University of BRISTOL

University Hospitals Bristol NHS
NHS Foundation Trust

headandneck 5000

Genetic and Epigenetic Data as a Tool to Understand Oropharyngeal Cancer Incidence and Progression

Principle Applicant: Ryan Langdon

Co-Applicants: Caroline Relton, Richard Martin, Andy Ness, Steve Thomas, George Davey Smith, Rebecca Richmond

Scientific Outline

Introduction

Head and neck cancer (HNC) is the 6th most common cancer in the world (1). The primary risk factors identified for HNC incidence are exposure to tobacco and alcohol, and more recently, human papilloma virus (HPV) infection (in particular, HPV-16 and HPV-18) (2-4).

Public health efforts (e.g. progressive rises in cigarette price through excise tax increases) have reduced the global smoking prevalence from 11% to 7%, and 44% to 36% in adult females and males respectively from 2000-2012 (5-7). Coupled with a general decline in tobacco use over the past two decades, these efforts seem to have significantly reduced the incidence of cancers of the oral cavity, larynx, and hypopharynx (8). It was expected that a decrease in smoking rates would also reduce incidence of cancers of the oropharynx, but interestingly incidence merely plateaued before sharply increasing again; a pattern that is independent of an individual's smoking and drinking status (9).

As mentioned above, infection with HPV has been associated with HNC cancer aetiology; particularly with oropharyngeal cancer (OC) (4). When stratified by HPV infection status, two molecularly and epidemiologically distinct types of OC exist (10). HPV-negative OC is primarily associated with exposure to tobacco and alcohol products leading to development of malignancy. However, HPV-positive OC develops following exposure to HPV, independently of exposure to other risk factors (4, 10).

The prevalence of HPV-associated OC has been increasing at epidemic proportions in multiple populations including the United States, Western Europe, and Australia (11, 12). This increase in incidence of OC correlates with an increase in HPV infection, resulting from alterations in sexual practices (e.g. increased number of sexual partners, and increased orogenital sexual activity) over the past 40 years (13). It is generally accepted that the observed increase in OC incidence is due to that of HPV-positive OC, while HPV-negative OC is actually declining in incidence (perhaps explaining the aforementioned plateau) (14).

Interestingly, HPV-positivity in OC (despite being a risk factor for HNC incidence) is associated with improved prognosis over those that are HPV-negative (15, 16), due in part to a more favourable response to chemotherapy (17). There is otherwise very little literature investigating modifiable risk factors for any HNC progression. Prognosis for HNC is currently determined by anatomical site, stage and HPV status, with late stage HNC currently limited to a complete response of 50% in contrast to an early stage diagnosis with 80% survival (18). This underscores the need to discover molecular markers with the potential to unravel the pathogenesis of HNC; particularly OC.

We are currently in an era where a vast amount of genetic and epigenetic data are publically available, as well as a multitude of bioinformatic techniques to curate these. These data can be used (in both a hypothesis-generating, and subsequent hypothesis-driven manner) to improve elucidation and understanding of the causal pathways of OC progression.

Genetic variants can be associated with a phenotype, and analogous to arms of a randomized control trial (RCT), are largely independent of confounding factors due to the random nature of their allocation within a population (19). They are also not modified by the later development of disease or health outcome and, with the increasing accuracy of genotyping arrays, display very little measurement error. As such, when used in instrumental variable analyses such as Mendelian randomization (MR), a genetic variant proxying for a phenotype is free of the limitations that would otherwise weaken causal inference in observational studies (i.e. direct measurement of the phenotype) (20).

Epigenetic signatures are another resource proving extremely valuable in public health research, particularly when understanding potential mechanisms through which an exposure causes an outcome, and for objectively assessing the phenotype of an individual (21-23). Environmental factors (i.e. tobacco smoking) may have long-lasting effects on DNA methylation patterns, which can alter gene expression, and thus the development or progression of a disease (e.g. OC). This knowledge can be used to construct transferrable methylation “predictors” of a phenotype as an objective alternative to questionnaire responses, which may not capture true exposure to a risk factor, and are prone to misreporting.

Using a combination of these hypothesis-generating methodology, including genome-wide and epigenome-wide association studies (GWAS and EWAS, respectively), comprehensive literature text mining, epigenetic phenotype predictors, and MR analyses (hypothesis-free (24), two-sample (25), and two-step-two-sample MR (26)), this project will aim to establish and appraise robust causal pathways associated with OC progression.

Sample and data requirements

Epigenetic (genome-wide DNA methylation) profiling is currently being performed using the Illumina Infinium HumanMethylationEPIC array, on blood samples from a number of HN5000 participants who

were diagnosed with OC and who have OncoChip genetic data, baseline questionnaire and data capture. We would like to request access to this methylation data along with the available genetic, serology and phenotypic data for these samples. We would also like to access saliva and tissue (tumour) samples from this sub-group of HN5000 participants. We appreciate that tissue access and DNA extraction needs to be agreed and funded; we propose to work with HN5000 to achieve this, and undertake pilot feasibility work to establish optimal DNA extraction protocols in the first instance. We will seek to identify funding through the Integrative Cancer Epidemiology Programme (ICEP) to then fund HumanMethylationEPIC beadchip data generation on these samples.

Methods

We will first conduct EWAS for cancer stage and prognosis (recurrence, mortality and survival) in OC using DNA methylation data obtained from peripheral blood samples. We will then quantitatively assess, using regression models and ROC curve analyses, the ability of identified methylation biomarkers to predict progression of OC.

We will next identify methylation signatures related to risk factors implicated in altering OC progression and we will evaluate the utility of methylation signatures as biomarkers of exposure by generating “scores” that index specific exposures. We have been developing statistical methods to construct methylation scores which act as a proxy for historical exposures (e.g. smoking). We propose to extend this work to alcohol using data from other sources, but importantly need to test the performance of this exposure predictor in HN5000. This work compliments and informs the proposed application of the constructed methylation scores by Rhona Beynon in using them to predict HNC progression.

In order to link epigenetic biomarkers with both exposure history and cancer stage and progression, we will conduct “meet-in-the-middle” style approaches to aid the identification of these intermediate biomarkers (27). We will also apply MR to discriminate between methylation marks that are causally implicated with cancer prognosis, and hence are possible targets for intervention, from non-causal biomarkers which may nonetheless have predictive utility.

When the data are available, we will seek to identify cross-tissue and tissue-specific methylation marks related to prognosis.

We will also assess the clinical implications of our findings in terms in terms of whether causal effects identified are clinically meaningful and whether the biomarkers identified offer utility as exposure and prognosis predictors.

Summary

This project falls under the scope of a previous application written by Dr Rebecca Richmond (“Identifying epigenetic biomarkers in cancers of the head and neck”), and is part of a PhD proposal which sits within the University of Bristol’s Integrative Cancer Epidemiology Programme (ICEP). ICEP is funded by CRUK, aiming to reduce the burden of cancer through the identification of causal risk factors, mechanistic targets and predictive biomarkers. Members of the Head and Neck 5000 cohort are currently being epigenetically profiled (DNA methylation) using the Illumina MethylationEPIC array; providing the greatest coverage of its type. The wider PhD proposal this project falls under aims to appraise novel and existing causal risk factors and their pathways in head and neck cancer incidence and progression using genetic and epigenetic data. Epigenetic variation has been linked to head and neck cancer risk factors and prognosis and may therefore serve as both a predictive biomarker and a

potential target for intervention. Using novel, robust bioinformatic analyses such as epigenome-wide association studies and Mendelian randomization, DNA methylation will be interrogated as a biomarker for risk factor exposure and cancer stage, and ultimately as a potential intermediate between a genetic risk factor and head and neck cancer.

Keywords

Epigenetic, methylation, biomarker, prognostic, causal, oropharyngeal cancer