Dysregulated Gene Expression Networks in Human Acute Myelogenous Leukemia Stem Cells

Ravindra Majeti[†], Michael W. Becker[§†], Qiang Tian [‡¶†], Tsung-Lu Michael Lee[‡||], Xiaowei Yan[‡], Rui Liu[**], Jung-Hsien Chiang[‡||], Leroy Hood[‡], Michael F. Clarke[*] & Irving L. Weissman[*]*

**Institute for Stem Cell Biology and Regenerative Medicine and Department of Internal Medicine, Divisions of Hematology and Oncology, Stanford University, Palo Alto, CA 94304 §Division of Hematology/Oncology, University of Rochester, Rochester, NY 14627 ‡Institute for Systems Biology, Seattle, WA 98103 ¶Department of Medicine/Hematology, University of Washington School of Medicine, and the Institute for Stem Cell and Regenerative Medicine, Seattle, WA 98103 ||Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan 70101,Taiwan **University of Michigan School of Dentistry, Ann Arbor, MI 48109*

†These authors contributed equally to this work.

Corresponding Author: Irving L. Weissman

279 Campus Drive West, B257 Beckman Center

Stanford University School of Medicine

Stanford, CA 94305

email: irv@stanford.edu

Classification: Biological Sciences – Medical Sciences

Manuscript Information: 25 pages, 3 figures, 1 table, 4 supplemental figures/tables

Abbreviations: AML, acute myeloid leukemia; HSC, hematopoietic stem cell; LSC, leukemia stem cell

ABSTRACT

We performed the first genome wide expression analysis directly comparing the expression profile of highly enriched normal human hematopoietic stem cells (HSC) and leukemic stem cells (LSC) from patients with acute myeloid leukemia (AML). Comparing the expression signature of normal HSC to that of LSC, we identified 3005 differentially expressed genes. Using two independent analyses, we identified multiple pathways that are aberrantly regulated in leukemic stem cells in comparison to normal HSC. Several pathways, including Wnt signaling, MAP Kinase signaling, and Adherens Junction, are well known for their role in cancer development and stem cell biology. Other pathways have not been previously implicated in the regulation of cancer stem cell functions, including Ribosome and T Cell Receptor Signaling Pathway. This study demonstrates that combining global gene expression analysis with detailed annotated pathway resources applied to highly enriched normal and malignant stem cell populations, can yield an understanding of the critical pathways regulating cancer stem cells.

\body

INTRODUCTION

With the ability to enrich for rare populations of cells using cell sorting techniques and the development of appropriate xenotransplant models, it has been possible to prospectively characterize the surface antigen phenotype of both normal and leukemic stem cell populations from normal and AML specimens. A number of lines of evidence have demonstrated that human HSC are contained in the Lin-CD34+CD38- fraction of hematopoietic progenitors (1, 2). Additional studies have demonstrated that human HSC also express CD90 (3-5). Perhaps the best demonstration of HSC function comes from human clinical trials of autologous mobilized peripheral blood in clinical transplantation, where long-term engraftment was provided by transplantation of purified CD34+CD90+ cells (6-8). In multiple published reports, a common phenotype for AML LSC has been identified and found to be negative for expression of lineage markers (Lin-), positive for expression of CD34, and negative for expression of CD38 (9-11). These Lin-CD34+CD38- LSC were further shown to be positive for expression of IL-3Rα (CD123) and negative for expression of CD90 (Thy-1) (12-14).

The cancer stem cell model has significant implications for the design of therapies for AML as well as other cancers. It postulates that in order to eradicate the tumor, therapies must target and eliminate the cancer stem cells. For the development of such cancer stem cell-targeted therapies, it is necessary to identify molecules and pathways that are preferentially expressed in these cancer stem cells compared to their normal counterparts.

DNA microarray technology has proven to be a powerful tool for the large scale analysis of gene expression differences, particularly in cancer investigations (15). There are many published reports of gene expression profiles of bulk human AML samples,

including several very large cohorts (16-18). However, few studies have directly compared AML to normal hematopoietic cells (19). Comparisons of unfractionated populations may fail to identify critical differentially expressed genes and pathways within infrequent LSC and rare HSC populations. One report has investigated gene expression differences between LSC and non-LSC from the same AML samples in order to identify genes critical to LSC function; however, no direct comparison was made to normal HSC (20). An additional study has investigated gene expression differences between normal HSC and HSC from patients with 5q- myelodysplastic syndrome, and identified genes potentially involved in the pathogenesis of this disorder (21). We report the first analysis of gene expression differences between AML LSC and normal bone marrow HSC using two independently derived data sets. We have applied the gene sets to a systems level pathway analysis and have identified molecular pathways and networks that are dysregulated between AML LSC and HSC.

RESULTS

Generation of gene expression profiles of human HSC and AML LSC

Two independent sets of microarray gene expression profiles of human bone marrow HSC and AML LSC were generated at the University of Michigan (n=3 HSC, n=7 AML) and Stanford University (n=4 HSC, n=9 AML) by fluorescence-activated cell sorting (FACS) of patient samples, followed by RNA purification, amplification, and hybridization to Affymetrix oligonucleotide-based microarrays. The clinical features of the AML samples are presented in Supplemental Table 1, and cover a range of subtypes of AML. These data sets were then combined for the bioinformatic analysis of differentially regulated pathways.

Dysregulated pathways between AML LSC and HSC

Although LSC and normal HSC share the common characteristics of unlimited self-renewal and multi-lineage differentiation, understanding how they differ from each other should reveal fundamental mechanisms governing leukemic transformation. While differentially expressed genes between LSC and HSC can be identified and validated individually, the full potential of genome-wide microarray analysis can be better realized in terms of gene regulatory networks, given that those genes, and the proteins they encode for, function in the context of intertwining pathways. We conducted an unbiased systems level pathway analysis without excluding any genes through a priori gene expression thresholds, by employing a newly developed algorithm dubbed integrative microarray analysis of pathways (IMAP) (22). This analysis first combines microarray data from multiple independent experiments through meta-analysis; a score for each pathway is assigned without specific gene expression cut-offs; the significance of each pathway is then computed by running 1 million iterations of randomly sampled genes.

We used the pathway information from Biocarta (http://www.biocarta.com), KEGG (http://www.genome.jp/kegg/), GeneGo (http://www.genego.com/), and Pathway Studio (http://www.ariadnegenomics.com/products/pathway-studio/) databases as references to derive the top dysregulated pathways. As shown in Table 1, among the top dysregulated pathways between LSC and HSC, are pathways involved in adherens junction, regulation of the actin cytoskeleton, apoptosis, MAPK signaling, and Wnt signaling. A full list of the top dysregulated pathways derived from all 4 databases, as well as up- or down-regulated pathways, is shown in Supplemental Table 2. Among the down-regulated pathways in LSC are those related to tumor suppressors (such as RB and ATM signaling), CXCR4->Stat3/5B pathways, and regulation of translation initiation. Thus, global pathway analysis has identified critical biological networks perturbed in LSC compared to normal HSC.

Identification of key molecular interactions within dysregulated pathways

To identify changing molecular interaction and reaction networks contributing to the top dysregulated pathways between LSC and HSC, we mapped the relative expression levels of all the genes found in a given pathway in the context of their signal transduction and cell communication processes as defined by the KEGG pathway database using the Advanced Pathway Painter program (http://www.gsa-online.de/eng/app.html). As shown in Figure 2, two of the top dysregulated pathways—adherens junction (2a) and Wnt signaling pathway (2b)—are illustrated with red color representing genes up-regulated in LSC, green color representing genes down-regulated in LSC, and yellow color representing genes with no significant change between LSC and HSC. Among the genes that are downregulated in LSC in the adherens junction pathway are α-Catenin, Afadin, and PAR3 (Figure 2a). Genes that are upregulated in LSC in the Wnt pathway include Axin and APC, whereas c-Jun, a

TCF/Lef target gene is downregulated in LSC (Figure 2b).

Validation of top dysregulated pathways using functional groups enrichment analysis

Independent of the IMAP analysis, which takes into account all the genes in a given pathway, we conducted a parallel analysis employing only differentially expressed genes (DEGs) from the combined datasets. The two AML microarray data sets were first combined as described above using z-scores. With $p<0.05$ we obtained 3005 DEGs (Supplemental Table 3). A heat map illustrating the expression of these genes across the Stanford samples is shown in Supplemental Figure 1. A functional enrichment analysis using this set of genes was performed using the DAVID program (the **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery; http://david.abcc.ncifcrf.gov/home.jsp)(23). The latest version of DAVID supports 40 annotation categories, including Gene Ontology terms, KEGG pathways, protein-protein interactions, protein functional domains, swissprot keywords, disease associations, bio-pathways, sequence general features, and literature. A number of functional groups or biological themes are identified as enriched in our AML dataset compared to the whole genome distribution of that specific group using DAVID ($p<0.02$). We developed a bioinformatic tool (Lee et al, unpublished) to visualize these enriched functional group associations using the Cytoscape program (http://www.cytoscape.org) (24). As shown in Figure 3, among the 3005 DEGs, genes involved in protein kinase activity, adherens junction, actin cytoskeleton, and apoptosis are specifically enriched. The size of each circle represents the number of genes in that specific functional group; the thickness of the lines represents the number of overlapping genes between the functional groups. Thus, two independent network analyses arrive at some common pathways differentially regulated between LSC and HSC, strongly implicating them in regulation of leukemic

stem cell functions.

DISCUSSION

While there is an abundance of data examining the gene expression profiles of normal and malignant bulk cell populations in AML, little has been done in the application of the stem cell model to gene expression analysis of AML samples. This is due in part to the difficulty in isolating a sufficient number of cells to perform these studies. Several studies have utilized CD34+ cells for microarray analyses, however, the expression of CD34 is often aberrant in AML and most studies to date have demonstrated there remains considerable heterogeneity in the CD34 positive fraction (9, 12, 13). We and others have previously demonstrated the value of applying macro- and microarray technology to highly enriched populations of normal stem cells and their committed progeny to identify key regulators of cell fate choices (25-27). In 2002, Guzman et al. applied macro-array technology to compare leukemic and normal stem cell populations and identified, among other genes, activation of the NFκB pathway in all the LSC samples examined (28). Recently Gal et al. examined the gene expression profile of enriched leukemic progenitors from 5 patients using microarray technology and compared the results of their study to a dataset previously published for normal hematopoietic progenitors (20). The current study represents the first effort to simultaneously compare the transcriptional profiles of highly enriched LSC and normal HSC from a wide variety of patients using modern microarray technology. The direct comparison of expression patterns of LSC to HSC, as opposed to other non-stem cell populations, enhanced our ability to identify genes and pathways which are disrupted at the stem cell level in AML. This analysis provides critical insights into the differences between normal and malignant stem cell populations which may be used for the development of targeted therapies, as well as tools for assessing the impact of therapy on the LSC population.

The first data set was generated from sorted leukemic stem cells (LSC) from 7 AML patients and normal hematopoietic stem cells (HSC) from 3 normal controls. The second data set was generated similarly from 9 AML patients and 4 normal individuals. We obtained 3005 differentially expressed genes with 1451 genes being up regulated and 1554 genes being down regulated in LSC when compared to normal HSC. Many of these genes have been previously identified as being playing a key role in normal stem cell biology as well as in leukemia. Despite the use of different microarray platforms, a detailed analysis of the two independently derived datasets demonstrated a significant degree of overlap between the gene signatures for the two cell types examined.

In addition to providing a genome wide survey for genes whose expression is disrupted in leukemic transformation at the level of the stem cell compartment, a comprehensive, unbiased pathway analysis has allowed the identification of critical pathways which are dysregulated in LSC in comparison to normal HSC. Again, focusing on the difference between normal and leukemic as opposed to stem cell versus progenitor has allowed us to screen out many of the pathways that are involved in stem cell function which are not dysregulated in the leukemic state. Hence, the limited number of pathways we identified as being dysregulated. Furthermore, the significant degree of overlap within and between the different pathway tools utilized suggests that these pathways are critical in the evolution of cancer stem cells from their normal counterparts.

To validate the data sets used, we performed an analysis employing only the 3005 differentially expressed genes (DEGs) from the combined datasets using the DAVID program. This independent and parallel analysis demonstrated a significant degree of overlap with the unbiased pathway analysis and confirmed the validity of the datasets employed in these analyses using a system wide approach.

Many of the pathways we identified as being aberrantly regulated in the LSC from the patients studied have already been established as playing key roles in leukemia

and/or leukemic stem cell biology including the Wnt canonical (29), the Adherens junction and NFκB pathways (28). Several of the pathways identified are involved in the interaction of the stem cells with their niche. There is a growing body of data demonstrating the importance of the interaction of stem cells with their niche in normal and malignant stem cell biology. Two classes of receptors and their ligands are critical in determining the nature of this interaction, cell adhesion molecules (CAMs) and chemokine receptors. Our current analysis demonstrates dysregulation of both gene families in the leukemic stem cells studied. Identification of dysregulation of these interactions is consistent with the hypothesis that alteration of the stem cell:niche interaction is a key step in the pathogenesis of cancer stem cells (30). In regard to CAMs, we identified several CAM related pathways that were aberrantly regulated in LSC including the Adherens junction and Tight junction pathways. In addition, pathway analysis of the data sets using the Biocarta, GeneGo, and Pathway Studio tools identified: How does salmonella hijack a cell, Adhesion Molecules on Lymphocyte, CXCR4 -> STAT3 signaling pathway, CXCR4 -> STAT5B signaling pathway, and the Angiopoietin - Tie2 signaling pathways (Supplemental Table 2).

The adherens junction has been demonstrated to be critical in the interaction of HSC and niche in both the fetal liver stage as well as adult stage of hematopoiesis (31-33). Our data confirms that N-cadherin and alpha-E catenin are expressed in normal human stem cells and that expression of these genes is disrupted in LSC. Several studies have also demonstrated that elements of the adherens junction are aberrantly expressed in leukemic cells compared to normal hematopoietic cells (34-38). We recently identified aberrant expression of CTNNA1 in highly enriched LSC from patients with advanced MDS and AML associated with abnormalities of chromosome 5 (39). Likewise, the Angiopoietin - Tie2 signaling pathway has been demonstrated to play an important role in normal and leukemic stem cell function (35, 36), and expression of Tie2 has been

shown to be lost in LSC. Finally, the selectins, as well as members of the tight junction complex are also involved in the interaction of normal stem cells with endothelial cells and are dysregulated during the leukemic transformation process.

Another family of proteins critical to the normal and leukemic stem cell: niche interaction is the chemokine family. CXCR4 is a chemokine receptor that plays a key role in regulating normal and leukemic stem cell homing to the bone marrow niche. The role of CXCR4 signaling in AML is of significant interest as several targeted therapies which disrupt the interaction of CXCR4 with its ligand CXCL12 are currently being investigated in clinical trials.

The Wnt canonical pathway was found to be dysregulated in this study between LSC and HSC, and has been implicated in the pathogenesis of several different types of human cancer, including leukemia. We have previously shown that the canonical Wnt pathway, signaling through nuclear beta-catenin, regulates the self-renewal of mouse HSC (40, 41). We have also shown that this pathway is aberrantly activated in downstream progenitors in the blast crisis phase of chronic myelogenous leukemia, resulting in the nuclear localization of beta-catenin where it likely acts to stimulate self-renewal and contributes to the formation of leukemia stem cells (29). Our data suggests that this pathway may be dysregulated in AML stem cells, possibly contributing to pathogenesis.

Many pathways not previously implicated in the regulation of leukemia stem cell functions were identified in our analysis. Several basic cellular biology pathways identified include Ribosome, Regulation of Actin Cytoskeleton, and Regulation of Translation Initiation. Numerous metabolic pathways were identified such as Glycosphingolipid, Androgen and Estogen, Glycerophospholipid, Regulation of Fatty Acid Synthase Activity, and Arginine Metabolism. Other pathways involved signal transduction including Angiotensin II, Oxidative Stress-Induced Gene Expression, T Cell Receptor, CD28, B Cell Receptor, and EGF

Signaling. Ultimately, detailed biological investigations will be necessary to determine the functional involvement of these pathways in leukemic pathogenesis.

In summary, we have utilized gene expression profiling and annotated pathway resources to identify biological networks that are dysregulated in AML stem cells compared to normal HSC. The application of the stem cell model for AML to a systems biology analysis of stem cell expression networks has confirmed the role of several pathways previously demonstrated to be important in cancer stem cell function. In addition, this approach has identified several pathways not previously studied in cancer stem cells. Such networks are candidates for involvement in the regulation of critical cancer stem cell functions, and as such may be targets for therapeutic intervention.

MATERIALS AND METHODS

Human Samples

Normal human bone marrow mononuclear cells were either purchased from AllCells Inc. (Emeryville, CA) or obtained from National Marrow Donor Program discarded filter units. For AML specimens, peripheral blood and/or bone marrow was obtained following informed consent at the time of clinical presentation according to IRB approved protocols at the University of Michigan or Stanford University. Mononuclear cells were prepared using Ficoll-Paque Plus (GE Healthcare, Fairfield, CT), and either used fresh or cryopreserved in 90% FBS/10% DMSO in liquid nitrogen.

Isolation and Purification of Normal HSC and AML LSC

Viably frozen cells were thawed and resuspended in IMDM with 2% heat-inactivated FBS and DNAse I (Sigma). For all normal specimens and leukemic samples with less than 10% of blasts expressing CD34, enrichment was performed prior to staining using CD34 positive selection (Stem Cell Technologies, Canada and Miltenyi Biotech, Germany). Cells were stained as previously described (39) (Supplemental Methods) and analyzed and sorted using FACSAria cytometers (BD Biosciences). 15,000-65,000 normal HSC and approximately 50,000 to 150,000 AML LSC were sorted for RNA purification.

RNA Purification, Amplification, and Microarray Analysis

Total RNA was extracted using Trizol reagent containing either glycogen or linear acrylamide according to the manufacturer's protocol, and then treated with DNaseI (Ambion, Austin, TX). For the University of Michigan samples only, RNA was re-extracted in Trizol. All RNA samples were quantified with the RiboGreen RNA

Quantitation Kit (Molecular Probes Inc., Eugene, Oregon), subjected to reverse transcription, two consecutive rounds of linear amplification, and production and fragmentation of biotinylated cRNA (Affymetrix). 15μg of cRNA from each sample was hybridized to Affymetrix HG U133 A or B (University of Michigan samples) or HG U133 Plus 2.0 (Stanford University samples) microarrays. Hybridization and scanning were performed according to the manufacturer's instructions (Affymetrix). The data from the University of Michigan and Stanford expression array studies was uploaded into ISB's Systems Biology Experiment Analysis Management System (SBEAMS) and normalized using the GCRMA algorithm.

Integration of Data Set

Since the two data sets of LSC and HSC were obtained from different platforms (3 different chips) and/or different probe sets, the ability to directly compare the data sets to each other was limited, and the correlation coefficient between the two data sets was low. To obtain a statistically more robust data set, we developed a strategy to combine both datasets for an increased sample size (Figure 1).

We performed meta-analysis for integration of the two different microarray data sets using a procedure similar to Setlur et al. (22). P-values were first calculated for all genes within each data set using the Wilcoxon rank-sum test to compare LSC against HSC samples. These p-values were then mapped onto a standard normal curve. That is, z-scores corresponding to the p-values were obtained by the following conversion:

$$z = \frac{p - \mu}{\sigma},$$

where p is the p-value to be standardized, μ the mean of all p-values within a data set, and σ the standard deviation of the p-values within the data set. After standardization, the two data sets were integrated to get a weighted z-score for each gene by combining

all z-scores for the gene across the two data sets, using the following Liptak-Stouffer formula:

$$z_{comb,g} = \frac{\sum_{i=1}^{m} w_i z_{i,g}}{\sqrt{\sum_{i=1}^{m} w_i^2}},$$

where m=2 for this study, i.e. the number of data sets; w_i =1, the weight of the i-th data set; $z_{i,g}$ the z-score of gene g in the i-th data set. The less the value of $z_{comb,g}$, the less the probability of differential expression between HSC and LSC by chance. The probability can be calculated from the cumulative standard normal distribution:

$$P(z \le z_{comb,g}) = \int_{-\infty}^{z_{comb,g}} \frac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{2}} du.$$

Based on these probabilities, namely p-values, we could integrate our data without bias, and finally obtained 3005 potential differentially expressed genes with p-values less than 0.05, and 387 genes with p-values less than 0.01.

Dys-regulation of Pathway

To define the dys-regulation of a pathway P, we first assigned a score s_g to each gene g in the pathway P by the negative logarithm of its p-value p_g, that was

$$s_g = -\log(p_g).$$

We then gave a total score S_P to the pathway P by summing up all scores of genes in the pathway, namely

$$S_P = \sum_{g \in P} s_g.$$

To estimate a p-value for significance of this pathway, we iteratively computed similar scores one million times on randomly generated pathways of the same size as that of pathway P.

The frequency of scores which were larger than S_P was used as the p-value of pathway *P* to describe its dys-regulation. For calculation of up- or down-regulated pathways, a positive or negative s_g value was used respectively to compute the S_p.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Libuse Jerabek for excellent lab management, R.M. is supported by the Walter and Idun Y. Berry Foundation, the Doctors Cancer Foundation, and the AACR Postdoctoral Fellowship in Cancer Research. R.M. holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund. R.M., M.W.B, Q.T., T.M.L., X.Y., R.L., and J.C. have no conflicts of interest to disclose. I.L.W. was a member of the scientific advisory board of Amgen and owns significant Amgen stock. I.L.W. co-founded and consulted for Systemix, is a co-founder and director of Stem Cells Inc., and co-founded Cellerant, Inc. This research is supported by National Institutes of Health grants R01CA86017 to I.L.W and P01DK53074 to M.F.C, I.L.W., and L.H.

AUTHOR CONTRIBUTIONS

R.M, M.W.B., and Q.T. conducted research and wrote the manuscript. T.M.L., X.Y., R.L., and J.C. assisted in preparation of the data. L.H., M.F.C., and I.L.W. supervised the research and assisted in writing the manuscript.

REFERENCES

1. Bhatia, M., Wang, J. C., Kapp, U., Bonnet, D. & Dick, J. E. (1997) Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice. *Proc Natl Acad Sci U S A* 94**,** 5320-5.
2. Kondo, M., Wagers, A. J., Manz, M. G., Prohaska, S. S., Scherer, D. C., et al. (2003) Biology of hematopoietic stem cells and progenitors: implications for clinical application. *Annu Rev Immunol* 21**,** 759-806.
3. Baum, C. M., Weissman, I. L., Tsukamoto, A. S., Buckle, A. M. & Peault, B. (1992) Isolation of a candidate human hematopoietic stem-cell population. *Proc Natl Acad Sci U S A* 89**,** 2804-8.
4. Craig, W., Kay, R., Cutler, R. L. & Lansdorp, P. M. (1993) Expression of Thy-1 on human hematopoietic progenitor cells. *J Exp Med* 177**,** 1331-42.
5. Murray, L., Chen, B., Galy, A., Chen, S., Tushinski, R., et al. (1995) Enrichment of human hematopoietic stem cell activity in the CD34+Thy-1+Lin- subpopulation from mobilized peripheral blood. *Blood* 85**,** 368-78.
6. Michallet, M., Philip, T., Philip, I., Godinot, H., Sebban, C., et al. (2000) Transplantation with selected autologous peripheral blood CD34+Thy1+ hematopoietic stem cells (HSCs) in multiple myeloma: impact of HSC dose on engraftment, safety, and immune reconstitution. *Exp Hematol* 28**,** 858-70.
7. Negrin, R. S., Atkinson, K., Leemhuis, T., Hanania, E., Juttner, C., et al. (2000) Transplantation of highly purified CD34+Thy-1+ hematopoietic stem cells in patients with metastatic breast cancer. *Biol Blood Marrow Transplant* 6**,** 262-71.
8. Vose, J. M., Bierman, P. J., Lynch, J. C., Atkinson, K., Juttner, C., et al. (2001) Transplantation of highly purified CD34+Thy-1+ hematopoietic stem cells in patients

with recurrent indolent non-Hodgkin's lymphoma. *Biol Blood Marrow Transplant* 7**,** 680-7.

9. Bonnet, D. & Dick, J. E. (1997) Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med* 3**,** 730-7.
10. Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., et al. (1994) A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* 367**,** 645-8.
11. Wang, J. C. & Dick, J. E. (2005) Cancer stem cells: lessons from leukemia. *Trends Cell Biol* 15**,** 494-501.
12. Blair, A., Hogge, D. E., Ailles, L. E., Lansdorp, P. M. & Sutherland, H. J. (1997) Lack of expression of Thy-1 (CD90) on acute myeloid leukemia cells with long-term proliferative ability in vitro and in vivo. *Blood* 89**,** 3104-12.
13. Jordan, C. T., Upchurch, D., Szilvassy, S. J., Guzman, M. L., Howard, D. S., et al. (2000) The interleukin-3 receptor alpha chain is a unique marker for human acute myelogenous leukemia stem cells. *Leukemia* 14**,** 1777-84.
14. Miyamoto, T., Weissman, I. L. & Akashi, K. (2000) AML1/ETO-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation. *Proc Natl Acad Sci U S A* 97**,** 7521-6.
15. Quackenbush, J. (2006) Microarray analysis and tumor classification. *N Engl J Med* 354**,** 2463-72.
16. Bullinger, L., Dohner, K., Bair, E., Frohling, S., Schlenk, R. F., et al. (2004) Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med* 350**,** 1605-16.
17. Bullinger, L. & Valk, P. J. (2005) Gene expression profiling in acute myeloid leukemia. *J Clin Oncol* 23**,** 6296-305.

18. Valk, P. J., Verhaak, R. G., Beijen, M. A., Erpelinck, C. A., Barjesteh van Waalwijk van Doorn-Khosrovani, S., et al. (2004) Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med* 350**,** 1617-28.

19. Stirewalt, D. L., Meshinchi, S., Kopecky, K. J., Fan, W., Pogosova-Agadjanyan, E. L., et al. (2008) Identification of genes with abnormal expression changes in acute myeloid leukemia. *Genes Chromosomes Cancer* 47**,** 8-20.

20. Gal, H., Amariglio, N., Trakhtenbrot, L., Jacob-Hirsh, J., Margalit, O., et al. (2006) Gene expression profiles of AML derived stem cells; similarity to hematopoietic stem cells. *Leukemia* 20**,** 2147-54.

21. Nilsson, L., Eden, P., Olsson, E., Mansson, R., Astrand-Grundstrom, I., et al. (2007) The molecular signature of MDS stem cells supports a stem-cell origin of 5q myelodysplastic syndromes. *Blood* 110**,** 3005-14.

22. Setlur, S. R., Royce, T. E., Sboner, A., Mosquera, J. M., Demichelis, F., et al. (2007) Integrative microarray analysis of pathways dysregulated in metastatic prostate cancer. *Cancer Res* 67**,** 10296-303.

23. Dennis, G., Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., et al. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4**,** P3.

24. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13**,** 2498-504.

25. Forsberg, E. C., Prohaska, S. S., Katzman, S., Heffner, G. C., Stuart, J. M., et al. (2005) Differential expression of novel potential regulators in hematopoietic stem cells. *PLoS Genet* 1**,** e28.

26. Ivanova, N. B., Dimos, J. T., Schaniel, C., Hackney, J. A., Moore, K. A., et al. (2002) A stem cell molecular signature. *Science* 298**,** 601-4.

27. Park, I. K., He, Y., Lin, F., Laerum, O. D., Tian, Q., et al. (2002) Differential gene expression profiling of adult murine hematopoietic stem cells. *Blood* 99**,** 488-98.

28. Guzman, M. L., Neering, S. J., Upchurch, D., Grimes, B., Howard, D. S., et al. (2001) Nuclear factor-kappaB is constitutively activated in primitive human acute myelogenous leukemia cells. *Blood* 98**,** 2301-7.

29. Jamieson, C. H., Ailles, L. E., Dylla, S. J., Muijtjens, M., Jones, C., et al. (2004) Granulocyte-macrophage progenitors as candidate leukemic stem cells in blast-crisis CML. *N Engl J Med* 351**,** 657-67.

30. Clarke, M. F. & Fuller, M. (2006) Stem cells and cancer: two faces of eve. *Cell* 124**,** 1111-5.

31. Calvi, L. M., Adams, G. B., Weibrecht, K. W., Weber, J. M., Olson, D. P., et al. (2003) Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* 425**,** 841-6.

32. Kim, I., Yilmaz, O. H. & Morrison, S. J. (2005) CD144 (VE-cadherin) is transiently expressed by fetal liver hematopoietic stem cells. *Blood* 106**,** 903-5.

33. Zhang, J., Niu, C., Ye, L., Huang, H., He, X., et al. (2003) Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* 425**,** 836-41.

34. Acs, G. & LiVolsi, V. A. (2001) Loss of membrane expression of E-cadherin in leukemic erythroblasts. *Arch Pathol Lab Med* 125**,** 198-201.

35. Arai, F., Hirao, A., Ohmura, M., Sato, H., Matsuoka, S., et al. (2004) Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche. *Cell* 118**,** 149-61.

36. Schliemann, C., Bieker, R., Padro, T., Kessler, T., Hintelmann, H., et al. (2006) Expression of angiopoietins and their receptor Tie2 in the bone marrow of patients with acute myeloid leukemia. *Haematologica* 91**,** 1203-11.

37. Shimamoto, T., Ohyashiki, J. H. & Ohyashiki, K. (2005) Methylation of p15(INK4b) and E-cadherin genes is independently correlated with poor prognosis in acute myeloid leukemia. *Leuk Res* 29**,** 653-9.

38. Wang, L., O'Leary, H., Fortney, J. & Gibson, L. F. (2007) Ph+/VE-cadherin+ identifies a stem cell like population of acute lymphoblastic leukemia sustained by bone marrow niche cells. *Blood* 110**,** 3334-44.

39. Liu, T. X., Becker, M. W., Jelinek, J., Wu, W. S., Deng, M., et al. (2007) Chromosome 5q deletion and epigenetic suppression of the gene encoding alpha-catenin (CTNNA1) in myeloid cell transformation. *Nat Med* 13**,** 78-83.

40. Reya, T., Duncan, A. W., Ailles, L., Domen, J., Scherer, D. C., et al. (2003) A role for Wnt signalling in self-renewal of haematopoietic stem cells. *Nature* 423**,** 409-14.

41. Willert, K., Brown, J. D., Danenberg, E., Duncan, A. W., Weissman, I. L., et al. (2003) Wnt proteins are lipid-modified and can act as stem cell growth factors. *Nature* 423**,** 448-52.

FIGURE LEGENDS

Figure 1. Schematic representation of dysregulated pathway identification

Figure 2. Visualization of molecular interaction and reaction networks in the KEGG database: (2a) Adherens junction; (2b) Wnt signaling pathway. Red color represents upregulation in LSC, green color represents upregulation in HSC, and yellow color represents no significant change.

Figure 3. Functional Groups Association Networks (FGAN) visualized using the Cytoscape program. Each node represents one enriched functional group (GO molecular function, GO cellular component, and Swissprot keywords) with p-value < 0.02 performed by DAVID. A total of 3005 differentially expressed genes were evaluated by the functional enrichment analysis. The size of the node is proportional to the number of genes in each functional group. The largest functional group is the protein kinase activity of GO molecular function, which contains 103 genes. The edge width represents the number of shared genes between any two functional groups.

Table 1. Top 10 dysregulated pathways using the KEGG database

Pathway	source	gene #	hit #	dys-pvalue	up-pvalue	down-pvalue
Adherens junction	KEGG	84	79	0	0.140219	0.859781
Ribosome	KEGG	117	94	0	1	0
Regulation of actin cytoskeleton	KEGG	221	214	0.000002	0.000015	0.999985
Tight junction	KEGG	133	128	0.000012	0.959437	0.040563
Focal adhesion	KEGG	239	234	0.000014	0.000776	0.999224
Apoptosis	KEGG	99	98	0.000136	0.073148	0.926852
MAPK signaling pathway	KEGG	251	249	0.000298	0.017295	0.982705
T cell receptor signaling pathway	KEGG	104	103	0.000694	0.021248	0.978752
Jak-STAT signaling pathway	KEGG	164	162	0.000753	0.028108	0.971892
Wnt signaling pathway	KEGG	154	149	0.002304	0.509085	0.490915

This table summarizes the top 10 dysregulated KEGG pathways according to their dys-regulation score (labeled dys-pvalue). The number of genes referenced in the KEGG pathways were labeled as gene #, and the number of genes that were found in our data set were labeled hit #. For each KEGG pathway, an unbiased systems level pathway analysis without excluding any genes was implemented by employing integrative microarray analysis of pathways (IMAP). The p-values of each pathway reflect the significance of dys-regulation (dys-pvalue), up-regulation (up-pvalue), and down-regulation (down-pvalue) of the pathway.

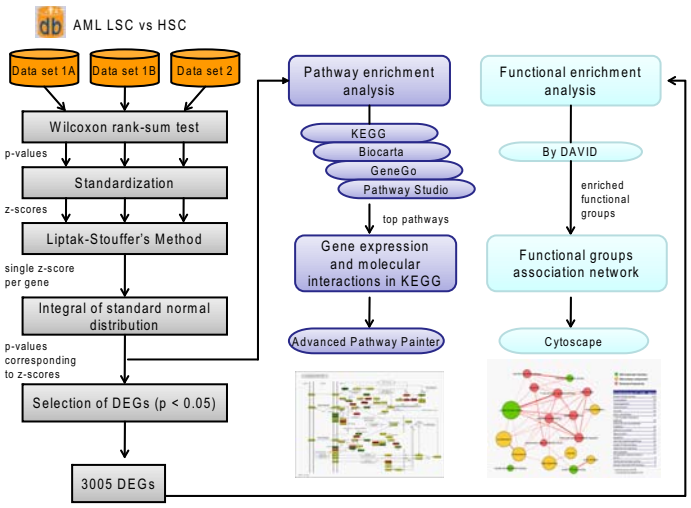

AML LSC vs HSC
Data set 1A
Data set 1B
Data set 2
Wilcoxon rank-sum test
p-values
Standardization
z-scores
Liptak-Stouffer's Method
single z-score per gene
Integral of standard normal distribution
p-values corresponding to z-scores
Selection of DEGs (p < 0.05)
3005 DEGs
Pathway enrichment analysis
KEGG
Biocarta
GeneGo
Pathway Studio
top pathways
Gene expression and molecular interactions in KEGG
Advanced Pathway Painter
Functional enrichment analysis
By DAVID
enriched functional groups
Functional groups association network
Cytoscape

Figure 1

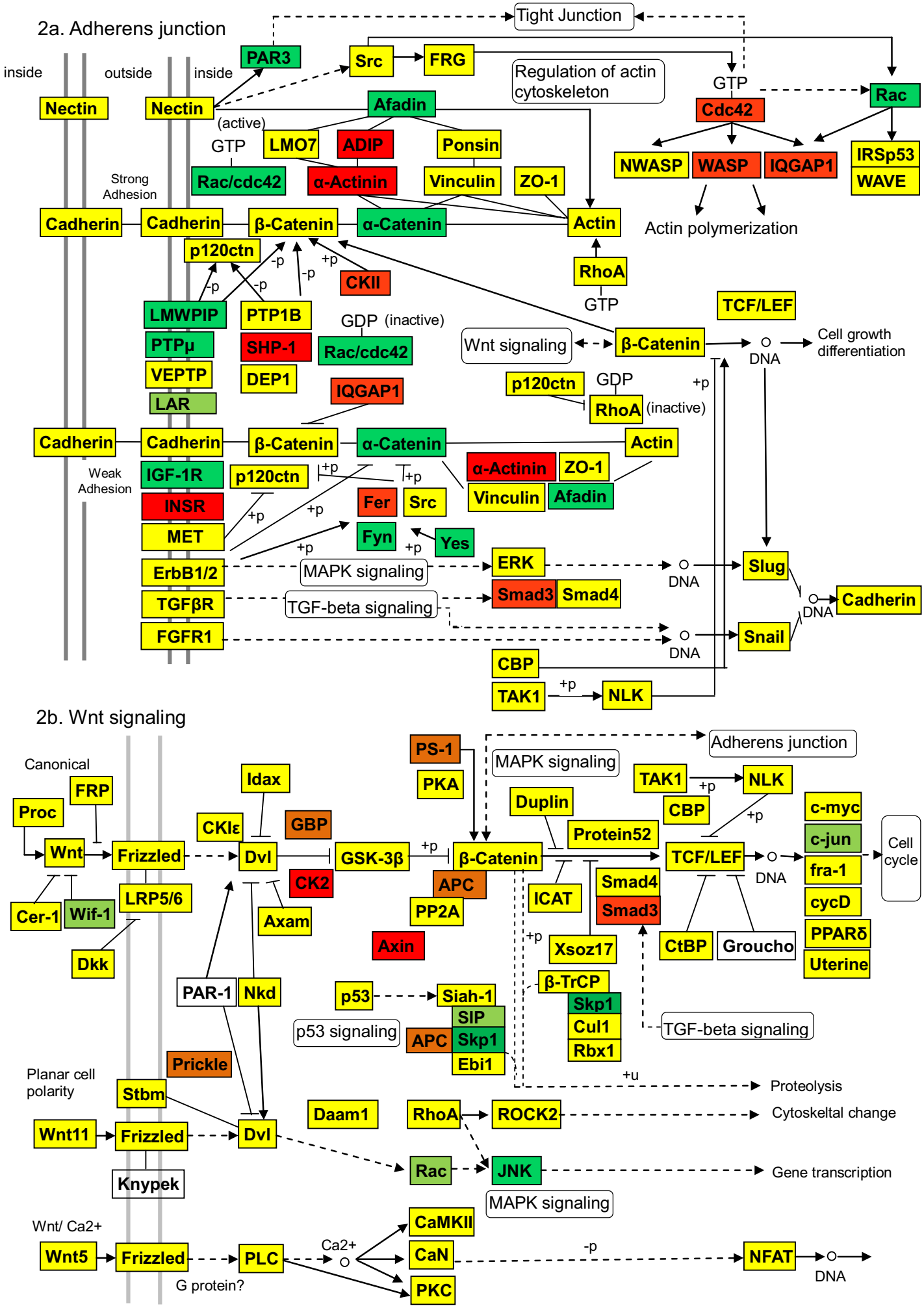
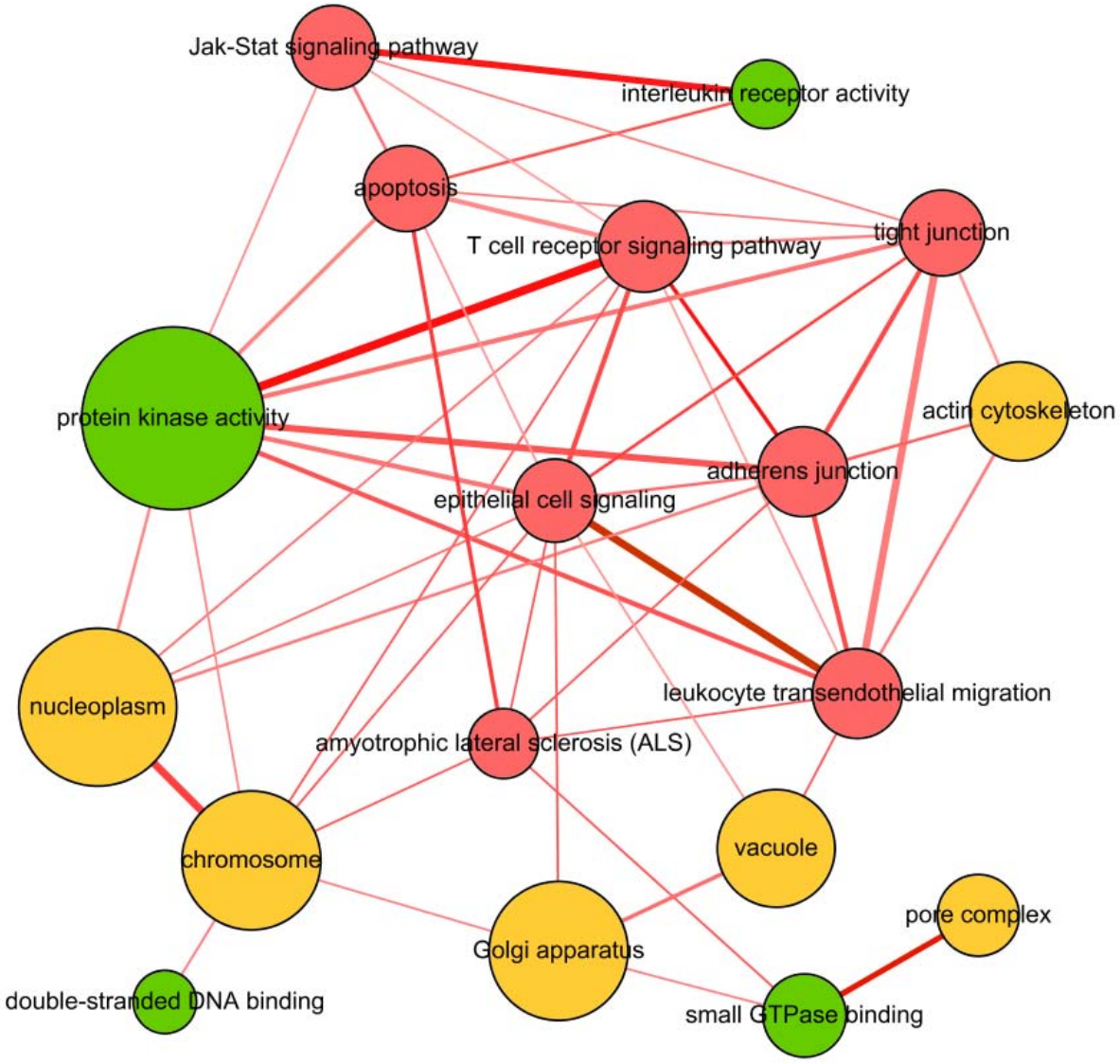

2a. Adherens junction
Tight Junction
inside
outside
inside
PAR3
Src
FRG
Regulation of actin cytoskeleton
GTP
Rac
Cdc42
Nectin
Nectin
Afadin
(active)
GTP
LMO7
ADIP
Ponsin
NWASP
WASP
IQGAP1
IRSp53
WAVE
Strong Adhesion
Rac/cdc42
α-Actinin
Vinculin
ZO-1
Cadherin
Cadherin
β-Catenin
α-Catenin
Actin
Actin polymerization
p120ctn
-p
+p
-p
-p
-p
CKII
RhoA
GTP
LMWPIP
PTP1B
GDP (inactive)
TCF/LEF
PTPμ
SHP-1
Rac/cdc42
Wnt signaling
β-Catenin
Cell growth differentiation
VEPTP
DEP1
DNA
LAR
IQGAP1
p120ctn
GDP
+p
RhoA (inactive)
Cadherin
Cadherin
β-Catenin
α-Catenin
Actin
Weak Adhesion
IGF-1R
p120ctn
+p
+p
α-Actinin
ZO-1
INSR
+p
+p
Fer
Src
Vinculin
Afadin
MET
+p
Fyn
+p
Yes
ErbB1/2
MAPK signaling
ERK
DNA
Slug
TGFβR
TGF-beta signaling
Smad3
Smad4
DNA
Cadherin
FGFR1
DNA
Snail
CBP
TAK1
+p
NLK
2b. Wnt signaling
Canonical
FRP
Idax
PS-1
MAPK signaling
Adherens junction
Proc
PKA
TAK1
+p
NLK
Duplin
CBP
+p
c-myc
CKIε
GBP
Protein52
c-jun
Wnt
Frizzled
Dvl
GSK-3β
+p
β-Catenin
TCF/LEF
DNA
fra-1
Cell cycle
APC
Smad4
CK2
ICAT
cycD
Cer-1
Wif-1
LRP5/6
Axam
PP2A
Smad3
PPARδ
Axin
+p
Xsoz17
CtBP
Groucho
Uterine
Dkk
β-TrCP
PAR-1
Nkd
p53
Siah-1
Skp1
SIP
p53 signaling
APC
Skp1
Cul1
TGF-beta signaling
Prickle
Ebi1
Rbx1
+u
Proteolysis
Planar cell polarity
Stbm
Daam1
RhoA
ROCK2
Cytoskeltal change
Wnt11
Frizzled
Dvl
Rac
JNK
Gene transcription
Knypek
MAPK signaling
Wnt/ Ca2+
CaMKII
Wnt5
Frizzled
PLC
Ca2+
CaN
-p
NFAT
DNA
G protein?
PKC

Figure 3



GO molecular function
GO cellular component
Swissprot keywords

Functional Groups (*P < 0.02)	#genes
protein kinase activity	103
nucleoplasm	76
Golgi apparatus	61
chromosome	61
vacuole	45
actin cytoskeleton	31
T cell receptor signaling pathway	25
leukocyte transendothelial migration	24
adherens junction	24
tight junction	21
apoptosis	21
Jak-Stat signaling pathway	20
small GTPase binding	19
epithelial cell signaling	19
pore complex	18
amyotrophic lateral sclerosis (ALS)	9
interleukin receptor activity	8
double-stranded DNA binding	4

* Performed by DAVID Functional Annotation Tool