UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Setting the scale

The balance between complement factor H and its related proteins in health and disease

Pouw, R.B.

Publication date
2018
Document Version
Other version
License
Other

Link to publication

Citation for published version (APA):
Pouw, R. B. (2018). *Setting the scale: The balance between complement factor H and its related proteins in health and disease*. [Thesis, externally prepared, Universiteit van Amsterdam].

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:18 Nov 2026

7

Variation in the *CFHR3* gene determines susceptibility to meningococcal disease by controlling factor H levels

Vikrant Kumar[†], Richard B. Pouw[†], Matias I. Autio[†], Zai Yang Phua, Lisa Borghini, Victoria J. Wright, Bangfen Pan, Antson Kiat Yee Tan, Alexander Binder, Mieke C. Brouwer, Ellie Pinnock, Ronald de Groot, Jan Hazelzet, Marieke Emonts, Michiel van der Flier, Luregn J. Schlapbach, Evangelos Bellos, Suzanne Anderson, Fatou Secka, Federico Martinon-Torres, Antonio Salas, Colin Fink, Enitan D. Carrol, Andrew J. Pollard, Lachlan J. Coin, Werner Zenz, Diana Wouters, Martin L. Hibberd, Lay Teng Ang, Michael Levin[#], Taco W. Kuijpers[#], and Sonia Davila[#] on behalf of the EUCLIDS consortium

In preparation

[†] Equal contribution, [#] Equal contribution

Abstract

Previous studies associated susceptibility to meningococcal disease (MD) with variation in the gene region containing the *complement factor H* (*FH*) and five factor H-related protein genes (*CFHRs*) but the causal variant and underlying mechanism remain unknown. To identify the causal gene mutations controlling MD susceptibility we deep sequenced the *CFH-CFHR* region in 238 Caucasian MD patients and 237 controls and validated the associated SNPs in 1,522 MD patients and 2,672 controls. We measured FH and FHR-3 serum concentrations, and related genotype to concentration. We used dual-luciferase studies, embryonic stem cells differentiated into hepatocytes and genome editing to confirm genetic control of FH expression by *CFHR3.* Fine-mapping of the *CFH-CFHR* region identified a SNP in *CFHR3* (rs75703017) as the most significant association with susceptibility to MD (p=1.1x10^{-16}, OR=0.63, 95% CI 0.55-0.71). Blood levels of FH, but not FHR-3, were higher in MD survivors (p=0.001) and were associated with both *CFHR3* SNP genotype (p=1.41x10^{-11}) and copy number variation (CNV) of the *CFHR3* gene. Deletion of the *CFHR3* gene region containing the associated SNP by genome editing in human embryonic stem cells differentiated to hepatocytes increased FH expression (P<0.05). Risk of MD is determined by FH blood levels, which are regulated by a distal variant in the *CFHR3* gene rather than in the *CFH* gene. As complement inhibition by FH is used for immune evasion by several bacteria, viruses and parasites, and is involved in inflammatory diseases, genetic regulation of FH through *CFHR3* may be relevant to many diseases.

Introduction

Neisseria meningitidis meningitis and sepsis remain amongst the most feared bacterial infections world-wide. Although immunization has decreased the incidence of meningococcal disease (MD) in some countries, there are no vaccines effective against all serogroups, and the emergence of new serogroups and strains[1,2] poses new challenges to international vaccination strategies. Epidemics and outbreaks continue to occur in many countries particularly in the meningitis belt of sub-Saharan Africa[3–6].

A remarkable feature of *N. meningitidis* is that it is a harmless commensal in the majority of the world's population, being carried in the nasopharynx repeatedly throughout life. Invasive disease occurs in 2-20 per 100,000 population in developed countries, but there is wide variation in incidence, and MD affects a higher proportion of the population during epidemics[1,2,7].

There is good evidence that genetic factors play a role in MD[8–10]. Rare Mendelian defects in complement pathway genes are associated with familial MD[11,12]. Our previous genome-wide association study (GWAS) identified an association between MD and a broad genomic region spanning *Complement Factor H* (*CFH*) and the *Complement Factor H-Related protein* (*CFHR*) genes[8]. Identification of the causal gene, and characterization of the functional variant(s) has been difficult due to the complexity of the region as *CFH* shows sequence similarity to five adjacent *CFHR* genes on chromosome 1[13].

Factor H (FH) is a serum glycoprotein, synthesized mostly in the liver, which acts as a negative regulator of alternative complement pathway activation[14]. FH is a crucial factor in preventing host cell damage by uncontrolled complement activation[15], and genetic variation in *CFH* or *CFHR* genes are associated with several diseases including systemic lupus erythematosus[16], glomerulonephritis[17], IgA nephropathy[18], atypical hemolytic uremic syndrome[19] and age-related macular degeneration[20].

N. meningitidis is highly sensitive to complement-mediated killing, and binding of *N. meningitidis* FH-binding protein (fHbp) to FH facilitates evasion of complement-mediated killing[21]. Host variation in the *CFH-CFHR* region may influence the success of this "Trojan horse" evasion by *N. meningitidis*[8]. Furthermore, inhibition of complement by "hijacking" FH has been adopted as an immune evasion strategy by several bacteria[22–26] as well as plasmodia[27]. We aimed to fine-map the *CFH-CFHR* region and identify the causal variants and mechanisms underlying the association with MD.

Materials and Methods

Study sample sets

The strategy for our study is shown in **Figure 1**. Clinical details of patients with MD in UK, Spanish and Western European cohorts have been reported previously, as have the diagnostic criteria, recruitment procedure, and ethical approvals[8]. In brief meningococcal disease was diagnosed based on isolation of *N. meningitidis* from blood or CSF, or detection of meningococcal genome by PCR in children presenting with purpuric rash, meningitis or septic shock. Cases without microbiological confirmation were included if they presented with the characteristic picture of purpura fulminans, and other bacterial or viral causes of critical illness had been excluded. Samples for DNA extraction (EDTA blood) and plasma (recovered from EDTA Blood) and serum both at the time of admission and after recovery were stored at -80°C until used. Informed consent was obtained from parents or guardians. The study was approved by the research ethics bodies of each collaborating partner (UK EC3263; Netherlands: 37986.091.11 / RvB12.51320; Austria: 24-116 ex 11/12; Spain: 2011/298). 238 MD patients and 237 controls from the Western European cohort were used to deep-sequence the *CFH-CFHR* region. Replication was undertaken in 755 cases and 1,253 controls from the UK, 279 cases and 395 controls from Western Europe, and 488 cases and 1,024 controls from Spain[8]. Convalescent serum was available from 367 patients and 124 healthy unrelated controls for measurement of FH and FHR-3 levels, and 295 UK patients and 56, healthy unrelated controls from Western Europe were used in protein quantitative trait loci (pQTL) analysis.

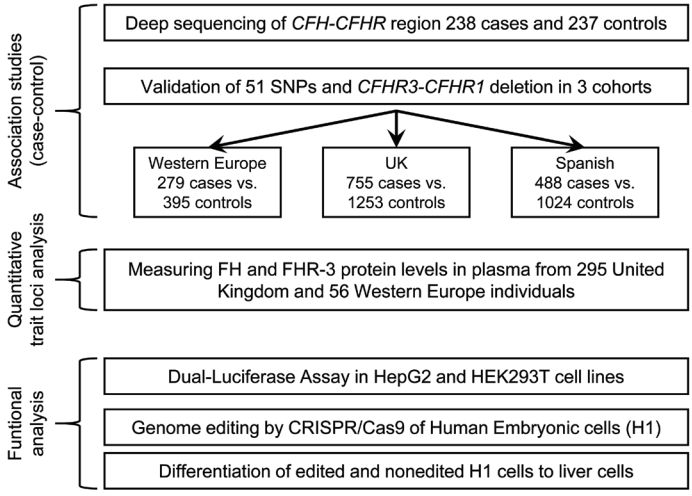


Figure 1 Schematic representation of study design.

Study design

We followed a multiple-stage study design to conduct a genetic association study (**Fig. 1**). In stage 1, we deep-sequenced the *CFH-CFHR* region in 475 individuals comprising of 238 MD patients and 237 controls from a Western European cohort. Heterogeneity scores indicated that there was no significant population admixture between cases and controls. Replication analyses were carried out on 4,194 individuals which includes three cohorts from the UK (755 cases, 1,253 controls), Western Europe (279 cases, 395 controls) and Spain (488 cases, 1,024 controls) that have been previously described[1].

In addition, quantitative trait loci (QTL) analysis was performed in 295 UK samples and replicated in a cohort of 56 Western Europe Individuals. Serum concentrations of FH and FHR-3 were measured in 295 UK and 59 Dutch survivors of MD (at least 6 months after the acute illness) and 124 Caucasian unrelated controls.

Sequencing and genotyping of the *CFH-CFHR* region

The *CFH* region spanning ~360 kb in chr1 (chr1:196722205-196808506, hg19) consists of *CFH* and the *CFHR* genes which show high sequence identity and are a great challenge to sequence. With Nimblegen SeqCap EZ choice library we made probes to capture ~80% of the whole region and performed targeted resequencing (**Supplemental Fig. 1a**). Deep sequencing was carried out on Illumina HiSeq 2000 using 100bp paired-end read sequencing protocol. Image analysis and base calling was done by Illumina pipeline with default parameters. Sequence reads were mapped to the reference genome (hg19) using the ELAND module of CASAVA v1.8.2 program. Multi-mapped reads were not considered for further analysis and duplicate reads were identified by Picard (http://picard.sourceforge.net/) and removed. This was followed by local realignment and base quality score recalibration of the reads and finally variant calling by GATK 3.3. The mean depth of coverage for all samples was estimated to be ~227X ± 47, with ~95% of the region having 5X or more depth of coverage. The average depth of sequencing was 227x (**Supplemental Fig. 1**). The most significant SNPs were validated using a Sequenom Multiplex MassArray (San Diego, USA).

Single nucleotide polymorphisms (SNPs) were called for all Western European samples together using the Unified Genotyper module of GATK 3.3. To obtain high quality variants SNPs with i) a quality score < 50 or ii) QD < 2 or iii) FS > 60 or iv) MQ < 40 or v) haplotype score > 13 or vi) MQRankSum < -12.5 or vii) ReadPosRankSum < -8 were removed. The top 51 SNPs were selected for validation in three cohorts of European descent using a Sequenom Multiplex MassArray (San Diego, USA). Six SNPs failed to be designed due to the mentioned complexity of the region. We also genotyped these SNPs on 295 UK individuals and 56 Western Europeans (mentioned above) included on a quantitative trait analysis.

Detection of copy number variation

A novel method, cnvCapSeq[2] that can detect copy number variation (CNV) in long-range targeted resequencing data, was applied to call CNVs in our sequencing effort. CNV results were validated using quantitative PCR (qPCR) on a randomly chosen subset of samples (n=67), in triplicates. To confirm that the primers (**Supplemental Table 1**) were targeting the intended region, we performed PCR and Sanger sequencing in one sample. As internal control a 5th set of primers for *PRKG1* (a house keeping gene) were designed. Following qPCR, copy number estimates were obtained using the ΔΔCt method of relative quantification. In the second stage of CNV validation all three European cohorts were assessed by Taqman assays (1,463 controls and 1,302 cases). To call CNVs we used CopyCaller v2.0 program and a known reference with two copies. A subset of samples (~5%) was validated using qPCR. We also typed additional 295 UK samples using qPCR (described above) and 56 Western European samples using qPCR and Multiplex Ligation-dependent Probe Amplification (MLPA) (**Supplemental Table 2**). The MLPA kit SALSA MLPA probemix P236-A3 ARMD mix-1 was used to genotype the *CFH-CFHR* gene region according to the manufacturer's instructions (MRC-Holland, Amsterdam, the Netherlands).

Serum concentrations of FH and FHR-3

FH and FHR-3 levels in serum were determined by highly specific ELISA's previously described[3]. In short, FH was measured by ELISA using an in-house developed, highly specific mouse monoclonal antibody (mAb) directed against SCR domains 16/17 of FH (clone anti-FH.16, Sanquin Research, Amsterdam, The Netherlands), as the capture mAb and using polyclonal goat anti-human FH (Quidel, San Diego, CA, USA), which was HRP-conjugated, as the conjugate. For the specific detection of FHR-3, an in-house developed mAb directed against FHR-3 and FHR-4A was used as the capture antibody. A biotinylated mAb directed against FHR-3 and FH was used as conjugate.

Identification of regulatory elements

Using haploreg v3 (19), we annotated the liver specific regulatory element in the *CFH-CFHR* region. To further confirm it Road Map ChIP-seq data showing active promoter region predicted only in the E066 liver cell-line by H3K4me3 and H3K9ac (in duplicates) were analyzed (http://epigenomegateway.wustl.edu/browser/). To understand the plausible mechanism by which the regulator elements govern the gene expression of *CFH*, we analyzed UCSD Hi-C H1 and IMR90 data (http://epigenomegateway.wustl.edu/browser/) with 20 kilobases resolutions (in duplicates).

Luciferase assay

For plasmid construction (primers listed in **Supplemental Table 3**), the putative 2.8 kb regulatory sequence was amplified from H1 hES genomic DNA for the wildtype (pGL3-WT)

using luci 5' F and 3' R primers. Regulatory sequences containing the lead SNP rs75703017 (pGL3-M) and the lead SNP and the two nearest SNPs identified (rs446868 and rs385390) to the lead SNP (pGL3-3M) were amplified from gBlocks designed with the specified SNPs and H1 gDNA (luci 5' F, 5' R, 3' F and 3' R primers). The amplified fragments were ligated to the SalI cut site of pGL3-Promoter (Promega, E1761) plasmid. We chose rs446868 and rs385390 to determine if we could observe additional regulatory effect or independent effect (if lead SNP does not show any regulatory effect). These two SNPs were in perfect LD (based on 1000 genome data; http://archive.broadinstitute.org/mammals/haploreg/haploreg.php), most proximal (one upstream in the 5'-UTR region and one before that) with the lead SNP.

The luciferase assays were conducted with Dual-Glo® Luciferase Assay System (Promega, E2920) using the GloMax®-Multi+ Detection system according to manufacturer's instructions. The pGL3-Promoter constructs were transfected into HepG2 and HEK293T cells with Lipofectamine 3000 (ThermoFisher, L3000015), according to manufacturer's instructions. Briefly, 40,000 and 25,000 cells per well of HepG2 and HEK293T respectively were seeded in 96-well plates. After 24h the cells were transfected with 0.5 μL of lipofectamine and 100 ng of each pGL3-Promoter plasmid and 10 ng of pRL-SV40 (renilla) plasmid per well. Luciferase analysis was done 48h after transfection. At least three replicate wells were analysed for each test condition and the experiment was repeated three times (twice for HEK293).

Each experiment, measured in triplicates or more, was normalized by dividing the test reporter (Firefly luciferase) activity by the internal control reporter (Renilla luciferase) activity. The normalized values were then averaged. To compare the estimates between different experiments we further normalized fold-change between different construct/test groups by dividing the average value of a construct with the average value of another construct. To determine the significance of fold–change we calculated P-value using single sample t-test.

Human embryonic stem cells differentiation

Wild type and CRISPR/Cas targeted H1 hES cells were maintained on Geltrex coated plates (Thermo Fisher, A1413302) in mTeSR1 or TeSR E8 (Stemcell technologies, #05850 &#05940). Cells were differentiated to a hepatic progenitor phenotype over 12 days as previously described[4] and further differentiated to a more mature hepatic cell type until day 18 (Ang LT; manuscript under revision).

Genome editing of differentiated hepatocytes by CRISPR/Cas9

Guide RNAs flanking the liver specific regulatory region were designed using the online CRISPR design tool at http://crispr.mit.edu/ and the gRNAs with the best available on target and off target scores were selected after *in vitro* testing (**Supplemental Table 4**). The guide RNAs selected for targeting experiments were used with 2sg-SpCas92AClover and 4sg-

nSpCas92AClover plasmids and electroporated into H1 hES cells with Amaxa 4D nucleofector (Lonza). Following the treatment cells were plated out for two days to recover. After this the cells underwent FACS for Clover+ cells to enrich for CRISPR/Cas targeted cells. Clover+ cells were seeded at a low density (500 -1,000 cells per one 6-well). After 2-3 weeks of culture single colonies were picked from the wells and expanded for screening and frozen for stock. Deletion of liver specific regulatory region was determined by PCR with primers spanning the targeted region (**Supplemental Table 5**). Confirmed deletion clones and wild type controls were used in downstream analysis.

gRNA testing

The designed gRNAs were tested using a GFP reconstitution assay[7]. Briefly, the liver specific regulatory region was amplified from H1 gDNA and inserted into SalI cut site of pCAG-EGxxFP plasmid (Addgene plasmid # 50716), a gift from Masahito Ikawa. Designed gRNAs were cloned into the BbsI cut sites of px330 plasmid. The target plasmid and each of the tested gRNA plasmids were transfected into HEK293T cells at a 1:1 ratio using 5 µg Polyethylenimine per 1µg of DNA. The expression of GFP was observed after 48h and the most efficient gRNAs were selected for targeting of hES.

RT-qPCR

Total RNAs were extracted from undifferentiated and liver differentiated cells with the RNeasy kit (Qiagen) according to manufacturer's instruction. RNA integrity and quantity were checked by running the samples on a Bioanalyzer RNA 6000 Nano chip (Agilent). One microgram of total RNAs were used for reverse transcription using QuantiTect Reverse Transcription kit (Qiagen) followed by qPCR with LightCycler 480 SYBR Green I Master (Roche) according to the manufacturers' instructions. The following PCR program was used: 5 minutes at 95°C; 45x (10 seconds at 95°C, 1 minute at 65°C, 30 seconds at 72°C). The ΔΔCt method was applied to determine the fold change in expression of the genes of interest, by normalizing with the housekeeping gene *GADPH*. Liver markers were tested on liver differentiated cells versus undifferentiated cells. *CFH* expression was investigated in *CFHR3*-deleted cells versus wild type. Sequences of the primers used for qPCR are shown in **Supplemental Table 5**. To determine the significance of fold–change we calculated P-value using single sample t-test.

Statistical analysis

Genetic association testing was carried out using Fisher's exact test for rare SNPs (Minor allele frequencies <1%) and logistic regression analysis for common SNPs and CNV under an additive genetic model. To mitigate the effect of population stratification, association of SNPs with MD was analyzed separately in all three replication cohorts under the additive model, and meta-analysis for both SNPs and CNV was performed by combining summary statistics

of stage 1 (deep sequencing) and stage 2 (sequenom validation) using the Cochran-Mantel-Haenszel (CMH) test. For CNV all samples were combined and analyzed under a genotypic and additive model.

Genotype-phenotype correlation of SNPs and CNV was analyzed for FH and FHR-3 protein level using linear regression analysis. We used ANCOVA, with sex as a covariate, to estimate the overall difference in the protein levels across six genotype groups. Differences in protein levels between two genotype groups were evaluated by t-test.

Results

Fine-mapping of the *CFH-CFHR* region identifies *CFHR3* as lead association

Deep sequencing of the *CFH-CFHR* region in 238 MD cases and 237 healthy controls identified 4,369 SNPs after applying stringent quality control filters (**Supplemental Table 6**). The strongest signal was identified on *CFHR3* in a region with high linkage disequilibrium (LD, D'=0.92) with the previously reported lead variant, rs1065489 in *CFH*[8–10]. The 51 SNPs with the strongest association with MD were selected for validation and 45 SNPs were successfully typed in the UK, Spanish and Western European cohorts (n=4,194, **Supplemental Table 7**). Thirteen SNPs, in a tight LD block within *CFHR3*, achieved genome-wide significance in the meta-analysis (**Fig. 2a**, **Table 1**) confirming the genetic association with *CFHR3*. The lead SNP (Table 1), rs75703017 (p=1.1x10^{-16}) located in intron 1 of *CFHR3*, showed consistent odds ratios (OR=0.62) for susceptibility to MD in all cohorts, indicating a protective effect (**Fig. 2b**).

Serum concentrations of FH, but not FHR-3, are higher in MD than in controls

To explore the relationship between serum concentrations of FH, FHR-3 and MD, we measured concentrations in serum from survivors of MD at least 6-24 months after the acute illness. Serum concentrations of FH were significantly increased in MD survivors as compared with healthy controls (**Fig. 3a**). In contrast, FHR-3 serum concentrations were not significantly different between MD and controls (**Fig. 3a**).

Serum concentrations of FH are associated with both SNP and CNV in *CFHR3*

To investigate the effect of the top associated SNPs on the serum levels of FH and FHR-3 we undertook pQTL, relating protein levels and genotype. The minor allele (A) of the lead SNP (rs75703017), shown to confer protection against MD (OR=0.63), was associated with lower FH serum levels (p=1.4x10^{-11}, **Fig. 3b**), as confirmed in a pairwise comparison between genotypes carrying the minor allele (A) of rs75703017 (**Fig. 3b** and **3c**, **Supplemental Table 8**).

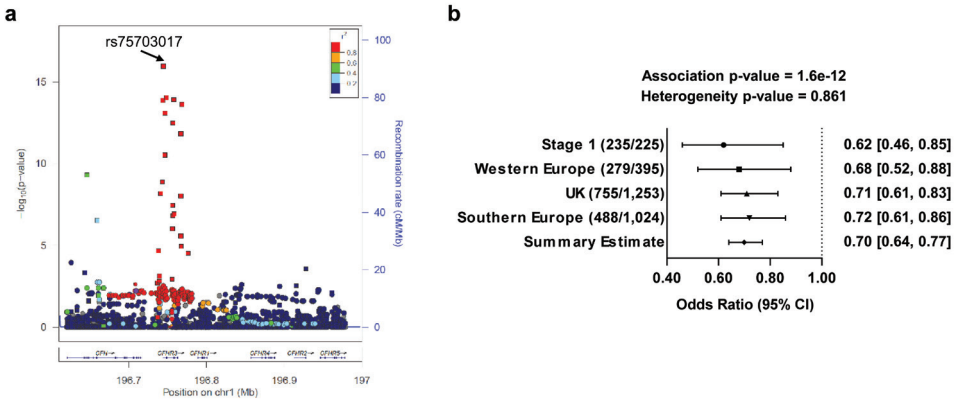


Figure 2 Fine-mapping of the *CFH-CFHR* gene locus. (**a**) Plot showing association results of all the SNPs (arranged according to their hg19 build chromosomal position on x axis) from deep sequencing (circle) and from meta-analysis (square) using combined stage 1 and stage 2 cohorts. The top SNP from the analysis is labeled (rs75703017). The color intensity of each symbol reflects the extent of LD with the top GWAS SNP[8]. Genetic recombination rates (cM/Mb) estimated using 1000genomes samples, are depicted by a blue line. (**b**) Forest plot of the top SNP rs75703017 showing odds ratios in each cohort (squares) with the summary measure (center of diamond) and associated confidence intervals (lateral tips of diamond), and dotted vertical line of no effect. Names of cohort are on the left along with number of cases/controls, odds ratios and confidence intervals on the right.

A common deletion spanning the *CFHR3* and *CFHR1* genes (including rs75703017), has been shown to influence susceptibility to several inflammatory diseases[16,18,20,31]. To establish whether this deletion was also associated with susceptibility to MD, we determined CNVs in the sequenced patients using cnvcapSeq[28], which permits detection of CNVs in long-range targeted sequencing data. We then validated the findings by MLPA or qPCR in a subset of samples (1,302 MD cases, 1,463 controls) from three European cohorts. Meta-analysis of CNV data revealed an overall lack of association (p=0.76) between the *CFHR3/CFHR1* deletion and susceptibility to MD (**Supplemental Table 9**), as previously reported[8,9]. Considering that the pQTL data indicated a dominant effect of the minor allele of rs75703017, we performed a second comparison consisting of minor allele carriers (A) vs. *CFHR3* deletion allele carriers. In contrast to the overall lack of association, this second comparison revealed that deletion of *CFHR3* was in fact associated with higher genetic risk of MD (p=0.0081, **Supplemental Table 10**) and increased FH serum levels. This positive genetic association with *CFHR3* deletion was only detected when the combination of three alleles (wild type C, minor allele A and deletion D) in rs75703017 were taken into account. Previous studies[8,9] have missed this effect due to combination of those with the minor allele (A), proven here to have a protective effect, with those with the wild type allele (C), shown here to have a risk effect. Meta-analysis of quantitative trait association removing all samples with the *CFHR3* deletion did not modify the results (**Supplemental Fig. 2**), indicating that the association of rs75703017 persists regardless of *CFHR3* deletion status (**Supplemental Table 11**).

Table 1 Unadjusted P-values and ORs of top SNPs in the combined analysis

	Sequence				Replication												Combined CMH*	
	Western European				Western European				UK				Spain					
SNP ID	MAF Cases	MAF Ctrls	P-value	OR	MAF Cases	MAF Ctrls	P-value	OR	MAF Cases	MAF Ctrls	P-value	OR	MAF Cases	MAF Ctrls	P-value	OR	P-value	OR
rs75703017	0.12	0.21	$3.11x10^{-3}$	0.62	0.14	0.21	$3.60x10^{-3}$	0.68	0.14	0.21	$8.34X10^{-6}$	0.71	0.18	0.26	$2.26x10^{-4}$	0.72	$1.11x10^{-16}$	0.62
rs620015	0.13	0.21	$6.18x10^{-3}$	0.65	0.14	0.22	$3.62x10^{-3}$	0.68	0.15	0.21	$2.27X10^{-5}$	0.73	0.2	0.26	$1.01x10^{-3}$	0.75	$9.55x10^{-15}$	0.64
rs387107	0.14	0.21	$9.35x10^{-3}$	0.67	0.14	0.21	$4.49x10^{-3}$	0.68	0.15	0.21	$4.97X10^{-5}$	0.74	0.2	0.27	$3.11x10^{-4}$	0.73	$1.24x10^{-14}$	0.65
rs385390	0.14	0.22	$6.04x10^{-3}$	0.65	0.15	0.22	$9.74x10^{-3}$	0.71	0.15	0.21	$1.50X10^{-5}$	0.72	0.2	0.27	$7.11x10^{-4}$	0.75	$1.35x10^{-14}$	0.65
rs12409571	0.13	0.21	$5.04x10^{-3}$	0.64	0.14	0.21	$8.20x10^{-3}$	0.7	0.15	0.2	$1.10X10^{-4}$	0.75	0.19	0.26	$2.46x10^{-4}$	0.73	$2.38x10^{-14}$	0.65
rs425524	0.13	0.21	$7.46x10^{-3}$	0.65	0.36	0.45	$4.97x10^{-3}$	0.73	0.31	0.42	$2.63X10^{-9}$	0.66	0.46	0.5	$9.96x10^{-2}$	0.87	$8.26x10^{-14}$	0.69
rs401188	0.13	0.21	$5.98x10^{-3}$	0.65	0.22	0.25	$3.36x10^{-1}$	0.87	0.2	0.34	$1.14X10^{-13}$	0.46	0.34	0.38	$3.22x10^{-2}$	0.76	$3.32x10^{-13}$	0.64
rs1738741	0.14	0.21	$9.35x10^{-3}$	0.67	0.14	0.21	$1.06x10^{-2}$	0.71	0.15	0.21	$1.39X10^{-5}$	0.72	0.22	0.27	$1.43x10^{-2}$	0.81	$1.54x10^{-12}$	0.67
rs376841	0.13	0.21	$4.42x10^{-3}$	0.64	0.1	0.16	$1.55x10^{-3}$	0.56	0.1	0.15	$1.29X10^{-5}$	0.63	0.14	0.18	$4.74x10^{-4}$	0.65	$3.06x10^{-11}$	0.66
rs1329423	0.19	0.27	$4.10x10^{-3}$	0.63	0.21	0.26	$3.43x10^{-2}$	0.75	0.21	0.25	$9.72X10^{-4}$	0.78	0.23	0.3	$7.31x10^{-5}$	0.7	$5.07x10^{-10}$	0.73
rs11807997	0.13	0.21	$5.12x10^{-3}$	0.64	0.19	0.18	$8.97x10^{-1}$	1.02	0.15	0.21	$1.24X10^{-4}$	0.74	0.19	0.25	$1.12x10^{-3}$	0.75	$1.34x10^{-9}$	0.71
rs12408446	0.13	0.2	$7.96x10^{-3}$	0.65	0.2	0.18	$5.19x10^{-1}$	1.08	0.15	0.21	$8.96X10^{-5}$	0.74	0.19	0.25	$1.36x10^{-3}$	0.76	$6.86x10^{-9}$	0.72
rs116249058	0.14	0.21	$7.53x10^{-3}$	0.66	0.12	0.14	$2.54x10^{-1}$	0.83	0.12	0.15	$1.74X10^{-2}$	0.81	0.14	0.21	$8.85x10^{-5}$	0.67	$9.59x10^{-9}$	0.7

* Cochran-Mantel-Haenszel test.
MAF = minor allele frequency
Ctrls = Controls

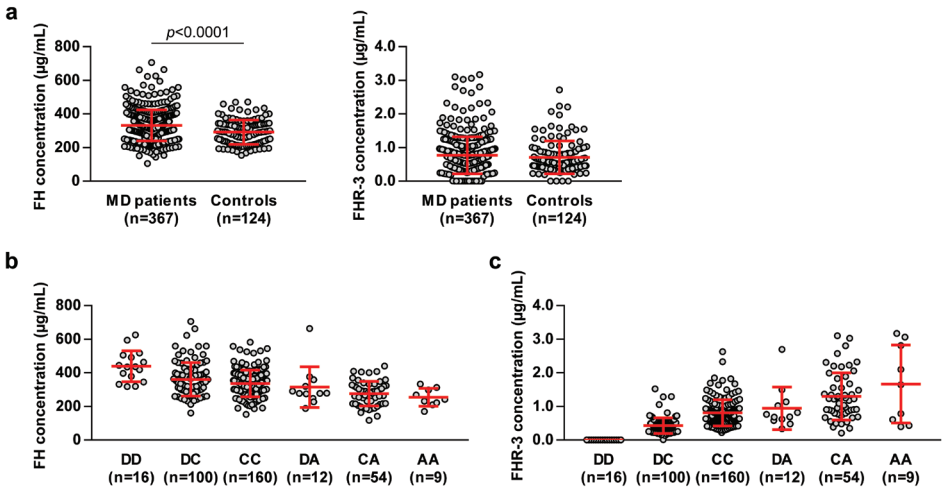


Figure 3 FH and FHR-3 levels in MD survivors. (**a**) FH serum levels (left panel), determined by ELISA, are increased in former MD patients (n=367) compared to healthy controls (n=124) whereas FHR-3 serum levels (right panel), determined by ELISA, are not significantly different between both groups. (**b**) Plots for the genotypes of the top associated SNP (rs75703017) of FH levels with P-value of 1.41×10^{-11}. (**c**) Plots of FHR-3 levels with P-value $<2 \times 10^{-16}$. P-values were estimated by ANCOVA. Each dot represent one individual, indicated with mean and SD (error bars), 'D' = *CFHR3/CFHR1* deletion, 'A' = mutant allele and 'C' = wild type allele.

CFHR3 controls *CFH* expression through epigenetic long range interaction

Having established the correlation between "protective genotypes" and lower serum levels of FH, and "risk genotypes" with higher levels of FH, we investigated the epigenetic histone marks in various cell lines to provide information on the putative regulatory role of the potential functional SNP in *CFHR3*. Histone marks (H3K4me3 and H3K9ac) from the Roadmap epigenomics database indicated that all investigated hepatic cell lines have an active regulatory site within *CFHR3*. Furthermore, no other cell types (non-hepatic) tested showed any indication of regulatory regions, suggesting that this functional site might be specifically active in liver cells[32,33], which is concordant with the main FH-producing organ being the liver[13]. In line with our hypothesis that there is a regulatory interaction between *CFHR3* and *CFH*, we examined whether the homozygous deletion of *CFHR3*, carried by 3% of the European population, affected FH protein levels. Indeed, the absence of *CFHR3*, identified by the lack of FHR-3 in serum, was associated with increased genetic susceptibility to MD and significantly higher FH protein levels (**Fig. 3b**). Interestingly, this deletion is far more common in African populations affected by MD outbreaks[34].

Dual-luciferase assays confirm liver specific activity

To confirm the role of the *CFHR3*-mutant allele identified in our fine-mapping in regulating *CFH* gene activation, we compared luciferase activity of a liver cell line (HepG2) versus a line

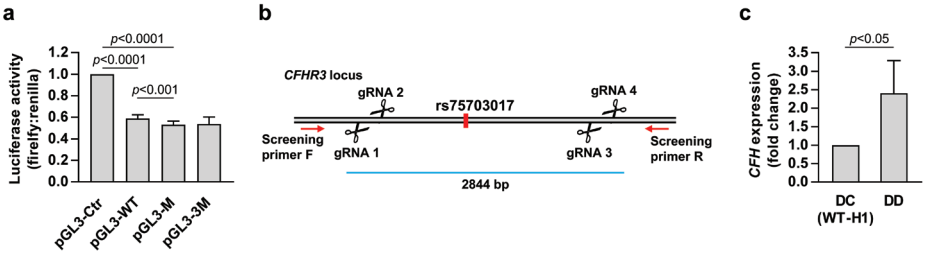


Figure 4 Functional validation of the top associated variant. (**a**) Effect of 2.8 kb regulatory sequence and the lead SNP rs75703017 on reporter (firefly luciferase) activity in HepG2 Human liver cell line. pGL3-Ctr is an empty construct only containing a promoter, while pGL3-WT contains the 2.8 kb regulatory sequence with allele C at rs75703017, pGL3-M contains allele A at rs75703017, and pGL3-3M contains minor alleles at rs75703017, rs446868 and rs385390. Firefly luciferase levels were normalized to renilla luciferase activity for each sample, and all values were plotted relative to the pGL3-Ctr (empty) construct. The graph is representative of six independent experiments. Level of significance, calculated by t-test, is indicated. (**b**) Schematic of CRISPR/Cas9 targeting of liver specific regulatory region of *CFHR3* comprising of 2,844 base pairs (bp). Excision sites of each guide RNA, located around the SNP of interest (rs75703017) are indicated by scissors. Position of screening primers designed to select for positive clones with excision of targeted region are indicated by red arrows. (**c**) Fold change of *CFH* transcript expression levels of CRISPR edited *CFHR3* (CRISPR edited-H1; genotype DD) relative to wild type, carrying one copy of *CFHR3* with allele C (WT-H1; genotype DC), liver differentiated H1 Human Embryonic Stem cells. Expression was measured by qRT-PCR. The graph represents three independent experiments with two biological replicates (having different sets of gRNA used for targeting; KO1 WT-Cas9 gRNA 1 & 3 and KO2 nickase-Cas9 gRNA 1, 2, 3 & 4; see **Supplementary Table S2**) and one technical replicate of KO1. Level of significance, calculated by t-test, is indicated, 'D' = *CFHR3*/*CFHR1* deletion and 'C' = wild type allele.

originating from embryonic kidney (HEK293T). We compared cells containing an empty vector, (pGL3-Ctr), and three constructs containing: wild type lead SNP, rs75703017 (pGL3-WT); mutant allele only (pGL3-M); and mutant allele and its two nearest SNPs (rs446868 and rs385390, pGL3-3M) (**Supplemental Fig. 3a**). Differential expression of a test reporter was detected in HepG2 (pGL3-M vs. pGL3-Ctr; $p<0.0001$, **Fig. 4a**) whereas HEK293T showed no significant change in expression (**Supplemental Fig. 3b**), supporting the liver-specific activity of the regulatory region.

Genome editing of *CFHR3* region using CRISPR/Cas9 confirms regulatory role in FH expression

To confirm that the *CFHR3* region identified regulates FH expression we undertook genome editing using CRISPR/Cas technology (**Fig. 4b**, **4c**). This required a liver cell line which constitutively expressed FH and carried at least one copy of *CFHR3*. Since none of the tested cell lines complied with both requirements, we differentiated human embryonic stem cells (H1 cell line) to hepatocytes[30]. H1 cells do not express FH nor FHR-3 and carry only one copy of *CFHR3*. Upon differentiation to hepatocytes (**Supplemental Fig. 4**) we detected FH expression (**Fig. 4c**) supporting the liver-specific expression reported previously[33]. Deletion

of a 2.8 kb region within *CFHR3* containing rs75703017 using CRISPR/Cas9 in H1 cells (**Fig. 4b**), followed by differentiation to hepatocytes, revealed enhanced FH expression, confirming the regulatory function of this region (**Fig. 4c**). This finding is consistent with our t-test analysis (**Supplemental Table 8**) of rs75703017 genotypes showing significant differential FH expression between DC vs. DD genotypes (p=6.8x10^{-3}) and further supported by Hi-C sequencing data, a strategy by which three-dimensional architecture of the genome can be studied by coupling proximity-based ligation with massive parallel sequencing and allows identification of long range genomic interactions[35]. In two cell lines a long range interaction could be observed between *CFH* and the association interval in *CFHR3* (**Supplemental Fig. 5**). Moreover, these results were concordant with *in vivo* data of individuals homozygous for the *CFHR3/CFHR1* deletion who showed increase FH levels (**Fig. 3b**).

Discussion

Genetic variants within *CFH* and the *CFHR* genes have been associated with genetic susceptibility to a range of human diseases[16–20,36]. Concordant with our work, deletion of *CFHR3-CFHR1* has been reported to alter FH levels in serum and modify genetic susceptibility to disease[16,36,37], suggesting that a regulatory region controlling FH levels may exist at this locus[16,36]. Identification of the causal variants underlying these associations has been difficult due to the complexity of the region, with CNVs and sequence homology hampering genotyping efforts. Thus, previous reports relied on surrogate markers to identify the deletion. Our strategy allowed us to type the CNV and polymorphisms in the *CFH-CFHR* region, to narrow the regulatory element to a short sequence in intron 1 of *CFHR3,* and to identify the complex interplay of six possible genotypes, including the lead SNP and CNV, with FH serum levels. Development of specific monoclonal antibodies for FH and FH-related proteins[29] made possible an accurate detection of serum levels of FH and FHR-3.

We have fine-mapped the complex *CFH-CFHR* region in patients of European ancestry with MD and found that susceptibility and resistance to the disease is associated with both, the SNP and CNV in intron 1 of *CFHR3*. Interestingly, the intronic lead SNP in *CFHR3*, rs75703017 (p=1.1x10^{-16}, OR=0.63, 95% CI 0.55-0.71), lies in a liver-specific regulatory region that has been shown to loop and interact with *CFH* at the genomic level. This interaction seems to control *CFH* transcription activity. Protective *CFHR3* genotypes were associated with low FH serum levels (pQTL=1.41x10^{-11}), whereas deletion of this region through genome editing in human embryonic stem cells differentiated to hepatocytes, showed a substantial increase (p<0.05) of *CFH* transcript levels and expression of FH protein, explaining the protective effect of rs75703017 in our analyses.

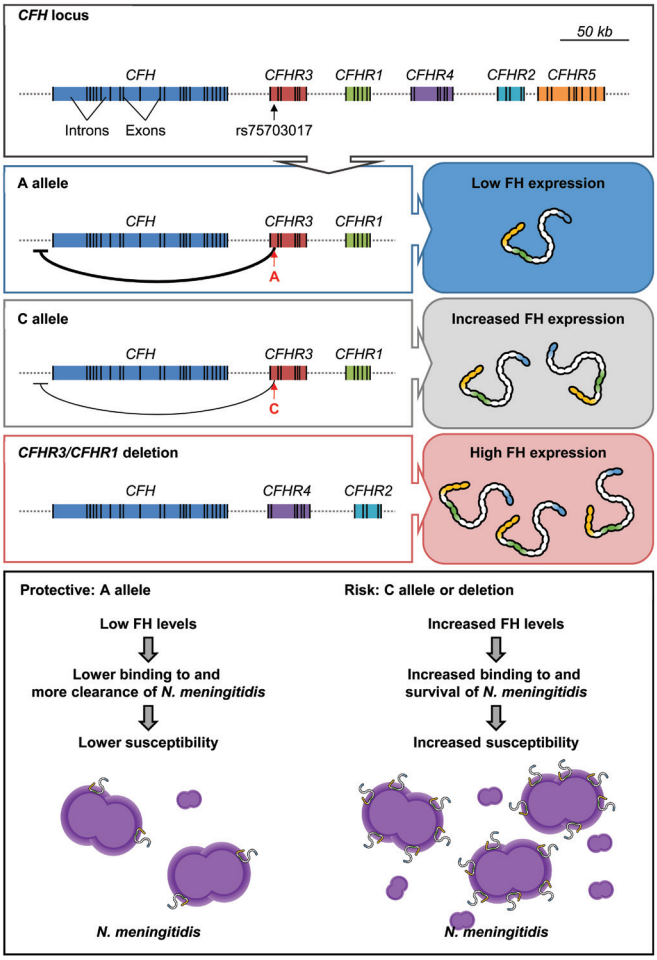


Figure 5 Schematic model of the effect of the *CFHR3* deletion and SNP (rs75703017) on FH levels, interactions with *N. meningitidis* and susceptibility to MD. Schematic diagram shows the structure of the gene region containing *CFH* and five *CFHR* genes. Carriers of the mutant allele, A, on rs75703017 show the lowest FH levels. Increased concentrations of FH are found with wild-type allele, C, on rs7570317, whereas deletion of *CFHR3* is associated with the highest levels of FH. Susceptibility to MD is driven by availability of FH in serum that increases binding to the meningococcal surface protein fHbp. This binding results in FH impairing complement mediated killing and allowing survival and growth of meningococci in blood.

We showed surviving MD patients had higher serum concentrations of FH than controls, and that high serum concentrations of FH increased risk of MD whereas low concentrations are protective. This is concordant with our previous report showing addition of excess FH to blood increases the survival of *N. meningitidis*[38]. Surprisingly, variation in serum concentration of FH corresponded with protective or risk genotypes in *CFHR3* rather than

CFH. Our data demonstrate that FH is a critical complement regulatory protein associated with MD susceptibility and that its serum concentrations are controlled through a cis-regulatory element in intron 1 of *CFHR3,* independent of FHR-3 protein levels.

Whereas previous studies have suggested that competition between FHR-3 and FH for the fHbp on the surface of *N. meningitidis* may be the mechanism controlling susceptibility to MD[21], we suggest that serum concentrations of FHR-3 are too low to affect binding of the (on average) 132-fold higher molar concentration of FH to fHbp[29] (**Supplemental Fig. 6**). Our genetic analysis confirms this assumption. In fact, our data indicate that the effect on MD susceptibility is predominantly defined by regulation of FH levels in serum by genetic variation in *CFHR3*, irrespective of serum FHR-3 levels. A schematic explanation of the inhibition of meningococcal bactericidal activity of complement in human blood by FH and its regulation by genetic variation in *CFHR3* is shown in **Figure 5**.

Our finding that serum concentrations of FH are genetically regulated by both SNPs and CNV in *CFHR3* have implications for many other infectious and inflammatory diseases. Complement activation is an important immune protection mechanism against infections, but uncontrolled or excessive complement activation is potentially damaging to host cells and tissues. FH is a major regulator of complement-mediated damage to host cells[39] as highlighted by the severe diseases associated with inadequate concentration or function of FH, including TTP/aHUS[19,31], glomerulonephritis[17], other inflammatory diseases[18] and age-related macular degeneration[20]. Many pathogens in addition to *N. meningitidis,* including *S. pneumoniae*[25], Group A streptococcus[26], *Borrelia burgdorferi*[40], and *Plasmodium falciparum*[27] possess FH-binding proteins and appear to use FH to evade complement-mediated killing. The genomic regulation of serum FH concentration that we have identified through genetic variation in *CFHR3* may thus be relevant to many other infectious and inflammatory diseases.

Author contributions

Study concept and design: all authors. Patient inclusion and collection of samples: VJW, BP, AKYT, AB, EP, RdG, JH, ME, MvdF, LJS, SA, FS, FM-T, AS, EC, AP and WZ. Data acquisition and analysis: VK, RBP, MA, ZYP, LB, MCB, EB and LJC. Interpretation of data: all authors. Drafting of manuscript: VK, RBP, ML, TWK, and SD. All authors critically revised the manuscript and contributed to the final version.

Conflict of interest disclosure

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

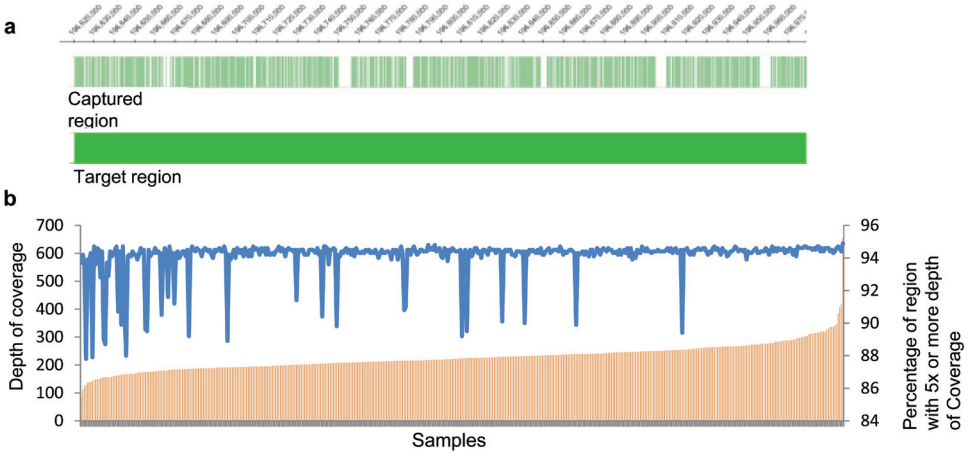
We would like to thank all children and parents who participated in this study. All samples have been collected under country-specific Institutional Review Boards (UK EC3263; Netherlands: 37986.091.11 / RvB12.51320; Austria: 24-116 ex 11/12; Spain: 2011/298). This work has been partially supported by the European Seventh Framework Program for Research and Technological Development (FP7) under EUCLIDS Grant Agreement nº. 279185, by funding from the Agency for Science Technology and Research of Singapore (A*STAR). The UK meningococcal cohort was established with support from Meningitis Research Foundation through grants to Imperial College London MRF 0402 and 1004.0. Patients were recruited through a grant from the Imperial College Biomedical Research Center (DMPED_P26077).
The ESIGEM Research group activities were supported by grants from Instituto de Salud Carlos III (Proyecto de Investigación en Salud, Acción Estratégica en Salud: proyecto GePEM PI16/01478) (A.S.); Consellería de Sanidade, Xunta de Galicia (RHI07/2-intensificación actividad investigadora, PS09749 and 10PXIB918184PR), Instituto de Salud Carlos III (Intensificación de la actividad investigadora 2007–2016), Convenio de colaboración de investigación (Wyeth España-Fundación IDICHUS 2007–2011), Convenio de colaboración de investigación (Novartis España-Fundación IDICHUS 2010–2011), Fondo de Investigación Sanitaria (FIS; PI070069/PI1000540) del plan nacional de I + D + I and 'fondos FEDER' (F.M.T.). The Swiss Pediatric Sepsis study was funded by grants from the Swiss National Science Foundation (342730_153158/1), the Swiss Society of Intensive Care, the Bangerter Foundation, the Vinetum and Borer Foundation, and the Foundation for the Health of Children and Adolescents. The Western Europe Meningococcal Study was supported by grants no 8842, 10112 and 12710 of the Oesterreichische Nationalbank (Austria), grants A3-16.K-8/2008-11 and A3-16.K-8/2006-9 of the Department for Science and Research of the Styrian federal government (Austria) and the non for profit association 'In Vita', Graz (Austria).

References

1. Borrow, R. *et al.* The Global Meningococcal Initiative: global epidemiology, the impact of vaccines on meningococcal disease and the importance of herd protection. *Expert Rev. Vaccines* 1–16 (2016). doi:10.10 80/14760584.2017.1258308
2. Sridhar, S. *et al.* Global incidence of serogroup B invasive meningococcal disease: a systematic review. *Lancet* **15,** 1334–1346 (2015).
3. Mustapha, M. M., Marsh, J. W. & Harrison, L. H. Global epidemiology of capsular group W meningococcal disease (1970-2015): Multifocal emergence and persistence of hypervirulent sequence type (ST)-11 clonal complex. *Vaccine* **34,** 1515–1523 (2016).
4. Xie, O., Pollard, A. J., Mueller, J. E. & Norheim, G. Emergence of serogroup X meningococcal disease in Africa: need for a vaccine. *Vaccine* **31,** 2852–2861 (2013).
5. Chow, J. *et al.* Invasive meningococcal meningitis serogroup C outbreak in Northwest Nigeria, 2015 – Third consecutive outbreak of a new strain. *PLoS Curr. Outbreaks* **8,** pii: ecurrents.outbreaks.06d10b6b4e6 90917d8b0a0426 (2016).
6. Campbell, H. *et al.* Presentation with gastrointestinal symptoms and high case fatality associated with group W meningococcal disease (MenW) in teenagers, England, July 2015 to January 2016. *Euro Surveill.* **21,** pii=30175 (2016).
7. Bårnes, G. K. *et al.* Prevalence and epidemiology of meningococcal carriage in Southern Ethiopia prior to implementation of MenAfriVac, a conjugate vaccine. *BMC Infect. Dis.* **16,** 639 (2016).
8. Davila, S. *et al.* Genome-wide association study identifies variants in the CFH region associated with host susceptibility to meningococcal disease. *Nat. Genet.* **42,** 772–776 (2010).
9. Bradley, D. T. *et al.* Susceptibility to invasive meningococcal disease: Polymorphism of complement system genes and Neisseria meningitidis factor H binding protein. *PLoS ONE* **10,** e0120757 (2015).
10. Martinón-Torres, F. *et al.* Natural resistance to meningococcal disease related to CFH loci: Meta-analysis of genome-wide association studies. *Sci. Rep.* **6,** 35842 (2016).
11. Sjöholm, A. G., Jönsson, G., Braconier, J. H., Sturfelt, G. & Truedsson, L. Complement deficiency and disease: an update. *Mol. Immunol.* **43,** 78–85 (2006).
12. Fijen, C. A. P., Kuijper, E. J., te Bulte, M. T., Daha, M. R. & Dankert, J. Assessment of complement deficiency in patients with meningococcal disease in the Netherlands. *Clin. Infect. Dis.* **28,** 98–105 (1999).
13. Józsi, M. & Zipfel, P. F. Factor H family proteins and human diseases. *Trends Immunol.* **29,** 380–387 (2008).
14. Blaum, B. S. The lectin self of complement factor H. *Curr. Opin. Struct. Biol.* **44,** 111–118 (2017).
15. Parente, R., Clark, S. J., Inforzato, A. & Day, A. J. Complement factor H in host defense and immune evasion. *Cell. Mol. Life Sci.* **74,** 1605–1624 (2016).
16. Zhao, J. *et al.* Association of genetic variants in complement factor H and factor H-related genes with systemic lupus erythematosus susceptibility. *PLoS Genet.* **7,** e1003079 (2011).
17. Abrera-Abeleda, M. A. *et al.* Variations in the complement regulatory genes factor H (CFH) and factor H related 5 (CFHR5) are associated with membranoproliferative glomerulonephritis type II (dense deposit disease). *J. Med. Genet.* **43,** 582–589 (2006).
18. Kiryluk, K. *et al.* Discovery of new risk loci for IgA nephropathy implicates genes involved in immunity against intestinal pathogens. *Nat. Genet.* **46,** 1187–1196 (2014).
19. Pickering, M. C. *et al.* Spontaneous hemolytic uremic syndrome triggered by complement factor H lacking surface recognition domains. *J. Exp. Med.* **204,** 1249–1256 (2007).
20. Hughes, A. E. *et al.* A common CFH haplotype, with deletion of CFHR1 and CFHR3, is associated with lower risk of age-related macular degeneration. *Nat. Genet.* **38,** 1173–1177 (2006).
21. Caesar, J. J. E. *et al.* Competition between antagonistic complement factors for a single protein on N. meningitidis rules disease susceptibility. *Elife* **3,** e04008 (2014).
22. Schneider, M. C. *et al.* Neisseria meningitidis recruits factor H using protein mimicry of host carbohydrates. *Nature* **458,** 890–893 (2009).
23. Hellwage, J. *et al.* The complement regulator factor H binds to the surface protein OspE of Borrelia burgdorferi. *J. Biol. Chem.* **276,** 8427–8435 (2001).
24. Haupt, K. *et al.* The Staphylococcus aureus protein Sbi acts as a complement inhibitor and forms a tripartite complex with host complement factor H and C3b. *PLoS Pathog.* **4,** e1000250 (2008).
25. Janulczyk, R., Iannelli, F., Sjöholm, A. G., Pozzi, G. & Björck, L. Hic, a novel surface protein of Streptococcus pneumoniae that interferes with complement function. *J. Biol. Chem.* **275,** 37257–37263 (2000).
26. Haapasalo, K. *et al.* Complement factor H allotype 402H is associated with increased C3b opsonization and

phagocytosis of Streptococcus pyogenes. *Mol. Microbiol.* **70,** 583–594 (2008).
27. Kennedy, A. T. *et al.* Recruitment of factor H as a novel complement evasion strategy for blood-stage Plasmodium falciparum infection. *J. Immunol.* **196,** 1239–1248 (2016).
28. Bellos, E. *et al.* cnvCapSeq: detecting copy number variation in long-range targeted resequencing data. *Nucleic Acids Res.* **42,** e158 (2014).
29. Pouw, R. B. *et al.* Complement factor H-related protein 3 serum levels are low compared to factor H and mainly determined by gene copy number variation in CFHR3. *PLoS ONE* **11,** e0152164 (2016).
30. Loh, K. M. *et al.* Efficient endoderm induction from human pluripotent stem cells by logically directing signals controlling lineage bifurcations. *Cell Stem Cell* **14,** 237–252 (2014).
31. Zipfel, P. F. *et al.* Deletion of complement factor H-related genes CFHR1 and CFHR3 is associated with atypical hemolytic uremic syndrome. *PLoS Genet.* **3,** e41 (2007).
32. Ward, L. D. & Kellis, M. HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40,** D930–D934 (2012).
33. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518,** 317–330 (2015).
34. Hart, C. A. & Cuevas, L. E. Meningococcal disease in Africa. *Ann. Trop. Med. Parasitol.* **91,** 777–785 (1997).
35. van de Werken, H. J. G. *et al.* Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* **9,** 969–972 (2012).
36. Zhu, L. *et al.* Variants in complement factor H and complement factor H-related protein genes, CFHR3 and CFHR1, affect complement activation in IgA nephropathy. *J. Am. Soc. Nephrol.* **26,** 1195–1204 (2014).
37. Ansari, M. *et al.* Genetic influences on plasma CFH and CFHR1 concentrations and their role in susceptibility to age-related macular degeneration. *Hum. Mol. Genet.* **22,** 4857–4869 (2013).
38. Haralambous, E. *et al.* Factor H, a regulator of complement activity, is a major determinant of meningococcal disease susceptibility in UK Caucasian patients. *Scand. J. Infect. Dis.* **38,** 764–771 (2006).
39. Ferluga, J., Kouser, L., Murugaiah, V., Sim, R. B. & Kishore, U. Potential influences of complement factor H in autoimmune inflammatory and thrombotic disorders. *Mol. Immunol.* **84,** 84–106 (2017).
40. Kraiczy, P., Skerka, C., Kirschfink, M., Brade, V. & Zipfel, P. F. Immune evasion of Borrelia burgdorferi by acquisition of human complement regulators FHL-1/reconectin and factor H. *Eur. J. Immunol.* **31,** 1674–1684 (2001).

Supplemental materials



Supplemental Figure 1 Sequencing (**a**) target region (solid green) and captured region (chr1:196620000-196979000) including *CFH* to *CFHR5* genes. (**b**) Depth of coverage of each samples (average 227x ± 47) and percentage of sequenced region with depth of coverage 5x or more per sample.

Supplemental Table 1 Sequence of screening primers used to validate copy number variants within *CFHR1* and *CFHR3*

Name	Sequence
CFHR1 F	AAATGCAGGTCCACTGGTAAGT
CFHR1 R	GAGATGATGATGCTACCGGTTT
CFHR3 F	GGAGAAAGGCTGGTCTCCTACT
CFHR3 R	CTGAGACTGTCGTCCGTGTTAC

Supplemental Table 2 List of samples, with deletions estimated by sequencing data, validated using qPCR with *CFHR3* and *CFHR1* primers

		Sets of Primer				Sets of Primer	
No.	Sample ID	*CFHR3*	*CFHR1*	No.	Sample ID	*CFHR3*	*CFHR1*
1	CHH3607	1	1	35	CHH3678	1	1
2	CHH3608	1	1	36	CHH3681	1	1
3	CHH3611	1	1	37	CHH3685	1	1
4	CHH3614	1	1	38	CHH3695	1	1
5	CHH3615	1	1	39	CHH3696	1	1
6	CHH3616	1	1	40	CHH3700	1	1
7	CHH3617	1	1	41	CHH3702	1	1
8	CHH3618	1	1	42	CHH3703	1	1
9	CHH3619	1	1	43	CHH3705	1	1
10	CHH3620	1	1	44	CHH3706	1	1
11	CHH3621	1	1	45	CHH3708	1	1
12	CHH3622	0	0	46	CHH3709	1	1
13	CHH3624	0	0	47	CHH3715	1	1
14	CHH3626	1	1	48	CHH3716	1	1
15	CHH3628	1	1	49	CHH3717	1	1
16	CHH3629	0	0	50	CHH3718	1	1
17	CHH3630	1	1	51	CHH3722	1	1
18	CHH3631	1	1	52	CHH3725	1	1
19	CHH3633	1	1	53	CHH3729	0	0
20	CHH3635	1	1	54	CHH3730	1	1
21	CHH3637	1	1	55	CHH3731	1	1
22	CHH3639	1	1	56	CHH3732	1	1
23	CHH3641	1	1	57	CHH3736	1	1
24	CHH3642	1	1	58	CHH3739	1	1
25	CHH3643	0	0	59	CHH3750	0	1
26	CHH3645	1	1	60	CHH3756	0	0
27	CHH3646	1	1	61	CHH3757	1	1
28	CHH3647	1	1	62	CHH3764	1	1
29	CHH3656	1	1	63	CHH3765	1	1
30	CHH3661	1	1	64	CHH3767	1	1
31	CHH3664	1	1	65	CHH3768	1	1
32	CHH3665	1	1	66	CHH3799	1	1
33	CHH3675	0	0	67	CHH3802	1	1
34	CHH3676	1	1				

Samples in red indicate non concordance in copy number inferred by sequencing data and qPCR.

Supplemental Table 3 Sequence of primers and oligos for plasmid construction

Name	Sequence
EF1a F	GACAAATGGCTCTAGATTCACTCGAGGTGCCCGT
EF1a R	TGAAAAAAAGTGATTATCCTCACGACACCTGAAA
Amp pUC F	GAATTTTAACAAAATTAGACGTCAGGTGGCACT
Amp pUC R	CTCGAGTGAATCTAGAATGTAACGGACCTCGAGC
sg chimeric pUC F	TCCGTTACATTCTAGGAGGGCCTATTTCCCATGA
sgRNA cassette F	GACAAATGGCTCTAGGAGGGCCTATTTCCCATGA
sgRNA cassette R	CTCGAGTGAATCTAGAGCCATTTGTCTGCAGAATTGG
Cas9 3' F	AGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAA
Cas9 3' 2sg R	AGTGGCTCCGCTTCCCTTTTTCTTTTTTGCCTGGCCGGCCTTT
Cas9 3' 4sg R	AGTGGCTCCGCTTCCGCTGGCCTCCACCTTTCTCT
2A Clover F	ggaagcggagccactaacttctccctgttgaaac
2A Clover R	CGAGCTCTAGGAATTCttaCTTGTACAGCTCGTCCAT
luci 5' F	gtaaaatcgataaggatccgCATTTTGTAAAACTCTCTGAAC
luci 5' R	atatagttctTTAATATGCAAGGCCTGC
luci 3' F	tgcatattaaAGAACTATATGGTAACATTAGCAG
luci 3' R	aaggctctcaagggcatcggAGTCAAGGAAATCCTGGG

Supplemental Table 4 Sequence of guided RNAs and screening primers used to target the region of interest

Name	Sequence
gRNA 1	GCTTCAATAATTCCAGTTAG
gRNA 2	CCTTAAAGCCCCTAGCTTCA
gRNA 3	AGCGTCCCAAAGAAGTACAA
gRNA 4	TTACACTTTGGCATAACAAA
Screening primer F	GAAGCCTTTCGTTCCCTT
Screening primer R	CACTAAAAGGCACTGGAA

Supplemental Table 5 Sequences of primers used for qPCR for quantification of liver specific transcripts and *CFH* in H1 and H1-differentiated liver cells

Gene	Forward	Reverse
TAT	CCTTTACTTGGAAGGCTTCGCT	CGTCCAGAATTGAGGGGAGGTT
CYP3A4	AAGTGTGGGGCTTTTATGATGGT	GGTGAAGGTTGGAGACAGCAATG
AAT	CCCCTGTTTGCTCCTCCGATAA	CGGCATTGTCGATTCACTGTCC
ALB	ACCCCACACGCCTTTGGCACAA	CACACCCCTGGAATAAGCCGAGCT
CFH	TGCACAAGTACTGGCTGGAT	TCCCAGTAACTTCCTGACGGA

Supplemental Table 6 Number of SNPs which failed the different filtering criteria

	No. of SNPs
Total	6020
PASS	4369
Strand Bias >60	733
Haplotype Score >13	249
Read Position Rank Sum <-8	354
Quality <50	144
MQ Rank Sum < -12.5	674
Quality by depth <2	539
Mapping Quality < 40	543

Supplemental Table 7 List of SNPs genotyped by Sequenom on chromosome 1

196633739	rs10922093	rs10922096
rs34181066	rs11580821	rs453912
rs72734328	196701018	rs2878713
rs12568400	rs1854499	rs11807686
rs385390	rs75703017	rs12408446
rs376841	rs620015	rs11807997
rs425524	rs1738741	rs370789
196755801	rs10754207	rs391537
rs401188	rs12240148	rs390154
rs387107	rs4086175	rs377298
rs116249058	rs1329423	rs1694442
rs12409571	rs12127759	rs532507
rs57426012	rs16840419	rs396051
rs114260576	rs34331968	196801353
rs41310132	rs12134975	19691576

Supplemental Table 8 Adjusted P-values estimated using the t-test between different genotype categories of rs75703017 for FHR-3 and FH levels

	AA	CA	DA	CC	DC	DD
AA	-	0.82	0.45	0.04	4.8×10^{-3}	6.9×10^{-6}
CA	0.05	-	0.48	1.5×10^{-4}	2.7×10^{-7}	2.3×10^{-9}
DA	1.6×10^{-3}	0.04	-	0.82	0.44	2.1×10^{-3}
CC	4.8×10^{-7}	3.5×10^{-10}	0.32	-	0.18	1.1×10^{-4}
DC	5.8×10^{-13}	$<2.0 \times 10^{-16}$	1.4×10^{-3}	1.6×10^{-9}	-	6.8×10^{-3}
DD	7.6×10^{-16}	$<2.0 \times 10^{-16}$	7.4×10^{-7}	5.5×10^{-10}	2.2×10^{-3}	-

Values for FHR-3 are indicated in blue. Values for FH are indicted in black. The P-value is adjusted for multiple comparison correction using Holm's method. The description of the alleles is in Figure 2 of the main manuscript.

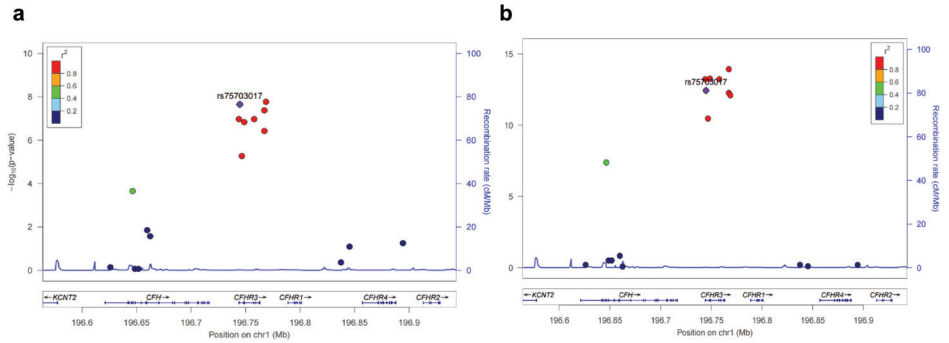
Supplemental Table 9 Unadjusted Odds ratio (P-values) of the *CFHR3/CFHR1* deletion with susceptibility to MD

Sequence	Replication			Meta Analysis
Western European	Western European	UK	Spain	
1.49 (0.028)	1.1 (0.50)	1.01 (0.94)	1.07 (0.50)	1.02 (0.76)

Supplemental Table 10 Unadjusted P-values for MD estimated using logistic regression analysis for top associated SNP (rs75703017)

	Genotypes			Total	P-Values	
	AA	A/Del	Del/Del		Genotype model	Dominant model
Cases/Controls	35/50	69/140	73/78	177/268	0.013	$8.10x10^{-3}$

To eliminate the effect of risk allele 'C', only samples with mutant 'A' allele were considered. Dominant model refers to AA + A/del vs del/del comparison. 'del' refers to *CFHR3/CFHR1* deletion

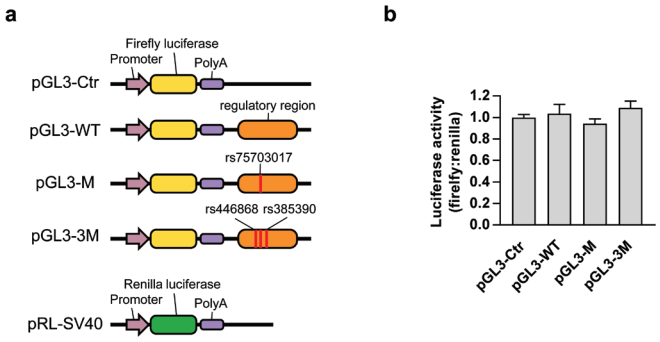


Supplemental Figure 2 Plot showing meta-analysis of quantitative trait association results based on samples without *CFHR3/CFHR1* deletions [combining the P-values, obtained by linear regression, of the UK (n=186) and the Western Europe (n=37) SNPs genotyped at the replication stage]. Genetic recombination rates (cM/Mb), estimated using 1000g samples, are depicted by a blue line] and the color intensity of each symbol reflects the extent of LD with the labeled SNP rs75703017 (SNPs arranged according to their hg19 build chromosomal position on x axis and genetic recombination rates (cM/Mb), estimated using 1000g samples), for protein levels (µg/mL) of (**a**) FHR and (**b**) FHR-3.

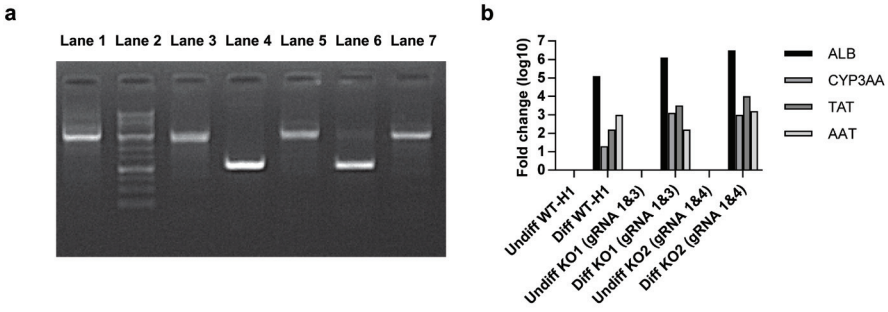
Supplemental Table 11 Beta and P-values for top associated SNP (rs75703017 with minor allele 'A') obtained using phenotype-genotype correlation from FH and FHR-3 protein dataset

	UK			Western Europe			Meta-Analysis			
	Beta	SE	P	Beta	SE	P	Beta	P	Q	I^2
FHR-3	0.4781	0.06926	$8.11E^{-11}$	0.3887	0.1674	0.0262	0.465	$3.70E^{-13}$	0.6217	0.00
FH	-47.46	9.135	$5.41E^{-07}$	-69.1	31.8	0.03666	-49.1096	$2.23E^{-08}$	0.5131	0.00

To remove the effect of copy number change, samples without copy number change were considered (n=185 and 37 for the UK and the Western Europe samples, respectively) for the analysis.

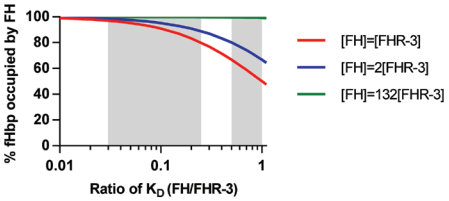


Supplemental Figure 3 (**a**) Schematic illustration of the luciferase constructs used. Four test firefly plasmids pGL3-Ctr, pGL3-WT, pGL3-M & pGL3-3M and the control renilla pRL-SV40 plasmid, each with key features and differences indicated. (**b**) Effect of 2.8 kb regulatory sequence and the lead SNP rs75703017 (Wild type and mutant) on reporter (firefly luciferase) activity in HEK293T (Human embryonic kidney cell line). We compared cells containing an empty vector, mimicking CNV, (pGL3-Ctr), and three constructs containing: wild type lead SNP, rs75703017 (pGL3-WT); mutant allele only (pGL3-M); and mutant allele and its two nearest SNPs (rs446868 and rs385390, pGL3-3M). The graph is representative of five independent experiments.



Supplemental Figure 4 (**a**) Screening of CRISPR deleted clones. Lane 1: wild-type H1, lane 2: DNA ladder, lanes 3, 5 and 7: unsuccessful clones, lane 4: KO1(gRNA 1&3) and lane 6: KO2(gRNA 1-4). Expected size of wild type H1 is 3087bp. (**b**) Fold change [cells after liver differentiation (Diff)/undifferentiated (Undiff)] of expression levels for four liver markers of two CRISPR edited (KO) cell lines and one unedited (WT) H1 hES cell line (each bar for a protein represents one data point). Expression was measured by RT-qPCR.

Supplemental Figure 5 Road Map ChIP-seq data (top panel) showing active regulatory region predicted only in the E066 liver cell-line by H3K4me3 (orange) and H3K9ac (purple) overlapping with the top SNP (rs75703017 depicted by yellow bar) and the SNPs in LD with it. UCSD Hi-C data (bottom panel) with 20 Kb resolution predicts cis-interaction (pink curves) between *CFHR3* and *CFH* gene in H1 and IMR90 cell lines. The 20 kb fragment showing this interaction includes intron 1 of *CFHR3*, location of rs75703017 (shadowed in yellow in the picture). For the purpose of reproducibility all the data were generated in duplicates.



Supplemental Figure 6 Graph depicting the percentage of fHbp occupied with FH, according to the model as published in Caesar *et al*. 2014 (eLIFE). Caesar et al. proposed an equal molar concentration or a two-fold higher concentration of FH compared to FHR-3 and calculated the % fHbp being occupied by FH. The green line depicts the physiological ratio between FH and FHR-3 (132-fold difference in molar concentration) and what that would mean for FH binding to fHbp. Grey boxes indicate the range in affinity ratio for FH/FHR-3 of fHbp v.1 (left box) and fHbp v.2 and v.3 (right box).