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

Application Summary	
GeCIP domain name	**Immune Disorders**
Project title *(max 150 characters)*	**Towards a comprehensive genetic architecture for heritable immune disorders**

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

The overarching aim of the proposed programme is to elucidate molecular mechanism in novel inherited disorders of immunity. This work will address a series of interrelated scientific, clinical and strategic objectives:

Scientific objectives:
1. Enrich understanding of the phenotypic diversity and natural history of rare disorders of the immune system associated with known genes
2. Discover causative variants in novel disease genes & their pathogenic mechanism
3. Discover pathogenic variants in non-coding space and their disease mechanism

Clinical objectives:
1. Enable early diagnosis and genetic counselling
2. Inform therapy for affected patients to achieve better health outcomes
3. Facilitate the gathering of information regarding natural history of individual disorders

Strategic objectives:
1. Extend knowledge and understanding of the genetic basis of immune disorders and implications for health and disease
2. Harmonise, coordinate and build national capacity for genomic diagnostics and research in these disorders

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 immune system is among the most complicated parts of the human body. We need an array of specialized white blood cells to fight infection and protect against cancer. To produce this complex system, each cell relies on detailed instructions in the form of DNA. Spelling mistakes in DNA can cause disease such as infection or inflammation, particularly in children but also increasingly recognised in adults. To treat patients better and advise families we need to uncover the underlying genetic spelling mistakes. This used to be an almost impossible task, but with new techniques we can read the entire DNA sequence of

individual patients. This is called their genome; every person's genome is unique. We are still learning how to make sense of all the information it contains and especially how to pick out which genetic changes cause disease. In this proposal, scientists will focus on studying the genomes of patients with immune problems. We hope to learn more about which genes are important for immunity. Scientists will work with doctors to bring better DNA tests to the NHS so that families can be helped quicker. Eventually new treatments may become available but this generally takes a very long time.

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)*

Inherited disorders of the immune system represent a rich resource for hypothesis-generating research that improves human health. By comparison with genetically engineered mouse models, a forward genetic approach offers inherent scientific advantages such as (i) increased relevance of findings to human health in the natural environment, (ii) a definite and clinically important phenotype, (iii) freedom from preconceptions as to the genetic basis and disease mechanism. Furthermore, **both the severity of associated disease and the potential for curative precision medicine mandate a molecular approach to diagnosis for individual patients**. The 100,000 Genomes Project now provides the opportunity to address these twin imperatives by interrogating variation across the entire genome of individuals with otherwise unexplained immune disorders.

In our capacity as Clinical Interpretation Partners, **our overriding aim is to deliver the benefits of molecular diagnosis to participating patients, their close relatives and the NHS**. Allied to this, our immediate research objectives are as follows:

1. Enriched understanding of phenotypic diversity and natural history of disorders associated with known disease genes. We will perform detailed clinical and laboratory assessments of patients with variants in shared disease genes, including the outcome of any therapeutic interventions. By integrating this new knowledge with prior experience (including the UKPIN registry of PID patients), we will enhance understanding of genotype-phenotype correlation and discover allelic disorders.

2. Discovery of causal coding variants in novel disease genes & their pathogenic mechanism. Using established and novel methods to interrogate and filter variants within coding regions, we will identify novel disease genes in patients grouped by disease phenotype or familial relationship. We will perform confirmatory studies at mRNA, protein and functional level to validate the link between variant and disease and to generate hypotheses for further research, including experimental medicine studies.

3. Discovery of pathogenic variants in non-coding space and their disease mechanism. A major strength within our GeCIP is the availability of data and expertise

regarding the annotation of non-coding space in immune and blood cells to International Human Epigenome Consortium standards. The early integration of these data into our analyses of variants in the non-coding space will accelerate the discovery of novel disease-causing or disease-modifying variants. Disease mechanism will be explored through transcriptional analysis in relevant patient material as well as model *in vitro* systems that have already proved powerful for the interrogation of gene regulation.

In order to carry out this research, we have assembled a large team of clinicians, clinical scientists, wet lab and computational biologists, drawn from both the NHS and academia. We will function as an extended research network with centres of expertise in particular disease areas. To benefit maximally from the potential research synergy, our intention will be to secure collaborative funding to support a shared analytic and administrative infrastructure, including dedicated bioinformatic support, a communications hub and a programme of regular virtual and face-to-face meetings.

Expected start date	**February 2017**
Expected end date	**January 2022**

Lead Applicant 1

Name	Sophie Hambleton
Post	Professor of Paediatrics and Immunology, Honorary Consultant in Paediatric Immunology and Infectious Diseases
Department	Institute of Cellular Medicine
Institution	Newcastle University Great North Children's Hospital, Newcastle upon Tyne
Current commercial links	None

Administrative Support

Name	Roz Gale
Email	roz.gale@ncl.ac.uk

Telephone	0191 282 5234

Subdomain leads		
Name	**Subdomain**	**Institution**
To be agreed, subject to merger of 100,000 Genomes and NIHR-BR-RD projects		

Detailed research plan

Full proposal (total max 1500 words per subdomain)	
Title *(max 150 characters)*	**Towards a comprehensive genetic architecture for heritable immune disorders**

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).*

Primary immune deficiencies (PID) cause substantial morbidity and mortality and demand significant healthcare resources. Although many of these disorders are curable by haemopoietic stem cell transplantation (HSCT), this remains a risky procedure and many patients instead endure chronic ill-health despite supportive therapy.

Familial occurrence of these diseases, as well as their enrichment in consanguineous families, indicate that many are monogenic disorders, albeit a considerable fraction is caused by *de novo* variants. Disease-causing variants have already been identified in >300 genes, leading to major advances in scientific understanding as well as manifest clinical benefits. A confirmed molecular diagnosis increases the confidence with which both natural history and treatment response can be predicted, enabling timely and tailored therapy with improved outcomes. For the affected family, a molecular diagnosis also brings with it the important possibilities of early (even pre-symptomatic or prenatal) diagnosis, pre-emptive therapy and genetic counselling.

With improved mechanistic understanding there may also be opportunities for targeted interventions such as small molecule inhibitors or biologic therapies. PID are therefore emerging as an important area for the development of precision medicine. A classic example is SCID (severe combined immunodeficiency), in which knowledge of the underlying gene defect enables personalised treatment regimens such as gene therapy for

ADA and X-SCID as well as reduced intensity conditioning for HSCT in patients with radiosensitive disorders like DNA Ligase 4 deficiency. Other examples include immunoglobulin replacement for patients with XLA and Cd79a deficiency, Sirolimus for patient with PI3Kd mutations among others and CTLA4-mimetics for those with deficiency of CTLA4 or LRBA.

GeCIP members have led research resulting in the discovery of >25 new disease genes through technologies including whole exome sequencing (WES), yielding important advances in understanding of disease mechanisms, clinical prognosis and therapy. For example, the discovery of *GATA2* haploinsufficiency, NHS provision of diagnostic testing and linked natural history studies have led to its recognition as a major cause of immune deficiency and familial myelodysplasia and an important indication for HSCT.

While WES has proved very informative, a substantial proportion (~10%) of the coding genome is typically not captured and potential pathogenic variants involving noncoding regulatory DNA causing or contributing to presenting immune phenotypes remain unresolved. There is significant potential to boost diagnostic yield and scientific discovery by harnessing the power of whole genome sequencing (WGS) in heritable immune disorders to advance the following areas:

- understanding of the phenotypic diversity and natural history of known inherited immune disorders (including response to therapy)
- novel disease phenotypes associated with known disease genes (i.e. allelic disorders associated with alternate effects on protein function)
- causal coding variants in novel disease genes & their pathogenic mechanism
- causal non-coding variants, their effect on gene regulation and their disease mechanism
- epistatic effects on the expression of disease phenotype
- integration with data arising from mapping genetic modulators of multifactorial immune traits involving common and rare genetic variants together with systems biology approaches to understand immune dysfunction
- drug target discovery and validation

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.*

Phenotypes of interest:
We are interested in interrogating the full range of immune disorders fulfilling eligibility criteria for the 100,000 Genomes Project.

Aim 1: Enriched understanding of phenotypic diversity and natural history of disorders associated with known disease genes

Members of this GeCIP have extensive experience of compiling and publishing multi-centre

case series to share clinical knowledge and experience. An expected outcome of the widespread application of WGS to patients with rare disorders is the identification of unanticipated defects in known genes. This will enable improved understanding of the phenotypic spectrum and natural history of individual disorders. Sometimes this may amount to a demonstration that entirely different diseases can be associated with alternative effects on the same gene (e.g. contrasting allelic disorders associated with gain- or loss-of-function variants). We are keen to capture this clinical information, matching it with results of refined analysis at the laboratory level and coding these data with Human Phenotype Ontology (HPO) terms. Individual centres have established capacity to perform extended laboratory investigations for particular areas of interest. There is also considerable experience of clinical and laboratory data capture including the development of new HPO terms. Together, this additional information may help to illuminate both final common pathways and potential diagnostic biomarkers of particular disease behaviours.

Aim 2: Discovery of causal coding variants in novel disease genes & their pathogenic mechanism

Experience to date suggests that the genetic architecture underlying as yet unresolved disorders is dispersed over a large number of new loci, many of which will encode proteins that are highly connected in protein-protein interaction networks that control immunity. The overall goal of this aim will be to advance understanding of the genetic basis of specific diseases, the function of the immune system in health and disease, and to enable drug target discovery and validation. As well as classical monogenic disorders we may encounter more complex inheritance patterns such as two rare alleles encoding proteins in the same pathway which create bi-genic pathway insufficiencies, combinations of rare CNVs and relatively common SNVs and occasionally, epigenetic mechanisms such as imprinting.

Open-ended analysis of variants throughout the virtual exome will be undertaken in light of the pedigree and clinical context (see below). Novel candidate disease-causing variants will be screened on the basis of expression pattern and function before being taken forward for functional validation. Putative disease-causing variants will first be verified by dideoxy sequencing of PCR amplicons of the relevant region of patient genomic DNA (or by MLPA for deletions). Segregation of the variant within the family will be studied by similar analysis of genomic DNA from parents, siblings &, where relevant, other family members, to confirm the expected pattern of inheritance. Subsequent assays will be designed to explore the effects of the variant(s) on mRNA and protein abundance by quantitative RT-PCR and immunoblotting +/- immunofluorescence microscopy/cytometry, respectively. Ideally these analyses will be carried out in patient cells (dermal fibroblasts, peripheral blood or bone marrow, B lymphoblastoid cell lines, induced pluripotent stem cells (iPSC) or iPSC-derived blood cells), and/or the variant(s) will be modelled in cultured or primary cells by standard techniques. For missense mutations with residual protein expression, the structural effect of amino acid substitutions will be modelled into known molecular structures – this may give insight into their effect on protein function. Further, functional studies will be tailored to the

protein of interest and draw on a wide range of core assays already established within the laboratories of the GeCIP participants. These assays will be applied in a selective manner and focus on subsets of immune or blood cells most likely affected by the altered protein function.

These assays include but are not limited: to flow cytometric assessment of cell proliferation and survival; apoptosis and cell death; cytokine production and capture, ELIspot and ELISAs; detailed assays of platelet aggregation and coagulation; qPCR (including nanofluidic methodology); clonogenic and stem cell differentiation assays; signalling assays including NFKB, JAK-STAT, TLR and MAPK pathways (Western blotting, flow cytometry; ImageStream); radiosensitivity analysis, assays of energy metabolism, cell adhesion and chemotaxis; Treg suppression, B7 transendocytosis, EMSA, luciferase enhancer and promoter assays, etc. Expertise in more sophisticated disease models, including patient iPSC (+/- CRISPR gene correction), xenografts, mouse knock-ins/knock-outs, Drosophila, Xenopus and zebrafish, is also available within the GeCIP as well as by collaboration and may have an important part to play, particularly for those genes not previously linked to immunity.

Aim 3 Discovery of pathogenic variants in non-coding space and their disease mechanism

Although we anticipate discovery of some conventional disease genes as above ('low hanging fruit'), it appears likely that remaining unresolved disorders are enriched for non-conventional genetic aetiology. In particular, we expect to encounter rare pathogenic *cis*-acting non-coding variants, the disease mechanism of which relates to an impaired transcriptional programme. Several such regulatory variants have already been identified in relation to known disease genes; for example the disruption of key enhancer sequences may prevent gene expression in *cis*, amounting to a null allele (eg *GATA2*, *UNC13D*) or insufficient transcription /translation (eg *RBM8A*). However we might imagine other, less gene-centric effects, for example with respect to the long-range spatial organisation of chromatin within the nucleus.

Key to our ability to identify and investigate noncoding variants will be the adoption of an integrated approach that leverages functional annotation for specific immune cell types implicated in disease pathogenesis. We will use publically accessible –omic datasets generated for example by the International Human Epigenome Consortium, the Roadmap Epigenomics Project and Human Functional Genomics Project. We will complement this by data generation for patient cells in specific disease phenotypes. We will analyse differential transcriptomic, epigenomic and proteomic profiles among affected and unaffected individuals for relevant immune cell populations. This will be analysed using a systems biology-led approach to define modulated gene sets and pathways, together with gene interaction networks by leveraging knowledge of protein-protein interactions. These data will enable focused functional annotation and interpretation of the observed genetic variation at the DNA sequence level, complemented by annotation of regulatory elements such as enhancers and of non-coding RNAs.

Data arising will be further interpreted by functional assays such as already described together with genome editing, including high throughput CRISPR loss- and gain-of-function screens for specific genes and regulatory elements, and to test hypotheses regarding causal relationships with specific sequence variants by allelic substitution and rescue. We will interpret data in the context of noncoding variants involved in multifactorial immune traits and observed variation in immune response among healthy individuals. Results will inform clinical interpretation and feedback of findings for individual patients, and be taken forward to advance drug target discovery and validation based on bioinformatic approaches that draw upon genetic data, human and mouse ontology annotations, protein-protein interactions and pathways being developed by members of the GeCIP.

We suggest that Mendelian disorders of the immune and blood systems have an important role to play as models in which to explore the functions of the non-coding space. However it is clear that such enquiry pushes the frontiers of current scientific understanding and substantial bioinformatic and statistical capacity will be required to make this aspect of the research successful.

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.*

Members of our domain are active in a variety of funded research programmes, the aims of which mirror the 100,000 Genomes Project, as well as several overlapping GeCIP domains including haematology, paediatric sepsis and quantitative methods. These individuals embody the collaborative relationship between projects and include representation from national initiatives on rare disease research as they relate to immune disorders. Thus immune disorders leads of the NIHRBR-RD and the NIHR Rare Diseases Translational Research Collaboration (Immune Disorders) are active members of the GeCIP as are several members of the 3Is consortium. External collaborators who have specifically expressed interest in working cooperatively with the GeCIP domain include Lenardo, Su, Milner and Holland (NIH), Latour (Institut Imagine, Paris), Oksenhendler (Hopital Saint-Louis, Paris), Westerberg and Bryceson (Karolinska Institutet, Stockholm) and Brinkman (Vancouver).

Through our overlapping membership we reach out to the wider clinical community of immunologists (UK Primary Immunodeficiency Network) and paediatricians. Susan Walsh, lead of the UK patient organisation PID-UK is a member of our GeCIP and many of our clinical members are actively engaged in outreach activities, ensuring that patient voices are heard.

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

Members of this GeCIP are already actively engaged in delivering the Masters of Genomic Medicine at their local Genomic Medicine Centres (GMCs) and a Clinical Sciences MPhil in Rare Diseases at Cambridge University. Furthermore we are raising awareness of genomic medicine and the 100,000 Genomes Project within our clinical communities, through

involvement with local GMCs and by participation in clinical and scientific meetings of national and international 'learned' societies.

There is widespread recognition across the GeCIP of the opportunities provided by the 100,000 Genomes Project to engender an interdisciplinary approach to translational research involving clinicians, bench researchers, bioinformaticians, statisticians and experts in machine learning. We aim to train a new cadre of individuals with skills and understanding across traditional subject boundaries. Activities associated with this GeCIP will provide ample opportunity to host research projects of varying complexity and length. We will thereby deliver an ideal training and research environment for fellows in the biological and quantitative sciences and clinical fellows supported by already existing schemes of the main UK research funders such as MRC, NIHR, the Wellcome Trust and other medical charities. Movement of clinical training fellows and trainee clinical scientists between centres will be encouraged wherever this enhances training opportunities (whether research or NHS).

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

This GeCIP includes a large number of members with established track records in clinical and bench-based research relevant for the subdomains of immune disorders. These individuals are drawn from coherent and collaborative research communities centred on highly specialised clinical practice. They include international leaders in their field, who run well-funded research groups often linked to and supported by an NIHR Biomedical Research Centres/Units (BRC/BRU). The quality of their research is outstanding with regular publication in high-impact journals, like Cell, Nature, Nature Genetics, Nature Medicine, New England Journal of Medicine and Science amongst others. Members within each subdomain have often worked and published with each other before and will function as a collaborative unit. Key investigators include Hambleton and Collin (Newcastle), Thrasher (UCL-London), Smith (Cambridge), Sansom and Burns (UCL-RFH), Patel and Knight (Oxford), Arkwright (Manchester), Williams (Southampton), Savic (Leeds) and Ibrahim (Kings), between them representing the majority of GMC-linked academic centres and bringing complementary expertise and interests.

In addition, the GeCIP has a credible and talented membership representing the quantitative sciences with experts in bioinformatics (both clinical and genomics), computational biology and statistical genomics. These individuals are drawn from a range of academic centres including the Wellcome Trust Centre for Human Genetics (Oxford, eg Knight, Taylor) and the Wellcome Trust Sanger Institute (Cambridge, eg Teichmann, Anderson). Collaboration with the non-malignant and malignant haematology domains has also been established with two residential meetings and overlapping membership. Our GeCIP will foster this working relationship as the development of new analytical methods will be critical for the successful delivery of the proposed programme of research and the translation of research findings into clinical care delivery.

We will function as an extended research network. To benefit maximally from the potential research synergy, our intention will be to secure collaborative funding to support a shared analytic and administrative infrastructure, including dedicated bioinformatic support, a communications hub and a programme of regular virtual and face-to-face meetings.

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

The validation and reporting of variants within **known disease genes** is a core activity of existing clinical diagnostic services, which are strongly represented within the GeCIP, including the lead for the *Validation and Reporting Working Group* (Gilmour, GOSH).

Diagnostic next generation sequencing (NGS) platforms for panels of genes relevant to the immune subdomains are available at GOSH (~70 PID genes along with others causing isolated gut inflammation; MDT chaired by Dr Kimberly Gilmour). The NGS panels are available as an accredited clinical diagnostic test with defined validation and reporting processes. Establishing and administering these tests has yielded useful insights into the bioinformatic and interpretational strengths and challenges of NGS in a clinical diagnostic setting, with relevance to reporting services provided by the GMCs. In particular this experience highlights the added value of integrating corroboratory evidence at protein and/or functional level and strong representation from a range of expert clinicians including clinical geneticists, providing the correct forum for patient-centric interpretation of genetic data. We are keen to offer support to GMCs in reporting all Tier 1 variants in immune-associated genes and in turn will seek their support in the confirmation of putative novel pathogenic variants by Sanger sequencing.

Judging when variants in **novel disease genes or non-coding regions** reach the status of actionable clinical findings is an ongoing challenge within the rare disease community. The GeCIP will adhere to international and national guidelines and follow a conservative approach before genes and variants therein are moved to Tier 1 status. We are aware of the importance of balancing the desire to deliver patient benefit by rapid reporting with the potential risks of incorrect cataloguing of variants. The GeCIP and its subdomains provide a forum in which to consider evidence for disease causation as it is generated. In practice we expect to assemble a case for the designation of novel variants as disease-causing and present it to the *Validation and Feedback* domain for their approval prior to publication and reporting.

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

We know from experience that accurate molecular diagnosis empowers patients, families and their healthcare teams to deliver best care. These benefits will attach not only to those enrolled in the project but also to patients with the same disorder globally, who will be more likely to achieve a molecular diagnosis as a result of our discoveries. As clinicians and

clinical scientists already working with patients in the NHS, and fully engaged with international learned societies, members of this GeCIP are well-placed to ensure the incorporation of new knowledge into routine diagnostics. At a minimum, a molecular diagnosis enables screening of family members and genetic counselling, including the possibilities of prenatal and preimplantation genetic diagnosis where appropriate. Molecular diagnosis is required for gene therapy, trials of which are already underway for ADA SCID, XSCID, WAS and XCGD with other being developed. Over time, clinicians learn more about the behaviour of individual disorders, which informs understanding of prognosis and hence the appropriateness of risky but potentially curative therapies such as stem cell transplantation – including knowing when not to transplant. Early ascertainment of the molecular basis of disease improves the accuracy of therapeutic intervention including the potential for precision medicine. A growing literature describes the experimental application of molecularly targeted therapies to immune disorders, for example supplemental magnesium in boys with MAGT1 deficiency and CTLA4-mimetics in deficiency of CTLA4 or LRBA. We predict further examples of 'first in man' experimental medicine will follow as a direct consequence of the 100,000 Genomes Project. Although often stereotyped as highly costly therapies, recent examples all represent repurposing of drugs already licensed for other indications, yet offering transformative improvement in patients' wellbeing. Furthermore, discoveries within our subdomain have the potential to inform scientific understanding of fundamental processes relevant to common disorders such as autoimmune, malignant and inflammatory conditions.

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?*

GeCIP members will be guided by GEL and their host institutions to optimise the 'return of investment' in the 100,000 Genomics Project for UK PLC, including through commercial partnership.

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

Angulo I, Vadas O, Garçon F, Banham-Hall E, Plagnol V, Leahy TR, Baxendale H, Coulter T, Curtis J, Wu C, Blake-Palmer K, Perisic O, Smyth D, Maes M, Fiddler C, Juss J, Cilliers D, Markelj G, Chandra A, Farmer G, Kielkowska A, Clark J, Kracker S, Debre M, Picard C, Pellier I, Jabado N, Morris JA, Barcenas-Morales G, Fischer A, Stephens L, Hawkins P, Barrett JC, Abinun M, Clatworthy M, Durandy A, Doffinger R, Chilvers ER, Cant AJ, Kumararatne D, Okkenhaug K, Williams RL, Condliffe A, Nejentsev S. Phosphoinositide 3-kinase δ gene mutation predisposes to respiratory infection and airway damage. Science. 2013;342(6160):866-71.

Bolze A, Byun M, McDonald D, Morgan NV, Abhyankar A, Premkumar L, Puel A, Bacon CM, Rieux-Laucat F, Pang K, Britland A, Abel L, Cant A, Maher ER, Riedl SJ, Hambleton S, Casanova JL. Whole-exome-sequencing-based discovery of human FADD deficiency.

Am J Hum Genet. 2010 Dec 10;87(6):873-81.

Booth C, Gilmour KC, Veys P, Gennery AR, Slatter MA, Chapel H, Heath PT, Steward CG, Smith O, O'Meara A, Kerrigan H, Mahlaoui N, Cavazzana-Calvo M, Fischer A, Moshous D, Blanche S, Pachlopnik Schmid J, Latour S, de Saint-Basile G, Albert M, Notheis G, Rieber N, Strahm B, Ritterbusch H, Lankester A, Hartwig NG, Meyts I, Plebani A, Soresina A, Finocchi A, Pignata C, Cirillo E, Bonanomi S, Peters C, Kalwak K, Pasic S, Sedlacek P, Jazbec J, Kanegane H, Nichols KE, Hanson IC, Kapoor N, Haddad E, Cowan M, Choo S, Smart J, Arkwright PD, Gaspar HB. X-linked lymphoproliferative disease due to SAP/SH2D1A deficiency: a multicenter study on the manifestations, management and outcome of the disease. Blood. 2011 Jan 6;117(1):53-62.

Brent J, Guzman D, Bangs C, Grimbacher B, Fayolle C, Huissoon A, Bethune C, Thomas M, Patel S, Jolles S, Alachkar H, Kumaratne D, Baxendale H, Edgar JD, Helbert M, Hambleton S, Arkwright PD. Clinical and laboratory correlates of lung disease and cancer in adults with idiopathic hypogammaglobulinaemia. Clin Exp Immunol. 2016 Apr;184(1):73-82.

Burns SO, Plagnol V, Gutierrez BM, Al Zahrani D, Curtis J, Gaspar M, Hassan A, Jones AM, Malone M, Rampling D, McLatchie A, Doffinger R, Gilmour KC, Henriquez F, Thrasher AJ, Gaspar HB, Nejentsev S. Immunodeficiency and disseminated mycobacterial infection associated with homozygous nonsense mutation of IKKβ. J Allergy Clin Immunol. 2014 Jul;134(1):215-8.

Burns SO, Zenner HL, Plagnol V, Curtis J, Mok K, Eisenhut M, Kumararatne D, Doffinger R, Thrasher AJ, Nejentsev S. LRBA gene deletion in a patient presenting with autoimmunity without hypogammaglobulinemia. J Allergy Clin Immunol. 2012 Dec;130(6):1428-32.

Coulter TI, Chandra A, Bacon CM, Babar J, Curtis J, Screaton N, Goodlad JR, Farmer G, Steele CL, Leahy TR, Doffinger R, Baxendale H, Bernatoniene J, Edgar JD, Longhurst HJ, Ehl S, Speckmann C, Grimbacher B, Sediva A, Milota T, Faust SN, Williams AP, Hayman G, Kucuk ZY, Hague R, French P, Brooker R, Forsyth P, Herriot R, Cancrini C, Palma P, Ariganello P, Conlon N, Feighery C, Gavin PJ, Jones A, Imai K, Ibrahim MA, Markelj G, Abinun M, Rieux-Laucat F, Latour S, Pellier I, Fischer A, Touzot F, Casanova JL, Durandy A, Burns SO, Savic S, Kumararatne DS, Moshous D, Kracker S, Vanhaesebroeck B, Okkenhaug K, Picard C, Nejentsev S, Condliffe AM, Cant AJ. Clinical spectrum and features of activated phosphoinositide 3-kinase δ syndrome: A large patient cohort study. J Allergy Clin Immunol. 2016 Jul 16. pii: S0091-6749(16)30623-6.

Davenport EE, Burnham KL, Radhakrishnan J, Humburg P, Hutton P, Mills TC, Rautanen A, Gordon AC, Garrard C, Hill AV, Hinds CJ, Knight JC. Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study. Lancet Respir Med. 2016 Apr;4(4):259-71.

Dickinson RE, Griffin H, Bigley V, Reynard LN, Hussain R, Haniffa M, Lakey JH, Rahman T, Wang XN, McGovern N, Pagan S, Cookson S, McDonald D, Chua I, Wallis J, Cant A, Wright M, Keavney B, Chinnery PF, Loughlin J, Hambleton S, Santibanez-Koref M, Collin M. Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. Blood. 2011;118(10):2656-8.

Dickinson RE, Milne P, Jardine L, Zandi S, Swierczek SI, McGovern N, et al. The evolution of cellular deficiency in GATA2 mutation. Blood. 2014;123(6):863-74.

Duncan CJ, Mohamad SM, Young DF, Skelton AJ, Leahy TR, Munday DC, et al. Human IFNAR2 deficiency: Lessons for antiviral immunity. Sci Transl Med. 2015;7(307):307ra154.

Edgar JD, Buckland M, Guzman D, Conlon NP, Knerr V, Bangs C, Reiser V, Panahloo Z, Workman S, Slatter M, Gennery AR, Davies EG, Allwood Z, Arkwright PD, Helbert M, Longhurst HJ, Grigoriadou S, Devlin LA, Huissoon A, Krishna MT, Hackett S, Kumararatne DS, Condliffe AM, Baxendale H, Henderson K, Bethune C, Symons C, Wood P, Ford K, Patel S, Jain R, Jolles S, El-Shanawany T, Alachkar H, Herwadkar A, Sargur R, Shrimpton A, Hayman G, Abuzakouk M, Spickett G, Darroch CJ, Paulus S, Marshall SE, McDermott EM, Heath PT, Herriot R, Noorani S, Turner M, Khan S, Grimbacher B. The United Kingdom Primary Immune Deficiency (UKPID) Registry: report of the first 4 years' activity 2008-2012. Clin Exp Immunol. 2014 Jan;175(1):68-78.

Engelhardt KR, Gertz ME, Keles S, Schaeffer AA, Sigmund EC, Glocker C, Saghafi S, Pourpak Z, Ceja R, Sassi A, Graham LE, Massaad MJ, Mellouli F, Ben-Mustapha I, Khemiri M, Kilic SS, Etzioni A, Freeman AF, Thiel J, Schulze I, Al-Herz W, Metin A, Sanal √ñ, Tezcan I, Yeganeh M, Niehues T, Dueckers G, Weinspach S, Patiroglu T, Unal E, Dasouki M, Yilmaz M, Genel F, Aytekin C, Kutukculer N, Somer A, Kilic M, Reisli I, Camcioglu Y, Gennery AR, Cant AJ, Jones A, Gaspar BH, Arkwright PD, Pietrogrande MC, Baz Z, Al-Tamemi S, Lougaris V, Lefranc G, Megarbane A, Boutros J, Galal N, Bejaoui M, Barbouche MR, Geha RS, Chatila TA, Grimbacher B. The extended clinical phenotype of 64 patients with dedicator of cytokinesis 8 deficiency. J Allergy Clin Immunol. 2015 Aug;136(2):402-12.

Gao Y, Workman S, Gadola S, Elliott T, Grimbacher B, Williams AP. Common variable immunodeficiency is associated with a functional deficiency of invariant natural killer T cells. J Allergy Clin Immunol. 2014 May;133(5):1420-8

Gaspar HB, Cooray S, Gilmour KC, Parsley KL, Zhang F, Adams S, Bjorkegren E, Bayford J, Brown L, Davies EG, Veys P, Fairbanks L, Bordon V, Petropoulou T, Kinnon C, Thrasher AJ. Hematopoietic stem cell gene therapy for adenosine deaminase-deficient severe combined immunodeficiency leads to long-term immunological recovery and metabolic correction. Sci Transl Med. 2011 Aug 24;3(97):97ra80.

Hacein-Bey-Abina S, Pai SY, Gaspar HB, Armant M, Berry CC, Berry CC, Blanche S, Bleesing J, Blondeau J, de Boer H, Buckland KF, Caccavelli L, Cros G, De Oliveira S, Fernández KS, Guo D, Harris CE, Hopkins G, Lehmann LE, Lim A, London WB, van der Loo JC, Malani N, Male F, Malik P, Marinovic MA, McNicol AM, Moshous D, Neven B, Oleastro M, Picard C, Ritz J, Rivat C, Schambach A, Shaw KL, Sherman EA, Silberstein LE, Six E, Touzot F, Tsytsykova A, Xu-Bayford J, Baum C, Bushman FD, Fischer A, Kohn DB, Filipovich AH, Notarangelo LD, Cavazzana M, Williams DA, Thrasher AJ. A modified γ-retrovirus vector for X-linked severe combined immunodeficiency. N Engl J Med. 2014;371(15):1407-17.

Hambleton S, Salem S, Bustamante J, Bigley V, Boisson-Dupuis S, Azevedo J, Fortin A, Haniffa M, Ceron-Gutierrez L, Bacon CM, Menon G, Trouillet C, McDonald D, Carey P, Ginhoux F, Alsina L, Zumwalt TJ, Kong XF, Kumararatne D, Butler K, Hubeau M, Feinberg J, Al-Muhsen S, Cant A, Abel L, Chaussabel D, Doffinger R, Talesnik E, Grumach A, Duarte A, Abarca K, Moraes-Vasconcelos D, Burk D, Berghuis A, Geissmann F, Collin M, Casanova JL, Gros P. IRF8 mutations and human dendritic-cell immunodeficiency. N Engl J Med. 2011;365(2):127-38.

Hambleton S, Goodbourn S, Young DF, Dickinson P, Mohamad SM, Valappil M, McGovern N, Cant AJ, Hackett SJ, Ghazal P, Morgan NV, Randall RE. STAT2 deficiency and susceptibility to viral illness in humans. *Proc Natl Acad Sci U S A*. 2013;110:3053-3058.

Jones RA, Feng Y, Worth AJ, Thrasher AJ, Burns SO, Martin P. Modelling of human Wiskott-Aldrich syndrome protein mutants in zebrafish larvae using in vivo live imaging. J Cell Sci. 2013 Sep 15;126(Pt 18):4077-84.

Liu Y, Hanson S, Gurugama P, Jones A, Clark B, Ibrahim MA. Novel NFKB2 mutation in early-onset CVID. J Clin Immunol. 2014 Aug;34(6):686-90.

Lu W, Zhang Y, McDonald DO, Jing H, Carroll B, Robertson N, Zhang Q, Griffin H, Sanderson S, Lakey J, Morgan NV, Reynard LN, Zheng L, Murdock HM, Turbey SE, Hackett SJ, PrestidgeT, Hall JM, Cant AJ, Matthews HF, Santibanex Koref MF, Simon AK, Korolchiuk VI, Lenardo MJ, Hambleton S*, Su HC* [*joint senior author]. Dual proteolytic pathways govern glycolysis and immune competence. Cell. 2014;159(7):1578-90.

Martin E, Palmic N, Sanquer S, Lenoir C, Hauck F, Mongellaz C, Fabrega S, Nitschk√© P, Esposti MD, Schwartzentruber J, Taylor N, Majewski J, Jabado N, Wynn RF, Picard C, Fischer A, Arkwright PD, Latour S. CTP synthase 1 deficiency in humans reveals its central role in lymphocyte proliferation. Nature. 2014 Jun 12;510(7504):288-92

Milner JD, Vogel TP, Forbes L, Ma CA, Stray-Pedersen A, Niemela JE, Lyons JJ, Engelhardt KR, Zhang Y, Topcagic N, Roberson EDO, Matthews H, Verbsky JW, Dasu T, Vargas-Hernandez A, Varghese N, McClain KL, Karam LB, Nahmod K, Makedonas G,

Mace EM, Sorte HS, Perminow G, Rao VK, O'Connell MP, Price S, Su HC, Butrick M, McElwee J, Hughes J, Willet J, Swan D, Xu Y, Santibanez Koref M, Slowik V, Dinwiddie DL, Ciaccio CE, Saunders CJ, Septer S, Kingsmore SF, White AJ, Cant AJ, Hambleton S, Cooper MA. Early-onset lymphoproliferation and autoimmunity caused by germline STAT3 gain-of-function mutations. Blood. 2015;125(4):591-9.

Morgan NV, Goddard S, Cardno TS, McDonald D, Rahman F, Barge D, Ciupek A, Straatman-Iwanowska A, Pasha S, Guckian M, Anderson G, Huissoon A, Cant A, Tate WP, Hambleton S, Maher ER. Mutation in the TCR-α subunit constant gene (TRAC) leads to a human immunodeficiency disorder characterized by a lack of TCRαβ+ T cells. J Clin Invest. 2011 Feb;121(2):695-702.

Naranbhai V, Fairfax BP, Makino S, Humburg P, Wong D, Ng E, Hill AV, Knight JC. Genomic modulators of gene expression in human neutrophils. Nat Commun. 2015 Jul 7;6:7545.

Ramakrishnan KA, Pengelly RJ, Gao Y, Morgan M, Patel SV, Davies EG, Ennis S, Faust SN, Williams AP. Precision Molecular Diagnosis Defines Specific Therapy in Combined Immunodeficiency with Megaloblastic Anemia Secondary to MTHFD1 Deficiency. J Allergy Clin Immunol Pract. 2016 Nov - Dec;4(6):1160-1166.e10.

Record J, Malinova D, Zenner HL, Plagnol V, Nowak K, Syed F, Bouma G, Curtis J, Gilmour K, Cale C, Hackett S, Charras G, Moulding D, Nejentsev S, Thrasher AJ, Burns SO. Immunodeficiency and severe susceptibility to bacterial infection associated with a loss-of-function homozygous mutation of MKL1. Blood. 2015;126(13):1527-35.

Schubert D, Bode C, Kenefeck R, Hou TZ, Wing JB, Kennedy A, Bulashevska A, Petersen BS, Schäffer AA, Grüning BA, Unger S, Frede N, Baumann U, Witte T, Schmidt RE, Dueckers G, Niehues T, Seneviratne S, Kanariou M, Speckmann C, Ehl S, Rensing-Ehl A, Warnatz K, Rakhmanov M, Thimme R, Hasselblatt P, Emmerich F, Cathomen T, Backofen R, Fisch P, Seidl M, May A, Schmitt-Graeff A, Ikemizu S, Salzer U, Franke A, Sakaguchi S, Walker LS, Sansom DM, Grimbacher B. Autosomal dominant immune dysregulation syndrome in humans with CTLA4 mutations. Nat Med. 2014 Dec;20(12):1410-6.

Taylor JC, Martin HC, Lise S, Broxholme J, Cazier JB, Rimmer A, Kanapin A, Lunter G, Fiddy S, Allan C, Aricescu AR, Attar M, Babbs C, Becq J, Beeson D, Bento C, Bignell P, Blair E, Buckle VJ, Bull K, Cais O, Cario H, Chapel H, Copley RR, Cornall R, Craft J, Dahan K, Davenport EE, Dendrou C, Devuyst O, Fenwick AL, Flint J, Fugger L, Gilbert RD, Goriely A, Green A, Greger IH, Grocock R, Gruszczyk AV, Hastings R, Hatton E, Higgs D, Hill A, Holmes C, Howard M, Hughes L, Humburg P, Johnson D, Karpe F, Kingsbury Z, Kini U, Knight JC, Krohn J, Lamble S, Langman C, Lonie L, Luck J, McCarthy D, McGowan SJ, McMullin MF, Miller KA, Murray L, Németh AH, Nesbit MA, Nutt D, Ormondroyd E, Oturai AB, Pagnamenta A, Patel SY, Percy M, Petousi N, Piazza P, Piret SE, Polanco-Echeverry

G, Popitsch N, Powrie F, Pugh C, Quek L, Robbins PA, Robson K, Russo A, Sahgal N, van Schouwenburg PA, Schuh A, Silverman E, Simmons A, Sørensen PS, Sweeney E, Taylor J, Thakker RV, Tomlinson I, Trebes A, Twigg SR, Uhlig HH, Vyas P, Vyse T, Wall SA, Watkins H, Whyte MP, Witty L, Wright B, Yau C, Buck D, Humphray S, Ratcliffe PJ, Bell JI, Wilkie AO, Bentley D, Donnelly P, McVean G. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. Nat Genet. 2015;47(7):717-26.

Turley AJ, Gathmann B, Bangs C, Bradbury M, Seneviratne S, Gonzalez-Granado LI, Hackett S, Kutukculer N, Alachkar H, Hambleton S, Ritterbusch H, Kralickova P, Marodi L, Seidel MG, Dueckers G, Roesler J, Huissoon A, Baxendale H, Litzman J, Arkwright PD. Spectrum and management of complement immunodeficiencies (excluding hereditary angioedema) across Europe. J Clin Immunol. 2015 Feb;35(2):199-205.

van Schouwenburg PA, Davenport EE, Kienzler AK, Marwah I, Wright B, Lucas M, Malinauskas T, Martin HC; WGS500 Consortium, Lockstone HE, Cazier JB, Chapel HM, Knight JC, Patel SY. Application of whole genome and RNA sequencing to investigate the genomic landscape of common variable immunodeficiency disorders. Clin Immunol. 2015 Oct;160(2):301-14.

Wong GK, Millar D, Penny S, Heather JM, Mistry P, Buettner N, Bryon J, Huissoon AP, Cobbold M. Accelerated Loss of TCR Repertoire Diversity in Common Variable Immunodeficiency. J Immunol. 2016 Sep 1;197(5):1642-9.

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 welcome data on all patients with immune disorders as classified by IUIS, together with their family members. This includes all those recruited under immune disorder eligibility criteria for the Main Programme, as well as autoinflammatory disorders (currently mapped to “periodic fevers” in the renal domain). We would also like to access data on other patients with immune dysregulation, for example early onset diabetes or early onset inflammatory bowel disease, which are recognised presentations of underlying immune disorders. These data models are currently mapped to the Endocrinology and Gastroenterology domains respectively and we are happy to work with those GeCIPs to analyse these patients. In addition we have a reciprocal arrangements with the Paediatric

Sepsis and Malignant Haematology domains re data-sharing where relevant. We will require all available quantitative and qualitative phenotype data, the appended HPO codes, BAM and VCF files.

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

We will use standard methods to interrogate the coding space, to identify candidate SNVs, indels, copy number variants and structural variants. (1) Variants within known disease genes will be interrogated first using HGMD, ClinVar and gene-specific reference databases as well as allele count catalogues such as ExAC, UK10K and other large scale genome sequencing projects. (2) After completing this analysis we will consider putative novel association signals within coding space. We will be keen to apply alternative analytic strategies in a patient-centric manner as well as by grouping similar patients together. We will filter variants on the basis of rarity, predicted deleteriousness, conformation to expected inheritance pattern, segregation within families, and similarity to variants in other affected patients sharing the same phenotype. We will take into account metrics such as the gene damage index (degree to which a given gene tolerates deleterious mutations within a population), and biological factors including known interactions with relevant biomolecules. (3) Interrogation of the noncoding space will be led by GeCIP members with relevant expertise and in collaboration with colleagues in the haematology GeCIP domains. We are beginning to define the levels of functional annotation required to overlay the genome using cell-type specific IHEC reference epigenome maps and regulome builds that relate promoters and regulatory elements. We also intend to integrate information from other sources such as genome-wide association studies in a variety of immune disease states.

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)*
All available phenotype data related to patients and their family members within our domain (according to the data models we have helped to develop).

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)*

In the first instance we anticipate using the GEL analysis pipeline.

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)*

Development of a test ('sandbox') system that mirrors the setup of the real data centre, but does not contain the data of the Main Programme and thus need not be restricted to the same level with respect to data and code import and export. This test system should contain benchmarking data (e.g. the Illumina's platinum genomes, the 8,000 WGS-BAM

files from the NIHR BR-RD pilot phase or Genome In A Bottle (GIAB) consortium data) that may then be used for measuring the accuracy of the developed analysis pipeline and would enable incremental pipeline development and optimization.

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)*

NIHR BR-RD (BRIDGE-PID)
TCGA
UK Biobank GWAS datasets for blood and immune cells,
Blueprint, Roadmap and other IHEC Epigenome datasets
Interaction data between promoters and regulatory elements
Expression QTL data for blood and immune cells (joint Oxford-Cambridge initiative)
A fast suite of analytical application tools which have been developed by NIHRBR-RD and GEL bioinformaticians

Computing resource requirements. *Describe any analyses that would place high demand on computing resources and specific storage or processing implications. (max 200 words)*
We do envisage issues that will be significantly different from the rest of the rare disease community because of the potential to integrate information about the functional annotation of the genome in different blood and immune cell types. We look forward to working with Augusto Rendon and team in the new compute environment to implement this functionality.

Omics samples

Analysis of omics samples. *Summarise any analyses that you are planning using omics samples taken as part of the Project. (max 300 words)*
We currently do not envisage systematic use of 'omics samples although it is possible that serologic analyses (specific/total immunoglobulins, cytokine levels, shed receptor determinations etc) may be useful in certain circumstances. Were cryopreserved samples of peripheral blood cells to become available then these would certainly be of use for extended immunophenotypic studies as described above. Furthermore certain transcriptional analyses may be useful in whole blood RNA eg to examine extreme allelic imbalance or gene silencing.

Data access and security

GeCIP domain name	**Immune disorders**
Project title	**Towards a comprehensive genetic architecture for heritable**

(max 150 characters)	**immune disorders**

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).*

X *Clinical care*

□ *Clinical trials feasibility*

X *Deeper phenotyping*

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

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

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

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

X *Non hypothesis driven R&D - health*

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

□ *Other health use - clinical audit*

□ *Public health purposes*

X *Subject access request*

X *Tool evaluation and improvement*

Information Governance

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

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