Galactosylation of IgA1 is associated with common variation in *C1GALT1*

Daniel P Gale[1], Karen Molyneux[2], David Wimbury[2], Patricia Higgins[2], Adam P Levine[3], Ben Caplin[1], Anna Ferlin[1], Peiran Yin[4], Christopher Nelson[5], Horia Stanescu[1], Nilesh Samani[5], Robert Kleta[1], Xueqing Yu[4], Jonathan Barratt[2]

Supplementary Material

[1]Centre for Nephrology, University College London, UK

[2]Department of Infection, Immunity & Inflammation, University of Leicester, UK

[3]Division of Medicine, University College London, UK

[4]Department of Nephrology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong, China

[5]Department of Cardiovascular Sciences, University of Leicester & National Institute for Health Research Leicester Cardiovascular Biomedical Research Unit, Leicester, UK

Measurement of Gd-IgA1 by Lectin binding ELISA

IgA was captured on 96-well immunoplates, coated overnight at 4°C with 10 µg/ml anti-human IgA antibody (DAKO A0262), washed, and blocked with 2% BSA in PBS. Serum samples, diluted 1:100 in PBS, were applied to the plates (50 µL/well), in duplicate, and incubated overnight at 4°C. At this dilution, all wells on the plate were completely saturated with IgA1. This enables equivalent amounts of IgA1 from each sample to be tested by the assay. IgA bound to each well was desialylated with 2 units of neuraminidase diluted in 0.5M sodium acetate buffer (pH5) with 100 IU/ml penicillin, and 100 mg/ml streptomycin, overnight at 37°C. After washing to remove the neuraminidase, Gd-IgA1 was detected by incubation for 90 minutes with biotinylated HA (Sigma L8764) followed by HRP-conjugated avidin (DY 998). The reaction was developed with OPD/H2O2 substrate and the results read as absorbance at 492 nm. The same batch of HA was used to perform all the analyses and three standard serum samples with high, medium and low HA lectin binding were included on all plates to allow plate to plate normalisation. Results are reported as normalised optical absorbance in as Arbitrary Units (AU). Intra- and inter-assay variations were <10%.

To determine the stability of Gd-IgA1 levels within individuals and whether the level of Gd-IgA1 was influenced by renal function we separately analysed a cohort of 31 individuals (16 IgAN cases (3 female) and 15 healthy subjects (5 female)) from the Leicester IgAN Research Archive who had sequential samples collected over the past 19 years. Levels of Gd-IgA1 were measured at two timepoints (T1 and T2) using HA lectin and peanut agglutinin (PNA) based ELISA methods as described above. The median time between T1 and T2 was 64 months, range 9 months to 10 years (IgAN) and 59 months, range 3 months to 19 years (healthy subjects). eGFR calculated using the modified MDRD formula was available at T1 and T2 for all IgAN patients. 9/16 IgAN patients progressed between T1 and T2 with 4 patients reaching end stage kidney disease.

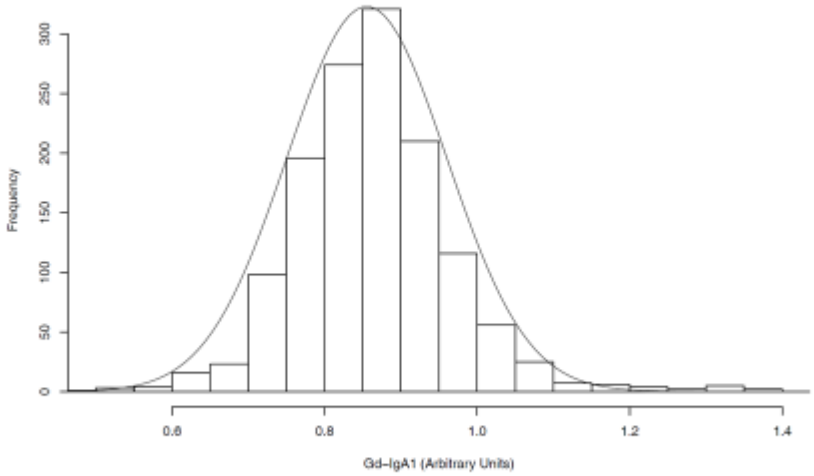
Genetic analyses

Analyses, including quality control steps, covariate and conditional analyses, were performed using Plink[49, 50]. Genomic inflation factor (λ) was estimated based on the median Chi^2 distribution. After quality control of genotype data, as described above, λ was 1.00196 in the discovery cohort, 0.999 in the GRAPHIC cohort and 1.000 in the MN cohort, indicating that significant population substructure was unlikely to be responsible for the associations observed. Combined analysis of the IgAN and MN cohorts was performed using only the 297,873 SNPs that were successfully (>90%) genotyped in both cohorts and passed quality control filters described. To validate the procedure and confirm data integrity, the IgA cohort was scored as 'cases' and the MN 'controls' for a case-control study. As expected, significant differences in allele frequencies were observed at loci on chromosomes 6 and 2 (harboring alleles within the major histocompatibility complex and *PLA2R1* loci previously reported in these cohorts[9, 46]) but not elsewhere in the genome, and λ was 1.043 in this analysis, suggesting that significant population stratification effects were not present.

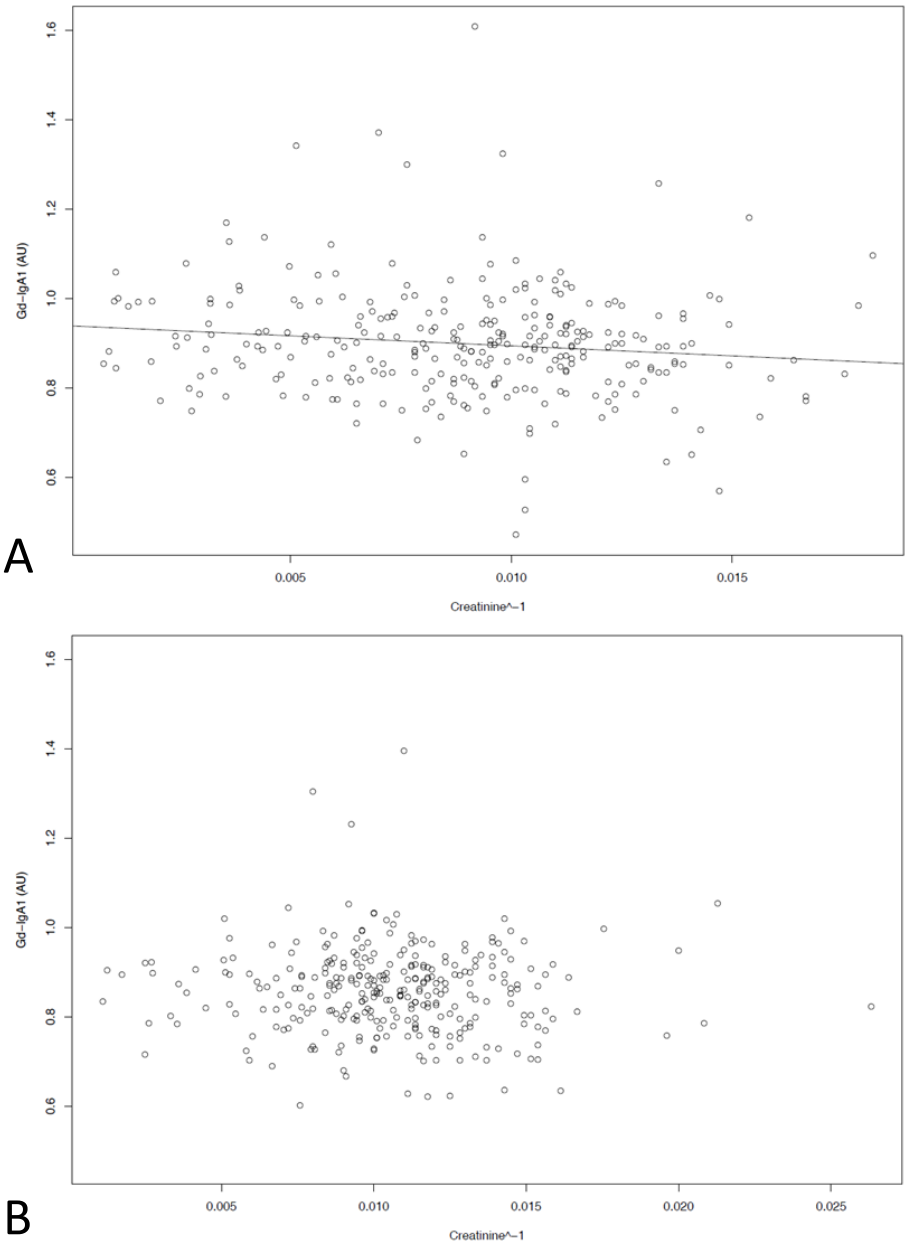
CHR	SNP1	SNP2	HAPLOTYPE	Freq.	BETA	STAT	P	Bonferroni
7	rs4720724	rs2190935	ATCTAT	0.207	0.186	7.27	0.00717	0.057
7	rs4720724	rs2190935	AGCTAC	0.17	-0.174	5.39	0.0205	0.164
7	rs4720724	rs2190935	AGTCAT	0.134	-0.168	3.67	0.0557	0.446
7	rs4720724	rs2190935	GGTCAT	0.0479	-0.059	0.182	0.669	1
7	rs4720724	rs2190935	AGTTAC	0.0429	-0.0926	0.473	0.492	1
7	rs4720724	rs2190935	ATCCGC	0.0154	0.151	0.424	0.515	1
7	rs4720724	rs2190935	GTCTAT	0.012	0.445	2.94	0.0871	0.697
7	rs4720724	rs2190935	GTCCAC	0.0113	0.449	2.95	0.0861	0.689

Supplementary Table 1 When haplotype analysis was performed conditioning on the presence of the H1 haplotype, no other haplotypes reached genome-wide significance when tested for association with Gd-IgA1 level. Bonferroni: Corrected for the 8 haplotypes tested.

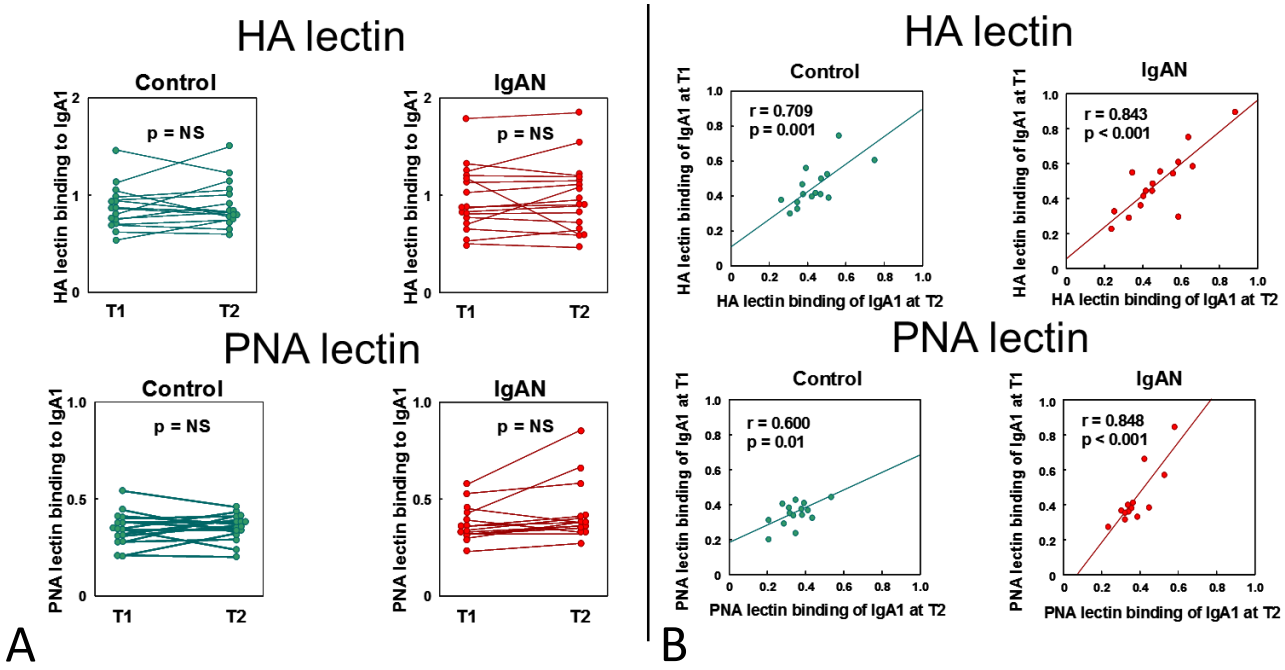
Supplementary Figures



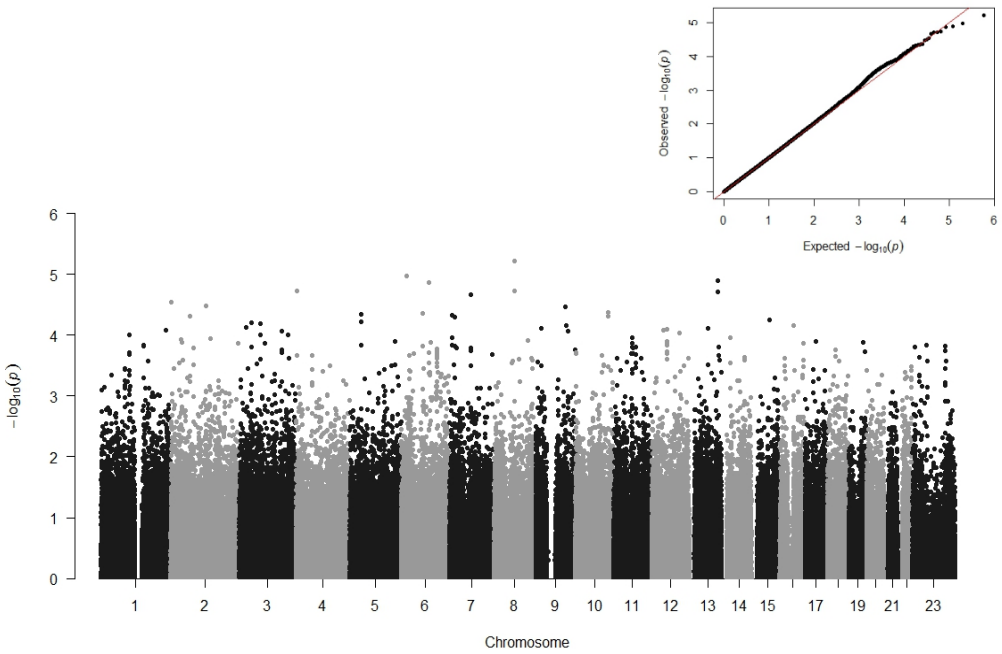
Supplementary Figure S1. Gd-IgA1 levels in the discovery cohort were normally distributed.



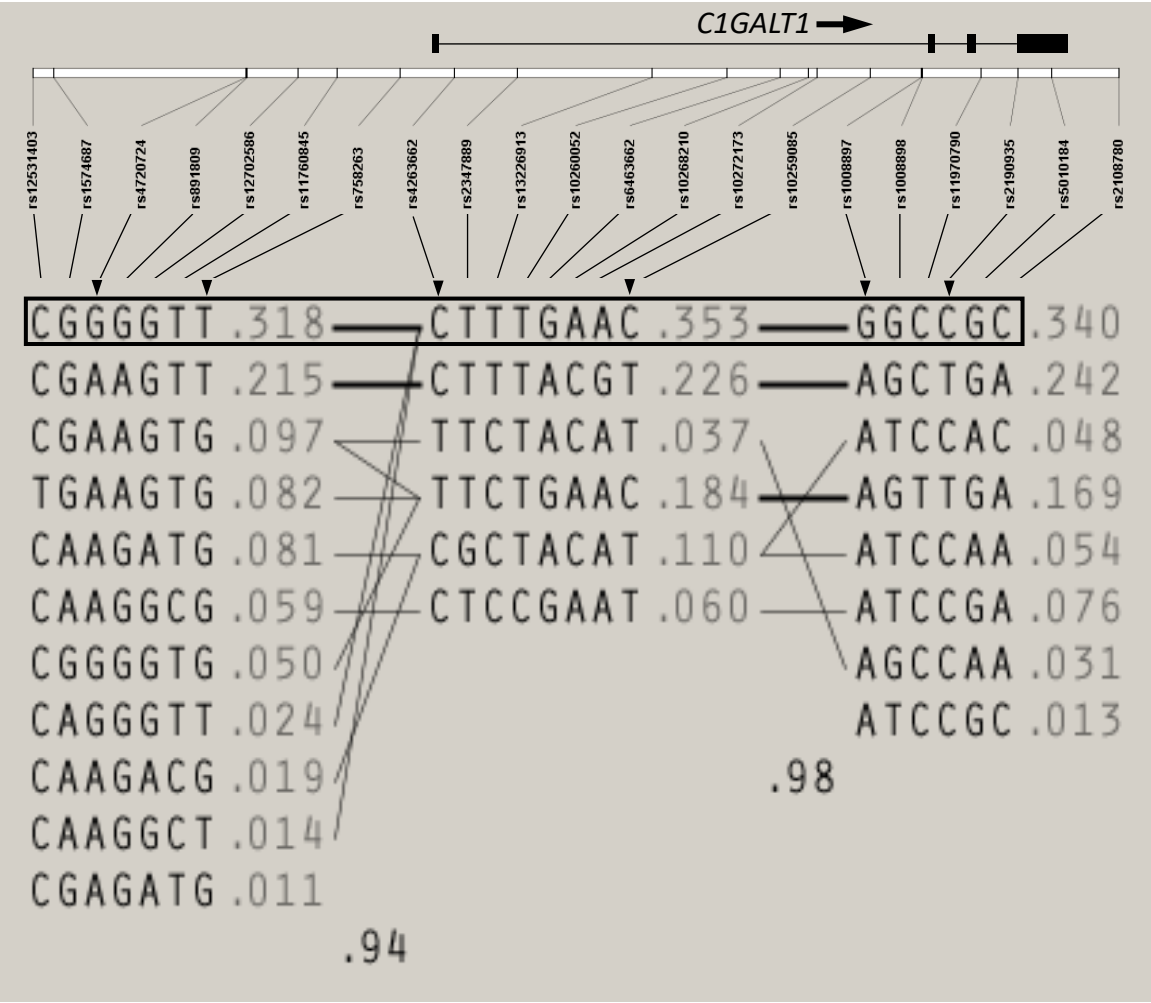
Supplementary Figure S2. A, Correlation was observed between Gd-IgA1 level and creatinine^{-1} in patients with IgAN (R^2 = 0.017, p = 0.031); B, Gd-IgA1 levels were not significantly correlated with creatinine^{-1} in 308 UK patients with membranous nephropathy, among whom >30% had eGFR <60 ml/min at time of blood sampling (R^2 = 0.0024, p = 0.42).



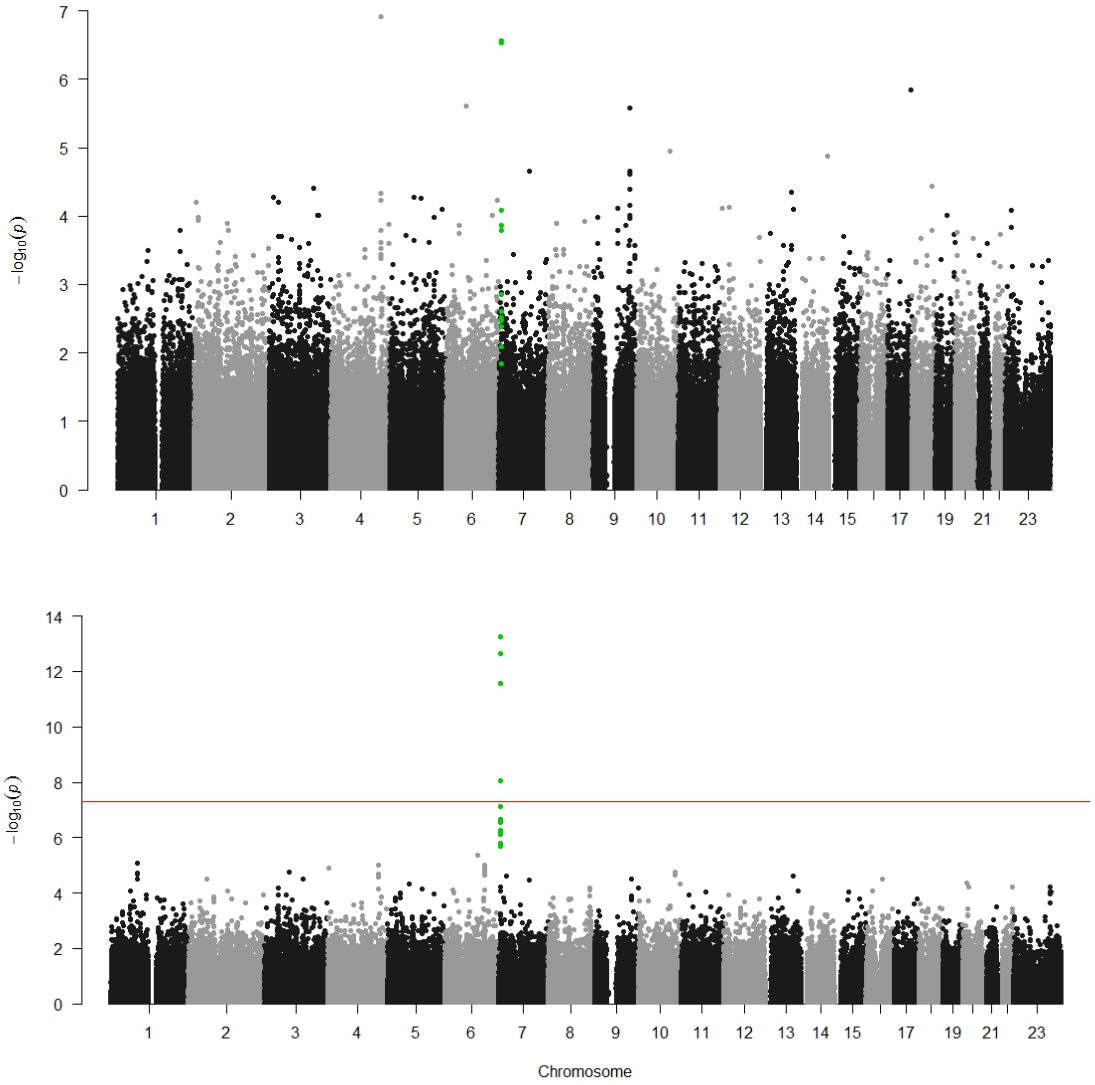
Supplementary Figure S3. Levels of Gd-IgA1 were measured in a cohort of 31 individuals (16 IgAN cases and 15 healthy subjects) using lectin (*Helix aspersa* (HA) and peanut agglutinin (PNA)) based ELISAs. Levels of Gd-IgA1 were measured at two time points (T1 and T2) to determine the stability of Gd-IgA1 levels within individuals over time and are expressed in arbitrary units (AU). The median time between T1 and T2 was 64 months, range 9 months to 10 years (IgAN) and 59 months, range 3 months to 19 years (healthy subjects). There was no significant change in the levels of Gd-IgA1 in either IgAN patients or healthy subjects over time (panel A) and there were significant correlations between T1 and T2 levels of Gd-IgA1 in both healthy subjects and IgAN patients (panel B).



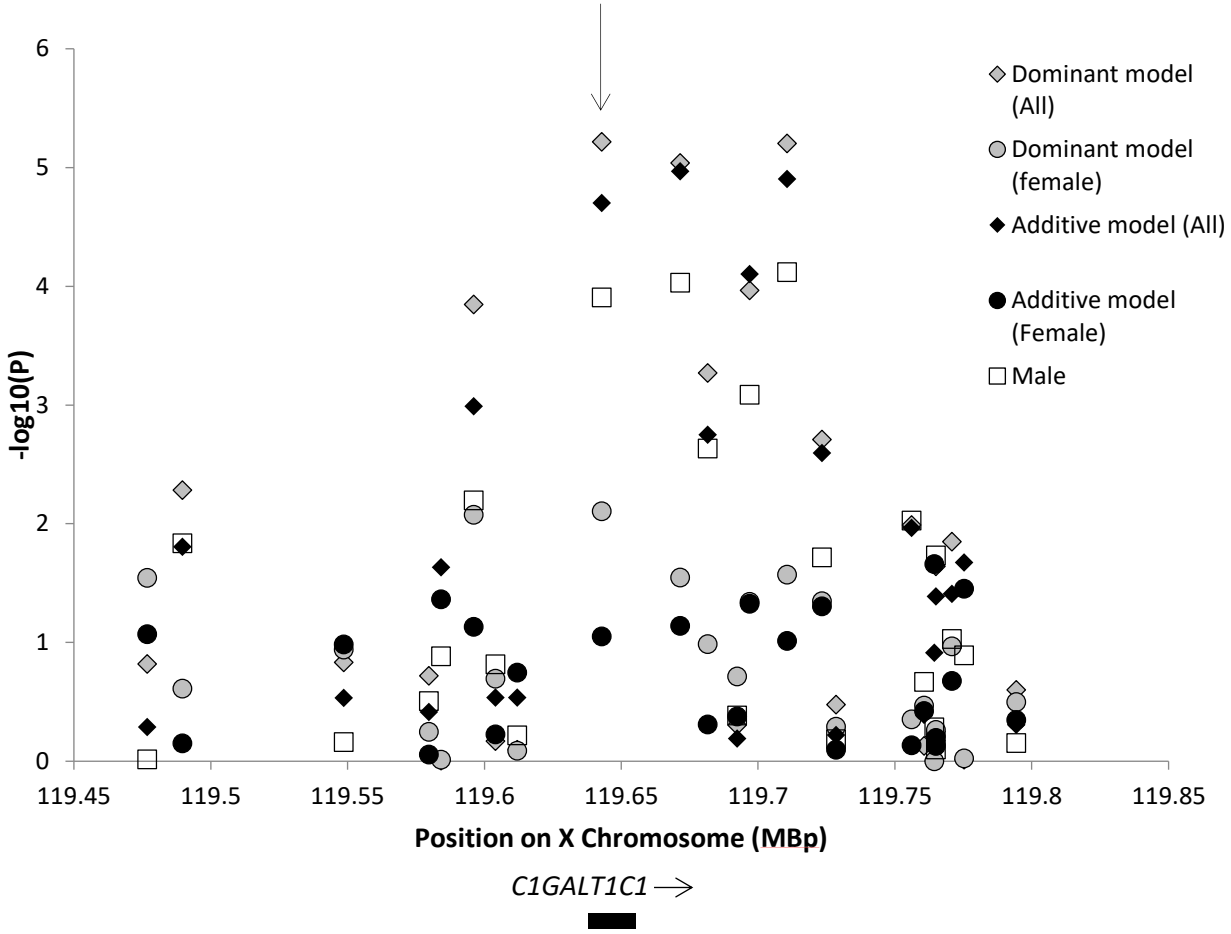
Supplementary figure S4. Manhattan plot showing genome wide association in discovery cohort conditioned on SNP rs1008897 showing no independent alleles at *C1GALT1*, or at other loci elsewhere, are significantly associated with Gd-IgA1 levels, with no deviation from p values expected under the null distribution (inset).



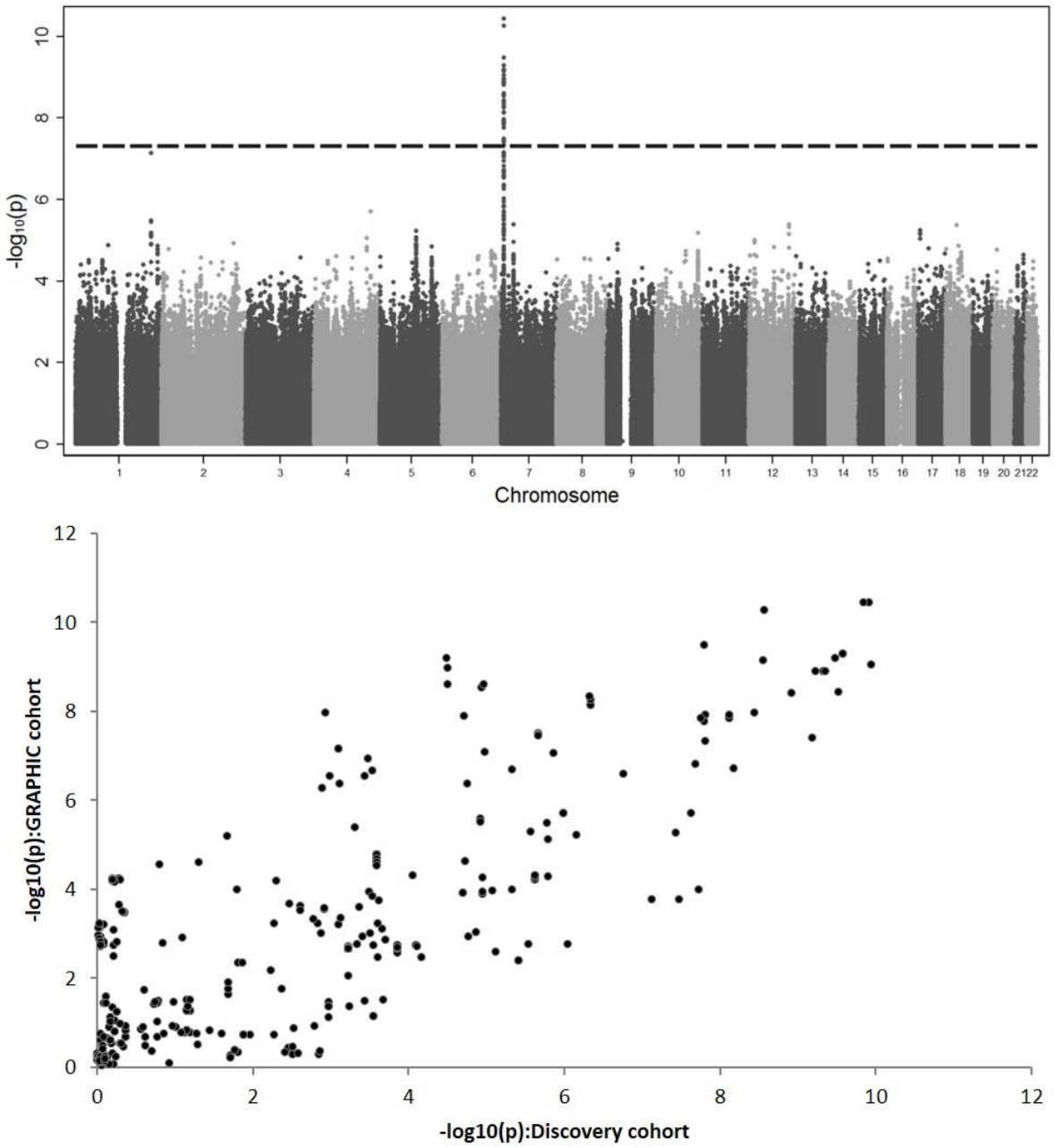
Supplementary figure S5. Haplotypes occurring across the *C1GALT1* locus in UK subjects. Alleles of neighbouring SNPs along a chromosome tend to be inherited together in haplotype blocks. The blocks occurring across the *C1GALT1* locus (position of the gene is shown at the top of the figure) in the Caucasian population are shown. The three blocks that span the gene and its upstream region are almost always inherited together as an extended haplotype (multiallelic D', which measures how closely correlated the blocks are with each other, of 0.94 and 0.98). The box indicates the 'H1' haplotype that occurs at a frequency of >0.3 in the Caucasian population. Inverted triangles show the SNPs also genotyped in Chinese patients that tag the H1 haplotype (see Table 2).



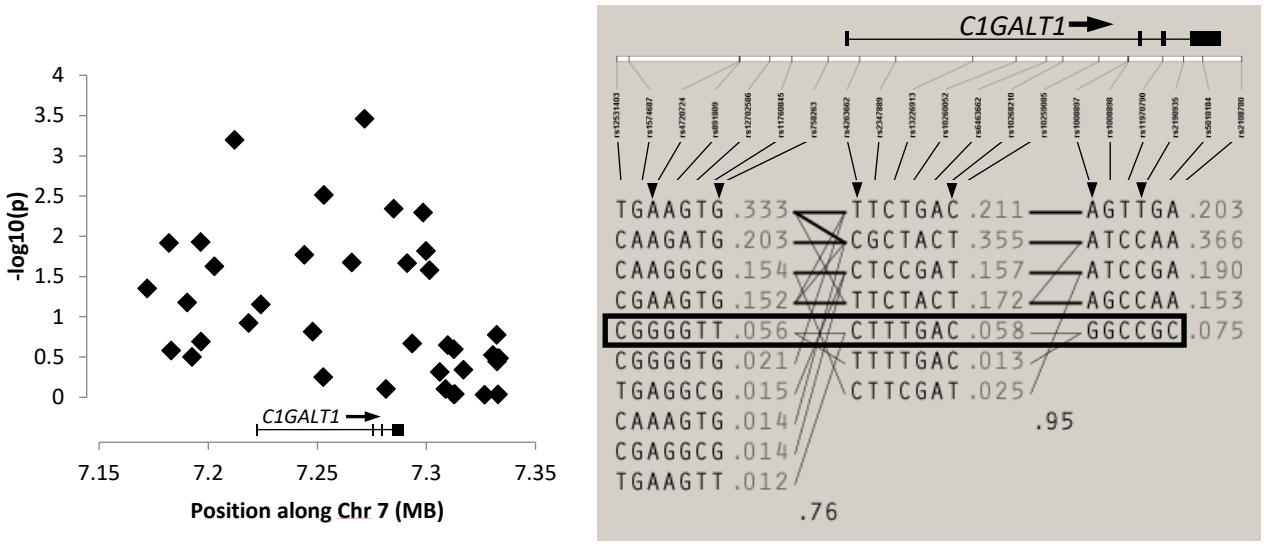
Supplementary figure S6. GWAS for Gd-IgA1 levels among 308 UK patients with membranous nephropathy (top panel) and in combined membranous nephropathy and IgAN cohorts (N=821, lower panel). SNPs that were strongly associated in the discovery cohort are highlighted in green. Genomic inflation (λ) was estimated to be 1.0068 in the combined analysis. No other loci contained alleles that approached genome wide significance (horizontal line).



Supplementary figure S7. Weak evidence of association was seen with alleles at the X chromosomal locus harboring *C1GALT1C1*, which encodes Cosmc. This association fell well short of genome-wide significance level whether analysed under additive or (in females) dominant models, and was not replicated in the GRAPHIC cohort, where p >0.01 for all SNPs in this region. The most strongly associated allele was rs5910948(T) (vertical arrow).



Supplementary Figure S8. Genome-wide association study of Gd-IgA1 level in 622 healthy Caucasians in the GRAPHIC cohort. The association observed in the discovery cohort with alleles across the *C1GALT1* gene and nowhere else in the genome is replicated, top panel. SNP-by-SNP comparison of the IgAN and GRAPHIC cohorts across the locus, lower panel, showing that the same alleles are associated in both cohorts.



Supplementary figure S9. Association between *C1GALT1* and Gd-IgA1 level in the Chinese population. The left panel shows the –log10(p) value for each SNP genotyped in the Chinese cohort. Right panel shows the haplotypes and their frequencies in the Chinese cohort. The H1 haplotype (boxed) is present but at substantially lower frequency than in Caucasians (see supplementary figure S5). SNPs tagging the H1 haplotype are marked with inverted triangles (see Table 2).