



HAL
open science

Variants in the non-coding region of the TLR2 gene associated with infectious subphenotypes in pediatric sickle cell anemia

Susana David, Pedro A. Aguiar, Liliana Antunes, Alexandra Dias, Anabela Morais, Anavaj Sakuntabhai, João Lavinha

► To cite this version:

Susana David, Pedro A. Aguiar, Liliana Antunes, Alexandra Dias, Anabela Morais, et al.. Variants in the non-coding region of the TLR2 gene associated with infectious subphenotypes in pediatric sickle cell anemia. *Immunogenetics*, 2018, 70 (1), pp.37-51. 10.1007/s00251-017-1013-7 . pasteur-02003876

HAL Id: pasteur-02003876

<https://pasteur.hal.science/pasteur-02003876>

Submitted on 19 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

Variants in the non-coding region of the *TLR2* gene associated with infectious subphenotypes in pediatric sickle cell anemia

Susana David^{1,2*}, Pedro Aguiar³, Liliana Antunes⁴, Alexandra Dias⁵, Anabela Morais⁶, Anavaj Sakuntabhai^{7,8}, João Lavinha^{1,9}

¹Departamento de Genética Humana, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisboa, Portugal

²Centro de Estudos de Ciência Animal, Instituto de Ciências e Tecnologias Agrárias e Agro-Alimentares, Universidade do Porto, Porto, Portugal

³Centro de Investigação em Saúde Pública, Escola Nacional de Saúde Pública, Universidade Nova de Lisboa, Lisboa, Portugal

⁴Departamento de Epidemiologia, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisboa, Portugal

⁵Departamento de Pediatria, Hospital Prof Doutor Fernando Fonseca, Amadora, Portugal

⁶Departamento de Pediatria, Hospital de Santa Maria, Lisboa, Portugal

⁷Unité de la Génétique Fonctionnelle des Maladies Infectieuses, Institut, Pasteur, Paris, France

⁸Centre National de la Recherche Scientifique, URA3012 Paris, France

21 ⁹BiolSI, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

22

23

24 **Key words:** Sickle cell anemia; *TLR2*; genetic variants; viral and bacterial infection;
25 hemolytic component; genotype-to-phenotype association

26

27 **Running title:** TLR2 and infection in pediatric sickle cell anemia

28

29 * Corresponding author. Address: Departamento de Genética Humana, Instituto Nacional
30 de Saúde Doutor Ricardo Jorge, Av. Padre Cruz, 1649-016, Lisboa, Portugal. Tel.: (+351)
31 217519200, fax: (+351) 217526400.

32 *E-mail address:* suzana.david@insa.min-saude.pt

33

34 ABSTRACT

35 Sickle cell anemia (SCA) is characterized by chronic hemolysis, severe vasoocclusive
36 crises (VOC) and recurrent often severe infections. A cohort of 95 SCA pediatric patients
37 was the background for genotype-to-phenotype association of the patient's infectious
38 disease phenotype and three noncoding polymorphic regions of the *TLR2* gene, the -196 to
39 -174 indel, SNP rs4696480 and a (GT)_n short tandem repeat. The infectious
40 subphenotypes included: (A) recurrent respiratory infections, and (B) severe bacterial
41 infection at least once during the patient's follow-up. The absence of the haplotype [Del]-T-
42 [n≥17](Hap7) in homozygosity protected against subphenotype (B), in a statistically
43 significant association, resisting correction for multiple testing. For the individual loci, the
44 same association tendencies were observed as in the haplotype, including a deleterious
45 association between the SNP rs4696480 T allele and subphenotype (A), whereas the A/A
46 genotype was protective, and a deleterious effect of the A/T genotype with subphenotype
47 (B), as well as including the protective effect of -196 to -174 insert (Ins) and deleterious
48 effect of the deletion (Del) in homozygosity, against subphenotype (B). Moreover, a
49 reduction in the incidence rate of severe bacterial infection was associated to a rise in the
50 hemolytic score, fetal hemoglobin levels (prior to hydroxyureia treatment) and 3.7-kb alpha-
51 thalassemia. Interestingly, differences between the effects of the two latter covariables
52 favoring a reduction in the incidence rate of subphenotype (B) contrast with a resulting
53 increase in relation to subphenotype (A). These results could have practical implications in
54 health care strategies to lower the morbidity and mortality of SCA patients.

55 **1. Introduction**

56 As the first line of host defense, the innate immune system discriminates between
57 varieties of pathogens *inter alia* via the pattern-recognition receptors (PRRs). These
58 receptors recognize as ligands essential microbial components, known as pathogen-
59 associated molecular patterns (PAMPs), horizontally expressed either across all the major
60 classes of infectious agents or more specifically in a particular class. Different PRRs react
61 with specific PAMPs, leading to distinct immunological responses in the early phase of
62 infection (Akira *et al.*, 2006).

63 Toll-like receptors (TLRs) are a pluripotent family of PRRs recognizing PAMPs from a
64 large spectrum of infectious agents. In humans, the TLR family of transmembrane proteins
65 comprises 10 functional members (Takeda *et al.*, 2003). Amongst these, TLR2 has been
66 associated with the innate immune response against bacteria, fungi, parasites and viruses
67 through the recognition of a variety of specific lipids of the cell envelopes and some viral-
68 envelope proteins (Kawai and Akira, 2005; Misch and Hawn, 2008; Thompson and Iwasaki
69 2008). Moreover, TLR2 dimer formation with TLR1 and TLR6 and interaction with non-TLR
70 receptors further diversifies its recognition potential (Kawai and Akira, 2011).

71 Genetic variation in the human *TLR2* gene (located on chromosome 4q32), can influence
72 the host susceptibility to infection (O'Neill *et al.*, 2009; Lin *et al.*, 2012; de Oliveira
73 Nascimento *et al.*, 2012; Medvedev, 2013) including by bacterial agents (Ma *et al.*, 2007;
74 Hawn *et al.*, 2009; Chen *et al.*, 2010; Alter *et al.*, 2011; Casanova *et al.*, 2011; Nischalke *et*
75 *al.*, 2011; Song *et al.*, 2011; Esposito *et al.*, 2014; Zhang *et al.*, 2013) and enveloped
76 viruses (Cai and Zheng, 2012; Nischalke *et al.*, 2012; Novis *et al.*, 2013; Bagheri *et al.*,
77 2014; Triantafilou *et al.*, 2014; Ma *et al.*, 2014; Reuven *et al.*, 2014; Chen *et al.*, 2015;
78 Hernandez *et al.*, 2015). Moreover, genetic polymorphic variants heterogeneously

79 distributed between populations of different ethnic or geographical background have been
80 described in association with inter-population susceptibility differences (Ioana et al., 2012;
81 Mukherjee et al., 2014).

82 The genetic variants in human *TLR2* reported in the literature as being related with
83 susceptibility to infection include:

- 84 (i) The -196 to -174 insertion or deletion (indel) of 22 bp localized in the *TLR2* first
85 untranslated exon, associated with susceptibility to infections of various
86 etiologies (Nischalke et al., 2011, 2012; Velez et al., 2010; Greene et al., 2012).
- 87 (ii) The single nucleotide polymorphism (SNP) rs4696480, c.-16934A>T
88 (g.153685974T>A), localized in the *TLR2* promoter, in a preliminary study
89 associated to susceptibility to pulmonary bacterial infections in sickle cell disease
90 (Rostane et al., 2012).
- 91 (iii) The microsatellite, short tandem repeat (STR), (GT)_n polymorphism in the
92 *TLR2* intron 2 ($12 \leq n \leq 28$), described as highly variable between populations
93 (Yim et al., 2004) and associated with different susceptibilities to tuberculosis in
94 distinct populations (Yim et al., 2006), as well as being implicated in other
95 infections (Folwaczny et al., 2011).

96 Sickle cell anemia (SCA) is a life threatening clinically heterogeneous monogenic
97 autosomal recessive chronic anemia (Driss et al., 2009). The disease is characterized by
98 recurrent episodes of severe vasoocclusion, chronic hemolysis and recurrent often severe
99 infection. It is caused by the homozygosity for a single nucleotide substitution in the human
100 beta-globin gene (*HBB*) (located on chromosome 11p15.5). This mutation (*HBB*:c.20A>T)
101 gives rise to a hemoglobin structural variant (HbS), that under low oxygen partial pressure

102 polymerizes, leading to red blood cell sickling and a clinically heterogeneous
103 pathophysiological cascade.

104 Development and validation of early predictors of disease severity is a strategy oriented
105 towards lowering morbidity and mortality in the affected patients. In this context, it has been
106 our purpose to investigate the association of the above mentioned noncoding genetic
107 variation in *TLR2* on the infection phenotype of SCA pediatric patients.

108

109

110 **Methods**

111 *Study population*

112 Ninety-five unrelated patients with SCA were recruited on the basis of the criterion of having
113 more than three and under 20 years. This 95 patient subgroup was part of a total of 99 SCA
114 patients, attending the pediatric wards of two general hospitals in Greater Lisbon, enrolled
115 in a previous longitudinal study with a total follow-up of 557 person*years (Coelho et al.,
116 2014). Patient information relative to gender, alpha-thalassemia, *HBB**S haplotypes and
117 fetal hemoglobin level was available from this study, and as previously described, Sub-
118 Saharan African ancestry accounted for the majority of the subjects (97%). The study
119 procedures were approved by INSA's ethics committee and are in accordance with the
120 Helsinki Declaration of 1975. Written informed consent, was given by the patients' legal
121 representatives.

122

123 *Genotyping*

124 Three *TLR2* variants having been associated with infection susceptibility were selected for
125 genotyping. These included the -196 to -174 a 22 bp nucleotide indel in the first
126 untranslated exon, promoter SNP rs4696480, c.-16934A>T (g.153685974T>A), and the
127 (GT)_n STR polymorphism in the second intron. Genotyping of these loci was performed for
128 all 95 patients.

129 All PCR reactions were carried out in a total reaction volume of 25 µL, containing 22,5 µL in
130 house PCR Buffer (Table S1); 0.25 µL of each primer (working solution 25 µM); 0,1 µL Taq
131 DNA polymerase (5 U/µL, GoTaq, Promega, Madison, WI, USA); 0.9 µL dH₂O; 1 µL DNA
132 (100 ng/µL).

133 The genotyping of the *TLR2* c.-196 to -174 indel was carried out using the forward 5'-
134 CACGGAGGCAGCGAGAAA-3' and reverse 5'-CTGGGCCGTGCAAAGAAG-3' primers
135 (Tahara et al., 2007). The PCR program consisted in an initial denaturation step at 95 °C for
136 5 min, 35 cycles of 95 °C for 30 s, 60 °C for 40 s, and 72 °C for 40 s, and a final extension
137 step at 72 °C for 7 min. The PCR products were visualized on a 4% agarose gel stained
138 with ethidium bromide and interpreted as follows: wild type, a single 286 bp band
139 (Insertion/Insertion, Ins/Ins); homozygous for the deletion, one 264 bp band
140 (Deletion/Deletion, Del/Del); heterozygous, two bands of 286 bp (Ins) and 264 bp (Del).

141 The genotyping of the rs4696480 SNP was carried out using restriction fragment length
142 polymorphism (RFLP) analysis. Digestion with the restriction enzyme HphI (NEB® New
143 England Biolabs, Ipswich, MA, USA) was carried out after PCR amplification of a 329 bp
144 fragment around the SNP. For the PCR the forward 5'-CCCCCAAATTTAAAAGAGGGC-3'
145 and reverse 5'-TGTTATCACCAAGGGAGCAG-3' primers were used. The PCR program
146 consisted in an initial denaturation step at 95 °C for 5 min, 30 cycles of 95 °C for 30 s, 60 °C
147 for 40 s, and 72 °C for 40 s, and a final extension step at 72 °C for 7 min. HphI enzymatic

148 restriction of the PCR products was carried out at 37 °C for 3hours. The digested products
149 were visualized on a 2% agarose gel stained with ethidium bromide and interpreted as
150 follows: homozygous T/T three bands of 212 bp, 68 bp and 47pb; heterozygous T/A four
151 bands of 260 bp, 212 bp, 47 bp, 68 pb; homozygous A/A two bands of 260 bp and 68 pb.

152 For genotyping of the STR polymorphism, the surrounding region in intron 2 was amplified
153 using the FAM (fluorescein marker 6-carboxyfluorescein) labeled forward primer 5'-FAM-
154 GCATTGCTGAATGTATCAGGGA-3', and the reverse primer 5'-
155 CTTGAGAAATGTTTTCTAGGC-3' (Folwaczny et al., 2011). The PCR program consisted in
156 an initial denaturation step at 95 °C for 5 min, 35 cycles of 95 °C for 30 s, 56 °C for 30 s,
157 and 72 °C for 30 s, and a final extension step at 72 °C for 10 min. The number of (GT)_n
158 repeats was identified by sizing the PCR products with the ABI 310 sequencer and the
159 Genescan Analysis 2.1 software (Applied Biosystems, Foster City, CA, USA). The following
160 subdivision based on the number of repeats (n) was used: Allele S, $n \leq 16$; Allele M, $17 \leq n$
161 ≤ 22 , and Allele L, $n \geq 23$, which was further simplified as a binary variable S versus others
162 (M or L) (Yim et al., 2006).

163 Hardy Weinberg equilibrium was verified using the standard asymptotic (1 df genotypic chi-
164 squared test) in all instances. For the STR the Hardy-Weinberg Equilibrium was verified
165 considering the S allele relative to the others (M or L).

166 Haplotype reconstruction from the genotyping data was carried out using the PHASE,
167 version 2.1, program (Stephens et al., 2001, 2003) and is summarized in Table 1. The
168 distances between the variants (relative to the arbitrary position c.-18619, 5' - indel
169 (untranslated exon 1) – SNP (promoter) – microsatellite (exon 2) – 3': 66/1685/18475 bp
170 respectively), and the binary subdivision of the *TLR2* (GT)_n locus allowed for their
171 approximation as single loci by the PHASE program.

172

173 *Derivation of a hemolytic component*

174 The derivation of a hemolytic component followed the strategy described by Nouraie and
175 collaborators (2013). These authors derived a hemolytic component using principal
176 component analysis (PCA) with 4 degrees of freedom from data on the reticulocyte count
177 (%), and serum lactate dehydrogenase (U/L), aspartate aminotransferase (U/L) and total
178 bilirubin concentrations (mg/dL). Levels of aspartate aminotransferase were not available in
179 our patient database, but as this is a more distal marker of hemolysis and the other three
180 are more direct markers, we worked with these remaining three degrees of freedom. For
181 the PCA only steady state values were used. The association of this component with
182 clinical phenotypes was assessed.

183

184 *Genotype-to-phenotype association studies*

185 A stepwise approach in the statistical analysis of genotype-to-phenotype association was
186 used (Fig. S1). In this manner the open-source whole genome association analysis toolset
187 PLINK v1.07 program (Purcell et al., 2007), particularly adapted to cross-sectional
188 association analysis of SNPs, was applied to the analysis of SNP rs4696480, c.-16934A>T,
189 in exploratory preliminary association testing for the selection between various possible
190 infectious disease subphenotypes. This preliminary analysis was followed by other cross-
191 sectional analytical methods, for association testing of the variants and haplotypes of the all
192 the different *TLR2* polymorphic loci under consideration (indel, SNP and STR). These
193 informative tests included descriptive statistics for the comparison of rates or the chi-square
194 test applied dichotomous traits. Lastly, a quasipoisson regression model was applied in the

195 association studies between the different *TLR2* polymorphic loci and the selected
196 quantitative or rare event (dichotomous) disease subphenotypes. This allowed for analysis
197 of the time based cohort and the prediction of correlated dependent variables.

198 For analysis using the PLINK program, genotype-to-phenotype association was carried out
199 redefining the infectious status as a binary trait, considering the number of infections/ time
200 at risk in person*years and “time at risk” equal to the time under observation during the
201 follow-up. A positive relationship was obtained when “the average rates of the phenotype
202 per patient (number of infections/ time at risk)/ the average rates of the phenotype for all the
203 patients (number of infections/ time at risk) (n = 95)” >1. The allelic, recessive and dominant
204 inheritance models were tested and the Cochran-Armitage trend test performed. No
205 correction for multiple comparisons was conducted at this level of analysis.

206 For descriptive statistics, the binary variables were expressed as proportions and the
207 continuous variables as means \pm standard deviation. For the comparison of means, the
208 two-sample t-test was used.

209 Descriptive statistics was carried out on the phenotypes showing a statistically significant
210 result in the preliminary PLINK analysis. For the quantitative infectious phenotype variables,
211 we compared the average rates of the phenotype for all the patients (number of infections/
212 time at risk), in the presence and in the absence of the *TLR2* variants or their haplotypes. If
213 the presence of the genotype was associated with an increase in the incidence of
214 respiratory infections (positive statistical effect), this indicated a deleterious genotype with a
215 potentially unfavorable clinical effect. On the contrary, if the presence of the variant was
216 associated with a decrease in the incidence of respiratory infections (negative statistical
217 effect), this indicated a protective genotype with a potentially favorable clinical effect. For
218 the dichotomous infectious phenotype variables (presence vs. absence), the differences in

219 allele frequencies and genotype distributions between the 2 groups was evaluated using
220 the Fisher's exact test (two sided) and the odds ratio (OR) was calculated under different
221 genetic models to measure the risk of infection.

222 The inheritance models for the *TLR2* variants and their haplotypes were tested using a
223 generalized linear model assuming a quasipoisson regression model for the SCD patients
224 ($n = 95$): *Phenotype* ~ *Genotype* + *steady state hemoglobin level* + *steady state leukocyte*
225 *count* + *fetal hemoglobin (prior to hydroxyurea treatment)* + *haemolytic component* + *alpha-*
226 *thalassemia*. The allelic and genotypic inheritance models were tested for each variant.
227 This statistical analysis was carried out in R (Team, R.C., 2013). The quasipoisson
228 regression model adjusts for possible deviations from the poisson assumptions (Ver Hoef
229 and Boveng, 2007). Risk estimates were obtained using incidence rate ratio calculations for
230 those inheritance models showing statistically significant associations to infection after
231 forward optimization.

232 The haplotypes defined by the three *TLR2* loci were analyzed for association by
233 considering each haplotype as an allele.

234 Genotype-to-phenotype associations were considered statistically significant when the $p \leq$
235 0.05. However, to control type I errors, the Bonferroni correction for multiple comparisons
236 (Bland and Altman 1995), was performed for the genetic models tested. For subphenotype
237 (A) the adjusted alpha was 0.001667, for a number of tests performed per allele $n = 30$ (six
238 covariables for two inheritance models, allelic and genotypic, with 2 tests for the allelic and
239 3 for the genotypic) and for subphenotype (B) the adjusted alpha was 0.001471, for $n = 34$
240 (adding to the above four Fisher's exact tests).

241

242

243 **Results**

244 *Study population*

245 In this study, genetic association studies for various infectious disease phenotypes were
246 performed for 95 pediatric SCA patients (mostly of Sub-Saharan origin) and infectious
247 episodes occurring within an age range of over three and under 20 years (Fig. 1). The total
248 observation period for the cohort (longitudinal study) was 433.4 person*year. The
249 distribution of the observations per age group is represented in Fig.1 and the average age
250 at observation was 7.7 ± 3.7 years. The mean follow-up per patient was 4.6 ± 1.5 years. The
251 male/ female (M/F) sex ratio was 1.21. Values for the other population data included: the
252 average steady state levels of hemoglobin 7.9 ± 0.8 (g/dL), fetal hemoglobin levels prior to
253 eventual hydroxyurea treatment 11.0 ± 7.6 (%) and alpha-thalassemia ($-\alpha 3.7$ -kb deletion
254 alleles) 43.6% (41/94) (data unavailable for one patient). The major *HBB**S haplotype allele
255 was Bantu (Bantu 127; Benin 32; Senegal 26; other 4), and the major genotype
256 Bantu/Bantu (Bantu/Bantu (n = 54), Bantu/[Other than Bantu] (n = 19) and [Other than
257 Bantu]/[Other than Bantu] (n = 22)).

258 Phenotypic characterization of the infection status during follow-up, was determined by
259 standard procedures and collected from the patients' hospital records. The 10th revision of
260 the International Statistical Classification of Diseases and Related Health Problems (ICD10)
261 <http://apps.who.int/classifications/icd10/browse/2016/en>, was used to classify infectious
262 episodes. The latter were observed within vasoocclusive crisis (VOC) and between crises
263 (steady-state) (IC) (Fig.2 to 3). The majority of the patients, 76% (72/95), were regularly
264 under prophylactic antibiotic treatment and virtually all of them completed an extended

265 vaccination program. In a total of 1351 observations (484 during crisis and 867 in steady-
266 state), 527 infectious episodes, 722 observations without infection and 102 observations
267 without reference to infection were reported (Fig.2). Reference to the etiologic agent of
268 infection was limited to the clinically more severe episodes. From the available clinical data
269 however, viral infections appeared characteristic of the inter-crisis period whereas during
270 vasoocclusive crisis (VOC) the distribution between infections caused by viruses and those
271 by bacteria was more equitable. Moreover, a viral etiology also appeared characteristic of
272 the respiratory infections (Fig. 3).

273

274 *Hemolytic component*

275 A hemolytic component derived by principal component analysis (PCA) with 3 degrees of
276 freedom from steady-state serum lactate dehydrogenase and total bilirubin, and peripheral
277 blood reticulocyte count in the 95 SCA patients had a mean of 0 (SD=1.23) and predicted
278 51% of the variation among all three variables (Eigenvalue=1.51). These results are
279 comparable to those obtained by Nouraie and collaborators (2013) using a PCA with 4
280 degrees of freedom in 415 SCA patients, predicting 55% of the variation among all four
281 variables (Eigenvalue=2.20), having a mean of 0 (SD=1.50).

282 The mean hemolytic component value was significantly lower with the alpha-thalassemia
283 genotype ($p = 0.023$) as has been previously reported by Nouraie and collaborators (2013).
284 However, we neither observed a significant association with gender ($p = 0.922$), with a
285 mainly pre puberty aged population, nor a positive correlation with increasing percentage of
286 fetal hemoglobin, prior to hydroxyurea treatment, or hydroxyurea treatment (Fig. S2A). This
287 may possibly be due to the fact that we did not study the fetal hemoglobin levels posterior
288 to hydroxyurea treatment but used the administration of hydroxyurea, in 18% (17/95) of the

289 patients as a gross, approximation for higher fetal hemoglobin levels. However, a significant
290 negative correlation with increasing percentage of steady state hemoglobin was observed
291 ($p = 0.001$) (Fig. S2B).

292

293 *TLR2 genotype and haplotype analysis*

294 Allele frequencies in the study population, included, for the -196 to -174 indel, 0.268
295 (51/190) and 0.732 (139/190) for the Del and Ins alleles, respectively, for the SNP
296 rs4696480, c.-16934A>T 0.426 (81/190) and 0.574 (109/190) for the A and T alleles,
297 respectively, and for the (GT)_n STR polymorphism 0,121 (23/190) and 0.879 (167/190) for
298 the S and Other (M or L) alleles respectively.

299 *TLR2* haplotype reconstruction data is summarized in Table 1. Of the eight haplotypes
300 identified, three had frequencies $\geq 5\%$, namely Hap3 (49%), Hap5 (18%) and Hap1 (17%).
301 These three haplotypes have in common a number of GT repeats ≥ 17 (Alleles M or L), a
302 characteristic also shared by Hap7.

303

304 *Genotype-to-phenotype association studies*

305 *Data analysis strategy*

306 Genotype-to-phenotype association studies followed a data analysis strategy (Fig. S1). The
307 stepwise approach included an extensive exploratory method used for subphenotype
308 selection based on SNP analysis using the Open-source whole genome association v1.07
309 program (Purcell et al., 2007), followed by exploratory methods to detect statistical
310 tendencies based on descriptive statistics, in the analysis of means of proportions, or the

311 Fisher's exact test, in the analysis of dichotomous traits. Finally, a longitudinal analysis of
312 dichotomous and quantitative traits using a quasipoisson regression model offered higher
313 statistical power and allowed the prediction of correlated dependent variables.

314 *Phenotype selection*

315 Six subphenotypes of the patients' infectious status, were originally defined for statistical
316 testing: (A) "number of respiratory infections during the patient's follow-up"; (B) "severe
317 bacterial infections at least once during the patient's follow-up"; (C) "number of all infectious
318 episodes during the patient's follow-up"; (D) "number of infectious episodes without
319 hospitalization during the patient's follow-up"; (E) "number of infectious episodes with
320 hospitalization during the patient's follow-up"; (F) "number of infections other than
321 respiratory during the patient's follow-up". The number of infections per time at risk was
322 measured in person*years considering the "time at risk" equal to the time under observation
323 during the follow-up. The phenotypic group (B) was the only originally constructed as a
324 binary trait. The following bacterial infections, affecting a total of nine patients in spite of
325 antibiotic prophylaxis, were observed for the phenotypic group (B): Buruli ulcer and
326 salmonellosis (observed in the same patient), pneumococcal meningitis, pyodermitis,
327 bacteremia and osteomyelitis.

328 Preliminary subphenotype selection was carried out using the PLINK v1.07 program for
329 genotype to phenotype association analysis of the *TLR2*SNP rs4696480, c.-16934A>T. The
330 allelic, recessive and dominant inheritance models were tested and the Cochran-Armitage
331 trend test performed (Table 2). Statistically significant results (without corrections for
332 multiple comparisons) were only obtained for the SNP's association with the subphenotype
333 (A)"number of respiratory infections during the patient's follow-up".

334 *Statistical analysis of subphenotype (A) “number of respiratory infections during the*
335 *patient’s follow-up”*

336 Descriptive statistics, carried out on the single infectious subphenotype, showed a
337 statistically significant result in the preliminary analysis, i.e. subphenotype (A) “number of
338 respiratory infections during the patient’s follow-up” (Table S2). For this we compared the
339 average rates of respiratory infections for all the patients (number of respiratory infections/
340 time at risk), in the presence and in the absence of the *TLR2* variants or their haplotypes. If
341 the presence of the genotype was associated with an increase in the incidence of
342 respiratory infections (positive statistical effect), this indicated a deleterious genotype with a
343 potentially unfavorable clinical effect. The descriptive statistics suggested a number of
344 deleterious tendencies, including for SNP rs4696480, c.-16934, T allele and the -196 to -
345 174 Del allele.

346 The association, of the subphenotype (A) “number of respiratory infections during the
347 patient’s follow-up” with the *TLR2* variants and their haplotypes, was then statistically tested
348 using a quasipoisson regression model. No statistically significant results were detected
349 with the univariable model, *Phenotype ~ Genotype*. With the multivariable model (see
350 methods), statistical confirmation of the suggestive results from the descriptive statistics
351 was obtained only for *TLR2* SNP rs4696480, c.-16934A>T (Table3). Risk estimates
352 obtained as a result of the allelic test showed that, in relation to the reference group (allele
353 A), the presence of allele T caused a deleterious effect resulting in an increase of the
354 incidence rate ratio ($p= 0.00903$). A protective effect of allele A was observed in
355 homozygosity ($p = 0.01328$). For allele T (allelic model) and allele A in homozygosity
356 (genotypic model) the deleterious effects of the covariables 3.7-kb alpha-thalassemia and
357 fetal hemoglobin levels (prior to hydroxyurea treatment) were also observed, whereas the

358 steady state hemoglobin level was protective. The Bonferroni correction for multiple
359 comparisons of these results however did not confirm the statistical significance of the
360 increase of the incidence rate ratio in the number of respiratory infections in the presence of
361 allele T (adjusted $\alpha = 0.001667$).

362 Using the quasipoisson regression analyses, no statistically significant associations were
363 obtained for this phenotype in relation to haplotypes.

364 *Statistical analysis of subphenotype (B)* “severe bacterial infections at least once during the
365 patient’s follow-up”

366 Although the PLINK analysis did not reveal a statistically significant association of the SNP
367 rs4696480, c.-16934A>T, with the subphenotype (B), “severe bacterial infections at least
368 once during the patient’s follow-up”, we considered it relevant to perform the Fisher’s exact
369 test and quasipoisson regression model on the *TLR2* variants and their haplotypes for this
370 rare phenotype (frequency < 0.10) (Tables 4 to 7), since this SNP has been reported
371 associated with severe bacterial infections (Rostane et al., 2012).

372 For the individual *TLR2* variants, no statistically significant results were detected with the
373 quasipoisson regression univariable model, *Phenotype* ~ *Genotype*. With the multivariable
374 model (see methods), statistically significant results were obtained for the SNP rs4696480,
375 c.-16934A>T (Table 4), and the -196 to -174 indel (Table 5). A deleterious effect of the A/T
376 genotype ($p = 0.00311$) was observed along with that of the covariable steady state
377 leukocyte count, whereas the covariables hemolytic component and fetal hemoglobin levels
378 (prior to hydroxyurea treatment) were protective. A protective effect for the -196 to -174 Ins
379 allele ($p = 0.0308$) and a deleterious effect of the Del/Del ($p = 0.0381$) genotype was also
380 observed, along with the protective effects of the covariables hemolytic component and

381 fetal hemoglobin levels (prior to hydroxyurea treatment). These results however did not
382 resist the Bonferroni correction for multiple comparisons (adjusted $\alpha = 0.001471$).

383 Considering haplotype analysis, the Fisher's exact test revealed a deleterious effect of
384 Hap7 (OR = 24.29; 95%CI 1.95-302.20; $p = 0.023$) (Table S3). However, the result did not
385 resist the correction for multiple comparisons. Moreover, the risk estimates obtained
386 assuming a univariable quasipoisson regression model (*Phenotype ~ Genotype*) ($p =$
387 0.0060) did not resist the Bonferroni correction (adjusted $\alpha = 0.001471$). However, the risk
388 estimates obtained using the multivariable model (see methods) (Table 6) resisted the
389 Bonferroni correction for multiple comparisons. Results showed that, in relation to the
390 reference group [Other than Hap7], the presence of Hap7 in the allelic model caused an
391 increase of the incidence rate ratio ($p = 0.00145$). On the contrary, in relation to
392 Hap7/[Other than Hap7] (no individuals presented the Hap7/Hap7 genotype), the absence
393 of Hap7 in the homozygous genotype [Other than Hap7]/[Other than Hap7] caused a
394 reduction of the incidence rate ratio ($p = 0.00145$).

395 With the quasipoisson regression model, a statistically significant association independent
396 of genotype was also obtained for the phenotype (B), "severe bacterial infections at least
397 once during the patient's follow-up", relative to the hemolytic component (Table 7). An
398 increase in one hemolytic component unit caused a decrease of 82% in the incidence rate
399 ratio ($p = 0.000366$). This result was positively influenced by the covariables fetal
400 hemoglobin levels (prior to hydroxyurea treatment) and 3.7-kb alpha-thalassemia.

401

402

403 **Discussion**

404 The clinical course of SCA is an alternation of periods of steady state and painful VOCs.
405 VOCs ultimately result from the reduced solubility of deoxygenated HbS, its subsequent
406 polymerization and the formation of the characteristic sickled red blood cells resulting in life
407 threatening pathophysiological changes (Kaul et al., 1996). These changes contribute to a
408 predisposition to infections, which constitute the major cause of morbidity and mortality in
409 these patients. Moreover, infective factors may in themselves trigger VOCs (Ahmed, 2011).
410 In developed countries, particular care is taken to reduce the infectious episodes through
411 extensive immunization and the use of prophylactic or therapeutic antimicrobial agents. The
412 pediatric patients in our cohort were under close and regular medical supervision and
413 subject to comprehensive preventive or curative measures. Nevertheless, severe and
414 recurrent infectious episodes were reported.

415 Both environmental and genetic factors appear to accrete in the determination of sickle cell
416 disease phenotypes (Driss et al., 2009). Genotype-phenotype association studies have
417 been conducted to identify genetic modulators of SCA phenotypes (Steinberg, 2008;
418 Steinberg and Sebastiani 2012), helping to understand the environmental and physiological
419 interactions at play in the onset, course and outcome of the disease.

420 The role of TLRs against invading organisms has been the subject of several literature
421 reviews (O'Neill et al., 2009; Lin et al., 2012; de Oliveira Nascimento et al., 2012;
422 Medvedev, 2013). These type I transmembrane glycoproteins belong to the so-called
423 pattern recognition receptors (PRRs) of the innate immune system. They are expressed by
424 macrophages, neutrophils, dendritic cells (DCs), natural killer (NK) cells, and epithelial and
425 endothelial cells (Kawai and Akira, 2011; Hippenstiel et al., 2006) and recognize
426 evolutionarily conserved molecular structures essential to pathogen survival. Amongst the
427 ten members of the human TLR family, TLR2 is a particularly pluripotent receptor due to

428 dimer formation with TLR1 and TLR6 in addition to interaction with non-TLR receptors
429 (Kawai and Akira, 2005). It is therefore not surprising that susceptibility to infections of
430 diverse etiologies has been associated to *TLR2* and its genetic variants.

431 In this work, a longitudinal study on a cohort of 95 SCA pediatric patients, was the
432 background for analyzing the genetic association of six infectious subphenotypes and three
433 loci in the noncoding regions of the *TLR2* gene, namely the -196 to -174,22 bp deletion in
434 the first untranslated exon, the promoter SNP rs4696480, c.-16934A>T (g.153685974T>A),
435 and the (GT)_n STR polymorphism in the second intron.

436 A stepwise approach to the genotype-to-phenotype association statistical analysis allowed
437 firstly, in preliminary analysis of the *TLR2* SNP c.-16934A>T variants, the identification of a
438 relevant infectious subphenotype. This was followed by the observation of deleterious or
439 protective tendencies of the SNP, indel and STR variants and their haplotypes in cross
440 sectional analyses relative to the selected subphenotype. Finally, the use of a quasipoisson
441 regression model optimized the statistical analysis through the incorporation of covariable
442 effects.

443 In the preliminary genotype-to-phenotype association analysis using the PLINK program a
444 potential association of the average rates per patient of the subphenotype “number of
445 respiratory infections during the patient’s follow-up” with *TLR2* SNP c.-16934A>T was
446 identified. This analysis was followed by descriptive statistics on this subphenotype,
447 comparing the average rates of respiratory infections for all the patients (number of
448 respiratory infections/ time at risk in person*year) with the variants and their haplotypes.
449 The trends observed in the preliminary and descriptive analyses were statistically confirmed
450 in a quasipoisson regression model for the locus variants. The deleterious effect of the T

451 allele as opposed to a protective effect of the A/A genotype in relation to respiratory
452 infections was thus statistically confirmed.

453 In our cohort, virus infections appear to be more frequent than bacterial in the episodes of
454 respiratory infections. In an extensive review Wu and Chu (2009) document on how the
455 host's capability to recognize microbes can influence susceptibility to asthma. Certain viral
456 infections including those affecting the respiratory tract, such as RSV are strongly
457 associated with asthma prevalence in children. Moreover, as reviewed by Klaassen and
458 collaborators (2013), the minor allele T of *TLR2*c.-16934A>T was related to childhood
459 asthma and current asthma symptoms in 3 out of 5 studies (Eder et al., 2004; Kormann et
460 al., 2008; Bottema et al., 2010; Kerkhof et al., 2010; Miedema et al., 2012). Therefore, our
461 results concerning the possible implication of this variant in the susceptibility to infections of
462 the respiratory tract, mainly of viral etiology, do not appear to contradict these previous
463 reports.

464 The analysis of the occurrence of at least one severe bacterial infection during the patient's
465 follow-up, was also tested relative to *TLR2* SNP c.-16934A>T, prompted by a previous
466 study showing that the homozygous genotypes A/A and T/T were associated with severe
467 bacterial infections in sickle cell patients whereas the heterozygous A/T genotype was more
468 frequent in patients without infections ($p < 0.001$) (Rostane et al., 2012). However, contrary
469 to the reported results (Rostane et al., 2012), we observed a deleterious effect of the
470 heterozygous A/T genotype in patients with severe bacterial infections.

471 The descriptive analyses in relation to the number of respiratory infections suggested a
472 deleterious effect of the -196 to -174 Del/Del homozygote. Using the quasipoisson
473 regression model, this was not statistically confirmed for respiratory infections. As concerns
474 the subphenotype of a severe bacterial infection at least once during the patient's follow-up,

475 the quasipoisson regression revealed the tendencies for a protective effect of the Ins allele
476 and a deleterious effect of the Del/Del genotype. The Del allele has been implicated in the
477 occurrence of infections of various etiologies (Nischalke et al., 2011, 2012; Velez et al.,
478 2010; Greene et al., 2012; Hishida et al., 2010; Zeng et al., 2011). Functionally, the -196 to
479 -174 Del/Del genotype has been reported to show decreased transcriptional activity of the
480 *TLR2* gene (Nischalke et al., 2012; Noguchi et al., 2004).

481 Concerning the (GT)_n STR polymorphism in the *TLR2* intron 2, the three major *TLR2*
482 haplotypes with frequencies $\geq 5\%$ (Hap1, Hap3, Hap5) have in common a number of
483 repeats $n \geq 17$ (Alleles M or L). This characteristic was also shared by Hap7, a less
484 frequent haplotype in our study population (3%). The sub-classification of the alleles of this
485 polymorphic marker into three subclasses (Alleles S, M and L) has unveiled a probable
486 functional impact since lower expression of *TLR2* due to weaker promoter activity was
487 reported in individuals with smaller repeat numbers (S allele) (Yim et al., 2004, 2006, 2008,
488 Lee et al., 2006; Folwaczny et al., 2011).

489 The Hap7 haplotype ([Del]-T-[M or L]) was relevant to the infectious subphenotypes in SCA
490 considered in this study. The deleterious effect of Hap7 was observed through a statistically
491 significant increase in the incidence rate in severe bacterial infections in SCA, resisting the
492 Bonferroni correction for multiple testing, with no influence of the covariables on the risk
493 estimates. This haplotype presents both the -196 to -174 22 bp deletion, localized in the
494 *TLR2* first untranslated exon, that was been associated with susceptibility to infections of
495 various etiologies (Nischalke et al., 2011, 2012; Velez et al., 2010; Greene et al., 2012),
496 and the SNP rs4696480 T allele, in the *TLR2* promoter, associated in a preliminary study to
497 susceptibility to pulmonary bacterial infections in sickle cell disease (Rostane et al., 2012).
498 Relative to the (GT)_n STR polymorphism in the *TLR2* intron 2, it would appear to have a

499 'normal' promoter activity. The observed effect of Hap7 was independent of statistically
500 significant effects of the covariables.

501 The high statistical effects observed in our association studies were accompanied by wide
502 confidence intervals. This lack of precision attests to the low number of cases in our study
503 that could be resolved in future studies by contemplating a higher number of cases.

504 In earlier studies, TLR2 signaling was believed to have a role in mediating proinflammatory
505 cytokines in a type I IFN independent inflammatory response rather than a specific type I
506 IFN antiviral response as is the case for TLR3, TLR4, TLR7, and TLR9 stimulation (Akira et
507 al., 2006; Thompson and Iwasaki, 2008). However, in more recent investigations, a type I
508 IFN response to viral infection, not requiring nucleic acids as ligands, has been reported
509 (Kawai and Akira, 2008; Barbalat et al., 2009; Takeuchi and Akira, 2010). In response to
510 viral ligands but not bacterial ligands, and depending on the cell type, TLR2 was shown to
511 induce either the production of type I IFN (in inflammatory monocytes) or of inflammatory
512 cytokines, but not type I IFN (in macrophages and other types of DCs). Vaccinia virus
513 stimulation of inflammatory monocytes induced TLR2 localization to endosomal
514 compartments and type I IFN production (Barbalat et al., 2009). As most of the nucleic acid
515 sensing TLRs (TLR3, TLR7, TLR8 and TLR9) are localized to the endosome, this model is
516 consistent with the paradigm that the endosomal system plays a pivotal role in virus
517 recognition and signaling (Liu et al., 2011).

518 Differences between the *TLR2* variants implicated in susceptibility/resistance to viral versus
519 bacterial infections could be expected from genetic variants located in the coding region of
520 the gene, due to the differences in the PAMPs between these two classes of pathogens.
521 This is predictable by the existing dichotomy between type I IFN mediation of antiviral
522 responses, i.e. whereas type I IFNs are crucial to survival from acute viral infection, the

523 toxicity of type I IFN alters innate and adaptive immune defenses against other
524 opportunistic infections including of bacterial origin (Trinchieri, 2010). As such, the
525 protection against viral infections through this innate defense mechanism may increase the
526 risk of bacterial opportunistic infections. Likewise, the genetic variants or haplotypes of the
527 genetic determinants of the innate immune system that protect against one class of
528 pathogens may be deleterious in relation to another. In our study however, the variants
529 analyzed were in non coding regions and most likely implicated in transcriptional activity.
530 The phenotype-to-genotype association results from our study indicated that these variants
531 worked in the same manner relative to susceptibility/resistance whether the infections
532 analyzed were of bacterial origin or primarily viral. Viral etiology was assigned to most of
533 the respiratory infections in our study. In pediatric ages, viral infections are frequent and
534 almost always benignant. However, from these results their frequency may be an indicator
535 of susceptibility to more severe bacterial infections. It would be pertinent to evaluate the IFN
536 responses in the presence of the *TLR2* variants and haplotypes in our study, for example in
537 a human relevant cell/tissue *ex vivo* model. This result is particularly pertinent in the light of
538 considerations on TLRs as a class of targets for therapies against infectious diseases
539 (O'Neill et al., 2009).

540 The hemolytic component reflects direct markers of intravascular hemolysis in patients and
541 allows for adjusted analysis of associations between hemolytic severity and clinical
542 phenotypes (Nouraie et al., 2013). The levels of cell-free hemoglobin concentrations, a
543 direct marker of hemolysis, not being available to us we worked with a hemolytic
544 component derived by PCA from three proximal hemolytic markers namely steady state
545 serum lactate dehydrogenase and total bilirubin, and reticulocyte count. A statistically
546 significant association of the subphenotype “severe bacterial infections at least once during
547 the patient’s follow-up” was obtained with the hemolytic component, regardless of the

548 patient *TLR2* genotype. An increase in the hemolytic component unit caused a decrease in
549 the incidence rate of these infections. This is an unexpected observation. SCA is a
550 hemolytic disease in which there is premature destruction of red blood cells. In a recent
551 review Orf and Cunningham (2015) discussed how hemolysis, by liberating the heme moiety
552 of hemoglobin and its iron, an essential element for both the host and the pathogen,
553 induces pathophysiological changes that increase the risk of bacterial infection. The
554 mechanisms by which this occurs are not fully understood and appear to differ amongst the
555 different causes of hemolysis. However, these authors suggested that a general
556 mechanism may be the impairment of neutrophil oxidative burst that has also been
557 observed in sickle cell disease. The covariable effects as revealed in the quasipoisson
558 regression model also need discussion. The covariables positively influencing this outcome
559 were fetal hemoglobin levels (prior to hydroxyurea treatment) and 3.7-kb alpha-
560 thalassemia. Fetal hemoglobin is a well known modulator of sickle cell disease severity, the
561 protective effect being due to a higher oxygen binding capacity than that of adult
562 hemoglobin A (Bridges et al., 1996). Alpha-thalassemia, another modulator reduces the
563 formation of dense and irreversibly deoxyHbS sickled cells through the reduction in overall
564 hemoglobin concentration, as a consequence of the absent alpha genes (Embury et al.,
565 1984). Moreover, alpha-thalassemia has been associated with reduced hemolysis as
566 measured by total steady state bilirubin and reticulocyte count (Coelho et al., 2014).
567 Supporting the hemolytic component's relationship with the intensity of total hemolysis as
568 previously reported (Driss et al., 2009), in our study the mean hemolytic component value
569 was significantly lower with the 3.7-kb alpha-thalassemia genotypes. Moreover, a
570 statistically significant negative association between the hemolytic component and steady
571 state hemoglobin levels was observed. Therefore although hemolysis and 3.7-kb alpha-
572 thalassemia are two factors generally working in opposition in SCA, in relation to the risk of

573 infection they appear to contribute jointly towards its reduction. These results once more
574 denote the complex genetic architecture in the SCA hemolytic process.

575 In our study, the hemolytic component was shown to be an important modulator of severe
576 bacterial infection predisposition along with other more commonly investigated modulators
577 of SCA. Our results showed a positive influence of the covariables hemolytic component
578 and fetal hemoglobin levels (prior to hydroxyureia treatment) to the protective effect of the
579 *TLR2* SNP c.-16934A>T allele A as well as the -196 to -174 indel Ins/Del genotype against
580 severe bacterial infection.

581 The hemolytic component had no *TLR2* genotype independent effect on the advent of
582 respiratory infections as a whole, which in our study are more frequently of viral origin. A
583 statistical tendency that pointed to the deleterious effect of *TLR2* SNP c.-16934A>T allele T
584 was observed, resulting in an increase in the incidence rate of respiratory infections.

585 Contrary to what was observed with steady state hemoglobin levels, both covariables fetal
586 hemoglobin levels (prior to hydroxyureia treatment) and 3.7-kb alpha-thalassemia acted to
587 increase the incidence rate of these infections. The differences observed between the
588 protective effects of 3.7-kb alpha-thalassemia against severe bacterial infection contrasts
589 with its deleterious effect against respiratory infections where it appears to favor an
590 increase in the incidence rate. Although this latter result is unexpected, negative epistasis
591 between sickle hemoglobin and 3.7-kb alpha-thalassemia or between sickle hemoglobin
592 and fetal hemoglobin levels resulting in an increased risk of malaria infection have been
593 reported in epidemiologic studies in Kenya (Mmbando et al., 2015).

594

595 **Conclusion**

596 This genotype-to-phenotype association study revealed important aspects relative to
597 statistical methodology and the pathophysiology of infection in pediatric SCA.

598 Methodologically, our study revealed the usefulness of time based cohorts in genotype-to-
599 phenotype association studies of infectious diseases at a time when whole genome cross-
600 sectional analysis is in vogue.

601 Our results evidenced the deleterious effect of the Hap7 haplotype ([Del]-T-[M or L]) in the
602 incidence rate in severe bacterial infections in pediatric SCA. This haplotype presents both
603 the -196 to -174 22 bp deletion, localized in the *TLR2* first untranslated exon, and the SNP
604 rs4696480 T allele, in the *TLR2* promoter, that have previously been associated with
605 susceptibility to infections of various etiologies.

606 Another interesting aspect was the analysis of a hemolytic component. Hemolysis is a
607 major pathophysiological feature of SCA. Our results show its relevance in relation to
608 severe bacterial infections in a positive epistasis with $-\alpha^{3.7}$ thalassemia. An unexpected
609 dichotomous relationship with $-\alpha^{3.7}$ thalassemia, suggested that a role in the protection
610 against severe bacterial might favor recurrent viral infections, highlighting the complexity of
611 the SCA hemolytic process in relation to infection.

612 These results should be further explored by replicating the findings in other cohorts to
613 obtain definite conclusions; conducting prospective studies to prove the usefulness of *TLR2*
614 information; applying detailed analysis of *TLR2* loci to pick-up the most useful markers to be
615 used in future clinical practice.

616 As we further and more deeply understand the genetic basis and the pathophysiological
617 mechanisms involved, targeted and personalized interventions may come to reduce the
618 morbidity and mortality attributable to infection in these particularly vulnerable SCA patients.

619

620 **Conflict of interests**

621 The authors declare no conflict of interest.

622

623 **Acknowledgements**

624 We are grateful to the patients and their families. We also thank Dominique Labie for having
625 suggested this topic of research. We thank INSA's "Unidade de Tecnologia e Inovação" as
626 well as Andreia Coelho and Emanuel Ferreira for technical support. We thank Eric David
627 Bosne for insight in PCA. We also thank Audrey V. Grant, Baltazar Nunes, Paula Faustino
628 and Vera C.M. David for their helpful suggestions in the elaboration of the study. This study
629 was carried out with financial support from FCT/MEC through national funds and co-
630 financed by FEDER, under the Partnership Agreement PT2020, in the project with
631 reference UIDMULTI/00211/2013, and was partially funded by FCT grants
632 PIC/IC/83084/2007 and the Centro de Investigação em Genética Molecular Humana
633 (CIGMH).

634

635

636 **Author contribution**

637 Conceived and designed the experiments: S David, A Dias, A Morais, J Lavinha.

638 Suggested methodologies: S David, A Dias, A Morais, A Sakuntabhai, J Lavinha.

639 Genotyped: S David. Reorganized the database: S David. Analyzed the data: S David, P

640 Aguiar. Carried out the investigation: S David. Provided the resources: A Dias, A Morais, J

641 Lavinha. Wrote the paper: S David. Reviewed the paper: P Aguiar, J Lavinha, A
642 Sakuntabhai. All authors approved the final version of the paper.

643

644 **Ethical approval:** All procedures performed in studies involving human participants were in
645 accordance with the ethical standards of the institutional and/or national research
646 committee and with the 1964 Helsinki declaration and its later amendments or comparable
647 ethical standards.

648

649

650 **References**

651

652 Ahmed, SG (2011) The role of infection in the pathogenesis of vaso-occlusive crisis in
653 patients with sickle cell disease. *Mediterranean journal of hematology and infectious*
654 *diseases* 3(1): p.2011028

655 Akira, S, Uematsu, S, Takeuchi, O (2006) Pathogen recognition and innate immunity. *Cell*
656 124(4): 783-801

657 Alter, A, Grant, A, Abel, L, Alcaïs, A, Schurr, E (2011) Leprosy as a genetic disease.
658 *Mammalian genome* 22(1-2): 19-31

659 Bagheri, V, Askari, A, Arababadi, MK, Kennedy, D (2014) Can Toll-Like Receptor (TLR) 2
660 be considered as a new target for immunotherapy against hepatitis B infection?.
661 *Human immunology* 75(6): 549-554

662 Barbalat, R, Lau, L, Locksley, RM, Barton, GM (2009) Toll-like receptor 2 on inflammatory
663 monocytes induces type I interferon in response to viral but not bacterial ligands.
664 Nature immunology 10(11): 1200-1207

665 Bland, JM, Altman, DG (1995) Multiple significance tests: the Bonferroni method. Bmj
666 310(6973): 170

667 Bottema, RW, Kerkhof, M, Reijmerink, NE, Thijs, C, Smit, HA, van Schayck, CP,
668 Brunekreef, B, van Oosterhout, AJ, Postma, DS, Koppelman, GH (2010) Gene-gene
669 interaction in regulatory T-cell function in atopy and asthma development in
670 childhood. Journal of Allergy and Clinical Immunology 126(2): 338-346

671 Bridges, KR, Barabino, GD, Brugnara, C, Cho, MR, Christoph, GW, Dover, G, Ewenstein,
672 BM, Golan, DE, Guttmann, CR, Hofrichter, J, Mulkern, RV (1996) A multiparameter
673 analysis of sickle erythrocytes in patients undergoing hydroxyurea therapy. Blood
674 88(12): 4701-4710

675 Cai, MS, Li, ML and Zheng, CF (2012) Herpesviral infection and Toll-like receptor 2. Protein
676 & cell 3(8): 590-601

677 Casanova, JL, Abel, L, Quintana-Murci, L (2011) Human TLRs and IL-1Rs in host defense:
678 natural insights from evolutionary, epidemiological, and clinical genetics. Annual
679 review of immunology 29: 447-491

680 Chen, J, Ng, MML, Chu, JJH (2015) Activation of TLR2 and TLR6 by dengue NS1 protein
681 and its implications in the immunopathogenesis of dengue virus infection. PLoS
682 Pathog 11(7): e1005053

683 Chen, YC, Hsiao, CC, Chen, CJ, Chin, CH, Liu, SF, Wu, CC, Eng, HL, Chao, TY, Tsen, CC,
684 Wang, YH, Lin, MC (2010) Toll-like receptor 2 gene polymorphisms, pulmonary
685 tuberculosis, and natural killer cell counts. *BMC medical genetics* 11(1): 17

686 Coelho, A, Dias, A, Morais, A, Nunes, B, Ferreira, E, Picanço, I, Faustino, P, Lavinha, J
687 (2014) Genetic variation in CD36, HBA, NOS3 and VCAM1 is associated with chronic
688 haemolysis level in sickle cell anaemia: a longitudinal study. *European journal of*
689 *haematology*, 92(3): 237-243

690 de Oliveira Nascimento, L, Massari, P, Wetzler, LM (2012) The role of TLR2 in infection and
691 immunity. *Frontiers in immunology*, 3: 79

692 Driss, A, Asare, KO, Hibbert, JM, Gee, BE, Adamkiewicz, TV, Stiles, JK (2009) Sickle cell
693 disease in the post genomic era: a monogenic disease with a polygenic phenotype.
694 *Genomics insights* 2: 23

695 Eder, W, Klimecki, W, Yu, L, von Mutius, E, Riedler, J, Braun-Fahrländer, C, Nowak, D,
696 Martinez, FD, ALEX Study Team (2004) Toll-like receptor 2 as a major gene for
697 asthma in children of European farmers. *Journal of allergy and clinical immunology*
698 113(3): 482-488

699 Embury, SH, Clark, MR, Monroy, G, Mohandas, N (1984) Concurrent sickle cell anemia and
700 alpha-thalassemia. Effect on pathological properties of sickle erythrocytes. *Journal of*
701 *Clinical Investigation*, 73(1): 116

702 Esposito, S, Zampiero, A, Pugini, L, Tabano, S, Pelucchi, C, Ghirardi, B, Terranova, L,
703 Miozzo, M, Mosca, F, Principi, N (2014) Genetic polymorphisms and sepsis in
704 premature neonates. *PloS one* 9(7): e101248

705 Folwaczny, M, Glas, J, Tonenchi, L, Török, HP (2011) Microsatellite GT polymorphism in
706 intron 2 of human Toll-like receptor (TLR) 2 gene and susceptibility to periodontitis.
707 Clinical oral investigations 15(3): 435-441

708 Greene, JA, Sam-Agudu, N, John, CC, Opoka, RO, Zimmerman, PA, Kazura, JW (2012)
709 Toll-like receptor polymorphisms and cerebral malaria: TLR2 Δ 22 polymorphism is
710 associated with protection from cerebral malaria in a case control study. Malaria
711 journal 11(1): 47.

712 Hawn, TR, Scholes, D, Li, SS, Wang, H, Yang, Y, Roberts, PL, Stapleton, AE, Janer, M,
713 Aderem, A, Stamm, WE, Zhao, LP (2009) Toll-like receptor polymorphisms and
714 susceptibility to urinary tract infections in adult women. PloS one 4(6): e5990

715 Hernandez, JC, Giraldo, DM, Paul, S, Urcuqui-Inchima, S (2015) Involvement of neutrophil
716 hyporesponse and the role of Toll-like receptors in human immunodeficiency virus 1
717 protection. PloS one 10(3): e0119844

718 Hippenstiel, S, Opitz, B, Schmeck, B, Suttorp, N (2006) Lung epithelium as a sentinel and
719 effector system in pneumonia—molecular mechanisms of pathogen recognition and
720 signal transduction. Respiratory research 7(1): 97

721 Hishida, A, Matsuo, K, Goto, Y, Naito, M, Wakai, K, Tajima, K, Hamajima, N (2010) No
722 associations of Toll-like receptor 2 (TLR2)-196 to-174del polymorphism with the risk of
723 Helicobacter pylori seropositivity, gastric atrophy, and gastric cancer in Japanese.
724 Gastric Cancer 13(4): 251-257

725 Ioana, M, Ferwerda, B, Plantinga, TS, Stappers, M, Oosting, M, McCall, M, Cimpoeu, A,
726 Burada, F, Panduru, N, Sauerwein, R and Doumbo, O (2012) Different patterns of

727 Toll-like receptor 2 polymorphisms in populations of various ethnic and geographic
728 origins. *Infection and immunity* 80(5): 1917-1922

729 Kaul, DK, Fabry, ME, Nagel, RL (1996) The pathophysiology of vascular obstruction in the
730 sickle syndromes. *Blood reviews* 10(1): 29-44

731 Kawai, T, Akira, S(2005) Pathogen recognition with Toll-like receptors. *Current opinion in*
732 *immunology* 17(4): 338-344

733 Kawai, T, Akira, S(2011) Toll-like receptors and their crosstalk with other innate receptors in
734 infection and immunity. *Immunity* 34(5): 637-650

735 Kerkhof, M, Postma, DS, Brunekreef, B, Reijmerink, NE, Wijga, AH, De Jongste, JC,
736 Gehring, U, Koppelman, GH(2010) Toll-like receptor 2 and 4 genes influence
737 susceptibility to adverse effects of traffic-related air pollution on childhood asthma.
738 *Thorax* 65(8): 690-697

739 Klaassen, EM, Thönissen, BE, van Eys, G, Dompeling, E, Jöbsis, Q(2013) A systematic
740 review of CD14 and toll-like receptors in relation to asthma in Caucasian children.
741 *Allergy, Asthma & Clinical Immunology* 9(1): 10

742 Kormann, MS, Depner, M, Hartl, D, Klopp, N, Illig, T, Adamski, J, Vogelberg, C, Weiland,
743 SK, von Mutius, E, Kabesch, M(2008) Toll-like receptor heterodimer variants protect
744 from childhood asthma. *Journal of Allergy and Clinical Immunology* 122(1): 86-92

745 Lee, EY, Yim, JJ, Lee, HS, Lee, YJ, Lee, EB, Song, YW(2006) Dinucleotide repeat
746 polymorphism in intron II of human Toll-like receptor 2 gene and susceptibility to
747 rheumatoid arthritis. *International journal of immunogenetics* 33(3): 211-215

748 Lin, YT, Verma, A, P Hodgkinson, C(2012) Toll-like receptors and human disease: lessons
749 from single nucleotide polymorphisms. *Current genomics* 13(8): 633-645

750 Liu, SY, Sanchez, DJ, Cheng, G(2011) New developments in the induction and antiviral
751 effectors of type I interferon. *Current opinion in immunology* 23(1): 57-64

752 Ma, X, Liu, Y, Gowen, BB, Graviss, EA, Clark, AG, Musser, JM(2007) Full-exon
753 resequencing reveals toll-like receptor variants contribute to human susceptibility to
754 tuberculosis disease. *PloS one* 2(12): e1318

755 Ma, Y, He, B (2014) Recognition of herpes simplex viruses: toll-like receptors and beyond.
756 *Journal of molecular biology* 426(6): 1133-1147

757 Medvedev, AE(2013) Toll-like receptor polymorphisms, inflammatory and infectious
758 diseases, allergies, and cancer. *Journal of Interferon & Cytokine Research* 33(9): 467-
759 484

760 Miedema, KGE, Tissing, WJE, Te Poele, EM, Kamps, WA, Alizadeh, BZ, Kerkhof, M, De
761 Jongste, JC, Smit, HA, De Pagter, AP, Bierings, M, Boezen, HM(2012)
762 Polymorphisms in the TLR6 gene associated with the inverse association between
763 childhood acute lymphoblastic leukemia and atopic disease. *Leukemia* 26(6): 1203-
764 1210

765 Misch, EA, Hawn, TR(2008) Toll-like receptor polymorphisms and susceptibility to human
766 disease. *Clinical Science* 114(5): 347-360

767 Mmbando, BP, Mgaya, J, Cox, SE, Mtatiro, SN, Soka, D, Rwezaula, S, Meda, E, Msaki, E,
768 Snow, RW, Jeffries, N, Geller, NL(2015) Negative Epistasis between sickle and foetal
769 haemoglobin suggests a reduction in protection against malaria. *PloS one* 10(5):
770 e0125929

771 Mukherjee, S, Ganguli, D, Majumder, PP(2014) Global footprints of purifying selection on
772 toll-like receptor genes primarily associated with response to bacterial infections in
773 humans. *Genome biology and evolution* 6(3): 551-558

774 Nischalke, HD, Berger, C, Aldenhoff, K, Thyssen, L, Gentemann, M, Grünhage, F,
775 Lammert, F, Nattermann, J, Sauerbruch, T, Spengler, U, Appenrodt, B(2011) Toll-like
776 receptor (TLR) 2 promoter and intron 2 polymorphisms are associated with increased
777 risk for spontaneous bacterial peritonitis in liver cirrhosis. *Journal of hepatology* 55(5):
778 1010-1016

779 Nischalke, HD, Coenen, M, Berger, C, Aldenhoff, K, Müller, T, Berg, T, Krämer, B, Körner,
780 C, Odenthal, M, Schulze, F, Grünhage, F(2012) The toll-like receptor 2 (TLR2)-196
781 to-174 del/ins polymorphism affects viral loads and susceptibility to hepatocellular
782 carcinoma in chronic hepatitis C. *International journal of cancer* 130(6): 1470-1475

783 Noguchi, E, Nishimura, F, Fukai, H, Kim, J, Ichikawa, K, Shibasaki, M, Arinami, T(2004) An
784 association study of asthma and total serum immunoglobulin E levels for Toll-like
785 receptor polymorphisms in a Japanese population. *Clinical & Experimental Allergy*
786 34(2): 177-183

787 Nouraie, M, Lee, JS, Zhang, Y, Kanias, T, Zhao, X, Xiong, Z, Oriss, TB, Zeng, Q, Kato, GJ,
788 Gibbs, JSR, Hildesheim, ME(2013) The relationship between the severity of
789 hemolysis, clinical manifestations and risk of death in 415 patients with sickle cell
790 anemia in the US and Europe. *Haematologica* 98(3): 464-472

791 Novis, CL, Archin, NM, Buzon, MJ, Verdin, E, Round, JL, Lichterfeld, M, Margolis, DM,
792 Planelles, V, Bosque, A(2013) Reactivation of latent HIV-1 in central memory CD4+ T
793 cells through TLR-1/2 stimulation *Retrovirology* 10(1): 119

794 O'Neill, LA, Bryant, CE, Doyle, SL(2009) Therapeutic targeting of Toll-like receptors for
795 infectious and inflammatory diseases and cancer. *Pharmacological reviews* 61(2):
796 177-197

797 Orf, K, Cunnington, A(2015) Infection-related hemolysis and susceptibility to Gram-negative
798 bacterial co-infection. *Frontiers in microbiology* 6: 666

799 Purcell, S, Neale, B, Todd-Brown, K, Thomas, L, Ferreira, MA, Bender, D, Maller, J, Sklar,
800 P, De Bakker, PI, Daly, MJ, Sham, PC(2007) PLINK: a tool set for whole-genome
801 association and population-based linkage analyses. *The American Journal of Human*
802 *Genetics* 81(3): 559-575

803 Reuven, EM, Ali, M, Rotem, E, Schwarzter, R, Gramatica, A, Futerman, AH, Shai, Y(2014)
804 The HIV-1 envelope transmembrane domain binds TLR2 through a distinct
805 dimerization motif and inhibits TLR2-mediated responses. *PLoS Pathog* 10(8):
806 e1004248

807 Rostane, H, Busson, M, Diallo, D(2012) Le risque d'infections bactériennes sévères au
808 cours de la maladie drépanocytaire est influencé par un polymorphisme du promoteur
809 du gène TLR-2. *Congrès Soc Franc Hématol 2012; Globule rouge et fer# 138.*
810 *Hématologie* 18(suppl 1): 26

811 Song, Z, Yin, J, Yao, C, Sun, Z, Shao, M, Zhang, Y, Tao, Z, Huang, P, Tong, C(2011)
812 Variants in the Toll-interacting protein gene are associated with susceptibility to sepsis
813 in the Chinese Han population. *Critical care* 15(1): R12

814 Steinberg, MH, Embury, SH(1986) Alpha-thalassemia in blacks: genetic and clinical
815 aspects and interactions with the sickle hemoglobin gene. *Blood* 68(5): 985-990

816 Steinberg, MH, Sebastiani, P(2012) Genetic modifiers of sickle cell disease. American
817 journal of hematology 87(8): 795-803

818 Steinberg, MH(2008) Sickle cell anemia, the first molecular disease: overview of molecular
819 etiology, pathophysiology, and therapeutic approaches. The Scientific World Journal
820 8: 1295-1324

821 Stephens, M, Donnelly, P(2003) A comparison of bayesian methods for haplotype
822 reconstruction from population genotype data. The American Journal of Human
823 Genetics 73(5): 1162-1169

824 Stephens, M, Smith, NJ, Donnelly, P(2001) A new statistical method for haplotype
825 reconstruction from population data. The American Journal of Human Genetics 68(4):
826 978-989

827 Tahara, T, Arisawa, T, Wang, F, Shibata, T, Nakamura, M, Sakata, M, Hirata, I, Nakano, H
828 (2007) Toll-like receptor 2–196 to 174del polymorphism influences the susceptibility of
829 Japanese people to gastric cancer. Cancer science 98(11): 1790-1794

830 Takeda, K, Kaisho, T, Akira, S(2003) Toll-like receptors. Annual review of immunology
831 21(1): 335-376

832 Takeuchi, O, Akira, S(2010) Pattern recognition receptors and inflammation. Cell 140(6):
833 805-820

834 Team, RC(2013) Vienna, Austria: R Foundation for Statistical Computing; 2013. R: A
835 language and environment for statistical computing

836 Thompson, JM, Iwasaki, A(2008) Toll-like receptors regulation of viral infection and
837 disease. Advanced drug delivery reviews 60(7): 786-794

838 Triantafilou, K, Eryilmazlar, D, Triantafilou, M(2014) Herpes simplex virus 2–induced
839 activation in vaginal cells involves Toll-like receptors 2 and 9 and DNA sensors DAI
840 and IFI16. American journal of obstetrics and gynecology 210(2): 122-e1

841 Trinchieri, G(2010) Type I interferon: friend or foe?. Journal of Experimental Medicinejem-
842 20101664

843 Velez, DR, Wejse, C, Stryjewski, ME, Abbate, E, Hulme, WF, Myers, JL, Estevan, R,
844 Patillo, SG, Olesen, R, Tacconelli, A, Sirugo, G(2010) Variants in toll-like receptors 2
845 and 9 influence susceptibility to pulmonary tuberculosis in Caucasians, African-
846 Americans, and West Africans. Human genetics 127(1): 65-73

847 Ver Hoef, JM, Boveng, PL(2007) Quasi-Poisson vs. negative binomial regression: how
848 should we model overdispersed count data?. Ecology 88(11): 2766-2772

849 Wu, Q, Chu, HW(2009) Role of infections in the induction and development of asthma:
850 genetic and inflammatory drivers. Expert review of clinical immunology 5(1): 97-109

851 Yim, JJ, Ding, L, Schäffer, AA, Park, GY, Shim, YS, Holland, SM(2004) A microsatellite
852 polymorphism in intron 2 of human Toll-like receptor 2 gene: functional implications
853 and racial differences. FEMS Immunology & Medical Microbiology 40(2): 163-169

854 Yim, JJ, Kim, HJ, Kwon, OJ, Koh, WJ(2008) Association between microsatellite
855 polymorphisms in intron II of the human Toll-like receptor 2 gene and nontuberculous
856 mycobacterial lung disease in a Korean population. Human immunology 69(9): 572-
857 576

858 Yim, JJ, Lee, HW, Lee, HS, Kim, YW, Han, SK, Shim, YS, Holland, SM(2006) The
859 association between microsatellite polymorphisms in intron II of the human Toll-like
860 receptor 2 gene and tuberculosis among Koreans. Genes and immunity 7(2): 150-155

861 Zeng, HM, Pan, KF, Zhang, Y, Zhang, L, Ma, JL, Zhou, T, Su, HJ, Li, WQ, Li, JY, Gerhard,
862 M, Classen, M(2011) Genetic variants of toll-like receptor 2 and 5, helicobacter pylori
863 infection, and risk of gastric cancer and its precursors in a chinese population. Cancer
864 Epidemiology and Prevention Biomarkers 20(12): 2594-602

865 Zhang, Y, Jiang, T, Yang, X, Xue, Y, Wang, C, Liu, J, Zhang, X, Chen, Z, Zhao, M, Li,
866 JC(2013) Toll-like receptor-1,-2, and-6 polymorphisms and pulmonary tuberculosis
867 susceptibility: a systematic review and meta-analysis. PloS one 8(5): e63357

868

869 **Table 1**870 Haplotype distribution of the three *TLR2* loci observed in the SCA pediatric patients (n = 95)

Name	Haplotype ¹ (two notations)	Frequency ²	S.E
Hap1	101 [Ins]-A-[M or L]	0,174	0,011
Hap2	100 [Ins]-A-[S]	0,028	0,006
Hap3	111 [Ins]-T-[M or L]	0,491	0,013
Hap4	110 [Ins]-T-[S]	0,048	0,010
Hap5	001 [Del]-A-[M or L]	0,182	0,012
Hap6	000 [Del]-A-[S]	0,039	0,008
Hap7	011 [Del]-T-[M or L]	0,030	0,001
Hap8	010 [Del]-T-[S]	0,008	0,007

871 ¹ Haplotypes were defined for all three loci detected in the *TLR2* using the PHASE, version
872 2.1, program [64,65]. Relative to the arbitrary position c.-18619, the genetic distance in bp
873 of the three loci was 5' – -196 to -174 22 bp nucleotide indel (untranslated exon 1) – SNP
874 rs4696480, c.-16934A>T (promoter) – the (GT)_n STR polymorphism (intron 2) – 3':
875 66/1685/18475 bp respectively. Using the PHASE program eight haplotypes were
876 identified: Hap1, 101 ([Ins]-A-[M or L]); Hap2, 100 ([Ins]-A-[S]); Hap3, 111 ([Ins]-T-[M or L]);
877 Hap4, 110 ([Ins]-T-[S]); Hap5, 001 ([Del]-A-[M or L]); Hap6, 000 ([Del]-A-[S]); Hap7, 011
878 ([Del]-T-[M or L]); Hap8, 010 ([Del]-T-[S]). For the *TLR2* STR, the following subdivision
879 based on the number of repeats (n) was used: Allele S, n ≤ 16; Allele M, 17 ≤ n ≤ 22, and
880 Allele L, n ≥ 23, which was further simplified as a binary variable S versus others (M or L)
881 (Yim et al., 2006).

882 ² Haplotype frequency from the present study

883 Note: For the indel: 1 (Ins) = insertion; 0 (Del) = deletion; for the SNP: 1 = T allele (major); 0
884 = A allele (minor); for the microsatellite (STR): 0 = S allele [(GT) ≤ 16], 1 = M or L alleles [(17
885 ≤ (GT) ≤ 22] and [(GT) ≥ 23], respectively)

886 Abbreviations: SCA, sickle cell anemia; S.E., standard error; SNP, single nucleotide
887 polymorphism; Hap, haplotype; Indel, insertion or deletion variation.

888

889 **Table 2**

890 Results of the preliminary genotype-to-phenotype association study of TLR2 SNP

891 rs4696480 in SCA pediatric patients (n = 95)

Phenotype ¹	TEST ²	p-value
A	ALLELIC	0.007167
A	TREND	0.009488
A	ADD	0.01152
B	ALLELIC	0.1805
B	TREND	0.1964
B	ADD	0.2056
C	ALLELIC	0.2226
C	TREND	0.2393
C	ADD	0.2416
D	ALLELIC	0.3874
D	TREND	0.4044
D	ADD	0.4056
E	REC	0.6686
F	ALLELIC	0.7261
F	TREND	0.7355
F	ADD	0.7356
E	ALLELIC	0.762
E	TREND	0.7702
E	ADD	0.7703
E	GEN	0.9098
E	DOM	0.9321

892 ¹ Infectious phenotype: (A) “number of respiratory infections during the patient’s follow-up”;
893 (B) “severe bacterial infections at least once during the patient’s follow-up”; (C) “number of
894 all infectious episodes during the patient’s follow-up”; (D) “number of infectious episodes
895 without hospitalization during the patient’s follow-up”; (E) “number of infectious episodes
896 with hospitalization during the patient’s follow-up”; (F) “number of infections other than
897 respiratory during the patient’s follow-up”. For the analysis the quantitative variables were
898 transformed into binary traits; a positive relationship was obtained when “the average rates
899 of the phenotype per patient (number of infections/ time at risk)/ the average rates of the

900 phenotype for all the patients (number of infections/ time at risk) (n = 95)" >1. The only
901 originally binary trait was (B).

902 ²Using the open-source whole genome association analysis toolset PLINK v1.07 program
903 (Purcell et al., 2007), six association tests were carried out for the SNP in analysis: the
904 Cochran-Armitage trend test (TREND) and five genotypic tests including the Additive test,
905 measuring additive effects of allele dosage (ADD), Genotypic (2 df) test (GEN), Allelic test
906 (ALLELIC), Dominant gene action (1df) test (DOM), Recessive gene action (1df) test (REC).
907 Only those results for which the chi-squared test was applicable are shown.

908 Abbreviations: SCA, sickle cell anemia.

Table 3

Quasi-poisson regression analysis¹ for the phenotype (A) “number of respiratory infections during the patient’s follow-up”, in SCA pediatric patients (n = 95) relative to *TLR2* variants for rs4696480

	Estimate	S.E.	Z value	p-value ⁴	Incidence rate ratio	CI Lower Limit	CI Upper Limit	Incidence rate ratio reduction/increase (%)
Allelic model								
T	2.98	1.06	2.80	0.00903 **	19.6128131	2.452477460	156.8464717	1861
Fetal hemoglobin ²	0.11	0.05	2.38	0.02451 *	1.1146879	1.019245354	1.2190677	11
- $\alpha^{3.7}$ thalassemia alleles	1.30	0.40	3.22	0.00321 **	3.6534094	1.661834046	8.0317288	265
Steady state hemoglobin ³	-0.65	0.29	-2.26	0.03169 *	0.5200436	0.295078447	0.9165201	-48
Genotypic model								
A/A	-2.84	1.07	-2.65	0.01328 *	0.05815951	0.0070963	0.4766610	-94
Fetal hemoglobin ²	0.11	0.05	2.43	0.02226 *	1.11546060	1.0211839	1.2184410	12
- $\alpha^{3.7}$ thalassemia alleles	1.40	0.42	3.38	0.00224 **	4.07171727	1.8024320	9.1980620	307
Steady state hemoglobin ³	-0.68	0.28	-2.39	0.02392 *	0.50816920	0.2919040	0.8846606	-49

¹ Statistical model: generalized linear model, *Phenotype* ~ *Genotype* + *steady state hemoglobin level* + *steady state leukocyte count* + *fetal hemoglobin (prior to hydroxyurea treatment)* + *haemolytic component* + *alpha-thalassemia*. The *Phenotype* = “number of respiratory infections during the patient’s follow-up”.

The reference group for the allelic model was allele A. The reference group for $-\alpha^{3,7}$ thalassemia was the absence of $-\alpha^{3,7}$ thalassemia. For the fetal hemoglobin there was no reference group. For steady state hemoglobin there was no reference group.

² Fetal hemoglobin (%).

³ Steady state hemoglobin (g/dL).

⁴ The Bonferroni adjusted $\alpha = 0.001667$, where the number of tests performed per inheritance model per locus $n = 30$.

Note 1: The interactions between factors were tested but because they were not statistically significant they were not included in the final model.

Note 2: The statistical analysis was carried out in R (Team, R.C., 2013). Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Abbreviations: SCA, sickle cell anemia; S.E., Standard Error; CI, Confidence interval.

Table 4

Significant results of the quasi-poisson regression analysis¹ for the phenotype (B) “severe bacterial infections at least once during the patient’s follow-up”, in SCA pediatric patients (n = 95) relative to *TLR2* variants for rs4696480

	Estimate	S.E.	Z value	p-value ⁵	Incidence rate ratio	CI Lower Limit	CI Upper Limit	Incidence rate ratio reduction/increase (%)
Genotypic model								
A/T	2.7173	0.8397	3.236	0.00311 **	1.513988e+01	2.920059e+00	78.49705859	1413
Hemolytic component ²	-1.4491	0.4238	-3.420	0.00194 **	2.347796e-01	1.023152e-01	0.53874169	-77
Fetal hemoglobin ³	-0.7724	0.2350	-3.287	0.00273 **	4.619032e-01	2.914266e-01	0.73210412	-54
Steady state leukocyte count ⁴	0.5257	0.1545	3.404	0.00202 **	1.691717e+00	1.249840e+00	2.28981750	69

¹ Statistical model: generalized linear model, *Phenotype* ~ *Genotype* + *steady state hemoglobin level* + *steady state leukocyte count* + *fetal hemoglobin (prior to hydroxyurea treatment)* + *haemolytic component* + *alpha-thalassemia*. The *Phenotype* = “respiratory infections during the patient’s follow-up”. The reference group for the allelic model was T. For the hemolytic component there was no reference group. For the fetal hemoglobin there was no reference group. The reference group for $-\alpha^{3,7}$ thalassemia was the absence of $-\alpha^{3,7}$ thalassemia. The reference group for the *HBB**S haplotype [Bantu/Bantu] was [Other than Bantu]/ [Other than Bantu]. For the steady state leukocyte count there was no reference group.

² Hemolytic component per score unit.

³ Fetal hemoglobin (%).

⁴ Steady state leukocyte count ($10^9/L$).

⁵ The Bonferroni adjusted $\alpha = 0.001471$, where the number of tests performed per inheritance model per locus $n = 34$.

Note 1: The interactions between factors were tested but because they were not statistically significant they were not included in the final model.

Note 2: The statistical analysis was carried out in R (Team, R.C., 2013). Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Abbreviations: SCA, sickle cell anemia; S.E., Standard Error; CI, Confidence interval.

Table 5

Significant results of the quasi-poisson regression analysis¹ for the phenotype “severe bacterial infections at least once during the patient’s follow-up”, in SCA pediatric patients (n = 95) relative to *TLR2* variants for the -196 to 174 indel

	Estimate	S.E.	Z value	p-value ⁴	Incidence rate ratio	CI Lower Limit	CI Upper Limit	Incidence rate ratio reduction/increase (%)
Allelic model								
Ins	-1.7843	0.7897	-2.260	0.0308 *	0.1679149	0.03572185	0.7893046	-83
Hemolytic component ²	-0.7306	0.3117	-2.344	0.0255 *	0.4816236	0.26144550	0.8872261	-52
Fetal hemoglobin ³	-0.4591	0.1989	-2.307	0.0277 *	0.6318783	0.42784553	0.9332111	-37
Genotypic model								
Del/Del	1.9527	0.8972	2.176	0.0381 *	7.04802251	1.2143868794	40.9051038	605
Hemolytic component ²	-1.2606	0.4914	-2.566	0.0159 *	0.28347897	0.1082098581	0.7426341	-72
Fetal hemoglobin ³	-0.5551	0.2365	-2.347	0.0262 *	0.57402478	0.3610729097	0.9125704	-43

¹ Statistical model: generalized linear model, *Phenotype* ~ *Genotype* + *steady state hemoglobin level* + *steady state leukocyte count* + *fetal hemoglobin (prior to hydroxyurea treatment)* + *haemolytic component* + *alpha-thalassemia*. The *Phenotype* = “severe bacterial infections at least once during the patient’s follow-up”. The reference group for the genotypic model was Ins/Ins. For the hemolytic component there was no reference group. For the fetal hemoglobin

there was no reference group. The reference group for $-\alpha^{3,7}$ thalassemia was the absence of $-\alpha^{3,7}$ thalassemia. The reference group for the *HBB**S haplotype [Bantu/Bantu] was [Other than Bantu]/ [Other than Bantu]. The reference group for male gender was female gender.

² Hemolytic component per score unit.

³ Fetal hemoglobin (%).

⁴ The Bonferroni adjusted $\alpha = 0.001471$, where the number of tests performed per inheritance model per locus $n = 34$.

Note 1: The interactions between factors were tested but because they were not statistically significant they were not included in the final model.

Note 2: The statistical analysis was carried out in R (Team, R.C., 2013). Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Abbreviations: SCA, sickle cell anemia; S.E., Standard Error; CI, Confidence interval; Indel, insertion or deletion variation; Del, deletion; Ins, insertion.

Table 6

Quasi-poisson regression model¹ for the phenotype “severe bacterial infections at least once during the patient’s follow-up” in SCA pediatric patients (n = 95) relative to *TLR2* haplotype 7

	Estimate	S.E.	Z value	p-value ²	Incidence rate ratio	CI Lower Limit	CI Upper Limit	Incidence rate ratio reduction/increase (%)
Allelic model								
Hap7	2.52	0.75	3.37	0.00145 **	12.36699701	2.866318564	53.3585544	1137
Genotypic model								
[Other than Hap7]/[Other than Hap7]	-2.52	0.75	-3.37	0.00145 **	0.08086037	0.01874114	0.3488796	-92

¹ Statistical model: quasipoisson regression model in a generalized linear model, *Phenotype* ~ *Genotype* + *steady state hemoglobin level* + *steady state leukocyte count* + *fetal hemoglobin (prior to hydroxyurea treatment)* + *haemolytic component* + *alpha-thalassemia*. The *Phenotype* = “severe bacterial infections at least once during the patient’s follow-up”. The reference group for the allelic model was the absence of Hap7 ([Other than Hap7]). The reference group for the genotypic model was Hap7/Hap7.

² The Bonferroni adjusted $\alpha = 0.001471$, where the number of tests performed per inheritance model per locus $n = 34$.

Note 1: The interactions between factors were tested but because they were not statistically significant they were not included in the final model.

Note 2: The statistical analysis was carried out in R (Team, R.C., 2013). Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Abbreviations: SCA, sickle cell anemia; S.E., standard error; Hap, haplotype; CI, Confidence interval.

Table 7

Quasi-poisson regression model¹ for the infectious phenotype “severe bacterial infections at least once during the patient’s follow-up” in SCA pediatric patients (n = 95) relative to the hemolytic component

	Estimate	S.E.	Z value	p-value	Incidence rate ratio	CI Lower Limit	CI Upper Limit	Incidence rate ratio reduction/increase (%)
Hemolytic component ²	-1.7206	0.4285	-4.015	0.000366 ***	0.17896141	0.077266233	4.145043e-01	-82
Fetal hemoglobin ³	-0.7612	0.2457	-3.098	0.004205 **	0.46708616	0.288567509	7.560431e-01	-53
- α^{37} thalassemia alleles	-2.6593	1.1501	-2.312	0.027811 *	0.07000001	0.007347522	6.668916e-01	-93

¹ Statistical model: quasipoisson regression model in a generalized linear model, *Phenotype* ~ *Genotype* + *steady state hemoglobin level* + *steady state leukocyte count* + *fetal hemoglobin (prior to hydroxyurea treatment)* + *haemolytic component* + *alpha-thalassemia*. The *Phenotype* = “severe bacterial infections at least once during the patient’s follow-up”. For the hemolytic component there was no reference group. For the fetal hemoglobin there was no reference group. The reference group for - α^{37} thalassemia was the absence of - α^{37} thalassemia. The reference group for the *HBB**S haplotype [Bantu/Bantu] was [Other than Bantu]/ [Other than Bantu].

² Hemolytic component per score unit

³ Fetal hemoglobin (%).

Note 1: The interactions between factors were tested but because they were not statistically significant they were not included in the final model.

Note 2: The statistical analysis was carried out in R (Team, R.C., 2013). Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Abbreviations: SCA, sickle cell anemia; S.E., standard error; CI, Confidence interval.

Legends to Figures

Fig. 1. Age distribution curve of the study cohort consisting of a total of 99 unrelated pediatric patients with SCA enrolled in a previous longitudinal study (Coelho et al., 2014). A sub-group of 95 patients was selected for the genotype-to-phenotype association studies based on the criterion of infectious episodes occurring within a range of over three and under 20 years of age. In the figure, the cutoff is illustrated by a vertical grey line. The total observation period for the cohort was 433.4 person*years, the average age at observation was 7.7 ± 3.7 years and the mean follow-up per patient was 4.6 ± 1.5 years. The M/F sex ratio was 1.21. Each bar represents one patient.

Fig. 2. Classification of observational periods as either infectious episodes, episodes without infection or episodes without reference to infection, within vasoocclusive crisis (VOC) and between crises (steady-state) (IC). Infectious disease phenotypes were reported for the SCA pediatric patients ($n = 95$) where the infectious episodes occurred within a range of over three and under 20 years of age.

Fig. 3. Phenotypic characterization of the infectious episodes according to the etiological agent for: (A) all the infectious episodes ($n = 527$); (B) episodes of respiratory infections ($n = 102$). The infectious status of the patients during the follow-up was determined by standard procedures and collected from the patients' hospital records. The 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD10) <http://apps.who.int/classifications/icd10/browse/2016/en>, was used to classify the infectious episodes.

Figure 1

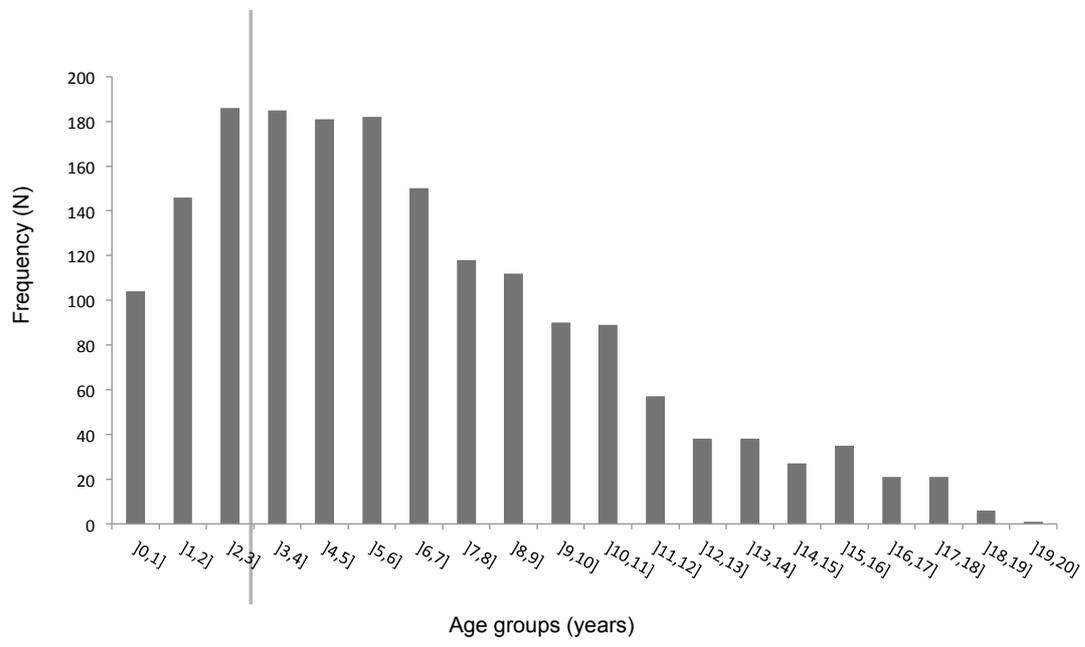


Figure 2

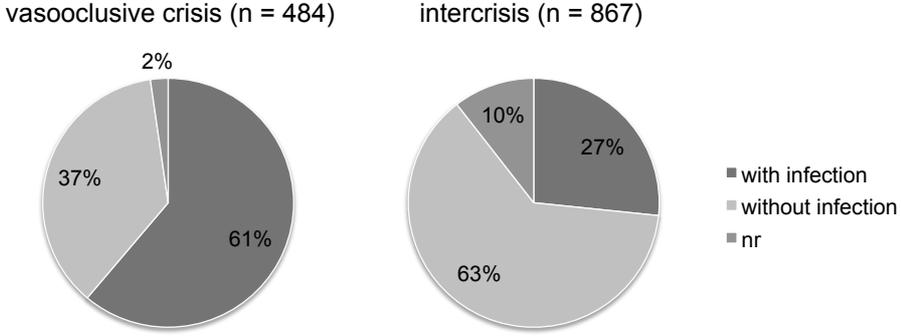


Figure 3

