Data in Brief

Gene expression profiling to define the cell intrinsic role of the SKI proto-oncogene in hematopoiesis and myeloid neoplasms

Alistair M. Chalk [a], Brian J.J. Liddicoat [a], Carl R. Walkley [a], Sofie Singbrant [a,b,*]

[a] St. Vincent's Institute of Medical Research and Department of Medicine, University of Melbourne, Fitzroy, Australia
[b] Department of Molecular Medicine and Gene Therapy, Lund University, Lund, Sweden

ARTICLE INFO

ABSTRACT

The proto-oncogene SKI is highly expressed in human myeloid leukemia and also in murine hematopoietic stem cells. However, its operative relevance in these cells remains elusive. We have over-expressed SKI to define its intrinsic role in hematopoiesis and myeloid neoplasms, which resulted in a robust competitive advantage upon transplantation, a complete dominance of the stem and progenitor compartments, and a marked enhancement of myeloid differentiation at the expense of other lineages. Accordingly, enforced expression of SKI induced gene signatures associated with hematopoietic stem cells and myeloid differentiation. Here we provide detailed experimental methods and analysis for the gene expression profiling described in our recently published study of Singbrant et al. (2014) in Haematologica. Our data sets (available at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39457) provide a resource for exploring the underlying molecular mechanisms of the involvement of the proto-oncogene SKI in hematopoietic stem cell function and development of myeloid neoplasms.

Specifications

Organism	*Mus musculus*
Array type	Illumina MouseWG-6 v2.0 expression bead chip
	qRT-PCR
Data format	Raw data: TAR. Normalized data: SOFT, MINIML, TXT, XLS, ODF, RDA
Experimental factors	Murine stem- and progenitor cells overexpressing SKI vs. MIG empty control vector.
Experimental features	Human SKI was overexpressed using a retroviral MIG-vector (MCV-based vector containing IRES and the marker gene GFP) in sorted murine stem- and progenitor cells (lin$^-$cKit$^+$Sca-1$^+$) from C57Bl6 animals. Cells were infected and cultured for two days prior to isolation of GFP$^+$ cells, followed by RNA isolation and microarray. Cells transduced with empty control vector (MIG) or overexpressing SKI were compared. Each sample represents an independent infection.
Consent	n/a

Direct link to deposited data

Raw and processed microarray data is available in GEO under accession GSE39457 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39457).

* Corresponding author at: Department of Molecular Medicine and Gene Therapy, Lund University. BMC A12, 221 84 Lund, Sweden. Tel.: +46 46 2220442.
E-mail address: sofie.singbrant@med.lu.se (S. Singbrant).

Materials and methods

Isolation and transduction of primary murine stem- and progenitor cells

The SKI construct [2] was cloned into the Murine Stem Cell Vector (MSCV)-based retroviral MIG-vector containing IRES and the marker gene GFP, and virus was produced as previously described [1].

Hematopoietic stem- and progenitor cells (negative for the mature lineage markers Mac1, Gr1, CD3, CD4, CD8, B220, Ter119, and positive for the primitive surface markers c-Kit and Sca1; LKS+) were FACS-sorted from the bone marrow of C57Bl6 wild-type mice and pre-stimulated with 50 ng/ml murine Stem Cell Factor (mSCF; Amgen, Thousand Oaks, CA), 10 ng/ml murine Interleukin 3 (mIL-3; PeproTech), and 50 ng/ml human IL-6 (PeproTech) at 37 °C, 5% CO$_2$ for 48 h to induce cycling. Pre-stimulated cells were subsequently cultured in the above described cytokines with the addition of 8 µg/ml Polybrene (Sigma) on retronectin-coated plates (Takara Bioscience, Clontech) pre-loaded with viral supernatant (1–2 hits), and GFP$^+$ cells were isolated after an additional 48 h (n = 6 independent transductions).

Mice were obtained from Animal Resources Centre, Perth, Australia and kept at BioResources Centre, St Vincent's Hospital, Australia. All experiments were performed under ethical approval of St Vincent's Health Melbourne Animal Ethics Committee.

cDNA microarray

Data normalization and analysis

Total RNA was extracted using Trizol followed by RNeasy Micro kit (Qiagen). RNA was hybridized to the Illumina Mouse WG-6 v2.0 Expression BeadChip at the Ramaciotti Centre for Gene Function Analysis at the University of New South Wales, Australia. Illumina Data was processed in GenePattern (GP) [3] using IlluminaExpressionFileCreator (GP, v2.1) with background correction using manifest MouseWG-6_V2_0_R3_11278593_A.txt (GPL6887). Detection p-values were calculated as the proportion of negative control probes with expression greater than the regular probe in question using the calculateDetected function in the bioconductor bead array package (v2.8.1) [4]. The data was log transformed in GenePattern (logTransform, GP, v3.1) and quantile normalized using NormalizeColulms (GP, v4.2.1).

Filtering

Filtering of the data to eliminate "bad probes" as defined by Barbosa-Morais NL et al. [5] resulted in the removal of 6324 probes, leaving 38,958 probes. Furthermore, probes that were not detected in at least 5 out of 6 SKI or MIG samples were discarded (75%), leaving 9780 probes (7514 unique gene symbols). The filtered data is available as supplementary file "SKI_normalized_filtered.gct".
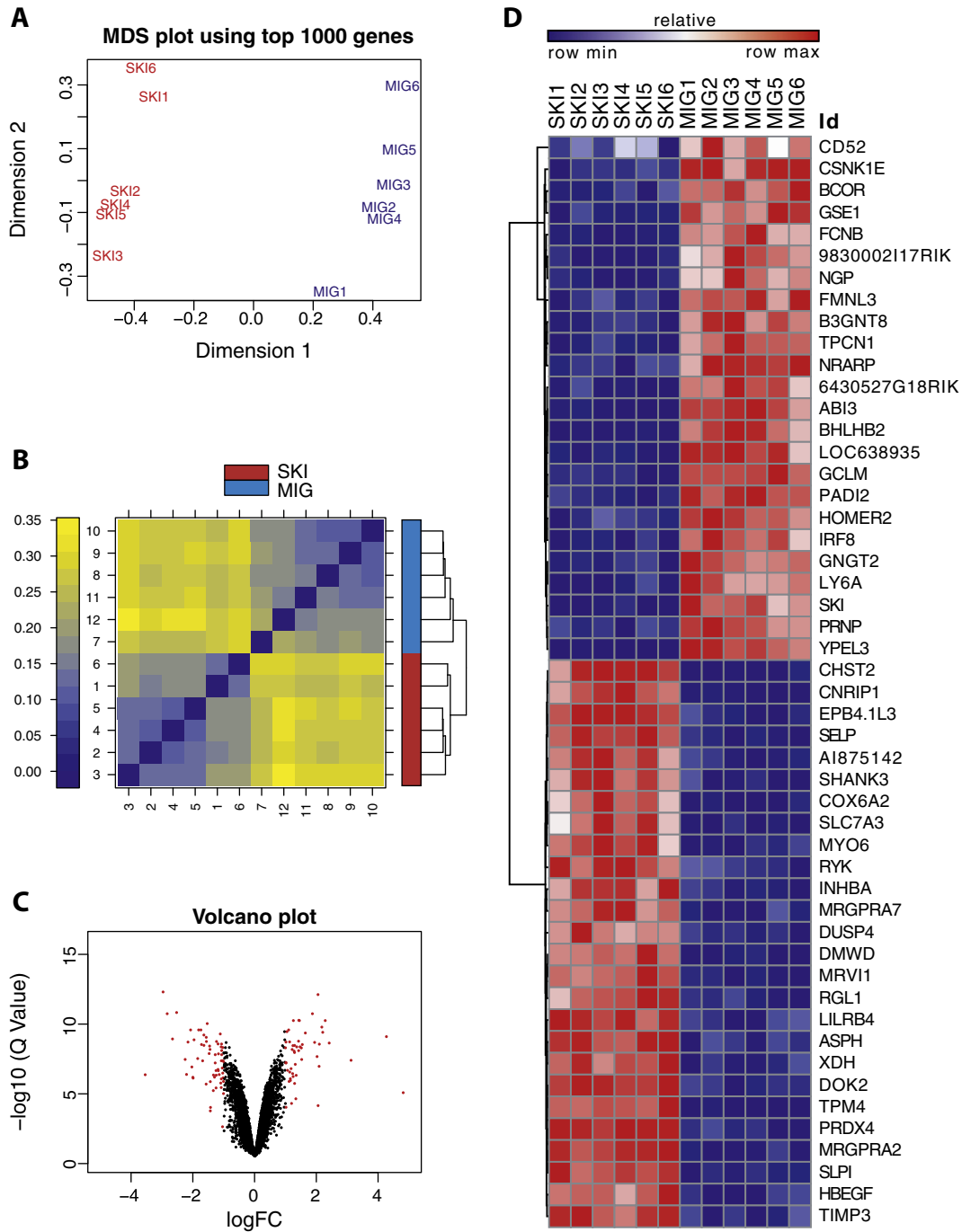

Fig. 1. Microarray data overview. A) MDS plot of samples demonstrating a clear separation between the SKI and MIG samples. B) A dissimilarity matrix of the normalized arrays (blue: high similarity, yellow: low similarity) reflecting the distance between each pair of arrays. C) Volcano plot showing the distribution of the gene expression changes. Genes with absolute logFC > 1 and Q value < 0.05 are indicated in red. D) Heat map of the top 50 genes deregulated by SKI overexpression.

Quality control

In order to assess microarray quality we ran ArrayQualityMetrics [6]. All microarrays passed quality control and the output is available as supplementary files "ArrayQualityMetrics". There was a clear separation between the SKI and MIG samples, as depicted by an MDS plot in Fig. 1A. The dissimilarity matrix of the normalized arrays (Fig. 1B, blue: high similarity, yellow: low similarity) reflects the distance between each pair of arrays — the mean absolute difference between the array data.

Differential expression

Differentially expressed genes were determined using the LimmaGP module [7] with the most informative probe for each gene used (supplementary files SKI_limmaGP_SKIvsMIG.odf, SKI_limmaGP_SKIvsMIG.rnk, SKI_limmaGP_SKIvsMIG.xls, SKI_limmaGP_list.Rda). The distribution of the fold change and q-value is shown in Fig. 1C. Using a cutoff of absolute logFC > 1 and Q.value < 0.05, 113 genes were considered differentially expressed, 51 genes up-regulated and 62 genes down-regulated in response to SKI overexpression. A heat map of the top 50 genes is shown in Fig. 1D, generated using GENE-E [8].

Pathway analysis using GSEA

We performed Gene Set Enrichment Analysis (GSEA) [9] using GSEAPreRanked against the gene set database C2 (c2.all.v3.0.symbols.gmt, curated gene sets). We used the ranked gene list from LimmaGP as input on all "good" probes with the number of permutations = 1000. The Gene Pattern modules LimmaGP and GSEAPreRanked were developed by Mark Cowley (m.cowley@garvan.org.au) from Peter Wills Bioinformatics Centre, Garvan Institute of Medical Research, Sydney Australia.

Validation of gene expression with quantitative real time PCR

The expression of a cohort of genes that were up- or down-regulated by SKI in the microarray was validated using real time PCR on transduced FACS sorted LKS cells from 3 transductions separate from those used for the array.

Conclusion

Collectively our results demonstrate that enforced expression of SKI results in activation of hematopoietic stem cell transcriptional programs and a hematopoietic stem cell autonomous repopulating advantage that favors myeloid development from an early progenitor stage, ultimately resulting in a myeloproliferative disorder.

Disclosures

None of the authors declare any competing financial interests.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2014.06.022.

References

[1] S. Singbrant, M. Wall, J. Moody, G. Karlsson, A.M. Chalk, B. Liddicoat, et al., The SKI proto-oncogene enhances the in vivo repopulation of hematopoietic stem cells and causes myeloproliferative disease. Haematologica 99 (4) (2014) 647–655, http://dx.doi.org/10.3324/haematol.2013.093971 PMID: 24415629.

[2] S. Akiyoshi, H. Inoue, J. Hanai, K. Kusanagi, N. Nemoto, K. Miyazono, et al., c-Ski acts as a transcriptional co-repressor in transforming growth factor-beta signaling through interaction with smads. J. Biol. Chem. 274 (49) (1999) 35269–35277.

[3] M. Reich, T. Liefeld, J. Gould, J. Lerner, P. Tamayo, J.P. Mesirov, GenePattern 2.0. Nat. Genet. 38 (5) (2006) 500–501.

[4] M.J. Dunning, M.L. Smith, M.E. Ritchie, S. Tavare, BeadArray: R classes and methods for Illumina bead-based data. Bioinformatics 23 (16) (2007) 2183–2184.

[5] N.L. Barbosa-Morais, M.J. Dunning, S.A. Samarajiwa, J.F. Darot, M.E. Ritchie, A.G. Lynch, et al., A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data. Nucleic Acids Res. 38 (3) (2010) e17.

[6] A. Kauffmann, R. Gentleman, W. Huber, ArrayQualityMetrics—a bioconductor package for quality assessment of microarray data. Bioinformatics 25 (3) (2009) 415–416.

[7] J.M. Wettenhall, G.K. Smyth, LimmaGUI: a graphical user interface for linear modeling of microarray data. Bioinformatics 20 (18) (2004) 3705–3706.

[8] GENE-E, [Internet]. [cited 16th of June 2014]. Available from: http://www.broadinstitute.org/cancer/software/GENE-E/index.html .

[9] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. 102 (43) (2005) 15545–15550.