Title: Predicting invasive aspergillosis in haematology patients by combining clinical and genetic risk factors with early diagnostic biomarkers

Authors: P. Lewis White[1], PhD, Christian Parr[2], PhD, Rosemary A. Barnes[2], MD.

Affiliations: [1]Public Health Wales, Microbiology Cardiff, UHW, Cardiff, UK; [2]Infection, Immunity and Biochemistry, School of Medicine, Cardiff University, Cardiff, UK.

Corresponding author: Dr P. Lewis White, Public Health Wales, Microbiology Cardiff, UHW, Cardiff, UK, CF14 4XW.

Telephone Number: +44 (0)29 2074 6581

Fax Number: +44 (0)29 2074 2161

Email: lewis.white@wales.nhs.uk

Key Words: Invasive aspergillosis, risk factors, SNPs, *Aspergillus* PCR

Summary: Stratification of haematology patients according to clinical, genetic and mycological risk of invasive aspergillosis (IA) showed 50% of patients were low risk (<5%) for IA, and <10% were at >50% risk. Personalized medicine provides a strategic approach for managing IA.

ABSTRACT

The incidence of invasive aspergillosis (IA) in high risk haematology populations, is relatively low (<10%), despite unavoidable exposure *Aspergillus* in patients with potentially similar clinical risk. Non-clinical variables including genetic mutations that increase susceptibility to IA could explain why only certain patients develop disease. This study aimed to screen for mutations in 322 haematology patients classified according to IA status, and to develop a predictive model based on genetic risk, established clinical risk factors and diagnostic biomarkers.

Genetic markers were determined by real-time PCR, and with clinical risk factors and *Aspergillus* PCR results were analysed by multi-logistic regression analysis to identify a best-fit model for predicting IA. Probability of IA was calculated and an optimal threshold determined.

Mutations in Dectin-1 (rs7309123) and DC-SIGN (rs11465384 and rs7248637), allogeneic stem cell transplantation, respiratory virus infection and *Aspergillus* PCR positivity were all significant risk factors for developing IA and combined in a predictive model. An optimal threshold requiring three positive factors generated a mean sensitivity/specificity of 70.4%/89.2%, and a probability of developing IA of 56.7%. In patients with no risk factors the probability of developing IA was 2.4%, compared to >79.1% in patients with four or more factors. Using a risk threshold of 50%, pre-emptive therapy would have been prescribed in 8.4% of the population.

Summary

This pilot study shows that patients can be stratified according to risk of IA, providing personalized medicine, based on strategic evidence, for the management of IA. Further studies are required to confirm this approach.

INTRODUCTION

The limited ability to accurately diagnosis invasive aspergillosis (IA) has led to an overreliance on empirical antifungal therapy (1). In recent years the incorporation of highly sensitive non-culture diagnostics (PCR, galactomannan EIA (GM) and β-D-Glucan) has increased diagnostic accuracy enabling disease to be excluded, decreasing unnecessary antifungal use (2-4).Early diagnosis is important for good prognosis and pre-emptive approaches, utilising non-culture based tests, can provide early evidence of infection (2-5). However, biomarkers alone are insufficient to initiate therapy as false positive results occur. Biomarkers are best used to exclude disease when negative, with positivity used to trigger further clinical investigations (*e.g.* Bronchcoscopy and HRCT) to confirm disease.

A pre-emptive strategy of managing IA may have advantages. By definition a pre-emptive approach involves taking action against an anticipated outcome, dependent on the likelihood of the patient developing future disease as determined by risk factors. While this is similar to a prophylactic strategy it differs due to the availability of sensitive biomarker assays, and the necessity for biomarker positivity during the early infective processes, prior to clinically overt disease. The threshold at which pre-emptive action is taken is critical; too low and the number of patients treated may be comparable empirical policies, too stringent and the opportunity to prevent disease is missed. Thresholds can be dynamic, dependent on perceived and potential outcomes of which may also alter the risk.

There are multiple risk factors for IA in the haematology population (6-8). Many are associated with the underlying haematological malignancy and treatment, including prolonged neutropenia, lymphopenia, allogeneic stem cell transplantation (SCT), iron overload, graft-versus-host disease (GVHD), prolonged corticosteroid use, monoclonal antibody use. Infection by other microbes (CMV and/or respiratory viruses), can also increase the risk of IA. Exposure to *Aspergillus* is unavoidable,

but exposure to high levels of the organism during building construction or housing in contaminated environments increases risk of disease.

Recently, genetic predisposition has been recognized as a risk factor for fungal infection. Research has focused on specific nucleotide polymorphisms (SNPs) in genes coding for proteins involved in innate and adaptive immune responses. C-type lectin receptors (Dectin-1, Dectin-2, Mannose binding lectin, DC-SIGN and Mincle) play a primary role in fungal immunity and much research has concentrated on these targets (9). Mutations in other pattern recognition receptors (Toll Like receptors TLRs) have also been associated with increased risk of IA (10). These genetic risk factors although relatively non-specific are present before infection and provide an opportunity to stratify patients according to risk. Risk factor stratification may improve patient management but requires a combination of host, clinical and early diagnostics markers.

Non-culture diagnostic assays can detect early infective processes, and the aim is to detect infection before overt tissue damage occurs. GM is mainly released from actively growing hyphae and presence in the circulation indicates hyphal growth and invasion (11). *Aspergillus* PCR has been shown to be positive earlier than GM and (1-3)-β-D-Glucan (2, 12, 13). In an animal model of IA, blood was *Aspergillus* PCR positive at a time scale that related to exposure rather than disease (13). While this may reduce PCR specificity it may be more beneficial in pre-emptive roles in patients stratified according to risk.

The aim of the study was not to identify novel clinical and genetic risk factors for IA, but, more so, combine risk factors previously associated with the disease, along with a well validated and *Aspergillus* PCR assay, standardized according to international recommendations, into a predictive model to determine the probability of developing IA. Genetic markers in Dectin-1 and DC-SIGN previously associated with IA, were retrospectively combined with established clinical risk factors and *Aspergillus* PCR screening results in a cohort of high risk patient to determine whether a management strategy could stratify patients according to risk, providing personalised medicine to

91 pre-empt disease (14).

MATERIALS AND METHODS

Patient population and study design

As part of the local neutropenic fever care pathway, haematology patients are routinely monitored for invasive fungal disease (IFD) by PCR and GM (2, 15). Patients were included on the certainty of diagnosis assigned by the EORT-MSG criteria (16). If patients were defined as proven, probable or possible IA they were included as cases. If patients had absolutely no radiological or mycological evidence of IA they were included as controls. Subsequently, patients with non-specific radiology and mycological evidence were excluded in order to maintain definitive case/control categorization. Retrospectively, a total of 322 haematology patients were anonymised and stratified according to IA diagnosis, using the revised EORTC-MSG definitions (16).There were 6 proven IA, 48 probable IA, 20 possible IA and 268 patients with no evidence of fungal disease (NEF). Given the lower certainty of diagnosis cases of possible IA were excluded from statistical analysis. Patients were treated for malignancy and other haematological /autoimmune conditions between October 2005 and June 2009. Known risk factors for developing IA were linked to the disease (Table 1.). As all patients are screened as part of the neutropenic fever care pathway neutropenia or fever were not included as risk factors. Information on underlying disease, clinical course (SCT, GVHD, and other infections), radiological history, results of all IA biomarkers, and mortality was gathered as part of the routine clinical management. No additional information outside this remit was sought. Clinical features (Histology/HRCT) or mycological (GM /culture) evidence used to define disease within consensus criteria were not included as risk factors to avoid incorporation bias.

The presence of SNPs, previously associated with increased risk of IA, was determined (14). SNPs were chosen on their availability as commercially available TaqMan assays, providing both methodological simplicity and standardisation, and quality control. The SNP assays were retrospectively performed on stored genomic DNA, previously extracted as bi-product of *Aspergillus* PCR screening. No additional samples specific for this purpose were requested and the local ethics

board ruled the project did not require ethical approval and was approved by the research and development board.

Galactomannan EIA testing

Testing by Platelia *Aspergillus* EIA (Bio-Rad) was undertaken as per the manufacturer's instructions. Optical densities (OD) were read at 450/620nm (Thermo Scientific Multiscan FC). Indices were calculated by dividing the OD of the sample by the mean OD of two threshold controls included in the kit. All samples were considered positive if the sample index was ≥0.5.

***Aspergillus* PCR testing**

Molecular testing was performed as previously described using well established methods that were compliant with international recommendations (2, 15, 17). In brief, DNA was extracted from a minimum of 3ml EDTA whole blood subjected to red and white cell lysis, prior to bead-beating and followed by automated DNA purification/precipitation using the Qiagen EZ1. Both human and *Aspergillus* DNA, when present, are co-extracted. *Aspergillus* PCR targeted the 28S rRNA gene with a limit of detection 3 input copies per reaction, and an internal control was used to monitor for inhibition. Clinical performance of the PCR has been previously described (15, 18).

Determination of genetic risk factors

A total of five SNPs previously associated with IA were available as commercially sourced TaqMan SNP assays (Applied Biosystems) (Table 2) (14, 19). Genomic DNA was extracted as a bi-product of routine *Aspergillus* PCR testing as described above. Prior to testing the concentration and quality (A_{260}/A_{280} ration) of the extracted DNA was determined by nanophotometer (P-300 nanophotometer, Implen, Germany). Genotyping was performed as per manufacturer's recommendations using an ABI7500 fast instrument.

Each sample was anonymously tested in duplicate, when sufficient DNA was available. For patients undergoing allogeneic SCT one sample prior to and one sample several weeks after SCT were tested.

If discordant results were generated additional samples, pre and post SCT, were tested to confirm genotypic switch post SCT.

Statistical analysis

Genotype and clinical risk factor frequencies were compared between proven/probable IA and NEF groups with the use of Fisher's exact test and Pearson X^2 tests. Odds ratios (OR) and 95% confidence intervals (CI) were calculated for the presence (combined homozygous/heterozygous SNP) or absence (homozygous wild-type allele) of the polymorphisms and clinical risk factors. The thresholds for significance for clinical and genetic risk factors were separately adjusted for multiple variables by performing the Bonferroni adjustment method (Clinical factors Significance *P*: 0·0125, Genetic factors significance *P*: 0·01). Pair-wise logistic regression was used to identify any associations between independently significant risk factors and when necessary adjustments were performed by Mantel-Haenszel OR and X^2 test of heterogeneity used to determine the significance of the adjustment. For each SNP consistency with the Hardy-Weinberg Equilibrium was determined by a standard observed-expected X^2 tests with *P* value at 1 degree of freedom. In order to determine any genetic linkage between significant SNPs linkage disequilibrium was calculated and an exclusion threshold of r^2 ≥0·8 was set. In order to provide a clinical utility the probability of IA post-risk factor was also calculated. Receiver operator characteristic (ROC) analysis was performed to determine an optimal threshold for the final model.

RESULTS

Genetic markers and IA

Three SNPs (Dectin-1 rs7309123, DC-SIGN rs11465384 and DC-SIGN rs7248637) showed a significant correlation with proven/probable IA (Table 2). All five SNPs were consistent with Hardy-Weinberg equilibrium. Linkage disequilibrium (LD) analysis showed that SNPs rs11465384 and rs7248637 were in LD with each other (r^2 = 0·6) but did not reach the pre-requisite threshold (r^2 ≥0·8) for exclusion Given the association between DC-SIGN rs11465384 and DC-SIGN rs7248637 (OR: 54·3, 95% CI: 22·8-129·2, *P*: <0·0001) it was decided to consider them a haplotype and combine data for further analysis; the association with IA remained significant (OR: 3·0, 95% CI: 1·6-5·7, *P*: 0·0011). Logistic regression identified only limited associations between clinical risk factors and genetic markers and the data was adjusted accordingly (Table 2). There were no associations between any of the SNPs and other infections (CMV/respiratory virus) included in the study.

As individual assays, none of the significant SNPs could be used to confidently diagnose (highest positive likelihood ratio (LR +tive): 2·48) or exclude IA (lowest negative likelihood ratio (LR –tive): 0·37). Even if all three significant SNPs were present in an individual (9·0% of the population) the LR +tive for developing IA was only 4·0, and if they were negative for all significant SNPs (27·6% of the population) the LR -tive was 0·2.

A multivariate regression model (Model fit *P*: 0·0003) of genetic markers factors confirmed significant associations between the DC-SIGN haplotype (*P*: 0·0126) and Dectin-1 rs7309123 (*P*: 0·0035) and the development of IA. These were retained for inclusion in the final model.

Clinical risk factors for IA

The analysis in this population confirmed previous findings. Significant associations between IA and acute leukaemia/myelodysplastic syndrome (AL/MDS OR: 2·095 CI: 1·115-3·941), use of allogeneic

187 SCT (OR: 2·90, CI: 1·25-6·69) and infections with respiratory viruses (OR: 3·22, CI: 1·44-7·21), mainly
188 driven by respiratory syncytial virus infection (OR: 3·47, CI: 1·38-8·69) (Table 1.).
189 From table 1 it is evident that there is a difference in age between patients with proven/probable IA
190 and those with NEF, but when adjusting for allogeneic SCT, this was not significant. The same was
191 true for CMV infection and GVHD where initial analysis showed a significant association with IA
192 (CMV: OR: 4·301 CI: 1·929-9·561, *P*: <0·0001; GVHD OR: 10·07, CI: 3·78-27·3, *P*: <0·0001), but this
193 was removed after adjustment for SCT. Respiratory virus infections retained a significant
194 independent associated with IA even after adjustment (Table 1).
195 A multivariate regression model (Model fit *P*: <0·0001) of clinical factors confirmed significant
196 associations between developing IA and patients diagnosed with AL/MDS (*P*: 0·0314), patients
197 receiving allogeneic SCT (*P*: 0·0021) and patients with respiratory virus infection (*P*: 0·0043). The
198 odds of developing IA when associated with the various significant clinical factors are shown in Table
199 3. For the AL/MDS population allogeneic SCT and respiratory virus infection were again significantly
200 associated with IA. For the non AL/MDS population allogeneic SCT remained significant whereas
201 respiratory virus infection was not significant, albeit numbers were limited. The following variables,
202 underlying haematological disease, allogeneic SCT and respiratory virus infection, were retained for
203 inclusion in the final model.
204
205 ***Aspergillus* PCR**
206 The sensitivity, specificity, LR +tive, LR –tive and DOR for *Aspergillus* PCR using a single positive result
207 as significant were 92·6% (95% CI: 82·0-97·6), 65·3%, (95% CI: 63·2-66·3), 2·7, 0·1 and 27,
208 respectively. Although there was a strong association between PCR result and IA status (OR: 23·5
209 95% CI: 7·8-79·2, *P*: <0·0001) PCR can be used to confidently exclude IA but diagnosis is hampered by
210 false positivity. Using a multiple positive PCR threshold significantly increased specificity (85·4%, 95%

CI: 83·0-87·3), although it still could not be used solely to confirm a diagnosis of IA (LR +tive: 5·3). *Aspergillus* PCR testing was retained for inclusion in the final model.

Combined prediction Model

In determining the final model clinical risk factors (underlying haematological malignancy (AL/MDS), Allogeneic SCT , respiratory virus infection), presence of significant genetic markers (DC-SIGN haplotype and Dectin-1 rs7309123), and mycological evidence not used to categorise IA disease (*Aspergillus* PCR result) were combined.

On combination of these variables, underlying haematological malignancy was no longer significantly associated with IA (*P*: >0·5772) and was removed from the model. The final model fit was X^2: 106·4 (*P*: <0·0001). The probability of developing IA associated with various combinations of clinical risk factors, genetic markers and biomarker positivity is shown in Figures 1a and 1b. For patients not receiving, or prior to, an allogeneic SCT the risk of developing IA, even in the presence of genetic markers was low (<5%). Consequently, 59·6% (162/272) of the patients without allogeneic SCT could be considered low risk and only patients with multiple genetic markers, who were *Aspergilus* PCR positive and had a respiratory virus infection (1·8% (5/272) of this population), had a risk of IA that exceeded 50% (Figures 1a and Figure 2).

The majority (96%) of patients post-allogeneic SCT were at greater risk of IA (>5%), with significantly higher risk (>50%) in 44% (22/50) of allogeneic SCT patients who were both *Aspergillus* PCR positive and had a respiratory infection, or who had multiple genetic markers and were either *Aspergillus* PCR positive or had a respiratory infection (Figures 1b and Figure 2). The risk of IA in 10% (5/50) of allogeneic patients with multiple genetic markers who were PCR positive and had respiratory virus infection was 89·9%.

When applying this model to the entire haematology population (n=322) 50·9% would be considered at low (<5%) and 8·4% at high (>50%) risk of developing IA. A breakdown of risk is shown in Figure 2.

ROC analysis of the final model generated an AUC of 0·8633 (Figure 3). If one variable was positive the sensitivity was 98·2% (95% CI; 90·1-99·9) and the LR-tive was 0·12, whereas if four or more variables were positive the specificity was >98·15 (95% CI: 95·7-99·4) and the LR+tive >18·8. The mean probability of developing IA in a patient with four risk factors was 79.2%. The optimal threshold to start pre-emptive therapy required three variables to be positive and the sensitivity and specificity were 70·4% (95% CI: 56·4-82·0) and 89·2% (95% CI: 84·8-92·6). The corresponding positive and negative likelihood ratios were 6.50 and 0.332 and in patients with three risk factors the mean probability of developing IA was 56.7%, compared to 6.3% in patients with <3 risk factors present. The number needed to treat was 1.69 (95% CI: 1.6-2.6). Using a threshold of two risk factors positive the sensitivity and specificity of the model were 90·7% (95% CI: 79·7-96·9) and 62·3% (95% CI: 56·2-68·1) and the positive and negative likelihood ratios were 2.41 and 0.149, respectively. The mean probability of developing IA in patient with 32.6%, compared to 2.9% in patients with <2 risk factors present.

DISCUSSION

The purpose of this study was to develop a model combining genetic markers with established clinical risk factors and biomarker screening to stratify haematology patients at risk of developing IA. By linking the high sensitivity of PCR testing (Sensitivity: 92·6%, LR-tive: 0·11) with specificity driven by combined clinical factors and genetic susceptibility (Pair-wise specificity range: 94·0-99·6; LR+tive: 6·2-37) (Figures 1a/1b) diagnostic performance is improved. ROC analysis showed the model was able to both exclude and diagnose IA (Figure 3). Using the optimal threshold of three positive risk factors generated the mean probability of developing IA across the combined haematology population was 56.7%. However, the probability of developing IA varied according to the combination of the three risk factors and in non-allogeneic patients (already missing one risk factor) the presence of four risk factors (Both genetic markers, *Aspergillus* PCR positivity and respiratory virus infection) were required to achieve a risk greater than 50% (N=5 patients). In this population the risk associated with three risk factors ranged from 14.1% to 45.5% (Figure 1a). If a minimum of 3 risk factors were required the mean probability of developing IA in the non-allogeneic population alone was 48.9%. In the allogeneic population the probability of IA developing IA when a patient had three risk factors ranged from 17.9% to 52.5% (Figure 1b). For the allogeneic population alone the mean probability of developing IA using a threshold of at least three positive risk factors was 63.9%. If at least four risk factors were required the probability of developing IA in the allogeneic population alone was 83.3%.

When combining the significant variables it was decided to take a strategic approach based on the probable timeline of evidence available in the clinic. SNP analysis could identify patients with genetic susceptibility to IA who could benefit from disease preventative strategies (either mould active prophylaxis or PCR/GM screening and pre-emptive treatment). Alternatively, a low probability of IA (<5%), could lead to a diagnostic approach with biomarker testing only when disease was clinically suspected.

In this study 41 SCT patients changed alleles in loci where SNPs were significantly associated with IA (Figure 4). Nineteen patients had SNPs removed post SCT and five developed IA, whereas 22 patients had SNPs introduced post SCT and significantly more (17, *P*: 0·0017) developed IA. Further studies are needed but this suggests the need to screen either donors or patients post SCT to determine risk. There is potential to remove high-risk SNPs by finding a suitable donor and to prevent the introduction of SNPs in patients with wild-type alleles pre-SCT.

Post SCT the presence of high-risk SNPs increased the risk of IA >5% and biomarker screening is required (Figure 2). If the patient is subsequently PCR positive, the risk of IA exceeds 48%, and respiratory virus infection increased this further to >75·9%. These patients should be screened by PCR /GM and treated pre-emptively to reduce associated mortality due to delayed diagnosis of IA disease (20).

Choosing optimal genetic markers can be difficult. Many have ethnic or geographical linkage and maybe unique to local population, limiting the widespread potential of the generic strategies. The SNPs used in this study, targeting a ≈95% Caucasian Welsh population (n=322, IA=54) confirm the findings of a previous Spanish study (n=182, IA=57) (14). Further confirmation of the applicability of these SNPs in additional ethnicities is required, and the application of novel SNPs must be investigated. Genome wide association studies and high throughput next generation sequencing will identify further SNPs associated with the IA and provide large amounts of specific information with minimal effort.

The DC-SIGN SNPs in both this and the previous Spanish study were associated with IA, but it is important to exclude the possibility of confounding factors (14).DC-SIGN is reported to have roles in HIV, Hepatitis C and CMV (21-23). This current study found no correlation between CMV infection and the presence of either DC-SIGN SNP (rs114 *P:* 0·9314 and rs724 *P*: 0·4370), but larger studies are needed. The heterogeneity of the patient population, but more so the limited number of certain clinical conditions, including allogeneic stem cell transplant, will limit the ability of the study in

299 determining all risk factors. Large scale multi-centre centres targeting specific populations (e.g.
300 transplant versus no-transplant) are required in order to determine conclusive findings. A further
301 limitation of the study is the limited number of proven cases (n=6). However, the number of
302 probable cases compensates for this and while the level of confidence of diagnosis is less than that
303 for proven IA, it highlights the difficulty in diagnosing IA and represents the usual level of diagnosis
304 attained in the clinical setting. Furthermore, cases of probable IA are regularly accepted in clinical
305 trials of antifungal therapy and their inclusion is equally applicable in this current study. PCR is
306 currently excluded from the revised EORTC/MSG definitions for IFD (16). Data from the audit of this
307 cohort showed most possible IA cases (GM negative be definition) to be *Aspergillus* PCR positive (2)
308 The likely inclusion of PCR in the second revision of the definitions will therefore increase the
309 number of probable cases in this study, but also provide an alternative mycological criterion allowing
310 us to incorporate GM testing into the model without the concern of incorporation bias.
311 In conclusion, this proof of concept study shows that genetic markers in combination with clinical
312 risk and early biomarker positivity can facilitate stratification of patients according to risk of IA. If
313 the probability of developing IA passes 50% then the use of pre-emptive therapy is justified and in
314 this study only 8.3% (27/322) of the population, mainly allogeneic patients (n=22) would receive
315 therapy (Figure 2). It highlights what can be achieved to improve management of patients at risk of
316 difficult to manage fungal diseases, and is an attempt to translate pre-clinical research into a clinical
317 setting (24). For this type of approach to be of use, it must be easily accessible, dynamic to change
318 and simple to use with respect to data input and processing. These systems can be applied to
319 portable electronic devices to permit real-time clinical decision making, individual patient
320 management and a strategic evidence based approach to the management of IA, or other difficult to
321 diagnose diseases. Further studies are needed to determine the range of genetic markers associated
322 with IA and while a recent large scale study by Fisher *et al*. investigated the relevance of previously

documented SNPs using micro-array the application of whole genome sequencing will be essential for this purpose (25).

Conflicts of Interest

PLW is a founding member of the EAPCRI, received project funding from Myconostica, Luminex, and Renishaw diagnostics, was sponsored by Myconostica, MSD and Gilead Sciences to attend international meetings, on a speaker's bureau for Gilead Sciences, and provided consultancy for Renishaw Diagnostics Limited.

RAB is a founding member of the EAPCRI, received an educational grant and scientific fellowship award from Gilead Sciences and Pfizer, is a member of the advisory board and speaker bureau for Gilead Sciences, MSD, Astellas, and Pfizer, and was sponsored by Gilead Sciences and Pfizer to attend international meetings.

CP has no conflicts of interest.

Acknowledgements

This work was funded by a Pfizer Air grant (Grant No WS1836576)

REFERENCES

1) Donnelly JP, Maertens J. 2013. The end of the road for empirical antifungal treatment? Lancet Infect Dis 13:470–2.

2) Barnes RA, Stocking K, Bowden S, Poynton MH, White PL. 2013. Prevention and diagnosis of invasive fungal disease in high-risk patients within an integrative care pathway. J Infect 67:206–14

3) Morrissey CO, Chen SC, Sorrell TC, Milliken S, Bardy PG, Bradstock KF, Szer J, Halliday CL, Gilroy NM, Moore J, Schwarer AP, Guy S, Bajel A, Tramontana AR, Spelman T, Slavin MA; Australasian Leukaemia Lymphoma Group and the Australia and New Zealand Mycology Interest Group. 2013. Galactomannan and PCR versus culture and histology for directing use of antifungal treatment for invasive aspergillosis in high-risk haematology patients: a randomised controlled trial. Lancet Infect Dis. 13:519–28.

4) Aguado JM, Vázquez L, Fernández-Ruiz M, Villaescusa T, Ruiz-Camps I, Barba P, Silva JT, Batlle M, Solano C, Gallardo D, Heras I, Polo M, Varela R, Vallejo C, Olave T, López-Jiménez J, Rovira M, Parody R, Cuenca-Estrella M; PCRAGA Study Group; Spanish Stem Cell Transplantation Group; Study Group of Medical Mycology of the Spanish Society of Clinical Microbiology and Infectious Diseases; Spanish Network for Research in Infectious Diseases. 2014. Serum galactomannan versus a combination of galactomannan and PCR-based *Aspergillus* DNA detection for early therapy of invasive aspergillosis in high-risk hematological patients: a randomized controlled trial. Clin Infect Dis. 60:405–414.

5) Schwarzinger M, Sagaon-Teyssier L, Cabaret O, Bretagne S, Cordonnier C; PREVERT Investigators. 2013. Performance of serum biomarkers for the early detection of invasive aspergillosis in febrile, neutropenic patients: a multi-state model. PLoS One 14: e65776.

6) Barnes PD, Marr KA. 2007. Risks, diagnosis and outcomes of invasive fungal infections in haematopoietic stem cell transplant recipients. *Br J Haematol*. 139:519–31.

7) Baddley JW. 2011. Clinical risk factors for invasive aspergillosis. Med Mycol. 49 Suppl 1: S7–S12.

8) Castagnola E, Viscoli C. 2009. Invasive aspergillosis in malignancy and stem cell transplant recipients. Pages 519–30 in *Aspergillus fumigatus* and Aspergillosis, Eds Latge J-P and Steinbach WJ; ASM Press, Washington DC.

9) Hardison SE, Brown GD. 2012. C-type lectin receptors orchestrate antifungal immunity. Nat Immunol. 13:817–22.

10) Caira M, Mancinelli M, Leone G, Pagano L. 2011. Invasive aspergillosis in acute leukemias: old and new risk factors and new epidemiological trends. Med Mycol. 49 Suppl 1: S13–16.

11) Mennink-Kersten MA, Ruegebrink D, Wasei N, Melchers WJ, Verweij PE. 2006. In vitro release by *Aspergillus fumigatus* of galactofuranose antigens, 1,3-beta-D-glucan, and DNA, surrogate markers used for diagnosis of invasive aspergillosis. J Clin Microbiol. 43: 1711-8.

12) Springer J, Morton CO, Perry M, Heinz WJ, Paholcsek M, Alzheimer M, Rogers TR, Barnes RA, Einsele H, Loeffler J, White PL. 2013. Multicenter comparison of serum and whole-blood specimens for detection of Aspergillus DNA in high-risk hematological patients. J Clin Microbiol. 51: 1445-50.

13) White PL, Wiederhold NP, Loeffler J, Najvar LK, Melchers W, Herrera M, Bretagne S, Wickes B, Kirkpatrick WR, Barnes RA, Donnelly JP, Patterson TF. 2016. Comparison of Nonculture Blood-Based Tests for Diagnosing Invasive Aspergillosis in an Animal Model. J Clin Microbiol. 54(4):960-6.

14) Sainz J, Lupianez CB, Segura-Catena J, Vazquez L, Ríos R, Oyonarte S, Hemminki K, Försti A, Jurado M. 2012. Dectin-1 and DC-SIGN Polymorphisms Associated with Invasive Pulmonary Aspergillosis Infection. Plos One 7: e32273.

15) Barnes RA, White PL, Bygrave C, Evans N, Healy B and Kell J. 2009. Clinical impact of enhanced diagnosis of invasive fungal disease in high-risk haematology and stem cell transplant patients. J Clin Pathol . 62: 64-69.

16) De Pauw B, Walsh TJ, Donnelly JP, Stevens DA, Edwards JE, Calandra T, Pappas PG, Maertens J, Lortholary O, Kauffman CA, Denning DW, Patterson TF, Maschmeyer G, Bille J, Dismukes WE, Herbrecht R, Hope WW, Kibbler CC, Kullberg BJ, Marr KA, Muñoz P, Odds FC, Perfect JR, Restrepo A, Ruhnke M, Segal BH, Sobel JD, Sorrell TC, Viscoli C, Wingard JR, Zaoutis T, Bennett JE; European Organization for Research and Treatment of Cancer/Invasive Fungal Infections Cooperative Group; National Institute of Allergy and Infectious Diseases Mycoses Study Group (EORTC/MSG) Consensus Group. 2008. Revised definitions of invasive fungal disease from the European Organization for Research and Treatment of Cancer/Invasive Fungal Infections Cooperative Group and the National Institute of Allergy and Infectious Diseases Mycoses Study Group (EORTC/MSG) Consensus Group. Clin Infect Dis. 46: 1813–21.

17) White PL, Bretagne S, Klingspor L, Melchers WJ, McCulloch E, Schulz B, Finnstrom N, Mengoli C, Barnes RA, Donnelly JP, Loeffler J; European *Aspergillus* PCR Initiative.2010. *Aspergillus* PCR: one step closer to standardization. J Clin Microbiol. 48: 1231–40.

18) White PL, Linton CJ, Perry MD, Johnson EM, Barnes RA. 2006. The evolution and evaluation of a whole blood polymerase chain reaction assay for the detection of invasive aspergillosis in hematology patients in a routine clinical setting. Clin Infect Dis. 42: 479–86.

19) Chai LY, de Boer MG, van der Velden WJ, Plantinga TS, van Spriel AB, Jacobs C, Halkes CJ, Vonk AG, Blijlevens NM, van Dissel JT, Donnelly PJ, Kullberg BJ, Maertens J, Netea MG. 2011.The Y238X stop codon polymorphism in the human beta-glucan receptor dectin-1 and susceptibility to invasive aspergillosis. J Infect Dis. 203: 736–743.

20) von Eiff M, Roos N, Schulten R, Hesse M, Zühlsdorf M, van de Loo J. 1995. Pulmonary aspergillosis: early diagnosis improves survival. Respiration 62: 341–7.

411 21) Plazolles N, Humbert J-M, Vachot L, Verrier B, Hocke C, Halary. 2011. Pivotal Advance: The
412 promotion of soluble DC-SIGN release by inflammatory signals and its enhancement of
413 cytomegalovirusmediated *cis*-infection of myeloid dendritic cells. Journal of Leukocyte
414 Biology 89: 329–342
415 22) Lozach PY, Amara A, Bartosch B, Virelizier JL, Arenzana-Seisdedos F, Cosset FL, Altmeyer R.
416 2004. C-type lectins L-SIGN and DC-SIGN capture and transmit infectious hepatitis C virus
417 pseudotype particles. J Biol Chem 279: 32035–45
418 23) Geijtenbeek TB, Kwon DS, Torensma R, van Vliet SJ, van Duijnhoven GC, Middel J,
419 Cornelissen IL, Nottet HS, KewalRamani VN, Littman DR, Figdor CG, van Kooyk Y. 2000. DC-
420 SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells.
421 Cell 100: 587–97.
422 24) Head MG, Fitchett JR, Atun R, May RC. 2014. Systematic analysis of funding awarded for
423 mycology research to institutions in the UK, 1997–2010. BMJ Open 4: e004129.
424 25) Fisher CE, Hohl TM, Fan W, Storer BE, Levine DM, Zhao LP, Martin PJ, Warren EH, Boeckh M
425 and Hansen JA. 2017. Validation of single nucleotide polymorphisms in invasive aspergillosis
426 following hematopoietic cell transplantation. Blood 129: 2693-2701.

Table 1. Patient demographics, underlying disease, haematopoetic stem cell transplantation status, biomarker (PCR) positivity and additional viral infections stratified by invasive aspergillosis (IA) classification. Any significant differences between proven/probable IA and patients with no evidence of fungal disease are highlighted in bold text.

	Disease classification				
Parameter	Proven (n=6)	Probable (N=48)	Possible (N=20)	NEF (N=268)	Significance[f] (proven/probable vs NEF)
Ratio male/female	2:1	1.7:1	1.9:1	1.5:1	*P*: 0·7667
Median Age (years)	58	52	64	61	*P*: 0·31[g]
Underlying disease (n)					
Acute leukaemia/MDS	3	25	15	91	***P*: 0·011**
Lymphoma	1	13	3	90	*P*: 0·3389
Myeloma	1	4	0	50	*P*: 0·1134
Chronic Leukaemia	1	4	2	18	*P*: 0·5605
Other[a]	0	2	0	19	*P*: 0·5521
Mortality (%)	66.7	54.2	60.0	45.5	*P*: 0·1302
Transplant rate (%)					
Combined	50.0	58.3	20.0	33.6	***P*: 0·0018**
Allogeneic	50.0	45.8	20.0	9.3	***P*:<0.0001**[h]
Autologous	0.0	12.5	0.0	24.3	***P*: 0·00321**
No transplant	50.0	41.7	80.0	66.4	***P*: 0·0018**
GVHD (n)	2	12	2	9	*P*:0·2757[i]
PCR positivity (n)					
1positive threshold[b]	6	44	15	93	***P*:<0·0001**
≥2 positive threshold[c]	6	36	10	39	***P*:<0·0001**
Additional infections					
CMV (n)	4	11	3	22	*P*:0·3494[j]
Respiratory Virus (n)	1	34	4	38	***P*: 0·0025**[k]
Influenza A or B	0	9	1	11	*P*: 0·2113[k]
Parainfluenza	1	7	1	8	*P*: 0·0988[k]
Rhinovirus	0	5	0	10	*P*: 0·4879[k]
RSV	0	12	2	9	**P: 0·0046**[k]
Adenovirus	0	1	0	0	*P*: 0·2996[k]
Patients with multiple respiratory virus infections	0	8[d]	0	8[e]	*P*: 0·1483[k]

[a] Includes: 8 cases of aplastic anaemia, 3 cases of myelofibrosis, 3 cases of unspecified leukaemia, 2 cases of Paroxysmal nocturnal hemoglobinuria, 2 cases of unspecified haematological malignancy, 1 case of amyloidosis, 1 case of unspecified autoimmune disorder, 1 case of IGM paraproteinanaemia , and 1 case of post transplant lymphoproliferative disorder.

[b] A single PCR positive result is considered significant.

[c] Multiple (≥2) PCR positive results are required to be considered significant.

[d] Includes: 20 infections from eight cases of proven/probable IA. Multiple infections include: 2 cases with influenza/parainfluenza, 1 case with Adenovirus/Rhinovirus, 1 case with influenza/parainfluenza/RSV/rhinovirus, 1 case with influenza/RSV/rhinovirus, 1 case with influenza/rhinovirus, 1 case with parainfluenza/rhinovirus and 1 case with influenza/parainfluenza/rhinovirus. These multiple infections are included in the individual breakdown of respiratory virus infection.

[e] Includes: 18 infections from eight cases with no evidence of invasive fungal disease. Multiple infections include: 4 cases with parainfluenza/rhinovirus, 1 case with influenza/parainfluenza/RSV, 1 case with influenza/RSV/rhinovirus, 1 case with influenza/parainfluenza and 1 case with RSV/rhinovirus. These multiple infections are included in the individual breakdown of respiratory virus infection.

[f] Significance adjusted to 0·0125 according the Bonferroni adjustment method to account for the presence of 4 potential risk factors (underlying disease, stem cell transplantation, GVHD and viral infection).

[g] Adjusted to account for association with allogeneic stem cell transplantation, GVHD and respiratory virus status.

[h] Adjusted to account for associations with GVHD, CMV and respiratory virus status

[i] Adjusted to account for associations with allogeneic stem cell transplantation, CMV and respiratory virus status

[j] Adjusted to account for association with allogeneic stem cell transplantation, GVHD and respiratory virus status

[k] Adjusted to account for association with allogeneic stem cell transplantation, GVHD and CMV status

Key:

NEF: No evidence of fungal infection CMV: Cytomegalovirus

MDS: myelodysplastic syndrome RSV: Respiratory syncital virus

GVHD: Graft versus host disease

Table 2. The prevalence of single nucleotide polymorphisms associated with invasive aspergillosis (IA) in haematology patients with EORTC-MSG defined proven/probable IA (n=54) and patients with no evidence of fungal disease (NEF, n=268). The prevalence of the SNP is based on a combined homozygous/heterozygous allele rate. Significant associations are shown in bold text with significance adjusted to *P:* 0.01 in accordance with the Bonferroni adjustment method.

Gene	Gene locus	ABI-Assay	Proven/Probable IA	NEF	Odds-ratio (95% CI)	*P* value
Dectin-1	rs16910526	C_33748481_10	12/54	51/268	1·2 (0·6-2·6)	0·576
Dectin-1	**rs7309123**	**C_3130832_10**	**47/54**	**173/268**	**3·7 (1·5-9·3)**	**0·001**
DC-SIGN	**rs11465384**	**C_25996399_10**	**15/54**	**30/268**	**3·1 (1·4-6·5)**	**0·004**
DC-SIGN	rs7252229	C_29620333_10	17/54	55/268	1·8 (0·9-3·6)	0·106
DC-SIGN	**rs7248637**	**C_29710787_10**	**18/54**	**37/268**	**2·9 (1·4-5·9)[a]**	**0·001[a]**
Combination	**Haplotype and rs7309123**	**C_3130832_10 C_25996399_10 C_29710787_10**	**20/54**	**31/268**	**4.5 (2.2-9.2)**	**<0.001**

[a] Adjusted for significant associations with AL/MDS.

Table 3. The probability of invasive aspergillosis in the presence of clinical conditions and risk factors.

Risk factor	Percentage of population (n=322)	Odds ratio (CI)	Significance (*P* value)
AL/MDS [a]	37·0%	2·1 (1·1-3·9)	0·011
AL/MDS – Allo SCT + respiratory virus[b]	4·3%	4·2 (1·0-17·7)	0·036
AL/MDS + Allo SCT[c]	6·2%	10·4 (3·2-35·1)	<0·001
AL/MDS + Allo SCT + respiratory virus[d]	2·8%	14·4 (1·1-412·5)	0·029
Other [e]	63·0%	0·5 (0·3-0·9)	0·011
Other – Allo SCT + respiratory virus[f]	4·7%	3·3 (0·6-15·7)	0·107
Other + Allo SCT[g]	9·3%	7·6 (2·8-20·8)	<0·001
Other + Allo SCT + respiratory virus[h]	3·7%	4·0 (0·6-31·6)	0·127

[a] Population used for analysis AL/MDS (N=119, 28 with IA) versus other haematological malignancy (N=203, 26 with IA).

[b] Population used for analysis AL/MDS without allogeneic SCT but positive for respiratory virus infection (N= 14, 5 with IA) versus AL/MDS without allogeneic SCT and no evidence of respiratory virus infection (N=85, 10 with IA).

[c] Population used for analysis AL/MDS with allogeneic SCT, (N=20, 13 with IA) versus AL/MDS without allogeneic SCT (N=99, 15 with IA).

[d] Population used for analysis AL/MDS with allogeneic SCT with respiratory virus infection (N=9, 8 with IA) versus AL/MDS without allogeneic SCT but with respiratory infection (N=14, 5 with IA).

[e] Population used for analysis other haematological malignancy (N=203, 26 with IA) versus AL/MDS (N=119, 28 with IA).

[f] Population used for analysis other haematological malignancy without allogeneic SCT but positive for respiratory virus infection (N= 15, 3 with IA) versus other haematological malignancy without allogeneic SCT and no evidence of respiratory virus infection (N=158, 11 with IA).

[g] Population used for analysis other haematological malignancy with allogeneic SCT, (N=30, 12 with IA) versus other haematological malignancy without allogeneic SCT (N=159, 14 with IA).

[h] Population used for analysis other haematological malignancy with allogeneic SCT and respiratory virus infection (N=12, 6 with IA) versus other haematological malignancy without allogeneic SCT but with respiratory infection (N=15, 3 with IA).

Key:

AL/MDS: Acute leukaemia/myelodysplastic syndrome

Allo SCT: Allogeneic stem cell transplantation

Other: Includes Lymphoma, Myeloma, Chronic leukaemia, aplastic anaemia, myelofibrosis, unspecified leukaemia, Paroxysmal nocturnal hemoglobinuria, unspecified haematological malignancy, amyloidosis, unspecified autoimmune disorder, IGM paraproteinanaemia , and post transplant lymphoproliferative disorder.

GVHD: Graft versus host disease.

Figure 1a. The risk of IA associated with haematology patients not receiving, or prior, to allogeneic SCT (n=272) and the influence of genetic markers, respiratory virus infection and *Aspergillus* PCR result as determined by multi-logistic regression analysis. Percentage population (Pop) refers to the proportion of the total haematology cohort (n=322). PCR positivity was defined using a single PCR positive result as significant.

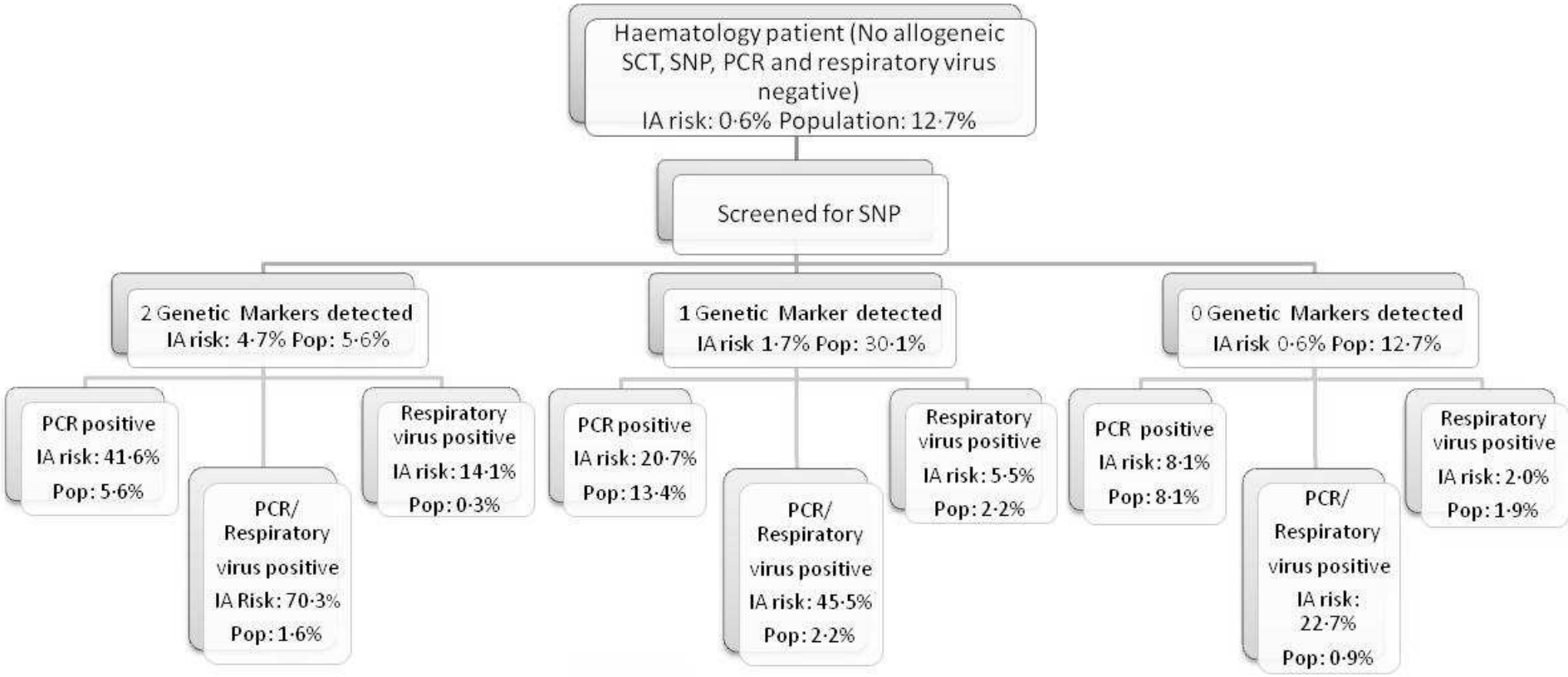
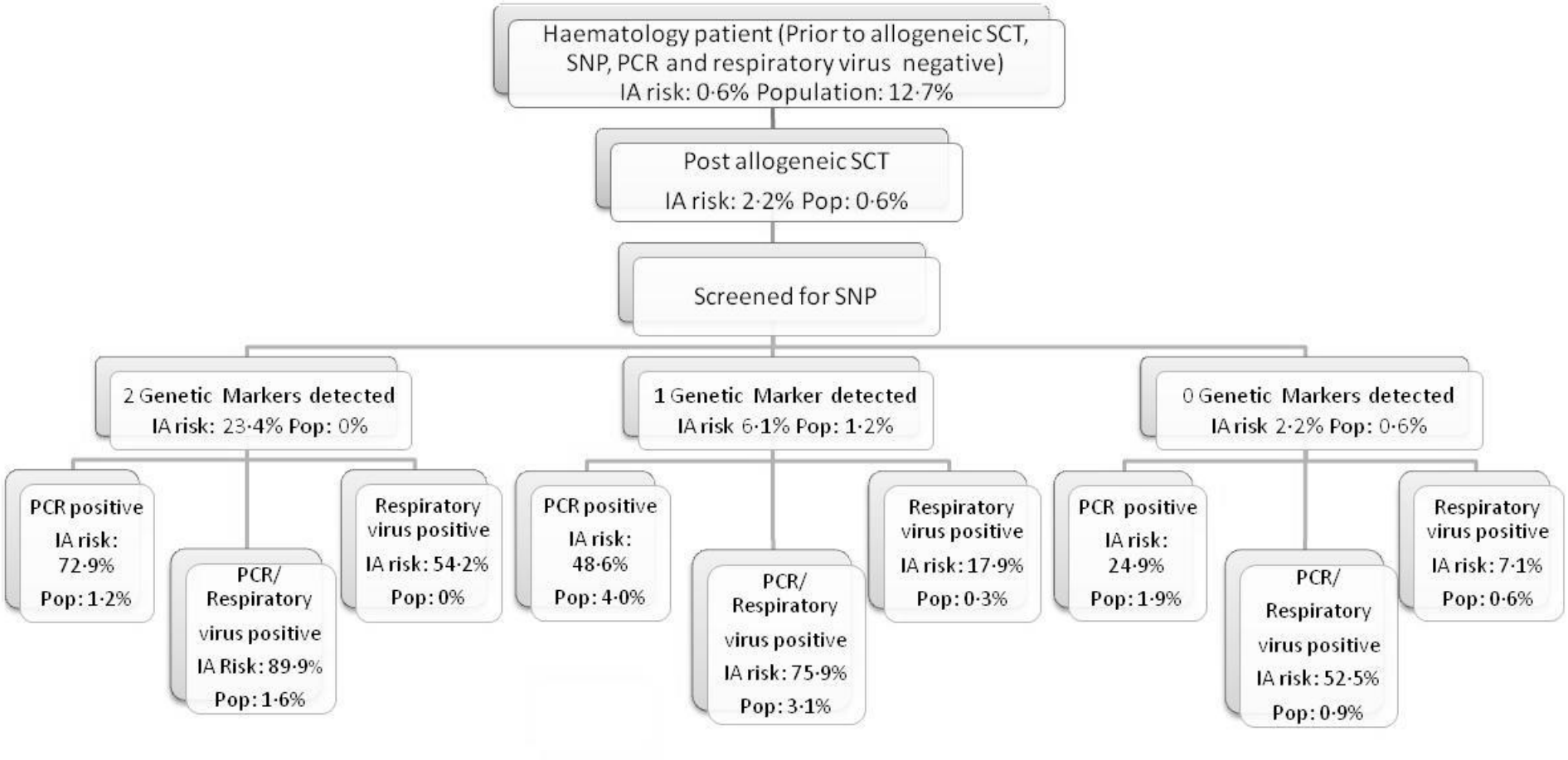
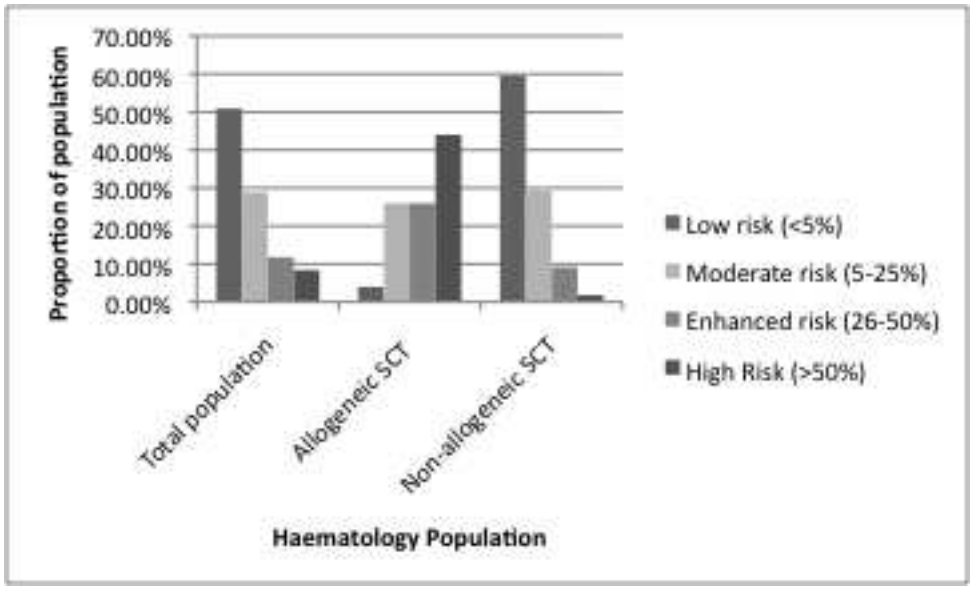


Figure 1b. The risk of IA associated with haematology patients receiving allogeneic SCT (n=50) and the influence of genetic markers, respiratory virus infection and *Aspergillus* PCR result as determined by multi-logistic regression analysis. Percentage population (Pop) refers to the proportion of the total haematology cohort (n=322). PCR positivity was defined using a single PCR positive result as significant.

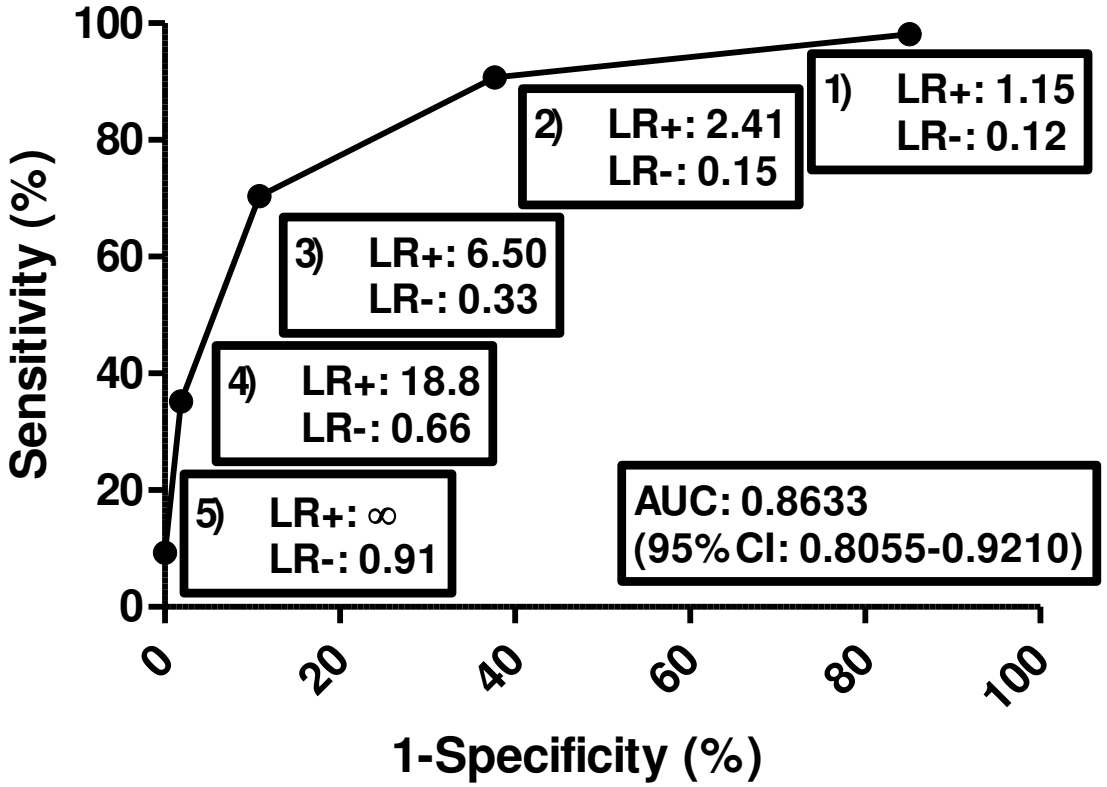


513 Figure 2. The distribution of risk of developing IA in a combined haematology population (n=322),
514 haematology patient posts allogeneic stem cell transplantation (n=50) or haematology without
515 allogeneic stem cell transplantation (n=272).

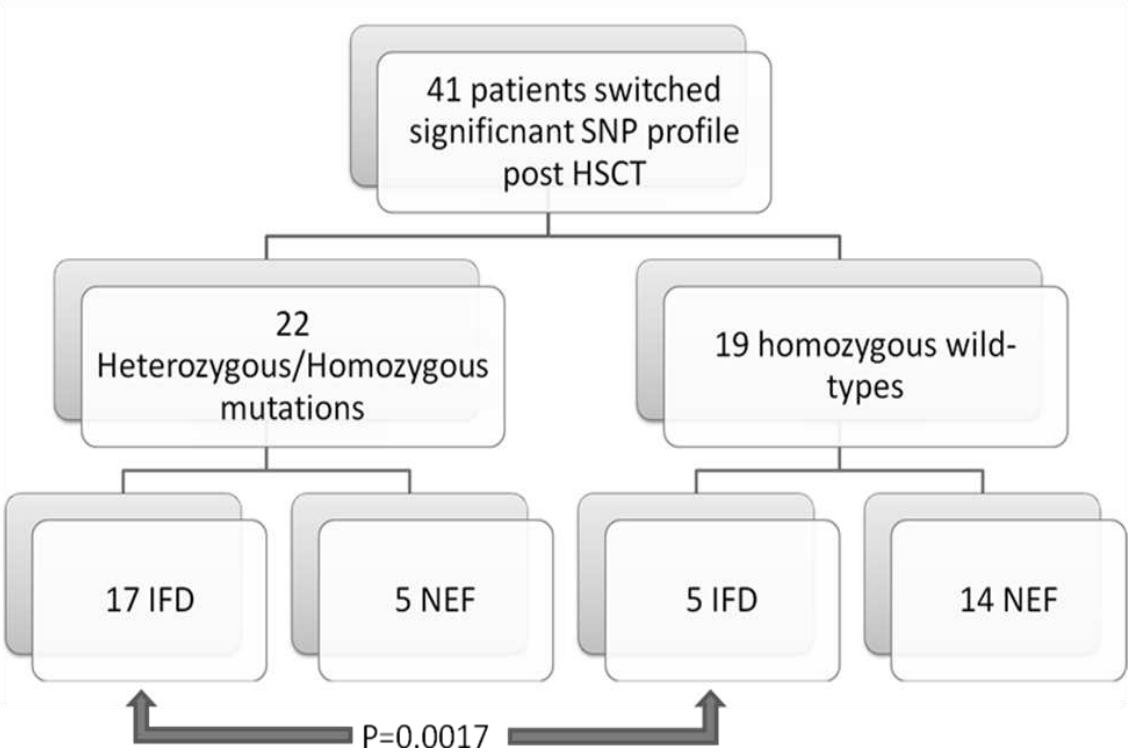


525

Figure 3. Receiver Operator Characteristic Curve for the final predictive model containing allogeneic stem cell transplantation, single nucleotide polymorphisms in Dectin 1 and DC-SIGN haplotype, respiratory virus infection and *Aspergillus* PCR positivity as risk factors associated with developed invasive aspergillosis . The thresholds for determining sensitivity/specificity are as follows: 1) Only one variable positive, 2) two variables positive, 3) three variables positive, 4) four variables positive and 5) all variables positive.



533 Figure 4. Switches in allele type at the single nucleotide polymorphisms (SNPs) significantly associated
534 with increased risk of developing IA



535