GeCIP Detailed Research Plan Form

Background

The Genomics England Clinical Interpretation Partnership (GeCIP) brings together researchers, clinicians and trainees from both academia and the NHS to analyse, refine and make new discoveries from the data from the 100,000 Genomes Project.

The aims of the partnerships are:

1. To optimise:

- clinical data and sample collection
- clinical reporting
- data validation and interpretation.

2. To improve understanding of the implications of genomic findings and improve the accuracy and reliability of information fed back to patients. To add to knowledge of the genetic basis of disease.

3. To provide a sustainable thriving training environment.

The initial wave of GeCIP domains were announced in June 2015 following a first round of applications in January 2015. Shortly after this, the GeCIP domains produced detailed research plans after working closely with Genomics England. The research plans will be used to ensure that the various domains are complimentary and add real value across the GeCIP portfolio and address the aims and objectives of the 100,000 Genomes Project. They will be shared with the MRC, Wellcome Trust, NIHR and Cancer Research UK as existing members of the GeCIP Board to give advance warning and manage funding requests to maximise the funds available to each domain. However, formal applications will then be needed by individual funders. The research plans will allow Genomics England to plan shared core analyses and the required research and computing infrastructure to support the proposed research. They will also form the basis of assessment by the Project's Access Review Committee, to permit access to data.

Genomics England Clinical Interpretation Partnership (GeCIP) Detailed Research Plan Form

Application Summary	
GeCIP domain name	**Head and Neck Cancer**
Project title *(max 150 characters)*	**Head and Neck Cancer Research in the 100,000 Genomes Project**

Objectives. *Set out the key objectives of your research. (max 200 words)*

GeCIP Aims:

The stated aims of the Head and Neck GeCIP, using prospectively collected tissue, defined retrospective tissue cohorts and samples collected as part of clinical trial protocols, include:

1. The development of a personalised therapy approach using WGS data to provide molecular mutation profiles for patients which may inform treatment selection and prognostication
2. To identify novel therapeutic targets, facilitate drug repurposing and define clinically informative biomarkers of risk stratification and treatment response.
3. In collaboration with national and international experts, to further understand the biological basis of head and neck tumour development and behaviour.
4. To utilise the H&N Domain infrastructure to facilitate the successful transition of molecular profiling from the research domain into routine NHS clinical care.
5. To train the next generation of clinical academic and scientific leaders in the field of `omics` based precision medicine for the management of head and neck cancer.

Lay summary. *Information from this summary may be displayed on a public facing website. Provide a brief lay summary of your planned research. (max 200 words)*

Head and neck cancers comprise a mixed group of tumours derived from the lining of the mouth and throat and the thyroid and salivary glands. Whilst the majority are related to cigarette smoking and alcohol use, a significant minority have no definable cause or result from viral infection, particularly Human papillomavirus (HPV) or Epstein Barr Virus (EBV). Over recent decades the incidence of throat cancer caused by HPV (HPV+) has increased markedly. Why this has happened and how HPV causes throat cancer is unclear. HPV+ cancer responds better to treatment than cancer not caused by HPV (HPV-) and so the emergence of HPV+ disease has created a treatment paradox: Improved survival is still necessary for HPV- disease whilst reduction of treatment related side-effects is an urgent aim for HPV+ cases. HPV+/- discrimination is insufficient to personalise treatment and better ways to define risk are required.
We intend to use Whole Genome Sequencing to better understand the differences between HPV+ and HPV– cancers and to define molecular profiles which will allow us to personalise treatment, as well as identify molecular targets for treatment with new drugs or existing drugs which are not normally used for head and neck cancer patients.

Technical summary. *Information from this summary may be displayed on a public facing website. Please include plans for methodology, including experimental design and expected outputs of the research. (max 500 words)*

This research plan will be delivered by a GeCIP membership that includes national and international experts in research and the clinical management of head and neck cancer.

The group will include: NHS clinical oncologists and head neck surgeons, geneticists, genetic counsellors, clinical scientists and patient and public representatives.

Our proposed GeCIP structure is explained in detail below. Each Subdomain will include relevant experts who will define clinically meaningful research questions, which will be addressed in collaboration with the highly experienced Data Analysis Core Team, which will act as a crosscutting data-analysis Subdomain. Collaborative links with other GeCIP Domains, industry and research funders have already been established and will further increase our chances of successfully achieving our GeCIP aims.

Using >300 samples of squamous cell carcinoma of the head and neck (SCCHN) collected prospectively as part of the 110k Genome project; a cohort (n ~80) of oral cavity cancer samples collected since 1 Jan 2015 and associated with a rich clinical dataset including follow-up data and tissue samples collected as part of two ongoing randomised controlled clinical trials (NICO, BMS funded & PATHOS, CR-UK funded), we intend to

- Define molecular profiles of HPV+ve and HPV-ve SCCHN to predict risk and inform de-intensification and intensification treatment strategies.
- Understand further the molecular mechanisms of HPV-driven SCCHN
- Identify novel therapies for SCCHN
- Improve the classification of thyroid & parathyroid cancers
- Understand the molecular aetiology of sinonasal tumours

In order to realise our aims, the following core analysis strategies will be employed

Somatic and germline variants will be called using the GEL pipeline; annotation would be with VEP; Sift, polyphen, ExAc and CADD would be used for the assessment of deleteriousness of variants with COSMIC for additional information on somatic variants; Canvas and Manta will be used for CNVs and SVs, respectively; mutation signatures will be called with R package deconstructSigs (https://cran.r-project.org) and with the COSMIC Mutation Signatures Consensus list and viral genome analysis will be with a bespoke pipeline which would be imported into the embassy.
We will also analyse somatic variants in non-coding regions of the tumour genomes with a particular emphasis on recurrent mutations (using MutSigCV to define the statistical significance of these) and mutations in promoter regions, UTRs, transcription factor binding sites and other non-coding elements defined by the ENCODE and GENCODE projects.
Furthermore, in an attempt to enhance the accuracy of our aims as outlined above, we will endeavour to liaise with expert groups in Artificial Intelligence in order to exploit further the potential clinical utility of our WGS and other `omics data.
(www.cancerresearchuk.org/funding-for-researchers/how-we-deliver-research/grand-challenge-award/artificial-intelligence#details60)

In addition, training the next generation of clinician and surgeon scientists to deliver on the existing challenges of head and neck surgery and oncology through this innovative programme is a high GeCIP priority. Researchers, students and clinicians will be encouraged to apply for funding to take up postdoc, MD/PhD positions and MSc and MRes programmes in genetic medicine, genetic counselling, and head and neck oncology. We will also seek funding to

provide short courses tailored to the development of "crossover" skills and knowledge which we have found necessary for multidisciplinary research e.g. basic head and neck anatomy for molecular scientists; fundamentals of bioinformatics databases for clinicians.

Expected start date	**Q4 2018**
Expected end date	**Q4 2020**

Lead Applicant(s)	
Name	Terry M Jones (Lead Applicant), Liam Masterson, Matt Lechner, Jenny Taylor (Co-Leads)
Post	Professor of Head and Neck Surgery, Hon. Lecturer, Hon. Lecturer, Associate Professor
Department	Molecular & Clinical Cancer Medicine, Dept. of Pathology, Cancer Institute, Wellcome Centre for Human Genetics
Institution	University of Liverpool, University of Cambridge, University College London, University of Oxford
Current commercial links	nil

Subdomain leads		
Name	**Subdomain**	**Institution**
T.M. Jones / Mererid Evans	Head and Neck SCC (including clinical trial cohorts)	University of Liverpool
M. Lechner/ Valerie Lund	Sinonasal Cancer	University College London
R. Metcalf / Mark McGurk	Salivary Gland Cancer	University of Manchester
L. Masterson / Dae Kim	Thyroid Cancer	University of Cambridge / Institute of Cancer Research
R. Hewitt	Paediatric Head and Neck Cancer	

The lead applicants are fully supported by the Data Analysis Core Team (in alphabetical order):

Tim Fenton (University of Kent), Trevor Graham (Barts Cancer Institute), N. David Hayes (University of North Carolina), Nischalan Pillay (UCL Cancer Institute), Andrew Schache (University of Liverpool), Jenny Taylor (University of Oxford).

This team will develop and apply the most effective pipeline for the analysis of WGS data on both cancer DNA and ctDNA, when available, working closely with the expert panels for each cancer subdomain. Consequently, the expert panels are fully supported by the Data Analysis Core team. Collaboration is key. Anyone with a specific interest in Head and Neck cancer is welcome to approach the Head and Neck Cancer Domain GeCIP team to join the existing expert panels.

Detailed research plan

Full proposal (total max 1500 words per subdomain)	
Title *(max 150 characters)*	**Head and Neck Cancer Research in the 100,000 Genomes Project**

Importance. *Explain the need for research in this area, and the rationale for the research planned. Give sufficient details of other past and current research to show that the aims are scientifically justified. Please refer to the 100,000 Genomes Project acceptable use(s) that apply to the proposal (page 6).*

1. **Squamous cell carcinomas of the head and neck (SCCHN)**, which arise from the mucosal surfaces of the upper aerodigestive tract, are the most common cancers of the head and neck region constituting the 6th most common cancer worldwide (~10,000 cases per annum in the UK).[1] Contemporary curative treatments include surgery, radiotherapy +/- cisplatin or combinations thereof, although anti-EGFR drugs and immunotherapeutic agents are licenced for specific clinical indications including the recurrent metastatic setting.[2,18] SCCHN exhibits variable responses to conventional treatment. Clinical response rates correlate with survival and are inversely related to primary tumour size, presence and volume of metastatic disease in the cervical lymph nodes and pathological evidence of tumour spread through the lymph node capsule (Extracapsular spread - ECS). Variability in treatment responsiveness has increased over recent decades with the emergence of a new discrete disease entity – Human papillomavirus associated oropharyngeal squamous cell carcinoma (HPV+ OPC) the incidence of which has doubled in the UK, US and Northern Europe. It occurs in patients who are younger, fitter, drink less alcohol and smoke less than patients with HPV-negative cancer.[3] Despite presenting with clinico-pathological features suggestive of aggressive tumour behaviour and poor outcome, survival rates are much higher for patients with HPV+ OPC although a significant minority (15-20%) will succumb to their disease.[4] Current epidemiological data suggest that poor outcome correlates tightly with cigarette smoking which is likely a surrogate of the underlying carcinogen-induced mutational load and genetic instability. However, mainstay cisplatin-based chemoradiotherapy (CRT) results in life-changing long-term swallowing disability: up to 20% of patients undergoing CRT require long-term gastrostomy tube feeding.[2] Consequently, there is an urgent need to identify patients who are destined for poor outcome and those for whom treatment de-intensification, with a view to avoiding long term swallowing difficulty, is an option. In contrast, despite advances in surgical and non-surgical treatments, five-year survival rates for HPV- SCCHN remain at 60%.[4] Therefore, there is an urgent need to develop treatments that enhance survival: minimising adverse effects, whilst important, remains a secondary consideration for many patients. Stratification based on HPV status in the context of oropharynx cancer is the only strategy in current clinical practice. However, whilst data confirm HPV status utility for prognostication, no data currently exist to suggest that treatment decision-making based on HPV status is safe and effective.[5] Moreover, little is known about the molecular underpinning of HPV in the development and progression of HPV+ OPC with much currently inferred from data derived from cervix cancer research: Whether such inference is appropriate and relevant is currently unclear.
   It follows that there is an urgent need to define molecular profiles which determine risk and have utility as predictive biomarkers of outcome as well as novel therapeutic targets for new drugs and the repurposing of existing drugs which are not currently used in the treatment of SCCHN.[6]

To date, data from three large genome-wide sequencing studies of HNSCCs are available.[7-9] The results from sequencing data have been disappointing from a therapeutic target standpoint because of relativity paucity of oncogene mutations in comparison to the more-frequent tumour suppressor gene (TSG) mutations.

TSGs are poor therapeutic targets because restoring loss-of-function in these genes is more difficult than inhibiting increased activity resulting from gain-of-function in oncogenes. Of the 15 most-frequent mutations in HNSCC, only two are known oncogenes, *PIK3CA* and *HRAS*, and only *PIK3CA* is currently considered to be therapeutically targetable.[10]

HPV-positive tumours have a lower average number of mutations per tumour, and rarely have *TP53* mutation and loss of p16INK4A function compared with HPV-negative tumours, reflecting the biological differences in these tumours.[7,8]

HNSCC has an immunosuppressive influence with patients demonstrating low absolute lymphocyte counts, decreased antigen-presenting function and elevated T-regulatory cells.[11] HNSCC achieves immune evasion by down regulating expression of the antigen-processing molecules, TAP1/2 and MHC1. In addition, co-inhibitory receptors, cytotoxic T-lymphocyte antigen 4 (CTLA-4) and programmed death ligand-1 (PD-L1) which induce immune tolerance to HNSCC, are frequently expressed on tumours.[12] The cytokine microenvironment promotes tumorigenesis with excessive immunosuppressive cytokines such as VEGF, IL-6, TGF-β, and IL-10. 18 The success of immune-checkpoint inhibitors in solid tumours, along with the increases in HPV-positive HNSCC incidence, has raised enthusiasm for novel immunotherapeutic approaches and development of corresponding biomarkers.[13]

2. **Differentiated Thyroid Carcinoma & Parathyroid Carcinoma:** Since the early 1990s, thyroid cancer incidence rates have increased by almost three-quarters in the UK. This trend is set to continue with a current incidence of 11 cases per 100,000 people which equates to >3000 cases per year in England.[14] Thyroid cancer is the most frequent endocrine cancer and contributes to 1-2% of all malignancies diagnosed each year. Thyroid nodules have a reported prevalence varying from 3% to 68% (depending on the evaluated population and the applied screening method), while carcinoma occur in less than 1% of nodules.[15] Most thyroid carcinomas originate from follicular epithelial cells and include well-differentiated papillary thyroid carcinomas (~80% of all thyroid malignancies) and follicular thyroid carcinomas (~15% of thyroid malignancies). Moreover, anaplastic (undifferentiated) carcinoma (1-2% of thyroid cancers) is the most aggressive thyroid malignancy to develop from epithelial cells. Medullary thyroid carcinoma in contrast, is derived from parafollicular C-cells and has neuroendocrine features (3-5% of thyroid cancers). Most thyroid tumours are diagnosed initially by morphological assessment of cells procured following fine needle aspiration cytology (FNAC) of a thyroid nodule. The resulting classification (Thy1-5) informs selection of a treatment and prognosis.[16] However, the final diagnosis is based on histopathological examination of thyroid tissue after surgical removal. It is often the case that cytological and histological patterns are ambiguous and definitive classification is problematic. Certain thyroid tumours can be challenging to diagnose even on histologic specimens due to a follicular growth pattern that includes a broad range of lesions that are difficult to distinguish. These lesions include benign follicular thyroid adenomas, hyperplastic nodules, follicular carcinomas, medullary thyroid carcinomas and follicular variant of papillary thyroid carcinomas.[16] Papillary thyroid carcinoma is now the most over diagnosed thyroid lesion, since it shares multiple

features with benign lesions (e.g. Hashimoto's thyroiditis, nodular hyperplasia) as well as with other malignant lesions (e.g. follicular / medullary carcinoma. Hence, additional tests are essential to classify thyroid tumours that exhibit non-standard morphological patterns. Molecular diagnostic tools have only a minimal role in the routine assessment of thyroid cancers. At present, no specific molecular biomarkers are used in current clinical practice which would allow discrimination from other differentiated thyroid cancers, although in most cases the expression of calcitonin could reliably differentiate medullary from follicular tumours. Although some possible biomarkers were proposed to confirm the diagnosis of PTC, they are not routinely used in clinical practice.[17] Early diagnosis is a challenge and recent research has shed light on various genetic abnormalities associated with papillary thyroid carcinoma (PTC) mutations in the MAPK pathway involving RET/PTC and BRAF. A possible link with familial adenomatous polyposis (FAP) has also been suggested.[18] As yet, none of the biomarkers isolated have been positively validated, hence histologic evaluation remains the gold standard to distinguish between these lesions.

New biomarkers identified with the use of high-throughput genomic investigation could support diagnosis of thyroid cancers using core biopsy techniques and thus may avoid the traditional diagnostic hemi-thyroidectomy approach. Not only will whole-genome sequencing of thyroid cancer and parathyroid cancer tissue significantly contribute to the understanding of the underlying cancer biology, but it will also prepare the participating centres for the introduction of routine diagnostic pathways in the future.

3. **Sinonasal Cancers, incl. Olfactory Neuroblastoma, Paediatric Head and Neck Tumours & Salivary Gland Cancers:**

   Tumours in this group are extremely rare and very little, if anything, is known about their molecular biology. The 100k genome project provides a perfect opportunity to start to fill this void by sequencing a small series of these tumours. No such datasets currently exist in the world to date.

   Olfactory neuroblastoma is a rare, aggressive tumour of the sinonasal region originating from olfactory neuroepithelium. Its incidence is estimated to be 0.5-1/1,000,000 population per year and, its incidence peaks in the second and fifth decades of life.[19] Its aetiology is unknown. An infectious cause has been suggested because the tumour contains viral particles,[20,21] but inadequate data are available to either confirm or refute a viral cause. Our proposed WGS analysis of olfactory neuroblastoma will shed light on this; non-human reads will be mapped to both known or predicted RNA and DNA viruses to detect novel viral contributors. Paediatric patients appear to have a more aggressive presentation than adults with a larger proportion of cases with advanced disease.

   Salivary gland tumours represent 1—4% of all human neoplasms and fewer than 5% occur in children and adolescents. In adults, 15—25% of all epithelial salivary gland tumours are malignant, however, this increase to 50—60% in children and adolescents.[15] Variations in protein expression and genetic alterations in tumoural cells has been shown in some paediatric salivary gland tumors, but further studies are necessary to evaluate their clinical significance.

4. **National Genomic Test Directory for Cancer Version 1: Head and Neck Cancer**

Head & Neck Squamous Cell Carcinoma	Prognostic / Predictive	n/a	Small variant dete	*CDKN2A, EGFR, TP53*	Panel	Core	Core
Head & Neck Squamous Cell Carcinoma	Prognostic / Predictive	n/a	Structural variant	*RET*	Panel	Core	Core
Head & Neck Squamous Cell Carcinoma	Prognostic / Predictive	n/a	Copy number vari	*CDKN2A, TP53*	Panel	Core	Core
Head & Neck Squamous Cell Carcinoma	Prognostic / Predictive	n/a	Structural variant	*RET*	FISH	Core	Core
Head & Neck Squamous Cell Carcinoma	Prognostic / Predictive	n/a	Copy number vari	*TP53*	FISH	Core	Core
Adenoid Cystic Carcinoma	Prognostic / Predictive	n/a	Structural variant	*MYB-NFIB*	FISH/Simple targeted muta	Core	Core
Secretory Carcinoma (Salivary Gland)	Prognostic / Predictive	n/a	Structural variant	*ETV6-NTRK3*	FISH/Simple targeted muta	Core	Core

Following an initial application, it has been agreed, subject to locally commissioned funding arrangements, that the above molecular markers can be tested in the context of routine clinical care. Our experience would suggest that the clinical utility of these markers has yet to be firmly established and so it is an important aim of this research plan to fill this void.

Research plans. *Give details of the analyses and experimental approaches, study designs and techniques that will be used and timelines for your analysis. Describe the major challenges of the research and the steps required to mitigate these.*

Head and Neck Cancer Tissue Resources:

1. **100K Prospective Tissue Cohort:**

The Head and Neck Cancer Domain within the 100K Genomes project was established in 2017 and now actively recruits at multiple GMCs across the country.
With the support of the local clinicians, robust patient identification and sample collection pathways have been implemented in the participating centres and ~300 samples have been processed for downstream analysis.

2. **Retrospective Banked Tissue Cohorts:**

Mersey Head and Neck Oncology Group (MHNORG), the largest integrated head and neck research group in the UK, have prospectively been collecting fresh frozen tumour tissue and matched blood samples from patients undergoing surgery since ~2005. Taking advantage of the opportunity to include appropriate matched tumour & blood samples collected since 1st Jan 2015, we have secured permission for WGS of a cohort (n ~80) of oral cavity cancer patients. This cohort is associated with a rich clinical dataset including follow-up data

A further cohort of olfactory neuroblastoma samples, collected under UCL REC 04/0099 is available. With the permission from Genomics England we have consented patients for the 100K project and retrospectively included fresh frozen and blood samples following QC.

3. **Clinical Trial Cohorts*:**

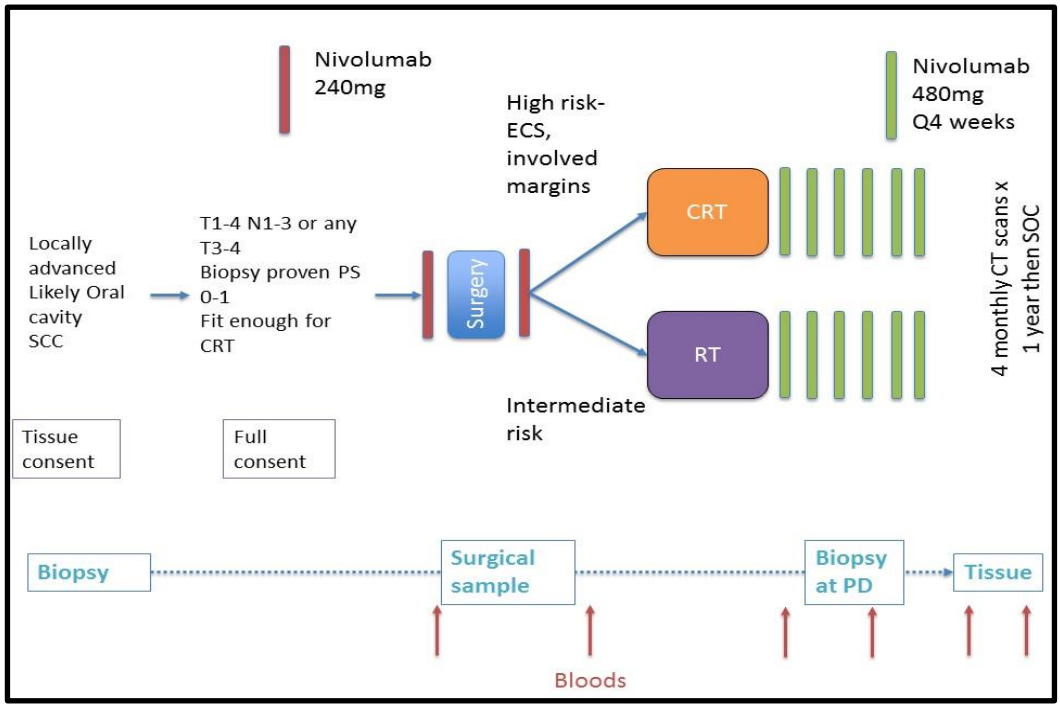
*Following discussions with Professor Mark Caulfield, Clinical Director of Genomics England, an application has been submitted to the Life Science Initiative to secure matched funding for the sequencing of the relevant bio-resource samples collected as part of the protocols of the following two clinical trials. Should funding for WGS be secured, these data will be available as part of the 100K Genomes Project Dataset, we have therefore included their analysis in this research proposal.

NICO:

The NICO study is a two-arm, non-randomised study of neoadjuvant and adjuvant nivolumab in locally advanced oral cavity cancer. Patients will be enrolled prior to surgery and will be

treated with a single dose of nivolumab prior to surgery, followed by a further dose prior to radiotherapy (with or without chemotherapy, depending on pathological risk factors). The patients will be treated for a further 6 months with nivolumab after radiotherapy. The study will recruit 120 patients, of whom approximately 40 are expected to be in the high-risk group and who will receive chemoradiotherapy. The study has co-primary endpoints; 1 year PFS in the high-risk group and feasibility of the treatment strategy.

NICO will enable the collection of invaluable samples from patients, including biopsies prior to enrolment, surgical samples post treatment with nivolumab, and biopsy samples at progression. This will allow investigation into early changes in lymphocyte populations in response to nivolumab, as well as investigation into potential markers predictive of outcome. We will plan to perform RNAseq on samples before and after nivolumab treatment. The availability of sequence data would add significantly to the power of the study, enabling assessment of the mutational load and potentially the mutanome, and their correlation with outcome.



PATHOS: (**P**ost-operative **A**djuvant **T**reatment for **H**PV associated **O**ropharyngeal **S**quamous Cell Carcinoma) is a CR-UK funded Phase II/III trial of risk-stratified, reduced intensity adjuvant treatment in patients undergoing transoral surgery for Human papillomavirus (HPV)-positive oropharyngeal cancer. All patients will receive post-operative intensity modulated radiotherapy (IMRT) as part of two post-surgery randomisations: Patients assigned to a pathological intermediate risk group (R0(intermediate risk)) surgical margins and no extracapsular spread (ECS)) will be randomised to receive 50 or 60Gy IMRT. Patients considered high risk (R0(high risk) margins and/or ECS) will be randomised to receive 60Gy IMRT +/- cisplatin (100mg/m2 on days 1, 22 & 43).
The phase II (n=148; 74/randomisation) is powered to detect enhanced swallowing function (10 point difference using the MD Anderson Dysphagia Inventory) in patients receiving de-intensified treatment. It is important to note that post treatment swallowing function is directly related to IMRT +/- cisplatin normal tissue tolerance. The non-inferiority phase III (n=1000; 500/randomisation) will confirm functional improvement as well as patient safety. The phase

III which has been funded and is in set up, will require the involvement of ~25 European centres.

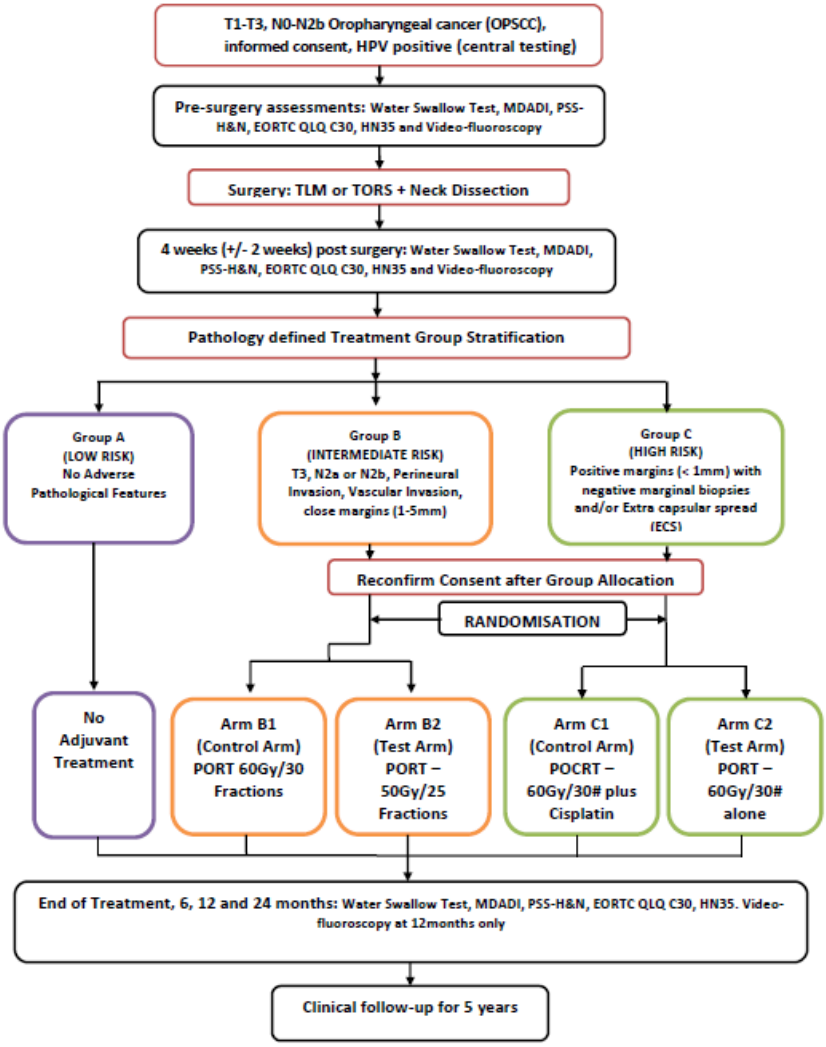
PATHOS-T: is the CR-UK funded bioresource collection associated with PATHOS. In patients recruited in the UK (the majority), five geographically distinct fresh tissue biopsies from the primary tumour, 2 from associated involved lymph nodes as well as blood samples pre-treatment, 6 weeks and 6, 12, 18 and 24 months post-treatment will be taken from all trial patients and stored to GCLP standard in Liverpool. Tissue samples are collected and frozen in Allprotect medium. FFPE sampes will be available for all recruited patients (UK & Europe)

PATHOS-TRANS: is a planned programme of work underpinned by the clinical data resulting from PATHOS and the bioresources collected as part of PATHOS-T. As explained above, PATHOS-TRANS will comprise 5 interconnected workstreams.

i. Investigation of the Tumour/Host immune synapse
ii. Investigation of the prevalence and consequences of viral genome integration and gene expression on the host gene expression profile and clinicopathological phenotype.
iii. Identification and validation of a tissue-based 'omics fingerprints of risk stratification, including outcome and treatment-related toxicity
iv. Investigation of the impact of intra-tumoural heterogeneity and the identification of sub-clones resulting in treatment resistance and/or metastasis.
v. Identification and validation of blood-based biomarkers (cfDNA and exosomes) of outcome and treatment response.

DNA and RNA will be extracted from the tissue samples. WGS, RNAseq and ChIPseq technology will then be employed to define `omic` fingerprints of swallowing outcome and therefore IMRT+/-cisplatin normal tissue tolerance. The geographically distinct primary tumour biopsies will allow investigation of the impact of intratumour heterogeneity and the lymph node tissue will allow the identification of expanded clones critical for metastasis development and treatment failure. The sequential blood samples will be used to define correlative liquid biomarkers based on ctDNA and/or exosome analysis as well as HPV serology. Any putative biomarker identified using the high-quality PATHOS-T fresh tissue bioresources (n=250-300) will be assessed for clinical utility using the higher quantity FFPE tissue samples (n=600-700) derived from all phase III patients.
The aim of this programme of work will be to define pre-treatment tissue and/or blood based molecular biomarkers of IMRT tolerance, metastasis and treatment failure. A successful outcome will ensure personalised prognostication and treatment allocation for patients presenting with HPV related OPSCC.

PATHOS Trial Schema:

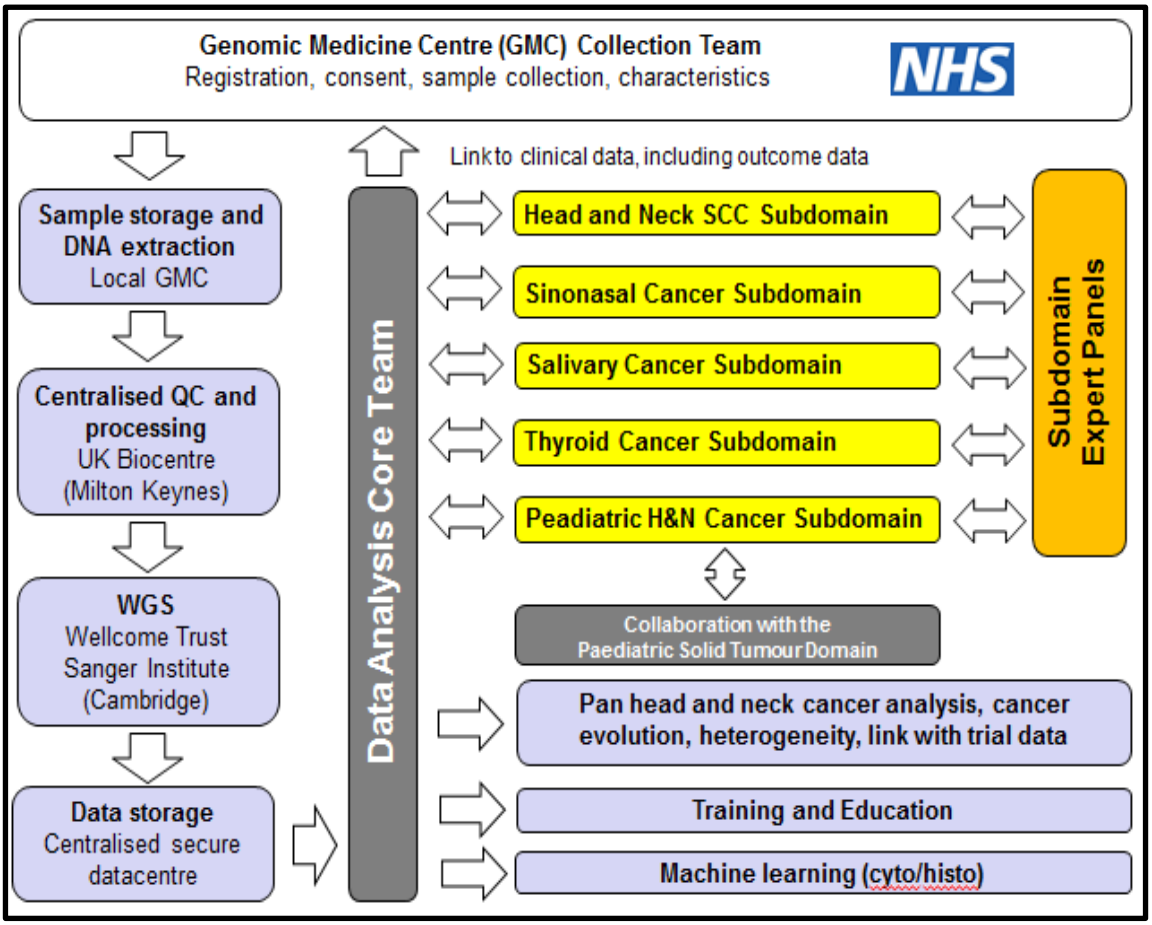


DELIVERY:

GeCIP Structure & Subdomains:

It is important to stress that when relative case incidence and duration of participation in the 100K Genome project is considered, recent recruitment data (March 2018) confirm that the Head and Neck GECIP domain is significantly outperforming most, if not all, other participating tumour groups. The Head and Neck domain is therefore well placed to achieve the central aims outlined above which are core to the overall project aims.

We propose the following GeCIP structure. We plan that the research questions defined by Subdomain members will collaboratively feed directly into the Data Analysis Core Team: A free flow of ideas and feedback between the two is anticipated. Professor Jenny Taylor, Wellcome Centre for Human Genetics, University of Oxford, will act as the lead for the analysis group for the GecIP, although each Subdomain will have a lead Bioinformatician selected based on their individual expertise. The Data Analysis Core Team will therefore act as a cross-cutting data-analysis Subdomain.



Collaborations including with other GeCIPs. *Outline your major planned academic, healthcare, patient and industrial collaborations. This should include collaborations and data sharing with other GeCIPs. Please attach letters of support.*

Paediatric Solid Tumour Domain (Lead: Dr. Tom Jacques, UCL) – please see letter of support attached

Endocrine & metabolism: Prof Stephen O`Rahilly, Cambridge. (Thyroid & parathyroid)

We will also endeavour to actively collaborate with GeCIPs who have included squamous cell carcinomas for other anatomical sites in order to exploit the opportunity to investigate squamous cell carcinomas as a common theme, in anticipation that they may well have common roots and oncogenic mechanisms.

Bristol Myers Squibb: Funders of NICO

Cancer Research-UK: Funders of PATHOS & PATHOS-T

Training. *Describe the planned involvement of trainees in the research and any specific training that will form part of your plan.*

The Head & Neck Cancer Domain GeCIP team members have a strong track record in clinical / molecular genetics, bioinformatics and statistical genetics. Researchers, students and clinicians will be encouraged to apply for funding to take up postdoc positions, MD/PhD positions and MSc and MRes programmes in genetic medicine, genetic counselling, and head and neck oncology. We aim to train the next generation of clinician and surgeon scientists to deliver on the existing challenges of head and neck surgery and oncology through this innovative programme.

Group members will contribute projects to the Health Education England MSc in Genomic Medicine through their affiliations with the HEIs that are selected to provide that programme. National trainees' days will be organised annually. We will seek funding to provide short courses tailored to the development of "crossover" skills and knowledge which we have found necessary for multidisciplinary research e.g. basic head and neck anatomy for molecular scientists; fundamentals of bioinformatics databases for clinicians.

People and track record. *Explain why the group is well qualified to do this research, how the investigators would work together.*

The UK has world-leading expertise in many of the above conditions, and this research plan effectively incorporates that national expertise, following a broad and inclusive consultation process. The conditions of interest are relatively numerous; to reflect this, the domain is structured into five disease-specific subdomains, each of which has been convened by experts in the specific area.

The group includes: NHS clinical oncologists and head neck surgeons, geneticists, genetic counsellors, and clinical scientists. The domain is involved across the designated NHS Genomic Medicine Centres and more broadly within the NHS and academic community in the UK. Domain members are already involved in extensive collaborations both nationally and internationally in head and neck genetics.

Analysis Core:
Prof Jenny Taylor (JT) has made a leading or substantive contribution to WGS500 and HICF2 genome sequencing programmes in rare diseases and cancer. The HICF2 WGS programme has undertaken analysis of active cancer patient cases, returning results which have informed diagnosis, prognosis or treatment selection. Whilst the HICF2 cohort contains only a small number of Head and Neck Cancer cases, a bioinformatics pipeline to assess viral integration status in cancers has been developed. Her group based at the Wellcome Trust Centre for Human Genetics, will provide analysis support to the sub-domains.

Prof Anna Schuh (AS) has also made a leading contribution to the HICF2 project and has experience in analysis of cancer genomes. She is lead on the CLL GECIP sub domain and the Haemato-oncology GECIP. Her group based at the Dept of Oncology, University of Oxford will work with Prof Taylor to undertake analyses for this GECIP Head and Neck sub domain. JT and AS are currently collaborating with GECIP members Prof Stuart Winter and Dr Ketan Shah on a circulating tumour DNA project for Head and Neck Cancer to use a targeted sequencing approach to assess utility of ctDNA analysis to inform assessment of treatment response for

oropharyngeal cases. This would complement the WGS analyses of oropharyngeal primary tumours being undertaken by GEL.

Clinical interpretation. *(Where relevant to your GeCIP) Describe your plans to ensure patient benefit through clinical interpretation relevant to your domain. This should specifically address variant interpretation and feedback and your interaction with the cross-cutting Validation and Feedback domain.*

The analysis plan will complement the analyses being undertaken by the participating GMCs and we will undertake to review the current practice at each GMC.

It is anticipated that the GMCs will undertake validation of SNVs in a small set of known cancer genes for their individual patients (as per the National Genomics Test Directory, above), using an orthogonal method (eg targeted NGS cancer panel), and that this will be undertaken in accredited clinical labs at the participating GMCs. Validation of CNVs and structural rearrangements may require further discussion with the GMCs and Validation and Feedback domain since array based methods for validation of CNVs lack the resolution of WGS. It may be necessary to design /use high resolution SNP arrays for CNV validation.

The GECIP sub domain will extend this by conducting more cohort-wide analyses across sub domains i) assessing recurrently mutated known cancer genes (eg the entire COSMIC gene list) for the specific sub phenotypes in the H&N cohort, with analysis to include SNVs, CNVs, structural anomalies and mutation signatures ii) assessing whether there is evidence for viral integration and any consequent activation of genes at viral integration sites iii) in due course undertaking analysis of non-coding regions to look for recurrent mutations which may have a regulatory effect.

Results which may have an impact on individual patients will be fed back to the referring GMC.

Beneficiaries. *How will the research benefit patients and healthcare institutions including the NHS, other researchers in the field? Are there other likely beneficiaries?*

There are established national networks of collaborating clinical centres researching within most subdomains, which are fully represented in this application (e.g. Newcastle, Manchester, Cambridge, Liverpool, Sheffield, Oxford). These networks have experience of standardised phenotypic ascertainment and documentation, handling issues surrounding patient and family recontact, and interpretation and feedback of potentially significant findings. The H&N GeCIP domain will facilitate the further development of such networks and the resultant sharing of expertise in each subdomain will ensure enrolment of the most appropriate patients in the 100K genomes project and will optimize the quality of variant interpretation and feedback to participants. The development and implementation of standard shared phenotypic datasets for the conditions of interest (where this has not already been achieved) will be critical to maximize the potential for outputs to be generated from the 100K genomes dataset; the national expertise that will be required to achieve this in each of the proposed subdomains has, we believe, been effectively captured in this application.

Commercial exploitation. *(Where relevant to your GeCIP) Genomics England has a very explicit intellectual property policy. We and other funders need to know if the proposed research likely to generate commercially exploitable results. Do you have commercial partners in place?*

It is understood that GeCIPs are not the vehicle whereby industry may access 100K Genomes data. The domain has well developed interactions with a variety of industrial partners that could form the basis of pre-competitive interactions (please see above).

References. *Provide key references related to the research you set out.*

1. De Martel, C. Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *Lancet Oncol,* 2012; 13**:** 607-15.
2. Masterson, L. De-escalation treatment protocols for HPV associated oropharyngeal squamous cell carcinoma: a systematic review and meta-analysis of current clinical trials. *Eur J Cancer* 2014; 50**:** 2636-48.
3. Ang K. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med,* 2010; 363**:** 24-35.
4. Fakhry C. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. *J Natl Cancer Inst,* , 2008; 100**:** 261-9
5. Schache AG, Liloglou T, Risk JM. Evaluation of human papilloma virus diagnostic testing in oropharyngeal squamous cell carcinoma: sensitivity, specificity, and prognostic discrimination. *Clin Cancer Res,* 2011; 17**:** 6262-71.
6. Psyrri A, Sasaki C. Future directions in research, treatment and prevention of HPV-related squamous cell carcinoma of the head and neck. *Head Neck Pathol,* 2012; 6**:** S121-8
7. Stransky, N. *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* 333, 1157–1160 (2011).
8. Agrawal, N. *et al.* Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in *NOTCH1*. *Science* 333, 1154–1157 (2011).
9. The Cancer Genome Atlas Network *. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*. 2015;517(7536):576-582. doi:10.1038/nature14129.
10. Kang, H. *et al.* Emerging biomarkers in head and neck cancer in the era of genomics. *Nat. Rev. Clin. Oncol*. 12, 11–26 (2015)
11. Masterson L, Lechner M, Loewenbein S, Davies-Husband C, Fenton T, Sudhoff H, Jani P, Goon P, Sterling J. CD8+ T cell response to HPV16 E7 can predict survival outcome in oropharyngeal cancer. *Eur J Cancer* 2016; 67; 141-151
12. Ferris RL. Immunology and Immunotherapy of Head and Neck Cancer. *Journal of Clinical Oncology*. 2015;33(29):3293-3304. doi:10.1200/JCO.2015.61.1509.
13. Ferris RL, Blumenschein G, Fayette J, et al. Nivolumab for Recurrent Squamous-Cell Carcinoma of the Head and Neck. *The New England journal of medicine*. 2016;375(19):1856-1867. doi:10.1056/NEJMoa1602252
14. Cancer Research UK, http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/thyroid-cancer, Accessed May 2018
15. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated

Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*. 2016;26(1):1-133.
16. Mitchell AL, Gandhi A, Scott-Coombes D, Perros P. Management of thyroid cancer: United Kingdom National Multidisciplinary Guidelines. *The Journal of Laryngology and Otology*. 2016;130: S150-S160.
17. Valderrabano P, McIver B. Evaluation and Management of Indeterminate Thyroid Nodules: The Revolution of Risk Stratification Beyond Cytological Diagnosis. *Cancer Control : Journal of the Moffitt Cancer Center*. 2017;24(5):1073274817729231.
18. Lee S, Hong SW, Shin SJ, Kim YM, Rhee Y, Jeon BI*, et al.* Papillary thyroid carcinoma associated with familial adenomatous polyposis: molecular analysis of pathogenesis in a family and review of the literature. *Endocrine journal* 2004, 51(3)**:** 317-323.
19. Bisogno G, Soloni P, Conte M, Podda M, Ferrari A, Garaventa A, Luksch R, Cecchetto G Esthesioneuroblastoma in pediatric and adolescent age. A report from the TREP project in cooperation with the Italian Neuroblastoma and Soft Tissue Sarcoma Committees. *BMC Cancer*. 2012; 12: 117
20. Weiss WA, Aldape K, Mohapatra G, Feuerstein BG, Bishop JM. Targeted expression of MYCN causes neuroblastoma in transgenic mice. *EMBO J*. 1997; 16(11): 2985-95.
21. Koike K, Jay G, Hartley JW, Schrenzel MD, Higgins RJ, Hinrichs SH. Activation of retrovirus in transgenic mice: association with development of olfactory neuroblastoma. *J Virol* 1990, 64: 3988-3991.
22. da Cruz Perez D, Pires FR, Alves F, Almeida OP, Kowalski LP. Salivary gland tumors in children and adolescents: a clinicopathologic and immunohistochemical study of fifty-three cases. *Int J Pediatr Otorhinolaryngol*. 2004; 68(7): 895-902

Data requirements

Data scope. *Describe the groups of participants on whom you require data and the form in which you plan to analyse the data (e.g. phenotype data, filtered variant lists, VCF, BAM). Where participants fall outside the disorders within your GeCIP domain, please confirm whether you have agreement from the relevant GeCIP domain. (max 200 words)*

The Head and Neck GECIP will undertake analyses of all samples submitted to this GECIP sub-domain. The total number of such samples at the current rate of recruitment is estimated to > 300 at the end of the current recriutment extension. The phenotypic / clinical subgroups are expected to comprise the following;

- Squamous cell carcinomas: HPV+ve & -ve
- Thyroid and parathyroid cancers
- Sinonasal tumours, incl. olfactory neuroblastomas
- Salivary gland tumours
- Paediatric head and neck cancers

Clinical, phenotypic and epidemiological data

The available clinical data from all participants is requested to inform the analyses described in the proposal. Ideally, this would include the following data fields, although it is not clear whether all this data has been collected, either prospectively or retrospectively, and may differ according to GMC.

- Cancer sub-type
- Local tumour or Metastasis and site of metastasis
- Pathological classification including grade and stage
- HPV status (+ve or –ve), method of detection (eg IHC) and information on strain if available
- Treatment history (surgical resection, radiotherapy, chemotherapy with details of dose and type of radiotherapy/ chemotherapy)
- Timing of biopsy / surgical resection for WGS sample in patient's treatment pathway
- Age
- Smoker / non-smoker
- Alcohol

Sequencing data

BAM and VCF files, including data on SNVs, CNVs and structural anomalies are requested. If mutation signatures have been analysed as part of the routine analysis this would be helpful. It would be useful to have the filtered lists by domain (1,2,3) to assess whether variants in likely actionable genes have been identified and reported back to the patients.

Data analysis plans. *Describe the approaches you will use for analysis. (max 300 words)*

The key clinical questions to be addressed are detailed in the main proposal but summarised here with attendant analysis approach:

1. **Molecular profiling of HPV+ve and HPV-ve squamous cell carcinomas of head and neck cancers to predict risk and inform de-intensification or intensification of treatment;** mutation profiling of key cancer driver genes, including SNVs, CNVs and SAs, plus mutation signatures and mutation burden will be undertaken to quantify the anticipated difference in mutational profile of HPV+ve and HPV-ve SCCHN which may, in part, explain the differential prognosis of these cancer sub types.
   We would anticipate that there would be significant numbers of both HPV positive and negative SCCHN in the GEL GeCIP collection. Hypotheses generated in the GeCIP dataset could be subsequently tested, subject to funding -as above, in trial cohorts eg

PATHOS & NICO thereby linking to high quality clinical outcome data available in these trials. Of particular note, the inclusion of the PATHOS clinical trial samples will significantly augment the world-wide number of HPV positive tumours for which WGS data will be available.

2. **Understanding the molecular mechanisms of HPV-driven SCCHN**: the analysis will focus on assessment of presence of viral genome in HPV+ve cases, whether integrated or episomal, whether there appears to be a common integration site (e.g. compare with myc insertion site for cervical cancers) and whether there is evidence of downstream activation of genes. As above, the clinical trial cohorts described earlier in the proposal will provide additional samples for testing hypotheses generated from the GEL dataset.

3. **Identifying novel molecular therapies for SCCHN**: the mutational profiling may reveal recurrently mutated driver genes which may be targeted by existing therapies or indicate novel therapeutic targets, or indeed may explain resistance to therapy. For example, it is recognised that radioiodine may be contra-indicated for SCCHN patients with mutations in BRAF. Other exemplars of such treatment modalities may emerge from comprehensive mutation profiling combined with clinical data on treatment responses.

4. **Improving the classification of thyroid and parathyroid carcinomas**; cytological and histological classification of these tumours can be ambiguous and comprehensive mutation profiling when combined with pathological data, may improve this classification. It would be important therefore to have access to the pathology datasets and combine this clinical information with the mutation profiles to address this question.

5. **Understanding the aetiology of sinonasal cancers:** as little is known about the aetiology of these, molecular profiling to identify any recurrently mutated genes and assessment of presence of integrated / episomal viral genomes (since viral particles have been reported in these tumours) will be the mainstay of the analysis for these tumours.

Analysis plans:

- Somatic and germline variants will be called using the GEL pipeline.
- Annotation would be with VEP.
- Sift, polyphen, ExAc and CADD would all be used for the assessment of deleteriousness of variants with COSMIC for additional information on somatic variants.
- We understand that the GEL pipeline encompasses Canvas and Manta for CNVs and SVs, respectively.
- Mutation signatures will be called with R package deconstructSigs (https://cran.r-project.org) and with the COSMIC Mutation Signatures Consensus list (Alexandrov et al. 2013) (http://cancer.sanger.ac.uk/COSMIC /signatures).
- Viral genome analysis will be with a bespoke pipeline which would be imported into the embassy (see below).
- Non-coding genome analyses: we will analyse somatic variants in non-coding regions of the tumour genomes with a particular emphasis on recurrent mutations (using MutSigCV to define the statistical significance of these) and mutations in promoter regions, UTRs, transcription factor binding sites and other non-coding elements defined by the ENCODE and GENCODE projects.

Alignment and calling requirements. *Please refer to the attached file (Bioinformatics for 100,000 genomes.pptx) for the existing Genomics England analysis pipeline and indicate whether your requirements differ providing explanation. (max 300 words)*

Our understanding is that the alignment and calling algorithms are fixed for the GEL programme (Illumina Variant Studio with recalling on Platypus) and that alternative pipelines cannot be used.

Tool requirements and import. *Describe any specific tools you require within the data centre with particular emphasis on those which are additional to those we will provide (see attached excel file List_of_Embassy_apps.xlsx of the planned standard tools). If these are new tools you must discuss these with us. (max 200 words)*

At this stage, we primarily anticipate using the tools available in the Research Embassy. However, we have a bespoke pipeline for analysis of viral genomes that would need to be imported.
It may be that we identify additional software packages as the analyses progress, although recognise that these must be OpenSource and not accessible from the internet.

Data import. *Describe the data sets you would require within the analysis environment and may therefore need to be imported or accessible within the secure data environment. (max 200 words)*

Pending the outcome of discussion regarding sequencing of clinical trials cohorts (e.g. PATHOS, NICO, described above), it may be necessary to import these datasets to provide additional power for analyses.

It may be necessary to import data on analysis of other markers analysed on samples submitted to the GECIP for example matched circulating tumour DNA analyses for primary tumours submitted to GEL (if such ctDNAs have not been analysed as part of GEL), trasncriptomics, methylomics and microarray data.

Computing resource requirements. *Describe any analyses that would place high demand on computing resources and specific storage or processing implications. (max 200 words)*

At this stage, it is not anticipated that there would be extraordinary requirements for processing or storage, particularly as the numbers of samples across the cohort would be modest, but this may change once we have experience of the research embassy environment capability.

Omics samples

Analysis of omics samples. *Summarise any analyses that you are planning using omics samples taken as part of the Project. (max 300 words)*

We are not clear whether samples for other 'omics analyses have routinely been collected by GMCs and processed by GEL. If samples are available we would propose to undertake transcriptomics analyses to confirm expression effects of mutations identified through genetic analysis.

Data access and security	
GeCIP domain name	Head and Neck Cancer
Project title *(max 150 characters)*	WGS analysis of Head and Neck Cancer

Applicable Acceptable Uses. *Tick all those relevant to the request and ensure that the justification for selecting each acceptable use is supported in the 'Importance' section (page 3).*

BOLD = anticipated Applicable Acceptable Uses

☐ ***Clinical care***

☐ ***Clinical trials feasibility***

☐ ***Deeper phenotyping***

☐ ***Education and training of health and public health professionals***

☐ ***Hypothesis driven research and development in health and social care - observational***

☐ ***Hypothesis driven research and development in health and social care - interventional***

☐ ***Interpretation and validation of the Genomics England Knowledge Base***

☐ ***Non hypothesis driven R&D - health***

☐ *Non hypothesis driven R&D - non health*

☐ *Other health use - clinical audit*

☐ *Public health purposes*

☐ *Subject access request*

☐ *Tool evaluation and improvement*

Information Governance

☐ The lead and sub-leads of this domain will read and signed the Information Governance Declaration form provided by Genomics England and will submit by e-mail signed copies to Genomics England alongside this research plan.

Any individual who wishes to access data under your embassy will be required to read and sign the above form. Access will only be granted to said individuals when a signed form has been processed and any other vetting processes detailed by Genomics England are completed.

Other attachments

Attach other documents in support of your application here including:

- a cover letter (optional)
- CV(s) from any new domain members which you have not already supplied (required)
- other supporting documents as relevant (optional)