A. SPECIFIC AIMS

The specific aims for this proposal have not been modified from the original competing application, and have been largely accomplished within the funding period.

B. STUDIES AND RESULTS

Specific Aim 1: Develop a quantitative measure of population heterogeneity to evaluate specific HSC and MSC populations from WT and diabetic mice.

In **Specific Aim 1**, we developed a quantitative measure of population heterogeneity based on the distribution of single cell gene expression profiles. We applied microfluidic analysis employing high-throughput real-time qPCR to measure precise mRNA content for individual mesenchymal stem cells (MSCs) and hematopoietic stem cells (HSCs) isolated from the murine bone marrow compartment. The transcriptional heterogeneity of each population was then evaluated using a novel adaptation of phi-entropy based divergence analysis. Briefly, Pearson product-moment correlation coefficients were calculated for each pair of gene targets based on expression across all cells in the population. In order to avoid artifacts due to data scaling, a shrinkage transformation was applied to the resulting matrix, with shrinkage intensity determined according to sample size and variance. Differential entropy, an extension of Shannon's entropy to continuous probability distributions, was then computed for this modified covariance matrix about a multivariate lognormal probability density function for gene expression. This provides a generalizable, indirect measure of information content and randomness based on maximum likelihood estimates, which functions as our absolute measure of heterogeneity for a population (**Figure 1**).

Differential Entropy: $h(f) = -\int_X f(x)\log f(x)dx$

Multivariate PDF: $f(x) = (2\pi)^{-\frac{p}{2}}\left|\sum\right|^{-\frac{1}{2}}\left(\prod_{i=1}^{p} x_i\right)^{-1} e^{-\frac{1}{2}(\log x-\mu)'\sum^{-1}(\log x-\mu)}$

Heterogeneity Metric: $h(f(x)) = \log(2\pi e)^{-\frac{p}{2}} + \frac{1}{2}\log\left|\sum\right| + \sum_{i=1}^{p}\mu_i$

Figure 1. Transcriptional heterogeneity metric based on the multivariate generalization of Shannon's entropy.

Specific Aim 2: Identify characteristic expression profiles for discrete MSC subpopulations.

Wild-type (WT) and leptin receptor-deficient (db/db) diabetic mice were euthanized and BM cells harvested. A well-characterized HSC population (Lin-/Sca-1+/cKit+ [LSK cells]) and a less well-characterized MSC population (Lin-/Sca-1+/CD45-) were FACS-isolated from primary (P0) and passage 3 (P3) cells, respectively. Isolated cells were evaluated by phi-entropy based analysis (as in Aim 1), and differences in individual gene expression determined using Bonferroni-corrected Kolmogorov-Smirnov statistics. Fuzzy c-means clustering was used to further partition populations based on correlations in median-normalized gene expression. Model parameters were optimized using Akaike Information Criterion (AIC), and the autonomy of each cluster evaluated by Kullback-Leibler divergence.

Greater heterogeneity was observed in both wild-type (WT) and diabetic MSC populations than among LSK cells, and MSCs from diabetic mice exhibited less heterogeneity than their wild-type counterparts (Figure 2). Diabetic MSCs also demonstrated significantly reduced expression of multiple known progenitor and stem cell-associated genes. Cluster analysis of pooled MSC data, without a priori knowledge of phenotype, identified four

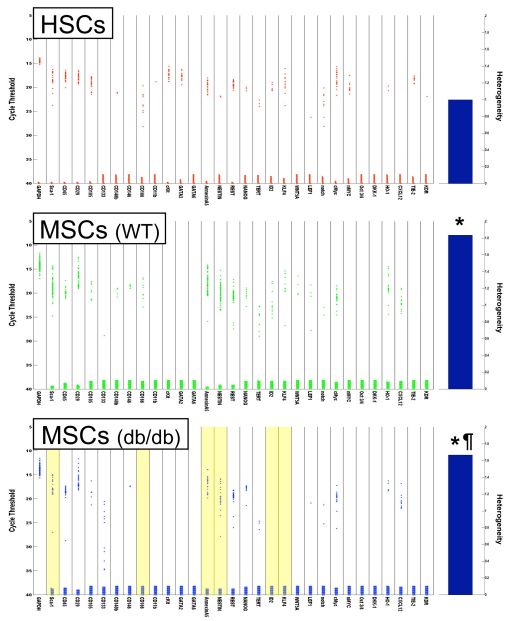

Figure 2. qPCR data (**left**). Each point in the histogram represents a single cell. Genes are shown in vertical columns, with yellow bars highlighting decreased expression in db/db relative to WT by KS-testing ($p<.05$). Heterogeneity for each population is evaluated as differential entropy [h(f)] about a lognormal distribution [f(x)] (per Fig. 1). Blue vertical bars represent heterogeneity [h(f(x))] values (normalized to HSCs [**right**]). **, p<.05 vs HSCs; ¶, p<.05 vs WT MSCs.*

distinct subgroups, which were objectively parameterized over cluster- and fuzzy- space using AIC. Notably, one cluster (1) was composed entirely of WT cells. Autonomy assessment using Kullback-Leibler divergence found the independence of this cluster statistically significant when compared with the remaining pooled MSC population (Figure 3). In a repeat comparison of WT and diabetic gene expression, omitting this single cluster corrected the disparities in all but one gene.

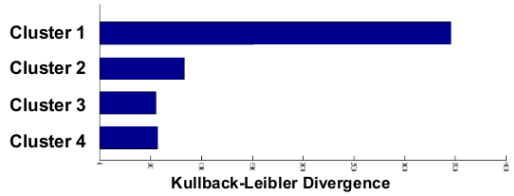

Figure 3. Kullback-Leibler divergence test statistics. Only cluster 1 demonstrated significant divergence. **, $p < 0.05$*

C. SIGNIFICANCE

Transcriptional changes in diabetic mesenchymal stem cells were attributable to the absence of a specific progenitor subpopulation, rather than global shifts in gene expression, suggesting a potential mechanism for vasculogenic dysfunction in diabetes. These findings support the development of translational studies to determine whether this specific pathophysiologic mechanism is also characteristic of human diabetic patients, the clinical implications of which would be considerable.