MiTCR modules are checked by more than 80 comprehensive unit tests, which improved the reliability and correctness of the code. The MiTCR API package can be used in Java projects through Maven and in Groovy scripts using Groovy Grapes. MiTCR is regularly updated; Windows installer, cross-platform binaries and source code are available from http://mitcr.milaboratory.com/ under the terms of the GNU GPL v3 license.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper (doi:10.1038/nmeth.2555).

ACKNOWLEDGMENTS
We are grateful to T. Schumacher and C. Linnemann for valuable technical discussions and to M. Eisenstein for English editing. This work was supported by the Molecular and Cell Biology program of the Russian Academy of Sciences and Russian Foundation for Basic Research (12-04-33139-mol-a, 12-04-33065-mol-a, 12-04-00229-a, 13-04-00998-a and 13-04-40251).

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Dmitriy A Bolotin[1,3], Mikhail Shugay[1,3], Ilgar Z Mamedov[1], Ekaterina V Putintseva[1], Maria A Turchaninova[1], Ivan V Zvyagin[1,2], Olga V Britanova[1] & Dmitriy M Chudakov[1,2]

[1]Shemiakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia. [2]Central European Institute of Technology, Masaryk University, Brno, Czech Republic. [3]These authors contributed equally to this work.
e-mail: chudakovdm@mail.ru

PUBLISHED ONLINE 28 JULY 2013; DOI:10.1038/NMETH.2555

1. Nguyen, P. *et al. BMC Genomics* **12**, 106 (2011).
2. Robins, H.S. *et al. Blood* **114**, 4099–4107 (2009).
3. Venturi, V. *et al. J. Immunol.* **186**, 4285–4294 (2011).
4. Warren, R.L. *et al. Genome Res.* **21**, 790–797 (2011).
5. Bolotin, D.A. *et al. Eur. J. Immunol.* **42**, 3073–3083 (2012).
6. Britanova, O.V. *et al. Bone Marrow Transplant.* **47**, 1479–1481 (2012).

ESTOOLS Data@Hand: human stem cell gene expression resource

To the Editor: We developed ESTOOLS Data@Hand, a resource to facilitate exploration of published gene expression array data in stem cell research. The resource, updated four times a year, offers efficient sample identification, preprocessing that enables cross-experiment comparisons and computational analysis.

High-throughput studies provide large amounts of data on human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hiPSCs), their parent cells as well as their differentiated progeny cells[1]. Published genome-wide expression data on stem cells can be exploited to answer questions other than those addressed in the original studies. This is one reason why such data are actively stored in the public repositories ArrayExpress[2] and Gene Expression Omnibus (GEO)[3]. Nevertheless, the means to perform such reanalyses are currently limited. Often, sample information is only available as free text in publications or public databases, hindering identification of appropriate samples. Moreover, the measurement data are often available in heterogeneous formats, and the lack of systematic preprocessing hampers cross-experiment comparisons. Some databases have been developed for stem cell research (**Supplementary Note 1** and **Supplementary Table 1**), but none provide both a wide range of human pluripotent stem cell gene expression data and typical analysis tools.

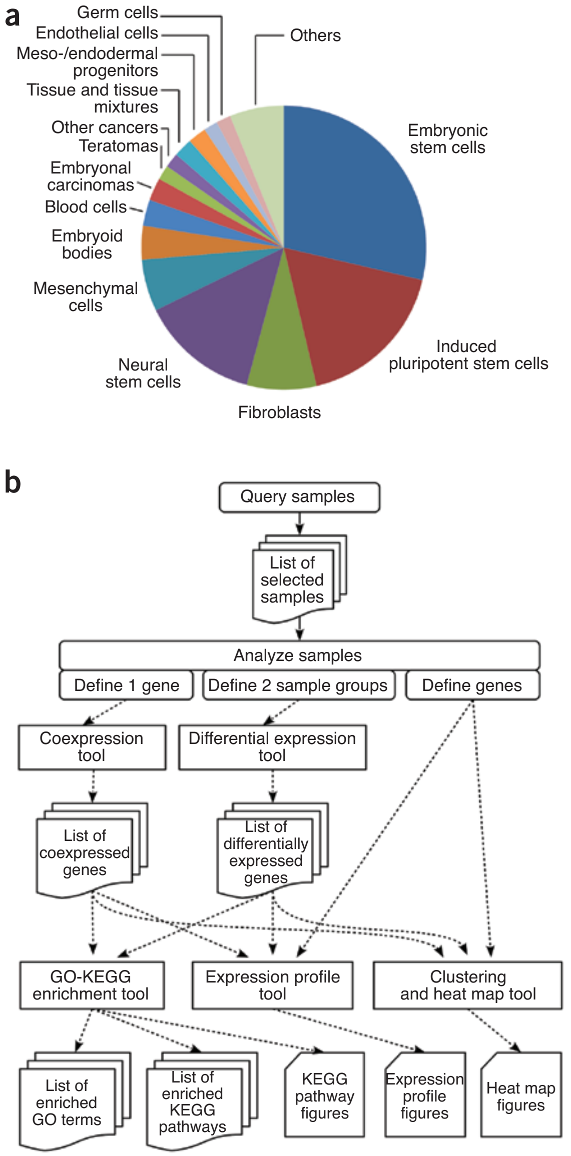

Figure 1 | Gene expression data and analysis workflows in ESTOOLS Data@Hand. (**a**) Cell and tissue types of the samples. (**b**) The data analysis steps available on the user interface.

ESTOOLS Data@Hand regroups gene expression array data and annotations primarily from experiments including hESCs or hiPSCs and thus involves stem cell pluripotency, differentiation and cell dedifferentiation. We selected data from GEO and ArrayExpress manually, preferentially including large experimental series; the selection covers pluripotent cells as well as dozens of other cell and tissue types reported in the same studies (**Fig. 1a**, **Supplementary Methods** and **Supplementary Tables** 2–4). Meta-analysis, a statistical approach to combine results from independent but related studies is a relatively inexpensive option that has the potential to increase both the statistical power and generalizability of single-study analysis[4]. We established two sample meta-datasets of 408 and 245 jointly normalized samples using the two most common array types for this data collection, Affymetrix and Illumina

(**Supplementary Methods**). These meta-datasets are updated when more data are added to the database, and they represent the biggest available collection of expression data of pluripotent stem cells and cell differentiation, to our knowledge. We preprocessed the data to summarize probe-wise measurements as gene-wise expression values, linking the data points to the Ensembl gene and to the standard gene symbols according to the HUGO Gene Nomenclature Committee recommendations. A single expression value for each gene enables more straightforward analyses and comparisons between different sample sets[4]. We curated all sample annotations manually based on information in the literature or obtained from the authors (**Supplementary Methods** and **Supplementary Table 5**). Annotations in Data@Hand extend well beyond those provided by GEO and ArrayExpress, and abide to a strict and homogenous formalism (**Supplementary Note 2**). Thus, a special value of this resource is in these more than 60 annotation dimensions on the characteristics of the measured cells, treatments, culture media, sample processing, platforms and citations.

The graphical user interface of this web resource guides the user for data access and analysis. One can access samples and sample sets by any sample annotation dimension, visualize or download annotations and view a summary report for each sample set. Data querying, selection and analysis tools can be organized into consecutive, easy-to-use logical workflow steps (**Fig. 1b**).

Any user-selected samples within a jointly normalized set of samples can be analyzed with several R Bioconductor-based[5] computational tools (**Supplementary Methods**). The differential expression tool identifies genes, whose expression differs between two sample groups. The coexpression tool finds genes whose expression profiles correlate with that of a selected gene across a sample group. The clustering tool sorts the samples or user-provided genes within a selected sample group by expression similarity and presents the results and expression levels as a dendrogram and a heat map. The expression profile tool plots the expression levels of genes of interest within a collected sample group. The enrichment tool maps genes to Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. In all visualizations, the samples can be arranged according to any annotation, enabling various approaches to data interpretation.

To validate the resource, we used the Affymetrix sample meta-dataset using two approaches to acquire lists of pluripotency marker genes (**Supplementary Methods**). We applied the differential expression tool to find differences between hESCs and cells in three selected clusters of differentiated cells (**Supplementary Fig. 1**), revealing 74 genes with greater expression (fold change > 8, adjusted $P < 0.001$) in hESCs in all three comparisons. Applying the coexpression tool, we found 73 genes, the expression of which correlates with that of the pluripotency marker gene *LIN28A* (correlation > 0.85, $P < 0.001$). An enrichment analysis associated the revealed genes mainly with developmental processes and gene expression regulation ($P < 0.05$) (**Supplementary Tables 6–9**). The two approaches found altogether 117 genes (31 shared) highly overlapping with a previous hESC signature gene set[6] (**Supplementary Fig. 2** and **Supplementary Tables 6, 8** and **10**). The expression of these genes in the Affymetrix sample meta-dataset and in another sample set clearly correlated with the degree of differentiation (**Supplementary Figs. 3–7**). Thus the resource both reproduced earlier results and generated new hypotheses.

ESTOOLS Data@Hand (**Supplementary Fig. 8**; http://estools.cs.tut.fi/, freely available for noncommercial research) should be a powerful one-stop site for integrative data reanalysis in stem-cell research.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper (doi:10.1038/nmeth.2576)

ACKNOWLEDGMENTS
See **Supplementary Methods.**

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Lingjia Kong[1,2,6], Kaisa-Leena Aho[1,6], Kirsi Granberg[1,6], Riikka Lund[2,6], Laura Järvenpää[1], Janne Seppälä[1], Paul Gokhale[3], Kalle Leinonen[1], Lauri Hahne[1,2], Jarno Mäkelä[1], Kirsti Laurila[1], Heidi Pukkila[1], Elisa Närvä[2], Olli Yli-Harja[1,4], Peter W Andrews[3], Matti Nykter[1], Riitta Lahesmaa[2], Christophe Roos[1,5] & Reija Autio[1,2]

[1]Department of Signal Processing, Tampere University of Technology, Tampere, Finland. [2]Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, Turku, Finland. [3]Centre for Stem Cell Biology and the Department of Biomedical Science, University of Sheffield, Western Bank, Sheffield, UK. [4]Institute for Systems Biology, Seattle, Washington, USA. [5]Present address: Euformatics Oy, Espoo, Finland. [6]These authors contributed equally to this work.
e-mail: j.ch.roos@gmail.com

1. Young, R.A. *Cell* **144**, 940–954 (2011).
2. Parkinson, H. *et al. Nucleic Acids Res.* **39**, D1002–D1004 (2011).
3. Barrett, T. *et al. Nucleic Acids Res.* **39**, D1005–D1010 (2011).
4. Ramasamy, A., Mondry, A., Holmes, C.C. & Altman, D.G. *PLoS Med.* **5**, e184 (2008).
5. Gentleman, R.C. *et al. Genome Biol.* **5**, R80 (2004).
6. Lowry, W.E. *et al. Proc. Natl. Acad. Sci. USA* **105**, 2883–2888 (2008).