strand as the coding strand annotated in RefSeq) and in discordance (motifs found in the non-coding strand annotated in RefSeq) (Supplementary figure S2.5b, Supplementary table S2.3). For UVB irradiated XP-C cells, it was detected a SC of 2, which indicates a markedly strand bias for mutations in the untranscribed strand (non-coding strand), as expected, since these cells are proficient in TCR.

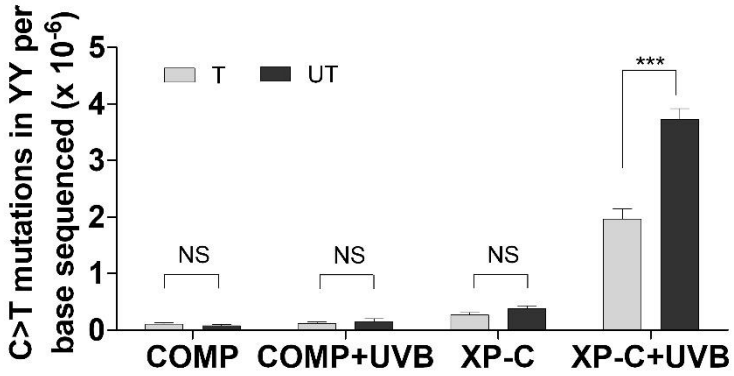


Figure 2.7 Transcriptional strand bias for C>T point mutations. a) Frequency of C>T transitions occurring inside the YY motif at transcribed (T) or untranscribed (UT) strands. Statistically significant differences were calculated by 2way ANOVA: P < 0.001 (***).

2.4.5 UVB light generates a considerable frequency of T>N point mutations

All kinds of base substitutions on thymine nucleotide were significantly increased after UVB irradiation in XP-C cells, and, thus, T>N mutations were also further investigated (Figure 2.8a). UVB-induced mutations at T were equally represented to the other three nucleotides, with both transitions and transversions. Curiously, after UVB irradiation these mutations were not significantly increased within dipyrimidine sites (Figure 2.8b). Also, no clear consensus sequence context was observed for the thymine mutations (Figure 2.8c). Therefore, the conclusion is that T>N mutations cannot be associated to pyrimidine dimer lesions even in UVB irradiated XP-C cells. However, it is not possible to completely discard a minor mutagenic effect of T containing pyrimidine dimers.

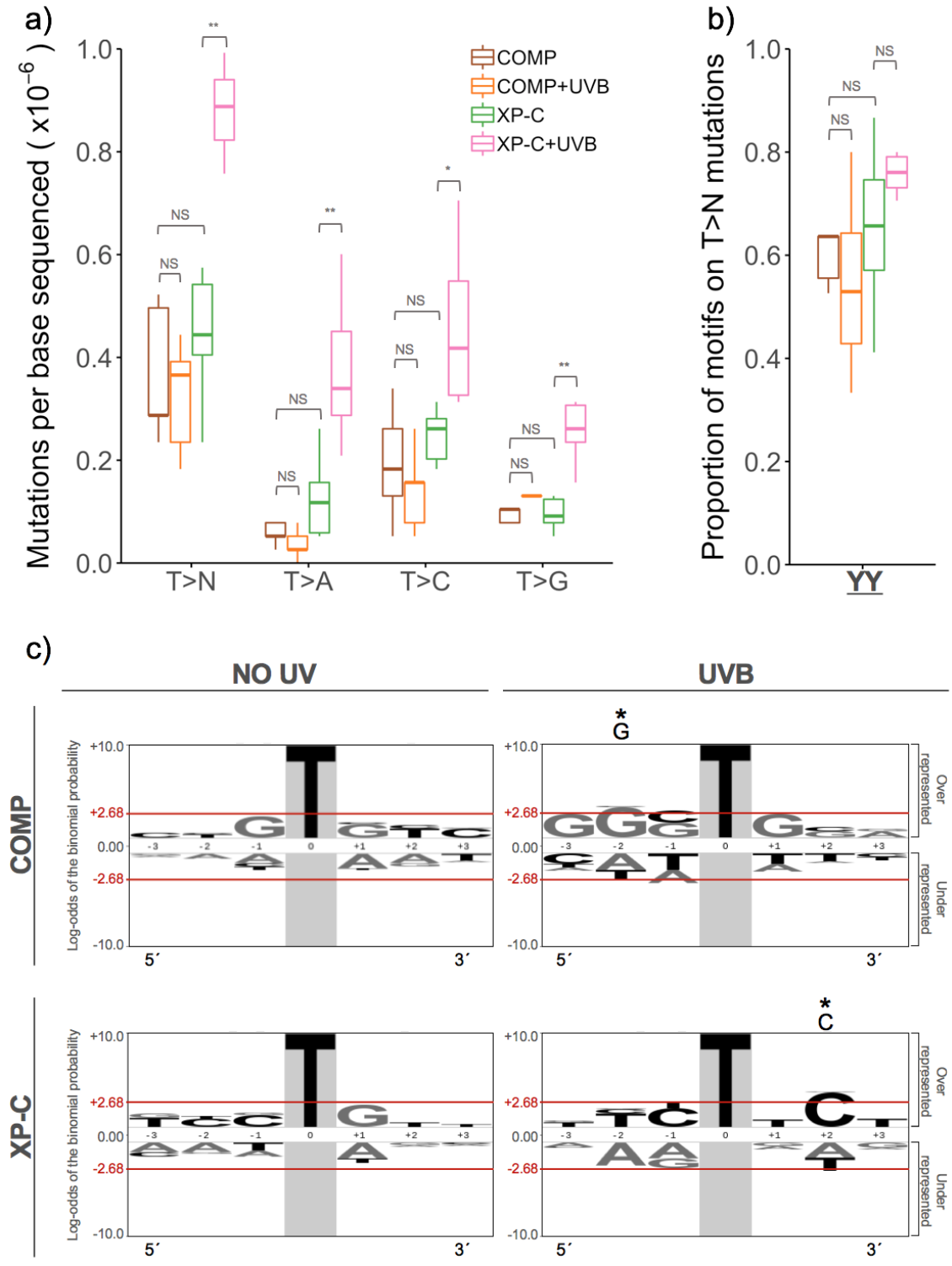


Figure 2.8 Nucleotide changes and mutagen-associated motif patterns in T>N point mutations. a) Analyses of T>N point mutations and each of the independent substitutions on T base after UVB or non-irradiation. **b)** Analyses of T>N mutations on YY motifs. Statistically significant differences were calculated by non-parametric permutation tests: P < 0.01 (**), P < 0.05 (*), P > 0.05 (NS). **c)** Sequence context logo of all T>N mutations on all clones. pLogo generator was used to generates the sequence conservation. Purine nucleotides were represented in gray and pyrimidine nucleotides in black. Significant enrichment of a particular nucleotide in a specific position is highlighted (*) at the top of the logo figure. Number of aligned foreground (n(fg)) sequences used to generate the image: COMP (n(fg)=70, COMP+UVB

(n(fg)=62, XP-C (n(fg)=102, XP-C+UVB (n(fg)=256. Number of background (n(bg)) 4096 sequences used.Irradiation dose: UVB 120 J/m^2.

2.4.6 Induction of C>A transversions by UVB irradiation is not related to oxidatively generated damage

C>A transversions are mostly associated with 8oxo-dG lesions, a product of oxidation normally related to UV irradiation. In fact, UVB irradiation increases significantly the frequency of C>A transversions on XP-C clones (Figure 2.3 and 2.9a). However, there was no clear context mutation signature where these mutations were occurring (Figure 2.4). 8oxo-dG seems to be more efficiently removed in a pyrimidine than in a purine rich sequence (Hatahet et al. 1998, Sassa et al. 2012), resulting in the proposal of a RGR motif for mutations induced by this lesion. Nonetheless, UVB light does not significantly increase the proportion of C>A transversions within the RGR motif in none of the irradiated clone groups (Figure 2.9b). On the other hand, YY motif was found to be increased for C>A point mutations in UVB-irradiated XP-C cells (Figure 2.9c). Similar to C>T transitions, C>A transversion mainly occurs at the second pyrimidine of a TC sequence but it is also represented in a smaller proportion on the first pyrimidine of CC dimer (Figure 2.9d). Moreover, the sequence context logo analysis reveals that irradiated XP-C clones had a statistically significant enrichment bias of the T base at the -1 position of the C>A mutation (Figure 2.10). Followed by a C base, but in a lesser and not significant extend, similar to observed for UVB-irradiated COMP clones. These results indicated that C>A mutations generated by UVB radiation in XP-C cells seem to also be targeted to pyrimidine dimers lesions, mainly the 5'TC dimer. In fact, the C>A point mutations represent the 7% of the total mutations generated by UVB in XP-C cells within YY motifs.

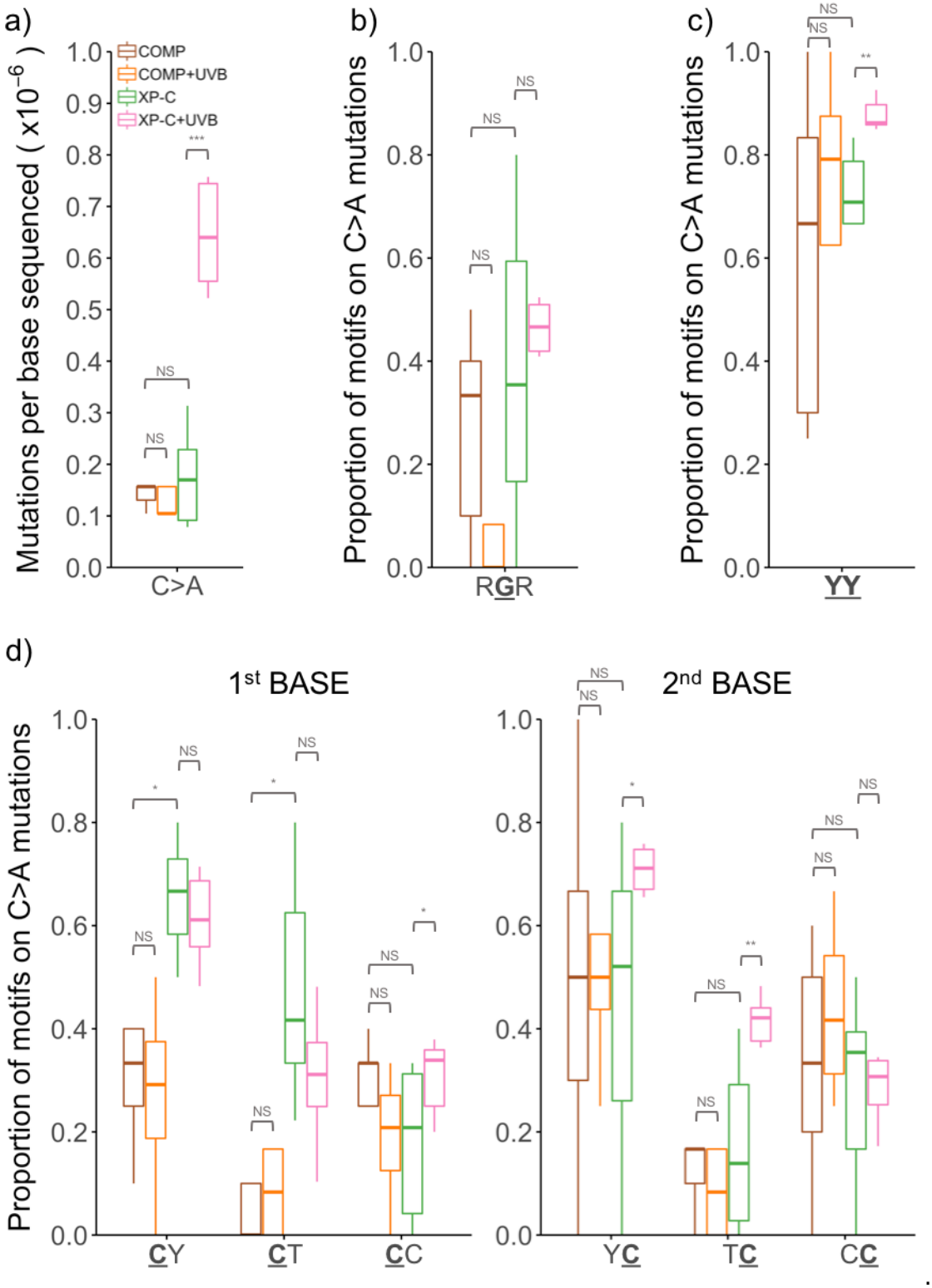


Figure 2.9 Nucleotide changes and mutagen-associated motif patterns in C>A point mutations. a) Analyses of C>A point mutations after UVB or non-irradiation. **b, c)** Analyses of motifs related to oxidized guanines (RGR sequence) and dipyrimidine sites (YY). **d)** Analysis of which base within a dipyrimidine is mutated, the first or the second one. Irradiation dose: UVB 120 J/m^2. Statistically significant differences were calculated by non-parametric permutation tests: P < 0.001 (***), P < 0.01 (**), P < 0.05 (*), P > 0.05 (NS).

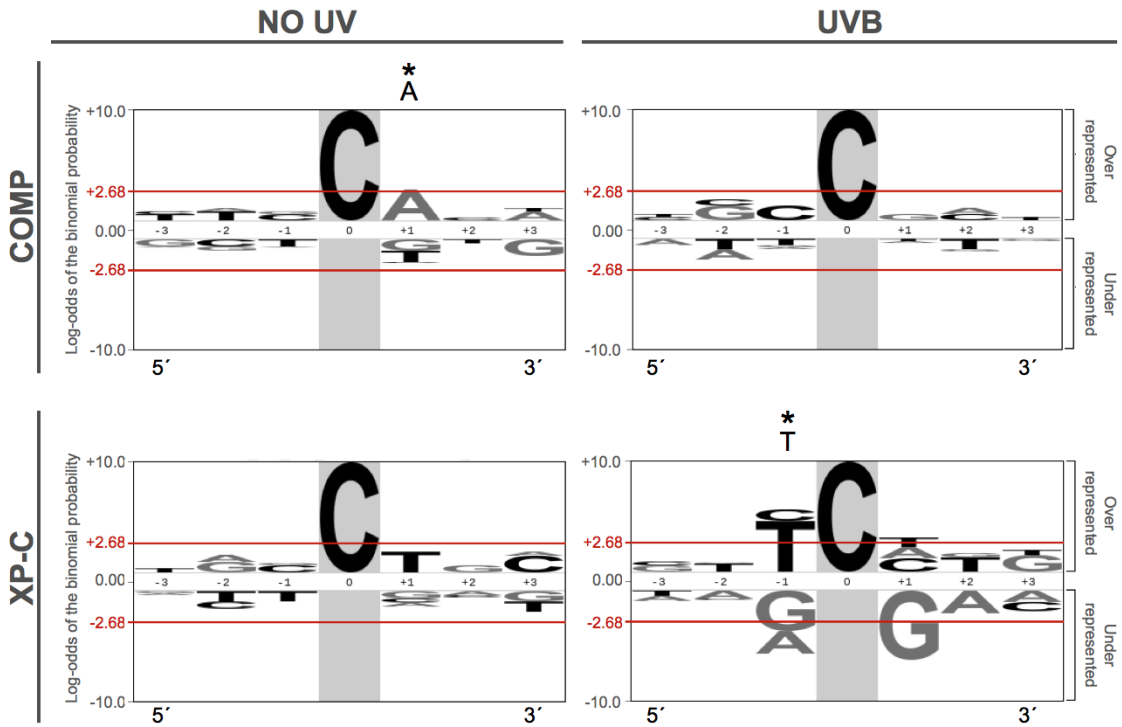


Figure 2.10 Sequence context for the C>A point mutations. Sequence context logo of all C>A mutations on all clones grouped. pLogo generator was used to generates the sequence conservation. Purine nucleotides were represented in gray and pyrimidine nucleotides in black. Significant enrichment of a particular nucleotide in a specific position is highlighted (*) at the top of the logo figure. Number of aligned foreground (n(fg)) sequences used to generate the image: COMP (n(fg)=31, COMP+UVB (n(fg)=20, XP-C (n(fg)=40, XP-C+UVB (n(fg)=148. Number of background (n(bg)) 4096 sequences used. Irradiation dose: UVB 120 J/m^2.

2.4.7 XP-C cells irradiated with UVB light display a mutational signature related to cutaneous skin cancer

Mutational Signatures are specific and unique combinations of the 96 types of base substitutions in a tri-nucleotide context classification, generated as a consequence of particular mutational processes (Alexandrov et al. 2013a). The somatic mutation signatures was extracted by the non-negative matrix factorization (NMF) method using all the base substitutions detected: only two signatures describe approximately ~86% of the mutations detected. Signature S-I is mainly characterized by a concentration of C>T mutations in a context of dipyrimidine sequences YCN, ACY and GCY, while the S-II signature is more-or-less equally represented by the 96 possible substitutions without apparent preference for any of the tri-nucleotide sequences, and consequently it could be more related with the mutagenic spectrum of non treated control clones (Figure 2.11a). Non-irradiated COMP clones are represented by the S-II signature and when UVB-irradiated a mixture of S-I and S-II signatures explains the mutation spectrum of this cell line, with an evident prevalence

of S-II. That is in agreement with the previous analysis that even without an increment in the absolute amount of mutations, UVB irradiation leads to a change in the mutation spectrum of COMP cell line with a slight increase of C>T transitions. On the other hand, non-irradiated XP-C clones exhibit a more homogeneous mixture of both, S-I and S-II signatures. In this case, the higher contribution of S-I seems to be reflect significant differences in the basal number of mutations, particularly C>T substitutions. UVB Irradiation, however, changes completely the contribution of these signatures and just S-I represents the mutation spectrum of XP-C irradiated clones (Figure 2.11b).

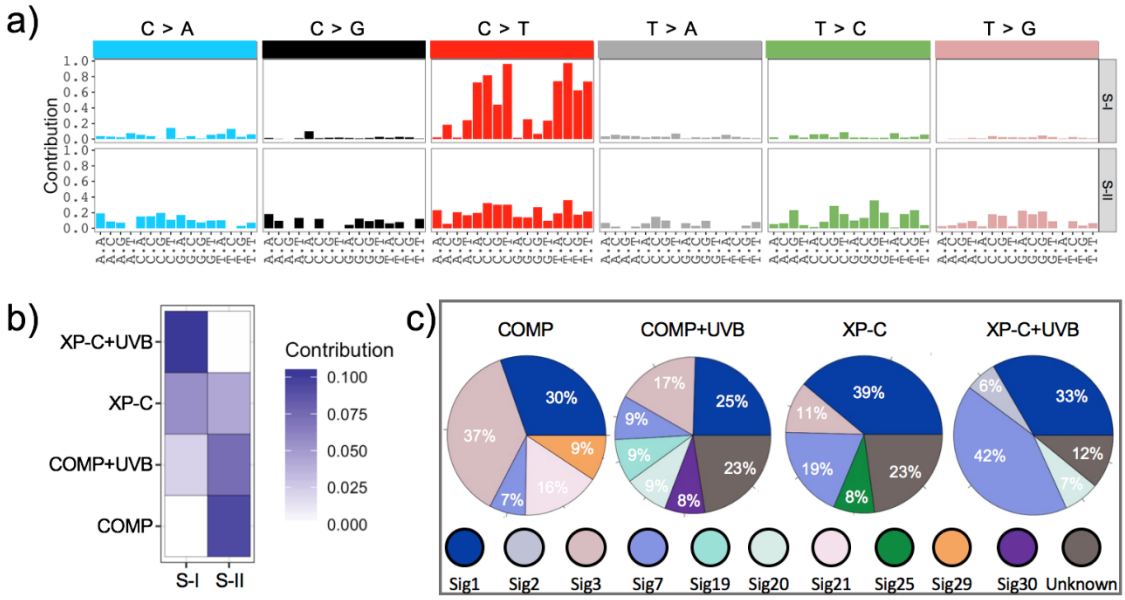


Figure 2.11 Mutational somatic signatures. (a) Mutational signatures extracted using NMF-method implemented in SomaticSignatures package based on a method developed by Alexandrov, 2012. **(b)** The contribution of each signature to explain the mutational pattern of irradiated or non-irradiated cell lines. **(c)** The contribution of the 30 published signatures from Sanger COSMIC project to these data calculated using deconstructSigs package.

Finally, the mutational spectrum of our data was compared with the pre-stablished signatures in the catalogue of somatic mutations in cancer, COSMIC version2 (https://cancer.sanger.ac.uk/cosmic/signatures_v2), that established a set of 30 signatures after been analyzed more than ten thousand samples of 40 distinct types of human cancers by whole genome or exome sequences (Figure 2.11c). Signature 1 found in all of our conditions is characterized by a C>T mutation in an NCG context. This signature was found in all types of cancer and its aetiology suggests an endogenous mutational process of spontaneous deamination of 5-methylcytocine (5mC), a very common process in cells. Signature 3, also found in some of the experimental conditions, is similar to the S-II signature. It is characterized by a more-

or-less equal distribution of the 96 possible substitutions, but with a clear lack of C>N mutations in an NCG context. This signature mainly found in breast, ovarian and pancreatic cancers, and has been associated with failure of DNA double-strand break-repair by homologous recombination. Signature 7, also found in all of our experimental conditions, is characterized by C>T mutations at dipyrimidines, has been linked to ultraviolet light exposure and mainly found in skin cancers. Interestingly, the contribution of signature 7 is significatively increased in XP-C cells when irradiated with UVB light, going from 19% to 42% on the total mutational spectrum of these clones. Additionally, signature 7 of COSMIC is similar to S-I signature, that was directly related to XP-C cells irradiated with UVB light. These results indicate that even a unique and small dose of UVB contribute to generate the typical mutation spectrum of skin cancer in cells of patients without repair.

2.5 Discussion

Sunlight radiation is essential for life on earth, but it can also be harmful to DNA due to its UV component that causes structural damage in DNA, such as pyrimidine dimers and oxidized bases. These lesions when unrepaired generate basically transitions of the C>T type and transversions of the C>A type, respectively. Also, the C>T mutations have been widely related to skin cancer (Epe 1991, Cheng et al. 1992, Ziegler et al. 1993, Steenken & Jovanovic 1997, Ikehata & Ono 2011). To date, most of the studies are carried out using UVC light as a model, but only UVA and UVB lights are considered biologically relevant, as they are part of the sunlight spectrum that reaches the Earth's surface (Matsumura & Ananthaswamy 2002, Schuch & Menck 2010, Schuch et al. 2012). Thus, this study focused investigation on the mutagenesis induced by the more energetic UVB wavelengths, which although correspond to only 5% of the solar UV, it its 10^3 fold more effective than UVA in inducing CPDs (Kuluncsics et al. 1999, Sage et al. 2012), the most mutagenic lesion induced by UV radiation (You et al. 2001, Courdavault et al. 2005). Moreover, we investigated the effects of UVB on human cells based on a model cell line, deficient on the XPC protein, which does not perform GGR of UV-induced DNA damage. As control, we used an isogenic cell line complemented with the wild type gene.

In order to obtain more relevant and comparable data, these cell lines were cloned previous to UVB-irradiation, and tested for basic phenotype after UVB-irradiation. As expected, XPC deficient cells exhibit a dose-dependent decrease on

cell survival when exposed to UVB, compared to control cells. This increase in sensitivity involves genotoxic stress, as indicated by the increased levels of γH2AX, as well as an arrest of the cell cycle on S/G2 (Figure 2.1). This is highly similar to previous reports for XP-C cells after UVB and UVC irradiation (Feraudy et al. 2010, Dupuy et al. 2013, Quinet et al. 2014, 2016, Andrade-Lima et al. 2015). The induction of death and the arrest of the cell cycle to favor repair are considered strategies of cells to prevent divisions when its DNA contains lesions. Also, the proliferation of human fibroblasts and keratinocytes was inhibited 24 h after UVB exposure, but a recovery was observed 48 h after treatment (Courdavault et al. 2004, 2005). This higher sensitivity of XP-C cells is certainly due to its inability to recognize and remove the UV-induced DNA damage in the global genome, not affecting TCR repair responsible for the repair in actively transcribed regions. Due to the importance of XPC protein on UV sensitivity protection, the XP-C cells + UVB light model allows to study directly the relationship between DNA damage, mutation and cancer, which can help to clarify the mechanisms involved in this process.

A UVB dose of 120 J/m^2 was selected for the mutagenesis experiments. This dose can be considered environmentally relevant, as it corresponds approximately to one minute and half, of exposure to UV-sunlight in the period of 10:00 am to 2:00 pm during summer in a tropical latitude, as is the case of Sao Paulo, Brazil (Schuch et al. 2012). Based on the experiments for DNA damage detection (Supplementary figure S2.2) and considering previous reports indicating that induction of bipyrimidine photoproducts by UVB is not modulated by cellular content when compared to naked DNA (Cadet et al. 2005). It was calculated that, this UVB irradiation dose generates approximately 140 CPDs per Mbp on DNA, which is similar to previously reported (Schuch et al., 2009), corresponding to approximately 8.46x10^5 CPDs per cell. This UVB radiation dose increased the mutation prevalence on XP-C cells from 1.5 to 8.15 per Mbp, without significantly modifying the prevalence of mutations in control cells (Figure 2.2a). This is consistent with the mutation prevalence in exons (8.33 per Mbp) and introns (9.93 per Mbp) estimated in the sequenced genome of a malignant melanoma (Pleasance et al. 2010a). And the mutation frequency of *cII* and *lacI* transgenes in mouse embryonic fibroblasts irradiated with 500 J/m^2 of UVB light, where for non-irradiated cells had a mutation frequency of 0.6-2 per Mbp and UVB irradiated ones a frequency of 11-19 per Mbp (You et al. 2001). Considering only C>T transitions in <u>YY</u> motif (6.25 mutations per Mb), we estimated that approximately 3.5% of the

CPDs induced by UVB irradiation are fixed as mutations in XPC deficient cells.

A unique dose of UVB increased the amount of SNVs on XP-C cells, being the C>T transition and CC>TT tandem mutation the main point mutations generated (Figure 2.3, 2.5a). Similar results were reported in the *hprt* gene of primary human fibroblast (NER proficient) and *cII* transgenes in mouse embryonic fibroblast after UVB irradiation (You et al. 2001, Kappes et al. 2006). Both, C>T and CC>TT are known as UVC/UVB signature mutations (Ikehata & Ono 2011), generated as consequence of unrepaired pyrimidine dimers (Ziegler et al. 1993). Using the mutation spectrum and motif analysis, the C>T substitutions were shown to be at dipyrimidine, where pyrimidine dimers could be induced. These transitions were observed mainly within a context sequence of 5´CCN and 5´TCN, and with a preference to mutate the second pyrimidine of a dimer, 3´Y (Figure 2.4 and 2.5). In a melanoma sequenced genome, most of the somatic base substitutions were C>T transitions, and 92% of this mutations occurred at the 3´base of a dipyrimidine site (Pleasance et al. 2010a). For UVB-irradiated XP-C cells, this prevalence was 86.5% with preference to occur at the TC dimer, whereas in COMP cells with functional repair, the preference was for the CC dimer.

HPLC-MS/MS analyses of DNA from the skin or primary cell culture of keratinocytes and fibroblasts after UVB exposure established that the main photoproducts induced were TT(CPD) > TC (CPD) ≈ TC (6-4PP) > CT (CPD) > TT (6-4PP), with CC and CT photolesions, as well as Dewar valence isomers, below the detection limit at the studied doses (Courdavault et al. 2004, 2005, Mouret et al. 2006). In naked DNA, a similar distribution was established both with UVB and UVC light (Douki & Cadet 2001). Interestingly, the removal efficiency of the four possible CPD dimers is CT > CC > TC > TT (Mouret et al. 2008), approximately matching mutagenesis data, since most mutations found on 5´TC sites while few mutations are found in 5´CT ones (Cadet & Douki 2018). Also, it was proposed that on TC sites equal frequencies of CPDs and 6-4PP dimers are generated (Cadet et al. 1992). However, the TC (6-4PP) dimer and its Dewar isomer were proposed as highly mutagenic in experimental model of *E. coli* (Horsfall & Lawrence 1994). While in mammalian cells, CPDs are proposed as responsibles for the vast majority of mutations (You et al. 2001), a contribution of TC 6-4PP cannot be ruled out. We cannot state definitively what type of lesion is generating the C>T transitions, however the data support the idea that TC dimers at 3´ base are the most mutagenic photoproducts.

Although the TT (CPD) is the major photoproduct induced by UV, mutations occur more frequently at dimers containing cytosine. The low contribution of TT dimers to the UV mutations is related to the translesion synthesis (TLS), a damage tolerance mechanism of cells, mainly performed by the Polη that bypass TT dimers in error-free manner (Stary et al. 2003, Pfeifer et al. 2005). Some authors argue that the two thymines that conform the TT (CPD) dimer keep their coding properties since they retain the Watson-Crick base pairing portion, and thus, TT (CPD) would behave more like a blockage lesson than as a mutagenic one (Taylor 2002, Cadet & Douki 2018). The A rule has been traditionally used to explain the high occurrence of the C>T mutation, this rule basically proposes the insertion of A more often than other nucleotide in the absence of a templating base (abasic sites) or in the presence of an instructional lesson that cannot be accommodated by the active site of the replicative polymerases (Taylor, 2002). Fit is considered a very important factor to select the nucleotide to be inserted, thereby, the greatest pi-stacking energy of A followed by G, T and C, favors the selection of A. This is consistent with the order in which the Y family of TLS polymerases prefers to insert nucleotides opposite abasic sites (Taylor 2002). Indeed, the highly error-free activity of Polη was also extended to CPDs of various dipyrimidine combinations (Yoon et al. 2009).

TLS mechanisms incorporate errors that increase the amount of mutations, however the absence of TLS could lead to the collapse of replication fork that ends in double strand breaks and require the activation of repair pathways that can generate the induction of chromosomal rearranges. Therefore, TLS mechanisms help to avoid chromosomal instability that have a higher contribution to the tumorogenesis process (Yoon et al. 2019). Also, it has been proposed that Polη protects from UV induced mutations *in vivo*, in fact the lack of Polη leads to XP variant syndrome (XP-V) characterized by an increased sensitivity to sunlight and elevated cancer prone (Stary et al. 2003, Moreno et al. 2020). On this sense, the second base of the CPDs and 6-4PPs dimers is the first one to be replicated by the TLS polymerases that can be error free or error prone (Polη, Polθ or Polι) depending on the nature of the dimer and the structure of the polymerase active site (Taylor 2002, Yoon et al. 2009, 2019), while the extension is apparently being performed in an error free manner (Polη, Polκ, Polζ or Polθ).

The frequency of CPD formation was reported to be increased in regions with

methylated cytosine (5mC) after sunlight and UVB irradiation, with subsequent generation of C>T mutations, particularly in methylated CpGs regions (Douki & Cadet 1994, Tommasi et al. 1997, Pfeifer et al. 2005, Pleasance et al. 2010a), possibly because at 290 nm the 5mC has a molar absorption coefficient fivefold higher than unmethylated cytosine (reviewed in Tomkova & Schuster-Böckler 2018). Additionally, the half-life for spontaneous deamination of C and 5mC goes from the order of thousands of years in undamaged DNA, to days and even hours when these bases are within pyrimidine dimers. Also, the sequence context appears to affect the deamination rate of 5mC, being faster in 5´TmCG than 5´CmCG sites (Cannistraro & Taylor 2009, Tomkova & Schuster-Böckler 2018). Based on those studies, we evaluated the motif 5´YCG and the specific 5´TCG and 5´CCG motifs. However, no increase in the frequency of mutations in these motifs were detected on UVB irradiated XP-C cells. Our data support the statements by Douki and Cadet (2001), who previously highlighted that it is not possible to extrapolate the partial information about dipyrimidine formation containing 5mC bases found within exons of p53 gene to whole genomes. In the same way, Poulos and collaborators (2017a) by implementing regression models expected a positive and linear association between the CpG mutation rate and methylation on skin cancer subtypes. However, they found a negative association at methylation fractions greater than ~0.5 and related this result with NER induction after UV exposure, since TCR subpathway early works in active genes that have highly methylated regions. On the other hand, 80% of the CpG methylation in the whole genome is not informative and it was suggested that 5-hydroxymethylcytosine (5hmC), detected mainly in actively transcribed genes, exons, and enhancers regions, could have a role in protecting CpG sites from mutations (Gifford et al. 2013, Tomkova & Schuster-Böckler 2018).

The results shown here indicate that a considerable part of the mutations occurred at T bases, since UVB light also increases all T mutations (T>A, T>C and T>G) without preference towards which base will change, but these mutations seem to be unrelated to pyrimidine dimers. The C>A transversions were also significantly increased, with this mutation known as the mutagenic hallmark of the 8oxo-dG, generated by guanine oxidation (Epe 1991, Cheng et al. 1992). However, under our experimental conditions the C>A transversions were more related with 5´TC and 5´CC dimers than with the RGR motif established for oxidized guanines (Figure 2.9 and 2.10). It is important to mention that on human skin exposed to UVB the yield of 8-

oxoG was estimated at least in two orders of magnitude less than CPDs (Mouret et al. 2006). Additionally, the base excision repair (BER) is the main pathway that deals with oxidatively generated damage and XP-C cells are BER proficient. Thus, it is possible that the mutations generated by 8oxo-dG were not detected because they are quickly and efficiently repaired. Also, it was related that while C>T mutations are early mutations in the evolution of cancer, C>A mutations are considered late mutations. Moreover, it was suggested that the process underlying C>A mutations is not related to UV light exposure (Pleasance et al. 2010a). Finally, the only type of mutation that did not increase after UVB irradiation of XP-C cells was C>G transversion, that curiously is the less registered on a sequenced malignant melanoma (Pleasance et al. 2010a).

The spontaneous mutation rate of XP-C cells is higher than that of COMP cells, we relate this to an increase in the proportion of transitions to transversions (Figure 2.2), mainly explained by C>T mutations (Figure 2.3). The elevated number of transitions in all experimental groups, in part, can be explained by the higher deamination rate of cytosine to uracil, 5-mC to thymine and adenine to hypoxanthine, which results in spontaneous mutations of C>T and A>G types, respectively (Lindahl et al. 1993, Tomkova & Schuster-Böckler 2018). The presence of COSMIC´s signature 1 in all of the experimental groups supports this idea (Figure 2.11c), as this signature has been related to the C>T mutation by spontaneous deamination of 5mC, the most common type of mutations observed in cancer (Tomkova & Schuster-Böckler 2018). Also, the particular increase of C>T transitions in XP-C cells may be related to spontaneous deamination of cytosine, since the percentage of SNVs explained by the signature 1 keeps constant in all conditions, while in non-irradiated XP-C cells increased by almost 10%. Additionally, due to the elevated incidence of cancer in XP-C patients, it has been suggested that mutations or loss of the XPC gene may be an early event during skin carcinogenesis. In fact, it was demonstrated that the XPC gene is inactivated or lost in at least 50% of squamous skin cancers of both immunocompetent and immunosuppressed non-XPC patients (Feraudy et al. 2010). Also, it was reported a 30-fold increase of the spontaneous mutation frequency in the *hprt* gene of T-lymphocytes in 1 year old $XPC^{-/-}$ mice (Wijnhoven et al. 2000).

A study that analyzes the three most frequent base substitutions, C>T, A>G and G>T, in more than 4500 human genes revealed a strand asymmetry, with substitution rates on average higher on the coding strand and lower on the transcribed strand

(Mugal et al. 2009). Also, it was evidenced an inverse relation between mutation prevalence and gene expression on both coding and transcribed strands in the analysis of the whole genome of a malignant melanoma and human lung cancer cell line (Pleasance et al. 2010a, 2010b). We observed a clear strand bias for mutations occurring on the coding strand of UVB irradiated XP-C cells, and to a lesser extent in non-irradiated XPC cells. This bias can be explained as in XP-C cells the TCR subpathway is functional, and it was established that repair of actively transcribed genes occurs at a same rate that in normal cells (Venema et al. 1991). So, the TCR subpathway promotes the generation of a bias of base substitutions at the coding strand, due to the efficient lesion removal at the transcribed (and non-coding) strand.

Our results point out that in cells deficient in GGR, a unique UVB dose as low as 120 J/m^2 contribute to generate the typical mutation spectrum of skin cancer. More than 40% of the mutational spectrum of the irradiated XP-C clones is explained by the COSMIC´s signature 7, considered a hallmark predominantly found in different types of skin cancers. The use of NER deficient cells allowed us to amplify the effects of UVB radiation on human skin fibroblasts. This experimental design is innovative as it allows to detect the mutations generated in the first round of replication after UV radiation (Moreno et al. 2020). Also, it takes into account the clonal diversity, an important characteristic of the carcinogenic process. Using exome data, a more global picture gives us good ideas of what can happen at the genome. Avoiding some problems, as the use of genes under selection and the generation of clustering information, found in previous works that analyze mutagenesis in only one gene, or DNA sequences carried in vectors derived from virus (Alexandrov & Stratton 2014). This work supports that UVB is highly mutagenic and in XPC patients, this is exacerbated. Eventually, the data may be used for comparison with the mutational profiles of skin tumors obtained from XP patients. As well as they may help to understand the mutational processes in skin tumors of individuals not affected with XP.

Acknowledgements: The authors thank for the financial support to the Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP, São Paulo, Brazil, Grants #2019/19435-3, #2013/08028-1). Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Grant # 308868/2018-8) and Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior (CAPES, Brasília, DF, Brazil, financial code 001) and Administrative Department of Science, Technology and Innovation of Colombia (COLCIENCIAS, Bogota, Colombia). We are also grateful to the support with NGS, at

the core Facility for Scientific Research – USP (CEFAP-USP/GENIAL), and Multi-user genomic Section of the Human Genome & Stem Cell Research Center (HUG-CELL).

2.6 References

ALEXANDROV, L.B. et al. Signatures of mutational processes in human cancer. **Nature**, v. 500, p. 415-421, 2013a.

ALEXANDROV, L.B. et al. Deciphering Signatures of Mutational Processes Operative in Human Cancer. **Cell Reports**, v. 3, p. 246–259, 2013b.

ALEXANDROV, L.B., STRATTON, M.R. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. **Current Opinion in Genetics & Development**, v. 24, p. 52–60, 2014.

ANDRADE-LIMA, L.C., ANDRADE, L.N., MENCK, C.F.M. ATR suppresses apoptosis after UVB irradiation by controlling both translesion synthesis and alternative tolerance pathways. **Journal of Cell Science**, v. 128, p.150–159, 2015.

ARMSTRONG, B.K., KRICKER, A. The epidemiology of UV induced skin cancer. **Journal of Photochemistry and Photobiology B: Biology**, v. 63, p. 8–18, 2001.

BOWDEN, N.A. et al. Understanding *xeroderma Pigmentosum* complementation groups using gene expression profiling after UV-light exposure. **International Journal of Molecular Sciences**, v. 16, p.15985–15996, 2015.

CADET, J., ANSELMINO, C., DOUKI, T., VOITURIEZ, L. Photochemistry of nucleic acids in cells. **Journal of Photochemistry and Photobiology B: Biology**, v. 15, p. 277–298, 1992.

CADET, J., SAGE, E., DOUKI, T. Ultraviolet radiation-mediated damage to cellular DNA. **Mutation Research**, v. 571, p. 3–17, 2005.

CADET, J., DOUKI, T. Formation of UV-induced DNA damage contributing to skin cancer development. **Photochemical & Photobiological Sciences**, v. 17, p. 1816–1841, 2018.

CANNISTRARO, V.J., TAYLOR, J.S. Acceleration of 5-methylcytosine deamination in cyclobutane dimers by g and its implications for UV-induced c-to-t mutation hotspots. **Journal of Molecular Biology**, v. 392, p. 1145–1157, 2009.

CHENG, K.C., CAHILL, D.S., KASAI H., NISHIMURA, S., LOEB, L.A. 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G→T and A→C substitutions. **The Journal of biological chemistry**, v. 267, p.166–172, 1992.

CLEAVER, J.E. Defective repair replication of DNA in xeroderma pigmentosum. **DNA repair**, v. 218, p. 652–656, 1968.

CLEAVER, J.E. DNA repair in human xeroderma pigmentosum group C cells involves a different distribution of damaged sites in confluent and growing cells. **Nucleic acids research**, v. 14, p. 8155–8165, 1986.

CLEAVER, J.E., LAM, E.T., REVET, I. Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. **Nature reviews. Genetics**, v. 10, p.756–768, 2009.

COURDAVAULT, S. et al. Unrepaired cyclobutane pyrimidine dimers do not prevent

proliferation of UV-B-irradiated cultured human fibroblasts. **Photochemistry and Photobiology**, v. 79, p. 145–151, 2004.

COURDAVAULT, S. et al. Repair of the three main types of bipyrimidine DNA photoproducts in human keratinocytes exposed to UVB and UVA radiations. **DNA Repair**, v. 4, p. 836–844, 2005.

DANECEK, P. et al. The variant call format and VCFtools. **Bioinformatics**, v. 27, p.2156–2158, 2011.

DAYA-GROSJEAN, L., JAMES, M.R., DROUGARD, C., SARASIN, A. An immortalized xeroderma pigmentosum, group C, cell line which replicates SV40 shuttle vectors. **Mutation Research DNA Repair Reports**, v. 183, p.185–196, 1987.

De GRUIJL F.R. et al. Wavelength dependence of skin cancer induction by ultraviolet irradiation of albino hairless mice. **Cancer research**, v. 53, p. 53–60, 1993.

DePRISTO, M.A. et al. A framework for variation discovery and genotyping using nextgeneration DNA sequencing data. **Nature Genetics**, v. 43, p.491–498, 2011.

DOUKI, T., CADET, J. Formation of cyclobutane dimers and (6-4photoproducts upon far-UV photolysis of 5-methylcytosine-containing dinucleoside monophosphates. **Biochemistry**, *v.* 33, p. 11942-11950, 1994.

DOUKI, T., CADET, J. Individual determination of the yield of the main UV-induced dimeric pyrimidine photoproducts in DNA suggests a high mutagenicity of CC photolesions. **Biochemistry**, v. 40, p. 2495–2501, 2001.

DUPUY, A. et al. Targeted gene therapy of xeroderma pigmentosum cells using meganuclease and TALENTM. **PLoS ONE**, v. 8, p.1–8, 2013.

EL GHISSASSI, F. et al. A review of human carcinogens—Part D: radiation. **The Lancet Oncology**, v. 10, p.751–752, 2009.

EPE, B. Genotoxicity of singlet oxygen. **Chemico-Biological Interactions**, v. 80, p.239–260, 1991.

FERAUDY, S. et al. The DNA damage-binding protein XPC is a frequent target for inactivation in squamous cell carcinomas. **The American Journal of Pathology**, v. 177, p. 555–562, 2010.

FRANKEN, N.A.P., RODERMOND, H.M., STAP, J., HAVEMAN, J., BREE, C.V. Clonogenic assay of cells *in vitro*. **Nature protocols**, v. 1, p. 2315–2319, 2006.

GEHRING, J.S., FISCHER, B., LAWRENCE, M., HUBER, W. SomaticSignatures: Inferring mutational signatures from single-nucleotide variants. **Bioinformatics**, v. 31, p. 3673–3675, 2015.

GIFFORD C.A. et al. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. **Cell**, v. 153, p. 1149–1163, 2013.

HATAHET, Z., ZHOU, M., REHA-KRANTZ, L.J., MORRICAL, S.W., WALLACE, S.S. In search of a mutational hotspot. **Proceedings of the National Academy of Science of *the USA*,** v. 95, p. 8556–8561, 1998.

HORSFALL, M.J., LAWRENCE. C.W. Accuracy of replication past the T-C (6-4) adduct.

Journal of Molecular Biology. v. 235, p. 465–471, 1994.

IKEHATA, H., ONO, T. The Mechanisms of UV Mutagenesis. **Journal of Radiation Research**, v. 52, p.115–125, 2011.

KAPPES, U.P., LUO, D., POTTER, M., SCHULMEISTER, K., RUNGER, T.M. Short- and long-wave UV light (UVB and UVA) induce similar mutations in human skin cells. **The Journal of investigative dermatology**, v. 126, p. 667–675, 2006.

KULUNCSICS, Z., PERDIZ, D., BRULAY, E., MUEL B., SAGE, E. Wavelength dependence of ultraviolet-induced DNA damage distribution: involvement of direct or indirect mechanisms and possible artefacts. **Journal of Photochemistry and Photobiology B: Biology**, v. 49, p. 71–80, 1999.

LEHMANN, A.R. et al. Xeroderma pigmentosum cells with normal levels of excision repair have a defect in DNA synthesis after UV-irradiation. **Proceedings of the National Academy of Sciences of the USA**, v. 72, p. 219–223, 1975.

LEITE, R.A. et al. Identification of XP complementation groups by recombinant adenovirus carrying DNA repair genes. **Journal of Investigative Dermatology**, v. 129, p. 502-506, 2009.

LI, L., BALES, E.S., PETERSON, C.A., LEGERSKI, R.J. Characterization of molecular defects in xeroderma pigmentosum group C. **Nature Genetics**, v. 5, p. 413–417, 1993.

LI, H. DURBIN, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. **Bioinformatics**, v. 26, p. 589–595, 2010.

LINDAHL, T. Instability and decay of the primary structure of DNA. **Nature**, v. 362, p. 709-715, 1993

MATSUMURA, Y. ANANTHASWAMY, H.N. Molecular mechanisms of photocarcinogenesis. **Frontiers in Bioscience**, v. 7, p. 765-783, 2002.

MENCK, C.F., MUNFORD, V. DNA repair diseases: what do they tell us about cancer and aging? **Genetics and Molecular Biology**, v. 37, p. 220–233, 2014.

MORENO, N. et al. Whole-exome sequencing reveals the impact of UVA light mutagenesis in xeroderma pigmentosum variant human cells. **Nucleic Acid Research**, v. 48, p. 1941-1953, 2020.

MOURET, S., BAUDOUIN, C., CHARVERON, M., FAVIER, A., CADET, J., DOUKI, T. Cyclobutane pyrimidine dimers are predominant DNA lesions in whole human skin exposed to UVA radiation. **Proceedings of the National Academy of Sciences of the USA**, v. 103, p.13765–13770, 2006.

MOURET, S., CHARVERON, M., FAVIER, A., CADET, J., DOUKI, T. Differential repair of UVB-induced cyclobutane pyrimidine dimers in cultured human skin cells and whole human skin. **DNA Repair**, v. 7, p. 704–712, 2008.

MUNFORD V., CASTRO L.P. et al. A genetic cluster of patients with variant xeroderma pigmentosum with two different founder mutations. **British Journal of Dermatology**, v. 176, p. 1270–1278, 2017.

MUGAL, C.F., VON GRÜNBERG, H.H., PEIFER, M. Transcription-induced mutational strand bias and its effect on substitution rates in human genes. **Molecular Biology Evolution**, v. 26, p. 131-142, 2009.

NIK-ZAINAL, S. et al. Mutational processes molding the genomes of 21 breast cancers. **Cell**, v. 149, p. 979–993, 2012.

O'LEARY, N.A. et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. **Nucleic Acids Research**, v. 44, p. D733–D745, 2016.

O'SHEA, J.P., CHOU, M.F., QUADER, S.A., RYAN, J.K., CHURCH, G.M., SCHWARTZ, D. pLogo: a probabilistic approach to visualizing sequence motifs. **Nature methods**, v. 10, p. 1211-1212, 2013.

PFEIFER, G.P., YOU Y.H., BESARATINIA, A. Mutations induced by ultraviolet light. **Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis**, v. 571, p.19–31, 2005.

PLEASANCE, E.D. et al. A comprehensive catalogue of somatic mutations from a human cancer genome. **Nature**, v. 463, p. 191–196, 2010a.

PLEASANCE, E.D. et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. **Nature**, v. 463, p. 184–190, 2010b.

POULOS, R.C., OLIVIER, J., WONG, J.W.H. The interaction between cytosine methylation and processes of DNA replication and repair shape the mutational landscape of cancer genomes. **Nucleic Acids Research**. v.45, p. 7786–7795, 2017.

QUINET, A. et al. Gap-filling and bypass at the replication fork are both active mechanisms for tolerance of low-dose ultraviolet-induced DNA damage in the human genome. **DNA Repair**, v. 14, p.27–38, 2014.

QUINET, A. et al. Translesion synthesis mechanisms depend on the nature of DNA damage in UV-irradiated human cells. **Nucleic Acids Research**, v. 44, p. 5717–5731, 2016.

RAFEHI, H, ORLOWSKI, C., GEORGIADIS, G.T., VERVERIS, K., EL-OSTA, A., KARAGIANIS, T.C. Clonogenic Assay: Adherent Cells. **Journal of Visualized Experiments**, v. 49, p.1–4, 2011.

RASTOGI, R.P., RICHA, KUMAR, A., TYAGI, M.B., SINHA, P. Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. **Journal of nucleic acids**, v. 2010, p. 1-32, 2010.

RATUSHNY, V., GOBER M.D., HICK R., RIDKY T.W., SEYKORA J.T. From keratinocyte to cancer: the pathogenesis and modeling of cutaneous squamous cell carcinoma. **Journal of Clinical Investigation**. V. 122, p. 464–472, 2012.

RAVANAT, J.L., DOUKI, T., CADET, J. Direct and indirect effects of UV radiation on DNA and its components. **Journal of Photochemistry and Photobiology B: Biology**, v. 63, p. 88–102, 2001.

ROSENTHAL, R., Mc.GRANAHAN, N., HERRERO, J., TAYLOR, B.S., SWANTON, C. DeconstructSigs: Delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. **Genome Biology**, v. 17, p. 1-11, 2016.

SAGE, E., GIRARD, P.M., FRANCESCONI, S. Unravelling UVA-induced mutagenesis. **Photochemical & Photobiological Sciences**, v. 11, p.74–80, 2012.

SANCAR, A. DNA Excision Repair. **Annual Review of Biochemistry**, v. 65, p.43–81, 1996.

SASSA, A., BEARD, W.A., PRASAD, R., WILSON, S.H. DNA sequence context effects on the glycosylase activity of human 8-oxoguanine DNA glycosylase. **Journal of Biological Chemistry**, v. 287, p. 36702–36710, 2012.

SCHNEIDER, C.A., RASBAND, W.S., ELICEIRI, K.W. NIH Image to ImageJ: 25 years of image analysis. **Nature Methods**, v. 9, p. 671-675, 2012.

SCHUCH, A.P., GALHARDO, R.S., LIMA-BESA, K.M., SCHUCH, N.J., MENCK, C.F.M. Development of a DNA-dosimeter system for monitoring the effects of pulsed ultraviolet radiation. **Photochemical & Photobiological Sciences**, v. 8, p.111–120, 2009.

SCHUCH, A.P., MENCK, C.F.M. The genotoxic effects of DNA lesions induced by artificial UV-radiation and sunlight. **Journal of photochemistry and photobiology. B, Biology**, v. 99, p.111– 116, 2010.

SCHUCH, A.P. et al. DNA damage profiles induced by sunlight at different latitudes. **Environmental and Molecular Mutagenesis**, v. 53, p.198–206, 2012.

SETLOW, R.B. The wavelengths in sunlight effective in producing skin cancer: a theoretical analysis. **Proceedings of the National Academy of Sciences of the USA**, v. 71, p.3363–3366, 1974.

SOUFIR N. et al. A prevalent mutation with founder effect in xeroderma pigmentosum group C from North Africa. **Journal of Investigative Dermatology**, v. 130, p.1537-1542, 2010.

SOUZA, T.A., DEFELICIBUS, A., MENCK, C.F.M. **in preparation**. Available at: github.com/tiagoantonio/woland.

STARY, A., SARASIN, A. The genetics of the hereditary xeroderma pigmentosum syndrome. **Biochimie**, v. 84, p. 49–60, 2002.

STARY, A., KANNOUCHE, P., LEHMANN, A.R., SARASIN, A. Role of DNA polymerase eta in the UV mutation spectrum in human cells. **Journal of Biological Chemistry**, v. 278, p. 18767–18775, 2003.

STEENKEN, S., JOVANOVIC, S.V. How easily oxidizable is DNA? One-electron reduction potentials of adenosine and guanosine radicals in aqueous solution. **Journal of the American Chemical Society**, v. 119, p. 617–618, 1997.

SUGASAWA, K. Xeroderma Pigmentosum genes: Functions inside and outside DNA repair. **Carcinogenesis**, v. 29, p. 455–465, 2008.

TAYLOR, J.S. Unraveling the Molecular Pathway from Sunlight to Skin Cancer. **American Chemical Society**, v. 27, p. 76–82, 1994.

TAYLOR, J.S. New structural and mechanistic insight into the A-rule and the instructional and non-instructional behavior of DNA photoproducts and other lesions. **Mutation Research**, v. 510, p. 55-70, 2002.

TOMKOVA, M., SCHUSTER-BÖCKLER, B. DNA Modifications: Naturally More Error Prone? **Trends Genetics**. 34, 627–638, 2018.

TOMMASI, S., DENISSENKO, M.F., PFEIFER, G.P. Sunlight Induces Pyrimidine Dimers Preferentially at 5-Methylcytosine Bases. **Cancer Research**, v. 57, p. 4727-4730, 1997.

VAN DER AUWERA, G.A. et al. From fastq data to high-confidence variant calls: the genome

analysis toolkit best practices pipeline. **Current Protocols in Bioinformatics**. V. 43. p. 11.10.1-11.10.33, 2013.

VENEMA, J., VAN HOFFEN, A., KARCAGI, V., NATARAJAN, A.T., VAN ZEELAND, A.A., MULLENDERS, L.H.F. Xeroderma pigmentosum complementation group C cells remove pyrimidine dimers selectively from the transcribed strand of active genes. **Molecular and Cell Biology**, v. 11, p. 4128-4134, 1991.

WANG, K., LI, M., HAKONARSON, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. **Nucleic Acids Research**, v. 38, p. 1–7, 2010.

WIJNHOVEN, S.W. et al. Age-dependent spontaneous mutagenesis in Xpc mice defective in nucleotide excision repair. **Oncogene**, v. 19, p. 5034–5037, 2000.

WILLIAMSON, C.E. et al. Solar ultraviolet radiation in a changing climate. **Nature Climate Change**, v. 4, p. 434–441, 2014.

YAGURA, T., Makita, K., YAMAMOTO, H., MENCK, C.F.M., SCHUCH, A.P. Biological Sensors for solar ultraviolet radiation. **Sensors**, v. 11, p. 4277-4294, 2017.

YOON, J.H., PRAKASH, L., PRAKASH, S. Highly error-free role of DNA polymerase eta in the replicative bypass of UV-induced pyrimidine dimers in mouse and human cells. **Proceedings of the National Academy of Sciences of the USA**, v. 106, p. 18219–18224, 2009.

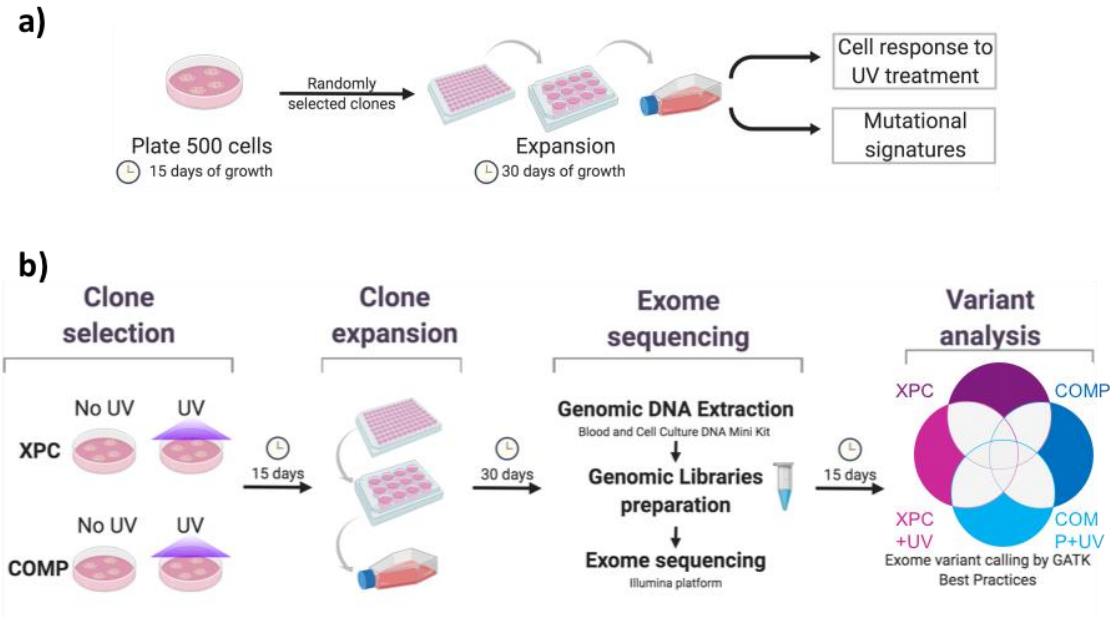
YOON, J.H. et al. Error-Prone Replication through UV Lesions by DNA Polymerase theta Protects against Skin Cancers. **Cell**, v. 176, p. 1295-1309.e15, 2019.

YOU, Y.H., LEE, D.H., YOON, J.H., NAKAJIMA, S., YASUI, A., PFEIFER, G.P. Cyclobutane pyrimidine dimers are responsible for the vast majority of mutations induced by UVB irradiation in mammalian cells. **Journal of Biological Chemistry**, 276, p. 44688-44694, 2001.

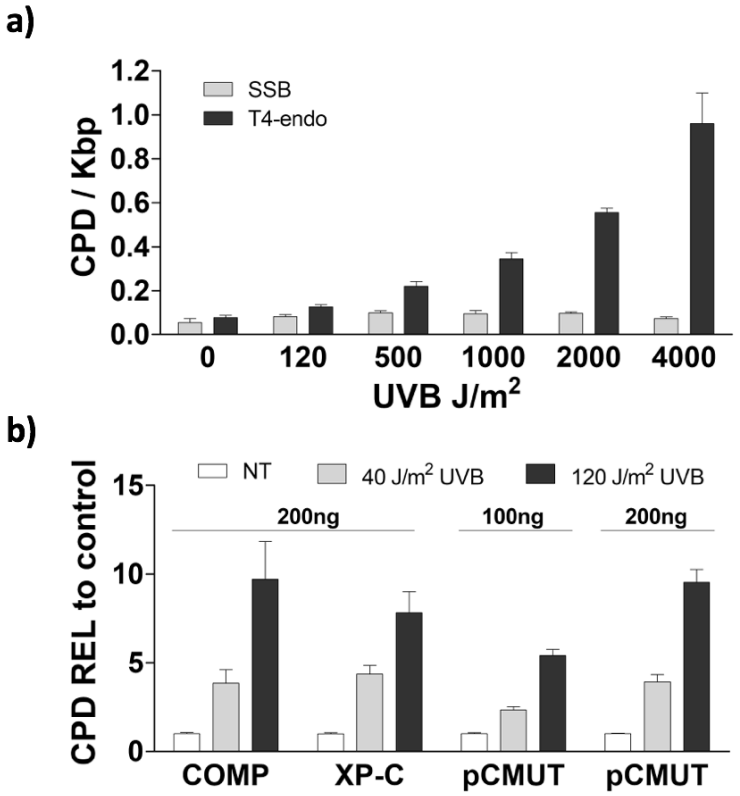
YU, S., TANG, S., MAYER, G.D., COBB, G.P., MAUL, J.D. Interactive effects of ultraviolet-B radiation and pesticide exposure on DNA photo-adduct accumulation and expression of DNA damage and repair genes in *Xenopus laevis* embryos. **Aquatic Toxicology**., v. 159, p. 256–266, 2015.

ZIEGLER, A. et al. Mutation hotspots due to sunlight in the p53 gene of nonmelanoma skin cancers. **Proceedings of the National Academy of Sciences of the USA**, v. 90, p. 4216–20, 1993.

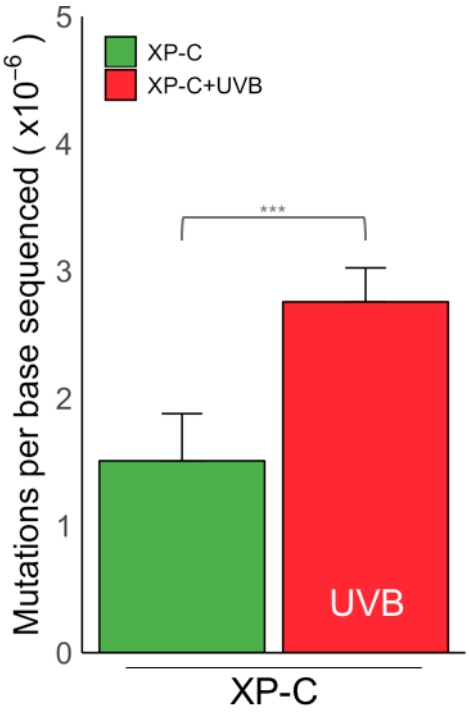
2.7 Supplementary Figures



Supplementary figure S2.1 Cloning strategy used to detect point mutations induced on the first round of replication using whole-exome sequencing. a) Selection of clones in order to obtain a more homogeneous population. XP-C and COMP cells were plated in a very low confluence and allowed to grow by two weeks until the formation of isolated colonies. Clones were randomly selected and transferred manually to a 96 multiwell plates, 1 colony per well. Upon reaching confluence, cells were expanded consecutively into larger plates always renewing the medium every 3 to 4 days. This was done successively until obtaining enough cells to freeze and to continue the experiments. **b)** Experimental design for Mutational signatures. Cells of previously selected clones were plated in a very low confluence, 500 cells in 100 mm Petri dishes, the next day cells were or not irradiated with 120 J/m^2 of UVB light and were allowed to grow as described in a. An 80% confluent 100 mm plate was used for genomic DNA extraction from clones of each cell type and treatment. Library preparation and exon capture were performed using the Agilent's SureSelect QXT Human All Exon v6 and were sequenced in paired-end mode on the Illumina NextSeq or HiSeq platforms. The data analysis was performed based on Genome Analysis Toolkit (GATK) best Practices for exome variant calling. Finally, we selected only exclusive variants by performing comparisons between all clones of the experiment and eliminating variants detected in two or more independent clones to reduce the false positive SNVs.

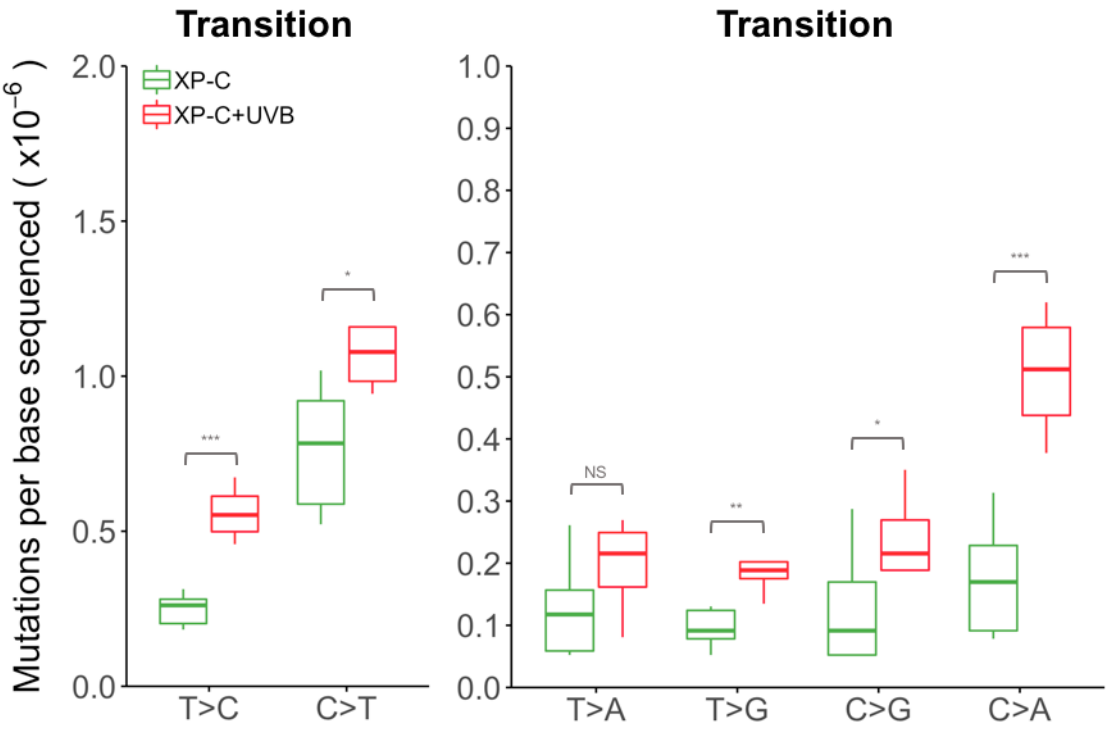


Supplementary figure S2.2 Quantification of CPDs induced by UVB irradiation. a) Plasmid DNA samples were exposed to increasing doses of UVB and then treated or not with T4-endo enzyme that recognizes and cleaves in sites of CPD lesions; **b)** CPD determination by immunoblot assay: amount of CPDs generated by UVB light *in vitro* and *in vivo* samples (fold increase in relation to non-exposed control samples).

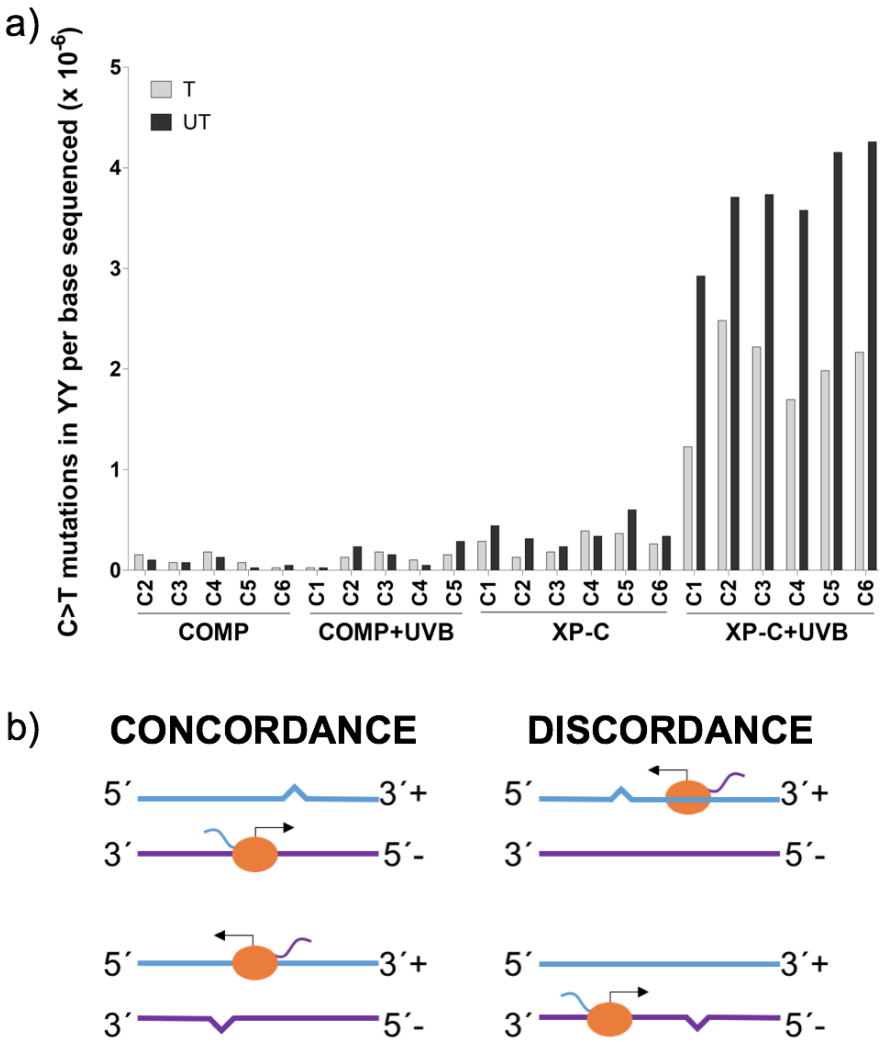


Supplementary figure S2.3 UVB light increases point mutations in XP-C cells, even at a low dose of irradiation. Unique point mutations within exons and splicing sites per sequenced

base (n = 4 or 6 clones per group), after UVB irradiation. The data represent mean and standard deviation (SD). Irradiation dose: 40 J/m^2. Statistical differences were calculated by non-parametric permutation tests: *P* < 0.001 (***), *P* < 0.01 (**), *P* < 0.05 (*), *P* > 0.05 (NS).



Supplementary figure S2.4 Point mutations, according to the nature of the base substitution type: transitions or transversions. The data represent the mean and standard deviation (SD) of exclusive substitutions found within exonic and splicing sites in the sequenced clones for each condition. Cells were irradiated with 40 J/m^2 of UVB. Statistically significant differences were calculated by non-parametric permutation tests: P < 0.001 (***), P < 0.01 (**), P < 0.05 (*), P > 0.05 (NS).



Supplementary figure S2.5 Transcriptional strand bias. a) Frequency of C>T transitions occurring inside the YY motif at transcribed (T) and untranscribed (UT) strands, by each individual sequenced clone. **b)** Graphic representation of mutational motifs found in concordance (motifs found on the same strand as the coding strand, as annotated in RefSeq) and in discordance (motifs found in the non-coding strand, as annotated in RefSeq). The ratio of mutations on concordance:discordance measured the transcriptional strand bias score (SC).

2.8 Supplementary Tables

Supplementary table S2.1 Detailed list of sequenced clones, library preparation kit and sequencing platform used.

Cell line	Treatment	Clon	Exome capture platform	Sequencing platform	Total SNVs
COMP	NO UV	COMP_NT_C2	Agilent - Sureselect QXT V6.	HiSeq 2500 System - Illumina	53
		COMP_NT_C3	Agilent - Sureselect QXT V6.	HiSeq 2500 System - Illumina	46
		COMP_NT_C4	Agilent - Sureselect QXT V6.	HiSeq 2500 System - Illumina	35
		COMP_NT_C5	Agilent - Sureselect QXT V6.	NextSeq 500 Instrument - Illumina	21
		COMP_NT_C6	Agilent - Sureselect QXT V6.	NextSeq 500 Instrument - Illumina	24
	UVB 120J/m2	COMP_B120_C1	Agilent - Sureselect QXT V6.	NextSeq 500 Instrument - Illumina	18
		COMP_B120_C2	Agilent - Sureselect QXT V6.	NextSeq 500 Instrument - Illumina	33
		COMP_B120_C3	Agilent - Sureselect QXT V6.	NextSeq 500 Instrument - Illumina	38
		COMP_B120_C4	Agilent - Sureselect QXT V6.	NextSeq 500 Instrument - Illumina	27
		COMP_B120_C5	Agilent - Sureselect QXT V6.	HiSeq 2500 System - Illumina	46
XP-C	NO UV	XPC_NT_C1	Agilent - Sureselect QXT V6.	HiSeq 2500 System - Illumina	67
		XPC_NT_C2	Agilent - Sureselect QXT V6.	NextSeq 500 Instrument - Illumina	48
		XPC_NT_C3	Agilent - Sureselect QXT V6.	NextSeq 500 Instrument - Illumina	40
		XPC_NT_C4	Agilent - Sureselect QXT V6.	HiSeq 2500 System - Illumina	56
		XPC_NT_C5	Agilent - Sureselect QXT V6.	HiSeq 2500 System - Illumina	80
		XPC_NT_C6	Agilent - Sureselect QXT V6.	HiSeq 2500 System - Illumina	56
	UVB 120J/m2	XPC_B120_C1	Agilent - Sureselect QXT V6.	NextSeq 500 Instrument - Illumina	233
		XPC_B120_C2	Agilent - Sureselect QXT V6.	HiSeq 2500 System - Illumina	341
		XPC_B120_C3	Agilent - Sureselect QXT V6.	NextSeq 500 Instrument - Illumina	307
		XPC_B120_C4	Agilent - Sureselect QXT V6.	NextSeq 500 Instrument - Illumina	270
		XPC_B120_C5	Agilent - Sureselect QXT V6.	NextSeq 500 Instrument - Illumina	317
		XPC_B120_C6	Agilent - Sureselect QXT V6.	HiSeq 2500 System - Illumina	371
XP-C	UVB 40J/m2	XPC_B40_C1	Illumina - Nextera Rapid Capture Exome	NextSeq 500 Instrument - Illumina	90
		XPC_B40_C2	Illumina - Nextera Rapid Capture Exome	NextSeq 500 Instrument - Illumina	105
		XPC_B40_C3	Illumina - Nextera Rapid Capture Exome	NextSeq 500 Instrument - Illumina	114
		XPC_B40_C4	Illumina - Nextera Rapid Capture Exome	NextSeq 500 Instrument - Illumina	100

Supplementary table S2.2 Sequencing alignment statistics

Cell line	**COMP**									
Treatment	**NO UV**					**UVB (120 J/m^2)**				
Codigo clone	COMP_NT_C2	COMP_NT_C3	COMP_NT_C4	COMP_NT_C5	COMP_NT_C6	COMP_B120_C1	COMP_B120_C2	COMP_B120_C3	COMP_B120_C4	COMP_B120_C5
kit - platform	Agilent-HiSeq	Agilent-HiSeq	Agilent-HiSeq	Agilent-NextSeq	Agilent-NextSeq	Agilent-NextSeq	Agilent-NextSeq	Agilent-NextSeq	Agilent-NextSeq	Agilent-HiSeq
% duplicate reads	17.8%	18.5%	19.9%	14.8%	15.1%	14.9%	14.8%	15.9%	15.6%	17.0%
Total reads	44804824	58167716	66268198	62124480	61051060	60037976	59169310	66096346	66912690	45729726
Unique reads	36822355	47433644	53086800	52927668	51810367	51081491	50384913	55567465	56460246	37939746
Number of reads mapped	36715333	47114547	52712128	51796270	50699224	50057517	49379138	54388547	55232927	37839638
% reads mapped	99.7%	99.3%	99.3%	97.9%	97.9%	98.0%	98.0%	97.9%	97.8%	99.7%
Bases on target	3066514880	3856312548	4383608823	2269799358	2219378367	2197645485	2116898507	2432797059	2437800404	2979737569
% bases on target	93.0%	93.4%	93.3%	88.8%	88.9%	88.9%	86.6%	88.9%	88.3%	92.0%
Bases off target	376078192	468759881	539403251	505891036	491752045	485181105	573522329	530121463	568949595	450175496
% off target	7.0%	6.6%	6.7%	11.2%	11.1%	11.1%	13.4%	11.1%	11.7%	8.0%
% of target bp not covered	3.2%	3.1%	2.9%	1.3%	1.3%	1.3%	1.3%	1.3%	1.3%	3.1%
% of target bp covered at ≥ 1X	96.4%	96.6%	96.8%	97.6%	97.7%	97.6%	97.6%	97.7%	97.7%	96.6%
% of target bp covered at ≥ 10X	83.4%	86.3%	87.7%	86.1%	86.6%	86.0%	84.4%	87.9%	87.2%	83.5%
% of target bp covered at ≥ 20X	68.1%	74.1%	76.7%	66.6%	66.5%	65.8%	63.0%	69.6%	68.7%	68.1%
Average depth of coverage	80	101	114	61	60	59	57	66	66	78
Raw variants called	118519	143216	149855	81787	82373	81329	80683	83163	84654	141278
Novel SNPs	7398	8832	9530	2376	2335	2283	2393	2537	2582	9602
Percent SNPs (dbSNP)	93.8%	93.8%	93.6%	97.1%	97.2%	97.2%	97.0%	96.9%	96.9%	93.2%
Raw Indels called	16355	17564	19013	6028	6287	6108	6001	6158	6432	17228
Novel Indels	2696	2261	2580	339	380	361	335	377	406	2381
Percent Indels (dbSNP)	83.5%	87.1%	86.4%	94.4%	94.0%	94.1%	94.4%	93.9%	93.7%	86.2%

Continuation supplementary table 2.2 Sequencing alignment statistics

Cell line	**XPC**									
Treatment	NO UV						**UVB (120 J/m^2)**			
Codigo clone	XPC_NT_C1	XPC_NT_C2	XPC_NT_C3	XPC_NT_C4	XPC_NT_C5	XPC_NT_C6	XPC_B120_C1	XPC_B120_C2	XPC_B120_C3	XPC_B120_C4
kit - platform	Agilent-HiSeq	Agilent-NextSeq	Agilent-NextSeq	Agilent-HiSeq	Agilent-HiSeq	Agilent-HiSeq	Agilent-NextSeq	Agilent-HiSeq	Agilent-NextSeq	Agilent-NextSeq
% duplicate reads	19.2%	13.8%	14.9%	17.0%	28.5%	16.3%	12.9%	21.9%	14.4%	15.4%
Total reads	66252510	55853698	60024760	41917512	98074826	44827128	48409004	52543292	55773134	66427354
Unique reads	53520507	48157191	51081473	34796111	70126092	37523451	42169366	41062077	47717634	56230251
Number of reads mapped	53047645	47105659	50013116	34461596	69914558	37424291	41326393	40950868	46702345	54948681
% reads mapped	99.1%	97.8%	97.9%	99.0%	99.7%	99.7%	98.0%	99.7%	97.9%	97.7%
Bases on target	4326533084	2023636863	2216766493	2815149094	6755706268	3002745706	1767359861	3651545408	1969349037	2336557334
% bases on target	93.1%	88.0%	88.2%	94.0%	92.9%	92.5%	88.9%	92.8%	86.1%	86.7%
Bases off target	557889874	483809862	514409279	310788866	850404633	410963804	388543370	447883239	561882979	637351578
% off target	6.9%	12.0%	11.8%	6.0%	7.1%	7.5%	11.1%	7.2%	13.9%	13.3%
% of target bp not covered	2.9%	1.3%	1.3%	3.1%	2.8%	3.4%	1.4%	5.0%	1.3%	1.2%
% of target bp covered at ≥ 1X	96.8%	97.6%	97.6%	96.5%	96.9%	96.4%	97.5%	94.8%	97.6%	97.8%
% of target bp covered at ≥ 10X	87.5%	85.0%	85.7%	84.0%	90.4%	82.5%	82.2%	75.2%	82.4%	89.0%
% of target bp covered at ≥ 20X	76.3%	63.0%	65.2%	68.4%	81.7%	67.3%	57.5%	60.7%	59.3%	70.7%
Average depth of coverage	113	55	60	74	176	78	48	95	53	63
Raw variants called	154347	83964	80015	119841	173269	131090	80927	117651	79788	84441
Novel SNPs	9664	2537	2383	6307	13605	8468	2969	10127	2891	3108
Percent SNPs (dbSNP)	93.7%	97.0%	97.0%	94.7%	92.1%	93.5%	96.3%	91.4%	96.4%	96.3%
Raw Indels called	19036	6624	5976	13980	25809	17075	6269	15857	5718	6490
Novel Indels	2505	373	334	1757	4464	2643	354	2992	333	358
Percent Indels (dbSNP)	86.8%	94.4%	94.4%	87.4%	82.7%	84.5%	94.4%	81.1%	94.2%	94.5%

Continuation supplementary table 2.2 Sequencing alignment statistics

Cell line	**XPC**					
Treatment	**UVB (120 J/m^2)**		**UVB (40 J/m^2)**			
Codigo clone	XPC_B120_C5	XPC_B120_C6	XPC_B40_C1	XPC_B40_C2	XPC_B40_C3	XPC_B40_C4
kit - platform	Agilent-NextSeq	Agilent-HiSeq	illumina-NextSeq	illumina-NextSeq	illumina-NextSeq	illumina-NextSeq
% duplicate reads	15.7%	15.1%	31.6%	30.4%	28.6%	30.9%
Total reads	62592238	40710836	135504962	60647580	84335732	77840514
Unique reads	52780652	34573687	92712692	42215033	60203185	53788322
Number of reads mapped	51604723	34495588	92178429	41879720	59724700	53219201
% reads mapped	97.8%	99.8%	99.4%	99.2%	99.2%	98.9%
Bases on target	2292987910	2717159272	7846149357	3093712736	4520157990	4031799291
% bases on target	87.9%	7.7%	69.2%	69.1%	68.6%	68.4%
Bases off target	547337898	380705828	5106433277	1995796738	3079329683	2659231845
% off target	12.1%	7.7%	30.8%	30.9%	31.4%	31.6%
% of target bp not covered	1.3%	3.6%	0.9%	1.0%	0.9%	1.0%
% of target bp covered at ≥ 1X	97.6%	96.2%	98.7%	98.5%	98.7%	98.5%
% of target bp covered at ≥ 10X	85.2%	80.3%	96.7%	89.9%	95.0%	92.7%
% of target bp covered at ≥ 20X	65.1%	64.2%	92.9%	73.7%	86.3%	81.7%
Average depth of coverage	34	71	211	83	122	109
Raw variants called	80457	126748	407396	156809	243932	200827
Novel SNPs	2967	9174	41026	17871	25580	24456
Percent SNPs (dbSNP)	96.3%	92.8%	89.9%	88.6%	89.5%	87.8%
Raw Indels called	5775	15787	76265	32833	47931	45274
Novel Indels	336	2422	12864	6856	9278	10384
Percent Indels (dbSNP)	94.2%	84.7%	83.1%	79.1%	80.6%	77.1%

Supplementary table 2.3 Transcriptional strand bias score (SC) associated to mutagenic motifs (YY) after UVB irradiation.

Cell line	SC mean (SD)	Mutations
COMP	0.9 (0.6)	54
COMP+UVB	1.2 (0.6)	69
XPC	1.5 (0.5)	176
XPC+UVB	2.0 (0.3)	1393

Transcriptional strand bias score (SC): measured by concordance:discordance ratio for the number of mutational motifs detected in concordance that are not subjected to TCR (motifs found in the coding strand, as annotated in RefSeq), and motifs detected in discordance that are potentially subjected to TCR (motifs found in the non-coding strand, as annotated in RefSeq). A SC of 1 means no transcriptional strand bias. Values > 1 means more motifs in concordance. Values < 1 means more motifs in discordance. The data represent mean and standard deviation (SD). The YY motif was used for this analysis, Y = IUPAC code for T or C nucleotides.

CHAPTER 3 – UVA light induced mutagenesis in the exome of human nucleotide excision repair deficient cells

Nathalia Quintero Ruiz[1], Camila Corradi[1], Natalia Cestari Moreno[1], Tiago Antonio de Souza[1], Carlos Frederico Martins Menck[1*].

[1]DNA repair Laboratory, Department of Microbiology, Institute of Biomedical Sciences, São Paulo University - USP, São Paulo, SP.

*Corresponding author
E-mail: cfmmenck@usp.br

3.1 Abstract

Skin cancer is associated with genetic mutations caused by exposure to sunlight radiation, mainly due to ultraviolet (UV) component that damages the DNA molecule. Although UVA has lower energy, it is the main solar UV component that reach the Earth´s surface. However, UVA-induced mutagenesis and its role in the induction of skin cancer are still unclear. A strategy of whole exome sequencing of clones, from human XP-C cells (deficient on nucleotide excision repair, NER), was used to characterize somatic mutations induced by the UVA exposure. DNA sequence analysis of UVA-irradiated XP-C cells revealed a significative increment in the frequency of mutations in almost all types of base substitutions, but mostly at C>T transitions in the CCN and TCN trinucleotide context, which are potential sites for pyrimidine dimer formation. The C>T mutation mainly occurs at the 3´ base of the 5´TC dimer and the CC>TT tandem mutations were also enriched. Additionally, it was possible to recover on the XP-C sequenced clones the SBS7 signature established in the COSMIC catalogue and related to skin cancer mutations. On the other hand, C>A transversions, traditionally related to oxidized guanines, was the second most frequently induced mutation. Therefore, this study provides evidence that pyrimidine dimers are the main type of lesion contributing to UVA induced mutagenesis on NER deficient human cells and that UVA generates mutational signatures similar to UVB irradiation.

Keywords: Ultraviolet-A radiation, Xeroderma Pigmentosum, XP-C, mutagenesis, exome, NGS, mutational signature, nucleotide excision repair.

3.2 Introduction

Exposure to solar radiation cause sunburn, photo-aging, ocular damage on the cornea and lens, and melanoma and non-melanoma skin cancer (Diffey 1991, de Gruijl et al. 1993, Armstrong & Kricker 2001, Matsumura & Ananthaswamy 2002). The ultraviolet component of sunlight is commonly divided into three spectral ranges based on the wavelength and therefore on the energy it contains: UVA (315-400 nm), UVB (280-315 nm), and UVC (200-280 nm), each one with different biological effects that strongly depend on light energy. However, just UVA and UVB are considered biologically relevant, since wavelengths below 300 nm are filtered by the stratospheric ozone layer (Floyd et al. 2002, Williamson et al. 2014). Thereby, approximately 90-95% of the total spectrum that reaches the Earth´s surface corresponds to UVA radiation, while the remaining 5-10% is represented by UVB (Sage et al. 2012). It has been widely established that UVB has an important participation in the production of vitamin D, as well as in the aforementioned solar deleterious effects and induces the direct formation of photoproducts, since DNA directly absorb incident UVB photons, which lead to DNA mutagenesis and skin cancer (Diffey 1991, Ravanat et al. 2001, Cadet et al. 2005, Cadet & Douki 2018). In contrast, it is still unclear the contribution of the UVA component to the solar deleterious effects, although it may be implicated on mutagenesis and skin cancer (Sage et al. 2012).

For a long time, UVA radiation was considered relatively harmless due to its lower energy properties, consequently the monitoring of UVA levels received less attention compared to UVB (Diffey 1991) and its biological effects were underestimated. Likewise, the use of artificial tanning chambers (Fears et al. 2011, Boniol et al. 2012) and longer exposure times to sunlight by the use of UVB-absorbing sunscreens that prevented sunburn (Garland et al. 1993, Gasparro 2000) were up the average exposure of human populations to UVA. Epidemiologic studies related sunbeds exposure with an increment in the risk of developing skin cancer (Fears et al. 2011, Boniol et al. 2012), mainly melanoma (Westerdahl 2000). However, artificial tanning chambers have a broad emission spectrum, so this relationship cannot be attributed solely to the UVA effect (Sage et al. 2012). On the other hand, there is increasing evidence of the UVA contribution to the development of skin tumors by using hairless mice as models (de Gruijl et al. 1993, Berg et al. 1993, de Laat et al. 1997, Kranen et al. 1997, Sterenborg & van der Leun 1990). Studies in a hybrid fish (Setlow et al. 1993) and in the South American opossum (Ley 2001) support the idea

that UVA plays a major role in the induction of melanoma skin cancer. However, in mouse model UVA increased the metastatic ability of melanoma (Pastila & Leszczynski 2005), but did not induce melanoma (de Fabo et al. 2004). Also, Mitchell and collaborators (2010) failed to replicate the results on fish. There are clear evidences for the involvement of UVA rays in skin cancer development, with an estimated contribution of 10-20% to the carcinogenic dose of sunlight (de Laat et al. 1997). However, it is still unclear how UVA light contributes to this process.

A series of work demonstrated that UVA affect different biomolecules in exposed cells generating reactive oxygen species (ROS), lipid peroxidation, oxidized proteins and DNA damage (Ridley et al. 2009, Sage et al. 2012, Marionnet et al. 2014, Schuch et al. 2017). Harmful effects of UV can be related both with direct and indirect induced damage to DNA, which lead to different and specific types of mutations. Cyclobutane pyrimidine dimers (CPDs) and pyrimidine-pyrimidone (6,4) photoproducts (6-4PPs) are generated at dipyrimidine sites by direct excitation of the DNA molecule. When un-repaired, both types of lesions mainly generate C>T and CC>TT mutations due to mispairing of A with C (Ziegler et al. 1993, Rastogi et al. 2010). Indirect excitation of the molecule is performed through photosensitization reactions, which can involve oxygen, and generate oxidized bases (Ravanat et al. 2001, Pfeifer et al. 2005, Rastogi et al. 2010, Sage et al. 2012). Recently, the formation of oxidized bases was demonstrated also by direct excitation of the DNA molecule (Yagura et al., 2017). The most abundant oxidized base is the 8-oxoguanine (8oxoG), since apparently guanine is the base that has the lowest reduction potential (Steenken and Jovanovic 1997). When unrepaired, 8oxoG pairs with adenine generating transversions of the G>T (C>A) type (Epe 1991, Cheng et al. 1992).

Direct UVA energy absorption by the DNA molecule is weak (Sutherland and Griffin 1981). Therefore, it was believed that mutations generated by UVA were only indirect and a consequence of oxidative stress (Mouret et al. 2006, Nichols and Katiyar 2010). However, several studies detected the induction of CPDs, mainly in the dipyrimidine sites TT followed in less proportion by TC and CT sites, rather than the induction of 8oxoG (Douki et al. 2003, Rochette et al. 2003, Courdavault et al. 2004b, Mouret et al. 2006). A previous *in vivo* study had detected thymine dimers in human skin (Burren et al. 1998). Even the induction of 6-4PP, in addition to CPDs, into the DNA molecule irradiated by UVA in the absence of any photosensitizer was

demonstrated (Schuch et al. 2009), as well as on human cells deficient in DNA repair (Cortat et al. 2013).

Mutagenesis studies are also contradictory. Some of them reported the G>T transversion as the main base substitution in the spectrum of UVA-induced mutations using rodent cells or human keratinocytes, and related the mutation with oxidative damage (Persson et al. 2002, Besaratinia et al. 2004). However, other studies indicate that 8oxoG does not have a significant contribution to UVA-induced mutagenesis and suggest a main role of pyrimidine dimers (Kappes and Runger 2005). For other groups, the T>G transversion was the main induced mutation in the *aprt* locus of CHO cells. This mutation was attributed to thymine dimers (Drobetsky et al. 1995, Rochette et al. 2003) and was even proposed as a fingerprint for UVA light (Agar et al. 2004). On the other hand, the C>T transition was reported as the main type of mutation generated both for UVA and UVB lights on the *hprt* gene of primary human fibroblasts (Kappes et al. 2006), in the *aprt* locus of NER deficient CHO cells (Sage et al.1996) and, more recently, by exome analysis on XP-V deficient cells (Moreno et al. 2020). In addition, these results in cell culture models were confirmed *in vivo* by irradiation of mouse skin (Ikehata et al. 2008). Thus, the C>T transition has been proposed as a characteristic signature of UV radiation in general, and not specifically for UVB and UVC wavelengths (Runger 2008, Sage et al. 2012).

Therefore, the role of UVA light in the generation of mutations and how it contributes to the development of skin cancer are matters of debate. The main objective of this work is to establish the profile of mutagenesis, type and frequency of mutations, induced by UVA in human skin fibroblast through the identification of single nucleotide variants (SNVs) in the exome of two cell lines, one deficient and one proficient in DNA repair. The XP-C cell line was selected as a model, since it is deficient in nucleotide excision repair (NER), as this pathway is the main mechanism used by human cells to remove UV induced damage (Sancar 1996, Menck & Munford 2014). Moreover, xeroderma patients suffer high sunlight sensitivity and have ten thousand more chances to develop skin cancer on sunlight-exposed areas due to a deficiency in one of the NER proteins, establishing a strong correlation between DNA damage and cancer (Cleaver et al. 2009, Menck & Munford 2014, Bowden et al. 2015). XPC protein is part of the complex that recognizes damage on the NER pathway (Menck & Munford 2014), and XP-C patients are the most prevalent among XP patients in Europe, USA and North Africa (Stary & Sarasin 2002, Leite et al. 2009, Soufir et al.

2010). More recently, they also have been reported as highly frequent in Brazil (Santiago et al., 2020). Also, we take advantage of the everyday more accessible next generation sequencing (NGS) technology (van Dijk et al. 2014, Goodwin et al. 2016), which allows the investigation of base substitutions in the entire exome or genome. This analysis generates a more global picture of the UVA effect on cells and avoids the bias observed in the analysis of only some genes that could be under selection and generate cluster information (Alexandrov & Stratton 2014). Additionally, the access to a higher number of data allows the analysis of the sequence context of base substitutions and of mutation signatures generated by UVA irradiation with higher confidence. Mutational signatures are patterns of mutagenesis useful for unravelling and understanding the underlying mutagenic process caused by different agents. Thus, we expect to contribute to a better understanding of the origin of skin tumors and to clarify the role of UVA in this process.

3.3 Material and Methods

3.3.1 Cell lineages and culture conditions

Three human fibroblasts immortalized by SV40 virus were used: the XP4PA (XP-C) (Daya-Grosjean et al. 1987), a NER deficient skin fibroblast cell line that carries a ΔTG mutation in exon 9 of the XPC gene (c1643-1644delTG), which translates into an inactive truncated protein (p.Val548AlafsX25) (Li et al. 1993, Soufir et al. 2010). As an isogenic control, the same cell line but complemented with the XPC gene, by a lentivirus-based strategy, was employed in these experiments (XP4PA complemented, clone 1-1, named COMP cells). The XP-C cells were kindly provided by Drs. Alain Sarasin and Anne Stary (Institut Gustave Roussy, France). Cultures were maintained at 37°C in a humid atmosphere, 5% of CO2 on Dulbecco's Modified Eagle Medium High Glucose (DMEM, LGC Biotechnologies, Cotia, SP, Brazil) supplemented with 10% fetal bovine serum (FBS, Cultilab, Campinas, SP, Brazil) and 1% antibiotic/antimycotic solution (0.1 mg/ml penicillin, 0.1 mg/ml streptomycin and 0.25 mg/ml fungizone) (Life Technologies, Carlsbad, CA, USA). The cells grew adhered to the surface of plastic bottles and were replated in periods of 3 to 4 days.

3.3.2 Selection of clones for sequencing

To identify the UVA induced mutations by NGS, a two-step cloning strategy was established and improved by our group. This strategy allows the detection of mutations

induced on the first round of replication (for details see Moreno et al. 2020, this work chapter 2). The first cloning step was carried out to homogenize the population's genetic background and obtain parental clones. Cells of each cell line were plated in a very low density ($0.5x10^3$ cells in 100 mm diameter plates) and allowed to grow by two weeks until the formation of isolated colonies that were randomly selected and transferred manually into 96-well plates. Then, clones were expanded consecutively into larger plates until obtaining enough material for the functional and mutagenic experiments. The second cloning step was carried out from one of the parental clones selected in the first step. Cells from the selected clone were plated ($0.5x10^3$ cells in 100 mm diameter plates), submitted to UVA irradiation and allowed to form the subclones that were randomly selected and expanded to obtain enough material for DNA extraction. This second step took approximately 45 days.

3.3.3 Irradiation conditions

UVA radiation was carried out with a Osram Ultramed FDA KY10s 1000 W (Munich, Germany) lamp in a dose rate average of 67.2 J/m2s, associated with Schott BG39 filter with 3 mm thick (Schott Glass, Mainz, Rheinland-Pfalz, Germany), which shuts off wavelengths lower than 320 nm, preventing cells from receiving any remaining UVB or UVC radiation (Schuch et al. 2009). The intensity of lamp emission was measured at each experiment with a VLX 3 W UV radiometer (Vilber Lourmat, Torcy, France) and the exposure time adjusted, so the cells received the desired dose. All experiments were performed in a time window between 16 to 24 h after plating, culture medium was discarded, cells were washed with PBS and irradiated in PBS with calcium. Non-irradiated control plates were submitted to the same process and maintained in the dark during the UV irradiation time.

3.3.4 Clonogenic assay

Cells were plated at low density, $1.5x10^3$ cells in 60 mm diameter plates, and, 16 h later, the plates were irradiated (or not). Cells were cultivated for two weeks with medium replacement every 4 or 5 days, until the formation of isolated colonies. Colonies were fixed with formaldehyde (10%) and stained with violet crystal (1%). Differences in the cell viability after irradiation were estimated in relation to the non-irradiated control group.

3.3.5 Flow cytometry

$1x10^5$ cells were plated in 60 mm diameter plates, irradiated (16 h later) and allowed to recover for 24 and 72 h. Both, detached dying cells and trypsinized adherent cells were collected, fixed with formaldehyde 1% on ice, washed with PBS, resuspended in cold ethanol 70% and stored at -20°C for, at least, 24 h. Then, cells were permeabilized with PBS-T-BSA buffer (0.2% Triton X- 100, 1% bovine serum albumin, BSA-Sigma–Aldrich, Saint Louis, MO, USA- in PBS), incubated overnight at 4ºC with the antibody for γH2AX (05-636 Merck Millipore, Burlington, MA, USA- at 1:500 dilution) and 1 h at room temperature in dark with the secondary antibody (anti-mouse fluorescein iso-thiocyanate (FITC) antibody, Sigma–Aldrich, at 1:200 dilution). Thereafter, samples were washed and incubated in propidium iodide (PI) solution (200 µg/mL RNAse A, Invitrogen-Life Technologies, 20 µg/ml PI, 0.1% (v/v) Triton X-100 in filtered PBS) for 30 min at room temperature, avoiding the incidence of direct light. Finally, samples were washed and resuspended in filtered PBS.

3.3.6 Whole exome sequencing (WES)

Genomic DNA was extracted using the Blood & cell culture DNA mini kit (Qiagen, Hilden, Germany), adhering the manufacturer's protocol. Genomic libraries were built using one of the following commercial human exome capture platforms: i) the Illumina's Nextera Rapid Capture Exome protocol (San Diego, CA, USA), ii) the Agilent's SureSelect QXT Human All Exon v6 (Agilent Technologies, Santa Clara, CA, USA), and iii) the genomic library preparation using the TruSight One Sequencing kit (Illumina, San Diego, CA, USA) with capture performed using the xGen hybridization capture of DNA libraries, IDT (IDT, Coralville, IA, EUA). Always following the protocol provided by the respective manufacturer. Libraries were sequenced in paired-end mode on the Illumina Nextseq or HiSeq platform (Illumina, San Diego, California, EUA). The details about kit used for DNA library preparation, platform of sequencing, coverage and statistics of sequencing for all subclones are provided in Supplementary tables S3.1 and S3.2.

3.3.7 Alignment and calling of variants

Best practices for exome variant calling of the Genome Analysis Toolkit (GATK) were followed for the data analyses (Depristo et al. 2011, Van der Auwera et al. 2013). Briefly, short reads were aligned to human UCSC hg19/ucsc.hg19 reference genome

using the burrows-Wheeler aligner (BWA) (Li and Durbin 2010) and duplicates marked with MarkDuplicates (Picard). Remaining reads were realigned around indels, and sequencing errors and experimental artifacts were corrected by base quality score recalibration (BQSR) method. HaplotipeCaller was used to identify raw variants relative to the reference genome and, to refine variants and reduce false positives we apply the variant quality score recalibration (VQSR) method. The annotation of SNVs was made by ANNOVAR (Wang et al. 2010) using RefSeq (O´leary et al. 2016), including a filter to select only SNVs annotated as “exonic” or “splicing”. Then, a call of exclusive variants for each subclone was performed using VCFTools (Danecek et al. 2011). For this, SNVs in all subclones obtained from the isogenic XP-C and COMP cell lines were compared, taking into account position and nucleotide change, and variants occurring in more than one subclone were excluded.

3.3.8 Exploratory analysis of exclusive variants

General patterns of nucleotide changes were explored using Woland (available at https://github.com/tiagoantonio/woland; Souza et al. in preparation). The SNVs generated were considered in the six possible types of base substitutions (C:G>A:T, C:G>G:C, C:G>T:A, T:A>A:T, T:A>C:G and T:A>G:C; which refers to the pyrimidine of the canonical Watson-Crick base pair). The associated motif sequence pre-established for potential lesions induced by irradiation was considered for the identification of both transcribed and non-transcribed DNA strand. In a general way, the mutated base (X) was always located in the middle of the context sequence, having three nucleotides in the 5´ and the 3´ flanking region of the mutated position, NNNXNNN. For the analysis of pyrimidine dimers, the YY and YCG motifs were explored. The YY motif represent any combination of adjacent pyrimidine bases, and this motif was investigated for the first (NNNYYNN) and the second base (NNYYNNN) mutated. The YCG motif explores whether the dimer includes a potentially methylated cytosine (NNTCGNN, NNCCGNN) (Ikehata et al. 2013). The RGR motif (NNRGRNN) has been related to 8oxoG, since recognition and remotion of 8oxoG seems to be more efficient in a pyrimidine than in purine rich sequence (Hatahet et al. 1998, Sassa et al. 2012). The somatic spectra analysis of exclusive variants in a tri-nucleotide context (which consider 96 possible context sequences for substitutions, the six possible substitution types and their immediately 5´and 3´ sequence context) and the mutational signature analyses (Alexandrov et al. 2013a, 2020) were performed using the

SomaticSignatures (Gehring et al. 2015), an implementation of the non-negative matrix factorization (NMF) method in R (Nik-Zainal et al. 2012, Alexandrov et al. 2013b). The graphical representation of context sequences, where exclusive variants occur, was created by the pLogo generator using exome as background (available at: https://plogo.uconn.edu/, O´Shea et al. 2013), in which the residue heights are scaled relative to their statistical significance. Finally, the deconstructSigs package was used to evaluate the published signatures from Sanger COSMIC in our data (Rosenthal et al. 2016).

3.3.9 Statistical analyses

We used a non-parametric permutation test to infer group differences under the null hypothesis that mean value of the experimental group is not greater than the mean value of the control group. The mean values of both groups generated using Monte-Carlo simulated data from a total of 4999 permutations, were compared and the p-values calculated. Considering a one-sided test, we determine P-values <0.001 as highly statistically significant (***), 0.01 < P-values as high statistically significant (**), P-values < 0.05 as statistically significant (*) and P-values > 0.05 as non-significant (ns).

3.4 Results

3.4.1 Selection of parental clones and cell responses to UVA light

In the first cloning step, single clones of the isogenic cell lines were randomly selected, in order to homogenize the cells genetic background and, thus, achieve a more accurate identification of mutations generated by UVA exposure. The parental clones were identified as XP-C and COMP clones, and were used for the functional and mutagenic experiments.

First, the parental clones were characterized for their response to UVA. Survival rates of clones were determined by the clonogenic assay (Figure 3.1a), which measures cells' viability and capacity to divide and replicate. XP-C cells were more sensitive than COMP cells to UVA irradiation. A dose of 60 kJ/m^2 was selected for next experiments, since this is the highest dose that allows recovering enough colonies required to continue the UVA induced mutagenesis experiments. This dose can be considered environmentally relevant, as it corresponds, approximately, to twenty minutes of exposure to sunlight at midday during summer in a tropical latitude (Schuch

et al. 2012). UVA induces in XP-C cells a significant increase in genotoxic stress (measured as H2AX phosphorylation), while in COMP cells the increase recorded was statistically irrelevant (Figure 3.1b). Damage induced by UVA did not cause a cell cycle arrest of these cells (Figure 3.1c), but generates an increase in apoptosis levels on both cell lines (Figure 3.1d) detected by the sub-G1 content, which indicates nuclei fragmentation.

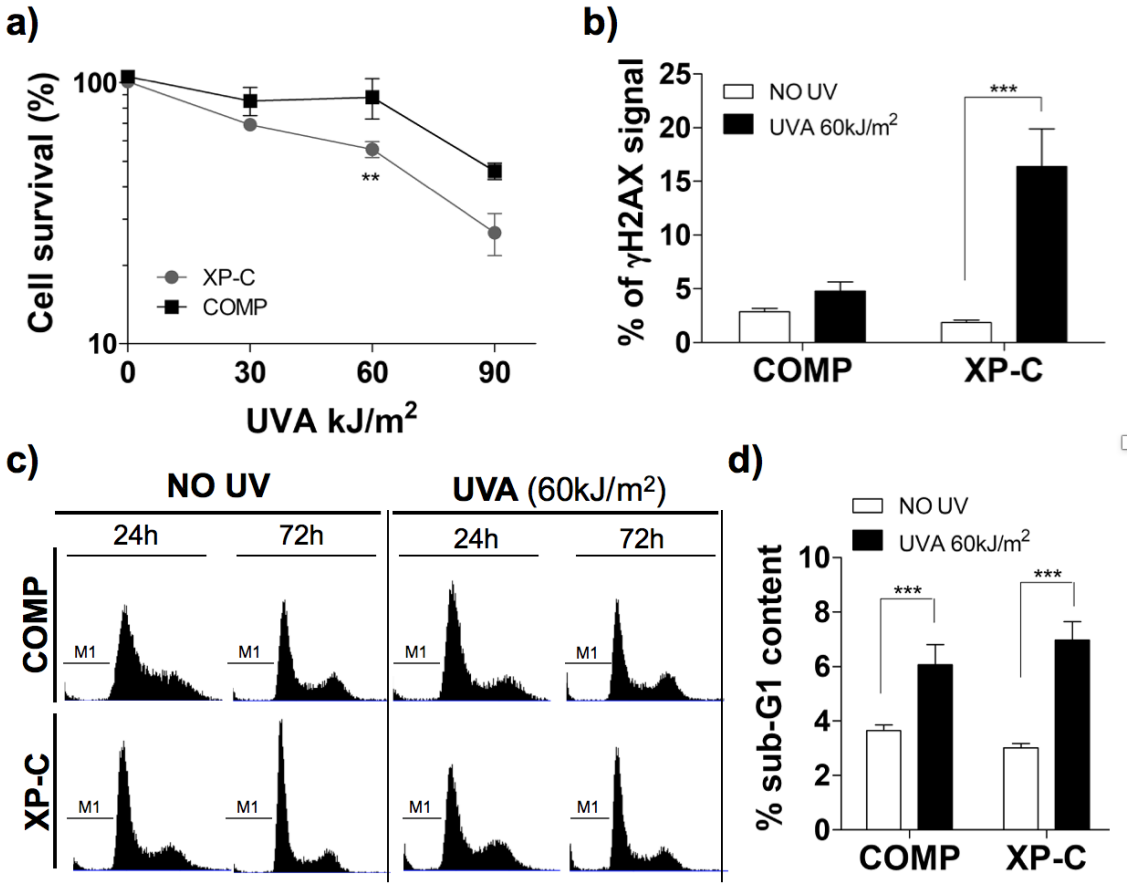


Figure 3.1. Responses of the parental clones to UVA irradiation. a) Clonogenic assay: dose response curves of cells exposed to UVA light, at the indicated doses; **b)** induction of H2AX histone phosphorylation (percentage) 24 h after UVA exposure; **c)** cell cycle progression 24 and 72 h after UVA irradiation: representative histograms; M1 indicates sub-G1 cell population; **d)** induction of sub-G1 content (indicative of apoptosis) induced 72 h after UVA irradiation ($P < 0.001$ (***), $P < 0.01$

3.4.2 UVA light induces an increase in the frequency of point mutations in NER deficient skin fibroblasts

A second cloning step was conducted to study the mutagenic potential of UVA light in skin fibroblasts. Subclones were obtained from the selected parental clones that were plated in very low densities, irradiated with 60 kJ/m^2 of UVA 16 h after plating,

with the surviving clones expanded and whole-exome sequenced. A total of 28 subclones were sequenced; sixteen of XP-C, eight non irradiated and eight irradiated, as well as twelve of COMP, six non irradiated and six irradiated, for details see Supplementary Table S3.1.

After identification of exclusive single nucleotide variation (SNVs) for each clone, the data was carefully analysed. UVA irradiation increased the mutation frequency (SNVs per 10^6 bases sequenced) by 3.6 times in XP-C subclones, while for the isogenic control COMP subclones no significant increase was detected (Figure 3.2a). In the absence of any irradiation, the lack of XPC protein led to a small but significant increase, by 1.7-fold, in the frequency of mutation when compared to complemented cells. Transitions were predominant over transversions in all conditions, but a significant increase in this type of base substitution was detected only in UVA irradiated XP-C subclones (Figure 3.2b). These results highlight the mutagenic capacity of UVA wavelength in NER deficient cells.

3.4.3 Base substitutions and sequence context for the detected mutations

To understand where mutations occur, SNVs generated in the six possible types of base substitutions (C:G>A:T, C:G>G:C, C:G>T:A, T:A>A:T, T:A>C:G and T:A>G:C) were investigated. Additionally, the somatic spectrum of point nucleotide mutations in the tri-nucleotide contexts was analysed to check for the sequence context.

UVA irradiation increased almost all types of base substitutions in XP-C subclones, but with a clear preference for C>T transitions, followed by C>A transversions with an increase rate of 4.9 and 3.3 times, respectively (Figure 3.3). The frequency of the T>C (2.1 times), T>A (2.3 times) and C>G (2.1 times) base substitutions were also significantly increased, while the T>G transversion, previously proposed as a fingerprint mutation for UVA light, was the only type of base substitution without significant increase detected. No increase in the mutation frequency, for any of the types of base substitution, was detected when DNA repair proficient cells (COMP) were irradiated. Finally, the increment in the basal mutation frequency of XP-C subclones respect to its isogenic control, seems to be explained by a significant increase in the C>T (2.2 times) and C>G (1.3 times) base substitutions.

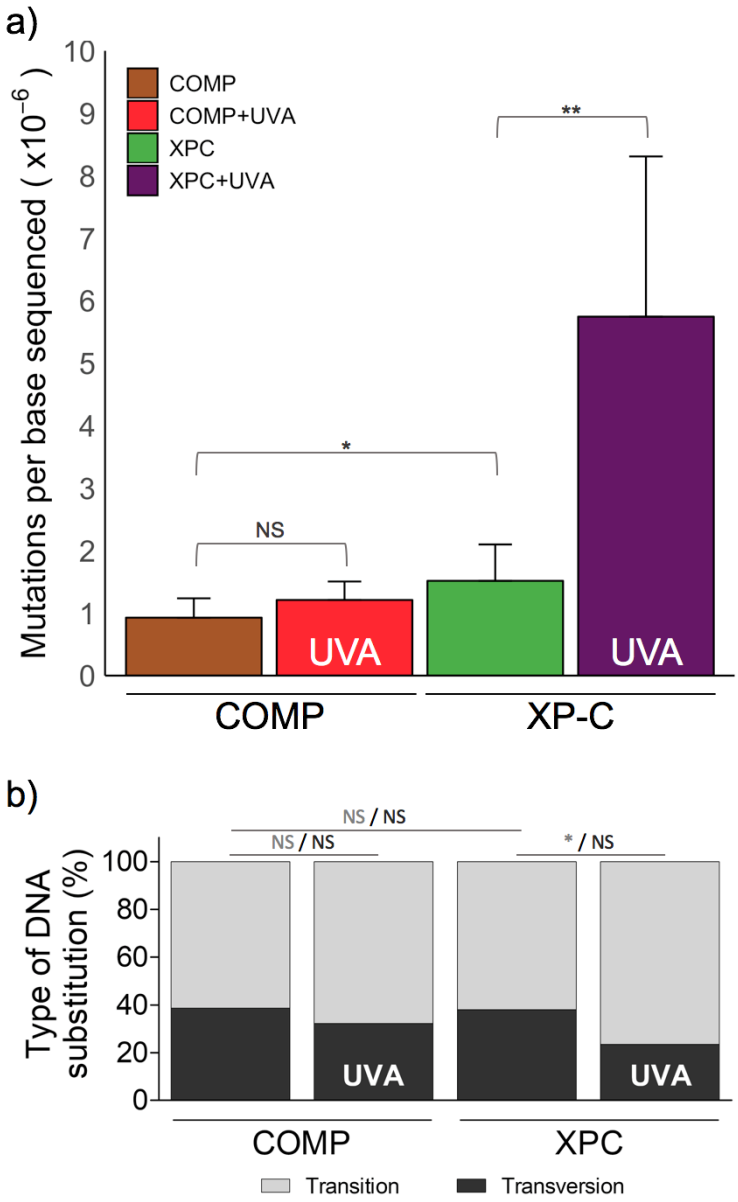


Figure 3.2. UVA light induces frequency of point mutations increase in NER deficient (XP-C) skin fibroblasts. (a) Unique point mutations within exons and splicing sites per sequenced base (n = 6 or 8 clones per group), after UVA irradiation. The data represent mean and standard deviation (SD); **(b)** contribution of transitions (in gray) and transversions (in black) to total mutagenesis induced by UVA light. Irradiation dose: 60 kJ/m^2. Statistical differences were calculated by non-parametric permutation tests: $P < 0.001$ (***), $P < 0.01$ (**), $P < 0.05$ (*), $P > 0.05$ (NS).

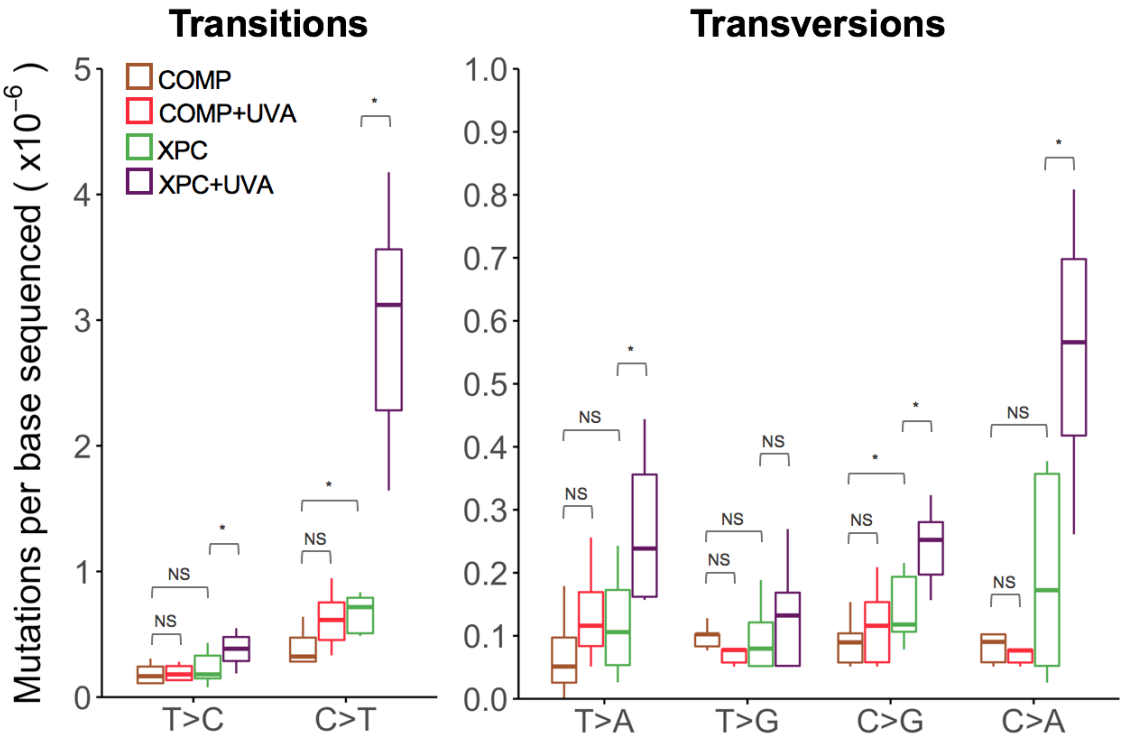


Figure 3.3. Different types of base substitution were induced by UVA light in XP-C cells. The data represent the mean and standard deviation (SD) of exclusive substitutions found within exonic and splicing sites in the sequenced clones for each condition. Cells were irradiated with 0 or 60 kJ/m^2 of UVA light. Statistically significant differences were calculated by non-parametric permutation tests: $P < 0.05$ (*), $P > 0.05$ (NS).

The somatic spectrum analysis (that includes the information on the immediately 5´and 3´ sequence context) indicates that non-irradiated COMP subclones have a more or less homogeneous spectrum with a slight concentration of mutations on the C>T base substitution, but without preference for any specific sequence context (Figure 3.4a). When irradiated with UVA light, COMP cells maintained a similar spectrum, but with a small increase in the frequency of C>T mutations on the YCC and TCA context sequences (Figure 3.4b), which are potential sites for pyrimidine dimer formation. On the other hand, non-irradiated XP-C subclones have a similar spectrum as observed for COMP subclones without any significant change comparing both cell lines (Figure 3.4c). Finally, XP-C cells irradiated with UVA light revealed an evident concentration of its mutations at the C>T transition mainly in the CCN and TCN trinucleotide contexts (Figure 3.4d), which are potential sites for pyrimidine dimer formation. For the other types of base substitutions that presented a significant increase after UVA irradiation, no preference for a particular DNA sequence context was detected: the spectra are flat among the possible combinations of tri-nucleotide. This analysis was also performed for each of the independent subclones, although they presented some small variations, in general they kept the pattern

described for the analysis of the experimental groups (supplementary figure S3.1). This is an indication that indeed, one dose of UVA light is enough to generate the characteristic mutation spectrum rich in C>T mutations on dipyrimidine sequence context.

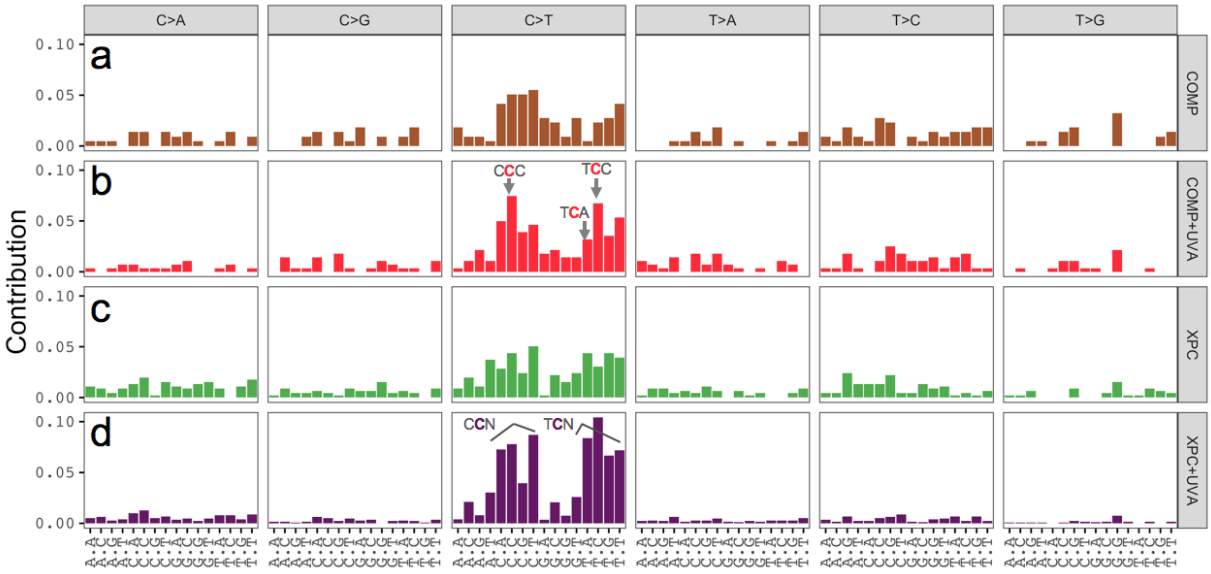


Figure 3.4. Somatic mutation spectra of irradiated or non-irradiated cells. Subclones was merged according to treatments **(a)** non irradiated COMP cell line subclones; **(b)** UVA irradiated COMP cell line subclones; **(c)** non irradiated XP-C cell line subclones; **(d)** UVA irradiated XP-C cell line subclones; this was used to characterize the somatic spectra of trinucleotide motifs using the SomaticSignatures package, an implementation in R. Trinucleotide contribution was calculated by its frequency. The most frequent mutated trinucleotides were highlighted by arrows. Irradiation dose: 60 kJ/m^2 of UVA light.

3.4.4 UVA light generates essentially C>T point mutations, mainly on the second base of potential pyrimidine dimers sites and in the non transcribed strand

It has been shown before that UVA light, as well as UVB and UVC, mainly induces mutations of the C>T type. Therefore, this type of base substitution was further investigated. The CC>TT tandem mutation, considered a UV light hallmark, was the first sequence to be contemplated. UVA irradiated XP-C subclones presented a 16.2 fold increased occurrence of CC>TT tandem mutations, while for the irradiated subclones of the control cell line no increase was detected (Figure 3.5a). Similarly, no change in CC>TT mutations were detected in non-irradiated cells (Figure 3.5a). Well-known motifs of UV induced lesions were also evaluated in order to establish the possible contribution of pyrimidine dimers to the UVA induced mutations. The YY motif was used for detecting any combination of adjacent pyrimidine bases where a mutation

occurs. At random, this motif should be at 75% of the C>T mutations. Experimentally, 82% and 88% of the C>T transitions occur inside YY sites on non-irradiated COMP and XP-C subclones, respectively. On irradiated XP-C subclones, this percentage increases significantly to 95% (Figure 3.5b). On the other hand, the percentage of C>T transitions inside the YCG motif that explores potential pyrimidine dimers in sites of CpG context (eventually targeted for methylated cytosines) did not increase after UVA irradiation. Also, no increase inside the specific TCG and CCG motifs was detected (Figure 3.5b), which indicates that under these experimental conditions, UVA light does not generate, preferentially, C>T transitions in CpG sites. Also, for XP-C irradiated cells the C>T transitions were mainly found in the second pyrimidine of the potential dimer, with a clear preference for the TC dimer sequence (Figure 3.5c).

Furthermore, the sequence contexts where the C>T mutations occur were analysed by the pLogo web software (Figure 3.6). In this representation, the height of the base names is scaled relative to their statistical significance (O´Shea et al. 2013). In non-irradiated cells, it was observed a small but statistically significant enrichment bias for a pyrimidine base at the position -1, being C for COMP and T for XP-C cells. Nonetheless, after UVA irradiation, there were clear significant increases in the preference for T, followed by C, nucleotides at the position -1 for both cell lines, although this was more evident in XP-C cells. Interestingly, this logo also evidenced that for the C>T mutations in UVA irradiated XP-C cells there is a bias to a pyrimidine-rich sequence context, CYYCYNC.

It was also evaluated if C>T mutations at potential dipyrimidine sites (YY) had any strand bias by comparing the mutations found on the transcribed strand (T) related to the untranscribed (UT) strand (Figure 3.7). For non irradiated XP-C and COMP cell lines and for UVA irradiated COMP cells, no transcriptional strand bias for mutations on the YY motif was found. However, for UVA irradiated XP-C cells, it is clear that mutations occur more frequently on the untranscribed strand, which is not subjected to damage removal by TCR activity. When the analysis is performed by each individual clone, this pattern of more mutations on the untranscribed strand is maintained both for C>T transitions, and for all types of base substitutions at the YY motif (Supplementary Figure S3.2).

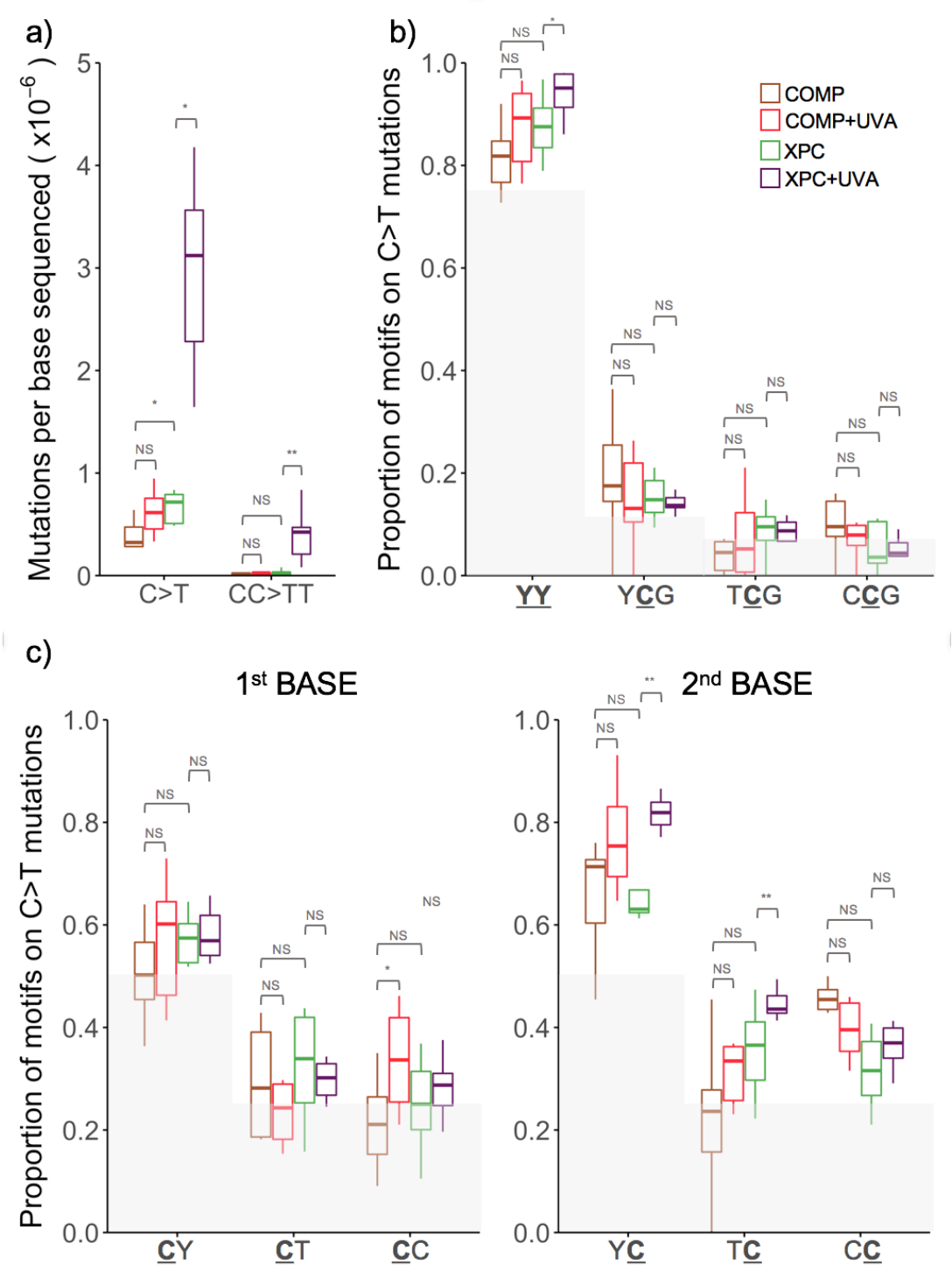


Figure 3.5. Nucleotide changes and UV-associated motif patterns in C>T point mutations. a) Analyses of C>T point mutations and CC>TT tandem mutations after UVA exposure; **b)** analyses of C>T mutations occurring at dipyrimidine sites (YY) and at dipyrimidine sites with a potentially methylated cytosine (YCG, CCG and TCG); **c)** analyses of C>T mutations at the first or the second pyrimidine of a dipyrimidine site. Irradiation dose: UVA 60 kJ/m^2. The IUPAC code was used to represent the nucleotides; Y = C or T. Statistically significant differences were calculated by non-parametric permutation tests: P < 0.01 (**), P < 0.05 (*), P > 0.05 (NS). Grey area on figure 5b and 5c correspond to the expected probability of C>T mutations inside the evaluated motif, 75% for YY, 12.5% for YCG, 6.25% for TCG and CCG, 50% for CY and YC, and 25% for CT, CC, TC and CC.

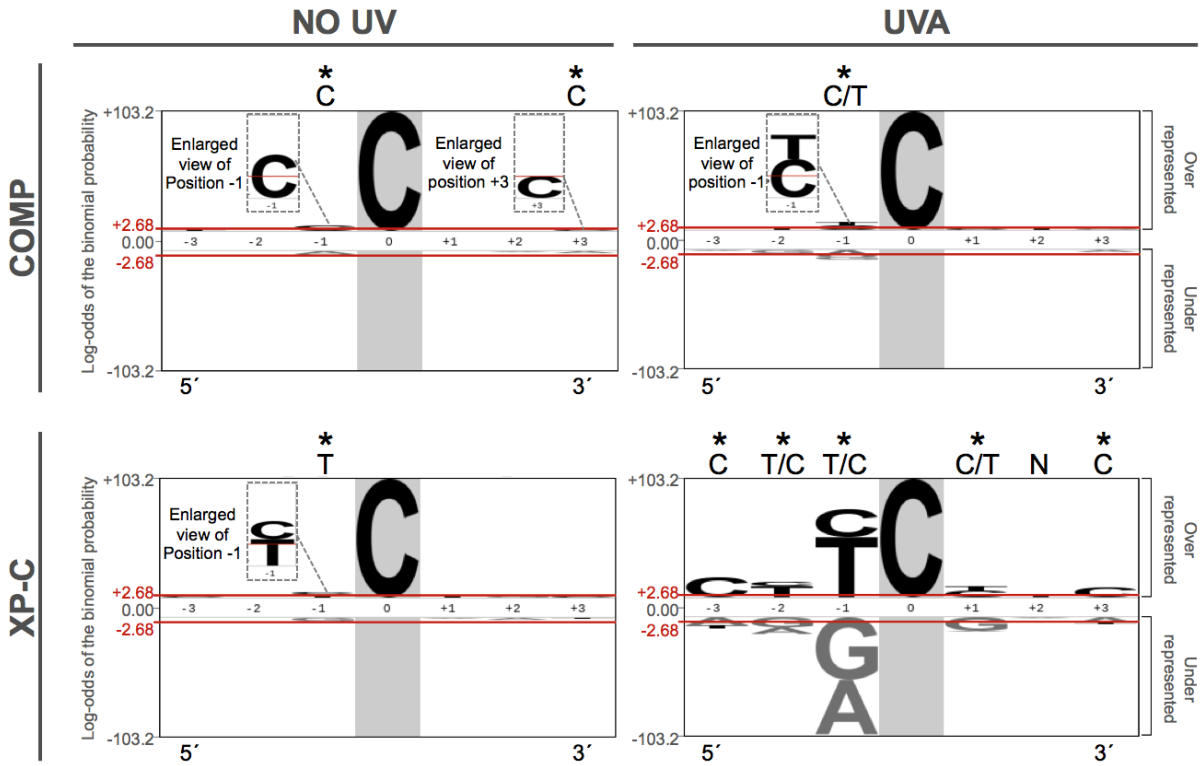


Figure 3.6. Sequence contexts for the C>T mutations. Sequence context logo of all C>T mutations grouped by cell line and treatment, adjusted by strand to mutations at C. The graphic representation was generated by the probability logo -pLogo- web software (https://plogo.uconn.edu/), which illustrates the log-odds of the binomial probability for each nucleotide at each position with respect to the human exome used as background. The pLogo scales the nucleotide residue heights proportional to their statistical significance, and not to their frequency, over- and underrepresented nucleotides are drawn above and below the x-axis, respectively. The red horizontal bar denotes the threshold of the Bonferroni-corrected statistical significance values (p=0.05), nucleotides highlighted (*) at the top of the logo indicates a significant enrichment in that specific position, the closer the nucleotides are drawn to the x-axis, the higher is its statistical significance and the "fixed" position (that allows the use of conditional probabilities) is highlighted by a vertical grey bar. Purine nucleotides are presented in gray and pyrimidine nucleotides in black. Number of aligned foreground (n(fg)) sequences used to generate the image: COMP (n(fg)=92, COMP+UVA (n(fg)=144, XP-C (n(fg)=203, XP-C+UVA (n(fg)=1260. For the background (n(bg)) 4096 sequences were used.

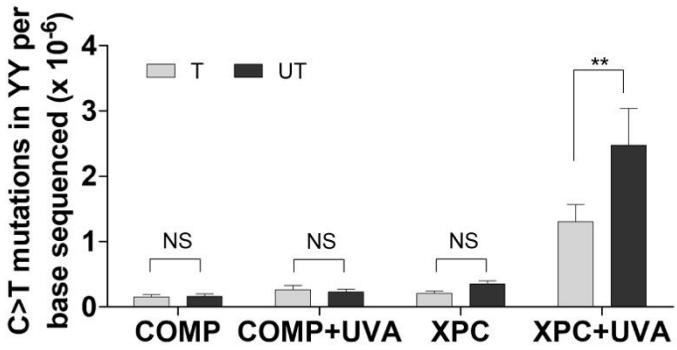


Figure 3.7. Transcriptional strand bias for the C>T mutations. C>T transitions at the YY motif were evaluated if presence at the transcribed (T) and untranscribed (UT) strands.

Statistically significant differences were calculated 2way ANOVA: P < 0.01 (**), P > 0.05 (NS) and normality tested by Kolmogorov-Smirnov.

3.4.5 C>A mutations induced by UVA do not exhibit a preferential sequence context.

The C>A transversion was the second most frequent induced mutation by UVA irradiation in XP-C cells (Figure 3.8a), thus it was further explored. This base substitution has been traditionally related to the 8oxoG lesion, a product of guanine oxidation. It has been shown that the efficient removal of 8oxoG its directly related to the sequence context it is in, since the *E. coli* protein MutM removes this lesion more competently in a pyrimidine rich context, and it is expected that this is also fulfilled for the eukaryotic protein OGG1 (Hatahet et al. 1998, Sassa et al. 2012), therefore, the C>A mutations within the RGR motif was evaluated. However, for all experimental conditions a similar proportion of C>A mutations were found within the RGR motif (Figure 3.8b), suggesting that UVA does not significantly increase the frequency of C>A mutations at this motif. The YY and YCG motif were also evaluated to establish a possible contribution of pyrimidine dimers, but the percentage of C>A transversions within these motifs also did not increase after UVA irradiation (Figure 3.8c). Moreover, the graphical representation of the sequence context adjusted by strand to mutations in G (Figure 3.8d) evidenced that non irradiated COMP and XP-C subclones, as well as the UVA irradiated COMP subclones did not present a significant enrichment bias in the neighbouring positions of the mutated G. However, for UVA irradiated XP-C cells a significant preference for G base at the position +1 was evidenced, but without bias in the position -1. Interestingly, the reverse complement of this sequence result in the 5´CC3´ sequence, which could suggest a relationship between the C>A mutation and the 3´ base of the CC dimer. Finally, strand bias for C>A mutations was also evaluated on the RGR or YY motifs (Figure 3.8e), but no strand bias was detected.

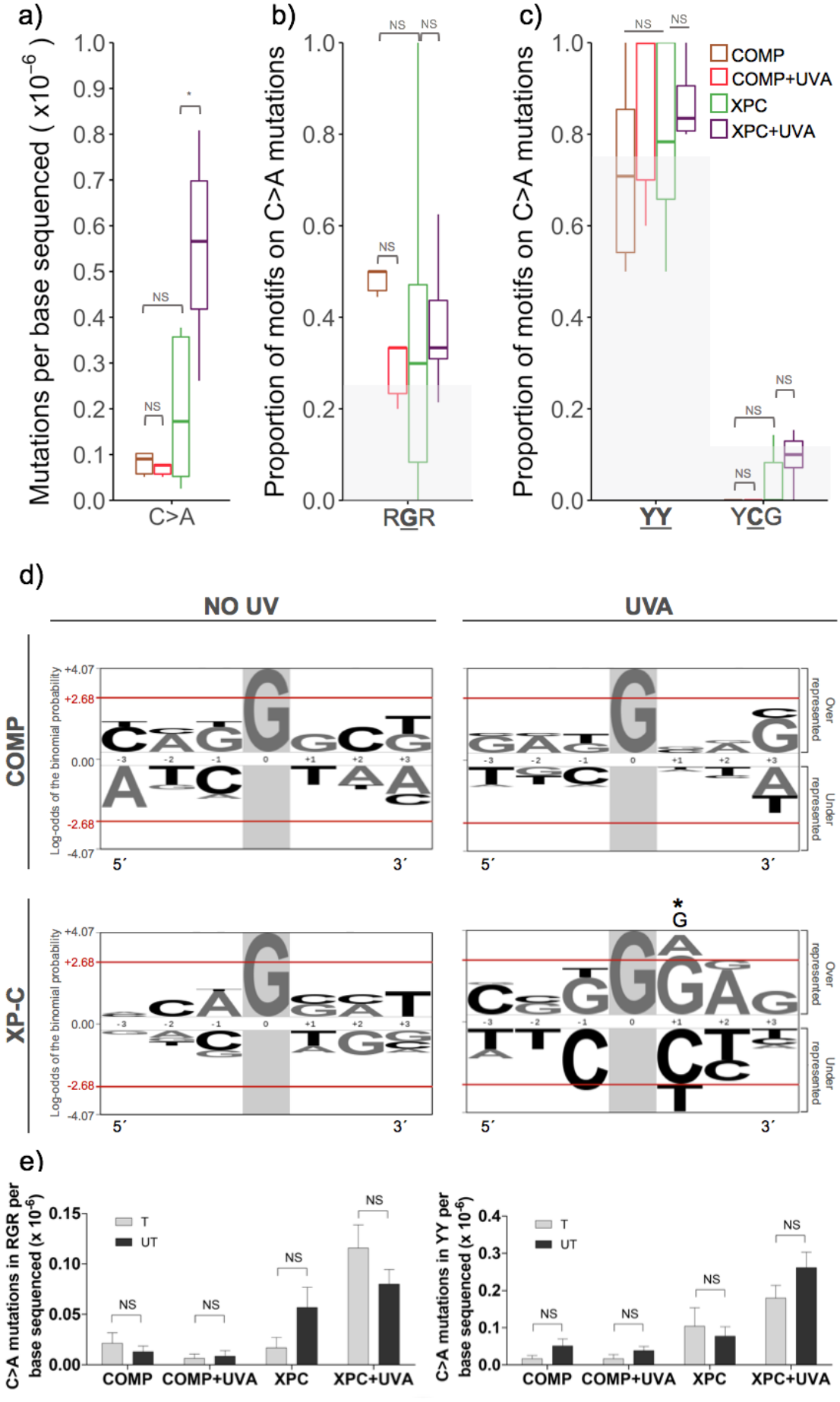

a)
Mutations per base sequenced (x10⁻⁶)
C>A
NS
*
b)
Proportion of motifs on C>A mutations
RGR
c)
Proportion of motifs on C>A mutations
YY
YCG
COMP
COMP+UVA
XPC
XPC+UVA
d)
NO UV
UVA
COMP
XP-C
Log-odds of the binomial probability
+4.07
+2.68
0.00
-2.68
-4.07
5´
3´
Over represented
Under represented
e)
C>A mutations in RGR per base sequenced (x 10⁻⁶)
C>A mutations in YY per base sequenced (x 10⁻⁶)
T
UT
COMP
COMP+UVA
XPC
XPC+UVA

Figure 3.8. Lack of clear sequence contexts for the C>A mutations. (a) Exploratory analyses of C>A mutations after UVA exposure; **(b)** analyses of C>A mutations occurring at oxidized guanines (RGR sequence) and **(c)** at potential dipyrimidine sites (YY, YCG). Irradiation dose: UVA 60 kJ/m^2. The IUPAC code was used to represent the nucleotides; R = A or G and Y = C or T. Statistical differences were calculated by non-parametric permutation tests: P < 0.05 (*), P > 0.05 (NS). Grey area on figures 8b and 8c correspond to the expected probability of C>A mutations inside the evaluated motif, 25% for RGR, 75% for YY and 12.5% for YCG. **(d)** Graphical representation of the context sequence where the C>A mutations occurs adjusted by strand to mutations at G on all clones grouped by the cell line. Purine nucleotides were colored in gray and pyrimidine nucleotides in black. Number of aligned foreground (n(fg)) and background (n(fg)) sequences used to generate the image: COMP (n(fg)=24, n(bg)=4096), COMP+UVA (n(fg)=18, n(bg)=4096), XP-C (n(fg)=77, n(bg)=4096), XP-C+UVA (n(fg)=168, n(bg)=4096). **(e)** Transcriptional strand bias: Frequency of C>A tranversions occurring inside the RGR and YY motif by transcribed (T) and untranscribed (UT) strands. Statistically significant differences were calculated 2way ANOVA : P < 0.05 (*), P > 0.05 (NS) and normality tested by Kolmogorov-Smirnov.

3.4.6 XP-C cells irradiated with UVA light recover the characteristic pattern of skin cancer signatures.

The similarity of mutations generated in the trinucleotide context between experimental groups (and all sequenced samples independently) was evaluated based on Euclidean distances (Supplementary Figures S3a and S3b). The analyses by experimental groups reveal the shortest distance between the UVA irradiated XP-C and COMP cell lines followed by the non-irradiated XP-C and COMP, respectively. Suggesting that UVA generates mutations in similar contexts in proficient and deficient NER cells. Clearly, observing the data for the independent clones (Figure S3b), the UVA-irradiated XP-C cells have the most homogeneous pattern of mutations.

The mutation signatures that explain data were extracted by the NMF method, using all the base substitutions detected (Figure 3.9a). Three signatures describe approximately ~96% of the detected mutations: signature S1 is mainly characterized by a concentration of C>T mutations in a context of dipyrimidine sequences CCN and TCN with the other possible combinations of substitutions in the trinucleotide context more-or-less equally and poorly represented. On the other hand, signatures S2 and S3, which are more similar to each other than with S1, are characterized by a concentration of mutations at C>T transition with a more or less homogeneous distributions among the 16 possible combinations of trinucleotides, but with some peculiarities. The S2 signature has an enrichment of C>T mutations on CCN and GCA tri-nucleotide context, as well as in the CTY context for T>A mutations, CTC, CTG and TTN context for T>C mutations and in the CTC, CTG and GTG context for T>G

mutations. While the S3 signature has a concentration of C>T mutations at TCA and ACT sequence context.

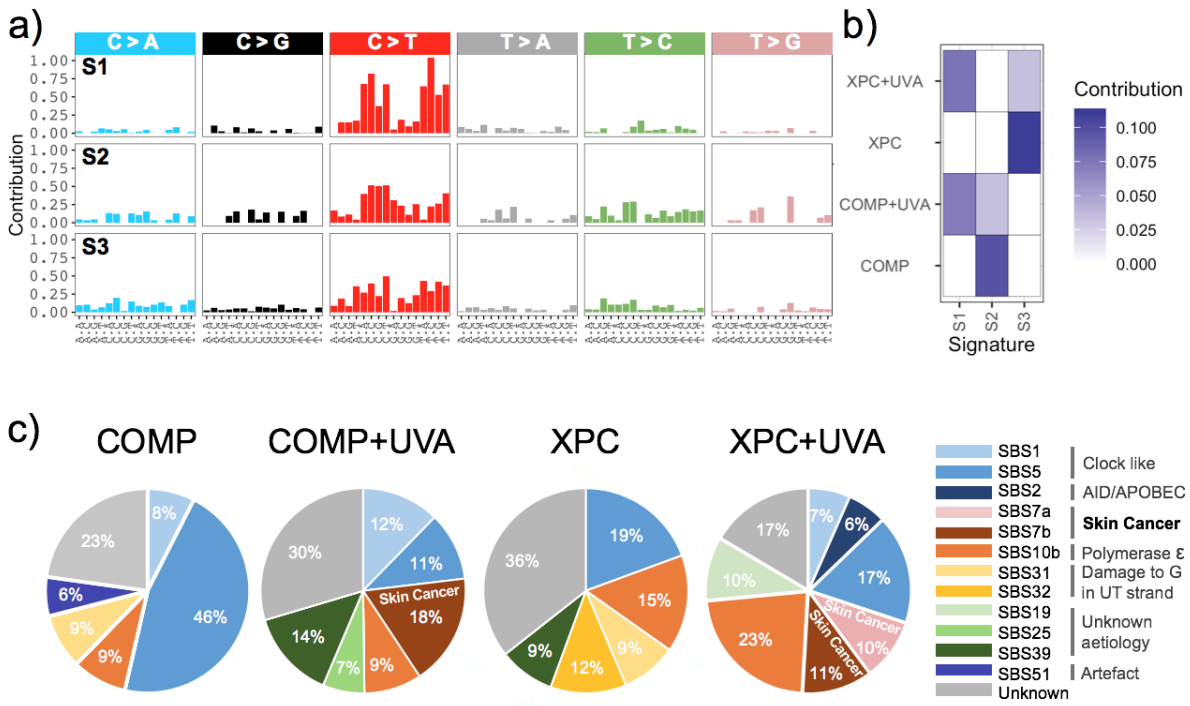


Figure 3.9. Mutational somatic signatures: **(a)** Mutational signatures extracted using NMF-method implemented in SomaticSignatures package based on a method developed by Alexandrov, 2013b. **(b)** Contribution of each signature to explain the mutational pattern of irradiated or non-irradiated cell lines. **(c)** Contribution of the 30 published signatures from Sanger COSMIC project to these data calculated using deconstructSigs package.

Non-irradiated COMP subclones are represented mainly by the S2 signature and when irradiated with UVA a mixture of S2 and S1 signatures explains the mutation spectrum observed, with prevalence for the S1 signature (Figure 3.9b). These results are in agreement with the previous analysis that even without an increment in the frequency of mutations, UVA irradiation leads to a change in the mutation spectrum of COMP cell line with a slight increase of C>T transitions. On the other hand, non-irradiated XP-C subclones are represented by the S3 signature and when irradiated with UVA light mainly the S1 signature and, in small proportion, the S3 signature explain the mutation spectrum of this experimental group. Thus, the S1 signature can be related to the effect of UVA radiation on both cell lines, while the signature S2 is related to the non-irradiated COMP cells and the S3 to the non-irradiated XP-C cells. Subclones analysed independently maintain the general pattern of the experimental

group to which they belong, mainly in the irradiated conditions, but with some variations in the combination of signatures that explain the mutation spectra (Supplementary Figure 3c). Therefore, data confirm that UVA light induce a robust mutational signature independently of the NER repair capacity of cells.

Finally, the exclusive variants of all subclones were compared with the single base substitution (SBS) signatures of the catalogue of somatic mutations in cancer, COSMIC version3 (https://cancer.sanger.ac.uk/cosmic/signatures/SBS/) (Figure 3.9c). A total of 12 COSMIC pre-established signatures contribute to explain the data, 3 of which have an unknown aetiology (SBS19, SBS25 and SBS39), one (SBS51) is considered a possible sequencing artefact and additionally between 17 to 36% of the data in each experimental condition did not match with any of the pre-established signatures. The SBS1 and SBS5 are clock-like signatures found in many types of tumors, these signatures were detected together in almost all of the experimental groups. The SBS1 is characterized by C>T mutations in an NCG context and its aetiology suggests an endogenous mutational process of deamination of 5-methylcytocine (5mC), a very common process in cells. While the SBS5 is characterized by a more or less homogeneous distribution of mutations in the 96 possible combinations with a small enrichment at the ATN context of T>C mutations, its aetiology is unknown. Interestingly, the SBS7b signature characterized by C>T mutations at CCN and to a lesser extent at TCN, ACC and GCC context and related to skin cancers of sun exposed areas was detected only in the UVA-irradiated cells. On the other hand, the SBS7a signature also characterized by C>T mutations, but mainly in TCN context and also related to skin cancer, was only detected in the UVA-irradiated XP-C group. Both, the SBS7a and SBS7b signatures of the COSMIC database, have a transcriptional strand bias of mutations on the C base of the un-transcribed strand. Also, these signatures are very similar to the S1 signature extracted from the data, related to XP-C and COMP UVA irradiated groups. Finally, the SBS10b signature was found in all of the experimental groups, but with an increment of 1.6 times on XP-C subclones and 2.5 times on UVA irradiated XP-C subclones. The SBS10b is characterized by C>T mutations mainly on TCG followed by TCT context, its aetiology relates to polymerase epsilon exonuclease domain mutations and samples with this signature had been considered hypermutators.

These results together suggest that UVA has an important participation in the generation of the typical mutation spectrum of skin cancer, which has been traditionally

attributed to UVB radiation. Thus, the mutagenicity of UVA is related to pyrimidine dimer formation, especially in the absence of NER, in XP-C deficient cells.

3.5 Discussion

The well-established relationship between solar UV, DNA damage, mutations and skin cancer has been possible thanks to studies of XP patients, who present high levels of skin cancer in areas exposed to sunlight due to deficiencies in the NER pathway. The XPC protein participates in the recognition of DNA damage in non-transcribed regions of the genome while lesions in actively transcribed genes are recognized by another route, thus in XP-C deficient cells the lesion removal decreases to 20-30% (Cleaver 1986, Cleaver et al. 2009, Menck & Munford 2014, Bowden et al. 2015). For XP-C cell lines, it has been widely shown their high sensitivity to UVB and UVC light (Feraudy et al. 2010, Dupuy et al. 2013, Quinet et al. 2014, Andrade-Lima 2015, and this work, Chapter 1). However, this work reveals, for the first time, a high sensitivity of XP-C cells also to UVA irradiation, the main component of the UV spectrum that reaches the Earth´s surface.

The elevated rates of skin cancer in XPC patients and the differences obtained in the survival rates and genotoxic stress of XP-C cells compared to complemented cells after UVA irradiation, support the need to perform an analysis of the mutation pattern induced by this wavelength. A unique dose of UVA (60 kJ/m2) increased the frequency of mutations in the sequenced exome of XP-C cells from 1.4 to 5.2 per Mbp, while in control cells the increase in the detected UVA mutagenesis was not significant. These results are in agreement with previous reports under similar experimental conditions for XP-V cells irradiated with UVA light (120 kJ/m^2) and for XP-C cells irradiated with UVB light (120 J/m^2), treatments that increased the mutation frequency to ~7.9 per Mpb for the former and 8.15 per Mpb for the latter, respectively (this work chapter 2, Moreno et al. 2020). These mutation frequencies are also consistent with the mutation prevalence estimated in the sequenced genome of a human malignant melanoma of 8.33 per Mbp in exons and 9.93 per Mbp in introns (Pleasance et al. 2010a). On the other hand, on mouse models predisposed to melanoma the frequency of SNVs slightly increased, but not statistically significant after UVA-irradiation, while it was significant after UVB exposure (Trucco et al. 2019, Hennessey et al. 2019).

The C>T transition in a pyrimidine dimer context is by far the predominant point mutation induced by UVA light on XP-C cells. Interestingly, also the CC>TT tandem

mutations were significantly increased. Similarly, as seen in the WES of XP-V cells (Moreno et al. 2020), and in the *hprt* gene of NER proficient primary human fibroblast (Kappes et al. 2006), in both cases with predilection for the nontranscribed strand as also evidenced in this work.

Due to the high efficiency with which 5mC form dimers in natural sunlight irradiated cells (Tommasi et al. 1997), and the faster spontaneous deamination of the 5mC inside pyrimidine dimers when compared to non methylated cytosine (Cannistraro & Taylor 2009, Tomkova & Schuster-Böckler 2018), it has been proposed that CPDs containing 5mC plays an important role in sunlight induced mutagenesis in mammalian cells (Tommasi et al. 1997, Pfeifer et al. 2005). This is why mutations occurring inside CpG-associated dipyrimidine sites are commonly explored.

A study using as a mutagenic reporter gene the *lacZ* transgene in the skin of mice irradiated with UV light, proposed the induction of C>T transition in the 5´TCG sequence as the “UVA signature” (Ikehata 2018). However, in mouse models predisposed to melanoma most UVA-induced C>T mutations did not occur at CpG sites and, in fact, the UVB-induced C>T mutations were enriched in the TCG and CCA sites, suggesting that cytosine deamination plays a key role in UVB-mediated mutagenesis (Hennessey et al. 2019). In this work, we found the C>T mutation on UVA-irradiated XP-C cells occurred preferentially at 5´TC sequence, but not at 5´TCG. Also, the somatic spectrum analysis evidenced a concentration of this base substitution on CCN and TCN context, but without a particular preference for any of the eight possible trinucleotide combinations. This result highlights that at least in XP-C human cells UVA-irradiation does not induce pyrimidine dimers preferentially in CpG regions, as well as previously reported for UVB irradiation (this work chapter 2), not confirming the 5´TCG sequence as a general signature for "UVA mutagenesis". This result reinforces the fact that it is not possible to directly extrapolate information about dipyrimidine formation containing 5mC bases on one gene to whole exomes or genomes (Douki and Cadet 2001). Also, DNA repair studies of highly methylated regions indicate that methylation may have a protective role from mutations in CpG sites (Tomkova & Schuster-Böckler 2018).

Interestingly, signatures from the COSMIC SBS7 signature group (SBS7a, SBS7b, SBS7c and SBS7d) were only detected on UVA-irradiated XP-C and COMP. This signature group is typically found in skin cancer and related to areas exposed to solar-UV (Alexandrov et al. 2020). Therefore, the data recapitulate experimentally the

mutation patterns seen in tumor analysis, validating the SBS7 mutational signature established for solar-UV. Alexandrov and coworkers (2020) proposed that solar-UV radiation could initiate multiple distinct mutational processes that generate the four SBS7 sub-signatures and suggested that the SBS7a (C>T at T<u>C</u>N) and SBS7b (C>T mainly at C<u>C</u>N and to a lesser extent at T<u>C</u>N) may be the result of mutations generated by one of the two main photoproducts induced by UV light, CPDs or 6,4-PPs. While SBS7c (T>A at N<u>T</u>T) and SBS7d (T>C at N<u>T</u>T) may be due to wrong incorporation of T and G by the TLS polymerases in front T containing dimers. In fact, in WES experiments, both UVA or UVB light, the SBS7a signature was detected only on NER or TLS deficient clones, while SBS7b also appears in irradiated DNA repair proficient cells (this work chapter 1, Moreno et al. 2020). Taking into account these results, and that 6,4-PPs are more quickly repaired than CPDs, it is feasible to think that SBS7b could be related to CPD lesions while SBS7a would be related to both CPDs and 6,4-PPs. On the other hand, the SBS7c and 7d signatures did not appear on UVA or UVB irradiated XP-C cells (this work chapter 2). The lack of SBS7c reinforces the importance of pol eta to bypass T containing dimers since this signature, characterized mainly by T>A mutation at TTT sites, is present on UVA-irradiated XP-V cells (Moreno et al. 2020), but not on UVA or UVB-irradiated XP-C cells (this work chapter 2). Thus, the absence of SBS7c and SBS7d confirms the low mutagenic potential of T when in dimers, confirming that although TT dimers are the most commonly lesion induced by UVA-irradiation (TT(CPD) >> TC (CPD) > CT (CPD)) (Perdiz et al. 2000, Douki et al. 2003, Cadet et al. 2005), it is the least mutagenic. Thus, the data reaffirm, that at least in NER deficient human cells, UVA wavelengths have a high mutagenic potential, similar to UVB and UVC irradiation.

On the other hand, it has been demonstrated that UVA generates oxidative stress that damages DNA (Sage et al. 2012, Schuch et al. 2017). The main repair pathway that deals with oxidatively generated damage is the base excision repair (BER). However, several studies suggest that the XPC protein in addition to the NER pathway also participates in the repair of oxidatively generated DNA damage, probably by interacting with BER (Shimizu et al. 2003, Nahari et al. 2004, D'Errico et al. 2006, Hazra et al. 2007, Kassam et al. 2007, Melis et al. 2008, 2013). The main mutation generated as a consequence of non-repaired oxidized bases are C>A transversions (Epe 1991, Cheng et al. 1992). However, other types of base substitutions can also be induced by oxidated bases, for example: products of cytosine oxidation as 5-

hydroxypyrimidine may induce C>T and C>G mutations, while uracil glycol may induce C>T mutations. Products of adenine oxidation as 2-hydroxy-adenine may induce A>G (T>C) mutations and products of guanine oxidation as fapy·dG may induce G>T (C>A) mutations (Purmal et al. 1994, 1998, Kamiya & Kasai 1997, Kalam et al. 2006).

In this work, UVA-irradiated XP-C cells had significant induction of T>C, T>A, C>G and, mainly, C>A mutations. All of them could be associated to oxidated base damage, and in this case, as they occur basically in XP-C cells, the results confirm that XPC protein may in fact participate in the removal of the DNA damage responsible for these mutations. Interestingly, comparison with similar experiments, performed with UVB-irradiated XP-C cells (this work chapter 2), indicates that C>G mutations are specific for UVA irradiation, while T>G are specific for UVB light.

Moreover, C>A transversions, considered standard types of mutation due to oxidated bases (mainly 8oxoG), were more frequent after UVA irradiation, only in XP-C cells. However, these mutations were not found at potential RGR motifs, where removal of 8oxoG would be less efficient (Hatahet et al. 1998, Sassa et al. 2012). Additionally, none of the recovered COSMIC mutation signatures in the data were related to oxidatively induced DNA damage. The contribution of oxidatively DNA damage on UVA induced mutagenesis has been in debate (Kappes and Runger 2005, Besaratinia et al. 2008). Previous results with XP-V cells demonstrated that, besides C>T transition, C>A transversions were frequently induced by UVA light (Moreno et al., 2020). Interestingly, the simple lack of the DNA polymerase eta (pol eta), under no UVA irradiation conditions, led to an increase of C>A mutations and the recovery of the SBS18 and SBS36 COSMIC signatures, which are typically related to damage possibly induced by reactive oxygen species. These results suggest that pol eta correctly bypass the oxidated bases (Moreno et al. 2019, 2020). Also, the C>A transversions were considered late mutations in the evolution of skin cancer, and it has been suggested that the underlying process of its formation is not necessarily related to UV light exposure (Pleasance et al. 2010a). Although, the results here did not reveal direct indications that these C>A transversions are due to oxidated damage, most likely in the lack of XPC protein, lesions like 8oxoG are not effectively repaired and thus may still be responsible for the mutations observed in this work.

It has been proposed that the high mutagenic capacity of UVA light is, in part, due to the slow removal rate of CPDs, which maybe slower that in cells irradiated with UVB light (Courdavault et al. 2005, Besaratinia et al. 2008, Mouret et al. 2006, Runger

et al. 2012). Most probably, this slower repair maybe due to protein oxidation consequently a general reduction in NER capacity (Guven et al. et al. 2015, Karran & Brem 2016, Brem et al. 2017). This is supported by the protective effect of some antioxidants on UVA-irradiated cells, as is the case of vitamin E, which reduces the level of oxidative generated DNA damage as well as pyrimidine dimers in keratinocytes (Delinasios et al. 2018). Similarly, N-acetylcystein (NAC) protects against protein carboxylation, replication fork stalling and cell cycle arrest in UVA irradiated XP-V cells, but not after UVC irradiation (Moreno et al. 2019). Interestingly, NAC also improves the survival rate of XP-C cells when irradiated with UVA but not when irradiated with UVB (data not shown). These results suggest that the use of antioxidants could help in the protection of some deleterious effects generated by UVA-irradiation in both TLS and NER deficient cells. However, it will be necessary to determine if the type and frequency of mutations induced by UVA light changes and/or decrease in cells treated with antioxidants. Therefore, the cellular oxidative process induced by UVA irradiation, and in a lesser extend by UVB light, may be responsible for part of the mutagenesis observed here, and play deleterious roles on cell metabolism, promoting human skin tumors and photoaging (Schuch et al. 2017).

Other COSMIC signatures enriched on UVA-irradiated cells included the SBS2 and SBS10b signatures, which are characterized by C>T mutation at T<u>C</u>N sites and related to AID/APOBEC family activity and to polymerase epsilon exonuclease domain mutations, respectively. The SBS2 signature was only recovered on UVA or UVB-irradiated XP-C cells, but not on repair proficient cells. Also, it was recovered on UVA-irradiated XP-V cells and on tumors from mouse models predisposed to melanoma irradiated either with UVA or with UVB (this work chapter 2, Moreno et al. 2020, Hennessey et al. 2019). On the other hand, the SBS10b was recovered on both XP-C and COMP cells, independent of irradiation, although this signature represents a higher percentage of the data in UVA irradiated XP-C cells. Similarly, SBS10b was also recovered on UVB-irradiated XP-C cells and on tumors from mouse models, but not on UVA-irradiated XP-V cells (this work chapter 2, Moreno et al. 2020, Hennessey et al. 2019). These results suggest a relation of SBS2 and SBS10b signatures and UVA induced mutations, and the eventual link of SBS10b in NER deficient cells could be related to the participation of polymerase epsilon in the gap-filling step of the NER pathway (Pospiech & Syväoja 2003, Ogi et al. 2010).

The results obtained in this work indicate, for the first time, the high sensitivity of XP-C cells to UVA irradiation. A unique and environmentally relevant UVA dose of irradiation is also able to induce high levels of mutations in these NER deficient cells, mostly C>T transitions in a dipyrimidine context. This confirms that UVA induced biological consequences are highly to lesions such as pyrimidine dimers. Other types of mutations (including the more common C>A transversions) were also detected, particularly in UVA-irradiated XP-C cells, probably due to lesions induced by oxidative stress. Interestingly, C>G transversions seem to be specific for UVA light, when compared to similar experiments performed with UVB. The increase in these mutations not only confirm a role of oxidation due to UVA irradiation, but also provide further evidence on the participation of XPC protein in the removal of such lesions. Finally, it was possible to recover the typical mutation spectrum of skin cancer on UVA-irradiated XP-C cells. The C>T mutation is widely considered as the mutational UV-signature when found at C-containing pyrimidine dimers, since it is commonly found in skin cancers but not other types of cancers (Ziegler et al. 1993, Wikondahl and Brash 1999, Ikehata & Ono 2011, Alexandrov et al. 2013a, 2020). Therefore, this work contributes to the understanding of the effects of UVA light on human cells, and demonstrating the relevance of these wavelengths on skin tumor development not only on XP patients, but also on the human population in general.

Acknowledgements: The authors thank for the financial support to the Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP, São Paulo, Brazil, Grants #2014/15982-6, #2013/08028-1). Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Grant # 308868/2018-8) and Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior (CAPES, Brasília, DF, Brazil, financial code 001) and Administrative Department of Science, Technology and Innovation of Colombia (COLCIENCIAS, Bogota, Colombia). We are also grateful to the support with NGS, at the core Facility for Scientific Research – USP (CEFAP-USP/GENIAL), and Multi-user genomic Section of the Human Genome & Stem Cell Research Center (HUG-CELL).

3.6 References

AGAR, N.S., HALLIDAY, G.M., BARNETSON, R.S., ANANTHASWAMY, H.N., WHEELER, M., JONES, A.M. The basal layer in human squamous tumors harbors more UVA than UVB fingerprint mutations: a role for UVA in human skin carcinogenesis. **Proceedings of the**

National Academy of Sciences of USA, v. 101, p. 4954–4959, 2004.

ALEXANDROV, L.B. et al. Signatures of mutational processes in human cancer. **Nature**, v. 500, p. 415-421, 2013a.

ALEXANDROV, L.B. et al. Deciphering Signatures of Mutational Processes Operative in Human Cancer. **Cell Reports**, v. 3, p. 246–259, 2013b.

ALEXANDROV, L.B., STRATTON, M.R. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. **Current Opinion in Genetics & Development**, v. 24, p. 52–60, 2014.

ALEXANDROV, L.B. et al. The Repertoire of Mutational Signatures in Human Cancer. **Nature**, v. 578, p. 94-101, 2020.

ANDRADE-LIMA, L.C., ANDRADE, L.N., MENCK, C.F.M. ATR suppresses apoptosis after UVB irradiation by controlling both translesion synthesis and alternative tolerance pathways. **Journal of Cell Science**, v. 128, p. 150–159, 2015.

ARMSTRONG, B.K., KRICKER, A. The epidemiology of UV induced skin cancer. **Journal of Photochemistry and Photobiology B: Biology**, v. 63, p. 8–18, 2001.

BESARATINIA, A., SYNOLD, T.W., XI, B., PFEIFER, G.P. G-to-T transversions and small tandem base deletions are the hallmark of mutations induced by ultra-violet A radiation in mammalian cells. **Biochemistry**, v. 43, p. 8169–8177, 2004.

BESARATINIA, A., KIM, S., PEFEIFER, G.P. Rapid repair of UVA-induced oxidized purines and persistence of UVB-induced dipyrimidine lesions determine the mutagenicity of sunlight in mouse cells. **FASEB Journal**, v. 22, p. 2379-2392, 2008

BERG, R.J.W., DE GRUIJL, F.R. VAN DER LEUN, J.C. Interaction between Ultraviolet A and Ultraviolet B Radiations in Skin Cancer Induction in Hairless Mice. **Cancer research**, v. 53, p. 4212-4217, 1993.

BONIOL, M., AUTIER, P., BOYLE, P., GANDINI, S. Cutaneous melanoma attributable to sunbed use: systematic review and meta-analysis. **British Medical Journal**, v. 345, p. E4757, 2012

BOWDEN, N.A. et al. Understanding xeroderma Pigmentosum complementation groups using gene expression profiling after UV-light exposure. **International Journal of Molecular Sciences**, v. 16, p.15985–15996, 2015.

BREM, R., GUVEN, M., KARRAN, P. Oxidatively-generated damage to DNA and proteins mediated by photosensitized UVA. **Free Radical Biology and Medicine**, v. 107, p. 101-109, 2017.

BURREN, R., SCALETTA, C., FRENK, E., PANIZZON, R.G., APPLEGATE, L.A. Sunlight and carcinogenesis: Expression of p53 and pyrimidine dimers in human skin following UVA I, UVA I + II and solar simulating radiations. **International Journal of cancer**, v. 76, p. 201-206, 1998.

CADET, J., DOUKI, T. Formation of UV-induced DNA damage contributing to skin cancer development. **Photochemical & Photobiological Sciences**, v. 17, p. 1816–1841, 2018.

CADET, J., SAGE, E., DOUKI, T. Ultraviolet radiation-mediated damage to cellular DNA. **Mutation Research**, v. 571, p. 3– 17, 2005.

CANNISTRARO, V.J., TAYLOR, J.S. Acceleration of 5-methylcytosine deamination in cyclobutane dimers by g and its implications for UV-induced c-to-t mutation hotspots. **Journal of Molecular Biology**, v. 392, p. 1145–1157, 2009.

CHENG, K.C., CAHILL, D.S., KASAI H., NISHIMURA, S., LOEB, L.A. 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G→T and A→C substitutions. **The Journal of biological chemistry**, v. 267, p.166–172, 1992.

CLEAVER, J.E. Defective repair replication of DNA in xeroderma pigmentosum. **DNA repair**, v. 218, p. 652–656, 1968.

CLEAVER, J.E. DNA repair in human xeroderma pigmentosum group C cells involves a different distribution of damaged sites in confluent and growing cells. **Nucleic acids research**, v. 14, p. 8155–8165, 1986.

CLEAVER, J.E., LAM, E.T., REVET, I. Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. **Nature reviews. Genetics**, v. 10, p. 756–768, 2009.

CORTAT, B. et al. 2013. The relative roles of DNA damage induced by UVA irradiation in human cells. **Photochemical & Photobiological Sciences**, v. 12, p. 1483-1495, 2013.

COURDAVAULT, S. et al. Repair of the three main types of bipyrimidine DNA photoproducts in human keratinocytes exposed to UVB and UVA radiations. **DNA Repair**, v. 4, p. 836–844, 2005.

COURDAVAULT, S. et al. Larger yield of cyclobutane dimers than 8-oxo-7,8-dihydroguanine in the DNA of UVA-irradiated human skin cells. **Mutation Research**, v. 556, p. 135–142, 2004b.

D'ERRICO M, et al. New functions of XPC in the protection of human skin cells from oxidative damage. **EMBO J**, v. 25, p. 4305–4315, 2006.

DANECEK, P. et al. The variant call format and VCFtools. **Bioinformatics**, v. 27, p.2156–2158, 2011.

DAYA-GROSJEAN, L., JAMES, M.R., DROUGARD, C., SARASIN, A. An immortalized xeroderma pigmentosum, group C, cell line which replicates SV40 shuttle vectors. **Mutation Research DNA Repair Reports**, v. 183, p. 185–196, 1987.

DeFABO, E.C., NOONAN, F.P., FEARS, T., MERLINO G. Ultraviolet B but not Ultraviolet A Radiation Initiates Melanoma. **Cancer research**, v. 64, p. 6372-6376, 20004.

DeGRUIJL F.R. et al. Wavelength dependence of skin cancer induction by ultraviolet irradiation of albino hairless mice. **Cancer research**, v. 53, p. 53–60, 1993.

DeLAAT, A., VAN DER LEUN, J.C., DeGRUIJL, F.R. Carcinogenesis induced by UVA (365-nm) radiation: the dose–time dependence of tumor formation in hairless mice. **Carcinogenesis**, v. 18, p. 1013-1020, 1997.

DELINASIOS, G.J., KARBASCHI, M., COOKE, M.S., YOUNG, A. Vitamin E inhibits the UVAI induction of "light" and "dark" cyclobutane pyrimidine dimers, and oxidatively generated DNA damage, in keratinocytes. **Scientific reports**, v. 8, p. 423-435, 2018.

DePRISTO, M.A. et al. A framework for variation discovery and genotyping using next generation DNA sequencing data. **Nature Genetics**, v. 43, p. 491–498, 2011.

DIFFEY, B.L. Solar ultraviolet radiation effects on biological systems. **Physics in Medicine and Biology**, v. 36, p. 299-328, 1991.

DOUKI, T., CADET, J. Individual determination of the yield of the main UV-induced dimeric pyrimidine photoproducts in DNA suggests a high mutagenicity of CC photolesions. **Biochemistry**, v. 40, p. 2495–2501, 2001.

DOUKI, T., REYNAUD-ANGELIN, A., CADET, J., SAGE, E. Bipyrimidine photoproducts rather than oxidative lesions are the main type of DNA damage involved in the genotoxic effect of solar UVA radiation. **Biochemistry**, v. 42, p. 9221–9226, 2003.

DROBETSKY, E.A., TURCOTTE, J., CHATEAUNEUF, A. A role for ultraviolet A in solar mutagenesis. **Proceedings of the National Academy of Sciences of USA**, v. 92, p. 2350–2354, 1995.

DUPUY, A. et al. Targeted gene therapy of xeroderma pigmentosum cells using meganuclease and TALENTM. **PLoS ONE**, v. 8, p. 1–8, 2013.

EPE, B. Genotoxicity of singlet oxygen. **Chemico-Biological Interactions**, v. 80, p. 239–260, 1991.

FEARS, T.R. et al. Sunbeds and sunlamps: who used them and their risk for melanoma. **Pigment Cell & Melanoma Research**, v. 24, p. 574-581, 2011.

FERAUDY, S. et al. The DNA damage-binding protein XPC is a frequent target for inactivation in squamous cell carcinomas. **The American Journal of Pathology**, v. 177, p. 555–562, 2010.

FLOYD, L., TOBISKA, W.K., CEBULA, R.P. Solar UV irradiance, its variation, and its relevance to the Earth. **Advances in Space Research**, v. 29, p. 1427-1440, 2002.

GARLAND, C.F., GARLAND, F.C., GORHAM, E.D. Rising trends in melanoma an hypothesis concerning sunscreen effectiveness. **Annals of Epidemiology**, v. 3, p. 103-110, 1993.

GASPARRO, F.P. Sunscreens, Skin Photobiology, and Skin Cancer: The Need for UVA Protection and Evaluation of Efficacy. **Environmental Health Perspective,** v. 108, p. 71–78, 2000.

GEHRING, J.S., FISCHER, B., LAWRENCE, M., HUBER, W. SomaticSignatures: Inferring mutational signatures from single-nucleotide variants. **Bioinformatics**, v. 31, p. 3673–3675, 2015.

GOODWIN, S., McPHERSON, J.D., MCCOMBIE, W.R. Coming of age: ten years of next-generation sequencing Technologies. **Nature reviews**, v. 17, p. 333-351, 2016.

GUVEN, M., BREM, R., MACPHERSON, P., PEACOCK, M., KARRAN, P. Oxidative damage to RPA limits the nucleotide excision repair capacity of human cells. **Journal of Investigative Dermatology**, v. 135, p. 2834–2841, 2015.

HATAHET, Z., ZHOU, M., REHA-KRANTZ, L.J., MORRICAL, S.W., WALLACE, S.S. In search of a mutational hotspot. **Proceedings of the National Academy of Science of the USA,** v. 95, p. 8556–8561, 1998.

HAZRA, T.K., DAS, A., DAS, S., CHOUDHURY, S., KOW, Y.W., ROY, R. Oxidative DNA damage repair in mammalian cells: a new perspective. **DNA Repair**, v. 6, p. 470–80, 2007.

HENNESSEY, R.C., et al. UVA and UVB elicit distinct mutational signatures in melanoma. **BioRxiv** 778449, 2019

HUSCHTSCHA, L.I., HOLLIDAY, R. Limited and unlimited growth of SV40-transformed cells from human diploid MRC-5 fibroblasts. **Journal of cell science**, v. 63, p. 77–99, 1983.

IKEHATA, H. et al. UVA1 genotoxicity is mediated not by oxidative damage but by cyclobutane pyrimidine dimers in normal mouse skin. **Journal of Investigative Dermatology**, v. 128, p. 2289–2296, 2008.

IKEHATA, H. KUMAGAI, J., ONO T., MORITA, A. Solar-UV-signature mutation prefers TCG to CCG: extrapolative consideration from uva1-induced mutation spectra in mouse skin. **Photochemical & Photobiological Sciences**, v. 12, p. 1319–1327, 2013.

IKEHATA, H. Mechanistic considerations on the wavelength-dependent variations of UVR genotoxicity and mutagenesis in skin: The discrimination of UVA-signature from UV-signature mutation. **Photochemical & Photobiological Sciences**, v. 17, p. 1861–1871, 2018.

IKEHATA, H., ONO, T. The Mechanisms of UV Mutagenesis. **Journal of Radiation Research**, v. 52, p. 115–125, 2011.

KALAM, M.A. et al. Genetic effects of oxidative DNA damages: comparative mutagenesis of the imidazole ring-opened formamidopyridines (Fapy lesions) and 8-oxo-purines in simian kidney cells. **Nucleic Acids Research,** v. 34, p. 2305–2315, 2006.

KAMIYA, H., KASAI, H. Mutations induced by 2-hydroxyadenine on a shuttle vector during leading and lagging strand syntheses in mammalian cells. **Biochemistry**, v. 36, p. 11125–11130, 1997.

KAPPES, U.P., LUO, D., POTTER, M., SCHULMEISTER, K., RUNGER, T.M. Short- and long-wave UV light (UVB and UVA) induce similar mutations in human skin cells. **The Journal of investigative dermatology**, v. 126, p. 667–675, 2006.

KAPPES, U.P., RUNGER, T.M. No major role for 7,8-dihydro-8-oxoguanine in ultraviolet light-induced mutagenesis. **Radiation Research**, v. 164, p. 440–445, 2005.

KARRAN, P., BREM, R. Protein oxidation, UVA and human DNA repair. **DNA Repair**, v. 44, p. 178–185, 2016.

KASSAM, S.N., RAINBOW, A.J. Deficient base excision repair of oxidative DNA damage induced by methylene blue plus visible light in xeroderma pigmentosum group C fibroblasts. **Biochemical and Biophysical Research Communications**, v. 359, p. 1004–9, 2007.

LEITE, R.A. et al. Identification of XP complementation groups by recombinant adenovirus carrying DNA repair genes. **Journal of Investigative Dermatology**, v. 129, p. 502-506, 2009.

LEY, R.D. Dose response for ultraviolet radiation A–induced focal melanocytic hyperplasia and nonmelanoma skin tumors in *Monodelphis domestica*. **Photochemistry and Photobiology**, v. 73, p. 20-23, 2001.

LI, H., DURBIN, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. **Bioinformatics**, v. 26, p. 589–595, 2010.

LI, L., BALES, E.S., PETERSON, C.A., LEGERSKI, R.J. Characterization of molecular defects in xeroderma pigmentosum group C. **Nature Genetics**, v. 5, p. 413–417, 1993.

MARIONNET, C., PIERRARD, C., GOLEBIEWSKI, C., BERNERD, F. Diversity of Biological Effects Induced by Longwave UVA Rays (UVA1) in Reconstructed Skin. **PLOS ONE**, v. 9, p. e105263, 2014.

MATSUMURA, Y. ANANTHASWAMY, H.N. Molecular mechanisms of photocarcinogenesis. **Frontiers in Bioscience**, v. 7, p. 765-783, 2002.

MELIS, J.P., et al. Mouse Models for xeroderma pigmentosum Group A and Group C Show Divergent Cancer Phenotypes. **Cancer Research**, v. 68, p. 1347-1353, 2008.

MELIS, J.P., et al. Slow accumulation of mutations in Xpc-/- mice upon induction of oxidative stress. **DNA Repair**, v. 12, p. 1081–1086, 2013.

MENCK, C.F., MUNFORD, V. DNA repair diseases: what do they tell us about cancer and aging? **Genetics and Molecular Biology**, v. 37, p. 220–233, 2014.

MILLER, S.A., HAMILTON, S.L., WESTER, U.G., HOWARD CYR, W. An Analysis of UVA Emissions from Sunlamps and the Potential Importance for Melanoma. **Photochemistry and Photobiology**, v. 68, p. 63-70, 1998

MITCHELL, D.L. et al. Ultraviolet A does not induce melanomas in a Xiphophorus hybrid fish model. **PNAS**, v. 107, p. 9329-9334, 2010.

MORENO, N. et al. The key role of UVA-light induced oxidative stress in human xeroderma Pigmentosum Variant cells. **Free Radical Biology and Medicine**, v. 131, p. 432-442, 2019.

MORENO, N. et al. Whole-exome sequencing reveals the impact of UVA light mutagenesis in xeroderma pigmentosum variant human cells. **Nucleic Acid Research**, v.48, p. 1941-1953, 2020.

MOURET, S., BAUDOUIN, C., CHARVERON, M., FAVIER, A., CADET, J., DOUKI, T. Cyclobutane pyrimidine dimers are predominant DNA lesions in whole human skin exposed to UVA radiation. **Proceedings of the National Academy of Sciences of the USA**, v. 103, p.13765–13770, 2006.

NAHARI, D., MCDANIEL, L.D., TASK, L.B., DANIEL, R.L., VELASCO-MIGUEL, S., FRIEDBERG, E.C. Mutations in the Trp53 gene of UV-irradiated Xpc mutant mice suggest a novel Xpc- dependent DNA repair process. **DNA Repair** (Amst), v. 3, p. 379–386, 2004.

NICHOLS, J., KATIYAR, S.K. Skin photoprotection by natural polyphenols: anti- inflammatory, antioxidant and DNA repair mechanisms. **Archives of dermatological research**, v. 302, p. 71– 83, 2010.

NIK-ZAINAL, S. et al. Mutational processes molding the genomes of 21 breast cancers. **Cell**, v. 149, p. 979–993, 2012.

OGI, T. et al. Three DNA Polymerases, Recruited by Different Mechanisms, Carry Out NER Repair Synthesis in Human Cells. **Molecular Cell**, v. 37, p. 714–727, 2010.

O'LEARY, N.A. et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. **Nucleic Acids Research**, v. 44, p. D733–D745, 2016.

O'SHEA, J.P., CHOU, M.F., QUADER, S.A., RYAN, J.K., CHURCH, G.M., SCHWARTZ, D. pLogo: a probabilistic approach to visualizing sequence motifs. **Nature methods**, v. 10, p. 1211-1212, 2013.

PASTILA, P., LESZCZYNSKI, D. Ultraviolet A exposure might increase metastasis of mouse melanoma: a pilot study. **Photodermatology Photoimmunology & Photomedicine**, v. 21, p. 183-190, 2005.

PERDIZ, D., GROF, P., MEZZINA, M., NIKKAIDO, O., MOUSTACCHI, E., SAGE, E. Distribution and repair of bipyrimidine photoproducts in solar UV-irradiated mammalian cells. **The Journal of Biological Chemistry**, v. 275, p. 26732–26742, 2000.

PERSSON, A.E. et al. The mutagenic effect of ultraviolet-A1 on human skin demonstrated by sequencing the p53 gene in single keratinocytes. **Photodermatology Photoimmunology Photomedicine**, v. 18, p. 287–293, 2002.

PFEIFER, G.P., YOU Y.H., BESARATINIA, A. Mutations induced by ultraviolet light. **Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis**, v. 571, p.19–31, 2005.

PLEASANCE, E.D. et al. A comprehensive catalogue of somatic mutations from a human cancer genome. **Nature**, v. 463, p. 191–196, 2010a.

POSPIECH, H., SYVÄOJA, J.E., DNA Polymerase ε - More Than a Polymerase. **The Scientific World Journal**, v. 3, p. 87–104, 2003.

PURMAL, A.A., KOW, Y.W., WALLACE, S.S. Major oxidative products of cytosine, 5-hydroxycytosine and 5-hydroxyuracil, exhibit sequence context-dependent mispairing *in vitro*. **Nucleic Acids Research**, v. 22, p. 72–78, 1994.

PURMAL, A.A., LAMPMAN, G.W., BOND, J.P., HATAHET, Z., WALLACE,S.S. Enzymatic processing of uracil glycol, a major oxidative product of DNA cytosine. **Journal of Biological Chemistry**, v. 273, p. 10026–10035, 1998.

QUINET, A. et al. Gap-filling and bypass at the replication fork are both active mechanisms for tolerance of low-dose ultraviolet-induced DNA damage in the human genome. **DNA Repair**, v. 14, p. 27–38, 2014.

RASTOGI, R.P., RICHA, KUMAR, A., TYAGI, M.B., SINHA, P. Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. **Journal of nucleic acids**, v. 2010, p.1-32, 2010.

RAVANAT, J.L., DOUKI, T., CADET, J. Direct and indirect effects of UV radiation on DNA and its components. **Journal of Photochemistry and Photobiology B: Biology**, v. 63, p. 88–102, 2001.

ROCHETTE, P.J. et al. UVA-induced cyclobutane pyrimidine dimers form predominantly at thymine-thymine dipyrimidines and correlate with the mutation spectrum in rodent cells. **Nucleic Acids Research**, v. 31, p. 2786–2794, 2003.

RIDLEY, A.J., WHITESIDE, J.R., MCMILLAN T.J., ALLINSON, S.L. Cellular and sub-cellular responses to UVA in relation to carcinogenesis. **International Journal of Radiation Biology**, v. 85, p. 177-195, 2009.

ROSENTHAL, R., Mc.GRANAHAN, N., HERRERO, J., TAYLOR, B.S., SWANTON, C. DeconstructSigs: Delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. **Genome Biology**, v. 17, p. 1-11, 2016.

RUNGER, T.M. C-T transition mutations are not solely UVB-signature mutations, because they are also generated by UVA. **Journal of Investigative Dermatology**, V. 128, P. 2138-2140,

2008.

RUNGER, T.M., FARAHVASH, B., HATVANI, Z., REES, A. Comparison of DNA damage responses following equimutagenic doses of UVA and UVB: A less effective cell cycle arrest with UVA may render UVA-induced pyrimidine dimers more mutagenic than UVB-induced ones. **Photochemical & Photobiological Sciences**, v. 11, p. 207-215, 2012.

SAGE, E., GIRARD, P.M., FRANCESCONI, S. Unravelling UVA-induced mutagenesis. **Photochemical & Photobiological Sciences**, v. 11, p. 74–80, 2012.

SAGE, E., LAMOLET, B., BRULAY, E. MOUSTACCHI, E., CHATEAUNEUF, A., DROBETSKY, E.A. Mutagenic specificity of solar UV light in nucleotide excision repair-deficient rodent cells. **Proceedings of the National Academy of Sciences of USA**, v. 93, p. 176–180, 1996.

SANCAR, A. DNA Excision Repair. **Annual Review of Biochemistry**, v. 65, p. 43–81, 1996.

SANTIAGO, K.M. et al. Comprehensive germline mutation analysis and clinical profile in a large cohort of Brazilian xeroderma pigmentosum patients. **Journal of the European Academy of Dermatology and Venereology**, (*submitted*).

SASSA, A., BEARD, W.A., PRASAD, R., WILSON, S.H. DNA sequence context effects on the glycosylase activity of human 8-oxoguanine DNA glycosylase. **Journal of Biological Chemistry**, v. 287, p. 36702–36710, 2012.

SCHUCH, A.P. et al. DNA damage profiles induced by sunlight at different latitudes. **Environmental and Molecular Mutagenesis**, v. 53, p.198–206, 2012.

SCHUCH, A.P., GALHARDO, R.S., LIMA-BESA, K.M., SCHUCH, N.J., MENCK, C.F.M. Development of a DNA-dosimeter system for monitoring the effects of pulsed ultraviolet radiation. **Photochemical & Photobiological Sciences,** v. 8, p. 111–120, 2009.

SCHUCH, A.P., MORENO, N.C., SCHUCH, N.J., MENCK, C., GARCIA, C.C.M. Sunlight damage to cellular DNA: focus on oxidatively generated lesions. **Free Radical Biology & Medicine**, v. 107, p. 110–124, 2017.

SETLOW. R.B., GRIST, E., THOMPSON, K., WOODHEAD, A.D. Wavelengths effective in induction of malignant melanoma. **Proceedings of the National Academy of Sciences of USA**, v. 90, p. 6666–6670, 1993.

SHIMIZU, Y., IWAI, S., HANAOKA, F., SUGASAWA, K. xeroderma pigmentosum group C protein interacts physically and functionally with thymine DNA glycosylase. **EMBO J**, v. 22, p.164–73, 2003.

SOUFIR N. et al. A prevalent mutation with founder effect in xeroderma pigmentosum group C from North Africa. **Journal of Investigative Dermatology**, v. 130, p. 1537-1542, 2010.

SOUZA, T.A., DEFELICIBUS, A., MENCK, C.F.M. in preparation. Available at: github.com/tiagoantonio/woland.

STARY, A., SARASIN, A. The genetics of the hereditary xeroderma pigmentosum syndrome. **Biochimie**, v. 84, p. 49–60, 2002.

STEENKEN, S., JOVANOVIC, S.V. How easily oxidizable is DNA? One-electron reduction potentials of adenosine and guanosine radicals in aqueous solution. **Journal of the American Chemical Society**, v. 119, p. 617–618, 1997.

STERENBORG, H.J.C.M., VAN DER LEUN, J.C. Tumorogenesis by a long wavelength UV-A source. **Photochemistry and Photobiology**, v. 51, p. 325-330, 1990.

SUTHERLAND, J.C., GRIFFIN, K.P. Absorption spectrum of DNA for wavelengths greater than 300 nm. **Radiation research**, v. 86, p. 399-409, 1981.

TAYLOR, J.S. Unravelling the Molecular Pathway from Sunlight to Skin Cancer. **American Chemical Society**, v. 27, p. 76–82, 1994.

TOMKOVA, M., SCHUSTER-BÖCKLER, B. DNA Modifications: Naturally More Error Prone?. **Trends Genetics**. 34, 627–638, 2018.

TOMMASI, S., DENISSENKO, M.F., PFEIFER, G.P. Sunlight Induces Pyrimidine Dimers Preferentially at 5-Methylcytosine Bases. **Cancer Research**, v. 57, p. 4727-4730, 1997.

TRUCCO, L.D. et al. Ultraviolet radiation-induced DNA damage is prognostic for outcome in melanoma. **Nature Medicine**, v. 25, p. 221–224, 2019.

VAN DER AUWERA, G.A. et al. From fastq data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. **Current Protocols in Bioinformatics**, v. 43. p. 11.10.1-11.10.33, 2013.

VAN DIJK, E. L., AUGER, H., JASZCZYSZYN, Y., THERMES, C. Ten years of next-generation sequencing technology. **Trends in Genetics**, v. 30, p. 418-426, 2014.

VAN KRANEN, H.J. et al. Low incidence of p53 mutations in UVA (365-nm)-induced skin tumors in hair less mice. **Cancer research**, v. 57, p. 1238-1240, 1997.

VAN WEELDEN, H., DE GRUJIL, F.R., VAN DER PUTTE, S.C.J., TOONSTRA, J., VAN DER LEUN, J.C. The carcinogenic risks of modern tanning equipment: Is UV-A safer than UV-B?. **Archives of Dermatological Research**, v. 280, p. 300-307, 1988.

WANG, K., LI, M., HAKONARSON, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. **Nucleic Acids Research**, v. 38, p. 1–7, 2010.

WESTERDAHL, J., INGVAR, C., MASBACK, A., JONSSON, N., OLSSON, H. Risk of cutaneous malignant melanoma in relation to use of sunbeds: further evidence for UV-A carcinogenicity. **British Journal of Cancer,** v. 82, p. 1593–1599, 2000.

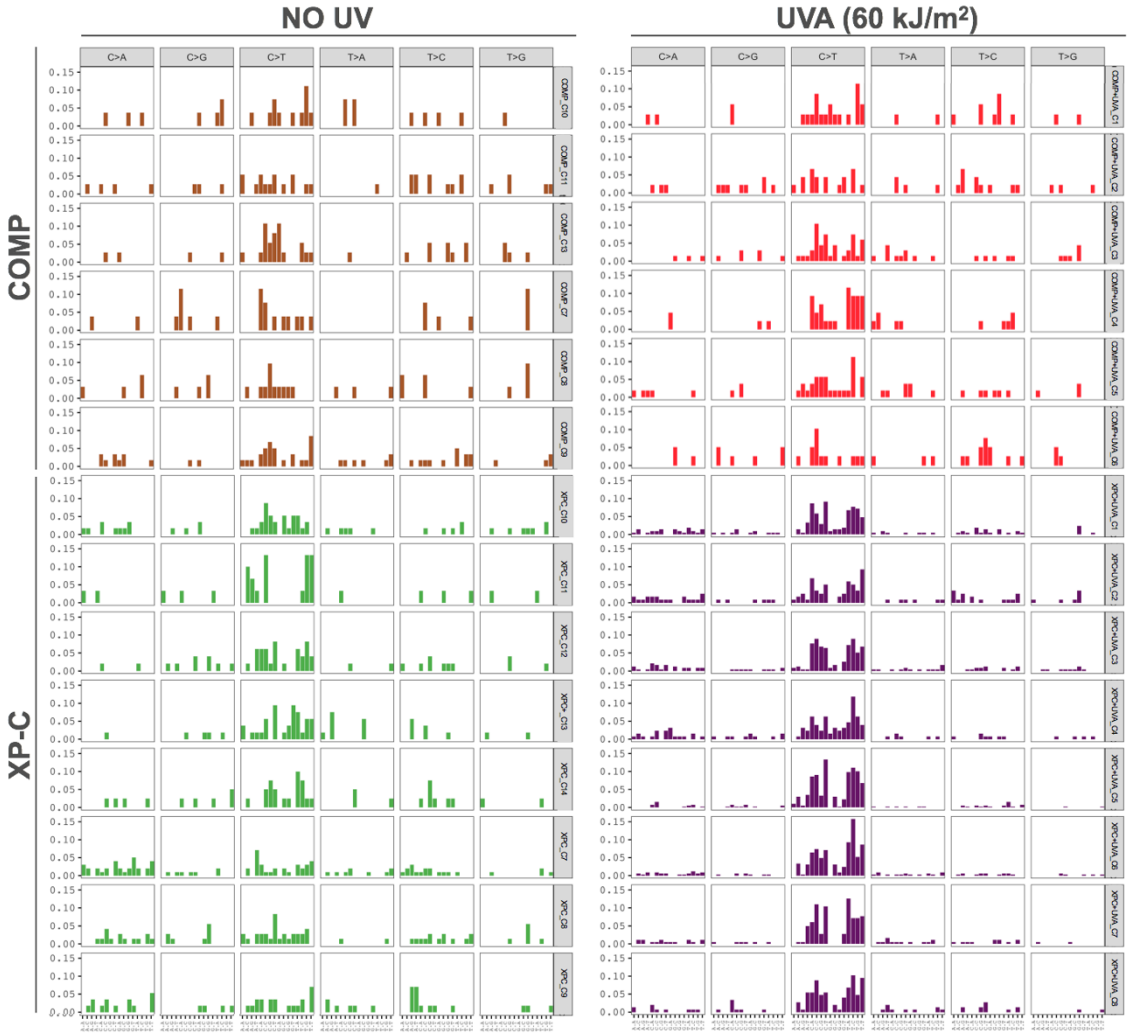
WIKONDAHL, N.M., BRASH, D.E. Ultraviolet radiation induced signature mutations in photocarcinogenesis. **Journal of Investigative Dermatology Symposium Proceedings**, v. 4, p. 6–10, 1999.

WILLIAMSON, C.E. et al. Solar ultraviolet radiation in a changing climate. **Nature Climate Change**, v. 4, p. 434–441, 2014.

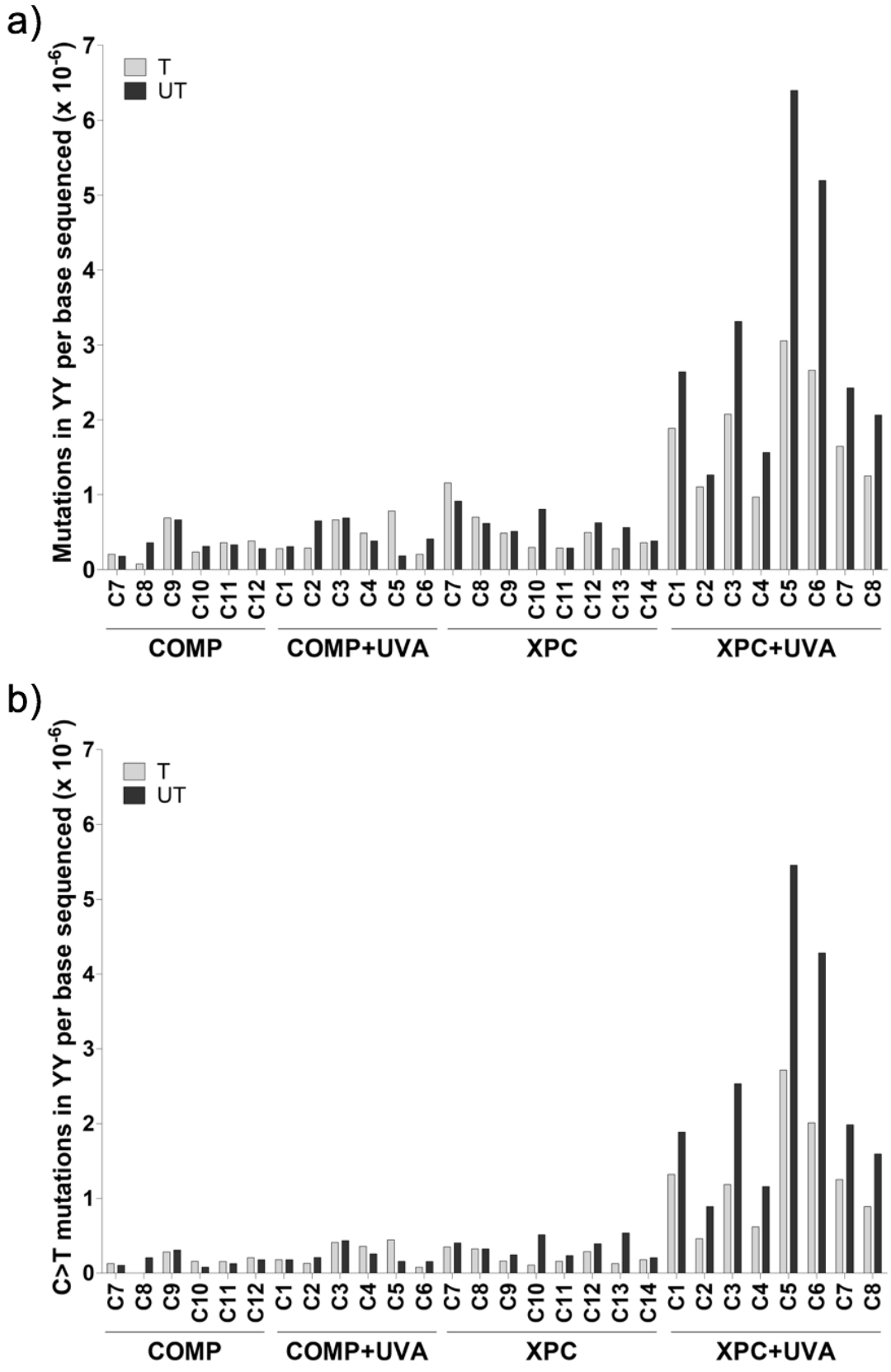
YAGURA, T., MAKITA, K., YAMAMOTO, H., MENCK, C.F.M., SCHUCH, A.P. Biological Sensors for solar ultraviolet radiation. **Sensors**, v. 11, p. 4277-4294, 2017.

ZIEGLER, A. et al. Mutation hotspots due to sunlight in the p53 gene of non'melanoma skin cancers. **Proceedings of the National Academy of Sciences of the USA**, v. 90, p. 4216–20, 1993.

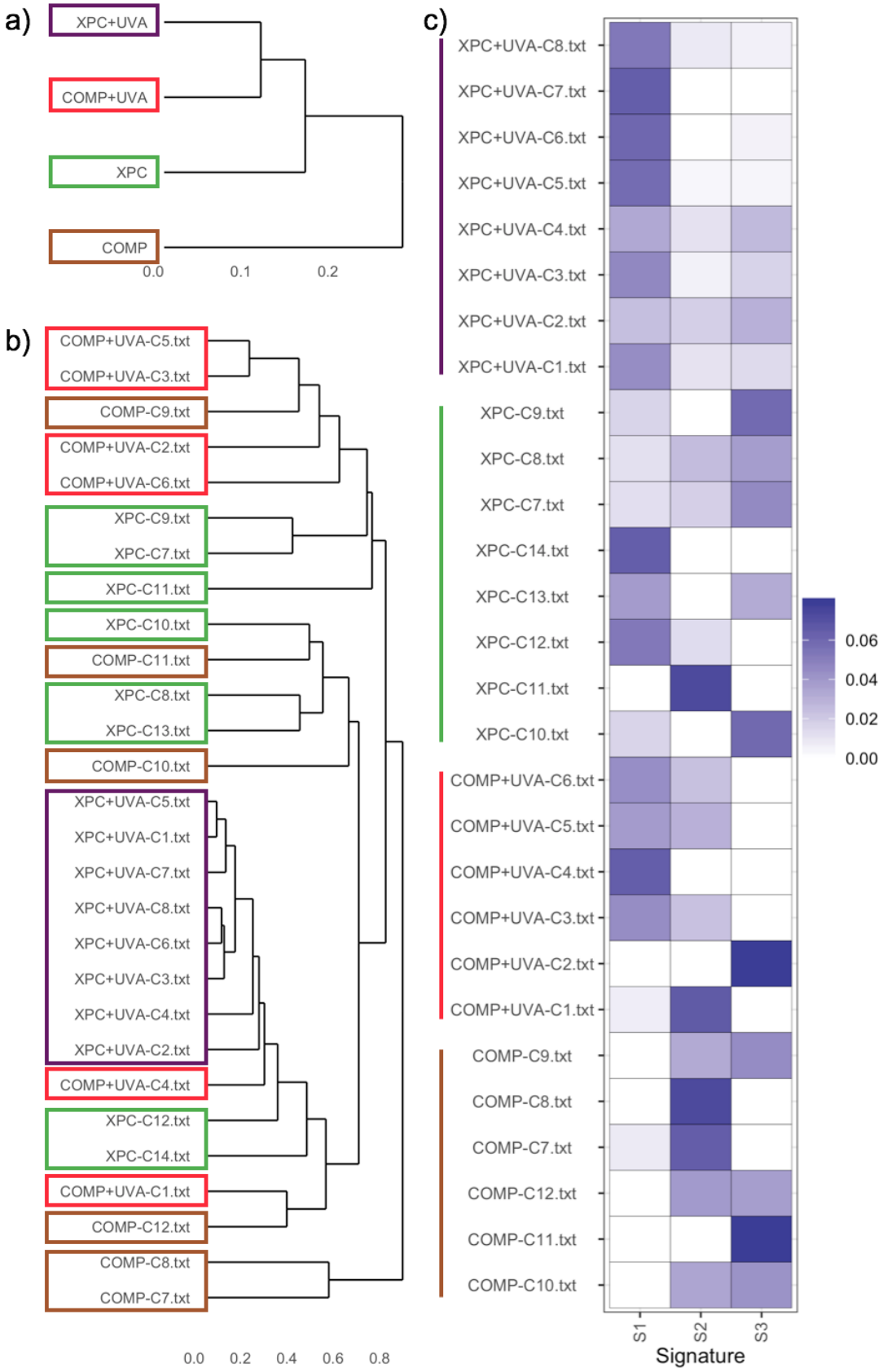
3.7 Supplementary Figures



Supplementary figure S3.1. Somatic mutation spectra for each independent subclone. Exclusive variants filtered for each subclone were employed to characterize the somatic spectra of tri-nucleotide motifs using the SomaticSignatures package. The contribution was calculated by its frequency. Irradiation dose: 60 kJ/m^2 of UVA light.



Supplementary figure S3.2. Transcriptional strand bias for induced mutations in each clone. (a) All mutations at YY motifs per million bases sequenced; and **(b)** C>T transitions at YY motifs present at the transcribed (T) and untranscribed (UT) strands, considering each individual clone.



Supplementary figure S3.3. Similarity of mutations generated in the tri-nucleotide context. **(a)** between experimental groups, **(b)** between all sequenced samples independently, evaluated based on Euclidean distances. **(c)** Contribution of each mutational somatic signature established in this work (Figure 3.9a) to explain the mutational pattern of irradiated or non-irradiated clones independently.

3.8 Supplementary Tables

Supplementary Table S3.1. Detailed list of sequenced clones, library preparation kit and sequencing platform used.

CELL LINE	TREATMENT	CLON	EXOME CAPTURE PLATFORM	SEQUENCING PLATFORM	SNVs
COMP	NO UV	COMP_NT_C7	xGen hybridization capture of DNA libraries, IDT	HiSeq 2500 System - Illumina	26
		COMP_NT_C8	xGen hybridization capture of DNA libraries, IDT	HiSeq 2500 System - Illumina	31
		COMP_NT_C9	xGen hybridization capture of DNA libraries, IDT	HiSeq 2500 System - Illumina	59
		COMP_NT_C10	Agilent - Sureselect QXT V6.	HiSeq 2500 System - Illumina	27
		COMP_NT_C11	xGen hybridization capture of DNA libraries, IDT	HiSeq 2500 System - Illumina	37
		COMP_NT_C12	xGen hybridization capture of DNA libraries, IDT	HiSeq 2500 System - Illumina	37
	UVA 60 kJ/m^2	COMP_A60_C1	xGen hybridization capture of DNA libraries, IDT	HiSeq 2500 System - Illumina	35
		COMP_A60_C2	Agilent - Sureselect QXT V6.	HiSeq 2500 System - Illumina	45
		COMP_A60_C3	xGen hybridization capture of DNA libraries, IDT	HiSeq 2500 System - Illumina	67
		COMP_A60_C4	xGen hybridization capture of DNA libraries, IDT	HiSeq 2500 System - Illumina	43
		COMP_A60_C5	Agilent - Sureselect QXT V6.	HiSeq 2500 System - Illumina	53
		COMP_A60_C6	xGen hybridization capture of DNA libraries, IDT	HiSeq 2500 System - Illumina	39
XP-C	NO UV	XPC_NT_C7	Illumina - Nextera Rapid Capture Exome	NextSeq 550 Instrument - Illumina	99
		XPC_NT_C8	Illumina - Nextera Rapid Capture Exome	NextSeq 550 Instrument - Illumina	72
		XPC_NT_C9	Illumina - Nextera Rapid Capture Exome	NextSeq 550 Instrument - Illumina	57
		XPC_NT_C10	Illumina - Nextera Rapid Capture Exome	NextSeq 550 Instrument - Illumina	57
		XPC_NT_C11	Agilent - Sureselect QXT V6.	HiSeq 2500 System - Illumina	30
		XPC_NT_C12	Agilent - Sureselect QXT V6.	HiSeq 2500 System - Illumina	49
		XPC_NT_C13	xGen hybridization capture of DNA libraries, IDT	HiSeq 2500 System - Illumina	53
		XPC_NT_C14	xGen hybridization capture of DNA libraries, IDT	HiSeq 2500 System - Illumina	40
	UVA 60 kJ/m^2	XPC_A60_C1	Illumina - Nextera Rapid Capture Exome	NextSeq 550 Instrument - Illumina	207
		XPC_A60_C2	Illumina - Nextera Rapid Capture Exome	NextSeq 550 Instrument - Illumina	118
		XPC_A60_C3	Illumina - Nextera Rapid Capture Exome	NextSeq 550 Instrument - Illumina	235
		XPC_A60_C4	Illumina - Nextera Rapid Capture Exome	NextSeq 550 Instrument - Illumina	127
		XPC_A60_C5	Agilent - Sureselect QXT V6.	HiSeq 2500 System - Illumina	397
		XPC_A60_C6	Agilent - Sureselect QXT V6.	HiSeq 2500 System - Illumina	324
		XPC_A60_C7	Agilent - Sureselect QXT V6.	HiSeq 2500 System - Illumina	182
		XPC_A60_C8	Agilent - Sureselect QXT V6.	HiSeq 2500 System - Illumina	147

Supplementary Table S3.2. Sequencing alignment statistics.

Cell line	**COMP**									
Treatment	**NO UV**						**UVA (60 kJ/m^2)**			
Codigo clone	COMP_NT_C7	COMP_NT_C8	COMP_NT_C9	COMP_NT_C10	COMP_NT_C11	COMP_NT_C12	COMP_A60_C1	COMP_A60_C2	COMP_A60_C3	COMP_A60_C4
kit - platform	IDT-HiSeq	IDT-HiSeq	IDT-HiSeq	Agilent-HiSeq	IDT-HiSeq	IDT-HiSeq	IDT-HiSeq	Agilent-HiSeq	IDT-HiSeq	IDT-HiSeq
% duplicate reads	10.4%	13.3%	11.9%	14.1%	11.7%	14.2%	13.0%	32.9%	12.7%	12.5%
Total reads	58290512	69800740	64687684	42733950	62219422	77985694	74274088	40152704	70551582	66156986
Unique reads	52211052	60543787	56992923	36713917	54919863	66945747	64629098	26946780	61609576	57870329
Number of reads mapped	52128883	60463277	56879550	36399188	54733063	66812496	64445986	26656505	61492308	57816912
% reads mapped	99.8%	99.9%	99.8%	99.1%	99.7%	99.8%	99.7%	98.9%	99.8%	99.9%
Bases on target	4506498312	5396798833	5002951156	2735842765	4793859620	6082811822	5724217734	2517804272	5446668881	5172412558
% bases on target	88.6%	89.0%	88.7%	92.8%	88.7%	87.8%	88.8%	91.3%	89.0%	88.0%
Bases off target	906824178	1003589912	979874948	375272240	934345421	1340551807	1110928601	428663732	1021664246	1125166889
% off target	11.4%	11.0%	11.3%	7.2%	11.3%	12.2%	11.2%	8.7%	11.0%	12.0%
% of target bp not covered	1.6%	1.7%	1.6%	3.2%	1.7%	1.6%	1.6%	3.5%	1.7%	1.6%
% of target bp covered at ≥ 1X	98.2%	98.2%	98.2%	96.5%	98.2%	98.3%	98.3%	96.2%	98.2%	98.3%
% of target bp covered at ≥ 10X	97.2%	97.6%	97.5%	82.8%	97.4%	97.4%	98.3%	78.0%	97.6%	97.2%
% of target bp covered at ≥ 20X	93.3%	95.6%	94.9%	67.0%	94.3%	93.9%	95.8%	57.7%	95.8%	92.7%
Average depth of coverage	116	139	129	71	123	156	147	66	140	133
Raw variants called	163558	154353	161698	135919	154383	191182	167699	135883	159418	186362
Novel SNPs	9471	9324	9534	7776	9088	12658	9901	9547	9614	11993
Percent SNPs (dbSNP)	94.2%	94.0%	94.1%	94.3%	94.1%	93.4%	94.1%	93.0%	94.0%	93.6%
Raw Indels called	22623	22426	22716	15268	20904	26000	23852	14786	23152	25544
Novel Indels	2248	2427	2346	1708	2137	2822	2518	1755	2430	2742
Percent Indels (dbSNP)	89.8%	89.2%	89.7%	88.8%	89.8%	89.1%	89.4%	88.1%	89.5%	89.3%

Continuation Supplementary Table S3.2. Sequencing alignment statistics.

Cell line	**COMP**		**XPC**							
Treatment	**UVA (60 kJ/m^2)**		**NO UV**							
Codigo clone	COMP_A60_C5	COMP_A60_C6	XPC_NT_C7	XPC_NT_C8	XPC_NT_C9	XPC_NT_C10	XPC_NT_C11	XPC_NT_C12	XPC_NT_C13	XPC_NT_C14
kit - platform	Agilent-HiSeq	IDT-HiSeq	illumina-NextSeq	illumina-NextSeq	illumina-NextSeq	illumina-NextSeq	Agilent-HiSeq	Agilent-HiSeq	IDT-HiSeq	IDT-HiSeq
% duplicate reads	17.6%	10.9%	43.0%	39.6%	28.5%	28.9%	21.1%	19.3%	10.8%	10.7%
Total reads	57987348	55971242	97664760	111648730	68832370	98521516	78345206	57833330	49784418	58223756
Unique reads	47763134	49862978	55700432	67386608	49183253	70011759	61831686	46651634	44396315	52003912
Number of reads mapped	47424424	49806887	55410487	66968617	48868392	69151899	61271351	46260320	44346890	51899395
% reads mapped	99.3%	99.9%	99.5%	99.4%	99.4%	98.8%	99.1%	99.2%	99.9%	99.8%
Bases on target	3758994437	4360525830	6343060699	6667678381	3936698470	5560335552	4958444398	3818707552	3821121804	4469003853
% bases on target	93.0%	87.6%	72.2%	67.2%	71.0%	71.3%	92.7%	93.6%	93.6%	93.6%
Bases off target	497336772	986449664	3546543880	4995686820	2428301737	3445205324	704942666	452180209	755370630	891998234
% off target	7.0%	12.4%	27.8%	32.8%	29.0%	28.7%	7.3%	6.4%	11.5%	11.5%
% of target bp not covered	2.9%	1.6%	1.0%	0.9%	0.9%	0.9%	2.6%	2.7%	1.7%	1.7%
% of target bp covered at ≥ 1X	96.8%	98.3%	98.7%	98.8%	98.7%	98.8%	97.1%	96.9%	98.2%	98.2%
% of target bp covered at ≥ 10X	86.3%	96.9%	94.7%	96.1%	94.3%	96.6%	90.5%	88.7%	97.0%	97.3%
% of target bp covered at ≥ 20X	73.9%	91.5%	85.9%	88.3%	83.5%	90.9%	81.4%	77.3%	92.1%	94.2%
Average depth of coverage	98	112	171	180	106	150	129	100	98	115
Raw variants called	148938	178175	246865	399525	194798	269891	177562	145321	145974	154345
Novel SNPs	9072	11274	29620	40027	19518	26213	11523	8323	8626	9108
Percent SNPs (dbSNP)	93.9%	93.7%	88.0%	90.0%	90.0%	90.3%	93.5%	94.3%	94.1%	94.1%
Raw Indels called	17684	23830	53802	71585	34990	47983	21538	17627	20076	21770
Novel Indels	2137	2444	11314	12544	6112	9223	2538	2203	2099	2249
Percent Indels (dbSNP)	87.9%	89.7%	79.0%	82.5%	82.5%	80.8%	88.2%	87.5%	89.5%	89.7%

Continuation Supplementary Table S3.2. Sequencing alignment statistics.

Cell line	**XPC**							
Treatment	**UVA (60 kJ/m^2)**							
Codigo clone	XPC_A60_C1	XPC_A60_C2	XPC_A60_C3	XPC_A60_C4	XPC_A60_C5	XPC_A60_C6	XPC_A60_C7	XPC_A60_C8
kit - platform	illumina-NextSeq	illumina-NextSeq	illumina-NextSeq	illumina-NextSeq	Agilent-HiSeq	Agilent-HiSeq	Agilent-HiSeq	Agilent-HiSeq
% duplicate reads	43.7%	29.3%	41.9%	30.4%	17.9%	24.7%	17.9%	21.6%
Total reads	122856326	89549622	115115552	82002716	55762046	43807386	42778024	80324554
Unique reads	69194281	63305129	66849595	57056855	45777117	32983652	35122562	62992940
Number of reads mapped	68649779	62789432	66472373	56680943	45406529	32584237	34783737	62411744
% reads mapped	99.2%	99.2%	99.4%	99.3%	99.2%	98.8%	99.0%	99.1%
Bases on target	7862681289	4841760699	7622495179	4715100039	3693398826	2699698131	2709501651	5069534031
% bases on target	71.1%	69.1%	72.5%	71.4%	93.6%	92.2%	92.9%	92.6%
Bases off target	4676945943	3156260259	4268258154	2800690540	439503344	420936329	375257808	736208766
% off target	28.9%	30.9%	27.5%	28.6%	6.4%	7.8%	7.1%	7.4%
% of target bp not covered	0.9%	0.9%	1.0%	0.9%	2.9%	2.9%	2.9%	2.5%
% of target bp covered at ≥ 1X	98.8%	98.7%	98.8%	98.7%	96.8%	96.8%	96.8%	97.2%
% of target bp covered at ≥ 10X	96.1%	95.0%	95.4%	95.1%	87.4%	83.4%	84.9%	90.8%
% of target bp covered at ≥ 20X	89.8%	86.6%	87.6%	86.4%	75.2%	66.3%	69.4%	81.8%
Average depth of coverage	212	130	205	127	96	71	71	132
Raw variants called	326390	247927	306801	218891	141613	145820	140193	180735
Novel SNPs	38870	27866	35958	23076	9309	9040	7851	12529
Percent SNPs (dbSNP)	88.1%	88.8%	88.3%	89.5%	93.4%	93.8%	94.4%	93.1%
Raw Indels called	69376	49658	58118	41286	17091	15867	15776	21940
Novel Indels	14422	9244	11179	7182	2147	1558	1739	2692
Percent Indels (dbSNP)	79.2%	81.4%	80.8%	82.6%	87.4%	90.2%	89.0%	87.7%

CHAPTER 4 – Results not included in the manuscripts

4.1 Production and validation of the isogenic control cell line COMP

In order to obtain an isogenic control for mutagenesis experiments, a cell line that constitutively express the *XPC* gene was produced using recombinant lentivirus derived vectors. Briefly, the XPC overexpressed cell was generated using the lentiviral vector pLV[Exp]-Puro-CMV>hXPC (Vector Builder, Guangzhou, China. Figure 4.1a). This plasmid was co-transfected with the lentiviral auxiliary plasmids pRRE, pREV and pVGVS into HEK 293 T cells using the polyethyleneimine method (kit). Then, the recombinant lentivirus was used to transduce XP-C cells (XP4PA-SV40), and selection was performed by puromycin (1.5 µg/mL) for 2 weeks, resulting in the stably expressing XPC+ cell line, denominated COMP. Then the population (COMP_pop) and three clones randomly selected (COMP_C1, COMP_C2, and COMP_C3) were used to test phenotype. It is worth to remember that the XP-C cell line was also sub-cloned in order to obtain a homogeneous population that would allow a more accurate detection of the UV-radiation-induced mutation signatures. Likewise, the cell line MRC5-SV40 was included as a wild-type fibroblast control expressing the *XPC* gene.

First, cell viability upon UVC exposure was determined by using Cell Proliferation Kit II (XTT), according to the manufacturer's specifications (Sigma-Aldrich). The assay is based on the ability of the metabolically active cells to cleave the tetrazolium salt, producing the orange dye formazan. Formazan concentration is proportional to the number of metabolically active cells and is quantified by spectrometry at 490 and 715 nm using a GloMax®-Multi Detection System spectrophotometer (Promega, Madison, WS, USA). The percentage of cell viability was calculated as the ratio of absorbance values of treatments in relation to the non-irradiated controls. In all cases, a dose-dependent decrease in cell viability was evidenced, being XP-C and COMP_C3, the most sensitive to UVC and MRC5, the most resistant (Figure 4.1b). However, the functional complementation of XP-C cells by lentivirus increased the cell viability of COMP_pop, COMP_C1 and COMP_C2 cells. Remarkably, the MRC5 cell line only showed a slightly better viability regarding COMP_pop and COMP_C1, demonstrating that the NER phenotype rescue in the complemented cell line was obtained.

Then, the expression of the *XPC* gene was investigated by quantitative RT-PCR. Briefly, total RNA was extracted using the PureLink™ RNA Mini Kit

(ThermoFisher, Waltham, MA, USA), quantified by spectrophotometer and 1 µg of RNA was used to obtain cDNA using the HighCapacity cDNA Reverse Transcription kit (ThemoFisher), according to manufacturers' instructions. The qRT-PCR reactions were performed using 200 nM of each primer (forward 5'-CATCGTGGGAGCCATCGTAAG-3' and reverse 5'-TCTCACCATCGCTGCACATTTT-3', kindly provided by Dr. Nadja de Souza Pinto, Instituto de Química, Universidade de São Paulo) and the Fast SYBR Green Master Mix (Thermo Fisher,) following the manufacturer instructions. The PCR reaction was performed in triplicate and analyzed using the 7500 real time PCR system (Thermo Fisher). Expression data were generated using the delta-delta Ct method ($2^{-\Delta\Delta Ct}$) and the *GAPDH* as a housekeeping gene (figure 4.1c). First, the expression of *XPC* gene was not observed in XP-C cell (XP4PA-SV40). The increased expression of the *XPC* gene was detected in all complemented cells examined, except in the clone COMP-C3. The complemented cells COMP_pop, COMP_C1 and COMP_C2 expressed the *XPC* gene at high levels when, compared to MRC5 cell line.

Moreover, the cellular protein detection was performed by *western blot* (figure 4.1d) following the standard protocol. Briefly, 50 µg of protein extract were separated on a Bolt® 4-12% Bis-Tris Plus SDS polyacrylamide gel (Life Technologies Corporation) and transferred to a nitrocellulose membrane using the iBlot® Gel Transfer Stacks Nitrocellulose kit, according to the manufacturer's specifications (Life Technologies Corporation). The following primary antibodies were used: anti-XPC (1:1000; rabbit 14768, Cell signaling) and anti-GAPDH (1:2000; mouse sc-32233, Santa Cruz Biotechnology). Finally, the expression of the specific proteins was detected using a horse-radish peroxidase method (Luminata™ Forte Western HRP substrate kit Millipore Corporation, Billerica, MA, USA) followed by the detection in a chemiluminescence detector system (Alliance Q9, Uvitec, Cambridge, England, UK). The XPC protein was detected in the COMP_pop and in the COMP_C1 clone, but not on COMP_C2 clone despite the gene expression detected by qPCR and an improved sensitivity to UVC irradiation. On the other hand, as expected, the XP-C and COMP_C3 cells did not show expression of the XPC protein.

Taken all the results into account, the COMP_C1 clone was chosen as a better control and was selected to continue the mutagenesis experiments. From here on (and in chapters 2 and 3 of this document), this clone was denominated COMP cell line, the isogenic control. Finally, the complementation of the XP-C cell line was confirmed by

the exome sequencing which showed the deletion of the CT nucleotides in exon 9 of the *XPC* gene in heterozygosity for COMP cell line, but in homozygosity for XP-C cell line (figure 4.2).

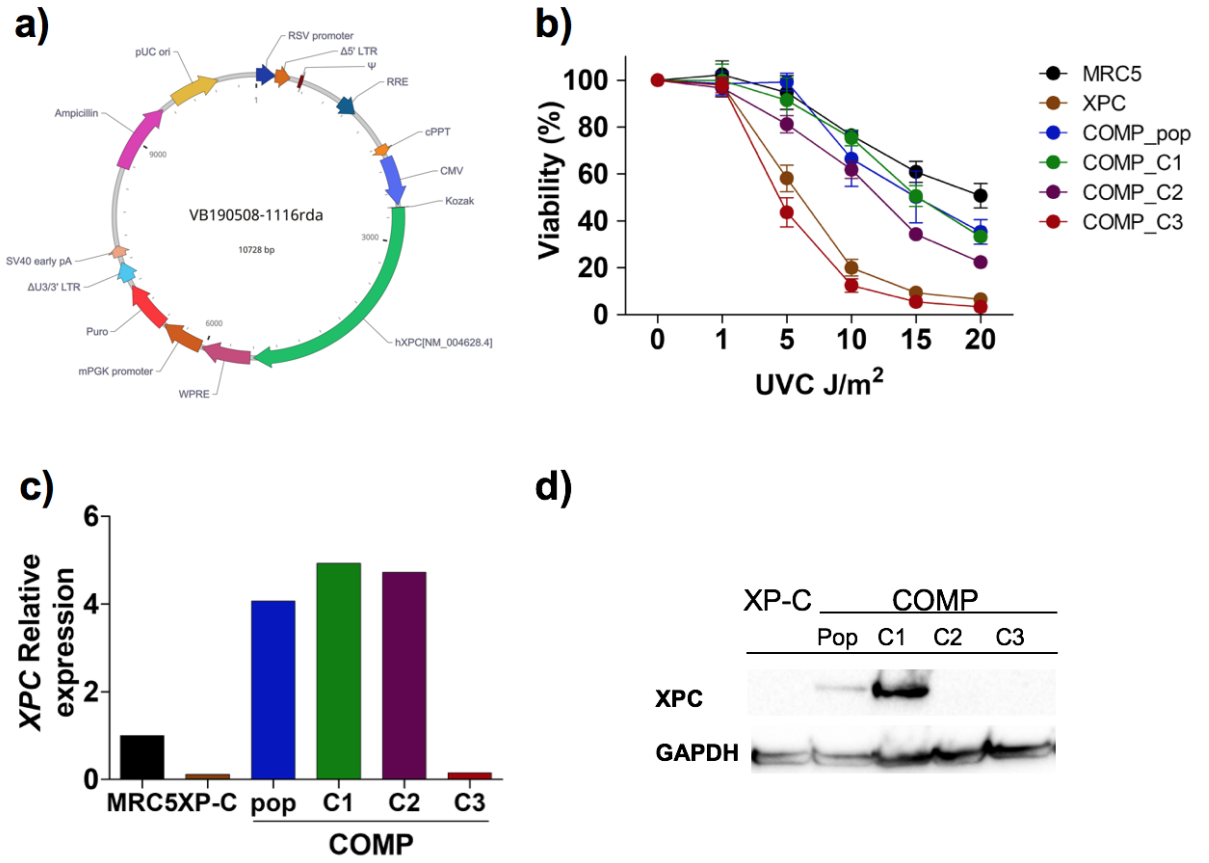


Figure 4.1. Production of the isogenic model expressing *XPC*. a) Representative map of the vector pLV[Exp]-Puro-CMV>hXPC, which was used to obtain lentivirus used to transduce the cell line XP-C (XP4PA-SV40). The validation and selection of clone expressing XPC used to the experiments shown in chapters 2 and 3 was determined by phenotype and the XPC expression. **b)** Dose-response curves of cells using XTT assay after increasing doses of UVC irradiation. **c)** qRT-PCR of the *XPC* gene expression. **d)** Cellular protein detection of XPC and GAPDH proteins by western Blot.

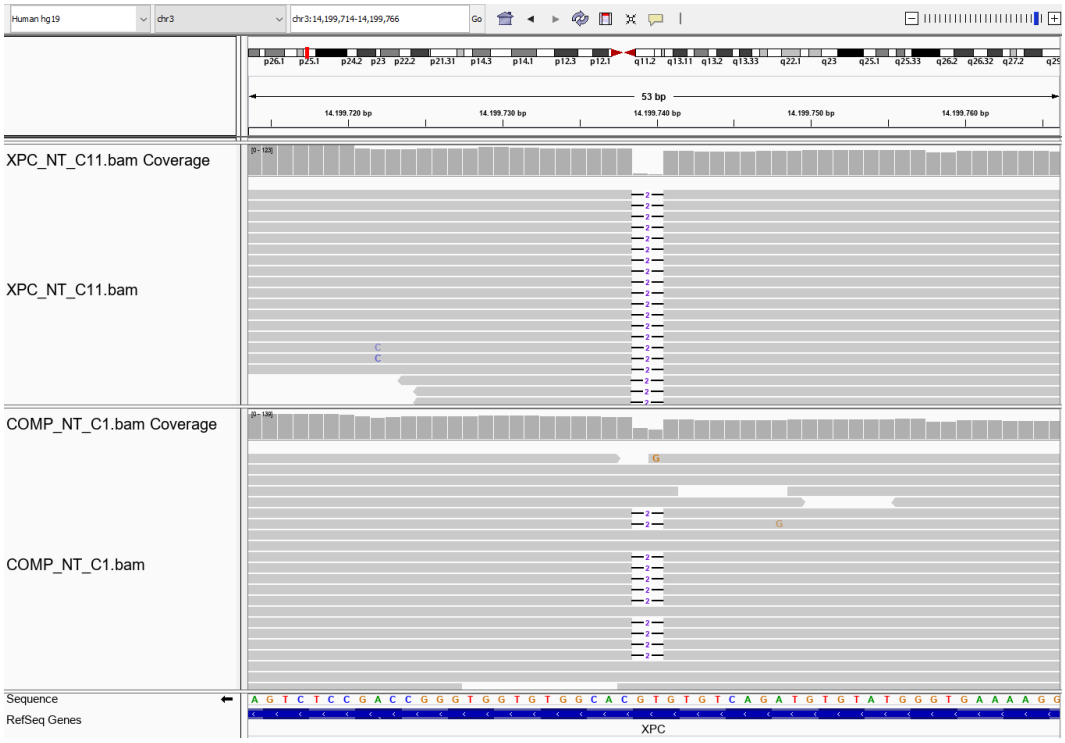


Figure 4.2. Confirmation of the complementation of COMP cell line by the detection of the heterozygosity of XPC mutation (c.1643_1644delTG).

4.2 The absence of XPC protein leads to an increase of C>T transitions

The lack of XPC protein leads to XP syndrome, characterized by a high incidence of cancer in sunlight exposed areas due to a deficient NER pathway. However, until now it has not been explored if the lack of the XPC protein leads to the increase of a specific base substitution or a change in the mutation spectra of spontaneous mutations. To answer this question, all non-irradiated sequenced sub-clones of the isogenic cell line COMP (11 sub-clones) and XP-C (14 sub-clones) were joined to increase the experimental n, and thus, achieve a more accurate identification of mutations generated as consequence of the absence of the XPC protein in non-irradiated conditions. The call of exclusive variants was performed for 25 sub-clones as previously described (2.3.9 and 3.3.7 sections).

The absence of the XPC protein lead to a small but statistically significant increase, by 1.7-fold, in the mutation frequency (SNVs per 10^6 bases sequenced) of XP-C cells when compared to the COMP cell line (figure 4.3a). Then, SNVs generated in the six possible types of base substitutions (C:G>A:T, C:G>G:C, C:G>T:A, T:A>A:T, T:A>C:G and T:A>G:C) were investigated in order to explore the effect of the absence of XPC protein. Just the C>T transitions were significantly increased in the XP-C cell line. The T>A transversions exhibit a slight increase but this was not significant and

the other four types of base substitution showed no changes (figure 4.3b). Finally, the somatic spectrum of point nucleotide mutations in the tri-nucleotide contexts was analyzed to check for the sequence context of the induced mutations (figure 4.3c). Both cell lines exhibit a homogeneous spectrum between the 96 possible tri-nucleotide combinations without large changes between them, but with mutations slightly concentrated in the C>T transition, as expected. A small concentration of C>T mutations in the CCN sequence context for the COMP sub-clones and on CCN, ACY and TCN in the XP-C sub-clones were observed. Similar to the results previously obtained in chapters 2 and 3. These results support the idea that the increased mutagenesis of XP-C cells could be related to the spontaneous deamination of 5mC, the most common type of mutations observed in cancer (previously discussed in chapter 2, section 2.5) (Lindahl 1993, Tomkova & Schuster-Böckler 2018).

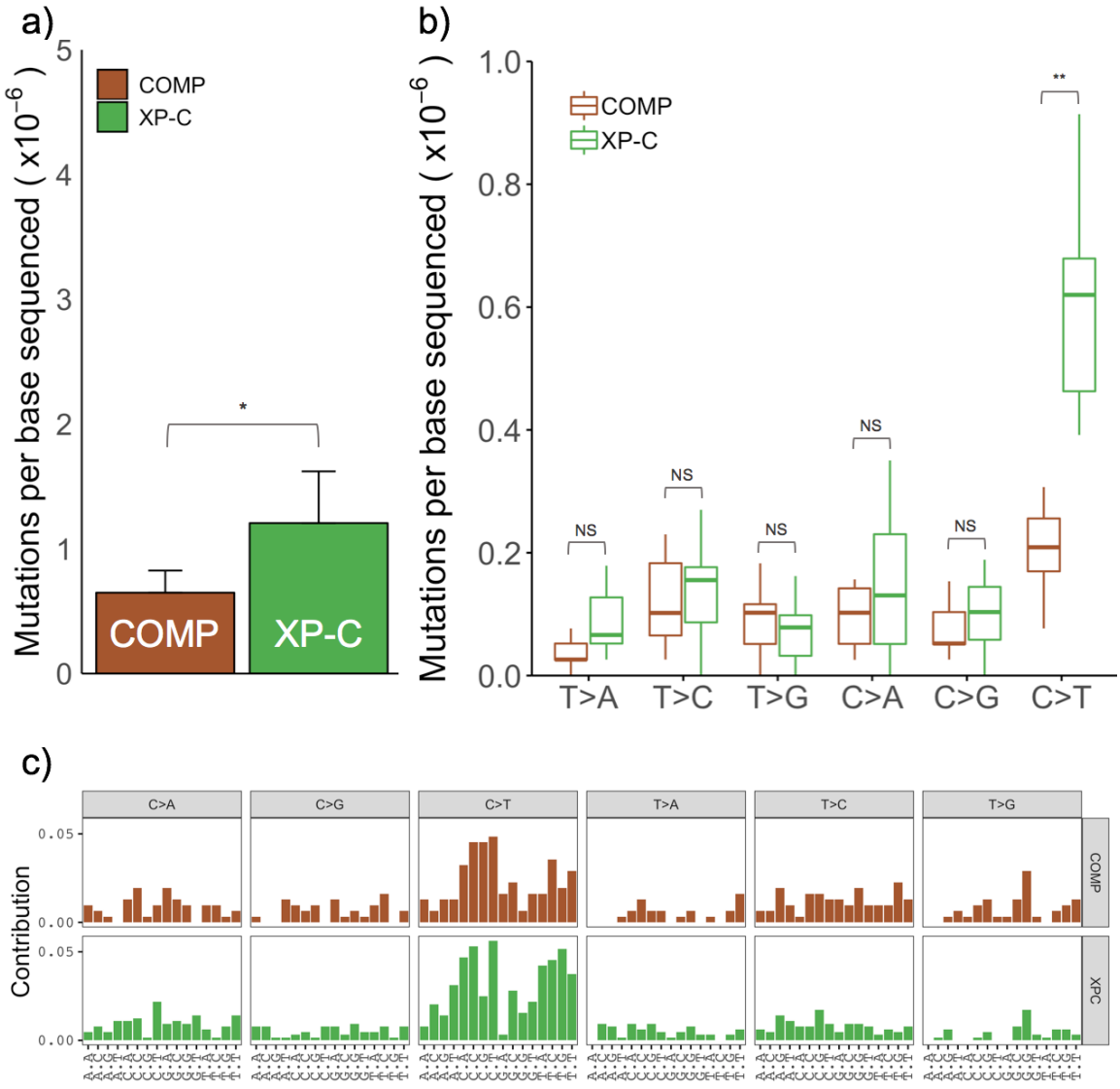


Figure 4.3. The absence of XPC protein leads to an increase of C>T transitions. a) Unique point mutations within exons and splicing sites per sequenced base (n = 11 for COMP and 14

for XP-C sub-clones). **b)** Exclusive mutations generated in the six possible types of base substitutions. The data represent mean and standard deviation (SD); statistical differences were calculated by non-parametric permutation tests: $P < 0.001$ (***), $P < 0.01$ (**), $P < 0.05$ (*), $P > 0.05$ (NS). **c)** Somatic mutation spectra of XPC protein proficient and deficient isogenic cell lines.

4.3 MRC5 as an additional control

A completely independent set of mutagenesis experiments was performed by using the MRC5 cell line, a wild-type lung fibroblast, as an additional control expressing the XPC gene. Four clones of each condition were sequenced: 4 non-irradiated, 4 UVA-irradiated (60 kJ/m^2) and 4 UVB-irradiated (120 J/m^2) for a total of 12 clones (Table 4.1). The exclusive variants selected and analysed by following the protocol previously described, but without the first cloning step (2.3.9 and 3.3.7 sections). The details about kit used for DNA library preparation, platform of sequencing, coverage and statistics of sequencing for all sequenced clones are provided in Supplementary table S4.1.

Table 4.1. Detailed list of sequenced clones. 12 Clones of the MRC5 cell line were sequenced by using the Nextera Rapid Capture Exome kit and the NextSeq 550 Instrument - Illumina.

CELL LINE	TREATMENT	CLON	SNVs
MRC5	NO UV	MRC5_C1	101
		MRC5_C2	95
		MRC5_C3	80
		MRC5_C4	101
	UVA 60 kJ/m^2	MRC5_UVA_C1	127
		MRC5_UVA_C2	84
		MRC5_UVA_C3	165
		MRC5_UVA_C4	109
	UVB 120 J/m^2	MRC5_UVB_C1	132
		MRC5_UVB_C2	96
		MRC5_UVB_C3	114
		MRC5_UVB_C4	138

UVA and UVB irradiation increased the mutation frequency in MRC5 cells by 1.3 and 1.2 fold (figure 4.4a). The basal mutation frequency of MRC5 clones (non-irradiated) was of more than triple when compared to non-irradiated COMP cells (chapter 2 and 3). Probably this was due because the first cloning stage was not carried out for MRC5 cell line (2.3.3 and 3.3.2 sections), so the population's genetic background was not homogenized before the selection of clones for exome sequencing. This supports the necessity of include a pre-cloning step to lead to a more accurate detection of mutations.

The analysis of the six types of base substitutions exhibits an increase in the

T>C transitions and C>G transversions after UVA or UVB irradiation. While, the main UV-induced mutation, the C>T transition was only significantly increased by UVB irradiation, although slightly (figure 4.4b). Interestingly, for the COMP cell line, the isogenic control, no increase was observed in any of the 6 types of base substitutions after both types of UV-irradiation. Finally, the somatic spectrum of point nucleotide mutations in the tri-nucleotide contexts exhibit a more or less flat and homogeneous spectrum between the 96 possible tri-nucleotide combinations, with mutations slightly concentrated in the C>A, C>T and T>C base substitutions (figure 4.3c), but without large changes between control and irradiates clones.

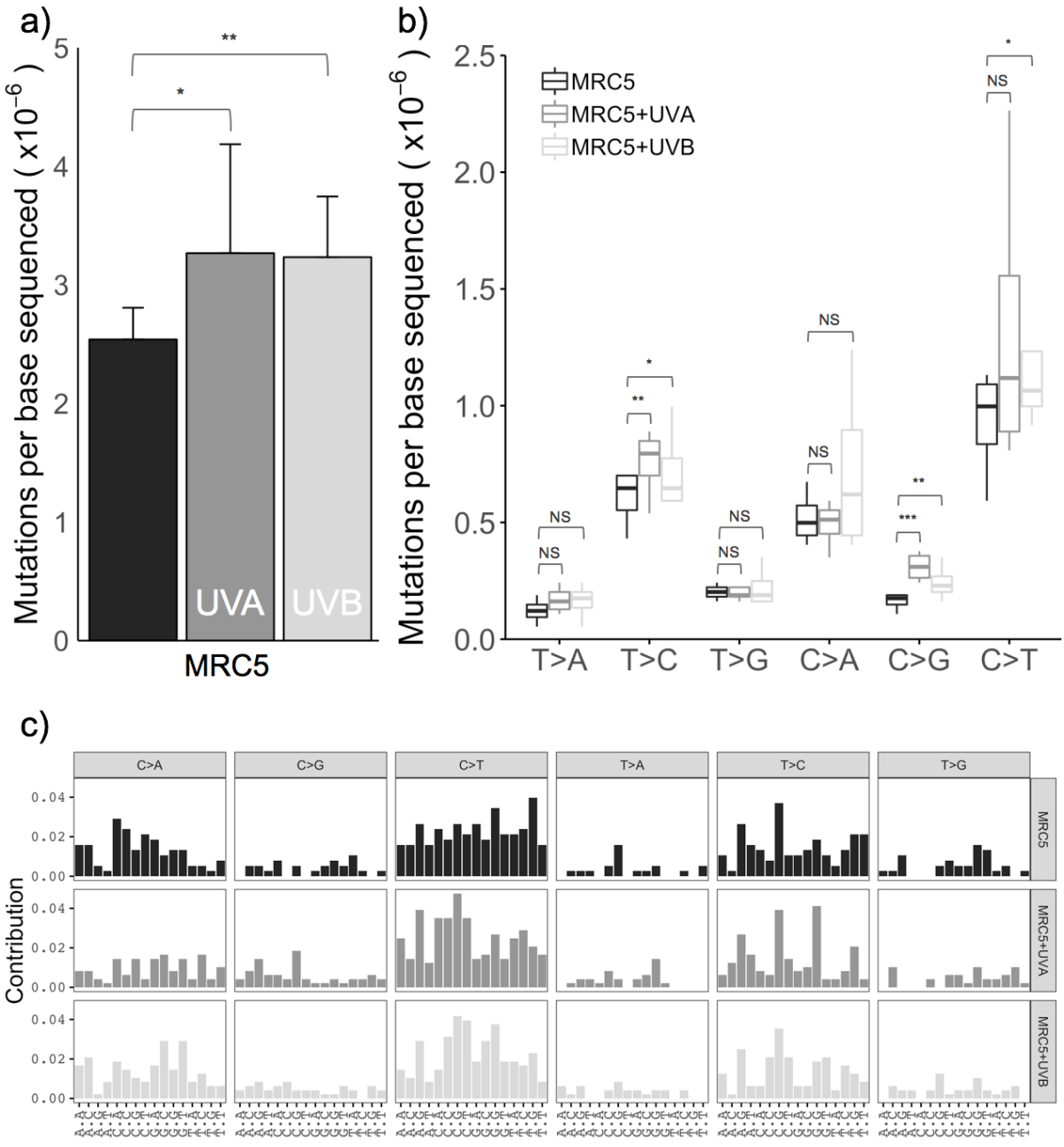


Figure 4.4. MRC5, as an additional control. a) Unique point mutations within exons and

splicing sites per sequenced base (n = 12 clones, 4 for each condition), after UVA (60 kJ/m^2) or UVB (120 J/m^2) irradiation. **b)** Exclusive mutations generated in the six possible types of base substitutions. The data represent mean and standard deviation (SD); statistical differences were calculated by non-parametric permutation tests: $P < 0.001$ (***), $P < 0.01$ (**), $P < 0.05$ (*), $P > 0.05$ (NS). **c)** Somatic mutation spectra of XPC protein proficient and deficient isogenic cell lines.

4.4 Co-treatment of UVA-irradiated cells with the antioxidant NAC improves cell survival rates of XP-C cells.

It has been well established that UVA light induce oxidative stress in irradiated cells. Evidence of ROS generation, lipid peroxidation, oxidized proteins and DNA damage induced by UVA-irradiation has been generated (Ridley et al. 2009, Sage et al. 2012, Marionnet et al. 2014, Schuch et al. 2017). Mutagenesis induced by UVA light is mainly related to the induction of oxidized bases and pyrimidine dimers. However, recently it was proposed that oxidation of proteins involved in DNA repair pathways could reduce the DNA repair capacity of cells and therefore make the repair of the induced damage slower, so it is an important factor to be considered (Runger et al. 2012, Karran & Brem 2016). This idea is supported by the protective effect of antioxidants as vitamin E and NAC against UVA deleterious effects in irradiated keratinocytes and fibroblasts, respectively (Morley et al. 2003, Delinasios et al. 2018, Moreno et al. 2019).

Cell survival experiments (clonogenic assay) were performed as indicated in sections 2.3.5 and 3.3.4, including a co-incubation with 2.5 mM of NAC before (16–18 h) and after UVA irradiation (as previously described in Moreno et al. 2019). Just for this experiment, the cell line XP4PA-T1 was used as control. This lineage was derived from the XP-C cells, for which its ΔTG mutation was corrected directly through a TALE nuclease. Such genetic correction allows the recovery of UV resistance by the restoration of full-length XPC protein that is active (Dupuy et al. 2013).

The survival experiments showed a significant improvement for XPC deficient and proficient cells after UVA irradiation when co-incubated with NAC, but not when irradiated with UVB (Figure 4.5). However, NAC did not completely restore the cell survival to the levels of non-irradiated cells. This indicates that not only indirect damage generated by ROS, but also direct damage (pyrimidine dimers) are interfering in cell survival after UVA irradiation on both cell lines. This result confirms that on NER deficient cells the use of antioxidants could help in the protection from some deleterious effects generated by UVA-irradiation. However, it is still necessary to study

if with the use of antioxidants, the amount and type of UVA-induced mutations will be modified.

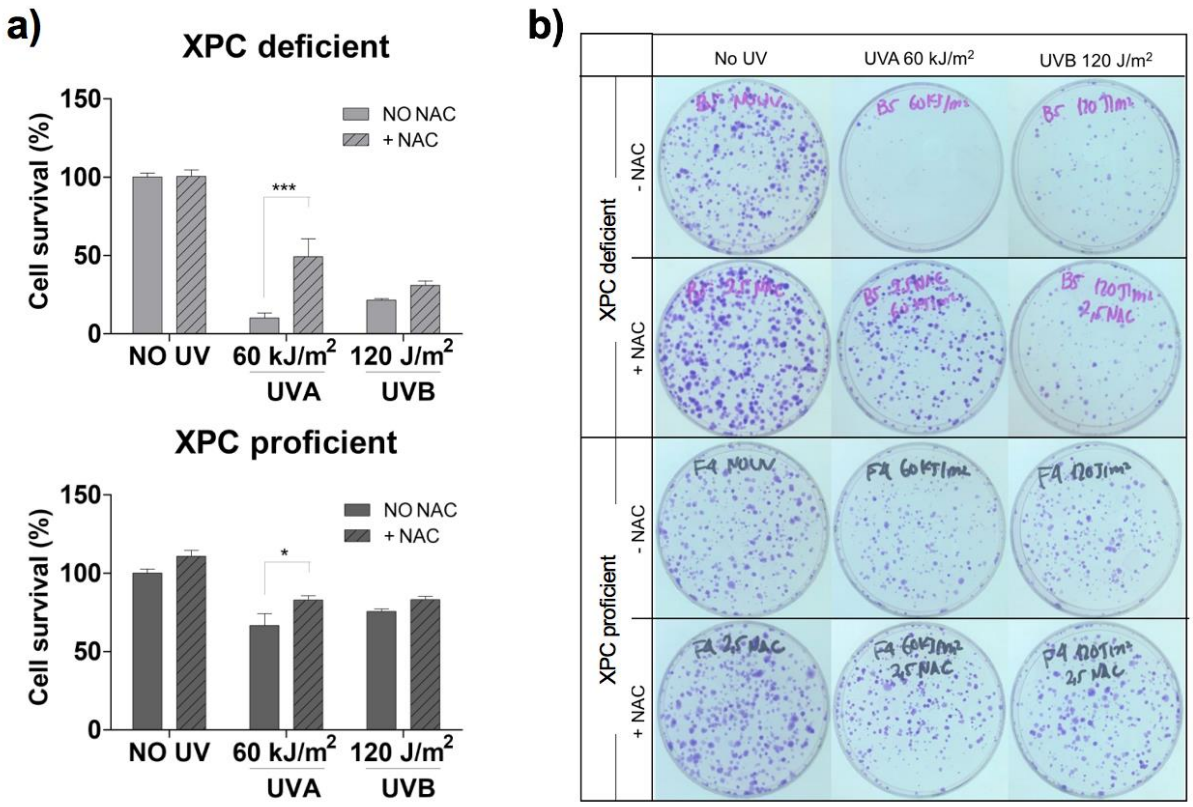


Figure 4.5. NAC improves cell survival rate of XP-C cells when irradiated with UVA. a) Cell survival of XP-C proficient and deficient sub-cloned populations after irradiation with UVA (60 kJ/m^2) or UVB (120 J/m^2) in the presence and absence of NAC antioxidant (2.5 mM). (***), $P < 0.01$ (**), $P < 0.05$ (*). **b)** Representative pictures of cell clones after each experimental condition.

4.5 Supplementary Tables

Supplementary Table S4.1. Sequencing alignment statistics.

Cell line	**MRC5**							
Treatment	**NO UV**				**UVA (60 kJ/m^2)**			
Codigo clone	MRC5_C1	MRC5_C2	MRC5_C3	MRC5_C4	MRC5_UVA_C1	MRC5_UVA_C2	MRC5_UVA_C3	MRC5_UVA_C4
kit - platform	illumina-NextSeq	illumina-NextSeq	illumina-NextSeq	illumina-NextSeq	illumina-NextSeq	illumina-NextSeq	illumina-NextSeq	illumina-NextSeq
% duplicate reads	52.3%	50.0%	51.0%	54.2%	53.7%	52.5%	51.9%	51.7%
Total reads	77602486	52327456	52880810	70019448	107221956	71728384	69515168	109624946
Unique reads	37007872	26153082	25934943	32056142	49664551	34105793	33464038	53001236
Number of reads mapped	36329135	25712324	25487114	31544919	48949154	33546346	32932988	52128911
% reads mapped	98.2%	98.3%	98.3%	98.4%	98.6%	98.4%	98.4%	98.4%
Bases on target	4365624017	2928012149	3034260621	3885687758	6241607096	3903637248	3833865911	6378578176
% bases on target	69.4%	68.3%	70.3%	68.3%	69.5%	67.7%	68.2%	69.8%
Bases off target	2824702234	2087371081	1926549265	2671224468	4062919026	2785575582	2560035659	4128683891
% off target	30.6%	31.7%	29.7%	31.7%	30.5%	32.3%	31.8%	30.2%
% of target bp not covered	0.9%	1.0%	1.0%	1.0%	0.9%	0.9%	1.0%	0.9%
% of target bp covered at ≥ 1X	98.5%	98.2%	98.2%	98.4%	98.7%	98.5%	98.2%	98.7%
% of target bp covered at ≥ 10X	85.5%	74.2%	75.5%	82.4%	91.4%	82.9%	82.5%	92.6%
% of target bp covered at ≥ 20X	62.4%	42.3%	43.9%	55.3%	75.5%	56.4%	58.0%	77.9%
Average depth of coverage	118	79	82	105	168	105	103	172
Raw variants called	106461	87221	81614	102356	145464	105365	97114	147661
Novel SNPs	8457	5987	5647	8187	11571	8095	8821	11497
Percent SNPs (dbSNP)	92.1%	93.1%	93.1%	92.0%	92.0%	92.3%	90.9%	92.2%
Raw Indels called	15915	10709	10133	15099	21658	14838	15743	20497
Novel Indels	2298	1450	1468	2332	3231	2410	2772	2913
Percent Indels (dbSNP)	85.6%	86.5%	85.5%	84.6%	85.1%	83.8%	82.4%	85.8%

Continuation Supplementary Table 4.1. Sequencing alignment statistics.

Cell line	**MRC5**			
Treatment	**UVB (120 J/m^2)**			
Codigo clone	MRC5_UVB_C1	MRC5_UVB_C2	MRC5_UVB_C3	MRC5_UVB_C4
kit - platform	illumina-NextSeq	illumina-NextSeq	illumina-NextSeq	illumina-NextSeq
% duplicate reads	54.9%	54.9%	53.3%	51.9%
Total reads	50069158	114459824	75478190	110414168
Unique reads	22588940	51611328	35263134	53140417
Number of reads mapped	22266086	50826871	34638255	52323007
% reads mapped	98.6%	98.5%	98.2%	98.5%
Bases on target	2804971444	6391065882	4222625823	6490258933
% bases on target	70.1%	68.7%	69.8%	70.4%
Bases off target	1710440344	4074772204	2610841447	4094132268
% off target	29.9%	31.3%	30.2%	29.6%
% of target bp not covered	1.1%	0.9%	1.0%	0.9%
% of target bp covered at ≥ 1X	97.7%	98.5%	98.3%	98.7%
% of target bp covered at ≥ 10X	71.7%	90.2%	84.0%	92.5%
% of target bp covered at ≥ 20X	41.5%	76.5%	61.0%	77.9%
Average depth of coverage	76	172	114	175
Raw variants called	73798	140716	98005	149071
Novel SNPs	6529	13867	8458	11363
Percent SNPs (dbSNP)	91.2%	90.1%	91.4%	92.4%
Raw Indels called	11478	27736	15688	20906
Novel Indels	1997	4873	2723	2984
Percent Indels (dbSNP)	82.6%	82.4%	82.6%	85.7%

CHAPTER 5 – General discussion and conclusions

Since the discovery of UV irradiation as an environmental mutagen, numerous studies attempt to understand its effects by using different biological models (DeMarini et al. 2020). Each one with a particular focus but all with the intuition of trying to put the pieces together and understand the consequences of exposure to UV irradiation in human health as a whole. This work seeks to provide information about the mutagenic effects of the UV irradiation spectrum that reaches the Earth's surface, UVA and UVB. To get this, we took advantage from the NGS technology that allowed to access point mutations generated in the exome of human cells, and therefore revealed a more global approach of the mutagenic effects of UV irradiation in the human genome. Also, the use of cells deficient in the main repair pathway that recognizes and remove UV induced damages, the NER pathway, allows increasing the sensitivity of the model to detect the mutagenic effect of both UVA and UVB light wavelengths.

Epidemiologically, it is important to study UVA since it corresponds to 95% of the UV-sunlight spectrum and penetrates deep in the dermis, which increases the amount and cellular types that it reaches. Also for a long time it was considered harmless and little studied compared to UVC and UVB, so less information is available (Sage et al. 2012). Meanwhile, studying UVB is also of great importance because although proportionally UVB corresponds to only a small fraction of the UV spectrum, it is much more energetic and generates 10^3 times more direct damage than UVA (Kuluncsics et al. 1999, Cadet et al. 2005). Therefore, small increases in UVB levels that reaches the Earth´s surface due to stratospheric ozone depletion translates into large effects on different ecosystems and consequently in a wide range of life forms (de Gruijl & van der Leun 2000). The implementation of Montreal protocol has improved the ozone levels by controlling the production of diverse ozone depleting substances (Williamson et al. 2014). However, to reach pre-industrial levels of UVB it is still a challenge, as UV incidence is also affected by the Earth´s complex climate systems and by non-stratospheric factors (Williamson et al. 2014, Chipperfield et al. 2017). For those reasons, it is important to pay special attention to the consequences of both UVA and UVB exposure in human health, especially its effect on skin cancer (Bais et al. 2015).

According to the literature, XPC deficient cells are more sensitive to UVB and UVC irradiation than proficient cells (Feraudy et al. 2010, Dupuy et al. 2013, Andrade-

Lima 2015). Our data confirm this sensitivity to UVB light and reveal, for the first time, an elevated sensitivity also to UVA light, since XP-C cells showed a dose-dependent decrease on cell survival experiments, where higher doses were cytotoxic. In both cases, the XPC deficient cells were more sensitive than its isogenic control, COMP cells, obtained by complementation of the XP-C cells, in this work, with a lentiviral vector carrying the *XPC* gene.

Based on cell survival experiments, irradiation doses for mutagenesis were determined, 60 kJ/m^2 for UVA and 120 J/m^2 for UVB. These doses can be considered environmentally relevant, as they correspond approximately to twenty minutes and one minute and a half, respectively, of the sunlight exposure at midday during summer in a tropical latitude (Schuch et al. 2012). At these doses, UVA and UVB irradiation induces genotoxic stress in XP-C cells, as indicated by the increased levels of γH2AX after irradiation, while in COMP cells the increase recorded was statistically irrelevant. This relationship between phosphorylation of H2AX as a consequence of UVB and UVC irradiation has been previously reported (Limoli et al. 2002, Revet et al. 2011, Quinet et al. 2014, Andrade-Lima et al. 2015). High levels of sub-G1 content in UVA and UVB irradiated XP-C cells were also detected, which indicates an increase in apoptosis levels.

Our data showed that UVB-irradiation but not UVA-irradiation causes an S / G2 arrest in XP-C cells. Previously an accumulation in S-phase 24 h after UVB-irradiation had already been reported. And similar to our results, cells that do not die by apoptosis recover and continue cycling (Andrade-Lima et al. 2015). Additionally, it was reported an arrest at late S and/or G2 phases, depending on the UVC dose used (Quinet et al. 2014). These results suggest that although UVA generates sufficient damage to activate DDR it is not enough to affect cell cycle progression, at least not at the doses employed. An accumulation of single-stranded DNA (ssDNA) after UVC irradiation indicates that this response may be due to the persistence of 6-4PP lesions (Quinet et al. 2014). Also XP-V cells irradiated with double the dose of UVA used in this work suffer replication fork stalling and cell cycle arrest in the S-phase (Moreno et al. 2019). Then, the arrest in the cell cycle could be related to the amount of pyrimidine dimers caused by each type of UV radiation that were not repaired.

The deficiency in the XPC protein is responsible for the sun-sensitivity phenotype of the XP-C cells as demonstrated in the survival and cell cycle experiments. XP-C cells were more sensitive to UVA and UVB irradiation when

compared to control cells. This higher sensitivity of XP-C cells is certainly due to its incapacity to recognize and remove the UV- induced DNA damage in the global genome (not affecting TCR repair responsible for the repair in actively transcribed regions). Thus, the XP-C + UV light model allows to study directly the relationship between DNA damage, mutation and cancer, which could help to clarify the mechanisms involved in this process.

UVA and UVB light increased the mutation frequency (SNVs per million base pair sequenced) on XP-C cells from 1.4 to 5.2 and 8.1, respectively. However, there were no significant changes in the irradiated controls. Interestingly, the absence of the XPC protein also leads to an significant increase in the basal mutation frequency of the deficient cells compared to the complemented ones, from 0.9 to 1.4. This seems to be explained by a significant increase in the C>T transition, that maybe related to the spontaneous deamination of cytosine, one of the most common mutation in human cells. The presence of COSMIC´s signature 1 in all of our experimental groups supports this idea.

The data clearly demonstrate that for both types of UV light evaluated, the mainly induced mutations were the C>T transitions. Also, it was observed a significant increase in the CC>TT tandem mutations, considered as the UV light hallmark. By using the mutation spectrum and motif analysis, it was demonstrated that more than 95% of the C>T mutations induced by UVA or UVB occur preferentially in potential sites for the pyrimidine dimer formation, within the T<u>C</u>N and C<u>C</u>N sequence context and with a predilection for the non-transcribed strand, which is consistent with these cells being able to remove pyrimidine dimers by TC-NER. Interestingly, for cells with functional NER, this mutation was mainly found at CC dimers, without preference for mutation at the first or the second cytosine. In XP-C cells, without functional GG-NER, the pyrimidine 3' of the dimer was the most mutated and T<u>C</u> was the most mutagenic dimer. Then, the analysis of the sequence context where the C>T transition occurs exhibits an enrichment bias for pyrimidine-rich sequence context, revealing almost exactly the same logo for both the UVA and the UVB light, C(T/C)(T/C)C(Y)NC. The difference is in the +1 position, so for UVA is C followed by T, while for UVB is T followed by C, however in both cases the preference is for a pyrimidine base. Finally, our data point out that a unique dose of UVA- or UVB-irradiation recapitulate experimentally the typical mutation pattern of skin cancer, signature 7 of the COSMIC catalog, established by the analysis of more than 12.000 genomic data of 40 different

cancer types (Alexandrov et al. 2013b, 2020).

The C>A (G>T) transversion was the second most increased type of mutation in XP-C cells after both UVA- and UVB-irradiation. This base substitution is known as the mutagenic hallmark of the 8oxoG, generated by oxidation of guanines (Epe 1991, Cheng et al. 1992). However, after exploring these mutations within the RGR motif it was not possible to relate them with oxidized guanines. It is important to remark that the RGR motif was established in *E. coli* and maybe it is not the best one to explore mutations induced by oxidized guanine in eukaryotes. Thus, the induced mutations in G base should be analyzed more carefully, in fact according to the logo analyses it could be interesting to explore a motif that consider a purine either on the 3' or 5' of the oxidized guanine. Additionally, mutations in the other four types of base substitutions were also significantly increased by UVA- and UVB- irradiation, and could be consequences of base oxidation processes. Interestingly, this work indicates that the induction of C>G transversions is specific for UVA, while T>G transversions seems to be specific for UVB. Probably, this result is related to the oxidative damage induced by each type of light.

The results together, provide evidence that pyrimidine dimers are the main type of lesion contributing both to UVA and UVB induced mutagenesis in NER deficient cells and support the idea that both types of UV light generates the same mutational signature: the C>T transition at C-containing pyrimidine dimers, considered the UV signature since it is commonly found in skin cancer, but not in other types of cancers. Other types of mutations (including the more common C>A transversions) were also detected, probably due to lesions induced by oxidative stress. The data evidenced that not only UVB light but also UVA are highly mutagenic and in XPC patients this is exacerbated. Thus, this work discloses the UVA light participation in the high sunlight sensitivity and the elevated rate of skin carcinogenesis, in the complementation C group patients. The information generated in this work may be used for comparison with the mutational profiles of skin tumors obtained from XP patients, but also from the general population, since it has been suggested that mutations or loss of the XPC gene may be an early event during skin carcinogenesis (Feraudy et al. 2010). Thus, it is expected that these findings may help to understand the mutational processes of skin tumors in general. Finally, our results highlight the importance of photoprotection against solar UV radiation and other artificial sources of UVA radiation, since UVA is clearly not innocuous.

REFERENCES

ALEXANDROV, L.B. et al. Deciphering Signatures of Mutational Processes Operative in Human Cancer. **Cell Reports**, v. 3, p. 246–259, 2013a.

ALEXANDROV, L.B. et al. Signatures of mutational processes in human cancer. **Nature**, v. 500, p. 415-421, 2013b.

ALEXANDROV, L.B. et al. The Repertoire of Mutational Signatures in Human Cancer. **Nature**, v. 578, p. 94-101, 2020.

American Cancer Society. Global Cancer Facts & Figures 4th Edition. **Atlanta: American Cancer Society**, 2018.

ANDERSON, R.R. AND PARRISH, J.A. The Optics of Human Skin. **Journal of Investigative Dermatology**, v. 77, p.13–19, 1981.

ANDRADE-LIMA, L.C., ANDRADE, L.N., MENCK, C.F.M. ATR suppresses apoptosis after UVB irradiation by controlling both translesion synthesis and alternative tolerance pathways. **Journal of Cell Science**, v. 128, p.150–159, 2015.

ARMSTRONG, B.K., KRICKER, A. The epidemiology of UV induced skin cancer. **Journal of Photochemistry and Photobiology B: Biology**, v. 63, p. 8–18, 2001.

BAIS, A.F. et al. Ozone depletion and climate change: impacts on UV radiation. **Photochemical and Photobiological Sciences**, v. 14, p.19-52, 2015.

BAYATI, M. et al. CANCERSIGN: a user-friendly and robust tool for identification and classification of mutational signatures and patterns in cancer genomes. **Scientific reports nature research**, v. 10, p. 1-11, 2020.

BEN REKAYA, M. et al. High frequency of the V548A fs X572 XPC mutation in Tunisia: implication for molecular diagnosis. **Journal of human genetics**, v. 54, p. 426-429, 2009.

BOWDEN, N.A. et al. Understanding *xeroderma pigmentosum* complementation groups using gene expression profiling after UV-light exposure. **International Journal of Molecular Sciences**, v. 16, p.15985–15996, 2015.

BUKOWSKA, B., KARWOWSKI, B.T. Actual state of knowledge in the field of diseases related with defective nucleotide excision repair. **Life Sciences**, v. 195, p. 6-18, 2018.

BUNICK, C.G., MILLER, M.R., FULLER, B.E., FANNING, E., CHAZIN, W.J. Biochemical and Structural Domain-Analysis of XPC. **Biochemistry**, v. 45, p. 14965-14979, 2006.

CADET, J., SAGE, E., DOUKI, T. Ultraviolet radiation-mediated damage to cellular DNA. **Mutation Research**, v. 571, p. 3–17, 2005.

CHENG, K.C., CAHILL, D.S., KASAI H., NISHIMURA, S., LOEB, L.A. 8- Hydroxyguanine, an abundant form of oxidative DNA damage, causes G→T and A→C substitutions. **The Journal of biological chemistry**, v. 267, p.166–172, 1992.

CHIPPERFIELD, M. P. et al. Detecting recovery of the stratospheric ozone layer. **Nature**, v. 549, p. 211–218, 2017.

CLEAVER, J.E. Defective repair replication of DNA in xeroderma pigmentosum. **DNA repair**, v. 218, p.652–656, 1968.

CLEAVER, J. E., THOMPSON, L. H., RICHARDSON, A. S., STATES, J. C. A summary of mutations in the UV-sensitive disorders: xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy. Human Mutation, v. 14, p. 9-22, 1999.

CLEAVER, J.E., LAM, E.T., REVET, I. Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. **Nature reviews. Genetics**, v. 10, p.756–768, 2009.

CORTAT, B. et al. The relative roles of DNA damage induced by UVA irradiation in human cells. **Photochemical & Photobiological Sciences**, v. 12, p.1483-1495, 2013.

DAYA-GROSJEAN, L., JAMES, M.R., DROUGARD, C., SARASIN, A. An immortalized xeroderma pigmentosum, group C, cell line which replicates SV40 shuttle vectors. **Mutation Research DNA Repair Reports**, v. 183, p.185–196, 1987.

DeGRUIJL, F.R., VAN DER LEUN, J.C. Environment and health: 3. Ozone depletion and ultraviolet radiation. **Canadian Medical Association Journal**, v. 163, p. 851–5, 2000.

DeGRUIJL F.R. et al. Wavelength dependence of skin cancer induction by ultraviolet irradiation of albino hairless mice. **Cancer research**, v. 53, p. 53–60, 1993.

D'ERRICO M, et al. New functions of XPC in the protection of human skin cells from oxidative damage. **EMBO J**, v. 25, p. 4305–15, 2006.

DELINASIOS, G.J., KARBASCHI, M., COOKE, M.S., YOUNG, A. Vitamin E inhibits the UVAI induction of "light" and "dark" cyclobutane pyrimidine dimers, and oxidatively generated DNA damage, in keratinocytes. **Scientific reports**, v. 8, p. 423-435, 2018.

DeMARINI, D.M. The Mutagenesis Moonshot: The Propitious Beginnings of the Environmental Mutagenesis and Genomics Society. **Environmental and Molecular Mutagenesis**, V. 61, p. 8-24, 2020.

DE WEERD-KASTELEIN, E. A., KEIJZER, W., BOOTSMA, D. Genetic heterogeneity of xeroderma pigmentosum demonstrated by somatic cell hybridization. **Nature New Biology,** v. 238, p. 80–83, 1972.

DIGIOVANNA, J.J., KRAEMER, K.H. Shining a light on xeroderma pigmentosum. **Journal of Investigative Dermatology**, v. 132, p. 785–796, 2012.

DOUKI, T., REYNAUD-ANGELIN, A., CADET, J., SAGE, E. Bipyrimidine photoproducts rather than oxidative lesions are the main type of DNA damage involved in the genotoxic effect of solar UVA radiation, **Biochemistry**, v. 42, p. 9221–9226, 2003.

DUPUY, A. et al. Targeted gene therapy of xeroderma pigmentosum cells using meganuclease and TALEN™. ***PLoS ONE***, V.8, pp.1–8, 2013.

EL GHISSASSI, F. et al. A review of human carcinogens-Part D: radiation. **The Lancet Oncology**, v. 10, p. 751–752, 2009.

EPE, B. Genotoxicity of singlet oxygen. **Chemico-Biological Interactions**, v. 80, p.239– 260, 1991.

FERAUDY, S. et al. The DNA damage-binding protein XPC is a frequent target for inactivation in squamous cell carcinomas. **The American Journal of Pathology**, v. 177, p.555–562, 2010.

FONG, Y.W., INOUYE, C., YAMAGUCHI, T., CATTOGLIO, C., GRUBISIC, I., TJIAN, R. A DNA repair complex functions as an Oct4/Sox2 coactivator in embryonic stem cells. **Cell**, v. 147, p. 120-131, 2011.

FRIEDBERG, E.C., WALKER, G.C., SIEDE, W., WOOD, R.D., SCHULTZ, R.A., ELLENBERGER, T. DNA repair and Mutagenesis, 2nd Edition. 2nd. **ASM Press**, v. 2, p. 1118, 2006.

GHOSAL, G.; CHEN, J. DNA damage tolerance: a double-edged sword guarding the genome. **Translational Cancer Research**, v. 2, p. 107–129, 2013.

GIGLIA-MARI, G., ZOTTER, A., VERMEULEN, W. DNA Damage Response. **Cold Spring Harbor Perspectives in Biology**, v. 3, p. 1-19, 2011.

GILLET, L.C.J., SCHÄRER, O.D. Molecular Mechanisms of Mammalian Global Genome Nucleotide Excision Repair. **Chemical Reviews**, v.106, p. 253-276, 2006.

GREAVES, M. Evolutionary Determinants of Cancer. **Cancer Discovery**, v. 5, p. 806-820, 2015.

HOEIJMAKERS, J.H.J. DNA damage, aging, and cancer. **The New England Journal of Medicine**, v. 361, p. 1475–1485, 2009.

IKEHATA, H., ONO, T. The Mechanisms of UV Mutagenesis. **Journal of Radiation Research**, v. 52, p.115–125, 2011.

KARRAN, P., BREM, R. Protein oxidation, UVA and human DNA repair. **DNA Repair**, v. 44, p. 178–185, 2016.

KAWANISHI, S., HIRAKU, Y., OIKAWA, S. Mechanism of guanine-specific DNA damage by oxidative stress and its role in carcinogenesis and aging. **Mutation Research/Reviews in Mutation Research**, v. 488, p. 65-76, 2001.

KHAN, S.G. Reduced XPC DNA repair gene mRNA levels in clinically normal parents of xeroderma pigmentosum patients. **Carcinogenesis**, v. 27, p. 84-94, 2006.

KULUNCSICS, Z., PERDIZ, D., BRULAY, E., MUEL B., SAGE, E. Wavelength dependence of ultraviolet-induced dna damage distribution: involvement of direct or indirect mechanisms and possible artefacts. **Journal of Photochemistry and Photobiology B: Biology**, v. 49, p. 71–80, 1999.

LAAT, W.L., Jaspers, N.G.J., Hoeijmakers, J.H.J. Molecular mechanism of nucleotide excision repair. **Genes & Development**, v. 13, p. 768–785, 1999.

LAVAL, J., JURADO, J., SAPARBAEV, M., SIDORKINA, O. Antimutagenic role of base-excision repair enzymes upon free radical-induced DNA damage. **Mutation Research**, v. 402, p. 93–102, 1998.

LEGERSKI, R. J., PETERSON, C. Expression cloning of a human DNA repair gene involved in xeroderma pigmentosum group C. Nature, v. 359, p. 70-73, 1992.

LEGERSKI, R.J., LIU, P., LI, L., PETERSON, C.A., ZHAO, Y., LEACH, R.J., NAYLOR, S.L., SICILIANO, M.J. Assignment of xeroderma pigmentosum group C (XPC) gene to chromosome 3p25. **Genomics**, v. 21, p. 266-269, 1994.

LEITE, R.A. et al. Identification of XP complementation groups by recombinant adenovirus

carrying DNA repair genes. **Journal of Investigative Dermatology**, v. 129, p. 502-506, 2009.

LE MAY, N., MOTA-FERNANDES, D., VÉLEZ-CRUZ, R., ILTIS, I., BIARD, D., EGLY, J.-M. NER factors are recruited to active promoters and facilitate chromatin modification for transcription in the absence of exogenous genotoxic attack. **Molecular Cell**, v. 38, p. 54- 66, 2010.

LI, L., BALES, E.S., PETERSON, C.A., LEGERSKI, R. J. Characterization of molecular defects in xeroderma pigmentosum group C. **Nature Genetics**, v. 5, p. 413-417, 1993.

LIMOLI, C.L. et al. UV-induced replication arrest in the Xeroderma Pigmentosum variant leads to DNA double-strand breaks, gamma -H2AX formation, and Mre11 relocalization. **Proceedings of the National Academy of Sciences of the USA**, v. 99, p. 233–238, 2002.

LINDAHL, T. Instability and decay of the primary structure of DNA. **Nature**, v. 362, p. 709-715,1993.

MARIONNET, C., PIERRARD, C., GOLEBIEWSKI, C., BERNERD, F. Diversity of Biological Effects Induced by Longwave UVA Rays (UVA1) in Reconstructed Skin. **PLOS ONE**, v. 9, p. e105263, 2014.

MASUTANI, C., SUGASAWA, K., YANAGISAWA, J., SONOYAMA, T., UI, M., ENOMOTO, T., TAKIO, K., TANAKA, K., VAN DER SPEK, P. J., BOOTSMA, D. Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23. **The EMBO Journal**, v. 13, n. 8, p. 1831-1843, 1994.

MATSUMURA, Y., ANANTHASWAMY, H.N. Molecular mechanisms of photocarcinogenesis. **Frontiers in Bioscience**, v. 7, p. 765-783, 2002.

MATSUMURA, Y., ANANTHASWAMY, H.N. Toxic effects of ultraviolet radiation on the skin. **Toxicology and Applied Pharmacology**, v. 195, p. 298-308, 2004.

MENCK, C.F., MUNFORD, V. DNA repair diseases: what do they tell us about cancer and aging? **Genetics and Molecular Biology**, v. 37, p. 220–233, 2014.

MIN, J.H., PAVLETICH, N. P. Recognition of DNA damage by the Rad4 nucleotide excision repair protein. **Nature**, v. 449, p. 570-575, 2007.

MORENO, N. et al. The key role of UVA-light induced oxidative stress in human xeroderma Pigmentosum Variant cells. **Free Radical Biology and Medicine**, v. 131, p. 432-442, 2019.

MORI, M.P. et al. Lack of XPC leads to a shift between respiratory complexes I and II but sensitizes cells to mitochondrial stress. **Scientific reports**, v. 7, p. 1-15, 2017.

MORLEY, N., CURNOW, A., SALTER, L., CAMPBELL, S., GOULD, D. N-acetyl-L-cysteine prevents DNA damage induced by UVA, UVB and visible radiation in human fibroblasts. **Journal of Photochemistry and Photobiology B: Biology**, v. 72, p. 55–60, 2003.

MOURET, S., BAUDOUIN, C., CHARVERON, M., FAVIER, A., CADET, J., DOUKI, T. Cyclobutane pyrimidine dimers are predominant DNA lesions in whole human skin exposed to UVA radiation. **Proceedings of the National Academy of Sciences of the USA**, v. 103, p.13765–70, 2006.

MUNFORD, V., CASTRO, L.P. et al. A genetic cluster of patients with variant xeroderma pigmentosum with two different founder mutations. **British Journal of Dermatology**, v. 176, p. 1270-1279, 2017.

NG, J. M. Y., VERMEULEN, W., VAN DER HORST, G.T.J., BERGINK, S., SUGASAWA, K., VRIELING, H., HOEIJMAKERS, J. H. J. A novel regulation mechanism of DNA repair by damage-induced and RAD23-dependent stabilization of xeroderma pigmentosum group C protein. **Genes & Development**, v. 17, p. 1630-1645, 2003.

NICHOLS, J., KATIYAR, S.K. Skin photoprotection by natural polyphenols: anti- inflammatory, antioxidant and DNA repair mechanisms. **Archives of dermatological research**, v. 302, p. 71– 83, 2010.

NIK-ZAINAL, S. et al. Mutational processes molding the genomes of 21 breast cancers. **Cell**, v. 149, p. 979–993, 2012.

NISHI, R., OKUDA, Y., WATANABE, E., MORI, T., IWAI, S., MASUTANI, C., SUGASAWA, K., HANAOKA, F. Centrin 2 stimulates nucleotide excision repair by interacting with xeroderma pigmentosum group C protein. **Molecular Cell Biology**, v. 25, p. 5664-5674, 2005.

PELTOMÄKI, P. DNA mismatch repair and cancer. **Mutation Research/Reviews in Mutation Research**, v. 488, p. 77-85, 2001.

PETLJAK, M., ALEXANDROV, L.B. Understanding mutagenesis through delineation of mutational signatures in human cancer. **Carcinogenesis**, v. 37, p. 531–540, 2016.

PFEIFER, G.P., YOU Y.H., BESARATINIA, A. Mutations induced by ultraviolet light. **Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis**, v. 571, p.19–31, 2005.

QUINET, A. et al. Gap-filling and bypass at the replication fork are both active mechanisms for tolerance of low-dose ultraviolet-induced DNA damage in the human genome. **DNA Repair**, v. 14, p.27–38, 2014.

RASTOGI, R.P., RICHA, KUMAR, A., TYAGI, M.B., SINHA, P. Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. **Journal of nucleic acids**, v. 2010, p. 1-32, 2010.

RAVANAT, J.L., DOUKI, T., CADET, J. Direct and indirect effects of UV radiation on DNA and its components. **Journal of Photochemistry and Photobiology B: Biology**, v. 63, p. 88–102, 2001.

REVET, I. et al. Functional relevance of the histone gammaH2Ax in the response to DNA damaging agents. **Proceedings of the National Academy of Sciences of the USA**, v. 108, p. 8663–8667, 2011.

RIDLEY, A.J., WHITESIDE, J. R., MCMILLAN T. J. & ALLINSON, S.L. Cellular and sub-cellular responses to UVA in relation to carcinogenesis. **International Journal of Radiation Biology**, v. 85, p. 177-195, 2009.

ROCHETTE, P.J. et al. UVA-induced cyclobutane pyrimidine dimers form predominantly at thymine-thymine dipyrimidines and correlate with the mutation spectrum in rodent cells. **Nucleic Acids Research**, v. 31, p. 2786–2794, 2003.

RUBIN, A.F., GREEN, P. Mutation patterns in cancer genomes. **Proceedings of the National Academy of Sciences**, v. 106, p. 21766–21770, 2009.

RUNGER, T.M., FARAHVASH, B., HATVANI, Z., REES, A. Comparison of DNA damage responses following equimutagenic doses of UVA and UVB: A less effective cell cycle arrest with UVA may render UVA-induced pyrimidine dimers more mutagenic than UVB- induced

ones. **Photochemical & Photobiological Sciences**, v. 11, p. 207-215, 2012.

SAGE, E., GIRARD, P.M., FRANCESCONI, S. Unravelling UVA-induced mutagenesis. **Photochemical & Photobiological Sciences**, v. 11, p.74–80, 2012.

SANCAR, A. DNA Excision Repair. **Annual Review of Biochemistry**, v. 65, p. 43–81, 1996.

SANTIAGO, K.M. et al. Comprehensive germline mutation analysis and clinical profile in a large cohort of Brazilian xeroderma pigmentosum patients. **Journal of the European Academy of Dermatology and Venereology**, (*submitted*).

SCHUCH, A.P., GALHARDO, R.S., LIMA-BESA, K.M., SCHUCH, N.J., MENCK, C.F.M. Development of a DNA-dosimeter system for monitoring the effects of pulsed ultraviolet radiation. **Photochemical & Photobiological Sciences**, v. 8, p.111– 120, 2009.

SCHUCH, A.P. et al. DNA damage profiles induced by sunlight at different latitudes. **Environmental and Molecular Mutagenesis**, v. 53, p.198–206, 2012.

SCHUCH, A.P., MORENO, N.C., SCHUCH, N.J., MENCK, C., GARCIA, C.C.M. Sunlight damage to cellular DNA: focus on oxidatively generated lesions. **Free Radical Biology & Medicine**, v. 107, pp. 110–124, 2017.

SETLOW, R.B. The wavelengths in sunlight effective in producing skin cancer: a theoretical analysis. **Proceedings of the National Academy of Sciences of the USA**, v. 71, p. 3363–3366, 1974.

SHIMIZU, Y., IWAI, S., HANAOKA, F., SUGASAWA, K. xeroderma pigmentosum group C protein interacts physically and functionally with thymine DNA glycosylase. **The EMBO Journal**, v. 22, p. 164-173, 2003.

SOUFIR N. et al. A prevalent mutation with founder effect in xeroderma pigmentosum group C from North Africa. **Journal of Investigative Dermatology**, v. 130, p.1537- 1542, 2010.

STEENKEN, S., JOVANOVIC, S.V. How easily oxidizable is DNA? One-electron reduction potentials of adenosine and guanosine radicals in aqueous solution. **Journal of the American Chemical Society**, v. 119, p.617–618, 1997.

SUGASAWA, K., et al. Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. **Molecular Cell**, v. 2, p. 223-232, 1998.

SUGASAWA, K. Xeroderma pigmentosum genes: Functions inside and outside DNA repair. **Carcinogenesis**, v. 29, p. 455–465, 2008.

SUTHERLAND, J.C., GRIFFIN, K.P. Absorption spectrum of DNA for wavelengths greater than 300 nm. **Radiation research**, v. 86, p. 399-409, 1981.

TAYLOR, J.S. Unraveling the Molecular Pathway from Sunlight to Skin Cancer. **American Chemical Society**, v. 27, p.76–82, 1994.

TOMASETTI, C., MARCHIONNI, L., NOWAK, M.A., PARMIGIANI, G., VOGELSTEIN, B. Only three driver gene mutations are required for the development of lung and colorectal cancers. **Proceedings of the National Academy of Sciences**, v. 112, p. 118-123, 2015.

TOMKOVA, M., SCHUSTER-BÖCKLER, B. DNA Modifications: Naturally More Error Prone?. **Trends Genetics**, v. 34, p. 627–638, 2018.

VOGELSTEIN, B., KINZLER, K.W. Carcinogens leave fingerprints. **Nature**, v. 355, p. 209-2010, 1992.

WILLIAMSON, C.E., et al. Solar ultraviolet radiation in a changing climate. **Nature Climate Change**, v. 4, p. 434–441, 2014.

YAGURA, T., MAKITA, K., YAMAMOTO, H., MENCK, C.F.M., SCHUCH, A.P. Biological Sensors for solar ultraviolet radiation. **Sensors**, v. 11, p. 4277-4294, 2017.

ZIEGLER, A. et al. Mutation hotspots due to sunlight in the p53 gene of nonmelanoma skin cancers. **Proceedings of the National Academy of Sciences of the USA**, v. 90, p.4216–20, 1993.

5. Curricular summary

Congress awards

Crodowaldo Pavan prize for the best work in the Mutagenesis area. 65th Brazilian Congress of Genetics, GENÉTICA 2019, from september 17th to 20th in Hotel Monte Real Resort, Águas de Lindóia, São Paulo, Brasil. 2019.

Published articles

Quintero, N., Corradi, C., Moreno, N.C., Souza, T.A., Rocha, C.R.R., Menck, C.F.M. Mutagenicity profile induced by UVB light in human Xeroderma Pigmentosum group C cells. (*In prep*).

Quintero, N., Corradi, C., Moreno, N.C., Souza, T.A., Menck, C.F.M. UVA light induced mutagenesis in the exome of human nucleotide excision repair deficient cells. (*In prep*).

Fuentes-León, F., Oliveira, A.P., **Quintero, N.,** Munford, M., Kajitani, G.S., Brum, A.C., Schuch, A.P., Colepicolo, P., Sánchez-Lamar, A., Menck, C.F.M.M. DNA damage induced by late spring sunlight in Antarctica. (*Submited*)

Moreno, N.C., Souza, T.A., Corradi, C., Garcia, C.C.M., **Quintero, N.**, Castro, L.P., Munford, V., Lenne, S., Menck, C.F.M. (2020). Whole-exome sequencing reveals the impact of UVA light mutagenesis in xeroderma pigmentosum variant human cells. *Nucleic acid research*. 48(4), 1941-1953. doi:10.1093/nar/gkz1182.

Fuentes, J.L., García, A., **Quintero, N.**, Prada, C.A., Rey, N., Franco, D.A., Contreras, D.A., Córdoba, Y., Stashenko, E.E. (2017). The SOS chromotest applied for screening plant antigenotoxic agents against ultraviolet radiation. *Photochemical & Photobiological Sciences*. 16(9), 1424-1434. doi:10.1039/c7pp00024c.

Quintero, N., Córdoba, Y., Stashenko, E. E., & Fuentes, J. L. (2017). Antigenotoxic effect against ultraviolet radiation-induced DNA damage of the essential oils from Lippia species. *Photochemistry and Photobiology*, 93(4), 1063-1072. doi: 10.1111/php.12735.

Prada, C.A., Tessmer, E.T., **Quintero, N.**, Serment-Guerrero, J., & Fuentes, J.L. (2016). Survival and SOS response induction in ultraviolet B irradiated Escherichia coli cells with defective repair mechanisms. International Journal of Radiation Biology, 92(6), 321-8. doi: 10.3109/09553002.2016.1152412.

Quintero, N., Stashenko, E. E., & Fuentes, J. L. (2012). The influence of organic solvents on estimates of genotoxicity and antigenotoxicity in the SOS chromotest. *Genetics and Molecular Biology*, 35(2), 503–14. doi:10.1590/S1415-47572012000300018.