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 was announced in June 2015 following a first round of applications with expressions of interest in January 2015. In April 2016 we invited the inaugurated Cancer GeCIP domains to develop research plans for 'Gear 2' working closely with Genomics England. Within the Cancer Main Programme, the 'Gear 2' phase of the project refers to recruitment of specific cohorts of patients, inclusion of biopsy tissue (diagnostic/recurrence) and ctDNA in selected cohorts and the initiation of clinical trials in early stage (adjuvant/consolidation) setting. These will be used to ensure that the plans 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 required to be submitted to individual funders. They 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.

Domain leads are asked to complete all relevant sections of the GeCIP Detailed Research Plan Form, ensuring that you provide names of domain members involved in each aspect so we or funders can see who to approach if there are specific questions or feedback and that you provide details if your plan relies on a third party or commercial entity. You may also attach additional supporting documents as relevant (optional).

Additional members can apply to join the GeCIP domain by completing the form on our website found here: http://www.genomicsengland.co.uk/join-a-gecip-domain/.

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

Application Summary	
GeCIP domain name	**Colorectal cancer**
Project title *(max 150 characters)*	**Colorectal cancer research in the 100,000 Genomes Project**

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

Our Overall Aim is to utilise whole-genome sequence data to provide a step-change in patient management over current molecular tests that are limited in their scope and utility. We shall principally achieve this through research projects, including studies that aim for a fundamentally improved understanding of colorectal tumorigenesis. Our specific Aims are:

1. Technical optimisation
2. Driver gene discovery and exploitation.
3. Analyses complementary to whole-genome DNA sequencing.
4. Clinical reporting.
5. New classifiers.
6. Specific focus areas.
7. Cross-cutting analysis.
8. Training and education.

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)*

Colorectal cancer (CRC), also known as bowel cancer, is the development of cancer from the colon or rectum (parts of the large intestine). It is one of the most common types of cancer, making up about 10% of all cancer. Around 38,000 people in the UK develop CRC annually, and 16,000 die from the disease each year. A person's chance of surviving CRC is highly dependent on how early or late the cancer is identified. Although diet and lifestyle impact on risk of developing CRC, no single modifiable risk factor has been shown to have a major effect on personal risk. The CRC GeCIP domain will aim to study the DNA of CRC patients in both their normal tissue (germline) and the tumour itself (somatic), and identify spelling mistakes (variants) that may contribute to the cancer itself, or may determine how the tumour, once established, develops and evolves over time. By studying these variants and understanding how they have their effects, and the impact on the patients response to treatment, they hope to be able to identify new targets for drugs, or cases where existing drugs could be repurposed to improve outcomes. They are also hoping to catalogue variants that have significant effects on outcomes so that these can be used to classify patients and identify beforehand those who are more/less likely to respond to treatment so that their clinical care can be adjusted accordingly.

Expected start date	**Q2 2017**
Expected end date	**Q2 2020**

Lead Applicant(s)	
Name	Ian Tomlinson
Post	Professor of Molecular and Population Genetics, Group Head/PI and Hon. Consultant in Clinical Genetics

Department	Oxford Centre for Cancer Gene Research
Institution	University of Oxford
Current commercial links	None

Gear 2 Substudies
CR01: Identifying critical events in benign to malignant transition through the analysis of colorectal carcinomas arising in polyps
CR02: An exploration of ctDNA in colorectal cancer: a pilot project designed to lead to full projects
CR03: Identifying the effects of known risk factors on colorectal cancer genomes
CR04: Identifying the effects of prior chronic aspirin/NSAID use on colorectal cancers
CR05: Colorectal cancers arising on a background of chronic inflammation
CR06: Profiles of colorectal cancers arising in diabetics
CR07: Small bowel cancer genomes
CR08: Analysis of synchronous primary colorectal cancers and their origins
CR09: Identifying new molecular sub-types and pathways of colorectal cancer
CR10: Cross-cancer analyses based on shared aetiology or molecular pathways
CR11: Identification of actionable mutations in primary and secondary colorectal cancers
CR12: Colorectal cancer profiling in non-European ethnic groups living in the UK
CR13: Carcinogenesis in specific morphological or molecular sub-types of colorectal cancer
CR14: Evolution of primary and secondary colorectal cancer in time and space
CR15: Non-human genomes and the gut microbiome in colorectal cancers
CR16: Molecular profiling of early (T1-2N0M0) colorectal cancers and local lymph nodes
CR17: Determining the effects of neo-adjuvant chemo/radiation therapy on colorectal cancer genomes
CR18: Identifying and characterising intrinsic and extrinsic mutation signatures and mutator phenotypes in colorectal cancers
CR19: New therapeutic or imaging targets
CR20: Integrating genomics and colorectal cancer clinical trials
CR21: Exceptional or highly unusual colorectal cancer cases (age, germline predisposition, excellent response, previous treatment with radiotherapy or chemotherapy, *etc*)
CR22: Colorectal cancer driver mutations outside the exome
CR23: Inherited variation
CR24: Functional evaluation and interpretation of potential driver mutations in colorectal cancer, and follow-up analyses in additional data sets and model systems

GeCIP domain - Expression of interest

Full proposal	
Title *(max 150 characters)*	**Colorectal cancer**

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.*

Around 38,000 of the UK population develop colorectal cancer (CRC) annually and 16,000 die from the disease. With an estimated 1 million new cases annually worldwide, CRC is the third most common cancer and second leading cause of cancer-related death and its incidence is increasing. Prognosis is highly stage-dependent. Advances in early detection and treatment have led to somewhat improved outcomes, with 57% of all patients surviving 5 years post-diagnosis. However, many patients present with late stage disease and 5-year survival is less than 10% in this group. Thus, there is unmet need to improve early detection, curative treatments and clinical outcome from localised and metastatic CRC.

Although diet and lifestyle impact on CRC risk, no specific modifiable environmental risk factors are known that have a major effect on personal risk. Targeted surveillance and chemoprevention in those at high genetic risk have been shown to highly effective in reducing CRC mortality and incidence. Although population screening for the early detection of bowel cancer is available and cost-effective, its take-up is only about 50% and its sensitivity and specificity are currently sub-optimal. National screening has, however, caused a stage-shift in CRC, with far more early-stage cancers being found. Whilst this is evidently a desirable outcome, these cancers present their own problems of management

Our Overall Aim is to utilise whole-genome sequence data to provide a step-change in patient management over current molecular tests that are limited in their scope and utility. Specific Aims are:

1. Technical optimisation

It is currently not clear whether FFPE samples will be utilised for sequencing. Several valuable CRC sample sets (e.g. clinical trials) are principally available in FFPE form and we would be keen to contribute to the assessment of these samples. We will also assess CRC-specific features such as the need for microdissection, desirability of additional sequencing depth for selected samples, and deviations from a standard analysis pipeline (e.g. using multiple mappers and callers, assessing structural variation, MSI-specific analysis, copy number and ploidy analysis, and intra-tumour heterogeneity).

2. **Driver gene discovery and exploitation.**

Somatic driver gene identification will rely on statistical methods of assessing mutation over-representation in genes and pathways, such as Intogen, MutSigCV, and Hotspotter. These will be backed up by complementary analyses performed for Aim 3. We shall also work with other GeCIP Domains to develop methods of assessment for mutations in non-coding regions. Moreover, while the focus of discovery will be on the identification of somatic driver mutations, the data will also allow investigation of germline variants that affect CRC susceptibility, treatment response and toxicity, outcome and pathways of tumorigenesis. We aim to translate findings into patient benefit, a critical aspect of which will be new targets for therapy and, perhaps, prevention.

3. Complementary analyses.

Further analyses of DNA (outside genome sequencing), RNA and protein studies (including

immunohistochemistry), specialist histopathology, and the use of model systems and bespoke in silico assessments all have several potential uses downstream of the genome sequencing. For example, it may inform functional annotation, provide actionable targets (e.g. pathway-based), and allow more sensitive driver identification (e.g. CNVs, structural variants).

4. Clinical reporting.
Clinical reporting of somatic variants and other molecular features with potential importance for patient treatment and prognosis will be developed to take account of multiple sources of information and the uncertainty inherent in some assessments. We shall adhere to the highest standards for conventional measures such as histopathology and may technically validate some findings. We will explore methods of disseminating data to the end user in a format which is easy to understand for patient management.

5. New classifiers.
Integrated analysis and machine learning methods will be used to identify new classifiers of CRC based on molecular features, histology and patient characteristics. Refined genetic pathways will also be determined. These classifiers will be tested as predictors of features such as prognosis and response to treatment in the GeL data sets and validated in collaborators' samples.

6. Specific focus areas.
See section below.

7. Cross-cutting analysis.
We anticipate extensive cross-GeL collaborations, including both cross-cutting and disease-specific themes. Examples include putative Domains in data analysis, inherited cancer, ELSI, health economics, population genetics, and other cancer types. We shall play a central role in cross-cancer analyses (e.g. pan-cancer drivers, method comparisons, new analytical tools, collaboration with inherited cancers GeCIP) and pan-GeL analyses (e.g. population genetics, incidental findings).

8. **Training and education**.
To ensure long-term durability and success, we will train a new generation of geneticists, molecular pathologists, clinical bioinformaticians, statisticians and biologists. A critical part of training will be in the clear written and oral presentation of results to clinicians. We will build on existing efforts in several centres. For example, the ICR/Royal Marsden/Imperial College are developing an MSc in Genomic Medicine, at Oxford University a MSc/DPhil programme in Clinical Omics for medical post-graduates is in preparation, and the Sanger Institute PhD Programme already trains both science and clinical PhDs. We shall work with the trainees in other programmes, such as the MRC Pathology Nodes. Trainees from these and other programmes participating in allied research projects will be invited to take part in the CRC Domain and gain access to the data as appropriate. The training lead will also be responsible for wider education of the medical professions (e.g. presentations to Oncologists, GPS and senior nurses) and provision of information to patients and the general public.

Exemplar focused projects
We plan that the Domain will undertake a number of specific projects on particular CRC types, each run by small groups of individuals with special interests, supplemented where necessary by researchers and clinicians outside the Domain. These will be added to as time goes on and will require specific funding which may be obtained by the Domain or with the Domain's support from UK and international funders. Examples include

- CRC in ethnic minorities
- CRC in inflammatory bowel disease

- carcinomas arising in adenomas carcinoma in situ (high-grade villous adenomas)
- unusual CRC morphologies, e.g. serrated cancers or carcinomas with neuroendocrine features
- cancer evolution in space and/or time, or as a result of treatment
- improving the utility of clinical trials (e.g. FOCUS4, SCOT)
- molecular staging and optimal biopsy analysis, e.g. occult disease in lymph nodes, circulating cancer cells or DNA
- non-human genomes in CRCs
- screen-detected CRCs
- high mutation burden cancers
- exceptional treatment responders

Domain role in specifying samples to be sequenced

At the time of writing, mechanisms for deciding on which samples will be sequenced have not been finalised, but we shall provide guidelines if selected to provide a detailed Domain application. We emphasise that the Domain will work within any GeL guidelines that are produced. However, we need to be cognisant of international efforts in CRC sequencing, including TCGA project which is sequencing large numbers of "conventional" CRCs. We shall contact TCGA and other CRC sequencing projects to explore the benefits of forming alliances. We propose that we will sequence a large number (~50-75%) conventional CRCs, but our collective view is that an "all-comers" or "first come-first served" approach will not maximise the potential of the GeL project. We shall therefore liaise with GMCs to guide collection towards the most biologically and clinically interesting samples. Some examples include samples for the specific projects in the previous section.

Funding

Leveraged existing funding will include:

- CR-UK Cancer Centres (e.g. ICR/Royal Marsden, Leicester, Oxford, Cardiff, Edinburgh, UCL, etc)
- NIHR Biomedical Research Centres (UCLH, Royal Marsden, Oxford,)
- Health Innovation Challenge 2 Programme (Oxford)
- Wellcome Trust Strategic Awards (ICR Centre for Evolution and Cancer)
- EU (e.g. Tomlinson, Houlston, Ilyas)
- CR-UK programme grants (e.g. Tomlinson, Dunlop)
- S-CORT/FOCUS4 trial stratified medicine (e.g. Maughan/Tomlinson/Quirke/Wilson)
- SCOT trial (Iveson, Kerr)
- FOXTROT trial (Morton)
- FACS and New EPOC trials (Primrose)

New funding through grant applications and strategic awards will be required for several of the proposed activities, including

- Basic administration, travel and meetings
- Training (including patient and clinician communication and specialist histopathology)
- Sample collection outside basic Genomic Medicine Centre remit
- Molecular work outside genomic DNA sequencing (e.g. RNAseq, MethylSeq, etc)
- High-depth sequencing, whether genome-wide, focussed or targeted (e.g. molecular staging)
- Technical and biological validation
- Immunohistochemistry
- Statistical analysis and mathematical modelling

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.*

Although all cancers evidently share some features, each also has highly specific features. For example, some mutations of the APC gene, which is the most common driver of CRC growth, are repeatedly wrongly annotated as pathogenic in public data sets owing to a lack of specific expertise in CRC biology. Similar considerations apply to many other genes. For such reasons, the Domain Steering Committee will largely consist of CRC specialists, including NCRI overall lead, the chair of the NCRI CRC Study Group and other clinical trialists, academics with an interest in CRC, bioinformaticians, biostatisticians, medical statisticians, diagnosticians, CRC biologists, clinical scientists, trainees and representatives of the pharmaceutical industry. Many of these individuals already collaborate closely.

David Adams (Wellcome Trust Sanger Institute): functional genetics, mouse models
Deborah Alsina (Bowel Cancer UK): funder, patient and public liaison
Sally Benton (Surrey): CRC screening
John Bridgewater (UCL): medical oncology, clinical trials, clinical translation
Rachel Butler (Cardiff): NHS molecular genetics (Wales)
Jean-Baptiste Cazier (Birmingham): bioinformatics, computing
Malcolm Dunlop (Edinburgh): surgery, molecular genetics (Scotland)
Jackie Gath: patient representative
Marco Gerlinger (ICR/Royal Marsden): medical oncology, cancer evolution, implementation and training
Trevor Graham (Bart's and the London): cancer evolution
Richard Houlston (ICR): molecular and clinical genetics, statistics
Mohammad Ilyas (Nottingham): histology, molecular pathology
Tim Iveson (Southampton): medical oncology, clinical trials (adjuvant setting)
Rachel Kerr (Oxford): medical oncology, clinical trials (adjuvant setting) and translational medicine
Tim Maughan (Oxford): clinical oncology, clinical trials (metastatic setting)
Dion Morton (Birmingham): surgery, clinical trials (neo-adjuvant setting), chemoprevention
John Primrose (Southampton): surgery of metastasis, clinical trials, translational medicine
Phil Quirke (Leeds): histology, molecular pathology
Andrew Ramwell (St George's, London): colorectal surgery
Manuel Salto-Tellez (Belfast): histology, molecular pathology (NI)
Matt Seymour (Leeds): NCRI Overall Lead
Andrew Silver (Bart's and the London): molecular genetics, cancer in ethnic minorities
Anne Thomas (Leicester): medical oncology, chemoprevention
Ian Tomlinson (Oxford): molecular, functional and clinical genetics, cancer evolution
Axel Walther (Bristol): medical oncology
Richard Wilson (Belfast): NCRI Colorectal Cancer lead
Sarah Wordsworth (Oxford): health economics
Chris Yau (Oxford): statistics, sequencing data analysis methods, machine learning

If the bid is successful, we shall recruit additional Domain members, including, but not limited to: major funders; international experts; pharmaceutical industry and SMEs; trainees; ethics (e.g. from ELSI GeCIP Domain); "GeL central", including the interpretation, validation and feedback domain; MRC Molecular Pathology Nodes; regulatory bodies; and those Genomics Medicine Centres, CR-UK Cancer Centres and Biomedical Research Centres not currently represented.

The Steering Committee will be the senior decision-making body. It will provide oversight, decide strategy, prioritise samples and projects, co-ordinate grant applications and initiate external liaison with groups such as TCGA. It will meet every three months and the Chair will rotate yearly. Sabine Tejpar (Leuven), an expert in molecular genetics and clinical trials, has agreed to be lead

external committee member and will recruit further colleagues from outside the UK. Independent experts may be asked to advise, comment on and review the Domain's work as required.

Sub-Committees will be smaller (5-10 members) and may comprise both Steering Committee members and additional members. Their responsibilities will map to one or more of the specific project aims.

- Training and Education (Leads: Marco Gerlinger, RMH/ICR and Deborah Alsina, Bowel Cancer UK)
- Driver gene discovery (Lead: TBA)
- Cancer evolution (Lead: Trevor Graham, Bart's and the London)
- Germ line (Lead: Richard Houlston, ICR)
- Functional studies: (Lead: David Adams, WTSI)
- Variant interpretation, diagnostics and clinical reporting (Lead: David Church, Oxford)
- Statistics, analysis and classification (Leads: Jean-Baptiste Cazier, Birmingham and Phil Quirke, Leeds)
- Data, records, website and informatics (Lead: TBA)

Proposed UK Leader

The Steering Committee will initially be chaired by Ian Tomlinson who will provide overall Domain leadership for the first year. Prof. Tomlinson has worked in colorectal cancer genetics for over 20 years, and is expert in both inherited predisposition and somatic mutation. He has a long-standing interest in cancer evolution and has published several recent manuscripts on genome and exome sequencing of cancer patients.

Potential international collaborators

We have not yet formally approached international collaborators, since we believe that this would best be done subsequent to a successful Domain application when we are better able to negotiate with other groups, especially large consortia. However, we have identified several target groups with similar or complementary interests, including Raju Kucherlapati (TCGA), Bert Vogelstein (Johns Hopkins), Oliver Sieber (WEHI, Melbourne), Ulrike Peters (Seattle), Steve Gruber (USC), Monica Bertagnolli (Boston), Frederico Innocenti (UNC), and a number of UK and European clinical trials of colorectal cancer therapy and prevention.

Mechanisms for pre-competitive interaction with partners from industry

The Domain membership, in particular the Oncologists and Histopathologists, have well established collaborations with industrial partners and hence have early access to an active pipeline of novel agents. Researchers at the ICR through links with the CR-UK Centre for Drug discovery are particularly well placed to exploit GeCip discoveries. Other Domain members have active partnerships with the Biotech. industry including companies working in the field of high-throughput sequencing and translational genetic/molecular diagnostics. For example, Mohammad Ilyas works with industrial collaborators Leica, Astra-Zeneca, GE/Omnyx, TissuGnostics, and Barco.

Gear 2 substudy proposals

Gear 2 Eligibility Criteria

Pre- and post-operative cfDNA collection - All Dukes' B-D undergoing primary resection.
Multiregion (multiple samples/cores taken ex-vivo from a single resected tumour mass) - Multiregion sampling including lesions with both benign adenomatous and invasive areas: Dukes' A-D tumours, up to 4 regions encompassing any combination of pre-malignant and malignant regions. Multiregion sampling in inflammatory bowel disease: Up to 4 regions of cancers +/- associated pre-malignant lesions (polyps, DALMS) from patients with IBD (UC and Crohns').
Multitumour (disparate sites of tumour deposition collected synchronously) - Multiple primary lesions, (a) 2 or more primary colorectal cancers, (b) 1 primary colorectal and 1 or more dysplastic adenomas.
Longitudinal - Paired samples pre-and post- administration of neoadjuvant chemotherapy; Paired samples pre-and post- administration of 1st line treatment in metastatic setting; Sampling of tumour recurrence or metastasis is invited for any patient with a previous successfully sequenced tumour.

CR01: Identifying critical events in benign to malignant transition through the analysis of colorectal carcinomas arising in polyps

Genes that drive the progression of a benign precursor, such as a colorectal adenoma, to a malignant carcinoma are largely unknown in most cancer types. In CRC, driver mutations such as p53 are thought to occur close to the benign-malignant transition, although p53 itself is not generally regarded as responsible for invasiveness. Since no single driver of malignancy has yet been found, many other genetic and non-genetic explanations are possible. These could include one or more of the following: multiple mutations in a polygenic model; genomic instability; and changes in the microenvironment. The use of paired benign-malignant samples from the same lesion will provide as controlled a comparison as possible for determining differences between the benign and invasive states.

CR02: An exploration of ctDNA in colorectal cancer: a pilot project designed to lead to full projects

Circulating tumour DNA (and, to a lesser extent circulating tumour cells) are mooted as powerful ways of monitoring for persistence of disease or relapse after curative treatment, for progression in the metastatic setting or even for early detection of cancer. For CRC, relatively little is known about ctDNA burden in these settings, although some authorities suggest that the levels are very low in disease that is confined to the bowel. We shall investigate this by collecting multiple samples for ctDNA extraction from a set of about 30 patients of various CRC stages at diagnosis, post-surgery and on follow-up. Mutations in the primary cancer (and any metastases available) will be used as targets to assess the ctDNA. Based on these findings, we shall design further experiments, for example to assess the utility of ctDNA in primary bowel cancer screening.

CR03: Identifying the effects of known risk factors on colorectal cancer genomes

A number of common genetic and environmental/lifestyle risk factors are expected to influence the cancer genome. These range from exposure to mutagens to common SNP alleles. One simple example is that cancers tend to have a high burden of passenger mutations and more C>T changes resulting from cytosine deamination. Many other hypotheses are testable, based on two underpinning notions: (i) mutagens (or epimutagens) will leave detectable signatures of their presence in DNA; and (ii) that non-mutagenic cancer-promoting agents will influence the (micro)environment causing selection of different driver mutations. For example, it is plausible that individuals with a high burden of risk SNPs require fewer, or a different spectrum of, driver mutations for a cancer to grow, compared with individuals with few risk SNP alleles; and other cancers may acquire somatic mutations at the same loci at which predisposition SNPs occur if they have low-risk genotypes. Putative cancer-promoting risk factors, such as weight, may cause rapid evolution. This project may provide important insights into how risk factors act and even whether they are causal for cancer or markers of risk.

CR04: Identifying the effects of prior chronic aspirin/NSAID use on colorectal cancers

NSAIDs are protective against colorectal polyps and cancer. We wish to investigate whether this affects the cancer genome. We might find, for example, find genetic pathways that differ from other sporadic colorectal cancers, less diversity, a different mutation spectrum or no difference at all. We would endeavor to relate any differences to mechanisms of NSAID action, for example in organoid models. The underlying reasons for taking NSAIDs are a potential confounder, and hence we will explore whether any NSAID effects vary with the reason for their use (e.g. primary prevention, chronic inflammation outside the GI tract, etc). Note: any CRCs with a known or suspected specific aetiology would be excluded from this analysis.

CR05: Colorectal cancers arising on a background of chronic inflammation

Bowel cancers arising on a background of Crohn disease, ulcerative colitis, PSC and related conditions are an important sub-group of tumours that arise from particular precursor lesions and follow molecular pathways that overlap only partially with those of sporadic CRC. Although some data exist on the clonal structure and evolution of these cancers, very sparse whole genome sequencing data are available, and basic knowledge is lacking of: driver mutations and copy number changes; clonal complexity and bottlenecking; effects of therapy; causes of malignant progression; mutational processes; and changes over time, including relation to severity of the underlying inflammatory process.

CR06: Profiles of colorectal cancers arising in diabetics

Pre-existing diabetes is a moderately strong risk factor for CRC, although certain oral hypoglycaemics may protect against increased risk. It is postulated that the raised risk results from chronically increased IGF signaling, which prima facie is tumour-promoting rather than mutagenic. However, several other explanation remain possible. This project will allow a comparison to be made between the genomic landscapes of colorectal cancers arising in type 1 or 2 diabetes are distinct from other sporadic CRCs, providing clues regarding the mechanism of risk. Selected DNA methylomics, gene expression profiling, proteomics and metabolomics would be of great benefit alongside genome sequencing here.

CR07: Small bowel cancer genomes

Small bowel cancers are rare. They can arise in inherited conditions such as Lynch syndrome and the various adenomatous polyposis syndromes, but most have no clear genetic basis. Small bowel cancer usually presents late and has a poor prognosis. We know very little about their genomes, from driver mutations to evolution, and this project will perform a comprehensive analysis that may aid in the management of these tumours. Cryptic or undiscovered germline mutations in bowel cancer predisposition genes may also be found.

CR08: Analysis of synchronous primary colorectal cancers and their origins

Multiple primary colorectal cancers can occur in Mendelian predisposition syndromes, but most such cases arise with no clear underlying genetic cause. Since CRC is a common disease, multiple cancers may arise largely by chance in some patients. However, synchronous carcinomas (or a carcinoma and severely dysplastic polyp) in the same region of the large bowel raise the suspicion of some sort of underlying “field defect”. A similar factor may cause the well-known phenomenon of a carcinoma with several discrete, but adjacent “satellite” polyp(s). We shall determine whether the tumours from such patients have evidence of a common origin based on their complement of driver and passenger (epi)mutations. We shall also determine whether the normal-appearing bowel between the tumours appears (epi)genetically normal, or is a mosaic, having acquired changes associated with an undetected clonal expansion that has aided carcinogenesis in the synchronous tumours. If detected, these findings would enhance our understanding of colorectal carcinogenesis, raising the possibility of undetected, but potentially pre-cancerous clones, underlying the growth of many sporadic CRCs.

CR09: Identifying new molecular sub-types and pathways of colorectal cancer

Although most of the major CRC driver genes are likely to have been found by groups such as TCGA, rare or weak-effect “mini-driver” mutations in coding regions may remain to be found, and

we shall search assiduously for these. However, a more fruitful task may be to refine the current molecular classification of colorectal cancer. Ideally, this would be based on multi-omic (poly-omic?) approaches that can be used should funding be available. Initially, working with the Machine Learning domain, we shall search for new mutation-based CRC groupings beyond the triad of hypermutation, ultramutation and chromosomal instability – and arguably CpG island methylation – that currently holds. A variety of tools will be used including both conventional hierarchical and Kmeans clustering and principal component analysis, and specialist Bayesian methods (e.g. regression, network analysis). The molecular pathways (small mutations, copy number changes, etc) underlying these cancer groups will be identified and validated in independent data sets. Associations with clinico-pathological variables, such as survival, will be assessed in clinical trial data sets, such as SCOT and FOCUS4.

CR10: Cross-cancer analyses based on shared aetiology or molecular pathways

Many cancers share genetic and other features, and TCGA and others have identified a number of driver genes mutated in more than one cancer type (sometimes referred to inaccurately as "pan-cancer drivers". For colorectal cancer, the natural comparative analyses are with (i) other luminal GI malignancies such as adenocarcinomas of the oesophagus and stomach (some shared driver mutations and forms of genomic instability), and (ii) endometrioid cancers of the uterus and ovary (shared genetic aetiology and forms of genomic instability). Many of the analyses detailed elsewhere for CRC will be performed across these cancer types. Exemplar questions of note include whether shared driver mutations tend to occur at similar stage of carcinogenesis, whether there are alternative means of (in)activation of the same pathway in different tissue types, and whether clinicopathological-molecular associations hold across cancers. We will specifically work with the ovarian cancer domain and the prospective upper GI and endometrial cancer domains.

CR11: Identification of actionable mutations in primary and secondary colorectal cancers

We shall identify mutations and forms of genomic instability that have potential relevance for patient management. We shall work with Validation and Feedback to provide CRC-specific expertise in this regard. We shall assess selected variants of uncertain pathological significance using more extensive bionformatic assessments and wet lab. functional assays.

CR12: Colorectal cancer profiling in non-European ethnic groups living in the UK

There is good evidence that some cancers have features that vary with ethnic origin, even when individuals live in the same locale. For example, MSI+ CRCs have been reported to be more common in non-European groups. However, such studies are prone to confounding by factors such as social class. Moreover, some features such as cancer stage at presentation may depend on cultural factors rather than underlying biology. Nevertheless, useful information to enable better treatment of CRC in non-European ethnic groups may be derived from a comparison within GeL. An important factor is that self-reported ethnicity and genetic ancestry can be determined and analysed separately. The analysis may, for example, identify a higher prevalence of certain molecular sub-types, different mutation spectra and signatures, different frequencies of actionable mutations, and an increased prevalence of certain predisposition mutations that are specific to particular ethnicities.

CR13: Carcinogenesis in specific morphological or molecular sub-types of colorectal cancer

Increasingly fine scale classification of cancer will be required if reality is to catch up with the hype of personalised or precision medicine. Breast cancer provides a prime example of how different molecular and morphological subtypes behave differently and require different treatment regimens. For CRC, similar considerations probably apply, but analysis has been more limited (e.g. types of genomic instability, consensus mRNA expression). Although TCGA project has identified major CRC drivers, it is not yet powered to identify subtype-specific drivers. Here, we shall enrich our collection for morphological sub-types (to be chosen). We shall also perform post hoc subgroup analyses on specific molecular sub-types. We shall principally aim to identify driver mutations and copy number changes, but will also perform all other standard genome analyses

such as assessment of clonality, mutation spectra, etc. This project will link in with that on molecular pathways described elsewhere.

CR14: Evolution of primary and secondary colorectal cancer in time and space

A burst of NGS-based studies has transformed cancer evolution analysis from a backwater to mainstream as a result of the excitement it has engendered in the Oncology community. Much remains to be done, however. In part this will consist of more detailed understanding of tumorigenesis in time and space, especially as regards mechanisms of resistance to targeted, genotoxic and immunotherapies, and linking the findings into therapeutic strategies and prognostic markers. Our overall strategy in the short term is to continue to describe cancer evolution at the highest possible level of complexity within the GeL project. We wish to use cutting edge statistical methods to identify sub-clones within biopsies, to correct for copy number and tumour cell fraction, to integrate copy number changes into evolutionary trees, to time mutation events, to detect bottlenecks and selection, to relate mutations to microenvironment and examine germline influences on invasion and metasatasis. Gradually, we will move to validation studies and hypothesis generation/testing outside GeL

CR15: Non-human genomes and the gut microbiome in colorectal cancers

Cancer genomes are known to contain viral, and sometime bacterial genomes of uncertain significance. In colorectal cancer, for example, JC virus has long been mooted as a causal agent. The gut microbiota is also plausibly a potent risk for factor for gastrointestinal tumours, and different flora can have profound influences on tumour burden in animal models. Whilst we do not propose to sequence gut flora from stool or other gut contents, we shall search for non-human DNA integrated into the cancer DNA. If necessary, we shall work with cross-cutting domains with expertise in this area. We envisage that the non-human genes could be expressed or act as mutagens.

CR16: Molecular profiling of early (T1-2N0M0) colorectal cancers and local lymph nodes

Very early CRCs are increasingly found as bowel cancer screening becomes more common. These are historically a small subgroup and are not fully represented in efforts such as TCGA. A full molecular profile of these tumours would help to answer questions such as which lesions have been found early on their path to rapid metastasis and which are intrinsically indolent. Deep molecular profiling of lymph nodes could allow the detection of occult metastases.

CR17: Determining the effects of neo-adjuvant chemo/radiation therapy on colorectal cancer

Although surgical resection remains the initial treatment for most colorectal cancer cases, neo-adjuvant therapy is increasingly used, for example where surgery is made technically easier. It is known that such therapy can produce profound responses in some patients, with a spectrum from good response to progression in others. Although the effects of neo-adjuvant therapy are principally on local control and preservation of bowel function rather than overall survival, the fact that neo-adjuvant therapy can be so effective strongly suggests that a better understanding of how and when it works has the potential to improve its use. This is especially true given recent technical advances, whether in radiotherapy delivery or in new agents. For example, are the cancer cells remaining after radiotherapy unscathed by treatment, or are they grossly mutated or chromosomally rearranged? The choice of secondary therapies would depend greatly on answering such questions.

CR18: Identifying and characterising intrinsic and extrinsic mutation signatures and mutator phenotypes in colorectal cancers

The factors contributing to a cancer's mutation burden and spectrum are potentially many. Using the exceptionally large, high quality data set afforded by GeL, we shall perform a deep analysis as outlined above, with the ultimate aim of explaining all the mutations found in each tumour.

CR19: New therapeutic or imaging targets

This work is largely implicit within other projects, and will presumably be a focus of commercial organisations accessing the GeL data. We will work with these organisations to annotate data and identify the targets with the most potential for clinical use.

CR20: Integrating genomics and colorectal cancer clinical trials

FOCUS4, a multi-arm adaptive trial of second line treatment in metastatic CRC, will be the pathfinder project. Other proposed trials and studies, such as GIOTTO, may emerge and we shall actively seek to collaborate with any new CRC trials, whether of therapy, imaging or prevention. In at least a subset of FOCUS4 patients, biopsies/samples will be taken at multiple stages in the patient pathway, including pre-therapy, after resection (if performed), after chemotherapy, after targeted therapy and so on depending on how many different alternating treatments are used. Monitoring with ctDNA will also be performed, FOCUS4 comprises allocation to one of several treatment arms (e.g. immune checkpoint inhibition, anti-EGFR, aspirin, etc) and a decision will be made depending on the success of sampling to concentrate on a small number of arms or spread the genomics across the whole trial.

CR21: Exceptional or highly unusual colorectal cancer cases (age, germline predisposition, excellent response, previous treatment with radiotherapy or chemotherapy, etc)

We shall examine these specific groups of patients for unusual features, mostly in a hypothesis-driven fashion. For example, do exceptional responders to a particular targeted therapy have unusual mutations in the target genes, are Lynch syndrome cancers polyclonal, do the genomes of very young patients (<30 years) indicate a cryptic Mendelian predisposition or special aetiology?

CR22: Colorectal cancer driver mutations outside the exome

It is likely that enrichment for specific features will be required, rather than an agnostic global analysis, focusing on features with an elevated prior risk of functionality, like the following.
Copy number, e.g. recurrent or focal changes
Translocations/fusion genes/inversions (may be coding but included for completeness)
Indels, e.g. validated recurrent but globally unusual changes involving >10bp
Non-coding RNA
Promoter and UTRs, e.g.miRNA binding sites
Regulatory regions, e.g. binding of specific transcription factors
Conserved regions
Open chromatin
Reactivated pseudogenes
Regions around known cancer driver genes

Multiple strands of evidence are likely to be needed to demonstrate driver status and statistical methods must be adapted to this. Set-based or burden tests may be required owing to high levels of genetic heterogeneity. Some work will be hypothesis driven, e.g. Wnt pathway modulation by mutations affecting binding of CRC-specific transcription factors such as TCF7L2.

CR23: Inherited variation

The genomes sequenced from the constitutional DNA sample will be useful for assessing inherited influences on

(i) susceptibility (in concert with the InCaP domain familial CRC and polyposis cases)
(ii) cancer features such as stage, grade, etc
(iii) prognosis, response to therapy and toxicity
(iv) somatic mutation burden, spectrum, etc
(v) anti-cancer immune response
(vi) driver mutations and (epi)genetic pathways
(vii) other features, such as predilection for specific sites of metastasis

In addition, we will identify undetected Mendelian CRC mutations, perhaps including some mosaics derived from the sequencing of the tumour. These findings are likely to be reported back

to many participants via the Validation and Feedback domain, although in difficult cases ("so-called Type III) these may require functional assessment which we will undertake if feasible (see project below).

CR24: Functional evaluation and interpretation of potential driver mutations in colorectal cancer, and follow-up analyses in additional data sets and model systems

Certain variants that we detect will require (i) validation/replication in additional data sets. (ii) functional effect assessment using multiple approaches including laboratory analysis, and (iii) further studies in cell, organoid and animal models (for example to elucidate pathogenic mechanisms, epistasis, pleiotropy and co-evolution). The CRC GeCIP domain already includes individuals with expertise in these areas, and we shall recruit additional functional biologists as the programme progresses.

(i) We have access to clinical trial data sets such as S-CORT, VICTOR, QUASAR2 and SCOT for validation/replication.

(ii) We strongly believe that a computational approach to variant effect prediction is essential and there already exist excellent tools and databases for this purpose. However, such as approach is limited – not least because the specific functions that need to be deranged to cause cancer are unknown for many mutated genes. We also know that many mutations are cancer type and allele-specific, and we have computational skills and the specific laboratory expertise to perform the necessary assessments for colorectal tumorigenesis.

(iii) We propose that individuals to be recruited would include CRC experts such as Owen Sansom and Inke Nathke.

Detailed research plan – Cancer Main Programme Gear 2 studies

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR01: Identifying critical events in benign to malignant transition through the analysis of colorectal carcinomas arising in polyps**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	Tumours with clear morphological demarcation between regions of benign neoplasia and malignancy, yielding sufficient DNA for sequencing of genomes of each component Note that specialist Histopathological assessment will be required, as distinction between benign and malignant parts of the same lesion can be difficult.
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	Single sample dissected into two parts (will require frozen section to assess morphology if suspected to fulfil inclusion criterion); second sample for additional 'omics analysis would be very helpful
# cores per tumour (if multi-region biopsying proposed)	As above
Follow-up samples following first ascertainment	None
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To identify mutations that drive tumour invasion (i.e. malignancy)
Scientific case and insights that will be gained from this cohort (more details, as indicated)	Genes that drive the progression of a benign precursor, such as a colorectal adenoma, to a malignant carcinoma are largely unknown in most cancer types. In CRC, driver mutations such as p53 are thought to occur close to the benign-malignant transition, although p53 itself is not generally regarded as responsible for invasiveness. Since no single driver of malignancy has yet been found, many other genetic and non-genetic explanations are possible. These could include one or more of the following: multiple mutations in a polygenic model; genomic instability; and changes in the microenvironment. The use of paired benign-malignant samples from the same lesion will provide as controlled a comparison as possible for determining differences between the benign and invasive states.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration).	No

Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	
Is this sub-study a new therapeutic trial?	No
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	No
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR03: Identifying the effects of known risk factors on colorectal cancer genomes**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	Any sporadic colorectal cancer. Reliable data on lifestyle, environmental, demographic and medical risk factors.
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	Single sample; second sample for additional 'omics analysis would be very helpful
# cores per tumour (if multi-region biopsying proposed)	As above
Follow-up samples following first ascertainment	None
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To identify associations of the CRC genetic landscape with the presence or strength of known genetic and non-genetic CRC risk factors (excluding specific predisposing diseases and medications)
Scientific case and insights that will be gained from this cohort (more details, as indicated)	A number of common genetic and environmental/lifestyle risk factors are expected to influence the cancer genome. These range from exposure to mutagens to common SNP alleles. One simple example is that cancers tend to have a high burden of passenger mutations and more C>T changes resulting from cytosine deamination. Many other hypotheses are testable, based on two underpinning notions: (i) mutagens (or epimutagens) will leave detectable signatures of their presence in DNA; and (ii) that non-mutagenic cancer-promoting agents will influence the (micro)environment causing

	selection of different driver mutations. For example, it is plausible that individuals with a high burden of risk SNPs require fewer, or a different spectrum of, driver mutations for a cancer to grow, compared with individuals with few risk SNP alleles; and other cancers may acquire somatic mutations at the same loci at which predisposition SNPs occur if they have low-risk genotypes. Putative cancer-promoting risk factors, such as weight, may cause rapid evolution. This project may provide important insights into how risk factors act and even whether they are causal for cancer or markers of risk.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	No
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR04: Identifying the effects of prior chronic aspirin/NSAID use on colorectal cancers**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	Any sporadic colorectal cancer (of no special aetiology). Reliable data on chronic use of aspirin and other NSAIDs.
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	Single sample; second sample for additional 'omics analysis would be very helpful
# cores per tumour (if multi-region biopsying proposed)	N/A
Follow-up samples following first ascertainment	Relapse samples if available
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To identify whether CRCs developing despite NSAID protection differ from other sporadic CRCs
Scientific case and insights that will be gained from this cohort (more details, as indicated)	NSAIDs are protective against colorectal polyps and cancer. We wish to investigate whether this affects the cancer genome. We might find, for example, find genetic pathways that differ from other sporadic colorectal cancers, less diversity, a different mutation spectrum or no difference at all. We would endeavor to relate any differences to mechanisms of NSAID action, for

	example in organoid models. The underlying reasons for taking NSAIDs are a potential confounder, and hence we will explore whether any NSAID effects vary with the reason for their use (e.g. primary prevention, chronic inflammation outside the GI tract, *etc*). Note: any CRCs with a known or suspected specific aetiology would be excluded from this analysis.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	No
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR05: Colorectal cancers arising on a background of chronic inflammation**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	Cancers and pre-malignant lesions from patients with inflammatory bowel disease and related conditions
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	single sample; second sample for additional 'omics analysis would be very helpful – samples may include polyps, DALMs and other neoplastic lesions
# cores per tumour (if multi-region biopsying proposed)	typically 10 per cancer
Follow-up samples following first ascertainment	where possible, longitudinal sampling over 2-3 years
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To identify mutations that drive carcinogenesis in this patient sub-group. To examine clonal structure and evolution of pre-malignant/malignant lesions in this sub-group.
Scientific case and insights that will be gained from this cohort (more details, as indicated)	Bowel cancers arising on a background of Crohn disease, ulcerative colitis, PSC and related conditions are an important sub-group of tumours that arise from particular precursor lesions and follow molecular pathways that overlap only partially with those of sporadic CRC. Although some data exist on the clonal structure and evolution of these cancers, very sparse whole genome sequencing data are available, and basic knowledge is lacking of: driver mutations and copy number changes;

	clonal complexity and bottlenecking; effects of therapy; causes of malignant progression; mutational processes; and changes over time, including relation to severity of the underlying inflammatory process.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	No
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR06: Profiles of colorectal cancers arising in diabetics**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	Colorectal cancers with no special aetiology arising on a background of type 1 or type 2 diabetes mellitus
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	Single sample; second sample for additional 'omics analysis would be very helpful
# cores per tumour (if multi-region biopsying proposed)	N/A
Follow-up samples following first ascertainment	Relapse samples if available
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To identify whether CRCs developing in diabetics differ from other sporadic CRCs
Scientific case and insights that will be gained from this cohort (more details, as indicated)	Pre-existing diabetes is a moderately strong risk factor for CRC, although certain oral hypoglycaemics may protect against increased risk. It is postulated that the raised risk results from chronically increased IGF signaling, which *prima facie* is tumour-promoting rather than mutagenic. However, several other explanation remain possible. This project will allow a comparison to be made between the genomic landscapes of colorectal cancers arising in type 1 or 2 diabetes are distinct from other sporadic CRCs, providing clues regarding the mechanism of risk. Selected DNA methylomics, gene expression profiling, proteomics and metabolomics would be of great benefit alongside genome sequencing here.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to	No

an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR07: Small bowel cancer genomes**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	Carcinomas of the small bowel (duodenum, jejunum and ileum)
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	Single sample
# cores per tumour (if multi-region biopsying proposed)	N/A
Follow-up samples following first ascertainment	None
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To identify mutations that drive carcinogenesis in this sub-group of bowel cancer patients
Scientific case and insights that will be gained from this cohort (more details, as indicated)	Small bowel cancers are rare. They can arise in inherited conditions such as Lynch syndrome and the various adenomatous polyposis syndromes, but most have no clear genetic basis. Small bowel cancer usually presents late and has a poor prognosis. We know very little about their genomes, from driver mutations to evolution, and this project will perform a comprehensive analysis that may aid in the management of these tumours. Cryptic or undiscovered germline mutations in bowel cancer predisposition genes may also be found.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	No
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR08: Analysis of synchronous primary colorectal cancers and their origins**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	Patients presenting with 2 or more primary colorectal carcinomas (or one carcinoma and one or more severely dysplastic adenomas) in the same region of the large bowel, with macroscopic and microscopic morphological demarcation between lesions. (Mendelian predisposition syndromes excluded). Cancers with "satellite" polyps may also be included.
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	Single sample from each cancer; will also require >1 sample of intervening normal bowel if cancers are in close proximity. Samples of polyps as per inclusion criteria.
# cores per tumour (if multi-region biopsying proposed)	5
Follow-up samples following first ascertainment	None
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To identify the relationships between multiple cancers from the same patient and to search for "field effect" (epi)mutations as an underlying cause
Scientific case and insights that will be gained from this cohort (more details, as indicated)	Multiple primary colorectal cancers can occur in Mendelian predisposition syndromes, but most such cases arise with no clear underlying genetic cause. Since CRC is a common disease, multiple cancers may arise largely by chance in some patients. However, synchronous carcinomas (or a carcinoma and severely dysplastic polyp) in the same region of the large bowel raise the suspicion of some sort of underlying "field defect". A similar factor may cause the well-known phenomenon of a carcinoma with several discrete, but adjacent "satellite" polyp(s). We shall determine whether the tumours from such patients have evidence of a common origin based on their complement of driver and passenger (epi)mutations. We shall also determine whether the normal-appearing bowel between the tumours appears (epi)genetically normal, or is a mosaic, having acquired changes associated with an undetected clonal expansion that has aided carcinogenesis in the synchronous tumours. If detected, these findings would enhance our understanding of colorectal carcinogenesis, raising the possibility of undetected, but potentially pre-cancerous clones, underlying the growth of many sporadic

	CRCs.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	No
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR09: Identifying new molecular sub-types and pathways of colorectal cancer**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	All colorectal cancers and benign polyps
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	All available
# cores per tumour (if multi-region biopsying proposed)	N/A
Follow-up samples following first ascertainment	All available
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To identify molecular groups of colorectal cancer based on unsupervised cluster analyses and machine learning
Scientific case and insights that will be gained from this cohort (more details, as indicated)	Although most of the major CRC driver genes are likely to have been found by groups such as TCGA, rare or weak-effect "mini-driver" mutations in coding regions may remain to be found, and we shall search assiduously for these. However, a more fruitful task may be to refine the current molecular classification of colorectal cancer. Ideally, this would be based on multi-omic (poly- omic?) approaches that can be used should funding be available. Initially, working with the Machine Learning domain, we shall search for new mutation-based CRC groupings beyond the triad of hypermutation, ultramutation and chromosomal instability – and arguably CpG island methylation – that currently holds. A variety of tools will be used including both conventional hierarchical and Kmeans clustering and principal component analysis, and specialist Bayesian methods (e.g. regression, network analysis). The molecular

	pathways (small mutations, copy number changes, *etc*) underlying these cancer groups will be identified and validated in independent data sets. Associations with clinico-pathological variables, such as survival, will be assessed in clinical trial data sets, such as SCOT and FOCUS4.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	No
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR10: Cross-cancer analyses based on shared aetiology or molecular pathways**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	All colorectal cancers not of specific aetiology.
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	Single sample from each cancer
# cores per tumour (if multi-region biopsying proposed)	N/A
Follow-up samples following first ascertainment	None
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To examine genomic similarities and differences between pairs of cancers with shared aetiology and/or molecular features Please note that colorectal polyp-cancer comparisons will be made subject to notification of the status of the GI Polyp GeCIP Domain.
Scientific case and insights that will be gained from this cohort (more details, as indicated)	Many cancers share genetic and other features, and TCGA and others have identified a number of driver genes mutated in more than one cancer type (sometimes referred to inaccurately as "pan- cancer drivers". For colorectal cancer, the natural comparative analyses are with (i) other luminal GI malignancies such as adenocarcinomas of the oesophagus and stomach (some shared driver mutations and forms of genomic instability), and (ii) endometrioid cancers of the uterus and

	ovary (shared genetic aetiology and forms of genomic instability). Many of the analyses detailed elsewhere for CRC will be performed across these cancer types. Exemplar questions of note include whether shared driver mutations tend to occur at similar stage of carcinogenesis, whether there are alternative means of (in)activation of the same pathway in different tissue types, and whether clinicopathological-molecular associations hold across cancers. We will specifically work with the ovarian cancer domain and the prospective upper GI and endometrial cancer domains.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	No
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR11: Identification of actionable mutations in primary and secondary colorectal cancers**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	All colorectal cancers
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	Single sample from each cancer
# cores per tumour (if multi-region biopsying proposed)	N/A
Follow-up samples following first ascertainment	None
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To identify mutations that may influence therapy, and/or are predictive of prognosis
Scientific case and insights that will be gained from this cohort (more details, as indicated)	We shall identify mutations and forms of genomic instability that have potential relevance for patient management. We shall work with Validation and Feedback to provide CRC-specific expertise in this regard. We shall assess selected variants of uncertain pathological significance using more extensive bionformatic assessments and wet lab.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to	No

an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR12: Colorectal cancer profiling in non-European ethnic groups living in the UK**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	All patients who have colorectal cancer of no specific aetiology and who self-identify as of Asian, African or mixed race origins.
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	Single sample
# cores per tumour (if multi-region biopsying proposed)	N/A
Follow-up samples following first ascertainment	None
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To determine whether molecular changes within cancers vary among ethnic groups in the UK
Scientific case and insights that will be gained from this cohort (more details, as indicated)	There is good evidence that some cancers have features that vary with ethnic origin, even when individuals live in the same locale. For example, MSI+ CRCs have been reported to be more common in non-European groups. However, such studies are prone to confounding by factors such as social class. Moreover, some features such as cancer stage at presentation may depend on cultural factors rather than underlying biology. Nevertheless, useful information to enable better treatment of CRC in non-European ethnic groups may be derived from a comparison within GeL. An important factor is that self-reported ethnicity and genetic ancestry can be determined and analysed separately. The analysis may, for example, identify a higher prevalence of certain molecular sub-types, different mutation spectra and signatures, different frequencies of actionable mutations, and an increased prevalence of certain predisposition mutations that are specific to particular ethnicities.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to	No

an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR13: Carcinogenesis in specific morphological or molecular sub-types of colorectal cancer**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	Cancer subtypes fulfilling certain specific criteria, but without special aetiology. Details to be decided, but to include morphological sub-types (e.g. neuroendocrine, mucinous) and molecular sub-types (e.g. ultramutator, "all wildtype" for Ras/Raf/Mek/Erk pathway, "triple negative" for genomic instability)
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	Single sample – a focussed collection to enrich for the chosen subgroups may be required
# cores per tumour (if multi-region biopsying proposed)	N/A
Follow-up samples following first ascertainment	None
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To identify mutations that drive tumorigenesis in sub-groups of CRCs and that may have clinicopathological importance
Scientific case and insights that will be gained from this cohort (more details, as indicated)	Increasingly fine scale classification of cancer will be required if reality is to catch up with the hype of personalised or precision medicine. Breast cancer provides a prime example of how different molecular and morphological subtypes behave differently and require different treatment regimens. For CRC, similar considerations probably apply, but analysis has been more limited (e.g. types of genomic instability, consensus mRNA expression). Although TCGA project has identified major CRC drivers, it is not yet powered to identify subtype-specific drivers. Here, we shall enrich our collection for morphological sub-types (to be chosen). We shall also perform *post hoc* subgroup analyses on specific molecular sub-types. We shall principally aim to identify driver mutations and copy number changes, but will also perform all other standard genome analyses such as assessment of clonality,

	mutation spectra, *etc.* This project will link in with that on molecular pathways described elsewhere.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	No
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR14: Evolution of primary and secondary colorectal cancer in time and space**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	Specific and selected patients from whom several distinct cancer samples can be obtained, to include (i) multiple samples from primary carcinoma, (ii) samples from different metastatic sites and (iii) samples at different times (to include before and after therapy, at presentation, relapse or death, *etc*)
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	These will all have to be decided on a case-by-case basis.
# cores per tumour (if multi-region biopsying proposed)	These will all have to be decided on a case-by-case basis.
Follow-up samples following first ascertainment	These will all have to be decided on a case-by-case basis.
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To determine how colorectal cancers evolve in different natural or artificial environments To identify sub-clonal and spatial driver mutations and copy number changes To examine heterogeneous treatment resistance mechanisms
Scientific case and insights that will be gained from this cohort (more details, as indicated)	A burst of NGS-based studies has transformed cancer evolution analysis from a backwater to mainstream as a result of the excitement it has engendered in the Oncology community. Much remains to be done, however. In part this will consist of more detailed understanding of tumorigenesis in time and space, especially as regards mechanisms of resistance to targeted, genotoxic and immunotherapies, and linking

	the findings into therapeutic strategies and prognostic markers. Our overall strategy in the short term is to continue to describe cancer evolution at the highest possible level of complexity within the GeL project. We wish to use cutting edge statistical methods to identify sub-clones within biopsies, to correct for copy number and tumour cell fraction, to integrate copy number changes into evolutionary trees, to time mutation events, to detect bottlenecks and selection, to relate mutations to microenvironment and examine germline influences on invasion and metasatasis. Gradually, we will move to validation studies and hypothesis generation/testing outside GeL.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	No, although patients in trials are excellent candidates. We expect that these studies will be restricted to GMCs in which there exist Histopathologists with special interests in these analyses.
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR15: Non-human genomes and the gut microbiome in colorectal cancers**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	All colorectal patients
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	Single sample
# cores per tumour (if multi-region biopsying proposed)	N/A
Follow-up samples following first ascertainment	None
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To identify the presence of non-human genomes within cancers and relate that to clinicopathological features
Scientific case and insights that will be gained from this cohort (more details, as indicated)	Cancer genomes are known to contain viral, and sometime bacterial genomes of uncertain significance. In colorectal cancer, for example, JC virus has long been mooted as a causal agent. The gut microbiota is also plausibly a potent risk for factor for gastrointestinal tumours, and different flora can have profound

	influences on tumour burden in animal models. Whilst we do not propose to sequence gut flora from stool or other gut contents, we shall search for non- human DNA integrated into the cancer DNA. If necessary, we shall work with cross-cutting domains with expertise in this area. We envisage that the non-human genes could be expressed or act as mutagens.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	No
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR16: Molecular profiling of early (T1-2N0M0) colorectal cancers and local lymph nodes**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	This is a subgroup analysis of the main programme, although we may request enrichment for such lesions
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	Single sample (and apparently involved LNs), although note that some of these lesions may be mixed adenomas/carcinomas (see separate protocol)
# cores per tumour (if multi-region biopsying proposed)	In a minority of cases, multiple cores may be requested
Follow-up samples following first ascertainment	Recurrent cancers and metastases where possible
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To perform a full molecular profile and cancer evolution analysis of these tumours, especially with respect to timings of mutations and selective sweeps, mutation signatures when the cancer is within a near-normal environment, presence of sub-clonal drivers, presence of morphologically undetectable precursor adenoma and differences between the tumour core and invasive edge.
Scientific case and insights that will be gained from this cohort (more details, as indicated)	Very early CRCs are increasingly found as bowel cancer screening becomes more common. These are historically a small subgroup and are not fully represented in efforts such as TCGA. A full molecular profile of these tumours would help to answer questions such as which lesions

	have been found early on their path to rapid metastasis and which are intrinsically indolent. Deep molecular profiling of lymph nodes could allow the detection of occult metastases.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	No
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR17: Determining the effects of neo-adjuvant chemo/radiation therapy on colorectal cancer genomes**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	All colorectal cancers treated with neo-adjuvant therapy (within or outside clinical trials) from which pre- treatment sample and post-treatment sample are available
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	Most likely biopsies (as many as possible)
# cores per tumour (if multi-region biopsying proposed)	Variable
Follow-up samples following first ascertainment	Post-treatment sample (ideally resection) as long as response is not pathologically complete (good responders may require sequencing at additional depth). On-treatment (samples) if possible.
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To determine the effects of neo-adjuvant therapies on cancer genomes, including their evolution, the identification of resistance (epi)mutations and the identification of post-therapy driver mutations, all in relation to therapeutic response. Non-genetic factors with influences on resistance (stem cell fraction, hypoxia, immune/inflammatory response) will also be assessed *via* RNA.
Scientific case and insights that will be gained from this cohort (more details, as indicated)	Although surgical resection remains the initial treatment for most colorectal cancer cases, neo- adjuvant therapy is increasingly used, for example where surgery is made technically easier. It is known that such therapy can produce profound responses in some patients,

	with a spectrum from good response to progression in others. Although the effects of neo-adjuvant therapy are principally on local control and preservation of bowel function rather than overall survival, the fact that neo-adjuvant therapy can be so effective strongly suggests that a better understanding of how and when it works has the potential to improve its use. This is especially true given recent technical advances, whether in radiotherapy delivery or in new agents. For example, are the cancer cells remaining after radiotherapy unscathed by treatment, or are they grossly mutated or chromosomally rearranged? The choice of secondary therapies would depend greatly on answering such questions.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	No, although it is likely that such alignment will occur.
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR18: Identifying and characterising intrinsic and extrinsic mutation signatures and mutator phenotypes in colorectal cancers**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	All colorectal cancer patients
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	All available
# cores per tumour (if multi-region biopsying proposed)	All available
Follow-up samples following first ascertainment	All available
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To perform a deep analysis of somatic mutations, including differences in: gross features (e.g. copy number *v* SNV); burden/spectrum (e.g. in relation to DNA repair defects or specific environmental aetiologies); signatures; locations with respect to chromatin features (transcribed regions, open chromatin, late/early-replicating, cohesin binding, *etc*);

	their timing during carcinogenesis; driver mutation spectra and selective consequences; underlying genomic instability; clonal structure; DNA modifications (e.g. methylation, atypical bases); replication (origins, strand, use of lesion bypass error-prone polymerases); and several other features. To examine whether any mutations are present in tumour stroma (including clonal TCR rearrangements, *etc*). To relate the findings to cancer behaviour and aetiology.
Scientific case and insights that will be gained from this cohort (more details, as indicated)	The factors contributing to a cancer's mutation burden and spectrum are potentially many. Using the exceptionally large, high quality data set afforded by GeL, we shall perform a deep analysis as outlined above, with the ultimate aim of explaining all the mutations found in each tumour.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	No
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR19: New therapeutic or imaging targets**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	All colorectal cancer patients
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	All as available
# cores per tumour (if multi-region biopsying proposed)	All as available
Follow-up samples following first ascertainment	All as available
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To identify mutations that could provide new targets for therapy, prevention or imaging
Scientific case and insights that will be gained from this cohort (more details, as indicated)	This work is largely implicit within other projects, and will presumably be a focus of commercial organisations accessing the GeL data. We will work with these organisations to

	annotate data and identify the targets with the most potential for clinical use.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	No
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR20: Integrating genomics and colorectal cancer clinical trials**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	Tumours with clear morphological demarcation between regions of benign neoplasia and malignancy, yielding sufficient DNA for sequencing of genomes of each component
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	primary, optionally with samples of other cancer sites
# cores per tumour (if multi-region biopsying proposed)	generally one, but may be more than one for a sub-group of patients
Follow-up samples following first ascertainment	usually up to 5
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To determine how cancer evolve in response to the therapies used within trials, to define molecular sub-types (in part for choice of therapy) and to identify actionable mutations
Scientific case and insights that will be gained from this cohort (more details, as indicated)	FOCUS4, a multi-arm adaptive trial of second line treatment in metastatic CRC, will be the pathfinder project. Other proposed trials and studies, such as GIOTTO, may emerge and we shall actively seek to collaborate with any new CRC trials, whether of therapy, imaging or prevention. In at least a subset of FOCUS4 patients, biopsies/samples will be taken at multiple stages in the patient pathway, including pre-therapy, after resection (if performed), after chemotherapy, after targeted therapy and so on depending on how many different alternating treatments are used. Monitoring with ctDNA will also be performed, FOCUS4 comprises allocation to one of several treatment arms (e.g. immune checkpoint inhibition, anti-EGFR, aspirin, *etc*) and a

	decision will be made depending on the success of sampling to concentrate on a small number of arms or spread the genomics across the whole trial.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	Yes. Phase 4, CR-UK funded, principally UK Yes.
Is this sub-study a new therapeutic trial?	It may result in a trial amendment. It is possible that data from GeL will be used to allocate patients to treatment arms or to bespoke treatment.

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR21: Exceptional or highly unusual colorectal cancer cases (age, germline predisposition, excellent response, previous treatment with radiotherapy or chemotherapy, etc)**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	We may request enrichment for these cancer types, where ascertainment and GeL recruitment are feasible.
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	As many as available
# cores per tumour (if multi-region biopsying proposed)	To be decided (e.g. multiple cores from Lynch syndrome)
Follow-up samples following first ascertainment	Recurrences or second primary cancers
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To identify mutations and other molecular features specific to these classes of patient
Scientific case and insights that will be gained from this cohort (more details, as indicated)	We shall examine these specific groups of patients for unusual features, mostly in a hypothesis- driven fashion. For example, do exceptional responders to a particular targeted therapy have unusual mutations in the target genes, are Lynch syndrome cancers polyclonal, do the genomes of very young patients (<30 years) indicate a cryptic Mendelian predisposition or special aetiology, *etc.*
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner,	No

geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR22: Colorectal cancer driver mutations outside the exome**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	All patients
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	All available samples
# cores per tumour (if multi-region biopsying proposed)	All available samples
Follow-up samples following first ascertainment	All available samples
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To identify and characterise non-coding mutations that drive tumorigenesis
Scientific case and insights that will be gained from this cohort (more details, as indicated)	It is likely that enrichment for specific features will be required, rather than an agnostic global analysis, focusing on features with an elevated prior risk of functionality, like the following. Copy number, e.g. recurrent or focal changes Translocations/fusion genes/inversions (may be coding but included for completeness) Indels, e.g. validated recurrent but globally unusual changes involving >10bp Non-coding RNA Promoter and UTRs, e.g.miRNA binding sites Regulatory regions, e.g. binding of specific transcription factors Conserved regions Open chromatin Reactivated pseudogenes Regions around known cancer driver genes Multiple strands of evidence are likely to be needed to demonstrate driver status and statistical methods must be adapted to this. Set-based or burden tests may be required owing to high levels of genetic heterogeneity. Some work will be hypothesis driven, e.g. Wnt pathway modulation by mutations affecting binding of CRC-specific transcription factors such as

	TCF7L2.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	No
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR23: Inherited variation**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	All colorectal patients
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	As available
# cores per tumour (if multi-region biopsying proposed)	As available
Follow-up samples following first ascertainment	As available
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To identify germline factors that are important for colorectal tumorigenesis
Scientific case and insights that will be gained from this cohort (more details, as indicated)	The genomes sequenced from the constitutional DNA sample will be useful for assessing inherited influences on (i) susceptibility (in concert with the InCaP domain familial CRC and polyposis cases) (ii) cancer features such as stage, grade, *etc* (iii) prognosis, response to therapy and toxicity (iv) somatic mutation burden, spectrum, *etc* (v) anti-cancer immune response (vi) driver mutations and (epi)genetic pathways (vii) other features, such as predilection for specific sites of metastasis In addition, we will identify undetected Mendelian CRC mutations, perhaps including some mosaics derived from the sequencing of the tumour. These findings are likely to be reported back to many participants via the Validation and Feedback domain, although in difficult cases ("so- called Type III) these may require functional assessment which we will

	undertake if feasible (see project below).
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	No
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**CR24: Functional evaluation and interpretation of potential driver mutations in colorectal cancer, and follow-up analyses in additional data sets and model systems**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	All colorectal cancers.
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	N/A
# cores per tumour (if multi-region biopsying proposed)	N/A
Follow-up samples following first ascertainment	N/A
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To perform functional assessment of and follow up selected findings in additional data sets and model systems
Scientific case and insights that will be gained from this cohort (more details, as indicated)	Certain variants that we detect will require (i) validation/replication in additional data sets. (ii) functional effect assessment using multiple approaches including laboratory analysis, and (iii) further studies in cell, organoid and animal models (for example to elucidate pathogenic mechanisms, epistasis, pleiotropy and co-evolution). The CRC GeCIP domain already includes individuals with expertise in these areas, and we shall recruit additional functional biologists as the programme progresses. (i) We have access to clinical trial data sets such as S-CORT, VICTOR, QUASAR2 and SCOT for validation/replication. (ii) We strongly believe that a computational approach to variant effect prediction is essential and there already exist excellent tools

	and databases for this purpose. However, such as approach is limited – not least because the specific functions that need to be deranged to cause cancer are unknown for many mutated genes. We also know that many mutations are cancer type and allele- specific, and we have computational skills and the specific laboratory expertise to perform the necessary assessments for colorectal tumorigenesis. (iii) We propose that individuals to be recruited would include CRC experts such as Owen Sansom and Inke Nathke.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	Yes, e.g. FOCUS4, VICTOR, QUASAR2, SCOT
Is this sub-study a new therapeutic trial?	No

Data and informatics requirements
All GeCIP Domains have contributed to the construction of the data model for their tumour type and have contributed to ongoing efforts within Genomics England to develop the clinical and research informatics infrastructure.

Data access and security	
GeCIP domain name	Colorectal cancer
Project title *(max 150 characters)*	Gear 2 substudies

Applicable Acceptable Uses. *Tick all those relevant to the request and ensure that the justification for selecting each acceptable use is supported above.*

☐X *Clinical care*

☐X *Clinical trials feasibility*

☐X *Deeper phenotyping*

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

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

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

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

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

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

☐X *Other health use - clinical audit*

☐X *Public health purposes*

☐X *Tool evaluation and improvement*

Information Governance

☐X The lead for each domain will be responsible for validating and assuring the identity of the researchers. The lead may be required to support assurance and audit activities by Genomics England.

Any research requiring access to the embassy will be required to complete IG Training and read and sign a declaration form. Access will only be granted once these requirements have been met.