Original paper

Massively parallel targeted resequencing reveals novel genetic variants associated with aspergillosis in paediatric patients with haematological malignancies

Katarzyna Skonieczna[1], Jan Styczyński[2], Anna Krenska[2], Piotr Stawiński[3,4], Rafał Płoski[3], Katarzyna Derwich[5], Wanda Badowska[6], Mariusz Wysocki[2], Tomasz Grzybowski[1]

[1]Department of Forensic Medicine, Division of Molecular and Forensic Genetics, Faculty of Medicine, Ludwik Rydygier Collegium Medicum, Nicolaus Copernicus University, Bydgoszcz, Poland
[2]Department of Paediatrics, Haematology and Oncology, Faculty of Medicine, Ludwik Rydygier Collegium Medicum, Nicolaus Copernicus University, Bydgoszcz, Poland
[3]Department of Medical Genetics, Medical University in Warsaw, Poland
[4]Department of Genetics, Institute of Physiology and Pathology of Hearing, Warsaw, Poland.
[5]Department of Paediatric Oncology, Haematology and Transplantology, Poznan University of Medical Sciences, Poland
[6]Division of Paediatric Haematology and Oncology, Children Hospital, Olsztyn, Poland

This study aimed to find novel genetic variants of susceptibility to aspergillosis in paediatric patients with haematological malignancies. Complete sequences of fifteen genes of human innate immunity (CCL2, CCR2, CD209, CLEC6A, CLEC7A and ten TLR genes) were studied in 40 patients diagnosed with haematological disorders (20 unaffected and 20 affected by aspergillosis). All samples were sequenced with MiSeq (Illumina) and 454 (Roche Diagnostics) technologies. Statistical significance of the differences between studied groups was determined using the two-tailed Fisher's exact test. Sixty variants of potential importance were identified, the vast majority of which are located in non-coding parts of the targeted genes. At the threshold of $p < 0.000005$, one intergenic (TLR2 rs4585282) and one intronic variant (CLEC6A rs12099687) were found significant between the case and control groups for genotype and allele frequencies, respectively. Rs12099687 in CLEC6A was predicted to constitute an alternative isoform or cryptic splice site, which potentially changes activity of the Dectin-2 protein. Overall, we assume that the two strongest associations reported in this study are expected to be reproducible even in the absence of other evidence, while another twelve associations may be strong enough to justify additional research in larger cohorts.

Key words: aspergillosis, children, human innate immunity genes, massively parallel sequencing, SNP.

Introduction

Invasive aspergillosis (IA) and other infections caused by the filamentous fungus *A. fumigatus* as well as several related pathogenic species of *Aspergillus* essentially concern immunocompromised individuals [1]. Among paediatric patients, primarily affected are those suffering from haematological malignancies and those who were subjected to haemopoietic stem cell transplantation (HSCT) [2]. Approximately

250 children from the Polish population are diagnosed with leukaemia annually. The incidence of IA in these patients is estimated to be 5-24% [3]. However, both diagnosis and treatment of IA are still challenging.

Along with genomic, transcriptomic and proteomic research focused on the infection process and drug resistance of the pathogen alone, studies were also conducted on the pathogen-host interactions [4] and the host response to fungal infections [5].

Additionally, the host's genetic susceptibility to infections is of particular interest and complexity. Since aspergillosis primarily affects immunocompromised patients, most research into the genetic factors possibly associated with the condition has employed a candidate gene strategy and focused on the immune-modulating genes, including the genes for Toll-like receptors (TLRs), the C-type lectin receptors (CLRs), NOD-like receptors (NLRs), cytokines, chemokines and immune receptors. As a result of those studies, a limited number of genetic polymorphisms were reported whose associations with aspergillosis were replicated and/or proved to have biological significance [reviewed in 6]. However, the majority of the studies concerning candidate genes concentrated on a limited number of single nucleotide polymorphisms (SNPs), employing low- or medium-throughput genotyping strategies [6, 7, 8].

In recent years, novel technologies of massively parallel sequencing (MPS) have certainly sped up the process of discovering new genetic variants of potential clinical significance and thus facilitated association analyses at a genomic level [9]. In particular, the targeted re-sequencing approach has made analysis of entire sequences of candidate genes relatively straightforward and cost-effective. Nevertheless, data including complete sequences of the human innate immunity genes in paediatric patients of significant risk factors for aspergillosis are practically non-existent. Searching for new variants possibly associated with the condition, in this study we analyzed with high coverage the sequences of fifteen human innate immunity genes (~230 kb in total) in two groups of paediatric patients with haematological malignancies – affected and not affected by aspergillosis. For targeted re-sequencing of all samples, we used two different MPS approaches.

Material and methods

Protection of human subjects

The study was approved by the Bioethics Committee of the Ludwik Rydygier Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Toruń, Poland (statement no. KB 605/2011). Parents of all patients gave informed written consent to participate in the study.

Patient characteristics

Altogether, 40 unrelated paediatric patients from Poland, treated for haematological malignancies at the Department of Paediatrics, Haematology and Oncology of the Nicolaus Copernicus University in Toruń, Collegium Medicum in Bydgoszcz, Department of Paediatric Oncology, Haematology and Transplantology, Poznań University of Medical Sciences, and Division of Paediatric Haematology and Oncology, Children Hospital, Olsztyn, Poland, were recruited for the study. The case group consisted of 20 individuals affected by aspergillosis, while the control group included 20 persons unaffected by the condition. None of the patients underwent HSCT prior to the study. Clinical characteristics of the studied individuals are given in Table I.

DNA extraction and quantitation

DNA was extracted from buccal swabs using the GeneMatrix Bio-Trace DNA Purification Kit according to the manufacturer's protocols (Eurx, Gdansk, Poland). Quality of DNA extracts was assessed by agarose gel electrophoresis. DNA was quantitated both spectrophotometrically and by real-time PCR, using the Quantifiler Duo DNA Quantification Kit and the ViiA 7 Real-Time PCR System (Thermo Fisher Scientific, Waltham, USA).

Targeted re-sequencing

A total of fifteen candidate genes for human innate immunity (*CCL2* – C-C motif chemokine ligand 2; *CCR2* – C-C motif chemokine receptor 2; *CD209*; *CLEC6A* – C-type lectin domain family 6 member

Table I. Demographic and clinical characteristics of paediatric patients included in the study

Parameter	Cases	Controls
Sex		
Male (%)	13 (65)	14 (70)
Female (%)	7 (35)	6 (30)
Mean age in years (SD)	12.5 (4.6)	7.3 (5.0)
Diagnosis		
AML (%)	8 (40)	1 (5)
ALL (%)	12 (60)	19 (95)
Site of invasive aspergillosis		
Pulmonary (%)	16 (80)	0 (0)
Pulmonary + extrapulmonary (%)	4 (20)	0 (0)

Cases – patients with invasive aspergillosis; Controls – patients without invasive aspergillosis; SD – standard deviation; AML – acute myeloid leukaemia; ALL – acute lymphoblastic leukaemia

A; *CLEC7A* – C-type lectin domain family 7 member A; and ten *TLR* [Toll-like receptor] genes) were chosen for targeted resequencing (Supplementary Table I).

All samples were analyzed using two different target enrichment strategies and respective massively parallel sequencing approaches – HaloPlex Target Enrichment System for Illumina sequencing and NimbleGen SeqCap EZ Target Enrichment System for 454 sequencing.

The construction of a custom gene library from 225 ng of DNA was performed using the HaloPlex Target Enrichment System Kit (Agilent Technologies, Santa Clara, USA) according to the manufacturers' instructions. PCR-amplified target libraries were quantified using the Bioanalyzer High Sensitivity DNA Assay kit and a 2100 Bioanalyzer (Agilent Technologies). After pooling equimolar amounts of differentially indexed samples, the final HaloPlex enrichment pool was sequenced following the protocol for the Illumina MiSeq V3 chemistry 2 × 150-bp paired-end reads (Illumina, San Diego, USA).

NimbleGen SeqCap EZ Choice libraries were prepared according to the manufacturer's instructions (Roche Diagnostics GmbH, Mannheim, Germany). NimbleGen libraries of samples were subsequently labelled with different RLMID adaptors (Roche Diagnostics GmbH) and subjected to 454 rapid library construction (Roche Diagnostics GmbH). Ten libraries (labelled with different RLMIDs) were mixed in equimolar ratios and sequenced on a PTP plate using a GS FLX Sequencer (Roche Diagnostics GmbH). Rapid Library preparation, emulsion PCR (emPCR), and 454 sequencing were performed according to the manufacturer's instructions using Titanium reagents (Roche Diagnostics GmbH).

Analysis of sequence data

Bcl2fastq (1.8.4) software (Illumina) was used for generating sample-specific fastq files. After assessing base quality scores, fastq files were subsequently aligned to the reference human genome hg19 (NCBI build 37) using the BWA tool [10]. The GATK programming framework [11] was used for recalibrating base quality scores and performing local realignment around indels. The Haplotype Caller algorithm from the GATK framework was employed for calling nucleotide variants.

The raw data obtained by the 454 platform were analysed with GS Run Processor v.2.5.3 and GS Reporter v.2.5.3 (Roche Diagnostics GmbH). The DNA sequences obtained by the 454 were analysed with GS Reference Mapper v.2.5.3 (Roche Diagnostics GmbH). For the GS Reference Mapper v.2.5.3 results, all three files (454AllDiffs.txt, 454HCDifffs.txt, and 454AlignmentInfo.tsv) were used to determine each sample's genotype. Several quality control steps were applied to detect DNA polymorphic positions. Heterozygosity was confirmed when the 454 unique reads had a high quality score, when the nucleotide variant that differed from the reference sequence was observed in 20-80% of all reads, and if the ratio of forward to reverse reads for the changed variant was similar to that calculated for the unchanged variant.

Nucleotide variants are presented in subsequent sections of the paper in reference to the original GenBank records for the targeted genes.

Statistical analysis of nucleotide variants

Nucleotide variants were filtered against the dbSNP database [12], 1000 Genomes Project database [13], GWAS central database [14] and ENSEMBL database [15]. Statistical significance of the differences between case and control groups in both allele and genotype frequencies was calculated using two-tailed Fisher's exact test. The odds ratio (OR) and 95% confidence intervals (95% CI) were calculated. For the purpose of initial screening, p values < 0.05 were considered as statistically significant. The SNPs found to be associated with aspergillosis at a conventional statistical significance level ($p < 0.05$) were subsequently scrutinized against the Better Associations for Disease and Genes (BADGE) classification system to assess their potential for reproducibility of associations. The BADGE system organizes genetic associations by five classes based on p-values, from $p < 0.0000002$ (a first-class association) to $p < 0.05$ (a fifth-class association) [16]. Calculations were performed using *STATISTICA* v. 12.5 (StatSoft Inc.) software. The ASSP tool [17] was used to determine the implication of the changes in the noncoding regions of the targeted genes on the splicing process. This web application employs backpropagation networks trained for classification of constitutive and alternative isoform/cryptic splice sites, with confidence ranging from 0 (undecided) to 1 (perfect classification) [17].

Results

Two different enrichment strategies enabled the resequencing of a total of 233 589 base pairs (bp) out of 325 771 bp of the fifteen targeted genes (~72% of the total gene sequence, Supplementary Table II). The uncaptured regions mainly constituted highly repetitive DNA. A total of 1961 genetic variants were identified (Supplementary Table III), out of which 60 polymorphisms showed differences at the conventional level of $p < 0.05$ between the case and control groups for allele and/or genotype frequencies (Table II). Out of the 60 SNPs showing potential association ($p < 0.05$), 55% of the SNPs were not reported previously in the dbSNP database. Thir-

Table II. List of nucleotide variants which were found significant between cases and controls for allele and/or genotype frequencies at the conventional level of $p < 0.05$. P-values are further classified according to the BADGE system (Manly, 2005). One SNP, which was previously found to be associated with pulmonary aspergillosis, is shaded grey

Gene	Position (rs number)	Reference allele {R}	Alternate allele {A}	Localization (amino acid change)	p-value for R vs. A*	p-value for RR vs. RA+AA*	p-value for RA vs. RR+AA*	p-value for AA vs. AR+AA*
TLR2	10437	T	insA	intron	NS	0.0202	NS	NS
	10439	T	A	intron	NS	0.0202	NS	NS
	11186	A	G	intron	NS	0.0248	0.0248	NS
	11445	A	G	intron	0.0015^{IV}	0.0436	NS	0.0436
	12124	T	C	intron	0.0129	NS	0.0012	0.0033
	12166	G	A	intron	NS	0.0484	0.0187	NS
	12196	G	C	intron	NS	NS	0.0104	0.0256
	12206 (rs4585282)	T	C	intergenic	NS	0.0000033II	0.0083	NS
	12317	G	A	intron	NS	0.0138	NS	NS
	12607	G	A	intron	NS	0.0471	NS	NS
	12767	T	G	intron	0.0145	NS	NS	0.0225
	12771	A	G	intron	0.0013^{IV}	0.0004 IV	NS	NS
	12779 (rs115889304)	C	A	intergenic	NS	NS	0.0057	0.0057
	12812	T	C	intron	0.0001^{IV}	0.0036	0.0036	NS
	12820	T	C	intron	0.0001^{IV}	NS	0.0033	0.0001^{IV}
	12822	T	G	intron	0.0015^{IV}	NS	NS	0.0057
	12833	T	G	intron	NS	NS	0.0033	0.0138
TLR2	12913	T	C	intron	0.0289	NS	0.0197	0.0197
	13028	G	C	intron	NS	NS	0.0436	0.0436
	13054	C	T	intron	NS	0.0309	0.0138	NS
	13233	C	T	intron	0.0339	0.0095	0.0079	NS
	22209 (rs1439164)	C	T	upstream region	0.0269	NS	NS	0.0407
TLR3	647:648 (rs71593671)	CAAAAAAA-AAAAAAA	del14A, insA, ins2A	upstream region	0.0210 for ins2A vs. insA + R	NS	0.031 for „R/insA vs. others"; 0.0202 for „ins2A/ins2A vs. others"	
	10081	G	A	intron	NS	0.0471	0.0471	NS
	10082	C	G	intron	NS	0.0471	0.0471	NS
TLR5	706 (rs2487835)	C	T	upstream region	NS	NS	NS	0.0407
	15519 (rs851181)	G	A	intron	NS	NS	NS	0.0407
	16482 (rs851187)	C	T	intron	0.0435	0.041	NS	NS
	32787	G	A	intron	NS	NS	0.0484	NS
	33310	G	C	intron	NS	0.0407	0.0407	NS
	39915 (rs4661281)	C	T	downstream region	NS	0.0407	0.0248	NS
	39916 (rs4661280)	C	A	downstream region	NS	0.0407	0.0248	NS

Table II. Cont.

Gene	Position (rs number)	Reference allele {R}	Alternate allele {A}	Localization (amino acid change)	p-value for R vs. A*	p-value for RR vs. RA+AA*	p-value for RA vs. RR+AA*	p-value for AA vs. AR+AA*
TLR7	3946	G	insA	intron	NS	0.0436	0.0436	NS
	3947	C	G	intron	NS	0.0436	0.0436	NS
	4838	G	delG	intron	NS	0.0471	0.0471	NS
	4840	A	G	intron	NS	0.0471	0.0471	NS
	29277 (rs1634318)	A	G	promoter flanking	0.0143	NS	NS	NS
	29281 (rs1616583)	C	G	promoter flanking	0.0143	NS	NS	NS
	29918 (rs771765953)	C	ins 5A, 6A or 12A	near the end of the enhancer	0.0129 for C vs. others	NS	NS	NS
TLR8	962 (rs178994)	G	T	upstream region	0.0476	NS	NS	NS
	7448 (rs2109134)	T	A	promoter flanking	0.0103	NS	NS	NS
	21788:21789	TCCCCC	del5C	intron	NS	0.0202	0.0202	NS
	21795	A	G	intron	NS	0.0202	0.0202	NS
TLR9	4149 (rs187084)	T	C	promoter flanking	0.0020	NS	NS	0.0197
	8483 (rs352140)	G	A	exon 2 (P545P)	0.0129	0.0484	NS	NS
CCL2	5974 (rs4586)	T	C	exon 2 (C35C)	NS	NS	0.0079	NS
CCR2	9564 (rs1799865)	T	C	exon 2 (N260N)	NS	0.0436	NS	NS
CLEC7A	10345 (rs6488263)	T	C	intron	0.0476	NS	NS	NS
	16814 (rs7959451)	A	G	exon 6 (3’ UTR)	0.0476	NS	NS	NS
	19825 (rs7314021)	C	T	upstream region	0.0482	0.0407	NS	NS
CLEC6A	7012 (rs12099687)	T	A	intron	0.00000023$^{\mathrm{II}}$	0.0033	NS	0.00014$^{\mathrm{IV}}$
	13393	T	C	intron	NS	0.0083	0.0083	NS
CD209	3602 (rs4804804)	C	T	transcription factor binding site	NS	NS	0.0104	NS
	5241	T	C	exon 1 (F25S)	NS	0.0484	0.0484	NS
	5242 (rs762628366)	C	T	exon 1 (F25L)	NS	0.0484	0.0484	NS
	8378 (rs745660447)	T	A	exon 6 (S307T)	NS	0.0471	NS	NS
	8386 (rs17159887)	A	G	exon 6 (R309R)	NS	0.0471	NS	NS
	8399	A	T	exon 6 (T314S)	NS	0.0471	NS	NS
	8532	A	G	intron	NS	NS	NS	0.0202
	12033 (rs560344977)	G	A	exon 7 (3'UTR)	NS	0.0471	0.0471	NS

**p-values for two-tailed Fisher exact test*

R – reference allele; A – alternate allele; NS – non-significant; II – second-class association according to the BADGE system ($p < 0.000005$); IV – fourth-class association according to the BADGE system ($p < 0.002$); p-values without any superscript denote fifth-class associations according to the BADGE system ($p < 0.05$)

ty-five nucleotide variants were located in intronic regions, eight were found in exons (four of which were non-synonymous), seven were located in upstream or downstream regions of genes and two were identified in 3′-untranslated regions (UTRs). Two SNPs were intergenic, whereas six other polymorphisms were found in well-characterized regulatory sequences (promoter flanking sites, a transcription factor binding site and near the end of an enhancer) (Table II).

Only one SNP, identified in this study as showing a significant differences for allele frequencies between cases and controls at $p < 0.05$ (rs7959451 at 3′UTR of *CLEC7A*), was recently found to be associated with aspergillosis in a population of European ancestry [18]. However, 16 other SNPs previously identified as possibly determining susceptibility to the condition did not reach a level of significance in our research (Supplementary Table IV). Conversely, two common SNPs in the *TLR9* gene (promoter flanking rs187084 and rs352140 in the second exon) which showed significance for both allele and genotype frequencies ($p < 0.05$) in our analysis (Table II) did not reach the level of significance in an earlier association study on the risk of invasive aspergillosis among recipients of allogeneic haematopoietic cell transplants [19]. Nevertheless, many other studies have proved associations of rs187084 and rs352140 with numerous immune-related diseases, response to therapies, as well as bacterial and viral infections in populations of both Western and Eastern Eurasian ancestry [20, 21, 22, 23, 24, 25]. It is also worth noting that the common rs4586 SNP, flanking the promoter of *CCL2* gene, which appeared significant for genotype frequencies between our cases and controls ($p = 0.0079$, OR = 8.5, Table II), was recently found to be significantly correlated with anti-tumour immune reaction in Korean patients with colorectal cancer [26] and associated with the clinical outcome in Japanese individuals with locoregional gastric cancer [27]. Moreover, the *TLR2* intergenic polymorphism rs4585282 (previous rs numbers: rs6535939, rs61329579), which was found in our study as significant for genotype frequencies (at $p = 0.0000033$ or $p = 0.0083$, depending on the model of association, Table II), was recently indicated as displaying modest associations with some vitiligo subgroups [28], while the T allele of this SNP was previously reported to be associated with pulmonary tuberculosis in a West African but not a European population sample [29].

The ASSP analysis did not point to the vast majority of the huge intronic variation found in this study to alter the splicing mechanism. However, rs12099687 in *CLEC6A*, which was found to be significant for allele frequencies between the two groups of patients ($p = 0.00000023$, Table II), constitutes notable exception. The ASSP application predicted the rs12099687 position as an alternative isoform/cryptic splicing donor site with very high confidence (0.927).

After filtering the associations against the very conservative BADGE system, the vast majority of associations were classified as fifth class ($p < 0.05$, Table II). However, eight fourth-class ($p< 0.002$) and two second-class associations ($p < 0.00005$) were also identified in our study (in *TLR2* and *CLEC6A* genes, Table II). The strongest second-class association ($p = 0.00000023$) concerns intronic SNP rs12099687 in *CLEC6A* (Table II). According to the BADGE recommendations, the p-value for this association is very close to that of a first-class association ($p < 0.0000002$).

Discussion

According to our knowledge, this is the first study employing massively parallel targeted resequencing in quest of potential susceptibility variants to aspergillosis in complete sequences of the candidate genes. We attempted to capture both exonic and intronic variation of fifteen genes of human innate immunity in two groups of paediatric haematooncological patients of the same ethnic origin. Since considerable homogeneity within the Polish population was previously reported for different genetic markers, a hidden population substructure is unlikely to affect the results of this study. Among the targeted genes, we included all ten genes for Toll-like receptors (TLRs), some of which were recently reported to have alleles critical to the innate immune response, introgressed from archaic humans [30]. The five remaining genes included in our research (*CCL2*, *CCR2*, *CD209*, *CLEC6A*, *CLEC6A*) were previously subjected to case-control studies concerning aspergillosis. However, even the most comprehensive of those studies included only a limited number of the known SNPs [7].

As a result of targeted enrichment and resequencing of all samples using two different MPS strategies, we captured ~70% of the studied genes, excluding highly repetitive DNA sequences located mostly in non-coding regions. In our analysis of resulting sequence data, we concentrated on both known and unknown variation, subsequently filtering differences in allele and genotype frequencies between the two groups of patients against five thresholds of significance, according to the BADGE recommendations [16].

Most studies in medical genomics still address coding parts of the human genome [9]. However, intronic variation has continually been attracting more attention, since introns are currently acknowledged to possess regulatory elements that can influence both gene expression and splicing [reviewed in 31]. Expectedly, we identified huge variations in the introns of the targeted genes, a limited proportion of which were previously characterized and presented in genomic databases. In particular, we found five novel intronic variants in the *TLR2* gene apparently showing differences for allele frequencies between cases and controls at $p < 0.002$. However, their functional significance is difficult to assess at this point in time without other comparable genomic data. Meanwhile, out of all intronic SNPs described in this study, rs12099687 in *CLEC6A*, which showed the strongest association with aspergillosis for allele frequencies (at $p = 0.00000023$), was predicted to constitute an alternative isoform or cryptic splice site. It is worth noting that the predictive power of alternatively or cryptically spliced sites using ASSP application is generally lower than that of constitutively spliced sites [17]. Nevertheless, the confidence of prediction for rs12099687 was very high (0.927). This strongly supports a potential role of non-canonical splicing of *CLEC6A* gene coding for Dectin-2 protein in susceptibility to aspergillosis. Interestingly, it was recently proved that Dectin-2 (but not Dectin-1) present on plasmacytoid dendritic cells (pDCs) directly recognizes hyphae of *A. fumigatus*, which triggers cytokine release and subsequent antifungal activity [32]. Considering this, non-canonical splicing of the *CLEC6A* gene may result in changing activity of Dectin-2, ultimately affecting a host's response to fungal infection.

Among the other SNPs implicated in potential association with aspergillosis in our study, of particular interest is a common promoter flanking polymorphism rs187084 in the *TLR9* gene. While it was not previously reported to be associated with the condition, many studies have revealed its involvement in the immune response in a variety of diseases, including auto-immune disorders, infectious diseases and cancers, as well as pharmacotherapies. These observations suggest the common mechanism that underlies the acting of this functional variant, presumably by down-regulating TLR9 expression and thus affecting the immune cascade initiated by TLR9 activation [33]. Similar regulatory functions at the level of transcription may be attributed to *TLR9* synonymous variants rs352140 and rs4586 flanking the promoter of the *CCL2* gene, of which the latter was reported to affect the immune response to some tumours. Overall, given the location of non-coding variants identified in our study as potentially associated with aspergillosis (Table II, excluding intronic variation), at least 28% of SNPs may be considered as potentially or actually affecting gene regulation. Interestingly, variation in *TLR* genes in modern humans that was recently reported as introgressed from Neanderthals and pivotal for human innate immunity also emphasizes the importance of differences in gene expression. Importantly, all *TLR* SNPs introgressed from Neanderthals are located in non-coding parts of the human genome (of which 47% lie in introns) and many of them overlap known regulatory sites [30]. It is also worth noting that gene regulation is often influenced by epistatic interactions. Possibility of SNP-SNP epistasis in invasive aspergillosis was actually suggested for selected variants of *CLEC7A*, *CCL2* and *CCR2* [34]. Overall, the findings of this and other studies contribute to our understanding of the complicated nature of potential susceptibility to aspergillosis, underlining the importance of non-canonical splicing, gene regulation and gene-gene interactions.

We acknowledge that association studies employing small population samples are usually underpowered in terms of detecting weak associations [17]. However, we also consider recruiting large groups of paediatric patients affected by aspergillosis as essentially unrealistic at the first, discovery stage of analysis. Therefore, for assessing the replication potential and actual strength of the reported associations, we applied very conservative BADGE recommendations based on p-values of associations. Two of the reported associations, classified as second class (*TLR2* rs4585282 and *CLEC6A* rs12099687), are most promising in terms of replication. In fact, the p-value for *CLEC6A* rs12099687 is very close to that of a first-class association ($p < 0.0000002$), suggesting replication, under conservative assumptions, without any prior evidence [17]. Seven fourth-class associations ($p < 0.002$) reported for the novel intronic SNPs in TLR2 may be strong enough to justify additional replication efforts, although their actual potential is difficult to assess without other comparable data. The remaining associations reported in this study are classified as fifth class ($p < 0.05$). While a fifth-class association itself does not provide assurance of reproducibility, it may be considered as much more reliable when replicated in another population sample [17]. For example, SNPs rs187084, rs352140 (*TLR9*) and rs4586 (*CCL2*), which are formally implicated in fifth-class associations in our study, were previously proved to be involved in the immune response in a variety of conditions. Similarly, rs7959451 at the 3'UTR of *CLEC7A*, showing differences for allele frequencies between our groups of patients (at $p < 0.05$), was recently found to be associated with bronchopulmonary aspergillosis in asthma. These independent findings constitute additional evidence justifying further replication studies.

Conclusions

In summary, our research should be considered as initial discovery and screening of complete genes in the search for genetic variants which deserve to be further replicated in larger cohorts. We attempted to cover for the first time the complete sequences of fifteen genes of innate immunity in two groups of paediatric patients affected by haematological disorders. Using two different MPS approaches, we identified a number of variants of potential importance in host susceptibility to aspergillosis, the vast majority of which are located in intronic and other non-coding parts of the targeted genes. Most of the variants indicated in this study were not previously reported in relation to aspergillosis, although several common SNPs were previously implicated in associations with other immune-related conditions. Some of these common variants may have functional significance, influencing splicing and gene expression.

This work was supported by the National Science Centre (Poland) {grant number 2011/03/B/NZ6/01728}.

The authors declare no conflicts of interest.

References

1. Walsh TJ, Anaissie EJ, Denning DW, et al. Treatment of aspergillosis: clinical practice guidelines of the Infectious Diseases Society of America. Clin Infect Dis 2008; 46: 327-360.
2. Morgan JE, Hassan H, Cockle JV, et al. Critical review of current clinical practice guidelines for antifungal therapy in paediatric haematology and oncology. Support Care Cancer 2017; 25: 221-228.
3. Herbrecht R, Bories P, Moulin JC, Ledoux MP, Letscher-Bru V. Risk stratification for invasive aspergillosis in immunocompromised patients. Ann N Y Acad Sci 2012; 1272: 23-30.
4. Schleicher J, Conrad T, Gustafsson M, et al. Facing the challenges of multiscale modelling of bacterial and fungal pathogen-host interactions. Brief Funct Genomics 2017; 16: 57-69.
5. Dix A, Czakai K, Springer J, et al. Genome-Wide Expression Profiling Reveals S100B as Biomarker for Invasive Aspergillosis. Front Microbiol 2016; 7: 320.
6. Wójtowicz A, Bochud PY. Host genetics of invasive Aspergillus and Candida infections. Semin Immunopathol 2015; 37: 173-186.
7. Lupiañez CB, Villaescusa MT, Carvalho A, et al. Common Genetic Polymorphisms within NFκB-Related Genes and the Risk of Developing Invasive Aspergillosis. Front Microbiol 2016; 7: 1243.
8. Pana ZD, Farmaki E, Roilides E. Host genetics and opportunistic fungal infections. Clin Microbiol Infect 2014; 20: 1254-1264.
9. Auer PL, Lettre G. Rare variant association studies: considerations, challenges and opportunities. Genome Med 2015; 7: 16.
10. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009; 25: 1754-1760.
11. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010; 20: 1297-1303.
12. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001; 29: 308-311.
13. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al.. A global reference for human genetic variation. Nature 2015; 526: 68-74.
14. Beck T, Hastings RK, Gollapudi S, et al. GWAS Central: a comprehensive resource for the comparison and interrogation of genome-wide association studies. Eur J Hum Genet 2014; 22: 949-952.
15. Yates A, Akanni W, Amode MR, et al. Ensembl 2016. Nucleic Acids Res 2016 44: D710-D716.
16. Manly KF. Reliability of statistical associations between genes and disease. Immunogenetics 2005, 57: 549-558.
17. Wang M, Marín A. Characterization and prediction of alternative splice sites. Gene 2006; 366: 219-227.
18. Overton NL, Denning DW, Bowyer P, Simpson A. Genetic susceptibility to allergic bronchopulmonary aspergillosis in asthma: a genetic association study. Allergy Asthma Clin Immunol 2016; 12: 47.
19. Bochud PY, Chien JW, Marr KA, et al. Toll-like receptor 4 polymorphisms and aspergillosis in stem-cell transplantation. N Engl J Med 2008; 359: 1766-1777.
20. Bank S, Andersen PS, Burisch J, et al. Associations between functional polymorphisms in the NFκB signaling pathway and response to anti-TNF treatment in Danish patients with inflammatory bowel disease. Pharmacogenomics J 2014; 14: 526-534.
21. Bharti D, Kumar A, Mahla RS, et al. The role of TLR9 polymorphism in susceptibility to pulmonary tuberculosis. Immunogenetics 2014; 66: 675-681.
22. He D, Tao S, Guo S, et al. Interaction of TLR-IFN and HLA polymorphisms on susceptibility of chronic HBV infection in Southwest Han Chinese. Liver Int 2015; 35: 1941-1949.
23. Karody V, Reese S, Kumar N, et al. A toll-like receptor 9 (rs352140) variant is associated with placental inflammation in newborn infants. J Matern Fetal Neonatal Med 2016; 29: 2210-2216.
24. Lee J, Kim YJ, Lee J. Gene-set association tests for next-generation sequencing data. Bioinformatics 2016; 32: i611-i619.
25. Paradowska E, Jabłońska A, Studzińska M, et al. TLR9 -1486T/C and 2848C/T SNPs Are Associated with Human Cytomegalovirus Infection in Infants. PLoS One 2016; 11: e0154100.
26. Yoon S, Kang BW, Park SY, et al. Prognostic relevance of genetic variants involved in immune checkpoints in patients with colorectal cancer. J Cancer Res Clin Oncol 2016; 142: 1775-1780.
27. Sunakawa Y, Stintzing S, Cao S, et al. Variations in genes regulating tumor-associated macrophages (TAMs) to predict outcomes of bevacizumab-based treatment in patients with metastatic colorectal cancer: results from TRIBE and FIRE3 trials. Ann Oncol 2015; 26: 2450-2456.
28. Traks T, Keermann M, Karelson M, et al. Polymorphisms in Toll-like receptor genes are associated with vitiligo. Front Genet 2015; 6: 278.
29. Velez DR, Wejse C, Stryjewski ME, et al. Variants in toll-like receptors 2 and 9 influence susceptibility to pulmonary tuberculosis in Caucasians, African-Americans, and West Africans. Hum Genet 2010; 127: 65-73.
30. Dannemann M, Andrés AM, Kelso J. Introgression of Neandertal- and Denisovan-like haplotypes contributes to adaptive variation in human toll-like receptors. Am J Hum Genet 2016; 98: 22-33.
31. Hubé F, Francastel C. Mammalian introns: when the junk generates molecular diversity. Int J Mol Sci, 2015; 16: 4429-4452.
32. Loures FV, Röhm M, Lee CK, et al. Recognition of Aspergillus fumigatus hyphae by human plasmacytoid dendritic cells is mediated by dectin-2 and results in formation of extracellular traps. PLoS Pathog 2015; 11: e1004643.
33. Krieg AM. Toll-like receptor 9 (TLR9) agonists in the treatment of cancer. Oncogene 2008; 27: 161-167.
34. Sainz J, Lupiáñez CB, Segura-Catena J, et al. Dectin-1 and DC-SIGN polymorphisms associated with invasive pulmonary Aspergillosis infection. PLoS One 2012; 7: e32273.

Address for correspondence

Katarzyna Skonieczna
Department of Forensic Medicine
Division of Molecular and Forensic Genetics
Collegium Medicum in Bydgoszcz
9 Curie-Sklodowskiej St.
85-094 Bydgoszcz, Poland
tel. +48 12 585 35 52
fax +48 12 585 35 53
e-mail: k.skonieczna@gmail.com