

## **The Genetic Epidemiology of Developmental Dysplasia of the Hip: A Genome-Wide Association Study Harnessing National Clinical Audit Data**

Konstantinos Hatzikotoulas<sup>1\*</sup>, Andreas Roposch<sup>2\*</sup> for the †DDH Case Control Consortium, Karan M Shah<sup>3</sup>, Matthew J Clark<sup>3</sup>, Selina Bratherton<sup>3</sup>, Vasanti Limbani<sup>4</sup>, Julia Steinberg<sup>1</sup>, Eleni Zeggini<sup>1</sup>, Kaltuun Warsame<sup>2</sup>, Madhushika Ratnayake<sup>5</sup>, Maria Tselepi<sup>5</sup>, Jeremy Schwartzentruber<sup>1</sup>, John Loughlin<sup>5</sup>, Deborah M Eastwood<sup>4</sup>, Eleftheria Zeggini<sup>1\*</sup>, J Mark Wilkinson<sup>3\*</sup>

<sup>1</sup>Wellcome Trust Sanger Institute, Morgan Building, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1HH, UK. <sup>2</sup>Institute of Child Health, University College London, 30 Guildford Street, London, WC1N 3EH, UK. <sup>3</sup>Department of Oncology and Metabolism, University of Sheffield, Medical School, Beech Hill Road, Sheffield, S10 2RX, UK. <sup>4</sup>Royal National Orthopaedic Hospital, Brockley Hill, Stanmore, Middlesex, HA7 4LP, UK. <sup>5</sup>Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK.

\*These authors contributed equally, †collaborators in the DDH Case Control Consortium are listed in support document and supplementary table S2:

Correspondence to:

Prof J Mark Wilkinson, University of Sheffield, The Medical School, Beech Hill Road, Sheffield, S10 2RX, UK, Phone +44 114 2714705, email [j.m.wilkinson@sheffield.ac.uk](mailto:j.m.wilkinson@sheffield.ac.uk)

Or, Prof Eleftheria Zeggini, Wellcome Trust Sanger Institute, Morgan Building, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1HH, Phone +44 1223 655235, email [eleftheria@sanger.ac.uk](mailto:eleftheria@sanger.ac.uk)

## ABSTRACT

**Background** Developmental dysplasia of the hip (DDH) is a common, heritable condition characterised by abnormal formation of the hip joint, but has a poorly understood genetic architecture due to small sample sizes. We apply a novel case-ascertainment approach using national clinical audit (NCA) data to conduct the largest DDH genome-wide association study (GWAS) to date, and replicate our findings in independent cohorts.

**Methods** We used the English National Joint Registry (NJR) dataset to collect DNA and conducted a GWAS in 770 DDH cases and 3364 controls. We tested the variant most strongly associated with DDH in independent replication cohorts comprising 1129 patients and 4652 controls.

**Results** The heritable component of DDH attributable to common variants was 55% and distributed similarly across autosomal and the X-chromosomes. Variation within the *GDF5* gene promoter was strongly and reproducibly associated with DDH (rs143384, OR 1.44 [95% CI 1.34-1.56],  $p=3.55 \times 10^{-22}$ ). Two further replicating loci showed suggestive association with DDH near *NFIB* (rs4740554, OR 1.30 [95% CI 1.16-1.45],  $p=4.44 \times 10^{-6}$ ) and *LOXLA* (rs4919218, 1.19 [1.10-1.28]  $p=4.38 \times 10^{-6}$ ). Through gene-based enrichment we identify *GDF5*, *UQCC1*, *MMP24*, *RETSAT* and *PDRG1* association with DDH ( $p < 1.2 \times 10^{-7}$ ). Using the UK Biobank and arcOGEN cohorts to generate polygenic risk scores we find that risk alleles for hip osteoarthritis explain  $< 0.5\%$  of the variance in DDH susceptibility.

**Conclusion** Using the NJR as a proof-of-principle, we describe the genetic architecture of DDH and identify several candidate intervention loci and demonstrate a scalable recruitment strategy for genetic studies that is transferrable to other complex diseases.

**Key Words:** Developmental dysplasia of the hip, genomic variation, heritability, GDF5

## Key Messages

- We report the first genome-wide scan for DDH in a European population, and the first to use national clinical audit data for case-ascertainment in complex disease.
- The heritable component of DDH attributable to common variants is 55% and is distributed similarly across autosomal and the X-chromosomes.
- Variation within the *GDF5* gene promoter is strongly and reproducibly associated with DDH, with fine-mapping indicating rs143384 as the likely casual variant.
- Enrichment analyses implicate *GDF5*, *UQCC1*, *MMP24*, *RETSAT* and *PDRG1* as candidate targets for intervention in DDH.
- DDH shares little common genetic aetiology with idiopathic osteoarthritis of the hip, despite sharing variation within the *GDF5* promoter as a common risk factor.

## INTRODUCTION

Developmental dysplasia of the hip (DDH) is characterised by abnormal development of the hip joint and presents with varying severity from mild uncovering of the femoral head to complete dislocation of the joint.<sup>1</sup> It is the most common developmental musculoskeletal anomaly, with population-weighted average incidence that ranges strongly with ethnic background from 0.06 per 1000 live births in Black Africans to 76.1 per 1000 live births in Native Americans,<sup>2</sup> and with an incidence in the UK European population of 3.6 per 1000. DDH is a complex disorder, with known associations including female sex, first-born, breech presentation, and family history.<sup>2</sup> There is a 7-fold increase in the incidence between siblings and a 10-fold increase in the parents of probands compared to the general population,<sup>3</sup> and a concordance rate of 41% between identical twins versus 3% in dizygotic twins.<sup>4</sup>

Whilst DDH is heritable, its genetic architecture remains poorly characterised. Several linkage scans and candidate gene studies have implicated possible associated genetic variants, including in *GDF5*,<sup>5</sup> but to date no replicated loci of genome-wide significance have been identified.<sup>2, 6-9</sup> A recent genome-wide association study (GWAS) of 386 patients and 558 controls in the Han Chinese population suggested an association with variation in *UQCC* (odds ratio (OR) 1.35,  $P=3.63 \times 10^{-6}$ ), a gene adjacent to *GDF5*.<sup>10</sup> Morphological abnormalities of the hip such as DDH are also recognised as risk factors for the development of secondary degenerative change of the hip,<sup>11, 12</sup> and commonly result in hip replacement in adult life.<sup>12, 13</sup> Idiopathic hip osteoarthritis (OA) is very common and has a substantial heritable component,<sup>14</sup> and may share common genetic aetiology with DDH.<sup>2, 7, 15</sup>

Systematic examination of genome-wide variation in DDH in larger sample sizes is necessary to clarify its heritable biology and inform mechanism-driven preventative strategies. However, sample collection for uncommon diseases can be challenging and time-consuming, as attested by the relative paucity of sample sets for genetic studies of DDH. Routinely-collected national clinical audit datasets (NCAs) present an opportunity for efficient case ascertainment in genetic epidemiology. The National Joint Registry for England, Wales, Northern Ireland, and the Isle of Man (NJR, <http://www.njrcentre.org.uk/>) was established in 2003 to collect audit data on all hip and knee replacement surgery in these regions,

for which it has a completeness rate of 97% (<http://www.njrreports.org.uk/Data-Completeness-and-quality>). As at December 31<sup>st</sup> 2014, the dataset held information on 711 765 primary hip replacement procedures, including data on the diagnostic indication for surgery. We used the NJR as a case-ascertainment tool to conduct a nationwide genome-wide association scan to characterise the genetic architecture of DDH. We examine the heritable contribution to DDH, and use fine-mapping and gene-based enrichment approaches to identify likely causal variants and genes. Finally, we use independent OA-datasets to examine for shared genetic aetiology between DDH and hip OA.

## **PATIENTS AND METHODS**

### **NCA-based recruitment strategy**

Between January 2013 and April 2014, 5411 adults (4095 female) from a total NJR database of 711 765 procedures (0.76%) had undergone a hip replacement for the recorded indication of hip dysplasia and were invited to participate in the screening program (figure 1, and supplementary section 1a.). Subsequent screening comprised filtering by self-reported questionnaires for UK European ancestry and a childhood history of idiopathic DDH (supplementary section 1.a and table S1). Following screening, 1091 individuals confirming both UK European ancestry and idiopathic DDH diagnosed in childhood were invited to return by post a saliva sample for DNA extraction and genotyping. Of these, 907 (803 female) returned a sample. The sex distribution of the responding subjects was similar to that of the DDH population in the UK.<sup>2</sup> Radiographic validation of the DDH phenotype was conducted using a plain pelvic radiograph preceding hip replacement retrieved electronically via the National Data Sharing Network (<http://www.image-exchange.co.uk/>) in a convenience subset of 231 (25%) subjects completing saliva return. The age and sex distribution of the validation subset was similar to the 907 that returned a saliva sample. Radiographic analysis confirmed DDH in 224/231 (97%, supplementary section 1.a). Non-DDH cases were excluded from further analysis. Samples from 900 subjects proceeded to DNA extraction and quality control. In 834 samples DNA yield was sufficient to progress to genome-wide genotyping.

The replication cohort comprised 838 children (725 female) from the UK DDH case control consortium recruited at 18 English hospitals (supplementary section 1.a and table S2) and 225 adults (220 female) with DDH recruited at the Royal National Orthopaedic Hospital. In addition, 66 cases (59 female) from the NJR discovery cohort of insufficient DNA yield for genome-wide genotyping were also included in the replication cohort. Subject overlap between the discovery and replication populations, and between the replication cohorts was excluded using individual patient identifiers and by cross-referencing to the NJR dataset. Case duplication was excluded in the discovery cohort at both subject ID and genotype quality control stages.

The control population comprised 8016 participants from the United Kingdom Household Longitudinal Study (UKHLS, <https://www.understandingsociety.ac.uk/about>). A sample of 3364 UKHLS controls was included in the discovery stage and an independent sample of 4652 UKHLS controls was used in the replication stage at a case:control ratio of ~1:4. The proportion of female controls in the discovery stage reflected the distribution in cases. All replication and control participants were of UK European ancestry. The study was approved by the National Research Ethics Service in England (NRES 12/YH/0390, October 30<sup>th</sup> 2012). All subjects provided written informed consent prior to participation.

### **Genotyping, quality control, and association analyses**

DNA from 834 NJR DDH cases and UKHLS controls was genotyped using the Illumina HumanCoreExome BeadChip (Illumina, San Diego, USA; supplementary section 1.b and table S3). Power was calculated by fixing the sample size to our post-quality control (QC) sample size and gave >80% power to detect common variants with moderate effect size at genome-wide significance ( $p=5.0 \times 10^{-8}$ , supplementary section 1.b and figure S1). Following QC checks at both the sample and single nucleotide polymorphism (SNP) level (supplementary section 1.b), 256 867 variants were tested using the likelihood ratio test under an additive genetic model for association with DDH in 770 (693 female) DDH cases and 3364 (3048 female) controls using SNPTEST v2.2.<sup>16</sup>

Independent variants ( $p < 1 \times 10^{-4}$ ) (supplementary section 1.b) in the discovery set were selected for replication. *De novo* genotyping for these SNPs in 1129 (1004 female) replication cases was conducted

using the iPLEX® Assay and the MassARRAY® System (Agena Bioscience, Inc). Case-control association analysis versus 4652 (2527 female) controls was performed under an additive genetic model. Finally, a fixed-effects meta-analysis was conducted across the discovery and replication datasets, comprising a total of 1899 cases and 8016 controls. Conditional single-variant association analyses were used to identify statistical independence. A variant was considered independent of the index SNP if the pre- and post-conditioning p-value difference was lower than two orders of magnitude.

### **Heritability analysis**

We estimated the heritability of DDH using the genome-wide analysis dataset using genetic complex trait analysis (GCTA) and phenotype correlation genotype correlation (PCGC), assuming a DDH prevalence of 3.6 per 1000 live births in the UK,<sup>2</sup> and using a minor allele frequency (MAF) cut-off of 0.01 (supplementary section 1.c).

### **Fine-mapping**

In order to fine-map the replicating association signals, we imputed variant genotypes locally using 1000 Genomes Project<sup>17</sup> combined with UK10K Project<sup>18</sup> data as the reference panel (supplementary section 1.d.). We then applied a method that assigns a relative “probability of regulatory function” (PRF) score among candidate causal variants, reweighting association statistics based on epigenomic annotations, and delineating the 95% probability set of likely causal variants (supplementary section 1.d).

### **Gene-based enrichment analyses**

In order to quantify the degree of association each gene has with the phenotype and understand the potential functional effects of the identified variants, we used the GWAS results to perform gene-based enrichment analyses in MAGMA v1.03 to identify the joint association of all markers in a given gene with the phenotype (supplementary section 1.e).

### **Genetic overlap with hip OA**

Polygenic risks scores were created for all genotyped participants to test for shared genetic aetiology with idiopathic hip OA. Two independent UK cohorts were used for these analyses: 3266 cases with hip OA versus 11 009 controls from the arcOGEN study; and 2396 cases with International Classification of Diseases (ICD) 10 coded hip OA versus 9593 controls from the UK Biobank (supplementary section 1.f). High-resolution polygenic risk scores were made using evenly spaced p-value thresholds (Pt) between 0.001 to 0.5, at increments of 0.0001, to produce 4991 risk thresholds.

### **Data Availability**

Genome-wide genotype data of the DDH cases and UKHLS controls have been deposited to the European Genome-Phenome Archive ([www.ebi.ac.uk/ega/home](http://www.ebi.ac.uk/ega/home), DDH accession numbers: EGAD00010000766 and EGAS00001000916; UKHLS accession numbers: EGAD00010000891 and EGAS00001001232) and of the DDH NJR cases to the NJR data archive ([www.njrcentre.org.uk](http://www.njrcentre.org.uk)).

## **RESULTS**

### **The heritable component of DDH**

Using GCTA we found that common-frequency autosomal SNPs explain 55% ( $\pm se=6\%$ ,  $p<0.0001$ ) of the liability-scale heritability of DDH. The heritability estimate by PCGC was very similar to that identified by GCTA (supplementary table S4). The heritability estimate was also similar ( $54.7\pm 5.8\%$ ) when the analysis was repeated using sex as a covariate.

### **Discovery GWAS**

Genome-wide analysis showed an excess of signals in the discovery case-control population (figure 2). A total of 53 SNPs comprising 25 independent signals showed suggestive evidence for association with DDH at  $p<1\times 10^{-4}$  compared to 3 independent variants expected under the null hypothesis of no association (binomial  $p=9.57\times 10^{-17}$ , supplementary table S5). Eleven correlated variants reached genome-wide significance ( $p<5.0\times 10^{-8}$ ) and all reside in the same region. The lead variant, rs143384



(effect allele A, effect allele frequency (EAF) 0.60, OR[95%CI] 1.57[1.3-1.77]  $p=1.72 \times 10^{-14}$ ), is located in the 5' untranslated region of *GDF5* (20q11.22). Since DDH is sex-biased toward females, we also repeated the analysis using gender as a covariate and found no qualitative difference (supplementary table S6).

## Replication

Replication of the independent signals was conducted in three further DDH populations of UK European ancestry, totalling 1129 cases and 4652 UKHLS controls. The variant rs143384 reached genome-wide significance in the replication dataset (effect allele A, EAF 0.61, OR 1.37[1.24 to 1.51]  $p=1.33 \times 10^{-10}$ ). Four further SNPs showed nominal association ( $p < 0.05$ ) with the same direction of effect as the discovery cohort, compared with 1.35 under the null expectation (binomial test  $p=0.01$ , supplementary table S7). When the replication analysis was performed excluding the 66 NJR subjects that failed genome-wide genotyping QC, the findings were similar (data not shown).

## Meta-analysis

At meta-analysis, 20 SNPs were associated with DDH with the same direction of effect in both the discovery and replication analysis. Three SNPs in the same region of chromosome 20 showed association with DDH at genome-wide significance: rs143384 in *GDF5* (figure 3A, effect allele A, OR 1.44[1.34 to 1.56]  $p=3.55 \times 10^{-22}$ ), rs12479765 in *MMP24* (effect allele G, OR 1.33[1.20 to 1.47]  $p=3.18 \times 10^{-08}$ ), and rs2050729 in *RMB29* (effect allele G, OR 1.41[1.25 to 1.58]  $p=1.15 \times 10^{-08}$ ). Conditional analysis of rs12479765 and rs2050729 on the lead variant rs143384 attenuated their association with DDH, indicating that they were conditionally dependent upon the lead variant. rs143384 explained 0.96% of DDH variance on the liability scale ( $h^2_L$ ) and the sibling relative risk ratio ( $\lambda_s$ ) attributable to this variant was  $\lambda_s=1.04$  (supplementary figure S2).

Two further independent signals showed evidence of replicating association with DDH, although this did not reach genome-wide significance at meta-analysis: rs4740554 (effect allele C, EAF 0.10, OR 1.30[1.16 to 1.45]  $p=4.44 \times 10^{-6}$ ) and rs4919218 (effect allele C, EAF 0.31, OR 1.19[1.10 to 1.28]

$p=4.38 \times 10^{-6}$ ) (figure 3B and 3C, and Table 1). rs4740554 and rs4919218 explained 0.36% and 0.24% of phenotypic variance ( $\lambda_s=1.02$  and 1.01), respectively (supplementary figure S2).

### **Fine-mapping**

For the *GDF5* signal, five variants were included in the 95% probability set of being causal. Among these, both rs143384 and rs143383 have high epigenomic scores as they are located in the 5' UTR of *GDF5*, have high nucleotide sequence conservation, and overlap DNase hypersensitivity and activating histone marks. Considering annotations and association statistics p-values together, rs143384 had a >99% likelihood of being the causal variant, assuming only 1 variant as causal.

### **Gene-based enrichment analyses**

After adjusting for the family-wise error rate (FWER), we found five genes that were associated with DDH susceptibility: *GDF5*, *UQCC1*, *MMP24*, *RETSAT* and *PDRG1* ( $p=9.24 \times 10^{-12}$ ,  $p=1.86 \times 10^{-10}$ ,  $p=3.18 \times 10^{-9}$ ,  $p=3.70 \times 10^{-8}$  and  $p=1.06 \times 10^{-7}$ , respectively; Supplementary table S8). To exclude association through *GDF5*, we repeated the analysis by removing *GDF5* and all genes remained associated with DDH.

### **Shared genetic aetiology with hip OA**

We found little evidence for shared genetic aetiology between DDH and hip OA. The best fit polygenic risk score for hip OA in the arcOGEN and ICD-10 coded hip OA in the UK Biobank datasets explained ~0.05% and 0.35%, respectively, of the variation in DDH susceptibility (FWER  $P>0.05$ , supplementary figure S3).

## **DISCUSSION**

We used case-ascertainment by NCA database search and postal recruitment to facilitate the largest genome-wide association scan for DDH to date. We found that the heritable component of DDH due to common autosomal variants is approximately 55%, consistent with the complex nature of the disease.

We establish variation within *GDF5* on chromosome 20 as robustly associated with DDH susceptibility, with the variant rs143384 as the causal signal by fine-mapping, although *GDF5* variation makes only a small contribution to overall DDH heritability. Through gene-based enrichment analyses we identify *GDF5*, *UQCC1*, *MMP24*, *RETSAT* and *PDRG1*, as independently associated with DDH susceptibility.

Previous studies examining the epidemiology of DDH within families using epidemiological linkage approaches have demonstrated that heritable factors contribute between 50% and 85% of the total liability of disease,<sup>3, 19, 20</sup> depending on population ethnicity. Here, we used dense genome-wide genotyping data to estimate the heritable risk of DDH amongst unrelated individuals of UK European ancestry in the general English population. Our estimate of DDH heritability attributable to variation within autosomal chromosomes was broadly similar to the estimates derived from linkage analyses.<sup>2</sup> When the heritability analysis was stratified by chromosome, we found that the X-chromosome contributed a similar amount to overall DDH heritability as the autosomes (Supplementary Fig. 4), and we identified no individual chromosome X signals. Taken together, these data suggest that although DDH is strongly sex-linked, this is not substantially due to variation in genomic DNA.

The gene *GDF5* encodes growth differentiation factor 5 (GDF5), belonging to the transforming growth factor beta superfamily.<sup>21</sup> GDF5 is required for normal bone and joint development by promoting cartilage condensation and increasing the size of the skeletal elements through proliferation within epiphyseal cartilage.<sup>22</sup> Two *GDF5* variants, rs143383 and 143384, have previously shown suggestive association with DDH in unreplicated candidate gene studies in Asians and in Europeans.<sup>5, 23</sup> Here we validate and extend these findings at the genome-wide level and with robust replication. These variants are in high linkage disequilibrium ( $r^2=0.82$ ,  $D'=0.99$ ), and were both directly genotyped in this study. Our data implicate rs143384 as the lead variant, with the rs143383 association with DDH being two orders of magnitude weaker. Subsequent fine-mapping of the *GDF5* locus using Roadmap Epigenome annotations with imputation using 1000 Genomes and UK10K Project data also identified rs143384 as the likely causal variant. Chen et al,<sup>24</sup> have shown that the *GDF5* locus contains many separate regulatory elements that control expression of the gene at different joint sites, and that these flanking regions are large. Interrogation of the GTEx database indicates that rs143384 is an expression

quantitative trait locus (eQTLs) for multiple genes across various tissues (supplementary table 9).<sup>25, 26</sup>

This annotated function provides novel opportunities for investigation of the role of *GDF5* as a candidate target for DDH prevention.

Gene-based enrichment analysis can boost power to identify contributing loci by combining information across multiple SNPs co-localised at the gene level. Here, we found associations with the *GDF5*, *UQCCL1*, *MMP24*, *RETSAT* and *PDRG1* genes. *UQCCL1* gene encodes a trans-membrane protein ubiquinol-cytochrome c reductase complex chaperone, which is structurally similar to the mouse basic fibroblast growth factor repressed ZIC-binding protein.<sup>27</sup> In humans, polymorphisms in this gene have been associated with DDH,<sup>10</sup> bone size,<sup>28</sup> height,<sup>29</sup> and hip axis length.<sup>30</sup> *MMP24* encodes a member of the peptidase M10 family of matrix metalloproteinases (MMPs) involved in the breakdown of extracellular matrix in normal physiological processes, such as embryonic development and tissue remodelling.<sup>31</sup> Polymorphisms within *MMP24* have also been associated with height variation in childhood.<sup>32, 33</sup> *RETSAT* codes for retinol saturase, an enzyme centrally involved in the metabolism of vitamin A.<sup>34</sup> Retinoic acid signalling is essential for normal limb bud development, including bone and cartilage formation,<sup>35</sup> and provides a further candidate target in DDH pathogenesis.

This study has limitations. The patients recruited in the discovery cohort were adults who had undergone a hip replacement for progressive symptoms related to an underlying diagnosis of hip dysplasia, as identified through the NJR. To mitigate against misclassification bias in case-ascertainment we employed a stepwise filtration approach that subsequently validated with high concordance against examination of plain radiographic images. The female-to-male case ratio in our final NJR DDH cohort was similar to that expected in the general DDH population in the UK,<sup>2</sup> and contrasts with that found in idiopathic OA of the hip.<sup>36</sup> Previous studies have also demonstrated the contribution of DDH as a risk factor for secondary OA,<sup>11, 12, 37</sup> and that OA susceptibility genes influence the association between hip morphology and OA.<sup>15, 38, 39</sup> We therefore examined whether we were simply detecting OA-associated loci in our DDH discovery cohort, as case identification required NJR registration for a hip replacement. Polygenic risk scores derived using independent hip OA cohorts predicted only a small amount of the genetic variability in the DDH discovery population, suggesting only limited common

genetic aetiology between the diseases. Finally, our replication cohorts mainly comprised independent populations of children recruited prospectively with a firm diagnosis of DDH, and in whom the odds-ratio of association for the primary variant signal was similar to that found in the NJR-discovery sample.

In summary, the results of this largest study of DDH to date provide a comprehensive picture of its complex genetic architecture, the contribution of common variants to its heritability, and establishes the first robust DDH genetic locus. We also demonstrate the utility of national clinical audit-based case ascertainment and recruitment for genetics and genomics studies using a strategy that is transferrable to other complex diseases in which large-scale datasets are held.

## **Funding**

This work was supported by the Healthcare Quality Improvement Partnership, the National Institute for Health Research, The Wellcome Trust (098051), Medical Research Council and Arthritis Research UK as part of the MRC-Arthritis Research UK Centre for Integrated Research into Musculoskeletal Ageing (CIMA), The European Union's Seventh Framework Program for research, technological development and demonstration (D-BOARD grant #305815), The Great Ormond Street Hospital National Institute for Health Research Biomedical Research Centre, and The Special Trustees of the Royal National Orthopaedic Hospital. The UK Household Longitudinal Study is led by the Institute for Social and Economic Research at the University of Essex and funded by the Economic and Social Research Council. The survey was conducted by NatCen and the genome-wide scan data were analysed and deposited by the Wellcome Trust Sanger Institute. Information on how to access the data can be found on the Understanding Society website <https://www.understandingsociety.ac.uk/>.

## **Acknowledgments**

We thank the patients and staff of all the hospitals who have contributed data to the National Joint Registry. We are grateful to the Healthcare Quality Improvement Partnership (HQIP), the National Joint

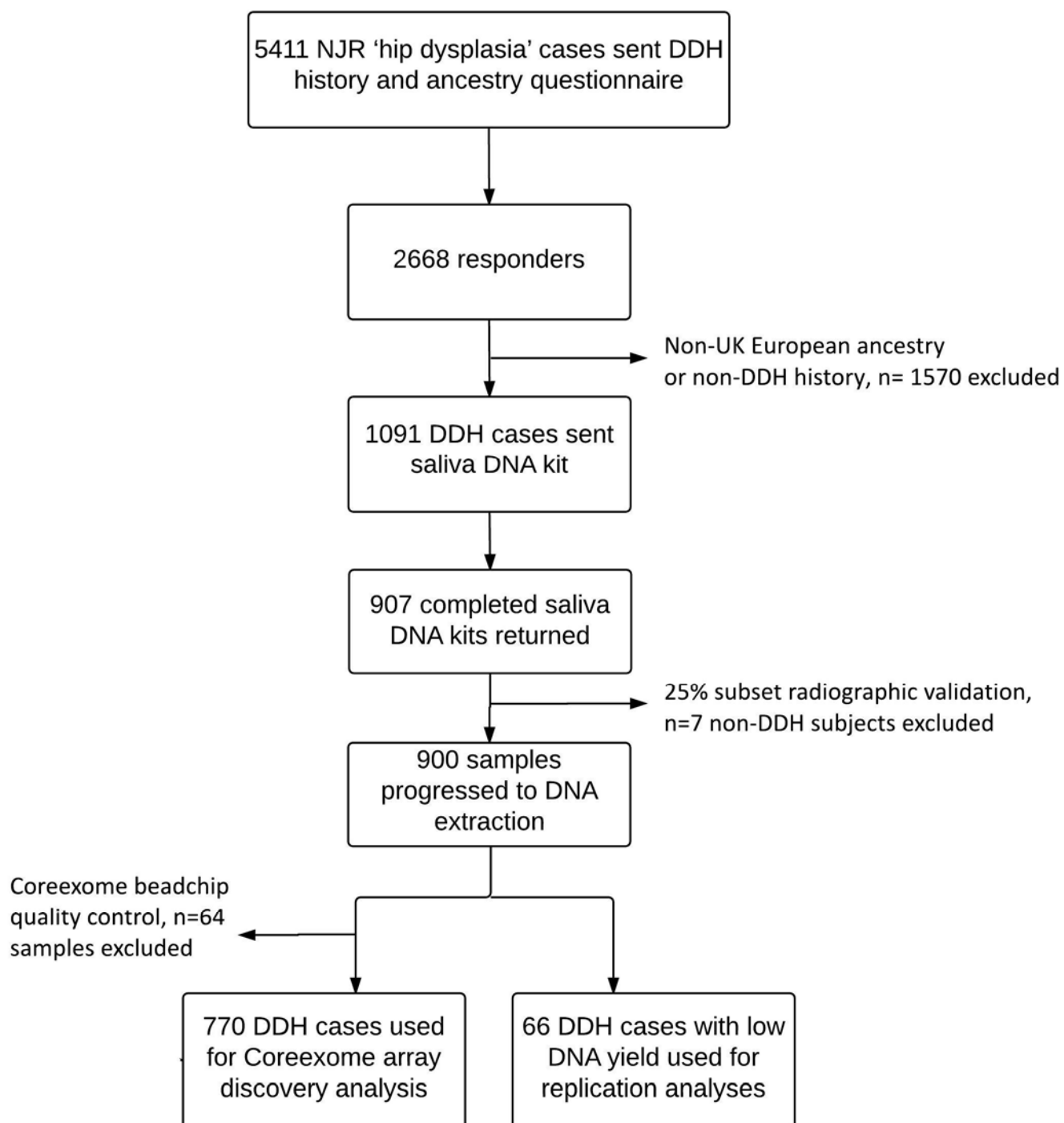
Registry Steering Committee (NJRSC), and staff at the NJR Centre for facilitating this work. The views expressed represent those of the authors and do not necessarily reflect those of the NJRSC or HQIP who do not vouch for how the information is presented. This research has been conducted using the UK Biobank Resource under Application Number 9979. The study also utilised summary genotype data from arcOGEN (<http://www.arcogen.org.uk/>) funded by a special purpose grant from Arthritis Research UK (grant 18030).

## REFERENCES

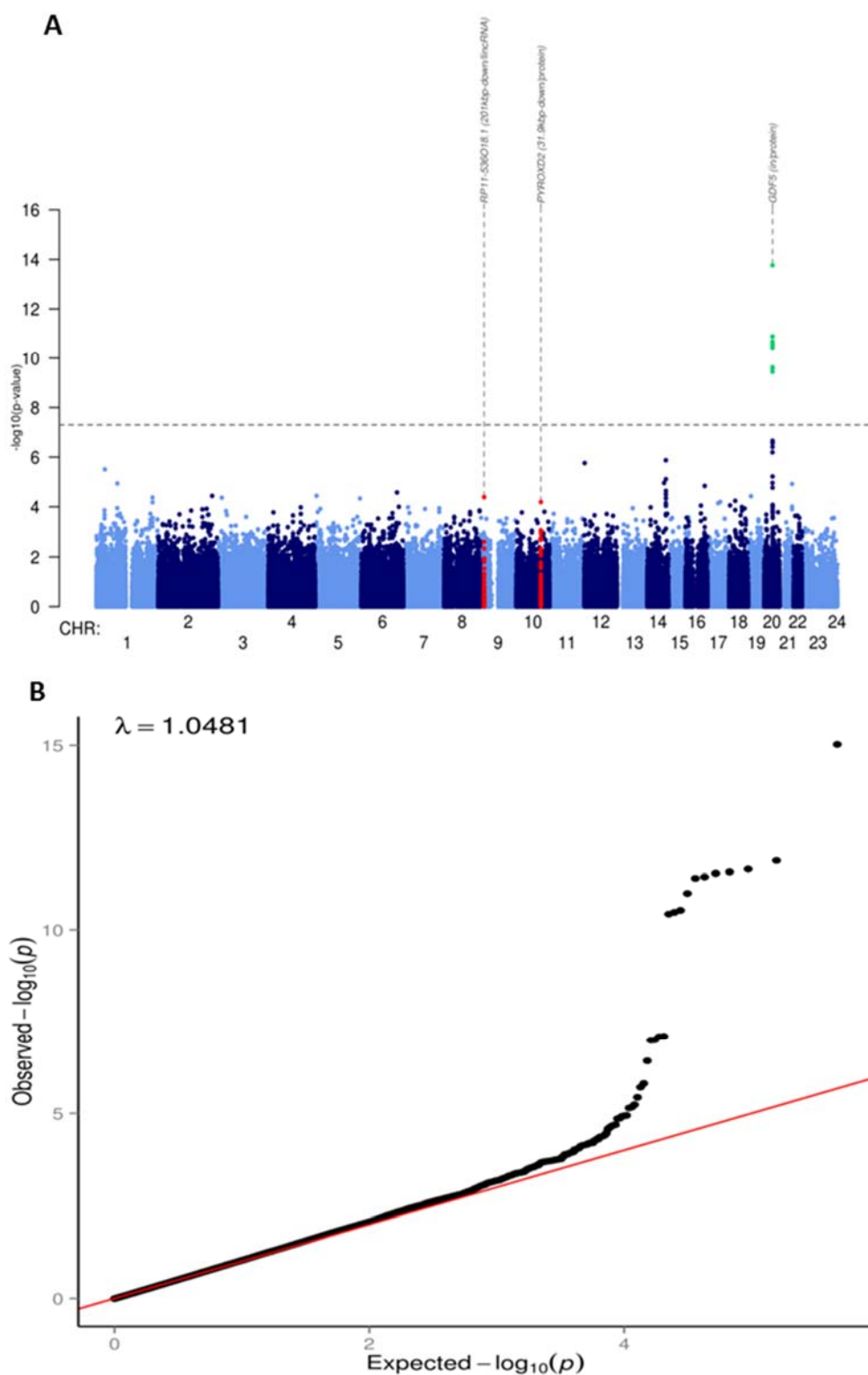
1. Harcke HT. Developmental dysplasia of the hip: a spectrum of abnormality. *Pediatrics* 1999; **103**: 152.
2. Loder RT, Skopelja EN. The epidemiology and demographics of hip dysplasia. *ISRN Orthop* 2011; **2011**: 238607.
3. Bjerkreim I, van der Hagen CB. Congenital dislocation of the hip joint in Norway. V. Evaluation of genetic and environmental factors. *Clin Genet* 1974; **5**: 433-48.
4. Idelberger KH. [Orthopedic genetics and family counseling (proceedings)]. *ZOrthop Ihre Grenzgeb* 1978; **116**: 552-4.
5. Rouault K, Scotet V, Autret S, et al. Evidence of association between GDF5 polymorphisms and congenital dislocation of the hip in a Caucasian population. *Osteoarthritis Cartilage* 2010; **18**: 1144-9.
6. Cilliers HJ, Beighton P. Beukes familial hip dysplasia: an autosomal dominant entity. *Am J Med Genet* 1990; **36**: 386-90.
7. Mabuchi A, Nakamura S, Takatori Y, Ikegawa S. Familial osteoarthritis of the hip joint associated with acetabular dysplasia maps to chromosome 13q. *Am J Hum Genet* 2006; **79**: 163-8.
8. Li LY, Sun XK, Zhao Q, et al. [Gene mapping of developmental dysplasia of the hip in chromosome 17q21 region]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 2010; **27**: 620-5.
9. Feldman G, Dalsey C, Fertala K, et al. The Otto Aufranc Award: Identification of a 4 Mb region on chromosome 17q21 linked to developmental dysplasia of the hip in one 18-member, multigeneration family. *Clin Orthop Relat Res* 2010; **468**: 337-44.
10. Sun Y, Wang C, Hao Z, et al. A common variant of ubiquinol-cytochrome c reductase complex is associated with DDH. *PLoS one* 2015; **10**: e0120212.
11. Murray RO. The aetiology of primary osteoarthritis of the hip. *Br J Radiol* 1965; **38**: 810-24.
12. Solomon L. Patterns of osteoarthritis of the hip. *Journal of Bone & Joint Surgery - British Volume* 1976; **58**: 176-83.
13. Murphy SB, Ganz R, Muller ME. The prognosis in untreated dysplasia of the hip. A study of radiographic factors that predict the outcome. *J Bone Joint Surg Am* 1995; **77**: 985-9.
14. Spector TD, MacGregor AJ. Risk factors for osteoarthritis: genetics. *Osteoarthritis Cartilage* 2004; **12 Suppl A**: S39-44.
15. Pollard TC, Batra RN, Judge A, et al. The hereditary predisposition to hip osteoarthritis and its association with abnormal joint morphology. *Osteoarthritis Cartilage* 2013; **21**: 314-21.
16. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007; **39**: 906-13.
17. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; **491**: 56-65.
18. Consortium UK, Walter K, Min JL, et al. The UK10K project identifies rare variants in health and disease. *Nature* 2015; **526**: 82-90.
19. Atasu M, Akkoyunlu U, Tokgozoglu N, Say B. The heritability of liability to congenital dislocation of the hip. *Turk J Pediatr* 1972; **14**: 23-6.
20. Li L, Sun K, Zhang L, Zhao Q, Cheng X, Dang Y. Heritability and sibling recurrent risk of developmental dysplasia of the hip in Chinese population. *Eur J Clin Invest* 2013; **43**: 589-94.
21. Hotten G, Neidhardt H, Jacobowsky B, Pohl J. Cloning and expression of recombinant human growth/differentiation factor 5. *Biochem Biophys Res Commun* 1994; **204**: 646-52.
22. Francis-West PH, Abdelfattah A, Chen P, et al. Mechanisms of GDF-5 action during skeletal development. *Development* 1999; **126**: 1305-15.
23. Dai J, Shi D, Zhu P, et al. Association of a single nucleotide polymorphism in growth differentiate factor 5 with congenital dysplasia of the hip: a case-control study. *Arthritis Res Ther* 2008; **10**: R126.

24. Chen H, Capellini TD, Schoor M, Mortlock DP, Reddi AH, Kingsley DM. Heads, Shoulders, Elbows, Knees, and Toes: Modular Gdf5 Enhancers Control Different Joints in the Vertebrate Skeleton. *PLoS Genet* 2016; **12**: e1006454.
25. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015; **348**: 648-60.
26. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013; **45**: 580-5.
27. Imabayashi H, Mori T, Gojo S, et al. Redifferentiation of dedifferentiated chondrocytes and chondrogenesis of human bone marrow stromal cells via chondrosphere formation with expression profiling by large-scale cDNA analysis. *Exp Cell Res* 2003; **288**: 35-50.
28. Deng FY, Dong SS, Xu XH, et al. Genome-wide association study identified UQCC locus for spine bone size in humans. *Bone* 2013; **53**: 129-33.
29. Sanna S, Jackson AU, Nagaraja R, et al. Common variants in the GDF5-UQCC region are associated with variation in human height. *Nat Genet* 2008; **40**: 198-203.
30. Soranzo N, Rivadeneira F, Chinappan-Horsley U, et al. Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size. *PLoS Genet* 2009; **5**: e1000445.
31. Borkakoti N. Matrix metalloproteases: variations on a theme. *Prog Biophys Mol Biol* 1998; **70**: 73-94.
32. Fontenele EG, Moraes ME, d'Alva CB, et al. Association Study of GWAS-Derived Loci with Height in Brazilian Children: Importance of MAP3K3, MMP24 and IGF1R Polymorphisms for Height Variation. *Horm Res Paediatr* 2015; **84**: 248-53.
33. Zhao J, Li M, Bradfield JP, et al. The role of height-associated loci identified in genome wide association studies in the determination of pediatric stature. *BMC Med Genet* 2010; **11**: 96.
34. Moise AR, Kuksa V, Imanishi Y, Palczewski K. Identification of all-trans-retinol:all-trans-13,14-dihydroretinol saturase. *J Biol Chem* 2004; **279**: 50230-42.
35. Ross SA, McCaffery PJ, Drager UC, De Luca LM. Retinoids in embryonal development. *Physiol Rev* 2000; **80**: 1021-54.
36. Felson DT. Epidemiology of hip and knee osteoarthritis. *Epidemiol Rev* 1988; **10**: 1-28.
37. McWilliams DF, Doherty SA, Jenkins WD, et al. Mild acetabular dysplasia and risk of osteoarthritis of the hip: a case-control study. *Ann Rheum Dis* 2010; **69**: 1774-8.
38. Waarsing JH, Kloppenburg M, Slagboom PE, et al. Osteoarthritis susceptibility genes influence the association between hip morphology and osteoarthritis. *Arthritis Rheum* 2011; **63**: 1349-54.
39. Agricola R, Leyland KM, Bierma-Zeinstra SM, et al. Validation of statistical shape modelling to predict hip osteoarthritis in females: data from two prospective cohort studies (Cohort Hip and Cohort Knee and Chingford). *Rheumatology (Oxford)* 2015; **54**: 2033-41.



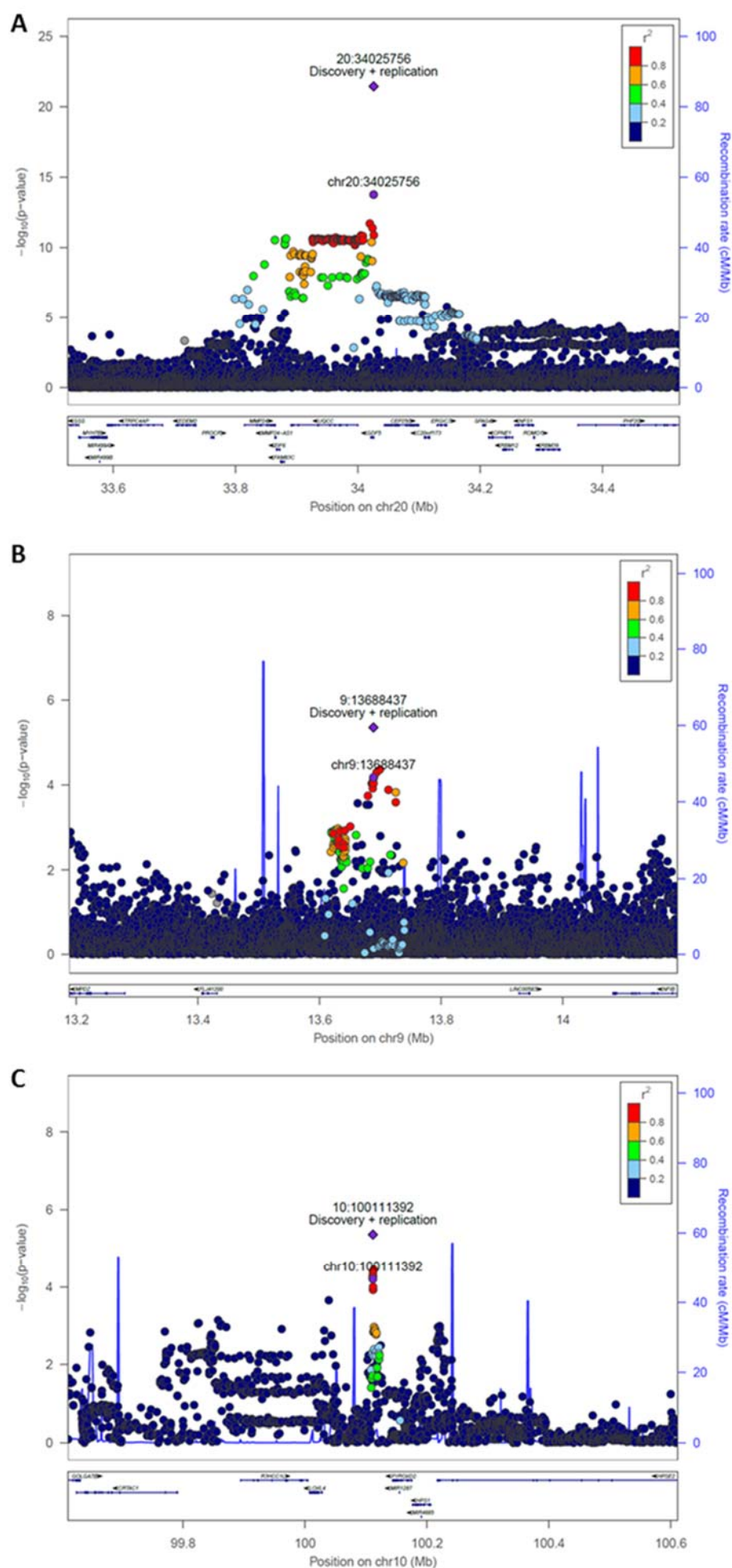


**Figure 1: Recruitment flowchart**



**Figure 2: Manhattan and quantile-quantile plot of the DDH genome-wide association scan.**

Panel A is a Manhattan plot showing the  $-\log_{10} p$  values for each variant (y axis) plotted against their respective chromosomal position (x axis). The dashed line indicates the genome-wide significance threshold ( $p=5.0\times 10^{-8}$ ). Green dots represent variants for which p values reached the genome-wide significance threshold where red dots illustrate the other two replicating signals. Panel B is a quantile-quantile plot. The x-axis indicates the expected  $-\log_{10} p$  values and the y-axis the observed ones. The red line represents the null hypothesis of no association at any locus.



**Figure 3: Region association plots of the three replicating signals.** Regional association plot of rs143384 (20:34025756, panel A), rs4740554 (9:13688437, panel B) and rs4919218 (10:100111392, panel C). Each filled circle represents the p value of analysed variants (as  $-\log_{10}$  values) plotted against their physical position (NCBI Build 37). The p value at the discovery stage and combined discovery and replication cohorts is represented by a purple circle and diamond, respectively. The rest variants in the region are coloured depending on their degree of correlation ( $r^2$ ) with the index variant according to a scale from  $r^2=0$ (blue) to  $r^2=1$ (red). Underneath, gene annotations from the UCSC genome browser.

<b>SNP ID (chromosome:position)</b>	<b>Effect Allele/Other Allele</b>	<b>Effect Allele Frequency (discovery stage)</b>	<b>Discovery Odds Ratio (95% CI)</b>	<b>Discovery p value</b>	<b>Replication Odds Ratio (95% CI)</b>	<b>Replication p value</b>	<b>Meta- analysis Odds Ratio (95% CI)</b>	<b>Meta- analysis p value</b>
rs143384 (20:34025756)	A/G	0.60	1.57 (1.4-1.77)	1.72E-14	1.37 (1.24-1.51)	1.33E-10	1.44 (1.34-1.56)	3.55E-22
rs4740554 (9:13688437)	C/A	0.10	1.44 (1.22-1.71)	4.09E-05	1.20 (1.03-1.39)	0.01	1.30 (1.16-1.45)	4.44E-6
rs4919218 (10:100111392)	C/T	0.69	1.27 (1.13-1.43)	6.48E-05	1.14 (1.03-1.25)	0.0089	1.19 (1.10-1.28)	4.38E-6

**Table 1: Replicating variants associated with DDH.**