Immunochip SNP array identifies novel genetic variants conferring susceptibility to candidaemia

Vinod Kumar1,*, Shih-Chin Cheng2,*, Melissa D. Johnson3,4,*, Sanne P. Smeekens2, Agnieszka Wojtowicz5, Evangelos Giamarellos-Bourboulis6,7, Juha Karjalainen1, Lude Franke1, Sebo Withoff1, Theo S. Plantinga2, Frank L. van de Veerdonk2, Jos W.M. van der Meer2, Leo A.B. Joosten2, Harry Sokol8,9, Hermann Bauer10, Bernhard G. Herrmann10, Pierre-Yves Bochud5, Oscar Marchetti5, John R. Perfect3,4, Ramnik J. Xavier8,9, Bart Jan Kullberg2, Cisca Wijmenga1,§ & Mihai G. Netea2,§

Candidaemia is the fourth most common cause of bloodstream infection, with a high mortality rate of up to 40%. Identification of host genetic factors that confer susceptibility to candidaemia may aid in designing adjunctive immunotherapeutic strategies. Here we hypothesize that variation in immune genes may predispose to candidaemia. We analyse 118,989 single-nucleotide polymorphisms (SNPs) across 186 loci known to be associated with immune-mediated diseases in the largest candidaemia cohort to date of 217 patients of European ancestry and a group of 11,920 controls. We validate the significant associations by comparison with a disease-matched control group. We observe significant association between candidaemia and SNPs in the CD58 (P = 1.97 × 10−11; OR = 4.68), LCE4A-C1orf68 (P = 1.98 × 10−10; OR = 4.25) and TAGAP (P = 1.84 × 10−8; OR = 2.96) loci. Individuals carrying two or more risk alleles have an increased risk for candidaemia of 19.4-fold compared with individuals carrying no risk allele. We identify three novel genetic risk factors for candidaemia, which we subsequently validate for their role in antifungal host defence.
Candidaemia is the fourth most common bloodstream infection\(^1\), with known risk factors such as neutropenia, mucosal barrier injury, transplantation, immunosuppressive drugs, intravascular catheters and extended intensive care unit stay\(^2-4\). Despite the availability of potent antifungal drugs, the mortality of patients with candidaemia remains high (up to 37–44%)\(^5,6\). It has therefore been proposed that only intensified patient care, using risk assessment and adjutant immunotherapy, may improve the outcome\(^7\).

The host immune status is crucial for the outcome of Candida infection, and identifying genetic variation in immune genes that confer susceptibility to Candida infection may aid in designing effective preventive strategies. Several small-scale candidate gene association studies suggest a role in candidaemia risk for single-nucleotide polymorphisms (SNPs) in toll-like receptors (TLR-1, TLR-2, TLR-3 and TLR-4), interleukins (ILs) (IL-12B and IL-10) and lymphoid protein tyrosine phosphatase (PTPN22) (ref. 3). Interestingly, ~10 monogenic disorders have been reported to be associated with chronic mucocutaneous candidiasis, and almost all are caused by defects in genes of the immune system\(^3\).

Intriguingly, common SNPs in 8 out of these 10 monogenic disorder genes (AIRE, CARD9, STAT1, STAT3, TYK2, CD25, IL17RA and IL17F) are also associated with susceptibility to different immune-mediated diseases (NHGRI GWAS catalog\(^8\)), implying that genes necessary for immune regulation are strong candidates in determining susceptibility to fungal infections.

Here we report the first genome-wide screen of ~200,000 SNPs in 186 loci in the largest candidaemia cohort to date. By using the ImmunoChip SNP array\(^9\), we identify three novel genetic risk factors for candidaemia that were validated using transcriptomics, pathway analysis and immunological studies.

**Results**

**Association analysis identifies three candidaemia risk loci.**

After filtering the ImmunoChip data using standard quality parameters, we obtained 118,889 SNPs from the 217 candidaemia cases and 11,920 healthy controls, which we analysed using logistic regression analysis. The results of this analysis revealed significant association \(P < 5 \times 10^{-8}\) to three independent loci with candidaemia (Fig. 1a). The top SNPs from these three loci were rare in healthy controls, with risk allele frequencies <2%, whereas the risk allele frequencies were >5% in candidaemia cases (Table 1). The top-associated SNP, rs17035850 \(\text{CD58}^\text{0.97} \times 10^{-11}; \text{oDDS ratio} = 4.68,\) is located in a block of linkage disequilibrium (LD) of ~30 kb at 1p13.1, which contains the CD58 gene (Fig. 1b), two long non-coding RNAs (lncRNA), RPS-1086K13.1 and RP4-655J12.4, and the pseudogene NAP1L4P1 (Supplementary Fig. 1). The second hit was with rs4845320 \(\text{CD58}^\text{0.198} \times 10^{-10}; \text{oDDS ratio} = 4.25,\) which lies in an LD block of 150 kb at 1q21.3 (Fig. 1c) that contains a cluster of genes encoding late cornified envelope (LCE) proteins\(^9\). The third significantly associated SNP, rs3127214 \(\text{CD58}^\text{1.84} \times 10^{-7}; \text{oDDS ratio} = 3.96,\) is located at the fifth end of TAGAP (Fig. 1d), encoding T-cell activation RhoGTPase-activating protein\(^1\), and in an LD block of 120 kb at 6q25.3.

**Validation and replication of three associated SNPs.**

Since we used a large population-based control cohort for discovering candidaemia susceptibility loci, we tested whether these associations could be confirmed using 146 disease-matched controls (Table 2). We observed a significant difference between cases and controls at two loci (LCE4A-C1orf68 and TAGAP), whereas at the CD58 locus we observed a trend of association (Table 1). This latter effect could be explained by the small number of controls and/or low frequency of the risk allele at rs17035850 (Table 1).

The top CD58 SNP (rs17035850) is a rare variant with a minor allele frequency (MAF) of 0.012, and the second CD58 SNP (rs12025416) is a frequent one with MAF of 0.13. We assessed the pairwise LD between these two CD58 SNPs in our population-based controls. We observed both a low correlation \(r^2 = 0.09\) and a high \(D' = 0.96\) between these two SNPs, indicating the existence of rare risk haplotype that carry the risk alleles of these two SNPs. Therefore, we also tested for association at the second CD58 top SNP, rs12025416, which is more frequent, and we observed a significant association with susceptibility to candidaemia \(P = 0.022\). Therefore, these results suggest that the observed associations are true genetic associations.

To further evaluate these associations genetically, we carried out replication analysis by genotyping the four SNPs in two independent candidaemia cohorts (one of 75 African-American patients and one of 27 European patients from Switzerland). Because of the small size of these additional cohorts, we found no successful replication in two independent cohorts, although we observed only a trend of association for rs4845320 SNP in the TAGAP locus in the African-American cohort (Supplementary Table 1). In the Swiss candidaemia replication cohort included matched controls, we performed a joint analysis of the discovery cohort with the matched controls and the Swiss candidaemia replication cohort. This analysis revealed an improved association for two loci: rs12025416 in the CD58 locus \((P = 0.015)\) and rs4845320 in the LCE4A-C1orf68 locus \((P = 0.0036)\), with the same allelic direction and without any evidence for heterogeneity between the two cohorts (Supplementary Table 2).

**Functional annotation of candidaemia-associated SNPs.**

To understand how the candidaemia-associated SNPs from these three loci affect disease, we intersected the three top SNPs as well as their proxies \(r^2 \geq 0.8\) and \(D' = 1\) with functional information from ENCODE using HaploReg\(^12\). We found 7, 19 and 5 SNPs to be in high LD with rs17035850, rs4845320 and rs3127214 SNPs, respectively (Supplementary Table 3). All SNPs that were in LD with the three candidaemia top SNPs were located in non-coding regions and not within exons of protein-coding genes. Some of these SNPs were overlapping with ENCODE-characterized regulatory regions, such as enhancers and/or DNase hypersensitive sites, suggesting that these SNPs may regulate gene expression. Next, we tested whether these SNPs affect expression of nearby genes using publicly available blood expression quantitative trait loci data from 5,000 samples\(^13\). However, we found no significant expression quantitative trait loci for these SNPs. Thus, it could be that the identified SNPs are likely in LD with regulatory variants, but these may only be functional in cell types such as macrophages and under stimulatory conditions with Candida infection. This will be important to test in the future studies.

Interestingly, the LD block around the top-associated SNP, rs17035850, contains not only the CD58 protein-coding gene, but also non-coding genes, namely RP5-1086K13.1, RP4-655J12.4 and the pseudogene NAP1L4P1 (Supplementary Fig. 1). Since many lncRNAs are co-regulated with their protein-coding gene in cis, there is a possibility that RP5-1086K13.1 lncRNA is co-regulating CD58. To test this possibility, we extracted all co-regulated genes with RP5-1086K13.1 using the GeneNetwork database\(^14\). We found that RP5-1086K13.1 is significantly co-regulated with CD58 compared with all other genes in the human genome (Supplementary Table 4). This observation suggests the possibility that even if the SNP affects the expression of RP5-1086K13.1, that may in turn affect CD58 expression levels.

**Association of CD58 and TAGAP SNPs with severity of disease.**

In addition to increasing susceptibility to candidaemia, the
The transcriptomes of CD58- and TAGAP-deficient macrophages showed a total of 169 (at 6 h) and 93 (at 24 h) transcripts with at least a 1.25-fold differential expression (Supplementary Fig. 3). Enrichment analysis showed that the differentially expressed genes are enriched for innate immune function, regulation of cytokine production and general cellular responses to bacterial infections (Supplementary Fig. 4). Furthermore, we investigated whether mutations in these differentially expressed genes show any common phenotypes in the mouse using the Mouse Genome Informatics phenotype data integrated in the GeneNetwork database. Indeed, we observed a significant enrichment for genes in which mutations may cause altered levels of IL-6 and tumor necrosis factor (TNF)-α secretion, and impaired macrophage phagocytosis (Fig. 2a). These data suggest that altered expression levels of CD58 may regulate Candida phagocytosis on one hand, and indirectly regulate IL-6 and TNF-α secretion on the other hand.

Functional validation of CD58 for anti-Candida host defence. We next tested the CD58 mRNA levels in macrophages at 6 and 24 h upon Candida infection using microarray data. We observed polymorphisms identified above may also influence the severity of the disease. Indeed, assessment of the effect of these SNPs revealed that CD58 SNP rs17035850 associated with persistent fungemia, defined as positive blood cultures for >5 days despite adequate therapy (P = 0.005), while TAGAP SNP rs3127214 associated with disseminated disease in the organs (P = 0.017) (see also Supplementary Table 5).

Figure 1 | Genetic association with candidaemia. (a) Manhattan plot showing the genome-wide P values of association with candidaemia. The y axis represents the – log_{10} P values of 118,989 SNPs and their chromosomal positions are shown on the x axis. The horizontal red line shows the genome-wide significant threshold of P < 5 × 10^{-8}. The P-values were obtained by logistic regression test after adjusting for the first four components from the multidimensional scaling analysis. Regional association plots at (b) 1p13.1, (c) 1q21.3 and (d) 6q25.3. The P values of genotyped SNPs are plotted as – log_{10} P values against their physical chromosomal positions (hg19). Estimated recombination rates from the 1000 Genomes European population show the local LD structure. Inset, the colours of the SNPs indicate LD with the top-associated SNPs according to a scale from r^2 = 0 to r^2 = 1 based on pairwise r^2 values from the 1000 Genomes European population. Lower panel, gene annotations from the University of California Santa Cruz genome browser (hg19).
Table 1 | Significantly associated loci with candidaemia that were validated using disease-matched controls.

<table>
<thead>
<tr>
<th>rs ID</th>
<th>Chr</th>
<th>Cohorts</th>
<th>Case (AA, AB, BB)</th>
<th>Controls (AA, AB, BB)</th>
<th>MAF</th>
<th>P-value (log10)</th>
<th>OR (95% CI)</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs17035850</td>
<td>1</td>
<td>Cases vs population-based controls</td>
<td>3, 16, 198</td>
<td>0, 277, 11,442</td>
<td>0.051</td>
<td>1.97 x 10^-11</td>
<td>4.68 (2.98-7.35)</td>
<td>CDS8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cases vs disease-matched controls</td>
<td>3, 16, 198</td>
<td>0, 8, 138</td>
<td>0.027</td>
<td>0.15</td>
<td>1.58 (0.66-3.78)</td>
<td></td>
</tr>
<tr>
<td>rs12025416</td>
<td>1</td>
<td>Cases vs population-based controls</td>
<td>11, 57, 147</td>
<td>179, 2,643, 9,097</td>
<td>0.184</td>
<td>0.00035</td>
<td>1.57 (1.22-2.01)</td>
<td>CDS8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cases vs disease-matched controls</td>
<td>11, 57, 147</td>
<td>1, 32, 111</td>
<td>0.118</td>
<td>0.022</td>
<td>1.69 (1.08-2.56)</td>
<td></td>
</tr>
<tr>
<td>rs4845320</td>
<td>1</td>
<td>Cases vs population-based controls</td>
<td>3, 16, 198</td>
<td>2, 290, 11,627</td>
<td>0.051</td>
<td>1.98 x 10^-10</td>
<td>4.25 (2.72-6.65)</td>
<td>LCE4A-Clof68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cases vs disease-matched controls</td>
<td>3, 16, 198</td>
<td>0, 3, 143</td>
<td>0.010</td>
<td>0.008</td>
<td>4.09 (1.21-13.85)</td>
<td></td>
</tr>
<tr>
<td>rs3127214</td>
<td>6</td>
<td>Cases vs population-based controls</td>
<td>3, 23, 191</td>
<td>11, 522, 11,386</td>
<td>0.067</td>
<td>1.84 x 10^-8</td>
<td>2.96 (2.03-4.33)</td>
<td>TAGAP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cases vs disease-matched controls</td>
<td>3, 23, 191</td>
<td>0, 8, 137</td>
<td>0.027</td>
<td>0.026</td>
<td>2.51 (1.13-5.55)</td>
<td></td>
</tr>
</tbody>
</table>

Chr, chromosome; CI, confidence interval; MAF, minor allele frequency; MDS, multidimensional scaling; OR, odds ratio.

P-values are derived from logistic regression analysis by including MDS components as covariates.

Table 2 | Clinical characteristics of the candidaemia cohort.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Controls</th>
<th>Patients</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (s.d.) (years)</td>
<td>60.2 (17.5)</td>
<td>54.7 (20.2)</td>
<td>0.01</td>
</tr>
<tr>
<td>Male gender (%)</td>
<td>49.3</td>
<td>64.0</td>
<td>0.008</td>
</tr>
<tr>
<td>Immunocompromised state (%)</td>
<td>36.3</td>
<td>61.7</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Hematopoietic stem cell transplantation (%)</td>
<td>0</td>
<td>2.8</td>
<td>0.06</td>
</tr>
<tr>
<td>Solid organ transplant (%)</td>
<td>0.7</td>
<td>15.4</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Active malignancy* (%)</td>
<td>21.9</td>
<td>35.0</td>
<td>0.008</td>
</tr>
<tr>
<td>Solid tumour (%)</td>
<td>12.3</td>
<td>26.0</td>
<td></td>
</tr>
<tr>
<td>Leukaemia (%)</td>
<td>6.2</td>
<td>5.7</td>
<td></td>
</tr>
<tr>
<td>Lymphoma (%)</td>
<td>4.1</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>Chemotherapy within the past 3 months (%)</td>
<td>12.3</td>
<td>18.9</td>
<td>0.11</td>
</tr>
<tr>
<td>Neutropenia (ANC &lt;500 cells per mm³) (%)</td>
<td>2.7</td>
<td>10.3</td>
<td>0.008</td>
</tr>
<tr>
<td>HIV-infected (%)</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>Surgery within the past 30 days (%)</td>
<td>56.2</td>
<td>49.1</td>
<td>0.21</td>
</tr>
<tr>
<td>Receipt of total parental nutrition (%)</td>
<td>3.4</td>
<td>21.5</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Dialysis dependent (%)</td>
<td>4.1</td>
<td>11.6</td>
<td>0.02</td>
</tr>
<tr>
<td>Acute renal failure (%)</td>
<td>15.8</td>
<td>31.6</td>
<td>0.001</td>
</tr>
<tr>
<td>Liver disease (%)</td>
<td>2.8</td>
<td>19.4</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Intensive care unit admission within the past 14 days (%)</td>
<td>34.2</td>
<td>54.4</td>
<td>0.0003</td>
</tr>
<tr>
<td>Candida spp.† (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>albicans</td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>glabrata</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>parapsilosis</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tropicalis</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>krusei</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Candida spp.</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline serum creatinine, mean (s.d.) (mg dl⁻¹)</td>
<td>1.3 (1.0)</td>
<td>1.6 (1.4)</td>
<td>0.008</td>
</tr>
<tr>
<td>Baseline WBC count, mean (s.d.) (cells per mm³)</td>
<td>10.6 (7.8)</td>
<td>13.5 (13.7)</td>
<td>0.02</td>
</tr>
</tbody>
</table>

ANC, absolute neutrophil count; HIV, human immunodeficiency virus; WBC, white blood cell.
A few patients had positive cultures with >1 Candida species, explaining the slightly higher sum of percentages than 100%.
*Subjects could have >1.
†8 Subjects had >1 species isolated.

A significant upregulation of CD58 in response to Candida infection at 6 h (Fig. 2b; P = 0.04), whereas no difference was seen at 24 h, suggesting CD58 as an early-response gene in host defence against Candida. To validate the functional role of CD58 in anti-Candida response further, we investigated the phenotypes of macrophages with silenced CD58. It has been demonstrated that yeast-to-hyphal transition is one of the virulence factors for Candida to escape macrophage phagocytosis. As expected, live Candida through germ-tube formation could escape macrophage phagocytosis in the control siRNA-transfected group, as well as in the TAGAP siRNA group. Strikingly, a massive fungal outgrowth with extensive fungal hyphae formation was observed in the CD58 siRNA-transfected group (Fig. 2c), showing that CD58 is important for Candida phagocytosis and inhibition of germination. To validate this possibility further, the colocalization of CD58 and Candida in the phagosome was examined by fluorescence microscopy. Upon Candida phagocytosis, a clear recruitment of CD58 (green) around the calcifluor white-labeled Candida (blue) was observed, indicating co-localization of CD58 and Candida during phagocytosis (Supplementary Fig. 5). When a yeast-locked Δkgl1 C. albicans strain was used, there was no defect in the control of fungal growth in cells transfected with CD58 siRNA (Fig. 2d).

As the transcriptome and pathway analyses in CD58-deficient macrophages implicated altered levels of IL-6 and TNF-α secretion, we tested the role of CD58 SNPs with these cytokine levels. The cytokines IL-6 and TNF-α were quantified from macrophages stimulated with either lipopolysaccharide (LPS) or Candida. The top CD58 SNP, rs17035850, is a rare SNP with very low risk allele frequency. Therefore, we tested the second CD58 top SNP, rs12025416, which occurs more frequently, for association with macrophages stimulated with either lipopolysaccharide (LPS) or Candida. The top CD58 SNP, rs17035850, is a rare SNP with very low risk allele frequency. Therefore, we tested the second CD58 top SNP, rs12025416, which occurs more frequently, for association with cytokine levels. Functional genetic validation showed that CD58 SNPs rs12025416 genotypes modulated altered levels of IL-6 and TNF-α secretion, confirming the specific role of CD58 polymorphisms in response to Candida infection.

Functional validation of TAGAP for anti-Candida host defence. Because TAGAP siRNA inhibition in primary
macrophages was not efficient, we interrogated the GeneNetwork database to predict TAGAP function based on co-expression data extracted from ~80,000 Affymetrix microarray experiments (see www.genenetwork.nl/genenetwork). Several gene ontology molecular functions, such as chemoattractant receptor activity and cytokine production, were strongly influenced by TAGAP (Fig. 3a). To validate these pathways, the role of TAGAP was tested in an in vivo disseminated candidiasis model in Tagap−/− mice. While the fungal loads did not differ between control and knockout mice early during infection, the Tagap−/− mice were incapable of clearing the fungus from organs at the later stages of infection (Fig. 3b). In addition, splenocytes isolated from Tagap−/− mice produced significantly less TNF-α compared with control animals (Fig. 3c).

Discussion

This first genome-wide association study assessing genetic susceptibility to a fungal infection identified three novel risk factors for susceptibility to candidaemia, namely CD58, the LCE4A-C1orf68 locus and TAGAP. Carrying two or more risk alleles from these loci increases the risk for candidaemia by >19-fold (Supplementary Fig. 6). Using transcriptomic analysis and immunological validation, we identified unknown roles of CD58 and TAGAP in host defence against Candida species.

Genetic association studies are extremely challenging in systemic fungal infections owing to the inherent difficulty of the relatively low number of patients available. The difficulty of enrolling a large number of candidaemia patients is exemplified by the fact that earlier cohorts had no more than 60 patients16,17. To surpass these difficulties, we took several steps. First, we increased the number of patients by combining patients of European descent from Duke University and Radboud University Nijmegen Medical Center. Second, for the discovery analysis we used a large cohort of 11,920 population-based controls. Analysing a large number of population-based controls against a small number of cases could potentially yield spurious associations owing to population substructure. Therefore, we strictly relied on confirming the significant findings by comparing the patients with the underlying disease-matched control cohort. Additional validation studies were performed in independent cohorts. Third, we focused on genetic variants with the strongest signal from the genome-wide association study of Candida bloodstream infections16. Finally, the gene expression analysis identified 93 genes differentially expressed in control and patient macrophages.

Figure 2 | Functional validation of CD58 in antifungal host defense. (a) Common phenotype enrichment in the mouse using the Mouse Genome Informatics (MGI) phenotype data of the 93 genes differentially expressed at 24 h between control and CD58 siRNA cells. (b) CD58 gene expression levels in wild-type macrophages at 0, 6 and 24 h upon Candida infection were extracted from the microarray experiment. (c) Human monocyte-derived macrophages were transfected with control and CD58 siRNA, respectively, for 48 h. Macrophages were co-incubated with live C. albicans for an additional 24 h. Representative microscopic photos of the Candida outgrowth through the macrophage are shown. The magnification is × 40. (d) First panel: macrophages were transfected with control and TAGAP siRNA, respectively, for 48 h. Macrophages were co-incubated with live C. albicans for an additional 24 h. Second panel: control and CD58 siRNA-transfected macrophages were co-incubated with live C. albicans hgc1 strain for an additional 24 h. Representative microscopic photos of the Candida outgrowth through the macrophage are shown. (e) Human monocyte-derived macrophages from 66 healthy volunteers with different CD58 SNP rs12025416 genotypes were stimulated with C. albicans for 24 h (grey bars) or with LPS for 24 h (blue bars), and the supernatant was collected for IL-6 and TNF-α measurements. Average values of Candida-induced IL-6 levels among CT individuals is 62.68 and among TT is 218.19; whereas for LPS it is 592.71 in CT group and 650.26 in TT group. Average values of Candida-induced TNF-α levels among CT individuals is 603.57 and among TT is 1,122.88; whereas for LPS it is 570.02 in CT group and 648.36 in TT group. The error bars indicate s.e.m.. The correlation between cytokine production and genotypes was tested by the Wilcoxon rank sum test. NS; not significant.
The role of the *LCE4A*-C1orf68 locus for anti-*Candida* host defence could have been anticipated from its involvement in the barrier function of the epithelium, as *Candida* colonization of the mucosae is one of the main risk factors for candidaemia in at-risk patients. In contrast, nothing was known regarding a potential role for TAGAP in antifungal host defence. Using a GeneNetwork microarray database and co-expression analysis, and immunological validation in mice with a genetic defect in *Tagap*, we demonstrate its role in *Candida*-induced inflammation and antifungal host defence.

Interestingly, the genes we have identified as being involved in the susceptibility to candidaemia are also involved in the genetic susceptibility to immune-mediated diseases. Polymorphisms in *CD58* have been reported to increase susceptibility to multiple sclerosis and rheumatoid arthritis, *LCE4A-C1orf68* locus variants are associated with rheumatoid arthritis and psoriasis, and TAGAP polymorphisms influence several autoimmune inflammatory diseases. These data point to strong similarities between the immune-mediated mechanisms involved in the host defence against fungal pathogens and those for immune-mediated pathology. This hypothesis is strengthened by the associations described between anti-*Candida*-specific antibodies and Crohn’s disease. Similar shared relationships have been proposed for other pathologies such as leprosy and...
Crohn’s disease\(^2\), and it has even been hypothesized that the genetic susceptibility to autoimmune diseases in modern human populations was shaped by evolutionary pressure exerted by infections\(^3,12\).

In conclusion, this first, unbiased, genetic association study in a fungal infection demonstrates the potential of functional genomics approaches to identify novel risk factors in infections even in clinical conditions, in which a limited number of patients are available. Our study highlights genetic variants in three novel antifungal host defence mechanisms that increase susceptibility to candidemia.

**Methods**

**Discovery cohort.** Adult candidemia patients were enrolled after informed consent at the Duke University Hospital (DUMC, Durham, North Carolina, USA) and Radboud University Nijmegen Medical Centre (Nijmegen, The Netherlands). Disease-matched controls were enrolled after informed consent at the Duke University Hospital (DUMC). The study was approved by the institutional review boards (Supplementary Table 6) at each study centre, and enrolment occurred between January 2003 and January 2009 (refs 33,34). Patients must have had at least one positive blood culture for a *Candida* species. A total of 217 patients of European ancestry (36 of the Netherlands origin and 181 of USA origin) were included in the study (for clinical characteristics, see Table 2).

Two different control groups were employed. One, consisting of 11,920 population-based individuals of European ancestry\(^26,27\), was used for the discovery phase of the study. The second control group consisted of 146 candidiasis-free but otherwise matched patients. Non-infected patients were recruited from the same hospital wards as infected patients so that co-morbidities and clinical risk factors for infection were similar. This second control group was used to confirm the candidate SNPs identified in the discovery phase. The infected subjects at DUMC were followed prospectively for up to 12 weeks following diagnosis of candidaemia to determine their clinical outcome. Disseminated infection was defined as the presence of *Candida* species at normally sterile body sites other than the bloodstream or urine. Persistent fungemia was defined as ≥5 days of persistently positive blood cultures.

**Validation cohorts.** Two independent cohorts of patients with candidemia were used for validation\(^28\). First, a cohort of African-American descent with candidemia was recruited at DUMC, and 54 patients of African-American descent without candidaemia from the same wards served as a disease-matched control group. Second, a cohort of 69 European surgical intensive care unit patients, including 29 cases of invasive candidiasis (without candidaemia) and 40 non-infected controls, were recruited for the Funginos study, as described elsewhere\(^29\).

**Genotyping and quality control.** The case and the control samples of the discovery cohort were genotyped using the Illumina chip according to Illumina’s protocols\(^28\). We applied SNP quality-control filters to exclude SNPs with (a) a low call rate (<99%), (b) a Hardy–Weinberg equilibrium of \(P < 0.01\) in control samples only and (c) a MAF of <0.01. In the end, 118,989 SNPs were used for case-control analysis. We also excluded 54 samples with a low genotyping rate (<98%) and 40 ethnic outlier samples based on multidimensional scaling analysis (Supplementary Fig. 7\(^2\)). We included 217 candidaemia cases and 11,920 controls of Caucasian descent in the discovery phase of the case-control association analysis. The replication cohorts were genotyped at four SNPs using the Competitive Allele-Specific PCR (KASP) system according to the manufacturer’s protocol (LG Genomics; http://www.lgggenomics.com, formerly KBioscience). The KASP allele-specific forward primers and common reverse primer were designed by Kraken assay design and workflow management software (LG Genomics, formerly KBioscience). Results were analysed on ClusterKiller software (LG Genomics, formerly KBioscience) according to standard protocols and quality controls.

**Statistical analysis.** In the discovery phase, the associations between Immunochip SNPs and candidaemia susceptibility were tested by logistic regression after adjusting for the first four components from the multidimensional scaling analysis using PLINKv1.07 (ref. 37). A strong inflation in Immunochip studies has been observed, as the selection of SNPs is biased towards only loci associated with immune-mediated traits. Therefore, we considered SNPs that map to non-immune regions but are present on the Immunochip to calculate the inflation factor. Comparison of the genetic inflation factor of all SNPs (\(λ = 1.22\)) with the genetic inflation factor of non-immune SNPs (\(λ = 1.102\)) indicated that there was little population stratification effect (Supplementary Fig. 8). We considered \(P < 5 \times 10^{-8}\) as the threshold for significant association. The association to the top-associated SNPs within the validation and replication cohorts was tested using the Cochran–Armitage trend test, and meta-analysis was conducted using the Mantel–Haenszel method. Heterogeneity across the two cohorts was examined using the Breslow–Day test.

**CD58/Tagap knockdown and phagocytosis experiments.** Human monocyte-derived macrophages were obtained by first allowing peripheral blood mononuclear cells to adhere to the plate for 90 min and the non-adherent cells were washed away by phosphate-buffered saline. The remaining adherent monocytes were differentiated into macrophages by incubating with RPMI containing M-CSF (50 ng ml\(^{-1}\)) and 10% human pooled serum for 6 days\(^38\). Macrophages from five different volunteers were transfected with CD58 siRNA (L-004538-00-0005), Tagap siRNA (L-008711-01-0003) or control siRNA (D-001810-10-20) by Dharmafekt 4 for 2 days (Thermo Scientific). Total RNA was isolated at 6 and 24 h, and global gene expression was profiled using an Illumina Human HT-12 Expression BeadChip\(^39\). Differentially expressed genes by at least 1.25-fold between control and CD58 siRNA cells were subjected to pathway enrichment analysis using GeneNetwork analysis\(^40\). After siRNA transfection, macrophages were exposed to live *C. albicans* at multiplicity of infection of 1 for 24 h, after which the phagocytosis and fungal outgrowth was determined by microscopy. The role of fungal germination was assessed using the yeast-locked Hgc1-deficient *C. albicans* strain (provided by Dr Bernhard Hube, Jena University). Cytokine concentrations were determined by enzyme-linked immunosorbent assay.

**In vitro macrophage stimulation assays.** The effect of SNPs in CD58 on cytokine production was studied in monocyte-derived macrophages isolated from a cohort of 66 healthy Europeans. Macrophages were incubated at 37 °C for 24 h with RPMI culture medium, LPS (10 ng ml\(^{-1}\), Sigma-Aldrich, MO, USA), or heat-killed *C. albicans* yeasts or hyphae (1 × 10\(^5\) microorganisms/ml). Cytokines were measured using an enzyme-linked immunosorbent assay (R&D Systems, MN, USA), and the correlation between cytokine production and genotypes was tested by the Wilcoxon rank sum test.

**Systemic *C. albicans* infection in Tagap-deficient mice.** C57BL/6J and Tagap loss-of-function female mice\(^41\) (8–12 weeks) were used for assessing their susceptibility to *C. albicans*. The Tagap\(^−/−\) mouse was reported in the original publication by Bauer et al.\(^42\), the accession number for the gene targeted in this knockout model is NM_145986, which corresponds to the Tagap gene. However, Bauer et al.\(^43\) refer to this gene as Tagap1. To clarify the gene targeted in these mice, we developed a quantitative reverse transcriptase-PCR method and showed that the targeted mice lacked Tagap mRNA, consistent with the accession number referenced in the original publication. See also MGI ID 3615484 for gene information and MGI ID 3630308 for mouse strain information. The experiments were approved by the Ethics Committee on Animal Experiments of the University of Athens. Mice were injected with live *C. albicans* blastoconidia 5 × 10\(^5\) CFU per mouse. The fungal loads in the liver and kidneys were assessed by microbiological dilution plating on days 3 and 7 after infection. Cytokine production capacity was assessed after stimulation of splenocytes (1 × 10\(^5\) per well) with *C. albicans* (1 × 10\(^5\) microorganisms per well).

**References**


Acknowledgements
This work was supported by the Netherlands Organization for Scientific Research (NWO Zernike grant 916.10.135 to L.F.), the European Research Council (Consolidator Grant, ERC-310372 to M.G.N., and Advanced Grant, ERC-671274 to C.W.) and the Dutch Digestive Diseases Foundation (MLDS W011–30 to C.W.). Research reported in this publication was also supported by the National Institutes of Health under award number K23AI51537 (to M.D.J.). We thank Mathieu Platteau, Astrid Maatman and Gosia Trynka for assisting in the RNA and DNA analysis, and array experiments (Immunochip and gene expression). We also thank Thierry Calandra, Jacques Bille, Frederic Tissot, Frederic Lamoth, Christina Orasch, Philippe Eggermann (Lausanne), Chloe Kaeck, Martin Siegemund, Ursula Flückiger (Basel), Stefan Zimmerman (Bern) and all other members of the Funginos group involved in the validation study. We thank Jackie Senior for editing the final version of the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions

Additional information
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Kumar, V. et al. Immunochip SNP array identifies novel genetic variants conferring susceptibility to candidaemia. Nat. Commun. 5:4675 doi: 10.1038/ncomms5675 (2014).