

# Multiple rare genetic variants co-segregating with familial IgA nephropathy all act within a single immune-related network

■ S. N. Cox<sup>1,2,\*</sup>, F. Pesce<sup>1,3,\*</sup>, J.S. El-Sayed Moustafa<sup>3,4</sup>, F. Sallustio<sup>1</sup>, G. Serino<sup>1,5</sup>, C. Kkoufou<sup>3</sup>, A. Giampetruzzi<sup>6</sup>, N. Ancona<sup>7</sup>, M. Falchi<sup>3,4,†</sup> & F. P. Schena<sup>1,2,†</sup> on behalf of the European IgAN Consortium

<sup>1</sup>Department of Emergency and Organ Transplantation, University of Bari Aldo Moro; <sup>2</sup>C.A.R.S.O. Consortium, University of Bari, Bari, Italy; <sup>3</sup>Department of Genomics of Common Disease, Imperial College London; <sup>4</sup>Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; <sup>5</sup>IRCCS 'de Bellis', Laboratory of Experimental Immunopathology; <sup>6</sup>Department of Soil, Plant and Food Sciences, University of Bari Aldo Moro; and <sup>7</sup>ISSIA, CNR, Bari, Italy

**Abstract.** Cox SN, Pesce F, El-Sayed Moustafa JS, Sallustio F, Serino G, Kkoufou C, Giampetruzzi A, Ancona N, Falchi M, Schena FP (University of Bari Aldo Moro; University of Bari, Bari, Italy; Imperial College London; King's College London, London, UK; Laboratory of Experimental Immunopathology; University of Bari Aldo Moro; and ISSIA, CNR, Bari, Italy). Multiple rare genetic variants co-segregating with familial IgA nephropathy all act within a single immune-related network. *J Intern Med* 2017; **281**: 189–205.

**Background.** IgA nephropathy (IgAN) is a common complex disease with a strong genetic involvement. We aimed to identify novel, rare, highly penetrant risk variants combining family-based linkage analysis with whole-exome sequencing (WES).

**Methods.** Linkage analysis of 16 kindreds of South Italian ancestry was performed using an 'affected-only' strategy. Eight most informative trios composed of two familial cases and an intrafamilial control were selected for WES. High-priority variants in linked regions were identified and validated using Sanger sequencing. Custom TaqMan assays were designed and carried out in the 16 kindreds and an independent cohort of 240 IgAN patients and 113 control subjects.

**Results.** We found suggestive linkage signals in 12 loci. After sequential filtering and validation of WES data, we identified 24 private or extremely rare (MAF <0.0003) linked variants segregating with IgAN status. These were present within coding or regulatory regions of 23 genes that merged into a common functional network. The genes were interconnected by *AKT*, *CTNNB1*, *NFKB*, *MYC* and *UBC*, key modulators of WNT/ $\beta$ -catenin and PI3K/Akt pathways, which are implicated in IgAN pathogenesis. Overlaying publicly available expression data, genes/proteins with expression notably altered in IgAN were included in this immune-related network. In particular, the network included the glucocorticoid receptor gene, *NR3C1*, which is the target of corticosteroid therapy routinely used in the treatment of IgAN.

**Conclusion.** Our findings suggest that disease susceptibility could be influenced by multiple rare variants acting in a common network that could provide the starting point for the identification of potential drug targets for personalized therapy.

**Keywords:** family medicine, gene polymorphism, genetics, glomerulonephritis, kidney disease.

## Introduction

Immunoglobulin A nephropathy (IgAN) is the most common form of primary glomerulonephritis worldwide amongst patients undergoing renal biopsy [1, 2]. The gold standard for diagnosis

depends entirely on this invasive procedure. The genetic component of IgAN, supported by familial clustering, may influence the disease on various levels and may also contribute to the variation observed in prevalence [3]. The phenotypic variability of IgAN, characterized by recurrent episodes of gross haematuria concomitant with upper respiratory tract infections or persistent microscopic haematuria and/or mild proteinuria, suggests that

\*These authors contributed equally.

†These authors share senior authorship.

this glomerulonephritis is a complex disease in which multiple genes may be involved and the kidneys are simple bystanders [2].

Various genomewide genetic approaches have been used to study IgAN with the aim to identify specific markers involved in its susceptibility. Genomewide association studies (GWASs) testing common variants with relatively small effects [4] have led to the identification of various susceptibility loci [5–9]. Furthermore, these studies have provided insights into the complex genetic architecture of IgAN and attempted to explain how susceptibility loci correlate with differences in disease prevalence amongst world populations [8, 10]. Most of the identified loci have been rigorously replicated suggesting a strong genetic component, but still a large proportion of disease risk remains unexplained and there are likely to be additional loci that have not been discovered [11].

On the other hand, genomewide linkage analyses (GWLAs), evaluating the co-segregation of genetic markers with the disease in multiplex families, have identified several linked chromosomal regions including 2q36, 4q26–31, 6q22–23 and 17q12–22 [12–14]. These loci are distinct from those identified by GWASs and ultimately both approaches have been unable to identify the exact causal genes involved in IgAN susceptibility [15].

Rare variants, which are not interrogated by GWASs, could be responsible for a substantial proportion of complex human diseases [16] and may explain a considerable portion of their missing heritability [4]. This concept has been recently demonstrated in an innovatively designed study of age-related macular degeneration in which rare causative coding variants were pinpointed within known associated genetic loci [17]. Because 85% of the disease-causing mutations are likely to be located in well-annotated coding regions of the genome [18], the cost-effective whole-exome sequencing (WES) strategy is a useful approach in the identification of rare causative variants.

Here, we report the first study of IgAN employing a combined multistep approach for the identification of rare variants associated with this disease. Using the family-based GWLA approach, we first identified genomic regions containing potential susceptibility loci. Then, WES of selected IgAN familial cases and controls was used to fine map the linked regions and identify rare segregating functional

genomic variants potentially involved in the pathogenesis of familial IgAN.

## Materials and methods

### Sample donors

The genomewide linkage analysis involved 34 biopsy-proven familial IgAN patients and 112 relatives from 16 Italian kindreds of South Italian ancestry (Table 1 and Figure S1). Recruitment strategies are available on the European IgAN Consortium website ([www.igan.net](http://www.igan.net)) [19]. Briefly, familial IgAN was diagnosed when at least two family members had biopsy-proven IgAN; the remaining family members underwent urinalysis. Unaffected family members had at least three documented normal urinalyses (Figure S1). An independent cohort of 240 biopsy-proven IgAN patients and 113 healthy blood donors (HBDs) were included in the study for custom TaqMan SNP genotyping assays. Written informed consent was obtained from all study participants. The study was carried out according to the principles of the Declaration of Helsinki and was approved by the Policlinico di Bari Ethics Review Board.

### GWLAs of IgAN families

Microarray genotyping is available under the GEO accession number GSE44974 (<http://www.ncbi.nlm.nih.gov/geo/>) [20], and a detailed description can be found in the Supplementary Methods. Individuals with more than 5% missing genotypes and more than 5% Mendelian errors were excluded. Markers that failed the Hardy–Weinberg equilibrium test ( $P \leq 1 \times 10^{-6}$ ) and those with a minor allele frequency (MAF) of  $\leq 0.05$  were excluded. Genotyping errors were also detected and removed using Merlin error detection analysis (–error option). Nonparametric linkage (modelling linkage disequilibrium with  $r^2 > 0.10$ ) analysis was performed on 146 individuals for a total of 16 families and 227 114 SNPs.

The Whittemore and Halpern NPL all statistics [21] was used to test for allele sharing amongst affected individuals, calculating the LOD score using the Kong and Cox linear model (–npl option) [22]. A  $P$ -value  $< 5.0 \times 10^{-3}$  (LOD  $> 1.5$ ) [23] was deemed significant. The study was powered to detect a maximum possible LOD score of 5.052 ( $P = 7.05 \times 10^{-7}$ ) for this data set. The analyses were carried out using Merlin software (version 1.1.2) [24].

**Table 1** Clinical and demographic features of IgAN patients, relatives and HBD included in the study<sup>a</sup>

	IgAN <sup>b</sup>	Relatives <sup>b</sup>	IgAN <sup>c</sup>	HBD <sup>c</sup>
Number	34	112	240	113
Male/female	23/11	46/66	164/76	84/29
Age of onset (years)	25 ± 11	n.d.	27 ± 11	n.d.
sCr (mg/dL)	1.36 ± 0.98	n.d.	1.18 ± 0.59	0.87 ± 0.01
eGFR(mL min <sup>-1</sup> per 1.73 m <sup>2</sup> )	88.3 ± 18	n.d.	88 ± 29	n.d.
CKD stage 1 (%)	55	n.d.	53	n.d.
CKD stage 2 (%)	24	n.d.	29	n.d.
CKD stage 3 (%)	15	n.d.	12	n.d.
CKD stage 4 (%)	3	n.d.	6	n.d.
CKD stage 5 (%)	3	n.d.	0	n.d.
Proteinuria (g/24 h)	1.09 ± 0.75	n.d.	1.07 ± 1.28	n.d.
Hypertension (%)	37%	n.d.	36%	n.d.
Histological classification (%)	G1 41% G2 24% G3 35%	n.d.	G1 38% G2 41% G3 20%	n.d.

<sup>a</sup>Values are expressed as mean±S.D. Abbreviations: sCr, serum creatinine; eGFR, estimated glomerular filtration rate (eGFR) has been calculated using the CKD-EPI creatinine formula (mL min<sup>-1</sup> per 1.73 m<sup>2</sup>); G1, Grade 1 mild; G2, Grade 2 moderate; G3 Grade 3 severe; HBS, Healthy Blood Donors; n.d., not determined. Demographic, clinical and histological data refer to the time of biopsy-proven diagnosis.

<sup>b</sup>These patients and relatives have been included in the in the Microarray Genotyping cohort and in the exome-sequencing study.

<sup>c</sup>These patients and controls have been included TaqMan genotyping cohort.

### WES of IgAN families

We performed WES on 16 most informative IgAN patients belonging to eight nonconsanguineous families and eight intrafamilial controls. For the selection of the internal (intrafamilial) negative controls, we performed an identical by descent (IBD) analysis on each of these eight families and identified the closest relative (for each affected individual) with the least IBD sharing (genetically discordant) in the region of interest [24, 25]. Exome libraries were prepared using 3 µg of genomic DNA. Sequence reads were mapped to the reference human genome (UCSC Genome Browser hg19) using the Burrows–Wheeler aligner (BWA; version 0.5.9-r16) [26], with default parameters. The Best Practices Workflow of Genome Analysis Toolkit (GATK, <http://www.broadinstitute.org>) [27] was used for improving the alignments and for genotype calling with recommended parameters. Genotypes were called at first with the GATK Unified Genotyper (UG, version 2.7-4), and the GATK VariantRecalibrator tool was used to score variant calls by a machine-learning algorithm and to

identify a set of high-quality variants using the variant quality score recalibration (VQSR) procedure. GATK was used to filter high-quality variants with hard filtering criteria (variant confidence score ≥30, mapping quality ≥40, read depth ≥5 and strand bias FS filter <60). At a later date, our exome data were also reprocessed using the newer GATK algorithm Haplotype Caller (HC, version v3.3). To be conservative, we decided to retain and evaluate both variant lists generated by HC and UG. These two tools are based on different algorithms [28], but both evaluate haplotypes using an affine gap penalty pair hidden Markov model [29]. Variants were then annotated with the software snpEFF (snpEff\_v2\_0\_5, <http://snpeff.sourceforge.net/download.html>) [30] and categorized into four classes (high, moderate, low and modifier), and the functional impact of coding variants was also predicted. Low-impact variants were predicted by snpEFF and filtered out as they were synonymous coding and ‘assumed to be mostly harmless or unlikely to change protein behaviour’ as described in the manual ([http://snpeff.sourceforge.net/SnpEff\\_manual.html](http://snpeff.sourceforge.net/SnpEff_manual.html)).

Sequence data were filtered against multiple databases using annovar (<http://annovar.openbioinformatics.org>, version 2013Aug23), and MAFs of the called variants were compared against dbSNP 137 (<ftp://ftp.ncbi.nih.gov/snp/>) and 1000 Genomes Project, where we filtered against the European cohort (EUR.MAF, April 2002 release, <ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp>). Then, variants were visualized with Integrative Genomics Viewer (IGV, version 2.3.36). The concordance between the genotypes of the variants identified through WES and bead SNP genotyping was evaluated using the PLINK program. Concordance was of 96% and 94% for UG and HC, respectively. A more detailed description can be found in the Supplementary Methods.

#### *Validation study: Sanger sequencing and variant segregation analysis*

Segregating variants in affected individuals were prioritized based on their scaled C-scores from the Combined Annotation Dependent Depletion (CADD) webserver (<http://cadd.gs.washington.edu>) [31]. Priority variants were validated using Sanger sequencing on IgAN patients and controls from the WES group. Validated variants were further tested in extended families to evaluate their segregation in other pedigree members with urinary abnormalities. See Supplementary Methods for detailed information.

#### *Frequency evaluation: database query and TaqMan SNP genotyping*

The frequencies of variants validated by Sanger sequencing were evaluated using the Exome Aggregation Consortium (ExAC; <http://exac.broadinstitute.org>) and the 1000 Genomes phase 3 (<http://browser.1000genomes.org>) databases. Genotyping was performed using TaqMan<sup>®</sup> assays. Briefly, appropriate amplification primers were designed for the genomic region containing the variant of interest. TaqMan reactions were carried out in 25  $\mu$ L volumes containing 10 ng DNA, according to the manufacturer's protocols. Amplification parameters were as follows: hot start at 95 °C for 10 min; 40 amplification cycles (95 °C for 15 s, 60 °C for 1 min). Results were obtained using a StepOnePlus Real-Time PCR system and genotypes were called using STEPONE Software v2.2 (Life Technologies, Monza, Italy). Genotype clustering patterns were examined and confirmed to be of good quality. Assay controls with known genotype from Sanger sequencing were included in each plate.

#### *Network analysis to identify rare family variants belonging to known pathways*

To assess biological relationships amongst genes, we used Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems, Redwood City, CA, USA; <http://www.ingenuity.com>). We performed canonical pathway analysis. IPA generated networks based on the connectivity of the genes and computed a score for each network. These scores indicated the likelihood of focus genes belonging to a network versus those obtained by chance. A score of 2 indicates a 1 in 100 chance that the focus genes are together in a network because of random chance. The canonical pathway analysis identified the pathways from the IPA library of canonical pathways that were most significant to the input data set. Significance of the canonical pathway was calculated using Fisher's exact test. The whole data set composed of the 52 *in silico*-identified segregating variants was not evaluated as it contained nonconfirmed variants or variants that were also present in the intrafamilial control. For network analysis, we uploaded 23 genes which contained 24 co-segregating variants with the IgAN status. IPA associates a list of drug targets with generated networks based on knowledge base data (US Food and Drug Administration, Goodman and Gilman's The Pharmacological Basis of Therapeutics and Mosby's Drug Consult). The upstream regulator analysis has been used to identify a list of transcriptional regulators that explains observed gene expression changes in our data sets. The calculated overlap *P*-value measures whether there is a statistically significant overlap between the dataset genes and the genes that are regulated by the transcription regulator. It is calculated using Fisher's exact test and significance is generally attributed to *P*-values <0.01.

## Results

### *GWLA of IgAN families*

Sixteen multiplex families comprising 34 IgAN patients and 112 relatives were included in the linkage study (Figure S1). Nonparametric linkage 'affected-only' analysis confirmed and refined the evidence for linkage to previously described susceptibility loci for IgAN on chromosomes 4q24–28 (LOD 2.4,  $P = 4.5 \times 10^{-4}$ ), 6q22–23 (LOD 1.6,  $P = 3.2 \times 10^{-3}$ ) and 17q12–21 (LOD 1.6,  $P = 3.3 \times 10^{-3}$ ) [12, 13]. Our previous linkage study was performed using 400 microsatellites [13]; here, we adopted a SNP-based strategy using a finer genetic



map of 300 000 SNPs (see Supplementary material for detailed description). We found nine additional suggestive linked regions ( $\text{LOD} > 1.5$  or  $P < 5.0 \times 10^{-3}$ ) in other areas of the genome (Table 2 and Figure S2). Interestingly, different families shared multiple linkage signals spread across 11 chromosomes. This prompted us to focus the sequencing on eight of these families that were providing the strongest contribution to multiple linkage signals in novel and known regions (Figure S3).

#### WES of IgAN families

From each of the eight kindreds, we selected a trio comprising two affected individuals and an internal negative control showing in the regions of interest the least IBD sharing with the affected individuals [24, 25]. This resulted in a total of 16 cases and eight familial controls to be sequenced (Figure S3).

We performed a target capture procedure able to target 62 Mb of the genome including the exome, microRNA sites and noncoding upstream and downstream regulatory regions. Parallel sequencing generated a mean of 3.6 billion bases of raw data per sample (range: 0.745–5.9 Gb). After mapping, 98% of reads mapped to the human reference genome with a mean coverage of  $20\times$ . Two distinct algorithms from the GATK were used to call variants: UG and HC (Table S1), identifying an average of 42 975 and 31 390 variants per individual, respectively.

The genetic variant annotation and the effect prediction toolbox SnpEff [30] were used to annotate variants based on their genomic locations, and the coding effects were also predicted.

To identify susceptibility variants potentially involved in IgAN, we adopted a filtering strategy, illustrated in Fig. 1, based on (i) MAF (excluding common variants with  $\text{EUR.MAF} > 0.01$ ), (ii) 'low' effect prediction (excluding all synonymous coding SNPs) and (iii) quality of the variant based on GATK's Variant Call Quality Recalibration (VCQR). This approach led to a total of 21 001 and 27 597 called variants for UG and HC, respectively (Fig. 1). Some variants were called uniquely by one or the other algorithm so we decided to retain both UG and HC lists (Figure S4). Furthermore, we found that the vast majority of variants in each gene region were comparable between the two callers

(Figure S5). Linkage data were then incorporated in the filtering pipeline and only variants located within linked genomic intervals (i.e. showing a  $\text{LOD score} > 1.5$  or  $P < 5 \times 10^{-3}$ ) were considered further (Table S2), leading to the identification of 3112 and 5654 variants for UG and HC, respectively (Fig. 1a). Finally, priority was given to variants that co-segregated in affected individuals and were absent in the intrafamilial control. This procedure was adopted to exclude co-segregating variations due to the relatedness of the trios, thus highlighting the ones that were effectively involved with IgAN. Each set of co-segregating variants per family was then visually inspected using the Integrated Genomics Viewer (Fig. 1b). Visual inspection permitted assessment of the variants in their genomic context, evaluation of repeat sequences and ascertainment that the variation was not an artefact. Furthermore, read pair orientation and overall depth of sequencing in cases and controls were evaluated. We decided to retain low coverage variants in the list of candidates to follow up only if detected in both affected individuals, as the likelihood of this occurring by chance is limited. Likewise, we decided to retain variants that were not contributing to the partial LOD in the linked region if co-segregating, giving priority to the fact that variants were present in both affected individuals and absent in the intrafamilial control. This procedure identified 52 co-segregating variants in affected individuals narrowing down to less than 1% of all variations called within the linked regions (Fig. 1b and Table S3).

Variants were annotated for scaled C-scores using the unsupervised CADD score which prioritizes functional, deleterious and pathogenic variants across many functional categories. We found that none of the co-segregating rare variants was present in more than one family. However, two genes, *BCLAF1* and *CHD5*, were found to contain distinct segregating variants in families 1, 36 and 385 (Fig. 1c).

#### Validation study: Sanger sequencing and variant segregation analysis

We performed Sanger sequencing to validate the rare variants identified by WES and to confirm their segregation in affected individuals and their absence in the unaffected controls. Due to limited DNA availability for some of our families, we decided to validate the 33 most promising variants (Fig. 1c, Table 3 and Tables S3 and S4). Variant

**Table 2** Sixteen families linked to each chromosomal region<sup>a</sup>

Family	ID	N	Chr1	Chr2	Chr3p	Chr3q	Chr4	Chr5	Chr6	Chr8	Chr9	Chr17	Chr20	Chr22
1*	4	0.23	0.23	0.22	0.23	0.23	0.23	0.23	0.23	0.19	0.23	-0.53	0.23	0.23
3	16	-0.37	0.20	-0.32	-0.37	0.00	-0.37	0.20	0.00	0.16	0.20	-0.37	0.20	0.20
4*	18	0.00	0.30	0.28	0.00	0.30	0.30	0.09	0.30	0.25	0.00	0.30	0.00	0.00
7*	5	0.00	0.30	0.00	0.30	0.00	0.30	0.00	0.30	0.25	0.30	0.00	0.30	0.30
14	4	0.19	0.00	0.00	0.00	0.30	0.00	0.30	0.30	0.25	0.30	0.00	0.30	0.30
15*	12	0.23	0.23	-0.45	0.23	0.23	0.23	0.23	0.23	0.19	0.23	0.23	0.23	0.23
16	8	0.00	0.00	0.28	0.00	0.00	0.00	0.00	0.30	0.25	0.00	0.00	0.30	0.30
17	10	0.00	0.00	0.28	0.00	0.00	0.30	0.00	0.00	0.00	0.30	0.30	0.30	0.00
36*	7	0.35	0.35	0.33	-0.23	-0.23	0.35	-0.23	-0.23	0.29	-0.23	0.35	-0.23	0.35
206*	8	0.30	0.30	0.28	0.00	0.00	0.30	0.00	0.00	0.25	0.00	0.00	0.00	0.30
343	10	0.38	-0.06	0.36	0.38	0.38	-0.06	0.28	-0.06	0.26	0.28	-0.06	-0.44	-0.06
344	12	0.20	-0.37	0.18	0.20	0.20	-0.37	0.20	-0.37	-0.26	0.20	0.20	0.20	-0.37
385*	16	0.35	-0.23	0.33	0.35	0.35	0.35	-0.23	-0.23	-0.17	-0.23	0.35	0.35	-0.23
386	5	0.00	0.30	0.00	0.00	0.30	0.30	0.30	0.30	0.25	0.00	0.30	0.00	0.00
393	6	0.00	0.00	0.28	0.30	0.00	0.30	0.00	0.30	-0.67	0.30	0.30	0.00	0.00
483*	5	0.23	0.23	-0.45	0.23	0.23	0.23	0.23	0.23	0.19	0.23	0.23	0.23	0.23
SNP		<b>rs6577472</b>	<b>rs12613771</b>	<b>rs6550478</b>	<b>rs12629557</b>	<b>rs17006113</b>	<b>rs2731665</b>	<b>rs2064687</b>	<b>rs11166903</b>	<b>rs11792985</b>	<b>rs11792985</b>	<b>rs2256020</b>	<b>rs915039</b>	<b>rs2252528</b>
LOD		2.1	1.8	1.6	1.6	2.4	1.6	1.6	1.7	1.7	2.1	1.6	2.0	1.8
P-value		$9.7 \times 10^{-4}$	$2.0 \times 10^{-3}$	$3.4 \times 10^{-3}$	$3.1 \times 10^{-3}$	$4.5 \times 10^{-3}$	$3.2 \times 10^{-3}$	$3.2 \times 10^{-3}$	$2.5 \times 10^{-3}$	$2.5 \times 10^{-3}$	$9.0 \times 10^{-4}$	$3.3 \times 10^{-3}$	$1.3 \times 10^{-3}$	$2.1 \times 10^{-3}$

<sup>a</sup>Positive and negative linkage signals are highlighted in red and blue colour, respectively. The partial LOD (pLOD) contribution for each family is reported in each box. For each chromosome, the top LOD score and P-Value are also reported.

\*Sequenced families; N = genotyped per family. Families 1, 3, 4, were present in the first linkage study performed by Gharavi AG *et al.* identifying the locus IgAN1 (Nat Genet 2000; 26: 354-7). Excluding families 1, 3 and 4, IgAN1 at 6q22-23 is replicated in the remaining 13 families with a LOD score of 1.05 ( $P = 0.014$ ). Families 3, 4, 7, 15, 16, 206, 343, 344, 385, 386, 393 were also included in our previous linkage study performed by Biscaglia L *et al.* (Am J Hum Genet 2006; 79: 1130-4). Families 14, 17, 36, 483 were new and were added exclusively for this study.

prioritization was performed in an unsupervised way based principally on CADD score (28 variants, score >5), gene expression in peripheral blood leucocytes of IgAN patients [32–35] (*JADE1* 4:g129783008t>a, *UBE2G1* 17:g4173166g>a and *DLRAP1* 1:g25894878c>g) and rare co-segregating variants in the same gene in more than one family (*BCLAF1* 6:g136579552a>g and *BCLAF1* 6:g136579558a>g) (Table 3). Of the 33 variants, 28 (85%) were confirmed by direct sequencing (Fig. 1d,e) and the remaining five variants (*USP6* 17:g5036210t>g, *CAAP1* 9:26841936c>a, *USP22* 17:g20931986g>t, *BCLAF1* 6:g136579552 and *BCLAF1* 6:g136579558) were not present. The two variants predicted within the *BCLAF1* gene were excluded (see Figures S6 and S7). Four of the 28 variants (Fig. 1d) were validated by Sanger sequencing but also detected in the unaffected intrafamilial control (*B4GALT5* 20:g48250578t>c, *TOM1L2* 17:g17748047g>a, *ERL1* 17:g27188606a>c and *TG* 8:g133925492c>t).

On the whole, WES data were confirmed with Sanger sequencing on 28 variants of which 24 were found to co-segregate with the IgAN affection status. Most of these variants (63%, 15/24) fell within untranslated or noncoding regulatory regions of the genome.

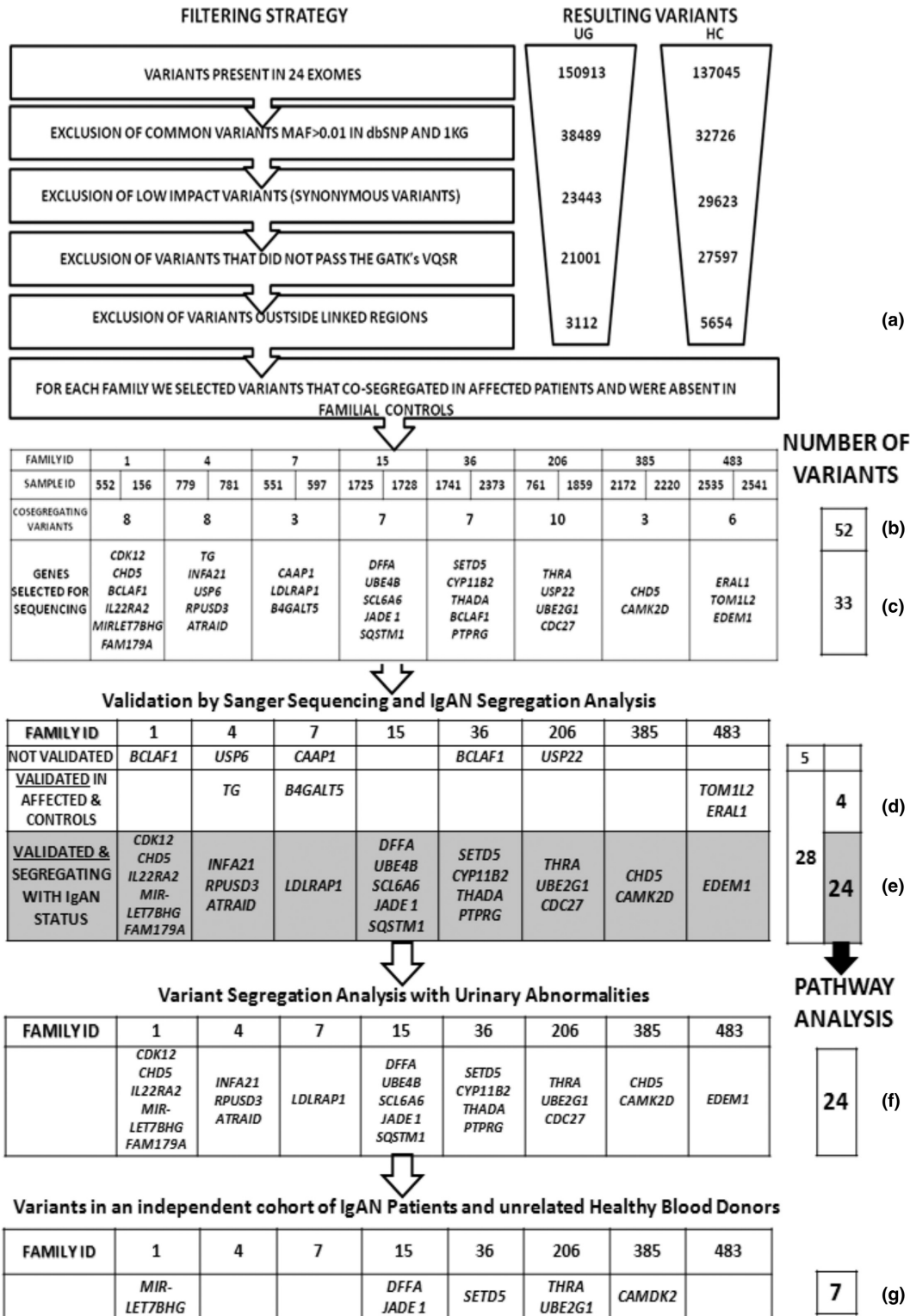
We investigated whether all the 24 variants that were segregating with biopsy-proven IgAN status were also segregating with urinary abnormalities, that are persistent microscopic haematuria and/or mild proteinuria, in extended family members included in WES (Figs 1f and 2, and Table S4). A complete description of the variant segregation in extended families can be found in the Supporting information. Overall, all tested variants segregated both with the IgAN status and with at least one other family member with urinary abnormalities. Moreover, in some families we detected putative obligate carriers. These subjects were characterized by normal urinalysis but carried all the mutations (2397, 2162 and 2542); the presence of putative obligate carriers within pedigrees is consistent with the autosomal dominant mode of transmission of this disease with incomplete penetrance for IgAN [13–15, 36] and suggests that environmental triggers together with genetic factors are needed for disease onset [15, 37]. The identification of multiple mutations in affected individuals suggests that each one may cumulatively concur in determining the disease.

#### *Frequency evaluation: database query and TaqMan SNP genotyping*

Next, we compared the frequencies of all 28 Sanger sequencing-validated variants using two different databases, the ExAC and the 1000 Genomes phase 3 databases; the former is the largest publicly accessible exome-sequencing data set reporting population frequency data of 60 706 unrelated individuals and the latter contains genomewide genotype data of 2504 individuals. We found that amongst the 28 validated variants, 15 were private and the remaining 13 were present in the databases with a low frequency (MAF <0.003) (Table 3). Then, we tested whether the frequencies of seven randomly selected variants were higher in an independent cohort of 240 IgAN patients compared with 113 unrelated HBDs (Fig. 1g). Furthermore, we also tested all families who did not undergo WES, but contributed to the linkage peaks from the original GWLA study. The custom TaqMan SNP genotyping assays were successfully designed for the following variants: *CAMK2D* 4:g114374628t>a, *THRA* 17:g38233146c>t, *MIRLET7BHG* 22:g46423973t>c, *JADE1* 4:g129783008t>a, *DDFA* 1:g10527277g>c, *SETD5* 3:g9515095c>a and *UBE2G1* 17:g4173166g>a (Table S5). For each variant tested, we found that it was only detected in the WES cohort in which it was first discovered. No other IgAN patients or HBDs carried the tested mutations. These results suggest that the tested variants are rare with a frequency of <0.002 in our IgAN population.

#### *Network analysis to identify rare family variants belonging to known pathways*

We investigated whether the 24 variants that co-segregate with the IgAN affection status within 23 apparently unconnected genes were functionally related and concurred to modulate a common pathway. We tested this hypothesis *in silico* using IPA Knowledge Base. This unsupervised analysis identified a single network characterized by a high connectivity score of 59 (Fig. 3a). Of 23 uploaded genes, 22 were within this network (Fig. 3a, red shaded genes). We found that *AKT*, *CTNBN1*, *NFKB*, *MYC* and *UBC* were included as central hubs of the network and all belonged to WNT/ $\beta$ -catenin and PI3K/Akt signalling (Fig. 3a, cyan coloured symbols). All these genes and pathways are implicated in disease pathogenesis and are aberrantly expressed both at gene and protein levels in IgAN patients [35, 38–40]. In particular,





when we overlaid our previously published expression data onto this network, we found that genes and complexes with altered expression in IgAN were also included in this network (*ABHD2*, *CTNNB1*, *CTNN $\beta$ -LEF1*, *DFFA*, *HNRNPA1*, *JADE1*, *KRTCAP2*, *LDLRAP1*, *let-7*, *MS4A4A*, *NR3C1*, *SETD5*, *TCF4-CTNN $\beta$* , *TPRA1*, *Ube2-ubiquitin*, *UBE2G1* and *Ubiquitin* (Fig. 3a, grey shaded symbols) [32, 33, 35, 40, 41]. The representative canonical pathways associated with the uploaded genes are mainly involved in innate and adaptive immunity (Fig. 3b).

Notably, the network included the gene encoding the glucocorticoid receptor, also known as *NR3C1*, which is the target of steroid therapy routinely used in the treatment of IgAN. Corticosteroids are able to lower the risk of kidney failure and reduce proteinuria [2, 42, 43] (Fig. 3B). Furthermore, all gene expression studies performed on IgAN patients showed that the receptor *NR3C1* significantly influences various downstream molecules aberrantly expressed in these patients [33–35] (Fig. 3c–e:  $P = 9.46 \times 10^{-6}$ ,  $P = 2.03 \times 10^{-7}$  and  $P = 1.88 \times 10^{-4}$ , respectively). The complete list of potential drugs that act on this specific IgAN-related network is shown in Table S6.

In conclusion, our results suggest that IgAN-segregating variants interact with each other generating a single network in which altered immune-related pathways have a central role. This network could highlight potential targets for a targeted and personalized therapy.

## Discussion

This is the first study designed to identify rare, highly penetrant risk variants involved in IgAN

integrating two different genomewide approaches, GWLA and WES. We identified 24 variants segregating with IgAN patients and absent in their intrafamilial controls. Furthermore, we confirmed that these variants were either rare or private and were characterized by a strong functional interconnectivity. In addition, these variants appeared in a single network connected by central hubs highly relevant to IgAN.

We first performed a GWLA using a finer SNP map compared with previous linkage studies carried out by microsatellites. The cumulative LOD scores generated supported the previous linkage signals on chromosome 6q22–23 [12] and on chromosomes 4q26–31 and 17q12–22 [13]. The identification of additional linked loci is consistent with a multiple susceptibility gene model for familial IgAN, as suggested in previous linkage studies in IgAN [12–14]. Furthermore, these linked loci were once again distinct from the genomic areas identified by GWASs [5–9]. This discrepancy may be ascribed to (i) differences in the genetic approach, the former identified through a family-based approach and the latter using population-based approaches [44], and (ii) potential underlying differences in the genetic determinants in familial and sporadic IgAN [45].

GWLA was guidance for next-generation sequencing, and the results were used to provide an informative, strictly selected subset of familial cases enriched with putative highly penetrant genetic risk variants, and genetically discordant intrafamilial controls to be sequenced. This approach was adopted to minimize the impact of population stratification, as the quality control methods developed for GWASs of common variants should also be applied to rare variant studies [46].

**Fig. 1** Overview of variant filtering and analysis strategy used to identify co-segregating variants in affected patients with IgA nephropathy (IgAN). Two distinct algorithms from the Genome Analysis Toolkit (GATK) were used to call variants: Haplotype Caller (HC) and Unified Genotyper (UG). These algorithms generated 150 913 and 137 045 variants, respectively, in 24 exomes. Filtering was performed by removing (i) common variants defined by  $EUR.MAF > 1\%$ , (ii) low-impact variants and (iii) variants that did not pass GATK filters. Linkage data were then included in the filtering step, and only variants located within linked genomic intervals were further considered (a). We identified 52 variants that co-segregated in affected individuals and were absent in the intrafamilial control generating a set of co-segregating variants (b). These variants were visually inspected using the Integrated Genomics Viewer, and 33 were selected for Sanger sequencing (c). Of these 33 selected variants, 28 were confirmed by Sanger sequencing. Four of these validated variants were also detected in unaffected controls (*B4GALT5*, *TOM1L2*, *ERAL1* and *TG*) (d). Thus, 24 variants segregated with the IgAN status and were used for pathway analysis (e). The same 24 validated variants were used to determine whether they were also co-segregating with urinary abnormalities (f). Finally, we tested the frequency of seven randomly selected validated variants in an independent cohort of 240 IgAN patients and 113 unrelated healthy blood donors (g).

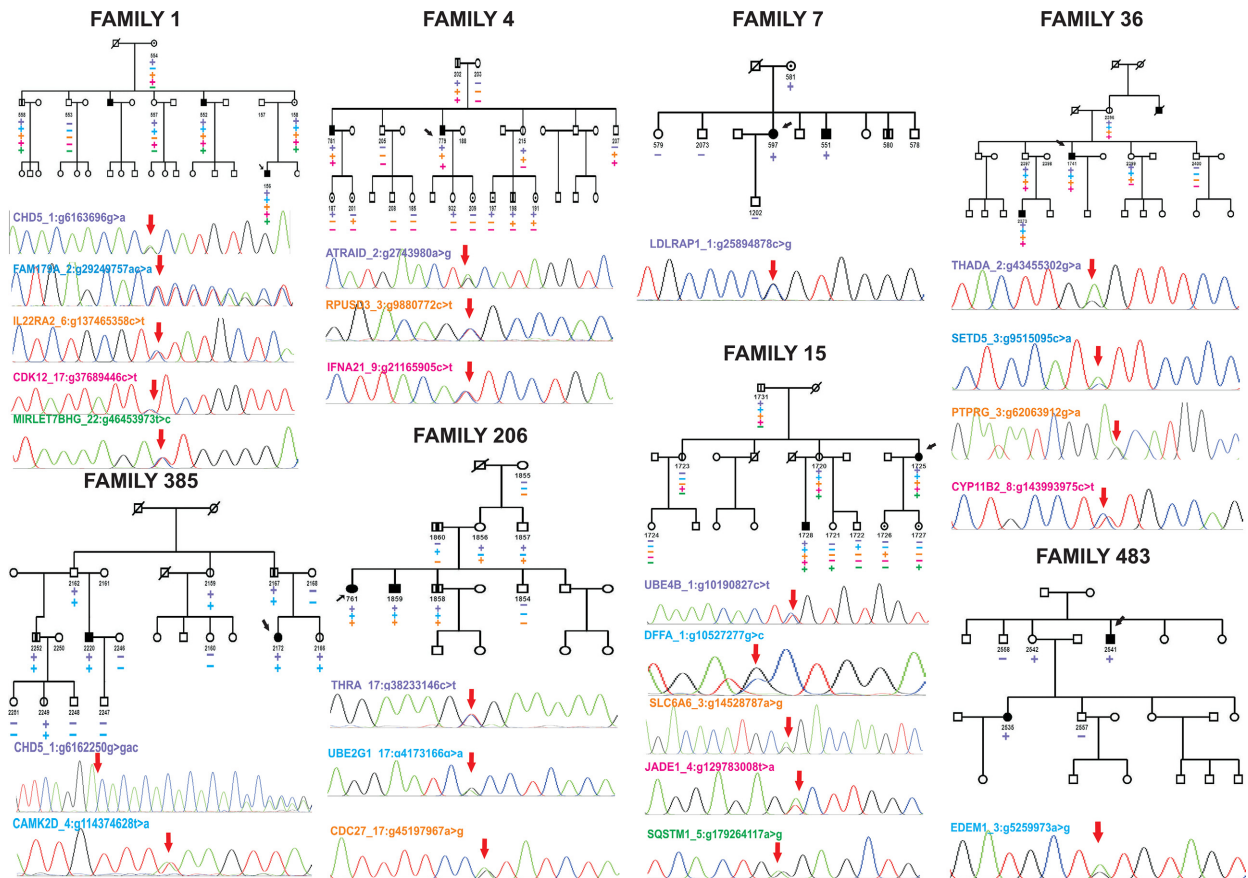
**Table 3** Candidate gene variants selected for Sanger sequencing

Family ID	Gene symbol	Gene name	CHR	POS	REF	ALT	SnpEff effect	Sanger validation	ExAC	1000G phase 3
1	<i>CHD5</i>	Chromodomain helicase DNA binding protein 5	1	6 163 696	G	A	DOWNSTREAM	Yes	NC	Absent
1	<i>FAM179A</i>	Family with sequence similarity 179, member A	2	29 249 757	AC	A	FRAME_SHIFT	Yes	Absent	Absent
1	<i>IL22RA2</i>	Interleukin 22 receptor, alpha 2	6	137 465 358	C	T	UTR_3_PRIME	Yes	NC	Absent
1	<i>CDK12</i>	Cyclin-dependent kinase 12	17	37 689 446	C	T	UTR_3_PRIME	Yes	NC	0.003
1	<i>MIRLET7BHG</i>	MIRLET7B host gene	22	46 453 973	T	C	INTRON	Yes	NC	Absent
4	<i>ATRAID</i>	All-trans retinoic acid-induced differentiation factor	2	27 439 820	A	G	DOWNSTREAM	Yes	NC	Absent
4	<i>RPUSD3</i>	RNA pseudouridylylate synthase domain containing 3	3	9 880 772	T	C	NON_SYNONYMOUS_CODING=T255A	Yes	0.00015	Absent
4	<i>IFNA21</i>	Interferon, alpha 21	9	21 165 905	C	T	UTR_3_PRIME	Yes	NC	Absent
7	<i>LDLRAP1</i>	Low-density lipoprotein receptor adaptor protein 1	1	25 894 878	C	G	DOWNSTREAM	Yes	NC	0.0001
15	<i>UBE4B</i>	Ubiquitination factor E4B	1	10 190 827	C	T	NON_SYNONYMOUS_CODING=R378C	Yes	0.000016	Absent
15	<i>DFFA</i>	DNA fragmentation factor, 45 kDa, alpha polypeptide	1	10 527 277	G	C	NON_SYNONYMOUS_CODING=S137R	Yes	0.0000082	Absent
15	<i>SLC6A6</i>	Solute carrier family 6 (neurotransmitter transporter), member 6	3	14 528 787	A	G	UTR_3_PRIME	Yes	NC	0.00059
15	<i>JADE1</i>	Jade family PHD finger 1	4	129 783 008	T	A	NON_SYNONYMOUS_CODING=S365R	Yes	0.00001648	Absent
15	<i>SQSTM1</i>	Sequestosome 1	5	179 264 117	A	G	DOWNSTREAM	Yes	NC	Absent
36	<i>THADA</i>	Thyroid adenoma associated	2	43 455 302	G	A	DOWNSTREAM	Yes	NC	Absent
36	<i>SETD5</i>	SET domain containing 5	3	9 515 095	C	A	NON_SYNONYMOUS_CODING=S1026Y	Yes	Absent	Absent
36	<i>FTPRG</i>	Protein tyrosine phosphatase, receptor type, G	3	62 063 912	G	A	NON_SYNONYMOUS_CODING=A199T	Yes	0.000008341	Absent
36	<i>CYP11B2</i>	Cytochrome P450, family 11, subfamily B, polypeptide 2	8	143 993 975	C	T	NON_SYNONYMOUS_CODING=E457K	Yes	Absent	Absent
206	<i>THRA</i>	Thyroid hormone receptor, alpha	17	38 233 146	C	T	STOP_GAINED=R26*	Yes	Absent	Absent
206	<i>UBE2G1</i>	Ubiquitin-conjugating enzyme E2G 1	17	4173 166	G	A	UTR_3_PRIME	Yes	NC	Absent
206	<i>CDC27</i>	Cell division cycle 27	17	45 197 967	A	G	DOWNSTREAM	Yes	NC	0.000199681
385	<i>CAMK2D</i>	Calcium/calmodulin-dependent protein kinase II delta	4	114 374 628	T	A	DOWNSTREAM	Yes	NC	0.00259585

Table 3 (Continued)

Family ID	Gene symbol	Gene name	CHR	POS	REF	ALT	SnpEff effect	Sanger validation	ExAC	1000G phase 3
385	<i>CHD5</i>	Chromodomain helicase DNA binding protein 5	1	6 162 250	G	GAC	DOWNSTREAM	Yes	NC	0.0071885
483	<i>EDEMI</i>	ER degradation enhancer, mannosidase alpha-like 1	3	5 259 973	A	G	UTR_3_PRIME	Yes	NC	Absent
4	<i>TG</i>	Thyroglobulin	8	133 925 492	C	T	STOP_GAINED=Q1454*	Yes + Control	0.000008248	Absent
7	<i>B4GALT5</i>	UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide 5	20	48 250 578	T	C	UTR_3_PRIME	Yes + Control	NC	Absent
483	<i>TOM1L2</i>	Target of myb1-like 2 (chicken)	17	17 748 047	G	A	DOWNSTREAM	Yes + Control	NC	0.000599042
483	<i>ERALI/MIR144/</i>	Era-like 12S mitochondrial rRNA chaperone 1	17	27 188 606	A	C	DOWNSTREAM	Yes + Control	NC	Absent
1	<i>ECLAF1</i>	BCL2-associated transcription factor 1	6	136 579 552	A	G	DOWNSTREAM	Repeat region	NC	Absent
36	<i>BCLAF1</i>	BCL2-associated transcription factor 1	6	136 579 558	A	G	DOWNSTREAM	Repeat region	NC	Absent
4	<i>USP6</i>	USP6 ubiquitin specific peptidase 6	17	5 036 210	T	G	NON_SYNONYMOUS_CODING=I67M	Not confirmed	NC	Absent
7	<i>CAAP1</i>	Caspase activity and apoptosis inhibitor 1	9	26 841 936	C	A	DOWNSTREAM	Not confirmed	NC	Absent
36	<i>USP22</i>	Ubiquitin specific peptidase 22	17	20 931 986	G	T	NON_SYNONYMOUS_CODING=A126D	Not confirmed	NC	Absent

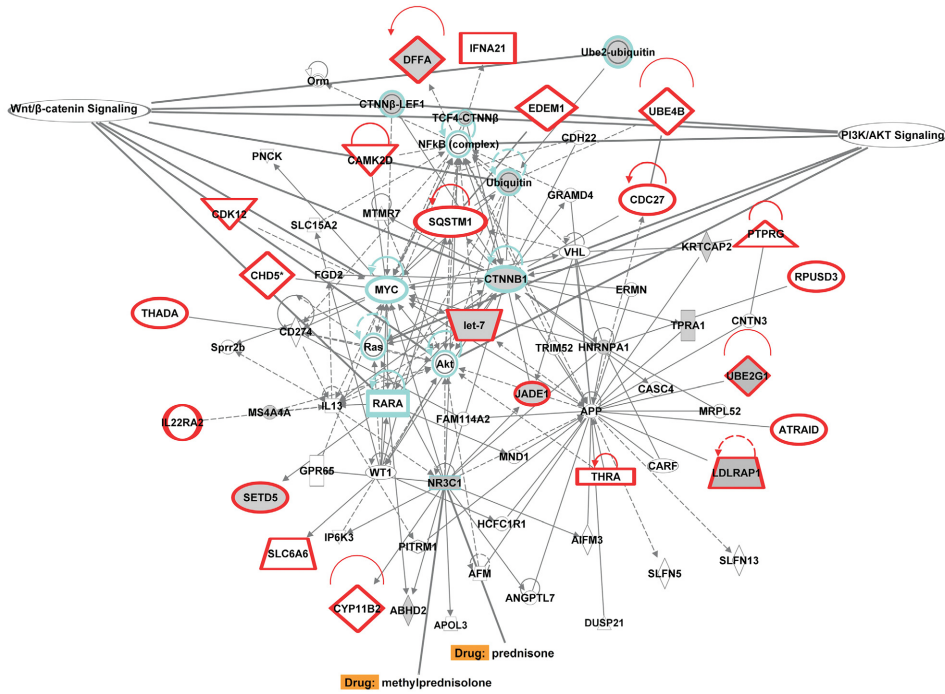
Chr, chromosome; Pos, position; Ref, reference allele; Alt, alternate allele; SnpEff EFFECT, prediction of variant effect on gene based on SnpEff software (<http://snpeff.sourceforge.net/download.html>); ExAC, Exome Aggregation Consortium variant frequency data (<http://exac.broadinstitute.org>); NC, no variant detected because not in coding regions; Absent, no variant detected. 1000G phase 3, variant frequency data from the 1000 Genomes phase 3 database (<http://browser.1000genomes.org>).



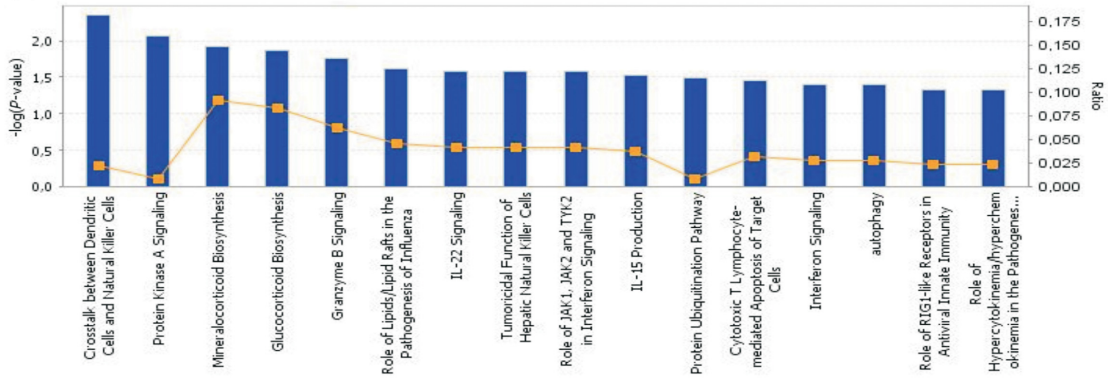
**Fig. 2** Rare variant segregation patterns in families with IgA nephropathy (IgAN). Heterozygous rare mutations that segregated with the affection status were evaluated in extended family members. Squares and circles represent males and females, respectively; arrows indicate probands, and slashes indicate deceased individuals. Filled and unfilled symbols indicate individuals affected and unaffected by IgAN, respectively. Symbols with a dot indicate individuals with an unknown phenotype (individuals without urinalysis or who had intermittent microscopic haematuria). Symbols with a vertical line indicate individuals with documented urinary abnormalities (persistent microscopic haematuria and/or proteinuria). The plus symbol indicates individuals who carry the heterozygous mutation, and the minus symbol indicates those who carry the wild-type allele at this locus. The panels under each pedigree show representative mutant electropherograms of the sequencing products obtained in the index patient (red arrow).

**Fig. 3** Functional analysis of the network containing validated segregating variants. The network was algorithmically constructed using Ingenuity Pathway Analysis (IPA) software based on the functional and biological connectivity of genes (a). The network is graphically represented as nodes (genes) and edges (the biological relationship between genes). Red nodes represent genes containing identified variants; others (empty nodes) are those that IPA automatically includes because they are biologically linked to the studied genes based on evidence in the literature. This is the only network that was generated by the software (score 59,  $n = 22$  associated genes). UBC, AKT, CTNNB1 and NFkB are central nodes computed by the software (cyan colour). We overlaid onto this network previously published expression data obtained from IgAN patients and found that differentially expressed genes and complexes were included in this network (ABHD2, CTNNB1, CTNN $\beta$ -LEF1, DFFA, HNRNPA1, JADE1, KRTCAP2, LDLRAP1, let-7, MS4A4A, NR3C1, SETD5, TCF4-CTNN $\beta$ , TPRA1, Ube2-ubiquitin, UBE2G1 and Ubiquitin; grey shaded symbols). The most representative canonical pathways associated with the uploaded genes are mainly involved in innate and adaptive immunity (b). The glucocorticoid receptor gene (NR3C1) is the target of steroid therapy (prednisone and methylprednisolone) routinely used in IgAN treatment (a). The upstream regulator analysis identified NR3C1 as a transcriptional regulator that explains the observed gene expression changes in differently expressed genes in peripheral blood cells (c,  $P = 9.46 \times 10^{-6}$ ), during the acute phase of the disease (d,  $P = 2.03 \times 10^{-7}$ ) and in monocytes isolated from IgAN patients (e,  $P = 1.88 \times 10^{-4}$ ).

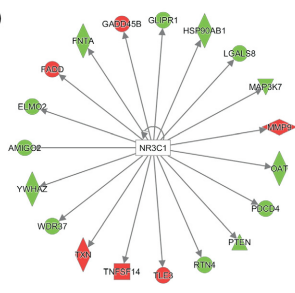
(a)



(b)

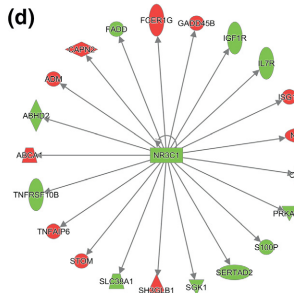


(c)



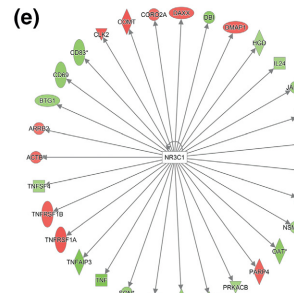
Overlap  $P$  value = 946E-06

(d)



Overlap  $P$  value = 235E-07

(e)



Overlap  $P$  value = 188E-04



After applying filtering criteria and having visually inspected the genomic context, we identified 52 co-segregating variants. They were selected due to their presence in both affected individuals and absence in unaffected intrafamilial sequencing controls. Because the number of rare variants in the human genome is very high [47], this approach excluded co-segregating variants due to the relatedness of the trios and rapidly narrowed down the potentially causative and co-segregating variants. The segregation analysis in our extended families demonstrated that variants segregated with the IgAN status and in individuals with persistent microscopic haematuria and/or mild proteinuria. Furthermore, we also detected putative obligate carriers in unaffected family members. The presence of putative obligate carriers within pedigrees is consistent with the autosomal dominant mode of transmission of this disease with incomplete penetrance for IgAN [13–15, 36]. These obligate carriers provide evidence that IgAN may be determined by multiple genes and environmental factors that interact in a complex network [15, 37]. A multihit model has been postulated for IgAN [11, 48] in which other environmental factors are needed for disease onset, with genetic risk having recently been shown to correlate strongly with variation in local pathogens. In particular, helminth diversity confers an increased risk in IgAN [8].

Then, we sought to evaluate whether the variants identified in our WES cohort were also present in other well-characterized IgAN patients and unrelated HBDs, but none carried the tested variants. The failure to identify a single rare segregating variant in more than one family or any of the rare variants in the other tested familial or sporadic IgAN cases supports the hypothesis that this complex and fairly common disease may be ascribed not to a single gene or variant but to a cumulative effect of rare variants [49].

Finally, we focused on the 24 validated variants within 23 genes that fulfilled the segregation criteria in affected individuals and were not carried by unaffected controls. The genes were interconnected in a single functional network with *AKT*, *CTNNA1*, *NFKB*, *MYC* and *UBC* as central nodes. These genes appear to be biologically relevant because they are key modulators of WNT/ $\beta$ -catenin and PI3K/Akt pathways, which are altered in this disease [35, 39–41, 50].

The network included the glucocorticoid receptor gene, *NR3C1*, which is the target of corticosteroid therapy recommended by the KDIGO guidelines for the treatment of IgAN [2, 43, 51]. We found that *NR3C1* significantly influences various downstream molecules aberrantly expressed in IgAN patients and its modulation may contribute to the beneficial effects of NR3C1 agonists in this disease [2, 42, 43]. Due to the increased risk of adverse events, the use of corticosteroids in IgAN has been recently debated [52–55] and evokes the need for a more targeted therapy. The analysis of other drug targets included in this IgAN-related network could be a starting point for future studies on personalized therapies (Table S6).

Recently, a WES study has been performed on individuals of Chinese Han ethnicity with familial IgAN [56] and seven very common co-segregating deleterious variants within five genes have been identified. These variants were not present in our filtered data set as we applied the stringent filtering criteria of MAF <1%. We further checked our data and found that there were no rare segregating variants within these genes. This difference could also be ascribed to the different ethnicity.

Our study has several limitations. First, because of the limited availability of DNA, it is not possible to validate all the segregating variants identified bioinformatically. Nevertheless, we pursued 64% of the variants (33 of 52) based mainly on the systematic selection of their higher CADD score, which is able to prioritize variants across functional categories and effect sizes. Secondly, the effect of each variation was not evaluated experimentally. As an alternative approach, we overlaid our previously validated experimental gene and protein expression data onto the functional network generated by IPA and found that statistically significant differently expressed genes/complexes in IgAN were included in the network. Further studies are needed to understand whether these variants can effectively disrupt normal function of their relevant genes and whether such a disruption may contribute to our phenotype of interest.

Despite these limitations, our study has several strengths. First, our results suggest a multiple susceptibility gene model for familial IgAN. Secondly, we have identified many rare segregating variants that are involved in a common pathway mostly with a role in innate and adaptive immunity. Thirdly, the identified network contains genes

and pathways previously demonstrated to be dysregulated in IgAN and drug targets that could be exploited in the treatment of the disease.

IgAN is characterized by an altered innate immunity [50] with a hyper-responsiveness of the IgA immune system and an increased IgA-secreting plasma cell number in both bone marrow and tonsils [57, 58]. These cells show a reduced susceptibility to Fas-mediated apoptosis with marked expression of bcl-2 [59]. Furthermore, WNT/ $\beta$ -catenin and PI3K/Akt pathways are hyperactivated in peripheral blood mononuclear cells with increased Akt phosphorylation,  $\beta$ -catenin and NF $\kappa$ B nuclear translocation [35, 39] ultimately leading to an enhanced cell proliferation rate. The uncovering of 23 variant-containing genes, strictly connected to these aberrant networks, may in part explain the IgAN cell hyperactivated phenotype, thus partly revealing the pathogenic mechanisms of IgAN.

### Conclusions

Together, our results support the novel hypothesis that IgAN is caused by distinct rare segregating variants, which may act within a specific IgAN immune-related network. Our findings could be used in future studies to test novel drug targets for the treatment of this disease.

### Conflict of interest

The authors declare no conflicts of interest.

### Availability of data and materials

Microarray data of 146 IgAN patients and relatives genotyped on Illumina HumanCytoSNP-12 have been previously deposited in NCBI's Gene Expression Omnibus (GEO) database under the accession number GSE44974. WES data for all families have been deposited in NCBI's SRA database (<http://www.ncbi.nlm.nih.gov/sra>) and are available under the study accession number SRP061415.

### Funding

This work was supported by grants from the European Framework Programme (QLG1-CT-2000-00464, to FPS), MRC (MR/K01353X/1, to MF), MiUR (PON-REC ONEV 134/2011 and FIRB RBAP11B2SX to FPS), Regione Puglia (BISIMANE project, 44/2008 to FPS) and the Ministry of Health

(GR-2011-02350438 to GS). FP was supported by long-term fellowships from the European Renal Association–European Dialysis and Transplant Association (ERA-EDTA ALTF 72-2010 and ERA-EDTA ALTF 89-2011). We are grateful to the Schena Foundation for scientific and financial support.

### Acknowledgments

We are grateful to the IgAN patients, their family members and the HBDs for their cooperation in this study. We thank Arturo Argentieri (ISSIA-CNR) for his technical support.

### References

- Schena FP, Pesce F. Epidemiology and ancestral difference. In: Lai KN, ed. *Recent Advances in IgA Nephropathy*. Singapore: World Scientific, 2009.
- Wyatt RJ, Julian BA. IgA nephropathy. *N Engl J Med* 2013; **368**: 2402–14.
- McGrogan A, Franssen CF, de Vries CS. The incidence of primary glomerulonephritis worldwide: a systematic review of the literature. *Nephrol Dial Transplant* 2011; **26**: 414–30.
- Maher B. Personal genomes: the case of the missing heritability. *Nature* 2008; **456**: 18–21.
- Feehally J, Farrall M, Boland A *et al.* HLA has strongest association with IgA nephropathy in genome-wide analysis. *J Am Soc Nephrol* 2010; **21**: 1791–7.
- Gharavi AG, Kiryluk K, Choi M *et al.* Genome-wide association study identifies susceptibility loci for IgA nephropathy. *Nat Genet* 2011; **43**: 321–7.
- Yu XQ, Li M, Zhang H *et al.* A genome-wide association study in Han Chinese identifies multiple susceptibility loci for IgA nephropathy. *Nat Genet* 2012; **44**: 178–82.
- Kiryluk K, Li Y, Scolari F *et al.* Discovery of new risk loci for IgA nephropathy implicates genes involved in immunity against intestinal pathogens. *Nat Genet* 2014; **46**: 1187–96.
- Li M, Foo JN, Wang JQ *et al.* Identification of new susceptibility loci for IgA nephropathy in Han Chinese. *Nat Commun* 2015; **6**: 7270.
- Kiryluk K, Li Y, Sanna-Cherchi S *et al.* Geographic differences in genetic susceptibility to IgA nephropathy: GWAS replication study and geospatial risk analysis. *PLoS Genet* 2012; **8**: <http://dx.doi.org/10.1371/journal.pgen.1002765>.
- Kiryluk K, Novak J. The genetics and immunobiology of IgA nephropathy. *J Clin Invest* 2014; **124**: 2325–32.
- Gharavi AG, Yan Y, Scolari F *et al.* IgA nephropathy, the most common cause of glomerulonephritis, is linked to 6q22-23. *Nat Genet* 2000; **26**: 354–7.
- Bisceglia L, Cerullo G, Forabosco P *et al.* Genetic heterogeneity in Italian families with IgA nephropathy: suggestive linkage for two novel IgA nephropathy loci. *Am J Hum Genet* 2006; **79**: 1130–4.
- Paterson AD, Liu XQ, Wang K *et al.* Genome-wide linkage scan of a large family with IgA nephropathy localizes a novel susceptibility locus to chromosome 2q36. *J Am Soc Nephrol* 2007; **18**: 2408–15.

- 15 Kiryluk K, Julian BA, Wyatt RJ *et al.* Genetic studies of IgA nephropathy: past, present, and future. *Pediatr Nephrol* 2010; **25**: 2257–68.
- 16 Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* 2010; **11**: 415–25.
- 17 Fritsche LG, Igl W, Bailey JN *et al.* A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat Genet* 2016; **48**: 134–43.
- 18 Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you? *J Med Genet* 2011; **48**: 580–9.
- 19 Schena FP, Cerullo G, Torres DD *et al.* The IgA nephropathy Biobank. An important starting point for the genetic dissection of a complex trait. *BMC Nephrol* 2005; **6**: doi: 10.1186/1471-2369-6-14.
- 20 Sallustio F, Cox SN, Serino G *et al.* Genome-wide scan identifies a copy number variable region at 3p21.1 that influences the TLR9 expression levels in IgA nephropathy patients. *Eur J Hum Genet* 2015; **23**: 940–8.
- 21 Whittemore AS, Halpern J. A class of tests for linkage using affected pedigree members. *Biometrics* 1994; **50**: 118–27.
- 22 Kong A, Cox NJ. Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 1997; **61**: 1179–88.
- 23 Nyholt DR. All LODs are not created equal. *Am J Hum Genet* 2000; **67**: 282–8.
- 24 Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002; **30**: 97–101.
- 25 Powell JE, Visscher PM, Goddard ME. Reconciling the analysis of IBD and IBS in complex trait studies. *Nat Rev Genet* 2010; **11**: 800–5.
- 26 Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010; **26**: 589–95.
- 27 DePristo MA, Banks E, Poplin R *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; **43**: 491–8.
- 28 Pirooznia M, Kramer M, Parla J *et al.* Validation and assessment of variant calling pipelines for next-generation sequencing. *Hum Genom* 2014; **8**: doi: 10.1186/1479-7364-8-14.
- 29 Krogh A, Brown M, Mian IS, Sjolander K, Haussler D. Hidden Markov models in computational biology. Applications to protein modeling. *J Mol Biol* 1994; **235**: 1501–31.
- 30 Cingolani P, Platts A, Wang L *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 2012; **6**: 80–92.
- 31 Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014; **46**: 310–5.
- 32 Serino G, Sallustio F, Curci C *et al.* Role of let-7b in the regulation of N-acetylgalactosaminyltransferase 2 in IgA nephropathy. *Nephrol Dial Transplant* 2015; **30**: 1132–9.
- 33 Cox SN, Serino G, Sallustio F *et al.* Altered monocyte expression and expansion of non-classical monocyte subset in IgA nephropathy patients. *Nephrol Dial Transplant* 2015; **30**: 1122–232.
- 34 Cox SN, Sallustio F, Serino G *et al.* Activated innate immunity and the involvement of CX3CR1-fractalkine in promoting hematuria in patients with IgA nephropathy. *Kidney Int* 2012; **82**: 548–60.
- 35 Cox SN, Sallustio F, Serino G *et al.* Altered modulation of WNT-beta-catenin and PI3K/Akt pathways in IgA nephropathy. *Kidney Int* 2010; **78**: 396–407.
- 36 Beerman I, Novak J, Wyatt RJ, Julian BA, Gharavi AG. The genetics of IgA nephropathy. *Nat Clin Pract Nephrol* 2007; **3**: 325–38.
- 37 Hsu SI. Racial and genetic factors in IgA nephropathy. *Semin Nephrol* 2008; **28**: 48–57.
- 38 Tang SC, Lai KN. The ubiquitin-proteasome pathway and IgA nephropathy: a novel link? *Kidney Int* 2009; **75**: 457–9.
- 39 Coppo R, Camilla R, Alfarano A *et al.* Upregulation of the immunoproteasome in peripheral blood mononuclear cells of patients with IgA nephropathy. *Kidney Int* 2009; **75**: 536–41.
- 40 Sallustio F, Serino G, Cox SN *et al.* Aberrantly methylated DNA regions lead to low activation of CD4+ T-cells in IgA nephropathy. *Clin Sci (Lond)* 2016; **130**: 733–46.
- 41 Cox SN, Sallustio F, Serino G *et al.* Activated innate immunity and the involvement of CX3CR1-fractalkine in promoting hematuria in patients with IgA nephropathy. *Kidney Int* 2012; **82**: 548–60.
- 42 Manno C, Torres DD, Rossini M, Pesce F, Schena FP. Randomized controlled clinical trial of corticosteroids plus ACE-inhibitors with long-term follow-up in proteinuric IgA nephropathy. *Nephrol Dial Transplant* 2009; **24**: 3694–701.
- 43 Lv J, Xu D, Perkovic V *et al.* Corticosteroid therapy in IgA nephropathy. *J Am Soc Nephrol* 2012; **23**: 1108–16.
- 44 Bailey-Wilson JE, Wilson AF. Linkage analysis in the next-generation sequencing era. *Hum Hered* 2011; **72**: 228–36.
- 45 Schena FP, Cerullo G, Rossini M, Lanzilotta SG, D’Altri C, Manno C. Increased risk of end-stage renal disease in familial IgA nephropathy. *J Am Soc Nephrol* 2002; **13**: 453–60.
- 46 Ott J, Wang J, Leal SM. Genetic linkage analysis in the age of whole-genome sequencing. *Nat Rev Genet* 2015; **16**: 275–84.
- 47 Henn BM, Botigue LR, Bustamante CD, Clark AG, Gravel S. Estimating the mutation load in human genomes. *Nat Rev Genet* 2015; **16**: 333–43.
- 48 Suzuki H, Kiryluk K, Novak J *et al.* The pathophysiology of IgA nephropathy. *J Am Soc Nephrol* 2011; **22**: 1795–803.
- 49 Feehally John BJ. The genetics of IgA nephropathy: an overview from western countries. *Kidney Dis* 2015; **1**: 33–41.
- 50 Coppo R, Amore A, Peruzzi L, Vergano L, Camilla R. Innate immunity and IgA nephropathy. *J Nephrol* 2010; **23**: 626–32.
- 51 Kidney Disease: Improving Global Outcomes (KDIGO) Glomerulonephritis Work Group. KDIGO Clinical Practice Guideline for Glomerulonephritis. *Kidney Int Suppl* 2012; **2**: 139–274.
- 52 Rauen T, Eitner F, Fitzner C *et al.* Intensive supportive care plus Immunosuppression in IgA nephropathy. *N Engl J Med* 2015; **373**: 2225–36.
- 53 Schena FP, Manno C. Intensive supportive care plus immunosuppression in IgA nephropathy. *N Engl J Med* 2016; **374**: 992.
- 54 Ayoub I, Hebert L, Rovin BH. Intensive supportive care plus immunosuppression in IgA nephropathy. *N Engl J Med* 2016; **374**: 991–2.
- 55 Robert T, Cambier A, Hertig A. Intensive supportive care plus immunosuppression in IgA nephropathy. *N Engl J Med* 2016; **374**: 991.
- 56 Liu RH, Hu B, Li QB *et al.* Novel genes and variants associated with IgA nephropathy by co-segregating with the disease phenotypes in 10 IgAN families. *Gene* 2015; **571**: 43–51.

- 57 Harper SJ, Allen AC, Bene MC *et al.* Increased dimeric IgA-producing B cells in tonsils in IgA nephropathy determined by in situ hybridization for J chain mRNA. *Clin Exp Immunol* 1995; **101**: 442–8.
- 58 Harper SJ, Allen AC, Pringle JH, Feehally J. Increased dimeric IgA producing B cells in the bone marrow in IgA nephropathy determined by in situ hybridisation for J chain mRNA. *J Clin Pathol* 1996; **49**: 38–42.
- 59 Kodama S, Suzuki M, Arita M, Mogi G. Increase in tonsillar germinal centre B-1 cell numbers in IgA nephropathy (IgAN) patients and reduced susceptibility to Fas-mediated apoptosis. *Clin Exp Immunol* 2001; **123**: 301–8.

*Correspondence:* Sharon Natasha Cox, C.A.R.S.O. Consortium, University of Bari, 70100 Valenzano (Ba), Italy.  
(fax: 0039/080-4602121; e-mail: sharonnatasha.cox@uniba.it).

### Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Figure S1.** Sixteen multiplex families included in the linkage study, the red bars represent 146 genotyped subjects.

**Figure S2.** Plot of LOD score statistics from NPL analysis for the chromosomes in which the score exceeded the 1.5 level.

**Figure S3.** Pedigrees included in the exome sequencing study.

**Figure S4.** Number of variants called by Unified Genotyper (UG) and Haplotype Caller (HC).

**Figure S5.** Percentage distributions of gene variations as depicted in the output of SnpEff: Variant analysis.

**Figure S6.** The two variants predicted within the *BCLAF1* gene were excluded as they were actually determined by the presence of a deletion in a repeated (AAAAAC)<sub>n</sub> region (and not by a G/A substitution).

**Figure S7.** We evaluated BCLAF1 complex genomic region in six unrelated HBD

**Table S1.** Summary of the exome sequencing results.

**Table S2.** Genomic intervals considered for variant filtration.

**Table S3.** List of co-segregating gene variants identified by exome sequencing.

**Table S4.** Left and right PCR primers designed to amplify the genomic region containing the variant of interest used for Sanger validation.

**Table S5.** List of primers designed for SNP genotyping analysis using custom TaqMan assays.

**Table S6.** List of potential drugs targeting the network. ■