Genetic Markers and Predictive Factors Influencing the Aggressive Behavior of Cerebral Cavernous Malformation


ABSTRACT

Biological behavior of Cerebral Cavernous Malformation (CCM) is still controversial without clear-cut signature for biological mechanistic explanation of lesion aggressiveness. There is plenty evidence implicating dysregulated inflammatory and immune responses in vascular malformation pathogenesis, including CCM. In the present study, we evaluated the predictive capacity of the SNPs $VDR^{rs7975232}$, $VDR^{rs731236}$, $VDR^{rs11568820}$ as well as expanded the analysis of $PTPN2^{rs72872125}$ and $FCGR2A^{rs1801274}$ in relation to the aggressive behavior of CCM and its implications in biological processes. This was a single-site prospective observational cohort study with 103 patients enrolled, 42 had close follow-up visits for a period of 4 years, focused on 2 main aspects of the disease: (1) symptomatic event that composed both intracranial bleeding or epilepsy and (2) precocity of symptoms. We report a novel observation that the $PTPN2^{rs72872125}$ CT and the $VDR^{rs7975232}$ CC genotype were independently associated with an asymptomatic phenotype. Additionally, $PTPN2^{rs72872125}$ CC genotype and serum level of GM-CSF could predict a diagnostic association with symptomatic phenotype in CCM patients, while the $FCGR2A^{rs1801274}$ GG genotype could predict a symptomatic event during follow-up. The study also found a correlation between $VDR^{rs731236}$ AA and $VDR^{rs11568820}$ CC genotype to the time to first symptomatic event. In summary, this study provides valuable insights into the genetic markers that could potentially impact the development and advancement of CCM.

KEYWORDS: Cerebral Cavernous Malformation, Diagnostic Biomarker, Prognostic Biomarker, $FCGR2A$, $PTPN2$, $VDR$

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
INTRODUCTION

Cerebral cavernous malformation (CCM) are vascular malformations consisting of capillary-like channels with a single layer of endothelium and no intervening nervous tissue\textsuperscript{1,2}. CCM are among the most common vascular malformation of the central nervous system (CNS) affecting 0.5\%-1\% of the population, with magnetic resonance image (MRI) as the gold standard for the diagnosis and the lesions evaluation should include the hemosiderin-sensitive techniques as the susceptibility-weighted imaging (SWI)\textsuperscript{3,4}. Intracerebral hemorrhage (ICH) and seizure, are the main clinical manifestations, followed by neurological disability without evidence of bleeding. However, many patients remain asymptomatic throughout their life and it is not known what factors may predict aggressive manifestations in individual cases\textsuperscript{4,5}.

Usually, single cavernomas are detected in patients affected by sporadic forms and often associated with a developmental venous anomaly. Familial CCM is mostly associated with the occurrence of multiple lesions that might increase in number and size with aging\textsuperscript{6}. The familial form of the disease is linked to mutations on specific genes, (CCM1/KRIT1, CCM2/MGC4607 and CCM3/PDCD10), which encode distinct proteins involved in the endothelial cell junction function and in the interaction with cytoskeletal proteins. The most frequent and well-studied of the CCM genes is the CCM1/KRIT1, with mutational analyses showing a Hispanic-American ancestral haplotype\textsuperscript{7,8}. CCMs are characterized by an incomplete disease penetrance, being 80\% to the CCM1 form, near 100\% to the CCM2 and 60\% to the CCM3 form\textsuperscript{9,10}. Furthermore, the discovery of PIK3CA mutations in CCM lesions provides important new insights into the genetic factors that contribute to the variable expression and progression of the disease phenotype\textsuperscript{11}.

Biological behavior of CCM is still controversial without clear-cut signature for biological mechanistic explanation of lesion aggressiveness. There is plenty evidence implicating dysregulated inflammatory and immune responses in vascular malformation pathogenesis, including CCM. Several dysregulated pathways have been confirmed in the transcriptome of CCM lesions, which are out of balance having an impact on the permeability of the blood-brain barrier and the stability of endothelial tight junctions\textsuperscript{12}. Notably, inflammatory and immune
cells such as monocytes, macrophages, B and T cells, are present in human CCM lesions as well as in mouse models of CCM\textsuperscript{13–17}. Recently, our group have demonstrated an adaptive immune-cellular reaction to CCM within CD20+ and CD68+ in the pericavernous tissue of an aggressive pharmaco-resistant epilepsy patient\textsuperscript{18}. Moreover, the severity of the disease has been linked to genetic polymorphisms within inflammatory and immune response genes \textsuperscript{19}.

The significance of certain genetic variants in the development of CCM has been established, underscoring the need for further investigation into their specific roles. Tang et al 2017 showed allele T for SNP in \textit{TLR4} (rs10759930) gene is associated with increased CCM lesion number \textsuperscript{20}. Some studies have also identified genetic polymorphisms in genes related to oxidative stress with a significant impact on inter-individual variability in CCM disease onset and severity \textsuperscript{21}. Of particular interest, in a previous study, our group demonstrated the association of \textit{FCGR2A}\textsuperscript{rs1801274} GG and \textit{PTPN2}\textsuperscript{rs72872125} CT genotype with a symptomatic profile of CCM patients in a smaller cohort \textsuperscript{22}.

In the present study, we evaluated the predictive capacity of the SNPs \textit{VDR\textsuperscript{rs7975232}}, \textit{VDR\textsuperscript{rs731236}}, \textit{VDR\textsuperscript{rs11568820}} as well as expanded the analysis of \textit{PTPN2}\textsuperscript{rs72872125} and \textit{FCGR2A}\textsuperscript{rs1801274} in relation to the aggressive behavior of CCM and its implications in biological processes. We also looked for plasmatic inflammatory cytokines expressed in the patients, verifying a pattern of heterogeneity of plasmatic expression and any correlation with the genetic variation identified with different clinical phenotypes of CCM. Using a multi-step Bayesian approach, we thought to build a biomarker that could predict a diagnostic and a prognostic aggressive phenotype of CCM.

\section*{METHODS}

\textit{Study Design and Population}

This was a single-site prospective observational cohort study. Qualified researcher on human subjects conducted this study having been approved by the National Council for Ethics in Research (CAAE 69409617.9.0000.5258). Blood samples were collected between October 2017 to March 2023 and informed consent was obtained. Strobe criteria were used to report the findings of this study \textsuperscript{63}.
The epidemiological data were defined at each clinical assessment. Patients were stratified as multiform when they harbor multiples CCM on SWI or Gecho or as isolated form when they had a single lesion at the CNS. Both definitions were based on MRI study with SWI or Gecho and the absence of venous development anomalies in the multiple cases in proximity.

CCM clinical presentation were identified as symptomatic (hemorrhage or seizure) and asymptomatic. Intracerebral hemorrhage was defined based on the consensus of Cavernous Angiomas with Symptomatic Hemorrhage (CASH) \(^50\). The solely epileptic patients were defined as not having achieved previous hemorrhagic consensus and whom the MRI evaluation depicted no hint of hemosiderin beyond the usual smooth and regular deposit around a CCM lesion.

In this study, we focused in 2 mains aspects of the disease: (1) symptomatic event that composed both intracranial bleeding or epilepsy and (2) precocity of symptoms. We considered that these presentations may reflect an aggressive phenotype of the disease. We also designed six sub-group analysis in order to check for correlation - 60 years or older, female, familiar form, pure epileptic without signs of lesion bleeding, pure symptomatic bleeding and infratentorial lesion. Among the 103 consecutive cases, 42 had closely follow-up visit for a period of 4 years. The remaining patients either did not have a clinical follow-up visit or underwent surgical resection.

**Sample collection and processing**

Blood samples were collected with EDTA and processed by centrifugation at 720 \( \times g \) and 4 °C for 5 min to separate the plasma. The plasma supernatants were immediately stored at -80°C until further processing. The Genomic DNA (gDNA) extraction was performed using the PureLink Genomic DNA Mini Kit according to the manufacturer’s recommendations (ThermoFisher Scientific, Waltham, Massachusetts, USA). The quality of gDNA was determined by NanoDrop 2000 (ThermoFisher Scientific) followed by quantification using the Qubit dsDNA HS Assay Kit (ThermoFisher Scientific) and Qubit Fluorometer3.0 (Thermo Fisher Scientific).

**Plasma isolation and Inflammatory Modulators Assessment**
Proinflammatory cytokine and chemokine levels in plasma were measured using a multiparametric immunoassay based on XMap-labeled magnetic microbeads (Luminex Corp – Austin, TX, USA). A human ProcartaPlex™ Panel (Invitrogen – Waltham, MA, USA) was used to analyze a set of 18 cytokines and chemokines (IL-1β, IL-2, IL-4, IL-5, IL-6, IL-9, IL-10, IL-12p70, IL-13, IL-17A, IL-18, IL-21, IL-22, IL-23, IL-27, GMCSF and INF-γ). The samples were measured according to the manufacturer’s instructions and simultaneously to avoid potential batch, as described previously, using a BioPlex MAGPIX system (Biorad – Hercules, CA, USA)\textsuperscript{64,65}. Cyto/chemokine expression was measured in duplicate, and the levels of inflammatory modulators were analyzed using Xponent v. 3.0 software (Luminexcorp) and expressed in pg/ml.

**Analysis of polymorphism of PTPN2\textsuperscript{rs72872125}, VDR\textsuperscript{rs7975232, VDR\textsuperscript{rs731236}, VDR\textsuperscript{rs11568820} and FCGR2A\textsuperscript{rs1801274} genes**

The SNPs investigated in this study are detailed in Table 1. The SNPs were genotyped by allelic discrimination performed in QuantStudio™ 3 Real-Time PCR System (ThermoFisher) using TaqMan SNP genotyping assays (ThermoFisher). The characteristics of the PTPN2\textsuperscript{rs72872125}, VDR\textsuperscript{rs7975232, VDR\textsuperscript{rs731236}, VDR\textsuperscript{rs11568820} and FCGR2A\textsuperscript{rs1801274} genes were obtained from the SNP bank of the National Center of Biotechnology Information - NCBI (http://www.ncbi.nlm.nih.gov/). The probes for each SNP were produced by Applied Biosystems™ rs72872125 (C\textsubscript{__}98019281_10), rs7975232 (C\textsubscript{__}28977635_10), rs731236 (C\textsubscript{__}24040008_10), rs11568820 (C\textsubscript{__}2880808_10) and rs1801274 (9077561_20). As the assay was designed and standardized to work with the same thermal cycles, a single protocol was applied for both genes. Briefly, PCR was performed with a 25 uL reaction mixture containing 10 ng DNA, TaqMan® Universal PCR Master Mix (1X), Probe TaqMan® Gene Expression Assay (1X), and DNAse free water for the final volume. The Real Time PCR conditions were: initially 60°C for 30 seconds and then 95°C for 10 min, and subsequently 40 cycles of amplification (95 °C for 15 seconds and 60°C for 1 min), and then 60°C for 30 seconds. The five selected SNPs were amplified on separate plates.
**In silico Structural Prediction of FCGR2A<sup>rs1801274</sup> Protein and Phylogenetic Analysis**

Since the FCGR2A<sup>rs1801274</sup> SNP is the only with amino acid change, we performed a structural analysis using bioinformatic tools. FCGR2A<sup>rs1801274</sup> transcript structure was predict by directly modifications of the wild-type FCGR2A using FASTA sequence obtained on NCBI database<sup>66</sup>. The new FASTA file was submitted to the Chimera software (developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311) for overall structural analysis<sup>67</sup>. The evolutionary history was inferred using the MEGA-X software using the maximum likelihood method and Jones-Taylor-Thornton (JTT) matrix-based model. We first gathered the FCGR2A wild-type sequence from Homo sapiens and 19 other mammals to study the evolutionary conservation of this protein. Structural alignment was performed using the muscle algorithm that allows multiple sequencing alignment with high accuracy and high throughput<sup>68</sup>.

**Statistical Analysis**

Statistical analyses were performed with STATA 13 (StataCorp LP, TX, USA) and GraphPad Prism 9.0.0 MacOS (GraphPad Software, San Diego, California USA, [www.graphpad.com](http://www.graphpad.com)). Categorical variables were expressed as n (%) and continuous data were given as mean and standard deviation (SD). A 2-tailed 2-sample t test or Mann-Whitney test was applied for continuous variables while the association between two categorical variables was measured by Pearson Chi-square (χ²) test when appropriate.

Allelic and genotypic frequencies were calculated for patient and control subjects via direct gene counting. Genotypic distributions in Hardy–Weinberg equilibrium (HWE) were evaluated by two-tailed χ²-test. Linkage disequilibrium (LD) were analyzed using: [Linkage Disequilibrium Calculator](https://grch37.ensembl.org/Homo_sapiens/Tools/LD), [LDlink](https://ldlink.nih.gov/) and Heatmap was plotted by [https://www.bioinformatics.com.cn/en](https://www.bioinformatics.com.cn/en). The degree of LD between SNPs is represented by R² and D´.
Genetic polymorphisms were evaluated through a univariate and multivariable binary logistic regression adjusting for sex, age and familiar form when adequate in matter of phenotype (symptomatic and asymptomatic). Plasma cytokines were also evaluated between genotypes and phenotypes (symptomatic and asymptomatic). Values over 2 times de standard deviation were excluded from analysis due to possible bias results.

Genetic polymorphism and plasma cytokines verified as statistically different between groups (both diagnostic and prognostic) were combined through a canonical discriminant function analysis in order to build a biomarker-model of symptomatic profile. All the possible combinations showing significant association in symptomatic individuals were build. Receiver Operating Characteristic (ROC) curves were generated along with a computed area under the curve (AUC) for each combination individually and ROC curves were compared to identify if any of them were statically superior then the others. The best model-biomarker to differentiate symptomatic and asymptomatic patients in both diagnostic and prognostic scenario was selected according to the Akaike Information Criteria (AIC), representing the best fit parsimonious model to the data with the fewest number of predictors. The optimal cutoff point was generated from ROC curves utilizing the Youden index method. Mann-Whitney test was then used to verify the difference in values generated using the modeled equation among symptomatic and asymptomatic patients and a logistic regression was used to control for age and sex. A p value <0.05 was considered significant.

To balance baseline covariates between patients with and without symptoms during follow-up we used a propensity score matching (PSM) strategy. The strategy involved 1:1 pairing and nearest-neighbor methods. After PSM, the distribution of gender, familial form, and age was balanced between the groups and the biomarker analysis was remade.

Finally, A Kaplan-Meyer survivor analysis was carried out among patients who presented different polymorphism combinations. Hazard ratios were calculated with 95% confidence intervals through a log-rank test. Failure events were determined according to if there had been symptomatic presentation at some point in the patient’s lifetime.
Results

Demographic and CCM lesions characteristics
Out of 103 CCM patients enrolled in the study, 44 are multifocal/familial CCMs, 70 patients presented with symptomatic phenotype, which 48 presented a symptomatic hemorrhage secondary to their CCM, and 22 presented with seizures and no history or signs of CCM bleeding on MRI. The median age of enrolment was 45.6, within a mean age of 41.6 in the symptomatic group and 53.7 in epileptic sub-group. There were no significant differences in terms of sex, age and form (sporadic vs familiar/multifocal) between the outcome groups (Table 2).

Association of PTPN2<sup>rs72872125</sup> and VDR<sup>rs7975232</sup> to a Symptomatic Phenotype and Plasmatic Cytokine Levels

The distribution of the genotypic frequencies of the PTPN2<sup>rs72872125</sup>, VDR<sup>rs7975232</sup>, VDR<sup>rs731236</sup>, VDR<sup>rs11568820</sup> and FCGR2A<sup>rs1801274</sup> in symptomatic and asymptomatic patients are shown in Table 3. All genotypic distributions were in Hardy-Weinberg equilibrium. Haplotype analysis indicated linkage disequilibrium between VDR<sup>rs7975232</sup> and VDR<sup>rs731236</sup> (D'>0.99), corroborating what has already been found in other populations (Figure 1A and Supplementary material). While the other combinations segregate independently VDR<sup>rs11568820</sup> and VDR<sup>rs731236</sup> (D'>0.04), VDR<sup>rs11568820</sup> and VDR<sup>rs7975232</sup> (D'>0.317) (Figure 1A and Supplementary material). A higher frequency of the PTPN2<sup>rs72872125</sup> CT genotype (OR 0.34, 95% CI 0.11-0.99, p = 0.04) and the VDR<sup>rs7975232</sup> CC genotype (OR 0.06, 95% CI 0.006-0.612, p = 0.017) was observed in asymptomatic phenotype in the age, familiar and female adjusted multivariable analysis, when compared with symptomatic patients (Table 3). The ROC analysis (Figure 1B) revealed that patients with PTPN2<sup>rs72872125</sup> CT genotype had a modest Area Under de Curve (AUC 0.420, SE 0.05 CI 95% 0.334-0.506) and VDR<sup>rs7975232</sup> CC individuals (Figure 1C) had a pour accuracy (AUC 0.439, SE 0.03 CI 95% 0.370-0.508).

We also hypothesized that these genetic variants might influence the plasmatic inflammation signature of the CCM patients. We found that patients who harbor the PTPN2<sup>rs72872125</sup> CT genotype showed a higher plasma level of IL-10 (p = 0.0146) (Figure 1D) and low levels of IL-18 (p = 0.0450) (Figure 1E) and...
IFN-γ \((p = 0.0310)\) (Figure 1F), while individuals with \(VDR^{rs7975232}\) at least one C allele had low plasma levels of IL-27 \((p = 0.0055)\) and IL-23 \((p = 0.0034)\) (Figure 1G-H). For other SNPs no significant difference in the levels of immune markers was observed. We did not observe any significant differences in the levels of these plasma molecules in relation to patient sex and form (sporadic/solitary or familial/multifocal).

**Performance of Diagnostic Biomarker**

In order to build a diagnostic biomarker of CCM activity that could improve our previously published weighted biomarker formula, we tested new combinations of genetic and cytokines factors. First, the previously published diagnostic weighted biomarker was again confirmed to distinguish symptomatic and asymptomatic patients with 14.2% sensitivity and 97.4% specificity (AUC 0.641 SE 0.06 CI 95% 0.512 – 0.770, \(p = 0.03\)), worse than the range reported previously. A similar canonical discriminant analysis approach was then implemented to determine if a weighted combination of different SNPs and cytokines could improve the diagnostic association with symptomatic phenotype. The best weighted-biomarker included the \(PTPN2^{rs72872125}\) CC genotype and GMCSF plasma levels as formulated \(-0.89 \times (\text{GMCSF}) + 0.41 \times (PTPN2^{rs72872125} \text{CC})\) (AUC 0.663, SE 0.06 CI 95% 0.534 – 0.792, \(p = 0.01\)) with a specificity and sensitivity 85.7% and 41.3%, respectively (Figure 2A). This formula outperformed all the other possible formula (Table 1 Supplement) that had a statistically significant value. The median weighted combination value was 2.65-times increased \((p = 0.01)\) in symptomatic patients (median estimated value -0.36) compared with asymptomatic individuals (median estimated value -0.97) (Figure 2B). The formula performed fairly in the sub-group analysis of hemorrhagic patients (AUC 0.661, SE 0.07, CI 95% 0.517 – 0.806) but had a good accuracy in familiar form of the disease (AUC 0.774, SE 0.09, CI 95% 0.595 – 0.953, \(p = 0.01\)) and patients with infratentorial lesions (AUC 0.804, SE 0.08, CI 95% 0.642 – 0.966, \(p = 0.001\)) (Figure 2 C-D). Female symptomatic patients, pure epileptic patients and elderly patients were not statically significant.
Validation Cohort with Symptomatic Event during Follow-up: Building a Prognostic Biomarker

We then tested if the newly weighted-diagnostic biomarker could act as a prognostic biomarker in a sub-group analysis of 42 patients that were prospectively followed-up after initial blood collection. 4 patients experienced a symptomatic event (1 lesion growth, 2 symptomatic hemorrhage and 1 newly onset epilepsy without signs of bleeding) while the others 38 stayed asymptomatic during the next 4 years. The weighted-biomarker had a fair accuracy (AUC 0.644, SE 0.078, CI 95% 0.491-0.976) in distinguishing these patients (Figure 3A). In order to verify if other combination of genetic variants and cytokines could act as a prognostic biomarker, we first tested the 5 SNPs in the sub-group of 42 patients that were closely followed-up. We found an independent statistically significant association of the FCGR2A rs1801274 GG genotype (OR 16.10, 95% CI 1.32 – 195.52, p = 0.029) in the age and familiar adjusted multivariable analysis. This variant had an excellent accuracy (AUC 0.796, SE 0.12, CI 95% 0.631-0.897) in distinguishing these individuals (Figure 3B). We further tested the same gene in a propensity matched subgroup. The presence of the FCGR2A rs1801274 GG genotype could predict a symptomatic event with high accuracy (AUC 0.875, SE 0.12, CI 95% 0.473-0.996) (Figure 3C). We also tested if this mutation would maintain its accuracy power when comparing symptomatic patients during follow-up and individuals who have never experienced any symptoms. We also found that FCGR2A rs1801274 GG genotype could predict a symptomatic event with high accuracy (AUC 0.829, SE 0.14, CI 95% 0.519 - 0.956) (Figure 3D).

In Silico Structural Analysis of Transcript FCGR2A rs1801274

In silico structural analysis prediction reveals that the FCGR2A rs1801274 leads to a substitution of a histidine for an arginine at position 131 (Figure 3E). The structural alignment predicts that the variation c.500A>G (p.His131Arg) occurred in domain that appears to be conserved mainly in higher primates, within 30% of the species sharing this site (Figure 3F).

VDR variants relates to Precocity of symptoms
We tested if any of the VDR genetic variants could relate to a precocity of symptoms in the CCM cohort. Patients with the VDR rs731236 AA genotype tend to present their symptoms late in their life compared to those who did not have (HR 0.52, SE 0.14, CI 95% 0.299-0.909 p = 0.020) (Figure 4A). Also, individuals that harbor the VDR rs11568820 CC genotype tend to have symptoms earlier (HR 1.58, SE 0.46, CI 95% 1.092-2.282 p = 0.016 (Figure 3B).

Discussion
Neuroinflammation is increasingly a focus of research in symptomatic events in CCM patients. In the present study, we evaluated a large and phenotypically well characterized CCM cohort and provided some key information about the genetic influence on the behavior of this unique neurovascular disease. Here, we have found an individual higher frequency of PTPN2 rs72872125 CT and the VDR rs7975232 CC genotypes in asymptomatic patients, associated with changes in cytokine levels, suggesting a possible protective role. Our main new observation is that the combination of a balanced formula using the PTPN2 rs72872125 CC genotype and serum level of GM-CSF could predict a diagnostic association with symptomatic phenotype in CCM patients, while the FCGR2A rs1801274 GG genotype showed the best accuracy in predicting a symptomatic event in the next years, possibly functioning as a prognostic genetic biomarker. In previous studies, we reported the association of this same variant with an aggressive phenotype of cerebral cavernous malformation but in a smaller sample. Furthermore, another exciting discovery is that individuals with the VDR rs731236 AA and VDR rs11568820 CC genotypes may experience an earlier onset of symptoms in the CCM cohort.

To date, there have been few published studies on SNPs in the PTPN2 gene, and among them, a weak association with certain diseases or different outcomes was found. However, it is important to highlight that in combination with other variants they have been shown to possibly increase the susceptibility of chronic inflammatory disorders, including rheumatoid arthritis, type 1 diabetes and celiac disease. Furthermore, there have been research findings indicating the presence of epistasis between PTPN2 and VDR gene. Interestingly, patients who inherit the PTPN2 rs72872125 CT genotype have high
levels of IL-10 when compared with CC genotype (Figure 1). IL-10 is an anti-inflammatory molecule, with a well-established function in restraining and regulating both acute and chronic inflammatory processes. Lyne et al. (2019) demonstrated that IL-10 is an important molecule present in diagnostic CASH biomarker and the levels were decreased in cases of individuals who had experienced symptomatic CCM hemorrhage in the prior year. Furthermore, the CT genotype was associated with low levels of IL-18 and INF-γ (Figure 1), which are increased in patients with epilepsy and CCM hemorrhagic phenotype. The Canonical values derived from the best weighted formula combination with GM-CSF were 2,65X higher in patients who suffered a subsequent symptomatic event (Figure 2). Like this, the a weighted combination of PTPN2rs72872125 SNP and GM-CSF levels is a potential diagnostic genetic biomarker for the symptomatic phenotype in CCM patients.

We had also provided more evidence that the FCGR2A rs1801274 GG genotype could act as a prognostic biomarker of the disease. The FCGR2A gene is located on chromosome 1q23 and consists of seven exons, encoding a member of a family of Fcγ receptors for immunoglobulin G (IgG). Through it is expression in immune system cells such as macrophages, dendritic cells and neutrophils, it is possible to link cellular and humoral immunity. The FCGR2A rs1801274 variant leads to a substitution of a histidine (A allele) for an arginine (G allele) at position 131, also known as H131R. This polymorphism is capable to increase the binding affinity of FCGR2A to IgG, resulting in activation of the FCGR2A signaling pathway and upregulation of IgG2-dependent phagocytosis. The GG genotype has been associated with several autoimmune diseases, such as Systemic Lupus Erythematosus, Type 1 Diabetes Mellitus, Crohn's Disease and others, based on its relationship with the release and stimulation of responsive inflammatory processes. Protein structure prediction using bioinformatic tools is an important approach to understand the FCGR2A variants. In this study we also performed an in-silico analysis that demonstrated that the H131R position is strongly conserved among higher primates, which may be related to its clinical importance.

Lyne et al. 2019 showed an upregulation of FCGR2B gene in CASH transcriptomic, shedding light on the importance of Fcγ receptors in CCM disease...
FcγRII receptors mediates the C-reactive protein (CRP)-induced changes in endothelial function and inflammatory response. A higher binding avidity of CRP to FcγRIIa on immune cells was identified for allotype FcγRIIa-R131 compared to other genotypes. The variant homozygous genotype (GG) was able to increase the expression of ICAM-1 and E-selectin in HUVEC (Human Umbilical Vein Endothelial Cells) and the levels of tPA, MCP-1, and IL-6 secreted. In addition, the G allele resulted in a significant defect in endothelium-dependent vasodilatation and reduced NO activity during endothelial cell stimulation in patients with hypercholesterolaemia, corroborating with what has already been found in models of CCM. This data could potentially impact our comprehension of the pathophysiology of CCM disease as well as consequences with regard to the interpretation of prognostic biomarker for follow-up of patients.

Growing evidence suggests that vitamin D signaling role in cavernous malformation behavior. Peripheral plasma vitamin D has been shown to reflect the severity of CCM disease. Indeed, cholecalciferol (vitamin D3), was shown to decrease CCM lesion burden in a murine model of CCM, and to inhibit ROCK activity, known to affect CCM development. The effects of vitamin D on the immune system are accomplished by binding to the nuclear Vitamin D Receptor (VDR). Some single nucleotide polymorphisms (SNPs) in genes involved in vitamin D singling were reported to have association with vitamin D deficiency. The VDR\textsuperscript{rs731236} also known as TaqI, is a synonymous variant, while VDR\textsuperscript{rs7975232} (ApaI) and VDR\textsuperscript{rs11568820} are located in the 3’ and 5’-untranslated regions of the gene respectively. These SNPs do not alter the amino acid sequence in the VDR protein, but they can exert influence on mRNA stability and gene transcription. Especially in relation to VDR\textsuperscript{rs7975232}, the literature does not make clear its role in relation to vitamin D levels, on the other hand we showed that patients who carried the CC genotype have decreased levels of proinflammatory cytokines (Figure 1). Corroborating with a possible protective role of the VDR\textsuperscript{rs7975232}CC genotype, Jiang et al (2015) showed that A allele slightly increased the risk of temporal lobe epilepsy in children.

Age at the first symptom is also a fundamental information during the counselling of CCM patients. Distinguishing individuals with higher chances of precocity of symptoms that could influence their productive life is of great
importance to the CCM community. Our study provided the correlation between $VDR^{rs731236}$ AA and $VDR^{rs11568820}$ CC genotypes to the time to first symptomatic event (Figure 4). However, we could not demonstrate a straight correlation between the $VDR^{rs11568820}$ CC and $VDR^{rs731236}$ AA genotype presence and the blood concentration of any of the analyzed cytokines. The relationship between VDR polymorphisms and susceptibility to autoimmunity diseases have been conducted in different settings, while the results obtained so far are conflicting. We hypothesized that VDR polymorphism may affect the way proinflammatory cells react to the CCM in an autoimmunity fashion leading to molecular level disruption. To the best of our knowledge, this is the first study to investigate the associations between VDR polymorphisms and cerebral cavernous malformations.

Previously reports demonstrated that CCM1/KRIT1 regulates vascular permeability through interaction with CCM2/MGC4607 to stabilize endothelial cell–cell junctions, and together they suppress RhoA activity and, thus, activation of the RhoA effector ROCK. Mutation on these genes lead to augment of actin stress fiber formation and increased permeability of CCMs. Moreli et al (2007) have shown that BXL-628, a VDR agonist, prevented RhoA activation and inhibits RhoA/Rho Kinase signaling in rat and human bladder. Since vitamin D is a natural agonist of VDR and animal models have suggested that fasudil and high doses of atorvastatin promotes significant inhibition of CCM lesional development and hemorrhage through the RhoA/Rho Kinase rational, it is reasonable to suppose that the prognostic ability of low levels of serum vitamin D is related to these VD/VDR interactions and its genetic variants.

Our study pioneers a novel classification denoted as the "aggressive behavior" paradigm within cerebral cavernous malformations (CCM), incorporating both hemorrhage and epilepsy as integral components of its symptomatic criteria. This redefined paradigm marks a transformative departure, acknowledging the critical role of epilepsy alongside hemorrhage in defining the clinical spectrum of CCM. Unlike a diagnostic biomarker, which detects the presence of a specific medical condition, a prognostic biomarker assesses the likelihood of progression or stability of medical conditions. These biomarkers contribute to a more precise and expedited diagnosis, as well as improved patient...
follow-up. In summary, our study provides new pieces of evidence for possible genetic biomarkers that may influence the behavior of cerebral cavernous malformations. We report a novel observation that the $PTPN2^{rs72872125}$ CT and the $VDR^{rs7975232}$ CC genotypes were independently associated with an asymptomatic phenotype. Additionally, $PTPN2^{rs72872125}$ CC genotype and serum level of GM-CSF could predict a diagnostic association with symptomatic phenotype in CCM patients, while the $FCGR2A^{rs1801274}$ GG genotype could predict a symptomatic event during follow-up. The study also found a correlation between $VDR^{rs731236}$ AA and $VDR^{rs11568820}$ CC genotype to the time to first symptomatic event. Overall, this study provided valuable information on the genetic factors that may influence the development and progression of CCM. Moreover, beyond the conventional genetic biomarker purview, our investigation postulates a novel etiological framework. We posit an inflammatory cascade precipitating a self-non-self-interaction, instigating an autoimmune response against CCM. This theoretical construct delineates the changes of CCM from a benign anatomical variant—traditionally construed as a common disease—into an aggressive lesion, now occupying the realm of a rare affliction—the aggressive behavior of CCM. This theoretical construct challenges extant paradigms, accentuating the potential role of immune-mediated mechanisms in the clinical trajectory of CCM, thereby offering prospects for targeted immunomodulatory interventions. In essence, this study augments our comprehension of genetic determinants influencing the advancement of CCM while advancing a transformative hypothesis concerning the disease's underlying inflammatory dynamics. These insights herald a paradigm shift in conceptualizing the clinical spectrum of CCM, fostering an enriched landscape for future investigative pursuits and potential therapeutic interventions aimed at modulating the intricate inflammatory cascades underpinning the aggressive behavior of CCM.

Supplemental material

Table Supplementary 1 (S1)

Correspondence
Fabrícia Lima Fontes-Dantas: ORCID https://orcid.org/0000-0002-5201-0927, e-mail fabricia.fontesdantas@uerj.br;
Jorge Marcondes de Souza: ORCID https://orcid.org/0000-0003-2412-8239, e-mail jormarcondes@gmail.com.

Affiliations

Acknowledgements
The authors thank the Cavernoma Alliance Brazil Research Institute – Aliança Cavernoma Brasil for the logistic assistance.

Sources of Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Brazilian National Council for Scientific and Technological Development (CNPq Number 440779/2016-2), Senator Romario Faria parliamentary amendment (no. 37990007 EIND), financial support of Coordination for the Improvement of Higher Education Personnel (CAPES Number 88887.130752/2016-00), FAPERJ E-26/210.657/2021, E-26/210.273/2018 and E-26/201.040/2021 and Chamada Pública MCTI/FINEP/CT-INFRA-PROINFRA 02/2014 – Equipamentos Multiusuários – Ref. nº 0097/2016. Thanks to Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) and Casa Hunter for the financial help to this Project.
Disclosures

The authors declare no competing interests

References


26. Santin I, Moore F, Colli ML, Gurzov EN, Marselli L, Marchetti P, Eizirik DL. PTPN2, a Candidate Gene for Type 1 Diabetes, Modulates Pancreatic β-Cell Apoptosis via Regulation of the BH3-Only Protein Bim. *Diabetes* [Internet]. 2011;60:3279–3288. Available from: https://diabetesjournals.org/diabetes/article/60/12/3279/14490/PTPN2-a-Candidate-Gene-for-Type-1-Diabetes


40. Shrestha S, Wiener HW, Olson AK, Edberg JC, Bowles NE, Patel H,


Legend Figures

Figure 1: Association of PTPN2 and VDR to a Symptomatic Phenotype, Linkage Disequilibrium Map and Plasmatic Cytokine Expression. (A) Haplotype map showing the 3 SNPs analyzed and the range of R² value, suggesting that VDR<sup>rs7975232</sup> and VDR<sup>rs731236</sup> are in linkage disequilibrium (D'>0.99), while the
other combinations segregate independently $\text{VDR}^{rs11568820}$ and $\text{VDR}^{rs731236}$

(D’>0.4), $\text{VDR}^{rs11568820}$ and $\text{VDR}^{rs7975232}$ (D’>0.317). (B) ROC analysis
evidencing the performance of $\text{PTPN2}^{rs72872125}$ CT genotype (AUC 0.420) and (C)
$\text{VDR}^{rs7975232}$ CC (AUC 0.439) to distinguish symptomatic patients. (D) Plasmatic
cytokine expression between $\text{PTPN2}^{rs72872125}$ groups showing the $\text{PTPN2}^{rs72872125}$
CT genotype have higher plasma level of IL-10 ($p = 0.0146$), (E) low levels of IL-18 ($p = 0.0450$) and (F) low levels of IFN-γ ($p = 0.0310$) (Figure 1F), while
individuals with at least one $\text{VDR}^{rs7975232}$ C allele had (G) low plasma levels of IL-27 ($p = 0.0055$) and (H) low plasma levels of IL-23 ($p = 0.0034$).

Figure 2: Performance of Diagnostic Biomarker. (A) ROC curve of the best
weighted-biomarker included $\text{PTPN2}^{rs72872125}$ CC genotype and GMCSF plasma
levels as formulated $-0.89*(\text{GMCSF})+0.41*(\text{PTPN2}^{rs72872125}\text{CC})$ (AUC 0.663, SE
0.06 CI 95% 0.534 – 0.792, $p = 0.01$) with a specificity and sensitivity 85.7% and
41.3%, respectively. (B) The median weighted combination value was 2.65-times
increased ($p =0.01$) in symptomatic patients (median estimated value -0.36)
compared with asymptomatic individuals (median estimated value -0.97). Sub-
group analysis demonstrating that the formula performed fairly in the sub-group
analysis of (C) symptomatic bleeding patients (AUC 0.661, SE 0.07, CI 95%
0.517 – 0.806) but had a good accuracy in (D) familiar form of the disease (AUC
0.774, SE 0.09, CI 95% 0.595 – 0.953, $p = 0.01$) and (E) patients with
infratentorial lesions (AUC 0.804, SE 0.08, CI 95% 0.642 – 0.966, $p = 0.001$).

Figure 3: Performance of Prognostic Biomarker and In Silico Structural Analysis
of Transcript $\text{FCGR2A}^{rs180127}$. (A) The weighted-biomarker had a fair accuracy
(AUC 0.644, SE 0.078, CI 95% 0.491-0.976) as prognostic biomarker while
$\text{FCGR2A}^{rs1801274}$ GG (B) genotype had an excellent accuracy (AUC 0.796, SE
0.12, CI 95% 0.631-0.897) in distinguishing these patients. (C) ROC curve
evidencing the performance of the $\text{FCGR2A}^{rs1801274}$ GG genotype after PSM
analysis (AUC 0.875) and (D) in a sub-group of pure asymptomatic patients (AUC
0.829). (E) In silico structural analysis prediction reveals that the $\text{FCGR2A}^{rs1801274}$
leads to a substitution of a histidine for an arginine at position 131 leading to a
predicted conformational structure slightly different than the wild-type. (F)
Structural alignment predicts that the mutation c.500A>G (p.His131Arg) occurred in domain conserved in higher primates, within 30% of the species sharing this site.

Figure 4: Kaplan-Meier curve demonstrating that VDR variants are related to Precocity of symptoms. (A) $VDR^{*731236}$ AA genotype tend to have symptoms later in life compared to those who did not have (HR 0.52, $p = 0.020$). (B) Individuals that harbor the $VDR^{*11568820}$ CC genotype tend to have symptoms earlier (HR 1.58, $p = 0.016$).
Table 1: Studied SNPs and their basic characteristics and female sex adjusted multivariable analysis.

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP ID</th>
<th>Alternative Name</th>
<th>Chr/Location</th>
<th>Nucleotide Change</th>
<th>Aminoacid Change</th>
<th>MAF</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTPN2</td>
<td>rs72872125</td>
<td>18/intron</td>
<td>C&gt;T</td>
<td></td>
<td></td>
<td>0.11</td>
</tr>
<tr>
<td>VDR</td>
<td>rs7975232</td>
<td>Apa1</td>
<td>12/intron</td>
<td>C&gt;A</td>
<td></td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>rs731236</td>
<td>Taq1</td>
<td>12/exon</td>
<td>A&gt;G</td>
<td>Ile &gt; Ile</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>rs11568820</td>
<td></td>
<td>12/intron</td>
<td>C&gt;T</td>
<td></td>
<td>0.28</td>
</tr>
<tr>
<td>FCGR2A</td>
<td>rs1801274</td>
<td>1/exon</td>
<td>A&gt;G</td>
<td></td>
<td>His &gt; Arg</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Chr - chromosome; MAF - minimal allele frequency
Table 2: Summary of the Demographic and Clinical Characteristics

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>N (%)</th>
<th>Multifocal/Familiar</th>
<th>Age (years), mean</th>
<th>Sex: Female, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptomatic</td>
<td>33 (32%)</td>
<td>13</td>
<td>46.3</td>
<td>22</td>
</tr>
<tr>
<td>Symptomatic</td>
<td>70 (67%)</td>
<td>31</td>
<td>45.3</td>
<td>43</td>
</tr>
<tr>
<td>Hemorrhage</td>
<td>48 (46%)</td>
<td>20</td>
<td>41.6</td>
<td>33</td>
</tr>
<tr>
<td>Epilepsy</td>
<td>22 (21%)</td>
<td>11</td>
<td>53.7</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>103</td>
<td>44</td>
<td>45.6</td>
<td>65</td>
</tr>
</tbody>
</table>

N - absolute number
Table 3: Association between SNPs and Different Phenotypes in the age,
familiar and female sex adjusted multivariable analysis

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP ID</th>
<th>Allele</th>
<th>Symptomatic</th>
<th>Asymptomatic</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTPN2</td>
<td>rs72872125</td>
<td>CC</td>
<td>59</td>
<td>22</td>
<td>2.83 (1.04-7.67)</td>
<td><strong>0.04</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CT</td>
<td>8</td>
<td>9</td>
<td>0.34 (0.11-0.99)</td>
<td></td>
</tr>
<tr>
<td>VDR</td>
<td>rs7975232</td>
<td>CC</td>
<td>1</td>
<td>5</td>
<td>0.06 (0.006-0.6120)</td>
<td><strong>0.017</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AC</td>
<td>20</td>
<td>5</td>
<td>2.33 (0.782-6.949)</td>
<td>0.128</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AA</td>
<td>47</td>
<td>22</td>
<td>1.00 (0.403-2.500)</td>
<td>0.993</td>
</tr>
<tr>
<td></td>
<td>rs731236</td>
<td>GG</td>
<td>12</td>
<td>3</td>
<td>2.16 (0.525-8.942)</td>
<td>0.779</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AG</td>
<td>21</td>
<td>16</td>
<td>1.06 (0.408-2.789)</td>
<td>0.893</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AA</td>
<td>30</td>
<td>14</td>
<td>0.89 (0.349-2.220)</td>
<td>0.808</td>
</tr>
<tr>
<td></td>
<td>rs11568820</td>
<td>TT</td>
<td>15</td>
<td>5</td>
<td>1.36 (0.434-4.298)</td>
<td>0.893</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CT</td>
<td>25</td>
<td>14</td>
<td>0.73 (0.298-1.817)</td>
<td>0.507</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CC</td>
<td>30</td>
<td>12</td>
<td>0.71 (0.288-1.783)</td>
<td>0.474</td>
</tr>
<tr>
<td>FCG2A</td>
<td>rs1801274</td>
<td>GG</td>
<td>19</td>
<td>8</td>
<td>1.14 (0.438-3.007)</td>
<td>0.962</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AG</td>
<td>36</td>
<td>14</td>
<td>0.71 (0.288-1.771)</td>
<td>0.469</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AA</td>
<td>14</td>
<td>11</td>
<td>0.71 (0.288-1.775)</td>
<td>0.471</td>
</tr>
</tbody>
</table>

**OR**: Odds Ratio; **CI**: Confidential Interval. Data analyzed through binary logistic regression. **P < 0.05** are highlighted in bold.
Figure 2

Diagnostic Canonical Value = \(-0.89 \times \text{GMCSF} + 0.41 \times \text{PTPN2}^{272872125} \times \text{CC}\)

A

B

C

D

E

Symptomatic Bleeding Sub-group

Familial Form Sub-group

Infratentorial CCM Sub-group

AUC = 0.661

AUC = 0.774

AUC = 0.804
Figure 3

Figure 4