Genetic endothelial systems biology of sickle stroke risk

Liming Chang Milbauer,1 Peng Wei,1,2 Judy Enenstein,1 Aixiang Jiang,1,2 Cheryl A. Hillery,3 J. Paul Scott,3
Stephen C. Nelson,4 Vidya Bodempudi,1 James N. Topper,5 Ruey-Bing Yang,5 Betsy Hirsch,5 Wei Pan,2 and
Robert P. Hebbel1

1Vascular Biology Center and Division of Hematology-Oncology-Transplantation, Department of Medicine, University of Minnesota Medical School, Minneapolis;
2Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis; 3Blood Research Institute, Children’s Research Institute, Medical College of Wisconsin, Milwaukee; 4Minneapolis Children’s Hospital and Department of Pediatrics, University of Minnesota Medical School, Minneapolis;
5Millennium Pharmaceuticals, San Francisco CA and 6Department of Laboratory Medicine and Pathology, University of Minnesota Medical School, Minneapolis

Genetic differences in endothelial biology could underlie development of phenotypic heterogeneity among persons afflicted with vascular diseases. We obtained blood outgrowth endothelial cells from 20 subjects with sickle cell anemia (age, 4-19 years) shown to be either at-risk (n = 11) or not-at-risk (n = 9) for ischemic stroke because of, respectively, having or not having occlusive disease at the circle of Willis. Gene expression profiling identified no significant single gene differences between the 2 groups, as expected. However, analysis of Biological Systems Scores, using gene sets that were predetermined to survey each of 9 biologic systems, showed that only changes in inflammation signaling are characteristic of the at-risk subjects, as supported by multiple statistical approaches. Correspondingly, subsequent biologic testing showed significantly exaggerated RelA activation on the part of blood outgrowth endothelial cells from the at-risk subjects in response to stimulation with interleukin-1β/tumor necrosis factor. We conclude that the pathobiology of circle of Willis disease in the child with sickle cell anemia predominantly involves inflammation biology, which could reflect differences in genetically determined endothelial biology that account for differing host responses to inflammation. (Blood. 2008;111:3872-3879)

Introduction

Many human diseases present in a clinically variable manner, yet the basis for the biologic phenomenon of phenotypic heterogeneity, the variation in presentation of any given disease, is generally unknown. We have used a specific example of this phenomenon to address our overarching hypothesis that genetic, inherited differences in endothelial biology can underlie the phenotypic heterogeneity of human vascular disease.

Sickle cell anemia, caused by inherited homozygosity for the mutant sickle hemoglobin, is a disease characterized by anemia, vascular occlusions, and chronic organ damage. It has an exceedingly complex pathophysiology and incredibly diverse clinical complications.1 Among these, there are 3 stroke syndromes: clinically silent strokes occurring in children resulting from multifocal small vessel disease; hemorrhagic strokes occurring in adults; and clinical ischemic stroke, the classical stroke syndrome of sickle cell anemia.

Notably, approximately 10% of children with sickle cell anemia develop classic ischemic stroke, with peak age being approximately 5 years.2,3 Risk factors include elevated white count, low blood hemoglobin, hypertension, and a prior neurologic event.2-5 However, the primary risk factor is occlusive disease at the circle of Willis (CoW).6,7 the encircling structure of medium to large vessels at the base of the brain. CoW disease is thought to be causal, as the strokes tend to be due to thrombosis occurring over the area of vessel wall abnormality, and the extent of stroke correlates with degree of CoW stenosis.2,8 Stroke pathogenesis does not simply involve sickling in the vasa vasorum because vessels in the CoW do not have vasa vasorum.9

Our hypothesis predicts that those children with sickle cell anemia who develop CoW disease and are therefore at-risk for ischemic stroke have inherited different polymorphisms (affecting endothelial gene expression) from each other, but polymorphisms that exert similar downstream effects on the relevant biologic systems involved in CoW disease development. Indeed, HLA linkage10 and sib-pair analysis11 have suggested a familial predisposition to stroke in sickle cell anemia.

A limitation in understanding this problem is that the identity of the systems biology of CoW disease in sickle cell anemia is, indeed, not known. On the other hand, a great deal is known about the vascular pathobiology of the sickle cell anemia subject in general. Particularly notable is the fact that sickle disease is a systemic inflammatory state.12 Thus, we expected that our study would probably implicate inflammation signaling as able to discriminate between the 2 study groups: sickle cell anemia children at-risk for ischemic stroke, by virtue of having CoW disease (n = 11); and sickle cell anemia children not-at-risk for stroke by virtue of not having CoW disease (n = 9).

Our approach uses endothelial sampling from carefully selected persons, with subsequent global microarray profiling for endothelial gene expression. Thus, the method is blind, for example, to...
single nucleotide polymorphisms\(^\text{13}\) that might modify a gene’s function without affecting expression level. Nonetheless, application of this approach to subjects having precisely defined phenotypes allows an assessment of constitutive, genetically determined phenotype to be made at the gene expression level.

Because our approach is based on use of cultured blood outgrowth endothelial cells (BOECs) as generic endothelial reporter cells, we performed extensive preliminary validation studies to ensure that the method would be useful. Some of these are described in “Methods validation studies.”

**Methods**

**Subjects**

Subjects had genotype HbSS or HbSP\(^\text{0-thal}\) (both have the sickle cell anemia phenotype, but they are difficult to distinguish from each other). All subjects gave informed assent, and informed consent was obtained in accordance with the Declaration of Helsinki from parents. The study was supervised by the Institutional Review Boards at the participating institutions.

**Group assignment**

We defined the state of being at-risk for stroke by virtue of having demonstrable CoW disease. Evidence for this was either a transcranial Doppler study showing flow rates more than 200 cm/sec at the CoW on repeated study,\(^\text{7}\) or a magnetic resonance angiogram\(^\text{14}\) abnormal at the CoW, or development of a chronic stroke in childhood (since these are overwhelmingly ischemic because of CoW disease\(^\text{3}\)). Converesely, the not-at-risk group was identified by having transcranial Doppler with repeated flow rates of less than 150 cm/sec or a normal magnetic resonance angiogram, plus the clinical criterion of having reached age 10 (at least) without a stroke. Based on the literature, these criteria would result in an extremely low risk (\(\sim 2\%)\) of false group assignment. Data on individual subjects are shown in Table S1 (available on the Blood website; see the Supplemental Materials link at the top of the online article).

**BOEC culture**

We obtained citrated fresh peripheral blood (50-100 mL) from each donor. For donors outside the Minneapolis metropolitan area, blood was shipped by same-day express delivery in Saf-T-Pak cartons that had been pre-equilibrated to room temperature. Samples were processed immediately on arrival at the University of Minnesota. Blood buffy coat mononuclear cells were prepared and cultured on rat collagen I, in presence of endothelial cell arrival at the University of Minnesota. Blood buffy coat mononuclear cells BOEC culture (see the Supplemental Materials link at the top of the online article).

We prepared BOEC lysates by adding 10 mL trizol (Invitrogen, Carlsbad, CA) to each T75 flask to solubilize the sample. As necessary, trizol lysates were stored at \(-80^\circ\text{C}\). We then isolated total RNA, and cleaned it using an RNeasy kit (Qiagen, Valencia, CA). We used the Invitrogen SuperScript Choice system to reverse transcribe and synthesize double-stranded cDNA. For in vitro transcription and biotin-labeling, we synthesized biotin-labeled cRNA using the Enzo Life Sciences BioArray High Yield RNA Transcript Labeling kit (Farmingdale, NY). We ran gels to verify that the cRNA prep was the correct size and checked 260/280 absorbance to verify sample concentration and purity. Then fragmentation of cRNA was done by standard protocol, and fragmentation was verified by gel electrophoresis. Samples were run if they yielded more than 15 \(\mu\text{g}\) fragmented cRNA. We always prepared samples from both subject groups on a given day to avoid artifact because of batch effects.

We used the Affymetrix U133A chip for this study because it has the features we are interested in: it interrogates approximately 22 000 transcripts, most known human genes, and numerous Expressed Sequence Tags. Details of Affymetrix (Santa Clara, CA) technology are provided at www.affymetrix.com. After we prepared biotin-labeled cRNA fragments, samples were turned over immediately to our Microarray Core Facility, which used Gene Chip Hybridization Oven 640, Affymetrix FS4000 Fluidics Station for staining with a streptavidin-fluorochrome probe and washing, and an Affymetrix High Resolution Scanner. To eliminate one source of variability, we purchased one single lot of chips in sufficient quantity for all samples from this project, so there would be no effects from different lots of chips.

**RNA isolation and quantitative reverse-transcribed polymerase chain reaction**

The quality and quantity of total RNA were determined by UV spectrophotometer; 1 \(\mu\text{g}\) of total RNA from each sickle patient was subjected for cDNA synthesis in 20 \(\mu\text{L}\) of a reaction using SuperScript III First-Strand Synthesis SuperMix for quantitative reverse-transcribed polymerase chain reaction (RT-PCR; Invitrogen) according to the protocol supplied with the kit. The cDNA was then used for quantitative reverse-transcribed polymerase chain reaction in an I Cycler (Bio-Rad Laboratories, Hercules, CA) to detect each gene expression. The unique sequence regions from each gene were identified, and subjected for designing specific primers and fluorescence FAM-labeled probe using the program of “Beacon Designer” (Bio-Rad Laboratories). The kit of TaqMan Rodent GAPDH Control Reagents, which contains universal primers and fluorescence VIC-labeled probe from Applied Biosystems (Weiterstadt, Germany), was used for reference of each target gene. Each quantitative PCR reaction contained 2 \(\mu\text{L}\) of cDNA (50 ng of original total RNA), 600 nM of each sense and antisense primers, 240 nM of probe, and 1 \(\times\) of iQ Supermix (Bio-Rad Laboratories). The PCR condition for most of genes was as follows: activating the iTaq DNA polymerase 95°C for 5 minutes, followed by 50 cycles of 2-step PCR (95°C for 30 seconds and 60°C for 30 seconds). The annealing temperature for the genes JRK, CXCL2, and IL18 was 58°C. To validate the efficiency of target and reference, serially diluted cDNA samples from normal patient’s BOECs were run on quantitative PCR side by side with those of 20 sickle patients in each 96-well reaction plate of each gene. The Ct (Threshold Cycle) values from samples of each gene and the internal control GAPDH were obtained and then calculated the relative fold change between 20 samples based on the “Comparative Ct Method” describe in the “User Bulletin #2” from User’s Manual for ABI PRISM 7700 (Applied Biosystems).

**RelA activation assay**

BOECs were stimulated with both 30 ng/mL of interleukin-1β (IL-1β) and 30 ng/mL of tumor necrosis factors (TNFα) for 1 hour or 24 hours, after cells reached 90% confluence. Nuclear proteins from nonstimulated and stimulated BOECs were extracted using Nuclear Extraction Kit (Active Motif, Carlsbad, CA) according to the protocol supplied by the company. The protein concentration was determined using Bio-Rad Protein Concentration Assay.
Bioinformatics

For raw microarray data, we used the robust multiarray average\textsuperscript{17} method to summarize expression values from probe pair values; data were background-adjusted, quantile normalized (global scaling across multiple arrays), and expression measures were summarized based on log-transformed perfect match values using Median Polish algorithm. Then, we used locally weighted scatterplot smoothing\textsuperscript{18} to do within-array normalization. Robust multiarray average and locally weighted scatterplot smoothing were performed in Genedata Expressionist Pro3.1PP (Basel, Switzerland). Bioinformatics approaches are discussed in detail in “Biologic systems analysis.” Briefly, we applied several tests for significant differences in single gene expression: Welch test, Wilcoxon test, significance analysis of microarray.\textsuperscript{19} Our analysis of biologic systems used the Wald test.\textsuperscript{20} This is a t-like test that is vectorial based and allows analysis that multiple dimensions simultaneously contribute to a difference between groups. Gene set enrichment analysis\textsuperscript{21} was applied for permutation analyses, as was the permutation Wald test.

Results

Methods validation studies

BOECs were cultured by a method developed in our laboratory.\textsuperscript{15,16} This yields a single population (at the level of immunohistochemistry and fluorescence-activated cell sorter analysis) that has typical cobblestone morphology, takes up acetylated low density lipoprotein, stains positively for multiple endothelial markers but negatively or low for endothelial activation antigens, stains negatively for monocytes (CD14) and myeloid cells (CD45) and endothelial progenitors (AC133), and does not express markers of other cell types, such as fibroblasts or smooth muscle cells.\textsuperscript{15,16} BOEC cultures reach extremely high levels of expansion (Figure S1) if allowed to do so.\textsuperscript{15} For the present studies, cells were expanded only approximately 1 million-fold (to $\sim$3 $\times$ 10\textsuperscript{7} cells). Validation experiments focused on the central issues of using cultured BOEC.

To allow power calculations to be done, we performed microarray analysis on BOECs from 27 normal subjects of diverse ages (GEO accession GSE98777, http://www.ncbi.nlm.nih.gov/geo).\textsuperscript{22} This showed that, at all quintiles of gene expression, the group sizes used in our actual study have more than 99% power of detecting a 2-fold or greater change in expression of any single gene. There was no significant effect of donor age on gene expression.

To address whether BOECs might reveal transiently acquired phenotypes (eg, by virtue of the inflammation status of the subjects), we deliberately stimulated BOEC from 5 donors with IL-1$\beta$ plus TNF$\alpha$ and monitored gene expression. In response, 122 genes significantly changed expression level. After only a 10-fold expansion thereafter, expression of all genes had returned to baseline. One of these experiments is shown as Figure 1. Because our BOECs are expanded 1 million-fold after removal from the donor, we are confident that acquired phenotypes will have washed out and that observed expression therefore reflects only culture conditions and the genetics of the subjects. Culture conditions were, of course, scrupulously standardized and carried out so that all samples were treated identically.

The other most significant question pertains to whether BOECs are sufficiently stable in culture for this study to be done. To address this, we cultured BOECs from 5 donors for multiple passages and monitored variance for expression of each individual transcript, as a function of passage number (each passage representing a 10-fold additional expansion in cell number). The example shown as Figure 2 plots the variance at passage 3 minus variance at passage 7. Approximately 99% of transcripts show no significant change across this 10 000-fold increase in cell number. The rest, comprising 2 tails that show increased or decreased variance with sequential passage, could represent development of instability or washout of phenotype, respectively. However, analysis reveals that 40% of them are highly expressed ribosomal protein genes, and none are members of the crucial biologic system categories we study here. We did our present study on cells that are comparable with passage 3 in this stability experiment (ie, before substantial expansion), and we checked for normal cytogenetics at time of culture completion.

Single gene analysis

Then, we examined our actual study samples for gene expression (GEO accession GSE98777\textsuperscript{22}): 9 subjects not-at-risk and 11 subjects at-risk. Two-class analysis failed to identify significant differences in single gene expression, using several tests: the Welch t test, which assumes distribution normality, as is generally true for microarray data; the Wilcoxon test, which does not require data to be normally distributed; and statistical analysis of microarray, which is based on false discovery rate (FDR), developed for microarray data and small sample numbers, where normality is not known and multiple comparisons are taken into account. Thresholds for significance were $P$ less than .001 for the Welch and Wilcoxon tests, and the FDR cutoff was 5%. Indeed, our expectation was not to find single gene differences that define the 2 subject groups because it is vanishingly unlikely for sickle cell anemia to be accompanied at allelic frequency by another single gene disorder.
Biologic systems analysis

Rather, we expected the 2 subject groups to be distinguished in terms of biologic systems. Specifically, our concept has been that different persons have different combinations of polymorphisms that affect endothelial gene expression and, therefore, endothelial biology. Thereby, some persons would have common downstream patterns of gene expression, affecting genes that impact similarly on the systems biology of CoW disease.

Because the systems biology of CoW disease is, indeed, unknown, we identified 9 biologic systems that could be implicated. For each biologic system, we assembled gene sets that would survey that system, aiming for gene sets of approximately 125 items in size. (For this gene set assembly, we surveyed specific published articles on each of the 9 biologic systems; some of these showed pathways and/or microarray lists or referenced Gene Ontology lists; in addition, pathway sources such as KEGG, BIOCARTA, STKE, and INGENUITY were used. More information about sources used to establish gene sets is provided in Table S2.) Doing this, we attempted to identify enhanced “activity” of each biologic system by surveying both regulatory (proximate) and endpoint molecules. Examples would be, respectively, RelA and integrins. Thus, gene sets were neither specific for the given systems nor were they mutually exclusive. The biologic systems (number of genes in list) were adhesion (146), angiogenesis (131), apoptosis (79), coagulation (152), hypoxia response (109), inflammation (117), redox signaling (83), shear stress response (156), and vasoregulation (106). Composition of all 9 gene lists is shown in Table S3. It is important to emphasize that these gene sets were identified and assembled before any gene expression results were obtained, so they are entirely unbiased vis-à-vis results.

To express the gene expression of these biologic systems, we used a Biological Systems Score (BSS). This was determined for each of the 9 systems for each subject by summing the “corrected ratio” of gene expression for each item in the given systems gene set. To create the corrected ratio, the raw expression of each gene was divided by the average value of all control subjects (the not-at-risk group) for that gene; then, values less than 1 were converted proportionately to values more than 1 (eg, 0.5 becomes 2; 0.1 becomes 10). By these manipulations, we avoid several severe problems with constructing scores, for example: the risk that highly expressed genes would overwhelm the contribution of lower expressed genes, the risk that downward fold changes would contribute less to the score than upward fold changes, and the risk that downward changes would offset upward changes. Thus, each subject receives a unique BSS for each biologic system, a score that reflects changes in the system’s gene expression, regardless of direction. Note that in cases where a given gene is represented by more than one probe, the values derived from the different probes were averaged so that one value was available for each gene. For example, 220 transcripts contributed to the 117 genes on the inflammation list.

For statistical comparison of BSS, we first applied the Wald test,20 which is a large sample-based test that does not require the BSS to be normally distributed. For single biologic systems (a one-dimensional Wald test that is essentially a 2-sided z test), we found the strongest difference to be for inflammation biology ($P = .022$), the only system that reached $P$ less than .05 (Table 1). (Of interest, 22 genes in the inflammation set were among the 122 genes that responded to TNF/IL-1 stimulation in the above-noted experiment, thus demonstrating that the inflammation gene set broadly surveys inflammation biology, rather than focusing on a single pathway.) Adhesion biology and coagulation were both close to being significant, however. The same comparison of an independent group of different adult healthy normal donors, divided into whites ($n = 21$) and blacks ($n = 22$), showed no remotely significant differences (Table 1), nor did division of this normal donor group into males ($n = 23$) versus females ($n = 21$; Table 1). The failure to observe significant biologic systems differences for these control comparisons indirectly supports the veracity of the critical comparison of at-risk versus not-at-risk study subjects. The inflammation and coagulation and hypoxia BSS are normally distributed, according to the Shapiro-Wilk Normality Test, whereas those for the other systems are not.

Note that the fact that the $P$ values are marginal is because the gene sets in the biologic systems were relatively large and created with no knowledge of results (see “Permutation testing”).

Permutation testing

Considering our relatively small sample size ($n = 20$), we were concerned with whether the Wald test would give us valid results. We therefore sought corroboration by permutation test using the Wald statistic (ie, permutation Wald test), which is nonparametric.
and does not require a parametric form of the test statistic distribution under the null hypothesis. We thereby obtained P values very close to those from the original Wald test, which indicates that it is appropriate to use the Wald test in this context. These permutation Wald test P values were as follows: inflammation (0.032), adhesion (0.075), coagulation (0.098), and other systems (> 0.10). Using the inflammation system gene set to create random groups of n = 9 and n = 11, we found that the strongest Wald score was obtained when subjects 3, 8 to 14, and 16 to 18 were assigned to the 11-subject group. Perfectly correct assignment, assuming the clinical criteria are accurately separating the 2 groups, would be subjects 10 to 20 in the 11-subject group. Thus, there are only 3 false assignments, assuming the clinical criteria sorted phenotypes correctly. We conclude that inflammation biology is probably the dominant system involved in CoW disease.

As an additional test for veracity of results, we performed the Gene Set Enrichment Analysis to test whether randomly arranged groups of 9 and 11 subjects had significant enrichment of the gene set probes in the at-risk versus not-at-risk groups. For this nonparametric, distribution-independent test, we used 2 types of permutation. Using phenotype label permutation (random variations of group sorting by assigned phenotype), we see that only the inflammation gene set is so significantly enriched at a false discovery rate less than 25% (FDR = 17.5%), which is considered an acceptable cutoff for application of this to microarray data. Using gene tag permutation (group sorting based on our permutation Wald test it is appropriate to use the Wald test in this context. These groups of n (2-dimensional Wald Test, 3-dimensional Wald test, etc). Although we expected the combination of inflammation and certain other systems to show improved significance, this is not the case. For example, a 2-dimensional Wald test looking at both BSS for inflammation and BSS for adhesion (P = 0.012), coagulation (P = 0.016), hypoxia (P = 0.023), or redox signaling (P = 0.035) did not show much improvement of significance over inflammation alone (P = 0.022). Thus, this further argues that inflammation is the dominant factor on CoW disease.

Analysis of the relative contribution of individual genes to the overall BSS of the inflammation system was assessed by ranking the 220 probed transcripts (rather than the corresponding 117 probed genes) in order of individual r value. We thereby, identified the top 24 and next 59 and next 137 contributors to the subjects’ inflammation scores. Using these truncated biologic system gene sets, we found that examination of the Wald score for the 2 subject groups confirmed the result for the original, full gene set: for the top 24 probes, P = 4.48 × 10⁻¹⁰; for the next 59 probes, P = 0.093; and for the least-contributing 137 probes, P = 0.826. Thus, it appears that a truncated list of 24 inflammation probes yields the most efficient discrimination of the at-risk and not-at-risk groups. It may be that prospective use of this BSS method could use such a truncated gene set. The members of the top group of 21 genes in the inflammation gene set, represented by 24 probes, in order of descending contribution to the Wald test are: CXCL5 (↑), CSF1 (↑), ITGAL (↓), ITGB1 (↓), LAMB3 (↑), ITG1 (↑), ITGA5 (↑), ITGA6 (↑), FBXW7 (↑), LAMC2 (↑), PSMB2 (↑), IRF1 (↓), IL-18 (↑), GPR37 (↑), LAMC2 (↑), HIPK2 (↑), BTG2 (↑), CX3CL1 (↑), ITGB3 (↑), HSPA8 (↑), BTG2 (↓), CALM1 (↓), TNFAIP6 (↑), COL4A1 (↓), and CXCL11 (↑). The arrow indicates whether the change in expression is increased or decreased for the at-risk compared with the not-at-risk subjects. This is a generally proinflammatory profile.

### Biologic confirmation
Quantitative, real-time PCR performed on 19 genes, plus NFκB RELA, confirmed only small differences between the 2 groups, as expected from the r values, consistent with the fact that there were no individual genes that were significantly different, as assessed by raw microarray expression values. Notably, however, 4 of the genes did show significant differences for PCR based expression, and 2 others were close: CXCL5 (P = 0.047), LAMB3 (P = 0.025),
CXCL2 ($P = .009$), CXCL11 ($P = .072$), LAMC2 ($P = .086$), and IL27RA ($P = .048$).

The top 80 inflammation genes are highly biologically related, as shown by a biologic relationship map for the top 21 created with Ingenuity Pathway and Analysis software (Redwood City, CA) (shown partially in Figure 3; shown completely in Figure S2). Because most members of this universe are biologically either proximate or distal to NFκB component RelA/p65, we evaluated the microarray-determined expression level of RELA for our 20 subjects (Figure 4A). This showed no statistically significant difference. It is evident by inspection, however, that this is probably the result of the subject-to-subject heterogeneity within the at-risk group, which substantially exceeds that of the not-at-risk group (which also tends to have lower RELA expression). From this, it appears that one of the control subjects might be misassigned into the not-at-risk group based on clinical criteria. This RELA variability, we think, is consistent with our overarching hypothesis. Indeed, we expect different subjects to have different polymorphisms in inflammation signaling. The probability of this is also illustrated by microarray-determined levels of CXCL2 expression, this being an important inflammatory regulatory molecule (Figure 4B).

The implication of these results is that the subjects in the at-risk group have potentially exaggerated biologic responses to inflammation. To test this, we thawed and expanded all of the viable original cryopreserved BOEC samples and tested the RelA response to stimulation with IL-1β and TNFα. For this assay, intraindividual variability was 8.9%, a log lower than average interindividual variability (89%). The significantly exaggerated RelA functional protein response of the at-risk subjects (Figure 5) confirms the notion that at-risk status corresponds to exaggerated inflammation response.

**Discussion**

We have studied CoW disease as the primary risk factor (and causal factor) for the classic stroke syndrome of sickle cell anemia, which involves ischemic stroke in childhood via thrombosis over an area of CoW occlusive disease. Our results are consistent with our original hypothesis: that phenotypic heterogeneity (in this case, sickle cell anemia with vs without CoW disease) can have a genetic endothelial basis. The tolerance of BOEC for cryopreservation allowed us to test the cell biology of at-risk versus not-at-risk BOEC samples, and results seem to confirm an exaggerated inflammation signaling response on the part of the at-risk subjects. The data suggest the culmination of this is at NFκB component RelA signaling (and perhaps p50). Interestingly, it has been suggested by several investigators that RelA homodimers are a major operative NFκB signaling unit. Our results thus suggest that the systems biology of CoW disease, which had not been identified, most likely involves inflammation signaling. This is consistent with our understanding of the pathobiologic context of sickle cell anemia.
We note that a mainstay of our method involves use of preselected gene sets. These were constructed (using a variety of sources, per Table S2) to broadly survey biologic systems for enhanced “activity.” They were, thus, neither specific nor mutually exclusive. For example, adhesion and inflammation gene sets overlap. Although our analysis suggests that inflammation is the key process in CoW disease, we acknowledge that a larger study might have further identified coagulation and adhesion biomarkers as being relevant. And use of a differently constructed gene set may have influenced results. Thus, the current data need to be further tested by a new, prospective study. An illustrative example is provided by ITGAL, also known as LFA1 and not known as an endothelial molecule. However, we included it in the endothelial inflammation gene set because ITGAL is reported, although, to our knowledge, not yet verified, to be expressed by endothelial progenitor cells and therefore could be considered a legitimate part of the endothelial repertoire. It is not yet known if this RNA expression is accompanied by protein translation. We do know that ITGAL expression by BOEC is not the result of monocyte contamination of the cell preparations (see “Methods”).

Another feature of our method is the use of cultured BOECs. Because these are extensively cultured and are not naturally residing endothelial cells in the first place, they are free of tissue specification phenomena and acquired influences. We regard this as an advantage because the point of our study was to evaluate for possible generic influences on the endothelium itself. However, it is true that our method is blind to issues that are clearly related to endothelial biology through the blood milieu exposure and via tissue specification of phenotype. Some blood influences could be genetically determined and exerted by changes, for example, in plasma proteins; our method is not sensitive to the resulting acquired change of endothelium itself.

The current approach would perhaps be useful in exploring other vascular diseases in which well-defined clinical sub-phenotypes exist. Our experience with this approach points to inflammation biology, in the disease context at hand, via employment of multiple statistical approaches. But our analysis of partial subcomponents of the inflammation gene set highlights that results are much more robust (P < .001 vs P = .022) if a truncated gene set is used (24 probes vs full set of 220 probes). This would have to be determined by preliminary testing in any given situation.

Acknowledgments

The authors thank the Minnesota Supercomputing Institute for computer resources, Jim Kiley for technical assistance, and Sil Benkovic for editorial assistance.

This work was supported by the National Institutes of Health (HL68970 and HL076540) and by a grant from Millennium Pharmaceuticals.

Authorship

Contribution: L.C.M. performed the molecular biologic aspects of the paper, along with V.B., and cowrote the paper; P.W., A.J., and W.P. conducted the bioinformatics aspects of the research; J.E. conducted cell biology experiments, constructed the biologic system gene sets, and cowrote the text; C.A.H., J.P.S., and S.C.N. each identified appropriate study subjects and obtained peripheral blood from them; J.N.T. and R.-B.Y. performed some of the very initial validation experiments that demonstrated this method could work; B.H. conducted the cytogenetics analysis; R.P.H. conceived and supervised the overall project, validated subject eligibility, oversaw day-to-day conduct of the project, designed cell biology experiments, constructed the biologic system gene sets, and edited and cowrote the paper.

Conflict-of-interest disclosure: A patent on the BOEC culture method and use is held by the University of Minnesota and R.P.H., but there is no current financial interest. Otherwise, the authors declare no competing financial interests.

Correspondence: Robert P. Hebbel, MMC 480, University of Minnesota Medical School, 420 Delaware Street SE, Minneapolis, MN 55455; e-mail: hebbe001@umn.edu.

References


Genetic endothelial systems biology of sickle stroke risk