MacGrogan D, Martínez-Poveda B, Desvignes JP, Fernandez-Friera L, Gomez MJ, Gil Vilariño E, Callejas Alejano S, García-Pavía P, Solis J, Lucena J, Salgado D, Collod-Béroud G, Faure E, Théron A, Torrents J, Avierinos JF, Montes L, Dopazo A, Fuster V, Ibáñez B, Sánchez-Cabo F, Zaffran S, de la Pompa JL. (2020). Identification of a peripheral blood gene signature predicting aortic valve calcification. *Physiol Genomics*, 52(12), 563-574. doi: 10.1152/physiolgenomics.00034.2020

MS PG-00034-2020R1 revised: Identification of a peripheral blood gene signature predicting aortic valve calcification

Running title: A gene signature predicting aortic valve calcification

Supplemental Material and *Suppl. Tables* available at:

https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712 and https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3.

Donal MacGrogan[1,2,*], Beatriz Martínez Poveda[1,2], Jean-Pierre Desvignes[3], Leticia Fernandez-Friera[2,4,5], Manuel José Gomez[6], Eduardo Gil Vilariño[7], Sergio Callejas Alejano[7], Pablo Garcia-Pavia[2,8,9], Jorge Solis[10], Joaquín Lucena[11], David Salgado[3], Gwenaelle Collod-Béroud[3], Emilie Faure[3], Alexis Théron[12], Julia Torrents[13], Jean-François Avierinos[3,12], Lorena Montes[14], Ana Dopazo[7], Valentín Fuster[15,16], Borja Ibañez[2,17], Fátima Sánchez-Cabo[6], Stephane Zaffran[3] & José Luis de la Pompa[1,2,*]

[1] Intercellular Signaling in Cardiovascular Development & Disease laboratory, Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Melchor Fernández Almagro 3, 28029 Madrid, SPAIN
[2] Ciber de Enfermedades Cardiovasculares, Instituto de Salud Carlos III, Melchor Fernández Almagro 3, 28029 Madrid, SPAIN
[3] Aix Marseille University, MMG, INSERM U1251, Marseille, FRANCE
[4] Translational Laboratory for Cardiovascular Imaging and Therapy, CNIC, SPAIN
[5] HM Hospitales-Centro Integral de Enfermedades Cardiovasculares, Madrid, SPAIN
[6] Bioinformatics Unit, CNIC, SPAIN
[7] Genomics Unit, CNIC, SPAIN
[8] Departmento de Cardiología, Hospital Universitario Puerta de Hierro, Madrid, SPAIN
[9] Universidad Francisco de Vitoria, Pozuelo de Alarcón, Madrid, SPAIN
[10] Departmento of Cardiología, Hospital Universitario Doce de Octubre, Madrid, SPAIN
[11] Servicio de Patología Forense, Instituto de Medicina Legal y Ciencias Forenses, Sevilla, SPAIN
[12] Service de Cardiologie, Hôpital de la Timone, 13005, Marseille, FRANCE
[13] Service d'anatomie et Cytologie Pathologiques, Hôpital de la Timone, 13005, Marseille, FRANCE
[14] Hospital Clínico San Carlos, Madrid, SPAIN
[15] Cardiovascular Imaging and Population Studies Laboratory, CNIC, SPAIN
[16] Cardiology Department, Icahn School of Medicine at Mount Sinai, New York, New York, USA
[17] IIS-Fundación Jiménez Díaz Hospital Universitario, Madrid, SPAIN
* Corresponding authors

Correspondence:
José Luis de la Pompa
Intercellular Signaling in Cardiovascular Development & Disease laboratory
Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC)
Melchor Fernández Almagro 3, 28029 Madrid, SPAIN.
Tel: +34 91 4531200
Email: jlpompa@cnic.es

ABSTRACT

Calcific aortic valve disease (CAVD) is a significant cause of illness and death worldwide. Identification of early predictive markers could help optimize patient management. RNA-sequencing was carried out on human fetal aortic valves at gestational weeks 9, 13, and 22, and on a case-control study with adult non-calcified and calcified bicuspid and tricuspid aortic valves. In dimension reduction and clustering analyses, diseased valves tended to cluster with fetal valves at week 9 rather than normal adult valves, suggesting that part of the disease program might be due to re-iterated developmental processes. The analysis of groups of co-regulated genes revealed predominant immune-metabolic signatures, including innate and adaptive immune responses involving lymphocyte T cell metabolic adaptation. Cytokine and chemokine signaling, cell migration, and proliferation were all increased in CAVD, whereas oxidative phosphorylation and protein translation were decreased. Discrete immune-metabolic gene signatures were present at fetal stages and increased in adult controls, suggesting that these processes intensify throughout life and heighten in disease. Cellular stress-response and neurodegeneration gene signatures were aberrantly expressed in CAVD, pointing to a mechanistic link between chronic inflammation and biological ageing. Comparison of the valve RNA-sequencing dataset with a case-control study of whole blood transcriptomes from asymptomatic individuals with early aortic valve calcification identified a highly predictive gene signature of CAVD and of moderate aortic valve calcification in overtly healthy individuals. These data deepen and broaden our understanding of the molecular basis of CAVD and identify a peripheral blood gene signature for the early detection of aortic valve calcification.

Keywords

Human fetal valve; CAVD; inflammation; peripheral blood biomarker; gene signature

INTRODUCTION

Calcific aortic valve disease (CAVD) is caused by calcium build-up of the aortic valve leaflet and may be associated with hemodynamic obstruction, leading to different degrees of aortic stenosis (AS). Untreated severe AS leads to heart failure, with valve replacement remaining the only available treatment. Severe AS is preceded by aortic sclerosis, a subclinical form of CAVD that develops over many years without hemodynamic consequence (6). The prevalence of aortic valve disorders increases with age, with 25% of individuals >65 years with aortic sclerosis progressing to AS at a rate of 2% per year (4). Other than age, the principal CAVD risk factor for AS is a congenitally malformed bicuspid aortic valve (BAV), which is present in 0.5-2% of the general population (30). Patients with AS and a BAV are a decade younger at diagnosis than those with normal tricuspid aortic valve (TAV) morphology (2). Premature leaflet thickening and calcification in BAV patients may be caused by abnormal shear stress hemodynamics.

CAVD shares mechanistic features with atherosclerosis (6). The current accepted model implicates initial lipid oxidation, inflammatory infiltration by T-lymphocytes and macrophages, cytokine and pro-fibrotic signaling, and altered extracellular matrix (ECM) secretion, leading to ectopic bone formation and tissue mineralization (11, 22). The factors predicting transition from the “at risk” status to aortic sclerosis and subsequently to severe AS, are not known (25). Transcriptional profiling of CAVD has helped to identify hundreds of genes differentially expressed between calcified and normal valves (8). Multi-omics approaches combining transcriptomics and proteomics have provided a high-resolution expression atlas and established potential drivers of disease progression (29).

Despite this progress, thus far no medical therapy has been proven to be effective in preventing progression to severe disease, with randomized clinical trials demonstrating the lack of effect of statins on disease progression (35). It has been suggested that statin treatment would have to begin at an earlier disease stage to have a beneficial effect. Current imaging modalities, especially echocardiography, remain the gold standard for diagnosis of valve disease but are

mainly performed when symptoms appear. Non-contrast computed tomography may have significant predictive value in both asymptomatic and symptomatic patients, but remains a source of radiation exposure and may not be available in centers with limited resources. Biomarkers may complement imaging modalities to detect early changes, especially early calcification. The ideal biomarker would predict disease as well as improve prognosis, subclinical disease management, and risk stratification and reduce disease-associated morbidity and mortality. Already identified potential biomarkers include markers of myocardial injury, cardiac mechanical stretch, inflammation, and hemostasis imbalance; however, with the exception of lipoprotein(a), these are not specific to CAVD or have yet to be validated in large randomized studies (31).

The aortic valve develops from rudimentary endocardial cushions that are subsequently refined into thin leaflets well into postnatal life. This process is dependent on complex cell-ECM interactions and is modulated by hemodynamics, and its spatio-temporal dysregulation results in malformations and maladaptive remodeling (10). Several lines of evidence indicate that valve disease at any age is accompanied by some underlying structural defect (28), suggesting that latent CAVD has a developmental origin. Moreover, developmental pathways may also be re-activated during tissue repair processes in the disease setting. For example, molecular mechanisms controlling heart valve cell differentiation are shared with cartilage, tendon, and bone development (reviewed in (15), and are reactivated during endochondral ossification and mature lamellar bone formation taking place in end-stage aortic valve disease (18). Remarkably, only end-stage CAVD has been described at the molecular level, whereas the molecular underpinnings of subclinical disease are poorly understood, due in part to difficulties in accessing biological material from early lesions.

To gain mechanistic insight, we performed an RNA-seq analysis of human fetal and adult control and calcified bicuspid (cBAV) and tricuspid valves (cTAV). K-means clustering was used to identify dynamically regulated processes taking place simultaneously at crucial stages of

123 human valve development and in diseased valves. The prevailing gene signatures were
124 associated with immune-metabolic processes and heretofore uncharacterized pathways linked to
125 cellular stress and ageing. A search for genes shared between our gene expression datasets and
126 datasets generated from asymptomatic subjects from the Progression of Early Subclinical
127 Atherosclerosis (PESA) cohort (5) with aortic valve calcification and without AS identified
128 marker genes in peripheral blood of asymptomatic individuals. When combined in quantitative
129 RT-PCR assays, these marker genes showed good sensitivity and specificity for predicting
130 CAVD at a subclinical stage.

RESULTS

Diseased valves are transcriptionally more similar to developing valves than to controls

Genome-wide transcriptome profiling was carried out on normal adult control valves (n=8) and severely stenotic calcified BAV (cBAV; n=5) and TAV (cTAV; n=7) valves, as well as on fetal aortic valves at gestational weeks 9 (n=4), 12-13, (n=3) and 22 (n=2) (Supplementary Table S1, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). These developmental timepoints correspond to mouse embryonic days (E) 13.5-17.5 (13), during which valves are remodeled by morphogenic processes leading to thinning, elongation, and formation of a multi-laminar valve structure. To integrate data across studies and generalize our findings, we conducted a meta-analysis with publicly available expression data from normal adult control (n=8), cBAVs (n=10), and cTAVs (n=9) valves (8) (Figure 1A). After correcting for batch effects, Principal Component Analysis (PCA) indicated that sample differences did not correlate with batch origin (Supplementary Fig. S1A, see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712), but rather depended on disease or developmental stages (Figure 1B). The first of the PCA components (PC1) separated developing and adult valve samples according to disease severity, ranging from cTAV (right hand side) to controls (left hand side). Interestingly, fetal valves tended to cluster with diseased adult valves. There were two control samples (both from heart transplants) and one sample of embryonic samples at 9 weeks showing more heterogeneity than the rest according to the second component (y-axis) but still the range of the axis (-0.6 to 0.4) is very small, confirming the reproducibility of the samples in terms of global transcriptome. In the second PCA component (PC2), the adult control and disease samples were separated from fetal weeks 9 and 13 samples (Figure 1B). Therefore, and considering that the first PCA dimension explains 15.24% of the variance, while the second dimension explains only 7.73%, adult disease samples appeared to be more closely related to fetal than to control samples,

suggesting that some of the disease pathogenesis can be ascribed to the recapitulation of developmental processes.

To further support this finding, we used Euclidean distances between overall gene expression profiles to perform hierarchical clustering (HC) of the samples (Figure 1C). Similar to what was observed in the PCA, disease samples cluster further apart from controls and appear under the same branch with embryonic samples at 9 and 13 weeks.

Clustering identifies dynamic gene expression across disease and developmental states

Differential expression analysis was initially carried out between each pair of conditions of interest (Supplementary Table S2, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3; 9 pairwise comparisons). However, two reasons precluded pairwise analysis: (1) the existence of an extremely large number of genes that seemed to be changing across conditions (ANOVA identified 19573 changing genes, Supplementary Table S3, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3) and (2) the difficulty of identifying the most relevant genes across conditions by comparing lists of differentially expressed genes (2^6=32 potential comparisons). Therefore, we sought to explore the coordinated gene expression profiles across conditions using k-means clustering. K-means analysis computing n=50 yielded dynamic expression profiles across the different conditions (Supplementary Fig. S1B, see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712). Co-expressed genes (ie. genes undergoing similar disease/stage status expression changes) likely participate in the same process, and genes that function in the same process are regulated coordinately (20). There were 6 broad combinations of expression patterns: 1) genes down-regulated in developing valves and up-regulated in control adult valves (e.g. C34); 2) genes down-regulated in control and up-regulated in diseased valves (e.g. C7); 3) genes up-

182 regulated in developing valves, down-regulated in adult control valves, and up-regulated in
183 diseased valves (e.g. C45); 4) genes up-regulated in developing valves and down-regulated in
184 adult control and disease valves (e.g. C37); 5) genes up-regulated in adult control valves and
185 down-regulated in diseased valves (e.g. C16); and 6) genes down-regulated in developing valves,
186 up-regulated in adult control valves, and down-regulated in diseased valves (e.g. C40)
187 (Supplementary Fig. S1B, see
188 https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712). Thus, part of the disease process involves a gene signature present during fetal valve
189
190 development.
191
192 **Comparative functional analysis identifies multiple interconnected pathways**
193 Functional enrichment analysis of clusters with IPA identified significant associations to
194 canonical pathways in 28 clusters (Figure 2). Subsequent unsupervised hierarchical clustering of
195 clusters, based on the fraction of shared canonical pathways, identified 6 metaclusters (MCs),
196 termed MC1-MC6 (Figure 2). Cluster 10 did not overlap with the others in terms of functional
197 enrichment, and was not further considered. A heatmap was generated based on the MC genes
198 (Supplementary Fig. S1C, see
199 https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712) and broadly reproduced the hierarchical clustering on all genes (Figure 1C), except for
200
201 the relative positions of 13wk and 22wk samples. Generally, genes highly expressed in adult
202 disease showed low expression at gestational week 13 and vice-versa; however, there were
203 exceptions (clusters 9, 12, 23, and 45; Figure 2). Although metaclusters were defined as non-
204 overlapping collections of gene clusters, enriched functions could be shared between several
205 meta-clusters. For example, canonical pathways associated to clusters 46 and 32 were shared by
206 MC1 and MC2; those associated to cluster 12 were shared by MC2, MC3, MC4, and MC6; those
207 to clusters 12 to 45 were shared by MC2, MC3, and MC4; and those canonical pathways

enriched in clusters 5 and 45 were shared by MC4 and MC6 (Figure 2). Therefore, the biological functions represented in the different MCs are interconnected.

Innate and adaptive immune processes are present in valve disease and development

The collections of genes associated to metaclusters MC1-MC6 were reanalysed with IPA. Enriched canonical pathways associated to each MC are summarized in Figure 3. MC3 was associated to 168 pathways related to innate-adaptive immune processes (Supplementary Fig. S2, see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712). The top 5 pathways were Th1 and Th2 Activation Pathway, TREM1 Signaling, Communication between Innate and Adaptive Immune Cells, Phagosome Formation and Dendritic Cell Maturation (Figure 3; Supplementary Fig. S3, see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712; Supplementary Table S4, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). The remaining 163 enriched biological functions were also involved in critical immune-inflammatory cellular processes (Supplementary Fig. S2, see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712; Supplementary Table S4, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). Among the nine clusters combined into MC3, five of them had expression profiles characterized by increased gene expression in all adult samples, both control and CAVD (clusters 22, 9, 34, 36 and 2), whereas four clusters had increased expression in CAVD only (clusters 49, 44, 7, and 35; Figure 3). Discrete immune cell gene signatures were present at week 9 (clusters 44 and 35), week 13 (cluster 9), and week 22 (clusters 49 and 22) of development (Figure 3), but involved relatively fewer pathways than those found in control or calcified adult valves. The presence of immune

and pro-inflammatory gene signatures during fetal development and in apparently healthy valves is consistent with the progressive nature of the calcified valve disease pathology.

Cell migration and DNA replicative pathways are re-activated in CAVD

Metaclusters MC4 and MC6 were enriched in genes belonging to 202 and 23 IPA pathways, respectively (Supplementary Fig. S4 and S6A, see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712; Supplementary Table S4, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). In general, genes associated to these clusters have increased expression during development (notably at week 13), decreased in adulthood and re-activated in CAVD. Therefore, of the 7 clusters associated with these two MCs, six were associated with more elevated gene expression in the diseased state (clusters 25, 23, 5, 45, 17, and 21), whereas only one was associated with lower expression (cluster 40; Figure 3).The top five pathways in MC4 were Actin Cytoskeleton Signaling, Molecular Mechanisms of Cancer, Axonal Guidance Signaling, Epithelial Adherens Junction Signaling and Integrin Signaling (Figure 3, Supplementary Fig. S5, Supplementary Table S4, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). The remaining 197 IPA pathways, represented components of fundamental cell and developmental signaling pathways (Supplementary Fig. S4, see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712; Supplementary Table S4, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). The top five pathways in MC6 were Role of CHK Proteins in Cell Cycle Checkpoint Control, Sumoylation Pathway, Cell Cycle Control of Chromosomal Replication, Cell Cycle: G2/M DNA Damage Checkpoint Regulation, Role of BRCA1 in DNA Damage Response (Figure 3, Supplementary Fig. S6B, see

https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712, Supplementary Table S4, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). The remaining 18 IPA pathways were involved in similar processes (Supplementary Fig. S6A, see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712; Supplementary Table S4, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). Cellular proliferation and migration may be part of the ongoing immune-inflammatory response associated with processes represented in MC3.

Decreased OXPHOS and ribosomal biogenesis in CAVD

Genes in MC2 were significantly associated to 83 IPA pathways (Supplementary Fig. S7, see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712; Supplementary Table S4, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3), the top 5 of which were Mitochondrial Dysfunction, EIF2 Signaling, Oxidative Phosphorylation, Sirtuin Signaling Pathway (Supplementary Fig. S8A-C), Regulation of eIF4 and p70S6K Signaling (Supplementary Fig. S8D,E see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712; Supplementary Table S4, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). The remaining 78 pathways in MC2 were also implicated in energy production and protein metabolism (Supplementary Fig. S7, see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712; Supplementary Table S4, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3), showing strong

enrichment in genes for nuclear encoded subunits of the electron transport chain complexes and ribosomal biogenesis (Supplementary Fig. S8A-C, see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712; Supplementary Table S4, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). Other gene signatures were AMP-activated protein kinase (AMPK), ER stress (Unfolded protein response), oxidative stress response (NRF2-mediated Oxidative Stress Response), cellular longevity (Telomerase Signaling), and neurodegenerative disease (Huntington's Disease Signaling; Supplementary Fig. S7, see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712; Supplementary Table S4, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3), suggesting the modulation of cellular stress responses, and age-related disease gene signatures. Of the nine clusters in MC2, six had, on average, decreased expression in disease conditions (clusters 24, 31, 32, 16, 41, and 1), while one had increased expression in the diseased state (cluster 12) and three were unchanged (clusters 31, 26, and 37; Figure 3). In general, MC2 transcripts had peak expression at week 13, with lower expression at the other developmental stages (Figure 3), suggesting that metabolic processes are dynamically regulated during aortic valve development. Overall, this analysis indicates that CAVD is associated with decreased OXPHOS metabolism and altered proteostasis.

Stress response gene signatures are altered in CAVD

MC1 and MC5 did not overlap significantly with any of the other MCs but included similar biological processes. Genes in MC1 were associated to 7 pathways (Figure 3; Supplementary Fig. S9A, see

https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712; Supplementary Table S4, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3), the top 5 being Sphingosine-1-phosphate Signaling, Endocannabinoid Developing Neuron Pathway, Amyloid Processing, Huntington's Disease Signaling, Dermatan Sulfate Degradation (Figure 3; Supplementary Fig. S9B, see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712; Supplementary Table S4, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). The expression of MC1 genes was relatively increased in adult controls and decreased in CAVD (Figure 3). MC5 was enriched in a single pathway, Unfolded protein response (Figure 3; Supplementary Fig. S9C, see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712; Supplementary Table S4, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). MC5 genes were expressed at decreased levels in controls and increased in CAVD (Figure 3). These data suggest that chronic immune-inflammatory processes are linked to cellular stress and age-related disease in calcific aortic valve pathology.

Identification of circulating biomarkers predicting CAVD

To identify potential biomarkers useful for predicting CAVD, a sub-set of genes from the aortic valve study was compared with whole blood RNA-Seq data from asymptomatic individuals with aortic valve calcification, detected by non-contrast computed tomography in the *Progression of Early Subclinical Atherosclerosis* (PESA) study (5). PESA participants undergo simultaneous cardiac calcium scoring and blood sampling. Of the 455 PESA participants for which RNA-seq data was available in visit 2 (ie. three years after recruitment), 59 had evidence of valve

calcification and 52 of theses had sufficient integrity for qPCR analysis. None of them had valvular stenosis/regurgitation, evidencing an early stage of the disease. Their transcriptome was compared to the transcriptome of PESA individuals without valvular calcification (n=395). From the whole set of genes identified as differentially expressed ($p<0.05$, Supplementary Table S5, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3), six were selected that showed a disease-specific pattern in the RNA-seq data of human aortic valves (see Methods for selection criteria; Supplementary Fig. S1, see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712). Five of these genes were related to adaptive innate and adaptive responses (*CD28, ITK, PAG1, NLRC5, and ASCC3*) and a sixth was related to the WNT pathway (*LINS1*) (Supplementary Table S6, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). The expression of the six genes was measured in whole blood (WB) on 104 PESA participants [52 cases: asymptomatic individuals with calcification of the aortic valve (PESA-SD); 52 controls: individuals without calcification (PESA-CTL)] and 12 patients with CAVD recruited at Monteprincipe Hospital (WB-CAVD) by qRT-PCR and correlated with the degree of aortic valve calcification (Figure 4A, Supplementary Table S7, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). Considering all conditions, the 6 genes showed some degree of correlation when considered individually ($P \leq 0.27$; Figure 4A, Supplementary Table S8, see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712 and Supplementary Table S9, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). Given that, *NLRC5* and *PAG1* were also associated with total calcification of the coronaries (Supplementary Fig. S10A, see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712, Supplementary Table S8, see

363 https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3) and were
364 therefore excluded as potential specific markers of aortic valve calcification.

365 To determine the utility of *ASCC3*, *ITK*, *CD28* and *LINS1* as potential markers of valve
366 calcification, we first confirmed that they were differentially expressed between valve tissue
367 disease and control samples (Supplementary Table S10, see
368 https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). Then, a ROC
369 curve was built to represent the predictive value of the model to classify calcific disease patients
370 (WB-CAVD) vs PESA-SD individuals and vs PESA-CTL, yielding areas under the curve
371 (AUCs) of 0.97 and 0.91 respectively (Figure 4B, Supplementary Table S11, see
372 https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). An additional
373 model was built to assess the power of the gene signature to distinguish between individuals
374 without calcification (PESA-CTL) from those with subclinical disease (PESA-SD; AUC=0.71)
375 (Figure 4B, Supplementary Table S11, see
376 https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3).

377 This model was then validated in whole blood samples of an independent cohort of
378 CAVD patients (n=16), together with samples from 25 of the 52 PESA-SD individuals and 35 of
379 the 52 PESA-CTL individuals taken three years after visit 2 (Figure 4C). The model achieved
380 AUCs of 0.90 and 0.88 to discriminate CAVD patients from PESA-SD and PESA-CTL
381 individuals, respectively. The AUC of the model to distinguish PESA-SD from PESA-CTL was
382 again 0.71 (Figure 4C). Thus, our model identifies a gene expression signature in blood that is
383 highly predictive of valve disease and of aortic valve calcification levels in otherwise healthy
384 individuals.

DISCUSSION

This study provides, for the first time, a comprehensive molecular profile of shared and unique cellular activities taking place in aortic valve development and CAVD. Our data broadly confirms previous transcriptomics studies revealing immune-fibro-calcific nature of CAVD and similarities of the expression profiles of calcified BAV and TAV (8,19). The overall cBAV and cTAV gene expression profiles were essentially identical, suggesting that the mechanisms leading to BAV formation are independent of those leading to BAV calcification; in other words, the cause of the embryonic defect is distinct from the resulting calcification process. Furthermore, our study extends these findings to the remodeling stages of aortic valve development; disease samples were more closely related to fetal week 9 than to control samples or weeks 13 and 22, suggesting that some of the disease pathogenesis might be ascribed to the recapitulation of processes taking place in early development.

Functional analysis provided compelling evidence that innate and adaptive immune interactions take place in the aortic valve throughout life (Supplementary Fig. S10B, see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712). The fetal immune cell gene signature in MC3 was notably simpler than in adult control or CAVD samples, with fewer immune cell gene expression signatures present at gestational weeks 9 (3/9 clusters), 13 (1/9 clusters), or 22 (3/9 clusters) than in adult controls (5/9 clusters) or in CAVD (9/9 clusters), suggesting that aortic valve immune cell enrichment in the aortic valve increases over time. Moreover, the presence of only one immune cell gene signature at 13 weeks (cluster 9), compared with three at 9 and 22 weeks of gestation, implies that the immune landscape is dynamically regulated in developing valves. These observations are consistent with previous findings of multiphasic changes in shear stress-response, cell-ECM interaction and adaptive immunity pathways from week 14 to week 22 of human aortic valve development (7).

The predominant T cell gene signature in the RNA-seq dataset strengthens the notion that unresolved immune-inflammatory response is a major CAVD component (17). This association

411 is well established, with early immunohistological findings indicating the presence of T
412 lymphocytes in immune infiltrates (23, 24). More recent studies have revealed the existence of T
413 cell infiltration in BAV and TAV calcific disease, consisting of non-specific chemokine driven
414 polyclonal T cell recruitment, and memory-effector CD8+ T cell clonal expansions (33, 34),
415 supporting the interpretation of an antigen-induced immune response. Moreover, this antigen-
416 driven response involved circulating expanded CD8+ and, to a lesser extent, CD4+ T cell clones
417 between circulating blood and valve tissues suggesting a systemic disease (33). The gene
418 expression signatures of several other immune-competent cells, including B-lymphocytes,
419 macrophages, and dendritic cells were also present, confirming previous characterization of
420 cellular infiltrates in CAVD (32). Together with a body of literature suggesting a tight
421 inflammatory cell-valve interstitial cell interaction in CAVD (11, 17), our data are consistent
422 with the notion that unresolved inflammation ultimately drives myo-fibroblast and osteoblast-
423 like differentiation and leads to heterotopic bone formation. Therefore, targeted therapeutic
424 immunosuppression of calcification might be approachable by better characterizing the various
425 infiltrating immune-cell subtypes.

426 Functional analysis revealed that OXPHOS metabolism was decreased in CAVD, with
427 coordinated expression reduction of multiple nuclear genes encoding subunits of mitochondrial
428 electron transport chain complexes I, III, IV and V. These changes might be part of ongoing
429 adaptive metabolic reprogramming during T lymphocyte activation (26), when upon
430 encountering an antigen, T cells proliferate and acquire new functions, switching energy
431 metabolism from OXPHOS to glycolysis. However, both increased and decreased expression of
432 glycolytic pathway components were present, suggesting that decreased OXPHOS metabolism
433 cannot be ascribed solely to T cell activation. Moreover, adult aortic valves showed markedly
434 reduced transcript levels of multiple subunits of the small and large protein ribosomal complexes
435 (RPS and RPL respectively) and eukaryotic translation initiation factors (EIF1-4), possibly
436 reflecting co-ordinate dysregulation of OXPHOS activity and ribosomal biogenesis (19). This

437 occurred in uncalcified adult valves as well (clusters 31, 37, and 12), suggesting declining
438 function with age. In contrast signatures for cell death and apoptosis (for example in MC3) and,
439 energy-demanding cell processes such as proliferation, DNA repair, and cell movement (in MC4
440 and MC6) were increased, possibly reflecting cellular heterogeneity within the calcified valve.

441 Decreased mitochondrial metabolism and protein translation may be a manifestation of
442 altered physiological stress responses that occur during ageing (29), which is also related to the
443 inability to maintain adaptability to stress and re-establish homeostasis in response to
444 environmental perturbations. Thus, multiple stress-response gene signatures were present in
445 MC2 and to a lesser extent in MC1 and MC5 (Supplementary Fig. S10B, see
446 https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/120
447 86712). These included SIRT, a highly conserved family of NAD-dependent deacetylases,
448 linking transcriptional regulation to mitochondrial function, stress resistance, metabolism, cell
449 survival and organismal longevity (3) and mTOR (Mammalian target of rapamycin), a nutrient
450 sensing system which regulates protein synthesis, to promote cell growth and proliferation (19).
451 Moreover, deficiency of the unfolded protein response (UPR) pathway, which is activated by
452 misfolding of newly synthesized proteins as part of the ER stress response (9) contributes to
453 inflammation, disease and ageing. Gene signatures including amyloid processing and
454 Huntington's disease signaling were also featured implying shared calcific valve and
455 neurodegenerative disease mechanisms (Supplementary Fig. S10B, see
456 https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/120
457 86712). Therefore, chronic inflammation drives metabolic stress and biological ageing in
458 CAVD.

459 Given the phenomenon of cellular trafficking of T cells between the diseased valve
460 tissue, and circulating blood (33), we reasoned that aortic valve and circulating blood from
461 CAVD patients share common genes that would be absent in subclinical individuals. In the
462 PESA study we identified a subpopulation of asymptomatic individuals in this cohort with > 1

valve calcification detected by imaging. Therefore, we searched for circulating biomarkers by comparing the CAVD transcriptomes and whole blood from datasets of the subclinical PESA study population, and identified six differentially expressed genes, four of which were found to predict subclinical and overt valve disease. Therefore *ASCC3, ITK, CD28,* and *LINS* were specific to aortic valve calcification, whereas *PAG1* and *NLRC5* were also associated with coronary artery calcification, suggesting that valve and vascular calcification proceed by shared and separate mechanisms. These mechanisms have yet to be examined, but noteworthy is that five of six genes belong to T cell activation and NF-kB pathways. This gene signature could be used in combination with high-resolution imaging for the early diagnosis of valve calcification.

In summary, our computational approach, based on clustering of the expression profiles and on the concordance between enriched pathways across clusters allowed a detailed analysis of the vast amount of data generated by RNA-seq. This strategy allowed us to narrow down the initial 19,662 set of genes to 10,256 and, importantly, we straightforwardly established a correspondence between the entire expression profile across six conditions and the pathways enriched for each of the final meta-clusters. In particular, we identified fetal transcriptional profiles in calcific valve disease suggestive of developmental reprogramming. Very likely, a more traditional approach, involving the comparison of lists of differentially expressed genes followed by functional analysis, would have failed to extract all the relevant information hidden in the RNA-seq data.

Study Limitations. A limitation of our study is the low sample numbers especially for week 22 samples (n=2). Investigational studies involving human fetal issue are limited by ethical and legal considerations on medical abortion. Nevertheless, our bioinformatics analysis showed that the week 22 samples clustered together and separated well from the other groups by PCA and we are confident of the robustness of the results involving this group. Moreover, differential expression analysis was performed using limma, a statistical method developed ad-hoc for RNA-seq data with low sample size (27). However, our gene signature still needs to be validated in a

489 larger number of individuals, with detailed assessment of false positives and false negatives.
490 Gene expression analysis might be a limitation in clinical settings and more direct and low-cost
491 techniques (e.g. ELISA) need to be investigated for use in the clinic.
492

METHODS

Study population and aortic valve collection.

Fetuses. Aortic valves from human fetuses were obtained from electively- (weeks 9; n=4 and week 12/13; n=3) or medically (week 22; n=2) -terminated pregnancies after written informed consent in concordance with French legislation (PFS14-011) and prior protocol approval from the "Agence de la biomédecine" (Supplementary Table S1, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). Aortic valves were recognized by their anatomical landmarks under the microscope and leaflets were isolated with minimal aortic wall contamination Samples were immediately flash frozen in liquid nitrogen. The investigation conformed to the principles outlined in the *Declaration of Helsinki*.

Patients. The CAVD study population consisted of BAV (n=5) or TAV (n=7) patients referred for aortic valve or aortic root surgery at Monteprincipe Hospital and Hospital Clínico San Carlos (Madrid, Spain; Supplementary Table S1, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). Patients were prospectively recruited between December 2009 and February 2013. Patients with a history of rheumatic heart disease, infective endocarditis, or connective tissue disorders were excluded. To determine the morphology and functional state of the aortic valve, all patients underwent clinical evaluation and preoperative 2-dimensional echocardiography according to the *American Heart Association* and the *American College of Cardiology* guidelines (21). The aortic valve morphology and the presence of valvular degeneration, including the extent of leaflet calcification, thickening, prolapse, and/or redundancy, were documented at surgery. Calcification was graded as mild (1/3 leaflet area affected) or severe (2/3 leaflet area affected). Immediately following surgical removal, a portion of the aortic valve leaflet was placed in sterile saline buffer and stored at -80°C until processing. Control valve samples were obtained at autopsy of individuals without cardiac problems who suffered a traumatic death (n=6) or at time of transplantation from heart transplant receptors with normal aortic valves (n=2). All patients or

519 close relatives (for autopsy controls) gave written informed consent, and the study was approved
520 by the Health Service Ethics Committees (Comité Ético de Investigación Clínica) of participant
521 centers. The investigation conformed to the principles outlined in the *Declaration of Helsinki*.
522 *Whole-blood samples.* Genes of interest were quantified by quantitative reverse transcription-
523 polymerase chain reaction (qRT-PCR) in whole-blood (WB) samples from age-matched case-
524 control males (n=104) of the Progression of Early Subclinical Atherosclerosis (PESA).
525 Subclinical cases (PESA-SD) were defined as participants with aortic-valve calcification score
526 >1, as assessed by the Agatston method (1). Controls (PESA-CTL) had no calcification of the
527 aortic valves or coronary arteries. Patients with aortic valve stenosis and moderate to severe
528 calcification (WB-CAVD; n=12) were recruited from Monteprincipe Hospital (Supplementary
529 Table S7, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). For
530 validation, an independent cohort of mainly severe aortic valve stenosis patients (n=16) was
531 obtained from the MRVALVE study. Whole blood was collected in PAXGene tubes
532 (BDBiosciences) and stored at -80^{o}C before processing.
533 **RNA isolation**
534 For each valve, total RNA from a single aortic valve cusp was extracted using the TRIzol LS
535 (Invitrogen, Carlsbad, CA, USA) method and digested with RNase free DNase I (Invitrogen).
536 For PESA and CAVD validation total RNA was using either the PAXgene blood mRNA kit
537 (PreAnalytiX, Hombrechtikon, Switzerland) for manual isolation or the QIAsymphony PAXgene
538 blood RNA kit (PreAnalytiX) for automated isolation using a QIAsymphony SP liquid handling
539 robot (Qiagen, Venlo, Netherlands). RNA purity, concentration, and integrity were assessed by
540 Nanodrop 2000 spectrophotometry (Thermo Scientific) and automated electrophoresis in a 2100
541 Bioanalyzer (RNA6000 Nano LabChip; Agilent). Samples with a RIN > 6 were selected for
542 reverse transcription.
543 **RNA-seq data generation**

RNA-seq was performed on DNase I-treated RNA samples with a RIN > 7. RNA-seq libraries were created using the TruSeq Stranded Total RNA with Ribo-Zero Gold Prep Kit (Fetal samples) or TruSeq™ RNA Sample Prep Kit v2 (Patient samples; Illumina, Diego, CA). Final cDNA libraries were checked for quality and quantified in the 2100 Bioanalyzer (Agilent). The libraries were loaded in the flow cell at 8 pM, and clusters were generated in Cbot and sequenced as 50bp single-end reads on a Hiseq 2500 instrument (adult samples) or 75 base single-end reads on a Genome Analyzer IIx instrument (fetal samples; Illumina, Diego, CA). Image analysis and base calling were performed using RTA and CASAVA.

Bioinformatics analysis of RNA-seq.

Data preprocessing

FASTQ files were pre-processed with cutadapt (16) to eliminate Illumina adapter remains and reads were aligned to the human reference transcriptome (GRCh38.78) using RSEM 1.2.31 (14). Raw data from 27 RNA-seq samples from GSE76718 were downloaded from GEO and pre-processed and aligned using the same pipeline. Control samples were present in all experiments, allowing for batch correction of the data using ComBat (12). Data were normalized using TMM with log-ratio trimming=0.3 and sumTrim=0.05 and voom-transformed. Only genes with at least 1 cpm in at least 10 samples were considered for further analysis.

Differential Expression Analysis

Differential expression between groups was performed using mixed models with eBayes, as implemented in the Bioconductor limma package (27). A gene was classified as differentially expressed if its Benjamini-Hochberg adjusted *P*-value was < 0.05 in any of the contrasts under study (Supplementary Table S2, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3).

Clustering and functional analyses

To identify groups of genes with similar expression profiles across the six conditions under consideration, k-means clustering was performed with the R ComplexHeatmap package on all

expressed genes (19,962 genes). The Elbow method was used to estimate the optimal number of clusters, which was set as k=50. The number of genes per cluster was between 215 and 680. Ingenuity Pathway Analysis (IPA, Qiagen) was then used to identify significant associations between canonical pathways and clusters. Twenty-eight clusters, out of the original collection of 50, were found to be associated to at least one canonical pathway with Benjamini-Hochberg (B-H) adjusted *P*-value < 0.05. Concordance between canonical pathway profiles was then used to identify groups of clusters that could be associated to similar functions. One cluster (cluster 10) did not share any enriched pathway with any other cluster and was not considered further. The remaining 27 clusters were grouped into six meta-clusters, using the fraction of shared pathways as distance measure. Meta-clusters contained from 332 to 3304 genes (Figure 2; Supplementary Table S4, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). IPA was used again to re-annotate meta-clusters (Figure 3).

Data availability

Data are deposited in the NCBI GEO database under accession number GSE148219. The scripts can be acessed at Github: https://github.com/mjgommo/CAVD_COMP_PIPELINES.

qRT-PCR for validation studies in whole blood

Selection of candidate genes for qRT-PCR. We used whole blood transcriptomics data generated in the PESA study to identify genes that might potentially serve as predictors of CAVD. The criteria to select these candidate genes were: (i) Genes differentially expressed between individuals with (n=59) and without (n=396) calcification of the aortic valve (p-value<=0.05) but not differentially expressed between individuals with and without calcification in the coronaries (p-value>0.05; Supplementary Table S5, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3). (ii) Aortic valve-specific differentially expressed genes with a top 10% correlation between their expression level and the level of calcification of the aortic valves. (iii) Their expression levels in the RNA-seq samples of the aortic valves were either up-regulated in disease v.s. no disease samples or they

were upregulated in early embryonic valve samples and in the diseased samples but not in the control samples. From the 10 genes fulfilling the three criteria, 6 were selected based on their presumptive function (Supplementary Table S6, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3).

qRT-PCR. RNA samples from PESA and CAVD patients (75ng) was reverse transcribed using the High Capacity cDNA Reverse Transcription kit with RNase Inhibitor (ThermoFisher Scientific). Gene expression was quantified by qPCR using TaqMan probes (ThermoFisher Scientific; Supplementary Table S12, see https://figshare.com/articles/dataset/Supplementary_Tables_1-12/12086724/3) in 52 case-control pairs from the PESA study (case: positive calcification of the aortic valve) and in 12 CAVD patients from Monteprincipe Hospital. Each reaction was performed in triplicate and the final reaction volume was set to 10µl. qPCRs were conducted in 384-well plates (Cat. no. 4344345; Applied Biosystems) using an ABI PRISM 7900HT Sequence Detection System (Applied Biosystems). Each well included 3µl (2.25 ng) cDNA, 5µl 2x TaqMan Universal Mater Mix (ThermoFisher Scientific), 0.5µl 20x TaqMan assay (ThermoFisher Scientific), and 1.5µl H_2O. The cycling protocol was 10 min at 95°C followed by 60 cycles of 15 s at 95°C and 60 s at 60°C. Baseline and threshold were set automatically in SD S 2.3 (ThermoFisher Scientific) to calculate Cq values. The Cq values were then imported into qBasePlus (Biogazelle, Zwijnaarde, Belgium) to calculate relative quantities normalized to *GAPDH*. Out of the 116 individuals analyzed, 4 were considered outliers and removed from the final analysis. Data were scaled per sample to eliminate potential biases.

Statistical analysis

A linear mixed model was used to assess correlation between the expression of each gene in WB and the degree of calcification in the aortic valve for each disease group (PESA-SD, WB-CAVD) separately. Visualization was performed using the multiplot function and the ggplot2 R package. All samples were jointly analyzed using a linear mixed model with the origin of the

622 samples as random effect. The same analysis was performed to assess the correlation of the gene
623 expression with calcification of the coronaries. Four genes which showed no correlation with the
624 amount of calcium in the coronaries were selected to build a gene expression signature using a
625 multinomial model with interaction for the prediction of the three groups of individuals with
626 different disease severity WB-CAVD>PESA-SD>PESA-CTL. To quantify the predictive power
627 of the gene signature between different groups of individuals a logistic regression model was fit
628 to the data and the corresponding ROC curves were built using the ROCR R package.
629

630

Acknowledgments

We thank S. Bartlett for English editing. This work was supported by the Ministerio de Ciencia, Innovación y Universidades from Spain (MCIU, grants SAF2016-78370-R, CB16/11/00399, CIBER CV, and RD16/0011/0021, TERCEL) and the Fundación BBVA (grant BIO14_298) and Fundación La Marató TV3 (grant 20153431) to J.L.dlP; and the Fondation pour la Recherche Médicale, grant from l'Institut National de la Santé et de la Recherche Médicale (grant DPC20111123002) and l'Association Française contre les Myopathies (grant TRIM-RD) to S.Z. The cost of this publication was supported in part with funds from the ERDF. The CNIC is supported by the Ministerio de Ciencia, Innovación y Universidades (MCIU), the Instituto de Salud Carlos III (ISCIII), and the Pro CNIC Foundation, and is a Severo Ochoa Center of Excellence (grant SEV-2015-0505).

Author contributions

Donal MacGrogan, PhD: Design of work: Substantial; interpretation of data: Substantial; Writing: original draft: Lead; Writing: review & editing: Lead. final approval of the version to be published.

Beatriz Martínez Poveda, PhD: Investigation: Lead; Methodology: Substantial; final approval of the version to be published.

Jean-Pierre Desvignes, MD: Investigation: Supporting; final approval of the version to be published.

Leticia Fernandez-Friera, MD, PhD: Data curation: Substantial; Formal analysis: Substantial; Investigation: Supporting; final approval of the version to be published.

Manuel José Gomez, PhD: Formal analysis: Substantial; Visualization: Substantial; final approval of the version to be published.

Eduardo Gil Vilariño, BSc: Investigation: Supporting; final approval of the version to be published.

Sergio Callejas Alejano, BSc: Investigation: Supporting; final approval of the version to be published.

Pablo Garcia-Pavia, MD, PhD: Investigation: Supporting; final approval of the version to be published.

Jorge Solis, MD, PhD: Investigation: Supporting; final approval of the version to be published.

662 Joaquín Lucena, MD, PhD:Investigation: Supporting; final approval of the version to be
663 published.
664 David Salgado, PhD: Data curation: supporting; Formal analysis: supporting; Investigation:
665 supporting; final approval of the version to be published.
666 Gwenaelle Collod-Béroud, PhD: Formal analysis: Supporting; Investigation: Supporting; final
667 approval of the version to be published.
668 Emilie Faure, PhD: Formal analysis: Supporting; Investigation: Supporting; final approval of the
669 version to be published.
670 Alexis Théron, MD: Investigation: Supporting; final approval of the version to be published.
671 Julia Torrents, MD: Investigation: Supporting; final approval of the version to be published.
672 Jean-François Avierinos, MD: Investigation: Supporting; final approval of the version to be
673 published.
674 Lorena Montes, MD: Investigation: Supporting; final approval of the version to be published.
675 Ana Dopazo, PhD: Investigation: Supporting; Supervision: Supporting; final approval of the
676 version to be published.
677 Valentín Fuster, MD, PhD: Investigation: Supporting; Resources: Supporting; final approval of
678 the version to be published.
679 Borja Ibañez, MD, PhD: Investigation: Supporting; Writing – review & editing: Supporting;
680 final approval of the version to be published.
681 Fátima Sánchez-Cabo, PhD: Formal analysis: Equal; Writing – review & editing: Supporting;
682 final approval of the version to be published.
683 Stephane Zaffran, PhD: Investigation: Supporting; Supervision: Supporting; Writing –review &
684 editing: Supporting; final approval of the version to be published.
685 José Luis de la Pompa, PhD: Conceptualization: Lead; Funding acquisition: Lead; Project
686 administration: Lead; Resources: Lead; Supervision: Equal; Visualization: Equal; Writing –
687 review & editing: Supporting.
688

689 **Competing interests**

690 The authors declare no competing interests.

691

692

693

REFERENCES

1. **Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte M, Jr., and Detrano R**. Quantification of coronary artery calcium using ultrafast computed tomography. *J Am Coll Cardiol* 15: 827-832, 1990.
2. **Beppu S, Suzuki S, Matsuda H, Ohmori F, Nagata S, and Miyatake K**. Rapidity of progression of aortic stenosis in patients with congenital bicuspid aortic valves. *Am J Cardiol* 71: 322-327, 1993.
3. **Chang HC, and Guarente L**. SIRT1 and other sirtuins in metabolism. *Trends Endocrinol Metab* 25: 138-145, 2014.
4. **Coffey S, Cairns BJ, and Iung B**. The modern epidemiology of heart valve disease. *Heart* 102: 75-85, 2016.
5. **Fernandez-Ortiz A, Jimenez-Borreguero LJ, Penalvo JL, Ordovas JM, Mocoroa A, Fernandez-Friera L, Laclaustra M, Garcia L, Molina J, Mendiguren JM, Lopez-Melgar B, de Vega VM, Alonso-Farto JC, Guallar E, Sillesen H, Rudd JH, Fayad ZA, Ibanez B, Sanz G, and Fuster V**. The Progression and Early detection of Subclinical Atherosclerosis (PESA) study: rationale and design. *Am Heart J* 166: 990-998, 2013.
6. **Freeman RV, and Otto CM**. Spectrum of calcific aortic valve disease: pathogenesis, disease progression, and treatment strategies. *Circulation* 111: 3316-3326, 2005.
7. **Gottlieb Sen D, Halu A, Razzaque A, Gorham JM, Hartnett J, Seidman JG, Aikawa E, and Seidman CE**. The Transcriptional Signature of Growth in Human Fetal Aortic Valve Development. *Ann Thorac Surg* 106: 1834-1840, 2018.
8. **Guauque-Olarte S, Droit A, Tremblay-Marchand J, Gaudreault N, Kalavrouziotis D, Dagenais F, Seidman JG, Body SC, Pibarot P, Mathieu P, and Bosse Y**. RNA expression profile of calcified bicuspid, tricuspid, and normal human aortic valves by RNA sequencing. *Physiol Genomics* 48: 749-761, 2016.
9. **Hetz C**. The unfolded protein response: controlling cell fate decisions under ER stress and beyond. *Nat Rev Mol Cell Biol* 13: 89-102, 2012.
10. **Hinton RB, and Yutzey KE**. Heart valve structure and function in development and disease. *Annu Rev Physiol* 73: 29-46, 2011.
11. **Hulin A, Hego A, Lancellotti P, and Oury C**. Advances in Pathophysiology of Calcific Aortic Valve Disease Propose Novel Molecular Therapeutic Targets. *Front Cardiovasc Med* 5: 21, 2018.
12. **Johnson WE, Li C, and Rabinovic A**. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8: 118-127, 2007.
13. **Krishnan A, Samtani R, Dhanantwari P, Lee E, Yamada S, Shiota K, Donofrio MT, Leatherbury L, and Lo CW**. A detailed comparison of mouse and human cardiac development. *Pediatr Res* 76: 500-507, 2014.
14. **Li B, and Dewey CN**. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* 12: 323, 2011.
15. **Lincoln J, Lange AW, and Yutzey KE**. Hearts and bones: shared regulatory mechanisms in heart valve, cartilage, tendon, and bone development. *Dev Biol* 294: 292-302, 2006.
16. **Martin M**. CUTADAPT removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* 17: 10-12, 2011.
17. **Mathieu P, Bouchareb R, and Boulanger MC**. Innate and Adaptive Immunity in Calcific Aortic Valve Disease. *J Immunol Res* 2015: 851945, 2015.
18. **Mohler ER, 3rd, Gannon F, Reynolds C, Zimmerman R, Keane MG, and Kaplan FS**. Bone formation and inflammation in cardiac valves. *Circulation* 103: 1522-1528, 2001.

19. **Morita M, Gravel SP, Hulea L, Larsson O, Pollak M, St-Pierre J, and Topisirovic I**. mTOR coordinates protein synthesis, mitochondrial activity and proliferation. *Cell Cycle* 14: 473-480, 2015.
20. **Niehrs C, and Pollet N**. Synexpression groups in eukaryotes. *Nature* 402: 483-487, 1999.
21. **Nishimura RA, Otto CM, Bonow RO, Carabello BA, Erwin JP, 3rd, Guyton RA, O'Gara PT, Ruiz CE, Skubas NJ, Sorajja P, Sundt TM, 3rd, Thomas JD, Anderson JL, Halperin JL, Albert NM, Bozkurt B, Brindis RG, Creager MA, Curtis LH, DeMets D, Guyton RA, Hochman JS, Kovacs RJ, Ohman EM, Pressler SJ, Sellke FW, Shen WK, Stevenson WG, Yancy CW, American College of C, American College of Cardiology/American Heart A, and American Heart A**. 2014 AHA/ACC guideline for the management of patients with valvular heart disease: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Thorac Cardiovasc Surg* 148: e1-e132, 2014.
22. **O'Brien KD**. Pathogenesis of calcific aortic valve disease: a disease process comes of age (and a good deal more). *Arterioscler Thromb Vasc Biol* 26: 1721-1728, 2006.
23. **Olsson M, Dalsgaard CJ, Haegerstrand A, Rosenqvist M, Ryden L, and Nilsson J**. Accumulation of T lymphocytes and expression of interleukin-2 receptors in nonrheumatic stenotic aortic valves. *J Am Coll Cardiol* 23: 1162-1170, 1994.
24. **Otto CM, Kuusisto J, Reichenbach DD, Gown AM, and O'Brien KD**. Characterization of the early lesion of 'degenerative' valvular aortic stenosis. Histological and immunohistochemical studies. *Circulation* 90: 844-853, 1994.
25. **Otto CM, and Prendergast B**. Aortic-valve stenosis--from patients at risk to severe valve obstruction. *N Engl J Med* 371: 744-756, 2014.
26. **Pearce EL**. Metabolism in T cell activation and differentiation. *Curr Opin Immunol* 22: 314-320, 2010.
27. **Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, and Smyth GK**. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* 43: e47, 2015.
28. **Roberts WC, and Ko JM**. Frequency by decades of unicuspid, bicuspid, and tricuspid aortic valves in adults having isolated aortic valve replacement for aortic stenosis, with or without associated aortic regurgitation. *Circulation* 111: 920-925, 2005.
29. **Schlotter F, Halu A, Goto S, Blaser MC, Body SC, Lee LH, Higashi H, DeLaughter DM, Hutcheson JD, Vyas P, Pham T, Rogers MA, Sharma A, Seidman CE, Loscalzo J, Seidman JG, Aikawa M, Singh SA, and Aikawa E**. Spatiotemporal Multi-Omics Mapping Generates a Molecular Atlas of the Aortic Valve and Reveals Networks Driving Disease. *Circulation* 138: 377-393, 2018.
30. **Siu SC, and Silversides CK**. Bicuspid aortic valve disease. *J Am Coll Cardiol* 55: 2789-2800, 2010.
31. **Small A, Kiss D, Giri J, Anwaruddin S, Siddiqi H, Guerraty M, Chirinos JA, Ferrari G, and Rader DJ**. Biomarkers of Calcific Aortic Valve Disease. *Arterioscler Thromb Vasc Biol* 37: 623-632, 2017.
32. **Steiner I, Krbal L, Rozkos T, Harrer J, and Laco J**. Calcific aortic valve stenosis: Immunohistochemical analysis of inflammatory infiltrate. *Pathol Res Pract* 208: 231-234, 2012.
33. **Winchester R, Wiesendanger M, O'Brien W, Zhang HZ, Maurer MS, Gillam LD, Schwartz A, Marboe C, and Stewart AS**. Circulating activated and effector memory T cells are associated with calcification and clonal expansions in bicuspid and tricuspid valves of calcific aortic stenosis. *J Immunol* 187: 1006-1014, 2011.
34. **Wu HD, Maurer MS, Friedman RA, Marboe CC, Ruiz-Vazquez EM, Ramakrishnan R, Schwartz A, Tilson MD, Stewart AS, and Winchester R**. The lymphocytic

792 infiltration in calcific aortic stenosis predominantly consists of clonally expanded T cells. *J*
793 *Immunol* 178: 5329-5339, 2007.
794 35. **Zhao Y, Nicoll R, He YH, and Henein MY**. The effect of statins on valve function and
795 calcification in aortic stenosis: A meta-analysis. *Atherosclerosis* 246: 318-324, 2016.
796

797 **FIGURE LEGENDS**
798

799 **Figure 1.** Aortic valve collection summary, bioinformatics pipeline and mapping of sample
800 batches and biological conditions in gene expression profile-based principal component analysis
801 (PCA) plots. (*A*) Study outline summarizing aortic valve collection. During the study period,
802 aortic valve leaflets were collected from human fetuses at gestational week (week) 9 (n=4), 12-
803 13 (n=3), and 22 (n=2), adult controls (n=8) and patients with calcified BAV (cBAV; n=5) or
804 TAV (cTAV; n=7). Samples were processed and subject to RNA-seq to characterize their gene
805 expression profiles. Complementary expression data from control, cBAV and cTAV samples
806 was downloaded from GEO (GSE76718). The resulting data set collection was analyzed with a
807 bioinformatics pipeline that included preprocessing of fastq files with cutadapt, alignment
808 against a human transcriptome reference and estimation of gene expression levels with RSEM,
809 and differential expression testing with limma, considering that samples could be ascribed to five
810 experimental batches. Expression profiles were clustered using k-means to obtain a collection of
811 50 clusters, which were then annotated with the Canonical Pathway feature of IPA. Twenty-
812 seven clusters associated to at least one canonical pathway (with Benjamini-Hochberg adjusted
813 p-value < 0.05) were grouped into six meta-clusters on the basis of concordant annotations.
814 Meta-clusters were finally re-annotated with IPA. (*B*) Gene expression profile PCA plot mapping
815 of biological conditions. Sample points are colored according to fetal developmental stage or
816 adult disease state (no disease, n=16; cBAV, n=15 and cTAV, n=16). In PC1, control samples
817 separate from fetal samples and cBAV or cTAV samples. In PC2, early-stage fetal samples
818 (weeks 9 and 13) separate from adult control and disease samples. Percentages on axis labels
819 denote the amount of variance explained by each principal component. (*C*) Relative average
820 expression of 19962 genes with detectable expression, and hierarchical clustering of the sample
821 groups based on Euclidean distance.
822

Figure 2. Concordance analysis of k-means clustering. K-means-based gene clusters (see Supplementary Fig. S1B, see https://figshare.com/articles/figure/SUPPLEMENTARY_DATA_FIGURESMacGrogan_pdf/12086712) were grouped into metaclusters according to the agreement of their enriched canonical pathways. The white to violet squares in the central heatmap represent the percentage (from 0 to 100) of Canonical Pathways (CP) shared between each pair of 50 k-means clusters for which at least one enriched and shared term was found (k=27 out of the original 50 clusters). Hierarchical clustering of the concordance matrix identified six meta-clusters (MC1-6) The heatmap on the left represents the mean gene expression values for each cluster across biological conditions, on a normalized scale (from -2 to 2, for values below and above the mean, respectively). The dendrogram at the top represents functional relatedness between clusters, which was used to group them into meta-clusters. Meta-clusters are identified with colored boxes (MC1 to MC6, in red, blue, green, purple, orange and yellow, respectively) and represent groups of clusters associated to similar functions.

838 **Figure 3**. Functional analysis of meta-clusters. The six gene metaclusters (MC1-6) described in
839 Figure 2 were annotated with IPA. Barplots on the right-hand side represent enrichment values
840 for the top five IPA Canonical Pathways associated with each meta-cluster (MC), selected
841 according to enrichment significance and having an adjusted Benjamini-Hochberg $P<0.05$. Meta-
842 clusters are identified with colored boxes (MC1 to MC6, in red, blue, green, purple, orange and
843 yellow, respectively) and represent groups of clusters associated to similar functions. Central
844 boxes list all the genes associated with the selection of pathways associated with each
845 metacluster. The heatmap on the left represents normalized expression values for individual
846 genes across biological conditions, on a normalized scale (from -2 to 2, for values below and
847 above the mean, respectively).
848

849 **Figure 4.** A gene expression signature predictive of CAVD. (A) Correlation of qPCR data (y-
850 axis) vs aortic-valve calcification levels (x-axis) for the 6 selected genes in whole blood (WB).
851 Each point represents an individual´s level of gene expression (y-axis) and its aortic valve
852 calcification (x-axis). Dots are colored based on the group to which the individual belongs:
853 PESA-CTL (n=52), PESA-SD (n=52) or WB-CAVD (n=12). Solid lines represent the linear
854 model fit for PESA-SD and WB-CAVD patients. Shaded regions represent the confidence of the
855 linear regression estimate for each calcium level. (B) Receiver operating characteristic (ROC)
856 curve of multivariate logistic regression model predicting CAVD (WB-CAVD) vs preclinical
857 aortic-valve calcification (PESA-SD; thick discontinuous line, AUC=0.97) vs valve calcification
858 (PESA-CTL; thick solid line, AUC=0.91). A third ROC curve for the prediction of valve
859 calcification (PESA-SD) vs no calcification (PESA-CTL; thin solid line, AUC=0.71) is shown.
860 The discontinuous diagonal line indicates a fifty-fifty percent chance to predict each group. (C)
861 Follow-up of the model in PESA subgroups after 3 years. Validation in an independent group of
862 CAVD patients (WB-CAVD, n=16) vs a PESA-SD subgroup (n=25; thick discontinuous line,
863 AUC=0.90) vs a PESA-CTL subgroup (n=35; thick solid line, AUC=0.88). A third ROC curve
864 for the prediction of PESA-SD (n=25) vs PESA-CTL (n=35; thin solid line, AUC=0.71) is
865 shown. The thin discontinuous diagonal line indicates a fifty-fifty percent chance to predict each
866 group.
867
868
869

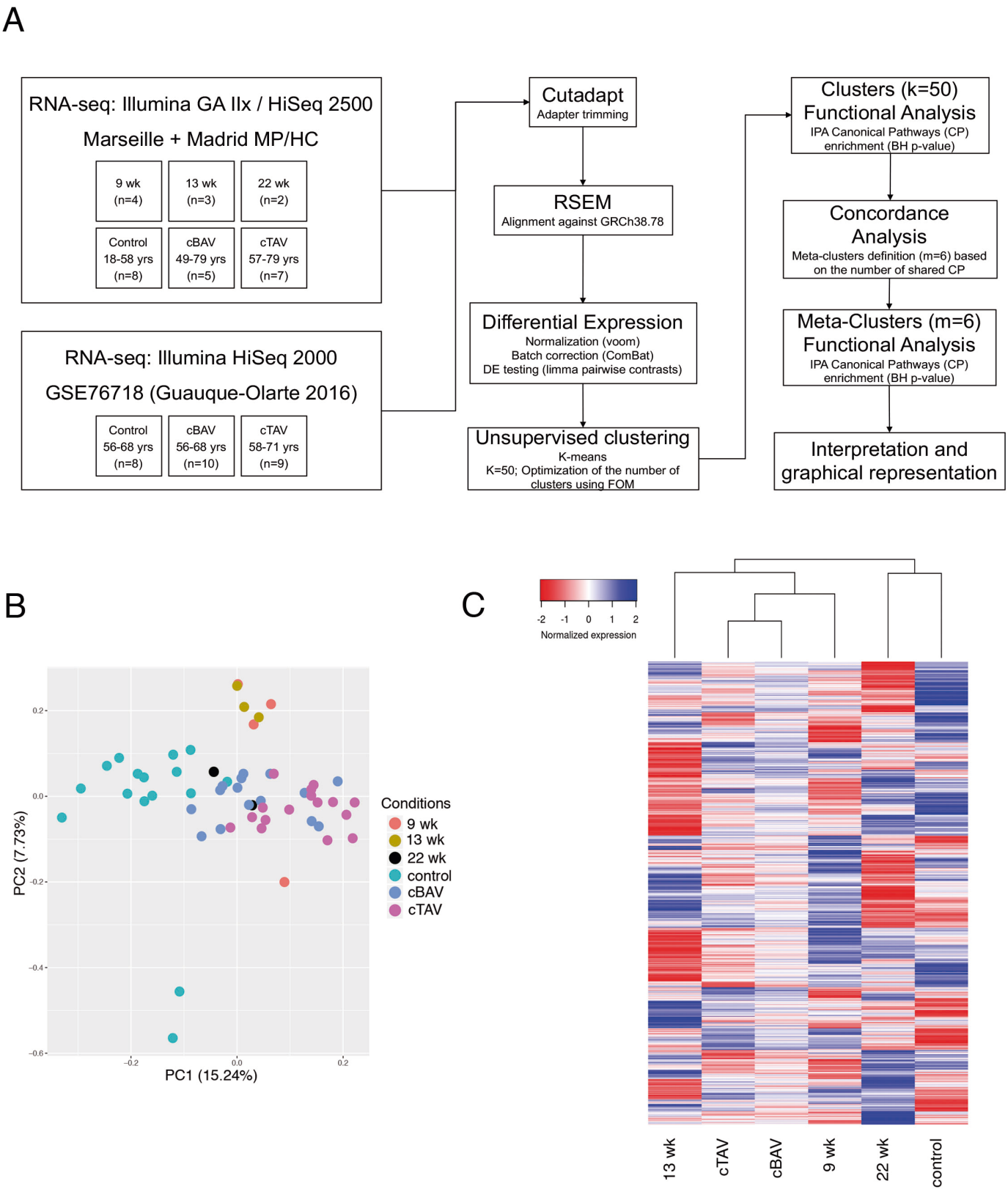

Figure_1_MacGrogan_et_al

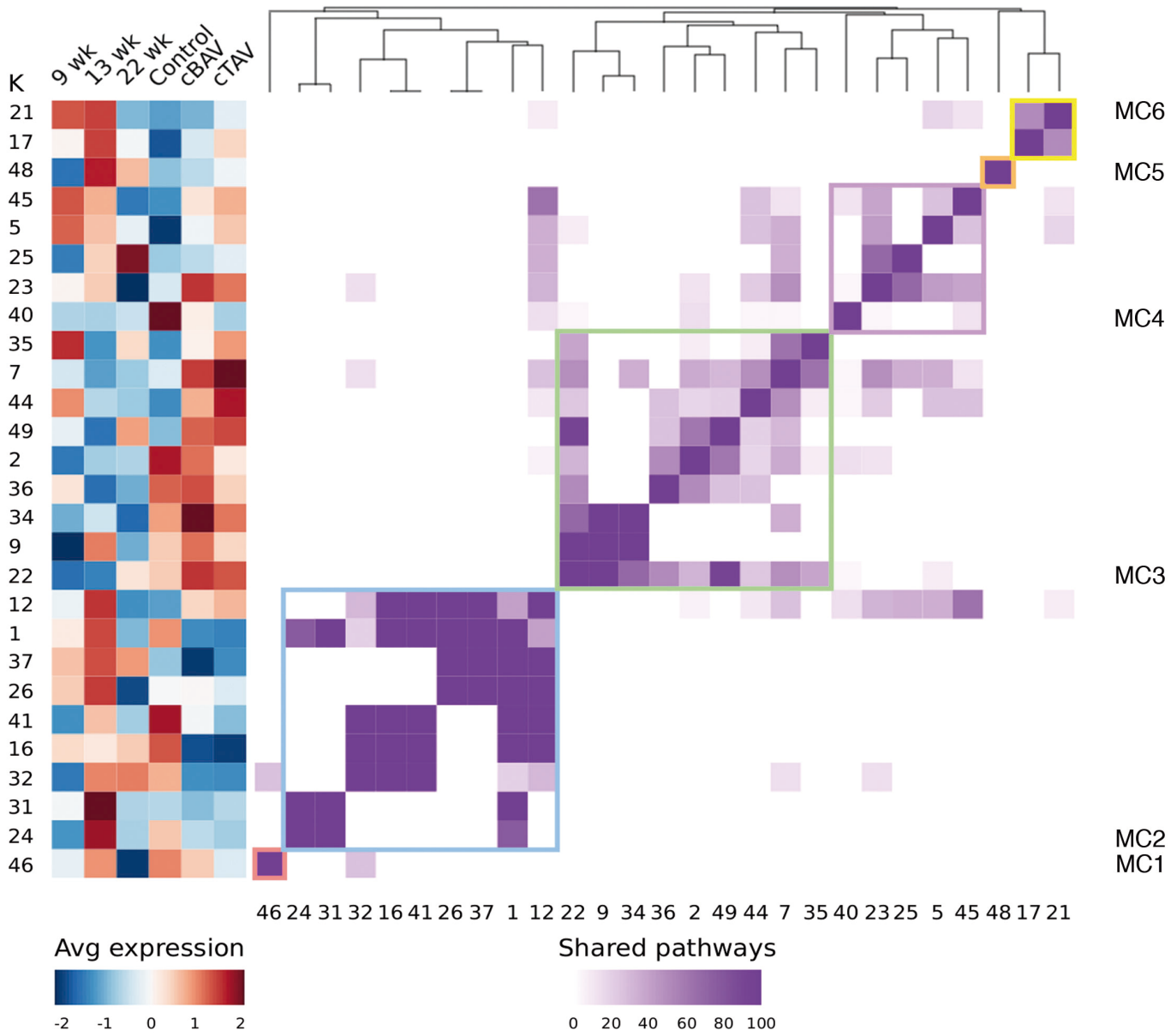

Figure_2_MacGrogan_et_al

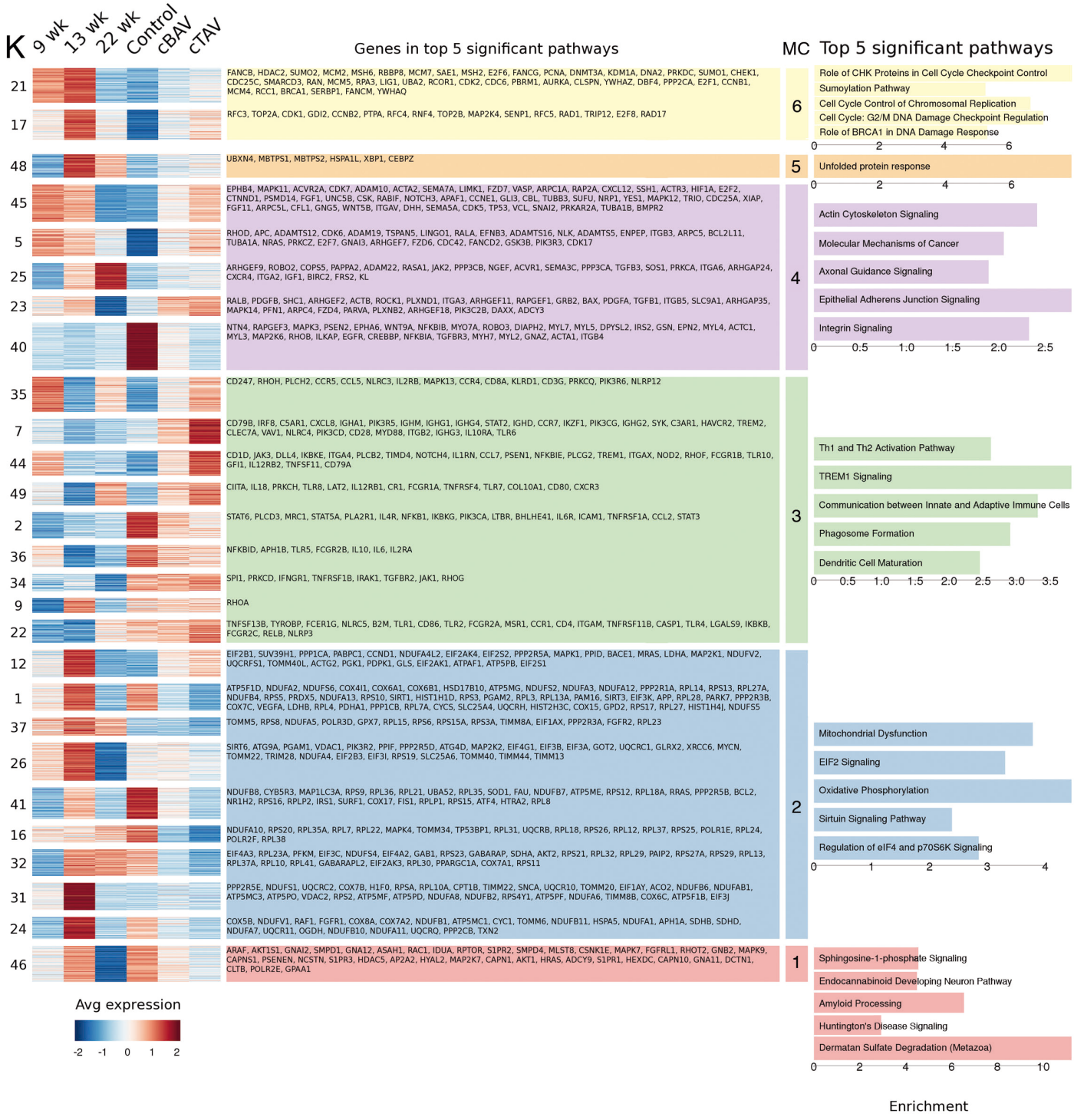

Figure_3_MacGrogan_et_al

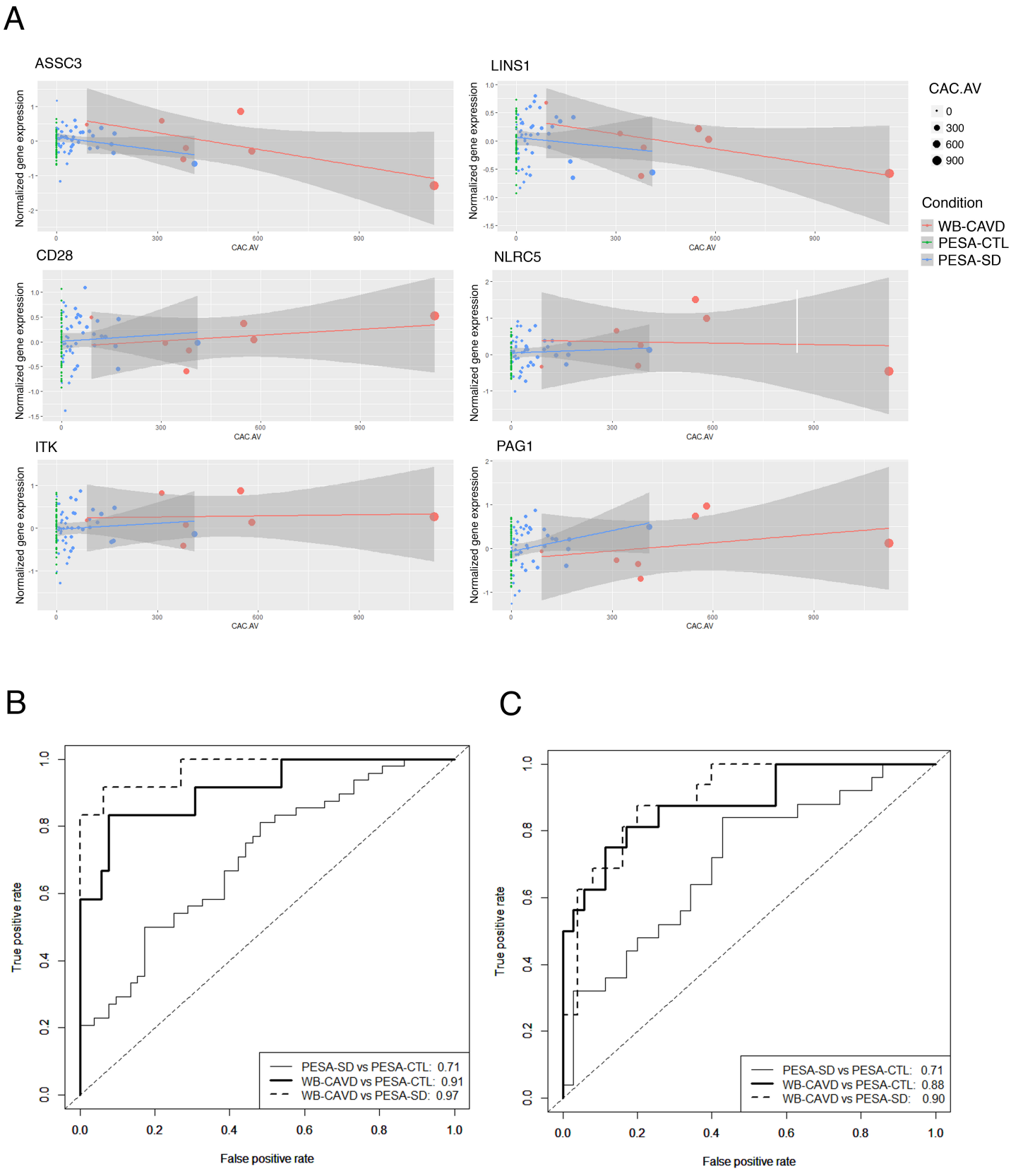

Figure 4_MacGrogan et al