PhD Project Proposal

Funder details

Studentship funded by:	MRC DTP

Project details

Project title:	Computational genomics and epigenomics to understand metastasis

Supervisory team

Primary Supervisor:	Trevor Graham
Associate Supervisor(s):	Annie Baker
Secondary Supervisor:	Janet Shipley Louis Chesler

Divisional affiliation

Primary Division:	Centre for Evolution and Cancer / Division of Molecular Pathology
Primary Team:	Genomics and Evolutionary Dynamics Group
Site:	Sutton

Project background

Metastasis arises because of Darwinian evolutionary selection acting on the milieu of genetic and epigenetic variation caused by DNA damage and genetic instability within cancers. Determining which genetic and epigenetic changes contribute directly to cancer evolution (“drivers”) and which are simply the by-products of genetic instability (“passengers”) is a long-standing and fundamental challenge for cancer biology, and which is compounded in metastasis due to the complexity of metastasis biology. Specifically, the computational genomics tools that have been successfully applied to identify drivers of primary cancers based on mutation frequency, fail in metastasis as they do not account for mutation frequency changes caused by the cell migration and intermixing that defines metastasis. Further, epigenetic changes in cancers remain poorly characterised in general, especially in metastasis, and so unsurprisingly we lack the tools to identify which epigenetic changes drive metastatic disease.

In a recent series of papers, we have shown that constructing mathematical models of DNA damage, genetic instability and evolutionary dynamics during cancer evolution provides a powerful and robust way to make sense of the complexity of the cancer genome and identify subclones that are experiencing clonal selection (Williams, Nature Genetics 2016; Williams, Nature Genetics 2018; Caravagna, Nature Genetics 2020; Househam & Heide, Nature 2022).

In the Graham lab, our recent Cancer Research UK funded programme “Metastatic Evolutionary Dynamics of Colorectal Cancer (MECCA)” has enabled us to generate a very large multi-omic multi-region sequencing dataset from metastatic colorectal cancer patients, which includes single cell sequencing data.

In this project, we will extend our mathematical models to fully describe metastasis evolution, and then apply the models to identify drivers of metastasis evolution in our MECCA dataset. We will then explore metastasis biology in other cancer types (with Shipley) including paediatric malignancy (with Chesler).

Project aims

- *Overall:* Develop and apply a computational framework to identify the genetic and epigenetic drivers of metastasis evolution.

  Specific aims:
- Construct a mathematical model describing the process of cell migration, growth and clonal selection during metastasis evolution.
- Establish a computational framework to simulate sequencing data from the mathematical model that mimics real-world data.
- Perform statistical analysis of the model to identify signatures of evolutionary selection in sequencing data.
- Apply the framework to identify drivers of metastasis in colorectal cancer metastasis sequencing data, and then further datasets as available.

Research proposal

This proposal aligns with the “Genome stability and DNA damage response” MRC theme.

The overall objective of this studentship is to develop computational tools that can be applied to uncover the genetic and epigenetic drivers of metastasis.

We have previously shown that insights from mathematical models of tumour evolution can greatly simplify the interpretation of cancer genomes (Turajlic, Nature Reviews Genetics, 2019). Specifically, constructing models that describe the processes of cell division, DNA damage and genome instability during cancer growth predicts the pattern of intra-tumour heterogeneity that is expected in a tumour under different evolutionary scenarios (Williams, Annual Reviews Genomics, 2018). For example, if a tumour subclone acquires a new driver alteration, it causes a characteristic pattern in the cancer genome by increasing the number of high frequency mutations in the tumour (Williams, Nature Genetics, 2018). Conversely, in the absence of a driver the tumour cell population evolves by neutral drift which produces a characteristic patterning of low frequency mutations (Williams, Nature Genetics, 2016). Further, predation of tumour cells by the immune system patterns the genome in a distinctive fashion (Lakatos, Nature Genetics, 2020). Our machine learning based methods can readily recognise these patterns and infer how a tumour has evolved (Caravagna, Nature Genetics, 2020).

In metastasis, the migration of cells to new metastatic sites, and between metastases, also influence the pattern of intra-tumour heterogeneity within and between metastatic lesions. There is also evidence that the mutation rate changes in metastasis, due to the influence of systemic therapies and potentially also the new tumour microenvironment. Our mathematical models do not currently take this metastasis-specific biology into account, and consequently, we cannot yet identify the evolutionary forces (mutation, selection, genetic drift, migration) that have shaped a metastasis genome.

Further, the role of changes to the epigenome (DNA methylation and chromatin structure) during cancer development has received relatively little attention. Our own recent work highlights that chromatin changes drive colorectal cancer evolution (Heide, Nature, 2022), but still we lack an understanding of how epigenetic changes occur and their role in metastasis biology. We recognise that epigenetic evolution can be modelled analogously to genome evolution.

This mathematical modelling and computational biology-based PhD project will address this shortfall by developing a mathematical framework that describes metastasis evolution. The student will then use their model to predict the pattern of intra- and inter-metastasis heterogeneity, and heterogeneity versus the primary tumour, that is expected under different evolutionary scenarios (clonal selection of a driver alteration, neutral drift, negative selection due to immunoediting, etc). Finally, they will compare their model predictions to patient cancer metastasis data to measure the evolutionary forces, and identify potential driver alterations, that have caused the evolution of the disease.

Detailed approach (approximate timings):

1. (Months 1-18) Construct a mathematical model, based on stochastic branching processes and agent-based modelling, of the process of primary tumour growth and metastatic dissemination, which includes the accrual of genetic and epigenetics alterations and describes the potential function consequence of these mutations on cell fitness.
2. (Months 18-24) Perform statistical analyses of the model, using inferential computational statistics approaches such as Approximate Bayesian Computation, to determine the signatures of clonal selection on the genome, epigenome and pattern of intra-tumour heterogeneity in metastatic tumours.

3. (Months 24-42) Apply the mathematical framework to explore clonal selection in metastatic colorectal cancer (mCRC). The Graham lab has generated a large dataset called “MECCA” (funded by Cancer Research UK) that documents the evolution of mCRC across space and time using multiple different omic technologies (whole genome sequencing, RNAseq, ATACseq, DNA methylation arrays, cyCIF and TCRseq). The genome and epigenome data should be amenable to mathematical analyses. The other datatypes can be then be used to corroborate selection of clones identified by the mathematical analyses, and to identify potential mechanisms (e.g. signalling pathway regulation, immune evasion, etc) underlying the advantage. Further applications of the methodology could be in paedatric malignancy, using data from Chesler’s Stratified Medicine Paediatrics (SMpaeds) study that includes genomic sequencing from primary and recurrent tumours, and in public datasets such as the Hartwig consortium data and MSK-impact cohorts (both pan-cancer metastasis).
4. (Months 42-48) Publish manuscript(s) and complete thesis. We envisage that the thesis will produce a mathematically-led paper describing the methodology and its applicability, and biology-led papers describing the findings in each dataset analysed.

Literature references

[1] Turajlic S, Sottoriva A, GRAHAM TA*, Swanton C*. 2019. Resolving genetic heterogeneity in cancer. Nature Reviews Genetics, 20, 404-416. https://www.nature.com/articles/s41576-019-0114-6
Review about the secret record of tumour evolution written within the pattern of intra-tumour heterogeneity.

[2] Williams MJ, Sottoriva A, GRAHAM TA. 2019. Measuring clonal evolution in cancer with genomics. Annual Reviews Genetics, 20 309-329 https://pubmed.ncbi.nlm.nih.gov/31059289/
Review about how to mathematically decipher the pattern of intra-tumour heterogeneity to measure the dynamics of cancer evolution.

[3] Househam J^, Heide T^, et al. Sottoriva A*, GRAHAM TA*. 2022 Phenotypic plasticity and genetic control in colorectal cancer evolution. Nature, accepted https://www.biorxiv.org/content/10.1101/2021.07.18.451272v1
Shows how computational modelling can detect clonal selection in multi-region sequencing data.

[4] Heide T^, Househam J^, et al. GRAHAM TA*, Sottoriva A*. 2022. The coevolution of the epigenome and genome in colorectal cancer. Nature, accepted
https://www.biorxiv.org/content/10.1101/2021.07.12.451121v1.full
Demonstrates an important role of epigenome change in colorectal cancer development.

[5] Lakatos E, Williams MJ, et al., Sottoriva A*, GRAHAM TA*. 2020. Evolutionary dynamics of neoantigens in growing tumours. Nature Genetics, 52, 1057-1066 https://www.nature.com/articles/s41588-020-0687-1
Mathematical modelling showing how immunediting shapes the cancer genome.

[6] Caravagna G, et al., GRAHAM TA*, Sottoriva A*. 2020. Subclonal reconstruction of cancers using machine learning and population genetics. Nature Genetics, 52, 898-907 https://www.nature.com/articles/s41588-020-0675-5
Development of a machine learning method to measure patterns in the cancer genome caused by evolutionary processes.

[7] Williams MJ, Werner B, Curtis C, Barnes C*, Sottoriva A*, GRAHAM TA*. 2018. Quantification of subclonal selection in cancer from bulk sequencing data. Nature Genetics, 50 895-903
https://www.nature.com/articles/s41588-018-0128-6
Mathematical modelling reveals the signature of positive clonal selection in the cancer genome.

[8] Williams M, Werner B, Barnes C, GRAHAM TA*, Sottoriva A*. 2016. Identification of neutral tumor evolution across cancer types. Nature Genetics, 44, 238-244 https://www.nature.com/articles/ng.3489
Mathematical modelling reveals the signature of neutral drift in the cancer genome.

Candidate profile

Note: the ICR’s standard minimum entry requirement is a relevant undergraduate Honours degree (First or 2:1).

Pre-requisite qualifications of applicants:	BSc or equivalent in a quantitative discipline (maths, physics, computer science, engineering, quantitative biology, etc) with First or 2:1. MSc preferred.
Intended learning outcomes:	• Skills in mathematical modelling of biological systems

- Skills in computational biology, including large simulations and statistical inference
- Skills in bioinformatics analysis, particularly of genomic data
- Knowledge of cancer biology, focusing on metastasis biology and colorectal cancer.
- Scientific writing and presentation skills, particularly in an interdisciplinary setting (maths -> biology)

Advertising details

Project suitable for a student with a background in:

- ☐ Biological Sciences
- ☒ Physics or Engineering
- ☐ Chemistry
- ☒ Maths, Statistics or Epidemiology
- ☒ Computer Science