THE GENETIC AETIOLOGY OF OESOPHAGEAL CANCER IN THE SOUTH AFRICAN BLACK POPULATION

Wenlong Carl Chen

ABSTRACT

Oesophageal squamous cell carcinoma (OSCC) is a common disease in eastern and southern Africa with a complex aetiology. Little is known about the contribution of genetic factors to OSCC risk in African populations. Recent genome-wide association studies (GWAS) in Asian and European populations identified multiple genetic risk loci for OSCC, but few findings from other populations have been replicated in African OSCC. This thesis aims to identify genetic risk factors for OSCC in the black South African population.

High molecular weight DNA of good quality and quantity is needed to conduct large-scale genetic studies. We had access to a large resource of frozen blood samples from OSCC cases and controls collected by the Johannesburg Cancer Study (JCS). The integrity and yield of genomic DNA isolated from 2,758 whole blood following long-term storage at -30°C were assessed using spectrophotometry, fluorometric assay for double-stranded DNA and agarose gel electrophoresis. The mean DNA yield per sample was 114 µg from whole blood volumes of 0.5 ml to 4 ml. The mean A260/280 ratio and median A260/280 ratios were both 1.8. This analysis showed that high-quality DNA can be extracted from whole blood samples that are stored at -30°C for up to 19 years.

The contributions of genetic factors to the risk of African OSCC were investigated for 7 risk loci previously identified in non-African populations. Association studies were performed in a total of 1,471 OSCC cases and 1,791 controls from Black South Africans in two study sample groups from the Western Cape and Gauteng province. Thereafter a meta-analysis was done for 11 variants that had been genotyped in both studies. A single nucleotide polymorphism (SNP) in the *CHEK2* gene, rs1033667, was significantly associated with OSCC (P = 0.002; OR = 1.176; 95% CI: 1.06 - 1.30). However, SNPs in the

CASP8/ALS2CR12, *TMEM173*, *PLCE1*, *ALDH2*, *ATP1B2/TP53* and *RUNX1* loci were not associated with the disease. The lack of association of these six loci with OSCC in South African populations may reflect different genetic risk factors in non-African and African populations or differences in the genetic architecture of African genomes. The association of the SNP in the *CHEK2* gene provided support for the contribution of a common genetic variant at this locus to the risk of OSCC.

A genome-wide association study (GWAS) was then conducted in 1,686 African OSCC patients and 3,217 controls to broaden the search for genetic risk factors. Samples were genotyped using the H3A custom microarray containing ~2.3 million genetic markers enriched for African content. One novel, genome-wide significant risk locus was identified on chromosome 9, upstream of the *FAM120A* gene (rs12379660, $P = 4.58x10^{-8}$, odds ratio = 1.28, 95% confidence interval = 1.22-1.34) and one African-specific risk locus was identified on chromosome 2 (rs142741123, $P = 5.49x10^{-8}$) within the *MYO1B* gene. A meta-analysis was performed for the African OSCC GWAS with a Chinese OSCC GWAS (2,013 OSCC patients, 2,701 controls) and identified three genome-wide significant loci on chromosome 9 at *FAM120A* (rs12379660, $P_{meta} = 9.36x10^{-10}$), chromosome 10 at *PLCE1* (rs7099485, $P_{meta} = 1.48x10^{-8}$) and chromosome 22 at *CHEK2* (rs1033667, $P_{meta} = 1.47x10^{-9}$).

This study, which is the first GWAS done in an African population, provided evidence of a substantial genetic contribution to OSCC risk in Africa and identified both shared and distinct risk loci for OSCC in populations of diverse ancestry. This work is the first indication of the molecular basis of OSCC in Africa and shows that larger studies in African populations are justified to identify more of the genetic contribution to the risk of this disease.