RESEARCH ARTICLE

Candidate genes-based investigation of susceptibility to Human African Trypanosomiasis in Côte d'Ivoire

Bernardin Ahouty¹, Mathurin Koffi²*, Hamidou Ilboudo³, Gustave Simo⁴, Enock Matovu⁵, Julius Mulindwa⁶, Christiane Hertz-Fowler⁷, Bruno Bucheton⁷, Issa Sidibé⁸, Vincent Jamonneau⁷, Annette MacLeod⁹, Harry Noyes⁶, Simon-Pierre N’Guetta¹, for the TrypanoGEN Research Group as members of The H3Africa Consortium¹⁰

1 Laboratoire de Génétique, Félix Houphouët Boigny University, Abidjan, Côte d’Ivoire, 2 Unité de Recherche en Génétique et Epidémiologie Moléculaire, Jean Lorougnon Guédé University, Daloa, Côte d’Ivoire, 3 Unité Maladies à Vecteurs et Biodiversité, Centre International de Recherche-Développement sur l’Elevage en zone Subhumide, Bobo-Dioulasso, Burkina Faso, 4 Department of Biochemistry, University of Dchang, Dchang, Cameroon, 5 School of Veterinary Medicine, Makerere University, Kampala, Uganda, 6 Centre for Genomic Research, University of Liverpool, Liverpool, United Kingdom, 7 Unité Mixte de Recherche 177 IRD-CIRAD, Institut de Recherche pour le Développement, Montpellier, France, 8 Unité de Recherche Glossines et Trypanosomes, Institut Pierre Richet, Bouaké, Côte d’Ivoire, 9 Wellcome Center for Molecular Parasitology, University of Glasgow, Glasgow, United Kingdom

¹© Copyright: Bernardin Ahouty et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Abstract

Human African Trypanosomiasis (HAT) or sleeping sickness is a Neglected Tropical Disease. Long regarded as an invariably fatal disease, there is increasing evidence that infection by *T. b. gambiense* can result in a wide range of clinical outcomes, including latent infections, which are long lasting infections with no parasites detectable by microscopy. The determinants of this clinical diversity are not well understood but could be due in part to parasite or host genetic diversity in multiple genes, or their interactions. A candidate gene association study was conducted in Côte d’Ivoire using a case-control design which included a total of 233 subjects (100 active HAT cases, 100 controls and 33 latent infections). All three possible pairwise comparisons between the three phenotypes were tested using 96 SNPs in 16 candidate genes (*IL1*, *IL4*, *IL4R*, *IL6*, *IL8*, *IL10*, *IL12*, *TNFA*, *INFγ*, *MIF*, *APOL1*, *HPR*, *CFH*, *HLA-A* and *HLA-G*). Data from 77 SNPs passed quality control. There were suggestive associations at three loci in *IL6* and *TNFA* in the comparison between active cases and controls, one SNP in each of *APOL1*, *MIF* and *IL6* in the comparison between latent infections and active cases and seven SNP in *IL4*, *HLA-G* and *TNFA* between latent infections and controls. No associations remained significant after Bonferroni correction, but the Benjamini Hochberg false discovery rate test indicated that there were strong probabilities that at least some of the associations were genuine. The excess of associations with latent infections despite the small number of samples available suggests that these subjects form a distinctive genetic cluster different from active HAT cases and controls, although no clustering by phenotype was observed by principle component analysis. This underlines the
complexity of the interactions existing between host genetic polymorphisms and parasite diversity.

Author summary

Since it was first identified, human African trypanosomiasis (HAT) or sleeping sickness has been described as invariably fatal. Recent data however suggest that infection by *T. b. gambiense* can result in a wide range of clinical outcomes in its human host including long lasting infections, that can be detected by the presence of antibodies, but in which parasites cannot be seen by microscopy; these cases are known as latent infections. While the factors determining, this varied response have not been clearly characterized, the effectors of the immune responses have been partially implicated as key players. We collected samples from people with active HAT, latent infections and controls in endemic foci in the Côte d’Ivoire. We tested the role of single nucleotide polymorphisms (SNPs) in 16 genes on susceptibility/resistance to HAT by means of a candidate gene association study. There was some evidence that variants of the genes for *IL4*, *IL6*, *APOL1*, *HLAG*, *MIF* and *TNFA* modified the risk of developing HAT. These proteins regulate the inflammatory response to many infections or are directly involved in killing the parasites. In this study, the results were statistically weak and would be inconclusive on their own, however other studies have also found associations in these genes, increasing the chance that the variants that we have identified play a genuine role in the response to trypanosome infection in Côte D’Ivoire.

Introduction

Sleeping sickness, or human African Trypanosomiasis (HAT), is caused by *Trypanosoma brucei gambiense* and *T. b. rhodesiense* and is transmitted by tsetse flies (*Glossina spp*). *T. b. gambiense* is associated with a more chronic disease that can take decades to become patent, and *T. b. rhodesiense* causes an acute disease within months of infection. The chronic form of the disease caused by the *T. b. gambiense* is classically characterized by an early hemolymphatic stage (stage 1) associated with non-specific symptoms such as intermittent fevers and headaches, followed by a meningo-encephalitic stage (stage 2) in which the parasite invades the central nervous system and causes neurological disorders and death if left untreated. This chronic form is found in Western and Central Africa while the acute form caused by *T. b. rhodesiense* is endemic to Eastern Africa [1,2]. However, recent observations are increasingly indicating that infection by *T. b. gambiense* can result in a wide range of clinical outcomes in its human host [3,4]. Self-cure processes have been described and reviewed in Checchi et al. [3]. Furthermore, some authors argued that HAT is not invariably fatal [5], supporting observations made by Garcia et al. [4] who followed individuals who remained seropositive for HAT but without detectable parasites for two years. This clinical diversity is not well understood but could be due to parasite genetic diversity [6,7], human immune gene variability [8] or their interaction [9]. It has been suggested that genetic polymorphisms of the parasite could be associated with asymptomatic and very chronic infections [10]. Nevertheless, genes involved in the host immune response have been implicated in the control of infection or susceptibility to HAT [11,12] and also *T. congolense* infections in experimental models [13,14]. Among the genes implicated in the pathogenesis of the disease, are interleukins 1, 4, 6, 8, 10 and 12 (*IL1, IL4,*...
IL6, IL8, IL10 and IL12), tumor necrosis factor A (TNFA), interferon G (INFG), complement factor H (CFH), macrophage inhibitory factor (MIF), [8,11,15–17].

Some studies have found associations between certain polymorphisms in genes encoding cytokines. For example, IL4 plays a role in susceptibility to T. brucei infection [18], while polymorphisms in the IL6 and IL10 genes have been associated with a decreased risk of developing HAT [11,15]. On the other hand, polymorphisms in TNFA genes have been associated with an increased risk of developing the disease [11,15]. Furthermore, MacLean et al. [19] reported in their study that plasma levels of IFNG significantly decrease during the late stage of the T. b. rhodesiense disease [19]. Human blood contains trypanolytic factors 1 and 2 (TLF1 and TLF2) that are lytic to almost all African trypanosomes except T. b. rhodesiense and T. b. gambiense [20]. Human TLF1 contains two primate-specific proteins, apolipoprotein L1 (APOL1) and haptoglobin-related protein (HPR). APOL1 expression is induced by T. b. gambiense infection but expression is not associated with susceptibility to sleeping sickness [21]. MIF contributes to inflammation-associated pathology in the chronic phase of T. brucei infection [16].

Although genes directly involved in the immune response, like genes encoding cytokines, are very important candidates, genes implicated in the regulation of immunity also have a critical role. Thus, Courtin et al. [12] have shown a genetic association between human leukocyte antigen G (HLA-G) polymorphisms and susceptibility to HAT.

There is now cumulative evidence that polymorphisms in genes involved in the control of the immune response and genes implicated in the regulation of immunity could play a role in HAT infection outcome [11,12].

We report a candidate gene association study of the role of single nucleotide polymorphisms (SNP) in IL1, IL4, IL6, IL8, IL10, IL12, interleukin 4 receptor (IL4R), interleukin 12 receptor (IL12R), TNF, INFG, MIF, haptoglobin-related protein (HPR), CFH, APOL1, HLA-A and HLA-G genes on susceptibility/resistance to HAT.

Results

Characteristics of the population

A total of 100 cases, 100 controls and 33 latent infections were enrolled into the study (Table 1). The mean age (range) of the study population was 38.8 (6–84) years. The sex ratio (male: female) was 0.88 (109/124).

Association study

Three association analyses were run on the three possible pairwise combinations of the controls, latent infections and active HAT cases. After filtering out SNP loci that were not in Hardy Weinberg equilibrium (HWE) (P < 0.001), or had > 10% missing genotypes or had minor allele frequencies of zero, 77 SNP loci remained in the latent infection vs active case comparison and 74 SNP loci remained in the other two comparisons. One individual was filtered out of the case sample and one out of the latent infection sample because of low genotyping rates (<90%).

FST between ethnic groups showed that allele frequency differences between the ethnic groups were small, indicating that ethnicity was unlikely to confound results (median Fst -0.00011763, Fst maximum 0.015). Scatter plots of the first two principal components between the different phenotypes (cases, latent infections and controls) and different ethnic groups show that this population is homogeneous and that the samples did not cluster by phenotype or by ethnic groups (Fig 1) and consequently the data were not stratified by ethnicity.

There were suggestive associations at three loci in IL6 and TNFA in the comparison between active cases and controls (Table 2), three loci in APOL1, MIF and IL6 in the
Table 1. Population characteristics.

<table>
<thead>
<tr>
<th></th>
<th>Cases (n = 100)</th>
<th>Controls (n = 100)</th>
<th>Latent infections (n = 33)</th>
<th>Whole population (n = 233)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age mean</td>
<td>41.8</td>
<td>36.8</td>
<td>36</td>
<td>38.8</td>
</tr>
<tr>
<td>Males</td>
<td>53</td>
<td>42</td>
<td>14</td>
<td>109</td>
</tr>
<tr>
<td>Females</td>
<td>47</td>
<td>58</td>
<td>19</td>
<td>124</td>
</tr>
<tr>
<td>TL+ve</td>
<td>65</td>
<td>0</td>
<td>18</td>
<td>83</td>
</tr>
<tr>
<td>TL-ve</td>
<td>35</td>
<td>100</td>
<td>15</td>
<td>150</td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pntd.0005992.t001

Fig 1. Multidimensional scaling (ms) plots of the genotype data by ethnicity (upper plot) and phenotype (lower plot). The plots show no evidence of clustering either by ethnicity or phenotype.

https://doi.org/10.1371/journal.pntd.0005992.g001
comparison between latent infections and active cases (Table 3) and five loci in IL4 (Fig 2), and one each in HLA-G and TNFA between latent infections and controls (Table 4). After Bonferroni correction, none of these associations remained significant, however Bonferroni correction is very conservative, particularly since there was some linkage between adjacent SNP marker in some genes. The Benjamini-Hochberg false discovery rate (FDR) test, which shows the probability that an observation is a false positive, indicated that at least some of the suggestively positive samples may be true positives. Under the FDR the rs62449495 SNP in IL6 had an 80% chance of being a true positive and the TNFA-308 rs1800629 SNP had a 71% chance of being a true positive (Table 2). The strongest association was with APOL1 rs73885319, which is also known as the G1 allele, which had a 90% chance of being a true positive in the comparison between latent infections and controls (Table 3). Complete results for all tests that passed quality control are shown in supplementary data tables (S1, S2 and S3 Tables).

**Discussion**

In this study, we investigated the role of SNPs on susceptibility/resistance to HAT. We genotyped 96 SNPs within sixteen genes that were investigated to test genetic association with HAT in Côte d’Ivoire. For consistency, all samples were commercially genotyped using two platforms: Genome Transcriptome de Bordeaux, France and LGC Genomics UK.

Our results suggest that three SNPs were associated with the development of HAT by controls, another three were associated with the development of active HAT by people with latent infections, and seven SNP were associated with the risk of developing latent infections five of them in IL4 although none of these results remained significant after Bonferroni correction. Only two of the SNP with suggestive associations were coding and missense (IL6 rs2069830 and APOL1 rs73885319) since, with some exceptions, SNP were primarily selected as tags that might be linked to functional SNP rather than for any putative function.

The most striking feature of the results is that most associations were found in the comparisons with latent infections, despite limited power, with only 33 of these samples being available compared with 100 each for the cases and controls. This illustrates the increased power to be gained from using well defined phenotypes. Given the evidence that some people in West Africa can self-cure their T. b. gambiense infections [5] it is likely that some of the control subjects are resistant to infection or have recovered from infection, whilst others are naïve and susceptible. These two groups may have very different genetic backgrounds which could confound the association studies. The data also suggest that the latent infections may represent a genetically distinctive group. Although the principle component analysis did not identify any

**Table 2. Association tests between cases (n = 99) and controls (n = 100).**

<table>
<thead>
<tr>
<th>CHR</th>
<th>BP</th>
<th>SNP</th>
<th>Gene</th>
<th>Cons</th>
<th>Min A</th>
<th>Cases</th>
<th>Controls</th>
<th>Maj A</th>
<th>P</th>
<th>Bonf_Corr</th>
<th>FDR_BH</th>
<th>OR</th>
<th>L95</th>
<th>U95</th>
<th>HWE_p</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>22,764,338</td>
<td>rs62449495</td>
<td>IL6</td>
<td>5′</td>
<td>A</td>
<td>0.04</td>
<td>0.13</td>
<td>G</td>
<td>0.0025</td>
<td>0.20</td>
<td>0.20</td>
<td>0.30</td>
<td>0.13</td>
<td>0.68</td>
<td>1.00</td>
</tr>
<tr>
<td>6</td>
<td>31,543,031</td>
<td>rs1800629</td>
<td>TNFA</td>
<td>5′</td>
<td>A</td>
<td>0.09</td>
<td>0.19</td>
<td>G</td>
<td>0.0071</td>
<td>0.58</td>
<td>0.29</td>
<td>0.44</td>
<td>0.24</td>
<td>0.80</td>
<td>0.51</td>
</tr>
<tr>
<td>7</td>
<td>22,767,137</td>
<td>rs2069830</td>
<td>IL6</td>
<td>Missense</td>
<td>0.11</td>
<td>0.06</td>
<td>C</td>
<td>0.0369</td>
<td>1.00</td>
<td>0.81</td>
<td>2.17</td>
<td>1.02</td>
<td>4.61</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pntd.0005992.t002

**Table 3. Association tests between latent infections (n = 32) and active HAT cases (n = 99).**

<table>
<thead>
<tr>
<th>CHR</th>
<th>BP</th>
<th>SNP</th>
<th>Gene</th>
<th>Cons</th>
<th>Min A</th>
<th>Cases</th>
<th>Latent</th>
<th>Maj A</th>
<th>P</th>
<th>Bonf_Corr</th>
<th>FDR_BH</th>
<th>OR</th>
<th>L95</th>
<th>U95</th>
<th>HWE_p</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>36,661,906</td>
<td>rs73885319</td>
<td>APOL1</td>
<td>Missense</td>
<td>0.20</td>
<td>0.41</td>
<td>A</td>
<td>0.0012</td>
<td>0.10</td>
<td>0.10</td>
<td>0.37</td>
<td>0.20</td>
<td>0.68</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>24,235,455</td>
<td>rs36086171</td>
<td>MIF</td>
<td>5′</td>
<td>G</td>
<td>0.29</td>
<td>0.13</td>
<td>A</td>
<td>0.0062</td>
<td>0.52</td>
<td>0.26</td>
<td>2.90</td>
<td>1.30</td>
<td>6.46</td>
<td>0.05</td>
</tr>
<tr>
<td>7</td>
<td>22,764,338</td>
<td>rs62449495</td>
<td>IL6</td>
<td>5′</td>
<td>A</td>
<td>0.04</td>
<td>0.13</td>
<td>G</td>
<td>0.0217</td>
<td>1.00</td>
<td>0.61</td>
<td>0.29</td>
<td>0.10</td>
<td>0.80</td>
<td>1.00</td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pntd.0005992.t003
cluster associated with latent infections (Fig 1) further studies of their genetic and immunological profiles are required to test this hypothesis.

Associations with \textit{IL6}

The minor (A) allele of \textit{IL6} rs62449495 appeared to protect against progression from latent infections to active HAT (Table 3) and against the development of active HAT by controls but these associations did not remain significant after Bonferroni correction (Table 2). The major allele of rs2069830 was also protective against the development of active HAT by controls before but not after Bonferroni correction (Table 2). The rs2069830 SNP causes a proline to serine change, but this is predicted to be benign by both Sift and Polyphen according to the Ensembl Variant Effect predictor. \textit{IL6} plays a key role in the acute inflammatory response and in regulation of the production of acute phase proteins such as C-reactive protein [22]. It contributes to the inflammatory response, regulates haematopoiesis, which may contribute to the anaemia associated with HAT, and modifies the permeability of the blood brain barrier, which may contribute to the development of the stage 2 invasion of the CNS [22,23]. The results of our association study are consistent with the involvement of \textit{IL6} variants in the susceptibility to the disease as Courtin \textit{et al} [15] have reported before. However, it should be noted that the polymorphisms of \textit{IL6} in our study (rs62449495 and rs2069830) are different of those found by Courtin \textit{et al}. [15] with \textit{IL6} \textit{4339} = rs2069849. They showed that in DRC, \textit{IL6} \textit{4339} SNP was significantly associated with a decreased risk of developing the disease with a \( P \)-value = 0.0006 before the Bonferroni correction (0.04 after Bonferroni Correction).

Table 4. Association tests between latent infections (n = 32) and controls (n = 100).

<table>
<thead>
<tr>
<th>CHR</th>
<th>BP</th>
<th>SNP</th>
<th>Gene</th>
<th>Cons</th>
<th>Min A</th>
<th>Latent</th>
<th>Controls</th>
<th>Maj A</th>
<th>P</th>
<th>Bonf_Corr</th>
<th>FDR_BH</th>
<th>OR (L95-U95)</th>
<th>L95</th>
<th>U95</th>
<th>HWE_p</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>132,009,154</td>
<td>rs2243250</td>
<td>IL4</td>
<td>5'</td>
<td>C</td>
<td>0.42</td>
<td>0.27</td>
<td>T</td>
<td>0.0159</td>
<td>1.00</td>
<td>0.48</td>
<td>2.02</td>
<td>1.13</td>
<td>3.64</td>
<td>0.80</td>
</tr>
<tr>
<td>5</td>
<td>132,010,726</td>
<td>rs734244</td>
<td>IL4</td>
<td>Intronic</td>
<td>T</td>
<td>0.25</td>
<td>0.41</td>
<td>C</td>
<td>0.0215</td>
<td>1.00</td>
<td>0.48</td>
<td>0.49</td>
<td>0.26</td>
<td>0.92</td>
<td>0.01</td>
</tr>
<tr>
<td>5</td>
<td>132,016,554</td>
<td>rs2243282</td>
<td>IL4</td>
<td>Intronic</td>
<td>A</td>
<td>0.30</td>
<td>0.45</td>
<td>C</td>
<td>0.0347</td>
<td>1.00</td>
<td>0.48</td>
<td>0.53</td>
<td>0.29</td>
<td>0.96</td>
<td>1.00</td>
</tr>
<tr>
<td>6</td>
<td>29,799,116</td>
<td>rs1611139</td>
<td>HLAG</td>
<td>3'</td>
<td>T</td>
<td>0.26</td>
<td>0.14</td>
<td>G</td>
<td>0.0354</td>
<td>1.00</td>
<td>0.48</td>
<td>2.13</td>
<td>1.04</td>
<td>4.36</td>
<td>0.39</td>
</tr>
<tr>
<td>5</td>
<td>132,009,710</td>
<td>rs2070874</td>
<td>IL4</td>
<td>5' UTR</td>
<td>T</td>
<td>0.34</td>
<td>0.50</td>
<td>C</td>
<td>0.0372</td>
<td>1.00</td>
<td>0.48</td>
<td>0.53</td>
<td>0.30</td>
<td>0.96</td>
<td>0.69</td>
</tr>
<tr>
<td>6</td>
<td>31,543,031</td>
<td>rs1800629</td>
<td>TNFA</td>
<td>5'</td>
<td>A</td>
<td>0.08</td>
<td>0.19</td>
<td>G</td>
<td>0.0396</td>
<td>1.00</td>
<td>0.48</td>
<td>0.37</td>
<td>0.14</td>
<td>0.99</td>
<td>0.51</td>
</tr>
<tr>
<td>5</td>
<td>132,014,109</td>
<td>rs2243270</td>
<td>IL4</td>
<td>Intronic</td>
<td>A</td>
<td>0.36</td>
<td>0.23</td>
<td>G</td>
<td>0.0415</td>
<td>1.00</td>
<td>0.48</td>
<td>1.88</td>
<td>1.02</td>
<td>3.45</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Cons = Consequence from Ensembl Variant predictor, Min A = minor allele, Maj A = major allele, Latent minor allele frequency of Latent cases; Controls minor allele frequency of controls; OR = odds ratio, \( P \) = \( p \)-value, [L95-U95] = confidence interval of odds ratio, Bonf_Corr = Bonferroni corrected \( p \) value, FDR_BH = False Discovery Rate Benjamini-Hochberg (It is probability of falsely rejecting the null hypothesis. The null hypothesis is that allele frequencies in cases and controls are the same), OR = odds ratio, [L95-U95] = confidence interval; HWE_p is the Hardy-Weinberg \( p \)-value.

https://doi.org/10.1371/journal.pntd.0005992.t004
Association with TNFA

Our results also indicate that subjects carrying the A allele of TNFA (rs1800629) A/G had a lower risk of developing active HAT (Table 2) or a latent infection (Table 4), suggesting the possibility of a protective effect. This SNP is also known as the TNF-308 SNP and the minor A allele is associated with higher plasma levels of TNFA [24]. In a previous study in Côte d’Ivoire, Courtin et al. [11] showed that the distribution of the TNF-308G/A polymorphism did not differ significantly between cases and controls in the total population, but found that, under a recessive model the AA genotype was associated with risk of HAT in the 39 cases and 57 controls who had been living in the endemic area of Sinfra for less than 10 years (before Bonferroni correction). In contrast, in our study, all participants had lived in the endemic area all their lives and the association was additive rather than recessive, furthermore the MAF in Courtin’s studies was 24%, whilst in ours it was 14%, suggesting significant differences in population structure between the two studies. Varying results in TNFA associations studies are not uncommon, multiple candidate gene association studies of TNFA-308 (rs1800629) and malaria have found inconsistent results in different populations [25,26]. Another explanation could be that the TNFA-308 has no effect but is in linkage disequilibrium (LD) with another unidentified polymorphism. Varying LD across populations might lead to different findings [27]. Despite the conflicting results from association study, animal models indicate that TNFA is likely to be a key mediator in the control of T. brucei infections [28] and a direct dose dependent lytic effect of TNFA on purified T. b. gambiense parasites has been reported suggesting an involvement in parasite growth control [29,30]. Given the experimental evidence for a role for TNFA, it is possible that inconsistent results of association studies are a consequence of different TNFA alleles only making a small difference to infection outcome and that therefore larger studies would be needed to detect associations and also to heterogeneity in regulation of TNFA between populations.

Association with IL4

Five of the seven SNP that were suggestively associated with the comparison between latent infections and controls were in IL4 (Table 4, Fig 2). Three of these SNP were adjacent to each other at the 5’ end of the gene (Fig 2), although linkage between them was modest ($r^2 < 0.5$). While the individual associations did not remain significant after Bonferroni correction the observation that five of the sixteen SNP tested in IL4 had suggestive associations increases the probability of a genuine association with IL4 and the risk of developing a latent infection. A study using three IL4 SNP in DRC did not find any associations, but that was in a comparison between cases and controls [15]. In B10.Q mice deletion in IL4 lead to increased T. brucei parasitaemia levels but longer survival time [18]. The relationship between IL4 polymorphisms and acute HAT infection requires further study.

Association with APOL1

The suggestive association with APOL1 G1 allele rs73885319 and protection against progression from latent infection to active HAT is consistent with the association found in Guinea [31,32], although in those studies this SNP was also associated with increased risk of controls developing a latent infection as well. Moreover, APOL1 expression is induced by T. b. gambiense infection but not associated with differential susceptibility to sleeping sickness [21]. APOL1 rs73885319 is also known as the G1 allele of APOL1 and is associated with kidney disease in African Americans and the relatively high frequency of this deleterious allele was assumed to be due to selection by HAT [33]. The data presented here is consistent with that hypothesis, but no support was found for the role of the G2 allele of APOL1 (rs71785313) in HAT although it is implicated in kidney disease.
Association with \textit{MIF}

Although, our data show that subjects carrying the G allele of \textit{MIF} (rs36086171) G/A had a risk of developing active HAT (Table 3), we did not find a significant difference after correction (BONF = 0.52). \textit{MIF} is a ubiquitously expressed protein that has proinflammatory, hormonal and enzymatic activities [34]. It is implicated in many inflammatory diseases [35]. It functions by recruiting myeloid cells to the site of inflammation [36], by inducing their differentiation towards M1 cells secreting TNF [37] and by suppressing p53-dependent apoptosis of inflammatory cells. While there are no human studies directly linking \textit{MIF} to HAT, murine studies show that \textit{MIF} plays a role as a mediator of the inflammation which is a key feature in trypanosomiasis-associated pathology [16,38].

Association with \textit{HLA-G}

\textit{HLA-G} molecule plays an important role on immune response regulation and has been associated with the risk of human immunodeficiency virus (HIV) infection [39], human papilloma virus [40] and herpes simplex virus type 1 [41]. The results of our association study are consistent with the involvement of \textit{HLA-G} genetic variants in the susceptibility to the disease [12]. Control subjects with rs1611139 T allele of the \textit{HLA-G} gene showed a suggestive association with increased risk of leading in latent infection in case they are infected. This result could be due to differences in selection pressures possibly driven by variability in the immuno-pathology of the diseases. It is known that cytokines such as IL-10 can induce \textit{HLA-G} expression by affecting mRNA transcripts and protein synthesis by human monocytes and trophoblasts, thus having a significant impact on parasitic infections [42].

Conclusion

The results discussed in this paper should be used with caution as no loci remained significantly associated with HAT after the Bonferroni correction for multiple testing, although some had high probabilities of being associated using a FDR test. Some of the SNP loci identified here were also significant in other studies. Multiple independent observations of marginally significant effects suggest that these effects may be genuine but that the effect size is not large enough to be detected by the numbers available to be tested.

Our data support the findings from Guinea about the role of the \textit{APOL1} G1 allele and also suggest that the polymorphisms of \textit{IL4, IL6, HLA-G} and \textit{TNFA} present interesting candidates for the investigation of the genetic susceptibility / resistance to HAT.

The large number of suggestive associations with latent infections despite the small number of samples indicate the people with latent infections form a genetically distinctive group that merit further investigation.

Materials and methods

Population study, definition of phenotypes and study design

The study took place in western-central Côte d’Ivoire in the main HAT foci of Bonon, Bouafia, Zoukougbeu, Oume, and Sinfra. Samples were collected in three stages: (i) previously archived samples; existing collections of previously archived samples were centralized to Centre International de Recherche-Développement sur l’Elevage en zone Subhumide (CIRDES) in Burkina-Faso, (ii) retrospectively collected samples for which study sites were revisited for consent and resample previously diagnosed and treated patients and (iii) prospectively collected samples including new HAT cases. Three phenotypes were considered: (i) cases, defined as individuals in whom the presence of trypanosomes was confirmed by microscopy and trypanolysis
test positive (TL+ve) but trypanolysis negative (TL-ve) for some cases who were retrospectively sampled for the purpose of this study and who had not been subject to the trypanolysis test (TL) when originally identified as HAT cases by microscopy; (ii) controls, defined as individuals living in the endemic area who are card agglutination test for trypanosomiasis (CATT) negative (CATT-ve) and trypanolysis test negative (TL-ve), and without evidence of previous HAT infection and (iii) latent infections, defined as subjects who were CATT positive (CATT+ve) with plasma dilution higher or equal to $\frac{1}{4}$ (CATTpl $\geq 1/4$) but in whom parasites could not be detected by microscopy at repeated tests for at least two years.

A total of 233 individuals with 100 HAT cases, 100 controls and 33 latent infections were included in the study. Samples were frozen directly in the field at -20˚C and kept at that temperature until use. For each individual, an aliquot of plasma was used to perform the immune TL test that detects Litat 1.3 and Litat 1.5 variable surface antigens specific for \textit{T.b. gambiense}.

This study was one of five populations studies of HAT endemic areas in Cameroon, Côte d’Ivoire, Guinea, Malawi and Uganda by the TrypanoGEN consortium [44]. The studies were designed to have 80% power to detect odds ratios (OR) > 2 for loci with disease allele frequencies of 0.15–0.65 and 100 cases and 100 controls with the 96 SNPs genotyped. Power calculations were undertaken using the genetics analysis package gap in R [45].

**Selection of SNPs and genotyping**

Genomic DNA was obtained from peripheral blood samples. Extraction was performed using the Qiagen DNA extraction kit (QIAamp DNA Blood Midi Kit) according to the manufacturer’s instructions and quantified by Nanodrop assay. DNA was stored at -20˚C until analysis.

Ninety-six SNPs were genotyped in \textit{IL1}, \textit{IL4}, \textit{IL4R}, \textit{IL6}, \textit{IL8}, \textit{IL10}, \textit{IL12}, \textit{IL12R}, \textit{TNFA}, \textit{INFG}, \textit{MIF}, \textit{HPR}, \textit{CFH}, \textit{APOL1}, \textit{HLA-A}, and \textit{HLA-G}. The SNPs were selected by two strategies: 1) SNP in \textit{IL4}, \textit{IL6}, \textit{IL8}, \textit{HLAG} and \textit{IFNG} were designed as markers for linkage disequilibrium (LD) scans of each gene [46], 2) SNPs in other genes were selected based on reports in the literature that they were associated with trypanosomiasis or other infectious diseases. The LD scans were designed using 1000 Genomes Project data [47], merged with low fold coverage (8-10x) whole genome shotgun data generated from 230 residents living in regions (DRC, Guinea Conakry, Côte d’Ivoire and Uganda) where trypanosomiasis is endemic (TrypanoGEN consortium, European Nucleotide Archive study EGAS00001002482). Loci with minor allele frequency < 5% in the reference data were excluded and an $r^2$ of 0.5 was used to select SNP in LD.

DNA was genotyped by two commercial service providers: INRA- Site de Pierroton, Plateforme Genome Transcriptome de Bordeaux, France and ii- LGC genomics Hoddesden UK. At INRA multiplex design (two sets of 40 SNPs) was performed using Assay Design Suite v2.0 (Agena Biosciences). SNP genotyping was achieved with the iPLEX Gold genotyping kit (Agena Biosciences) for the Mass Array iPLEX genotyping assay, following the manufacturer’s instructions. Products were detected on a Mass Array mass spectrophotometer and data were acquired in real time with Mass Array RT software (Agena Biosciences). SNP clustering and validation was carried out with Typer 4.0 software (Agena Biosciences). At LGC Genomics SNP were genotyped using the PCR based KASP assay[48].

**Statistical analysis**

All statistical analyses were performed using the Plink 1.9 and R v 3.2.1 software. Individuals were excluded who had > 10% missing SNP data. SNP loci were excluded that > 10% missing genotypes or if the control samples were not in Hardy-Weinberg equilibrium (p<0.001).

Case-control association analysis using SNP alleles/genotypes was undertaken using Fisher’s
exact test. The difference between the ethnic groups was estimated using the \( F_{ST} \), with values potentially varying from zero (no population differentiation) to one (complete differentiation) [47]. Ethnicity was not used as a risk factor for HAT because we observed no significant differences in \( F_{ST} \) between ethnic groups (\( F_{ST} \) maximum 0.015) or clustering by linguistic group by multidimensional scaling in Plink. Some studies have used ethnicity and age as risk factors for HAT but found no significant association [11,15]. 77 SNP remained in at least one comparison after filtering. \( P \) values were adjusted for multiple testing using the Bonferroni corrections and Benjamini Hochberg false discovery rate test as implemented in Plink.

Ethics statement

The population of the study was informed. All adult subjects provided written informed consent. For children, a parent or guardian of any child participant provided written informed consent on their behalf. The protocol of the study was approved by the traditional authorities (chief and village committee) and by National ethics committee hosted by the Public Health Ministry of Côte d’Ivoire with the number: N°38/MSLS/CNERm-dkn 5th May 2014. This study is part of the TrypanoGen project which aims to a better understanding of genetic determinism of human susceptibility to HAT and the TrypanoGen-CI samples were archived in the TrypanoGen Biobank hosted by CIRDES in Bobo-Dioulasso, Burkina Faso [44].

Supporting information

S1 Table. Quality control test between HAT cases and controls.
(XLSX)

S2 Table. Quality control test between latent infections and HAT cases.
(XLSX)

S3 Table. Quality control test between latent infections and controls.
(XLSX)

Acknowledgments

We thank subjects who generously donated their specimens and the field workers from the Ivorian HAT foci for their dedication in collecting and processing these specimens.


Author Contributions

Conceptualization: Mathurin Koffi.

Data curation: Christiane Hertz-Fowler, Harry Noyes.

Formal analysis: Julius Mulindwa, Harry Noyes.

Funding acquisition: Mathurin Koffi, Gustave Simo, Enoch Matovu, Christiane Hertz-Fowler, Bruno Bucheton, Issa Sidibé, Vincent Jamonneau, Annette MacLeod.


Writing – original draft: Bernardin Ahouty, Harry Noyes.

Writing – review & editing: Bernardin Ahouty, Mathurin Koffi, Hamidou Ilboudo, Harry Noyes, Simon-Pierre N’Guetta.

References


