Inês Martins Tomás
BSc in Biology

Oral Cancer: from genomic landscape to tumor immunobiology

This dissertation is submitted for the degree of
Master of Molecular Genetics and Biomedicine

Supervisor: Dr Inês Sequeira,
Centre for Stem Cells & Regenerative Medicine,
King's College London.

Second Supervisor: Prof Fiona Watt,
Centre for Stem Cells & Regenerative Medicine,
King's College London.

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

September, 2018

Inês Martins Tomás
BSc in Biology

Oral Cancer: from genomic landscape to tumor immunobiology

This dissertation is submitted for the degree of
Master of Molecular Genetics and Biomedicine

Supervisor: Dr Inês Sequeira,
Centre for Stem Cells & Regenerative Medicine,
King's College London

Second Supervisor: Prof Fiona Watt,
Centre for Stem Cells & Regenerative Medicine,
King's College London

September, 2018

Oral Cancer: from genomic landscape to tumor immunobiology

Setembro, 2018

Acknowledgments

In first place I would like to express my gratitude to my supervisor Dr Inês Sequeira for her endless support. Thanks to her I got the opportunity of having this amazing adventure in London, meeting fantastic people but most of all do great science and meet great scientists. Thank you for all the support and guidance both inside and out of the lab.

I'm very grateful to Prof Fiona Watt for accepting me in the Centre for Stem Cells & Regenerative Medicine, at King's College London. Firstly, as a summer intern and later as a master's student. I cannot express enough my gratitude to all the members and former members of the Centre and the Watt Lab. Thank you so much for welcoming me in such a warm way, for all the care and companionship. I would like to specially thank Ms Marina Torre that contributed directly to some of the work presented here as well as for the companionship and shared experiences outside the laboratory.

This dissertation would not be finished without the contribution of Dr David Adams and Mr Mamun Rashid. Their collaboration was crucial in sequence data analysis and incredibly stimulating to whom to learn from. On the same note, I thank Dr Alessandra Vigilante for sharing her knowledge on bioinformatics. I would like to thank Dr Joana Neves and Mr Tomasz Zabinski for their help on mastering some of the techniques used in my thesis. Flow cytometry and tissue dissociation was much simpler after their explanations and advice. As well as Dr Paola Bonfanti and her lab for collaborating with their resources and expertise in human immunology. I'm so grateful for the opportunity to learn from all these great scientists on diverse fields.

I would like to also acknowledge the team of the NIHR GSTT Biomedical Research Centre, Guy's Hospital London, as well as to all the staff of Biological Services Unit in the Hodgkin Building part of Guy's Campus, King's College London. Thank you for the training and endless support provided on flow cytometry and animal care, respectively. A special reference also to the Portuguese Association of Researchers and Students in the UK (PARSUK) for giving me the opportunity to start my international experience in science.

Finally, to my parents and grandparents. Thank you so much for believing in me more than I do myself. Your strength and support were my shelter and what kept me going. Without you, I could never dream as high as I do now.

Abstract

It is estimated that cancer will cause 9.6 million deaths and 18.1 million new patients diagnosed during 2018. Within this number, over 350 000 have oral tumors with tobacco and alcohol consumption identified as the biggest risk factors. Cancer research is still missing a comprehensive model that mimics human cancer as a whole. Therefore, the aim of this study was to analyse an experimental model that accurately mimics human cancer. To this end we treated mice with the carcinogen 4-nitroquinoline-1-oxide in the drinking water for over 16 weeks. This allowed us to induce differently graded tumors in mice tongue oral cavity. We performed whole exome sequencing of the tumors and the analysis confirmed similarities with human oral cancer genomic landscape. This study allowed us to gain new insight on the genomic progression of oral cancer and to explore an animal model that mimics not only the histological changes but also the genetic alterations observed in human oral cancer.

Previous work has shown that knockout mice for keratin 76 are more susceptible to develop oral cancer due to increased and over-suppressive regulatory T cells in the absence of keratin 76. However, the link between the loss of keratin 76 and these changes in the immune system remains unknown. Keratin 76 is progressively more expressed in mice thymus with aging and there is a parallel with the Hassall's corpuscles in human thymus. We showed that mice lacking keratin 76 present bigger thymic medullary regions and hypothesise one of the targets in the thymus to be Aire since its expression is reduced in the knockout. This study suggested an important role for keratin 76 in regulating the immune system.

Keywords

Head and neck squamous cell carcinoma (HNSCC)

Whole exome sequencing (WES)

4-nitroquinoline-1-oxide (4NQO)

Keratin 76 (Krt76)

Thymus

Regulatory T cells (Tregs)

Sumário

É estimado que 9.6 milhões de pessoas morram de cancro e 18.1 milhões de novos pacientes sejam diagnosticados só em 2018. Incluídos estão também mais de 350 000 pacientes com cancro oral cujo álcool e tabaco são indicados como fatores de risco. No entanto, continua em falta um modelo que mimetize o cancro humano de uma forma global. Desta forma, o objetivo deste estudo foi analisar um modelo experimental capaz de imitar o cancro em humanos. Para este fim, administrámos o carcinogénio 4-nitroquinolina-1-óxido na água de ratinhos durante 16 semanas. Consequentemente conseguimos induzir tumores de diferentes severidades na língua de ratinho com sucesso. Através da sequenciação total do exoma destes tumores concluímos que o padrão de mutações era semelhante ao característico de tumores humanos orais. Com este estudo compreendemos a progressão de cancro oral tumoral e não só explorámos um modelo experimental capaz de mimetizar as alterações histológicas, mas também as alterações genéticas envolvidas na tumorogénese.

Publicações anteriores comprovaram que ratinhos sem a expressão de keratina 76 são mais suscetíveis a cancro devido ao aumento de células T regulatórias, tal como mais supressivas. No entanto, a ligação entre a ausência de keratina 76 e as alterações no sistema imunitário continua desconhecida. Keratina 76 é progressivamente mais expressa no timo de ratinho com a idade, apresentando um paralelismo com os corpúculos de Hassall no timo humano. Demonstrámos que ratinhos sem keratina 76 apresentam maiores regiões medulares no timo. Adicionalmente, colocamos a hipótese de um dos alvos da keratina 76 ser o fator de transcrição Aire, uma vez que a sua expressão está reduzida em animais sem esta keratina. Este estudo sugere uma nova função de keratina 76 como reguladora do sistema imunitário.

Palavras-chave

Carcinoma de células esquamosas da cabeça e pescoço

Sequenciação total do exoma

4-nitroquinolina-1-óxido (4NQO)

Keratina 76

Timo

Células T regulatórias (Tregs)

Table of contents

List of figures

List of tables

Abbreviations

4HAQO	4-hydroxyaminoquinoline-1-oxide
4NQO	4-nitroquinoline-1-oxide
5-FU	5-fluorouracil
Aire	Autoimmune regulator
APECED	Autoimmune poly-endocrinopathy candidiasis ectodermal dystrophy
BSA	Bovine serum albumin
CCR7	C-C chemokine receptor 7
CD	Cluster of differentiation
cDNA	Complementary deoxyribonucleic acid
COSMIC	Catalogue of somatic mutations in cancer
cTEC	Cortical thymic epithelial cells
DAPI	4',6-diamidino-2-phenylindole
DC	Dendritic cells
DMBA	9,10-dimethyl-1, 2-benzathracene
DNA	Deoxyribonucleic acid
EDTA	Ethylenediamine tetra-acetic acid
EdU	5-ethynyl-2'-deoxyuridine
EGFR	Epithelial growth factor receptor
EPCAM	Epithelial cell adhesion molecule
FFPE	Formalin-fixed paraffin embedded
FMO	Fluorescence minus one
H&E	Haematoxylin and eosin
HC	Hassall's corpuscles
HNSCC	Head and neck squamous cell carcinoma
HPV	Human papillomavirus
IL	Interleukin
Inv	Involucrin
Krt	Keratin(s)
$Krt1^{-/-}$	Homozygous mice for the interruption of the keratin 1 gene.
$Krt17^{-/-}$	Homozygous mice for the interruption of the keratin 17 gene.
$Krt76^{-/-}$	Homozygous mice for the interruption of the keratin 76 gene.
$Krt76^{+/-}$	Heterozygous mice for the interruption of the keratin 76 gene.
Ly51	Glutamyl aminopeptidase
mTEC	Medullary thymic epithelial cells
NNK	4-methylnitrosamino-1-3-pyridyl-1-butanone
o.n	Overnight
OCT	Optimal cutting temperature compound
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
RT	Reverse transcription
SCC	Squamous cell carcinoma
SNV	Single-nucleotide variation
Tcells	Effector T cells
TEC	Thymic epithelial cells
TNM	Tumor-node-metastasis
Tregs	Regulatory T cells
TSLP	Thymic stromal lymphopoietin
UEA-1	Ulex europaeus agglutinin-1
WES	Whole exome sequencing
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Forefront

My journey in London started with a summer internship sponsored by PARSUK – Portuguese Association of Researchers and Students in UK. For one month, I joined the Centre for Stem Cells & Regenerative Medicine at King's College London to study the genomic landscape of oral squamous cell carcinoma. I then studied it as part of my master's thesis.

At the beginning of my stay in the Watt Lab, the samples involved in this project had just been sent for sequencing to our collaborators at the Wellcome Sanger Institute. This is a process that can be long and so I included another project in my thesis: tumor immunobiology, which I studied while waiting for the previous data.

In parallel with this first project, my supervisor had shown a new role of keratin 76 involving regulatory T cells maturation in cancer formation. Her work revealed that there is an increase of regulatory T cells related to an earlier appearance of oral cancer, in keratin 76 knockout mice. This discovery raised questions like why this population of T cells is increased in keratin 76 knockout mice, which lead to our research together.

Therefore, this dissertation includes two separate projects:

- The genomic landscape of oral cancer mouse model;
- The role of keratin76 in thymic organisation and immunoregulation.

Chapter 1 Introduction

1.1 Cancer

Accordingly to the World Health Organization, cancer is the leading cause of death in 48 countries, mostly socioeconomically developed (Bray *et al.*, 2018). It is estimated to cause 9.6 million deaths worldwide only in 2018 along with 18.1 million new cases (Bray *et al.*, 2018). Furthermore, the numbers continue raising due to population growth and aging adding to risk factors, despite advances in treatment and cancer research (Bray *et al.*, 2018). Therefore, it is necessary to understand the disease in order to develop a cure or improved treatments.

1.1.1 Hallmarks of Cancer

Tumorigenesis is a multistage process that includes several hallmark stages, which are characterized by the following tumorigenic states: dysplasia, carcinoma *in situ*, invasive carcinoma and intra and extravasation (Figure 1.1). Tumor cells are different from the others because they develop atypical characteristics, such as abnormal proliferation, evasion of growth suppressors, immortality, inducible angiogenesis and lastly invasion and metastasis (Hanahan and Weinberg, 2011).

Dysplastic cells characteristically harbour mutations in their DNA that normally leads to a biological advantage – oncogenes – followed by inactivating mutations in genes crucial to cell division – tumor suppressor genes, or *vice-versa*. Adding to these initial driver mutations, DNA instability results in cells harbouring progressively more mutations giving them the ability to proliferate and regulate different processes (Hanahan and Weinberg, 2011).

If/When a cell presents an abnormal behaviour and/or has too many growth-and-division cycles, the apoptotic pathway will be triggered leading to the destruction of the cell (Hanahan and Weinberg, 2011). In carcinoma *in situ*, cells have this pathway damaged, which will lead to a continuous uncontrollable division and growth. This incident is controlled by the cell cycle, during several check-points that prevent them from dividing abnormally (Figure 1.1). The aforementioned control is made at an intra-cellular level, however there are also control mechanisms at the extra-cellular level (Hanahan and Weinberg, 2011).

Cancerous cells at the stage of invasive carcinoma manipulate the system in which they reside in order to evade neighbouring and immune cells by expressing growth factors and specific ligands (Hanahan and Weinberg, 2011). This allows the cancerous cells to grow wildly and affect the correct function of the rest of the organs in the organism. These alterations demand a better supply of oxygen and nutrients combined with evacuation of the waste metabolites. To

accomplish this, an angiogenic switch is induced via signal proteins which induce the development of new vessels in the tissue surrounding the tumour cells (Hanahan and Weinberg, 2011).

This process is also crucial for the last step in tumorigenesis as the addition of vessels surrounding the tumour facilitates the transport of cancer cells to other regions and organs of the body – intravasation (Figure 1.1) (Hanahan and Weinberg, 2011). As these cancerous cells maintain biological advantages over the remaining non-cancerous cells, they are able to invade nearby tissues, be transported to distant ones – extravasation - and metastasise then establishing a new micro or macro tumor (Figure 1.1) (Hanahan and Weinberg, 2011). These alterations can occur in any organ of the body, but colorectal, female breast and lung cancer are the ones with bigger incidence (Torre *et al.*, 2015; Bray *et al.*, 2018; Reboux, 2018).

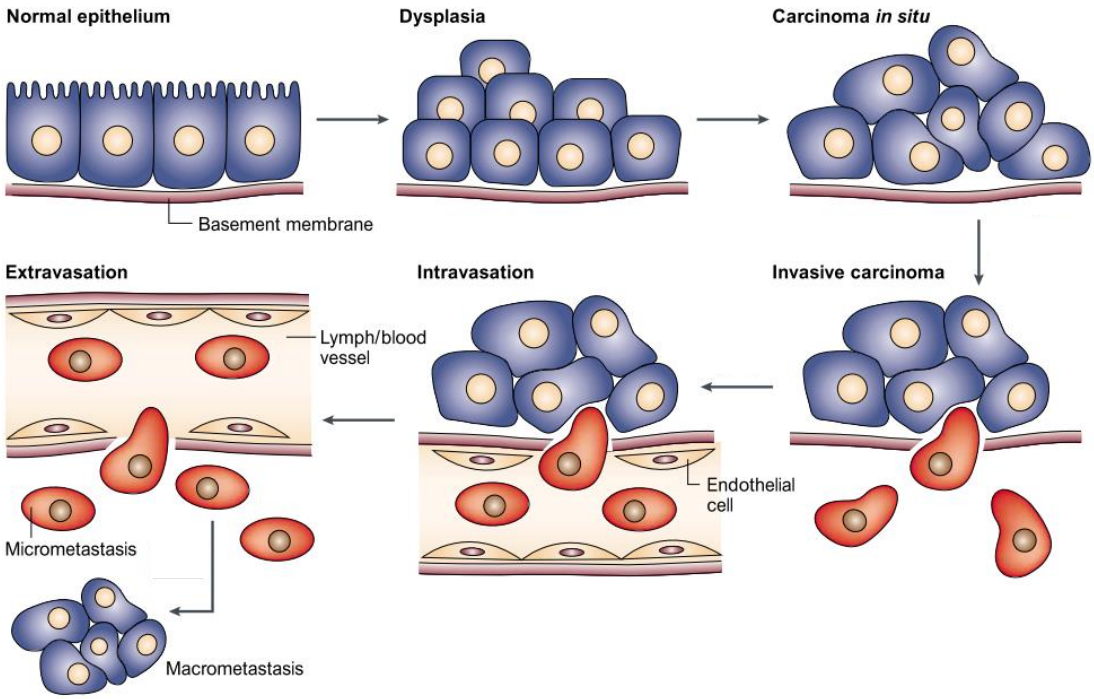

Figure 1.1 Stages of tumor development. In Thiery, 2002

1.1.2 Cancer diagnosis

Tumor development may be diagnosed based on the cellular alterations previously described, thus there are different classifications to be made on a patient's sample. The evaluation of the grade of each tumor is done with a biopsy tissue. The phenotype of the cells in the sample is then observed using microscopy techniques, which might include staining or not. The first stage, called hyperplasia, displays no alarming physiological changes, however the cells proliferate atypically which can lead to an accumulation of cell layers (Christopherson, 1977). At

this phase, the tumor may be classified as a grade 1 (Edge and Compton, 2010; National Cancer Institute, 2013). After several division cycles, and mostly due to molecular changes, the cells start to develop abnormal shapes and proportions. When a dysplasia starts to develop (Figure 1.1), normally the patient is still asymptomatic, however the lesions are clearly identifiable under the microscope. This stage is characterized by a disorganized growth and nuclear atypia which is clearly different from the remaining cells in respect to cellular features and maturation, being a grade 2 or 3 (Christopherson, 1977; Edge and Compton, 2010; National Cancer Institute, 2013).

Carcinoma *in situ* is the earliest stage of cancer (Figure 1.1). It is characterized by distinctive unusual figure, often associated with malignant proliferation but still restricted to the epithelium. If a patient is diagnosed at this stage it is very common to have a good prognosis. Unfortunately, diagnosis at this stage is not commonplace due to the lack of visible symptoms. Once these tumours evolve into an invasive carcinoma (Figure 1.1), the integrity of the tissue and the surrounding cells under the basement membrane becomes compromised (Christopherson, 1977). This makes treatment more difficult due to the lack of containment, it can develop metastasis in the lymph nodes or lungs due to vascularization. Tumours presenting a rapid growth with poorly differentiated or undifferentiated cells are classified as grade 3 or 4 respectively (Christopherson, 1977; Edge and Compton, 2010; National Cancer Institute, 2013).

Another clinical classification is the tumor-node-metastasis (TNM) denomination. T stands for primary tumor and it is normally followed by a number that indicates its size. When there are tumors in the lymph nodes, they are numbered according to the number of nodes affected. Metastasis in other regions of the body are labelled M1 with absence denoted as M0 (National Cancer Institute, 2015). This classification can be transposed to stage 0 while the carcinoma reaches an *in situ* state. Stage I to III represents the intermediate stages before metastasis, while stage IV is applied to metastatic cancer (National Cancer Institute, 2015).

There are genetic and behaviour factors that can increase the risk of cancer occurrence for example pollution and/or an unhealthy diet (Reboux, 2018). Nevertheless, different factors contribute differently to each cancer also affecting which therapy to be adopted.

1.2 **Head and Neck Squamous Cell Carcinoma (HNSCC)**

Cancer can be very heterogeneous and have different classifications depending on the organ it occurs in. Additionally, the type of cell affected can also vary. Thus, head and neck squamous cell carcinoma (HNSCC) is denominated when there is a tumor in the lips, oral and nasal cavity, paranasal sinuses, pharynx or larynx occurring mostly in epithelial cells, which account for 90% of head and neck cancer cases (Figure 1.2) (Tabatabaeifar *et al.*, 2014; WHO, 2014).

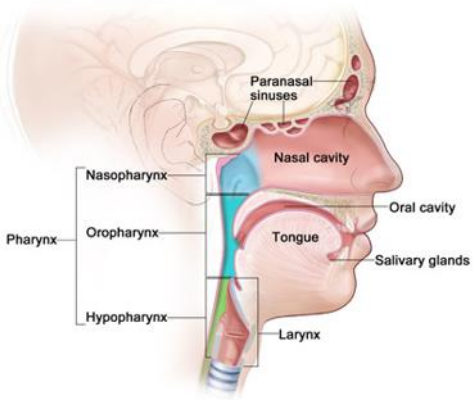

Figure 1.2 Tumor locations classified as head and neck cancer sites. In National Cancer Institute, 2017.

HNSCC is responsible for more than 300 000 patient deaths worldwide per year and is the sixth leading cancer in regards to incidence (Ragin and Taioli, 2007; Tabatabaeifar *et al.*, 2014; WHO, 2014; Morris *et al.*, 2017). In UK, 3% of cancer cases are head and neck cancers and this trend increased more than 30% since 1990 (Cancer Research UK, 2018) Unfortunately, the 50% 5-years survival rate of these patients has remained constant for the last decades with women having in general better prognosis (Ragin and Taioli, 2007; Morris *et al.*, 2017; Cancer Research UK, 2018). Given the location of these tumors, the risk factors alcohol consumption and tobacco abuse contribute largely and synergistically to the increase of the number of new cases of tumors greatly in the oral cavity (Kreimer *et al.*, 2005; Gaykalova *et al.*, 2014; Tabatabaeifar *et al.*, 2014; WHO, 2014; Cecco *et al.*, 2015; Torre *et al.*, 2015; Morris *et al.*, 2017).

Interestingly, HNSCC patients infected with the human papillomavirus (HPV) have a better prognostic and different clinical characteristics than those who are not (Ragin and Taioli, 2007). There is also differential occurrence within the genders as the chance of developing HNSCC can be up to 4x greater in men (WHO, 2014; Cancer Research UK, 2018). HNSCC can be classified according to tumor location as well as pattern of gene-expression.

1.2.1 Classification of HNSCC (Histological subtypes of HNSCC; grading of the tumors)

HNSCC is a heterogeneous disease not only in regard to prognosis and treatments but also at a molecular and genomic level. Therefore, it has been divided into subtypes which consider clinical, histological and genomic data. Originally 4 different subtypes were identified by Chung *et al.* based on gene expression: basal, mesenchymal, atypical and classical (Chung *et al.*, 2004; Walter *et al.*, 2013). These subtypes were later supported by Walter *et al.* (Walter *et al.*, 2013). However in 2015, 6 subtypes of HNSCC were identified using gene expression analysis of 1300 tumor samples (Cecco *et al.*, 2015). These subtypes were as follows: HPV-like, mesenchymal, hypoxia-associated, defence response, classical and immunoreactive and were divided mostly as a result of the pathway altered in each one. The classical subgroup was found to be directly correlated with heavy smoking, whilst the HPV-like subgroup was associated with the best clinical outcome (Cecco *et al.*, 2015). Most importantly, drug tests revealed differential susceptibility in each of the subgroups contributing to a possibly more personalized and adapted

treatment (Cecco *et al.*, 2015).

1.2.2 Mutation landscape in HNSCC

Head and neck tumors have been sequenced and widely analysed. Most of the studies analyse differently HPV-positive and HPV-negative, as well as tumors from smokers or non-smokers when the clinical data is available. Two comprehensive studies on the mutational landscape of HNSCC were published in 2011 (Agrawal *et al.*, 2011; Stransky *et al.*, 2011). Since then the number of known significantly mutated genes in HNSCC has been increasing (Tabatabaeifar *et al.*, 2014). Bigger number of samples analysed aligned with more evolved sequencing techniques give us more certainty determining genes mutated in smaller fractions of tumors. The different sequencing techniques commonly used for these studies are on review in chapter 1.4 Next Generation Sequencing.

Stransky *et al*. determined the mutational landscape of HNSCC by analysing 92 tumors (Stransky *et al.*, 2011). They concluded that HPV-negative tumors have a doubled mutation rate than HPV-positive. Also HPV-positive tumors never presented mutations in TP53, a commonly mutated oncogene in cancer (Stransky *et al.*, 2011; Gaykalova *et al.*, 2014; Tabatabaeifar *et al.*, 2014). Despite this observation, this gene was mutated in 62% of the total samples analysed (HPV-positive and negative), being the most mutated gene (Stransky *et al.*, 2011). Ranked by significance follows CDKN2A, mutated in 12% of the samples, CASP8, FAT1, NOTCH1, PTEN and SYNE1, each one mutated in 8%, 12%, 14%, 7% and 20%, respectively (Stransky *et al.*, 2011; Tabatabaeifar *et al.*, 2014). Still worth mentioning HRAS and PIK3CA which were newly identified in 5% and 8%, respectively (Stransky *et al.*, 2011).

Later that year, Agrawal, *et al*. analysed 32 HNSCC human samples (Agrawal *et al.*, 2011). The mutation rate regarding HPV infection stated before was confirmed (Agrawal *et al.*, 2011; Stransky *et al.*, 2011). They revealed that smokers' tumors have more mutations than non-smokers, contradicting Stransky. *et al*. (Agrawal *et al.*, 2011; Stransky *et al.*, 2011; Gaykalova *et al.*, 2014). Although smokers data was not associated to any signature (Agrawal *et al.*, 2011). TP53 mutations only appeared in HPV-negative tumors being present in 78% of the analysed tumors (Agrawal *et al.*, 2011; Stransky *et al.*, 2011; Gaykalova *et al.*, 2014). The top significantly mutated genes included TP53, NOTCH1, CDKN2A and PIK3CA with 47%, 15%, 9% and 6% of the patients having these respectively genes mutated (Gaykalova *et al.*, 2014; Tabatabaeifar *et al.*, 2014). Interestingly, in this analysis NOTCH1 was suggested to act as tumor suppressor, unlike other types of cancer (Tabatabaeifar *et al.*, 2014). Notice a new gene determined to be mutated in 5% of the patients - FBXW7 (Agrawal *et al.*, 2011).

The Cancer Genome Atlas Network combined bigger data sets by analysing 279 HNSCC

tumors (Kandoth *et al.*, 2013). Smoking was determined to be associated with TP53 loss-of-function mutations (Kandoth *et al.*, 2013). Top significantly mutated genes included CDKN2A, FAT1, TP53, AJUBA and CASP8, PIK3CA, NOTCH1 and KMT2D (Kandoth *et al.*, 2013). Interestingly, these first four genes were exclusively mutated in HPV-negative tumors being mutated in 22%, 23%, 72% and 6% of the total samples, respectively (Kandoth *et al.*, 2013). The presence of mutations in NOTCH2 and 3 were highlighted as present in 9% and 5% of the tumors, respectively (Kandoth *et al.*, 2013).

Regarding oral squamous cell lines, the mutational landscape is similar to the ones verified in human (Hayes *et al.*, 2016). It was concluded from the analysis of 16 oral SCC cell lines that genes like CDKN2A, FAT1, TP53, CASP8, AJUBA, PIK3CA, NOTCH1, MLL2, NSD1, HRAS, FBXW7, NFE2L2, CUL3 and PTEN are recurrently mutated (Hayes *et al.*, 2016). NOTCH1 was mutated in 44% of the samples (Hayes *et al.*, 2016). It was also determined an interaction between CASP8 and FAT1 leading to growth advantage in these cells (Hayes *et al.*, 2016). Studies specifically analysed 78 oral tumors (Vettore *et al.*, 2015). TP53 continued to be the most common mutation present in 38% of the samples (Vettore *et al.*, 2015). Vettore *et al*. identified other recurrent mutated genes as FAT1, 16%, ZFHX4, 15%, PLEC, 15% and SYNE1 with 13% of mutation occurrence (Vettore *et al.*, 2015). Tongue tumors had fewer mutations in TP53, NOTCH1 and CDKN2A when compared to HNSCC (Vettore *et al.*, 2015).

Curiously, smoking signatures have been discussed widely, varying with type of tobacco, tumor location and how is it induced (Sun and Califano, 2014; Tabatabaeifar *et al.*, 2014; Nassar *et al.*, 2015; Vettore *et al.*, 2015). Recent studies confirmed a differential mutational signature in smokers, as well as bigger mutational rate (Sun and Califano, 2014; Tabatabaeifar *et al.*, 2014; Morris *et al.*, 2017).

Besides the potential knowledge taken from the mutational landscape of tumor samples, it has been applied personalized medicine accordingly to which oncogenes and tumor suppressor genes are mutated (Morris *et al.*, 2017). This approach is essential for an effective therapy however while this application is not feasible, it continues to be important to adapt each treatment to each tumor (Morris *et al.*, 2017).

1.2.3 Treatment of HNSCC

Head and neck cancer treatment is a multi-disciplinary approach. Due to tumor location diversity included in this group, the therapy elected is considered taking several factors into account. Namely the stage, location and genetic background (Marur and Forastiere, 2008; Mehanna *et al.*, 2010; Leemans, Braakhuis and Brakenhoff, 2011).

Primary tumors in early stages are the first purposed for surgery. The removal of the tumor with good margins is crucial. Because head and neck cancer include regions involved in speaking, swallowing and chewing, a commitment between organ preservation and quality of life is necessary (Mehanna *et al.*, 2010; Galbiatti *et al.*, 2013).

Adjuvant chemotherapy is one of the most used therapies nowadays. Cancer research in HNSCC discovered that cisplatin and cetuximab are effective against the cancerous cells (Mehanna *et al.*, 2010). These two drugs are usually administrated together and as first-line treatment in recurrent tumors or metastasis. There are some patients also exposed to radiotherapy or alternative combinations with the chemotherapy agent 5-fluorouracil (5-FU) (Kandoth *et al.*, 2013b; Vettore *et al.*, 2015). In summary, head and neck treatment demands a multidisciplinary approach, including different therapies and adjusting them to the patients' needs and tumor specificities (Mehanna *et al.*, 2010).

1.3 Models of HNSCC research

In scientific research there is the need to create models that can mimic the symptoms of each disease and tissue conditions. Given that cancer, especially HNSCC, is a complex disease with a variety of factors contributing differently there is a need to complement the research done *in vitro* with an *in vivo* model. One of the most widely used animal models in research is the genetically engineered mice. The use of this model has enabled us to have a better understanding of gene function and has contributed greatly to diagnosis and treatments (Cheon and Orsulic, 2011). However, the mutated gene can cause alterations in other organs influencing tumour formation; therefore, other models using carcinogen drugs (such as 4-nitroquinoline-1-oxide [4NQO] and 9, 10-dimethyl-1, 2-benzathracene [DMBA]) have also been developed.

1.3.1 Animal model

Taxonomically, the human species belongs to the mammalian class. Within this group, the rodent order has many advantages which make it the preferred animal model. For example, reduced size, large litters, quick life cycles and the existence of genetic modification tools (Cheon and Orsulic, 2011).

To study tumor biology using a mouse line normally there are oncogenes activated or tumor-suppressors inactivated, or both, using gene-targeting techniques. Often these approaches result in knockins/outs or transgenic mice (for example *lox*P sites, reporter genes or transgene integration) which can also be complemented with a conditional and/or inducible expression (for examples Cre-*lox*P system, FLP-FRT system or tamoxifen-inducible system) (Cheon and Orsulic, 2011). Additionally, these mouse lines permit inter-crossing with other mouse lines which gives another tool to better understand gene function and regulation (Cheon and Orsulic, 2011).

Besides studying tumour development, there is also a need to study environmental factors, risk factors and potential therapies. In order to partly accomplish this, mice (genetically modified or not) should be exposed to different drugs and the resulting reactions and potential mechanisms triggered, should be investigated (Cheon and Orsulic, 2011). Comparative analysis between different species, perhaps mouse and human, is another method which can be used to aid in tumour understanding as it allows more accurate and relevant identification of genetic alterations important in tumorigenesis (Cheon and Orsulic, 2011).

1.3.2 4-nitroquinoline-1-oxide carcinogen (4NQO)

Carcinogens currently available for oral cancer research are 9,10-dimethyl-1, 2-benzathracene (DMBA) or 4-nitroquinoline-1-oxide (4NQO) (Figure 1.3). Out of the two, 4NQO has been the preferred choice for the last 50 years as it allows the study of early lesions, which DMBA does not (Kanojia and Vaidya, 2006; Vitale-Cross *et al.*, 2009). 4NQO models are the only ones that produce the sequential phases of tumorigenesis (complemented with similar histological and molecular changes as those in human HNSCC (Kanojia and Vaidya, 2006).

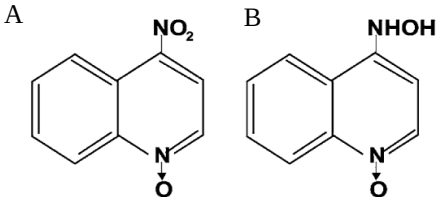

Figure 1.3 Structures of 4NQO. A) Chemical structure of 4NQO (4-nitroquinoline-1-oxide). B) Chemical structure of 4HAQO (4-hydroxyaminoquinoline-1-oxide). In from Kanojia and Vaidya, 2006.

The active compound of this carcinogen is actually its metabolic product – 4HAQO (4-hydroxyaminoquinoline-1-oxide) (Figure 1.3) (Kanojia and Vaidya, 2006). This transformation is done by consecutive reductions within the cell. It binds to DNA forming aducts which are prone to mismatches leading to mutations. Studies in different species had unravelled that the aducts occur more frequently at guanines, with adenines as the second most frequent nucleotide (Fronza *et al.*, 1992; Downes *et al.*, 2014). For example, in *E. coli* these aducts majorly happen at the positions C8, N2 or N6 of guanine nucleotide, with the last one proven to be the most mutagenic (Fronza *et al.*, 1992; Downes *et al.*, 2014). Due to this abnormal associations, the most frequent mutations are guanine to adenine transitions followed by guanine to pyrimidine (thymine and cytosine) – transversions (Kanojia and Vaidya, 2006).

Besides this direct interaction with the double-strand DNA, 4NQO induces intracellular oxidative stress by generating oxygen reactive species. The compound can enter the redox cycle contributing to the production of hydrogen peroxide and hydroxyl radicals putting the cells under a lot of stress (Kanojia and Vaidya, 2006; Osei-Sarfo *et al.*, 2013). Another consequence is an increase of the total intracellular β-catenin. This mechanism is included in the Wnt/β-catenin signalling pathway. Once dishevelled is sequestrated, it will regulate the destruction of β-catenin

downstream, which is responsible for the down-regulation of several transcription factors. Therefore, increased levels of this compound will translate into upregulation of genes involved in cell proliferation which once dysregulated can lead to tumorigenesis (Osei-Sarfo *et al.*, 2013). An additional feature of 4NQO treatment is that the stem cells existing in the tongue epithelium become reduced in number, whilst the existing and neighbouring clonal populations expand. Subsequently, if there is few clonal populations and those are prone to develop genetic mutations, it's highly probable that a cancer driver mutation will arise and then be present in most of the cells of that tissue, starting tumorigenesis (Osei-Sarfo *et al.*, 2013).

4NQO is used to mimic the use of tobacco in animal models (Osei-Sarfo *et al.*, 2013). Tobacco has over 60 carcinogens with different inorganic and organic compounds. Some of which have the same effects on DNA that 4NQO by creating instability on guanines, per example benzopyrene or NNK (4-methylnitrosamino-1-3-pyridyl-1-butanone) (Kanojia and Vaidya, 2006).

It has been proved that the usage of 4NQO mimics the alterations caused by HNSCC tumorigenesis, for example altered keratin expression in the epithelium, altered genetic expression and mutations, and epithelial growth factor receptor (EGFR) overexpression (Tang *et al.*, 2004, 2015; Ge *et al.*, 2016). Namely, it was determined that the changes in gene transcription were the same in tumors induced by 4NQO treatment and in HNSCC (Tang *et al.*, 2015). Moreover, the changes in genetic pattern both from down and upregulation are greater along tumor progression, meaning that hyperplasias have less genes being misregulated and invasive SCC greater number of altered genes (Ge *et al.*, 2016).

However, while the histological and morphological changes together with gene expression are similar between HNSCC and the tumors induced by 4NQO, there is no information about the genes mutated using this carcinogen. In this study we used next generation sequencing to determine the mutation landscape in mice tongue tumors induced by the administration of 4NQO in the drinking water.

1.4 Next generation sequencing

In the last decade we verified an exponential growth on sequencing technologies. Sanger sequencing, considered the first-generation, allowed the discovery of many mutations and their phenotype (Mardis, 2008). Despite its huge potential, this process was imposing big costs with small amounts of data. This problem was answered by next-generation sequencing techniques (Metzker, 2009).

There are several companies developing next-generation sequencing processes and each

of them found different technologies that differ in template preparation, sequencing, imaging and data analysis (Metzker, 2009). Regarding template preparation there are 2 options: DNA amplification, usually using a PCR method or single molecule templates. Sequencing and imaging can be done using single or multiple terminator nucleotides, removing a blocking group or several bases, measuring light, visualising several colours simultaneously or only one at the time (Metzker, 2009). Regarding data analysis, the alignment of the reads can be done comparing to a reference sequence or assembled *de novo*. In this topic we might consider time, effort, cost, coverage and biological application. The combination of the previous stages and options is what give rise to so many sequencing technics considering theirs advantages and disadvantages (Metzker, 2009).

The sequencing technology chosen for our study is commercialized by Illumina/Solexa. This technic uses solid-phase amplification – randomly distributed clusters that will amplify one DNA molecule leaving the end free for the ligation of the universal primer (Metzker, 2009). The sequencing and detection of the base is done recognizing 4 colours. Each nucleotide has a fluorescent dye with each correspondent colour (Metzker, 2009). The cycle continues after imaging, washing and cleavage. This process detects high-quality bases per run making it a good use for mutation detection in human genomes (Metzker, 2009).

Scientific knowledge is increasing greatly due to the amount of data produced by next-generation sequencing. These techniques gave access to whole genome, whole exome or targeted sequencing, answering to many questions raised in research (Mardis and Wilson, 2009). Along with bioinformatics progressing, cancer genomics analysed endless genes, tumors, samples, clarified mechanisms, pathways and its influence in therapy (Mardis and Wilson, 2009).

The comparison of tumor mutations with normal matched sequences has determined many driver mutations originating new treatment options and therapies. Continuing to gather information on this causative effect will take us a step closer to unravel the tumorigenic process completely (Mardis and Wilson, 2009). We used next generation sequencing to analyse the exome of 69 tumors. The mutational landscape determined from this data was filtered for inter-individual variability. Because we sequenced matched control tissue from all the animals we excluded somatic mutations. The aim of our study is to determine driver genes and mutations in mouse oral tumors. Comparison of whole exome sequencing data from our model and HNSCC patients will confirm similar mutational landscapes.

1.5 The role of the immune system in cancer development

One characteristic of tumorigenic tissue is an increase of immune cells (Hanahan and Weinberg, 2011). Accordingly, tumors have an increased number of cells from both innate and

adaptative immune response (Hanahan and Weinberg, 2011). This immune response is an attempt to eliminate the cancerous cells, however there are evidences on pro-inflammatory responses enhancing tumor progression (Hanahan and Weinberg, 2011). Indeed the immune system can be evaded and induced to produce bioactive molecules like growth and survival factors thus stimulating survival and facilitating angiogenesis (Hanahan and Weinberg, 2011).

Clearly the immune system has an important role in tumor progression and cancer response (Hanahan and Weinberg, 2011; Whiteside, 2018). However, it can be activated by several stimuli and regulated by many proteins. One example is keratin 1 (krt1), which is known to be crucial to skin barrier preservation (Roth *et al.*, 2012). It was proved that krt1 interacts closely with the inflammatory machinery. In mice genetically modified to be missing this keratin, there is an increase of interleukin 18 (IL-18), known to be associated with atopic eczema. Pharmacological depletion of this interleukin rescues partially the defective barrier phenotype from $krt1^{-/-}$. Thus krt1 was determined to be mediating inflammasome activity in skin as well as epidermal barrier formation (Roth *et al.*, 2012).

Likewise, it was demonstrated that krt17 has an immunomodulatory role (Depianto *et al.*, 2010). In the absence of krt17, it was detected a reduction of the inflammatory response and hyperplasia is retarded in ear skin. Confirming the hypothesis, rescue of this keratin induced an increase of chemokines which contributed to the tumorigenic process (Depianto *et al.*, 2010).

Another study has determined that krt76 also has an important immunomodulatory role (Sequeira *et al.*, 2018). $Krt76^{-/-}$ mice demonstrated to have an increased proinflammatory response detected in the form of splenomegaly, lymph node enlargement, increased number and more suppressive regulatory T cells (Tregs) in the thymus, and finally elevated cytokines production (Sequeira *et al.*, 2018). However, the thymus cellularity and size did not change (Sequeira *et al.*, 2018). Moreover, even though krt76 ablation is not sufficient to induce tumors, these animals are more susceptible to oral tumor development when exposed to 4NQO carcinogen due to increased action of the immune system (Sequeira *et al.*, 2018). Therefore krt76 mediates local and systemic immune response demonstrated to be important to tumor suppression (Sequeira *et al.*, 2018).

Gathering the previous statements, in this project we aim to better understand the role of krt76 in the regulation of the immune system. Taking together previous studies, this new function of krt76 can reveal new insights on tumor progression and susceptibility (Sequeira *et al.*, 2018).

1.6 **Thymus**

The thymus is one of the organs of the immune system (Sakaguchi *et al.*, 2008). It is responsible for the production of the immune cells but is also a reservoir of T cells (Sakaguchi *et*

al., 2008). Thymus architecture is divided into cortex (outer region) and medulla (inner region) (Figure 1.4) (Xing and Hogquist, 2014; Lopes *et al.*, 2015). Within the thymic cells we can name epithelial cells (TEC), T cells (Tcells) and dendritic cells (DC) (Figure 1.4) (Xing and Hogquist, 2014). In this first group they are distinguished by their location – medullary thymic epithelial cells (mTEC) or cortical thymic epithelial cells (cTEC).

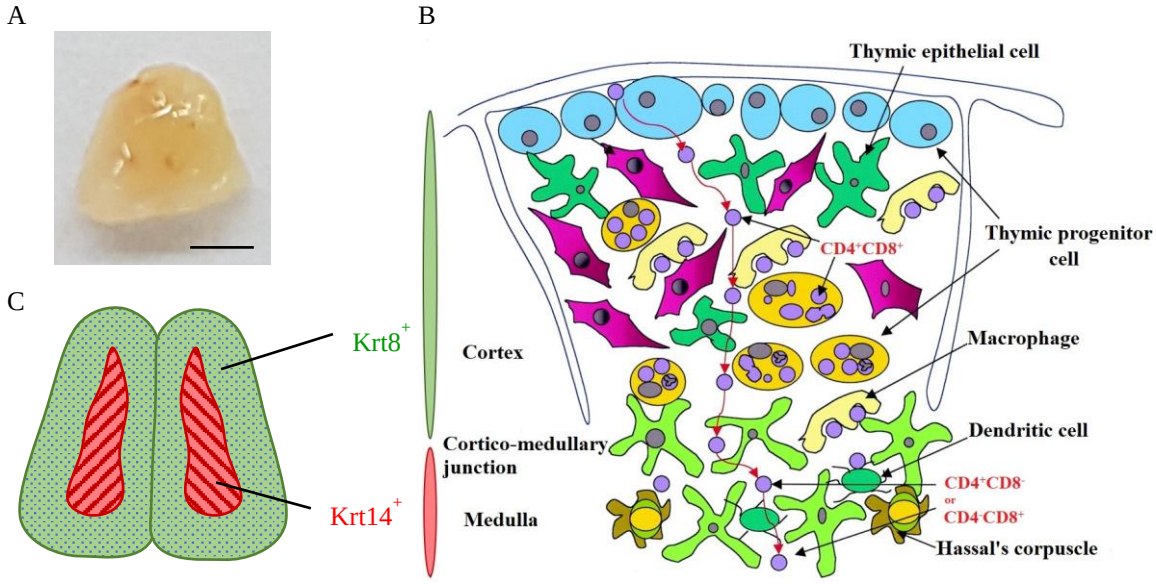

Figure 1.4 Thymic arquitecture and organization. A) Dissected adult mice thymus. Scale bar 200µm. B) Schematic of thymus organization along Tcells maturation. Edited from Pezzano *et al.*, 2001. C) Schematic of thymus arquictecture.

1.6.1 T cells

T (thymic) cells, referring to where they are formed and matured, can be classified accordingly to their different functions. The most commons are the T effectors which are the first ones to be activated over stimulus. Other types of T cells are the T helpers, natural killers and T regulatory (Tregs), among others (Caramalho *et al.*, 2015). This collection of immune cells gives us a wide range of protection against various threats, although sometimes it also means over-reactiveness as in allergies, or in worst cases, autoimmune diseases (Sakaguchi *et al.*, 2008).

The progenitor Tcells originate in the bone marrow and migrate into the thymus (Xing and Hogquist, 2014; Caramalho *et al.*, 2015). At this point, they are double negative cells, specifically $CD4^{-}CD8^{-}$ (Xing and Hogquist, 2014; Abramson and Anderson, 2017; Hu *et al.*, 2017). Once in the cortex, they are exposed to several stimuli by the epithelial cells and the dendritic cells (DC) inducing the expression of different genes (Xing and Hogquist, 2014; Abramson and Anderson, 2017; Hu *et al.*, 2017). As result they differentiate into $CD4^{+}CD8^{+}$ thymocytes, known as double positive (Figure 1.4) (Xing and Hogquist, 2014; Caramalho *et al.*, 2015; Lopes *et al.*, 2015; Abramson and Anderson, 2017; Hu *et al.*, 2017). Upon continued

interaction with the surrounding cells still in the cortex, these thymocytes then differentiate into single positive cells as $CD4^+$ or $CD8^+$ (Figure 1.4) (Xing and Hogquist, 2014; Caramalho *et al.*, 2015; Lopes *et al.*, 2015; Abramson and Anderson, 2017). The Tcells have a double step maturation process, namely positive and negative selection (Xing and Hogquist, 2014; Abramson and Anderson, 2017).

Firstly they are exposed to nonself-antigens by the cortical epithelial cells and rejected if not reactive (positive selection) (Xing and Hogquist, 2014; Caramalho *et al.*, 2015; Lopes *et al.*, 2015; Abramson and Anderson, 2017). The maturation of thymic cells is done from the periphery of the organ into the core so after this phase, the Tcells migrate into the medulla (Figure 1.4) (Xing and Hogquist, 2014; Lopes *et al.*, 2015; Abramson and Anderson, 2017; Hu *et al.*, 2017). The receptor CCR7 assures that the thymocytes are retained in the medullary region thus guaranteeing complete maturation of the Tcells (Hu *et al.*, 2017). Here, they are exposed by mTEC, and partially DC, to self-antigens and are negatively selected being signalled for apoptosis if reactive (Xing and Hogquist, 2014; Lopes *et al.*, 2015; Abramson and Anderson, 2017; Hu *et al.*, 2017). Tcells progenitors give origin to several lineages of cells (Xing and Hogquist, 2014; Caramalho *et al.*, 2015). Namely, the maturation and differentiation process of Tcells can be diverted into other paths and so give origin to Tregs, for example.

1.6.2 Regulatory T cells (Tregs)

This subgroup of thymic cells is responsible for the balance between immune cells (Caramalho *et al.*, 2015; Lopes *et al.*, 2015; Sequeira *et al.*, 2018). Tregs can manipulate the activity of $CD4^+$, $CD8^+$ cells, as well as B cells, natural killer cells, macrophages and DC subsequently able to recognise self and nonself-antigen responses, thus guaranteeing homeostasis (Sakaguchi *et al.*, 2008). Tregs need to be exposed to hundreds of antigens to guarantee its responsiveness against Tcells activity in all the organs (Lio and Hsieh, 2011; Abramson and Anderson, 2017).

Tregs differentiation begins from dysfunctional Tcells. This means that the over reactive Tcells reaching negative selection might not be signalled to apoptosis and instead go through Tregs maturation (Aschenbrenner *et al.*, 2007; Sakaguchi *et al.*, 2008; Caramalho *et al.*, 2015; Lopes *et al.*, 2015; Abramson and Anderson, 2017). Mostly from the poll of $CD4^+$ cells they are separated as $CD25^+FOXP3^-$ or $CD25^-FOXP3^+$ being thus the Tregs progenitor cells (Aschenbrenner *et al.*, 2007; Lio and Hsieh, 2011; Abramson and Anderson, 2017). This part of the process is known to be mainly done by the mTEC, but also influenced in many aspects by macrophages as well as IL-2 and IL-5 expression even though there are still some pathways to be clarify (Aschenbrenner *et al.*, 2007; Lio and Hsieh, 2011; Caramalho *et al.*, 2015; Abramson and Anderson, 2017).

Tregs are specifically identified by their unique and crucial expression of the transcription factor FOXP3 (Sakaguchi *et al.*, 2008; Caramalho *et al.*, 2015; Lopes *et al.*, 2015). It is able to upregulate CD25, as well as IL-2, among other factors and markers, which consequently leads to a $FOXP3^{+}CD25^{+}$ population of Tregs when mature and functional (Nomura and Sakaguchi, 2007; Sakaguchi *et al.*, 2008). Mice lacking FOXP3 present a severe phenotype of autoimmune disease by the absence or abnormal Tregs, proving its crucial involvement in Tregs maturation, with the same applying to CD25 (Sakaguchi *et al.*, 2008; Caramalho *et al.*, 2015). When mature they circulate into peripheral organs, however they can re-enter the thymus. This mechanism is a feedback controlled by the DC that recognized IL-2 thus inducing proliferation if needed, proving its role on Tregs maturation (Aschenbrenner *et al.*, 2007; Caramalho *et al.*, 2015; Malchow *et al.*, 2016; Abramson and Anderson, 2017; Hu *et al.*, 2017).

Tregs are capable of supressing the activity of Tcells (Caramalho *et al.*, 2015; Lopes *et al.*, 2015; Sequeira *et al.*, 2018). This interaction guarantee that Tcells are inhibited and eliminated if over responsive to self-antigens (Caramalho *et al.*, 2015; Lopes *et al.*, 2015). When Tregs are not functional, Tcells will responde inadequatly leading to a autoimmune disease (Caramalho *et al.*, 2015; Lopes *et al.*, 2015). The same principle applies to anti-tumor response (Hanahan and Weinberg, 2011; Whiteside, 2018). Likewise the Tcells responsible for anti-tumor response can be wrongly detected as over-responsive and eliminated by Tregs leading to tumorigenesis stimulus (Hanahan and Weinberg, 2011; Whiteside, 2018).

Curiously, it has been determined that $krt76^{-/-}$ have increased Treg population with over suppressive capacity (Sequeira *et al.*, 2018). This discovery was supported by local and systemic evidences of inflammation along with increased expression of chemokines and cytokines (Sequeira *et al.*, 2018). Due to this increase in Tregs, these mice were more susceptible to tumor development once exposed to a carcinogen thus proving the importance of Tregs in the immune system but also in tumor response (Sequeira *et al.*, 2018). While $FOXP3^{+}$ Tregs were more abundant in tumors, T effectors were reduced, suggesting that Tregs are evaded to inhibit the normal anti-tumor immune response (Sakaguchi *et al.*, 2008; Whiteside, 2018).

1.6.3 Thymic epithelial cells (TEC)

TEC can be generally identified by flow cytometry selecting to $CD45^{-}EPCAM^{+}$ since this excludes Tcells ($CD45^{+}$) and selects for epithelial thymic cells ($EPCAM^{+}$) (Xing and Hogquist, 2014; Lopes *et al.*, 2015) Because they are a rare population this staining combination is common for enrichment and purification from Tcells (Xing and Hogquist, 2014). Additionally, the separation of cTEC from mTEC can be done by staining with UEA-1 and Ly-51 respectively, because although they originate from the same progenitor cell the expressed markers are different (Xing and Hogquist, 2014; Lopes *et al.*, 2015; Abramson and Anderson, 2017)

Thymic epithelial cells are responsible for the antigen presentation necessary in Tcells maturation and also Tregs, however this process is more intense in the latter (Nomura and Sakaguchi, 2007; Xing and Hogquist, 2014). Both the cortical and medullary TEC are essential to these processes, although each area has different specific functions and therefore is populated by cells with different markers (Aschenbrenner *et al.*, 2007; Xing and Hogquist, 2014; Malchow *et al.*, 2016; Abramson and Anderson, 2017). Interestingly, mice with reduced mTEC have significantly reduced number of Tregs, and *vice versa* (Liu *et al.*, 2013; Lopes *et al.*, 2015).

Genetic regulation is controlled by transcription factors like Aire, one of most expressed in the mTEC (Nomura and Sakaguchi, 2007; Lopes *et al.*, 2015; Malchow *et al.*, 2016; Abramson and Anderson, 2017). It is a marker of TEC maturation hence it is expressed in later stages of differentiation (Lopes *et al.*, 2015). More importantly, its regulation is largely done by histone methylation which then leads to the activation of many genes – 3 000 to 4 000 (Malchow *et al.*, 2016; Abramson and Anderson, 2017). Aire is closely involved in FOXP3 expression so consequently regulates the production of Tregs also through the expression of other antigens in mTEC (Nomura and Sakaguchi, 2007; Malchow *et al.*, 2016; Hu *et al.*, 2017). Another transcription factor present in the mTEC is Fezf2 that can regulate antigen expression in an Aire independent way (Takaba *et al.*, 2015; Malchow *et al.*, 2016)

Autoimmune poly-endocrinopathy candidiasis ectodermal dystrophy (APECED) is characterized by the loss-of-function of Aire. This disease presents autoimmune destruction in several organs associated with a reduction of Tregs (Caramalho *et al.*, 2015; Malchow *et al.*, 2016). This scenario is similar in knockout mice for Aire thus supporting its role in Treg maturation since the cells are not exposed to the necessary stimuli (Caramalho *et al.*, 2015; Malchow *et al.*, 2016). There is further evidences indicating that Aire is also involved in negative selection of Tcells thus doing the selection into Tregs lineage (Nomura and Sakaguchi, 2007; Lopes *et al.*, 2015; Malchow *et al.*, 2016; Hu *et al.*, 2017).

Aire expression can be lost in later stages of mTEC differentiation and differentiate into structures characterized by involucrin expression - Hassall's Corpuscle (HC) (Figure 5.4 D) (Lopes *et al.*, 2015). Interestingly, these cells induce the production of TSLP in the medulla, which will then lead to a stimulus on Tregs differentiation (Caramalho *et al.*, 2015). It was observed *in vitro* that TSLP is able to induce dendritic cells to transform thymocytes into Tregs (Nomura and Sakaguchi, 2007).

1.7 **Keratins**

Keratins (Krt) are epithelial-specific intermediate filaments of 10nm coded in 54 human genes (Moll, Divo and Langbein, 2008; Herrmann and Aebi, 2016). Their main function is

assembling a network of filaments responsible for mechanical and structural stability. Keratin filaments are composed by heterodimers: one type I protein together with one type II protein (Moll, Divo and Langbein, 2008; Herrmann and Aebi, 2016). They can be classified as "soft", when in living cells, or as "hard", when in death cells like hair or nails (Moll, Divo and Langbein, 2008; Herrmann and Aebi, 2016). Given their diversity, each keratin have its specific expression pattern within cell type, differentiation and functional stage (Moll, Divo and Langbein, 2008).

There are several clinical pathologies associated with keratin mutations. Because the relation genotype-phenotype is not always straight forward there has been new hypothesis formulated on keratin having influence within intracellular pathways (Moll, Divo and Langbein, 2008; Coulombe, 2017). Due to keratins' epithelial expression, they are a key component on the skin, being the first defensive barrier against the exterior (Hobbs, Lessard and Coulombe, 2012). It is understandable that if a threat or lesion is detected, keratins might have a potential role in regulating immune response (Hobbs, Lessard and Coulombe, 2012). Hence tumor development is intimately related with the immune system, it is crucial to understand keratins' involvement (Hobbs, Lessard and Coulombe, 2012). Keratins are structural proteins however new studies have been revealing other functions. In summary, there are theories on keratin interaction with cell cycle machinery, apoptosis, wound healing, inflammation and tumor progression (Moll, Divo and Langbein, 2008; Hobbs *et al.*, 2015; Herrmann and Aebi, 2016).

Specifically, krt17 expression is correlated with more aggressive tumors and poor prognosis for some types of cancer (Hobbs *et al.*, 2015). In krt17$^{-/-}$ mouse skin, tumorigenesis is slower, as well as there is a reduction on inflammatory molecules present (Hobbs *et al.*, 2015). It was concluded that the transcription factor Aire and krt17 are in close and dependent interaction influencing immune response in skin tumors. Nonetheless, along with these discoveries it was observed for the first time the presence of a keratin in the nucleus (Hobbs *et al.*, 2015). This new innovative line of thinking has been analysed being considered a possible hypothesis given the diversity and variability of this family (Hobbs, Jacob and Coulombe, 2016).

Taking advantages of keratins' specificity and adding to their wide-ranging epithelial expression, they have been used in cell and tumor typing once their expression may be retain during tumorigenesis – mainly in epithelial tumors – and also prognostic markers (Moll, Divo and Langbein, 2008).

1.7.1 Keratin 76

Keratin 76 (krt76) is a type II protein expressed specifically in the suprabasal cell layers of oral epithelium, specifically the hard palate and orthokeratinized stratified squamous epithelium in the gingiva (Moll, Divo and Langbein, 2008; Ambatipudi *et al.*, 2013; Sequeira *et*

al., 2018). Krt76$^{-/-}$ mice presented flaky tail, darker footpads pigmentation as well as epidermal thickness and basal layer hyperproliferation confirmed in the tail (Liakath-Ali *et al.*, 2014).

Krt76 was shown to be gradually downregulated along tumor progression in human but not related to a smaller survival rate (Ambatipudi *et al.*, 2013). Although the loss of this protein lead to preneoplastic changes like leukoplakia, similarly in mouse and human, this was not enough to generate mouse spontaneous tumors (Ambatipudi *et al.*, 2013; Sequeira *et al.*, 2018).

Even though there is no disease identified to be caused/related by/to krt76 mutations, there are indirect interactions leading to known phenotypic. Precisely, loss of krt76 in mice skin leads to abnormal claudin 1 positioning. This protein is a critical component on tight junction formation, which are essential proteins for the epidermal barrier once they are selective seals between cells and layers of cells (DiTommaso *et al.*, 2014). Thus it is hypothesized that the skin barrier compromised phenotype of ktr76$^{-/-}$ is due to abnormal claudin 1 positioning which is verified in psoriasis and some cancers (DiTommaso *et al.*, 2014). Krt76 was determined to be also involved in wound healing processes (DiTommaso *et al.*, 2014).

1.8 **Aims of this study**

In this study we try to understand tumor progression and the biological changes associated. Using next generation sequencing, specifically whole exome sequencing, we determine the mutation landscape of mice tongue tumors induced by the carcinogen 4NQO. This study allows the comparison of this experimental model with the genetic mutations occurring in HNSCC thus leading to prospective model for cancer research and drug testing.

We also aimed to explore the role of krt76 in the immune system regulation. By studying the immune alterations in krt76$^{-/-}$ mice we pretend to better comprehend why these animals are more susceptible to tumor development due to an increased inflammatory response. By understanding the interaction between krt76 and thymic alterations we can deduce its involvement in tumor response leading to a new biomarker.

Chapter 2 Materials and Methods

2.1 **4NQO oral cavity lesions**

This experiment was done with adult C57 bl/6 *wild type* mice under the approval of the UK Home Office for the procedure project licence 70/8474. The mice were maintained with regular mouse chow and water *ad libitum*. The experimental group was treated with of 4NQO (Sigma) diluted at 100µg/mL in the drinking water, being this their only source of water (Sequeira *et al.*, 2018). The carcinogen treatment lasted for 16 weeks, and 4NQO solution was renewed once a week (Figure 2.1). This was the only difference in the conditions held between mice groups. Animals were harvested every 2 weeks, starting at week 4 after treatment until week 28, which is 12 weeks after treatment has stopped (Figure 2.1). Treated mice were screened for lesions in the oral cavity, once a week, while sedated by isoflurane (Sequeira *et al.*, 2018). One hour before culling, the mice were injected intraperitoneally with 500µg EdU (5-ethynyl-2'-deoxyuridine) diluted in phosphate-buffered saline (PBS), to access cell proliferation, according to the Click-It Edu Imaging Kit (Invitrogen) protocol. From 72 animals, we harvested 69 tumors. From every animal was harvested ear tissue (healthy) and the tongue. The first tissue was sent to the Wellcome Trust Sanger Institute to have its whole exome sequenced (WES). The tongue was preserved as formalin-fixed paraffin-embedded (FFPE) blocks which allowed the micro-dissection of tumorigenic tissue that was sent to be sequenced.

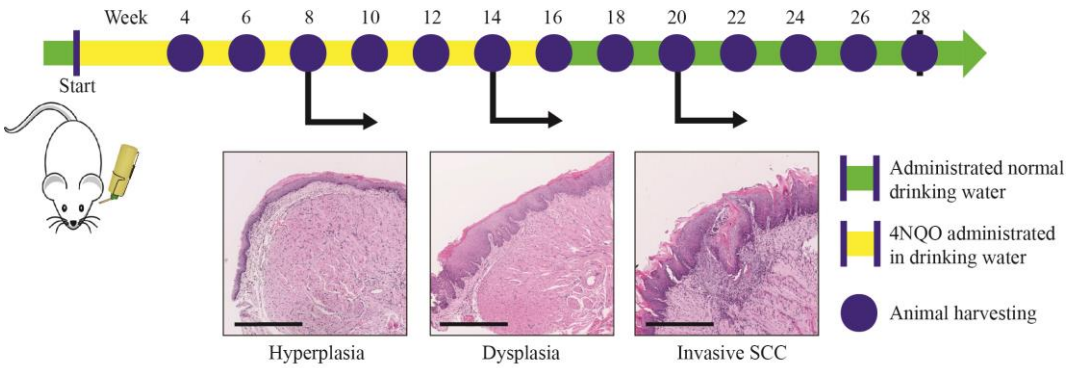

Figure 2.1 Experimental design of 4NQO treatment administrated in the drinking water. Treatment lasted for 16 weeks and oral cavity was screened for lesions during 28 weeks. H&E representative sections indicate the grading of the lesions expected in each week (Figure 1.1). Hyperplasias were occuring at week 8, while dysplasias were developing at week 14. Invasive SCC were identified after week 20. Scale bar 500µm.

2.2 **Whole exome sequencing and analysis**

To determine single-nucleotide variations we performed whole exome sequencing using Illumina/Solexa platform. Germline mutations were filtered with matched control tissue in all samples by SNV maximization-based mutation calling applying CaVEMan package.

2.3 **Keratin 76 knockout mice**

The animals used were approved by the UK Home Office for the procedure project licence 70/8474. The strain Krt76$^{tm1a(KOMP)Wtsi}$ were ordered to the Wellcome Trust Sanger Institute Mouse Genetics Programme, being part of the International Mouse Phenotype Consortium (Skarnes *et al.*, 2011). These krt76$^{+/-}$ mice were produced interrupting the krt76 gene as indicated in the Figure 2.2 (Cheon and Orsulic, 2011). Note that there is a LacZ reporter gene added between the first and second exons which allows tracking krt76 expression while having the protein disrupted. Krt76$^{-/-}$mice were obtained by mating heterozygous mice, which give origin to *wild type* and heterozygous animals used as littermate controls. Genotype was confirmed by ear punch biopsies in all the animals.

Figure 2.2 Schematic of the krt76 disrupted gene. LacZ insert is between exon 1 and 2 of krt76 gene. This construct leads to the transcription of a truncated protein due to the size of the insert. The FRT and loxP sites allow the rescue of the gene through the use of Flp recombinase and Cre, respectively (Sequeira *et al.*, 2018).

2.4 **Histology**

The tissues harvested from mice for histology purposes were collected into cold PBS. For cryosections, they were then fixed in paraformaldehyde (PFA) 4% for 1h at room temperature with agitation. After one PBS wash, the tissue stayed overnight (o.n) in sucrose 15% at 4°C. The tissues were directly embedded in optimal cutting temperature compound (OCT, VWR chemicals) and sectioned into Menzel-gläser Superfrost® Plus (Thermo Scientific) using a Cryostat (CryoStar NX70) with the platform at -15°C and cutting blade at -10°C. Sections of 7µm thickness were made from a ventral to dorsal perspective with longitudinal cuts. Sections were stored at -20°C. For paraffin sections, the tissues were processed to paraffin embedding using HistoStar (Thermo Scientific). Paraffin sections were sectioned at 5µm thickness into Menzel-gläser Polysine® Slides (Thermo Scientific) using a microtome (Thermo Scientific Microm HM 355S), which were posteriorly heated in an oven at 60°C for a minimum of 1hour.

The haematoxylin and eosin staining (H&E) was done following the standard methods. Deparaffinization was done by exposing the slides to xylene twice, alcohol 100% twice, alcohol 70% and dH_2O, for 2 minutes each solution. Thymus as frozen sections were let dry at room temperature for about 5 minutes and fixed in formalin 10% for 5 more minutes. Remaining steps are the same as for paraffin or frozen sections. It was used Haematoxylin Gill No.3 for 1 minute, followed by dH_2O for 2 minutes and Acid Alcohol 1% (acidic acid diluted in alcohol 70%) for 5 seconds followed by more than 3 minutes wash. Eosin 1% was applied for 1 minute, washed trice

and exposed to increasing percentages of alcohol with 70% between 3 to 5 minutes, alcohol 100% twice for 1 minute. The slides were dipped in xylene for at least 5 minutes before being mounted in DPX mounting medium. Section must dry o.n in the hood. The human thymus slides used were arranged by Dr Paola Bonfanti from The Francis Crick Institute in agreement with the UK legislation. Imaging was done using Hamamatsu Slide Scanner at 40x amplification with manual point focus.

2.5 **Immunofluorescence**

For immunofluorescence using paraffin sections, a deparaffinization step was done using xylene twice for 5 minutes and decreasing concentrations of alcohol as 100% twice for 2 minutes, alcohol 95% and 80%, one minute each. After washing, an heat-mediated antigen retrieval was done. This step was made using 1.2L of citrate buffer (citrate sodium 0.1M) boiling during 15 to 20 minutes. After cooling the slides must be washed twice in PBS. Fixation was done with PFA 2% diluted in PBS for 10 minutes at room temperature. Frozen sections dried for 5 minutes at room temperature, washed twice in PBS and fixed in PFA 2% diluted in PBS for 2 minutes. All sections were then permeabilized with Triton X-100 0.5% diluted in PBS at room temperature for 1 hour. Blocking of the samples was done using goat serum 10%, BSA 3%, fish skin gelatine 0.25%, triton X-100 0.05% diluted in PBS, for 1 hour at room temperature. After 3 PBS washes the primary antibodies were diluted in blocking buffer and incubated at 4°C o.n, protected from light. Washes with PBS must be done for at least 3 times, 10 minutes each. Secondary antibodies were incubated for 1 hour at room temperature, protected from light. The antibodies used are indicated in Table 2.1 . Nuclear counter-staining was done with DAPI (Life technologies). Finished the staining proceed to PBS washes at least 3 times, for 10 minutes. The mounting of the slides must be done with ProLong Gold anti-fade reagent (Life Technologies) and sealed with nail polish after dry. The human thymus slides used were arranged by Dr Paola Bonfanti from The Francis Crick Institute in agreement with the UK legislation. Imaging was done using Nikon A1 Upright Confocal microscope.

2.6 **X-gal staining**

Thymus slides were let to dry at room temperature for about 5 minutes and washed in PBS. Fixation was done with PFA 2% diluted in PBS for 10 minutes at room temperature. After washing with PBS twice the slide should be dipped in X-gal reaction solution. This solution is made of 100µL $MgCl_2$ (2M), 2mL FerrI (0.2M), 2mL FerrO (0.2M) and 1mL X-gal substrate (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, 40mg/mL diluted in DMSO) (Sequeira *et al.*, 2013). X-gal reaction incubated for 6 days, at 30°C, at 50rpm, protected from light. The protocol for H&E was then applied following to formalin fixation and skipping to Eosin staining directly.

Table 2.1 Antibodies used for immunoflourescence.

Primary Antibody	Reactiveness	Host	Reference/Clone	Company	Dilution
Anti-Aire	Mouse	Rat	53-5934-82 5H12	eBioscience	1/300
Anti-CD45	Mouse	Rat	103102 B204309	BioLegend	1/200
Anti-EPCAM	Mouse	Rat	12-5791-82 G8.8	eBioscience	1/400
Anti-Involucrin	Mouse	Mouse	Sy7 (in house)	CRUK	1/1000
Anti-Krt10	Mouse	Mouse	MMS-159S-250 DE-K10	Covance	1/100
Anti-Krt14	Human	Chicken	SIG-3476 POLY9060	Biolegend	1/5000
Anti-Krt14	Mouse	Rabbit	PRB-155P-100 POLY19053	Covance	1/200
Anti-Krt76	Human	Rabbit	R08080 HPA019656	Atlas Antibodies	1/200
Anti-Krt8	Mouse	Rat	TROMA-I	DSHB	1/200
Alexa Fluor 488	Chicken	Goat	A11039 1812246	Invitrogen	1/300
Alexa Fluor 488	Rabbit	Goat	A11008 1470706	Life Technologies	1/200
Alexa Fluor 488	Rat	Goat	A11006 1423045	Life Technologies	1/200
Alexa Fluor 555	Mouse	Donkey	A31570 1736967	Life Technologies	1/300
Alexa Fluor 555	Rabbit	Goat	A21428 1903133	Invitrogen	1/200
Alexa Fluor 594	Rat	Donkey	A21209 1744721	Life Technologies	1/300
Alexa Fluor 633	Rat	Goat	A21094 1602779	Life Technologies	1/300
Alexa Fluor 647 Azide	EdU	-	C10340 1524911	Life Technologies	1/400
Alexa Fluor 647	Mouse	Donkey	A31571 1900251	Invitrogen	1/500
Alexa Fluor 647	Rabbit	Donkey	A31573 1693297	Life Technologies	1/300

2.7 **Image analysis**

Medullary region quantification and tumor characterization was done using NanoZoomer Digital Pathology Software. Medullary areas in H&E images were measured using the software's tool freehand region, as well as total size of the lobes. Immunofluorescent images were analysed using ICY image analysis software. Medullary areas and total area of thymus was measured selecting "region of interest" tab and "polygon" available in the ICY software. The areas indicated for each region in each of the programs was analysed using Office Excel. Tumor characterization was done using the same two softwares. The parameters accessed in H&E images were measured using "freehand line" available with the software. Qualitative evaluation was made based on histological characteristics indicated by El-Naggar *et al.* (El-Naggar *et al.*, 2017). Lesion differentiation, immune infiltrate and proliferation was accessed using ICY image analysis software. The region of each tumor was identified in the section through H&E histology comparison. Lesion differentiation was determined based on krt14$^+$ staining and classified as indicated by El-Naggar *et al*. (El-Naggar *et al.*, 2017). Immune infiltration was determined by quantifying the number of DAPI$^+$CD45$^+$ cells per stromal area. This calculation was done on 3 regions of similar sizes evenly distributed along the tumor region. Total number of immune cells was added and divided by the added area of the 3 regions. Proliferation was determined by quantifying the number of DAPI$^+$EdU$^+$ cells per length of krt14$^+$ epithelium. Calculations were

done using Office Excel.

2.8 **Isolation of thymus cells**

The animals were euthanized by cervical dislocation and chest opened. The thymus was prepared using the protocol of Xing and Hogquist with the purpose of increasing the presence of TEC (Xing and Hogquist, 2014). The lobes were dissected, trimmed out of any fat and connective tissue and reserved in 5mL of RPMI-1640 medium. They were then transferred to 2.5mL of enzymatic solution which is Liberase TH 0.05% (w/v) and DNase I 100U/mL diluted in RPMI-1640 medium. After doing small cuts in the lobes, the 6-well plate was kept for 20 minutes at 37°C at 100rpm. After the first incubation the solution was aspirated up and down with a 5mL pipette. The supernatant was transferred to a 50mL falcon tube with 10mL cold albumin-rich buffer which is composed by bovine serum albumin (BSA) 0.5% and EDTA 2mM diluted in 1x PBS (Ca^{2+}/Mg^{2+}-free). It was added 2.5mL more of enzyme solution and incubated for 15 minutes, in the same previous conditions. Mechanical agitation was performed using a 3mL syringe with a 21G needle. The supernatant was once again transferred into the same albumin-rich buffer tube. We added 2.5mL of enzyme solution and incubated for 15 more minutes. We compressed the remaining tissue into a 3mL syringe with a 25G needle repeatedly and incubated the plate for more 10 minutes. Finally, we transferred the remaining solution entirely into the albumin-rich buffer tube. After digestion the solution is centrifuged between 6°C and 10°C at 1200rpm for 8 minutes to force the cells to sediment. Resuspend the pellet in 10mL FACS buffer (PBS + EDTA 4mM) and filter through a 100mm mesh in an attempt of having single cells. Wash the tube and filter with 5mL of the same solution. Add 2mL FACS buffer (wash) before centrifugation in the same conditions but extending the time to 8 minutes. Finally resuspend the pellet in the remaining liquid. This method results in an enrichment from 0.02% to 0.15% of $EPCAM^{+}$ cells.

2.9 **Magnetic separation**

Since the majority of the thymic cells are Tcells, to increase the number of TEC we excluded $CD45^{-}$ cells using the CD45 MicroBeads mouse (MACS – Miltenyi Biotec). After the last centrifugation, we mixed the cells in 200µL of albumin-rich buffer and incubate with 20µL of CD45 MicroBeads for 20 minutes, mixing every 5 minutes, at 4°C. After incubation we washed with FACS buffer and centrifuge in the same previous conditions. Before applying the incubated cells, the LD columns (Miltenyi Biotec) were washed with 2mL FACS buffer. The pellet of the centrifugation was resuspended in 500µL of FACS buffer and applied to the columns proceeding to the tube wash. The first elution, being $CD45^{-}$ cells, was reserved and advance to be stained. The remaining fraction was removed from the magnetic field by pressure, being flushed out and used as controls. Before labelling, we proceed with another wash with FACS buffer and

consequent resuspension of the pellet.

2.10 **Flow cytometry**

We reserved single cells to run as controls. We had unstained control, labelling with DAPI (Molecular Probes) as Live/Dead control. We used OneCompBeads (eBioscience) as fluorescence minus one (FMO) controls for each of the fluorophores. The single cell suspension was blocked with Fc block solution (Rat, anti-mouse CD16/CD32 BD Biosciences) 1µL/100µL and incubated for 10 minutes at 4°C and protected from light. We then stained the cell surface with the primary antibodies at 1µL/100µL for 30 minutes at 4°C and protected from light. The antibodies specifications can be found in Table 2.2. As UEA-1 is conjugated with biotin, we incubated with a secondary streptavidin antibody at 1µL/200µL for 30 minutes in the dark at 4°C, thus amplifying the signal. The cells were ready to analyse after being washed and resuspended. The samples were sorted at the NIHR GSTT Biomedical Research Centre, Guy's Hospital London. Sorted cells were collected into albumin-rich buffer. The cells were kept in ice and ran on a BD Fortessa II cell analyser or sorted on a BD Aria cell sorter. Data was then analysed using FlowJo software.

Table 2.2 Antibodies used for flow cytometry.

Antibody	Reactiveness	Host	Conjugation	Reference/Clone	Company
CD45	Mouse	Rat	FITC	103108 30-F11	Biolegend
EPCAM	Mouse	Rat	PE	12-5791-82 G8.8	eBioscience
UEA-1	Mouse	-	Biotin	B-1065 ZD0810	Vector Labs
Streptavidin	Mouse	-	BV605	563260 7166664	BD Biosciences
Ly51	Mouse	Rat	PerCp/Cy5.5	108316 6C3	Biolegend

2.11 **cDNA synthesis**

Sorted cells were lysed and RNA extracted. This process was done using TaqMan® Gene Expression Cell-to-C_t Kit (Ambion, Life Technologies). Sorted cells were centrifuged at 2000rpm for 5 minutes between 6°C and 10°C. Pellet was resuspended in lysis solution provided by the kit. After incubation, the reaction was stopped by mixing with the stop solution provided. The cell lysates were then stored at -80°C. After thawing, we assembled a reverse transcription (RT) master mix made of RT buffer and RT enzyme mix and added the lysate. A two-step RT thermal cycler incubated for 60 minutes at 37°C, followed by 5 minutes at 95°C. The cDNA formed by this reaction was stored at -20°C.

2.12 **Statistical analysis**

All graph and statistical analysis were done using GraphPad Prism 7 software. Significant differences were calculated automatically with each test indicated in the figure legend. *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$; ns non-significant.

Chapter 3 Results - The genomic landscape of oral cancer mouse model

3.1 Workflow of oral tumors induction

To analyse the genomic landscape of oral SCC, we induced tumorigenesis in the oral cavity. The experimental aim was to mimic the use of tobacco in human tumors and to replicate the alterations caused by tobacco mutagens. This was accomplished using the carcinogen 4NQO in the drinking water to affect areas within the oral cavity, specifically the tongue. 4NQO treatment was applied for a period of 16 weeks and its effect was observed until week 28 (Figure 2.1). The extent of the treatment was efficient to develop tumorigenic lesions and allow their progression from hyperplasias to dysplasias and to invasive SCC, at later stages. Harvesting was done at different time-points to elucidate tumors with different gradings (Figure 2.1). Interestingly, at time points after 4NQO treatment had ceased, more aggressive tumors were found. This proves that the carcinogen can have lag effects and the changes caused by it may influence later developments.

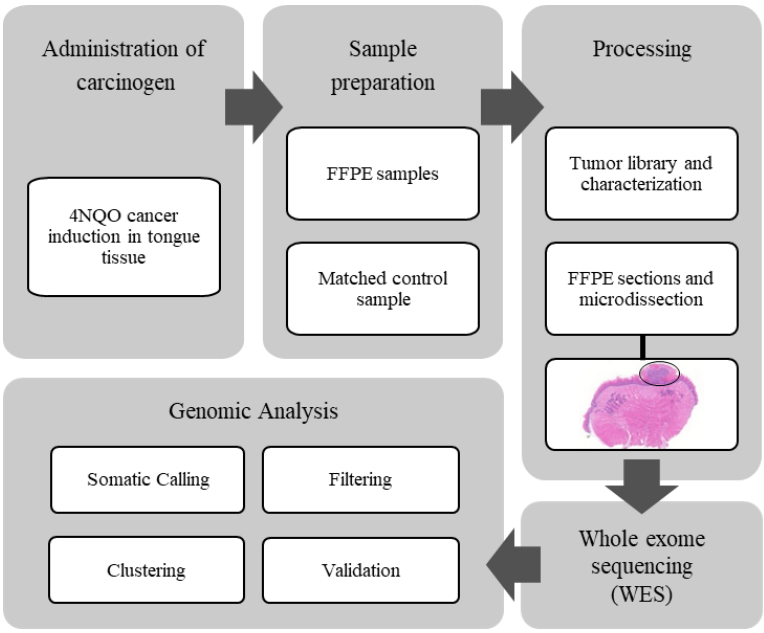

Figure 3.1 Workflow of mouse oral SCC induction using 4NQO. Administration of carcinogen: the carcinogen was added in the drinking water of mice for 16 weeks. Harvesting of tumors was done at different time-points during 28 weeks (Figure 2.1). Sample preparation: the tumors identified were preserved in FFPE blocks. All animals had a normal matched control tissue harvested from the ear as well. Processing: tumor histology features was observed microscopically and characterized. Tumor region was micro-dissected avoiding healthy tissue contamination. Both FFPE blocks and matched control tissue were sent for WES. Genomic Analysis: sequencing data was filtered for inter-individual variability with somatic calling. Filtering, clustering and data validation was completed.

Including control mice, we harvested a total of 69 lesions from 72 animals. Following the same protocol to harvest tumor, healthy matched control tissue were harvested from the ear of all the animals. Lesions were identified macro- and microscopically, and their histological features

characterized. The tumors were micro-dissected from FFPE-sections. Although FFPE might affect the nucleic acids present in the tissue, paraffin was the best fit for micro-dissection of the tumor and to separate the tumorigenic tissue from the remain healthy tissue.

Both tumors and matched controls tissue were sent to the Wellcome Trust Sanger Institute to have their WES. This process was then followed by a somatic calling, filtering, clustering and validation before analysis (Figure 3.1) (Sequeira *et al.*, unpublished).

3.2 **Tumor characterization**

We next wanted to determine the biological changes happening during tumorigenesis. We characterized and classified several biological features of all tumors harvested (Table 3.1). To evaluate grading and histological parameters, tumor paraffin sections were analysed using H&E. This allowed location, grading and to determine the dimensions of the lesions. Lesion differentiation, immune infiltrate and proliferation were quantified using immunofluorescence labelling for krt14, CD45 and EdU, respectively, and counterstaining with nuclear dye DAPI (Figure 3.3).

The experimental design enabled us to observe the progression of lesions from a very early time point at different gradings of hyperplasia to dysplasia and consequently into invasive SCC, at longer treatment periods. In our experimental design we harvested the animals every 2 weeks throughout the 28 weeks. Because of this we could not control the number of tumours obtained in each grading. This led to an uneven number of tumours in each group, which may explain the non-significant results in proliferation and immune infiltrate (Figure 3.3).

On this note, the quantification of $CD45^+$ cells revealed a tendency on increased immune infiltrate (number of immune cells per millimetre-square of the tumor stromal area) as tumors progress (Figure 3.3). This could be due to a crucial role of the immune system in tumorigenesis. It is known that during tumour formation, the immune cells try to signal and destroy the abnormal cells rising, however as tumor continues to grow and mutate, the immune system is evaded. Consequently there is signalling for growth factors and other ligands, thus supporting our hypothesis of an increasing number of immune cells within later tumor progression (Hanahan and Weinberg, 2011).

Table 3.1 Parameters analysed in each mouse tongue section.

Parameters analysed	Classification	Method description
Weeks of treatment	4 – 28	Number of weeks after starting treatment administration as indicated in Figure 2.1.
Lesions per animal	1 – 3	Number of lesions/tumors in each animal's tongue.
Site	D, L, V	Dorsal, lateral, ventral (Figure 3.2).
Lesion grading	1, 2, 3, 4, 5, 6	Ascendant grading indicates increased abnormalities in the tissue. Numbers correspond to normal tissue (1), hyperplasia (2), mild (3), moderate (4), or severe dysplasia (5) and invasive SCC (6).
Lesion dimensions	Length (µm), maximum diameter (µm), maximum invasiveness (µm)	Measurements were done using NanoZoomer Digital Pathology Software. Diameter was only measure in 3 – 4 graded lesions and when applicable. Invasiveness length, when applicable, was measured from the basement membrane.
Tumor growth exophytic	Y (Round, papilloma, verrucous), N	Assessments were made on observation of the H&E sections. Y stands for Yes, N for No.
Tumor invasiveness	Y, N	Assessments were made on observation of the H&E sections. Y stands for Yes, N for No.
Lesion differentiation	Well-differentiated, moderately differentiated	Assessments were made on observation of krt14 staining on immunofluorescence.
Immune infiltrate	1 - 4	Assessments were made on observation of $CD45^+$ staining on immunofluorescence. Index stands for <2 000, 2 000 - 5 000, 5 000 - 10 000, >10 000 cells/mm^2. Quantification was made in the micro-dissection region, in the stroma (Figure 3.3 B).
Proliferation	+, +, +++	Assessments were made on observation of EdU^+ staining on immunofluorescence. Index stands for <150, 150 - 600, >600 cells/mm^2. Quantification was made in the micro-dissection region, in the epithelium (Figure 3.3 B).

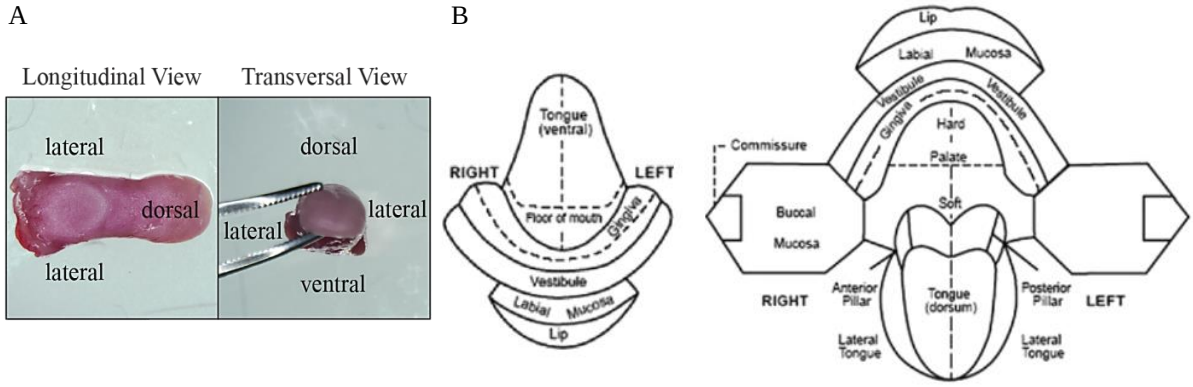

Figure 3.2 Tumor site. A) Mouse tongue views. B) Oral cavity's map. In from Laronde *et al.*, 2014.

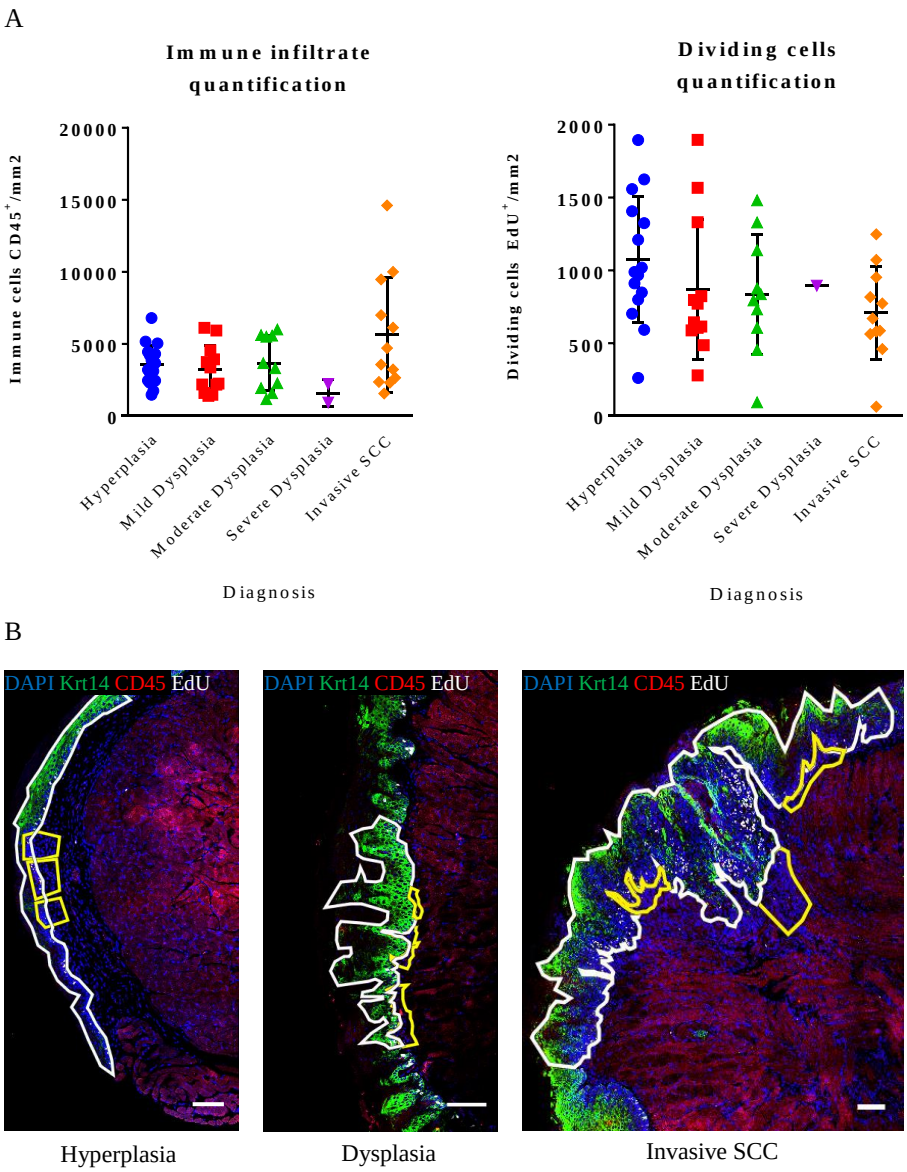

Figure 3.3 Immune infiltrate and proliferation quantification in the different gradings of tumors. A) Quantification of the immune cells ($CD45^+/mm^2$) and dividing cells (EdU^+/mm^2) (bar graphs represent mean ± SD, ns non significant). B) Immunofluorescence representative images of each tumor grading labelled with anti-krt14 (green), anti-CD45 (red), anti-EdU (white) and counterstained with nuclear dye DAPI. Dividing cells were quantifyed by the number of EdU^+ cells in the epithelium ($krt14^+$). Assessment was done per milimeters squared (white regions). Immune infiltrate was quantified by the number of $CD45^+$ cells in the stroma. Three equally distributed regions were quantified along the micro-dissection region (yellow regions). Assessment was done adding the number of cells in the three regions per total area. Scale bar 200µm.

Proliferation was analysed within tumor progression. To assess cell proliferation, we injected each mouse with EdU one hour prior harvesting. We quantified the total number of EdU^+ cells within the tumour region (measured as the number of EdU^+ cells within the epithelial region of the lesion, in milimeters square). The number of EdU^+ cells, identified as proliferative, did not reveal a significant difference. Although neither of the tumors grading groups had revealed differences between them, we observed different distributions (Figure 3.3).

When quantifying the number of proliferative cells within an area, hyperplasias appear to have a more heterogeneous distribution in comparison to more severe tumors. As tumors start to develop, they become larger, more aggressive tumors and tend to be more differentiated thus with more differentiated cells. We hypothesise this observation to be due to an increased number of cells per milimeters square since we did not access cell density. Therefore bigger tumors with more differentiated cells are presenting reduced ratio of $EdU^+/DAPI^+$ cells (Figure 3.3).

3.3 **Mutational signature**

Even though somatic mutations are common to occur, gathering multiple factors together as DNA replication and repair machinery mismatch or exposure to mutagens generates specific patterns of mutations. These combinations of mutation types have been widely analysed in human cancer types resulting in the catalogue of somatic mutations in cancer known as COSMIC (Wellcome Sanger Institute). This data base analysed almost 11 000 exomes and more than 1 000 whole genomes across dozens of cancer types.

To better understand the type of mutations occurring in our model we analysed the most present mutational signatures. As shown in Figure 3.4, the more common alterations are mutations from a cytosine to an adenine, indicated in blue, followed by alterations cytosine to guanine and then cytosine to thymine. According to COSMIC, it corresponds with 3 individual signatures: signature 4, 18 and 29 (Figure 3.5) (Wellcome Sanger Institute).

Signature 4 is characterized by a big contribution of C>A mutations and it was previous detected in head and neck cancers and lung SCC, among others. It is known to be associated with smoking thus being attributed to tobacco mutagens (Wellcome Sanger Institute). Signature 18 has a big influence of C>A and C>T, it is common in neuroblastomas and interestingly verified in breast and stomach carcinomas (Wellcome Sanger Institute). Signature 29 has the same main contributions as signature 4, however with a different pattern of mutations (Wellcome Sanger Institute). This allows the differentiation between the mutations caused by smoking or chewing tobacco, with signature 29 associated with the latter. Therefore, this signature is often present in gingivo-buccal oral SCC. These results are supporting our model as mimicking the mutations happening in oral human tumors.

3.4 **Genomic landscape**

To investigate the mutational landscape caused by the carcinogen 4NQO treatment, we analysed WES data from micro-dissected tumors. Mutations present in the tumors were filtered with the somatic mutations from the matched control tissue. Excluding inter-individual variability, we determined the genetic alterations present in the tumors and how they are changing

along tumor progression.

Firstly, we can confirm that tumors developed within different severity gradings. We determined a correspondence between longer treatment periods and more aggressive tumors (Figure 3.4). The number of single-nucleotide variation (SNV) revealed an increase from the samples on the left – hyperplasia and mild dysplasia – to the ones on the right – moderate to severe dysplasias and invasive SCC. These results confirm the hypothesis that tumor progression is result of increased number of genetic alterations (Hanahan and Weinberg, 2011).

Interestingly the pattern of genes mutated revealed a correlation between treatment duration and number of SNV. Early stage tumors, namely hyperplasia and mild dysplasia, appear to have their SNV dependent on treatment duration (Figure 3.4). The tumors of these gradings which were exposed to longer treatments have substantially more SNV than the ones developed in earlier weeks of experiment.

To understand if the mutations present in our induced mouse oral SCC model were similar to human HNSCC, we compared their genomic landscape. Revising the top significantly mutated genes in HNSCC, The Cancer Genome Atlas Network analysed 279 tumors (Cancer Genome Atlas Network, 2014). Namely TP53, FAT1, CDKN2A, PIK3CA and NOTCH1 were identified as mutated genes in most samples. Interestingly 4 out of 5 of these genes are present in our data (Figure 3.4) (Cancer Genome Atlas Network, 2014). Additionally, Stransky *et al.* stated SYNE1 and PTEN as significant in HNSCC tumorigenesis, confirming once again the relevance of our data (Stransky *et al.*, 2011). Tabatabaeifar *et al.* reviewed the use of next generation sequencing in HNSCC where the genes more frequently found to be mutated in these patients are specified (Tabatabaeifar *et al.*, 2014). The genes identified to be mutated in our experimental model for oral cancer match many genes previously identified in the literature thus highlighting the suitability of our model for use in oral cancer research.

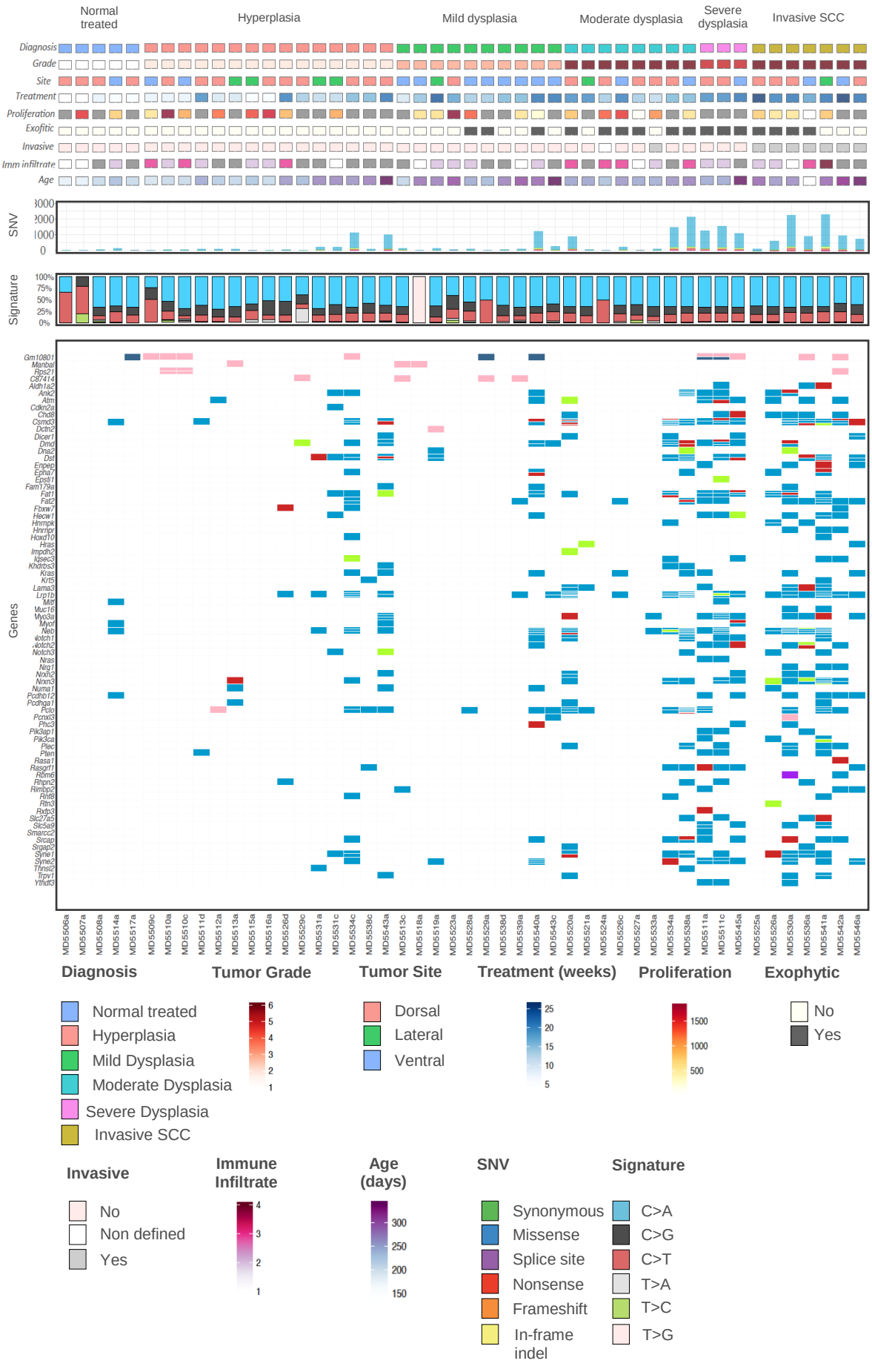
Normal treated
Hyperplasia
Mild dysplasia
Moderate dysplasia
Severe dysplasia
Invasive SCC
Diagnosis
Grade
Site
Treatment
Proliferation
Exofitic
Invasive
Imm infiltrate
Age
SNV
Signature
Genes
Diagnosis
Normal treated
Hyperplasia
Mild Dysplasia
Moderate Dysplasia
Severe Dysplasia
Invasive SCC
Tumor Grade
Tumor Site
Dorsal
Lateral
Ventral
Treatment (weeks)
Proliferation
Exophytic
No
Yes
Invasive
No
Non defined
Yes
Immune Infiltrate
Age (days)
SNV
Synonymous
Missense
Splice site
Nonsense
Frameshift
In-frame indel
Signature
C>A
C>G
C>T
T>A
T>C
T>G

Figure 3.4 Significantly mutated genes in mouse tongue tumors. Colour coding indicates diagnosis, tumor-grade, tumor-site, weeks of treatment, proliferation, exophytic yes/no, invasive yes/no, immune-infiltrate, age of the mice in days, type of SNV and signature contribution (Table 3.1). Genes (rows) with significantly mutated genes; samples (columns, *n*=69) are organized by diagnosis by increasing grading.

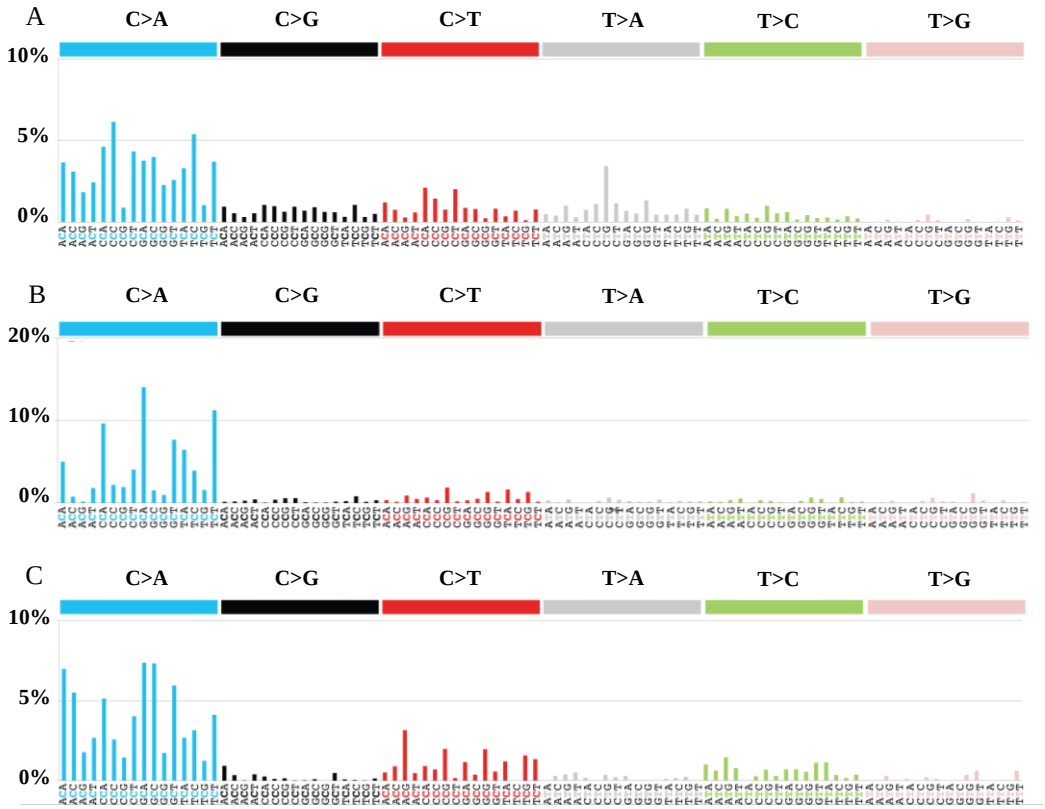

Figure 3.5 Mutational signatures gathered at the catalogue of somatic mutations in cancer - COSMIC (Wellcome Sanger Institute). Color coding represents the type of each mutation. A) Signature number 4. B) Signature number 18. C) Signature number 24.

Chapter 4 Discussion - The genomic landscape of oral cancer mouse model

The aim of this project was twofold to confirm the mouse model we utilised successfully mimicked the molecular and genomic changes that occur in human oral cancer, and to elucidate if these changes occur in early or late stages of tumour development.

The use of the water-soluble carcinogen 4NQO successfully allowed us to observe the gradual development of tumors in the oral cavity of the mouse, capturing the early events of tumour formation. This ongoing study revealed mutational signatures indicative of smoking exposure supporting our model aim. Additionally, we were able to determine some biological trends which correlated with the tumorigenic process.

Although this project provided us with some intriguing clues that can lead to new biomarkers helping on cancer research, we still need to confirm that our data in mouse is completely equivalent to human oral cancer. The critical genes identified in mice correlate to the human genes and consequently, their mutations and molecular implications. Therefore, our data can also give us clues on the mutations involved in the human tumorigenic process. Owing to our experimental design, the data gathered will allow us to determine which mutations are driving tumor growth and leading to a more aggressive tumor.

This study elucidates the changes happening due to cancer thus exposing new biomarkers for cancer therapy. We hypothesize that this experimental design using the carcinogen 4NQO can be implemented as a new model for drug testing once it more closely mimics the human system response to the treatment than a cancer line, per example.

Chapter 5 Results - The role of keratin 76 in thymic organization and immunoregulation

5.1 Thymus histology

Previous studies highlighted a novel role for krt76 (Sequeira *et al.*, 2018). Its immunomodulatory role was noticed in ktr76$^{-/-}$ mice as having increased over-suppressive Tregs (Sequeira *et al.*, 2018). These mice are also more susceptible to tumor development when treated with 4NQO due to an increased local and systemic inflammation (Sequeira *et al.*, 2018). Thus, we wanted to understand how this keratin is regulating the immune response.

To study thymic function, we first determined how is this organ organized. There are different markers that define thymus' anatomy. With the objective of differentiating the cortical from the medullary regions, we stained with anti-krt8 which identified the cortical regions and complemented with anti-krt14 labelling the medullary areas (Figure 5.1) (Anderson and Jenkinson, 2001).

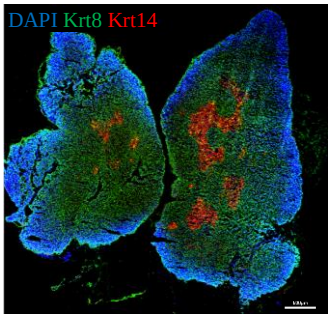

Figure 5.1 Mouse thymus. Immunofluorescence image of thymus labelled with anti-krt8 (green), anti-krt14 (red) and counterstained with nuclear dye DAPI. In red are identified the medullary region, populated by the mTEC, in contrast with the green areas (cortex) where are the cTEC (Figure 1.4). Scale bar 500µm.

5.2 Keratin 76 is expressed in the thymus

Sequeira *et al.* has revealed new information on the role of krt76 using krt76$^{-/-}$ (Sequeira *et al.*, 2018). Proceeding this, we investigated how is the absence of krt76 increasing the Tregs, which will consequently influence cancer formation. Taking advantage of the LacZ insert in the krt76 knockout mice model (Figure 2.2), we labelled the LacZ^{+} cells in the thymus of these mice. We identified blue cells (LacZ^{+}) after X-gal reaction and H&E staining as being in very low number, scattered and mostly in the medullary regions (Figure 5.2).

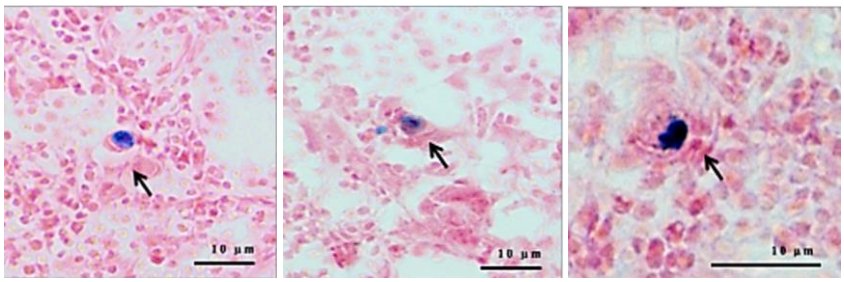

Figure 5.2 Krt76^{+} cells in the thymus. Representative images of adult mouse sections with LacZ^{+} cells (arrows) identified after H&E staining. Scale bar 10µm.

To confirm this expression pattern, we stained the thymus sections by immunofluorescence labelling with anti-krt14 (green), anti-krt76 (red), anti-krt8 (white) and nuclear counterstaining with DAPI. It is

known that the thymus changes with age, degenerating into an organ with a lot of fatty tissue. Our hypothesis was that krt76^{+} cells increased in number with age progression.

To confirm our hypothesis, we labelled krt76^{+}cells in thymus sections from embryonic mice, adult (3-4 months) and older mice (20 months) (Figure 5.3). We also verified that krt76^{+} cells are both present in the medullary and cortical regions (light orange arrowhead in Figure 5.3), whereby they co-express anti-krt14 and anti-krt8, respectively.

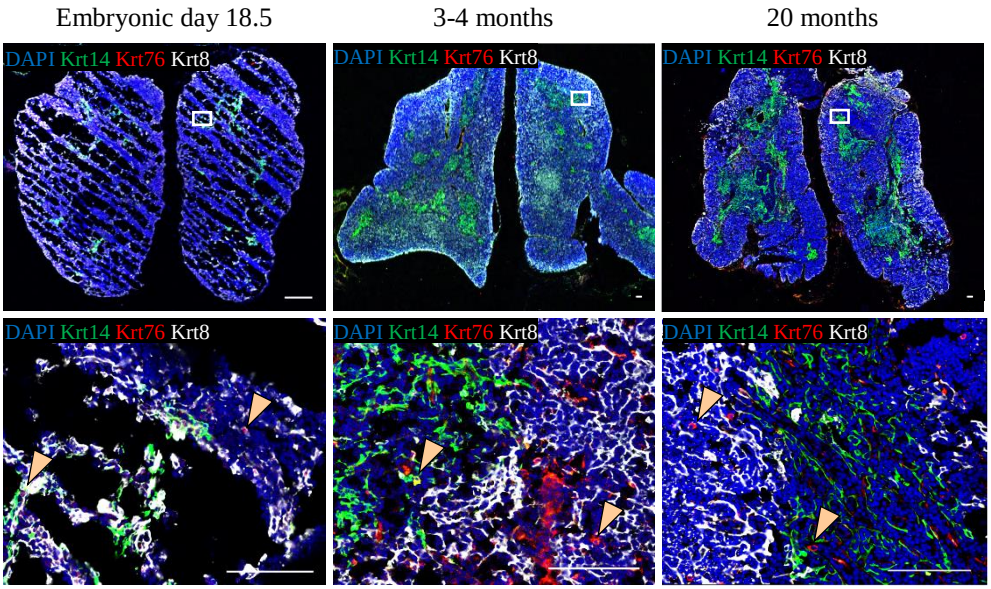

Figure 5.3 Krt76^{+} cells change with age. Immunofluorescence representative images of mouse thymus at embryonic day 18.5, 3 - 4 months and 20 months of age (left to right). The sections were labelled with anti-krt14 (green), anti-krt76 (red), anti-krt8 (white) and counterstained with nuclear dye DAPI. Light orange arroheads indicate krt76^{+} cells in the medulla, co-expressing krt14, and in the cortex, co-expressing krt8. Scale bar 100µm.

5.3 Keratin 76 is expressed in the thymus – comparation mouse with human

We observed that LacZ^{+} cells in H&E staining were often part of a more complex swirled epithelial structures, very similar to specific and exclusive cell structures in the human thymus - HC (Figure 5.2, Figure 5.4 D). To clarify the role of krt76^{+} in the thymus, we proceed to the comparison with human foetal thymus (Figure 5.4). Using immunofluorescence, we verified that human krt76 is co-localized with HC markers, while these structures are surrounded by krt14^{+} cells.

We confirmed the presence of krt76 in some rare mouse cells, mostly localized in the medullary regions (anti-krt14^{+}) which were also positive for EPCAM, a marker of epithelial cells. In mice we found a different structure from the typical swirled structures of HC - we observed rare isolated krt76^{+} cells surrounded by krt14^{+} cells (Figure 5.4).

Interestingly, in both species – human and mouse - krt76^{+} cells had a consistency of markers expressing involucrin and krt10, two specific markers of HC. Co-localization suggests that these structures could be the equivalent of HC in mouse. Since mice don’t have these

structures, krt76^{+} cells might be mimicking its function. Lacking the swirled layered circle, in mice they are isolated however, keeping krt14^{+} cells as neighbours. It is known that HC can increase dendritic cells maturation, which are capable of inducing differentiation of Tregs in the thymus (Nomura and Sakaguchi, 2007; Sakaguchi *et al.*, 2008; Caramalho *et al.*, 2015). Therefore, these findings can shed some light on the mechanism for Treg alteration in krt76$^{-/-}$.

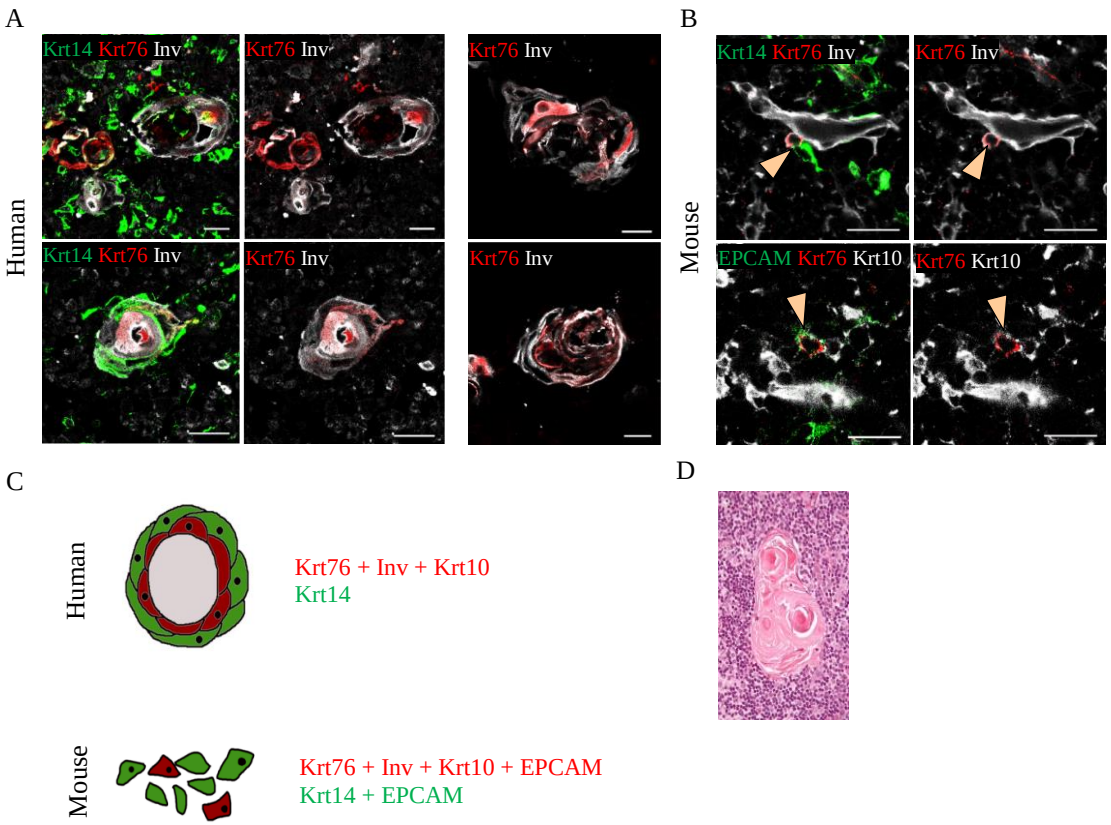

Figure 5.4 Krt76^{+} is co-localized with HC specific markers. A) Immuflourescence representative images of human thymus labelled with anti-krt14 (green), anti-krt76 (red) and anti-involucrin (white). Scale bar 30µm. B) Immuflourescence representative images of mouse thymus labelled with anti-krt14 or anti-EPCAM (green), anti-krt76 (red) and anti-involucrin or anti-krt10 (white). Scale bar 30µm. C) Schematic of krt76^{+} cell markers in human and mouse thymic cells. D) Example of human HC. Slides were kindly provided by Dr Paola Bonfanti from The Francis Crick Institute.

5.4 Alterations in the krt76$^{-/-}$ thymus

Previous work has shown that krt76$^{-/-}$ has no changes in global thymus architecture, size or total cellularity (Sequeira *et al.*, 2018). To access the changes also noted in Treg population by Sequeira *et al.*, we aimed to unravel how the Treg maturation process was being affected (Sequeira *et al.*, 2018).

Staining for H&E allowed us to measure the medullary regions (stained in light pink) as well as outline the thymus lobes (Figure 5.5). This quantification was done by Ms Marina Torre and permitted us to compare the architecture proportions of both phenotypes (control and ktr76$^{-/-}$

). We verified changes in the size of the medullary regions in krt76$^{-/-}$ mice. To confirm this difference, we used immunofluorescence to quantify the medullary regions (krt14^{+}) as well as the total area of the lobe (counterstained with DAPI) (Figure 5.5). Once again, the medullary regions are increased in krt76$^{-/-}$ thymus. Hence TEC present in the medulla are extremely important in Treg generation, this proportion could contribute to the increased number of Treg found with our mouse model.

The human thymus slides used were arranged by Dr Paola Bonfanti from The Francis Crick Institute in agreement with the UK legislation. The human thymus slides used were arranged by Dr Paola Bonfanti from The Francis Crick Institute in agreement with the UK legislation.

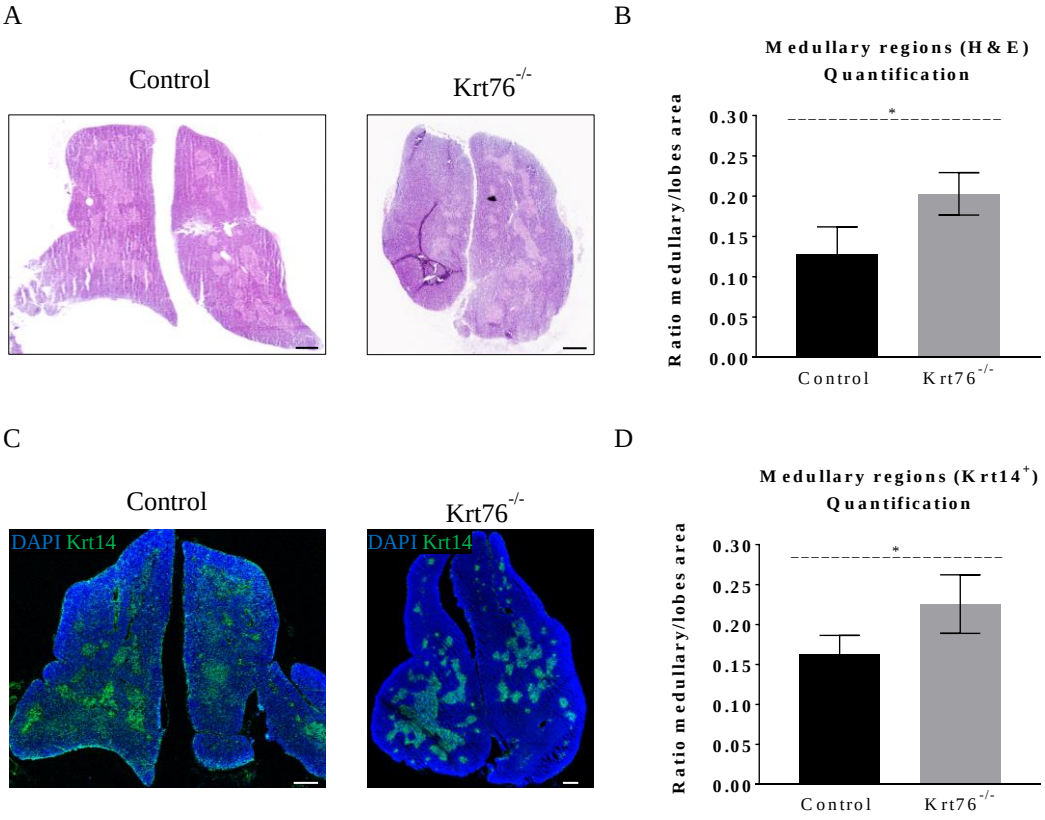

Figure 5.5 Medullary areas are increased in krt76$^{-/-}$ mouse thymus. A) Representative image of H&E thymus of both phenothypes with medullary and cortical regions differently stainined. Scale bar 500µm. B) Quantification of the ratio medullary/lobes area by H&E stainings. *n*=5 animals/group. C) Representative image of thymus of both phenotypes by immunoflourescence labelled with anti-krt14 identifying medulla and counterstained with nuclear dye DAPI. Scale bar 500µm. D) Quantification of the ratio medullary/lobes area by anti-krt14^{+} immunoflourescence. *n*=4 animals/group. C-D) Bar graphs represent mean ± SD, Statistic t-test, *p<0.05.

5.5 Thymic populations analysis

So far, we understand that there is a difference in the medullary regions. However, it remains to be discovered which cells and populations are causing it. To better understand the mechanism behind this architectural change, we analysed the different thymic populations. Using

flow cytometry, we analysed the immune cells ($CD45^+EPCAM^-$) and epithelial cells ($CD45^-EPCAM^+$) (Abramson and Anderson, 2017). Using both markers we were able to quantify the number of cells in each population from each phenotype: control and $krt76^{-/-}$ (Figure 5.6). In $krt76^{-/-}$ the total number of epithelial cells was approximately 1/3 smaller than those existents in the control mice.

Considering the decreased number of epithelial cells, we can deduce an increase on the immune cells ($CD45^+$), once these two populations make up for the similar total number of cells in the thymus (Sequeira *et al.*, 2018). To access the increase in immune cells ($CD45^+$) it is still to be examined if there is an abnormal proliferation of this population. Liakath-Ali *et al.* described $krt76^{-/-}$ as having a thickening of the epidermis and more importantly an hyperproliferation of the basal layer (Liakath-Ali *et al.*, 2014). We can then hypothesise that the loss of krt76 can lead to dysregulation of maturation of T cells in the thymus, justifying the imbalance found.

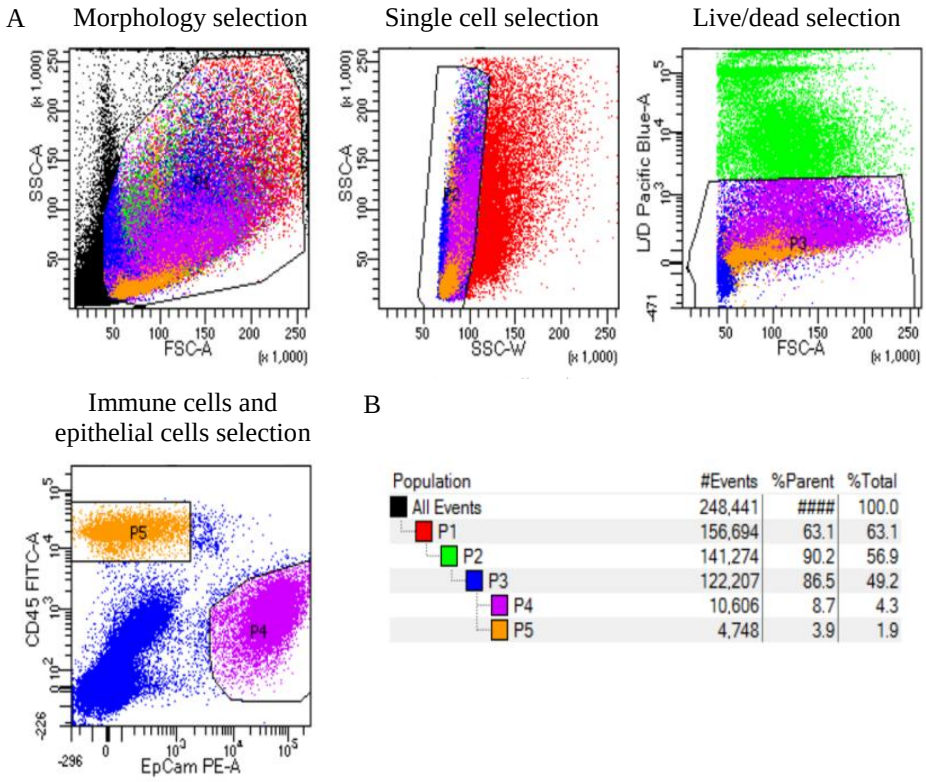

Figure 5.6 Gating strategy for analysing the immune cells and epithelial cells. A) The top left gating was done first and to separate the cells from debri (morphology selection), top middle is single cells selection to avoid doublets. DAPI staining was used to select for live/dead cells in top right. The populations $CD45^+EPCAM^-$ were identified as immune cells (P4), while $CD45^-EPCAM^+$ were epithelial cells (P5). B) Hierarchy of population gated.

To investigate why $krt76^{-/-}$ have increased medullary areas in addition to decreasing number of TEC we investigated which populations were altered and leading to this difference. To analyse if there was a disproportion of the medullary and cortical TEC, we did flow cytometry

using differential markers for these two populations (Xing and Hogquist, 2014). mTEC are identified by Ly51^{low}UEA-1^{+}, while cTEC are characterized by Ly51^{+}UEA-1^{-} (Xing and Hogquist, 2014; Lopes *et al.*, 2015; Abramson and Anderson, 2017). We applied a similar gating strategy (Figure 5.6). We did not observed a significant difference on the percentage of either TEC populations in krt76$^{-/-}$ (Figure 5.7 B). This experiment needs to be replicated in order to confirm these results and immunofluorescence should be conducted using anti-Ly51 and anti-EUA-1 confirming antibody specificity and these results.

The human thymus slides used were arranged by Dr Paola Bonfanti from The Francis Crick Institute in agreement with the UK legislation.

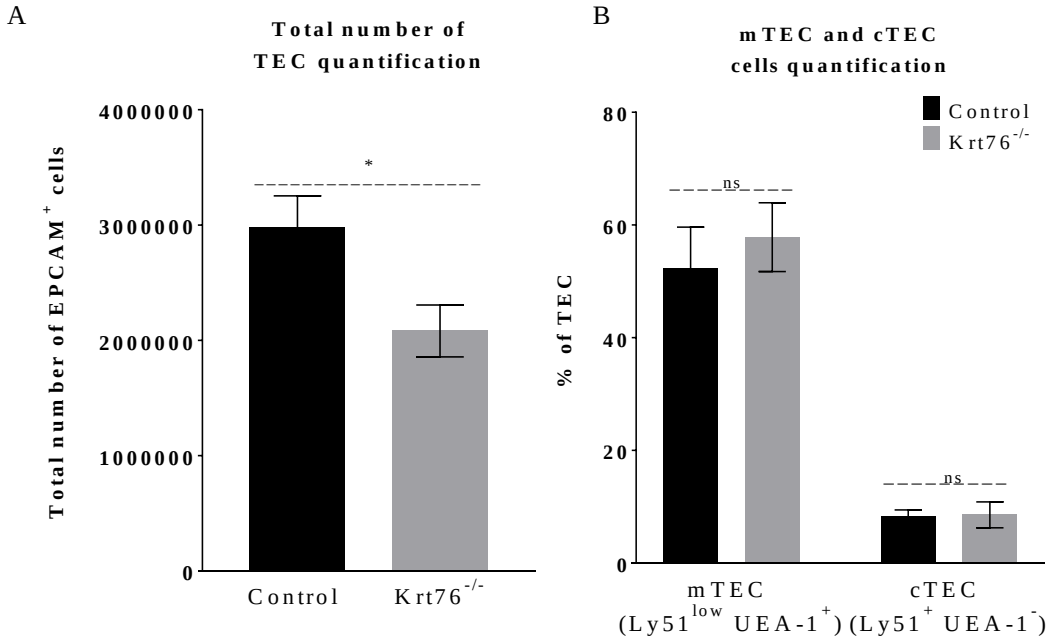

Figure 5.7 TEC populations are altered in krt76$^{-/-}$. A) Quantification of total number of TEC as EPCAM^{+} of both phenotypes. B) Quantification of mTEC and cTEC of thymus of both phenotypes. mTEC here identified Ly51^{low}UEA-1^{+}. cTEC identified Ly51^{+}UEA-1^{-}. A-B) Bar graphs represent mean ± SD, Statistic t-test, *p<0.05, ns non significative. *n*=3 animals/group.

5.6 **Targets of keratin 76**

To understand how is krt76 inducing these changes in our model, we questioned its possible targets. Keratins are known to interact with several other proteins. Krt76 is not an exception as it has been shown to interact with claudin 1, for example and is involved in important cell functions and wound mechanism (DiTommaso *et al.*, 2014). In the literature there are several interactions between keratins and the immune system involving tumour formation (Hobbs *et al.*, 2015).

Krt17 was confirmed to be able to regulate Aire, an immune transcription factor (Hobbs *et al.*, 2015). While both krt17 and Aire in their active forms are necessary to develop skin cancer

(Hobbs *et al.*, 2015). Another paper supporting our theory is from Malchow *et al*. where they confirm the importance of Aire in Treg maturation, differentiation and function (Malchow *et al.*, 2016). Because of Aire's involvement with the immune system, specifically Treg maturation, adding to the connection with tumor growth, we hypothesized its interaction with krt76.

Using immunofluorescence, we labelled the control and krt76$^{-/-}$ thymus sections with anti-aire, anti-involucrin, anti-krt76 and nuclear counterstain with DAPI (Figure 5.8). The absence of krt76 in the krt76$^{-/-}$ was clear, confirming a successful knockout and the specificity of our antibody. Additionally, at later age krt76^{+} cells (green) appear to be increased (Figure 5.8, 20 months).

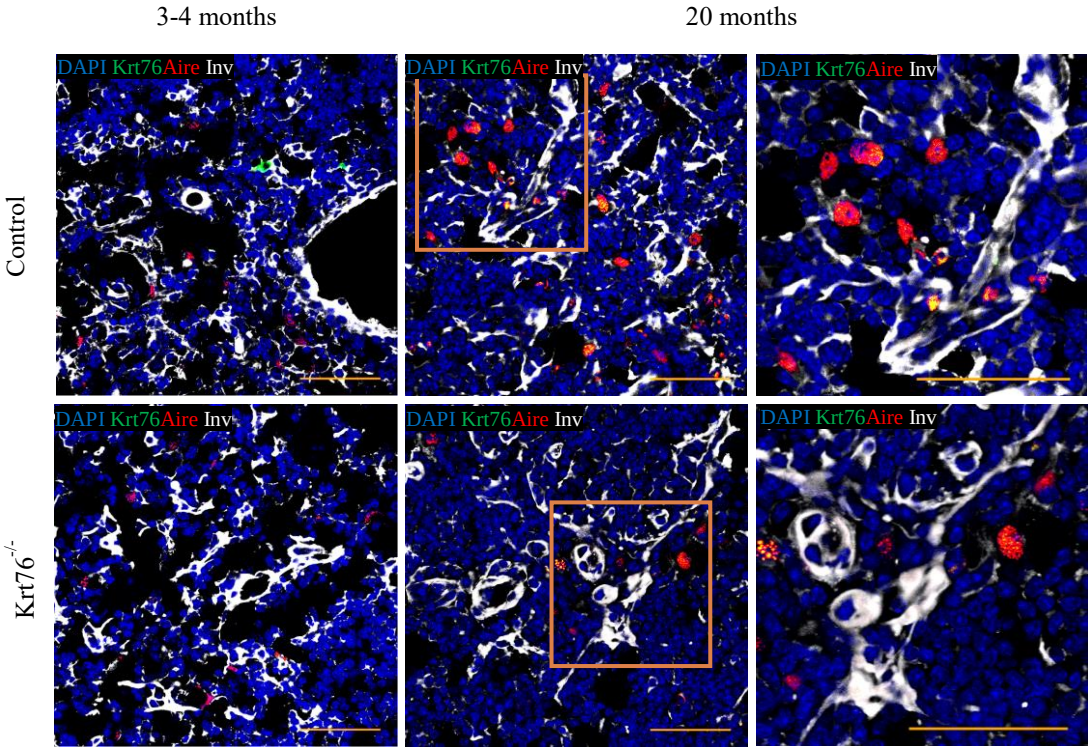

Figure 5.8 Aire^{+} cells are increased with age. Representative images of immunoflourescence labelled with anti-aire (green), anti-involucrin (red), anti-krt76 (white) and counterstained with the nuclear dye DAPI. Top row represents control animals in adulthood (left) and older with 20 months of age (center and right). Down row represents krt76$^{-/-}$ at the same time points. Scale bar 100µm.

It is important to note that aire – being a transcription factor – appears in the nucleus. Immunostaining labelling against aire shown an apparent decrease in ktr76$^{-/-}$ when comparing to the control. Although this trend was not quantified, it is interesting to note that krt76 was found to be co-localized with aire in the nucleus, mostly in older mice, which needs to be confirmed in a big resolution image. Krt76 has not been referenced to be in the nucleus however it is not an irrational hypothesis since Hobbs *et al.* stated that krt17 become nuclear acting as transcription factors interacting with Aire and as cell cycle regulators (Hobbs, Jacob and Coulombe, 2016).

The number of Aire^{+} cells appears to be increased in the older mice. Malchow *et al.* revealed that Aire$^{-/-}$ mice, with a phenotype of auto-immune disease, have the Treg altered in the thymus into conventional Tcell hence their correct differentiation is an Aire-dependent process (Malchow *et al.*, 2016).

The suggested decreased number of Aire^{+} cells is noticed from the control to krt76$^{-/-}$ suggesting that krt76 may be upregulating the transcription factor Aire. This alteration could be happening directly, by the entrance of krt76 into the nucleus, or indirectly through interactions with secondary proteins. We hypothesise that krt76 is causing a dysregulation of Aire, a major player in Treg differentiation, which might lead to an excessive Treg differentiation and abnormal selection as verified by Sequeira *et al*. (Sequeira *et al.*, 2018). To confirm our theory, we need to understand the mechanism behind the interactions by analysing other possible targets of krt76 involved in Treg maturation applying qPCR to cDNA synthesized from by flow cytometry sorted cells.

Chapter 6 Discussion - The role of keratin76 in thymic organization and immunoregulation

Taken together these results paired with the fact that krt76$^{-/-}$ mice have increased Treg and are more suppressive, led to the hypothesis that krt76 may play a role in the maturation of the Treg (Sequeira *et al.*, 2018). These cells, being responsible for the inhibition of Tcells, are extremely important in the immune system preventing autoimmune diseases.

We found new data on the expression of krt76 on the thymus, supporting once again that keratins are not only structural proteins but may play a role in regulating the immune system. We determined a reduction on the medullary regions caused by the absence of krt76 although this disproportion was not attributed either to mTEC neither to cTEC. As these cells are responsible for the maturation and differentiation of Tregs, there is still to be unravelled how krt76 is changing thymic function.

To unravel the mechanism behind the influence of this keratin and the maturation of immune cell, we access the interaction with Aire which is known to be involved in Tregs differentiation. Although we got some interesting clues from their interaction, we still need to clarify the mechanism by which krt76 is regulating this process. Adding to this, we need to investigate and explore other possible targets for krt76 that can be involved in TEC or Tregs maturation (Liu *et al.*, 2013; DiTommaso *et al.*, 2014; Hobbs *et al.*, 2015; Lopes *et al.*, 2015).

The next step could be using RNAseq to access a differential expression on a much bigger gene list between the control and krt76$^{-/-}$ mouse. On a follow-up of our discoveries it would be helpful to analyse the changes caused by the absence of this keratin *in vitro*, thus transposing our hypothesis to human cells, giving us clues on some biomarkers and possible immune therapies that could be used for cancer.

Chapter 7 Conclusions

In this study we explored tumor progression both histological and genetic mutations. We were able to understand the molecular changes as well as the mutations happening due to tumorigenesis. Continuing with the genomic analysis, we expect to do a temporal model to determine which mutations are leading to more aggressive tumors and are responsible for cancer hallmarks like invasiveness.

Along this line, we hypothesise that this experimental model will be applied in other studies. It mimics the global biological changes happening in tumorigenesis thus being an innovate cancer model providing new insights in tumor progression. Additionally, its complexity also allows drug therapy research.

We explored the role of krt76 in immunoregulation due to its involvement with tumor progression (Sequeira *et al.*, 2018). Our experiments reveal that keratin 76 is involved in thymic function and architecture. Further studies will unravel the mechanism behind this interaction that is affecting Tregs. This can lead to a better comprehension of the role of krt76 with the immune system in tumor progression contributing to better therapy approaches.

Chapter 8 References

Abramson, J. and Anderson, G. (2017) 'Thymic Epithelial Cells', *Annual Review of Immunology*, 35, pp. 85–118. doi: 10.1146/annurev-immunol.

Agrawal, N. *et al.* (2011) 'Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1', *Science*, 333(6046), pp. 1154–1156. doi: 10.1126/science.1206923.

Ambatipudi, S. *et al.* (2013) 'Downregulation of Keratin 76 Expression during Oral Carcinogenesis of Human, Hamster and Mouse', *PLoS ONE*, 8(7). doi: 10.1371/journal.pone.0070688.

Anderson, G. and Jenkinson, E. J. (2001) 'Lymphostromal interactions in thymic development and function', *Nature Reviews Immunology*, 1(1), pp. 31–40. doi: 10.1038/35095500.

Aschenbrenner, K. *et al.* (2007) 'Selection of Foxp3+ regulatory T cells specific for self antigen expressed and presented by Aire+ medullary thymic epithelial cells', *Nature Immunology*, 8(4), pp. 351–358. doi: 10.1038/ni1444.

Bray, F. *et al.* (2018) 'Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries', *A Cancer Journal for Clinicians*, 0(0), pp. 1–31. doi: 10.3322/caac.21492.

Cancer Genome Atlas Network (2014) 'Comprehensive genomic characterization of head and neck squamous cell carcinomas', *Nature*, 517. doi: 10.1038/nature14129.

Cancer Research UK (2018) *Head and neck cancers incidence statistics*, *Cancer Research UK*. Available at: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/head-and-neck-cancers/incidence.

Caramalho, Í. *et al.* (2015) 'Regulatory T-cell development in the human thymus', *Frontiers in Immunology*, p. 395. doi: 10.3389/fimmu.2015.00395.

Cecco, L. De *et al.* (2015) 'Head and neck cancer subtypes with biological and clinical relevance: Meta-analysis of gene-expression data', *Oncotarget*, 6(11), pp. 9627–9642. doi: 10.18632/oncotarget.3301.

Cheon, D.-J. and Orsulic, S. (2011) 'Mouse Models of Cancer', *Annual Review of Pathology: Mechanisms of Disease*, 6(1), pp. 95–119. doi: 10.1146/annurev.pathol.3.121806.154244.

Christopherson, W. M. (1977) 'Dysplasia, carcinoma in situ, and microinvasive carcinoma of the uterine cervix', *Human Pathology*, 8(5), pp. 489–501. doi: 10.1016/S0046-8177(77)80110-X.

Chung, C. H. *et al.* (2004) 'Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression', *Cancer Cell*, 5, pp. 489–500. doi: S1535610804001126 [pii].

Coulombe, P. A. (2017) 'The Molecular Revolution in Cutaneous Biology: Keratin Genes and their Associated Disease: Diversity, Opportunities, and Challenges', *Journal of Investigative Dermatology*, pp. e67–e71. doi: 10.1016/j.jid.2016.04.039.

Depianto, D. *et al.* (2010) 'Keratin 17 promotes epithelial proliferation and tumor growth by polarizing the immune response in skin', *Nature Genetics*, 42(10), pp. 910–914. doi: 10.1038/ng.665.

DiTommaso, T. *et al.* (2014) 'Keratin 76 Is Required for Tight Junction Function and Maintenance of the Skin Barrier', *PLoS Genetics*, 10(10), p. e1004706. doi: 10.1371/journal.pgen.1004706.

Downes, D. J. *et al.* (2014) 'Characterization of the Mutagenic Spectrum of 4-Nitroquinoline 1-Oxide (4-NQO) in Aspergillus nidulans by Whole Genome Sequencing', *G3 Genes|Genomes|Genetics*, 4(12), pp. 2483–2492. doi: 10.1534/g3.114.014712.

Edge, S. B. and Compton, C. C. (2010) ‘The american joint committee on cancer: The 7th edition of the AJCC cancer staging manual and the future of TNM’, *Annals of Surgical Oncology*, pp. 1471–1474. doi: 10.1245/s10434-010-0985-4.

El-Naggar, A. K. *et al.* (2017) *WHO Classification of Head and Neck Tumours*. Lyon: International Agency for Research on Cancer (IARC). ISBN-13: 978-92-832-2438-9.

Fronza, G. *et al.* (1992) ‘The 4-nitroquinoline 1-oxide mutational spectrum in single stranded DNA is characterized by guanine to pyrimidine transversions’, *Nucleic Acids Research*, 20(6), pp. 1283–1287. doi: 10.1093/nar/20.6.1283.

Galbiatti, A. L. S. *et al.* (2013) ‘Head and neck cancer: Causes, prevention and treatment’, *Brazilian Journal of Otorhinolaryngology*, 79(2), pp. 239–247. doi: 10.5935/1808-8694.20130041.

Gaykalova, D. A. *et al.* (2014) ‘Novel insight into mutational landscape of head and neck squamous cell carcinoma’, *PLoS ONE*, 9(3), p. e93102. doi: 10.1371/journal.pone.0093102.

Ge, S. *et al.* (2016) ‘Dynamic changes in the gene expression profile during rat oral carcinogenesis induced by 4-nitroquinoline 1-oxide’, *Molecular Medicine Reports*, 13(3), pp. 2561–2569. doi: 10.3892/mmr.2016.4883.

Hanahan, D. and Weinberg, R. A. (2011) ‘Hallmarks of Cancer: The Next Generation’, *Cell*, pp. 646–674. doi: 10.1016/j.cell.2011.02.013.

Hayes, T. F. *et al.* (2016) ‘Integrative genomic and functional analysis of human oral squamous cell carcinoma cell lines reveals synergistic effects of FAT1 and CASP8 inactivation’, *Cancer Letters*, 383, pp. 106–114. doi: 10.1016/j.canlet.2016.09.014.

Herrmann, H. and Aebi, U. (2016) ‘Intermediate filaments: Structure and assembly’, *Cold Spring Harbor Perspectives in Biology*, 8, p. a018242. doi: 10.1101/cshperspect.a018242.

Hobbs, R. P. *et al.* (2015) ‘Keratin-dependent regulation of Aire and gene expression in skin tumor keratinocytes’, *Nature Genetics*, 47(8), pp. 933–938. doi: 10.1038/ng.3355.

Hobbs, R. P., Jacob, J. T. and Coulombe, P. A. (2016) ‘Keratins Are Going Nuclear’, *Developmental Cell*, 38(3), pp. 227–233. doi: 10.1016/j.devcel.2016.07.022.

Hobbs, R. P., Lessard, J. C. and Coulombe, P. A. (2012) ‘Keratin intermediate filament proteins – novel regulators of inflammation and immunity in skin’, *Journal of Cell Science*, 125(22), pp. 5257–5258. doi: 10.1242/jcs.122929.

Hu, Z. *et al.* (2017) ‘CCR7 Modulates the Generation of Thymic Regulatory T Cells by Altering the Composition of the Thymic Dendritic Cell Compartment’, *Cell Reports*, 21, pp. 168–180. doi: 10.1016/j.celrep.2017.09.016.

Kandoth, C. *et al.* (2013) ‘Mutational landscape and significance across 12 major cancer types’, *Nature*, 502(7471), pp. 333–339. doi: 10.1038/nature12634.

Kanojia, D. and Vaidya, M. M. (2006) ‘4-Nitroquinoline-1-oxide induced experimental oral carcinogenesis’, *Oral Oncology*, 42, pp. 655–667. doi: 10.1016/j.oraloncology.2005.10.013.

Kreimer, A. R. *et al.* (2005) ‘Human Papillomavirus Types in Head and Neck Squamous Cell Carcinomas Worldwide : A Systematic Review Human Papillomavirus Types in Head and Neck Squamous Cell Carcinomas Worldwide : A Systematic Review’, *Cancer Epidemiology Biomarkers and Prevention*, 14(February), pp. 467–475. doi: 10.1158/1055-9965.EPI-04-0551.

Laronde, D. M. *et al.* (2014) ‘Influence of fluorescence on screening decisions for oral mucosal lesions in community dental practices’, *Journal of Oral Pathology and Medicine*, 43(1), pp. 7–13. doi: 10.1111/jop.12090.

Leemans, C. R., Braakhuis, B. J. M. and Brakenhoff, R. H. (2011) 'The molecular biology of head and neck cancer', *Nature Reviews Cancer*, 11, pp. 9–22. doi: 10.1038/nrc2982.

Liakath-Ali, K. *et al.* (2014) 'Novel skin phenotypes revealed by a genome-wide mouse reverse genetic screen', *Nature Communications*, 5, p. 3540. doi: 10.1038/ncomms4540.

Lio, C. W. J. and Hsieh, C. S. (2011) 'Becoming self-aware: The thymic education of regulatory T cells', *Current Opinion in Immunology*, 23, pp. 213–219. doi: 10.1016/j.coi.2010.11.010.

Liu, B. *et al.* (2013) 'Cbx4 regulates the proliferation of thymic epithelial cells and thymus function', *Development*, 140(4), pp. 780–788. doi: 10.1242/dev.085035.

Lopes, N. *et al.* (2015) 'Thymic crosstalk coordinates medulla organization and T-cell tolerance induction', *Frontiers in Immunology*. doi: 10.3389/fimmu.2015.00365.

Malchow, S. *et al.* (2016) 'Aire Enforces Immune Tolerance by Directing Autoreactive T Cells into the Regulatory T Cell Lineage', *Immunity*, 44(5), pp. 1102–1113. doi: 10.1016/j.immuni.2016.02.009.

Mardis, E. R. (2008) 'The impact of next-generation sequencing technology on genetics', *Trends in Genetics*, pp. 133–141. doi: 10.1016/j.tig.2007.12.007.

Mardis, E. R. and Wilson, R. K. (2009) 'Cancer genome sequencing: A review', *Human Molecular Genetics*, 18(R2), pp. R163–R168. doi: 10.1093/hmg/ddp396.

Marur, S. and Forastiere, A. A. (2008) 'Head and neck cancer: Changing epidemiology, diagnosis, and treatment', *Mayo Clinic Proceedings*, 83(4), pp. 489–501. doi: 10.4065/83.4.489.

Mehanna, H. *et al.* (2010) 'Head and neck cancer — Part 2 : Treatment and prognostic factors', *British Medical Association*, 341, pp. 721–725. doi: 10.1136/bmj.c4690.

Metzker, M. L. (2009) 'Sequencing technologies — the next generation', *Nature Reviews Genetics*, 11. doi: 10.1038/nrg2626.

Moll, R., Divo, M. and Langbein, L. (2008) 'The human keratins: Biology and pathology', *Histochemistry and Cell Biology*, pp. 705–733. doi: 10.1007/s00418-008-0435-6.

Morris, L. G. T. *et al.* (2017) 'The molecular landscape of recurrent and metastatic head and neck cancers insights from a precision oncology sequencing platform', *JAMA Oncology*, 3(2), pp. 244–255. doi: 10.1001/jamaoncol.2016.1790.

Nassar, D. *et al.* (2015) 'Genomic landscape of carcinogen-induced and genetically induced mouse skin squamous cell carcinoma', *Nature Medicine*, 21(8), pp. 946–954. doi: 10.1038/nm.3878.

National Cancer Institute (2013) *Tumor Grade Fact Sheet, 3 May 2013*. Available at: https://www.cancer.gov/about-cancer/diagnosis-staging/prognosis/tumor-grade-fact-sheet.

National Cancer Institute (2015) *Cancer Staging, 9 March 2015*. Available at: https://www.cancer.gov/about-cancer/diagnosis-staging/staging.

National Cancer Institute (2017) *Head and Neck Cancers, 29 March 2017*. Available at: https://www.cancer.gov/types/head-and-neck/head-neck-fact-sheet.

Nomura, T. and Sakaguchi, S. (2007) 'Foxp3 and Aire in thymus-generated Treg cells: a link in self-tolerance', *Nature Immunology*, 8(4), pp. 333–334. doi: 10.1038/ni0407-333.

Osei-Sarfo, K. *et al.* (2013) 'The molecular features of tongue epithelium treated with the carcinogen 4-nitroquinoline-1-oxide and alcohol as a model for HNSCC', *Carcinogenesis*, 34(11), pp. 2673–2681. doi: 10.1093/carcin/bgt223.

Pezzano, M. *et al.* (2001) 'Questionable thymic nurse cell.', *Microbiology and Molecular Biology*

Reviews, 65(3), pp. 390–403. doi: 10.1128/MMBR.65.3.390-403.2001.

Ragin, C. C. R. and Taioli, E. (2007) 'Survival of squamous cell carcinoma of the head and neck in relation to human papillomavirus infection: Review and meta-analysis', *International Journal of Cancer*, pp. 1813–1820. doi: 10.1002/ijc.22851.

Reboux, W. (2018) *Key facts; Risk factors for cancers; Reducing the cancer burden; Early detection; Early diagnosis, 1 February 2018*. Available at: http://www.who.int/en/news-room/fact-sheets/detail/cancer.

Roth, W. *et al.* (2012) 'Keratin 1 maintains skin integrity and participates in an inflammatory network in skin through interleukin-18', *Journal of Cell Science*, 125(22), pp. 5269–5279. doi: 10.1242/jcs.116574.

Sakaguchi, S. *et al.* (2008) 'Regulatory T Cells and Immune Tolerance', *Cell*, pp. 775–787. doi: 10.1016/j.cell.2008.05.009.

Sequeira, I. *et al.* (2013) 'Microdissection and Visualization of Individual Hair Follicles for Lineage Tracing Studies', *Article in Methods in Molecular Biology*. doi: 10.1007/7651_2013_48.

Sequeira, I. *et al.* (2018) 'Immunomodulatory role of Keratin 76 in oral and gastric cancer', *Nature Communications*, 9(1), p. 3437. doi: 10.1038/s41467-018-05872-4.

Sequeira, I. *et al.* (no date) *The genomic landscape of carcinogen-induced mouse oral squamous cell carcinoma*. Centre for Stem Cells & Regenerative Medicine: Unpublished.

Skarnes, W. C. *et al.* (2011) 'A conditional knockout resource for the genome-wide study of mouse gene function', *Nature*, 474. doi: 10.1038/nature10163.

Stransky, N. *et al.* (2011) 'The Mutational Landscape of Head Squamous Cell Carcinoma.', *Science*, 333(6046), pp. 1157–1160. doi: 10.1126/science.1208130.

Sun, W. and Califano, J. A. (2014) 'Sequencing the head and neck cancer genome: Implications for therapy', *Annals of the New York Academy of Sciences*, 1333(1), pp. 33–42. doi: 10.1111/nyas.12599.

Tabatabaeifar, S. *et al.* (2014) 'Use of next generation sequencing in head and neck squamous cell carcinomas: A review', *Oral Oncology*, 50(11), pp. 1035–1040. doi: 10.1016/j.oraloncology.2014.08.013.

Takaba, H. *et al.* (2015) 'Fezf2 Orchestrates a Thymic Program of Self-Antigen Expression for Immune Tolerance Article Fezf2 Orchestrates a Thymic Program of Self-Antigen Expression for Immune Tolerance', *Cell*, 163, pp. 975–987. doi: 10.1016/j.cell.2015.10.013.

Tang, X.-H. *et al.* (2015) 'Gene expression profiling signatures for the diagnosis and prevention of oral cavity carcinogenesis-genome-wide analysis using RNA-seq technology', *Oncotarget*, 6(27), pp. 24424–24435. doi: 10.18632/oncotarget.4420.

Tang, X. H. *et al.* (2004) 'Oral Cavity and Esophageal Carcinogenesis Modeled in Carcinogen-Treated Mice', *Clinical Cancer Research*, 10(1 I), pp. 301–313. doi: 10.1158/1078-0432.CCR-0999-3.

Thiery, J. P. (2002) 'Epithelial–mesenchymal transitions in tumour progression', *Nature Reviews Cancer*, 2(6), pp. 442–454. doi: 10.1038/nrc822.

Torre, L. A. *et al.* (2015) 'Global cancer statistics, 2012', *A Cancer Journal for Clinicians*, 65(2), pp. 87–108. doi: 10.3322/caac.21262.

Vettore, A. L. *et al.* (2015) 'Mutational landscapes of tongue carcinoma reveal recurrent mutations in genes of therapeutic and prognostic relevance', *Genome Medicine*. l, 7(1), p. 98. doi: 10.1186/s13073-015-0219-2.

Vitale-Cross, L. *et al.* (2009) 'Chemical carcinogenesis models for evaluating molecular-targeted

prevention and treatment of oral cancer', *Cancer Prevention Research*, 2(5), pp. 419–422. doi: 10.1158/1940-6207.CAPR-09-0058.

Walter, V. *et al.* (2013) 'Molecular Subtypes in Head and Neck Cancer Exhibit Distinct Patterns of Chromosomal Gain and Loss of Canonical Cancer Genes', *PLoS ONE*, 8(2). doi: 10.1371/journal.pone.0056823.

Wellcome Sanger Institute (no date) *Signatures of Mutational Processes in Human Cancer*. Available at: https://cancer.sanger.ac.uk/cosmic/signatures.

Whiteside, T. L. (2018) 'FOXP3+ Treg as a therapeutic target for promoting anti-tumor immunity', *Expert Opinion on Therapeutic Targets*, 22(4), pp. 353–363. doi: 10.1080/14728222.2018.1451514.

WHO (2014) 'Locally advanced squamous carcinoma of the head and neck', *Review of Cancer Medicines on the WHO List of Essential Medicines*, pp. 1–8. doi: 10.1093/annonc/mdl274.

Xing, Y. and Hogquist, K. (2014) 'Isolation, identification, and purification of murine thymic epithelial cells.', *Journal of Visualized Experiments*, (90), p. e51780. doi: 10.3791/51780.