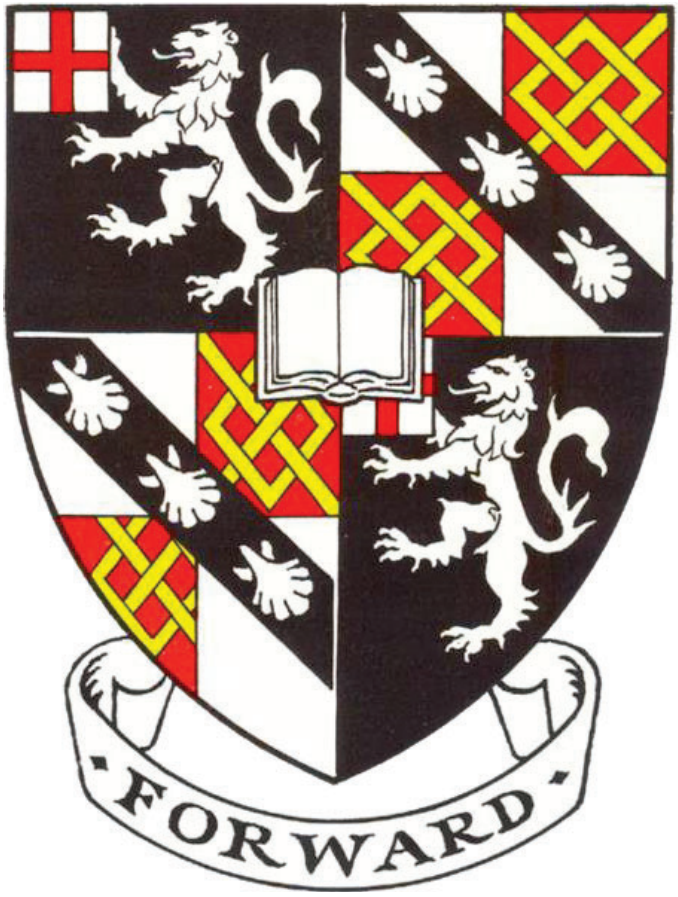
Genetic studies of cardiometabolic traits



Fernando Riveros Mckay Aguilera

Churchill College

University of Cambridge

Wellcome Sanger Institute

September 2018

This dissertation is submitted for the degree of Doctor of Philosophy

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the contributions section within each chapter and/or specified in the text. It is not being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution. It does not exceed the prescribed word limit for the Faculty of Biology.

Fernando Riveros Mckay Aguilera

September 2018

“The world is full of lies. Memory is fuzzy and unreliable. Words we say are often transformed and what ends up in the pages of history is an amalgamation of people’s perception of us through time. But science…man science is cool”

Winston Churchill

Abstract

Diet and lifestyle have changed dramatically in the last few decades, leading to an increase in prevalence of obesity, defined as a body mass index >30Kg/m^2, dyslipidaemias (defined as abnormal lipid profiles) and type 2 diabetes (T2D). Together, these cardiometabolic traits and diseases, have contributed to the increased burden of cardiovascular disease, the leading cause of death in Western societies.

Complex traits and diseases, such as cardiometabolic traits, arise as a result of the interaction between an individual's predisposing genetic makeup and a permissive environment. Since 2005, genome-wide association studies (GWAS) have been successfully applied to complex traits leading to the discovery of thousands of trait-associated variants. Nonetheless, much is still to be understood regarding the genetic architecture of these traits, as well as their underlying biology. This thesis aims to further explore the genetic architecture of cardiometabolic traits by using complementary approaches with greater genetic and phenotype resolution, ranging from studying clinically ascertained extreme phenotypes, deep molecular profiling, or sequence level data.

In chapter 2, I investigated the genetic architecture of healthy human thinness (N=1,471) and contrasted it to that of severe early onset childhood obesity (N=1,456). I demonstrated that healthy human thinness, like severe obesity, is a heritable trait, with a polygenic component. I identified a novel BMI-associated locus at *PKHD1*, and found evidence of association at several loci that had only been discovered using large cohorts with >40,000

individuals demonstrating the power gains in studying clinically ascertained extreme phenotypes.

In chapter 3, I coupled high-resolution nuclear magnetic resonance (NMR) measurements in healthy blood donors, with next-generation sequencing to establish the role of rare coding variation in circulating metabolic biomarker biology. In gene-based analysis, I identified *ACSL1*, *MYCN*, *FBXO36* and *B4GALNT3* as novel gene-trait associations ($P<2.5x10^{-6}$). I also found a novel link between loss-of-function mutations in the "regulation of the pyruvate dehydrogenase (PDH) complex" pathway and intermediate-density lipoprotein (IDL), low-density lipoprotein (LDL) and circulating cholesterol measurements. In addition, I demonstrated that rare "protective" variation in lipoprotein metabolism genes was present in the lower tails of four measurements which are CVD risk factors in this healthy population, demonstrating a role for rare coding variation and the extremes of healthy phenotypes.

In chapter 4, I performed a genome-wide association study of fructosamine, a measurement of total serum protein glycation which is useful to monitor rapid changes in glycaemic levels after treatment, as it reflects average glycaemia over 2-3 weeks. In contrast to HbA1c, which reflects average glucose concentration over the life-span of the erythrocyte (~3 months), fructosamine levels are not predicted to be influenced by factors affecting the erythrocyte. Surprisingly, I found that in this dataset fructosamine had low heritability (2% vs 20% for HbA1c), and was poorly correlated with HbA1c and other glycaemic traits. Despite this, I found two loci previously associated with glycaemic or albumin traits, *G6PC2* and *FCGRT* respectively *($P<5x10^{-8}$),* associated with fructosamine suggesting shared genetic influence.

Altogether my results demonstrate the utility of higher resolution genotype and phenotype data in further elucidating the genetic architecture of a range of cardiometabolic traits, and the power advantages of study designs that focus on individuals at the extremes of phenotype distribution. As large cohorts and national biobanks with sequencing and deep multi-dimensional phenotyping become more prevalent, we will be moving closer to understanding the multiple aetiological mechanisms leading to CVD, and subsequently improve diagnosis and treatment of these conditions.

Acknowledgements

Ok, so from my understanding this is pretty much the free flow section of this work. So yeah, firstly, all of these you're about to read (or skim through) would not be at all possible without the dynamic duo of Inês Barroso and Eleftheria Zeggini. Initially I thought I might even have to transform acknowledgements into its own chapter with a subsection for each because there's so much I want to say (yet so little that realistically would come out). Yet if I could present an abstract for the "acknowledgements chapter" that only exists in my head is this: Inês, thank you for the amazing support, for the constant challenges and helping me feel like this process and piece of work is truly an achievement. Ele, thank you so much for making me feel like part of your team, Volos and everything it represents for my career and life ambitions. Both of you are top notch mentors and role models and the science discussions I've had with both of you in and out of work have truly enriched my experience. It also must be said that I truly appreciate all the emotional support throughout the hardships endured this four years. And on the topic of mentorship and support, I cannot end this paragraph without thanking Eleanor Wheeler who was incredibly patient and supportive as my day to day supervisor and really helped me get started here and was the filter for my dumbest questions so that Inês or Ele wouldn't have to deal with those. Ellie, you rock.

Moving on to other people in my science life, thank you to my thesis committee Nicole Soranzo and Adam Butterworth. I really appreciate your input into my work and for integrating me into some of the research done by you to get an early feel on big collaborative research. Thanks also Carl Anderson for the discussions and his excellent role as head of Grad Programme. Thanks Darren Logan for being so helpful ever since the first interview. Thanks past and present members of team35 and team144 for all the help in multiple stages of my PhD. It is truly a privilege to be surrounded by such a diverse group of smart people with different expertise. At Churchill College, special shout out to Rebecca Sawalmeh who was incredibly important in my college life. Also Rita and Barry for the mentorship dinners.

Now let's get a bit more personal (just a bit). I formed an amazing group of friends here at Sanger, met a bunch of super cool people and these people made life sooooo much easier. I'll start kind of historically. But thanks Ximena and Martin, for being there for me since my first jetlagged interview and help me get settled here and teaching me what it's like to be a Mexican in Cambridge. Sophie, the first friend I made here (and very wisely done), the funniest woman I know and a pillar of emotional support throughout, and the cakes..oh god the cakes. Thank you Neneh, my big sister whose calming and soothing demeanour are incredibly contagious. Patrick Short: the man, the legend. What would life have been like if we hadn't shared that car to work. Thank you to Loukas, my life mentor who has taught me

so much and yet I'm still constantly learning more. Thank you to Arthur, I can't believe I didn't hang out with you more often in the early days; you're like jalapeno to my life adding a much needed spice. My best fitbit friend forever (bfff) Katharina for keeping me fit and all the fitbit walks. Miguel, thanks for the jamming sessions and the active lifestyle guidance. Thanks to all my friends back in Mexico, especially: Yoshi, Yann, Sebas and Palas. Then finally the gang: Dim, Lil, Veli, Alice, GM and Mash. The centre of my social life and the people that made me want to go to work every day. Thanks all of you for being there in my hour of need as well. Thanks GM for trying to get me out of my house constantly when all I want is sulk. Thanks Alice for being there for me when I needed to talk. Thanks Veli for the closeness we've achieved this past year and all the life advice. Dim..D-bone. Dude..duude. Thanks for being there through the best and the worst, literally. Lili, thanks for all the selfless acts of kindness, for the excellent company on my way to Sanger and for calling me a "stupid idiot" when I needed it. And thanks Mash…just coz.

Last but not least, I'd like to thank my family. I've been incredibly lucky to have been born with the parents I have. I'm lucky to have shared most of my life with them and my siblings. I've been really lucky to get the random combination of genetic variants and environment they provided to form a complex human being that is about to get a PhD (maybe?). So yeah…thank you infinitely.

Ok, so that should be it. But also, one thing I must say is that this thesis has a soundtrack. Every chapter was written with an album playing in the background. You don't have to listen to it yourself, but …I mean..if you want to :

- Chapter 1: Gaslight Anthem – Handwritten
- Chapter 2: Sum 41 – Underclass hero
- Chapter 3: PXNDX – Para ti con desprecio
- Chapter 4: Hamilton (the musical)
- Chapter 5: My Chemical Romance – Danger days …

Perfect. So..now..if you read all of this..sorry ..and thanks. Enjoy the rest of the thesis (or bye if you only read this).

P.S: I'm just kidding Mash, you know you're the best friend in the world. I would not have survived Cambridge, especially not the last year without you. You're key to my sanity. Thank you so much for everything.

Publications

From this thesis:

Riveros-Mckay F, Oliver-Williams C, Karthikeyan S, Walter K, Kundu K, et al. Sequencing reveals role of rare variation in circulating metabolic biomarkers. (In preparation)

Riveros-Mckay F*, Mistry V, Bounds R, Hendricks A, Keogh JM, et al. Genetic architecture of human thinness compared to severe obesity. *PLOS Genetics*. (in press)

Arising elsewhere:

Huckins LM*, Hatzikotoulas K*, Southam L, Thornton LM, Steinberg J, **Aguilera-McKay F**, et al.. (2018) Investigation of common, low-frequency and rare genome-wide variation in anorexia nervosa. *Mol Psychiatry*. 2018 May;23(5):1169-1180. doi: 10.1038/mp.2017.88. Epub 2017 Jul 25.

Astle WJ*, Elding H*, Jiang T*, Allen D, Ruklisa D, Mann AL, Mead D, Bouman H, **Riveros-Mckay F**, et al. (2016) The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease.*Cell*. 2016 Nov 17;167(5):1415-1429.e19.doi: 10.1016/j.cell.2016.10.042

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Full name
1000G	1000 Genomes
AC	Allele count
ALSPAC	Avon Longitudinal Study of Parents and Children
AN	Allele number
BMI	Body mass index
CD/CV	Common disease/common variant
CD/RV	Common disease/rare variant
CHD	Coronary heart disease
CI	Confidence intervals
CNV	Copy number variant
CVD	Cardiovascular disease
DZ	Dizygotic
EA	Effect allele
EAF	Effect allele frequency
eQTL	Expression quantitative trait locus
FG	Fasting glucose
FI	Fasting insulin
GG	Glycation gap
GOOS	Genetics Of Obesity Study
GRS	Genetic risk score
GWAS	Genome-wide association study
GxE	Gene-by-environment
H^2	Broad-sense heritability
h^2	Narrow-sense heritability
HbA1c	Glycated haemoglobin
HDL-C	High-density lipoprotein cholesterol
HOMA-B	Beta cell function by homeostasis model assessment
HOMA-IR	Insulin resistance by homeostasis model assessment
HRC	Hapoltype Reference Consortium
HWE	Hardy-Weinberg equilibrium
IMS	Institute of Metabolic Sciences
LD	Linkage disequilibrium
LDL-C	Low-density lipoprotein cholesterol
LoF	Loss-of-function
MAC	Minor allele count
MAF	Minor allele frequency
MI	Myocardial infarction
MR	Mendelian randomisation
MZ	Monozygotic
N	Sample size
NEA	Non-effect allele

NGS	Next-generation sequencing
NMR	Nuclear magnetic resonance
OGTT	Oral glucose tolerance test
OR	Odds ratio
p	P-value
PCA	Principal component analysis
PheWAS	Phenome-wide association study
QC	Quality control
r	Correlation coefficient
r^2	Coefficient of determination
RG	Genetic correlation
SCOOP	Severe Childhood Onset Obesity Project
SE	Standard error
SNP	Single nucleotide polymorphism
STILTS	STudy Into Lean and Thin Subjects
T2D	Type 2 diabetes
TC	Total cholesterol
TG	Triglycerides
UKBB	UK Biobank
UKHLS	UK household longitudinal study
WES	Whole-exome sequencing
WGS	Whole-genome sequencing
WHR	Waist-to-hip ratio
WSI	Wellcome Sanger Institute