


ORIGINAL ARTICLE

Segregation analysis identifies specific alpha-defensin (*DEFA1A3*) SNP–CNV haplotypes in predisposition to IgA nephropathy

Nzar A. A. Shwan^{1,2} | Eric C. Moise¹ | Paula E. Necsoiu¹ | Amy J. Farr¹ | Daniel P. Gale³  | Jonathan Barratt⁴ | John A. L. Armour¹

¹School of Life Sciences, University of Nottingham, Nottingham, UK

²Erbil Technical Health and Medical College, Erbil Polytechnic University, Erbil, Kurdistan Region, Iraq

³Department of Renal Medicine, Division of Medicine, University College London, London, UK

⁴Department of Cardiovascular Medicine, University of Leicester, Leicester, UK

Correspondence

John A.L. Armour, School of Life Sciences, University of Nottingham, Queen's Medical Centre Nottingham NG7 2UH UK.

Email: john.armour1@nottingham.ac.uk

Abstract

Background: Immunoglobulin A (IgA) nephropathy is a disorder of the immune system affecting kidney function, and genome-wide association studies (GWAS) have defined numerous loci with associated variation, all implicating components of innate or adaptive immunity. Among these, single nucleotide polymorphisms (SNPs) in a region including the multiallelic copy number variation (CNV) of *DEFA1A3* are associated with IgA nephropathy in both European and Asian populations. At present, the precise factors underlying the observed associations at *DEFA1A3* have not been defined, although the key alleles differ between Asian and European populations, and multiple independent factors may be involved even within a single population.

Methods: In this study, we measured *DEFA1A3* copy number in UK family trios with an offspring affected by IgA nephropathy, used the population distributions of joint SNP–CNV haplotypes to infer the likely segregation in trios, and applied transmission disequilibrium tests (TDT) to examine joint SNP–CNV haplotypes for over- or undertransmission into affected offspring from heterozygous parents.

Results and Conclusions: We observed overtransmission of 3-copy class 2 haplotypes (raw $p = 0.029$) and some evidence for under-transmission of 3-copy class 1 haplotypes (raw $p = 0.051$), although these apparent effects were not statistically significant after correction for testing of multiple haplotypes.

KEYWORDS

defensin, genetic association, haplotype, Immunoglobulin A nephropathy

1 | INTRODUCTION

Immunoglobulin A (IgA) nephropathy results from the generation of immune complexes between galactose-

deficient IgA1 and abnormal anti-glycan antibodies, leading to mesangial inflammation and damage, and ultimately resulting in renal failure. Numerous lines of evidence indicate strong genetic factors in predisposition

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Annals of Human Genetics* published by University College London (UCL) and John Wiley & Sons Ltd.

to IgA nephropathy, including a characteristically high frequency of the disorder specifically in East Asian populations, and the identification of associated variation at specific loci using genome-wide association studies (GWAS) (Kiryluk et al., 2010; Sanchez-Rodriguez et al., 2021; Xie et al., 2013).

The genetic factors predisposing to IgA nephropathy discovered via GWAS are connected to a range of functions in adaptive and innate immunity, including *HLA* alleles (Feehally et al., 2010; Gharavi et al., 2011; Kiryluk et al., 2014; Yu et al., 2012), complement factors (Gharavi et al., 2011; Xie et al., 2016), and components of innate immunity such as *CARD9* (Kiryluk et al., 2014), *TNFSF13* (Kiryluk et al., 2014; Yu et al., 2012) and the copy-variable alpha-defensin locus *DEFBIA3* (Ai et al., 2016; Kiryluk et al., 2014; Qi et al., 2015; Yu et al., 2012) encoding the neutrophil alpha-defensin peptides HNP1-3. The alpha-defensins encoded by *DEFBIA3* are short (29–30 amino acids) peptides that are abundant in human neutrophil granules, and which in addition to well-established microbicidal properties appear to have key signaling roles connecting the innate and adaptive immune systems (Ganz, 2003; Miles et al., 2009).

Although IgA nephropathy was the first example of a human phenotype significantly associated with *DEFBIA3* variation, single nucleotide polymorphisms (SNPs) neighboring *DEFBIA3* have also been shown to have significant associations with periodontitis (Munz et al., 2017) and with measured properties of human blood cells, including differential white blood cell counts and neutrophil granularity (Akbari et al., 2020; Astle et al., 2016). There are indications that the key allelic factors may overlap between the three phenotypes or may possibly be identical.

Genome-wide significant association of SNP alleles at *DEFBIA3* with IgA nephropathy has been observed in both East Asian (Ai et al., 2016; Kiryluk et al., 2014; Qi et al., 2015; Yu et al., 2012) and European (Kiryluk et al., 2014) populations, but the precise allelic factors involved appear to differ between Asian and European populations (Kiryluk et al., 2014). Furthermore, although *DEFBIA3* and the neighboring *DEFBIA4* are the only protein-coding genes in the associated linkage disequilibrium block, it is not clear in any population what the crucial functional difference is between risk alleles and others, or indeed whether there could be multiple alleles carrying different levels of risk or protection.

Analysis of *DEFBIA3* copy number and sequence variants in Chinese populations demonstrated that copy numbers of *DEFBIA3* and its sequence variants were more closely associated with IgA nephropathy than the lead single-copy SNPs (Ai et al., 2016), suggesting that the underlying causal variation was likely to be in the copy number variation (CNV) region rather than in single-copy

flanking DNA. Gene copy number at *DEFBIA3* can vary extensively, from a minimum of three copies per person up to more than 20 in some populations, and the contributions of the *DEFBIA1* and *DEFBIA3* gene variants to the total gene complement can also be highly variable (Black et al., 2014; Hughes et al., 2020; Khan et al., 2013). In other examples of multiallelic structural variation associated with human phenotypes, such as the association of complement C4 with schizophrenia (Sekar et al., 2016) or of haptoglobin with cholesterol levels (Boettger et al., 2016), different structural variants have different effects, in patterns not simply associated with individual flanking SNPs or proportional to gene dosage. In a detailed analysis of SNP-based association with IgA nephropathy at the *DEFBIA3* locus in Han Chinese, at least three independent signals were defined, with indications that further independent effects remained to be discovered (Li et al., 2015). Risk factors associated with the copy-variable *DEFBIA3* gene region may therefore show graded multiallelic effects, and the relationship between gene copy number and alpha-defensin function is not clear; although early work suggested a simple proportionality between gene number and peptide levels (Linzmeier & Ganz, 2005), subsequent analysis of leukocyte mRNA levels failed to show a correlation with gene number (Aldred et al., 2005).

To address these challenges, we attempted to identify key allelic factors at *DEFBIA3* in predisposition to IgA nephropathy in a European (UK) population using the combined linkage and association approach of the transmission disequilibrium test (TDT). In this work, we examined *DEFBIA3* SNP and CNV variation in DNA from family trios assembled at the UK MRC glomerulonephritis DNA Bank collection, each trio composed of an IgA nephropathy proband and both their parents.

2 | METHODS

2.1 | *DEFBIA3* copy number measurement by PRT

Paralogue ratio test (PRT) methods (Armour et al., 2007; Carpenter et al., 2015; Shwan et al., 2017; Walker et al., 2009) were used to determine gene copy numbers for *DEFBIA3*, including the determination of the relative contribution of sequence variants within the gene repeats, multiplexing PCR products from different assays for the same individual in a single electrophoresis run. The details of experimental methods followed those described (Ai et al., 2016; Khan et al., 2013), except that two new improved PRT measurements were adopted to measure total copy number for the *DEFBIA3* genes. The cen1PRT assay used primers DEFACen1F

(CCCAGAGAGCTCCTTCATT) and DEFAcen1R (TCCTA-GAAAGCTGGTTGCTC) to amplify a 444 bp product from the centromeric gene position, and a 332 bp product from other (noncentromeric) copies; the tel2PRT assay used primers DEFAtel2F (AGAGCAGCCGTGCACAAAC) and DEFAtel2R (GCATCTYGGGGTCCATTGT) to amplify a 260 bp product from the telomeric gene position and a 263 bp product from other copies. Measurements of *DEFA1A3* copy number deduced from cen1 and tel2 PRT data, and the splits of other sequence variants, were used to infer the most likely copy number for each sample as described (Khan et al., 2013).

2.2 | DNA samples and SNP data

DNA samples from IgA nephropathy families and controls were obtained from the MRC glomerulonephritis DNA Bank held at the CIMR Biobank (University of Manchester). Individual samples were genotyped for 318,127 autosomal and X-chromosomal SNPs with the Illumina Sentrix HumanHap300 BeadChip as previously reported (Feehally et al., 2010). For an additional 47 family trio samples not genotyped by those methods, SNP genotypes flanking *DEFA1A3* were determined using the PCR-RFLP assays described (Black et al., 2014; Khan et al., 2013).

2.3 | Common European SNP haplotypes

Previous work on variation in European population samples (Black et al., 2014; Khan et al., 2013) used sequencing of a 4.1 kb region immediately adjacent to the centromeric boundary of the repeat array to define five common haplotypes. The haplotype represented on the GRCh37/hg19 reference assembly was designated the “Reference” haplotype, and two other common variant haplotypes as Class 1 and Class 2. Sequencing also demonstrated that two haplotypes included segmental exchanges at the border of the repeat region, and these were therefore termed Exchange 1 and Exchange 2. Although the repeat exchanges are closely adjacent to the SNPs that define the major European haplotypes, the key SNPs are not themselves included in the replacement events. Each of the common haplotypes in European populations are associated with distinctive numbers and sequence variants of the repeat unit; for example, the Class 2(C2) haplotype is nearly always associated with CNV alleles containing two or three repeat units, of which the most telomeric repeat contains the *DEFA3* gene (Black et al., 2014). The composition and properties of these common European haplotypes are summarized in Table S1.

2.4 | Joint phasing of SNP–CNV haplotypes

SNP and *DEFA1A3* CNV data from 989 Europeans (combining data from (Khan et al., 2013) and the European population samples in the 1000 Genomes Project phase 3 (The 1000 Genomes Project Consortium, 2015)) were phased using MOCSphaser (Kato et al., 2008), which infers the likely joint SNP–CNV haplotypes underlying the observed variation. MOCSphaser analysis produced haplotype frequency (*.hpfq) outputs giving a frequency distribution of different *DEFA1A3* copy number states on each of the five common SNP haplotype backgrounds (Table S2). We then used this information about the common SNP–CNV haplotypes in European populations to infer the likely haplotype segregation in family trios. For each trio, a custom R script iterated across all possible haplotype segregation patterns consistent with the observed SNP and CNV profiles, and the frequencies of the implicated SNP–CNV haplotypes were used to assign relative probabilities to each of the possible solutions. To reduce the search space without appreciable loss of accuracy, we assumed that the minimum *DEFA1A3* copy number per haplotype was 2, in keeping with observations from large-scale typing (Black et al., 2014; Hughes et al., 2020; Khan et al., 2013), in which 1-copy haplotypes are rare; 0-copy haplotypes have not been observed. As a general index of confidence in the most likely segregation pattern, we calculated the probability of the most likely segregation pattern relative to the total for all possible patterns (Table S3).

2.5 | Transmission disequilibrium test

To test over- or undertransmission of specific haplotypes from heterozygous parents into affected offspring, after appropriate recoding of genotypes to target specific haplotypes we used transmission disequilibrium tests implemented via the “—tdt” option of PLINK v1.90b6.2 (Chang et al., 2015).

3 | RESULTS

3.1 | High-precision typing of *DEFA1A3* copy number

Building on previous methods for characterizing *DEFA1A3* CNV in genomic DNA (Black et al., 2014; Khan et al., 2013), we developed two new PRT (Armour et al., 2007) methods for assessing *DEFA1A3* gene copy number (Figure 1). The first (cen1) exploits a consistent sequence length difference

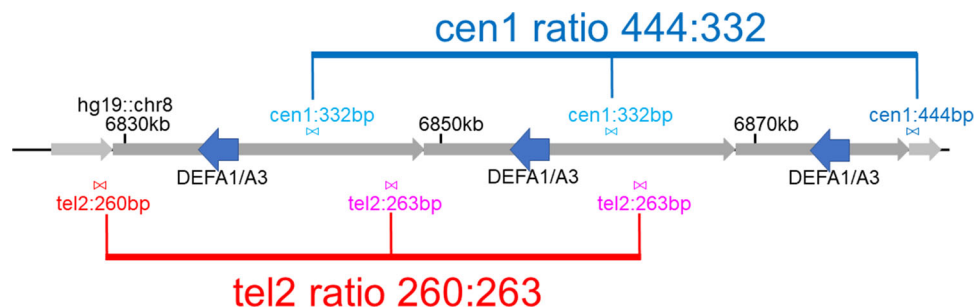


FIGURE 1 Overview of the *DEFA1A3* copy-variable region and the basis of the new measurement methods described in this report. The *DEFA1A3* genes are represented as three tandemly arranged copies (blue arrows) on the hg19 reference assembly. Each copy of the gene is included within near-identical repeats of a 19 kb full repeat unit or (at the centromeric end, shown here on the right) a shorter partial repeat (gray arrows). Flanking these copy-variable regions of near-identical sequence are diverged sequences that do not show common structural variation (lighter gray arrows). The cen1 PRT measurement uses a single primer pair to amplify a 444 bp product (dark blue) from a constant flanking site, and products of 332 bp from other repeats (light blue); the ratio of internal (332 bp) to flanking (444 bp) products can be used to deduce the total copy number for the sample. Similarly, the tel2 primers amplify a 260 bp product (red) from a flanking site and 263 bp products from copy-variable internal locations (pink)

between the copy of *DEFA1A3* found at the centromeric end of the array (shown on the right in Figure 1) and uses a single primer pair to amplify different-length products from the two centromeric copies (per diploid genome) and the variable number of noncentromeric copies. The total copy number can be estimated from the ratio of noncentromeric to centromeric PCR products, assuming that there are exactly two centromeric copies (see the Section 2 Methods above). Similarly, a consistent length difference between the gene position at the telomeric end of the array, shown on the left in Figure 1, is the basis for the tel2 PRT which compares the signal amplified from the (exactly two) telomeric positions with the variable number of nontelomeric positions.

Despite the risk of structural variants disrupting the spatial arrangement of these sequences, we found that in practice the cen1 and tel2 measurements gave consistent outcomes, clustered as expected around integer copy number values (Figure 2). In addition to the cen1 and tel2 PRT measurements, we included assays of other ratios between different gene sequence variants, including the *DEFA1* and *DEFA3* variants of the *DEFA1A3* gene. For example, a sample showing a *DEFA1:DEFA3* ratio of about 1.5 (i.e., 3:2) is consistent with a copy number of 5 or 10 but is harder to reconcile with a copy number of 4. Combining all these ratio measurements with knowledge of the observed magnitudes and distributions of measurement errors allowed us to derive a maximum likelihood copy number (MLCN) for each sample, as previously described (Black et al., 2014; Khan et al., 2013).

We applied these methods to 1146 measurements of *DEFA1A3* copy number; other samples typed included members of uninformative trios, other (nontrio) family members, or spouses of affected individuals. The DNA

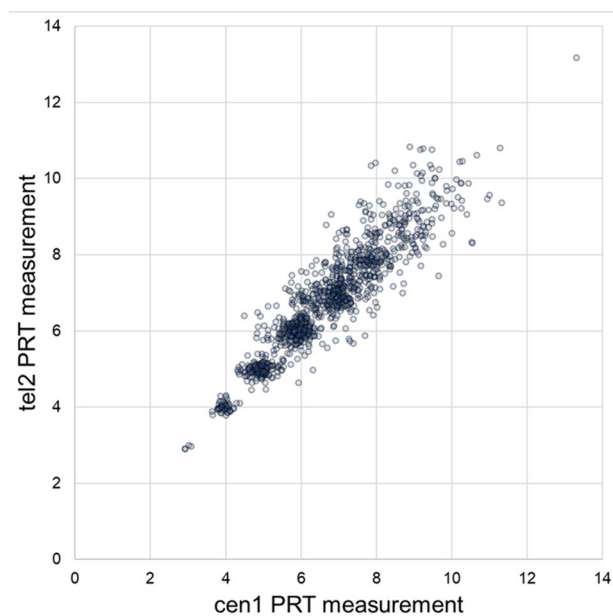


FIGURE 2 Results from 1146 PRT measurements of *DEFA1A3* copy number as part of this study, including members of 193 informative family trios. The two independent measurements using the cen1 or tel2 ratios are well correlated ($r^2 = 0.86$), and the underlying accuracy of the measurements is validated by the strong clustering of data points around integer values, most clearly at copy numbers of 6 or fewer

samples we initially typed included members of 210 family trios with an offspring with IgA nephropathy, of which flanking SNP segregation could be deduced unambiguously for 193 informative trios independently of the CNV status.

All of the 193 IgA nephropathy family trios for which the segregation of flanking SNP genotypes could be resolved

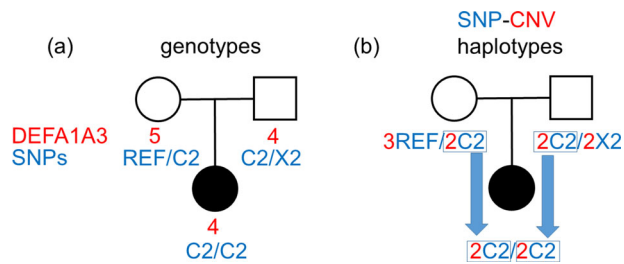


FIGURE 3 A simple example from IgA nephropathy family data illustrating the deduction from (a) SNP and CNV genotypes of (b) joint SNP–CNV haplotypes. In (a), the total *DEFA1A3* copy number as measured by PRT methods is shown in red, with the SNP haplotypes deduced from SNP genotypes shown in blue. In (b), the joint haplotypes are designated in the form (for example) “3REF,” which indicates a “Reference” SNP haplotype carrying three copies of *DEFA1A3*. Neglecting the possibility that a haplotype can have 0 or 1 copies of *DEFA1A3*, the only resolution for the child is to have two 2-copy class 2 (C2) haplotypes, from which the nature of the untransmitted haplotypes in each parent can be deduced by simple subtraction. This trio has low *DEFA1A3* copy numbers for all members, making the inferred haplotypes shown highly likely; by contrast, most trios include an offspring with higher copy number, allowing a wider range of compatible possibilities. In those cases, given the known relationships between SNP haplotype backgrounds and *DEFA1A3* content, the relative likelihoods for all the different possible solutions can be evaluated

had flanking SNP haplotypes corresponding to the five common haplotype classes defined in Europeans (Black et al., 2014). To identify the key features of risk or protective haplotypes for IgA nephropathy, we aimed to exploit

the known relationships between flanking SNP haplotypes and *DEFA1A3* copy numbers (Black et al., 2014) to infer the likely segregation of joint SNP–CNV haplotypes in family trios consisting of IgA nephropathy probands and their parents. To provide a systematic basis for the inferences, we phased SNP and *DEFA1A3* CNV data from 989 Europeans using MOCSphaser (Kato et al., 2008), from which the haplotype frequency (*.hpfq) outputs gave a frequency distribution of different *DEFA1A3* copy number states on each of the five common SNP haplotype backgrounds (Table S2). These distributions could then be used to calculate probabilities for each of the possible segregation patterns for each trio, and the definition of the most likely segregation pattern in each case (Figure 3 and Table S3). Although in some trios there were numerous possible patterns of SNP–CNV segregation, in most families there was a clear most likely segregation pattern, and in 167 trios (87%) the most likely segregation pattern had a posterior probability of 0.75 or greater (Table S3).

In evaluation of the transmission of specific haplotypes into affected offspring from heterozygous parents using TDTs, we first examined the four individual SNPs that between them define the five common haplotypes in Europeans, plus one SNP (rs2738048) highlighted as a lead SNP in several different association studies of IgA nephropathy (Ai et al., 2016; Kiryluk et al., 2014; Yu et al., 2012); none of these appeared to be significantly over- or undertransmitted from heterozygous parents to individuals affected with IgA nephropathy (Table 1).

We then examined (Table 2) the six most common joint SNP–CNV haplotypes defined in our segregation analysis,

TABLE 1 Summary statistics for TDT analysis of SNP segregation in 193 family trios with an IgA nephropathy proband offspring. The first four SNPs act as tags for common haplotypes, and the relevant haplotypes are indicated in the second column

Locus	Haplotype	Focal allele	Versus	Transmitted	Untransmitted	OR	P
rs7825750	C2	C	T	95	81	1.173	0.291
rs62487514	X1	A	C	44	46	0.957	0.833
rs4300027	REF/X1	T	C	98	117	0.838	0.195
rs7826487	C1	G	A	48	53	0.906	0.619
rs2738048	(none)	G	A	78	64	1.219	0.240

TABLE 2 TDT analysis of SNP segregation in 193 family trios with an IgA nephropathy proband offspring for combined SNP–CNV haplotypes at *DEFA1A3*; the naming of joint SNP–CNV haplotypes matches Figure 3, with (for example) “3C2” indicating a 3-copy Class 2 haplotype

Haplotype	Versus	Transmitted	Untransmitted	OR	p
5REF	(all others)	52	57	0.912	0.632
3C1	(all others)	22	37	0.595	0.051
2C2	(all others)	50	48	1.042	0.840
3C2	(all others)	57	36	1.583	0.029
4x1	(all others)	31	32	0.969	0.900
3x2	(all others)	23	14	1.643	0.139

which indicated overtransmission of 3-copy Class 2 haplotypes (“3C2,” odds ratio 1.59, raw $p = 0.029$) and possible undertransmission of 3-copy Class 1 haplotypes (“3C1,” odds ratio 0.60, raw $p = 0.051$). In analyses restricted to the 181 trios for which the most likely segregation pattern had a relative likelihood greater than 0.6 of the total, the apparent unequal transmission remained nominally significant for both overtransmission of 3-copy Class 2 haplotypes ($p = 0.024$) and undertransmission of 3-copy Class 1 haplotypes ($p = 0.026$). However, given that six common haplotypes were examined, these possible effects were no longer statistically significant after Bonferroni correction for multiple testing.

4 | DISCUSSION

In this work, we have applied new approaches to infer the detailed segregation of structural variants at the multiallelic CNV *DEFAIA3* in family trios, by combining accurate measurement of gene copy number with knowledge of the population relationships between SNP variation and CNV status. Although some trios with high copy number in the proband had numerous possible segregation patterns with similar likelihoods, the constraints imposed by the observed segregation and the known copy number spectrum on each SNP-defined haplotype led to the definition of segregation in many trios with good levels of confidence.

Despite this high level of informativeness for segregation, the signals of apparent overtransmission of 3-copy class 2 haplotypes and undertransmission of 3-copy class 1 haplotypes into affected offspring did not reach levels that were statistically significant after correction for testing of multiple haplotypes (Table 2). Because of recombination across the CNV repeat region the phylogenetic relationships of the major European SNP haplotypes are complex and do not shed light on the association findings. For example, the Exchange 1 haplotype appears to be of Neanderthal origin, and the ancestral sequence is represented by the reference haplotype, but these haplotypes do not seem to have a specific role in the associations observed.

If the observed over-transmission of 3-copy class 2 haplotypes into affected offspring is the true signal of a risk haplotype, the effect is not a property of class 2 haplotypes generally, as shown by the even segregation of the class 2-tagging SNP rs7825750 (Table 1); similarly, overtransmission is not seen for 2-copy class 2 haplotypes (Table 2). Functionally important differences between 3-copy and 2-copy class 2 haplotypes could arise from background-specific gene dosage effects, haplotype-specific gene sequence variants, or conceivably because of the presence of sequences important in longer-range control of gene expression distant from the *DEFAIA3* region.

The implication of 3-copy class 1 alleles as a potentially protective factor is interesting in the light of the appearance of the class 1-tagging SNP rs7826487 as a lead SNP in this region in studies of the cellular composition of blood, including associations with monocyte and basophil counts (see Table S4 of Astle et al., 2016). The most frequent sequences of 3-copy class 1 and 3-copy class 2 alleles differ not only in their flanking SNPs but also in the internal structure of the *DEFAIA3* gene array, with a copy of *DEFA3* found on class 2 but not class 1 backgrounds (Black et al., 2014).

Finally, if the 3-copy class 2 and class 1 alleles do in fact have opposing functional effects, the otherwise puzzling consistency with which GWAS data implicate SNPs like rs2738048, which have no close relationship with *DEFAIA3* structural variation, could be explained as an aggregate effect, effectively summing the varied effects of different haplotypes when analysed from the standpoint of a single SNP.

AUTHOR CONTRIBUTIONS

Study design DPG JB JALA; Data collection NS DPG JALA; Contribution of new analytical tools and methodology NS ECM PEN AJF JALA; Data analysis NS JALA; Manuscript preparation JALA, with input from all authors.

ACKNOWLEDGMENTS

This work was supported by an Innovation Grant from Kidney Research UK (IN_005_20170302).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the supporting information of this article.

ORCID

Daniel P. Gale  <https://orcid.org/0000-0002-9170-1579>

REFERENCES

- Ai, Z., Li, M., Liu, W., Foo, J.-N., Mansouri, O., Yin, P., Zhou, Q., Tang, X., Dong, X., Feng, S., Xu, R., Zhong, Z., Chen, J., Wan, J., Lou, T., Yu, J., Zhou, Q., Fan, J., Mao, H., ... Yu, X. (2016). Low α -defensin gene copy number increases the risk for IgA nephropathy and renal dysfunction. *Science Translational Medicine*, 8(345), 345ra388–345ra388. <https://doi.org/10.1126/scitranslmed.aaf2106>
- Akbari, P., Vuckovic, D., Jiang, T., Kundu, K., Kreuzhuber, R., Bao, E. L., Mayer, L., Collins, J. H., Downes, K., Georges, M., Grassi, L., Guerrero, J. A., Kaptoge, S., Knight, J. C., Meacham, S., Sambrook, J., Seyres, D., Stegle, O., Verboon, J. M., ... Astle, W. J. (2020). Genetic analyses of blood cell structure for biological and pharmacological inference. *bioRxiv*, 2020.2001.2030.927483. <https://doi.org/10.1101/2020.01.30.927483>

- Aldred, P. M. R., Hollox, E. J., & Armour, J. A. L. (2005). Copy number polymorphism and expression level variation of the human alpha-defensin genes DEFA1 and DEFA3. *Human Molecular Genetics*, *14*(14), 2045–2052.
- Armour, J. A. L., Palla, R., Zeeuwen, P. L. J. M., den Heijer, M., Schalkwijk, J., & Hollox, E. J. (2007). Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. *Nucleic Acids Research*, *35*(3), e19. <https://doi.org/10.1093/nar/gkl1089>
- Astle, W. J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A. L., Mead, D., Bouman, H., Riveros-Mckay, F., Kostadima, M. A., Lambourne, J. J., Sivapalaratnam, S., Downes, K., Kundu, K., Bomba, L., Berentsen, K., Bradley, J. R., Daugherty, L. C., Delaneau, O., ... Soranzo, N. (2016). The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell*, *167*(5), 1415–1429. e1419. <https://doi.org/10.1016/j.cell.2016.10.042>
- Black, H. A., Khan, F. F., Tyson, J., & Armour, J. A. (2014). Inferring mechanisms of copy number change from haplotype structures at the human DEFA1A3 locus. *BMC Genomics [Electronic Resource]*, *15*(1), 1–11. <https://doi.org/10.1186/1471-2164-15-614>
- Boettger, L. M., Salem, R. M., Handsaker, R. E., Peloso, G. M., Kathiresan, S., Hirschhorn, J. N., & McCarroll, S. A. (2016). Recurring exon deletions in the HP (haptoglobin) gene contribute to lower blood cholesterol levels. *Nature Genetics*, *48*, 359–366.
- Carpenter, D., Dhar, S., Mitchell, L., Fu, B., Tyson, J., Shwan, N., Yang, F., Thomas, M. G., & Armour, J. A. L. (2015). Obesity, starch digestion and amylase: Association between copy number variants at human salivary (AMY1) and pancreatic (AMY2) amylase genes. *Human Molecular Genetics*, *24*, 3472–3480. <https://doi.org/10.1093/hmg/ddv098>
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 7. <https://doi.org/10.1186/s13742-015-0047-8>
- Feehally, J., Farrall, M., Boland, A., Gale, D. P., Gut, I., Heath, S., Kumar, A., Peden, J. F., Maxwell, P. H., Morris, D. L., Padmanabhan, S., Vyse, T. J., Zawadzka, A., Rees, A. J., Lathrop, M., & Ratcliffe, P. J. (2010). HLA Has strongest association with IgA nephropathy in genome-wide analysis. *Journal of the American Society of Nephrology*, *21*(10), 1791–1797. <https://doi.org/10.1681/asn.2010010076>
- Ganz, T. (2003). Defensins: antimicrobial peptides of innate immunity. *Nature Reviews: Immunology*, *3*(9), 710–720.
- Gharavi, A. G., Kiryluk, K., Choi, M., Li, Y., Hou, P., Xie, J., Sanna-Cherchi, S., Men, C. J., Julian, B. A., Wyatt, R. J., Novak, J., He, J. C., Wang, H., Lv, J., Zhu, L., Wang, W., Wang, Z., Yasuno, K., Gunel, M., ... Lifton, R. P. (2011). Genome-wide association study identifies susceptibility loci for IgA nephropathy. *Nature Genetics*, *43*(4), 321–327. <https://doi.org/10.1038/ng.787>
- Hughes, T., Hansson, L., Akkouch, I., Hajdarevic, R., Bringsli, J. S., Torsvik, A., Inderhaug, E., Steen, V. M., & Djurovic, S. (2020). Run-away multi-allelic copy number variation at the α -defensin locus in African and Asian populations. *Scientific Reports*, *10*(1), 9101. <https://doi.org/10.1038/s41598-020-65675-w>
- Kato, M., Nakamura, Y., & Tsunoda, T. (2008). MOCSphaser: a haplotype inference tool from a mixture of copy number variation and single nucleotide polymorphism data. *Bioinformatics*, *24*(14), 1645–1646. <https://doi.org/10.1093/bioinformatics/btn242>
- Khan, F. F., Carpenter, D., Mitchell, L., Mansouri, O., Black, H. A., Tyson, J., & Armour, J. A. (2013). Accurate measurement of gene copy number for human alpha-defensin DEFA1A3. *Bmc Genomics [Electronic Resource]*, *14*(1), 1–12. <https://doi.org/10.1186/1471-2164-14-719>
- Kiryluk, K., Julian, B. A., Wyatt, R. J., Scolari, F., Zhang, H., Novak, J., & Gharavi, A. G. (2010). Genetic studies of IgA nephropathy: past, present, and future. *Pediatric Nephrology*, *25*(11), 2257–2268. <https://doi.org/10.1007/s00467-010-1500-7>
- Kiryluk, K., Li, Y., Scolari, F., Sanna-Cherchi, S., Choi, M., Verbitsky, M., Fasel, D., Lata, S., Prakash, S., Shapiro, S., Fischman, C., Snyder, H. J., Appel, G., Izzi, C., Viola, B. F., Dalleria, N., Del Vecchio, L., Barlassina, C., Salvi, E., ... Gharavi, A. G. (2014). Discovery of new risk loci for IgA nephropathy implicates genes involved in immunity against intestinal pathogens. *Nature Genetics*, *46*(11), 1187–1196. <https://doi.org/10.1038/ng.3118>
- Li, M., Foo, J. N., Wang, J. Q., Low, H. Q., Tang, X. Q., Toh, K. Y., Yin, P. R., Khor, C. C., Goh, Y. F., Irwan, I. D., Xu, R. C., Andiappan, A. K., Bei, J. X., Rotzschke, O., Chen, M. H., Cheng, C. Y., Sun, L. D., Jiang, G. R., Wong, T. Y., ... Liu, J. J. (2015). Identification of new susceptibility loci for IgA nephropathy in Han Chinese. *Nature Communication*, *6*, 7270. <https://doi.org/10.1038/ncomms8270>
- Linzmeier, R. M., & Ganz, T. (2005). Human defensin gene copy number polymorphisms: comprehensive analysis of independent variation in alpha- and beta-defensin regions at 8p22-p23. *Genomics*, *86*(4), 423–430.
- Miles, K., Clarke, D. J., Lu, W., Sibinska, Z., Beaumont, P. E., Davidson, D. J., Barr, T. A., Campopiano, D. J., & Gray, M. (2009). Dying and necrotic neutrophils are anti-inflammatory secondary to the release of α -defensins. *Journal of Immunology*, *183*(3), 2122–2132. <https://doi.org/10.4049/jimmunol.0804187>
- Munz, M., Willenborg, C., Richter, G. M., Jockel-Schneider, Y., Graetz, C., Staufenbiel, I., Wellmann, J., Berger, K., Krone, B., Hoffmann, P., van der Velde, N., Uitterlinden, A. G., de Groot, L., Sawalha, A. H., Direskeneli, H., Saruhan-Direskeneli, G., Guzeldemir-Akcanat, E., Keceli, H. G., Laudes, M., ... Schaefer, A. S. (2017). A genome-wide association study identifies nucleotide variants at SIGLEC5 and DEFA1A3 as risk loci for periodontitis. *Human Molecular Genetics*, *26*(13), 2577–2588. <https://doi.org/10.1093/hmg/ddx151>
- Qi, Y. Y., Zhou, X. J., Cheng, F. J., Hou, P., Zhu, L., Shi, S. F., Liu, L. J., Lv, J. C., & Zhang, H. (2015). DEFA gene variants associated with IgA nephropathy in a Chinese population. *Genes and Immunity*, *16*(3), 231–237. <https://doi.org/10.1038/gene.2015.1>
- Sanchez-Rodriguez, E., Southard, C. T., & Kiryluk, K. (2021). GWAS-Based discoveries in IgA nephropathy, membranous nephropathy, and steroid-sensitive nephrotic syndrome. *Clin J Am Soc Nephrol*, *16*(3), 458–466. <https://doi.org/10.2215/cjn.14031119>
- Sekar, A., Bialas, A. R., de Rivera, H., Davis, A., Hammond, T. R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V., Genovese, G., Rose, S. A., Handsaker, R. E., Daly, M. J., Carroll, M. C., Stevens, B., & McCarroll, S. A., Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2016). Schizophrenia risk from complex variation of complement component 4. *Nature*, *530*, 177–183.
- Shwan, N. A. A., Louzada, S., Yang, F., & Armour, J. A. L. (2017). Recurrent rearrangements of human amylase genes create multiple independent CNV series. *Human Mutation*, *38*(5), 532–539. <https://doi.org/10.1002/humu.23182>

- The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature* 526, 68–74. <https://doi.org/10.1038/nature15393>
- Walker, S., Janyakhantikul, S., & Armour, J. A. L. (2009). Multiplex paralogue ratio tests for accurate measurement of multiallelic CNVs. *Genomics*, 93(1), 98–103. <https://doi.org/10.1016/j.ygeno.2008.09.004>
- Xie, J., Kiryluk, K., Li, Y., Mladkova, N., Zhu, L., Hou, P., Ren, H., Wang, W., Zhang, H., Chen, N., & Gharavi, A. G. (2016). Fine mapping implicates a deletion of CFHR1 and CFHR3 in protection from IgA nephropathy in Han Chinese. *Journal of the American Society of Nephrology*, 27(10), 3187–3194. <https://doi.org/10.1681/asn.2015111210>
- Xie, J., Shapiro, S., & Gharavi, A. (2013). Genetic studies of IgA nephropathy: what have we learned from genome-wide association studies. *Contributions to Nephrology*, 181, 52–64. <https://doi.org/10.1159/000348652>
- Yu, X.-Q., Li, M., Zhang, H., Low, H.-Q., Wei, X., Wang, J.-Q., Sun, L. D., Sim, K. S., Li, Y., Foo, J. N., Wang, W., Li, Z. J., Yin, X. Y., Tang, X. Q., Fan, L., Chen, J., Li, R. S., Wan, J. X., Liu, Z. S., ... Liu, J.-J. (2012). A genome-wide association study in Han Chinese

identifies multiple susceptibility loci for IgA nephropathy. *Nature Genetics*, 44(2), 178–182. <https://doi.org/10.1038/ng.1047>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Shwan, N. A. A., Moise, E. C., Necsoiu, P. E., Farr, A. J., Gale, D. P., Barratt, J., & Armour, J. A. L. (2022). Segregation analysis identifies specific alpha-defensin (*DEFBIA3*) SNP–CNV haplotypes in predisposition to IgA nephropathy. *Annals of Human Genetics*, 1–8. <https://doi.org/10.1111/ahg.12481>