Anna Luiza Silva Almeida Vicente

Interação entre a exposição à radiação ultravioleta e o genoma e o epigenoma de melanomas cutâneos e acrais: fatores prognósticos condutores de câncer associados à histopatologia

Tese apresentada ao Programa de Pós-Graduação da Fundação Pio XII – Hospital de Câncer de Barretos para obtenção do Título de Doutora em Ciências da Saúde.
Área de Concentração: Oncologia

Orientador: Prof. Dr. Vinicius de Lima Vazquez
Coorientador: Prof. Dr. Akram Ghantous

Barretos, SP
2022

SUPORTE À PESQUISA POR AGÊNCIA DE FOMENTO

Este trabalho recebeu apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) através de bolsa individual de doutorado e bolsa estágio de pesquisa no exterior (processos número 2016/15941-3 e 2017/09612-0, respectivamente).

As opiniões, hipótese e conclusões ou recomendações expressas neste material são de responsabilidade dos autores e não necessariamente refletem a visão da FAPESP.

Esta tese foi elaborada e está apresentada de acordo com as normas da Pós-Graduação do Hospital de Câncer de Barretos – Fundação Pio XII, baseando-se no Regimento do Programa de Pós-Graduação em Oncologia e no Manual de Apresentação de Dissertações e Teses do Hospital de Câncer de Barretos. Os pesquisadores declaram ainda que este trabalho foi realizado em concordância com o Código de Boas Práticas Científicas (FAPESP), não havendo nada em seu conteúdo que possa ser considerado como plágio, fabricação ou falsificação de dados. As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade dos autores e não necessariamente refletem a visão da Fundação Pio XII – Hospital de Câncer de Barretos.

Os pesquisadores declaram não ter qualquer conflito de interesse relacionado a este estudo.

Dedico este trabalho aos meus pais e ao meu irmão, cujo amor me faz vivenciar a existência do infinito.

AGRADECIMENTOS

Eu sempre quis ser cientista, desde muito criança eu misturo coisas e observo o que acontece. Para mim sempre houve uma espécie de magia no poder de pensar coisas novas e investigar a interação entre elas. No final desses 4 anos de doutorado, tendo passado por dois anos de mestrado e mais dois anos de especialização, eu me sinto realizando um sonho de criança e nada mais puro do que alcançar aquilo que nós sonhamos com essa idade. O que eu não imaginava tão nova era que ao me aventurar e investigar o mundo lá fora, eu descobriria muito mais dentro de mim. E essa jornada foi tão intrigante, interessante e desafiadora, quanto a de pesquisar a influência dos raios de sol nos melanomas. Hoje, a única certeza que eu tenho é que a ciência se faz em muitas mãos, com muitos cérebros brilhantes e uma boa dose de generosidade com o próximo, misturada com resiliência. Somente assim, é possível experimentar uma jornada que a conclusão é um trabalho do qual nos orgulhamos e uma evolução pessoal que honra a nossa existência. Eu não poderia deixar de agradecer ninguém que me ajudou nesse caminho, portanto peço a sua compreensão enquanto leitor, de não se importar com a extensão destes agradecimentos.

Aos pacientes com melanoma, que gentilmente cederam o uso de suas informações genéticas e clínicas para o desenvolvimento deste trabalho. Todo o meu profundo respeito por, mesmo em um momento de grande fragilidade, vocês entenderem a importância da ciência e consentirem em contribuir com o desenvolvimento dela. Esse trabalho é por vocês e para vocês.

Ao Dr Vinicius de Lima Vazquez, meu orientador nesse trabalho e parceiro de tantos outros projetos acadêmicos, pela confiança em mim depositada há quase 9 anos atrás, de que eu me desenvolveria profissionalmente o suficiente para alcançar as suas expectativas. Todos os dias desde então e para sempre, não importa em que lugar do mundo eu esteja, eu vou lutar para honrar ter sido sua aluna, porque para mim você é um exemplo de líder, de médico, de pesquisador, mas, sobre todas as coisas, de ser humano ético, justo, respeitoso e bondoso. Você me inspira e tudo o que construímos juntos ao longo desses anos me orgulha muito. É o fim de um ciclo, mas eu espero que não o fim da nossa parceria. Minha gratidão é eterna.

Ao Dr Rui Manuel Reis, por todas as oportunidades profissionais que me foram dadas ao longo desses anos, que mudaram a minha vida, por confiar na minha capacidade, mesmo no meu início profissional, e por ser uma referência tão forte para mim de cientista qualificado, ético e íntegro. Sou grata pela oportunidade de ter aprendido tanto com você e eu admiro profundamente o seu legado na ciência.

To Dr Akram Ghantous, my advisor and mentor, for your constant support. I am grateful for your guidance and mentorship through all these years. From the very beginning you trusted on my potential even in those moments when I had many difficulties. You were committed to this project, committed to my learning and I am extremely grateful for the many opportunities you gave me so I could learn and grow as a scientist and as a person. I enjoyed a lot working with such a brilliant scientist like you, who always thinks outside the box. You are a leader who brings the best out of everyone always with respect and empathy. You are a huge inspiration for me.

Ao Dr André Carvalho e Dr Alberto Wainstein, por terem feito parte da minha banca de acompanhamento, dispensando tempo para contribuir com o meu desenvolvimento intelectual e acadêmico. Além disso, o senso crítico de vocês ao longo do andamento deste trabalho foi decisivo para a melhoria do mesmo. E eu sou muito grata por isso.

Ao Grupo de Pesquisa de Melanoma do Hospital de Câncer de Barretos, por ser um espaço ideal para crescer e aprender. É uma honra fazer parte deste grupo desde o início e me orgulho do que ele se tornou e do potencial que ele tem. **Drª Ana Carolina Laus**, obrigada pela paciência, toda a ajuda dispensada a mim ao longo desses anos e pelas contribuições com o meu desenvolvimento. **Drª Lidia Rebolho,** sou muito grata por todo o apoio que foram, muitas vezes, além deste trabalho. Você nunca hesitou em me ajudar e te agradeço enormemente por isso. **Drª Adriane Feijó,** obrigada pela colaboração neste trabalho e por todas as inúmeras conversas que sempre me ensinaram muito. **Msc Camila Crovador**, a sua postura profissional é admirável e eu sou grata por mais essa colaboração. Trabalhamos juntas há 9 anos e eu sou incapaz de dizer uma única vez em que você esteve abaixo da excelência. **Msc Bruna Sorroche,** obrigada pelas parcerias profissionais e por ser uma grande amiga. Eu te admiro enormemente na ciência e na vida, e agradeço o apoio durante todos esses anos. **Msc Raquel Desci**, seu comprometimento com o trabalho, sempre disposta a aprender, ao mesmo tempo que já chegou muito madura profissionalmente, me ensinam muito. Agradeço o apoio e a sua sempre disponibilidade em me ajudar.

À todas as pessoas que passaram pelo grupo ao longo desses anos, que me ensinaram muito e à todas as pessoas que constituem o NAP-melanoma, por toda a ajuda dispensada com carinho e competência.

To Epigenomics and Mechanism Branch of the International Agency for Research on Cancer, for the positive, friendly, and healthy work environment. **Dr Zdenko Herceg,** for hosting me in your lab where I was provided with many opportunities to learn. I really admire your journey in science. Your collaboration in this project was decisive for its success. **Elizabeth Page**, I am grateful for all your support through all these years and especially when I moved to France and needed your help many times. **Dr Alexei Novoloaca**, thank you for all the time you spent teaching me bioinformatics and also for your encouragement, which always kept me going. **Dr Zainab Awada**, I enjoyed working with such a curious and qualified scientist as you are. Thank you for the collaboration in this project and for the fun moments we spent together. **Vincent Cahais,** I have not enough words to say thank you for all contribution you provided to this work. I am extremely thankful for all your help during all these years. **Dr Felicia Chung and Aurelie Salle**, thank you both for being kind when I was a newcomer at IARC and we shared the same office. You contributed positively to my adaptation process, and I admire the scientists that both of you are. **Dr Andrea Halaburkova,** I am grateful for all the fun moments we shared during the time I was in France. Thank you for being a great coworker and for everything that I learned from you. **Dr Natália Spitz**, thank you for your collaboration in this project and for always being kind and supportive. I admire the scientist that you are.

Also, I would like to give a special thanks to **Dr Jiri Zavadil**, **Dr Michael Korenjak**, **Dr Rita Khoueiry** and **Dr Fazlur Talukdar** for having warmly welcomed me to IARC and for the opportunity to learn from you in several occasions. I really admire you all.

Ao Hospital de Câncer de Barretos, minha profunda gratidão à instituição por ter sido a minha primeira casa profissional e por isso ter se perdurado por tantos anos. Tenho profundo respeito pelo trabalho sério desenvolvido pela instituição há 60 anos e por tantas oportunidades que me foram dadas. Aproveito para agradecer todos os departamentos ligados a esse trabalho: Centro de Pesquisa em Oncologia Molecular, Biobanco, Instituto de Ensino e Pesquisa, Núcleo de Apoio ao Pesquisador e Departamento de Melanoma e Sarcoma.

À FAPESP e CAPES, pelo incentivo à ciência em nosso país, ainda mais nos tempos sombrios de ataques a mesma nos últimos anos. Sou muito grata pelo apoio e trabalho sério desenvolvimento por ambas as instituições.

À Liliana Silva Almeida Vicente, minha mãe, por ter sido o meu primeiro e mais forte exemplo de mulher. Eu agradeço por você ter me ensinado o que não te ensinaram, por ter me dado o que não te deram e por ter lutado por mim, para que eu pudesse ser, ter e conhecer tudo aquilo que a você não foi possível. O que você fez é poderoso, transformador e baseado em amor incondicional. Todos os dias da minha vida eu vou honrar isso e igualmente lutar por você. Eu devo absolutamente tudo a você. Eu te amo da forma mais genuína e pura que existe!

Ao Welton Vicente, meu pai, por não medir esforços para me apoiar e para eu realizar os meus sonhos. Eu sou grata pelo pai dedicado e amoroso que você é e sempre foi, bem como por você ter me ensinado valores tão essenciais que eu carregarei sempre comigo. Te orgulhar me realiza e eu te amo muito!

Ao Welliton Silva Almeida Vicente, meu irmão, por ser o laço mais forte de amor, cumplicidade e lealdade que eu já construí com alguém nessa vida. Eu sou grata por você ser o maior incentivador dos meus sonhos, o meu melhor amigo, meu confidente, meu ponto de equilíbrio, meu porto-seguro e, por ainda assim, você estar sempre comprometido em ser um irmão ainda melhor para mim. Eu admiro o seu coração, a sua bondade, a sua impressionante capacidade de se colocar no lugar das outras pessoas e, sem dúvidas, você é a maior inspiração que eu tenho na vida. Ao longo deste trabalho, quando tudo parecia impossível era você quem estava lá me dando forças para continuar e não desistir do meu sonho, então eu não teria chegado tão longe sem você. Eu te amo com todo o meu coração e com toda a minha alma.

Ao Guilherme Lino, meu cunhado, pelo apoio, incentivo, amor e por somar tanto desde o primeiro momento que você entrou para a nossa família. Obrigada por ser um grande amigo, por cuidar das pessoas que eu mais amo nessa vida com tanto carinho e por ser tão compreensivo. Eu te admiro e te amo.

À toda a minha família, especialmente a minha avó **Mirley**, por rezar, torcer por mim, me incentivar e por sempre me mandar mensagens que acalmam meu coração. Junto com o meu avô **Anagê** (*in memorian*) muitas vezes me deu a oportunidade de estudar e eu sou eternamente grata por isso, porque a educação transformou a minha vida. Agradeço especialmente também a minha avó Elmiria (*in memorian*), que com muito amor contribuiu para

que eu tenha tido uma infância feliz e saudável, e ao meu avô José (*in memorian*), por ter sido tão querido com a minha mãe, que o faz ser tão vivo na minha vida, mesmo sem eu sequer ter o conhecido.

Ao Leonardo Masquio, meu grande amigo, por trazer poesia e sensibilidade, em uma vida que só existia racionalidade. Depois da sua chegada eu nunca mais fui a mesma, porque você me inspira a sempre buscar pelo meu melhor. Eu admiro a amizade que a gente construiu baseada em tanta cumplicidade e respeito por quem cada um é. Além disso, é uma delícia poder dividir a minha vida com você e assistir a sua, tal como o melhor seriado já produzido pela Netflix. O seu apoio foi fundamental incontáveis vezes, principalmente nos momentos mais desafiadores desenvolvendo esse trabalho. E o amor que eu desenvolvi por você me transformou e me curou.

À Nathalia Campanella, minha grande amiga, pelo exemplo de mulher, profissional e mãe que você significa para mim. Obrigada pelo incentivo, apoio, conversas que me tiram da zona de conforto e que ao mesmo tempo me confortam, e por sempre acolher os meus sentimentos. Eu tenho sorte de ter a sua amizade, lealdade e poder dividir essa vida com você. Eu amo e admiro quem você é.

To Dr Jesús Rafael Rodríguez-Aguilera, for the friendship and for being a person who made me believe that genuine and kind hearts still exist. I am extremely grateful because you supported me during my worst moments when I lived in France, at the same time that you and I had memorable memories during our travels or even exploring the French culture. Also, I admire you as a scientist, for always being persistent, hard worker and for thinking outside the box.

À Tatiana Perecin, pela brilhante e dedicada profissional que você é, o que permite que você consiga ajudar as pessoas a alcançar o maior potencial delas. Eu sou muito grata pelo seu acompanhamento ao longo desses anos e nos momentos mais delicados que eu passei desenvolvendo esse trabalho.

À todos os meus amigos, especialmente a Ana Paula, Guilherme, Wellington, Danilo, Leticia, Fernanda, Danillo e Renan. Vocês tornam tudo possível. Agradeço especialmente a Renata, cujo apoio foi essencial, principalmente durante o período em que eu estive na França

À todas e todos os cientistas brilhantes que eu tive a honra de trabalhar junto ao longo de toda a minha trajetória acadêmica e a quem, certamente, eu devo quase tudo o que eu sei e sou.

Extremistas e conservadores mostram o que mais os assusta: uma menina curiosa segurando um livro.

adaptado de Malala Yousafzai

SUMÁRIO

LISTA DE FIGURAS

LISTA DE ABREVIATURAS

RNA	Ácido ribonucleico
DNA	Ácido desoxirribonucleico
RNAr	Ácido ribonucléico ribossômico
RNAm	Ácido ribonucleico mensageiro
RNAt	Ácido ribonucleico transportador
TFIID	Fator de transcrição IID
TBP	Subunidade de ligação ao TATA
TAF	Fator associado à subunidade de ligação ao TATA
TFIIH	Fator de transcrição IIH
CPSF	Fator de especificidade de clivagem e poliadenilação
IF	Fatores de iniciação
EF	Fatores de elongação
SAM	S-adenosil metionina
5mC	Citosina metilada
CpG	Dinucleotídeo citosina e guanina ligados por um grupo fosfato
DNMT3a	DNA metiltransferase 3 alfa
DNMT3b	DNA metiltransferase 3 beta
DNMT1	DNA metiltransferase 1
Bp	Pares de bases
Kb	kilobase
FGF8	Fator de crescimento de fibroblasto 8
MITF	Fator de transcrição associado à microftalmia
KIT	Kit proto-oncogene
ET3	Sinalização de endotelina-3
RUV	Radiação ultravioleta
DHI	Dopa,5,6,diidroxiindol
DHICA	5,6 diidroxiindol-2-ácido carboxílico
Tyrp 2-Dct	Dopacromo tautomerase
Tyrp 1	Tirosinase relacionada à proteína 1
ROS	Espécies reativas de oxigênio
BRAF	*V-Raf murine sarcoma viral oncogene homolog B*
TERT	*Telomerase reverse transcriptase*
NRAS	*Neuroblastoma RAS viral oncogene homolog*
CDKN2A	*Cyclin-dependent kinase inhibitor 2A*
PTEN	*Phosphatase and tensin homolog*
EUA	Estados Unidos da américa
DATASUS	Departamento de informática do sistema único de saúde
mm	Milímetro
PET	Tomografia por emissão de pósitrons
DHL	Desidrogenase lática
AJCC	*American joint committee on cancer*
IARC	Agência internacional de pesquisa em câncer
OMS	Organização mundial de saúde
Anvisa	Agência nacional de vigilância sanitária
cm	Centímetro

NGS	*Next generation sequencing*
TCGA	*The cancer genome atlas network*
ICGC	*International cancer genome consortium*
NF1	Neurofibromatose tipo 1
LoF	Perda de função
Mb	Mega base
MAPK	*Mitogen-activated protein kinase*
PI3K	*Phosphoinositol 3-kinase*
RTK	Tirosina quinase
FDA	*Food and Drug Administration*
IgC1	Imunoglobulina humana C1
CTLA-4	Linfócito citotóxico associado a antigeno-4
PD-1	Proteína de morte celular programada 1
HCB	Hospital de câncer de Barretos
REN	Via de reparo por excisão de nucleotídeos
STL	Síntese de translesão
EGFR	Receptor de fator de crescimento epidérmico

LISTA DE SÍMBOLOS

%	Porcentagem
β	Beta
>	Maior

RESUMO

A radiação ultravioleta (UV) está causalmente ligada ao desenvolvimento de melanoma cutâneo, mas os mecanismos epigenéticos, conhecidos como sensores moleculares de exposição, nunca foram caracterizados em amostras clínicas de melanoma. Nesse trabalho, integramos o perfil clínico, o epigenoma (metiloma do DNA), genoma e transcriptoma de 112 melanomas cutâneos de duas coortes com origens multiétnicas. Identificamos alterações moleculares relacionadas à UV em regiões regulatórias e vias imunológicas, com potencial multi-OMIC de serem condutores de câncer afetando a sobrevida dos pacientes. *TAPBP*, o principal gene, está criticamente envolvido na função imunológica e abrange vários locais de metilação alterados por exposição à UV, que foram validados por sequenciamento direcionado, oferecendo oportunidades com bom custo-benefício para futura aplicação clínica. A análise do metiloma do DNA também revela alterações não relacionadas à exposição à UV explicando as diferenças patológicas entre os melanomas cutâneos e 17 melanoma acrais. O mapeamento epigenômico não-supervisionado demonstrou que o melanoma cutâneo não exposto à UV se assemelha mais ao melanoma acral do que ao melanoma cutâneo exposto à UV, com o último apresentando melhor prognóstico do paciente do que as outras duas formas. Essas interações gene-ambiente revelam mecanismos translacional impactantes na melanomagênese.

Palavras-chave: melanoma, cutâneo, acral, radiação ultravioleta, patologia, metiloma do DNA, assinaturas de mutação, transcriptoma, sequenciamento de nova geração, genes condutores de câncer

ABSTRACT

Ultraviolet radiation (UV) is causally linked to cutaneous melanoma, yet the underlying epigenetic mechanisms, known as molecular sensors of exposure, have never been characterized in clinical biospecimen. Here, we integrate clinical and epigenome (DNA methylome), genome and transcriptome profiling of 112 cutaneous melanoma from two multi-ethnic cohorts. We identify UV-related alterations in regulatory regions and immunological pathways, with multi-OMICs cancer driver potential affecting patient survival. TAPBP, the top gene, is critically involved in immune function and encompasses several UV-altered methylation sites that were validated by targeted sequencing, providing cost-effective opportunities for clinical application. The DNA methylome also reveals non UV-related aberrations underlying pathological differences between the cutaneous and 17 acral melanomas. Unsupervised epigenomic mapping demonstrated that non UV-mutant cutaneous melanoma more closely resembles acral rather than UV-exposed cutaneous melanoma, with the latter showing better patient prognosis than the other two forms. These gene-environment interactions reveal translationally impactful mechanisms in melanomagenesis.

Key words: melanoma, cutaneous, acral, ultraviolet radiation, pathology, DNA methylome, mutation signatures, transcriptome, cross-OMICS, cancer driver genes

1. INTRODUÇÃO

1.1. Célula: A Unidade Fundamental

A célula é a unidade estrutural e funcional fundamental dos seres vivos, que se agrupam em tecidos, e estes, por sua vez, em órgãos, que juntos formam um organismo complexo (Figura 1). As células são divididas em dois grandes grupos, procariontes e eucariontes. A principal diferença entre ambos os grupos é que as células procariontes não possuem envoltório nuclear, nesses, os cromossomos se localizam no protoplasma, juntamente com os ribossomos, água, íons, outros tipos de ácidos ribonucleicos (RNA), proteínas estruturais e enzimáticas, diversas moléculas pequenas, entre outras estruturas[1].

Todos os organismos complexos, como por exemplo, os seres humanos, outras espécies de animais e plantas, são formados pela junção de células eucarióticas, que por sua vez se diferem de acordo com as funções específicas dos diferentes tecidos. Embora essa especialização possibilite que as células possuam características singulares de acordo com a sua localização no organismo, ainda assim, elas compartilham o mesmo modelo de organização[1]:

- Membrana Plasmática: trata-se de uma bicamada lipídica contínua e proteínas intercaladas ou aderidas a sua superfície, que delimita o conteúdo da célula com o ambiente externo. Além disso, essa estrutura controla a passagem de substâncias do exterior, no processo de endocitose e de exocitose, e possui receptores que interagem especificamente com moléculas provenientes do exterior, como hormônios e neurotransmissores, dentre outras atividades[2].
- Citoplasma: é a parte da célula que não corresponde ao núcleo e pode ser divida em dois espaços, sendo o primeiro correspondente ao citosol e outro que se encontra no interior das organelas. O citosol representa, em média,

cerca de 50% do volume do citoplasma, sendo considerado o verdadeiro meio intracelular[1].

- Citoesqueleto: é uma complexa rede de filamentos proteicos que estão espalhados por todo o citosol, incluindo os filamentos de actina, os filamentos intermediários, os filamentos de miosina e os microtúbulos. Essa rede dá forma às células, como resultado da interação dos três tipos de filamentos com diferentes proteínas acessórias[1, 2].
- Organelas: são diversas as estruturas localizadas no citosol, cada qual responsável por uma ação, que mantêm a funcionamento celular normal. Dentre elas incluem as mitocôndrias, responsáveis pela síntese de energia; os ribossomos, local onde a síntese proteica é realizada; os lisossomos, responsáveis pela digestão de macromoléculas e micro-organismos; complexo golgiense, onde se realiza o empacotamento e a distribuição das macromoléculas para a secreção, para a membrana plasmática ou para outras organelas; retículo endoplasmático, constituído pelo rugoso e liso, onde ocorrem, respectivamente a síntese de proteínas e a síntese de lipídios e biogênese de novas membranas celulares; os peroxissomos, que possuem enzimas da β-oxidação dos ácidos graxos e enzimas que oxidam aminoácidos; e proteassomos, complexo de proteases que digerem as proteínas marcadas com ubiquitina[1, 2].
- Envoltório nuclear: é composto por duas membranas concêntricas que delimitam o núcleo e se unem ao nível dos poros, que se acham distribuídos de certa forma regularmente por todo o envoltório. Esses poros permitem a comunicação entre núcleo e citoplasma[1].
- Núcleo: ocupa cerca de 10% das células eucarióticas e nele se encontra o ácido desoxirribonucléico (DNA), exceto os da mitocôndria. Neste compartimento nuclear encontram-se 46 cromossomos, o nucléolo, onde estão localizados os genes do RNAr (ácido ribonucléico ribossômico) e o RNAr recém-sintetizado,

bem como diversas proteínas que regulam as atividades dos genes, processamentos dos RNAs, dentre outras atividades[2].

- Ácidos nucléicos: são divididos em dois tipos, o DNA, que constitui todo o depósito da informação genética e que é o foco central deste trabalho; e o RNA, que é dividido entre mensageiro, ribossômico e transportador, todos envolvidos na síntese proteica. O RNA mensageiro (RNAm) compreende a transcrição da informação do DNA, que estabelece a sequência dos aminoácidos da proteína, enquanto o RNAr representa cerca de 50% da massa do ribossomo, que como dito anteriormente, é a organela onde ocorre a síntese proteica; e o RNA transportador (RNAt), responsável pelo transporte dos aminoácidos até os ribossomos. Mais recentemente se descobriu a existência de RNAs não-codificantes, como os RNAs de interferência, micro e longo-RNAs, ligados, em última instância, a regulação da expressão gênica[3].

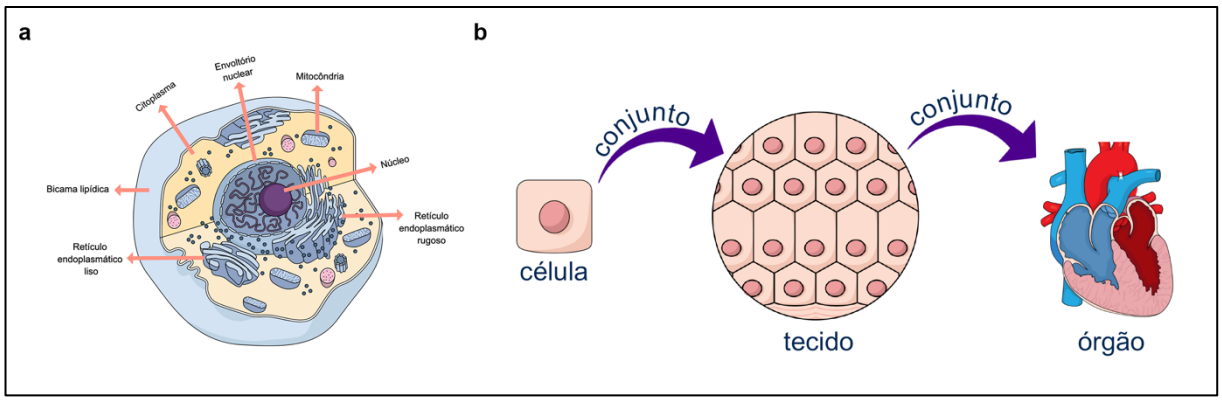


Figura 1 - Célula, a unidade fundamental da vida. A- Descrição dos componentes celular. B- Diferentes níveis de organização celular. Fonte: Autoria própria.

1.2. Ácidos Nucleicos, Expressão e Regulação Gênica

A molécula de DNA e RNA são polímeros cujos monômeros são nucleotídeos sucessivamente ligados por meio de ligações fosfo-diéster (Figura 2). Conceitualmente, o genoma é o conjunto total da sequência de nucleotídeos que compõe o DNA de um ser vivo, enquanto o epigenoma é o que regula se a informação contida na sequência de DNA

será ou não transformada em proteína. O epigenoma inclui metilação, modificação de histonas e RNAs não-codificantes[1].

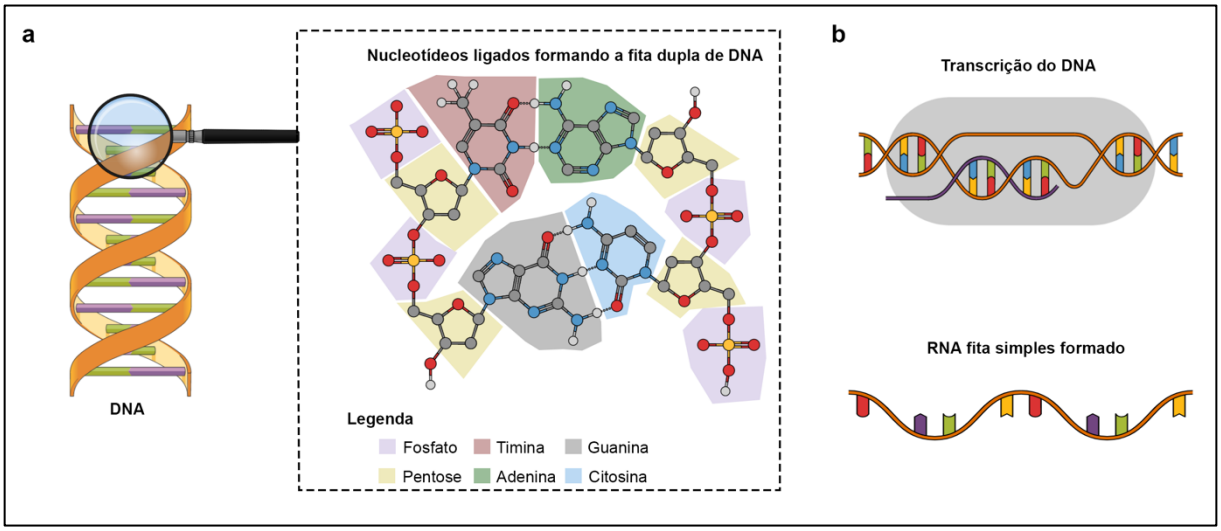


Figura 2 - Estrutura do DNA e RNA. A- Ligação entre nucleotídeos para formar a dupla fita de DNA. B- Síntese da fita de RNA a partir do DNA. Fonte: Autoria própria.

Cada nucleotídeo tem os três componentes seguintes (Figura 2A): 1- uma pentose, um açúcar cíclico de cinco átomos de carbono, que pode ser desoxirribose no DNA e ribose no RNA, e que se diferem pela ausência de um átomo de oxigênio na desoxirribose; 2- uma base nitrogenada, que é dividida em dois grupos, as pirimidinas e as purinas, sendo que as pirimidinas possuem um anel heterocíclico, enquanto as purinas têm dois anéis fundidos entre si. No DNA, as pirimidinas são a timina (T) e a citosina (C), e as purinas, a adenina (A) e a guanina (G). O RNA contém a uracila (U) no lugar de timina. Além das duas diferenças já citadas entre DNA e RNA, existe uma terceira, que se consiste no fato do DNA possuir duas cadeias polinucleotídicas, enquanto o RNA apenas uma; 3- um fosfato ligado ao carbono 5’ do açúcar por uma ligação fosfodiéster[4].

A molécula de DNA é formada por duas cadeias polinucleotídicas helicoidais com uma rotação para a direita, que compõem uma dupla hélice em torno de um mesmo eixo central (Figura 2A). As duas cadeias são antiparalelas, o que significa que as suas ligações

3´,5´-fosfodiéster seguem sentidos opostos. Ambas as cadeias estão unidas entre sim por pontes de hidrogênio estabelecido entre os pares de base[1].

O pareamento de bases é uma das características mais importantes da estrutura do DNA. Tendo em vista que entre as pentoses das cadeias opostas existe uma distância fixa, apenas certos pares de bases podem se estabelecer dentro da estrutura. Sendo assim, os únicos pares possíveis são A-T/T-A, formando-se duas pontes de hidrogênio entre as respectivas bases, e C-G/G-C, cuja interação é mais estável, existindo três pontes de hidrogênio entre as respectivas bases[4] (Figura 2A). O RNA tem muitas das características estruturais do DNA, sendo observado apenas três diferenças, que foram relatadas acima (Figura 2B).

Do ponto de vista molecular, a definição clássica de gene é uma sequência de DNA que contém as informações necessárias para produzir uma molécula de RNA e, se esta corresponder a um RNA mensageiro, a partir dele elaborar uma proteína[1]. Estruturalmente o gene possui 4 segmentos clássicos[1:]

- Promotor: região que inicia a transcrição e sinaliza a partir de qual nucleotídeo o gene deve ser transcrito. A combinação mais comum inclui sequências chamadas TATA e CAAT, situadas perto do codificador;
- Sequências reguladoras: determinam quando deve ocorrer a transcrição do gene e quantas vezes o mesmo deve ser transcrito;
- Região codificadora: Alterna-se entre éxons, que são fragmentos de DNA que de fato possuem informação para a transcrição em RNAm e, posteriormente a tradução em proteína, e os íntrons, sequências de DNA que não são transcritos em RNAm, mas que atualmente se sabe que são transcritos em RNAs não-codificantes, que possuem função regulatória;
- Sequência de terminação: se localiza na extremidade 3´ da região codificadora, geralmente sendo uma sequência AATAAA, importante para a conclusão da síntese do transcrito primário.

A transcrição do DNA (Figura 2B) ocorre na direção 3’→ 5’, para que o RNA seja sintetizado a partir da extremidade 5’ e progrida para a sua extremidade 3’. De forma simplificada, a transcrição do DNA em RNAm se inicia quando as sequências reguladoras e o promotor são ativados por proteínas especiais chamadas de fatores de transcrição, que se dividem em basais e específicos. Assim, quando o fator de transcrição IID (TFIID), um fator de transcrição basal, se une ao promotor por meio da sua subunidade de ligação ao TATA (TBP), alterando a estrutura da cromatina, essa mudança de conformação atrai tanto os fatores de transcrição basais restantes como a RNA polimerase II. Ao mesmo tempo, os fatores de transcrição específicos se ligam as sequências reguladoras, permitindo que o gene se curve em forma de uma forquilha, uma vez que os fatores de transcrição específicos unidos a região regulatória interagem com os fatores basais situados no promotor por meio da subunidade TAF (fator associado à TBP) do fator TFIID[1].

Uma vez unida ao promotor, a RNA polimerase II é fosforilada pelo TFIIH (fator de transcrição IIH), outro fator de transcrição basal, então um ATP (adenosina trifosfato) doa um fósforo, depois de ser hidrolisado pelo TFIIB, sendo este também um fator de transcrição basal. Em seguida, a RNA polimerase II fosforilada desprende-se dos fatores de transcrição e abre a dupla hélice do DNA no setor do gene adjacente ao promotor, formando a bolha de transcrição, iniciando a síntese de RNAm. Para alongar o RNAm, a RNA polimerase II necessita de dois fatores adicionais: os fatores de alongamento SII e SIII1. A transcrição é concluída quando a RNA polimerase alcança a sequência de terminação na extremidade 3’ do gene[1].

O processamento do RNAm compreende a remoção dos íntrons e a agregação de duas estruturas, chamadas cap e poli A, sendo que a primeira é adicionada na extremidade 5’ e a segunda na extremidade 3’ do RNAm. Essas modificações são importantes para que o RNAm possa sair do núcleo e ser traduzido no citosol. Em primeiro lugar, uma guanosina trifosfato é adicionada a extremidade 5’ do transcrito e a ligação dessa com o primeiro nucleotídeo do RNAm se dá por meio de uma ligação trifosfato. Em seguida, a metiltransferase adiciona um radical metil à guanina do cap, transformando-a em 7-metilguanosina, e adiciona um outro radical metil ao que passou a ser o segundo

nucleotídeo do RNAm. A adição de uma sequência de aproximadamente 250 adeninas, chamada poli A, na extremidade 3’ do RNAm pela enzima poli A polimerase é chamada poliadenilação. Para isso, é necessária a presença do *CPSF* (fator de especificidade de clivagem e poliadenilação) e de outro fator, o PABII[1].

A partir do momento que o RNAm está processado ele sai do núcleo e se dirige ao citoplasma para que ocorra a síntese proteica no ribossomo. A etapa de iniciação da síntese proteica é regulada por proteínas citosólicas denominadas fatores de iniciação (IF). O primeiro evento ocorre quando o fator IF-4 se liga a sequência localizada entre o cap e o códon de iniciação, em seguida o fator IF-3, com a ajuda do IF-4, coloca a extremidade 5’ do RNAm sobre a face da subunidade menor do ribossomo que tem os sítios E, P e A. Assim, a subunidade menor consegue deslizar pelo RNAm e detectar o códon AUG de iniciação, que se coloca no sítio P, fazendo com que o próximo códon do RNAm se aloque no sítio A. A etapa de iniciação termina quando a subunidade maior se une à subunidade menor e o ribossomo é formado, sendo este processo mediado pelo fator IF-5[1].

A etapa de alongamento da síntese proteica é regulada por fatores de elongação (EF), iniciando quando o RNAt cujo anticódon é complementar ao segundo códon do RNAm, ingressa no ribossomo, o que é mediado pelo fator EF-1. O RNAt do sítio P e o RNAt do sítio A ficam um do lado do outro, igualmente aos seus aminoácidos. Essa aproximação é importante para que ambos os aminoácidos possam se ligar entre si por meio de uma ligação peptídica. O ribossomo avança três bases ao longo do mRNA no sentido 5'→3'. Este passo é mediado pelo fator de elongação EF-G. Assim, há a liberação do RNAt presente no sítio P e que não se encontra mais carregado[1].

A etapa de terminação da síntese proteica é regulada por fatores de terminação e ocorre quando o códon de terminação do RNAm (UAA, UGA ou UAG) chega ao sítio A do ribossomo. Esse sítio é ocupado pelo fator eRF-1, que é capaz de reconhecer os três códons de terminação, uma vez que nenhum RNAt é capaz de se acoplar nessa região. Em virtude da ausência de um novo RNAt, o polipeptídeo se desliga do último RNAt e se torna

independente do RNAm e do ribossomo. Esse desprendimento é dependente do fator eRF-3. De imediato, as subunidades maior e menor do ribossomo se separam do RNAm[1].

A regulação da expressão gênica pode ser controlada por diversos mecanismos, incluindo a metilação. Conceitualmente, esse termo denota a adição de um radical metil de um S-adenosil metionina (SAM) para o quinto carbono de uma citosina (5mC), que é precedida por uma guanina (CpG), sendo essa reação catalisada pela família das DNA-metiltransferases. DNMT3a (DNA metiltransferase 3 alfa) e DNMT3b (DNA metiltransferase 3 beta) podem estabelecer um padrão de metilação *de novo*, enquanto que a DNMT1 (DNA metiltransferase 1) é responsável pela manutenção deste padrão nas fitas filhas de DNA durante a replicação. Essa marca epigenética possui um papel fundamental em outros mecanismos, incluindo embriogênese, inativação do cromossomo X e diferenciação celular[5] (Figura 3).

O mecanismo pelo qual a metilação do DNA regula a expressão gênica é restringir a acessibilidade dos fatores de transcrição (discutido anteriormente) às regiões importantes no processo de transcrição do DNA, tais como a região promotora e as sequências regulatórias. Paradoxalmente, a metilação do DNA é associada a ativação do gene, quando esta marca epigenética ocorre no corpo do gene, em regiões que são transcritas[6].

Os CpGs podem estar distribuídos ao longo do genoma de forma isoladas ou podem acontecer em ilhas CpGs, que são definidas como regiões no genoma com mais de 200 bp (pares de bases), onde os dinucleotídeos GC constituem mais de 50% dessas bases e uma razão entre o esperado e observado de CpG é maior do que 60%. Além disso, atualmente se classifica os CpGs ao longo do genoma de acordo com a distância em que eles se encontram das ilhas CpGs, sendo *shore*, *shelf* e *open sea*, se estiverem localizadas abaixo (*dowstream*) ou acima (*upstream*) das ilhas CpGs 2 Kb (*kilobase*), 4 Kb e mais do que 4 Kb, respectivamente[6] (Figura 4).

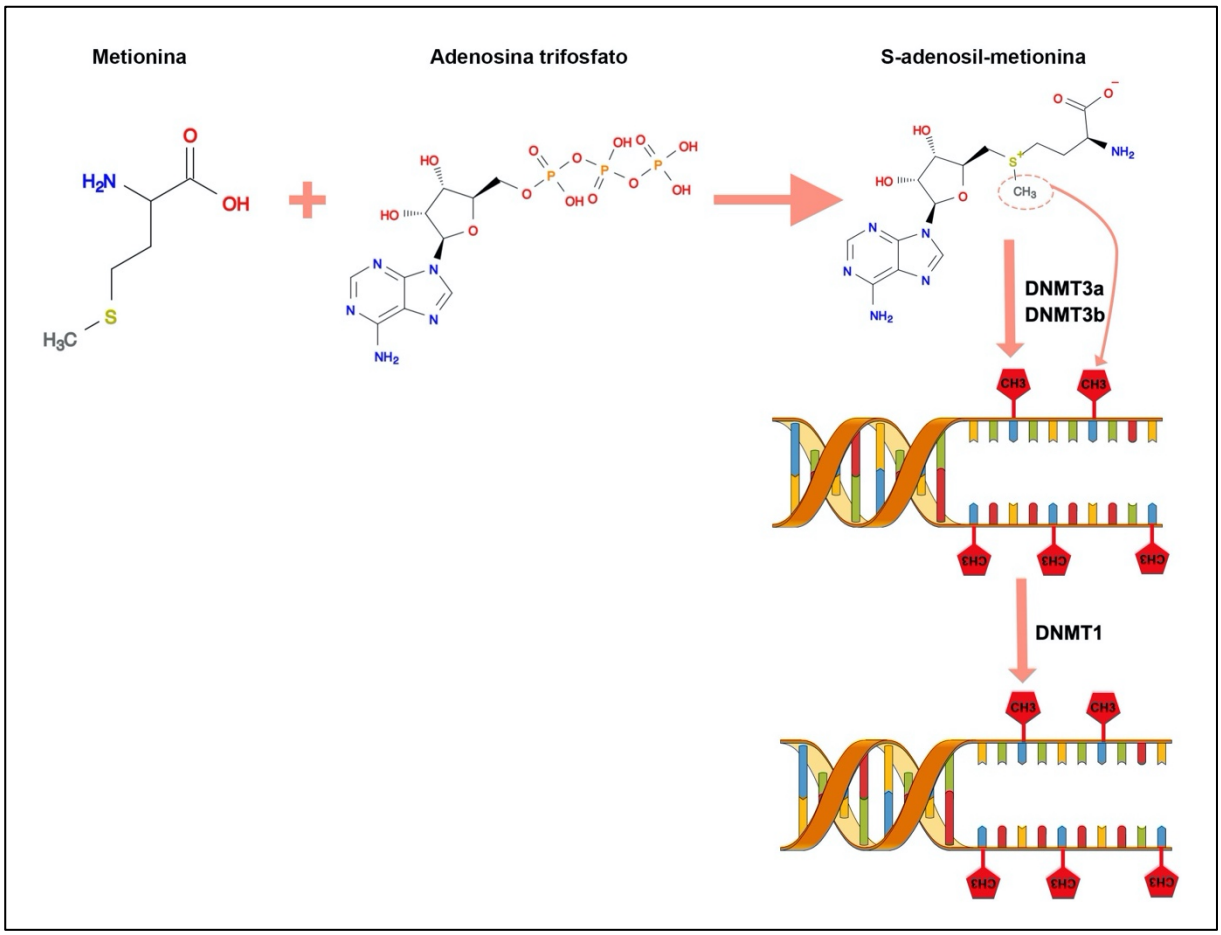


Figura 3 - Processo de doação do radical metil da molécula S-adenosil-metionina para a fita de DNA, processo conhecido como metilação. Fonte: Autoria própria.

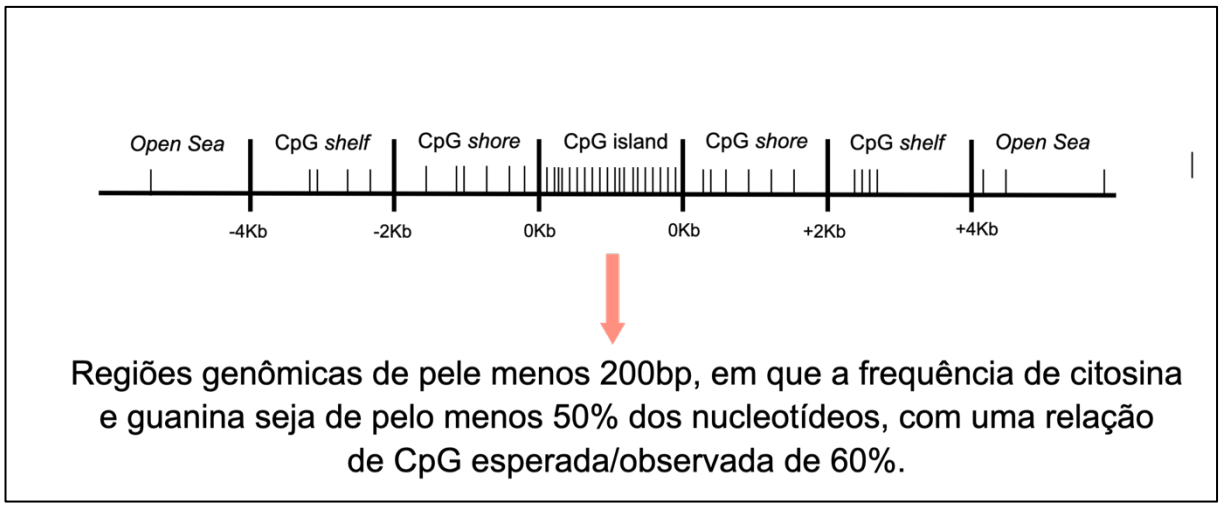


Figura 4 – Definição da distribuição dos CpGs ao longo do genoma. Fonte: Autoria própria.

1.3. A Origem: Melanócitos

Durante a embriogênese, por volta da terceira semana de gestação, ocorre a gastrulação, o processo que estabelece os três folhetos germinativos, sendo eles ectoderme, mesoderme e endoderme A gastrulação se inicia quando as células do epiblasto se proliferam e migram para o plano mediano do disco embrionário, formando a linha primitiva. Em seguida, as células do epiblasto migram em direção à linha primitiva, se diferenciam, se desprendem do epiblasto e deslizam por baixo dele. Esse movimento para dentro é conhecido como invaginação, resultando na formação do sulco primitivo e fosseta primitiva[7] (Figura 5).

A migração celular e a diferenciação são controladas pelo fator de crescimento de fibroblastos 8 (FGF8), que é sintetizado pelas próprias células. Este fator de crescimento controla o movimento da célula pela regulação negativa da E-caderina, uma proteína que normalmente une as células do epiblasto. FGF8 então controla a diferenciação da célula no mesoderma regulando a expressão de Brachyury. Uma vez que as células invaginaram, algumas deslocam o hipoblasto, criando a endoderme embrionária, e outras ficam entre o epiblasto e endoderme recém-criada para formar a mesoderme. As células que permanecem no epiblasto formam a ectoderme[7].

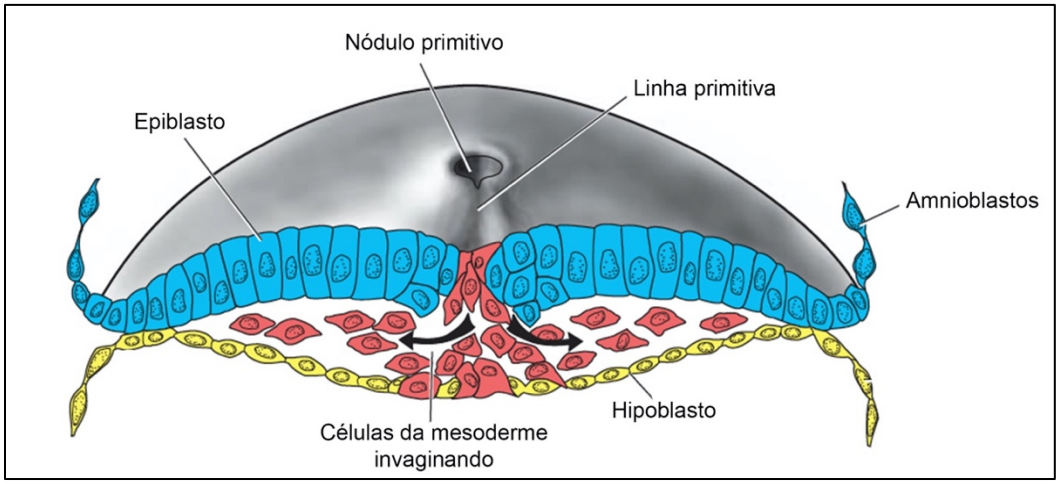


Figura 5- Formação dos três folhetos embrionários: endoderme, mesoderme e ectoderme. Fonte: Adaptado de Sadler e Langman[7].

O período embrionário seguinte é conhecido como organogênese, ocorrendo entre a terceira e oitava semana de gestação, e é definida como o período em que as três camadas germinativas dão origem aos órgãos. No início da terceira semana de desenvolvimento, a camada germinativa ectodérmica tem a forma de um disco que é mais largo na região cefálica do que na região caudal. O aparecimento da notocorda e do mesoderma pré-cordal induz o ectoderma sobrejacente a engrossar e formar a placa neural. As células da placa formam o neuroectoderma, e sua indução representa o evento inicial no processo de neurulação, mecanismo pelo qual a placa neural forma o tubo neural, que posteriormente se diferenciará em encéfalo e medula espinhal[7].

Até o final da terceira semana, as bordas laterais da placa neural se elevam para formar as pregas neurais, e a região média deprimida forma o sulco neural. Gradualmente, as pregas neurais se aproximam na linha média, onde se fundem. A fusão começa na região cervical (quinto somito) e prossegue no sentido cranial e caudal. Como resultado, o tubo neural é formado. Até que a fusão esteja completa, as extremidades cefálica e caudal do tubo neural se comunicam com a cavidade amniótica por meio dos neuróporos anterior e posterior, respectivamente. O fechamento do neuróporo craniano ocorre aproximadamente no dia 25 (estágio de 18 a 20 somitos), enquanto o neuróporo posterior se fecha no dia 28 (estágio de 25 somitos). A neurulação é então completa e o sistema nervoso central é representado por uma estrutura tubular fechada com uma porção caudal estreita, a medula espinhal, e uma porção cefálica muito mais ampla caracterizada por várias dilatações, as vesículas cerebrais[8] (Figura 6).

À medida que as pregas neurais se elevam e se fundem, as células da borda lateral ou crista do neuroectoderma começam a se dissociar de suas vizinhas. Essa população de células, a crista neural, sofrerá uma transição epitélio-mesenquimal à medida que deixa o neuroectoderma por migração ativa e deslocamento para entrar no mesoderma[7]. As células da crista da região do tronco deixam o neuroectoderma após o fechamento do tubo neural e migram ao longo de uma das duas vias: 1- uma via dorsolateral através da derme, onde entrarão na ectoderme através de orifícios na lâmina basal para formar melanócitos na pele e folículos pilosos; e 2- uma via ventral através da metade anterior

de cada somito para se tornarem gânglios sensoriais, neurônios simpáticos e entéricos, células de Schwann e células da medula adrenal. As células da crista neural são tão fundamentalmente importantes e contribuem para tantos órgãos e tecidos que às vezes são chamadas de quarta camada germinativa[9].

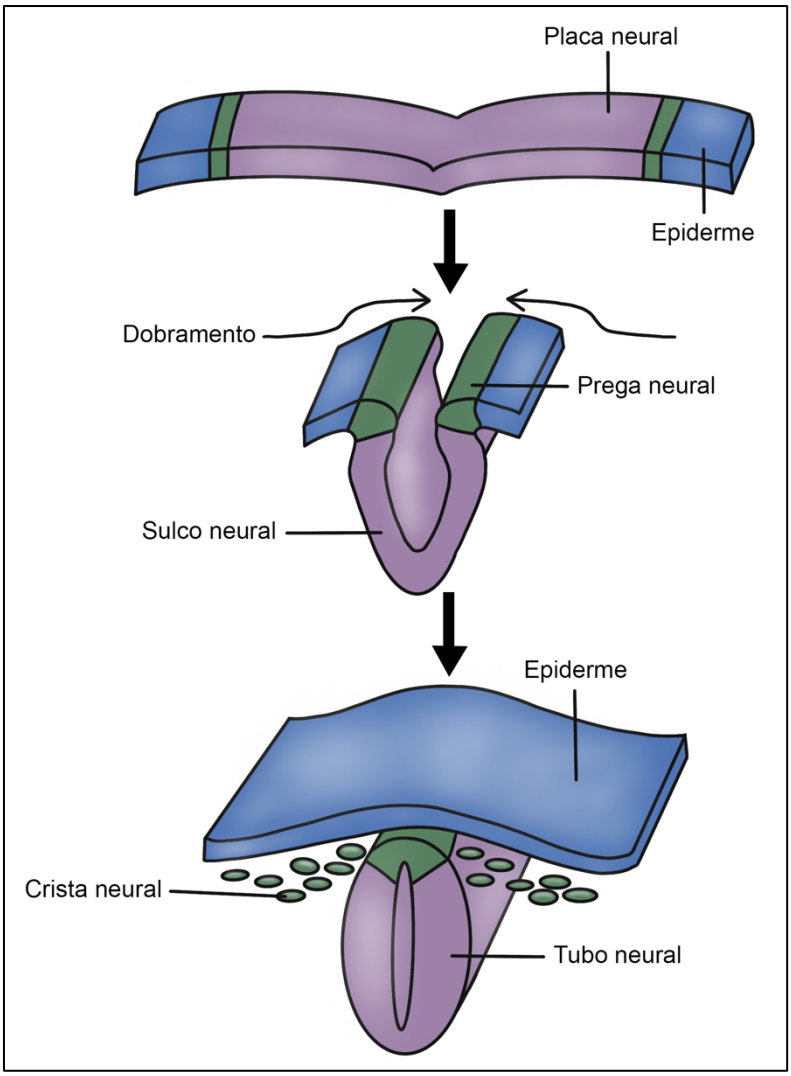


Figura 6- Processo de neurulação durante a embriogênese. Fonte: Autoria própria.

Um resumo dos principais eventos e marcadores moleculares chaves durante a diferenciação dos melanoblastos embrionários é mostrado na Figura 7A[10]. Nesta

dissertação os embriões foram classificados em dias de gestação identificados pela letra "E".

Os primeiros marcadores de melanoblastos aparecem entre E10.5 e E11.5. O primeiro gene melanócito-específico que é expresso em melanoblastos é o fator de transcrição associado à microftalmia (*MITF*), considerado como o principal gene regulador da linhagem de melanócitos. Isso porque esse gene é crucial para muitos processos celulares, incluindo sobrevivência, progressão do ciclo celular, proliferação e diferenciação de melanócitos, bem como regula a expressão dos genes necessários para a síntese do pigmento de melanina[11].

MITF é regulado por muitos genes envolvidos no desenvolvimento da crista neural e melanoblastos, como *PAX3*, *SOX10*, *FOXD3* e *CREB*. Inicialmente, os melanoblastos migram e proliferam na derme. Por volta de E12.5, quase 50% de todos os melanoblastos passaram pela membrana basal, a barreira física que separa a derme da epiderme. Enquanto os melanoblastos dentro da derme continuam a se dividir, o número total é mantido inalterável quando um fluxo constante através da membrana da derme para a epiderme é mantido. A epiderme é mais favorável à proliferação de melanoblastos em comparação com a derme, pois os melanoblastos que entram na epiderme começam a proliferar e a diferenciar-se massivamente[10].

De E15.5 a E16, os melanoblastos se distribuem por toda a epiderme e começam a se agrupar nos folículos pilosos em desenvolvimento. Os melanoblastos que atingem a epiderme e os folículos pilosos serão segregados em três populações distintas durante a morfogênese do folículo piloso (Figura 7B): 1- uma população se localizará na parte inferior do folículo piloso, onde se diferenciará em melanócitos maduros e contribuirá imediatamente para a pigmentação do cabelo durante o primeiro ciclo do cabelo; 2- uma outra população consiste em células-tronco de melanócitos que residem dentro de um nicho especializado de células-tronco, conhecido como protuberância do folículo piloso, e serve como reservatório para reabastecer os folículos pilosos adultos em regeneração com novos melanócitos; 3- a última população consiste em melanócitos que povoam a epiderme interfolicular, responsável pela pigmentação da pele. A colonização da

epiderme e dos folículos pilosos é um processo que requer a atividade de sinalização do KIT proto-oncogene (*KIT*). Em contraste, as populações de melanoblastos e melanócitos que se localizam dentro da derme são altamente dependentes da sinalização de endotelina-3 (*ET3*) em vez de sinalização KIT[10].

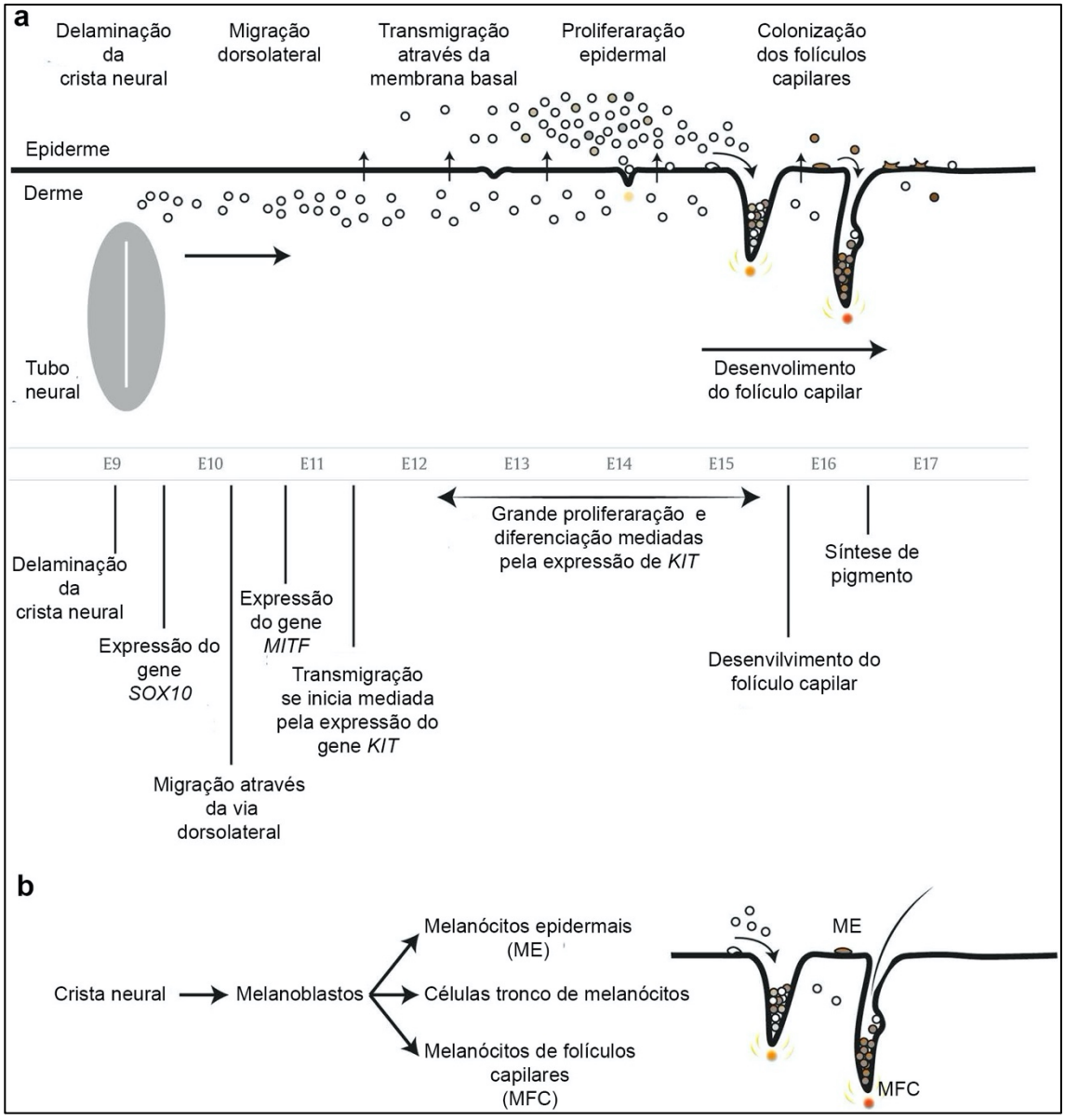


Figura 7- Processo de diferenciação dos melanoblastos em melanócitos durante o desenvolvimento embrionário. Fonte: Adaptado de Vandamme e Berx[10].

A pele humana adulta possui dois compartimentos denominados epiderme e derme, que são separados por uma barreira física, a membrana basal (Figura 8). A epiderme compreende múltiplas camadas de queratinócitos e inclui os melanócitos, glândulas sebáceas, glândulas sudoríparas e folículos pilosos. Enquanto isso, a derme é um agrupamento especializado de células de fibroblastos[12].

Existem aproximadamente 1.500 melanócitos por milímetros quadrados na epiderme humana, cerca de 3 bilhões dessas células na pele de um ser humano com estatura média. Os melanócitos são a minoria das células que compõem a pele e se dividem raramente, cerca de duas vezes por ano. A principal função dos melanócitos é sintetizar melanina para os queratinócitos vizinhos, promovendo proteção física contra a radiação ultra-violeta (RUV)[13].

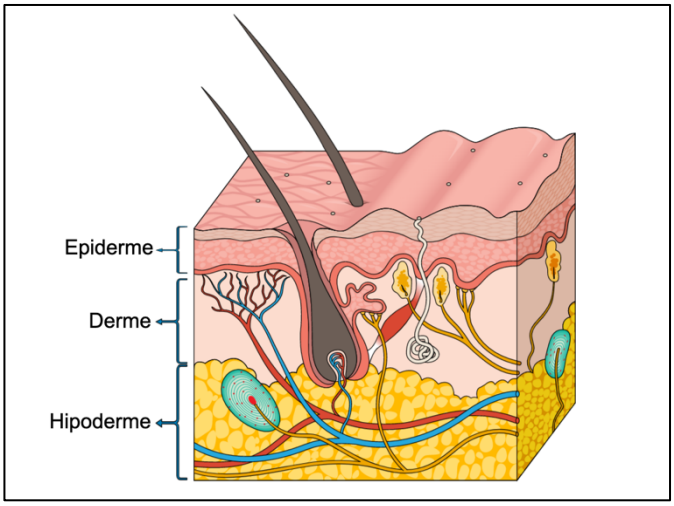


Figura 8- Desenho esquemático da pele humana adulta. Fonte: Autoria própria.

A melanogênese, via bioquímica responsável pela síntese da melanina, ocorre nos melanócitos, em organelas citoplasmáticas chamadas melanossomos (Figura 9). Dois tipos principais de melanina são produzidos: feomelanina, cores entre vermelho-amarelo; e eumelanina, cores entre marrom-preta. Em presença de oxigênio, a tirosinase oxida a tirosina em dopa (dioxifenilalanina) e essa em dopaquinona. A partir desse momento, a presença ou ausência de cisteína determina o rumo da reação para síntese de eumelanina

ou feomelanina. Na ausência de cisteína (glutationa), a dopaquinona é convertida em ciclodopa (leucodopacromo) e essa em dopacromo. Há duas vias de degradação de dopacromo: uma que forma DHI (dopa,5,6 diidroxiindol) em maior proporção; e outra que forma DHICA (5,6 diidroxiindol-2-ácido carboxílico) em menor quantidade. Esse processo é catalisado pela dopacromo tautomerase (Tyrp 2-Dct). Finalmente, esses diidroxiindóis são oxidados à melanina. A tirosinase relacionada à proteína 1 (Tyrp 1) parece estar envolvida na catalisação da oxidação da DHICA à eumelanina. Por outro lado, na presença de cisteína, dopaquinona rapidamente reage com tal substância para gerar 5-S-cisteinildopa, e, em menor proporção, a 2-S-cisteinildopa. Logo, as cisteinildopas são oxidadas em intermediários benzotiazínicos e, finalmente, produzem feomelanina[14].

A pele humana contém uma mistura de todos os tipos de melanina e em certo nível a proporção dessa mistura determina a pigmentação visível. A eumelanina, comparada à feomelanina, apresenta melhores propriedades de fotoproteção, incluindo maior resistência à degradação e capacidade de neutralização das espécies reativas de oxigênio (ROS).

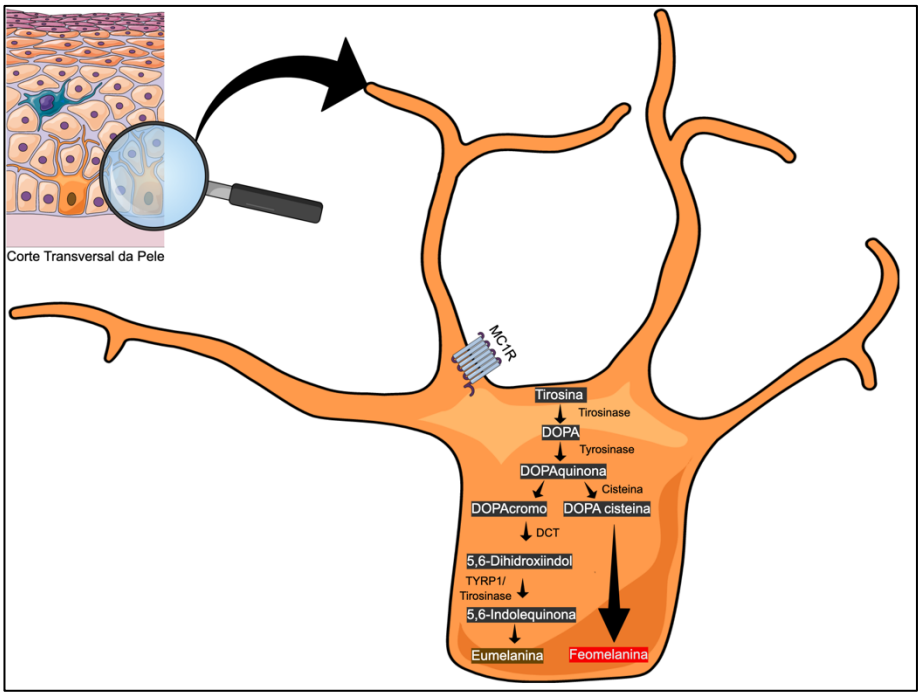


Figura 9- Panorama geral da síntese de melanina nos melanócitos. Fonte: Autoria própria.

1.4. Câncer: Doença Celular e Genética

O câncer é um termo genérico para definir um conjunto de mais de 100 tipos de doenças, cuja principal característica é o crescimento descontrolado das células, que se multiplica além dos seus limites e que podem invadir partes adjacentes e se espalhar para outros órgãos, processo conhecido como metástase. Esse crescimento anormal muitas vezes ocorre em decorrência de alterações genéticas, bem como por modificações epigenéticas. Conceitualmente, a primeira engloba alterações que mudam a sequência do DNA, incluindo duplicações, deleções, inserções ou mutações, que levarão a produção de um RNAm diferente e, consequentemente, a tradução em uma proteína anormal ou, até mesmo, não-funcional; a segunda são alterações que modificam o nível de expressão de uma proteína intracelularmente, seja pela impossibilidade dos fatores de transcrição reconhecerem as sequências da região promotora, seja por uma regulação pós-transcricional (discutido anteriormente), incluindo metilação, remodelamento de histonas e RNAs não-codificantes. Entretanto, estas alterações não modificam a sequência do DNA, portanto são reversíveis, o que não acontece com as alterações genéticas[5,15,16].

O processo de surgimento do câncer é chamado de oncogênese ou tumorigênese, sendo dividido em dois tipos: 1- câncer esporádico, iniciado a partir de alterações genéticas e/ou epigenéticas em células somáticas, que podem ocorrer espontaneamente ou em virtude da exposição a carcinógenos; 2- câncer hereditário, iniciado a partir de alterações genéticas nas células germinativas (óvulo ou espermatozoide) e que são herdadas entre as gerações de uma família, podendo levar a uma maior predisposição ao desenvolvimento do câncer. Estima-se que entre 5 a 10% de todos os tumores são herdados, sendo portanto, a grande maioria esporádico[17].

Durante a tumorigênese, essas alterações, sejam genéticas, ou epigenéticas, podem ocorrer em duas grandes classes de genes, os oncogenes ou os genes supressores tumorais. Os oncogenes são genes que codificam proteínas relacionadas ao crescimento celular, proliferação, sobrevivência e diferenciação, enquanto os genes supressores

tumorais, são genes que regulam todos esses processos, permitindo que a célula se mantenha em equilíbrio. No surgimento e desenvolvimento do câncer, alterações genéticas e/ou epigenéticas em oncogenes levam à excessividade destes eventos, enquanto em genes supressores tumorais, essas alterações promovem a perda de função do gene e, consequentemente, ocorre uma desregulação celular[15, 16].

É evidente a complexidade do câncer, portanto, não é uma doença que surge em decorrência de poucas alterações genéticas ou epigenéticas, uma vez que o próprio mecanismo celular consegue reparar muitos desses danos. Sendo assim, com a finalidade de compreender a diversidade dos tumores foi proposto por Hanahan e Weinberg (2000 e 2011) os *Hallmarks of Cancer*, que são capacidades biológicas adquiridas durante o desenvolvimento dos tumores. Dessa forma, as células normais precisariam adquirir uma sucessão dessas capacidades para o processo de tumorigênese se iniciar. Estes eventos incluem: manutenção da sinalização proliferativa, descontrole dos supressores de crescimento, resistência à morte celular, imortalidade, indução de angiogênese, ativação de invasão e metástase, desregulação energética, bloqueio da destruição do sistema imune, inflamação como indutor de tumores, instabilidade genômica e mutações[16, 18].

1.5. Melanoma: O Câncer Derivado dos Melanócitos

O melanoma é o tipo de câncer de pele mais agressivo, derivado dos melanócitos, células que se localizam na epiderme e que possuem a função de síntese, armazenamento e distribuição da melanina para as células epiteliais adjacentes, conferindo a essas proteção ao DNA contra os raios UV (discutido no ítem 1.3)[19].

Ainda não foi elucidado todos os mecanismos que levam o surgimento e desenvolvimento do melanoma, no entanto, Shain e colaboradores (2015)[20] mostraram com os dados de sequenciamento de 37 melanomas primários, em comparação com lesões precursoras adjacentes (nevos), que a evolução da doença a partir desses nevos se estabelece com o acúmulo de mutações somáticas, enquanto que variações no número de cópias do DNA é um processo importante para a metastização. Eles propuseram um

modelo de progressão da doença, em que a radiação UV possui papel importante desde o surgimento da lesão benigna, até a doença metastática e que a maioria dessas lesões possui mutações condutoras (*driver mutations*) importantes, que levam ao desenvolvimento da neoplasia. Eles postulam três possíveis caminhos de surgimento do melanoma. Na primeira hipótese, a lesão benigna já adquiriu a mutação V600E do gene *BRAF* (*v-Raf murine sarcoma viral oncogene homolog B*) e a mutação na região promotora do *TERT* (*telomerase reverse transcriptase*) é a responsável pela transição da lesão intermediária para o melanoma *in situ*. A segunda e terceira teoria é a existência de uma lesão intermediária entre a lesão benigna e o melanoma *in situ*, que surge por mutações no *NRAS* (*neuroblastoma RAS viral oncogene homolog*) e no *TERT*, ou pela mutação *BRAF* V600K concomitante com mutação em *TERT*, respectivamente. A transição do melanoma *in situ* para o melanoma invasivo, independente das três teorias, seria em virtude de alterações no gene *CDKN2A* (*cyclin-dependent kinase inhibitor 2A*), importante supressor tumoral, que leva a perda da função do gene e alterações na via SWI/SNF, responsável por mecanismos epigenéticos de enovelamento da cromatina nas histonas. E a metastização do tumor provavelmente se estabelece por perda de função de *PTEN* (*phosphatase and tensin homolog*) e P53.

Essas hipóteses sofreram críticas[21], especialmente porque ainda é controversa a ideia de que os nevos possam ser lesões precursoras, uma vez que a transformação maligna de um nevo individual ocorre em uma taxa de 0,00005 a 0,003% por ano[22, 23]. Além disso, apenas 20 a 30% dos melanomas são associados aos nevos[21], portanto, a grande maioria dos melanomas surge *de novo* (No tópico 1.9 isso é discutido em mais detalhes).

1.6. Epidemiologia do Melanoma

Segundo as estimativas do GLOBOCAN cerca de 324.635 novos casos de melanoma estavam previstos para serem diagnosticados em 2020, sendo 173.844 em homens e 150.791 em mulheres. Além disso, 57.043 mortes em decorrência deste tipo tumoral

estavam previstas para acontecer em todo o mundo durante este ano[24]. Levando-se em conta essa estimativa e a anterior de 2018, cuja previsão era de mais de 287.000 novos casos[25], a projeção global é que 510.000 novos de melanoma serão diagnosticados, um aumento de 50%, e 96.000 mortes decorrentes dessa doença ocorrerão em 2040, um aumento de 68%[26].

As estimativas na população brasileira, segundo o Instituto Nacional de Câncer, indicam que 8.450 novos casos de melanoma são esperados para cada ano do triênio 2020-2022, praticamente não existindo diferença em relação a incidência por sexo, sendo 4.200 em homens e 4.250 em mulheres. As maiores taxas estimadas para ambos os sexos encontram-se na região Sul[27].

Segundo o registro hospitalar de câncer do Hospital de Câncer de Barretos, 160 novos casos de melanoma foram diagnosticados na Instituição em 2019, sendo 46.9% homens e 53.1% mulheres. Além disso, cerca de 43% dos pacientes possuíam ensino fundamental incompleto, indicando o baixo nível de escolaridade desses pacientes. Cerca de 78% dos pacientes eram provenientes do Sudeste e 51% chegaram na Instituição com melanomas de baixo grau, *in situ* ou grau I[28].

Utilizando 3 décadas de registro de câncer de 6 populações com moderada até alta incidência de melanoma, incluindo a população branca dos Estados Unidos (EUA), Reino Unido, Suécia, Noruega, Austrália e Nova Zelândia, Whiteman e colaboradores[29] aplicaram modelos de idade-período-corte para projetar taxas de incidência futuras de melanoma até 2031. O número de pessoas com melanoma aumentou em cada população de 1982 a 2011, com aumentos relativos em novos casos de melanoma variando de 133% na Noruega a 278% no Reino Unido. A maior parte do aumento de novos diagnósticos de melanoma foi atribuível a mudanças na incidência de melanoma específico por idade, mais evidente no Reino Unido e na Suécia, sendo que em todas as seis populações, as maiores taxas de melanoma específico por idade foram observadas em idosos (> 80 anos).

Durante o período de previsão (2012-2031), todas as seis populações podem esperar grandes aumentos no número de melanomas diagnosticados. O número de melanomas diagnosticados em brancos nos EUA aumentará de cerca de 70.000 por ano entre 2007-

2011 para mais de 116.000 por ano entre 2026-2031. Ao longo de todo o intervalo de 1982-2031, o número de brancos norte-americanos diagnosticados com melanoma a cada ano quadruplicará, dos quais 79% podem ser atribuídos a aumentos nas taxas de melanoma específicas por idade e 21% podem ser atribuídos ao crescimento e envelhecimento da população. Mudanças de magnitude aproximadamente semelhante foram projetadas para as demais populações. Em relação a 1982, o número de pessoas com melanoma em 2031 aumentará substancialmente no Reino Unido (585%), Suécia (388%), Noruega (333%), Austrália (291%) e Nova Zelândia (362%). Em todas as populações, exceto na Austrália, cuja incidência do melanoma diminui desde 2005 em adultos jovens, a maior parte do aumento no número de pessoas diagnosticadas com melanoma será atribuível a aumentos nas taxas de melanoma específicas por idade, e não ao crescimento populacional[29].

A diminuição na projeção de incidência de melanoma na Austrália é corroborada por outro estudo mais recente que mostra que entre sete tipos de tumores, o melanoma teve a maior diminuição na porcentagem de casos projetados entre pessoas com menos de 50 anos, saindo de 20% em 2016 para 10% em 2031[30]. Entretanto, há discussões no meio acadêmico, se essa diminuição observada é em sua maioria em virtude da assertividade de campanhas de conscientização no país[31] ou causada pelo número crescente de jovens com baixo risco de melanoma que imigraram para o país[32, 33]. Para responder essa questão, um estudo modelou as possíveis mudanças na população suscetível ou em risco com menos de 30 anos, usando uma abordagem de simulação, na qual foi aplicado vários cenários para o papel da imigração no aumento populacional observado entre os australianos de 0 a 29 anos durante o período de estudo, corresponde de 1982 a 2010. Cada cenário foi modelado calculando o aumento populacional específico por idade e sexo de um ano para outro ano com base nos dados reais estimados da população residente. Em seguida, foi removido uma porcentagem desse aumento que se supunha ser devido à imigração de pessoas provenientes de países de baixo risco (Ásia, Oriente Médio e Ilhas do Pacífico). Os extremos foram que nenhum aumento populacional foi devido à imigração de pessoas provenientes de países de baixo risco (0% de imigração) e

que todo o aumento populacional foi devido à imigração dessas pessoas (100% de imigração). As estimativas de imigração de 10%, 25%, 50% e 75% foram incluídas como cenários intermediários[34].

Ao aplicar essa metodologia, encontrou-se um aumento significativo na incidência de melanoma em pessoas de 0 a 29 anos de 1982 até meados da década de 1990, seguido por uma diminuição significativa desde então em cada um dos seis cenários. Mesmo sob a suposição extrema de que todo o aumento líquido da população de jovens australianos foi devido à imigração de pessoas provenientes de países de baixo risco, a incidência de melanoma ainda diminuiu entre 1996 e 2010 em 3.2% ao ano. Um fator determinante disso é que o número de melanomas diagnosticados nessa faixa etária na Austrália atingiu um pico de cerca de 635 em meados da década de 1990, antes de diminuir para cerca de 400/ano a partir de 2007[34].

Além disso, essa mesma observação esperada na Austrália é esperada em outras regiões do mundo. Dados de projeções mostram que no período entre 2019-2043 a taxa de mortalidade geral e as taxas de mortalidade por melanoma padronizadas por idade em idosos de ambos os sexos aumentarão nos próximos anos na Espanha, enquanto as taxas em pessoas mais jovens se estabilizarão ou diminuirão progressivamente, sendo a faixa etária >85 anos o grupo com maior número de óbitos esperado[35]. Na Holanda, houve um aumento de mortalidade por melanoma de 0.39 por 100.000 habitantes em 1950 para 5.79 em 2018, sendo que as projeções apontam para um pequeno declínio de mortes por melanoma na população jovem do país, mas um aumento na população geral até 2045[36].

Entretanto, as projeções de incidências de câncer nos EUA mostram que em 2040 o melanoma será o segundo câncer com maior taxa de incidência em ambos os sexos, quando ajustado por variação percentual média anual da incidência de câncer, ficando atrás apenas do câncer de mama. Na população mais jovem incluindo pessoas até 49 anos, espera-se que o melanoma seja o quinto câncer mais incidente, atrás apenas de câncer de mama, colorretal, linfoma non-Hodgkin e rim[37].

As tendências e projeções de mortalidade por melanoma no Brasil são controvérsias. Dados do DATASUS (Departamento de Informática do Sistema Único de Saúde) mostram

que de 1998 a 2012 ocorreram 18.048 mortes por este câncer, sendo que a região sul apresenta a maior taxa de mortalidade, enquanto as regiões norte e nordeste apresentam as menores. Além disso, as projeções apontam para uma redução de mortalidade por este câncer no país até 2032 para ambos os sexos[38].

Por outro lado, utilizando dados de 23 registros de câncer cobrindo 315 hospitais de câncer do Brasil observou-se um total de 28.624 casos de melanoma diagnosticados entre os anos de 2000 e 2013. Houve um aumento de 2.52 casos de melanoma para 4.84 a cada 100.000 homens e um aumento de 1.94 casos de melanoma para 3.22 a cada 100.000 mulheres entre os anos de 2000 e 2013. Entretanto, em relação a mortalidade houve um discreto aumento em homens, partindo de 0.85 para 0.90 a cada 100.000 homens, enquanto foi observada uma discreta diminuição de mortalidade em mulheres de 0.56 para 0.53 a cada 100.000 mulheres no mesmo período de tempo[39].

Um outro estudo utilizando registros de câncer do estado de São Paulo entre 1996 a 2016 evidenciou um aumento de mortalidade em homens acima de 60 anos de idade, mas uma estabilidade na população geral[40]. A discrepância nos dados se deve a diversos fatores, mas principalmente aos estudos utilizarem dados hospitalares, o que não representa a totalidade da população brasileira.

1.7. Diagnóstico e Estadiamento do Melanoma

Para melhorar a detecção precoce do melanoma cutâneo, Friedman e colaboradores (1985)[41] desenvolveram o método "ABCD" como uma forma simples para avaliação de lesões cutâneas pigmentadas. Assim, considera-se a lesão como benigna quando o escore varia entre 1.0 e 4.75, suspeita entre 4.75 e 5.45, e altamente suspeita quando o escore é maior do que 5.45. Posteriormente, um "E" foi adicionado ao “ABCD” como um importante critério adicional na diferenciação de melanoma de lesões pigmentadas benignas[42] (Figura 10):

A – Assimetria: são traçados dois eixos perpendiculares entre si, para análise da simetria em cada metade, quanto à forma, estrutura e cor. Esse critério possui peso 1.3, então se a lesão for assimétrica em um eixo, multiplica-se um eixo de assimetria por 1.3 e se obtém 1.3. Entretanto, se a lesão for assimétrica em dois eixos, então multiplica-se dois eixos de assimetria por 1.3 e se obtém 2.6.

B – Borda (Irregularidade nas fronteiras): são traçados quatro eixos sobre a lesão, perpendiculares dois a dois, dividindo-se em 8. Então, para cada octante com interrupção brusca de borda é multiplicado o peso de 0.1.

C – Cores: muitas vezes os melanomas possuem cor marrom escuro ou preto, mas podem apresentar cores variadas com algumas áreas mais claras de marrom, vermelho, azulado ou acinzentado. Para cada cor que a lesão apresentar, multiplica-se por 0.5.

D – Diâmetro: geralmente maior do que 6 mm.

E – Evolução: mudanças ao longo do tempo, incluindo ampliação da lesão, mudança de cores ou novos sintomas.

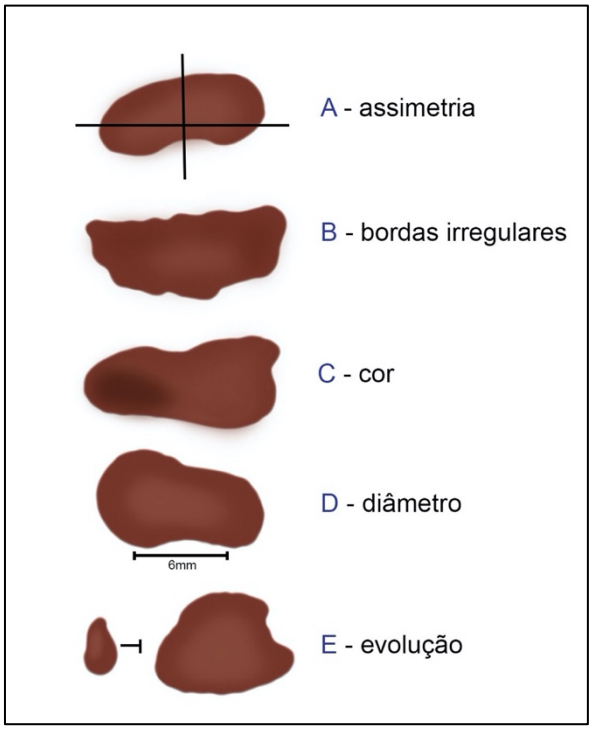


Figura 10- Sistema ABCDE para diagnóstico de melanoma.
Fonte: Autoria própria.

O diagnóstico precoce do melanoma primário foi significativamente melhorado com a introdução da dermatoscopia e dermatoscopia digital. Em pacientes com melanoma com sinais clínicos ou sintomas de doença metastática (estágio III ou IV), radiografias de tórax e exames de imagem como ultrassonografia, tomografia computadorizada, ressonância magnética, tomografia por emissão de pósitrons (PET) e cintilografia óssea são ferramentas diagnósticas necessárias. Além disso, no diagnóstico de melanoma metastático, é importante realizar avaliação laboratorial dos níveis de desidrogenase lática (DHL) e testes de função hepática[43].

Apesar da importância inicial dos achados dermatoscópicos, os principais fatores prognósticos estão relacionados aos dados histopatológicos, sendo o nível de Clark e espessura de Breslow os mais clássicos em melanoma (Figura 11). O primeiro determina a invasão do tumor nas camadas da pele e está relacionado, portanto, a extensão da espessura do tumor, sendo que quanto menor o nível histopatológico da lesão, melhor o prognóstico. Como a pele apresenta diferentes espessuras ao longo do corpo, o nível de Clark é menos preciso como fator prognóstico. Por outro lado, a espessura de Breslow associa diferentes prognósticos de acordo com a espessura vertical máxima do melanoma primário medido em milímetros através de uma régua acoplada ao microscópio na análise histopatológica, e compõe desde o seu desenvolvimento na década de 1970 até os dias atuais, o fator prognóstico de maior poder preditivo para tumores localizados[44].

Além destes dois fatores prognósticos clássicos, existem outros importantes, incluindo a ulceração, que está associada a maior agressividade tumoral, com maior probabilidade de metástase e menor sobrevida[45]. Assim como o índice mitótico, que é um importante fator prognóstico para melanomas “finos” (espessura inferior a 1mm)[46]. A DHL está envolvida na síntese de energia celular e a sua superexpressão correlaciona-se com o metabolismo anaeróbio tumoral e reduz a dependência celular de oxigênio. Há um tempo pesquisas demonstram que o nível de DHL encontra-se aumentado no soro de pacientes com melanoma e que está associado com uma menor sobrevida naqueles com doença avançada[47].

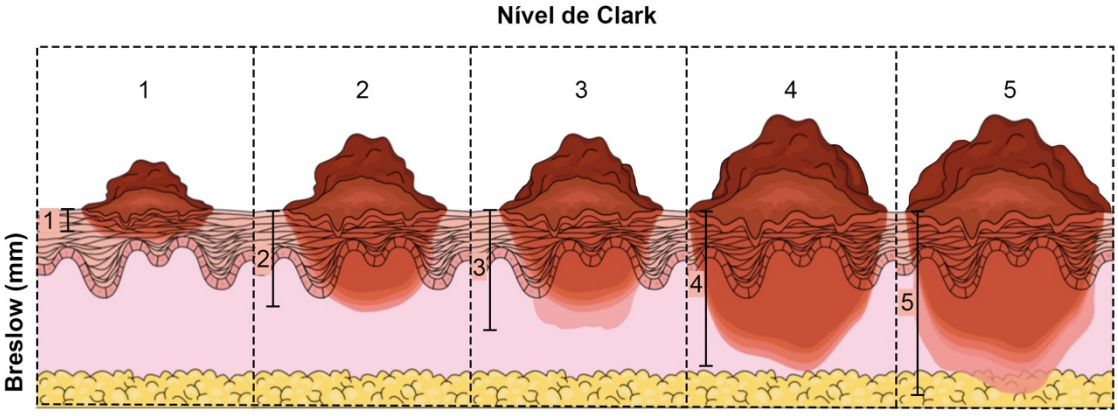


Figura 11- Fatores prognósticos mais importantes em melanoma, o nível de Clark e a espessura de Breslow.
Fonte: Autoria própria.

O estadiamento dos tumores malignos é regido por normais internacionais do AJCC *(American Joint Committee on Cancer*), que publica regularmente o livro "TNM – *Classification of Malignant Tumors"*, sendo que a mais recente é a 8ª edição, publicada em 2017[48].

O sistema TNM visa à análise da extensão anatômica da doença e possui como base a avaliação de três componentes: T - a extensão do tumor primário; N - a ausência ou presença e a extensão de metástase em linfonodos regionais; e M - a ausência ou presença de metástase à distância. A adição de números a estes três componentes indica a extensão da doença maligna.

As principais mudanças na oitava edição do Manual de Estadiamento do Câncer da AJCC incluem: 1) as medidas da espessura do tumor a com a aproximação de 0.1 mm, não 0.01 mm; 2) as definições de T1a e T1b são revisadas (T1a: <0.8 mm sem ulceração; T1b: 0.8-1.0 mm com ou sem ulceração ou <0.8 mm com ulceração), sendo que o índice mitótico não é mais critério da categoria T; 3) o estágio IA patológico (mas não clínico) é revisado para incluir T1b N0 M0 (anteriormente estágio patológico IB); 4) os descritores da categoria N "microscópicos" e "macroscópicos" para as metástases linfonodais regionais são redefinidos como "clinicamente ocultos" e "clinicamente aparentes"; 5) os grupos de prognósticos de estágio III são baseados nos critérios da categoria N e critérios da categoria T (isto é, espessura e ulceração do tumor primário) e aumentaram de 3 para

4 subgrupos (estágios IIIA-IIID); 6) as definições das subcategorias N são revisadas, com a presença de microssatélites, satélites ou metástases em trânsito agora categorizadas como N1c, N2c ou N3c com base no número de linfonodos regionais envolvidos no tumor, se houver; 7) os descritores são adicionados a cada designação de subcategoria M1 para o nível de DHL (elevação de DHL não mais em estágios superiores a M1c); e 8) uma nova designação de M1d é adicionada para metástases do sistema nervoso central[48, 49].

1.8. Regras para Classificação

Para a descrição precisa do melanoma, deve haver confirmação histológica da doença e os procedimentos para a avaliação das categorias N e M são exame físico e diagnóstico por imagem. Os linfonodos regionais são aqueles referentes à localização do tumor primário.

O sistema de estadiamento descrito abaixo é referente ao TNM 8ª edição[48]:

Estágio 0: Inclui o tumor que está contido na epiderme, a camada superficial da pele (Tis), que não se espalhou para os linfonodos próximos (N0) ou outros órgãos (M0). Esse estágio é conhecido como melanoma *in situ*.

Estágio I: Inclui o tumor que não tem mais do que 2 mm de espessura. T1a quando a espessura do tumor é menor do que 0.8 mm e não apresenta ulceração; T1b inclui tumor com espessura maior ou igual a 0.8 mm, enquanto T2a inclui tumor com espessura entre 1.1-2.0 mm sem presença de ulceração. Na impossibilidade de informação a cerca da ulceração, o tumor é T1 quando sua espessura é menor ou igual a 1.0 mm e T2 quando a mesma é maior do que 1.1 mm. O tumor não se espalhou para os linfonodos (N0) próximos ou outros órgãos (M0).

Estágio II: inclui o tumor entre 1.1-2.0 mm de espessura com ulceração (T2b), tumor com espessura entre 2.1-4.0 mm sem informação sobre ulceração (T3), tumor com espessura entre 2.1-4,0 mm sem presença de ulceração (T3a) e com o mesmo

intervalo de espessura, mas com a presença de ulceração (T3b). Ainda, inclui tumor com mais de 4.0 mm de espessura sem informação de ulceração (T4). O tumor não se espalhou para os linfonodos próximos (N0) ou outros órgãos (M0).

Estágio IIIA. O tumor tem até 2 mm de espessura e pode (ou não) ter ulceração (T1/T1a/T1b ou T2a). Ele se espalhou para três ou menos linfonodos, podendo ser N1a, se detectado em um linfonodo e clinicamente oculto; N1b, se detectado em um linfonodo e clinicamente detectável; N1c, se nenhum linfonodo é detectado, mas o tumor apresenta satelitose; N2a, se detectado em dois ou três nódulos e clinicamente oculto; N2b, se detectado em dois ou três nódulos e clinicamente detectável; e N2c, se detectado em um nódulo, clinicamente detectável ou oculto, mas com presença de instabilidade de microssatélite. O tumor não se espalhou para outros órgãos (M0).

Estágio IIIB. Uma das situações abaixo:

- Não existe sinal do tumor primário (T0) e está disseminado para apenas um linfonodo (N1b) ou para áreas muito pequenas da pele próxima (tumores satélites) ou para os canais linfáticos da pele ao redor do tumor (sem atingir os linfonodos) (N1c). Não se espalhou para outros órgãos (M0).
- O tumor tem até 4 mm de espessura e pode ou não ter ulceração (T1/T1a/T1b, T2/T2a/T2b, ou T3a). Se espalhou para apenas um linfonodo (N1a ou N1b) ou para áreas muito pequenas da pele próxima (tumores satélites) ou para os canais linfáticos da pele ao redor do tumor (sem atingir os nódulos linfáticos) (N1c) ou se espalhou para 2 ou 3 nódulos linfáticos (N2a ou N2b). Não se espalhou para outros órgãos (M0).

Estágio IIIC. Uma das situações abaixo:

- Não existe sinal do tumor primário (T0) e está disseminado para 2 ou mais linfonodos (N2b ou N3b) ou se espalhou para áreas muito pequenas da pele próxima (tumores satélites) ou para os canais linfáticos da pele ao redor do tumor e atingiu os gânglios linfáticos próximos (N2c ou N3c) ou se espalhou para os linfonodos do grupo (N3b ou N3c). Não se espalhou para outros órgãos (M0).
- O tumor tem até 4 mm de espessura e pode (ou não) ter ulceração (T1/T1a/T1b, T2/T2a/T2b ou T3a). Se espalhou para áreas muito pequenas da pele próxima (tumores satélites) ou para os canais linfáticos da pele ao redor do tumor e atingiu os linfonodos próximos (N2c ou N3c) ou se espalhou para 4 ou mais linfonodos próximos (N3a ou N3b) ou se espalhou para os linfonodos do grupo (N3b ou N3c). Não de espalhou para outros órgãos (M0).
- O tumor tem entre 2 e 4 mm de espessura ou pode ser mais espesso que 4,0 mm e tem ulceração (T3b) ou tem mais de 4 mm e não tem ulceração (T4a). Se espalhou para um ou mais linfonodos ou se espalhou para áreas muito pequenas da pele próxima (tumores satélites) ou para os canais linfáticos da pele ao redor do tumor (N1 ou superior). Não se espalhou para outros órgãos.
- O tumor tem mais do que 4 mm de espessura e tem ulceração (T4b). Se espalhou para até 3 linfonodos (N1a/N1b ou N2a/N1b) ou para áreas muito pequenas da pele próxima (tumores satélites) ou para os canais linfáticos da pele ao redor do tumor e pode (N2c) ou não (N1c) ter atingido os linfonodos próximos. Não se espalhou para outros órgãos (M0).

Estágio IIID. O tumor tem mais do que 4 mm de espessura e apresenta ulceração (T4b). Se espalhou para 4 ou mais linfonodos (N3a ou N3b) ou se espalhou

para áreas muito pequenas da pele próxima (tumores satélites) ou para os canais linfáticos da pele ao redor do tumor (sem atingir os nódulos linfáticos) ou se espalhou para os linfonodos do grupo (N3c). Não se espalhou para outros órgãos (M0).

Estágio IV. O tumor pode ser de qualquer espessura e pode (ou não) ter ulceração (Qualquer T). Pode (ou não) ter se espalhado para os linfonodos próximos (Qualquer N). Ele se espalhou para os linfonodos distantes ou outros órgãos, como pulmões, fígado ou cérebro (M0).

1.9. Subtipos Histológicos de Melanoma

Os melanomas são classificados como cutâneos, compreendendo três subtipos de tumores que ocorrem em locais expostos à radiação UV, denominados melanoma extensivo superficial, melanoma nodular e melanoma lentigo maligno. Além disso, compreendem outros subtipos mais raros de melanoma, que ocorrem em locais protegidos contra a UV, tais como os melanomas acrais lentiginosos, os melanomas de mucosa, o melanoma ocular, entre outros. Essa definição foi estabelecida por Clark em 1969 e se baseia nos padrões microscópicos de crescimento e está associada às características clínicas, tais como localização anatômica do tumor primário e idade do paciente[50].

Abaixo estão descritos em mais detalhes os subtipos de melanoma incluídos no presente trabalho (Figura 12):

- **Melanoma Extensivo Superficial (MES):** é o tipo mais comum de melanoma cutâneo, sendo responsável por aproximadamente 70% de todos os melanomas primários em pessoas de pele clara. O pico de incidência de MES é durante a quarta e quinta década de vida. Embora o MES possa ser localizado em qualquer sítio anatômico do corpo, os locais mais comuns são os membros inferiores nas mulheres e o tronco nos homens. Histologicamente este tumor é caracterizado pela presença de um componente intraepidérmico que exibe um padrão de crescimento pagetóide e aninhado em todos

os níveis da epiderme. A citologia de células tumorais intraepidérmicas predominante é a de uma célula epitelióide com abundante citoplasma eosinofílico, anfofílico ou com distribuição de finos grânulos citoplasmáticos de melanina. Os núcleos podem ser grandes e ter um ou mais nucléolos proeminentes. O componente dérmico também pode ser caracterizado por numerosos ninhos de tamanho variável e ocasionalmente um nódulo expansivo de tumor. Nos casos com tumor dérmico abundante, a heterogeneidade citológica torna-se mais aparente[43, 51].

- **Melanoma Nodular (MN):** é o segundo tipo mais comum de melanoma, sendo responsável por 10 a 15% de todos os melanomas. A principal característica do melanoma nodular é o crescimento rápido, sendo mais comumente visto no tronco, cabeça e pescoço. O MN é caracterizado por um predomínio de tumor invasivo dérmico. Um componente intraepidérmico pode estar presente, mas diretamente sobrepõe-se ao melanoma invasivo. Ocasionalmente, o componente epidérmico é tão pequeno que sugere a possibilidade de o tumor representar uma metástase dérmica. O tumor é composto de pequenos ninhos e agregados de células tumorais que juntos formam o nódulo tumoral global. Este tumor tem uma propensão a crescer verticalmente e exibir comportamentos agressivos[43, 51].
- **Melanoma Lentigo Maligno (MLM):** é responsável por aproximadamente 5-10% de todos os melanomas. É visto principalmente em pessoas mais velhas, com um pico de incidência na sétima década de vida. Locais típicos para o desenvolvimento de LMM são a face (particularmente bochechas e nariz) e o pescoço. O risco para o desenvolvimento de LMM está associado a uma exposição solar acumulativa ao longo da vida. Histologicamente, o componente intra-epidérmico é caracterizado por uma proliferação predominantemente individual de células localizadas nas camadas basais da epiderme. Este padrão de crescimento é descrito como lentiginoso em tumores melanocíticos benignos e malignos. As células tumorais têm grandes núcleos cromáticos densos e ocasionalmente são multinucleadas. As células tumorais estendem-se pelo epitélio do folículo piloso, mantendo a proximidade da camada

basal. Ninhos intraepidérmicos e disseminação pagetóide podem ser observados. O componente invasivo dérmico de MLM pode exibir células fusiformes[43, 51].

- **Melanoma Acral Lentiginoso (AM):** representa menos de 2% de todos os melanomas em pessoas de pele clara. No entanto, esse tipo é a forma mais comum de melanoma entre populações de pele escura com incidência relativamente baixa de melanoma (afro-americanos, asiáticos e hispânicos), onde pode ser responsável por 30-70% de todos os melanomas. ALM aparece nas palmas das mãos, solas dos pés e sob a região ungueal. O pico de incidência é na sétima década de vida. A fase de crescimento radial tipicamente mostra uma proliferação difusa de grandes melanócitos atípicos ao longo da junção derme-epiderme, os quais estão dispersos individualmente em um "padrão lentiginoso". Melanócitos únicos predominam sobre ninhos juncionais na maioria dos campos. Os melanócitos anormais em forma de fuso ou epitelióides exibem proeminentes processos dendríticos. Os núcleos são freqüentemente grandes, atípicos e hipercromáticos[43, 52].

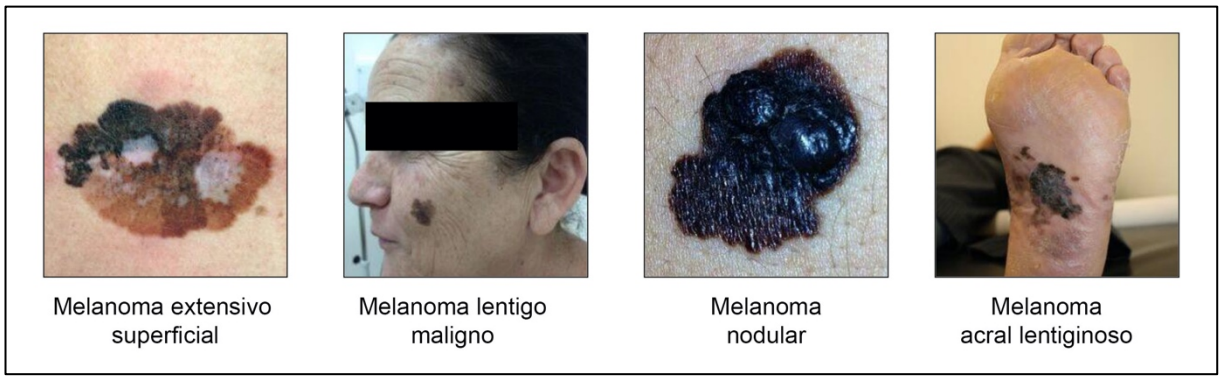


Figura 12 - Principais subtipos de melanoma. Fonte: Autoria própria.

1.10. Fatores de Risco e Prevenção do Melanoma

A Agência Internacional de Pesquisa em Câncer (IARC) da Organização Mundial de Saúde (OMS) classifica todo o espectro de radiação ultra-violeta dentro do grupo 1, a mais importante categoria de substâncias carcinogênicas aos seres humanos[53]. Juntamente

com o tabaco, obesidade, hábito alimentar, álcool e certos vírus, a exposição ao sol contribui com cerca de 45% das mortes por câncer que são evitáveis de acordo com a *American Cancer Society*[54].

A prevenção do melanoma se baseia principalmente em intervenções que buscam a informação e educação, por meio de políticas públicas e de órgãos não-governamentais, em ambientes que incluem desde locais de recreação ao ar livre, creches, escolas e locais de trabalho. Essas ações têm se baseado, principalmente, em precaver queimaduras solares, reduzir o tempo de exposição ao sol e o uso de fotoprotetor[55]. Inclusive, existem ensaios clínicos que demonstram que o uso regular de fotoprotetor reduz significativamente o risco de desenvolvimento de melanoma, mesmo décadas depois[56]. Entretanto, é importante precaver o número de queimaduras ao longo da vida, uma vez que este evento é associado a um aumento de risco de desenvolvimento de melanoma em todos os estágios da vida, não apenas na primeira infância, como já foi proposto[57].

Em 2016, Lazovich e colaboradores publicaram o primeiro estudo examinando associações entre bronzeamento artificial e o risco de desenvolvimento de melanoma, levando-se em consideração idade e sexo. Neste estudo de base populacional (681 casos e 654 controles), a idade de início de uso de bronzeamento artificial inferior a 25 anos foi associada a um aumento de 6 vezes no risco de desenvolvimento de melanoma entre mulheres com idade inferior a 30 anos, um aumento de 3.5 vezes em mulheres entre 30-39 anos e um aumento de 2.2 vezes em mulheres entre 40-49 anos. Uma resposta dose-dependente foi observada entre as mulheres, independentemente da idade[58].

Corroborando com essa observação, um ano mais tarde foi publicado um estudo prospectivo que incluiu 141.045 mulheres com uma média de acompanhamento de aproximadamente 14 anos, demonstrando que o risco de melanoma aumentou com o aumento do número cumulativo de sessões de bronzeamento artificial. Também foi observado que idade do início das sessões de bronzeamento inferior a 30 anos foi associada a um risco maior em comparação com nunca ter realizado esse procedimento ao longo da vida. Além disso, as mulheres que iniciaram o bronzeamento artificial antes dos 30 anos de idade eram, em média, 2.2 anos mais jovens ao diagnóstico do que aquelas

que nunca foram usuárias dessa prática[59]. Por fim, estima-se que cerca de 3.438 casos de melanoma na Europa sejam atribuídos ao bronzeamento artificial, sendo as mulheres abaixo dos 35 anos as mais afetadas[60].

Uma das justificativas mais utilizadas pela indústria de bronzeamento para apoiar e promover o uso de bronzeamento artificial é que a vitamina D é produzida na pele sob a influência da irradiação UVB e, portanto, o bronzeamento artificial deve ser considerado benéfico para os seres humanos, principalmente aqueles que vivem em lugares com inverno intenso. Esse tópico tem ganhado especial atenção da comunidade científica e da população em geral nos últimos anos, uma vez que a Vitamina D vem sendo associada como um fator protetivo contra inúmeras doenças, incluindo doenças cardiovasculares, auto-imunes e condições oncológicas[61].

Muitos estudos demonstram que de fato o bronzeamento artificial aumenta o nível sérico de Vitamina-D[62-65], entretanto, esse aumento é, na maioria dos casos, apenas transitório e não é sustentado a não ser pelo uso contínuo de bronzeamento artificial[66-68], o que já foi dito anteriormente como sendo fator de risco para desenvolvimento de melanoma. Além disso, estudos têm demonstrado que as câmaras de bronzeamento emitem espectros maiores de radiação UV, comparados com aqueles que naturalmente são emitidos pelo sol[69, 70]. Há evidências científicas suficientes que preenchem todos os critérios epidemiológicos de causalidade entre bronzeamento artificial e risco de desenvolvimento de melanoma, portanto, considerar essa prática como uma estratégia para aumentar os níveis séricos de Vitamina-D é totalmente desencorajado pela comunidade científica. Além disso, pode-se realizar a suplementação da Vitamina-D por via oral, com a vantagem de que não há exposição a nenhum carcinógeno[71].

Nos países em que o bronzeamento artificial é uma prática comum, a prevenção do melanoma inclui conscientizar sobre o risco de desenvolvimento de melanoma que essa prática causa, além da importância de regulamentar os espaços onde ela é realizada[72]. A Agência Nacional de Vigilância Sanitária (Anvisa) proibiu o bronzeamento artificial no Brasil para fins estéticos na resolução de número 56 de 09 de novembro de 2009[73], baseado em diversos estudos científicos, como os relatados aqui, que mostram

causalidade entre bronzeamento artificial e aumento de risco de desenvolvimento de câncer de pele.

Em termos de característica genética dos pacientes, número de nevos melanocíticos e história familiar, constituem os principais fatores de risco de desenvolvimento de melanoma[74]. Por definição, os nevos melanocíticos são acumulações benignas de melanócitos ou células névicas e podem ser congênitos ou adquiridos. A contagem total de nevos está positivamente correlacionada com o risco de melanoma e varia com base no número, tamanho e tipo de nevo[75, 76]. O resultado de uma meta-análise demonstrou que as pessoas com mais de 100 nevos têm um risco 7 vezes maior de desenvolvimento de melanoma[77]. Em relação ao tamanho, nevos maiores (> 5 mm) e gigantes (> 20 cm) estão associados a um risco significativamente maior de melanoma[78].

Além disso, estima-se que entre 34-59% dos melanomas esporádicos, que são tumores que não surgem por alterações herdadas, ocorrem no cenário de um nevo displásico pré-existente[79]. Esses nevos se caracterizam como lesões de no mínimo 5 mm, com características atípicas como pigmentação variável, contorno irregular assimétrico e bordas indistintas[80]. A presença de um único nevo atípico aumenta em duas vezes o risco de desenvolvimento de melanoma, enquanto que a presença de dez ou mais dessas lesões conferem um risco aumentado de doze vezes de desenvolver a doença[81]. Os melanomas que se desenvolvem no contexto de nevos displásicos pré-existentes, geralmente localizam-se no tronco em pacientes mais jovens e pertencem ao subtipo extensivo superficial[82].

Em relação a história familiar de melanoma como fator de risco, até o momento dois genes de predisposição ao melanoma foram identificados, sendo eles o *CDKN2A* e o *CDK4* (quinase dependente de ciclina 4). Ambos podem ser herdados no padrão autossômico dominante demonstrando uma transmissão vertical da doença, igualmente afetando homens e mulheres[83]. Não parece haver outro *locus* importante além dos citados, pois a prevalência dos novos genes de predisposição ao melanoma é bastante rara. Embora grandes avanços tenham sido feitos na identificação de outras novas variantes co-

segregantes dentro dos tipos de melanoma, é provável que muitas mutações raras causadoras de doenças permaneçam desconhecidas[84].

Por fim, algumas características fenotípicas como pele e olhos claros, sardas, incapacidade da pele de bronzear e cor dos cabelos, também são marcas de grande susceptibilidade ao desenvolvimento de melanoma[85]. Pessoas com essas características fenotípicas devem seguir as medidas de prevenção de exposição ao sol como as mencionadas anteriormente, enquanto pessoas com número alto de nevos melanocíticos e história familiar de melanoma devem procurar um serviço de saúde especializado para o acompanhamento contínuo e prevenção do surgimento da doença.

1.11. A Era OMIC do Melanoma

Há um esforço global para se caracterizar as principais alterações moleculares do melanoma, especialmente em virtude do avanço do conhecimento científico de que os tumores são entidades genéticas complexas e o surgimento de novas tecnologias que permitiram as chamadas análises em grande escala ou sequenciamento de nova geração (NGS, *Next Generation Sequencin*g). Essas tecnologias permitem o acesso integral das informações genéticas e epigenéticas dos pacientes e, por isso, são chamados de dados OMICs.

Assim, surgiram projetos multicêntricos, destacando-se o TCGA (*The cancer genome atlas network*) e o ICGC (*International Cancer Genome Consortium*), que visaram mapear todas as alterações moleculares envolvidas na patogênese de vários tumores, incluindo o melanoma, afim de caracterizar o aspecto molecular destes tumores, identificar possíveis alvos terapêuticos, biomarcadores de progressão, biomarcadores de repostas a diversos tratamentos, dentre tantas outras possibilidades clínicas que esse mapeamento permite.

O TCGA[86] foi o primeiro projeto a mapear as alterações moleculares exclusivamente de 331 melanomas cutâneos, incluindo análise de dados OMICs, como genoma, metiloma

e transcriptoma. Para isto, eles analisaram o exoma destes pacientes, bem como mais de 450.000 CpGs ao longo do genoma e a expressão de RNAm e microRNA.

A principal contribuição desse trabalho foi definir os quatro principais grupos moleculares do melanoma:

- **Subtipo BRAF:** compreende o maior subtipo genômico e é definido pela presença de mutações no *hot-spot* (regiões do DNA onde são mais frequentes o surgimento de mutações) do gene *BRAF*. Cerca de 52% dos pacientes abrigavam mutações somáticas neste gene, sendo a grande maioria a mutação V600E. Mutações nesse gene foram mutualmente exclusivas com mutações no *hot-spot* do gene *NRAS*. Em contraste, as mutações em *BRAF* em regiões não-*hot-spot* ocorreram em concomitância com as mutações em locais *hot-spot* em *RAS* (N/H/K) e *NF1*.
- **Subtipo RAS:** é o segundo subtipo principal, sendo definido pela presença de mutações nos *hot-spot* do gene *RAS*, incluindo alterações de aminoácidos conhecidas com consequências funcionais, em todos os três membros da família RAS (N-,K- e H-RAS). No geral, 28% dos pacientes apresentaram mutações somáticas em *NRAS*, sendo a mutação mais frequente a Q61R, seguida de Q61K, Q61L e Q61H. Além disso, foram identificadas mutações menos frequentes em outros membros da família *RAS*, incluindo quatro mutações no gene *HRAS* (G13D, G13S e Q61K) e três mutações no *KRAS* (G12D, G12R e Q61R); todas foram mutuamente exclusivas com as mutações *NRAS* e *BRAF* V600 e K601.
- **Subtipo NF1:** o NF1 (Neurofibromatose tipo 1) é o terceiro gene mais frequentemente mutado, cujas alterações apareceram em 14% dos pacientes. Mais da metade de suas mutações foram eventos de perda de função (LoF), incluindo 27 *nonsens*e, 9 sítios splice e 4 *frame-shif*. O subtipo NF1 apresentou a maior prevalência de mutação (39 mutações/Mb), mais que o dobro dos outros três subtipos. Como o NF1 é uma proteína ativadora de GTPase conhecida por regular negativamente a atividade de RAS através de sua atividade GTPase intrínseca, a mutação LoF de *NF1* pode ser vista como uma maneira alternativa de

ativar a via de sinalização canônica da MAPK (*Mitogen-activated protein kinase*). Além disso, as mutações em *NF1* foram mutuamente exclusivas com mutações em regiões *hot-spot* de *BRAF*, mas não com mutações em regiões *hot-spot* de RAS.

- **Subtipo Triplo-Negativo (TN):** é um grupo heterogêneo que se caracteriza pela falta de mutações no gene *BRAF*, *N*/*H*/*K-RAS* ou *NF1*.

Os pacientes do subtipo BRAF eram mais jovens do que os demais subtipos, enquanto os pacientes do subtipo NF1 eram significativamente mais velhos. Não foram observadas associações significativas na sobrevida de acordo com o subtipo molecular.

O subtipo TN se caracteriza com perda de assinatura UV, sendo que apenas 30% dos pacientes apresentaram essa assinatura, em comparação com 90.7% do subtipo BRAF, 93.5% do subtipo RAS e 92.9% do subtipo NF1. Em contraste, os pacientes TN tiveram mais alterações no número de cópias e arranjos estruturais complexos em comparação com os outros grupos. As mutações na região promotora do *TERT* foram observadas em 75.0% do subtipo BRAF, 71.9% do RAS e 83.3% do NF1, mas em apenas 6.7% do grupo TN, sugerindo um mecanismo alternativo de ativação do *TERT* (por exemplo, amplificação ou rearranjo) nesses melanomas.

As duas principais contribuições do ICGC[87] foi mapear o genoma dos pacientes com melanomas cutâneos, diferente do TCGA, que havia avaliado exclusivamente os éxons, bem como incluir subtipos mais raros da doença, os melanomas acrais e de mucosas. Este trabalho identificou assinaturas de exposição ao sol nos melanomas cutâneos, enquanto que nos melanomas acrais ou de mucosas seis das nove assinaturas observadas foram descritas em outros tipos de câncer, mas nunca em melanoma e provavelmente não estão relacionadas à exposição ao sol. A assinatura mais frequente nesse grupo foi a assinatura 1, resultado da desaminação endógena espontânea da 5-metilcitosina. Além disso, foi observada uma frequência maior de variantes estruturais, incluindo duplicações, deleções e inversões, bem como uma frequência maior de rearranjos complexos nos melanomas acrais e de mucosas, na comparação com melanomas cutâneos.

Este estudo do ICGC corroborou com outros artigos que evidenciaram uma frequência menor de mutações em *TERT* e maior em *KIT* em melanomas acrais e de mucosas, bem como a falta de mutação em *TP53*, *PTEN*, *DDX3X*, *RASA2*, *PPP6C*, *RAC1* ou *RB1* nestes subtipos, indicando que as vias moleculares que os conduzem diferem enormemente daqueles que conduzem ao desenvolvimento de melanoma cutâneo, mesmo que em todos os subtipos de melanoma haja um domínio de mutações na via de sinalização da MAPK, PI3K (*phosphoinositol 3-kinase*) e RTK (tirosina quinase).

O trabalho de Wouters e colaboradores[88] publicado em 2017 analisando o metiloma de 14 nevos, 33 tumores primários e 28 tumores metastáticos de melanoma é uma das publicações com dado OMIC até o momento que mais corroborou para o entendimento das alterações epigenéticas no surgimento e desenvolvimento do melanoma. Nesse trabalho eles evidenciaram o perfil de metilação diferencial entre lesão benigna e tumor, bem como entre tumor em fases iniciais e mais avançadas.

Além disso, eles identificaram e validaram a metilação do gene *HOXA*9 como evento importante no desenvolvimento do melanoma e a metilação do *TBC1D16* como importante na progressão da doença. Ademais, determinaram uma assinatura prognóstica da metilação de *PON3* e a expressão de *OVOL1.* Em relação a classificação funcional dos genes hipermetilados/*down*-regulados houve um envolvimento significativo de várias vias de sinalização relacionadas ao melanoma e surgimento de metástases, incluindo polaridade célula/tecido e adesão célula-célula, enquanto a superexpressão associada à hipometilação foi enriquecida em vias envolvendo regulação do sistema imune e processos inflamatórios[88].

Ao analisar o genoma dos pacientes com melanoma cutâneo utilizando diversos dados públicos, Trucco e colaboradores[89] descreveram 10 genes que são frequentemente mutados, comparando pacientes que abrigam assinatura molecular UV e aqueles que não a abrigam (assinatura UV é discutida em detalhe no ítem 1.12), sendo eles *LRP1B*, *ADGRV1*, *XIRP2*, *PKHD1L1*, *USH2A*, *DANH9*, *PCDH15*, *DNAH10*, *TP53* e *PCDHAC1*. Além disso, a mutação nesses genes foi associada à sobrevida global mais longa. Um outro trabalho deste mesmo grupo demonstrou que a presença de mutação nesses 10 genes

corretamente diferenciava melanomas de mucosas que abrigavam, daqueles que não abrigavam assinatura molecular UV, descrevendo, portanto, um painel que permite a triagem rápida daqueles pacientes com melanoma UV-expostos[90].

1.12. Tratamento do Melanoma

A remoção cirúrgica do melanoma é o tratamento padrão para os melanomas primários e consiste na excisão completa do tumor com margens de segurança, ou seja, uma margem tecidual contendo pele normal e tecido subcutâneo subjacente, sendo que a quantidade da margem é bem estabelecida e varia conforme a espessura do tumor [43, 91].

Como os linfonodos regionais são os locais mais frequentes de metástases, totalizando entre 45-65% de todas as recorrências, o tratamento dos linfonodos regionais é muito importante em pacientes com melanoma cutâneo[92]. O mapeamento linfático e a biópsia de linfonodo sentinela foram desenvolvidos por Morton e colaboradores[93] como uma técnica minimamente invasiva para o estadiamento de linfonodos regionais. A linfocintilografia mapeia com precisão a drenagem linfática do local do melanoma cutâneo até o primeiro linfonodo, chamado de "sentinela", na drenagem regional, que será removido e examinado pelo patologista. Se o linfonodo sentinela for positivo, os pacientes poderão ser submetidos à dissecção linfonodal completa ou encaminhados para tratamento adjuvante e observação quanto a cadeia linfonodal, enquanto se o linfonodo sentinela for negativo, a dissecção e tratamentos complementares serão evitados. A presença de maior acometimento linfonodal é indicador de menores taxas de sobrevida[94,95-99].

Nas últimas três décadas, a dacarbazina foi o *gold-standard* do tratamento do melanoma metastático, apesar de nunca ter sido demonstrado um aumento consistente da sobrevida global. Esse quimioterápico é um análogo das purinas, que inibe a síntese de DNA e RNA[100]. Em virtude do melhor conhecimento da patogênese molecular do

melanoma e da imunologia do câncer, houve uma revolução no tratamento de pacientes com estágio avançado e irressecável, portanto desde 2011 a dacarbazina deixou de ser considerada a primeira opção para terapia sistêmica do melanoma.

A descoberta da alta frequência de mutação no éxon 15 do gene *BRAF* em uma substituição da valina por ácido glutâmico na posição 600 (Val600Glu ou V600E) fez desta proteína oncogênica um alvo ideal para a terapia[101, 102]. Dessa forma surgiu o PLX4032 (Vemurafenibe), que é um inibidor que induz a parada do ciclo celular, a proliferação e inicia a apoptose em células exclusivamente portadoras da mutação V600E[100]. Os primeiros ensaios clínicos com essa terapia-alvo demonstraram até 80% de taxa de resposta clínica em portadores de melanoma metastático com esta mutação em *BRAF*[103]. Em 17 de agosto de 2011 a instituição governamental norte-americana "Food and Drug Administration" (FDA) aprovou o Zelboraf™ (Vemurafenibe) para o tratamento de melanomas inoperáveis ou metastáticos com a mutação *BRAF* V600E. Em 2013 houve a aprovação do Dabrafenib, um inibidor de BRAF similar ao Vemurafenib, por este mesmo órgão governamental estadunidense[104].

Enquanto essa alta taxa de resposta global foi sem precedentes, vários pacientes desenvolveram resistência após a resposta inicial, sendo observada recidiva e/ou progressão do tumor. A duração média da resposta à PLX4032 é de aproximadamente sete meses[105]. Sendo assim, surgiu o Trametinibe, um inibidor de *MEK*, que assim como o *BRAF*, faz parte da via das MAPK. A terapia de combinação de inibidores de *BRAF* e *MEK* aumentou a sobrevida e permitiu a diminuição do tumor dos pacientes[106, 107]. Então, em janeiro de 2014 a FDA aprovou a terapia de combinação destas duas drogas em pacientes com melanoma metastático com mutação V600E e V600K[104].

Além dos avanços no descobrimento de inibidores de alterações específicas relacionadas à tumorigênese do melanoma, houve mais recentemente grande evolução na imunoterapia para esse tipo tumoral. Os pontos de checagem (*checkpoint*) imunológicos são parte do importante mecanismo pelo qual as respostas imunes do organismo são reguladas e o dano autoimune dos tecidos normais é evitado. As células tumorais podem sequestrar esse mecanismo para desenvolver resistência ao ataque imunológico,

principalmente dos linfócitos T. Os checkpoints imunológicos envolvem a ligação de um ligante a um receptor, produzindo um sinal inibitório nas células imunes efetoras que freiam a resposta imune. Os anticorpos monoclonais foram projetados para se ligar aos receptores ou ligantes envolvidos, bloqueando a sinapse inibitória e liberando a ativação dos linfócitos T (Figura 13)[108].

O Ipilimumabe é um anticorpo monoclonal de imunoglobulina humana (IgG1) inibidor de ponto de checagem imunológico, que bloqueia totalmente CTLA-4 (linfócito citotóxico associado a antigeno-4) e promove imunidade antitumoral. Ele foi o primeiro agente a demonstrar uma melhoria na sobrevida global em um estudo randomizado de fase III em pacientes com melanoma avançado[109]. Mais recentemente, o desenvolvimento de anticorpos que bloqueiam PD-1 (proteína de morte celular programada 1) têm demonstrado uma alta eficácia e baixa toxicidade, e sendo considerado como o maior avanço no tratamento de câncer na última década[108]. O desenvolvimento de anticorpos anti-PD-1 mostrou benefício clínico em pacientes com melanoma metastático irresecável, bem como uma ampla variedade de tumores[110].

Para os pacientes de alto risco com melanoma ressecável, a terapia sistêmica adjuvante pode ser oferecida para reduzir o risco de recorrência do melanoma após a cirurgia[111]. Pacientes com melanoma de estágio II de alto risco (ou seja, aqueles com tumor de espessura > 4 mm ou > 2 mm e com ulceração, ou seja, estágio IIB-C) e pacientes com linfonodo positivo têm risco aumentado de recorrência e demonstram pior sobrevida específica por melanoma[49].

Em 2015, o Ipilimumabe foi aprovado para o tratamento adjuvante de todos os pacientes com melanoma em estágio III com base nos resultados do EORTC 18071 (CA184-029), que demonstrou que altas doses de Ipilimumabe (10mg/kg) administradas a cada três semanas, em um total de quatro doses, aumentou a sobrevida livre de recidiva em comparação com o placebo[112]. Em 2017, o estudo Checkmate-238 demonstrou aumento na sobrevida livre de recidiva com Nivolumab (anti-PD-1) em comparação com Ipilimumab e com um perfil de toxicidade mais favorável[113]. O estudo KEYNOTE-054, um ensaio clínico de fase III randomizado e controlado por placebo, demostrou que o Pembrolizumabe (anti-

PD-1) apresentava um aumento de sobrevida livre de recidiva comparável ao Ipilimumabe[98]. Com base nos resultados desses estudos, tanto o Nivolumabe, quanto o Pembrolizumabe foram aprovados para o tratamento adjuvante de pacientes com melanoma ressecável em estágio III e pacientes com melanoma ressecável em estágio IV oligometastático, um estágio no espectro do processo metastático em que a terapia local agressiva pode erradicar completamente a carga tumoral, levando a uma potencial cura[114].

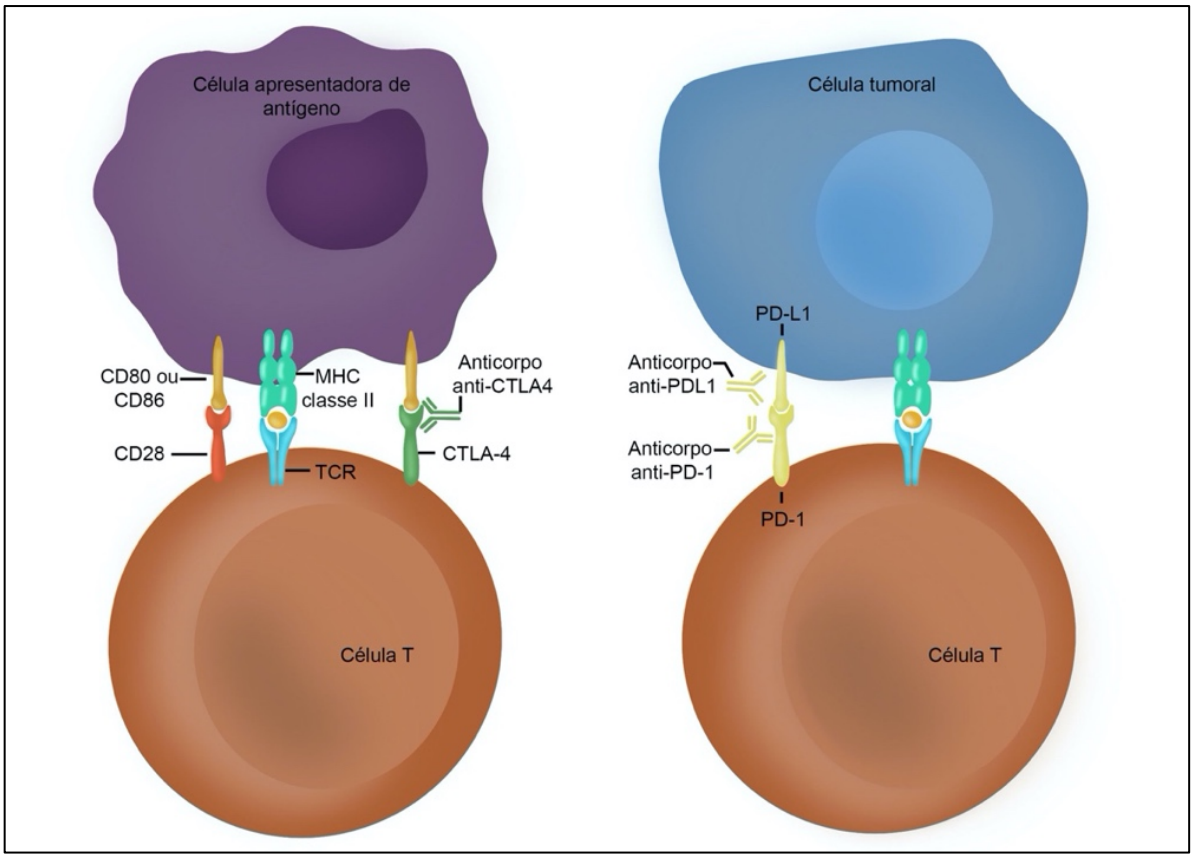


Figura 13 - Panorama geral da ação dos inibidores dos pontos de checagem imunológicos. Fonte: Autoria própria.

Mais recentemente um novo receptor associado a exaustão de células T linfocítica, o gene ativador de linfócito 3 (LAG-3), foi ativamente bloqueado com anticorpos anti-LAG-3 com resultados clínicos expressivos associados a anti-PD-1[115]. Em resumo, o tratamento de melanoma trata-se de uma área em constante evolução.

Entretanto, é importante destacar que mais de 30% dos pacientes tratados com essas terapias recidivam nos primeiros dois anos[98], e outros 15–20%[113] progridem rapidamente antes que a terapia adjuvante possa começar. Em muitos tumores avançados, os inibidores de pontos de checagem imunológicos induzem o controle durável do tumor e melhoram a sobrevida global[116].

Tanto para doença sistêmica, como para doença de alto risco de recorrência, o tratamento com imunoterapia e com terapia alvo tem evoluído rapidamente e constantemente. Atualmente é disponível de forma efetiva tratamento adjuvante para portadores de melanoma em estádio III e para pacientes estádio IV em várias modalidades e combinações. Assim, os tratamentos direcionados e novos avanços na terapia alvo e imunoterapia vêm se mostrando eficientes em melhorar a sobrevida e são agora uma parte do atendimento clínico de rotina dos pacientes com melanoma em vários países[117].

Em relação a realidade do Brasil em termos de tratamento do melanoma, um trabalho do nosso grupo de pesquisa em melanoma incluiu 1.848 pacientes com melanoma tratados no Hospital de Câncer de Barretos (HCB) entre os anos de 1996 e 2015. Nesse período, dos 313 pacientes que fizeram tratamento sistêmico, 298 foram submetidos a quimioterapia, enquanto 67 fizeram uso de imunoterapia e 19 de terapia-alvo. A sobrevida específica por câncer em 5 anos de pacientes tratados apenas com quimioterapia foi de 20%, enquanto essa taxa foi de 3.2% em pacientes que foram submetidos a imunoterapia ou terapia-alvo, uma diferença que foi estatisticamente significante. Considerando o tempo de início do tratamento sistêmico, a sobrevida mediana foi de 13.96 meses naqueles tratados com terapia-alvo ou imunoterapia e 4.40 naqueles que receberam apenas quimioterapia[118].

1.13. Radiação Ultravioleta (RUV)

A luz solar contém um espectro de RUV de comprimentos de onda entre 200-400 nm e é dividido em três comprimentos de onda principais: UVA (315-400 nm), UVB (280-315 nm) e UVC (200-280 nm). Estima-se que 95% da RUV que atinge a terra seja UVA e 5% seja UVB, sendo que a totalidade de UVC é absorvida pela camada de ozônio estratosférico[119]. O comprimento de onda UVB é mais curto e penetra apenas na camada epidérmica da pele, enquanto que o da UVA é mais longo e pode penetrar mais profundamente nas camadas epidérmica e dérmica da mesma[120] (Figura 14).

A RUV tem inúmeros efeitos no corpo humano que podem ser divididos em duas categorias: mutacionais e não-mutacionais. Os efeitos mutacionais estão ligados a produção de fotoprodutos, como os dímeros de ciclobutano de pirimidina (DCPs) e o fotoproduto (6-4)-piridimina-pirimidona (64PPs). Essas lesões induzidas por UV são formadas através de uma reação fotoquímica, cuja eficiência depende do comprimento de onda, seguido pela absorção direta da energia UV pelas bases de DNA[121] (Figura 15A). Se o dano induzido por RUV não for reparado e a replicação começar na célula, os dímeros podem causar paralisação e colapso da forquilha de replicação, levando a quebras de fita dupla do DNA. Esses eventos podem ser prejudiciais para a célula e podem levar à eventual morte celular[121].

Os efeitos mutagênicos da RUV foram extensivamente estudados em fagos, bactérias, leveduras e mamíferos, o que culminou na designação de uma assinatura UV molecular. Essa se caracteriza por mais de 60% de substituições de bases de citosina (C) → timina (T) em sítios de dipirimidina e mais de 5% de CC → TT, embora essas últimas raramente ocorram[122]. Os fotoprodutos são eficientemente reparados pela célula através da via de reparo por excisão de nucleotídeos (REN). No entanto, a desaminação espontânea dos DCPs forma uma base de uracila (Figura 15 B). Essa incompatibilidade resultante é reparada pela síntese de translesão (STL), emparelhando corretamente uma base de adenina onde deveria haver uma timina. A desaminação espontânea de

fotoprodutos de citosina metilada produz uma base de timina incompatível. A reparação desta lesão leva à incorporação de adenina[123].

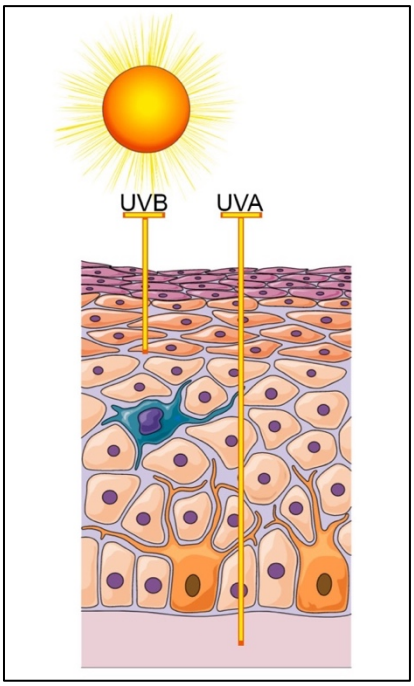


Figura 14 - Capacidade de penetrância na pele dos raios UVA e UVB. Fonte: Autoria própria.

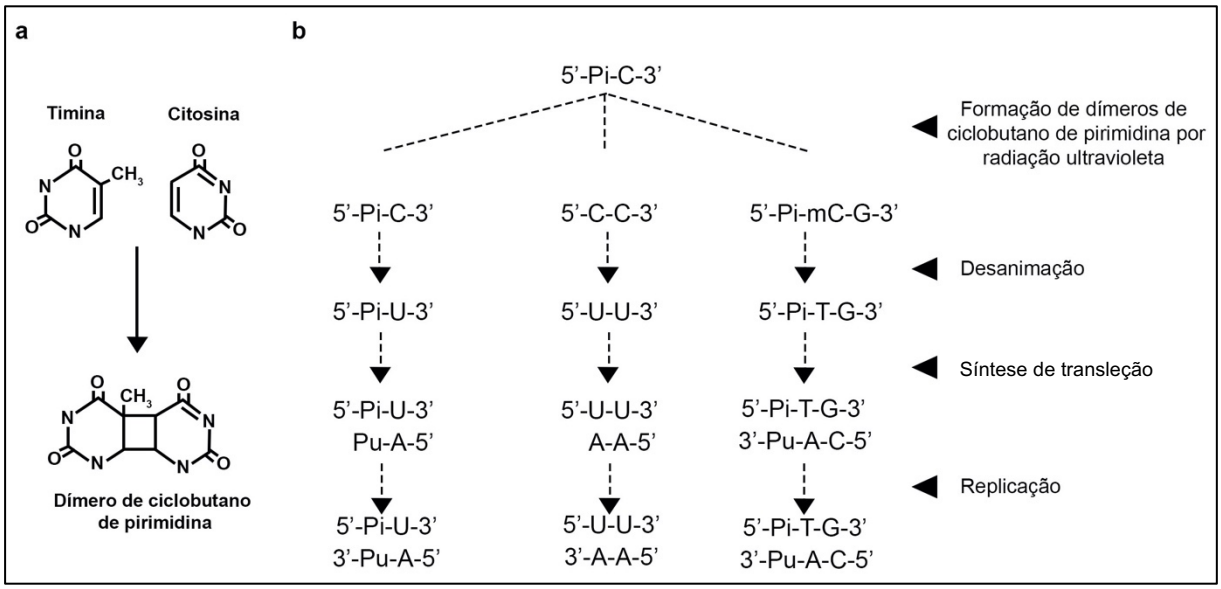


Figura 15 - Mecanismo de formação de mutações do DNA induzidas por exposição à radiação UV. A- Formação dos dímeros de ciclobutano de pirimidina. B- As três vias possíveis de formação dessas mutações, incluindo a mais clássica. Pi: pirimidina; Pu: purina e mC: metilcitosina. Fonte: Adaptado de Ikehata e Ono[121].

Tem sido demonstrado que a RUV induz CPDs preferencialmente em locais de dipirimidina contendo 5-metilcitosina (mC)[124-126]. No genoma dos mamíferos, as citosinas são metiladas em uma proporção alta nas sequências 5'-CG-3' (sítio CpG), fornecendo muitos locais de dipirimidina contendo mC. A abundância em mC sugere uma alta sensibilidade do genoma de mamífero em locais CpG aos comprimentos de onda UV mais longos, como UVB e UVA. Na verdade, mutações no gene p53 encontradas em tumores de pele humanos, que provavelmente foram induzidas por UV solar, mostram a assinatura UV e ocorrem frequentemente nos locais de dipirimidina associados ao CpG metilado (mCpG)[127-129].

Em estudos mutagênicos e carcinogênicos usando pele de camundongo exposta a RUV, foram relatados padrões de mutação semelhantes com frequente presença de transições C → T em locais de dipirimidina associados a mCpG[130, 131]. Assim, a faixa de comprimento de onda UVB e UVA parece induzir preferencialmente mutações de assinatura UV em locais de dipirimidina associada a mCpG (Py-mCpG).

Por outro lado, os efeitos não-mutacionais da RUV são danos causados ao DNA de forma indireta. A produção de espécies reativas de oxigênio ocorre através da absorção de UVA pelo pigmento melanina na pele. As ROS reagem com o DNA e podem produzir danos oxidativos na base, como 8-hidroxiguanina e timina glicol no DNA, ou podem induzir quebras de fita[132, 133]. Além disso, podem reagir com nucleotídeos celulares, produzindo nucleotídeos oxidados, como 8-hidroxideoxiguanosina-trifosfato, que ainda podem ser usados como precursores de nucleotídeos para a síntese de DNA[134]. Alguns desses tipos de dano oxidativo ao DNA e nucleotídeo são conhecidos por serem mutagênicos[133, 135].

A radiação ultravioleta não só causa danos ao DNA, mas também é um potente agente imunossupressor. Isso foi demonstrado em uma série de experimentos realizados por Kripke e colaboradores[136]. Em camundongos, tumores de pele induzidos por radiação ultravioleta foram transplantados em camundongos que foram irradiados com UVB e naqueles que não foram irradiados. Nos camundongos irradiados com UVB, os tumores continuaram a crescer, enquanto os camundongos não irradiados foram capazes de

rejeitar os tumores transplantados. Além disso, esse estudo demonstrou que as células T com função supressora ou reguladora são, de fato, um componente importante no processo de imunossupressão induzida por UV, uma vez que essa exposição ativa essas células.

Sabe-se atualmente que a imunossupressão induzida por RUV é complexa e não mediada apenas por um processo. Essa complexidade se baseia na capacidade da RUV de modular quatro famílias principais de fatores de crescimento: receptor de fator de crescimento epidérmico (EGFR), receptor de fator de crescimento derivado de plaquetas, receptor de fator de crescimento de fibroblastos e receptor de insulina, além de controlar a expressão de citocinas, cada uma das quais tem efeitos imunossupressores[137-139].

A imunossupressão direta localmente na pele ocorre quando o UVB impacta diretamente as células de Langerhans, sendo essas células dendríticas essenciais para a apresentação de antígenos ao sistema imunológico. A radiação ultravioleta induz a depleção das células de Langerhans mesmo em doses baixas e essa capacidade parece ser genética[140].

2. Justificativa

Frente ao que foi amplamente apresentado na introdução deste trabalho, nota-se que há um esforço na comunidade científica internacional para se mapear as alterações genômicas e epigenômicas do melanoma, o que abriu e continua abrindo um leque de oportunidades para a aplicação clínica, incluindo elucidação de potenciais novos alvos terapêuticos, resposta a diversas terapias, bem como a melhor classificação da doença. No entanto, ainda é desconhecida as mudanças no epigenoma atribuídos à exposição à UV em pacientes com melanoma cutâneo e o epigenoma não é atualmente considerado para classificar os diferentes subtipos patológicos da doença, mesmo que o epigenoma tenha um papel crucial no mecanismo adaptativo de diversos tipos de exposição, incluindo a UV, maior fator ambiental de risco de desenvolvimento de melanoma. Além disso, ainda não foi realizada uma

análise integrativa de dados OMICs com o intuito de elucidar genes com alto potencial de serem condutores e fatores prognósticos em melanoma, levando-se em consideração a presença ou ausência de exposição à UV.

3. Hipótese

Este estudo teve como hipótese que a exposição à radiação ultra-violeta desencadeia distúrbios epigenéticos que corroboram, junto com as modificações genéticas, no desenvolvimento de melanoma, e que a presença dessas alterações epigenéticas podem ter impacto clínico. Além disso, uma análise integrativa de dados genômicos e epigenômicos poderia melhorar a atual classificação da doença, culminando em melhor manejo dos pacientes e, consequentemente, maior sobrevida e qualidade de vida dos mesmos.

4. Objetivos

Objetivo Geral

Avaliar as mudanças no epigenoma atribuídos à exposição à UV em pacientes com melanoma cutâneo, bem como investigar se as alterações epigenéticas, interagindo com eventos transcricionais e mutacionais associados à exposição à UV, constituem mecanismos biológicos que sustentam as diferenças histopatológicas em melanomas e que conduzem o desenvolvimento da doença.

Objetivos Específicos:

- Realizar o metiloma de 54 pacientes com melanomas cutâneos diagnosticados no HCB e comparar o perfil de metilação daqueles que possuem a assinatura UV predita por *whole genome sequencing* com aqueles que não possuem a mesma;

- Validar os achados encontrados em nossa casuística usando os dados públicos do TCGA;
- Associar a metilação das CpGs selecionadas com a expressão dos genes no TCGA;
- Associar a metilação das CpGs selecionadas com a sobrevida dos pacientes tratados no HCB e incluídos no TCGA;
- Realizar análise integrativa utilizando transcriptoma, metiloma e genoma, com o intuito de elucidar potencial genes condutores de melanoma;
- Realizar análise integrativa utilizando metiloma e genoma, com o intuito de classificar dois diferentes subtipos de melanoma, cutâneo e acral, levando-se em consideração a presença de assinatura molecular de exposição à UV.

5. Resultados

5.1. Artigo Principal

Esse artigo é referente aos resultados fruto dessa tese de doutorado, que testa a hipótese e contempla os objetivos descritos na mesma. O artigo encontra-se publicado na *Nature Communications*, uma prestigiada revista científica classificada pela CAPES como Qualis A1 (Fator de impacto: 17.7), e a candidata ao título de doutora é a única primeira autora do trabalho.

Esse artigo foi realizado em colaboração com a IARC, no Grupo de Epigenética, sob co-orientação do Dr. Akram Ghantous, em um esforço institucional e governamental de internacionalização dos programas de pós-graduação do Brasil.

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-022-31488-w OPEN

Cutaneous and acral melanoma cross-OMICs reveals prognostic cancer drivers associated with pathobiology and ultraviolet exposure

Anna Luiza Silva Almeida Vicente[1,2✉], Alexei Novoloaca[2], Vincent Cahais[2], Zainab Awada[2], Cyrille Cuenin[2], Natália Spitz[2], André Lopes Carvalho[1,3], Adriane Feijó Evangelista[1], Camila Souza Crovador[1], Rui Manuel Reis[1,4,5], Zdenko Herceg[2,7], Vinicius de Lima Vazquez[1,6,7] & Akram Ghantous[2,7✉]

Ultraviolet radiation (UV) is causally linked to cutaneous melanoma, yet the underlying epigenetic mechanisms, known as molecular sensors of exposure, have not been characterized in clinical biospecimens. Here, we integrate clinical, epigenome (DNA methylome), genome and transcriptome profiling of 112 cutaneous melanoma from two multi-ethnic cohorts. We identify UV-related alterations in regulatory regions and immunological pathways, with multi-OMICs cancer driver potential affecting patient survival. *TAPBP*, the top gene, is critically involved in immune function and encompasses several UV-altered methylation sites that were validated by targeted sequencing, providing cost-effective opportunities for clinical application. The DNA methylome also reveals non UV-related aberrations underlying pathological differences between the cutaneous and 17 acral melanomas. Unsupervised epigenomic mapping demonstrated that non UV-mutant cutaneous melanoma more closely resembles acral rather than UV-exposed cutaneous melanoma, with the latter showing better patient prognosis than the other two forms. These gene-environment interactions reveal translationally impactful mechanisms in melanomagenesis.

[1] Molecular Oncology Research Center, Barretos Cancer Hospital, Barretos, São Paulo, Brazil. [2] Epigenomics and Mechanisms Branch, International Agency for Research on Cancer (IARC), Lyon, France. [3] Early Detection Prevention and Infections Branch, International Agency for Research on Cancer (IARC), Lyon, France. [4] Life and Health Sciences Research Institute (ICVS), Medical School, University of Minho, Braga, Portugal. [5] ICVS/3B's-PT Government Associate Laboratory, Braga, Guimarães, Portugal. [6] Department of Surgery—Melanoma and Sarcoma, Barretos Cancer Hospital, Barretos, São Paulo, Brazil. [7] These authors contributed equally: Zdenko Herceg, Vinicius de Lima Vazquez, Akram Ghantous. ✉email: annaluizaalmeida@hotmail.com; GhantousA@iarc.who.int

Melanoma is a neoplasm arising from melanocytes in the skin, mucosa, or uvea[1]. It accounts for more than 75% of skin cancer-related deaths though it represents <5% of all cutaneous malignancies[2]. The incidence of melanoma has been increasing worldwide[3] and this trend has been observed for decades in some populations (e.g., the US)[4].

Epidemiological studies have highlighted that the strongest risk factors for cutaneous melanoma development are severe sunburns during childhood and intense intermittent ultraviolet (UV) exposure, which consists of UVC (100–280 nm), UVB (290–320 nm), and UVA (320–400 nm)[5]. However, there are types of melanoma that arise in body parts protected from direct UV light, and these are acral, mucosal and uveal melanomas. These types represent uncommon cancers, among which the most frequent is the acral melanoma, which occurs on the glabrous skin (the skin of palms of the hands and the soles of the feet) and the subungual area[6,7]. Even though it is rare in the general population, acral melanoma is the most common melanoma among people with darker skin[8].

The melanoma genome has the highest mutation burden of any cancer and a predominant C>T nucleotide transition signature attributable to UV radiation[9,10]. Recently, ten mutated UV signature genes were identified in both clinical samples and animal models, and patients harboring the UV mutation signature presented longer disease-free and overall survival[11]. Although associations between genetic changes and UV exposure have been well characterized, the role of epigenetic modifications induced by UV exposure has never been investigated directly in human melanoma tissues (Supplementary Data 1). Epigenetic mechanisms function as central players in tumorigenesis and as molecular sensors to environmental factors[12]. In fact, CpG DNA methylation sites are highly sensitive to UV damage, as evidenced from experimental approaches of UV exposure using cell line and animal models[13].

Furthermore, the DNA methylation profile of acral melanomas is barely characterized, which could be due in part to its scarcity. It is also unclear whether molecular differences between UV-related and non UV-related melanoma types are due to intrinsic pathological characteristics, extrinsic responses to UV exposure or a combination of both. To address these questions, a comparative study encompassing both cutaneous and acral melanomas would represent an important step forward, with particular focus on epigenetic mechanisms as they can function as both sensors to exposures and key determinants of cell identity. The most recent melanoma classification by the World Health Organization (WHO), including the Blue Books by the International Agency for Research on Cancer (IARC), presented evidence based on epidemiologic, clinical, histopathologic and genomic features[14], while not yet encompassing epigenomics.

We hypothesize that epigenetic alterations, interplaying with transcriptional and mutational events, constitute critical biological mechanisms underpinning intrinsic pathological differences and extrinsic responses to UV exposure in cutaneous and acral melanomas. We perform differential DNA methylome-wide analysis in cutaneous melanoma patients comparing UV-exposed and non UV-exposed melanomas in two independent clinical cohorts, including a sample population from Brazil which encompasses the white and pigmented phenotypes (Fig. 1). UV exposure status is inferred from UV mutational signatures derived from whole genome sequencing (WGS) or whole exome sequencing (WES). This is followed by functional genomic, pathway and methylation-expression analysis of the identified DNA methylation alterations, assessment of their cancer driver roles using a multi-OMICs approach, investigation of their effect on patient survival, and validation of the top hits using bisulfite pyrosequencing. The methylome landscape of cutaneous melanoma is then compared to that of acral melanoma to elucidate the relative contributions of intrinsic pathological and extrinsic UV-related differences towards shaping the cancer epigenome of the two major UV-related and non UV-related melanoma types (Fig. 1).

Results

Cross genome-methylome analysis of UV exposure in cutaneous melanoma. UV mutation status was inferred in cutaneous melanoma patients using WGS and WES data from Barretos Cancer Hospital (BCH) in the context of International Cancer Genome Consortium (ICGC)-Brazil project and The Cancer Genome Atlas (TCGA) study, respectively (Fig. 1). Similar characteristics were observed in the BCH and TCGA-cutaneous melanoma patients, including larger proportions of the male sex, white skin phenotype, metastatic tumor type, UV mutation signature, and BRAF molecular group (Table 1). Primary tumors and *BRAF* mutations were relatively more enriched in BCH than in TCGA patients (Table 1, $p = 9.40e{-}03$, $p = 3.20e{-}03$, respectively).

We observed that UV-mutant cutaneous melanoma patients have higher melanoma-specific survival relative to non UV-mutant patients in both BCH and TCGA (Fig. 2a). In order to investigate whether the DNA methylome functions as a molecular sensor to UV exposure and underlies the difference in survival between UV-mutant *versus* non UV-mutant cutaneous melanoma patients, DNA methylome-wide analysis based on Infinium HumanMethylation450 (450 K) array was performed in BCH samples and compared with that in the TCGA cohort (the quality-control analysis and selection of the appropriate statistical model, including adjustment for potential confounders, are described in the Methods section and Supplementary Data 2–5).

In BCH melanomas, of the 2620 differentially methylated regions (DMRs), 1541 (58.8%) were hypermethylated and 1,079 (41.2%) were hypomethylated (Fig. 2b; Supplementary Data 2). A similar proportion of hypermethylated (62.8%; 378 out of 602) and hypomethylated (37.2%; 224 out of 602) DMRs was observed in TCGA (Fig. 2b; Supplementary Data 6). The enrichment distributions in CpG regulatory or density regions were also similar in both cohorts. Specifically, in CpG regulatory regions, the significant enrichments in both cohorts were those of hypomethylated DMRs in regions 1–5 Kb upstream of the transcription start site and of hypermethylated DMRs in promoters, exon/intron boundaries and 5′UTR ($p < 0.001$) (Fig. 2c). In CpG density regions, the significant enrichments were those of hyper- or hypomethylated DMRs in CpG islands or shores ($p < 0.001$) (Fig. 2d).

The DNA methylome marks UV exposure with effect on immunomodulation. In order to prioritize the top DMRs that distinguish UV-mutant and non UV-mutant cutaneous melanoma patients (Supplementary Data 7, 8), we applied the filters described in Supplementary Fig. 1a to focus on DMRs encompassing at least 3 CpGs, with consistent directions of effect, absolute effect sizes ≥10%, and not enriched in single-nucleotide polymorphisms (SNPs). The resultant methylome map distinctly clustered UV-mutant from non UV-mutant patients in both BCH and TCGA (Fig. 3a). In the BCH cohort, cluster C1 (as defined by Euclidean distance) was fully occupied by non UV-mutant samples (Fig. 3a) and exhibited a DNA methylation profile that was visually distinct, with an upper hypermethylation (red) stretch and a lower hypomethylation (blue) stretch, relative to the other clusters. Even if C1 is merged with the adjacent cluster C2, the non UV-mutant patients remain statistically enriched in this combined cluster ($p = 1.95e{-}03$), which now encompasses

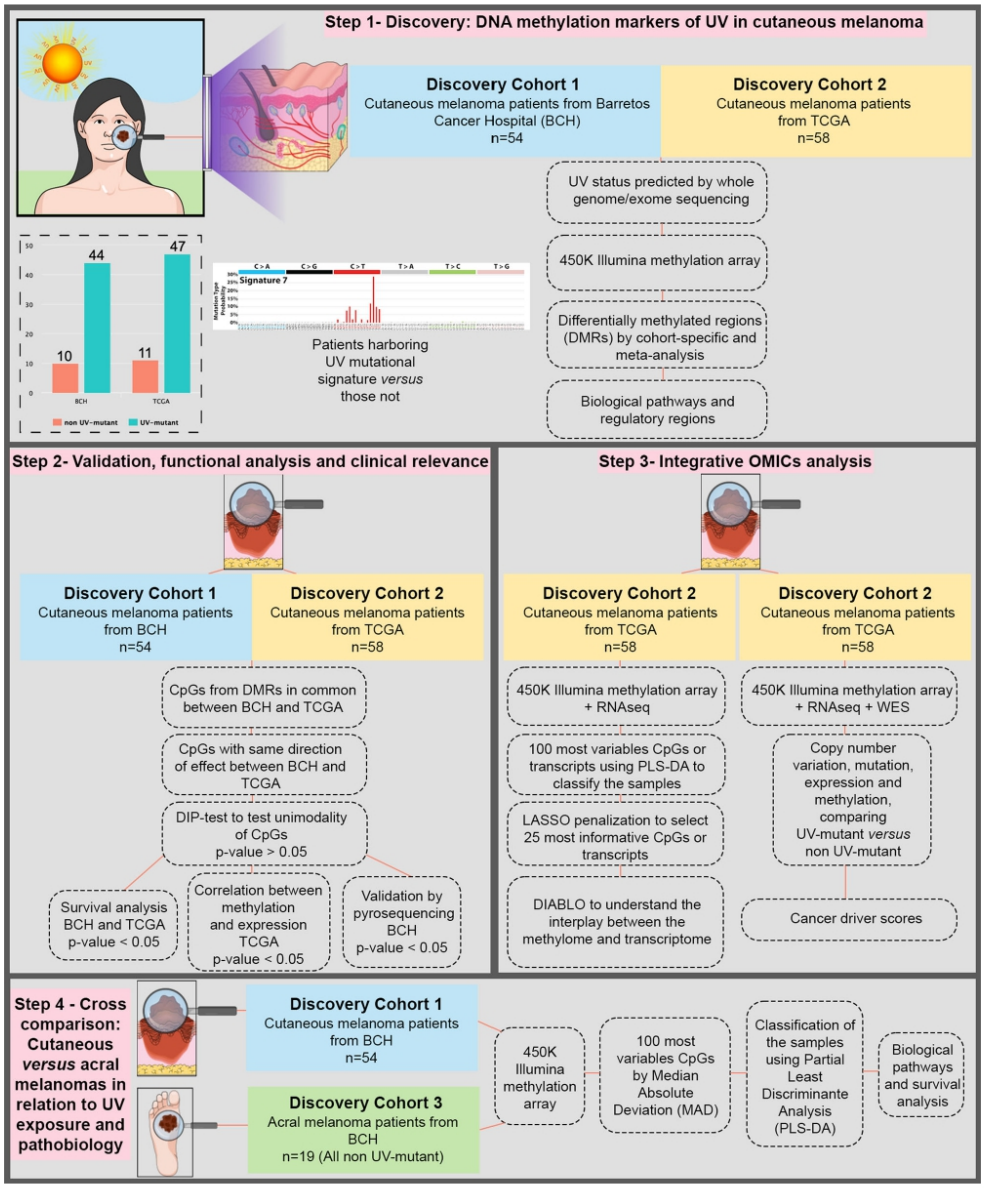


Fig. 1 Study design, resources and methodology. A major aim (steps 1–2) is the discovery and validation of genome-wide methylation alterations associated with the UV mutational signature in cutaneous melanoma, based on two independent cohorts. Another major aim (step 3) is assessing the discriminative potential of the DNA methylome *versus* transcriptome *versus* integrated methylome-transcriptome in differentiating between UV-mutant and non UV-mutant cutaneous melanomas. The integrative OMICs approach is expanded to include small nucleotide variants (SNVs) and copy number variants (CNVs) in order to assess cancer driver potential of prioritized differentially methylated genes. This is complemented by step 4, which investigates whether the DNA methylome could capture pathological and/or UV-related differences between major melanoma types predominantly associated with UV exposure (cutaneous melanoma) and those not (acral melanoma).

Table 1 Clinicopathological characteristics of the BCH-cutaneous melanoma patients, TCGA-cutaneous melanoma patients, and BCH-acral melanoma patients profiled with the DNA methylation array.

Characteristics	Cutaneous BCH (Discovery cohort 1)	Cutaneous TCGA (Discovery cohort 2)	Acral BCH (Discovery cohort 3)
	No. of patients (%)	No. of patients (%)	No. of patients (%)
Gender			
Male	33 (61.1)	31 (53.4)	13 (61.9)
Female	21 (38.9)	27 (46.6)	8 (38.1)
Skin§§§			
White	51 (94.4)	58 (100.0)	15 (71.4)
Pigmented	3 (5.6)	0 (0.0)	6 (28.6)
Tumor type**			
Primary	20 (37.0)	9 (15.5)	13 (61.9)
Superficial spreading	16 (65.0)	Missing	Not applicable
Nodular	7 (35.0)	Missing	Not applicable
Metastatic	34 (63.0)	49 (84.5)	8 (38.1)
Superficial spreading	16 (47.1)	Missing	Not applicable
Nodular	8 (23.5)	Missing	Not applicable
Missing	10 (29.4)	Missing	Not applicable
Breslow depth (mm)			
Up to 1.0	5 (9.2)	5 (8.6)	2 (9.5)
1.1–2.0	7 (13.0)	14 (24.1)	2 (9.5)
2.1–4.0	9 (16.7)	9 (15.5)	3 (14.3)
More than 4.0	22 (40.7)	15 (25.9)	14 (66.7)
Missing	11 (20.4)	15 (25.9)	Not applicable
TNM stage*			
In situ	0 (0.0)	0 (0.0)	0 (0.0)
I	7 (13.0)	8 (13.8)	1 (4.8)
II	16 (29.6)	18 (31.0)	8 (38.1)
III	14 (25.9)	19 (32.8)	10 (47.6)
IV	15 (27.8)	2 (3.4)	2 (9.5)
Missing	2 (3.7)	11 (18.0)	Not applicable
UV signature§§§			
No	10 (18.5)	11 (19.0)	17 (81.0)
Yes	44 (81.5)	47 (81.0)	4 (19.0)
Molecular group**§§§			
BRAF	39 (72.2)	23 (39.7)	4 (19.0)
RAS	6 (11.1)	20 (34.5)	6 (28.6)
NF1	2 (3.7)	6 (10.3)	1 (4.8)
TN	7 (13.0)	9 (15.5)	10 (47.6)
Age [Mean (SD)]	57.3 (17.0)	57.7 (14.7)	59.4 (12.5)

*, $^{\S}p < 0.05$; **, $^{\S\S}p < 0.01$; ***, $^{\S\S\S}p < 0.001$.
*Variables statistically different between BCH-cutaneous and TCGA-cutaneous melanoma patients using two-sided Chi-square test.
§Variables statistically different between BCH-cutaneous and BCH-acral melanoma patients using two-sided Chi-square test.

almost all non UV-mutants analyzed. A similar pattern was observed in the TCGA cohort. Cluster C1 was fully occupied by non UV-mutant samples and was visually distinct, exhibiting again an upper hypermethylation (red) stretch and a lower hypomethylation (blue) stretch, relative to the other clusters. Even if C1 is merged with the adjacent cluster C2, the non UV-mutant patients remain statistically enriched in this combined cluster ($p = 3.00e{-}10$), which now encompasses almost all non UV-mutants analyzed (Fig. 3a).

Recently, TCGA-cutaneous melanoma patients have been classified into four molecular mutation subgroups: 1—*BRAF*, associated with younger patients and with *BRAF* and *MITF* amplifications; 2—*RAS*, associated with MAPK activation and AKT3 overexpression; 3—*NF1*, associated with older patients and higher mutation burden; 4—The triple negative (TN), which is wild-type for *BRAF, RAS*, and *NF-1*, lacks the UV mutational signature and has higher copy number and complex rearrangements[9]. Indeed, we observed that the TN molecular subgroup is significantly enriched in non UV-mutant patients in both BCH and TCGA cohorts (Fig. 3a, b). We also observed that the *BRAF* mutant group was the most enriched in UV-mutant patients in both BCH and TCGA, though reaching statistical significance only in BCH (Fig. 3a, b). This was in line with other studies[15,16], further reinforcing the reproducibility potential of our data. Interestingly, *BRAF, NF1* and *RAS* were not significantly differentially methylated in melanoma tissues in relation to UV exposure (Supplementary Data 7, 8), highlighting that UV exposure produces DNA methylation changes in genes that can be different from critical ones mutationally altered by the same environmental exposure.

Jensen disease analysis of the filtered DMRs showed a significant implication of the differentially methylated genes in skin disorders, such as systemic scleroderma (BCH and TCGA), vitiligo (BCH and TCGA), melanoma (BCH), and skin cancer (BCH), particularly among the top and false discovery rate (FDR)-adjusted ontologies (FDR < 0.05) (Supplementary Data 9, 10). A number of other cancers and diseases were significantly enriched as well (Supplementary Data 9, 10). This was complemented by KEGG pathway analysis, revealing 28 and 30 significant pathways ($p < 0.05$) in BCH and TCGA, respectively. Among them, a large proportion (10 pathways) were identical between BCH and TCGA, 8 and 6 of which remained significant after adjustment for the number of CpGs per gene and FDR, respectively (Fig. 3c, Supplementary Data 11–14). These pathways constituted of differentially methylated genes implicated in immune system regulation: hematopoietic cell lineage, allograft rejection, graft-versus-host disease, intestinal immune network for IgA production, antigen-processing presentation, inflammatory bowel disease, and relatedly, autoimmune diseases, such as type 1 diabetes mellitus, autoimmune thyroid disease, systemic lupus erythematosus and rheumatoid arthritis (Fig. 3c).

The role of DNA methylation alterations in regulating immune system function was investigated in further depth and validated using RNA sequencing data (Methods section), demonstrating that immune cell composition was indeed different between UV-mutant and non UV-mutant cutaneous melanoma patients (Fig. 3d). Specifically, dendritic cells were significantly infiltrated in the non UV-mutant than in UV-mutant cutaneous melanoma (Fig. 3d, $p = 0.03$). Complementary analysis using differentially expressed genes comparing non UV-mutant and UV-mutant cutaneous melanoma patients ($p < 0.05$, Supplementary Data 15) also showed enrichment in immune disorders and skin-related diseases, though none reached FDR significance (Supplementary Data 16, 17).

DNA methylome markers of UV are prognostic of patient survival. In addition to the large proportion of overlap in biological pathways between BCH and TCGA described above, there was a significant overlap in DNA methylation alterations at the gene and CpG levels between BCH and TCGA (Fig. 4a). Out of the 458 CpGs from 169 genes significantly overlapping ($p = 2.3e{-}109$ and $p = 3.71e{-}29$, respectively) between BCH and TCGA cohorts (Supplementary Data 18, Fig. 4a), 6 CpGs (*HOXC9, KCNQ1DN* and *MGMT* genes) were hypermethylated and 30 CpGs (*TAPBP, ERICH3, FINL2, ZNF732, SLC6A18, MFSD13A, SLFN12L*, and *IFNLR1* genes) were hypomethylated in both BCH and TCGA, considering CpGs with absolute effect sizes ≥10% and with no significant enrichment in SNPs (Supplementary Fig. 1b, c).

We complemented the cohort-specific analyses with a DMR meta-analysis across the BCH and TCGA datasets (Fig. 4b). As the results demonstrate, there are 45,915 CpGs significantly differentially methylated across the two datasets between UV-mutant and non UV-mutant cutaneous melanomas (FDR < 0.05), of which a high proportion of CpGs (equal to 24,711 CpGs or equivalent to 53.8%) have the same direction of effect between

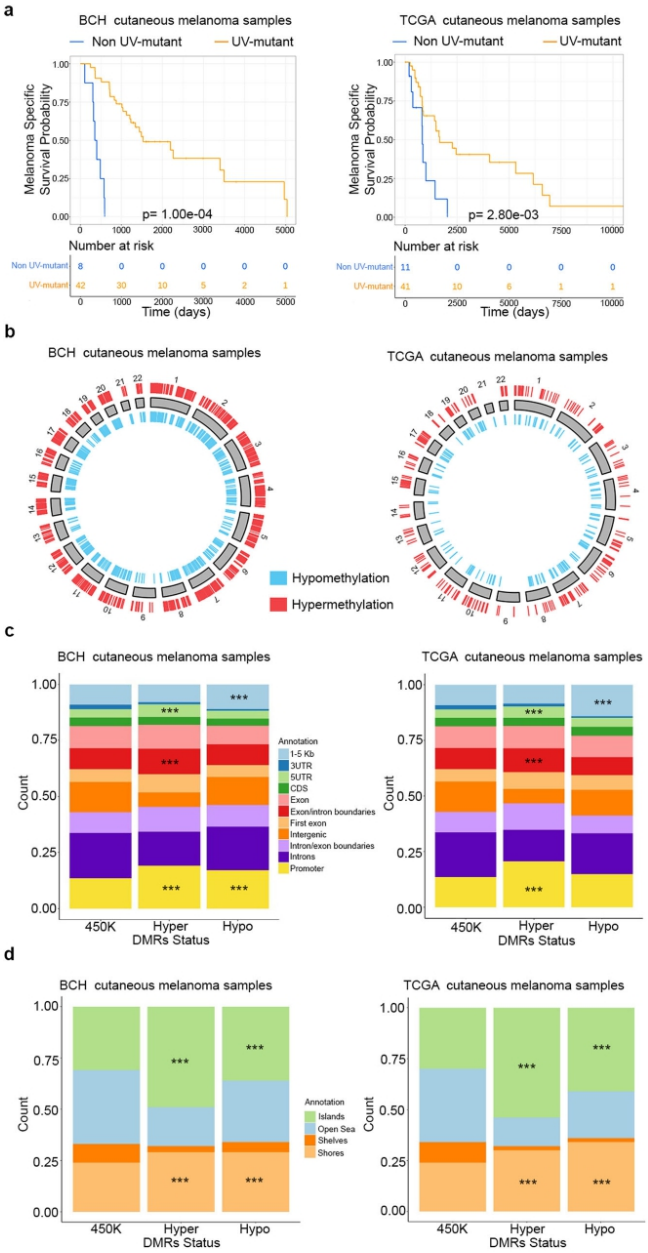


Fig. 2 Cross genome-methylome analysis of UV mutation signatures in cutaneous melanoma patients from BCH and TCGA cohorts. a Kaplan–Meier survival curves of melanoma patients by UV signature status in BCH ($n = 50$) and TCGA ($n = 52$). The *P* values were derived by log-rank test. Also shown are the DMR distributions from the crude model relative to chromosomal location (**b**), genomic regulatory regions (**c**), and CpG density regions (**d**) in both BCH ($n = 54$) and TCGA ($n = 58$). Enrichment analysis of hyper- and hypomethylated DMRs relative to the 450 K reference set in (**c**, **d**) was done using two-sided Chi-square test. $^{***}p < 0.001$.

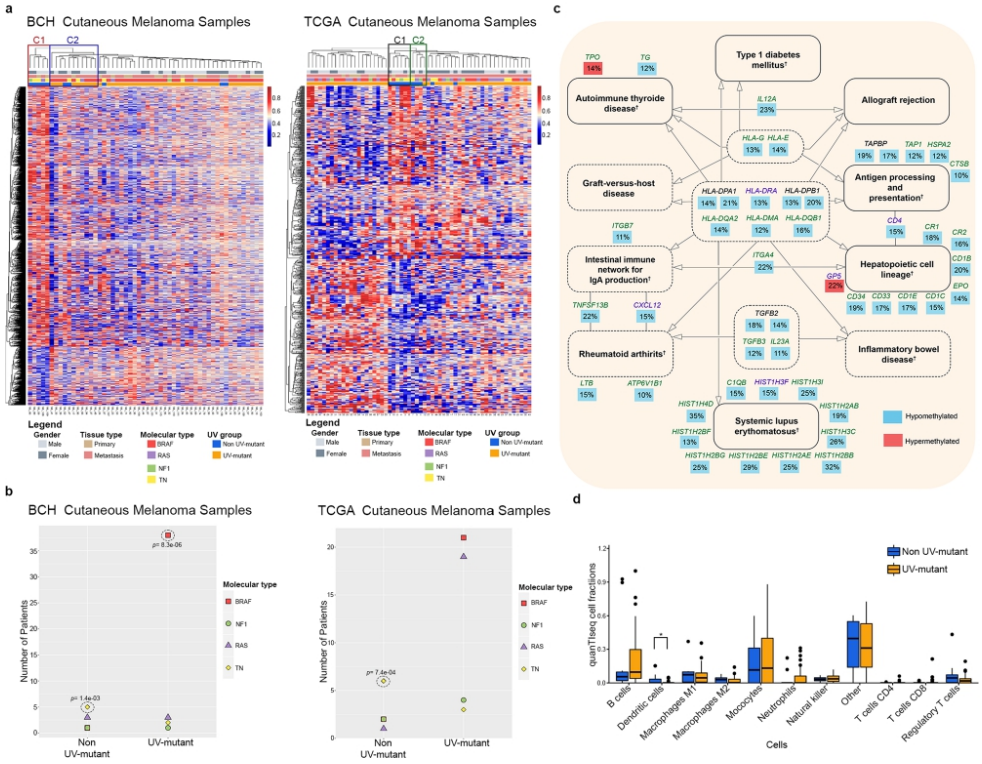
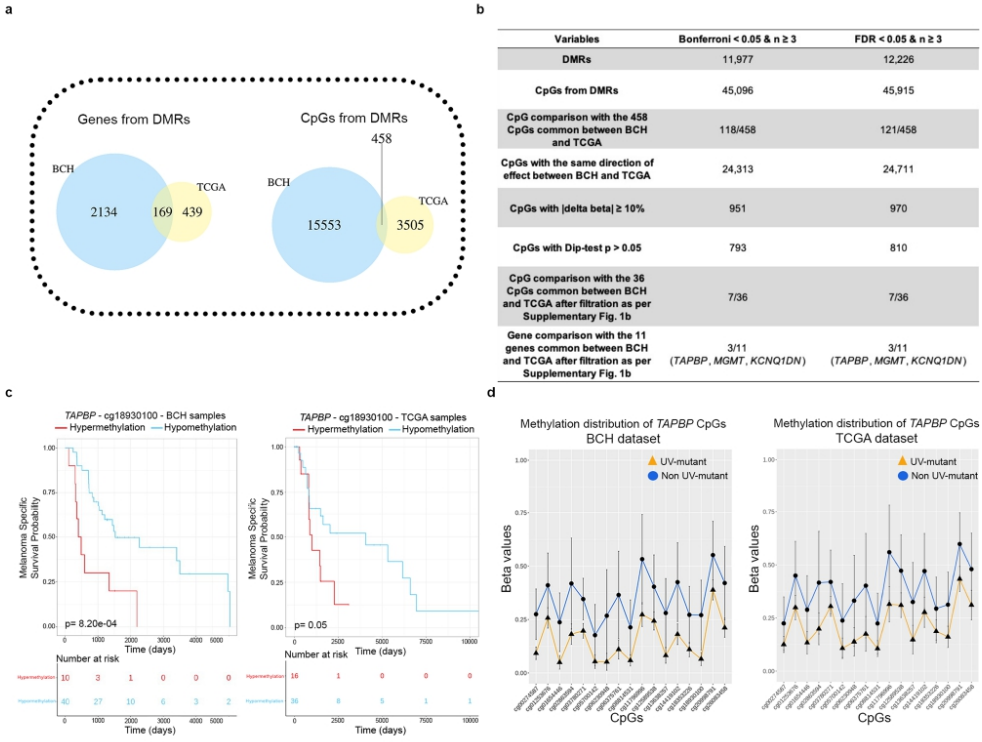


Fig. 3 The DNA methylome marks UV exposure and associated *BRAF/RAS/NF1* mutations, with effect on immunomodulation. a Hierarchical clustering of cutaneous melanoma patients in BCH and TCGA based on methylation levels of 4721 and 793 CpGs, respectively, as derived from Supplementary Fig. 1a. Enrichment analysis for non UV-mutant patients in clusters C1–C2 in each cohort was performed using two-sided Chi-square test while delimiting the cluster boundaries by the limits statistically specified by Euclidean distance. **b** Proportions of *BRAF, NF1, RAS*, and TN groups in UV-mutant and non UV-mutant melanomas. *P* values were derived by two-sided Fisher's exact test. **c** Common KEGG pathways between BCH and TCGA of genes differentially methylated between UV-mutant and non UV-mutant cutaneous melanoma patients, as derived from the prioritized CpGs in Supplementary Fig. 1a. Solid lines around the pathways' names indicate those with FDR < 0.05 in BCH and/or TCGA; whereas, dashed lines indicate those with $p < 0.05$. The percentage represents the average effect size across the CpGs of a given gene. Genes written in black are common between BCH and TCGA; whereas, green ones were found only in BCH and purple ones only in TCGA. †Pathways significant after adjustment for the number of CpGs associated with each gene. *P* value was delivered from two-sided Fisher exact test. **d** Immune cell composition inferred from RNA sequencing data comparing UV-mutant ($n = 47$) and non UV-mutant ($n = 11$) cutaneous melanoma patients from TCGA. Box center lines, bound of the box, and whiskers indicate medians, first and third quantiles, and minimum and maximum values within 1.5 × IQR (interquartile range) of the box limits, respectively. Each data point in the box plot represents the samples. *$p < 0.05$, by two-sided Mann–Whitney *U* Test.

BCH and TCGA (Supplementary Data 19, Fig. 4b). 121 meta-analysis CpGs (FDR < 0.05) overlapped with the 458 CpGs that are common between the BCH and TCGA cohort-specific analyses. As expected, the meta-analysis yields a larger number of significant hits (due to higher statistical power) than the cohort-specific analyses. However, the former is more prone to false positivity especially given some clinicopathological and ethnic dissimilarities (Table 1) and methodological differences between the two cohorts in inferring UV signature status (WGS versus WES, respectively). For this reason, (1) we additionally report the more stringent Bonferroni threshold, which yielded similar results as FDR (Supplementary Data 20, Fig. 4b), and (2) we present the meta-analysis results as a complementary method that reinforces the robustness of the findings across the different cohorts and analysis approaches, while prioritizing the more conservative cohort-specific analysis which yields signals that are common between BCH and TCGA and which, though less profuse, are less prone to error.

Thus, we further investigated whether the 36 CpGs in common between BCH and TCGA could be used to predict the survival of patients with cutaneous melanoma. Among them, cg06230948-*TAPBP*, cg18930100-*TAPBP*, cg19495013-*FIGNL2*, and cg26835312-*IFNLR1* were significantly associated with survival in BCH after adjustment for multiple testing (FDR < 0.05) (Supplementary Data 21). Among these four CpGs, cg18930100-*TAPBP* was also significantly associated with survival in a lookup analysis in TCGA. Specifically, patients in the low methylation groups at this CpG site had significantly higher melanoma-specific survival in both cohorts (Supplementary Data 21 and Fig. 4c). This was coherent with the hypomethylation at this CpG (Fig. 4d) and increased survival



Variables	Bonferroni < 0.05 & n ≥ 3	FDR < 0.05 & n ≥ 3
DMRs	11,977	12,226
CpGs from DMRs	45,096	45,915
CpG comparison with the 458 CpGs common between BCH and TCGA	118/458	121/458
CpGs with the same direction of effect between BCH and TCGA	24,313	24,711
CpGs with |delta beta| ≥ 10%	951	970
CpGs with Dip-test p > 0.05	793	810
CpG comparison with the 36 CpGs common between BCH and TCGA after filtration as per Supplementary Fig. 1b	7/36	7/36
Gene comparison with the 11 genes common between BCH and TCGA after filtration as per Supplementary Fig. 1b	3/11 (*TAPBP*, *MGMT*, *KCNQ1DN*)	3/11 (*TAPBP*, *MGMT*, *KCNQ1DN*)

Fig. 4 UV-related DNA methylome-wide alterations common between BCH and TCGA are prognostic for survival in cutaneous melanoma patients. **a** Venn diagrams showing that DMR-derived 169 genes or 458 CpGs are common between BCH and TCGA, based on the crude model. **b** DMR fixed effects inverse variance-weighted meta-analysis of BCH and TCGA, and comparison with the cohort-specific analysis. **c** Kaplan–Meier survival of melanoma patients in relation to methylation levels of cg18930100 (*TAPBP*) measured in the target tumors derived from BCH and TCGA. Patients were categorized into low- and high-methylation groups depending on whether the methylation value of a given CpG is lower or higher, respectively, than the mean methylation across the samples profiled for that CpG. *P* values were derived by two-sided log-rank test. **d** DNA methylation profiles of cg18930100 (*TAPBP*) that is associated with melanoma-specific survival, showing differential methylation between UV-mutant ($n = 40$ and 41 in BCH and TCGA, respectively) and non UV-mutant patients ($n = 10$ and 11 in BCH and TCGA, respectively). Data were expressed as the average values of each group (UV-mutant and non UV-mutant) for each single CpG with error bars indicating the 95% confidence interval (CI).

(Fig. 2a) observed in UV-mutant relative to non UV-mutant patients. Notably, *TAPBP* differential methylation is robustly significant in both the cohort-specific and meta-analyses of the BCH and TCGA cohorts (Fig. 4b and Supplementary Data 7, 8, 19, 20).

Validation and multi-OMICs functional roles of UV methylome markers. We next investigated the functional effect of UV-related DNA methylation alterations on gene expression using expression quantitative trait methylation (eQTM) analysis applied to DNA methylome and transcriptome data profiled on the same samples (Fig. 1). We first used a targeted approach focusing on cg18930100-*TAPBP* prioritized in the previous analysis (Fig. 4) and found that its methylation levels were significantly correlated with *TAPBP* RNA expression changes (Fig. 5a). We then performed eQTM analysis on all 458 CpGs that are common between TCGA and BCH (Supplementary Fig. 2a) in order to investigate whether the *TAPBP* gene could be still identified agnostically among the eQTMs. Out of the 458 CpGs, 10 (*TAPBP*: cg01253676, cg01654446, cg06230948, cg06375761, cg02863594, cg18930100, cg18353226; and *EIF2AK4*: cg20255370, cg16127683, cg01081584) were significantly correlated with expression, among which 7 CpGs were indeed located in the *TAPBP* gene, including its cg18930100 (Fig. 5a). All the significant correlations showed an inverse association between CpG methylation and RNA expression levels of each gene, with *TAPBP* showing hypomethylation while *EIF2AK4* showing hypermethylation in UV-mutant relative to non UV-mutant cutaneous melanoma patients (Fig. 5a). Notably, cg18930100 in *TAPBP* presented hypomethylation associated with both increased *TAPBP* RNA expression (Fig. 5a) and increased patient survival (Fig. 4c) in UV-mutant relative to non UV-mutant cutaneous melanoma patients. *TAPBP* and *EIF2AK4* RNA expression levels did not significantly associate with patient survival (Supplementary Fig. 2b), suggesting that their methylation levels may be stronger prognostic markers than their transcript levels.

Next, we pooled all 36 CpGs prioritized in Supplementary Fig. 1b (being common between BCH and TCGA) with the 10 CpGs prioritized in Supplementary Fig. 2a (being significant eQTMs) and investigated their cancer driver potential derived

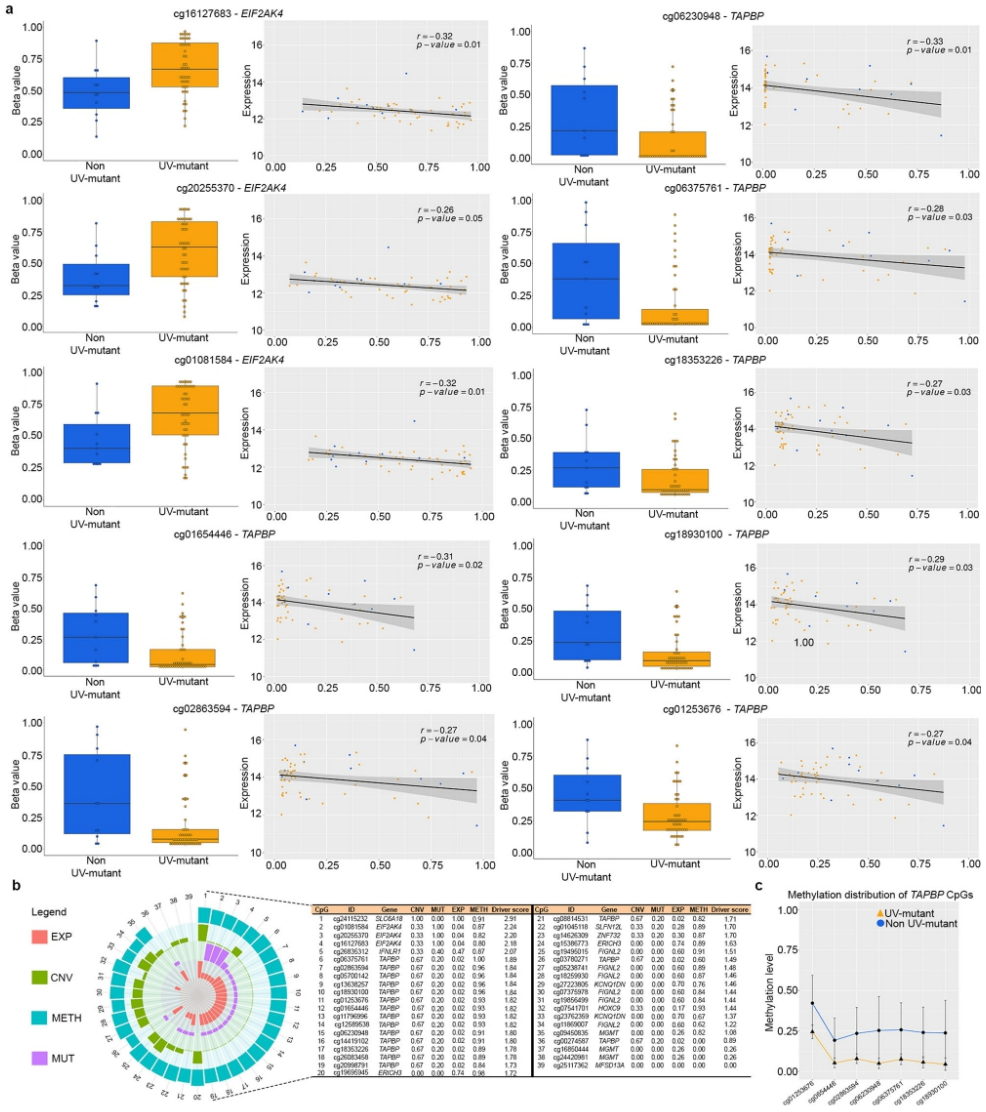


Fig. 5 Validation, eQTM, and multi-OMICs cancer driver analysis of UV methylome markers. a Pearson correlation was used to measure linear relationships between DNA methylation (Beta values) and gene expression levels measured in the same samples for the 10 selected CpGs (filtration step described in Supplementary Fig. 1a), using the TCGA dataset (UV-mutant = 11 and non UV-mutant = 47). Box center lines, bound of the box, and whiskers indicate medians, first and third quantiles, and minimum and maximum values within 1.5 × IQR (interquartile range) of the box limits, respectively. Each data point in the box plot represents the samples. The correlation *r* and *P* values were calculated by the two-sided correlation test and are shown for each CpG. **b** Multi-OMICs data integration from TCGA, encompassing copy number variation (CNV), expression (EXP), methylation (METH) and mutation (MUT), was performed in order to decipher the driver potential of the 12 prioritized genes (see Results) in cutaneous melanoma development following UV exposure. For each gene, scores of CNV, MUT, EXP, METH and multi-OMICs driver are indicated in the table and plotted in the associated circular diagram. **c** Validation of array-based DNA methylation by bisulfite pyrosequencing of the *TAPBP* gene in BCH samples (UV-mutant = 7 and non UV-mutant = 40). Data were expressed as the average values of each group (UV-mutant and non UV-mutant) for each single CpG with error bars indicating the 95% confidence interval.

from our recent multi-OMICs driver score[17]. This was performed using data on copy number variation, point mutations, RNA expression and DNA methylation profiled in cutaneous melanoma patients. We found that the top half of the CpGs with the highest cancer driver potential were largely predominated by CpGs of the *TAPBP* gene (Fig. 5b) and that this gene ranked among the top 4 driver genes when methylation levels were averaged across CpGs of a given gene (Supplementary Fig. 2c).

As a positive control, we used a list of genes known to play driver roles in cutaneous melanoma based on the Consensus-Driver score method (i.e., with Consensus-Driver >1.5)[18], which preferentially selects cancer driver genes that are frequently mutated in tumor tissues. We calculated the multi-OMICs driver scores for those genes, derived by measuring the extent of their OMICs alterations in UV-mutant relative to non UV-mutant melanomas (Supplementary Fig. 2d), as was done for the experimental gene set (Fig. 5b). We found that the multi-OMICs driver scores of the latter, including *TAPBP*, were predominantly in the same range as that of the positive control genes (1.24–2.50) (Supplementary Fig. 2d), reinforcing the cancer driver potential of the experimental gene set relative to known driver genes in melanoma.

Because of the biological and clinical relevance of *TAPBP* methylation, which was correlated with melanoma-specific survival and RNA expression and concurred with other genome-wide deregulations that led to its high multi-OMICs driver potential, we performed technical validation of *TAPBP* methylation using bisulfite pyrosequencing in the BCH cohort (Supplementary Data 22). Methylation by pyrosequencing validated that obtained with the methylome-wide array, confirming the observed *TAPBP* hypomethylation (including similar effect sizes and baseline methylation levels) in UV-mutant relative to non UV-mutant cutaneous melanoma (Fig. 5c).

Cutaneous and acral melanoma cross-OMICs: UV *versus* pathobiology. In addition to the genome- and methylome-wide analysis of UV exposure status in cutaneous melanoma, we next investigated whether the transcriptome landscape, taken alone or integrated with the methylome map, can better distinguish UV-mutant from non UV-mutant melanomas (Fig. 1). Based on PLS-DA modeling (Methods section), we observed that the DNA methylome alone predicts the two groups of patients (Fig. 6a—left panel) better than the transcriptome alone (Fig. 6a—right panel). The discriminative potential of the DNA methylome between UV-mutant and non UV-mutant cutaneous melanomas was sufficiently powerful, with slight or no improvement observed by the integrated methylome-transcriptome map (Fig. 6b and Supplementary Data 23) using LASSO coupled to DIABLO (Methods section). This complemented our earlier results showing that DNA methylation levels altered in UV-mutant melanomas are more prognostic to patient survival than the transcript levels of the corresponding genes (Fig. 4 and Supplementary Fig. 2b).

We further investigated whether the DNA methylome could also underlie differences between pathologically different melanomas, with interaction by UV mutation status, namely between melanoma types predominantly associated with UV exposure (cutaneous melanoma) and those not UV-associated (acral melanoma). Table 1 shows the clinical annotations of the acral samples collected at BCH. In contrast to cutaneous melanomas, out of 21 acral, only a few (19.0% compared to 81.5% in BCH-cutaneous melanoma, $p = 4.24e{-}07$) had the UV mutation signature as expected. The majority (47.6%) of the acral melanoma patients did not exhibit mutations in *BRAF*, *NRAS* or *NF1*, and a substantial portion (28.6% compared to 5.6% in BCH-cutaneous melanoma, $p = 2.28e{-}05$) presented a pigmented skin phenotype.

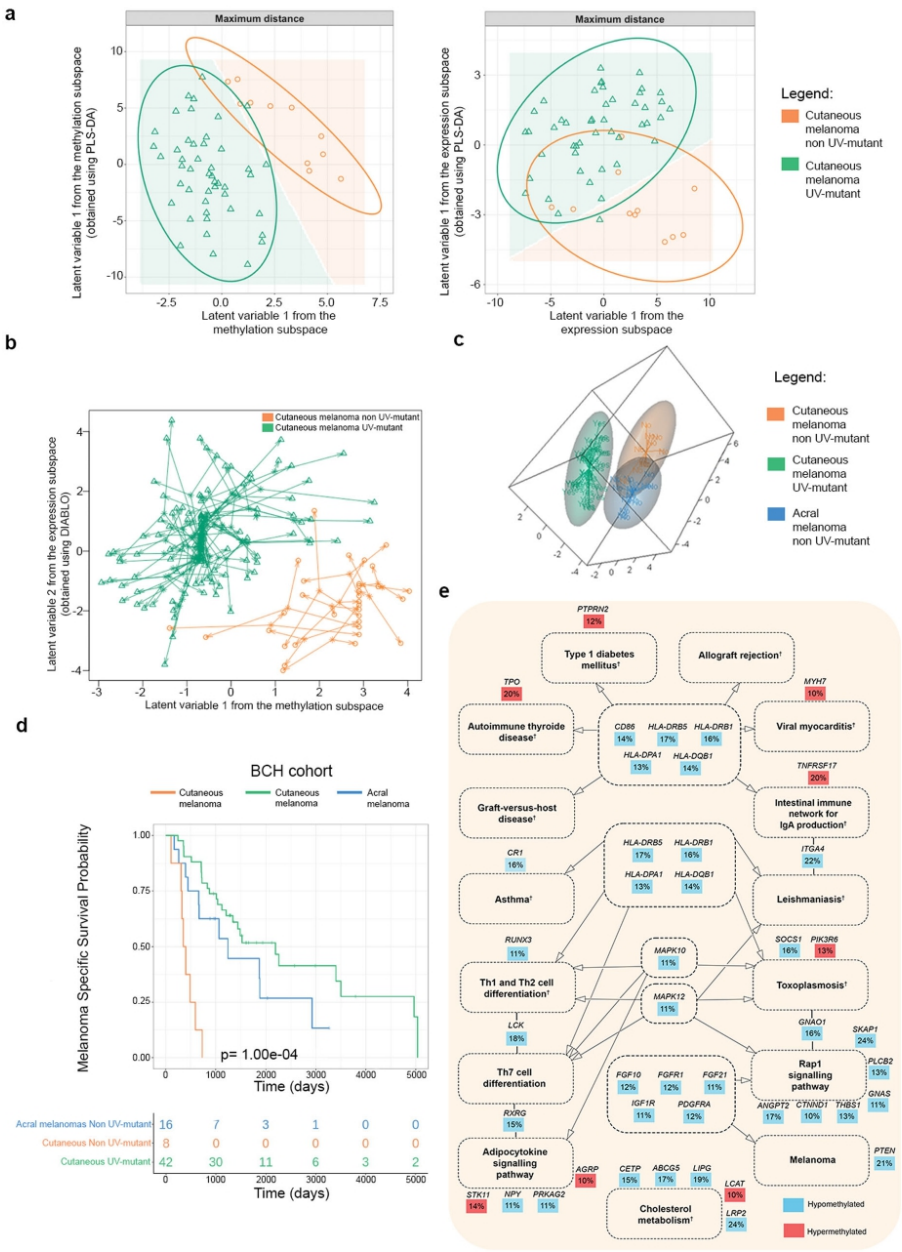
Based on PLS-DA modeling in BCH, we observed that the DNA methylome of the non UV-mutant cutaneous melanomas resembles more that of the pathologically different acral melanomas than the pathologically related UV-mutant cutaneous melanomas (Fig. 6c). This was in line with the survival analysis showing that the non UV-mutant cutaneous melanoma patients presented worse prognosis, more closely resembling that of acral melanoma patients (known to have poorer prognosis) rather than that of UV-mutant cutaneous melanoma patients (Fig. 6d, $p < 1.00e{-}04$).

The impact of tumor pathology on DNA methylome alterations was still observable, however, as evident by 1784 DMRs distinguishing non UV-mutant cutaneous and non UV-mutant acral melanomas (Supplementary Data 24). Jensen disease ontology and KEGG pathway analyses of the filtered DMRs (as described in Supplementary Fig. 2e and Supplementary Data 25) showed a significant implication ($p < 0.05$) of the differentially methylated genes in skin disorders (Supplementary Data 26) and immunological pathways (Fig. 6e, Supplementary Data 27, 28), respectively, but we interpret these results with caution as none of them remained significant after correcting for multiple testing (FDR > 0.05) (Supplementary Data 26–28).

Discussion

Melanoma is a type of skin cancer, which represents one of the most complex and heterogeneous cancers compared to other cancer types[19]. Although the positive association between UV exposure with melanoma development is well known, the underlying epigenetic mechanisms are largely unexplored in human melanoma tissues, as outlined by our systematic literature search (Supplementary Data 1). With the advent of new powerful technologies, such as WGS/WES, UV exposure status can now be inferred and analysed in human tissues. The present study investigated DNA methylome-wide alterations associated with UV mutation status in two cohorts of human cutaneous melanomas, with in-depth analysis of the functional and clinical implications of those alterations, including effects on regulatory regions, biological pathways, gene transcription, cancer driver potential, tumor classification and patient survival. This was complemented by testing whether the DNA methylome could also underlie differences between pathologically and molecularly different subtypes of melanomas with interaction by UV mutation status, namely between BRAF, RAS, NF1 and TN molecular groups and between melanoma types predominantly associated with UV exposure (cutaneous melanoma) and those not UV-associated (acral melanoma). To date, there is only one study that described the methylome landscape of acral melanomas[20], and our work additionally highlights genes and biological pathways that are epigenetically deregulated in this uncommon melanoma type in comparison with cutaneous melanoma analysed in the same cohort encompassing patients of European and Latin-American descents.

The only available melanoma dataset with methylome and genome data for replication of our BCH findings was from TCGA. The number of detectable signals was higher in the BCH relative to the TCGA cohort, and this is probably not due to statistical power differences as both datasets had similar sample sizes. This could be rather due in the BCH dataset to (1) the better quality of samples and/or their processing using our in-house optimized automated workflow to generate DNA methylome data, coupled to a priori designed sample distribution on the array that minimizes confounding with batch effects based on statistical semi-randomization, and (2) more accurate technology,



using WGS rather than WES, to assess UV exposure. Even though there were some clinicopathological and ethnic dissimilarities (Table 1) and methodological differences between the two cohorts in predicting UV signature status (WGS *versus* WES, respectively), we observed consistent findings in both at the CpG, gene and biological pathway levels. Moreover, UV-related DNA methylation alterations showed similar distributions between BCH and TCGA in hypo- and hypermethylated regions as well as similar enrichments in regulatory and CpG density regions, in skin disorders and in immunological pathways. Among the CpGs and genes differentially methylated in both BCH and TCGA, methylation levels of *TAPBP* (encompassing several differentially

Fig. 6 Comparative maps of the DNA methylome and transcriptome of cutaneous and acral melanomas in relation to UV mutational signatures, with associated patient survival and biological pathways. a PLS-DA modeling based on DNA methylome (left panel) or transcriptome (right panel) data derived from UV-mutant and non UV-mutant cutaneous melanoma from TCGA. **b** Diablo integrative analysis method with LASSO penalization were applied on methylome and transcriptome data from TCGA to select the most informative CpGs and transcripts that could improve the classification of the UV-mutant and non UV-mutant cutaneous melanomas. **c** Methylome matrices of acral (excluding the few UV-mutants), UV-mutant and non UV-mutant cutaneous melanomas based on the 100 most variables CpGs selected using median absolute deviation (MAD) and analysed with Partial Least Squares Discriminant Analysis (PLS-DA) in the BCH samples. **d** Melanoma-specific survival comparing the three groups of melanoma patients: cutaneous UV-mutant ($n = 42$), cutaneous non UV-mutant ($n = 8$) and acral non UV-mutant ($n = 16$). *P* value was delivered from log-rank test. **e** KEGG pathways of genes differentially methylated between acral *versus* cutaneous non UV-mutant melanomas in BCH, as derived from the prioritized CpGs in Supplementary Fig. 2a. Dashed lines around the pathways' names indicate those with $p < 0.05$; none were FDR-significant. The percentage represents the average effect size across the CpGs of a given gene. †Pathways significant after adjustment for the number of CpGs associated with each gene.

methylated CpGs) were significantly associated with RNA expression of this gene, concurred with other genome-wide deregulations to yield a high multi-OMICs driver potential and were significantly correlated with melanoma-specific survival. The array-based methylation results of *TAPBP* were independently validated by bisulfite pyrosequencing, further reinforcing the robustness of the findings and providing promising opportunities for clinical application via pyrosequencing as a cost-effective technique.

TAPBP is a member of the immunoglobulin superfamily, which mediates the interaction between newly assembled major histocompatibility complex class 1 (MHC-I) and the transporter associated with antigen processing[21]. Downregulation of TAPBP (tapasin) protein expression has been observed in multiple cancers as an immune escape mechanism of human tumors, which is restored after cytokine administration, indicating that deficient TAPBP expression might be due to dysregulation than to structural alterations[22]. Our findings show that *TAPBP* transcription is significantly inversely associated with its DNA methylation levels, and the latter are altered in relation to UV exposure rather than to melanoma pathological identity since this gene was not found to be differentially methylated in non UV-mutant cutaneous *versus* acral melanomas (Supplementary Data 25).

MHC-I complex expression on tumor cells has been described as an excellent surrogate marker of the overall tumor immunogenicity level as well as a predictor of response to immune checkpoint blockade therapy[23]. Moreover, MHC-I downregulation was identified as a common mechanism of resistance to PD-I inhibitor in melanoma clinical samples[24]. Restoring TAPBP expression can enhance MHC-I (HLA-B and -C) expression, as demonstrated in vitro, highlighting the possibility that patients with defects in MHC-I antigen-processing machinery may benefit from combining immunotherapeutic strategies with demethylating agents (such as those that could restore TAPBP expression)[25]. In complement with *TAPBP*, several HLA genes were also differentially methylated in UV-mutant *versus* non UV-mutant cutaneous melanoma in both BCH and TCGA, and these genes were centrally involved in the multiple immunological pathways identified (Fig. 3c). Taken together, *TAPBP* and MHC-I machinery genes, dysregulated by DNA methylation mechanisms as observed in our study, represent promising targets for epigenetic therapy and for predicting clinical response to immunotherapy.

TAPBP methylation significantly predicted patient prognosis in both BCH and TCGA. Even though the sample size of expression data was the same as that of methylation data, the association between TAPBP expression and survival was not significant. This suggests that the difference is not merely due to statistical power but could indeed reflect biological basis. DNA methylation does not act solely through affecting gene transcription but is known to also associate with chromosomal instability, the induction of splice variants, alterations in enhancer regions, changes in microRNA binding regions and expression control regions, and mutations[26–28]. Hence, DNA methylation may function as a prognostic marker per se or through these various non expression-related mechanisms. Our observation is in line with a multitude of studies highlighting the high sensitivity of the epigenome to exposure and risk factors[12,29].

The relation between *TAPBP* methylation and survival may not be necessarily causal. Our results, however, pinpoint to an increased likelihood of causality because (1) they were reproduced in two independent populations, including different ethnicities, which offer a natural means of effect randomization (hence, minimizing the influence of confounders), (2) they showed a dose-response (*TAPBP* hypomethylation was associated with increased survival relative to hypermethylation), and (3) they yielded a cancer driver potential for *TAPBP* that was comparable to that of known cancer driver genes. Still, more datasets will be needed to better reinforce the causality of the associations, for example, by using in larger sample sizes germline data as proxies for *TAPBP* methylation through Mendelian Randomization.

In addition to the findings focused on *TAPBP*, we reported that the UV mutational signature is associated with a high load (thousands) of epigenetic alterations affecting the methylome landscape of cutaneous melanoma. This highlights that, even though the UV-mutant and non UV-mutant cutaneous melanomas are supposed to share the same pathological/cellular origin, they may need to be classified separately, at least based on their underlying epigenomic landscape, which has the potential to capture markers of both exposures and cell identity. Moreover, the non UV-mutant cutaneous melanoma, by resembling in its epigenome the acral melanoma, may have a poorer prognosis and require a different therapeutic approach than the UV-exposed cutaneous melanoma. This is in line with our data showing that patient survival is worse for the non UV-mutant cutaneous and acral melanomas relative to the UV-mutant cutaneous melanomas. Our findings also corroborate those of another study showing that cutaneous melanoma patients harboring the UV mutation signature had higher disease-free and overall survival[11]. The consistency between findings is notable especially that our dataset included a different ethnic group and a more accurate methodology (WGS rather than WES) to infer the UV signature.

A recent study investigated the genetic changes related to UV in clinical melanoma samples and found 10 genes commonly mutated in UV-mutant relative to non UV-mutant cutaneous melanoma[11]. Among them, *PKHD1L1*, *LRP1B*, *ADGRV1* and *DNAH10* were hypomethylated in UV-mutant compared with non UV-mutant cutaneous melanoma patients in BCH (Supplementary Fig. 3a). Moreover, methylation of 6 CpGs of *LRP1B* were significantly associated with melanoma-specific survival (FDR < 0.05) (Supplementary Data 29); the most significant of those CpGs was cg02322989, the hypomethylation of which was associated with higher melanoma-specific survival (Supplementary Fig. 3b). The TCGA cohort did not corroborate the DMRs of these genes.

Consensus driver[30] and secondary driver genes have been recently described in cutaneous melanoma[31], among which several were differentially methylated in UV-mutant *versus* non UV-mutant cutaneous melanomas patients in our analysis (*COL5A1*, *DACH1*, *MECOM*, *PTEN*, *TP53*, *BRD9*, *BCL7*, *SPRED1*, *SIGLEC12*, and *SIGLEC10*) (Supplementary Data 2–6). These driver genes were derived mostly based on genomic data. We identified driver genes in melanoma and validated others already described, based on our multi-OMICs driver score encompassing genomics, transcriptomics and DNA methylome data. Our results suggest that genes differentially methylated in response to UV may play driver mechanisms in melanoma development.

In this work, we applied multiple powerful technologies, encompassing WGS, WES, RNA sequencing and DNA methylome-wide profiling, coupled to state-of-the-art bioinformatics tools onto a unique series of cutaneous and acral melanoma samples. Specifically, we leveraged publicly available data and complemented that with the generation of new datasets, with larger sample sizes, higher genomic coverage, more detailed phenotypic assessment, high-quality frozen tissue samples, and the inclusion of melanomas other than cutaneous and of ethnicities besides European-descent. In fact, less than 5% of genetic studies worldwide include participants with multiple ethnicity[32], specifically in acral melanoma research[33], and our work helps address this timely advocated need[32,33] by contributing to genomics and epigenomics data from populations of non-European descent. By investigating epigenetic markers of UV exposure in human melanoma tissues from two distinct populations and overlaying the DNA methylome landscape onto the transcriptome and genome maps of UV-mutant cutaneous relative to non UV-mutant cutaneous and acral melanomas, this work contributes to (1) uncovering potentially powerful exposure and cancer epigenetic biomarkers that can be exploited in risk stratification; (2) enhancing tumor classification within and across melanoma types; (3) revealing molecular drivers in melanomagenesis that could be at the origins of this cancer, hence, suitable for targeted therapy; and (4) diminishing population disparities and knowledge inequalities in melanoma pathobiology. The translational impact of the work covers common and less frequent melanomas and offers a roadmap guiding similar gene-environment investigations of other melanoma types.

Methods

Patient eligibility, biospecimen, and clinical data. The study was conducted according to the Brazilian national and institutional ethical policies, and it was previously approved by the Barretos Cancer Hospital Ethics Committee (716/2013). No compensation was provided to the participants in this study and informed consent was obtained by all participants included in BCH cohort. Patients were recruited at BCH in the context of the ICGC-Brazil project, which encompassed 100 melanoma patients prior to any systemic treatment and from whom paired tumor/blood tissues were profiled by WGS[34] and tumor tissues by 450 K DNA methylation array. We selected two subsets of patients from ICGC-Brazil (BCH cohort): first, Discovery cohort 1, encompassing 54 cutaneous melanomas patients harboring or not the UV mutation signature; second, Discovery cohort 3, constituting of 17 acral melanomas that are non UV-mutant (Fig. 1), after having excluded the 4 acral samples that were UV-mutant. All BCH samples were fresh frozen. Clinicopathological data were collected under ICGC guidelines. During the admission process at BCH, all patients self-report their skin type and ethnicity, and this information was extracted from medical records given the retrospective nature of the study. In addition, several studies were conducted on this patient population to determine their genetic-based ethnicity and correlate their ethnicity with clinical characteristics. These studies observed considerable admixture in the genetic composition[35–38].

The second cohort comprised 58 cutaneous melanoma samples from TCGA-SKCM for which information about UV signature was available (Fig. 1) based on WES. We excluded formalin fixed paraffin embedded samples and selected only fresh frozen samples for best quality of data and to eliminate sample processing bias in our comparisons with BCH samples. Clinicopathological data were downloaded from the TCGA-SKCM published study[9].

DNA isolation. DNA from fresh frozen BCH-cohort samples were isolated using the DNA Mini Qiasymphony kit (Qiagen catalog no 937236) following BCH Biobank procedures and the manufacturer's instructions[39]. Briefly, ~25 mg tissue in 180 µL ATL Buffer was homogenized (Precellys, Bertin-instruments) at 6500 1 × 10/10 s for three times. After samples were centrifuged for 1 min at 2867 × *g*, supernatants were transferred for another tube, 25 µL of Proteinase K were added per sample, and samples were incubated at 56 °C, 134 × *g* for 3 h. Then 4 µL RNase were added per sample and DNA was isolated using QIAsymphony (Qiagen catalog no 9001297).

Whole genome and exome sequencing and inference of UV mutational signatures. The WGS library construction and sequencing of BCH samples were performed at Mendelics (São Paulo, SP, Brazil). A total amount of one µg of each matched normal and tumor DNA was submitted to sonication fragmentation and further library preparation by Illumina TruSeq DNA PCR-Free Library Preparation kit (Illumina catalog no 20015963) using the 350 bp protocol. Libraries were quantified by Qubit Fluorometer (Thermofisher catalog no Q33238) and qualified by 2100 Bioanalyzer (Agilent catalog no G2939BA). The sequencing was carried out using Illumina HiSeq 2500 by paired-end strategy at a minimum of 30X coverage. WES data of TCGA samples was available from GDC Legacy Portal[9]. Molecular subgroups were defined by investigating somatic single-nucleotide mutations in *BRAF* hotspot, *RAS* hotspot and *NF1* throughout Mutect[40] algorithm and further annotated using Annovar[41]. The TN molecular subgroup denoted melanoma patients who did not harbor mutations in any of the three genes[34].

The UV mutational signature identification was performed using the SomaticSignatures Bioconductor package[42]. We used the Non-negative Matrix Factorization (NMF) algorithm[43] to determine the consensus signatures among the 71 patients. At the moment of the analyses, we used the 21 signatures[10] that were available and identified a consensus signature with more than 0.8 cosine similarity. For both BCH and TCGA cohorts, we classified samples as harboring an UV mutation signature (Cosmic Signature 7) based on the recommended criteria in which C > T transitions at dipyrimidine sites accounted for more than 60% or CC > TT mutations more than 5% of the total mutation burden[9].

Bisulfite conversion. The isolated DNA (500 ng) from BCH-cohort was bisulfite-modified using the EZ DNA Methylation Kit (Zymo Research catalog no #ZD5004) following the manufacturer's instructions for Illumina Infinium 450 K beadchip assay. Modified DNA was stored at −20 °C when short intervals were required between bisulfite conversion and further processing, and at −80 °C for long-term storage.

450K DNA methylome-wide array and analysis. The 450K data of BCH were generated in-house, and those of TCGA were downloaded from the GDC Legacy Portal[9]. For BCH, bisulfite converted DNA samples were profiled using 450K (Illumina catalog no WG-314-1003) and a well-established workflow optimized at IARC for high-throughput analyses through an automated robotic system (Freedom EVO 150 by Teca) that can process the chips with minimal human error. Chips are scanned using Illumina iScan to produce two-color raw data files (.idat format). Sample allocation to the arrays was based on a semi-randomization design that ensures minimum confounding by technical variation and minimizes the masking of biological covariates of interest by batch effects.

For the bioinformatics pre-processing, IDAT files from both cohorts were imported and processed using R software. Quality-control graphs and bimodal distributions for each dataset are shown in Supplementary Figs. 4, 5. We excluded cross-reactive probes and XY chromosomes, leaving a total of 459,761, 459,770, and 459,768 probes for the analysis in BCH-Cutaneous, BCH-Acral and TCGA cohort, respectively. The data were further normalized using the FunNorm function of the Bioconductor Minfi package[44] (Supplementary Fig. 4), which was shown to perform equally good or outperform existing normalization methods[45]. Inferred beta values were used to predict sex as a quality-control step using the Minfi function getSex. All samples were correctly predicted. The DNA methylation level β-values were logit transformed to M-values to map the range (0, 1) to (−inf, +inf) as it is more suitable for running regressions. Surrogate variable analysis (SVA) was performed on the methylome data to correct for potential batch effects, to adjust for differences in cell type composition as a reference-free method[46], and to adjust for latent variables, a choice validated by the findings of our benchmarking[47]. SVA also increases statistical power by removing (unwanted) variability through aggregating information at the data level and constraining the data's variability to the phenotype of interest[48].

For the statistical analysis, we used robust linear regression (robust to outliers) to test four statistical models in the discovery cohort 1 (BCH), including one crude and three adjusted models, comparing the DNA methylome of UV-mutant *versus* non UV-mutant cutaneous melanoma patients (Supplementary Figs. 5, 6): 1—Crude Model (Supplementary Data 2); 2—Adjusted Model 1 adjusted for sex (Supplementary Data 3); 3—Adjusted Model 2, adjusted for sex + age at diagnosis (Supplementary Data 4); and 4—Adjusted Model 3, adjusted for sex + age at diagnosis + tumor type (primary or metastasis) (Supplementary Data 5). Supplementary Fig. 5 shows quantile-quantile (Q–Q) plots of $-\log_{10}$P values, which deviate from their expected values under the null hypothesis across all models. Although the adjusted models yielded a larger number of significant findings relative to the crude model (Supplementary Fig. 6), we preferred to take a conservative approach and focus our analysis on the crude model, especially that it

showed the least genomic inflation and risk of false positives, with a lambda of 1.20 (i.e., approaching to 1.0 being the no inflation limit) (Supplementary Fig. 5). Moreover, the predominant proportion of significant findings in the crude model was actually common with any of the adjusted models (Supplementary Fig. 6b). We also compared for each model, two approaches of methylome-wide analysis: the Differentially Methylated CpG Probes (DMPs), analysing individual CpGs using the Bioconductor limma package[49], and the DMRs, analysing regions of genomically proximal CpGs using the Bioconductor DMRcate[50] package with the default proximity-based criteria (±1000 base pairs). At least 90% of DMP-based genes overlapped with those derived from DMRs across all models (Supplementary Fig. 6a). For this reason and because DMR analysis represents a dimension reduction approach with higher statistical power than DMP analysis, we focused downstream analyses in BCH and TCGA data onto the DMR approach applied to the crude model. This pipeline was equally applied to the DNA methylome comparison between non UV-mutant acral and non UV-mutant cutaneous melanomas. Statistically significant DMPs and DMRs were defined as those with FDR-adjusted P value < 0.05.

We complemented the cohort-specific analysis by a meta-analysis across the BCH and TCGA cohorts comparing UV-mutant *versus* non UV-mutant cutaneous melanoma patients. We used the Metal tool[51] and the Dmrff package in R to perform DMR fixed effects inverse variance-weighted meta-analysis[52], using the crude model as prioritized in the cohort-specific analysis. The meta-analysis lambda value was 1.16, showing low inflation. Statistically significant DMRs were defined as those with FDR-adjusted P value < 0.05. Due to the larger number of hits expected with the increased statistical power afforded by the meta-analyses, we also reported the more stringent Bonferroni-adjusted p values, especially considering the higher likelihood of false positivity due to clinicopathological, ethnic and methodological differences between meta-analysed BCH and TCGA.

In addition to generating DMPs and DMRs, methylation data from BCH and TCGA samples were further investigated using Partial Least Squares Discriminant Analysis (PLS-DA)[53]. This approach performs classification of samples using partial least squares regression of the categorical outcome Y (cancer subtype) on the predictor variables (DNA methylation). PLS-DA is a clustering technique that allows the quantification of the discrimination relevance of a given variable (CpG) and to predict the phenotype of new samples (independent of DMPs or DMRs). This method is especially suited to deal with a much larger number of variables than samples, as in next-generation microarray and sequencing data, and we aided this method further by a filtering step using median absolute deviation (MAD)[54]. We selected the 100 most variable CpGs and applied PLS-DA on the methylation matrix on this subset of sites to assess the discriminative potential of the DNA methylome between UV-mutant and non UV-mutant cutaneous melanomas (as well as acral melanomas in the case of BCH samples).

Pyrosequencing methylation analysis. For the quantitative measurement of DNA methylation levels in individual CpG sites of the *TAPBP* (7 CpGs) gene (Supplementary Data 22), we pyrosequenced the bisulfite converted DNA using the PyroMark Kit (Qiagen catalog no 978703) as per the manufacturer's instruction. Briefly, DNA was immobilized onto streptavidin-coated beads in binding buffer for 10 min. The biotin-labeled PCR template was isolated and denaturated using the pyrosequencing vacuum prep tool and incubated with 0.4 µM sequencing primer in annealing buffer (20 mM Tris-acetate, 2 mM MgAc2; pH 7.6). The reaction was incubated at 80 °C for 2 min and cooled down to room temperature for 20 min to allow sequencing primer annealing. The methylation levels at the target CpGs were evaluated by converting the resulting pyrograms to numerical values for peak heights and expressed as the average of all patients for a given CpG site analyzed.

RNA expression data and analysis. Transcriptome data, measured by RNA sequencing (RNAseq), were downloaded from TCGA-SKCM project[9] and normalized with DESeq package[55]. As with the DNA methylation data, the normalized RNAseq data was first filtered by MAD for the 100 most variable transcripts and then analysed by PLS-DA to assess the discriminative potential of the transcriptome between UV-mutant and non UV-mutant cutaneous melanomas. We used quanTiseq package[56] to estimate the fractions of ten immune cell types using the RNAseq from TCGA-SKCM project[9], comparing UV and non UV-mutant melanoma patients. Then Mann–Whitney U Test was performed to compare the two conditions as this non-parametric test is robust to outliers, which were detected in some data points of the various cell types.

Gene ontology and pathway analysis. Gene ontology and pathway analysis were performed using the Jensen Disease ontology and KEGG pathway databases available on Enrichr website[57,58] (https://maayanlab.cloud/Enrichr). Given that genes with larger numbers of probes are more likely to have significantly differentially methylated CpGs, potentially biasing gene set analysis, we implemented the *gometh* function of the missMethyl package[59] in R to adjust for the number of CpGs per gene, which ranges on the 450K array from 1 to 1299 CpGs.

Cross-OMICs and integrative analysis. Regarding eQTMs, we applied Pearson correlation between RNA expression and DNA methylation data of the 169 genes that are common between BCH and TCGA DMRs, while limiting the analysis of a given gene to its constituent CpGs.

For the integrated methylome-transcriptome analysis, we filtered each of the 450K and RNAseq datasets by MAD and applied PLS-DA independently to each OMICs, as described in previous sections. Next, we applied sparse PLS-DA[60] that uses LASSO[61] penalization technique to select the 25 most informative CpGs and transcript probes in each dataset (Supplementary Data 23). We then applied an integrative analysis on the subsets of methylation and transcriptomic data, together with UV exposure outcome. We used the DIABLO method to gain a better understanding of the interplay between the different levels of data that are measured[62]. All these analyses were done using mixOmics R package[63].

For the prediction of multi-OMICs driver score, genome, transcriptome and DNA methylome data were downloaded from TCGA-SKCM (473 cases)[9]. Then, we calculated for each gene: [CNV: Copy Number Variation], being the number of tumors affected by a deep insertion ($\geq +2$) or a deep deletion (≤ -2), and this number was divided by the maximum CNV value obtained across the analysed genes in order to generate the CNV score; [MUT: Mutation], being the number of tumors affected by at least one single-nucleotide alteration, and this number was divided by the maximum MUT value obtained across the analysed genes in order to generate the MUT score; [EXP: Expression], being the number of UV-mutant cases presenting variations in RNA expression ($|\log FC| > 2$) relative to non UV-mutant patients, and this number was divided by the maximum EXP value obtained across the analysed genes in order to generate the EXP score; and [METH: Methylation], being the number of UV-mutant cases presenting variations in methylation ($|\text{delta betas}| > 0.1$) relative to non UV-mutant patients, and this number was divided by the maximum METH value obtained across the analysed genes in order to generate the METH score. The driver score for each gene was then calculated as the sum of these four proportions, representing a derivation of our recently reported cancer driver score[17] by additionally including DNA methylation data on top of genomic and transcriptomic data.

Power estimates. The OMIC with the largest dimension (i.e., involving an agnostic approach with a multitude of statistical tests) would require the largest number of samples analysed to maintain a high statistical power. In this work, it would be the DNA methylome (~450,000 tests) followed by the transcriptome (~20,000 tests). Although WGS has a larger dimension than either, it is not being used agnostically in this work, but rather to screen for specific mutational signatures or mutated genes known to be genetically altered in melanoma. Accordingly, statistical power is estimated based on the methylome as such: the overall mean standard deviation (SD) of methylation probes in the BCH or TCGA data is 0.11 (for methylation values ranging 0-1). Given an effect size ≥10% methylation difference (a threshold used in our prioritization filters as reported in Supplementary Figs. 1a, b, 2a) and based on an alpha of 0.05, we will have >80% power with at least 20 exposed cases and 20 controls. Our sample size is larger and encompasses 21 UV-mutant cutaneous and 91 non UV-mutant cutaneous melanomas from BCH and TCGA. In addition to single-OMIC analysis, we performed integrative OMICs analysis, which can depict small effects shared between OMICs and not detected in the individual analyses and, hence, could be performed on smaller sample sizes than single-OMIC analysis. A recent study also proposes a joint power method for all OMICs being integrated[64]; however, we preferred to estimate the power based on the OMICs with the largest dimension as a more conservative approach. Statistical power was further enhanced by implementing dimension reduction and SVA approaches (as described in the 450K analysis), and the false positive likelihood was reduced by monitory and correcting for potential inflation, by adjusting for multiple-testing, and by replication of findings in two independent cohorts as well as by two different techniques (array- and pyrosequencing-based).

Other statistical analyses. Enrichment analyses was done using Chi-Square test or, when sample sizes were small, Fisher's exact test as proposed by R. For Kaplan–Meier melanoma-specific survival analyses, methylation data were dichotomized using the mean methylation level as cut-off, and log-rank testing was used to evaluate differences between curves[65]. The various plots in the manuscript were generated using ggplot2 package[66], except for the heatmaps, which were generated using Heatmap plus package. All analyses were performed on R. P values ≤ 0.05 were considered statistically significant. Adjustment for multiple testing was based on FDR < 0.05.

Systematic literature search. We performed a systematic literature search on PubMed to select papers published until May 2021 that analyzed DNA methylome-wide data in clinical melanoma samples. To this end, we used the following syntax: ((melanoma) OR (melanomas)) AND ((Global DNA methylation) OR (methylome) OR (methylome-wide) OR (DNA methylation)). In total, we found 867 studies of which 20 (Supplementary Data 1) were included in our analysis since they covered methylome-wide profiling rather than targeted DNA methylation assays, and they were also conducted on clinical samples rather than cell lines or animal models.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The 450K data generated in this study have been deposited in the GE0 database under accession code GSE202750. The WGS data are available in the ICGC database [https://dcc.icgc.org/projects/SKCA-BR]. The 450K, RNAseq and WES data on melanoma samples from TCGA were downloaded from the GDC Legacy Portal [https://portal.gdc.cancer.gov/legacy-archive/search/f] and cBioPortal [https://www.cbioportal.org/datasets]. All other data is available within the Supplementary Data. Source data are provided with this paper.

Code availability

Bioinformatics pipelines used in this study are available in [https://zenodo.org/record/6530343#.Ynuzki-tFTY][67] under the DOI number: https://doi.org/10.5281/zenodo.6530343.

Received: 16 October 2021; Accepted: 16 June 2022;
Published online: 15 July 2022

References

1. Uong, A. & Zon, L. I. Melanocytes in development and cancer. *J. Cell Physiol.* **222**, 38–41 (2010).
2. Tsao, H., Chin, L., Garraway, L. A. & Fisher, D. E. Melanoma: from mutations to medicine. *Genes Dev.* **26**, 1131–1155 (2012).
3. Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
4. Guy, G. P. Jr. et al. Vital signs: melanoma incidence and mortality trends and projections—United States, 1982–2030. *MMWR Morb. Mortal. Wkly Rep.* **64**, 591–596 (2015).
5. Boniol, M., Autier, P., Boyle, P. & Gandini, S. Cutaneous melanoma attributable to sunbed use: systematic review and meta-analysis. *BMJ* **345**, e4757 (2012).
6. Clark, W. H. Jr., From, L., Bernardino, E. A. & Mihm, M. C. The histogenesis and biologic behavior of primary human malignant melanomas of the skin. *Cancer Res* **29**, 705–727 (1969).
7. Bernardes, S. S. et al. More than just acral melanoma: the controversies of defining the disease. *J. Pathol. Clin. Res.* **7**, 531–541 (2021).
8. Bradford, P. T., Goldstein, A. M., McMaster, M. L. & Tucker, M. A. Acral lentiginous melanoma: incidence and survival patterns in the United States, 1986–2005. *Arch. Dermatol.* **145**, 427–434 (2009).
9. Cancer Genome Atlas, N. Genomic classification of cutaneous melanoma. *Cell* **161**, 1681–1696 (2015).
10. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
11. Trucco, L. D. et al. Ultraviolet radiation-induced DNA damage is prognostic for outcome in melanoma. *Nat. Med.* **25**, 221–224 (2019).
12. Herceg, Z. et al. Roadmap for investigating epigenome deregulation and environmental origins of cancer. *Int J. Cancer* **142**, 874–882 (2018).
13. Ikehata, H. & Ono, T. Significance of CpG methylation for solar UV-induced mutagenesis and carcinogenesis in skin. *Photochem. Photobiol.* **83**, 196–204 (2007).
14. Elder, D. E., Bastian, B. C., Cree, I. A., Massi, D. & Scolyer, R. A. The 2018 world health organization classification of cutaneous, mucosal, and uveal melanoma: detailed analysis of 9 distinct subtypes defined by their evolutionary pathway. *Arch. Pathol. Lab. Med.* **144**, 500–522 (2020).
15. Kim, S. Y. et al. Metaanalysis of BRAF mutations and clinicopathologic characteristics in primary melanoma. *J. Am. Acad. Dermatol.* **72**, 1036–46.e2 (2015).
16. Lee, J. H., Choi, J. W. & Kim, Y. S. Frequencies of BRAF and NRAS mutations are different in histological types and sites of origin of cutaneous melanoma: a meta-analysis. *Br. J. Dermatol.* **164**, 776–784 (2011).
17. Halaburkova, A. et al. Pan-cancer multi-omics analysis and orthogonal experimental assessment of epigenetic driver genes. *Genome Res.* **30**, 1517–1532 (2020).
18. Bertrand, D. et al. ConsensusDriver improves upon individual algorithms for predicting driver alterations in different cancer types and individual patients. *Cancer Res.* **78**, 290–301 (2018).
19. Andor, N. et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat. Med* **22**, 105–113 (2016).
20. Pradhan, D. et al. Aberrant DNA methylation predicts melanoma-specific survival in patients with acral melanoma. *Cancers* **11**, 2031 (2019).
21. Ortmann, B. et al. A critical role for tapasin in the assembly and function of multimeric MHC class I-TAP complexes. *Science* **277**, 1306–1309 (1997).
22. Seliger, B. et al. Downregulation of the constitutive tapasin expression in human tumor cells of distinct origin and its transcriptional upregulation by cytokines. *Tissue Antigens* **57**, 39–45 (2001).
23. Lechner, M. G. et al. Immunogenicity of murine solid tumor models as a defining feature of in vivo behavior and response to immunotherapy. *J. Immunother.* **36**, 477–489 (2013).
24. Lee, J. H. et al. Transcriptional downregulation of MHC class I and melanoma de- differentiation in resistance to PD-1 inhibition. *Nat. Commun.* **11**, 1897 (2020).
25. Chang, C. C. et al. Multiple structural and epigenetic defects in the human leukocyte antigen class I antigen presentation pathway in a recurrent metastatic melanoma following immunotherapy. *J. Biol. Chem.* **290**, 26562–26575 (2015).
26. Narayanan, S. P., Singh, S. & Shukla, S. A saga of cancer epigenetics: linking epigenetics to alternative splicing. *Biochem. J.* **474**, 885–896 (2017).
27. Cho, J. W. et al. The importance of enhancer methylation for epigenetic regulation of tumorigenesis in squamous lung cancer. *Exp. Mol. Med.* **54**, 12–22 (2022).
28. Anwar, S. L. & Lehmann, U. DNA methylation, microRNAs, and their crosstalk as potential biomarkers in hepatocellular carcinoma. *World J. Gastroenterol.* **20**, 7894–7913 (2014).
29. Bowers, E. C. & McCullough, S. D. Linking the epigenome with exposure effects and susceptibility: the epigenetic seed and soil model. *Toxicol. Sci.* **155**, 302–314 (2017).
30. Bailey, M. H. et al. Comprehensive characterization of cancer driver genes and mutations. *Cell* **173**, 371–385.e18 (2018).
31. Conway, J. R. et al. Integrated molecular drivers coordinate biological and clinical states in melanoma. *Nat. Genet.* **52**, 1373–1383 (2020).
32. Fatumo, S. et al. A roadmap to increase diversity in genomic studies. *Nat. Med.* **28**, 243–250 (2022).
33. Alicea, G. M. & Rebecca, V. W. Un-Fair Skin: racial disparities in acral melanoma research. *Nat. Rev. Cancer* **22**, 127–128 (2022).
34. Vazquez, V. D. L. et al. Brazilian melanoma genome project: mutational landscape based on whole-genome sequencing. *Cancer Res.* **77**, 3386 (2017).
35. Duraes, R. O. et al. Role of genetic ancestry in 1,002 Brazilian colorectal cancer patients from Barretos Cancer Hospital. *Front. Oncol.* **10**, 145 (2020).
36. Leal, L. F. et al. Mutational profile of Brazilian lung adenocarcinoma unveils association of EGFR mutations with high Asian ancestry and independent prognostic role of KRAS mutations. *Sci. Rep.* **9**, 3209 (2019).
37. Fernandes, G. C. et al. Prevalence of BRCA1/BRCA2 mutations in a Brazilian population sample at-risk for hereditary breast cancer and characterization of its genetic ancestry. *Oncotarget* **7**, 80465–80481 (2016).
38. Berardinelli, G. N. et al. Association of microsatellite instability (MSI) status with the 5-year outcome and genetic ancestry in a large Brazilian cohort of colorectal cancer. *Eur. J. Hum. Genet.* **30**, 824–832 (2022).
39. Neuber, A. C. et al. The biobank of barretos cancer hospital: 14 years of experience in cancer research. *Cell Tissue Bank* **23**, 271–284 (2021).
40. Lawrence, M. S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
41. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
42. Gehring, J. S., Fischer, B., Lawrence, M. & Huber, W. SomaticSignatures: inferring mutational signatures from single-nucleotide variants. *Bioinformatics* **31**, 3673–3675 (2015).
43. Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **401**, 788–791 (1999).
44. Aryee, M. J. et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
45. Fortin, J. P. et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* **15**, 503 (2014).
46. Kaushal, A. et al. Comparison of different cell type correction methods for genome-scale epigenetics studies. *BMC Bioinform.* **18**, 216 (2017).
47. Perrier, F. et al. Identifying and correcting epigenetics measurements for systematic sources of variation. *Clin. Epigenetics* **10**, 38 (2018).
48. Lin, X., Barton, S. & Holbrook, J. D. How to make DNA methylome wide association studies more powerful. *Epigenomics* **8**, 1117–1129 (2016).
49. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
50. Peters, T. J. et al. De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin* **8**, 6 (2015).
51. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

52. Suderman, M. et al. dmrff: identifying differentially methylated regions efficiently with power and control. *BioRxiv* 508556 (2018).
53. Perez-Enciso, M. & Tenenhaus, M. Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. *Hum. Genet* **112**, 581–592 (2003).
54. Donoho, D. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, **41**, 613–627 (1995).
55. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
56. Finotello, F. et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med.* **11**, 34 (2019).
57. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.* **14**, 128 (2013).
58. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
59. Phipson, B., Maksimovic, J. & Oshlack, A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics* **32**, 286–288 (2016).
60. Le Cao, K. A., Boitard, S. & Besse, P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinform.* **12**, 253 (2011).
61. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc.* **58**, 267–288 (1996).
62. Singh, A. et al. DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics* **35**, 3055–3062 (2019).
63. Rohart, F., Gautier, B., Singh, A. & Le Cao, K. A. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* **13**, e1005752 (2017).
64. Tarazona, S. et al. Harmonization of quality metrics and power calculation in multi-omic studies. *Nat. Commun.* **11**, 3092 (2020).
65. Terry Therneau, P. G. *Modeling Survival Data: Extending the Cox Model* (Springer, 2000).
66. Wickham, H. *ggplot2: Elegant graphics for data analysis* (Springer-Verlag, 2009).
67. Cahais, V., Ghantous, A. & Zdenko, H. IARC pipeline for methylome analysis. (Zenodo, 2022).

Acknowledgements

We thank all melanoma patients who contributed to the study. This work was supported by Public Ministry of Labor, Campinas (Research, Prevention, and Education of Occupational Cancer), Barretos Cancer Hospital and IARC. A.L.S.A.V. was supported by FAPESP (grant no 2016/15941-3 and 2017/09612-0). We thank the BCH biobank for DNA isolation and sample processing and storage. The initial development of the multi-OMICs driver score was partially supported by the grant from the Institut National du Cancer (INCa, France), La direction générale de l'offre de soins (DGOS), and INSERM (SIRIC LYriCAN, INCa-DGOS-Inserm_12563) to Z.H. Where authors are identified as personnel of the IARC/WHO, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the IARC/WHO.

Author contributions

Conceptualization: A.L.S.A.V., V.L.V., and A.G. Patient recruitment and clinical data collection in BCH: C.S.C., A.L.C., and V.L.V. Methylome array design: A.L.S.A.V. and A.G. Generation of array and pyrosequencing methylation data: C.C. Analysis of OMICs data: A.L.S.A.V., A.N., V.C., Z.A., N.S., A.F.E., and A.G. Data interpretation: A.L.S.A.V., Z.H., R.M.R., V.L.V., and A.G. Supervision: A.G., Z.H., and V.L.V. Writing original draft: A.L.S.A.V., V.L.V., and A.G. Editing and reviewing the paper: all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-31488-w.

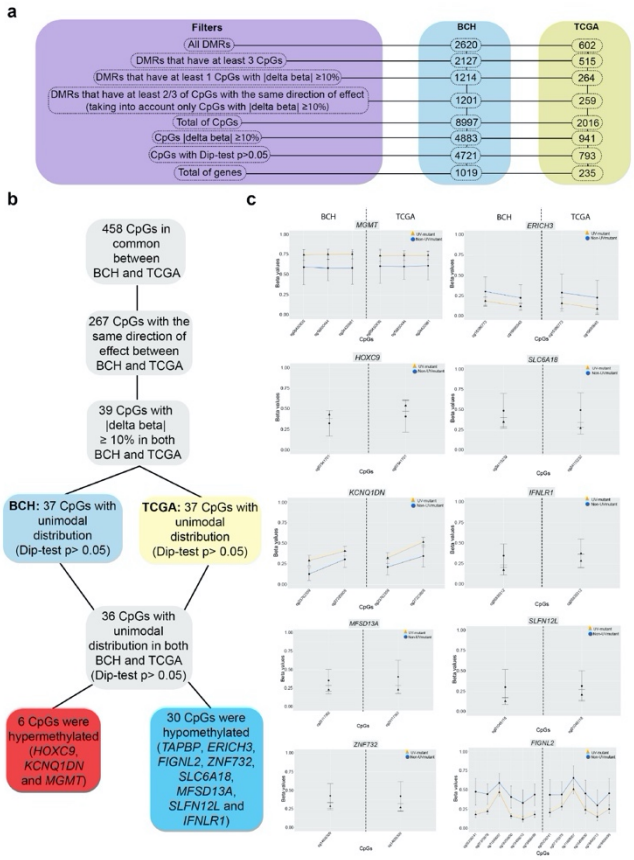
Correspondence and requests for materials should be addressed to Anna Luiza Silva Almeida Vicente or Akram Ghantous.

Peer review information *Nature Communications* thanks Carla Robles-Espinoza and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

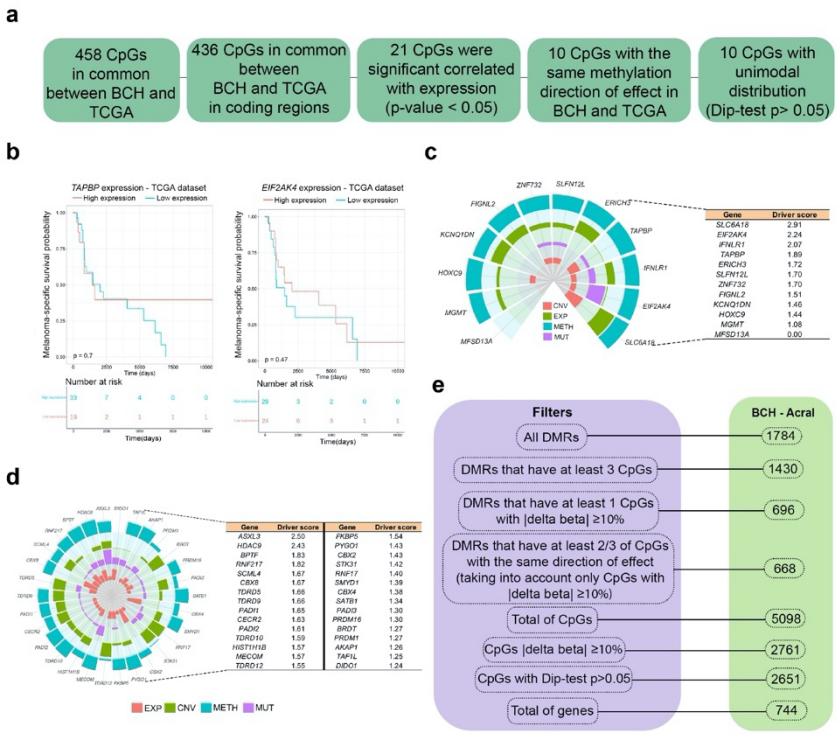
Supplementary Information

Cutaneous and acral melanoma cross-OMICs reveals prognostic cancer drivers associated with pathobiology and ultraviolet exposure
Vicente A.L.S.A. *et al*.

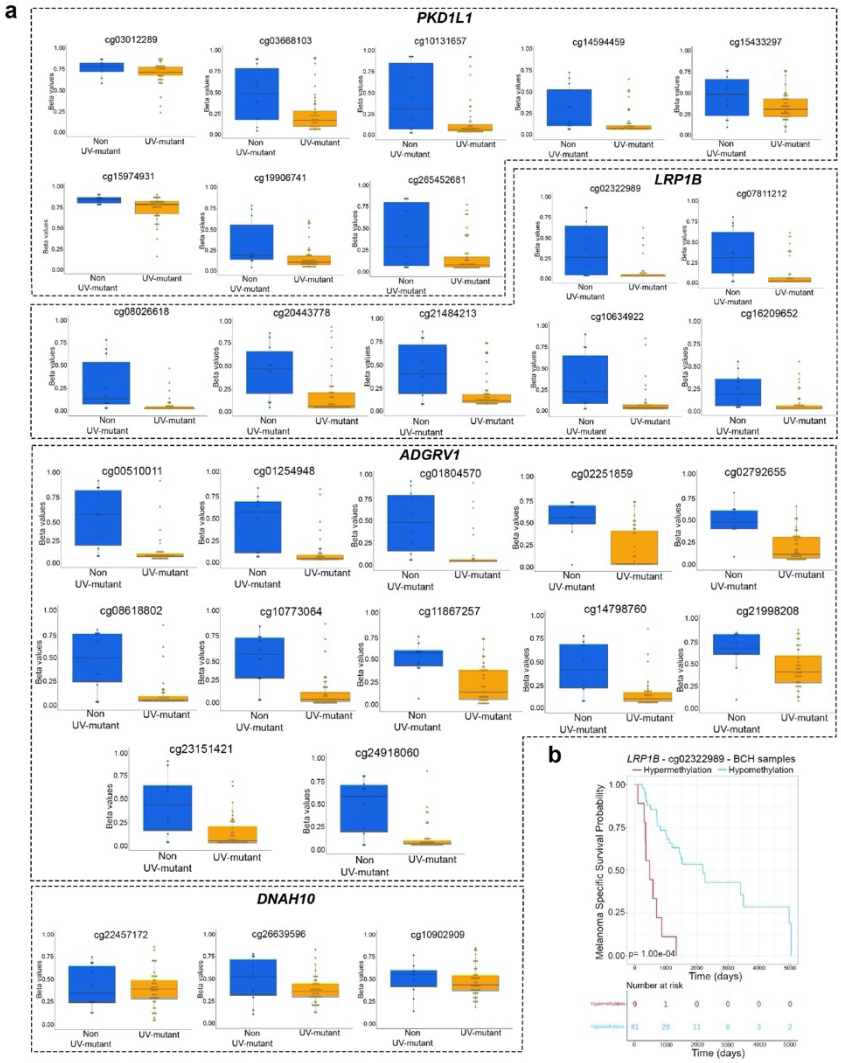
Supplementary Figures



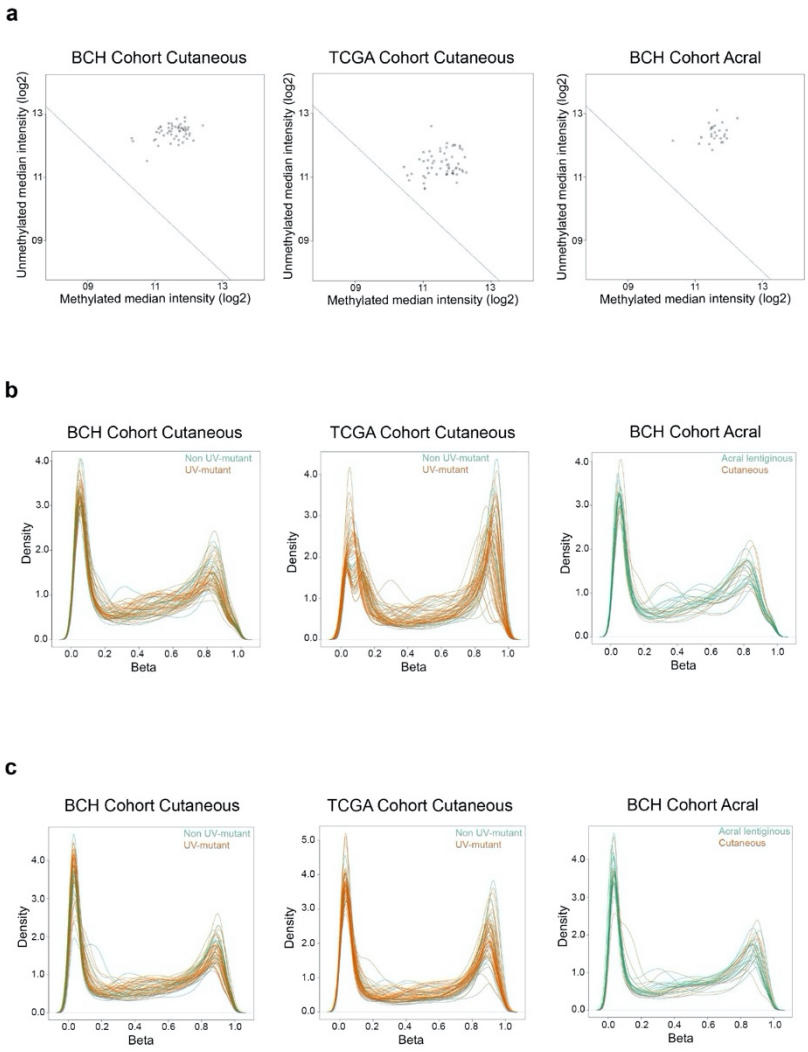
Supplementary Figure 1: Prioritization of UV methylation markers. a) Filtration steps applied to DMRs derived from the crude model in order to select the top CpGs and genes that are differentially methylated between UV-mutant *versus* non UV-mutant in BCH and TCGA. Significance was assessed using linear robust regression with FDR < 0.05. b) Prioritization criteria of UV-associated CpGs that are differentially methylated in both BCH and TCGA. c) DNA methylation levels of the 9 genes in common between BCH and TCGA showing differential methylation between UV-mutant (n= 44 and 47 in BCH and TCGA, respectively) and non UV-mutant patients (n= 44 and 47 in BCH and TCGA, respectively). Data were expressed as the average values of each group (UV-mutant and non UV-mutant) for each single CpG with error bars indicating the 95% confidence interval.



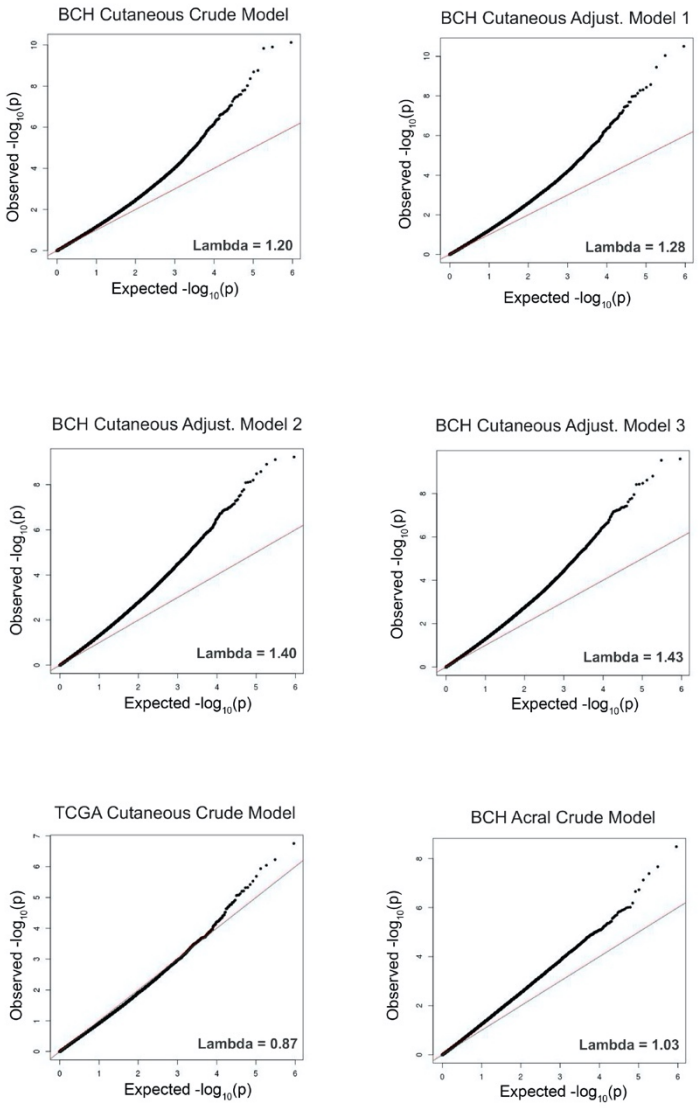
Supplementary Figure 2: Association of DNA methylation with transcription, transcription-mediated patient survival, cancer driver potential and pathobiology. a) Prioritization criteria of CpGs with significant expression quantitative trait methylation (eQTMs) ($p < 0.05$) in both BCH and TCGA. P-value was delivered from two-sided Pearson correlation and Dip-tests. b) Kaplan-Meier survival of melanoma patients in relation to expression levels of *TAPBP* and *EIF2AK4* measured in the target tumors derived from TCGA. Patients were categorized into low- and high-expression groups relative to the mean value of expression across profiled samples for a given gene. P values were derived by log-rank test. c) and d) Multi-omics data integration, encompassing copy number variation (CNV), expression (EXP), methylation (METH) and mutation (MUT), was performed in order to decipher the melanoma driver potential of the 12 prioritized genes (c) and of positive control genes previously identified in a recent study based on the ConsensusDriver method[1] (d). e) Filtration steps applied to DMRs derived from the crude model in order to select the top CpGs and genes that are differentially methylated between cutaneous and acral melanoma patients not harboring the UV mutational signature.



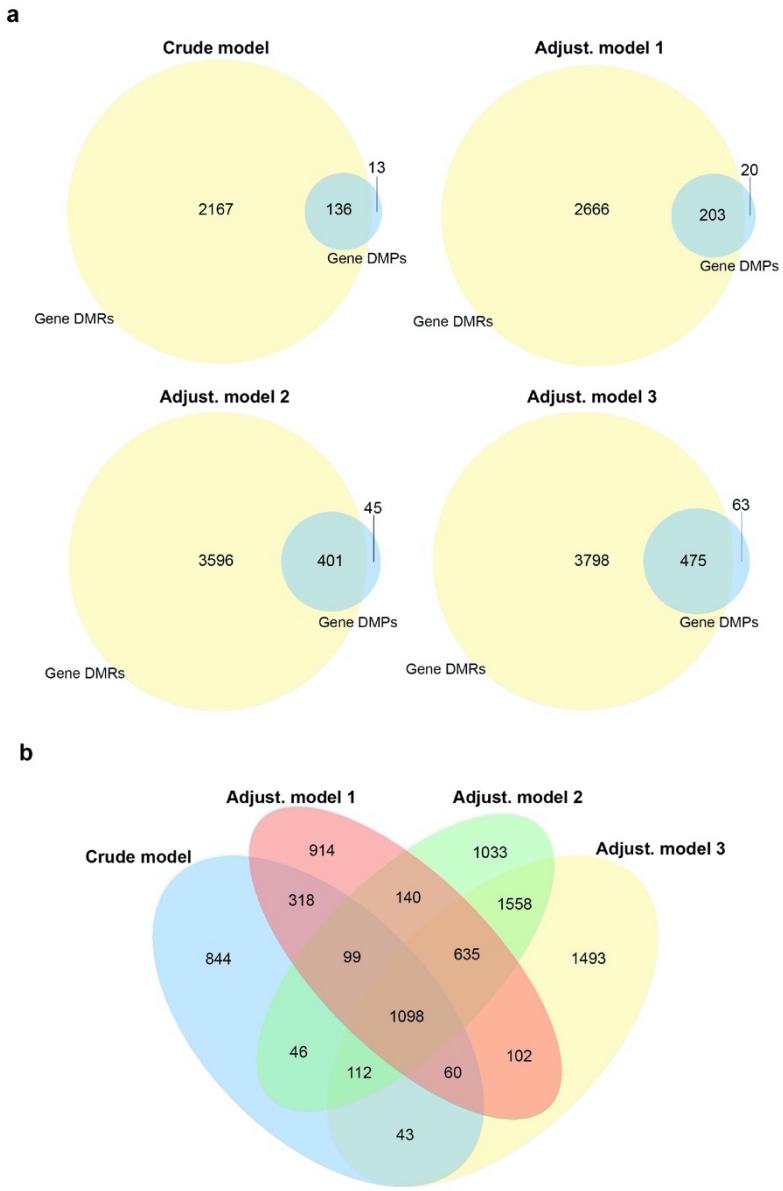
Supplementary Figure 3: DNA methylation alterations and clinical relevance of genes reported in the literature to be frequently mutated in UV-mutant *versus* non UV-mutant cutaneous melanoma patients[2]. a) *LRP1B*, *PKHD1L1*, *ADGRV1* and *DNAH10* were differentially methylated in UV-mutant (n= 44) relative to non UV-mutant (n= 10) patients in the BCH-cutaneous cohort. Box center lines, bound of the box, and whiskers indicate medians, first and third quantiles, and minimum and maximum values within 1.5xIQR (interquartile range) of the box limits, respectively. Each data point in the box plot represents the samples. b) Kaplan-Meier survival of cutaneous melanoma patients in relation to methylation levels of cg02322989 (*LRP1B*) measured in the target tumors derived from BCH. Patients were categorized into low- and high-methylation groups depending on whether the methylation value of a given CpG is lower or higher, respectively, than the mean methylation across the samples profiled for that CpG. The P-value was derived by log-rank test.



Supplementary Figure 4: Quality control of 450K methylation data. a) The plots show that all samples passed quality control (above diagonal threshold line) in each of the three indicated cohorts (BCH Cohort Cutaneous= 54, TCGA Cohort Cutaneous= 58 and BCH Cohort Acral= 21) ; b) Density plots of the beta methylation values for BCH-cutaneous, TCGA-cutaneous and BCH-acral datasets. In BCH-cutaneous and TCGA-cutaneous plots, orange and green densities represent samples harboring or not the UV signature, respectively. In BCH-acral, orange and green densities represent cutaneous (not harboring UV signature) and acral melanomas, respectively. c) Density plots of (b) after FunNorm normalization.



Supplementary Figure 5: Q-Q plots and lambda genomic inflation values of the statistical models used in BCH-cutaneous (n= 54), TCGA-cutaneous (n=58) and BCH-acral melanoma (n=17) cohorts. Crude and adjusted models are described in Methods.



Supplementary Figure 6: Comparison of results yielded from the tested crude and adjusted statistical models as well as the two approaches used for the analyses, DMP and DMR. a) Approximately 90% of DMP-based genes overlapped with DMR-based genes across all tested models. b) Venn diagram showing DMR overlaps among the 4 statistical models.

Supplementary References

1. Bertrand, D. *et al.* ConsensusDriver Improves upon Individual Algorithms for Predicting Driver Alterations in Different Cancer Types and Individual Patients. *Cancer Res* **78**, 290-301 (2018).
2. Trucco, L.D. *et al.* Ultraviolet radiation-induced DNA damage is prognostic for outcome in melanoma. *Nat Med* **25**, 221-224 (2019).

Description of Additional Supplementary Files

Cutaneous and acral melanoma cross-OMICs reveals prognostic cancer drivers associated with pathobiology and ultraviolet exposure
Vicente A.L.S.A. *et al*.

Supplementary Data 1. Systematic literature search conducted in the PubMed database until May 2021 in order to find out studies that investigated the DNA methylome profile of melanoma patients.

Supplementary Data 2. DMRs in relation to UV exposure in BCH-cutaneous crude model.

Supplementary Data 3. DMRs in relation to UV exposure in BCH-cutaneous adjusted for sex.

Supplementary Data 4. DMRs in relation to UV exposure in BCH-cutaneous adjusted for sex and age at diagnosis.

Supplementary Data 5. DMRs in relation to UV exposure in BCH-cutaneous adjusted for sex, age at diagnosis and tumor type (primary or metastatic).

Supplementary Data 6. DMRs in relation to UV exposure in TCGA-cutaneous crude model.

Supplementary Data 7. All CpGs differentially methylated comparing non UV-mutant *versus* UV-mutant in crude model that passed the filtration steps described in Supplementary Fig. 1a and used in pathway and heatmap cluster analysis (Figure 3a and 3c) in BCH.

Supplementary Data 8. All CpGs differentially methylated comparing non UV-mutant *versus* UV-mutant in crude model that passed the filtration steps described in Supplementary Fig. 1a and used in pathway and heatmap cluster analysis (Figure 3a and 3c) in TCGA.

Supplementary Data 9. Jensen Ontology analysis using genes prioritized in Supplementary Fig. 1a and described in Supplementary Data 7 in BCH cohort using Enrich-r website. P-value was delivered from two-sided Fisher's exact test.

Supplementary Data 10. Jensen Ontology analysis using genes prioritized in Supplementary Fig. 1a and described in Supplementary Data 8 in TCGA cohort using Enrich-r website. P-value was delivered from two-sided Fisher's exact test.

Supplementary Data 11. KEGG pathway analysis using genes prioritized in Supplementary Fig. 1a and described in Supplementary Data 7 in BCH cohort using Enrich-r website. P-value was delivered from two-sided Fisher's exact test.

Supplementary Data 12. KEGG pathway analysis using genes prioritized in Supplementary Fig. 1a and described in Supplementary Data 8 in TCGA cohort using Enrich-r website. P-value was delivered from two-sided Fisher's exact test.

Supplementary Data 13. KEGG Pathway analysis using CpGs prioritized in Supplementary Fig. 1a and described in Supplementary Data 7 in BCH cohort using missMethyl package, which adjusts for the number of CpG associated with each gene. P-value was delivered from two-sided Fisher's exact test.

Supplementary Data 14. KEGG Pathway analysis using CpGs prioritized in Supplementary Fig. 1a and described in Supplementary Data 8 in TCGA cohort using missMethyl package, which adjusts for the number of CpG associated with each gene. P-value was delivered from two-sided Fisher's exact test.

Supplementary Data 15. Differentially expressed genes comparing non UV-mutant and UV-mutant in TCGA cohort.

Supplementary Data 16. Jensen Ontology analysis using differentially expressed genes described in Supplementary Data 15 in TCGA cohort using Enrich-r website. P-value was delivered from two-sided Fisher's exact test.

Supplementary Data 17. KEGG Pathway analysis using differentially expressed genes described in Supplementary Data 15 in TCGA cohort using Enrich-r website. P-value was delivered from two-sided Fisher's exact test.

Supplementary Data 18. 458 CpGs in common between BCH and TCGA.

Supplementary Data 19. Meta-analysis of DMRs across the BCH and TCGA datasets considering FDR-adjusted $p < 0.05$ and DMRs with at least 3 CpGs. The approach used is fixed effects inverse variance-weighted meta-analysis. P-value was derived from two-sided test based on Z-score (obtained from the direction of effect and P-value observed in each DMR) and the standard normal cumulative distribution function.

Supplementary Data 20. Meta-analysis of DMRs across the BCH and TCGA datasets considering Bonferroni-adjusted $p < 0.05$ and DMRs with at least 3 CpGs. The approach used is fixed effects inverse variance-weighted meta-analysis. P-value was derived from two-sided test based on Z-score (obtained from the direction of effect and P-value observed in each DMR) and the standard normal cumulative distribution function.

Supplementary Data 21. Association between DNA methylation and melanoma-specific survival in BCH and TCGA cohorts of CpGs common between the two cohorts, after the filters applied in Supplementary Fig. 1b. P-values were derived from log rank test.

Supplementary Data 22. Primers used for pyrosequencing validation of the *TAPBP* gene. "For", "Rev" and "Seq" denote forward, reverse and sequencing primers, respectively.

Supplementary Data 23. Twenty-five most informative CpGs and transcripts using LASSO penalization for integrative analysis in the TCGA cohort.

Supplementary Data 24. DMRs in BCH cohort comparing acral *versus* cutaneous in non UV-exposed melanoma patients.

Supplementary Data 25. All CpGs differentially methylated comparing acral *versus* cutaneous BCH melanoma patients that passed the filtration steps described in Supplementary Fig. 2e and used in gene ontology analysis (Figure 6d).

Supplementary Data 26. Jensen Ontology analysis using genes prioritized in Supplementary Fig. 2e and described in Supplementary Data 25 in BCH-acral cohort using Enrich-r website. P-value was delivered from two-sided Fisher's exact test.

Supplementary Data 27. KEGG pathway analysis using genes prioritized in Supplementary Fig. 2e and described in Supplementary Data 25 in BCH-acral using Enrich-r website. P-value was delivered from two-sided Fisher's exact test.

Supplementary Data 28. KEGG Pathway analysis using CpGs prioritized in Supplementary Fig. 2e and described in Supplementary Data 25 in BCH cohort using missMethyl package, which adjusts for the number of CpG associated with each gene. P-value was delivered from two-sided Fisher's exact test.

Supplementary Data 29. Association between DNA methylation and melanoma-specific survival of genes frequently mutated in response to UV exposure[1]. P-values were derived from log rank test. Adjustment for multiple testing was done using FDR.

Supplementary Reference

1- Trucco, L.D. *et al.* Ultraviolet radiation-induced DNA damage is prognostic for outcome in melanoma. *Nat Med* **25**, 221-224 (2019).

5.2. Segundo Artigo

Esse artigo é fruto de um trabalho desenvolvido pela candidata ao título de doutora concomitante ao descrito no artigo 1, com o objetivo de desenvolver e padronizar um protocolo que permita a remoção de melanina de biópsias de melanoma. Esse estudo se iniciou paralelamente ao final do mestrado da candidata e a conclusão dele se deu durante o doutorado.

A melanina é um inibidor da reação em cadeia da polimerase, por reduzir a capacidade da polimerase em sintetizar DNA ou por diminuir o tempo de interação entre a enzima e o DNA molde. A reação em cadeia da polimerase é um passo essencial para a análise de mutação em tumores de pacientes com melanoma, cuja presença ou ausência muitas vezes define o melhor tratamento disponível para os pacientes. Portanto, a presença de melanina nessas biópsias muitas vezes impende que o paciente tenha o melhor tratamento disponível, em virtude de o laudo ser inconclusivo por falhas técnicas.

Dessa forma, por mais que esse tenha sido um estudo com técnicas mais simples e com objetivo menos complexos, ele é de extrema relevância para a comunidade que estuda melanoma e, sobretudo, para os pacientes que agora tem à sua disposição um protocolo que diminui as falhas técnicas das análises mutacionais que influenciam o seu tratamento. Sem dúvidas, isso tem enorme impacto no manejo dos pacientes com melanoma.

O artigo encontra-se publicado na *Histology and Histopatology*, revista científica classificada pela CAPES como Qualis A3 (Fator de impacto: 2.1), e a candidata ao título de doutora é a única primeira autora do trabalho.

Histol Histopathol (2019) 34: 1089-1096

http://www.hh.um.es

Histology and Histopathology
From Cell Biology to Tissue Engineering

Comparison of protocols for removal of melanin from genomic DNA to optimize PCR amplification of DNA purified from highly pigmented lesions

Anna Luiza Silva Almeida Vicente[1], Raquel Alves Bianchini[1], Ana Carolina Laus[1], Graziela Macedo[2], Rui Manuel Reis[1,3,4] and Vinicius de Lima Vazquez[1,5]

[1]Molecular Oncology Research Center, [2]Department of Pathology, Barretos Cancer Hospital, Barretos, São Paulo, Brazil, [3]Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, Braga, [4]ICVS/63B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal and [5]Department of Surgery Melanoma/Sarcoma, Barretos Cancer Hospital, Barretos, São Paulo, Brazil

Summary. Melanin is produced by melanocytes and protects against DNA damage by ultraviolet light. Unfortunately, the melanin protein present in melanoma tumor cells is often co-purified during DNA extraction, and this contamination may inhibit subsequent PCR methods, which directly impacts research applications and the molecular diagnostic tests needed for targeted therapeutics. There are presently no described purification protocols that efficiently remove melanin from genomic DNA. In this study, we compare six different methods for melanin removal from genomic DNA: Agarose Gel Electrophoresis, 1mg Chelex®-100, Chelex®-100 5%, centrifugation, OneStep™ PCR Inhibitor Removal Kit and centrifugation plus OneStep™ PCR Inhibitor Removal Kit. Each comparison was made using 16 formalin-fixed paraffin-embedded (FFPE) and 11 fresh cell line samples. All samples were initially tested using the multiplex PCR reaction for *GAPDH* gene that generates different sized amplified products: 100, 200, 300 and 400 base pairs, which could be inhibited by the addition of exogenous melanin. Six purification protocols were then applied, and all samples that amplified at least one *GAPDH* fragment were sequenced to analyze the presence of the *BRAF* V600E mutation. The efficiencies of amplification decreased for larger sized fragments in all methods. Our comparisons showed that centrifugation combined with the OneStep™ PCR Inhibitor Removal Kit was superior to all other methods for successful *BRAF* sequencing with 100% (100bp), 75% (200bp), 50% (300bp), and 31.3% (400bp) amplification efficiencies for the different amplicon sizes. In conclusion, this genomic DNA extraction method is highly efficient for successful PCR when tumor samples are contaminated with melanin.

Key words: Melanin, PCR inhibitor, Pigmented melanomas, Purification, Polymerase inhibition

Offprint requests to: Vinicius de Lima Vazquez, MD, PhD, , Molecular Oncology Research Center - Barretos Cancer Hospital, Antenor Duarte Villela, 1331, Zip Code: 14784 400, Barretos, São Paulo, Brazil. e-mail: viniciusvazquez@gmail.com

DOI: 10.14670/HH-18-112

Introduction

Melanocytes are dendritic cells of the neuroectoderm, and are responsible for the production of the two major types of pigments present in the skin. Activation of wild-type MC1R (melanocortin 1 receptor) promotes the production of darkly pigmented eumelanin, whereas variants with impaired signaling are associated with the production of red/yellow-sulfated pheomelanin. Activation of the MC1R by its endogenous agonists αMSH (alpha-melanocyte-stimulating hormone) and ACTH (melanocortin 2 receptor/adrenocorticotropic hormone) stimulates tyrosinase activity, converting DOPA (dihydroxyphenylalanine) to DOPAquinone (dihydroxyphenylalanine quinone) after the eumelanogenesis and pheomelanogenesis biosynthetic

pigment pathways have diverged. Melanogenesis is completed with the transfer of pigments to the keratinocytes through the melanosomes, thus protecting against DNA damage from ultraviolet (UV) radiation. The genetic damage caused by exposure to the sun and UV radiation leads to abnormal proliferation of the pigmented melanocytes leading to the dark color typical of melanoma.

The discovery that melanomas often have a specific somatic *BRAF* V600 mutation affecting the BRAF/MEK/ERK pathway led to the increasing use of BRAF and MEK therapeutic inhibitors in clinical practice for treatment of advanced metastatic disease in several countries, including North America and Europe (Flaherty et al., 2012). These new drugs are very effective, but before they can be used on patients, it is necessary to determine whether the mutation is present by performing polymerase chain reaction (PCR) tests of genomic DNA (gDNA) derived from a tumor biopsy. A technical challenge for performing PCR testing is that melanin from cancer cells can co-purify with the tumor gDNA inhibiting PCR amplification. It has been demonstrated that melanin is a potent inhibitor of thermostable DNA polymerase that is caused by interactions between the polymerase and melanin proteins. Also, PCR of large amplicons has been found to be more prone to melanin inhibition than shorter PCR tests, due the reduction of the capacity of the polymerase to extend longer synthesized DNA products (Eckhart et al., 2000).

A small number of studies have investigated different ways of removing the contaminating melanin during the DNA extraction process. Purification steps designed to exclude melanin have used NucleoSpin® DNA Clean-Up XS kits (Faber et al., 2013), PowerClean® DNA Clean-Up kits, DNA IQ™ System (Hu et al., 2015) or addition of BSA (bovine serum albumin) (Frouin et al., 2016). However, none of these clean-up methods have demonstrated a high purification efficiency. For example, the BSA assay for *BRAF* V600E analysis of gDNA from formalin-fixed paraffin-embedded (FFPE) samples was only successful for PCR in 61.5% of the samples (Frouin et al., 2016). To our knowledge, there have been no previous systematic studies comparing different methods of removing melanin from gDNA that considered the size of the amplicons and and using gDNA samples from cell lines, fresh tissue or from DNA derived from FFPE.

The aim of this study is to compare six different methods of melanin removal from gDNA (Agarose Gel Electrophoresis, 1mg Chelex®-100, Chelex®-100 5%, centrifugation, OneStep™ PCR Inhibitor Removal Kit and Spin plus OneStep™ PCR Inhibitor Removal Kit), and to evaluate their relative efficiencies for subsequent PCR and sequencing methods. Also, our experimental design allows us to evaluate the impact of different amplicon sizes on assay performance and to determine the efficiency of each method for both freshly collected and paraffin embedded samples.

Materials and methods

Cell culture

For the 11 cultured cell lines (U251, U373, GAMG, SIHA, HCT, NHA, HELA, CASKI, UW497, A375 and FADU), gDNA was extracted using BioPur (Biometrix) following the manufacturer´s recommendations. Concentration and spectrophotometric purity indicators of DNA were measured using NanoDrop 2000 (ThermoFisher) and are summarized in Table 1.

Tissue sample collection

Sixteen FFPE samples from patients with melanoma were retrospectively retrieved from the files of the Department of Pathology at Barretos Cancer Hospital. All the patients were diagnosed between 2000 and 2005 and all cases were re-evaluated by a pathologist who confirmed the diagnosis and identified the tumor region for the molecular analysis. gDNA was obtained from FFPE slides after microdissection using the QIAmp DNA Mini Kit (Qiagen) following the manufacturer´s

Table 1. Concentration and spectrophotometric purity indicators of isolated DNAs from FFPE samples and cultured cell lines measured using NanoDrop 2000 (ThermoFisher).

Sample type	Sample ID	Stock concentration (ng/uL)	A260/280 ratio	A260/230 ratio
FFPE* samples				
	P1	109.7	1.93	1.49
	P2	124.9	1.95	1.51
	P3	420.8	1.88	1.91
	P4	308.7	1.83	1.66
	P5	428.8	1.94	2.24
	P6	377.5	1.87	1.92
	P7	71.6	2.00	1.42
	P8	1572.8	1.87	2.17
	P9	608.2	1.90	1.20
	P10	1124.1	1.92	1.99
	P11	1170	1.91	1.80
	P12	939.7	1.85	1.96
	P13	1172.2	1.85	2.03
	P14	1849.5	1.93	2.16
	P15	798.7	1.78	1.97
	P16	896.1	1.91	2.05
Cultured cell lines				
	L1	1222.9	2.01	2.38
	L2	832.2	1.95	2.48
	L3	1419.1	2.04	2.30
	L4	887.3	2.00	2.22
	L5	614.1	1.83	2.43
	L6	1118.3	2.05	2.40
	L7	636.1	2.03	2.14
	L8	516.9	1.98	2.21
	L9	1178	2.02	2.42
	L10	47.5	2.10	2.00
	L11	28.6	2.00	3.40

*FFPE: formalin-fixed paraffin-embedded.

recommendations. All the samples were metastatic tumors and obtained from patients prior to any radiation or systemic therapy. The biopsies were clinically indicated and not tied to the study. Concentration and spectrophotometric purity indicators of DNA were measured using NanoDrop 2000 (ThermoFisher) and are summarized in Table 1.

Evaluation of the inhibitory effect of melanin on PCR assays

To evaluate the inhibition of melanin in the PCR, in each protocol 1µg of each gDNA and 550 ng (55ng/µL) of commercial melanin (Sigma) were used, according to Hu and collaborators (Hu et al., 2015). Briefly, pure melanin was diluted in ultrapure water at room temperature and vortexed so that known amounts of the exogenous melanin could then be spiked into the gDNA samples being tested to investigate PCR inhibition.

Agarose gel electrophoresis

Gel electrophoresis was used to separate the gDNA from the faster-migrating melanin. All samples were submitted to 1.5% agarose gel electrophoresis for 40 minutes at 90V. After that, gDNA was extracted using GeneJET Gel Extraction Kit (Thermo Fisher Scientific) following the recommendations of the manufacturer.

1 mg Chelex®-100 (Bio-Rad):

One milligram of Chelex-100 was added per sample, mixed and incubated at room temperature for 15 minutes. After centrifugation at 10,000g for 5 minutes, the supernatant containing gDNA was used for PCR reactions (Beadling et al., 2008).

Chelex®-100 5% (Bio-Rad):

An equal volume of Chelex-100 5% (Chelex-100 equilibrated in Qiagen AE buffer) was added per sample and incubated at room temperature for 10 minutes. The mixture was heated to 95°C for 2 minutes, centrifuged at 10,000g for 5 minutes and the supernatant gDNA used for PCR reactions (Beadling et al., 2008).

Centrifugation

This melanin removal protocol was established by our research group. All samples being tested in the study were centrifuged at 20,000g 14°C for 15 minutes so that the contaminating melanin collects in the pellet at the bottom of the tube. The supernatant containing the purified gDNA was used for subsequent PCR reactions.

OneStep™ PCR Inhibitor Removal Kit (Zymo Research)

All samples were purified using the fractionation columns according to the manufacturer's instructions. Before using the column, it was necessary to break the base, remove the cap, and insert it into a collection tube before centrifuging at 8,000g for 3 minutes. The column was then transferred to a 1.5 mL tube, the DNA was added and the samples, which were centrifuged for 8000g for 1 minute.

Centrifugation plus OneStep™ PCR Inhibitor Removal Kit (Zymo Research)

All samples were centrifuged at 20,000g 4°C for 15 minutes and the supernatant gDNA was added in the pre-prepared columns and the samples were centrifuged at 8,000g for 1 minute. This protocol was established by our research group.

PCR conditions

To verify the quality and integrity of the extracted DNA, a multiplex PCR reaction with four pairs of primers for the *GAPDH* gene was performed as described by Van Beers et al. (2006). This multiplex reaction amplifies 4 different fragments of 100 base pairs (bp), 200 bp, 300 bp, and 400 bp, enabling a rapid assessment of DNA quality based on fragment length. Table 2 shows the primers used and the size of the fragments. Conventional PCR multiplex amplification was performed in a volume of 30 µl containing 1.6X PCR buffer, 10 µM deoxyribonucleoside triphosphate (dNTP), 5 µM each of the forward and reverse primers for each fragment, 1 unit of Platinum Taq DNA

Table 2. Primers sequences used for PCR.

Gene	Forward	Reverse	Fragment size
GAPDH	GTTCCAATATGATTCCACCC	CTCCTGGAAGATGGTGATGG	100 bp
GAPDH	AGGTGGAGCGAGGCTAGC	TTTTGCGGTGGAAATGTCCT	200 bp
GAPDH	AGGTGAGACATTCTTGCTGG	TCCACTAACCAGTCAGCGTC	300 bp
GAPDH	ACAGTCCATGCCATCACTGC	GCTTGACAAAGTGGTCGTTG	400 bp
BRAF	TTCATGAAGACCTCACAGTAAAAA	CCACAAAATGGATCCAGACA	107 bp

bp: base pair.

polymerase (Life Technologies), and 100 ng of genomic DNA. The following cycling conditions were used: 94°C for 4 min, 35 cycles of 94°C for 1 min, 56°C for 1 min and 72°C for 1 min, and a final extension at 72°C for 7 min, and reactions were performed in Veriti Thermal Cycler (ThermoFisher). A positive control (blood DNA from a healthy donor) and no template control were used in all reactions. Amplification of PCR products was confirmed by 1.5% agarose gel electrophoresis for 60 minutes at 100V.

Analysis of BRAF V600E mutations

The BRAF V600E mutation analysis was performed by PCR followed by direct Sanger sequencing. Briefly, specific primer pairs (Table 2) that flank the target mutated region were amplified by PCR in a reaction containing 10x PCR buffer, 50mM of MgCl2, 10mM dNTP, 10 µM each of forward and reverse primers, 1 unit of Platinum Taq DNA polymerase (Life Technologies), and 50 ng of genomic DNA. The cycling condition was an initial denaturation at 95°C for 15 min, followed by 40 cycles of 95°C denaturation for 45s, 55.5°C for 45s and 72°C elongation for 45s, and 72°C final elongation for 10 min, in Veriti Thermal Cycler (ThermoFisher). Amplification of PCR products was confirmed by gel electrophoresis. Sequencing PCR products was performed using a Big Dye terminator v3.1 cycle sequencing ready reaction kit (Applied Biosystems) and the ABI PRISM 3500 xL Genetic Analyzer (Applied Biosystems). Figure 1 show a sample considered to be mutated.

Results

We first determined the initial amplification pattern of *GAPDH* of 16 FFPE and 11 cell line gDNA samples by PCR followed by electrophoresis. As expected, all of the four PCR products (100 bp, 200 bp, 300 bp and 400 bp) (Fig. 2A) derived from the DNA samples from the cell lines could be amplified efficiently. For FFPE samples, 16/16 DNA samples amplified the 100 bp PCR product (16/16 FFPE - 100%), 7/16 FFPE DNA samples amplified the 200 bp fragment (43.7%) (three samples had a band with moderate intensity and four had a weak intensity band), only 1/16 samples had weakly stained amplification product for the 300 bp (18.7%) and none of the samples amplified the 400 bp fragment (Fig. 2B). All initial amplification results using the six purification protocols for our subsequent comparisons of efficiencies of melanin removal from gDNA are summarized in Table 3. When 60.0 ng/µL of pure melanin was added in the reaction mixture, PCR was totally inhibited for gDNA from all FFPE and cell line samples (Fig. 2C,D).

Agarose gel electrophoresis

After the contaminating melanin had been removed from the gDNA by agarose gel electrophoresis the efficiency of PCR was evaluated. We observed that there was amplification of 100 bp fragment in 7/16 FFPE samples (43.7%), for the 200 bp in 1/16 samples (6.3%) and for 1/16 samples (6.3%), for 300 bp fragment. No samples amplified the 400 bp fragment (Fig. 3A). We found that in one sample P7, the gel-purified gDNA amplified both 200 bp and 300 bp fragments. Previously this sample had just amplified the 200 bp fragment in the above initial PCR experiment. For cell line gDNAs, 1/11

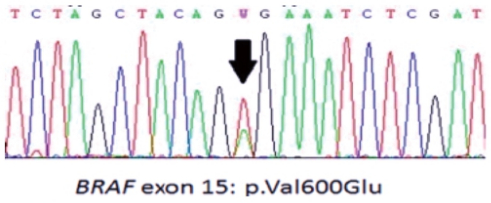


Fig. 1. Example of identification of the V600E mutation (arrow) in a post-purification sample of melanin.

Table 3. Initial amplification of GAPDH and after six protocols for melanin purification, stratified by amplicon size and gDNA origin.

Fragments size	Amplification pre-melanin		Gel Electrophoresis		1 mg Chelex®-100		Chelex-100® 5%		Centrifugation		OneStep™ Kit		Spin + OneStep™ Kit	
	FFPE	Cell Line	FFPE	Cell Line	FFPE	Cell Line	FFPE	Cell Line	FFPE	Cell Line	FFPE	Cell Line	FFPE	Cell Line
100 bp	100.0% (16/16)	100.0% (11/11)	43.8% (7/16)	90.9% (10/11)	0.0% (0/16)	0.0% (0/11)	0.0% (0/16)	0.0% (0/11)	100.0% (16/16)	100.0% (11/11)	93.8% (15/16)	100.0% (11/11)	100.0% (16/16)	100.0% (11/11)
200 bp	43.8% (7/16)	100.0% (11/11)	6.3% (1/16)	63.6% (7/11)	0.0% (0/16)	0.0% (0/11)	0.0% (0/16)	0.0% (0/11)	12.5% (2/16)	90.9% (10/11)	43.8% (7/16)	100.0% (11/11)	75.0% (12/16)	100.0% (11/11)
300 bp	6.3% (1/16)	100.0% (11/11)	6.3% (1/16)	54.5% (6/11)	0.0% (0/16)	0.0% (0/11)	0.0% (0/16)	0.0% (0/11)	0.0% (0/16)	90.9% (10/11)	12.5% (2/16)	100.0% (11/11)	50.0% (8/16)	100.0% (11/11)
400 bp	0.0% (0/16)	100.0% (11/11)	0.0% (0/16)	27.3% (3/11)	0.0% (0/16)	0.0% (0/11)	0.0% (0/16)	0.0% (0/11)	0.0% (0/16)	90.9% (10/11)	6.3% (1/16)	100.0% (11/11)	31.3% (5/16)	100.0% (11/11)

bp: base pair; FFPE: formalin-fixed paraffin-embedded.

(9.1%) did not amplify any fragments, 10/11 amplified the 100 bp fragment (90.9%%), 7/11 (63.6%) amplified the 200 bp fragment, 3 amplified the 300 bp (27.3%) and 3/11 samples (27.3%) amplified the four fragments (Fig. 3B) (Table 3).

1 mg Chelex®-100 and Chelex®-100 5%

Neither of these purification protocols was efficient at amplifying any of the four fragments, for both FFPE and cell line gDNA (data not show) (Table 3).

Centrifugation

After applying our simple centrifugation protocol to the FFPE gDNA the 100 bp fragment was observed in 16/16 FFPE samples (100%) but only 2/16 samples amplified the 200 bp fragment and neither the 300 bp or

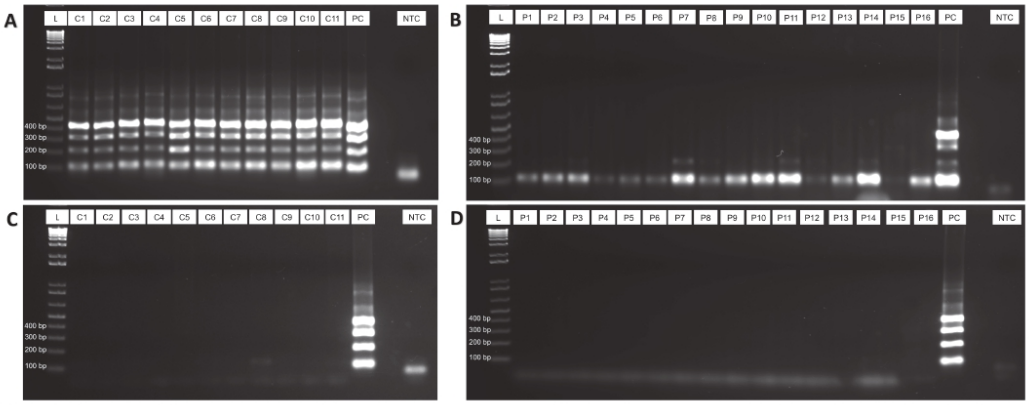


Fig. 2. Initial standard PCR amplification and effects of addition of melanin on PCR. Amplification pattern of GAPDH (100, 200, 300 and 400 bp) for gDNA from cell lines **(A)** and FFPE samples **(B)**. Inhibitory effects of 55 ng/μL of melanin in the PCR for samples of cell lines **(C)** and FFPE **(D)**. L: ladder; P: FFPE samples; C: cell lines; bp: base pair; PC: positive control; NTC: no template control.

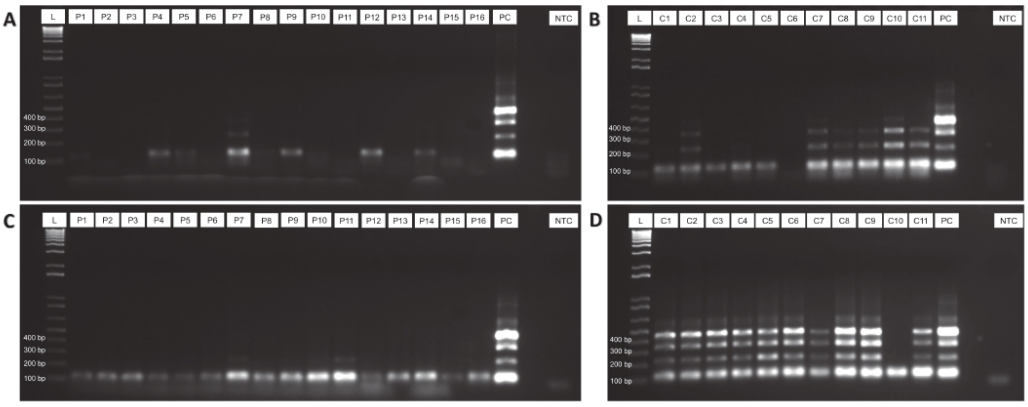


Fig. 3. Protocols for removal of the inhibitory effects of melanin on PCR amplification. The effect of electrophoretic separation of gDNA from melanin and PCR results on purified gDNA from FFPE samples **(A)** and from cell lines **(B)**. Amplification post-purification with the spin protocol for FFPE **(C)** and cell lines **(D)**. L: ladder; P: FFPE samples; C: cell lines; bp: base pair; PC: positive control; NTC: no template control.

400 bp fragment could be amplified (Fig. 3C). In contrast, for cell line gDNA samples 10/11 samples (90.9%) of 200 bp,300 bp, and 400 bp fragments could be amplified. For the smaller 100 bp amplicon, all FFPE and cell line gDNA samples had successful amplification (Fig. 3D) (Table 3).

OneStep™ PCR Inhibitor Removal Kit:

We found that 15/16 (93.8%) of FFPE samples amplified the 100 bp sized amplicon after gDNA purification with OneStep™ PCR Inhibitor Removal kit. For the larger sized 200 bp fragment 7/16 (43.8%) samples had successful amplification, and 2/16 (12.5%) amplified the 300 bp and 1/16 (6.3%) amplified the 400 bp amplicon (Fig. 4A). For the gDNA purified from cell line samples, 11/11 (100%) amplified all four fragments, but in two of them (sample C4 and C11), the 200 bp fragment was weakly amplified (Fig. 4B) (Table 3).

Centrifugation plus OneStep™ PCR Inhibitor Removal Kit

We reasoned that we could maximize the overall efficiency of purification by combining the two protocols with the highest success rates. Our results showed that the combination of centrifugation protocol followed by OneStep™ PCR Inhibitor Removal Kit was efficient in successful amplification of 100 bp fragment in 100% (16/16) of the FFPE samples. Interestingly, there was also an increase of amplification of the 200 bp (12/16 samples - 75%) and the 300 bp fragment (8/16 - 50%) in comparison to our experiments in which we spiked melanin into the initial gDNA templates (Table 3). Furthermore, 5/16 gDNA samples (31.3%) successfully amplified the 400 bp fragment, which was not observed in the previous experiment (Fig. 4C). Also, this protocol was efficient in recovering 100% of cell line samples, maintaining the same amplification pattern prior to the addition of melanin (Fig. 4D) (Table 3).

Finally, all samples in which there was successful amplification post-purification of melanin, regardless of the protocol used, could be successfully used for molecular analyses, such as the identification of the *BRAF* V600E mutation by sequencing (Fig. 1).

Discussion

Melanin inhibits PCR, which is an essential molecular technique for most investigative methods of cancer biology. Co-purification of melanin with gDNA isolated from lesions of patients with melanoma can not only compromise molecular diagnostic testing for the management of patients in need of targeted therapeutics but also limits the use of the tumor gDNA for many research applications. There has been no previous comparison of the impact of different gDNA separation protocols from melanin on the quality of PCR efficiencies gDNA templates from various biological sources with different amplicon sizes.

The gel electrophoresis protocol was the least efficient in maintaining the amplification pattern of the samples pre-addition of melanin, independent of the size of the amplicon and the origin of gDNA. We found that Chelex®-100 was inefficient for melanin purification. For the centifugation and the centifugation plus

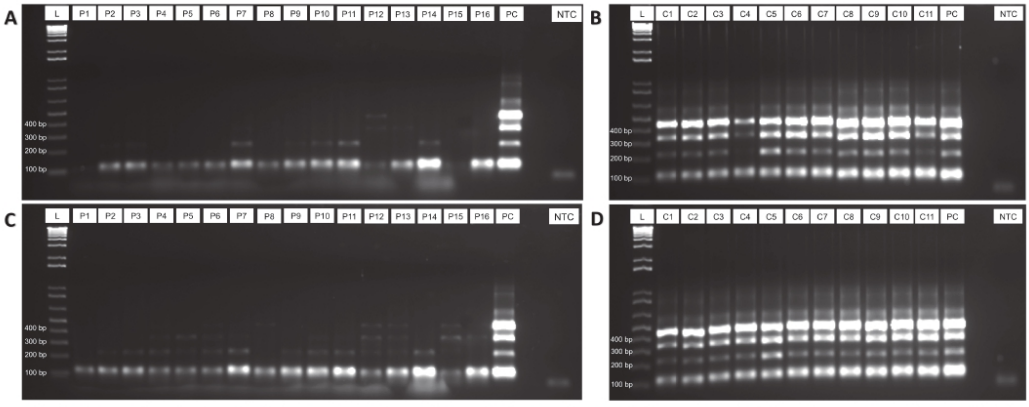


Fig. 4. The two most efficient protocols for removal of the inhibitory effects of melanin on PCR amplification. The efficiency of the OneStep™ PCR Inhibitor Removal Kit to amplify the FFPE **(A)** and cell line samples **(B)**. Combination of the Spin plus OneStep™ PCR Inhibitor Removal Kit purification protocols to recover the amplification pattern of the FFPE **(C)** and cell line samples **(D)**. L: ladder; P: FFPE samples; C: cell lines; bp: base pair; PC: positive control; NTC: no template control.

OneStep™ PCR Inhibitor Removal Kit protocols, we found that the 100 bp fragment for both FFPE-gDNA and cell line-gDNA amplified with the same percentage of success that we obtained before addition of exogenous melanin. However, only one FFPE sample did not amplify this fragment using only OneStep™ PCR Inhibitor Removal Kit. The amplification success rate of the 200, 300 and 400 bp fragments for the cell line-gDNA using purification with centrifugation and centifugation plus OneStep™ PCR Inhibitor Removal Kit was equal to the rate we obtained before addition of exogenous melanin. Interestingly, the combination of centifugation with OneStep™ PCR Removal Kit increases the amplification rate of these fragments in the paraffin samples. Also noteworthy is that the use of our simple centrifugation protocol resulted in amplification at high rates in gDNA cell line samples for the four fragments (Table 3).

In agreement with our results, Hu and colleagues (Hu et al., 2015) showed that the use of Chelex®-100 was inefficient for the removal of melanin from a commercial DNA sample (Hu et al., 2015). However, the two protocols that we used were also briefly described in two publications for the removal of melanin from pigmented samples of patients with melanoma (Beadling et al., 2008; Torres-Cabala et al., 2009). There is no specific information on the percentage of recovery of samples using these methods, as well as a lack of detailed descriptions of the procedures used. It is therefore difficult to draw conclusions regarding the different findings reported in these studies.

Frouin and colleagues (Frouin et al., 2016) showed the efficiency of the BSA and another separation column, the NucleoSpin® Kit (Macherey-Nagel GmbH & Co) to generate conclusive results in the analysis of the *BRAF* V600E mutation in 50 pigmented samples with three levels of melanin contamination, weak, moderate and high. However, none of these samples had very high levels of melanin contamination since without any treatment mutation was detected in 48% of the samples (Frouin et al., 2016). After the BSA treatment the success rate increased to 80%, an increase of 61.5%, and conclusive results were possible in 89.0% of the samples with high melanin levels. They found that the use of the NucleoSpin® Kit maintained the same 48.0% success rate for detecting the presence or absence of the mutation that was found before treatment. However, the use of this Kit was not as efficient for purification at higher levels of melanin contamination. The NucleoSpin® Kit seems to be efficient at retaining larger amounts of contaminating melanin during the clean-up procedure, but it is possible that the column resin interacts with the gDNA in some way so that some conclusive pre-treatment samples were inconclusive after purification.

Our experimental design addressed the possibility of additional inhibitors being introduced by the reagents or processes used for purification. We only used gDNA-FFPE and gDNA-cell line samples that were known to be efficiently amplified by the PCR. We decided to follow the most common scenario, in which the diagnosis routines are performing in FFPE samples, since frozen samples are still rarely available. The culture cells were included for validation in a situation with high quality gDNA.

Then, we added commercial melanin to show that there was no other inhibition in amplification reaction by the different methods of purification. Thus, our findings relate exclusively to the removal of melanin from gDNA and the relative efficiencies of each method. It was then possible to compare the percentage that each method produced and determine whether the pattern of amplification based on the various fragment sizes could be restored when the exogenous melanin was removed from the gDNA. On the other hand, this experimental design made inviable the comparison of the six different protocols when melanin is endogenous in melanoma specimen. However, Frouin and colleagues (Frouin et al., 2016) showed agreement of their results between exogeneous and endogeneous melanin.

Our findings show that the use of the centifugation plus OneStep™ PCR Inhibitor Removal Kit amplified fragments of FFPE samples that had not been previously detected. The samples preserved in paraffin are exposed to fixation and embedded in paraffin resulting in partial degradation of the gDNA. The substances used in this procedure may not be completely removed by gDNA extraction and purification, interfering with subsequent molecular analyses. In our results it is clear that the column of the OneStep™ PCR Inhibitor Removal Kit was also capable of removing some of these contaminants and removal seems to be further improved by sedimentation during centrifugation. Our centrifugation protocol is efficient, simple and low cost, and it can recover approximately 91.0% of the samples for amplicon sizes up to 400 bp, and 100.0% of the 100 bp PCR amplicons. However, for the samples derived from FFPE, we recommend using amplicon sizes up to 100 bp, with a success rate of 100.0%, since we found that when the target amplicon size was 200 bp, the efficiency decreased by 12.5%. The use of agarose gel electrophoresis was also efficient in the separation of melanin and gDNA, especially in the gDNA from cell lines. However, we do not encourage the use of this protocol because it involves extensive manipulation, which may lead to contamination between adjacent samples during electrophoresis. Also, the total concentration of post-purification gDNA was very low.

We chose a multiplex PCR that amplified four fragments of 100 bp, 200 bp, 300 bp, and 400 bp to test the efficiency of different amplicon sizes for the purification gDNA. We found that the amplification of larger fragments was more affected by melanin inhibition than smaller fragments (Eckhart et al., 2000). It is well described that a balance between DNA template and all other components of the PCR reaction, especially the primers, is necessary because PCR

performance is strongly influenced by internal stability, melting temperature, secondary structure or interference with each other (Sint et al., 2012). This template-primer balance is essential when amplifying degraded DNA samples, to avoid the preferential amplification of more efficient primers (Sint et al., 2012). Finally, the idea of using a multiplex PCR was that if a purification protocol were able to restore the amplification of all amplicons, independent of the size, it probably would be efficient in common PCR protocols routinely used in diagnosis laboratories, as in fact we described in our results for *BRAF* evaluation.

In conclusion, we compared six different protocols of purification of contaminating melanin in gDNA samples of different quality with varying sizes of the amplified fragments. Our protocol that combined the separation of melanin from gDNA by centrifugation with subsequent passage through the column of the OneStep™ PCR Inhibitor Removal Kit, recovered 100% of gDNA from cell line and 93.8% of gDNA from FFPE, with amplification of more different sized fragments in comparison to the results obtained before melanin was added to the gDNA. This protocol is feasible and may be used as a routine for research and molecular diagnostic to eliminate the melanin present in heavily pigmented lesions of patients with melanoma.

Acknowledgements. The authors would like to express their sincere thanks to the São Paulo Research Foundation (FAPESP) for its funding of this research (Vazquez's Grant #2012/04194-1, Vicente's Doctoral Fellowship Grant #2016/15941-3 and Vicente's International Research Fellowship Grant #2017/09612-0). The authors would like to thank Dr. Jeremy Squire for checking the English and for his critical review of this manuscript.

Competing financial interest. The author(s) declare no competing financial interests.

Ethical statement. The study was conducted according to the national and institutional ethical policies and it was approved by the Ethic Committee of Barretos Cancer Hospital (Process number: CEP-548/2011). Informed consent was obtained from all participants.

References

Beadling C., Jacobson-Dunlop E., Hodi F.S., Le C., Warrick A., Patterson J., Town A., Harlow A., Cruz F., 3rd, Azar S., Rubin B.P., Muller S., West R., Heinrich M.C. and Corless C.L. (2008). Kit gene mutations and copy number in melanoma subtypes. Clin. Cancer Res. 14, 6821-6828.

Eckhart L., Bach J., Ban J. and Tschachler E. (2000). Melanin binds reversibly to thermostable DNA polymerase and inhibits its activity. Biochem. Biophys. Res. Commun. 271, 726-730.

Faber K.L., Person E.C. and Hudlow W.R. (2013). Pcr inhibitor removal using the nucleospin(r) DNA clean-up xs kit. Forensic Sci. Int.. Genet. 7, 209-213.

Flaherty K.T., Robert C., Hersey P., Nathan P., Garbe C., Milhem M., Demidov L.V., Hassel J.C., Rutkowski P., Mohr P., Dummer R., Trefzer U., Larkin J.M., Utikal J., Dreno B., Nyakas M., Middleton M.R., Becker J.C., Casey M., Sherman L.J., Wu F.S., Ouellet D., Martin A.M., Patel K., Schadendorf D. and Group M.S. (2012). Improved survival with mek inhibition in braf-mutated melanoma. N. Engl. J. Med. 367, 107-114.

Frouin E., Maudelonde T., Senal R., Larrieux M., Costes V., Godreuil S., Vendrell J.A. and Solassol J. (2016). Comparative methods to improve the detection of braf v600 mutations in highly pigmented melanoma specimens. PLoS One 11, e0158698.

Hu Q., Liu Y., Yi S. and Huang D. (2015). A comparison of four methods for pcr inhibitor removal. Forensic science international. Genetics 16, 94-97.

Sint D., Raso L. and Traugott M. (2012). Advances in multiplex PCR: Balancing primer efficiencies and improving detection success. Methods Ecol. Evol. 3, 898-905.

Torres-Cabala C.A., Wang W.L., Trent J., Yang D., Chen S., Galbincea J., Kim K.B., Woodman S., Davies M., Plaza J.A., Nash J.W., Prieto V.G., Lazar A.J. and Ivan D. (2009). Correlation between kit expression and kit mutation in melanoma: A study of 173 cases with emphasis on the acral-lentiginous/mucosal type. Mod. Pathol. 22, 1446-1456.

van Beers E.H., Joosse S.A., Ligtenberg M.J., Fles R., Hogervorst F.B., Verhoef S. and Nederlof P.M. (2006). A multiplex PCR predictor for acgh success of ffpe samples. Br. J. Cancer 94, 333-337.

Accepted April 5, 2019

6. Colaborações

O doutorado, de forma geral, é um estágio acadêmico reconhecido por estudos profundos em uma área do conhecimento, visando entender questões da sociedade. Além disso, também é um período em que o profissional, no seu desenvolvimento enquanto cientista, deve construir conhecimentos sólidos em uma área de pesquisa, bem como em diversas técnicas que possam ajudá-lo a ser um pesquisador independente nos próximos estágios. Sendo assim, o período de doutorado é um momento ideal para colaborar com outros projetos de pesquisa afim de aumentar o nível de aprendizado tanto quanto possível, além de contribuir com o teste de outras hipóteses.

A Oncologia Molecular é uma vasta área do conhecimento que engloba, além de diversos tipos tumorais, diversas técnicas que vão desde procedimentos bem estabelecidos e corriqueiros no laboratório, como a extração de DNA, passando por ensaios pré-clínicos e análises OMICs. Portanto, colaborar com outros projetos é uma oportunidade de desenvolver uma visão mais ampla dessa área que é tão complexa, aumentando a probabilidade de maior sucesso enquanto um pesquisador independente no futuro.

Sendo assim, o primeiro artigo deste tópico é um estudo em colaboração com o grupo de Estudos em Câncer de Cabeça e Pescoço do Hospital de Câncer de Barretos e o segundo artigo é um outro trabalho fruto da parceria com o Grupo de Epigenética, da IARC.

6.1. Primeiro Artigo

Esse artigo encontra-se publicado na *Oncotarget*, revista científica classificada pela CAPES como Qualis A1 (Fator de impacto: 5.2), e a candidata ao título de doutora é a terceira autora do trabalho.

As principais contribuições de aprendizado proporcionada por essa colaboração incluem: procedimentos básicos de cultivo celular (linhagens primárias e comerciais/estabelecidas), técnica de Elisa, citometria de fluxo, western blot e teste de viabilidade cellular.

www.oncotarget.com Oncotarget, 2018, Vol. 9, (No. 43), pp: 27233-27241

Research Paper

Construction and characterization of a new TRAIL soluble form, active at picomolar concentrations

Matias Eliseo Melendez[1], Renato José Silva-Oliveira[1], Anna Luiza Silva Almeida Vicente[1], Lidia Maria Rebolho Batista Arantes[1], Ana Carolina de Carvalho[1], Alberto Luis Epstein[2], Rui Manuel Reis[1,3,4] and André Lopes Carvalho[1]

[1]Molecular Oncology Research Center, Barretos Cancer Hospital, Barretos, São Paulo, Brazil

[2]UMR1179, INSERM-UVSQ, Handicap Neuromusculaire, Biotherapie et Pharmacologie Appliquées, Université de Versailles-Saint Quentin en Yvelines, Versailles, France

[3]Life and Health Sciences Research Institute (ICVS), Health Sciences School, University of Minho, Braga, Portugal

[4]ICVS/3B's-PT Government Associate Laboratory, Braga/Guimarães, Portugal

Correspondence to: *André Lopes Carvalho,* ***email:*** *carvalhoal@gmail.com*

Keywords: *TRAIL; apoptosis; cancer treatment; amplicon vectors*

Received: *August 12, 2016* ***Accepted:*** *May 14, 2018* ***Published:*** *June 05, 2018*

ABSTRACT

Apoptosis induction has emerged as a treatment option for anticancer therapy. Tumor necrosis factor-related apoptosis-inducing ligand (TRAIL), a type II transmembrane protein, is a potent and specific pro-apoptotic protein ligand, which activates the extrinsic apoptosis pathway of the cell death receptors. Here we describe the construction and characterization of a new soluble TRAIL, sfTRAIL, stabilized with the trimerization Foldon domain from the Fibritin protein of the bacteriophage T4. Supernatants of 0.22 µM-filtered supernatants were produced in Vero-transduced cells with HSV1-derived viral amplicon vectors. Experiments were undertaken in two known TRAIL-sensitive (U373 and MDA.MB.231) and two TRAIL-resistant (MCF7 and A549) cell lines, to determine (i) whether the sfTRAIL protein is synthetized and, (ii) whether sfTRAIL could induce receptor-mediated apoptosis. Our results showed that sfTRAIL was able to induce apoptosis at concentrations as low as 1899.29 pg/mL (27.71 pM), independently of caspase-9 activation, and reduction in cell viability at 998.73 fM.

INTRODUCTION

Apoptosis induction has emerged as a treatment option in anticancer therapy. Different from non-apoptotic forms of cell death, apoptosis does not result in the release of cellular pro-inflammatory molecules into the tumor microenvironment, which may lead to further tissue damage [1]. Apoptosis is initiated by two different mechanisms: the death receptor-mediated (extrinsic pathway) and the mitochondria-triggered one (intrinsic pathway). Over the last decade, activation of extrinsic pathway of apoptosis has been considered an attractive therapeutic strategy to promote apoptosis of tumor cells [2–4]. Among the death receptors explored for cancer treatment, receptors belonging to the pro-apoptotic pathway of the TNF-related apoptosis-inducing ligand (TRAIL) has gained interest and even entered in clinical trials in combination with cytotoxic chemotherapy.

Tumor necrosis factor-related apoptosis-inducing ligand (TRAIL), a type II transmembrane protein is a potent and specific pro-apoptotic protein ligand, able to bind to five different receptors. Upon trimerization, TRAIL promotes apoptosis by binding to their effector death receptors, TRAIL-R1 and TRAIL-R2, triggering the recruitment of the adaptor molecule Fas-associated death domain and procaspase 8 to the cytoplasmic death domain of the receptor [5, 6]. Activated caspase-8 then cleaves caspase-3, which in turn promotes cleavage of

critical downstream cellular proteins, finally leading to DNA fragmentation and cell death. Decoy receptor 1 (TRAIL-R3), decoy receptor 2 (TRAIL-R4) and the soluble receptor osteoprotegerin lack active cytoplasmic death domains, blocking the apoptotic machinery [7].

TRAIL-induced apoptosis leaves non-tumoral cells unharmed, underlying its potential as a candidate therapeutic protein in the carcionogenesis cascade [2, 8]. While this oncotargeted potential is poorly understood, it has made TRAIL pathway an attractive target for cancer treatment, leading to the development of clinical trials of recombinant human TRAIL proteins. Among the tumor types already tested in clinical trials, Non-Small Cell Lung Cancer (Phase II), Colorectal Cancer (Phase I) and Non-Hodgkin's Lymphomas (Phase I). Although, despite the therapeutic potential of cell death receptor agonists already tested in preclinical models, the translation of these effects into the clinic remains disappointing, mainly due to protein instability and high cost of production. Here, we describe the construction and characterization of a novel recombinant soluble TRAIL protein, named sfTRAIL stabilized with the trimerization Foldon domain, showing its capabilities of inducing robust apoptosis, even at picomolar concentrations.

RESULTS

Expression of TRAIL receptors

We further determined the expression of TRAIL receptors, TRAIL-R1 and TRAIL-R2, in A549, MCF7, U373 and MDA.MB.231 cell lines by flow cytometry (Supplementary Figure 1). Resistance/sensitivity to TRAIL-mediated apoptosis was previously described for these cell lines [9–12]. Both sensitive cell lines (U373 and MDA.MB.231) showed higher expression of TRAIL-R1 and TRAIL-R2 than the MCF7 cell line. Interestingly, the A549 cell line showed high expression of both TRAIL receptors.

Sensitivity to rhTRAIL induced apoptosis

In order to analyze the sensitivity and resistance of cell lines to TRAIL induced apoptosis, commercially available rhTRAIL was used (Supplementary Figure 2). Both TRAIL-sensible cell lines, U373 and MDA.MB.231 showed an increased apoptosis, when compared to control groups. Interestingly, MCF7 cell line also showed an increase in apoptosis. Only A549 cell line showed no effect when exposed to rhTRAIL.

Expression of sfTRAIL

In silico protein mass analysis (http://web.expasy.org/compute_pi/) indicates that the sfTRAIL protein constructed had a 22845.66 Da mass. To detect the sfTRAIL protein expression, we initially transduced Vero cells with pA.EUA1 or pA.sfTRAIL amplicon vectors. As expected, western blot results showed a band of ~22 KDa, which represents the monomeric form of the sfTRAIL protein (Figure 1).

Quantification of sfTRAIL in pA.sfTRAIL transduced Vero cells

Secretion of sfTRAIL protein was quantified by ELISA, in Vero cells transduced with pA.EUA1 or pA.sfTRAIL amplicon vector at MOI1. Subconfluent Vero cells were transduced in T75 cell culture flask with pA.EUA1 or pA.sfTRAIL amplicon vectors, at MOI1. Transduced Vero cell lines were able to produce filtered supernatants (SN) with a concentration of up to 1899.29 pg/mL (27.71 pM), which was the highest sfTRAIL production ever obtained by this method. Despite the concentration produced by this approach (about 2 ng/mL) not being enough for *in vivo* pre-clinical or clinical testing, it was sufficient for the cell culture assays performed, as shown below. Supernatants from pA.EUA1-transduced Vero cells were used as negative control, with undetectable production of TRAIL protein (Supplementary Figure 3).

Bystander effect of sfTRAIL protein

Next, we evaluated the capability of apoptosis induction of the sfTRAIL protein, using conditioned media from pA.sfTRAIL-transduced Vero cells described above, which sfTRAIL concentration was 1899.29 pg/mL (27.71 pM). Supernatants from pA.EUA1-transduced Vero cells were used as negative control. The MTS assay showed a statistically different viability reduction in both TRAIL-sensitive cell lines, U373 (p-value=0.0016) and MDA.MB.231 (p-value<0.0001), treated with sfTRAIL-containing supernatants (Figure 2). Statistical differences were not observed for TRAIL-resistant cell lines, A549 and MCF7.

Apoptosis induction of sfTRAIL protein

We further investigated the sfTRAIL protein pro-apoptotic properties. We also analyzed caspase-9 cleavage, which is one of the molecular components of the intrinsic mitochondrial apoptosis pathway, using filtered conditioned media from Vero transduced cells (Figure 3). Western blot analysis indicates cleavage of pro-caspase 8, pro-caspase 3 and PARP, upon induction with 0.22 μm-filtered conditioned media, produced in Vero cells transduced with pA.EUA1 or pA.sfTRAIL (1899.29 pg/mL; 27.71 pM). As expected, cleavage of pro-caspase 9 was not observed at these time points. In this way, sfTRAIL apoptosis induction rapidly activates the extrinsic apoptosis pathway, in TRAIL-sensitive cell lines, as early as 2 hours after sfTRAIL induction. We also observed a cleavage of caspase-9 in the

MDA.MB.231 cell line treated with sfTRAIL-containing supernatants, which is consistent with its type II apoptotic behavior. We found that, sfTRAIL protein behaves as expected, promoting apoptosis at concentrations of no more than 2 ng/mL, in the sensitive cell lines used, but not in the resistant ones. At this point, it is worth to recall that we decided to include the analysis of caspase-3 cleavage in the MCF-7 cell line, despite the fact this cell line does not express this protein.

Bystander effect of sfTRAIL protein at femtomolar concentrations

Next, we evaluated the ability to reduce cell viability of the sfTRAIL protein at femtomolar concentrations, using conditioned media from pA.sfTRAIL-transduced Vero cells described above, which sfTRAIL concentration was in mean 78.32 pg/mL. Supernatants from non-transduced cells (NTC) and pA.EUA1-transduced Vero cells were used as negative control. After 72 hours of incubation with conditioned supernatants, MTS assay showed a statistically different viability reduction (p-value=0.0032 or p-value<0.0001, depending on the supernatant stock used) in the U373 (TRAIL-sensitive) for all three protein productions, while the A549 (TRAIL-resistant) cell line did not respond to the sfTRAIL challenge (Figure 4). Thus, considering a trimeric sfTRAIL mass of 68.53698 KDa and a concentration of 68.45 pg/mL (sfTRAIL stock B), the sfTRAIL protein induced a decrease in cell viability at a concentration of 998.73 fM in U373 cells.

Reproducibility of sfTRAIL protein production

In order to address the consistency between each sfTRAIL protein production, we transduced Vero cells with a EGFP-expressing amplicon control vector or the sfTRAIL-expressing amplicon vector in a biological triplicate. Transductions were done at MOI0.1. As a result, after the ELISA quantification we obtained sfTRAIL protein supernatants with a mean concentration of 78.32 pg/mL (Table 1 and Supplementary Figure 4). Variation between the three independent sfTRAIL protein productions was minimal, demonstrating a good reproducibility in the sfTRAIL production approach.

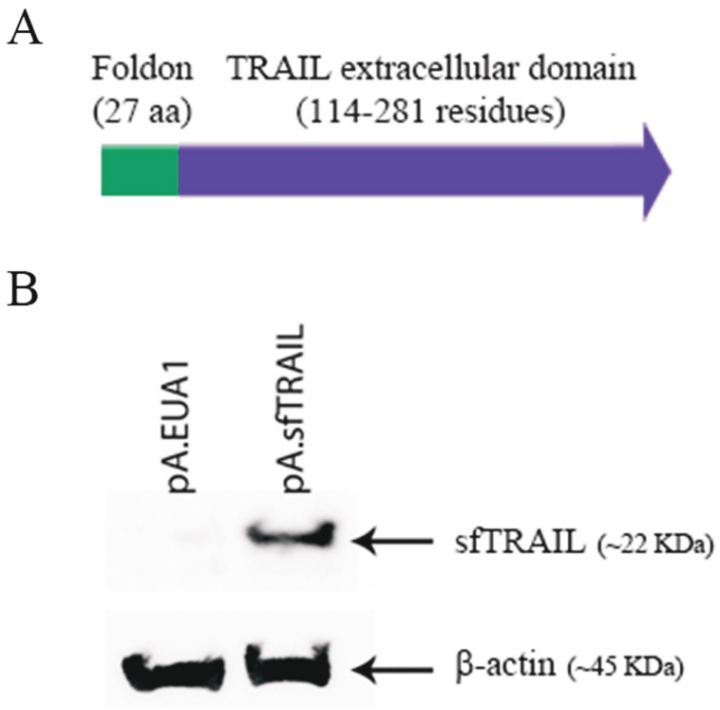


Figure 1: (A) Schematic representation of the sfTRAIL protein; **(B)** Western blot analysis of sfTRAIL protein expression. Vero cells were transduced with pA.EUA1 or pA.sfTRAIL amplicon vectors, at MOI1. Anti-TRAIL antibody and anti-β-actin were used.

DISCUSSION

The use of soluble TRAIL has been pronounced as an ideal therapeutic molecule exhibiting a selective and potent antitumor effect [8]. TRAIL-induced apoptosis leaves non-tumoral cells unharmed, underlying its potential as a therapeutic protein in oncology treatment. While this oncotargeted potential is poorly understood, it has encouraged the development of clinical trials of recombinant human TRAIL proteins. Among the tumor types already tested in clinical trials, Non-Small Cell Lung Cancer (Phase II), Colorectal Cancer (Phase I) and Non-Hodgkin's Lymphomas (Phase I). However, the instability of TRAIL protein and its high cost constitute major drawbacks to its clinical use.

Soluble TRAIL (sTRAIL) stabilization as a trimer is essential for its binding to the effector TRAIL receptors, TRAIL-R1 and TRAIL-R2, even if expression of both TRAIL receptors were not described as predictors of TRAIL response [13, 14]. Several versions of recombinant soluble TRAIL with different N-terminal fusions domains have been reported [15–18]. Most TRAIL clinical trials use a non-tagged version, containing amino acids 114–281, stabilized by the addition of zinc and reducing agent to the cell-culture media and extraction buffers, and by formulation of the purified protein at neutral pH [19]. These further modifications increases the production steps and time, and may explain the high cost of the therapeutic protein. In addition, further modifications may affect the pro-apoptotic capabilities of the resulting trimer, requiring high drug concentrations.

To our knowledge, this is the first study describing the stabilization of TRAIL protein with a synthetic N-terminal fusion domain based on the C-terminal Foldon domain of the Fibritin T4 bacteriophage. This recombinant protein (sfTRAIL) was produced in cell culture supernatants of transfected/transduced Vero cells. Although Vero cells are not ideal for secreted recombinant protein production, this approach allowed a rapid proof-

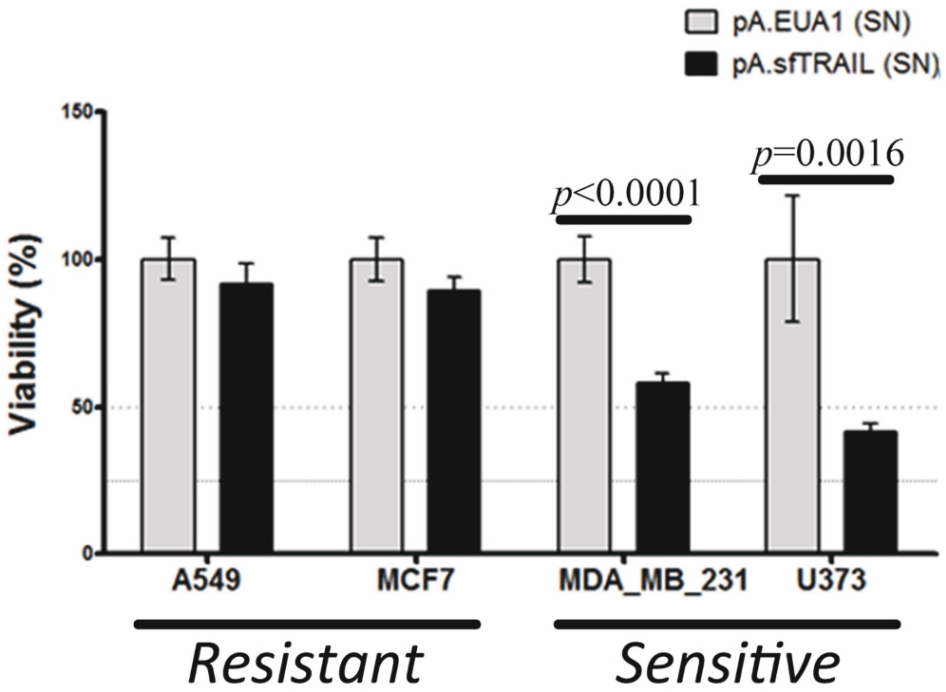


Figure 2: Cell viability (MTS) of cells incubated with sfTRAIL supernatants. Values shown are the mean value (± SDs) of triplicates. (SN) means supernatants. sfTRAIL concentration in supernatants was 1899.29 pg/mL. Data represents three experiments performed in triplicate. p-value were assessed by Student's *t*-test.

of-concept validation. Therefore, it is worth to note that sfTRAIL protein concentration may be further improved.

Regardless of the amount of protein produced, sfTRAIL protein was able to induce apoptosis at picomolar concentrations, and reduction in cell viability at femtomolar concentrations, without any special biochemical stabilization requirement. In this work, we produced recombinant TRAIL proteins simple by filtering cell culture supernatants. This may represent a major advantage over currently used recombinant TRAIL proteins available, even reducing its toxic side effects and improving patient outcome. Bioactive concentrations of current available soluble TRAIL variants are usually tested at concentrations of ng/mL or even μg/ml [10, 20–22].

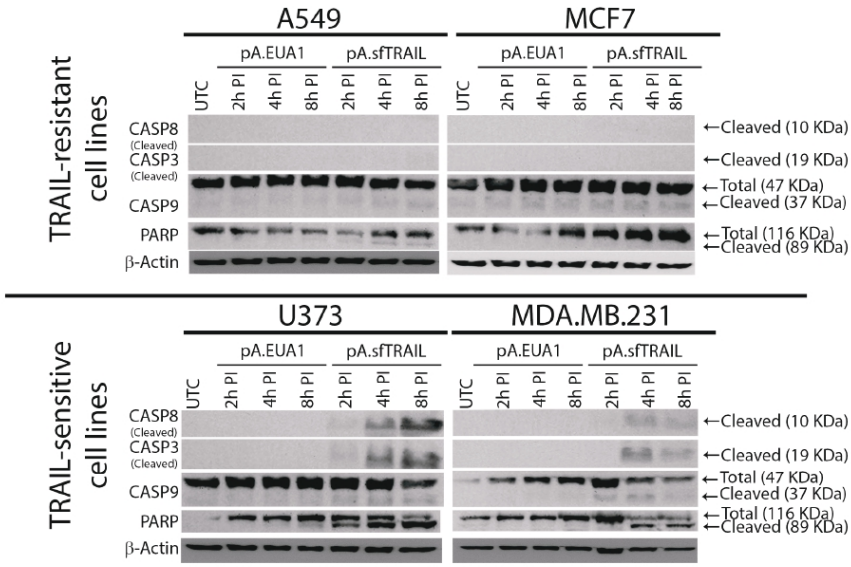


Figure 3: Western blot analysis of sfTRAIL-mediated apoptosis induction. sfTRAIL-mediated apoptosis induction of A549, MCF7, U373 and MDA-MB-231 cells, cultured in conditioned media, at 2h, 4h and 8h post-induction (PI). UTC, untreated control.

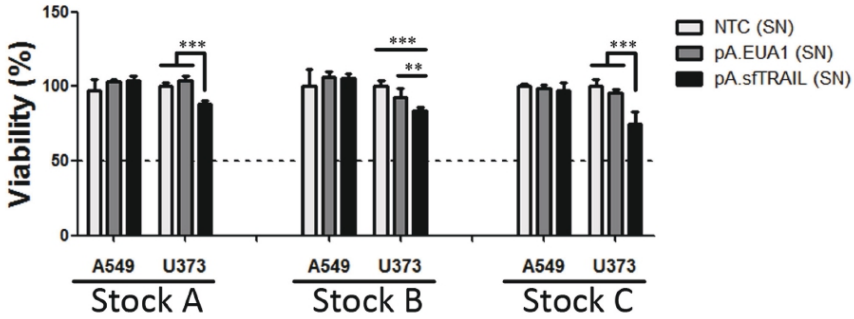


Figure 4: Cell viability (MTS) of cells incubated with sfTRAIL supernatants. Values shown are the mean value (± SDs) of three experiments performed in triplicate (Stock A, B and C). (SN) means supernatants; NTC (SN) represents supernatants of non-transduced cells; * indicates p-value (Student's *t*-test), where ** means p-value=0.0032 and *** p-value<0.0001.

Table 1: ELISA quantification of sfTRAIL protein in non-transduced controls (NTC), pA.EUA1 or pA.sfTRAIL-transduced supernatants

Samples	Concentration (pg/mL)
NTC_A	-0.48
NTC_B	-0.68
NTC_C	-3.17
pA.EUA1_A	-2.34
pA.EUA1_B	-6.02
pA.EUA1_C	-6.97
pA.TRAIL_A	**84.80**
pA.TRAIL_B	**68.45**
pA.TRAIL_C	**81.71**

A, B and C identify the biological triplicate of the experiment

The antitumor capabilities of the recombinant TRAIL variants here described were tested in four cancer cell lines with known TRAIL-sensitivity/resistance profiles, A549 (resistant), MCF7 (moderate-sensitive), U373 (sensitive) and MDA.MB.231. Although the MCF7 cell line was previously described as resistant, in our work it showed moderate sensitivity [10, 23]. So far, the bystander effect, transferred in the conditioned medium, was observed in the U373, MDA.MB.231 and MCF7 cell lines, not affecting the A549 cell line.

In conclusion, we showed that sfTRAIL protein exhibit a potent pro-apoptotic effect, triggering cell death at pico and femtomolar concentrations. In addition, as it is not necessary to further purify the supernatants, its low production requirement should reduce its final cost.

MATERIALS AND METHODS

Cell culture

Vero (African green monkey kidney), Vero-7b (Vero-derived cell line expressing ICP4/ICP27) [24], Gli36 (glioblastoma) were kindly provided by Dr. Alberto Epstein [25]. A549 (lung cancer) cell line was obtained from ECACC. Jurkat (leukemia), MCF7 (breast cancer), MDA.MB.231 (breast cancer) and HeLa (cervical cancer) cell lines were obtained from ATCC. U373 (glioma) cell line was kindly given by Dr. Joseph Costello. Cell lines were cultured in Dulbecco's minimum essential medium (DMEM; Invitrogen, USA) supplemented with 10% fetal bovine serum (FBS; Invitrogen, USA), and 1% antibiotic-antimycotic solution (Invitrogen, USA). Vero-7b cells were selected with 1 mg/mL of G418 (Sigma-Aldrich, USA) at every three passages. All cell lines were maintained in a humidified 37°C atmosphere of 95% air and 5% CO_2.

Cell line authentication

Authentication of A549, MCF7, U373 and MDA.MB.231 cell lines was performed by short tandem repeat (STR) DNA typing according to the International Reference Standard for Authentication of Human Cell Lines using a panel of 8 (D5S818, D13S317, D7S820, D16S539, vWA, TH01, TPOX and CSF1P0) STR loci plus gender determination (AMEL), using the fluorescent labeling primers as reported by Dirks et al [26]. Briefly, 50ng of DNA was amplified in a multiplex PCR reaction carried out in a total volume of 10uL with the Qiagen multiplex kit (Qiagen) comprising 0.5 μM of all fluorescent primer pairs plus 1 μM of TH01 primer reinforcement. DNA amplification products were diluted 1:100 in ultrapure nuclease free water (USB, Cleveland, USA) and combined with 0.3 μL of the internal size standard 500ROX (Applied Biosystems, Foster City, CA, USA) in 8.7 μL of formamide and loaded automatically in a capillary electrophoresis system using the Genetic Analyzer ABI PRISM 3500 (Applied Biosystems). The analysis was performed in the GeneMapper software version 4.1 (Applied Biosystems). Genotyping confirmed the identity of all cell lines, with the exception of U373, which was shown to be a sub-clone of U251 cell line.

Plasmid construction

cDNA from Jurkat cells was used to amplify the extracellular domain of the human TRAIL (aa 114-281). Total RNA was obtained from 1×10^6 cells and converted in cDNA using RevertAid H Minus Reverse Transcriptase (Thermo Scientific) and oligo(dT)18 primer. The resulting cDNA was subsequently PCR amplified with the Phusion High-Fidelity DNA Polymerase (Thermo Scientific) using the primers sTRAIL_F and sTRAIL_R (Table 1) and the 550 bp PCR fragment was gel-purified, digested with EcoRI (Thermo Scientific) and ligated with T4 DNA ligase (Thermo Scientific) at the EcoRI restriction site of pCI.Neo plasmid (Promega), generating the pCI.Neo.sTRAIL plasmid.

The coding sequence of the Fibritin foldon domain, of the T4 Bacteriophage, was humanized using Sequence Manipulation Suite [27]. Both Foldon_F and Foldon_R ssDNA oligos were hybridized and ligated into the NheI site of pCI.Neo.sTRAIL, with T4 DNA ligase (Thermo Scientific). The resulting plasmid, named pCI.Neo.sfTRAIL, was sequenced to confirm the correct fusion gene structure. Escherichia coli DH5α cells (New England Biolabs, USA) were used for cloning experiments and plasmid propagation. Bacterial strains were routinely

grown at 37°C in Luria-Broth (Sigma-Aldrich) or LB-agar (Sigma-Aldrich) containing medium and supplemented with 100 μg/mL ampicillin (Sigma).

HSV-1-derived amplicon vector production

Amplicon plasmid pA.EUA1 [28] containing one HSV-1 replication origin and one HSV-1 package signal "a" was used to derive the pA.sfTRAIL amplicon plasmid. Firstly, pCI.Neo.sfTRAIL was digested with AseI and DraI (New England Biolabs), generating an 1804 bp fragment. After being blunt-ended with the Klenow enzyme (Thermo Scientific), this 1804 bp AseI/DraI fragment was subcloned into the SnaBI (New England Biolabs) site of the pA.EUA1 plasmid. The resulting plasmid, named pA.sfTRAIL, was used to produce amplicon vectors, as described below.

The defective cre-loxP-based helper virus, HSV-1 LaLΔJ [25] was used as helper to produce amplicon vectors that were propagated and titrated in Vero-7b cells. Amplicon vector stocks were produced in T25 flasks containing 3×10^6 Vero-7b cells, previously transfected with pA.EUA1 or pA.sfTRAIL plasmids, infected at multiplicity of infection (MOI) of 0.1 plaque forming unit (PFU)/cell in Medium 199 (Invitrogen, USA) supplemented with 1% FBS (M199 1% FBS). When complete cytophatic effect (CPE) was observed (48~72 h post-infection), viruses were harvested. Thus, infected cells were centrifuged at 1,000 × g for 10 min at 4°C. The formed pellet was diluted in 400 μL of M199 1% FBS and frozen/thawed three times to break down the infected cells and facilitate the viral particles release. The pellet solution was clarified at 1,000 × g for 10 min at 4°C, aliquoted and stored at -80°C until use. Amplicon/helper particles were tittered by plaque assay [25] in Gli36 and Vero-7b cells respectively.

Western blot analysis

Vero cells were seeded in 6 well plate at a density of 5×10^5 cells per well. Twenty-four hours later, cells were transduced with pA.EUA1 or pA.sfTRAIL amplicon particles at MOI1. After 48 h, supernatants and cells were collected separately. Supernatants were filtered in 0.22 μm filters. Cells were rinsed in ice-cold PBS then scraped and lysed in lysis buffer (50mM Tris pH7.6–8, 150mM NaCl, 5mM EDTA, 1mM Na3VO4, 10mM NaF, 10mM sodium pyrophosphate, 1% NP-40, and protease cocktail inhibitors). Western blotting was performed using standard SDS-PAGE gel, loading 20 μg of protein per lane. The antibodies used were: rabbit polyclonal anti-TRAIL (abcam, #ab2056), rabbit monoclonal anti-caspase 8 (Cell Signaling, #4790), rabbit polyclonal anti-caspase 3 (Cell Signaling, #9662), mouse monoclonal anti-caspase 9 (Cell Signaling, #9508), rabbit polyclonal anti-PARP (Cell Signaling, #9542) and β-actin (Cell Signaling, #3700). After incubation with the appropriate horseradish peroxidase (HRPO)-conjugated secondary antibody (Cell Signaling, dilution 1:5000), the immunoreactivity was visualized by chemiluminescence (Amersham ECL Western Blotting Detection Reagent, GE Healthcare).

Electroporation

Electroporations of 8×10^6 HeLa cells were carried out with two pulses of 200 V and 850 μF, in 2 mm cuvettes, using Gene Pulser II (Bio-rad). Electroporated cells were seeded in T75 cell culture flasks and further maintained in a humidified 37°C atmosphere with 5% CO_2. After 48 hours of incubation, supernatants were cleared by centrifugation followed by filtration with 0.22 μm filters and stored at 4°C.

ELISA protein quantification

sfTRAIL protein contained in filtered supernatants were quantified by ELISA, using Human TRAIL/TNFSF10 Quantikine ELISA kit (R&D Systems, Minneapolis, MN, USA), following the manufacturer's instructions. Serially diluted rhTRAIL (R&D Systems, Minneapolis, MN, USA) was used for indirect quantification. It is worth to note that along the experiments, sfTRAIL protein was produced in two opportunities. Protein quantification was performed by ELISA at each sfTRAIL production.

Cell viability assay

To determine the effect of sfTRAIL protein in cellular viability over time, A549, MCF7, MDA.MB.231 and U373 cells were seeded into 96-well plates, at a concentration of 5×10^3 cells per well, as previously described [29]. After 24 hours, cells were cultured in 0.22 μm-filtered conditioned media. After 48 hours of incubation, viable cells were quantified using MTS assay (CellTiter 96® AQueous One Solution Cell Proliferation Assay, Promega). Assays were performed in triplicate and repeated three times. In order to analyze the stability of the sfTRAIL protein, filtered supernatants were maintained at 4°C during 60 days, without any conservator. After this period, all four cell lines were rechallenged to TRAIL induced apoptosis and assayed by MTS analysis.

Apoptosis and TRAIL receptors detection by flow cytometry

For apoptosis analysis, cells were incubated with rhTRAIL (R&D Systems, Minneapolis, MN, USA) at 100 ng/mL for 4 hours. After incubation, cells were washed in cold PBS and apoptosis was assessed by flow cytometry using the Annexin V-PE/7-AAD apoptosis detection kit (BD Biosciences), according to the manufacturer's protocol. Briefly, 10^5 cells in 100 μL of 1x binding buffer were incubated with 5 μL Annexin V-PE and 5 μL 7-AAD

reagent for 15 minutes, at room temperature in the dark. After incubation 400 μL of 1x binding buffer were added and tubes were directly analyzed in a FACS Aria II (BD Bioscience).

Expression of TRAIL receptors was also evaluated by flow cytometry. Thus, incubation with antibodies was performed as described above, using human anti-DR4-PE (R&D Systems - Clone 69036) and anti-DR5-PE (R&D Systems - Clone 71908) antibodies. Non-stained cells were used as controls, setting gates at 1% positivity for comparisons.

Statistical analysis

The differences between groups were evaluated by calculating Student's *t*-test, one-way ANOVA with post hoc comparison Tukey's test. For all tests, p-value < 0.05 were considered statistically significant.

ACKNOWLEDGMENTS

The authors would like to thank to *Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP), and to *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq). MEM, LMRBA and ACL were recipient of FAPESP fellowship. ALC has a CNPq scholarship.

CONFLICTS OF INTEREST

Authors MEM and ALC have a patent for this technology.

GRANT SUPPORT

This project was supported by FAPESP (Grant 2012/04697-3).

REFERENCES

1. Savill J, Fadok V. Corpse clearance defines the meaning of cell death. Nature. 2000; 407:784–8. https://doi.org/10.1038/35037722.
2. Ashkenazi A, Holland P, Eckhardt SG. Ligand-based targeting of apoptosis in cancer: the potential of recombinant human apoptosis ligand 2/Tumor necrosis factor-related apoptosis-inducing ligand (rhApo2L/TRAIL). J Clin Oncol. 2008; 26:3621–30. https://doi.org/10.1200/JCO.2007.15.7198.
3. Wiezorek J, Holland P, Graves J. Death receptor agonists as a targeted therapy for cancer. Clin Cancer Res. 2010; 16:1701–8. https://doi.org/10.1158/1078-0432.CCR-09-1692.
4. Holland PM. Death receptor agonist therapies for cancer, which is the right TRAIL? Cytokine Growth Factor Rev. 2014; 25:185–93. https://doi.org/10.1016/j.cytogfr.2013.12.009.
5. Hymowitz SG, O'Connell MP, Ultsch MH, Hurst A, Totpal K, Ashkenazi A, de Vos AM, Kelley RF. A unique zinc-binding site revealed by a high-resolution X-ray structure of homotrimeric Apo2L/TRAIL. Biochemistry. 2000; 39:633–40. https://doi.org/10.1021/bi9922421.
6. Verbrugge I, Johnstone RW, Smyth MJ. SnapShot: Extrinsic apoptosis pathways. Cell. 2010; 143:1192. e1–2. https://doi.org/10.1016/j.cell.2010.12.004.
7. LeBlanc HN, Ashkenazi A. Apo2L/TRAIL and its death and decoy receptors. Cell Death Differ. 2003; 10:66–75. https://doi.org/10.1038/sj.cdd.4401187.
8. Ashkenazi A, Pai RC, Fong S, Leung S, Lawrence DA, Marsters SA, Blackie C, Chang L, McMurtrey AE, Hebert A, DeForge L, Koumenis IL, Lewis D, et al. Safety and antitumor activity of recombinant soluble Apo2 ligand. J Clin Invest. 1999; 104:155–62. https://doi.org/10.1172/JCI6926.
9. Jin CY, Moon DO, Lee JD, Heo MS, Choi YH, Lee CM, Park YM, Kim GY. Sulforaphane sensitizes tumor necrosis factor-related apoptosis-inducing ligand-mediated apoptosis through downregulation of ERK and Akt in lung adenocarcinoma A549 cells. Carcinogenesis. 2007; 28:1058–66. https://doi.org/10.1093/carcin/bgl251.
10. Zhang Y, Zhang B. TRAIL resistance of breast cancer cells is associated with constitutive endocytosis of death receptors 4 and 5. Mol Cancer Res. 2008; 6:1861–71. https://doi.org/10.1158/1541-7786.MCR-08-03136/12/1861.
11. Yoshida T, Zhang Y, Rivera Rosado LA, Zhang B. Repeated treatment with subtoxic doses of TRAIL induces resistance to apoptosis through its death receptors in MDA-MB-231 breast cancer cells. Mol Cancer Res. 2009; 7:1835–44. https://doi.org/10.1158/1541-7786.MCR-09-0244.
12. Chen JJ, Knudsen S, Mazin W, Dahlgaard J, Zhang B. A 71-gene signature of TRAIL sensitivity in cancer cells. Mol Cancer Ther. 2012; 11:34–44. https://doi.org/10.1158/1535-7163.MCT-11-0620.
13. Plummer R, Attard G, Pacey S, Li L, Razak A, Perrett R, Barrett M, Judson I, Kaye S, Fox NL, Halpern W, Corey A, Calvert H, et al. Phase 1 and pharmacokinetic study of lexatumumab in patients with advanced cancers. Clin Cancer Res. 2007; 13:6187–94. https://doi.org/10.1158/1078-0432.CCR-07-0950.
14. Wakelee HA, Patnaik A, Sikic BI, Mita M, Fox NL, Miceli R, Ullrich SJ, Fisher GA, Tolcher AW. Phase I and pharmacokinetic study of lexatumumab (HGS-ETR2) given every 2 weeks in patients with advanced solid tumors. Ann Oncol. 2010; 21:376–81. https://doi.org/10.1093/annonc/mdp292.
15. Pitti RM, Marsters SA, Ruppert S, Donahue CJ, Moore A, Ashkenazi A. Induction of apoptosis by Apo-2 ligand, a new member of the tumor necrosis factor cytokine family. J Biol Chem. 1996; 271:12687–90.

16. Walczak H, Miller RE, Ariail K, Gliniak B, Griffith TS, Kubin M, Chin W, Jones J, Woodward A, Le T, Smith C, Smolak P, Goodwin RG, et al. Tumoricidal activity of tumor necrosis factor-related apoptosis-inducing ligand *in vivo*. Nat Med. 1999; 5:157–63. https://doi.org/10.1038/5517.

17. Bodmer JL, Meier P, Tschopp J, Schneider P. Cysteine 230 is essential for the structure and activity of the cytotoxic ligand TRAIL. J Biol Chem. 2000; 275:20632–7. https://doi.org/10.1074/jbc.M909721199.

18. Mohr A, Lyons M, Deedigan L, Harte T, Shaw G, Howard L, Barry F, O'Brien T, Zwacka R. Mesenchymal stem cells expressing TRAIL lead to tumour growth inhibition in an experimental lung cancer model. J Cell Mol Med. 2008; 12:2628–43. https://doi.org/10.1111/j.1582-4934.2008.00317.x.

19. Ashkenazi A, Dixit VM. Apoptosis control by death and decoy receptors. Curr Opin Cell Biol. 1999; 11:255–60. https://doi.org/10.1016/S0955-0674(99)80034-9.

20. Hougardy BM, Reesink-Peters N, van den Heuvel FA, ten Hoor KA, Hollema H, de Vries EG, de Jong S, van der Zee AG. A robust ex vivo model for evaluation of induction of apoptosis by rhTRAIL in combination with proteasome inhibitor MG132 in human premalignant cervical explants. Int J Cancer. 2008; 123:1457–65. https://doi.org/10.1002/ijc.23684.

21. Meijer A, Kruyt FA, van der Zee AG, Hollema H, Le P, ten Hoor KA, Groothuis GM, Quax WJ, de Vries EG, de Jong S. Nutlin-3 preferentially sensitises wild-type p53-expressing cancer cells to DR5-selective TRAIL over rhTRAIL. Br J Cancer. 2013; 109:2685–95. https://doi.org/10.1038/bjc.2013.636.

22. Jalving M, de Jong S, Koornstra JJ, Boersma-van Ek W, Zwart N, Wesseling J, de Vries EG, Kleibeuker JH. TRAIL induces apoptosis in human colorectal adenoma cell lines and human colorectal adenomas. Clin Cancer Res. 2006; 12:4350–6. https://doi.org/10.1158/1078-0432.CCR-05-2487.

23. Bozza WP, Tolleson WH, Rosado LA, Zhang B. Ricin detection: tracking active toxin. Biotechnol Adv. 2015; 33:117–23. https://doi.org/10.1016/j.biotechadv.2014.11.012.

24. Krisky DM, Marconi PC, Oligino T, Rouse RJ, Fink DJ, Glorioso JC. Rapid method for construction of recombinant HSV gene transfer vectors. Gene Ther. 1997; 4:1120–5. https://doi.org/10.1038/sj.gt.3300497.

25. Zaupa C, Revol-Guyot V, Epstein AL. Improved packaging system for generation of high-level noncytotoxic HSV-1 amplicon vectors using Cre-loxP site-specific recombination to delete the packaging signals of defective helper genomes. Hum Gene Ther. 2003; 14:1049–63. https://doi.org/10.1089/104303403322124774.

26. Dirks WG, Faehnrich S, Estella IA, Drexler HG. Short tandem repeat DNA typing provides an international reference standard for authentication of human cell lines. ALTEX. 2005; 22:103–9.

27. Stothard P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques. 2000; 28:1102, 4.

28. Adrover MF, Guyot-Revol V, Cheli VT, Blanco C, Vidal R, Alche L, Kornisiuk E, Epstein AL, Jerusalinsky D. Hippocampal infection with HSV-1-derived vectors expressing an NMDAR1 antisense modifies behavior. Genes Brain Behav. 2003; 2:103–13.

29. Silva-Oliveira RJ, Silva VA, Martinho O, Cruvinel-Carloni A, Melendez ME, Rosa MN, de Paula FE, de Souza Viana L, Carvalho AL, Reis RM. Cytotoxicity of allitinib, an irreversible anti-EGFR agent, in a large panel of human cancer-derived cell lines: KRAS mutation status as a predictive biomarker. Cell Oncol (Dordr). 2016; 39:253–63. https://doi.org/10.1007/s13402-016-0270-z.

6.2. Segundo Artigo

Esse artigo encontra-se em revisão na *Environment International*, revista científica classificada pela CAPES como Qualis A1 (Fator de impacto: 13.4), e a candidata ao título de doutora é a quinta autora do trabalho.

Nesse trabalho foi possível solidificar e ampliar as habilidades na análise de metilação em larga escala.

DNA Methylome-Wide Alterations Associated with Waterpipe and Cigarette Smoking in Cancer Development

Awada Z.[1], Cahais V.[1], Cuenin C.[1], Akika R.[2], Vicente A.L.[1], Makki M, Tamim H.[3], Herceg Z.[1], Zgheib N.K.[2*], Ghantous A.[1*]

[1]Epigenomics and Mechanisms Branch, International Agency for Research on Cancer, Lyon, France

[2]Department of Pharmacology and Toxicology, Faculty of Medicine, American University of Beirut, Beirut, Lebanon

[3]Department of Internal Medicine and Clinical Research Institute, American University of Beirut Faculty of Medicine, Beirut, Lebanon.

*Co-corresponding

Author Contribution

Conceptualization (Z.A., N.K.Z., A.G.); methodology (Z.A., N.K.Z., Z.H., A.G.); sample and data acquisition and processing (H.T., N.K.Z., M.M., R.A.); experiments (C.C., Z.A.); analysis (Z.A., V.C., V.A.L., A.G.); interpretation (Z.A., N.K.Z., A.G.); writing of first draft (Z.A., N.K.Z., A.G.); funding (N.K.Z., Z.H., A.G.); supervision (N.K.Z., A.G.). All authors read and approved the final version of the manuscript.

Running title: Epigenetic markers of waterpipe and cigarette smoking

Key terms: waterpipe, cigarette, DNA methylation, biomarker, câncer

Disclaimer: "Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization."

ABSTRACT

Waterpipe smoking, historically widespread in the Middle East and North Africa, is becoming increasingly prevalent in western countries. Like cigarette, it is associated with increased risk of several cancers, yet the underlying mechanisms require elucidation. The epigenome functions as a molecular sensor of exposure and a central player in cancer. Thus, epigenetic biomarkers of waterpipe smoking and their role in cancer development are of interest. DNA methylome-wide profiling (850K array) was performed on 96 blood samples split equally between never, cigarette-only and waterpipe-only smokers from a cohort in Lebanon. Dimension reduction-based analysis yielded 160 and 87 Differentially Methylated CpG Regions (DMRs) for cigarette and waterpipe smoking, respectively, with effect size ≥ 5%, FDR < 0.05 and power > 90%. Top CpG markers were replicated by pyrosequencing in 194 additional samples, with time-dose response analysis showing irreversibility of some markers over time. Minimal overlap was observed between waterpipe and cigarette smoking markers, and artificial intelligence-based epigenome mapping accurately (> 92%) predicted waterpipe *versus* cigarette smoking status. Functional genomics showed enrichment of waterpipe and cigarette smoking markers in regulatory regions and cancer hallmarks, specifically in genes epigenetically deregulated in lung, breast and/or pancreatic cancers based on in-depth screening of the DNA methylome of 11,258 tumor and 1,106 normal tissues spanning 33 tumor types from The Cancer Genome Atlas (TCGA). Among them, several genes were predicted to play cancer driver roles based on integrated methylation, transcription, copy number and mutation alterations, and to affect patient prognosis. Our findings validate previously reported epigenetic markers of cigarette smoking, identify novel ones through increased genomic coverage, and report for the first time DNA methylome-wide markers for waterpipe smoking, with translationally impactful mechanisms in carcinogenesis. These results set a basis for the quest for robust epigenetic biomarkers of different forms of tobacco exposure.

INTRODUCTION

Waterpipe also referred to as "*narghile"*, "*arguile*", "*hookah"*, "*shisha"*, or *"hubble bubble"* is a tobacco-use method that was historically prevalent in India, Middle East and North Africa in the 1990s (1). The past two decades have witnessed an exponential rise in waterpipe use to the extent that it surpassed cigarette use among young people in the Eastern Mediterranean region (2, 3). In parallel, several epidemiological studies reported global spread in waterpipe use in other countries such as Canada, European countries and US, particularly among the youth (4-7).

Despite the common unsubstantiated perception that waterpipe is less harmful than cigarette, waterpipe involves roughly 10 times greater number of puffs and 100 times larger smoke volume when compared to cigarette (8). A meta-analysis of 17 studies showed that the inhaled amounts of nicotine, tar and carbon monoxide were higher from a single waterpipe session than from a single cigarette (9). Chemical characterization of waterpipe smoke revealed significant concentrations of several toxicants previously identified in cigarette smoke, including 27 established or presumed carcinogens (10). Although the adverse effects of waterpipe are less investigated than cigarette, available studies reported that similarly to cigarette, long-term use of waterpipe is associated with lung, esophageal, and head and neck cancers (11), as well as pulmonary and cardiovascular diseases (12).

DNA methylation is a well-characterized epigenetic mechanism that involves the addition of methyl groups to the cytosine residues of DNA and increases the likelihood of silencing of gene expression when present in gene promoters (13). Despite their potentially reversible nature, DNA methylation changes can sometimes serve as long-term biomarkers for environmental exposures (14, 15). Large-scale studies, including our own, characterized several DNA methylation changes in peripheral blood as biomarkers for cigarette smoking (15-17). Some of these biomarkers continue to capture the exposure several years after cessation of smoking and may explain the long-term adverse effects of smoking observed decades after cessation (15, 18).

Moreover, studying the DNA methylation alterations associated with cigarette smoking has allowed better understanding of the molecular mechanisms underlying cigarette-

associated health outcomes. As such, some DNA methylation signatures induced by cigarette smoking were reported to be enriched in lung cancer tissues (19, 20) and associated with lung cancer risk, incidence or mortality (21-24). In addition, it was postulated that *aryl hydrocarbon receptor repressor* (*AHRR*) hypomethylation and overexpression induced by cigarette smoking inhibit the detoxification pathway mediated by the aryl hydrocarbon receptor (AHR), thereby, leading to accumulation of environmental pollutants and increase in cancer risk among smokers (25). In contrast to the multiple studies in the literature investigating the epigenetic effects of cigarette smoking, no study has so far investigated epigenome-wide associations with waterpipe smoking in accessible human tissues, such as blood. With the rapid increase in the prevalence of waterpipe use, it is timely to investigate waterpipe-associated DNA methylation alterations and their role in disease development.

To screen for DNA methylation markers associated with cigarette and waterpipe smoking, we performed microarray-based genome-wide methylation profiling on DNA extracted from peripheral blood of never, current cigarette-only and current waterpipe-only smokers (Figure 1; Discovery set). Top findings were technically validated by pyrosequencing and replicated in additional samples (Figure 1; Validation set), including their potential to accurately predict each of the two tobacco smoking forms. Functional genomics analysis of the identified tobacco markers was evaluated in relation to overall age acceleration, mortality risk, and cancer development, with in-depth investigation of their effects on cancer hallmarks, driver events and patient prognosis, based on pan-cancer multi-omics analysis of 11,258 tumor and 1,106 normal tissues spanning 33 cancer types (Figure 1).

MATERIALS AND METHODS

Recruitment and study population

Five hundred and one participants from several districts of Greater Beirut area were recruited by the American University of Beirut between February and June 2014 (26). All subjects signed an informed consent form in alignment with the institutional review board rules and regulations. The majority of participants agreed that their samples be used in further

studies and shared by external investigators. Blood samples were withdrawn from all individuals, and most of the participants had remaining blood samples at the time of the study (see Figure 1).

Anthropometric measurements (height, weight, etc.) were taken at the time of recruitment. In addition, data related to participants' demographics, lifestyle-related characteristics and medical history were collected using a multicomponent questionnaire. Information about smoking status, duration, and frequency (i.e., number of cigarettes or waterpipe sessions per day) of participants were abstracted from the questionnaire. Out of 501 samples, 476 consented their blood be used in further studies, were cancer-free and had available blood samples (Figure 1). In the overall cohort, waterpipe-only smokers were younger than cigarette-only and never- smokers (Supp Figure 1A), and there were more females who are never or current smokers than males (Supp Figure 1B).

Power analysis

Statistical power was estimated based on the DNA methylome data as such: the overall mean standard deviation (SD) of methylation probes in our data is 0.026 (for methylation values ranging 0-1). Based on an alpha of 4.52×10^{-7} (Bonferroni significance threshold for 110,560 regions which is roughly the total number of regions with at least two CpGs separated by a maximum gap of 1000 bases, as used in the analysis of differentially methylated CpG regions (DMRs), we will have > 90% power to detect an effect size ≥ 5% with 32 exposed and 32 non-exposed subjects (Figure 1).

The 5% is the minimum accepted methylation change of at least one CpG in the DMR based on our filtration criteria. In addition, power analysis was performed based on the R^2 coefficient of determination retrieved from the robust linear regression model of every CpG in the array in association with smoking status (Supp Figure 2). Similarly, results showed that this model when performed on a sample size of 32 individuals per group has a power ranging from 80 to > 90%, contingent on the CpG (Supp Figure 2).

DNA isolation and bisulfite conversion

DNA was isolated from peripheral blood using the Flexigene DNA isolation kit from Qiagen (cat # 51206, Hilden, Germany) as per manufacturer's protocol and stored in aliquots at -20°C. DNA was quantified using Quant-iT PicoGreen® dsDNA reagent (Thermo Fisher Scientific, P7589) as per manufacturer's instructions. Then, 600 ng of DNA was bisulfite converted using EZ DNA Methylation Kit (Zymo Research, CA, USA) as per manufacturer's protocol, and its efficiency was tested using PCR primers that were specific to bisulfite-converted DNA and spanning the region of *RASSFF1A* gene (Supp Table 2). All samples passed this quality control step.

Genome-wide methylation analysis using Infinium Methylation EPIC BeadChip arrays (Discovery set)

Discovery set design

The Discovery set constituted of DNA methylome-wide array profiling of 32 blood samples taken from each of the never, current cigarette-only and current waterpipe-only smokers who were non-alcohol drinkers (Figure 1; Table 1). The sample sizes were determined based on the power estimates described earlier, and the samples were selected by age and sex frequency matching of the three groups, particularly that age and sex influence both the type and extent of smoking. Hence, each group consisted of 20 females and 12 males representing a similar female/male ratio as the entire cohort (Figure 1 and Supp Figure 1).

Statistical analysis confirmed no FDR-significant difference in age and sex distribution among the never, current cigarette-only and current waterpipe-only smokers (Table 1). There were also no detectable FDR-significant differences in BMI, coffee intake and physical activity between each of cigarette and waterpipe smokers *versus* never smokers (Table 1).

DNA methylome-wide data Generation

Genome-wide methylation analysis was performed using the Infinium Methylation EPIC microarray (850K array) that covers over 850,000 CpGs, following manufacturer's protocol (Illumina Inc., San Diego, CA, USA) and optimized in-house for high-throughput

analyses through an automated robotic system (Freedom EVO 150 by Tecan, Mannedorf, Switzerland) that can process the chips with minimal human error. Each chip covers eight samples, and samples on chips were semi-randomized so that batch effects (e.g. sample position and intra and inter-variability in arrays and chips) do not completely confound with biological covariates of interest. Each chip included similar proportions of each of never, cigarette and waterpipe smokers as well as a male: female ratio of 3: 5 reflective of the cohort proportions. Bisulfite-converted DNA (250 ng) was used for hybridization on Infinium Methylation EPIC bead arrays, following the manufacturer's protocol. Chips were scanned using Illumina iScan to produce two-color raw data files (.idat format).

DNA methylome analyses

1. DNA methylome data pre-processing

Data files obtained from array analysis in ".idat format" were analyzed using R/Bioconductor packages (http://www.bioconductor.org/). Raw data constituted of 866,091 CpG probes from the 850K array and were loaded from IDAT files and handled in the minfi package (27). Red and green IDAT files were transformed into methylation values (betas, β); β = intensity of the methylated allele (M)/[intensity of the unmethylated allele (U) + intensity of the methylated allele (M) + 100]. Data quality was assessed using box plots for the distribution of methylated and unmethylated signals, and array-wide multi-dimensional scaling plots and unsupervised clustering was used to identify potential sample outliers (none observed).

Multidimensional scaling plots clustered samples into two groups representing sex status (confirming that recorded sex was correct). Then, using minfi package, the probes were filtered from 850K cross-reactive probes (n = 43254), probes overlapping genetic variants at targeted CpG sites (n = 414) or at the probe body (n = 12510), and XY chromosome probes (n = 19681) (28), resulting in a total of 791,494 CpGs to analyze. Beta values were then normalized with Functional normalization (Funnorm, minfi R package) that was shown to perform equally good or outperform existing normalization methods (29). It removes

unwanted variation by regressing out variability explained by the control probes present on the array.

The distribution of β values became more homogenous after filtration and normalization of data (Supp Figure 3A). Filtered and normalized β values were log transformed to M values (calculated as the $\log_2$ ratio of the intensities of methylated probe *versus* unmethylated probe). Finally, we used surrogate variable analysis (SVA) (30) coupled to Bonferroni adjustment for multiple testing to correct for batch effects and adjust for latent variables, a choice validated by the findings of our benchmarking (31). SVA also increases statistical power by removing (unwanted) variability through aggregating information at the data level and constraining the data's variability to the phenotype of interest (32), being smoking status in this case. Our detailed genome-wide methylation analysis code is available on GitHub: https://github.com/IARCbioinfo/methylkey.

Principal component analysis (PCA) was performed before and after SVA correction of β values in order to characterize overall and particular effects of variables including sentrix ID, sentrix position, sex, age, smoking status, BMI, coffee intake, physical activity and blood cell composition on DNA methylation results. The overall proportion of variance explained on the first ten principal components (PCs) was decreased after SVA correction of β values (Supp Figure 3B). In addition, after SVA, all analyzed variables were statistically non-significantly associated with the 10 PCs, except B lymphocyte proportion which was significantly associated with PC-5 (Bonferroni-adjusted P value = 0.034).

Hence, our statistical regression models consisted of a crude model and another one adjusted for B lymphocyte proportion (in addition to age and sex) (Supp Figure 4A). Even though there was no detectable significant effect of age and sex on the PCs after SVA correction, we still included these two variables in the adjusted model to account for any potential residual confounding by age and sex, which are known to affect both smoking exposure and DNA methylation. We additionally adjusted for body mass index (BMI) in a third model. For either cigarette or waterpipe smoking, results from the three models were largely similar (Supp Figure 4A).

Hence, we reported in downstream analyses the model adjusting for age, sex and B lymphocyte proportion since additional adjustment for BMI did not substantially change the results. Methylation beta-values were used for data plotting and interpretation. However, the distribution of beta-values is bimodal, so their logarithmic transformation (M-value) was used for most statistical analyses due to its approximate parametric distribution (33).

2. Assessment of cell mixture distribution

White blood cell (WBC) composition was derived from the IDAT files using the FlowSorted.Blood.EPIC package version 1.2.0 which applies the Houseman model (34). Proportions of six cell types (B, CD4+ T, CD8+ T, granulocytes, monocytes, natural killers (NK)) were determined. SVA was also used as a reference-free method to adjust for differences in WBC composition (35). Residual effects from specific WBC types that persisted after SVA (Supp Figure 3B) were adjusted for in the subsequent regression models.

3. Dimension reduction and CpG region-level methylome-wide regression analysis

A dimension reduction approach (DMRcate) (36) was implemented to reduce the matrix size of ~850,000 individual sites into clusters of genetically proximal and correlated CpGs to enhance statistical power (32) and aid biological inferences (as single CpG sites often have more subtle functional relevance than CpG clusters), as per our previous work (15, 37). We used the recommended DMRcate proximity-based criteria and a minimum of two methylation sites per region with a maximum gap of 1000 bp.

DMRs were obtained using robust logistic regression analysis. To reduce the number of false positive results, we controlled for multiple testing using Benjamini Hochberg (FDR) procedure (38). Identified differentially methylated regions (DMRs) were considered significant if their FDR-adjusted P values were < 0.05. In both comparisons (cigarette versus *never* smokers and waterpipe *versus* never smokers), there were no major differences in the inflation factor (lambda, λ) between the three models (Supp Figure 5A-B).

DMR filtration was performed with the following filtration criteria: 1) Each DMR should include at least one CpG with $|\Delta\beta_{filtered}|$ (i.e. $|\beta_{filtered\ in\ smoker\ group} - \beta_{filtered\ in\ never\ smoker\ group}|$) at least 5%, 2) At least two-thirds of the probes with $|\ \Delta\beta_{filtered}\ | \geq 5\%$ within each DMR should have the same direction of change in $\Delta\beta_{filtered}$, and 3) All statistically significant CpGs within DMRs should show unimodal distribution when tested by dip test ($P > 0.05$). CpGs (of $|\Delta\beta_{filtered} \geq 5\%|$) within DMRs associated with cigarette or waterpipe smoking were clustered in heatmaps using Nonnegative Matrix Factorization (NMF) package version 0.22.0 on R studio version 3.6.1 (39). Z-score scaling of rows was performed after clustering of FunNorm $\beta_{filtered}$ values using Canberra distance dissimilarity measurement and Ward's hierarchal clustering.

4. Machine learning on the DNA methylome data to predict smoking status and form

We used the R package, EpiSmokEr (Epigenetic Smoking status Estimator), for prediction of current cigarette and never smoker status. This package includes a lasso regression model trained for prediction of cigarette smoking status and reported to possess high prediction accuracy in several datasets (40).

For the prediction of current and never waterpipe smokers, we performed lasso regression from the glmnet R package on the probes with at least 3% methylation difference between waterpipe and never smokers that were used as predictors, as well as sex. After that, samples were divided into training set (80%) and test set (remaining 20%). Sampling and lasso regression were repeated 1000 times, and the model with the highest accuracy in the training set and validated in the test set was reported and visualized using ggplot. For the calculation of accuracy, cigarette smokers were considered as never waterpipe smokers in this model, and waterpipe smokers were considered as never cigarette smokers in the former model predicting for cigarette smoking status.

Validation of array results by pyrosequencing and in additional samples (Validation set)

Validation set design

For validation of waterpipe-associated probes, we first calibrated the statistics of the DNA methylome-wide analysis of our Discovery set to targeted CpG-level analysis using two-sample independent t-test (Supp Table 1) and then performed pyrosequencing on the probes, cg26113488 [NA] and cg15327692 [*RASGRP1*], representing the most significantly (FDR < 0.05) hypomethylated and hypermethylated CpGs with the largest effect sizes, respectively (with the exception of the hypermethylated cg11752927 [*PTPRN2*] which could not be pyrosequenced due to a dense CpG-region).

The selected CpGs also happen to be specifically associated with waterpipe but not cigarette smoking (Supp Table 1). Moreover, the probe cg26113488 (NA) was close to another probe cg14667685 that was also associated with waterpipe smoking according to the 850K analysis, so both were analyzed by pyrosequencing (Figure 1). Validation was performed in 24 sex-matched samples of each of never, current cigarette-only and current waterpipe-only smokers. We were not able to additionally frequency match the validation samples by age, as was done for the discovery samples, due to the limited number of smokers remaining from the overall cohort.

Accordingly, we performed sensitivity analysis on the topmost probe, NA (cg26113488 and cg14667685), by pyrosequencing all cohort blood samples with sufficient DNA quantity comprising all remaining never (N= 86) and current waterpipe-only (N= 52) smokers that had not been profiled by 850K arrays and a subset of current cigarette-only smokers (N = 36); we additionally included all former waterpipe-only smokers (N= 14).

As expected, and in line with the demographics of the overall cohort, waterpipe smokers were significantly younger (FDR < 0.05) than never smokers in this set (Table 1B). There were no detectable FDR-significant differences in BMI, physical activity and alcohol consumption between each of current cigarette-only and current waterpipe-only smokers *versus* never smokers (Table 1B). As for coffee intake, it was associated with cigarette smoking in the validation set, but there was no association between coffee intake and waterpipe

smoking, in line with the expected behavioral discordance between simultaneous waterpipe and coffee use (unlike with cigarette smoking).

The validation and sensitivity analyses both confirmed that NA (cg26113488 and cg14667685) was significantly hypomethylated in current waterpipe-only relative to never and current cigarette-only smokers (Supp Figure 6A). For technical validation, we additionally pyrosequenced the three probes (cg26113488 [NA], cg14667685 [NA] and cg15327692 [*RASGRP1*]) in 5-6 samples analyzed with microarrays (Figure 1).

Pyrosequencing DNA methylation data Generation

Bisulfite converted DNA was amplified using DNA polymerase and forward and reverse primers listed in Supp Table 2. Polymerase chain reaction (PCR) involved enzyme denaturation at 95°C for 15 min, followed by 50 cycles of 95°C for 30 sec, primer annealing at 57°C for cg26113488 and cg26113488 and at 55°C for cg15327692 for 30 sec, and DNA extension at 72°C for 30 sec. After that, pyrosequencing was performed with the sequencing primers listed in Supp Table 1 using the Pyrosequencer PyroMark Q96, and signals were analyzed with the PyroMark Q96 software (41).

Pyrosequencing DNA methylation analysis

Linear regression slope and standard error were determined for each CpG by linear regression analysis with and without adjustment of age and sex. After that, region analysis was performed by inverse variance based meta-analysis method using metal software (42). Results of the samples that were assayed by both 850K arrays and pyrosequencing were evaluated for correlation by Pearson chi-square test. In representation of results, the e symbol was used which stands for “× 10”.

Time-dose analysis of DNA methylation with the extent of waterpipe smoking exposure

Time-dose analysis of DNA methylation with the extent of waterpipe smoking exposure was performed in the Validation set since it includes a larger sample size. Extent of waterpipe smoking exposure was represented as waterpipe session-year, calculated as number of waterpipe sessions per day multiplied by smoking duration in years. The methylation values of CpGs were correlated with the extent of waterpipe smoking exposure in waterpipe smokers using Pearson chi-square test. DNA methylation was also visualized in relation to tertiles of extent of waterpipe exposure, and linear regression was performed to compare DNA methylation across the tertiles. In the time analysis, we also compared current, former and never smokers based on the extended sample size and exposure coverage afforded by the Validation set.

Smoking in association with epigenetically predicted biological and clinical readouts

We predicted biological (epigenetic age acceleration*,* and *LINE-1* and *ALU* methylation as surrogates for global DNA methylation) and clinical readouts (blood cell composition and mortality risk score) from the methylome-wide data.

LINE-1 methylation at 13,057 loci and *ALU* methylation at 88,299 loci were predicted from available epigenetic information of neighboring loci using remp package version 1.8.2 on R studio version 3.6.1 (43). Average DNA methylation (average $\beta_{filtered}$) of the 791,494 filtered probes was calculated and considered as representative of global DNA methylation.

Observed DNA methylation age was derived from Horvath 2018 epigenetic clock for skin and blood cells (44) and plotted against chronological age. The following linear regression equation was derived from the plot: DNA $methylation_{predicted}$ = 0.9552 x chronological age - 1.353. Epigenetic age acceleration was calculated from the subtraction of predicted from observed DNA methylation whereby the former was obtained from the residuals of the linear regression equation.

Mortality risk score was predicted from β-values at eight CpG sites based on the following combination formula: cg01612140 x (-0.38253) + cg05575921 x (-0.92224) + cg08362785 x (2.71749) + cg10321156 x (-0.02073) + cg14975410 x (-0.04156) + cg19572487 x (-0.28069) + cg24704287 x (-2.98637) + cg25983901 x (-1.80325). The formula was obtained from Zhang *et al.* study in which the weight of every CpG site was derived from the least absolute shrinkage and selection operator (LASSO) regression (45).

Average methylation of the 850K probes, *LINE-1* and *ALU* methylation, and epigenetic age acceleration and mortality risk scores were plotted as mean + SD in each of the smoking groups. Differences in mean values across the three smoking groups were tested for statistical significance using ANOVA followed by Dunnett *post-hoc* test. In addition, the association between each of these readouts and the extent of smoking exposure was analyzed using linear regression. An FDR-adjusted P value of less than 0.05 was considered statistically significant.

Enrichment analysis of DNA methylation alterations within CpG density and regulatory regions, biological pathways and cancer hallmarks

The enrichment of probes in CpG density and regulatory regions was tested by conducting Pearson chi-square test on 2×2 contingency tables. FDR-adjusted P value less than 0.05 was considered statistically significant. Pathway enrichment analysis was performed using genecodis software (46). Specifically, genes annotating for CpGs within DMRs were tested for enrichment in Gene Ontology Biological Process (GO BP), KEGG, and Reactome databases. Statistically significant pathways with FDR < 0.05 and at least three overlapping genes were tabulated, and top 10 significant pathways were visualized using igraph on R studio. As for enrichment within cancer hallmarks, the genes annotating for all, hypomethylated and hypermethylated CpGs were tested for enrichment in the gene lists of 10 cancer hallmarks (47) using GeneOverlap R package. Hallmarks with P < 0.05 were considered statistically significant (no FDR adjustment was applied as the ten hallmarks are in principle inter-correlated).

Screening of cigarette- or waterpipe-associated CpGs within those epigenetically deregulated in multiple cancer types

DNA methylation analysis of tumor versus normal samples from various cancer types

DNA methylation (beta files), biospecimen and clinical data were downloaded for all cancer patients in The Cancer Genome Atlas (TCGA) database using the National Cancer Institute Genomics Data Common Data Portal. The following were excluded from the analysis: 1) samples with 27K data, 2) formalin-fixed paraffin embedded samples, 3) samples with prior treatment, and 4) cancer types with less than 10 normal tissue samples.

For breast cancer cases, males were excluded from the analysis since they had a very small sample size relative to females, and their exclusion decreased inflation (Supp Table 3); however, for the remaining cancer types, sex was adjusted for in the analysis. In order to decrease the heterogeneity of samples and prioritize early-stage tumors that better enable the detection of early-onset driver events, we filtered out patients with stage greater than 1 for cancer types with sufficient remaining sample sizes ($N \geq 10$). This was feasible for six cancer types.

For the remaining cancers, analysis was performed on all available stages, and therefore, adjusted for the stage covariate. Filtration and/or adjustment of the stage contributed to a decrease in sample heterogeneity and improvement in quality of results (see lambda, Supp Table 3). The probes were filtered from 450K cross-reactive probes (n = 29233), probes overlapping genetic variants at targeted CpG sites or at the probe body (n = 8241), and XY chromosome probes (n = 792) (48). β values were log transformed to M values. Linear regression analysis was performed with adjustment for age in addition to sex and stage contingent on sample characteristics in every cancer type (Supp Table 3). Three cancer types (kidney renal clear cell carcinoma, lung squamous cell carcinoma and uterine corpus endometrial carcinoma) still had a relatively high lambda value and were excluded from the downstream analysis.

To further reduce the likelihood of false-positive results, dimension reduction was applied, and DMRs were analyzed using DMRcate package and filtered using the following

filtration criteria: 1) each DMR should include at least one CpGs with $|\Delta\beta_{filtered}|$ (i.e. $|\beta_{filtered\ in\ tumor\ tissues} - \beta_{filtered\ in\ normal\ tissues}|$) at least 20% (high effect sizes are expected in target tumor samples and further reduce false-positivity), 2) at least two-thirds of the probes with $|\Delta\beta_{filtered}| \geq 20\%$ within each DMR should have the same direction of change in $\Delta\beta_{filtered}$, and 3) all statistically significant CpGs within DMRs should show unimodal distribution when tested by dip test ($P > 0.05$).

Enrichment analysis

CpGs within DMRs associated with each of the tobacco smoking or cancer types were filtered out from CpGs exclusively present in either 450K or 850K arrays. After filtration, the number of overlapping CpGs whose methylation was altered in the same direction within each of cigarette or waterpipe smoking and each of the cancer types was determined and used to construct 2×2 contingency tables. Enrichment of cigarette- or waterpipe-associated CpGs within each cancer type was tested using Fisher's exact test, and FDR < 0.05 was considered statistically significant.

Expression Quantitative Trait Methylation Analysis (eQTM) of overlapping CpGs in enriched cancer types

RNA sequencing counts were downloaded for all samples with cancer types whose differential methylation showed significant overlap with CpGs within DMRs associated with cigarette or waterpipe smoking. Using R studio, differential gene expression was analyzed and normalized using DESeq2 Bioconductor package. After that, normalized expression data were correlated with filtered β using Pearson chi-square test, and FDR < 0.05 was considered statistically significant.

Multi-Omics Driver score of overlapping CpGs in enriched cancer types

In order to assess the cancer driver potential of the CpGs shared between each of the tobacco smoking and cancer types, Multi-Omics Driver score was determined for these CpGs in each of the enriched cancer types. The score was derived only from samples that were included in the DMR analysis. Multi-Omics Driver score within each cancer type was visualized as per the following. For single nucleotide alterations (SNA), the number of tumor samples with SNAs for each gene (encompassing the CpG site) was calculated.

After that, the SNA score was calculated as the number of samples having a given SNA over the sample size of the most common SNA in any of the genes analyzed of the respective cancer type. As for transcriptomic variations, the total number of samples with |z score| > 1.96 was calculated for a given gene. Then, the expression score of that gene was determined as the ratio of this number over the maximum among all genes in the corresponding cancer type.

For copy number alterations (CNAs), the sum of all deep insertions and deletions (i.e., with |CNA| > 1) in each gene in all samples was calculated. The CNA score was determined as the ratio of this CNA number to the maximum CNA of the corresponding cancer type. For methylation, the sum of samples with |Δβ| greater than or equal to 20% was calculated.

The methylation score was determined as the ratio of this value to the maximum value across all CpGs in the corresponding cancer type. A set of positive control driver CpGs in every cancer type was selected encompassing CpGs with the highest $\Delta\beta_{\text{tumor tissues - normal tissues}}$ in the gene having the highest ConsensusDriver score for a given cancer type according to Bertrand et al. (49). The combined Multi-Omics Driver score in all enriched cancers was similarly calculated, with the inclusion of enriched cancer types altogether.

Survival analysis in enriched cancer types

Survival analysis was determined for CpGs overlapping between each of the tobacco smoking types and each of the enriched cancer types, by aggregating for the corresponding cancer patients the DNA methylation and clinical data (vital status, time to death and time to

last follow-up). Samples were filtered with the same filtration criteria used for DNA methylation analysis, except for the stage variable, whereby all stages were included in this analysis to widen the survival outcome gradient. Then, the association of these CpGs with 5-year survival was analyzed using Cox regression with adjustment for age, sex (when applicable) and stage, and hazard ratios (95% CI) were reported. FDR < 0.05 was considered statistically significant. DNA methylation of significant CpGs were categorized using the median as a cut-off and visualized with Kaplan-Meier curves using R package survminer.

Multi-omics integration using multi-omics factor analysis (MOFA) followed by survival analysis

For enriched cancer types, multi-omics factor analysis (MOFA) was performed for DNA methylation and expression data of genes whose CpGs were associated with cigarette or waterpipe smoking, followed by survival analysis to investigate the effect of latent factors (derived from DNA methylation and gene expression data of smoking markers) on survival in these particular cancer types. In brief, for DNA methylation, the following were excluded from the analysis: 1) samples with 27K data, 2) formalin-fixed paraffin embedded samples and 3) samples with prior treatment.

For breast cancer cases, males were excluded since they had a very small sample size relative to females. β values were log transformed to M values and rescaled using scales package on R studio. For gene expression, RNA sequencing counts were normalized using DESeq2 Bioconductor package. After that, we focused on the CpGs that were analyzed by the 450K array and associated with cigarette or waterpipe smoking in our sample; these were 717 CpGs (115 genes) associated with cigarette smoking and 383 CpGs (63 genes) associated with waterpipe smoking.

MOFA was performed to integrate the DNA methylation and the RNA expression data of these markers into 10 latent factors using the MOFA2 R bioconductor package. Cox regression was performed to determine the association of the 10 factors with the overall survival in cancer patients, after adjustment for age, sex (when applicable) and stage (when

applicable). Results were visualized as hazard ratios ± confidence intervals, and FDR < 0.05 was considered statistically significant.

Data mining of Epigenome-Wide Association Studies (EWAS) on cigarette and waterpipe smoking

Several studies, including our own, have investigated the blood-based DNA methylation markers associated with cigarette smoking (15-17). To investigate whether there is any such study performed using 850K arrays, we carried out literature and data mining on PubMed and the EWAS data hub (50, 51).

In PubMed, the following key and MeSH terms were used in the literature search in Jan 2022:
((((("Illumina Infinium MethylationEPIC"[All Fields]) OR (“MethylationEPIC"[All Fields]) OR (“HumanMethylationEPIC"[All Fields]) OR ("EPIC array*"[All Fields]) OR ("850k"[All Fields]) OR ("epic methylation"[All Fields]) OR ("epic DNA methylation"[All Fields]) OR ("epic microarray*"[All Fields]) OR ("epic (illumina)*"[All Fields])) AND ((“tobacco”[All Fields]) OR ("cigarette" [All Fields]) OR ("cigarette smoking"[MeSH Terms]))) AND (("blood*"[All Fields]) OR ("blood"[MeSH Terms]))).

Ten studies were retrieved: four on cord blood (52-55), one on saliva and peripheral blood leukocytes (56), one on bronchoalveolar cells (57), one on nucleus accumbens brain cells (58), two on peripheral blood (59, 60), and one study was irrelevant. The two studies involving 850K profiling on peripheral blood are not yet reported in the EWAS data hub, and, among them, the study by Christiansen et al. (59) was used as our 850K reference because it conducted a meta-analysis of multiple cohorts, covering as well the second study by Domingo-Relloso et al. (60). Accordingly, our 850K-specific probes associated with cigarette smoking were compared with those reported by Christiansen et al. (59).

As for the 450K-inclusive probes associated with cigarette smoking in our study, they were compared with those in the EWAS data hub, and statistical overlap was determined using Joehanes et al. (61) as a 450K reference since it is a large consortium meta-analysis performed

on 16 cohorts. Enrichment analysis was performed using Fisher's exact test to determine the significance of the overlap between our findings and those of others.

Regarding waterpipe smoking, to investigate whether there are genome-wide methylation studies performed on blood samples of waterpipe smokers, the following key and MESH terms were used:
((((("genome-wide DNA methylation"[All Fields]) OR ("27k"[All Fields]) OR ("450k"[All Fields]) OR ("Illumina Infinium HumanMethylation450"[All Fields]) OR ("Illumina Infinium MethylationEPIC"[All Fields]) OR ("MethylationEPIC"[All Fields]) OR ("HumanMethylationEPIC"[All Fields]) OR ("EPIC array*"[All Fields]) OR ("850k"[All Fields]) OR ("epic methylation"[All Fields]) OR ("epic DNA methylation"[All Fields]) OR ("epic microarray*"[All Fields]) OR ("epic (illumina)*"[All Fields])) AND (("narjile*"[All Fields]) OR ("arguile*"[All Fields]) OR ("hookah"[All Fields]) OR ("hubble bubble"[All Fields]) OR ("shisha"[All Fields]) OR ("chicha"[All Fields]) OR ("waterpipe" [All Fields]))) AND (("blood*"[All Fields]) OR ("blood"[MeSH Terms]))).

No published study was available on this topic in PubMed or the EWAS data hub.

RESULTS

Discovery phase

Dimension reduction-based DNA methylome-wide analysis was performed on the Discovery set for each of the current cigarette-only and current waterpipe-only smokers relative to never smokers (see Materials and Methods).

For cigarette smoking, in the prioritized model adjusting for age, sex and B lymphocyte proportion, 2,776 DMRs were obtained (FDR < 0.05), of which 160 DMRs passed our filtration criteria (see Materials and Methods; Supp Figure 4A and Supp Table 4A). A similar number of hypomethylated (N = 80) and hypermethylated (N = 80) DMRs was observed (Figure 2A).

These DMRs encompassed 972 differentially methylated CpG sites, among which 281 had an effect size of at least 5 % (Supp Figure 4A-B) and ranging from -22.6 % to 16.2 % (Figure

2B-C). The CpGs with the highest negative and positive effect sizes were cg05575921 (-22.58%) in the *AHRR* gene and cg09232555 (16.19%) in a non-annotated region, respectively (Figure 2B-C).

For waterpipe smoking, in the prioritized model adjusting for age, sex and B lymphocyte proportion, 2,139 DMRs were obtained (FDR < 0.05), of which 87 DMRs passed our filtration criteria (see Materials and Methods; Supp Figure 4A and Supp Table 4B). There were more hypermethylated (N= 53) than hypomethylated DMRs (N= 34) (Figure 2A). These DMRs encompassed 491 differentially methylated CpG sites, among which 142 had an effect size of at least 5 % (Supp Figure 4A-B) and ranging from -9.5 % to 11.1 % (Figure 2B-C). The CpGs with the highest negative and positive effects sizes were cg26113488 (-9.50%) in a non-annotated region and cg06201514 (11.10%) in the *MYT1L* gene, respectively (Figure 2B-C).

Heatmap clustering of the 281 cigarette-associated CpGs (effect size ≥ 5 %) or the 142 waterpipe-associated CpGs (effect size ≥ 5 %) accurately distinguished cigarette smokers and waterpipe smokers from never smokers, respectively (Figure 3A-B). Moreover, machine learning-based prediction of cigarette smoking status not only accurately predicted subjects who are current cigarette smokers but also correctly categorized waterpipe smokers as being never cigarette smokers (Figure 3C-D).

Similarly, machine learning-based modelling trained on our data to predict waterpipe smoking (train set, Supp Table 5) was able to accurately predict subjects who are current waterpipe smokers in both training and test sets, as well as correctly categorize cigarette smokers as being never waterpipe smokers (Figure 3C-D). The overall accuracy of the cigarette and waterpipe prediction models were 92.6% and 95%, respectively (Figure 3C-D).

Technical validation and biological replication (validation phase)

We performed two separate validations of our results, one for the cigarette- and one for the waterpipe-associated CpGs (Figure 1). For validating the cigarette-associated CpGs, we compared our findings to previous reports, given that there have been several DNA

methylome-wide studies on blood of cigarette smokers (see further details on data mining in Materials and Methods).

Among our cigarette-associated CpGs with |Δβ| more than 5% (the minimum effect size used in our study), 79 of 281 CpGs were not previously analyzed by 450K arrays and were, hence, exclusive to the 850K array; the remaining 202 CpGs were 450K-inclusive. Of the 79 (850K-specific) CpGs, 19 significantly overlapped with ($P = 1.48\ e^{-27}$) and had the same direction of effect as 850K-specific cigarette-associated CpGs reported in the literature (using Christiansen et al. (59) as a reference set – see Materials and Methods) (Supp Table 6A).

Of the 202 (450K-inclusive) CpGs, 55 overlapped with 450K-inclusive cigarette-associated CpGs reported in the literature (using the EWAS Data Hub (50, 51) as a reference set – see Materials and Methods). Among those 55 CpGs, only 3 had opposite directions of effect compared to our findings and were all uniquely associated with maternal cigarette smoking during pregnancy and measured in the neonatal blood of the offspring (Supp Table 6B).

Moreover, 51 CpGs significantly overlapped with ($P = 1.99\ e^{-38}$) and had the same direction of change as 450K analyzed cigarette associated CpGs in the literature (using Joehanes et al. (61) as a reference set – see Materials and Methods). The well-established *AHRR* cg05575921 reported by several studies (61-64) was also the top hit associated with cigarette smoking in our study and had a similar direction and magnitude of effect as reported in the other studies (~ -20%).

Regarding the waterpipe-associated CpGs, we performed a technical validation of the array-based results using pyrosequencing as well as a biological replication in additional samples. The CpGs cg26113488 (NA) and cg15327692 (*RASGRP2*) were selected for this purpose, representing the most significantly (FDR < 0.05) hypomethylated and hypermethylated CpGs with the largest effect sizes, respectively (Supp Table 1, Figure 4A), with the exception of cg11752927 (*PTPRN2*) omitted for technical reasons (see further details on validation in Materials and Methods). In the NA gene, another proximal and significant CpG (cg14667685) was covered by pyrosequencing (Supp Table 1, Figure 4A).

The 850K- and pyrosequencing-based methylation values of the three CpGs were significantly correlated (Pearson correlation coefficient (r) = 0.91, P = 4.75 e^{-07}) (Figure 4B). Similarly to the results in the Discovery set (Figure 4C), pyrosequencing of cg26113488 (NA) and cg14667685 (NA) showed a significant hypomethylation in new samples of current waterpipe-only compared to never smokers (P = 0.002 before adjustment and 0.007 after adjustment for age and sex), and this difference was significantly maintained in former waterpipe smokers (Figure 4D).

Pyrosequencing of cg15327692 also showed results consistent with those in the Discovery set, that is, a significant hypermethylation in new samples of current waterpipe-only *versus* never smokers (P = 9.28 e^{-06} before and 9.46 e^{-04} after adjustment for age and sex) (Figure 4C-D). Time-dose analysis showed that methylation of cg26113488 (Figure 4E-LEFT) but not the other two CpGs (Supp Figure 6B) was negatively correlated with waterpipe session-year in current waterpipe-only smokers (r = -0.36, P = 0.020).

Moreover, waterpipe smokers belonging to the highest tertile of extent of waterpipe exposure had significantly lower (P = 0.043) methylation of cg26113488 when compared to the lowest tertile, and the methylation levels in the latter were similar to those in never smokers (Figure 4E-RIGHT).

Functional genomic analysis of smoking associated CpGs in relation to CpG density clusters, regulatory regions and biological readouts

Analysis of the genomic distribution of CpGs within either cigarette- or waterpipe-associated DMRs revealed that they were enriched in CpG islands (FDR = 5.74 e^{-37} for cigarette and 3.04 e^{-21} for waterpipe), shores (FDR = 6.19 e^{-24} for cigarette and 3.76 e^{-18} for waterpipe) and promoters (FDR = 2.17 e^{-04} for cigarette and 2.85 e^{-10} for waterpipe) in comparison to remaining CpGs on the 850K array (Figure 5A).

In addition, CpGs within cigarette-associated DMRs were enriched in intron-exon boundaries (FDR = 0.047), first exons (FDR = 0.030), coding DNA sequence (FDR = 0.005) and 5'UTR (FDR = 0.017) (Figure 5A). There were no significant changes between cigarette or

waterpipe smokers compared to never smokers in the mean DNA methylation using all 850K probes, CpGs within cigarette-associated DMRs and CpGs within waterpipe-associated DMRs, when plotted as overall or grouped by CpG density or regulatory regions (Supp Figure 7A-C).

Consistently, there were no significant changes in global DNA methylation (represented by average overall 850K probes and predicted *LINE-1* or *ALU* methylation) in association with either cigarette or waterpipe smoking categories or extent of smoking exposure (Supp Figure 8A-C). Similarly, no changes in epigenetically predicted age acceleration were significantly associated with smoking (Supp Figure 8A-C). As for epigenetically predicted mortality risk score, it was positively associated with extent of cigarette smoking exposure (FDR-adjusted P = 0.003) (Supp Figure 8B), and this association remained significant after adjustment for age, sex, and BMI (P = 0.010). No significant changes in mortality risk score were associated with waterpipe smoking (Supp Figure 8A, C).

Gene Ontology, KEGG and Reactome analyses highlighted a significant (FDR < 0.05) implication of the genes differentially methylated by cigarette and waterpipe smoking in multiple biological pathways, including pathways commonly altered by both tobacco forms (Supp Table 7A-B). The top ten significant pathways for each exposure are shown in Figure 5B, including the common pathways, being regulation of transcription, cell differentiation, signal transduction, hemostasis and cell adhesion. Moreover, cancer hallmark enrichment analysis showed that differentially methylated genes associated with cigarette or waterpipe smoking were enriched in the cancer hallmark, activating invasion motility (Figure 5C). Differentially methylated genes associated with cigarette smoking were additionally enriched in cancer hallmarks avoiding immune destruction and inducing angiogenesis (Figure 5C).

Enrichment analysis of smoking-induced DNA methylome alterations in different cancer types DMR analysis comparing tumor *versus* normal tissues was performed in nine cancer types from TCGA database after exclusion of three cancers with large inflation (lambda) values (see further details in Materials and Methods). Supp Table 8 shows the filtration criteria of top DMRs obtained, with focus on CpGs common between 450K and 850K arrays; these DMRs are reported in Supp Tables 9A-K. Both cigarette- and waterpipe- associated CpGs were significantly (FDR < 0.05) enriched in the DNA methylation alterations associated with lung

adenocarcinoma (Figure 6A). Cigarette- and waterpipe-associated CpGs were additionally enriched (FDR < 0.05) in the DNA methylation alterations associated with pancreatic adenocarcinoma and breast carcinoma, respectively (Figure 6A). The number of overlapping CpGs between cigarette- or waterpipe- associated CpGs and their enriched cancers is shown in Supp Figure 9A.

eQTM analysis of overlapping DNA methylation alterations between smoking and câncer

We next performed eQTM analysis on DNA methylation levels of CpGs overlapping between smoking and enriched cancers in association with RNA expression of the genes harboring the corresponding CpGs. For CpGs overlapping between cigarette smoking and enriched cancers, negative correlation between DNA methylation and gene expression was observed in the majority (89%) of the 38 FDR-significant CpGs in lung cancer, while all 7 FDR-significant CpGs showed positive correlation between DNA methylation and gene expression in pancreatic cancer (Supp Figure 9B).

As for CpGs overlapping between waterpipe smoking and enriched cancers, negative correlation between DNA methylation and gene expression was observed in 14% of the 29 FDR-significant CpGs in breast cancer and 33% of the 21 FDR-significant CpGs in lung cancer (Supp Figure 9B). In particular, the eQTMs with moderate-to-high correlation (r > 0.5) are shown in Figure 6B.

Multi-Omic Driver roles of smoking-induced DNA methylome alterations and effect on cancer prognosis

The methylation-expression omics analysis performed earlier on CpGs overlapping between smoking and enriched cancer types was integrated with CNA and SNA genomics data in order to derive Multi-Omics Driver scores for those CpGs, based on our previously reported algorithm (65). The Multi-Omics Driver scores of those CpGs across their enriched cancer types were comparable to the Multi-Omics Driver scores of positive control CpGs derived from

known cancer driver genes, that is, genes having ConsensusDriver scores of at least 1.5 in each analyzed cancer type (49) (Figure 6C). Similar results were obtained by analyzing Multi-Omics Driver scores in each cancer type separately (Supp Figure 10 and Supp Table 10). Despite the small number of overlapping CpGs initially identified between cigarette and waterpipe smoking (Figure 4B), several of these CpGs were shared with the enriched cancer types and showed relatively high Multi-Omics Driver scores (Figure 6C and Supp Figure 10).

Survival analysis of cigarette-associated cancer-enriched CpGs revealed no detectable effect on patient survival in lung and pancreatic cancers (FDR > 0.05) (Supp Table 11A-B). MOFA-based integration of the DNA methylation and transcriptomics data of cigarette-associated CpGs and their annotated genes showed that one latent factor (# 2) was significantly associated (FDR < 0.05) with survival in lung cancer patients (Figure 6D-LEFT), and none in pancreatic cancer patients (Supp Figure 11A-B). This latent factor was weighted more for transcriptional than DNA methylation changes (Figure 6D-LEFT), in line with the observed undetectable effect between cigarette-associated CpGs and lung cancer patient survival.

Survival analysis of waterpipe-associated cancer-enriched CpGs revealed that the methylation levels of three CpGs in *WNK4* had a significantly increased hazard ratio (FDR < 0.05) in association with survival in lung but not breast cancer patients (Supp Table 11C-D). Among them, cg05886546 had more than 5% mean methylation difference in waterpipe *versus* never smokers (Supp Table 11D), and its hypermethylation was significantly associated (P = 0.012) with worse patient survival (Figure 6E).

MOFA-based integration of the DNA methylation and transcriptomic data of waterpipe-associated CpGs and their annotated genes showed that one latent factor (# 6) was significantly associated (FDR < 0.05) with survival in lung cancer patients (Figure 6D-RIGHT), and none in breast cancer patients (Supp Figure 11A-B). This latent factor was weighted more for DNA methylation than gene expression changes, indicating a potentially larger contribution of DNA methylation alterations in affecting cancer patient survival for waterpipe-associated markers. This result was in line with the observed significant effects of several waterpipe-associated CpGs on lung cancer patient survival.

DISCUSSION

In this study, we assessed the DNA methylation alterations in peripheral blood associated with waterpipe smoking and compared them to those of cigarette, based on a highly exposed population from Lebanon, representing one of the few adult cohorts in the Middle East and Africa.

Multiple studies investigated the blood-based DNA methylome-wide alterations associated with cigarette smoking (15, 17, 18, 59-61, 63, 64, 66, 67), all of which were 450K-based, except two which included 850K (59, 60) and of which one (59) conducted a meta-analysis that covered the other (60). As described in Results, our findings significantly replicated 450K-inclusive and 850K-specific CpGs associated with cigarette smoking.

Moreover, when compared to the largest meta-analysis evaluating 450K data of 16 cohorts using CpG-by-CpG (instead of DMR) analysis, our study replicated seven of the ten CpGs having the highest effect size (61). These CpGs are annotated to *AHRR*, *GFI1*, *GNG12*, *F2RL3* and *MYO1G* genes. When compared to another large study (n = 910) using a DMR approach similar to ours, albeit in a population constituting only of women, our results also replicated four of the five DMRs that had the strongest association with cigarette smoking (15). These are annotated to *GFI1*, *MYO1G*, *AHRR* and near *ALPPL2*.

In addition, the top hit with the highest effect size, *AHRR* cg05575921, reported by several studies (61-64) was also the top hit associated with cigarette smoking in our study, with the same direction and magnitude of effect (~ -20%). The reproducibility of our findings in relation to cigarette smoking reinforces the estimated power of our study, upon which we also identified 79 cigarette-associated CpGs exclusive to the 850K array (not shared with 450K) and encompassing 60 novel CpGs annotated to genes listed in Supp Table 6A. Although these require further validation in more studies, they could provide additional markers for cigarette smoking and better understanding of the molecular mechanisms underlying the effect of cigarette smoking on health outcomes.

With regards to waterpipe smoking, it had been investigated in two studies in relation to DNA methylation using, however, only gene-centered or low-coverage approaches. In one

study, non-cancerous small airway epithelium tissues of seven waterpipe and seven never smokers were analyzed using HpaII tiny fragment Enriched by Ligation-mediated PCR (HELP) assay (68). In the second, oral epithelium cells and blood lymphocytes were derived from 150 waterpipe and 150 never smokers and analyzed for *MLH1* promoter methylation using methylation-specific PCR technique (69).

Since genome-wide methylation studies, particularly in surrogate tissues (e.g. blood), of waterpipe smoking had been lacking so far, we successfully validated the top two 850K-based markers of waterpipe smoking, cg26113488 (NA) and cg15327692 (*RASGRP2*), by pyrosequencing in additional samples. Moreover, cg26113488 hypomethylation was correlated with increased dose and duration of waterpipe smoking and was maintained in former waterpipe smokers, suggesting that methylation alterations at this site are not reversed with waterpipe smoking cessation.

Longitudinal study designs with collection of biospecimen at multiple time points are required to enhance the resolution of the dose- and time-response relationships in smokers. These promising biomarkers also necessitate further validation in independent populations, and their pyrosequencing-based measurement represents a simple, cost-effective and rapid screening tool in epidemiological and clinical settings.

We observed a significant enrichment between the DNA methylation markers of waterpipe and cigarette smoking (Supp Figure 4B), but the markers of each smoking form were predominantly distinct from one another, at both the CpG and gene levels, and they could accurately predict for and differentiate between the two types of smoking, which, nevertheless, are both tobacco-based. Even the well-established *AHRR* marker of cigarette smoking had no detectable CpG(s) significantly associated with waterpipe smoking.

The unique epigenetic alterations associated with waterpipe and cigarette smoking in this study could be partly explained by differences in the amounts and/or nature of chemical constituents in the two types of tobacco smoke (70-72). Puff topography data show that one waterpipe session releases greater amounts of carbon monoxide, nicotine, volatile aldehydes and polycyclic aromatic hydrocarbons than a single cigarette (70-72). Nevertheless, differences in modes and patterns of smoking within and across waterpipe and cigarette

smokers need to be considered, so direct comparison between cigarette and waterpipe smokers in an epidemiological setting needs to be interpreted with caution.

A cross-over study on 31 cigarette and waterpipe smokers partly addressed this issue and confirmed that a single waterpipe session produced three times greater peak of plasma carboxyhemoglobin levels relative to a single cigarette (73). *In vitro* and *in vivo* experimental assays can complement these settings and enhance the resolution of the effect of smoking exposure, including experimental manipulation of its specific constituents, flavors, doses and durations (74), all of which may vary between and within waterpipe and cigarette smoking forms. These assays can also help further characterize the functional roles of epigenetic markers, particularly that the latter not only act as sensors to exposures (like other well-established tobacco markers such as cotinine) but also constitute heritable mechanisms driving disease development and are potentially stable with the ability to capture exposure dose and duration over several years (15, 61, 75).

Still, in terms of functional genomics, DNA methylation alterations in both cigarette-only and waterpipe-only smokers were enriched in CpG islands and shores, in promoter regions, and in the cancer hallmark activating invasion motility. Moreover, among the top ten most significant biological pathways that were epigenetically deregulated in either smoking form, half were common.

Based on in-depth screening of the DNA methylome of 11,258 tumor and 1,106 normal tissues spanning multiple tumor types from TCGA, we found that epigenetic markers of cigarette and waterpipe smoking were also both enriched in genes epigenetically deregulated in lung cancer. Our results are considered complementary to other findings showing that the lung adenocarcinoma tissues harbor cigarette-associated epigenetic signature that overlaps with cigarette-associated epigenetic modifications observed in blood of adults or newborns (20).

The waterpipe and cigarette markers were additionally epigenetically deregulated in breast and pancreatic cancers, respectively. These markers may, hence, explain, at least in part, the epigenetic mechanisms underlying the consistently reported associations of cigarette and waterpipe smoking with several cancer types (76-79). They may also facilitate

the focus on specific genomic regions. For instance, some of the overlapping markers between smoking and lung cancer were shared between the two tobacco types, functionally impacted gene expression and had relatively high Multi-Omics Driver scores; these were annotated to *ZNF423* and *HOPX* genes. Consistently, another study showed that *ZNF423* was hypomethylated in cancerous compared to contralateral tumor-free bronchoscopic biopsies (80); hence, its role in lung cancer pathogenesis warrants further investigation.

MOFA-based integration of the DNA methylation and expression data of cigarette- or waterpipe-associated CpGs and genes showed that one of the top ten latent factors in each smoking form was significantly associated with lung cancer survival, hence, providing a mechanistic basis for the association between smoking and worse lung cancer survival or increased cancer mortality reported in several studies (81-86).

The cigarette-specific latent factor was weighted more for transcriptional relative to DNA methylation changes, in line with the observed undetectable effect between cigarette-associated CpGs and lung cancer patient survival. On the contrary, the waterpipe-specific latent factor was weighted more for DNA methylation relative to gene expression changes, suggesting that DNA methylation of waterpipe-associated markers might be more implicated in lung cancer prognosis than gene expression. This result was in line with the observed significant effects of waterpipe-associated *WNK4* CpGs on lung cancer patient survival.

This first high-coverage EWAS study of waterpipe smokers, is based on a well-powered design and an accessible tissue matrix (blood) for biomarker utility, compares waterpipe and cigarette smokers in the same population derived from a geographic region historically rich in these exposures, identifies novel epigenetic markers of waterpipe- and cigarette-only smoking, validates findings by independent techniques and replicates them in additional samples, and investigates the functional and clinical relevance of the identified markers in relation to cancer using functional genomics analysis across multiple omic layers and cancer types.

Moreover, the DMRs derived from our pan-cancer analysis comparing tumor *versus* normal tissues provide *per se* a research repertoire of genes epigenetically deregulated in each

of several cancer types. In addition, artificial intelligence-based DNA methylome mapping was able to accurately predict and differentiate cigarette from waterpipe smokers and *vice versa*.

While these models warrant further validation in independent cohorts, they can be potentially useful in the future for accurate prediction of smoking type, duration and dose. The implications can be multifold, such as validating questionnaires in epidemiological studies, predicting molecular diaries of exposure in archived biospecimen or their derived omics data, and identifying patients at risk especially when these exposure markers also happen to affect disease outcomes, such as cancer. Our findings are also expected to provide fundamental insights to various research and implementation settings interested in the utility of epigenetic markers as powerful molecular sensors of smoking exposure and its various evolving forms (87), including the increasingly prevalent waterpipe smoking.

Acknowledgement

The authors thank the participating subjects of the Greater Beirut Cohort and Dr Alexei Novoloaca for sharing some statistical expertise.

Funding

American University of Beirut Medical Practice Plan (MPP), Lebanon; l'Institut National du Cancer (INCa) and l'Institut pour la Recherche en Santé Publique (IReSP), INCa/IReSP 16132, France. Cette recherche a été financée par l'IReSP et l'INCa dans le cadre de l'appel à projets 2021 de recherche pour lutter contre les usages et les addictions aux substances psychoactives. The work reported in this paper was undertaken during the tenure of an IARC Postdoctoral Fellowship (awarded to Awada Z) at the International Agency for Research on Cancer.

Ethical Approval

This study builds on subjects who were previously recruited onto the Greater Beirut Cohort after signing an AUB IRB approved informed consent (IM.HT.03). The DNA methylation aspect of the study is also approved by the same IRB (BIO-2019-0243).

Conflict of interest

The authors declare no competing interests.

REFERENCES

1. World Health Organization (WHO). Advisory note: waterpipe tobacco smoking: health effects, research needs and recommended actions by regulators, 2nd edition [press release]. 15 April 2015 15 April 2015.
2. Akl EA, Gunukula SK, Aleem S, Obeid R, Jaoude PA, Honeine R, et al. The prevalence of waterpipe tobacco smoking among the general and specific populations: a systematic review. BMC Public Health. 2011;11:244.
3. Tucktuck M, Ghandour R, Abu-Rmeileh NME. Waterpipe and cigarette tobacco smoking among Palestinian university students: a cross-sectional study. BMC Public Health. 2017;18(1):1.
4. Jawad M, Charide R, Waziry R, Darzi A, Ballout RA, Akl EA. The prevalence and trends of waterpipe tobacco smoking: A systematic review. PLoS One. 2018;13(2):e0192191.
5. Chan WC, Leatherdale ST, Burkhalter R, Ahmed R. Bidi and hookah use among Canadian youth: an examination of data from the 2006 Canadian Youth Smoking Survey. J Adolesc Health. 2011;49(1):102-4.
6. Filippidis FT, Jawad M, Vardavas CI. Trends and Correlates of Waterpipe use in the European Union: Analysis of Selected Eurobarometer Surveys (2009–2017). Nicotine & Tobacco Research. 2017;21(4):469-74.
7. Hammal F, Wild TC, Nykiforuk C, Abdullahi K, Mussie D, Finegan BA. Waterpipe (Hookah) Smoking Among Youth and Women in Canada is New, not Traditional. Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco. 2016;18(5):757-62.
8. Yadav S, Rawal G. Waterpipe Tobacco Smoking: A Mini-review. J Transl Int Med. 2018;6(4):173-5.
9. Primack BA, Carroll MV, Weiss PM, Shihadeh AL, Shensa A, Farley ST, et al. Systematic Review and Meta-Analysis of Inhaled Toxicants from Waterpipe and Cigarette Smoking. Public Health Rep. 2016;131(1):76-85.

10. Shihadeh A, Schubert J, Klaiany J, El Sabban M, Luch A, Saliba NA. Toxicant content, physical properties and biological activity of waterpipe tobacco smoke and its tobacco-free alternatives. Tobacco Control. 2015;24(Suppl 1):i22-i30.
11. Mamtani R, Cheema S, Sheikh J, Al Mulla A, Lowenfels A, Maisonneuve P. Cancer risk in waterpipe smokers: a meta-analysis. Int J Public Health. 2017;62(1):73-83.
12. El-Zaatari ZM, Chami HA, Zaatari GS. Health effects associated with waterpipe smoking. Tob Control. 2015;24 Suppl 1:i31-i43.
13. Ghantous A, Saffery R, Cros MP, Ponsonby AL, Hirschfeld S, Kasten C, et al. Optimized DNA extraction from neonatal dried blood spots: application in methylome profiling. BMC Biotechnol. 2014;14:60.
14. Shenker NS, Ueland PM, Polidoro S, van Veldhoven K, Ricceri F, Brown R, et al. DNA methylation as a long-term biomarker of exposure to tobacco smoke. Epidemiology (Cambridge, Mass). 2013;24(5):712-6.
15. Ambatipudi S, Cuenin C, Hernandez-Vargas H, Ghantous A, Le Calvez-Kelm F, Kaaks R, et al. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. Epigenomics. 2016;8(5):599-618.
16. Sikdar S, Joehanes R, Joubert BR, Xu CJ, Vives-Usano M, Rezwan FI, et al. Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking. Epigenomics. 2019;11(13):1487-500.
17. Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, et al. DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. Am J Hum Genet. 2016;98(4):680-96.
18. McCartney DL, Stevenson AJ, Hillary RF, Walker RM, Bermingham ML, Morris SW, et al. Epigenetic signatures of starting and stopping smoking. EBioMedicine. 2018.
19. Teschendorff AE, Yang Z, Wong A, Pipinikas CP, Jiao Y, Jones A, et al. Correlation of Smoking-Associated DNA Methylation Changes in Buccal Cells With DNA Methylation Changes in Epithelial Cancer. JAMA Oncol. 2015;1(4):476-85.
20. Bakulski KM, Dou J, Lin N, London SJ, Colacino JA. DNA methylation signature of smoking in lung cancer is enriched for exposure signatures in newborn and adult blood. Scientific reports. 2019;9(1):4576.
21. Zhang Y, Elgizouli M, Schottker B, Holleczek B, Nieters A, Brenner H. Smoking-associated DNA methylation markers predict lung cancer incidence. Clin Epigenetics. 2016;8:127.
22. Zhang Y, Breitling LP, Balavarca Y, Holleczek B, Schottker B, Brenner H. Comparison and combination of blood DNA methylation at smoking-associated genes and at lung cancer-related genes in prediction of lung cancer mortality. Int J Cancer. 2016;139(11):2482-92.
23. Fasanelli F, Baglietto L, Ponzi E, Guida F, Campanella G, Johansson M, et al. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. Nat Commun. 2015;6:10192.
24. Baglietto L, Ponzi E, Haycock P, Hodge A, Bianca Assumma M, Jung CH, et al. DNA methylation changes measured in pre-diagnostic peripheral blood samples are associated with smoking and lung cancer risk. Int J Cancer. 2017;140(1):50-61.
25. Lee KW, Pausova Z. Cigarette smoking and DNA methylation. Front Genet. 2013;4:132.

26. Mouneimne Y, Nasrallah M, Khoueiry-Zgheib N, Nasreddine L, Nakhoul N, Ismail H, et al. Bisphenol A urinary level, its correlates, and association with cardiometabolic risks in Lebanese urban adults. Environmental monitoring and assessment. 2017;189(10):517.
27. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics (Oxford, England). 2014;30(10):1363-9.
28. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome biology. 2016;17(1):208.
29. Fortin JP, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol. 2014;15(12):503.
30. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics (Oxford, England). 2012;28(6):882-3.
31. Perrier F, Novoloaca A, Ambatipudi S, Baglietto L, Ghantous A, Perduca V, et al. Identifying and correcting epigenetics measurements for systematic sources of variation. Clin Epigenetics. 2018;10:38.
32. Lin X, Barton S, Holbrook JD. How to make DNA methylome wide association studies more powerful. Epigenomics. 2016;8(8):1117-29.
33. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics. 2010;11(1):587.
34. Salas LA, Koestler DC, Butler RA, Hansen HM, Wiencke JK, Kelsey KT, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. Genome Biology. 2018;19(1):64.
35. Kaushal A, Zhang H, Karmaus WJJ, Ray M, Torres MA, Smith AK, et al. Comparison of different cell type correction methods for genome-scale epigenetics studies. BMC Bioinformatics. 2017;18(1):216.
36. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, R VL, et al. De novo identification of differentially methylated regions in the human genome. Epigenetics & chromatin. 2015;8:6.
37. Woo HD, Fernandez-Jimenez N, Ghantous A, Degli Esposti D, Cuenin C, Cahais V, et al. Genome-wide profiling of normal gastric mucosa identifies Helicobacter pylori- and cancer-associated DNA methylome changes. Int J Cancer. 2018;143(3):597-609.
38. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300.
39. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics. 2010;11:367.
40. Bollepalli S, Korhonen T, Kaprio J, Anders S, Ollikainen M. EpiSmokEr: a robust classifier to determine smoking status from DNA methylation data. Epigenomics. 2019;11(13):1469-86.

41. Busato F, Dejeux E, El Abdalaoui H, Gut IG, Tost J. Quantitative DNA Methylation Analysis at Single-Nucleotide Resolution by Pyrosequencing®. Methods Mol Biol. 2018;1708:427-45.
42. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics (Oxford, England). 2010;26(17):2190-1.
43. Zheng Y, Joyce BT, Liu L, Zhang Z, Kibbe WA, Zhang W, et al. Prediction of genome-wide DNA methylation in repetitive elements. Nucleic Acids Res. 2017;45(15):8697-711.
44. Horvath S, Oshima J, Martin GM, Lu AT, Quach A, Cohen H, et al. Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. Aging (Albany NY). 2018;10(7):1758-75.
45. Zhang Y, Wilson R, Heiss J, Breitling LP, Saum KU, Schottker B, et al. DNA methylation signatures in peripheral blood strongly predict all-cause mortality. Nat Commun. 2017;8:14617.
46. Carmona-Saez P, Chagoyen M, Tirado F, Carazo JM, Pascual-Montano A. GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. Genome biology. 2007;8(1):R3.
47. Monk D. Deciphering the cancer imprintome. Briefings in Functional Genomics. 2010;9(4):329-39.
48. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013;8(2):203-9.
49. Bertrand D, Drissler S, Chia BK, Koh JY, Li C, Suphavilai C, et al. ConsensusDriver Improves upon Individual Algorithms for Predicting Driver Alterations in Different Cancer Types and Individual Patients. Cancer Res. 2018;78(1):290-301.
50. Li M, Zou D, Li Z, Gao R, Sang J, Zhang Y, et al. EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. Nucleic Acids Res. 2019;47(D1):D983-D8.
51. Xiong Z, Li M, Yang F, Ma Y, Sang J, Li R, et al. EWAS Data Hub: a resource of DNA methylation array data and metadata. Nucleic Acids Res. 2020;48(D1):D890-D5.
52. Howe CG, Zhou M, Wang X, Pittman GS, Thompson IJ, Campbell MR, et al. Associations between Maternal Tobacco Smoke Exposure and the Cord Blood [Formula: see text] DNA Methylome. Environ Health Perspect. 2019;127(4):47009.
53. Bergens MA, Pittman GS, Thompson IJB, Campbell MR, Wang X, Hoyo C, et al. Smoking-associated AHRR demethylation in cord blood DNA: impact of CD235a+ nucleated red blood cells. Clin Epigenetics. 2019;11(1):87.
54. De Queiroz Andrade E, Gomes GMC, Collison A, Grehan J, Murphy VE, Gibson P, et al. Variation of DNA Methylation in Newborns Associated with Exhaled Carbon Monoxide during Pregnancy. Int J Environ Res Public Health. 2021;18(4).
55. Danielewicz H, Gurgul A, Debinska A, Myszczyszyn G, Szmatola T, Myszkal A, et al. Maternal atopy and offspring epigenome-wide methylation signature. Epigenetics. 2021;16(6):629-41.
56. Barcelona V, Huang Y, Brown K, Liu J, Zhao W, Yu M, et al. Novel DNA methylation sites associated with cigarette smoking among African Americans. Epigenetics. 2019;14(4):383-91.

57. Ringh MV, Hagemann-Jensen M, Needhamsen M, Kular L, Breeze CE, Sjoholm LK, et al. Tobacco smoking induces changes in true DNA methylation, hydroxymethylation and gene expression in bronchoalveolar lavage cells. EBioMedicine. 2019;46:290-304.
58. Markunas CA, Semick SA, Quach BC, Tao R, Deep-Soboslay A, Carnes MU, et al. Genome-wide DNA methylation differences in nucleus accumbens of smokers vs. nonsmokers. Neuropsychopharmacology. 2021;46(3):554-60.
59. Christiansen C, Castillo-Fernandez JE, Domingo-Relloso A, Zhao W, El-Sayed Moustafa JS, Tsai PC, et al. Novel DNA methylation signatures of tobacco smoking with trans-ethnic effects. Clin Epigenetics. 2021;13(1):36.
60. Domingo-Relloso A, Riffo-Campos AL, Haack K, Rentero-Garrido P, Ladd-Acosta C, Fallin DM, et al. Cadmium, Smoking, and Human Blood DNA Methylation Profiles in Adults from the Strong Heart Study. Environ Health Perspect. 2020;128(6):67005.
61. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic Signatures of Cigarette Smoking. Circulation Cardiovascular genetics. 2016;9(5):436-47.
62. Monick MM, Beach SR, Plume J, Sears R, Gerrard M, Brody GH, et al. Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics. 2012;159b(2):141-51.
63. Sayols-Baixeras S, Lluís-Ganella C, Subirana I, Salas LA, Vilahur N, Corella D, et al. Identification of a new locus and validation of previously reported loci showing differential methylation associated with smoking. The REGICOR study. Epigenetics. 2015;10(12):1156-65.
64. Harlid S, Xu Z, Panduri V, Sandler DP, Taylor JA. CpG sites associated with cigarette smoking: analysis of epigenome-wide data from the Sister Study. Environ Health Perspect. 2014;122(7):673-8.
65. Halaburkova A, Cahais V, Novoloaca A, da Silva Araujo MG, Khoueiry R, Ghantous A, et al. Pan-cancer multi-omics analysis and orthogonal experimental assessment of epigenetic driver genes. Genome research. 2020;30:1-16.
66. Su D, Wang X, Campbell MR, Porter DK, Pittman GS, Bennett BD, et al. Distinct Epigenetic Effects of Tobacco Smoking in Whole Blood and among Leukocyte Subtypes. PLoS One. 2016;11(12):e0166486.
67. Mishra PP, Hanninen I, Raitoharju E, Marttila S, Mishra BH, Mononen N, et al. Epigenome-450K-wide methylation signatures of active cigarette smoking: The Young Finns Study. Biosci Rep. 2020;40(7).
68. Walters MS, Salit J, Ju JH, Staudt MR, Kaner RJ, Rogalski AM, et al. Waterpipe smoking induces epigenetic changes in the small airway epithelium. PLoS One. 2017;12(3):e0171112.
69. Sabi SH, Khabour OF, Alzoubi KH, Cobb CO, Eissenberg T. Changes at global and site-specific DNA methylation of MLH1 gene promoter induced by waterpipe smoking in blood lymphocytes and oral epithelial cells. Inhal Toxicol. 2020;32(3):124-30.
70. Sepetdjian E, Shihadeh A, Saliba NA. Measurement of 16 polycyclic aromatic hydrocarbons in narghile waterpipe tobacco smoke. Food Chem Toxicol. 2008;46(5):1582-90.
71. Al Rashidi M, Shihadeh A, Saliba NA. Volatile aldehydes in the mainstream smoke of the narghile waterpipe. Food Chem Toxicol. 2008;46(11):3546-9.

72. Shihadeh A, Saleh R. Polycyclic aromatic hydrocarbons, carbon monoxide, "tar", and nicotine in the mainstream smoke aerosol of the narghile water pipe. Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association. 2005;43(5):655-61.
73. Eissenberg T, Shihadeh A. Waterpipe tobacco and cigarette smoking: direct comparison of toxicant exposure. Am J Prev Med. 2009;37(6):518-23.
74. Jackson A, Grobman B, Krishnan-Sarin S. Recent findings in the pharmacology of inhaled nicotine: Preclinical and clinical in vivo studies. Neuropharmacology. 2020;176:108218.
75. Herceg Z, Ghantous A, Wild CP, Sklias A, Casati L, Duthie SJ, et al. Roadmap for investigating epigenome deregulation and environmental origins of cancer. Int J Cancer. 2018;142(5):874-82.
76. Gandini S, Botteri E, Iodice S, Boniol M, Lowenfels AB, Maisonneuve P, et al. Tobacco smoking and cancer: a meta-analysis. Int J Cancer. 2008;122(1):155-64.
77. Montazeri Z, Nyiraneza C, El-Katerji H, Little J. Waterpipe smoking and cancer: systematic review and meta-analysis. Tob Control. 2017;26(1):92-7.
78. Waziry R, Jawad M, Ballout RA, Al Akel M, Akl EA. The effects of waterpipe tobacco smoking on health outcomes: an updated systematic review and meta-analysis. International Journal of Epidemiology. 2016;46(1):32-43.
79. Lee PN, Forey BA, Coombs KJ. Systematic review with meta-analysis of the epidemiological evidence in the 1900s relating smoking to lung cancer. BMC Cancer. 2012;12:385.
80. Goldmann T, Schmitt B, Müller J, Kröger M, Scheufele S, Marwitz S, et al. DNA methylation profiles of bronchoscopic biopsies for the diagnosis of lung cancer. Clinical Epigenetics. 2021;13(1):38.
81. Yi JH, Choi PJ, Jeong SS, Bang JH, Jeong JH, Cho JH. Prognostic Significance of Cigarette Smoking in Association with Histologic Subtypes of Resected Lung Adenocarcinoma. Korean J Thorac Cardiovasc Surg. 2019;52(5):342-52.
82. Lee SJ, Lee J, Park YS, Lee CH, Lee SM, Yim JJ, et al. Impact of smoking on mortality of patients with non-small cell lung cancer. Thorac Cancer. 2014;5(1):43-9.
83. Tammemagi CM, Neslund-Dudas C, Simoff M, Kvale P. Smoking and lung cancer survival: the role of comorbidity and treatment. Chest. 2004;125(1):27-37.
84. Bryant A, Cerfolio RJ. Differences in epidemiology, histology, and survival between cigarette smokers and never-smokers who develop non-small cell lung cancer. Chest. 2007;132(1):185-92.
85. Etemadi A, Khademi H, Kamangar F, Freedman ND, Abnet CC, Brennan P, et al. Hazards of cigarettes, smokeless tobacco and waterpipe in a Middle Eastern Population: a Cohort Study of 50 000 individuals from Iran. Tob Control. 2017;26(6):674-82.
86. Wu F, Chen Y, Parvez F, Segers S, Argos M, Islam T, et al. A prospective study of tobacco smoking and mortality in Bangladesh. PLoS One. 2013;8(3):e58516.
87. Kim SC, Friedman TC. A New Ingenious Enemy: Heat-Not-Burn Products. Tobacco use insights. 2022;15:1179173x221076419.

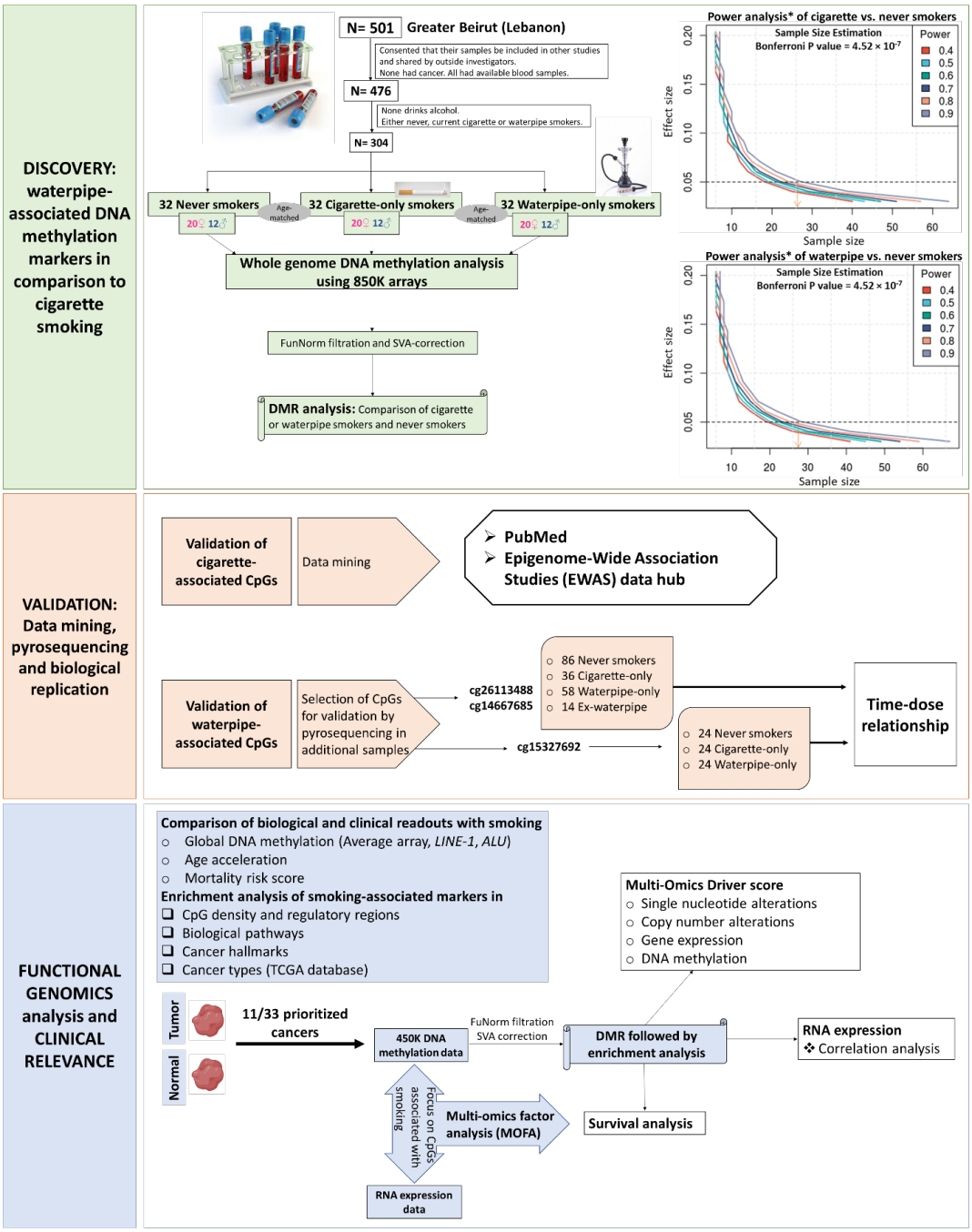


Figure 1. Flow chart of the study design.

*Power analysis and sample size estimation was performed for the comparison between smokers and never smokers using two-tailed t-test. Cigarette = current cigarette-only smokers, waterpipe = current waterpipe-only smokers, ex-waterpipe = former waterpipe-only smokers.

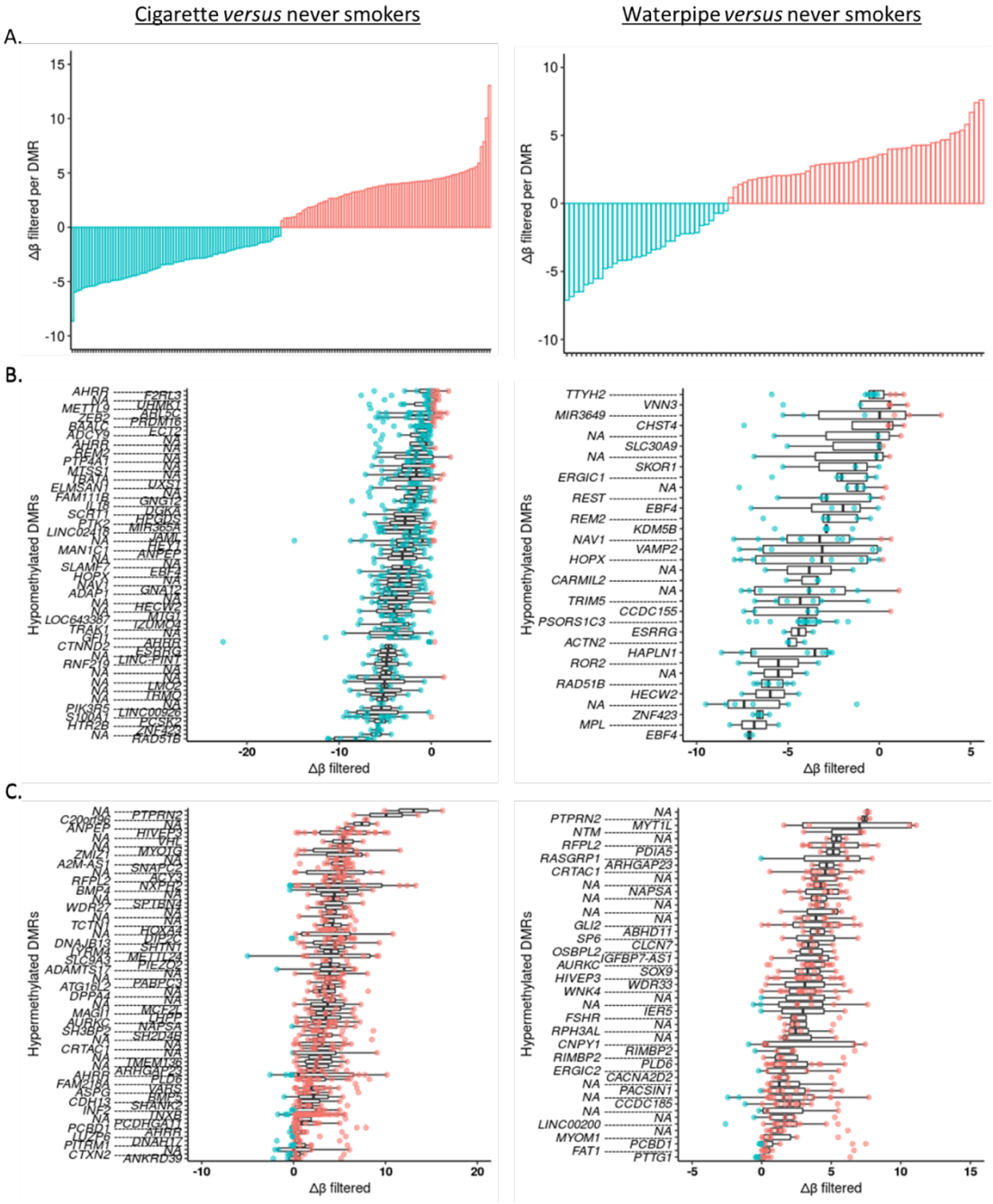


Figure 2. Characteristics of the DMRs associated with cigarette or waterpipe smoking.

Illustration of the DMRs associated with cigarette (LEFT) or waterpipe smoking (RIGHT) showing A: mean $\Delta\beta_{filtered}$ of DMRs; B: box plots of $\Delta\beta_{filtered}$ of CpGs within hypomethylated DMRs; C: box plots of $\Delta\beta_{filtered}$ of CpGs within hypermethylated DMRs. Green and red dots in B and C indicate the CpG sites that are hypomethylated and hypermethylated; respectively. Cigarette = current cigarette-only smokers, waterpipe = current waterpipe-only smokers.

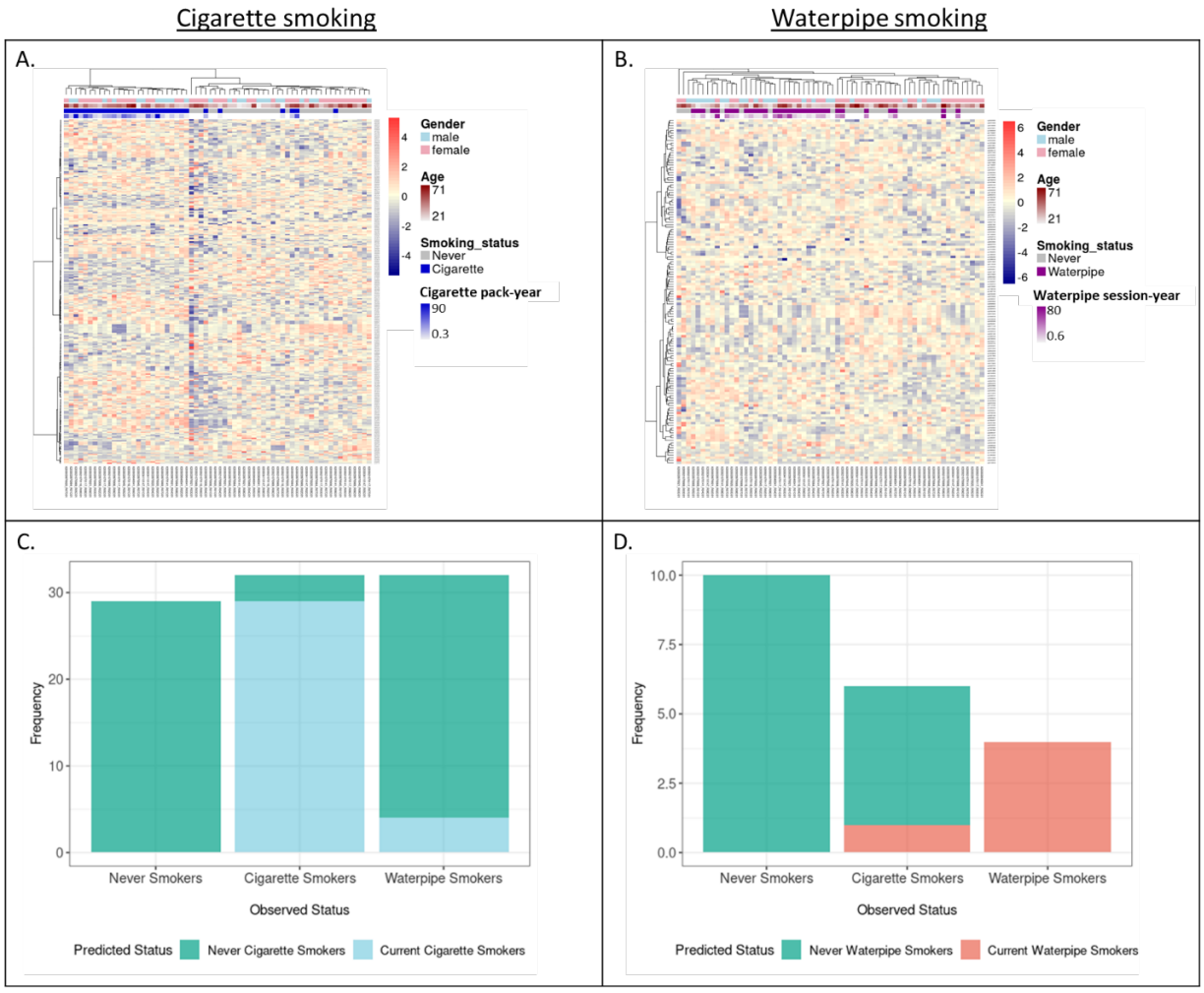


Figure 3. Heatmap clustering, machine learning-based prediction and time-dose response of smoking status.

A-B: Heatmap clustering of CpGs (with $|\Delta\beta_{filtered}| \geq 5\%$) present within the DMRs associated with either cigarette (A) or waterpipe smoking (B). Z-score scaling of rows was performed after clustering of FunNorm $\beta_{filtered}$ using Canberra distance dissimilarity measurement and Ward's hierarchal clustering. C-D: Visualization of observed versus predicted smoking status derived from lasso regression-based models trained to predict current cigarette smokers (C) and current waterpipe smokers (D). Cigarette = current cigarette-only smokers, waterpipe = current waterpipe-only smokers.

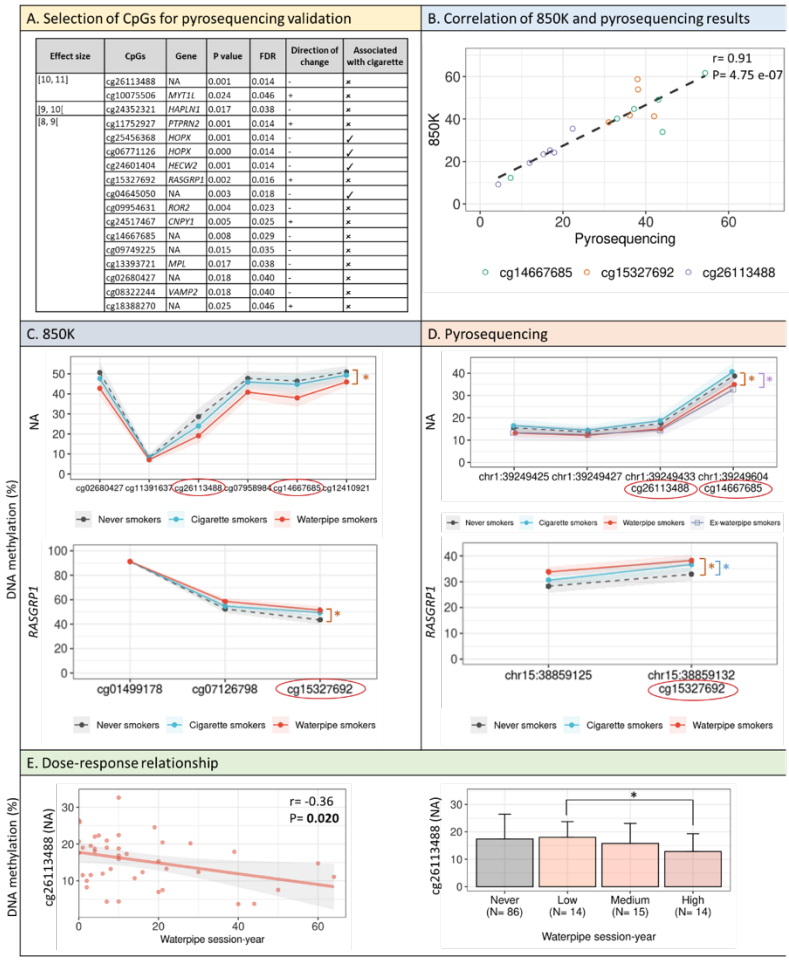


Figure 3. Technical validation and biological replication of waterpipe-associated DMRs
A: Characteristics of top significant CpGs with the highest effect size in waterpipe compared to never smokers. Student's t-test was performed for CpGs within DMRs associated with waterpipe smoking and having |Δβ| ≥ 5%. FDR-significant CpGs with the highest methylation change were reported. B: Array and pyrosequencing-based results of three CpGs assayed in 5-6 samples with the two methods. For each CpG, Spearman correlation coefficient (r) and P-value are shown. C-D: mean DNA methylation ± 95% confidence interval (CI) of the two waterpipe-associated DMRs covering the top CpGs prioritized for validation; LEFT: 850K results with validated CpGs in circles; RIGHT: pyrosequencing results on new samples for the NA region (86 never, 36 current cigarette-only, 58 current waterpipe-only smokers and 14 former waterpipe-only smokers) and the *RASGRP1* region (24 samples of each of never, current cigarette-only, and current waterpipe-only smokers). E: Pyrosequencing results of cg26113488 as a function of waterpipe session-year as continuous (LEFT) or categorized into tertile (RIGHT). LEFT: Pearson correlation coefficient and P value are shown, RIGHT: comparison of cg26113488 methylation across the tertiles was performed using linear regression. The e symbol stands for "×10". Cigarette smokers = current cigarette-only smokers; waterpipe smokers = current waterpipe-only smokers, ex-waterpipe smokers = former waterpipe-only smokers.

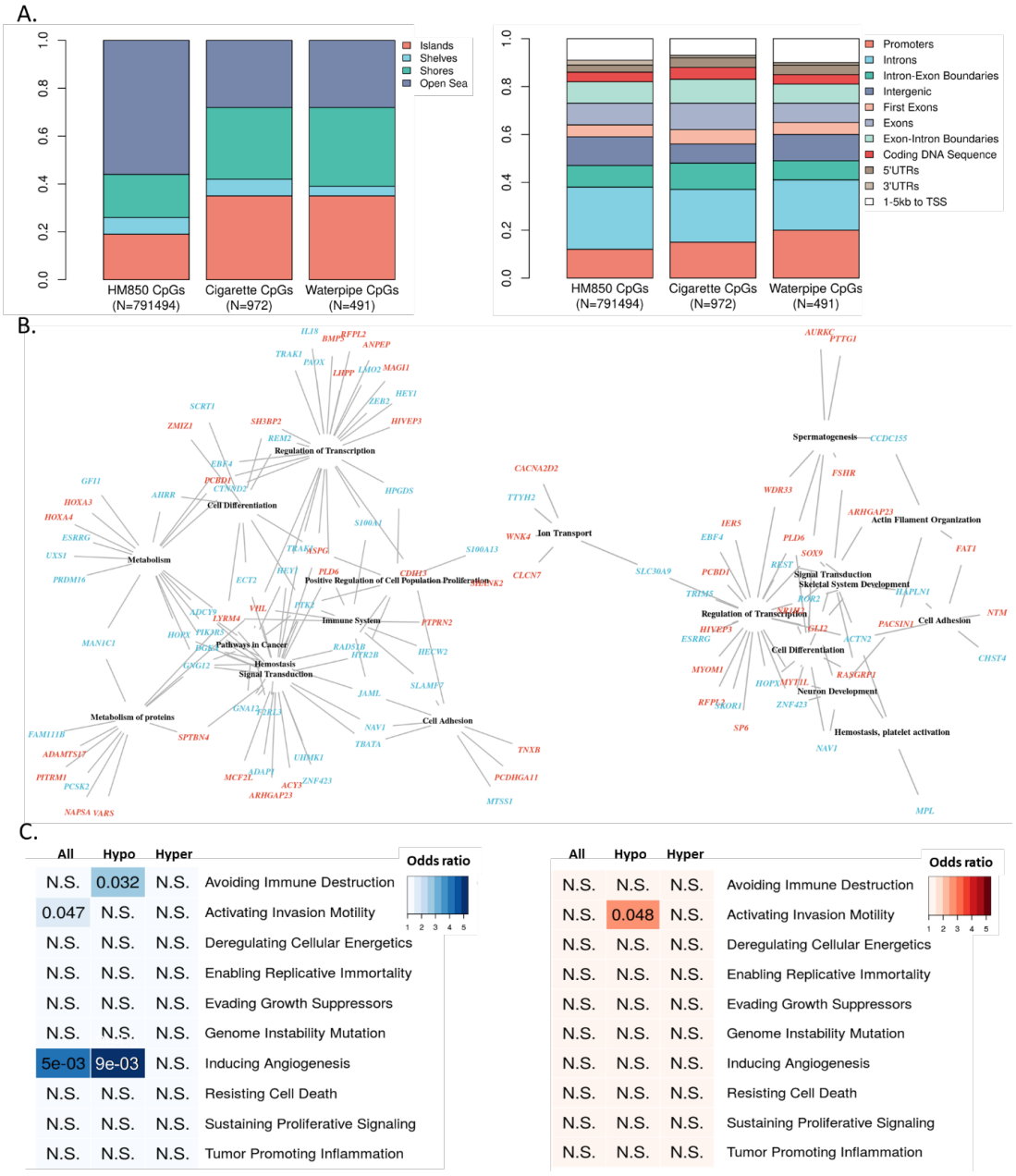
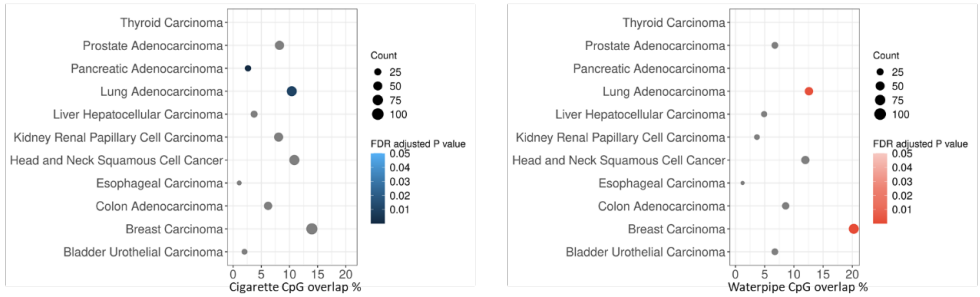
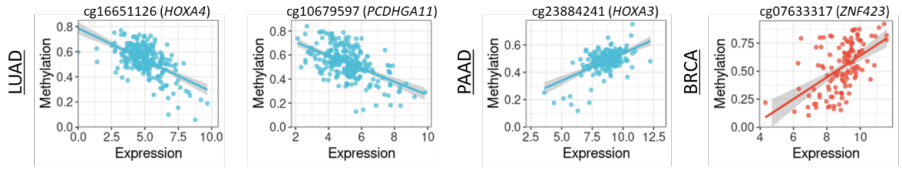


Figure 5. Enrichment of the DMRs associated with either cigarette or waterpipe smoking in CpG density cluster, regulatory regions, biological pathways and cancer hallmarks.
A: Distribution of the CpGs within the overall 850K array, cigarette DMRs and waterpipe DMRs by CpG density (LEFT) and regulatory regions (RIGHT). B: Visualization of the top ten FDR-significant pathways associated with differentially methylated genes of cigarette (LEFT) and waterpipe (RIGHT) using three databases available on GeneCodis web-based tool (https://genecodis.genyo.es/) (blue: hypomethylated; red: hypermethylated). C: Enrichment of the differentially methylated genes associated with cigarette (LEFT) and waterpipe (RIGHT) in 10 cancer hallmarks derived from Kiefer et al. (2017) (49). P-values are shown in the text. The e symbol stands for "× 10". Cigarette = current cigarette-only smokers; waterpipe = current waterpipe-only smokers. Hypo: hypomethylated, hyper: hypermethylated

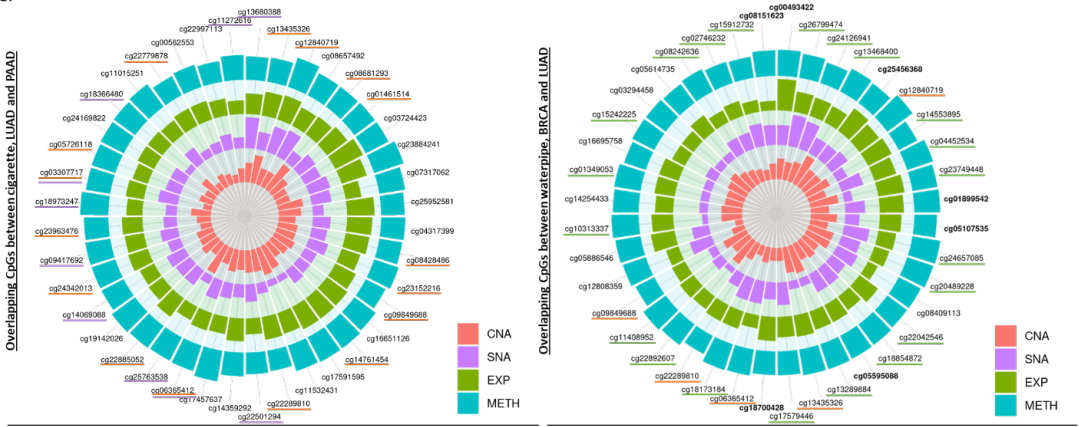
A.



B.



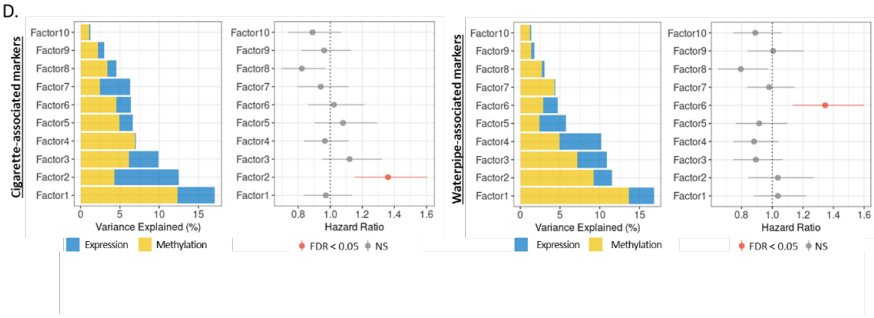
C.



CpG	Gene	Driver Score	CpG	Gene	Driver Score	CpG	Gene	Driver Score
cg13680388	*GNAS*	3.30	cg09849688	*ARID1A*	2.90	cg14069088	*CDKN2A*	2.71
cg13435326	*PIK3CA*	3.26	cg16651126	*HOXA4*	2.88	cg24342013	*RB1*	2.64
cg12840719	*CDKN2A*	3.22	cg14761454	*ATM*	2.86	cg09417692	*TGFBR2*	2.56
cg08657492	*HOXA4*	3.14	cg17591595	*HOXA4*	2.85	cg23963476	*SMARCA4*	2.55
cg08681293	*STK11*	3.13	cg11532431	*HOXA3*	2.83	cg18973247	*ARID1A*	2.51
cg01461514	*EGFR*	3.09	cg22289810	*NF1*	2.83	cg03307717	*U2AF1*	2.51
cg03724423	*HOXA4*	3.03	cg22501294	*SMAD4*	2.82	cg05726118	*CTNNB1*	2.47
cg23884241	*HOXA3*	3.02	cg14359292	*HOXA4*	2.81	cg24169822	*HOXA4*	2.46
cg07317062	*HOXA3*	2.98	cg17457637	*HOXA4*	2.78	cg18366480	*RNF43*	2.44
cg25952581	*HOXA4*	2.97	cg06365412	*TP53*	2.78	cg11015251	*HOXA4*	2.40
cg04317399	*HOXA3*	2.93	cg25763538	*KRAS*	2.77	cg22779878	*KEAP1*	2.37
cg08428486	*BRAF*	2.91	cg22885052	*SETD2*	2.77	cg00562553	*HOXA3*	2.35
cg23152216	*KRAS*	2.91	cg19142026	*HOXA3*	2.76	cg22997113	*HOXA3*	2.33
						cg11272616	*EEF2*	2.26

CpG	Gene	Driver Score	CpG	Gene	Driver Score	CpG	Gene	Driver Score
cg00493422	*HOPX*	3.30	cg08409113	*WNK4*	2.91	cg09849688	*ARID1A*	2.57
cg26799474	*CASP8*	3.24	cg22042546	*ARID1A*	2.86	cg12808359	*WNK4*	2.57
cg24126941	*FOXA1*	3.19	cg18854872	*SF3B1*	2.84	cg05886546	*WNK4*	2.56
cg13468400	*TP53*	3.12	cg05595088	*NA*	2.84	cg10313337	*CDH1*	2.54
cg25456368	*HOPX*	3.10	cg13289884	*CDKN1B*	2.80	cg14254433	*PACSIN1*	2.52
cg12840719	*CDKN2A*	3.07	cg13435326	*PIK3CA*	2.80	cg01349053	*FBXW7*	2.51
cg14553895	*PIK3CA*	3.04	cg17579446	*GPS2*	2.80	cg16695758	*WNK4*	2.51
cg04452534	*MAP3K1*	3.04	cg18700428	*NA*	2.73	cg15242225	*RUNX1*	2.49
cg23749448	*NF1*	3.04	cg06365412	*TP53*	2.65	cg03294458	*WNK4*	2.48
cg01899542	*ZNF423*	2.99	cg18173184	*CHD4*	2.64	cg05614735	*WNK4*	2.47
cg05107535	*NA*	2.98	cg22289810	*NF1*	2.61	cg08242636	*CBFB*	2.47
cg24657085	*ERBB2*	2.98	cg22892607	*GATA3*	2.60	cg02746232	*MAP2K4*	2.46
cg20489228	*TBX3*	2.97	cg11408952	*RB1*	2.57	cg15912732	*AKT1*	2.45
						cg08151623	*NA*	2.44

D.



E.

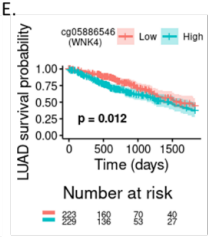


Figure 6. Pan-cancer and integrative omics analysis of cigarette- and waterpipe-associated DMRs in relation to cancer gene enrichment, functional impact on gene expression, cancer driver potential, and effect on cancer patient survival

A: Enrichment analysis of CpGs associated with cigarette (LEFT) or waterpipe (RIGHT) in those differentially methylated in tumor *versus* normal tissues of patients derived from The Cancer Genome Atlas (TCGA) database. Count: number of CpGs associated with both (cigarette or waterpipe) smoking and cancer; CpG overlap % is the percentage of smoking-associated CpGs also associated with cancer. B: Linear correlation plots between RNA expression (normalized counts) and methylation (beta, β) of CpGs overlapping between smoking and enriched cancers. Only FDR-significant CpGs with moderate Pearson correlation coefficient ($r > 0.5$) were plotted, and 1 CpG is shown per gene. C: Combined Multi-Omics Driver scores of CpGs common between smoking and enriched cancers across all enriched cancer types (LEFT: Cigarette; RIGHT: Waterpipe). Bold CpGs are common between cigarette and waterpipe smoking, and underlined CpGs are positive controls, that is, annotated to genes with a ConsensusDriver score of at least 1.5 (green: BRCA, orange: LUAD, purple: PAAD). D: Survival analysis of the top 10 latent factors derived using multi-omics factor analysis (MOFA) of methylation and expression data in lung cancer patients. Only methylation data of CpGs associated with smoking and the expression of their annotated genes were included in MOFA analysis (LEFT: cigarette; RIGHT: waterpipe). E: Kaplan-Meier Curve showing the effect of the top CpG significantly (FDR < 0.05) affecting cancer patient survival, and the CpG overlaps between waterpipe smoking and lung cancer DMRs. Abbreviations: BRCA: breast cancer, LUAD: lung adenocarcinoma, PAAD: pancreatic adenocarcinoma, CNA: copy number alteration, EXP: expression, METH: methylation; SNA: single nucleotide alteration, cigarette = current cigarette-only smokers, waterpipe = current waterpipe-only smokers.

7. Perspectivas Futuras

A observação de que a exposição à radiação ultra-violeta produz alterações epigenéticas abre um leque enorme de possibilidades de futuros trabalhos. Um próximo passo óbvio e necessário a partir dos resultados obtidos nesse trabalho é tentar validar o papel da metilação dos principais genes (*HOXC9*, *KCNQ1DN*, *MGMT*, *TAPBP*, *ERICH3*, *FINL2*, *ZNF732*, *SLC6A18*, *MFSD13A*, *SLFN12L* e *IFNLR1*) na tumorigênese de melanoma devido a exposição à radiação ultra-violeta em uma corte maior e em grupos mais homogêneos de pacientes (por exemplo, com etnicidade similar e com subtipos tumorais únicos). Também seria interessante aprofundar o estudo do papel da metilação desses genes como biomarcadores clínicos, seja prognóstico ou de resposta a diversos tratamentos.

Além disso, ensaios funcionais seriam de grande ajuda para entender o mecanismo pelo qual a metilação dos genes validados, especialmente o *TAPBP*, afeta o melanoma. Alguns dos principais experimentos seriam testar como o bloqueio da metilação do DNA influencia a expressão do *TAPBP* e se isso pode influenciar ainda mais a proliferação, invasão e/ou metástase.

Até o momento da escrita dessa tese, é pouco conhecido o papel da metilação do DNA na tumorigênese de melanoma, portanto, seria valioso estudar o que difere em termos epigenéticos diferentes nevos associados ao melanoma, bem como diferentes estágios da doença. Isso permitiria uma melhor classificação da doença, se a metilação refletisse a identidade de diferentes células de origem que iniciam os diferentes tipos de melanoma existentes. A consequência seria um tratamento muito mais eficaz para os pacientes e uma qualidade de vida muito melhor.

8. Divulgação na Mídia Nacional e Internacional

O presente trabalho teve ampla divulgação e alcance nas mídias nacionais e internacionais. Os links das principais matérias encontram-se abaixo (todos os links acessados em 23/04/2023):

Agência FAPESP:
https://agencia.fapesp.br/estudo-mapeia-alteracoes-no-dna-que-tornam-o-melanoma-mais-grave/39806/

Site G1:
https://g1.globo.com/sp/ribeirao-preto-franca/noticia/2022/10/04/melanoma-entenda-como-estudo-no-hospital-de-amor-de-barretos-pode-ajudar-em-novos-tratamentos.ghtml

Entrevista para o programa "Bom dia Cidade", da TV Globo:
https://globoplay.globo.com/v/10994571/

Revista Veja:
https://veja.abril.com.br/saude/brasileiros-mapeiam-alteracoes-no-dna-que-tornam-melanoma-mais-grave/

Site UOL:
https://www.uol.com.br/vivabem/noticias/redacao/2022/10/14/estudo-mapeia-alteracoes-no-dna-que-tornam-o-melanoma-mais-grave.htm

Revista Galileu:
https://revistagalileu.globo.com/saude/noticia/2022/10/estudo-mapeia-alteracoes-no-dna-que-tornam-o-melanoma-mais-grave.ghtml

Site Terra:
https://www.terra.com.br/byte/estudo-mapeia-mudancas-no-dna-que-aumentam-chances-de-cancer-de-pele,0248cb92f321d5e0f14d5e42581837c789e91ys8.html

Site da Organização Mundial de Saúde:

https://www.iarc.who.int/news-events/multi-ethnic-study-uncovers-unique-origins-of-melanoma-types-and-actionable-molecular-targets/

Site na Alemanha:

https://biermann-medizin.de/cross-omics-bei-kutanen-und-akralen-melanomen-mit-pathobiologie-und-uv-exposition-assoziierte-krebstreiber-identifiziert/

Sites em Inglês:

https://scienmag.com/study-maps-alterations-that-make-melanoma-more-severe-could-lead-to-better-therapies-and-diagnosis/

https://medicalxpress.com/news/2022-10-melanoma-severe-therapies-diagnosis.html

https://www.news-medical.net/news/20221011/Study-offers-new-insights-into-other-melanomas-not-caused-by-the-effects-of-UV-radiation.aspx

https://www.technologynetworks.com/tn/news/study-maps-genetic-markers-linked-to-melanoma-survival-366493

https://www.miragenews.com/research-maps-alterations-that-make-melanoma-872318/

https://newsbulletin247.com/healthcare/189289.html

https://justbalancinghealth.com/health-news/study-offers-new-insights-into-other-melanomas-not-caused-by-the-effects-of-uv-radiation/

https://newsatw.com/study-maps-alterations-that-make-melanoma-more-severe-could-lead-to-better-therapies-and-diagnosis/

9. Impacto Social

O trabalho principal fruto dessa tese de doutorado descreve pela primeira vez que a exposição à radiaçâo ultra-violeta modifica o DNA de pacientes com melanoma por meio de mecanismos epigenéticos. Sendo assim, os impactos desse trabalho para a sociedade incluem desde a contribuição de ciência básica, elucidando pela primeira vez esse mecanismo, bem como a descrição de genes envolvidos nessa resposta à radiação ultra-violeta, que podem

servir de alvos moleculares para futuras terapias e serem alvos para detecção precoce de melanoma.

Além disso, descrevemos pela primeira vez com base em uma abordagem integrada de vários dados genômicos, que a classificação atual do melanoma por meio de histopatologia deveria ser revista, uma vez que molecularmente esses subtipos da doença não condizem com a atual classificação, o que poderia impactar no manejo e, consequentemente, na sobrevida dos pacientes.

Como demonstrado, o presente trabalho foi publicado em uma revista científica de alto impacto e amplamente divulgado pelas mídias, mostrando a relevância do mesmo para a população.

Ainda, a descrição de um protocolo viável de ser implementado em uma rotina de diagnóstico molecular que aumenta a possibilidade que os pacientes com melanoma com lesões altamente pigmentadas tenham um laudo molecular conclusivo para melhor decisão terapêutica, é um ganho importante para a sociedade.

Por fim, frente ao descaso governamental nos últimos anos com a ciência e a descrença nela por certa parte da população, essa tese de doutorado se soma a tantas outras para mostrar como há ciência de excelência realizada em nosso país, com impactos reais para a sociedade. Infelizmente, esse tipo de prova talvez seja o melhor impacto social a curto prazo que se pode oferecer ao país onde atualmente vivemos.

10. Referências Bibliográficas

1. Robertis Ed. *Bases da Biologia Celular e Molecular*. **Guanabara Koogan**2006. 389 p.

2. Montanari T. *Histologia: Texto, atlas e roteiro de aulas práticas*. **Editora da UFRGS**2016. 229 p.

3. Kaushik SB, Kaushik N. *Non-coding RNAs in skin cancers: An update*. **Noncoding RNA Res**. 2016;1(1):83-6.

4. Malacinski G. *Fundamentos de Biologia Molecular*. **Guanabara Koogan**2005. 439 p.

5. Moore LD, Le T, Fan G. *DNA methylation and its basic function*. **Neuropsychopharmacology**. 2013;38(1):23-38.

6. Tirado-Magallanes R, Rebbani K, Lim R, Pradhan S, Benoukraf T. *Whole genome DNA methylation: beyond genes silencing*. **Oncotarget**. 2017;8(3):5629-37.

7. Sadler TW, Jan L. *Langma's Medical Embriology*. 12 ed**: Philaderphia**; 2012.

8. Singh V. *Clinical Embriology***: Elsevier**; 2012.

9. Ernfors P. *Cellular origin and developmental mechanisms during the formation of skin melanocytes*. **Exp Cell Res**. 2010;316(8):1397-407.

10. Vandamme N, Berx G. *From neural crest cells to melanocytes: cellular plasticity during development and beyond*. **Cell Mol Life Sci**. 2019;76(10):1919-34.

11. Park HY, Kosmadaki M, Yaar M, Gilchrest BA. *Cellular mechanisms regulating human melanogenesis*. **Cell Mol Life Sci**. 2009;66(9):1493-506.

12. Junqueira LCU, Carneiro J. *Histologia Básica*. 12 ed**: Guanabara Koogan**; 2013.

13. Shain AH, Bastian BC. *From melanocytes to melanomas*. **Nat Rev Cancer**. 2016;16(6):345-58.

14. Cichorek M, Wachulska M, Stasiewicz A, Tyminska A. *Skin melanocytes: biology and development*. **Postepy Dermatol Alergol**. 2013;30(1):30-41.

15. Peltomaki P. *Mutations and epimutations in the origin of cancer*. **Exp Cell Res**. 2012;318(4):299-310.

16. Hanahan D, Weinberg RA. *Hallmarks of cancer: the next generation*. **Cell**. 2011;144(5):646-74.

17. Nagy R, Sweet K, Eng C. *Highly penetrant hereditary cancer syndromes*. **Oncogene**. 2004;23(38):6445-70.

18. Hanahan D, Weinberg RA. *The hallmarks of cancer*. **Cell**. 2000;100(1):57-70.

19. Houghton AN, Polsky D. *Focus on melanoma*. **Cancer Cell**. 2002;2(4):275-8.

20. Shain AH, Yeh I, Kovalyshyn I, Sriharan A, Talevich E, Gagnon A, et al. *The Genetic Evolution of Melanoma from Precursor Lesions*. **N Engl J Med**. 2015;373(20):1926-36.

21. Criscito MC, Polsky D, Stein JA. *The Genetic Evolution of Melanoma*. **N Engl J Med**. 2016;374(10):993.

22. Tsao H, Bevona C, Goggins W, Quinn T. *The transformation rate of moles (melanocytic nevi) into cutaneous melanoma: a population-based estimate*. **Arch Dermatol**. 2003;139(3):282-8.

23. Duffy K, Grossman D. *The dysplastic nevus: from historical perspective to management in the modern era: part I. Historical, histologic, and clinical aspects*. **J Am Acad Dermatol**. 2012;67(1):1 e-16; quiz 7-8.

24. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. *Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries*. **CA Cancer J Clin**. 2021;71(3):209-49.

25. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries*. **CA Cancer J Clin**. 2018;68(6):394-424.

26. Arnold M, Singh D, Laversanne M, Vignat J, Vaccarella S, Meheus F, et al. *Global Burden of Cutaneous Melanoma in 2020 and Projections to 2040*. **JAMA Dermatol**. 2022;158(5):495-503.

27. INCA. *Estimativa 2020 : incidência de câncer no Brasil* 2020.

28. HCB. *Infográficos 2019 do Registro de Câncer do Hospital de Câncer de Barretos*. [Internet] 2019;Available from: **https://infogram.com/copy-hospital-de-amor-melanoma-2019-1hkv2nkkvyzp2x3**.

29. Whiteman DC, Green AC, Olsen CM. *The Growing Burden of Invasive Melanoma: Projections of Incidence Rates and Numbers of New Cases in Six Susceptible Populations through 2031*. **J Invest Dermatol**. 2016;136(6):1161-71.

30. Cameron JK, Baade P. *Projections of the future burden of cancer in Australia using Bayesian age-period-cohort models*. **Cancer Epidemiol**. 2021;72:101935.

31. Baade PD, Green AC, Smithers BM, Aitken JF. *Trends in melanoma incidence among children: possible influence of sun-protection programs*. **Expert Rev Anticancer Ther**. 2011;11(5):661-4.

32. Czarnecki D, Meehan CJ. *Is the incidence of malignant melanoma decreasing in young Australians?* **J Am Acad Dermatol**. 2000;42(4):672-4.

33. Czarnecki D. *The incidence of melanoma is increasing in the susceptible young Australian population*. **Acta Derm Venereol**. 2014;94(5):539-41.

34. Baade PD, Youlden DR, Youl P, Kimlin M, Sinclair C, Aitken J. *Assessment of the effect of migration on melanoma incidence trends in Australia between 1982 and 2010 among people under 30*. **Acta Derm Venereol**. 2015;95(1):118-20.

35. Garcia-Souto F, Duran-Romero AJ, Pereyra-Rodriguez JJ. *Melanoma mortality in Spain: predictions up to 2043*. **Int J Dermatol**. 2021;60(7):844-50.

36. van Niekerk CC, Groenewoud HMM, Verbeek ALM. *Trends and projections in cutaneous melanoma death in the Netherlands from 1950 to 2045*. **Medicine (Baltimore)**. 2021;100(48):e27784.

37. Rahib L, Wehner MR, Matrisian LM, Nead KT. *Estimated Projection of US Cancer Incidence and Death to 2040*. **JAMA Netw Open**. 2021;4(4):e214708.

38. Santos CAD, Souza DLB. *Melanoma mortality in Brazil: trends and projections (1998-2032)*. **Cien Saude Colet**. 2019;24(4):1551-61.

39. de Melo AC, Wainstein AJA, Buzaid AC, Thuler LCS. *Melanoma signature in Brazil: epidemiology, incidence, mortality, and trend lessons from a continental mixed population country in the past 15 years*. **Melanoma Res**. 2018;28(6):629-36.

40. Nader Marta G, Munhoz RR, Teixeira MP, Waldvogel BC, Pires de Camargo V, Feher O, et al. *Trends in Melanoma Mortality in Brazil: A Registry-Based Study*. **JCO Glob Oncol**. 2020;6:1766-71.

41. Friedman RJ, Rigel DS, Kopf AW. *Early detection of malignant melanoma: the role of physician examination and self-examination of the skin*. **CA Cancer J Clin**. 1985;35(3):130-51.

42. Abbasi NR, Shaw HM, Rigel DS, Friedman RJ, McCarthy WH, Osman I, et al. *Early diagnosis of cutaneous melanoma: revisiting the ABCD criteria*. **JAMA**. 2004;292(22):2771-6.

43. Lens M. *Current clinical overview of cutaneous melanoma*. **Br J Nurs**. 2008;17(5):300-5.

44. Menzies SW, Ingvar C, McCarthy WH. *A sensitivity and specificity analysis of the surface microscopy features of invasive melanoma*. **Melanoma Res**. 1996;6(1):55-62.

45. Balch CM, Gershenwald JE, Soong SJ, Thompson JF, Atkins MB, Byrd DR, et al. *Final version of 2009 AJCC melanoma staging and classification*. **J Clin Oncol**. 2009;27(36):6199-206.

46. Bartlett EK, Karakousis GC. *Current staging and prognostic factors in melanoma*. **Surg Oncol Clin N Am**. 2015;24(2):215-27.

47. Perrotta R, Bevelacqua Y, Malaguarnera G, Paladina I, Giordano M, Malaguarnera M. *Serum markers of cutaneous melanoma*. **Front Biosci (Elite Ed)**. 2010;2:1115-22.

48. Amin MB, Edge SB, FL G. *AJCC Cancer Staging Manual*. 8, editor. **New York: Springer International Publishing**; 2017.

49. Gershenwald JE, Scolyer RA, Hess KR, Sondak VK, Long GV, Ross MI, et al. *Melanoma staging: Evidence-based changes in the American Joint Committee on Cancer eighth edition cancer staging manual*. **CA Cancer J Clin**. 2017;67(6):472-92.

50. Clark WH, Jr., From L, Bernardino EA, Mihm MC. *The histogenesis and biologic behavior of primary human malignant melanomas of the skin*. **Cancer Res**. 1969;29(3):705-27.

51. Duncan LM. *The classification of cutaneous melanoma*. **Hematol Oncol Clin North Am**. 2009;23(3):501-13, ix.

52. Merkel EA, Gerami P. *Malignant melanoma of sun-protected sites: a review of clinical, histological, and molecular features*. **Lab Invest**. 2017;97(6):630-5.

53. El Ghissassi F, Baan R, Straif K, Grosse Y, Secretan B, Bouvard V, et al. *A review of human carcinogens--part D: radiation*. **Lancet Oncol**. 2009;10(8):751-2.

54. Islami F, Goding Sauer A, Miller KD, Siegel RL, Fedewa SA, Jacobs EJ, et al. *Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States*. **CA Cancer J Clin**. 2018;68(1):31-54.

55. Berwick M, Buller DB, Cust A, Gallagher R, Lee TK, Meyskens F, et al. *Melanoma Epidemiology and Prevention*. **Cancer Treat Res**. 2016;167:17-49.

56. Green AC, Williams GM, Logan V, Strutton GM. *Reduced melanoma after regular sunscreen use: randomized trial follow-up*. **J Clin Oncol**. 2011;29(3):257-63.

57. Dennis LK, Vanbeek MJ, Beane Freeman LE, Smith BJ, Dawson DV, Coughlin JA. *Sunburns and risk of cutaneous melanoma: does age matter? A comprehensive meta-analysis*. **Ann Epidemiol**. 2008;18(8):614-27.

58. Lazovich D, Isaksson Vogel R, Weinstock MA, Nelson HH, Ahmed RL, Berwick M. *Association Between Indoor Tanning and Melanoma in Younger Men and Women*. **JAMA Dermatol**. 2016;152(3):268-75.

59. Ghiasvand R, Rueegg CS, Weiderpass E, Green AC, Lund E, Veierod MB. *Indoor Tanning and Melanoma Risk: Long-Term Evidence From a Prospective Population-Based Cohort Study*. **Am J Epidemiol**. 2017;185(3):147-56.

60. Boniol M, Autier P, Boyle P, Gandini S. *Cutaneous melanoma attributable to sunbed use: systematic review and meta-analysis*. **BMJ**. 2012;345:e4757.

61. Pierret L, Suppa M, Gandini S, Del Marmol V, Gutermuth J. *Overview on vitamin D and sunbed use*. **J Eur Acad Dermatol Venereol**. 2019;33 Suppl 2:28-33.

62. Thieden E, Jorgensen HL, Jorgensen NR, Philipsen PA, Wulf HC. *Sunbed radiation provokes cutaneous vitamin D synthesis in humans--a randomized controlled trial*. **Photochem Photobiol**. 2008;84(6):1487-92.

63. Lagunova Z, Porojnicu AC, Aksnes L, Holick MF, Iani V, Bruland OS, et al. *Effect of vitamin D supplementation and ultraviolet B exposure on serum 25-hydroxyvitamin D concentrations in healthy volunteers: a randomized, crossover clinical trial*. **Br J Dermatol**. 2013;169(2):434-40.

64. Langdahl JH, Schierbeck LL, Bang UC, Jensen JE. *Changes in serum 25-hydroxyvitamin D and cholecalciferol after one whole-body exposure in a commercial tanning bed: a randomized study*. **Endocrine**. 2012;42(2):430-5.

65. Kimball SM, Lee J, Vieth R. *Sunbeds with UVB radiation can produce physiological levels of serum 25-Hydroxyvitamin D in healthy volunteers*. **Dermatoendocrinol**. 2017;9(1):e1375635.

66. de Gruijl FR, Pavel S. *The effects of a mid-winter 8-week course of sub-sunburn sunbed exposures on tanning, vitamin D status and colds*. **Photochem Photobiol Sci**. 2012;11(12):1848-54.

67. Weber B, Bachmann CC, Braun R, Abraham AG, Serra AL, Hofbauer GFL. *25-Hydroxyvitamin-D3 serum modulation after use of sunbeds compliant with European Union standards: A randomized open observational controlled trial*. **J Am Acad Dermatol**. 2017;77(1):48-54.

68. Moan J, Lagunova Z, Cicarma E, Aksnes L, Dahlback A, Grant WB, et al. *Sunbeds as vitamin D sources*. **Photochem Photobiol**. 2009;85(6):1474-9.

69. Gerber B, Mathys P, Moser M, Bressoud D, Braun-Fahrlander C. *Ultraviolet emission spectra of sunbeds*. **Photochem Photobiol**. 2002;76(6):664-8.

70. Tierney P, Ferguson J, Ibbotson S, Dawe R, Eadie E, Moseley H. *Nine out of 10 sunbeds in England emit ultraviolet radiation levels that exceed current safety limits*. **Br J Dermatol**. 2013;168(3):602-8.

71. Suppa M, Gandini S. *Sunbeds and melanoma risk: time to close the debate*. **Curr Opin Oncol**. 2019;31(2):65-71.

72. Holm RP. *Skin cancer prevention and screening*. **S D Med**. 2015;Spec No:75-7, 9-81.

73. ANVISA. *Resolução número 56*. [Internet] 2009;Available from: **https://bvsms.saude.gov.br/bvs/saudelegis/anvisa/2009/res0056_09_11_2009.html**.

74. Saginala K, Barsouk A, Aluru JS, Rawla P, Barsouk A. *Epidemiology of Melanoma*. **Med Sci (Basel)**. 2021;9(4).

75. Holly EA, Kelly JW, Shpall SN, Chiu SH. *Number of melanocytic nevi as a major risk factor for malignant melanoma*. **J Am Acad Dermatol**. 1987;17(3):459-68.

76. Halpern AC, Guerry Dt, Elder DE, Clark WH, Jr., Synnestvedt M, Norman S, et al. *Dysplastic nevi as risk markers of sporadic (nonfamilial) melanoma. A case-control study*. **Arch Dermatol**. 1991;127(7):995-9.

77. Gandini S, Sera F, Cattaruzza MS, Pasquini P, Abeni D, Boyle P, et al. *Meta-analysis of risk factors for cutaneous melanoma: I. Common and atypical naevi*. **Eur J Cancer**. 2005;41(1):28-44.

78. Watt AJ, Kotsis SV, Chung KC. *Risk of melanoma arising in large congenital melanocytic nevi: a systematic review*. **Plast Reconstr Surg**. 2004;113(7):1968-74.

79. Silva JH, Sa BC, Avila AL, Landman G, Duprat Neto JP. *Atypical mole syndrome and dysplastic nevi: identification of populations at risk for developing melanoma - review article*. **Clinics (Sao Paulo)**. 2011;66(3):493-9.

80. Friedman RJ, Farber MJ, Warycha MA, Papathasis N, Miller MK, Heilman ER. *The "dysplastic" nevus*. **Clin Dermatol**. 2009;27(1):103-15.

81. Tucker MA, Halpern A, Holly EA, Hartge P, Elder DE, Sagebiel RW, et al. *Clinically recognized dysplastic nevi. A central risk factor for cutaneous melanoma*. **JAMA**. 1997;277(18):1439-44.

82. Purdue MP, From L, Armstrong BK, Kricker A, Gallagher RP, McLaughlin JR, et al. *Etiologic and other factors predicting nevus-associated cutaneous malignant melanoma*. **Cancer Epidemiol Biomarkers Prev**. 2005;14(8):2015-22.

83. Goldstein AM, Tucker MA. *Genetic epidemiology of cutaneous melanoma: a global perspective*. **Arch Dermatol**. 2001;137(11):1493-6.

84. Soura E, Eliades PJ, Shannon K, Stratigos AJ, Tsao H. *Hereditary melanoma: Update on syndromes and management: Genetics of familial atypical multiple mole melanoma syndrome*. **J Am Acad Dermatol**. 2016;74(3):395-407; quiz 8-10.

85. Dzwierzynski WW. *Melanoma Risk Factors and Prevention*. **Clin Plast Surg**. 2021;48(4):543-50.

86. Cancer Genome Atlas N. *Genomic Classification of Cutaneous Melanoma*. **Cell**. 2015;161(7):1681-96.

87. Hayward NK, Wilmott JS, Waddell N, Johansson PA, Field MA, Nones K, et al. *Whole-genome landscapes of major melanoma subtypes*. **Nature**. 2017;545(7653):175-80.

88. Wouters J, Vizoso M, Martinez-Cardus A, Carmona FJ, Govaere O, Laguna T, et al. *Comprehensive DNA methylation study identifies novel progression-related and prognostic markers for cutaneous melanoma*. **BMC Med**. 2017;15(1):101.

89. Trucco LD, Mundra PA, Hogan K, Garcia-Martinez P, Viros A, Mandal AK, et al. *Ultraviolet radiation-induced DNA damage is prognostic for outcome in melanoma*. **Nat Med**. 2019;25(2):221-4.

90. Mundra PA, Dhomen N, Rodrigues M, Mikkelsen LH, Cassoux N, Brooks K, et al. *Ultraviolet radiation drives mutations in a subset of mucosal melanomas*. **Nat Commun**. 2021;12(1):259.

91. Sladden MJ, Nieweg OE, Howle J, Coventry BJ, Thompson JF. *Updated evidence-based clinical practice guidelines for the diagnosis and management of melanoma: definitive excision margins for primary cutaneous melanoma*. **Med J Aust**. 2018;208(3):137-42.

92. Piepkorn M, Weinstock MA, Barnhill RL. *Theoretical and empirical arguments in relation to elective lymph node dissection for melanoma*. **Arch Dermatol**. 1997;133(8):995-1002.

93. Morton DL, Hoon DS, Cochran AJ, Turner RR, Essner R, Takeuchi H, et al. *Lymphatic mapping and sentinel lymphadenectomy for early-stage melanoma: therapeutic utility and implications of nodal microanatomy and molecular staging for improving the accuracy of detection of nodal micrometastases*. **Ann Surg**. 2003;238(4):538-49; discussion 49-50.

94. van der Ploeg AP, van Akkooi AC, Haydu LE, Scolyer RA, Murali R, Verhoef C, et al. *The prognostic significance of sentinel node tumour burden in melanoma patients: an international, multicenter study of 1539 sentinel node-positive melanoma patients*. **Eur J Cancer**. 2014;50(1):111-20.

95. Leiter U, Stadler R, Mauch C, Hohenberger W, Brockmeyer N, Berking C, et al. *Complete lymph node dissection versus no dissection in patients with sentinel lymph node biopsy positive melanoma (DeCOG-SLT): a multicentre, randomised, phase 3 trial*. **Lancet Oncol**. 2016;17(6):757-67.

96. Fioranelli M, Roccia MG, Pastore C, Aracena CJ, Lotti T. *Completion dissection or observation for sentinel-node metastasis in melanoma*. **Dermatol Ther**. 2017;30(6).

97. Wong SL, Faries MB, Kennedy EB, Agarwala SS, Akhurst TJ, Ariyan C, et al. *Sentinel Lymph Node Biopsy and Management of Regional Lymph Nodes in Melanoma: American Society of Clinical Oncology and Society of Surgical Oncology Clinical Practice Guideline Update*. **J Clin Oncol**. 2018;36(4):399-413.

98. Eggermont AMM, Blank CU, Mandala M, Long GV, Atkinson V, Dalle S, et al. *Adjuvant Pembrolizumab versus Placebo in Resected Stage III Melanoma*. **N Engl J Med**. 2018;378(19):1789-801.

99. Long GV, Hauschild A, Santinami M, Atkinson V, Mandala M, Chiarion-Sileni V, et al. *Adjuvant Dabrafenib plus Trametinib in Stage III BRAF-Mutated Melanoma*. **N Engl J Med**. 2017;377(19):1813-23.

100. Romano E, Schwartz GK, Chapman PB, Wolchock JD, Carvajal RD. *Treatment implications of the emerging molecular classification system for melanoma*. **Lancet Oncol**. 2011;12(9):913-22.

101. Curtin JA, Fridlyand J, Kageshita T, Patel HN, Busam KJ, Kutzner H, et al. *Distinct sets of genetic alterations in melanoma*. **N Engl J Med**. 2005;353(20):2135-47.

102. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, et al. *Mutations of the BRAF gene in human cancer*. **Nature**. 2002;417(6892):949-54.

103. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, et al. *Improved survival with vemurafenib in melanoma with BRAF V600E mutation*. **N Engl J Med**. 2011;364(26):2507-16.

104. Domingues B, Lopes JM, Soares P, Populo H. *Melanoma treatment in review*. **Immunotargets Ther**. 2018;7:35-49.

105. Davies MA, Gershenwald JE. *Targeted therapy for melanoma: a primer*. **Surg Oncol Clin N Am**. 2011;20(1):165-80.

106. Flaherty KT, Robert C, Hersey P, Nathan P, Garbe C, Milhem M, et al. *Improved survival with MEK inhibition in BRAF-mutated melanoma*. **N Engl J Med**. 2012;367(2):107-14.

107. Flaherty KT, Infante JR, Daud A, Gonzalez R, Kefford RF, Sosman J, et al. *Combined BRAF and MEK inhibition in melanoma with BRAF V600 mutations*. **N Engl J Med**. 2012;367(18):1694-703.

108. Carlino MS, Larkin J, Long GV. *Immune checkpoint inhibitors in melanoma*. **Lancet**. 2021;398(10304):1002-14.

109. Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, et al. *Improved survival with ipilimumab in patients with metastatic melanoma*. **N Engl J Med**. 2010;363(8):711-23.

110. Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, et al. *Safety, activity, and immune correlates of anti-PD-1 antibody in cancer*. **N Engl J Med**. 2012;366(26):2443-54.

111. Eggermont AMM, Robert C, Ribas A. *The new era of adjuvant therapies for melanoma*. **Nat Rev Clin Oncol**. 2018;15(9):535-6.

112. Eggermont AM, Chiarion-Sileni V, Grob JJ, Dummer R, Wolchok JD, Schmidt H, et al. *Prolonged Survival in Stage III Melanoma with Ipilimumab Adjuvant Therapy*. **N Engl J Med**. 2016;375(19):1845-55.

113. Weber J, Mandala M, Del Vecchio M, Gogas HJ, Arance AM, Cowey CL, et al. *Adjuvant Nivolumab versus Ipilimumab in Resected Stage III or IV Melanoma*. **N Engl J Med**. 2017;377(19):1824-35.

114. Jenkins RW, Fisher DE. *Treatment of Advanced Melanoma in 2020 and Beyond*. **J Invest Dermatol**. 2021;141(1):23-31.

115. Tawbi HA, Schadendorf D, Lipson EJ, Ascierto PA, Matamala L, Castillo Gutierrez E, et al. *Relatlimab and Nivolumab versus Nivolumab in Untreated Advanced Melanoma*. **N Engl J Med**. 2022;386(1):24-34.

116. Versluis JM, Long GV, Blank CU. *Learning from clinical trials of neoadjuvant checkpoint blockade*. **Nat Med**. 2020;26(4):475-84.

117. Singh BP, Salama AK. *Updates in Therapy for Advanced Melanoma*. **Cancers (Basel)**. 2016;8(1).

118. da Costa LMM, Crovador CS, de Carvalho CEB, Vazquez VL. *Characteristics of Brazilian melanomas: real-world results before and after the introduction of new therapies*. **BMC Res Notes**. 2019;12(1):296.

119. McKenzie RL, Aucamp PJ, Bais AF, Bjorn LO, Ilyas M. *Changes in biologically-active ultraviolet radiation reaching the Earth's surface*. **Photochem Photobiol Sci**. 2007;6(3):218-31.

120. Wang SQ, Setlow R, Berwick M, Polsky D, Marghoob AA, Kopf AW, et al. *Ultraviolet A and melanoma: a review*. **J Am Acad Dermatol**. 2001;44(5):837-46.

121. Sun X, Zhang N, Yin C, Zhu B, Li X. *Ultraviolet Radiation and Melanomagenesis: From Mechanism to Immunotherapy*. **Front Oncol**. 2020;10:951.

122. Brash DE. *UV signature mutations*. **Photochem Photobiol**. 2015;91(1):15-26.

123. Ikehata H, Ono T. *The mechanisms of UV mutagenesis*. **J Radiat Res**. 2011;52(2):115-25.

124. Tommasi S, Denissenko MF, Pfeifer GP. *Sunlight induces pyrimidine dimers preferentially at 5-methylcytosine bases*. **Cancer Res**. 1997;57(21):4727-30.

125. Martinez-Fernandez L, Banyasz A, Esposito L, Markovitsi D, Improta R. *UV-induced damage to DNA: effect of cytosine methylation on pyrimidine dimerization*. **Signal Transduct Target Ther**. 2017;2:17021.

126. Mitchell DL. *Effects of cytosine methylation on pyrimidine dimer formation in DNA*. **Photochem Photobiol**. 2000;71(2):162-5.

127. Ziegler A, Leffell DJ, Kunala S, Sharma HW, Gailani M, Simon JA, et al. *Mutation hotspots due to sunlight in the p53 gene of nonmelanoma skin cancers*. **Proc Natl Acad Sci U S A**. 1993;90(9):4216-20.

128. You YH, Li C, Pfeifer GP. *Involvement of 5-methylcytosine in sunlight-induced mutagenesis*. **J Mol Biol**. 1999;293(3):493-503.

129. Ikehata H, Ono T. *Significance of CpG methylation for solar UV-induced mutagenesis and carcinogenesis in skin*. **Photochem Photobiol**. 2007;83(1):196-204.

130. Ikehata H, Masuda T, Sakata H, Ono T. *Analysis of mutation spectra in UVB-exposed mouse skin epidermis and dermis: frequent occurrence of C-->T transition at methylated CpG-associated dipyrimidine sites*. **Environ Mol Mutagen**. 2003;41(4):280-92.

131. Ikehata H, Kudo H, Masuda T, Ono T. *UVA induces C-->T transitions at methyl-CpG-associated dipyrimidine sites in mouse skin epidermis more frequently than UVB*. **Mutagenesis**. 2003;18(6):511-9.

132. Kielbassa C, Roza L, Epe B. *Wavelength dependence of oxidative DNA damage induced by UV and visible light*. **Carcinogenesis**. 1997;18(4):811-6.

133. Kino K, Sugiyama H. *UVR-induced G-C to C-G transversions from oxidative DNA damage*. **Mutat Res**. 2005;571(1-2):33-42.

134. Sekiguchi M, Tsuzuki T. *Oxidative nucleotide damage: consequences and prevention*. **Oncogene**. 2002;21(58):8895-904.

135. Grollman AP, Moriya M. *Mutagenesis by 8-oxoguanine: an enemy within*. **Trends Genet**. 1993;9(7):246-9.

136. Fisher MS, Kripke ML. *Systemic alteration induced in mice by ultraviolet light irradiation and its relationship to ultraviolet carcinogenesis*. **Proc Natl Acad Sci U S A**. 1977;74(4):1688-92.

137. Zheng ZS, Chen RZ, Prystowsky JH. *UVB radiation induces phosphorylation of the epidermal growth factor receptor, decreases EGF binding and blocks EGF induction of ornithine decarboxylase gene expression in SV40-transformed human keratinocytes*. **Exp Dermatol**. 1993;2(6):257-65.

138. Sachsenmaier C, Radler-Pohl A, Zinck R, Nordheim A, Herrlich P, Rahmsdorf HJ. *Involvement of growth factor receptors in the mammalian UVC response*. **Cell**. 1994;78(6):963-72.

139. Peus D, Vasa RA, Beyerle A, Meves A, Krautmacher C, Pittelkow MR. *UVB activates ERK1/2 and p38 signaling pathways via reactive oxygen species in cultured keratinocytes*. **J Invest Dermatol**. 1999;112(5):751-6.

140. Murphy GM, Norris PG, Young AR, Corbett MF, Hawk JL. *Low-dose ultraviolet-B irradiation depletes human epidermal Langerhans cells*. **Br J Dermatol**. 1993;129(6):674-7.

141. Uong A, Zon LI. *Melanocytes in development and cancer*. **J Cell Physiol**. 2010;222(1):38-41.

142. Tsao H, Chin L, Garraway LA, Fisher DE. *Melanoma: from mutations to medicine*. **Genes Dev**. 2012;26(11):1131-55.

143. Guy GP, Jr., Thomas CC, Thompson T, Watson M, Massetti GM, Richardson LC, et al. *Vital signs: melanoma incidence and mortality trends and projections - United States, 1982-2030*. **MMWR Morb Mortal Wkly Rep**. 2015;64(21):591-6.

144. Bernardes SS, Ferreira I, Elder DE, Nobre AB, Martinez-Said H, Adams DJ, et al. *More than just acral melanoma: the controversies of defining the disease*. **J Pathol Clin Res**. 2021.

145. Bradford PT, Goldstein AM, McMaster ML, Tucker MA. *Acral lentiginous melanoma: incidence and survival patterns in the United States, 1986-2005*. **Arch Dermatol**. 2009;145(4):427-34.

146. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. *Signatures of mutational processes in human cancer*. **Nature**. 2013;500(7463):415-21.

147. Herceg Z, Ghantous A, Wild CP, Sklias A, Casati L, Duthie SJ, et al. *Roadmap for investigating epigenome deregulation and environmental origins of cancer*. **Int J Cancer**. 2018;142(5):874-82.

148. Elder DE, Bastian BC, Cree IA, Massi D, Scolyer RA. *The 2018 World Health Organization Classification of Cutaneous, Mucosal, and Uveal Melanoma: Detailed Analysis of 9 Distinct Subtypes Defined by Their Evolutionary Pathway*. **Arch Pathol Lab Med**. 2020;144(4):500-22.

149. Kim SY, Kim SN, Hahn HJ, Lee YW, Choe YB, Ahn KJ. *Metaanalysis of BRAF mutations and clinicopathologic characteristics in primary melanoma*. **J Am Acad Dermatol**. 2015;72(6):1036-46.e2.

150. Lee JH, Choi JW, Kim YS. *Frequencies of BRAF and NRAS mutations are different in histological types and sites of origin of cutaneous melanoma: a meta-analysis*. **Br J Dermatol**. 2011;164(4):776-84.

151. Halaburkova A, Cahais V, Novoloaca A, Araujo M, Khoueiry R, Ghantous A, et al. *Pan-cancer multi-omics analysis and orthogonal experimental assessment of epigenetic driver genes*. **Genome Res**. 2020;30(10):1517-32.

152. Bertrand D, Drissler S, Chia BK, Koh JY, Li C, Suphavilai C, et al. *ConsensusDriver Improves upon Individual Algorithms for Predicting Driver Alterations in Different Cancer Types and Individual Patients*. **Cancer Res**. 2018;78(1):290-301.

153. Andor N, Graham TA, Jansen M, Xia LC, Aktipis CA, Petritsch C, et al. *Pan-cancer analysis of the extent and consequences of intratumor heterogeneity*. **Nat Med**. 2016;22(1):105-13.

154. Pradhan D, Jour G, Milton D, Vasudevaraja V, Tetzlaff MT, Nagarajan P, et al. *Aberrant DNA Methylation Predicts Melanoma-Specific Survival in Patients with Acral Melanoma*. **Cancers (Basel)**. 2019;11(12).

155. Ortmann B, Copeman J, Lehner PJ, Sadasivan B, Herberg JA, Grandea AG, et al. *A critical role for tapasin in the assembly and function of multimeric MHC class I-TAP complexes*. **Science**. 1997;277(5330):1306-9.

156. Seliger B, Schreiber K, Delp K, Meissner M, Hammers S, Reichert T, et al. *Downregulation of the constitutive tapasin expression in human tumor cells of distinct origin and its transcriptional upregulation by cytokines*. **Tissue Antigens**. 2001;57(1):39-45.

157. Lechner MG, Karimi SS, Barry-Holson K, Angell TE, Murphy KA, Church CH, et al. *Immunogenicity of murine solid tumor models as a defining feature of in vivo behavior and response to immunotherapy*. **J Immunother**. 2013;36(9):477-89.

158. Lee JH, Shklovskaya E, Lim SY, Carlino MS, Menzies AM, Stewart A, et al. *Transcriptional downregulation of MHC class I and melanoma de- differentiation in resistance to PD-1 inhibition*. **Nat Commun**. 2020;11(1):1897.

159. Chang CC, Pirozzi G, Wen SH, Chung IH, Chiu BL, Errico S, et al. *Multiple structural and epigenetic defects in the human leukocyte antigen class I antigen presentation pathway in a recurrent metastatic melanoma following immunotherapy*. **J Biol Chem**. 2015;290(44):26562-75.

160. Narayanan SP, Singh S, Shukla S. *A saga of cancer epigenetics: linking epigenetics to alternative splicing*. **Biochem J**. 2017;474(6):885-96.

161. Cho JW, Shim HS, Lee CY, Park SY, Hong MH, Lee I, et al. *The importance of enhancer methylation for epigenetic regulation of tumorigenesis in squamous lung cancer*. **Exp Mol Med**. 2022;54(1):12-22.

162. Anwar SL, Lehmann U. *DNA methylation, microRNAs, and their crosstalk as potential biomarkers in hepatocellular carcinoma*. **World J Gastroenterol**. 2014;20(24):7894-913.

163. Bowers EC, McCullough SD. *Linking the Epigenome with Exposure Effects and Susceptibility: The Epigenetic Seed and Soil Model*. **Toxicol Sci**. 2017;155(2):302-14.

164. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. *Comprehensive Characterization of Cancer Driver Genes and Mutations*. **Cell**. 2018;173(2):371-85 e18.

165. Conway JR, Dietlein F, Taylor-Weiner A, AlDubayan S, Vokes N, Keenan T, et al. *Integrated molecular drivers coordinate biological and clinical states in melanoma*. **Nat Genet**. 2020;52(12):1373-83.

166. Fatumo S, Chikowore T, Choudhury A, Ayub M, Martin AR, Kuchenbaecker K. *A roadmap to increase diversity in genomic studies*. **Nature Medicine**. 2022;28(2):243-50.

167. Alicea GM, Rebecca VW. *Un-Fair Skin: racial disparities in acral melanoma research*. **Nature Reviews Cancer**. 2022;22(3):127-8.

168. Vazquez VdL, Evangelista A, Silveira H, Crovador C, Carvalho AL, Scapulatempo-Neto C, et al. *Brazilian melanoma genome project: mutational landscape based on whole-genome sequencing*. In: **Cancer Research**; 2017. 77(13).

169. Duraes RO, Berardinelli GN, da Costa AM, Scapulatempo-Neto C, Pereira R, Oliveira MA, et al. *Role of Genetic Ancestry in 1,002 Brazilian Colorectal Cancer Patients From Barretos Cancer Hospital*. **Front Oncol**. 2020;10:145.

170. Leal LF, de Paula FE, De Marchi P, de Souza Viana L, Pinto GDJ, Carlos CD, et al. *Mutational profile of Brazilian lung adenocarcinoma unveils association of EGFR mutations with high Asian ancestry and independent prognostic role of KRAS mutations*. **Sci Rep**. 2019;9(1):3209.

171. Fernandes GC, Michelli RA, Galvao HC, Paula AE, Pereira R, Andrade CE, et al. *Prevalence of BRCA1/BRCA2 mutations in a Brazilian population sample at-risk for hereditary breast cancer and characterization of its genetic ancestry*. **Oncotarget**. 2016;7(49):80465-81.

172. Berardinelli GN, Duraes R, Mafra da Costa A, Bragagnoli A, Antonio de Oliveira M, Pereira R, et al. *Association of microsatellite instability (MSI) status with the 5-year outcome and genetic ancestry in a large Brazilian cohort of colorectal cancer*. **Eur J Hum Genet**. 2022.

173. Neuber AC, Tostes CH, Ribeiro AG, Marczynski GT, Komoto TT, Rogeri CD, et al. *The biobank of barretos cancer hospital: 14 years of experience in cancer research*. **Cell Tissue Bank**. 2021.

174. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. *Mutational heterogeneity in cancer and the search for new cancer-associated genes*. **Nature**. 2013;499(7457):214-8.

175. Wang K, Li M, Hakonarson H. *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data*. **Nucleic Acids Res**. 2010;38(16):e164.

176. Gehring JS, Fischer B, Lawrence M, Huber W. *SomaticSignatures: inferring mutational signatures from single-nucleotide variants*. **Bioinformatics**. 2015;31(22):3673-5.

177. Lee DD, Seung HS. *Learning the parts of objects by non-negative matrix factorization*. **Nature**. 1999;401(6755):788-91.

178. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. *Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays*. **Bioinformatics**. 2014;30(10):1363-9.

179. Fortin JP, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. *Functional normalization of 450k methylation array data improves replication in large cancer studies*. **Genome Biol**. 2014;15(12):503.

180. Kaushal A, Zhang H, Karmaus WJJ, Ray M, Torres MA, Smith AK, et al. *Comparison of different cell type correction methods for genome-scale epigenetics studies*. **BMC Bioinformatics**. 2017;18(1):216.

181. Perrier F, Novoloaca A, Ambatipudi S, Baglietto L, Ghantous A, Perduca V, et al. *Identifying and correcting epigenetics measurements for systematic sources of variation*. **Clin Epigenetics**. 2018;10:38.

182. Lin X, Barton S, Holbrook JD. *How to make DNA methylome wide association studies more powerful*. **Epigenomics**. 2016;8(8):1117-29.

183. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. *limma powers differential expression analyses for RNA-sequencing and microarray studies*. **Nucleic Acids Res**. 2015;43(7):e47.

184. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, R VL, et al. *De novo identification of differentially methylated regions in the human genome*. **Epigenetics Chromatin**. 2015;8:6.

185. Willer CJ, Li Y, Abecasis GR. *METAL: fast and efficient meta-analysis of genomewide association scans*. **Bioinformatics**. 2010;26(17):2190-1.

186. Matthew Suderman, James Staley, Robert French, Ryan Arathimos, Andrew Simpkin, Tilling K. *dmrff: identifying differentially methylated regions efficiently with power and control*. [Internet] 2022.

187. Perez-Enciso M, Tenenhaus M. *Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach*. **Hum Genet**. 2003;112(5-6):581-92.

188. Donoho D. *De-noising by soft-thresholding*. *IEEE Transactions on Information Theory*1995.

189. Love MI, Huber W, Anders S. *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2*. **Genome Biol**. 2014;15(12):550.

190. Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, et al. *Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data*. **Genome Med**. 2019;11(1):34.

191. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. *Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool*. **BMC Bioinformatics**. 2013;14:128.

192. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. *Enrichr: a comprehensive gene set enrichment analysis web server 2016 update*. **Nucleic Acids Res**. 2016;44(W1):W90-7.

193. Phipson B, Maksimovic J, Oshlack A. *missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform*. **Bioinformatics**. 2016;32(2):286-8.

194. Le Cao KA, Boitard S, Besse P. *Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems*. **BMC Bioinformatics**. 2011;12:253.

195. Tibshirani R. *Regression Shrinkage and Selection via the Lasso*. **Journal of the Royal Statistical Society**. 1996;58(1):267-88.

196. Singh A, Shannon CP, Gautier B, Rohart F, Vacher M, Tebbutt SJ, et al. *DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays*. **Bioinformatics**. 2019;35(17):3055-62.

197. Rohart F, Gautier B, Singh A, Le Cao KA. *mixOmics: An R package for 'omics feature selection and multiple data integration*. **PLoS Comput Biol**. 2017;13(11):e1005752.

198. Tarazona S, Balzano-Nogueira L, Gómez-Cabrero D, Schmidt A, Imhof A, Hankemeier T, et al. *Harmonization of quality metrics and power calculation in multi-omic studies*. **Nature Communications**. 2020;11(1):3092.

199. Terry Therneau PG. *Modeling Survival Data: Extending the Cox Model***: Springer, New York**; 2000.

200. Wickham H. *ggplot2: Elegant graphics for data analysis***: Springer-Verlag New York**; 2009.

201. *IARC pipeline for methylome analysis* [database on the Internet]. 2022. Available from: https://zenodo.org/record/6530343#.Ynuzki-tFTY.