Sample collection

Sample	Sample Collection Details	Timepoint	Initial transport	Processing	Storage	Numbers
Blood	16ml venous blood. Collected in EDTA tubes (1 x 6ml and 1 x 10ml)	Pre treatment	Posted at ambient temperature	Centrifuged at 3500 rpm for 10 minutes. Buffy coat collected and DNA extracted. Plasma stored in 200 µl & 500 µl aliquots.	Samples frozen & stored at −80°C	**All sites: 4,587 (4,630)** Oral cavity: 1,147 Oropharynx: 1,611 Larynx: 924
Saliva	At least 1ml (where possible) collected in sterile screw top container. No fixative added.	Pre treatment	Posted at ambient temperature	Samples divided into 7 x 1 ml tubes.	Samples frozen and stored at −80°C	**All sites: 4,899 (4,321)** Oral cavity: 1,199 Oropharynx: 1,752 Larynx: 968
Tissue	Formalin fixed paraffin embedded blocks	Tissue collected from diagnostic procedure or surgery to remove primary tumour	Posted at ambient temperature	H&E sections were cut and examined to ensure representative tumour samples and confirm diagnosis. Currently scanning slides to create a digital library. Tissue microarray blocks planned for salivary gland tumours.	Ambient temperature	**All sites: 2,518 (2,781)** Oral cavity: 710 Oropharynx: 839 Larynx: 486

Blood assays and measures - 1

Assay/measure	Details of assay	Number	Notes
Human papillomavirus (HPV) serology	HPV antibodies were measured using a glutathione S-transferase multiplex assay carried out at the German Cancer Research Centre (DKFZ) in Heidelberg, Germany. [Waterboer T et al 2005]	All sites =4,538 (4,611) N by sites	For HPV16 and HPV18 the following antibodies were analysed: L1, E1, E2, E4, E6 E7. For HPV31, HPV33, HPV35, HPV45, HPV52 and HPV58 the following antibodies were analysed: L1, E1, E2, E6, E7. This method and seropositivity definition used have been described: • Waterboer T, Sehr P, Michael KM, et al. Multiplex human papillomavirus serology based on in situ-purified glutathione s-transferase fusion proteins. Clin Chem. 2005;51(10):1845-1853. • Schroeder L, Pring M, Ingarfield K, et al. HPV driven squamous cell head and neck cancer of unknown primary is likely to be HPV driven squamous cell oropharyngeal cancer Oral Oncology 2020; 107: 104721. Additionally, antibodies for HPV16 and HPV33 have been analysed at higher dilutions for greater sensitivity.
Genome wide association study (GWAS) batch 1	Genotyped at the Centre for Inherited Disease Research on the Illumina OncoArray. Genotypes were called alongside other cases and controls from the oral and pharynx OncoArray study [Lesseur et al 2016], in GenomeStudio software (Illumina, USA).	All sites: 1,079 Oral cavity: 426 Oropharynx: 596*	Quality control Conducted in PLINK 1.934. Imputation performed using the Michigan Imputation Server: SHAPEIT was used for pre-phasing and Minimac3 for imputation using the HRC reference panel. Details are in: • Lesseur C, Diergaarde B, Olshan AF, et al. Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. Nat Genet. 2016;48(12):1544-1550.

* IARC definitions of oral cavity and oropharynx tumours were used, which differ from the definitions used in HN5000. The availability of GWAS data by tumour site is by HN5000 definition.

Blood assays and measures – 2

Assay/measure	Details of assay	Number	Notes
GWAS batch 2	Genotyped at Bristol Bioresource Laboratory on the Illumina Global Screening Array.		Currently in process. Estimated to be available Q1 2021.
Epigenome wide association study (EWAS	DNA was bisulphite-converted using the Zymo EZ DNA MethylationTM kit (Zymo, Irvine, CA, USA). Genome-wide methylation data were generated using the Infinium MethylationEPIC BeadChips (EPIC array) (Illumina, USA). Arrays were scanned using an Illumina iScan (version 2. 3).	440 408 confirmed OPC	Analyses of these data have been reported: • Langdon RJ, Beynon RA, Ingarfield K, et al. Epigenetic prediction of complex traits and mortality in a cohort of individuals with oropharyngeal cancer. Clin Epigenetics. 2020;12(1):58. • Langdon, R., Richmond, R., Elliott, H.R. et al. Identifying epigenetic biomarkers of established prognostic factors and survival in a clinical cohort of individuals with oropharyngeal cancer. Clin Epigenetics. 2020;12(1):58
Metabolomics	Metabolic profiling performed using a high-throughput serum nuclear magnetic resonance (NMR) metabolomics platform (Nightingale Health®, Helsinki, Finland), originally described by Soininen et al. This was hosted by the Department of Chemistry, University of Bristol.	1,595 analysed, 1,483 confirmed OPC	The platform provides quantification of routine lipids, 14 lipoprotein subclasses, including particle concentration and lipids transported by these particles, various fatty acids and fatty acids traits (e.g., chain length, degree of unsaturation), amino acids, ketone bodies, glycolysis and gluconeogenesis-related metabolites, fluid balance, and one inflammation-related metabolite in molar concentration units. • Soininen P et al. High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *Analyst*. 2009; **134**: 1781-5.
Epstein Barr virus (EBV) serology	IgG antibodies against 14 EBV antigens. Multiplex serology conducted at the German Cancer Research Centre (DKFZ) in Heidelberg, Germany.	4,566	Analysis of the data for nasopharyngeal cancer has been reported: • Simon J, Schroeder L, Ingarfield K, et al. Epstein-Barr virus and human papillomavirus serum antibodies define the viral status of nasopharyngeal carcinoma in a low endemic country. Int J Cancer. 2020;147(2):461-471.

Tissue assays and measures - 1

Assay/measure	Details	Number	Notes
Human papillomavirus (HPV) DNA and RNA	Multiplex Papillomavirus Genotyping was used to analyse DNA for the presence of 51 HPV types. [Schmitt et al 2006] HPV type-specific reverse transcription polymerase chain reaction (RT-PCR) and hybridization assays were used to detect HPV RNA. [Halec G et al 2013] These analyses were conducted at the German Cancer Research Centre (DKFZ) in Heidelberg, Germany.	Oropharynx: 264 Oral cavity: 272 Hypopharynx: 30 Nasopharynx: 27 Nasal cavity: 9 Larynx: 5 Sinuses: 4 Salivary glands: 1	Hematoxylin and eosin stained slides were reviewed to ensure tumour was present. Negative controls were included to monitor cross-contamination. The methods and the results in people with cancer unknown primary and nasopharyngeal cancer have been described: • Schmitt M, Bravo IG, Snijders PJ, Gissmann L, Pawlita M, Waterboer T. Bead-based multiplex genotyping of human papillomaviruses. J Clin Microbiol. 2006 Feb;44(2):504-12 • Halec G, Schmitt M, Dondog B, et al. Biological activity of probable/possible high-risk human papillomavirus types in cervical cancer. Int J Cancer. 2013;132(1):63-71. • Simon J, Schroeder L, Ingarfield K, et al. Epstein-Barr virus and human papillomavirus serum antibodies define the viral status of nasopharyngeal carcinoma in a low endemic country. Int J Cancer. 2020;147(2):461-471. • Schroeder L, Pring M, Ingarfield K, et al. HPV driven squamous cell head and neck cancer of unknown primary is likely to be HPV driven squamous cell oropharyngeal cancer Oral Oncology 2020;107:104721.
P16 and ISH	P16 protein expression (cell cycle inhibitor) as a surrogate marker of HPV infection. Protein expression determined using immunohistochemistry on formalin fixed paraffin embedded tissue sections. High risk Human papilloma virus DNA detection using in-situ hybridisation.		P16 and ISH were recorded in clinical notes where available. There is a subset of 504 cases with P16, ISH, TILS and Survivin and pathologist (calibrated) scoring. For information on techniques/standards recommended in the UK, see: • https://www.nice.org.uk/guidance/ng36/evidence/full-guideline-pdf-2307980269 • https://www.rcpath.org/uploads/assets/6201bef5-79df-4107-ba6a42833377457f/g111_pharynxmucosaldataset_nov13.pdf P16, ISH, TILS and Survivin were conducted on the same participants. Only around 55 participants also have HPV DNA and RNA.

Tissue assays and measures - 2

Assay/measure	Details	Number	Notes
Tumour infiltrating lymphocytes (TILs)	Assessment and scoring of Tumour infiltrating lymphocytes (TILs) from haematoxylin and eosin stained formal fixed paraffin embedded tissue sections.		Analysis in process. P16, ISH, TILS and Survivin were conducted on the same participants. Only around 55 participants also have HPV DNA and RNA. • Ward MJ, Thirdborough SM, Mellows T, et al. Tumour-infiltrating lymphocytes predict for outcome in HPV-positive oropharyngeal cancer. Br J Cancer. 2014;110(2):489-500.
Survivin	Assessment of Survivin protein expression (Bio SB Survivin RMab Clone: EP119) using immunohistochemistry on formal fixed paraffin embedded tissue sections		Calibrated pathologists. Scoring in process. P16, ISH, TILS and Survivin were conducted on the same participants. Only around 55 participants also have HPV DNA and RNA. • Preuss SF, Weinell A, Molitor M, et al. Nuclear survivin expression is associated with HPV-independent carcinogenesis and is an indicator of poor prognosis in oropharyngeal cancer. Br J Cancer. 2008;98(3):627-632.
Epstein Barr virus encoded ribonucleic acid in situ hybridisation (EBER-ISH)	Detection of Epstein Barr virus using Epstein Barr encoding region (EBER) in situ hybridisation from formal fixed paraffin embedded tissue sections		Nasopharyngeal cancers only. Data from pathology reports and some data based on analysis by UKAS accredited NHS pathology laboratory (Severn pathology). • https://www.ukas.com/services/accreditation-services/medical-laboratory-accreditation-iso-15189/ For more information and the data reported in H&N5000 see: • https://www.rcpath.org/uploads/assets/6201bef5-79df-4107-ba6a42833377457f/g111_pharynxmucosaldataset_nov13.pdf • Nakao K, Mochiki M, Nibu K, Sugasawa M, Uozaki H. Analysis of prognostic factors of nasopharyngeal carcinoma: impact of in situ hybridization for Epstein-Barr virus encoded small RNA 1. Otolaryngol Head Neck Surg. 2006;134(4):639-645. • Simon J, Schroeder L, Ingarfield K, et al. Epstein-Barr virus and human papillomavirus serum antibodies define the viral status of nasopharyngeal carcinoma in a low endemic country. Int J Cancer. 2020;147(2):461-471.

Tissue assays and measures – 3

Somatic mutations	Whole exome and targeted sequencing are being carried by Dr Neil Hayes at University of Tennessee, Memphis, USA.	Oropharynx: 264 Oral cavity: 272 Hypopharynx: 30 Nasopharynx: 27 Nasal cavity: 9 Larynx: 5 Sinuses: 4 Salivary glands: 1	Sample selection criteria included OPC (including all OPC that had an atypical seropositive response), OC (including all OC that were HPV seropositive), all larynx cancers that were HPV seropositive, all SCC at other sites. Analysis is ongoing. Hybrid capture library preparation takes tumour DNA and fragments it into roughly 250 bp fragments that are then baited with targeted DNA probes. The probes target approximately 800 genes as well as HPV 16 and 18 regions and germline mutations in clinically relevant genes and drug metabolizing enzymes. These are then sequenced with paired end Illumina brand sequencing and perform variant calling on tumors. Some references to the methods being used by Dr Hayes can be found in the following references: PMID 30102605; PMID 29158372; PMID 27083775; PMID 26371432; PMID: 26076459 Further OC and larynx samples may be provided for genotyping by Dr Hayes' laboratory. CUP tissue samples (approximately 40) will be included in a Cancer Genome Atlas consortium led by Manel Esteller, Cancer Epigenetics and Biology programme, Barcelona.