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 in January 2015. On the 18th June 2015 we invited the inaugurated GeCIP domains to develop more detailed research plans working closely with Genomics England. 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. Some of you have requested a template for the research plan which we now provide herewith.

We are only expecting one research plan per domain and have designed this form to contain common features with funder application systems to minimise duplication of effort. Please do not hesitate to contact us if you need help or advice.

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 including:

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

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

Application Summary	
GeCIP domain name	**Functional Effects**
Project title *(max 150 characters)*	**Functional Effects Consortium:** Integrated Data for interpretation of individual genomes.

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

The consortium will bring together the internationally excellent research of sixteen diverse UK resources with the world class bioinformatics infrastructure maintained at the European centre for bio-molecular data delivery, EMBL-EBI, to form a Functional Effects Consortium.

1. The consortium will develop new methods to iteratively integrate the predicted genomic, structural, functional, metabolite, protein interaction, and pathway data currently maintained individually by consortium members, with experimental, curated data at EMBL-EBI.

2. These methods will be implemented on the Variant Annotation Platform (VAP), which will be developed by the consortium.

3. VAP data will be integrated with machine-learning predictors, developed by the consortium, to extend and improve the existing Variant Effect Predictor (VEP) tool.

4. Improved VEP and VAP datasets will be used to analyse datasets from the 100,000 Genomes project to facilitate rational selection of strains and help to identify drug targets, aid drug design, and inform therapeutic strategies. VAP and VEP will continuously improve as the consortium partners develop novel analytical and predictive methods.

5. We will promote and disseminate these resources and results for use by the research and clinical community to more accurately predict the impact of individual genomic variants and the relationship to disease progression.

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)*

The UK has possibly the strongest concentration of structural bioinformatics researchers in the world: UK researchers have pioneered developments for predicting protein structures, functional annotations, and the impacts of genetic changes. For these complex and intensive tasks spanning fundamental areas of biophysics and computer science the UK has internationally renowned resources, extensively used in specific research communities. However, currently no resource connects data from this vibrant research ecosystem on protein structure and function to genome variation and clinical practice. Creating a foundational project that leverages existing extensive knowledge about proteins, in combination with the highest quality protein predictions from known diseases, has the potential for huge advances in this field. These data must be integrated with robust tools suitable for high-throughput use and that scale to millions of genomes. Even incremental improvements for access and dissemination of existing tools will have a profound and transformative impact if the full promise of these statistical methods are realised at scale. Building connections within this basic-to-applied consortium will bridge research communities and drive fundamental research, stimulating new ideas.

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

Structural and functional analyses provide critical insights into the probable impact of a variant on human health, but fewer than 10% of proteins have experimental characterisation. The prediction data required to interpret genetic variation exists in disparate resources and seriously hinders queries for large-scale genome analysis.

We will develop a new Variant Annotation Platform (VAP) to integrate predictions from all consortium partners to transform interoperability between annotations and genomic data. The VAP will include systematic updates from consortium data resources and will develop and implement consensus and confidence metrics for predication methods. The VAP platform will significantly improve understanding of the impacts of genetic variation. The platform will also include expression data to highlight conditionally activated pathways and will annotate the specific baseline tissues and cell lines in which protein-coding variants are present, indicating which experiments the gene(s) are involved in. Druggable site information will be integrated with variant information to evaluate impacts on drug binding and resistance.

The VAP will be developed by the Universal Protein Resource (UniProt) team at EMBL-EBI. For each protein we will display known and predicted annotations, and for genetic variants we will show predicted pathogenicity and possible structural and functional impacts. We will expand the visualisation[1] of sequence annotations to allow the highlighting of incorporated data, customised data views, and specialised visualisations from other resources. VAP data will also be available for download and programmatic access for local processing, and all visualisation tools will be publicly accessible.

We will extend the VEP and improve prediction by integrating modern predictors and data from the VAP. The Ensembl Variant Effect Predictor (VEP[2]) is a robust and powerful tool for variant annotation. We will extend the VEP's coding variation analysis capabilities, currently based on SIFT and Polyphen, by accessing data from the VAP and embedding analyses from consortium partner algorithms. This aim leverages VEP's highly configurable and extensible architecture. This strategy consolidates these many different algorithms into a consistent framework for the first time. Where these predictors depend on additional annotations, we will build this into the VAP.

Consortium partners have shown[3] that combining structure-based with predictors outperforms purely sequence-based approaches. We will systematcally explore meta-predictor approaches, starting with relatively naïve methods, then moving to increasingly sophisticated combinations, to enhance prediction accuracy. The input of confirmed disease-causing mutations will provide many true positive disease-causing mutations. We will use frequent non-synonymous variants as true negatives, which cannot be involved in highly penetrant disease, and remain always aware that false positives are particularly undesirable in a clinical context.

Integration of consortium pathogenicity predictors will drive innovation but challenge development of a sustainable infrastructure for scaling from thousands to millions of genomes. We will optimise of the VEP code to handle the greater number of predictors, and strengthen the testing and QC framework around VEP. This will include code profiling to identify bottlenecks, and a new automated test harness. We will convene a workshop to define workflows, file formats and pipelines across the consortium. The “plug-in” architectural design of VEP provides maximal flexibility to facilitate quality control for iterative refinements and improvements.

Expected start date	Dependent on funding, decision due summer 2016

Expected end date	3 years from start date

Lead Applicant(s)	
Name	Dr Ewan Birney
Post	Director
Department	
Institution	EMBL-EBI
Current commercial links	CCTV, Oxford Nanopore Technologies

Administrative Support	
Name	Stacy Knoop
Email	knoop@ebi.ac.uk
Telephone	01223 492645

Subdomain leads		
Name	**Subdomain**	**Institution**
Dr. Colin Campbell	*Functional Effects*	University of Bristol
Prof. Michael Sternberg	*Protein pathogenicity predictors*	Imperial College
Prof. Christine Orengo	Protein structural bioinformatics	University College London
Prof. Janet Thornton	*Steering Committee member*	EMBL-EBI

Detailed research plan

Full proposal (total max 1500 words per subdomain)	
Title *(max 150 characters)*	**Functional Effects Consortium:** Integrated Data for interpretation of individual genomes

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

Other health use – disease diagnosis

Rare diseases: An estimated 1 in 8 people suffer from a rare disease in England, yet in a recent developmental delay study almost 30 percent of rare disease patients were diagnosed using exome sequencing[4]. These diagnoses transform the lives of patients and families as no further medical tests are required and hope of a cure increases. *Cancer:* With lifetime estimates in the UK of 1 in 3 diagnosed with cancer, even modest changes in diagnosis rates or treatment improvements for cancers will have profound impacts on practicing healthcare.

Interpretation and validation of the Genomics England Knowledge Base

Despite the ease of generating genomic sequences, interpretation is far harder. The majority of known severe disease or phenotype-causing mutations are within proteins, but interpreting observed variation involves combining sequence with structure to reveal the conservation and mutability of individual amino acids; their local environment within the structure of the protein;

their interactions with other proteins/DNA or small molecules; their flexibility; the variations in different populations; their pathway context; and their role in specific protein functions. These data are critical for predicting the functional effect of a single mutation. Depressingly, despite the numerous innovations in bioinformatics and structural biology, most clinical diagnoses rely only on genetic data, often using only the most severe changes (such as disabling stop mutations), and limited protein information. The problem is exasperated by the fact that in most organisms only a small proportion of the proteins have been structurally characterised.

Tool evaluation and improvement
This project will provide state-of-the-art derived data for annotating variants and powerful tools for assessing the contexts and impacts of mutations. This will be achieved by integrating the data collected and maintained within the consortium's resources into the new **Variant Annotation Platform (VAP)**. This will include as well the development of methods to iteratively integrate predicted genomic, structural, functional, metabolite, protein interaction, and pathway data with the existing experimental, curated data at EMBL-EBI already maintained in the Variant Effect Predictor (VEP). The VAP and VEP can be applied to any genome to rationalise any phenotype. Our platforms will facilitate rational selection of strains and help to identify drug targets, aid drug design and inform therapeutic strategies. VAP and VEP will continuously improve as FE partners develop novel analytical and predictive methods.

The core of this project is to develop the VAP integration engine and embed the results from VAP into the VEP prediction tool. After these components are developed, we expect these projects to be "business as usual" for the Universal Protien Resource (UniProt) and Ensembl teams respectively. The coordination and deployment of resources will generate new innovations in methods, and we would aim to have "best in class" tools, quite likely being meta-predictors of some kind. The project will also provide a national network of experts combining tools and resources to interpret variants in genomes linked to health and food security, and will link this network to applied researchers.

Education and training of health and public health professionals
We will dedicate researchers from each team who will continuously interact with all partners to integrate their data and give them feedback on analyses of the datasets. This project will hugely benefit the consortium's partners by disseminating their resources much more widely and by giving them direct links to large scale sequencing initiatives around health to enable them to tune and improve their predictors for these important datasets.

The knowledge of and uptake of the platforms will be further increased by other collaborations that FE partners have established with experimental and clinical groups. Furthermore, the aggregated information on the impacts of genetic variants will not only guide interventions for improving health, but will also yield important insights into the structural and other mechanisms by which proteins function and accelerate progress in structural bioinformatics and systems biology.

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.*

A three-year programme of work is described below.

Objective 1 Developing the VAP to integrate predicted and experimental annotations

Structural and functional analyses provide critical insights into the probable impact of a variant, for example, whether mutations destabilise a protein by causing structural changes or affect the charge state that mediates a protein's function. However, fewer than 10% of proteins in the UniProt database have experimental characterisation. The FE partners' existing prediction data has huge potential for interpreting genetic variation, but these are currently dispersed at partners' institutions, making comprehensive high-throughput, large-scales genome analysis queries difficult.

To drive a fundamental change in the interpretation potential we will **develop a new Variant Annotation Platform (VAP)** to integrate predictions from all FE partners with EMBL-EBI resources. This will transform interoperability between innovative annotations and genomics data. Furthermore, we will develop a system to manage regular updates from the consortium members, within the same coordinate system and with harmonised semantics.

Systematic incorporation of predicted annotations from the FE Partners will significantly aid in understanding the impacts of genetic variations. For example, using predicted models, structural coverage of the human proteome can be increased from ~6% to over 60% and, because damaging nsSNPs are frequently found in or close to protein interfaces, predicted interactions from all the FE partners will increase information on protein interactions by twenty-fold or more. We will also include expression data to highlight pathway modules activated in certain conditions. To make this relevant for systems analyses we will annotate the specific baseline tissues and cell lines in which protein-coding variants are present, and indicate which differential experiments the gene(s) are involved in. Importantly, druggable site information will be integrated with variant information to shed light on the specific impact of variants on drug binding sites and possible drug-resistant variants.

As well as extending coverage, the aim of combining the predicted data and developing the VAP is to provide consensus and confidence metrics based on the number of methods in agreement. The VAP concept will be extended at later stages to provide summary annotation and impact prediction pages for non-coding regions as this field develops.

Access to the data and visualisation**:** VAP will be developed by the UniProt team at EMBL-EBI. Information will flow from VAP to and from both UniProt and Ensembl, with careful attribution of the source data. The combined information will be displayed via highly intuitive webpages accessible from both UniProt and Ensembl. VAP pages will improve these two international resources and also provide a showcase and increased access to resources from all partners. For each protein we will display known and predicted annotations (including for alternatively spliced proteins). For existing genetic variants, we will show predicted pathogenicity and possible structural and functional impacts calculated using the consortium partners' methods and the enhanced VEP .

We will expand the visualisation of sequence annotations in a flexible, embeddable JavaScript widget[1] that will display each type of annotation as a linear track above the protein sequence. Researchers will be able to customise the view to show only annotations of interest. The viewer will allow highlighting incorporated data and specialised visualisations from other resources. VAP data will also be available for download and programmatic access for further processing by other groups and through an API so that all consortium visualisation tools can be made available to large user communities.

Objective 2. Integrating algorithms: Extending the VEP and improving prediction by integrating modern predictors and data from the VAP

The Ensembl Variant Effect Predictor (VEP) is a robust and powerful tool for variant annotation. For coding variants the output includes predictions from SIFT, and PolyPhen for human, and whilst these are popular with an extensive track record they are no longer the best performers. To exploit the broad range of modern predictors we will extend the VEP's analysis capabilities by accessing data from the VAP and embedding analyses from the consortium partner algorithms. This aim leverages one of VEP's major strengths: it is highly configurable and extensible. With this strategy, we build on the value of pulling these many different algorithms together and, for the first time, have these consolidated within one consistent framework. Where these predictors depend on additional annotations, we will build this into the VAP. As the VEP is regularly updated with new data, particularly from Ensembl and VAP, keeping data resources synchronised with update cycles from all partners will require calculated planning and coordination across the consortium.

We will also implement a robust meta-predictor. The consortium partners have shown[3] that combining structure-based data in SAAPpred with predictors such as the well established SIFT and PolyPhen outperforms purely sequence-based approaches. We expect a broader meta-predictor, combining information from all partners to show even more enhanced performance. We will use the results of systematic analyses to explore meta-predictor approaches, starting with relatively naïve methods, and exploring increasingly more sophisticated combinations. Here the input of confirmed disease causing mutations will provide many true positive disease causing mutations. We will use reasonably frequent non-synonymous variants, which cannot be involved in highly penetrant disease, as true negatives. Results from these analyses will drive innovations in the individual predictors of each partner. We will take particular care to assess the predictive value of these methods given the screening criteria, aware that false positives are particularly undesirable in a clinical context.

Integration of pathogenicity predictors from all partners will drive innovation but at the same time challenge development of a sustainable infrastructure for scaling from thousands to millions of genomes. We will investigate the necessary optimisations of the VEP code to handle the greater number of predictors, and strengthen the testing and QC framework around VEP. This will include code profiling to identify bottlenecks and a new automated test harness. We will convene a workshop to define workflows, file formats and pipelines across the Function Effect consortium. The “plug-in” architectural design of VEP provides maximal flexibility to facilitate quality control for iterative refinements and improvements.

We have already undertaken preliminary studies[4], using a bespoke approach, to combine predicted structural biology data with genomic information for an improved diagnosis (DDD study Fig. 1 below). We will expand this work to apply our platforms across all genes and all predictors to test our improvements to VEP and the use of data from VAP. The partners will draw immense value from this work by gaining access to these important human datasets and more direct communication with the experimentalists. The feedback they obtain from the validation of their predictions will be important for enabling further research to develop and improve their methods.

Individual data will remain in the secure GeL environment but resulting annotations will be made available to GeL and other GeCIPs. Summary and aggregate data will be exported for analysis.

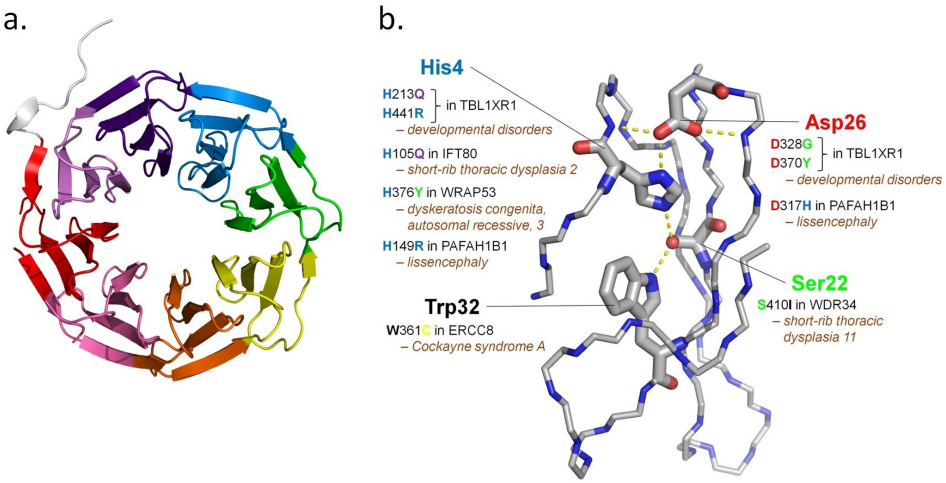

Figure 1. Disease-associated mutations affecting the DHSW tetrad in the WD40 motif of different proteins implicated in rare diseases being studied by the DDD Developmental Diseases Consortium. a) The β-propeller structure of the WD40 domain b) The hydrogen-bonding network (dashed yellow lines) between the DHSW residues of a typical propeller blade which, when disrupted by mutation, can lead to disease - as in the gene/disease examples listed[5].

Some FE consortium partners are pursuing disease-specific predictors, which, given sufficient data, are more accurate than non-specific predictors. The Bristol group have already investigated 17 disease-specific predictors[6] and are devising predictors for estimating disease association for short insertion and deletions events (up to 20 nucleotides). Current developments with SAAPpred also include custom versions for differential phenotype prediction for individual proteins in specific diseases and the use of protein models rather than crystal structures.

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.*

It is a core objective within this project to build a lasting collaborating between EMBL-EBI and the UK's world class research teams.

EMBL-EBI has a track record of building, maintaining and delivering infrastructure for genomics as well as for protein structures, interactions and function (i.e. PDBe, UniProt[7] and Reactome). Below we describe the world-leading resources that will provide annotations for our VAP platform grouped in their subdomains.

Structural biology:
Protein Data Bank in Europe (PDBe) at EMBL-EBI curates experimental protein structural data; FE partners provide SCOP-SUPERFAMILY/CATH-Gene3D structural domain classifications and their sequence families, and Pfam has the best available descriptors of functional domains. For unknown protein structures, the consortium include leaders in 3D structure prediction, who regularly rank high in the international CASP assessment[4] (PHYRE, GenTHREADER, Fugue) or provide comprehensive genome annotations (SUPERFAMILY, Gene3D). The consortium also includes world-leading resources for prediction of secondary structure, transmembrane and disordered regions in proteins (PSIPRED, JPred, DISOPred, D2P2, MEMSAT, Medeller/Memoir); and resources (SAAPdab, PINsnps) predicting whether genetic variations damage protein

structures or interfaces.

Gene/protein function:
Ensembl[8] at EMBl-EBI provides experimental-based gene, regulatory, comparative and variant annotation in chordates. UniProt at EMBL-EBI integrates curated protein information, including phenotypes or diseases associated with variants; functional sites; naturally variant protein forms; RNA editing; proteolytic processing; and post-translational modifications. Together, UniProt and Ensembl provide the most integrated view connecting functional protein data with genetic variants and mutations.

FE Consortium partners also provide resources on predicted function (FFpred, dcGO, Combfunc, FunFams), derived by combining homology data with sources ranging from gene expression data to text mining. All methods were highly ranked in recent international assessments of protein function prediction (CAFA1, CAFA2[9]). The consortium partners will also provide information on known catalytic sites (CSA); predicted functional sites (FunSite, POPScomp); data from unique and internationally renowned resources on pathways (MACiE); protein complexes (3DComplex); and predicted pathway and protein interaction data (SNAPPI, PIPs, FunL, PinSNPs). These complement known protein interaction data from world-leading resources at EMBL-EBI: IntAct and Reactome. The VAP platform will also combine unique partner data on protein drug activity relationships (CanSAR) with world-class EMBL-EBI resources on protein compound and drug interactions (ChEMBL).

Protein pathogenicity predictors:
The FE consortium partners have developed state-of-the-art algorithms that exploit structural and sequence features, conserved sites, protein networks and gene expression data to predict whether a mutation is likely to be pathogenic. The SAAPpred data analysis pipeline exploits a range of sequence and structural analysis data from SAAPdab. SDM, also incorporates environment-specific residue propensities, while mCSM, exploits graph-based signatures, also for protein, nucleic acid and ligand interactions. VarMod determines if variants are located in protein functional regions using binding sites and protein-protein interface sites. SuSPect uses interactome features for a significant improvement in accuracy. The FATHMM tools can predict impacts of variants in non-coding regions. This is important since many variant-disease trait associations are in these regions. All these approaches significantly outperform the 1990 pedigree algorithms commonly used in translational work, such as PolyPhen2 and SIFT (see for example recent references[10,11]).

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

Appropriate training of healthcare professionals on making the most of clinical genomics data will be vital if the UK is going to capitalise fully on the wealth of data emerging from the 100,000 Genomes Project. EMBL-EBI is well placed to orchestrate this, having already participated in Health Education England's task and finish group on clinical bioinformatics, in which we developed competency profiles in clinical bioinformatics for a wide range of clinical roles. EMBL-EBI is also a partner on the recently shortlisted Cambridge application for the HEE Master's in Genomic Medicine. We will ensure that the research outcomes from the Functional Effect GeCIP are rapidly translated into advanced training for the clinical bioinformatics community, through the existing network of clinical bioinformaticians and through EMBL-EBI's input into the new HSST for clinical bioinformatics.

Training will be key to the success of this project, with regular pan-consortium meetings and targeted workshops with the applied genomics community. Ideally, there will be a full time GeCIP trainer on the Train Online team at EMBL-EBI who will organise workshops and develop, with the help of the partners, training workflows to enable biologists and biomedical researchers to exploit VEP and VAP when analysing variants. The trainer will consult broadly with each FE group through visits, workshops and video calls, to advise on and help develop training material content. He or she will work closely with developers of the ELIXIR TeSS training platform to organise the training material into web-based workflows that address major questions being asked – both in industry and in academia, such as, for example, what effect would a mutation at a given position likely have on the stability, activity or function of a specific protein? The workflows will allow biologists to move seamlessly through each step of the analysis to obtain a structure or model of a protein, extract functional site information, and visualising in a 3D viewer to aid interpretation. Training of researchers will be vital if the UK is going to capitalise fully on the wealth of sequencing data. Later in the project we will organise four 2-day workshops to demonstrate our work to relevant researchers.

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

Governance: The FE consortium will be led by Ewan Birney, EMBL-EBI. A steering committee has been established (Birney, Campbell, Orengo, Sternberg, & Thornton) who will oversee the project. We will appoint a project manager to coordinate all scientific activities and manage frequent (at least monthly) all hands teleconferences and an annual meeting of all partners to plan integration of FE tools/resources and discuss innovations. We will promote the network nationally and internationally, and recruit additional groups with appropriate tools or resources as opportunities arise. Longer term, we will work with pharmaceutical companies to explore therapeutic interventions from our novel insights.

The FE consortium members and their teams are experts in their fields. Together they manage key databases and prediction tools across in structural biology that will, when integrated, provide tremendous added value to variant prediction.

FE Consortium members

Dr Ewan Birney	Director (Consortium lead)	EMBL-EBI
Dr Fiona Cunningham	Ensembl, VEP specialist	EMBL-EBI
Dr Maria Martin	UniProt manager	EMBL-EBI
Dr Anne Hersey	ChEMBL manager	EMBL-EBI
Dr Bissan Al-Lazikani	Team Leader	Institute of Cancer Research
Prof. Michael Sternberg	Director of the Centre for Integrative Systems Biology and Bioinformatics (CISBIO);	Imperial College
Prof. Christine Orengo	Professor of Bioinformatics	University College London
Prof. Janet Thornton	Team Leader	EMBL-EBI
Dr. Colin Campbell	Director of the Intelligent Systems Laboratory	University of Bristol
Prof. Geoff Barton	Professor of Bioinformatics	University of Dundee
Prof.Tom Blundell	Research Group Leader	University of Cambridge
Dr Tom Gaunt	Reader in Bioinformatics	University of Bristol

Prof. Julian Gough	Professor of Bioinformatics	University of Bristol
Prof. Charlotte Deane	Professor of Structural Bioinformatics	University of Oxford
Prof. Franca Fraternali	Professor in Bioinformatics and Computational Biology	Kings College London
Dr. Sarah Teichmann	Senior Group Leader	Wellcome Trust Sanger Institute
Dr Mark Wass	Senior Lecturer in Computational Biology	University of Kent
Prof. David Jones	Head of the Bioinformatics Group	University College London
Dr Andrew Martin	Reader in Bioinformatics and Computational Biology	University College London
Dr John Mitchell	Reader of Chemistry	University of St. Andrews

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

As outlined in research plans section, preliminary work using the Variant Effect Predictor (VEP) as part of the Deciphering Development Disorders (DDD) study provided confidence that the presence of a broader range of predictors will provide direct impact on disease diagnosis. We expect to increase the proportion of confidently predicted disease causing mutations by at least 5%. This information then needs to be verified and studied by the clinical genetics groups in GeL and other GeCIPs, but this would provide the potential for at least 2,500 more diagnoses in the rare-disease GeL context. This is a confident lower bound of the expected impact of just this single component.

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

There are four main groups of beneficiaries affected by this proposal:

Healthcare professionals will have another invaluable tool to utilise when investigating rare diseases and the impact on human health of genetic variation, resulting in additional diagnoses of previously undiagnosed conditions. These diagnoses transform the lives of **patient's population** and their families, as no further medical tests are then required and hope of a cure increases. The improved VEP and new VAP tools also have the potential to aid in unravelling mutation significance in cancer genomes. With lifetime estimates in the UK of 1 in 3 diagnosed with cancer, even modest improvements in understanding cancers could lead to significant treatment improvements.

Additionally, the coordination and deployment of resources will provide **Life Science researchers** a "best in class" tool which can be applied to any genome to rationalise any phenotype, yielding important insights into the structural and other mechanisms by which proteins function and accelerate progress in structural bioinformatics and systems biology.

The platforms will facilitate the interpretation of human genomes and help to identify drug targets, aid drug design and inform therapeutic strategies while also providing a national network

of experts to interpret variants in genomes which extends to applied researchers

The **FE consortium partners** will themselves draw immense value from this proposal by gaining access to these important datasets and more direct communication with the experimentalists. The feedback they obtain from the validation of their predictions will be important for enabling further research to develop and improve their methods.

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

There is currently no commercial partner in place but EMBL-EBI has strong links to the pharmaceutical industry through its Industry Programme and an ongoing large scale collaborations in the Centre for Therapeutic Target Validation.

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

1. (wwwdev.ebi.ac.uk/uniprot/features-viewer/).
2. McLaren, W. et al. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinforma. Oxf. Engl. 26, 2069–2070 (2010).
3. Al-Numair, N.S. and Martin, A.C.R. (2013) The SAAP pipeline and database: tools to analyze the impact and predict the pathogenicity of mutations, BMC Genomics, (ISMB 2012 SNP-SIG Special Issue) 14(Suppl 3):S4.
4. Wright, C. F. et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. The Lancet 385, 1305–1314 (2015).
5. Laskowski, R.A., Tyagi, N., Johnson, D., Joss, J., Kinning, E., McWilliam, C., Splitt, M.,, Thornton, J. M., Firth, H. V., DDD study, Wright, C. F. Integrating population variation and protein structural analysis to improve clinical interpretation of missense variation: application to the WD40 domain. Human Molecular Genetics, in press (2016).
6. Shihab HA, Gough J, Mort M, Cooper DN, Day IN, Gaunt TR.
Ranking non-synonymous single nucleotide polymorphisms based on disease concepts.
Human Genomics. 2014 Jun 30;8:11. doi: 10.1186/1479-7364-8-11.
7. Yates, A. et al. Ensembl 2016. Nucleic Acids Res. gkv1157 (2015).
8. Consortium, T. U. UniProt: a hub for protein information. Nucleic Acids Res. 43, D204–D212 (2015).
9. Radivojac, P….Sternberg, M, Gough, J, Jones, D, Orengo, C…Friedberg, I (2013) A Large Scale Evaluation of Protein Function prediction. Nature Methods, 10:221-7.
10. Shihab H, Rogers M, Gough J, Mort M, Cooper D, Day I, Gaunt T, Campbell C. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. Bioinformatics. (2015).
11. Yates, C. M., Filippis, I., Kelley, L. A. & Sternberg, M. J. E. (2014). SuSPect: Enhanced Prediction of Single Amino Acid Variant (SAV) Phenotype Using Network Features. J Mol Biol 426, 2692-2701.

Data requirements

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

We will run the analysis across all VCFs using the Varient Effect Predictor.

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

Data analyses are described in detail above in the main proposal text.

Key phenotype data. *Describe the key classes of phenotype data required for your proposed analyses to allow prioritisation and optimisation of collection of these. (max 200 words)*

We will use ontological terms (HPO) and groupings to determine whether any of the predicted severe variants are associated with disease.

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

We will use the standard GeL approaches, for example consuming standard VCF files.

Tool requirements and import. *Describe any specific tools you require within the data centre with particular emphasis on those which are additional to those we will provide (see attached excel file List_of_Embassy_apps.xlsx of the planned standard tools). If these are new tools you must discuss these with us. (max 200 words)*
We will need a read-only internet connection to the core EMBL-EBI resources, including Ensembl, UniProt, and VAP, ideally executed via a secure connection to EMBL-EBI data centres. All other tools are already available within EMBL-EBI or consortium partner systems.

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

Assuming a read-only connection to EMBL-EBI we will not require additional datasets.

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

The comprehensive variant effect prediction proposed with this breadth of algorithms is processor intensive. Comprehensive annotation of the 4.5 million variants expected from a whole genome analysis currently requires 30-120 minutes of a quad core processor depending on the depth and nature of the annotation required. We expect this time to remain roughly stable as methods to speed up the core algorithms will likely to offset by increasingly deep annotation resources.

For developmental stages of the project, processing a few thousands of genomes, the required resources are modest. However, if the system is expanded to include all of the 100,000 genomes additional computational resource would likely be required.

Omics samples
Analysis of omics samples. *Summarise any analyses that you are planning using omics samples taken as part of the Project. (max 300 words)* None planned.

Data access and security	
GeCIP domain name	**Functional Effects Consortium**
Project title *(max 150 characters)*	Integrated Data for interpretation of individual genomes

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

☐ *Clinical care*

☐ *Clinical trials feasibility*

☐ *Deeper phenotyping*

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

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

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

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

☐ *Non hypothesis driven R&D - health*

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

√ *Other health use - clinical audit*

☐ *Public health purposes*

√ *Tool evaluation and improvement*

Information Governance

☐ 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.

Other attachments

Attach other documents in support of your application here including:

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

Dr John Blayney Owen MITCHELL

Addresses and Contact Information:

Dr John Mitchell, EaStCHEM School of Chemistry and Biomedical Sciences Research Complex, University of St Andrews, North Haugh, St Andrews, Fife, Scotland KY16 9ST.

E–mail: jbom@st-andrews.ac.uk **Phone:** 01334 467259
Web: http://chemistry.st-andrews.ac.uk/staff/jbom/group/
Born: 14th April 1965 (London) **Nationality:** British

Academic Positions:

University of St Andrews, UK; Reader in Translational Biology, School of Chemistry, 1.8.09-

University of Cambridge, UK; Lecturer in Molecular Informatics, Department of Chemistry, 1.9.00-31.7.09. Fellow of Churchill College (1.11.00 – 31.7.09)

Birkbeck College, University of London, UK; Temporary Lecturer in Crystallography (one year appointment); 31.3.96-30.3.97.

Postdoctoral Positions:

University College London, Chemistry (1.12.98 – 31.8.00)
Research Fellow in the group of Dr Sally Price.
University College London, Biochemistry (1.1.91 – 30.3.96 & 1.4.97 – 30.11.98)
Research Fellow in the group of Prof. Janet Thornton.

Further Education:

Churchill College, Cambridge University (UG 1984–1987, PG 1987–1990)
Entrance Scholarship in Natural Sciences
First Class Honours in Tripos Part II Chemistry, 1987. Graduated B.A., 1987.
Ph.D. in Theoretical Chemistry under the supervision of Dr Sally Price, 1987-1990.

Principal Research Interests:

- Protein target prediction for druglike molecules
- Classification and computer-based representation of enzyme reaction mechanisms
- Empirical knowledge-based potentials for protein-ligand interactions
- Protein function prediction by machine learning
- Informatics and molecular evolution
- Prediction of solubility
- Modelling of the crystalline state for organic molecules

Research Output:

91 publications in leading peer-reviewed journals, generating 3259 citations and an H-index of 32 (32 papers cited 32 or more times; Google Scholar). Author of #1 most accessed article in *Molecular Pharmaceutics* for Jan-Mar 2008. iGEM Gold Medallist 2010, 2011 & 2012.

Speaker at: CompLife 2006, Cambridge; British Crystallographic Association, York, 2008; NSCCS, London, 2008; Improving Solubility, London, 2008; ScotCHEM Computational Chemistry Symposium, Edinburgh 2009; SULSA annual symposium, Edinburgh, 2009; Improving Solubility, London, 2009; ADMET, London, 2009; Solvation of Bioactive Compounds, Leipzig, 2010; ADMET Europe, Munich, 2010; UK-QSAR, Glasgow, 2010; Physical Chemical Aspects of Biomolecular Solvation, Leipzig, 2011, Computational Chemogenomics, Geneva, 2012; COST CM-1103 working group, Madrid, 2012; Discovery Chemistry, Munich, 2013; Open Drug Discovery, Strathclyde, 2013; Life Sciences seminar, Dundee, 2015.

Research Sponsors since 2000:

EPSRC, BBSRC, Government of Chile, Swedish Research Council, Unilever, Pfizer, Avecia, Arrow Therapeutics, Insight Faraday, Gates Cambridge Trust, SULSA, Wellcome Trust, World Anti-Doping Agency, Cambridge Crystallographic Data Centre.

Relevant Recent Publications:

HY Mussa, JBO Mitchell & RC Glen, *A note on utilising binary features as ligand descriptors*, **J Cheminformatics, 7:58 (2015)** http://dx.doi.org/10.1186/s13321-015-0105-3

HY Mussa, JBO Mitchell & AM Afzal, *The Parzen Window method: In terms of two vectors and one matrix*, **Patt Recog Letts, 63, 30-35 (2015)** http://dx.doi.org/10.1016/j.patrec.2015.06.002

JL McDonagh, T van Mourik & JBO Mitchell, *Predicting Melting Points of Organic Molecules: Applications to Aqueous Solubility Prediction Using the General Solubility Equation*, **Molecular Informatics 24, 715-724 (2015)** http://dx.doi.org/10.1002/minf.201500052

W Kew & JBO Mitchell, *Greedy and Linear Ensembles of Machine Learning Methods Outperform Single Approaches for QSPR Regression Problems*, **Molecular Informatics 34, 634-647 (2015)** http://dx.doi.org/10.1002/minf.201400122

K Nikolic, L Mavridis, OM Bautista-Aguilera, J Marco-Contelles, H Stark, M do Carmo Carreiras, I Rossi, P Massarelli, D Agbaba, RR Ramsay & JBO Mitchell, *Predicting targets of compounds against neurological diseases using cheminformatic methodology*, **Journal of Computer-Aided Molecular Design, 29, 183-188 (2015)** http://dx.doi.org/10.1007/s10822-014-9816-1

RG Alderson, D Barker & JBO Mitchell, *One origin for metallo-β-lactamase activity, or two? An investigation assessing a diverse set of reconstructed ancestral sequences based on a sample of phylogenetic trees,* **Journal of Molecular Evolution, 79, 117-129 (2014)** http://dx.doi.org/10.1007/s00239-014-9639-7

L De Ferrari & JBO Mitchell, *From sequence to enzyme mechanism using multi-label machine learning*, **BMC Bioinformatics, 15:150 (2014)** http://dx.doi.org/10.1186/1471-2105-15-150

N Nath, JBO Mitchell & G Caetano-Anollés, *The Natural History of Biocatalytic Mechanisms*, **PLoS Comput Biol, 10, e1003642 (2014)** http://dx.doi.org/10.1371/journal.pcbi.1003642

D Barker, DEK Ferrier, PWH Holland, JBO Mitchell, H Plaisier, MG Ritchie & SD Smart, *4273π: Bioinformatics education on low cost ARM hardware*, **BMC Bioinformatics, 14:243 (2013)** http://dx.doi.org/10.1186/1471-2105-14-243

L Mavridis, N Nath & JBO Mitchell, *PFClust: a novel parameter free clustering algorithm*, **BMC Bioinformatics, 14:213 (2013)** http://dx.doi.org/10.1186/1471-2105-14-213

L Mavridis & JBO Mitchell, *Predicting the protein targets for athletic performance-enhancing substances*, **J Cheminformatics, 5:31 (2013)** http://dx.doi.org/10.1186/1758-2946-5-31

HY Mussa, JBO Mitchell & RC Glen, *Full "Laplacianised" posterior naive Bayesian algorithm*, **J Cheminformatics, 5:37 (2013)** http://dx.doi.org/10.1186/1758-2946-5-37

RG Alderson, L De Ferrari, L Mavridis, JL McDonagh, JBO Mitchell & N Nath, *Enzyme Informatics,* **Current Topics in Medicinal Chemistry, 12, 1911-1923 (2012)** http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3605803/

PJ Ballester, M Mangold, NI Howard, RL Marchese Robinson, C Abell, J Blumberger & JBO Mitchell, *Hierarchical Virtual Screening for the Discovery of New Molecular Scaffolds in Antibacterial Hit Identification,* **J. R. Soc. Interface, 9, 3196-3207 (2012)** http://dx.doi.org/10.1098/rsif.2012.0569

DS Palmer, JL McDonagh, JBO Mitchell, T van Mourik & MV Fedorov, *First-Principles Calculation of the Intrinsic Aqueous Solubility of Crystalline Druglike Molecules*, **J. Chem. Theor. Comput., 8, 3322-3337 (2012)** http://dx.doi.org/10.1021/ct300345m

N Nath & JBO Mitchell*, Is EC Class Predictable from Reaction Mechanism?* **BMC Bioinformatics, 13:60 (2012)** http://dx.doi.org/10.1186/1471-2105-13-60

R Lowe, HY Mussa, F Nigsch, RC Glen & JBO Mitchell, *Predicting the Mechanism of Phospholipidosis*, **J. Cheminformatics, 4:2 (2012)** http://dx.doi.org/10.1186/1758-2946-4-2

Mark N Wass

Personal Information

• Current position:
Senior Lecturer in Computational Biology, University of Kent , 2012

• Contact information:
Stacey Building
School of Biosciences
University of Kent
+44 (0)1227 827627
m.n.wass@kent.ac.uk

Education

• Imperial College London, UK. 2004-2008
- PhD in structural Bioinformatics.

• Imperial College London, UK. 2000-2001
- MSc in Computing Science.

• University of Cambridge, UK. 1997-2000.
- BA(Hons) Natural Sciences (Biochemistry).

Previous Positions

• 2011-2012 FEBS Long Term Fellow, Spanish National Cancer Research Centre (CNIO), Spain.
• 2008-2008. Postdoctoral researcher, Imperial College London, UK.
• 2002-2004. Technical Consultant, IFS UK.

Publications

1. Wong K.A.*, **Wass M.***, Thomas K. (2015)The Role of Protein Modelling in Predicting the Disease Severity of Cystinuria. European Urol doi: 10.1016/j.eururo.2015.10.039. *joint 1st authors.

2. Kelley L.A., Mezulis, S., Yates, C.M., **Wass, M.N**., Sternberg, M.J.E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* 10:845-858.

3. Michaelis, M. et al., **Wass, M.N**., Cinatl, J. Jr. (2015) Identification of flubendazole as potential anti-neuroblastoma compound in a large cell line screen. *Sci Rep* 5:8202..

4. Friedberg, I, **Wass, M.N**., Mooney, S.D., Radivojac, P. (2015) Ten Simple Rules for a Community Computational Challenge. PLOS Comp Biol 11:e1004150.

5. Wong, K.A., Mein, R., **Wass, M.N**., Flinter, F., Pardy, C., Bultitude, M., Thomas, K. (2015) The genetic diversity of cystinuria in a UK population of patients. *BJU Int*, 116:109-116.

6. Pappalardo M. and **Wass M.N.** (2014) VarMod: modelling the functional effects of non-synonymous variants. *Nucleic Acids Res*. 42:W331-6.

7. **Wass, M.N.**, Mooney, S.D., Linial, M., Radivojac, P., Friedberg, I. (2014) The automated function prediction SIG looks back at 2013 and prepares for 2014. *Bioinformatics*

8. Talman et al., (2014) Proteomic analysis of the Plasmodium male gamete reveals the key role for glycolysis in flagellar motility. *Malar J* 13:315.

9. Chambers, J.C., et al., (2014) The South Asian Genome. *PLOS one* 9:e102645.

10. Radivojac et al.,(2013) A large-scale evaluation of computation function prediction. *Nature Methods*. 10(3):221-7.

11. **Wass, M.N.,** Barton, G., Sternberg M.J. (2012) CombFunc: predicting protein function using heterogeneous data sources. *Nucleic Acids Res*. 40:W466-70.

12. **Wass, M.N**., Stanway, R., Blagborough, A., Lal, K., Prieto, J.H., Raine, D., Sternberg, M.J., Talman, A.M., Tomley F., Yates III, J., Sinden, R.E. (2012) Proteomic analysis of Plasmodium in the mosquito: progress and pitfalls. *Journal of Parasitology*. 139, 1131–1145.

13. David, A., Razali, R., **Wass, M.N.§**, Sternberg, M.J.**§** (2012) Protein-protein interaction sites are hot spots for disease-associated non-synonymous SNPs. *Human mutation*, *33*, 359–363. **§joint senior authors**

14. Chambers, J.C., Zhang, W., Sehmi, J., Li, X., **Wass M.N.§** et al., (2011) Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat Genet* 43:1131–1138. **§co-first author**

15. **Wass, M.N**., David A, Sternberg MJE (2011) Challenges for the prediction of macromolecular interactions. *Curr. Opin. Struct. Biol.* 21:382–390

16. **Wass, M.N.,** Fuentes, G., Pazos, F. and Valencia A. (2011). Towards the prediction of protein interaction partners using physical Docking. *Mol Syst Biol*. 7:469

17. Sinden, R.E., Talman A., Marques, S.R., **Wass, M.N**. and Sternberg M.J. (2010) The Flagellum in Malarial Parasites**.** *Curr Opin Microbiol* 13:491–500.

18. **Wass, M.N**., Kelley, L.A. and Sternberg, M.J. (2010). 3DLigandSite: Predicting ligand binding sites using homologous structures. *Nucleic Acids Res.* 38:W469–73.

19. Chambers JC, Zhang W, Lord GM, van der Harst P, Lawlor DA, Sehmi JS, Gale DP, **Wass M.N.**, *et al.*, (2010) Genetic loci influencing kidney function and chronic kidney disease. *Nat Genet*, 42, 373-5.

20. Chambers JC, Zhao J, Terracciano CM, Bezzina CR, Zhang W, Kaba R, Navaratnarajah M, Lotlikar A, Sehmi JS, Kooner MK, Deng G, Siedlecka U, Parasramka S, El-Hamamsy I, **Wass M.N.**, *et al.,* (2009) Genetic variation in SCN10A influences cardiac conduction. *Nat Genet*, 42**,** 149-52.

21. Chambers, J.C., Zhang, W., Li, Y., Sehmi, J., **Wass, M.N.** et al., (2009) Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels. *Nat Genet*, 41, 1170-2

22. **Wass, M.N.** and Sternberg, M.J. (2009) Prediction of ligand binding sites using homologous structures and conservation at CASP8. *Proteins*, 77 Suppl 9:147-51.

23. **Wass, M.N**. and Sternberg, M.J. (2008) ConFunc - Functional Annotation in The Twilight Zone, *Bioinformatics*, 24, 798-806.

24. Gherardini, P.F., **Wass, M.N**., Helmer-Citterich, M. and Sternberg, M.J. (2007) Convergent Evolution of Enzyme Active Sites Is not a Rare Phenomenon, *J Mol Biol*, 372, 817-845

Professor Tom Blundell FRS, FMedSci: Curriculum Vitae

Personal Information

• Current position: Professor Emeritus & Director of Research, University of Cambridge,
2009 - 2021 VRA at University Cambridge, Space and facilities assured.
• Contact information: Department of Biochemistry, Tennis Court Road, Cambridge CB2 1GA

Education

1961 Open Scholarship in Natural Sciences; Oxford University
1964 First Class Honours, Oxford University.
1967 D.Phil. Oxford University.

Previous Positions

1967 College Lecturer, Hertford College; JRF Linacre College, Oxford;
1973 Lecturer, Biological Sciences, Sussex University;
1976-1996 Professor of Crystallography, Birkbeck College, London;
1991-96 DG, AFRC; CEO, BBSRC;
1996, Sir William Dunn Professor of Biochemistry, University of Cambridge;
1996-2009 Head of Biochemistry, Cambridge;
2003-2009 Head, School of Biological Sciences, Cambridge;

Honours

1984 FRS; 1985 Member of EMBO; 1986 Alcon Award; 1987 Gold Medal, Institute Biotechnology; 1987 Krebs Medal, FEBS; 1988 Ciba Medal, Biochemical Society; 1988 Feldberg Prize in Biology and Medicine; 1993 Member Academia Europaea; 1995 Fellow, Indian National Science Academy; 1996 Gold Medal, Society for Chemical Industry; 1997 Knighthood; 1998 First Recipient, Pfizer European Prize for Innovation; 1998 Bernal Medal, Royal Society; 1999 Founding Member, Academy of Medical Sciences; 2005-2008 President UK Biosciences Federation; 2006 Hon Memb, British Biophysical Society; 2006 Honorary Fellow Royal Society Chemistry; 2008 Foreign Member, Third World Academy of Sciences; 2008 Ramachandran Professor, IISc Bangalore; 2009-2011 President, Biochemical Society; 2011, Foreign Member, Chilean Academy of Sciences: Biochemical Society Award: Prize and Lecture; Cambridge Philosophical Society Fellows Prize 2014. Honorary Fellowships, Brasenose and Linacre Colleges, Oxford; Honorary Fellowship, Birkbeck College, London University; Honorary Doctorates from 16 Universities

Industry: companies founded, boards etc

1999 Astex Technology Ltd: Co-founder, 1999-2011 Board Member, Chair SAB, 2013, sold $886million to Otsuka 1996-2004 Celltech: Non-executive Director, Chair SAB, 2005 - UCB Celltech, Science Advisory Board, 1997-1999 SmithKline Beecham: International R&D Board, 1983-1990 & 2014- Pfizer: Consultant in UK and USA, 1987-1990 Parke Davis: US Consultant;
1986-1990, 1996- Abingworth Management Ltd, Science Advisory Board, 2008-Syntaxin/Ipsen: member of SAB, 2008-2014 Isogenica, Scientific Consultant,

Public Bodies, Charities etc: Research Councils: Council SERC (1998-2000); AFRC (1985-1990); DG, AFRC (1991-1994); Founding CEO, BBSRC, 1994-1996. Non-executive Chairman of BBSRC: July 2009 - 2015; Royal Commission on the Environment Chairman (1998-2005); Advisory Council on Science and Technology, Cabinet Office (ACOST) 1988-1990. Chaired by Margaret Thatcher; Council of Royal Society, 1997-1999; Board Member, Parliamentary Office of Science and Technology, 1997- 2006 Trustee, Daphne Jackson Trust, 1995- 2011; President, Science Council, 2011 - 2016.

Research: Architecture of macromolecules and multi-component assemblies, focus on relation to biological function, knowledge-based prediction of structure & discovery of new proteins & chemical entities. Published >560 papers, >30 in Nature

Publications: Directly related to proposal

Knowledge-based approaches to modelling and homology recognition:

1. Sibanda BL, Sternberg MJE, Thornton JM and Blundell TL (1987) Knowledge-based prediction of protein structures and design of novel molecules **Nature**, 326, 347-352;
2. Sutcliffe MJ, Haneef I, Carney D and Blundell TL (1987) Composer: knowledge-based

modelling of homologous proteins. **Prot. Eng**, 1, 377-384

3. Sali A & Blundell TL. (1993) Modeller: comparative modelling by satisfaction of spatial restraints. **J. Mol. Biol**. 234: 779-815. (~ 8000 citations)
4. Shi J, Blundell TL, and Mizuguchi K (2001) FUGUE: Sequence-structure Homology Recognition Using Environment-specific Substitution Tables and Structure-dependent Gap Penalties. **J. Mol. Biol.** 310, 243-257 (over 1100 citations).
5. Topham, C.M., Srinivasan, N. and Blundell, T.L. SDM: Prediction of stability of protein mutants based on structural environment-dependent amino acid substitution (1997) **Prot. Eng**.10: 7-21.
6. Worth CL, Preissner R & Blundell TL (2011) SDM—a server for predicting effects of mutations on protein stability and malfunction. **Nucleic Acids Research** 39: W215-W222

Machine Learning approaches

7. Pires DEV, Ascher DB, Blundell TL (**2013**) mCSM: predicting the effects of mutations in proteins using graph-based signatures. **Bioinformatics** 30(3):335-342
8. Gossage L, Pires DE, Olivera-Nappa A, Asenjo J, Blundell TL, Eisen T (2014) Classifying VHL missense mutations according to risk Clear Cell Renal Carcinoma **Hum Mol Genet** 23: 5976-88
9. Jafri M, Ascher DB, Pires DEV et al, Blundell TL, Latif F, Maher ER (2015) Germline Mutations in CDKN2B tumor suppressor gene predispose to renal cell carcinoma. **Cancer Disc** 5:723-9.

Databases of protein interactions:

10. Higueruelo A et al. (2009) Profile of Small Molecules Disrupting Protein-Protein Interfaces: the TIMBAL Database. **Chem. Biol. Drug Des**. 74, 457 – 467;
11. Schreyer, A. & Blundell (2009) T. Credo: Protein-ligand database for drug discovery. **Chem Biol Drug Des** 73: 157–167.

Publications: Experimental and Translational Research

Polypeptide hormones, growth factors and Receptors

12. Blundell TL, Cutfield JF, Cutfield SM, Dodson GG, Dodson EJ, Hodgkin DC, Mercola D, Vijayan M (1971) Atomic positions in 2-Zinc insulin crystals. **Nature** 231, 506-511. (226 citations)
13. Sasaki K, Dockerill S, Adamiak DA, Tickle IJ and Blundell TL (1975) X-ray analysis of glucagon and its relationship to receptor binding **Nature** 257, 751-757.
14. McDonald N, Lapatto R, Murray-Rust J, Gunning J, Wlodawer A and Blundell TL (1991) New protein fold revealed by a 2.3Å resolution crystal structure of nerve growth factor. **Nature** 345: 411-414. (452 citations)
15. Pellegrini L, Burke D.F., von Delft F., Mulloy B., Blundell TL (2000) Crystal Structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. **Nature** 407, 1029-1034. (561 citations) PMID: 11069186
16. Hepburn L, Prajsnar TK, Klapholz C, Moreno P, …Blundell TL, Wang L, Ligoxygakis P, Minichiello L Floto RA (2014) Innate immunity. A Spaetzle-like role for nerve growth factor β. **Science** 346(6209): 641-646

Multiprotein systems mediating DNA Repair

17. Pellegrini L, Yu DS, Lo T, Anand S, Lee M, Blundell TL, Venkitaraman AR (2002) Insights into DNA recombination from the structure of a RAD51-BRCA2 complex. **Nature** 420, 287-293
18. Sibanda BL, Chirgadze DY, Blundell TL. (**2010**) Crystal Structure of DNA-PKcs Reveals a Large Open-Ring Cradle. **Nature** 7; 463: 118-121. (125 citations) PMCID: PMC2811870
19. Ochi T, Blackford AN, Coates J, Jhujh S, Mehmood S, Wu Q, Draviam VM, Draviam VM, Robinson CV, Blundell TL, Jackson SP (**2015**) PAXX, a paralog of XRCC4 & XLF, interacts with Ku to promote DNA double-strand break repair. **Science** 347(6218):185-188

Structure-guided drug discovery

20. Foundling SI, Cooper J, Watson FE, Cleasby A, Pearl LH, Sibanda BL, Hemmings A, et al. Blundell TL (1987) X-ray analyses of renin inhibitor complexes. **Nature**, 327, 349-352
21. Lapatto R, Hemmings A, Overington J, Wilderspin A, Wood S, Whittle PJ, Hobart PM and Blundell TL (1989) X-ray analysis of HIV-1 proteinase **Nature** 342, 299-302 (332 citations)
22. Blundell, T.L., Jhoti, H. and Abell, C. (2002). High-Throughput crystallography for lead discovery in drug design. **Nature Reviews Drug Discovery**. 1, 45-54.
23. Scott DE, Ehebauer MT, Pukala T, Marsh M, Blundell TL, Venkitaraman AR, Abell C, Hyvönen M (**2013**) Using a Fragment-Based Approach To Target PPI **ChemBioChem** 14, 332-342.

Geoffrey J. Barton: Curriculum Vitae

Personal Information

• Current position:
Professor of Bioinformatics (from 2001) and Head of Division of Computational Biology (since 2013), School of Life Sciences, University of Dundee

• Contact information:
Division of Computational Biology, School of Life Sciences, University of Dundee,
Dow Street, Dundee DD1 5EH, UK., Tel: +44 (0)1382 385860,
email: gjbarton@dundee.ac.uk, web: www.compbio.dundee.ac.uk

Education

• University of London (Birkbeck College), UK. 1987
- PhD: "Computer Analysis of Protein Sequence and Structure".
• University of Manchester, UK. 1980-1984
- BSc(Hons 2^1) Biochemistry.

Previous Positions

Jan. 1998-Jul. 2001: Head of European Macromolecular Structure Database (now PDBe).
Oct. 1997-Jul. 2001: Research and Development Team Leader, European Molecular Biology Laboratory Outstation - European Bioinformatics Institute (EMBL-EBI) Hinxton, Cambridge, UK.
Apr. 1995-Sep. 1997: Head of Genome Informatics, Wellcome Trust Centre for Human Genetics, University of Oxford, UK.
Oct. 1989-Sep. 1997: Royal Society University Research Fellow in the Laboratory of Molecular Biophysics, Dept. Biochemistry, University of Oxford
Nov. 1987-Sep. 1989: ICRF Research Fellow (ICRF, now CRUK) London, UK

Awards and Honours

2011 - Fellow of the Royal Society of Biology (FRSB)
2006 - Honorary Fellow of James Hutton Institute
1993 – 1995 Junior Research Fellow, Linacre College Oxford, UK
1989 – 1997 Royal Society University Research Fellow
1987 – 1989 Imperial Cancer Research Fund Fellow (London, UK)

Current Grants

1/10/2014 – 30/9/2019: The Jalview Resource for Sequence Analysis and Annotation. BBSRC. £607,407

1/09/2014 – 31/8/2019: Extending the Jalview Resource for protein sequence alignment and analysis (www.jalview.org). Wellcome Trust. £815,000

1/03/2013 – 28/2/2018: The Dundee Resource for protein structure prediction and sequence analysis. BBSRC. £776,000.

1/10/2012 – 30/9/2017: Dermatology and Genetic Medicine: A multidisciplinary research initiative aimed at translating basic science discoveries in genetic skin disease into clinical application. Co-I (PI Irwin McLean; other Co-Is, Irene Leigh, Paul Campbell, Paul Wyatt and J. McGrath). Wellcome Trust Strategic Award: £5.7 Million

1/07/2012 – 31/12/2015: The non-coding Arabidopsis genome. Co-I with Gordon Simpson BBSRC: £998,000.

1/01/2015 – 31/12/2017: Diversifying transcription termination function. Co-I with Gordon Simpson, BBSRC: £778,000.

1/04/2015 – 31/03/2017: The Arabidopsis epitranscriptome. Co-I with Gordon Simpson, BBSRC: £796,000.

Selected Recent Publications

[1-18]

1. Seifert, A., Schofield, P., Barton, G.J., and Hay, R.T., *Proteotoxic stress reprograms the chromatin landscape of SUMO modification.* Sci Signal, 2015. **8**(384): p. rs7.
2. Madeira, F., Tinti, M., Murugesan, G., Berrett, E., Stafford, M., Toth, R., Cole, C., MacKintosh, C., and Barton, G.J., *14-3-3-Pred: Improved methods to predict 14-3-3-binding phosphopeptides.* Bioinformatics, 2015.
3. Gierlinski, M., Cole, C., Schofield, P., Schurch, N.J., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G., Owen-Hughes, T., Blaxter, M., and Barton, G.J., *Statistical models for RNA-seq data derived from a two-condition 48-replicate experiment.* Bioinformatics, 2015.
4. Drozdetskiy, A., Cole, C., Procter, J., and Barton, G.J., *JPred4: a protein secondary structure prediction server.* Nucleic Acids Res, 2015.
5. Schurch, N.J., Cole, C., Sherstnev, A., Song, J., Duc, C., Storey, K.G., McLean, W.H., Brown, S.J., Simpson, G.G., and Barton, G.J., *Improved annotation of 3' untranslated regions and complex loci by combination of strand-specific direct RNA sequencing, RNA-Seq and ESTs.* PLoS One, 2014. **9**(4): p. e94270.
6. Cole, C., Kroboth, K., Schurch, N.J., Sandilands, A., Sherstnev, A., O'Regan, G.M., Watson, R.M., Irwin McLean, W.H., Barton, G.J., Irvine, A.D., and Brown, S.J., *Filaggrin-stratified transcriptomic analysis of pediatric skin identifies mechanistic pathways in patients with atopic dermatitis.* J Allergy Clin Immunol, 2014. **134**(1): p. 82-91.
7. Lyons, R., Iwase, A., Gansewig, T., Sherstnev, A., Duc, C., Barton, G.J., Hanada, K., Higuchi-Takeuchi, M., Matsui, M., Sugimoto, K., Kazan, K., Simpson, G.G., and Shirasu, K., *The RNA-binding protein FPA regulates flg22-triggered defense responses and transcription factor activity by alternative polyadenylation.* Sci Rep, 2013. **3**: p. 2866.
8. Duc, C., Sherstnev, A., Cole, C., Barton, G.J., and Simpson, G.G., *Transcription termination and chimeric RNA formation controlled by Arabidopsis thaliana FPA.* PLoS Genet, 2013. **9**(10): p. e1003867.
9. Sherstnev, A., Duc, C., Cole, C., Zacharaki, V., Hornyik, C., Ozsolak, F., Milos, P.M., Barton, G.J., and Simpson, G.G., *Direct sequencing of Arabidopsis thaliana RNA reveals patterns of cleavage and polyadenylation.* Nat Struct Mol Biol, 2012. **19**(8): p. 845-52.
10. Scott, M.S., Ono, M., Yamada, K., Endo, A., Barton, G.J., and Lamond, A.I., *Human box C/D snoRNA processing conservation across multiple cell types.* Nucleic Acids Res, 2012. **40**(8): p. 3676-88.
11. Miranda-Saavedra, D., Gabaldon, T., Barton, G.J., Langsley, G., and Doerig, C., *The kinomes of apicomplexan parasites.* Microbes Infect, 2012. **14**(10): p. 796-810.
12. Troshin, P.V., Procter, J.B., and Barton, G.J., *Java bioinformatics analysis web services for multiple sequence alignment--JABAWS:MSA.* Bioinformatics, 2011. **27**(14): p. 2001-2.
13. Scott, M.S., Troshin, P.V., and Barton, G.J., *NoD: a Nucleolar localization sequence detector for eukaryotic and viral proteins.* BMC Bioinformatics, 2011. **12**: p. 317.
14. Scott, M.S., Boisvert, F.M., Lamond, A.I., and Barton, G.J., *PNAC: a protein nucleolar association classifier.* BMC Genomics, 2011. **12**: p. 74.
15. Ono, M., Scott, M.S., Yamada, K., Avolio, F., Barton, G.J., and Lamond, A.I., *Identification of human miRNA precursors that resemble box C/D snoRNAs.* Nucleic Acids Res, 2011. **39**(9): p. 3879-91.
16. Scott, M.S., Boisvert, F.M., McDowall, M.D., Lamond, A.I., and Barton, G.J., *Characterization and prediction of protein nucleolar localization sequences.* Nucleic Acids Res, 2010. **38**(21): p. 7388-99.
17. Procter, J.B., Thompson, J., Letunic, I., Creevey, C., Jossinet, F., and Barton, G.J., *Visualization of multiple alignments, phylogenies and gene family evolution.* Nat Methods, 2010. **7**(3 Suppl): p. S16-25.
18. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J., *Jalview Version 2--a multiple sequence alignment editor and analysis workbench.* Bioinformatics, 2009. **25**(9): p. 1189-91.

Curriculum Vitae, Ewan Birney

Full Name: John Frederick William Birney
Date of Birth: 12 December 1972
Nationality: UK
Email: birney@ebi.ac.uk

77 Lancaster Road
London N4 4PL

Employment:

2015-Current: Director, European Bioinformatics Institute
2012-2015: Associate Director, European Bioinformatics Institute
2000-2012: Head of Nucleotide data, European Bioinformatics Institute
Current supervisor for 4 PhD students

On a variety of SAB boards (include Riken Institute, BCGSC, Leipzig MPI, Roslin Institute, IMP)

1996-2000: PhD at the Sanger Centre (Supervisor, Richard Durbin)

Other positions held:

- A number of consultancy contracts, both strategic and technical in the biotech and pharmaceutical industry, including funding and finance orientated roles.
- Equity Research in SBC Warburg Pharmaceutical division (summer 1995).
- Freelance journalist (Economist) (1995).
- Research Assistant at Cold Spring Harbor Laboratory and EMBL Heidelberg.

Prizes and Awards:

EMBO Member, Elected 2012
Winner of the Royal Society's Francis Crick Lecture in 2003
Winner of the Overton Award from the International Computational Biology Society, 2005
Winner of the Benjamin Franklin Award from Bioinformatics.org/BioIT in 2005

Education:

1996-1999: PhD, St John's College Cambridge. Awarded a Scholarship
1992-1996: BA Biochemistry, Balliol College Oxford. 1st Class degree. Awarded a Scholarship

Publications:

180 Peer reviewed publications, 18 in Nature, 7 Science. H-index: 70 (Google Scholar)

Selected Publications:

Keane TM *et al*. (2011) Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*, 477(7364): 289-94.

ENCODE Project Consortium, Myers RM, Stamatoyannopoulos J, Snyder M, Dunham I, Hardison RC, Bernstein BE, Gingeras TR, Kent WJ, **Birney E**, Wold B, Crawford GE. (2011) A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol*, 9(4): e1001046.

Hsi-Yang Fritz M, Leinonen R, Cochrane G, **Birney E**. (2011) Efficient storage of high throughput DNA sequencing data using reference-based compression. *Genome Res.* 21(5): 734-40.

McDaniell R, Lee BK, Song L, Liu Z, Boyle AP, Erdos MR, Scott LJ, Morken MA, Kucera KS, Battenhouse A, Keefe D, Collins FS, Willard HF, Lieb JD, Furey TS, Crawford GE, Iyer VR, **Birney E**. (2010) Heritable individual-specific and allele-specific chromatin signatures in humans. *Science,* 328(5975): 235-9.

Zerbino DR, **Birney E.** (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res,* 18: 821-9.

The ENCODE Project Consortium (**Lead Author** from 308 authors) (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447(7146): 799-816.

Nene *et al* (one of 52 authors). (2007) Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science*, 316(5832): 1718-23.

Rhesus Macaque Genome Sequencing and Analysis Consortium (one of 114 authors) (2007) Evolutionary and biomedical insights from the rhesus macaque genome. *Science*, 316(5822): 222-34.

Birney E, Clamp M, Durbin R. (2004) GeneWise and Genomewise. *Genome Res*, 14(5): 988-95.

Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, **Birney E** *et al*. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915): 520-62.

Lander ES *et al*. (2001) Initial sequencing and analysis of the human genome. *Nature*, 409(6822): 860-921.

Fiona Cunningham: Curriculum Vitae

Personal Information

• Current position:
- Variation Annotation Coordinator, European Bioinformatics Institute, 2011
- Ensembl Variation Project Leader, European Molecular Biology Laboratory, 2008

• Contact information:

European Bioinformatics Institute
European Molecular Biology Laboratory
Wellcome Trust Genome Campus
Hinxton, Cambridge CB10 1SD
United Kingdom
+44 (0)1223 494612 (voice); +44 (0)1223 494494 (fax)
fiona@ebi.ac.uk

Education

• University of Cambridge, UK. 2014
- PhD: "*Bioinformatics tools for Understanding Genomic Variation*".

• University of Exeter, UK. 2000-2001
- MSc in Bioinformatics.

- University of Cambridge, UK. 1997-2000.
- BA(Hons) Natural Sciences (Genetics) / (MA Cantab 2003).

Previous Positions

• 2007-2008. Senior Web Developer and Website Manager, Wellcome Trust Sanger Institute, UK.
• 2004-2007. Senior Web Developer, Wellcome Trust Sanger Institute, UK.
• 2002-2004. Scientific Programmer, Stein Lab, Cold Spring Harbor, NY, USA.
• 2001. Research Internship, deCODE Genetics Inc., Iceland.

Publications

1. **Cunningham**, F., Moore, B., Ruiz-Schultz, N., Ritchie, G.R., and Eilbeck, K. (2015). Improving the Sequence Ontology terminology for genomic variant annotation. J Biomed Semantics *6*.

2. MacArthur, J.A.L., Morales, J., Tully, R.E., Astashyn, A., Gil, L., Bruford, E.A., Larsson, P., Flicek, P., Dalgleish, R., Maglott, D.R., and **Cunningham**, F. (2014). Locus Reference Genomic: reference sequences for the reporting of clinically relevant sequence variants. Nucl. Acids Res. *42*, D873–D878.

3. **Cunningham**, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2014). Ensembl 2015. Nucl. Acids Res. gku1010.

4. Wright, C.F., Middleton, A., Burton, H., **Cunningham**, F., Humphries, S.E., Hurst, J., Birney, E., and Firth, H.V. (2013). Policy challenges of clinical genome sequencing. BMJ *347*, f6845–f6845.

5. Flicek, P., Ahmed, I., Amode, M.R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., et al. (2013). Ensembl 2013. Nucleic Acids Res. *41*, D48–D55.

6. Flicek, P., Amode, M.R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., et al. (2012). Ensembl 2012. Nucleic Acids Res. *40*, D84–D90.

7. Flicek, P., Amode, M.R., Barrell, D., Beal, K., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., et al. (2011). Ensembl 2011. Nucleic Acids Res. *39*, D800–D806.

8. Rios, D., McLaren, W.M., Chen, Y., Birney, E., Stabenau, A., Flicek, P., and **Cunningham**, F. (2010). A database and API for variation, dense genotyping and resequencing data. BMC Bioinformatics *11*, 238.

9. Reese, M.G., Moore, B., Batchelor, C., Salas, F., **Cunningham**, F., Marth, G.T., Stein, L., Flicek, P., Yandell, M., and Eilbeck, K. (2010). A standard variation file format for human genome sequences. Genome Biol. *11*, R88.

10. McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P., and **Cunningham**, F. (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics *26*, 2069–2070.

11. Kohonen-Corish, M.R.J., Al-Aama, J.Y., Auerbach, A.D., Axton, M., Barash, C.I., Bernstein, I., Béroud, C., Burn, J., **Cunningham**, F., Cutting, G.R., et al. (2010). How to catch all those mutations-the report of the Third Human Variome Project Meeting, UNESCO Paris, May 2010. Human Mutation *31*, 1374–1381.

12. Flicek, P., Aken, B.L., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S., et al. (2010). Ensembl's 10th year. Nucleic Acids Res. *38*, D557–D562.

13. Dalgleish, R., Flicek, P., **Cunningham**, F., Astashyn, A., Tully, R.E., Proctor, G., Chen, Y., McLaren, W.M., Larsson, P., Vaughan, B.W., et al. (2010). Locus Reference Genomic sequences: an improved basis for describing human DNA variants. Genome Medicine *2*, 24.

14. Chen, Y., **Cunningham**, F., Rios, D., McLaren, W.M., Smith, J., Pritchard, B., Spudich, G.M., Brent, S., Kulesha, E., Marin-Garcia, P., et al. (2010). Ensembl variation resources. BMC Genomics *11*, 293.

CV – Franca Fraternali

PERSONAL DETAILS

Full Name Franca **Fraternali**
Professor in Bioinformatics and Computational Biology
Randall Division of Cellular and Molecular Biophysics
King's College London
New Hunt's House, Guy's Campus, London SE1 1UL
Tel: +44-(0)20-78486843
E-mail:franca.fraternali@kcl.ac.uk

CURRENT POSITION

2013 - date Professor in Bioinformatics and Computational Biology

QUALIFICATIONS

1991 MSc in Chemistry University of Naples, Italy
1994 PhD Physical Chemistry University of Naples, Italy and ETH Zurich

EMPLOYMENT HISTORY

since April 2013 Professor in Bioinformatics and Computational Chemistry
since 2009 Reader in Bioinformatics, Randall Division King's College London
2005:2009 Lecturer in Bioinformatics, Randall Division
1999:2005 Permanent staff Scientist at the MRC Division of Mathematical Biology NIMR London
1997:1998 Research Associate at the EMBL Heidelberg, Group of Dr. A. Pastore
1995:1997 European Grant at the University of Strasbourg, Group of Prof. G. Wipff
1994:1995 EMBO Fellowship at the ETH Zurich, Group of Prof. W.F. van Gunsteren

GRANTS – RELATED TO GECIP-SB ACTIVITY

2015-2020 BHF 'A comprehensive approach to the genetic, molecular and functional impact of rare titin variants in hypertrophic cardiomyopathy'. Prof. M. Gautel, Prof. F. Fraternali KCL, Prof. P. Elliot UCL GBP 1,175,357; second stage.
2014- 2019 MRC programme grant 'Multi-scale analysis of B-cell response in ageing' Dunn-Walters, D., Coolen, A., Fraternali, F. GBP 1,785,778.
2013- 2017 Leukemia and Lymphoma Research Studentship 'Protein interaction abnormalities in leukaemia: effects of patient-specific genetic variation' Thomas, S., Fraternali, F. GBP 142,588.
2012 - 2015 BHF `Biophysical Investigation of Nesprin Spectrin Repeats: Implications for Cardiovascular Cell Function.' FF(Principal Applicant), Prof. C. Shananan, Dr. S. Garcia Manyes, Dr. M. Pfuhl, King's College London GBP 290,000.

SELECTED RECENT PUBLICATIONS

- Setta-Kaffetzi N, Simpson MA, Navarini AA, Patel VM, Lu HC, Allen MH, Duckworth M, Bachelez H, Burden AD, Choon SE, Griffiths CE, Kirby B, Kolios A, Seyger MM, Prins C, Smahi A, Trembath RC, **Fraternali F**, Smith CH, Barker JN, Capon F. Am J Hum Genet. 2014 May 1;94(5):790-7
- Scharner J, Lu HC, **Fraternali F**, Ellis JA, Zammit PS.. Proteins. 2014 Jun;82(6):904-15.
- Vaz F, Hanenberg H, Schuster B, Barker K, Wiek C, Erven V, Neveling K, Endt D, Kesterton I, Autore F, **Fraternali F**, Freund M, Hartmann L, Grimwade D, Roberts RG, Schaal H, Mohammed S, Rahman N, Schindler D, Mathew CG.
- Buffa P, Romano C, Pandini A, Massimino M, Tirrò E, Di Raimondo F, Manzella L, Fraternali F, Vigneri PG.. FASEB J. 2014 Mar;28(3):1221-36.
- Fornili A, Pandini A, Lu HC, Fraternali F. J Chem Theory Comput. 2013 Nov 12;9(11):5127-5147.
- Carlin LM, Evans R, Milewicz H, Fernandes L, Matthews DR, Perani M, Levitt J, Keppler MD, Monypenny J, Coolen T, Barber PR, Vojnovic B, Suhling K, Fraternali F, Ameer-Beg S, Parker PJ, Thomas NS, Ng T. Sci Signal. 2011 Nov 29;4(201):ra81. doi: 10.1126/scisignal.2001729.

CV - Professor D. T. Jones BSc, ARCS, MSc, PhD

PERSONAL DETAILS

Full name Professor David Tudor **JONES**

Address Dept. of Computer Science and Dept. of Structural and Molecular Biology, University College London, Gower Street, London WC1E 6BT
E-mail: d.t.jones@ucl.ac.uk
Lab URL: http://bioinf.cs.ucl.ac.uk
Centre URL: http://www.bcb.lon.ac.uk

CURRENT POSITIONS

Jan 2008	Co-director, Wellcome Trust 4 Year PhD Programme in Chemical, Computational and Structural Biology, Institute for Structural Molecular Biology
June 2005	Director, Bloomsbury Centre for Bioinformatics
July 2001	Professor of Bioinformatics (Non-clinical HEFCE funded post), jointly between Dept. Of Computer Science and Dept. of Structural and Molecular Biology, University College London

QUALIFICATIONS

1993	PhD Biochemistry (Bioinformatics), University College London
1989	MSc General Biochemistry, Kings College, London
1988	B.Sc. (Hons) Physics, Imperial College, London

HONOURS & AWARDS

1995-1999	Royal Society University Research Fellowship

PREVIOUS POSITIONS

1999 - 2001	Professor of Bioinformatics, Brunel University
1998 - 1999	Reader, Dept. of Biological Sciences, University of Warwick
1995 - 1998	Lecturer, Dept. of Biological Sciences, University of Warwick
1995 - 1999	Royal Society University Research Fellowship, University of Warwick
1993 – 1995	Wellcome Trust Biomathematics Fellowship, UCL

EDITORIAL BOARDS

Proteins: Structure, Function and Bioinformatics
PLoS One
Advances in Bioinformatics
BioData Mining
Intrinsically Disordered Proteins
Faculty of 1000 Head of Protein Folding Section

RESEARCH COMMITTEES SINCE 2000

BBSRC Research Committee E (2014-), MRC Bioinformatics and Neuroinformatics Training and Career Development Panel (2006-2009). BBSRC EBS Panel (2004-2006). Other ad hoc panels include: BBSRC BBR Panel, RCUK Fellowships Panel, BBSRC e-Science Development Fund Panel, EPSRC Advanced Fellowships Panel

PHD STUDENTS AND POSTDOCS SUPERVISED

To date I have supervised 17 PhD students (14 passed, 3 in progress) and 15 postdocs.

GRANTS SINCE 2001

I have held grants with a total value of over £16.7M since 2001, not including fellowships and studentships awarded directly to lab members.

SELECTED RECENT PUBLICATIONS (OUT OF 130)

Kosciolek, T., & Jones, D.T. (2014) De Novo Structure Prediction of Globular Proteins Aided by Sequence Variation-Derived Contacts. PloS one 9 (3), e92197.

Nugent, T., Cozzetto, D., & Jones, D.T. (2014) Evaluation of predictions in the CASP10 model refinement category. Proteins. 82 (S2), 98-111.

Radivojac, P., Clark, W.T., Oron, T., Schnoes, A.M., Wittkop, T., Sokolov, A., Graim, K., Funk, C., Verspoor, K., Ben-Hur, A., Pandey, G., Yunes, J.M., Talwalkar, A.S., Repo, S., Souza, M.L., Piovesan, D., Casadio, R., Wang, Z. Cheng, J., Fang, H., Gough, J., Koskinen, P., Toronen, P., Nokso-Koivisto, J., Holm, L., Cozzetto, D., Buchan, D.W.A., Bryson, K., Jones, D.T. et al. (2013) A large-scale evaluation of computational protein function prediction. Nature Methods. doi:10.1038/nmeth.2340.

Cozzetto, D., Buchan, D.W.A, Bryson, K., & Jones, D.T. (2013) Protein function prediction by massive integration of evolutionary analyses and multiple data sources. BMC Bioinformatics. 14 (Suppl. 3), S1.

Lewis, T.E., Sillitoe, I., Andreeva, A., Blundell, T.L., Buchan, D.W.A., Chothia, C., Cuff, A., Dana, J.M., Filippis, I., Gough, J., Hunter, S., Jones, D.T., Kelley, L.A., Kleywegt, G.J., Minneci, F., Mitchell, A., Murzin, A.G., Ochoa-Montaño, B., Rackham, O.J.L, Smith, J., Sternberg, M.J.E., Velankar, S., Yeats, Orengo, C.A. (2013) Genome3D: a UK collaborative project to annotate genomic sequences with predicted 3D structures based on SCOP and CATH domains. Nucleic. Acid. Res. 41, D499-D507.

Nugent, T. & Jones, D.T. (2012) Accurate de novo structure prediction of large transmembrane protein domains using fragment-assembly and correlated mutation analysis. Proc Natl Acad Sci U S A. 109, E1540-E1547.

Jones, D.T., Buchan, D.W., Cozzetto, D. & Pontil, M. (2012) PSICOV: Precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. Bioinformatics. 28, 184-190.

Lise, S., Buchan, D., Pontil, M. & Jones, D.T. (2011) Predictions of hot spot residues at protein-protein interfaces using support vector machines. PLoS One. e16774.

Nugent, T., Ward, S & Jones, D.T. (2011) The MEMPACK alpha-helical transmembrane protein structure prediction server. Bioinformatics. 27, 1438-1439.

Cuff, A.L., Sillitoe, I., Lewis, T., Clegg, A.B., Rentzsch, R,, Furnham, N., Pellegrini-Calace, M., Jones, D., Thornton, J. & Orengo, C.A. (2011) Extending CATH: increasing coverage of the protein structure universe and linking structure with function. Nucl. Acids Res. 39, D420-D426.

Buchan, D.W., Ward, S.M., Lobley, A.E., Nugent, T.C., Bryson, K. & Jones, D.T. (2010) Protein annotation and modelling servers at University College London. Nucl. Acids Res. 38 Suppl, W563-W568.

Nugent, T. & Jones, D.T. (2010) Predicting transmembrane helix packing arrangements using residue contacts and a force-directed algorithm. Plos Comput. Biol. 6, e1000714604.

Pentony, M. & Jones, D.T. (2010) Modularity of intrinsic disorder in the human proteome. Proteins. 78, 212-221.

Lise, S., Archambeau, C., Pontil, M. & Jones, D.T. (2009) Prediction of hot spot residues at protein-protein interfaces by combining machine learning and energy-based methods. BMC Bioinformatics. 10, 365. Epub.

Nugent, T. & Jones, D.T. (2009) Transmembrane protein topology prediction using support vector machines. BMC Bioinformatics. 10, 159. Epub.

Edwards, Y.J., Lobley, A., Pentony, M.M. & Jones, D.T. (2009) Insights into the regulation of intrinsically disordered proteins in the human proteome by analysing sequence and gene expression data. Genome Biol. 10, R50. Epub.

Lobley, A., Sadowski, M.I. & Jones, D.T. (2009) pGenTHREADER and pDomTHREADER: New Methods For Improved Protein Fold Recognition and Superfamily Discrimination. Bioinformatics. 25, 1761-1767.

BIBLIOMETRIC DATA (AS OF NOVEMBER 2015)

Published papers: 142; Total citations: 27276; Max citations per paper: 3950; H-index: 59.

Curriculum Vitae: Maria J. Martin

Education

1990 Complutense University, Madrid, Spain, MSc Veterinary medicine
2003 Autonoma University, Madrid, Spain, PhD Bioinformatics

Employment history

2009-present	Team Leader Protein Function - Development, EMBL European Bioinformatics Institute (EMBL-EBI), UK
1999-2009	Sequence Database Coordinator, EMBL European Bioinformatics Institute, UK
1996-1999	Database Programmer, EMBL European Bioinformatics Institute, UK
1994-1996	PhD. Fellowship, Biotechnology Center, Autonoma University, Madrid, Spain
1993-1994	PhD. Fellowship, Animal Health Research Center, Valdeolmos, Madrid, Spain
1992-1993	Software Developer, TDI S.A. Research Dept., Madrid, Spain
1990-1992	Research Assistant, Animal Health Research Center, Madrid, Spain
1990-1991	Veterinary Assistant, Private Veterinary Clinic, Madrid, Spain
1988-1990	Medical Residency, Dept. Infectious Diseases, Complutense Veterinary School, Madrid, Spain

Current employment and scientific profile

Maria J. Martin is the Team Leader of the Protein function developments at the EMBL European Bioinformatics Institute (EMBL-EBI) in Cambridge, UK. She is a world expert in the management of database resources with nineteen years of experience in developing protein sequence resources at EMBL-EBI. Since she joined the EMBL-EBI in 1996, she has been involved in the strategic planning, data analysis and development of many database resources in this institute with special focus in protein data and leading the bioinformatics infrastructure of the Universal Protein Resource (UniProt), the world leading database of classified and functionally annotated protein sequences. In this role, she has large experience in developing database software and tools for the UniProt and GO biocurators and researchers. Her interests include the study of novel methods for protein function prediction, and protein annotation and representation, the provision of quality annotation Reference proteomes for the scientific community, and data analysis and visualization of protein data in the context of other –omics data i.e. proteomics and variation. She is key staff of the UniProt Consortium and an active member of the GO and QfO Consortiums. Maria has served a number of scientific advisory committees, and she is Co-PI in a number of grants. She has well-established working relationships with many groups in UK and Consortium members, and a broad network of contacts that include the NCBI in the USA, various sequencing centres and users specializing in a variety of biological domains.

Publications

I have over 70 publications with the UniProt publication in 2012 cited over 3800 times, and 14 publications cited more than 200 times. Some relevant peer-reviewed publications within the last 5 years include:

UniProt Consortium. UniProt: a hub for protein information. Nucleic Acids Res Volume 43 (2015) p.d204-12

Heinzel A, Mühlberger I, Stelzer G, Lancet D, Oberbauer R, Martin M, Perco P. Molecular disease presentation in diabetic nephropathy. Nephrol Dial Transplant Volume 30 Suppl 4 (2015) p.iv17-25

Huntley RP, Sawford T, Mutowo-Meullenet P, Shypitsyna A, Bonilla C, Martin MJ, O'Donovan C Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res Volume 43 (2015) p.d1049-56

Pundir S, Magrane M, Martin MJ, O'Donovan C, UniProt Consortium. Searching and Navigating UniProt Databases. Curr Protoc Bioinformatics Volume 50 (2015) p.1.27.1-1.27.10 DOI: 10.1002/0471250953.bi0127s50

Famiglietti ML, Estreicher A, Gos A, Bolleman J, Géhant S, Breuza L, Bridge A, Poux S, Redaschi N, Bougueleret L, Xenarios I, UniProt Consortium. Genetic variations and diseases in UniProtKB/Swiss-Prot: the ins and outs of expert manual curation. Hum Mutat Volume 35 (2014) p.927-935

Alpi E, Griss J, da Silva AW, Bely B, Antunes R, Zellner H, Ríos D, O'Donovan C, Vizcaíno JA, Martin MJ. Analysis of the tryptic search space in UniProt databases. Proteomics Volume 15 (2015) p.48-57

Barrera A, Alastruey-Izquierdo A, Martín MJ, Cuesta I, Vizcaíno JA. Analysis of the protein domain and domain architecture content in fungi and its application in the search of new antifungal targets. PLoS Comput Biol Volume 10 (2014) p.e1003733

Garcia L, Yachdav G, Martin MJ. FeatureViewer, a BioJS component for visualization of position-based annotations in protein sequences. F1000Res Volume 3 (2014) p.47

Huntley RP, Sawford T, Martin MJ, O'Donovan C. Understanding how and why the Gene Ontology and its annotations evolve: the GO within UniProt. Gigascience Volume 3 (2014) p.4

Sonnhammer EL, Gabaldón T, Sousa da Silva AW, Martin M, Robinson-Rechavi M, Boeckmann B, Thomas PD, Dessimoz C, Quest for Orthologs consortium. Big data and other challenges in the quest for orthologs. Bioinformatics Volume 30 (2014) p.2993-2998

Poux S, Magrane M, Arighi CN, Bridge A, O'Donovan C, Laiho K, UniProt Consortium. Expert curation in UniProtKB: a case study on dealing with conflicting and erroneous data. Database (Oxford) Volume 2014 (2014) p.bau016

Velankar S, Dana JM, Jacobsen J, van Ginkel G, Gane PJ, Luo J, Oldfield TJ, O'Donovan C, Martin MJ, Kleywegt GJ. SIFTS: Structure Integration with Function, Taxonomy and Sequences resource. Nucleic Acids Res Volume 41 (2013) p.d483-9

Gómez J, García LJ, Salazar GA, Villaveces J, Gore S, García A, Martín MJ, Launay G, Alcántara R, Del-Toro N, Dumousseau M, Orchard S, Velankar S, Hermjakob H, Zong C, Ping P, Corpas M, Jiménez RC. BioJS: an open source JavaScript framework for biological data visualization. Bioinformatics Volume 29 (2013) p.1103-1104

Schneider M, UniProt Consortium. UniProtKB amid the turmoil of plant proteomics research. Front Plant Sci Volume 3 (2012) p.270

Eberhardt RY, Haft DH, Punta M, Martin M, O'Donovan C, Bateman A. AntiFam: a tool to help identify spurious ORFs in protein annotation. Database (Oxford) Volume 2012 (2012) p.bas003

Andrew C.R. Martin: Curriculum Vitae

Personal Information

•Current position

- Reader in Bioinformatics and Computational Biology, UCL, 2014–

•Contact information

Institute of Structural and Molecular Biology
Division of Biosciences
Darwin Building
Gower Street
London WC1E 6BT
+44(0)207 679 7034
andrew.martin@ucl.ac.uk

Education

•Christ Church and Laboratory of Molecular Biophysics, University of Oxford. 1986–1990
D.Phil. *Molecular Modelling of Antibody Combining Sites*

•Christ Church, University of Oxford. 1982–1986
BA Honours (First Class) in Biochemistry,

•Therfield Comprehensive, Leatherhead, Surrey. 1975-1982

Previous Positions

- 2005–2014 Senior Lecturer in Bioinformatics, UCL.
- 2004–2005 Lecturer in Bioinformatics, University College London.
- 1999–2003 Lecturer in Bioinformatics, The University of Reading.
- 1998–1999 Temporary Lecturer, University College London and Technical Director, Inpharmatica Ltd.
- 1994–1998 Research Associate, University College London, with Prof. Janet Thornton.
- 1993–1993 Guest scientist, Deutsches Krebsforschungzentrum, Heidelberg, Germany.
- 1990–1994 Self employed scientific and technical software development and contracting.
- 1990–1990 M.R.C. Training Fellowship at National Institute for Medical Research.

Consultancy and Visiting Posts

- 2014–date: Consultant to Rees Consulting AB (antibody humanization).
- 1997–date: Expert witness on various patent disputes including • **Novartis** v Merck, Sharp and Dohme (2013-2014, opposition to patent EP1725261); • Lilly v **Human Genome Sciences** (2007-2009, opposition to patent EP0939804B1 – the first gene patenting case to be tried in the UK courts); • **UCB-Celltech** v MedImmune , plc. (2002-2005, non-payment of royalties for patent US6632927, EP626390, US5859205); • **UCB-Celltech** v Protein Design Labs (2009-2011, Interference 105,705, US Application 10/938,117 v US6180370); • PDL v **Celltech-Chiroscience** (1997-1999, Interference 105,688, US5585089 v US5859205 and Opposition to grant of patent EP451216)
- 2008–2013: Visiting Professor, University of Cagliari, Sardinia
- 1999–2006: Consultant to Inpharmatica, Ltd.

Selected Publications

1. Porter, C.T. and Martin, A.C.R. (2015) **BiopLib and BiopTools - a C programming library and toolset for manipulating protein structure** (Application Note), Bioinformatics, btv482 (epub ahead of print).
2. Martin, A.C.R. (2014) **Viewing multiple sequence alignments with the JavaScript Sequence Alignment Viewer (JSAV)**, *F1000Research* **3,** 249.
3. Martin, A.C.R. (2014) **Structural Biology of Moonlighting - Lessons from Antibodies**, *Biochem Soc Trans* **42,** 1704-1708.
4. Henderson, B. and Martin, A.C.R. (2014) **Protein Moonlighting: A New Factor in Biology and Medicine**, *Biochem Soc Trans* **42,** 1671-1678.
5. Leigh, S.E.A., Lee, Y.P., Whittall, R.A., Dawson, N., Das, S., Martin, A.C.R., Orengo, C.A. and

Humphries, S.E. (2014) **Update and analysis of the UCL Pro-protein convertase subtilisin/kexin Type9 gene (PCSK9) variant database** *Atherosclerosis* **235,** E98-E98.
6. Al-Numair, N.S. and Martin, A.C.R. (2013) **The SAAP database and pipeline: tools to analyze the impact and predict the pathogenicity of mutations**, *BMC Genomics, ISMB 2012 SNP-SIG Special Issue,* **14(Suppl 3)**:S4
7. Jayaram, N., Bhowmick, P. and Martin, A.C.R. (2012) **Germline VH/VL Pairing in Antibodies**, *Protein Engineering Design and Selection* **25**, 523-530.
8. Usifo, E., Leigh, S.E.A., Whittall, R.A., Lench, N., Taylor, A., Yeates, C., Orengo, C.A., Martin, A.C.R., Celli, J. and Humphries, S.E. (2012) **Low Density Lipoprotein Receptor Gene Familial Hypercholesterolemia Variant Database: update and pathological assessment**, *Annals of Human Genetics*, **76**, 387-401.
9. Izarzugaza, J.M.G., McMillan, L.E.M., Baresic, A., Orengo, C.A., Martin, A.C.R. and Valencia, A. (2011) **Characterizing Pathogenic Germline Deviations in Human Protein Kinases**. *BMC Bioinformatics* **10 Suppl 8**, S5.
10. Baresic, A. and Martin, A.C.R. (2011) **Compensated Pathogenic Mutations**, *Biomolecular Concepts* **2**, 281-292.
11. Abhinandan, K.R. and Martin, A.C.R. (2010) **Analysis and Prediction of VH/VL Packing in Antibodies** *Protein Engineering Design and Selection*, **23**, 689-697.
12. Baresic, A. McMillan, L.E.M., Rogers, H.H., Hurst, J.M. and Martin, A.C.R (2010) **Compensated pathogenic deviations: analysis of structural effects**, *Journal of Molecular Biology*, **396**, 19-30.
13. Thullier, P., Huish, O., Pelat, T. and Martin, A.C.R. (2010) **The humanness of macaque antibody sequences**, *Journal of Molecular Biology*, **396**, 1439-1450.
14. Izarzugaza, J.M.G., McMillan, L.E.M., Baresic, A., Orengo, C.A., Martin, A.C.R. and Valencia, A. (2010) **Characterizing Pathogenic Germline Deviations in Human Protein Kinases** in **Proceedings of the Workshop on Annotation, Interpretation and Management of Mutations (AIMM-2010); a workshop at ECCB10.** Ghent, Belgium, September 26th, 2010. Ed. Christopher J. O. Baker, Rene Witte, Dietrich Rebholz-Schuhmann. [Online at http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-645/]
15. Izarzugaza, J.M.G., Baresic, A., McMillan, L.E.M., Yeats, C., Clegg, A., Orengo, C.A., Martin, A.C.R. and Valencia, A. (2009) **An integrated approach to the interpretation of Single Amino Acid Polymorphisms within the framework of CATH and Gene3D** *BMC Bioinformatics*. **10 Suppl 8**, S5-S5.
16. Hurst, J. M., McMillan, L. E. M., Porter, C. T., Allen, J. Fakorede, A. and Martin, A. C. R. (2008) **The SAAPdb web resource: a large scale structural analysis of mutant proteins**, *Human Mutation,* **30**, 616-624.
17. Cuff, A.L., Janes, R.W. and Martin, A.C.R. (2006) **Analysing the Ability to Retain Sidechain Hydrogen-bonds in Mutant Proteins**, *Bioinformatics*, **22**,1464-1470.
18. Cuff, A.L. and Martin, A.C.R. (2004) **Analysis of void volumes in proteins and application to stability of the p53 tumour suppressor protein**, *J. Mol. Biol*., **344**, 1199-1209.
19. Martin, A. C. R., Facchiano, A. M., Cuff, A. L., Hernandez-Boussard, T., Olivier, M., Hainaut, P. and Thornton, J. M. (2002) **Integrating Mutation Data and Structural Analysis of the p53 Tumour-Suppressor Protein**, *Human Mutation*, **19**, 149-164.

Major Software Releases

- **AbM** (antibody modelling) Released by Oxford Molecular as a commercial product, 1991;
- **abYsis** (antibody database) Sold via UCL Business, 2010–date;
- **apat** (Sequence annotation) >1200 users;
- **avp** (Void calculation) >1600 users;
- **BiopLib and BiopTools** (C library and tool set) Released via GitHub, 2015;
- **Cluster** (hierarchical clustering) >2800 users;
- **MINT** (graphical interface to Modeller) >3150 users;
- **nw** (Needleman and Wunsch sequence alignment) >2000 users;
- **ProFit** (protein least squares fitting) >7600 users;
- **QTree** (rendered molecular graphics) >2400 users;
- **Qlite** (simple queuing system for computer farms) >1200 users;
- **saptf** (Tools architecture for web sequence analysis) >300 users;
- **torsions** (backbone torsion angles) >3900 users.

CV – Christine Orengo

PERSONAL DETAILS

Full Name	Christine Anne **Orengo**
e-mail	c.orengo@ucl.ac.uk

CURRENT POSITION

2002 - date	Professor of Structural Bioinformatics, University College London

QUALIFICATIONS

1984	PhD – Enzyme Kinetics, University College London
1977	MSc – Medical Physics, Aberdeen University
1976	BSc - Chemical Physics, Bristol University

PREVIOUS POSITIONS

1995 – 2005	MRC Senior Fellow, Non-Clinical
1991 – 1995	Postdoctoral Research Fellow, University College London
1987 – 1991	Research Fellow, National Institute for Medical Research, London
1985 – 1987	Mathematical Modeller, Humphreys & Glasgow, London
1982 – 1985	Chief Chemist, FCI International, Brussels, Belgium
1977 – 1978	Research Assistant, Guys Hospital, London, UK

AWARDS, HONOURS AND POSITIONS

2014 -	Member of EMBO
2013 -	Scientific Advisory Board, The Genome Annotation Centre (TGAC)
2013 -	BBSRC strategy advisory panel on New Ways of Working
2013 -	Vice President International Society of Computational Biologists
2012 -	Editor Current Opinions in Structural Biology
2008 -	Associate Editor PLoS Computational Biology
2005 -	Scientific Advisory Board, Swiss Institute of Bioinformatics,
2006 - 2011	Executive Committee EU ENFIN Network for Systems Biology
1995 - 2005	Senior Fellowship award, Medical Research Council, UK.

SELECTED RECENT GRANTS

2015 – 2020	FunVar: Impacts of Genetic Variations on Protein Functions and Functional Pathways. Wellcome Trust £536,722.
2014 – 2018	A Greatly Expanded CATH-Gene3D with Functional Fingerprints to Characterise Proteins. BBSRC. £612,409.
2014 – 2018	The Developing Fly Interactome. Consortium of 6 groups. Prof. Simon Hubbard is PI. BBSRC. £475,000.
2015 – 2016	CATH-FunL: Improving Gene Target Selection by Predicting Functional Modules in Biological Systems. BBSRC £113,199.
2010 – 2015	Midwest Centre for Structural Genomics. NIH funded. £300,000.

SELECTED RECENT PUBLICATIONS

- **Orengo**, CA, Michie, AD, Jones DT, Swindells MB, & Thornton. JM (1997) CATH: A Hierarchic Classification of Protein Domain Structures. *Structure*, 5, 1093-1108.

- Buchan, D, Shepherd, A, Lee, D, Bray, J, Pearl, F, Thornton J & **Orengo** C (2002) Gene3D: Structural Assignments for Whole Genes and Genomes Using the CATH Domain Structure Database. *Genome Res*, 12, 503-514.
- Ranea, JA, Yeats, C., Grant, A. & **Orengo**, CA. (2007) Predicting protein function: the Phylo-Tuner method applied to eukaryotic genomes. *PLoS Comput Biol,* 3, e237.
- Lee, D, Redfern, O & **Orengo**, C (2007) Predicting protein function from sequence and structure. *Nat Rev Mol Cell Biol,* 8, 995-1005.
- Yeats, C, Lees, J, Reid, A, Kellam, P, Martin, N, Liu, X & **Orengo**, C. (2008). Gene3D: comprehensive structural and functional annotation of genomes. *NAR,* 36, D414-8.
- Rentzsch, R. & **Orengo**, CA. (2009) Protein function prediction - the power of multiplicity. *Trends Biotechnol*.
- Redfern, O, Dessailly, B, Dallman, T & **Orengo**, C (2009) Flora a Novel Method to Predict Protein Function from Structure. *PLoS Comput Biol*, 5 (8) : e1000485.
- Reid, A, Ranea, JA & **Orengo**, C (2010) CODA: Accurate detection of functional associations by domain fusion in higher eukaryoytes. *PLoS One,* 5, :e 10908.
- Ranea, JA, Reid, A, Yeats, C & **Orengo**, CA (2010) Study of the Dark Matter in Protein Network Prediction and Modelling. *PLoS Comp. Biol*, 6, pii: e1000945.
- Lee DA, Rentzsch R, **Orengo** C. (2010) GeMMA: functional subfamily classification within superfamilies of predicted structural domains. *Nucleic Acids Res*, 38, 720-37.
- Lees J, Yeats C, Redfern O, Clegg A, **Orengo C**. (2010) Gene3D: merging structure and function for a Thousand genomes. *Nucleic Acids Res*, 38, D296-300.
- Dessailly BH, Redfern OC, Cuff AL, **Orengo** CA (2010) Detailed analysis of function divergence in a large and diverse domain superfamily: toward a refined protocol of function classification. *Structure,* 18, 1522-35.
- Furnham N, Sillitoe I, Holliday GL, Cuff AL, Laskowski RA, **Orengo** CA, Thornton JM. (2012) Exploring the evolution of novel enzyme functions within structurally defined protein superfamilies. *PLoS Comput Biol*, 8, :e1002403.
- Rentzsch, R & **Orengo**, C. (2103) Protein Function Prediction Using Domain Families. *BMC Bioinformatics*. Suppl 3:S5.
- Radivojac, P….Jones, D, **Orengo**, C…Friedberg, I (2013) A Large Scale Evaluation of Protein Function prediction. *Nature Methods*, 10:221-7.
- Studer, R, Williams, M, Christin, A, Orengo, C (2014) Stability-activity trade-offs constrain the adaptive evolution of RubisCO. *Proc. Nat. Acad. Sci.* 111:2223-8.
- Lehtinen S, Lees J, Bähler J, Shawe-Taylor J, **Orengo C.** Gene Function Prediction from Functional Association Networks Using Kernel Partial Least Squares Regression. PLoS One. (2015) 10, e0134668.
- Das S, Lee D, Sillitoe I, Dawson NL, Lees JG, **Orengo C.** Functional classification of CATH superfamilies: a domain-based approach for protein function annotation. Bioinformatics. (2015) pii: btv398.
- Das S, Sillitoe I, Lee D, Lees JG, Dawson NL, Ward J, **Orengo C**. CATH FunFHMMer web server: protein functional annotations using functional family assignments. Nucleic Acids Res. (2015) 43(W1):W148-53.
- Mitchell A, ……Wu CH, **Orengo C,** Sillitoe I, Mi H, Thomas PD, Finn RD. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res. (2015) 43(Database issue):D213-21.
- Sillitoe I, Lewis TE, Cuff A, Das S, Ashford P, Dawson NL, Furnham N, Laskowski RA, Lee D, Lees JG, Lehtinen S, Studer RA, Thornton J, **Orengo C**. CATH: comprehensive structural and functional annotations for genome sequences. Nucleic Acids Res. (2015) 43(Database issue):D376-81.
- Lewis TE, Sillitoe I, …Velankar S, **Orengo C.** Genome3D: exploiting structure to help users understand their sequences. Nucleic Acids Res. (2015) 43:D382-6.
- Rallapalli P, **Orengo C**, Studer R, Perkins S. Positive selection in blood coagulation factors in the context of disease-causing mutations. Mol Biol Evol. (2014) 31:3040-56.

BIBLIOMETRIC DATA (AS OF DECEMBER 2015)

Total publications 245. Peer reviewed papers: 160; Total citations: 18,604; Max citations per paper: 2420; H-index = 68 (Google Scholar)

Curriculum Vitae: Michael J.E. Sternberg

Personal Details

Full Name	Michael Joseph Ezra **Sternberg**
e-mail	m.sternberg@imperial.ac.uk
URLs	www.imperial.ac.uk/people/m.sternberg
	www.imperial.ac.uk/cisbio
Present Position	Director Centre for Integrative Systems Biology and Bioinformatics

Professional History

2011	Director Centre for Integrative Systems Biology and Bioinformatics, Imperial College
2009 - 2011	Director Centre for Integrative Systems Biology at Imperial College
2005 - date	Deputy Head, Division of Molecular Biosciences, Imperial College
2001 - 2011	Director Centre for Bioinformatics, Imperial College
2001 - date	Professor of Structural Bioinformatics, Imperial College
1988 - 2001	Head of Laboratory of Biomolecular Modelling, ICRF, London
1985 - 1988	Lecturer, Dept of Crystallography, Birkbeck College, London
1983 - 1985	Royal Society University Research Fellowship, Birkbeck College, London
1977 - 1983	Royal Society Fellowships, Laboratory of Molecular Biophysics, Oxford.

Qualifications

1977	D Phil - Laboratory of Molecular Biophysics, Oxford.
1974	MSc - Computer Science, Imperial College
1972	BA - 1st Natural Sciences, Cambridge.

Selected Grants

- BBSRC - BB/M011526/1 Enhancing the Phyre2 protein modelling portal for the community Sternberg (PI) and Kelley. £700K 01/05/2015 – 31/04/2020
- Wellcome Trust - WT104955MA Development and dissemination of a community tool for structure-based annotations in proteins in disease networks The Wellcome Trust -. Sternberg (PI) & Houlston £830K. 01/12/14 – 30/11/2019
- BBSRC BB/L005247/1 Development and launch of a crowd-sourced serious-games platform for protein docking for use by the public and the scientific community Sternberg (PI) and Leymarie DockIt: Total £ 980K £430K to Sternberg
- BBSRC BB/J019240/1 The Gut Health and Food Safety Integrated Strategic Programme Grant (lead by IFR) BBSRC (£740K to Sternberg) 01/04/2012 – 31/03/2017

Prior

- BBSRC - BB/J019240/1 Sternberg (PI) Maintaining and extending PHYRE2 to deliver an internationally-recognised resource for protein model BBSRC £440K 01/08/2012-31/07/2015
- BBSRC - BB/I025271/1 GENOME-3D: A UK Network Providing Structure-Based Annotations for Genotype to Phenotype Studies. (£109K To Sternberg). 01/09/2011 – 31/07/2014

Selected Awards, Honours and Offices

2007 - 2015	Scientific Advisory Board Institute of Food Research
2005 - date	Editorial Board *Journal of Molecular Biology*
2004 - 2006	BBSRC representative on E-Science Steering Committee
2003 - 2004	Member of BBSRC Strategy Board
2003 - 2005	Chair of BBSRC Bioinformatics Coordinating Committee
2003 - date	Editorial Board *Protein Engineering Design and Selection*
2001 - date	Elected Fellow of the Institute of Biology

Research Interests

Structural and systems bioinformatics, especially protein modelling, and the development of logic-based chemoinformatics.

Selected Recent Publications

- Wass, M.N. and Sternberg, M.J. (2008) ConFunc--functional annotation in the twilight zone, **Bioinformatics**, 24, 798-806
- Lesk, V.I. and Sternberg, M.J. (2008) 3D-Garden: a system for modelling protein-protein complexes based on conformational refinement of ensembles generated with the marching cubes algorithm, **Bioinformatics,** 24, 1137-1144.
- Dobbins, S. E., Lesk, V. I. & Sternberg, M. J. E. (2008). Insights into protein flexibility: The relationship between normal modes and conformational change upon protein-protein docking. **Proc Natl Acad Sci U S A** 105, 10390-5.
- Kelley, L. A. & Sternberg, M. J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. **Nat Protoc** 4, 363-71.
- Wass, M. N. & Sternberg, M. J. (2009). Prediction of ligand binding sites using homologous structures and conservation at CASP8. **Proteins**. 77, 147-51.
- Jefferys BR, Kelley LA, Sternberg MJ (2010) Protein Folding Requires Crowd Control in a SimulatCell**. J Mol Biol** 397, 1329-38. 1.
- Wass, M. N., Kelley, L. A. & Sternberg, M. J. E. (2010). 3DLigandSite: predicting ligand-binding sites using similar structures. **Nucleic Acid Res** 38, W469 – W437.
- Reynolds CR, Amini AC, Muggleton SH, Sternberg MJ (2012) Assessment of a Rule-Based Virtual Screening Technology (INDDEx) on a Benchmark Data Set. ***J Phys Chem B*** 116: 6732-6739
- Wass, M. N., Barton, G. & Sternberg, M. J. (2012). CombFunc: predicting protein function using heterogeneous data sources. **Nucleic Acids Res** 40, W466-70.
- Phan, H.T. & Sternberg, M.J. (2012). PINALOG: a novel approach to align protein interaction networks--implications for complex detection and function prediction. **Bioinformatics** 28, 1239-45.
- David, A., Razali, R., Wass, M. N. & Sternberg, M. J. (2012). Protein-protein interaction sites are hot spots for disease-associated nonsynonymous SNPs. **Hum Mutat** 33, 359-63
- Yates, C. M. & Sternberg, M. J. (2013). Proteins and domains vary in their tolerance of non-synonymous single nucleotide polymorphisms (nsSNPs). **J Mol Biol 425**, 1274-86.
- Yates, C. M., Filippis, I., Kelley, L. A. & Sternberg, M. J. E. (2014). SuSPect: Enhanced Prediction of Single Amino Acid Variant (SAV) Phenotype Using Network Features. **J Mol Biol** 426, 2692-2701.
- Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ: **The Phyre2 web portal for protein modeling, prediction and analysis.** *Nature protocols* 2015, **10:**845-858.
- Cornish AJ, Filippis I, David A, Sternberg MJ. Exploring the cellular basis of human disease through a large-scale mapping of deleterious genes to cell types. Genome Medicine. 2015;7:1-18.
- Mezulis S, Sternberg MJ, Kelley LA. PhyreStorm: A web server for fast structural searches against the PDB. Journal of Molecular Biology. 2015. Advanced publication

CV – Charlotte Deane

PERSONAL DETAILS

Full Name	Charlotte Deane
e-mail	deane@stats.ox.ac.uk

CURRENT POSITIONS

2002 - date	Professor of Structural Bioinformatics, Oxford University
Since Oct 2105	Head of Department of Statistics, Oxford University
Since Oct 2014	Associate Head of Mathematical Physical and Life Sciences Division
Since Jan 2009	Director - Systems Approaches to Biomedical Research CDT

AWARDS, HONOURS AND POSITIONS

Since 2009	Director of the SABS industrial consortium
Since 2003	ISMB and ECCB Program Committee
Since 2006	EPSRC peer review college
Since 2013	External Examiner Imperial College
Since 2013	BBSRC Tools and Resources Development Fund Panel

GRANTS – RELATED TO GECIP ACTIVITY

2009 - 2019	Systems Biology Doctoral Training Centre EPSRC £6,197,164
2009 – 2018	Industrial Doctorate Centre: Systems Approaches to Biomedical Science EPSRC£6,623,296 (Plus a contribution of ~£1.5M from industry)
2013 - 2016	Computational Structural Biology UCB £131,713
2013 - 2016	Network comparison EPSRC £506,211
2014 - 2023	Centre for Doctoral Training in Systems Approaches to Biomedical Science EPSRC and MRC ~£3.6M (Plus a contribution from industry of at least £2.7M)

SELECTED RECENT PUBLICATIONS

- Dunbar, J and Deane CM ANARCI: antigen receptor numbering and receptor classification, *Bioinformatics,* 2015. 10.1093/bioinformatics/btv552
- Bradley, AR. Wall, ID. Green, DV. Deane, CM. Marsden, BD OOMMPPAA: A Tool To Aid Directed Synthesis by the Combined Analysis of Activity and Structural Data, *J. Chem. Inf. Model.,* 2014, 54 (10), pp 2636–2646
- Wilman, HR. Ebejer, JP. Shi, J. and Deane, CM. Knapp, B. Crowdsourcing yields a new standard for kinks in protein helices, *Journal of Chemical Information and Modeling*, 2014, 54(9), 2585-2593
- Ali, W. Rito, T. Reinert, G. Sun, F. Deane, CM. Alignment-free protein interaction network comparison, *Bioinformatics*, 2014, 30(17), i430-i437
- K Krawczyk, K. Liu, X. Baker, T. Shi, J Deane, CM Improving B-cell epitope prediction and its application to global antibody-antigen docking, *Bioinformatics*, 2014, 30(16), 2288-2294
- Edwards, H. Abeln, S. Deane, CM. Exploring Fold Space Preferences of New-born and Ancient Protein Superfamilies, *Plos Computational biology*, 2013, 9(11), e1003325
- J. Dunbar, J. Krawzyk, K. Leem, J. Baker, T. Fuchs, A. Georges, G. Shi, J. Deane, CM. SAbDab: the Structural Antibody Database, *NAR*, 2014, 42, D1140-D1146
- Lewis AC, Jones NS, Porter MA, Deane CM, What evidence is There for the Homology of Protein-Protein interactions*, PloS Comp. Bio*., 2012, 8(9), e1002645
- Ebejer JP, Morris GM, Deane CM., Freely Available Conformer Generation Methods: How Good Are They?, *J Chem Inf Model.* 2012 May 25;52(5):1146-58
- Ebejer, JP. Hill, J. Kelm, S. Shi, J. Deane, CM. Memoir: template based structure prediction for membrane proteins, *Nucleic Acid Research*, 2013, 41(W1), W379-W383
- Kelm S, Shi J, Deane CM, MEDELLER: Homology-Based Coordinate Generation for Membrane Proteins, *Bioinformatics*, 2010, 26(22), 2833-40
- Hamer R, Luo Q, Armitage JP, Reinert G, Deane CM, i-Patch: Interprotein contact prediction using local network information, *Proteins*, 2010, 78(13), 2781-97
- Saunders R, Deane CM, Synonymous codon usage influences the local protein structure observed, *Nucleic Acids Res*., 2010, 38(19), 6719-28

Anne Hersey: Curriculum Vitae

Personal Information

• Current position:
- Acting Chemogenomics Group Leader, European Bioinformatics Institute, 2015

• Contact information:

European Bioinformatics Institute
European Molecular Biology Laboratory
Wellcome Trust Genome Campus
Hinxton, Cambridge CB10 1SD
United Kingdom
+44 (0)1223 492689 (voice); +44 (0)1223 494494 (fax)
ahersey@ebi.ac.uk

Education

• University of Kent at Canterbury, UK. 1982
- PhD: "*Substrate Binding to Cyclodextrin*".

• University of Kent at Canterbury 1975 - 1978
- BSc (Hons) Chemistry

Previous Positions

- 2009 – 2015 Group Coordinator, ChEMBL Group, EMBL-EBI, Wellcome Genome Campus, Hinxton,UK
- 2002 – 2009 Team Leader, ADMET Modelling Group, Department of Computational Chemistry, GlaxoSmithKline, Medicines Research Centre, Gunnels Wood Road, Stevenage, Herts
- 1998 – 2002 Scientific Manager, Technology and Informatics Group, Department of DMPK, GlaxoWellcome R&D, Park Road, Ware, Herts
- 1995-1998 Principal Scientist, Technology and Informatics Group, Department of DMPK, GlaxoWellcome R&D, Park Road, Ware, Herts
- 1990-1995 Leader of Physical Organic Chemistry Group, Department of Physical Sciences, Wellcome Research Laboratories, Langley Park, Beckenham, Kent
- 1982-1990 Research Scientist, Department of Physical Sciences, Wellcome Research Laboratories, Langley Park, Beckenham, Kent

Publications

1. Papadatos G, Mark Davies M, Dedman N, Chambers J, Gaulton A, Siddle J, Koks R, Irvine SA, Pettersson J, Goncharoff N, **Hersey A**, Overington JP(2016) SureChEMBL: a large-scale, chemically annotated patent document database Nucl. Acids Res. 44 (D1): D1220-D1228

2. Gaulton A, Kale N, van Westen GJP, Bellis LJ, Bento AP, Davies M, **Hersey A**, Papadatos G, Forster M, Wege P, Overington JP (2015) A large-scale crop protection bioassay data set *Scientific Data* **2** 150032

3. Papadatos G, Gaulton A, **Hersey A**, Overington JP (2015) Activity, assay and target data curation and quality in the ChEMBL database *J. Comput. Aided Mol. Des.* DOI:10.1007/s10822-015-9860-5 PMID:26201396 PMC4607714

4. **Hersey A**, Chambers J, Bellis LJ, Bento AP, Gaulton A, Overington JP, (2015) Chemical databases: curation or integration by user-defined equivalence? *Drug Discovery Today: Technologies* 14,17-24.

5. Davies M, Dedman N, **Hersey A**, Papadatos G, Hall MD, Cucurull-Sanchez L, Jeffrey P, Hasan S, Eddershaw PJ, Overington JP (2015) ADME SARfari: Comparative Genomics of Drug Metabolising Systems *Bioinformatics* btv010v2.

6. Bento AP, Gaulton A, **Hersey A**, Bellis LJ, Chambers J, Davies M, Krueger FA, Light Y, Mak L, McGlinchey S, Nowotka M, Papadatos G, Santos R, & Overington JP (2014) The ChEMBL bioactivity database: an update *Nucl. Acids Res. Database Issue*. 42 D1083-1090.

7. Chambers J., Davies M, Gaulton A, Papadatos G, **Hersey A**, Overington JP (2014) UniChem: extension of InChI-based compound mapping to salt, connectivity and stereochemistry layers *J. Cheminf.* 6(1): 43.

8. Chambers J, Davies M, Gaulton A, **Hersey A**, Velankar S, Petryszak R, Hastings J, Bellis LJ, McGlinchey S, & Overington JP (2013) UniChem: A Unified Chemical Structure Cross-Referencing and Identifier Tracking System *J. Cheminf.* 5(1): 3.

9. **Hersey A**, Senger S, Overington JP, (2012) Open data for drug discovery: learning from the biological community, *Future Medicinal Chemistry* 4 (15), 1865-1867.

10. Zdrazil B, Pinto M, Vasanthanathan P, Williams AJ, Balderud LZ, Engkvist O, Chichester C, **Hersey A**, Overington JP, Ecker GF (2012) Annotating Human P-Glycoprotein Bioassay Data, *Molecular informatics* 31 (8), 599-609.

11. Gaulton A, Bellis LJ, Chambers J, Davies M, **Hersey A**, Light Y, McGlinchey S, Akhtar R, Bento AP, Al-Lazikani B, Michalovich D, & Overington JP (2012) ChEMBL: A Large-scale Bioactivity Database For Chemical Biology and Drug Discovery *Nucl. Acids Res. Database Issue*. 40 D1100.

12. Taboureau, O, **Hersey A**, Audouze, KML Gautier, L, Jacobsen, U P Akhtar, R, Atkinson, F; Overington, J P; Brunak, S (2012) Toxicogenomics investigation under the eTOX Project *J. Pharmacogenom & Pharmacoproteomics* S7.

13. Gleeson MP, **Hersey A**, Montanari D & Overington JP (2011) Probing the links between in vitro potency, ADMET and physicochemical parameters *Nature Rev. Drug Discov*. 10 197-208.

14. Bellis LJ, Akhtar R, Al-Lazikani B, Atkinson F, Bento AP, Chambers J, Davies M, Gaulton A, **Hersey A**, Ikeda K, Krüger FA, Light Y, McGlinchey S, Santos R, Stauch B, Overington JP (2011) Collation and data-mining of literature bioactivity data for drug discovery. *Biochem Soc Trans* Oct; 39(5): 1365-70.

15. Gleeson MP, **Hersey A**, Hannongbua S (2011) In-silico ADME models: a general assessment of their utility in drug discovery applications *Curr. Topics Med. Chem.*,11(4), 358-381

16. Hills EE, Abraham MH, **Hersey A**, Bevan CD (2011) Diffusion coefficients in ethanol and in water at 298K: Linear free energy relationships *Fluid Phase Equilibria* 303 (1), 45-55

17. Myatt JW, Healy MP, Bravi GS, Billinton A, Johnson CN, Matthews KL, Jandu KS, Meng W, **Hersey A**, Livermore DG, Douault CD, Witherington J, Bit RA, Rowedder JE, Brown JD, Clayton NM (2010) Pyrazolopyridazine alpha-2-delta-1 ligands for the treatment of neuropathic pain *Bioorg. & Med. Chem. Letters* 20 (15), 4683-4688

18. **Hersey A**, Blaney FE, Modi S Transporters. Published in “Gene Family Focused Molecular Design”, Ed. Karen Lackey, Wiley 2008.

Sarah Teichmann: Curriculum Vitae

Personal Information

• Current position:
- Research Group Leader, EMBL-European Bioinformatics Institute, 2012
& Senior Group Leader, Wellcome Trust Sanger Institute
& Director of Research, Dept Phyics/Cavendish Laboratory, University of Cambridge

• Contact information:

European Bioinformatics Institute
European Molecular Biology Laboratory
Wellcome Trust Genome Campus
Hinxton, Cambridge CB10 1SD
United Kingdom
+44 (0)1223 492520 (voice); +44 (0)1223 494494 (fax)
saraht@ebi.ac.uk

Education

• University of Cambridge, UK & MRC Laboratory of Molecular Biology. 2000
- PhD: "*Genome Evolution*".

- University of Cambridge, UK. 1993-96.
- BA(Hons) Natural Sciences (Biochemistry)

Previous Positions

1999-2005:	Trinity College Research Fellowship.
2000-2001:	Postdoctoral research with Prof. Dame J. Thornton CBE FRS FMedSci in the Department of Biochemistry & Molecular Biology, University College London on a Beit Memorial Fellowship for Medical Research.
2001-2005:	MRC Career Track Programme Leader at the MRC Laboratory of Molecular Biology, Cambridge, UK & Research Fellow of Trinity College, Cambridge.
2006-2013:	MRC Programme Leader at the MRC Laboratory of Molecular Biology, Cambridge, UK
2005-2015:	Fellow, Trinity College, Cambridge.
02/2013-09/2015:	Principal Research Associate, Cavendish Laboratory/Physics Dept., Cambridge University

Top Ten Publications out of over 130:

1. Ahnert, S.E., Marsh, J.M., Hernandez, H., Robinson, C.V. & Teichmann, S.A. (2015) Principles of assembly reveal a periodic table of protein complexes. ***Science***.
2. Kolodziejczyk A.A., Kim, J.K., Tsang, J. C. H., Illicic, T., Henriksson, J., Natarajan K. N. , Tuck A.C., Gao X., Bühler M., Liu, P., Marioni, J.C. & Teichmann, S.A. (2015) Single-cell mRNA-sequencing of pluripotent states unlocks modular transcriptional variation. ***Cell Stem Cell***, 17, 471-485.
3. Perica, T., Kondo, Y., Tiwari, S.P., McLaughlin, S.H., Kemplen, K.R., Zhang, X., Steward, A., Reuter, N., Clarke, J. & Teichmann, S.A. (2014) Evolution of oligomeric state through allosteric pathways that mimic ligand binding. ***Science***, 346, 1254346.
4. Mahata B., Zhang X., Kolodziejczyk A.A, Proserpio, V., Haim-Vilmovsky, L., Taylor, A.E., Hebenstreit, D., Dingler, F.A., Moignard, V., Gottgens, B., Arlt, W., McKenzie, A.N.J. &

Teichmann, S.A. (2014) Single Cell RNA-Sequencing Reveals T helper Cells Synthesizing Steroids *de novo* to Contribute to Immune Homeostasis. ***Cell Reports*, 7**, 1130-42.
5. Marsh, J.A., Hernández, H., Hall, Z., Ahnert, S.E., Perica, T., Robinson, C.V. & Teichmann S.A. (2013) Protein Complexes Are under Evolutionary Selection to Assemble via Ordered Pathways. ***Cell***, 153, 461-70.
6. Charoensawan, V., Janga, S.C., Bulyk, M.L., Babu, M.M., & Teichmann S.A. (2012) DNA sequence preferences of transcriptional activators correlate more strongly than repressors with nucleosomes. ***Mol Cell,*** 47, 183-92.
7. Hebenstreit, D., Fang, M., Gu, M. Charoensawan, V., van Oudenaarden, A. & Teichmann, S.A. (2011) RNA sequencing reveals two major classes of gene expression levels in metazoan cells. ***Mol. Sys. Biol.,*** 7:497
8. Levy, E.D., Boeri-Erba, E., Robinson, C.V. & Teichmann, S.A. (2008) Assembly reflects evolution of protein complexes. ***Nature,*** 453, 1262-5.
9. Teichmann, S.A. & Babu, M.M. (2004) Gene Regulatory Network Growth by Duplication. ***Nature Genet.***, 36**,** 492-496
10. Chothia, C., Gough, J., Vogel, C. & Teichmann S.A. (2003) Evolution of the Protein Repertoire. ***Science***, 300, 1701-1703.

Curriculum Vitae: Professor J. Gough MSci, PhD (Cantab)

PERSONAL DETAILS
Full name Professor Julian **GOUGH**
Address Dept. of Computer Science, University of Bristol
E-mail: Julian.Gough@bris.ac.uk
Lab URL: bioinformatics.bris.ac.uk

CURRENT POSITION
Aug 2012 Professor of Bioinformatics, Computer Science, University of Bristol

QUALIFICATIONS
2002 PhD University of Cambridge
1998 MSci joint hon.s Mathematics and Physics, University of Bristol

HONOURS & AWARDS
2002 – 2004 Projects in Maths and Molecular Biology Fellowship
2012 – 2013 University Research Fellowship (Bristol)

PREVIOUS POSITIONS
2007 – 2012 Reader in Bioinformatics, University of Bristol
2005 – 2006 Visiting Scientist, Institut Pasteur, Paris
2003 – 2005 Adjunct Professor, Tokyo Medical and Dental University
2003 – 2005 Research Scientist, RIKEN Genomic Sciences Centre, Tokyo
2002 – 2003 Research Fellow, Stanford University (Levitt)
2001 – 2002 Research Associate, MRC Laboratory of Molecular Biology, Cambridge
1993 – 1998 Research Assistant, Schlumberger Cambridge Research

EDITORIAL BOARDS
BioEssays (2012-present)
Proteins: Structure, Function and Bioinformatics (2007-2012)
Current Opinion in Structural Biology (guest editor 2010, 2012)

PROGRAMME COMMITTEES
ISMB 2004-2014
ECCB 2004,2005,2007,2010-2011
Bioinformatics 2010-2014

RESEARCH COMMITTEES
BBSRC Bioinformatics and Biological Resources (2010,2012)
MRC Human Genetics QQR review committee (2011)
BBSRC Tools and Resources Development Fund (2010)
BBSRC Committee C (2010)

PHD STUDENTS SUPERVISED
9 students (4 completed PhD, 1 submitted, 1 about to submit, 3 in progress)

GRANTS
I have held 9 as PI and 3 as Co-I. I have been PI on funding of over £1.8M, including support from: BBSRC, EPSRC, EU, NERC, Google, Amazon, JSPS and the Royal Society.

EXTERNAL RESPONSIBILITIES
Scientific Advisor and co-founder of Genetrainer (2013-present)
External Examiner at the University of Cardiff MSc in Bioinformatics (2013-present)

BIBLIOMETRIC DATA (AS OF NOVEMBER 2013)
Publications: 80; Total citations: 10762; Since 2010: 5820; Single paper: 1770; H-index: 36

Bissan Al-Lazikani

THE INSTITUTE OF CANCER RESEARCH15 COTSWOLD ROAD• LONDON • SM2 5NG • ✆: 020 8722 4300 • ✆: 07904 239 565 •✉: BISSAN.AL-LAZIKANI@ICR.AC.UK

Education

1995-1998	*Ph.D. MRC Laboratory of Molecular Biology and Newnham College, Cambridge University.*
1994-1995	*M.Sc. Computer Science, Imperial College London*
1991-1994	*B.Sc. (Hons) Molecular Biology. University College London*

Experience

Jan 2009 - current	*Head Integrative Informatics, The Institute of Cancer Research* *Team Leader Computational Biology and Chemogenomics, Cancer Therapeutics, The Institute of Cancer Research*
Sep-Dec 2008	*Consultant, European Bioinformatics Institute*
Jan 2001-Jul 2008	*Group Leader/Associate Director Discovery Informatics, Inpharmatica Ltd.*
Jan 1999-Jan 2001	*Postdoctoral Fellow (HH fellowship) in the laboratory of Prof. Barry Honig HHMI* *Dept. Biochemistry & Molecular Biophysics, Columbia University, New York*
Oct-Dec 1998	Postdoctoral scientist at the MRC Laboratory of Molecular Biology

Publictions

- Tym JE, Mitsopoulos C, Coker EA, Razaz P, Schierz AC, Antolin AA, **Al-Lazikani B**. canSAR: an updated cancer research and drug discovery knowledgebase. (2016) Nucleic Acids Res. 44:D938-43.
- Mitsopoulos C, Schierz AC, Workman P, **Al-Lazikani B.** Distinctive Behaviors of Druggable Proteins in Cellular Networks. (2015) PLoS Comput Biol. 11:e1004597. doi: 10.1371/journal.pcbi.1004597.
- Pearl LH, Schierz AC, Ward SE, **Al-Lazikani B**, Pearl FM. Therapeutic opportunities within the DNA damage response. (2015) Nat Rev Cancer. Vol 15(3): pp166-80.
- **Al-Lazikani**, Chemoinformatics for Bioinformaticians (2014) Concise Encyclopaedia of Bioinformatics and Computational Biology, 2nd Edition, Ed. Hancock & Zvelebil. Wiley-Blackwell - ISBN: 978-0-470-97871-9
- Bulusu KC, Tym JE, Coker EA, Schierz AC, **Al-Lazikani B**. canSAR: updated cancer research and drug discovery knowledgebase. (2014) Nucleic Acids Res. 42(Database issue):D1040-7. Epub 2013 Dec 3.
- Workman P, **Al-Lazikani B**. Drugging cancer genomes. (2013) Nat Rev Drug Discov. Vol 12(12), pp889-90
- Workman P, **Al-Lazikani B**, Clarke PA. Genome-based cancer therapeutics: targets, kinase drug resistance and future strategies for precision oncology. (2013) Curr Opin Pharmacol. Vol 13(4):486-96
- Box C, Mendiola M, Gowan S, Box GM, Valenti M, Brandon AD, **Al-Lazikani B**, Rogers SJ, Wilkins A, Harrington KJ, Eccles SA. A novel serum protein signature associated with resistance to epidermal growth factor receptor tyrosine kinase inhibitors in head and neck squamous cell carcinoma. (2013) Eur J Cancer. 9. doi:pii: S0959-8049(13)00213-X.
- Walters, ZS., Villarejo-Balcells, B., Olmos, D., Buist, TW., Missiaglia, E., Allen, R., **Al-Lazikani, B.,** Garrett, MD., Blagg, J. & Shipley, J. (2013) JARID2 is a direct target of the PAX3-FOXO1 fusion protein and inhibits myogenic differentiation of rhabdomyosarcoma cells. Oncogene.
- Gonzalez de Castro, D., Clarke, PA., **Al-Lazikani, B.** & Workman, P. (2013) Personalized cancer medicine: molecular diagnostics, predictive biomarkers, and drug resistance. Clin Pharmacol Ther, Vol.93(3), pp.252-259.
- **Al-Lazikani, B.** & Workman, P. (2013) Unpicking the combination lock for mutant BRAF and RAS melanomas. Cancer Discov, Vol.3(1), pp.14-19.
- Patel, MN., Halling-Brown, MD., Tym, JE., Workman, P. & **Al-Lazikani, B.** (2013) Objective assessment of cancer genes for drug discovery. Nat Rev Drug Discov, Vol.12(1), pp.35-50.
- **Al-Lazikani, B.**, Banerji, U. & Workman, P. (2012) Combinatorial drug therapy for cancer in the post-genomic era. Nat Biotechnol, Vol.30(7), pp.679-692.
- Workman, P., Clarke, PA. & **Al-Lazikani, B.** (2012) Personalized medicine: patient-predictive panel power. Cancer Cell, Vol.21(4), pp.455-458.
- Halling-Brown, MD., Bulusu, KC., Patel, M., Tym, JE. & **Al-Lazikani, B**. (2012) canSAR: an integrated cancer public translational research and drug discovery resource. *Nucleic Acids Res*, **40**(Database issue), pp.D947-D956

THE INSTITUTE OF CANCER RESEARCH15 COTSWOLD ROAD• LONDON • SM2 5NG • ✆: 020 8722 4300 • ✆: 07904 239 565 •✉: BISSAN.AL-LAZIKANI@ICR.AC.UK

- Gaulton, A., Bellis, LJ., Bento, AP., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., **Al-Lazikani, B**., et al. (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res*, **40**(Database issue), pp.D1100-D1107
- Bellis, LJ., Akhtar, R., **Al-Lazikani, B.,** Atkinson, F., Bento, AP., Chambers, J., Davies, M., Gaulton, A., Hersey, A., Ikeda, K., et al. (2011) Collation and data-mining of literature bioactivity data for drug discovery. *Biochem Soc Trans*, **39**(5), pp.1365-1370
- Orchard, S., **Al-Lazikani, B.,** Bryant, S., Clark, D., Calder, E., Dix, I., Engkvist, O., Forster, M., Gaulton, A., Gilson, M., et al. (2011) Minimum information about a bioactive entity (MIABE). *Nat Rev Drug Discov*, **10**(9), pp.661-669
- Suwaki, N., Vanhecke, E., Atkins, KM., Graf, M., Swabey, K., Huang, P., Schraml, P., Moch, H., Cassidy, AM., Brewer, D., **Al-Lazikani, B.,** et al. (2011) A HIF-regulated VHL-PTP1B-Src signaling axis identifies a therapeutic target in renal cell carcinoma. *Sci Transl Med*, **3**(85), pp.85ra47-
- Abad-Zapatero, C., Perišić, O., Wass, J., Bento, AP., Overington, J., **Al-Lazikani, B.** & Johnson, ME. (2010) Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an atlas-like representation. *Drug Discov Today*, **15**(19-20), pp.804-811
- Berriman, M., Haas, BJ., LoVerde, PT., Wilson, RA., Dillon, GP., Cerqueira, GC., Mashiyama, ST., **Al-Lazikani, B.**, Andrade, LF., Ashton, PD., et al. (2009) The genome of the blood fluke Schistosoma mansoni. *Nature*, **460**(7253), pp.352-358
- Agüero, F., **Al-Lazikani, B.,** Aslett, M., Berriman, M., Buckner, FS., Campbell, RK., Carmona, S., Carruthers, IM., Chan, AW., Chen, F., et al. (2008) Genomic-scale prioritization of drug targets: the TDR Targets database. *Nat Rev Drug Discov,* **7**(11), pp.900-907
- **Al-Lazikani, B.,** Hill, EE. & Morea, V. (2008) Protein structure prediction. *Methods Mol Biol*, **453** pp.33-85, ISSN: 1064-3745
- Overington JP, **Al-Lazikani, B** and Hopkins AL. "How many drug targets are there?" Nat. Rev. Drug. Discov. (2006) **5**:993-6.
- **Al-Lazikani, B**, Gaulton, A, Paolini, G, Lanfear, J, Overington, JP and Hopkins. AL "The Molecular Basis of Predicting Druggability." Chemical Biology, Ed: Gunther Wess & Stuart Schreiber, Wiley, 2007
- **Al-Lazikani, B**, Gaulton, A, Paolini, G, Lanfear, J, Overington, JP and Hopkins. AL "The Molecular Basis of Predicting Druggability" Bioinformatics – From Genomes to Therapies, Ed: Thomas Lengauer, Wiley-VCH, 2007
- Marks DJ, Harbord MW, MacAllister R, Rahman FZ, Young J, **Al-Lazikani B**, Lees W, Novelli M, Bloom S, Segal AW. "Defective acute inflammation in Crohn's disease: a clinical investigation." *Lancet*. (2006). **367**:668-78.
- George RA, Spriggs RV, Bartlett GJ, Gutteridge A, MacArthur MW, Porter CT, **Al-Lazikani B**, Thornton JM, Swindells MB. "Effective function annotation through catalytic residue conservation." *Proc Natl Acad Sci* U S A. (2005). **102**:12299-304.
- Freilich S, Spriggs RV, George RA, **Al-Lazikani B**, Swindells M, Thornton JM. "The complement of enzymatic sets in different species." *J Mol Biol*. (2005). **349**:745-63.
- George RA, Spriggs RV, Thornton JM, **Al-Lazikani B**, Swindells MB. "SCOPEC: a database of protein catalytic domains." *Bioinformatics*. (2004). **4**; Suppl 1:I130-I136.
- Sheinerman FB, **Al-Lazikani B**, Honig B. "Sequence, structure and energetic determinants of phosphopeptide selectivity of SH2 domains." *J Mol Biol*. (2003). **334**:823-41.
- **Al-Lazikani B,** Sheinerman FB, Honig B. "Combining multiple structure and sequence alignments to improve sequence detection and alignment: application to the SH2 domains of Janus kinases." *Proc Natl Acad Sci USA*. (2001). **98**:14796-801.
- **Al-Lazikani B,** Jung J, Xiang Z, Honig B "Protein structure prediction." *Curr Opin Chem Biol*. (2001). **5**:51-6.
- **Al-Lazikani B**, Lesk AM, Chothia C. "Canonical structures for the hypervariable regions of T cell alphabeta receptors." *J Mol Biol*. (2000). **295**:979-95.
- **Al-Lazikani, B**., C. Chothia, A.M. Lesk, V. Morea, M. Rustici & A. Tramontano. "Immunoglobulin Structure." *In: Encyclopedia of Molecular Biology*, (1999). T.E. Creighton (ed.) John Wiley and Sons, NY.
- **Al-Lazikani B,** Lesk AM, Chothia C. "Standard conformations for the canonical structures of immunoglobulins." *J Mol Biol*. (1997). **273**:927-48.

CV – Janet Thornton

PERSONAL DETAILS

Full Name	Janet M Thornton
e-mail	thornton@ebi.ac.uk

CURRENT POSITION

2015 - Senior Scientist, EMBL-EBI

EDUCATION

1970 – BSc Physics Class I, University of Nottingham

1973 – PhD Biophysics King's College, London & National Institute of Medical Research

AWARDS, HONOURS AND POSITIONS

2015 - Senior Scientist and Director Emeritus

2001 - 2015 Director, EMBL-EBI
1990 - 2001 Professor of Biomolecular Structure, UCL
1995 - 2001 Bernal Professor at Birkbeck College, London

2014 Fellow of the Academy of Medical Sciences, UK
2012 Dame Commander of the Order of the British Empire
2005 ISCB (International Society of Computational Biology) Award & Senior Fellow
2003 Foreign Associate of the US National Academy of Science
2002 Honorary Professor University of Cambridge, UK
1999 Fellow of the Royal Society, UK

GRANTS – RELATED TO GeCIP-SB ACTIVITY

2014 – 2018 Wellcome Trust Strategic Award
Biological determinants of ageing in late life health, and their pharmacological manipulation from model organisms to humans
Collaborators: Professor L. Partridge, D. Gems & D. Withers

2014 – 2018 A Greatly Expanded CATH-Gene3D with Functional Fingerprints to Characterise Proteins. Wellcome Trust. £612,409 with C. Orengo.

2012-2016 EU FP7 "BioMedBridges : Building data bridges between biological and medical infrastructures in Europe". , Euros: 3,564,236; Project Coordinator

SELECTED RECENT PUBLICATIONS

Martin, A.C.R., Facchiano, A.M., Cuff, A.L., Hernandez-Boussard, T., Oliveier, M., Hainaut, P. & **Thornton, J.M.** (2002) Integrating Mutation Data and Structural Analysis of the TP53 Tumor-Suppressor Protein. *Human Mutation*, **19:** 149-164.

Steward, R.E., MacArthur, M.W., Laskowski, R.A. & **Thornton, J.M.** (2003) Molecular basis of inherited diseases: A structural perspective. *Trends in Genetics*. **19**, 505-513.

George, R.A., Spriggs, R.G., Bartlett, G.J., Gutteridge, A., MacArthur, M.W., Porter, C.T., Al-Lazikani, B., **Thornton, J.M.** & Swindells, M.B. (2005) Effective function annotation through residue conservation. *PNAS*, **102**, 12299-12304.

Talavera, D., Taylor, M.S., and **Thornton, J.M.** (2010). The (non)malignancy of cancerous amino acidic substitutions. *Proteins* 78, 518-29.

Furnham, N., de Beer, T.A., and **Thornton, J.M.** (2012). Current challenges in genome annotation through structural biology and bioinformatics. *Curr. Op. Struct. Biol.* 22, 594-601.

de Beer TA, Laskowski RA, Parks SL, Sipos B, Goldman N, Thornton JM. (2013) Amino acid changes in disease-associated variants differ radically from variants observed in the 1000 genomes project dataset. *PLoS Comput Biol* Volume 9 (2013) p.e1003382

Martinez Cuesta S, Furnham N, Rahman SA, Sillitoe I, Thornton JM.(2014) The evolution of enzyme function in the isomerases. *Curr Opin Struct Biol* Volume 26 (2014) p.121-130

Ivanov DK, Escott-Price V, Ziehm M, Magwire MM, Mackay TF, Partridge L, Thornton JM.(2015)
Longevity GWAS Using the Drosophila Genetic Reference Panel. *J Gerontol A Biol Sci Med Sci* Volume 70 p.1470-1478

RESEARCH FOCUS

Professor Janet Thornton FRS has worked in the field of computational structural biology for over 30 years. She was Director of the European Bioinformatics Institute (EMBL-EBI) at Hinxton from 2001-2015 and now leads an active research group of post-docs and students. This group is focussed on computational studies in the areas of structural and functional genomics. Her main contribution (published in more than 240 peer-reviewed articles) has been in the systematic analysis of protein sequences, structures and functions. Initially our work focussed on understanding the relationship between protein sequences and structures, resulting in the classification of protein structures according to their evolutionary origins and their structural characteristics, as represented in the CATH and PDBsum data resources, which are heavily used worldwide. More recently we have focussed on protein interactions with other proteins and with small molecules. In collaboration with others she has developed novel approaches to predicting protein structure and function. Of direct relevance to this grant application, her work has included some analysis of disease causing mutations, with several directly related publications over the last 10 years.

Curriculum Vitae: Dr. Thomas Gaunt

Current appointment

Aug 2015-: Reader in Bioinformatics, School of Social and Community Medicine, University of Bristol.

Education and qualifications

BSc (Hons)	2:1	Biology	University of Southampton	1995
PhD		Human Genetics	University of Southampton	2002

Not clinically qualified or active

Source of funding

HEFCE

Previous appointments

2012-2015	MRC Integrative Epidemiology Unit in the School of Social and Community Medicine, University of Bristol	Senior Lecturer in Bioinformatics and Molecular Genetics (HEFCE funded)
2006-2012	MRC CAiTE Centre, School of Social and Community Medicine (originally Dept Social Medicine), University of Bristol	Lecturer in Bioinformatics and Molecular Genetics (HEFCE funded)
2005-2008	University of Southampton 2005-2006 University of Bristol 2006-2008	British Heart Foundation Intermediate Fellow
1997-2006	Human Genetics and University Medicine, University of Southampton	Research Assistant, then Research Fellow in Human Genetics
1995-1997	Molecular Microbiology, University of Southampton	Research Assistant in Molecular Microbiology

Research Interests

- *Bioinformatics*: bioinformatics & data mining theme co-lead in the MRC IEU, bioinformatics lead for ALSPAC. Machine learning applications in molecular epidemiology. Analysis of exome-seq, BS-seq, 16S-seq, metabolomics. Development of web-based utilities for statistical genetics
- *Data integration*: integrating different types of omics data using database, machine learning and graph-based approaches. Web-based data exploration/sharing platforms.
- *Cardiovascular genetic and molecular epidemiology*: including BHF funded cardiochip and metabochip genotype and metabolomic analyses in several cohorts (the UCLEB consortium)
- *Systems medicine*: Pathway and pleiotropy analyses and gene classification/annotation.
- *Epigenetic epidemiology*: BBSRC-funded HM450k analysis of 1000 mother/child pairs at five time-points in the ALSPAC cohort (ARIES). BS-Seq analysis in a subset of ARIES. GWAS of methylation (GoDMC consortium).

Publications 2012-2015

1. Simpkin AJ, Hemani G, Suderman M, **Gaunt TR**, Lyttleton O, McArdle WL, Ring SM, Sharp GC, Tilling K, Horvath S *et al* **[Total: 18 authors]** *Prenatal and early life influences on epigenetic age in children: A study of mother-offspring pairs from two cohort studies.* Hum Mol Genet. 2015 Nov 5. pii: ddv456.
2. Davies NM, **Gaunt TR**, Lewis SJ,), Holly J, Donovan JL, Hamdy FC, Kemp JP *et al* **[Total: 58 authors]** *The effects of height and BMI on prostate cancer incidence and mortality: a Mendelian randomization study in 20,848 cases and 20,214 controls from the PRACTICAL consortium.* Cancer Causes Control. 2015 Nov;26(11):1603-16.
3. Erzurumluoglu AM, Rodriguez S, Shihab HA, Baird D, Richardson TG, Day IN, **Gaunt TR** *Identifying Highly Penetrant Disease Causal Mutations Using Next Generation Sequencing: Guide to Whole Process.* Biomed Res Int. 2015;2015:923491.
4. Männik K, Mägi R, Macé A, Cole B, Guyatt AL, Shihab HA, Maillard AM, Alavere H, Kolk A, Reigo A **(Gaunt TR position 20)** *et al* **[Total: 27 authors]** *Copy number variations and cognitive phenotypes in unselected populations.* JAMA. 2015 May 26;313(20):2044-54.
5. Relton CL, **Gaunt T**, McArdle W, Ho K, Duggirala A, Shihab H, Woodward G, Lyttleton O, Evans DM, Reik W *et al* **[Total: 19 authors]** *Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES).* Int J Epidemiol. 2015 Aug;44(4):1181-90.

6. Rodriguez S, **Gaunt TR**, Guo Y, Zheng J, Barnes MR, Tang W, Danish F, Johnson A, Castillo BA, Li YR *et al* **[Total: 23 authors]** *Lipids, obesity and gallbladder disease in women: insights from genetic studies using the cardiovascular gene-centric 50K SNP array.* Eur J Hum Genet. 2015 29.
7. Simpkin AJ, Suderman M, **Gaunt TR**, Lyttleton O, McArdle WL, Ring SM, Tilling K, Davey Smith G, Relton CL *Longitudinal analysis of DNA methylation associated with birth weight and gestational age.* Hum Mol Genet. 2015 Jul 1;24(13):3752-63.
8. Erzurumluoglu AM, Alsaadi MM, Rodriguez S, Alotaibi TS, Guthrie PA, Lewis S, Ginwalla A, **Gaunt TR**, Alharbi KK, Alsaif FM *et al* **[Total: 12 authors]** *Proxy molecular diagnosis from whole-exome sequencing reveals Papillon-Lefevre syndrome caused by a missense mutation in CTSC.* PLoS One. 2015 Mar 23;10(3):e0121351.
9. Taylor PN, Porcu E, Chew S, Campbell PJ, Traglia M, Brown SJ, Mullin BH, Shihab HA, Min J, Walter K **(Gaunt TR position 31)** *et al* **[Total: 50 authors]** *Whole-genome sequence-based analysis of thyroid function.* Nat Commun. 2015 Mar 6;6:5681.
10. King DA, Jones WD, Crow YJ, Dominiczak AF, Foster NA, **Gaunt TR**, Harris J, Hellens SW, Homfray T, Innes J *et al* **[Total: 28 authors]** *Mosaic structural variation in children with developmental disorders.* Hum Mol Genet. 2015 May 15;24(10):2733-45.
11. Shihab HA, Rogers MF, Gough J, Mort M, Cooper DN, Day IN, **Gaunt TR**, Campbell C *An integrative approach to predicting the functional effects of non-coding and coding sequence variation.*Bioinformatics. 2015 May 15;31(10):1536-43.
12. Seoane JA, Campbell C, Day IN, Casas JP, **Gaunt TR** *Canonical correlation analysis for gene-based pleiotropy discovery.* PLoS Comput Biol. 2014 Oct 16;10(10):e1003876.
13. Alsaadi MM, Erzurumluoglu AM, Rodriguez S, Guthrie PA, **Gaunt TR**, Omar HZ, Mubarak M, Alharbi KK, Al-Rikabi AC, Day IN. *Nonsense mutation in coiled-coil domain containing 151 gene (CCDC151) causes primary ciliary dyskinesia.* Hum Mutat. 2014 Dec;35(12):1446-8.
14. Shihab HA, Gough J, Mort M, Cooper DN, Day IN, **Gaunt TR** *Ranking non-synonymous single nucleotide polymorphisms based on disease concepts.* Hum Genomics. 2014 Jun 30;8:11.
15. **Gaunt TR**, Zabaneh D, Shah S, Guyatt A, Ladroue C, Kumari M, Drenos F, Shah T, Talmud PJ, Casas JP *et al* **[Total: 18 authors]** *Gene-centric association signals for haemostasis and thrombosis traits identified with the HumanCVD BeadChip.* Thromb Haemost. 2013 Nov;110(5):995-1003.
16. Shihab HA, Gough J, Cooper DN, Day IN, **Gaunt TR.** *Predicting the functional consequences of cancer-associated amino acid substitutions.* Bioinformatics. 2013 Jun 15;29(12):1504-10.
17. Zheng J, **Gaunt TR**, Day IN. *Sequential sentinel SNP Regional Association Plots (SSS-RAP): an approach for testing independence of SNP association signals using meta-analysis data.* Ann Hum Genet. 2013 Jan;77(1):67-79.
18. **Gaunt TR**, Lowe GD, Lawlor DA, Casas JP, Day IN. *A gene-centric analysis of activated partial thromboplastin time and activated protein C resistance using the HumanCVD focused genotyping array.* Eur J Hum Genet. 2013 Jul;21(7):779-83.
19. **Gaunt TR**, Shah S, Nelson CP, Drenos F, Braund PS, Adeniran I, Folkersen L, Lawlor DA, Casas JP, Amuzu A *et al* **[Total: 25 authors]** *Integration of genetics into a systems model of electrocardiographic traits using HumanCVD BeadChip.* Circ Cardiovasc Genet. 2012 Dec;5(6):630-8.
20. Asselbergs FW, Guo Y, van Iperen EP, Sivapalaratnam S, Tragante V, Lanktree MB, Lange LA, Almoguera B, Appelman YE, Barnard J **(Gaunt TR position 15)** *et al* **[Total: 170 authors]** *Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci.* Am J Hum Genet. 2012 Nov 2;91(5):823-38.
21. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, Day IN, **Gaunt TR.** *Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models.* Hum Mutat. 2013 Jan;34(1):57-65.
22. Alsaadi MM, **Gaunt TR**, Boustred CR, Guthrie PA, Liu X, Lenzi L, Rainbow L, Hall N, Alharbi KK, Day IN. *From a single whole exome read to notions of clinical screening: primary ciliary dyskinesia and RSPH9 p.Lys268del in the Arabian Peninsula.* Ann Hum Genet. 2012 May;76(3):211-20.