

# Gene-expression patterns reveal underlying biological processes in Kawasaki disease

Stephen J Popper<sup>\*</sup>, Chisato Shimizu<sup>†</sup>, Hiroko Shike<sup>†</sup>, John T Kanegaye<sup>†</sup>, Jane W Newburger<sup>‡</sup>, Robert P Sundel<sup>‡</sup>, Patrick O Brown<sup>§</sup>, Jane C Burns<sup>†</sup> and David A Relman<sup>\*¶</sup>

Addresses: <sup>\*</sup>Departments of Microbiology and Immunology, and Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>†</sup>Department of Pediatrics, University of California San Diego School of Medicine, La Jolla, CA 92093 and Rady Children's Hospital San Diego, San Diego, CA 92123, USA. <sup>‡</sup>Departments of Cardiology and Pediatrics, Children's Hospital Boston, and Harvard Medical School, Boston, MA 02115, USA. <sup>§</sup>Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>¶</sup>Veterans Affairs Palo Alto Health Care System, Palo Alto, CA 94304, USA.

Correspondence: Stephen J Popper. Email: [spopper@stanford.edu](mailto:spopper@stanford.edu)

Published: 11 December 2007

*Genome Biology* 2007, **8**:R261 (doi:10.1186/gb-2007-8-12-r261)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2007/8/12/R261>

Received: 2 April 2007

Revised: 13 July 2007

Accepted: 11 December 2007

© 2007 Popper *et al.*; licensee BioMed Central Ltd.

## Abstract

**Background:** Kawasaki disease (KD) is an acute self-limited vasculitis and the leading cause of acquired heart disease in children in developed countries. No etiologic agent(s) has been identified, and the processes that mediate formation of coronary artery aneurysms and abatement of fever following treatment with intravenous immunoglobulin (IVIG) remain poorly understood.

**Results:** In an initial survey, we used DNA microarrays to examine patterns of gene expression in peripheral whole blood from 20 children with KD; each was sampled during the acute, subacute, and convalescent phases of the illness. Acute KD was characterized by increased relative abundance of gene transcripts associated with innate immune and proinflammatory responses and decreased abundance of transcripts associated with natural killer cells and CD8<sup>+</sup> lymphocytes. There was significant temporal variation in transcript levels during the acute disease phase and stabilization thereafter. We confirmed these temporal patterns in a second cohort of 64 patients, and identified additional inter-individual differences in transcript abundance. Notably, higher levels of transcripts of the gene for carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1) were associated with an increased percentage of unsegmented neutrophils, fewer days of illness, higher levels of C-reactive protein, and subsequent non-response to IVIG; this last association was confirmed by quantitative reverse transcription PCR in a third cohort of 33 patients, and was independent of day of illness.

**Conclusion:** Acute KD is characterized by dynamic and variable gene-expression programs that highlight the importance of neutrophil activation state and apoptosis in KD pathogenesis. Our findings also support the feasibility of extracting biomarkers associated with clinical prognosis from gene-expression profiles of individuals with systemic inflammatory illnesses.

## Background

Over the past decade, genome-wide profiles of the host response to disease have generated important new clues about etiology and opportunities for clinically relevant classification of patients. Among the types of profiles and categories of disease subjected to this approach, DNA microarray-based patterns of gene expression in cancer have been most heavily exploited. Studies in this area have led to patient classification on the basis of molecular pathology and clinical outcome, with benefits to patient care [1-3]. This approach has been far less exploited in the setting of acute infectious diseases and other acute inflammatory conditions in humans [4-7].

In this study, we examined patterns of gene expression in Kawasaki disease (KD), an acute self-limited vasculitis and prominent emerging disease of young children [8]. There are many compelling reasons to focus on KD. First, it is the most common cause of acquired heart disease among children in developed nations. Although the acute symptoms of rash, fever, and mucosal changes resolve spontaneously within 2-3 weeks, the acute vasculitis results in permanent damage to the coronary arteries in 20-25% of untreated patients. Second, an infectious cause is suspected, but despite 30 years of intensive study, the etiologic agent(s) remain(s) unknown. Third, high-dose intravenous immunoglobulin (IVIG) reduces the rate of coronary artery aneurysms (CAA) to 3-5% if administered early in the course of the illness, but without a specific test for the disease, many affected children go undiagnosed and untreated. In addition, IVIG infusion fails to abrogate fever in approximately 10-15% of affected children [9]; the ability to identify children at risk for resistance to IVIG could alert clinicians to institute other therapies.

We postulated that gene-expression patterns would provide insights into KD pathogenesis and might also provide clues about the mechanisms of IVIG response. In this study, we first examined gene-expression patterns in sequential blood

samples from children with KD to characterize the molecular features of the transition from acute to convalescent KD. We then examined samples from a second, larger cohort of KD patients, collected before treatment, to focus on early patterns of gene expression and to search for potential early associations with subsequent CAA formation and IVIG resistance. A molecular marker that emerged from this second cohort in association with IVIG response was evaluated in a third cohort of KD patients.

## Results

### Gene-expression patterns associated with acute Kawasaki disease

We collected whole blood from 20 individuals (cohort 1) at each of three time points during the course of their illness: immediately before treatment (acute samples), 11-24 days post-onset of fever (5-19 days post-IVIG treatment, subacute samples), and more than 24 days post-onset (convalescent samples). Five patients each provided one additional sample at various times: subacute phase (1), convalescent phase (1), and 1 or 2 days after IVIG treatment (3). A total of 65 whole-blood RNA samples were analyzed using DNA microarrays (Table 1, cohort 1); to identify genes whose expression varied during the course of KD, we selected the 1,255 genes (1,594 clones) whose transcript abundance varied at least threefold from the median in two or more of the 65 samples, and organized both genes and samples using hierarchical clustering (Figure 1).

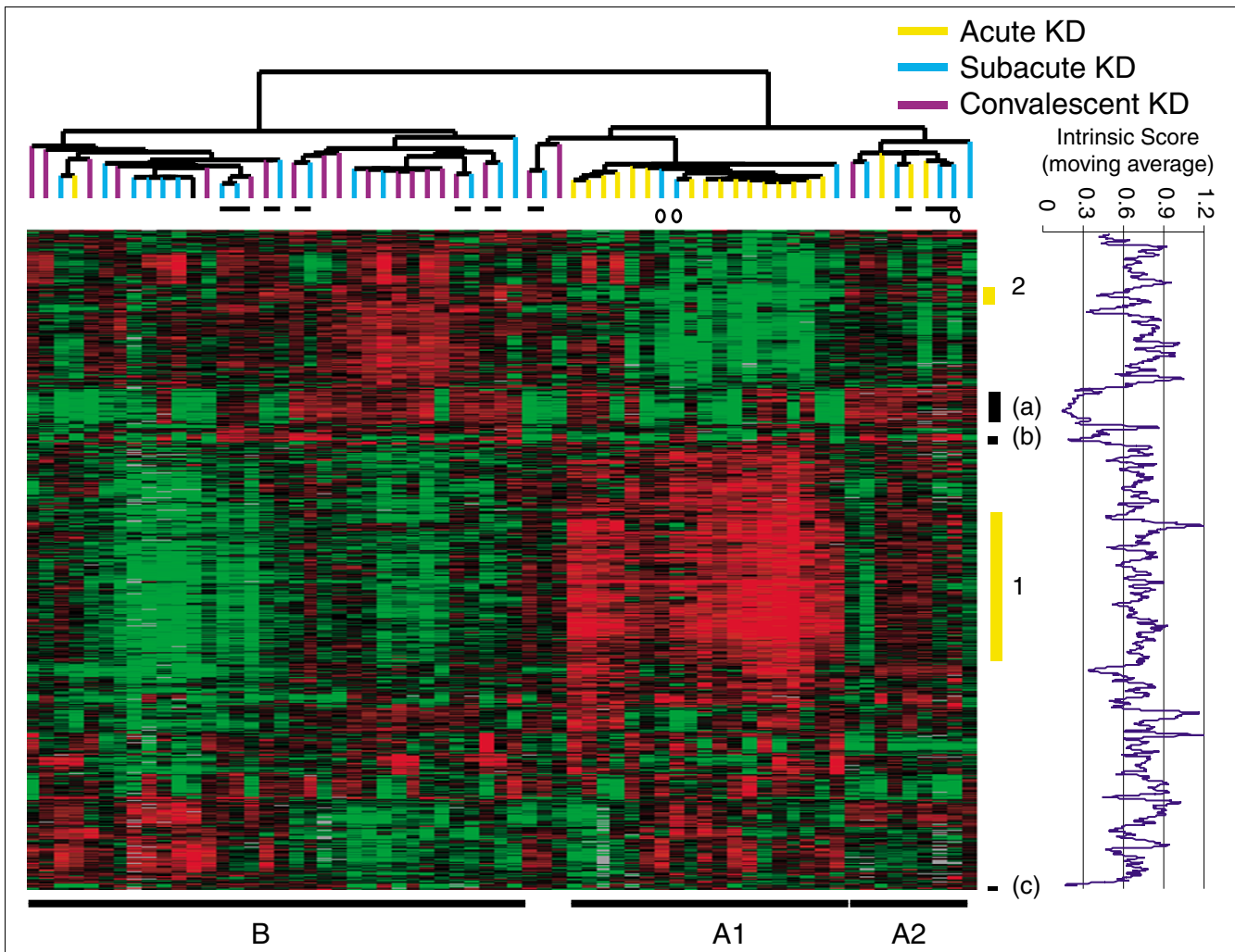
Acute and convalescent samples had distinct patterns of gene expression: 19 of the 20 acute samples in cohort 1 clustered together (clusters A1 and A2, Figure 1), whereas 33 of the 45 subacute and convalescent samples formed the other major cluster (B, Figure 1). Sixteen of the acute samples segregated into a sub-cluster (A1, median illness day 6 (illness day 1 = first day of fever), range 4-12). Three other acute samples segregated in a different sub-cluster with five subacute and

**Table 1**

#### Characteristics of patients

	Cohort 1 (n = 20)*	Cohort 2 (n = 64)*	Cohort 3 (n = 33)*
Sex	14 male, 6 female	44 male, 20 female	19 male, 14 female
Age in years <sup>†</sup>	3.5 (0.4-12.6)	2.1 (0.35-15.6)	1.7 (0.1-4.5)
Coronary artery status <sup>‡</sup>	11 N, 7 D, 2 A	33 N, 22 D, 9 A	26 N, 3 D, 4 A
IVIG response <sup>§</sup>	14 R, 4 NR	45 R, 8 NR	22 R, 11 NR
Illness day, acute sample <sup>†</sup>	Day 6 (4-12)	Day 6 (2-14)	Day 5 (3-7)
Illness day, subacute sample <sup>†</sup>	Day 20 (12-24) <sup>¶</sup>	NA	NA
Illness day, convalescent sample <sup>†</sup>	Day 47 (32-64) <sup>¶</sup>	NA	NA

\*Seven individuals were in both cohort 1 and cohort 2; three additional individuals were in both cohort 1 and cohort 3. <sup>†</sup>Data presented as median (range). <sup>‡</sup>Arterial diameter determined by echocardiogram. N, normal; D, dilated; A, aneurysm. <sup>§</sup>R, IVIG responder; NR, IVIG non-responder. Two patients in cohort 1 and 11 in cohort 2 were treated with IVIG after illness day 10 and excluded from analysis of IVIG response. <sup>¶</sup>Additional samples were obtained from three patients less than 48 h after treatment with IVIG on illness days 7, 8, and 13; one additional sample was obtained from a fourth patient at a 1-year follow-up. NA, not applicable.



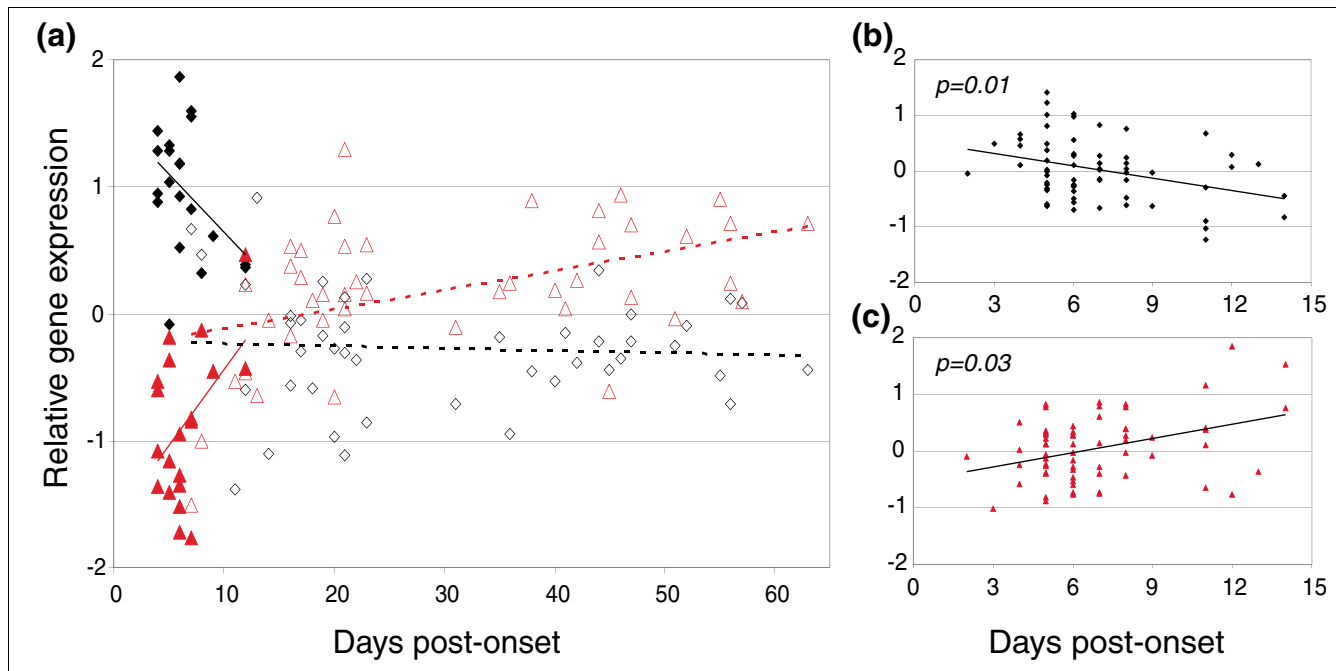
**Figure 1**

Patterns of transcript abundance in whole-blood samples from children with KD (cohort 1). Genes and samples were organized using hierarchical clustering; each row represents a single gene, and each column a single sample. Black indicates the median level of expression, red indicates greater expression than the median, green less expression, and gray missing data. Horizontal bars under the sample dendrogram at the top indicate samples from the same individual that cluster together; open circles represent samples obtained 1 or 2 days after IVIG treatment. Pretreatment acute KD samples are in yellow; early post-treatment subacute samples in light blue; late post-treatment convalescent samples in purple. Horizontal bars at the bottom of the figure indicate the sample groups mentioned in the text. A1 and A2 are clusters of acute samples; B is a cluster of subacute and convalescent samples. Vertical yellow bars (1 and 2) on the right mark the two gene clusters most strongly associated with the grouping of the pretreatment samples. The intrinsic score was graphed as a moving average of nine genes along the y-axis. The black vertical bars on the right lettered a, b, and c mark the three gene clusters with intrinsic scores more than 2 standard deviations away from the mean score of 0.742.

one convalescent sample (A2, median illness day 15, range 8-24); these samples were collected from three individuals diagnosed and sampled more than 7 days after the onset of fever.

Two sets of genes (see Figure 1, gene clusters 1 and 2, vertical yellow bars) with an average expression correlation coefficient of greater than 0.8 were significantly associated with the segregation of samples into clusters A and B (clusters 1 and 2:  $p = 0.005$  and  $0.01$ , respectively). Average transcript levels in cluster 1 were high early in the illness and declined as a function of illness day, whereas cluster 2 transcripts were less

abundant early in the illness and increased as a function of illness day, although these trends did not achieve statistical significance in this group of specimens (Figure 2a,  $p = 0.08$  for both clusters). Both sets of genes were stable over time in the convalescent phase. The 245 genes (354 clones) in cluster 1 included many whose expression has been associated with neutrophils and inflammatory processes, including adrenomedullin (*ADM*), B-cell lymphoma 6 (*BCL6*), colony stimulating factor 2 receptor, beta (*CSF2RB*), formyl peptide receptor 1 (*FPR1*), grancalcin (*GCN*), granulins (*GRN*), interferon-induced transmembrane protein 2 (*IFITM2*), IL1 receptor 2 (*IL1R2*) and IL1 receptor antagonist (*IL1RN*), matrix metallo-



**Figure 2**  
 Patterns of transcript abundance as a function of illness day (days post-onset of fever). **(a)** Average transcript abundance of the two gene clusters associated with segregation of the acute and convalescent samples in cohort 1 (yellow bars in Figure 1). Closed symbols indicate pretreatment samples; open symbols indicate post-treatment samples. Black squares and red triangles represent transcripts expressed at higher (cluster 1) and lower (cluster 2) levels in acute KD, respectively. Solid (acute) and dashed (subacute and convalescent) black lines and red lines show the trend in expression levels of the genes in cluster 1 and cluster 2, respectively. **(b,c)** Average transcript abundance of (b) gene cluster 1 and (c) gene cluster 2 in cohort 2. The significance of the trend of expression over time in cohort 2 ( $p = 0.01$  in (b), and  $p = 0.03$  in (c)) was determined using Spearman rank correlations.

proteinase-9 (*MMP9*), pre-B cell colony-enhancing factor 1 (*PBEF1*), and the S100 calcium binding protein A11 (*S100A11*). The 10 genes in cluster 2 included those for the T-cell receptor gamma subunit (*TRG*), killer cell lectin-like receptor subfamily G, member 1 (*KLRG1*), the IL2 receptor beta subunit (*IL2RB*), and V-myb myeloblastosis viral oncogene homolog (avian)-like 1 (*MYBL1*) - all genes preferentially expressed in CD8 T cells and NK cells [10,11] (see Additional data file 1 for a complete list of the genes in clusters 1 and 2).

**Individuality of gene expression in KD patients**

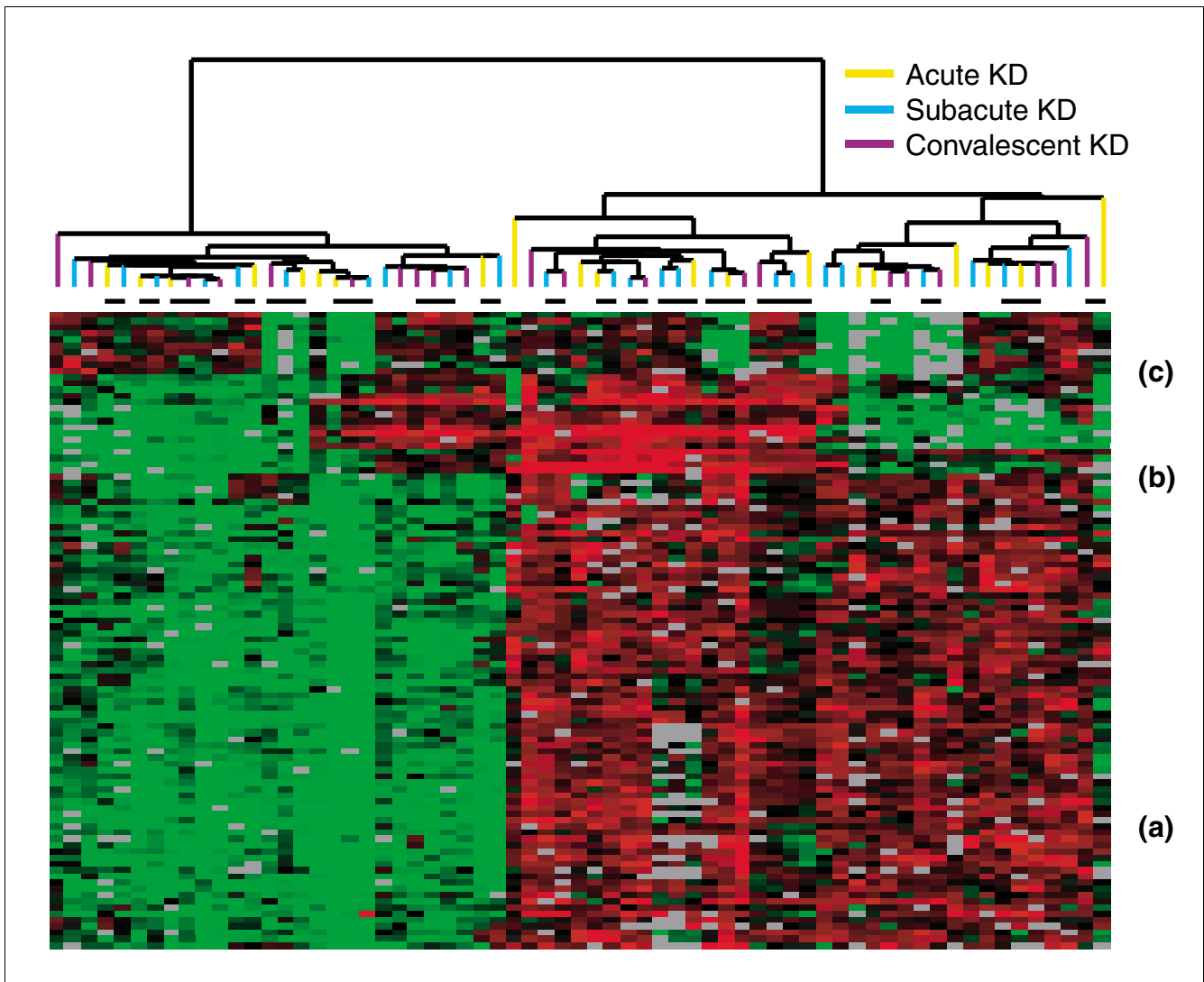
There was no evidence of sample clustering by day of illness among the convalescent samples. However, we observed pairing of subacute and convalescent samples from six of the 20 individuals. This pattern suggested the presence of intrinsic differences in gene expression of individuals that might reflect underlying DNA polymorphisms [7]. We calculated an 'intrinsic' score for each gene, based on the ratio of within-individual to between-individual variation among the subacute and convalescent samples. Three gene clusters had an average intrinsic score or more than two standard deviations from the mean score of all genes in the original dataset that met our quality criteria (see Figure 1, gene clusters a-c; see Additional data file 2 for a complete list of the genes). Two of these intrinsic gene sets were composed of HLA loci (both

class I HLA-B and class II HLA-DR and DQ) and Y-linked genes that were expressed at higher levels in the male patients (see Figure 1, gene clusters b and c, respectively). The third cluster (see Figure 1, gene cluster a) was composed of 65 genes (78 clones), including those that encode proteins involved in antigen processing (HLA-DOA and the immunoglobulin heavy-chain mu locus) and regulation of the immune response (cytotoxic T-lymphocyte associated protein 4, the Ets transcription factor Spi-B, and IL24).

When we re-clustered the 65 samples using these three gene sets, specimens from all three time periods (acute, subacute, and convalescent) from each of six subjects clustered together. For 14 of 20 subjects, the acute sample was most similar to a subacute or convalescent sample from the same individual, indicating that expression of the most person-intrinsic genes, including those involved in antigen presentation, did not differ in the acute and convalescent phases of the disease (Figure 3).

**Variation in gene expression in acute KD and correlations with clinical parameters**

Cohort 1 provided an overview of gene-expression patterns in KD, from the acute phase through to convalescence. The observed patterns emphasized the differences between early and late phases of the disease and suggested that gene expres-



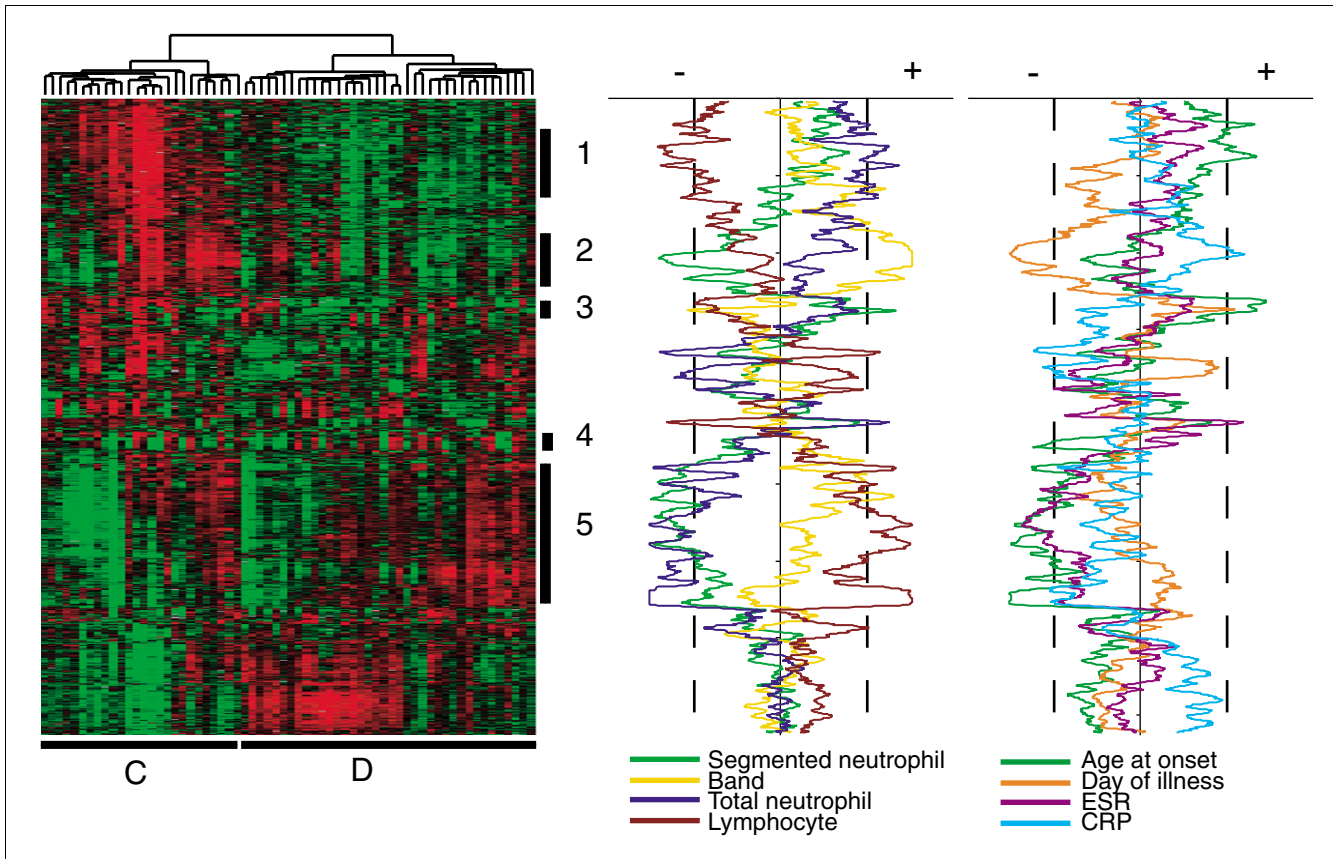
**Figure 3**  
Clustering of KD samples on the basis of intrinsic gene-expression patterns. The 65 samples from 20 patients in Figure 1 were reorganized using the three gene clusters identified as having the lowest (strongest) intrinsic scores. a, b, and c indicate the same gene clusters as in Figure 1. Colors are as in Figure 1. Horizontal black bars indicate samples from the same patient that clustered together.

sion in the early phase is highly dynamic. To characterize early events that shape the pathogenic process, we examined gene-expression patterns in a larger set of blood samples, collected before treatment (acute phase) from 64 patients (Table 1, cohort 2). This second set included seven acute-phase samples from cohort 1 that were reprocessed in parallel with acute-phase samples from 57 new patients; comparison of the original and re-derived expression profiles from the seven patients demonstrated reproducibility of these microarray data (Additional data file 3).

The relative transcript levels of the 1,241 most variably expressed genes (more than threefold variation from the median, in three or more of the 64 samples: 1,652 clones) are displayed in Figure 4. In contrast to the acute samples in

cohort 1, the acute samples in cohort 2 did not segregate by day of illness. Instead, the two main sample clusters (Figure 4, clusters C and D) were most strongly associated with differences in lymphocyte percentage and the age of the patient at onset of illness. Patients in cluster C had a lower percentage of lymphocytes and tended to be older than those in cluster D ( $p = 0.05$  and  $0.06$ , respectively, rank-sum test).

To identify gene-expression patterns associated with the distinctions between clusters C and D and to elucidate the molecular processes underlying clinical parameters used to assess and predict the course of KD, we calculated the correlation coefficient of the expression levels of each gene with each available demographic and clinical laboratory



**Figure 4**  
 Association of transcript abundance and clinical parameters in patients with acute KD (cohort 2). **(a)** Genes and samples are organized using hierarchical clustering as in Figure 1. Sample clusters C and D are explained in the text. Numbered vertical bars at the right of the heat map indicate gene sets described in the text and in Additional data file 4. **(b)** Correlation coefficients were calculated for the expression of each gene and each of the following clinical parameters: percentage segmented neutrophils, percentage unsegmented neutrophils (band), percentage total neutrophils, percentage lymphocytes, age at onset, illness day, erythrocyte sedimentation rate (ESR), and level of C-reactive protein (CRP). A *p* value was calculated using permutation, and assigned a negative or positive value corresponding to the direction of the correlation. Results are portrayed as a moving average along the *y*-axis, with a window size of 15 clones, and were plotted for a given parameter if the *p* value was less than 0.01 (marked by dotted line) over 10 or more consecutive array elements.

parameter. Eight parameters showed particularly strong associations ( $p < 0.01$ ) with subsets of genes analyzed in cohort 2 (see Figure 4).

Two large sets of genes were associated with differences in the relative abundance of the two major blood cell subpopulations, neutrophils (234 genes, 321 clones; cluster 1 in Figure 4) and lymphocytes (319 genes, 387 clones; cluster 5 in Figure 4). The neutrophil-associated gene set was also positively correlated with age at onset of illness, and contained genes involved in the innate immune response, including those for CD14, Toll-like receptors 1, 2, 4, 6 and 8, and Fc-receptor II subunits A and B. Other genes in the cluster included those for proteins involved in cell adhesion (PECAM1, CEACAM4, CD36 and metalloproteases ADAM8, ADAM19, and MMP9), leukocyte chemotaxis (FPR1, HCK, PF4, SELP, SELL, VNN2 and VNN3), and genes encoding proinflammatory cytokines (IL1A and B, S100A12, and

ADM). Names and symbols of both sets of genes, as well as those discussed below, are listed in Additional data file 4.

*Age-associated patterns of expression*

The gene clusters whose expression varied with the relative abundance of neutrophils and lymphocytes were also associated with the age of the patient ( $r = 0.59$  and  $-0.56$ , respectively). Two other sets of genes (clusters 3 and 4) were strongly associated with age but not with any of the other measured parameters. The 24 genes (42 clones) in cluster 3 (see Figure 4) were more highly expressed in older children, and included membrane metalloendopeptidase (*MME/CD10*) and protease inhibitor 3 (*PI3/SKALP*), as well as a number of antimicrobial peptides (see Additional data file 4). Cluster 4 (see Figure 4) consisted of 11 genes more highly expressed in younger children, and included the immunoglobulin lambda locus (*IGL*), the gene for the pre B-cell receptor light chain subunit (*IgLL1*), as well as the gene for

pro-apoptotic caspase adaptor protein (*PACAP*), an anti-inflammatory protein that negatively regulates *IL6*.

#### Patterns related to day of illness

Many of the genes that distinguished acute and convalescent KD in cohort 1 appeared to have time-dependent patterns of expression within the acute phase of the disease (see Figure 2a). With the larger set of samples in cohort 2, we confirmed these associations using DNA microarray-based data (see Figure 2b,c). Genes highly expressed in acute KD in cohort 1 also comprised most (54 of 60) of those found in the day-related cluster identified in cohort 2 (cluster 2, Figure 4). These 60 day-correlated genes were also associated with a higher proportion of unsegmented neutrophils (bands) in the blood; regression analysis indicated that average transcript abundance for this set of genes was independently associated with both illness day and band percentage ( $p < 0.001$  for each parameter). Among the 60 genes were a number that have been implicated in regulation of apoptosis (*BCL2A1*, *BIRC1*, *IKBK*, *SFRP1*, and *SH3GLB1*), as well as others associated with proinflammatory processes (*ANXA3*, *HP*, *IL13*, *IL18R1*, *CEACAM1*). Kobayashi *et al.* [12] recently demonstrated that transcriptional programs leading to inhibition of apoptosis and an increase in expression of proinflammatory genes are features of neutrophils exposed to granulocyte-macrophage colony stimulating factor (GM-CSF). Of the 60 genes, 23 were among those upregulated in response to GM-CSF, compared with 123 of the 1,241 genes in the larger set of most variable genes ( $p < 10^{-8}$ , hypergeometric).

#### Genes associated with response to IVIG and coronary artery outcome

Echocardiograms of the 64 patients in cohort 2 revealed normal coronary arteries in 33, dilatation in 22 and CAA in nine. Eight of the 53 patients treated on or before illness day 10 did not defervesce within 48 hours following treatment with IVIG, and were classified as non-responders (Table 1).

We looked for potential prognostic biomarkers associated with response to IVIG and with coronary artery outcome using significance analysis of microarrays (SAM) [13]. We did not identify any genes whose expression differed significantly using the criterion of coronary artery outcome. However, we identified five genes whose level of expression was associated with a non-response to IVIG (median false-discovery rate 20%): these were the genes for FK506 binding protein 5 (*FKBP5*), beta-actin (*ACTB*), carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein, *CEACAM1/CD66*), vesicle-associated membrane protein 5 (*VAMP5*), and polo-like kinase 2 (*PLK2*, or serum-inducible kinase *SNK*). We performed quantitative reverse transcription PCR (RT-PCR) for all five genes using 11 RNA samples that had also been analyzed with microarrays (six responders and five non-responders). *CEACAM1* and *FKBP5* showed a significant association with treatment response on the basis

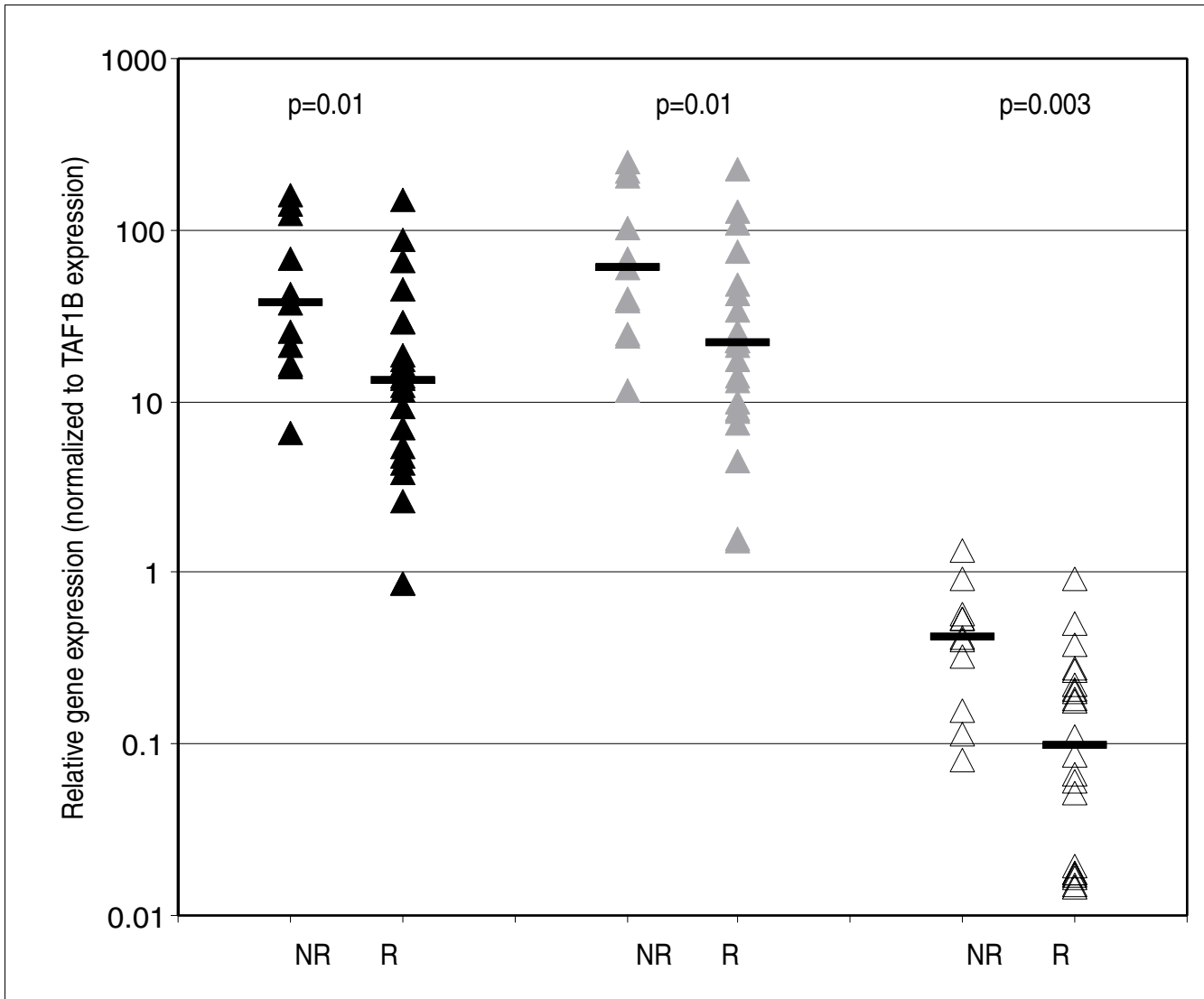
of RT-PCR data, but only *CEACAM1* showed a strong correlation with the array results (data not shown).

To confirm this finding for *CEACAM1* in an independent group of patients, we measured *CEACAM1* mRNA levels in pre-treatment samples from 11 additional IVIG non-responders and 22 responders, matched for illness day and age (cohort 3, Table 1). *CEACAM1* was again expressed at significantly higher levels in the non-responders (Figure 5). The *CEACAM1* transcript is alternatively spliced; *CEACAM1* mRNA exists either in a long (L) form with a cytoplasmic tail that contains immunoreceptor tyrosine-based inhibitory motifs (ITIMs), or a short (S) form in which the tail is truncated (reviewed in [14]). The two forms have different signaling properties, and occur at different ratios in different cell types. To determine if both isoforms were associated with IVIG non-response, we measured levels of each of these transcripts separately, using specific quantitative RT-PCR assays; both transcripts were found at higher levels in the patients who subsequently failed to respond to IVIG treatment (see Figure 5).

#### Discussion

Our unsupervised analysis of expression profiles from sequential samples of KD patients (cohort 1) revealed segregation of acute and convalescent samples, driven by higher levels of gene transcripts associated with neutrophils and innate immune responses during the acute phase of the illness. The subacute and convalescent phases of KD were characterized by increased abundance of transcripts associated with CD8 T cells and NK cells. These patterns of expression in part reflect previously observed fluctuations in leukocyte subpopulations during the course of KD; expansion of the neutrophil population is a characteristic feature of acute KD, and previous studies have reported an elevated CD4:CD8 ratio in the peripheral blood during the acute phase, and an increase in both CD8 T cells and CD16 NK cells in the convalescent phase [15-17]. Our findings are consistent with the hypothesis that CD8 T cells and NK cells are sequestered in the arterial walls or bound to endothelial cells during acute KD, and return to the circulation later in convalescence [15,16]. All patients were treated with IVIG after the acute samples were obtained; it is therefore possible that the subsequent shifts in transcript abundance were due to the direct effects of IVIG on cellular processes. However, three of four pre-treatment acute samples obtained after illness day 7 clustered with subacute samples, and the two post-treatment samples obtained before illness day 10 clustered with acute samples. This suggests that the changes in the levels of most transcripts associated with acute KD were more closely associated with the natural history of the disease than with the effects of IVIG.

These acute disease-associated gene-expression patterns were superimposed on underlying person-specific features that did not vary significantly throughout the course of KD.



**Figure 5**  
Levels of *CEACAM1* mRNA in pretreatment whole-blood samples from KD patients who subsequently responded (R) or failed to respond (NR) to IVIG treatment. Transcript levels were measured by quantitative RT-PCR, and normalized to *TAF1B* transcript levels. Black triangles, assay for both long (L) and short (S) splice forms of *CEACAM1*; gray triangles, L-form; open triangles: S-form. Horizontal bars indicate median relative expression level.

Transcript abundance for genes identified as 'intrinsic' to an individual were not associated with treatment response or development of aneurysms, but may be linked to specific genetic polymorphisms, which are believed to play an important role in determining susceptibility to KD [8,18].

Two previous studies examined genome-wide gene-expression patterns in acute and convalescent samples from four patients [19,20]. Both reported increased transcript abundance in acute KD for a number of the genes we identified in our study, including *ADM* and *S100A12*. *ADM* encodes a vasodilatory peptide, and *S100A12* contributes to proinflammatory cascades; both gene products are known to be elevated in the serum of KD patients [19,21]. While previous studies focused on the increased expression of these genes in

monocytes, they are also expressed at high levels in granulocytes [11]. Given the abundance of neutrophils in acute KD, and the correlation of these genes with neutrophil count in our study, it is likely that much of the expression of *ADM* and *S100A12* in acute KD can be attributed to neutrophils.

By sampling whole blood, we were able to identify the substantial contribution of neutrophils to the overall transcriptional program associated with KD; the sampling of whole blood also offers important practical advantages in the clinical workplace. Yet the use of mixed cell populations also limited our ability to evaluate the relative importance and activation state of specific cell types [4]. The differences in neutrophil and lymphocyte abundance in the sequential blood samples from KD patients (cohort 1) accounted for



approximately half of the variation in average expression of the genes in clusters 1 and 2 ( $r^2 = 0.53$  and  $0.50$ , respectively), suggesting that differences in activation, or other sources of heterogeneity, also play a role in the observed variation in transcript abundance. Further studies of transcript patterns in purified neutrophils and lymphocyte populations from KD patients will help clarify the role of activation states for specific cell types in the different phases of the illness.

Because our study included many more patients and samples than in previous studies, we were able to examine patterns of gene expression at much higher resolution and able to discern inter-patient variation within the acute phase of the disease. The sets of neutrophil- and CD8 T cell- or NK-associated genes varied with day of illness, emphasizing the dynamic and variable nature of the acute phase of the disease. Gene expression during the acute phase of KD also correlated with fluctuations in the major leukocyte populations, age, and concentrations of C-reactive protein (CRP). All these parameters have been associated with the risk of coronary artery damage [22-24]. The sets of correlated genes provide molecular signatures for these clinical features, and may also provide insight into physiological properties of the immune system. For example, the increased abundance of immunoglobulin  $\lambda$  chain transcripts in younger children with KD mirrors the inverse correlation we previously reported between  $\lambda$  expression and age among healthy adults [7], suggesting that this may be a more general phenomenon. Conversely, a set of genes primarily expressed in neutrophils was positively correlated with age. These genes include several that encode antimicrobial peptides, MME, and PI3, and may reflect differences in neutrophil function as children age. Levels of a number of neutrophil surface proteins, including MME, are known to change with age, and changes in neutrophil function have been identified in elderly individuals, but have not been examined in younger age groups [25,26]. This study was limited to patients with KD; studies that include age-matched healthy and febrile controls would clarify whether the observed associations of gene expression and clinical parameters are unique to KD.

The genes that vary in expression during acute KD also provide leads for identifying biomarkers associated with clinical outcome and with biological processes underlying KD pathology. Of particular note, increased abundance of *CEACAM1* transcripts was associated with subsequent non-response to IVIG in two independent sets of patients. The clinical utility of *CEACAM1* transcript levels for predicting IVIG non-response remains to be determined; there was a 200-fold range in the level of *CEACAM1* transcript, but our sample size was too small to obtain a precise estimate of the risk associated with a 10-fold increase in transcript abundance (odds ratio = 5.9, 95% CI = 1.2-30.5). The results of previous studies that have attempted to build a decision instrument for predicting IVIG response suggest that no single marker is likely to suffice [27]; the overlap in *CEACAM1*

expression for IVIG responders and non-responders suggests it may be most useful when combined with other markers (see Figure 4).

*CEACAM1* was among the genes whose expression was correlated with day of illness and the relative abundance of unsegmented neutrophils, and *CEACAM1* expression levels were also correlated with levels of CRP. Fewer days of illness at diagnosis, an increase in unsegmented neutrophils, and higher levels of CRP have all been associated with non-response to IVIG in multiple studies [27-30]; thus, *CEACAM1* may provide clues about a common biological process underlying treatment response. The *CEACAM1* protein is a surface molecule expressed on a wide variety of cells, but at particularly high levels on neutrophils, and increases in *CEACAM1* expression have been associated with delayed apoptosis in both human and rat neutrophils [12,31]. Neutrophil apoptosis is an important feature in the resolution of inflammatory processes; the association of *CEACAM1* transcript abundance with IVIG response in this study supports the idea that neutrophil apoptosis may be an important factor in the mechanism of action of IVIG in KD patients. Two studies reported that immunoglobulin induces apoptosis *in vitro* in neutrophils obtained from pre-treatment KD patients, but not from healthy controls [32,33]. One of these studies also found a significant association between the fraction of neutrophils undergoing apoptosis and defervescence following IVIG treatment. Further studies will be required to determine whether neutrophils in KD patients expressing *CEACAM1* are less susceptible to apoptosis following treatment with IVIG and whether *CEACAM1* mediates this process.

## Conclusion

KD has been a leading cause of acquired heart disease in children for more than 30 years, but the pathogenic mechanisms remain obscure. Our goal was to build a comprehensive portrait of the host response in KD by examining genome-wide transcript abundance patterns in whole blood. The results emphasize two features: first, the dynamic and variable nature of the acute phase of the disease, and second, the role of neutrophils and their activation programs. Variation in transcript abundance was associated not only with day of illness, but also with clinical parameters that have been used in the management of KD and in efforts to predict the course of disease. The sets of correlated genes provide molecular signatures for these clinical features and help to identify biological processes with a role in KD pathogenesis. Higher *CEACAM1* transcript levels were also associated with subsequent non-response to IVIG treatment, suggesting the feasibility of identifying prognostic markers for an infectious or acute inflammatory process on the basis of host gene-expression profiles.

## Materials and methods

### Study population and sample collection

KD patients ( $n = 107$ ) were enrolled at two clinical sites (Rady Children's Hospital, San Diego, CA and Children's Hospital Boston, MA) after parental informed consent. All patients had fever and four or more of the five principal clinical criteria for KD (rash, conjunctival injection, cervical lymphadenopathy, changes in the oral mucosa, and changes in the extremities) or three criteria plus coronary artery abnormalities documented by echocardiography [34]. The human subjects protocol was reviewed and approved by the Institutional Review Boards at UCSD, Children's Hospital Boston, and Stanford University. Clinical data including gender, ethnicity, race, age, day of illness (first day of fever = illness day 1), results of laboratory testing, response to IVIG therapy, and coronary artery status were recorded for all subjects. IVIG non-response was defined as persistent or recrudescent fever ( $T \geq 100.4^\circ\text{C}$  F rectally or orally) 48 h following completion of the IVIG infusion (2 g/kg). The mean duration of fever in untreated KD patients is 11 days; patients treated more than 10 days after the onset of fever have therefore been excluded from trials of IVIG efficacy, and were excluded from the analysis of IVIG response in this study [34]. The internal diameters of the right coronary and left anterior descending arteries were classified by echocardiography as normal ( $< 2.5$  standard deviations ( $z$  score) from the mean, normalized for body surface area [35]), dilated ( $2.5 \leq z$  score  $< 4.0$ ), or aneurysmal (saccular or fusiform dilatation of a coronary artery segment with  $z$  score  $\geq 4.0$ ).

Blood was obtained for determination of complete blood count and differential, CRP, erythrocyte sedimentation rate (ESR), and gamma glutamyltranspeptidase (GGT), and for RNA studies (Paxgene Blood RNA System, PreanalytiX GmbH, 8634 Hombrechtikon, CH). PAXgene tubes were stored at  $4^\circ\text{C}$  for no more than 5 days and RNA was purified according to the manufacturer's instructions.

### DNA microarray hybridization

RNA transcripts in the samples and a standard reference RNA (Universal Human Reference RNA, Stratagene, La Jolla, CA) were amplified using the MessageAmp aRNA amplification kit (Ambion, Austin, TX). Sample and reference transcripts were then reverse-transcribed, labeled with fluorescent dyes (Cy5 and Cy3, respectively), mixed together, and hybridized to cDNA microarrays as previously described [36,37]. The arrays used for these studies contain 37,632 spots derived from cDNA clones representing approximately 18,000 unique human genes [2]. Images of hybridized arrays were obtained using a Genepix 4000B microarray scanner, and analyzed with Genepix 5.0 (Axon Instruments, Union City, CA). The data were submitted to the Stanford Microarray Database [38].

### Microarray data filtering and analysis

Data were filtered to include only clones that met the following criteria for at least 80% of the samples tested: signal intensity 2.5-fold above background in either the Cy5 (sample) or Cy3 (reference) channel, and a regression correlation for the two channels of at least 0.6 across each measured element. A normalization factor was applied so that the mean  $\log_2$  ratio for each array (sample) was zero, and data for each clone were then median-centered across all observations. Selected data were organized using a hierarchical clustering algorithm based on a Pearson correlation metric, with average linkage clustering [39], and visualized using Java Treeview [40].

The EASE software program was used to identify Gene Ontology (GO) categories enriched in specified subsets of the data [41]. An EASE score was calculated for each GO term represented in the dataset; EASE scores represent an adjustment to the Fisher exact probability that penalizes terms containing only a few genes. A global false-discovery rate was calculated to adjust for multiple testing, and to generate the reported  $p$  values. Enrichment within gene clusters for genes found in other gene expression datasets was determined using the hypergeometric distribution.

Intrinsic scores were calculated as previously described [7]. For each gene, the ratio of the mean squared pairwise difference in transcript levels among multiple samples from a single individual to the mean squared pairwise difference among individuals. The mean intrinsic score in cohort 1, for the 20,361 clones (12,186 genes) well measured in at least 80% of the subacute and convalescent samples, was 0.742, with a standard deviation of 0.473.

The association of expression levels of a given gene and a given clinical parameter was defined by the Pearson correlation coefficient. Parameter values were then randomly permuted 1,000 times among the patients, while the expression data were held constant. The fraction of times the permuted data yielded a correlation coefficient greater than that obtained with the true data was used to calculate a  $p$  value. The negative logarithm (base 10) of the  $p$  value, with the sign of the original coefficient to indicate direction, represented this association graphically.

The association of laboratory parameters with the occurrence of either a coronary artery lesion or failure to respond to IVIG therapy was assessed using a Student's  $t$ -test; CRP levels were log-transformed to approximate a normal distribution more closely. SAM was used to identify genes significantly associated with treatment response [13]. Linear regression models and Spearman rank correlations were tested using STATA v.7 (STATA, College Station, TX).

### Quantitative RT-PCR

Levels of *CEACAM1*, *FKBP5*, *ACTB*, *VAMP5*, and *PLK2* mRNA were measured using TaqMan 5'-nuclease Gene Expression Assays (Applied Biosystems, Foster City, CA). Three different primer sets were used to measure *CEACAM1* transcript levels for the alternatively spliced long (L) form, the short (S) forms missing the cytoplasmic tail, and a subset of both long and short forms, respectively. Relative abundance of the target transcripts was calculated by comparison to a standard curve, and normalized to the expression level of TATA box binding protein-associated factor, RNA polymerase I, B (*TAF1B*). Full details and Applied Biosystems catalog numbers for all assays are available in Additional data file 5. The data are also available in NCBI's Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) and are accessible through GEO Series accession number GSE9873.

### Additional data files

Additional data is available with this paper online. Additional data file 1, genes elevated in acute and convalescent KD, is a spreadsheet containing a list of the genes in clusters 1 and 2 in Figure 1 (cohort 1). Gene symbol, gene name, and gene cluster are listed for each gene. Additional data file 2, person-intrinsic genes, is a spreadsheet containing the list of genes identified as having person-specific patterns of expression (see Figure 3).

Additional data file 3 is a figure showing the reproducibility of whole blood genome-wide transcript abundance patterns. Additional data file 4 is a spreadsheet listing gene sets correlated with clinical parameters (see Figure 4). Additional data file 5 contains detailed information on the quantitative RT-PCR methods used in this study.

### Authors' contributions

SJP, JCB, and DAR conceived and designed the study; JCB, JTK, JWN, and RPS contributed samples and clinical information; SJP, CS, and HS carried out the experiments; SJP performed the statistical analysis; SJP, CS, HS, POB, JCB, and DAR interpreted the data and wrote the manuscript; JTK, JWN, and RPS provided critical reviews of the manuscript.

### Acknowledgements

We thank Joseph Marquis and DeeAnna Scherrer for technical assistance, Amy Kapp and Elizabeth Joyce for helpful discussion, and Joan Pancheri and Annette L. Baker, for data and sample collection. This work was supported in part by the National Heart, Lung, Blood Institute of the NIH (HL69413 to J.C.B., J.K., and D.A.R., and K24 HL074864 to J.C.B.), the Farb Family Fund to J.W.N., and a gift from the Horn Foundation to P.O.B. and D.A.R. P.O.B. is an investigator of the Howard Hughes Medical Institute.

### References

1. Chang HY, Nuyten DS, Sneddon JB, Hastie T, Tibshirani R, Sorlie T,

- Dai H, He YD, van't Veer LJ, Bartelink H, et al.: **Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival.** *Proc Natl Acad Sci USA* 2005, **102**:3738-3743.
2. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, et al.: **Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling.** *Nature* 2000, **403**:503-511.
3. Lapointe J, Li C, Higgins JP, van de Rijn M, Bair E, Montgomery K, Ferrari M, Egevad L, Rayford W, Bergerheim U, et al.: **Gene expression profiling identifies clinically relevant subtypes of prostate cancer.** *Proc Natl Acad Sci USA* 2004, **101**:811-816.
4. Liu M, Popper SJ, Rubins KH, Relman DA: **Early days: genomics and human responses to infection.** *Curr Opin Microbiol* 2006, **9**:312-319.
5. Griffiths MJ, Shafi MJ, Popper SJ, Hemingway CA, Kortok MM, Wathen A, Rockett KA, Mott R, Levin M, Newton CR, et al.: **Genomewide analysis of the host response to malaria in Kenyan children.** *J Infect Dis* 2005, **191**:1599-1611.
6. Simmons CP, Popper SJ, Dolocek C, Chau TNB, Griffiths M, Dung NTP, Long TH, Hoang DM, Chau NV, Thao LTT, et al.: **Patterns of host genome-wide gene transcript abundance in the peripheral blood of patients with acute dengue haemorrhagic fever.** *J Infect Dis* 2007, **195**:1097-1107.
7. Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, Brown PO: **Individuality and variation in gene expression patterns in human blood.** *Proc Natl Acad Sci USA* 2003, **100**:1896-1901.
8. Burns JC, Glode MP: **Kawasaki syndrome.** *Lancet* 2004, **364**:533-544.
9. Burns JC, Capparelli EV, Brown JA, Newburger JW, Glode MP: **Intravenous gamma-globulin treatment and retreatment in Kawasaki disease. US/Canadian Kawasaki Syndrome Study Group.** *Pediatr Infect Dis J* 1998, **17**:1144-1148.
10. Obata-Onai A, Hashimoto S, Onai N, Kurachi M, Nagai S, Shizuno K, Nagahata T, Matsushima K: **Comprehensive gene expression analysis of human NK cells and CD8(+) T lymphocytes.** *Int Immunol* 2002, **14**:1085-1098.
11. Palmer C, Diehn M, Alizadeh AA, Brown PO: **Cell-type specific gene expression profiles of leukocytes in human peripheral blood.** *BMC Genomics* 2006, **7**:115.
12. Kobayashi SD, Voyich JM, Whitney AR, DeLeo FR: **Spontaneous neutrophil apoptosis and regulation of cell survival by granulocyte macrophage-colony stimulating factor.** *J Leukoc Biol* 2005, **78**:1408-1418.
13. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci USA* 2001, **98**:5116-5121.
14. Gray-Owen SD, Blumberg RS: **CEACAM1: contact-dependent control of immunity.** *Nat Rev Immunol* 2006, **6**:433-446.
15. Brown TJ, Crawford SE, Cornwall ML, Garcia F, Shulman ST, Rowley AH: **CD8 T lymphocytes and macrophages infiltrate coronary artery aneurysms in acute Kawasaki disease.** *J Infect Dis* 2001, **184**:940-943.
16. Finberg RW, Newburger JW, Mikati MA, Heller AH, Burns JC: **Effect of high doses of intravenously administered immune globulin on natural killer cell activity in peripheral blood.** *J Pediatr* 1992, **120**:376-380.
17. Terai M, Kohno Y, Niwa K, Toba T, Sakurai N, Nakajima H: **Imbalance among T-cell subsets in patients with coronary arterial aneurysms in Kawasaki disease.** *Am J Cardiol* 1987, **60**:555-559.
18. Burns JC, Shimizu C, Shike H, Newburger JW, Sundel RP, Baker AL, Matsubara T, Ishikawa Y, Brophy VA, Cheng S, et al.: **Family-based association analysis implicates IL-4 in susceptibility to Kawasaki disease.** *Genes Immun* 2005, **6**:438-444.
19. Abe J, Jibiki T, Noma S, Nakajima T, Saito H, Terai M: **Gene expression profiling of the effect of high-dose intravenous Ig in patients with Kawasaki disease.** *J Immunol* 2005, **174**:5837-5845.
20. Nomura I, Abe J, Noma S, Saito H, Gao B, Wheeler G, Leung DY: **Adrenomedullin is highly expressed in blood monocytes associated with acute Kawasaki disease: a microarray gene expression study.** *Pediatr Res* 2005, **57**:49-55.
21. Foell D, Ichida F, Vogl T, Yu X, Chen R, Miyawaki T, Sorg C, Roth J: **SI00A12 (EN-RAGE) in monitoring Kawasaki disease.** *Lancet* 2003, **361**:1270-1272.
22. Beiser AS, Takahashi M, Baker AL, Sundel RP, Newburger JW: **A predictive instrument for coronary artery aneurysms in Kawasaki disease. US Multicenter Kawasaki Disease Study Group.**

- Am J Cardiol* 1998, **81**:1116-1120.
23. Honkanen VE, McCrindle BW, Laxer RM, Feldman BM, Schneider R, Silverman ED: **Clinical relevance of the risk factors for coronary artery inflammation in Kawasaki disease.** *Pediatr Cardiol* 2003, **24**:122-126.
  24. Harada K: **Intravenous gamma-globulin treatment in Kawasaki disease.** *Acta Paediatr Jpn* 1991, **33**:805-810.
  25. Wenisch C, Patruta S, Daxbock F, Krause R, Horl W: **Effect of age on human neutrophil function.** *J Leukoc Biol* 2000, **67**:40-45.
  26. Elghetany MT, Lacombe F: **Physiologic variations in granulocytic surface antigen expression: impact of age, gender, pregnancy, race, and stress.** *J Leukoc Biol* 2004, **75**:157-162.
  27. Kobayashi T, Inoue Y, Takeuchi K, Okada Y, Tamura K, Tomomasa T, Morikawa A: **Prediction of intravenous immunoglobulin unresponsiveness in patients with Kawasaki disease.** *Circulation* 2006, **113**:2606-2612.
  28. Muta H, Ishii M, Egami K, Furui J, Sugahara Y, Akagi T, Nakamura Y, Yanagawa H, Matsuishi T: **Early intravenous gamma-globulin treatment for Kawasaki disease: the nationwide surveys in Japan.** *J Pediatr* 2004, **144**:496-499.
  29. Durongpisitkul K, Soongswang J, Laohaprasitiporn D, Nana A, Prachuabmoh C, Kangkagate C: **Immunoglobulin failure and retreatment in Kawasaki disease.** *Pediatr Cardiol* 2003, **24**:145-148.
  30. Fukunishi M, Kikkawa M, Hamana K, Onodera T, Matsuzaki K, Matsumoto Y, Hara J: **Prediction of non-responsiveness to intravenous high-dose gamma-globulin therapy in patients with Kawasaki disease at onset.** *J Pediatr* 2000, **137**:172-176.
  31. Singer BB, Klaiile E, Scheffrahn I, Muller MM, Kammerer R, Reutter W, Obrink B, Lucka L: **CEACAMI (CD66a) mediates delay of spontaneous and Fas ligand-induced apoptosis in granulocytes.** *Eur J Immunol* 2005, **35**:1949-1959.
  32. Tsujimoto H, Takeshita S, Nakatani K, Kawamura Y, Tokutomi T, Sekine I: **Intravenous immunoglobulin therapy induces neutrophil apoptosis in Kawasaki disease.** *Clin Immunol* 2002, **103**:161-168.
  33. Sugita K, Hirao J, Arisaka O, Eguchi M: **Gamma-globulin-induced modulation with necrotic-like morphology of peripheral blood neutrophils.** *Eur J Pharmacol* 2005, **513**:141-144.
  34. Newburger JW, Takahashi M, Gerber MA, Gewitz MH, Tani LY, Burns JC, Shulman ST, Bolger AF, Ferrieri P, Baltimore RS, et al.: **Diagnosis, treatment, and long-term management of Kawasaki disease: a statement for health professionals from the Committee on Rheumatic Fever, Endocarditis and Kawasaki Disease, Council on Cardiovascular Disease in the Young, American Heart Association.** *Circulation* 2004, **110**:2747-2771.
  35. de Zorzi A, Colan SD, Gauvreau K, Baker AL, Sundel RP, Newburger JW: **Coronary artery dimensions may be misclassified as normal in Kawasaki disease.** *J Pediatr* 1998, **133**:254-258.
  36. **Brown Lab Protocols** [<http://cmgm.stanford.edu/pbrown/protocols/index.html>]
  37. Eisen MB, Brown PO: **DNA arrays for analysis of gene expression.** *Methods Enzymol* 1999, **303**:179-205.
  38. **Stanford MicroArray Database** [<http://genome-www5.stanford.edu>]
  39. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci USA* 1998, **95**:14863-14868.
  40. Saldanha AJ: **Java Treeview - extensible visualization of microarray data.** *Bioinformatics* 2004, **20**:3246-3248.
  41. Hosack DA, Dennis G Jr, Sherman BT, Lane HC, Lempicki RA: **Identifying biological themes within lists of genes with EASE.** *Genome Biol* 2003, **4**:R70.