|  |  |
| --- | --- |
| Job Description |  |
| Data Scientist in Cancer: Research Fellow |  |
| Department: UCL Institute of Health Informatics | Grade: 7 £35,328 - £42,701 inclusive of London Allowance |
| Location: 222 Euston Road, London, NW1 2DA | Hours: 1 FTE (5 days per week) |
| Duration: Available immediately for 2 years with the possibly to extend subject to further funding |  |

#### Reports to:

**LONDON’S GLOBAL UNIVERSITY**

Professor Harry Hemingway, Director, UCL Institute of Health Informatics

Professor Simona Parrinello, UCL Cancer Institute (indirect report)

#### Context

**Who we are**

The Cancer Research UK Edinburgh-UCL Glioma Centre of Excellence is being established to foster the building of a new neuro-oncology community, with integration into our strong cancer research and training infrastructure at both University College London (UCL) and the University of Edinburgh (UoE). We will establish a new translational pipeline built on a deeper biological understanding of glioblastoma biology, translating biological and data-intensive health research and phenotypic-led drug discovery (in disease-relevant and predictive pre-clinical models) via a strong clinical community equipped to perform state-of-the-art adaptive clinical studies.

UCL Institute of Health Informatics has research teams of data analysts and clinical scientists who can harness the move into new digital pathology and data-driven innovation at scale around electronic health records (EHR), image informatics and related disciplines, leveraging the world-class infrastructure for medical informatics.

The Cancer Institute is a £40 million investment in central London based at UCL, one of the world’s top universities and a founding member of the Francis Crick Institute, aimed at developing world-class basic and translational cancer research. Bringing together over 300 academics and clinical cancer researchers, and through its associations with partner NHS Trusts (UCL Hospitals NHS Foundation Trust, Great Ormond Street Hospital NHS Foundation Trust and the Royal Free Hospital NHS Foundation Trust), the Institute has greater clinical links than any comparable centre in the UK, creating a unique opportunity for significant impact on the delivery of clinical service to cancer patients.

**HDR UK**

Health Data Research UK (HDR UK) is the new national institute for data science for health, which was established in 2018 with long-term funding support from research councils, UK Research and Innovation (UKRI), and charitable and governmental research funders. The HDR UK seeks to drive improvements in the health of patients and populations through research at regional and national scale. The triple mission of HDR UK includes:

1. Research spanning precision medicine, trials and public health;
2. Building platforms underlying infrastructure and expert services to enable research at national scale;
3. Developing training opportunities.

To deliver this mission, the five major London universities – UCL (coordinating), Imperial College London, King’s College London, London School of Hygiene &Tropical Medicine and Queen Mary University of London – have come together as a pan-London site. The terms of this partnership are set out in the Site Agreement. The six HDR UK sites include London, Cambridge, Oxford, the Midlands, Scotland and Wales/ Northern Ireland.

**Accountability:** The line manager for this post is Professor Harry Hemingway, Director at UCL Institute of Health Informatics and HDR UK London Director.

Accession number for this post is **HDRUKLondon10\_UCL**.

The post is available immediately for 2 years with the possibly to extend subject to further funding.

**Scientific problem**

The CRUK Glioma Centre of Excellence will provide the foundation for further integration between the partners, allowing for the maximisation of new cross-disciplinary opportunities in data-driven innovation, imaging technologies and growing strengths in phenotypic drug discovery using patient-derived gliobastoma multiforme (GBM) cellular models. In 5-10 years, we aim to have transformed the landscape across the UK for discovery and clinical evaluation of new brain cancer therapeutics.

**Who you are**

The post-holder will ignite new biological- and data-intensive health research and collaborations, through cross-disciplinary technological advances, e.g. in data-driven innovation, imaging and phenotypic screening.

We believe that there is a timely opportunity to derive value from patient-focused clinical data, e.g. determining the response or relapse of any residual tumour after surgery or after drug or radiation therapy.

The post-holder will work closely with their counterpart in Edinburgh to map out numbers of brain tumour patients of different types by age and date across Scotland/London and, in addition, they will build longitudinal trajectories of consented Scottish/London-based sub-cohorts to examine co-morbidities, tumour recurrence, repeat surgery, deaths by cause etc.

It is expected that the appointee will have an ORCID ID and maintain their profile.

#### Main purpose of the job

The primary purpose of this job is to develop large-scale detailed EHR in imaging pathology and genomic resources for glioma research.

The appointee will draw on the wealth of research excellence and translational experience brought together in UCL partner organisations. The post-holder will foster the development of a vibrant culture of translational science and will act as a role model for scientists at all stages in their careers. The successful candidate will be expected to leverage the multidisciplinary culture that characterises Health Data Research UK London, and for which UCL is also renowned.

#### Key Working Relationships

**At UCL:** Professor Harry Hemingway (UCL Institute of Health Informatics, HDR UK London Site) and Professor Simona Parrinello, UCL Cancer Institute and UCL Principal Investigator for Glioma Centre of Excellence.

The informatics and bioinformatics expertise at UCL:

• Dr Spiros Denaxas (structured EHRs, CALIBER, UCL Institute of Health Informatics);

• Professor Richard Dobson (unstructured and structured EHRs, UCL Institute of Health Informatics);

• Dr Wai Keong Wong (Haematological Oncologist and Chief Research Clinical Informatics Officer, UCLH).

**At Edinburgh:** Importantly, the post-holder will be expected to collaborate with Professor Cathie Sudlow, who leads the Health Data Science theme, in the CRUK Glioma Centre of Excellence at Edinburgh University.

**At HDR UK:** It is expected that the appointee contributes towards the emerging theme in data science for cancer and phenomics.

#### Duties and responsibilities

The post-holder will work closely with and be supported in their role by the Principal Investigator. The main responsibilities include:

**Research**

* Develop and validate EHR phenotyping algorithms relevant to understanding the (early) detection, classification and progression of glioma
* Access large-scale EHR resources across UK (e.g. CALIBER and Scotland) and develop and execute reusable analytic approaches
* Develop appropriate tools for analysis (e.g. federated) of EHR directly relevant to emerging hypotheses in relation to e.g. drug repurposing
* Access unstructured health record data at scale e.g. from imaging or pathology reports and develop appropriate analytic pipelines
* Catalogue and control metadata parameters and metadata administration towards the automation of data extraction
* Provide a point of contact to growing collaborations in cancer informatics.

In pursuit of these research responsibilities, the post-holder will work well in an interdisciplinary team and be a self-starter, taking the initiative to develop and lead the development of infrastructure and research projects, with the support of the wider team. S/he will:

* Submit research papers to refereed journals on a regular basis and present findings at national and international conferences
* Develop tools and methods which are FAIR (Findable, Accessible, Interoperable and Reusable)
* Support ongoing research projects within the Institute, providing supervision and advice to other team members and work closely with colleagues at other universities in the Glioma Centre of Excellence
* Contribute to the development of grant applications and manuscripts in collaboration with others in the Institute
* Participate in emerging cancer and phenomics themes in HDR UK as appropriate.

**Teaching, training and supervision**

* Participate in existing and new teaching activities across the field of using EHR for research. This may involve MSc level teaching (e.g. supervision of MSc dissertations), and involvement in the new MSc in Health Data Science at UCL Institute of Health Informatics.
* Developing web-based resources and seminar series e.g. around the HDR UK proposed national repository for EHR phenotypes.

Subject to the regulations of the funding bodies, appropriately qualified members of research staff will be expected to contribute to teaching or other departmental activities. At the present time, this is estimated to amount to an absolute maximum of 30 contact teaching hours per annum for full-time staff, with opportunities to develop and deliver teaching on our Short Course programme, as well as providing informal advice and training to other research staff and students, as appropriate.

**Personal development opportunities**

We will strongly encourage and support the post-holder to develop their career, either through a research route or through a technical specialist/ research software engineer route. This may include personal fellowship applications where appropriate. The post-holder will be expected to formulate a personal development plan and to capitalise on training opportunities at the Institute of Health Informatics and/or through UCL’s Graduate School as appropriate

***This job description reflects the present requirements of the post. As duties and responsibilities change and develop, the job description will be reviewed and will be subject to amendment in consultation with the post-holder.***

# Person specification

| Criteria | Essential or Desirable |
| --- | --- |
| **Qualifications and Education** |  |
| PhD in relevant quantitative discipline in bioinformatics, health data science, information systems engineering, computer science or related subjects or relevant experience | Essential |
| **Skills and abilities** |  |
| Programming skills in at least one of the languages e.g. Perl/ Python/ BASH/ SQL | Essential |
| Advanced skills in at least one statistical software package e.g. R | Essential |
| Electronic health record (EHR) specifications and standards, health information exchange standards, FIHR, SNOMED-CT, ICD-10 codes, ISB standards and controlled clinical terminologies | Desirable |
| Fully executed system development life cycles of Relational Database Management Systems for electronic health records in an SQL environment | Desirable |
| Writing, presenting and explaining technical and/or scientific reports to a wide range of scientific and lay audiences | Essential |
| Skills in natural language processing, particularly as applied to NHS imaging and pathology reports | Desirable |
| Skills in machine learning, particularly as applied to identifying reproducible clusters of disease sub-types | Desirable |
| Peer-reviewed publication track record  | Desirable |
| Managing, linking and analysing large, >10M records, primary and/or secondary health care and administrative datasets  | Desirable |
| Computationally intensive analysis using HPC clusters | Desirable |
| Data visualisation to advance data science projects | Desirable |
| Project management and using relevant tools to support successful delivery of projects | Desirable |
| Teaching at undergraduate or post-graduate level | Desirable  |
| **Experience** |  |
| Working in multi-disciplinary teams and multi-institutional and international collaborations | Desirable |
| Publication record as lead or co-author in relevant area of research | Desirable |
| **Knowledge / Aptitude** |  |
| Experience of information governance, privacy and security issues using confidential NHS health records in a secure setting | Essential |
|  |  |
| Sound understanding of neuro-oncology, in particular brain tumour detection, prognosis, symptoms, diagnosis, aetiology, treatment plan and quality of life and disease classification using ICD-10 codes | Desirable |
| Awareness of current debates in big data health research and electronic health records | Desirable |
| Working knowledge of metadata standards (e.g. DDI, SDMX etc) | Desirable |

# Apply

|  |
| --- |
| To apply for this position visit:ucl.ac.uk/jobs |