Defining novel biomarkers and therapeutic targets for type 2 diabetes through genetic analysis in massive human data sets

Supervisors: Prof Mark McCarthy [1,2]
Associate Prof Cecilia Lindgren [2,3]
Prof Gil McVean [2,3]

Departments: [1] Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM), Radcliffe Department of Medicine
[2] Wellcome Trust Centre for Human Genetics (WTCHG), Nuffield Department of Medicine
[3] Big Data Institute

Background

Type 2 diabetes (T2D) and related traits, including obesity, are major contributors to ill health globally. Lifestyle modification has proven to be of limited efficacy, and there is a critical need for novel strategies that provide effective prevention and treatment, supported by robust stratification of individual risk and disease subtype. However, progress is hampered by poor understanding of the mechanisms responsible for disease development and progression. Human genetics offers a singularly powerful way to highlight pathways of relevance to human disease, and recent developments have focused attention on the value of such information to guide drug and biomarker development.

McCarthy & Lindgren are world leaders in the genetics and genomics of T2D and obesity. They have led global consortia that identified and characterised most of the (~300) risk variants detected, and, with colleagues, have made substantial strides towards delivering functional insights into disease development. This proposal seeks to build on this platform of discovery by advancing understanding of disease mechanisms in ways that have direct translational relevance. McVean brings world-leading strengths in the analysis of large genetic and genomic data sets ("big data").

Hypothesis

This project revolves around the hypothesis that **the integration of diverse types of large-scale human genetic and genomic ("big") data, in a framework of robust causal inference, provides a powerful strategy for the identification and characterisation of genes, proteins, metabolites and pathways that are directly responsible for the development and progression of T2D and related traits (including obesity).** Such discoveries offer particular translational value because the novel biomarkers and therapeutic targets highlighted by such research speak primarily to their relevance to human disease, avoiding the assumptions implicit in the use of preclinical models.

The broad aims of the research to be performed by the fellow are to:

- Identify genetic "predictors" of molecular and genomic traits (e.g. RNA and metabolomic phenotypes) through analyses of genomic data sets informative for tissue-specific expression and for circulating analytes;
- Implement "two sample mendelian randomisation" and related methods to evaluate these predictors within massive human genetic data sets, identifying molecular phenotypes with variant association profiles that overlap those of T2D and/or obesity;
- Evaluate and validate the causal nature of those relationships using informatic, statistical, epidemiological and empirical analyses; and
- Advance confirmed validated biomarkers and therapeutic targets for further mechanistic evaluation and translational development, benefitting from collaborations with the newly established **Novo Nordisk Research Centre Oxford**.

Description of the work

The primary research focus will be computational and statistical. The initial focus of activities will involve: (a) analysis of molecular phenotype data gathered from tissues of direct relevance to T2D and obesity (including RNA expression data from human pancreatic islets, fat and skeletal muscle; serum proteomic and metabolomics data), to characterise single- and multi-variant predictors of those phenotypes; (b) interrogation of already generated human GWAS and sequencing data-sets, including those described above, accessing data from >1M individuals; and (c) application of statistical methods (two sample mendelian randomisation, latent variable model selection) to define causal relationships between these different data domains.

To achieve these goals, the supervisors are able to assemble a unique constellation of data including:

- Unparalleled collections of human genetic data describing the relationships between genetic variation and T2D and obesity phenotypes, including: (a) genome wide association (GWAs) and/or exome array data gathered from over 1M individuals (including data from several large biobanks including UK Biobank [500k], GERA [100k], and in the medium-term Chinese Kadoorie Biobank [510K] and Mexico Diabetes Study [200K]); (b) whole genome and exome sequence data from >100K individuals (via the T2D-GENES and TOPMED projects, amongst others);
- Increasingly rich and complex descriptions of relationships between genetic variation and (a) RNA expression from key tissues relevant to T2D and obesity (islets, fat, muscle, liver, brain, via GTEx and INSPIRE); and (b) complex 'omic' phenotypes including the plasma proteome, metabolome, and the microbiome (e.g. from Oxford Biobank, DIRECT, UK Twins).

The objective will be to identify molecular phenotypes and pathways "causal" for disease: in other words, those that are responsible for mediating the action of genetic (and possibly non-genetic) risk factors on disease development and progression. The key advance here is that the inclusion of high density genetic data provides a "root to the tree of causation" that enables primary (causal) and secondary (reactive) changes in phenotype level to be disentangled. The validated molecules which emerge from these analyses have enhanced translational potential with respect to: (a) stratification of individual disease risk; (b) stratification of disease subtype; (c) monitoring need for and response to intervention; and (d) prioritisation of validated therapeutic targets for drug development. A

particularly apposite example of the clinical impact of such a molecular biomarker is the use of cholesterol in relation to vascular disease.

The first phase of these analyses should be complete in year 1 of the fellowship, generating candidate mechanisms, targets and biomarkers. Depending on interest and aptitude, the fellow will then be able to pursue follow-up in a variety of alternative directions including: (a) extension of these methods to additional traits and reference data sets (e.g. within UKBiobank); (b) confirmation of T2D/obesity biomarkers in ongoing, prospective studies; (c) mechanistic studies in cellular or animal models; and/or (d) proof-of-principle intervention studies in clinical samples. We will be particularly interested in pursuing mechanistic follow-up with colleagues in the new Novo Nordisk Research Centre in Oxford.

This project will provide the fellow with training in a wide range of computational and statistical techniques, exposure to the biology of T2D and obesity, and powerful opportunities in the emerging field of genomic medicine.

Recent relevant publications (McCarthy)

1. Small KS, Hedman AK, Grundberg E, Nica AC, Thorleifsson G, Kong A, Thorsteindottir U, Shin SY, Richards HB, GIANT consortium, MAGIC investigators, DIAGRAM consortium, Soranzo N, Ahmadi KR, **Lindgren CM**, Stefansson K, Dermitzakis ET, Deloukas P, Spector TD, **McCarthy MI**, MuTHER consortium (2011). Identification of an imprinted master trans-regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nature Genetics* **43**:561-564. PMID: 21572415.

2. Morris AP, Voight BF, Teslovich TM, [204 authors including **Lindgren CM**] Altshuler D, Boehnke M, **McCarthy MI** for the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genetics* **44**: 981-990. PMID: 22885922.

3. Gaulton KJ, Ferreira T, Lee Y, [213 authors including **Lindgren CM**], **McCarthy MI**, Morris AP; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium (2015). Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nature Genetics* **47**: 1415-25. PMID: 26551672.

4. Horikoshi M, Beaumont RN, Day FR, [159 authors], Evans DM, **McCarthy MI**, Freathy RM (2016). Genome-wide associations for birth weight and correlations with adult disease. *Nature*. **538**: 248-252. PMID: 27680694.

5. Fuchsberger C, Flannick J, Teslovich TM, [295 authors including **Lindgren CM** and **McVean G**] Boehnke M, Altshuler D, **McCarthy MI** (2016). The genetic architecture of type 2 diabetes. *Nature* **536**: 41-47. PMID: 27398621.

Recent relevant publications (Lindgren)

1. Surendran P, Drenos F, Young R, [215 authors including **McCarthy MI**], **Lindgren CM**, Danesh J, Wain LV, Butterworth AS, Howson JM, Munroe PB (2016). Trans-ancestry meta-

analyses identify rare and common variants associated with blood pressure and hypertension. *Nat Genet* **48**: 1151-61. PMID: 27618447.

2. Fuchsberger C, Flannick J, Teslovich TM, [295 authors including **Lindgren CM** and **McVean G**], Boehnke M, Altshuler D, **McCarthy MI** (2016). The genetic architecture of type 2 diabetes. *Nature* **536**: 41-47. PMID: 27398621.

3. Lessard S, Manning AK, Low-Kam C, [32 authors], Tardif JC, **Lindgren CM**, Lettre G (2016). Testing the role of predicted gene knockouts in human anthropometric trait variation. *Hum Mol Genet* **25**: 2082-2092. PMID: 26908616.

4. Shungin D, Winkler TW, Croteau-Chonka DC, [413 authors including **McCarthy MI**], Morris AP, **Lindgren CM**, Mohlke KL (2015). New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**: 187-196. PMID: 25673412.

5. Locke AE, Kahali B, Berndt SI, [490 authors including **McCarthy MI** and **Lindgren CM**], Hirschhorn JN, Loos RJ, Speliotes EK (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**: 197-206. PMID: 25673413.

Recent relevant publications (McVean)

1. Moutsianas L, Jostins L, Beecham AH, [50 authors], McCauley JL, Sawcer S, **McVean G** for the International Multiple Sclerosis Genetics Consortium (2015). Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nat Genet* **47**: 1107-13. PMID: 26343388.

2. Taylor JC, Martin HC, Lise S, [104 authors], Bentley D, Donnelly P, **McVean G** (2015). Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat Genet* **47**: 717-26. PMID: 25985138.

3. 1000 Genomes Project Consortium., Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, **McVean GA,** Abecasis GR (2015). A global reference for human genetic variation. *Nature* **526**: 68-74. PMID: 26432245.

4. Dilthey A, Cox C, Iqbal Z, Nelson MR, **McVean G** (2015). Improved genome inference in the MHC using a population reference graph. *Nat Genet* **47**: 682-8. PMID: 25915597.

5. Venn O, Turner I, Mathieson I, de Groot N, Bontrop R, **McVean G** (2014). Nonhuman genetics. Strong male bias drives germline mutation in chimpanzees. *Science* **344**: 1272-5. PMID: 24926018.